

Trabalho 1 BDGEO - Diogo Neiss

Estado de Santa Catarina

1. Fontes de dados utilizadas

1.1. Dados administrativos

1.2. Dados eleitorais

1.3. Indicadores socioeconômicos

1.4. Base de dados extra: Mortalidade

1.5. Dados geométricos e dados associados as regiões

1.6. Anexo: Fontes de dados estudadas e desconsideradas

 1.6.1. Dados de alfabetização

 1.6.2. Setor censitário

 1.6.3. Renda e trabalho

2. Diagrama OMT-G

3. Análise de dados

 3.1. Verificação de qualidade dos dados

 3.1.1. Corretude da associação de dados geométricos em diferentes níveis

 3.1.2. Inconsistência de chaves estrangeiras entre tabelas

 3.1.3. Corretude da agregação espacial eleitoral

 Exemplo candidato Edson Renato Dias entre as duas fontes

 3.2. Tabelas e análise

4. Transformações de dados

 4.1. Metodologia da engenharia de dados

 4.2. Dados eleitorais

 4.3. Dados de educação

 4.4. Preparação de dados geográficos com município

 4.5. Otimizações gerais

 4.5.1. Limpeza de tabelas e colunas

 4.5.2. Índices

5. Amostras QGIS

 5.1. Agrupamento espacial de níveis de região

 5.2. Município SC

 5.3. Município em overlay com regiões

1. Fontes de dados utilizadas

De acordo com o processo de aquisição e análise exploratória dos dados, decidi utilizar as seguintes fontes de dados

1.1. Dados administrativos

Para a junção de dados de domínios diferentes é importante ter tabelas com dados administrativos de cada município, permitindo a junção adequada posterior.

Diretórios Brasileiros: Tabelas ligando diversos códigos institucionais e informações de entidades brasileiras

<https://basedosdados.org/dataset/33b49786-fb5f-496f-bb7c-9811c985af8e?table=dffb65ac-9df9-4151-94bf-88c45bfcb056>

A tabela gerada mais útil é a *br_bd_diretorios_brasil_municipio*, que contém identificadores para quase todas as outras

Diretório ligando diversos códigos institucionais de municípios brasileiros: IBGE, Receita Federal, TSE, BCB, regiões, comarcas, região de saúde, etc.

1.2. Dados eleitorais

Baixei tabelas com dados a nível da votação do candidato em uma seção eleitoral e outros mais gerais, como perfil do eleitorado, votos totais por seção, etcs. Acabei mantendo apenas a tabela de candidato em seção, as outras foram mantidas mas não tem uso, isso veio de uma tentativa infrutífera de fazer a análise combinando seções em zonas eleitorais e juntando-as com as vizinhanças dos municípios.

Resultados - 2022: Contém totalizações, minha principal fonte de informações ao utilizar <https://dadosabertos.tse.jus.br/dataset/resultados-2022>

Perfil do eleitorado:

<https://dadosabertos.tse.jus.br/dataset/eleitorado-2022>

Dados extras para referência:

Resultados por Candidato e Município: Dados de votação em eleições brasileiras a nível candidato-município. Usando para conferência dos resultados

https://basedosdados.org/dataset/eef764df-bde8-4905-b115-6fc23b6ba9d6?table=741973_11-35e3-4800-88ea-e0eec2b1b764

1.3. Indicadores socioeconômicos

Selecionei duas fontes de dados interessantes

Indicador Nível Socioeconômico (INSE) do município: A tabela indica o Indicador de Nível Socioeconômico (INSE) ao nível do município a partir das informações de posse de bens domésticos, renda e contratação de serviços pela família dos alunos e pelo nível de escolaridade de seus pais nas respostas do questionário socioeconômico de exames como SAEB e ENEM

<https://basedosdados.org/dataset/bff9bdbd-d15e-45f1-83dc-a2ea3f1ee340?table=3775acd-10cd-45bd-9a23-09eb44df29f3>

Município: Tabela de indicadores no nível de município

<https://basedosdados.org/dataset/08a1546e-251f-4546-9fe0-b1e6ab2b203d?table=707fd42e-95e0-4856-922f-fcbb55db913a>

1.4. Base de dados extra: Mortalidade

Sistema de Informações sobre Mortalidade (SIM): Foi criado pelo DATASUS para a obtenção regular de dados sobre mortalidade no país. A partir da criação do SIM foi possível a captação de dados sobre mortalidade, de forma abrangente, para subsidiar as diversas esferas de gestão na saúde pública. Com base nessas informações é possível realizar análises de situação, planejamento e avaliação das ações e programas na área.

<https://basedosdados.org/dataset/5beec93-cbf3-43f6-9eea-9bee6a0d1683?table=0e00424f-abc3-4ef3-b72c-2428696b4e70>

1.5. Dados geométricos e dados associados as regiões

Com os dados de município adequadamente tratados, podemos pensar na agregação espacial em níveis superiores. Utilizei a seguinte fonte de dados

Dados Geográficos Brasileiros (geobr): Um pacote R que permite que os usuários acessem facilmente os shapefiles do Instituto Brasileiro de Geografia e Estatística (IBGE) e outros conjuntos oficiais de dados espaciais do Brasil. O pacote inclui uma ampla variedade de conjuntos de dados geográficos como recursos simples, disponíveis em várias escalas geográficas e por vários anos.

<https://basedosdados.org/dataset/49ace9c8-ae2d-454b-bed9-9b9492a3a642?table=3880670f-eceb-47ec-802b-4579ee62ae3b>

Recuperei as seguintes tabelas geométricas:

- Limite Vizinhança

- Município
- Região
- Região Imediata
- Região Intermediária

Com isso, é possível juntar as tabelas de dados simples e obter as visualizações desejadas.

1.6. Anexo: Fontes de dados estudadas e desconsideradas

Seção pode ser pulada, estão aqui para eventual referência futura

1.6.1. Dados de alfabetização

Alfabetização por Sexo, Raça e Grupo de Idade: Tabela 9542 - Pessoas de 15 anos ou mais de idade, total e as alfabetizadas, por sexo, cor ou raça e grupos de idade:

<https://basedosdados.org/dataset/08a1546e-251f-4546-9fe0-b1e6ab2b203d?table=cf9537b5-6198-455f-a8b0-7c762e94d79c>

É bem ruim, falta dado populacional. Tentei recorrer ao Sidra mas o parse disso daria muito trabalho

Exemplo no Sidra para brasil:

<https://sidra.ibge.gov.br/tabela/9542#/n1/all/v/allxp/p/all/c59/all/c2/6794/c86/all/c287/2999,3000,9482,9483,9484,93086,93087,100362/l/v,p+c59+c2,t+c86+c287>

Apenas SC e tudo:

<https://sidra.ibge.gov.br/Tabela/k/727659884>

Dados de todos os municípios

<https://sidra.ibge.gov.br/tabela/9542#/g/44/v/allxp/p/all/c59/all/c2/6794/c86/all/c287/2999,3000,9482,9483,9484,93086,93087,100362/l/v,p+c59+c2,t+c86+c287>

1.6.2. Setor censitário

Aqui temos duas tabelas bem interessantes, porém o dicionário de dados é complexo e as informações não são tão interessantes, por isso descartei

Setor censitário com dados geométricos: Informações tabulares sobre os totais de pessoas e de domicílios por espécie, bem como sobre a média de moradores em domicílios particulares ocupados, por Setores Censitários preliminares

<https://basedosdados.org/dataset/08a1546e-251f-4546-9fe0-b1e6ab2b203d?table=c5c43530-0c76-4226-b67c-9e0d73b66d12>

Pesada demais para baixar diretamente. Selecionei apenas estado SC no big query, exportei para csv, montei no container docker e usei o seguinte comando para importar

```
COPY eleicoes22.br_ibge_censo_2022_setor_censitario_sc FROM  
'/var/local/censitario/br_ibge_censo_2022_setor_censitario_sc.csv' DELIMITER ''  
CSV HEADER;
```

Diretório de dados do Setor censitário: Diretório de referência para os setores censitários. Contem o id_setor_censitario, que permite ligar informações (município, distrito, etc) com as tabelas do Censo 2022, bem mais compacta que a anterior

<https://basedosdados.org/dataset/33b49786-fb5f-496f-bb7c-9811c985af8e?table=2250a925-60fa-4a37-983a-c513b018265c>

1.6.3. Renda e trabalho

Setor Censitário Domicílio Renda 2010: Informações de renda dos domicílios por setor censitário

<https://basedosdados.org/dataset/b8e8bd62-4eb9-42f9-9ffa-b5cca093f58e?table=260602ef-5106-4dbf-9f13-15a01b34b1d3>

Dados incompletos, descartei o uso, todas as colunas úteis vem com null

Tentei usar essa do IBGE mas deu muito trabalho o parse

<https://sidra.ibge.gov.br/tabela/10265#/g/87/v/allxp/p/all/c2/5/c1568/all/c11913/31721,31724,31727,31731,79369,79374,79377,79378,96165/l/v,p+c2+c1568,t+c11913>

2. Diagrama OMT-G

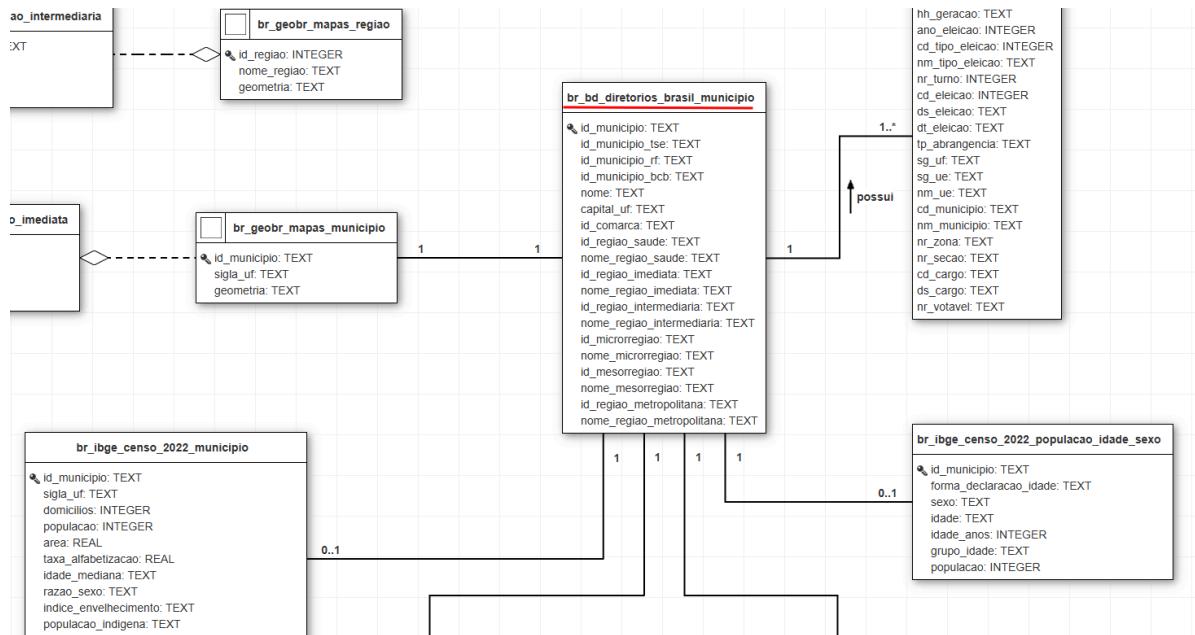
Fiz um script que a partir do banco de dados e seu schema gera as classes do diagrama omt-g, automaticamente organizando-os por número de colunas e altura de cada linha.

Depois disso, excluí tabelas desnecessárias e agrupei elementos geométricos.

Por fim, criar os relacionamentos/agregações e organizar o diagrama.

Amostra:

(xml e pdf completos em anexo)



A tabela *br_bd_diretorios_brasil_municipio* é o centro de tudo. As relações com as outras tabelas é direta por *id_município*, e praticamente todas também poderiam ter associação espacial ao conectar-se com *br_geobr_mapas_municipio*, não criei as conexões pela potencial complexidade do diagrama.

As cardinalidades são “0..1” pois não há uma restrição formal de associação formal entre elas, algumas tabelas associadas foram filtradas para ter dados apenas de Santa Catarina e economizar espaço. A tabela central sempre tem cardinalidade 1 pois é semanticamente a fonte de verdade do banco de dados.

3. Análise de dados

3.1. Verificação de qualidade dos dados

3.1.1. Corretude da associação de dados geométricos em diferentes níveis

Para isso utilizei o QGis, importando diferentes camadas e verificando sua posição no espaço e associação com outros elementos geográficos

3.1.2. Inconsistência de chaves estrangeiras entre tabelas

Foi feita uma análise de inconsistência entre cruzamentos de dados de diferentes fontes para garantir a consistência, de forma que nada fique de fora nas junções. Como as tabelas foram construídas por importação e conversão de dados, sem chaves estrangeiras explícitas, essa etapa é necessária.

Por exemplo, para ver o mapeamento entre dados de votação do TSE e uma base de dados processada baixada de outra fonte, foi utilizada a seguinte consulta

```
-- Query achar ids TSE sem mapeamento para id municipio
SELECT m.*, v.* FROM votacao_secao_2022_sc v
    full outer join br_bd_diretorios_brasil_municipio m
        on v.cd_municipio = m.id_municipio_tse
    where m.sigla_uf = 'SC' and
        (v.cd_municipio is null or m.id_municipio_tse is null);
```

Essa outra tem a mesma verificação, porém para uma base de dados censitária

```
-- Query achar ids de municipio na tabela censitaria sem mapeamento
SELECT m.* FROM br_bd_diretorios_brasil_municipio m
    full outer join br_ibge_censo_2022_populacao_idade_sexo c
        on m.id_municipio = c.id_municipio
    where
        m.sigla_uf = 'SC' and
        (m.id_municipio is null or c.id_municipio is null);
```

3.1.3. Corretude da agregação espacial eleitoral

Um ponto importante é verificar se a agregação espacial das seções eleitorais na tabela do TSE utilizada está correta em termos do **total de votos** final por município.

Para isso, utilizei outra tabela para comparação

```
--Comparativo de votos entre minha view agregada e tabela referênciawith oficial as (
select
    id_municipio,
    sequencial_candidato,
    sum(votos) as votos_oficial,
    count(*) as entradas_oficiais
from eleicoes22.br_tse_eleicoes_resultados_candidato_municipio_sc
where ano = 2022
    and sigla_uf = 'SC'
```

```

        and cargo ilike 'deputado estadual'
        group by id_municipio, sequencial_candidato
    ),
    agg as (
        select
            id_municipio,
            sq_candidato,
            sum(qt_votos) as votos_agg,
            count(*) as entradas_agg
        from eleicoes22.vw_eleicao_votacao_agg_por_municipio
        group by id_municipio, sq_candidato
    )
    select
        coalesce(o.id_municipio, a.id_municipio) as id_municipio,
        o.sequencial_candidato as sequencial_candidato,
        o.votos_oficial,
        a.votos_agg,
        (coalesce(a.votos_agg,0) - coalesce(o.votos_oficial,0)) as votos_excedentes,
        o.entradas_oficiais,
        a.entradas_agg
    from oficial o
    join agg a
        on o.id_municipio = a.id_municipio
        and o.sequencial_candidato = a.sq_candidato
        and (coalesce(a.votos_agg,0) - coalesce(o.votos_oficial,0)) != 0
    order by abs((coalesce(a.votos_agg,0) - coalesce(o.votos_oficial,0))) desc, id_municipio,
    sq_candidato
;

```

Isso resultou em ~550 pares de candidato e município reportados como inconsistentes de um total de 60k, margem de erro tolerável

id_municipio	sequencial_candidato	votos_oficial	votos_agg	votos_excedentes	entradas_oficiais	entradas_agg
4202008	240001610221	0	8892	8892	1	1
4203204	240001610221	0	4077	4077	1	1
4208906	240001614355	0	2639	2639	1	1
4207502	240001679647	0	1729	1729	1	1
4202909	240001616213	0	940	940	1	1
4205704	240001605915	0	661	661	1	1

id_municipio	sequencial_candidato	votos_oficial	votos_agg	votos_excedentes	entradas_oficiais	entradas_agg
4208203	240001610221	0	395	395	1	1
4207304	240001610223	0	363	363	1	1
4206504	240001614355	0	261	261	1	1
4208302	240001610221	0	253	253	1	1

Exemplo candidato Edson Renato Dias entre as duas fontes

Um exemplo interessante de inconsistência de votos que encontrei investigando dados.

Esse candidato foi impugnado!

https://pt.wikipedia.org/wiki/Edson_Renato_Dias#~text=Nas%20ele%C3%A7%C3%B5es%20de%202022%20teve%20aplica%C3%A7%C3%A3o%20de%20recursos%20federais.

A base de dados de votos por seção eleitoral ainda o considera, porém a base de votos por município não. Isso possivelmente explica a divergência encontrada de votos, assim como questões de voto em trânsito

1. Minha tabela agregada, busca por id_municipio = '4202008' and sq_candidato = '240001610221'
`vw_eleicao_votacao_agg_por_municipio`

Número de votos foi calculado com groupby juntando votos por zona e seção para um mesmo candidato de um município

id_municipio	4202008
nm_municipio	BALNEÁRIO CAMBORIÚ
qt_votos	8892
nr_votavel	10115
nm_votavel	EDSON RENATO DIAS
sq_candidato	240001610221
ds_cargo	DEPUTADO ESTADUAL

2. Tabela já processada e agregada a nível de município. Busca por id_municipio = '4202008' and sequencial_candidato = '240001610221'
br_tse_eleicoes_resultados_candidato_municipio_sc

Curiosamente 0 votos

ano	2022
turno	1
id_eleicao	546
tipo_eleicao	eleicao ordinaria
data_eleicao	2022-10-02
sigla_uf	SC
id_municipio	4202008
id_municipio_tse	80390
cargo	deputado estadual
numero_partido	10
sigla_partido	REPUBLICANOS
titulo_eleitoral_candidato	019691620990
sequencial_candidato	240001610221
numero_candidato	10115
resultado	nao eleito
votos	0

3. Tabela base de seções e zonas. Busca por cd_municipio = '80390' and sq_candidato = '240001610221'
votacao_secao_2022_sc

Observe que todas as seções tem voto

dt_geracao	01/11/2022	01/11/2022	01/11/2022	01/11/2022
hh_geracao	16:05:25	16:05:25	16:05:25	16:05:25

ano_eleicao	2022	2022	2022	2022
cd_tipo_eleicao	2	2	2	2
nm_tipo_eleicao	ELEIÇÃO ORDINÁRIA	ELEIÇÃO ORDINÁRIA	ELEIÇÃO ORDINÁRIA	ELEIÇÃO ORDINÁRIA
nr_turno	1	1	1	1
cd_eleicao	546	546	546	546
ds_eleicao	ELEIÇÕES GERAIS ESTADUAIS 2022	ELEIÇÕES GERAIS ESTADUAIS 2022	ELEIÇÕES GERAIS ESTADUAIS 2022	ELEIÇÕES GERAIS ESTADUAIS 2022
dt_eleicao	02/10/2022	02/10/2022	02/10/2022	02/10/2022
tp_abrangencia	E	E	E	E
sg_uf	SC	SC	SC	SC
sg_ue	SC	SC	SC	SC
nm_ue	SANTA CATARINA	SANTA CATARINA	SANTA CATARINA	SANTA CATARINA
cd_municipio	80390	80390	80390	80390
nm_municipio	BALNEÁRIO CAMBORIÚ	BALNEÁRIO CAMBORIÚ	BALNEÁRIO CAMBORIÚ	BALNEÁRIO CAMBORIÚ
nr_zona	56	56	56	56
nr_secao	380	331	415	133
cd_cargo	7	7	7	7
ds_cargo	DEPUTADO ESTADUAL	DEPUTADO ESTADUAL	DEPUTADO ESTADUAL	DEPUTADO ESTADUAL
nr_votavel	10115	10115	10115	10115
nm_votavel	EDSON RENATO DIAS	EDSON RENATO DIAS	EDSON RENATO DIAS	EDSON RENATO DIAS
qt_votos	22	16	10	42
nr_local_votacao	1090	1198	1228	1031
sq_candidato	240001610221	240001610221	240001610221	240001610221
nm_local_votacao	NÚCLEO DE EDUCAÇÃO INFANTIL CARROSEL	CENTRO EDUCACIONAL MUNICIPAL VEREADOR SANTA	COLÉGIO LICEU CATARINENSE	CENTRO EDUCACIONAL MUNICIPAL PROFESSOR ANTÔNIO LÚCIO
ds_local_votacao_endereco	RUA GRÉCIA, N. 205	RUA 2450, N. 420	RUA 3.020, 167	RUA ITÁLIA, N. 977

3.2. Tabelas e análise

Esse relatório carece de mostras de dados ou mais exemplos de análise, a razão disso é por que fiz tudo com scripts automáticos

O primeiro é um “**apresentador” do banco**, que se conecta nele e recupera todas as tabelas, views e views materializadas, extrai seu DDL e mostra 5 linhas aleatórias, retornando um arquivo markdown agrupado por tabela.

O resultado de sua execução está no arquivo html ***eleicoes22_introspection.html***

O segundo é um **script analisador de dados**, que itera todas as tabelas do banco (descartando views) e para cada coluna, recupera seu tipo e faz um select apropriado para inferir sua distribuição, com isso executando análises estatísticas pré definidas.

O resultado de sua execução está no arquivo html ***eleicoes22_stats.html***

Ambos os relatórios gerados foram inseridos no zip

4. Transformações de dados

4.1. Metodologia da engenharia de dados

Para otimizar a execução de consultas e simplificar o tratamento de dados, vários tratamentos foram feitos utilizando views.

Pipeline de extração, transformação e carga funcionou dessa forma

1. Baixar csv dos dados, seja diretamente ou um export do BigQuery (onde já filtro preliminarmente)
2. Analisar estrutura de colunas e tipos para conversão ao Postgres
3. Criar tabelas com tipos/encodings adequados e copiar dados dos arquivos csv
4. Validar consistência de chaves entre tabelas
5. Preparar views materializadas de operações custosas
6. Preparar views simples derivadas para filtros simples
7. Integrar script de análise de qualidade dos dados

4.2. Dados eleitorais

Tabela de resultados por seção eleitoral do TSE: *vw_eleicao_votacao_agg_por_municipio*

Agrupei os dados por informações a nível de candidato e município, somando as votações de zona e seção, criando uma view materializada agrupada

```
--vw_eleicao_votacao_agg_por_municipio
WITH base AS (SELECT v.id_municipio,
                     v.cd_municipio,
                     v.nm_municipio,
                     v.nr_zona,
                     v.nr_secao,
                     v.cd_cargo,
                     v.ds_cargo,
                     v.nr_votavel,
                     v.nm_votavel,
                     v.sq_candidato,
                     v.qt_votos::integer AS qt_votos
      FROM eleicoes22.vw_eleicao_votacao_por_secoes v)
SELECT id_municipio,
       nm_municipio,
       sum(qt_votos) AS qt_votos,
       nr_votavel,
       nm_votavel,
       sq_candidato,
       ds_cargo
  FROM base
 GROUP BY id_municipio, nm_municipio, ds_cargo, nr_votavel, nm_votavel, sq_candidato
 ORDER BY id_municipio, nm_municipio, qt_votos DESC, nm_votavel;
```

Tabela referência: *br_inep_indicador_nivel_socioeconomico_municipio*

Veio do BigQuery, filtrei rudemente por lá e exportei o csv, onde fiz uma limpeza mais aprofundada de dados desnecessários. Ao final será utilizada apenas para comparação com os dados que agreguei de outra fonte.

4.3. Dados de educação

Complementei os dados selecionados do INEP com informações extraídas de um dicionário de dados por código e ano para os campos de “rede” e “tipo_localizacao” e criei uma view materializada filtrando dados desnecessários, facilitando a análise.

Consulta base que gera a view

```
with rede_dict as (select *
                   from (values (0, 2021, 'Total (Federal, Estadual, Municipal e Privada)'),
                           (1, 2021, 'Federal'),
                           (2, 2021, 'Estadual'),
                           (3, 2021, 'Municipal'),
                           (4, 2021, 'Privada'),
                           (5, 2021, 'Total (Estadual e Municipal)'),
                           (6, 2021, 'Total (Federal, Estadual e Municipal)')) as t(cod_rede, ano,
desc_rede)),
     tipo_loc_dict as (select *
                        from (values (0, 2021, 'Total (Urbana e Rural)'),
                               (1, 2021, 'Urbana'),
                               (2, 2021, 'Rural')) as t(cod_tipo_localizacao, ano,
desc_tipo_localizacao))
select b.id_municipio,
       m.nome,
       quantidade_alunos_inse,
       rede,
       r.desc_rede,
       tipo_localizacao,
       tl.desc_tipo_localizacao,
       inse,
       percentual_nivel,
       percentual_nivel_2,
       percentual_nivel_3,
       percentual_nivel_4,
       percentual_nivel_5,
       percentual_nivel_6,
       percentual_nivel_7,
       percentual_nivel_8,
       m.sigla_uf
  from eleicoes22.br_inep_indicador_nivel_socioeconomico_municipio b
    left join rede_dict r
      on r.ano = b.ano
      and r.cod_rede = nullif(b.rede, "")::int
    left join tipo_loc_dict tl
      on tl.ano = b.ano
```

```

        and tl.cod_tipo_localizacao = nullif(b.tipo_localizacao, '')::int
    left join eleicoes22.br_bd_diretorios_brasil_municipio m
        on b.id_municipio = m.id_municipio
    where b.sigla_uf = 'SC'
        and b.ano = '2021'
    order by b.id_municipio, b.sigla_uf, b.rede, b.tipo_localizacao;

```

Inicialmente fiz com vários anos, mas como só nos interessa dados recentes descartei ao final e inclui esse filtro no where.

Um ponto de cuidado é que ele contém roll ups de dados intermediários!

Exemplo tipo de localização urbana fixa

	<input type="checkbox"/> id_municipio	<input type="checkbox"/> nome	<input type="checkbox"/> quantidade_alunos_inse	<input type="checkbox"/> rede	<input type="checkbox"/> desc_rede	<input type="checkbox"/> tipo_localizacao	<input type="checkbox"/> desc_tipo_localizacao
1	4216255	São João do Oeste		127 0	Total (Federal, Estadual, Municipal e Privada)	1	Urbana
2	4216255	São João do Oeste		0 1	Federal	1	Urbana
3	4216255	São João do Oeste		88 2	Estadual	1	Urbana
4	4216255	São João do Oeste		39 3	Municipal	1	Urbana
5	4216255	São João do Oeste		127 5	Total (Estadual e Municipal)	1	Urbana
6	4216255	São João do Oeste		127 6	Total (Federal, Estadual e Municipal)	1	Urbana

Exemplo rede municipal fixa

	<input type="checkbox"/> id_municipio	<input type="checkbox"/> nome	<input type="checkbox"/> quantidade_alunos_inse	<input type="checkbox"/> rede	<input type="checkbox"/> desc_rede	<input type="checkbox"/> tipo_localizacao	<input type="checkbox"/> desc_tipo_localizacao
1	4216255	São João do Oeste		50 3	Municipal	0	Total (Urbana e Rural)
2	4216255	São João do Oeste		39 3	Municipal	1	Urbana
3	4216255	São João do Oeste		11 3	Municipal	2	Rural

Para resolver esse problema, criei uma view materializada crua (*vw_indicador_nivel_socioeconomico_municipio_raw*) e duas views a partir dela:

1. Totalizada por município: *vw_indicador_nivel_socioeconomico_municipio_agregado*
2. Segmentada por tipos de rede e localização:
vw_indicador_nivel_socioeconomico_municipio_segmentado

Dessa forma, é possível analisar estatísticas agregadas ou não no município para análises.

4.4. Preparação de dados geográficos com município

A tabela *br_bd_diretorios_brasil_municipio* contém boa parte dos dados úteis, assim como identificadores e informações de outras tabelas, sejam em níveis acima, como região imediata / intermediária, ou tabelas de outros domínios, como identificadores de municípios do TSE e receita federal.

Criei uma view materializada *vw_diretorio_municipio_e_geom* juntando essa tabela com a de mapas de município *br_geobr_mapas_municipio*, permitindo fáceis agregações simples ou espaciais posteriores

```
-- vw_diretorio_municipio_e_geom
```

```

SELECT dbm.id_municipio,
       dbm.nome,
       dbm.sigla_uf,
       dbm.capital_uf,
       dbm.id_comarca,
       dbm.id_regiao_saude,
       dbm.nome_regiao_saude,
       dbm.id_regiao_imediata,
       dbm.nome_regiao_imediata,
       dbm.id_regiao_intermediaria,
       dbm.nome_regiao_intermediaria,
       dbm.id_microrregiao,
       dbm.nome_microrregiao,
       dbm.id_mesorregiao,
       dbm.nome_mesorregiao,
       dbm.id_regiao_metropolitana,
       dbm.nome_regiao_metropolitana,
       dbm.ddd,
       dbm.amazonia_legal,
       dbm.id_municipio_tse,
       dbm.id_municipio_rf,
       dbm.id_municipio_bcb,
       br_geobr_mapas_municipio.geometria
FROM eleicoes22.br_geobr_mapas_municipio
      JOIN eleicoes22.br_bd_diretorios_brasil_municipio dbm
        ON br_geobr_mapas_municipio.id_municipio = dbm.id_municipio

```

Existem outras tabelas com dados geométricos para regiões imediatas e intermediárias, mas preferi mantê-las separadas da view para economizar espaço. Durante a fase de visualização pode ser interessante criar uma única view com tudo

Para junção não é necessário utilizar funções espaciais, as “chaves estrangeiras” funcionam muito bem.

4.5. Otimizações gerais

4.5.1. Limpeza de tabelas e colunas

Durante a etapa exploratória várias tabelas desnecessárias foram importadas, é desejado para a parte 2 removê-las e também eliminar colunas redundantes ou não utilizadas. No momento tudo que faço é ignorá-las no cliente de banco de dados que utilizo.

4.5.2. Índices

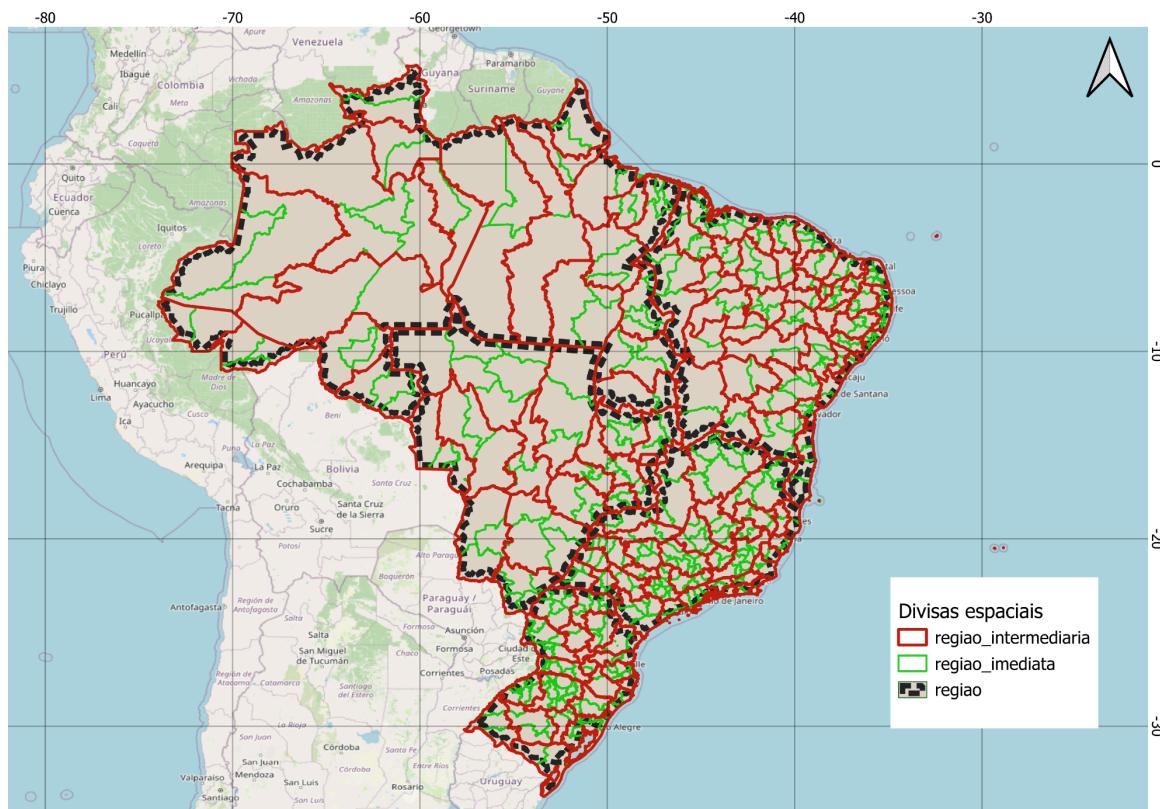
Não implementei índices nas tabelas, isso deve acelerar bastante o postgis. Da mesma forma também não configurei chaves primárias para os ids de município nas tabelas criadas, isso permitirá usar indexação automática

5. Amostras QGIS

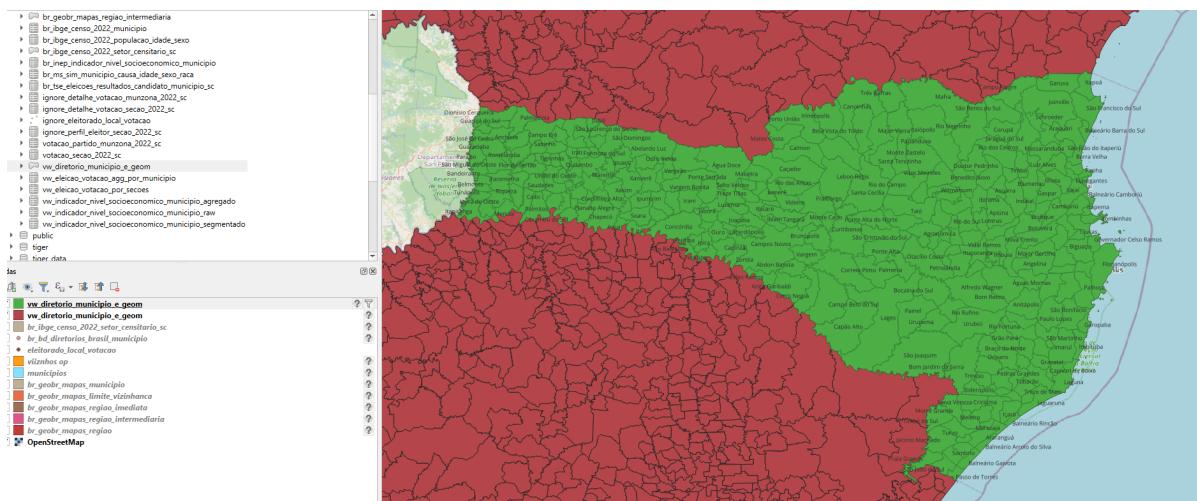
5.1. Agrupamento espacial de níveis de região

Regiões do país -> regiões intermediárias -> regiões imediatas

Criei um layout no qgis para fazer o mapa bonitinho, os outros apenas tirei print



5.2. Município SC



5.3. Município em overlay com regiões

Regiões intermediárias e imediatas em tons de branco, apenas para ver a combinação delas com municípios

