

Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz



Davi da Silveira Barroso Alves

**Mineração de dados na identificação de padrões de mortalidade no Brasil de 1979 a
2013**

Rio de Janeiro

2017

Davi da Silveira Barroso Alves

Mineração de dados na identificação de padrões de mortalidade no Brasil de 1979 a 2013

Tese apresentada ao Programa de Pós-graduação em Epidemiologia em Saúde Pública, da Escola Nacional de Saúde Pública Sergio Arouca, na Fundação Oswaldo Cruz, como requisito parcial para obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. Oswaldo Gonçalves Cruz.

Coorientadora: Prof.^a Dra. Enirtes Caetano Prates Melo.

Rio de Janeiro

2017

Catálogo na fonte
Fundação Oswaldo Cruz
Instituto de Comunicação e Informação Científica e Tecnológica em Saúde
Biblioteca de Saúde Pública

A474m Alves, Davi da Silveira Barroso.
Mineração de dados na identificação de padrões de mortalidade no Brasil de 1979 a 2013 / Davi da Silveira Barroso Alves. – 2017.
209 f. : il. color. ; graf. ; tab.

Orientador: Oswaldo Gonçalves Cruz.
Coorientadora: Enirtes Caetano Prates Melo.
Tese (doutorado) – Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública Sergio Arouca, Rio de Janeiro, 2017.

1. Mineração de Dados. 2. Aprendizado de Máquina.
3. Transição Epidemiológica. 4. Mortalidade. I. Título.

CDD – 22.ed. – 304.64

Davi da Silveira Barroso Alves

Mineração de dados na identificação de padrões de mortalidade no Brasil de 1979 a 2013

Tese apresentada ao Programa de Pós-graduação em Epidemiologia em Saúde Pública, da Escola Nacional de Saúde Pública Sergio Arouca, na Fundação Oswaldo Cruz, como requisito parcial para obtenção do título de Doutor em Ciências.

Aprovada em: 24 de outubro de 2017

Banca Examinadora

Prof.^a Dra. Cláudia Medina Coeli, UFRJ

Prof.^a Dra. Maria Tereza Serrano Barbosa, Universidade Federal do Estado do Rio de Janeiro

Prof.^a Dra. Luciane de Souza Velasque, Universidade Federal do Estado do Rio de Janeiro

Prof. Dr. Joaquim Gonçalves Valente, ENSP/FIOCRUZ

Prof. Dr. Oswaldo Gonçalves Cruz, PROCC/FIOCRUZ

Rio de Janeiro
2017

Este estudo é dedicado à minha Família.

Amo Vocês!

AGRADECIMENTOS

Ao Oswaldo pela ideia que deu origem a este estudo e por toda atenção e dedicação, refletidas em muitas horas de auxílio para lidar com os diversos desafios técnicos e metodológicos para o desenvolvimento das análises aqui presentes. Agradeço pela oportunidade de ter sido seu orientando e pela generosidade em compartilhar sua genialidade, fundamental para tornar esta Tese possível.

À Enirtes pelo carinho, apoio e paciência, por ter aceito o desafio de me orientar e por todo o suporte nos momentos críticos desta jornada. Agradeço pelas diversas correções de rumo que permitiram organizar e estruturar o caos das minhas ideias e pela dedicação na reta final. Sua ajuda foi determinante para a conclusão de mais esta etapa de minha formação.

Às professoras Cláudia Medina Coeli e Maria Tereza Serrano Barbosa pelas contribuições durante a qualificação que foram decisivas para o direcionamento do estudo.

Ao professor Claudio José Struchiner por generosamente permitir a utilização da máquina na qual parte das análises desta Tese foram implementadas.

À Raíla, amiga de todas as horas, por continuar me aturando, pela atenção e todo o incentivo na reta final, quando pensei que a inspiração havia acabado. Muito obrigado pelas ideias, boa vontade, dedicação e disponibilidade.

Aos meus pais, Clivilaine e Plínio, pelo apoio incondicional, pela paciência e por compreenderem a ausência cada vez mais constante. Não tenho palavras para expressar a gratidão por tudo que sempre fizeram por mim.

Ao meu irmão, Daniel, minha fonte de inspiração para a vida.

Ao meu primo e amigo Thiago, por ter me ajudado a montar o *desktop* utilizado em todo pré-processamento e parte do processamento de dados desta Tese, e pela solicitude em diversos momentos cruciais nestes últimos quatro anos.

À Ediene, pelo companheirismo, carinho, paciência, amor e compreensão. Sua presença e apoio foram decisivos para superar os momentos mais difíceis nos últimos três anos e meio.

A todos os amigos que fizeram parte desta caminhada, tornando-a mais aprazível e descontraída, em especial Adriano, Eduardo, Gabriel, Rodrigo e Pedro.

Você vê, mas não observa.

DOYLE, 1928

RESUMO

Nos últimos 60 anos observou-se no Brasil um processo de transição demográfica e epidemiológica caracterizado por desigualdades regionais que proporcionou o aparecimento de padrões de mortalidade, gerando demandas de saúde específicas e representando um desafio para a saúde pública. As informações sobre mortalidade somam mais de 32 milhões de óbitos registrados no Sistema de Informação sobre e Mortalidade (SIM) que geralmente são analisados de maneira agregada e fragmentada. A mineração de dados é o instrumental preferencial para a identificação de padrões em grandes massas de dados e sua implementação junto ao grande volume de informações do SIM pode ajudar na compreensão do panorama amplo da mortalidade no país, subsidiando o desenvolvimento de políticas de saúde pública equitativas que possibilitem uma gestão mais efetiva do Sistema Único de Saúde. O objetivo desta tese é analisar padrões regionais e temporais da mortalidade no Brasil no período entre 1979 e 2013. Trata-se de um estudo ecológico utilizando dados do SIM, do Censo e das Estimativas Populacionais do IBGE. As análises foram realizadas em três etapas, cada uma abordando diferentes indicadores de mortalidade, unidades de observação, períodos e métodos de análise. Na primeira foi analisada a evolução da mortalidade proporcional e do padrão etário de óbitos nas regiões, e das taxas de mortalidade segundo grupos de causa e causas selecionadas nas unidades da federação. Na segunda foi analisada a evolução das taxas de mortalidade segundo causas selecionadas e dos indicadores sociais, econômicos e demográficos, nas microrregiões e a correlação entre eles. Na terceira foram implementadas técnicas de mineração de dados nas informações individuais do SIM. Os resultados mostram tendências regionais que representam a continuidade da transição da mortalidade iniciada na década de 1940, com Norte e Nordeste atingindo nos últimos 30 anos panorama similar ao observado no Sul e Sudeste entre as décadas de 1960 e 1980. Destaca-se a redução das diferenças regionais das taxas de mortalidade por doenças cardiovasculares e neoplasias que diminuíram no Sul e Sudeste, onde foram mais altas e aumentaram no Norte e Nordeste, onde foram mais baixas. Os achados da mineração de dados com as informações individuais do SIM foram prejudicados pela baixa disponibilidade de variáveis e não representam novidade sobre o tema, contudo, são condizentes com a literatura, o que aponta para o potencial de sua utilização em estudos futuros.

Palavras-chave: Mineração de dados, Aprendizado de Máquina, Transição Epidemiológica, Mortalidade

ABSTRACT

In the last 60 years a demographic and epidemiological transition process has been observed in Brazil, resulting in mortality patterns that generate specific health demands that are challenges to public health. Mortality information accounts for more than 32 million registered deaths in the Mortality Information System (SIM), generally analyzed in aggregate and fragmented approaches. Data mining is the preferred tool to identify patterns in big data sets and the implementation of these techniques in the large information volume from SIM can help to understand the big picture of mortality in Brazil, guiding equity health policies development and allowing effective management of SUS. The objective of this thesis is to analyze regional and temporal patterns of mortality in Brazil between 1979 and 2013. This is an Ecological study using data from the SIM, Census and Population Estimates of IBGE. The analysis was performed in three stages, addressing different mortality indicators, observation units, periods and analysis methods. The first one was the proportional mortality evolution and the death age in regions, the mortality rates by groups and selected causes in the federation units. The second one was the mortality rates evolution by selected causes and social, economic and demographic indicators by microregions and the correlation between them. In the third, data mining techniques were implemented in SIM individual information. The results point to regional trends that represent the continuity of mortality transition started in the 1940s, with the north and northeast regions reaching, in the last 30 years, a similar scenario with the one observed in the south and southeast regions in the 1960s to the 1980s. Regional differences in mortality rates by cardiovascular diseases and neoplasias decreased, reducing in the South and Southeast, where it was higher, and increasing in the North and Northeast, where it was lower. Individual information data mining of the SIM was complicated by the availability of variables and the results of that approach do not represent a big innovation about the subject, however, were consistent with the literature, suggesting a potential use in future studies.

Palavras-chave: Data Mining, Machine Learning, Health Transition, Mortality

LISTA DE FIGURAS

Figura 1 – Exemplo de representação gráfica de uma árvore de classificação e suas estruturas	38
Figura 2 – Exemplo da representação gráfica de uma Análise de Correspondência Múltipla.	41
Figura 3 – Exemplo de Diagrama de Kiviat utilizado para representação da mortalidade proporcional segundo grupos de causa nas regiões do Brasil.....	42
Figura 4 – Exemplo de <i>heatmap</i> utilizado para descrição do padrão espaço-temporal das taxas de mortalidade nas Unidades da Federação no período estudado.....	44
Figura 5 – Estratégia para estruturação do Banco de Dados.....	53
Figura 6 – Esquematização dos procedimentos para estruturação do Banco de Dados.....	54
Figura 7 – Detalhamento do preparo dos dados e para construção dos Fluxogramas do ranking da mortalidade proporcional segundo grupos de causa.....	61
Figura 8 – Detalhamento do preparo dos dados para construção dos Diagramas de Kiviat.....	62
Figura 9 – Organização dos dados para análise do padrão etário de óbito segundo grupo de causa no Brasil e Regiões.....	63
Figura 10 – Procedimentos para obtenção das taxas padronizadas de mortalidade e um exemplo de <i>heatmap</i>	64
Figura 11 – Procedimentos para obtenção das taxas de mortalidade segundo sexo padronizadas por faixa etária e dos indicadores sociais, econômicos e demográficos nas microrregiões em 1991, 2000 e 2010.....	66
Figura 12 – Dados utilizados em cada abordagem da Etapa III.....	68
Figura 13 – Esquematização das Análises Realizadas na Etapa III.....	71
Figura 14 – Distribuição da Mortalidade Proporcional Segundo Grupo de Causa no Brasil entre 1979 e 2013.....	74
Figura 15 – Distribuição da Mortalidade Proporcional Segundo Grupo de Causa na Região Norte entre 1979 e 2013.....	74
Figura 16 – Distribuição da Mortalidade Proporcional Segundo Grupo de Causa na Região Nordeste entre 1979 e 2013.....	75
Figura 17 – Distribuição da Mortalidade Proporcional Segundo Grupo de Causa na Região Sudeste entre 1979 e 2013.....	76
Figura 18 – Distribuição da Mortalidade Proporcional Segundo Grupo de Causa na Região Sul entre 1979 e 2013.....	76
Figura 19 – Distribuição da Mortalidade Proporcional Segundo Grupo de Causa na Região Centro-Oeste entre 1979 e 2013.....	77
Figura 20 – Diagramas de Kiviat da Mortalidade Proporcional por Grupos de Causa selecionados segundo Região nos períodos de 1979 a 1984, 1985 a 1989, 1990 a 1995, 1996 a 1999, 2000 a 2004 e 2005 a 2013.....	79
Figura 21 – Evolução da distribuição da idade dos óbitos em geral no Brasil e Regiões de 1979 a 2013.....	81

Figura 22 – Evolução da distribuição da idade dos óbitos por Doenças Infecciosas e Parasitárias no Brasil e Regiões de 1979 a 2013.....	83
Figura 23 – Evolução da distribuição da idade dos óbitos por Doenças do Aparelho Respiratório no Brasil e Regiões de 1979 a 2013.....	84
Figura 24 – Evolução da distribuição da idade dos óbitos por Neoplasias no Brasil e Regiões de 1979 a 2013.....	86
Figura 25 – Evolução da distribuição da idade dos óbitos por Doenças do Aparelho Circulatório no Brasil e Regiões de 1979 a 2013.....	87
Figura 26 – Evolução da distribuição da idade dos óbitos por Causas Externas no Brasil e Regiões de 1979 a 2013.....	88
Figura 27 – Evolução da Taxa de Mortalidade por Doenças Infecciosas e Parasitárias segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012.....	90
Figura 28 – Evolução da Taxa de Mortalidade por Doenças Infecciosas Intestinais segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012.....	91
Figura 29 – Evolução da Taxa de Mortalidade por Neoplasias segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012.....	93
Figura 30 – Evolução da Taxa de Mortalidade por Neoplasias de Pulmão segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012.....	94
Figura 31 – Evolução da Taxa de Mortalidade por Doenças do Aparelho Circulatório segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012.....	96
Figura 32 – Evolução da Taxa de Mortalidade por Doenças Cerebrovasculares segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012.....	97
Figura 33 – Evolução da Taxa de Mortalidade por Doenças do Aparelho Respiratório segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012.....	99
Figura 34 – Evolução da Taxa de Mortalidade por Pneumonia segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012.....	100
Figura 35 – Evolução da Taxa de Mortalidade por Causas Externas segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012.....	103
Figura 36 – Evolução da Taxa de Mortalidade por Homicídios segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012.....	104
Figura 37 – <i>Boxplots</i> das Taxas de Mortalidade por Doenças Infecciosas Intestinais no Sexo Feminino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010.....	106
Figura 38 – <i>Boxplots</i> das Taxas de Mortalidade por Doenças Infecciosas Intestinais no Sexo Masculino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010.....	106
Figura 39 – <i>Boxplots</i> das Taxas de Mortalidade por Neoplasias de Pulmão no Sexo Feminino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010.....	107
Figura 40 – <i>Boxplots</i> das Taxas de Mortalidade por Doenças Cerebrovasculares no Sexo Feminino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010.....	107
Figura 41 – <i>Boxplots</i> das Taxas de Mortalidade por Doenças Cerebrovasculares no Sexo Feminino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010.....	108
Figura 42 – <i>Boxplots</i> das Taxas de Mortalidade por Doenças Cerebrovasculares no Sexo Masculino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010.....	108

Figura 43 – <i>Boxplots</i> das Taxas de Mortalidade por Pneumonia no Sexo Feminino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010.....	109
Figura 44 – <i>Boxplots</i> das Taxas de Mortalidade por Pneumonia no Sexo Masculino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010.....	110
Figura 45 – <i>Boxplots</i> das Taxas de Mortalidade por Homicídios no Sexo Feminino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010.....	110
Figura 46 – <i>Boxplots</i> das Taxas de Mortalidade por Homicídios no Sexo Masculino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010.....	111
Figura 47 – <i>Boxplots</i> da Renda Per Capita Média segundo Região em 1991, 2000 e 2010...	113
Figura 48 – <i>Boxplots</i> da Razão de Dependência de Idosos segundo Região em 1991, 2000 e 2010.....	113
Figura 49 – <i>Boxplots</i> da Taxa de Desemprego segundo Região em 1991, 2000 e 2010.....	114
Figura 50 – <i>Boxplots</i> do Percentual de Domicílios com Acesso à Rede Geral de Esgoto ou Pluvial segundo Região em 1991, 2000 e 2010.....	115
Figura 51 – Coeficientes de Correlação entre as Taxas de Mortalidade por Causas Seleccionadas e os Indicadores Sociais Econômicos e Demográficos Seleccionados em 1991, 2000 e 2010.....	116
Figura 52 – Árvore de Classificação obtida pelo modelo <i>CART</i> Abordagem 1: Mortalidade Geral Segundo Grupos de Causa.....	118
Figura 53 – Decréscimo Médio do Índice de Impureza de Gini das Variáveis Independentes na Abordagem 1 obtidas pelo <i>Random Forest</i>	119
Figura 54 – Análise de Correspondência Múltipla na Abordagem 1 com Grupos de Causa de Óbito, Local de Óbito, Faixa Etária e Unidade da Federação.....	120
Figura 55 – Análise de Correspondência Múltipla na Abordagem 1 com Grupos de Causa de Óbito, Local de Óbito e Faixa Etária.....	120
Figura 56 – Árvore de Classificação obtida pelo modelo <i>CART</i> na Abordagem 2: Doenças Infecciosas e Parasitárias.....	121
Figura 57 – Decréscimo Médio do Índice de Impureza de Gini das Variáveis Independentes na Abordagem 2 obtidas pelo <i>Random Forest</i>	123
Figura 58 – Análise de Correspondência Múltipla na Abordagem 2 com Causa Específica de Óbito, Faixa Etária, Ano de Óbito e Unidade da Federação.....	123
Figura 59 – Decréscimo Médio do Índice de Impureza de Gini das Variáveis Independentes na Abordagem 3 obtidas pelo <i>Random Forest</i>	124
Figura 60 – Análise de Correspondência Múltipla na Abordagem 3 com Causa Específica de Óbito, Sexo, Faixa Etária e Ano de Óbito.....	125
Figura 61 – Árvore de Classificação obtida pelo modelo <i>CART</i> na Abordagem 4: Doenças do Aparelho Circulatório.....	126
Figura 62 – Decréscimo Médio do Índice de Impureza de Gini das Variáveis Independentes na Abordagem 4 obtidas pelo modelo <i>Random Forest</i>	127
Figura 63 – Análise de Correspondência Múltipla na Abordagem 4 com Causa Específica de Óbito, Sexo, Faixa Etária e Ano de Óbito.....	128

Figura 64 – Árvore de Classificação obtida pelo modelo <i>CART</i> na Abordagem 5: Doenças do Aparelho Respiratório.....	129
Figura 65 – Decréscimo Médio do Índice de Impureza de Gini das Variáveis Independentes na Abordagem 5 obtidas pelo modelo <i>Random Forest</i>	131
Figura 66 – Análise de Correspondência Múltipla na Abordagem 5 com Faixa Etária, Ano de Óbito e Unidade da Federação.....	131
Figura 67 – Árvore de Classificação obtida pelo modelo <i>CART</i> na Abordagem 6: Causas Externas.....	132
Figura 68 – Decréscimo Médio do Índice de Impureza de Gini das Variáveis Independentes na Abordagem 6 obtidas pelo modelo <i>Random Forest</i>	134
Figura 69 – Análise de Correspondência Múltipla na Abordagem 6 com Causa Específica de Óbito, Faixa Etária, Local de ocorrência do Óbito e Unidade da Federação.....	134

LISTA DE QUADROS

Quadro 1 – Resumo das Etapas de desenvolvimento do estudo segundo Indicadores Analisados, Unidade de Observação, Período e Métodos de Análise.....	48
Quadro 2 – Indicadores Sociais, Econômicos e Demográficos a serem utilizados nas análises por microrregião e seus respectivos conceitos e métodos de cálculo.....	51
Quadro 3 – Categorias da variável “Estado Civil” nos anos da CID 9 e CID 10 e compatibilização adotada.....	55
Quadro 4 – Categorias da variável “Local de Ocorrência” nos anos da CID 9 e CID 10 e compatibilização adotada.....	55
Quadro 5 – Categorias das variáveis “Grau de Instrução” e “Escolaridade” nos anos da CID 9 e CID 10 e compatibilização adotada.....	56
Quadro 6 – Compatibilização dos Grupos de Causa na CID-9 e na CID-10.....	56
Quadro 7 – Compatibilização das Causas Específicas de Óbito segundo CID-9 e CID-10.....	57
Quadro 8 – Percentual somado de registros faltantes (NA) e informação “Ignorada” das variáveis.....	59
Quadro 9 – Variáveis mantidas na Tabela Base.....	59
Quadro 10 – Causas Específicas mantidas em cada Abordagem.....	69
Quadro 11 – Categorização das Variáveis utilizadas em cada abordagem.....	70

LISTA DE TABELAS

Tabela 1 – Resultado da Análise Fatorial Exploratória com os Indicadores Sociais, Econômicos e Demográficos em 1991, 2000 e 2010.....	112
Tabela 2 – Medidas de Qualidade de Predição do Modelo <i>CART</i> da Abordagem 1.....	118
Tabela 3 – Medidas de Qualidade de Predição do Modelo <i>CART</i> da Abordagem 2.....	122
Tabela 4 – Medidas de Qualidade de Predição do Modelo <i>CART</i> da Abordagem 4.....	126
Tabela 5 – Medidas de Qualidade de Predição do Modelo <i>CART</i> da Abordagem 5.....	130
Tabela 6 – Medidas de Qualidade de Predição do Modelo <i>CART</i> da Abordagem 6.....	133

SUMÁRIO

1 INTRODUÇÃO	16
2 REVISÃO DE LITERATURA	19
2.1 TRANSIÇÃO DEMOGRÁFICA E EPIDEMIOLÓGICA E O PANORAMA BRASILEIRO.....	19
2.2 DINÂMICA DOS PADRÕES DE MORTALIDADE NO BRASIL.....	26
2.3 <i>BIG DATA, DATA MINING E MACHINE LEARNING – O POTENCIAL DE APLICAÇÃO EM SAÚDE PÚBLICA</i>	31
2.4 O SISTEMA DE INFORMAÇÃO SOBRE MORTALIDADE – SIM.....	35
2.5 MÉTODOS DE MINERAÇÃO PARA GRANDES DADOS E ANÁLISES GRÁFICAS.....	37
2.5.1 <i>MÉTODOS BASEADOS EM ÁRVORES: ÁRVORES DE DECISÃO (CART) E RANDOM FOREST</i>	37
2.5.2 <i>ANÁLISE DE CORRESPONDÊNCIA</i>	40
2.5.3 <i>DIAGRAMAS DE KIVIAT</i>	42
2.5.4 <i>HEATMAPS</i>	43
3 JUSTIFICATIVA	45
4 OBJETIVOS	46
4.1 OBJETIVO GERAL.....	46
4.2 OBJETIVOS ESPECÍFICOS.....	46
5 MATERIAL E MÉTODOS	47
5.1 VARIÁVEIS ANALISADAS E FONTES DE DADOS.....	47
5.2 PROCEDIMENTOS DE ESTRUTURAÇÃO DA BASE DE DADOS E SELEÇÃO DE VARIÁVEIS – PRÉ-PROCESSAMENTO.....	52
5.3 ANÁLISE DE DADOS.....	60
5.3.1 <i>ETAPA I: ANÁLISE DOS PADRÕES E TENDÊNCIAS DE MORTALIDADE SEGUNDO GRUPOS DE CAUSA E CAUSAS SELECIONADAS NO BRASIL</i>	60
5.3.2 <i>ETAPA II: ANÁLISE DA TENDÊNCIA DAS TAXAS DE MORTALIDADE POR CAUSAS SELECIONADAS E IDENTIFICAÇÃO DOS INDICADORES SOCIAIS, ECONÔMICOS E DEMOGRÁFICOS RELACIONADOS</i>	65
5.3.3 <i>ETAPA III: IMPLEMENTAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS NAS INFORMAÇÕES INDIVIDUAIS DO SIM PARA IDENTIFICAÇÃO DE PADRÕES DE MORTALIDADE</i>	67
5.5 CONSIDERAÇÕES ÉTICAS.....	72
6 RESULTADOS	73
6.1 PADRÕES E TENDÊNCIAS DE MORTALIDADE SEGUNDO GRUPOS DE CAUSA.....	73

6.1.1 MORTALIDADE PROPORCIONAL SEGUNDO GRUPOS DE CAUSA NO BRASIL E REGIÕES.....	73
6.1.2 EVOLUÇÃO DO PADRÃO ETÁRIO DOS ÓBITOS SEGUNDO GRUPOS DE CAUSA NO BRASIL E REGIÕES.....	80
6.1.3 EVOLUÇÃO DAS TAXAS DE MORTALIDADE SEGUNDO GRUPO DE CAUSA E CAUSAS SELECIONADAS NAS UNIDADES DA FEDERAÇÃO.....	89
6.2 TENDÊNCIAS DAS TAXAS DE MORTALIDADE POR CAUSAS SELECIONADAS E INDICADORES SOCIAIS, ECONÔMICOS E DEMOGRÁFICOS.....	105
6.2.1 COMPORTAMENTO DAS TAXAS DE MORTALIDADE POR CAUSAS SELECIONADAS.....	105
6.2.2 SELEÇÃO E COMPORTAMENTO DOS INDICADORES SOCIAIS, ECONÔMICOS E DEMOGRÁFICOS.....	111
6.2.3 CORRELAÇÃO ENTRE OS INDICADORES SELECIONADOS E AS TAXAS DE MORTALIDADE.....	115
6.3 MINERAÇÃO DE DADOS DAS INFORMAÇÕES INDIVIDUAIS DO SIM E A IDENTIFICAÇÃO DE PADRÕES DE MORTALIDADE NO BRASIL.....	117
6.3.1 ABORDAGEM 1 – MORTALIDADE GERAL SEGUNDO GRUPOS DE CAUSA.....	117
6.3.2 ABORDAGEM 2 – DOENÇAS INFECCIOSAS E PARASITÁRIAS.....	121
6.3.3 ABORDAGEM 3 – NEOPLASIAS.....	124
6.3.4 ABORDAGEM 4 – DOENÇAS DO APARELHO CIRCULATÓRIO.....	126
6.3.5 ABORDAGEM 5 – DOENÇAS DO APARELHO RESPIRATÓRIO.....	129
6.3.5 ABORDAGEM 6 – CAUSAS EXTERNAS.....	132
7 DISCUSSÃO.....	135
8 CONSIDERAÇÕES FINAIS.....	154
9 REFERÊNCIAS.....	159
APÊNDICE A – AVALIANDO OS REGISTROS DE ÓBITO COM INFORMAÇÃO FALTANTE OU IGNORADA DA ESCOLARIDADE/INSTRUÇÃO.....	177
APÊNDICE B – DISTRIBUIÇÃO DOS ÓBITOS POR CAUSAS ESPECÍFICAS NA ABORDAGEM 3 – NEOPLASIAS DE PULMÃO, NA ABORDAGEM 4 – DOENÇAS DO APARELHO CIRCULATÓRIO E NA ABORDAGEM 6 – CAUSAS EXTERNAS DA ETAPA III SEGUNDO FAIXA ETÁRIA.....	178
APÊNDICE C – DISTRIBUIÇÃO DA PROPORÇÃO DE ÓBITO DE CADA CATEGORIA DE CAUSA ESPECÍFICA SEGUNDO ABORDAGENS DA ETAPA III.....	181
APÊNDICE D – ARTIGO.....	186

1 INTRODUÇÃO

As estatísticas de mortalidade são de fundamental importância para compreensão do perfil epidemiológico de uma população, refletindo sua condição de vida e saúde (DRUMOND JR et al., 1999; LAURENTI; MELLO JORGE; GOTLIEB, 2004a). Estas informações apresentam baixo custo de obtenção, são abrangentes e geralmente cobrem longos períodos de tempo (JOHANSSON, 2008), podendo ser utilizados para definir agendas políticas, direcionar os recursos e ações governamentais, bem como para monitorá-las (ROSENBERG, 1999), de maneira que todo o planejamento, implementação e avaliação de programas de saúde necessitam de informações válidas sobre a mortalidade para que se possa avaliar a natureza e a magnitude dos problemas de saúde (HILL, 2006).

A mortalidade vem diminuindo em um processo intrínseco às transformações demográficas pelas quais diversas populações passaram nos últimos dois séculos. De maneira geral, o decréscimo expressivo da mortalidade por doenças infecciosas e parasitárias impactou a mortalidade infantil e de mulheres jovens, sendo gradualmente substituída por doenças crônicas e degenerativas que passaram a representar as principais causas de morbimortalidade, em um processo conhecido como transição epidemiológica (OMRAN, 1971; LEBRÃO, 2007; MENDES, 2011).

Concomitantemente ocorreu uma transição demográfica, intrinsecamente relacionada à transição epidemiológica, demarcada por dois momentos de equilíbrio populacional entremeados por um período de crescimento acelerado, resultado da redução da mortalidade e da fecundidade em velocidades desiguais. Esta transição é marcada por ganhos na expectativa de vida e aumento da participação proporcional das faixas etárias adulta e idosa em detrimento de faixas etárias mais jovem na população (TEITELBAUM, 1975; KIRK, 1996).

O ritmo no qual estas transições ocorreram é caracterizado por desigualdades inerentes às especificidades das mudanças sociais, culturais e tecnológicas de cada região (PRESTON, 1975; LLERAS-MUNEY; CUTLER; DEATON, 2006; LEE, 2007).

No Brasil o processo de transição demográfica ocorreu de maneira acelerada, tendo início na década de 1940 com a queda na mortalidade, fruto principalmente de avanços médicos e sanitários, que resultou em um período de grande crescimento populacional até o final da década de 1950 e início da década de 1960, quando se iniciou a queda na fecundidade. Este processo se desenvolveu de maneira desigual, estando em um estágio mais avançado nas regiões Sul e Sudeste e Centro-Oeste, quando comparado às regiões Norte e

Nordeste, apesar das mudanças em relação às principais causas de óbito se mostrarem mais agudas nas duas últimas (CARVALHO; RODRÍGUEZ-WONG, 2008; CHAIMOWICZ, 1997; LEBRÃO, 2007; PRATA, 1992; VASCONCELOS; GOMES, 2012).

Numerosos sistemas de informação em saúde foram desenvolvidos e mantidos pelo poder público com finalidades diversas no país. Embora persistam grandes diferenças entre esses sistemas em relação à estrutura dos dados, é possível compatibilizá-los a partir de bases territoriais e grupos populacionais específicos (MELO, 2004).

Uma grande massa de dados referente aos óbitos é produzida e disponibilizada pelo Sistema de Informação sobre a Mortalidade (SIM), somando mais de 32 milhões de registros gerados a partir da Declaração de Óbito (DO) (MINISTÉRIO DA SAÚDE, 2011). Considera-se como pontos positivos deste sistema a universalidade e padronização dos registros que permitem comparar dados de diferentes locais e assegura confiabilidade ao sistema. Como limitações devem ser considerados o sub-registro, preenchimento incorreto ou incompleto, ausência de retorno das DO com erros para a unidade informante, erros de codificação, digitação terceirizada, falta de relatórios de inconsistências e elevado número de óbitos por causas mal definidas em algumas localidades (AMARAL, 2002; SANTOS, 2011).

Informações sobre a mortalidade têm sido analisadas de maneira fragmentada, agregando-as segundo causas, locais, faixas etárias e/ou períodos específicos, mas raramente são abordadas como um todo e não foram encontrados estudos utilizando-as de maneira individualizada, tendo como unidade de observação cada óbito registrado no sistema. Contudo, considerando o cenário de transição epidemiológica do país e as diferentes expressões do contexto social, econômico e demográfico sobre a mortalidade, a saúde pública pode se beneficiar da exploração destas informações individuais sobre óbitos, na identificação de estruturas que apontem padrões e tendências desconhecidos ou na confirmação daqueles já consagrados.

Avanços computacionais recentes, bem como o desenvolvimento e aprimoramento de técnicas de análise possibilitam a abordagem de grandes bases de dados, o que ainda é pouco explorado na área da Saúde Pública (PENTLAND; REID; HEIDBECK, 2013; BORDER, 2014; OLA; SEDIG, 2014). O conjunto de técnicas conhecido como “*data mining*”, mineração de dados em tradução livre, compõem o arsenal preferencial para identificação de padrões em grandes massas de dados. Estes padrões, além de possibilitar a extração de informações a partir do próprio dado, podem ser úteis em predições quando aplicados em novos dados (BERRY; LINOFF, 2004; WITTEN; FRANK; HALL, 2011).

A produção científica abordando a temática da mineração de grandes massas de dados em Saúde Pública é recente e pouco extensa. Nos últimos dez anos poucos estudos analisaram ou descreveram padrões de mortalidade em amplitude nacional, de maneira que a maioria o fizeram como instrumento de análise da carga de doenças, descrição do processo de transição epidemiológica, ou abordando a mortalidade por causas e grupos etários específicos (ISHITANI et al., 2006; SOUZA et al., 2006; SOUZA; LIMA, 2006; BARRETO; CARMO, 2007; LEBRÃO, 2007; CESSÉ et al., 2009; FONSECA; ELUF-NETO; WUNSCH FILHO, 2010; MENDES et al., 2012; VASCONCELOS; GOMES, 2012; LEITE et al., 2015).

Contudo, são escassos os estudos abordando diversas causas de mortalidade simultaneamente, descrevendo padrões e tendências regionais e associando a mortalidade com indicadores sociais, econômicos e demográficos, o que pode oferecer visão ampla do panorama da mortalidade no país e da influência do contexto socioeconômico e demográfico na expressão dos padrões e tendências da mortalidade.

Diferenças marcantes nos padrões de mortalidade podem estar associadas às desigualdades e iniquidades observadas no país. Neste cenário, a mineração da grande massa de dados que compõe o SIM desde sua criação, abordando as informações sobre a mortalidade de diferentes perspectivas, relacionando-a com indicadores sociais, econômicas e demográficas, pode favorecer a identificação de padrões e tendências espaciais e temporais que pode contribuir com o planejamento e gestão do Sistema Único de Saúde (SUS), subsidiando o desenvolvimento de políticas públicas equitativas visando atender as demandas de uma população em constante transformação.

2 REVISÃO DE LITERATURA

2.1 TRANSIÇÃO DEMOGRÁFICA E EPIDEMIOLÓGICA E O PANORAMA BRASILEIRO

A dinâmica de crescimento populacional passou por transformações nos últimos duzentos anos. No início do século XIX a população mundial era composta por um bilhão de pessoas, dobrando por volta de 1920 em um processo que demorou cerca de cento e vinte anos. Este número voltou a dobrar em menos de sessenta anos, por volta de 1975, quando atingiu a marca de 4 bilhões (VAN BAVEL, 2013) e dobrará novamente antes de 2030 (UNITED NATIONS, 2015). Além do aumento no ritmo de crescimento, a população vem sofrendo alterações em relação a sua composição etária. A proporção de jovens de até 14 anos de idade na população mundial decaiu no período entre 1950 e 2015 de 34,3% para 26,1% e tende a diminuir ainda mais, chegando a 17,7% em 2100. Por outro lado a população com 50 ou mais anos de idade, que aumentou de 15,8% para 22,3% entre 1950 e 2015, atingirá a marca de 40,0% em 2100 (UNITED NATIONS, 2015).

Estas mudanças no cenário populacional mundial são conhecidas como Transição Demográfica (VAN BAVEL, 2013) e descritas por um modelo teórico homônimo. A teoria da Transição Demográfica propõe um modelo genérico para descrever as mudanças na população mundial entre três momentos distintos: dois momentos de equilíbrio entremeados por um período de acelerado crescimento populacional, resultado da redução da mortalidade e da fecundidade em velocidade desigual.

De maneira geral, considerando o modelo transicional clássico dos países desenvolvidos da Europa ocidental, quedas na mortalidade, resultado dos avanços na agricultura que possibilitaram a redução da morte por desnutrição e do processo de urbanização que teve grande impacto na mortalidade por diversas doenças infecciosas, permitiram um rápido crescimento populacional devido à elevada fecundidade (NOTESTEIN, 1953; TEITELBAUM, 1975). Contudo, com o aumento na sobrevivência os custos envolvidos para concepção e criação também aumentaram, levando os pais a optarem por investir mais na saúde e bem-estar de um número menor de filhos (NERLOVE, 1974). Além disto, o progresso tecnológico promoveu a valorização do tempo de trabalho, e como cuidar dos filhos estava historicamente associado às obrigações da mãe, com a inserção da mulher no mercado de trabalho ter filhos passou a representar perda da força produtiva (LEE, 2007). O desenvolvimento e popularização de novos e eficientes métodos contraceptivos também

foram importantes na redução da fecundidade (NOTESTEIN, 1953; TEITELBAUM, 1975), que provocou uma desaceleração no crescimento populacional.

Neste ínterim, a massa de indivíduos que nasceu durante as primeiras etapas da transição, resultado do desbalanço inicial entre mortalidade e fecundidade, alcançou idades elevadas e a proporção de adultos e idosos aumentaram. Com o controle das doenças infecciosas novos grupos de doenças, aos quais indivíduos mais velhos são mais suscetíveis, e relacionados ao estilo de vida urbano e industrial ganharam cada vez mais importância na mortalidade da população (LLERAS-MUNEY; CUTLER; DEATON, 2006; LEE, 2007).

Logo, observa-se que o comportamento da mortalidade e a dinâmica populacional estão intimamente ligadas, de maneira que assim como a mortalidade influenciou a transição demográfica, esta transição favoreceu alterações no processo de adoecimento e óbito da população, em um fenômeno descrito por Omran em 1971 como Transição Epidemiológica, com postulados baseados nos modelos e etapas da transição demográfica (OMRAN, 1971):

- i. A mortalidade é um fator fundamental na dinâmica populacional;
- ii. Durante a transição epidemiológica a mortalidade por doenças infecciosas é gradualmente substituída por doenças degenerativas como principal causa de óbito, em processo dividido em três estágios:
 - a) Era da pestilência e da fome: característico do estágio pré transição demográfica, quando a maioria dos óbitos é atribuída a doenças infecciosas, desnutrição e complicações na maternidade. Neste estágio a mortalidade geral é elevada, a expectativa de vida é baixa e o crescimento populacional é praticamente nulo
 - b) Era do retrocesso das pandemias: quando a mortalidade decai progressivamente, com aumento na expectativa de vida, e o crescimento populacional é acelerado. Nesta etapa observa-se o declínio da mortalidade por doenças infecciosas, e ligeiro aumento na mortalidade por câncer e doenças circulatórias
 - c) Era das doenças degenerativas e causadas pelo homem: as doenças degenerativas, junto com as causas externas, se tornam as principais causas de mortalidade, a expectativa de vida aumenta cada vez mais, a mortalidade chega aos valores mais baixos e se estabiliza, enquanto o crescimento populacional passa a ser explicado principalmente pelo nível de fecundidade
- iii. As mudanças nos padrões de saúde e doença durante a transição afetaram principalmente crianças e mulheres jovens;

- iv. As mudanças nos padrões saúde e doenças que caracterizam a transição epidemiológica estão diretamente associadas com a transição demográfica e socioeconômica que fazem parte da modernização;
- v. Variações peculiares no padrão, ritmo, determinantes e consequências das mudanças populacionais diferenciam três modelos básicos de transição epidemiológica:
 - a) Modelo clássico ou ocidental: transição lenta e gradual da mortalidade e fecundidade que acompanhou o processo de modernização na maioria dos países da Europa Ocidental, tendo os fatores socioeconômicos como determinantes primários
 - b) Modelo acelerado: transição rápida da mortalidade, em que a progressão entre a era da pestilência e da fome e a era do retrocesso das epidemias se deu em espaço de tempo mais curto e em um momento histórico posterior ao da transição clássica, caracterizado por um processo de modernização lento no qual a queda na mortalidade foi determinada por avanços médicos, sanitários e por melhorias sociais em geral. Medidas nacionais para o controle do crescimento populacional favoreceram uma rápida queda na fecundidade
 - c) Modelo contemporâneo ou atrasado: típico da maioria dos países em desenvolvimento, onde a queda na mortalidade ocorreu de maneira rápida, determinada principalmente por medidas de saúde pública, mantendo-se níveis elevados de fecundidade que resultou em crescimento populacional muito acelerado. Este modelo é típico de muitos países da América Latina, África e Ásia

Como destacado nos postulados da teoria da transição epidemiológica, as transformações na mortalidade foram tipicamente diferentes nos países desenvolvidos e nos países em desenvolvimento (PRESTON, 1975; TEITELBAUM, 1975; LLERAS-MUNNEY; CUTLER; DEATON, 2006; LEE, 2007), demarcando as desigualdades neste processo. Nos países desenvolvidos a queda da mortalidade ocorreu de maneira gradual, tendo se iniciado ainda no final do século XVIII e início do século XIX, com anos de intervalo entre cada estágio, relacionada aos avanços sociais e econômicos provenientes do desenvolvimento do país de maneira geral e da industrialização (TEITELBAUM, 1975; PRESTON, 2007). A melhoria na nutrição, resultado dos avanços na produção agrícola, junto com o crescimento econômico tiveram importante papel na saúde e na queda da mortalidade, posteriormente a saúde pública ganhou importância nesta redução, sobretudo nas medidas de saneamento e na promoção de práticas de saúde e higiene pessoal, e por último os avanços na medicina,

iniciado com a vacinação e antibióticos, seguida do avanço nas intervenções individuais (LLERAS-MUNEY; CUTLER; DEATON, 2006; LEE, 2007).

Por outro lado, nos países em desenvolvimento esta queda ocorreu de maneira rápida, tendo início no século XX, resultado principalmente de avanços de caráter exógeno, que causaram grande impacto na mortalidade não relacionados com o desenvolvimento do país de maneira geral (ARRIAGA; DAVIS, 1969; TEITELBAUM, 1975; CANNING, 2011), como as ações de saúde pública e os avanços médicos sanitários, como a vacinação, a utilização de antibióticos e melhorias no saneamento (LLERAS-MUNEY; CUTLER; DEATON, 2006; CANNING, 2011).

Neste contexto, o perfil de saúde marcado por grande heterogeneidade em diversos países da América Latina aponta para uma nova experiência de transição que não se encaixa nos modelos propostos por Omran, denominado modelo polarizado prolongado. Este modelo tem como características a superposição de etapas, pois doenças marcantes das diversas etapas da transição coexistem durante longo período de tempo, a contra transição, marcada pelo aumento da importância de algumas doenças infecciosas, a transição prolongada, com manutenção de elevada morbimortalidade por doenças infecciosas e parasitárias apesar do domínio absoluto das causas crônicas e polarização epidemiológica, pois estas características apresentadas afetam de maneiras diferentes grupos sociais e regiões geográficas distintas, agravando as desigualdades em saúde (FRENK et al., 1991). É justamente neste modelo que se encaixa o padrão de transição demográfica e epidemiológica observado no Brasil (CHAIMOWICZ, 1997; ARAÚJO, 2012).

A transição demográfica no país teve início na década de 1940 (BARATA et al., 1997). Até então o país apresentava elevadas taxas de fecundidade e mortalidade, com uma expectativa de vida de 41,53 anos de idade (IBGE, 2010). A partir desta década, a mortalidade sofreu uma queda brusca, sobretudo a mortalidade infantil, enquanto a fecundidade se manteve elevada até a década de 1960, período no qual a população brasileira cresceu rapidamente (PRATA, 1992; BARATA et al., 1997). Segundo dados do IBGE entre 1940 e 1970 a população brasileira aumentou em mais de 100%, de cerca de 40 milhões para cerca de 93 milhões (IBGE, 2007), e a expectativa de vida aumentou em mais de 10 anos de idade (IBGE, 2010).

A partir da década de 1970 observou-se queda aguda na fecundidade e natalidade, com manutenção da queda na taxa de mortalidade, o que gerou uma desaceleração no crescimento populacional e marcou o início do processo de envelhecimento populacional (BARATA et al., 1997). No período entre 1970 e 2010 a proporção de jovens diminuiu gradativamente, enquanto

a de adultos aumentou moderadamente e a idosa praticamente dobrou. O índice de envelhecimento populacional, que corresponde ao número de idosos para cada cem jovens, quadruplicou, a expectativa de vida ao nascer aumentou mais de vinte anos, enquanto a fecundidade diminuiu em quase um terço, ficando abaixo da taxa de reposição, a taxa bruta de natalidade foi reduzida pela metade e a taxa de crescimento populacional alcançou seus menores valores em toda a história (VASCONCELOS; GOMES, 2012).

Estas mudanças ocorreram de maneira desigual no país. Enquanto as regiões Sul e Sudeste experimentaram quedas agudas na mortalidade e ganhos expressivos em esperança de vida ainda na década de 1970, as regiões Norte e Nordeste apresentavam taxas elevadas de mortalidade geral e infantil e os ganhos na esperança de vida nestas regiões foram percebidos tardiamente, a partir das décadas de 1980 e 1990. Contudo, nestas duas últimas regiões, as mudanças destes indicadores foram agudas, de maneira que se aproximaram das demais regiões até 2010 (VASCONCELOS; GOMES, 2012).

Em relação à estrutura etária, a proporção de idosos apresentou aumento igual ou superior a 100% entre 1970 e 2010 em todas as regiões do país, sendo o maior aumento proporcional observado na região Centro-Oeste, onde triplicou de 3% para 9% no período. Em 2010 as regiões Sul e Sudeste apresentavam a mesma distribuição da população por grandes grupos etários, sendo 22% com até quatorze anos, 66% entre quinze e cinquenta e nove anos e 12% com sessenta ou mais anos de idade. A região Norte era única em que indivíduos com menos de quinze anos de idade ainda representavam mais de 30% da população total em 2010, e aquela com a menor proporção de adultos e idosos, 61% e 8%, respectivamente (VASCONCELOS; GOMES, 2012).

As desigualdades também podem ser observadas pela complexa relação intergeracional no país (BRITO, 2008). A Razão de Dependência, que indica o número de indivíduos em idade inativa para cada 100 indivíduos em idade ativa, diminuiu em todas as regiões no período entre 1970 e 2010, um indicativo do aumento generalizado na proporção de adultos. Contudo, enquanto na região Norte em 2010, a relação era de um jovem para cada dois indivíduos em idade ativa e um idoso para cada oito indivíduos em idade ativa, na região Sul, a relação era de um jovem para cada três indivíduos em idade ativa e um idoso para cerca de cinco indivíduos em idade ativa (VASCONCELOS; GOMES, 2012). A variação do Índice de Envelhecimento Populacional entre 1970 e 2010 também foi distinta nas regiões, sendo mais elevada e com aumento mais acelerado nas regiões Sul e Sudeste, mais desenvolvidas socioeconomicamente. Os menores valores, e o menor crescimento deste índice em todo o período foram observados na região Norte (CLOSS; SCHWANKE, 2012).

Neste sentido, Brito (2008) destaca a desigualdade que marca este processo de transição no país ao afirmar que:

“A originalidade da transição demográfica no Brasil está definida pelas particularidades históricas onde ela se insere, permeadas pelos fortes desequilíbrios regionais e sociais. Nessa perspectiva ela é única, enquanto um processo global que atinge toda a sociedade brasileira, mas, ao mesmo tempo, múltipla, pois se manifesta diferentemente segundo as diversidades regionais e, principalmente, sociais” (BRITO, 2008, p.6)

Como as transformações no perfil demográfico estão intrinsecamente relacionadas com as transformações no perfil epidemiológico, as mudanças no processo de adoecimento e óbito da população também se deram de maneira desigual. O que deflagrou a queda da mortalidade no país foram as ações sanitárias e de saúde pública do Estado, como a vacinação, campanhas sanitárias e avanços da atenção médica na rede pública, que permitiram um melhor controle de algumas doenças infecciosas e parasitárias com alta letalidade no período neonatal, reduzindo gradativamente a importância deste grupo de doenças na mortalidade da população (CHAIMOWICZ, 1997).

Entre 1930 e 2010 a taxa de mortalidade infantil reduziu de 135 para 16,2 óbitos para cada mil nascidos vivos, contudo, enquanto a região Sul concentrou os valores mais baixos de mortalidade infantil desde 1970, próximos aos observados na região Sudeste, seguida da região Centro-Oeste, a região Nordeste apresentou os maiores valores até a década de 2000, sendo ultrapassada pela região Norte em 2010 (BARRETO; CARMO, 2007; VASCONCELOS; GOMES, 2012).

Esta redução ocorreu de maneira mais acentuada para os óbitos pós-neonatais, que concentram maior proporção de óbitos por doenças infecciosas, em relação ao componente neonatal, relacionado à assistência pré e pós-natal (BARRETO; CARMO, 2007). Ao comparar a mortalidade infantil no período de 1985 a 1987 e de 1995 a 1997, Victora & Barros apontaram para um aumento de 27,6% no percentual de óbitos neonatais em relação aos demais óbitos infantis no Brasil, sendo o aumento superior a 40% nas regiões Norte e Nordeste, destacando que nestas duas regiões se concentravam, em 1996, os piores indicadores referentes ao atendimento pré-natal e ao parto no país (VICTORA; BARROS, 2001).

De maneira concomitante à queda na mortalidade infantil e ao controle das doenças infecciosas e parasitárias, o aumento na expectativa de vida provocou um aumento proporcional da população idosa, mais sensível ao desenvolvimento de doenças crônicas e degenerativas (MENDES *et al.*, 2012). Logo, as doenças infecciosas e parasitárias que eram responsáveis por mais de 40% dos óbitos do país em 1930 (PRATA, 1992) passaram a representar cerca de 4,3% dos óbitos em 2015 (SIM, 2017c), enquanto as doenças cardiovasculares aumentaram de cerca de 10% (PRATA, 1992) para cerca de 27% (SIM, 2017c), com grandes diferenças regionais.

Neste contexto, observou-se no Brasil nos últimos 60 anos um processo acelerado e desigual, no qual a queda na mortalidade provocou um ganho significativo na esperança de vida, e a queda na fecundidade contribuiu para desacelerar o crescimento populacional e reduzir a população jovem, enquanto a população adulta e idosa aumentou expressivamente. Ao mesmo tempo o perfil epidemiológico mudou, emergindo novos padrões de mortalidade, que geraram demandas de saúde específicas e tornaram-se um desafio. Compreender e identificar estes padrões e sua evolução ao longo dos anos pode contribuir com o desenvolvimento de políticas e ações equitativas visando garantir uma melhor sustentabilidade do SUS.

2.2 DINÂMICA DOS PADRÕES DE MORTALIDADE NO BRASIL

Os padrões de mortalidade no Brasil, marcados por desigualdades regionais, têm sido amplamente discutidos na literatura nacional. Já em meados da década de 1970 Yunes e Rochenzel (1974) descreveram queda nas taxas de mortalidade em todo o país entre 1941 e 1970, sendo o decréscimo mais agudo no Norte e Centro-Oeste, onde os valores eram mais elevados, indicando que nas regiões onde a transição da mortalidade teve início tardio o processo ocorria de maneira mais acelerada.

Duas décadas depois, em estudo publicado originalmente em 1992, Araújo (2012) descreveu intensa queda da mortalidade proporcional por doenças parasitárias de 45,7% para 8% entre 1930 e 1986, e aumento constante das doenças do aparelho circulatório, de 11,8% para 33,5%, das causas externas, de 2,6% para 14,8%, e das neoplasias malignas, de 2,7% para 11,9%, alertando para polarização epidemiológica, geográfica e social caracterizadas pelos contrastes dos principais tipos de agravo de saúde entre e dentro das regiões. No mesmo ano, Prata (1992) identificou maior proporção de mortalidade por doenças cardiovasculares, neoplasias e causas externas, e menor por doenças infecciosas nas regiões Sul e Sudeste entre 1930 e 1985, apesar de acelerado aumento das doenças crônicas e redução das doenças infecciosas no Norte e Nordeste no final do período, reforçando o cenário de transição epidemiológica desigual no Brasil.

Entre 1980 e 2000 as doenças do aparelho circulatório se mantiveram como principal causa de óbito no país, seguidas por causas mal definidas e neoplasias, que ultrapassaram as causas externas, enquanto as doenças infecciosas e parasitárias continuaram em declínio (SVS/MS, 2004). Os padrões e tendências da mortalidade segundo grupos de causa nas regiões Norte e Nordeste em 2001 representavam a continuidade dos achados de Prata (1992) enquanto nas regiões Sul e Sudeste ocorreu queda na mortalidade por doenças do aparelho circulatório e relativa estagnação das neoplasias e causas externas (SVS/MS, 2004).

Segundo dados do SIM (2017a, 2017b) as doenças infecciosas e parasitárias, neoplasias, doenças do aparelho circulatório, do aparelho respiratório e causas externas representaram quase 70% dos óbitos no Brasil entre 1979 e 2015, contudo, poucos estudos analisaram as desigualdades regionais na mortalidade por este grupo de causa no país nos últimos anos.

Paes e Silva (1999) descreveram queda nas taxas de mortalidade por doenças infecciosas em todas as regiões do Brasil entre 1980 e 1990, com maior redução percentual

nas unidades da federação do Sul e Sudeste onde as taxas eram menores, e constataram também aumento das taxas de mortalidade por septicemia na maioria das unidades da federação no período, com destaque para as que compõem Norte e Nordeste. Estudos mais recentes mostraram aumento das taxas de mortalidade por doenças infecciosas específicas, como Dengue em jovens e idosos entre 2000 e 2015, com maior aumento da carga no Norte e Nordeste (ARAÚJO et al., 2017), e HIV/Aids, que também aumentaram no Norte e Nordeste e diminuíram no Sudeste no mesmo período (GUIMARÃES et al., 2017).

Em relação às neoplasias, Guerra *et al.* (GUERRA et al., 2017) identificaram redução das taxas de mortalidade no Brasil como um todo, assim como nas unidades da federação do Sul e Sudeste, e aumento na maioria das que compõem o Norte e Nordeste entre 1990 e 2015. Observaram aumento das taxas de mortalidade por neoplasias de mama, cólon/reto e pulmão entre mulheres, e das neoplasias de próstata e cólon/reto entre homens, e de maneira geral as maiores taxas se concentraram nas unidades da federação do Sul e Sudeste e os maiores aumentos nas do Norte e Nordeste.

Segundo Souza *et al.* (SOUZA et al., 2001) as taxas de mortalidade por doenças do aparelho circulatório diminuíram em todas as faixas etárias no Norte, Sul e Sudeste e aumentaram no Centro-Oeste e Nordeste nas faixas etárias de 40 ou mais anos de idade entre 1979 e 1996. De maneira geral, constataram declínio das taxas de mortalidade por doenças cerebrovasculares e isquêmicas do coração nas regiões Sul, Sudeste, onde eram mais elevadas, e no Norte, e aumento no Nordeste e Centro-Oeste no período. Em estudo mais recente, Souza *et al.* (2006) descreveram tendências semelhantes das taxas de mortalidade por doenças cerebrovasculares e isquêmicas do coração entre 1980 e 2005, com exceção da região Norte, onde observaram aumento no período.

Faltam estudos analisando a tendência das taxas de mortalidade por doenças do aparelho respiratório e por causas externas de maneira geral nas regiões do país. Especificamente sobre a mortalidade de idosos por doenças do aparelho respiratório, Paes (2016) constatou aumento das taxas no sexo masculino em todas as regiões e aumento das taxas no sexo feminino nas unidades da federação do Sul, Sudeste e Nordeste bem como redução no Centro-Oeste entre 1980 e 1995, sendo as taxas mais elevadas no Sul e Sudeste em ambos os sexos. Acerca dos homicídios, causa externa mais frequente no país, Reichenheim *et al.* (2015) observaram taxas de mortalidade mais elevadas no Norte, Nordeste e Centro-Oeste, com tendência de queda no Norte, Centro-Oeste e Sudeste e aumento no Nordeste e no Sul entre 1991 e 2007.

Além das mudanças nas principais causas de óbito observou-se também acelerada e desigual mudança no padrão etário da mortalidade no país. Entre 1990 e 2004 os óbitos na faixa etária de 0 a 4 anos de idade diminuíram de 14,1% para 6,2% dos óbitos, contudo, enquanto no Sul e Sudeste diminuíram de cerca de 11% para menos de 5% no período, na região Norte, em 2004, a proporção de óbitos nesta faixa etária ainda era maior do que a constatada no Brasil em 1990. No outro extremo, os óbitos entre indivíduos com 60 ou mais anos de idade aumentaram em todas as regiões, com os maiores valores concentrados no Sul e Sudeste, seguidos do Nordeste, Centro-Oeste e os menores no Norte (DATASUS/MS, 2012).

As transformações do padrão etário e a transição das principais causas de óbito ocorreram de maneira concomitante, com a queda da mortalidade infantil acompanhando a redução das doenças infecciosas e parasitárias, e o aumento da mortalidade de adulto e idosos o avanço da mortalidade por doenças crônicas e não transmissíveis.

Neste contexto, nos últimos anos os óbitos em faixas etárias elevadas e por doenças crônicas e não transmissíveis apresentaram um impacto cada vez maior na mortalidade do país. Segundo dados do SIM mais de 65% dos óbitos no Brasil em 2015 ocorreram entre idosos, dos quais quase 70% correspondiam a neoplasias, doenças do aparelho circulatório, do aparelho respiratório e causas externas (SIM, 2017b). Mesmo a mortalidade por doenças infecciosas entre os mais jovens foi superada pela de idosos no período de 1980 a 2000, com redução constante da mortalidade por tuberculose pulmonar e gastroenterite infecciosa e aumento da septicemia nas faixas etárias mais elevadas (SVS/MS, 2004).

A distribuição da mortalidade por diferentes grupos de causa segundo faixas etárias foi marcado por desigualdades regionais no país, com maior proporção de anos de vida perdido por morte prematura por doenças infecciosas e parasitárias, maternas, perinatais e nutricionais no Norte e Nordeste e por doenças não transmissíveis no Sul e Sudeste. Em 2008 as taxas de anos de vida perdido por morte prematura por todos os grupos de causa foram mais elevadas no Norte, Nordeste e Centro-Oeste (SCHRAMM et al., 2004; LEITE et al., 2015).

Existem diferenças também na mortalidade de homens e mulheres no país. Em 2001 as taxas de mortalidade entre homens foram superiores à de mulheres em todas as faixas etárias, sendo a diferença mais aguda nas faixas etárias de 10 a 39 anos de idade, e pela maioria dos grupos de causa, com exceção das doenças endócrinas, nutricionais e metabólicas, das doenças de pele e tecido subcutâneo e das doenças do sistema osteo muscular, maiores entre mulheres. (LAURENTI; MELLO JORGE; GOTLIEB, 2005). De maneira geral, as taxas de anos de vida perdido foram mais elevadas entre homens, sobretudo

no Norte e Nordeste, e menores entre mulheres, principalmente no Sul e Sudeste (LEITE et al., 2015).

Neste contexto, entre as principais características da dinâmica da mortalidade no Brasil, destacam-se:

- i. Intensa transformação das principais causas de mortalidade no país, com aumento proporcional das doenças crônicas, como as neoplasias e doenças do aparelho circulatório, e das causas externas e redução da mortalidade por doenças infecciosas e parasitárias;
- ii. Estas transformações na mortalidade por grandes grupos de causa ocorreram em ritmo e intensidade desiguais nas grandes regiões do país, sendo um processo mais antigo e avançado nas regiões Sul, Sudeste e Centro Oeste, e mais recente e acelerado nas regiões Norte e Nordeste;
- iii. As mudanças afetaram de maneira desigual a mortalidade de diferentes grupos etários, de maneira que mesmo com as intensas transformações nos padrões regionais de mortalidade, e a tendência de aproximação entre as regiões, as doenças infecciosas e parasitárias ainda são mais elevadas no Norte e Nordeste e as doenças crônicas e degenerativas no Sul e Sudeste;
- iv. As desigualdades propiciaram o aparecimento de padrões regionais específicos, como o aumento da mortalidade por neoplasias e doenças do aparelho circulatório no Norte e Nordeste em contraste com a redução das mesmas no Sul e Sudeste;
- v. Ocorreram mudanças no padrão etário da mortalidade no país, com a mortalidade de idosos e por doenças crônicas e não transmissíveis aumentando continuamente em detrimento da mortalidade infantil;
- vi. A transição da idade de óbito também ocorreu de maneira desigual e, mesmo em anos recentes, a mortalidade por todos os grupos de causa é mais precoce nas regiões Norte e Nordeste;
- vii. A expressão da mortalidade se deu de maneira desigual entre homens e mulheres;

Contudo, há lacunas de conhecimento em relação à mortalidade no país, pois fica evidente a necessidade de se abordar as desigualdades regionais na mortalidade por grandes grupos de causas e causas específicas, e sua distribuição por diferentes grupos etários, bem como a melhor compreensão da influência do contexto social, econômico e demográfico na dinâmica da mortalidade.

Neste sentido, a despeito de estudos e relatórios que abordam a temática no Brasil ao longo dos anos, este é um campo ainda fértil para a saúde pública, possibilitando o uso de técnicas que permitam tanto descrever o panorama mais amplo da mortalidade quanto identificar os diversos cenários locais, elucidando padrões e tendências que permitam apontar as necessidades atuais e realizar previsões sobre estes cenários nos próximos anos.

2.3 *BIG DATA*, *DATA MINING* E *MACHINE LEARNING* – O POTENCIAL DE APLICAÇÃO EM SAÚDE PÚBLICA

Todo o planejamento, tomada de decisões e desenvolvimento de políticas para atender às diferentes necessidades de saúde da população são baseadas nos conhecimentos extraídos de diversos sistemas de informação na área da Saúde Pública que atualmente atingiram volume, variedade e são gerados em velocidade jamais vistos (OLA; SEDIG, 2014). Esta grande massa de dados, com complexidade cada vez maior, organizados e divulgados por diferentes fontes e apresentando diversos níveis de confiabilidade, caracterizam o que é chamado de *big data*, grandes dados em tradução literal, e sua utilização se mostra um desafio e uma oportunidade para a Saúde Pública (BORDER, 2014).

Assim como o volume e a complexidade dos dados evolui, os métodos para analisá-los também evoluem, e os avanços computacionais bem como o desenvolvimento e o aprimoramento das técnicas de análise permitem cada vez mais uma abordagem abrangente destas informações, apesar de ainda pouco exploradas na área da saúde (PENTLAND; REID; HEIDBECK, 2013; OLA; SEDIG, 2014). Neste contexto, os desafios para a saúde pública se encontram na necessidade de explorar ferramentas e técnicas de análise que não são convencionais, na abordagem dos dados disponíveis nos sistemas de informação sob uma nova perspectiva e no instrumental tecnológico necessário para fazê-lo, enquanto a oportunidade emerge de possibilidades ainda pouco exploradas de extração de informações de grandes dados que não poderiam ser obtidas em “dados menores”.

A mineração de dados, tradução livre do termo original *data mining*, se refere ao conjunto de técnicas e ferramentas de análise exploratória mais apropriadas para abordar dados em larga escala na identificação de potenciais regras e padrões (BERRY; LINOFF, 2004; WITTEN; FRANK; HALL, 2011). O processo de mineração de dados é composto pelas fases de coleta, onde os dados são obtidos e armazenados em uma base de dados, extração de recursos e limpeza de dados, também chamada de pré-processamento, em que os dados são padronizados e adequados para o processamento, e etapa de processamento ou analítica, quando um ou mais algoritmos são aplicados para identificação de estruturas (AGGARWAL, 2015).

O pré-processamento é a fase mais crucial e que mais consome tempo no processo de mineração de dados, contudo, geralmente não recebe a devida atenção pois o foco principal está na análise. Esta fase é composta por uma etapa de extração de recursos, na qual os dados

mais pertinentes a cada análise são extraídos da base de dados, uma etapa de limpeza de dados, onde as informações faltantes ou inconsistentes são retiradas, corrigidas ou estimadas, e uma etapa de seleção e transformação, em que os dados extraídos na primeira etapa são selecionados e padronizados, quando necessário (WITTEN; FRANK; HALL, 2011; AGGARWAL, 2015; TORGO, 2017).

Uma ferramenta importante para mineração de dados que pode ser utilizada tanto no pré-processamento quanto na fase analítica são as representações visuais, que codificam as estruturas dos dados através de marcadores visuais, como pontos, linhas e formas, que ganham complexidade quando combinadas e tiram proveito do sistema de percepção visual humano, adaptado ao processamento rápido de dados e ao reconhecimento de padrões (OLA; SEDIG, 2014; TORGO, 2017). Gráficos exploratórios mais simples podem ser implementados para a identificação de inconsistências na base dados ou para permitir a visualização de padrões que serão explorados mais profundamente na fase analítica, enquanto representações gráficas mais complexas geralmente são utilizadas para apresentar e explicar os padrões e tendências evidenciados em todo o processo de mineração (BARLOW, 2015).

Qualquer tipo de visualização depende de quatro elementos chaves da apresentação de dados: comparação, permitindo comparar diferenças entre diversos itens em um determinado momento ou as mudanças de uma variável em um período, relacionamento, para elucidar a correlação entre variáveis, distribuição, que representa o espalhamento dos dados, e composição, que se refere a representação de múltiplos componentes, como os gráficos de setores ou colunas (RAMASUBRAMANIAN; SINGH, 2017). Segundo Yau (2013) uma boa visualização de dados é aquela que ajuda a descobrir algo que de outra maneira não seria observado, permitindo a identificação de tendências, padrões e valores discrepantes que despertam o êxtase de ver algo pela primeira vez. Afirma ainda que a representação gráfica deve ser encarada mais como um meio do que como ferramenta, e ainda que tenha de ser lida de maneira rápida e precisa, deve despertar a curiosidade.

Neste sentido, gráficos representam uma das maneiras mais acessíveis, convincentes e expressivas de identificar e descrever padrões em dados, capturando melhor os detalhes e mostrando características que não poderiam ser observadas com estatísticas descritivas ou de maior complexidade, facilitando a compreensão das informações (VISWANATHAN et al., 2016; MAILUND, 2017; BAUMER; KAPLAN; HORTON, 2017). No presente estudo foram propostos gráficos não convencionais, que serão melhores detalhados nos capítulos a seguir, para descrição de padrões e tendências de mortalidade no Brasil.

As técnicas de *machine learning*, aprendizado de máquina em tradução livre, fazem parte do instrumental utilizado na fase analítica na mineração de dados e se baseiam em algoritmos adaptáveis que “aprendem” com os dados analisados, de maneira que quanto maior o volume de dados disponíveis maior o potencial de “aprendizado” para identificar estruturas e correlações desconhecidas (RAMASUBRAMANIAN; SINGH, 2017).

Segundo Chiavegatto Filho (2015) a metodologia de aprendizado de máquina mais utilizada em análise de dados é a árvore de decisão. Diversos estudos utilizaram esta técnica em dados relacionados à saúde, sobretudo na identificação de preditores de eventos específicos em clínica médica, abordando um número de observações não muito grande mas lidando com uma extensa variedade de variáveis explicativas (KUSIAK; DIXON; SHAH, 2005; PATEL et al., 2014; PUSKARICH, 2015; ESTEBAN et al., 2015; TAYLOR et al., 2016). Abordando uma base de dados da Organização Mundial de Saúde sobre mortalidade com muitas observações (mais de 1 milhão de registros), mas com menos variáveis, Paoin (2011) utilizou árvores de decisão e outros algoritmos de aprendizagem de máquina para prever a causa específica de óbito.

Estes estudos dedicaram grande atenção ao potencial das árvores de decisão para predição, contudo, uma das vantagens da utilização desta técnica é a expressão dos resultados através gráficos de fácil compreensão, pois refletem de maneira mais próxima a tomada de decisão humana do que os modelos de regressão (JAMES et al., 2013). Neste sentido, diante da importância da representação visual e gráfica na mineração de dados, sua utilização pode ser especialmente útil na identificação e descrição de padrões e tendências.

No Brasil uma grande quantidade de dados referentes aos óbitos são gerados diariamente através da DO e disponibilizados pelo SIM. As informações referentes a este evento vital somam mais de 32 milhões de registros desde 1979, e englobam informações sobre os indivíduos e as circunstâncias do óbito, como idade, sexo, ocupação, raça/cor, local de moradia, causa de óbitos entre outros. Estas informações são constantemente analisadas de maneira agregada e fragmentada, abordando os óbitos segundo causas, locais, faixa etária e/ou período específico, mas raramente são analisados como um todo ou de uma perspectiva em que cada óbito é uma unidade de análise.

Considerando o cenário de transição apresentado inicialmente neste estudo, pode ser de grande valia para a saúde pública explorar as informações individuais sobre a mortalidade no país de maneira integral, na expectativa de elucidar padrões e tendências na mortalidade através de estruturas que só podem ser extraídas quando a massa de dados é analisada como um todo. Apesar do elevado volume de dados do SIM, a velocidade de divulgação e a

variedade das informações disponíveis não são grandes o suficiente para ser considerado um grande dado, contudo métodos convencionais de análise podem não ser suficientes para a abordagem das informações individuais, fazendo necessária a implementação de técnicas de mineração de dados.

Neste contexto, dado o amplo instrumental estatístico e computacional disponível para exploração de grandes massas de dados, sua utilização ainda escassa e recente na área de saúde pública e o grande volume de informação do SIM, evidencia-se uma janela de oportunidade para a implementação de técnicas de mineração de dados na perspectiva de identificar padrões de mortalidade no país.

2.4 O SISTEMA DE INFORMAÇÃO SOBRE MORTALIDADE – SIM

Com a criação do Sistema Nacional de Vigilância Epidemiológica (SVE) em 1975 tornou-se mister o desenvolvimento de um sistema de informação sobre mortalidade que permitisse a obtenção de dados confiáveis e abrangentes, o que culminou com a criação do Sistema de Informação Sobre Mortalidade (SIM). Para tanto, identificou-se a necessidade de adoção de um formulário padrão para coleta de informação, pois naquele momento existiam mais de 40 modelos oficiais de Atestado de Óbito utilizados em diferentes estados e municípios, havendo registros de óbito em papel sem timbre, sem formato definido ou qualquer identificação do notificante (JORGE; LAURENTI; GOTLIEB, 2007; MINISTÉRIO DA SAÚDE, 2011).

Uma comissão formada por Ruy Laurenti, José Leão Costa e João Antônio Neto Caminha, foi designada para a definição das variáveis principais contempladas no formulário padrão e que seriam processadas de maneira central pelo Ministério da Saúde, sendo possível a cada estado ampliar as variáveis trabalhadas. Este formulário padrão, chamado de Declaração de Óbito (DO), tornou-se a fonte de alimentação do SIM, que seria implementado neste mesmo ano (JORGE; LAURENTI; GOTLIEB, 2007; MINISTÉRIO DA SAÚDE, 2011).

Para a implementação e utilização da DO foi realizado treinamento de pessoal e produzido manuais para preenchimento, codificação de variáveis, processamento, crítica e correção de dados. Além disto, a Lei nº 6.015, de 31/12/73, com as alterações introduzidas pela Lei nº 6.126 de 30/07/75, caracterizou incentivo legal para utilização da DO, tornando-a essencial para a obtenção da Certidão de Óbito, documento indispensável para a realização de sepultamentos no país (MINISTÉRIO DA SAÚDE, 2009).

O documento sofreu alterações desde sua criação e em 1994 passou a ser composta de três vias numeradas e de cores diferentes. As DO atualmente são distribuídas pelo Ministério da Saúde às Secretarias Estaduais de Saúde que repassam para as Secretarias Municipais de Saúde, responsáveis por fornecê-las aos estabelecimentos de Saúde, Institutos de Medicina Legal, Serviços de Verificação de Óbito e Cartórios de Registro Civil, e evitar que funerárias tenham acesso às mesmas. A primeira via da DO (branca) preenchida é captada das unidades notificadoras mensalmente pelos órgãos responsáveis pelas estatísticas de mortalidade através de busca ativa, a segunda via (amarela) é entregue à família, que deverá apresentá-la ao cartório de registro civil, onde ficará arquivada e serão utilizadas nos procedimentos legais, e

a terceira via (rosa) fica anexada à documentação do paciente na unidade notificadora (JORGE; LAURENTI; GOTLIEB, 2007; MINISTÉRIO DA SAÚDE, 2009).

O processamento dos dados das DO é realizado localmente, sendo criticado em relação à consistência das informações essenciais. O Sistema Seletor de Causa Básica (SCB), que seleciona automaticamente a causa básica do óbito, se encontra incorporado ao SIM, sendo um importante instrumento na uniformização da seleção da causa básica de óbito. São consideradas indispensáveis as variáveis ano e tipo de óbito, de maneira que DO sem estas informações não são registrados pelo sistema, e como essenciais as variáveis sexo, idade, município de residência e ocorrência, causa de óbito e tipo de violência, tendo a prioridade em relação à crítica e correção. As demais variáveis são consideradas secundárias e não possuem prioridade em relação a crítica e correção, mais realizadas nas áreas onde o contingente de informação “ignorada” é elevada (MINISTÉRIO DA SAÚDE, 2011).

Apesar das desigualdades regionais, as informações do SIM têm melhorado nos últimos anos. Segundo informações da “Consolidação da Base de Dados do SIM em 2011” (MINISTÉRIO DA SAÚDE, 2013) a cobertura, dada pela razão entre os óbitos coletados pelo sistema e os óbitos projetados pelo IBGE, no Brasil foi de 96,1%, o maior valor desde 2000, sendo superior a 100% no Rio de Janeiro, São Paulo, Paraná, Santa Catarina, Rio Grande do Sul e Mato Grosso do Sul. O percentual de óbitos por causas mal definidas, que representam uma lacuna em relação a distribuição por grupos de causas, têm diminuído, assim como os registros faltantes de diversos campos (MELLO JORGE; LAURENTI; GOTLIEB, 2007).

O SIM é o sistema de informação em saúde em uso mais antigo do país, que representa uma importante fonte de dados secundários de acesso aberto, com elevado volume, cobrindo todo o território nacional e de boa qualidade. Os microdados somam mais de 32 milhões de registros individuais acerca da mortalidade desde 1979 e no decorrer dos anos cada vez mais variáveis foram adicionadas, proporcionando uma grande diversidade de informação. O Sistema é alimentado constantemente e as informações disponíveis apresentam uma defasagem média de dois anos, caracterizando uma grande massa de dados que pode ser analisada para melhor compreensão dos padrões e tendências da mortalidade no país.

2.5 MÉTODOS DE MINERAÇÃO PARA GRANDES DADOS E ANÁLISES GRÁFICAS

2.5.1 MÉTODOS BASEADOS EM ÁRVORES: ÁRVORES DE DECISÃO (CART) E RANDOM FOREST

Os modelos de árvore, conhecidos como *CART – Classification And Regressions Trees*, são algoritmos de classificação de aprendizado de máquina que utilizam a estratégia de “dividir e conquistar” para identificar valores ou categorias de variáveis que determinam a predição de um desfecho de interesse, estratificando o espaço de predição, que corresponde aos dados analisados, em regiões mais simples através de um conjunto de “regras” que definem “nós”, “ramos” e “folhas” que podem ser representadas graficamente por uma “árvore de cabeça para baixo”. Este tipo de método pode ser aplicado tanto para predição de variáveis quantitativas, quando são estabelecidas Árvores de Regressão, quanto variáveis qualitativas, através das Árvores de Classificação (JAMES et al., 2013; WITTEN; FRANK; HALL, 2011; BAKAR et al., 2011; MITCHELL, 1997).

A Figura 1 mostra um exemplo de representação gráfica, adotada no presente estudo, de uma árvore de classificação ajustada a partir das informações de cada óbito registrado no SIM onde a variável resposta é o grupo de causa de óbito (Doenças do Aparelho Circulatório, Causas Externas, Doenças Infecciosas e Parasitárias, Causas Mal Definidas, Neoplasias, Demais Causas e Doenças do Aparelho Respiratório). Cada “nó”, destacado em azul na figura, representa uma variável explicativa e os “ramos”, destacados em verde, são compostos por divisões desta variável em valores ou categorias mutuamente exclusivas. As “folhas”, destacadas em vermelho na parte inferior do gráfico, representam o valor ou categoria da variável resposta predito pelas divisões do espaço amostral definidas pelos “nós” e “ramos” desde a “raiz” da árvore (MITCHELL, 1997; JAMES et al., 2013).

Os números apresentados nas duas linhas nas “folhas” da árvore são referentes às proporções de óbitos por grupo de causa enquanto o percentual mais abaixo representa a proporção aproximada do conjunto total de dados alocado naquela “folha”. Neste sentido, no exemplo apresentado (Figura 1), o grupo de dados alocado na “folha” mais à esquerda na figura, classificado como “Causas Externas”, determinada pelo “ramo” referente aos óbitos ocorridos em via pública e outros locais, representa 8% do total de registros analisados pelo modelo. Neste grupo, referente aos óbitos ocorridos em via pública e outros locais, 12% dos óbitos foram por Doenças do Aparelho Circulatório, 68% por Causas Externas, 1% por

Doenças Infecciosas e Parasitárias, 11% por Causas Mal Definidas, 2% por Neoplasias, 4% por Demais Causas e 2% por Doenças do Aparelho Respiratório.

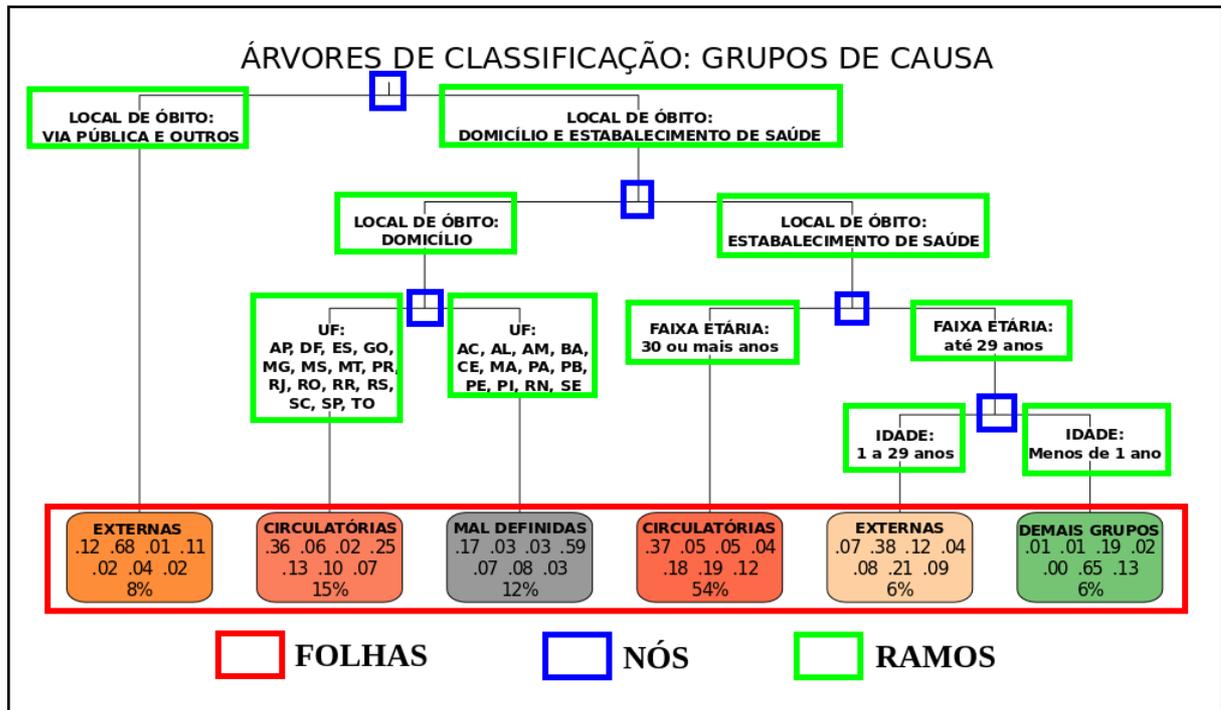


Figura 1: Exemplo de representação gráfica de uma árvore de classificação e suas estruturas

Existem diversos algoritmos para construção de árvores de classificação, mas de maneira geral, as regras que definem a divisão do espaço amostral são aplicadas “de cima para baixo”, de forma que na “raiz” da árvore, na parte superior do gráfico, todas as observações analisadas se encontram em uma mesma região de predição e, a partir de divisões binárias recursivas, o espaço preditor é estratificado sucessivamente (MITCHELL, 1997; JAMES et al., 2013).

Neste estudo foi utilizada a biblioteca *rpart* do *software* estatístico R cujo algoritmo realiza a construção da árvore em dois estágios. No primeiro estágio a melhor variável preditora para divisão dos dados é selecionada, processo que é repetido em cada subgrupo originado da divisão anterior, até que os grupos atinjam um tamanho mínimo ou até que não ocorram mais ganhos explicativos. No segundo estágio são selecionadas as divisões que irão compor a árvore através de validação cruzada (THERNEAU; ATKINSON; MAYO, 2017).

A melhor variável preditora para cada divisão é selecionada através da “impureza”, definida pela distribuição das categorias da variável predita nos subgrupos resultantes das possíveis divisões (TAN, 2013; THERNEAU; ATKINSON; MAYO, 2017). Diferentes medidas de “impureza” podem ser utilizadas como critério para selecionar as variáveis que

definem as divisões do espaço preditor em uma árvore de classificação (MUCHAI; ODONGO, 2014) e no presente estudo optou-se pela utilização do Índice de Impureza de Gini, que assume valor zero quando apenas uma categoria da variável predita está presente no subgrupo resultante da divisão, e valor máximo quando a distribuição das categorias é uniforme (TAN, 2013). Neste contexto, divisões com menor impureza são melhores preditoras porque subgrupos formados por observações com a mesma característica, ou com a mesma categoria da variável predita, são mais “puros” ou “melhor definidos” do que subgrupos onde as características são distribuídas de maneira uniforme.

O índice de Gini é calculado por $1 - \sum_{i=0}^{c-1} [p(i|t)]^2$ onde c é o número de categorias da variável predita e $p(i|t)$ corresponde à fração de registros ou observações pertencentes à categoria i em um determinado subgrupo t . Para a construção da árvore as divisões são selecionadas de acordo com a média ponderada do Índice de Gini dos subgrupos, obtida através da soma dos produtos entre Índice de Gini e as proporções das observações presentes em cada subgrupo (TAN, 2013).

O número de subdivisões, ou nós, que compõem uma árvore estão relacionados ao seu nível de complexidade. Árvores menos complexas são adequadas quando o objetivo é a tomada de decisão, por serem mais compreensíveis, contudo, árvores mais complexas geralmente são mais acuradas e devem ser priorizadas quando o objetivo principal da aplicação deste modelo é a predição (ROKACH; MAIMON, 2005).

Os modelos *Random Forest*, floresta aleatória em tradução livre, são obtidos por um conjunto de árvores de decisão ajustadas através de um processo de *bootstrap* garantindo ganhos no poder de predição e permitindo identificar de maneira mais clara a importância das variáveis preditoras. Como é ajustado a partir de um grande número de árvores, seus resultados são mais difíceis de se interpretar, uma vez que não podem ser representados por meios gráficos como as árvores. Neste sentido, o *Random Forest* aumenta a qualidade da predição ao custo da perda da interpretabilidade (JAMES et al., 2013).

A principal aplicação para este tipo de modelo, além da predição, é identificar a importância das variáveis preditoras através do montante do Índice de Gini que é reduzido pelas divisões segundo cada variável considerando todas as árvores construídas para sua implementação. A este montante de Índice de Gini reduzido é dado o nome de *Mean Decrease Gini Index*, Decréscimo Médio do Índice de Gini em tradução literal (JAMES et al., 2013). Como Índices de Gini mais baixos representam melhores preditores, quanto maior for

o decréscimo médio do índice atribuído a uma variável, mais importante aquela variável foi na composição do conjunto de árvores que originaram o modelo *Random Forest*.

2.5.2 ANÁLISE DE CORRESPONDÊNCIA

A Análise de Correspondência é um método de análise multivariada indicado para descrever a relação entre duas variáveis qualitativas através de gráficos representando a proximidade de suas categorias (INFANTOSI; COSTA; ALMEIDA, 2014; ABDI; VALENTIN, 2007; CARVALHO; STRUCHINER, 1992). Os conceitos principais deste método foram descritos inicialmente por Fisher na década de 1940 e aprimorados e desenvolvidos por diversos pesquisadores partir da década de 1970 (INFANTOSI; COSTA; ALMEIDA, 2014; DORÉ et al., 1996).

De maneira simplificada, a partir de uma tabela de contingência composta por duas variáveis, que pode ser representada por dois conjuntos I e J constituídos por i e j elementos que correspondem às categorias destas variáveis, respectivamente, extrai-se uma matriz de correspondência P_{ixj} onde cada elemento $p_{i,j}$ é dado pela divisão de n_{ij} pelo total

principal da tabela $\sum_{i=1}^I \sum_{j=1}^J n_{ij}$. Desta matriz são obtidas massas e perfis que associados formam uma nuvem de pontos que podem ser projetados em um plano para representar as relações entre as categorias de variáveis através da distância entre elas no gráfico. Os eixos que compõem o plano bidimensional para a representação da nuvem de pontos são chamados de autovetores e representam parte da variância total da matriz. Quanto maior a variância da matriz representada por um eixo, melhor descreve a projeção dos pontos da nuvem no plano e, consequentemente, melhor representa a relação entre as variáveis (CARVALHO; STRUCHINER, 1992; SOUZA, 2004; CYRINO, 2011).

A Análise de Correspondência Múltipla (ACM) é uma generalização da Análise de Correspondência para diversas variáveis. Partindo de uma tabela $N \times Q$ onde as linhas se referem a N observações e as colunas a Q variáveis, cada uma composta por J_q categorias, obtém-se uma matriz indicadora $N \times J$ onde as colunas representam variáveis *dummy* para cada uma das categorias de Q variáveis, sendo $J = \sum J_q$ (GREENACRE, 2010).

A Figura 2 apresenta um exemplo de representação gráfica de uma análise de correspondência múltipla executada no presente estudo. Foram atribuídas cores iguais para as categorias referentes a mesma variável com o intuito de facilitar a visualização da relação as mesmas, e quanto mais próximas as categorias se encontram no plano, maior é a associação. Os eixos horizontal (Dim1) e vertical (Dim2) são os autovetores e os percentuais relacionados aos mesmos se referem à variância explicada por cada um.

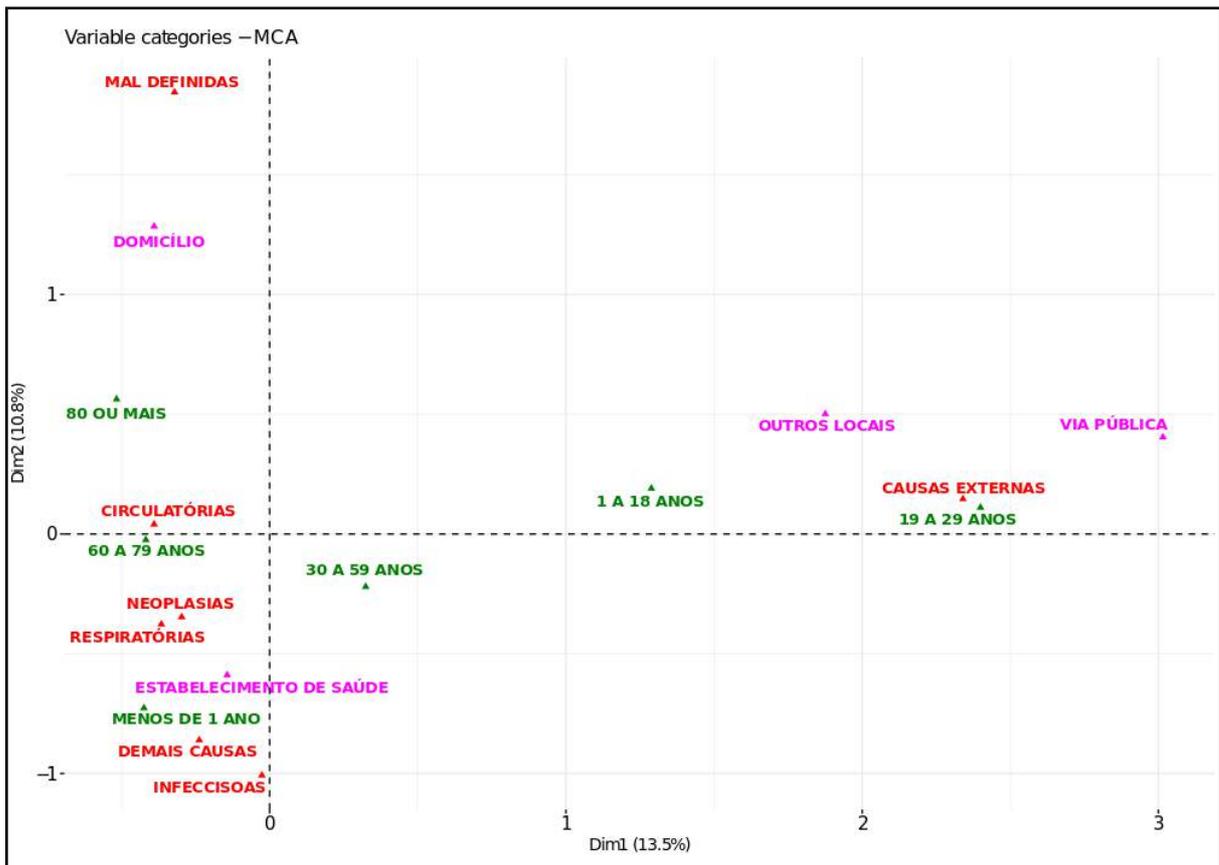


Figura 2: Exemplo da representação gráfica de uma Análise de Correspondência Múltipla

2.5.3 DIAGRAMAS DE KIVIAT

(Doenças do Aparelho Circulatório, Causas Externas, Doenças Infecciosas e Parasitárias, Causas Mal Definidas, Neoplasias, Demais Causas e Doenças do Aparelho Respiratório)

O Diagrama de Kiviat, também chamado de “Gráfico de Radar” ou “Polígonos Irregulares”, é um método gráfico para apresentação de dados multivariados. Neste, três ou mais categorias de uma variável (fatores) são representadas em um plano composto por eixos originados no mesmo ponto ao longo dos quais podem ser sobrepostos polígonos formados pela união de pontos referentes a medida de outras variáveis segundo cada fator (ELLIOTT, 1976).

A Figura 3 apresenta um exemplo Diagrama de Kiviat desenvolvido neste estudo, utilizando informações sobre os óbitos ocorridos no Brasil entre 1979 e 1984 por Doenças Infecciosas e Parasitárias, Neoplasias, Doenças do Aparelho Circulatório, Doenças do Aparelho Respiratório e Causas Externas. Cada eixo se refere às 5 regiões do Brasil e cada polígono representa uma causa de óbito de maneira que quanto mais próximo da extremidade exterior do eixo se encontra o ponto que forma o polígono maior a mortalidade proporcional pelo grupo de causa correspondente. Polígonos mais distorcidos refletem maiores diferenças regionais.

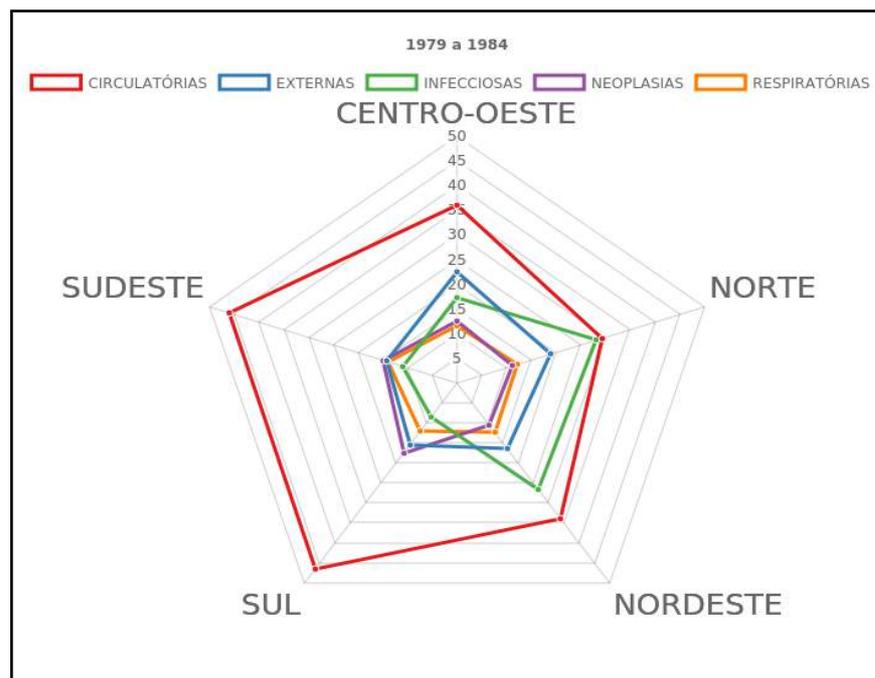


Figura 3: Exemplo de Diagrama de Kiviat utilizado para representação da mortalidade proporcional segundo grupos de causa nas regiões do Brasil

2.5.4 HEATMAPS

O termo *heatmap* pode ser traduzido de maneira literal como “mapa de calor”, o que pode induzir a ideia de que o espalhamento espacial de variáveis analisadas teria sido representado através de mapas coropléticos. Contudo, no presente estudo utilizou-se o termo *heatmap* para referir à representação gráfica desenvolvida para apresentar de maneira intuitiva a tendência espaço-temporal das taxas de mortalidade no país, sem utilizar mapas ou séries temporais.

A Figura 4 apresenta um exemplo de *heatmap*, utilizado para descrever as taxas de mortalidade por uma das causas analisadas no presente estudo. O gráfico foi composto por um mosaico de 918 células dispostas em 27 “linhas” (eixo horizontal), referentes às unidades da federação, e 34 “colunas” (eixo vertical) correspondentes a cada ano do período de 1980 a 2013. Neste sentido, cada célula do gráfico representou uma Unidade da Federação em determinado ano, sendo atribuída uma cor referente ao valor da taxa de mortalidade daquela unidade naquele ano.

Uma linha vertical foi traçada no ponto do eixo horizontal referente ao 1996, no qual a CID-10 substituiu a CID-9, com o intuito de facilitar a identificação de possíveis alterações nas taxas de mortalidade originadas pela mudança na classificação. A organização das Unidades da Federação agrupadas por Região seguiu a ordenação adotada nos arquivos disponibilizados pelo DataSUS, com o intuito de favorecer o entendimento do espalhamento espacial das taxas de mortalidade, mesmo sem a utilização de mapas.

Foi adotada uma escala de cores divergentes, em que os tons mais escuros representaram os valores mais extremos do indicador, verde os mais baixos e vermelho os mais elevados, intermediado por tons mais claros. Os pontos de corte utilizados para estabelecer o gradiente de cores em cada gráfico foram obtidos do mínimo, máximo, 1º e 2º quartil e mediana da distribuição composta pelos valores das taxas de mortalidade por causa, considerando todas as Unidades da Federação em todo o período. Desta maneira, células de cor vermelha e verde em tom mais escuro representam, respectivamente, os valores mais elevados e mais baixos daquela taxa de mortalidade em todo o período, enquanto células em tom mais claro representaram valores medianos.

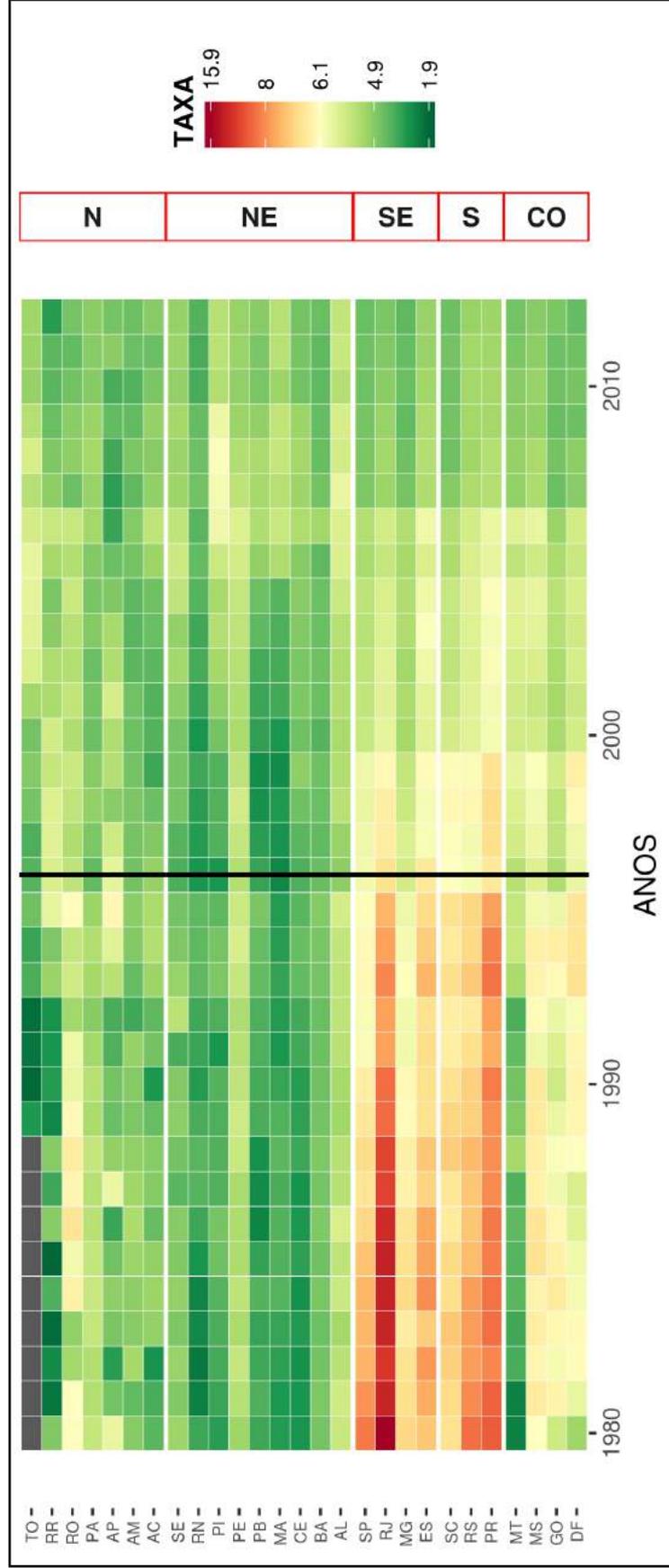


Figura 4: Exemplo de *heatmap* utilizado para descrição do padrão espaço-temporal das taxas de mortalidade nas Unidades da Federação no período estudado

3 JUSTIFICATIVA

O Brasil passou por um processo de transição epidemiológica rápido e polarizado. Este processo no país foi deflagrado principalmente por ações sanitárias e de saúde pública de grande impacto na mortalidade infantil, nem sempre acompanhadas por desenvolvimento e melhorias nas condições de vida e saúde da população, de maneira que a heterogeneidade do contexto social, econômico e demográfico influenciou a velocidade e a intensidade com o qual vem ocorrendo, afetando de maneiras diferentes grupos sociais e regiões geográficas distintas e, conseqüentemente, agravando as desigualdades em saúde.

Uma massa cada vez maior de dados referentes à mortalidade é produzido e divulgado anualmente no país, totalizando atualmente mais de 32 milhões de óbitos registrados. O conjunto de métodos computacionais e estatísticos conhecidos como mineração de dados podem ser implementados junto a este grande volume de dados, abordando-o de uma perspectiva diferente dos estudos clássicos, considerando diferentes níveis de agregação e também as informações individuais.

O número de estudos sobre a mortalidade no país é extenso, contudo, poucos abordam os diferentes padrões e tendências regionais de diversas causas de óbito ou as relacionam com indicadores sociais, econômicos e demográficos; e nenhum estudo abordando as informações individuais do SIM como uma grande massa de dados foi encontrado, o que caracteriza lacunas de conhecimento sobre a temática e evidencia uma janela de oportunidade para a implementação de mineração de dados em uma perspectiva alternativa à dos métodos clássicos para a análise da mortalidade.

A abordagem de informações individuais dos óbitos que compõem o SIM em conjunto com a análise da mortalidade por diversas causas em diferentes níveis de agregação, pode ajudar a descrever o panorama amplo da mortalidade, compreendendo a relação do contexto socioeconômico e demográfico com a dinâmica da mortalidade e a identificar especificidades regionais, elucidando padrões e tendências espaciais e temporais que permitam apontar necessidades atuais e a predizer o cenário nos próximos anos, subsidiando o desenvolvimento de políticas de saúde pública equitativas e proporcionando instrumentos que possibilitem uma gestão mais efetiva do SUS.

4 OBJETIVOS

4.1 OBJETIVO GERAL

Analisar padrões regionais e temporais da mortalidade no Brasil no período entre 1979 e 2013

4.2 OBJETIVOS ESPECÍFICOS

Analisar padrões e tendências de mortalidade segundo grupos de causa e causas selecionadas no Brasil.

Analisar a tendência das taxas de mortalidade por causas selecionadas e identificar os indicadores sociais, econômicos e demográficos relacionados.

Discutir a implementação de técnicas de mineração de dados na identificação de padrões de mortalidade no Brasil.

5 MATERIAL E MÉTODOS

Trata-se de um estudo ecológico, descritivo e exploratório que utilizou dados do Sistema de Informação Sobre a Mortalidade (SIM/MS) no período de 1979 a 2013, do Censo 1991, 2000 e 2010 e Estimativas Populacionais do Instituto Brasileiro de Geografia e Estatística (IBGE) para o período de 1980 a 2012, disponibilizados pelo Departamento de Informática do SUS (DataSUS).

Esta Tese foi desenvolvida em três etapas:

1. Na primeira etapa foi realizada a análise da evolução da mortalidade proporcional e do padrão etário de óbitos segundo grupos de causa, e das taxas de mortalidade segundo grupos de causas e causas selecionadas
2. Na segunda etapa foram analisadas taxas de mortalidade segundo causas selecionadas e indicadores sociais, econômicos e demográficos
3. Na terceira etapa foram aplicadas técnicas de Mineração de Dados nas informações individuais do SIM na identificação de padrões de mortalidade no Brasil

5.1 VARIÁVEIS ANALISADAS E FONTES DE DADOS

Em cada etapa do estudo um dos objetivos específicos foi contemplado abordando diferentes indicadores de mortalidade, unidades de observação, período e métodos de análise, como mostra o Quadro 1.

Na primeira etapa analisou-se evolução da mortalidade proporcional por Doenças Infecciosas e Parasitárias (códigos 001-139 da CID – 9 e A00-B99 da CID – 10), Neoplasias (códigos 140-239 da CID – 9 e C00-D48 da CID – 10), Doenças do Aparelho Circulatório (códigos 390-459 da CID – 9 e I00-I99 da CID – 10), Doenças do Aparelho Respiratório (códigos 460-519 da CID – 9 e J00-J99 da CID – 10), Causas Externas (códigos 800-999 da CID – 9 e V01-Y98 da CID – 10), grupos de causa selecionados por refletir os estágios do processo de transição epidemiológica e suas diferenças regionais. Para a análise da evolução do ranqueamento da mortalidade proporcional segundo grupos de causa foram consideradas também a mortalidade por Causas Mal Definidas (códigos E001-E139 da CID – 9 e R00-R99

da CID – 10), devido a seu potencial de indicar mudanças na qualidade do SIM, e Demais Causas de Óbito (Demais códigos das CID – 9 e CID – 10).

	OBJETIVO ESPECÍFICO	INDICADORES ANALISADOS	UNIDADE DE OBSERVAÇÃO	PERÍODO	MÉTODOS DE ANÁLISE
ETAPA I	Analisar padrões e tendências de mortalidade segundo grupos de causa e causas selecionadas no Brasil	Mortalidade Proporcional segundo grupos de causa	Brasil e Regiões	1979 a 1984, 1985 a 1989, 1990 a 1995, 1996 a 1999, 2000 a 2004, 2005 a 2013	Fluxograma do ranqueamento da Mortalidade Proporcional segundo grupo de causa
		Mortalidade Proporcional segundo grupos de causa	Brasil e Regiões	1979 a 1984, 1985 a 1989, 1990 a 1995, 1996 a 1999, 2000 a 2004, 2005 a 2013	Diagramas de Kiviat (Gráficos de Radar)
		Idade de Óbito	Cada Óbito Registrado no SIM	Cada ano de 1979 a 2013	Gráficos com <i>boxplot</i> da idade de óbito por ano segundo grupo de causa no Brasil e Regiões
		Taxas de Mortalidade Padronizadas segundo grupos de causa e causas selecionadas	Unidades da Federação	Cada ano de 1980 a 2013	<i>Heatmaps</i>
ETAPA II	Analisar a tendência das taxas de mortalidade por causas selecionadas e identificar os indicadores sociais, econômicos e demográficos relacionados	Taxas de Mortalidade Padronizadas segundo causas selecionadas e Indicadores Sociais, Econômicos e Demográficos	Microrregiões	1991, 2000 e 2010	Análise Fatorial para seleção de Indicadores; <i>Boxplots</i> das Taxas de Mortalidade e dos Indicadores organizados por Regiões em cada ano; Correlação Linear de Pearson
ETAPA III	Discutir a implementação de técnicas de mineração de dados na identificação de padrões de mortalidade no Brasil	Variáveis Individuais sobre cada óbito ocorrido no País	Cada Óbito Registrado no SIM	1980 a 2013 (Agregados por período)	Modelos <i>CART</i> (Árvores de Classificação)
		Variáveis Individuais sobre cada óbito ocorrido no País	Cada Óbito Registrado no SIM	1980 a 2013 (Agregados por período)	Modelos <i>Randon Forest</i>
		Variáveis Individuais sobre cada óbito ocorrido no País	Cada Óbito Registrado no SIM	1980 a 2013 (Agregados por período)	Análise de Correspondência Múltipla

Quadro 1 – Resumo das Etapas de desenvolvimento do estudo segundo Indicadores Analisados, Unidade de Observação, Período e Métodos de Análise

As unidades de observação foram o Brasil e as cinco regiões e a mortalidade proporcional foi calculada considerando todos os óbitos registrados no país entre 1979 e 2013, excluindo-se aqueles com informações faltantes ($n = 29.182.133$), agregados em intervalos de 5 anos. Para preservar a separação dos anos em que cada revisão da Codificação Internacional de Doenças (CID) era utilizado no país, e manter todos os anos na análise, considerou-se os períodos de 1979 a 1984, 1985 a 1989, 1990 a 1995 referentes à CID – 9 e de 1996 a 1999, 2000 a 2004, 2005 a 2013 referentes à CID – 10.

Para a análise do padrão etário a unidade de observação foi cada óbito ocorrido no país no período de 1979 e 2013 ($n = 32.259.128$). A idade de óbito no Brasil e nas cinco regiões foi analisada considerando-se os óbitos em geral e também segundo Doenças Infecciosas e Parasitárias, Neoplasias, Doenças do Aparelho Circulatório, Doenças do Aparelho Respiratório e Causas Externas em cada ano do período.

Ainda na primeira etapa, para analisar a evolução das taxas de mortalidade segundo grupo de causa e causas selecionadas as unidades de observação foram as Unidades da Federação ($n = 27$). As taxas de mortalidade em cada ano entre 1980 e 2012, período selecionado devido à disponibilidade de dados populacionais, foram calculadas para cada sexo e padronizadas por faixa etária considerando como padrão a população do Brasil de cada sexo em 2000.

Os grupos de causa analisados foram Doenças Infecciosas e Parasitárias, Neoplasias, Doenças do Aparelho Circulatório, Doenças do Aparelho Respiratório e Causas Externas e as causas selecionadas foram Doenças Infecciosas Intestinais (códigos 001-009 da CID – 9 e A01-A09 da CID – 10), Neoplasias de Traqueia, Brônquios e Pulmão (códigos 162 da CID – 9 e C33-C34 da CID – 9), Pneumonias (códigos 480-487 da CID – 9 e J09-J18 da CID – 10), Doenças Cerebrovasculares (códigos 430-438 da CID – 9 e I60-I69 da CID – 10) e Homicídios (códigos 960-978 da CID – 9 e X85-99, Y00-09 e Y35-36 da CID – 10). Estas causas foram selecionadas por serem as mais frequentes em cada grupo, e se relacionar com diferentes aspectos da transição da mortalidade.

Na segunda etapa foi analisada a relação entre as taxas de mortalidade segundo causas selecionadas e indicadores sociais, econômicos e demográficos. As unidades de observação foram as Microrregiões do Brasil ($n = 557$) nos anos censitários de 1991, 2000 e 2010. A Microrregião de Fernando de Noronha foi excluída da análise por se tratar de uma área de preservação com população residente reduzida, o que gera flutuações nas taxas de mortalidade. A divisão territorial do Brasil em Mesorregiões e Microrregiões Geográficas é adotada pelo Instituto Brasileiro de Geografia e Estatística – IBGE desde 1990 (IBGE, 1990).

Uma Microrregião é definida como um grupo de municípios contíguos com características similares em relação às estruturas de produção e relações locais (IBGE, 1990; RODRIGUES et al., 2002).

As causas selecionadas foram as mesmas utilizadas na primeira etapa: Doenças Infecciosas Intestinais, Neoplasias de Traqueia, Brônquios e Pulmão, Pneumonias, Doenças Cerebrovasculares e Homicídios. Para estabilizar o número de óbitos utilizou-se a média dos períodos de 1990 a 1992, 1999 a 2001 e de 2009 a 2011 no cálculo das taxas de mortalidade referentes a 1991, 2000 e 2010, respectivamente. Foram calculadas taxas segundo sexo padronizadas por idade pelo método direto considerando como referência a população do Brasil de cada sexo no ano 2000. A população segundo sexo e faixa etária foram retiradas dos censos dos anos 1991, 2000 e 2010, assim como os indicadores apresentados no Quadro 2, selecionados por sua disponibilidade por microrregião nos três anos analisados.

Na terceira etapa, para discutir a implementação de técnicas de mineração de dados nas informações individuais do SIM para identificar padrões de mortalidade no Brasil, a unidade de observação foi cada óbito registrado no Brasil no período de 1979 a 2013, excluindo-se aqueles com informações faltantes ($n = 29.182.133$). Foram definidas seis abordagens das informações sobre mortalidade; na primeira todos os óbitos foram incluídos na análise enquanto nas demais foram analisados apenas os óbitos referentes a Doenças Infecciosas e Parasitárias, Neoplasias, Doenças do Aparelho Circulatório, Doenças do Aparelho Respiratório e Causas Externas, respectivamente.

As variáveis independentes em todas as abordagens foram Ano de Óbito, Sexo, Faixa Etária, Estado Civil, Local de Ocorrência do Óbito e Unidade da Federação de Residência, enquanto as variáveis dependentes foram os Grupos de Causas quando considerados todos os óbitos, e as Causas Específicas de Óbito nas demais abordagens. A qualidade de predição dos modelos foi analisada através das medidas de Sensibilidade, Especificidade, Valor Preditivo Positivo, Valor Preditivo Negativo e Acurácia.

INDICADOR	CONCEITO	MÉTODO DE CÁLCULO
Taxa de Analfabetismo	Representa o percentual de indivíduos com 15 ou mais anos de idade que no sabem ler e escrever pelo menos um bilhete simples	$\frac{N^{\circ} \text{ de Indivíduos Analfabetos}}{\text{População Residente}} \times 100$
Proporção da População que Cursou Ensino Médio ou Mais	Representa o percentual de indivíduos com 15 ou mais anos de idade que estudou 8 mais anos ou 2º ciclo fundamental completo ou mais	$\frac{N^{\circ} \text{ de Indivíduos que cursou ensino médio ou mais}}{N^{\circ} \text{ de Indivíduos com 15 ou mais anos de idade}} \times 100$
Renda Média Domiciliar Per Capita	Média das rendas domiciliares per capita*	$\frac{\sum \text{ Renda Domiciliar Média}}{\text{População Residente}}$
Proporção de Pessoas com Baixa Renda	Representa o percentual da população residente com renda domiciliar mensal per capita* de até meio salário-mínimo	$\frac{N^{\circ} \text{ de Indivíduos com renda inferior a meio salário mínimo}}{\text{População Residente}} \times 100$
Taxa de Desemprego	Representa o percentual da população economicamente ativa, de 16 ou mais anos de idade, que se encontravam sem trabalho	$\frac{N^{\circ} \text{ de Indivíduos de 16 ou mais anos de idade sem trabalho}}{\text{População com de 16 ou mais anos de idade}} \times 100$
Percentual de Domicílios com acesso à Rede Geral de Esgoto ou Pluvial	Representa a proporção de domicílios que tinham como instalação sanitária a rede geral de esgoto ou pluvial	$\frac{N^{\circ} \text{ de Domicílios com acesso a Rede Geral de Esgoto ou Pluvial}}{N^{\circ} \text{ total de Domicílios}} \times 100$
Índice de Envelhecimento	Representa o número de idosos para cada cem jovens	$\frac{N^{\circ} \text{ de Indivíduos com 60 ou mais anos de idade}}{N^{\circ} \text{ de Indivíduos com menos de 15 anos de idade}} \times 100$
Razão de Dependência de Jovens	Mede a participação do contingente populacional inativo jovem que deveria ser sustentado pela população potencialmente produtiva	$\frac{N^{\circ} \text{ de Indivíduos com menos de 15 anos de idade}}{N^{\circ} \text{ de Indivíduos entre 15 e 59 anos de idade}} \times 100$
Razão de Dependência de Idosos	Mede a participação do contingente populacional inativo idoso que deveria ser sustentado pela população potencialmente produtiva	$\frac{N^{\circ} \text{ de Indivíduos com 60 ou mais anos de idade}}{N^{\circ} \text{ de Indivíduos entre 15 e 59 anos de idade}} \times 100$

Quadro 2 – Indicadores Sociais, Econômicos e Demográficos a serem utilizados nas análises por microrregião e seus respectivos conceitos e métodos de cálculo

* A renda domiciliar per capita é dada pela soma dos rendimentos mensais dos residentes do domicílio dividida pelo número de residentes do mesmo. Os valores referentes a renda são padronizados e corrigidas com base no Índice Nacional de Preço ao Consumidor – INPC de julho de 2010

5.2 PROCEDIMENTOS DE ESTRUTURAÇÃO DA BASE DE DADOS E SELEÇÃO DE VARIÁVEIS – PRÉ-PROCESSAMENTO

Os dados do Sistema de Informação Sobre Mortalidade (SIM) podem ser obtidos através do Departamento de Informática do SUS (DataSUS) no endereço <http://www2.datasus.gov.br/DATASUS/index.php?rea=0901&item=1&acao=26&pad=31655>. O sistema disponibiliza *links* para *download* de um arquivo compactado (com extensão “.zip”) contendo arquivos selecionados através de uma interface “apontar e clicar”, sendo o tamanho máximo permitido para este arquivo compactado 100 *MegaBytes* e todo o procedimento de definição do tipo de dado, ano e UF, e seleção dos arquivos desejados deve ser repetido após cada *download*.

A base de dados refere-se ao período de 1979 a 2013 e engloba 32.259.128 Declarações de Óbito, correspondendo a 935 tabelas que totalizam cerca de 1 *Gigabyte* de informação. Estas tabelas apresentam estruturas diferentes, uma vez que as mudanças nas Declarações de Óbito e no SIM desde sua criação acarretaram a inclusão, substituição e atualização das informações coletadas. A adoção da 10ª Revisão da Codificação Internacional de Doenças – CID-10 a partir 1996 em substituição à CID-9 é uma das mudanças mais marcantes e o DataSUS utiliza as versões da CID adotadas em cada período como referência para diferenciar os arquivos auxiliares para tabulação, com a codificação das variáveis do sistema.

Considerando os desafios impostos pela variedade, volume, e incompatibilidade das informações do sistema ao longo dos anos, foi definida uma estratégia, esquematizada na Figura 5, para a estruturação da base de dados utilizada. Os procedimentos foram realizados em um computador com processador *Core i5* de 4 núcleos com 32 *GigaBytes* de memória *RAM* e sistema operacional *Linux*.

Para contornar a limitação do tráfego de dados e evitar a repetição dos procedimentos relacionados à obtenção de arquivos referentes às DO, foi realizado o *download* recursivo diretamente dos servidores “FTP” do DataSUS, diretórios de acesso aberto e que contém todos os arquivos relativos ao SIM, disponíveis no endereço <ftp://ftp.datasus.gov.br/dissemin/publicos/SIM/>, utilizando o *GNU Wget*, *software* livre de transferência de arquivos que suporta diversos protocolos.

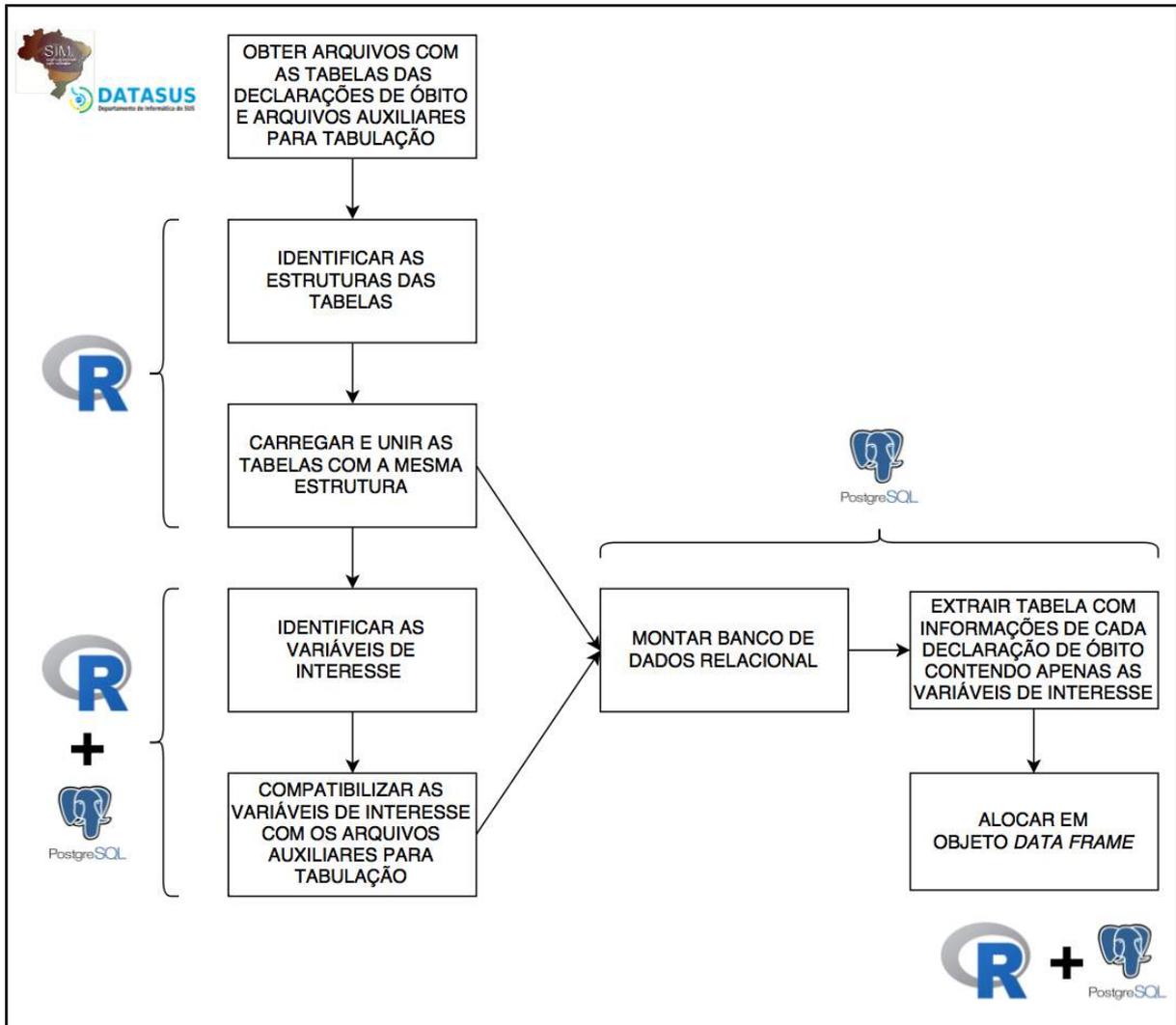


Figura 5 – Estratégia para estruturação do Banco de Dados

Os arquivos do SIM são disponibilizados com a extensão “.DBC”, formato compactado que geralmente deve ser expandido para o formato “.DBF”, através da aplicação *DBFtoDBC* disponibilizada pelo DataSUS, para que seja reconhecido pela maioria dos *softwares* de análise e manipulação de dados. Neste estudo foi utilizada a biblioteca *read.dbc* do R (PETRUZALEK, 2016) que permite carregar arquivos com a extensão “.DBC” sem a necessidade de expandi-los, poupando tempo de processamento e espaço em disco rígido.

Arquivos referentes a cada ano foram carregados para verificar as variáveis disponíveis e identificar a estrutura de dados. Das onze estruturas observadas, duas referem-se ao período em que se utilizava a CID-9 (1979 a 1995) e nove à CID-10 (1996 em diante), como mostra a Figura 6.

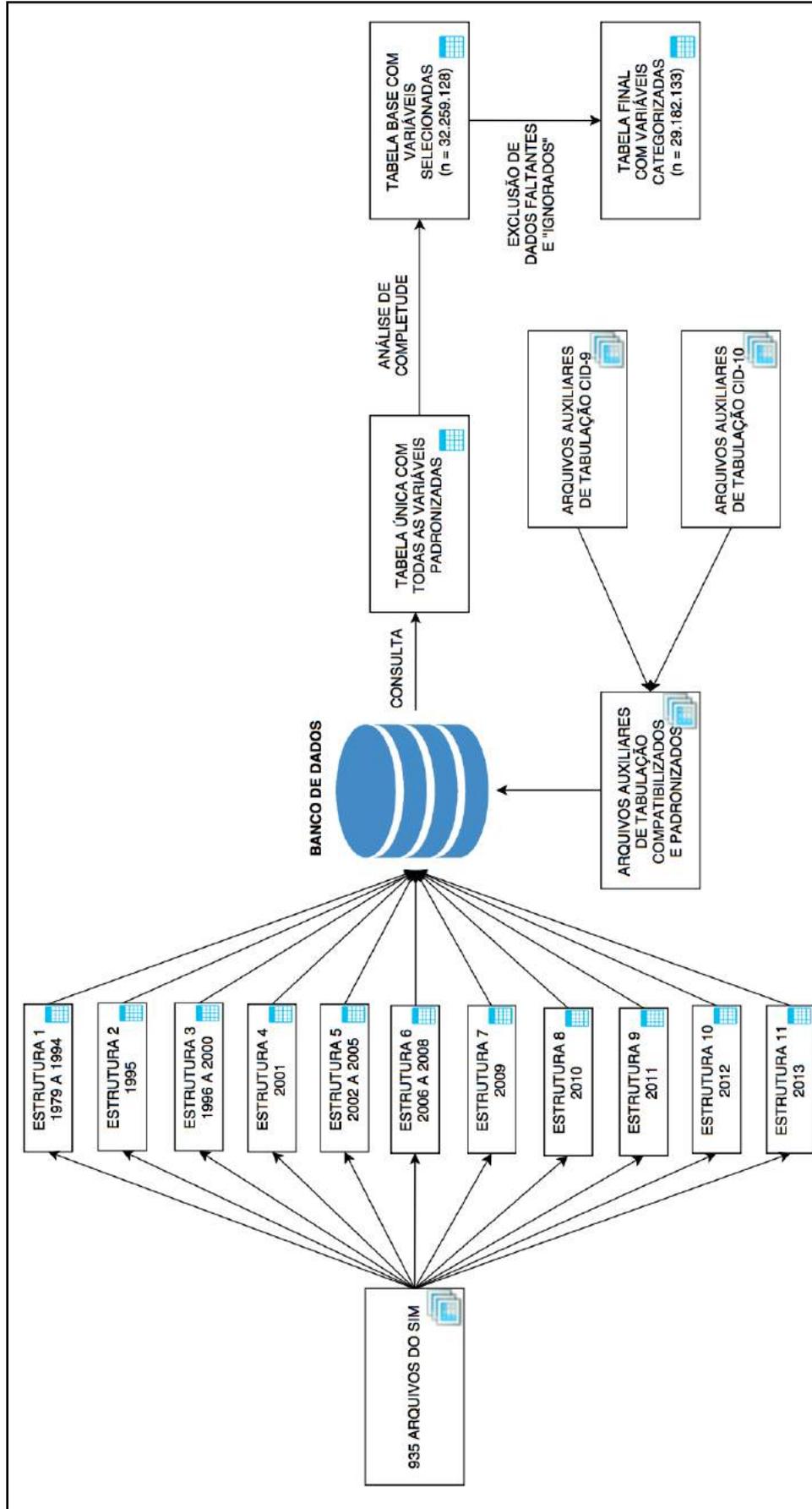


Figura 6 – Esquematização dos procedimentos para estruturação do Banco de Dados

Nesta etapa foram selecionadas as variáveis disponíveis para todo o período e que poderiam ser utilizadas nas análises deste estudo: Ocupação, Tipo de Óbito, Sexo, Idade do óbito, Data de Óbito, Estado Civil, Local de Ocorrência do Óbito, Instrução/Escolaridade, Assistência Médica, Cirurgias Realizadas, Exames Realizados, Ocupação da Mãe e Instrução/Escolaridade da Mãe, Município de Residência, Causa Básica do Óbito.

Uma rotina automatizada foi implementada no *software* R para carregar e unir as tabelas dos arquivos correspondentes a cada estrutura em uma tabela única por estrutura e enviar para o *PostgreSQL*, *software* gerenciador de banco de dados, utilizando a biblioteca *Rpostgresql*.

Visando abordar de maneira padronizada as variáveis selecionadas em todo o período analisado, foram criadas tabelas que relacionavam a codificação original da variável em seu período correspondente com novas categorias compatibilizadas, quando necessário (Quadro 3, Quadro 4 e Quadro 5).

ESTADO CIVIL		
CID-9	CID-10	COMPATIBILIZAÇÃO
Solteiro	Solteiro	Solteiro
Casado	Casado ou União Consensual	Casado
Viúvo	Viúvo	Viúvo
Separado Judicialmente	Separado	Separado
Outro ou Ignorado	Ignorado ou Não Informado	Ignorado ou Não Informado

Quadro 3 – Categorias da variável “Estado Civil” nos anos da CID 9 e CID 10 e compatibilização adotada

LOCAL DE OCORRÊNCIA		
CID-9	CID-10	COMPATIBILIZAÇÃO
Hospital	Hospital/Outro Estabelecimento de Saúde	Estabelecimentos de Saúde
Domicílio	Domicílio	Domicílio
Via Pública	Via Pública	Via Pública
Outros	Outros	Outros
Ignorado	Ignorado	Ignorado

Quadro 4 – Categorias da variável “Local de Ocorrência” nos anos da CID 9 e CID 10 e compatibilização adotada

GRAU DE INSTRUÇÃO E ESCOLARIDADE		
CID-9	CID-10	COMPATIBILIZAÇÃO
Nenhuma	De 1 a 3 anos de estudo Nenhum	Nenhuma
1º Grau	De 4 a 7 anos de estudo De 1 a 8 anos de estudo	1º Ciclo Fundamental
2º Grau	De 8 a 11 anos de estudo De 9 a 11 anos de estudo	2º Ciclo Fundamental
Superior	De 12 ou mais anos de estudo	Superior ou Técnico
Ignorado	Ignorado	Ignorado

Quadro 5 – Categorias das variáveis “Grau de Instrução” e “Escolaridade” nos anos da CID 9 e CID 10 e compatibilização adotada

Procedimento equivalente foi realizado para compatibilizar os grupos de causa e causas específicas de óbito referentes a CID-9 e CID-10 (Quadro 6 e Quadro 7, respectivamente) e os Municípios para Microrregiões, Unidades da Federação e Região. Para as variáveis que não houve mudança na categorização ao longo dos anos foram utilizados os arquivos auxiliares de tabulação.

GRUPOS DE CAUSA DE ÓBITO		
CID-9	CID-10	COMPATIBILIZAÇÃO
Capítulo I Códigos: 001-139	Capítulo I Códigos: A00-B99	Algumas Doenças Infecciosas e Parasitárias
Capítulo II Códigos: 140-239	Capítulo II Códigos: C00-D48	Neoplasias
Capítulo VII Códigos: 390-459	Capítulo IX Códigos: I00-I99	Doenças do Aparelho Circulatório
Capítulo VIII Códigos: 460-519	Capítulo X Códigos: J00-J99	Doenças do Aparelho Respiratório
Capítulo XV Códigos: 760-779	Capítulo XVI Códigos: P00-P96	Algumas Afecções Originadas no Período Perinatal
Capítulo XVII Códigos: 800-999	Capítulo XX Códigos: V01-Y98	Causas Externas
Capítulo XVI Códigos: 001-139	Capítulo XVIII Códigos: R00-R99	Causas Mal Definidas
Demais Capítulos e Códigos CID-9	Demais Capítulos e Códigos CID-10	Demais Causas de Óbito

Quadro 6 – Compatibilização dos Grupos de Causa na CID-9 e na CID-10

GRUPO	CAUSAS ESPECÍFICAS	CÓDIGOS CID-9	CÓDIGOS CID-10
INFECCIOSAS	Infecções Intestinais	0010 – 0099	A000 – A099
	Tuberculose	0100 – 0189	A150 – A199
	Outras Doenças Bacterianas	0200 – 0419	A200 – A649
	Doenças Virais	0450 – 0799	A800 – A999 B000 – B249
	Doenças devido a Protozoários	0800 – 0899	B500 – B649
	Demais Causas Infecciosas	Demais Códigos Referentes às Doenças Infecciosas	
NEOPLASIAS	Neoplasias de Esôfago	1500 – 1509	C150 – C159
	Neoplasia de Estômago	1510 – 1519	C160 – C169
	Neoplasia de Cólon	1530 – 1539	D180 – C219
	Neoplasia de Laringe	1610 – 1619	C320 – C329
	Neoplasia de Pulmão	1620 – 1629	C330 – C349
	Neoplasia de Mama	1740 – 1749	C500 – C509
	Neoplasia de Colo de Útero	1800 – 1809	C530 – C539
	Neoplasia de Próstata	1850 – 1859	C610
	Leucemia	2040 - 2089	C910 – C959
	Código Lixo	1950; 1990	C760 – C769; C80; C97
	Demais Neoplasias	Demais Códigos Referentes às Neoplasias	
DOENÇAS DO APARELHO CIRCULATÓRIO	Doenças Hipertensivas	4010 – 4059	I100 – I159
	Infarto Agudo do Miocárdio	4100 – 4109	I210 – I219
	Outras Isquêmicas do Coração	4110 – 4149	I200 – I209; I220 – I259
	Doenças Cerebrovasculares	4300 – 4389	I600 – I699
	Código Lixo	4271; 4274 – 4275; 4280; 4290 – 4292; 4299; 4409	I460 – I469; I472; I490; I500; I514 – I516; I519; I709
	Demais Doenças Circulatórias	Demais Códigos Referentes às Doenças Circulatórias	
DOENÇAS DO APARELHO RESPIRATÓRIO	Doenças Agudas VA Inferiores	4660 – 4669	J200 – J229
	Pneumonia	4800 – 4869	J120 – J189
	Gripe	4870 – 4879	J100 – J119
	Doenças Crônicas VA Inferiores	4900 – 4939	J400 – J479
	Demais Doenças Respiratórias	Demais Códigos Referentes às Doenças Respiratórias	
CAUSAS EXTERNAS	Acidentes de Transporte	8000 – 8489	V010 – V999
	Suicídios e Lesões Auto Infligidas	9500 – 9599	X600 – X849
	Homicídios	9600 – 9699	X850 – X999 Y000 – Y099
	Código Lixo	9800 – 9899	Y100 – Y349; Y872
	Demais Causas Externas	Demais Códigos Referentes às Causas Externas	

Quadro 7 – Compatibilização das Causas Específicas de Óbito segundo CID-9 e CID-10

Tabelas com a compatibilização das variáveis selecionadas foram enviadas para o *PostgreSQL* e relacionadas com as tabelas referentes às DO. Uma consulta para extração de uma Tabela única composta pelas variáveis de interesse já compatibilizadas permitiu agregar as informações dos cerca de 32 milhões de registros referentes às onze estruturas identificadas inicialmente, alocadas em um objeto *data frame* no *R* utilizando a biblioteca *Rpostgresql*.

A completude de variáveis e a concentração da categoria “ignorado” foram analisadas a partir da Tabela única. As variáveis Assistência Médica, Cirurgias Realizadas, Exames Realizados, Ocupação da Mãe e Instrução/Escolaridade da Mãe foram excluídas em função da elevada proporção de informação ignorada (Quadro 8). A variável Ocupação foi excluída por apresentar elevado número de categorias com baixa frequência de ocorrência, e a variável Instrução/Escolaridade foi excluída devido a perda seletiva de informação (Apêndice A).

VARIÁVEIS	PERCENTUAL DE NA E “IGNORADO”
Instrução/Escolaridade da Mãe*	87,3 %
Ocupação da Mãe*	86,9 %
Cirurgias*	57,3 %
Exames*	51,8 %
Assistência Médica*	43,4 %
Instrução/Escolaridade*	37,6 %
Ocupação*	23,7 %
Estado Civil	6,3 %
Local de Ocorrência do Óbito	1,4 %
Sexo	0,2 %

Quadro 8 – Percentual somado de registros faltantes (NA) e informação “Ignorada” das variáveis. As variáveis com (*) foram excluídas

O Quadro 9 apresenta as variáveis que compuseram a Tabela Base, a partir da qual foi obtida a contagem de óbitos por grupo de causa, estratificados por sexo e faixa etária, para o cálculo das taxas de mortalidade padronizadas. Os registros com dados faltantes (“ignorado”) foram excluídos para obtenção da Tabela Final, em que variáveis foram categorizadas para o cálculo da mortalidade proporcional segundo grupos de causa e para implementação de técnicas de Mineração de Dados.

VARIÁVEL	TIPO DE VARIÁVEL	VARIÁVEL BASE / CATEGORIAS
Idade	Variável Numérica	Criada a partir da Idade de Óbito
Ano de Óbito	Variável Numérica	Criada a partir da Data de Óbito
Mês de Óbito	Variável Categórica	Criada a partir da Data de Óbito
Microrregião de Residência	Variável Categórica	Criada a Partir do Município de Residência
Unidade da Federação de Residência	Variável Categórica	Criada a Partir do Município de Residência
Região de Residência	Variável Categórica	Criada a Partir do Município de Residência
Estado Civil	Variável Categórica	Compatibilizada conforme Quadro 1
Local de Ocorrência do Óbito	Variável Categórica	Compatibilizada conforme Quadro 2
Grupos de Causa de Óbito	Variável Categórica	Criada a Partir da Causa Básica de Óbito e Compatibilizada conforme Quadro 4
Causas Específicas de Óbito	Variável Categórica	Criada a Partir da Causa Básica de Óbito e Compatibilizada conforme Quadro 6

Quadro 9 – Variáveis mantidas na Tabela Base

Apesar do número reduzido de variáveis utilizado neste estudo, o procedimento de identificação das diferentes estruturas das tabelas do SIM ao longo dos anos permitiu a organização de um banco de dados estruturado com todas as informações já disponibilizadas pelo Sistema desde sua criação. Ao unir as tabelas contendo as informações das Declarações de Óbito por estrutura identificada o grande volume de dados foi particionado, o que diminuiu o custo computacional para carregar os arquivos no R e enviá-los para o *PostgreSQL*.

A Tabela Final aloca cerca de 5 *Gigabytes* de memória *RAM* quando carregado no R; a manipulação de variáveis deste objeto consomem no mínimo mais 2 *Gigabytes*, de maneira que para realizar análises simples destas informações é indicada a disponibilidade de pelo menos 16 *Gigabytes* de *RAM*. Parte das análises realizadas neste estudo foi implementada no mesmo computador onde os procedimentos para a estruturação do banco de dados foram realizados, contudo, para a implementação das técnicas da terceira etapa a configuração não foi suficiente, sendo utilizada uma máquina com 128 *GigaBytes* de memória *RAM*.

5.3 ANÁLISE DE DADOS

As informações utilizadas nas três etapas foram retiradas da Tabela Base e da Tabela Final e as análises realizadas com o *software* R. A seguir serão descritos os procedimentos para extração e estruturação dos dados, os métodos de análise e as bibliotecas do R utilizadas em cada etapa.

5.3.1 ETAPA I: ANÁLISE DOS PADRÕES E TENDÊNCIAS DE MORTALIDADE SEGUNDO GRUPOS DE CAUSA E CAUSAS SELECIONADAS NO BRASIL

A Figura 7 sumariza a estruturação de dados utilizados na criação de fluxogramas. A mortalidade proporcional no Brasil e nas cinco Regiões segundo grupos de causa analisados – Doenças Infecciosas e Parasitárias, Neoplasias, Doenças do Aparelho Circulatório, Doenças do Aparelho Respiratório, Causas Externas, Causas Mal Definidas e Demais Causas de Óbito – foi ordenada de maneira decrescente, sendo atribuído um *ranque* de 1^a a 7^a causa mais frequente em cada período e organizadas de maneira que “caixas” referentes a cada um dos grupos de causas foram dispostas em seis “colunas” representando os períodos. Os dados foram tabulados e ordenados no R e os fluxogramas criados utilizando a plataforma *Daw.io*, disponível em <https://www.draw.io/>.

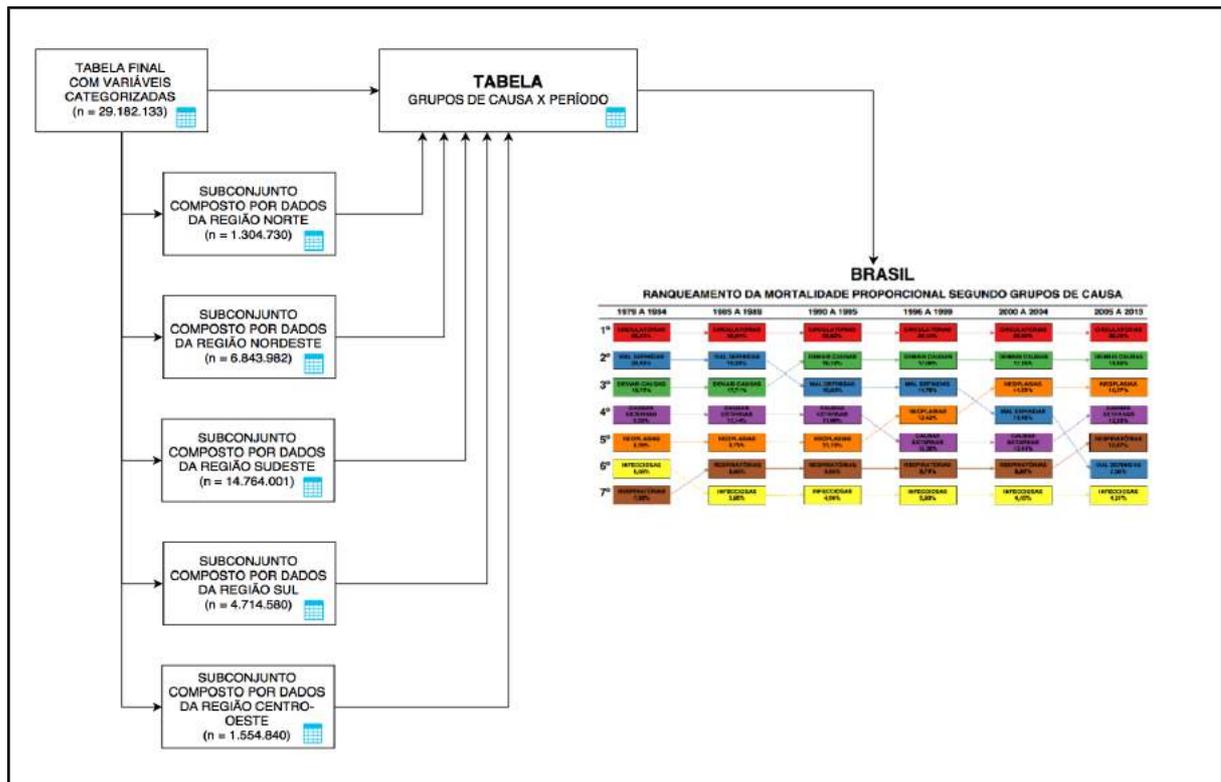


Figura 7 – Detalhamento do preparo dos dados e para construção dos Fluxogramas do *ranking* da mortalidade proporcional segundo grupos de causa

Na Figura 8 se encontra de maneira detalhada a organização dos dados para o desenvolvimento dos Diagramas de Kiviat. Foram obtidos seis subconjuntos da Tabela Final, cada um composto por dados referentes aos períodos de 1979 a 1984, 1985 a 1989, 1990 a 1995, 1996 a 1999, 2000 a 2004 e de 2005 a 2013, respectivamente, considerando apenas os óbitos por Doenças Infecciosas e Parasitárias, Neoplasias, Doenças do Aparelho Circulatorio, Doenças do Aparelho Respiratório e Causas Externas.

Foram criados seis diagramas referentes aos períodos analisados, onde os eixos representaram as regiões e os polígonos sobrepostos apresentaram a proporção de óbito por cada grupo de causa. Esta organização permitiu observar a evolução da mortalidade proporcional segundo cada grupo de causa no país e como as diferenças regionais foram configuradas neste processo. Os gráficos foram criados utilizando a biblioteca *radarchart* (ASHTON et al., 2016) do R.

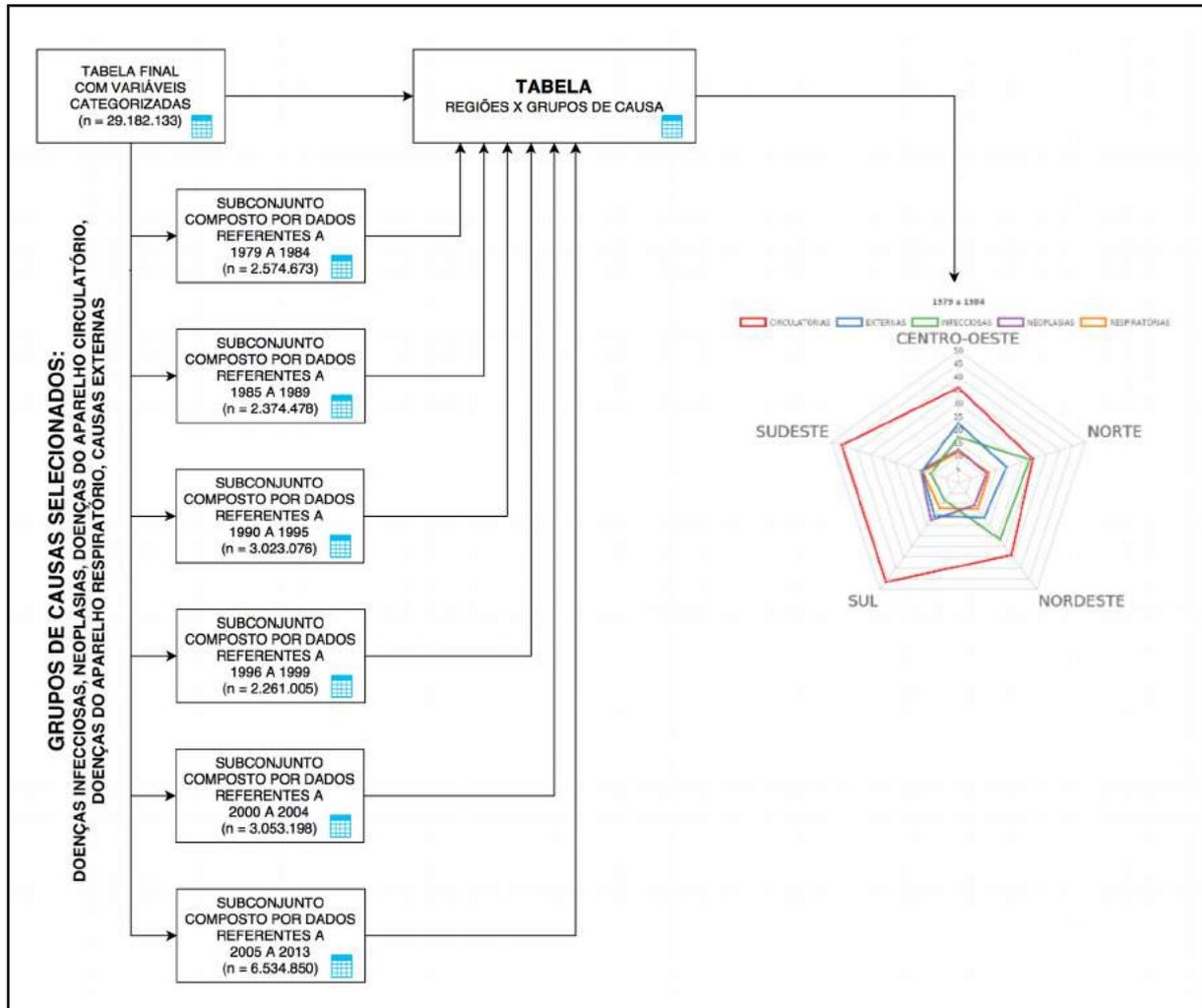


Figura 8 – Detalhamento do preparo dos dados para construção dos Diagramas de Kiviat

Para analisar o padrão etário dos óbitos segundo grupos de causa os dados foram organizados conforme detalhado na Figura 9. Cinco subconjuntos foram extraídos Tabela Base, cada um contendo apenas as informações dos óbitos segundo cada grupo de causa analisado – Doenças Infecciosas e Parasitárias, Neoplasias, Doenças do Aparelho Circulatório, Doenças do Aparelho Respiratório e Causas Externas – a partir dos quais foram criados seis gráficos compostos por 35 *boxplots*, cada um representando a distribuição da idade dos óbitos em cada ano no período de 1979 a 2013, referentes ao Brasil e cada Região. Os gráficos referentes ao Brasil e cada Região foram apresentados em conjunto, segundo grupo de causa, para facilitar a visualização das diferenças regionais.

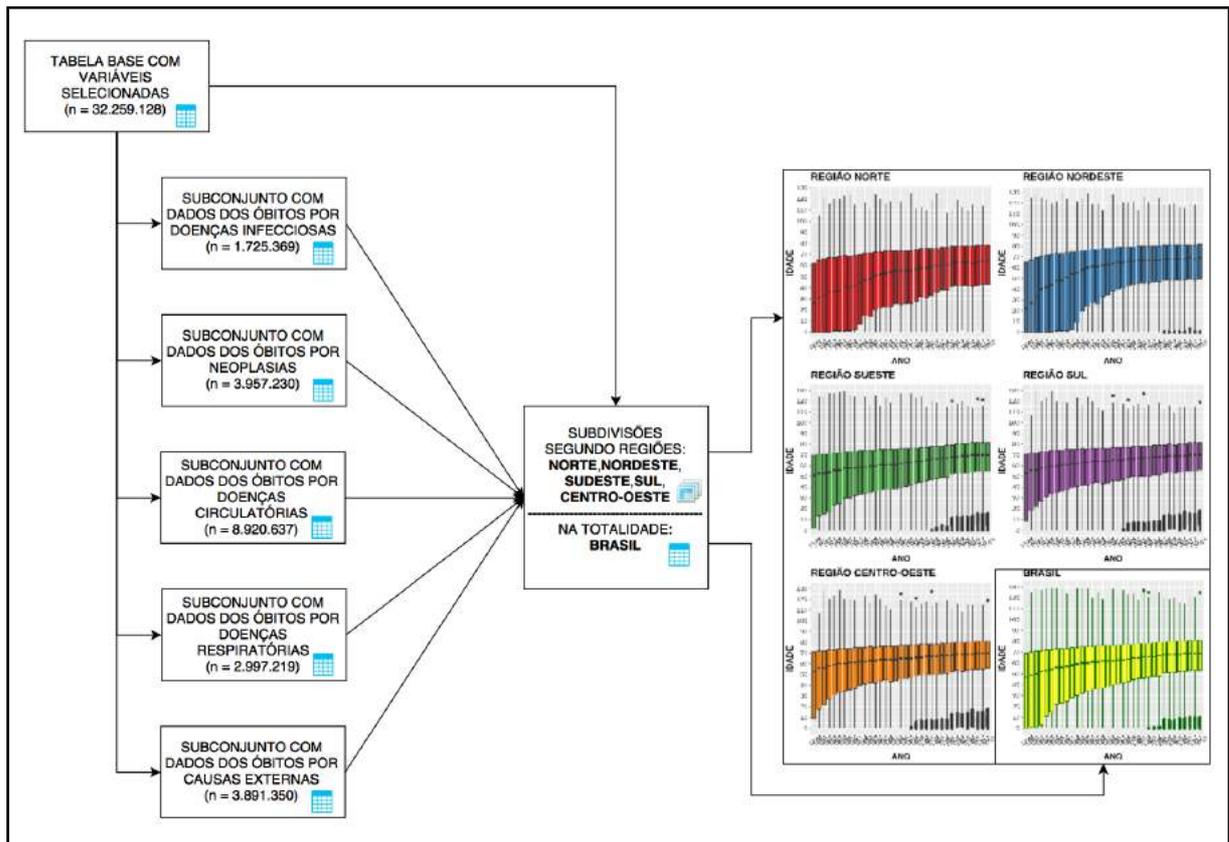


Figura 9 – Organização dos dados para análise do padrão etário de óbito segundo grupo de causa no Brasil e Regiões

A Figura 10 apresenta os procedimentos para obtenção das taxas padronizadas de mortalidade analisadas através de *heatmaps*. As contagens dos óbitos de cada um dos grupos de causa e causas selecionadas nas Unidades da Federação segundo faixa etária e sexo referentes a cada para cada ano do período de 1980 a 2013 foram extraídas da Tabela Base e as estimativas populacionais das Unidades da Federação no período segundo faixa etária e sexo produzidas pelo IBGE, foram retiradas do DataSUS. Taxas de Mortalidade, padronizadas segundo faixa etária pelo método direto, referentes a cada sexo e também para os óbitos em geral no período foram calculadas utilizando como padrão a população do Brasil segundo sexo em 2000. As taxas referentes ao Tocantins foram calculadas a partir de 1989, pois até 1988, ano da criação desta Unidade da Federação, não existiam óbitos registrados.

Foram criados três *heatmaps* referentes às taxas de mortalidade de cada sexo e em geral segundo grupo de causa – Doenças Infecciosas e Parasitárias, Neoplasias, Doenças do Aparelho Circulatório, Doenças do Aparelho Respiratório e Causas Externas – e causa selecionada – Doenças Infecciosas Intestinais, Neoplasias de Traqueia, Brônquios e Pulmão, Pneumonias, Doenças Cerebrovasculares e Homicídios. Os *heatmaps* foram construídos utilizando a biblioteca *ggplot2* (WICKHAM, 2009) do R.

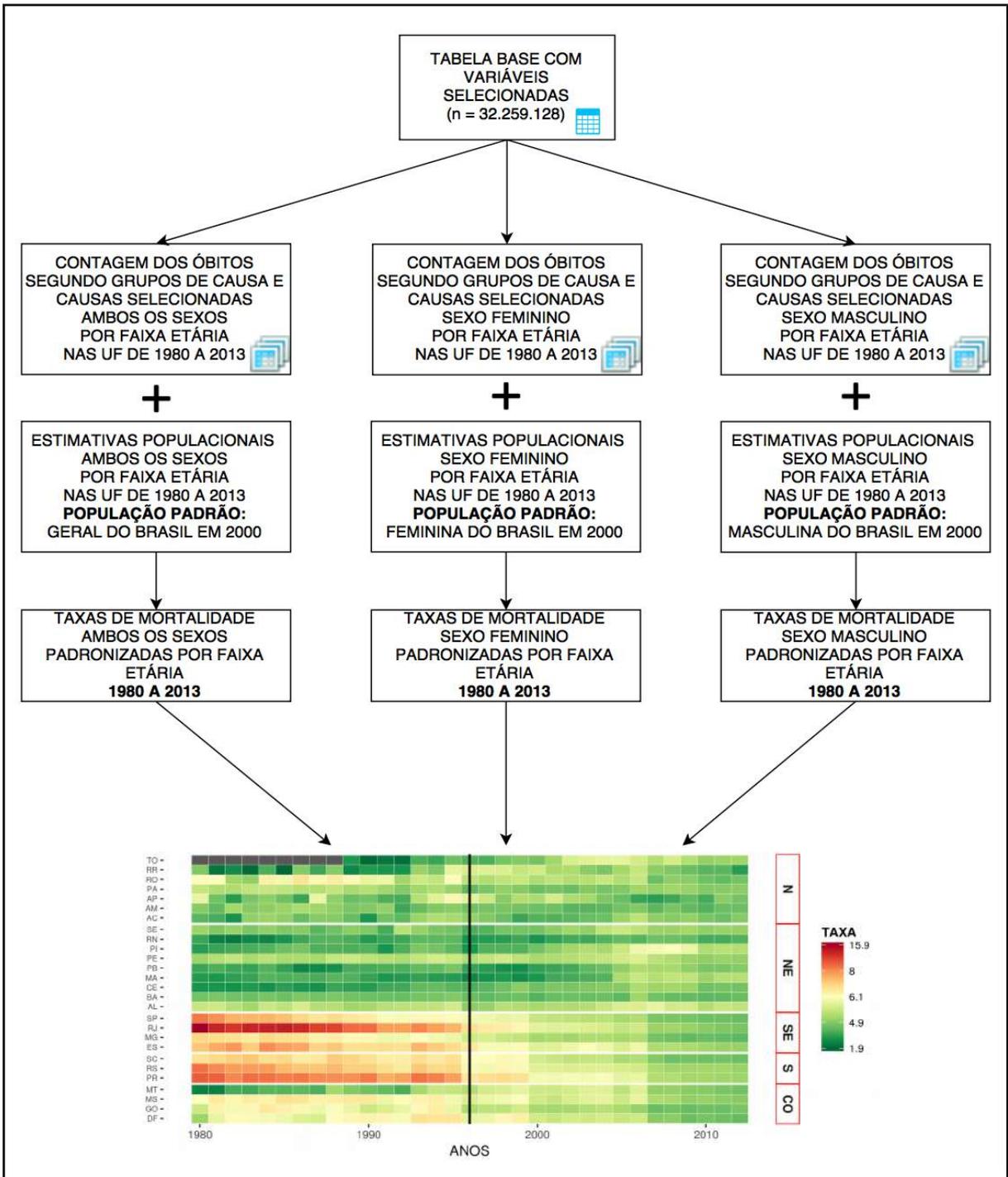


Figura 10 – Procedimentos para obtenção das taxas padronizadas de mortalidade e um exemplo de *HeatMap*

5.3.2 ETAPA II: ANÁLISE DA TENDÊNCIA DAS TAXAS DE MORTALIDADE POR CAUSAS SELECIONADAS E IDENTIFICAÇÃO DOS INDICADORES SOCIAIS, ECONÔMICOS E DEMOGRÁFICOS RELACIONADOS

Os procedimentos para obtenção das informações sobre mortalidade e os indicadores sociais, econômicos e demográficos estão sumarizados na Figura 11. As contagens dos óbitos pelas causas selecionadas – Doenças Infecciosas Intestinais, Neoplasias de Traqueia, Brônquios e Pulmão, Pneumonias, Doenças Cerebrovasculares e Homicídios – segundo microrregiões, sexo e faixa etária foram obtidas da Tabela Base e utilizadas no cálculo de taxas de mortalidade por sexo, padronizadas por idade pelo método direto, em 1991, 2000 e 2010. A população segundo sexo e faixa etária foram retiradas dos censos dos anos 1991, 2000 e 2010, assim como os indicadores sociais, econômicos e demográficos.

Foi realizada uma Análise Fatorial Exploratória, método que permite identificar grupos de variáveis correlacionadas representadas por fatores não correlacionados entre si (HAIR et al., 2009), para subsidiar a seleção dos indicadores socioeconômicos.

A definição do número de fatores para o ano de 1991 se deu pelo critério da raiz latente, extraído-se o mesmo número de fatores nos anos subsequentes. O critério da raiz latente indica a seleção dos fatores com autovalor (*eigenvalue*) superior a 1. O autovalor representa a variância explicada pelo fator de maneira que quanto maior seu valor, maior o número de variáveis correlacionadas ao fator (HAIR et al., 2009; DAMÁSIO, 2012). Os indicadores menos correlacionados entre si foram selecionados com base na carga fatorial e variância específica, que representa a proporção da variância de cada variável não explicada pela variabilidade das demais variáveis (DAMÁSIO, 2012). Quanto maior a variância específica do indicador, menor sua participação na composição dos fatores, indicando menor correlação com os demais indicadores.

As taxas de mortalidade e os indicadores selecionados foram analisados através de *boxplots* segundo ano e região. Para analisar a associação entre os indicadores e as taxas de mortalidade foi utilizado o Coeficiente de Correlação Linear de Pearson, que varia de -1 a +1 indicando a força e a direção da relação linear de duas variáveis (AYRES et al., 2007), apresentados em gráficos com círculos proporcionais ao valor dos coeficientes onde tons em vermelho representaram correlação positiva e em azul a negativa. Os gráficos com *boxplots* e representando os coeficientes de correlação foram criados com a biblioteca *ggplot2* (WICKHAM, 2009) do R.

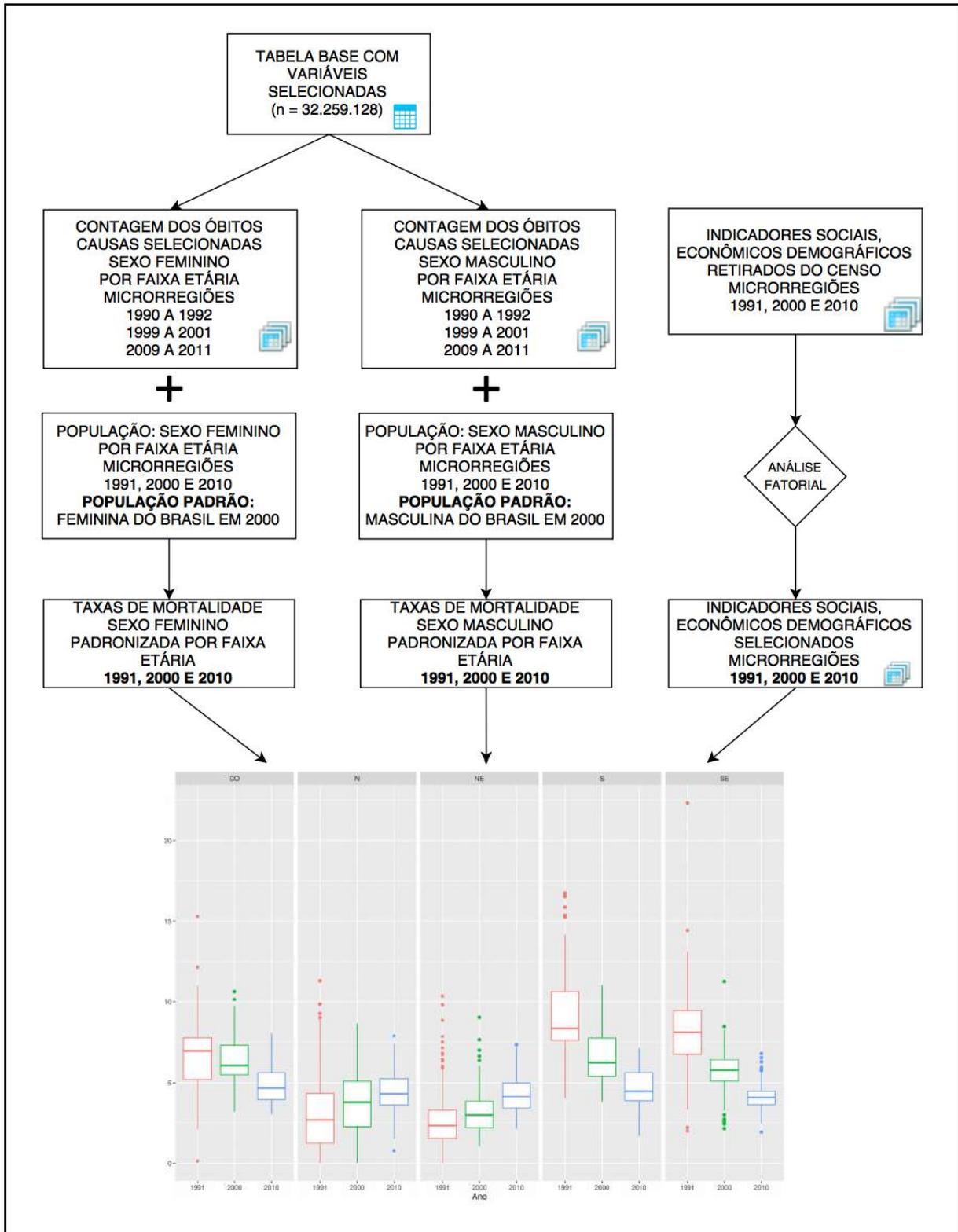


Figura 11 – Procedimentos para obtenção das taxas de mortalidade segundo sexo padronizadas por faixa etária e dos indicadores sociais, econômicos e demográficos nas microrregiões em 1991, 2000 e 2010

5.3.3. ETAPA III: IMPLEMENTAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS NAS INFORMAÇÕES INDIVIDUAIS DO SIM PARA IDENTIFICAÇÃO DE PADRÕES DE MORTALIDADE

Nesta etapa as unidades de observação foram os óbitos registrado no SIM no período de 1979 a 2013 extraídos da Tabela Final em sua totalidade e subconjuntos compondo as seis abordagens apresentadas na Figura 12. Foram ajustados modelos de árvore (*CART*) e *Random Forest* e realizada uma Análise de Correspondência Múltipla em cada abordagem.

Os óbitos nas faixas etárias de 0 a 29 anos foram excluídos das análises abordando os óbitos por Neoplasias e os óbitos por Doenças do Aparelho Circulatório, e na faixa etária de menos de 1 ano quando abordado os óbitos por Causas Externas devido à baixa frequência, conforme apresentado no Apêndice B. Os óbitos na faixa etária de 60 a 79 anos e de 80 anos ou mais foram somados na abordagem dos óbitos por Causas Externas.

Os Grupo de Causas de Óbito foram analisados na primeira abordagem como variável dependente dos modelos *CART* e *Random Forest*, tendo como categorias Doenças Infecciosas e Parasitárias, Neoplasias, Doenças do Aparelho Circulatório, Doenças do Aparelho Respiratório, Causas Externas, Causas Mal Definidas e Demais Grupos de Causa.

As Causas Específicas de Óbito foram utilizadas como variável dependente nas demais abordagens e para selecioná-las foi realizada uma análise exploratória dos dados utilizados em cada abordagem, considerando as categorias apresentadas no Quadro 5, sendo mantidas aquelas que representavam pelo menos 10% dos óbitos no grupo. As causas específicas com frequência inferior a 10% foram somadas às “Demais Causas”. As Causas Específicas utilizadas nestas abordagens estão discriminadas no Quadro 10, e a distribuição da proporção de cada uma das Causas Específicas se encontram no Apêndice C.

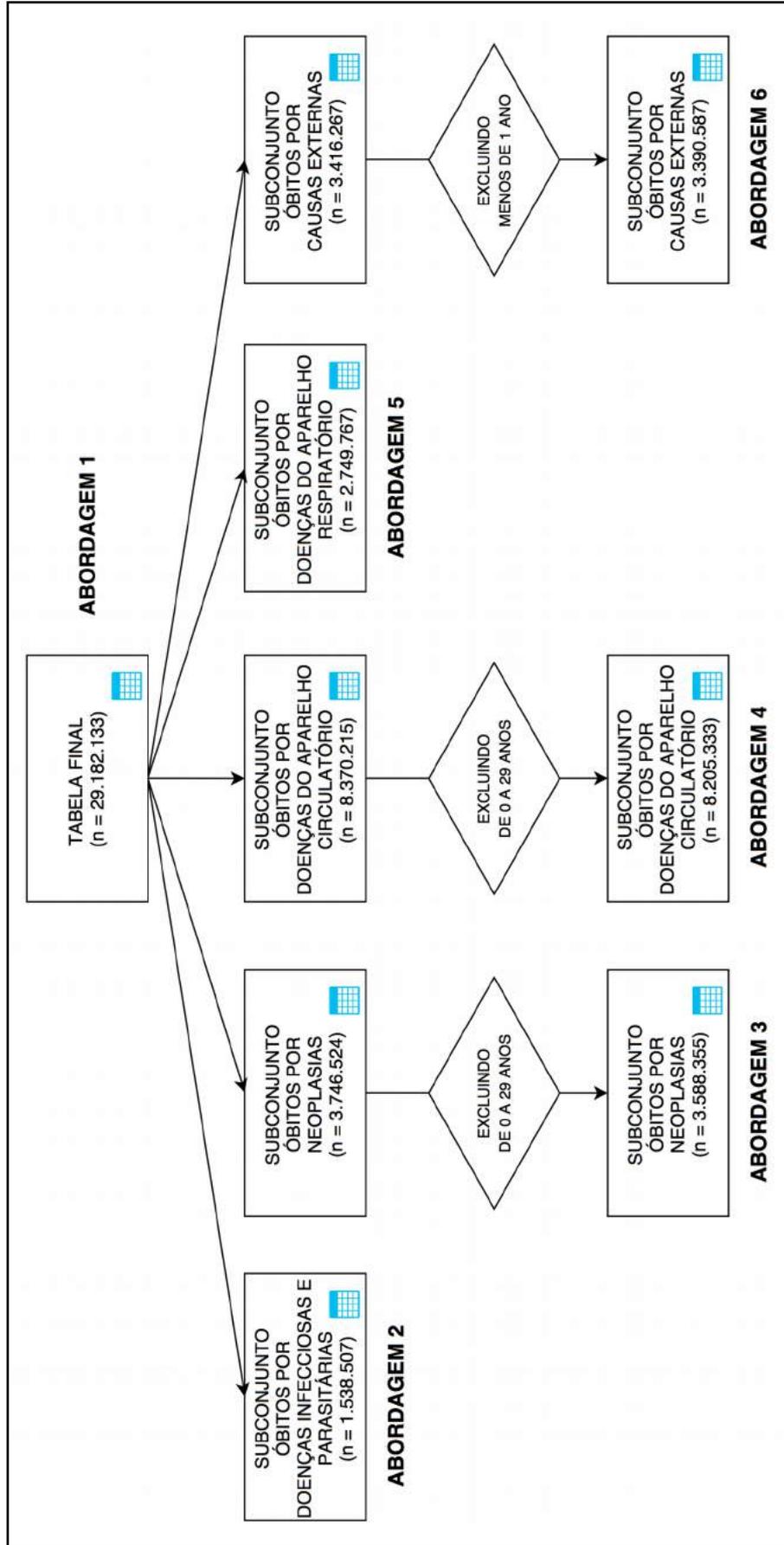


Figura 12 – Dados utilizados em cada abordagem da Etapa III

ABORDAGEM/GRUPO DE CAUSA	CAUSAS ESPECÍFICAS UTILIZADAS
ABORDAGEM 2: INFECCIOSAS E PARASITÁRIAS	Doenças Infecciosas Intestinais
	Tuberculose
	Outras Doenças Bacterianas
	Doenças Virais
	Doenças devido a Protozoários
	Demais Causas Infecciosas
ABORDAGEM 3: NEOPLASIAS	Neoplasia de Estômago
	Neoplasia de Pulmão
	Demais Neoplasias
ABORDAGEM 4: DOENÇAS DO APARELHO CIRCULATORIO	Infarto Agudo do Miocárdio
	Doenças Cerebrovasculares
	Código Lixo
	Demais Doenças Circulatórias
ABORDAGEM 5: DOENÇAS DO APARELHO RESPIRATÓRIO	Pneumonia
	Doenças Crônicas das Vias Aéreas Inferiores
	Demais Doenças Respiratórias
ABORDAGEM 6: CAUSAS EXTERNAS	Acidentes de Transporte
	Homicídios
	Código Lixo
	Demais Causas Externas

Quadro 10 – Causas Específicas mantidas em cada Abordagem

As variáveis independentes em todas as abordagens foram Ano de Óbito, Sexo, Faixa Etária, Estado Civil, Local de Ocorrência do Óbito, considerando as categorias apresentadas no Quadro 11, e a Unidade da Federação de Residência.

Para a obtenção dos modelos *CART* e *Random Forest* foram selecionados “Conjuntos de Treinamento”, que correspondem as informações efetivamente utilizadas no ajuste de cada modelo, a partir de amostras aleatórias compostas por um terço dos dados utilizados em cada abordagem. Os dois terços restantes dos dados de cada abordagem foram definidos como “Conjuntos de Teste” e suas informações foram utilizadas para testar a qualidade de predição dos modelos a partir das medidas de sensibilidade, especificidade, valores preditivos, taxa de detecção e acurácia. No modelo *Random Forest* da primeira abordagem foram ajustadas 500 árvores e nos modelos das demais abordagens foram ajustadas 1000 árvores. Os procedimentos de análise realizadas nesta etapa estão esquematizadas na Figura 13.

VARIÁVEL	ABORDAGENS	CATEGORIAS
Ano de Óbito	Todas as Abordagens	1979 a 1984 1985 a 1989 1990 a 1995 1996 a 1999 2000 a 2004 2005 a 2013
Sexo	Todas as Abordagens	Masculino Feminino
Faixa Etária	Abordagens 1, 2 e 5	Menos de 1 ano 1 a 18 anos 19 a 29 anos 30 a 59 anos 60 a 79 anos 80 anos ou mais
	Abordagem 3: Neoplasias Abordagem 4: Doenças do Aparelho Circulatório	30 a 59 anos 60 a 79 anos 80 anos ou mais
	Abordagem 6: Causas Externas	1 a 18 anos 19 a 29 anos 30 a 59 anos 60 anos ou mais
Estado Civil	Todas as Abordagens	Solteiro Casado Viúvo Separado
Local de Ocorrência do Óbito	Todas as Abordagens	Estabelecimentos de Saúde Domicílio Via Pública Outros

Quadro 11 – Categorização das Variáveis utilizadas em cada abordagem

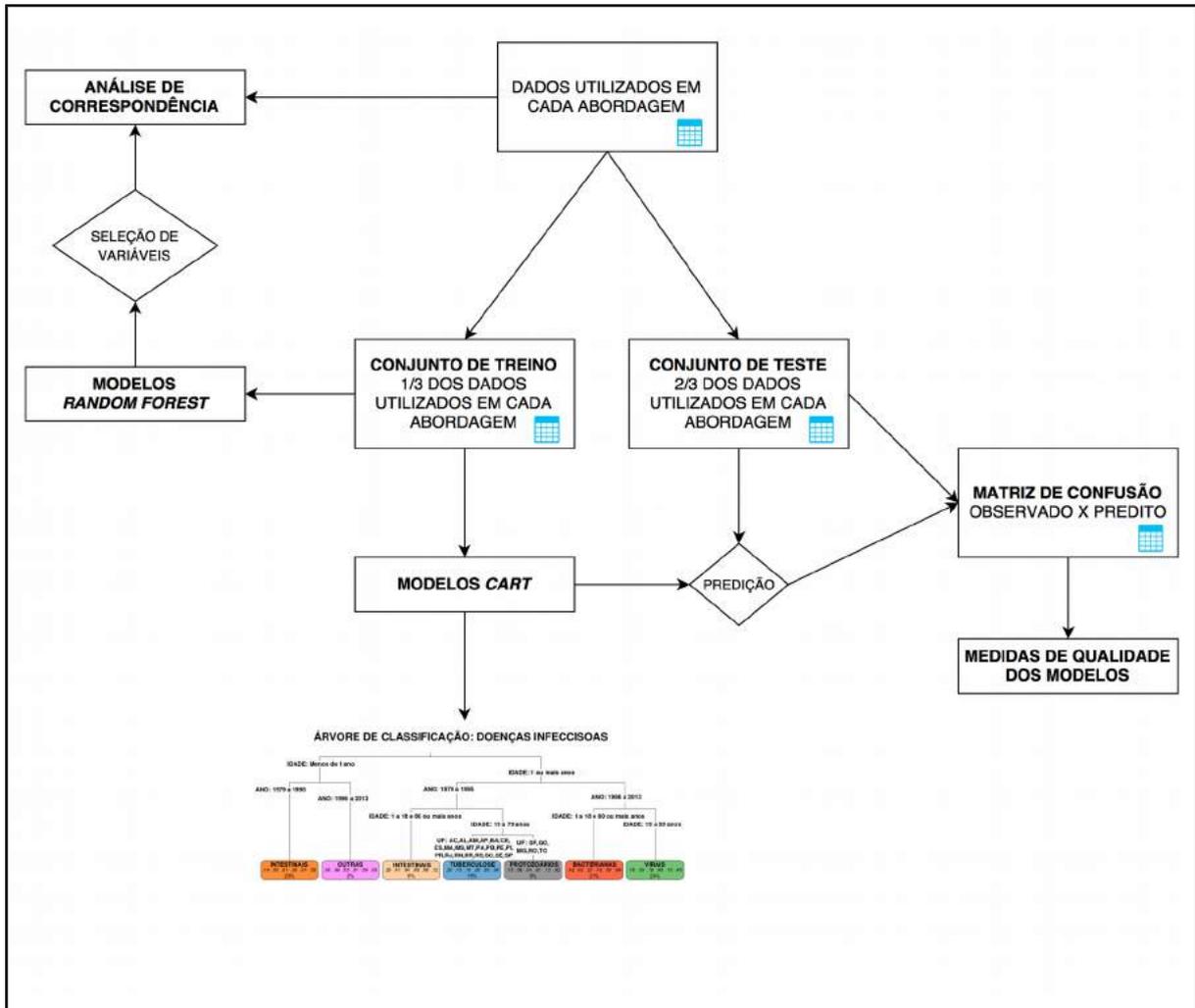


Figura 13 – Esquemática das Análises Realizadas na Etapa III

Os resultados do modelo *CART* foram apresentados através de gráficos. O Decréscimo Médio do Índice de Impureza de Gini, obtido dos modelos *Random Forest*, foi apresentado em gráficos de barras para subsidiar a seleção de variáveis para a Análise de Correspondência de Múltipla, utilizando a totalidade dos dados de cada abordagem.

A qualidade de predição dos modelos *CART* foi analisada através das medidas de Sensibilidade, Especificidade, Valor Preditivo Positivo, Valor Preditivo Negativo e Acurácia calculadas a partir da Matriz de Confusão obtida dos valores observados da variável resposta em cada Conjunto de Teste e dos valores preditos pela aplicação do modelo ao respectivo Conjunto de teste.

As bibliotecas *rpart* (THERNEAU; ATKINSON; PORT, 2017) e *ranger* (WRIGHT; ZIEGLER, 2017) foram utilizadas para implementação dos modelos *CART* e *Random Forest*, respectivamente, e *rpart.plot* (MILBORROW, 2017) para obtenção dos gráficos de árvore. Para o ajuste da Análise de Correspondência Múltipla utilizou-se a biblioteca *FactoMineR*

(LÊ; JOSSE; HUSSON, 2008) e os gráficos foram criados com a biblioteca *factoextra* (KASSAMBARA; MUNDT, 2017). Para obtenção da Matriz de Confusão e o cálculo das medidas de qualidade de predição dos modelos foi utilizada a biblioteca *caret* (WING et al., 2017).

5.5 CONSIDERAÇÕES ÉTICAS

O estudo utilizou dados de acesso aberto, sem identificação dos indivíduos, estando de acordo com os aspectos éticos definidos pela Resolução 466/12 e recebeu dispensa do Comitê de Ética em Pesquisa da Escola Nacional de Saúde Pública Sérgio Arouca – ENSP através do Parecer N° 03/2016 de 11 de agosto de 2016.

6 RESULTADOS

6.1 PADRÕES E TENDÊNCIAS DE MORTALIDADE SEGUNDO GRUPOS DE CAUSA

6.1.1 *MORTALIDADE PROPORCIONAL SEGUNDO GRUPOS DE CAUSA NO BRASIL E REGIÕES*

No Brasil, as doenças do aparelho circulatório se mantiveram como principal causa de óbito entre 1979 e 2013 (Figura 14), sendo responsável por cerca de 30% dos óbitos em todo o período. As Causas Mal Definidas apresentaram importante declínio no período, de 20,49% para 7,35%, definindo importante evolução no sistema de captação e registro de informações sobre mortalidade. Da mesma forma, as Doenças Infecciosas decaíram de maneira constante, variando de 8,48% (entre 1979 e 1984) para 4,21%, no período de 2005 a 2013.

Em contrapartida, as Neoplasias apresentaram expressivo aumento, destacando-se dentre as principais causas e passando de 8,75% dos óbitos no país, entre 1979 e 1984, para 16,57% entre 2005 e 2013. Padrão semelhante ao observado entre as Doenças do Aparelho Respiratório que variaram de 7,98% para 10,57% no período, e as Causas Externas que aumentaram de 9,33% para 12,52% (Figura 14).

Em relação à mortalidade proporcional não se verifica diferenças significativas entre as regiões brasileiras, ainda que a distribuição segundo grupos de causas se diferencie (Figuras 15, 16, 17, 18 e 19).

As regiões Norte e Nordeste (Figuras 15 e 16) apresentaram a maior queda proporcional de mortalidade por causas mal definidas, que representavam a principal causa de óbito até 2004, fato que interfere negativamente na análise do comportamento da mortalidade. Apesar do declínio observado nestas regiões, tais causas se mantêm mais elevadas quando comparadas às demais regiões.

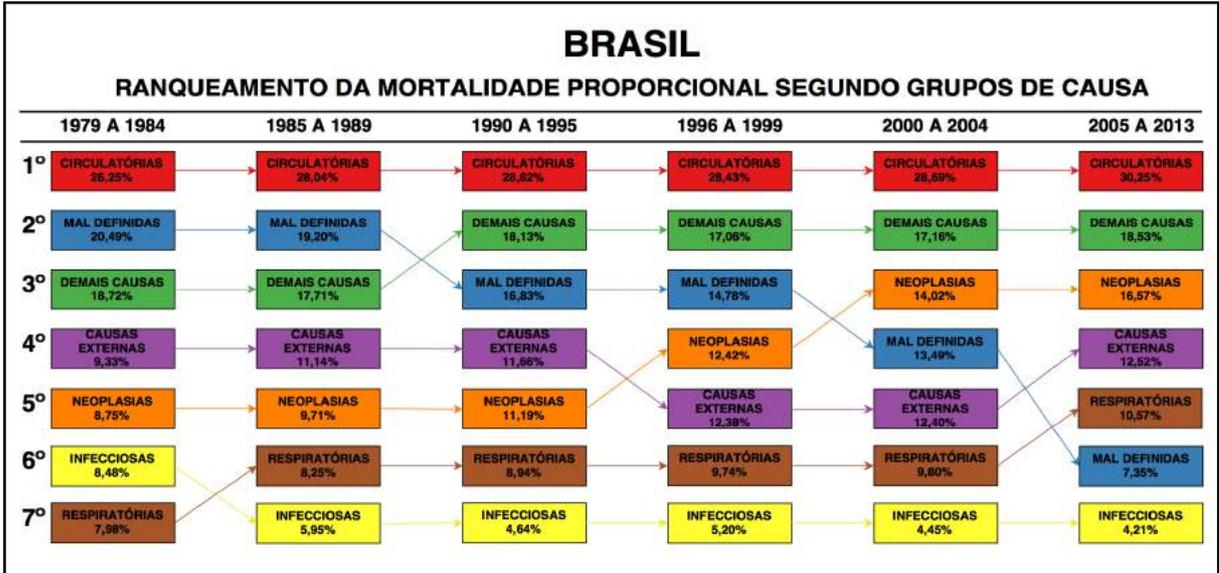


Figura 14 – Distribuição da Mortalidade Proporcional Segundo Grupo de Causa no Brasil entre 1979 e 2013

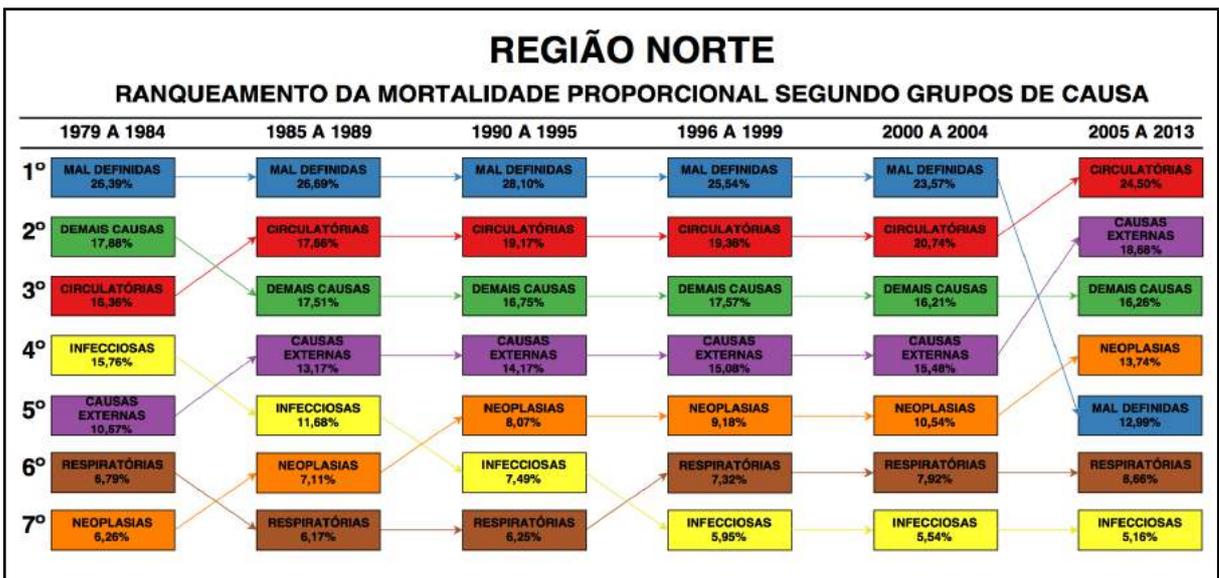


Figura 15 – Distribuição da Mortalidade Proporcional Segundo Grupo de Causa na Região Norte entre 1979 e 2013

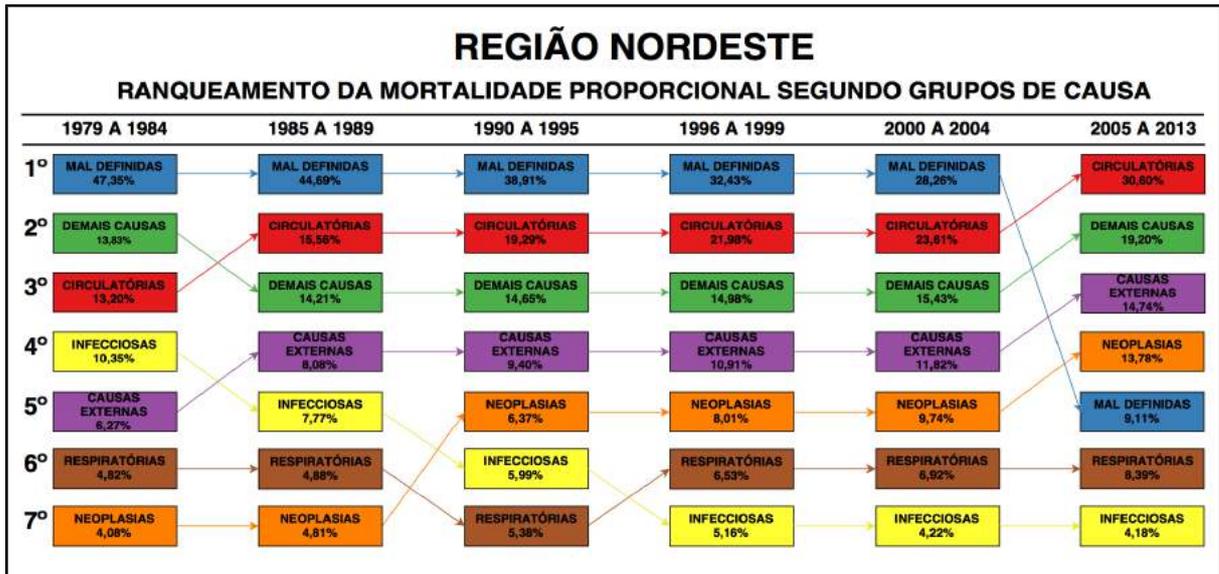


Figura 16 – Distribuição da Mortalidade Proporcional Segundo Grupo de Causa na Região Nordeste entre 1979 e 2013

Doenças do Aparelho Circulatório representam a principal causa relacionada à mortalidade proporcional em todas as regiões. Mudanças importantes no perfil de distribuição dos grupos de causa atingem particularmente as Neoplasias cujo peso passa a ser maior em relação à mortalidade total, constituindo-se no segundo grupo de causas de morte abaixo, de modo geral, apenas das Doenças do Aparelho Circulatório. As Doenças infecciosas apresentaram expressiva queda no período, passando a configurar dentre as causas menos frequentes; ainda assim mantêm papel relevante na distribuição de causas.

A partir da década de 80 as Causas Externas passam a se constituir uma das principais causas de óbito no Brasil, mantendo-se de modo geral no quarto lugar na mortalidade proporcional entre as regiões (Figuras 15, 16, 17, 18 e 19).

A importância relativa da mortalidade por Doenças do Aparelho Respiratório aumentou em todas as regiões, com destaque para Centro-Oeste e Sul, comportamento que pode estar associado à distribuição das outras causas.

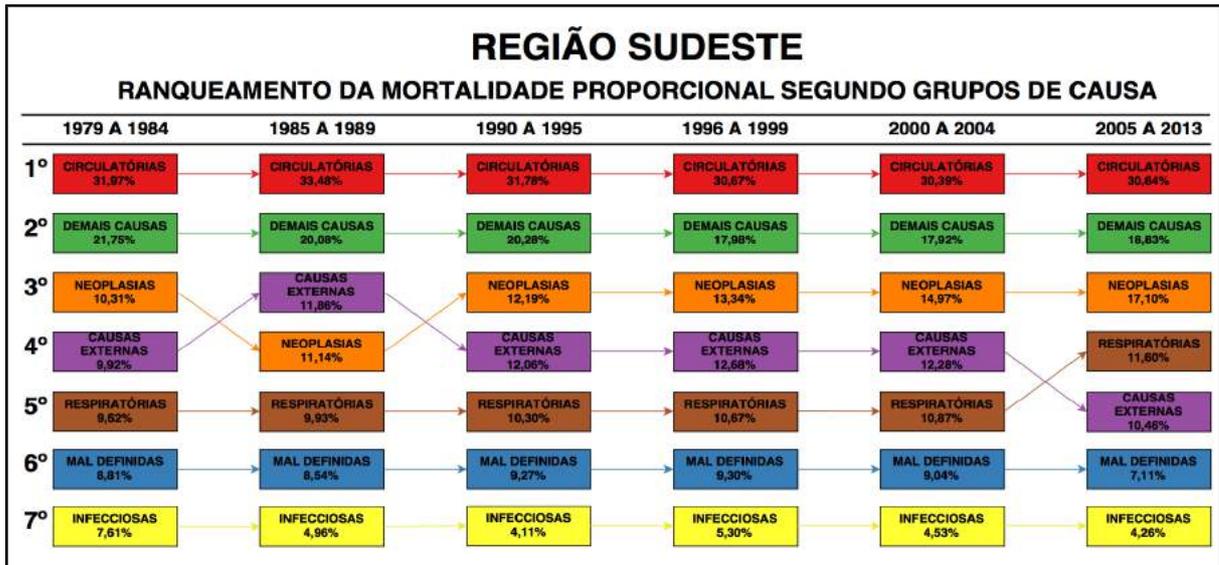


Figura 17 – Distribuição da Mortalidade Proporcional Segundo Grupo de Causa na Região Sudeste entre 1979 e 2013

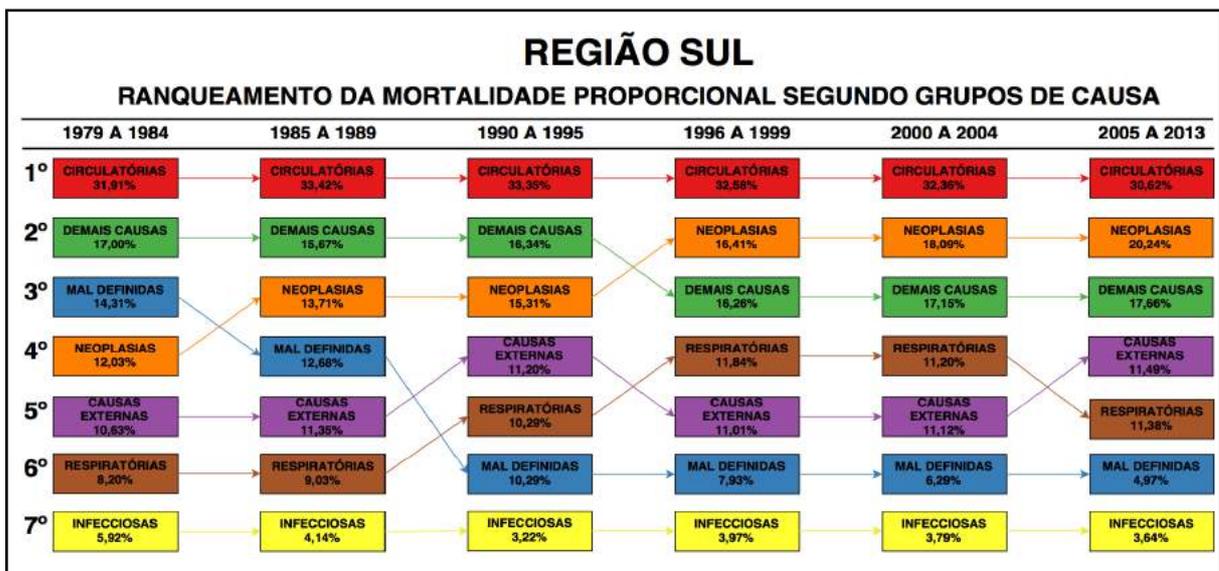


Figura 18 – Distribuição da Mortalidade Proporcional Segundo Grupo de Causa na Região Sul entre 1979 e 2013

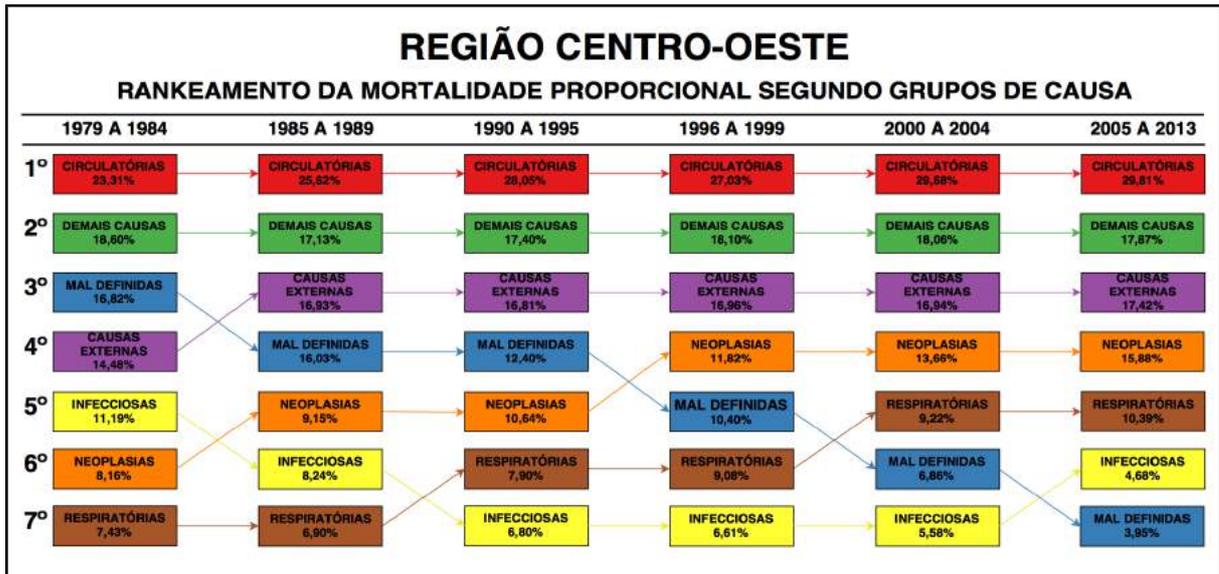


Figura 19 – Distribuição da Mortalidade Proporcional Segundo Grupo de Causa na Região Centro-Oeste entre 1979 e 2013

As desigualdades regionais na mortalidade proporcional segundo grupos de causa também podem ser observadas nos Diagramas de Kiviat (Figura 20). As Doenças do Aparelho Circulatorio (em vermelho nos gráficos) representaram o principal grupo de causa em todas as regiões no período estudado, sendo inicialmente mais representativa no Sul e Sudeste, expandindo-se nas demais regiões ao longo dos anos. No período de 2005 a 2013 ocorreu um ligeiro deslocamento deste grupo de causa para a região Nordeste, ultrapassando o valor das demais regiões.

A distribuição das Doenças Infecciosas (destacadas em verde) apresentou desigualdades regionais marcantes no período de 1979 a 1984, constituindo-se a causa mais frequente nas regiões Norte e Nordeste. Importante redução se observa a partir de 1996 quando tornou-se o grupo de causa menos frequente em todas as regiões, distanciando-se dos demais grupos de causa a partir de então.

O grupo de Causas Externas (em azul) foi desigual em relação às regiões em todo o período. Inicialmente mais representativa na região Centro-Oeste, avançou nas regiões Norte e Nordeste a partir de 1985, mantendo-se como segundo grupo de causa mais frequente nestas três regiões até 2013. Na região Sudeste este grupo de causa aumenta constante e consistentemente até 1999, menos frequente no Centro-Oeste, Norte e Nordeste e superior ao Sul, reduzindo-se a partir de 2000.

A mortalidade por Neoplasias (em roxo) aumentou de maneira constante em todas as regiões, apresentando conformação mais aguda na região Sul em todo o período, enquanto a

mortalidade por Doenças do Aparelho Respiratório (cor laranja) variaram pouco nas regiões e aumentaram ligeiramente nas regiões Sul e Sudeste (entre 2000 e 2013).

Neste contexto, considerando o ranqueamento da mortalidade proporcional segundo grupos de causa em cada região e o padrão exposto nos Diagramas de Kiviat; as Doenças do Aparelho Circulatório representaram principal grupo de causa de óbito no país, mantendo-se elevado nas regiões Sul, Sudeste e Centro-Oeste em todo o período. Verifica-se aumento nas regiões Norte e Nordeste nas últimas duas décadas. A mortalidade por Neoplasias cresce em todo o país, mas de maneira mais intensa nas regiões Sul e Sudeste, sobressaindo-se como segundo ou terceiro grupo de causa mais frequente em oposição às Doenças Infecciosas e Parasitárias que decaíram de maneira generalizada em todas as regiões.

A mortalidade por Doenças do Aparelho Respiratório aumentou ligeiramente em todas as regiões, de maneira um pouco mais intensa nas regiões Sul e Sudeste; já as Causas Externas se destacaram nas regiões Norte, Nordeste e Centro-Oeste.

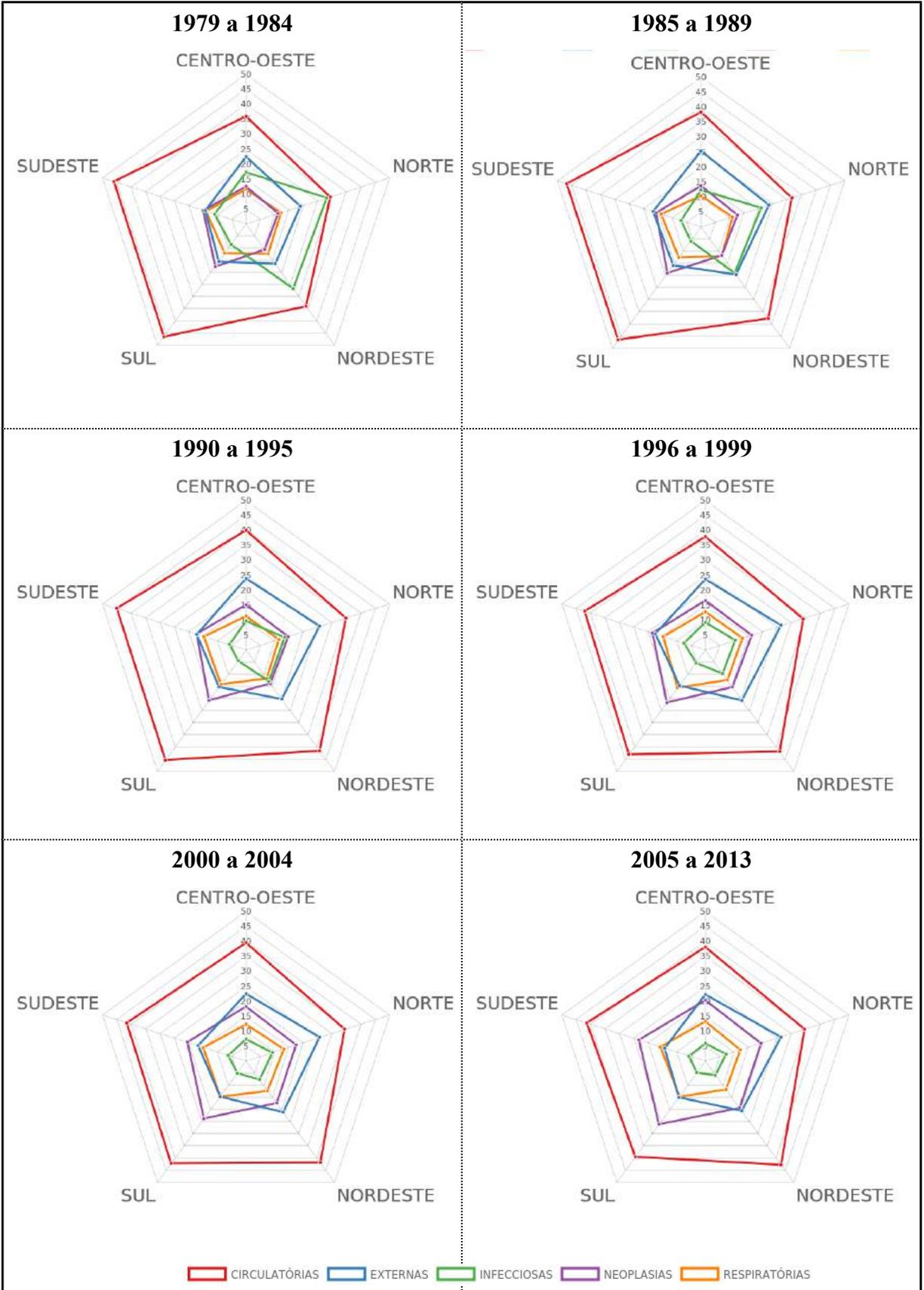


Figura 20 – Diagramas de Kiviat da Mortalidade Proporcional por Grupos de Causa selecionados segundo Região nos períodos de 1979 a 1984, 1985 a 1989, 1990 a 1995, 1996 a 1999, 2000 a 2004 e 2005 a 2013

6.1.2 EVOLUÇÃO DO PADRÃO ETÁRIO DOS ÓBITOS SEGUNDO GRUPOS DE CAUSA NO BRASIL E REGIÕES

A Figura 21 apresenta os *boxplots* da idade dos óbitos em geral entre 1979 e 2013 no Brasil e Regiões. Observa-se que a idade de óbito aumentou no Brasil e em todas as regiões em um processo desigual.

Nas regiões Norte e Nordeste as idades de óbito eram as mais baixas do país, com mediana de 20 e 28 anos em 1979, respectivamente, aumentando de maneira acelerada. Na região Norte, até 1987 o primeiro quartil da distribuição da idade de óbito foi zero, indicando que ao menos 25% dos óbitos ocorridos nesta região eram por mortalidade infantil, e somente a partir de 2008 a mediana da idade alcançou os 60 anos de idade, avançando pouco a partir de então. Já na região Nordeste o aumento da idade de óbito foi mais acelerado, de maneira que a mediana da idade de óbito aumentou constantemente em todo o período, alcançando 60 anos em 1991 e se aproximando de 70 anos em 2013.

No outro extremo, as regiões Sul e Sudeste se destacaram com idade de óbito mais elevadas em todo o período. A principal diferença de comportamento do indicador nestas regiões foi o primeiro quartil mais baixo na região Sudeste em todo o período. Na região Sul os óbitos infantis (menos de 1 ano de idade) passaram a ser considerados *outliers* a partir de 1998 e os óbitos antes de 10 anos a partir de 2009, enquanto na região Sudeste os óbitos entre menores de um ano aparecessem como *outliers* em 2004 e somente em 2013 para o grupo antes de 10 anos.

Vale notar que o comportamento da idade de óbito na região Centro-Oeste foi similar às regiões Sudeste e Sul, evoluindo de maneira mais suave do que nas regiões Norte e Nordeste. Assim, a mediana de óbito era de 40 anos em 1979, atingiu cerca de 60 anos em 2000 e cerca de 65 anos em 2013.

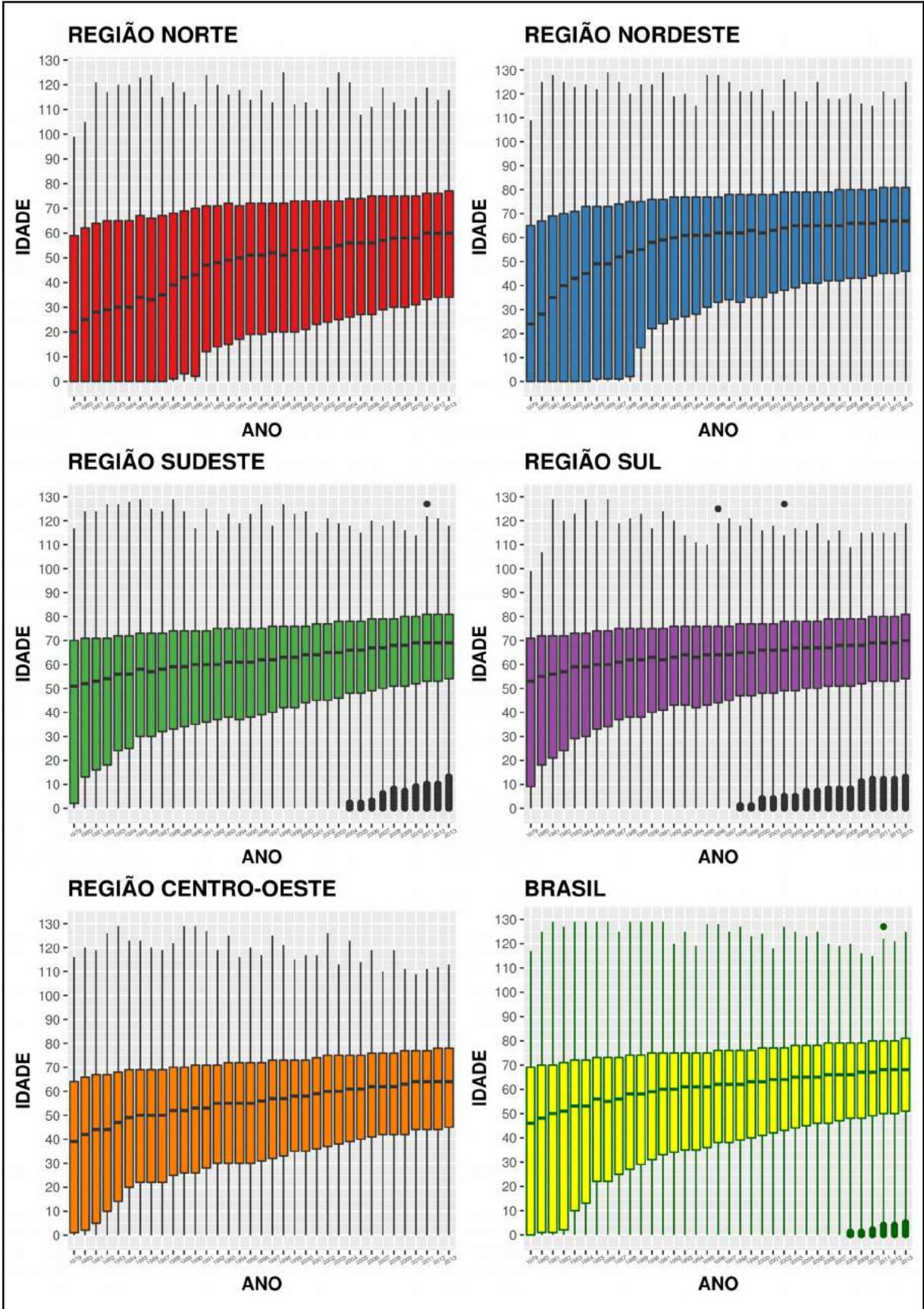


Figura 21 – Evolução da distribuição da idade dos óbitos em geral no Brasil e Regiões de 1979 a 2013

Diferenças regionais também podem ser observadas na mortalidade por Doenças Infecciosas e Parasitárias (Figura 22) e do Aparelho Respiratório (Figura 23), com impacto entre as faixas etárias mais jovem no início do período analisado e entre os idosos ao longo de todo o período.

Até o início da década de 1990 ao menos metade dos óbitos por Doenças Infecciosas nas regiões Norte e Nordeste ocorria antes de 5 anos, e, adicionalmente, mantinha-se concentração de *outliers* entre as idades mais avançadas (Figura 22).

O aumento da idade de óbito foi lento na região Norte, de maneira que o terceiro quartil foi inferior a 10 anos de idade até 1980, elevando-se gradualmente ao longo dos anos. Já o primeiro quartil permaneceu abaixo dos 10 anos até 2005, mantendo-se abaixo de 30 anos em 2013. A mediana ultrapassou os 10 anos em 1993, e os 30 anos em 2000 e aproximou-se dos 50 anos em 2013.

Em contrapartida, na região Nordeste as mudanças foram aceleradas. O terceiro quartil da distribuição esteve abaixo de 10 anos de idade até 1983, ultrapassou 30 anos em 1985 e 50 anos em 1990, chegando a quase 80 anos em 2013. O primeiro quartil permaneceu próximo de 0 até o ano 2000, subindo rapidamente a partir de então, até atingir 40 os anos em 2013. A mediana aumentou pouco até 1994, subindo para cerca de 15 anos em 1995 e ultrapassando 60 anos em 2013.

Em contrapartida, nas regiões Sudeste e Sul os óbitos por Doenças do Aparelho Respiratório atingiam aqueles em idade mais avançada, quando comparados às regiões Norte e Nordeste. Em 1979 a mediana da idade de óbito era um pouco menos de 20 anos no Sudeste e mais de 40 anos no Sul, ultrapassando 60 anos em 1985 em ambas as regiões. O primeiro quartil das distribuições superou os 30 anos em 1989 na região Sudeste e em 1986 na região Sul, e 65 anos em ambas em 2010. Óbitos abaixo de 5 anos tornaram-se *outliers* em 1990 na região Sul e em 1995 na região Sudeste. Em 2013 os óbitos de indivíduos com menos de 40 e 35 anos foram considerados *outliers* no Sul e Sudeste, respectivamente.

Na região Centro-Oeste o comportamento da distribuição foi similar ao observado no Sudeste e Sul. Embora a mediana estivesse abaixo de 5 anos em 1979, no ano seguinte supera 20 anos e em 1986 ultrapassa os 50 anos, chegando a 75 anos em 2013. O primeiro quartil também avança rapidamente, superando 30 anos já em 1990 e 60 anos em 2010; óbitos até 5 anos de idade passam a ser considerados *outliers* em 1997 e aqueles abaixo de 35 anos em 2012.

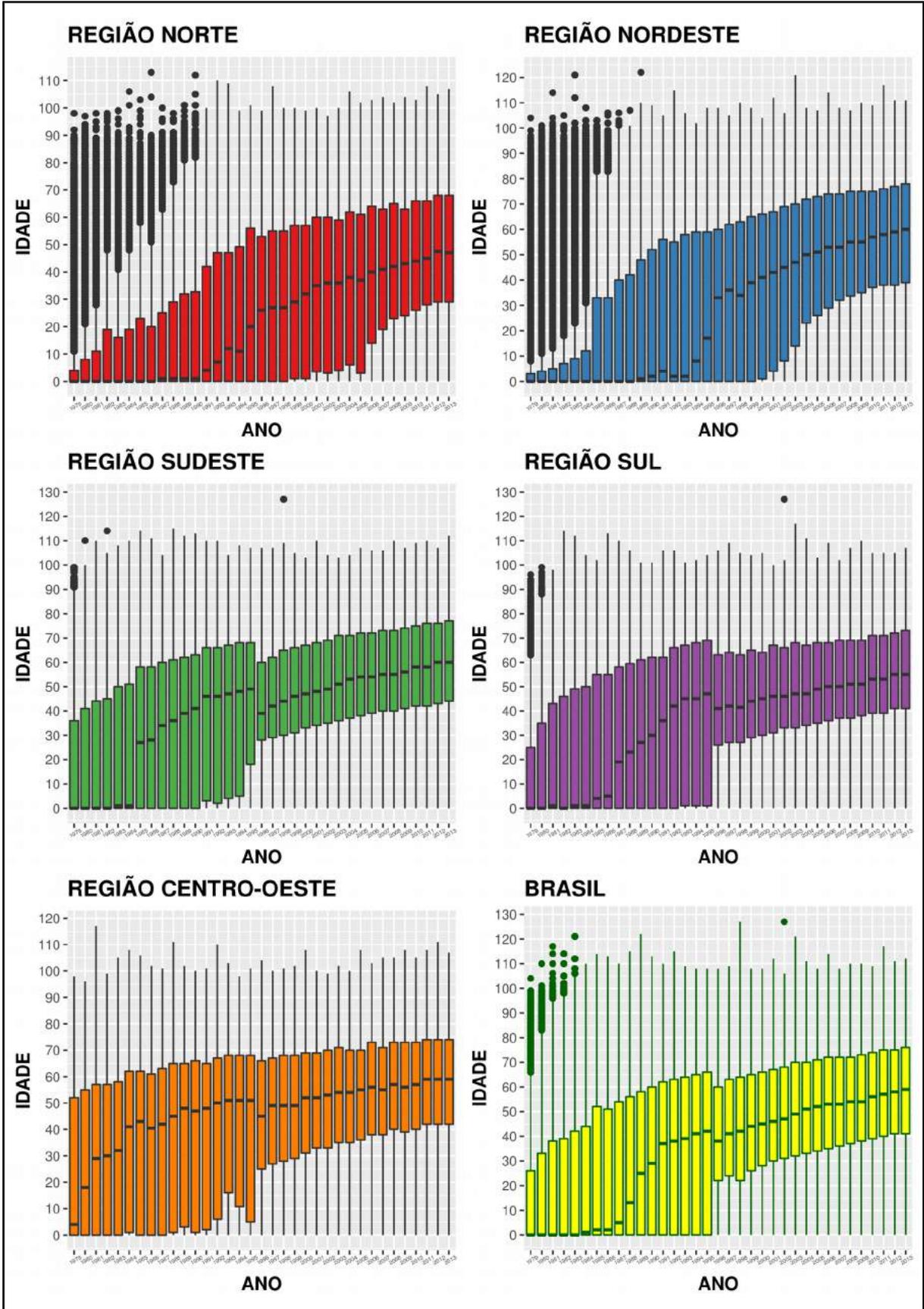


Figura 22 – Evolução da distribuição da idade dos óbitos por Doenças Infecciosas e Parasitárias no Brasil e Regiões de 1979 a 2013

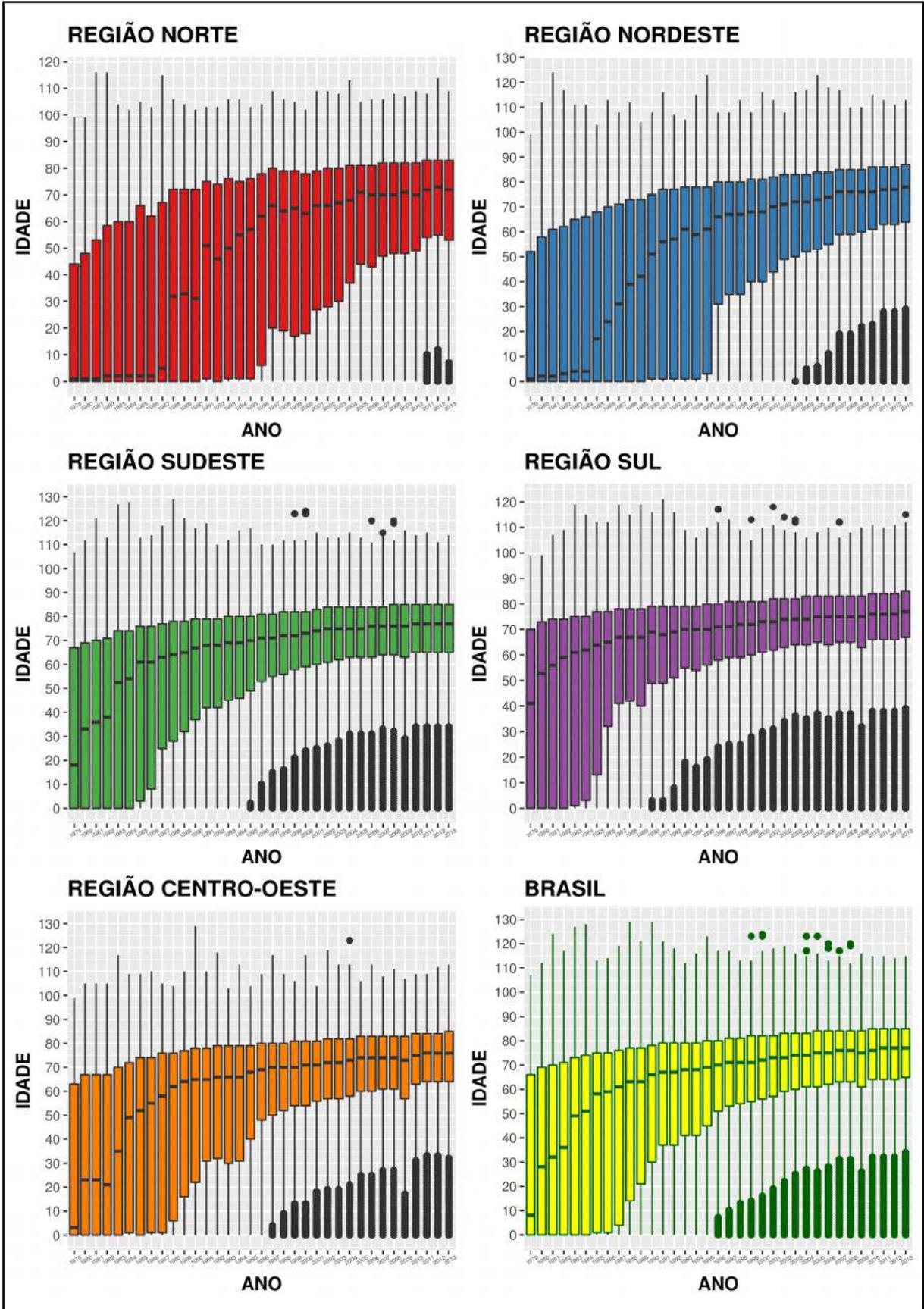


Figura 23 – Evolução da distribuição da idade dos óbitos por Doenças do Aparelho Respiratório no Brasil e Regiões de 1979 a 2013

Diferente da evolução do padrão geral da idade de óbito e dos óbitos por Doenças Infecciosas e Parasitárias e por Doenças do Aparelho Respiratório, onde é possível observar padrões regionais distintos, os óbitos por Neoplasias e Doenças do Aparelho Circulatório atingem faixas etárias mais elevadas no Brasil e em todas as regiões, com pequenas variações no período, enquanto os óbitos por Causas Externas se concentram em faixas etárias mais jovens.

No Brasil, os óbitos por Neoplasia atingiram principalmente indivíduos em torno dos 60 anos em todas as regiões nos anos iniciais, com ligeira tendência de crescimento na mediana da idade no período, enquanto as Doenças do Aparelho Circulatório foi inicialmente mais frequentes por volta dos 70 anos de idade. Em ambos os grupos de causa, as idades foram um pouco mais baixas na região Centro-Oeste, que apresentou a tendência de aumento mais agudo (Figura 24 e Figura 25).

Em relação às Causas Externas, os óbitos mantiveram-se por volta de 25 anos na região Norte e de 30 anos nas demais regiões (Figura 26). Nas regiões Sudeste e Sul observa-se um suave crescimento na mediana e no terceiro quartil a partir de 2000.

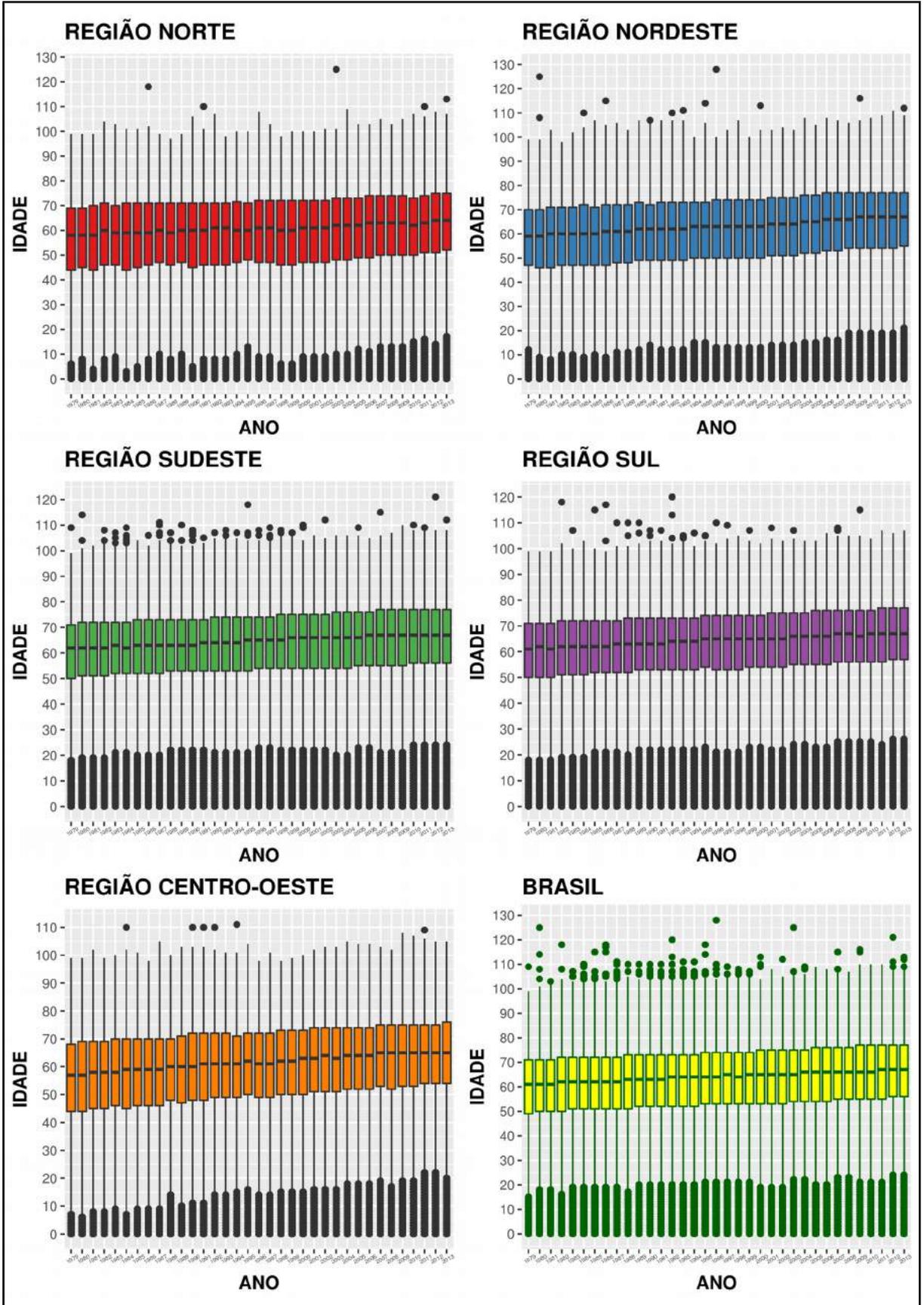


Figura 24: Evolução da distribuição da idade dos óbitos por Neoplasias no Brasil e Regiões de 1979 a 2013

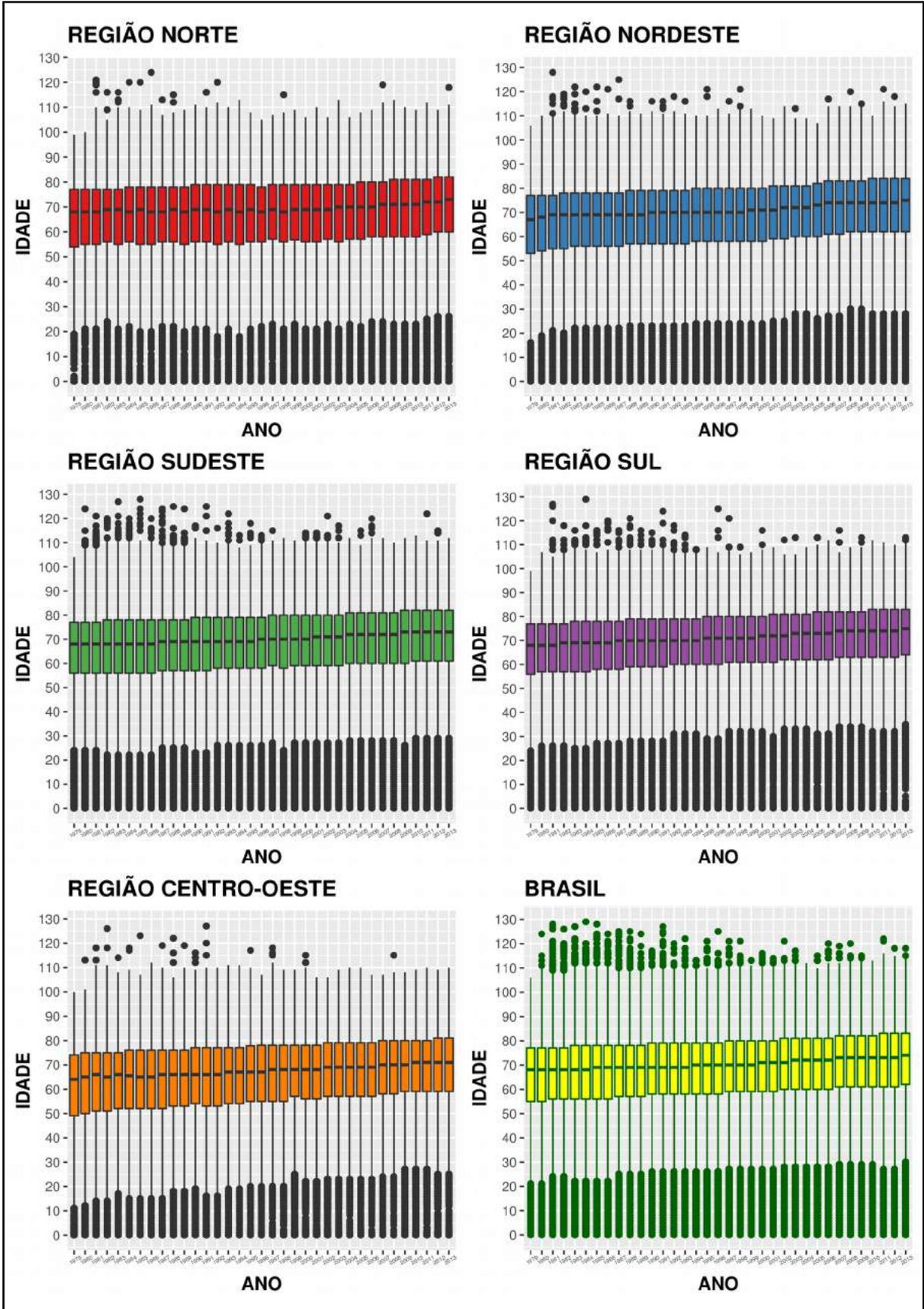


Figura 25 – Evolução da distribuição da idade dos óbitos por Doenças do Aparelho Circulatório no Brasil e Regiões de 1979 a 2013

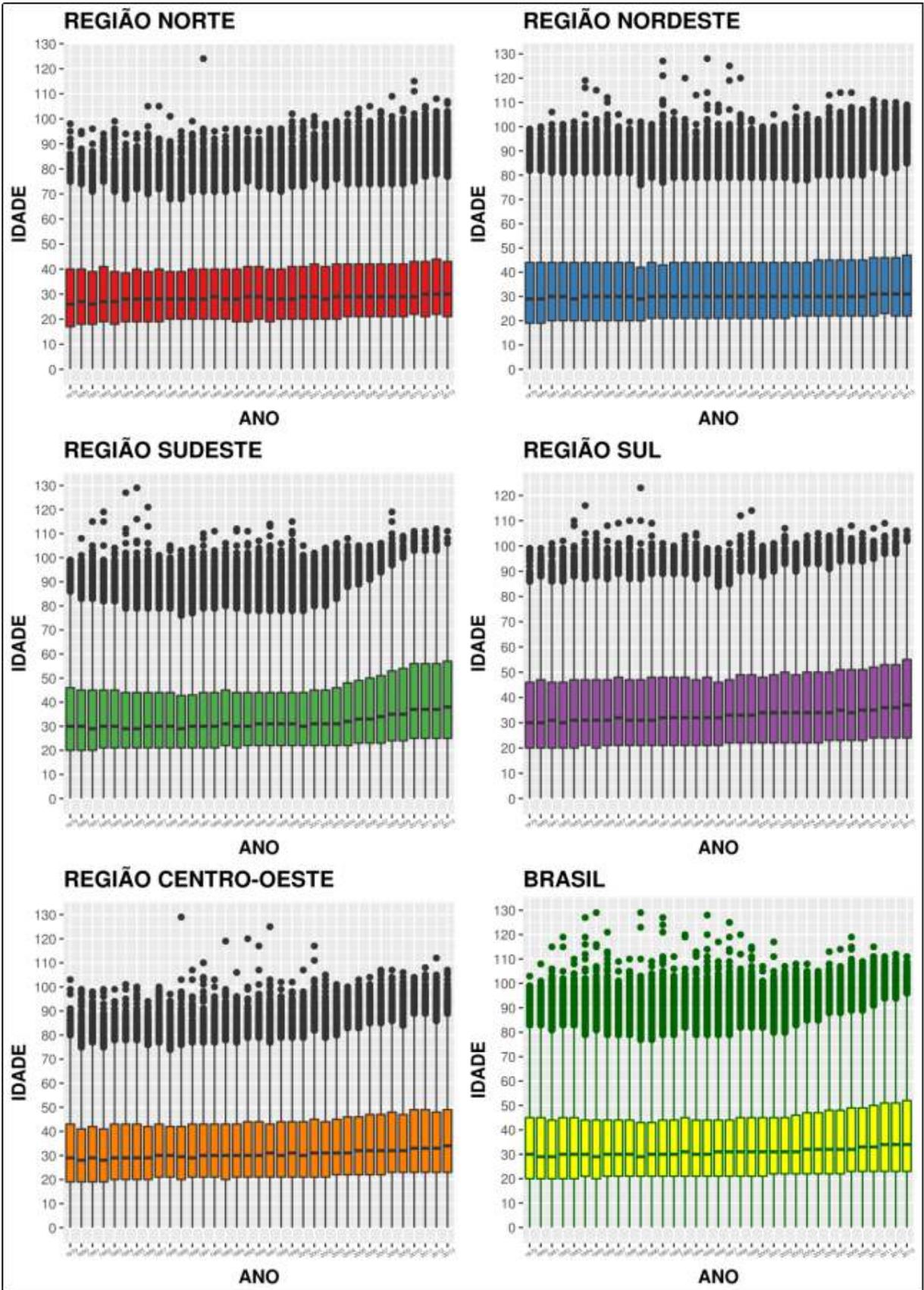


Figura 26 – Evolução da distribuição da idade dos óbitos por Causas Externas no Brasil e Regiões de 1979 a 2013

6.1.3 *EVOLUÇÃO DAS TAXAS DE MORTALIDADE SEGUNDO GRUPO DE CAUSA E CAUSAS SELECIONADAS NAS UNIDADES DA FEDERAÇÃO*

Observa-se na Figura 27 a evolução da Taxa de Mortalidade por Doenças Infecciosas e Parasitárias segundo sexo nas Unidades da Federação no período de 1980 a 2012. As taxas foram ligeiramente mais elevadas entre homens e os valores mais elevados se concentraram na região Norte em seu conjunto, em Goiás e no Distrito Federal no início do período estudado, com tendência geral de queda em todo país.

Na Figura 27 destacam-se em vermelho as altas taxas em Rondônia, no período de 1980 a 1988, e em Goiás de 1981 a 1985, e também em Roraima e Distrito Federal apenas para o sexo masculino. A alteração no padrão de cores indica uma queda mais rápida nas Unidades da Federação que compõem a região Norte, já em Goiás e no Distrito Federal o processo foi mais lento (em destaque com cores mais claras que predominam até 1995). Minas Gerais e Alagoas apresentaram valores mais próximas a mediana até meados da década de 80, decaindo a partir de então.

As Taxas de Mortalidade por Doenças Infecciosas Intestinais, causa mais frequente no grupo de Doenças Infecciosas foram mais elevadas entre homens e os valores mais altos concentraram-se na região Norte, com destaque para Roraima, Rondônia e Amazonas e na região Nordeste, onde sobressaíram Paraíba e Alagoas (Figura 28).

A tendência foi de queda em todo o país, mas diferente do comportamento das Doenças Infecciosas e Parasitárias em geral, taxas próximas da mediana foram observadas em quase toda região Norte e Nordeste até o início da década de 1990. Na região Sudeste, o Rio e Janeiro, São Paulo e Minas Gerais apresentaram taxas medianas até 1982, decaindo a partir de então, assim como o Paraná e o Mato Grosso do Sul.

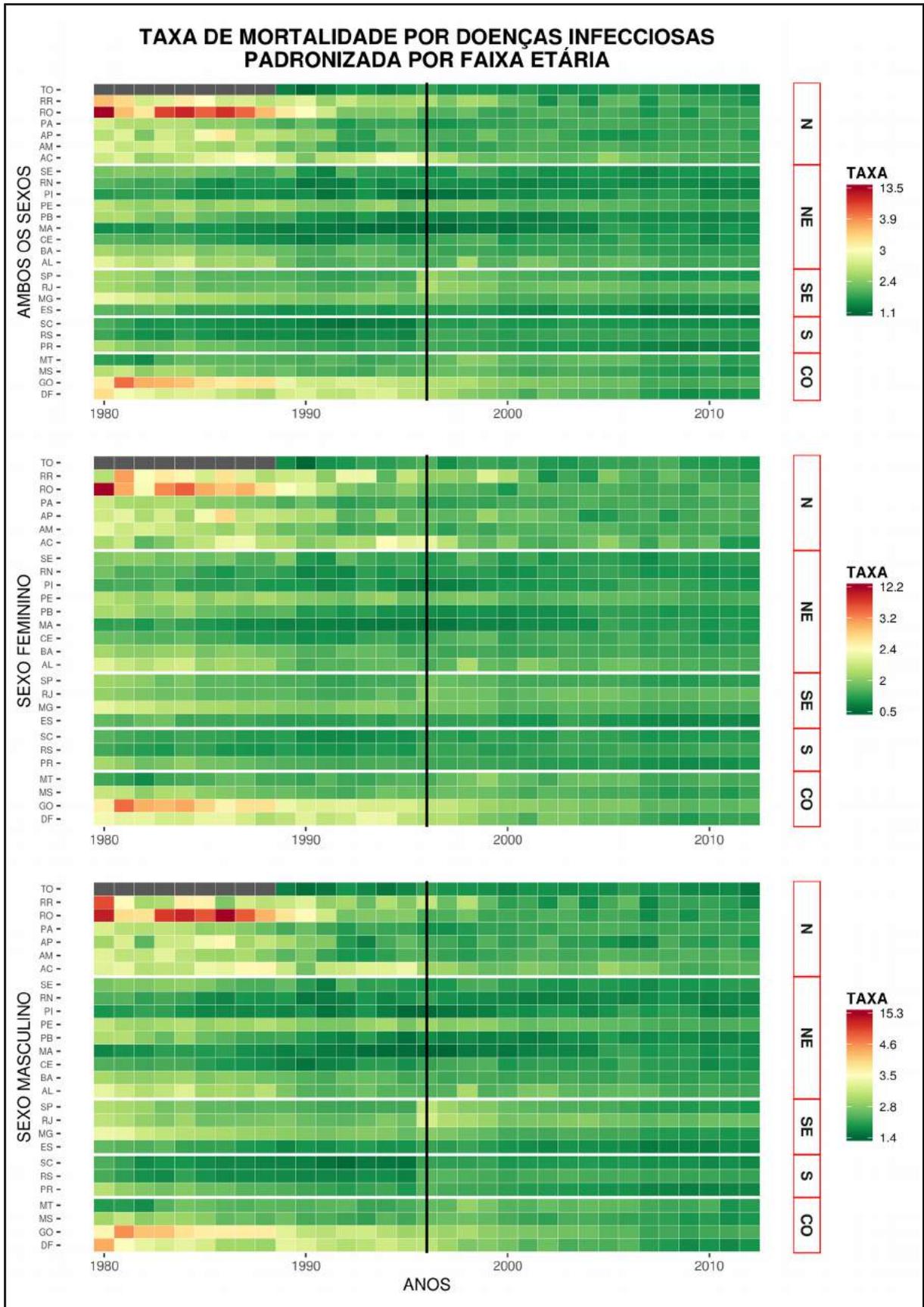


Figura 27 – Evolução da Taxa de Mortalidade por Doenças Infecciosas e Parasitárias segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012

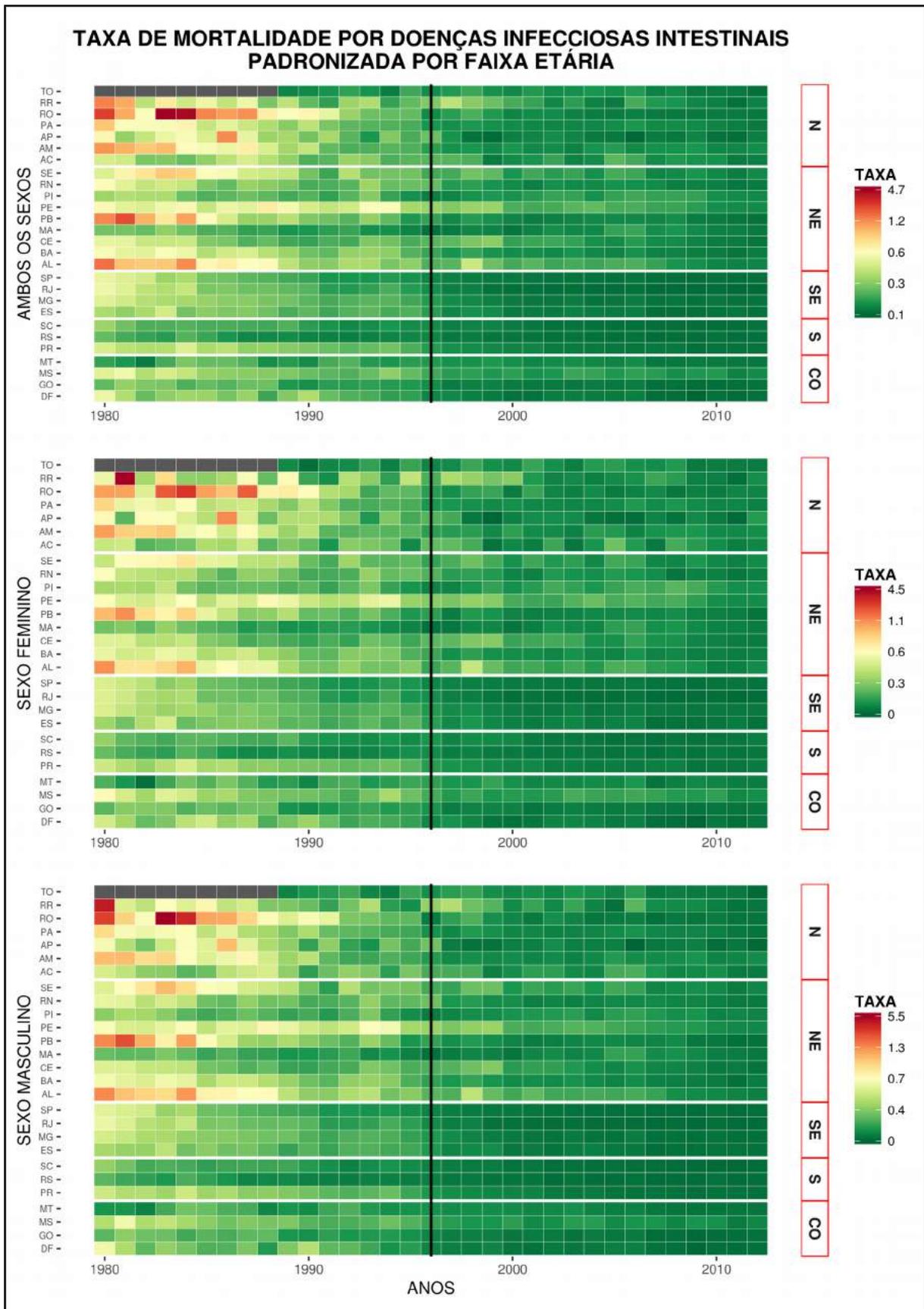


Figura 28 – Evolução da Taxa de Mortalidade por Doenças Infecciosas Intestinais segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012

As Taxas de Mortalidade por Neoplasia foram mais elevadas entre homens, subindo em todo o país (Figura 29). De maneira geral as maiores taxas se mantiveram nas regiões Sul, Sudeste e Centro-Oeste, com destaque para Rio de Janeiro, São Paulo, Rio Grande do Sul e Distrito Federal. O Mato Grosso (em ambos os sexos), Goiás e Mato Grosso do Sul no sexo masculino apresentaram padrão diverso com taxas mais baixas até 1995, aumentando de maneira mais intensa a partir de então.

Na maior parte da região Norte as taxas foram mais elevadas do que na região Nordeste, entretanto, a partir de meados dos anos 2000, as taxas crescem de maneira mais intensa na região Nordeste. Observa-se valores elevados dispersos na região Norte ao longo dos anos em Amapá e Roraima.

A Taxa de Mortalidade por Neoplasias de Pulmão aumenta sobretudo entre mulheres, ainda que permaneçam patamares mais elevados no sexo masculino (Figura 30). No gráfico referente aos óbitos em geral observa-se que na região Norte as cores apontam para flutuação dos mais valores baixos e próximos à mediana em todo o período; na região Nordeste a tendência foi de aumento (predominando a cor verde, com clareamento do tom a partir de 2005).

Na região Sudeste as taxas foram medianas em São Paulo e Rio de Janeiro em todo o período (prevalecendo tons mais claros destacados na Figura 30), baixas em Minas Gerais (sobressaindo a cor verde), e aumentam ligeiramente no Espírito Santo. Em contrapartida, na região Sul e na região Centro-Oeste observa-se tendência de aumento (mudança da cor dominante de verde para tons mais claros). O comportamento entre homens foi semelhante à dos óbitos na população total, contudo, com alteração mais sutil (gradação no padrão de cor).

Na região Norte, em relação ao sexo feminino, não há registro de óbitos por Neoplasias de Pulmão em alguns anos no Acre, Roraima e Amapá. Na região Nordeste as taxas foram mais baixas e, a partir de 2005, a tendência de aumento é evidenciada pelo padrão de clareamento de cores na Figura 30. No outro extremo, nas regiões Sudeste e Centro-Oeste o aumento ocorreu de maneira gradual ao longo do período. Na região Sul a elevação de taxas ocorreu de maneira acelerada.

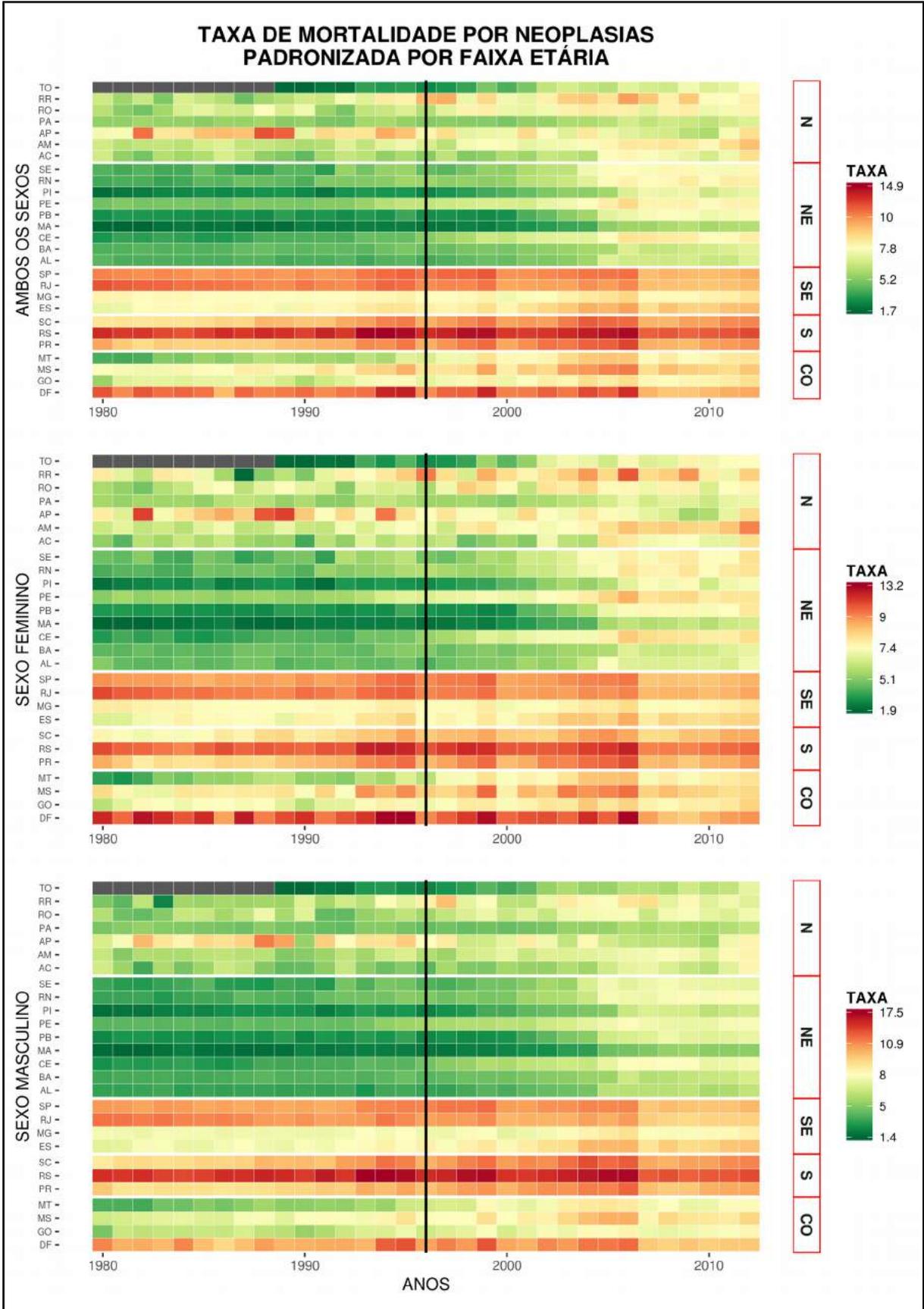


Figura 29 – Evolução da Taxa de Mortalidade por Neoplasias segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012

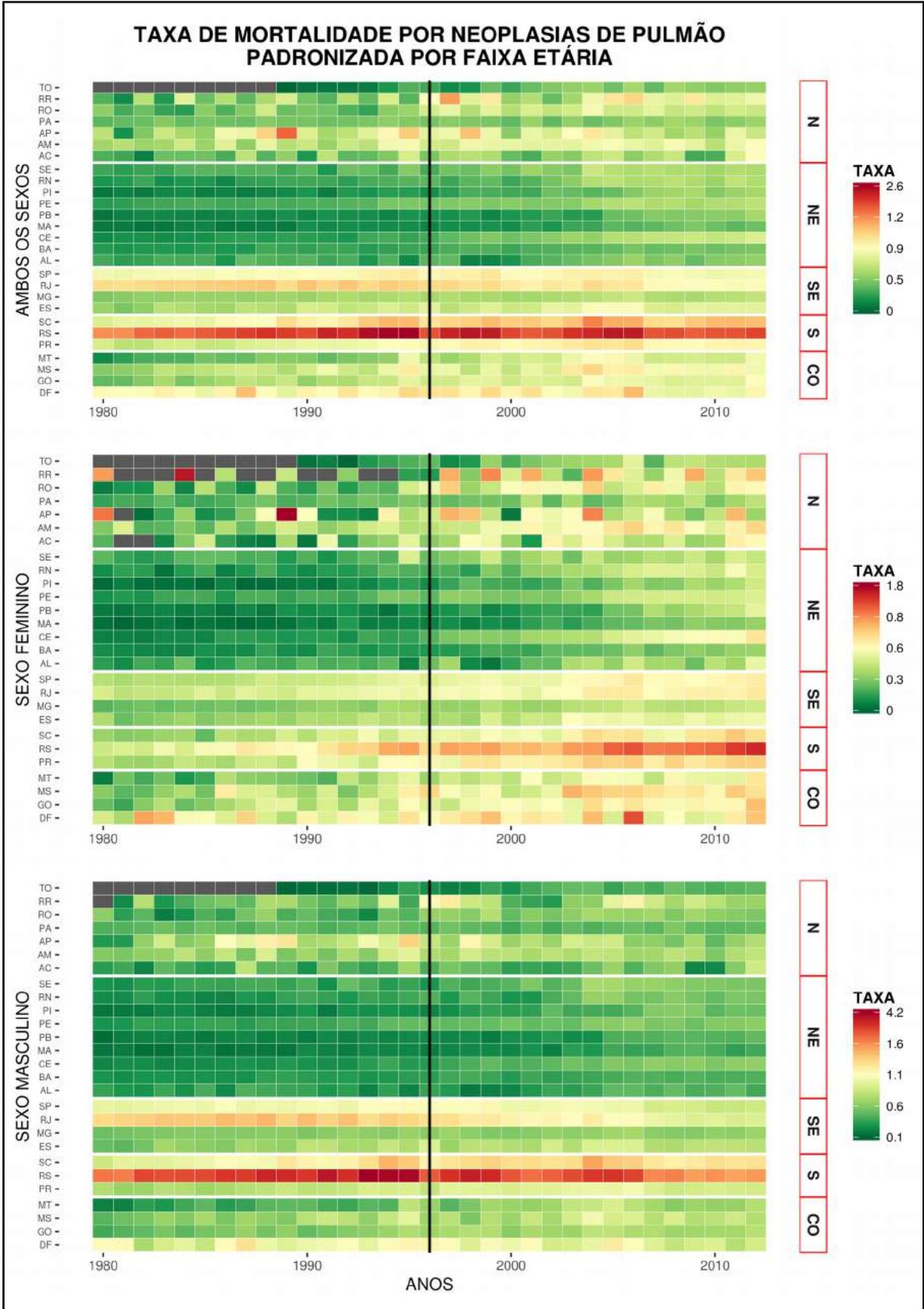


Figura 30 – Evolução da Taxa de Mortalidade por Neoplasias de Pulmão segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012

A Figura 31 mostra a evolução das Taxas de Mortalidade por Doenças do Aparelho Circulatório nas Unidades da Federação no período analisado. Com exceção de taxas mais elevadas entre homens, não se verifica diferenças nos padrões de distribuição segundo sexo e na população em geral.

Na região Norte verifica-se flutuações nas taxas de mortalidade, mantendo-se baixas ou com valores próximos à mediana em todo o período. Na região Nordeste as taxas mantiveram-se baixas (prevalendo o verde em tom mais escuro), mas a partir de 2007 observa-se uma tendência de aumento (com o clareamento do tom representado na Figura 31). Pernambuco destacou-se com valores mais próximos a mediana em todo o período.

Em contrapartida, nas regiões Sudeste e Sul as taxas foram mais elevadas no início do período, com os maiores valores concentrados em São Paulo, Rio de Janeiro, Rio Grande do Sul e Paraná, entre 1980 e 1995. e tendência de queda constante e gradual para o conjunto destas regiões. Na região Centro-Oeste a tendência também foi de queda no Mato Grosso do Sul, Goiás e Distrito Federal, partindo, de um patamar mais baixo ao observado no Sul e Sudeste nos anos iniciais; já no Mato Grosso as taxas flutuaram com valores baixos e próximos da mediana, diminuindo gradativamente a partir de 2007.

A distribuição da Taxa de Mortalidade por Doenças Cerebrovasculares apresenta padrão semelhantes ao observado no grupo de Doenças do Aparelho Circulatório, com taxas mais elevadas nas regiões Sudeste e Sul e próximos da mediana no Centro-Oeste até 1995, e tendência de queda constante nestas regiões. Observa-se na Figura 32 que, de maneira geral, o tom de verde tendeu à homogeneidade a partir de 2007, um indicando valores equivalentes de taxas.

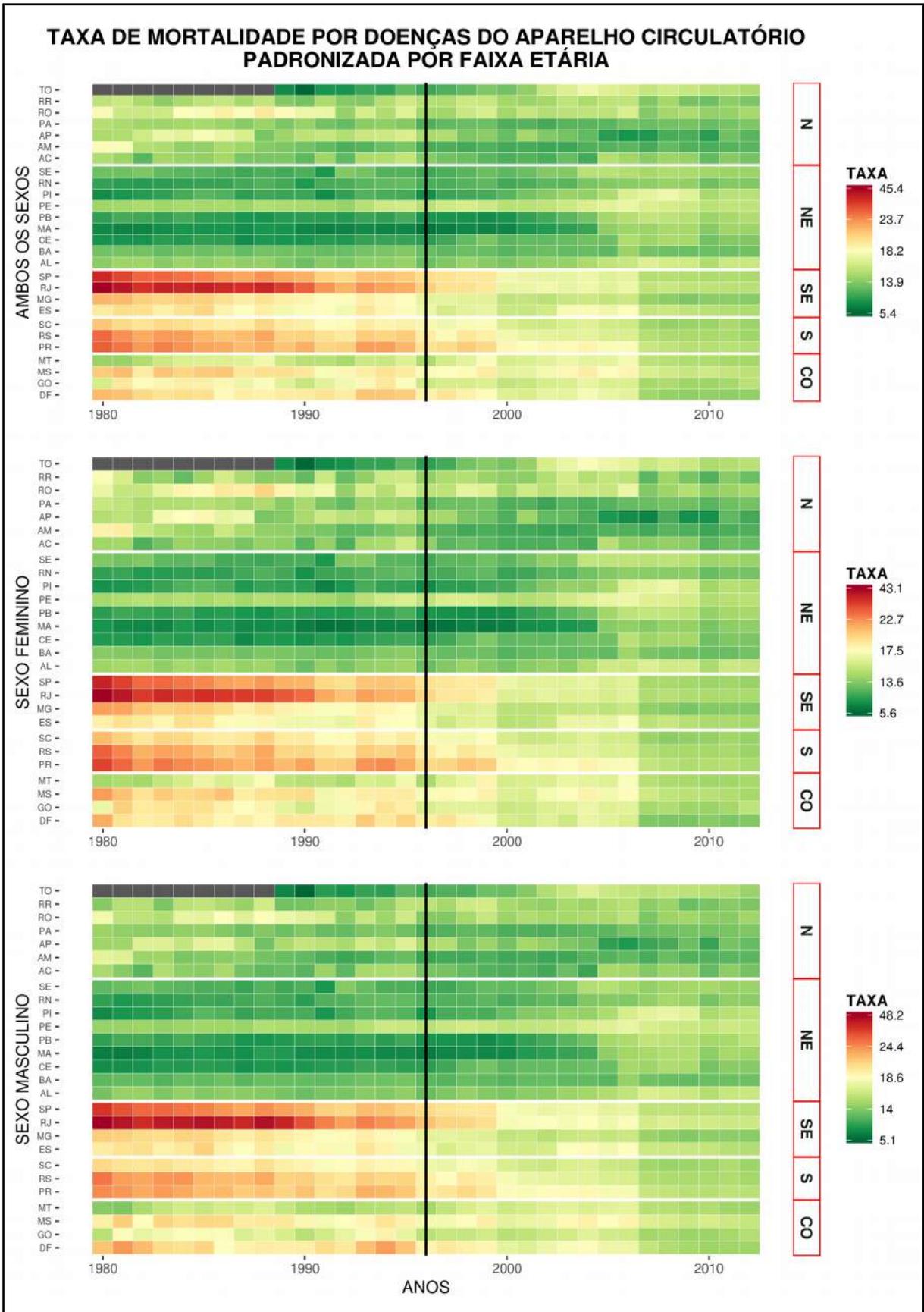


Figura 31 – Evolução da Taxa de Mortalidade por Doenças do Aparelho Circulatório segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012

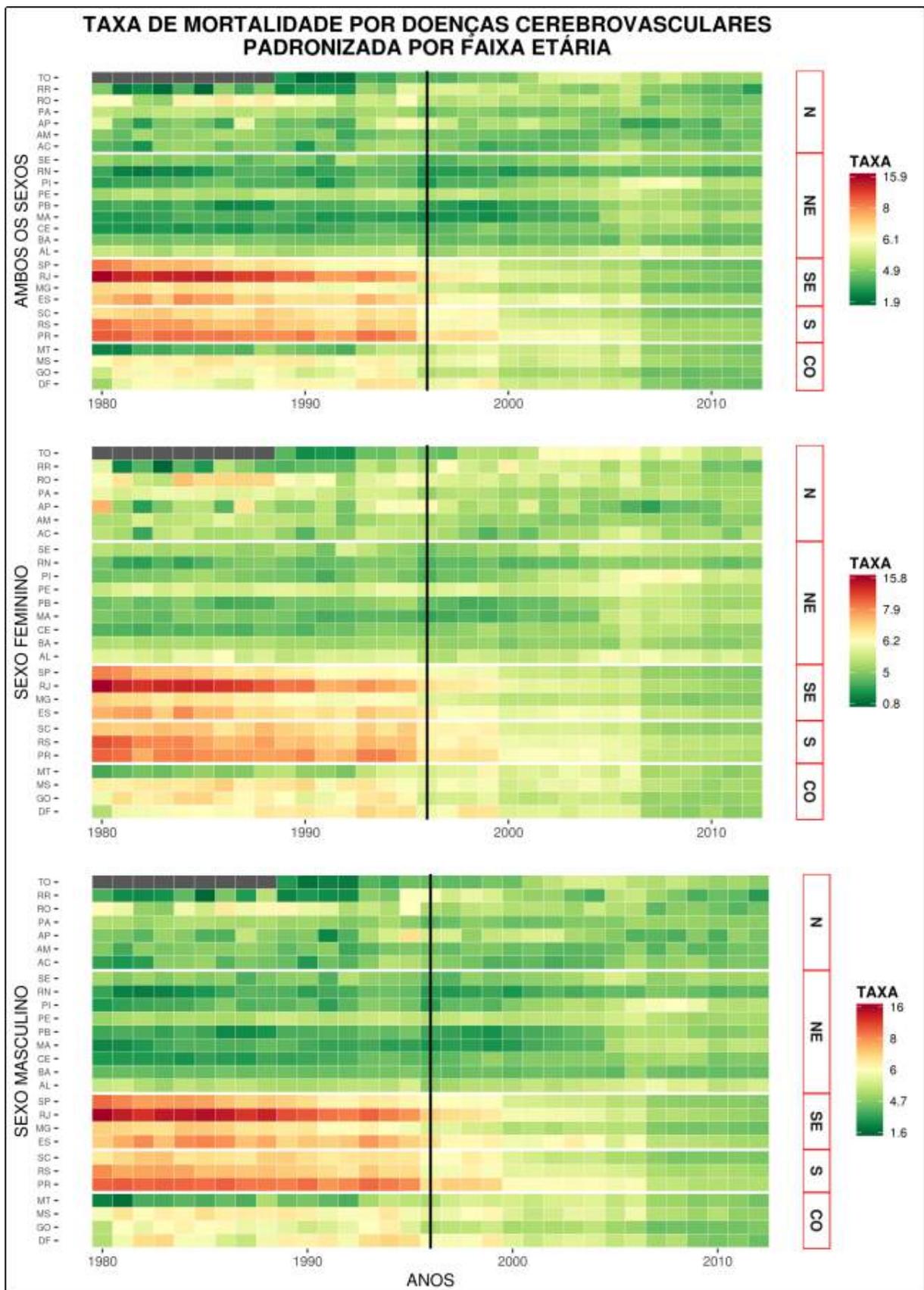


Figura 32 – Evolução da Taxa de Mortalidade por Doenças Cerebrovasculares segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012

Os gráficos da Figura 33 mostram a evolução das Taxas de Mortalidade por Doenças do Aparelho Respiratório. Em relação aos óbitos em geral e no sexo feminino observou-se aumento de taxas nas regiões Norte, e Nordeste, esta última com Pernambuco mantendo taxas mais elevadas. Na região Sudeste as taxas diminuíram em São Paulo e no Rio de Janeiro, que concentraram valores elevados e próximos da mediana em todo o período, e aumentaram no Espírito Santo, enquanto em Minas Gerais ocorreu aumento na primeira metade do período e diminuição na segunda metade.

Ainda considerando os óbitos em geral e do sexo feminino, observa-se que em toda a região Sul as taxas atingiram os patamares mais elevados em 1993 e decaindo a partir de 2006. Já na região Centro-Oeste as taxas aumentaram em Mato Grosso e Goiás, diminuíram no Distrito Federal e variaram entre valores próximos da mediana e elevados no Mato Grosso do Sul. No sexo masculino as taxas foram mais elevadas e apresentaram menos flutuação, sendo mais evidente o aumento nas regiões Norte e Nordeste e a diminuição no Rio de Janeiro, São Paulo, Rio Grande do Sul e Distrito Federal.

De maneira semelhante, a Taxa de Mortalidade por Pneumonia mostrou-se mais elevadas homens (Figura 34). Na região Norte destacam-se taxas elevadas em Roraima (nos dois primeiros e no último ano do período) e no Amapá (entre 1987 e 1993) e próximos da mediana em Roraima de 1983 a 1990. Nos demais estados da região verificam-se taxas mais baixas (destacadas em verde) com ligeiro clareamento do tom no Amapá, Amazonas e Acre a partir de 2005.

Na região Nordeste a variação de cores indica flutuações nas taxas de mortalidade, com valores baixos em todo o período e suave tendência de aumento nos últimos dois anos, com exceção do Maranhão e Bahia (o verde mais escuro prevalece).

Em contrapartida, nas regiões Sudeste, Sul e Centro-Oeste observa-se taxas elevadas em São Paulo em todo o período e no Distrito Federal nos anos iniciais. No Rio de Janeiro as taxas são elevadas no início do período nos óbitos em geral e no sexo masculino. Nas demais Unidades da Federação que compõem estas regiões as taxas variaram entre patamares mais baixos e próximos da mediana, com tendência de aumento nos últimos três anos.

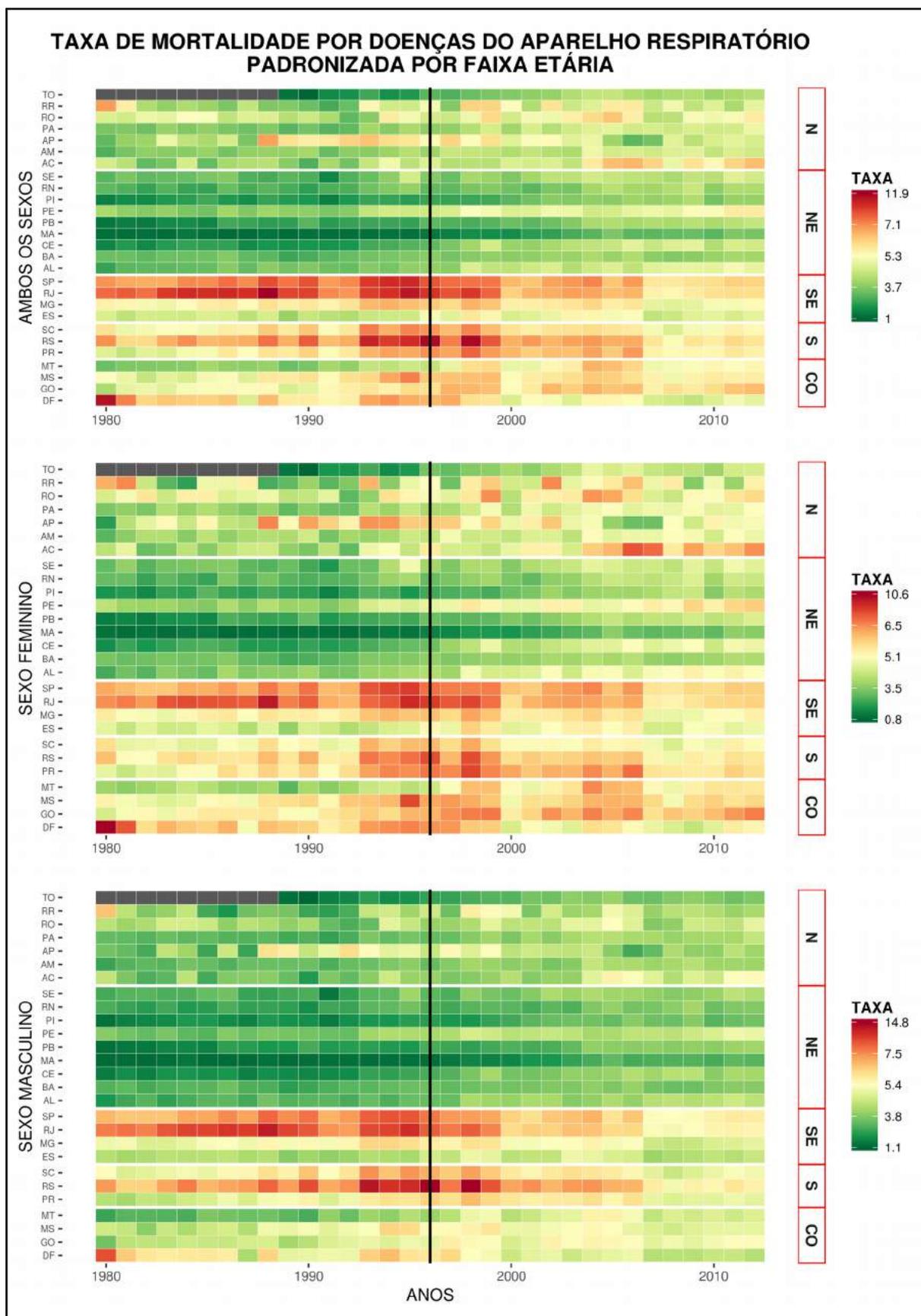


Figura 33 – Evolução da Taxa de Mortalidade por Doenças do Aparelho Respiratório segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012

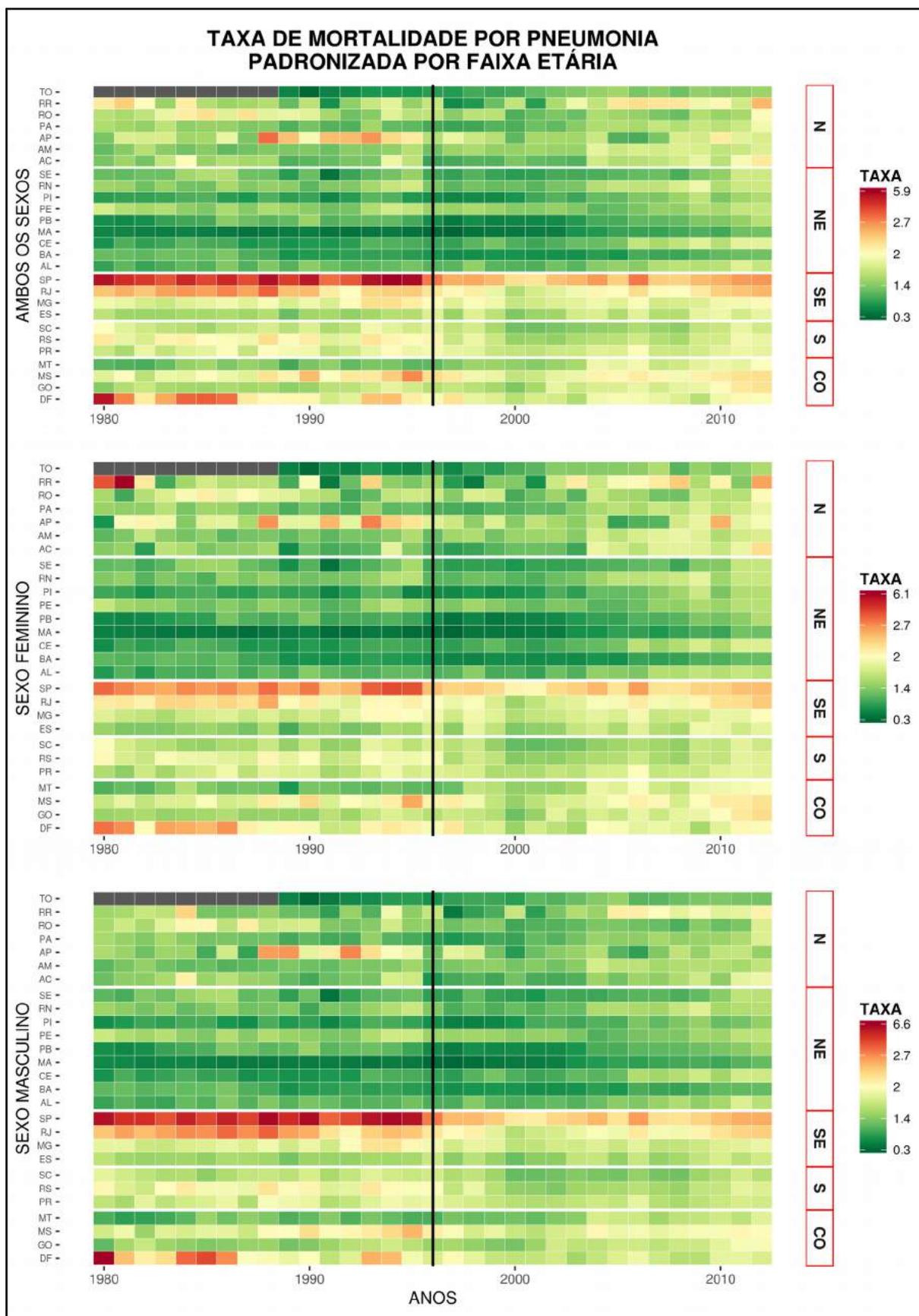


Figura 34 – Evolução da Taxa de Mortalidade por Pneumonia segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012

As taxas de mortalidade por Causas Externas foram expressivamente mais elevadas entre homens (Figura 35). No sexo feminino predominaram valores baixos em todo o período, com o Amapá em 1981 se destacando como *outlier*. O gradiente de cores indica taxas ligeiramente mais elevadas em Roraima e Rondônia e nas regiões Sudeste, Sul e Centro-Oeste em geral, onde os tons de verde foram mais claros, e valores baixos na região Nordeste em todo o período (prevalecendo o verde em tom escuro).

Quanto aos óbitos em geral e entre homens, observa-se que na Região Norte as taxas foram mais elevadas em Roraima, onde flutuaram com valores elevados entre os anos de 1988 e 1990 e entre 1996 e 2000 entremeados por valores próximos da mediana, e em Rondônia até 1996, decaindo a partir de então. Observa-se também taxas de mortalidade por Causas Externas elevadas no Amapá nos anos de 1981, 1994 e 1995, entretanto, prevaleceram valores mais baixos na primeira metade do período, elevando ligeiramente na segunda metade. Nos demais estados que compõe esta região as taxas eram mais baixas nos anos iniciais, aumentando nos últimos anos.

Entre homens e nos óbitos em geral, as taxas decorrentes de Causas Externas foram baixas na região Nordeste, aumentando gradativamente a partir de 1996, sendo o aumento mais agudo em Alagoas. Em Pernambuco as taxas foram as mais elevadas da região na primeira metade do período, aumentando ligeiramente entre 1996 e meados dos anos 2000, voltando a patamares mais próximos da mediana em seguida, com tendência de queda nos três últimos anos.

Na região Sudeste as taxas diminuíram gradativamente em São Paulo, enquanto no Rio de Janeiro aumentaram no início do período, atingindo o valor mais elevado em 1989, diminuindo de maneira constante a partir de então. As taxas aumentaram de maneira gradativa e contínua no Espírito Santo e permaneceram baixas em Minas Gerais durante todo o período analisado.

Já na região Centro-Oeste as taxas foram baixas nos anos iniciais no Mato Grosso, Goiás e Distrito Federal, oscilando para valores elevados, terminado o período com valores próximos da mediana nas duas primeiras e mais baixos na última. No Mato Grosso do Sul as taxas eram próximas da mediana no início do período, aumentando até 1996, a partir de então diminuindo e mantendo-se em patamares próximos à mediana nos anos subsequentes.

Diferente do grupo de Causa Externas como um todo, o contraste entre as taxas de mortalidade por Homicídios, foi mais intenso entre as Unidades da Federação (Figura 36). Os patamares das taxas foi mais elevado entre homens, com exceção de *outliers* em Roraima em 1990, 1996 e 2000.

Em relação aos óbitos em geral e no sexo masculino, as taxas em Roraima e Rondônia foram mais elevadas em meados da década de 1980 e no final dos anos 1990, atingindo valores próximos da mediana e baixos no final do período. Entre os estados desta região observa-se tendência de aumento suave nos últimos três anos analisados. Na região Nordeste o padrão foi semelhante, aumentando de taxas mais baixas para valores mais próximos da mediana na maioria dos estados, excetuando-se Pernambuco e Alagoas. Em Pernambuco as taxas apresentam aumento mais expressivo entre 1996 e 2009 quando atingem o patamar mais elevado, regredindo a partir de 2010, enquanto em Alagoas o aumento foi abrupto a partir de 2005.

Na região Sudeste as taxas oscilaram com valores baixos e próximos à mediana em São Paulo, com tendência de queda intensa a partir de meados dos anos 2000, assim como no Rio de Janeiro onde as taxas eram ainda mais baixas no início do período e atingiram patamares mais elevados, retornando a valores mais baixos mais ao final. No Espírito Santo as taxas aumentaram constantemente.

Na região Sul as taxas foram baixas em todo o período, enquanto no Centro-Oeste os valores eram mais baixos até 1995, aumentando para valores próximos da mediana na segunda metade do período analisado.

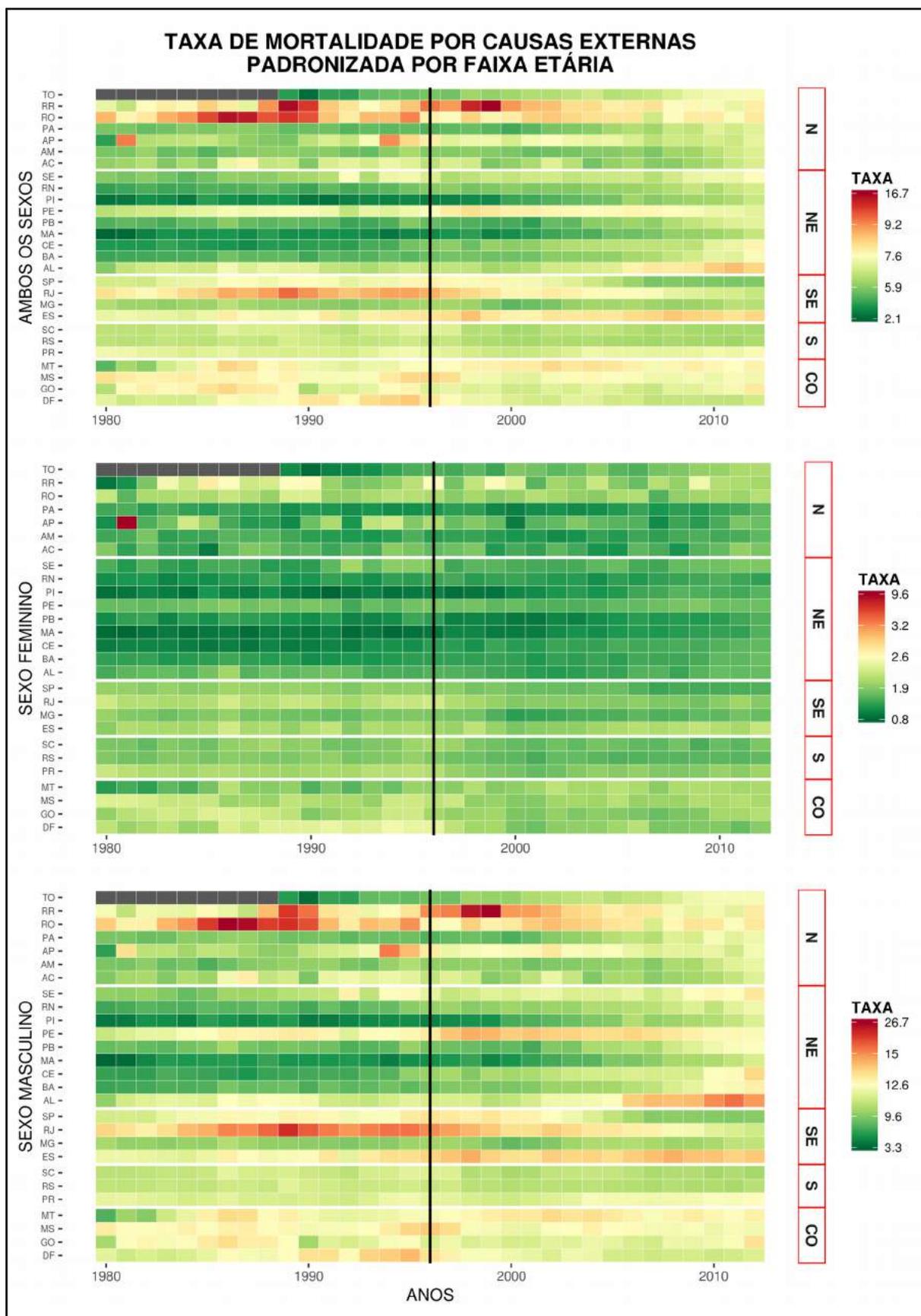


Figura 35 – Evolução da Taxa de Mortalidade por Causas Externas segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012

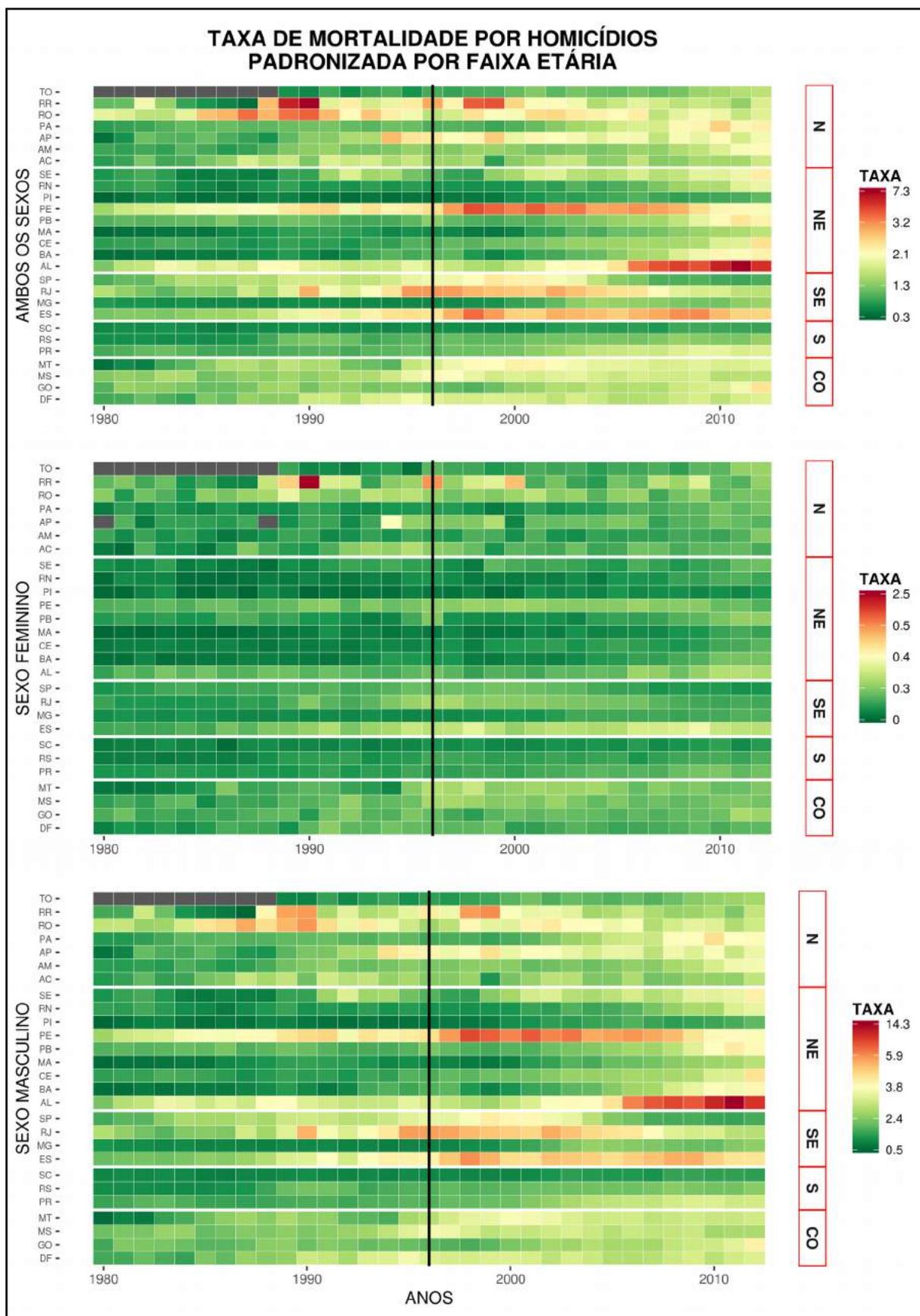


Figura 36 – Evolução da Taxa de Mortalidade por Homicídios segundo sexo, padronizada por faixa etária nas Unidades da Federação de 1980 a 2012

6.2 TENDÊNCIAS DAS TAXAS DE MORTALIDADE POR CAUSAS SELECIONADAS E INDICADORES SOCIAIS, ECONÔMICOS E DEMOGRÁFICOS

6.2.1 *COMPORTAMENTO DAS TAXAS DE MORTALIDADE POR CAUSAS SELECIONADAS*

Nesta etapa as causas de óbito selecionadas foram Doenças Infecciosas Intestinais, Neoplasias de Pulmão, Doenças Cerebrovasculares, Pneumonias e Homicídios. Todas as taxas de mortalidade, expressas por 100.000 habitantes, foram superiores entre homens sendo a maior razão de sexo observada em relação aos homicídios, em que a média da taxa entre homens (24,7, 27,5 e 37,2 em 1991, 2000 e 2010, respectivamente) foi cerca de nove vezes a média entre mulheres (2,7, 3,1 e 4,1).

A média das taxas de mortalidade por doenças infecciosas intestinais diminuiu de 8,9 para 2,5 entre homens e de 7,2 para 2,4 entre mulheres no período de 1991 a 2010. Da mesma forma, as doenças cerebrovasculares variaram de 58,6 para 43,6 entre homens e de 48,5 para 38,6 entre mulheres. Já as taxas de mortalidade por pneumonia aumentaram de 18,6 para 20,4 no sexo masculino e de 14,6 para 18,6 no feminino; comportamento equivalente à neoplasia de pulmão, que variou de 8,5 para 9,6 entre homens e de 3,2 para 3,9 entre mulheres.

Com exceção das doenças infecciosas intestinais, o comportamento das taxas nas regiões Sul e Sudeste foi similar para ambos os sexos, com padrão diferente do observado nas regiões Norte e Nordeste.

As taxas de mortalidade por doenças infecciosas intestinais no sexo feminino e masculino, foram ligeiramente mais elevadas nas regiões Norte e Nordeste, diminuindo em todas as regiões (Figura 37 e Figura 38). A redução da amplitude dos *boxplots* e do número de *outliers* aponta para uma distribuição mais homogênea destas taxas entre as microrregiões em 2010.

A taxa de mortalidade por neoplasia de pulmão no sexo feminino aumentou em todas as regiões e os patamares mais elevados se concentraram nas regiões Sul, Sudeste e Centro-Oeste; já no sexo masculino aumentou nas regiões Norte, Nordeste e Centro-Oeste, variando pouco nas regiões Sul e Sudeste, onde os valores foram mais elevados (Figura 39 e Figura 40).

As taxas de mortalidade por doenças cerebrovasculares aumentaram nas regiões Norte, Nordeste e Centro-Oeste e diminuíram nas regiões Sul e Sudeste, de maneira que as medianas de todas as regiões se aproximaram em 2010 em ambos os sexos (Figura 41 e Figura 42).

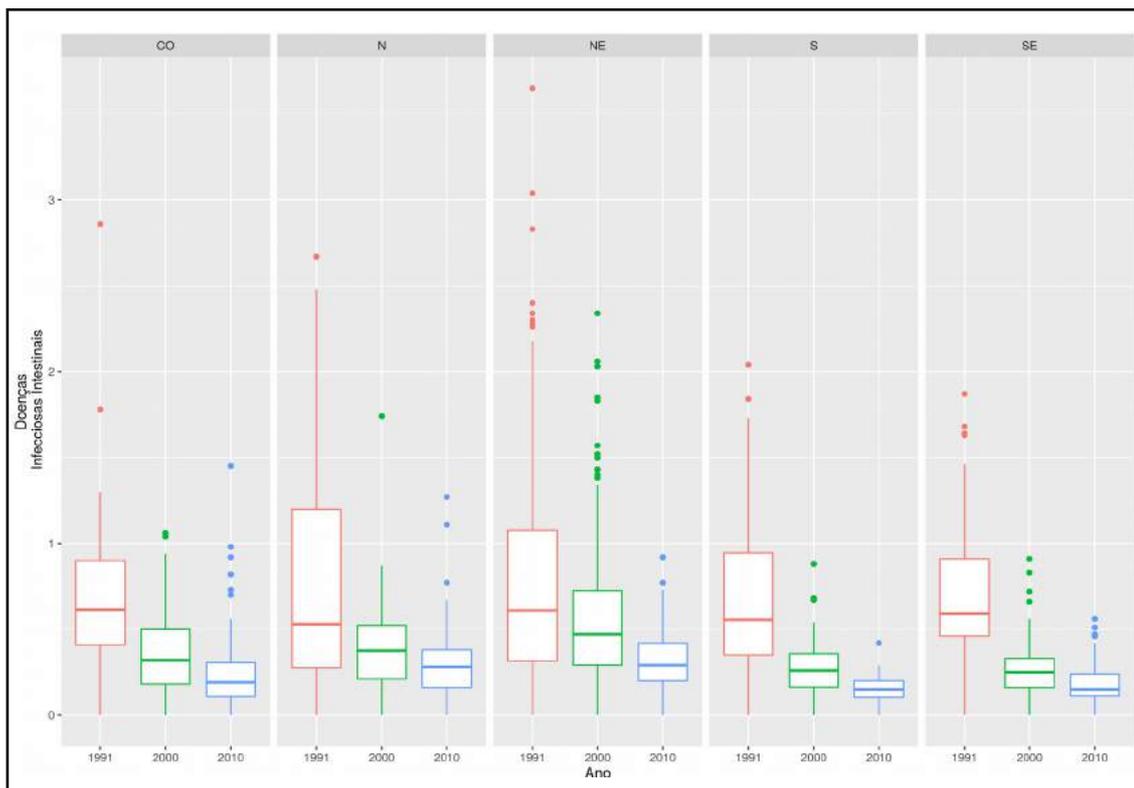


Figura 37 – *Boxplots* das Taxas de Mortalidade por Doenças Infecciosas Intestinais no Sexo Feminino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010

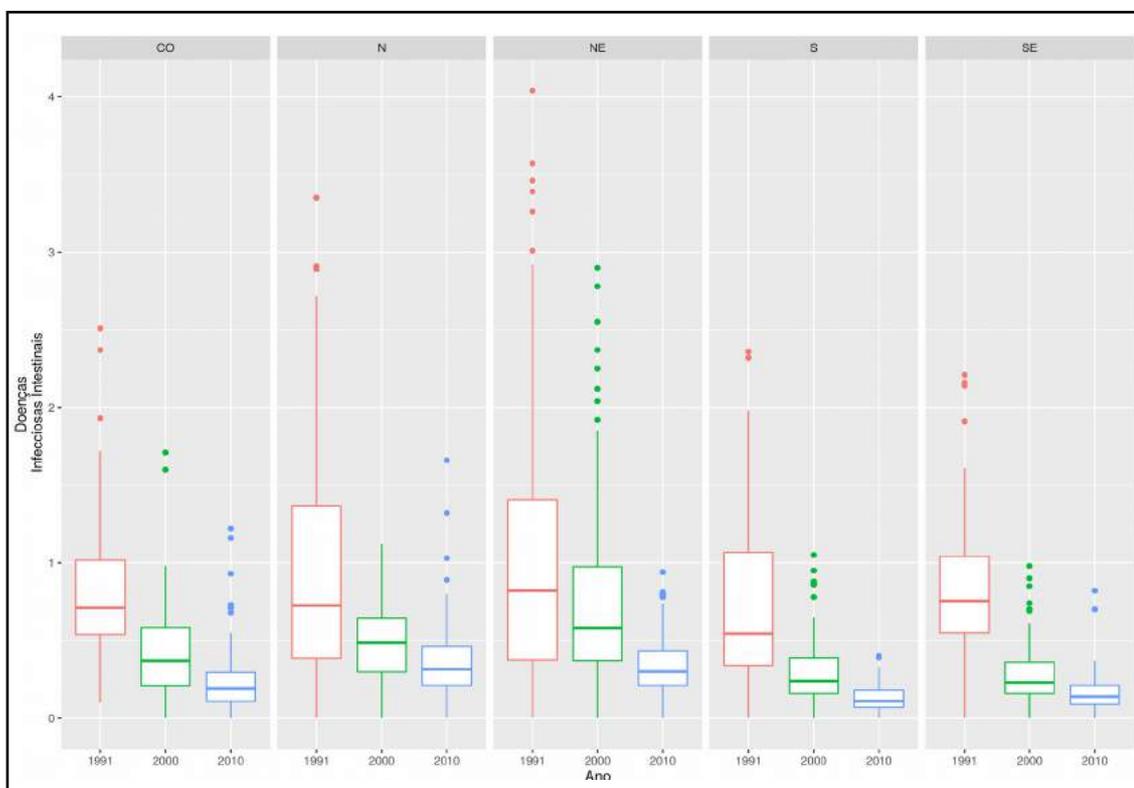


Figura 38 – *Boxplots* das Taxas de Mortalidade por Doenças Infecciosas Intestinais no Sexo Masculino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010

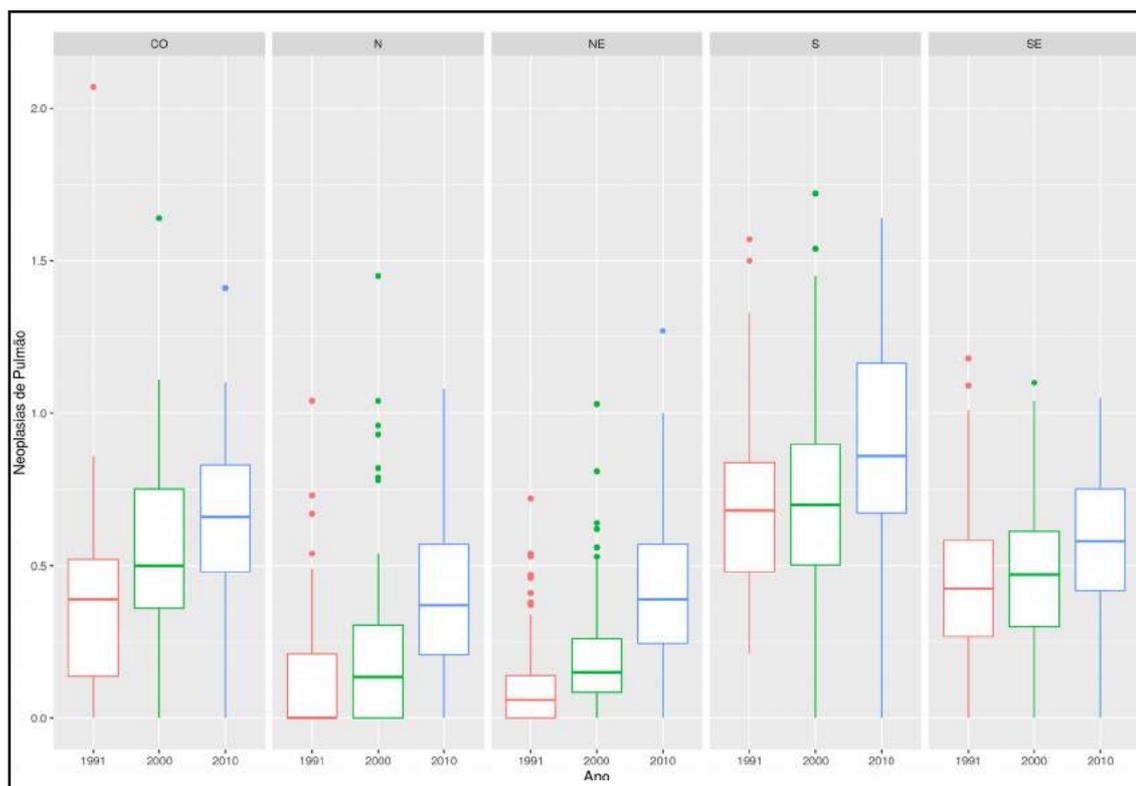


Figura 39 – *Boxplots* das Taxas de Mortalidade por Neoplasias de Pulmão no Sexo Feminino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010

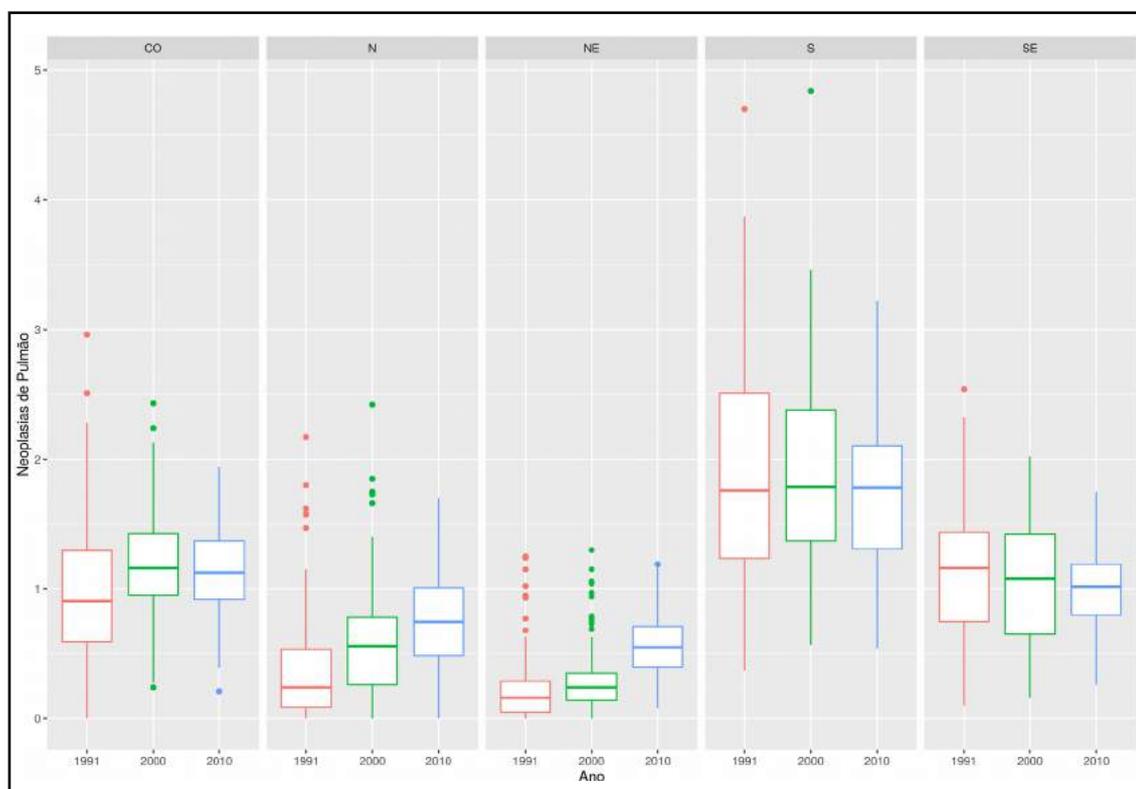


Figura 40 – *Boxplots* das Taxas de Mortalidade por Doenças Cerebrovasculares no Sexo Feminino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010

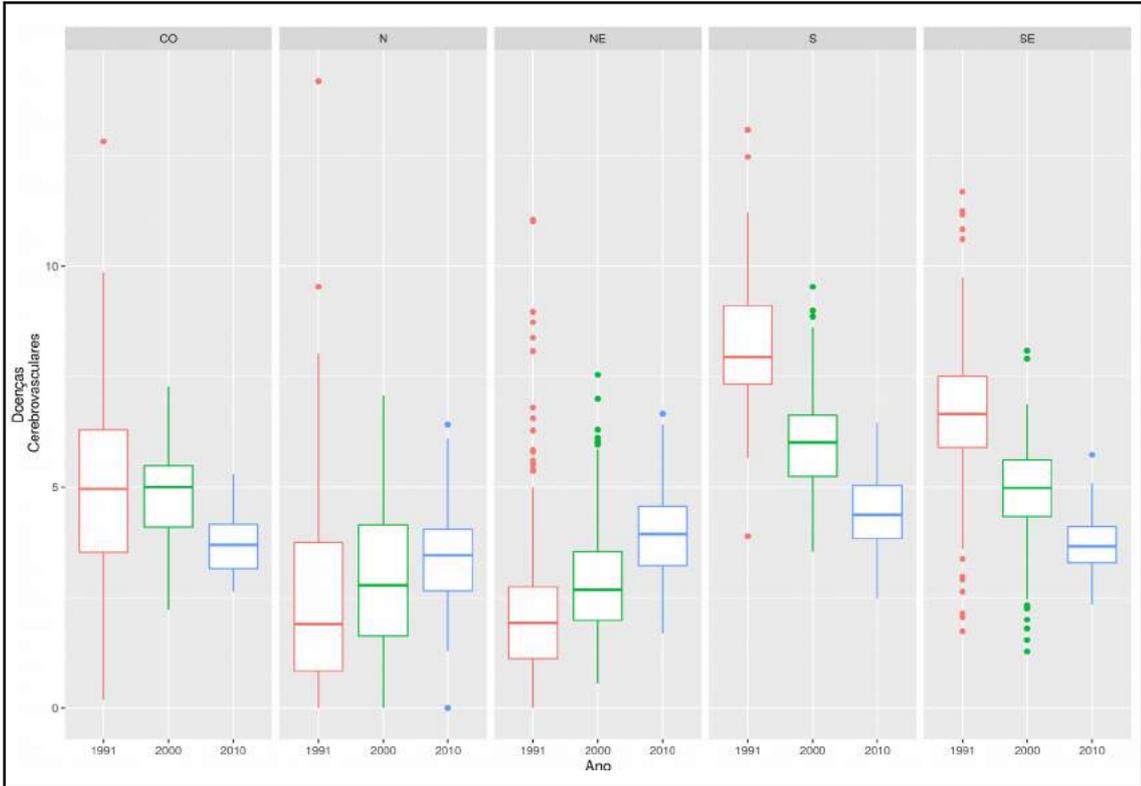


Figura 41 – *Boxplots* das Taxas de Mortalidade por Doenças Cerebrovasculares no Sexo Feminino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010

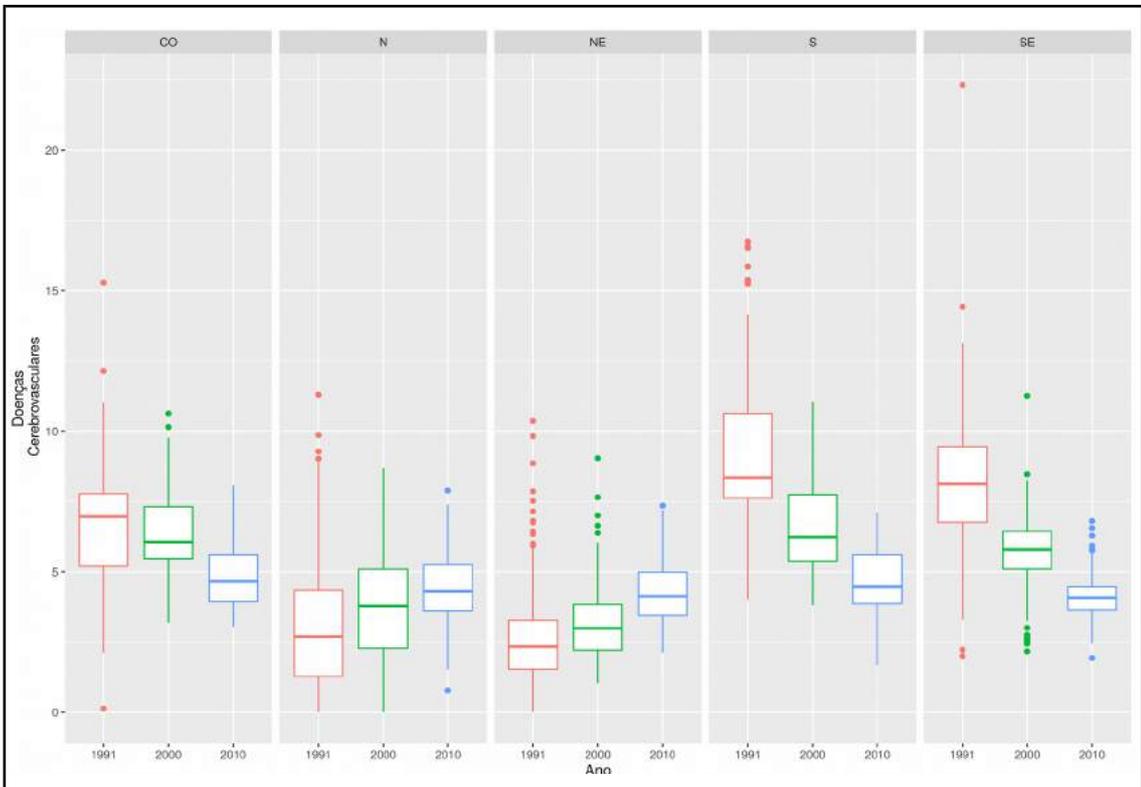


Figura 42 – *Boxplots* das Taxas de Mortalidade por Doenças Cerebrovasculares no Sexo Masculino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010

O comportamento das taxas de mortalidade por pneumonia foi de aumento nas regiões Norte, Nordeste e Centro-Oeste (Figura 43 e Figura 44). Nas regiões Sul e Sudeste, onde se concentraram os valores mais elevados em todo o período, observou-se diminuição no sexo masculino e estabilidade no sexo feminino, ocorrendo queda entre 1991 e 2000 em ambos os sexos.

Em relação aos homicídios, observou-se que as taxas de mortalidade no sexo feminino aumentaram em todo o país, de maneira mais aguda nas regiões Norte e Nordeste, com os valores mais elevados concentrados na região Centro-Oeste em todo o período (Figura 40). No sexo masculino as taxas de mortalidade por homicídios também aumenta em todas as regiões (Figura 46). Em 2010 as medianas nas regiões Norte, Nordeste e Centro-Oeste foram as mais elevadas do país, com valores próximos de 40. Ao contrário das demais causas, observou-se uma elevada concentração de *outliers*, com destaque para as regiões Nordeste e Sudeste, e uma manutenção de grande variabilidade da distribuição de taxas nas microrregiões em todo o período.

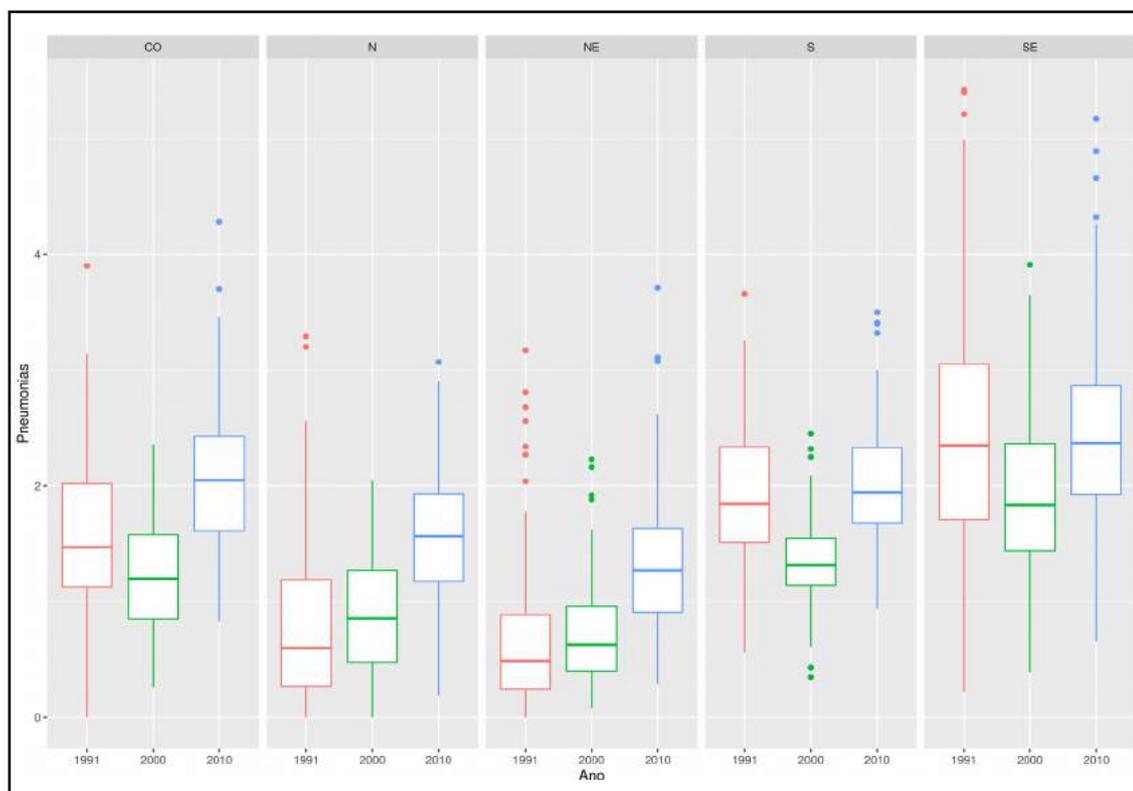


Figura 43 – *Boxplots* das Taxas de Mortalidade por Pneumonia no Sexo Feminino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010

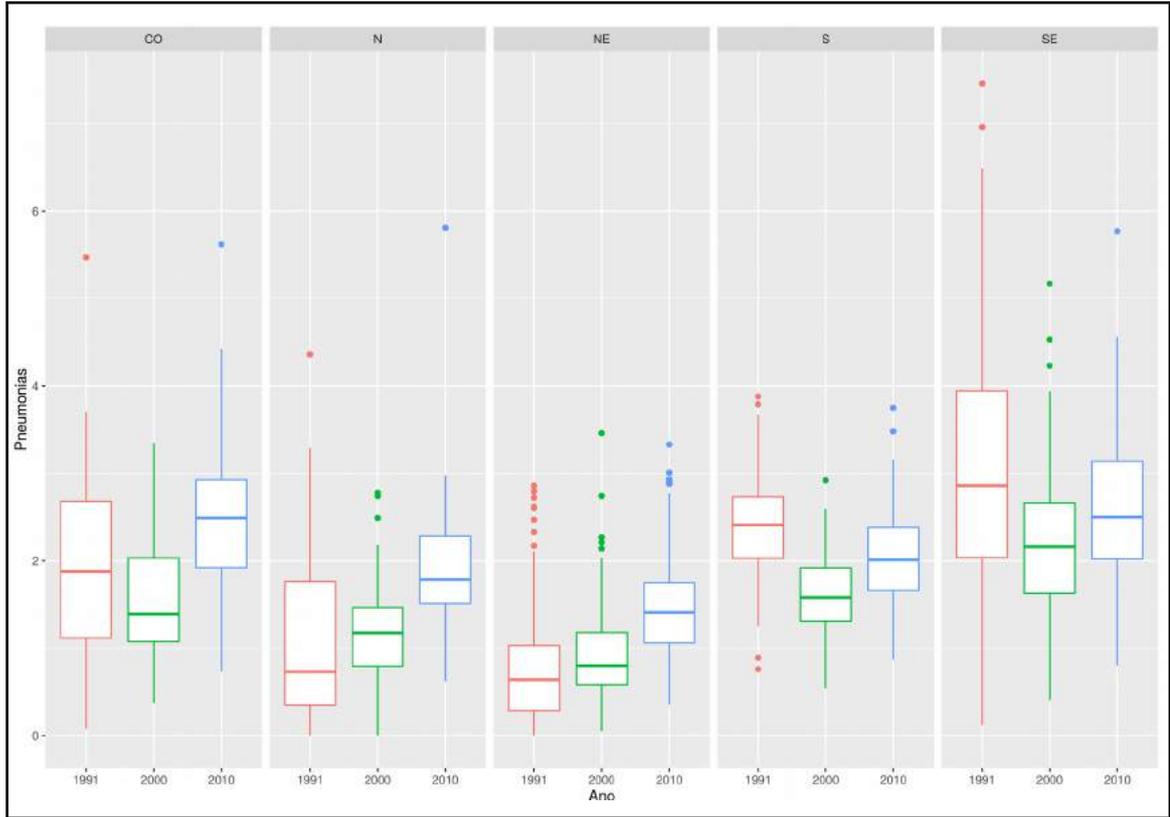


Figura 44 – *Boxplots* das Taxas de Mortalidade por Pneumonia no Sexo Masculino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010

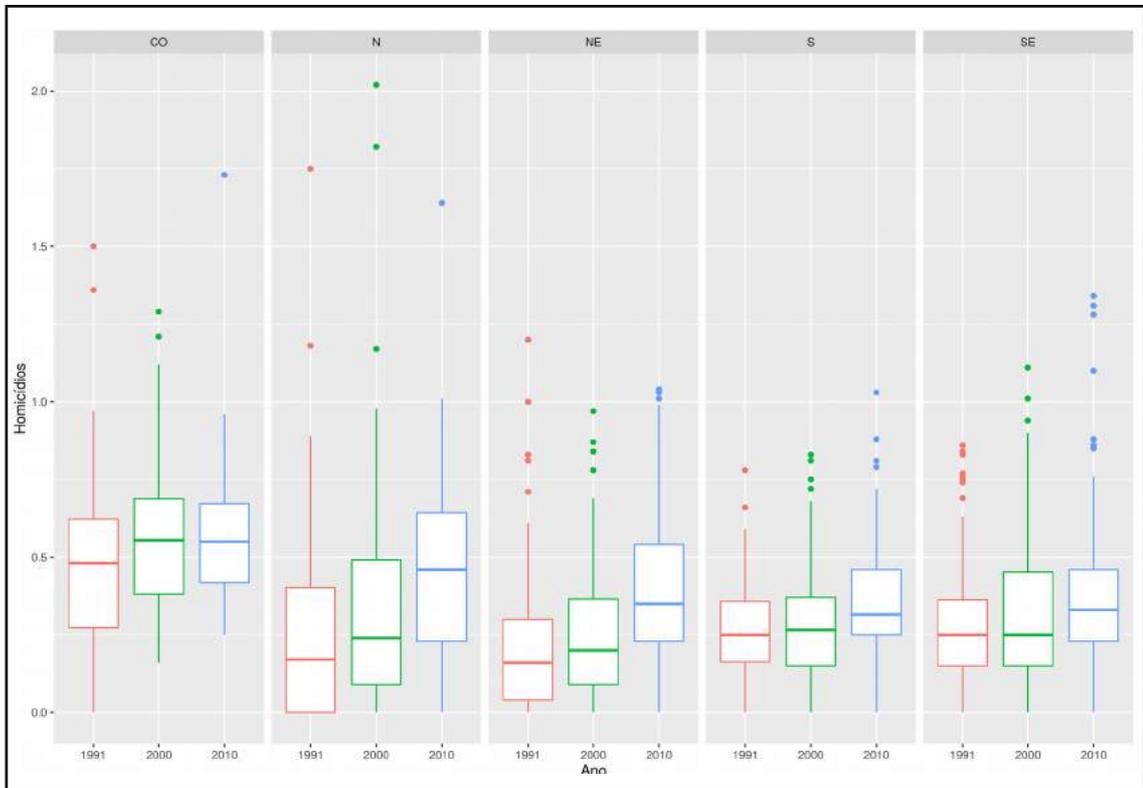


Figura 45 – *Boxplots* das Taxas de Mortalidade por Homicídios no Sexo Feminino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010

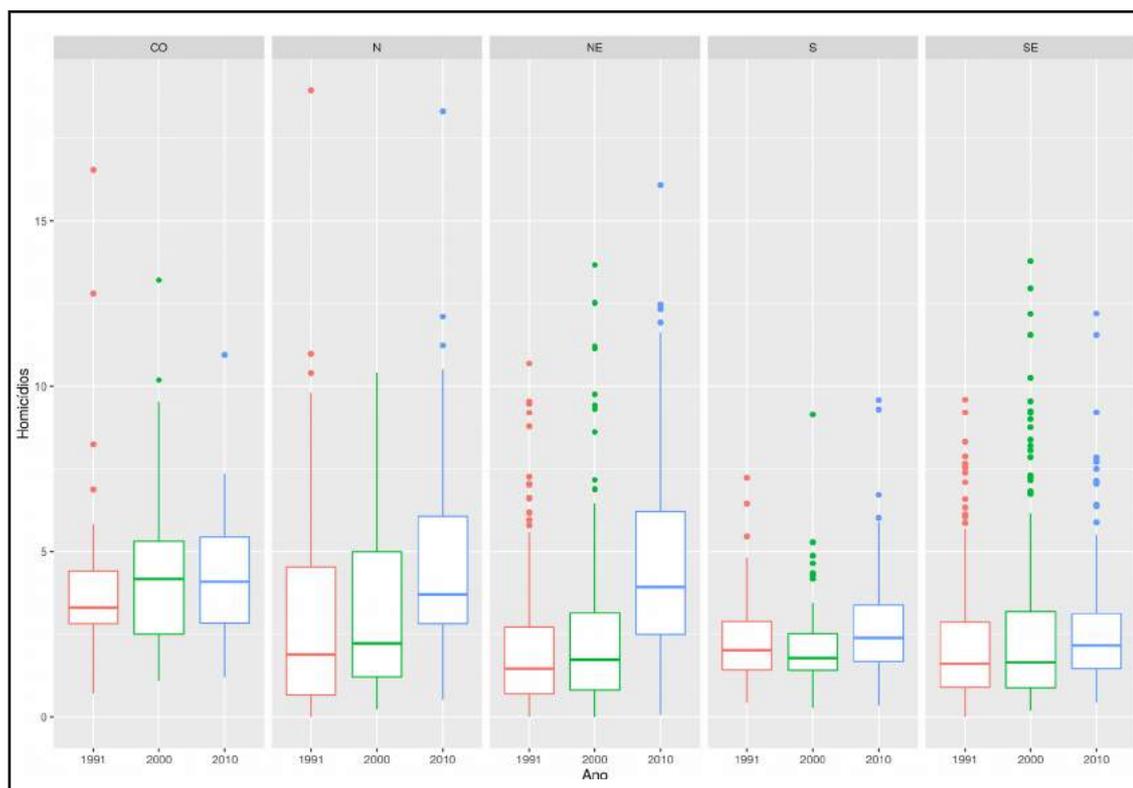


Figura 46 – *Boxplots* das Taxas de Mortalidade por Homicídios no Sexo Masculino Padronizadas por Faixa Etária segundo Região em 1991, 2000 e 2010

6.2.2 SELEÇÃO E COMPORTAMENTO DOS INDICADORES SOCIAIS, ECONÔMICOS E DEMOGRÁFICOS

Os resultados da Análise Fatorial se encontram na Tabela 1. Foram selecionados dois fatores em cada ano que explicaram 74,7%, 77,3% e 77,9% da variabilidade dos indicadores em 1991, 2000 e 2010, respectivamente, e em todos os anos a Taxa de Analfabetismo, a Proporção de Pessoas com Baixa Renda, a Razão de Dependência de Jovens, a Renda e a Proporção da População que cursou ao menos ensino médio estiveram fortemente correlacionados com o primeiro fator, enquanto a Razão de Dependência de Idosos e o Índice de Envelhecimento estiveram correlacionados com o segundo fator. A Taxa de Desemprego e o Percentual de Domicílios com acesso à Rede Geral de Esgoto ou Pluvial foram os indicadores menos correlacionados com os fatores.

Tabela 1 – Resultado da Análise Fatorial Exploratória com os Indicadores Sociais, Econômicos e Demográficos em 1991, 2000 e 2010

INDICADORES	CARGA FATORIAL					
	1991		2000		2010	
	FATOR		FATOR		FATOR	
	I	II	I	II	I	II
Taxa de Analfabetismo	0,85		- 0,90	0,16	0,92	
Taxa de Desemprego		- 0,31		- 0,25	0,35	- 0,36
Percentual de Domicílios com acesso a Rede Geral de Esgoto ou Pluvial	- 0,66	0,25	0,66	0,23	- 0,53	0,39
Índice de Envelhecimento Populacional	- 0,34	0,94	0,52	0,85	- 0,33	0,94
Proporção de Pessoas com Baixa Renda	0,98		- 0,99		0,93	- 0,30
Razão de Dependência de Idosos	0,26	0,89		0,96	0,13	0,97
Razão de Dependência de Jovens	0,77	- 0,42	- 0,85	- 0,34	0,68	- 0,60
Renda Média Domiciliar Per Capita	- 0,99		0,97	- 0,13	- 0,94	0,18
Proporção da População que Cursos ao Menos o Ensino Médio	- 0,90		0,91	- 0,12	- 0,90	
AUTOVALOR	4,70	2,02	5,00	1,95	4,41	2,60
VARIÂNCIA ACUMULADA	74,7%		77,3%		77,9%	

As variáveis selecionadas para análise de correlação com as taxas de mortalidade foram Renda Média Domiciliar Per Capita (Renda) e Razão de Dependência de Idosos (RDI), indicadores mais correlacionados com os fatores 1 e 2, Percentual de Domicílios com Acesso a Rede Geral de Esgoto ou Pluvial (Esgotamento), por se tratar de um indicador de serviços básicos de saneamento potencialmente relacionados com algumas causas de mortalidade, e Taxa de Desemprego (Desemprego), pela menor correlação com os demais indicadores.

Observou-se aumento na média da Renda Per Capita Média, no Brasil, de R\$ 220 para R\$ 541 entre 1991 e 2010, refletindo no aumento deste indicador nas cinco regiões (Figura 47). Em todo o período as medianas foram mais elevadas nas regiões Sul, Sudeste e Centro-Oeste enquanto as mais baixas se concentraram na região Nordeste, onde se destacaram *outliers* em todos os anos, referentes às microrregiões que compõem as capitais.

A Razão de Dependência de Idosos aumentou em todas as regiões, com elevação mais acentuada no Centro-Oeste (Figura 48). As maiores medianas se concentraram nas regiões Nordeste, Sul e Sudeste enquanto as menores estiveram sempre na região Norte, intensificando a diferença em relação às demais regiões.

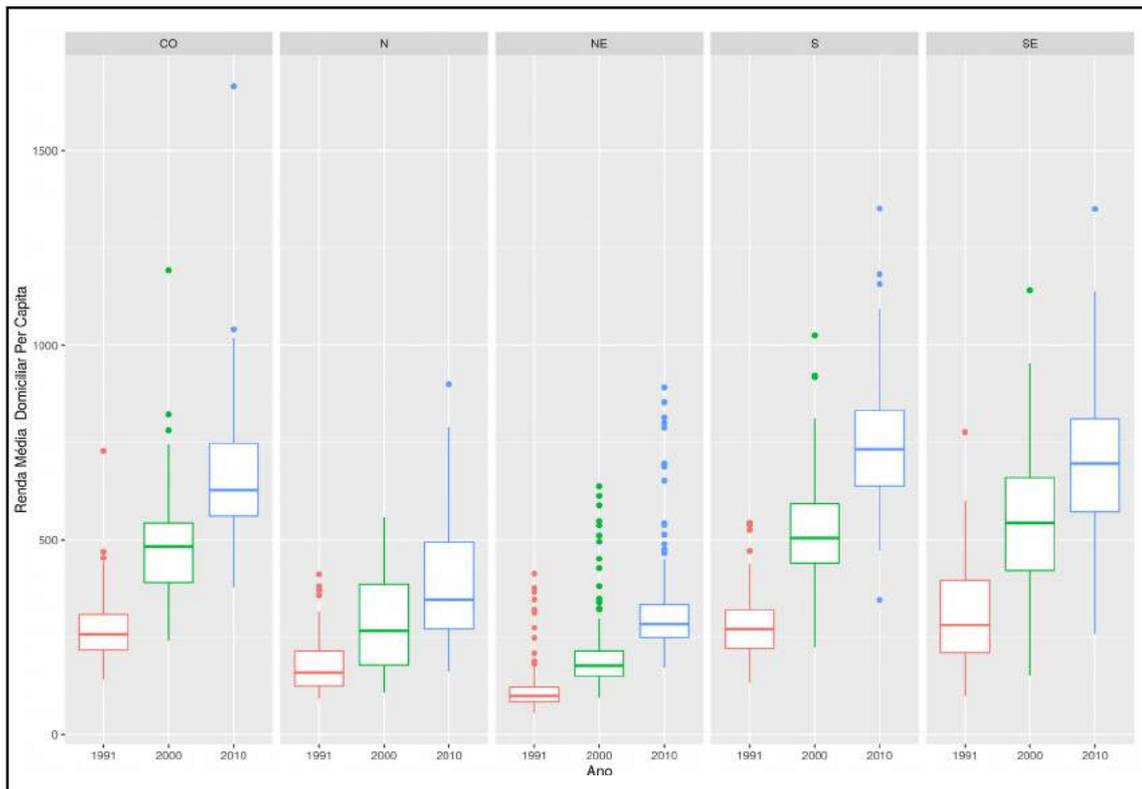


Figura 47 – *Boxplots* da Renda Per Capita Média segundo Região em 1991, 2000 e 2010

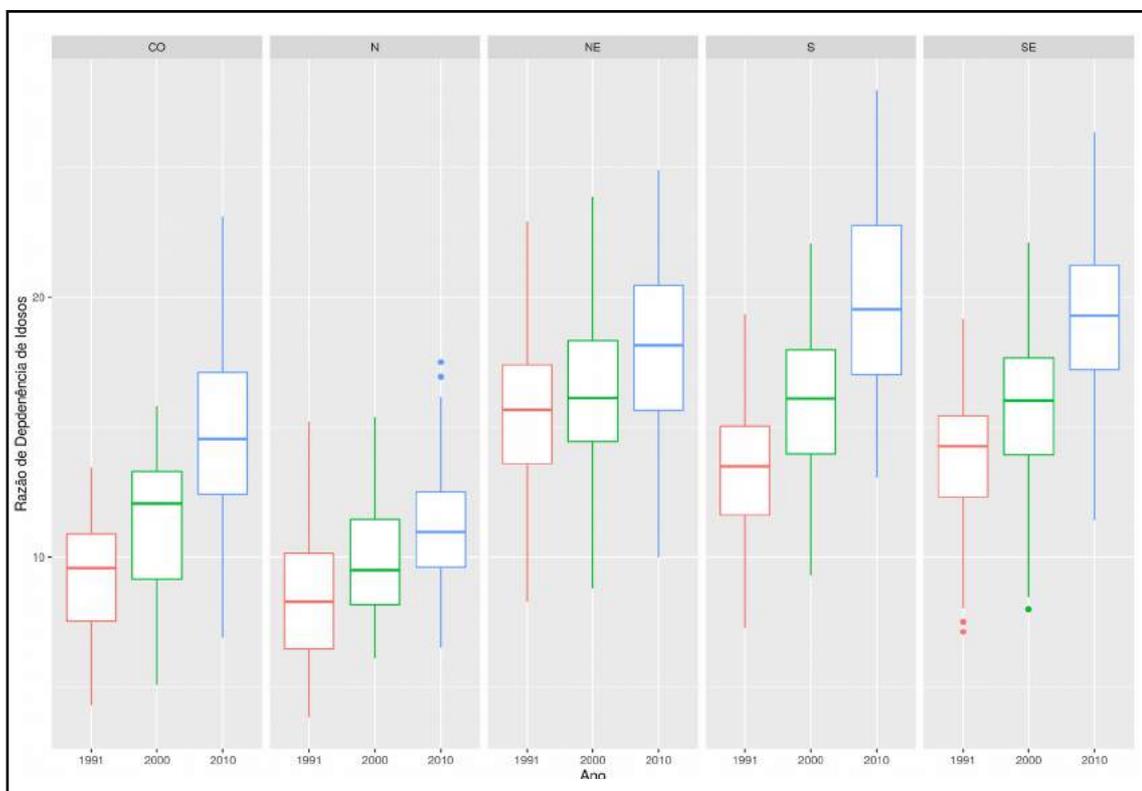


Figura 48 – *Boxplots* da Razão de Dependência de Idosos segundo Região em 1991, 2000 e 2010

Já o comportamento da Taxa de Desemprego foi diferente daquele observado nos demais indicadores selecionados, com tendência geral de aumento entre 1991 e 2010, e um pico em todas as regiões em 2000 (Figura 49). Os maiores valores foram observados nas regiões Norte e Nordeste, seguido das regiões Sudeste, Centro-Oeste e Sul. Observou-se também uma concentração de microrregiões *outliers* com valores elevados na região Nordeste.

O Percentual de Domicílios com Acesso à Rede Geral de Esgoto ou Pluvial apresentou as diferenças regionais mais intensa no período (Figura 50). Na região Sudeste, já em 1991, pelo menos metade das microrregiões tinham mais de 50% de domicílios com acesso a esgotamento sanitário, superando 75% em 2010. Nas demais regiões até 2010 a maioria das microrregiões não apresentavam sequer 40% dos domicílios com acesso à rede geral de esgoto ou pluvial.

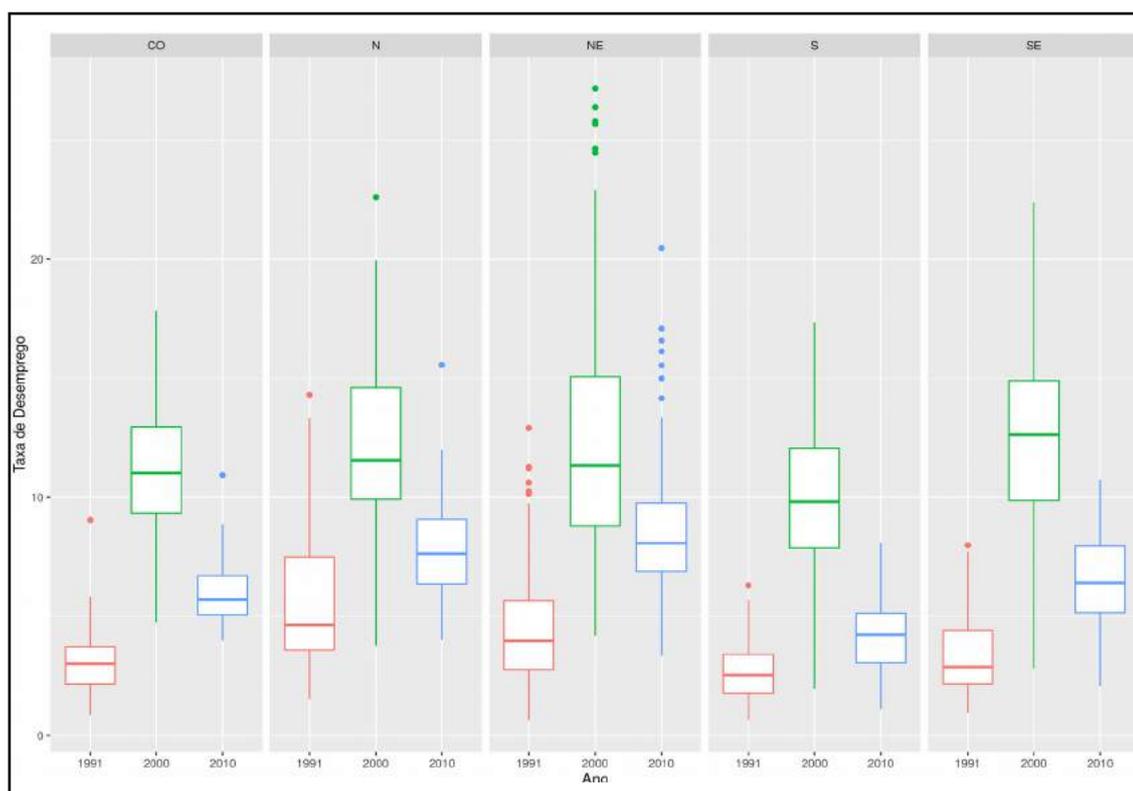


Figura 49 – Boxplots da Taxa de Desemprego segundo Região em 1991, 2000 e 2010

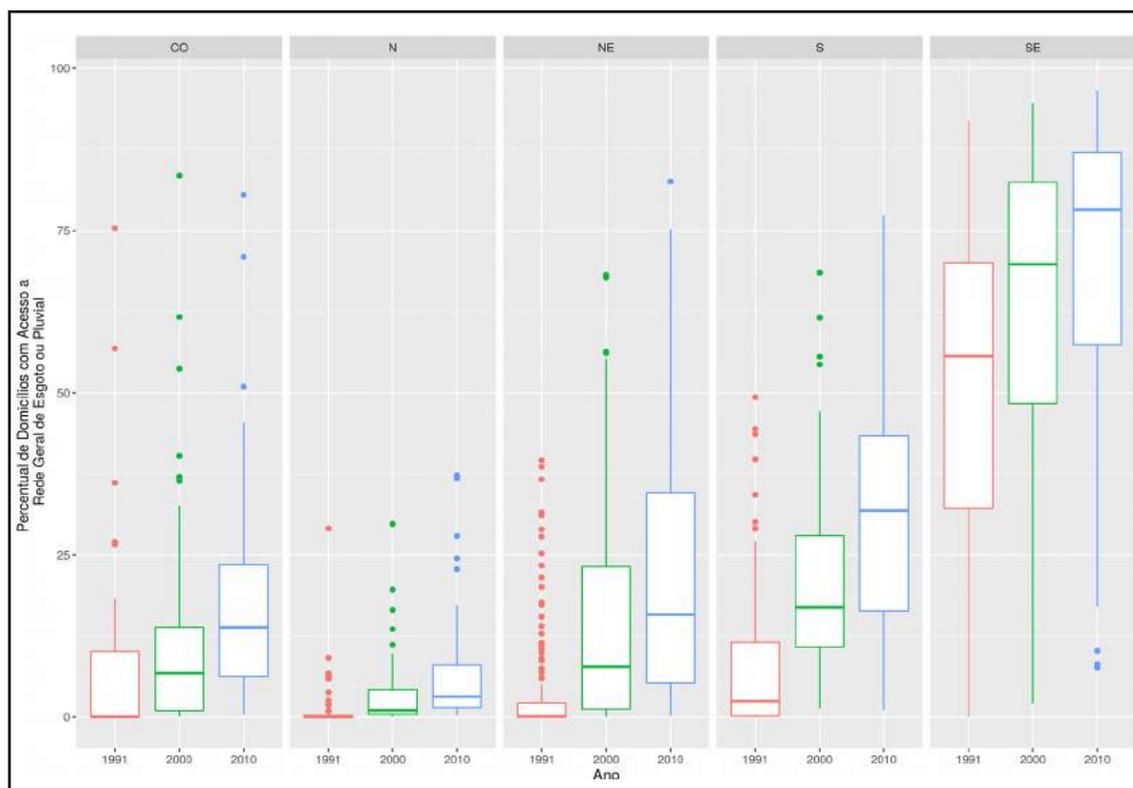


Figura 50 – *Boxplots* do Percentual de Domicílios com Acesso à Rede Geral de Esgoto ou Pluvial segundo Região em 1991, 2000 e 2010

6.2.3 CORRELAÇÃO ENTRE OS INDICADORES SELECIONADOS E AS TAXAS DE MORTALIDADE

A Figura 51 apresenta os coeficientes de correlação entre cada taxa de mortalidade, por sexo, e os indicadores selecionados, em cada ano. O diâmetro dos círculos e a tonalidade das cores são proporcionais ao valor do coeficiente de correlação enquanto a cor azul representa correlação negativa e a cor vermelha representa correlação positiva.

As Doenças Infecciosas Intestinais apresentaram correlação negativa e fraca/moderada com a renda e com o esgotamento em ambos os sexos, com no período, o que indica que as microrregiões com renda mais elevada e maior percentual de domicílios com acesso à rede geral e fluvial de esgoto apresentaram, em média, menores taxas de mortalidade por este grupo de doenças.

A correlação entre neoplasia de brônquios e pulmão foi semelhante em ambos os sexos, sendo: positiva e moderada/forte com a renda, flutuando no período; positiva e fraca com esgotamento, e com tendência de queda; e negativa e fraca com o desemprego. Neste

sentido, nas microrregiões onde a renda per capita foi mais elevada, as taxas de mortalidade por esta causa foram maiores.

As pneumonias também apresentaram correlação positiva e moderada com renda e com esgotamento, ambas decaindo entre 1991 e 2010, de maneira ligeiramente mais acelerada no sexo masculino. De maneira semelhante, a correlação das doenças cerebrovasculares foi positiva e moderada com a renda e positiva e fraca com o esgotamento até 2000, não estando correlacionada com nenhum dos indicadores em 2010.

A taxa de mortalidade por homicídios apresentou correlação negativa e fraca com a Razão de Dependência de Idosos no sexo feminino, diminuindo em intensidade no período, e negativa e moderada no sexo masculino, com ligeiro aumento na força da correlação. No sexo masculino se observou uma correlação positiva e fraca/moderada com o desemprego, com tendência de aumento.

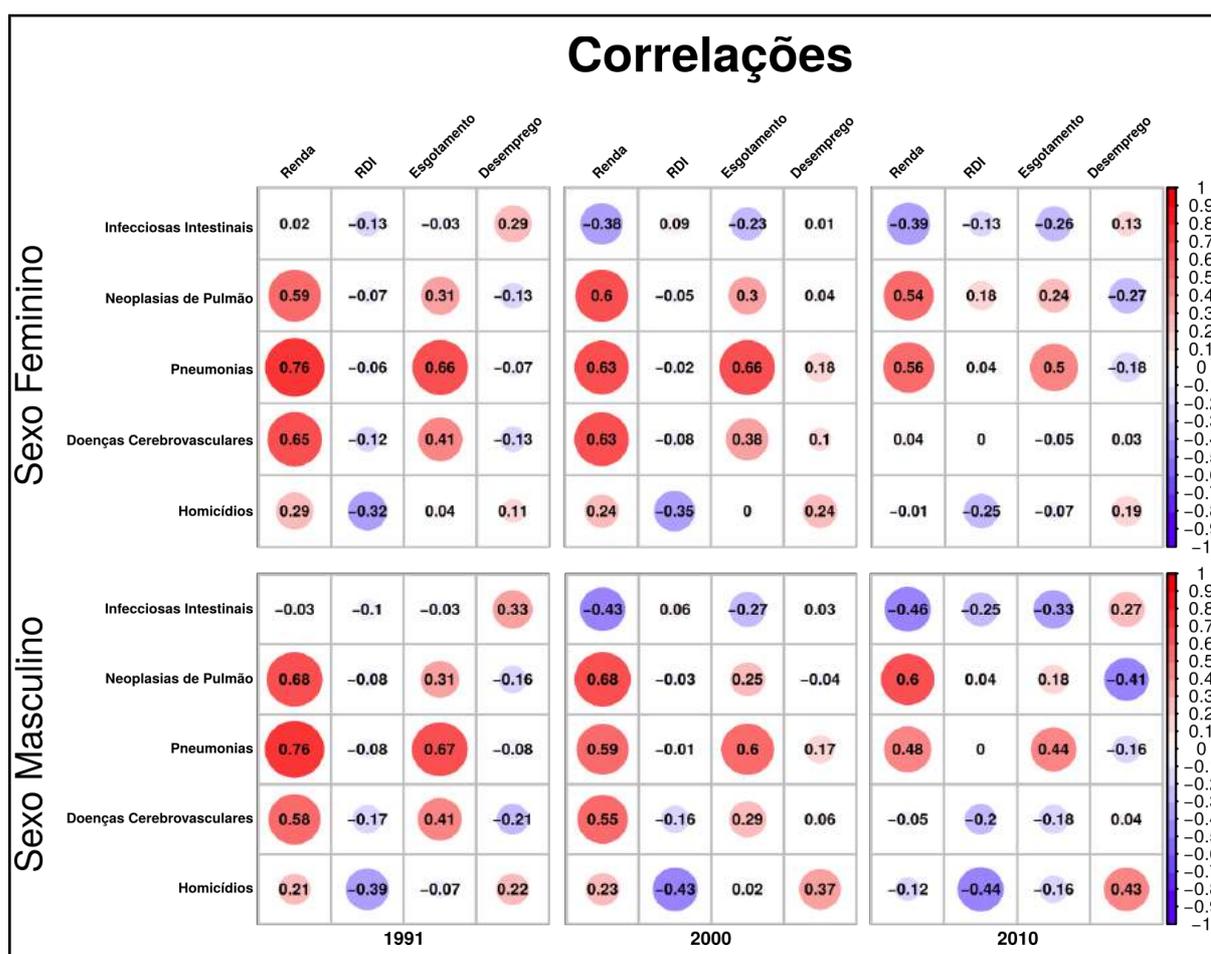


Figura 51 – Coeficientes de Correlação entre as Taxas de Mortalidade por Causas Seleccionadas e os Indicadores Sociais Econômicos e Demográficos Seleccionados em 1991, 2000 e 2010

6.3 MINERAÇÃO DE DADOS DAS INFORMAÇÕES INDIVIDUAIS DO SIM E A IDENTIFICAÇÃO DE PADRÕES DE MORTALIDADE NO BRASIL

6.3.1 ABORDAGEM 1 – MORTALIDADE GERAL SEGUNDO GRUPOS DE CAUSA

Os resultados do modelo de árvore de decisão da Abordagem 1, considerando a mortalidade geral segundo grupos de causa, se encontram na Figura 52. Nesta abordagem a variável mais influente para a classificação do grupo de causa de óbito foi o Local de Ocorrência do Óbito. O ramo à esquerda indica que óbitos ocorridos em via pública e em outros locais, que representaram 8% dos dados abordados, foram em sua maioria por Causas Externas, sendo este grupo de causa responsável por 68% dos óbitos ocorridos nestes locais. Para os óbitos ocorridos em domicílio a Unidade da Federação também aparece como variável importante para a classificação do grupo de causa, e cerca de 36% dos óbitos domiciliares ocorridos nas regiões Sudeste, Sul e Centro-Oeste, além de Amapá, Rondônia e Roraima foram por Doenças do Aparelho Circulatório. Em contrapartida, 59% dos óbitos domiciliares ocorridos nas demais Unidades da Federação das regiões Norte e Nordeste, foram por Causas Mal Definidas.

Entre os óbitos ocorridos em estabelecimentos de saúde, na Faixa Etária de 30 ou mais anos de idade, 37% dos óbitos foram por Doenças do Aparelho Circulatório, entre os indivíduos de 1 a 29 anos, 38% dos óbitos foram por Causas Externas e entre os óbitos infantis (menores de 1 ano de idade), 65% dos óbitos foram por Demais Causas.

A Tabela 2 apresenta as medidas de qualidade de predição do modelo. O modelo apresentou valores baixos de Acurácia (0,433) e Índice Kappa (0,286), com maior Sensibilidade para classificação dos óbitos por Doenças do Aparelho Circulatório (0,822) e maior Especificidade para Causas Externas (0,932), Causas Mal Definidas (0,944) e Demais Causas (0,975). A taxa de detecção mais elevada foi a de Doenças do Aparelho Circulatório (0,252).

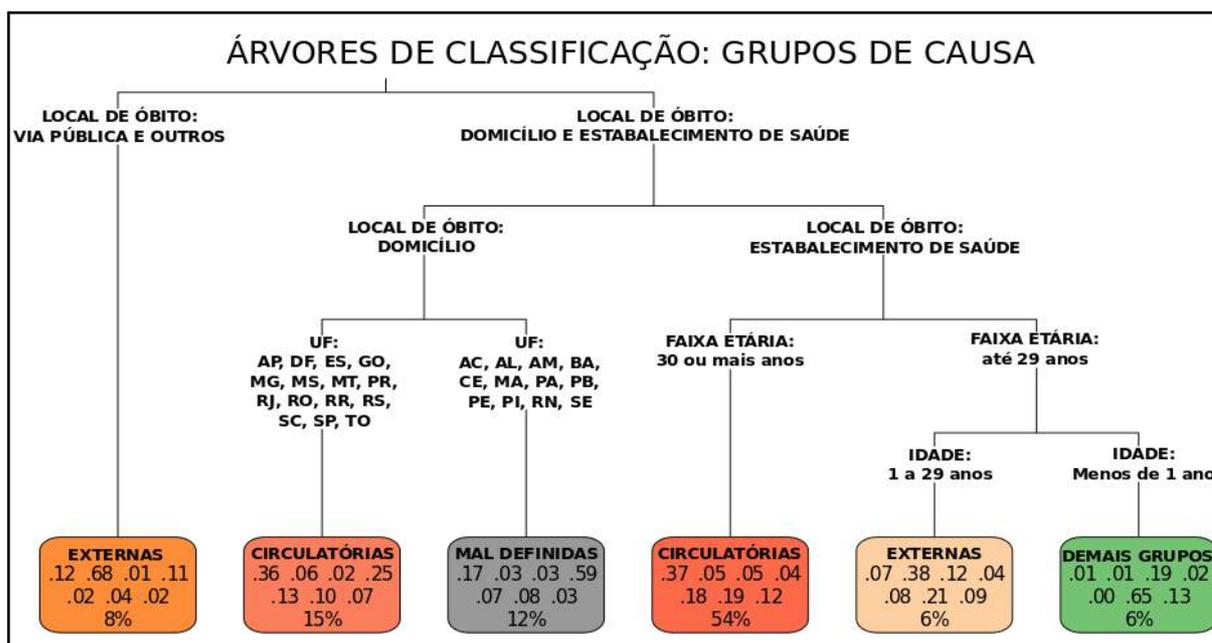


Figura 52 – Árvore de Classificação obtida pelo modelo *CART* na Abordagem 1: Mortalidade Geral Segundo Grupos de Causa

Tabela 2 – Medidas de Qualidade de Predição do Modelo *CART* da Abordagem 1

CAUSAS	MEDIDAS DE QUALIDADE					
	Sensibilidade	Especificidade	VPP	VPN	Prevalência	Taxa Detecção
Circulatórias	0,822	0,348	0,365	0,893	0,286	0,252
Externas	0,630	0,932	0,552	0,950	0,117	0,073
Infeciosas	0	1	-	0,947	0,052	0
Mal Definidas	0,491	0,944	0,590	0,918	0,140	0,069
Neoplasias	0	1	-	0,871	0,128	0
Demais Causas	0,207	0,975	0,649	0,848	0,180	0,037
Respiratórias	0	1	-	0,871	0,128	0

Acurácia: 0,433

Kappa: 0,286

O Decréscimo Médio do Índice de Impureza de Gini das variáveis independentes obtidos a partir do modelo *Random Forest*, apresentados na Figura 53, apontam para o Local de Ocorrência do Óbito e a Faixa Etária como as variáveis mais influentes, e as Unidades da Federação com uma importância um pouco menor.

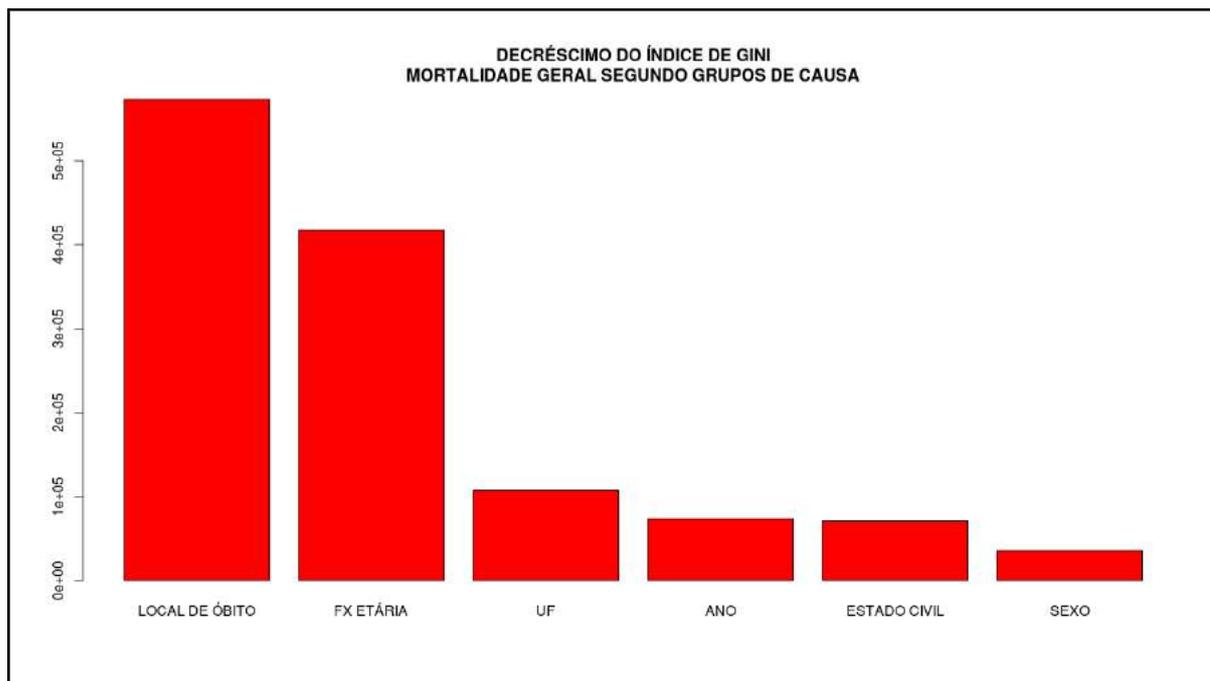


Figura 53 – Decréscimo Médio do Índice de Impureza de Gini das Variáveis Independentes na Abordagem 1 obtidas pelo *Random Forest*

O gráfico da Figura 54 mostra os resultados da Análise de Correspondência Múltipla utilizando o Grupo de Causa de Óbito e as três variáveis mais influentes, enquanto na Figura 55 se encontra os resultados da Análise de Correspondência apenas com Grupos de Causa, Local de Óbito e Faixa Etária.

A variância explicada na Análise de Correspondência Múltipla utilizando a Unidade da Federação (Figura 54) foi menor do que quando esta variável não foi utilizada (Figura 55). A elipse em azul na Figura 54 destaca a proximidade entre Causas Mal Definidas, Domicílio e um grupo de Unidades da Federação que compõem as regiões Norte e Nordeste, relação que também foi destacada pelo modelo de árvore.

As categorias Via Pública, Causas Externas e 19 a 29 anos, destacadas pela elipse em vermelho (Figura 54 e Figura 55), estiveram próximas independente da inclusão da Unidade da Federação, assim como as Doenças do Aparelho Circulatório, Doenças do Aparelho Respiratório, Neoplasias, Faixa Etária de 60 a 79 anos e Estabelecimento de Saúde, destacadas pela elipse em verde, e as Causas Mal Definidas, 80 ou mais anos e Domicílio, na elipse em preto.

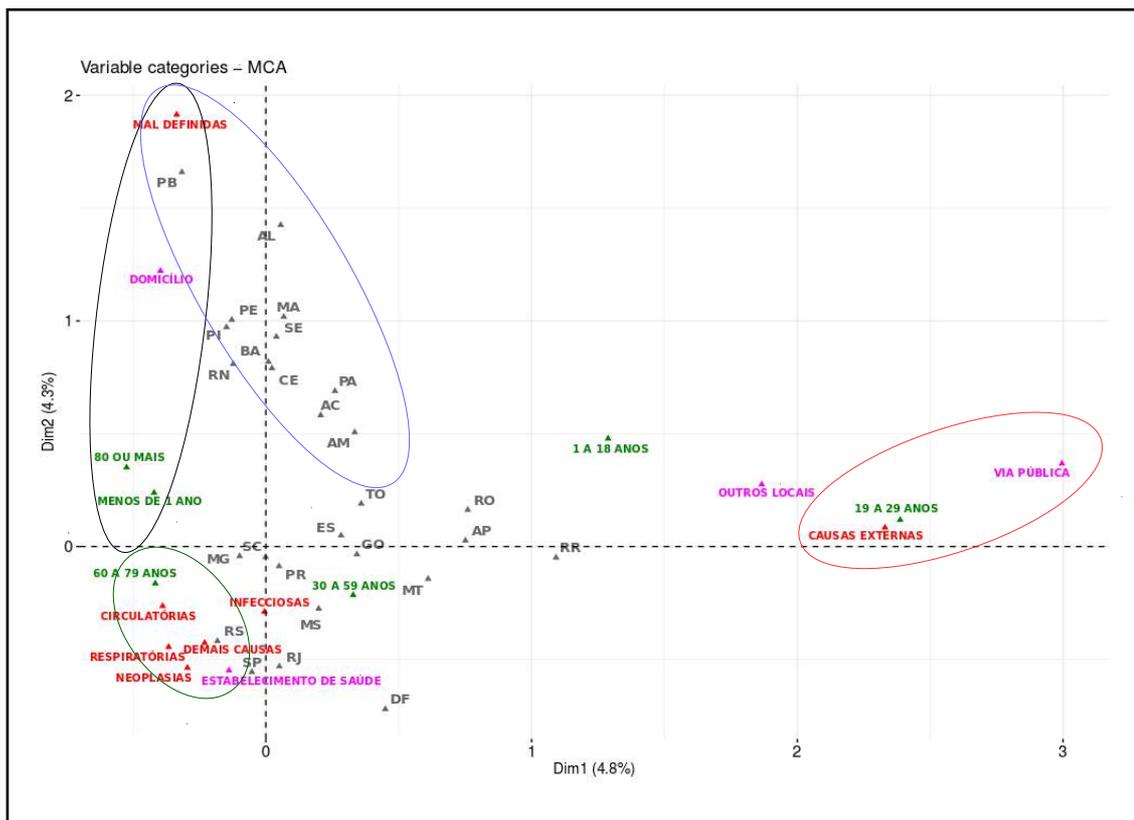


Figura 54 – Análise de Correspondência Múltipla na Abordagem 1 com Grupos de Causa de Óbito, Local de Óbito, Faixa Etária e Unidade da Federação

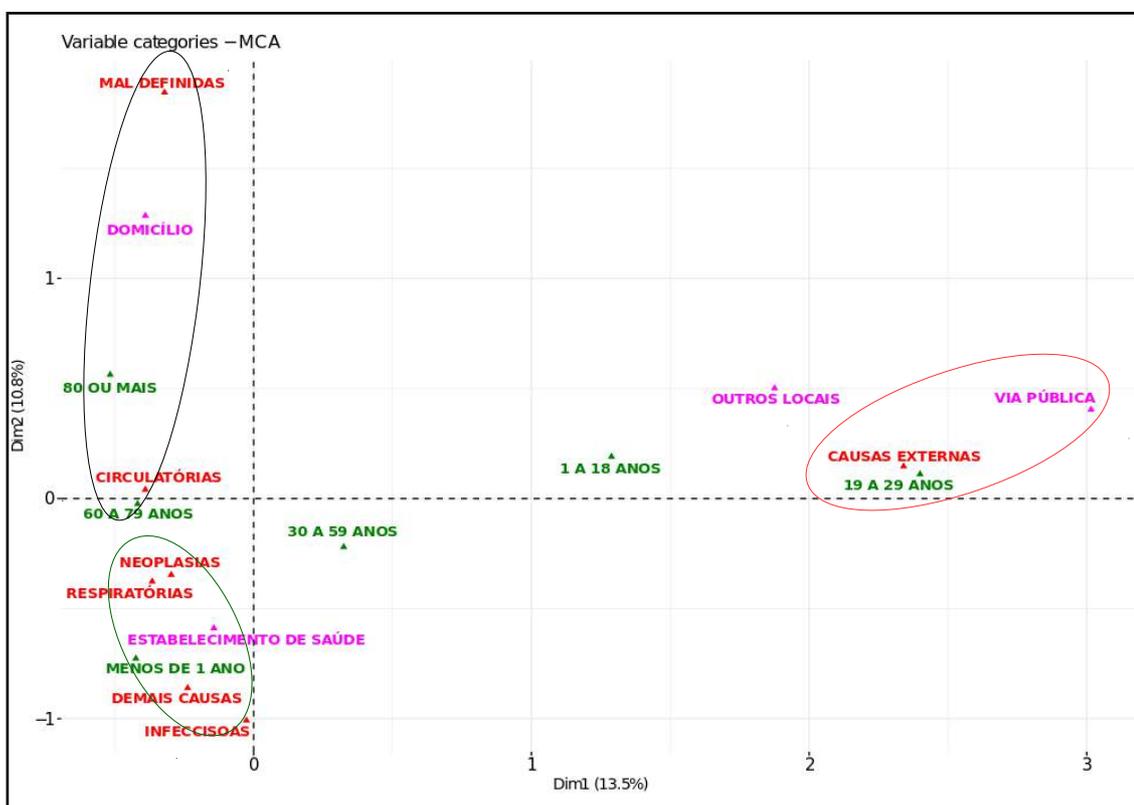


Figura 55 – Análise de Correspondência Múltipla na Abordagem 1 com Grupos de Causa de Óbito, Local de Óbito e Faixa Etária

6.3.2 ABORDAGEM 2 – DOENÇAS INFECCIOSAS E PARASITÁRIAS

A Figura 56 apresenta o resultado do modelo de árvore ajustado na Abordagem 2, na qual foram analisados os óbitos por Doenças Infecciosas e Parasitárias.

A Faixa Etária foi a variável mais importante na classificação da causa específica de óbito dentro deste grupo de causa. Entre os óbitos infantis (menores de 1 ano de idade) as Doenças Infecciosas Intestinais foram a principal causa de mortalidade, sendo responsáveis por 79% dos óbitos. Nas demais Faixas Etárias o Ano de Óbito e a Unidade da Federação também figuram como determinantes para a classificação.

Nas Faixas Etárias de 1 a 18 anos e de 60 ou mais anos no Distrito Federal, Goiás, Minas Gerais, Rondônia e Tocantins, as Doenças Devido a Protozoários foram a causa de óbito mais frequente (42%), enquanto nas demais Unidades da Federação sobressaíram as Doenças Bacterianas.

Nos óbitos na Faixa Etária de 19 a 59 anos, no período de 1996 a 2013, as Doenças Virais apareceram como principal causa de óbito (55%), enquanto no período de 1979 a 1995 no Amazonas, Alagoas, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte e Sergipe, Rio de Janeiro, Espírito Santo, Rio Grande do Sul e Santa Catarina a Tuberculose figurou como causa de óbito mais frequente (42%). Nas demais Unidades da Federação, majoritariamente componentes da região Norte e da região Centro-Oeste, as Doenças Devido a Protozoários se destacaram (43%).

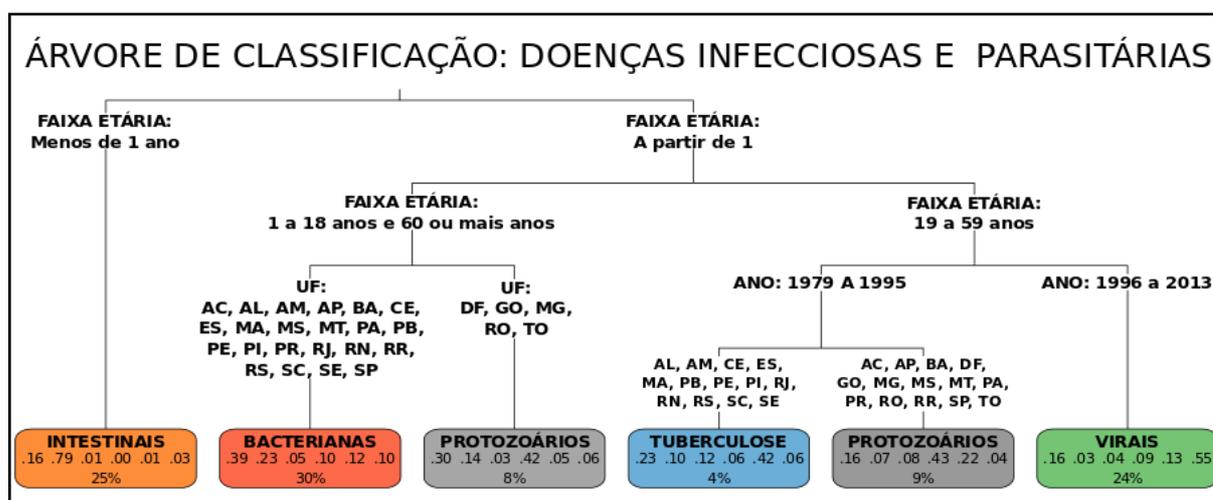


Figura 56 – Árvore de Classificação obtida pelo modelo *CART* na Abordagem 2: Doenças Infecciosas e Parasitárias

As medidas de qualidade de predição do modelo se encontram na Tabela 3. A Acurácia foi 0,532 enquanto o Índice Kappa foi 0,404 e o modelo mostrou elevada Sensibilidade (0,729), Especificidade (0,863) e Valor Preditivo Negativo (0,935) para Doenças Virais. A maior especificidade foi para Tuberculose (0,972) e a maior Taxa de Detecção foi para Doenças Infecciosas Intestinais (0,194).

Tabela 3 – Medidas de Qualidade de Predição do Modelo *CART* da Abordagem 2

CAUSAS	MEDIDA DE QUALIDADE					
	Sensibilidade	Especificidade	VPP	VPN	Prevalência	Taxa Detecção
Bacterianas	0,478	0,759	0,392	0,817	0,245	0,117
Intestinais	0,666	0,925	0,785	0,871	0,291	0,194
Demais Causas	0	1	-	0,957	0,042	0
Protozoários	0,551	0,889	0,426	0,930	0,129	0,071
Tuberculose	0,160	0,972	0,417	0,903	0,110	0,014
Virais	0,729	0,863	0,542	0,935	0,181	0,132

Acurácia: 0,532
Kappa: 0,404

O gráfico da Figura 57 mostra o Decréscimo Médio do Índice de Impureza de Gini das variáveis independentes extraídos do modelo *Random Forest*. A Faixa Etária foi a variável independente mais influente seguida do Ano de Óbito e da Unidade da Federação, neste contexto, foi realizada uma Análise de Correspondência Múltipla utilizando a Causa Específica de Óbito junto com estas três variáveis independentes (Figura 58).

Observa-se na Figura 58 que a variância explicada pela Análise de Correspondência foi baixa. O Ano de Óbito aumentou ao longo do eixo horizontal da direita para esquerda. As Doenças Infecciosas Intestinais estão próximas a Faixa Etária correspondente a menores de 1 ano, destacado pela elipse azul à esquerda, enquanto as outras categorias da Causa de Óbito se concentraram à direita do gráfico, espalhadas ao longo do eixo vertical. A elipse em vermelho no quadrante superior direito destaca Doenças devido a Protozoários próxima de Goiás, Distrito Federal, Minas Gerais e Tocantins, e a elipse em verde no quadrante inferior direito mostra o grupamento formado por São Paulo, Rio de Janeiro, Santa Catarina e Tuberculose.

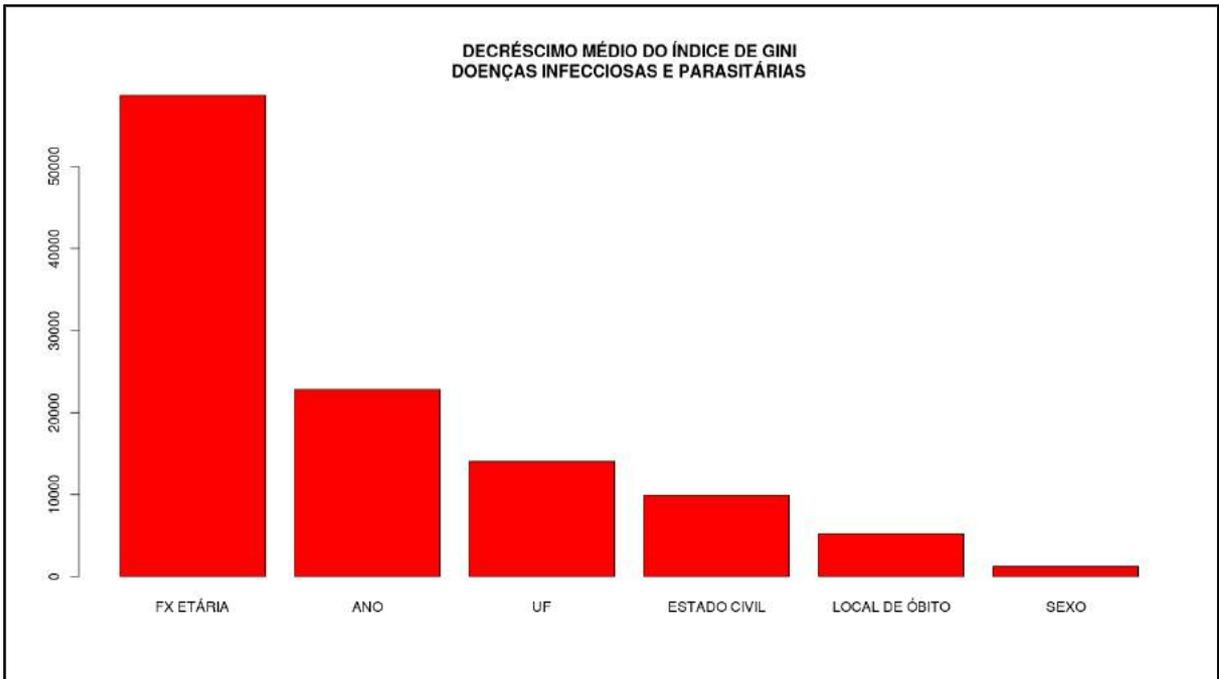


Figura 57 – Decréscimo Médio do Índice de Impureza de Gini das Variáveis Independentes na Abordagem 2 obtidas pelo *Random Forest*

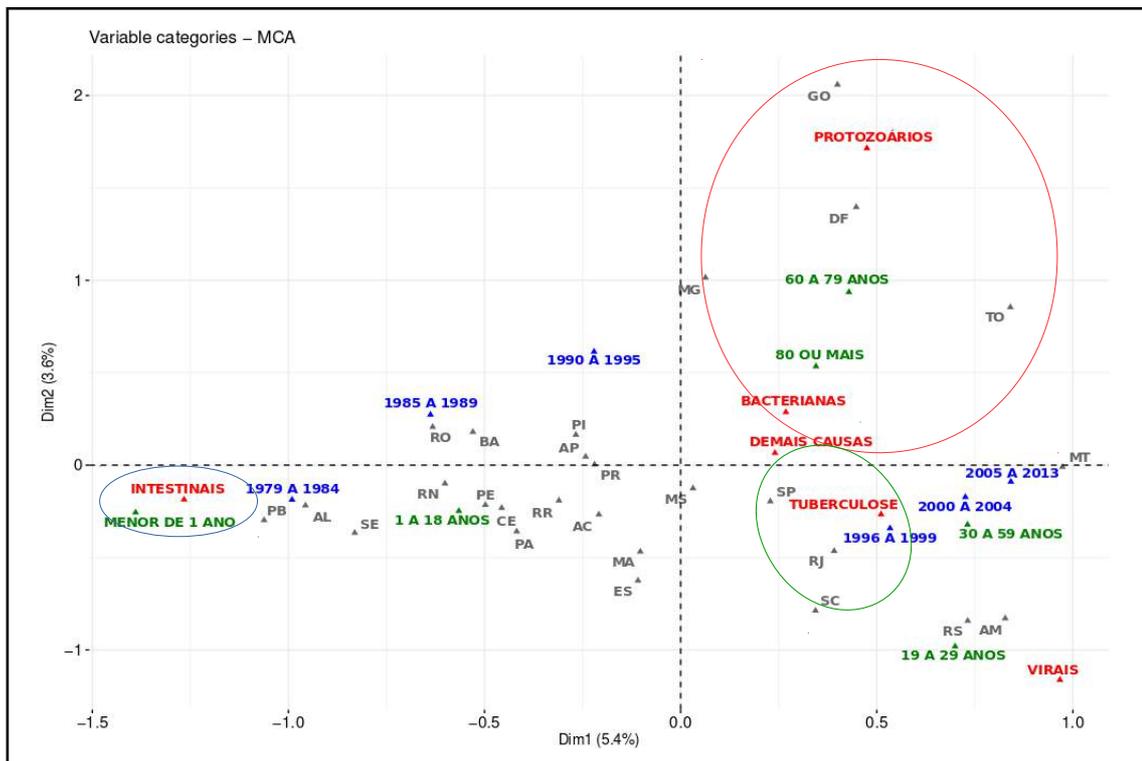


Figura 58 – Análise de Correspondência Múltipla na Abordagem 2 com Causa Específica de Óbito, Faixa Etária, Ano de Óbito e Unidade da Federação

6.3.3 ABORDAGEM 3 – NEOPLASIAS

O modelo de árvore ajustado na Abordagem 3 não apresentou ramificações, classificando todos os óbitos como Demais Causas de Óbito. Como a categoria Demais Causas de Óbito, neste caso as outras Neoplasias que não as de Pulmão e as de Estômago, sobressaíram como mais frequente neste grupo de doenças (77,4%), a Acurácia do Modelo foi elevada (77,4%) e a Sensibilidade e Taxa de Detecção para Demais Causas foi de 100%, tendo em vista que todos os óbitos desta categoria foram classificadas corretamente.

Na Figura 59 se encontra o Decréscimo Médio do Índice de Impureza de Gini das variáveis independentes na Abordagem 3 obtidas através do modelo *Random Forest*. As variáveis mais importantes foram Sexo, Faixa Etária e do Ano de Óbito.

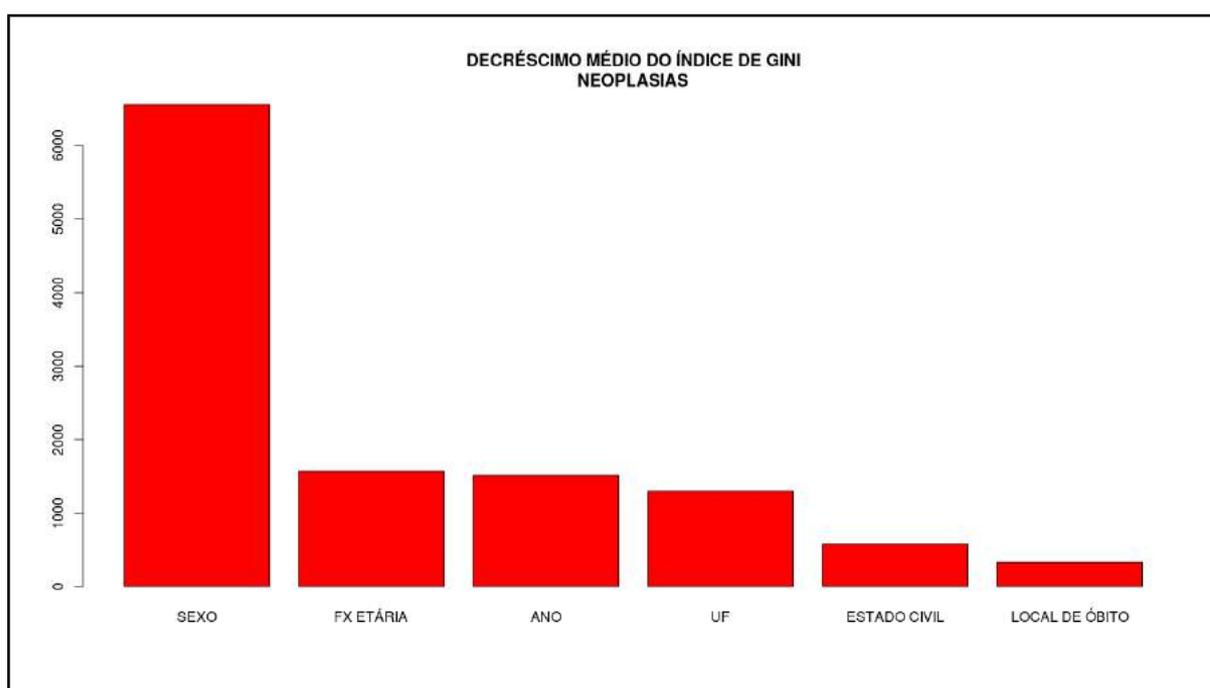


Figura 59 – Decréscimo Médio do Índice de Impureza de Gini das Variáveis Independentes na Abordagem 3 obtidas pelo *Random Forest*

Os resultados da Análise de Correspondência Múltipla com a Causa Específica de Óbito, Sexo, Faixa Etária e Ano se encontram na Figura 60. Destaca-se a proximidade das Demais Causas com sexo Feminino e a Faixa Etária de 30 a 59 anos, na elipse em azul, e a proximidade entre o sexo Masculino e a Faixa Etária de 60 a 79 anos de idade, na elipse em vermelho.

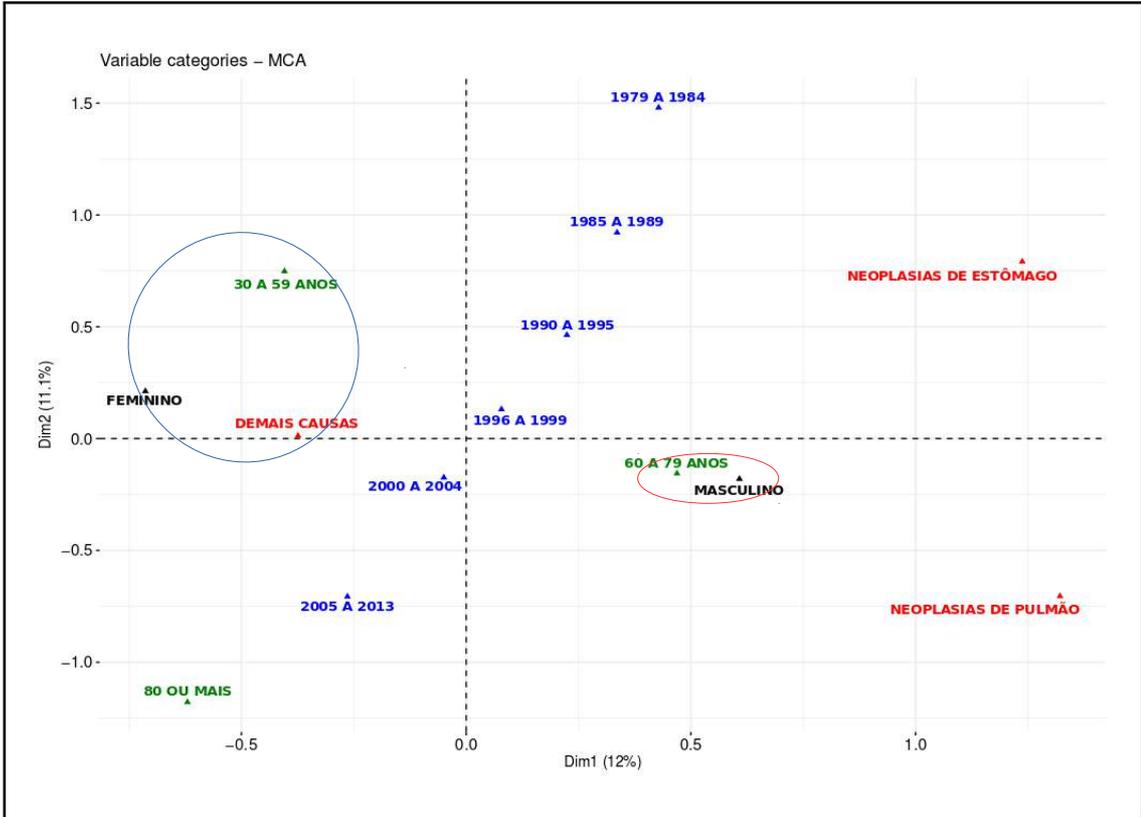


Figura 60 – Análise de Correspondência Múltipla na Abordagem 3 com Causa Específica de Óbito, Sexo, Faixa Etária e Ano de Óbito

6.3.4 ABORDAGEM 4 – DOENÇAS DO APARELHO CIRCULATÓRIO

Os resultados do modelo de árvore com os dados da Abordagem 4 se encontram na Figura 61. Observa-se que o modelo indica o Local de Ocorrência do Óbito como determinante para a classificação da causa específica de óbito, de maneira que em Estabelecimentos de Saúde, onde ocorreram 70% dos óbitos analisados, as Doenças Cerebrovasculares foram mais frequentes (37%). Já os óbitos nos demais Locais de Ocorrência do Óbito 35% dos óbitos foram por Demais Causas Circulatórias.

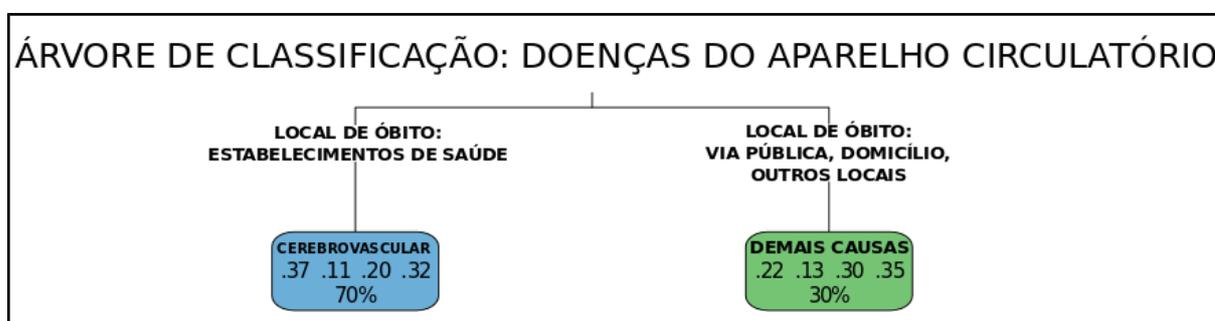


Figura 61 – Árvore de Classificação obtida pelo modelo *CART* na Abordagem 4: Doenças do Aparelho Circulatório

As medidas de qualidade de predição do modelo se encontram na Tabela 3. O modelo apresentou Acurácia (0,360) e Índice Kappa (0,051) baixos e se mostrou Sensível para as Doenças Cerebrovasculares (0,796) e Específico para as Demais Causas Circulatórias (0,310). A Taxa de Detecção destas duas Causas Específicas foi 0,258 e 0,101, respectivamente.

Tabela 4 – Medidas de Qualidade de Predição do Modelo *CART* da Abordagem 4

CAUSAS	MEDIDA DE QUALIDADE					
	Sensibilidade	Especificidade	VPP	VPN	Prevalência	Taxa Detecção
Cerebrovascular	0,796	0,340	0,367	0,776	0,324	0,258
Código Lixo	0	1	-	0,884	0,118	0
Infarto	0	1	-	0,771	0,228	0
Demais Causas	0,310	0,711	0,344	0,679	0,327	0,101
Acurácia: 0,360						
Kappa: 0,051						

A Figura 62 mostra o Decréscimo Médio do Índice de Impureza de Gini das variáveis independentes obtidas através do modelo *Random Forest*. O Local de Ocorrência do Óbito foi a variável mais influente, seguida da Faixa Etária e da Unidade da Federação.

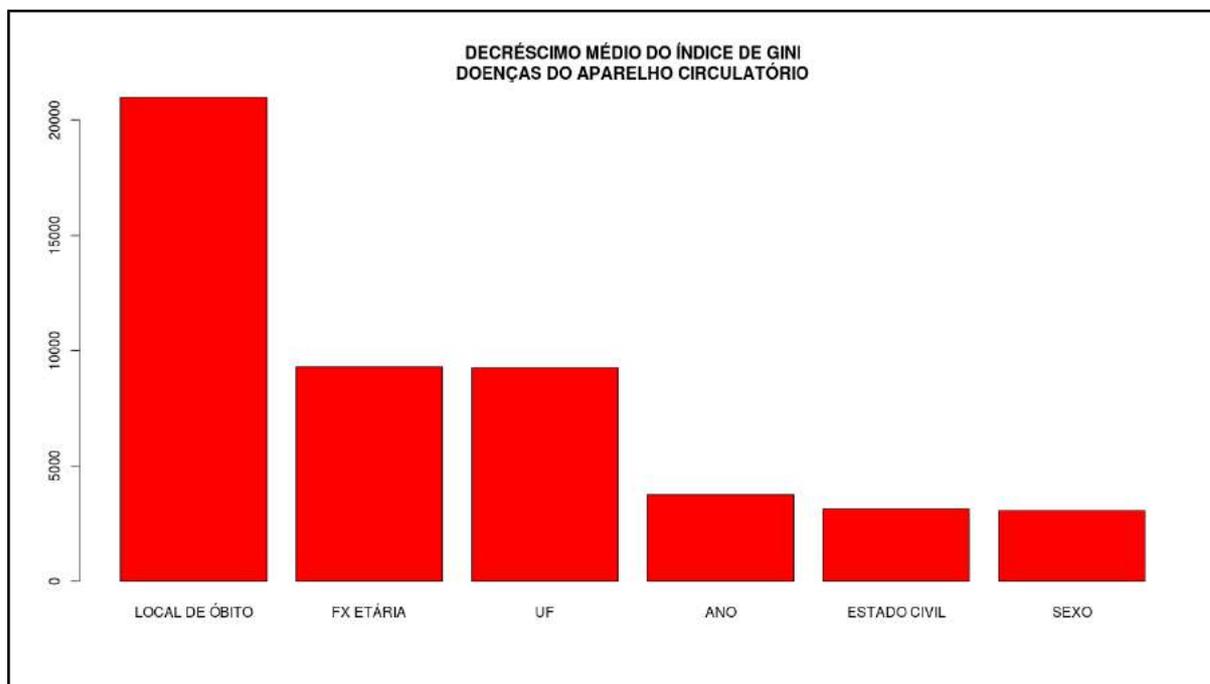


Figura 62 – Decréscimo Médio do Índice de Impureza de Gini das Variáveis Independentes na Abordagem 4 obtidas pelo modelo *Random Forest*

Os resultados da Análise de Correspondência Múltipla utilizando a Causa Específica de Óbito e as três variáveis mais influentes se encontra na Figura 63. Observa-se que a variância explicada pelas variáveis na Análise de Correspondência Múltipla foi pequena. No quadrante inferior esquerdo o Estabelecimento de Saúde e Doenças Cerebrovasculares estão próximas, circundadas pelas Unidades da Federação da região Sudeste, além de Acre, Amazonas, Distrito Federal e Goiás e ainda da Faixa Etária de 60 a 79 anos, destacados pela elipse em azul.

No quadrante inferior direito evidencia-se pela elipse em vermelho um grupamento formado por Código Lixo, a Faixa Etária de 80 ou mais anos, Bahia, Sergipe e Paraíba. Na área central destaca-se pela elipse em verde as Demais Causas Circulatórias próxima das Unidades da Federação da região Sul, Tocantins, Pará, Amapá e Roraima da região Norte, além de Pernambuco e Mato Grosso do Sul. As categorias Infarto, Outros Locais de ocorrência de óbito e Via pública estão isoladas no quadrante superior esquerdo.

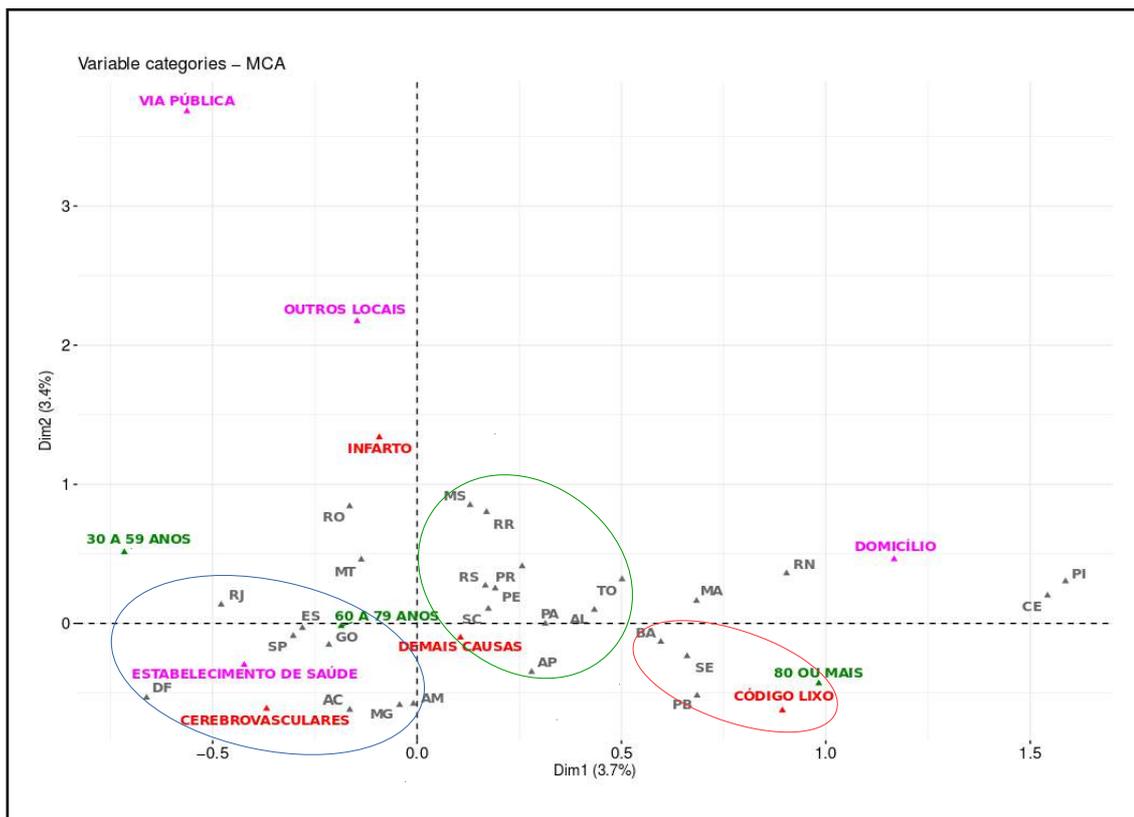


Figura 63 – Análise de Correspondência Múltipla na Abordagem 4 com Causa Específica de Óbito, Sexo, Faixa Etária e Ano de Óbito

6.3.5 ABORDAGEM 5 – DOENÇAS DO APARELHO RESPIRATÓRIO

A Figura 64 apresenta os resultados do modelo de árvore com os dados da Abordagem 5. A Faixa Etária foi a primeira variável determinante para a classificação da causa específica de óbito, apontando para a Pneumonia como mais frequente (77%) até os 18 anos de idade, estando Ano de Óbito e Unidade da Federação relacionadas às demais categorias da Faixa Etária.

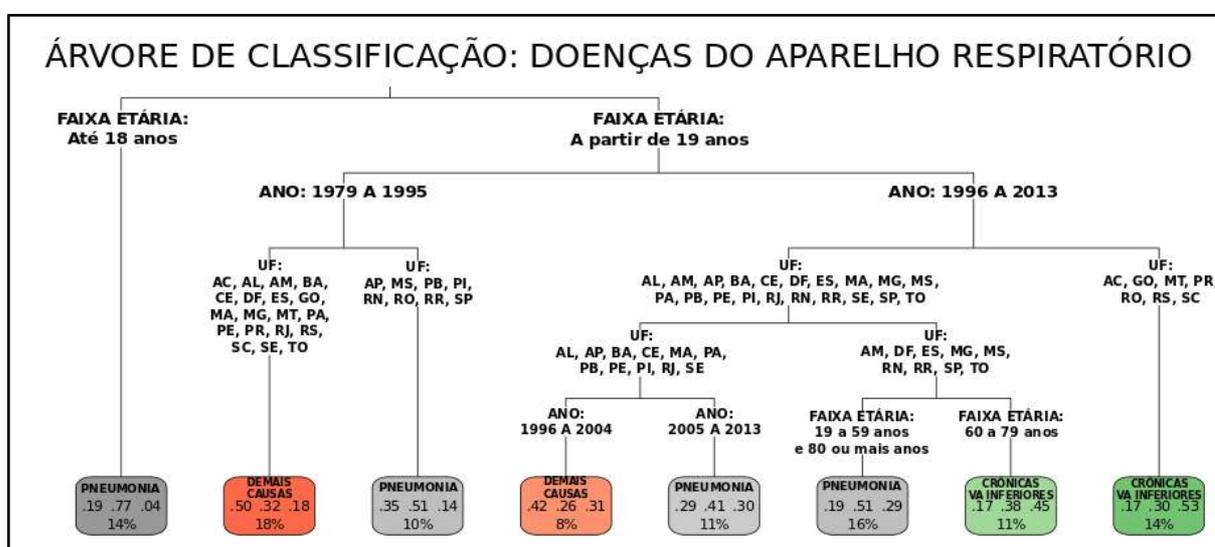


Figura 64 – Árvore de Classificação obtida pelo modelo *CART* na Abordagem 5: Doenças do Aparelho Respiratório

Observa-se que os óbitos na Faixa Etária de 19 ou mais anos de idade no período de 1979 a 1995 no Amapá, Mato Grosso do Sul, Paraíba, Piauí, Rio Grande do Norte, Rondônia, Roraima e São Paulo a Pneumonia foi mais frequente (51%) e nas demais Unidades da Federação sobressaíram as Demais Causas Respiratórias (32%). Já no período de 1996 a 2013 em toda região Sul, Acre e Rondônia na região Norte e Goiás e Mato Grosso sobressaíram as Doenças Crônicas das Vias Aéreas (53%).

Ainda em relação a Faixa Etária de 19 ou mais anos nas Unidades da Federação da região Nordeste, com exceção do Rio Grande do Norte, Alagoas, Amapá e Pará na região Norte e no Rio de Janeiro no período de 1996 a 2004 as Demais Causas Respiratórias se destacaram (42%) enquanto no período de 2005 a 2013 a Pneumonia foi mais frequente (41%). Nos demais estados no período de 1996 a 2013 nas Faixas Etárias de 19 a 59 anos e de 80 ou mais anos as Pneumonias foram responsáveis por 51% dos óbitos e na Faixa Etária de

60 a 79 anos de idade as Doenças Crônicas das Vias Aéreas Inferiores representaram 45% dos óbitos.

A Tabela 4 apresenta as medidas de qualidade de predição do modelo *CART*. O modelo apresentou Acurácia de 52,3% e Índice Kappa de 0,253 e maior Sensibilidade para Pneumonia (0,637) e Especificidade para as Demais Causas Respiratórias (0,817) e Doenças Crônicas das Vias Aéreas Inferiores (0,827). A maior Taxa de Detecção foi para Pneumonia (0,281).

Tabela 5 – Medidas de Qualidade de Predição do Modelo *CART* da Abordagem 5

CAUSAS	MEDIDA DE QUALIDADE					
	Sensibilidade	Especificidade	VPP	VPN	Prevalência	Taxa Detecção
Demais Causas	0,421	0,817	0,478	0,781	0,283	0,195
Pneumonia	0,637	0,603	0,559	0,678	0,441	0,281
Doenças Crônicas VA Inferiores	0,443	0,827	0,493	0,797	0,274	0,121
Acurácia: 0,523						
Kappa: 0,253						

O Gráfico da Figura 65 mostra o Decréscimo Médio do Índice de Impureza de Gini das variáveis da Abordagem 5 segundo o modelo *Random Forest*. A Faixa Etária foi a mais influente, seguida do Ano de Óbito e da Unidade da Federação. Os resultados da Análise de Correspondência Múltipla encontram na Figura 66.

Observa-se na que as Doenças Crônicas das Vias Aéreas Inferiores se encontra no quadrante inferior direito, próxima de Unidades da Federação da região Centro-Oeste, além de Paraná, Espírito Santo, Ceará, Piauí e Amazonas, o período de 2005 a 2013 e a Faixa Etária de 80 ou mais anos de idade, destacados pela elipse em azul (Figura 66).

A Pneumonia ficou no quadrante inferior direito, assim como as Faixas Etárias de menos de 1 ano e de 1 a 18 anos de idade e os anos de 1979 a 1984, como evidencia a elipse em vermelho. Neste quadrante estiveram também Roraima, Rondônia, Pará e Acre da Região Norte, Sergipe, além de São Paulo e Distrito Federal.

No quadrante superior direito estão dispostos as Demais Causas Respiratórias, os anos de 1985 a 1989 e de 1990 a 1995, as faixas etárias de 19 a 29 anos e de 30 a 59 anos além de Rio de Janeiro, Bahia, Alagoas e Amapá, destacadas na elipse em verde. No quadrante superior esquerdo o Rio Grande do Sul e Santa Catarina formaram um grupamento com a Faixa Etária de 60 a 79 anos de idade e os períodos de 1996 a 1999 e de 2000 a 2004.

Nenhuma causa específica foi alocada neste quadrante, mas o grupamento descrito esteve próximo das Doenças Crônicas das Vias Aéreas Inferiores, como mostra a elipse em preto.

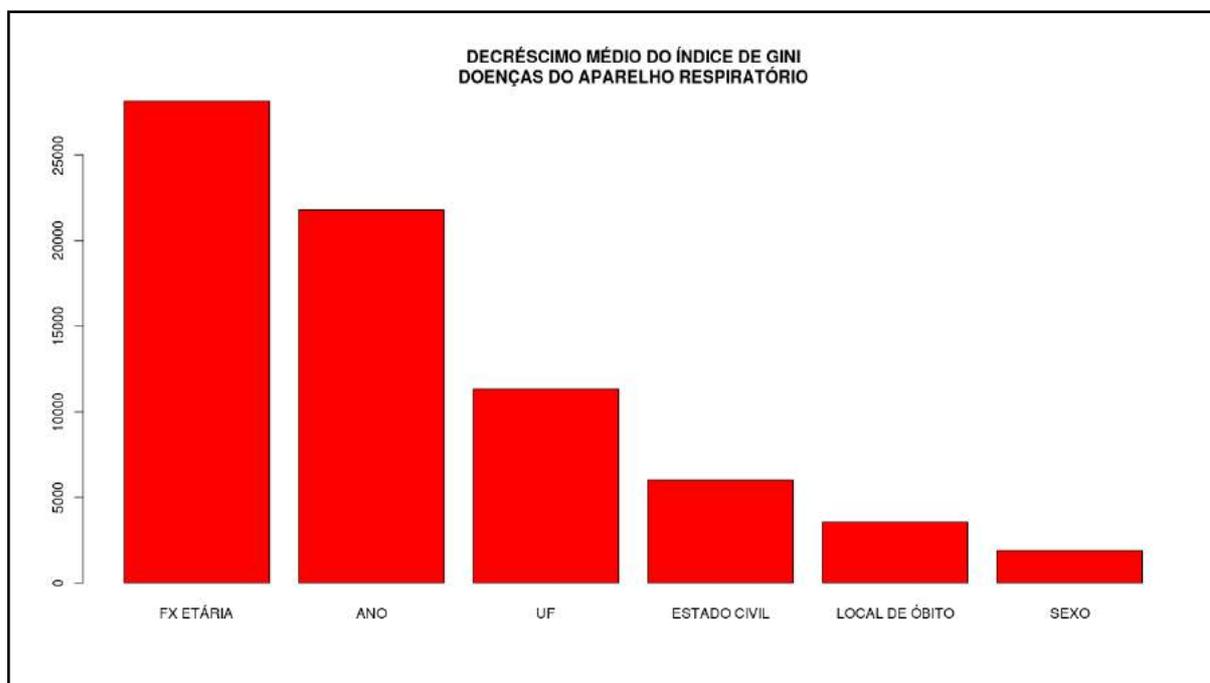


Figura 65 – Decréscimo Médio do Índice de Impureza de Gini das Variáveis Independentes na Abordagem 5 obtidas pelo modelo *Random Forest*

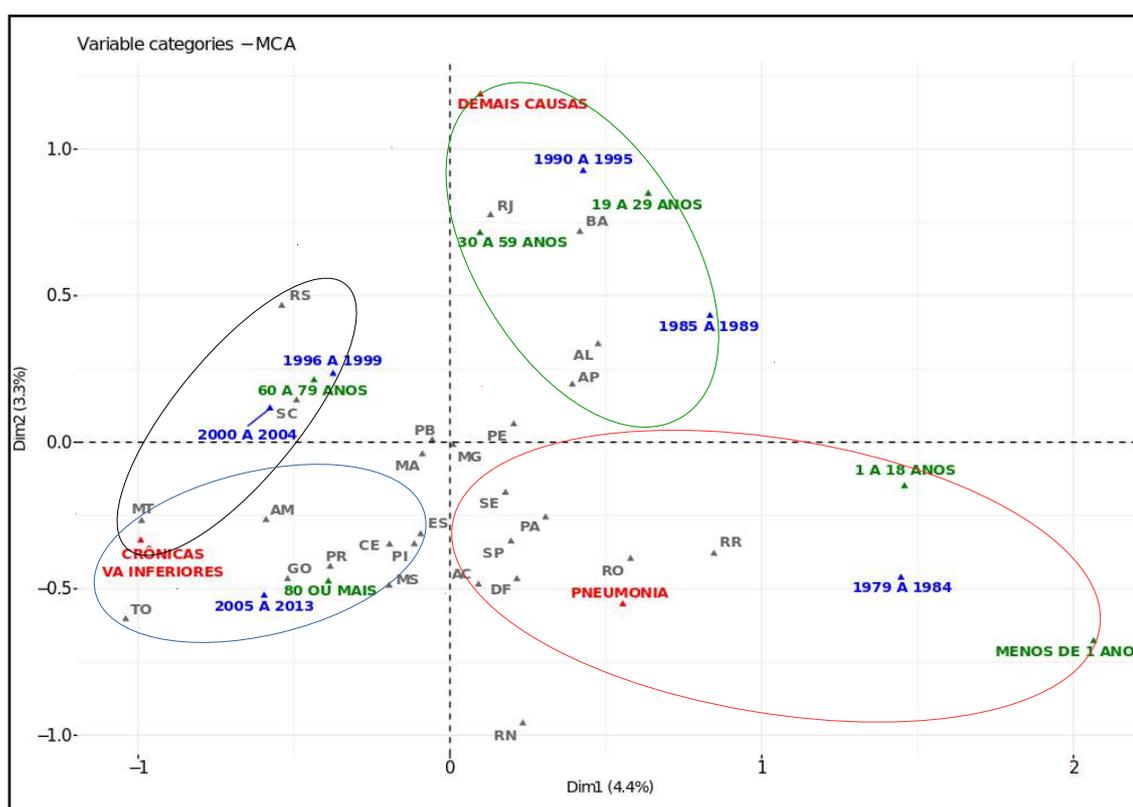


Figura 66 – Análise de Correspondência Múltipla na Abordagem 5 com Faixa Etária, Ano de Óbito e Unidade da Federação

6.3.5 ABORDAGEM 6 – CAUSAS EXTERNAS

A árvore obtida pelo modelo *CART* na Abordagem 6 se encontra na Figura 67. Nos óbitos por Causas Externas em Via Pública nas Unidades da Federação de Alagoas, Amazonas, Bahia, Espírito Santo, Pará, Pernambuco e Rio de Janeiro o Homicídio foi a causa mais frequente (52%), enquanto nas demais Unidades da Federação os Acidentes de Transporte sobressaíram com (51%). Já nos óbitos ocorridos em Domicílios e Outros Locais as Demais Causas Externas foram as mais frequentes (49%).

Entre os óbitos por Causas Externas ocorridos em Estabelecimentos de Saúde na Faixa Etária de 19 a 29 anos de idade os Homicídios foram mais frequentes (42%) enquanto nas Faixas Etárias de 1 a 18 e de 30 a 58 anos os Acidentes de Transporte foram a principal causa (34%).

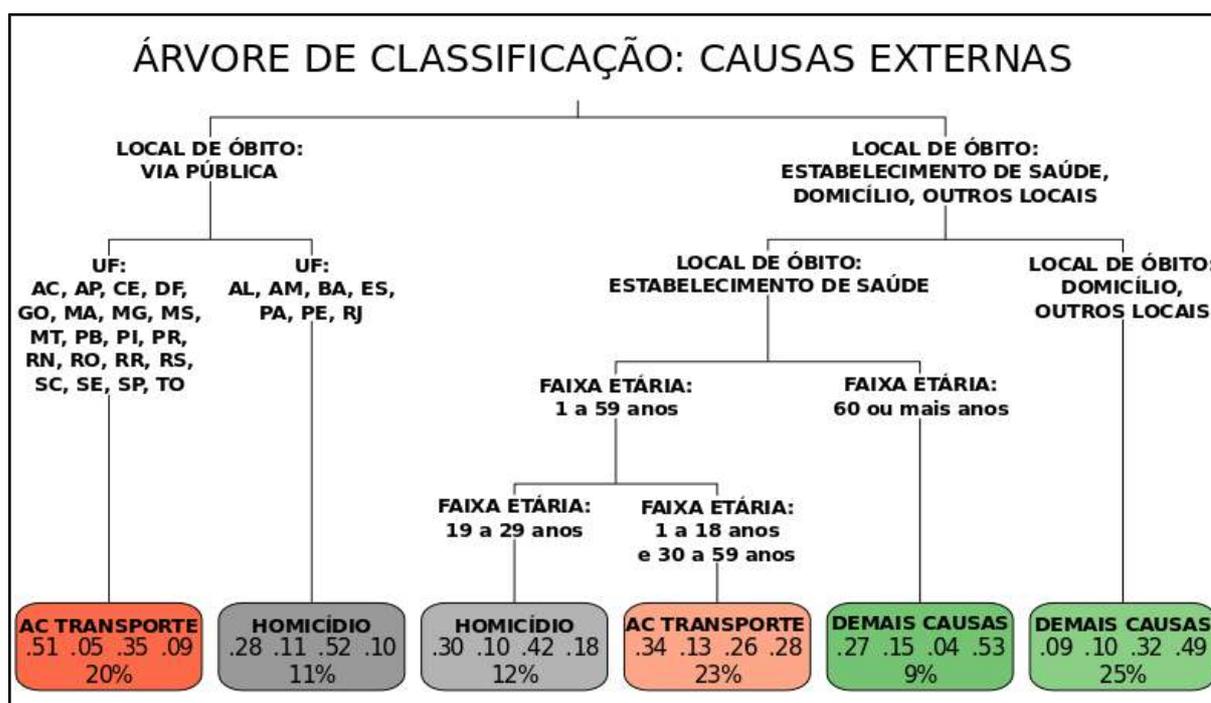


Figura 67 – Árvore de Classificação obtida pelo modelo *CART* na Abordagem 6: Causas Externas

As medidas de qualidade de predição do modelo da Abordagem 6 se encontram na Tabela 5. A Acurácia do modelo de árvore foi 0,456 e o Índice Kappa 0,227. A maior Sensibilidade do modelo foi para a classificação de Acidente de Transporte (0,611) enquanto a maior Especificidade foi para Homicídio (0,812). A Especificidade para as Demais Causas Externas também foi elevada (0,763) e a Taxa de Detecção das três causas classificadas foram

próximas: 0,178 para Acidente de Transporte, 0,110 para Código Lixo e 0,166 para Demais Causas Externas.

Tabela 6 – Medidas de Qualidade de Predição do Modelo *CART* da Abordagem 6

CAUSAS	MEDIDA DE QUALIDADE					
	Sensibilidade	Especificidade	VPP	VPN	Prevalência	Taxa Detecção
Acidente de Transporte	0,611	0,651	0,420	0,801	0,292	0,178
Código Lixo	0	1	-	0,898	0,101	0
Homicídios	0,342	0,812	0,465	0,721	0,323	0,110
Demais Causas	0,591	0,763	0,496	0,826	0,282	0,166
Acurácia: 0,456						
Kappa: 0,227						

O gráfico da Figura 68 mostra o Decréscimo Médio do Índice de Impureza de Gini das variáveis independentes obtidas pelo modelo *Random Forest* onde o Local de Ocorrência do Óbito se destacou como variável mais influente, seguida da Faixa Etária e da Unidade da Federação. Os resultados A Análise de Correspondência Múltipla utilizando a Causa Específica de Óbito, Local de Ocorrência do Óbito, Faixa Etária e Unidade da Federação se encontram na Figura 69.

Observa-se que a Via Pública ficou posicionada à esquerda assim como Homicídio e Acidente de Transporte em um padrão semelhante ao destacado pelo modelo *CART*. A elipse em preto destaca o quadrante superior esquerdo, onde se encontra o Homicídio e a maioria das Unidades da Federação da região Norte, Ceará, Pernambuco, Alagoas e Maranhão da região Nordeste, além de Espírito Santo e Tocantins e a Faixa Etária de 19 a 29 anos. Ainda à esquerda no gráfico, no quadrante inferior estão os Acidentes de Transporte assim como Sergipe, Goiás, Distrito Federal e Rio de Janeiro, destacados em azul (Figura 69).

No quadrante superior direito estão os Outros Locais de Óbito, os Domicílios e as Demais Causas Externas, mais dispersos. Também de maneira dispersa estão Estabelecimento de Saúde, Código Lixo e a Faixa Etária de 60 ou mais anos de idade.

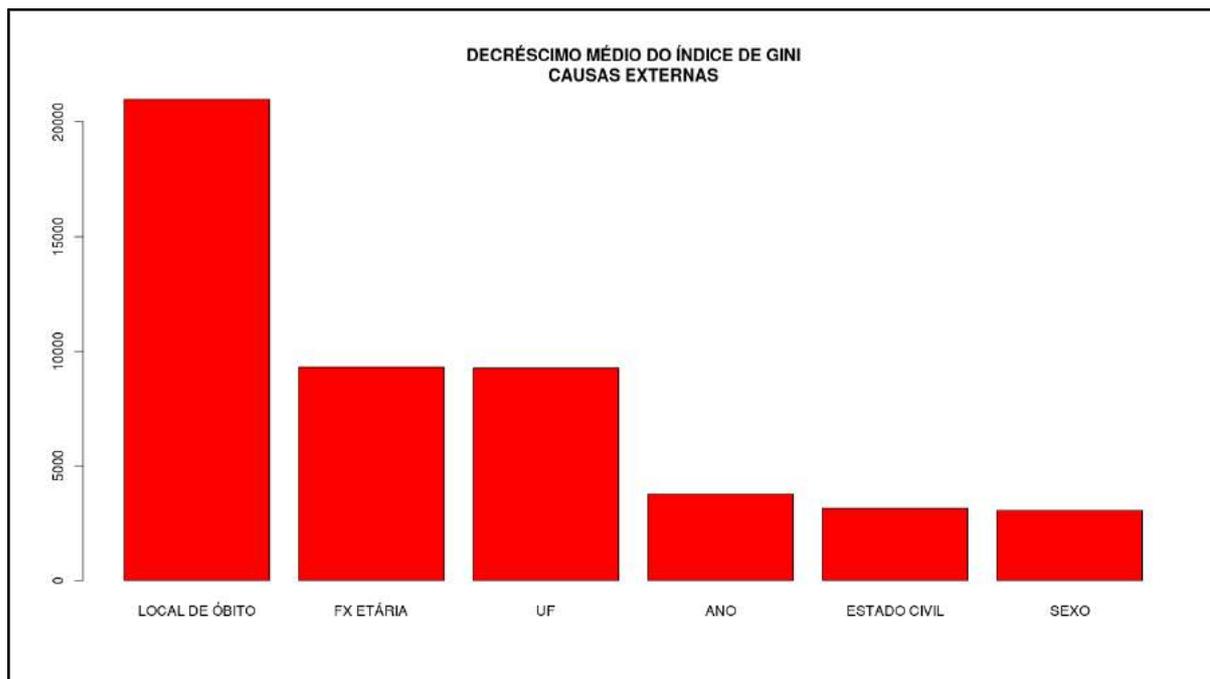


Figura 68 – Decréscimo Médio do Índice de Impureza de Gini das Variáveis Independentes na Abordagem 6 obtidas pelo modelo *Random Forest*

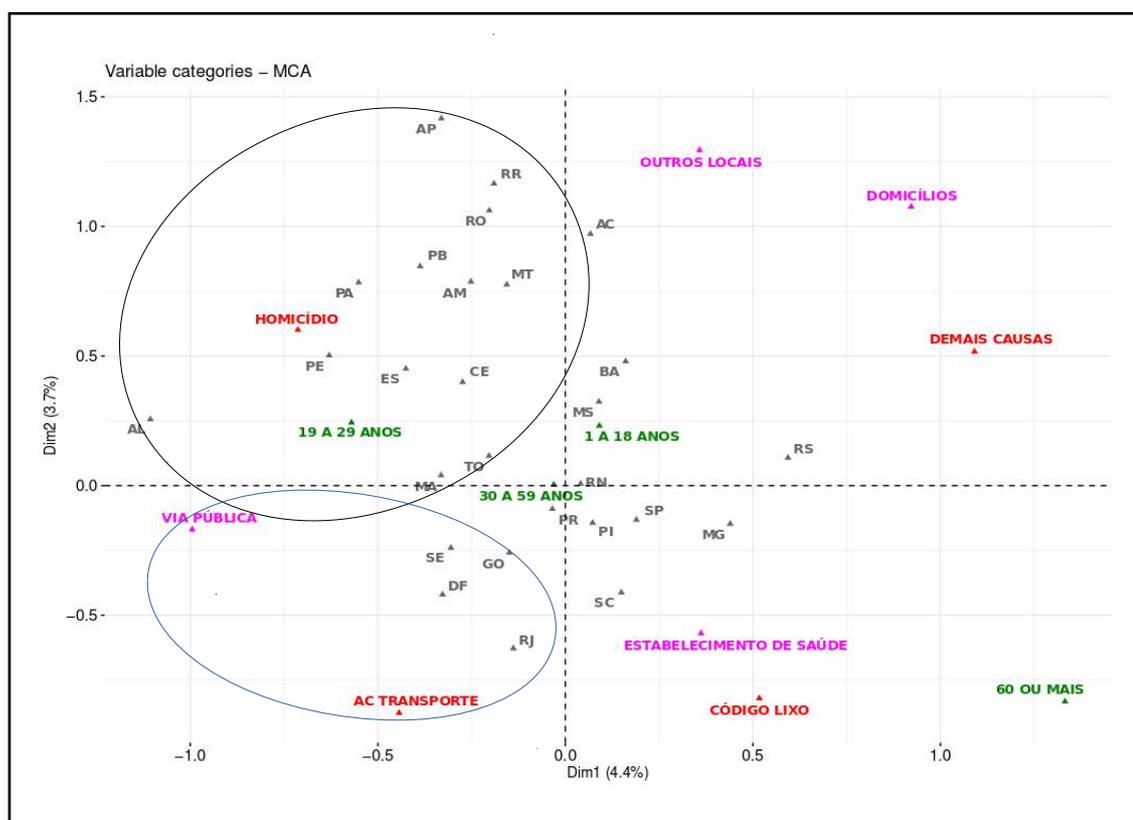


Figura 69 – Análise de Correspondência Múltipla na Abordagem 6 com Causa Específica de Óbito, Faixa Etária, Local de ocorrência do Óbito e Unidade da Federação

7 DISCUSSÃO

As mudanças do perfil demográfico e epidemiológico da população brasileira nas últimas décadas refletem condições associadas simultaneamente ao atraso e ao desenvolvimento socioeconômico do país e os padrões e tendências da mortalidade identificados nesta Tese confirmam o que já vem sendo identificado por diversos estudos.

Estas transições ocorreram de maneira desigual no país em um processo caracterizado por polarização epidemiológica, geográfica e social, permeado por padrões e tendências regionais distintos com impacto diverso entre os grupos populacionais de cada região como reflexos do desnível social, econômico e demográfico (ARAÚJO, 2012) – sinalizando “Vários Brasis”.

Os resultados do presente estudo demarcam as particularidades históricas e tendências regionais de continuidade da transição da mortalidade iniciada no país na década de 1940 (PRATA, 1992). As regiões Sul e Sudeste, que detém os melhores indicadores socioeconômicos, concentram óbitos por doenças crônico-degenerativas e em faixas etárias mais elevadas; já nas regiões Norte e Nordeste os indicadores socioeconômicos melhoraram no período e a transição da mortalidade mostrou-se acelerada, impactando particularmente os mais jovens, com elevação constante de doenças crônico-degenerativas e causas externas. Nestas últimas regiões o declínio de doenças infecciosas e parasitárias, mostrou-se ainda maior, comparado às demais.

A queda abrupta da mortalidade infantil, resultado de ações médico sanitárias de grande impacto na redução da mortalidade por doenças infecciosas e parasitárias, junto à redução da fecundidade e aumento na expectativa de vida, propiciou um aumento gradativo da população adulta e idosa, que se tornou mais expressiva do que a de jovens (CHAIMOWICZ, 1997; ARAÚJO, 2012; VASCONCELOS; GOMES, 2012).

Concomitantemente, mudanças socioeconômicas e culturais que permearam a transição demográfica, como a industrialização e a urbanização, popularizaram comportamentos não saudáveis como sedentarismo, tabagismo, consumo excessivo de álcool e o aumento do consumo de alimentos processados de elevado valor energético e baixo valor nutricional em detrimento de frutas, verduras e legumes. Estas mudanças demarcam o processo de transição nutricional, caracterizado pelo desaparecimento de quadros graves de desnutrição aguda, aumento da prevalência de sobrepeso e obesidade, diabetes e hipertensão arterial, etapa que favorece a transição epidemiológica configurada por elevada

morbimortalidade por doenças crônicas e não transmissíveis (BATISTA FILHO et al., 2008; COUTINHO; GENTIL; BERTOLIN, 2008; SCHMIDT et al., 2011).

As mudanças identificadas no Norte e Nordeste no período analisado foram semelhantes àquelas constatadas nas regiões Sul e Sudeste entre as décadas de 1960 e 1980. No Centro-Oeste o padrão foi semelhante ao das regiões Norte e Nordeste no início do período, aproximando-se do Sul e Sudeste com o passar dos anos. Aspectos refletem mais do que apenas características individuais ou de distribuição heterogênea da população e dos serviços de saúde no território. Indicam desigualdades entre atributos do lugar e no uso e qualidade dos serviços de saúde, passíveis de serem modificados por políticas sociais (TRAVASSOS; WILLIAMS, 2004; BLAS; KURUP; WORLD HEALTH ORGANIZATION, 2010; PAIM et al., 2011; MACINKO; LIMA-COSTA, 2012). Se por um lado as doenças crônicas e as causas externas definem um padrão de “modernidade”, já que representam causas de adoecimento e morte intimamente relacionadas à urbanização, industrialização e expansão capitalista, a persistência das causas infecciosas marca padrão relacionado à pobreza e ao subdesenvolvimento, dado ao seu caráter de evitabilidade e relação com as condições de saneamento e adensamento populacional. No Brasil, observa-se a simultaneidade destes padrões, resultando em uma combinação heterogênea, e por vezes perversa, já que ambos os conjuntos de enfermidades atingem particularmente a parcela da população em maior desvantagem social (PAIM et al., 2011; SCHMIDT et al., 2011; ARAÚJO, 2012).

A idade de óbito segundo as causas analisadas foi mais elevada no Sul e Sudeste e mais baixas no Norte e Nordeste em todo o período, padrão condizente com a expectativa de vida e com os anos de vida perdidos devido a óbito prematuro no país, possivelmente uma representação das desigualdades socioeconômicas e no acesso a serviços de saúde entre as regiões (SCHRAMM et al., 2004; LEITE et al., 2015; SZWARCOWALD et al., 2016).

A redução da mortalidade por doenças infecciosas e parasitárias faz parte da queda na mortalidade geral das fases iniciais da transição epidemiológica, e impactam principalmente as faixas etárias mais jovens, estando intimamente relacionada com a queda mortalidade infantil no país. A tendência, destacada neste e em outros estudos, mostra que a mortalidade proporcional por doenças infecciosas diminuiu de 50% para 5% entre 1930 e 2007 (BARRETO et al., 2015), chegando a quase 4% em 2013, enquanto paralelamente a mortalidade infantil decaiu de mais de 200 óbitos por 1.000 nascidos vivos em 1940 (YUNES; RONCHEZEL, 1974) para menos de 20 óbitos por 1.000 nascidos vivos em 2004 (FISCHER et al., 2007). Segundo Chaimowicz (1997) esta redução no Brasil foi inicialmente

resultado de políticas urbanas de saúde pública, campanhas sanitárias do início do século XX e do avanço da atenção à saúde na rede pública.

O Programa Nacional de Imunização é um exemplo de medida de grande impacto e êxito na redução da mortalidade por doenças infecciosas e da mortalidade infantil, atingindo altos índices de cobertura, mesmo em estratos sociais menos desenvolvidos, e contribuindo de maneira preponderante na erradicação e eliminação de doenças como poliomielite, sarampo, rubéola e tétano neonatal (QUEIROZ et al., 2013; DOMINGUES; TEIXEIRA, 2013; BARRETO, 2015).

A coordenação e a estruturação de ações do sistema de saneamento e a expansão da assistência médica nas décadas seguintes também resultaram na queda das taxas de mortalidade por doenças infecciosas no país. Políticas públicas de saneamento, inicialmente de caráter higienista, representaram resposta a situações endêmicas no final do século XIX, evoluindo com aumento da cobertura do abastecimento de água e do sistema de esgoto e melhorias no tratamento e controle da qualidade da água (SOARES; BERNARDES; NETTO, 2002).

Estudos realizados na década de 1980 já apontavam para menor mortalidade infantil em áreas de maior cobertura de abastecimento de água e esgotamento sanitário e de maior renda; adicionalmente, identificavam gradientes na esperança de vida ao nascer entre os diferentes estratos de rendimento considerando-se áreas com infraestrutura de saneamento básico semelhante (VETTER; SIMÕES, 1981; MONTEIRO, 1982). Tal panorama indica que os avanços sanitários, responsáveis pela diminuição abrupta da mortalidade por doenças infecciosas, e conseqüentemente da mortalidade infantil, não foram suficientes para o controle objetivado. A persistência de iniquidades sociais, refletidas em diferentes níveis de condição de vida, interfere na qualidade, acesso e uso dos serviços de saúde, na eficácia das intervenções e na disponibilidade de meios diagnósticos e de tratamento. Desta forma, as grandes desigualdades regionais que caracterizam o Brasil explicam em parte os impasses e a dificuldade do controle de agravos infecciosos.

Os óbitos por doenças infecciosas intestinais, causa que compõe o grupo de doenças redutíveis em menores 5 anos por ações adequadas de promoção e atenção à saúde (MALTA et al., 2007a, 2010), se destacaram como os mais frequentes entre as causas infecciosas, representando um terço daqueles do capítulo, acompanhando sua tendência. A persistência do padrão de mortalidade por esta causa, principalmente entre crianças, está associada ao estado nutricional, hábitos alimentares, saneamento e deficiência de cuidados primários que apontam para um contexto socioeconômico precário (SZWARCOWALD et al., 1997; PAES; SILVA,

1999), fato que justifica sua correlação negativa observada com os indicadores de renda e esgotamento em todo o período. Neste cenário, a queda generalizada na mortalidade por doenças infecciosas intestinais foi condizente com o avanço de indicadores socioeconômicos no país, ainda que com taxas ligeiramente mais elevadas nas regiões Norte e Nordeste ao final do período, onde se concentraram os piores indicadores. Como visto em diversos estudos, constata-se redução constante das taxas de mortalidade infantil vinculada a esta causa, associando-se esta redução aos avanços da cobertura de abastecimento de água e tratamento de esgoto, e à adoção de terapia de reidratação oral (GUIMARÃES, et al., 2001; TEIXEIRA, et al., 2002).

No presente estudo não foi calculada a taxa de mortalidade infantil por este grupo de causa, o que dificulta a comparação com a magnitude verificada em outros estudos e países. Contudo, a taxa de mortalidade infantil em menores de 5 anos por doenças infecciosas intestinais, em 2015 no Brasil, foi de 0,2 óbitos para cada 1.000 nascidos vivos (SIM, 2017c; SINASC, 2017), valor superior a países da Europa Ocidental, Estados Unidos e alguns da América Latina, como Chile e Uruguai, e equiparado a países como Albânia, Ucrânia, Oman, Líbano e Arábia Saudita (WHO, 2016), um cenário longe do ideal mas que representa um grande avanço comparado ao início dos anos 1990 (SZWARCOWALD et al., 1997).

Elevadas taxas de mortalidade por doenças infecciosas intestinais e óbitos precoces marcam o início do período analisado, mas a partir de meados dos anos 1980, com o avanço da mortalidade por doenças bacterianas, virais e tuberculose, observa-se óbitos em idades mais elevadas. Padrão equivalente foi observado na Espanha, com aumento recente na mortalidade por doenças infecciosas na faixa etária de 65 anos ou mais anos e tendência de aumento das taxas de mortalidade por septicemia, infecções renais e hepatites virais entre 1980 e 2011 (LÓPEZ-CUADRADO et al., 2014). Nos Estados Unidos o aumento das taxas de mortalidade por doenças infecciosas nos últimos anos tem sido associado ao aumento da mortalidade por HIV/Aids (ARMSTRONG et al., 1999; HANSEN et al., 2016). No Brasil, Paes & Silva (PAES; SILVA, 1999) descreveram padrão etário semelhante para os óbitos por doenças infecciosas intestinais, tuberculose e septicemia entre 1980 e 1990. Não foram encontrados estudos abordando o padrão etário de óbitos por doenças infecciosas específicas utilizando estatísticas de mortalidade mais atualizadas no país.

Os padrões destacados pelo modelo de árvore foram condizentes com as tendências de doenças infecciosas específicas no país. A elevada mortalidade por doenças causadas por protozoários em unidades da federação do Norte, Nordeste e Centro-Oeste, além de São Paulo, Minas Gerais e Paraná pode estar associada à elevada incidência de Leishmaniose

Visceral (GONTIJO & MELO, 2004; MAIA-ELKHOURY *et al.*, 2008) e Doença de Chagas (MARTINS-MELO *et al.*, 2014) nestes locais.

Estes dois agravos eram considerados endemias rurais no país até os anos 1980, contudo o intenso processo de migração, responsável por o deslocamento de um grande número de indivíduos infectados para as cidades, e a urbanização desordenada, que promoveu o desmatamento e a ocupação de áreas em condições sanitárias precárias, propiciando a interação entre vetor, reservatório e seres humanos, favoreceram a ascensão dos mesmos também no ambiente urbano (COSTA,; TAPETY; WERNECK, 2007; CERBINO NETO; WERNECK; COSTA, 2009; MONCAYO; SILVEIRA, 2009; LIMA *et al.*, 2012).

As principais formas de transmissão da Doença de Chagas em áreas endêmicas no país eram a vetorial e a transfusão sanguínea, mas com o aumento considerável da cobertura da triagem de doadores de sangue e a eliminação do principal vetor, a transmissão congênita e oral foram colocadas em evidência (GONTIJO *et al.*, 2009; MORAES-SOUZA; FERREIRA-SILVA, 2011; OSTERMAYER *et al.*, 2011). Já a leishmaniose visceral apresenta maior incidência em áreas periféricas, onde a área urbana e de mata se misturam, e em regiões com maior concentração de reservatórios (cachorros); um forte indicativo de que a interação vetor, reservatório e hospedeiro é preponderante para sua disseminação (COSTA; TAPETY; WERNECK, 2007). O controle de ambas é um desafio para a Saúde Pública no país, uma vez que não existem vacinas e as estratégias de combate envolvem monitoramento constante para diagnóstico e tratamento de casos novos, controle de vetor e de reservatórios e melhoria das condições de saneamento e moradia que demandam elevado grau de investimento (COSTA; TAPETY; WERNECK, 2007; MASSAD, 2008).

O destaque das doenças virais a partir de 1996 pode ser explicado pelo avanço recente das taxas de mortalidade por doenças como Dengue (ARAÚJO *et al.*, 2017) e HIV/Aids (GUIMARÃES *et al.*, 2017; MALTA *et al.*, 2017). A concentração de óbitos por tuberculose em unidades da federação do Sudeste, Sul, Norte e Nordeste entre 1979 e 1995 vão ao encontro dos achados de Biereenbach *et al.* (2007), e são compatíveis com distribuição de indicadores socioeconômicos e de coinfeção desta doença com HIV no país (LIMA, *et al.*, 2016; CECCON *et al.*, 2017).

A mortalidade proporcional por Neoplasias aumentou em todas as regiões do país, tendência do processo de transição epidemiológica, mostrando-se mais elevada no Sul e Sudeste, com aumento acelerado no Norte e Nordeste. As taxas de mortalidade apresentaram comportamento desigual entre estados, aumentando de maneira geral, com exceção do Rio de Janeiro, São Paulo, Rio Grande do Sul, Distrito Federal e Amazonas, onde foram mais

elevadas no período mas apresentaram tendência de queda, padrão semelhante ao observado por Guerra et al. (2017) ao analisar a mortalidade por neoplasias em geral e específicas nos estados brasileiros.

Segundo estimativas da Organização Mundial de Saúde as neoplasias causaram mais óbitos do que as doenças coronarianas e acidente vascular encefálico em 2011, ocorrendo mais de 14 milhões de casos novos e 8 milhões de óbitos em 2012 (FERLAY et al., 2015). A identificação de padrões e tendências na mortalidade por neoplasias é um desafio por se tratar de um conjunto de doenças com particularidades em relação a fatores de risco e etiológicos, características clínicas e epidemiológicas. Neste sentido, fatores como práticas de rastreamento, acesso aos cuidados de saúde e sua qualidade, a precocidade do diagnóstico e a prevalência de fatores de risco podem influenciar a incidência e a mortalidade pelos diferentes tipos de câncer (KAMANGAR *et al.*, 2006).

Parte do comportamento da mortalidade por neoplasias pode ser explicado pela tendência de neoplasias específicas que compõem a principal massa de óbitos do grupo. Estudos recentes evidenciaram tendências das taxas de mortalidade por neoplasias de mama e colo de útero e de próstata, entre os principais cânceres em cada sexo no Brasil, semelhantes às das neoplasias em geral observada no presente estudo, com taxas mais elevadas e decedentes no Sul e Sudeste e mais baixas e ascendentes no Norte e Nordeste (JEREZ-ROIG et al., 2014; GONZAGA et al., 2015; BARBOSA et al., 2016; GUERRA et al., 2017).

A mortalidade por neoplasias de mama e de colo de útero está associada a diferentes contextos socioeconômicos no país (GIRIANELLI; GAMARRA; AZEVEDO E SILVA, 2014), contudo, o acesso à realização de exames de rastreamento para ambos e o impacto na redução da mortalidade permanece crítico em áreas mais pobres (NOVAES; BRAGA; SCHOUT, 2006; BIM et al., 2010; MELO et al., 2016). Entretanto, o programa de rastreio não consegue minimizar as iniquidades em saúde. No Brasil, o rastreamento é realizado por uma rede de confirmação diagnóstica lenta e com barreiras de acesso. Ainda que a disponibilidade de recursos represente uma das principais barreiras ao uso de serviços de saúde, um conjunto de fatores pode explicar a variabilidade encontrada entre as regiões, dos quais se destacam capacidade de compra de serviços de saúde das populações locais, condição social, escolaridade, componentes sociodemográficos e epidemiológicos (SANTOS; MELO, 2011).

A prevalência de mamografia e do exame de papanicolau é mais elevada entre mulheres brancas, de maior renda e escolaridade, que vivem em zona urbana, possuem plano de saúde e utilizam de maneira regular serviços de saúde particulares e não pertencentes a

unidades públicas ou contratadas do SUS (NOVAES; BRAGA; SCHOUT, 2006). Segundo Melo et al. (2016) a probabilidade da mulher fazer mamografia é mais influenciada por condições socioeconômicas, como renda e escolaridade, do que pela disponibilidade de unidades para realização do exame, o que reflete uma distribuição desequilibrada dos serviços de saúde além de desigualdades no uso e qualidade dos serviços que podem ser melhoradas através de políticas sociais e de saúde.

No Brasil, não há rastreamento ativo de neoplasias de próstata, e não existem evidências suficientes que permitam recomendá-lo considerando que tal prática não reduz a mortalidade causada pelo agravo. Apesar disso, na prática clínica há profissionais que mantêm a recomendação de realização dos exames de dosagem do antígeno prostático específico (PSA) e de toque retal entre homens de 50 ou mais anos de idade, e muitos homens os realizam quando buscam atendimento médico de maneira espontânea (BUSATO; ALMEIDA, 2016).

Conforme esperado, estudos identificando os fatores associados à realização de exames de rastreamento do câncer de próstata no Brasil são escassos. Ainda assim, estudos realizados em Juiz de Fora e em alguns municípios de São Paulo constataram maior prevalência de realização de exames de rastreamento entre homens em faixas mais elevadas de escolaridade e renda e usuários de serviços de saúde privados (AMORIM et al., 2011; SANTIAGO et al., 2013).

A composição etária da população é um fator que pode ter contribuído com a distribuição desigual entre as regiões, pois taxas de mortalidade por neoplasias mais altas e de tendência ascendente são observadas em faixas etárias elevadas na Europa, Estados Unidos e Brasil (DOLL, 1990; BAILAR III; GORNIK, 1997; LIMA-COSTA; PEIXOTO; GIATTI, 2004). Neste contexto, é coerente o padrão observado no Sul e Sudeste, dado sua composição etária, e a tendência de aumento nas regiões onde o processo de envelhecimento populacional é mais recente e acelerado, como Norte e Nordeste.

Em relação aos fatores de risco e etiológicos, Danaei *et al.* (2005) observaram que 35% dos óbitos por neoplasias estão associados a nove fatores de risco potencialmente evitáveis: sobrepeso e obesidade, baixo consumo de frutas e vegetais, inatividade física, tabagismo, consumo de álcool, sexo sem proteção, fumaça por uso domiciliar de combustíveis fósseis e contaminação em ambientes de cuidado de saúde, sendo cada uma delas relacionadas a diferentes tipos de câncer.

Neste contexto, a tendência de aumento do risco de mortalidade por neoplasias nas regiões Norte e Nordeste, menos desenvolvidas, e de queda no Sul e Sudeste refletem a

prevalência de tabagismo, obesidade, baixo consumo de frutas e vegetais e atividade física insuficiente em grupos populacionais com piores indicadores socioeconômicos (MONTEIRO et al., 2007; MOURA et al., 2009; MALTA et al., 2011).

O consumo de álcool está associado a neoplasias de fígado, boca, orofaringe e esôfago (DANAIEI et al., 2005), com baixo impacto na mortalidade no país (WÜNSCH & MONCAU, 2002). O aumento das taxas de mortalidade por neoplasias no Norte e Nordeste também pode ter sido influenciado por este consumo, pois a mortalidade por câncer hepático aumentou nestas regiões nos últimos anos (MELO et al., 2017), onde a prevalência de consumo (MINISTÉRIO DA SAÚDE, 2007, 2017) e de dependência de álcool (GALDURÓZ & CAETANO, 2004) são mais elevadas.

As neoplasias de pulmão foram as mais frequentes no país, padrão observado mundialmente (FERLAY et al., 2015). As taxas de mortalidade por esta causa foram mais elevadas no Sul, seguidas pelo Sudeste e mais baixas no Norte e Nordeste para ambos os sexos, corroborando com a correlação positiva com a renda em todo o período. As taxas entre homens foram mais elevadas, com tendência de aumento no Norte e Nordeste, enquanto entre mulheres o aumento foi generalizado. No Rio Grande do Sul o padrão foi discrepante em comparação às demais unidades da federação, concentrando taxas mais elevadas, em ambos os sexos, e apresentando tendência de aumento em todo o período. Estes padrões e tendências são condizentes com diversos estudos nacionais (MALTA; MOURA *et al.*, 2007; AZEVEDO E SILVA *et al.*, 2011; GUERRA *et al.*, 2017) e refletem a prevalência de tabagismo descrita por Malta *et al.* (2013) e Barros *et al.* (2011), mais elevada no Sul e Sudeste, com tendência de queda entre homens e estratos sociais de maior renda e escolaridade, e estabilidade entre mulheres.

As doenças do aparelho circulatório foram a causa de óbito mais frequente em todas as regiões no período analisado e variaram de maneira desigual, diminuindo no Sul e Sudeste, onde eram mais elevadas, e aumentando no Norte e Nordeste. Diversos estudos constataram padrões e tendências regionais semelhantes no país (SOUZA et al., 2001, 2006; CESSE et al., 2009).

No final do período as taxas por doenças cardiovasculares se encontravam em patamares próximos entre as regiões, de maneira que mantidas as tendências descritas anteriormente, é esperado que no futuro o risco de óbito no Norte e Nordeste ultrapasse o das demais regiões. Contudo, estimativas do Estudo Carga Global de Doença (BRANT et al., 2017), baseadas em correções dos dados do Sistema de Informação sobre Mortalidade, mostraram um platô com pouca variação das taxas de mortalidade cardiovascular nos estados

brasileiros entre 2010 e 2015, indicando que as tendências observadas no presente estudo podem não se manter nos próximos anos, ou que a qualidade da informação sobre mortalidade pode tê-las influenciado.

Reforçando esta perspectiva, parte do aumento das taxas de mortalidade por estas doenças observado no Norte e Nordeste tem sido atribuído à queda da mortalidade por causas mal definidas, de maneira que a melhoria na qualidade do diagnóstico e da determinação da causa de óbito pode ter resultado em aumento “artificial” do número de óbitos por doenças do aparelho circulatório (SOUZA et al., 2001; CESSÉ et al., 2009). Este padrão de subnotificação prejudica de um modo geral a análise das principais causas de mortalidade.

O padrão da magnitude das taxas no Brasil foi contrastante com o observado na Europa, onde regiões mais desenvolvidas apresentaram menor mortalidade por doenças do aparelho cardiovasculares, em contrapartida, a tendência das taxas nas regiões Norte e Nordeste foi semelhante à de países do leste Europeu, menos desenvolvidos, enquanto no Sul e Sudeste o comportamento remete aos países da Europa Ocidental e aos Estados Unidos, mais desenvolvidos (CHOR et al., 1995; SANS et al., 1997; LOTUFO, 1998; MULLER-NORDHORN et al., 2007). Estudos têm evidenciado a forte associação entre condição social e doenças cardiovasculares. As desigualdades em saúde se manifestam através da expectativa de vida, das taxas de morbidade e mortalidade e das mortes precoces (WAMALA; LYNCH; KAPLAN, 2001; MACKENBACH et al., 2008).

Tanto o contexto socioeconômico quanto a dinâmica demográfica podem explicar estas similaridades e disparidades; segundo Roth et al. (2015), apesar de 80% do aumento na mortalidade cardiovascular no planeta entre 1990 e 2013 estar associada ao crescimento e envelhecimento populacional, em países mais desenvolvidos as melhorias na saúde e os avanços no tratamento e no controle dos fatores de risco destas doenças superaram o efeito demográfico, resultando em redução das mesmas, o que não ocorre países em desenvolvimento. Neste sentido, é possível que as taxas mais elevadas de doenças cardiovasculares no Sul e Sudeste tenham resultado do aumento constante, a partir da década de 1940, decorrente do crescimento e envelhecimento populacional e a queda recente das mesmas nestas regiões possa ser explicada por um contexto socioeconômico que propiciou melhores condições de saúde nos últimos anos, enquanto o avanço das taxas no Norte e Nordeste ainda seja um reflexo marcante de transição demográfica acelerada e tardia. O aumento das taxas de mortalidade em regiões menos desenvolvidas é condizente também com a maior prevalência de fatores de risco como tabagismo, obesidade, consumo de carnes com

excesso de gordura, inatividade física, hipertensão e diabetes entre indivíduos de menor escolaridade no Brasil (MOURA et al., 2009; MALTA et al., 2011; ISER et al., 2012).

As doenças cerebrovasculares se destacaram como principal causa de óbito entre as doenças do aparelho circulatório no Brasil no período analisado, apresentando padrões semelhantes ao das doenças do aparelho circulatório, com valores mais elevados e tendência de queda no Sul e Sudeste e aumento no Norte e Nordeste, sendo as taxas mais elevadas entre homens. Segundo Mansur et al. (2009) a elevada mortalidade por doenças cerebrovasculares reflete o controle ineficiente de fatores de risco no país, principalmente da hipertensão. Estudos que estimam a prevalência de hipertensão nas regiões e unidades da federação são escassos, o que dificulta a compreensão da relação entre sua distribuição e os padrões regionais da mortalidade por doenças cerebrovasculares.

Dados do Vigitel (MINISTÉRIO DA SAÚDE, 2007, 2017) mostraram maior prevalência de hipertensão diagnosticada por médico em capitais do Sudeste em 2006, padrão condizente com as taxas de mortalidade por doenças cerebrovasculares mais elevadas nesta região, enquanto o aumento acelerado nas capitais do Norte, Nordeste e Centro-Oeste e a queda no Sudeste até 2015 refletem as tendências regionais destacadas no presente estudo. Cabe ressaltar que, como a estimativa é obtida a partir evento autorreferido, em parte, as proporções mais elevadas podem representar maior acesso da população aos serviços de saúde, propiciando diagnóstico e controle da doença; tal fato pode ter ocorrido nas capitais da região Sul onde as prevalências de hipertensão aumentaram entre 2006 e 2015 e as taxas de mortalidade por doenças cerebrovasculares diminuíram. Além disto, o padrão observado nas capitais pode não representar desigualdades intra e interestaduais.

A correlação positiva entre a taxa de mortalidade por doenças cerebrovasculares e a renda nos anos iniciais é condizente com os padrões e tendências das taxas aqui destacadas, mais elevadas em regiões mais desenvolvidas. Os estudos de Paes (2016) e de Alves & Barbosa (2010) observaram relação similar entre as taxas de mortalidade de idosos por doenças do aparelho circulatório e indicadores socioeconômicos entre as unidades da federação no país, o que reforça a importância do componente etário. Tal componente está intimamente relacionado com o nível de desenvolvimento no comportamento geral da mortalidade neste grupo de causa e corrobora com padrões identificados no presente estudo.

Em contrapartida, os achados são contrastantes com outros estudos que abordaram a mortalidade por doenças cerebrovasculares e hipertensivas em contextos específicos, tal como bairros de Porto Alegre (BASSANESI *et al.*, 2008) e do Rio de Janeiro (ALVES, 2013), municípios com mais de cem mil habitantes (ISHITANI et al., 2006) e os estados de São

Paulo, Rio de Janeiro e Rio Grande do Sul (SOARES et al., 2013), evidenciando contrastes nos padrões de mortalidade que vem sendo acentuados pela disparidade social. Importante destacar que esta discordância é um indicativo de que a correlação das condições socioeconômicas com a mortalidade por doenças cardiovasculares em grupos específicos é diferente daquela observada entre grupos. Neste sentido, embora entre as unidades da federação e microrregiões a mortalidade por doenças cerebrovasculares mostre-se mais elevada em locais mais desenvolvidos, quando o nível de observação fica mais próximo do individual o contexto socioeconômico mais precário sobressai. Este efeito constitui maior risco de óbito para estas doenças, o que provavelmente está associado à menor disponibilidade, acessibilidade e qualidade dos serviços de saúde, acesso a medicamentos, entre outros.

A ausência de correlação entre a taxa de mortalidade por doenças cerebrovasculares e os indicadores socioeconômicos em 2010 pode estar associada à baixa variabilidade das taxas entre as regiões no final do período analisado. Este pode ser um indicativo de uma inversão na correlação da mortalidade por estas doenças com os indicadores socioeconômicos nos próximos anos ocorrerá; fato plausível considerando a tendência de queda das taxas em regiões mais desenvolvidas e aumento nas menos desenvolvidas.

O local de ocorrência do óbito, a faixa etária e a unidade da federação de residência sobressaíram como importante preditor na determinação da causa específica entre os óbitos por doenças do aparelho circulatório. As doenças cerebrovasculares foram mais frequente na faixa etária de 60 a 79 anos de idade e nas unidades da federação do Sul e Sudeste, achado condizente com outros estudos e demais achados da presente pesquisa (LOLIO; LAURENTI, 1986; SOUZA et al., 2001, 2006; MANSUR et al., 2009). A maior frequência de óbitos registrados com código lixo na faixa etária de 80 anos ou mais, unidades da federação do Nordeste e local de ocorrência domicílio reflete a qualidade da informação sobre o óbito no país (MELLO JORGE et al., 2008).

O local de ocorrência do óbito, a faixa etária e a unidade da federação de residência sobressaíram como importantes preditoras na determinação da causa específica entre os óbitos por doenças do aparelho circulatório. As doenças cerebrovasculares foi mais frequente na faixa etária de 60 a 79 anos de idade e nas unidades da federação do Sul e Sudeste, achado condizente com diversos estudos e com os demais achados do presente estudo . A maior frequência de óbitos registrados com código lixo na faixa etária de 80 anos ou mais, unidades da federação do Nordeste e local de ocorrência domicílio reflete a qualidade da informação sobre o óbito no país.

As taxas de mortalidade por doenças do aparelho respiratório foram maiores entre homens e, de maneira geral, mais elevadas no Sul, Centro-Oeste e Sudeste; nestas regiões observa-se aumento de taxas nos anos iniciais e decaimento no final do período, e patamares mais baixos em relação ao Norte e Nordeste, onde a tendência foi de aumento contínuo. Tais padrões e tendências mostram-se semelhantes aqueles observados por Paes (2016), ao analisar a mortalidade de idosos por este grupo de doenças entre 1990 e 1995.

As diferenças regionais nas taxas de mortalidade por pneumonia, agravo mais frequente dentro do grupo de doenças respiratórias, foram menos intensas do que as descritas anteriormente; revelando patamares mais baixos e com tendência de aumento no Norte e Nordeste, seguidas do Centro-Oeste, e ligeiramente mais elevadas no Sul e Sudeste, onde permaneceram estáveis. Tendências regionais semelhantes foram constatadas por Campagna et al. (2009), no período compreendido 1992 e 2005, em relação à mortalidade de idosos por doenças relacionadas à influenza, entre as quais estão as pneumonias. Padrão similar ao observado por Oliveira et al. (2013) em relação à mortalidade por pneumonia entre idosos nas regiões Sudeste e Nordeste entre 1996 e 2008.

O aumento constante da idade de óbito por doenças do aparelho respiratório pode estar associado às diversas transformações socioeconômicas e demográficas no país, tendo em vista a correlação entre os componentes deste grupo de causas e indicadores socioeconômicos e características domiciliares (PRIETSCH et al., 2003; CHIESA et al., 2008), cujo impacto na mortalidade mostra-se mais elevado entre os extremos etários. Tal se vê com as infecções respiratórias agudas e pneumonias em crianças (BENGUIGUI, 2002; MACEDO et al., 2007; CHIESA et al., 2008; RODRIGUES et al., 2011) e as doenças relacionadas à influenza, como pneumonias, bronquites e obstrução crônica das vias aéreas entre idosos (FRANCISCO et al., 2005; DONALISIO et al., 2006; CAMPAGNA et al., 2009).

Neste sentido, a baixa mediana da idade de óbito, no início da década de 1980, no Norte e Nordeste, é condizente com a maior mortalidade infantil, associada a um pior contexto socioeconômico no período e que persiste até o início dos anos 1990. Por outro lado, o aumento repentino da idade de óbito no Sul e Sudeste reflete um estágio de envelhecimento populacional avançado, em que a mortalidade de idosos se sobrepõe às demais faixas etárias, com elevada participação das pneumonias.

Esta perspectiva é corroborada por diversos estudos que descreveram redução da mortalidade por doenças respiratórias entre crianças (HOLCMAN et al., 2004; FISCHER et al., 2007a; NASCIMENTO et al., 2014) e aumento entre idosos (MS/SVS, 2004; CAMPAGNA et al., 2009), sobretudo em faixas etárias mais elevadas (FRANCISCO et al.,

2003; FRANCISCO et al., 2005) e podem explicar, em parte, as desigualdades na magnitude das taxas de mortalidade entre regiões. Além da composição etária, a prevalência de tabagismo no país também refletem estes padrões de mortalidade (BARROS, et al., 2011; MALTA et al., 2013).

As taxas de mortalidade por pneumonia apresentaram correlação positiva com os indicadores de renda e esgotamento sanitário, o que é condizente com taxas mais elevadas em regiões mais desenvolvidas. Paes (2016) observou relação semelhante entre as taxas de mortalidade por doenças do aparelho respiratório e diversos indicadores socioeconômicos.

Os principais preditores para o óbito por doenças do aparelho respiratório foram faixa etária, ano de óbito e unidade da federação de residência. As doenças crônicas das vias aéreas inferiores, em que se incluem bronquite crônica, enfisema e DPOC, sobressaíram entre 1996 e 2013 na região Sul nos grupos a partir de 19 anos de idade; já no Sudeste, Norte e Centro-Oeste entre aqueles com 60 e 79 anos. Padrões condizentes com a composição etária das regiões e prevalência de tabagismo no país, dado que estão associadas ao hábito de fumar e ao envelhecimento (PESSÔA & PESSÔA, 2009). Comportamento semelhante foi descrito por Campos (2003) em estudo sobre as taxas de mortalidade por DPOC, consideradas mais elevadas e em ascensão na faixa etária de 65 ou mais anos e na região Sul.

A mortalidade por pneumonia foi mais frequente nas faixas etárias mais jovens em todo o período, e entre 1979 e 1996 nas demais faixas etárias em alguns estados, particularmente no Nordeste e no Rio de Janeiro entre 2005 e 2013. Achados condizentes com o elevado impacto do agravo na mortalidade entre crianças (RODRIGUES, 2011), e com os demais resultados do presente estudo, revelando elevação de taxas no Nordeste.

A mortalidade por causas externas mostrou-se mais elevada entre homens e na região Centro-Oeste de maneira geral e patamares mais baixos no Sul. Verifica-se tendência de aumento de taxas no Norte, Nordeste e Centro-Oeste e queda nas demais regiões; a maioria dos óbitos ocorreram entre 20 e 50 anos de idade em todas as regiões. Os homicídios sobressaíram como a principal causa do grupo no período, com taxas de mortalidade maiores entre homens e tendência de aumento no Norte e Nordeste e estabilidade para as demais regiões.

Estes achados apontam para a manutenção do cenário observado nos estudos de Gawryszewski et al. (2004) e Souza & Lima (2006) que identificaram maior mortalidade por causas externas no sexo masculino, nas faixas etárias a partir dos 15 anos de idade e homicídios como principal causa de óbito entre as causas externas no país em 2000 e 2003, respectivamente. Reichenheim (2015) também constatou os homicídios como principal causa

de óbito entre as causas externas no Brasil em 2007, contudo descreve queda nas taxas de mortalidade por esta causa nas regiões Norte, Centro-Oeste e Sudeste e aumento no Nordeste e Sul, tendência discrepante à observada no presente estudo.

Algumas unidades da federação destacaram-se com elevadas taxas de mortalidade tanto por causas externas quanto por homicídios, como Rondônia e Roraima no Norte, Pernambuco e Alagoas no Nordeste e Rio de Janeiro e Espírito Santo no Sudeste. De maneira semelhante, Peres & Santos (2005), Souza & Lima (2006) e Bando & Lester (2014) observaram maiores taxas de mortalidade por homicídios nas capitais destas unidades da federação no período em diversos períodos entre 1991 e 2010.

Observou-se correlação negativa entre as taxas de mortalidade por homicídios e o indicador de envelhecimento populacional, achado coerente com a maior mortalidade por esta causa na faixa etária jovem e adulta como visto na literatura (GAWRYSZEWSKI et al., 2004; SOUZA & LIMA, 2006). A correlação com o desemprego foi positiva e ascendente, padrão identificado nas capitais brasileiras, em 2010, por Bando & Lester (2014) e compatível com a elevada frequência de homicídios em áreas mais pobres e de intensa desigualdade social (PEREIRA et al. 2015; SOUSA et al., 2014).

Na abordagem das informações individuais sobressaíram local de óbito, faixa etária e unidade da federação de residência como principais variáveis preditoras para a causa externa como causa específica de óbito. Além da maior mortalidade por homicídios na faixa etária de 19 a 29 anos e o destaque para o Espírito Santo, Rio de Janeiro e Pernambuco, sobressaíram-se também Alagoas, Amazonas, Pará e Bahia. Os acidentes de transporte mostraram-se mais concentrados em faixas etárias mais elevadas, sobretudo entre 30 e 59 anos de idade.

Segundo Andrade & Diniz (2013) a mortalidade por homicídios no Norte se concentra em áreas rurais devido a conflitos de ocupação e exploração territorial e em regiões metropolitanas das capitais, como Belém, Porto Velho e Rio Branco; já no Nordeste observa-se também bolsões em áreas de intensa atividade turística. Neste contexto, apesar da repercussão da violência transmitida pela mídia a respeito de grandes centros urbanos, como Rio de Janeiro e São Paulo, principalmente aquela envolvida com operações legais e combate ao tráfico de drogas, é no Norte e Nordeste que se concentra o maior risco relacionado à mortalidade por homicídios do país. Tal tendência se mostra em ascensão e está relacionada tanto às disputas por terra quanto à carência social nos centros urbanos.

Observou-se queda da mortalidade proporcional por causas mal definidas em todo o país, indicativo de melhoria da qualidade da informação sobre mortalidade. A redução geral foi mais acentuada nas regiões Norte, Nordeste e Centro-Oeste, cujos patamares mantinham-

se mais elevados em todo o período, comportamento já destacado em estudos anteriores e que pode estar relacionado ao maior acesso à assistência médica em situações fatais nestas regiões (LAURENTI et al., 2004; COSTA & MARCOPITO, 2008; FRANÇA et al., 2008; CGIAE, 2013).

A melhoria da qualidade dos registros de óbito no país aumenta a confiança nos padrões e tendências aqui descritos. Contudo, apesar dos avanços na qualidade da informação sobre a causa de óbito e as variáveis prioritárias do sistema, como data de óbito, local de residência, idade, sexo e causa de óbito, persistem limitações em relação ao preenchimento de alguns campos da Declaração de Óbito, particularmente em relação à completude satisfatória ao longo da série histórica disponível no Sistema de Informação sobre Mortalidade. Mesmo em anos recentes a proporção de dados faltantes e “ignorados”, relacionados à escolaridade, ocupação, realização de exames, cirurgias e assistência médica recebida, ainda é considerado elevada.

Neste sentido, a opção metodológica de abordar todo o Sistema de Informação sobre Mortalidade no país desde sua criação implicou em limitações relacionadas às informações disponíveis que poderiam ser padronizadas e abordadas em todo o período, de maneira que das 10 variáveis selecionadas inicialmente, desconsiderando as prioritárias e essenciais, apenas 3 foram mantidas na análise dos dados individuais. Entre as variáveis não utilizadas a escolaridade é a que provavelmente permitiria uma compreensão mais específica dos padrões de mortalidade, por representar, ainda que indiretamente, as condições socioeconômicas.

Em contrapartida, a escolha de selecionar apenas as variáveis com elevada completude e de retirar os registros com informação faltante ou ignorada parece ter sido acertada, pois menos de 10% dos registros foram excluídos da análise e não há indícios de perda seletiva, pois não foram identificados padrões discrepantes com a literatura que pudessem ter sido gerados pela exclusão.

As principais vantagens da utilização dos dados do SIM no presente estudo foram fácil acesso, baixo custo, grande volume, boa cobertura e longo período de seguimento. A completude das variáveis essenciais, como sexo, idade, município de ocorrência e causa do óbito foram elevadas e a proporção de óbitos por causas mal definidas diminuiu de maneira constante. Entretanto, diversas variáveis apresentaram baixo preenchimento, muitas foram introduzidas apenas em anos mais recentes e algumas sofreram alterações constantes, dificultando sua análise em um longo período de tempo, apesar da documentação e os arquivos de conversão disponibilizados permitirem a compatibilização das informações. Também é provável que a cobertura e a qualidade dos dados seja desigual entre as regiões.

A despeito de persistirem limitações do Sistema, cabe destacar que faltam estudos nacionais que utilizem procedimentos para mineração do grande volume de dados sobre mortalidade, fato que de alguma forma limita a comparabilidade de parte dos achados deste estudo com a literatura.

O único estudo identificado que aplicou técnicas de mineração a uma grande massa de dados sobre mortalidade, teve como um dos objetivos testar a utilização de algoritmos de classificação para predizer a causa de óbito abordando mais de um milhão de registros de uma base de dados sobre mortalidade da organização mundial de saúde referente a diversos países (PAOIN, 2011). Entre os algoritmos utilizados no referido estudo, a árvore de decisão foi aquele apresentou acurácia mais elevada, cerca de 40%, considerada baixa pelo autor, o que foi atribuído à pequena disponibilidade de variáveis explicativas (apenas 3: país de ocorrência do óbito, sexo e idade) e ao elevado número de classes da variável resposta, tendo em vista que obteve uma árvore com 440 folhas. Problemas semelhantes ocorreram no presente estudo, pois as análises que envolveram os métodos de mineração de dados a partir de informações individuais sobre mortalidade também foram prejudicadas pelo número reduzido de variáveis explicativas, de maneira que os modelos, em sua maioria, destacaram o ano de óbito, faixa etária, local de ocorrência e unidade da federação de residência como determinantes da causa específica de óbito. Neste sentido, é provável que uma maior diversidade de variáveis possibilitasse a descrição de padrões e tendências de maior valor informativo para a saúde pública.

Seria possível aumentar a diversidade de variáveis utilizando o mesmo banco de dados estruturado para o presente estudo através da criação de indicadores relacionados às informações disponíveis. Abordar exclusivamente dados referentes aos anos mais recentes, principalmente a partir de 2000, seria uma alternativa dado que se observa uma maior variedade de informações de melhor completude e confiabilidade no Sistema, fato que permitiria aumentar a variedade de indicadores. Desta forma seria mantido um grande volume de dados já que o número de registros a ser analisado ultrapassaria 14 milhões, apesar de a descrição da tendência temporal da mortalidade em todo o período de existência do SIM sofrer prejuízo.

Diversos estudos têm destacado o potencial do *linkage* entre diversos sistemas de informação para aumentar o volume e a variedade na análise de grandes dados na área da saúde (BORDER, 2014; CHIAVEGATTO FILHO, 2015; ASCHE et al., 2017; EHRENSTEIN et al., 2017). No caso do Brasil, o *linkage* de dados dos diversos sistemas de informação mantidos pelo Ministério da Saúde, como o próprio SIM, o Sinasc, o Sistema de

informações Hospitalares (SIH) entre outros, e também de outras pesquisas sociodemográficas como o Censo, a Pesquisa Nacional de Amostra de Domicílios (PNAD) e a Pesquisa de Orçamento Familiar (POF) proporcionaria um manancial ainda mais amplo de informação para a mineração de padrões que favorece a compreensão de maneira mais ampla de diversos eventos de saúde, norteando políticas públicas mais específicas e efetivas.

A opção por agrupar os diversos códigos referentes à causa básica de óbito em causas específicas anulou o possível efeito de um grande número de categorias para as variáveis resposta que poderiam prejudicar a adequação dos métodos utilizados. Além disto, como os procedimentos realizados se basearam na descrição de padrões de mortalidade através da representação gráfica dos modelos de árvore e da Análise de Correspondência Múltipla, como se propõe esta tese, a redução do número de classes possibilitou a obtenção de estruturas de compreensão mais simples e intuitiva.

Mesmo realizando este agrupamento de códigos de causa de óbito, a composição das causas específicas em determinados grupos dificultou a identificação de padrões de mortalidade a partir da mineração de dados. As neoplasias, por exemplo, são representadas por diversas doenças específicas que individualmente possuem pouca importância proporcional na composição geral da mortalidade do grupo. O critério adotado para seleção das causas específicas contribuiu para que neste caso a categoria “demais causas” fosse a mais frequente, representando mais de 70% dos óbitos por neoplasias. Como o algoritmo para definir os critérios de divisão dos ramos das árvores de decisão é baseado também no nível de acerto que cada divisão proporciona, a “simples” alocação de todos os casos em uma folha classificada como “demais causas” pelo modelo resultou em uma acurácia superior aos demais modelos apresentados, contudo, sem maior utilidade em relação à extração de informação do modelo.

Quanto às doenças do aparelho circulatório, é provável que a limitação do modelo de árvore de decisão, que apontou apenas o local de ocorrência como preditor da causa específica de óbito, esteja relacionada à homogeneidade da ocorrência dos óbitos pelas causas específicas considerando as variáveis explicativas utilizadas. Assim, é possível que alguma variável não disponível, principalmente um *proxi* de classe social como a escolaridade ou a ocupação, pudesse favorecer a elucidação de padrões mais interessantes na compreensão da mortalidade por este grupo de causa.

Não obstante as limitações acerca da qualidade de dados e particularidades relativas às escolhas metodológicas para compatibilização das informações, os achados oriundos da mineração de dados foram condizentes com a literatura disponível, apontando para o potencial

de utilização destas técnicas em dados da área da saúde no país. A aplicação dos métodos de mineração de dados no presente estudo foi direcionada à determinação de fatores individuais relacionadas à mortalidade por causas específicas considerando os óbitos segundo grupo de causa, perspectiva pela qual os padrões e tendências de mortalidade foram descritos principalmente através de representações gráficas. As mesmas técnicas podem ser utilizadas para identificar padrões de mortalidade no país a partir de perspectivas diferentes, assim como outros métodos de mineração podem complementar as análises aqui realizadas e permitir abordagens diferentes destas informações.

O avanço e o barateamento da tecnologia e o aprimoramento das técnicas de análise também favorecem a abordagem de grandes massas de dados, que tende a se tornar cada vez mais acessível. Prova disto é que os modelos *Random Forest* nas abordagens propostas só puderam ser implementadas a partir de novembro de 2016 quando a biblioteca *Ranger* do R foi disponibilizada, pois até então as bibliotecas existentes não permitiam a utilização de um volume tão grande de informação. Adicionalmente, todo o processo de pré-processamento e a maior parte do processamento, com exceção da Análise de Correspondência Múltipla em todas as abordagens e dos Modelos *Random Forest* na abordagem com os óbitos por todas as causas, foi realizado em um *desktop* de uso residencial, com configuração de *hardware* compatível com modelos acessíveis no mercado brasileiro.

Para as exceções exigidas foi utilizada uma máquina de poder de processamento mais elevado e com configurações fora da realidade comercial do país. Contudo, existem serviços de processamento em *nuvem*, contratados conforme a necessidade do usuário, que podem suprir a necessidade de uma máquina física com tal poder processamento e que seria aproveitada em plenitude em situações muito específicas como as do presente estudo, gerando um custo desnecessário.

A maior acessibilidade aos métodos de análise empregados em conjunto com a estruturação das informações do SIM como uma grande massa dados e a utilização de mineração de dados, podem promover um olhar revitalizado para este sistema de informação em um futuro próximo.

Os padrões e tendências aqui destacados não representam novidade sobre a temática da mortalidade no país, contudo, os estudos geralmente abordam causas específicas de óbito em períodos e espaço geográfico restritos, de maneira que as desigualdades regionais mais amplas muitas vezes não são evidenciadas, assim como *outliers* locais.

Os métodos de análise gráfica confirmaram-se como boas alternativas na descrição de padrões e tendências da mortalidade e suas vantagens devem ser destacadas. Os gráficos de

radar apresentaram a evolução regional da mortalidade proporcional segundo grupos de causa de maneira simples e robusta; os *boxplots* da idade de óbito permitiram a comparação da distribuição regional do indicador ao longo dos anos, apesar de não representar o risco de óbito por idade; os *boxplots* das taxas de mortalidade e dos indicadores sociais econômicos e demográficos agrupados por região possibilitaram a comparação das distribuições e sua evolução nos anos censitários; a representação da matriz de correlação em conjunto com elementos gráficos com escala contrastante de cores atraiu a atenção para associações mais intensas; os *heatmaps* apresentaram as tendências temporais das taxas de mortalidade nas unidades da federação de maneira mais clara do que um gráfico de linhas com tantas unidades de análise permitira, possibilitaram a noção do espalhamento espaço-temporal do indicador, apesar de o espaço não ser representado de maneira tão precisa quanto em mapas temáticos, e foram úteis para identificar *ouliers*; os gráficos resultantes da Análise de Correspondência Múltipla representaram de maneira objetiva e intuitiva a relação entre as categorias das variáveis.

No presente estudo, as informações foram analisadas em diferentes níveis de observação, abordando desde os óbitos registrados individualmente no SIM até as macrorregiões, passando pelas microrregiões e unidades da federação. Utilizou-se diferentes técnicas de análise e explorou-se ferramentas visuais alternativas de apresentação de achados, condensando um grande e variado volume de informação processada sobre a mortalidade, sob uma perspectiva inédita sobre o tema no país, que pode auxiliar a compreensão de desigualdades regionais na mortalidade por diferentes grupos de doença, bem como sua evolução.

8 CONSIDERAÇÕES FINAIS

A presente Tese buscou analisar padrões e tendências da mortalidade por diversos grupos de causa e causas específicas, abordando os registros do Sistema de Informação Sobre a Mortalidade entre 1979 e 2013 de maneira individual, como uma grande massa de dados, e segundo diferentes níveis de agregação, implementando diversos métodos de análise em uma abordagem de mineração de dados.

Três aspectos do desenvolvimento deste estudo devem ser destacados: a utilização dos dados do Sistema de Informação Sobre a Mortalidade em diferentes abordagens; a implementação de métodos mineração para a identificação de padrões de mortalidade no país; e os padrões e tendências de mortalidade.

A utilização das microrregiões como unidade de observação se mostrou uma alternativa interessante por garantir um nível de agregação mais baixo do que as unidades da federação e por se comportar de maneira estável ao longo do período, pois a composição das microrregiões não foi alterada.

Os achados permitiram uma visão ampla da mortalidade no país nos últimos 35 anos, caracterizando padrões e tendências regionais que refletem mudanças no processo de adoecimento da população. Neste sentido, é importante destacar o potencial de utilização dos dados deste sistema de informação, principalmente através de métodos e abordagens inovadoras que permitam subsidiar o planejamento da saúde pública no país. Apesar dos inegáveis avanços verificados na qualidade dos dados, medidas que ofereçam maior acurácia dos dados aportados no Sistema certamente favorecem a compreensão da dinâmica da mortalidade no Brasil.

Quanto aos métodos de análise, o uso de um conjunto de ferramentas de análise exploratória capaz de identificar padrões marca positivamente o alto potencial de aplicabilidade desta Tese. Diferente de outros estudos que analisaram a mortalidade no país, este utiliza métodos e técnicas diversificadas que permitiram abordar o sistema de mortalidade de maneira integral e sob diferentes perspectivas, explorando nuances de padrões relacionados a diferenças entre populações e mudanças nas populações ao longo do tempo.

A estruturação da base de dados mostrou-se procedimento essencial para a realização deste estudo, uma vez que todo o processo de obtenção dos microdados, seleção, transformação, compatibilização e padronização de variáveis foi concebido para permitir a abordagem das informações individuais como uma grande massa de dados, e para extrair os

dados necessários para cada etapa de análise, segundo diferentes níveis de agregação e referente a todo o período de tempo abordado.

Os resultados da utilização dos dados individuais foram menos informativos do que o esperado e não permitiram a identificação de padrões ainda não identificados ou que representassem novidades sobre o que se conhece da mortalidade no país, o que pode ser atribuído ao número pequeno de variáveis analisadas, às características dos algoritmos e a sua aplicação.

A baixa disponibilidade de variáveis representa uma limitação relacionada à qualidade dos dados e às opções metodológicas adotadas no estudo, para superá-la seria necessário mudar a abordagem concentrando as análises em dados mais recentes e atualizados, com um número maior de variáveis disponíveis e de melhor qualidade de preenchimento. A possibilidade de integração e relacionamento probabilístico de registros oriundos de outros estudos e sistemas a partir de bases territoriais e grupos populacionais específicos representa alternativa adicional à incorporação de dados.

Há que se considerar aspectos relacionadas ao algoritmo de árvore de decisão, uma vez que a análise de correspondência múltipla utilizou as variáveis destacadas como importantes pelo modelo *Random Forest*, baseado no ajuste de árvores, propagando provavelmente consequências relacionadas à limitação do método. A maneira como o algoritmo ajusta a árvore de decisão faz com que as relações mais frequentes sejam destacadas, o que pode ser considerado um ponto positivo por garantir consistência aos achados, mas “esconde” as associações menos frequentes, as exceções ou *outliers*, que podem ser importantes para a compreensão de padrões locais ou específicos da mortalidade. Outros algoritmos de aprendizado de máquina podem ser mais interessantes na detecção destas exceções em grandes massas de dados.

As limitações relativas à aplicação do algoritmo também têm relação com opções metodológicas do estudo. Os algoritmos de árvore foram direcionados para determinar os fatores individuais relacionados à mortalidade por causas específicas, o que diminui a perspectiva dos padrões a serem descritos relacionados a um subgrupo de causas dentro de um subconjunto total de óbitos registrados. Em estudos futuros, oriundos deste, devem emergir novas “perguntas” voltadas para a exploração de informações individuais sobre a mortalidade, utilizando os mesmos algoritmos deste estudo e a mesma base dados, apontando padrões mais interessantes. Um exemplo de pergunta que pode ser respondida com esta mesma base de dados, sem a necessidade de segmentá-la, implementando o modelo de árvores e complementar aos padrões da mortalidade aqui descritos diz respeito a “que fatores

individuais determinam a idade de óbito?”. Uma árvore de regressão poderia ser ajustada tendo a idade de óbito como variável resposta numérica e as mesmas variáveis explicativas utilizadas nas abordagens, além do próprio grupo de causa ou causa específica de óbito.

Ainda em relação à aplicação do algoritmo de árvore, poderiam ter sido testadas outras categorizações das variáveis contínuas como idade e ano de óbito, ou mesmo sua utilização como variáveis numéricas. Contudo, estas variáveis não poderiam ser utilizadas na análise de correspondência, que é específica para dados categóricos, e cada procedimento de ajuste nos dados, por menor que seja, eleva o custo computacional e demanda tempo de processamento, tanto na reclassificação da variável, quanto no ajuste de novos modelos.

É necessário reforçar que os padrões destacados a partir das informações individuais foram condizentes com a literatura e que o volume de dados analisados garante robustez aos achados, o que aponta para o potencial de utilização destes e outros métodos de mineração a esta grande massa de dados sobre a mortalidade no país. Entretanto, como qualquer outro método, sua aplicação pode ser limitada pela qualidade dos dados, ainda que em grande volume, e pelas características do problema analisado, além de envolver elevado custo de processamento, o que deve ser ponderado em cada situação.

Os achados das três etapas se complementaram, permitindo uma visão ampliada dos padrões de mortalidade no país, desde a tendência das taxas de mortalidade, mortalidade proporcional e idade de óbito, por diferentes grupos de causa nas regiões, unidade da federação e microrregiões, a evolução da correlação da mortalidade por causas específicas com os indicadores sociais, econômicos e demográficos e os preditores individuais relacionados a mortalidade por causas específicas em cada grupo de causa.

Observou-se tendências regionais distintas convergindo para padrões semelhantes da magnitude da mortalidade por doenças infecciosas, que decaíram em todas as regiões, neoplasias e doenças do aparelho circulatório, que aumentaram no Norte e Nordeste onde foram mais baixas e decaíram no Sul e Sudeste onde foram mais altas. Já o comportamento da mortalidade por doenças do aparelho respiratório e por causas externas apontaram para a definição e manutenção de padrões regionais específicos. A evolução da correlação das taxas de mortalidade por causas específicas com os indicadores socioeconômicos e demográficos revelaram intensas transformações e desigualdades na expressão da mortalidade em diferentes contextos no país. A abordagem dos dados individuais destacou padrões e tendências de mortalidade concordantes com a literatura.

A proximidade da magnitude da mortalidade entre as regiões provavelmente é resultado do processo de transição demográfica e epidemiológica generalizado e acelerado no

país, pois como o controle das doenças infecciosas foi realizado através de medidas de caráter hexógeno com grande impacto imediato em todo o território nacional diminuindo a mortalidade infantil e possibilitando o aumento da expectativa de vida, era natural que o processo de envelhecimento populacional atingisse todas as regiões e que as doenças crônicas e não transmissíveis sobressaíssem como as principais causas de morbimortalidade da população. Desta maneira, era de se esperar que no correr dos anos as transformações demográficas aceleradas resultassem em novo cenário de estabilidade, no qual as diferenças regionais na composição etária da população seriam pequenas ao ponto de não refletir padrões desiguais de mortalidade.

Não é possível afirmar que o Brasil tenha atingido tal estágio de estabilidade e esteja menos desigual demograficamente, contudo, parece caminhar nesta direção considerando um cenário menos favorável. Como as transições são permeadas por fatores sociais, econômicos e culturais, distribuídos de maneira desigual no país, as diferentes tendências regionais da mortalidade expressam o ritmo no qual as transformações demográficas e epidemiológicas ocorreram, reforçando a perspectiva de um processo de polarização geográfica e social da mortalidade no qual diversos grupos de doenças atingiram regiões e grupos populacionais em intensidade distinta. Esta diferença de ritmo foi constatada ao identificar que as mudanças na magnitude da mortalidade segundo grupos de causa no Norte e Nordeste nos últimos 30 anos foram semelhantes àquelas processadas no Sul e Sudeste há 40 ou 50 anos.

Contudo, apesar da melhoria nas condições socioeconômicas do país, os avanços não foram suficientes para reduzir desigualdade sociais, de maneira que a recente tendência de igualdade na magnitude da mortalidade entre regiões parece ser preenchida por um amplo espectro de iniquidade. Estas iniquidades foram percebidas quando se observou que apesar da aproximação das taxas de mortalidade por doenças do aparelho circulatório entre as regiões em anos recentes, tendência que provavelmente será repetida pelas neoplasias nos próximos anos, a idade de óbito por estas causas foi mais baixa no Norte e Nordeste quando comparadas a Sul e Sudeste, indicando uma maior carga destas doenças nas regiões menos desenvolvidas. De maneira semelhante, mesmo com queda generalizada da mortalidade por doenças infecciosas intestinais, ela permaneceu em patamares ligeiramente mais elevados no Norte e Nordeste. A mortalidade por causas externas também sobressaiu em algumas unidades da federação do Centro-Oeste, Norte e Nordeste onde as desigualdades sociais são mais intensas.

Este panorama indica que a polarização geográfica, que durante muitos anos caracterizou a transição epidemiológica no país através de diferenças marcantes nas taxas de mortalidade entre as regiões, perdeu intensidade, enquanto a polarização social se torna cada

vez mais significativa. Nesta perspectiva, as desigualdades na mortalidade passam a ser representadas por padrões distintos nos diferentes estratos sociais, sobressaindo as características do contexto socioeconômico na determinação da distribuição da mortalidade, de maneira que grupos populacionais menos favorecidos demandam maiores cuidados de saúde.

Por um lado, é possível que os indivíduos inseridos em contextos menos desenvolvidos apresentem piores condições de saúde e por isso sofram óbito prematuramente e por causas potencialmente evitáveis. Porém, é provável que o contexto socioeconômico, demográfico e cultural no qual estes indivíduos estão inseridos impeça que sejam satisfatoriamente beneficiados pelos cuidados de saúde e, nesta perspectiva, que as políticas públicas sejam mais efetivas, garantindo maior oferta e acesso a grupos populacionais com maiores necessidades. Considerando as tendências observadas, é possível ainda que os padrões de mortalidade nas regiões se invertam, com as doenças crônicas alcançando maior magnitude no Norte e Nordeste do que no Sul e Sudeste, neste cenário as desigualdades regionais se intensificariam novamente e as iniquidades sobressairiam ainda mais.

A mortalidade é um “extremo da equação” que representa o desfecho de uma trajetória de adoecimento, de maneira que padrões específicos de mortalidade podem representar desequilíbrios sistêmicos que interferem no ciclo de vida de uma população. Causas de mortalidade “agudas”, de rápida resolução como algumas doenças infecciosas e as causas externas, representam a perda estanque de anos potenciais de vida, enquanto a mortalidade por causas crônicas e degenerativas geralmente resultam de longos períodos de morbidade, gerando limitações e sequelas que acarretam anos de sobrevivência com baixa qualidade de vida, ocasionando anos de utilização recorrente dos serviços de saúde. Neste contexto, é necessário que as políticas públicas possibilitem a redução da mortalidade aguda, o que tem tido relativo sucesso no país, e oportunizem a expectativa de vida com qualidade, o que parece estar longe de se tornar realidade.

9 REFERÊNCIAS

- ABDI, H.; VALENTIN, D. Multiple correspondence analysis. **Encyclopedia of measurement and statistics**, p. 651–657, 2007.
- AGGARWAL, C. C. **Data Mining**. Cham: Springer International Publishing, 2015.
- ALVES, D. B.; BARBOSA, M. T. S. Desigualdades na mortalidade por doenças crônicas entre idosos e sua associação com indicadores socioeconômicos no Brasil. **Revista Brasileira de Ciências do Envelhecimento Humano**, v. 7, n. 1, p. 22–33, 2010.
- ALVES, D. da S. B. Análise espacial da mortalidade de idosos por doenças crônicas no município do Rio de Janeiro. 2013. Disponível em: <<http://bvsmms.saude.gov.br/bvs/publicacoes/premio2013/mestrado/Davi%20da%20Silveira%20Barroso%20Alves.pdf>>. Acesso em: 4 ago. 2017.
- AMARAL, T. C. L. Mortalidade hospitalar na rede SUS: espelho dos óbitos ocorridos na população brasileira. **Gestão de sistemas de saúde**, p. 169, 2002.
- AMORIM, V. M. S. L.; BARROS, M. B. de A.; CÉSAR, C. L. G.; GOLDBAUM, M.; CARANDINA, L.; ALVES, M. C. G. P. Fatores associados à realização dos exames de rastreamento para o câncer de próstata: um estudo de base populacional. **Cadernos de Saúde Pública**, p. 347–356, 2011.
- ANDRADE, L. T. de; DINIZ, A. M. A. Spatial reorganization of homicides in Brazil and the interiorization thesis. **Revista Brasileira de Estudos de População**, v. 30, p. S171–S191, 2013.
- ARAÚJO, J. D. de. Polarização Epidemiológica No Brasil. **Epidemiologia e Serviços de Saúde**, v. 21, n. 4, p. 533–538, dez. 2012.
- ARAÚJO, V. E. M. de; BEZERRA, J. M. T.; AMÂNCIO, F. F.; PASSOS, V. M. de A.; CARNEIRO, M. Aumento da carga de dengue no Brasil e unidades federadas, 2000 e 2015: análise do Global Burden of Disease Study 2015. **Revista Brasileira de Epidemiologia**, v. 20, n. suppl 1, p. 205–216, maio 2017.
- ARMSTRONG GL; CONN LA; PINNER RW. Trends in infectious disease mortality in the united states during the 20th century. **JAMA**, v. 281, n. 1, p. 61–66, 6 jan. 1999.
- ARRIAGA, E. E.; DAVIS, K. The Pattern of Mortality Change in Latin America. **Demography**, v. 6, n. 3, p. 223–242, ago. 1969.
- ASCHE, C. V.; SEAL, B.; KAHLER, K. H.; OEHRLEIN, E. M.; BAUMGARTNER, M. G. Evaluation of Healthcare Interventions and Big Data: Review of Associated Data Issues. **PharmacoEconomics**, v. 35, n. 8, p. 759–765, ago. 2017.
- ASHTON, D.; PORTER, S.; LIBRARY), N. D. (chart js; LIBRARY), T. L. (chart js; LIBRARY), W. E. (chart js. **radarchart: Radar Chart from “Chart.js”**. [s.l.: s.n.]

AYRES, M.; AYRES JÚNIOR, M.; AYRES, D. L.; SANTOS, A. de A. Aplicações estatísticas nas áreas das ciências bio-médicas. **Instituto Mamirauá, Belém**, 2007. Disponível em: <https://www.researchgate.net/profile/Alex_De_Assis_Dos_Santos/publication/263608962_BIOESTAT__aplicaes_estatsticas_nas_reas_das_Cincias_Bio-Mdicas/links/02e7e53b598e69ebfe000000.pdf>. Acesso em: 2 maio. 2016.

AZEVEDO E SILVA, G.; GAMARRA, C. J.; GIRIANELLI, V. R.; VALENTE, J. G.; OTHERS. Tendência da mortalidade por câncer nas capitais e interior do Brasil entre 1980 e 2006. **Revista de Saúde Pública**, v. 45, n. 6, p. 1009–1018, 2011.

BAILAR III, J. C.; GORNIK, H. L. “ N 1986, when one öfüs” report’ed öhtrenäsin”. **J Med**, v. 336, p. 1569–74, 1997.

BAKAR, A. A.; KEFLI, Z.; ABDULLAH, S.; SAHANI, M. Predictive models for dengue outbreak using multiple rulebase classifiers. In: Electrical Engineering and Informatics (ICEEI), 2011 International Conference on, **Anais...IEEE**, 2011. Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6021830>. Acesso em: 30 mar. 2016.

BANDO, D. H.; LESTER, D. An ecological study on suicide and homicide in Brazil. **Ciência & Saúde Coletiva**, v. 19, n. 4, p. 1179–1189, abr. 2014.

BARATA, R. de C. B.; TORRES, A. M.; FILHO, N. de A.; VERAS, R. P. (ed.). **Equidade e saúde: contribuições da epidemiologia**. Rio de Janeiro, RJ: Editora Fiocruz : Abrasco, 1997.

BARBOSA, I. R.; SOUZA, D. L. B. de; BERNAL, M. M.; COSTA, I. do C. C. Desigualdades regionais na mortalidade por câncer de colo de útero no Brasil: tendências e projeções até o ano 2030. **Ciência & Saúde Coletiva**, v. 21, n. 1, p. 253–262, jan. 2016.

BARLOW, M. **Data Visualization: A New Language for Storytelling**. 1. ed. United States of America: O’Reilly Media, Inc., 2015.

BARRETO, M. L.; CARMO, E. H. Padrões de adoecimento e de morte da população brasileira: os renovados desafios para o Sistema Único de Saúde. **Ciência & Saúde Coletiva**, v. 12, n. Supl 0, p. 1779–1790, 2007.

BARRETO, M. L.; TEIXEIRA, M. G.; BASTOS, F. I.; XIMENES, R. A.; BARATA, R. B.; RODRIGUES, L. C. Saúde no Brasil 3 Sucessos e fracassos no controle de doenças infecciosas no Brasil: o contexto social e ambiental, políticas, intervenções e necessidades de pesquisa. **The Lancet**, p. 47–60, 2015.

BARROS, A. J.; CASCAES, A. M.; WEHRMEISTER, F. C.; MARTÍNEZ-MESA, J.; MENEZES, A. M. B. Tobacco smoking in Brazil: regional inequalities and prevalence according to occupational characteristics. **Ciência & Saúde Coletiva**, v. 16, n. 9, p. 3707–3716, 2011.

BASSANESI, S. L.; AZAMBUJA, M. I. R.; ACHUTTI, A. C. Mortalidade precoce por doenças cardiovasculares e desigualdades sociais em Porto Alegre: da evidência à ação. **Arquivos Brasileiros de Cardiologia. Sao Paulo. Vol. 90, n. 6 (2008), p. 403-412**, 2008. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/20787>>. Acesso em: 5 jul. 2017.

BATISTA FILHO, M.; DE SOUZA, A. I.; MIGLIOLI, T. C.; DOS SANTOS, M. C. Anemia e obesidade: um paradoxo da transição nutricional brasileira Anemia and obesity: a paradox of the nutritional transition in Brazil. **Cad. saúde pública**, v. 24, n. Sup 2, p. S247–S257, 2008.

BAUMER, B. S.; KAPLAN, D. T.; HORTON, N. J. **Modern Data Science with R**. Edição: 1 ed. Boca Raton: Chapman and Hall/CRC, 2017.

BENIGUI, Y. As infecções respiratórias agudas na infância como problema de saúde pública. **Boletim de Pneumologia Sanitária**, v. 10, n. 1, p. 13–22, 2002.

BERRY, M. J. A.; LINOFF, G. **Data mining techniques for marketing, sales, and customer relationship management**. Indianapolis: Wiley, 2004.

BIERRENBACH, A. L.; DUARTE, E. C.; GOMES, A. B. F.; SOUZA, M. de F. M. de. Tendência da mortalidade por tuberculose no Brasil, 1980 a 2004. 2007. Disponível em: <<http://repositorio.unb.br/handle/10482/12914>>. Acesso em: 28 jul. 2017.

BIM, C. R.; PELLOSO, S. M.; CARVALHO, M. D. de B.; PREVIDELLI, I. T. S. Early diagnosis of breast and cervical cancer in women from the municipality of Guarapuava, PR, Brazil. **Revista da Escola de Enfermagem da USP**, v. 44, n. 4, p. 940–946, 2010.

BLAS, E.; KURUP, A. S.; WORLD HEALTH ORGANIZATION (ed.). **Equity, social determinants, and public health programmes**. Geneva, Switzerland: World Health Organization, 2010.

BORDER, P. **Big Data and Public Health**. jul. 2014. . Disponível em: <<http://www.parliament.uk/briefing-papers/POST-PN-474.pdf>>. Acesso em: 26 set. 2014.

BRANT, L. C. C.; NASCIMENTO, B. R.; PASSOS, V. M. A.; DUNCAN, B. B.; BENSENÖR, I. J. M.; MALTA, D. C.; SOUZA, M. de F. M. de; ISHITANI, L. H.; FRANÇA, E.; OLIVEIRA, M. S.; MOONEY, M.; NAGHAVI, M.; ROTH, G.; RIBEIRO, A. L. P. Variações e diferenciais da mortalidade por doença cardiovascular no Brasil e em seus estados, em 1990 e 2015: estimativas do Estudo Carga Global de Doença. **Revista Brasileira de Epidemiologia**, v. 20, n. suppl 1, p. 116–128, maio 2017.

BRITO, F. Transição demográfica e desigualdades sociais no Brasil. **Rev bras estud popul**, v. 25, n. 1, p. 5–26, 2008.

BUSATO JR., W. F. S.; ALMEIDA, G. L. Prostate cancer screening in Brazil: should it be done or not? **International braz j urol**, v. 42, n. 6, p. 1069–1080, dez. 2016.

CAMPAGNA, A. de S.; DOURADO, I.; DUARTE, E. C.; DAUFENBACH, L. Z. Mortalidade por causas relacionadas à influenza em idosos no Brasil, 1992 a 2005. 2009. Disponível em: <<http://repositorio.unb.br/handle/10482/12907>>. Acesso em: 10 maio. 2016.

CAMPOS, H. S. Mortalidade por DPOC no Brasil, 1980-1998. **Pulmão RJ**, v. 12, n. 4, p. 217–225, 2003.

CANNING, D. The Causes and Consequences of Demographic Transition. **Population Studies**, v. 65, n. 3, p. 353–361, nov. 2011.

CARVALHO, J. A. M.; RODRÍGUEZ-WONG, L. L. A transição da estrutura etária da população brasileira na primeira metade do século XXI. **Cadernos de saúde pública**, v. 24, n. 3, p. 597–605, 2008.

CARVALHO, M. S.; STRUCHINER, C. J. Análise de correspondência: uma aplicação do método à avaliação de serviços de vacinação. **Cad Saúde Pública**, v. 8, n. 3, p. 287–301, 1992.

CECCON, R. F.; MAFFACCIOLLI, R.; BURILLE, A.; MENEGHEL, S. N.; OLIVEIRA, D. L. L. C. de; GERHARDT, T. E.; CECCON, R. F.; MAFFACCIOLLI, R.; BURILLE, A.; MENEGHEL, S. N.; OLIVEIRA, D. L. L. C. de; GERHARDT, T. E. Mortalidade por tuberculose nas capitais brasileiras, 2008-2010. **Epidemiologia e Serviços de Saúde**, v. 26, n. 2, p. 349–358, mar. 2017.

CERBINO NETO, J.; WERNECK, G. L.; COSTA, C. H. N. Factors associated with the incidence of urban visceral leishmaniasis: an ecological study in Teresina, Piauí State, Brazil. **Cadernos de Saúde Pública**, v. 25, n. 7, p. 1543–1551, 2009.

CESSE, E. Â. P.; DE CARVALHO, E. F.; DE SOUZA, W. V.; LUNA, C. F. Tendência da mortalidade por doenças do aparelho circulatório no Brasil: 1950 a 2000. **cities**, v. 93, n. 5, p. 454–460, 2009.

CHAIMOWICZ, F. A saúde dos idosos brasileiros às vésperas do século XXI: problemas, projeções e alternativas. **Revista de Saúde Pública**, v. 31, n. 2, p. 184–200, 1997.

CHIAVEGATTO FILHO, A. D. P. Uso de Big Data Em Saúde No Brasil: Perspectivas Para Um Futuro Próximo. **Epidemiologia e Serviços de Saúde**, v. 24, n. 2, p. 325–332, jun. 2015.

CHIESA, A. M.; WESTPHAL, M. F.; AKERMAN, M.; OTHERS. Doenças respiratórias agudas: um estudo das desigualdades em saúde. **Cadernos de Saúde Pública**, v. 24, n. 1, p. 55–69, 2008.

CHOR, D.; FONSECA, M. de J. M. da; ANDRADE, C. R. de. Doenças cardiovasculares: comentários sobre a mortalidade precoce no Brasil. **Arq. bras. cardiol**, v. 64, n. 1, p. 15–9, jan. 1995.

CLOSS, E. V.; SCHWANKE, C. H. A. A evolução do índice de envelhecimento no Brasil, nas suas regiões e unidades federativas no período de 1970 a 2010. **Rev Bras Geriatr Gerontol**, v. 15, n. 3, p. 443–458, 2012.

COSTA, C. H. N.; TAPETY, C. M. M.; WERNECK, G. L. Control of visceral leishmaniasis in urban areas: randomized factorial intervention trial. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 40, n. 4, p. 415–419, 2007.

COSTA, M. R. da; MARCOPITO, L. F. Mortalidade por causas mal definidas, Brasil, 1979-2002, e um modelo preditivo para idade. **Cadernos de Saúde Pública**, v. 24, n. 5, p. 1001–1012, 2008.

COUTINHO, J. G.; GENTIL, P. C.; BERTOLIN, M. N. T. A desnutrição e obesidade no Brasil: o enfrentamento com base na agenda única da nutrição. 2008. Disponível em: <<http://www.scielo.org/pdf/csp/v24s2/18>>. Acesso em: 1 set. 2017.

CYRINO, C. D. **Análise Geométrica de Dados através de Análise de Correspondência Múltipla**. 2011. UFJF - Universidade Federal de Juiz de Fora, Juiz de Fora, MG, 2011. Disponível em: <<http://www.ufjf.br/cursoestatistica/files/2014/04/An%C3%A1lise-Geom%C3%A9trica-de-Dados-atrav%C3%A9s-de-An%C3%A1lise-de-Correspond%C3%Aancia-Multipla.pdf>>. Acesso em: 21 jun. 2017.

DAMÁSIO, B. F. Uso da análise fatorial exploratória em psicologia. **Avaliação psicológica**, v. 11, n. 2, p. 213–228, 2012.

DANAIEI, G.; VANDER HOORN, S.; LOPEZ, A. D.; MURRAY, C. J.; EZZATI, M.; (CANCERS, C. R. A. collaborating group; OTHERS. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. **The Lancet**, v. 366, n. 9499, p. 1784–1793, 2005.

DATASUS/MS. **Fichas de Qualificação da RIPSA - Mortalidade proporcional por idade - A.8 - 2012**. Disponível em: <<http://fichas.ripsa.org.br/2012/a-8/>>. Acesso em: 11 ago. 2017.

DOLL, R. Are we winning the fight against cancer? An epidemiological assessment: EACR—Mühlbock memorial lecture. **European Journal of Cancer and Clinical Oncology**, v. 26, n. 4, p. 500–508, 1990.

DOMINGUES, C. M. A. S.; TEIXEIRA, A. M. da S. Coberturas Vacinais e Doenças Imunopreveníveis No Brasil No Período 1982-2012: Avanços e Desafios Do Programa Nacional de Imunizações. **Epidemiologia e Serviços de Saúde**, v. 22, n. 1, p. 9–27, mar. 2013.

DONALISIO, M. R.; FRANCISCO, P. M. S. B.; LATORRE, M. do R. D. de; OTHERS. Respiratory mortality trend in the elderly before and after influenza vaccination campaigns, State of São Paulo-1980 to 2004. **Revista Brasileira de Epidemiologia**, v. 9, n. 1, p. 32–41, 2006.

DORÉ, J.-C.; OJASOO, T.; OKUBO, Y.; DURAND, T.; DUDOGNON, G.; MIQUEL, J.-F. Correspondence factor analysis of the publication patterns of 48 countries over the period 1981–1992. **Journal of the Association for Information Science and Technology**, v. 47, n. 8, p. 588–602, 1996.

DOYLE, A. C. **As aventuras de Sherlock Holmes**. [s.l.] Zahar, 1928.

DRUMOND JR, M.; LIRA, M. M.; FREITAS, M. de; NITRINI, T.; SHIBAO, K. Evaluation of the quality of mortality information by unspecified accidents and events of undetermined intent in a metropolis of Southeastern Brazil. **Revista de Saúde Pública**, v. 33, n. 3, p. 273–280, 1999.

EHRENSTEIN, V.; NIELSEN, H.; PEDERSEN, A. B.; JOHNSEN, S. P.; PEDERSEN, L. Clinical Epidemiology in the Era of Big Data: New Opportunities, Familiar Challenges. **Clinical Epidemiology**, v. Volume 9, p. 245–250, abr. 2017.

ELLIOTT, R. W. Kiviat-graphs as a means for displaying performance data for on-line retrieval systems. **Journal of the American Society for Information Science**, v. 27, n. 3, p. 178–182, 1976.

- ESTEBAN, C.; AROSTEGUI, I.; GARCIA-GUTIERREZ, S.; GONZALEZ, N.; LAFUENTE, I.; BARE, M.; FERNANDEZ DE LARREA, N.; RIVAS, F.; QUINTANA, J. M.; IRYSS-COPD. A Decision Tree to Assess Short-Term Mortality after an Emergency Department Visit for an Exacerbation of COPD: A Cohort Study. **Respiratory Research**, v. 16, n. 1, dez. 2015. Disponível em: <<http://respiratory-research.biomedcentral.com/articles/10.1186/s12931-015-0313-4>>. Acesso em: 20 ago. 2017.
- FERLAY, J.; SOERJOMATARAM, I.; DIKSHIT, R.; ESER, S.; MATHERS, C.; REBELO, M.; PARKIN, D. M.; FORMAN, D.; BRAY, F. Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012: Globocan 2012. **International Journal of Cancer**, v. 136, n. 5, p. E359–E386, 1 mar. 2015.
- FISCHER, T. K.; LIMA, D.; ROSA, R.; OSÓRIO, D.; BOING, A. F. A mortalidade infantil no Brasil: série histórica entre 1994-2004 e associação com indicadores socioeconômicos em municípios de médio e grande porte. **Medicina (Ribeirão Preto. Online)**, v. 40, n. 4, p. 559–566, 2007.
- FONSECA, L. A. M.; ELUF-NETO, J.; WUNSCH FILHO, V. Tendências Da Mortalidade Por câncer Nas Capitais Dos Estados Do Brasil, 1980-2004. **Revista da Associação Médica Brasileira**, v. 56, n. 3, p. 309–312, 2010.
- FRANÇA, E.; DE ABREU, D. X.; RAO, C.; LOPEZ, A. D. Evaluation of Cause-of-Death Statistics for Brazil, 2002–2004. **International Journal of Epidemiology**, v. 37, n. 4, p. 891–901, ago. 2008.
- FRANCISCO, P. M. S. B.; DONALISIO, M. R. C.; LATTORRE, M. do R. D. de O. Impacto da vacinação contra influenza na mortalidade por doenças respiratórias em idosos. **Revista de Saúde Pública**, v. 39, n. 1, p. 75–81, 2005.
- FRANCISCO, P. M. S. B.; DONALISIO, M. R. de C.; LATTORRE, M. do R. D. de; OTHERS. Tendência da mortalidade por doenças respiratórias em idosos do Estado de São Paulo, 1980 a 1998. **Revista de saúde pública**, 2003. Disponível em: <<http://www.scielosp.org/pdf/rsp/v37n2/15285>>. Acesso em: 2 ago. 2017.
- FRENK, J.; FREJKA, T.; BOBADILLA, J. L.; STERN, C.; LOZANO, R.; SEPÚLVEDA, J.; JOSÉ, M. La transición epidemiológica en América Latina. 1991. Disponível em: <<http://iris.paho.org/xmlui/handle/123456789/16560>>. Acesso em: 10 ago. 2017.
- GALDURÓZ, J. C. F.; CAETANO, R. Epidemiologia do uso de álcool no Brasil [Epidemiology of alcohol use in Brazil]. **Rev Bras Psiquiatr**, v. 26, n. Supl I, p. 3–6, 2004.
- GAWRYSZEWSKI, V. P.; KOIZUMI, M. S.; DE MELLO-JORGE, M. H. P. As causas externas no Brasil no ano 2000: comparando a mortalidade e a morbidade Morbidity and mortality from external causes in Brazil, 2000. **Cad. Saúde Pública**, v. 20, n. 4, p. 995–1003, 2004.
- GIRIANELLI, V. R.; GAMARRA, C. J.; AZEVEDO E SILVA, G. Disparities in cervical and breast cancer mortality in Brazil. **Revista de Saúde Pública**, v. 48, n. 3, p. 459–467, jun. 2014.
- GONTIJO, C. M. F.; MELO, M. N. Visceral leishmaniasis in Brazil: current status, challenges and prospects. **Revista Brasileira de Epidemiologia**, v. 7, n. 3, p. 338–349, 2004.

GONTIJO, E. D.; ANDRADE, G. M. Q. de; SANTOS, S. E.; GALVÃO, L. M. da C.; MOREIRA, E. F.; PINTO, F. S.; DIAS, J. C. P.; JANUÁRIO, J. N. Triagem neonatal da infecção pelo *Trypanosoma cruzi* em Minas Gerais, Brasil: transmissão congênita e mapeamento das áreas endêmicas. **Epidemiologia e Serviços de Saúde**, v. 18, n. 3, p. 243–254, 2009.

GONZAGA, C. M. R.; FREITAS-JUNIOR, R.; CURADO, M.-P.; SOUSA, A.-L. L.; SOUZA-NETO, J.-A.; SOUZA, M. R. Temporal Trends in Female Breast Cancer Mortality in Brazil and Correlations with Social Inequalities: Ecological Time-Series Study. **BMC Public Health**, v. 15, n. 1, dez. 2015. Disponível em: <<http://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-015-1445-7>>. Acesso em: 23 jul. 2017.

GREENACRE, M. J. Correspondence analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 2, n. 5, p. 613–619, 2010.

GUERRA, M. R.; BUSTAMANTE-TEIXEIRA, M. T.; CORRÊA, C. S. L.; ABREU, D. M. X. de; CURADO, M. P.; MOONEY, M.; NAGHAVI, M.; TEIXEIRA, R.; FRANÇA, E. B.; MALTA, D. C. Magnitude e variação da carga da mortalidade por câncer no Brasil e Unidades da Federação, 1990 e 2015. **Revista Brasileira de Epidemiologia**, v. 20, n. suppl 1, p. 102–115, maio 2017.

GUIMARÃES, M. D. C.; CARNEIRO, M.; ABREU, D. M. X. de; FRANÇA, E. B. Mortalidade por HIV/Aids no Brasil, 2000-2015: motivos para preocupação? **Revista Brasileira de Epidemiologia**, v. 20, n. suppl 1, p. 182–190, maio 2017.

GUIMARÃES, Z. A.; COSTA, M. da C. N.; PAIM, J. S.; SILVA, L. M. V. da. Declínio e desigualdades sociais na mortalidade infantil por diarreia. 2001. Disponível em: <<http://www.repositorio.ufba.br/ri/handle/ri/3352>>. Acesso em: 2 set. 2017.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise multivariada de dados - 6ed.** [s.l.] Bookman Editora, 2009.

HANSEN, V.; OREN, E.; DENNIS, L. K.; BROWN, H. E. Infectious Disease Mortality Trends in the United States, 1980-2014. **Jama**, v. 316, n. 20, p. 2149–2151, 2016.

HILL, K. Making deaths count. **Bulletin of the World Health Organization**, v. 84, n. 3, p. 162–162, 2006.

HOLCMAN, M. M.; SANTOS, J. L. F.; OTHERS. Evolução da mortalidade infantil na região metropolitana de São Paulo, 1980-2000. **Revista de Saúde Pública**, v. 38, n. 2, p. 180–186, 2004.

IBGE. **População presente e residente**. Disponível em: <http://seriesestatisticas.ibge.gov.br/lista_tema.aspx?op=0&no=10>. Acesso em: 2 mar. 2016.

IBGE. **Esperança de Vida ao Nascer**. Disponível em: <<http://seriesestatisticas.ibge.gov.br/series.aspx?no=10&op=0&vcodigo=POP210&t=esperanca-vida-nascer>>. Acesso em: 10 ago. 2017.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Divisão do Brasil em Mesorregiões e Microrregiões Geográficas**. [s.l.: s.n.]v. 1

INFANTOSI, A. F. C.; COSTA, J. C. da G. D.; ALMEIDA, R. M. V. R. de. Análise de Correspondência: Bases Teóricas Na Interpretação de Dados Categóricos Em Ciências Da Saúde. **Cadernos de Saúde Pública**, v. 30, n. 3, p. 473–486, mar. 2014.

ISER, B. P. M.; DE CARVALHO YOKOTA, R. T.; DE SÁ, N. N. B.; DE MOURA, L.; MALTA, D. C. Prevalência de fatores de risco e proteção para doenças crônicas nas capitais do Brasil-principais resultados do Vigitel 2010. **Ciência & Saúde Coletiva**, v. 17, n. 9, p. 2343–2356, 2012.

ISHITANI, L. H.; FRANCO, G. da C.; PERPÉTUO, I. H. O.; FRANÇA, E. Desigualdade social e mortalidade precoce por doenças cardiovasculares no Brasil. **Rev Saúde Pública**, v. 40, n. 4, p. 684–91, 2006.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning**. New York, NY: Springer New York, 2013. v. 103

JEREZ-ROIG, J.; SOUZA, D. L. B.; MEDEIROS, P. F. M.; BARBOSA, I. R.; CURADO, M. P.; COSTA, I. C. C.; LIMA, K. C. Future burden of prostate cancer mortality in Brazil: a population-based study. **Cadernos de Saúde Pública**, v. 30, n. 11, p. 2451–2458, nov. 2014.

JOHANSSON, L. A. **Targeting non-obvious errors in death certificates**. 2008. Acta Universitatis Upsaliensis : Universitetsbiblioteket [distributör], Uppsala, 2008.

JORGE, M. H. P. de M.; LAURENTI, R.; GOTLIEB, S. L. D. **Análise da qualidade das estatísticas vitais brasileiras: a experiência de implantação do SIM e do SINASC**. Disponível em: <<http://www.redalyc.org/articulo.oa?id=63012314>>. Acesso em: 24 fev. 2016.

KAMANGAR, F.; DORES, G. M.; ANDERSON, W. F. Patterns of Cancer Incidence, Mortality, and Prevalence Across Five Continents: Defining Priorities to Reduce Cancer Disparities in Different Geographic Regions of the World. **Journal of Clinical Oncology**, v. 24, n. 14, p. 2137–2150, 10 maio 2006.

KASSAMBARA, A.; MUNDT, F. **factoextra: Extract and Visualize the Results of Multivariate Data Analyses**. [s.l.: s.n.]

KIRK, D. Demographic transition theory. **Population studies**, v. 50, n. 3, p. 361–387, 1996.

KUSIAK, A.; DIXON, B.; SHAH, S. Predicting Survival Time for Kidney Dialysis Patients: A Data Mining Approach. **Computers in Biology and Medicine**, v. 35, n. 4, p. 311–327, maio 2005.

LAURENTI, R.; MELLO JORGE, M. H. P. de; GOTLIEB, S. L. D. Perfil epidemiológico da morbi-mortalidade masculina. **Ciênc saúde coletiva**, v. 10, n. 1, p. 35–46, 2005.

LAURENTI, R.; MELLO JORGE, M. H. P.; GOTLIEB, S. L. D. A confiabilidade dos dados de mortalidade e morbidade por doenças crônicas não-transmissíveis. **Ciência & Saúde Coletiva**, v. 9, n. 4, 2004a. Disponível em: <<http://www.redalyc.org/html/630/63043001012/>>. Acesso em: 29 ago. 2017.

LAURENTI, R.; MELLO JORGE, M. H. P.; GOTLIEB, S. L. D. A confiabilidade dos dados de mortalidade e morbidade por doenças crônicas não-transmissíveis. **Ciênc saúde coletiva**, v. 9, n. 4, p. 909–20, 2004b.

LÊ, S.; JOSSE, J.; HUSSON, F. FactoMineR: A Package for Multivariate Analysis. **Journal of Statistical Software**, v. 25, n. 1, p. 1–18, 2008.

LEBRÃO, M. L. O envelhecimento no Brasil: aspectos da transição demográfica e epidemiológica. **Saúde Coletiva**, v. 4, n. 17, p. 135–40, 2007.

LEE, R. The Demographic Transition: Three Centuries of Fundamental Change. **INTERNATIONAL LIBRARY OF CRITICAL WRITINGS IN ECONOMICS**, v. 202, p. 115, 2007.

LEITE, I. da C.; VALENTE, J. G.; SCHRAMM, J. M. de A.; DAUMAS, R. P.; RODRIGUES, R. do N.; SANTOS, M. de F.; OLIVEIRA, A. F. de; SILVA, R. S. da; CAMPOS, M. R.; MOTA, J. C. da. Carga de doença no Brasil e suas regiões, 2008. **Cadernos de Saúde Pública**, v. 31, n. 7, p. 1551–1564, jul. 2015.

LIMA, M. da S.; MARTINS-MELO, F. R.; HEUKELBACH, J.; ALENCAR, C. H.; BOIGNY, R. N.; RAMOS JÚNIOR, A. N. Mortality related to tuberculosis-HIV/AIDS co-infection in Brazil, 2000-2011: epidemiological patterns and time trends. **Cadernos de Saúde Pública**, v. 32, n. 10, out. 2016. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2016001005008&lng=en&tlng=en>. Acesso em: 28 jul. 2017.

LIMA, M. M.; SARQUIS, O.; OLIVEIRA, T. G. de; GOMES, T. F.; COUTINHO, C.; DAFLON-TEIXEIRA, N. F.; TOMA, H. K.; BRITTO, C.; TEIXEIRA, B. R.; D'ANDREA, P. S.; JANSEN, A. M.; BÓIA, M. N.; CARVALHO-COSTA, F. A. Investigation of Chagas Disease in Four Periurban Areas in Northeastern Brazil: Epidemiologic Survey in Man, Vectors, Non-Human Hosts and Reservoirs. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 106, n. 3, p. 143–149, mar. 2012.

LIMA-COSTA, M. F.; PEIXOTO, S. V.; GIATTI, L. Tendências da mortalidade entre idosos brasileiros (1980-2000). **Epidemiologia e serviços de saúde**, v. 13, n. 4, p. 217–228, 2004.

LLERAS-MUNEY, A.; CUTLER, D. M.; DEATON, A. S. **The Determinants of Mortality**. [s.l.] National Bureau of Economic Research, 2006.

LOLIO, C. A. de; LAURENTI, R. Tendência da mortalidade por doenças cerebrovasculares em adultos maiores de 20 anos de idade no município de São Paulo (Brasil), 1950 a 1981. **Rev. Saúde públ**, p. 343–6, 1986.

LÓPEZ-CUADRADO, T.; LLÁCER, A.; PALMERA-SUÁREZ, R.; GÓMEZ-BARROSO, D.; SAVULESCU, C.; GONZÁLEZ-YUSTE, P.; FERNÁNDEZ-CUENCA, R. Trends in Infectious Disease Mortality Rates, Spain, 1980–2011. **Emerging Infectious Diseases**, v. 20, n. 5, p. 782–789, maio 2014.

LOTUFO, P. A. Mortalidade precoce por doenças do coração no Brasil. Comparação com outros países. **Arq Bras Cardiol**, v. 70, n. 5, p. 321–5, 1998.

MACEDO, S. E. C.; MENEZES, A. M. B.; ALBERNAZ, E.; POST, P.; KNORST, M. Fatores de risco para internação por doença respiratória aguda em crianças até um ano de idade. **Revista de Saúde Pública**, v. 41, n. 3, p. 351–358, 2007.

MACINKO, J.; LIMA-COSTA, M. F. Horizontal equity in health care utilization in Brazil, 1998–2008. **International Journal for Equity in Health**, v. 11, n. 1, p. 33, 2012.

MACKENBACH, J. P.; STIRBU, I.; ROSKAM, A.-J. R.; SCHAAP, M. M.; MENVIELLE, G.; LEINSALU, M.; KUNST, A. E. Socioeconomic inequalities in health in 22 European countries. **New England Journal of Medicine**, v. 358, n. 23, p. 2468–2481, 2008.

MAIA-ELKHOURY, A. N. S.; ALVES, W. A.; SOUSA-GOMES, M. L. de; SENA, J. M. de; LUNA, E. A. Visceral leishmaniasis in Brazil: trends and challenges. **Cadernos de Saúde Pública**, v. 24, n. 12, p. 2941–2947, 2008.

MAILUND, T. **Beginning Data Science in R**. Berkeley, CA: Apress, 2017.

MALTA, D. C.; DUARTE, E. C.; ALMEIDA, M. F. de; DIAS, M. A. de S.; MORAIS NETO, O. L. de; MOURA, L. de; FERRAZ, W.; SOUZA, M. de F. M. de. Lista de causas de mortes evitáveis por intervenções do Sistema Único de Saúde do Brasil. 2007a. Disponível em: <<http://repositorio.unb.br/handle/10482/12913>>. Acesso em: 6 maio. 2016.

MALTA, D. C.; DUARTE, E. C.; ESCALANTE, J. J. C.; ALMEIDA, M. F. de; SARDINHA, L. M. V.; MACÁRIO, E. M.; MONTEIRO, R. A.; MORAIS NETO, O. L. de. Mortes evitáveis em menores de um ano, Brasil, 1997 a 2006: contribuições para a avaliação de desempenho do Sistema Único de Saúde. 2010. Disponível em: <<http://repositorio.unb.br/handle/10482/12902>>. Acesso em: 25 maio. 2017.

MALTA, D. C.; FRANÇA, E.; ABREU, D. M. X.; PERILLO, R. D.; SALMEN, M. C.; TEIXEIRA, R. A.; PASSOS, V.; SOUZA, M. de F. M.; MOONEY, M.; NAGHAVI, M. Mortality due to noncommunicable diseases in Brazil, 1990 to 2015, according to estimates from the Global Burden of Disease study. **Sao Paulo Medical Journal**, v. 135, n. 3, p. 213–221, jun. 2017.

MALTA, D. C.; ISER, B. P.; SÁ, N. N.; OTHERS. Tendências temporais no consumo de tabaco nas capitais brasileiras, segundo dados do VIGITEL, 2006 a 2011 [Trends in tobacco consumption from 2006 to 2011 in Brazilian capitals according to the VIGITEL survey]. **Cad Saude Publica**, v. 29, n. 4, p. 812–22, 2013.

MALTA, D. C.; MOURA, L. de; SOUZA, M. de F. M. de; CURADO, M. P.; ALENCAR, A. P.; ALENCAR, G. P. Lung cancer, cancer of the trachea, and bronchial cancer: mortality trends in Brazil, 1980-2003. **Jornal Brasileiro de Pneumologia**, v. 33, n. 5, p. 536–543, 2007b.

MALTA, D. C.; OLIVEIRA, M. R. de; MOURA, E. C. de; ARAÚJO SILVA, S.; SOARES ZOUAIN, C.; PEREIRA DOS SANTOS, F.; LIBANIO DE MORAIS NETO, O.; OLIVEIRA PENNA, G. de. Fatores de risco e proteção para doenças crônicas não transmissíveis entre beneficiários da saúde suplementar: resultados do inquérito telefônico Vigitel, Brasil, 2008. **Ciência & Saúde Coletiva**, v. 16, n. 3, 2011. Disponível em: <<http://www.redalyc.org/html/630/63018467035/>>. Acesso em: 11 jul. 2017.

MANSUR, A. de P.; LOPES, A. I. A.; FAVARATO, D.; AVAKIAN, S. D.; CÉSAR, L. A. M.; RAMIRES, J. A. F.; OTHERS. Transição epidemiológica da mortalidade por doenças circulatórias no Brasil. **Arquivos Brasileiros de Cardiologia**, v. 93, n. 5, p. 506–510, 2009.

MARTINS-MELO, F. R.; RAMOS, A. N.; ALENCAR, C. H.; HEUKELBACH, J. Prevalence of Chagas Disease in Brazil: A Systematic Review and Meta-Analysis. **Acta Tropica**, v. 130, p. 167–174, fev. 2014.

MASSAD, E. The Elimination of Chagas' Disease from Brazil. **Epidemiology and Infection**, v. 136, n. 09, set. 2008. Disponível em: <http://www.journals.cambridge.org/abstract_S0950268807009879>. Acesso em: 2 set. 2017.

MELLO JORGE, M. H. P. de; LAURENTI, R.; LIMA-COSTA, M. F.; GOTLIEB, S. L. D.; CHIAVEGATTO FILHO, A. D. P.; OTHERS. A mortalidade de idosos no Brasil: a questão das causas mal definidas. **Epidemiologia e Serviços de Saúde**, v. 17, n. 4, p. 271–281, 2008.

MELLO JORGE, M. de; LAURENTI, R.; GOTLIEB, S. L. D. Análise da qualidade das estatísticas vitais brasileiras: a experiência de implantação do SIM e do SINASC. **Ciênc Saúde Coletiva**, v. 12, n. 3, p. 643–54, 2007.

MELO, A. P. S.; FRANÇA, E. B.; MALTA, D. C.; GARCIA, L. P.; MOONEY, M.; NAGHAVI, M. Mortalidade por cirrose, câncer hepático e transtornos devidos ao uso de álcool: Carga Global de Doenças no Brasil, 1990 e 2015. **Revista Brasileira de Epidemiologia**, v. 20, n. suppl 1, p. 61–74, maio 2017.

MELO, E. C. P. **Infarto agudo do miocárdio no Município do Rio de Janeiro: qualidade dos dados, sobrevida e distribuição espacial**. 2004. Disponível em: <<http://www.arca.fiocruz.br/handle/icict/4566>>. Acesso em: 12 set. 2017.

MELO, E. C. P.; DE OLIVEIRA, E. X. G.; CHOR, D.; CARVALHO, M. S.; PINHEIRO, R. S. Inequalities in Socioeconomic Status and Race and the Odds of Undergoing a Mammogram in Brazil. **International Journal for Equity in Health**, v. 15, n. 1, dez. 2016. Disponível em: <<http://equityhealthj.biomedcentral.com/articles/10.1186/s12939-016-0435-4>>. Acesso em: 7 nov. 2016.

MENDES, A. C. G.; SÁ, D. A.; MIRANDA, G. M. D.; LYRA, T. M.; TAVARES, R. A. W. Assistência pública de saúde no contexto da transição demográfica brasileira: exigências atuais e futuras. **Cad. saúde pública**, v. 28, n. 5, p. 955–964, 2012.

MENDES, E. V. As redes de atenção à saúde. **Brasília: Organização Pan-Americana da Saúde**, v. 549, 2011. Disponível em: <<http://www.scielo.br/pdf/csc/v15n5/v15n5a05>>. Acesso em: 10 mar. 2016.

MILBORROW, S. **rpart.plot: Plot “rpart” Models: An Enhanced Version of “plot.rpart”**. [s.l: s.n.]

MINISTÉRIO DA SAÚDE. **VIGITEL BRASIL 2006 - VIGILÂNCIA DE FATORES DE RISCO E PROTEÇÃO PARA DOENÇAS CRÔNICAS POR INQUÉRITO TELEFÔNICO**. [s.l: s.n.]. Disponível em: <http://bvsmms.saude.gov.br/bvs/publicacoes/relatorio_vigitel_2006_marco_2007.pdf>. Acesso em: 12 jul. 2017.

MINISTÉRIO DA SAÚDE. **A experiência brasileira em sistemas de informação em saúde**. Brasília, DF: Editora MS, 2009.

MINISTÉRIO DA SAÚDE. **O Sistema de Informações de Mortalidade**, 24 ago. 2011. .

MINISTÉRIO DA SAÚDE. **Sistema de Informações sobre Mortalidade - SIM: Consolidação da base de dados de 2011.** [s.l.: s.n.]. Disponível em: <http://tabnet.datasus.gov.br/cgi/sim/Consolida_Sim_2011.pdf>. Acesso em: 9 mar. 2016.

MINISTÉRIO DA SAÚDE. **VIGITEL BRASIL 2015 - VIGILÂNCIA DE FATORES DE RISCO E PROTEÇÃO PARA DOENÇAS CRÔNICAS POR INQUÉRITO TELEFÔNICO.** [s.l.: s.n.]. Disponível em: <http://www.ans.gov.br/images/stories/Materiais_para_pesquisa/Materiais_por_assunto/2015_vigitel.pdf>. Acesso em: 12 jul. 2017.

MITCHELL, T. M. **Machine Learning.** 1st edition ed. [s.l.] McGraw-Hill, 1997.

MONCAYO, Á.; SILVEIRA, A. C. Current epidemiological trends for Chagas disease in Latin America and future challenges in epidemiology, surveillance and health policy. **Memórias do Instituto Oswaldo Cruz**, v. 104, p. 17–30, 2009.

MONTEIRO, C. A. Contribuição para o estudo do significado da evolução do coeficiente de mortalidade infantil no município de São Paulo, SP (Brasil) nas três últimas décadas (1950-1979). **Revista de Saúde Pública**, v. 16, n. 1, p. 7–18, 1982.

MONTEIRO, C. A.; CONDE, W. L.; POPKIN, B. M. Income-specific trends in obesity in Brazil: 1975–2003. **American journal of public health**, v. 97, n. 10, p. 1808–1812, 2007.

MORAES-SOUZA, H.; FERREIRA-SILVA, M. M. Control of transfusional transmission. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 44, p. 64–67, 2011.

MOURA, E. C.; MALTA, D. C.; MORAIS NETO, O. L. de; MONTEIRO, C. A. Prevalence and social distribution of risk factors for chronic noncommunicable diseases in Brazil. **Revista Panamericana de Salud Pública**, v. 26, n. 1, p. 17–22, 2009.

MUCHAI, E.; ODONGO, L. Comparison of crisp and fuzzy classification trees using gini index impurity measure on simulated data. **European Scientific Journal**, v. 10, n. 18, 2014. Disponível em: <<http://search.proquest.com/openview/b0a04e2d00dce6a0743e5f794511be84/1?pq-origsite=gscholar&cbl=1576353>>. Acesso em: 19 jun. 2017.

MULLER-NORDHORN, J.; BINTING, S.; ROLL, S.; WILLICH, S. N. An Update on Regional Variation in Cardiovascular Mortality within Europe. **European Heart Journal**, v. 29, n. 10, p. 1316–1326, 7 dez. 2007.

NASCIMENTO, S. G. do; OLIVEIRA, C. M. de; SPOSITO, V.; FERREIRA, D. K. da S.; BONFIM, C. V. do. Infant Mortality Due to Avoidable Causes in a City in Northeastern Brazil. **Revista Brasileira de Enfermagem**, v. 67, n. 2, 2014. Disponível em: <<http://www.gnresearch.org/doi/10.5935/0034-7167.20140027>>. Acesso em: 10 maio. 2016.

NERLOVE, M. Household and economy: toward a new theory of population and economic growth. In: **Marriage, Family, Human Capital, and Fertility.** [s.l.] Journal of Political Economy 82 (2), Part II, 1974. p. 200–221.

NOTESTEIN, F. W. **Economic problems of population change.** [s.l.] Oxford University Press London, 1953.

NOVAES, D. H. M.; BRAGA, P. E.; SCHOUT, D. Fatores associados à realização de exames preventivos para câncer nas mulheres brasileiras, PNAD 2003. **Ciência & Saúde Coletiva**, v. 11, n. 4, 2006. Disponível em: <<http://www.redalyc.org/html/630/63011423/>>. Acesso em: 28 jul. 2017.

OLA, O.; SEDIG, K. The Challenge of Big Data in Public Health: An Opportunity for Visual Analytics. **Online Journal of Public Health InformationsOJPHI**, 3. v. 5, n. 223, p. 21, 2014.

OLIVEIRA, J. de F. M. de; BOING, A. F.; WALDMAN, E. A.; ANTUNES, J. L. F. Ecological study on mortality from influenza and pneumonia before and after influenza vaccination in the Northeast and South of Brazil. **Cadernos de Saúde Pública**, v. 29, n. 12, p. 2535–2545, dez. 2013.

OMRAN, A. R. The epidemiologic transition: a theory of the epidemiology of population change. **The Milbank Memorial Fund Quarterly**, v. 49, n. 4, p. 509–538, 1971.

OSTERMAYER, A. L.; PASSOS, A. D. C.; SILVEIRA, A. C.; FERREIRA, A. W.; MACEDO, V.; PRATA, A. R. The national survey of seroprevalence for evaluation of the control of Chagas disease in Brazil (2001–2008). **Revista da Sociedade Brasileira de Medicina Tropical**, v. 44, p. 108–121, 2011.

PAES, N. A. A geografia da mortalidade por causas dos idosos no Brasil. **Anais**, p. 1–22, 2016.

PAES, N. A.; SILVA, L. A. A. Infectious and parasitic diseases in Brazil: a decade of transition. **Revista Panamericana de Salud Pública**, v. 6, n. 2, p. 99–109, 1999.

PAIM, J.; TRAVASSOS, C.; ALMEIDA, C.; BAHIA, L.; MACINKO, J. The Brazilian health system: history, advances, and challenges. **The Lancet**, v. 377, n. 9779, p. 1778–1797, 2011.

PAOIN, W. Lessons Learned from Data Mining of WHO Mortality Database: **Methods of Information in Medicine**, v. 50, n. 4, p. 380–385, 21 jun. 2011.

PATEL, R. B.; MATHUR, M. B.; GOULD, M.; UYEKI, T. M.; BHATTACHARYA, J.; XIAO, Y.; KHAZENI, N. Demographic and Clinical Predictors of Mortality from Highly Pathogenic Avian Influenza A (H5N1) Virus Infection: CART Analysis of International Cases. **PLoS ONE**, v. 9, n. 3, p. e91630, 25 mar. 2014.

PENTLAND, A.; REID, T. G.; HEIDBECK, T. **Revolutionizing and Transforming Health Science Using Big Data**: Revolutionizing medicine and Public Health Report of the Big Data and Health Working Group 2013. [s.l.: s.n.]. Disponível em: <<http://ijaem.org/papers/Vol.1-Issue1/4AS0119.pdf>>. Acesso em: 28 set. 2014.

PEREIRA, D. V.; MOTA, C. M.; ANDRESEN, M. A. Social disorganization and homicide in Recife, Brazil. **International journal of offender therapy and comparative criminology**, p. 0306624X15623282, 2015.

PERES, M. F. T.; SANTOS, P. C. Mortalidade por homicídios no Brasil na década de 90: o papel das armas de fogo. **Revista de Saúde Pública**, v. 39, n. 1, p. 58–66, 2005.

- PESSÔA, C. L. C.; PESSÔA, R. S. Epidemiologia da DPOC no presente—aspectos nacionais e internacionais. **Pulmão RJ**, v. 1, n. 1, p. 7–12, 2009.
- PETRUZALEK, D. **read.dbc: Read Data Stored in DBC (Compressed DBF) Files**. [s.l.: s.n.]
- PRATA, P. R. A transição epidemiológica no Brasil. **Cadernos de Saúde Pública**, v. 8, n. 2, p. 168–175, 1992.
- PRESTON, S. H. The Changing Relation between Mortality and Level of Economic Development. **Population Studies**, v. 29, n. 2, p. 231, jul. 1975.
- PRESTON, S. H. The Changing Relation between Mortality and Level of Economic Development. **International Journal of Epidemiology**, v. 36, n. 3, p. 484–490, 1 jun. 2007.
- PRIETSCH, S. O.; FISCHER, G. B.; CÉSAR, J. A.; LEMPEK, B. S.; BARBOSA, L. V.; ZOGBI, L.; OTHERS. Doença respiratória em menores de 5 anos no sul do Brasil: influência do ambiente doméstico. **Rev Panam Salud Publica**, v. 13, n. 5, p. 303–10, 2003.
- PUSKARICH, M. A Decision Tree Incorporating Biomarkers and Patient Characteristics Estimates Mortality Risk for Adults with Septic Shock. **Evidence-Based Nursing**, v. 18, n. 2, p. 42–42, 18 mar. 2015.
- QUEIROZ, L. L. C.; MONTEIRO, S. G.; MOCHEL, E. G.; VERAS, M. A. de S. M.; SOUSA, F. G. M. de; BEZERRA, M. L. de M.; CHEIN, M. B. da C. Coverage of the basic immunization schedule in the first year of life in State capitals in Northeast Brazil. **Cadernos de saude publica**, v. 29, n. 2, p. 294–302, 2013.
- RAMASUBRAMANIAN, K.; SINGH, A. **Machine Learning Using R**. Berkeley, CA: Apress, 2017.
- REICHENHEIM, M. E.; SOUZA, E. R. de; MORAES, C. L.; JORGE, M. H. P. de M.; SILVA, C. M. F. P. da; MINAYO, M. C. de S. Saúde no Brasil 5: Violência e lesões no Brasil: efeitos, avanços alcançados e desafios futuros. **The Lancet**, p. 75–89, 2015.
- RODRIGUES, C. M.; SILVA, J. K. T.; MARTINS, P. C.; TERRON, S. L.; DE SOUZA SILVA, R. L. **Divisão Territorial Brasileira**. [s.l.: s.n.]
- RODRIGUES, F. E.; TATTO, R. B.; VAUCHINSKI, L.; LE&ATILDE;ES, L. M.; RODRIGUES, M. M.; RODRIGUES, V. B.; CATHARINO, A.; CAINELLI, M.; PRATES, G. P.; CERQUEIRA, T. M.; ZHANG, L. Pneumonia mortality in Brazilian children younger than or equal to 4 years. **Jornal de Pediatria**, v. 0, n. 0, 17 jan. 2011. Disponível em: <http://jped.com.br/conteudo/Ing_resumo.asp?varArtigo=2160&cod=&idSecao=1>. Acesso em: 3 ago. 2017.
- ROKACH, L.; MAIMON, O. Decision trees. **Data mining and knowledge discovery handbook**, p. 165–192, 2005.
- ROSENBERG, H. M. Cause of Death as a Contemporary Problem. **Journal of the History of Medicine and Allied Sciences**, v. 54, n. 2, p. 133–153, abr. 1999.

ROTH, G. A.; FOROUZANFAR, M. H.; MORAN, A. E.; BARBER, R.; NGUYEN, G.; FEIGIN, V. L.; NAGHAVI, M.; MENSAH, G. A.; MURRAY, C. J. L. Demographic and Epidemiologic Drivers of Global Cardiovascular Mortality. **New England Journal of Medicine**, v. 372, n. 14, p. 1333–1341, 2 abr. 2015.

SANS, S.; KESTELOOT, H.; KROMHOUT, D. The Burden of Cardiovascular Diseases Mortality in Europe. Task Force of the European Society of Cardiology on Cardiovascular Mortality and Morbidity Statistics in Europe. **European Heart Journal**, v. 18, n. 8, p. 1231–1248, ago. 1997.

SANTIAGO, L. M.; LIMA LUZ, L.; SANTOS DA SILVA, J. F.; ECHENIQUE MATTOS, I. Prevalência e fatores associados à realização de exames de rastreamento para câncer de próstata em idosos de Juiz de Fora, MG, Brasil. **Ciência & Saúde Coletiva**, v. 18, n. 12, 2013. Disponível em: <<http://www.redalyc.org/html/630/63028840010/>>. Acesso em: 3 set. 2017.

SANTOS, K. M. **Distribuição espacial da mortalidade por câncer de colo de útero no Brasil, 1996 a 2009: Eco-cuidado de Enfermagem**. 2011. UNIRIO - Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, RJ, 2011. Disponível em: <<http://www2.unirio.br/unirio/ccbs/ppgenf/arquivos/dissertacoes-arquivo/dissertacoes-2011/keitt-martins-santos>>. Acesso em: 12 set. 2017.

SANTOS, R. de S.; MELO, E. C. P. Mortalidade e assistência oncológica no Rio de Janeiro: câncer de mama e colo uterino. **Escola Anna Nery Revista de Enfermagem**, v. 15, n. 2, p. 410–416, 2011.

SCHMIDT, M. I.; DUNCAN, B. B.; SILVA, G. A.; MENEZES, A. M.; MONTEIRO, C. A.; BARRETO, S. M. Doenças crônicas não-transmissíveis no Brasil: carga e desafios atuais. 2011. Disponível em: <<http://dms.ufpel.edu.br/ares/bitstream/handle/123456789/222/1%20%202011%20Doen%20E7as%20cr%20F4nicas%20n%20E3o%20transmiss%20EDveis%20no%20Brasil.pdf?sequence=1>>. Acesso em: 11 jul. 2017.

SCHRAMM, J. M. de A.; OLIVEIRA, A. F. de; LEITE, I. da C.; VALENTE, J. G.; GADELHA, Â. M. J.; PORTELA, M. C.; CAMPOS, M. R. Epidemiological transition and the study of burden of disease in Brazil. **Ciência & Saúde Coletiva**, v. 9, n. 4, p. 897, 2004.

SIM. **Mortalidade - Brasil - 1979 a 1995**. Disponível em: <<http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sim/cnv/obt09uf.def>>. Acesso em: 11 set. 2017a.

SIM. **Mortalidade - Brasil - 1996 a 2015**. Disponível em: <<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/obt10uf.def>>. Acesso em: 13 ago. 2017b.

SIM. **Mortalidade - Brasil**. Disponível em: <<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/obt10uf.def>>. Acesso em: 10 ago. 2017c.

SINASC. **Nascidos vivos - Brasil**. Disponível em: <<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinasc/cnv/nvuf.def>>. Acesso em: 2 set. 2017.

SOARES, G. P.; BRUM, J. D.; OLIVEIRA, G. M. M. de; KLEIN, C. H.; SILVA, N. A. S. e. Evolution of Socioeconomic Indicators and Cardiovascular Mortality in Three Brazilian States. **Arquivos Brasileiros de Cardiologia**, v. 100, n. 2, p. 147–156, 2013.

SOARES, S. R.; BERNARDES, R. S.; NETTO, O. de M. C. Relações entre saneamento, saúde pública e meio ambiente: elementos para formulação de um modelo de planejamento em saneamento Relationship between water supply, sanitation, public health, and environment: elements. **Cad. Saúde Pública**, v. 18, n. 6, p. 1713–1724, 2002.

SOUSA, G. dos S. de; MAGALHÃES, F. B.; GAMA, I. da S.; LIMA, M. V. N. de; ALMEIDA, R. L. F. de; VIEIRA, L. J. E. de S.; BEZERRA FILHO, J. G. Social determinants and their interference in homicide rates in a city in northeastern Brazil. **Revista Brasileira de Epidemiologia**, v. 17, n. suppl 2, p. 194–203, 2014.

SOUZA, A. C. **Análise de Correspondência Aplicada à ECINF: a diversidade do setor informal urbano no Brasil**. 2004. ENCE - Escola Nacional de Ciências Estatísticas, Rio de Janeiro, RJ, 2004. Disponível em: <http://www.ence.ibge.gov.br/images/ence/doc/mestrado/dissertacoes/2003/2004DISSERTACAO-Augusto_Souza.pdf>. Acesso em: 21 jun. 2017.

SOUZA, E. R. de; LIMA, M. L. C. de. The panorama of urban violence in Brazil and its capitals. **Ciência & Saúde Coletiva**, v. 11, p. 1211–1222, 2006.

SOUZA, M. de F. M. de; ALENCAR, A. P.; MALTA, D. C.; MOURA, L.; MANSUR, A. de P. Análise de séries temporais da mortalidade por doenças isquêmicas do coração e cerebrovasculares, nas cinco regiões do Brasil, no período de 1981 a 2001. **Arq Bras Cardiol**, v. 87, n. 6, p. 735–40, 2006.

SOUZA, M. F. M.; TIMERMAN, A.; SERRANO JR, C. V.; SANTOS, R. D.; MANSUR, A. P. Tendências do risco de morte por doenças circulatórias nas cinco regiões do Brasil no período de 1979 a 1996. **Arq Bras Cardiol**, v. 77, n. 6, p. 562–75, 2001.

SVS/MS. **Evolução da Mortalidade no Brasil**. [s.l.: s.n.]

SZWARCWALD, C. L.; DO CARMO LEAL, M.; DE CASTILHO, E. A.; DE ANDRADE, C. L. T. Mortalidade infantil no Brasil: Belíndia ou Bulgária? Infant mortality rate in Brazil: Belgium-India or Bulgaria? **Cad. Saúde Públ**, v. 13, n. 3, p. 503–516, 1997.

SZWARCWALD, C. L.; SOUZA JÚNIOR, P. R. B. de; MARQUES, A. P.; ALMEIDA, W. da S. de; MONTILLA, D. E. R. Inequalities in Healthy Life Expectancy by Brazilian Geographic Regions: Findings from the National Health Survey, 2013. **International Journal for Equity in Health**, v. 15, n. 1, dez. 2016. Disponível em: <<http://equityhealthj.biomedcentral.com/articles/10.1186/s12939-016-0432-7>>. Acesso em: 1 ago. 2017.

TAN, P.-N. **Introduction to Data Mining**. 2 edition ed. Boston, Mass; London: Addison-Wesley, 2013.

TAYLOR, R. A.; PARE, J. R.; VENKATESH, A. K.; MOWAFI, H.; MELNICK, E. R.; FLEISCHMAN, W.; HALL, M. K. Prediction of In-Hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach.

Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine, v. 23, n. 3, p. 269–278, mar. 2016.

TEITELBAUM, M. S. Relevance of Demographic Transition Theory for Developing Countries. **Science**, New Series. v. 188, n. 4187, p. 420–425, 5 fev. 1975.

TEIXEIRA, M. da G. L. C.; MEYER, M. de A.; COSTA, M. da C. N.; PAIM, J. S.; SILVA, L. M. V. da. Mortalidade por doenças infecciosas e parasitárias em Salvador-Bahia: evolução e diferenciais intra-urbanos segundo condições de vida. 2002. Disponível em: <<https://repositorio.ufba.br/ri/handle/ri/3357>>. Acesso em: 2 set. 2017.

THERNEAU, T.; ATKINSON, B.; PORT, B. R. (author of initial R. **rpart: Recursive Partitioning and Regression Trees**. [s.l.: s.n.]

THERNEAU, T. M.; ATKINSON, E. J.; MAYO. **An introduction to recursive partitioning using the RPART routines**. [s.l.] Technical report Mayo Foundation, 2017. . Disponível em: <<http://r.789695.n4.nabble.com/attachment/3209029/0/zed.pdf>>. Acesso em: 19 jun. 2017.

TORGO, L. **Data Mining with R: Learning with Case Studies, Second Edition**. 2 edition ed. Boca Raton: Chapman and Hall/CRC, 2017.

TRAVASSOS, C.; WILLIAMS, D. R. The concept and measurement of race and their relationship to public health: a review focused on Brazil and the United States. **Cadernos de saúde pública**, v. 20, n. 3, p. 660–678, 2004.

UNITED NATIONS. **World Population Prospects: The 2015 Revision, Key Findings and Advance Tables**. [s.l.: s.n.]

VAN BAVEL, J. The world population explosion: causes, backgrounds and projections for the future. **Facts, Views & Vision in ObGyn**, v. 5, n. 4, p. 281, 2013.

VASCONCELOS, A. M. N.; GOMES, M. M. F. Transição Demográfica: A Experiência Brasileira. **Epidemiologia e Serviços de Saúde**, v. 21, n. 4, p. 539–548, dez. 2012.

VETTER, D. M.; SIMÕES, C. Ca. da S. Acesso a infra-estrutura de saneamento básico e mortalidade. **Revista Brasileira de Estatística**, v. 42, p. 17–35, 1981.

VICTORA, C. G.; BARROS, F. C. Infant mortality due to perinatal causes in Brazil: trends, regional patterns and possible interventions. **São Paulo Medical Journal**, v. 119, n. 1, p. 33–42, 2001.

VISWANATHAN, V.; VISWANATHAN, S.; GOHIL, A.; YU-WEI, C. **R: Recipes for Analysis, Visualization and Machine Learning**. Edição: 1 ed. [s.l.] Packt Publishing, 2016.

WAMALA, S. P.; LYNCH, J.; KAPLAN, G. A. Women's Exposure to Early and Later Life Socioeconomic Disadvantage and Coronary Heart Disease Risk: The Stockholm Female Coronary Risk Study. **International Journal of Epidemiology**, v. 30, n. 2, p. 275–284, abr. 2001.

WHO. **Global Health Observatory data repository - Diarrhoeal diseases**. Disponível em: <<http://apps.who.int/gho/data/view.main.ghe2002015-CH3?lang=en>>. Acesso em: 2 set. 2017.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. [s.l.] Springer-Verlag New York, 2009.

WING, M. K. C. from J.; WESTON, S.; WILLIAMS, A.; KEEFER, C.; ENGELHARDT, A.; COOPER, T.; MAYER, Z.; KENKEL, B.; TEAM, the R. C.; BENESTY, M.; LESCARBEAU, R.; ZIEM, A.; SCRUCICA, L.; TANG, Y.; CANDAN, C.; HUNT, and T. **caret: Classification and Regression Training**. [s.l.: s.n.]

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data mining: practical machine learning tools and techniques**. Burlington, MA: Morgan Kaufmann, 2011.

WRIGHT, M. N.; ZIEGLER, A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. **Journal of Statistical Software**, v. 77, n. 1, p. 1–17, 2017.

WÜNSCH, F. V.; MONCAU, J. E. Mortalidade por câncer no Brasil 1980-1995: padrões regionais e tendências temporais. **Rev Assoc Med Bras**, v. 48, n. 3, p. 250–7, 2002.

YAU, N. **Data points: visualization that means something**. Indianapolis, IN: John Wiley & Sons, Inc, 2013.

YUNES, J.; RONCHEZEL, V. S. C. Evolução da mortalidade geral, infantil e proporcional no Brasil. **Rev. Saúde públ**, p. 3–48, 1974.

APÊNDICE A – AVALIANDO OS REGISTROS DE ÓBITO COM INFORMAÇÃO FALTANTE OU IGNORADA DA ESCOLARIDADE/INSTRUÇÃO

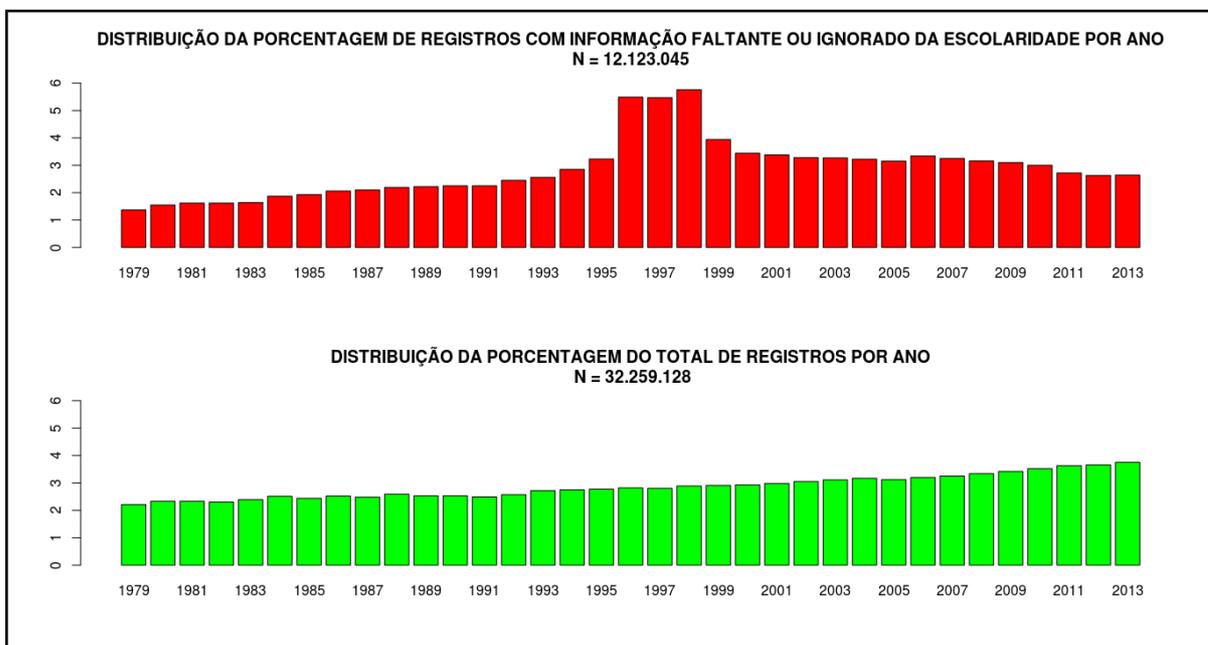


Figura A1 – Comparação entre a distribuição anual de registros com informação faltante ou ignorada da Escolaridade e o total de registros

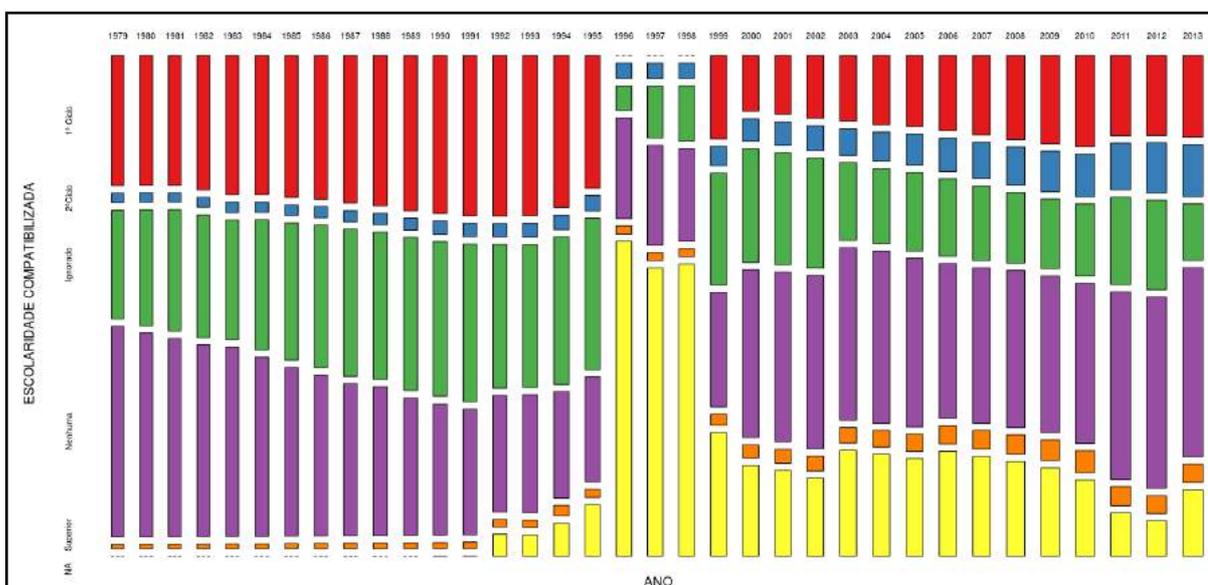


Figura A2 – Distribuição da categoria da Escolaridade por ano

APÊNDICE B – DISTRIBUIÇÃO DOS ÓBITOS POR CAUSAS ESPECÍFICAS NA ABORDAGEM 3 – NEOPLASIAS DE PULMÃO, NA ABORDAGEM 4 – DOENÇAS DO APARELHO CIRCULATÓRIO E NA ABORDAGEM 6 – CAUSAS EXTERNAS DA ETAPA III SEGUNDO FAIXA ETÁRIA

CAUSAS ESPECÍFICAS	FAIXA ETÁRIA					
	Menos de 1 ano	1 a 18 anos	19 a 29 anos	30 a 59 anos	60 a 79 anos	80 ou mais anos
Neoplasias de Estômago	11 (0,00%)	268 (0,01%)	2.947 (0,08%)	108.634 (2,90%)	193.829 (5,17%)	55.323 (1,48%)
Neoplasias de Pulmão	37 (0,00%)	783 (0,02%)	2.332 (0,06%)	134.531 (3,59%)	263.735 (7,04%)	53.519 (1,43%)
Demais Neoplasias	3.974 (0,11%)	76.422 (2,04%)	71.395 (1,91%)	997.862 (26,63%)	1.339.346 (35,75%)	441.576 (11,79%)
TOTAL	4.022 (0,11%)	77.473 (2,07%)	76.674 (2,05%)	1.241.027 (33,12%)	1.796.910 (47,96%)	550.418 (14,70%)

Tabela B1 – Distribuição dos Óbitos por Causas Específicas da Abordagem 3 – Neoplasias segundo Faixa Etária

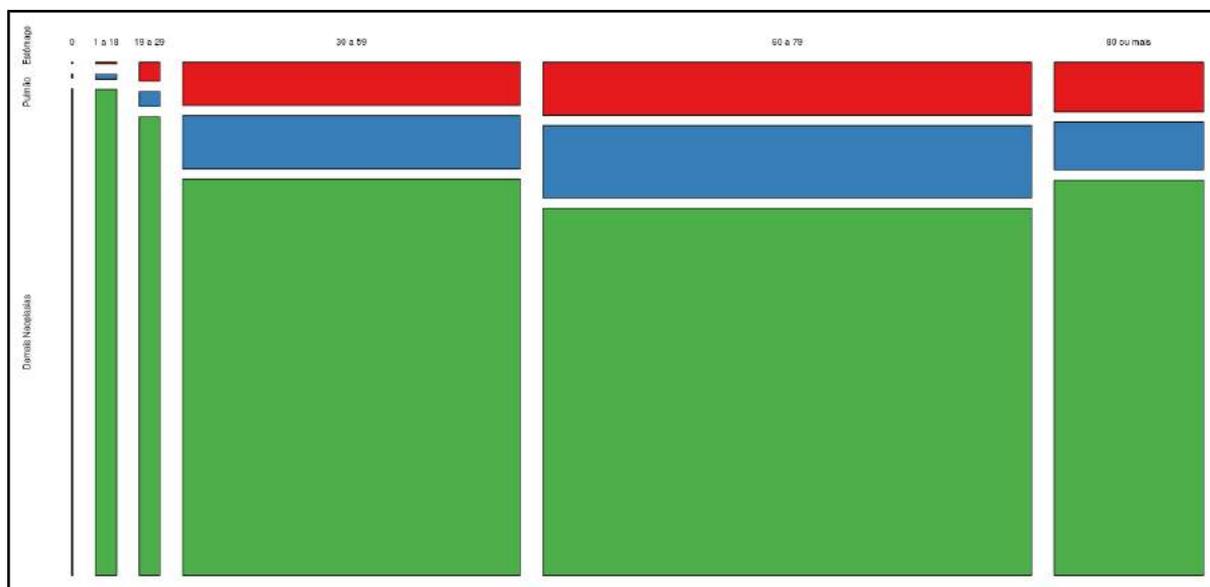


Figura B1 – Distribuição dos Óbitos por Causas Específicas da Abordagem 3 – Neoplasias – segundo Faixa Etária

CAUSAS ESPECÍFICAS	FAIXA ETÁRIA					
	Menos de 1 ano	1 a 18 anos	19 a 29 anos	30 a 59 anos	60 a 79 anos	80 ou mais anos
Cerebrovascular	1.323 (0,02%)	10.499 (0,13%)	28.800 (0,34%)	647.308 (7,73%)	1.286.978 (15,38%)	729.813 (8,72%)
Código Lixo	4.536 (0,05%)	7.800 (0,09%)	11.909 (0,14%)	164.505 (1,97%)	443.196 (5,29%)	363.893 (4,35%)
Infarto	17 (0,00%)	1.486 (0,02%)	14.185 (0,17%)	566.669 (6,77%)	952.944 (11,39%)	359.284 (4,29%)
Demais Causas Circulatórias	7.477 (0,09%)	27.402 (0,33%)	49.448 (0,59%)	640.856 (7,66%)	1.282.777 (15,33%)	767.110 (9,16%)
TOTAL	13.353 (0,16%)	47.187 (0,57%)	104.342 (1,24%)	2.019.338 (24,13%)	3.965.895 (47,39%)	2.220.100 (26,52%)

Tabela B2 – Distribuição dos Óbitos por Causas Específicas da Abordagem 4 – Doenças do Aparelho Circulatório segundo Faixa Etária

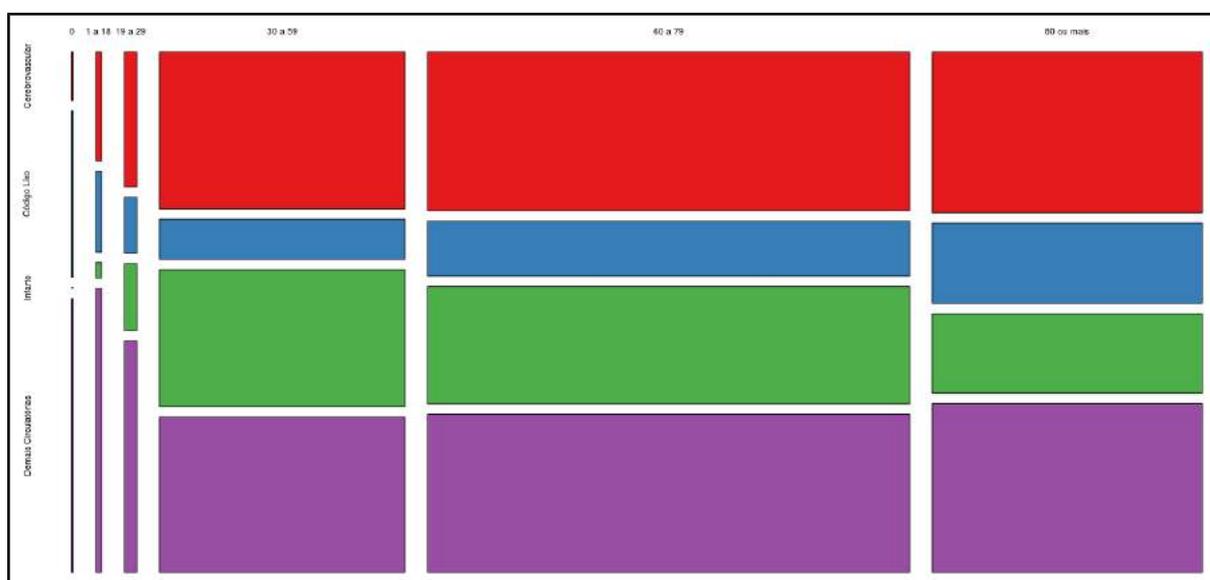


Figura B2 – Distribuição dos Óbitos por Causas Específicas da Abordagem 4 – Doenças do Aparelho Circulatório segundo Faixa Etária

CAUSAS ESPECÍFICAS	FAIXA ETÁRIA					
	Menos de 1 ano	1 a 18 anos	19 a 29 anos	30 a 59 anos	60 a 79 anos	80 ou mais
Acidente de Transporte	2.794 (0,08%)	148.295 (4,34%)	278.281 (8,15%)	433.412 (12,69%)	113.718 (3,33%)	18.458 (0,54%)
Código Lixo	3.197 (0,09%)	44.056 (1,29%)	96.070 (2,81%)	144.293 (4,22%)	40.982 (1,20%)	20.214 (0,59%)
Homicídios	1.382 (0,04%)	140.918 (4,12%)	484.662 (14,19%)	431.690 (12,64%)	35.393 (1,04%)	3.709 (0,11%)
Demais Causas Externas	18.307 (0,54%)	161.644 (4,73%)	204.561 (5,99%)	369.2247 (10,81%)	138.147 (4,04%)	82.837 (2,42%)
TOTAL	25.680 (0,75%)	491.913 (14,48%)	1.063.574 (31,14%)	1.379.642 (40,36%)	328.240 (9,61%)	125.258 (3,66%)

Tabela B3 – Distribuição dos Óbitos por Causas Específicas da Abordagem 6 – Causas Externas segundo Faixa Etária

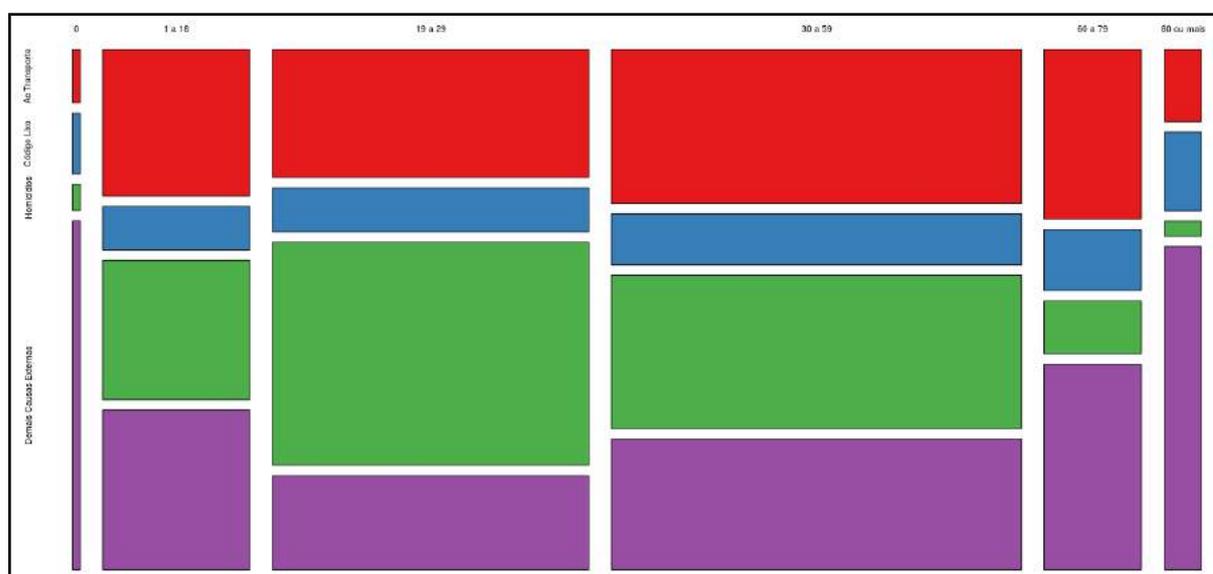


Figura B3 – Distribuição dos Óbitos por Causas Específicas da Abordagem 6 – Causas Externas segundo Faixa Etária

APÊNDICE C – DISTRIBUIÇÃO DA PROPORÇÃO DE ÓBITO DE CADA CATEGORIA DE CAUSA ESPECÍFICA SEGUNDO ABORDAGENS DA ETAPA III

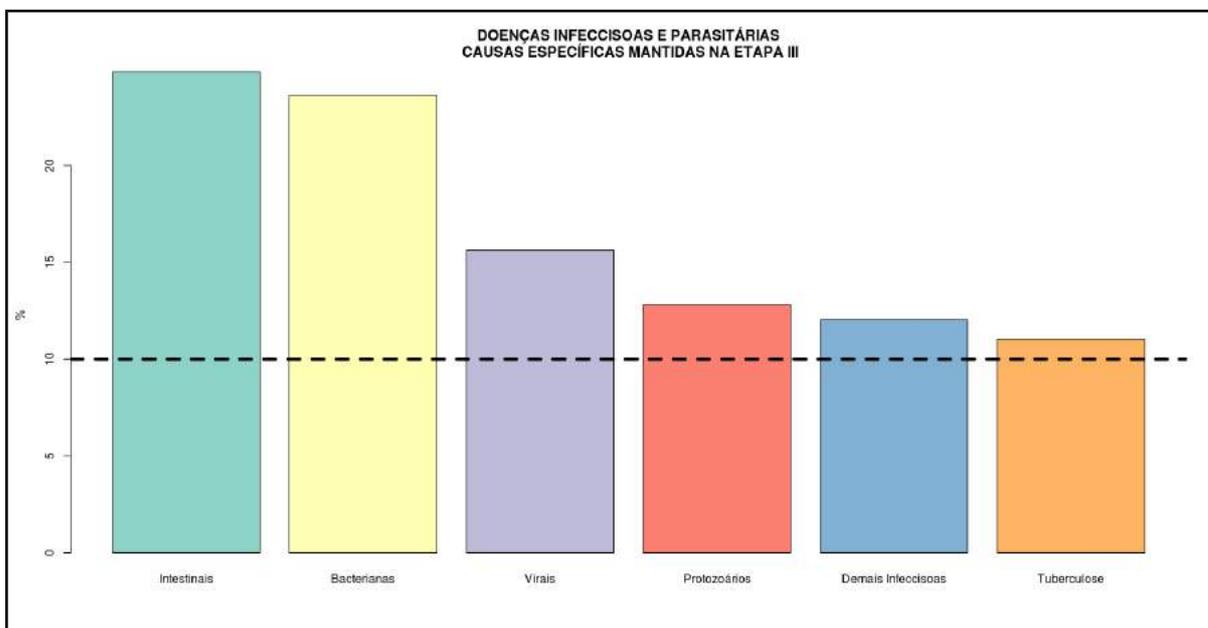


Figura C1 – Causas Específicas entre as Doenças Infecciosas e Parasitárias utilizadas na Abordagem 2 da Etapa III

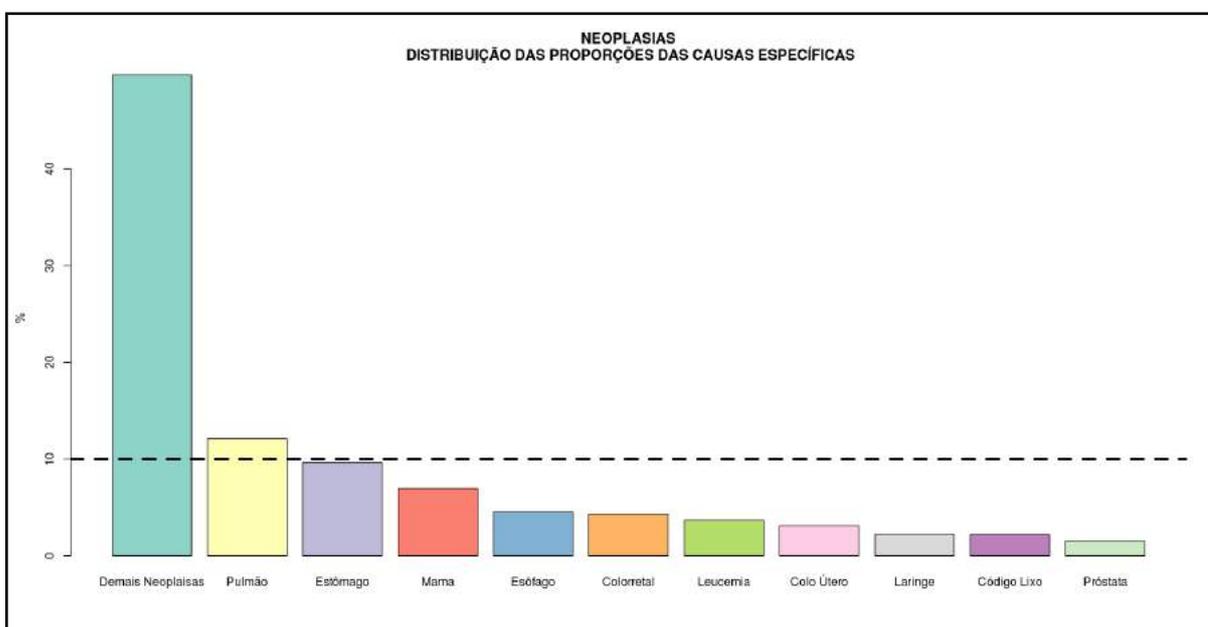


Figura C2 – Distribuição das proporções das Causas Específicas entre as Neoplasias

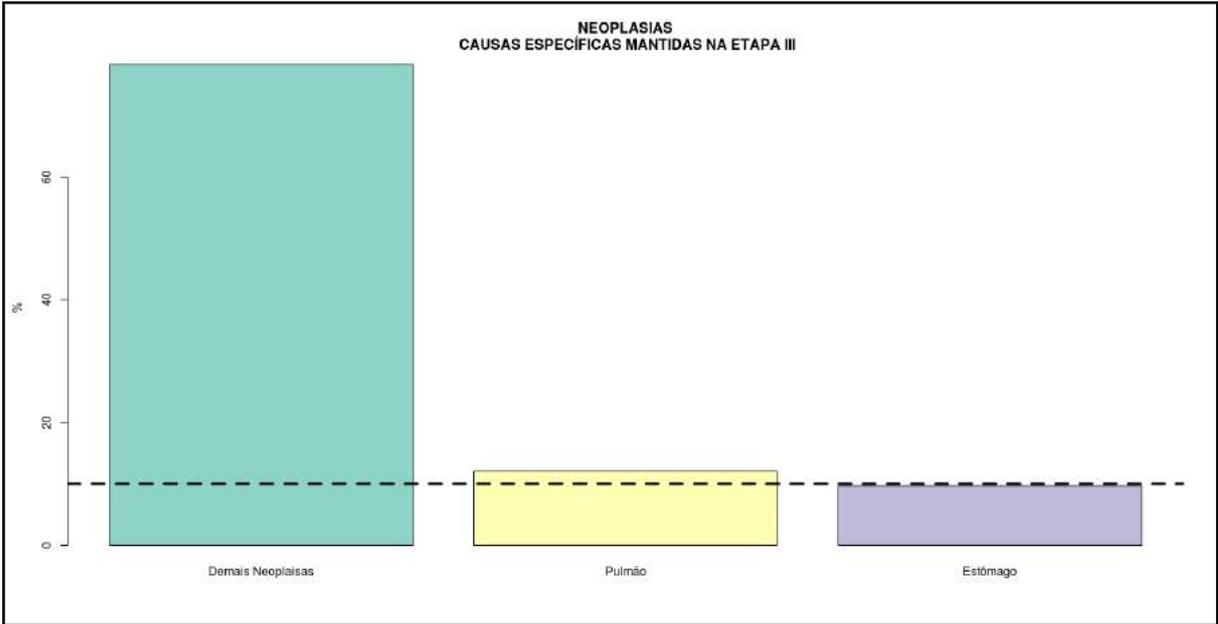


Figura C3 – Causas Específicas de Óbito entre as Neoplasias utilizadas na Abordagem 3 da Etapa III

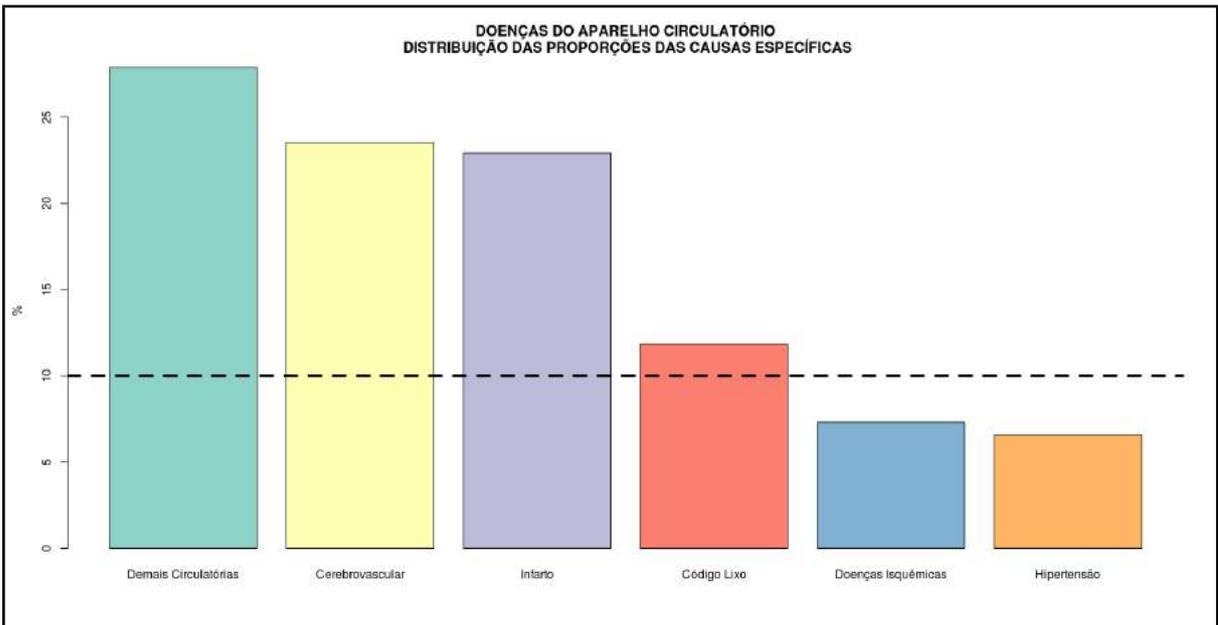


Figura C4 – Distribuição das proporções das Causas Específicas entre as Doenças do Aparelho Circulatório

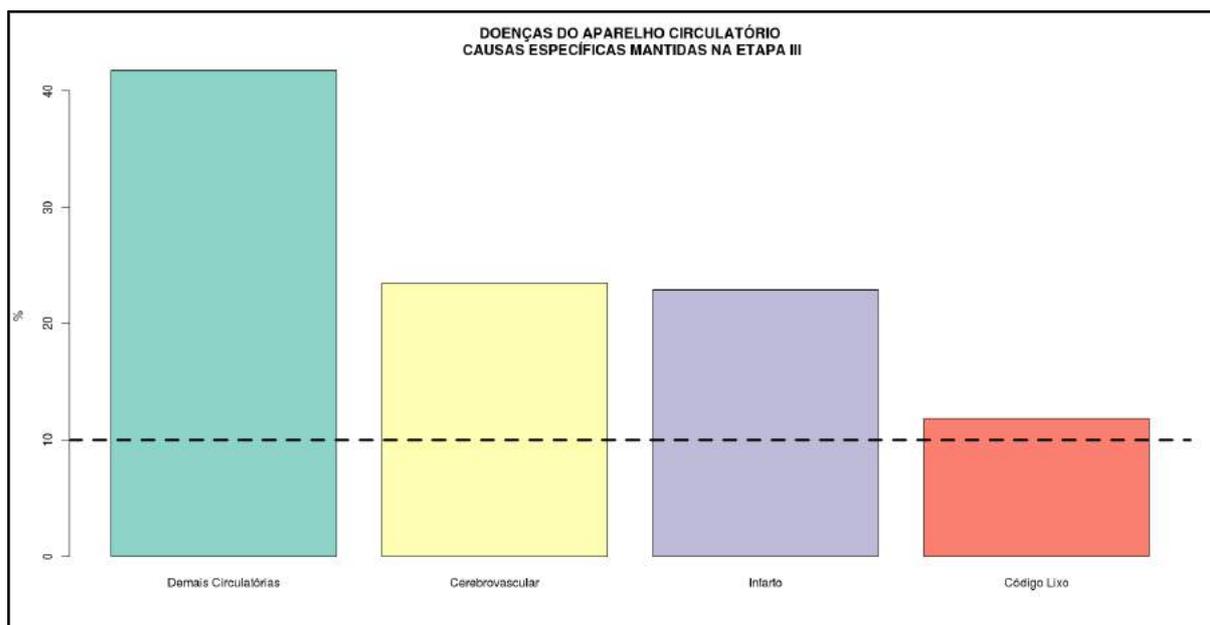


Figura C5 – Causas Específicas entre as Doenças do Aparelho Circulatório utilizadas na Abordagem 4 da Etapa III

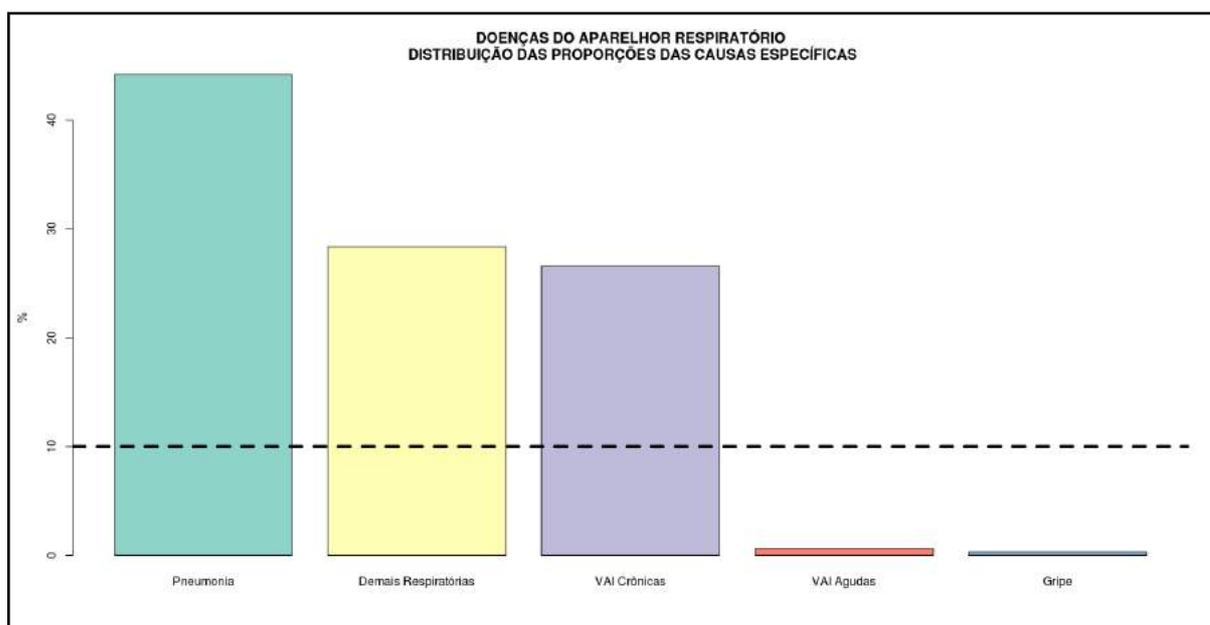


Figura C6 – Distribuição das proporções das Causas Específicas entre as Doenças do Aparelho Respiratório

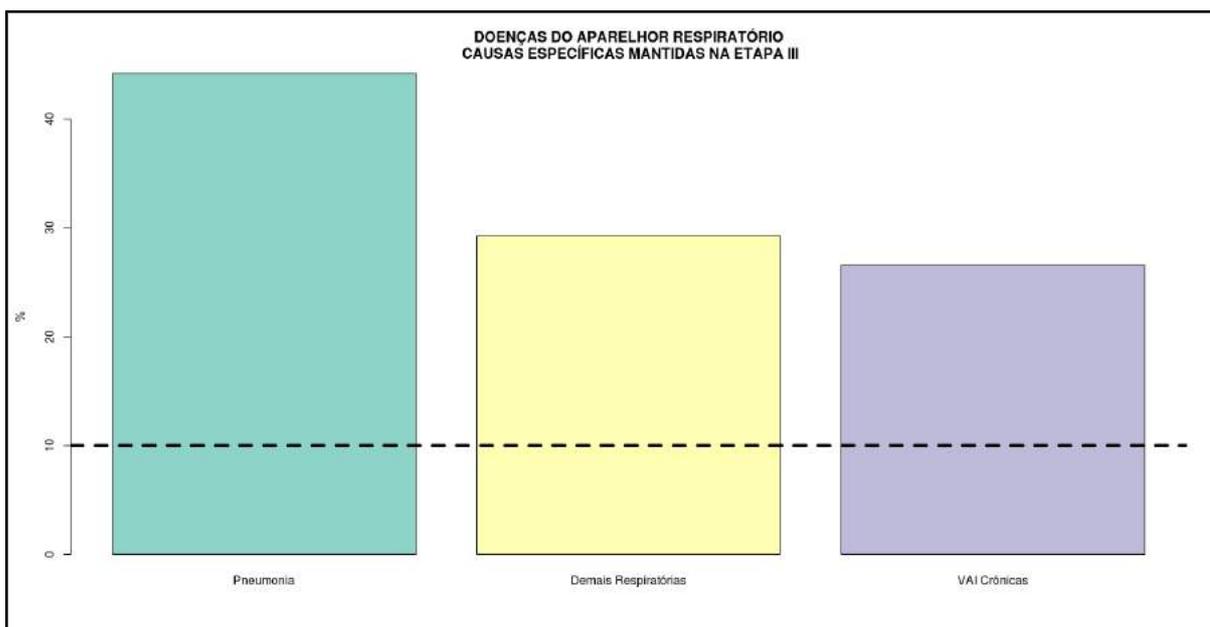


Figura C7 – Causas Específicas entre as Doenças do Aparelho Respiratório utilizadas na Abordagem 5 da Etapa III

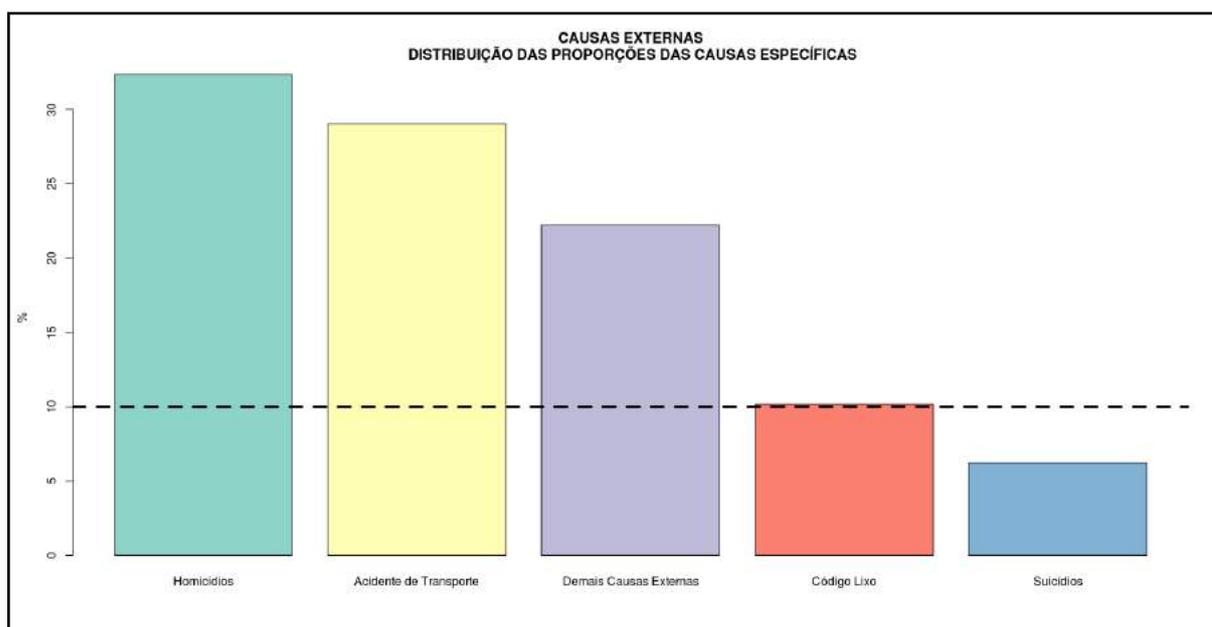


Figura C8 – Distribuição das proporções das Causas Específicas entre as Causas Externas

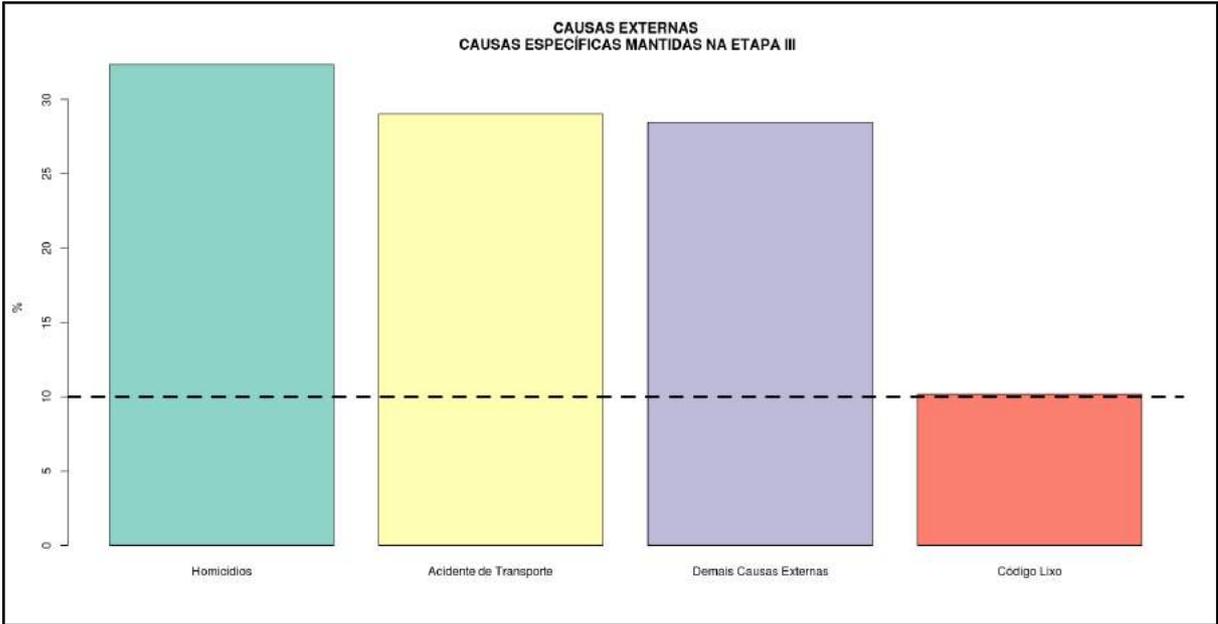


Figura C9 – Causas Específicas entre as Causas Externas utilizadas na Abordagem 6 da Etapa III

APÊNDICE D – ARTIGO

MORTALIDADE NO BRASIL: O QUE MUDOU NOS ÚLTIMOS 20 ANOS?

RESUMO

OBJETIVO Descrever tendências regionais das taxas de mortalidade por grupos de causas selecionadas no Brasil nos anos 1980, 1991 e 2010, e identificar os indicadores sociais, econômicos e demográficos associados. **METODOLOGIA** Trata-se de um estudo ecológico tendo como unidade de observação Microrregiões do Brasil (n = 557), nos anos censitários de 1991, 2000 e 2010. As causas de mortalidade selecionadas, foram as mais frequentes dentro dos principais grupos de causa de mortalidade no país, que atingem grupos etários específicos, e apontados na literatura como tendo relações distintas com diversos aspectos contextuais no Brasil. Os indicadores sociais, econômicos e demográficos correspondem a diversos aspectos contextuais do país utilizados com frequência em estudos epidemiológicos. Para a seleção de indicadores foi utilizado Análise Fatorial Exploratória, e a relação entre os indicadores e taxas foi analisada a partir do Coeficiente de Correlação de Pearson. *Boxplots* foram utilizados para apresentar os padrões e tendências regionais. **RESULTADOS** Observou-se tendências distintas nas taxas de mortalidade por algumas causas de óbito por região e por sexo, com destaque para as Doenças Cerebrovasculares, que aumentaram em ambos os sexos nas regiões Norte, Nordeste e Centro-Oeste, e diminuiu nas regiões Sul e Sudeste, para as Neoplasias de Pulmão, que aumentou de maneira mais aguda entre mulheres e para a mortalidade por Homicídios, cerca de 9 vezes maior entre homens. Destaca-se também a queda na correlação das taxas de mortalidade com alguns dos indicadores analisados sobretudo com as Doenças Cerebrovasculares, que passou de forte a inexistente no período. **CONCLUSÕES** Os achados do presente estudo apontam para a existência de diversas realidades em relação à mortalidade dentro do território nacional, que apesar das melhorias gerais nos indicadores socioeconômicos, mantém um cenário de iniquidades.ínima de 75%, avaliações formais e monografia de conclusão de Curso

INTRODUÇÃO

O Brasil passa por um processo de transição epidemiológica no qual a mortalidade por doenças infecciosas e parasitárias, que afetam grupos etários mais jovens, decaíram de maneira aguda permitindo que uma parcela cada vez maior da população infantil atingisse idade adulta, enquanto os avanços médicos, sociais e culturais permitiram um aumento na expectativa da população, sobressaindo assim as doenças crônicas e não transmissíveis, que afetam com maior frequência os adultos e idosos, como as principais causas de mortalidade¹⁻³.

Apesar de este processo ter sido deflagrado principalmente por ações sanitárias e de saúde pública de grande impacto na mortalidade⁴, a heterogeneidade do contexto social, econômico e demográfico do país influencia a velocidade e intensidade com o qual o mesmo vem ocorrendo, afetando de maneira diferente os grupos sociais e regiões geográficas distintas, agravando as desigualdades em saúde^{5,6}.

O Brasil é historicamente marcado por desigualdades sociais e econômicas^{7,8}. Diversas mudanças permitiram uma melhoria na condição de vida da população nos últimos anos^{9,10}, contudo, as iniquidades em saúde persistem no país^{11,12}.

Nas últimas quatro décadas as doenças do aparelho circulatório, as neoplasias, as causas externas, as doenças do aparelho respiratório e as doenças infecciosas e parasitárias foram responsáveis por 65,9% do total de óbitos no Brasil^{13,14}.

As doenças do aparelho circulatório se destacam como a principal causa de mortalidade no país nos últimos anos, representando cerca de um terço dos óbitos registrados desde a criação do Sistema de Informação sobre a Mortalidade – SIM, em 1979. A tendência das taxas de mortalidade por doenças cerebrovasculares e doenças isquêmicas do coração foi de queda entre 1980 e 2001 no Sul e Sudeste, tendência similar à observada nos Estados Unidos na década de 1960, e de aumento no Norte e Nordeste, assim como observado nos países menos desenvolvidos do leste Europeu a partir da 1970¹⁵⁻¹⁷.

Cerca de 55% dos casos novos de neoplasias se concentram em países em desenvolvimento, sendo as de pulmão, colorretal e estômago as cinco mais comuns para ambos os sexos¹⁸. A proporção de óbitos por neoplasias aumentou no Brasil entre 1980 e 2000, passando da quinta para a terceira principal causa de óbito atrás apenas de doenças do aparelho circulatório e causas mal definidas¹⁹. Constata-se uma sobreposição de neoplasias associadas à pobreza no país, como neoplasias de estômago, útero, pênis e cavidade oral e aquelas mais frequentes em regiões desenvolvidas, como mama, próstata, cólon e reto²⁰, fato provavelmente resultante da exposição a fatores de risco ambientais resultantes do processo de industrialização, mudanças culturais e das disparidades sociais²¹⁻²³.

A mortalidade por causas externas nos principais centros urbanos do Brasil tem aumentado desde a década de 1980²⁴, passando de 59 óbitos para cada 100.000 habitantes para 75,2 em 2002, valores muito superiores aos observados nos países a Europa Ocidental e Estados Unidos, onde foram cerca de 3 e 5 óbitos, respectivamente²⁵. Homicídios e acidentes de transporte são as principais causas externas de mortalidade no país, ambos mais frequentes em homens. A mortalidade por acidentes de trânsito é mais dispersa espacialmente, afetando de maneira similar as faixas etárias a partir dos 20 anos de idade^{25,26} enquanto a mortalidade por homicídios geralmente é concentrada em determinados espaços sociais^{27,28}, sendo a mortalidade mais elevada entre indivíduos com 20 a 29 e com cor da pele preta^{25,26,29}.

Entre as doenças do aparelho respiratório a pneumonia se destaca como a causa mais frequente de mortalidade. Este grupo de doenças impactam principalmente a morbimortalidade de crianças²¹⁻²³, e idosos³³⁻³⁵ no Brasil, estando associadas a indicadores socioeconômicas³², características domiciliares³⁶ e poluição atmosférica^{37,38}. A proporção de óbitos infantis por estas causas diminuiu no país entre 1994 e 2004³⁹, sendo mais elevada na região Sudeste⁴⁰. Entre idosos as taxas de mortalidade aumentaram em todas as regiões no período entre 1980 e 2000, com as taxas mais elevadas concentradas no Sul e Sudeste, e as mais baixas no Norte e Nordeste¹⁹.

As doenças infecciosas intestinais representam um terço dos óbitos por doenças infecciosas no Brasil, e a persistência da mortalidade pelas mesmas reflete condições de saneamento e disponibilidade de água tratada, de maneira que indivíduos inseridos em contextos menos favorecidos se mostram mais suscetíveis a estes eventos, potencialmente evitáveis^{41,42}. Observou-se redução de 66% na mortalidade infantil por doenças infecciosas intestinais no Brasil entre 1997 e 2006⁴³, estando as maiores taxas concentradas nas regiões Nordeste e Norte, com valores de duas a quatro vezes mais elevados que no Sul e Sudeste⁴⁴.

Logo, observam-se diferenças nos padrões e tendências da mortalidade das principais causas de óbito nas regiões do Brasil que podem estar associadas com as desigualdades e iniquidades no país. Apesar dos padrões de mortalidade serem amplamente analisados, pouco se sabe sobre sua relação com os indicadores sociais, econômicos e demográficos e da evolução desta relação nas últimas décadas. Esta compreensão pode ajudar no desenvolvimento de políticas e ações de saúde pública que contemplem necessidades regionais específicas.

Neste contexto, o objetivo deste estudo foi analisar as tendências regionais das taxas de mortalidade por grupos de causas selecionadas no Brasil nos anos 1980, 1991 e 2010 e de sua correlação com indicadores sociais, econômicos e demográficos.

MATERIAL E MÉTODOS

Trata-se de um estudo ecológico tendo como unidade de observação as Microrregiões do Brasil (n = 557), nos anos censitários de 1990 a 2010. A Microrregião de Fernando de Noronha foi excluída da análise por se tratar de uma área de preservação com população residente reduzida, o que gera flutuações nas taxas de mortalidade.

Cada Microrregião é composta por municípios contíguos com características similares em relação às estruturas de produção e relações locais^{45,46}. Optou-se pela utilização das Microrregiões como unidades de análise para a abordagem dos dados em espaço geográfico delimitado e constante em todo o período, tendo em vista o surgimento e extinção de municípios no país.

Os dados sobre mortalidade foram retirados do Sistema de Informação Sobre Mortalidade – SIM. As informações sobre óbitos para cada sexo e faixa etária, foram agregados por municípios e posteriormente alocados em cada Microrregião.

Para estabilizar a contagem de óbitos utilizou-se a média do número de óbitos de três anos, incluindo o ano anterior e posterior a cada censo, ou seja, para 1991 utilizou-se a média de óbitos de 1990, 1991 e 1992, para o ano 2000 a média de 1999, 2000 e 2001 e para 2010 a média de 2009, 2010 e 2011. Foram calculadas taxas de mortalidade por sexo, padronizadas por idade pelo método direto considerando como referência a população do Brasil de cada sexo no ano 2000.

Foram selecionadas as causas de óbito mais frequentes dentro dos principais capítulos da Codificação Internacional de Doenças – CID: Doenças Infecciosas Intestinais (códigos 001-009 da CID – 9 e A01-A09 da CID – 10), Neoplasias de Traqueia, Brônquios e Pulmão (códigos 162 da CID – 9 e C33-C34 da CID – 9), Pneumonias (códigos 480-487 da CID – 9 e J09-J18 da CID – 10), Doenças Cerebrovasculares (códigos 430-438 da CID – 9 e I60-I69 da CID – 10) e Homicídios (códigos 960-978 da CID – 9 e X85-99, Y00-09 e Y35-36 da CID – 10) para Causas Externas.

Os indicadores sociais, econômicos e demográficos foram retirados do Censo (IBGE) e foram selecionados de acordo com a disponibilidade da informação por microrregião nos três anos analisados: Taxa de Analfabetismo (Analfabetismo), Proporção da População que cursou ao Menos o Ensino Médio ou Correspondente (Anos de Estudo), Renda Média Domiciliar Per Capita (Renda), com os valores referentes a renda padronizados e corrigidas com base no Índice Nacional de Preço ao Consumidor – INPC de julho de 2010, Proporção de Pessoas com Baixa Renda (Baixa Renda), Taxa de Desemprego (Desemprego), Percentual de

Domicílios com acesso à Rede Geral de Esgoto ou Pluvial (Esgotamento), Índice de Envelhecimento (IEP), Razão de Dependência de Jovens (RDJ), Razão de Dependência de Idosos (RDI).

Optou-se por selecionar os indicadores menos correlacionados e que melhor representavam cada aspecto do contexto social, econômico e demográfico, tendo em vista sua utilização em uma análise de regressão em estudos posteriores. Para tanto, utilizou-se Análise Fatorial Exploratória, método que permite identificar grupos de variáveis correlacionadas representadas por fatores não correlacionados entre si⁴⁷.

Os resultados da Análise Fatorial foram considerados para compreender a correlação entre os indicadores no decorrer dos anos e selecioná-los de acordo com sua participação na composição dos fatores. Neste contexto, a definição do número de fatores se deu pelo critério da raiz latente⁴⁷ para o ano de 1991 extraindo-se o mesmo número de fatores nos anos subsequentes e a seleção dos indicadores foi baseada na sua carga fatorial e em sua variância específica.

Gráficos com *boxplots* da distribuição de cada taxa por sexo e de cada indicador selecionado foram utilizados para descrever a evolução e diferenças regionais. Para analisar a associação entre indicadores selecionados e as taxas de mortalidade utilizou-se o coeficiente de correlação linear de Pearson, que varia de -1 a +1 indicando a força e a direção da relação linear de duas variáveis⁴⁸. Todas as análises foram realizadas utilizando o software R⁴⁹.

O estudo utilizou dados de acesso aberto, sem identificação dos indivíduos, estando de acordo com os aspectos éticos definidos pela Resolução 466/2012.

RESULTADOS

Comportamento das Taxas de Mortalidade

As taxas de mortalidade analisadas, expressas por 100.000 habitantes, foram superiores entre homens em todo o período. A maior razão de sexo foi entre as taxas por homicídios, em que a média entre homens (24,7, 27,5 e 37,2 em 1991, 2000 e 2010, respectivamente) foi cerca de nove vezes a média entre mulheres (2,7, 3,1 e 4,1).

A média das taxas de mortalidade por doenças infecciosas intestinais diminuiu de 8,9 para 2,5 entre homens e de 7,2 para 2,4 entre mulheres no período de 1991 a 2010. Padrão similar foi observado para as doenças cerebrovasculares, que variaram de 58,6 para 43,6 entre homens e de 48,5 para 38,6 entre mulheres. Já as taxas de mortalidade por pneumonia aumentaram de 18,6 para 20,4 e de 14,6 para 18,6 no sexo masculino e feminino,

respectivamente. Da mesma forma, as neoplasias de pulmão variaram de 8,5 para 9,6 entre homens e de 3,2 para 3,9 entre mulheres.

Com exceção das doenças infecciosas intestinais, o comportamento das taxas no Sul e Sudeste foram similares para ambos os sexos, com padrão diferente do observado no Norte e Nordeste, também similares entre si.

A mortalidade por doenças infecciosas intestinais para ambos os sexos foi ligeiramente mais elevada no Norte e Nordeste e diminuiu em todas as regiões (**Figura 1 – A** e **Figura 2 – A**). A redução da amplitude dos *boxplots* e do número de *outliers* apontam para uma distribuição mais homogênea destas taxas entre as microrregiões em 2010.

A taxa de mortalidade por neoplasia de pulmão no sexo feminino aumentou em todas as regiões e os patamares mais elevados se concentraram nas regiões Sul, Sudeste e Centro-Oeste; já no sexo masculino aumentou nas regiões Norte, Nordeste e Centro-Oeste, variando pouco nas regiões Sul e Sudeste, onde os valores foram mais elevados (**Figura 2 – B** e **Figura 2 – B**).

As taxas de mortalidade por doenças cerebrovasculares aumentaram nas regiões Norte, Nordeste e Centro-Oeste e diminuíram nas regiões Sul e Sudeste, de maneira que as medianas de todas as regiões se aproximaram em 2010 em ambos os sexos (**Figura 1 – C** e **Figura 2 – C**).

O comportamento das taxas de mortalidade por pneumonia foi de aumento nas regiões Norte, Nordeste e Centro-Oeste (**Figura 1 – D** e **Figura 2 – D**). Nas regiões Sul e Sudeste, onde se concentraram os valores mais elevados em todo o período, observou-se diminuição no sexo masculino e estabilidade no sexo feminino, ocorrendo queda entre 1991 e 2000 em ambos os sexos.

Em relação aos homicídios, observou-se que as taxas de mortalidade no sexo feminino aumentaram em todo o país, de maneira mais aguda nas regiões Norte e Nordeste, com os valores mais elevados concentrados na região Centro-Oeste em todo o período (**Figura 1 – E**). No sexo masculino as taxas de mortalidade por homicídios também aumenta em todas as regiões (**Figura 2 – E**). Em 2010 as medianas nas regiões Norte, Nordeste e Centro-Oeste foram as mais elevadas do país, com valores próximos de 40. Ao contrário das demais causas, observou-se uma elevada concentração de *outliers*, com destaque para as regiões Nordeste e Sudeste, e uma manutenção de grande variabilidade da distribuição de taxas nas microrregiões em todo o período.

Seleção e Comportamento dos Indicadores Sociais, Econômicos e Demográficos

Os resultados da Análise Fatorial se encontram na Tabela 1. Foram selecionados dois fatores em cada ano que explicaram 74,7%, 77,3% e 77,9% da variabilidade dos indicadores em 1991, 2000 e 2010, respectivamente, e em todos os anos a Taxa de Analfabetismo, a Proporção de Pessoas com Baixa Renda, a Razão de Dependência de Jovens, a Renda e a Proporção da População que cursou ao menos ensino médio estiveram fortemente correlacionados com o primeiro fator, enquanto a Razão de Dependência de Idosos e o Índice de Envelhecimento estiveram correlacionados com o segundo fator. A Taxa de Desemprego e o Percentual de Domicílios com acesso à Rede Geral de Esgoto ou Pluvial foram os indicadores menos correlacionados com os fatores.

As variáveis selecionadas para análise de correlação com as taxas de mortalidade foram Renda Média Domiciliar Per Capita (Renda) e Razão de Dependência de Idosos (RDI), indicadores mais correlacionados com os fatores 1 e 2, Percentual de Domicílios com Acesso a Rede Geral de Esgoto ou Pluvial (Esgotamento), por se tratar de um indicador de serviços básicos de saneamento potencialmente relacionados com algumas causas de mortalidade, e Taxa de Desemprego (Desemprego), pela menor correlação com os demais indicadores.

Observou-se aumento na média da Renda Per Capita Média, no Brasil, de R\$ 220 para R\$ 541 entre 1991 e 2010, refletindo no aumento deste indicador nas cinco regiões (**Figura 3 – A**). Em todo o período as medianas foram mais elevadas nas regiões Sul, Sudeste e Centro-Oeste enquanto as mais baixas se concentraram na região Nordeste, onde se destacaram *outliers* em todos os anos, referentes às microrregiões que compõem as capitais.

A Razão de Dependência de Idosos aumentou em todas as regiões, com elevação mais acentuada no Centro-Oeste (**Figura 3 – B**). As maiores medianas se concentraram nas regiões Nordeste, Sul e Sudeste enquanto as menores estiveram sempre na região Norte, intensificando a diferença em relação às demais regiões.

Já o comportamento da Taxa de Desemprego foi diferente daquele observado nos demais indicadores selecionados, com tendência geral de aumento entre 1991 e 2010, e um pico em todas as regiões em 2000 (**Figura 3 – C**). Os maiores valores foram observados nas regiões Norte e Nordeste, seguido das regiões Sudeste, Centro-Oeste e Sul. Observou-se também uma concentração de microrregiões *outliers* com valores elevados na região Nordeste.

O Percentual de Domicílios com Acesso à Rede Geral de Esgoto ou Pluvial apresentou as diferenças regionais mais intensa no período (**Figura 3 – D**). Na região Sudeste, já em 1991, pelo menos metade das microrregiões tinham mais de 50% de domicílios com acesso a

esgotamento sanitário, superando 75% em 2010. Nas demais regiões até 2010 a maioria das microrregiões não apresentavam sequer 40% dos domicílios com acesso à rede geral de esgoto ou pluvial.

Na região Nordeste ocorreu um aumento acentuado no período do indicador de esgotamento, ainda que desigual, e apenas Salvador, onde mais de 80% dos domicílios tinham acesso ao esgotamento pela rede geral ou pluvial, foi considerado *outlier* em 2010. Contudo, na região Norte a variação foi discreta e até 2010 nenhuma de suas microrregiões apresentava ao menos 40% de domicílios com acesso a esgotamento.

Correlação entre os Indicadores Selecionados e as Taxas de Mortalidade

A **Figura 4** apresenta os coeficientes de correlação entre cada taxa de mortalidade, por sexo, e os indicadores selecionados, em cada ano. O diâmetro dos círculos e a tonalidade das cores são proporcionais ao valor do coeficiente de correlação enquanto a cor azul representa correlação negativa e a cor vermelha representa correlação positiva.

As Doenças Infecciosas Intestinais apresentaram correlação negativa e fraca/moderada com a renda e com o esgotamento em ambos os sexos, com no período, o que indica que as microrregiões com renda mais elevada e maior percentual de domicílios com acesso à rede geral e fluvial de esgoto apresentaram, em média, menores taxas de mortalidade por este grupo de doenças.

A correlação entre neoplasia de brônquios e pulmão foi semelhante em ambos os sexos, sendo: positiva e moderada/forte com a renda, flutuando no período; positiva e fraca com esgotamento, e com tendência de queda; e negativa e fraca com o desemprego. Neste sentido, nas microrregiões onde a renda per capita foi mais elevada, as taxas de mortalidade por esta causa foram maiores.

As pneumonias também apresentaram correlação positiva e moderada com renda e com esgotamento, ambas decaindo entre 1991 e 2010, de maneira ligeiramente mais acelerada no sexo masculino. De maneira semelhante, a correlação das doenças cerebrovasculares foi positiva e moderada com a renda e positiva e fraca com o esgotamento até 2000, não estando correlacionada com nenhum dos indicadores em 2010.

A taxa de mortalidade por homicídios apresentou correlação negativa e fraca com a Razão de Dependência de Idosos no sexo feminino, diminuindo em intensidade no período, e

negativa e moderada no sexo masculino, com ligeiro aumento na força da correlação. No sexo masculino se observou uma correlação positiva e fraca/moderada com o desemprego, com tendência de aumento.

DISCUSSÃO

O objetivo do estudo foi descrever tendências regionais das taxas de mortalidade por grupos de causas selecionadas no Brasil nos anos 1980, 1991 e 2010, e identificar os indicadores sociais, econômicos e demográficos relacionados. Apesar das melhoras nos indicadores socioeconômicos em todo o país ainda existem desigualdades regionais marcantes que podem estar relacionadas com o comportamento também desigual das taxas de mortalidade no período analisado.

As doenças infecciosas intestinais atingem principalmente crianças, e a persistência da mortalidade por esta causa está associada ao estado nutricional, hábitos alimentares, saneamento e deficiência de cuidados primários que apontam para um contexto socioeconômico precário^{41,42}, o que justifica sua correlação negativa com os indicadores de renda e esgotamento em todo o período. Neste sentido, a queda generalizada na mortalidade por doenças infecciosas intestinais foi condizente com os avanços dos indicadores socioeconômicos no país, assim como as taxas ligeiramente mais elevadas no final do período nas regiões Norte e Nordeste onde se concentraram os piores indicadores.

Esta causa compõe o grupo de doenças reduzíveis em menores 5 anos de idade por ações adequadas de promoção e adequada de atenção à saúde, e diversos estudos também constataram redução constante das taxas de mortalidade infantil por esta causa, associando esta redução aos avanços da cobertura de abastecimento de água e tratamento de esgoto, e a adoção de terapia de reidratação oral⁵⁰⁻⁵³.

No presente estudo não foi calculada a taxa de mortalidade infantil por este grupo de causa, o que dificulta a comparação da magnitude das taxas com outros estudos e outros países, contudo, segundo dados do SIM⁵⁴ e do SINASC⁵⁵ a taxa de mortalidade infantil em menores de 5 anos por doenças infecciosas intestinais em 2015 no Brasil foi de 0,2 óbitos para cada 1.000 nascidos vivos, valor superior ao da taxa de mortalidade infantil por doenças diarreicas em países da Europa Ocidental, Estados Unidos e alguns países da América Latina, como Chile e Uruguai e equiparando-se a países como Albânia, Ucrânia, Oman, Líbano e Arábia Saudita⁵⁶, um cenário longe do ideal mas que representa um grande avanço comparado ao início dos anos 1990⁴¹.

O aumento das taxas de mortalidade por neoplasias de pulmão entre mulheres é compatível com os padrões descritos em estudos anteriores no país, bem como a concentração de valores mais elevados da mesma na região Sul, seguida do Sudeste, e mais baixos no Norte e Nordeste^{21,57-59}. Este padrão reflete a prevalência de tabagismo, mais elevada no Sul e Sudeste, com tendência de queda entre homens e estratos sociais de maior renda e escolaridade, e de estabilidade entre mulheres^{60,61}. Em estudo abordando as microrregiões da região Sul observou-se tendências espaciais e temporais distintas das taxas de mortalidade por neoplasias de pulmão entre homens e mulheres⁶², apontando a existência de desigualdades intra regionais que não foram abordadas no presente estudo que demandam atenção.

As tendências observadas na mortalidade por pneumonias no período podem ter sido influenciadas pela estrutura etária da população em cada região, pois enquanto em 2010 mais de 70% dos óbitos por esta causa ocorreram entre idosos no Sul e Sudeste, mais envelhecidas, no Norte e Nordeste cerca de 50% dos óbitos por esta causa ocorreram nesta faixa etária. Esta perspectiva é corroborada por diversos estudos que descreveram redução da mortalidade por doenças respiratórias entre crianças⁶³⁻⁶⁵ e aumento entre idosos^{19,35}, sobretudo em faixas etárias mais elevadas^{33,66} e podem explicar, em parte, as desigualdades na magnitude das taxas de mortalidade nas regiões. Além da composição etária, a prevalência de tabagismo no país também reflete estes padrões de mortalidade^{60,61}. Padrões e tendências regionais semelhantes às constatadas no presente estudo foram observadas em relação a mortalidade de idosos por doenças relacionadas à influenza no período de 1992 a 2005³⁵, e por pneumonia nas regiões Sudeste e Nordeste entre 1996 e 2008⁶⁷.

Além disto, a queda nas taxas de mortalidade por pneumonias em todas as regiões do país no ano 2000 e voltando a crescer em 2010 foi similar ao observado entre idosos no estado de São Paulo, onde se acredita que tal flutuação é resultado impacto inicial da vacinação de idosos contra influenza^{33,34}. Neste contexto, uma hipótese plausível para esta flutuação na mortalidade por pneumonia nas regiões é que ela também é causada pelo impacto imediato da vacinação contra influenza em 1999.

As correlações entre as taxas de mortalidade por pneumonia e o indicador de renda e esgotamento foram positivas e merecem uma análise mais apurada em estudos posteriores, pois valores elevados destes dois indicadores expressam de maneira geral um contexto social e sanitário melhor, o que representaria mais acesso a recursos que concedem maior proteção para esta doença. Uma hipótese é de que o envelhecimento populacional é mais avançado nas microrregiões com melhores indicadores socioeconômicos, e como a tendência geral é de aumento na mortalidade por pneumonia entre idosos, esta correlação poderia ser justificada.

Esta possibilidade é reforçada por achados de estudo recente que mostrou relação semelhante entre as taxas de mortalidade de idosos por doenças do aparelho respiratório e diversos indicadores socioeconômicos⁶⁸. Contudo, como a razão de dependência de idosos, que representa o estágio de transição demográfica, não se mostrou correlacionado com estas taxas o envelhecimento populacional, esta pode não ser a causa indireta desta associação.

Em relação à mortalidade por doenças cerebrovasculares as tendências e padrões regionais vão ao encontro do observado em estudos anteriores^{15,69}. A elevada mortalidade por estas doenças reflete o controle ineficiente de fatores de risco no país, principalmente da hipertensão⁷⁰. Dados do Vigitel^{71,72} mostraram maior prevalência de hipertensão diagnosticada por médico em capitais do Sudeste em 2006, o que é condizente com as taxas de mortalidade por doenças cerebrovasculares mais elevadas nesta região, enquanto o aumento acelerado da prevalência de hipertensão nas capitais do Norte, Nordeste e Centro-Oeste e queda no Sudeste até 2015 refletem as tendências regionais destacadas no presente estudo.

A correlação positiva entre a taxa de mortalidade por doenças cerebrovasculares e a renda nos anos iniciais é condizente com os padrões e tendências das taxas aqui destacadas, mais elevadas em regiões mais desenvolvidas e mais envelhecidas. Outros estudos^{73,68} constataram relação similar entre as taxas de mortalidade de idosos por doenças do aparelho circulatório e indicadores socioeconômicos entre as unidades da federação no país, o que reforça a importância do componente etário, intimamente relacionado com o nível de desenvolvimento, no comportamento geral da mortalidade por esta doença no país e corrobora os padrões destacados no presente estudo.

Em contrapartida, os achados são contrastantes com diversos estudos que abordaram a mortalidade por doenças cerebrovasculares e hipertensivas em contextos específicos, como bairros de Porto Alegre⁷⁴ e do Rio de Janeiro⁷⁵, municípios com mais de cem mil habitantes⁷⁶ e os estados de São Paulo, Rio de Janeiro e Rio Grande do Sul⁷⁷. Esta discordância é um indicativo de que a correlação das condições socioeconômicas com a mortalidade por doenças cardiovasculares em grupos específicos é diferente daquela observada entre grupos. Neste sentido, embora entre as unidades da federação e microrregiões a mortalidade por doenças cerebrovasculares seja mais elevada em locais mais desenvolvidos, quando o nível de observação fica mais próximo do individual o contexto socioeconômico mais precário sobressai aumentando o risco de óbito para estas doenças, o que provavelmente está associado a menor disponibilidade, acessibilidade e qualidade dos serviços de saúde, ineficiência do controle dos fatores de risco, alimentação de pior qualidade, dificuldade para aquisição de medicamentos entre outros.

Os padrões e tendências regionais dos homicídios apontam para a manutenção do cenário observado em outros estudos^{78,79} que identificaram maior mortalidade por causas externas no sexo masculino, nas faixas etárias a partir dos 15 anos de idade e constataram os homicídios como principal causa de óbito entre as causas externas no país em 2000 e 2003, respectivamente. Em contrapartida, é discordante com o estudo de Reichenheim⁸⁰ que descreveu queda nas taxas de mortalidade por esta causa nas regiões Norte, Centro-Oeste e Sudeste e aumento no Nordeste e Sul, tendência discrepante à observada no presente estudo.

A correlação negativa entre as taxas de mortalidade por homicídios e o indicador de envelhecimento populacional é condizente com a maior mortalidade por esta causa na faixa etária jovem e adulta observada neste estudo e na literatura^{78,79}, assim como a correlação positiva e ascendente com o desemprego, também observado entre as capitais do país⁸¹, e condizente com a elevada frequência de homicídios em áreas mais pobres e de intensa desigualdade social^{82,83}.

Os homicídios estão associados a contextos distintos, ocorrendo em áreas de maior desigualdade social nos grandes centros urbanos⁸⁴⁻⁸⁶, e em conflitos de ocupação territorial⁸⁷ no interior da região Norte e em regiões rurais ou de fronteira entre unidades da federação. Neste contexto, apesar do grande repercussão midiática da violência nos grandes centros urbanos, como Rio de Janeiro e São Paulo, principalmente aquelas envolvidas com operações legais e combate ao tráfico de drogas, os maiores riscos de mortalidade por homicídios, e com tendência ao crescimento nos próximos anos, foram observados nas regiões menos desenvolvidas do país.

Os achados do presente estudo apontam para a existência de diversas realidades em relação à mortalidade dentro do território nacional, que apesar das melhorias gerais nos indicadores socioeconômicos, mantém um cenário de iniquidades. Evidenciou-se a influência do contexto socioeconômico e indicativos da influência da distribuição etária da população para a interpretação da magnitude da mortalidade por algumas causas. Contudo, as análises foram realizadas apenas nos anos censitários, o que pode não evidenciar flutuações no período intercensitário que podem estar relacionados com o comportamento das taxas a longo prazo, e a opção de organizar os resultados segundo regiões dificultou a compreensão dos cenários locais. Análises mais profundas que considerem as diferenças etárias nos óbitos pelas causas analisadas, a análise espaço-temporal que permitam uma melhor visualização das desigualdades e tendências locais, bem como a inserção de variáveis que representem também o acesso e o uso dos serviços de saúde podem favorecer a melhor compreensão da dinâmica da mortalidade no país.

É necessário reforçar que as doenças cerebrovasculares, neoplasias de pulmão e pneumonia merecem atenção especial nos próximos anos, pois mantendo-se as tendências e padrões regionais aqui observadas, é possível que já na próxima década as regiões Norte e Nordeste, que convivem com maior morbimortalidade por doenças transmissíveis e causas externas, apresentem também as maiores taxas de mortalidade por estas três doenças, tipicamente crônicas. Neste sentido, políticas públicas para o controle da hipertensão e adoção de hábitos saudáveis e para a redução antitabagismo devem ser intensificadas nestas regiões.

REFERÊNCIAS

1. Prata, P. R. A transição epidemiológica no Brasil. *Cad. Saúde Pública* **8**, 168–175 (1992).
2. Mendes, A. C. G., Sá, D. A., Miranda, G. M. D., Lyra, T. M. & Tavares, R. A. W. Assistência pública de saúde no contexto da transição demográfica brasileira: exigências atuais e futuras. *Cad Saúde Pública* **28**, 955–964 (2012).
3. Campolina, A. G., Adami, F., Santos, J. L. F. & Lebrão, M. L. A transição de saúde e as mudanças na expectativa de vida saudável da população idosa: possíveis impactos da prevenção de doenças crônicas. *Cad Saúde Pública* **29**, 1217–29 (2013).
4. Chaimowicz, F. A saúde dos idosos brasileiros às vésperas do século XXI: problemas, projeções e alternativas. *Rev. Saúde Pública* **31**, 184–200 (1997).
5. Lebrão, M. L. O envelhecimento no Brasil: aspectos da transição demográfica e epidemiológica. *Saúde Coletiva* **4**, 135–40 (2007).
6. Araújo, J. D. Polarização epidemiológica no Brasil. *Epidemiol. E Serviços Saúde* **21**, 533–538 (2012).
7. Paim, J., Travassos, C., Almeida, C., Bahia, L. & Macinko, J. The Brazilian health system: history, advances, and challenges. *The Lancet* **377**, 1778–1797 (2011).
8. Souza, C. Brazil's System of Local Government, Local Finance, and Intergovernmental Relations. *Pap. Part EngKaR Res. Proj.* **8070**, (2002).
9. *As causas sociais das iniquidades em saúde no Brasil*. (Editora FIOCRUZ, 2008).
10. IBGE - Instituto Brasileiro de Geografia e Estatística. *Síntese de Indicadores Sociais: Uma análise das condições de vida da população brasileira 2013*. (2013).
11. Neri, M. & Soares, W. Desigualdade social e saúde no Brasil. Social Inequality and Health in Brazil. *Cad. Saúde Pública* (2002).
12. *Epidemiologia das desigualdades em saúde no Brasil: um estudo exploratório*. (Ministério da Saúde, Fundação Nacional de Saúde, Centro Nacional de Epidemiologia ; Organização Pan-Americana da Saúde, Organização Mundial da Saúde, Programa Especial de Análises de Saúde, 2002).

13. SIM. Mortalidade - Brasil - 1979 a 1995. *DATASUS* (2017). Available at: <http://tabnet.datasus.gov.br/cgi/defptohtm.exe?sim/cnv/obt09uf.def>. (Accessed: 11th September 2017)
14. SIM. Mortalidade - Brasil - 1996 a 2015. *DATASUS* (2017). Available at: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/obt10uf.def>. (Accessed: 13th August 2017)
15. Souza, M. de F. M. de, Alencar, A. P., Malta, D. C., Moura, L. & Mansur, A. de P. Análise de séries temporais da mortalidade por doenças isquêmicas do coração e cerebrovasculares, nas cinco regiões do Brasil, no período de 1981 a 2001. *Arq Bras Cardiol* **87**, 735–40 (2006).
16. Cesse, E. Â. P., de Carvalho, E. F., de Souza, W. V. & Luna, C. F. Tendência da mortalidade por doenças do aparelho circulatório no Brasil: 1950 a 2000. *cities* **93**, 454–460 (2009).
17. Chor, D., Fonseca, M. de J. M. da & Andrade, C. R. de. Doenças cardiovasculares: comentários sobre a mortalidade precoce no Brasil. *Arq Bras Cardiol* **64**, 15–9 (1995).
18. World Health Organization (WHO). Policies and managerial guidelines for national cancer control programs. *Rev. Panam. Salud Pública* **12**, 366–370 (2002).
19. SVS/MS. *Evolução da Mortalidade no Brasil*. (2004).
20. Koifman, S. & Koifman, R. J. Environment and cancer in Brazil: an overview from a public health perspective. *Mutat. Res. Mutat. Res.* **544**, 305–311 (2003).
21. Wünsch, F. V. & Moncau, J. E. Mortalidade por câncer no Brasil 1980-1995: padrões regionais e tendências temporais. *Rev Assoc Med Bras* **48**, 250–7 (2002).
22. Guerra, M. R., Gallo, C. de M., Mendonça, G. A. & Silva, G. Risco de câncer no Brasil: tendências e estudos epidemiológicos mais recentes. *Rev Bras Cancerol* **51**, 227–34 (2005).
23. Fonseca, L. A. M., Eluf-Neto, J. & Wunsch Filho, V. Tendências da mortalidade por câncer nas capitais dos estados do Brasil, 1980-2004. *Rev. Assoc. Médica Bras.* **56**, 309–312 (2010).
24. De Souza, E. Homicídios no Brasil: o grande vilão da saúde pública na década de 80. *Cad. Saúde Pública* **10**, 45–60 (1994).
25. Souza, E. R. de & Lima, M. L. C. de. Panorama da violência urbana no Brasil e suas capitais. *Ciênc. Amp Saúde Coletiva* **11**, 363–373 (2006).
26. de Souza Minayo, M. C. Seis características das mortes violentas no Brasil. *Rev. Bras. Estud. Popul.* **26**, 135–140 (2013).
27. Freitas, E. D. de, Paim, J. S., Silva, L. M. V. da & Costa, M. da C. N. Trends and spatial distribution of mortality from external causes in Salvador, Bahia State, Brazil. *Cad. Saúde Pública* **16**, 1059–1070 (2000).

28. Beato Filho, C. C. *et al.* Conglomerados de homicídios e o tráfico de drogas em Belo Horizonte, Minas Gerais, Brasil, de 1995 a 1999 Homicide clusters and drug traffic in Belo Horizonte, Minas Gerais State, Brazil. *Cad Saúde Pública* **17**, 1163–1171 (2001).
29. de Mello Jorge, M. H. P., Gawryszewski, V. P. & Latorre, M. I-Análise dos dados de mortalidade. *Rev. Saúde Pública* **31**, 5–25 (1997).
30. Benguigui, Y. As infecções respiratórias agudas na infância como problema de saúde pública. *Bol. Pneumol. Sanitária* **10**, 13–22 (2002).
31. PostIII, P. & KnorstIV, M. Fatores de risco para internação por doença respiratória aguda em crianças até um ano de idade. *Rev Saúde Pública* **41**, 351–8 (2007).
32. Chiesa, A. M., Westphal, M. F. & Akerman, M. Doenças respiratórias agudas: um estudo das desigualdades em saúde Acute respiratory diseases: a study on health inequalities. *Cad Saúde Pública* **24**, 55–69 (2008).
33. Francisco, P. M. S. B., Donalisio, M. R. C. & Latorre, M. do R. D. de O. Impacto da vacinação contra influenza na mortalidade por doenças respiratórias em idosos. *Rev. Saúde Pública* **39**, 75–81 (2005).
34. Donalisio, M. R., Francisco, P. M. S. B., Latorre, M. do R. D. de & others. Respiratory mortality trend in the elderly before and after influenza vaccination campaigns, State of São Paulo-1980 to 2004. *Rev. Bras. Epidemiol.* **9**, 32–41 (2006).
35. Campagna, A. de S., Dourado, I., Duarte, E. C. & Daufenbach, L. Z. Mortalidade por causas relacionadas à influenza em idosos no Brasil, 1992 a 2005. (2009).
36. Prietsch, S. O. *et al.* Doença respiratória em menores de 5 anos no sul do Brasil: influência do ambiente doméstico. *Rev Panam Salud Publica* **13**, 303–10 (2003).
37. Bakonyi, S. M. C., Danni-Oliveira, I. M., Martins, L. C. & Braga, A. L. F. Poluição atmosférica e doenças respiratórias em crianças na cidade de Curitiba, PR. *Rev. Saúde Pública* **38**, 695–700 (2004).
38. Cançado, J. E. D. *et al.* Repercussões clínicas da exposição à poluição atmosférica. *J Bras Pneumol* **32**, S5–S11 (2006).
39. Fischer, T. K., Lima, D., Rosa, R., Osório, D. & Boing, A. F. A MORTALIDADE INFANTIL NO BRASIL: SÉRIE HISTÓRICA ENTRE 1994-2004 E ASSOCIAÇÃO COM INDICADORES SOCIOECONÔMICOS EM MUNICÍPIOS DE MÉDIO E GRANDE PORTE. *Med. Ribeirao Preto Online* **40**, 559–566 (2007).
40. Duarte, E. C., Schneider, M. C., Paes-Sousa, R., Silva, J. B. da & Castillo-Salgado, C. Expectativa de vida ao nascer e mortalidade no Brasil em 1999: análise exploratória dos diferenciais regionais. *Rev Panam Salud Publica* **12**, 436–44 (2002).
41. Szwarcwald, C. L., do Carmo Leal, M., de Castilho, E. A. & de Andrade, C. L. T. Mortalidade infantil no Brasil: Belíndia ou Bulgária? Infant mortality rate in Brazil: Belgium-India or Bulgaria? *Cad Saúde Públ* **13**, 503–516 (1997).

42. Paes, N. A. & Silva, L. A. A. Infectious and parasitic diseases in Brazil: a decade of transition. *Rev. Panam. Salud Pública* **6**, 99–109 (1999).
43. Malta, D. C. *et al.* Mortes evitáveis em menores de um ano, Brasil, 1997 a 2006: contribuições para a avaliação de desempenho do Sistema Único de Saúde. (2010).
44. DATASUS. Available at: <http://www2.datasus.gov.br/DATASUS/index.php?area=0205&id=6937>. (Accessed: 13th May 2016)
45. IBGE - Instituto Brasileiro de Geografia e Estatística. *Divisão do Brasil em Mesorregiões e Microrregiões Geográficas*. **1**, (1990).
46. Rodrigues, C. M., Silva, J. K. T., Martins, P. C., Terron, S. L. & de Souza Silva, R. L. *Divisão Territorial Brasileira*. (2002).
47. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. & Tatham, R. L. *Análise multivariada de dados - 6ed.* (Bookman Editora, 2009).
48. Ayres, M., Ayres Júnior, M., Ayres, D. L. & Santos, A. de A. Aplicações estatísticas nas áreas das ciências bio-médicas. *Inst. Mamirauá Belém* (2007).
49. R Development Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing).
50. Guimarães, Z. A., Costa, M. da C. N., Paim, J. S. & Silva, L. M. V. da. Declínio e desigualdades sociais na mortalidade infantil por diarreia. (2001).
51. Teixeira, M. da G. L. C., Meyer, M. de A., Costa, M. da C. N., Paim, J. S. & Silva, L. M. V. da. Mortalidade por doenças infecciosas e parasitárias em Salvador-Bahia: evolução e diferenciais intra-urbanos segundo condições de vida. (2002).
52. Malta, D. C. *et al.* Lista de causas de mortes evitáveis por intervenções do Sistema Único de Saúde do Brasil. (2007).
53. Malta, D. C. *et al.* Mortes evitáveis em menores de um ano, Brasil, 1997 a 2006: contribuições para a avaliação de desempenho do Sistema Único de Saúde. (2010).
54. SIM. Mortalidade - Brasil. *DATASUS* (2017). Available at: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/obt10uf.def>. (Accessed: 10th August 2017)
55. SINASC. Nascidos vivos - Brasil. (2017). Available at: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinasc/cnv/nvuf.def>. (Accessed: 2nd September 2017)
56. WHO. Global Health Observatory data repository - Diarrhoeal diseases. *WHO* (2016). Available at: <http://apps.who.int/gho/data/view.main.ghe2002015-CH3?lang=en>. (Accessed: 2nd September 2017)
57. Malta, D. C. *et al.* Lung cancer, cancer of the trachea, and bronchial cancer: mortality trends in Brazil, 1980-2003. *J. Bras. Pneumol.* **33**, 536–543 (2007).

58. Azevedo e Silva, G., Gamarra, C. J., Girianelli, V. R., Valente, J. G. & others. Tendência da mortalidade por câncer nas capitais e interior do Brasil entre 1980 e 2006. *Rev. Saúde Pública* **45**, 1009–1018 (2011).
59. Guerra, M. R. *et al.* Magnitude e variação da carga da mortalidade por câncer no Brasil e Unidades da Federação, 1990 e 2015. *Rev. Bras. Epidemiol.* **20**, 102–115 (2017).
60. Barros, A. J., Cascaes, A. M., Wehrmeister, F. C., Martínez-Mesa, J. & Menezes, A. M. B. Tobacco smoking in Brazil: regional inequalities and prevalence according to occupational characteristics. *Ciênc. Saúde Coletiva* **16**, 3707–3716 (2011).
61. Malta, D. C., Iser, B. P., Sá, N. N. & others. Tendências temporais no consumo de tabaco nas capitais brasileiras, segundo dados do VIGITEL, 2006 a 2011 [Trends in tobacco consumption from 2006 to 2011 in Brazilian capitals according to the VIGITEL survey]. *Cad Saude Publica* **29**, 812–22 (2013).
62. de Castro, M. S. M. & Vieira, V. A. Padrões espaço-temporais da mortalidade por câncer de pulmão no Sul do Brasil. *Rev Bras Epidemiol* **7**, (2004).
63. Holcman, M. M., Santos, J. L. F. & others. Evolução da mortalidade infantil na região metropolitana de São Paulo, 1980-2000. *Rev. Saúde Pública* **38**, 180–186 (2004).
64. Fischer, T. K., Lima, D., Rosa, R., Osório, D. & Boing, A. F. A mortalidade infantil no Brasil: série histórica entre 1994-2004 e associação com indicadores socioeconômicos em municípios de médio e grande porte. *Med. Ribeirao Preto Online* **40**, 559–566 (2007).
65. Nascimento, S. G. do, Oliveira, C. M. de, Sposito, V., Ferreira, D. K. da S. & Bonfim, C. V. do. Infant mortality due to avoidable causes in a city in Northeastern Brazil. *Rev. Bras. Enferm.* **67**, (2014).
66. Francisco, P. M. S. B., Donalisio, M. R. de C., Latorre, M. do R. D. de & others. Tendência da mortalidade por doenças respiratórias em idosos do Estado de São Paulo, 1980 a 1998. *Rev. Saúde Pública* (2003).
67. Oliveira, J. de F. M. de, Boing, A. F., Waldman, E. A. & Antunes, J. L. F. Ecological study on mortality from influenza and pneumonia before and after influenza vaccination in the Northeast and South of Brazil. *Cad. Saúde Pública* **29**, 2535–2545 (2013).
68. Paes, N. A. A geografia da mortalidade por causas dos idosos no Brasil. *Anais* 1–22 (2016).
69. Souza, M. F. M., Timerman, A., Serrano Jr, C. V., Santos, R. D. & Mansur, A. P. Tendências do risco de morte por doenças circulatórias nas cinco regiões do Brasil no período de 1979 a 1996. *Arq Bras Cardiol* **77**, 562–75 (2001).
70. Mansur, A. de P. *et al.* Transição epidemiológica da mortalidade por doenças circulatórias no Brasil. *Arq. Bras. Cardiol.* **93**, 506–510 (2009).
71. Ministério da Saúde. *VIGITEL BRASIL 2006 - VIGILÂNCIA DE FATORES DE RISCO E PROTEÇÃO PARA DOENÇAS CRÔNICAS POR INQUÉRITO TELEFÔNICO.* (2007).

72. Ministério da Saúde. *VIGITEL BRASIL 2015 - VIGILÂNCIA DE FATORES DE RISCO E PROTEÇÃO PARA DOENÇAS CRÔNICAS POR INQUÉRITO TELEFÔNICO*. (2017).
73. Alves, D. B. & Barbosa, M. T. S. Desigualdades na mortalidade por doenças crônicas entre idosos e sua associação com indicadores socioeconômicos no Brasil. *Rev. Bras. Ciênc. Envelhec. Hum.* **7**, 22–33 (2010).
74. Bassanesi, S. L., Azambuja, M. I. R. & Achutti, A. C. Mortalidade precoce por doenças cardiovasculares e desigualdades sociais em Porto Alegre: da evidência à ação. *Arq. Bras. Cardiol. Sao Paulo Vol 90 N 6 2008 P 403-412* (2008).
75. Alves, D. da S. B. Análise espacial da mortalidade de idosos por doenças crônicas no município do Rio de Janeiro. (2013).
76. Ishitani, L. H., Franco, G. da C., Perpétuo, I. H. O. & França, E. Desigualdade social e mortalidade precoce por doenças cardiovasculares no Brasil. *Rev Saúde Pública* **40**, 684–91 (2006).
77. Soares, G. P., Brum, J. D., Oliveira, G. M. M. de, Klein, C. H. & Silva, N. A. S. e. Evolution of Socioeconomic Indicators and Cardiovascular Mortality in Three Brazilian States. *Arq. Bras. Cardiol.* **100**, 147–156 (2013).
78. Gawryszewski, V. P., Koizumi, M. S. & de Mello-Jorge, M. H. P. As causas externas no Brasil no ano 2000: comparando a mortalidade e a morbidade Morbidity and mortality from external causes in Brazil, 2000. *Cad Saúde Pública* **20**, 995–1003 (2004).
79. Souza, E. R. de & Lima, M. L. C. de. The panorama of urban violence in Brazil and its capitals. *Ciênc. Saúde Coletiva* **11**, 1211–1222 (2006).
80. Reichenheim, M. E. *et al.* Saúde no Brasil 5: Violência e lesões no Brasil: efeitos, avanços alcançados e desafios futuros. *The Lancet* 75–89 (2015).
81. Bando, D. H. & Lester, D. An ecological study on suicide and homicide in Brazil. *Ciênc. Saúde Coletiva* **19**, 1179–1189 (2014).
82. Sousa, G. dos S. de *et al.* Social determinants and their interference in homicide rates in a city in northeastern Brazil. *Rev. Bras. Epidemiol.* **17**, 194–203 (2014).
83. Pereira, D. V., Mota, C. M. & Andresen, M. A. Social disorganization and homicide in Recife, Brazil. *Int. J. Offender Ther. Comp. Criminol.* 0306624X15623282 (2015).
84. Souza, E. de. Violência velada e revelada: estudo epidemiológico da mortalidade por causas externas em Duque de Caxias, Rio de Janeiro. *Cad. Saúde Pública* **9**, 48–64 (1993).
85. De Lima, M. L. C. & Ximenes, R. Violência e morte: diferenciais da mortalidade por causas externas no espaço urbano do Recife, 1991 Violence and death: differentials in mortality from external causes in Recife, Pernambuco. *Cad Saúde Pública* **14**, 829–840 (1998).

86. Freitas, E. D. de, Paim, J. S., Silva, L. M. V. da & Costa, M. da C. N. Trends and spatial distribution of mortality from external causes in Salvador, Bahia State, Brazil. *Cad. Saúde Pública* **16**, 1059–1070 (2000).
87. Andrade, L. T. de & Diniz, A. M. A. A reorganização espacial dos homicídios no Brasil e a tese da interiorização. *R Bras Est Pop* **30**, S171–S191 (2013).

Figura 1 – *Boxplots* das Taxas de Mortalidade Padronizadas por A. Doenças Infecciosas Intestinais, B. Neoplasias de Traqueia, Brônquios e Pulmões, C. Pneumonias, D. Doenças Cardiovasculares e E. Homicídios no Sexo Feminino por Região em 1991, 2000 e 2010

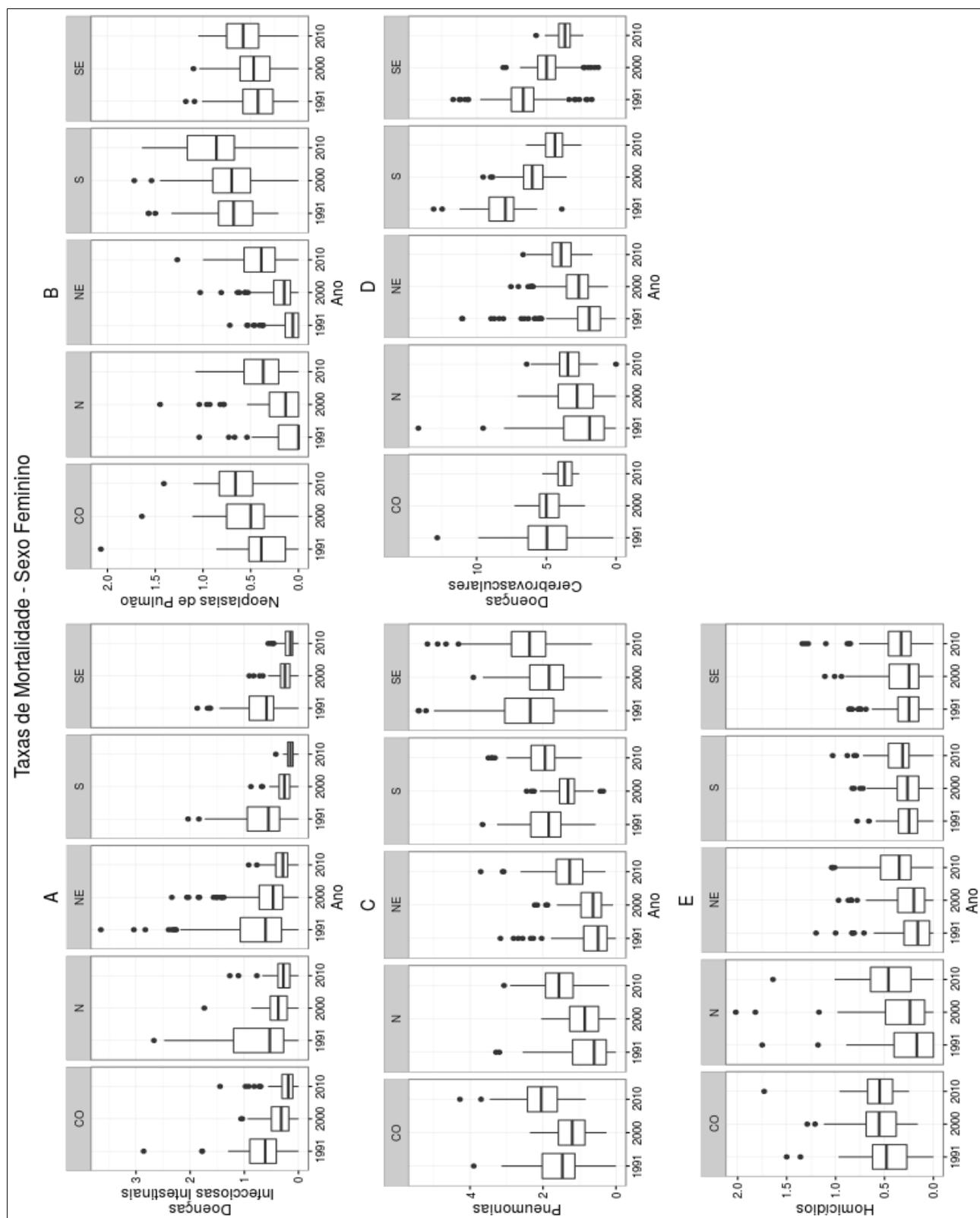
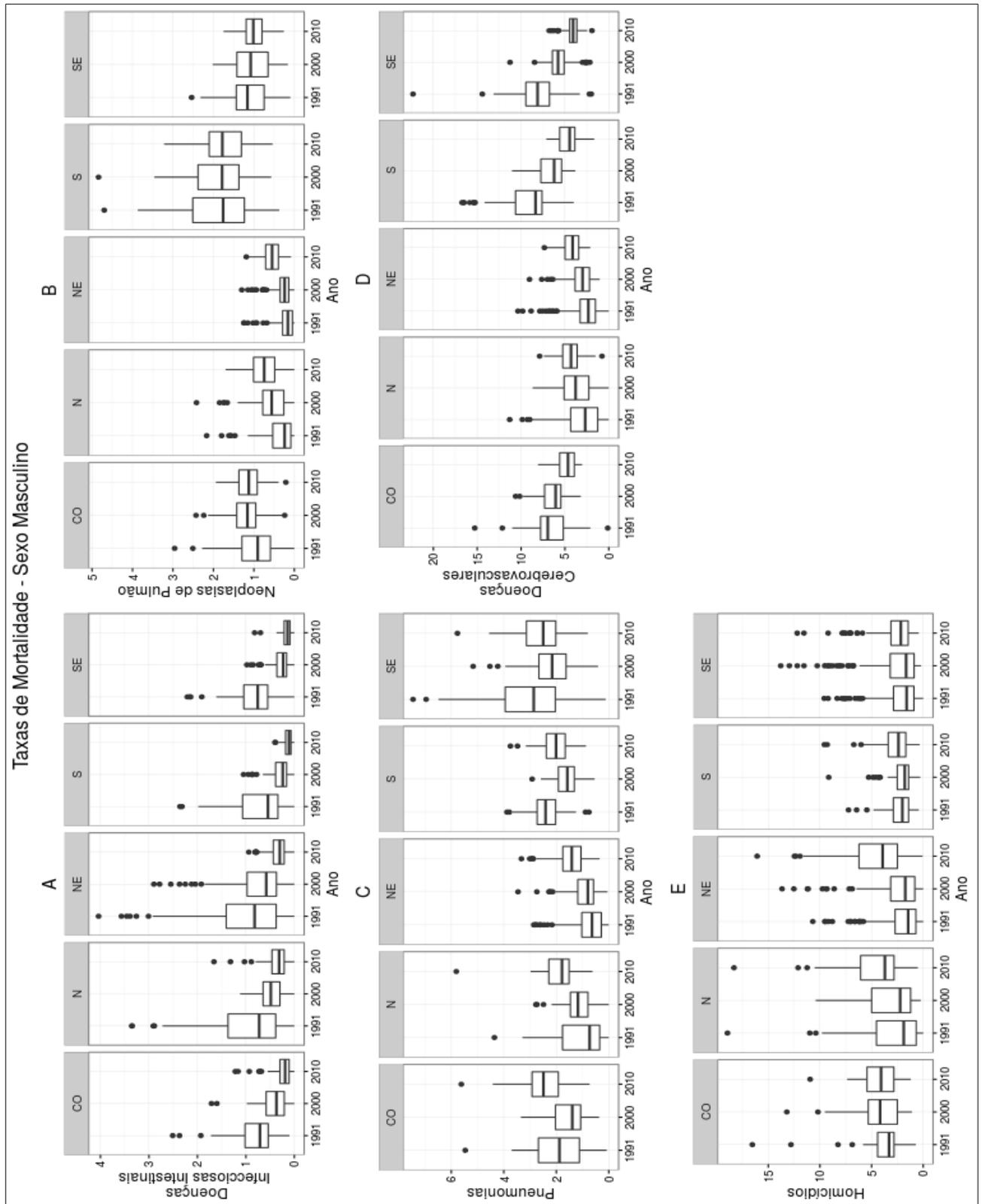


Figura 2 – *Boxplots* das Taxas de Mortalidade Padronizadas por A. Doenças Infecciosas Intestinais, B. Neoplasias de Traqueia, Brônquios e Pulmões, C. Pneumonias, D. Doenças Cardiovasculares e E. Homicídios no Sexo Masculino por Região em 1991, 2000 e 2010



INDICADORES	CARGA FATORIAL					
	1991		2000		2010	
	FATOR		FATOR		FATOR	
	I	II	I	II	I	II
Taxa de Analfabetismo	0,85		- 0,90	0,16	0,92	
Taxa de Desemprego		- 0,31		- 0,25	0,35	- 0,36
Percentual de Domicílios com acesso a Rede Geral de Esgoto ou Pluvial	- 0,66	0,25	0,66	0,23	- 0,53	0,39
Índice de Envelhecimento Populacional	- 0,34	0,94	0,52	0,85	- 0,33	0,94
Proporção de Pessoas com Baixa Renda	0,98		- 0,99		0,93	- 0,30
Razão de Dependência de Idosos	0,26	0,89		0,96	0,13	0,97
Razão de Dependência de Jovens	0,77	- 0,42	- 0,85	- 0,34	0,68	- 0,60
Renda Média Domiciliar Per Capita	- 0,99		0,97	- 0,13	- 0,94	0,18
Proporção da População que Coursou ao Menos o Ensino Médio	- 0,90		0,91	- 0,12	- 0,90	
AUTOVALOR	4,70	2,02	5,00	1,95	4,41	2,60
VARIÂNCIA ACUMULADA	74,7%		77,3%		77,9%	

Tabela 1 - Resultado da Análise Fatorial Exploratória com os Indicadores Sociais, Econômicos e Demográficos nos anos analisados

Figura 3 – *Boxplots* dos indicadores selecionados A. Renda Média Domiciliar Per Capita, B. Razão de Dependência de Idosos, C. Taxa de Desemprego, D. Percentual de Domicílios com Acesso a Rede Geral de Esgoto ou Pluvial, por Região nos anos 1991, 2000, 2010

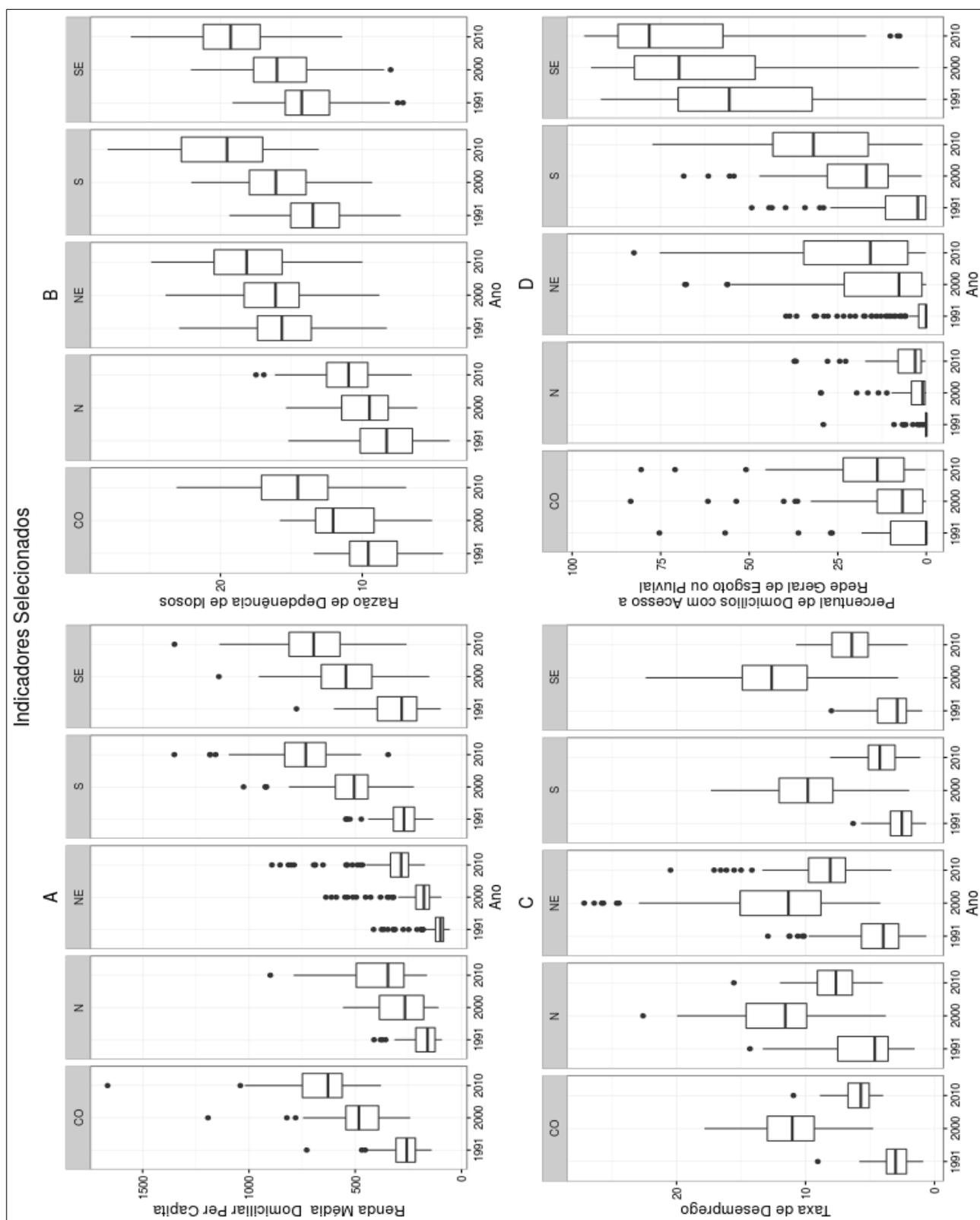


Figura 4 – Coeficientes de Correlação de Pearson das Taxas de Mortalidade por sexo com os Indicadores Sociais, Econômicos e Demográficos selecionados em 1991, 2000 e 2010

