



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Disciplina: Bancos de Dados Geográficos e
Ciência de Dados Geoespaciais

Professor: Clodoveu Davis

Trabalho Prático 2

Valor: Parte 1 – 15 pontos / Parte 2 – 25 pontos

DADOS ELEITORAIS ESPACIALIZADOS

O presente trabalho prático visa colocar em prática conceitos e princípios discutidos ao longo da disciplina, usando dados e situações reais. O trabalho será baseado em dados de votação, mantidos pelo Tribunal Superior Eleitoral, e publicados em seu portal de dados abertos (<https://dadosabertos.tse.jus.br/>).

Serão formados grupos de até 5 componentes, sendo que alunos de pós-graduação atuarão sozinhos. Cada grupo examinará os dados de um estado, e os relatórios finais serão compartilhados com todos os alunos via Moodle. A atribuição de estados a grupos será aleatória. Não serão admitidos mais de 20 grupos (o Distrito Federal e os estados com menos municípios – SE, AM, RO, AC, AP, RR – não serão trabalhados).

Os conjuntos de dados publicados pelo TSE são amplos e volumosos, portanto o TP se concentrará nas eleições de 2022, em votos para candidatos a Deputado Estadual por estado.

Os dados dos resultados eleitorais estão disponíveis para download em formato CSV. Existem datasets para cada estado da federação, porém estes trazem totais por seção eleitoral. Será necessário fazer a agregação espacial dos resultados em três níveis: município, região geográfica imediata e região geográfica intermediária. Em seguida, integrar os dados às divisas espaciais, de modo a permitir análises e visualizações geográficas.

Dados socioeconômicos deverão ser usados nas análises (ex. escolaridade, renda, etc.). Idealmente, deve-se usar dados municipais do Censo 2022, porém certos resultados não foram divulgados ainda. Em último caso, usar dados do Censo 2010. Dados censitários estão disponíveis no site do IBGE. O IDH municipal também pode ser usado (fonte usual: Fundação João Pinheiro)

Além dos dados eleitorais e socioeconômicos, o grupo deverá usar dados de mais uma fonte, à sua escolha, nas análises. Sugere-se definir essa fonte depois de fazer uma análise exploratória inicial. Exemplos: criminalidade, saúde pública, vacinação, educação, clima, meio ambiente, etc., com dados em nível municipal.

O objetivo do trabalho é que cada aluno de pós ou grupos de até 5 alunos de graduação construa um BDG, partindo dos dados eleitorais citados, integre esses dados com outros, e produza análises e visualizações ao final. Os itens a seguir detalham essas etapas e apresentam sugestões de dados a incorporar.

Especificações

Parte 1

1. Aquisição de dados

- a. Identificação e conhecimento da fonte primária: TSE, conforme descrito acima;
- b. Identificação de fontes de dados geográficos para espacialização dos dados eleitorais; obtenção e verificação de qualidade.
- c. Obtenção de dados censitários: escolher temas e fonte.
- d. Aquisição de dados do tema complementar.

2. Construção do BDG e Engenharia de Dados

Nesta parte, espera-se que os dados sejam analisados de forma exploratória, de modo a adquirir conhecimento e estatísticas básicas sobre o conteúdo. É uma atividade típica de engenharia de dados, dentro da Ciência dos Dados.

- a. Organizar, filtrar, agregar, etc. os dados adquiridos, formando tabelas geográficas e convencionais relacionadas a elas. Apresentar um esquema OMT-G do BD formado.
- b. Importar os dados, de modo que características geográficas possam ser usadas nas análises.
- c. Executar quaisquer transformações necessárias nos dados originais, de modo a prepará-los para as análises.
- d. Análise exploratória – Verificar, por exemplo:
 - i. Ocorrência de dados faltantes
 - ii. Limitações dos dados
 - iii. Necessidade de correção de problemas de padronização nas colunas
 - iv. Verificação das faixas de variação dos valores das variáveis, estatísticas básicas sobre as variáveis (média, desvio padrão, etc.), valores possíveis de atributos categóricos
 - v. Análise crítica do conteúdo disponível

3. **Entrega:** relatório descrevendo o processo de formação do BD para análise e resultados da análise exploratória.

Parte 2

4. Análise

- a. **Autocorrelação espacial:** dados de candidatos escolhidos, ex. um candidato com votação regionalizada, um candidato com votação ampla; candidato eleito versus candidato não-eleito
- b. **Autocorrelação espacial em nível agregado:** agregar os dados em regiões geográficas imediatas e em regiões geográficas intermediárias (IBGE), e refazer a análise de autocorrelação espacial. Observar e reportar preservação ou alteração do comportamento constatado na análise anterior
- c. **Correlação com dados socioeconômicos:** verificar se existe correlação entre a votação de alguns candidatos escolhidos pelo grupo e características socioeconômicas dos municípios, e se essa correlação é espacializada
- d. **Correlação com dados da fonte adicional** escolhida pelo grupo
- e. **Autocorrelação espacial por partido,** na votação para deputado federal: escolher um grupo pequeno de partidos mais representativos, e compará-los
- f. **Análise extra:** à escolha do grupo (p. ex. comparação entre duas ou mais eleições, formação de grupos de partidos alinhados politicamente, agrupamento de municípios de perfil semelhante)

5. Visualização

(Discutiremos em sala alternativas de visualização para esses tipos de dados.)

Incluir na descrição da visualização:

- a. Propósito da visualização
 - b. Escolha do plano de fundo
 - c. Escolha de parâmetros gráficos/visuais
 - d. Processo de produção (software, recursos, opções relevantes)
6. **Entrega:** relatório apresentando as análises realizadas e contendo os produtos de visualização. No relatório, incluir link para o conjunto de dados construído para o trabalho, em formato shapefile ou backup do PostGIS, ou ainda como um repositório numa plataforma de compartilhamento, como GitHub ou Zenodo; apresentação de slides para discussão em aula.

Avaliação

- **Primeira etapa:** itens 1 e 2 (aquisição de dados, engenharia de dados) – entrega até 30/10/2025, valor 15 pontos
- **Segunda etapa:** itens 3 e 4 (análise e visualização) – entrega até 25/11/2025, valor 25 pontos.
- **A ausência de entrega da segunda etapa anula a pontuação da primeira etapa**