

Visualização de Dados

ESTAT0011 – Estatística Aplicada

Prof. Dr. Sadraque E. F. Lucena

sadraquelucena@academico.ufs.br

Visualização de Dados

- A representação gráfica de dados tem a vantagem de trazer diversas informações relevantes sobre o comportamento dos dados.
- Os gráficos são utilizados para diversos fins:
 - Buscar padrões e relações;
 - confirmar (ou não) certas expectativas que se tinha sobre os dados;
 - descobrir novos fenômenos;
 - confirmar (ou não) suposições feitas sobre os procedimentos estatísticos usados; e
 - apresentar resultados de modo mais rápido e fácil.

Visualização de Dados

- Ao construir qualquer tipo de gráfico, é importante garantir que:
 - o gráfico receba um título adequado;
 - a legenda dever ser concisa e objetiva;
 - cada um dos eixos seja rotulado
 - uma escala sensata utilizada.

Visualização de Dados Categóricos

- Quando os dados são categóricos, o objetivo da visualização é representar a magnitude de valores quantitativos associados a categorias.
 - **Exemplos:** vendas de carros por marca, população por cidade, idade de atletas por esporte.
- Nesses casos, a visualização padrão é o gráfico de barras, com variações como barras simples, agrupadas ou empilhadas.

Visualização de Contagens

Gráfico de Barras

Considere o total de vendas de ingressos dos filmes mais populares em 2024 no Brasil.

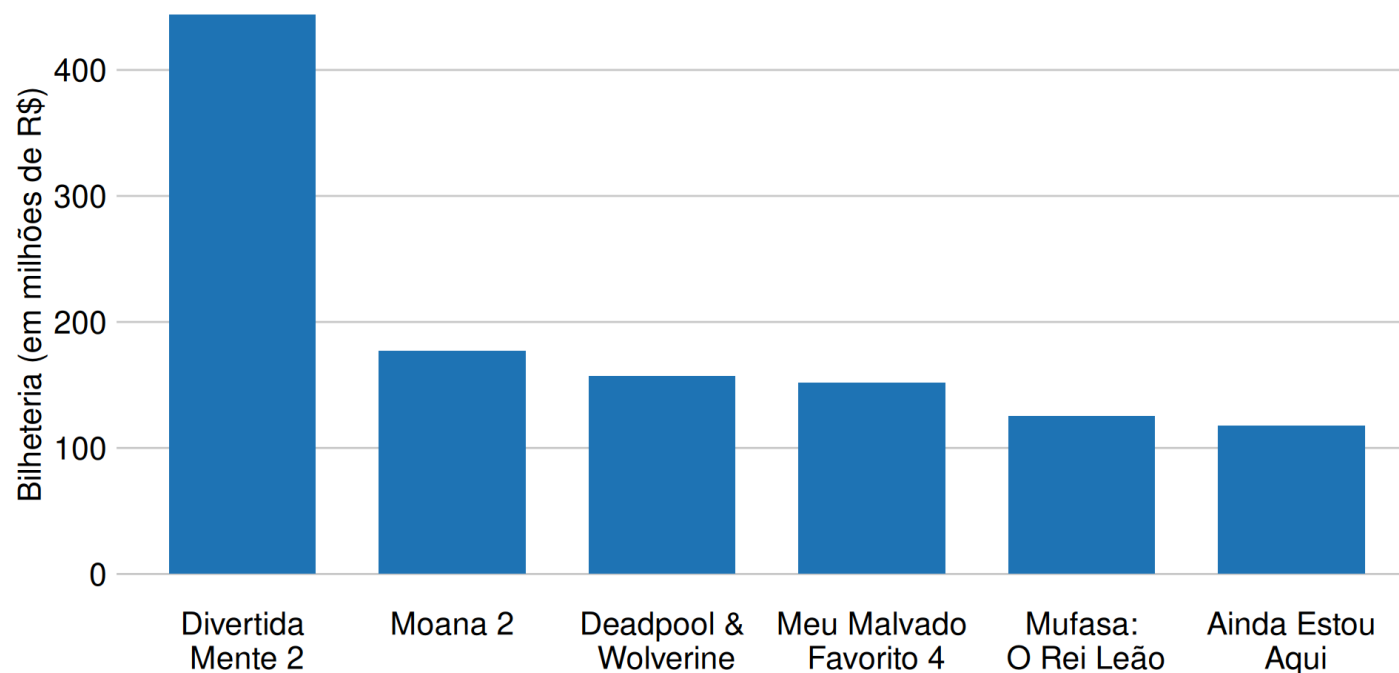
Posição	Título	Bilheteria
1	Divertida Mente 2	R\$ 443.789.000
2	Moana 2ª	R\$ 177.353.562
3	Deadpool & Wolverine	R\$ 156.920.000
4	Meu Malvado Favorito 4	R\$ 151.710.000
5	Mufasa: O Rei Leão	R\$ 125.414.354
6	Ainda Estou Aqui	R\$ 117.630.000

Fonte: Wikipedia.

Visualização de Contagens

Gráfico de Barras

- Esse tipo de dado é comumente visualizado com barras verticais.
- Para cada filme, desenhamos uma barra que começa no zero e se estende até o valor correspondente à bilheteria daquele ano.

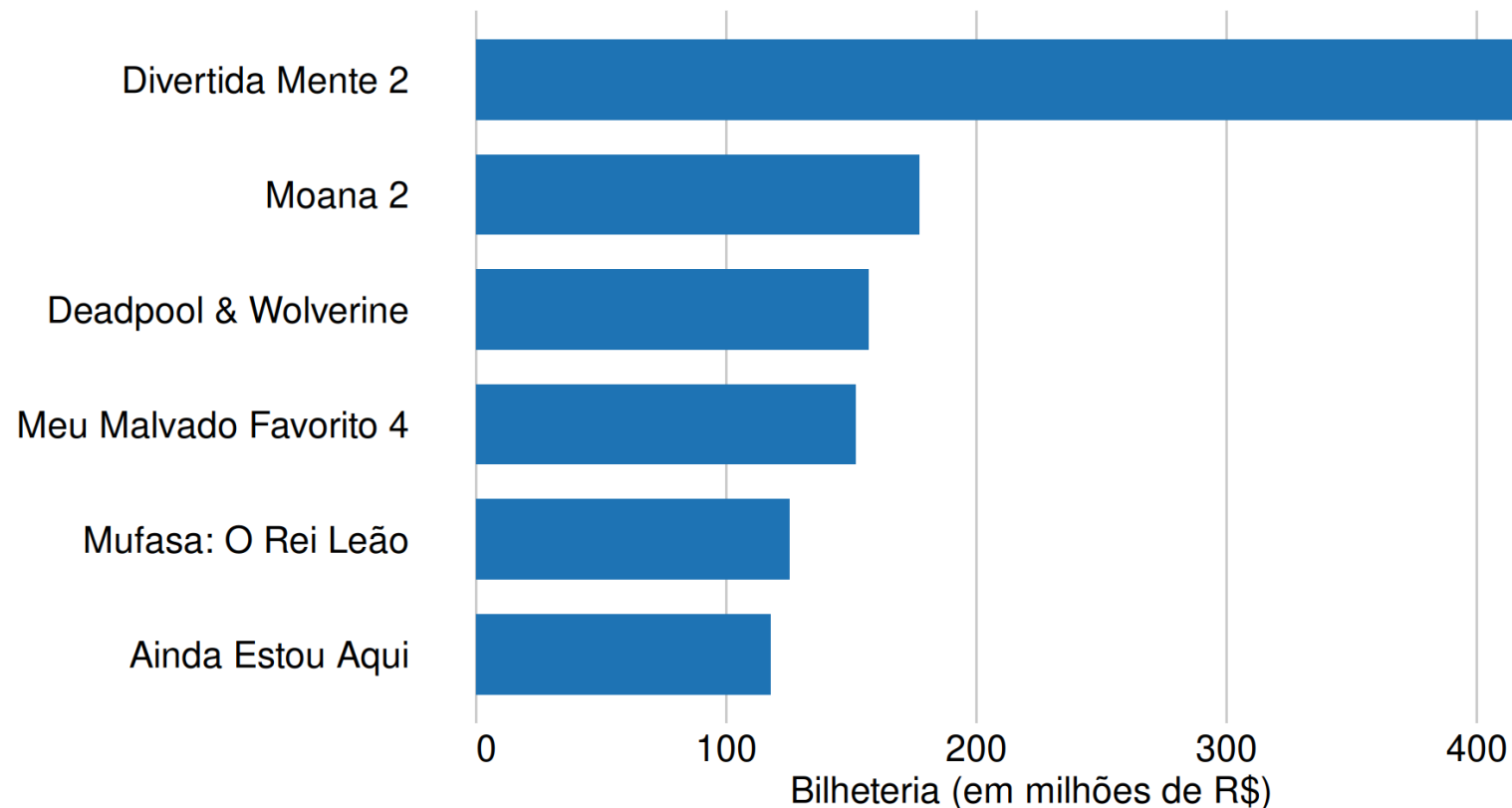


Fonte: Wikipedia.

Visualização de Contagens

Gráfico de Barras

- Quando os rótulos de cada barra ocupam muito espaço horizontal, você pode rotacionar o gráfico.

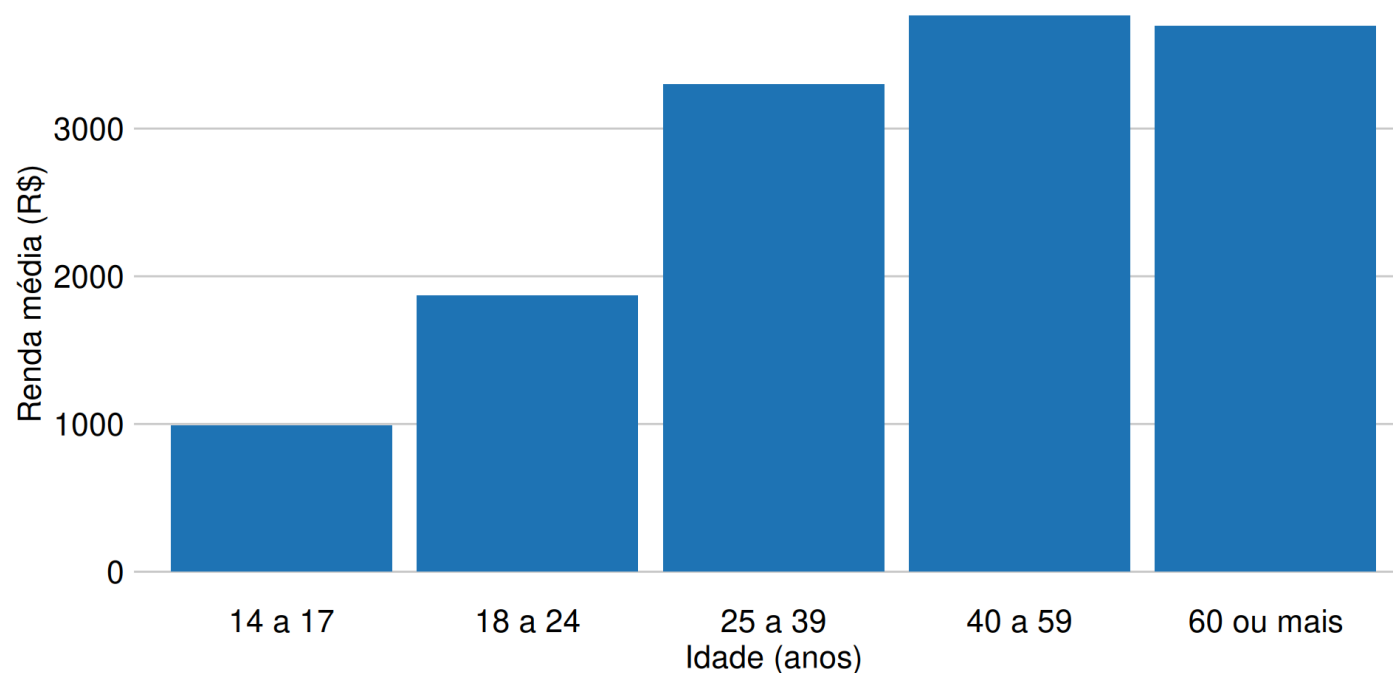


Fonte: Wikipedia.

Visualização de Contagens

Gráfico de Barras

- As barras do gráfico de barras devem estar sempre em ordem decrescente.
 - **Exceção:** variável cujos valores possuem uma ordem natural (variável ordinal).

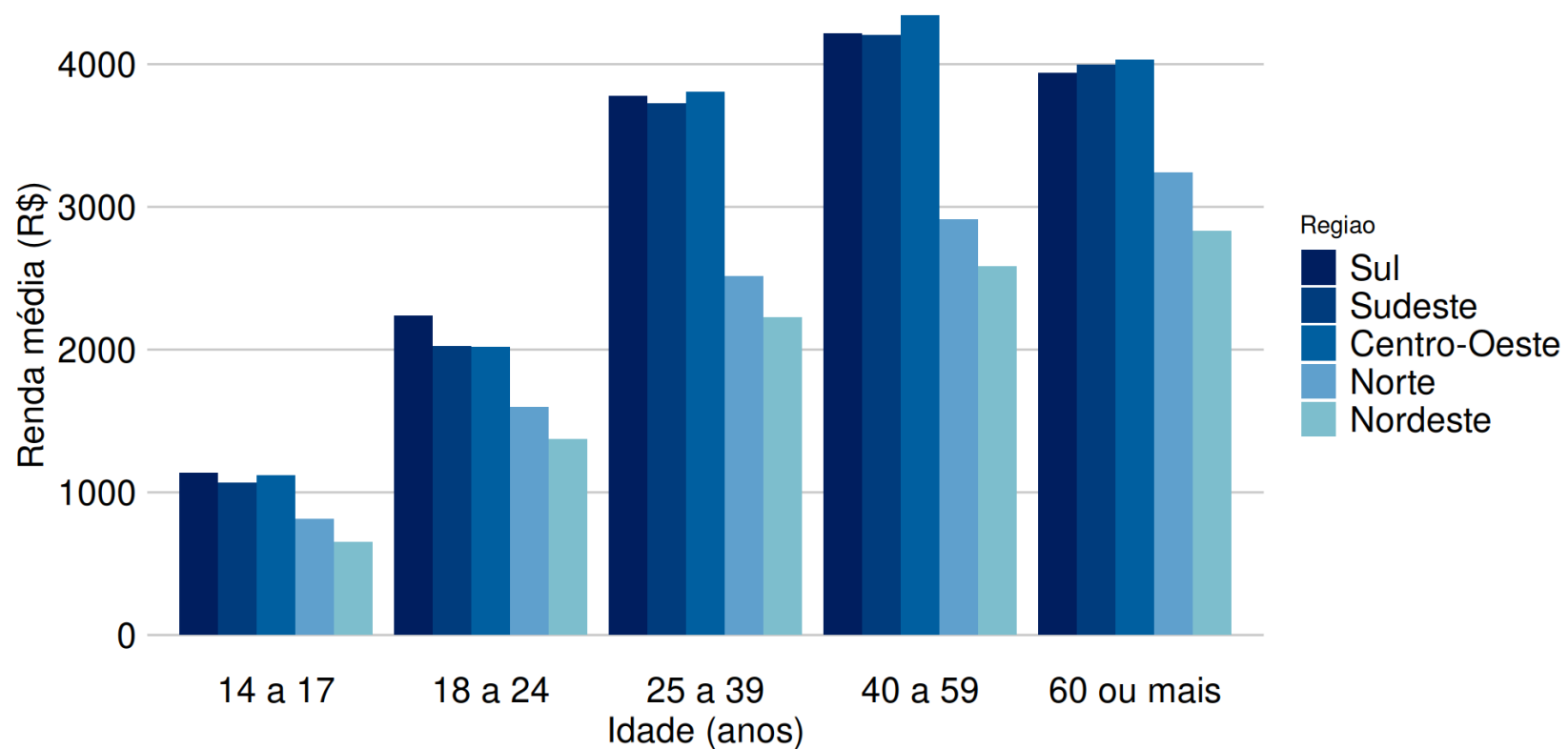


Fonte: IBGE.

Visualização de Contagens

Barras Agrupadas

- A visualização abaixo fornece comparação entre renda média de cada região na mesma faixa de idade.

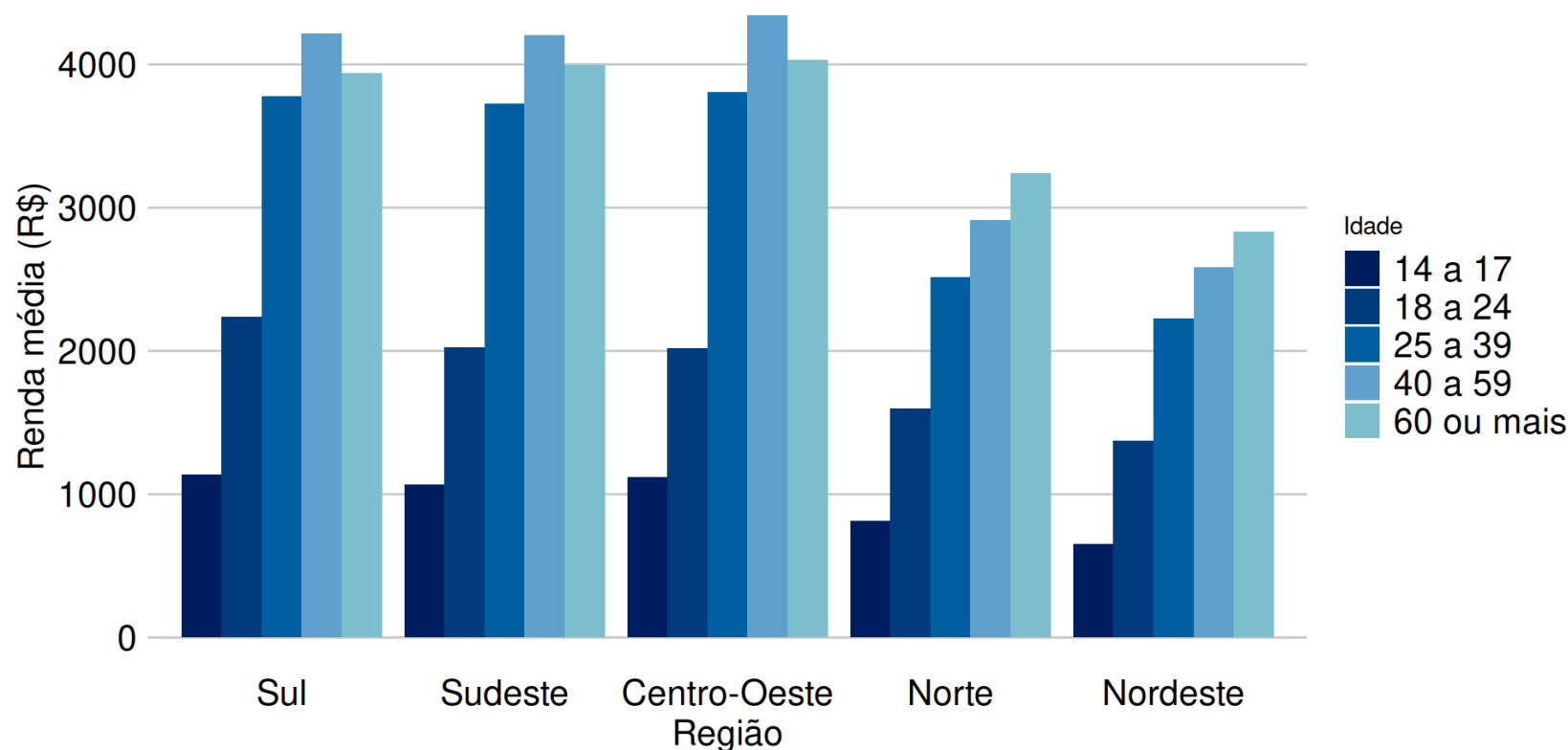


Fonte: IBGE.

Visualização de Contagens

Barras Agrupadas

- Se quisermos comparar a distribuição da renda média ao longo das faixas etárias para cada região, o gráfico abaixo é melhor.

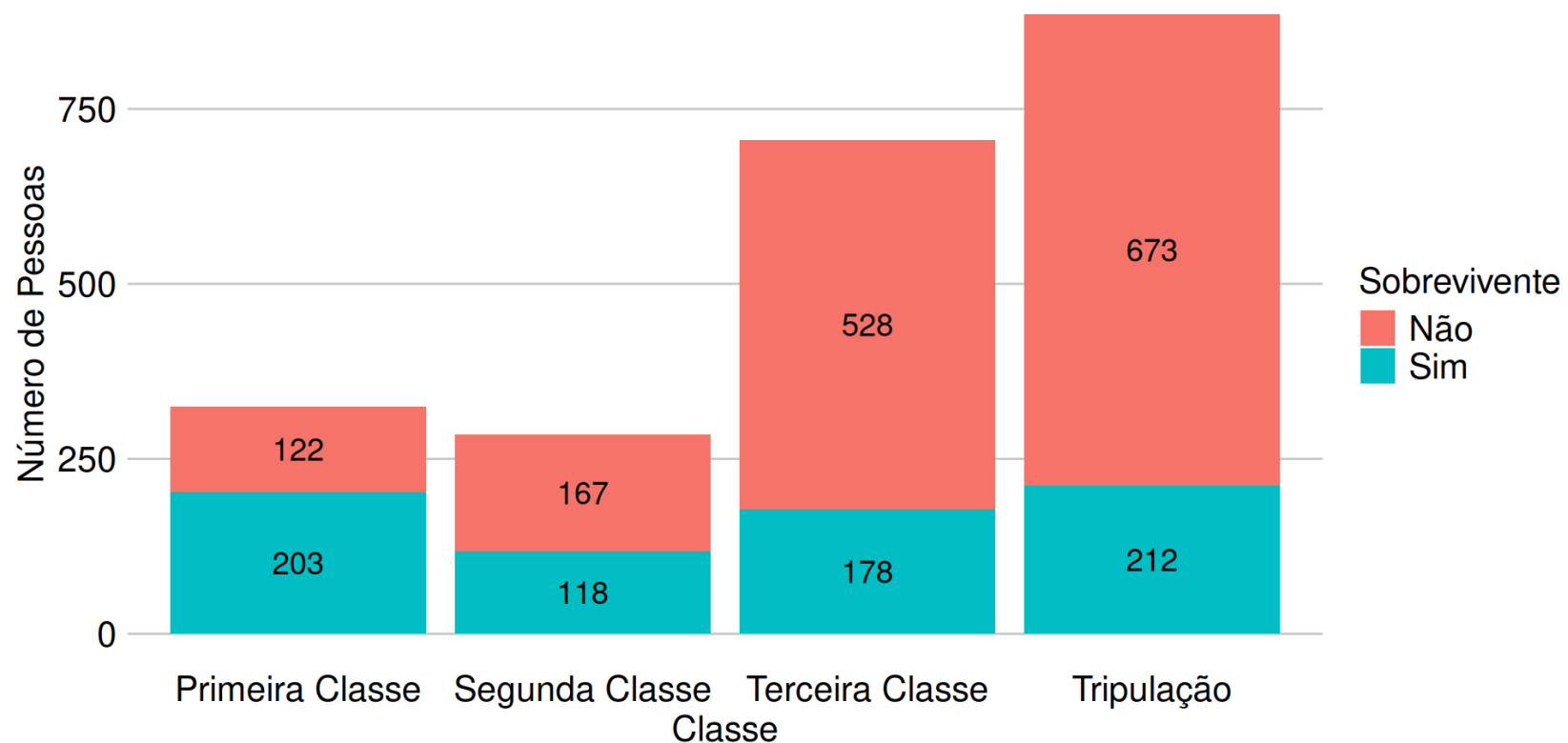


Fonte: IBGE.

Visualização de Contagens

Barras Empilhadas

- Cada segmento corresponde a uma subcategoria, e a altura total da barra representa a soma de todas as subcategorias.



Dados de passageiros e tripulantes do Titanic.

Visualização de Proporções

- Usado para mostrar como um todo se divide em partes que representam proporções individuais. Exemplos:
 - Proporção de gêneros em um grupo.
 - Percentual de votos para partidos.
 - Participação de mercado de empresas.
- **Desafio:** Pode ser complexo quando há muitas partes ou quando as proporções mudam ao longo do tempo.
- **Regra de Ouro:** Sempre escolha a visualização que melhor se adapta ao seu conjunto de dados e destaca as características que você quer mostrar.

Visualização de Proporções

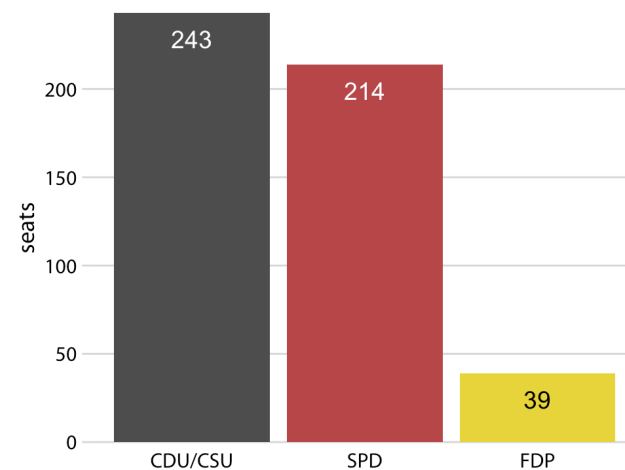
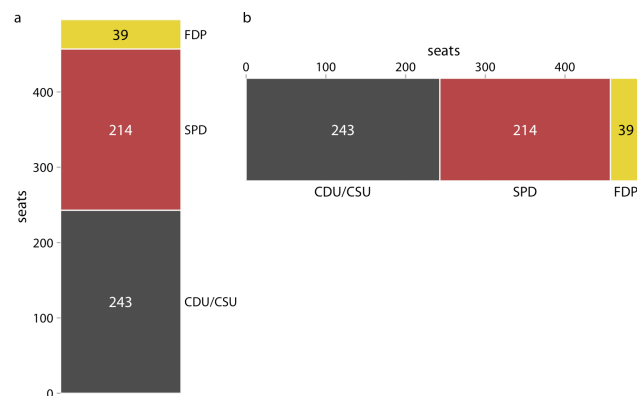
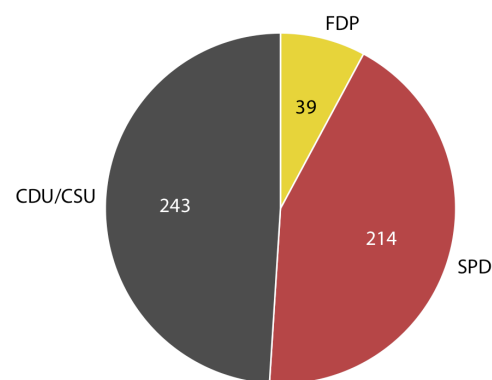
Gráfico de Setores (Pizza)

- Divide um círculo em fatias, onde a área de cada fatia é proporcional à fração do total.
- Quando usar:
 - Para enfatizar frações simples (ex: metade, um terço).
 - Com pequenos conjuntos de dados (poucas categorias).
 - São visualmente claros para mostrar a relação das partes com o todo em um cenário único.
- Muito usado em apresentações, mas criticado por cientistas de dados.
 - Dificulta a comparação precisa entre categorias (especialmente com muitas partes ou valores similares).

Visualização de Proporções

Gráfico de Setores (Pizza)

- Composição partidária do parlamento alemão entre 1976 e 1980.
- Note que o gráfico de barras pode ser usado.



Visualização de Distribuições

Histogramas

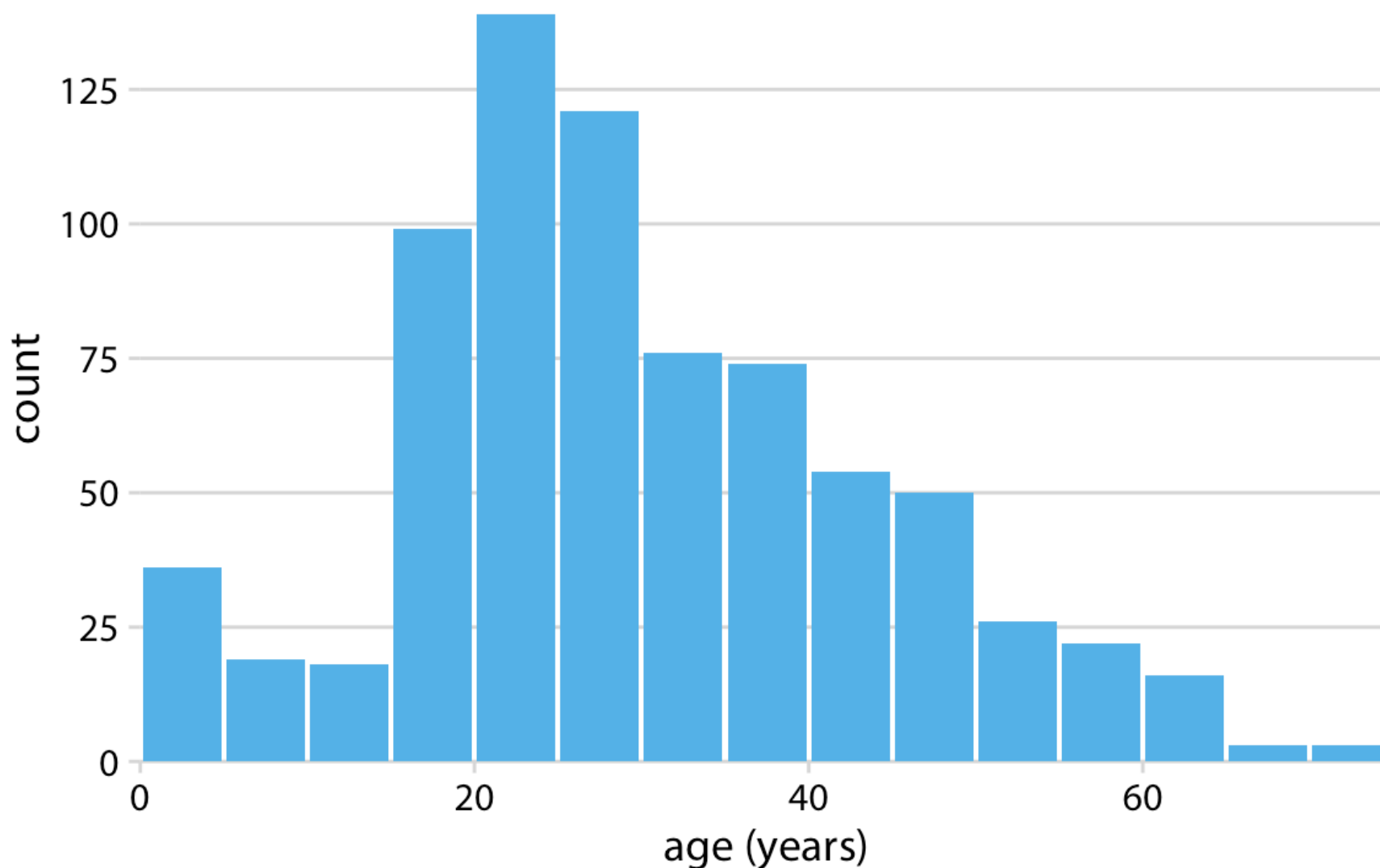
- Quando os dados são quantitativos, o histograma é uma ótima opção para compreender seu comportamento.
- Vamos considerar as idades dos passageiros do Titanic.

Idade	Frequência	Idade	Frequência	Idade	Frequência
0–5	36	31–35	76	61–65	16
6–10	19	36–40	74	66–70	3
11–15	18	41–45	54	71–75	3
16–20	99	46–50	50		
21–25	139	51–55	26		
26–30	121	56–60	22		

Visualização de Distribuições

Histogramas

- Podemos visualizar os dados por meio de um histograma.



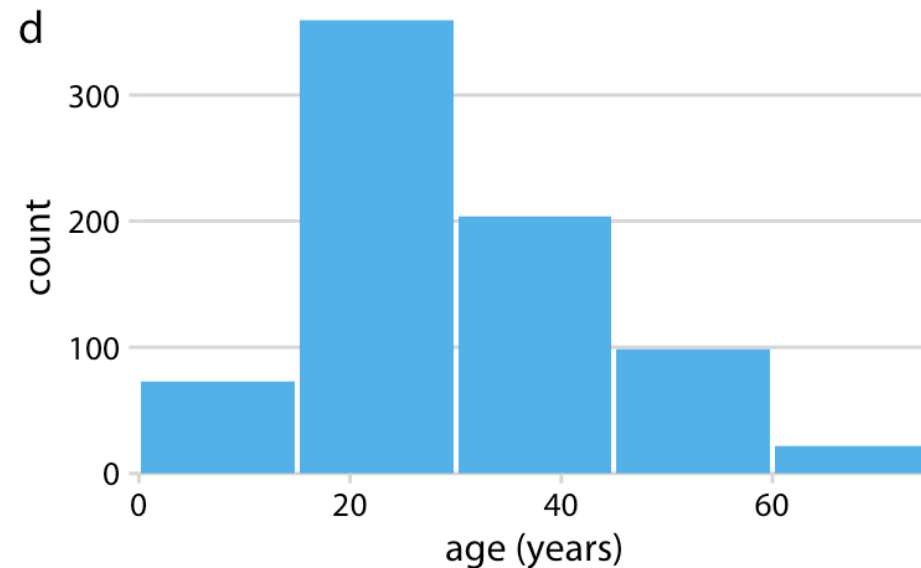
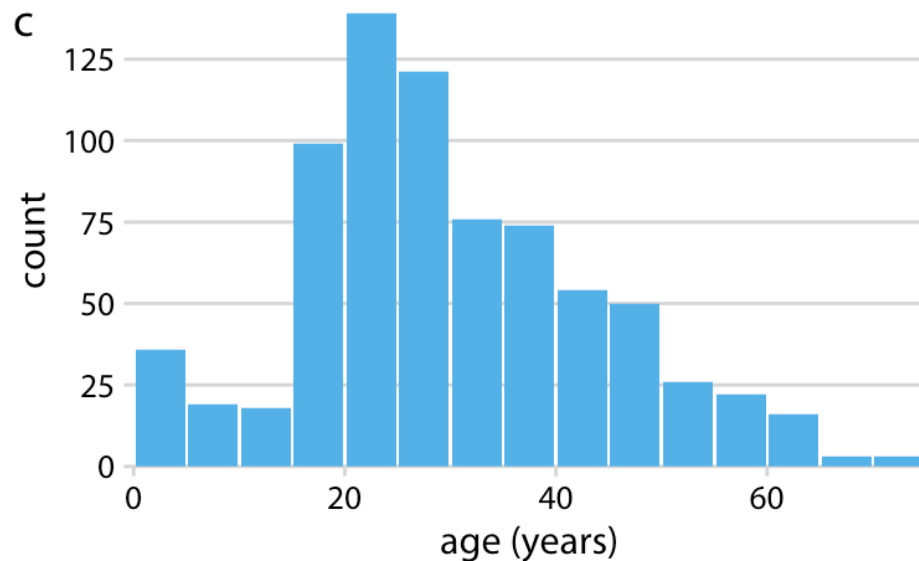
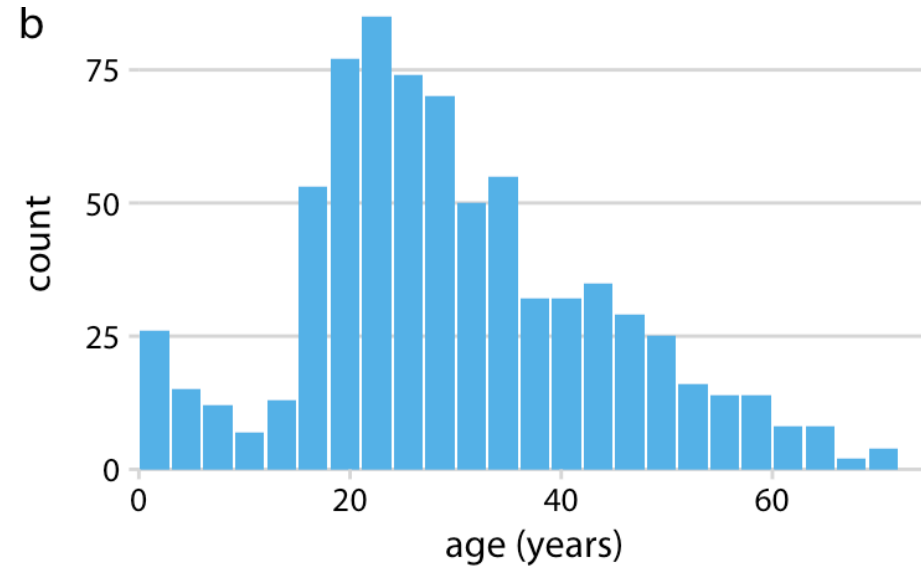
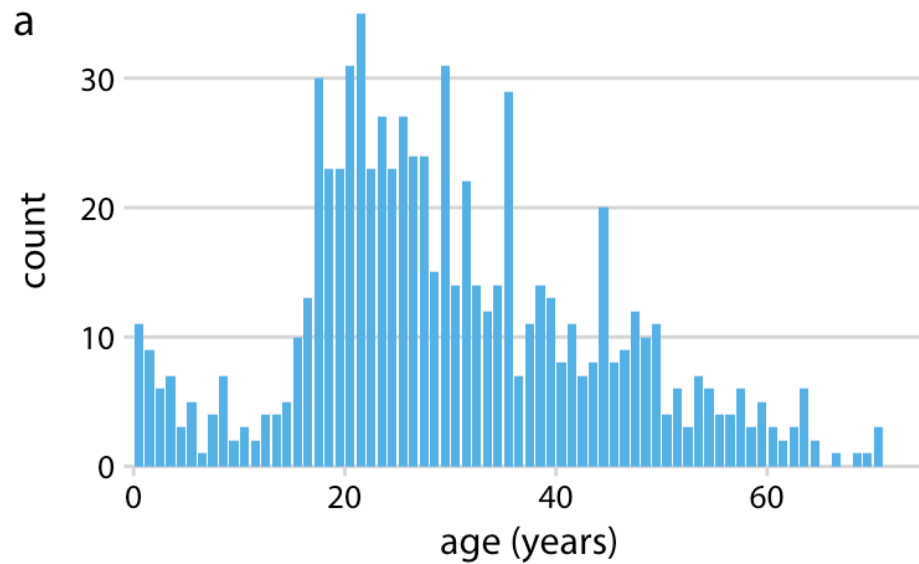
Visualização de Distribuições

Histogramas

- A aparência visual exata de um histograma depende do número de intervalos usados para sua construção.
 - **Largura do intervalo muito pequena:** O histograma se torna excessivamente “pontudo” e visualmente poluído, e as tendências principais nos dados podem ficar escondidas.
 - **Largura de bin muito grande:** Características menores na distribuição dos dados, como a queda em torno dos 10 anos de idade, podem desaparecer.

Visualização de Distribuições

Histogramas



Visualização de Distribuições

Histogramas

- Uma regra útil para determinar o número de intervalos é usar a **Regra de Sturges**. Segundo ela, o número de intervalos (k) deve ser:

$$k = 1 + \log_2(n)$$

ou

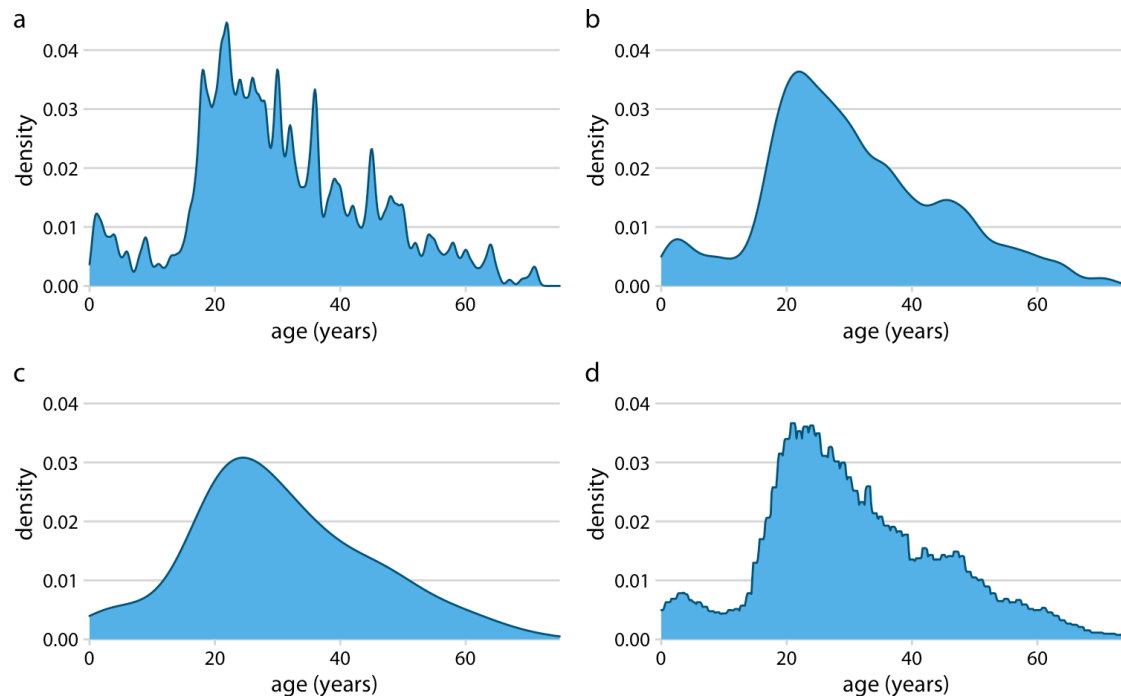
$$k = 1 + 3.322 \times \log_{10}(n),$$

em que n é o tamanho da amostra.

Visualização de Distribuições

Gráfico de densidade

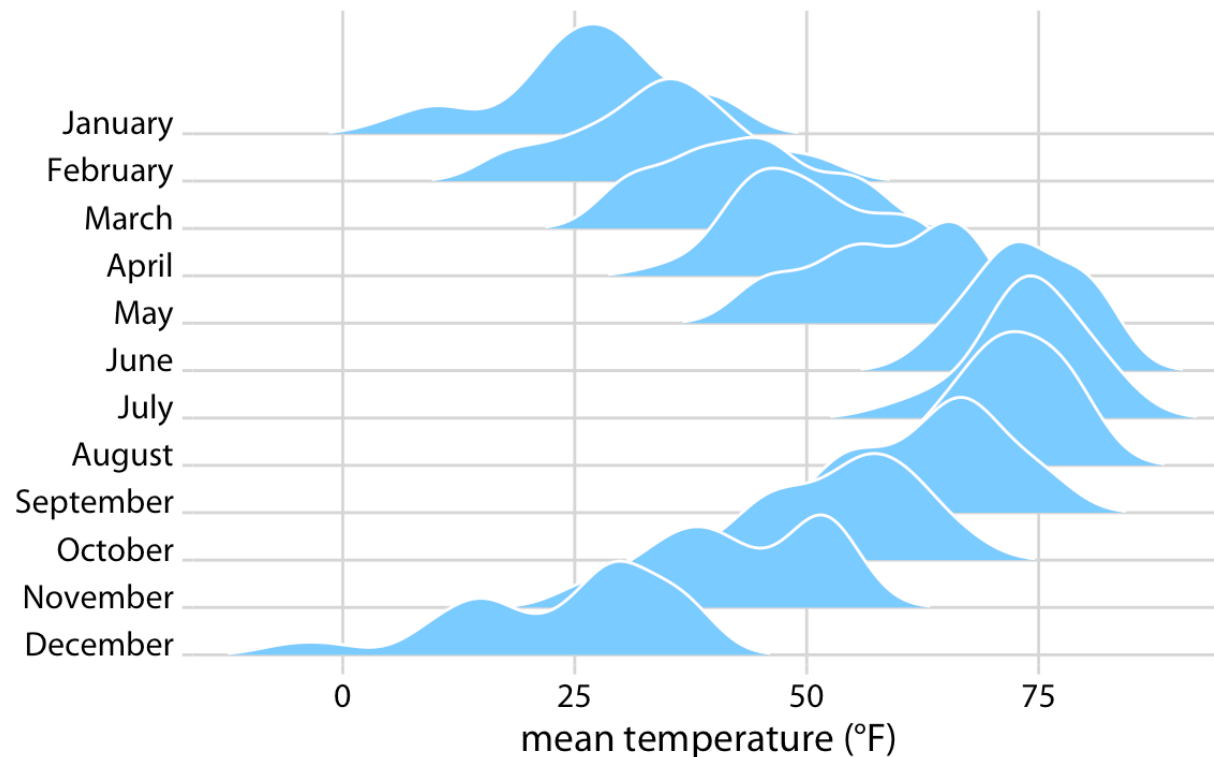
- Outra forma de visualizar a distribuição de dados quantitativos é usando um gráfico de densidade.
- A curva de representação dos dados é obtida via estimação de densidade kernel.



Visualização de Distribuições

Gráfico de densidade

- O gráfico de densidade também pode ser usado para visualizar o comportamento de vários grupos, geralmente ao mesmo tempo.

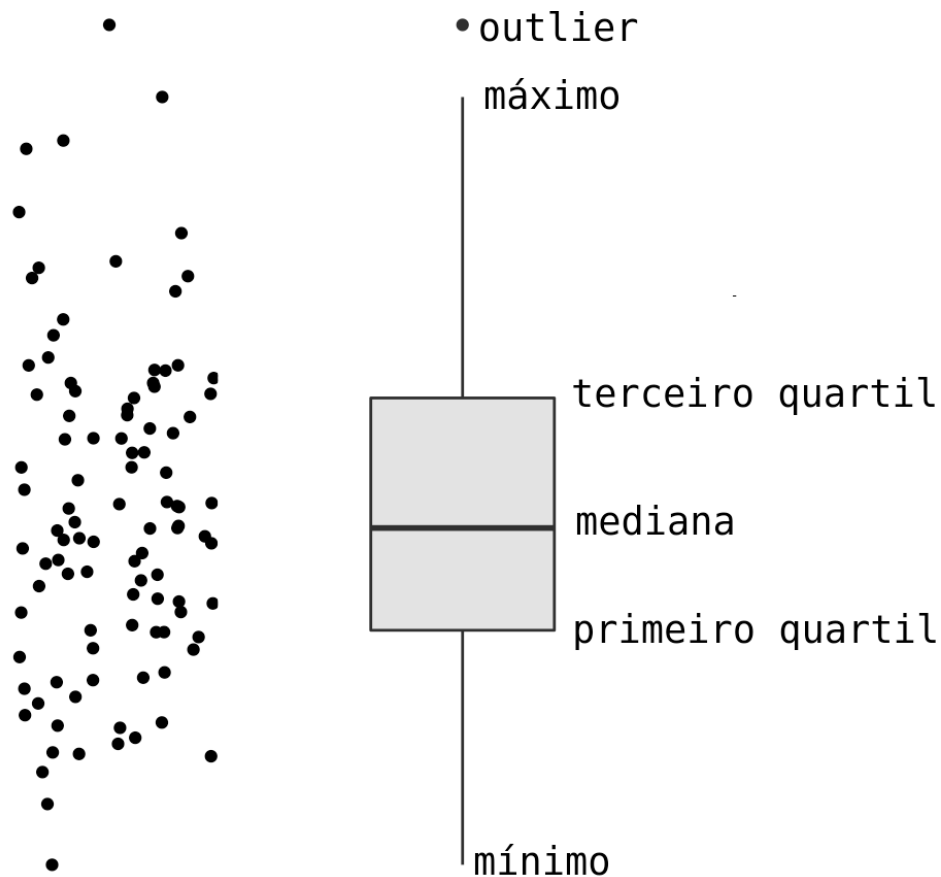


Temperatura diária média ao longo dos meses em Lincoln, Nebraska.

Visualização de Distribuições

Boxplot

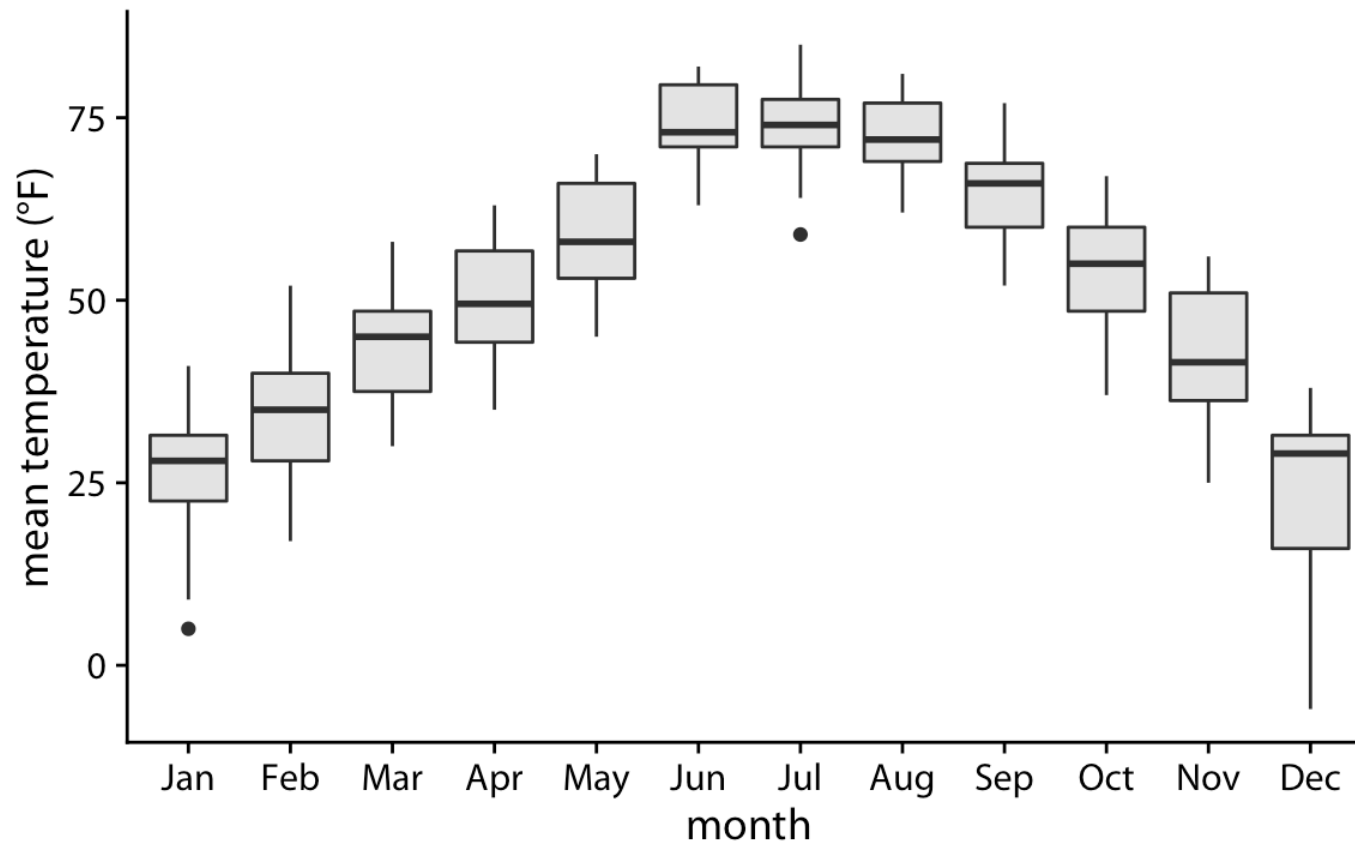
- Boxplot, ou diagrama de caixa, é um gráfico que divide os dados em quartis
- Ele exibe cinco estatísticas principais:



Visualização de Distribuições

Boxplot

- O boxplot é muito útil para comparar a distribuição de grupos diferentes.

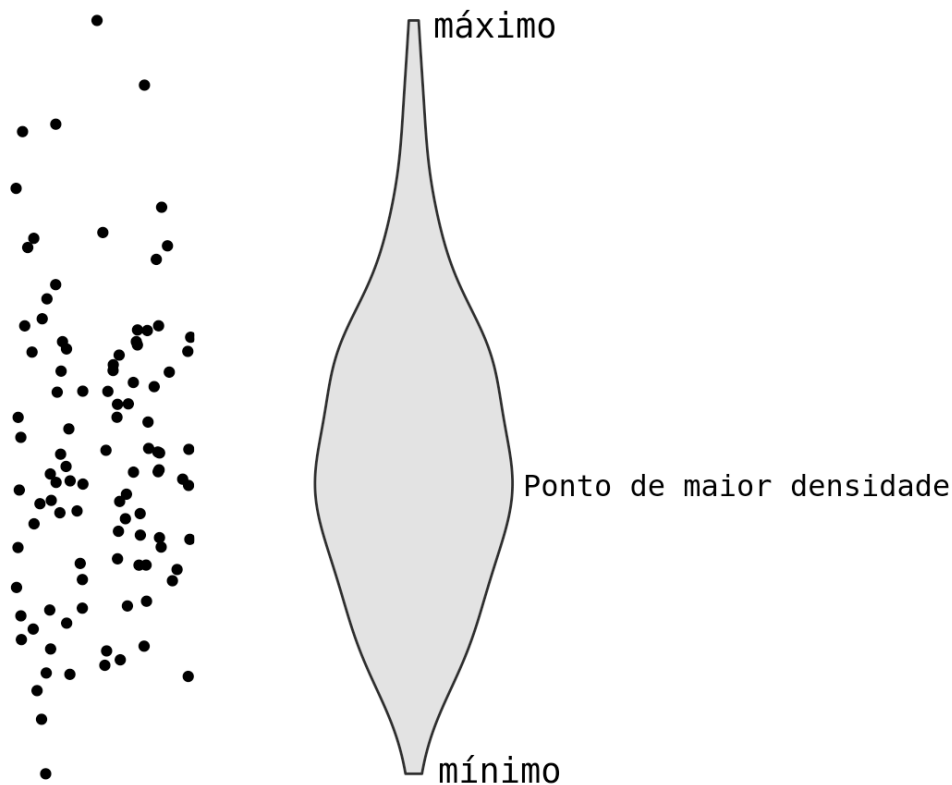


Temperatura diária média ao longo dos meses em Lincoln, Nebraska

Visualização de Distribuições

Gráfico de Violino

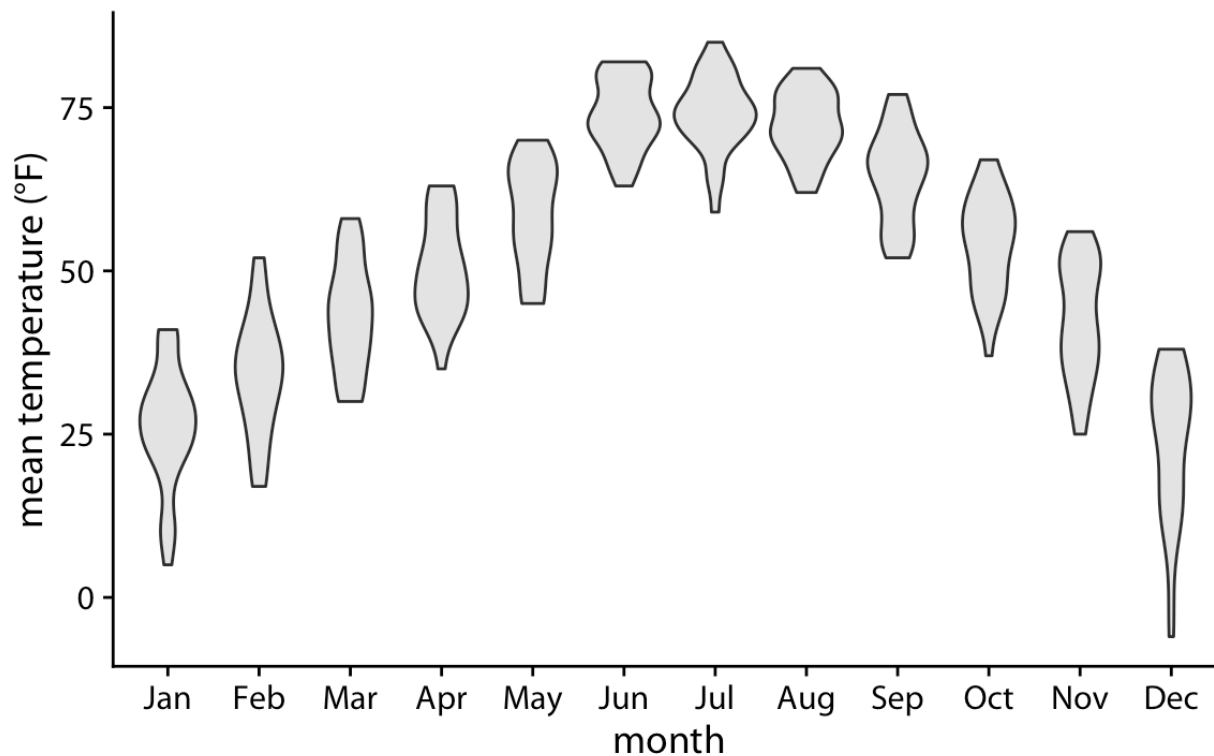
- Combina um boxplot com uma estimativa de densidade de kernel.
- Quanto mais largo o violino em um determinado ponto, maior a concentração de dados naquele valor.



Visualização de Distribuições

Gráfico de Violino

- Ele também é útil para visualizar o comportamento de uma variável e diferentes grupos.

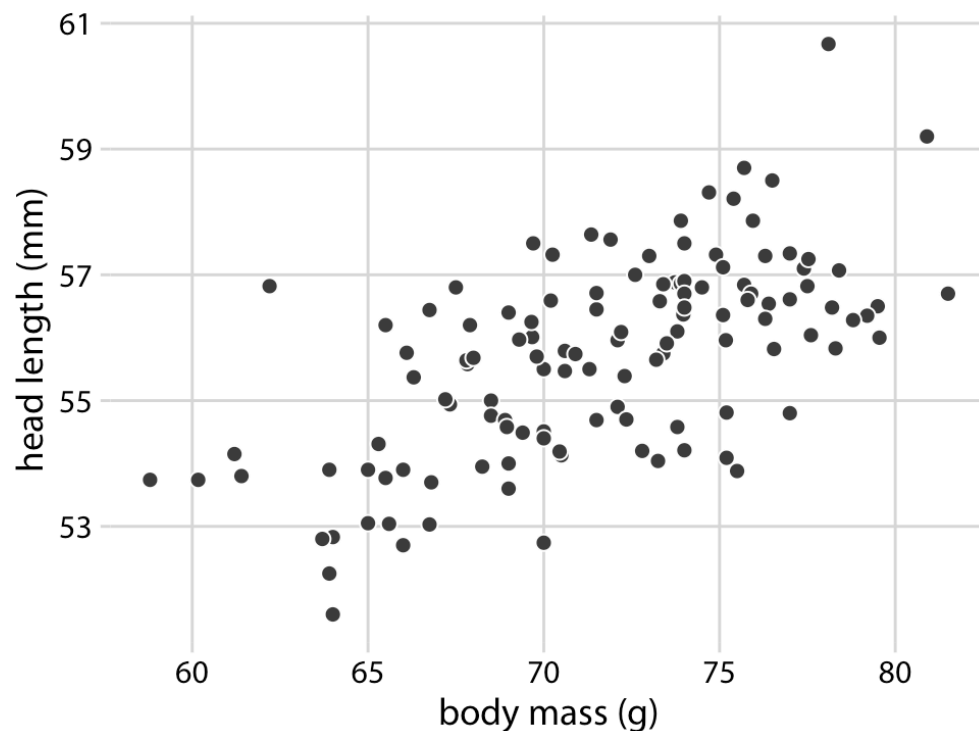


Temperatura diária média ao longo dos meses em Lincoln, Nebraska.

Visualização de Associações

Gráfico de dispersão

- Serve para visualização de duas variáveis ao mesmo tempo.
- Auxilia a compreensão da relação entre variáveis.

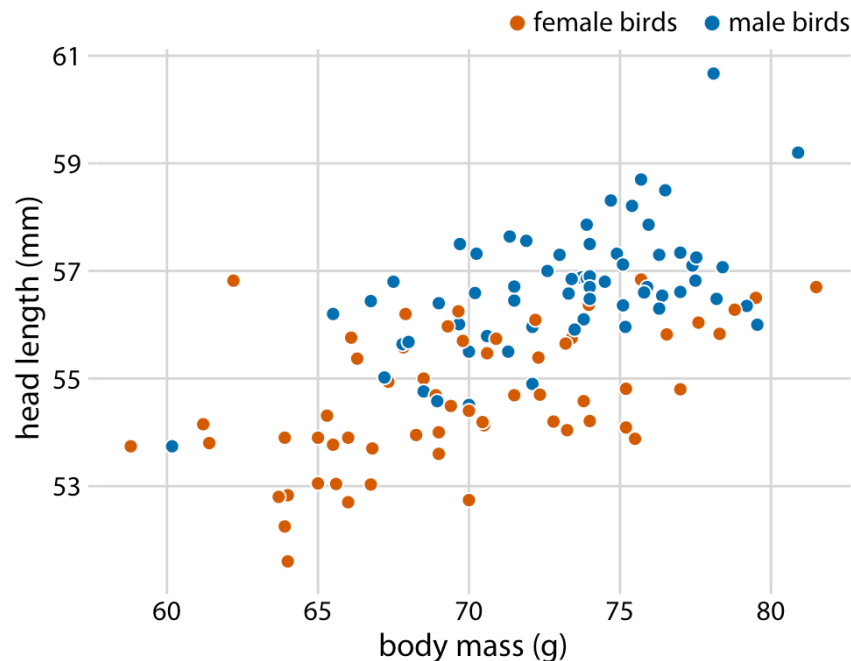


Tamanho da cabeça de uma espécie de pássaro e sua massa corporal.

Visualização de Associações

Gráfico de dispersão

- É possível adicionar uma terceira variável para ajudar numa melhor compreensão dos dados.

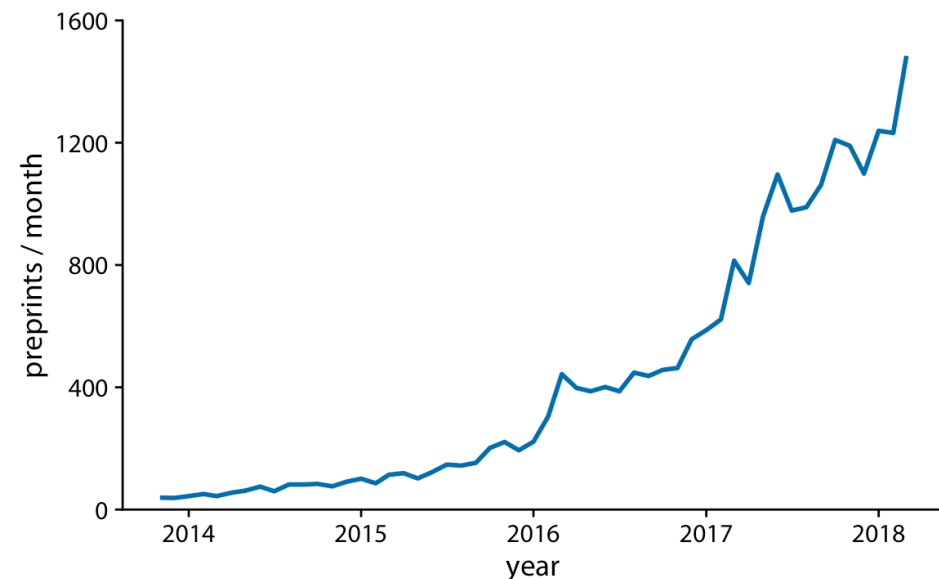
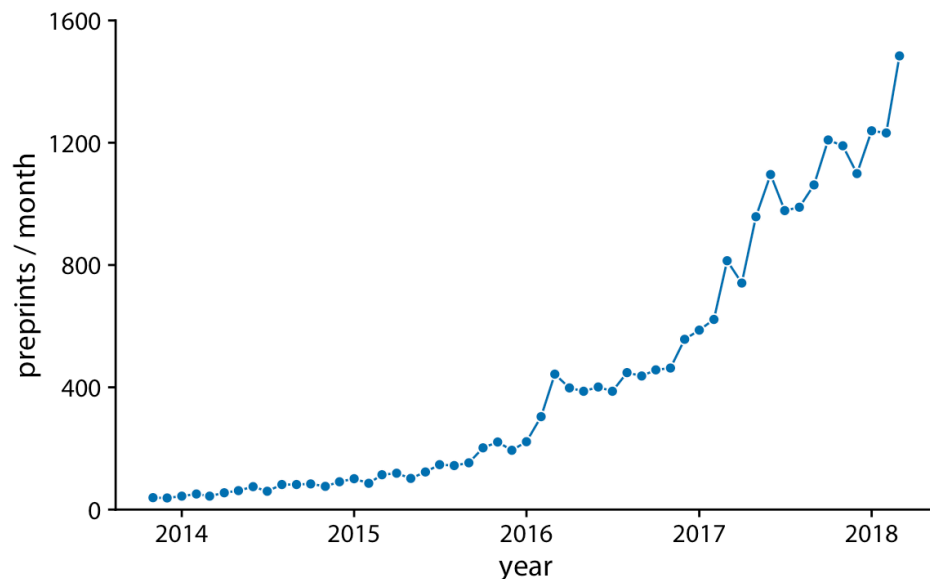


Tamanho da cabeça de uma espécie de pássaro e sua massa corporal segundo sexo.

Visualização de Dados no Tempo

Gráfico de Linha para Série temporal

- Mais comum e eficaz para mostrar como uma ou mais variáveis numéricas evoluem ao longo do tempo.



Submissões mensais ao servidor de pré-publicações bioRxiv.

Visualização de Dados no Tempo

Gráfico de Linha para Série temporal

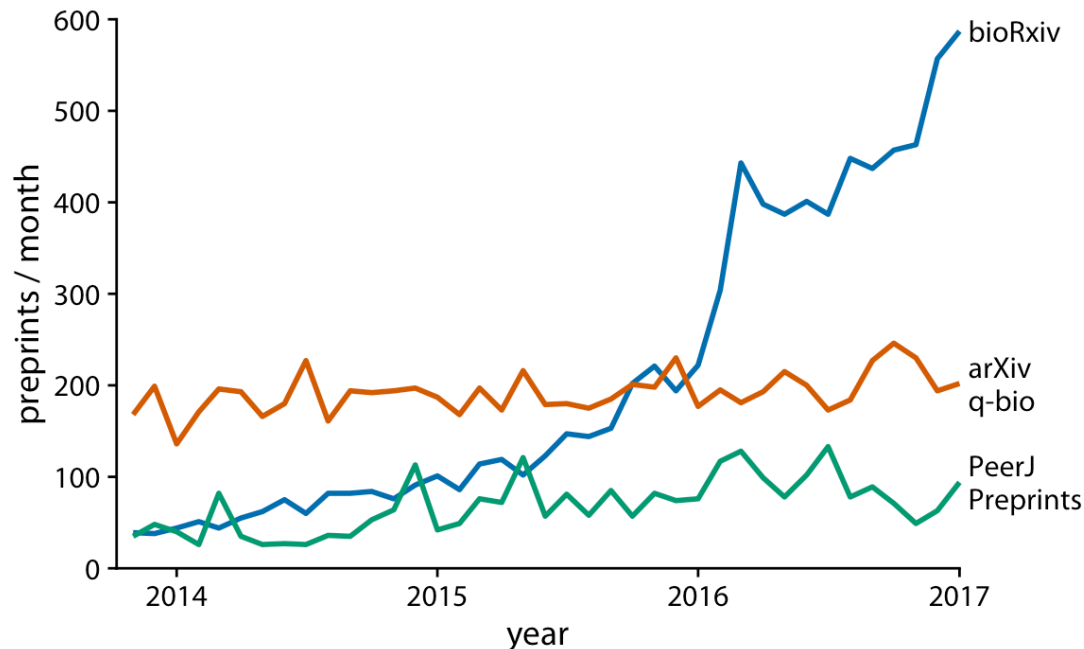
Auxilida na visualização de:

- **Tendências:** Permitem visualizar facilmente a tendência geral dos dados ao longo do tempo (crescimento, declínio, estabilidade).
- **Padrões:** Ajudam a identificar padrões recorrentes (sazonalidade, ciclos).
- **Flutuações:** Mostram variações e ruídos nos dados.
- **Pontos de Inflexão:** Destacam momentos de mudança significativa ou eventos importantes.
- **Comparação:** Quando múltiplas séries temporais são plotadas no mesmo gráfico (com linhas diferentes ou cores diferentes), eles são excelentes para comparar o comportamento de diferentes variáveis ou grupos ao longo do mesmo período.

Visualização de Dados no Tempo

Visualização Múltipla

- O gráfico de linha permite a visualização múltipla para facilitar comparações.



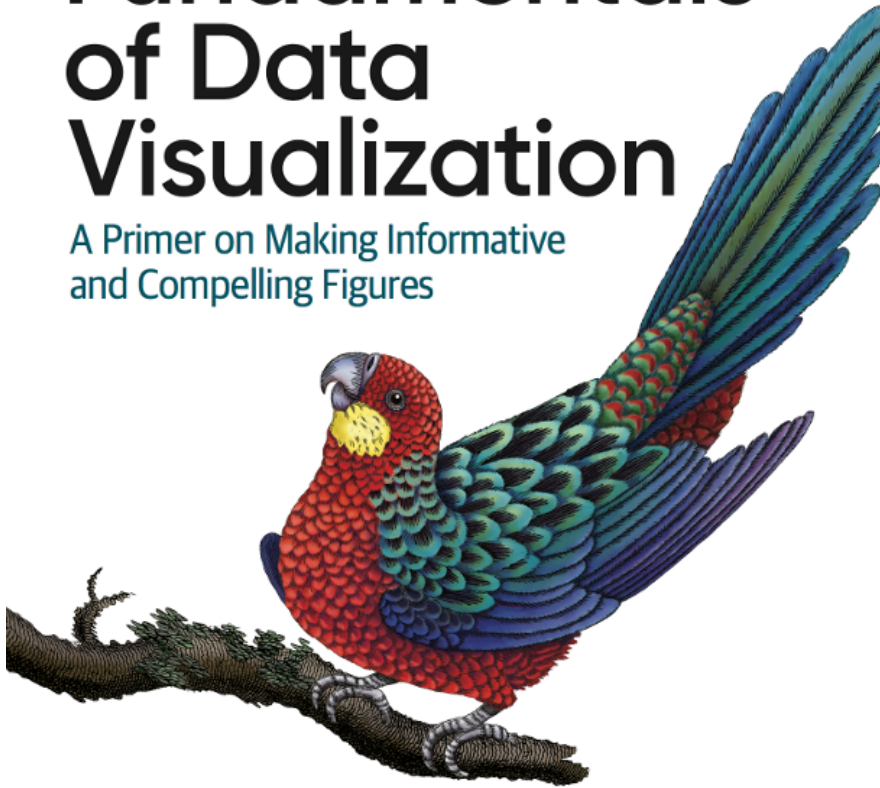
Submissões mensais a três servidores de pré-publicações cobrindo pesquisa biomédica.

Para saber mais

O'REILLY®

Fundamentals of Data Visualization

A Primer on Making Informative
and Compelling Figures



Claus O. Wilke

<https://clauswilke.com/dataviz/visualizing-trends.html>

Fim