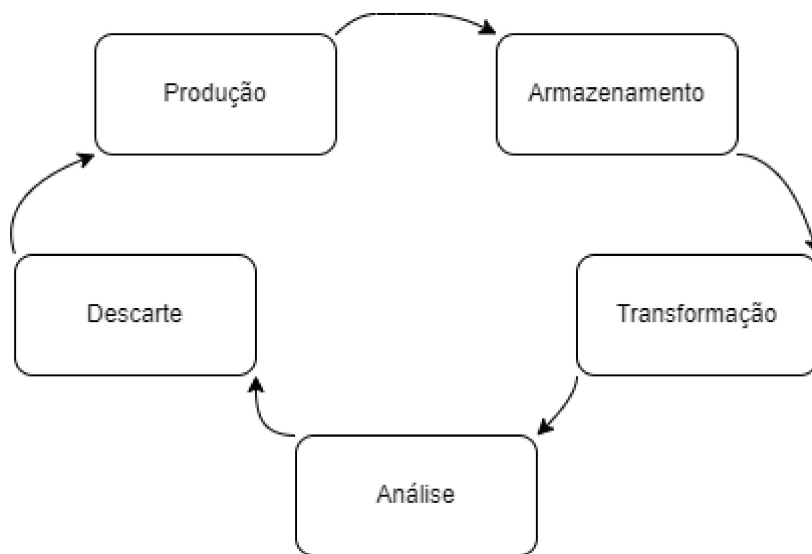


Mortalidade no Brasil no ano de 2020

A proposta deste projeto é aplicar o ciclo de vida do dado em uma base de dados do Ministério da Saúde do Brasil. A base escolhida para análise contém os registros de mortalidade no Brasil no ano de 2020.

O ciclo de vida do dado, citado anteriormente, diz respeito às fases que um conjunto de dados percorre dentro da ciência de dados. Cada fase deste ciclo é abordada em um capítulo deste relatório.



Produção

A base de mortalidade geral foi coletada do sistema OpenDataSus, por meio do pacote microdatasus, feito para linguagem R. A outra base, com os municípios brasileiros, foi baixada manualmente do Moodle Câmpus, ambiente virtual de apoio ao ensino presencial e a distância do IFSP. Os arquivos vieram em formato CSV.

Armazenamento

Ambas as bases foram armazenadas em datasets da Posit Cloud.

Transformação

A transformação necessária na base do Ministério da Saúde, como a codificação utilizada, ficou por parte do pacote microdatasus.

Análise de dados

A fase de análise foi baseada no modelo CRISP-DM, padrão internacional de mineração de dados. CRISP-DM é o acrônimo para Cross Industry Standard Process for Data Mining, que em tradução direta pode ser entendido como um padrão de processos de mineração de dados entre indústrias.

Ele especifica os passos necessários para o aproveitamento de dados a fim de se obter informações e conhecimento sobre eles.

Este padrão se constitui em seis fases:

- Entendimento do negócio
- Entendimento dos dados
- Preparação de dados
- Modelagem
- Avaliação
- Implantação

Entendimento do negócio

O Ministério da Saúde brasileiro desenvolveu o SIM, Sistema de Informação sobre Mortalidade, que unifica declarações de óbito emitidas no país desde 1979. Seu conjunto de informações serve de apoio para o desenvolvimento de políticas públicas com respeito à saúde da população.

Entendimento dos dados

Extraímos os dados de mortalidade geral em 2020, que conta com informações sobre a causa do óbito, local de ocorrência, e características físicas e socioeconômicas dos indivíduos de forma anonimizada.

Todos os óbitos registrados na base são do estado de São Paulo, não fetais.

Preparação de dados

Identificando os municípios

Durante a preparação dos dados, foi feita junção da base de mortalidade com uma base de dados sobre os municípios brasileiros. A ação foi necessária pois na base de mortalidade os municípios são referenciados pelo seu código, e a partir da junção foi possível identificá-los.

Gerenciamento de dados ausentes

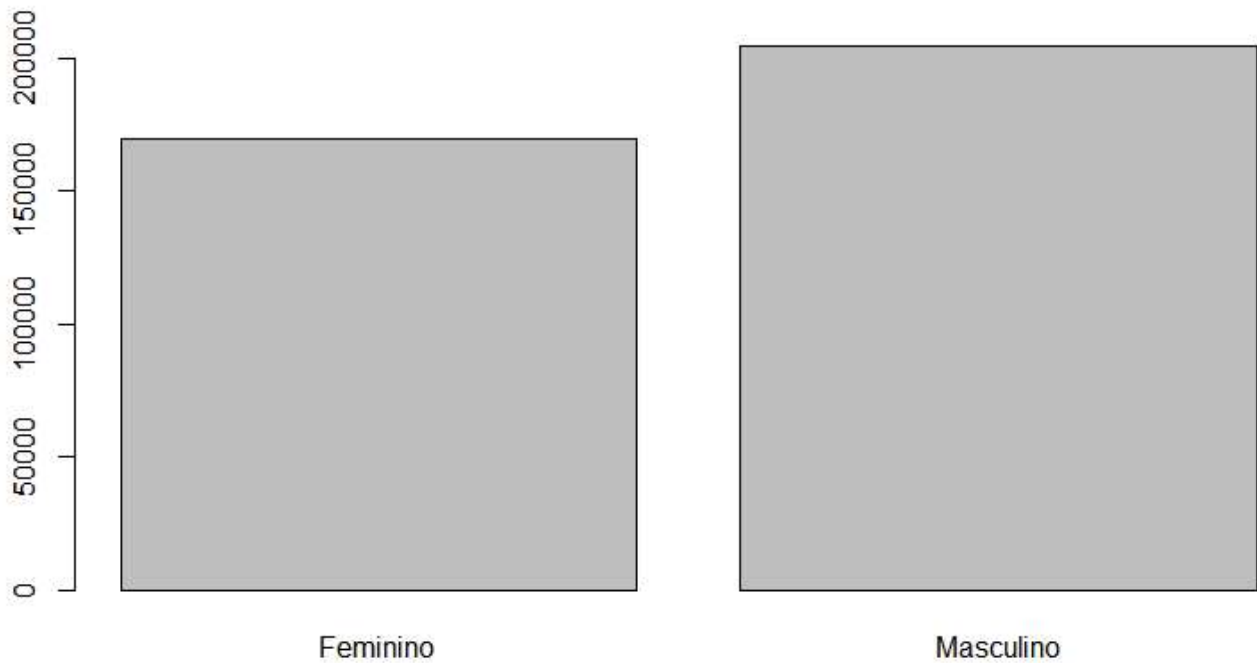
- apagar a linha?
- ou substituir valor?

Padronização dos dados

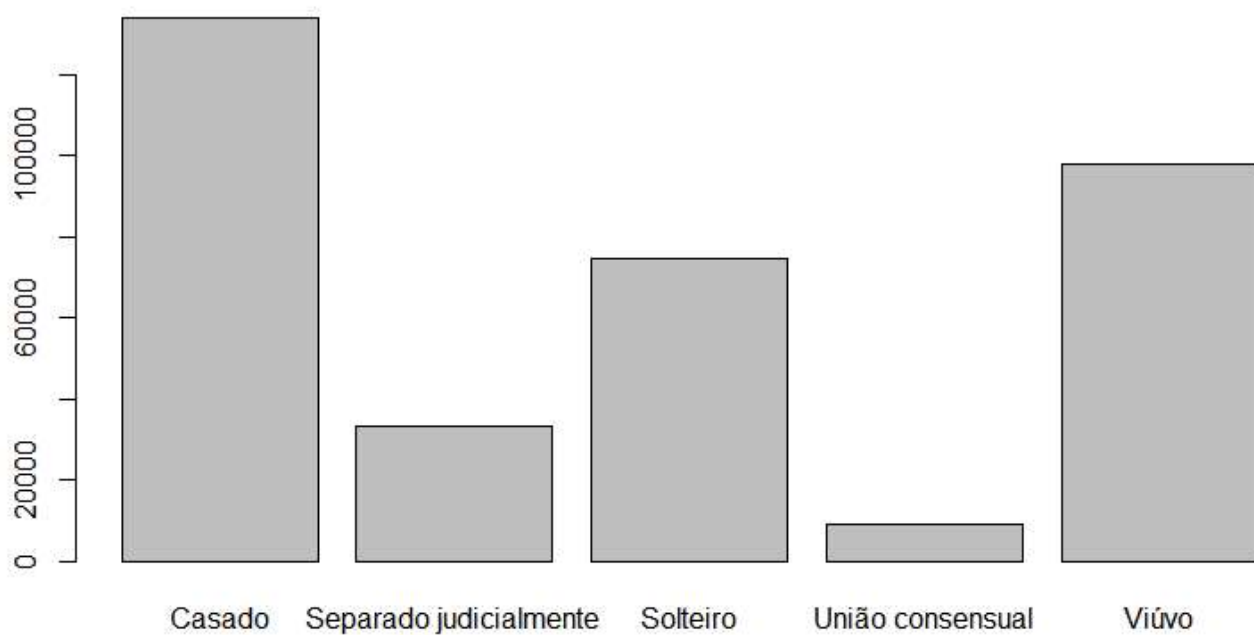
- IDADE: Todas as idades em dias, meses e de até um ano foram agrupadas na categoria "até um ano". Demais idades seguem em anos, até agrupar idades maiores ou iguais a 100 em "cem ou mais".

Modelagem

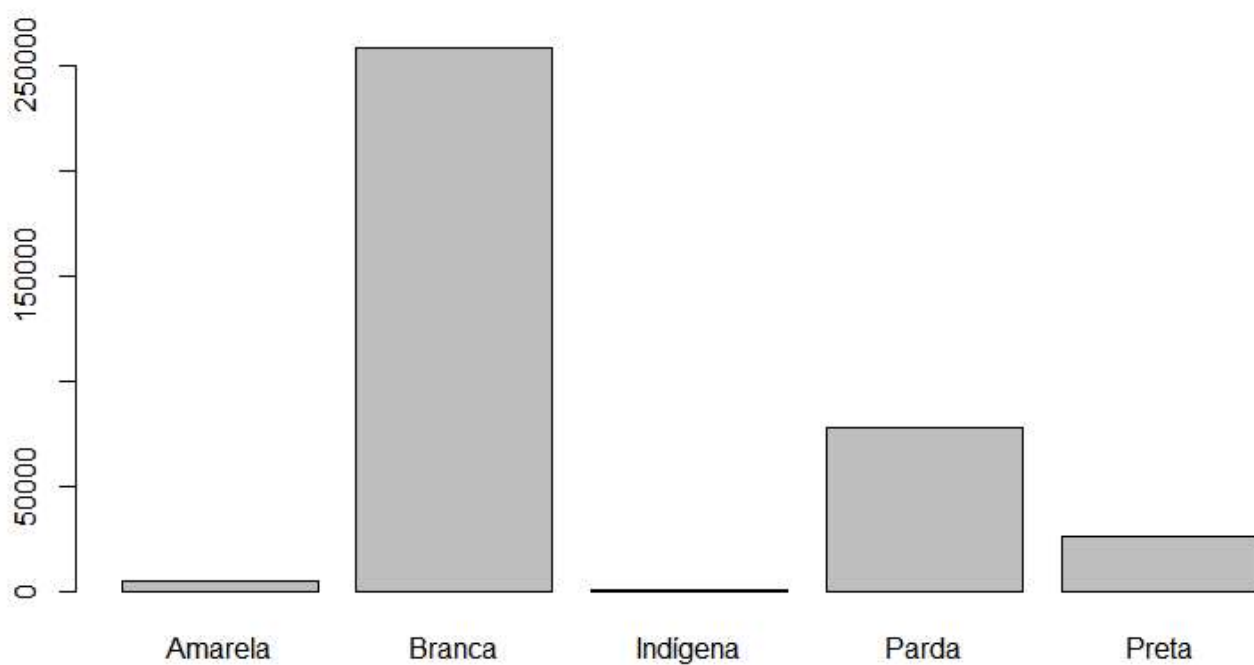
Homens morrem mais do que as mulheres



A maior parte da população morre casada



Apesar de no Brasil a maior parte da população se autodeclarar negra ou parda, são os brancos que mais morreram em São Paulo em 2020.



As principais causas de morte foram

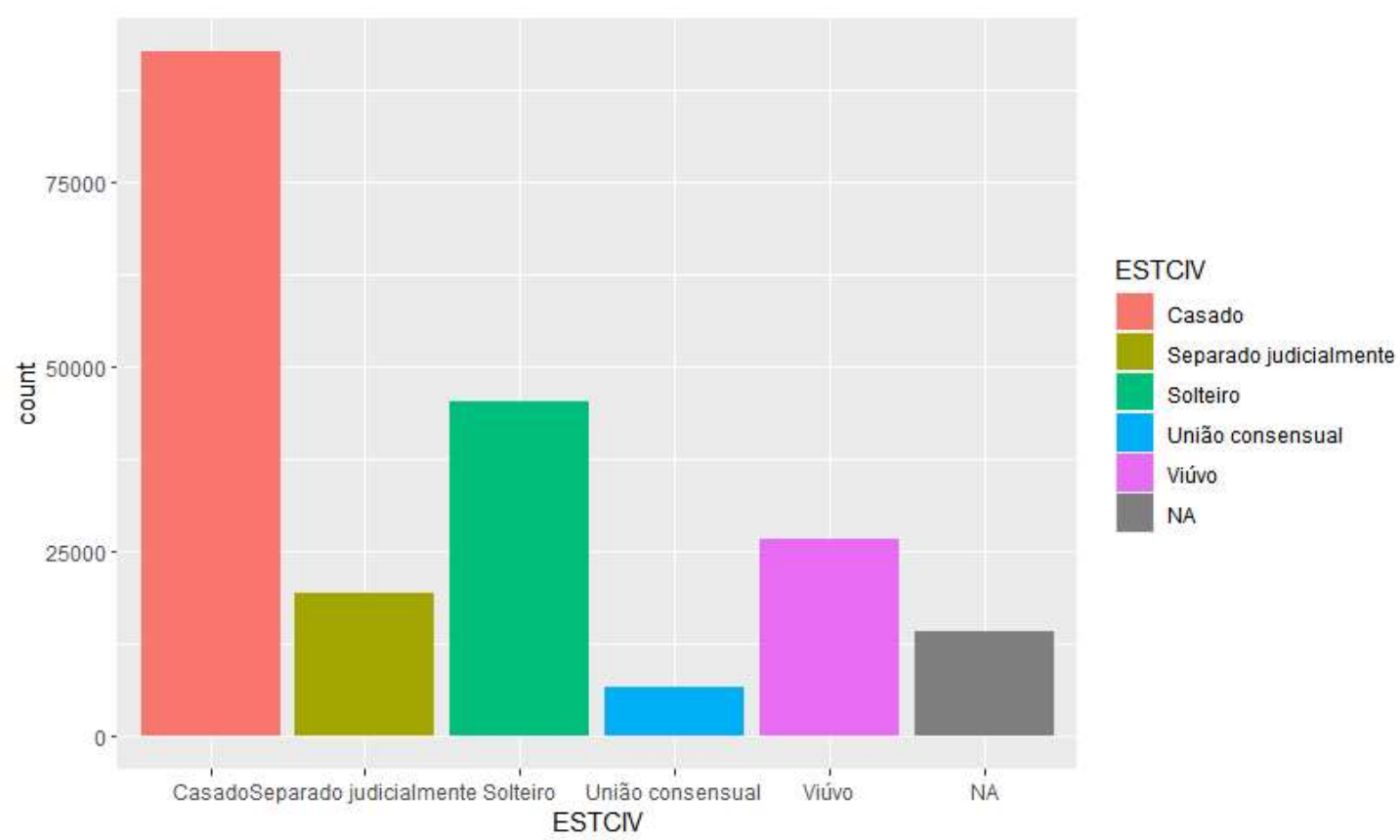
- Coronavírus

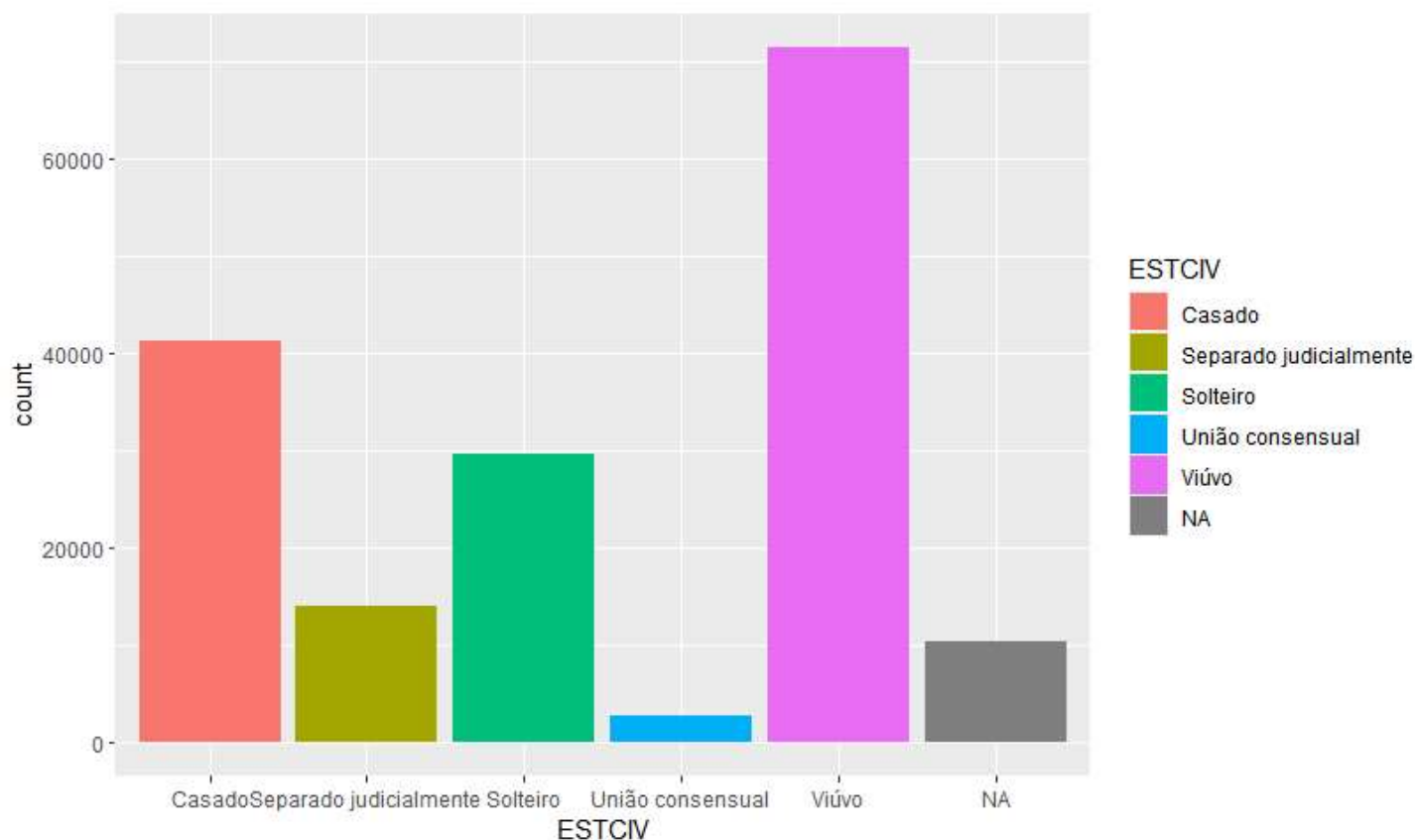
- Infarto
- Causas não especificadas,
- Demais transtornos respiratórios, diabetes,
- Neoplasia dos brônquios, infecção urinária
- AVC e alzheimer

Groups : CAUSABAS [3,627]

	CAUSABAS	n
	<fct>	<int>
1	B342	50722
2	I219	23354
3	R99	11305
4	J988	9085
5	J189	7920
6	E149	7050
7	C349	7005
8	N390	6934
9	I64	6767
10	G309	6574

A maioria dos homens que morrem, morrem casados, mas quanto as mulheres, em sua maioria ela morrem viúvas





As ocupações com mais óbitos são: Aposentados, donas de casa e pedreiros

# Groups: OCUP [1,631]	
OCUP	n
<fct>	<int>
1 Aposentado/Pensionista	<u>116126</u>
2 Dona de Casa	<u>59391</u>
3 Nao informada	<u>41393</u>
4 Ignorada	<u>17652</u>
5 Pedreiro	<u>8237</u>
6 Desempregado cronico ou cuja habitação habitual	<u>7750</u>
7 Comerciante varejista	<u>5787</u>
8 Empregado doméstico nos serviços gerais	<u>5118</u>
9 Representante comercial autônomo	<u>5084</u>
10 Motorista de carro de passeio	<u>4318</u>

Avaliação

Implantação

Descarte

Referências

Bases de dados utilizadas

- [Mortalidade Geral 2020](#)
- [Municípios brasileiros](#)
- [Estrutura do SIM](#)

Trabalhos semelhantes

- [Capítulos da Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde / CID-10](#)

Ferramentas utilizadas

- [Posit Cloud](#): Ambiente de desenvolvimento de modelos e armazenamento de datasets.
- [microdatasus](#): Pacote R para download de arquivos do OpenDataSUS.
- [read.dbc](#): Pacote necessário para usar a biblioteca microdatasus.
- [md2pdf](#): Conversão do README do depósito em um arquivo PDF.