

GENE7033 – Tópicos Especiais em Genética I:

Visualização de dados para publicações científicas

Profª Drª Chirlei Glienke

Drª Desirrê Petters-Vandresen

Tipos de dados, variáveis e elementos estéticos

Dr^a Desirrê Petters-Vandresen

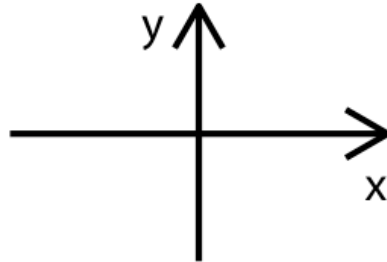
13/10/2022

Convertendo dados em elementos estéticos

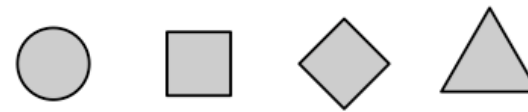
- Visualização de dados: conversão de valores em elementos visuais/estéticos de forma lógica e sistemática

- Alguns exemplos:

position



shape



size



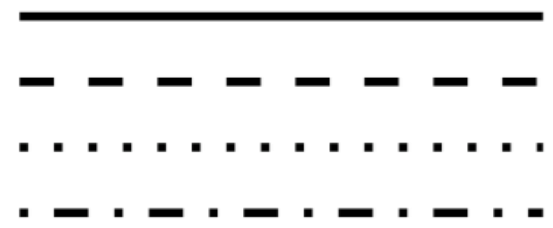
color



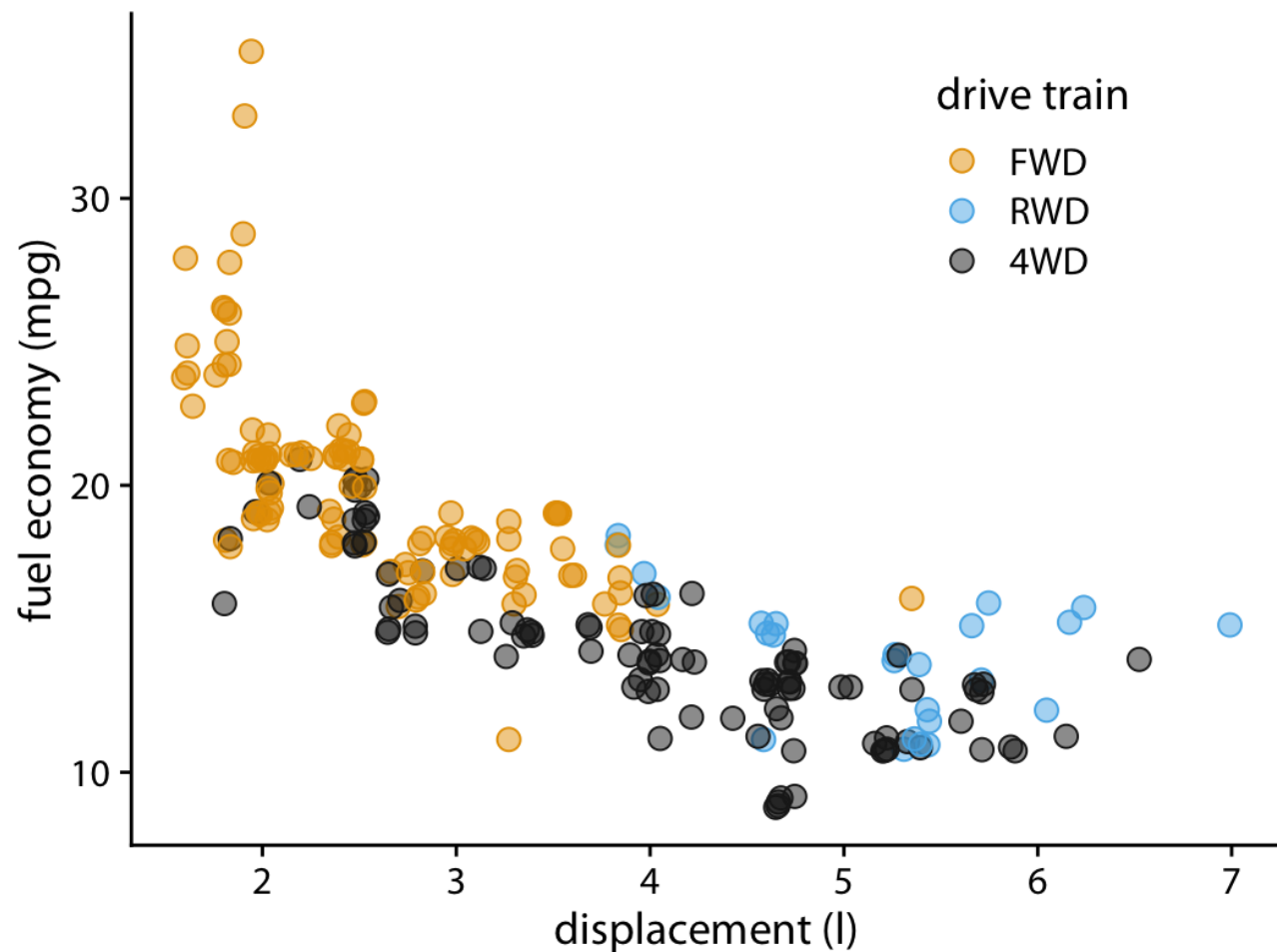
line width



line type



Convertendo dados em elementos estéticos

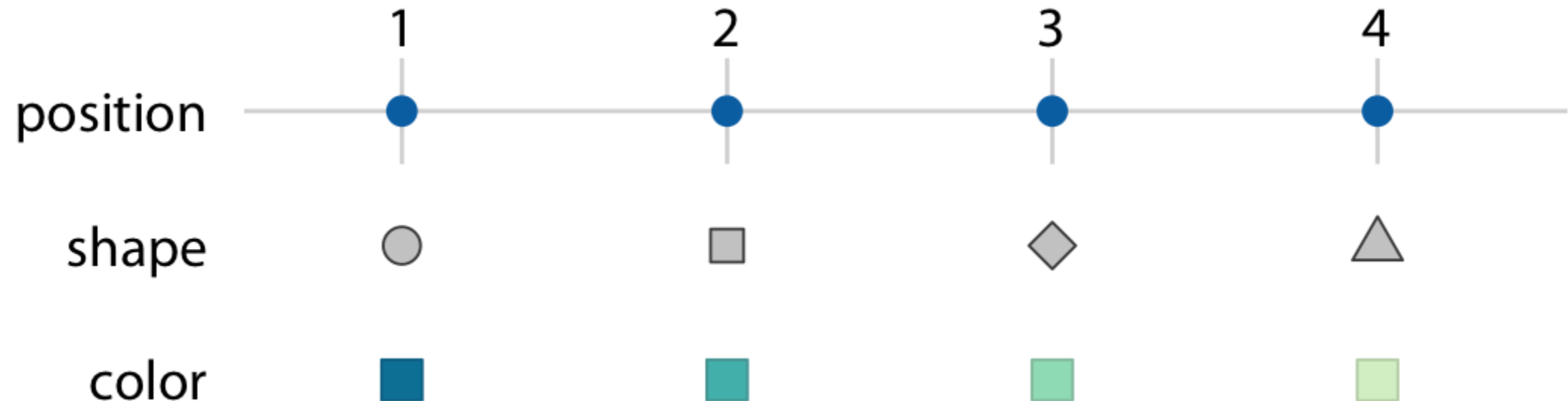


Adaptado de:
WILKE, C. O. **Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures**. O'Reilly Media, 2019.



Escalas

- Mapeamento de quais valores das variáveis correspondem a quais valores/estados dos elementos estéticos
- Um-para-um: somente um valor/estado dos elementos estéticos para cada valor de variável, para evitar ambiguidade



Tipos de variáveis

Tipo de variável		Exemplos	Escala adequada
Quantitativa	Numérica, contínua	1.3, 5.7, 83	Contínua
	Numérica, discreta	1, 2, 3, 4	Discreta
Qualitativa*	Categórica, ordinal	Gato, cachorro, peixe	Discreta
	Categórica, nominal	Ruim, médio, bom	Discreta
-	Data ou hora	5 de janeiro de 2018, 12:30	Contínua ou discreta
-	Texto	<i>Um pequeno jabuti xereta viu dez cegonhas felizes</i>	Nenhuma, ou discreta

Adaptado de:
WILKE, C. O. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. O'Reilly Media, 2019.

Variáveis com dados qualitativos são “fatores” com diferentes categorias (níveis)

Elementos estéticos vs. tipos de variáveis

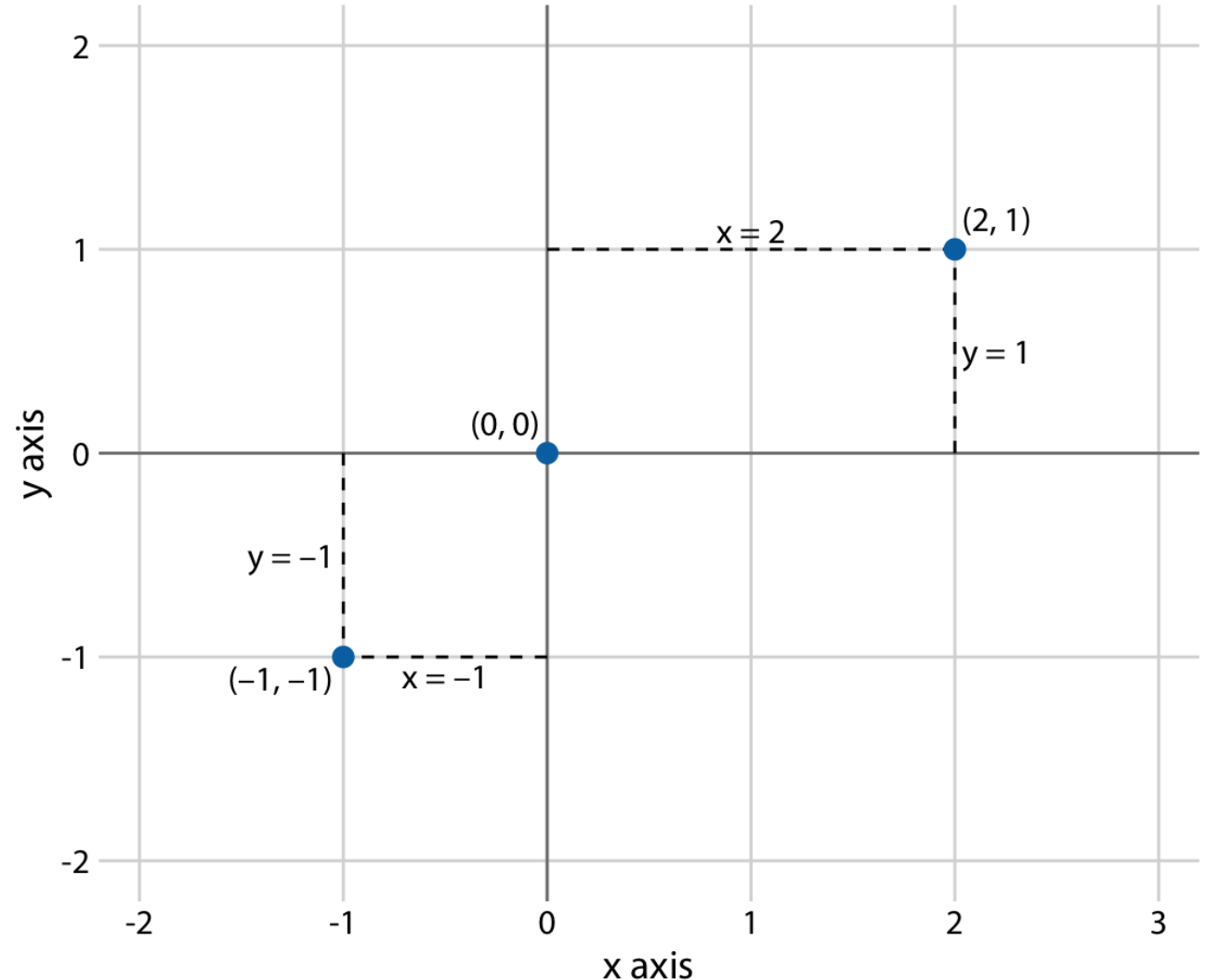
- Classificação “principal”: elementos estéticos capazes ou não se representar variáveis contínuas
- Variáveis contínuas: Posição, tamanho, cor e espessura da linha/traçado
- Variáveis discretas: Forma, tipo de linha/traçado

Escalas de posição: eixos e sistemas de coordenadas

- Escala de posição: “onde” diferentes valores de variáveis estão representados visualmente em uma figura
- Visualização 2D: dois valores numéricos são necessários para especificar uma posição individual para um ponto
- Uso de duas escalas de posição (normalmente “x” e “y”)
- Sistema de coordenadas: combinação de escalas de posição e seu respectivo arranjo geométrico

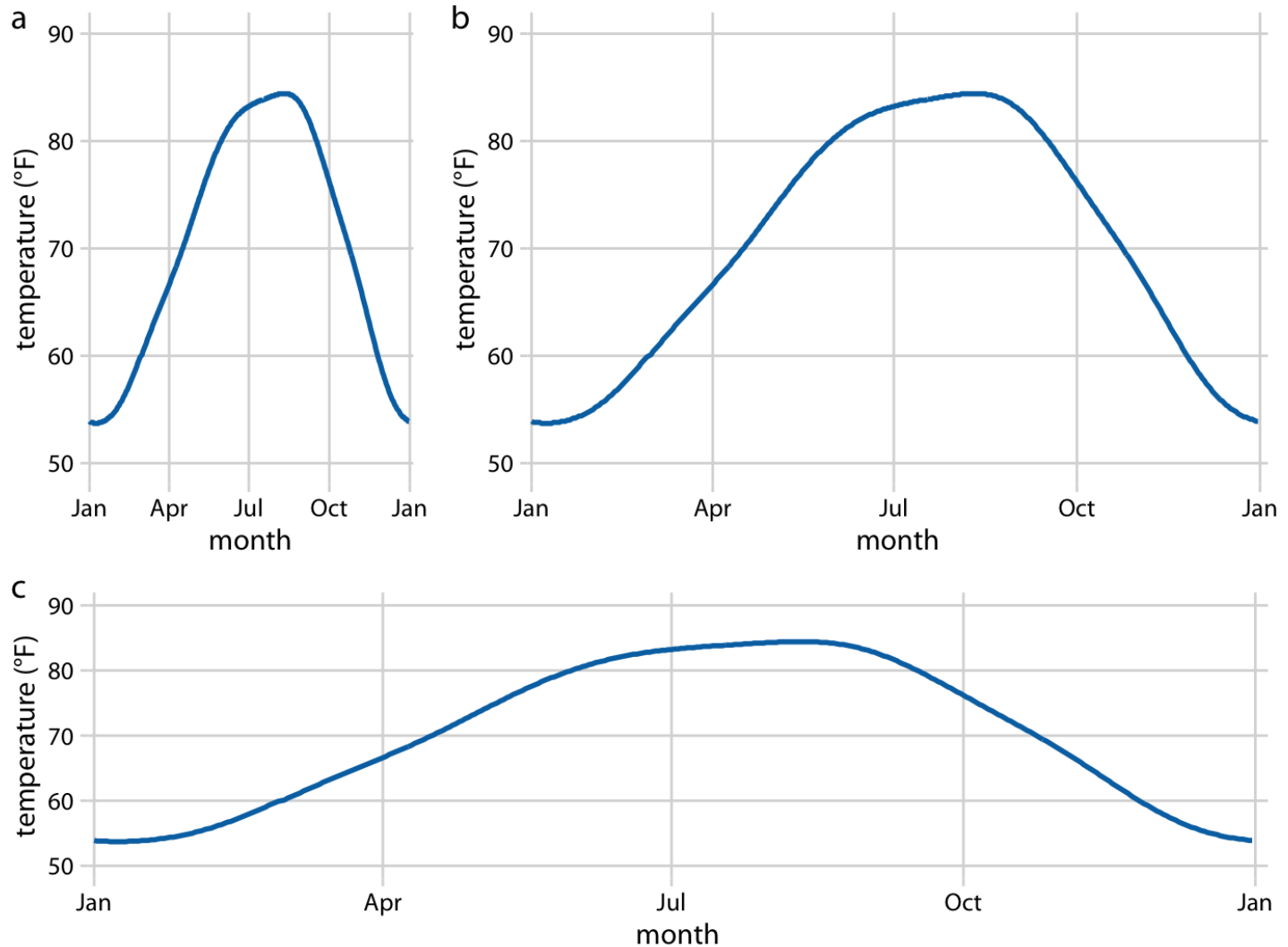
Coordenadas cartesianas

- Sistema de coordenadas mais utilizado em visualizações 2D
- A localização de cada ponto é especificada por um valor de “x” e “y”
- Eixos “x” e “y” perpendiculares entre si e os valores nos eixos estão espaçados de forma regular (sistema linear)
- Eixos são escalas contínuas de posição: podem representar números reais positivos ou negativos



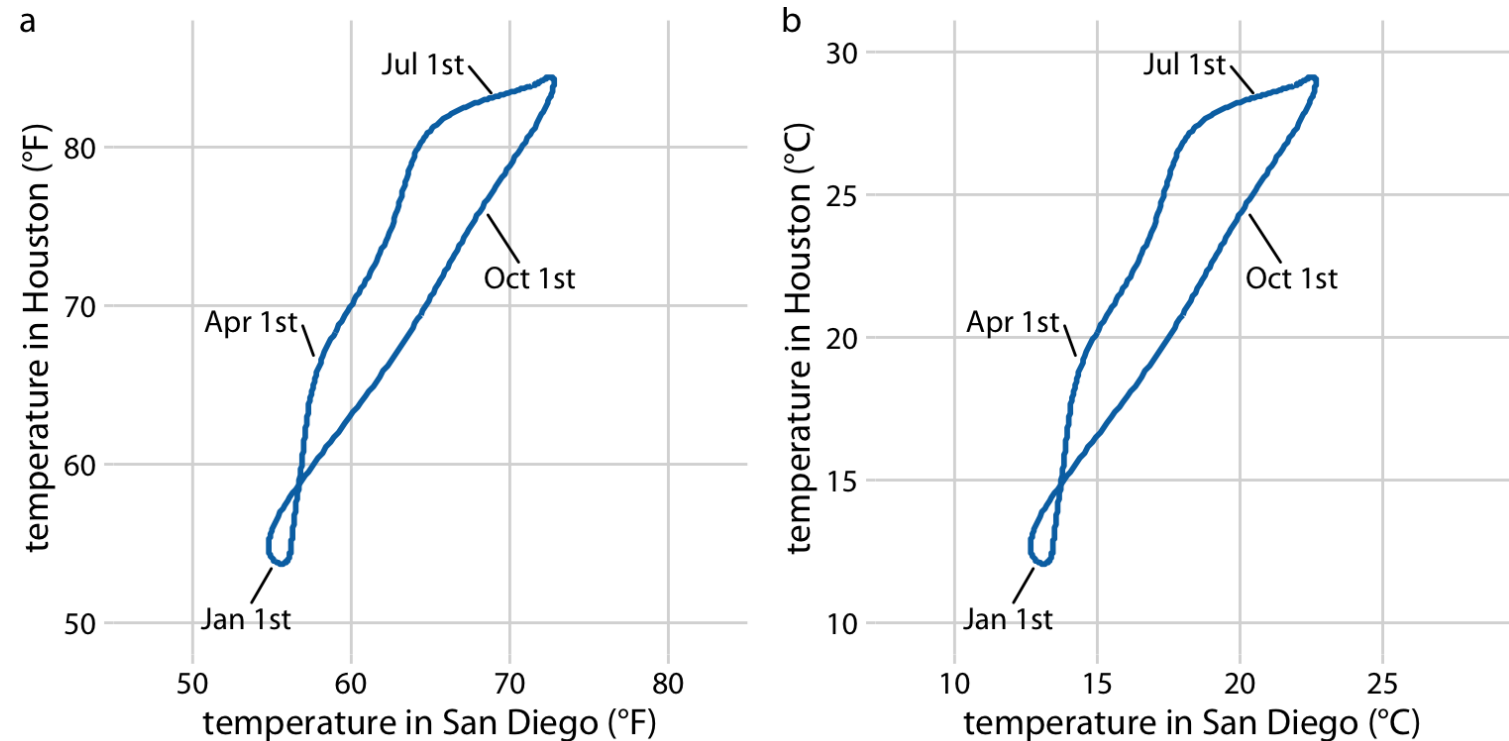
Coordenadas cartesianas

- Eixos representam unidades das variáveis sendo analisadas
- Se os eixos representarem variáveis de diferentes unidades, é possível espaçar ou comprimir uma delas em relação à outra e ainda ser fidedigno aos dados
- Escolher a proporção adequada à mensagem a ser enfatizada



Coordenadas cartesianas

- Se os eixos representam variáveis da mesma unidade, o espaçamento interno em cada eixo deve ser igual ao do outro eixo
- Transformações lineares: aspecto visual invariável em sistemas de coordenadas cartesianas



Adaptado de:
WILKE, C. O. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. O'Reilly Media, 2019.

Escalas não-lineares para eixos

- Escalas não-lineares: espaçamento interno no eixo não é proporcional ao espaçamento visualizado graficamente, e vice-versa
- Escala logarítmica (log scale):
 - Cada posição no espaçamento corresponde à multiplicação da variável por um valor fixo
 - Determinar o expoente (logaritmo) ao qual valor fixo (base) deve ser elevada para produzir o valor original (logaritmando)

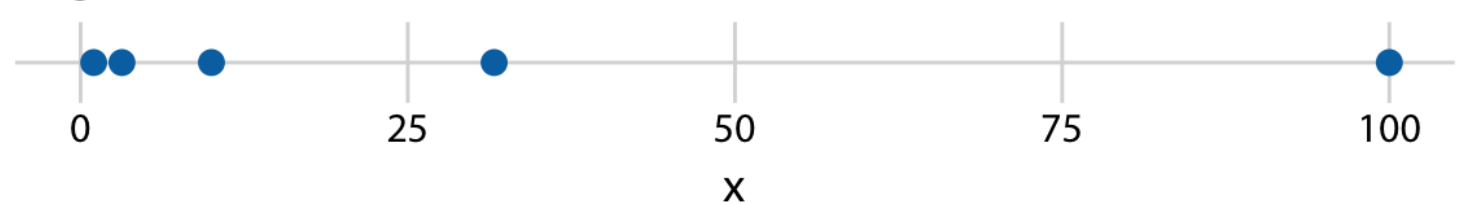
The diagram illustrates the components of the logarithmic equation $\text{Log}_a b = x \Leftrightarrow a^x = b$. Three light blue rectangular boxes are positioned around the equation: 'Logaritmando' (Logarithmand) at the top left, 'Logaritmo' (Logarithm) at the top right, and 'Base' (Base) at the bottom left. Curved arrows point from each box to its corresponding part in the equation: from 'Logaritmando' to b , from 'Logaritmo' to x , and from 'Base' to a .

$$\text{Log}_a b = x \Leftrightarrow a^x = b$$

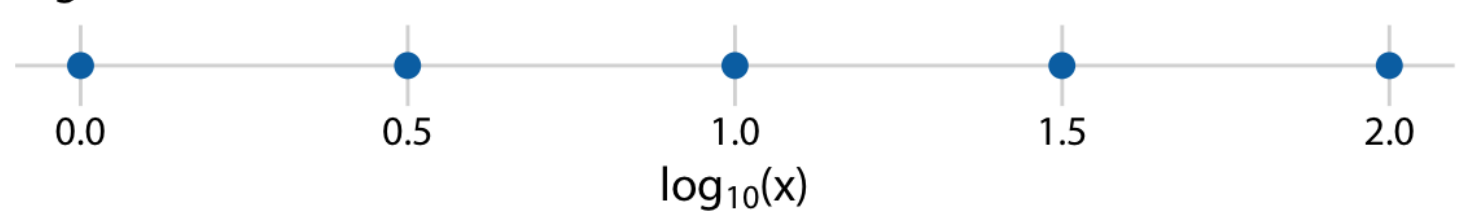
Escalas não-lineares para eixos

- Valores 1, 3.16, 10, 31.6 e 100 mapeados em escalas linear e logarítmica
- Matematicamente, sem diferença entre plotar os dados na escala linear ou logarítmica
- Preferencialmente utilizar logarítmica para evitar problemas de interpretação e sempre informar a base de transformação na legenda do eixo

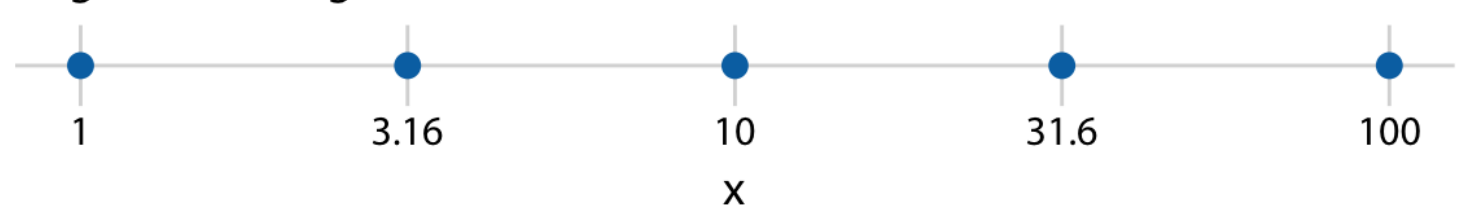
original data, linear scale



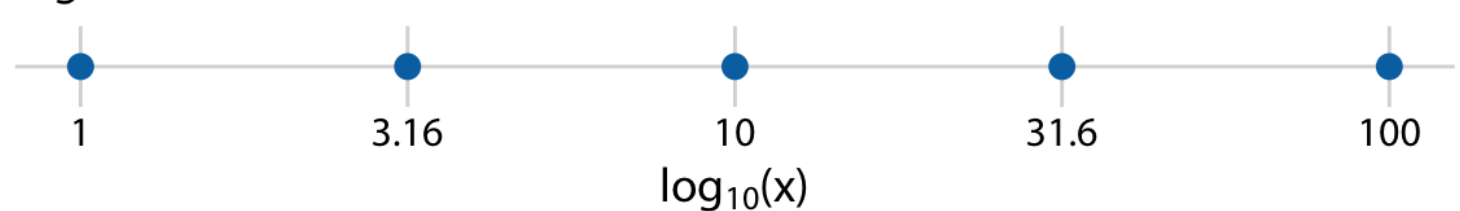
log-transformed data, linear scale



original data, logarithmic scale



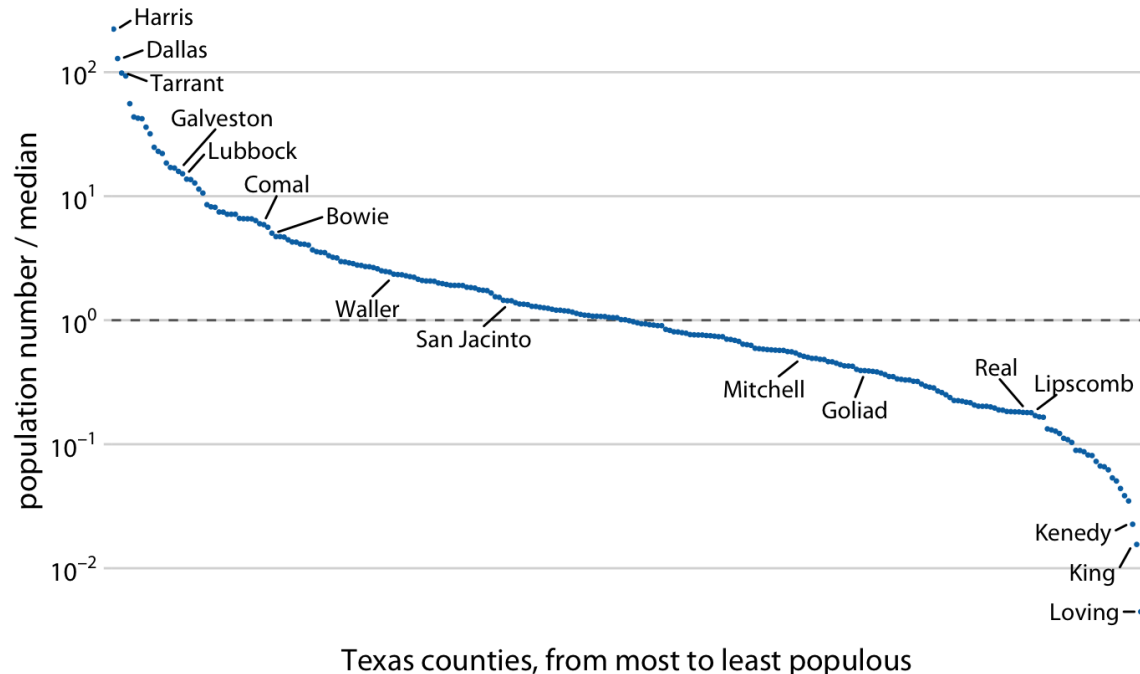
logarithmic scale with incorrect axis title



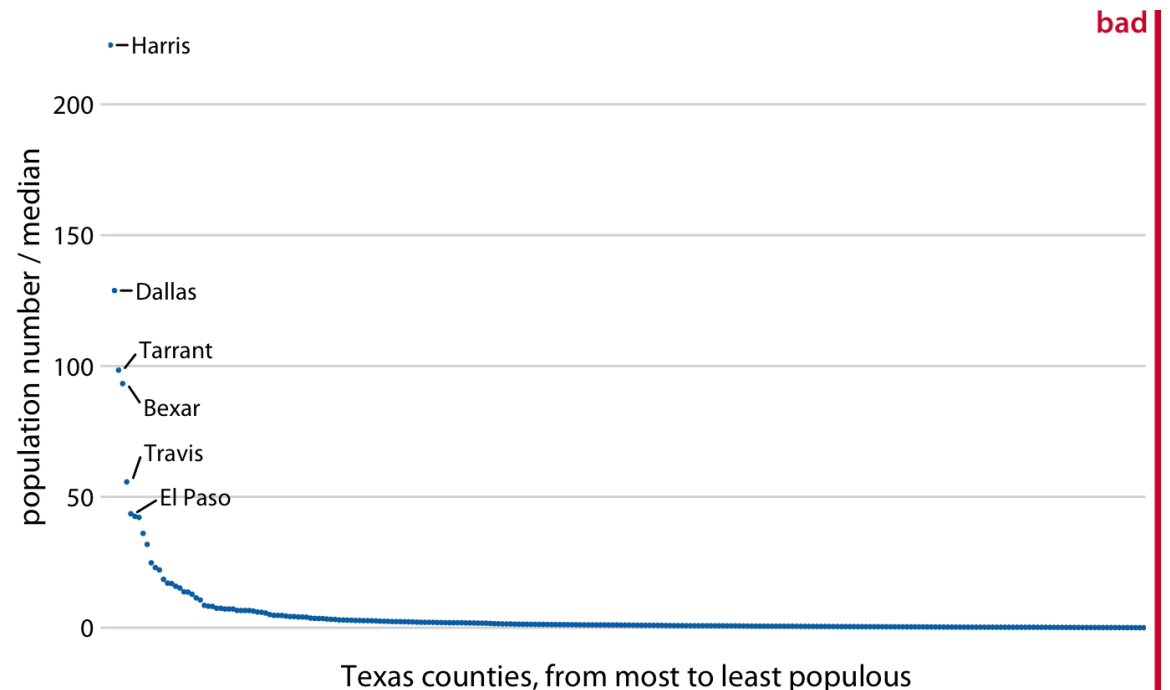
wrong

Escalas não-lineares para eixos

- Uso frequente de escalas logarítmicas para dados obtidos via multiplicação ou divisão, como razões entre duas variáveis
- Representação de razões em escalas lineares: ênfase em razões >1 e falta de resolução em razões <1

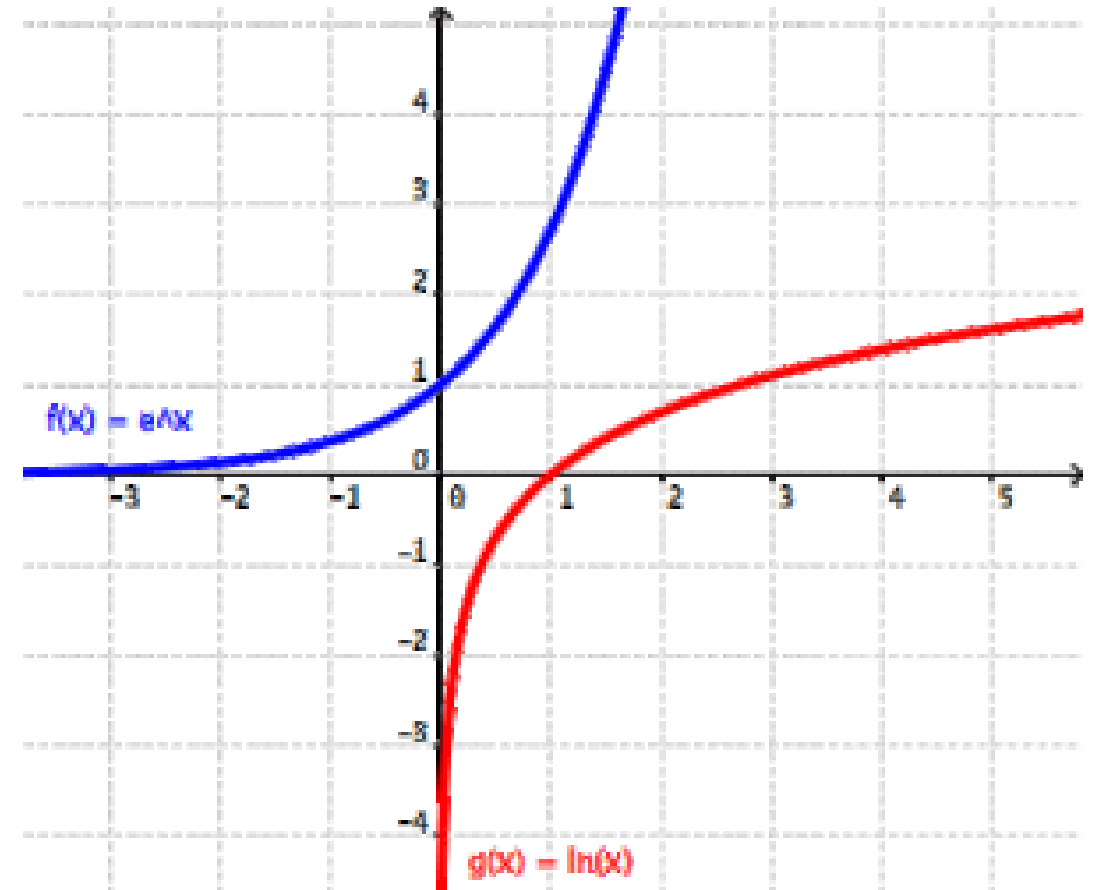


Adaptado de:
WILKE, C. O. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. O'Reilly Media, 2019.



Escalas não-lineares para eixos

- Ponto médio em escala logarítmica é 1, assim como o 0 seria em uma escala linear
- Valores superiores e inferiores a 1 como resultados de multiplicação e divisão, respectivamente
- Impossível transformar 0 para uso em uma escala logarítmica, não há número ao qual um expoente possa ser elevado para que o resultado seja 0
- Limite da função exponencial quando x se aproxima de zero é $-\infty$

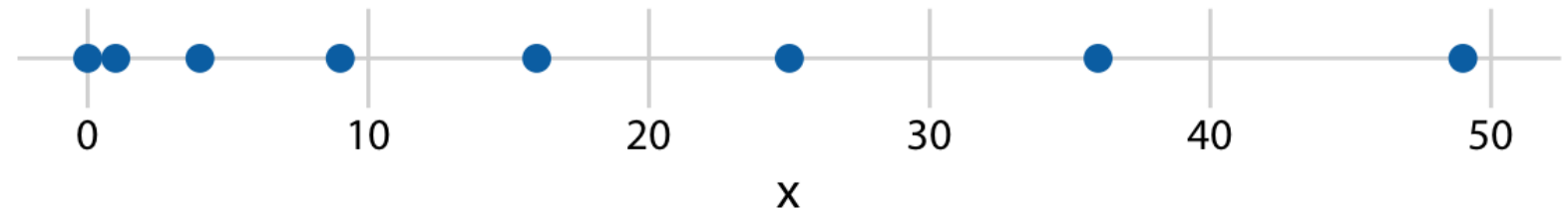


[Fonte](#)

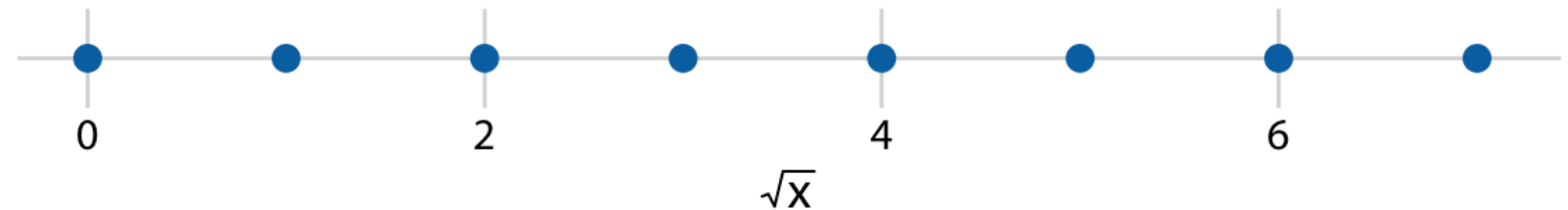
Escalas não-lineares para eixos

- Alternativa: usar escala baseada em transformação por raiz-quadrada, que permite uso e representação de valor zero
- Assim como na escala logarítmica, matematicamente, sem diferença entre plotar os dados na escala linear ou de raiz-quadrada

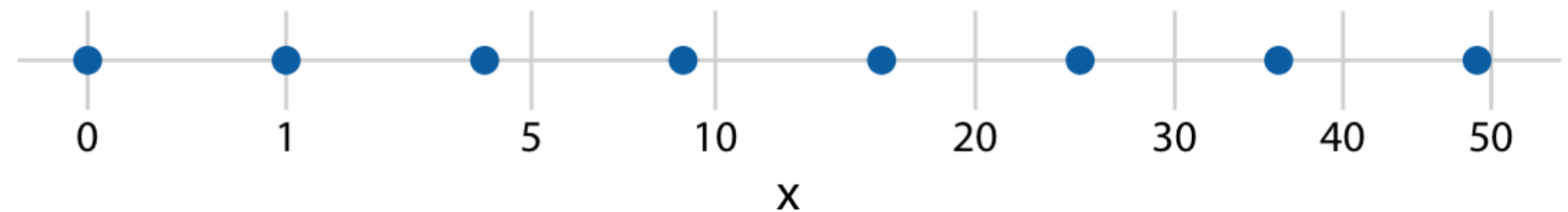
original data, linear scale



square-root-transformed data, linear scale



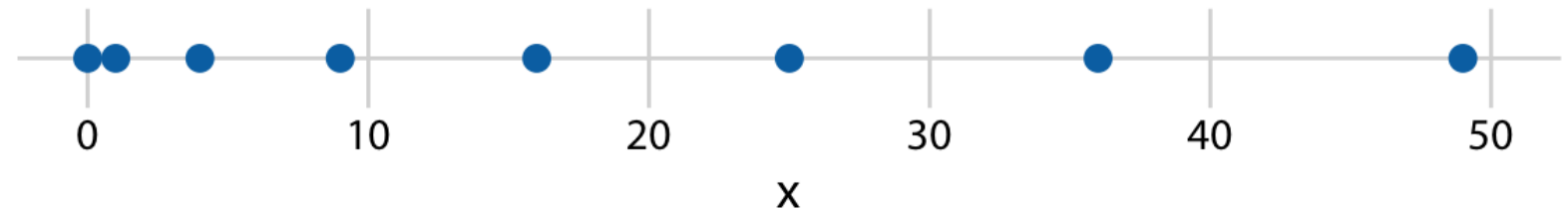
original data, square-root scale



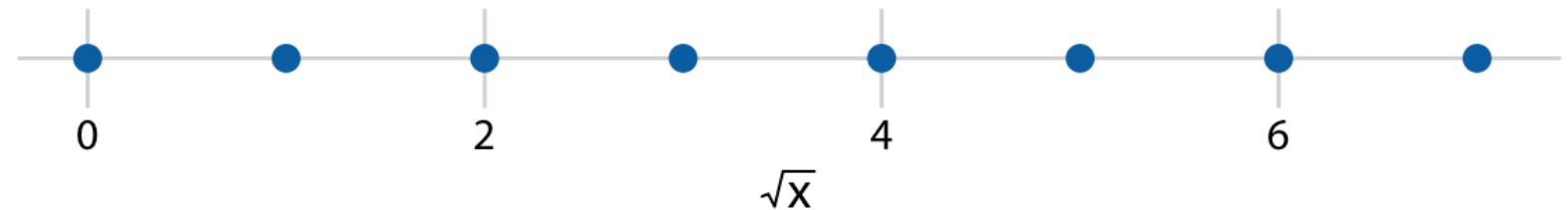
Escalas não-lineares para eixos

- Primeiro problema: falta de “proporcionalidade” nas unidades de espaçamento no eixo
 - Escala linear: adição/subtração
 - Escala logarítmica: multiplicação/divisão
 - Escala de raiz quadrada: depende do valor em que se inicia

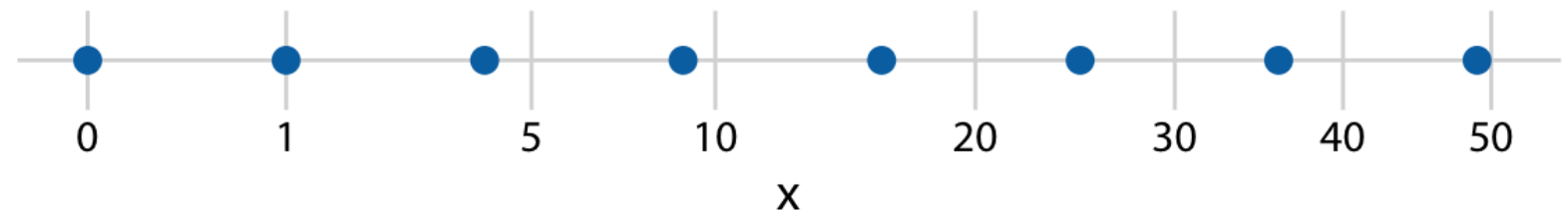
original data, linear scale



square-root-transformed data, linear scale



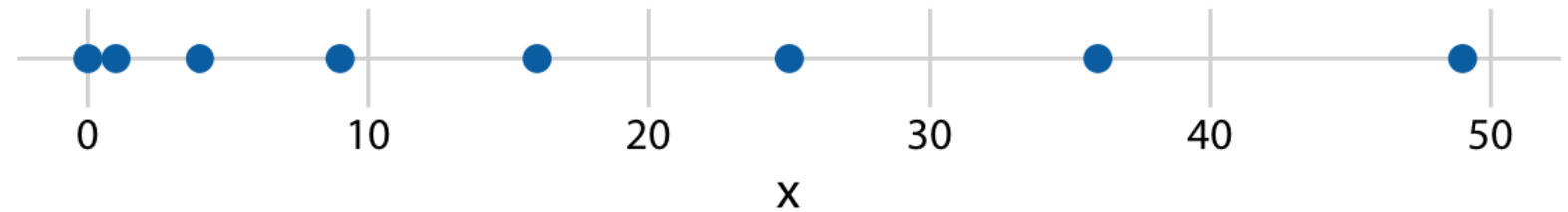
original data, square-root scale



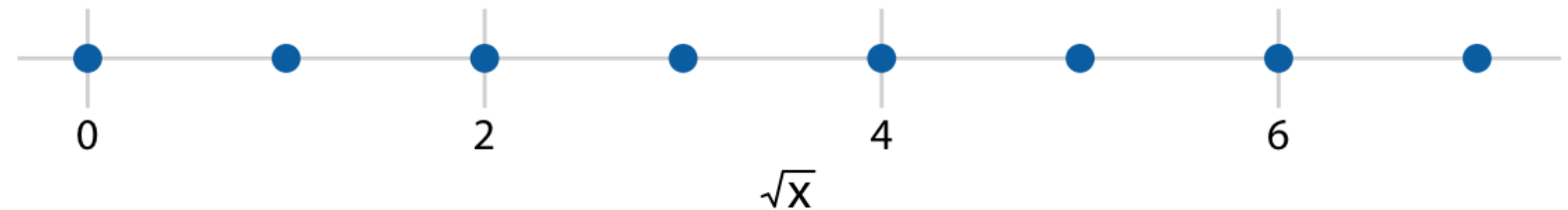
Escalas não-lineares para eixos

- Segundo problema: dificuldade de distribuir o espaçamento no eixo
 - Espaçamento uniforme: valores não-intuitivos (no exemplo, seriam 0, 4, 25, 49, 81)
 - Intervalos lineares: em alguns casos, número excessivo de intervalos em alguma das extremidades do eixo

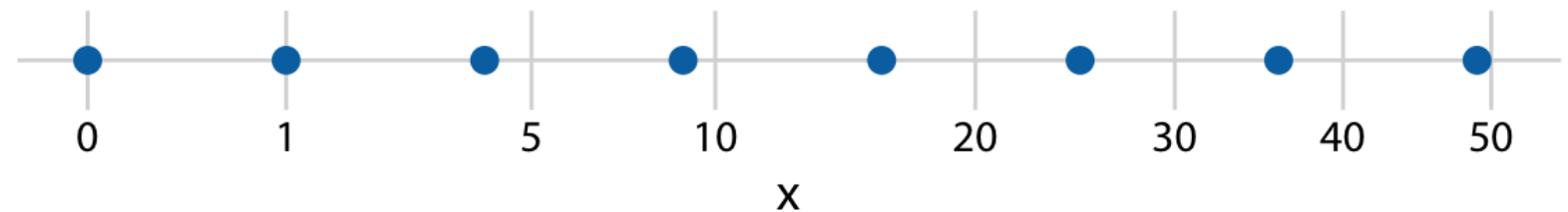
original data, linear scale



square-root-transformed data, linear scale

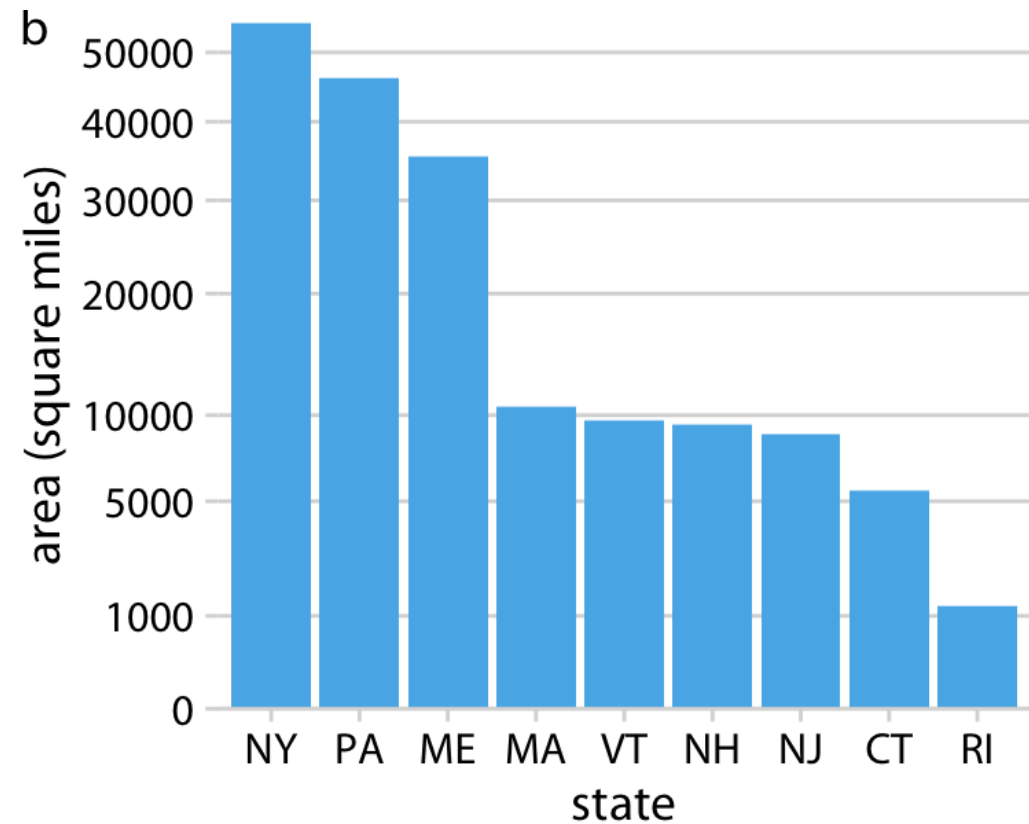
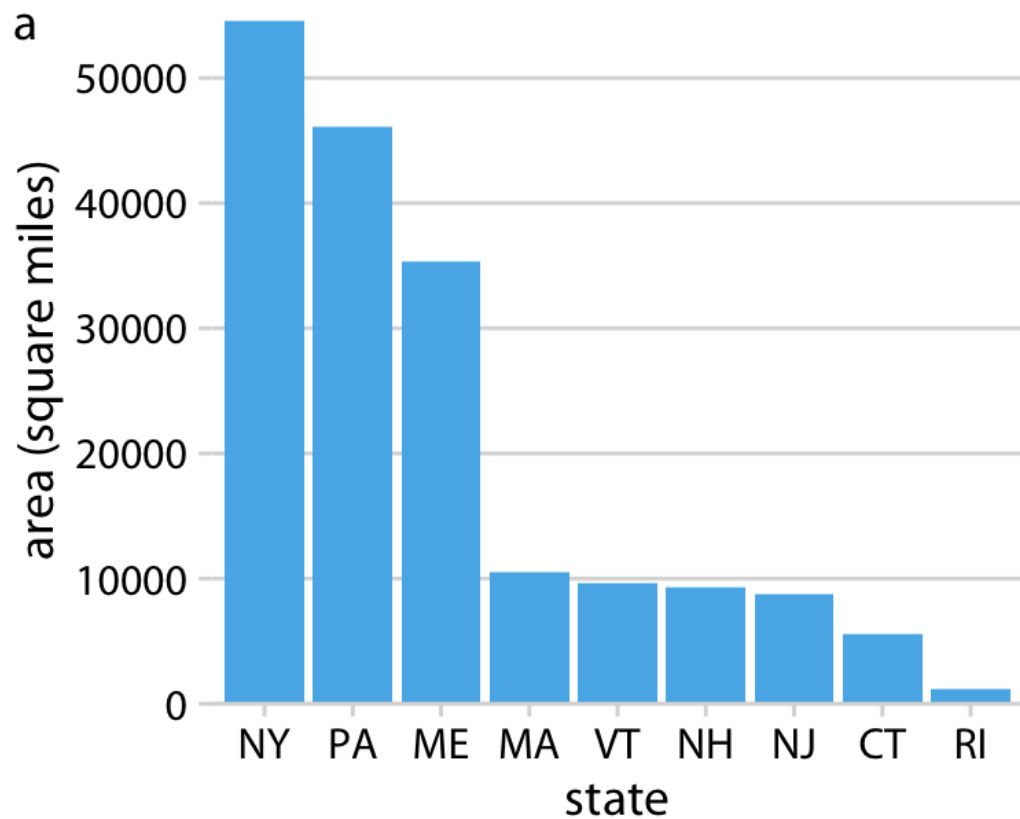


original data, square-root scale



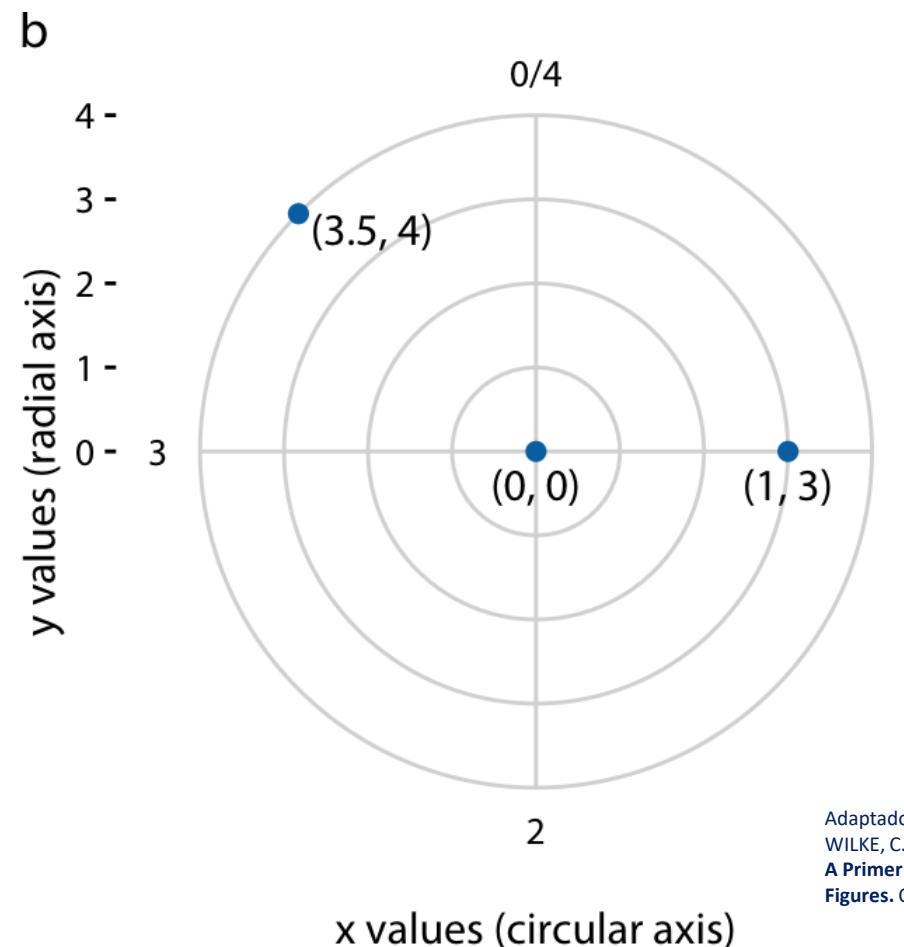
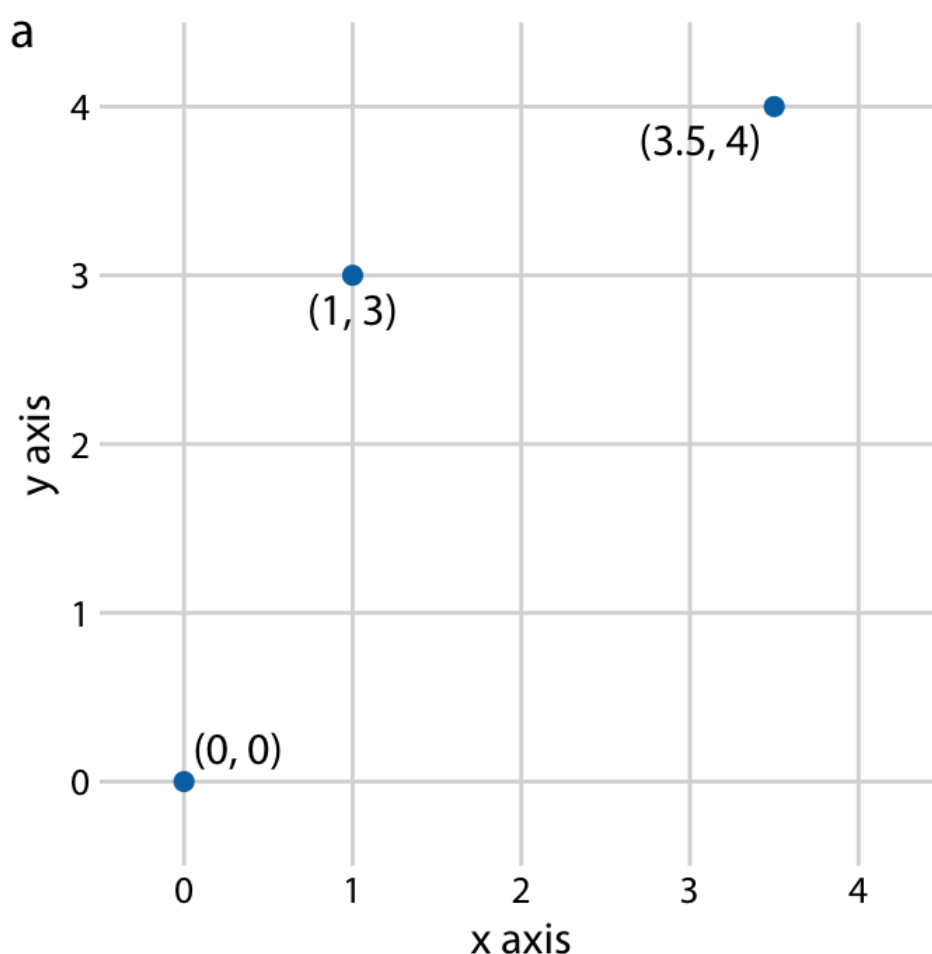
Escalas não-lineares para eixos

- Vantagens: representação de dados que naturalmente estejam relacionados à “quadrados”, como áreas em m^2 e km^2



Sistemas de coordenadas com eixos curvos

- Sistema de coordenadas polar: posições especificadas em ângulo, em distância radial a partir da origem, gerando um eixo circular



Adaptado de:
WILKE, C. O. **Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures**. O'Reilly Media, 2019.

Sistemas de coordenadas com eixos curvos

- Útil para dados de natureza periódica/cíclica (dados em uma extremidade da escala podem ser unidos ao dados da outra extremidade)
- 31 de Dezembro é o último dia de um ano, mas é também o dia anterior ao primeiro dia do ano seguinte
- Meio-dia é a última hora da manhã, mas também é a hora anterior à primeira hora da tarde



We are R-Ladies

@WeAreRLadies · Follow



Y'all there is a ggclock @rstudio #ggplot2 #rstats

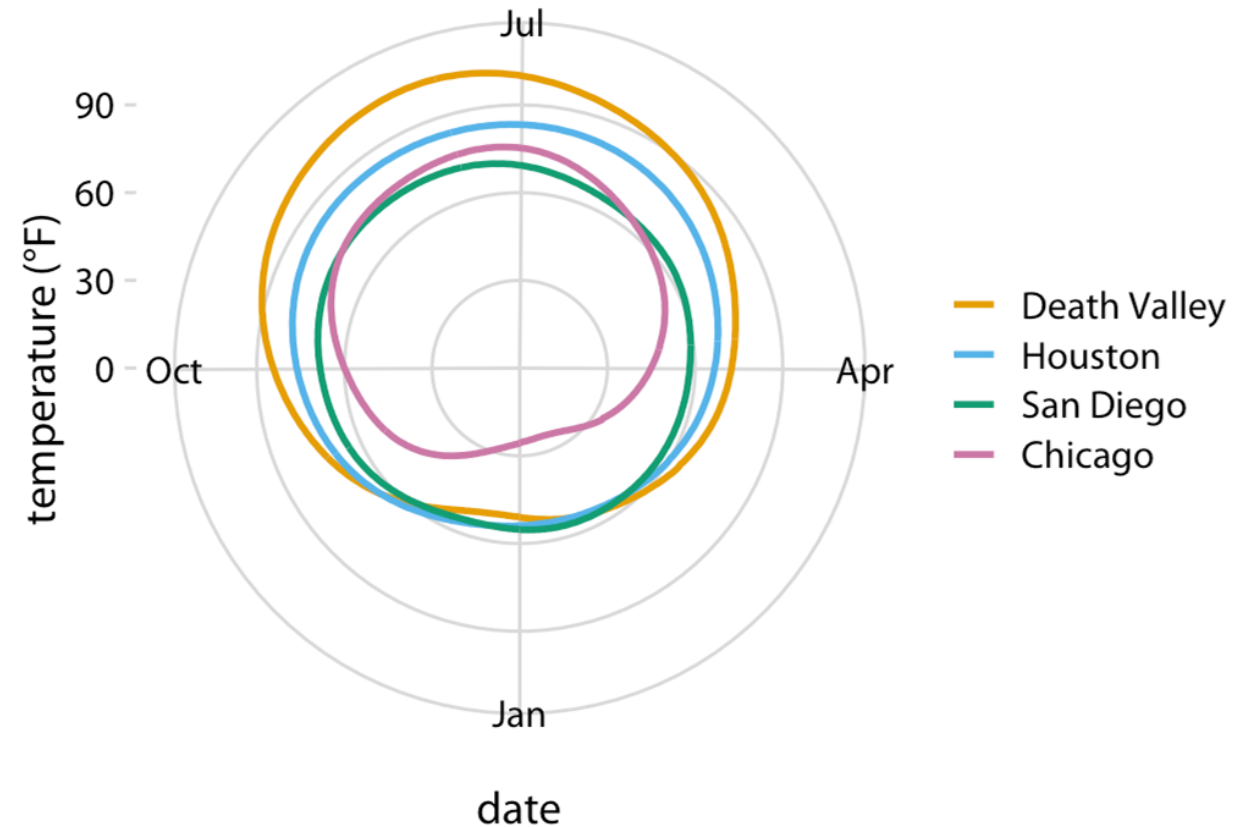


12:50 AM · Nov 9, 2018

[Fonte](#) ⓘ

Sistemas de coordenadas com eixos curvos

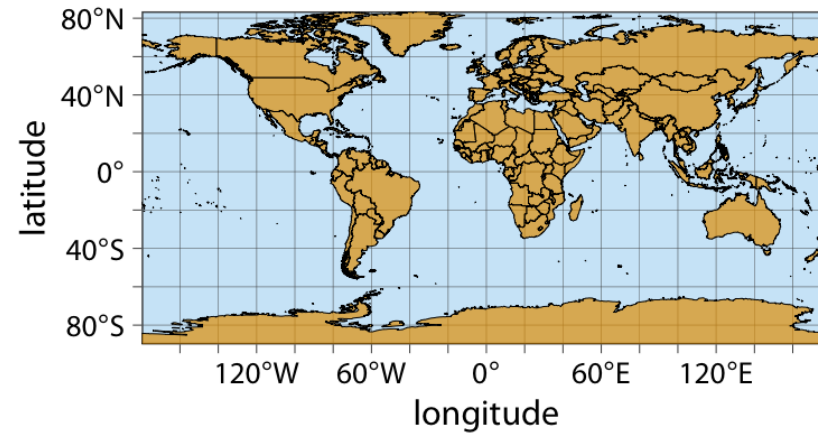
- Ênfase na característica cíclica dos dados
- Gráfico com sistema de coordenadas polar: visualização intuitiva de que as temperaturas são similares em Death Valley, Houston e San Diego entre o fim do outono e início da primavera
- Gráfico com sistema de coordenadas cartesianas: temperaturas de Dezembro e Janeiro estão em extremidades opostas, sem formar uma unidade visual



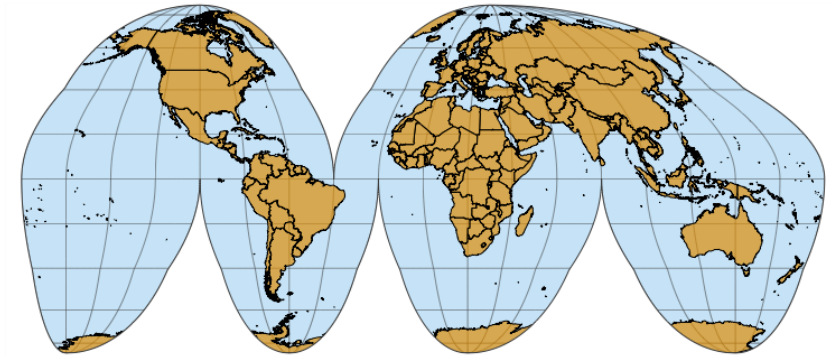
Sistemas de coordenadas com eixos curvos

- Utilidade em dados geoespaciais (mapas)
- Localizações no globo especificadas via latitude e longitude
- Representação em sistema cartesiano gera distorções em função do formato de geoide da Terra

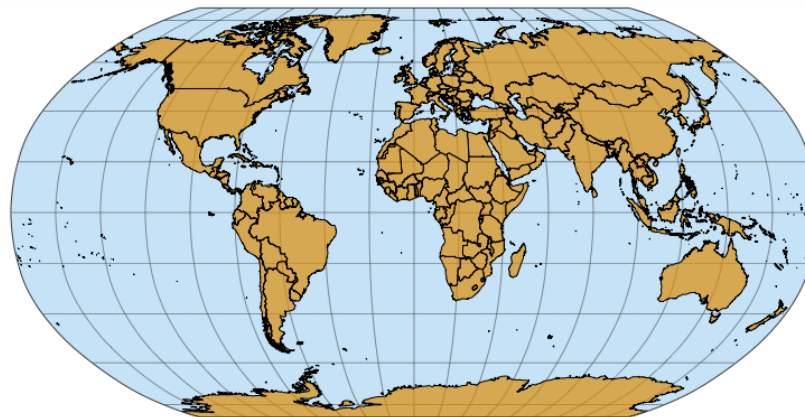
Cartesian longitude and latitude



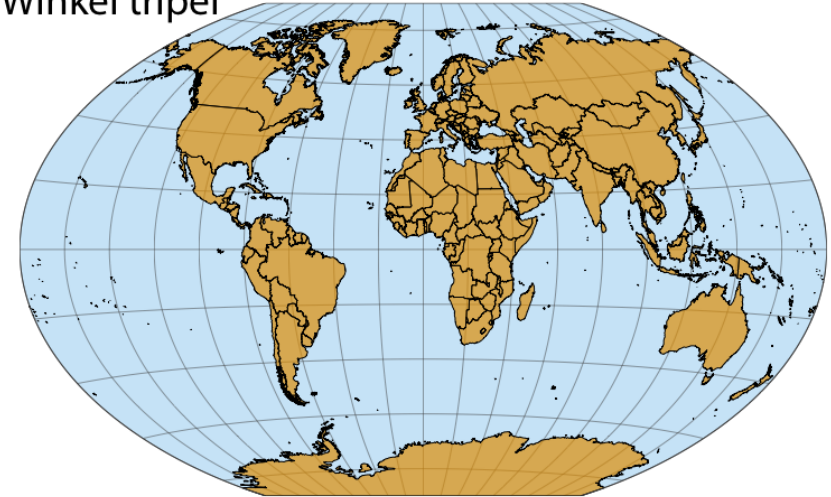
Interrupted Goode homolosine



Robinson



Winkel tripel



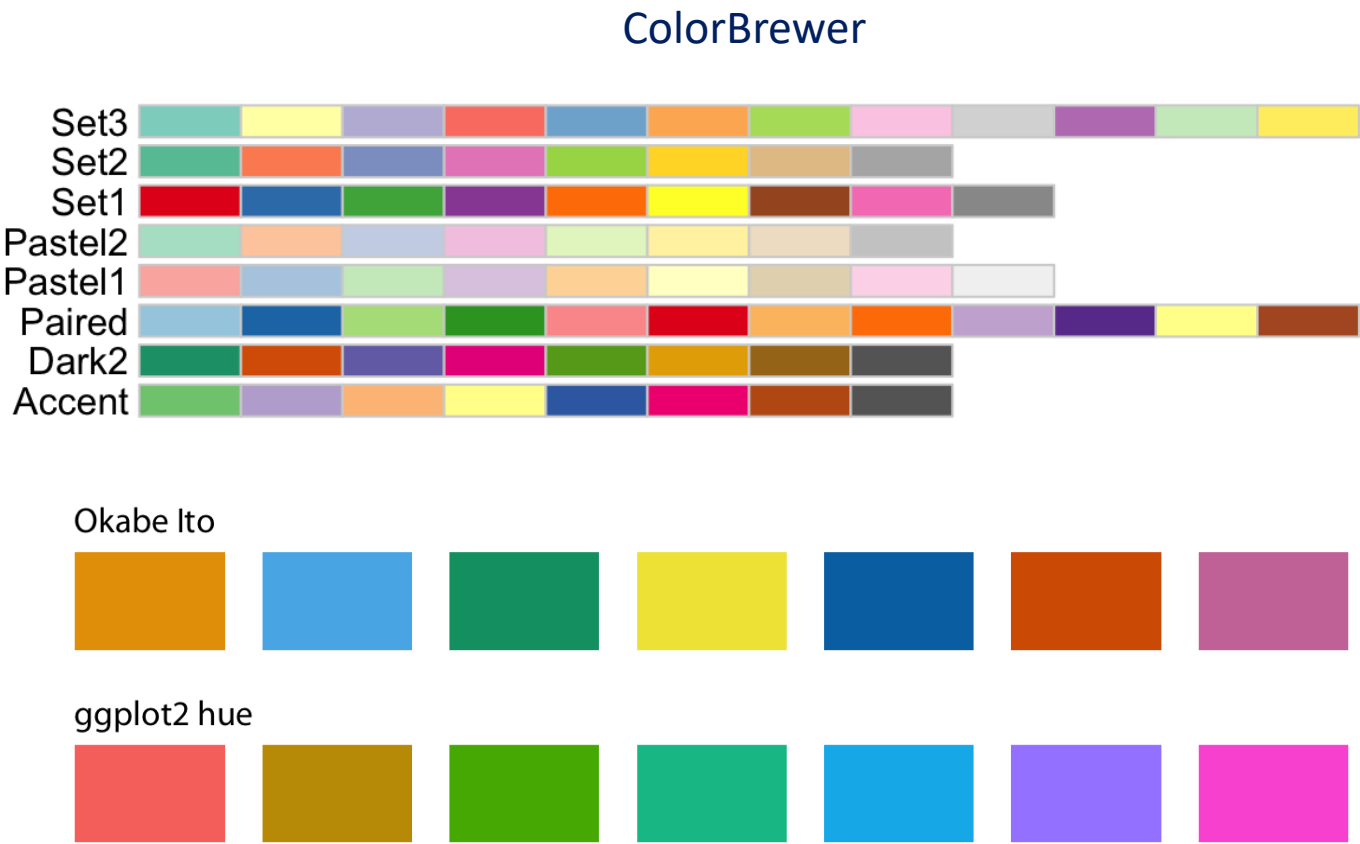
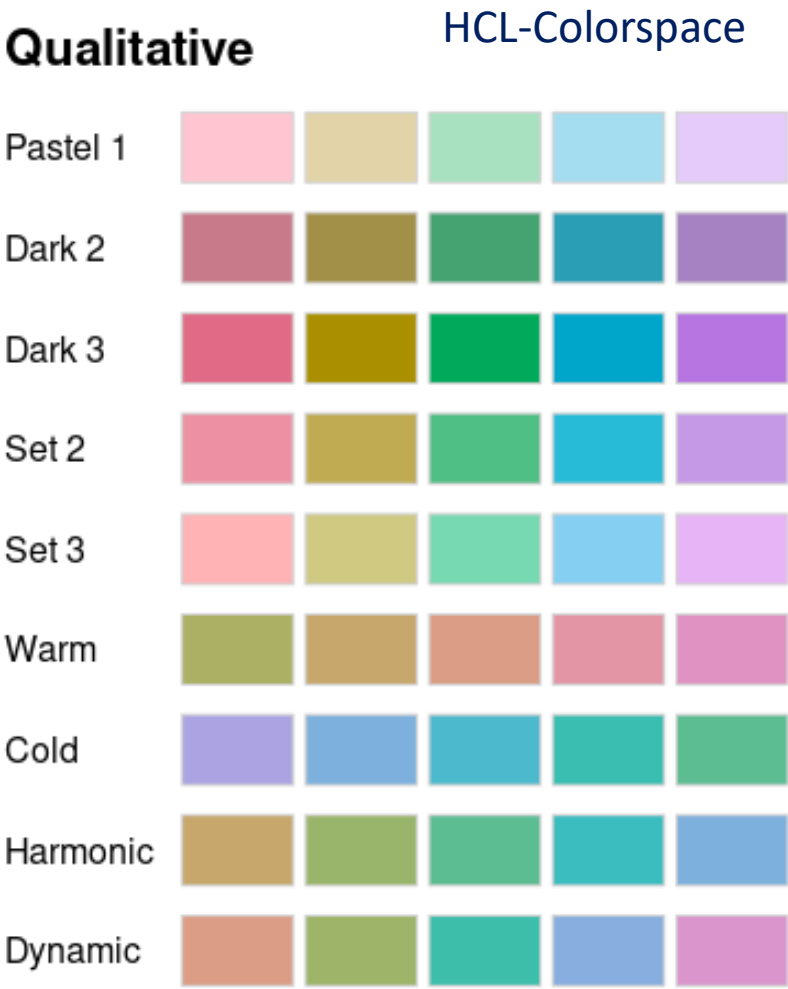
Escalas de cor

- Usos principais:
 - Distinção entre categorias
 - Representação de variáveis
 - Ênfase e realce em elementos específicos

Distinção entre categorias

- Distinção entre itens discretos ou grupos nominais sem ordem intrínseca
- Escalas qualitativas de cor: conjunto finito de cores que são visualmente distintas entre, mas visualmente equivalentes (diferenças de matiz, mas sem diferenças gritantes de saturação ou luminosidade)
- Sem cores que se destacam e sem criar a impressão de ordem (cores que se dispõem em sequência, ex: de mais claras para mais escuras)

Distinção entre categorias



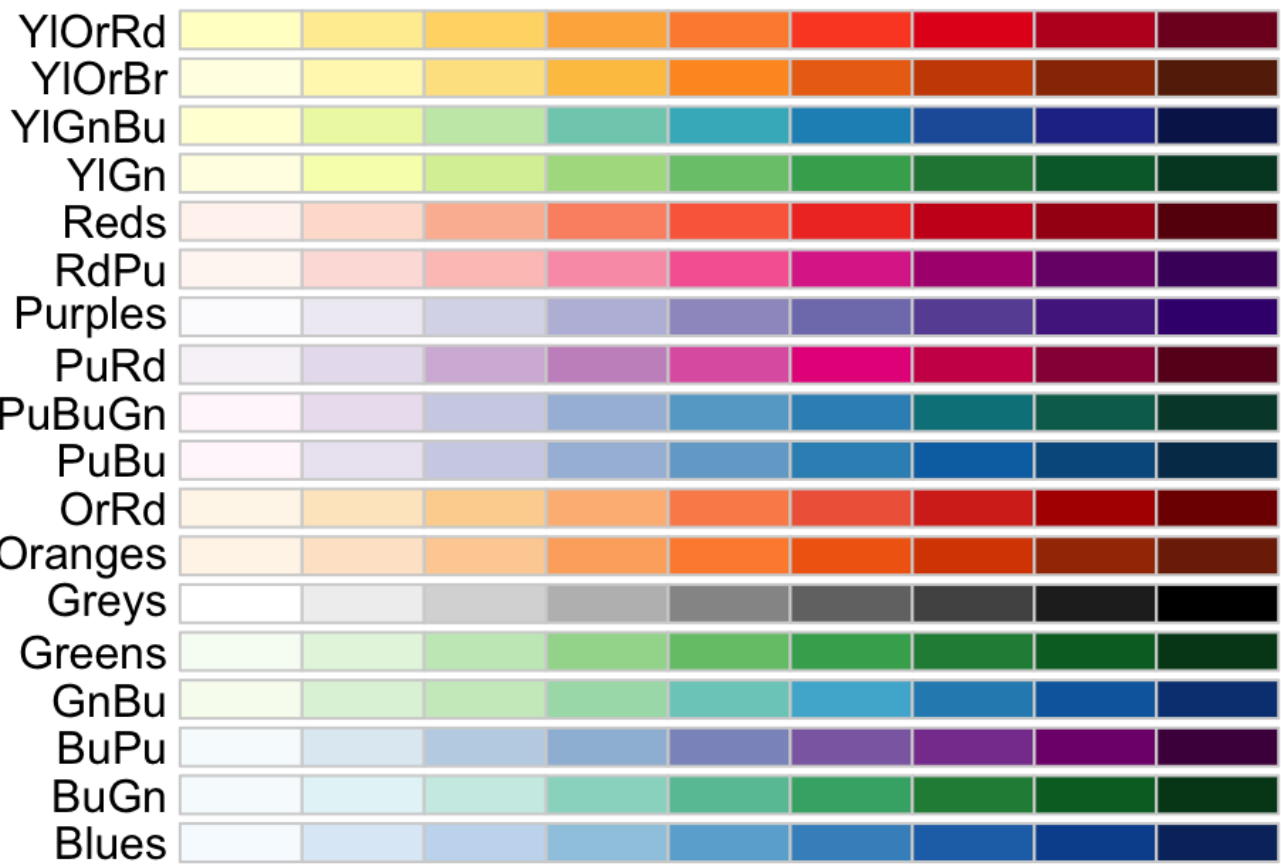
Adaptado de:
WILKE, C. O. **Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures**. O'Reilly Media, 2019.

Representação de variáveis

- Representação de valor contínuos por escalas contínuas
- Escalas quantitativas sequenciais de cor: indicação de valores maiores/menores, distância entre valores (uniformidade de percepção em toda a amplitude de cor)
- Variação de luminosidade/saturação em uma matiz ou múltiplas matizes (gradientes que existem na natureza)

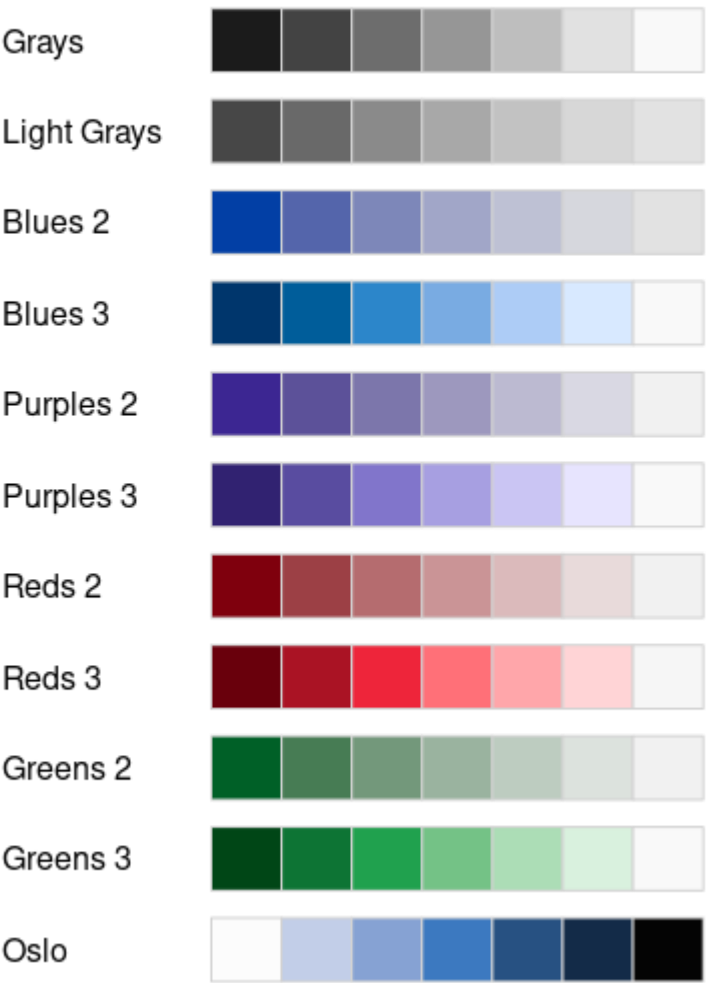
Representação de variáveis

HCL-Colorspace



ColorBrewer

Sequential (single-hue)

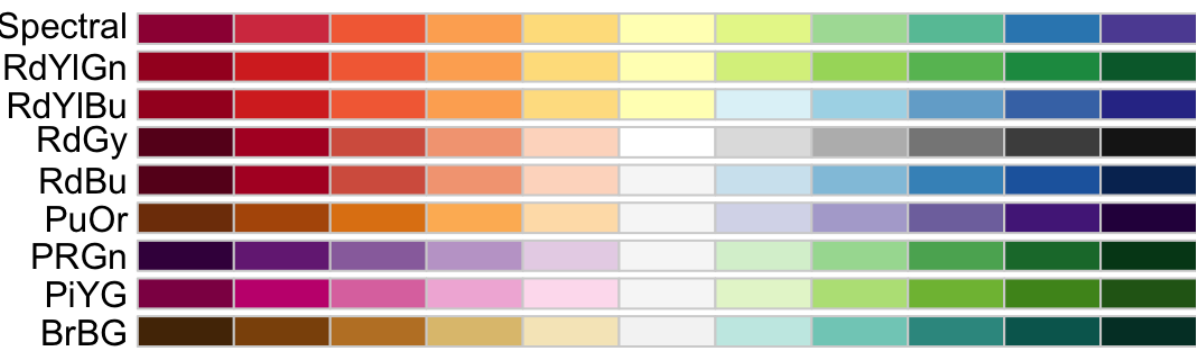


Representação de variáveis

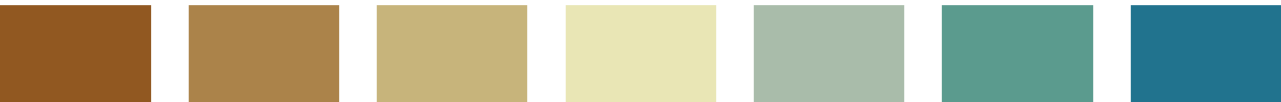
- Escalas quantitativas divergentes de cor: indicação da variação dos valores em relação a um ponto médio neutro
- Variação entre 0 e 100%, entre amplitude negativa e positiva
- Em geral, matizes distintas com um ponto médio em cor clara e neutra, progressão entre matizes de forma uniforme em ambas as direções

Representação de variáveis

ColorBrewer



CARTO Earth



Adaptado de:
WILKE, C. O. **Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures**. O'Reilly Media, 2019.

Diverging

HCL-Colorspace



Ênfase e realce em elementos específicos

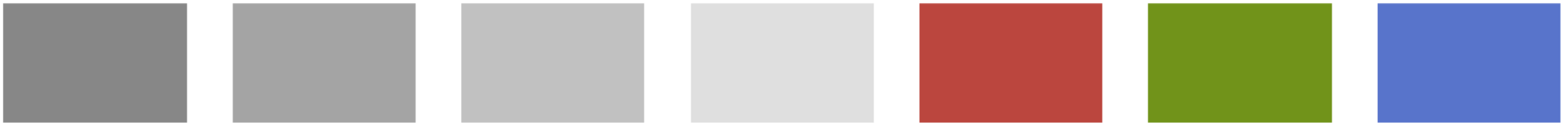
- Uso de cores que claramente se destacam em relação ao resto da figura
- Em geral, cores mais saturadas e com menor luminosidade combinadas a cores com maior luminosidade e menos saturadas
- Cores secundárias não devem competir por atenção entre si: em alguns casos, utilizar somente cinza e marcação por formas na informação que não deve ser destacada

Ênfase e realce em elementos específicos

Okabe Ito Accent



Grays with accents

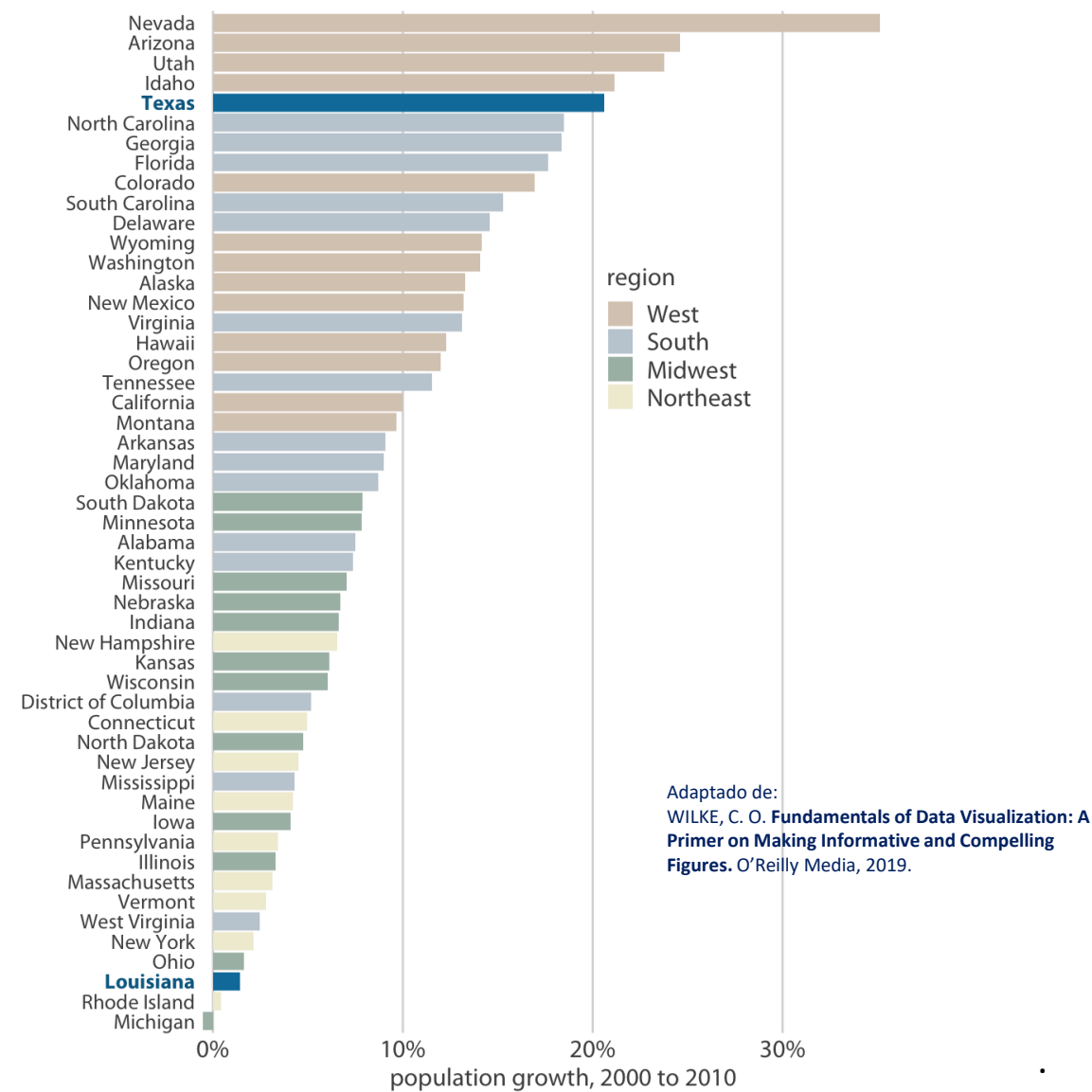
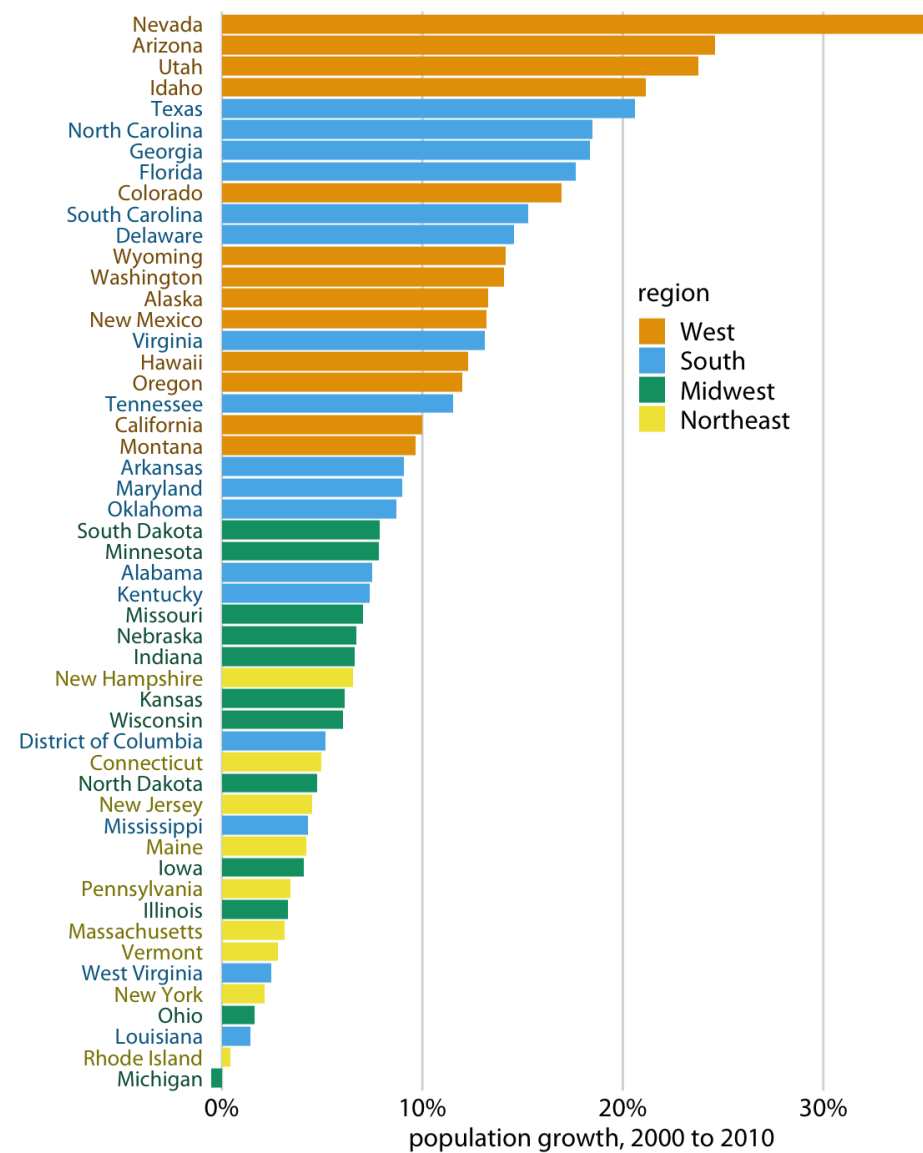


ColorBrewer Accent

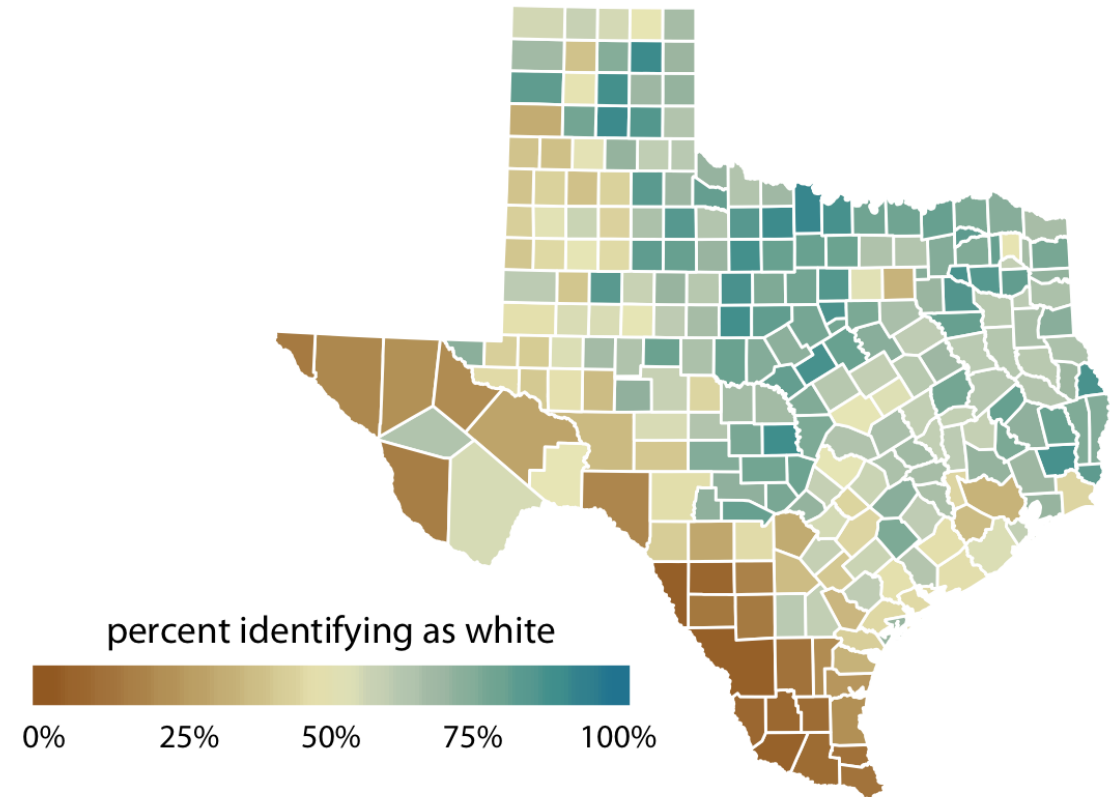
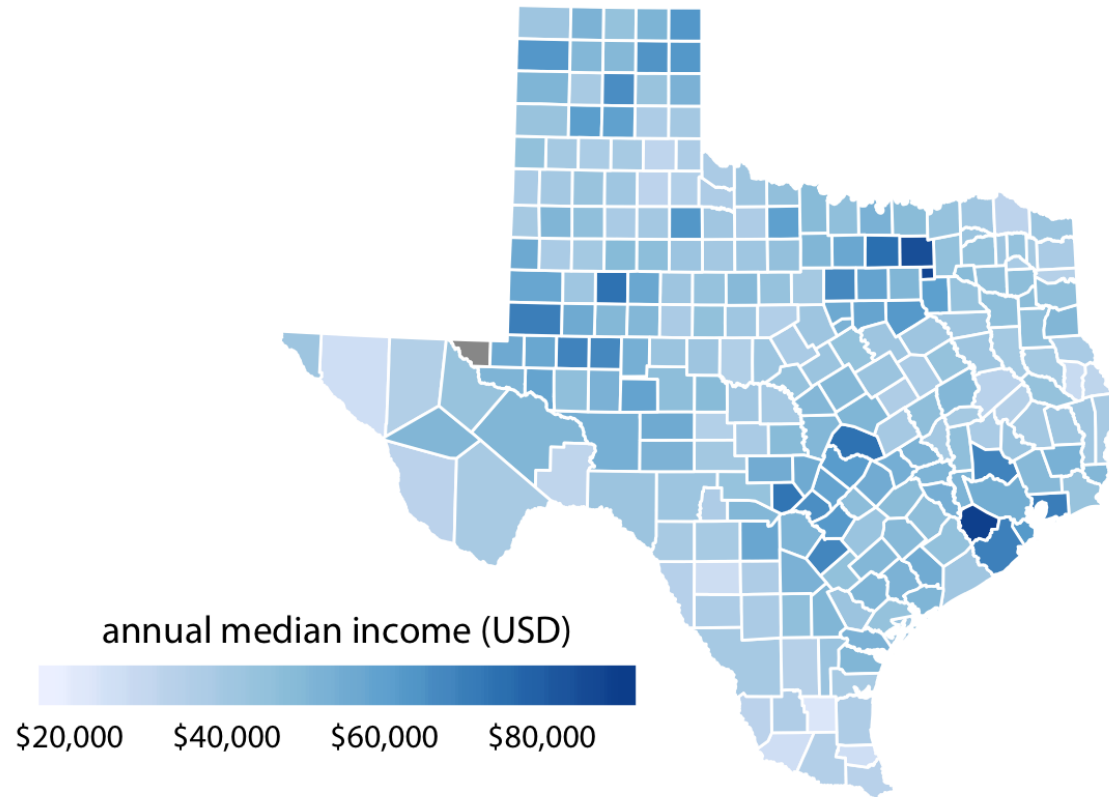


Adaptado de:
WILKE, C. O. **Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures**. O'Reilly Media, 2019.

Contando diferentes histórias com escalas de cor



Correlacionando diferentes histórias com escalas de cor



Adaptado de:
WILKE, C. O. **Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures**. O'Reilly Media, 2019.