Data Analysis and Visualization: Part 01

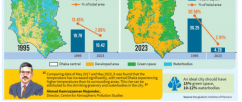
Unit 5 : Session 01

By Md. Meftaul Haque Mishu

An newspaper article

Heading: With greenery wiped out, the city turns into a heat island

Source



Infographic: TBS

Amid record high temperatures making the lives of Dhaka-dwellers miserable, the city corporations are felling trees in several places including Dhanmondi, further shrinking down greeneries to an alarming level.

Ideally, a city should have at least 15% green space and 10-12% wetland but a recent study conducted by the Bangladesh Institute of Planners (BIP) found that in the last 28 years, green space in Central Dhaka has shrunk to 7.09% while wetlands occupy as low as 2.9% of the city.

At present, central Dhaka has only 10.42 sq km of green space which was around 19.78 sq km in 1995. On the other hand, wetland areas have decreased significantly to stand at only 2.28 sq km compared to 30.24 sq km in 1995, the study finds.

The findings of the study titled "38 Years of Reservoir and greenery destruction in The Capital: Reality and Salvation Roadmap" was revealed at a roundtable, jointly organised by the BIP and Urban Development Journalists Forum Bangladesh, at the BIP auditorium in the capital on Saturday.

The event was organised on the occasion of World Environment Day.

BIP survey area is 146.98 sq km, which is the area of the old Dhaka city corporation (central Dhaka) encompassing Gabtoli, Rupnagar, Mirpur Cantonment, Airport, Kuril, Gulshan, Dhanmondi, Motijheel, Old Dhaka, Farmgate, Shampur, Hazaribag, and others.

BIP General Secretary Sheikh Muhammad Mehdi Ahsan presented the keynote speech on the occasion.

Syeda Rizwana Hasan, chief executive of the Bangladesh Environmental Lawyers Association (Bela), Sultana Kamal, president of Bangladesh Poribesh Andolon (Bapa) and LGRD Minister Tazul Islam were also present at the event.

Recently, Dhaka South fell trees on Sat Masjid Road in Dhanmondi. The cutting down of trees began in January 2023 as part of a developmental project of Dhaka South City despite protests by locals and environmentalists.

At the roundtable, BIP General Secretary Sheikh Muhammad Mehdi Ahsan said, "The city authorities fell trees and occupied the water bodies in the name of development. There is a huge lack of transparency in this kind of project."

Lack of awareness, weakness of legal framework and misuse of personal interests and power are the significant factors behind the destruction of green spaces and filling of reservoirs in the city, he said.

The BIP general secretary also added that Dhaka's two city corporations have taken some initiatives, including the restoration of canals, which deserve praise.

According to the BIP study report, attempts to set up new clubs, markets, and bazaars are eating up the space previously designated as parks and playgrounds. Besides, the concrete-covered area has increased in accordance with the latest development plans chalked out for parks.

Meanwhile, development works in Dhaka frequently result in erecting residential or commercial buildings by filling the wetlands.

The BIP study also noted that the existing land ownership in Dhaka is mostly private and despite approved guidelines for the conservation of specific land areas, it is not being followed here.

The study shows, in Dhaka city, there should be a two-acre playground and a one-acre park for 12,500 people but only a one-acre playground or park is available for every 37,900 people.

Dhaka North has a 251.7-acre park and playground while Dhaka South has a 340-acre park and playground. There is no playground in the 41 wards out of 129 wards of the two city corporations.

At the roundtable, Syeda Rizwana Hasan, chief executive of Bangladesh Environmental Lawyers Association (Bela), said Dhaka does not have the capacity to hold such a population as not only living space is enough for a sustainable lifestyle of a human being but wetlands and greenery are also needed.

"Rajuk is filling up water bodies, excavating forests and making plots in Purbachal. I think the Rajuk board should be reformed because everyone here is a bureaucrat," she added.

Sultana Kamal, president of Bangladesh Poribesh Andolon (Bapa), said, "Occupying the green spaces and water bodies is not only a technical, social, or cultural move but also political. Corrupt and wicked people are occupying these assets, making Dhaka as they wish. They do not want green spaces and parks because they and their children are more comfortable living abroad."

LGRD Minister Tazul Islam said, "There is also a need to have a certain limit on how many people can live in Dhaka city to protect the green spaces and wetlands. Any good plan and civic amenities are bound to collapse due to the pressure of too many people."

Dhaka turning into a heat island

Dhaka residents are having their hottest days in recent memories. The temperature, which crossed 40 degrees Celsius in April, was the highest in 58 years. But it feels 5-6 degrees higher than the actual reading.

According to field surveys of the Centre for Atmospheric Pollution Studies (Caps), summertime temperatures in Dhaka were 3-4 degrees Celsius higher in heavily urbanised areas than in areas with more green coverage and water bodies.

Caps Director Dr Ahmad Kamruzzaman Majumder told The Business Standard, "Comparing data of May 2017 and May 2023, it was found that the temperature has increased significantly, with central Dhaka experiencing higher temperatures than its surrounding areas. This rise can be attributed to the shrinking greenery and waterbodies in the city."

"Dhaka's air quality and temperature are particularly related to the existing greenery and water reservoirs. Roads and buildings (concrete areas) act as heat storage and increase the surrounding temperatures. As a result, Dhaka is turning into a heat island day by day. We need to plant a lot of trees and there should be some guidelines for the use of colours on buildings and roads," Kamruzzaman added.

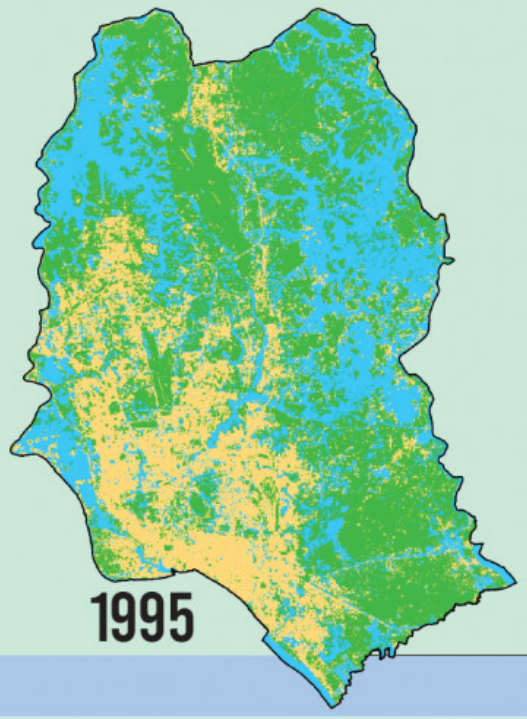
Among the cities of neighbouring countries, Dhaka City is placed at the bottom for greenery space and livable cities.

In Mumbai city, the ratio of green spaces to total area fell from 46.7% in 1988 to 26.67% in 2018. According to 2017 data, the green cover area of Singapore is 47%, which increased from 36% in the 1980s.

Professor Adil Mohammed Khan, the executive director of the Institute for Planning and Development (IPD), told TBS, "Our policymakers must find a way out of the current situation of Dhaka's greenery and water bodies. Trees should be planted wherever possible and water bodies should be conserved."

CENTRAL DHAKA

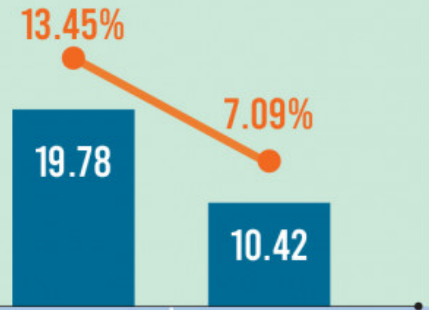
Total area: **146.98 sq km**

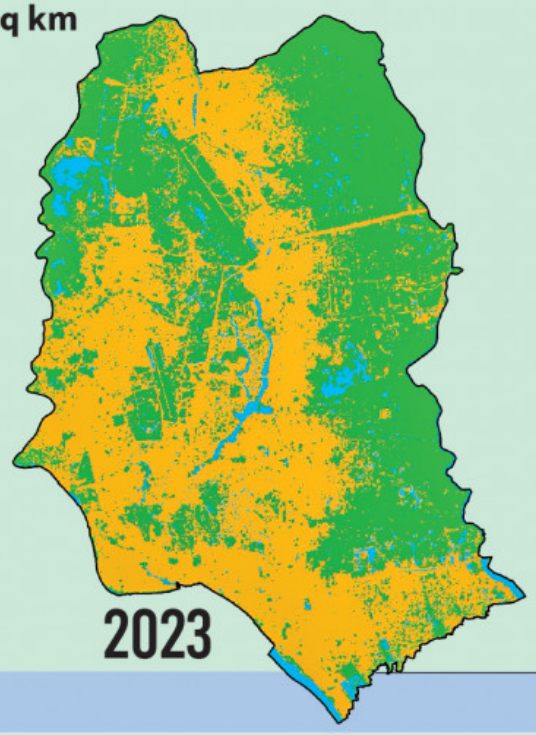


Green space

Area in sq km

% of total area

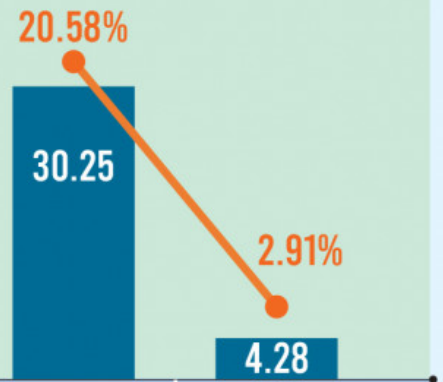




Waterbody

Area in sq km

% of total area



Dhaka central | Developed area | Green space | Waterbodies

"Comparing data of May 2017 and May 2023, it was found that the temperature has increased significantly, with central Dhaka experiencing higher temperatures than its surrounding areas. This rise can be attributed to the shrinking greenery and waterbodies in the city."

Ahmad Kamruzzaman Majumder,
Director, Centre for Atmospheric Pollution Studies

An ideal city should have **15%** green space, **10-12%** waterbodies

Source: Bangladesh Institute of Planners

Whats is Data Visualization?

- Visual / Graphical representation of data
- Knowledge is compressed into a picture

Why visualizing data ?

- To make sense of huge amount of data
- To communicate information clearly and efficiently
- To understand patterns, trends and outliers in data
- To identify relationships between data points

Disadvantages of Data Visualization

- Misinterpretation of data
- Biased or incorrect conclusions
- Correlation does not imply causation
- Core message can be lost in the visualization

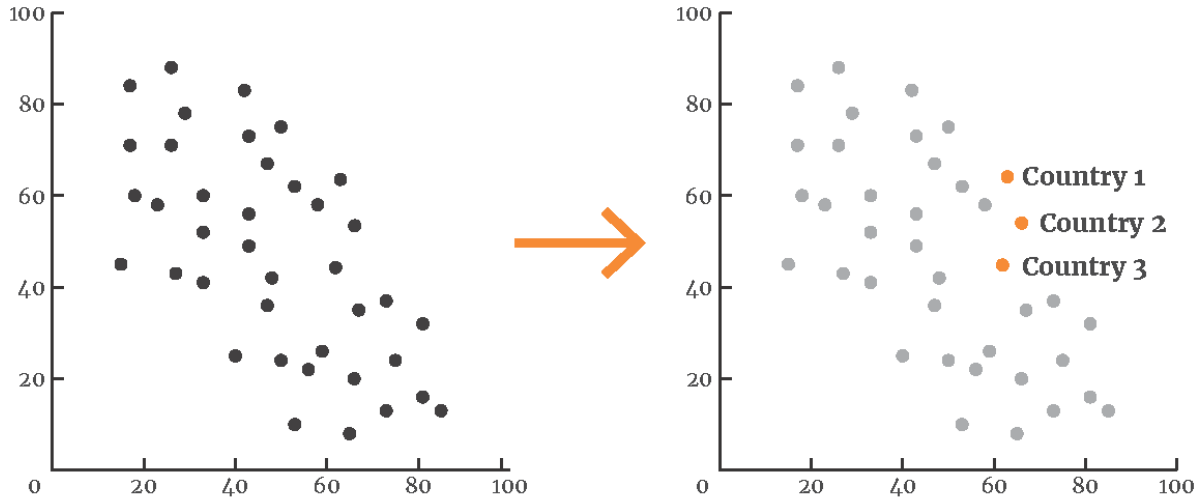
Core Principles of Data Visualization

Know your audience

- Scientific Journal or Newspaper Article
- Research Report or Blog
- Written or Digital Media

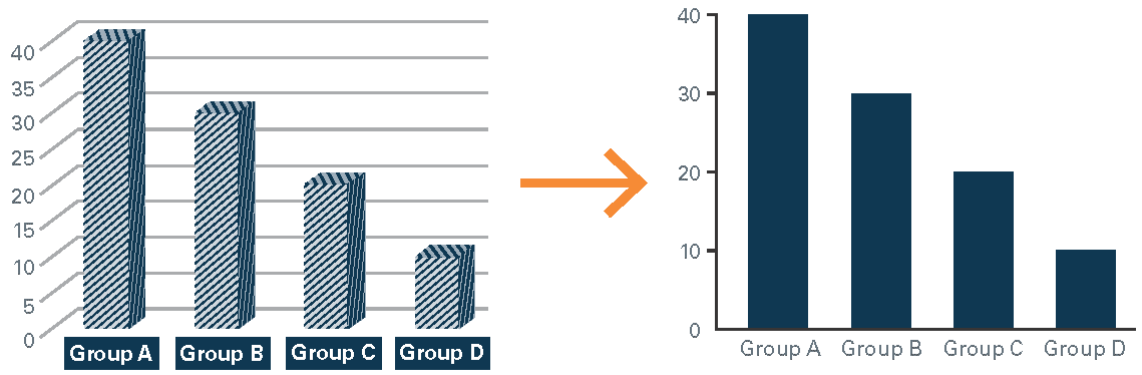
Show the data

- Highlight the what is important



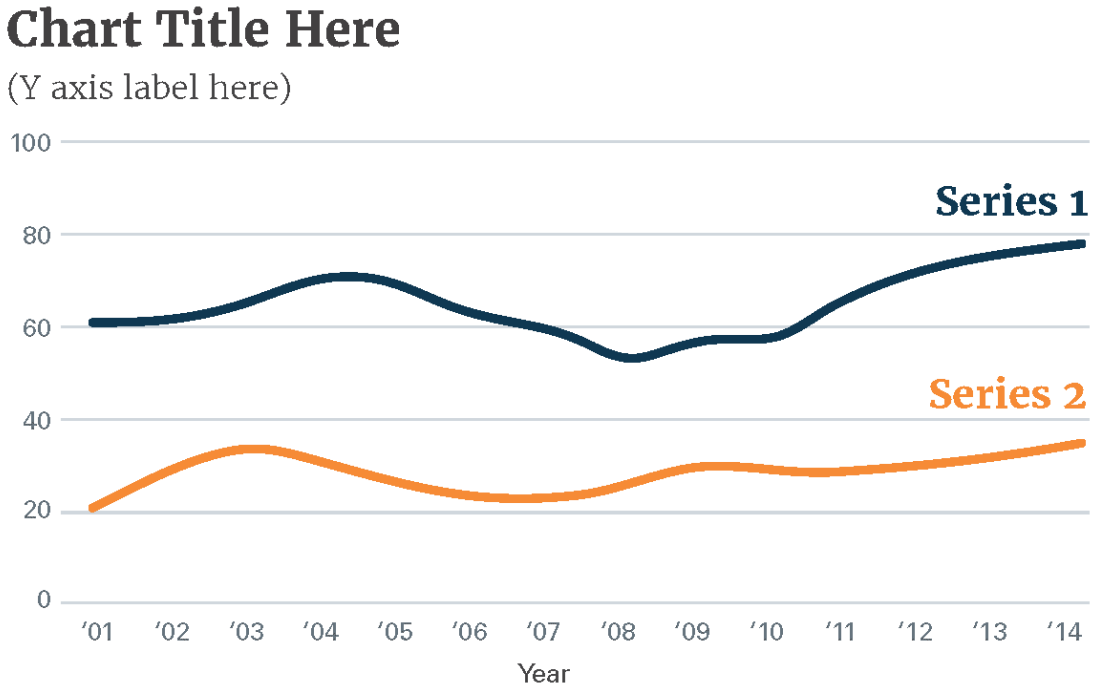
Reduce the clutter

- Remove unnecessary elements
- Avoid 3D, shadows, and other effects



Use text wisely

- Annotate properly
- Guide the reader
- Make labels readable

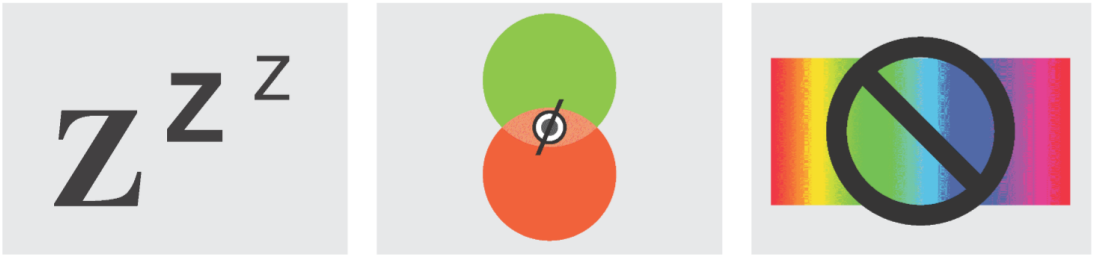


Try small multiples

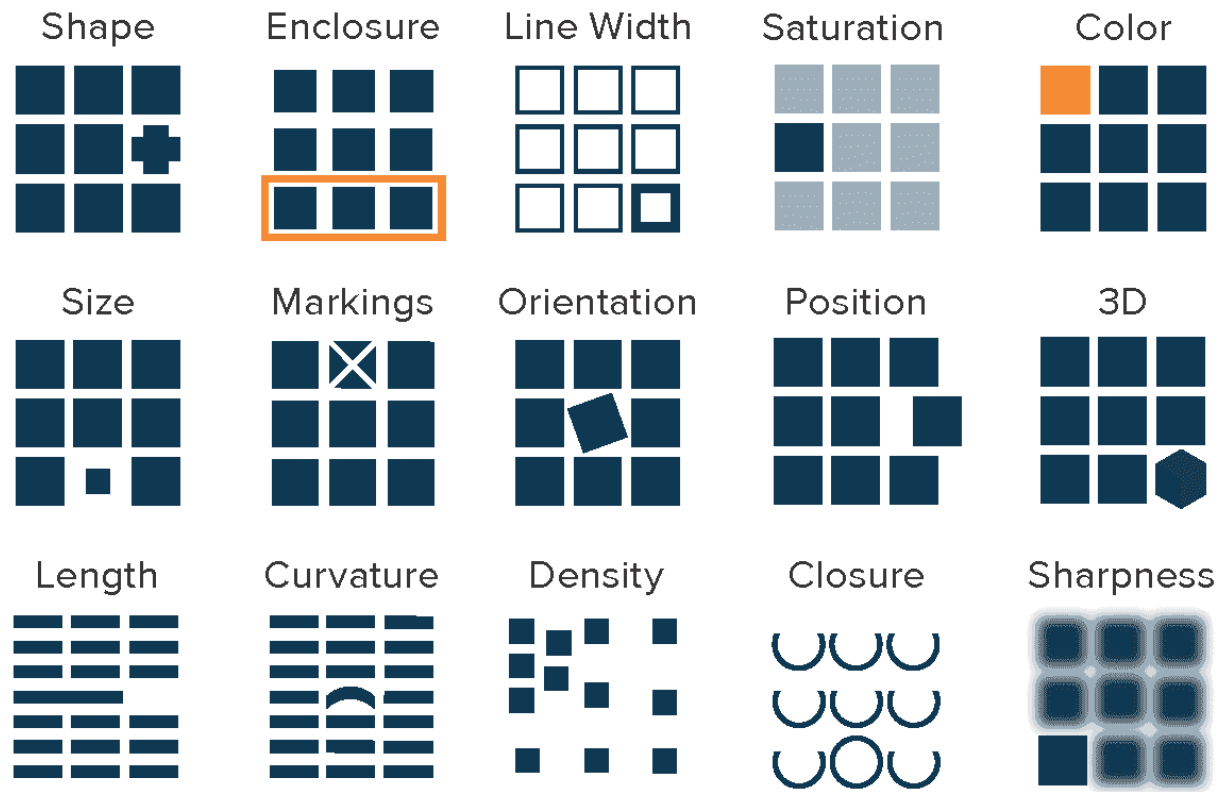
- Break up complicated charts into smaller chunks

Color and font considerations

- Avoid default colors and fonts
- Consider color blindness
- Avoid the rainbow color pallette



Preattentive Processing



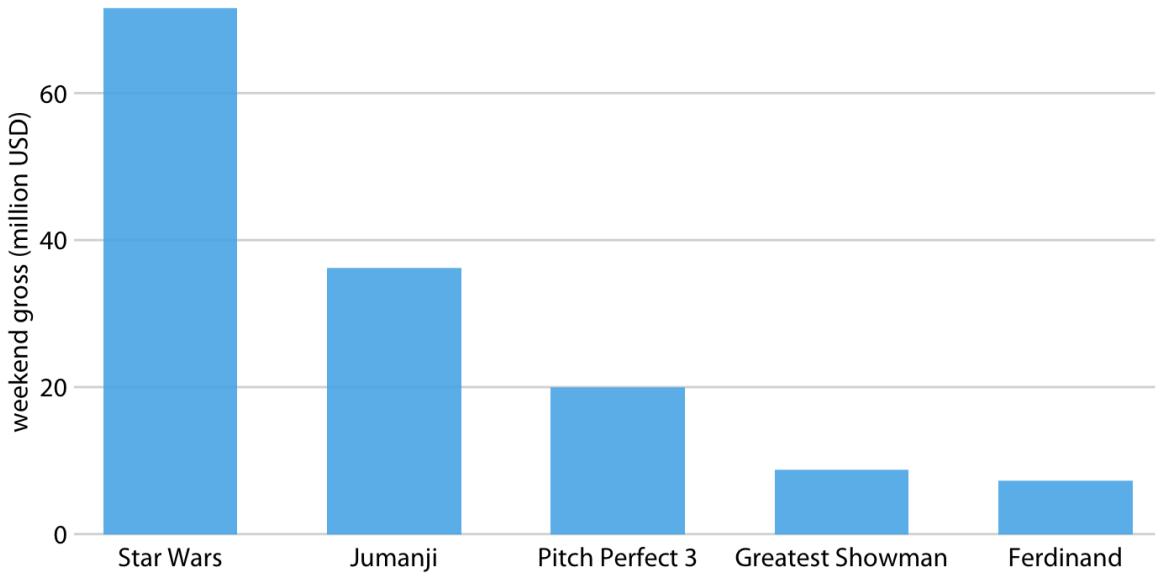
Directory of visualizations

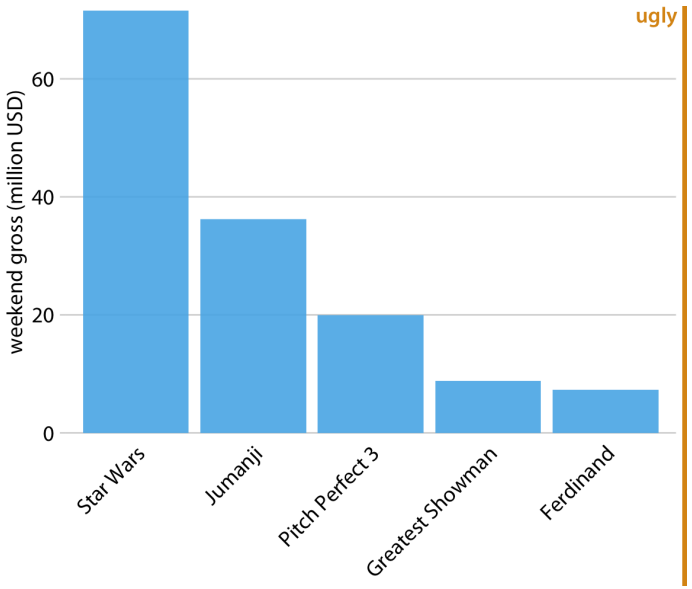
- Amount Visualization
- Visualizing Distributions
- Visualizing Proportions
- $x - y$ Relationships
- Geospatial Data
- Uncertainty Visualization

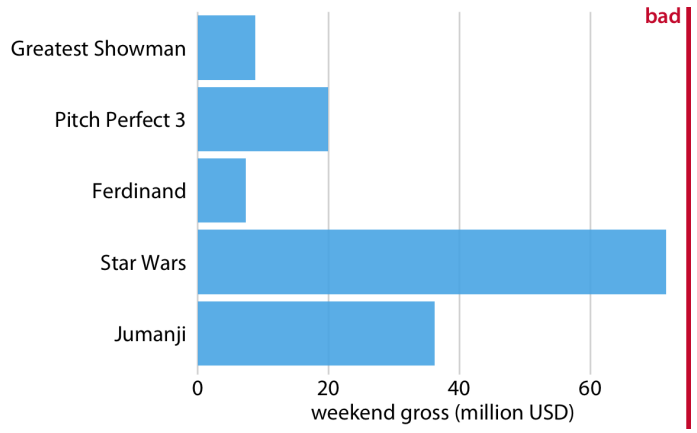
Visualizing Amounts

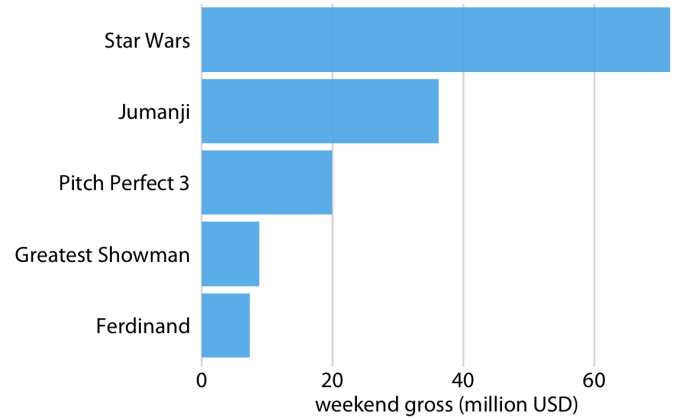
Bar plots

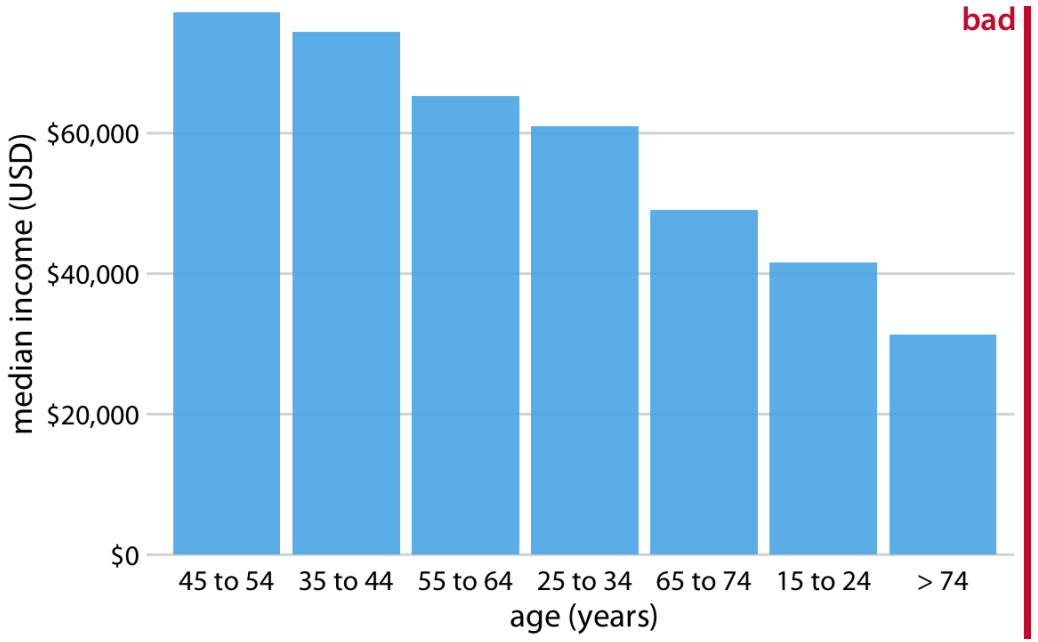
Highest grossing movies for the weekend of December 22-24, 2017

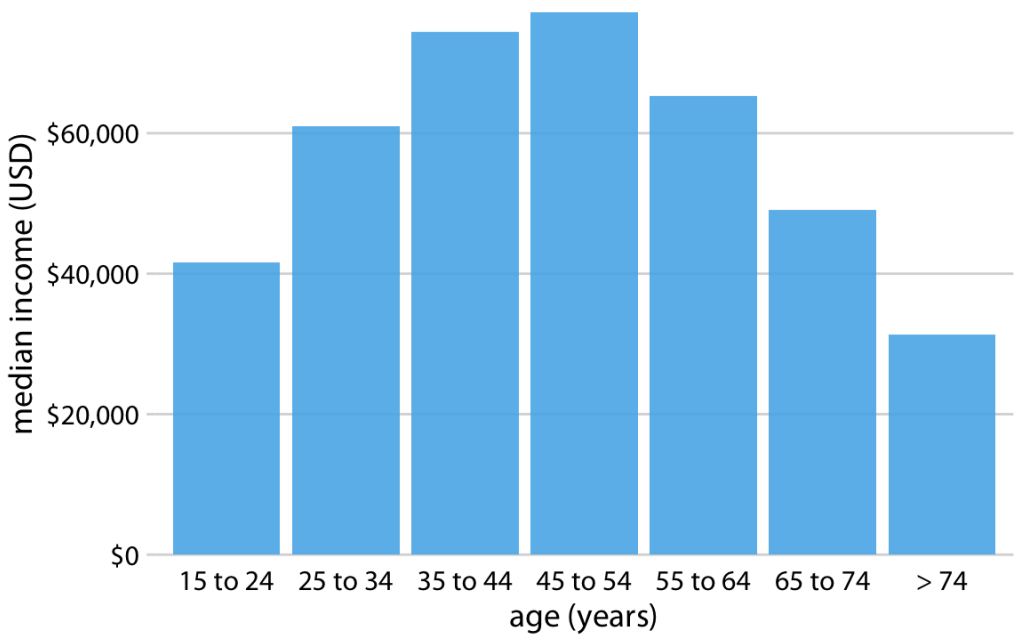



ugly
weekend gross (million USD)
0
20
40
60
Star Wars
Jumanji
Pitch Perfect 3
Greatest Showman
Ferdinand


bad
Greatest Showman
Pitch Perfect 3
Ferdinand
Star Wars
Jumanji
0
20
40
60
weekend gross (million USD)

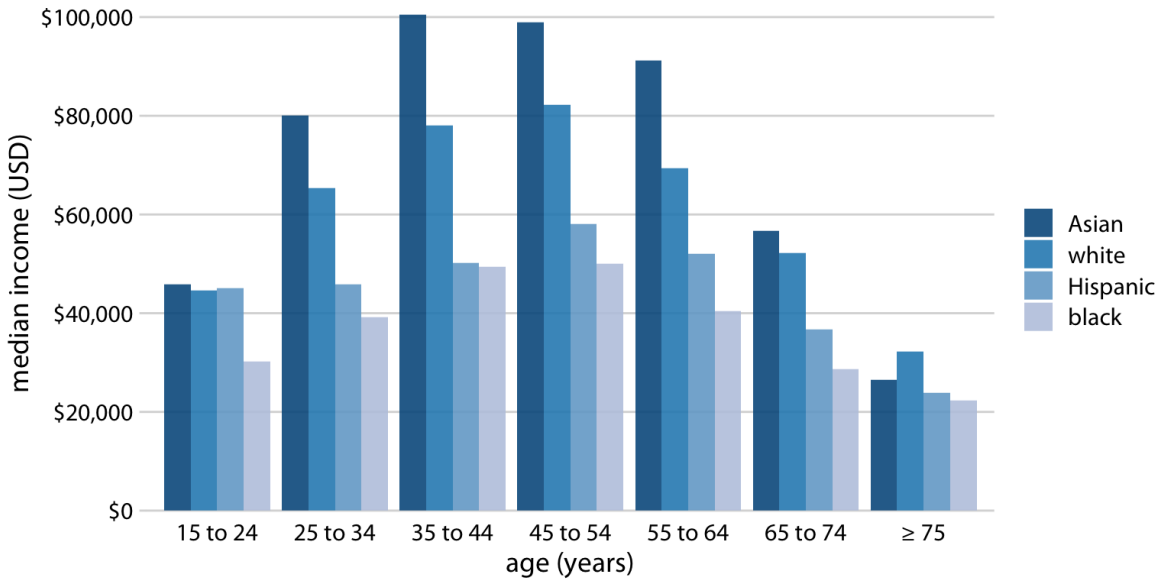

Star Wars
Jumanji
Pitch Perfect 3
Greatest Showman
Ferdinand
0
20
40
60
weekend gross (million USD)

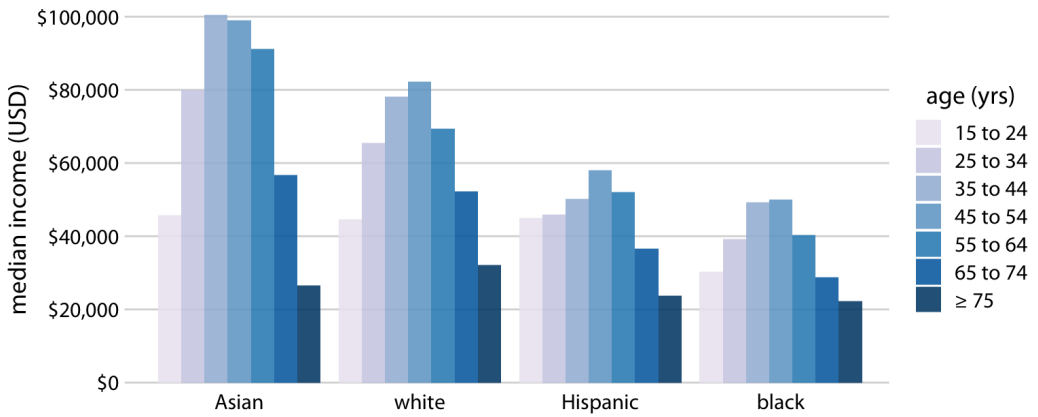

bad
median income (USD)
$60,000
$40,000
$20,000
$0
45 to 54
35 to 44
55 to 64
25 to 34
65 to 74
15 to 24
> 74
age (years)

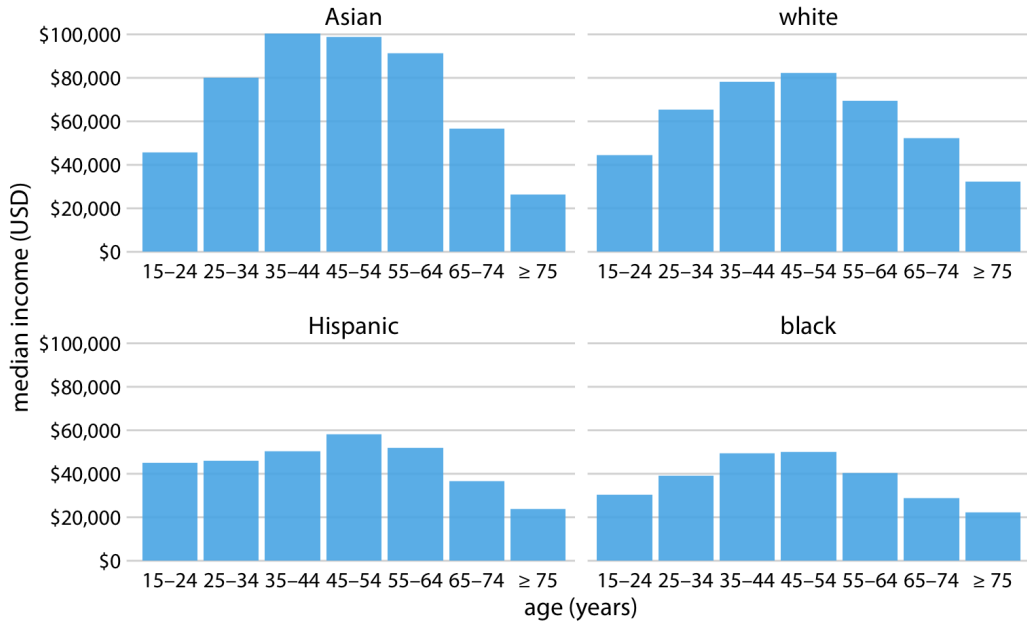

median income (USD)
$60,000
$40,000
$20,000
$0
15 to 24
25 to 34
35 to 44
45 to 54
55 to 64
65 to 74
> 74
age (years)

Grouped and stacked bars

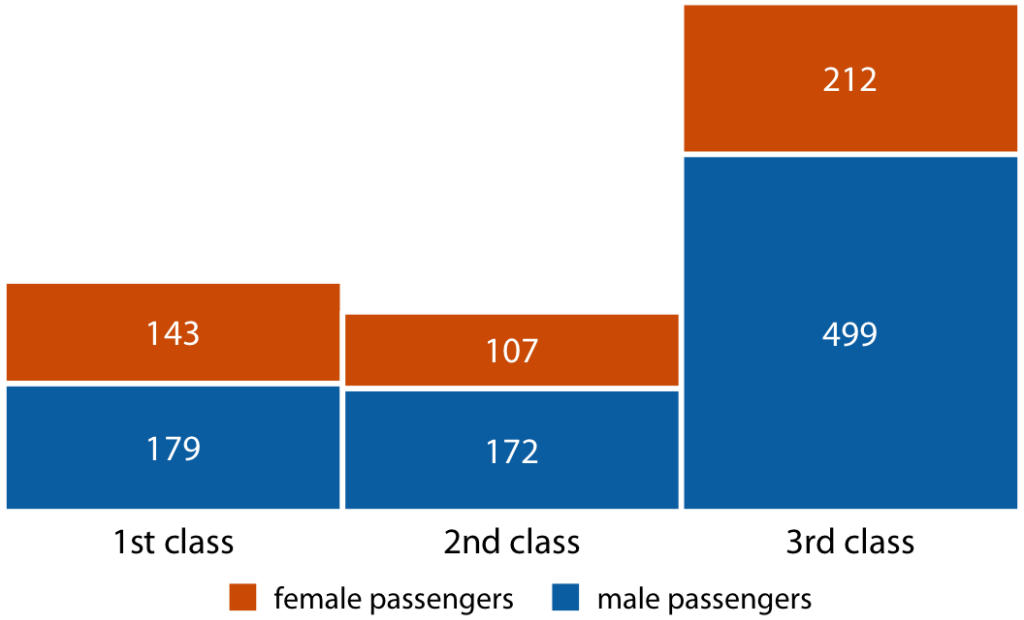
2016 median U.S. annual household income versus age group and race



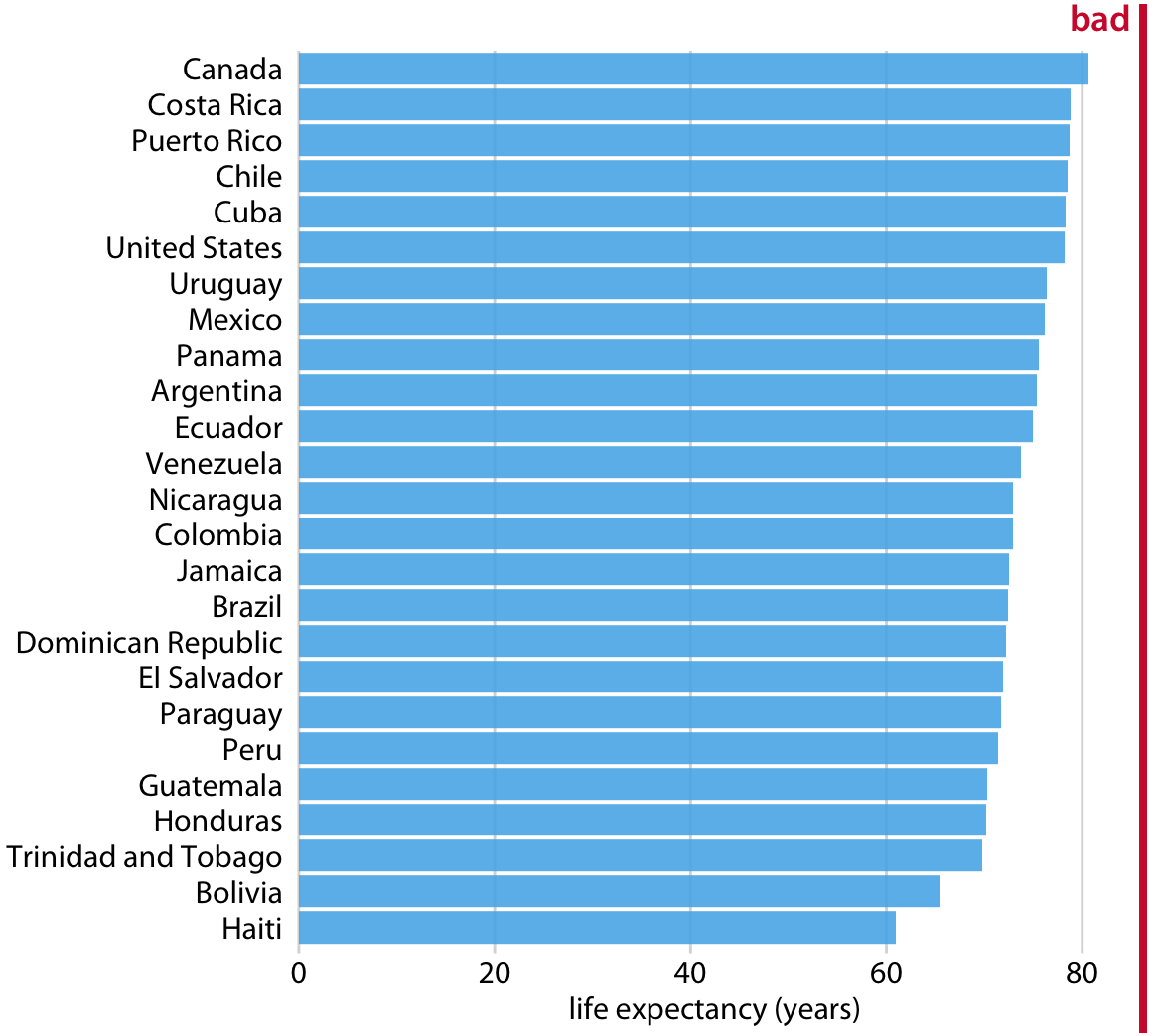

median income (USD)
$100,000
$80,000
$60,000
$40,000
$20,000
$0
Asian
white
Hispanic
black
age (yrs)
15 to 24
25 to 34
35 to 44
45 to 54
55 to 64
65 to 74
≥ 75


Asian
white
Hispanic
black
median income (USD)
$100,000
$80,000
$60,000
$40,000
$20,000
$0
15–24 25–34 35–44 45–54 55–64 65–74 ≥ 75
age (years)

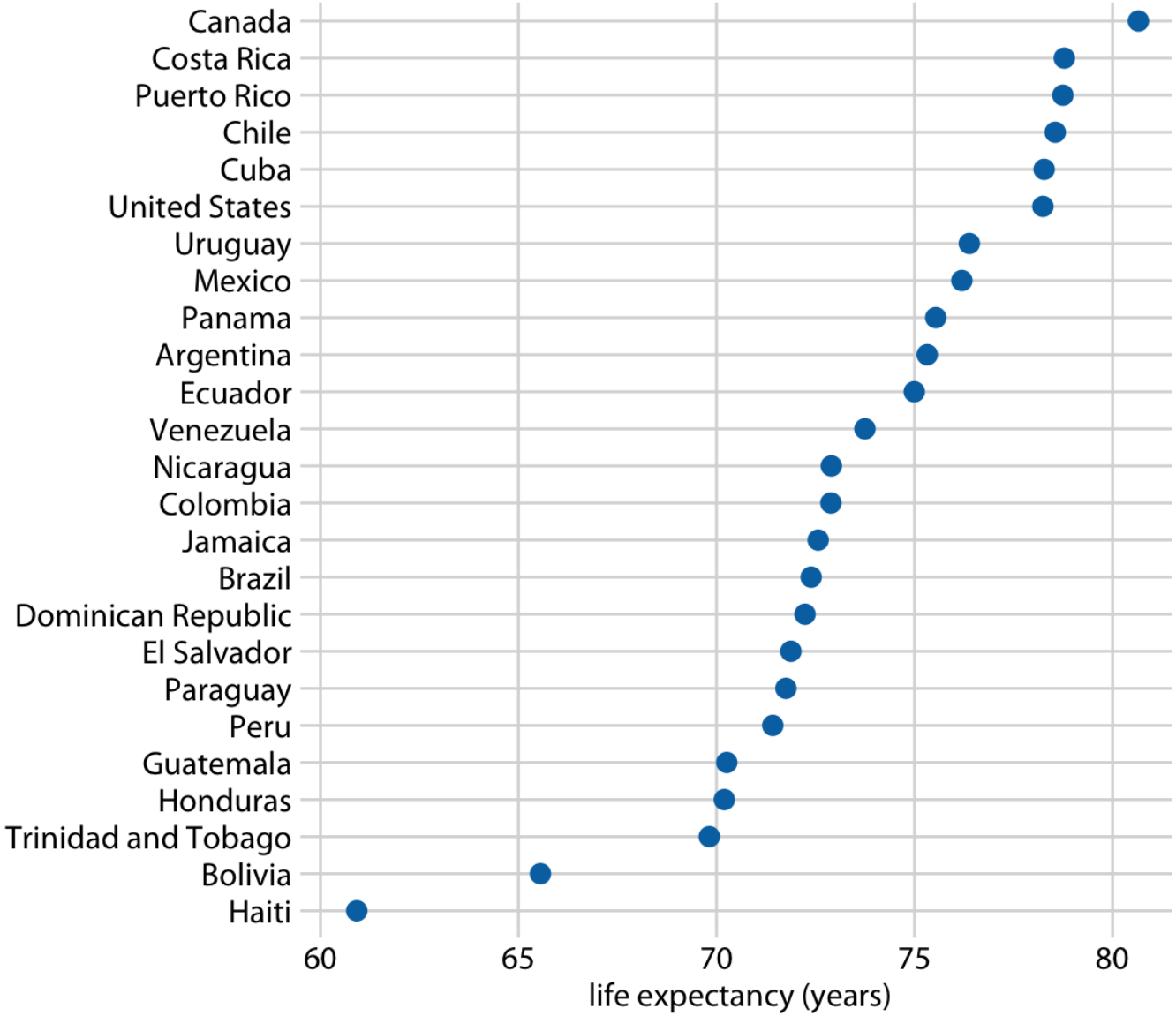
Whenever a plot is meant to display only a small number of different values, it makes sense to add the actual numbers to the plot. This substantially increases the amount of information conveyed by the plot without adding much visual noise, and it removes the need for an explicit y axis.



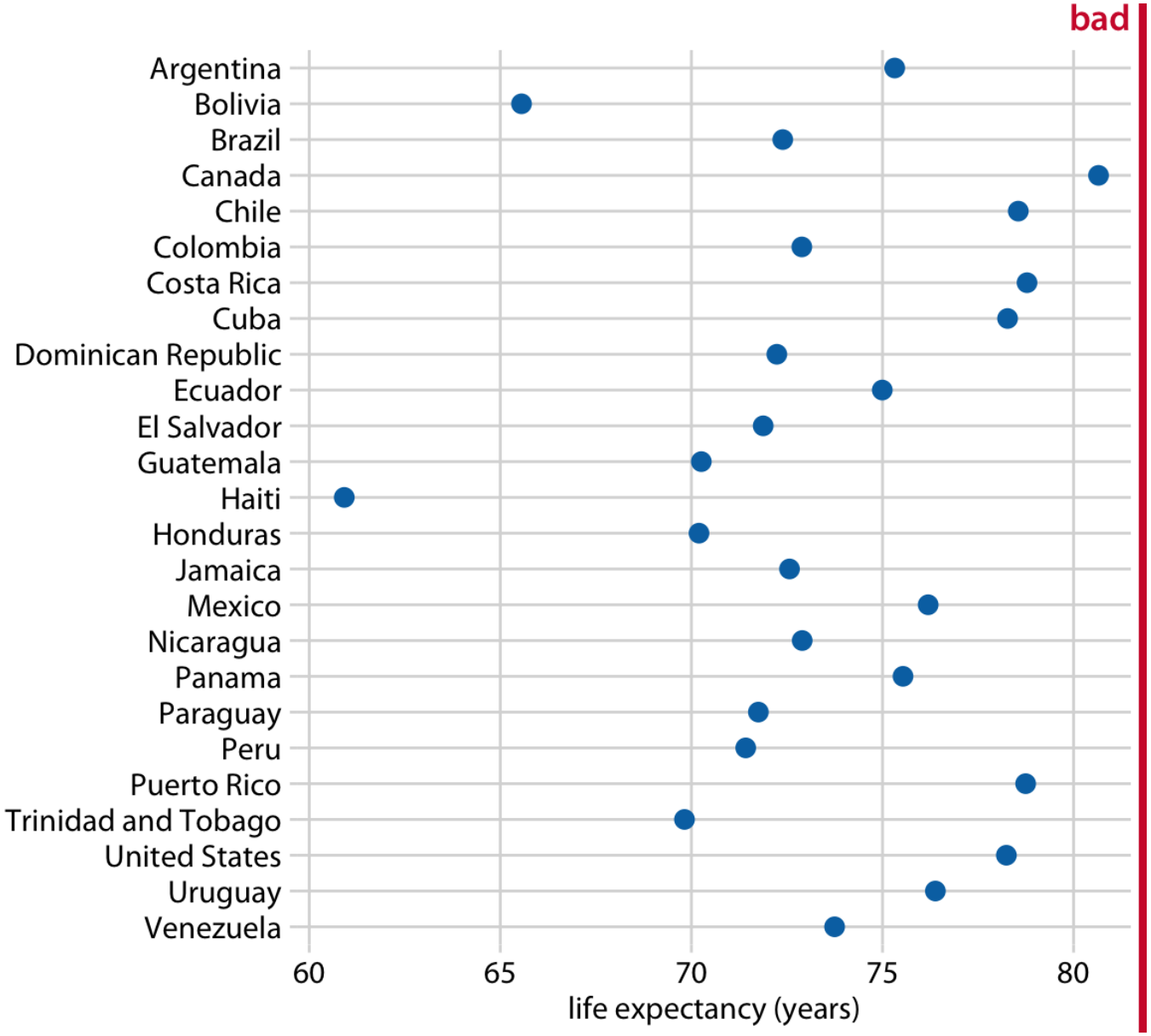
Life expectancies of countries in the Americas, for the year 2007, shown as bars.



Dot plots


Canada
Costa Rica
Puerto Rico
Chile
Cuba
United States
Uruguay
Mexico
Panama
Argentina
Ecuador
Venezuela
Nicaragua
Colombia
Jamaica
Brazil
Dominican Republic
El Salvador
Paraguay
Peru
Guatemala
Honduras
Trinidad and Tobago
Bolivia
Haiti
60
65
70
75
80
life expectancy (years)

Misleading Dot Plot



Visualizing Distributions

Histogram of a single variable

Histogram of the ages of Titanic passengers.

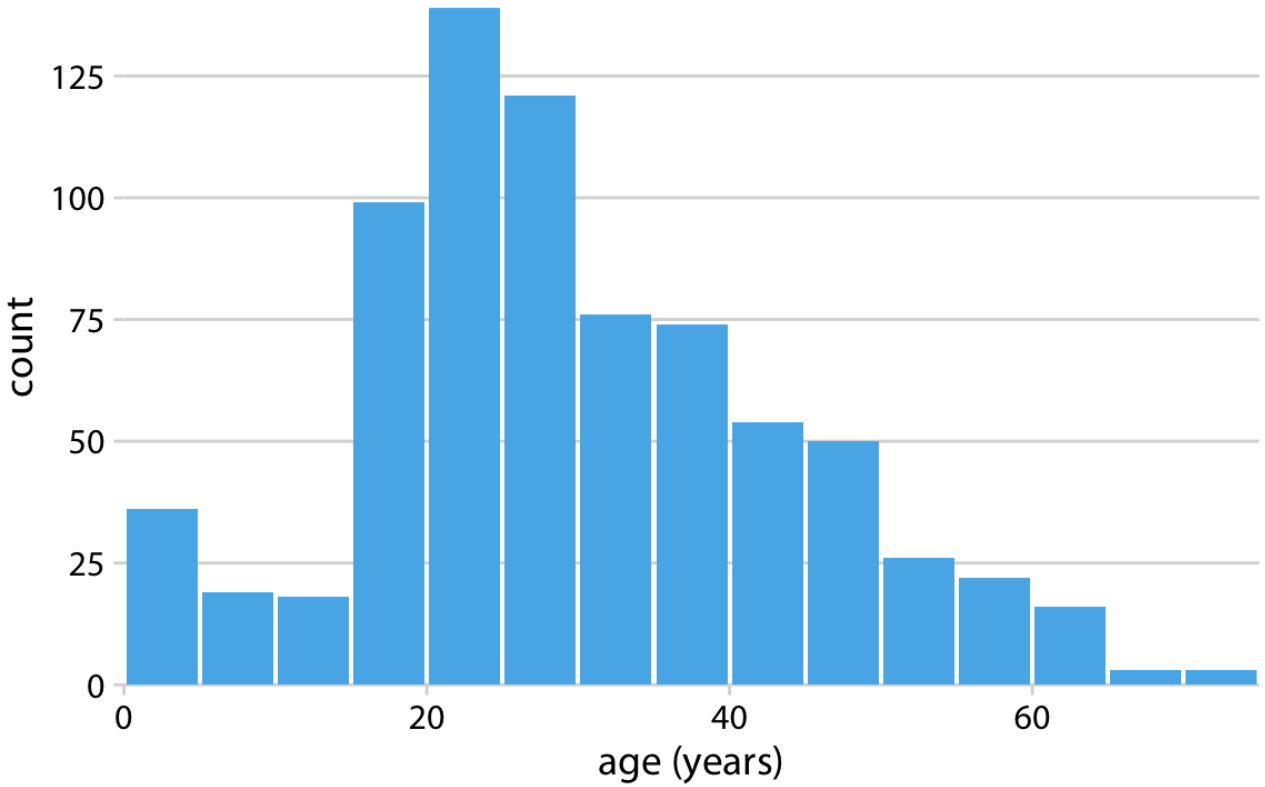
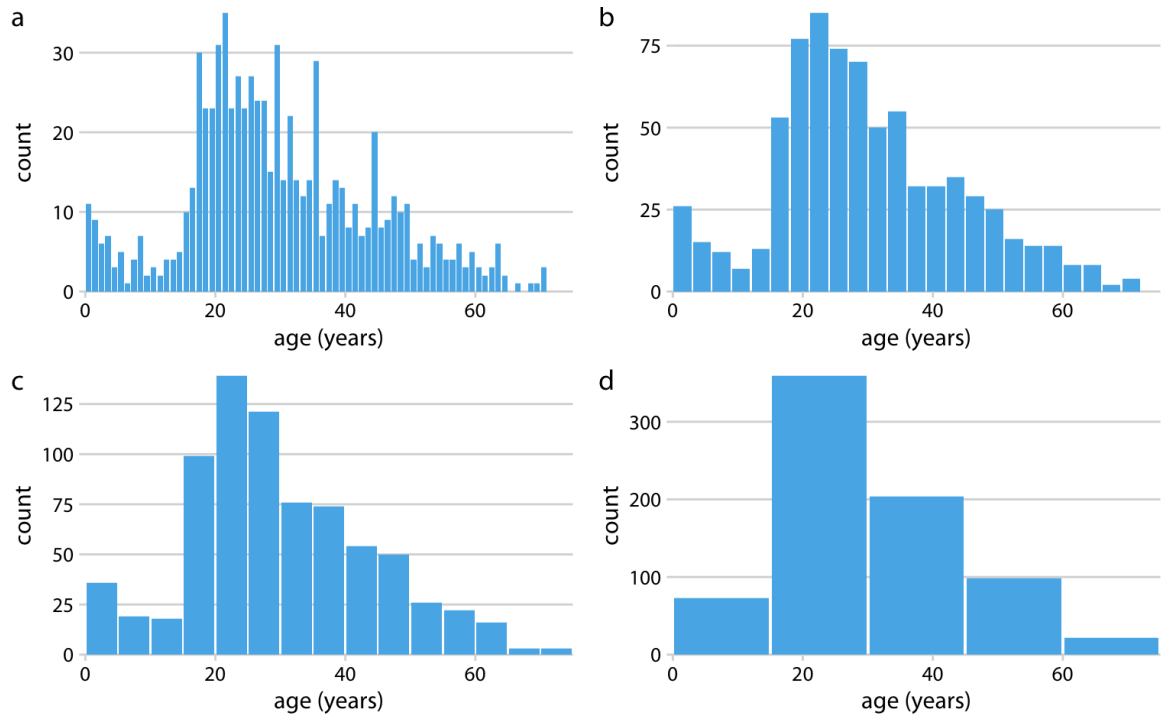
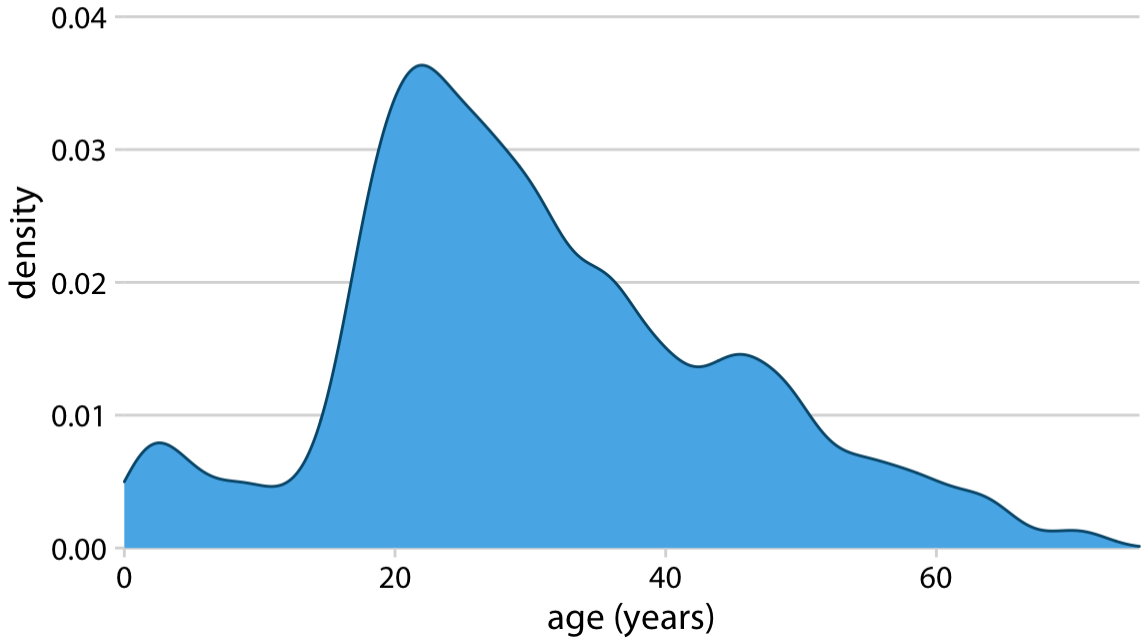


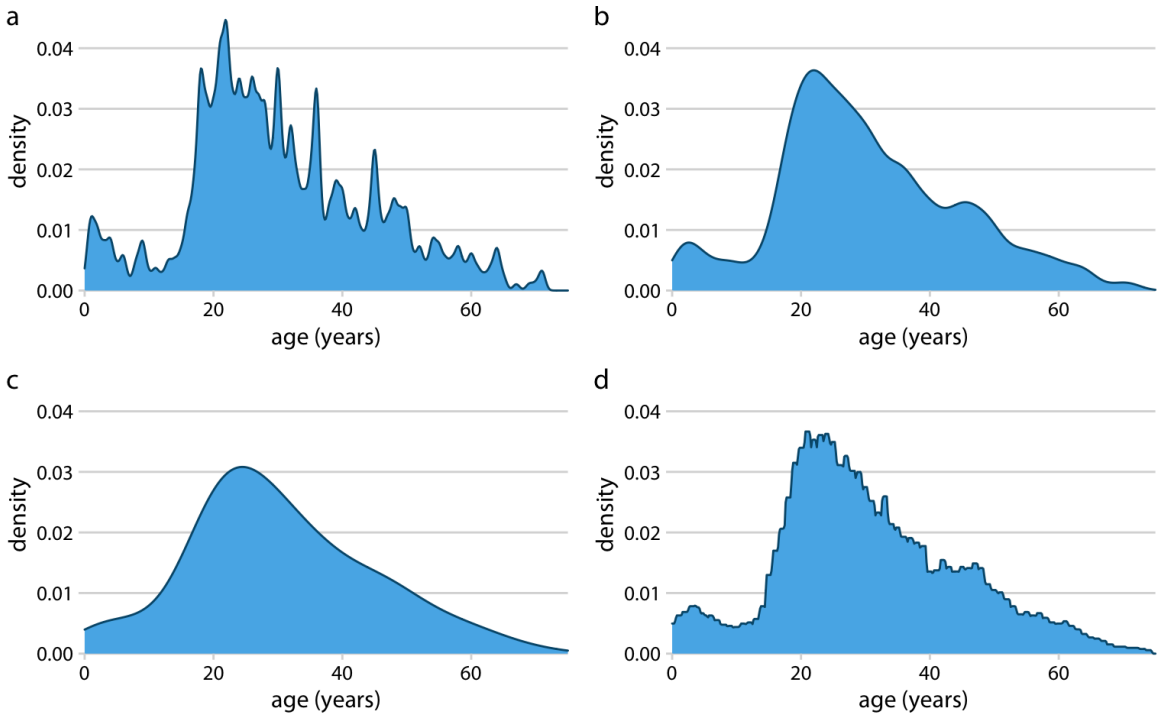
Table 7.1: Numbers of passenger with known age on the Titanic.

Age range	Count	Age range	Count	Age range	Count
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		

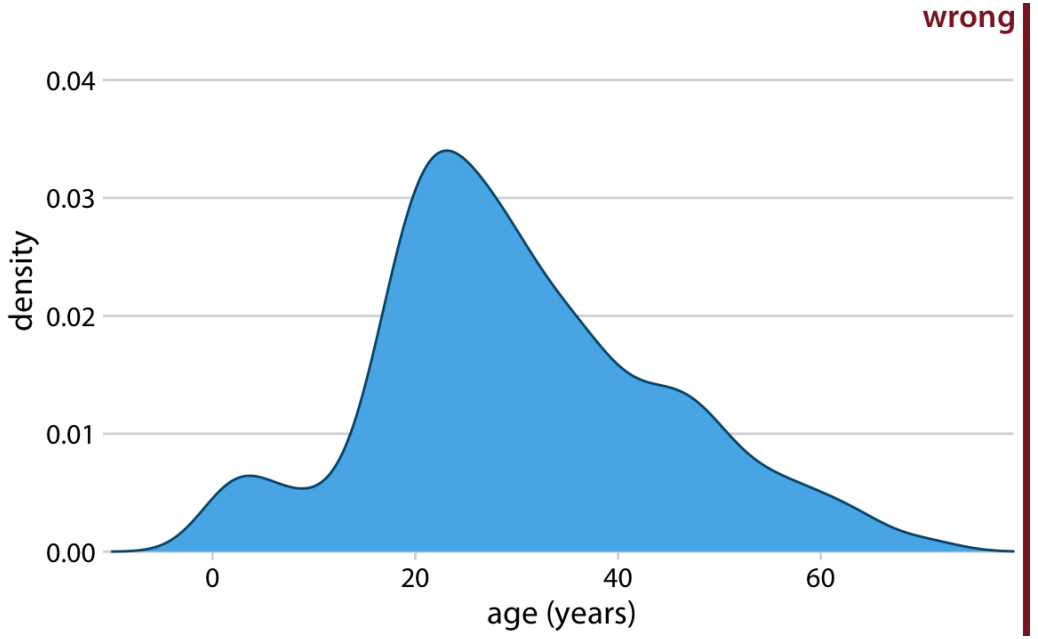


When making a histogram, always explore multiple bin widths.

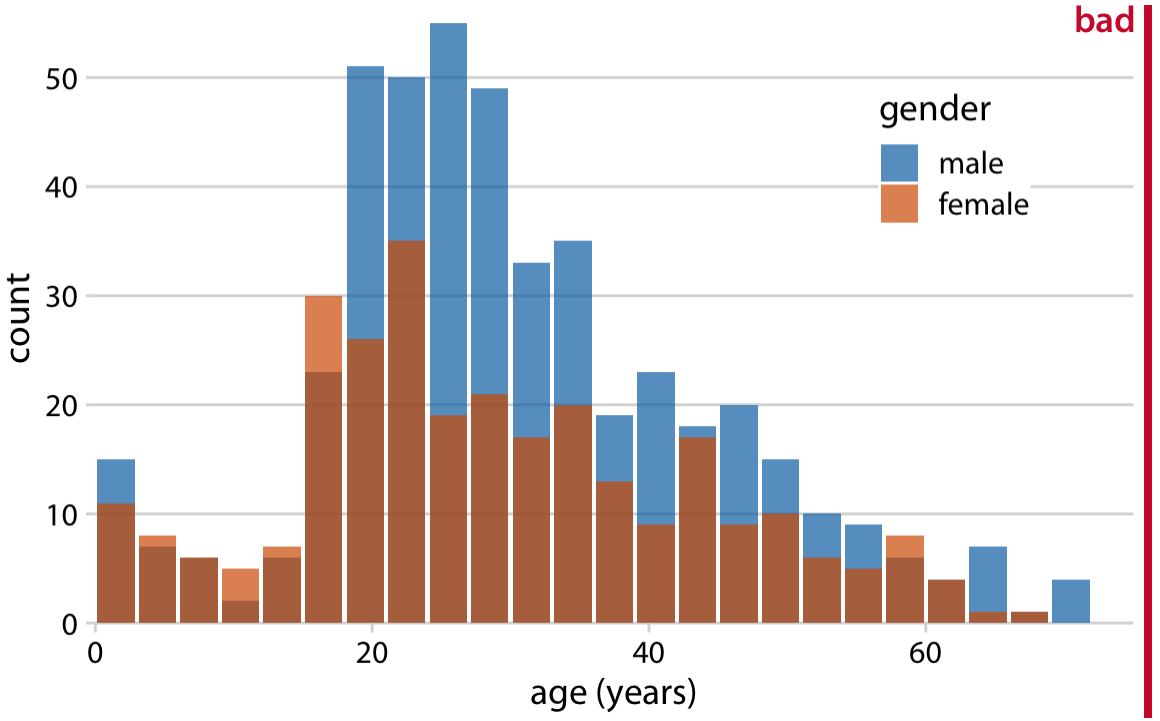

0.04
0.03
0.02
0.01
0.00
density
0
20
40
60
age (years)

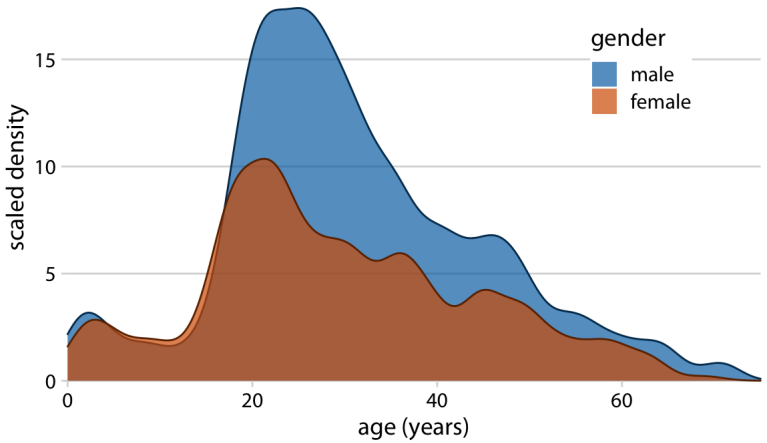

a
b
c
d
density
age (years)

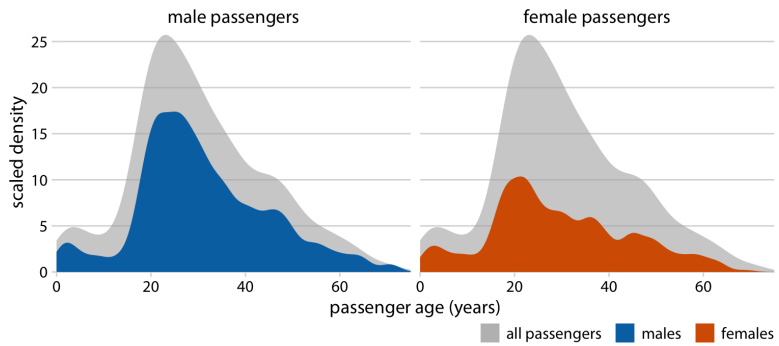
Always verify that your density estimate does not predict the existence of nonsensical data values.

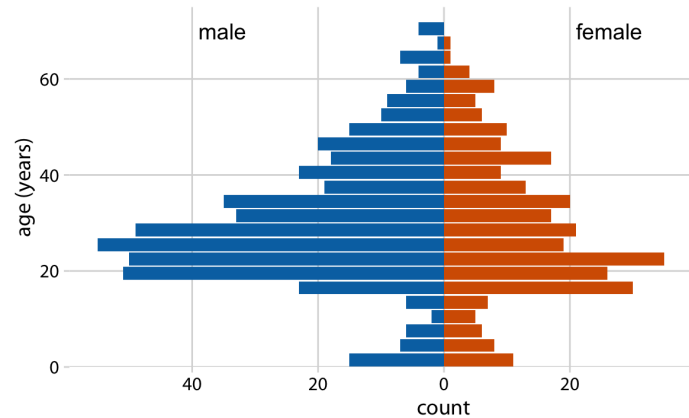


Visualize multiple distributions

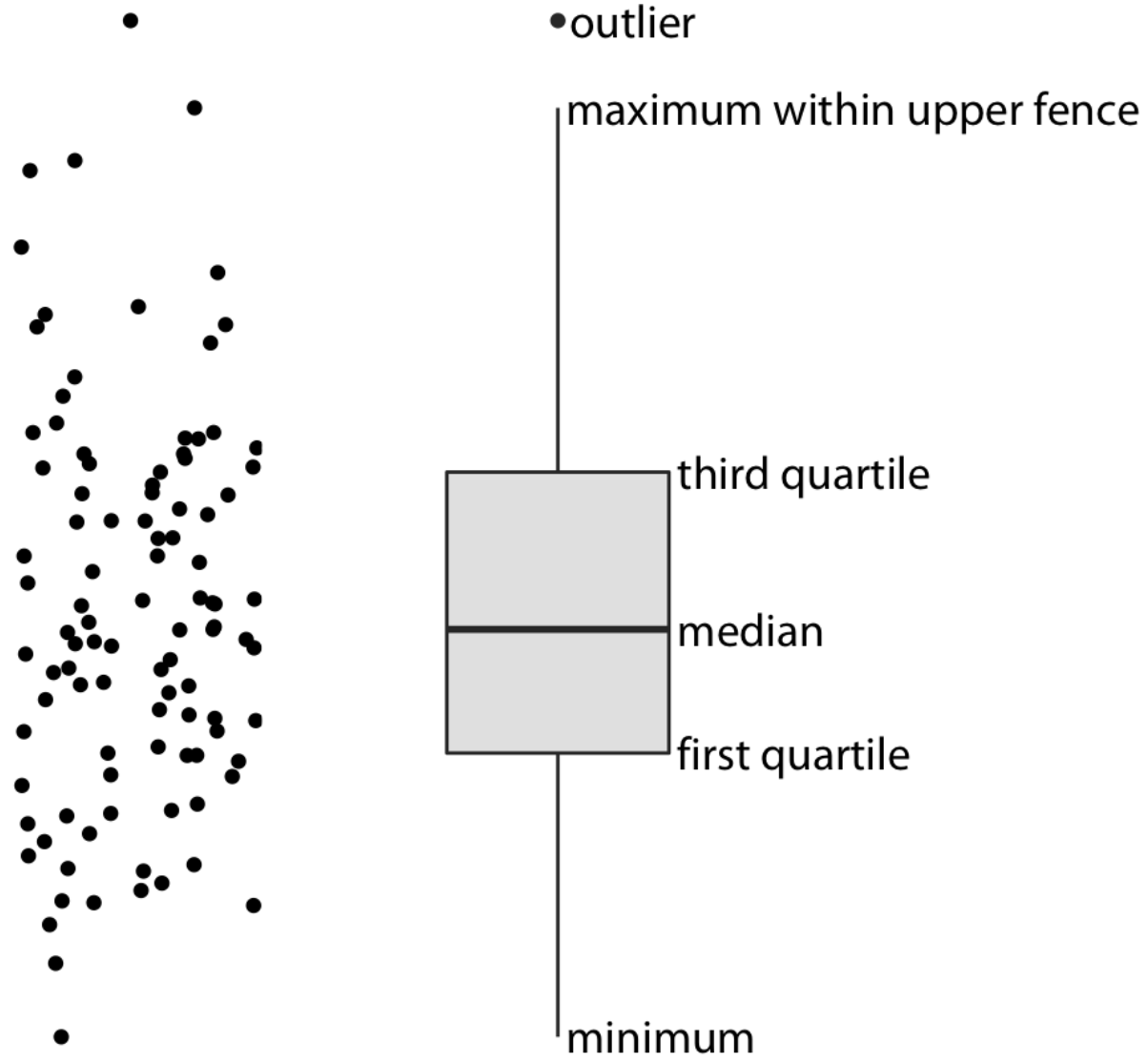



gender
male
female
scaled density
15
10
5
0
0
20
40
60
age (years)

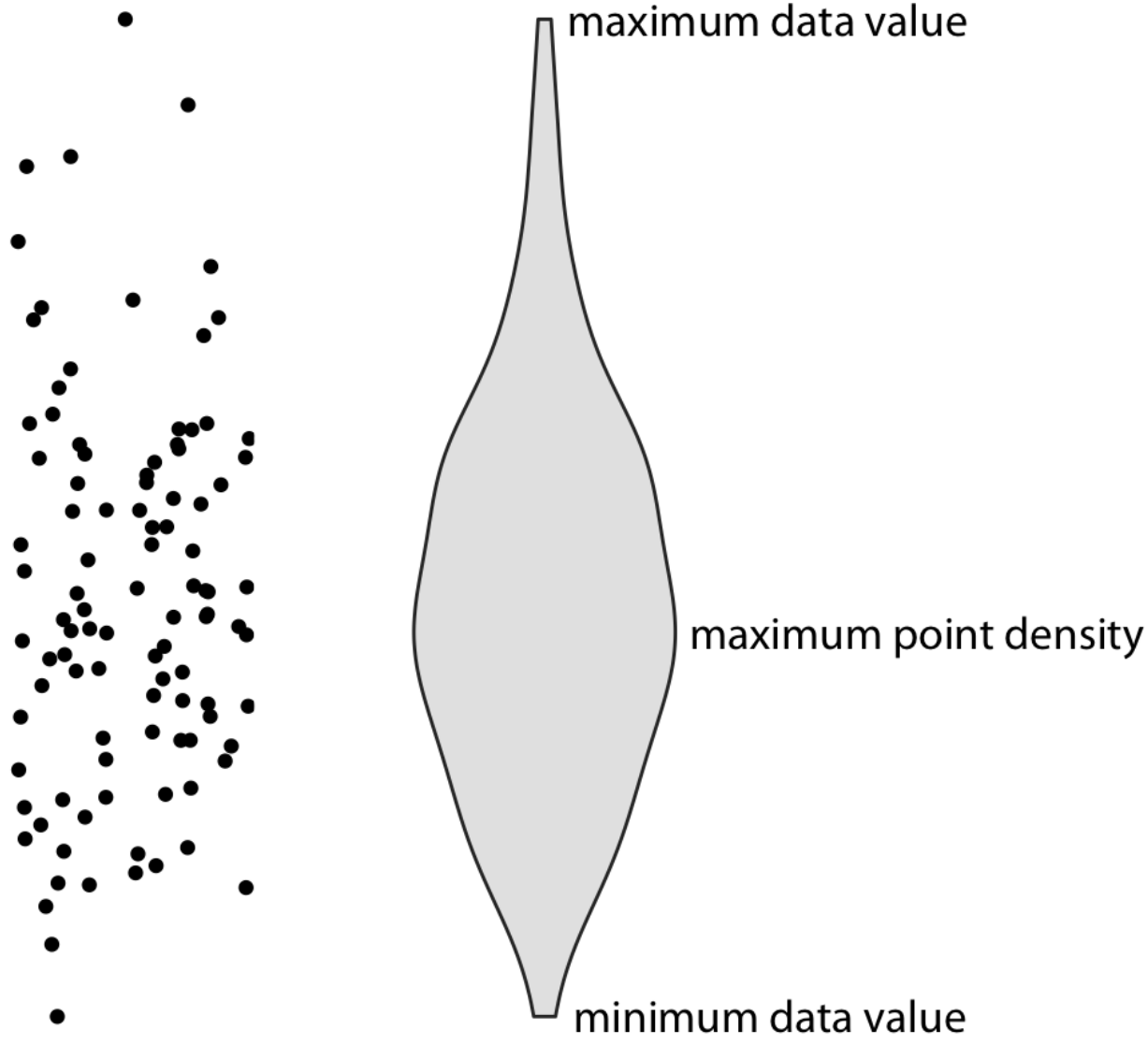

male passengers
female passengers
scaled density
25
20
15
10
5
0
0
20
40
60
passenger age (years)
all passengers
males
females


male
female
age (years)
60
40
20
0
40
20
0
20
count

Box plot

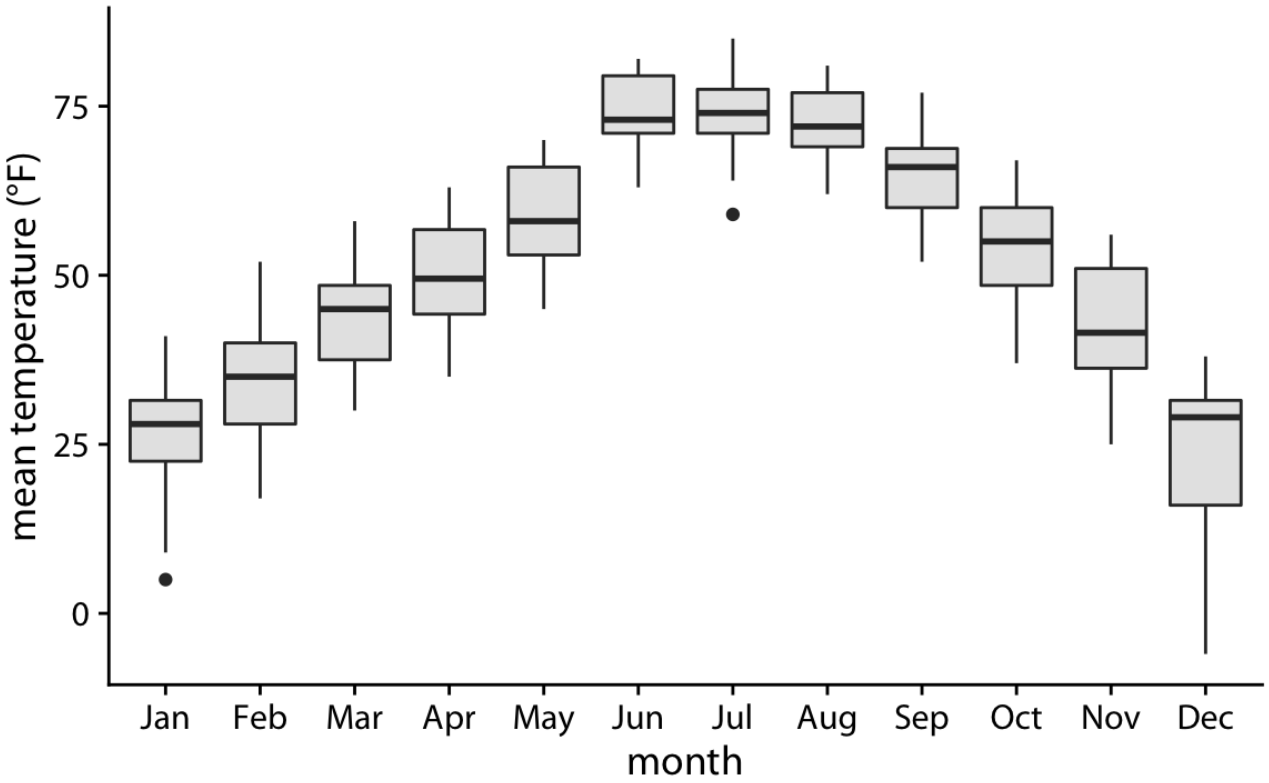

outlier
maximum within upper fence
third quartile
median
first quartile
minimum

Violin plot


maximum data value
maximum point density
minimum data value

Visualizing many distributions at once

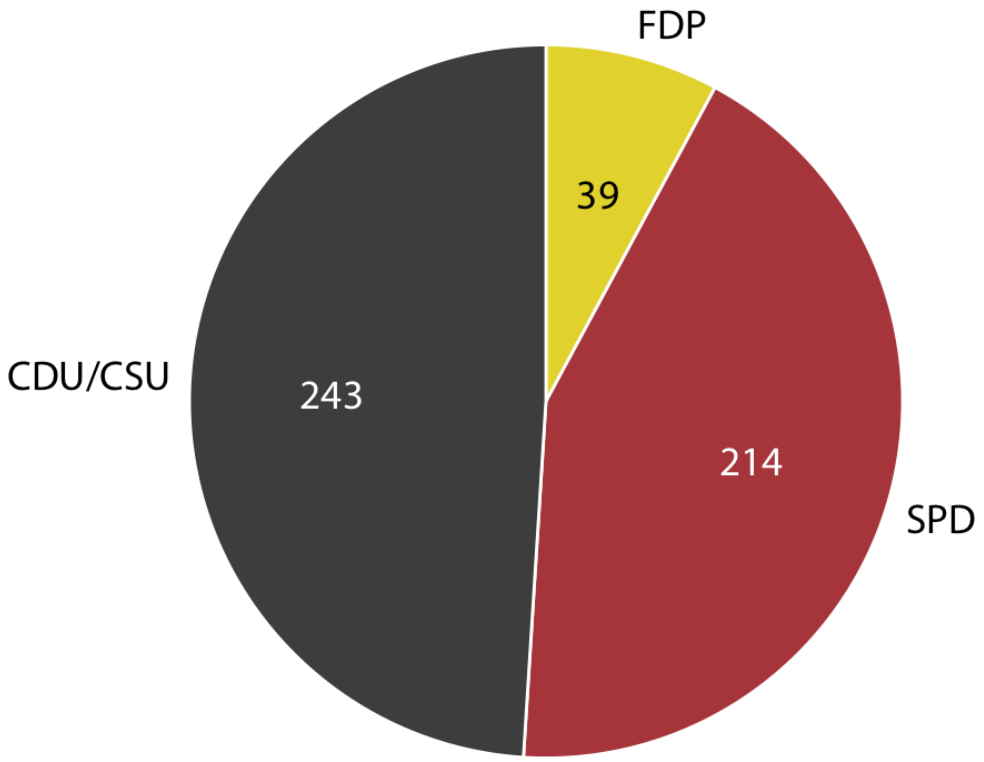
Mean daily temperatures in Lincoln, Nebraska, visualized as boxplots



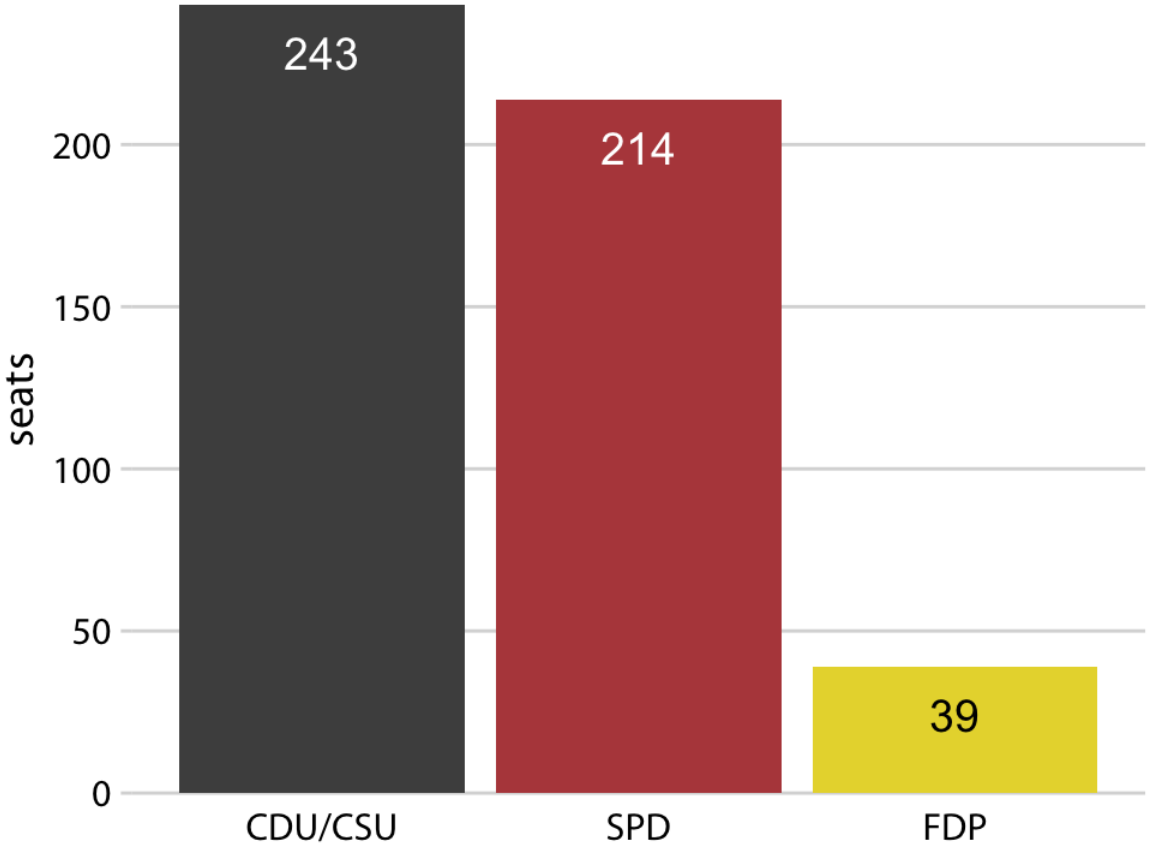
Visualizing Proportions

Pie chart

Party composition of the 8th German Bundestag, 1976–1980, visualized as a pie chart.

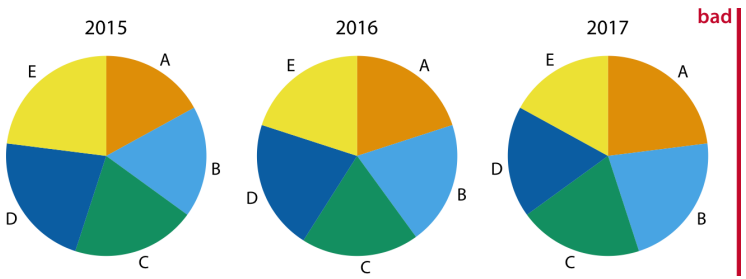


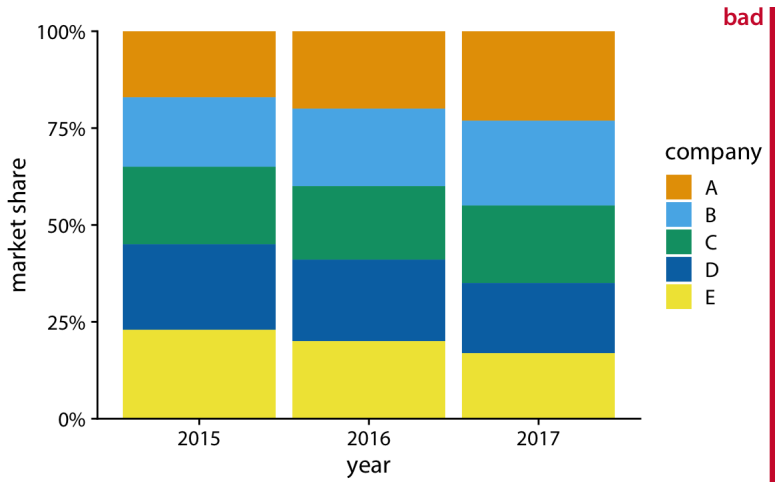
Many authors categorically reject pie charts and argue in favor of side-by-side or stacked bars

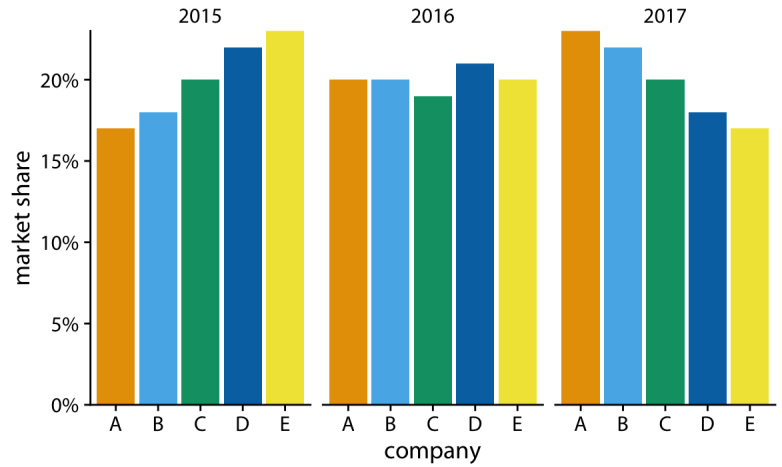


Pros and cons of common approaches to visualizing proportions: pie charts, stacked bars, and side-by-side bars.

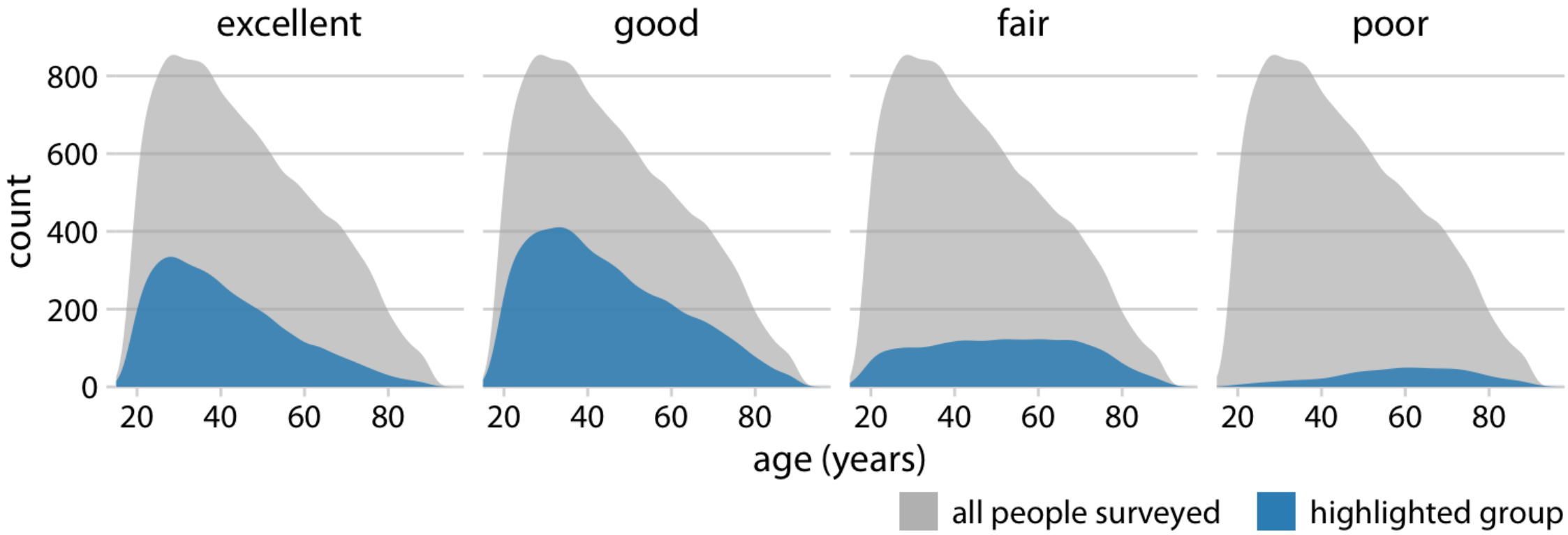
	Pie chart	Stacked bars	Side-by-side bars
Clearly visualizes the data as proportions of a whole	✔	✔	✖
Allows easy visual comparison of the relative proportions	✖	✖	✔
Visually emphasizes simple fractions, such as 1/2, 1/3, 1/4	✔	✖	✖
Looks visually appealing even for very small datasets	✔	✖	✔
Works well when the whole is broken into many pieces	✖	✖	✔
Works well for the visualization of many sets of proportions or time series of proportions	✖	✔	✖


bad
2015
2016
2017
A
B
C
D
E


bad
100%
75%
50%
25%
0%
market share
2015
2016
2017
year
company
A
B
C
D
E


2015
2016
2017
20%
15%
10%
5%
0%
market share
A
B
C
D
E
company

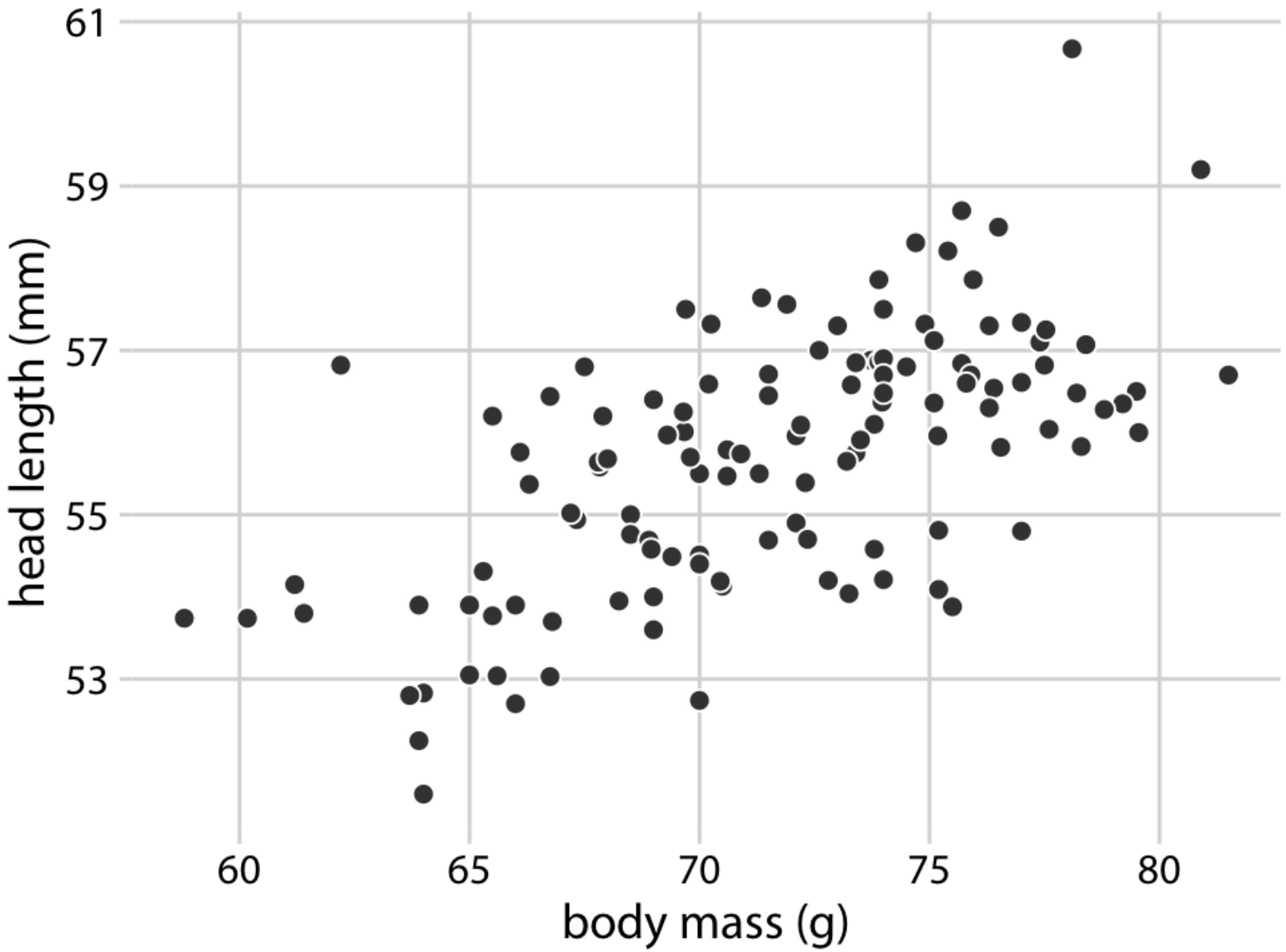
Health status by age, shown as proportion of the total number of people

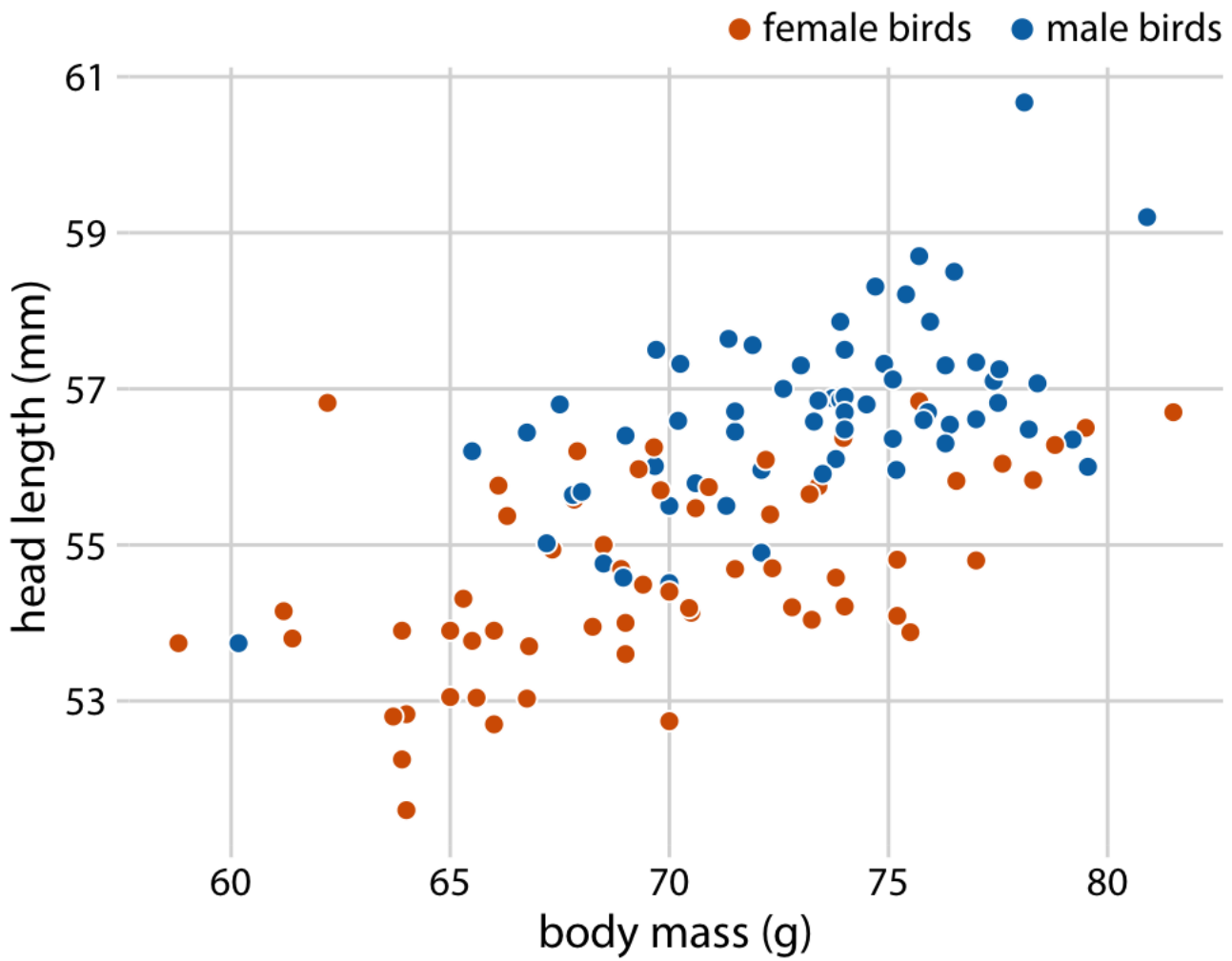


Visualizing Relationships

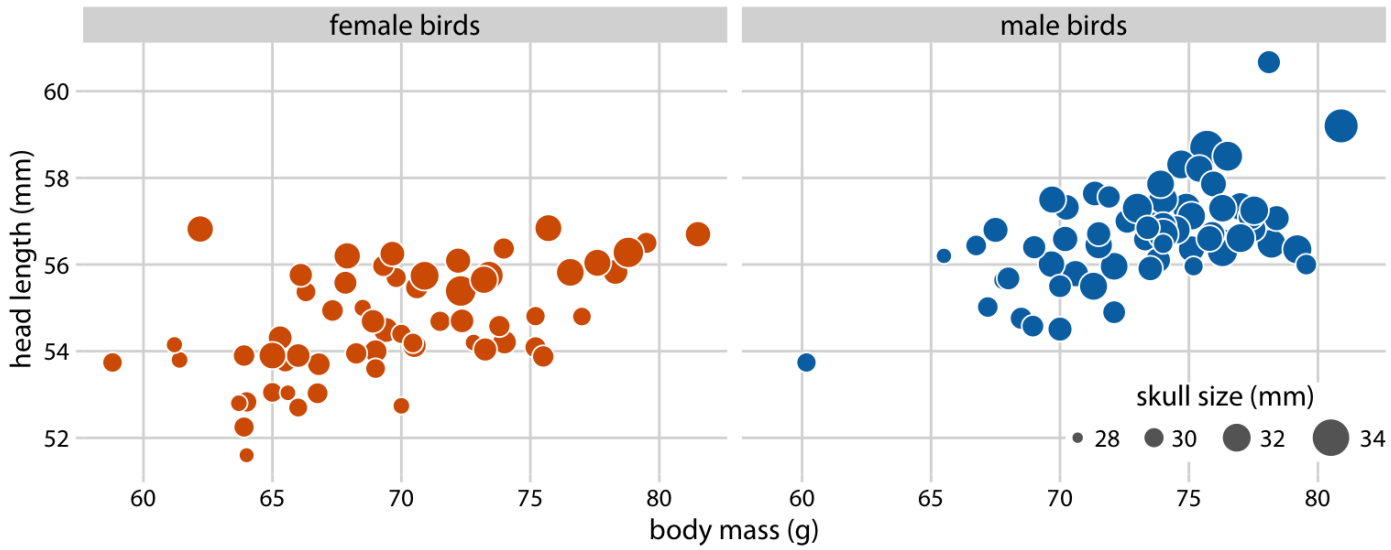
Scatter plot

Head length (measured from the tip of the bill to the back of the head, in mm) versus body mass (in gram), for 123 blue jays

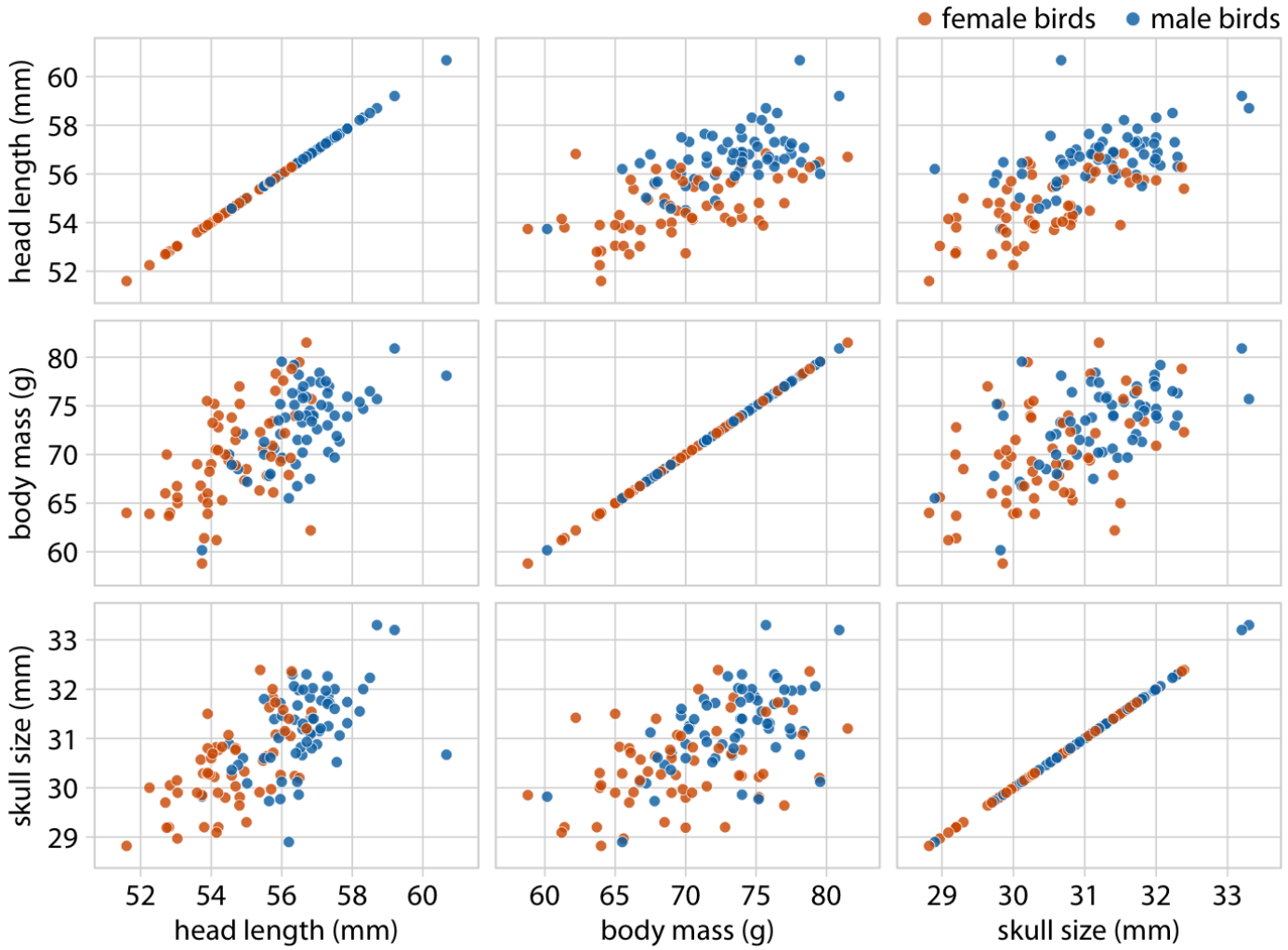



female birds
male birds
61
59
57
55
53
head length (mm)
60
65
70
75
80
body mass (g)

Bubble chart



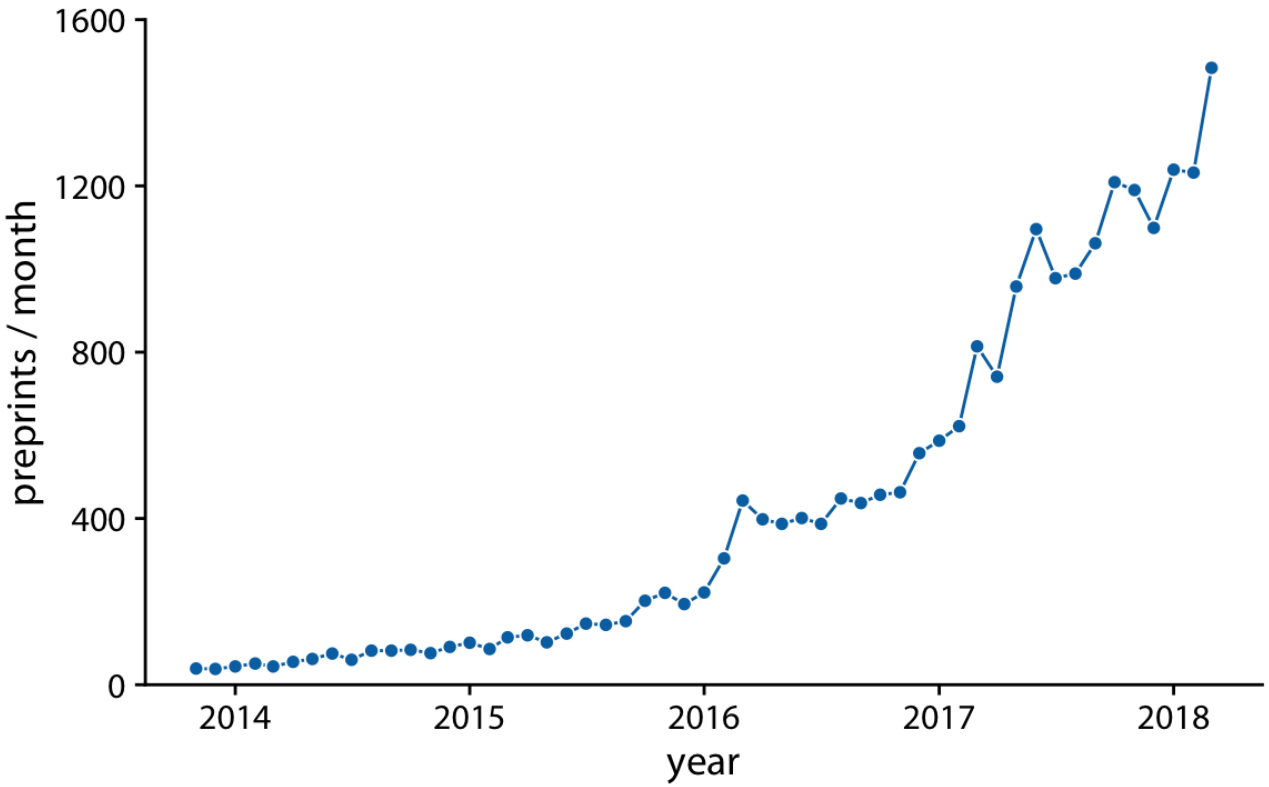
Scatter plot matrix



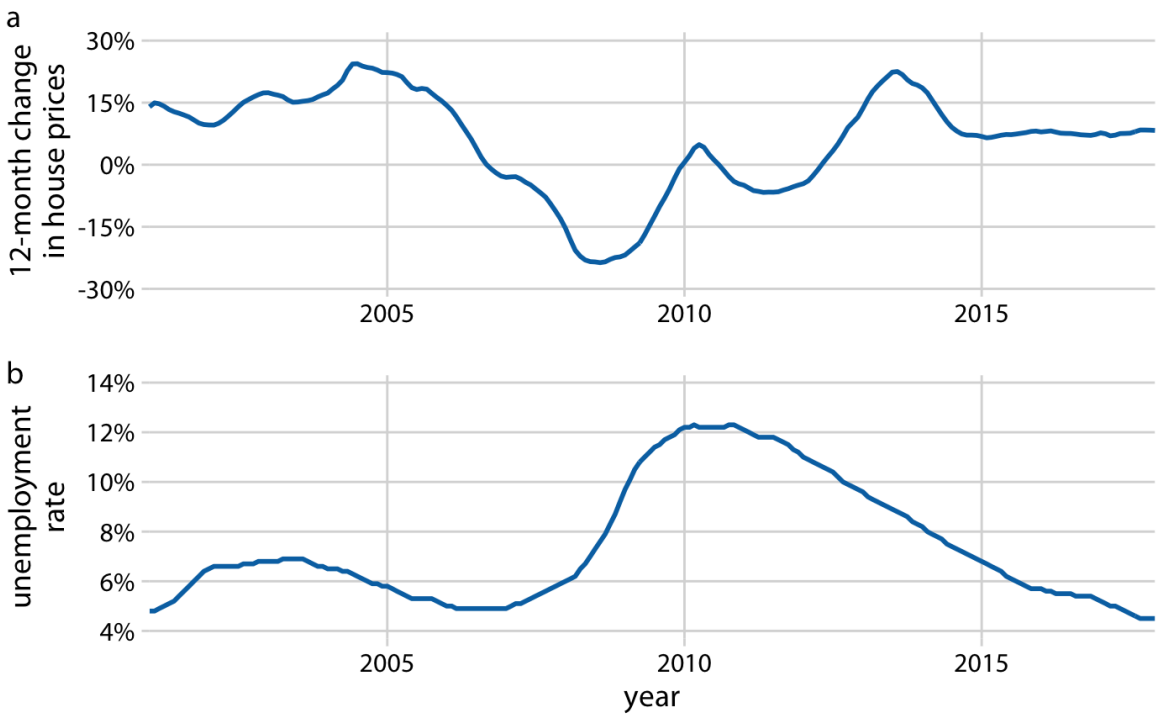
Line Chart

Helps to visualize trends over time

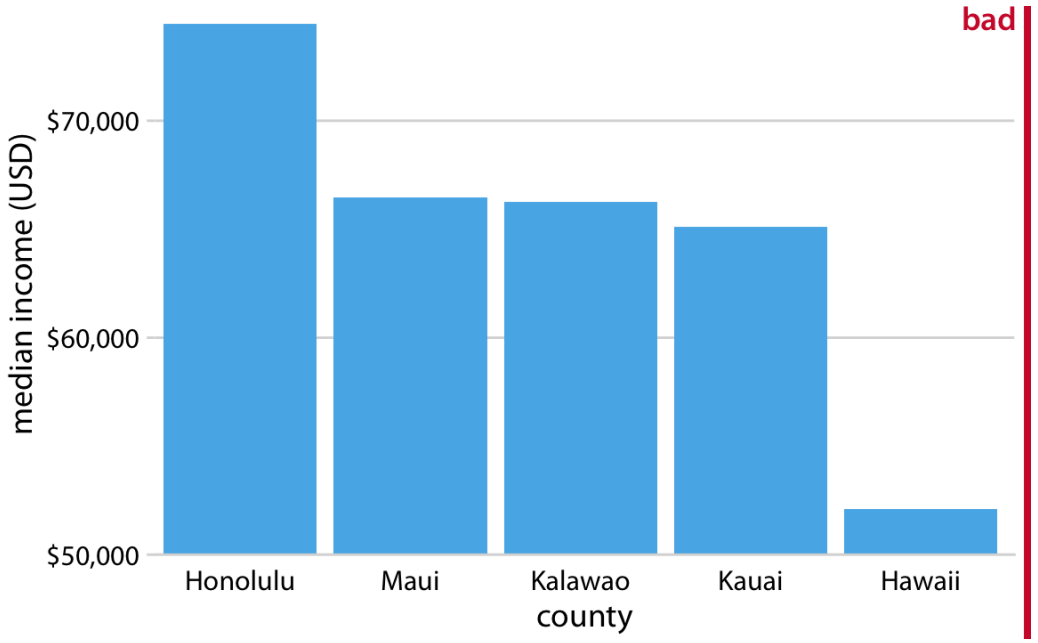
(Time Series)

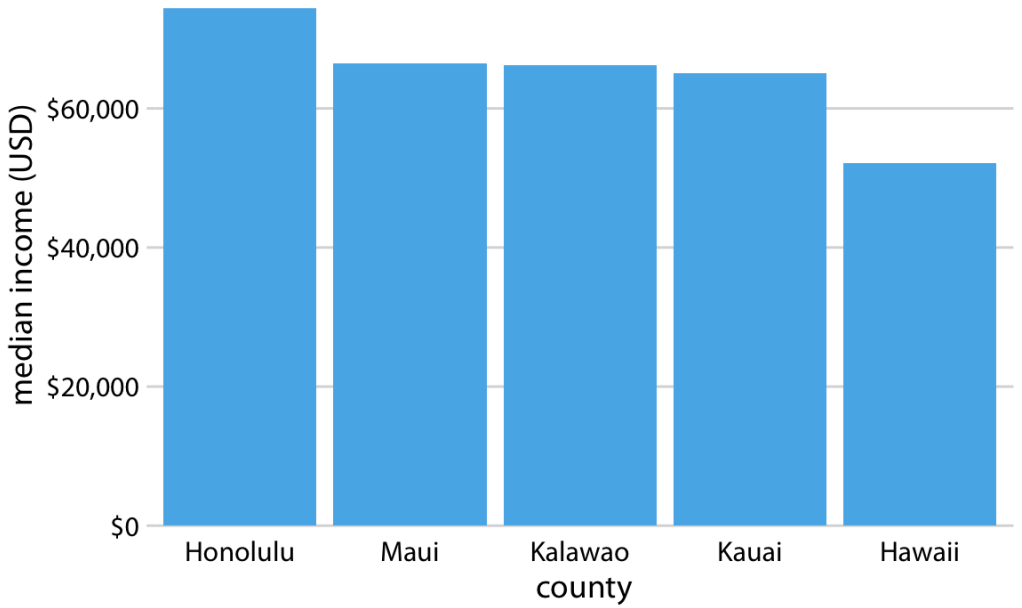


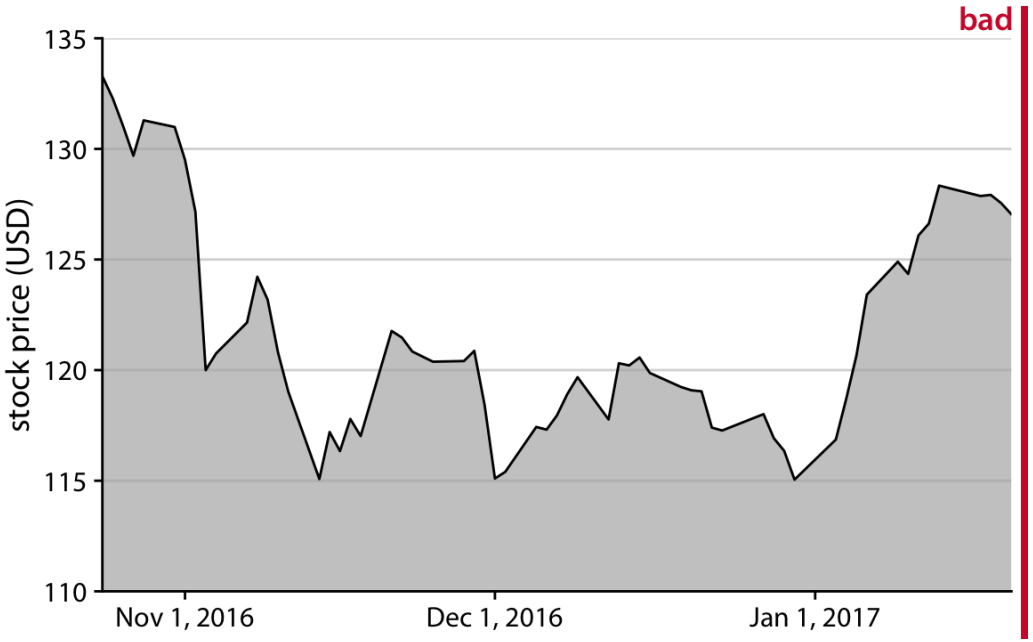
Time Series of two variables

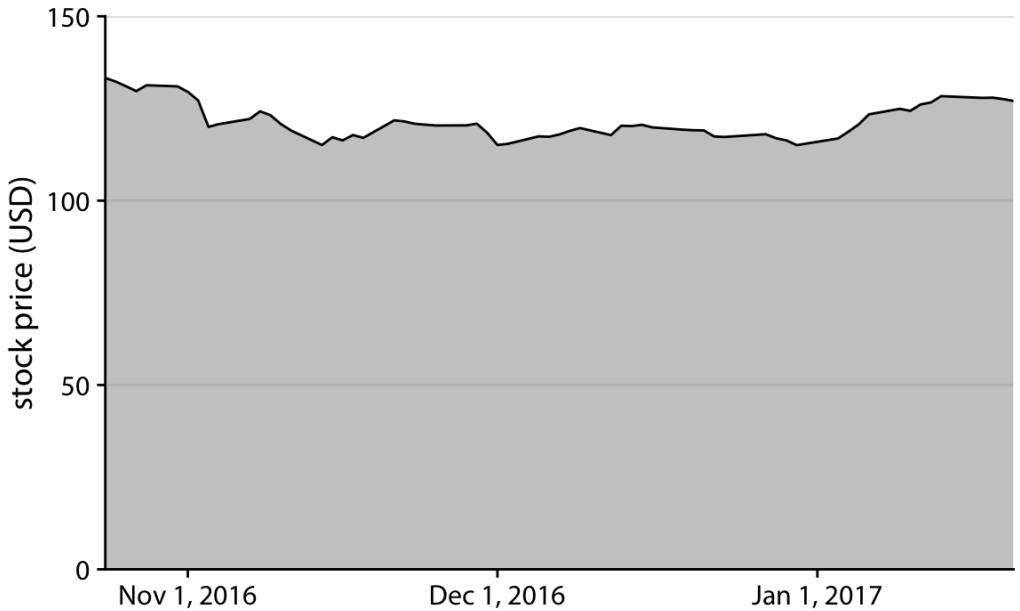


Common Mistakes


bad
median income (USD)
$70,000
$60,000
$50,000
Honolulu
Maui
Kalawao
Kauai
Hawaii
county


median income (USD)
$60,000
$40,000
$20,000
$0
Honolulu
Maui
Kalawao
Kauai
Hawaii
county





meftaul.com | Stock price of Facebook (FB) from Oct. 22, 2016 to Jan. 21, 2017

Tips:

- Bars on a linear scale should always start at 0.
- Use a log scale when the data spans several orders of magnitude.
- Always label the y-axis with the unit of measurement.

Popular Data Visualization Tools

Tableau

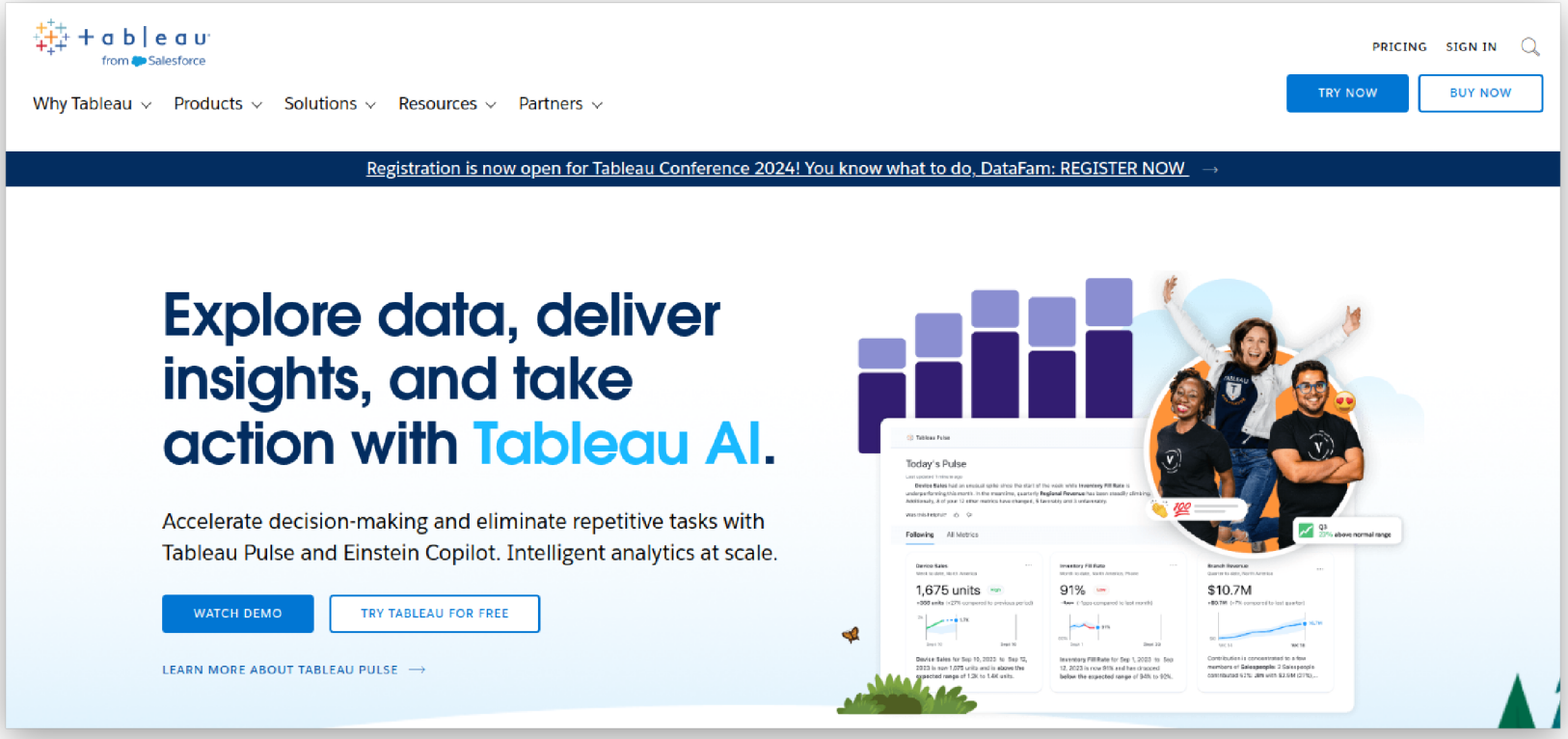
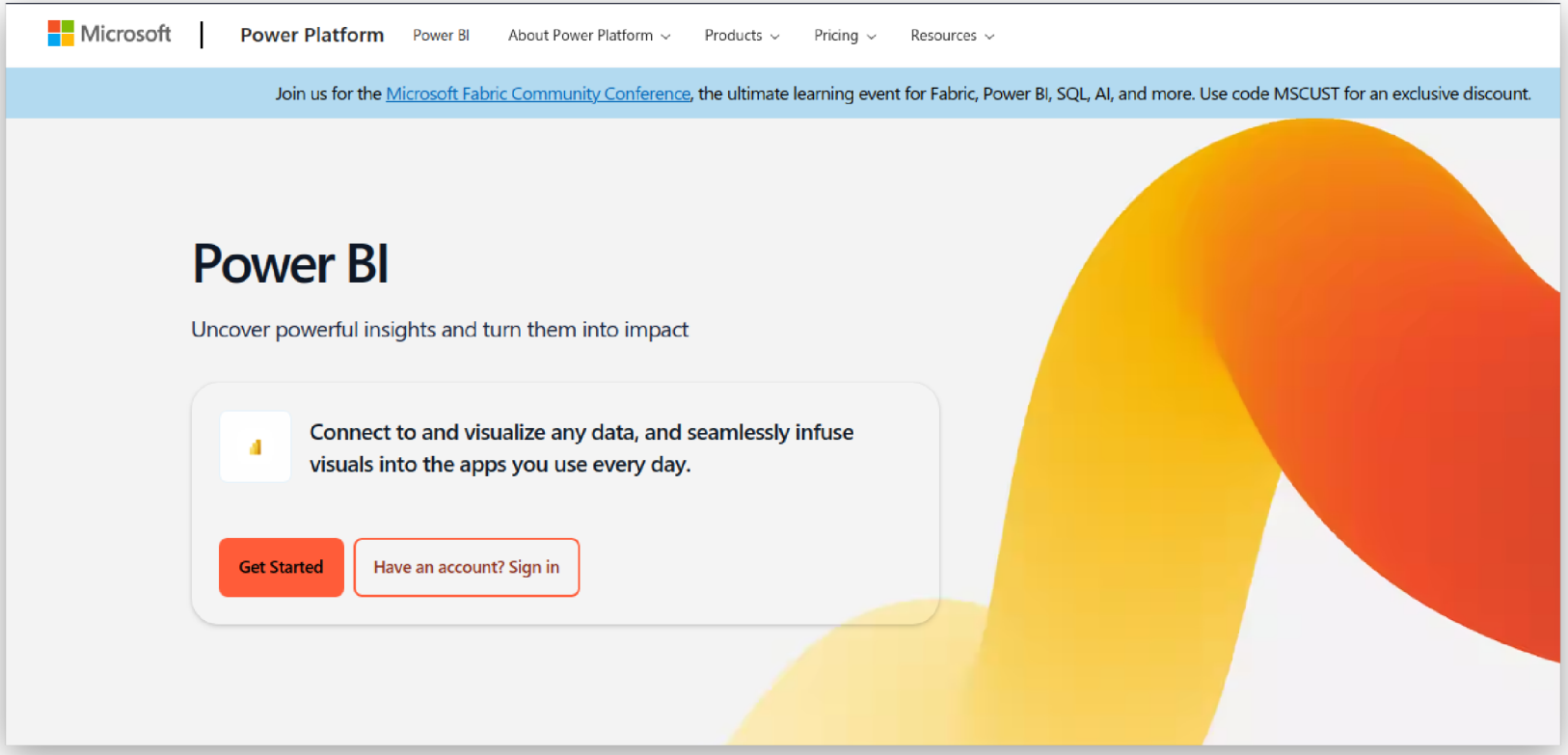


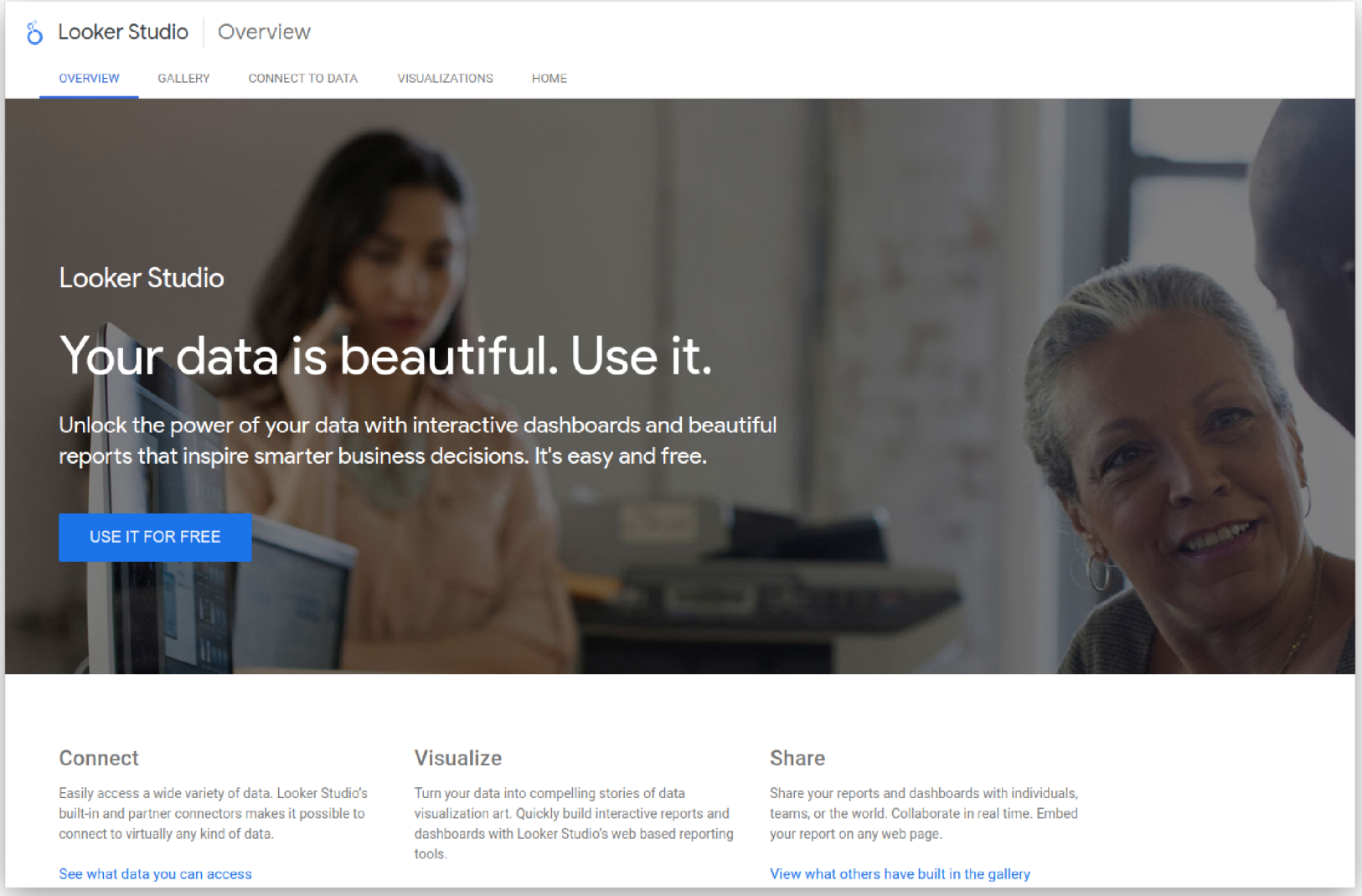
tableau
from Salesforce
PRICING
SIGN IN
TRY NOW
BUY NOW
Why Tableau
Products
Solutions
Resources
Partners
Registration is now open for Tableau Conference 2024! You know what to do, DataFam: REGISTER NOW
Explore data, deliver insights, and take action with Tableau AI.
Accelerate decision-making and eliminate repetitive tasks with Tableau Pulse and Einstein Copilot. Intelligent analytics at scale.
WATCH DEMO
TRY TABLEAU FOR FREE
LEARN MORE ABOUT TABLEAU PULSE
Today's Pulse
1,675 units
91%
$10.7M

Power BI

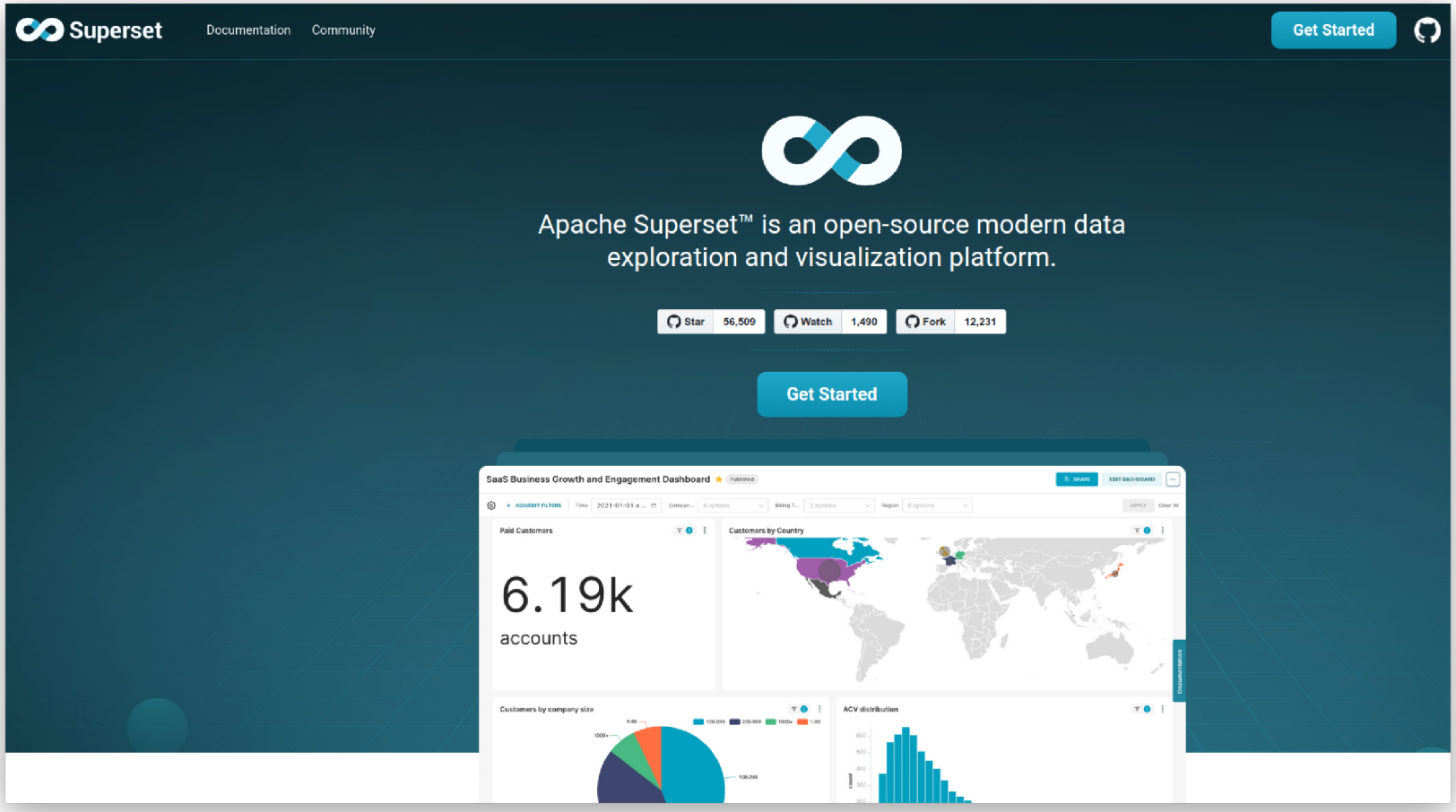

Microsoft
Power Platform
Power BI
About Power Platform
Products
Pricing
Resources
Join us for the Microsoft Fabric Community Conference, the ultimate learning event for Fabric, Power BI, SQL, AI, and more. Use code MSCUST for an exclusive discount.
Power BI
Uncover powerful insights and turn them into impact
Connect to and visualize any data, and seamlessly infuse visuals into the apps you use every day.
Get Started
Have an account? Sign in

Looker Studio

(Former Google Data Studio)



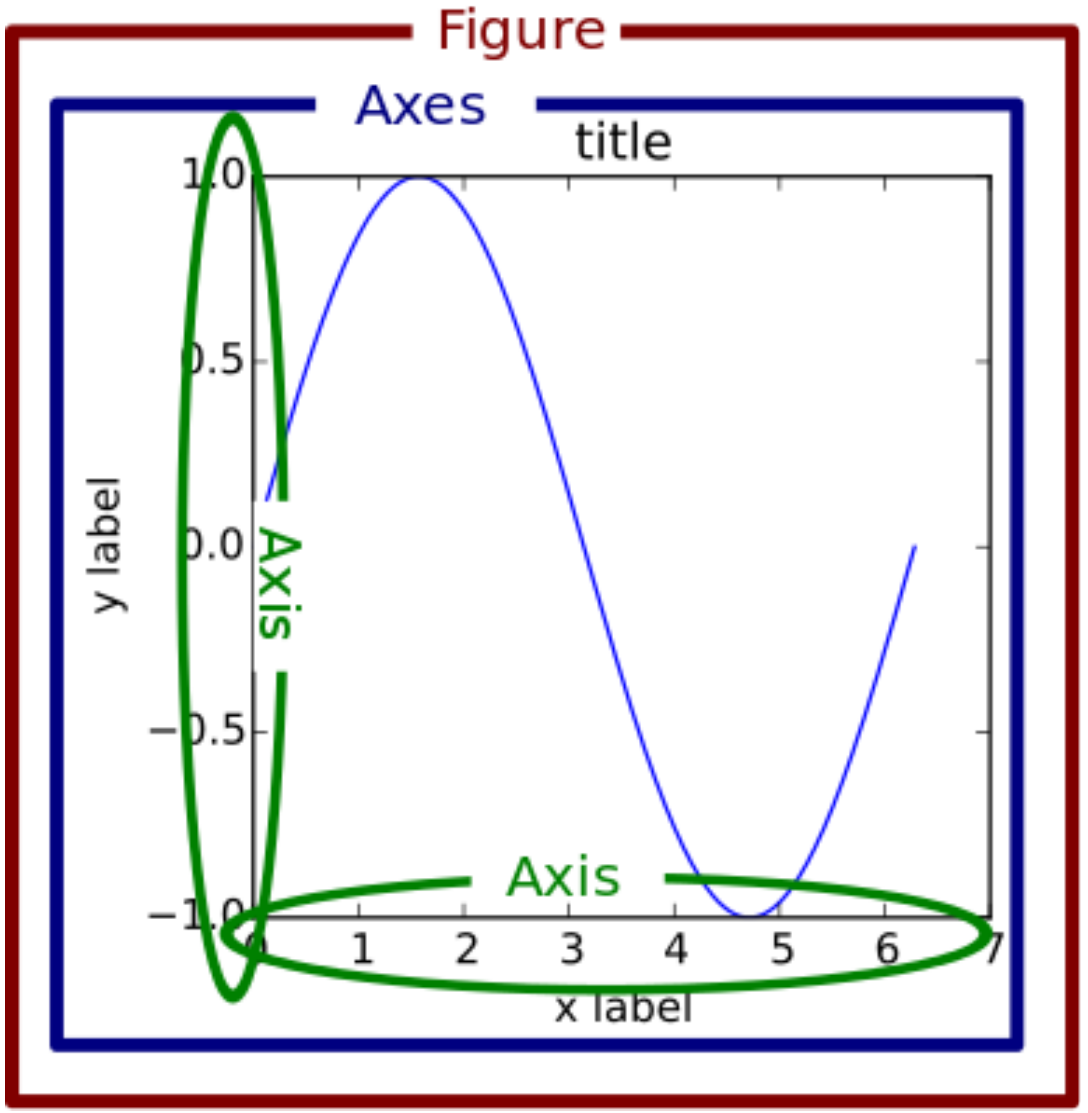
Apache Superset



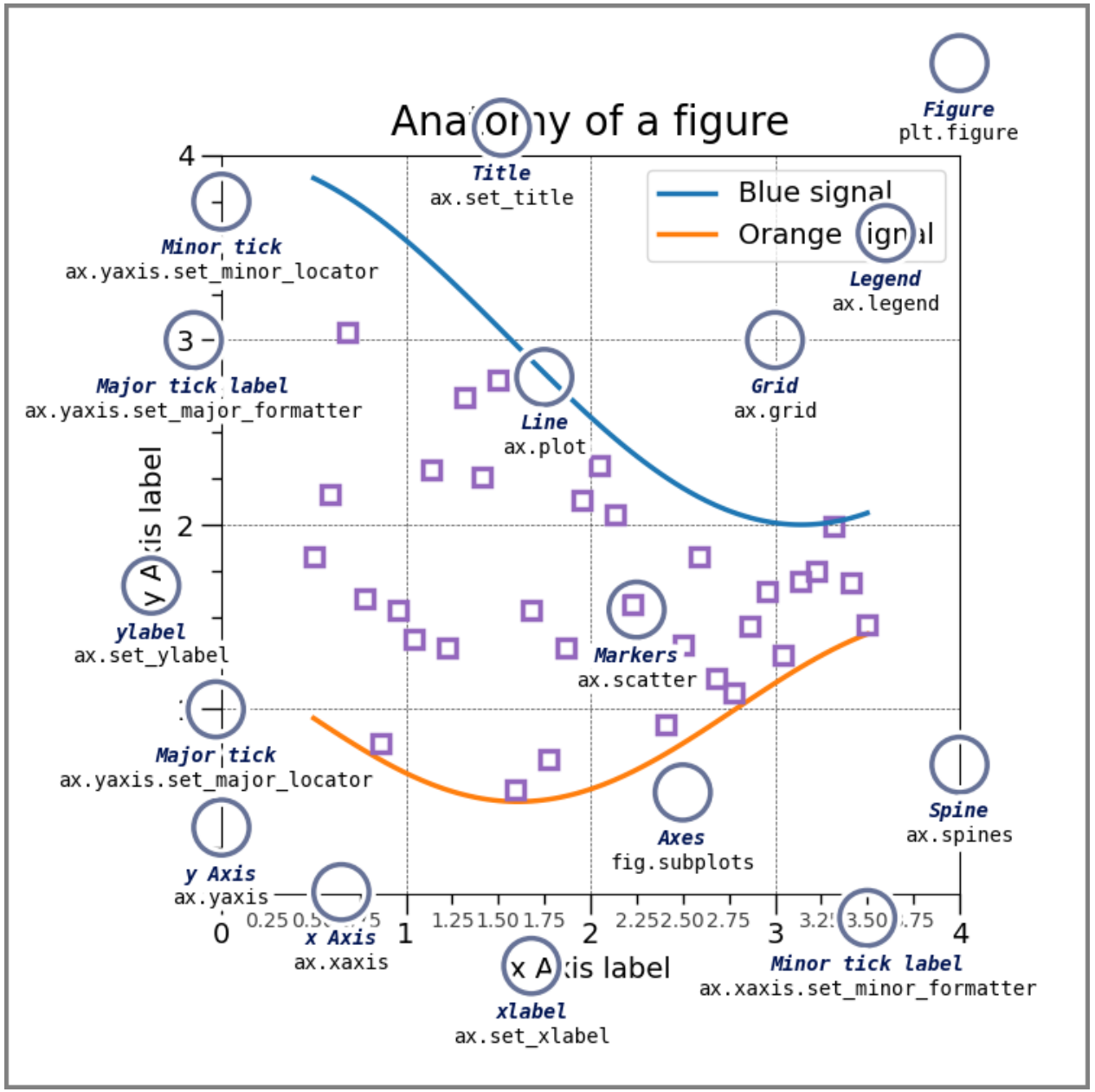
Popular charting libraries

- Matplotlib
- Seaborn
- Plotly
- Bokeh
- D3.js

Matplotlib Basics

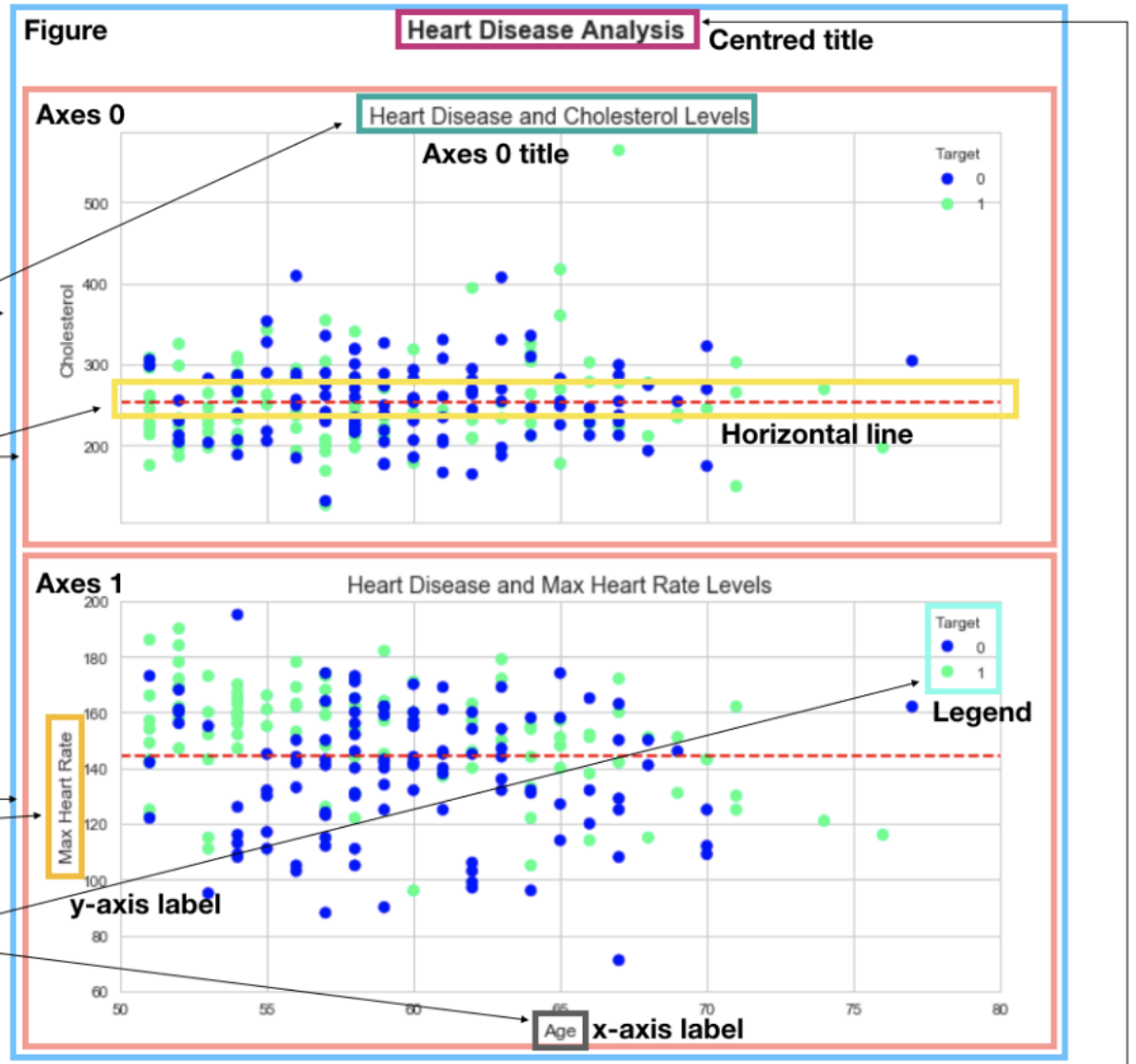


Anatomy of a Matplotlib plot



Anatomy of a Matplotlib plot

```
%matplotlib inline
import pandas as pd
import matplotlib.pyplot as plt
plt.style.use('seaborn-whitegrid') # set plot style

# Read in and manipulate data
heart_disease = pd.read_csv('../data/heart-disease.csv')
over_50 = heart_disease[heart_disease['age'] > 50]

# Create figure (plot) with 2 axes
fig, (ax0, ax1) = plt.subplots(nrows=2,
                               ncols=1,
                               sharex=True,
                               figsize=(10, 10))

# Add data, titles, meanline (axhline) and legend to axes 0
scatter = ax0.scatter(over_50["age"],
                      over_50["chol"],
                      c=over_50["target"],
                      cmap='winter')
ax0.set(title="Heart Disease and Cholesterol Levels",
        ylabel="Cholesterol",
        xlim=[50, 80])
ax0.axhline(y=over_50["chol"].mean(),
            color='r',
            linestyle='--',
            label="Average");
ax0.legend(*scatter.legend_elements(), title="Target")

# Add data, titles, meanline (axhline) and legend to axes 1
scatter = ax1.scatter(over_50["age"],
                      over_50["thalach"],
                      c=over_50["target"],
                      cmap='winter')
ax1.set(title="Heart Disease and Max Heart Rate Levels",
        xlabel="Age",
        ylabel="Max Heart Rate",
        ylim=[60, 200])
ax1.axhline(y=over_50["thalach"].mean(),
            color='r',
            linestyle='--',
            label="Average");
ax1.legend(*scatter.legend_elements(), title="Target")

# Title the figure
fig.suptitle('Heart Disease Analysis', fontsize=16, fontweight='bold');
```



References

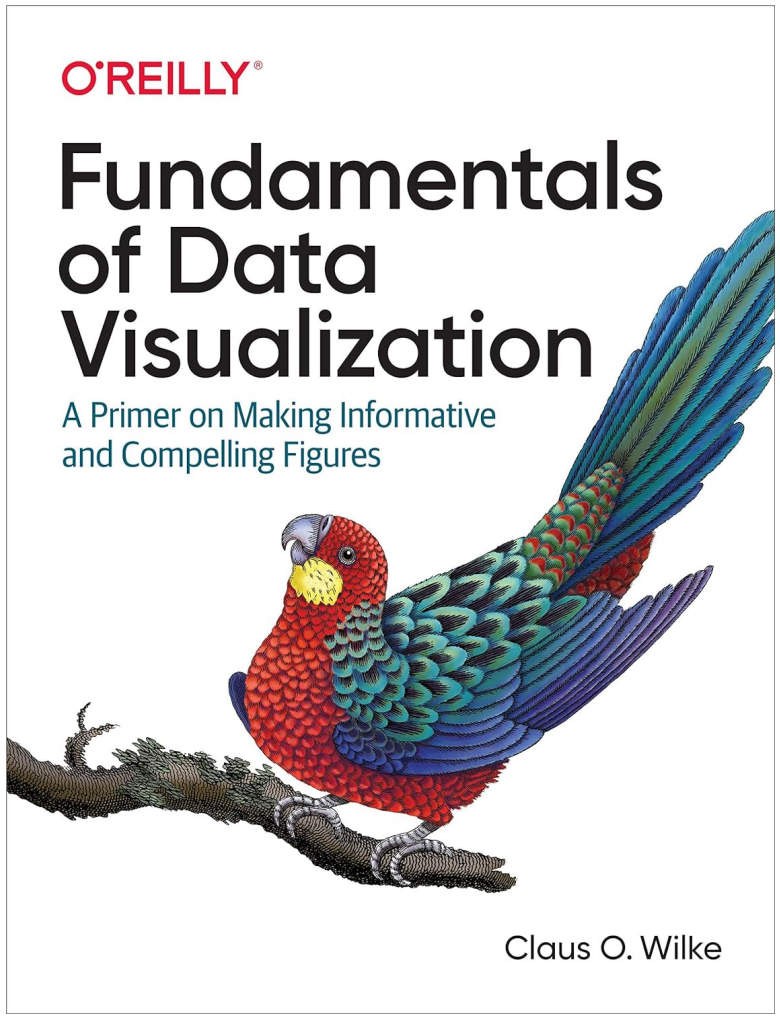
https://clauswilke.com/dataviz/ (Book)

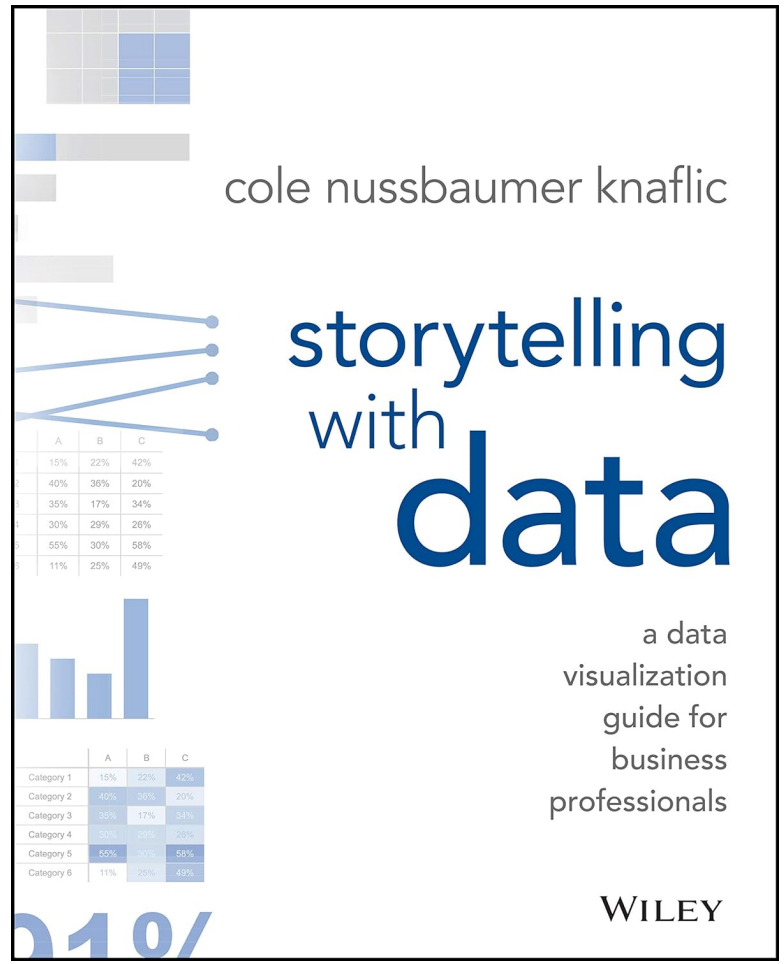
https://www.storytellingwithdata.com/books

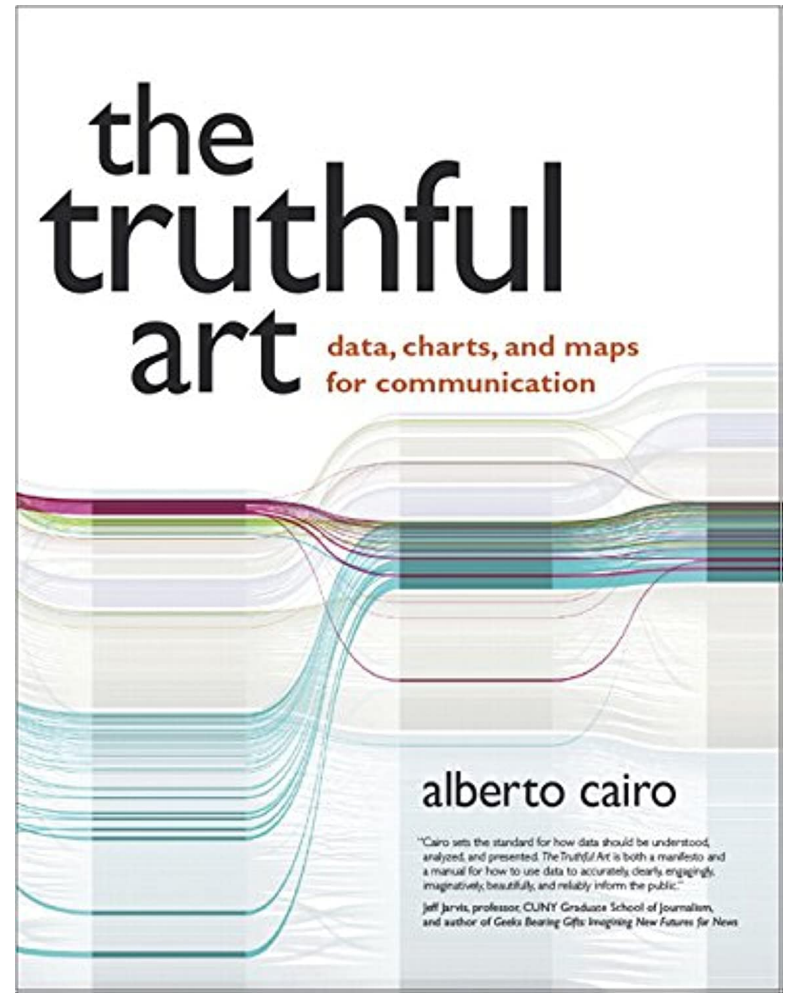
https://datavizcatalogue.com/

https://informationisbeautiful.net/

https://coolinfographics.com/dataviz-guides/


O'REILLY®
Fundamentals of Data Visualization
A Primer on Making Informative and Compelling Figures
Claus O. Wilke


cole nussbaumer knaflic
storytelling with data
a data visualization guide for business professionals
WILEY


the truthful art
data, charts, and maps for communication
alberto cairo
"Cairo sets the standard for how data should be understood, analyzed, and presented. The Truthful Art is both a manifesto and a manual for how to use data to accurately, clearly, engagingly, imaginatively, beautifully, and reliably inform the public."
Jeff Jarvis, professor, CUNY Graduate School of Journalism, and author of Geeks Bearing Gifts: Imagining New Futures for News

Thank You

Happy Learning 🚀 !!

meftaulhaque@gmail.com