

Trabalho Final

Amanda Carolina Wilmsen

Lazaro

PUCRS

1. Introdução

O seguinte trabalho consiste na coleta, preparação e análise de dados governamentais disponíveis através do DATASUS, tendo como objetivo a criação de uma visualização para esses dados em formato de *dashboard*, que dê suporte à análise da nossa hipótese

Para nossa análise levantamos a seguinte hipótese: Existe uma relação entre desigualdade de renda e taxas de analfabetismo? Com esse assunto começamos a investigar e encontramos dados relevantes do censo entre os anos de 1991, 2000 e 2010, assim conseguindo fazer uma comparação entre os anos e ver como foi a evolução dos números relacionados a isso. Com as bases encontradas conseguimos ter muitos dados referente a cada município e raça, tendo a oportunidade de ter uma investigação mais detalhada.

2. Bases de dados escolhidas

A base utilizada para a comprovação dessa hipótese foi o Sistema Demográfico e Socioeconômico que nos disponibiliza dados referentes à educação e renda dos censos realizados nos anos 1991, 2000 e 2010. Na base de educação tínhamos dados referentes ao total de pessoas por município e a quantidade de pessoas não alfabetizadas por município também, podendo fazer uma relação de taxa de analfabetismo. Por outro lado, na base de renda tínhamos a média de renda familiar per capita por município também, podendo assim relacionar esses dados.

Abaixo segue as duas fontes de dados que utilizamos:

- <https://datasus.saude.gov.br/educacao-censos-1991-2000-e-2010>
- <https://datasus.saude.gov.br/trabalho-e-renda-censos-1991-2000-e-2010>

3. Técnicas de integração, limpeza e transformação

Infelizmente a base utilizada para análise não estava disponível no PySUS, assim tendo que extrair os dados de forma manual para fazer a limpeza e transformação dos dados. Esse processo foi trabalhoso, porém tendo em vista que os dados que queríamos não possuíam algum tipo de padrão para coleta, não seria mais produtivo realizar alguma técnica de web scraping. Conseguimos fazer a exportação dos dados para arquivo CSV.

Com os dados em mãos utilizamos pandas para fazer a manipulação, limpeza e transformação dos dados. Foi feita análise para verificar se havia algum tipo de dado faltante ou não condizente com a realidade e foi constatado que a maioria dos dados estava preenchido e condizente, acreditamos muito que isso tenha ocorrido devido esses dados virem de pesquisas do censo, onde costumam ser mais assertivas.

A única dificuldade que encontramos foi nos dados em relação a cor e raça, devido ter a possibilidade de serem não definidos ou que não possuía informação para a raça. Ele tratava esses dados inserindo hífen ou três pontos, fizemos o tratamento desses dados para todos serem float, então para dados que não são definidos colocamos o número -1 e para quando não possui informação colocamos o equivalente a 0, assim preparando todos os dados necessários para a análise.

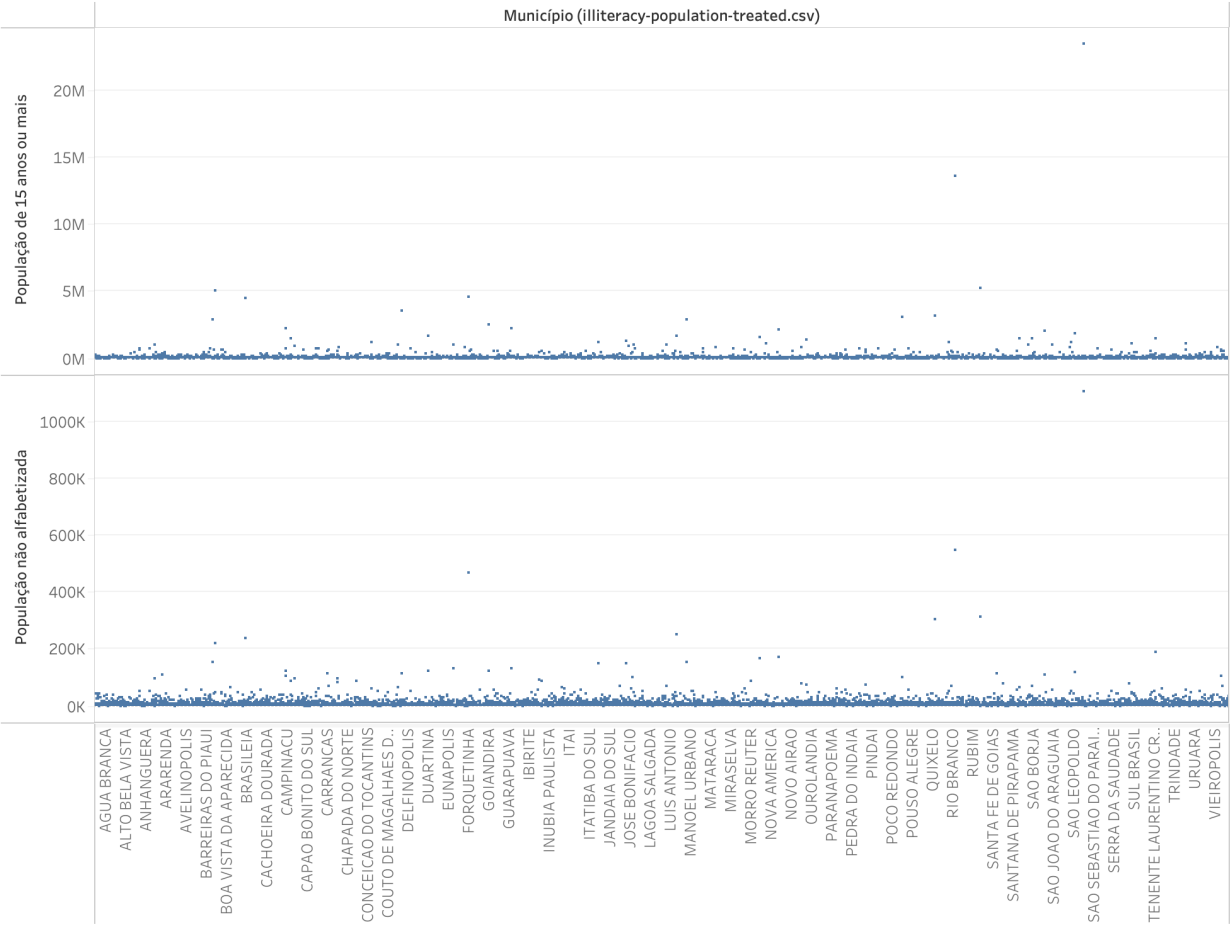
4. Conclusões

Com a análise dos dados dentro do software tableau chegamos a conclusão que a taxa de analfabetismo durante os anos aumentou consideravelmente, enquanto a renda per capita seguiu uma linha linear, assim concluindo que eles não possuem uma relação direta.

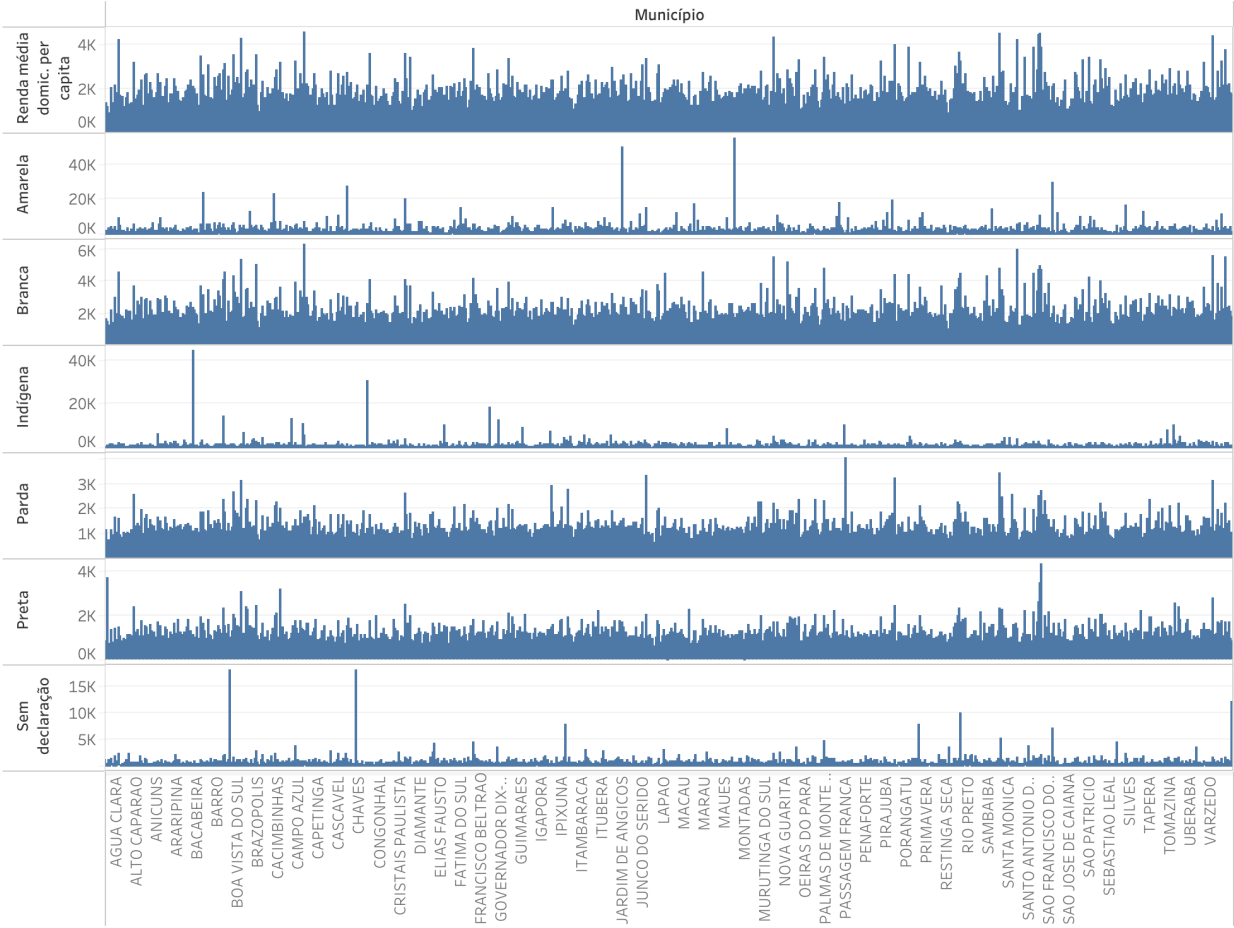
Todo o material utilizado está disponibilizado no github:

<https://github.com/amandacwilmsen/Analise-DATASUS>

População de 15 anos ou mais X População não alfabetizada por município



Renda média da população por cor



Taxa

