

Trabalho Prático - Banco de Dados Geográficos

Integrantes

- João Lucas Lage Gonçalves - 2020054552
- João Vítor Bicalho da Silva - 2020054579
- Pedro Renato Ferreira da Silva - 2020054757

Introdução

Este trabalho tem como objetivo colocar em prática conceitos e princípios discutidos ao longo da disciplina usando dados de votação, mantidos pelo Tribunal Superior Eleitoral.

Nosso grupo ficou responsável por analisar os dados de votação para Deputados Federais das eleições de 2022, no estado do Ceará, e relacioná-los com dados geográficos do estado.

A primeira etapa desse trabalho foi feita usando o [Github](#) para armazenamento dos arquivos e engenharia de dados, e o [OMT-G Designer](#) para a construção do Banco de Dados Geográficos.

Metodologia

1. Aquisição de dados

Nessa etapa realizamos a coleta dos dados necessários para as análises. Foram coletados os seguintes dados:

- [Dados de votação para Deputados Federais no estado do Ceará, nas eleições de 2022, disponíveis no site do TSE.](#)
- [Dados do Censo de 2022, disponíveis no site do IBGE.](#)
- [Dados do IDH municipal no ano de 2010, disponíveis no site da Fundação João Pinheiro.](#)
- [Dados geográficos do estado do Ceará disponíveis no site do IPEA.](#)
- [Dados de furtos no estado do Ceará entre 2013 e 2022, disponíveis no site da SSP-CE.](#)

Os dados foram coletados em formato CSV e estão disponíveis na pasta [dados](#).

2. Construção do BDG e Engenharia de Dados

Essa etapa foi dividida em duas partes: a Engenharia de Dados e a construção do Banco de Dados Geográficos.

Engenharia de Dados

A engenharia de dados foi realizada no arquivo [tratamento_de_dados.ipynb](#). Nesse arquivo, realizamos a limpeza e transformação dos dados de votação, censo, IDH e furtos, para que eles pudessem ser inseridos no Banco de Dados Geográficos. Além disso foi feita uma análise exploratória inicial dos dados a fim de verificar as estatísticas básicas sobre os dados.

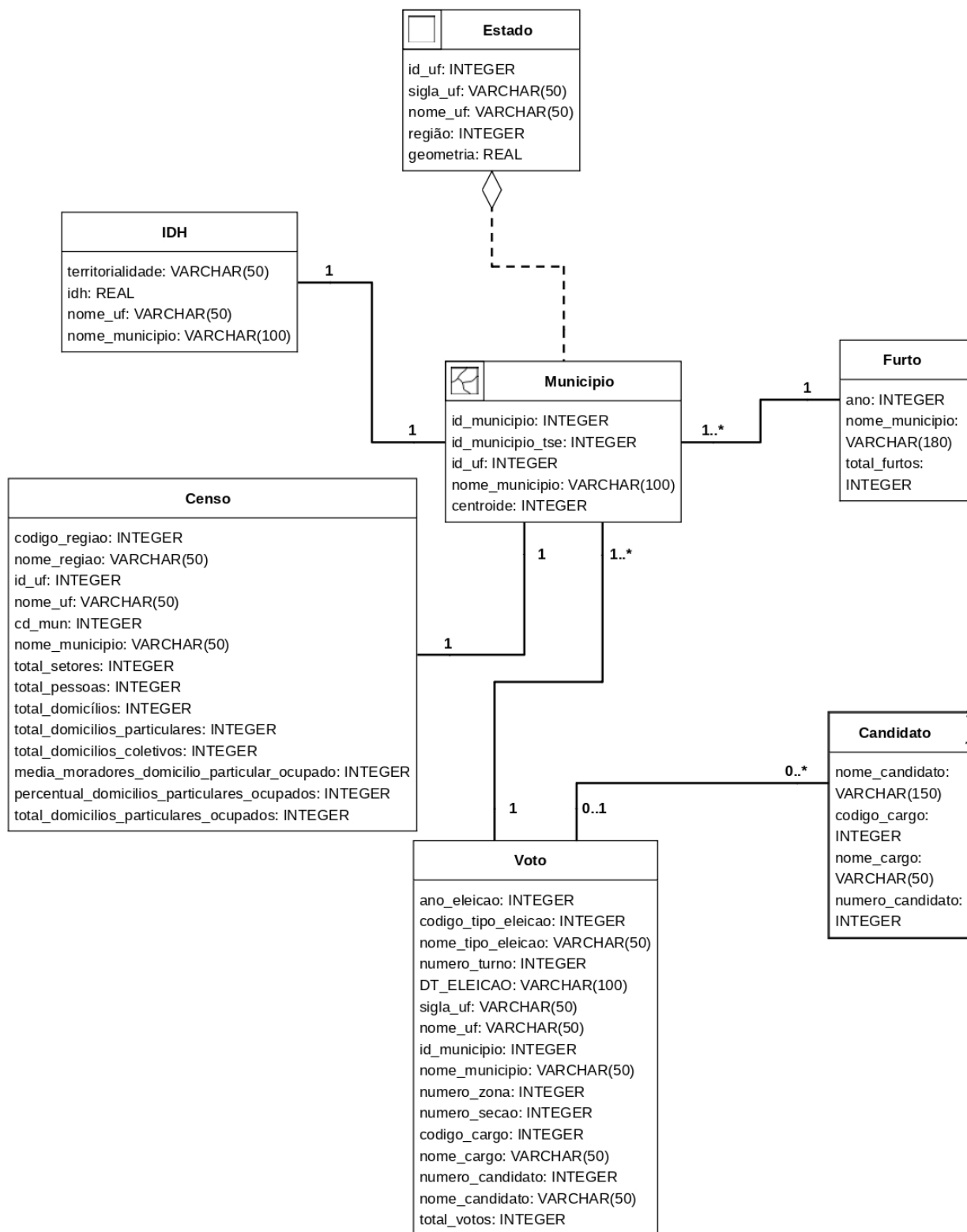
1. Carregamento dos dados
2. Padronização de nomes e tipos de dados das colunas
3. Filtragem dos dados

4. Agregação dos dados
5. Exploração dos dados
6. Construção das tabelas do Banco de Dados Geográficos

Esquema OMT-G

O diagrama construído pode ser melhor visualizado acessando o arquivo [OMT-G.pdf](#)

OMT-G Designer



1/1

3. Análises

Correlação com dados socioeconômicos

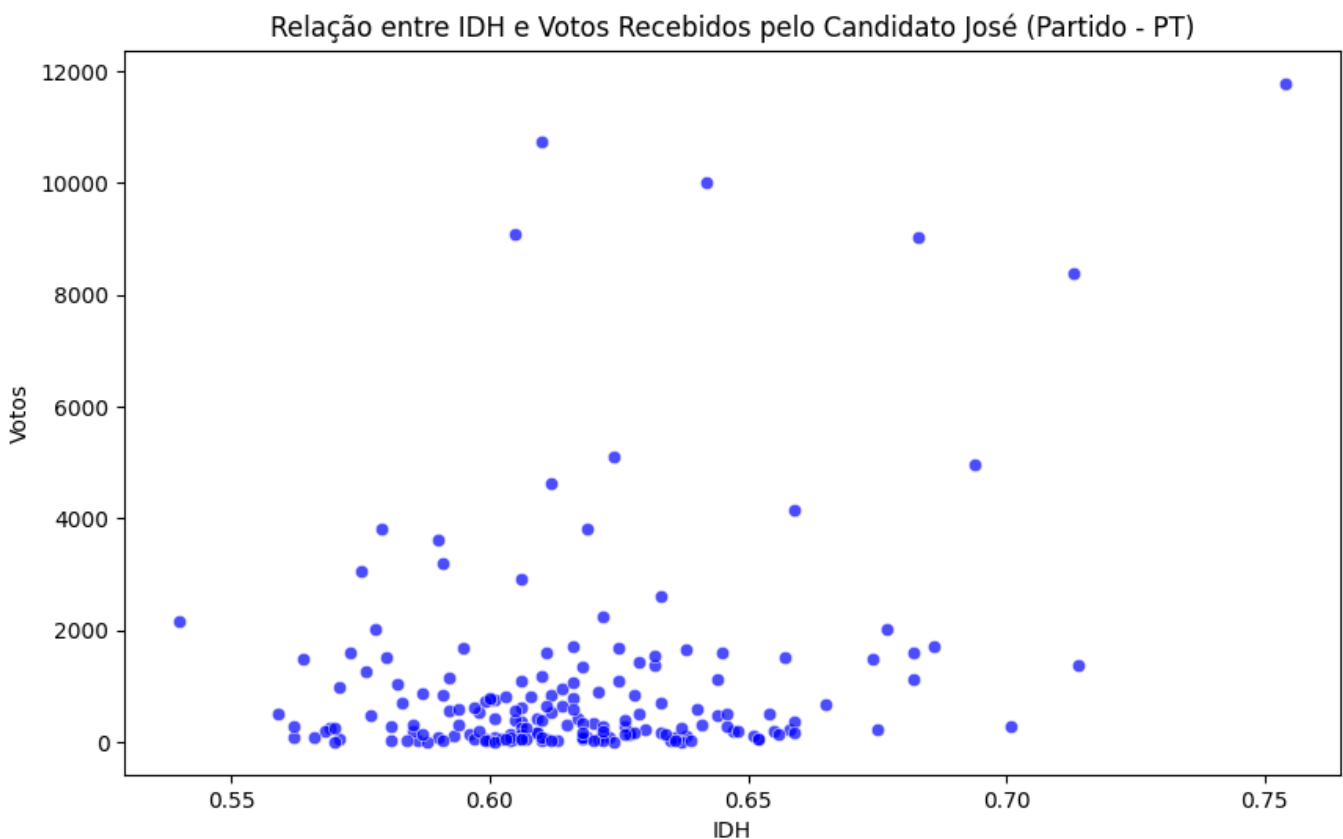
Para essa análise selecionamos um candidato a deputado federal pelo partido PL, André Fernandes, e outro do PT, José Nobre Guimarães. Ambos representam espectros políticos opostos e tiveram votações expressivas.

Considerando as variáveis **total de votos** e **idh** por cidade, obtivemos os seguintes coeficientes de correlação de Pearson:

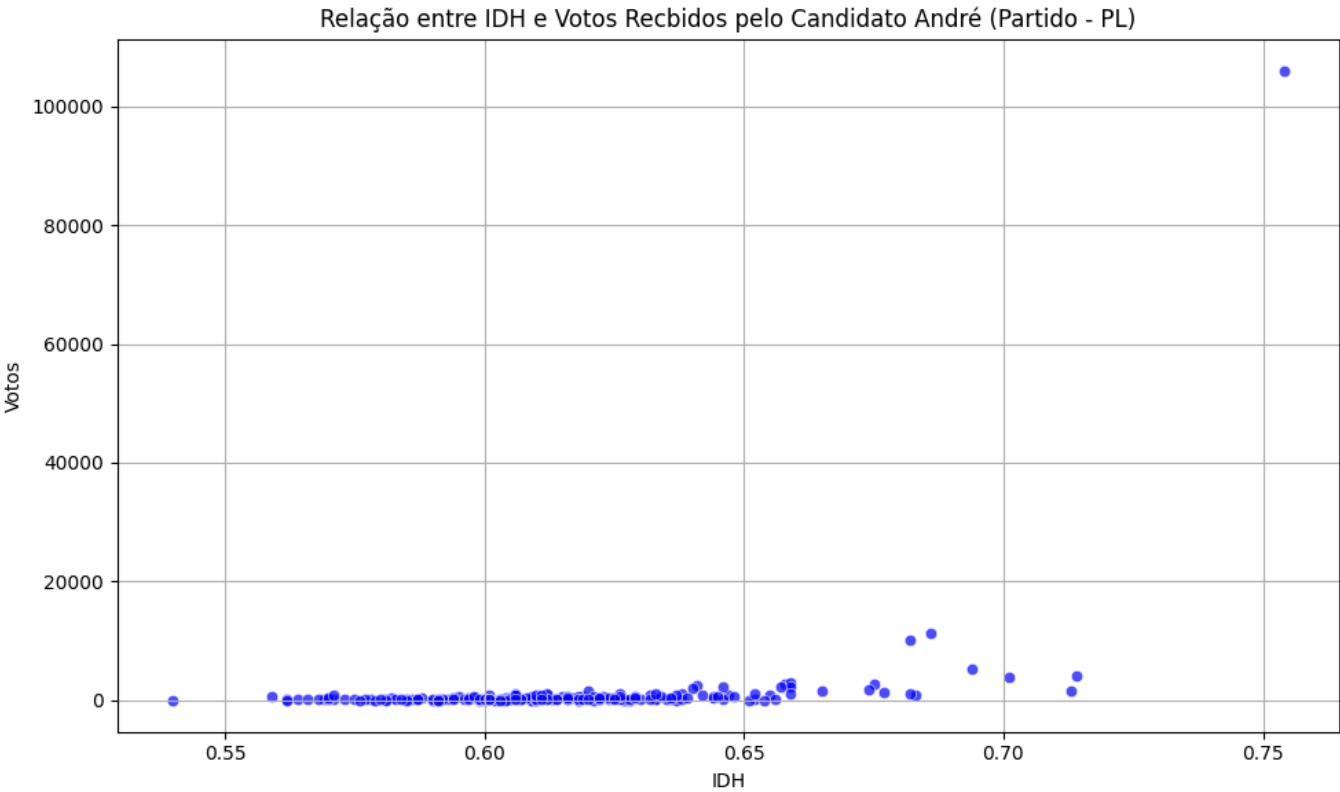
- Candidato Andre - Coeficiente de Correlação de Pearson: 0.40341885945941275
- Candidato José - Coeficiente de Correlação de Pearson: 0.30443282422165246

Os coeficientes obtidos indicam inicialmente que não há uma correlação entre as variáveis de idh e total de votos em cada cidade. No entanto, também foram obtidos p-values extremamente próximos de sero, ou seja, tais coeficientes não são estatisticamente significativos.

Observando os gráficos de dispersão, podemos extrair algumas outras informações interessantes:



Nota-se que o candidato José possui uma quantidade de votos maior em cidades com IDH menor



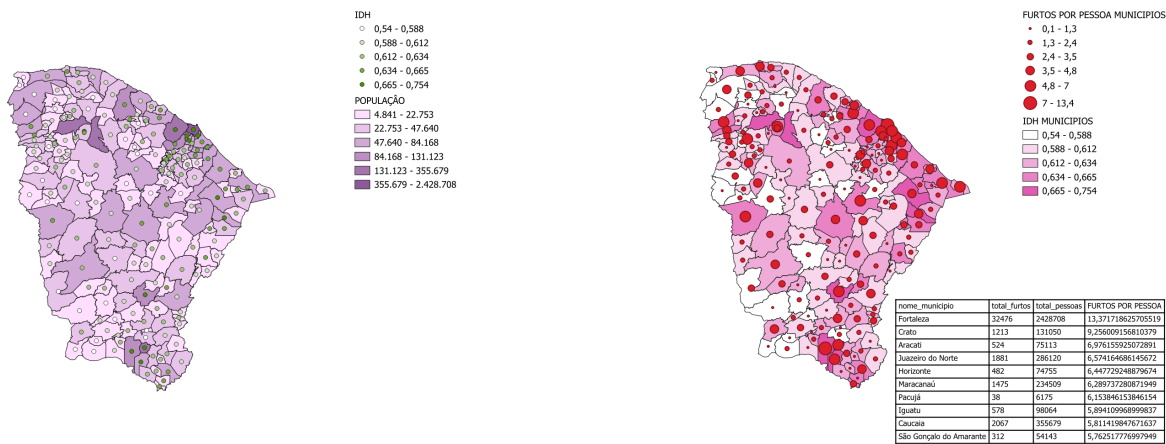
No caso do candidato André, há uma grande parcela de votos na cidade de maior IDH, Fortaleza.

4. Visualização

A parte de visualização foi feita usando o programa QGIS. Para isso, foi necessário importar os arquivos `.csv` na pasta `tabela` para o banco de dados PostgreSQL e, em seguida, conectar o banco de dados ao QGIS.

Os arquivos na pasta `camadas` os dados processados em camadas para serem utilizados para visualização no programa. Além disso os arquivos `.qgz` contêm os projeto do QGIS com as camadas já carregadas.

4.1 Dados Gerais



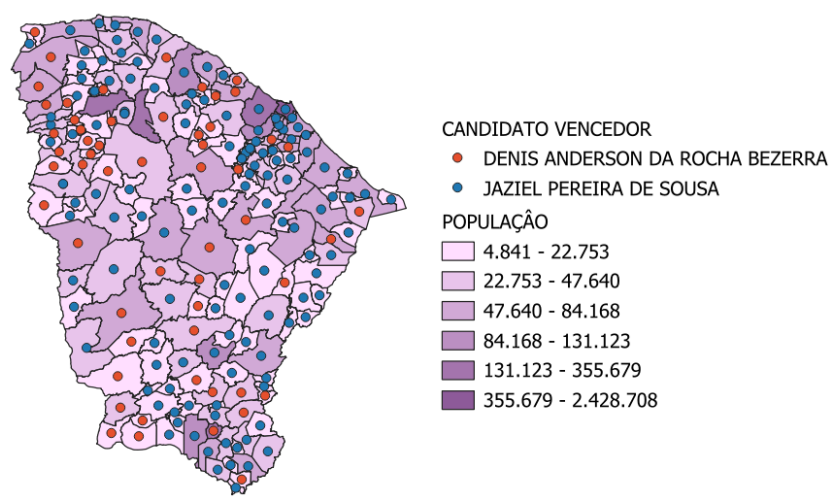
Visualização da relação entre tamanho da população e IDH dos municípios do Ceará

Visualização da relação entre quantidade de furtos e IDH dos municípios do Ceará

As visualizações acima mostram que população e IDH aparentam ter uma correlação positiva, dado que municípios com maior população parecem ter um IDH maior. Já a quantidade de furtos e IDH não apresentam

uma correlação clara, diferente do que se pode pensar, não são os municípios com menor IDH que possuem maior taxa de furtos.

4.2 Dados de Votação - Candidatos



Comparação de municípios com mais votos entre os candidatos Jeziel (eleito com menor número de votos) e Denis (não eleito com maior número de votos).



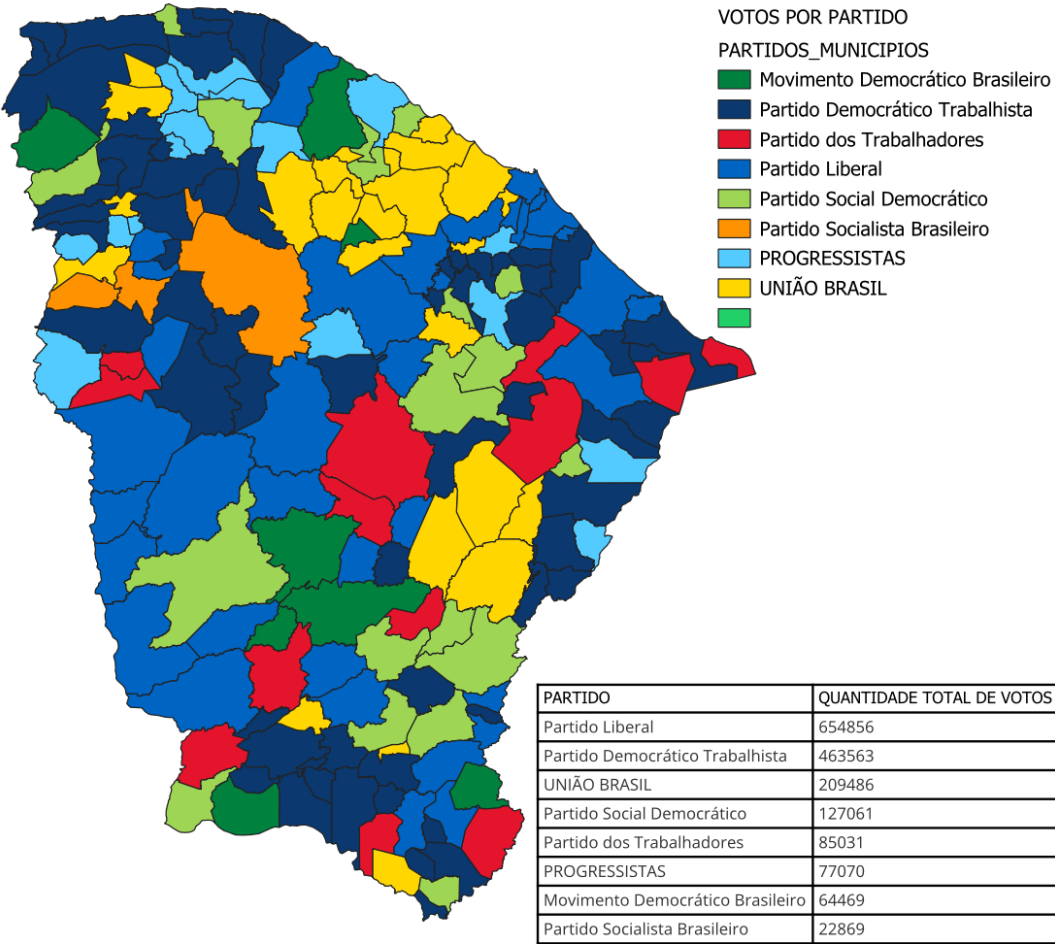
Quantidade de votos por município do candidato Jeziel ...

Quantidade de votos por município do candidato Denis

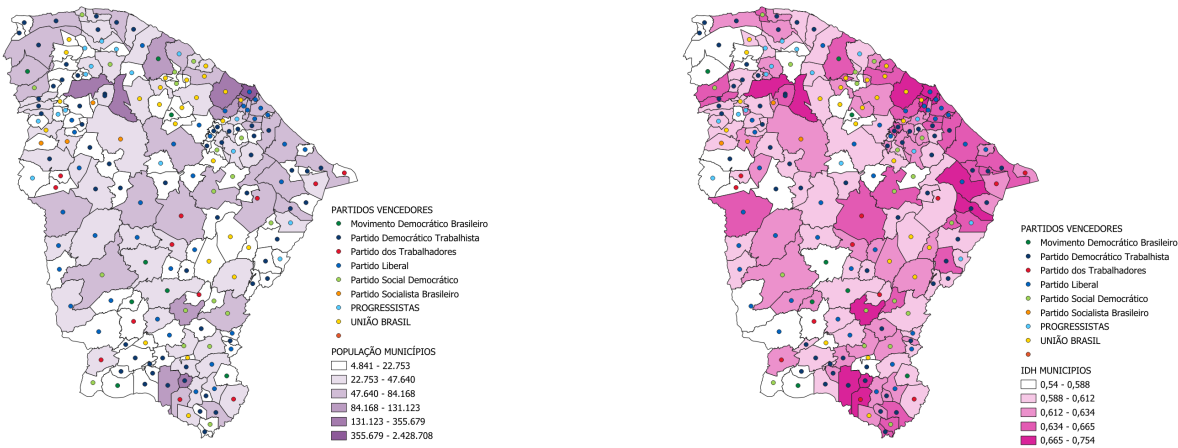
As visualizações acima mostram que o candidato Jeziel, eleito com menor número de votos, teve uma votação mais pulverizada, com menos municípios mas com maior quantidade de votos, além de "vencer" em

cimdades de maior população. Já o candidato Denis, não eleito com maior número de votos, teve uma votação mais concentrada, com poucos municípios com maior quantidade de votos, porém municípios de menor população.

4.3 Dados de Votação - Partidos



Partido vencedor por município



Comparação entre partido vencedor e população do município

Comparação entre partido vencedor e IDH do município

As visualizações acima mostram que o Partido Liberal (PL) foi o partido vencedor em vários municípios de população média para grande, ficando com o maior número de votos total. Já o Partido Democrático

Trabalhista (PDT) foi o partido vencedor em municípios com grande população, ficando em segundo no número total de votos. Com relação a IDH dos municípios, não há uma correlação clara entre o partido vencedor e o IDH.