

# Análise de Dados Espaciais

**Lucas Gelape (UFMG)**

<https://lgelape.github.io/>

*Formação Metodológica do MAPE-IESP*  
Instituto de Estudos Sociais e Políticos, UERJ  
14 de novembro de 2025

# Tópicos

## ① Introdução

- Um ponto de partida: a investigação de John Snow sobre o surto de cólera em Londres
- Por que análise espacial? A geografia como variável de fenômenos sociais e políticos

## ② Noções metodológicas

- Noções de cartografia
- Sistemas de informação geográfica (SIG, GIS)
- Tipos e formatos de dados espaciais
- Funcionalidades úteis

## ③ Mapas

- Orientações gerais
- Mapas coropléticos
- Densidade de pontos
- Símbolos proporcionais
- Cartogramas
- Mapas de fluxos

## ④ Análise espacial de dados

- Vizinhança
- Autocorrelação espacial

## ⑤ Cuidados na produção de inferências

- Falácia ecológica
- MAUP

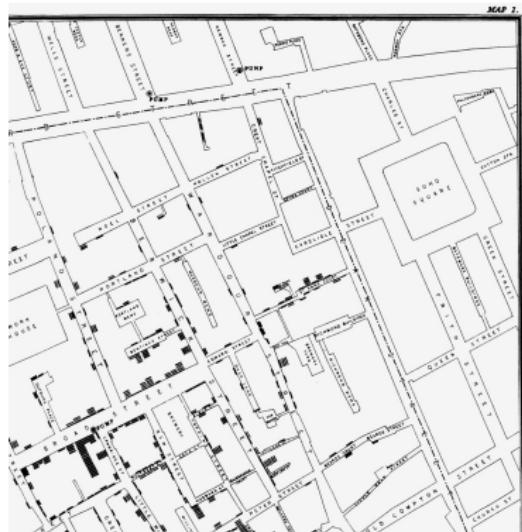
# Análise de Dados Espaciais

## *Introdução*

# John Snow e uma epidemia de cólera em Londres

- Como explicar uma epidemia de cólera em Londres?
- John Snow (1854)
- Explicações concorrentes:
  - Miasmas
  - Ingestão de água insalubre
- Experimento
  - Consumo de água fornecida por diferentes empresas e fontes
  - Identificação das residências das vítimas
  - Cálculo da distância das residências e bombas d'água

Mapa produzido por John Snow



Fonte: The John Snow Archive and Research Companion

# Heranças

- Métodos mistos
- Importância do espaço
- Centralidade na visualização de dados



Fonte: o autor.

# Por que análise espacial?

- O espaço influencia fenômenos sociais? Como?
- Problema de Galton (1988) / Primeira Lei da Geografia (Tobler, 1970):

*"Tudo está relacionado com as demais coisas, mas aquelas próximas estão mais relacionadas entre si do que as distantes."*

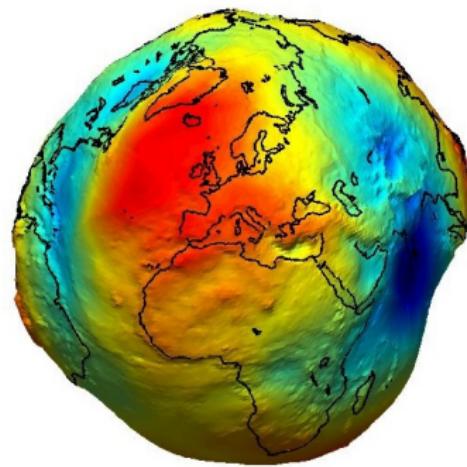
- Pensar para mapear (raciocínio dedutivo)
- Mapear para pensar (raciocínio indutivo)

# Análise de Dados Espaciais

## *Noções metodológicas introdutórias*

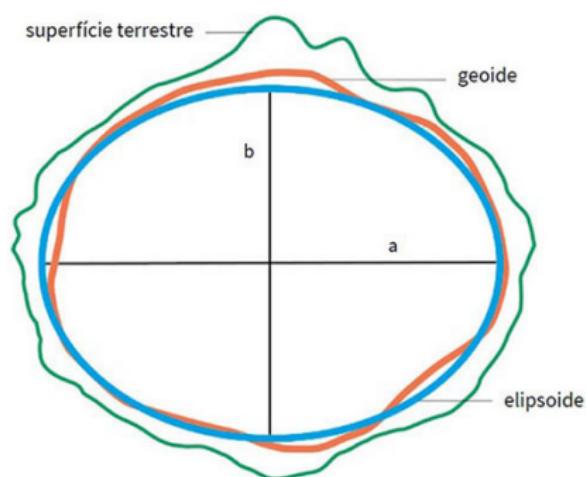
# Qual a forma da terra?

- Geoide



Fonte: ESA, 2004.

- Elipsoide

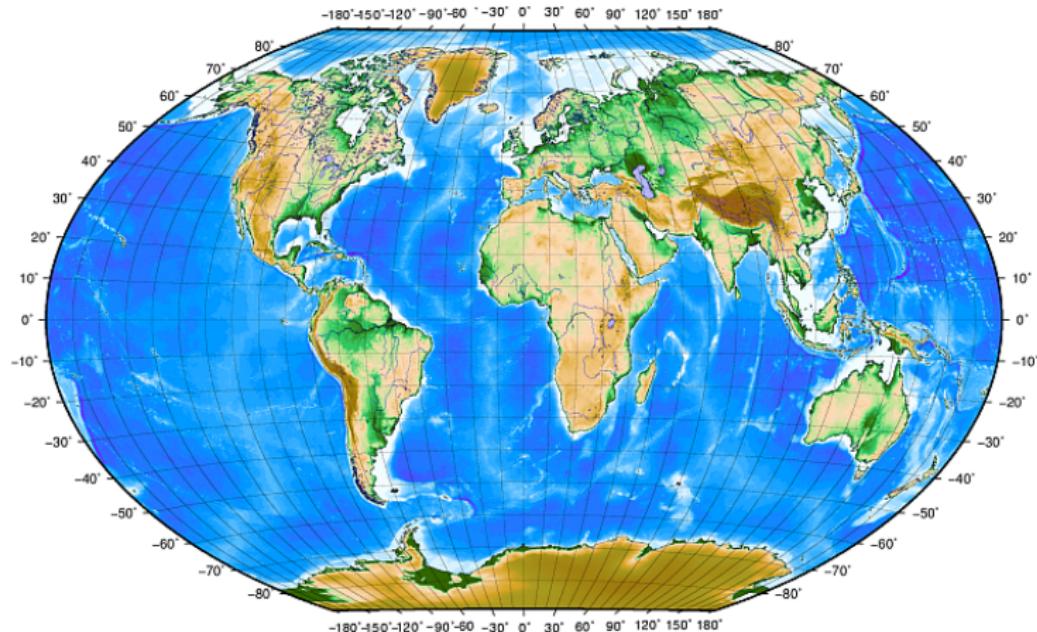


Fonte: IBGE, 2024.

# Sistemas de coordenadas (SRC)

- **Datum**: referências para a representação da superfície terrestre em certo mapa (forma, tamanho e posição)
- **SRC/CRS**: expressa os pontos sobre uma superfície, com base em um *datum*
- **Coordenadas geográficas**: representação baseada em esferoide (graus decimais)
- **Coordenadas projetadas**: representação em um plano (unidades lineares)

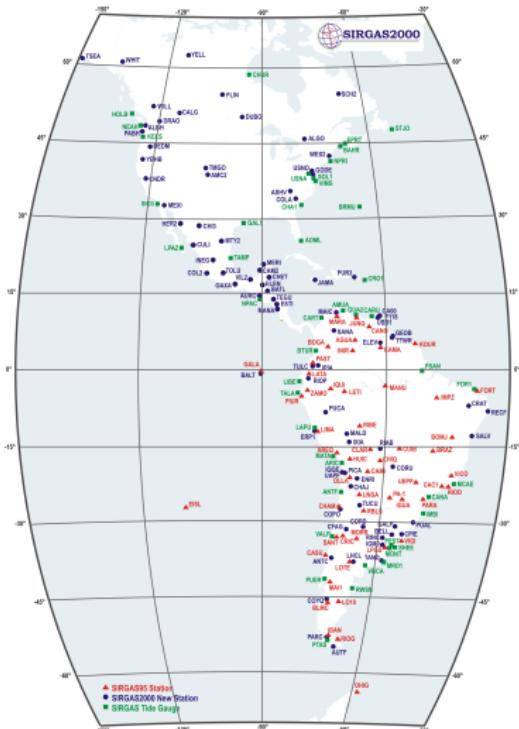
# Sistemas de coordenadas



Fonte: QGis Documentation, v 3.34.

# Datum oficial no Brasil

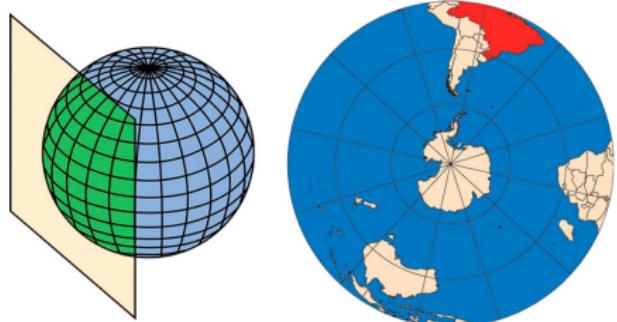
- *Datum planimétrico:*
  - SIRGAS-2000 (desde 2015)
  - SAD69 e CA (2005 a 2015)
- “Padrão” mundial: WGS 84 (World Geodetic System)



Fonte: TUM, 2024.

# Projeções cartográficas

- Tradução do elipsoide em um plano
- Por que importam?
  - Perspectivas sobre a representação terrestre
  - Cálculos de área, tempo e distância
- Como existem distorções, são buscadas características como:
  - Equidistância
  - Equivalência (área)
  - Conformidade (ângulos)



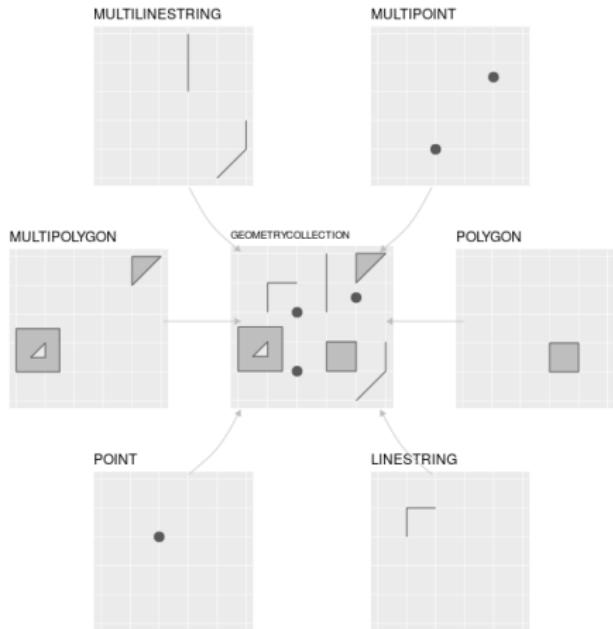
Fonte: IBGE, 2024

# Sistemas de informação geográficas (SIG, GIS)

- Conjunto de ferramentas para:
  - coletar
  - organizar
  - processar
  - visualizar
  - analisar
- dados espaciais.
- Possibilidades:
  - R
  - QGIS
  - ArcGIS
  - MapInfo
  - Python
  - GeoDa
  - E outros...

# R e pacote sf

- Adoção das *simple features* no R
- Modelo hierárquico de representação de tipos geométricos
- Padrão aberto endossado pela *Open Geospatial Consortium*.



Fonte: Lovelace et al., 2024.

# Pacote sf: o que nos importa

- Maior velocidade e simplicidade na leitura e exportação de dados
- Melhor performance na produção de visualização
- Podem ser tratados como “tabelas” (`data.frames`)
- Nomes de funções intuitivos (`st_`)
- Compatível com o `tidyverse`
- Maior interoperabilidade entre sistemas

- Software livre com código-fonte aberto
- Funciona em diversos sistemas operacionais
- Interface simplificada de uso
- Permite o uso de dados de diversos formatos
- Integra-se a outros SIG, existência de plugins criados em outras linguagens
- Ampla comunidade, mesmo em língua portuguesa

- Foco em análise espacial (econometria espacial)
- Funciona em diversos sistemas operacionais
- Interface ainda mais simples
- Permite o uso de dados de diversos formatos

# Fontes (não-exaustivas) de dados espaciais

- Dados nacionais:
  - IBGE
  - CEM-USP
- Dados estaduais:
  - SP: GEOSEADE
  - MG: IEGE
  - BA: IDE.Bahia
- Dados metropolitanos:
  - CEM-USP
  - Atlas Brasil
- Dados municipais:
  - BH: BHGEO
  - SP: Geosampa
  - RJ: Data.Rio
  - Curitiba: IPPUC
  - Recife: Dados Recife
- Pacotes em R:
  - geobr
  - hydrobr
  - rworldmap

# Dados espaciais e não espaciais

- Dados espaciais:

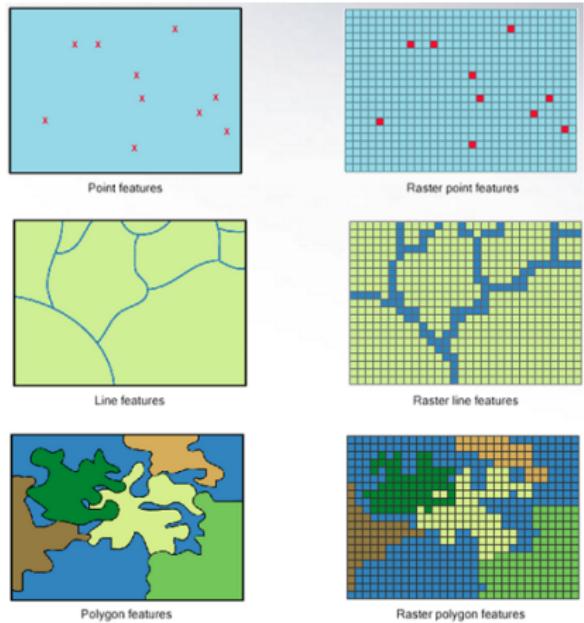
- “Todos os que representam recursos ou características físicas do espaço geográfico [...]. Também podem ser [...] as atividades humanas associadas ao território [...].” (Rodrigues-Silveira, 2013, p. 41)

- Dados não-espaciais:

- “Características de qualquer fenômeno consideradas como independentes (*a priori*) de qualquer consideração geométrica ou geográfica.” (Rodrigues-Silveira, 2013, p. 41).

# Formatos de dados espaciais

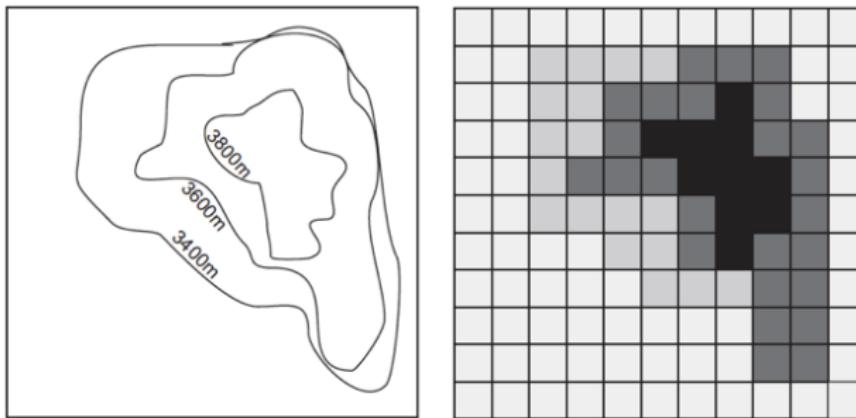
- *Raster* (ou matricial):
  - Organizados em pixels
  - Representação de dados contínuos no espaço
- *Vectorial*
  - Organizados em pontos, linhas ou polígonos
  - Representações discretas de dados espaciais



Fonte: Humboldt State University (2018)

# Dados espaciais *raster*

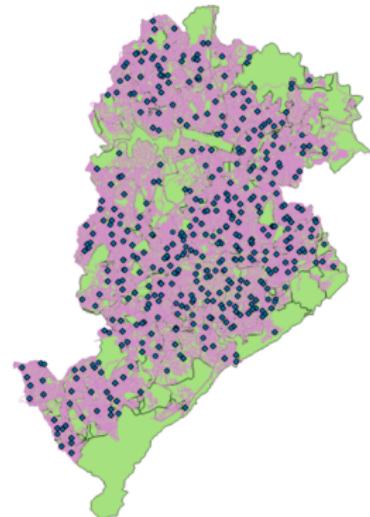
- Grid de pixels, composto por linhas e colunas
- O espaço é dividido no grid e a variável medida para cada célula
- Exemplos: topografia de um terreno ou a poluição em uma área



Fonte: Gleditsch e Weidmann (2012, p. 463).

# Dados espaciais vetoriais

- A escolha pelo tipo depende da característica dos dados e objetivo:
  - **Pontos:** localização exata do fenômeno.
  - **Linhas:** deslocamentos, fluxos, cuja largura não é relevante.
  - **Polígonos:** úteis para a representação de área.



Fonte: elaborado com dados do TSE e da Prefeitura de BH.

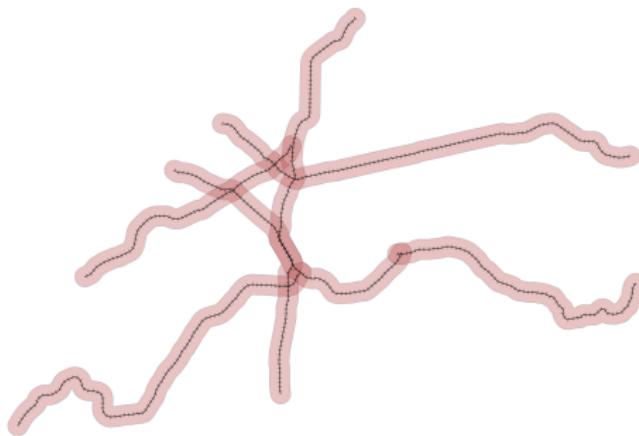
# *Shapefiles*

- O formato mais utilizado de dados espaciais vetoriais. Compostos ao menos por:
  - .shp: contém os dados geométricos;
  - .shx: índice de posição das geometrias;
  - .dbf: armazena os atributos;
  - .prj: o formato da projeção.

# Buffers

- Área que circunda uma geometria espacial, com raio/espressura pré-determinada
- Pode ser produzida ao redor de pontos, linhas ou polígonos

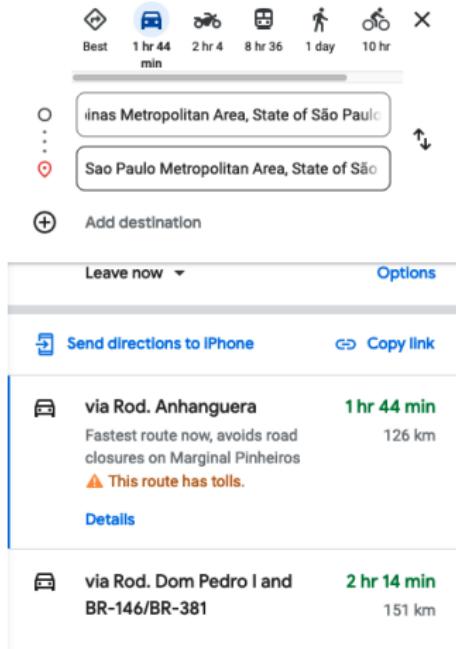
*Buffer de 500m ao redor das linhas de metrô de SP*



Fonte: Geosampa.

# Cálculos de distância

- O mais comum é o cálculo da distância euclidiana ou de grande círculo entre pares de geometrias.
- Na prática, contudo, essas distâncias podem ser maiores/menores. Podemos, então, calcular “rotas” entre pontos:
  - Há uma série de aplicativos pagos que auxiliam nisto. Porém, o *OpenStreetMap* é uma opção segura e gratuita para tanto.



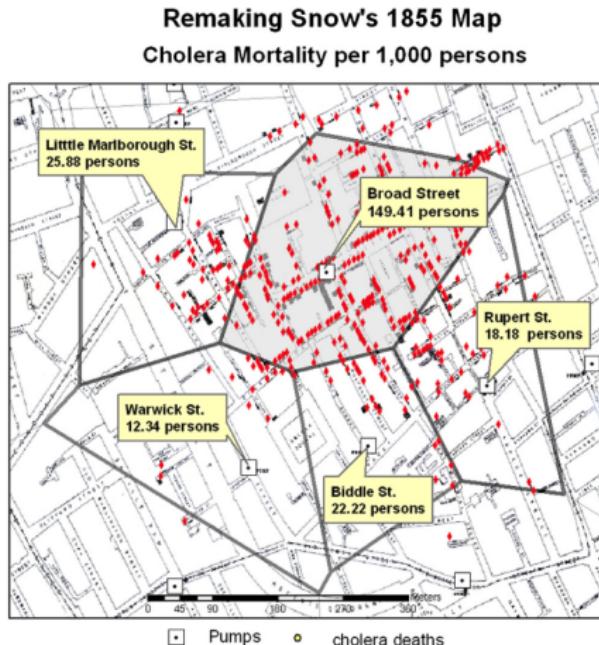
Fonte: Google Maps

# Geocodificação

- Procedimento para se obter coordenadas geográficas de endereços ou outras informações.
- Diversos serviços (gratuitos ou pagos): Google Maps, Here, Nominatin (OpenStreetMap), ArcGIS, geocodebr.
- Demanda testar **consistência** dos resultados:
  - N. de dados faltantes (*missing*).
  - Os pontos estão todos dentro dos limites municipais?
  - Dos que não estão, a qual distância estão das fronteiras?
  - Sortear uma amostra aleatória e verificar manualmente o erro/acerto.
- O **geocodebr** merece menção especial.

# Diagramas de Voronoi

- Uma área é dividida em polígonos equivalentes ao número de pontos de referência ali contidos
- O ponto dentro dele estará mais próximo do seu ponto de referência do que dos demais pontos de referência



Fonte: Koch e Denike (2009).

# Análise de Dados Espaciais

## *Mapas*

# Produção de mapas – Orientações gerais

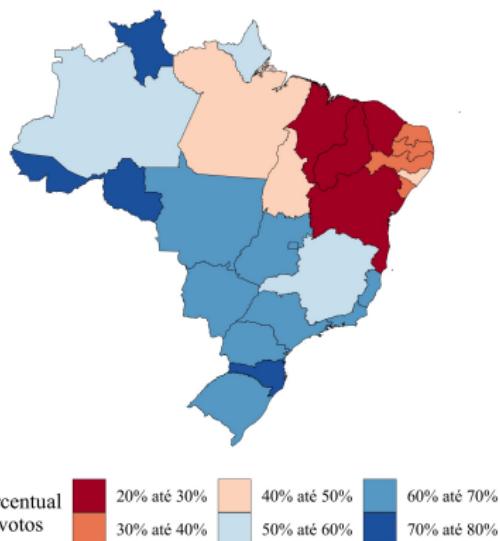
Com base em Rodrigues-Silveira (2013) e Cairo (2016):

- Posicione o título na parte superior do mapa.
- A figura do mapa deve vir abaixo do título, com algum espaço entre eles.
- A fonte da informação deve vir abaixo do mapa, em letra menor.
- Legenda e outros símbolos devem ser incluídos segundo o público e a informação transmitida, serem claros e fazer bom uso de espaços vazios.
- Não confie nas configurações padrão dos softwares!
- Lembrem-se: *menos é mais!*

# Mapas coropléticos

- Mapas que atribuem cores a áreas segundo os valores de uma variável

Votação de Jair Bolsonaro (2018, 2º turno)

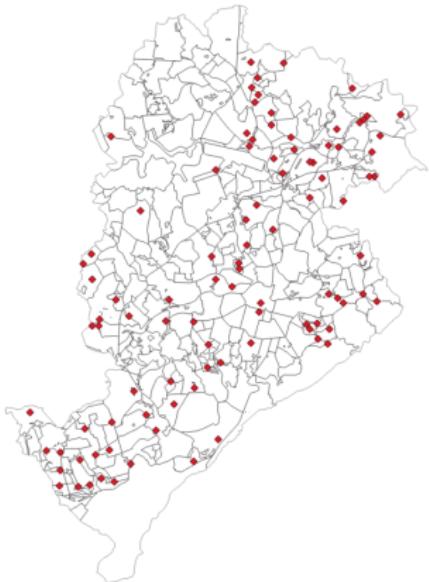


Fonte: Gelape (2021)

# Densidade de pontos

UMEIs representadas sobre bairros de BH

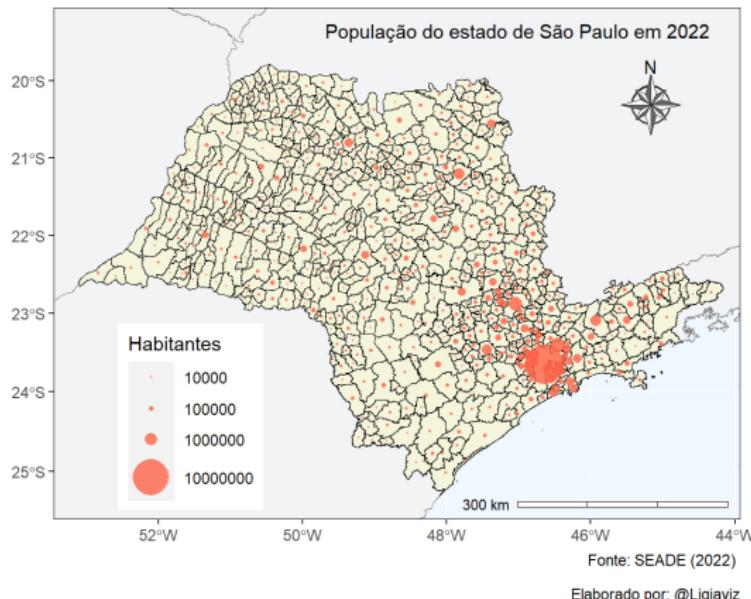
- Mapas que auxiliam na compreensão do grau de concentração espacial de algum elemento (Rodrigues-Silveira, 2013, p. 106).



Fonte: Prefeitura de BH

# Símbolos proporcionais

- Representam valores de uma (ou mais) variável associados a determinados pontos no espaço.

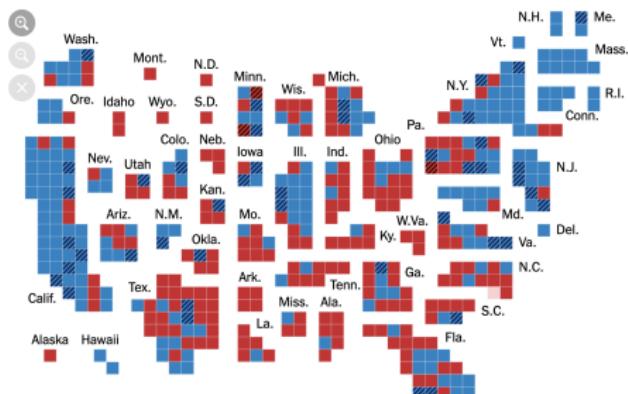


Fonte: Barrozo (2023)

# Cartogramas

- Mapas que distorcem a forma ou o tamanho dos objetos que representam as unidades geográficas, segundo alguma variável escolhida pelo pesquisador
- Objetivo: reduzir sobrerepresentação de territórios com grandes dimensões

Resultado das eleições *midterm* - EUA (2018)



Fonte: The New York Times (2018)

# Mapas de fluxos

- Útil para descrever e analisar o movimento, seja de pessoas ou informações, por exemplo.

Os refugiados da  
**Síria**  
foram para...



Fonte: Zanlorenssi e Almeida, Nexo Jornal (2018)

# Análise de Dados Espaciais

## *Análise espacial de dados*

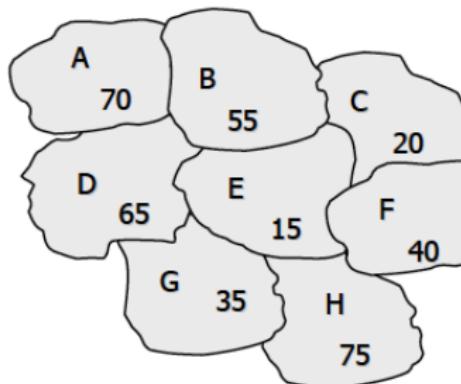
# Análise espacial de dados

- Centrada na ideia de vizinhança: operacionalização da influência do espaço.
- Baseada na teoria:
  - Difusão?
  - Composição?
- Conectividade mútua: a vizinhança não é só geográfica (Neumayer; Plümper, 2016)

# Critérios de definição de vizinhança

- Contiguidade
  - Rainha (*queen*)
  - Bispo (*bishop*)
  - Torre (*rook*)
- Vizinhos mais próximos
- Distância
- Ordem
  - Primeira-ordem
  - Segunda-ordem
  - Terceira-ordem...

Exemplo: vizinhança por contiguidade



Fonte: Terron (2009, p. 62)

# Matriz de pesos espaciais ( $W$ )

- Operacionalização do conceito de vizinhança
- Preenchida com valores a depender da existência ou não, conforme o critério adotado
- Dois tipos:
  - Exposição heterogênea: Matrizes binárias
  - Exposição homogênea: Matrizes padronizadas (*row-standardized*)

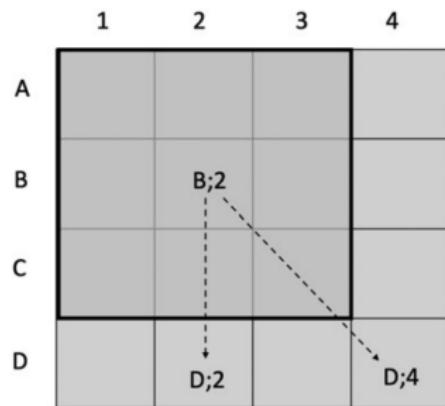
# Matriz de pesos espaciais ( $W$ )

Exemplo: matriz de vizinhança binária

	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4	D1	D2	D3	D4
A1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
A2	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0
A3	0	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0
A4	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0
B1	1	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0
B2	1	1	1	0	1	0	1	0	1	1	1	0	0	0	0	0
B3	0	1	1	1	0	1	0	1	0	1	1	1	0	0	0	0
B4	0	0	1	1	0	0	1	0	0	0	1	1	0	0	0	0
C1	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0
C2	0	0	0	0	1	1	1	0	1	0	1	0	1	1	1	0
C3	0	0	0	0	0	1	1	1	0	1	0	1	0	1	1	1
C4	0	0	0	0	0	1	1	0	0	1	0	0	0	1	1	1
D1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0
D2	0	0	0	0	0	0	0	0	1	1	1	0	1	0	1	0
D3	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	1
D4	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0

Fonte: Di Salvatore e Ruggeri (2021, p. 6)

Exemplo: contiguidade rainha de segunda-ordem



Fonte: Di Salvatore e Ruggeri (2021, p. 7)

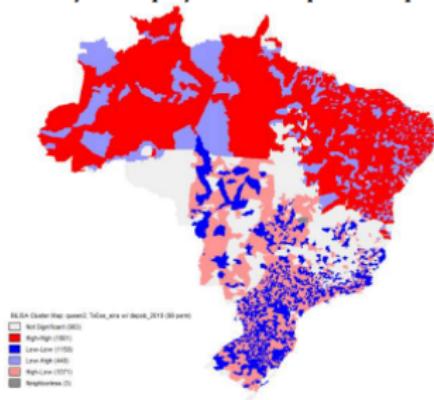
# Autocorrelação espacial

- Inclinação de humanos para detectar padrões onde eles não necessariamente existem (*apoefenia*)
- Testes estatísticos para avaliar se valores de uma variável em uma área estão associados aos valores de uma variável em áreas vizinhas
- Autocorrelação espacial *global*: o conjunto de dados como um todo apresenta dependência espacial?
- *I* de Moran: avalia os desvios em relação à média de uma variável, levando em conta seus vizinhos.
  - $H_0: (-1)/(N - 1)$  (tende a 0 em amostras grandes)
  - Resultado: em matrizes normalizadas, variam entre valores próximos a -1 e +1, com interpretação semelhante ao de um teste de correlação

# Autocorrelação espacial local

- Indicadores globais não possibilitam a avaliação de como áreas específicas estão relacionadas com seus vizinhos
- Indicadores locais de associação espacial (ou LISA: *Local Indicators of Spatial Association*)
- Cada unidade é classificada em:
  - Alto-alto, alto-baixo, baixo-alto, baixo-baixo e não-significante.

Mapa 1: Estatística LISA Bivariada: Taxa de Execução e Proporção de Pobres por Município



Fonte: D'Albuquerque (2017, p. 64)

# Análise de Dados Espaciais

## *Cuidado na produção de inferências*

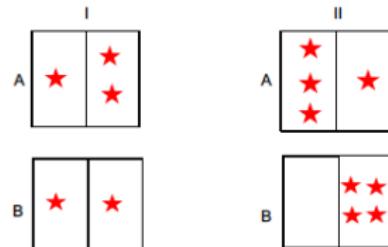
# Falácia ecológica

- “Inferir um comportamento individual a partir de características compiladas em um nível de grupo” (Rodrigues-Silveira, 2013, p. 43).
- Existem métodos de inferência ecológica que visam a análise do comportamento individual, como a transferência de votos (King, 1997; Calvo e Escolar, 2003).

# Problema da unidade de área modificável (MAUP)

- Riscos de se chegar a mais de uma conclusão (possivelmente divergentes!), se agregarmos os dados em diferentes escalas
- Podem ter duas fontes:
  - Escala
  - Partição
- Recomendações para a decisão da escala/partição:
  - Teoria
  - Comparação dos resultados entre escalas

MAUP: escala



MAUP: partição



Fonte: Gonçalves (2017).

# Sugestões de leitura

- Texto introdutório em português:
  - Gelape, L. *Possibilidades de pesquisa empírica: teorias e ferramentas para análises espaciais*. In: Silotto, G; Gelape, L; Silva, G; Castro, P. **Poder e Território: Uma abordagem a partir da Ciência Política**. Curitiba: Intersaberes, 2021. p. 229-292.
- Textos introdutórios em inglês:
  - Di Salvatore, J.; Ruggeri, A. *Spatial analysis for political scientists. Italian Political Science Review/Rivista Italiana di Scienza Politica*, v. 51, n. 2, 2021, p. 198–214.
  - Gleditsch, K.; Weidmann, N. *Richardson in the Information Age: Geographic Information Systems and Spatial Data in International Studies. Annual Review of Political Science*, v. 15, n. 1, p. 461-481

# Sugestões de leitura

- Livros:

- Darmofal, D. *Spatial analysis for the social sciences*. New York, NY: Cambridge University Press, 2015.
- Rodrigues-Silveira, R. *Representación espacial y mapas*. CIS - Centro de Investigaciones Sociológicas: Madrid, 2013. [esta obra tem aplicações em R, produzidas antes do pacote sf. Logo, recomendo mais pelos aspectos teóricos do que pelas aplicações.]

- Livros com aplicações em R:

- Barrozo, L. V. *Cartografia Temática em R para estudantes de Geografia: do zero aos principais tipos de representações cartográficas em R*. 2023.
- Gelape, L.; Silva, G. *Geografia Eleitoral: teoria e prática*. Curitiba: Intersaber, 2025.
- Lovelace, R.; Nowosad, J.; Muenchow, J. *Geocomputation with R*. CRC Press, 2025.
- Moraga, P. *Spatial statistics for data science: theory and practice with R*. Boca Raton, FL: CRC Press, 2023.