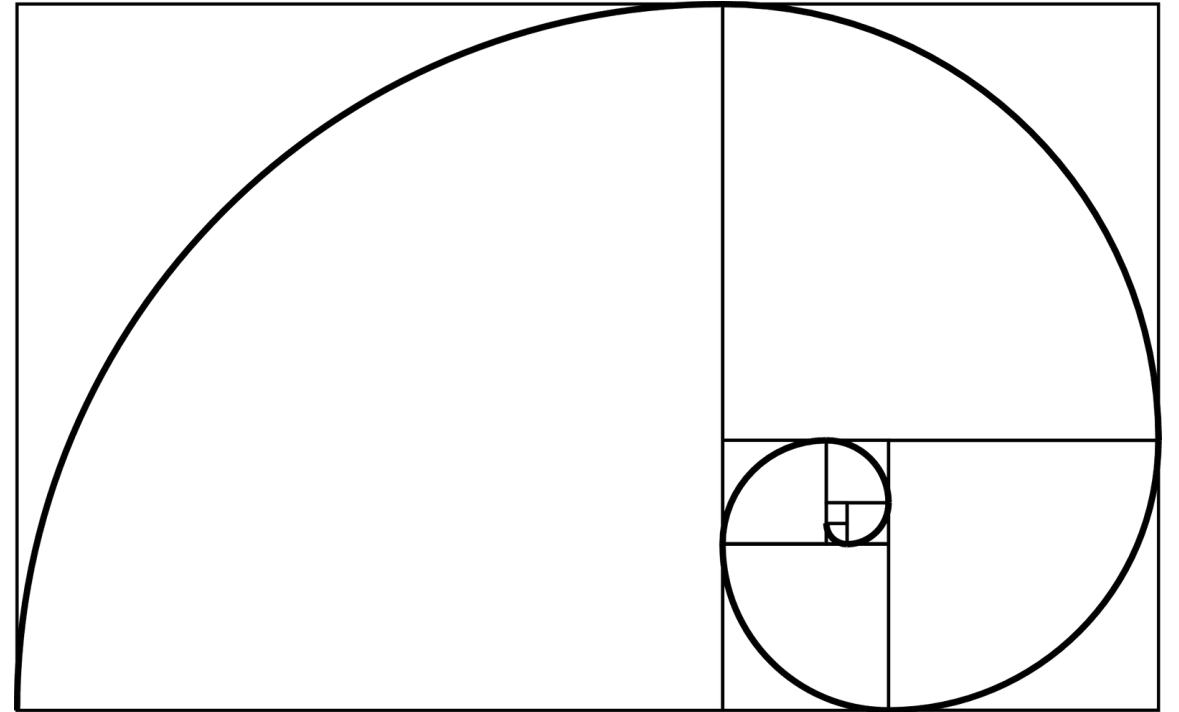


Designing **Effective** Data Visualizations

Pete Lawson, Ph.D.

Data and Visualization Librarian
Data Services
Johns Hopkins University




Data visualization

What is it?

Data visualization is the **creation** and **study** of the visual representation of data

Data visualization involves converting our data sources into visual representations.

These might be charts, maps, graphs



“The simple graph has brought more information to the data analyst’s mind than any other device”

- John Tukey

What makes an effective data visualization?

- It has a clear purpose and message
 - Complements and enhances the text

What makes an effective data visualization?

- It has a clear purpose and message
 - Complements and enhances the text
- It is easy to interpret

“A good figure is like a joke, if you have to explain it, it’s not that good.”

What makes an effective data visualization?

- It has a clear purpose and message
 - Complements and enhances the text
- It is easy to interpret

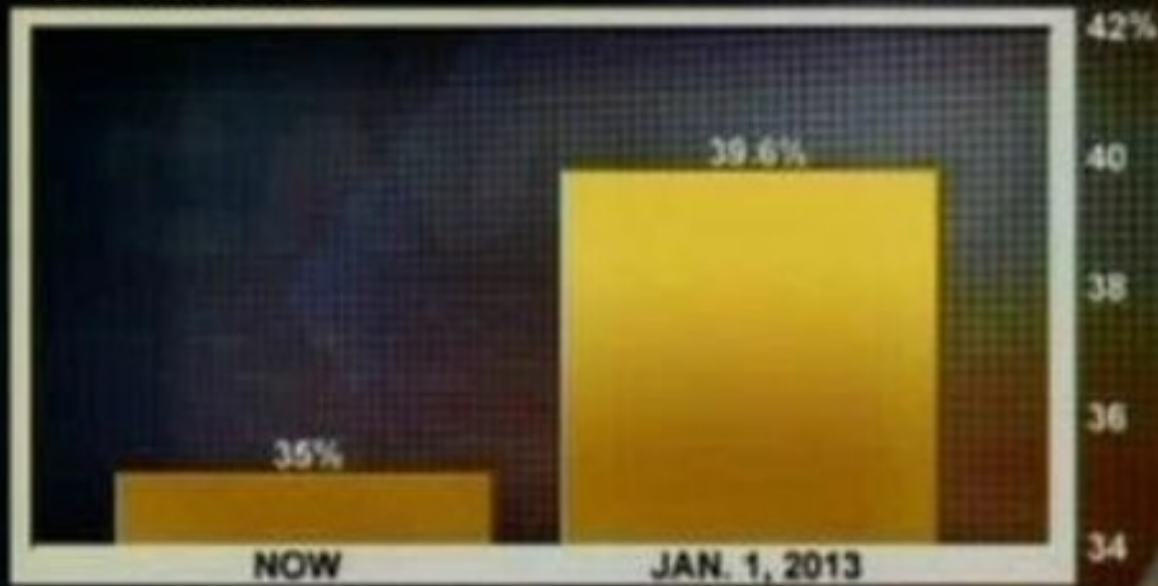
“A good figure is like a joke, if you have to explain it, it’s not that good.”

- It accurately reflects the data

What makes a **bad** data visualization?

IF BUSH TAX CUTS EXPIRE

TOP TAX RATE



8:01 p ET

FOX
BUSINESS

TOP STORIES

TECHNOLOGY

CONSUMER

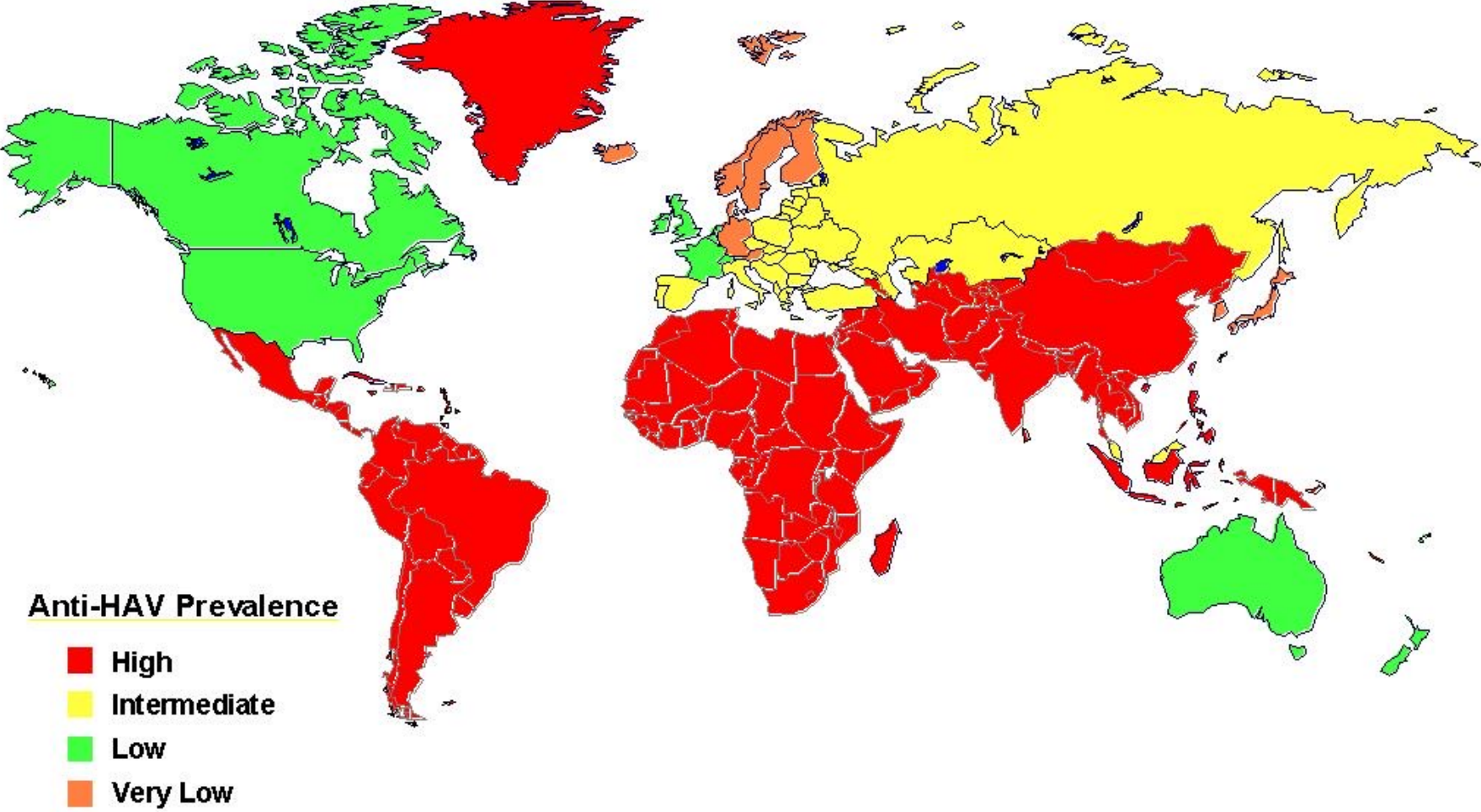
WITH THE JUSTICE DEPARTMENT AND ACQUIRES FULL T

DOW 13008.68 \downarrow 64.33

S&P 1379.32 \downarrow 5.98

NASDAQ 2939.52 \downarrow 6.32

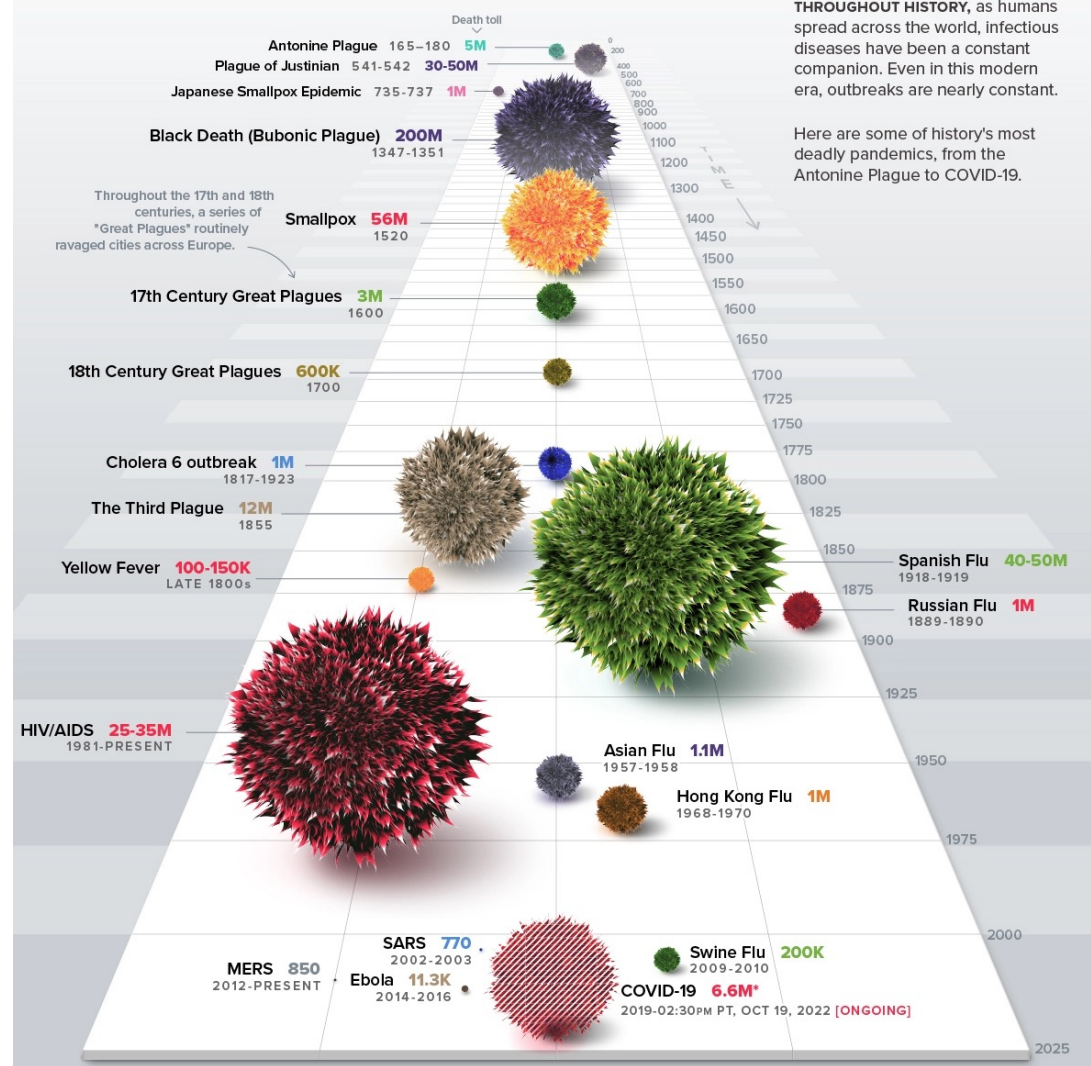
Geographic Distribution of HAV Infection



Source: <https://www.vislies.org/2021/gallery/>

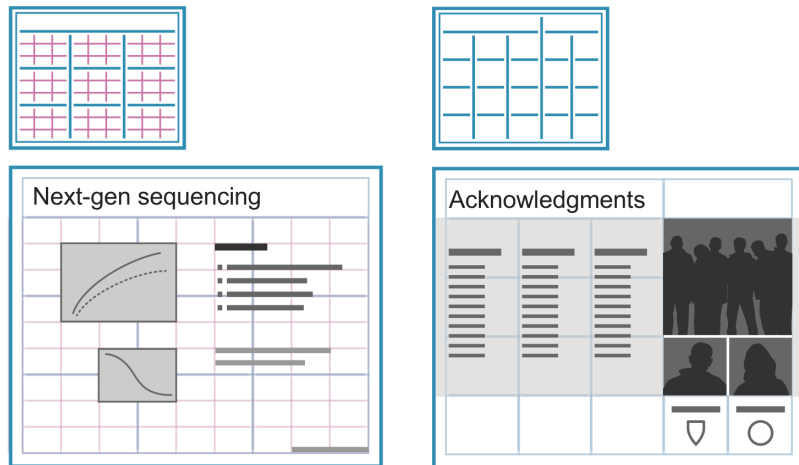
HISTORY OF PANDEMICS

PAN-DEM-IC (of a disease) prevalent over a whole country or the world.



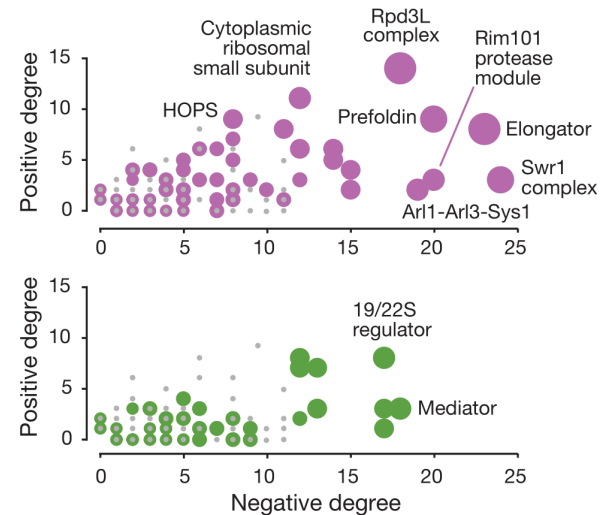
Introduction

design of data visualizations



Graphic design

Communicating ideas in an **engaging** and **appealing** way



Information Design

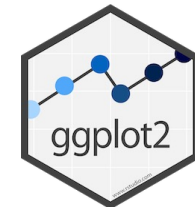
Communicating information **effectively** to facilitate **comprehension**

What this workshop doesn't cover

How to create visualizations in specific programs

What this workshop doesn't cover

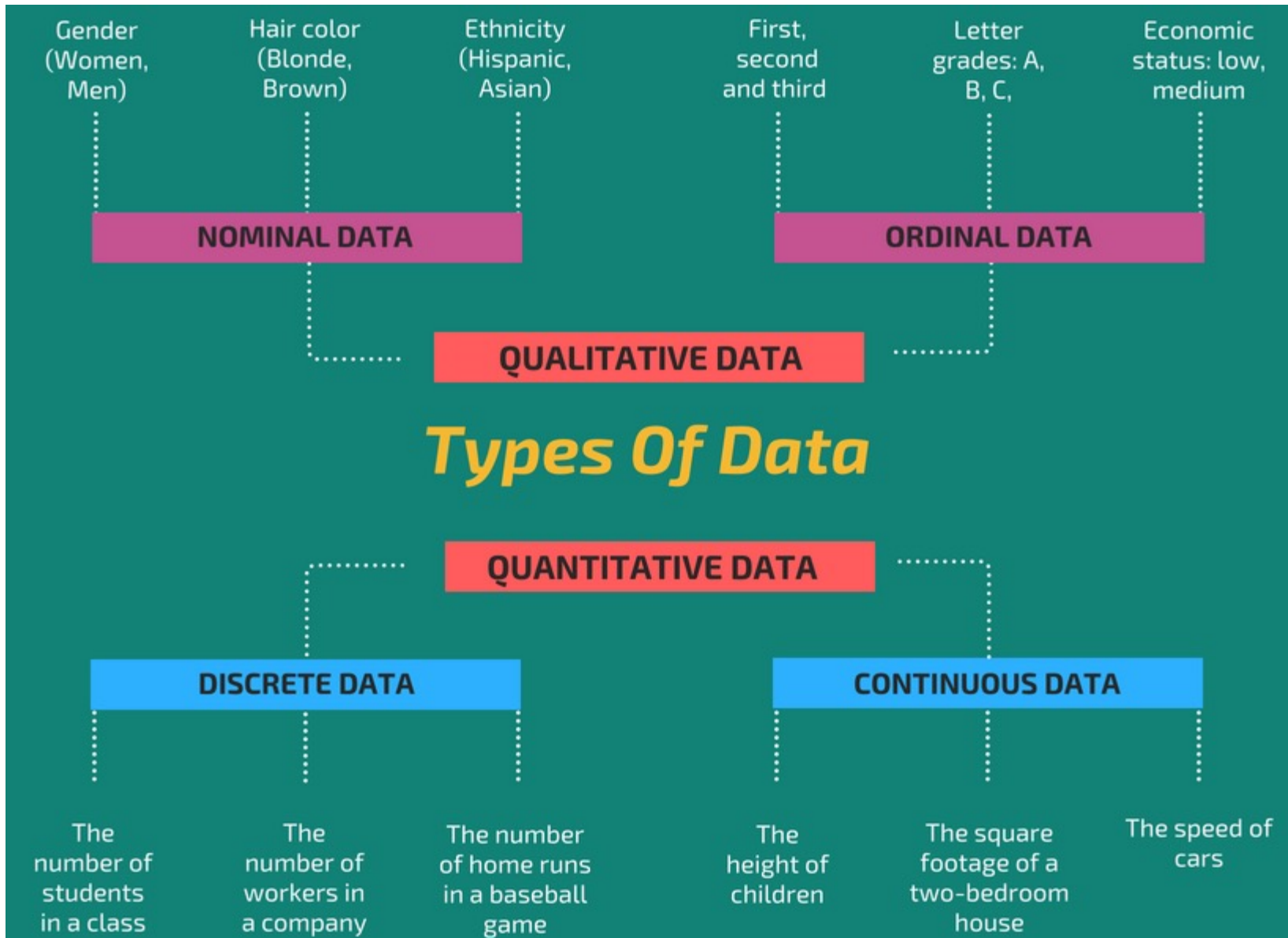
How to create visualizations in specific programs

The logo for STATA, featuring the word "STATA" in a bold, blue, sans-serif font.The logo for IBM SPSS, with "IBM" in its characteristic eight horizontal stripes and "SPSS" in a bold, black, sans-serif font below it.The logo for SAS, consisting of a blue stylized "S" icon followed by the word "sas" in a lowercase, black, sans-serif font.The logo for Prism, featuring a colorful triangular prism icon followed by the word "Prism" in a bold, black, sans-serif font.The logo for Python, showing the two interlocking snakes (one blue, one yellow) followed by the word "python" in a lowercase, grey, sans-serif font.The logo for Microsoft Excel, featuring a green square with a white "X" and a grid pattern, followed by the word "Excel" in a green, sans-serif font.The logo for the R programming language, featuring a blue "R" inside a grey circle.The logo for Julia, featuring the word "julia" in a lowercase, black, sans-serif font with four colored dots (blue, red, green, purple) above the letters.The logo for MATLAB, featuring a colorful 3D surface plot icon followed by the word "MATLAB" in a black, serif font.

Introduction

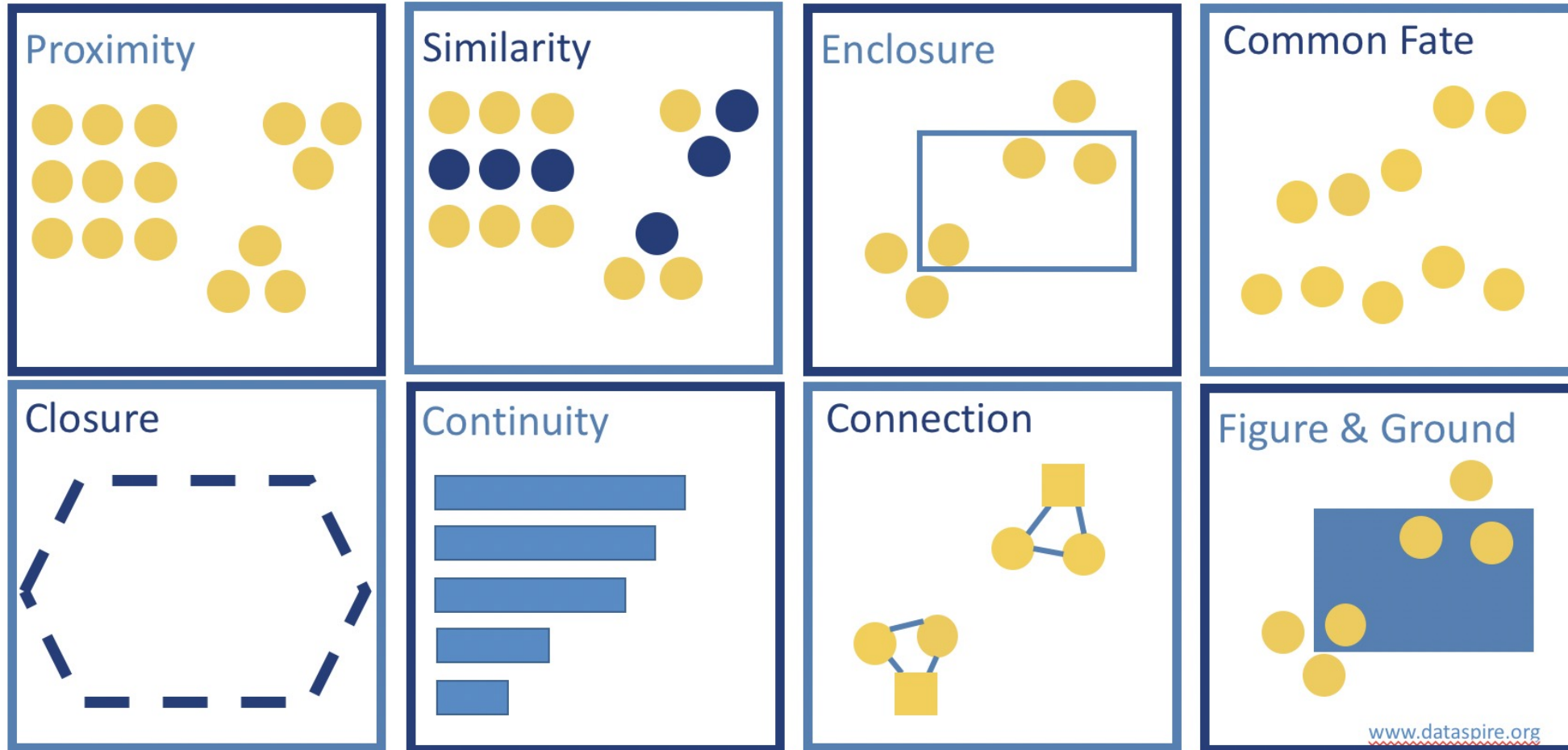
Before you begin

- **Understand your audience.**
 - What is their expertise?
 - What is your expertise?
 - Is there a mismatch between your expertise and your audiences?
- **What is the purpose of your visualization?**
 - Are you trying to communicate some pattern?
 - Are you exploring data to discover underlying trends?
 - What is the story you are trying to tell?



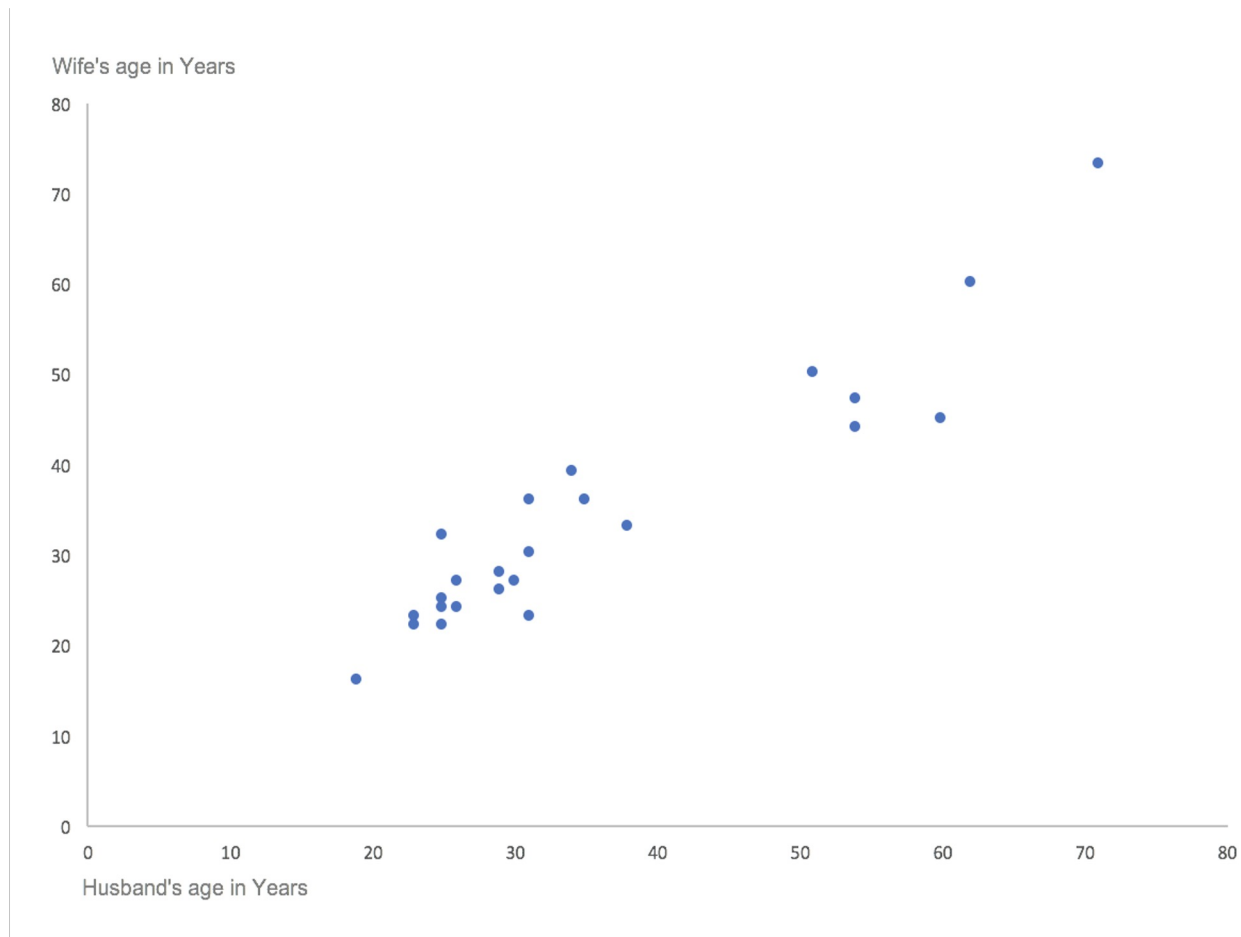
Principles of data visualization

Gestalt's Principles of Visual Perception



Law of Similarity

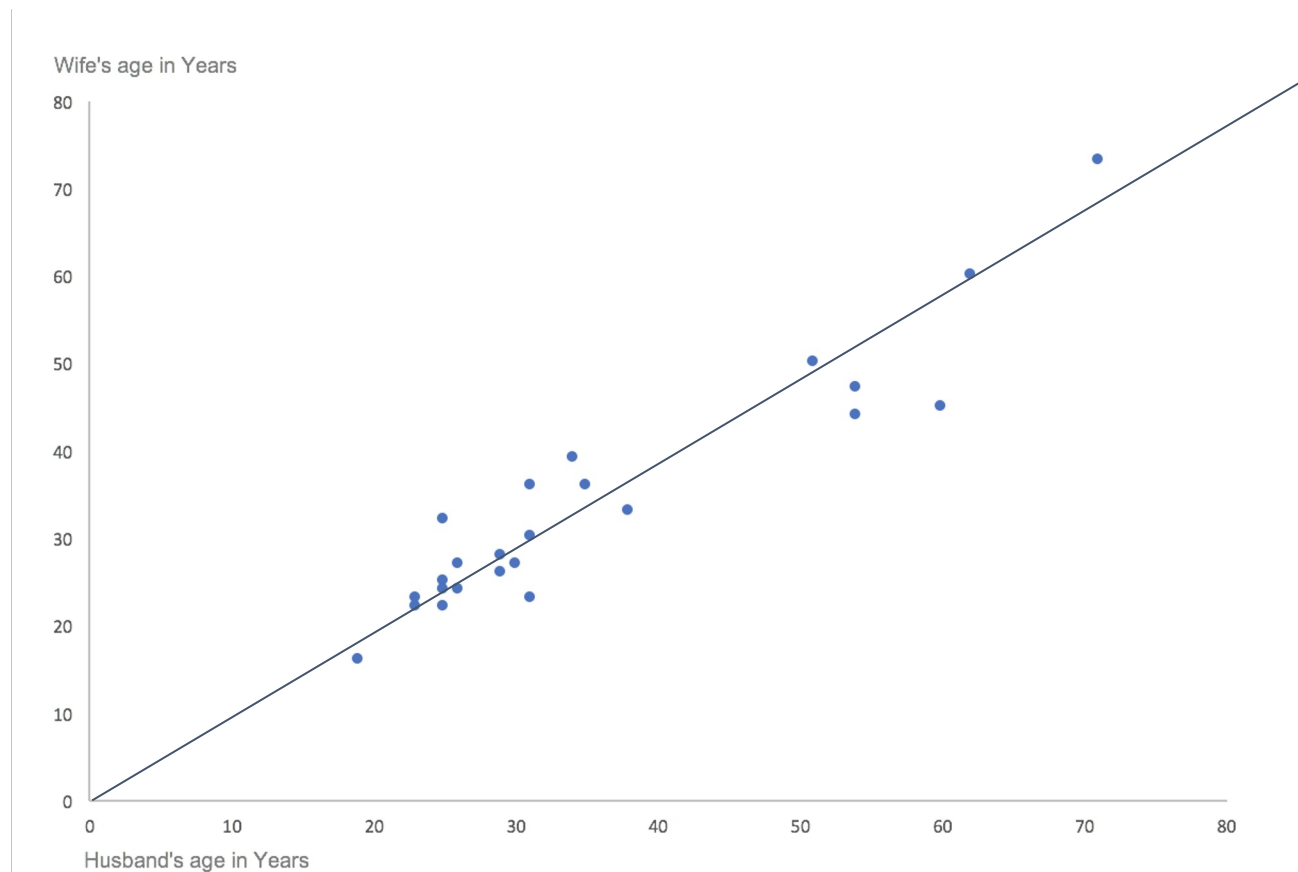
Our brains tend to make connections between elements with a similar design



Daydreaming Numbers, "[Gestalt Laws Applied to Data Visualization](#)"

Law of Similarity

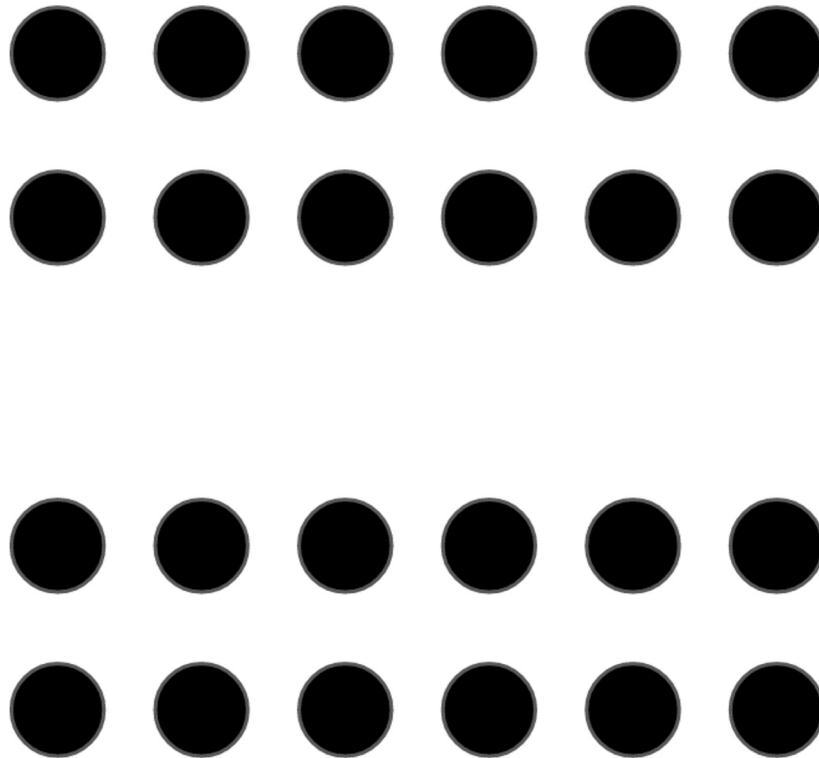
Our brains tend to make connections between elements with a similar design



Daydreaming Numbers, "[Gestalt Laws Applied to Data Visualization](#)"

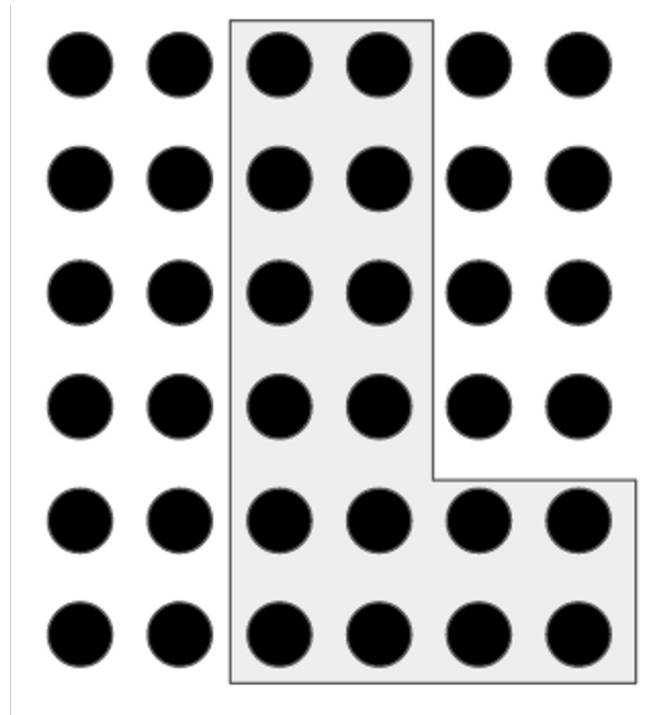
Law of Proximity

Elements that are closer together in a design are often linked in our minds



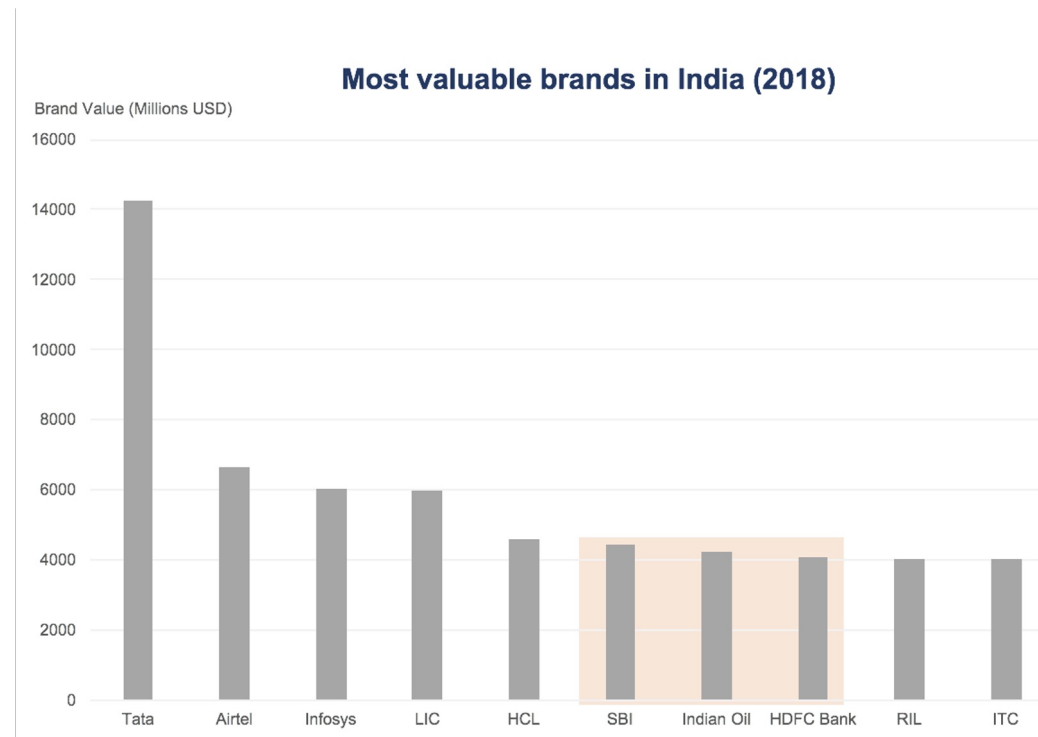
Law of Enclosure

Objects that appear to have a boundary around them are perceived as a group.



Law of Enclosure

Objects that appear to have a boundary around them are perceived as a group.



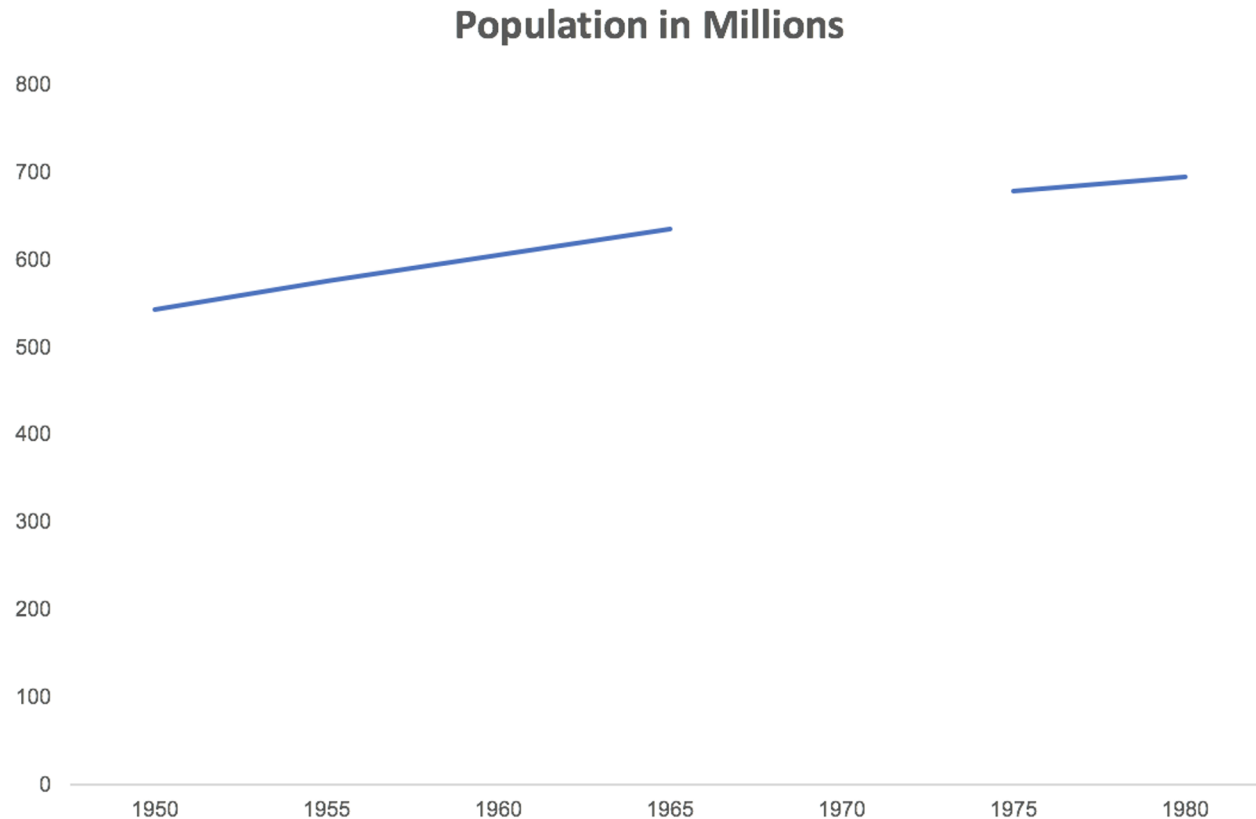
Law of Closure

Open structures are perceived as closed, complete, and regular whenever there is a way that they can be reasonably interpreted as such.



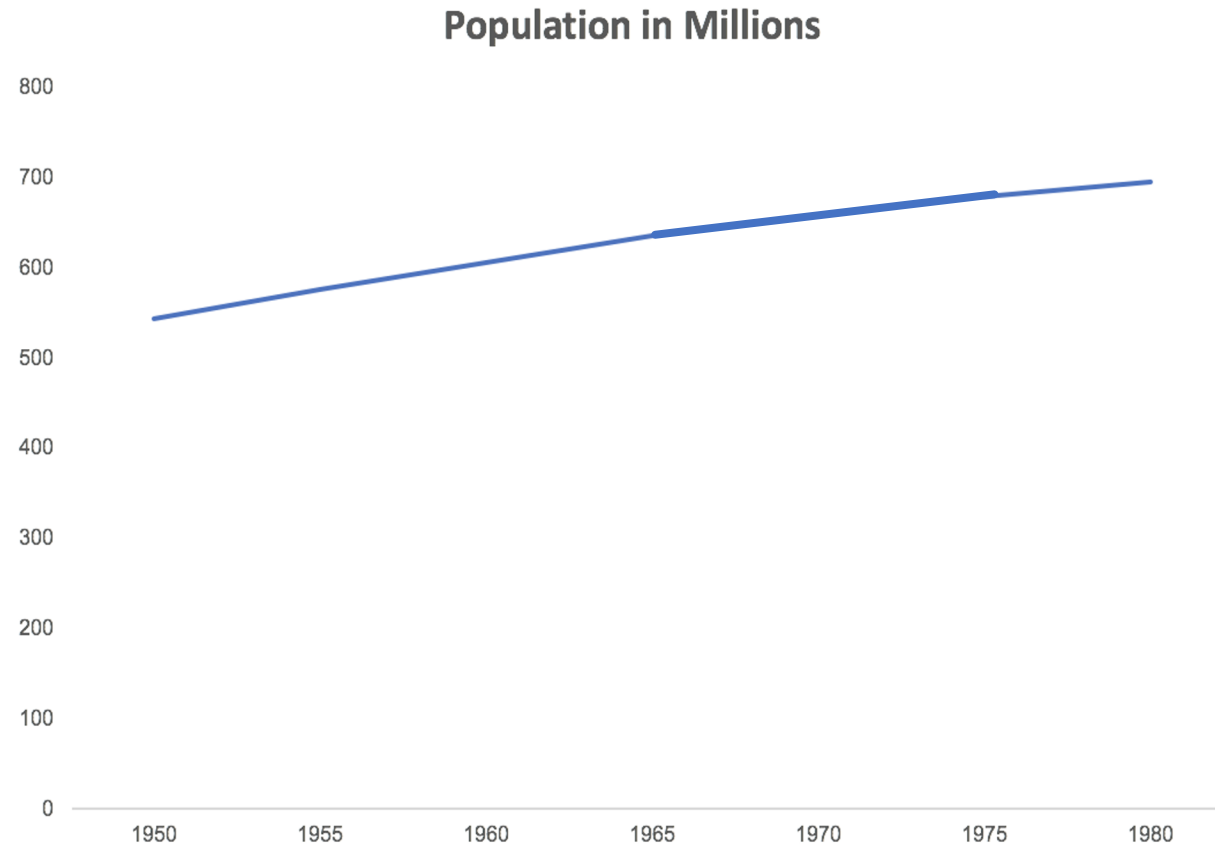
Law of Closure

Open structures are perceived as closed, complete, and regular whenever there is a way that they can be reasonably interpreted as such.



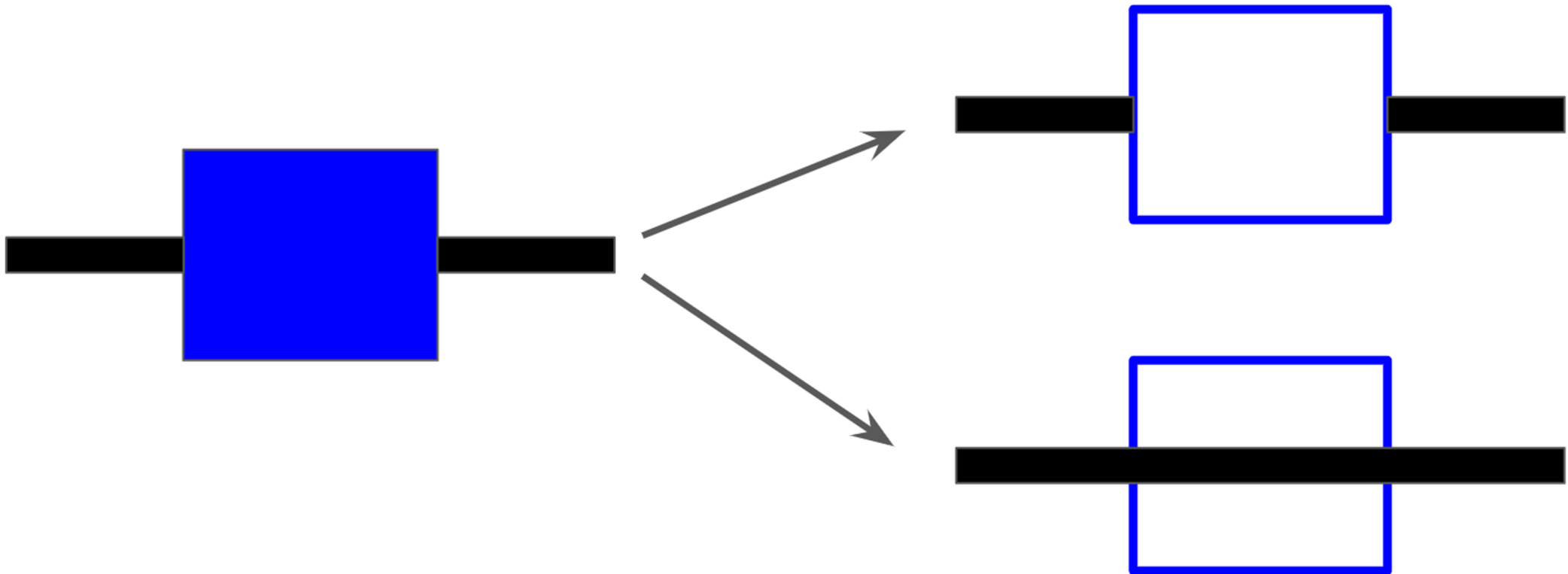
Law of Closure

Open structures are perceived as closed, complete, and regular whenever there is a way that they can be reasonably interpreted as such.



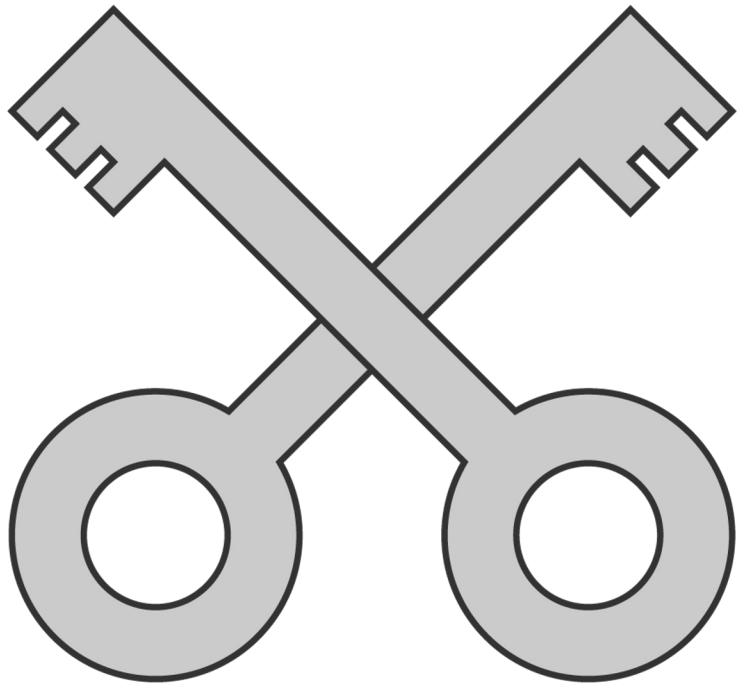
Law of Continuity

Objects that are aligned together or appear to be a continuation of one another are perceived as a group.

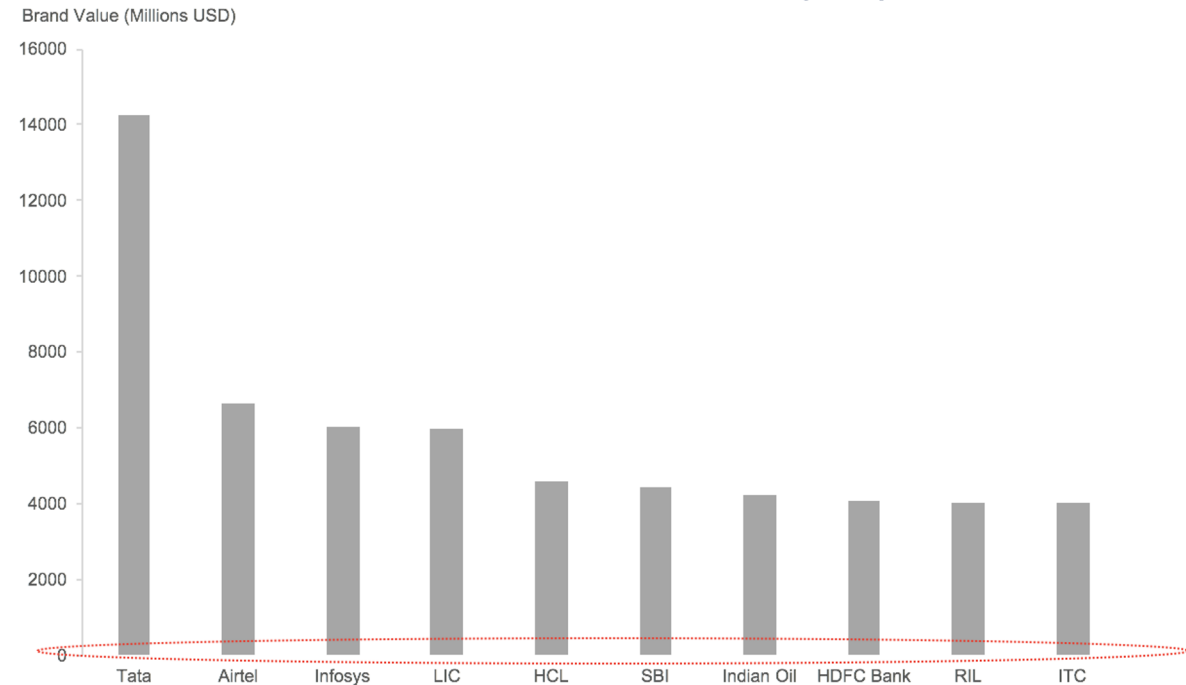


Law of Continuity

Objects that are aligned together or appear to be a continuation of one another are perceived as a group.

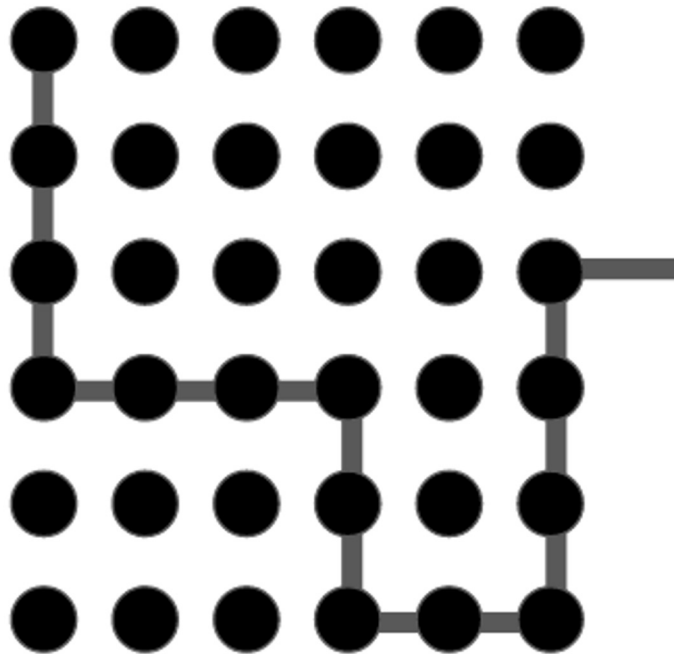


Most valuable brands in India (2018)



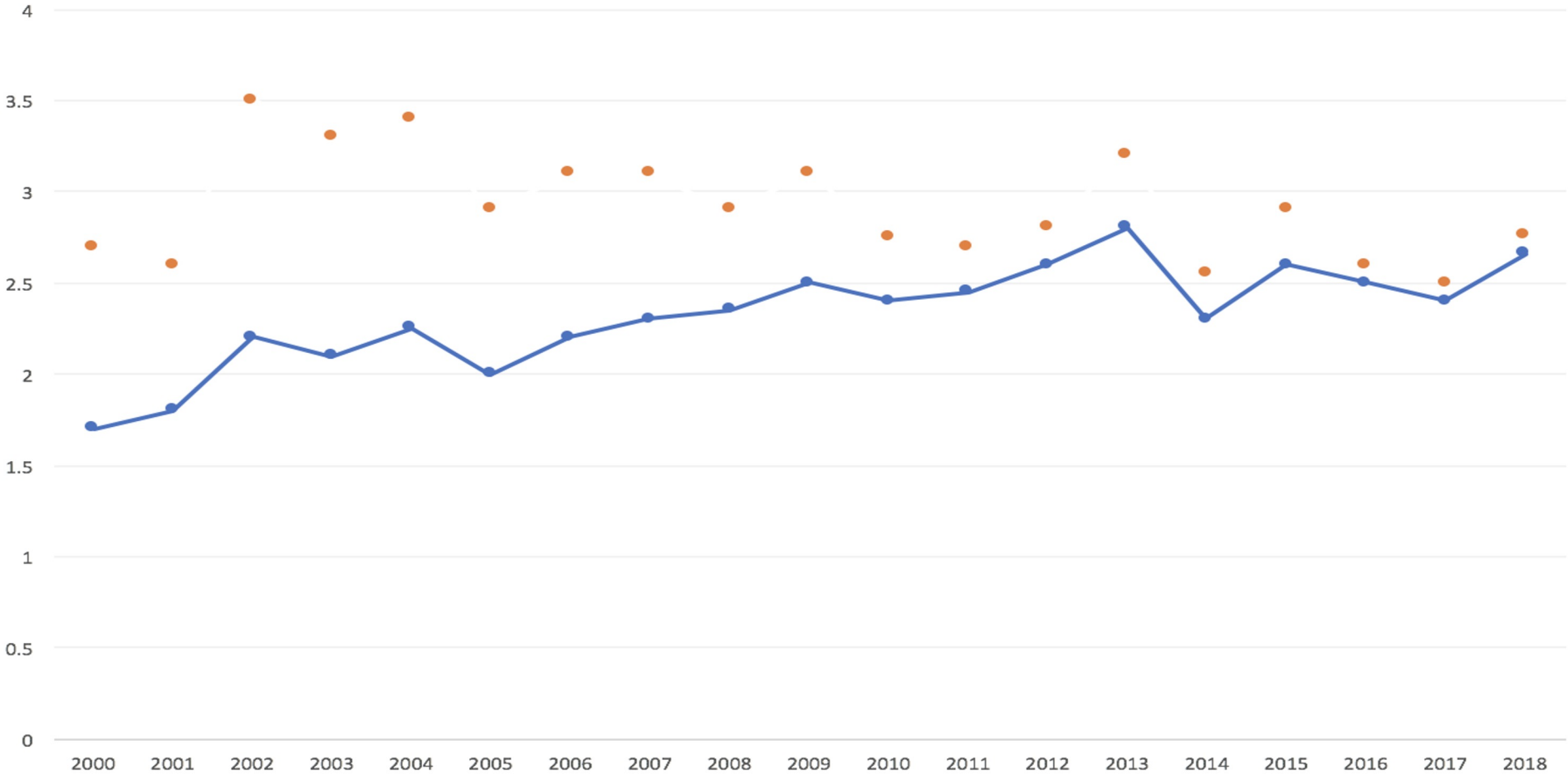
Law of Connection

Objects that are connected, usually by a line, are perceived as a group.



Law of Connection

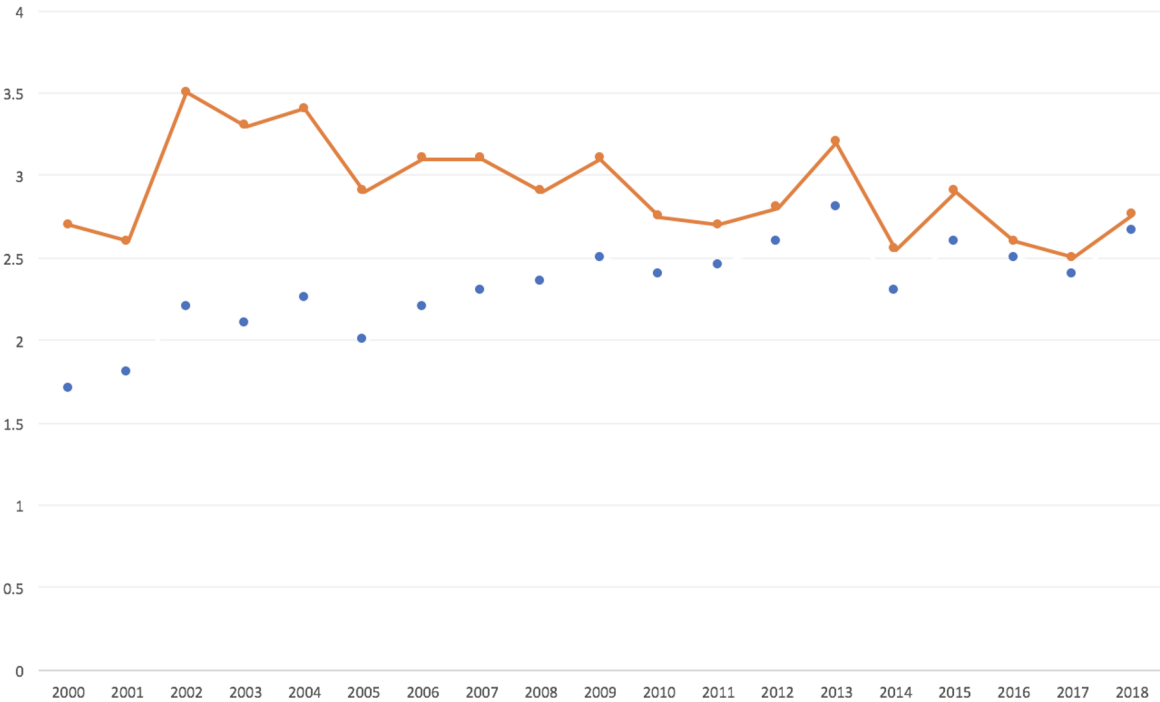
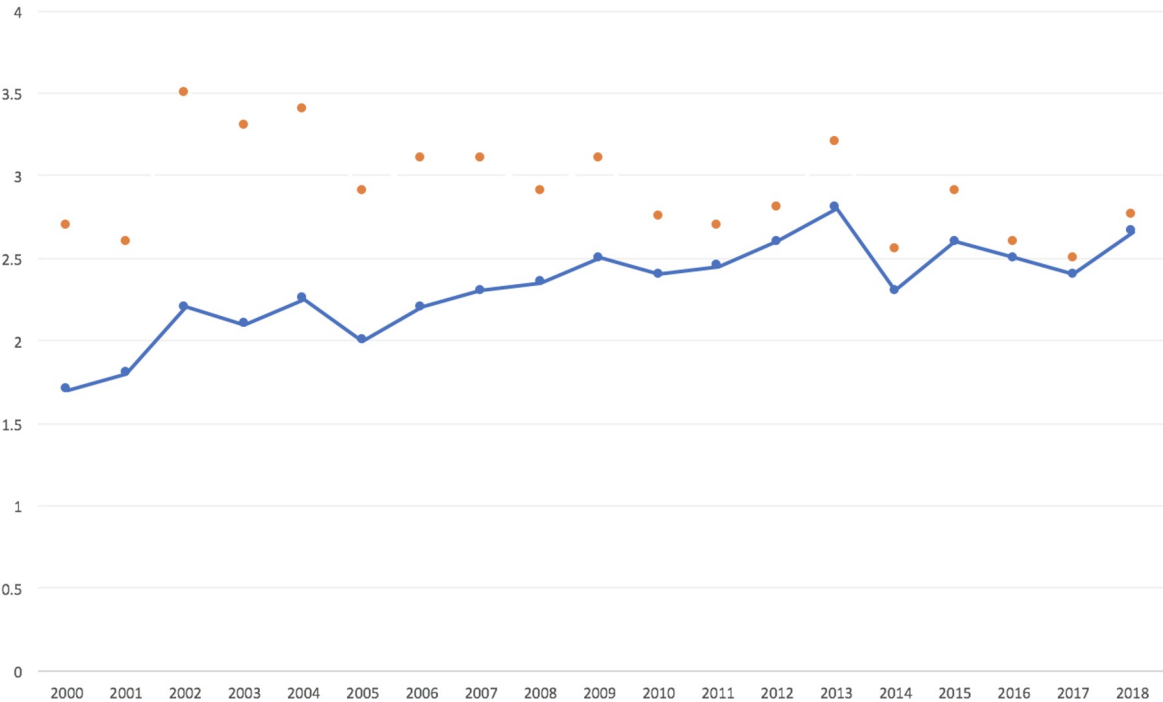
Objects that are connected, usually by a line, are perceived as a group.



Source: <https://visualizingthefuture.github.io/data-viz-101/>

Law of Connection

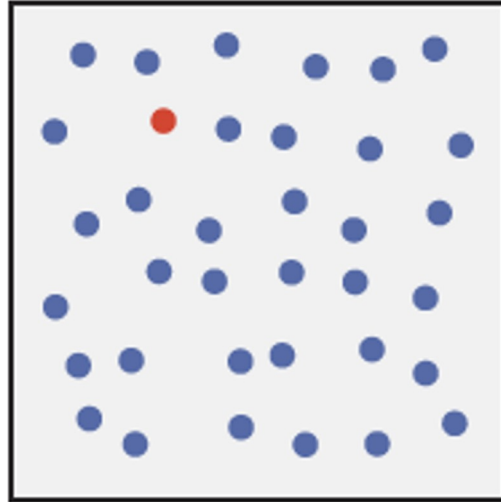
Objects that are connected, usually by a line, are perceived as a group.



173658103837575063348181736401
016254539319123938525616173943
987139874619319586716628309897
273164613984019358094285976205
897629835921873589321759871059
283198254781237598698127359812

173658103837575063348181736401
016254539319123938525616173943
987139874619319586716628309897
273164613984019358094285976205
897629835921873589321759871059
283198254781237598698127359812

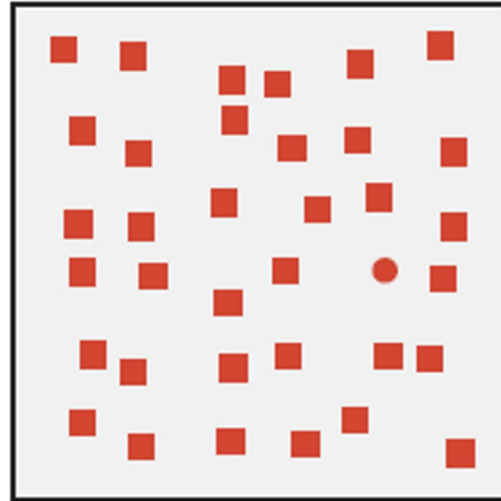
Pre-attentive Processing: Color



Healey, C. G., & Enns, J. T. (2012). Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7), 1170-1188.

<http://dx.doi.org.proxy.lib.duke.edu/10.1109/TVCG.2011.127>

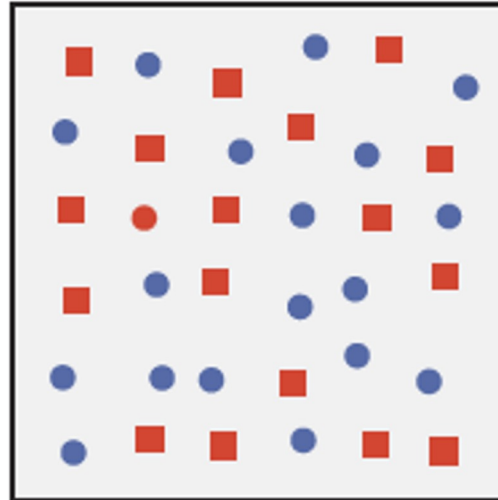
Pre-attentive Processing: Shape



Healey, C. G., & Enns, J. T. (2012). Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7), 1170-1188.

<http://dx.doi.org.proxy.lib.duke.edu/10.1109/TVCG.2011.127>

Pre-attentive Processing: Combined



Healey, C. G., & Enns, J. T. (2012). Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7), 1170-1188.

<http://dx.doi.org.proxy.lib.duke.edu/10.1109/TVCG.2011.127>

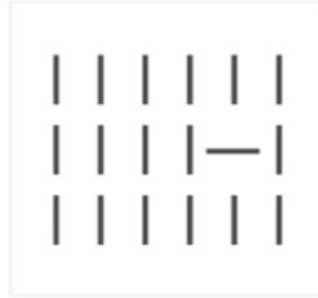
Pre-attentive attributes



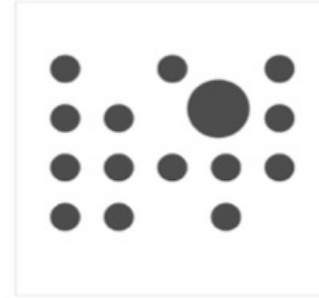
Length



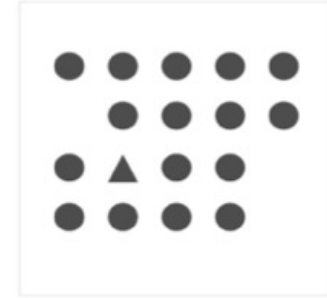
Width



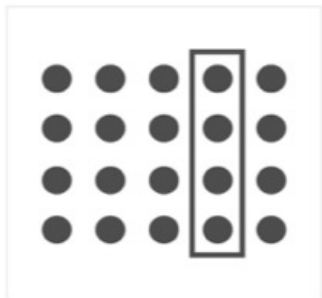
Orientation



Size



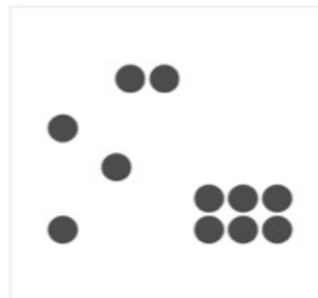
Shape



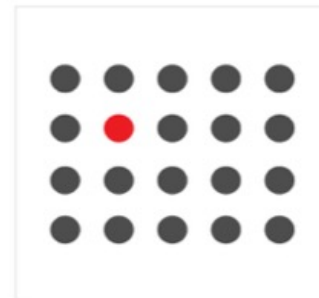
Enclosure



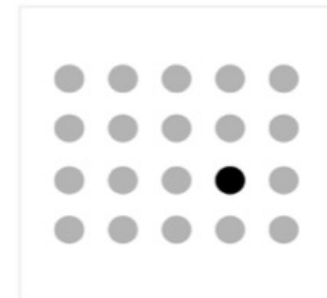
Position



Grouping



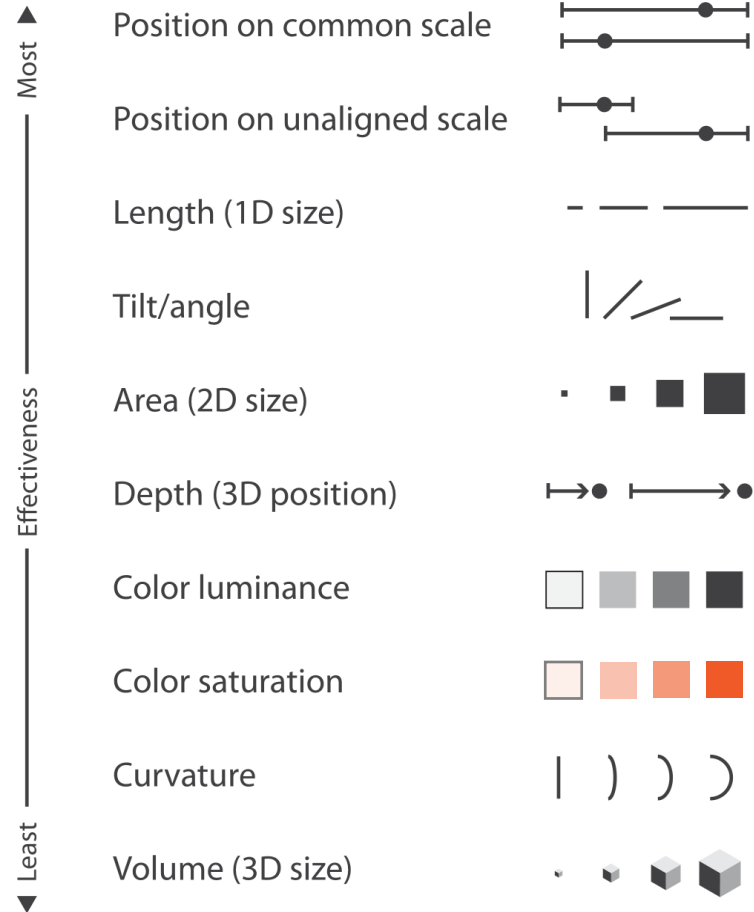
Color Hue



Color Intensity

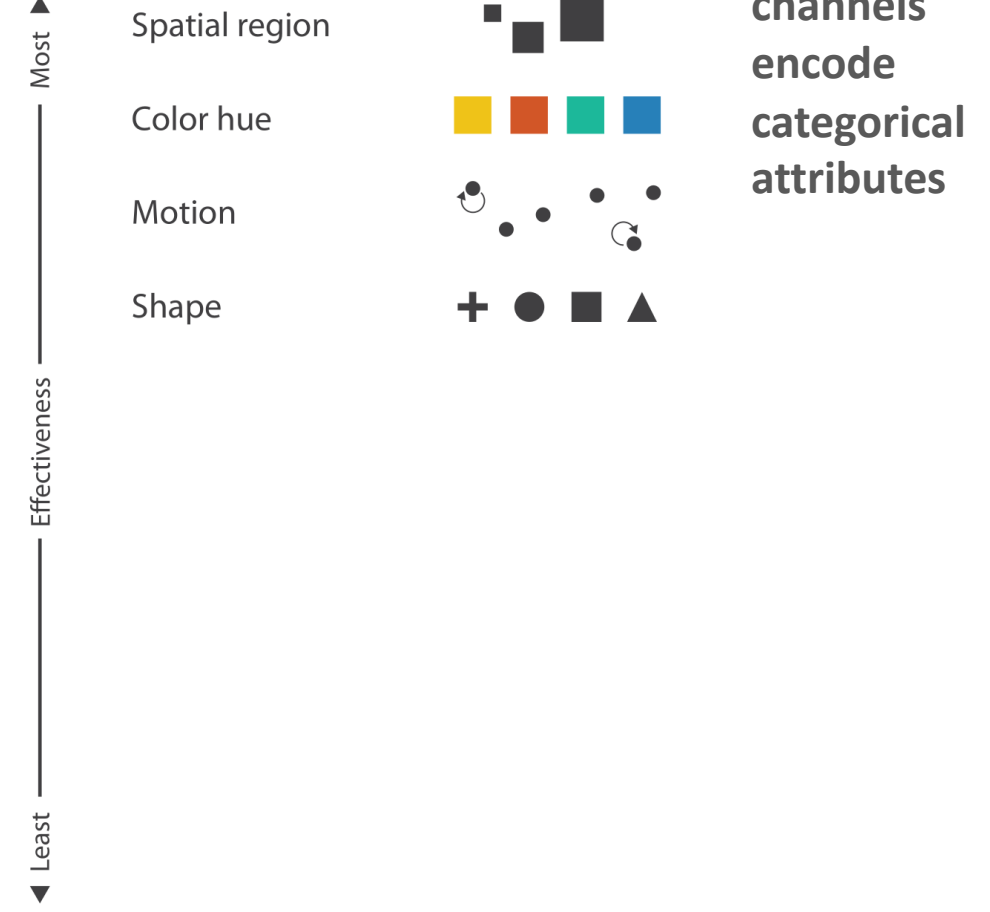
Ranking the effectiveness of visual encodings

Magnitude Channels



Magnitude channels encode ordered attributes

Identity Channels



Identity channels encode categorical attributes

Principles of visual encoding

Expressiveness principle

Visual encodings should express all of, and only, the information in the dataset attributes.

Effectiveness principle

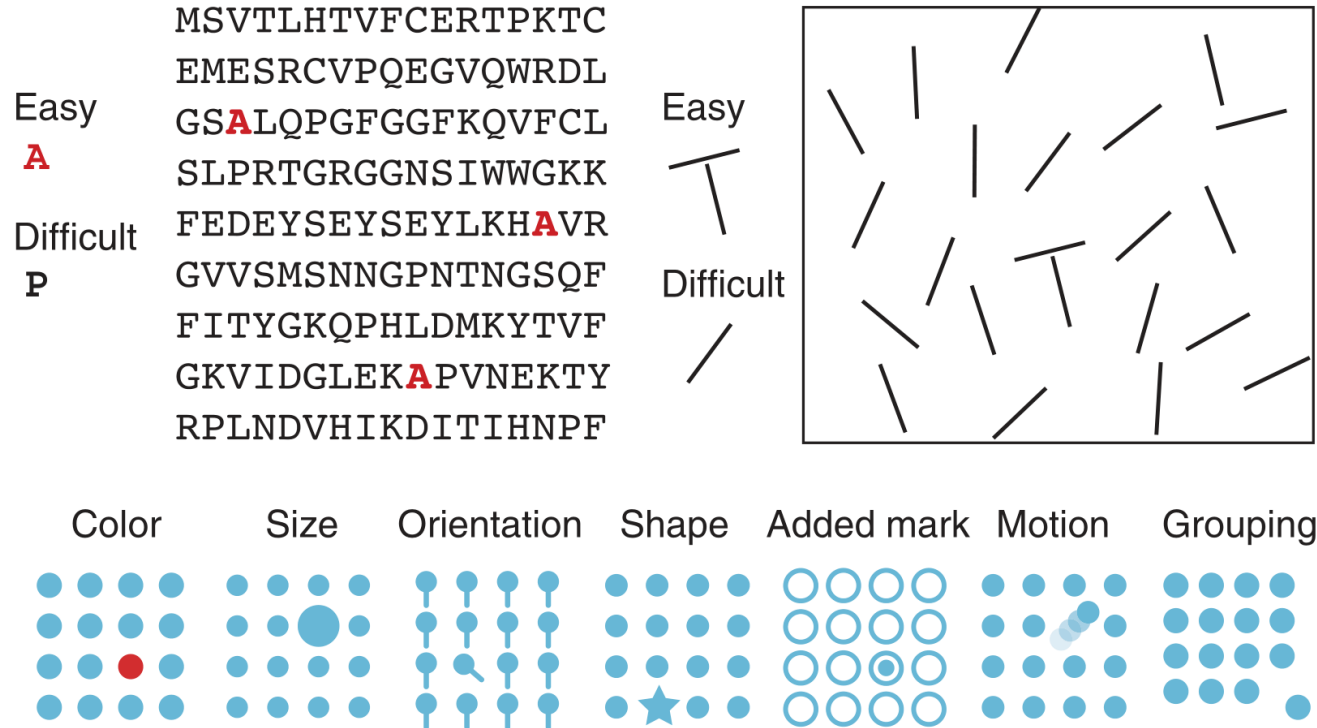
The importance of the attribute should match the **salience** of the channel.

What is encoded

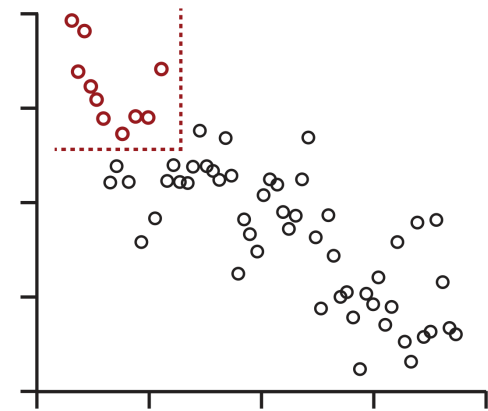
must be decoded

Saliency

Set an object apart from its surroundings to create contrast



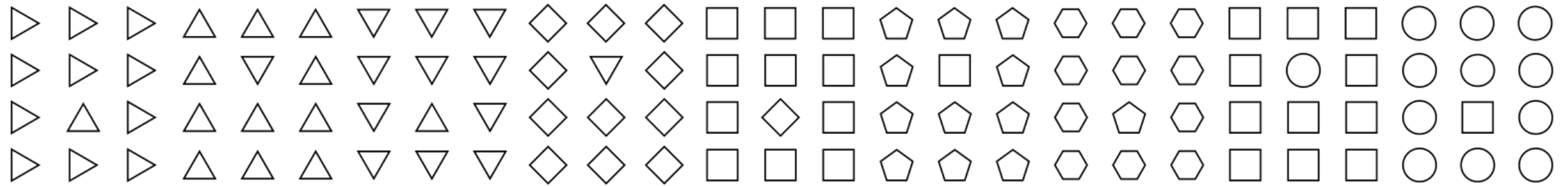
Color name	RGB (1-255)
Black	0, 0, 0
Orange	230, 159, 0
Sky blue	86, 180, 233
Bluish green	0, 158, 115
Blue	0, 114, 178
Vermillion	213, 94, 0



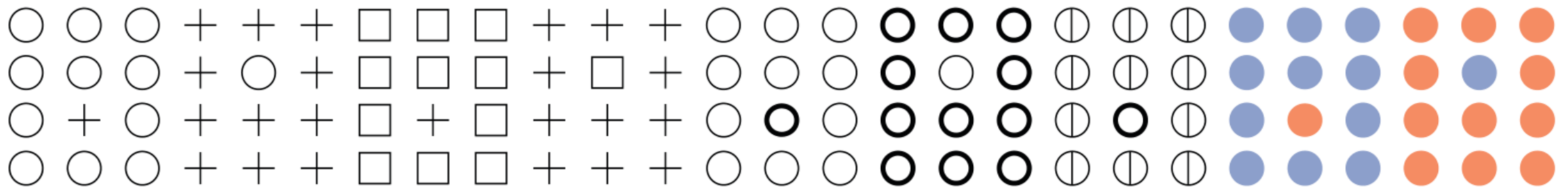
Saliency

With dense data, use visually distinctive shapes

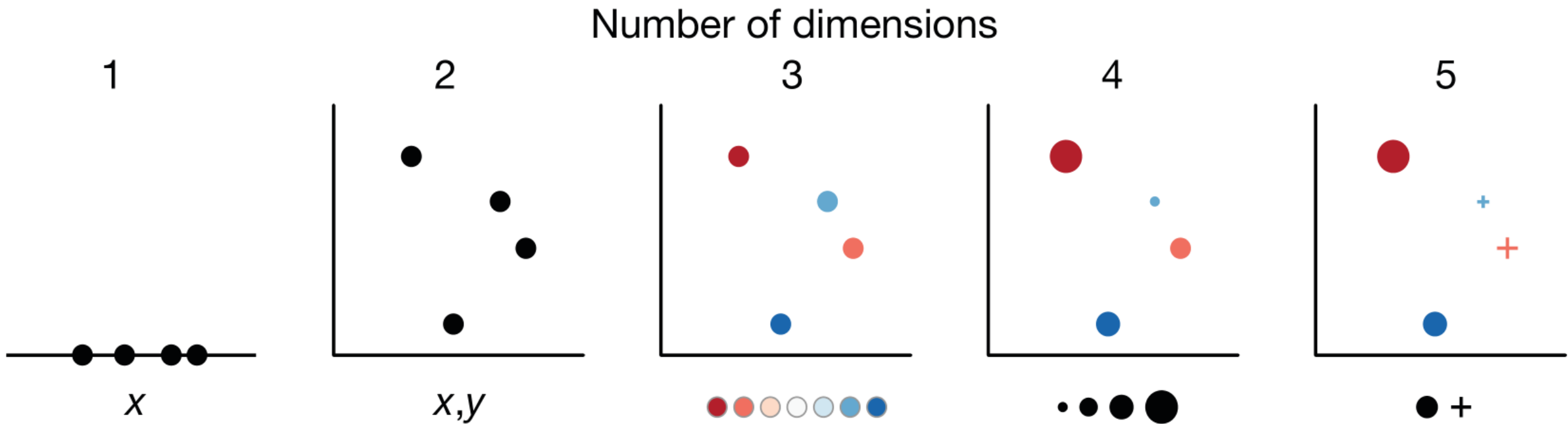
Weak visual boundaries



Strong visual boundaries



Encoding multivariate data



Effectiveness principle

The importance of the attribute should match the **salience** of the channel.

Choosing the most effective
visualization type

Creating an effective visualization type

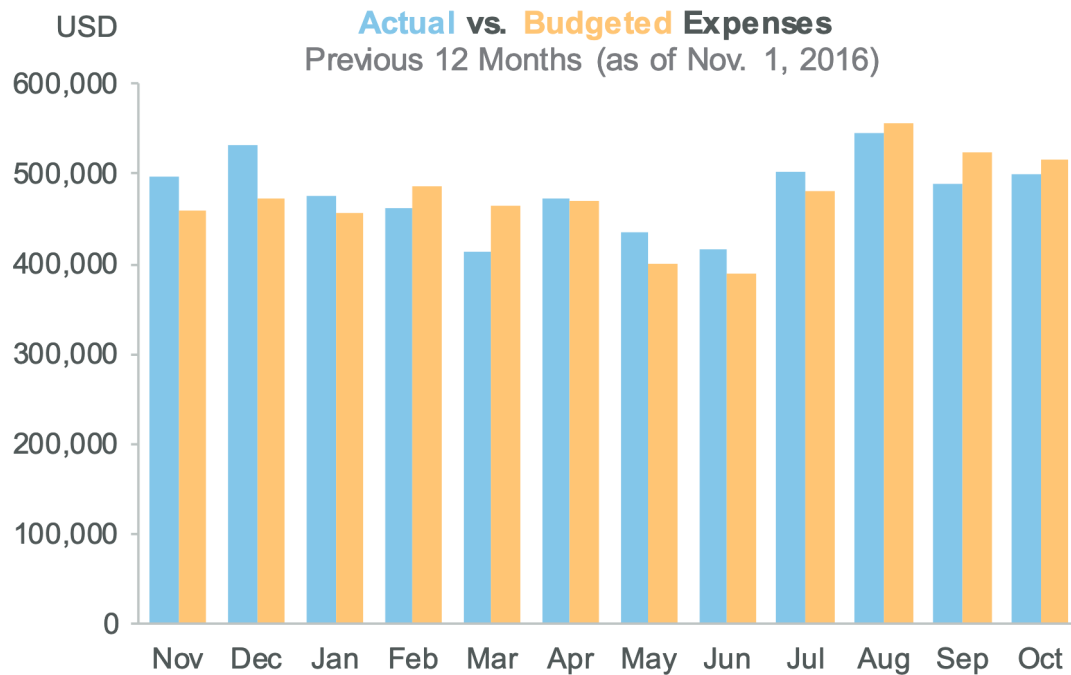
What are you trying to **highlight** in your data?

Are you trying to show a variable's overall pattern of variation or differences among individual values?

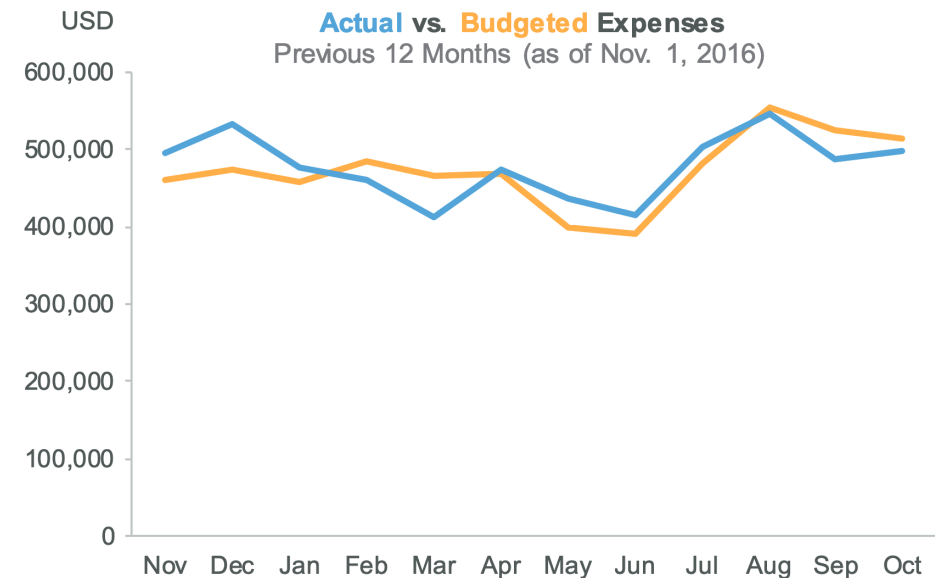
Creating an effective visualization type

What are you trying to highlight in your data?

Individual differences



Overall variation



Visualizing quantities

Bar plots

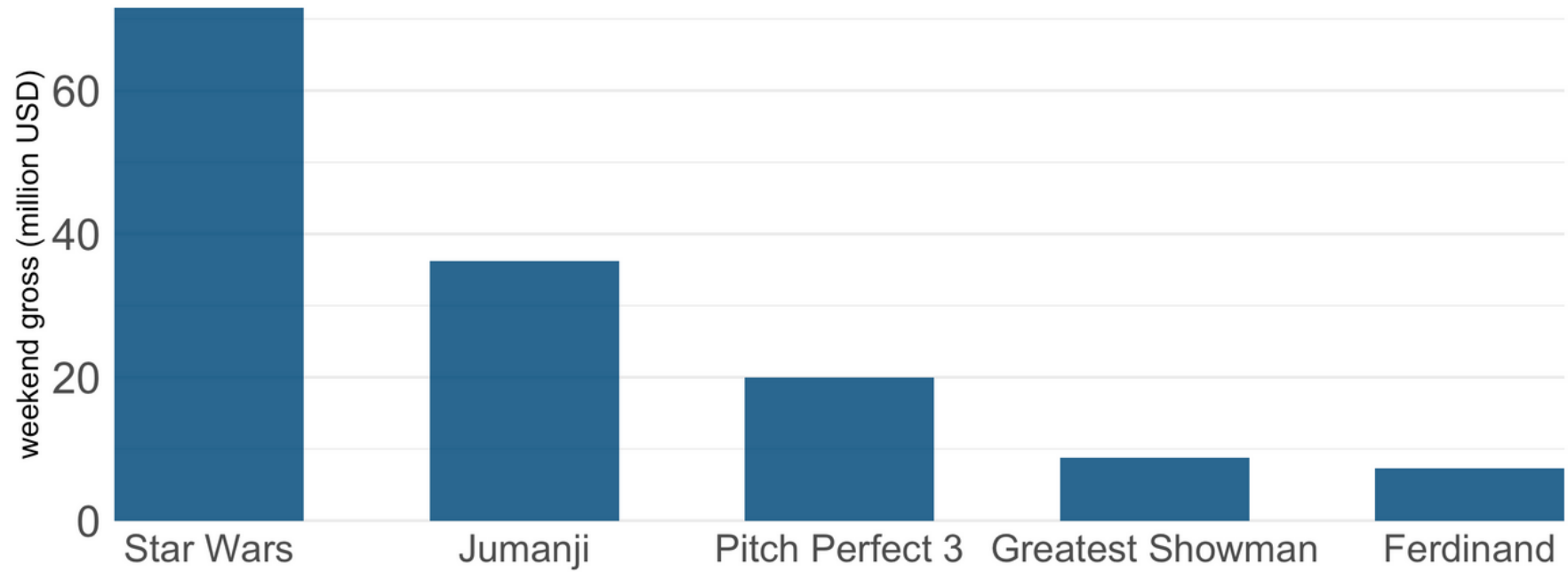
Show how a quantity varies with a categorical variable

Bar plots

Show how a quantity varies with a categorical variable

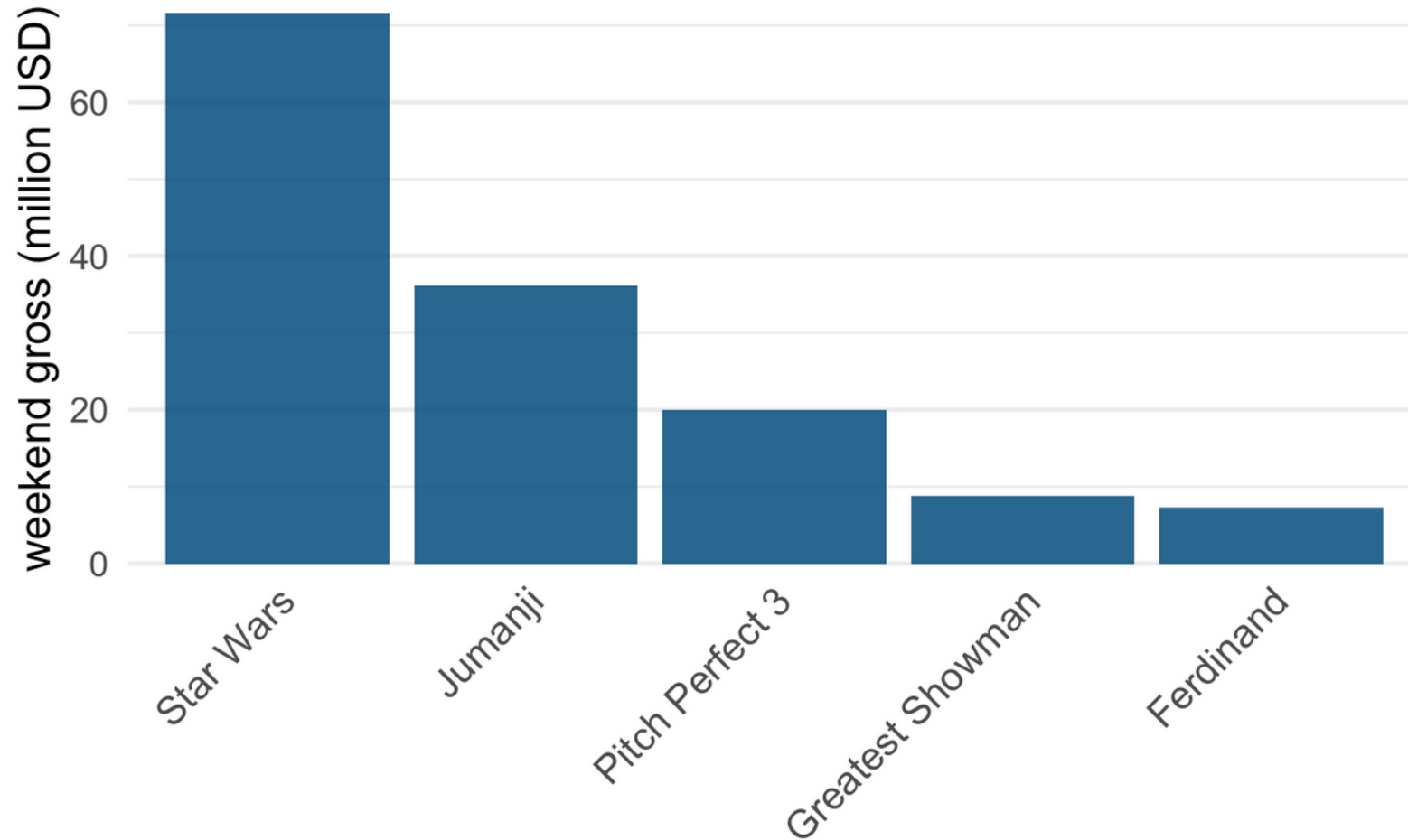
Rank	Movie Title	Amount
1	Star Wars: The Last Jedi	\$71,565,498
2	Jumanji: Welcome to the Jungle	\$36,169,328
3	Pitch Perfect 3	\$19,928,525
4	The Greatest Showman	\$8,805,843
5	Ferdinand	\$7,316,746

Bar plot

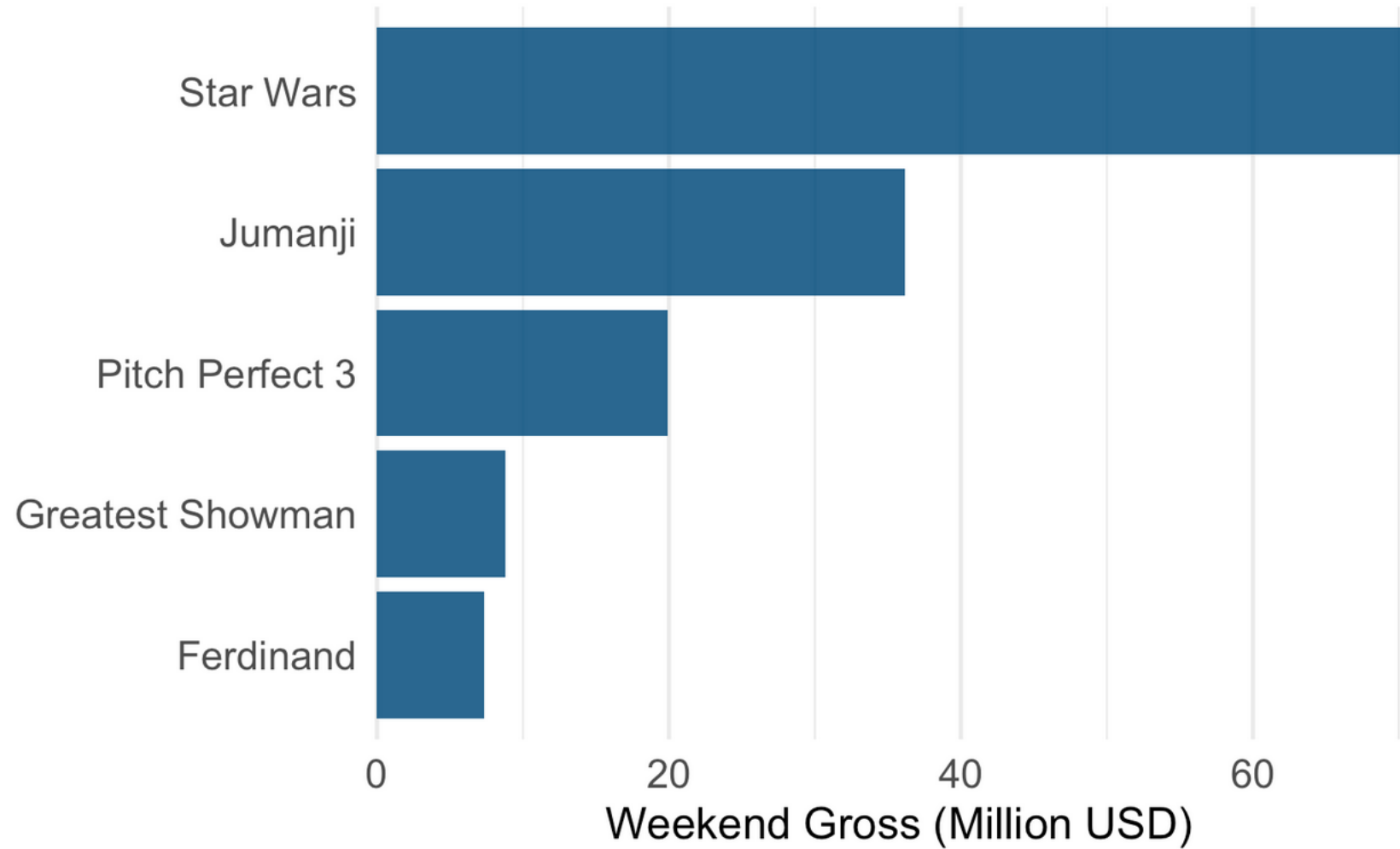


Bar plot

What **not** to do:

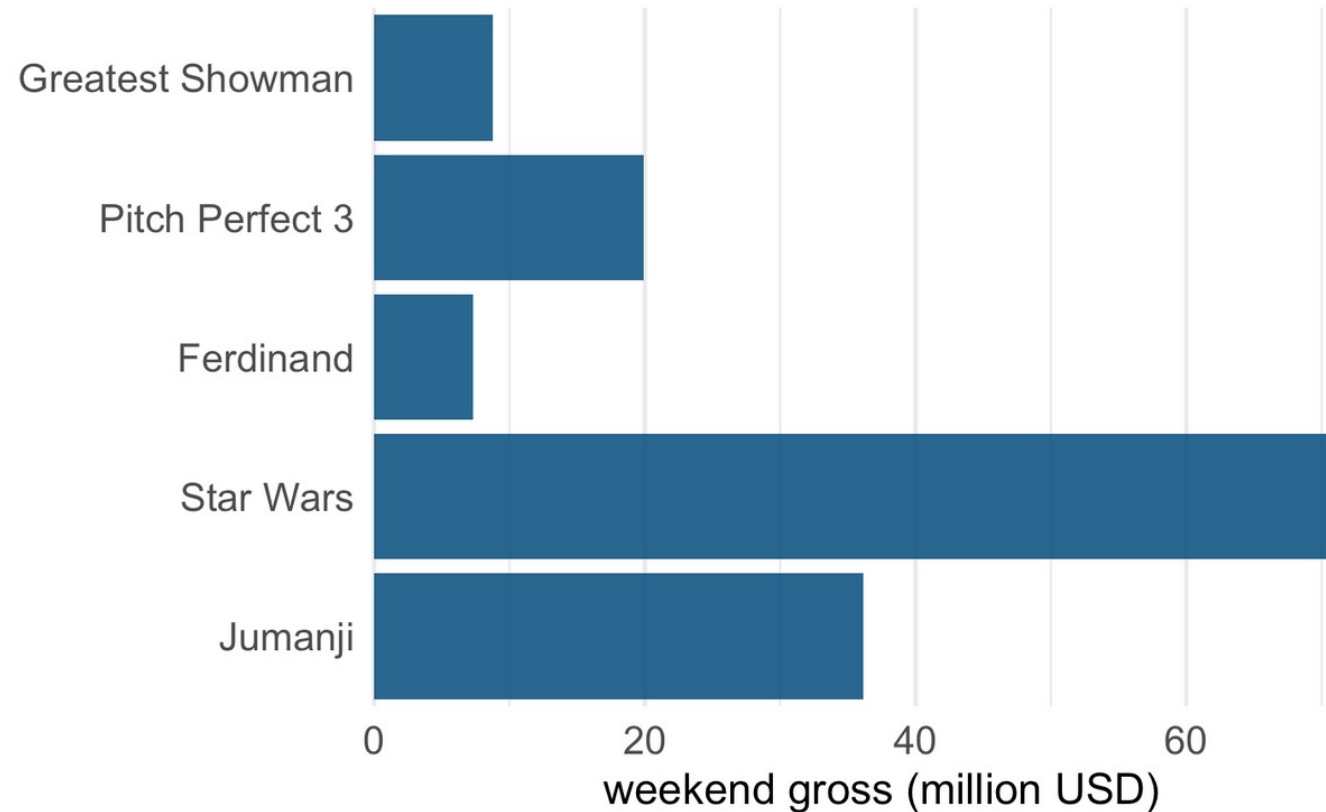


Rotated bar plot



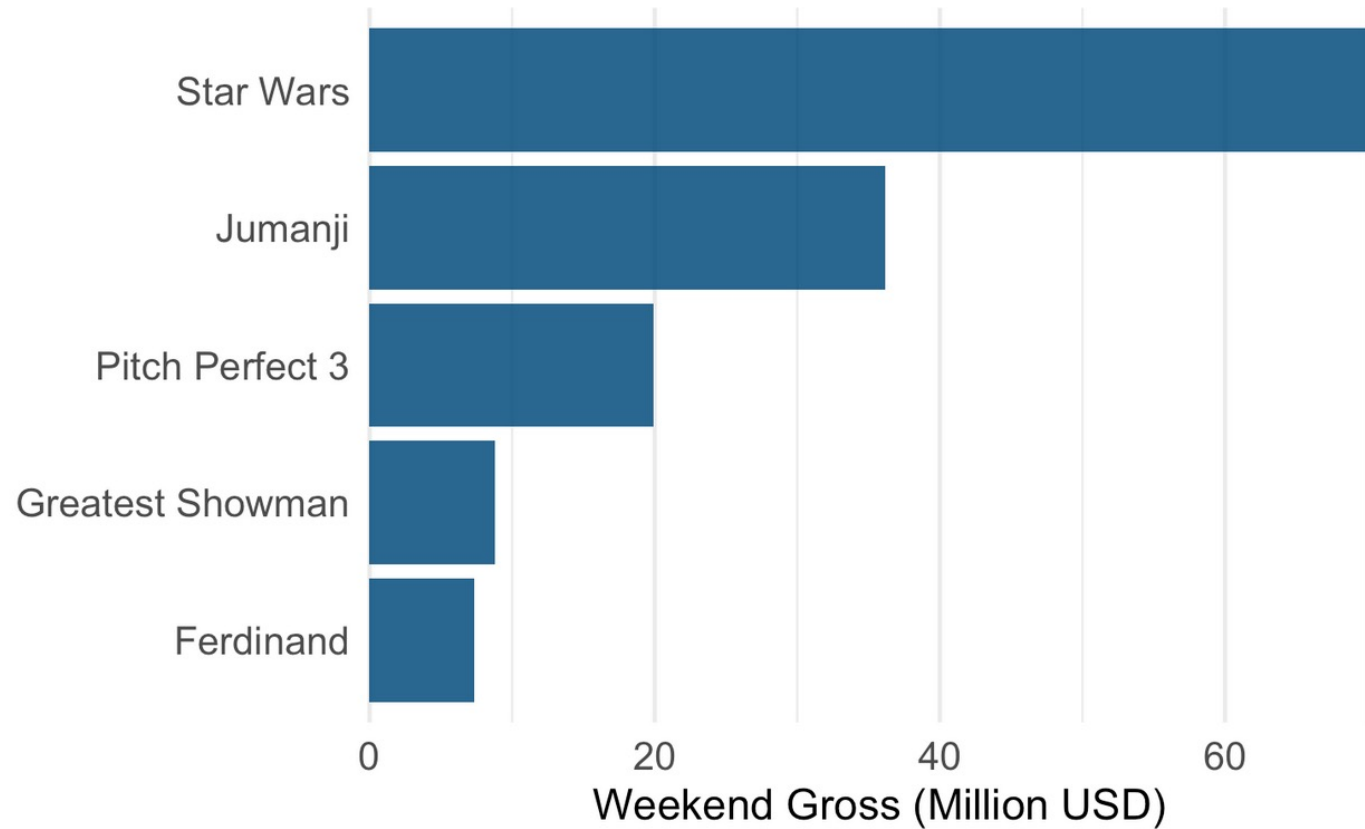
Ordering bars

Order bars by [ascending](#) or [descending](#) order *when* there is no natural ordering to the categories the bars represent



Ordering bars

Order bars by **ascending** or **descending** order *when* there is no natural ordering to the categories the bars represent

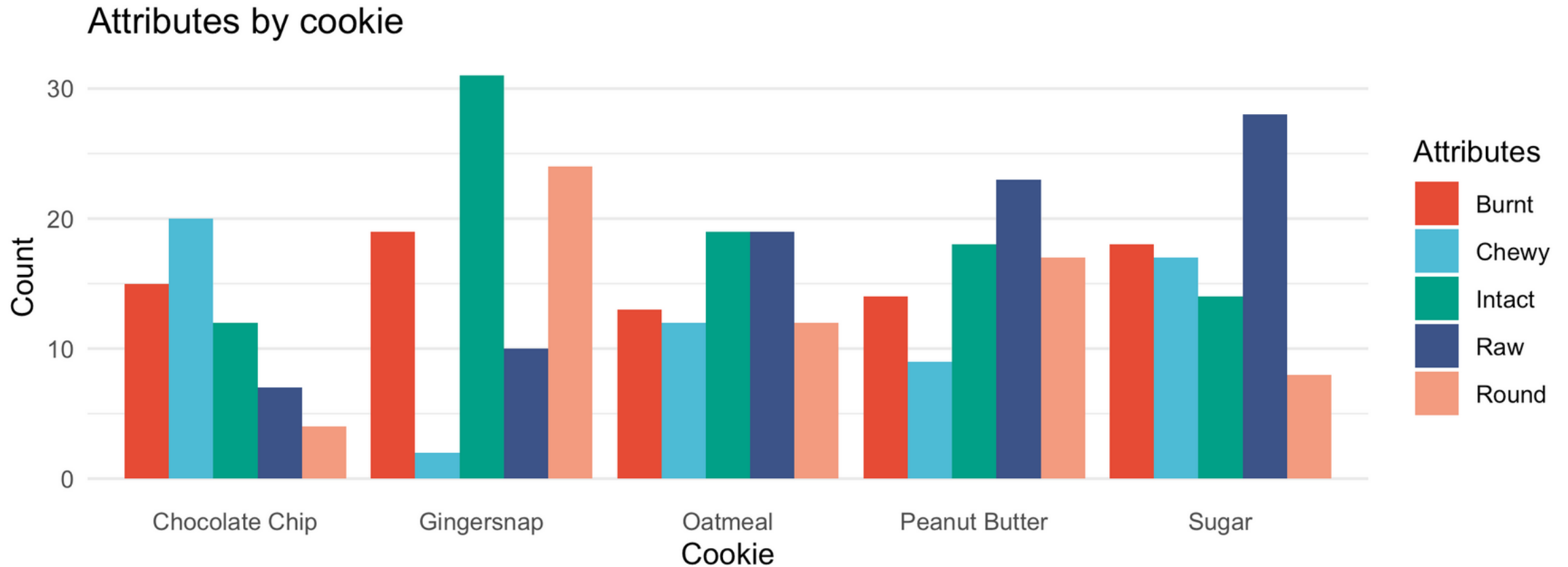


Visualizing multiple categorical variables

Cookie	Raw	Burnt	Chewy	Round	Intact
Chocolate Chip	7	15	20	4	12
Sugar	28	18	17	8	14
Oatmeal	19	13	12	12	19
Peanut Butter	23	14	9	17	18
Gingersnap	10	19	2	24	31

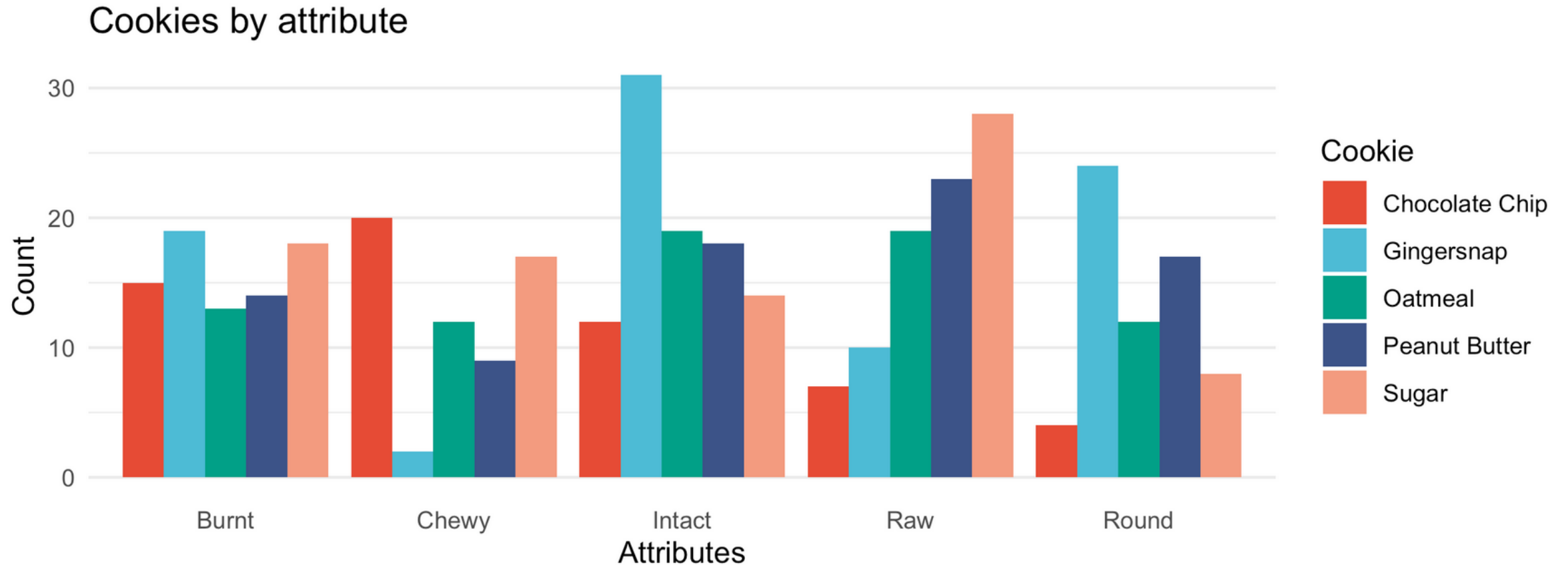
Grouped bar plots

Comparisons across individuals



Grouped bar plots

Comparisons across categories



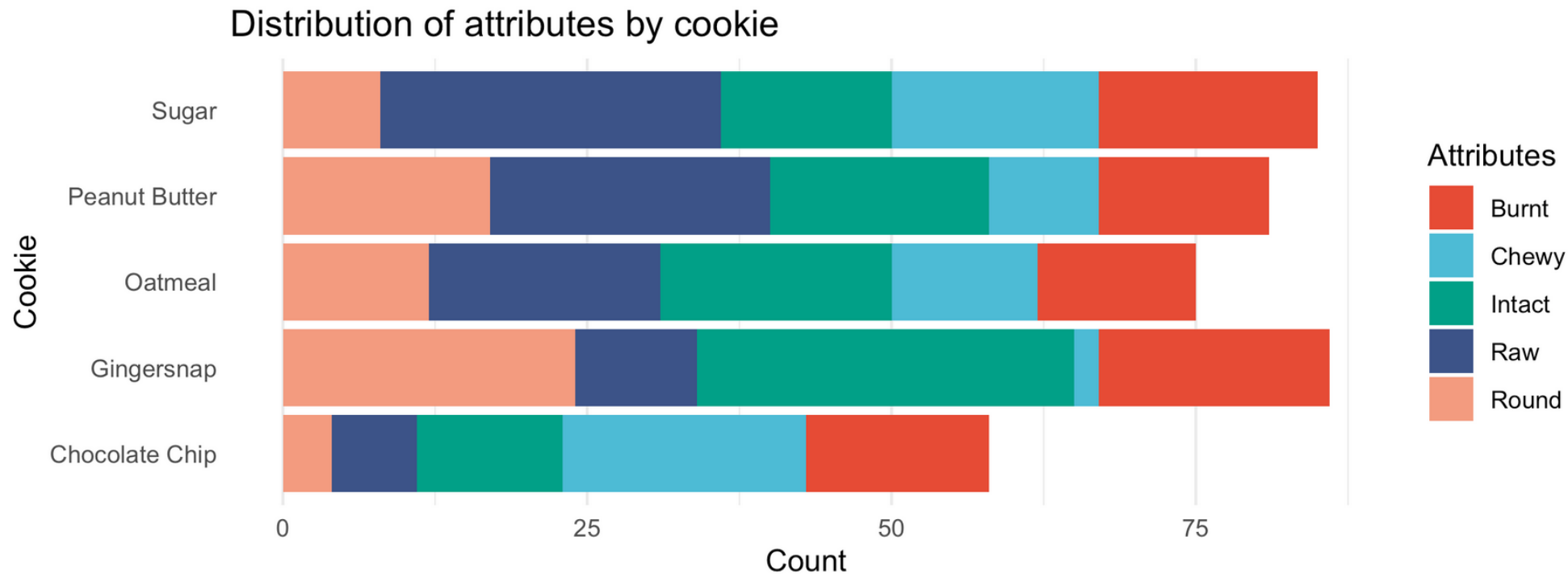
Stacked bar plots

Distribution of attributes *within* individuals

Stacking can be used when:

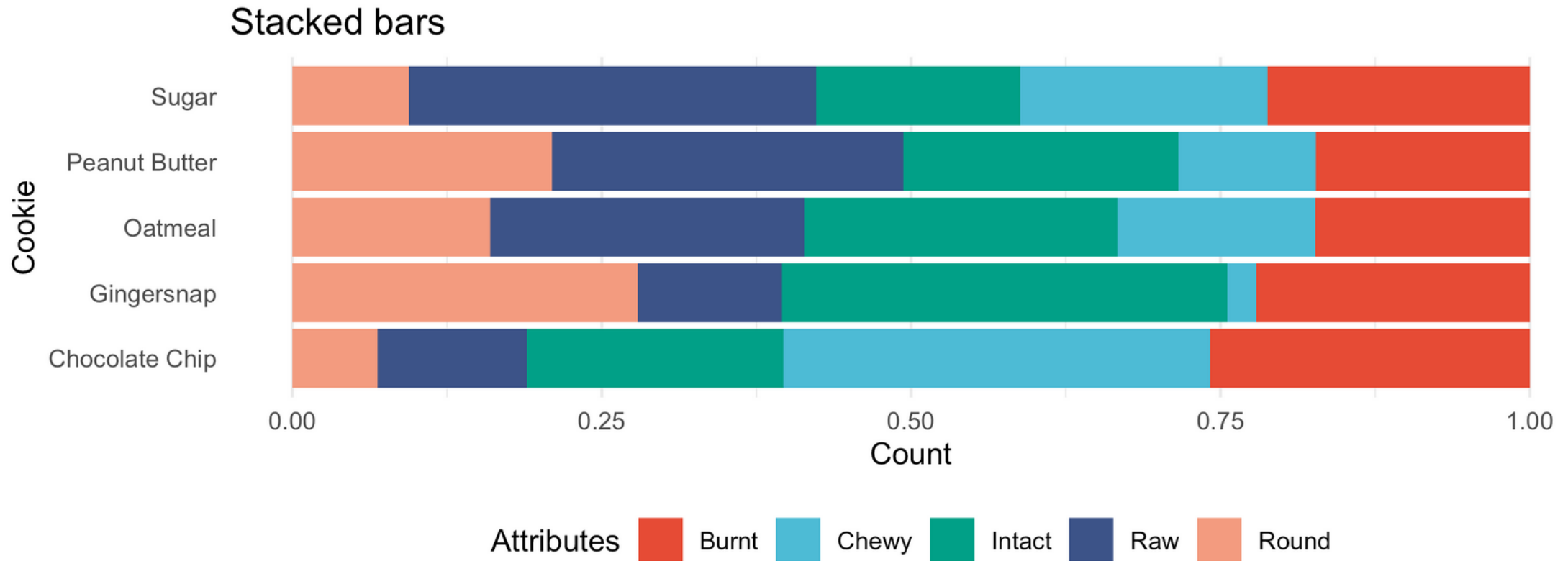
The sum of the individual stacked bars is a meaningful amount (total cookies per cookie type) or

The individual bars represent counts (number of cookies)



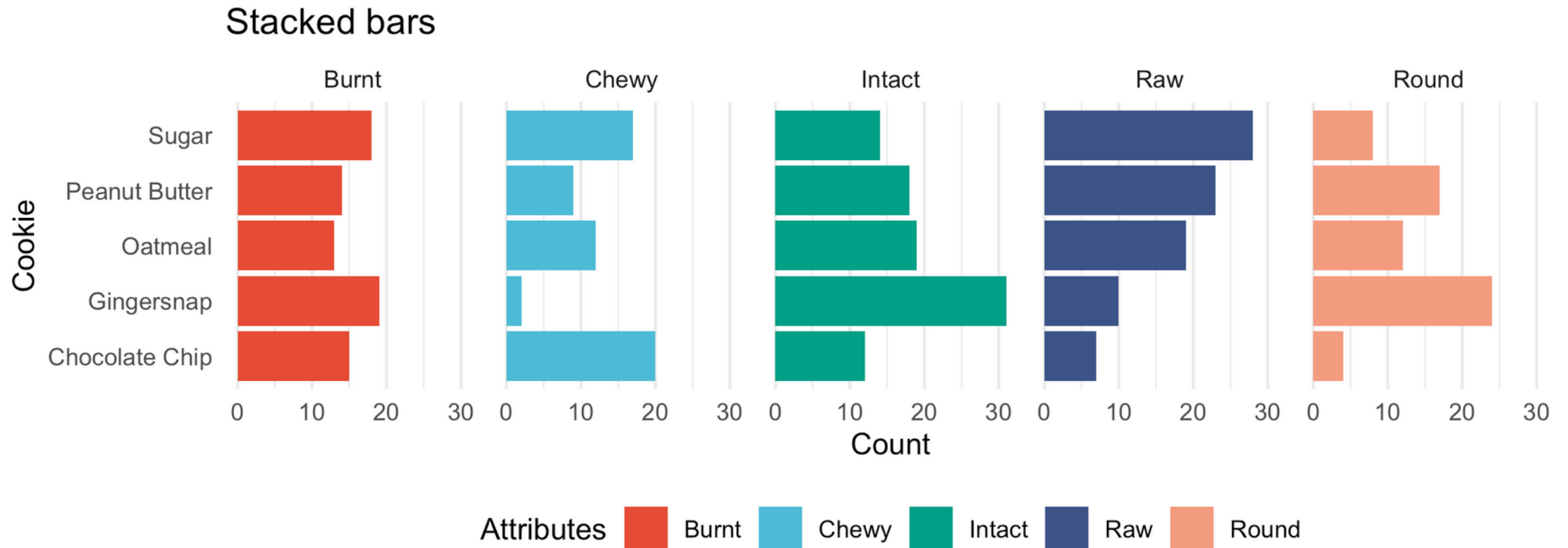
Percent stacked bar plots

Distribution across individuals



Facetted bar plots

Comparisons *within* attributes



Visualizing distributions

Visualizing a single distribution

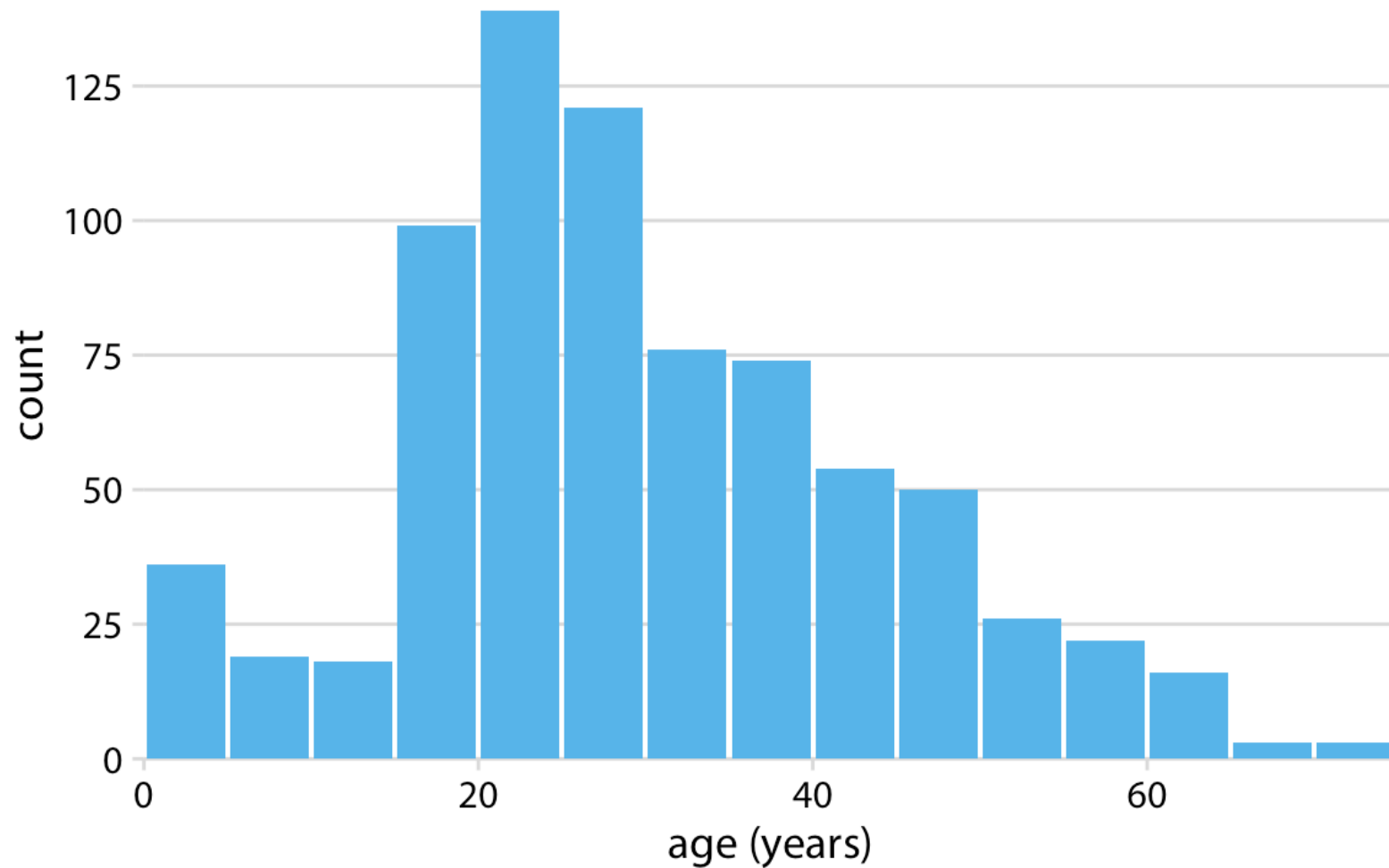
Numbers of passenger with known age on the Titanic

Age range	Count
0–5	36
6–10	19
11–15	18
16–20	99
21–25	139
26–30	121

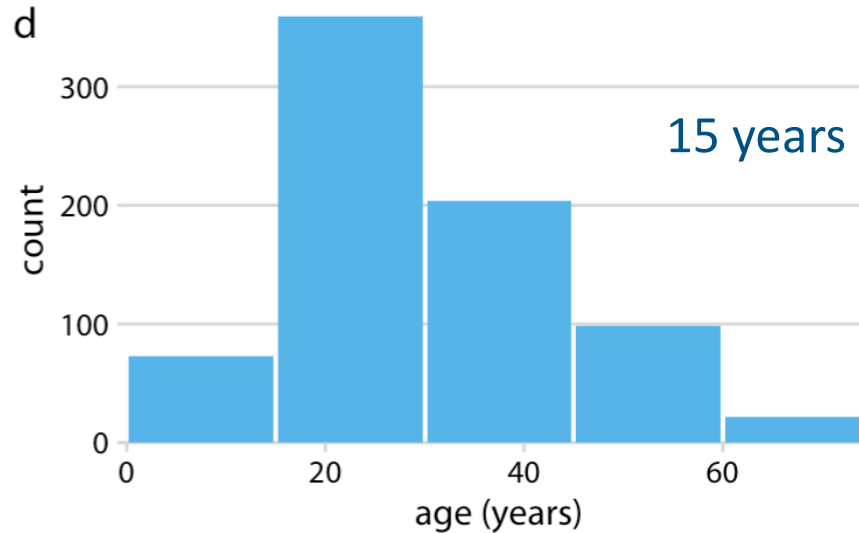
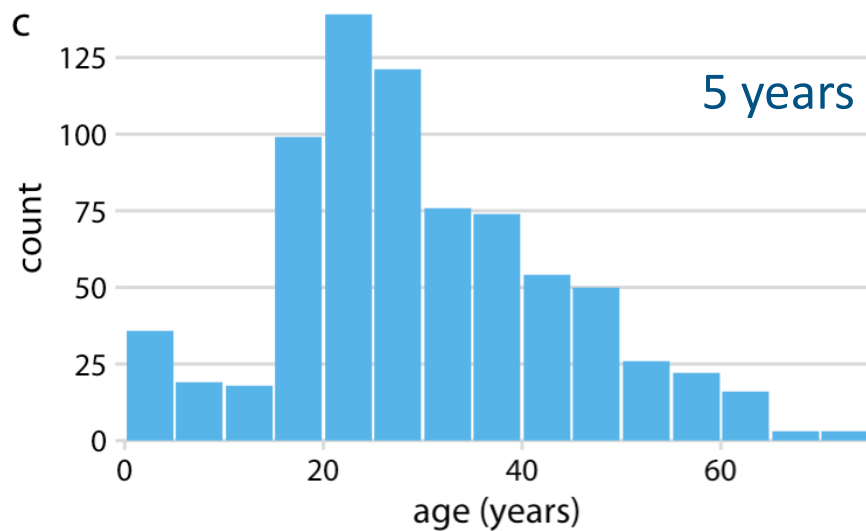
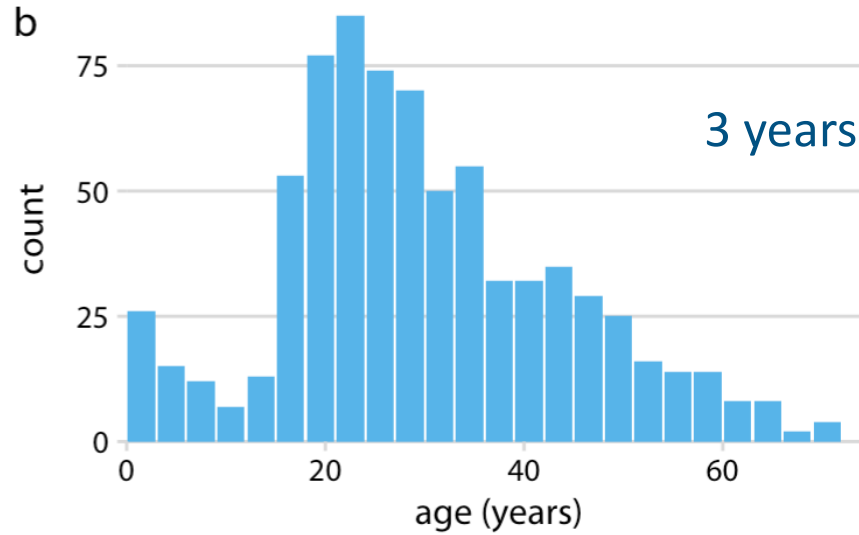
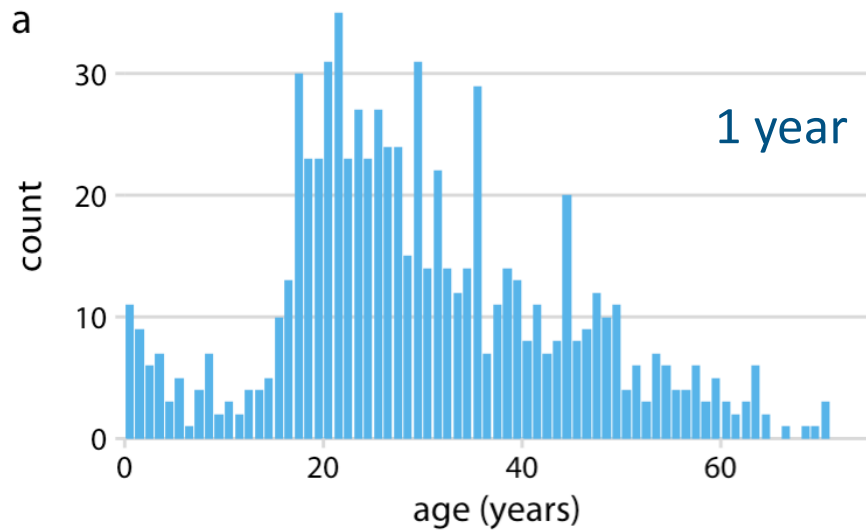
Age range	Count
31–35	76
36–40	74
41–45	54
46–50	50
51–55	26
56–60	22

Age range	Count
61–65	16
66–70	3
71–75	3

Histogram

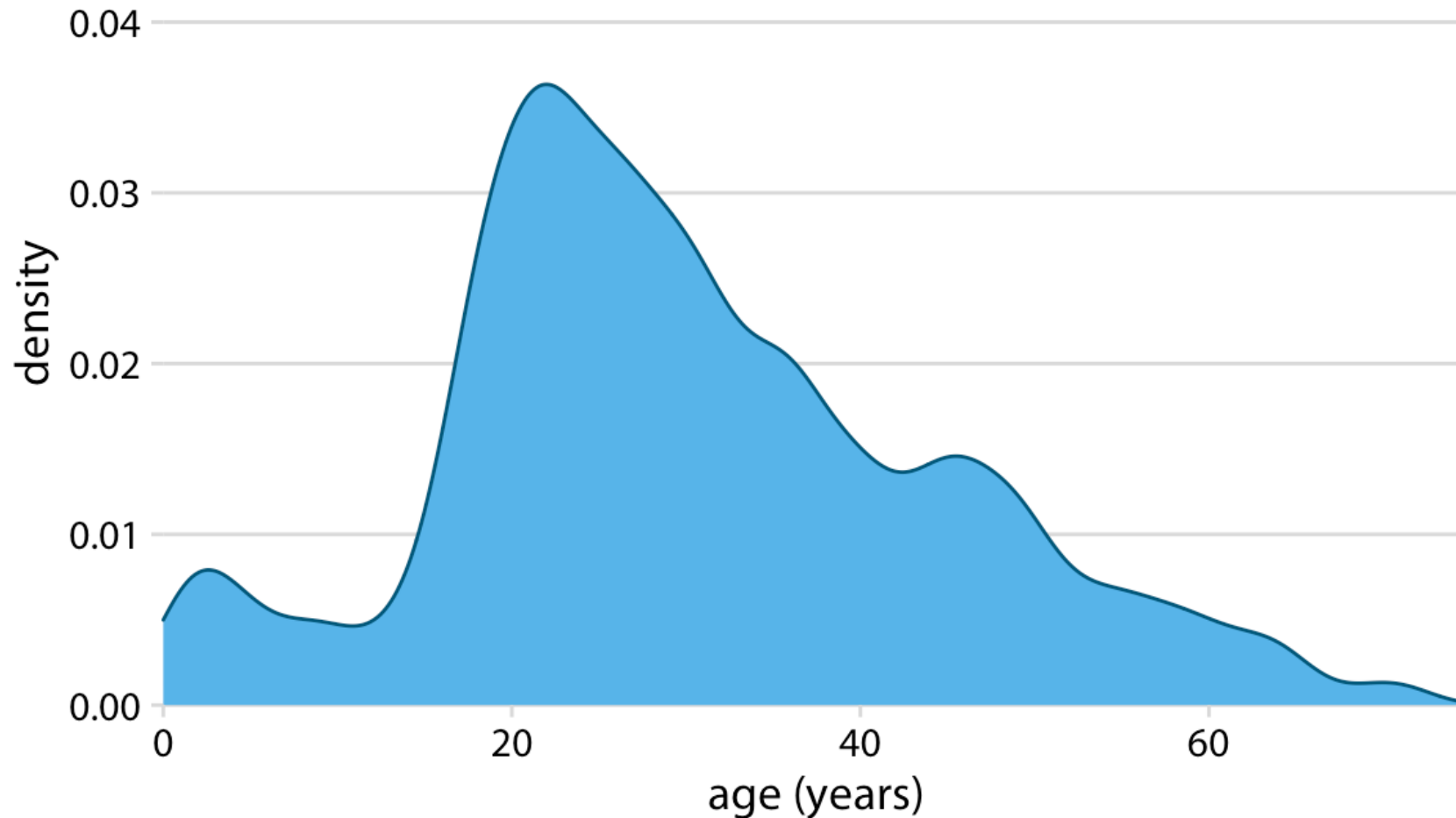


Histogram - varying bin width



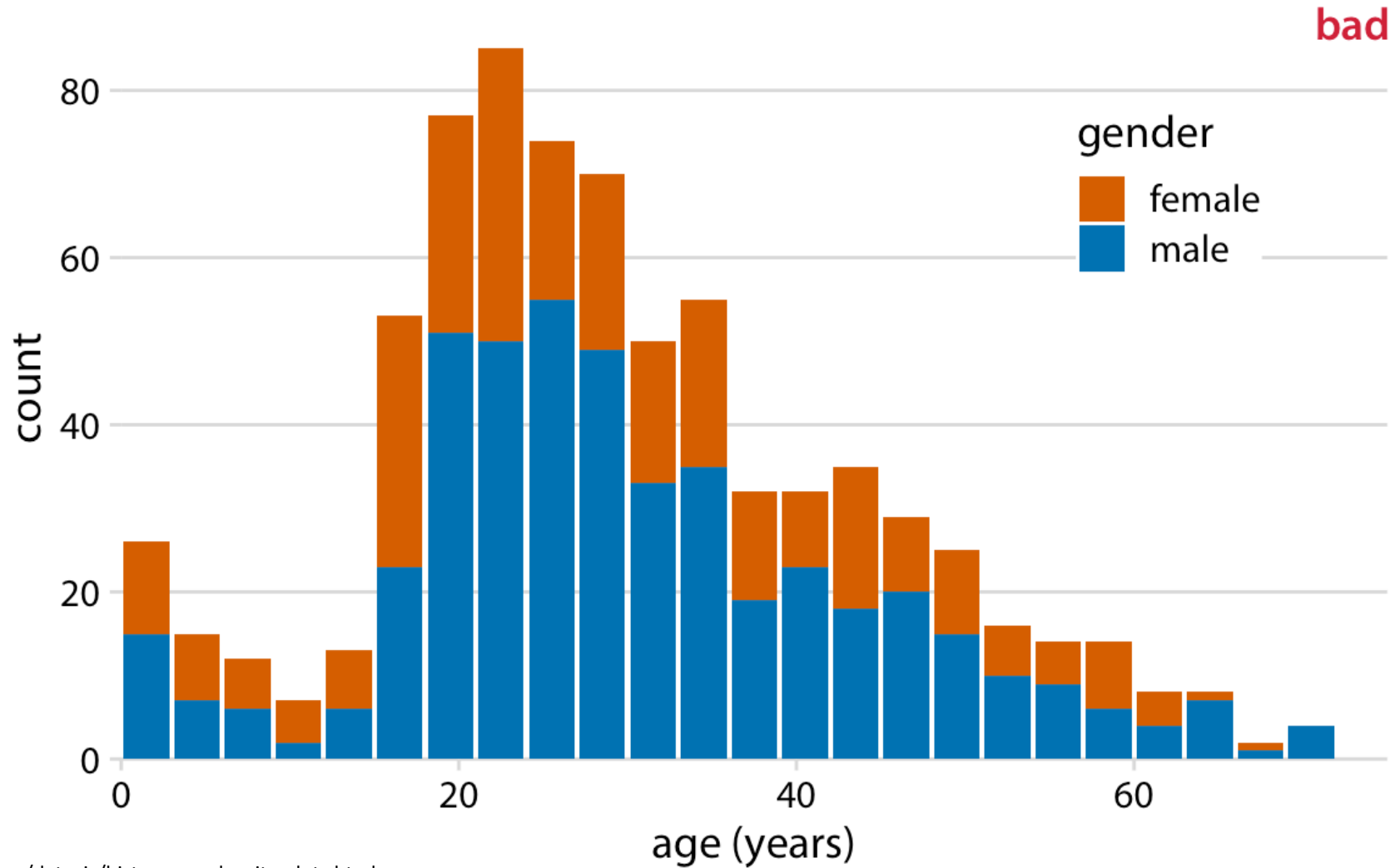
Kernel density estimate (KDE)

Approximate the underlying probability distribution of the data



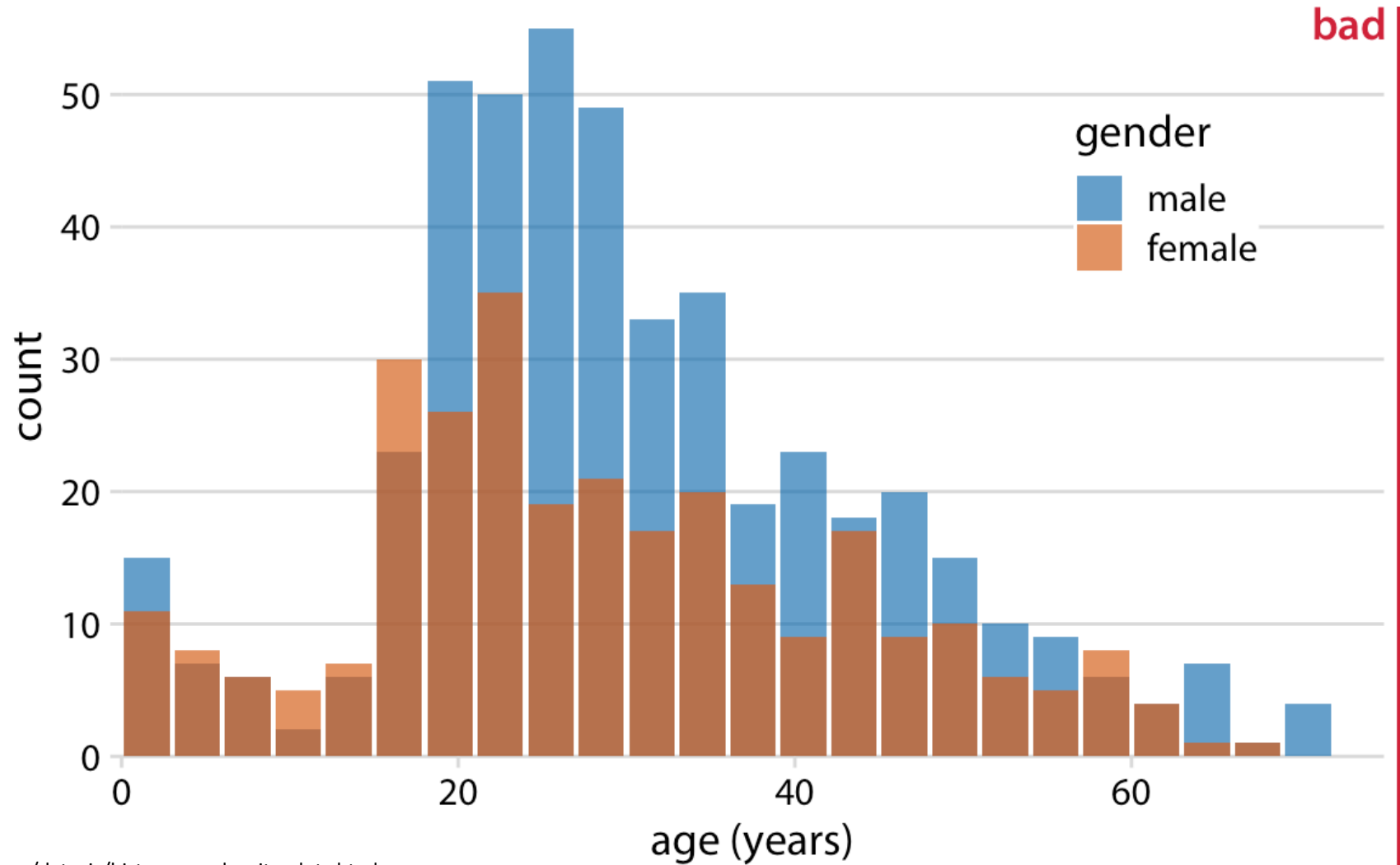
Visualizing multiple distributions

Stacked histograms are problematic

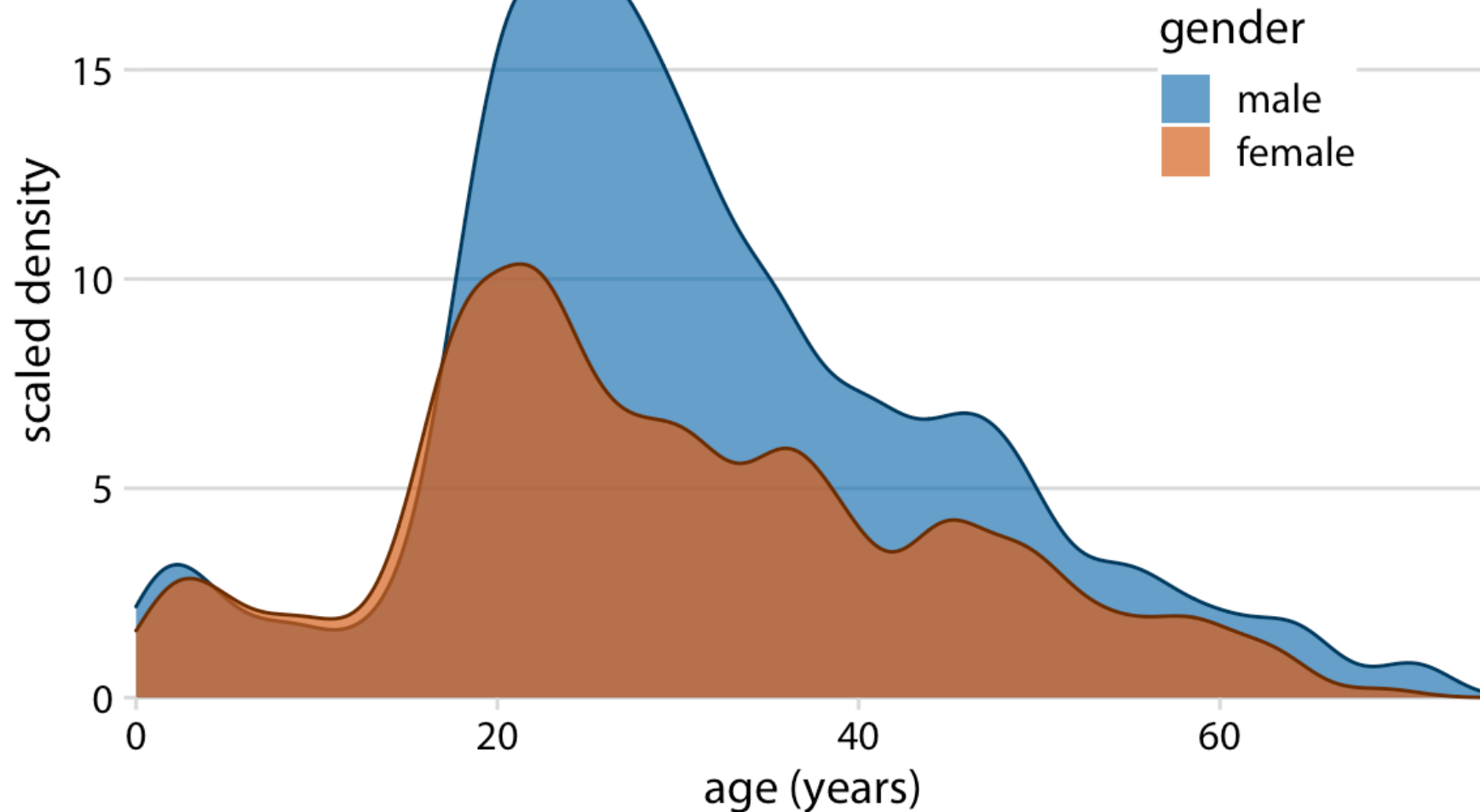


Visualizing multiple distributions

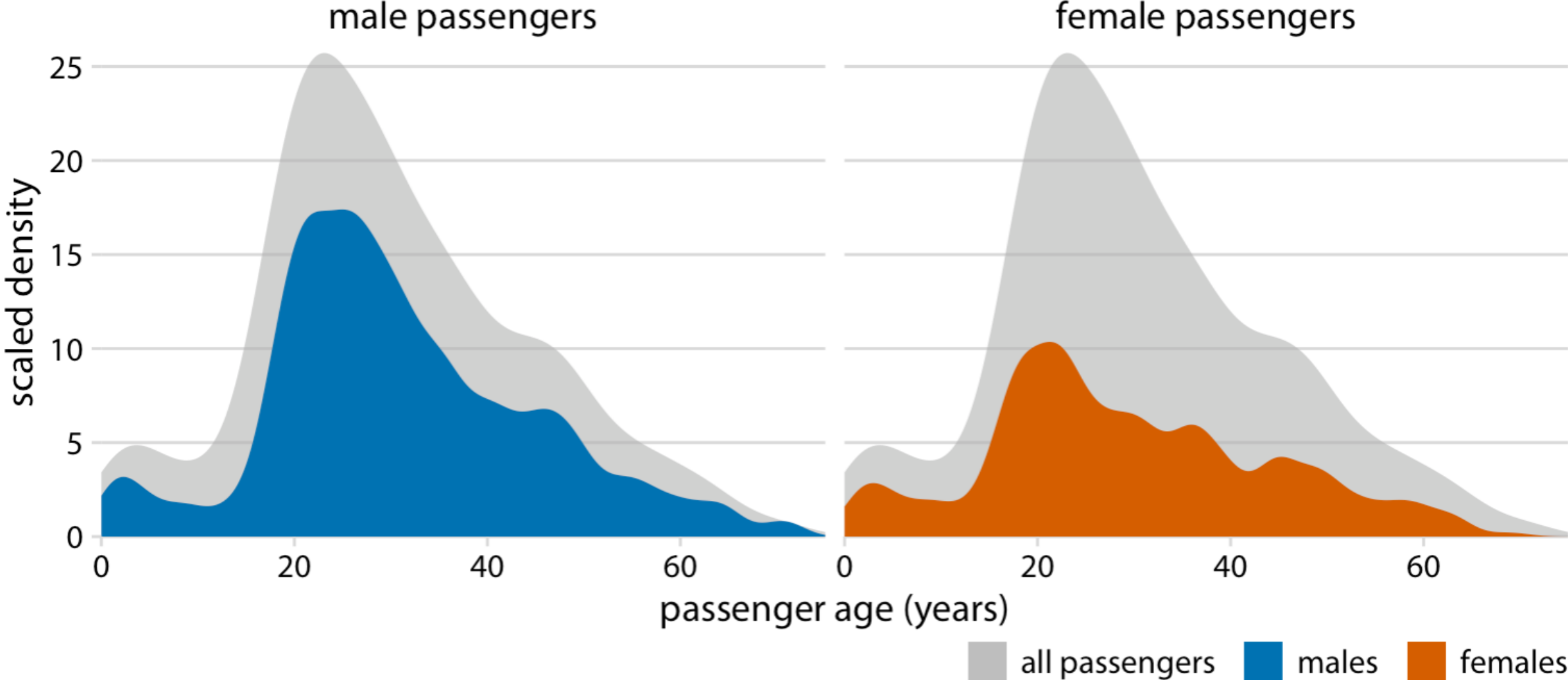
Overlapping histograms are problematic for different reasons



Overlapping density estimates



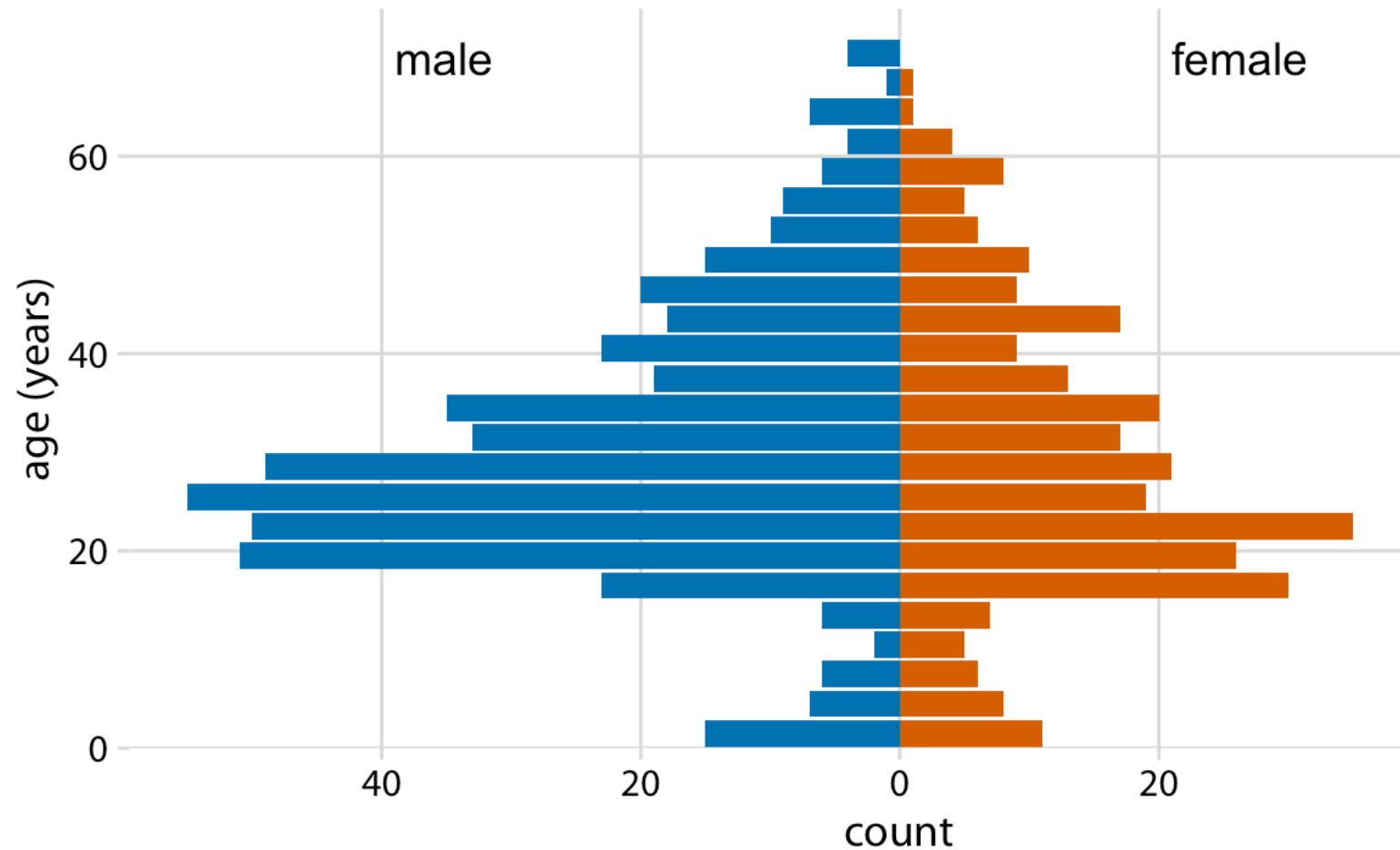
Small multiples



Source: <https://clauswilke.com/dataviz/histograms-density-plots.html>

Age pyramids

Effective for **exactly** two distributions

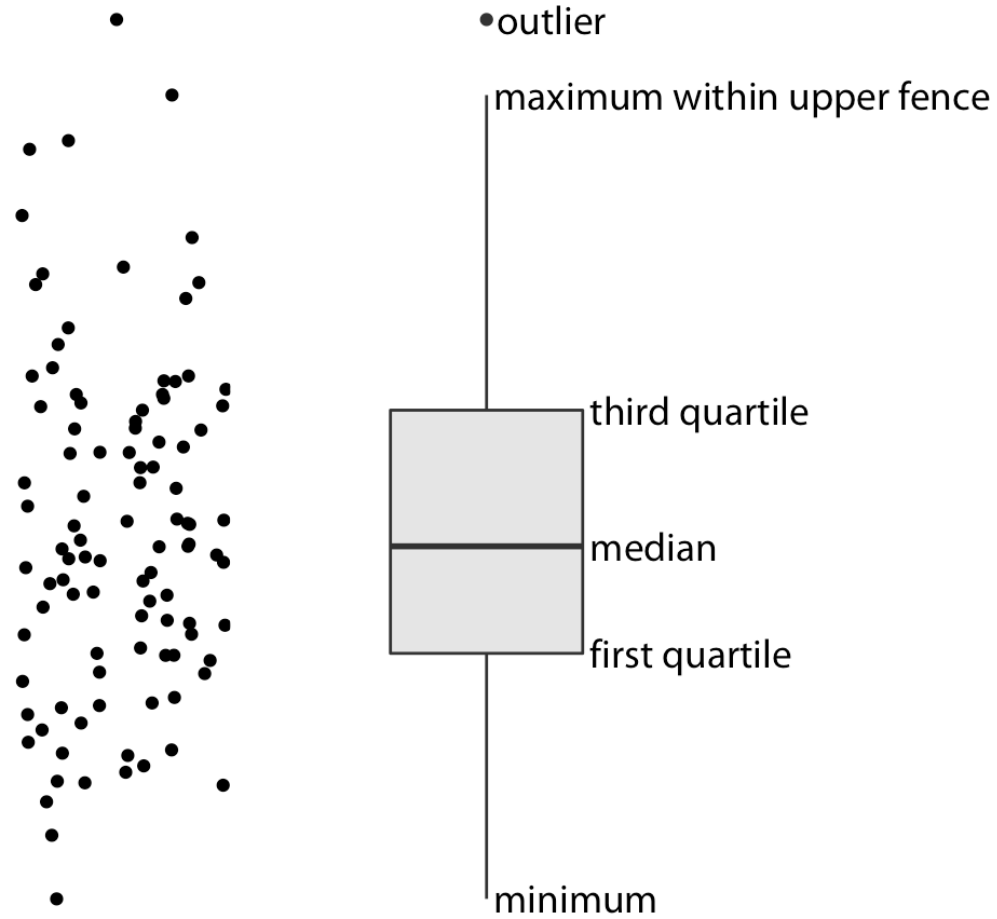


Visualizing many distributions

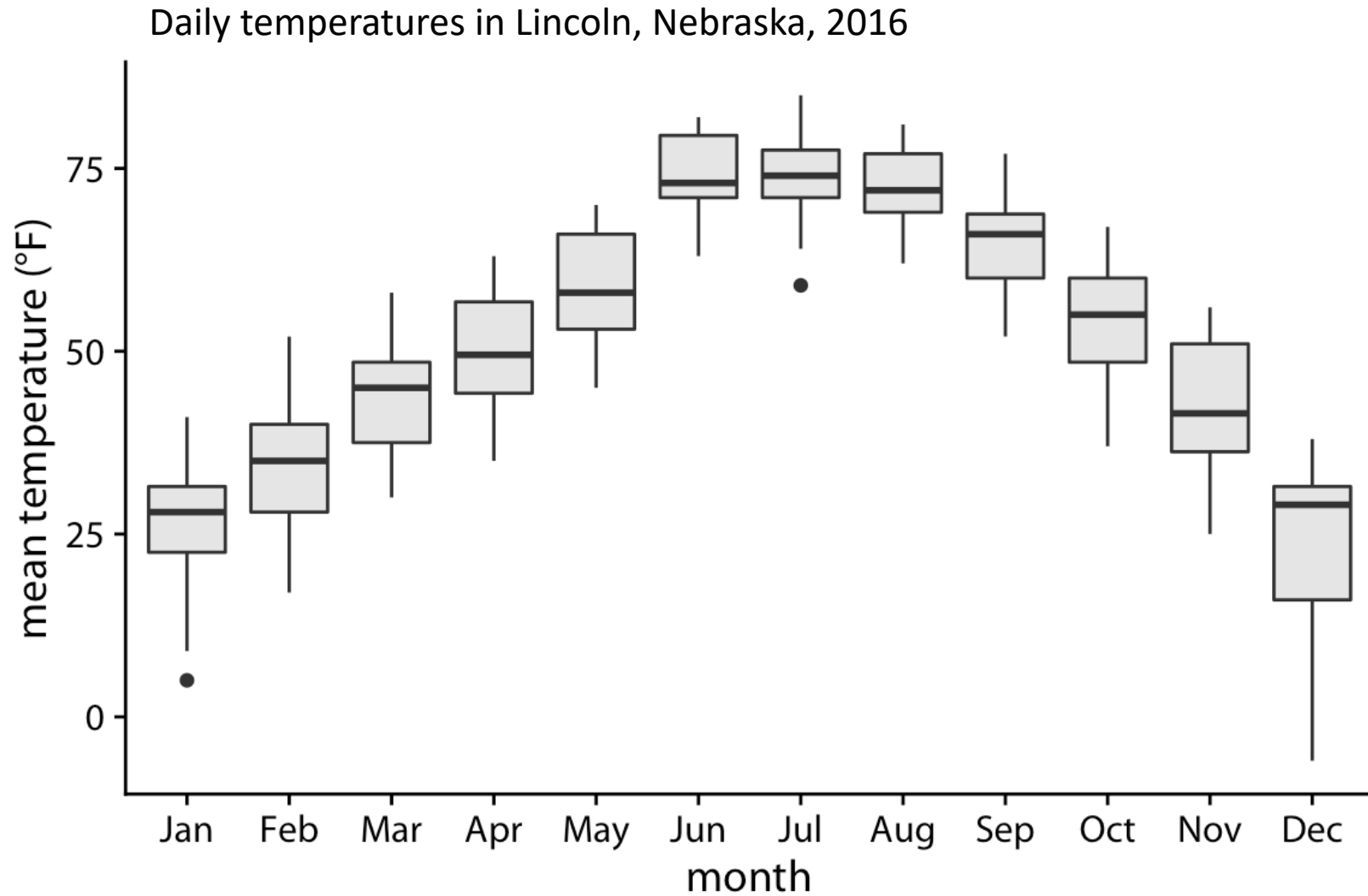
Visualizing many distributions

Visualizing distributions along the vertical axis

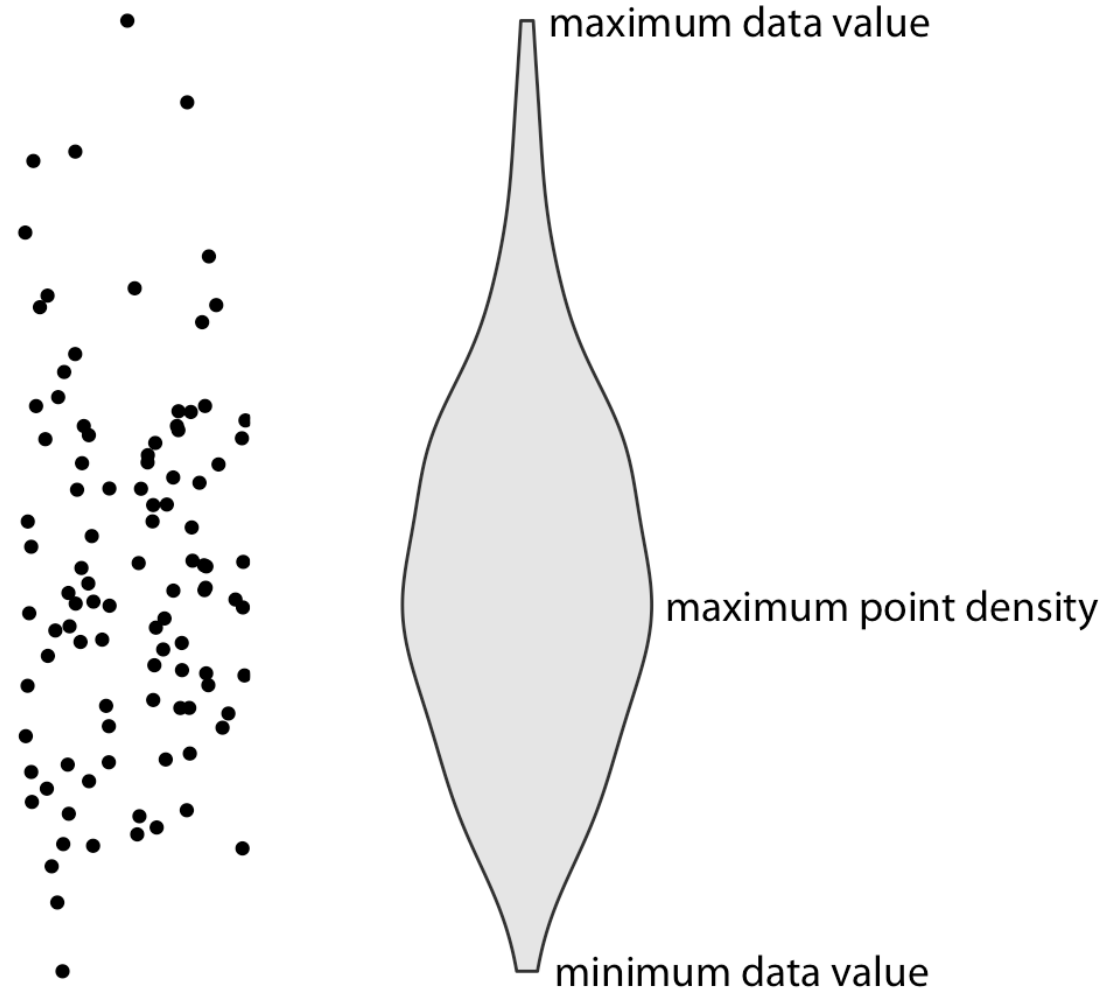
Boxplots



Boxplots



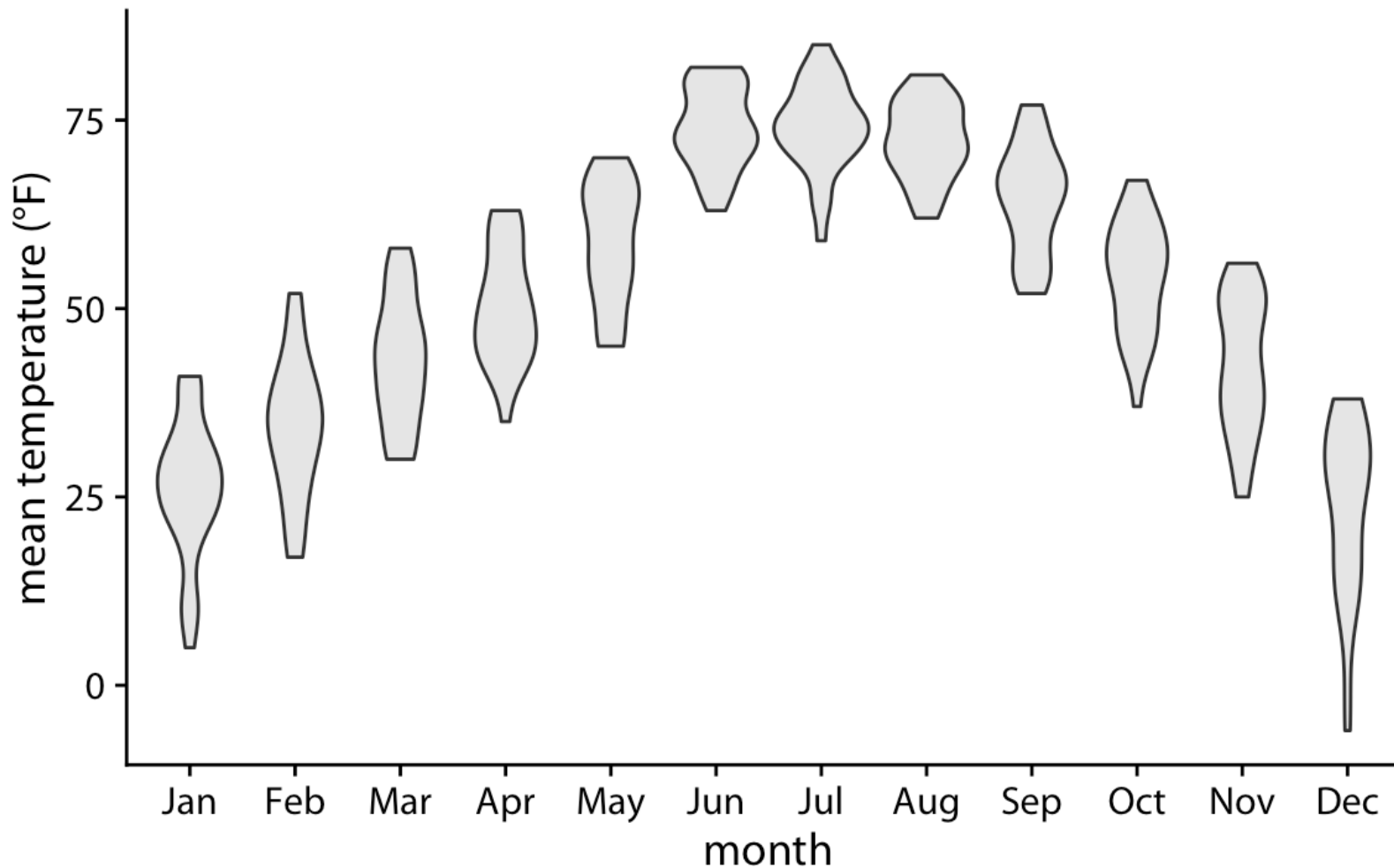
Violin plots



Violin plots

A density estimate alternative to boxplots

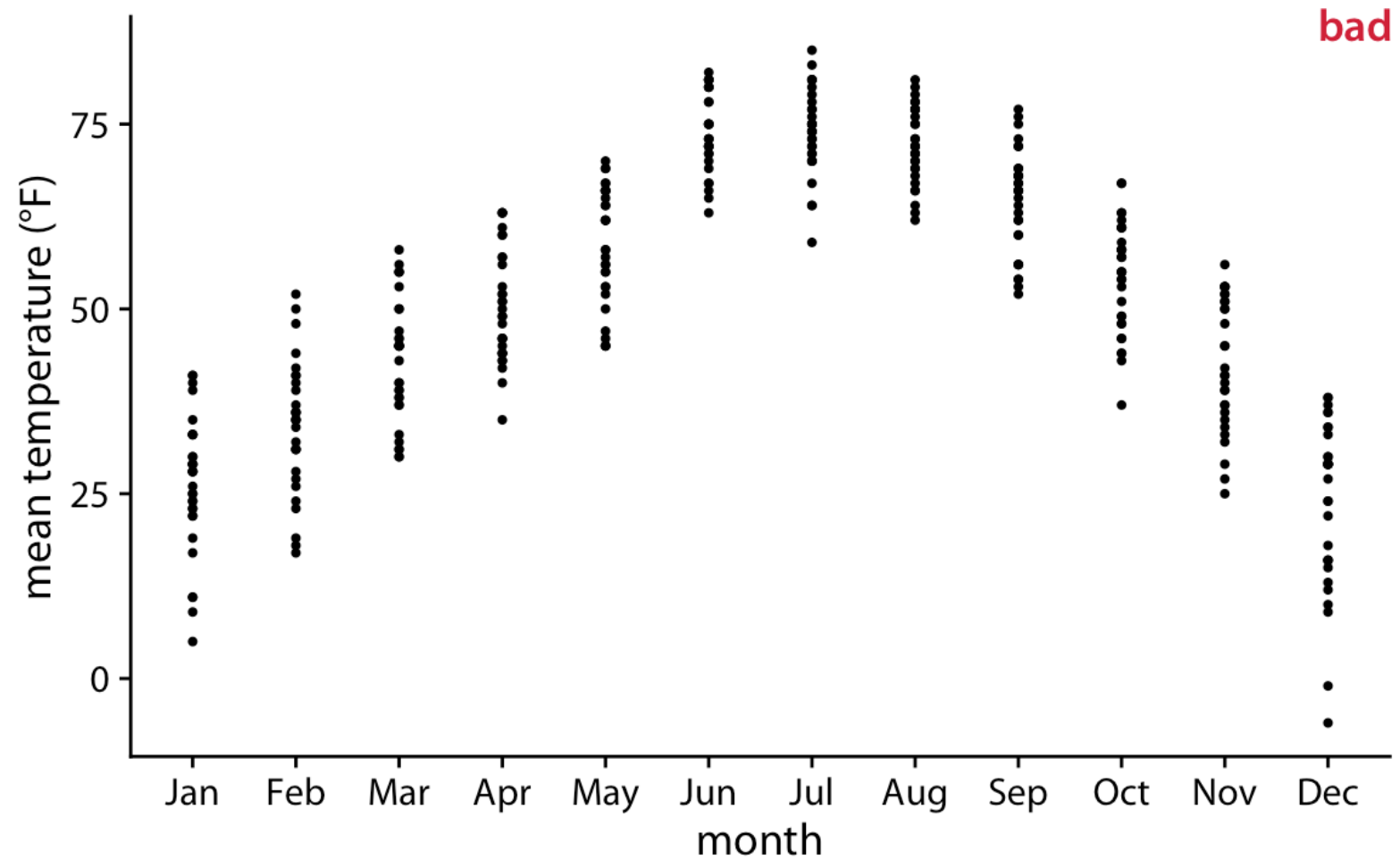
Daily temperatures in Lincoln, Nebraska, 2016



Strip charts

Show the raw data rather than an estimate or summary

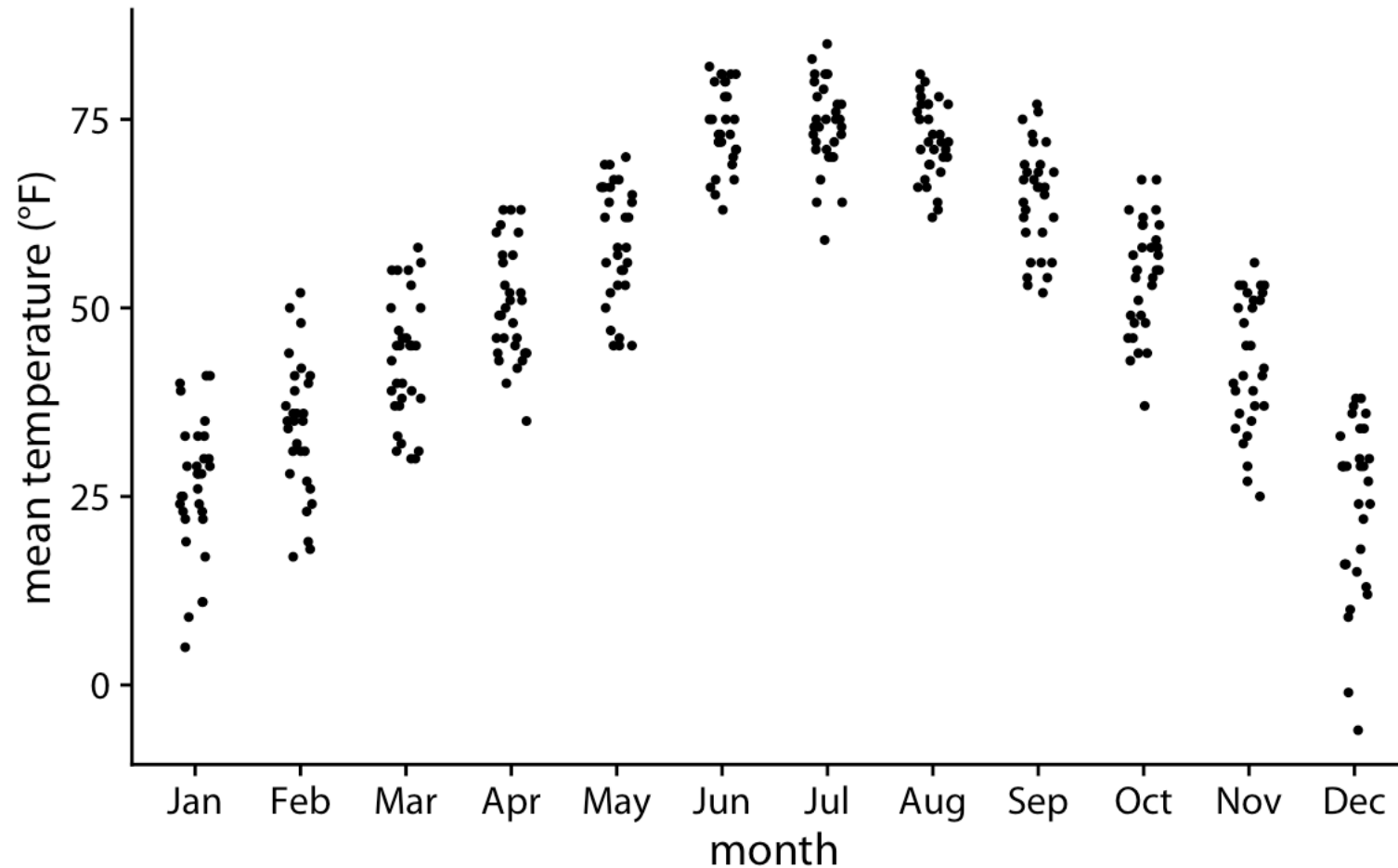
Daily temperatures in Lincoln, Nebraska, 2016



Strip charts

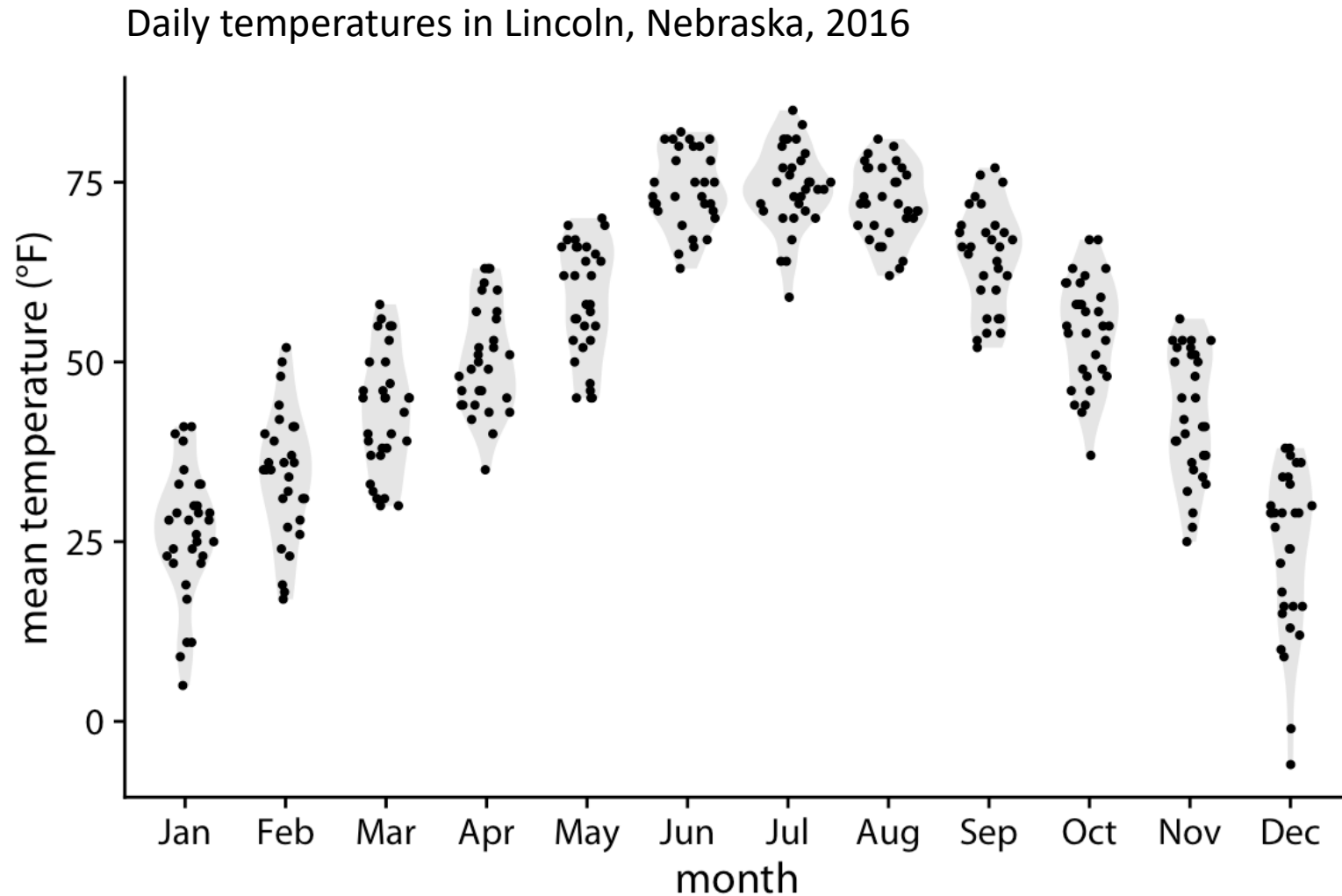
Overcome point occlusion with jittering

Daily temperatures in Lincoln, Nebraska, 2016



Sina plot

A strip chart with a width that corresponds to the density estimate

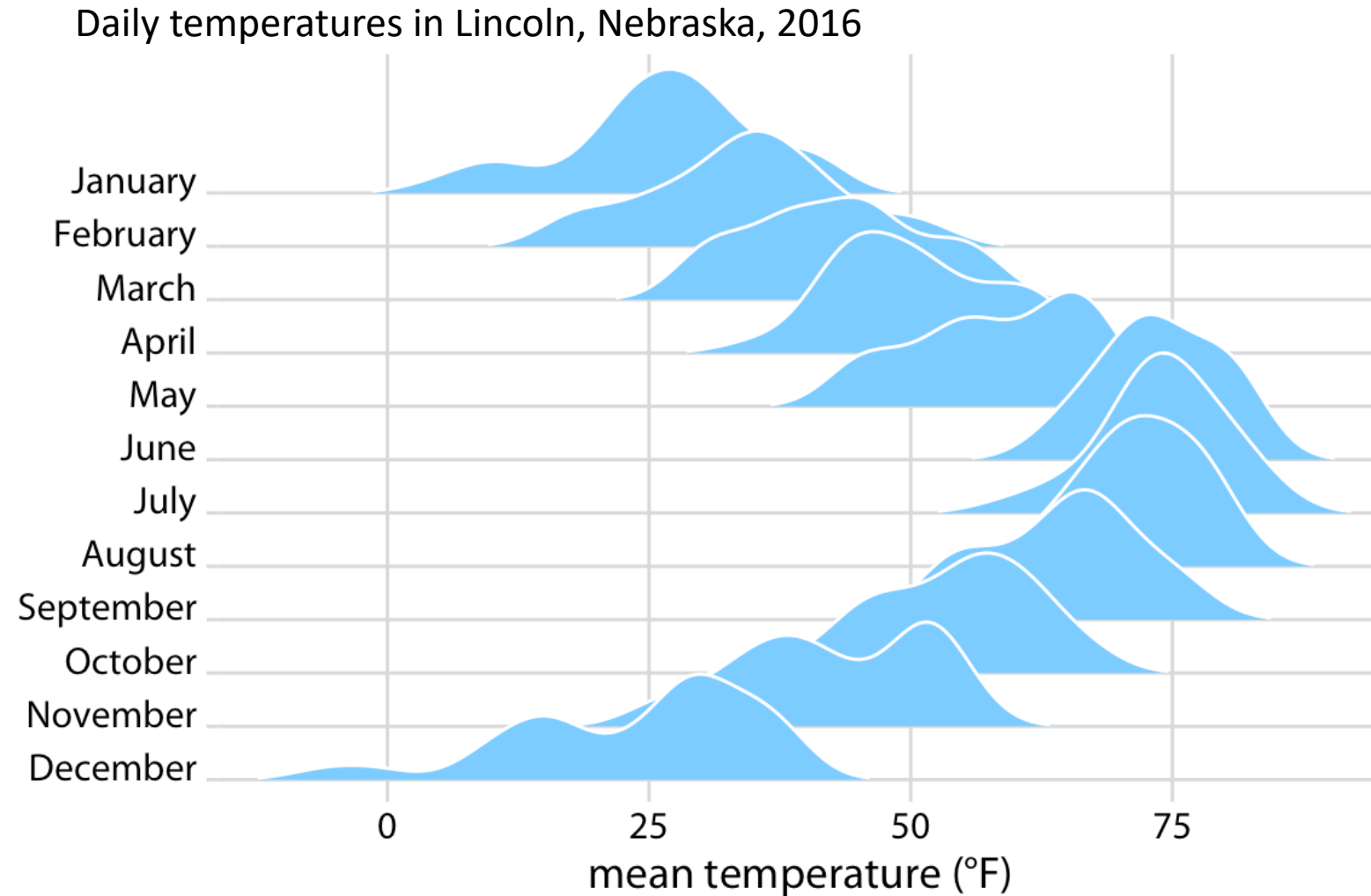


Visualizing many distributions

Visualizing distributions along the horizontal axis

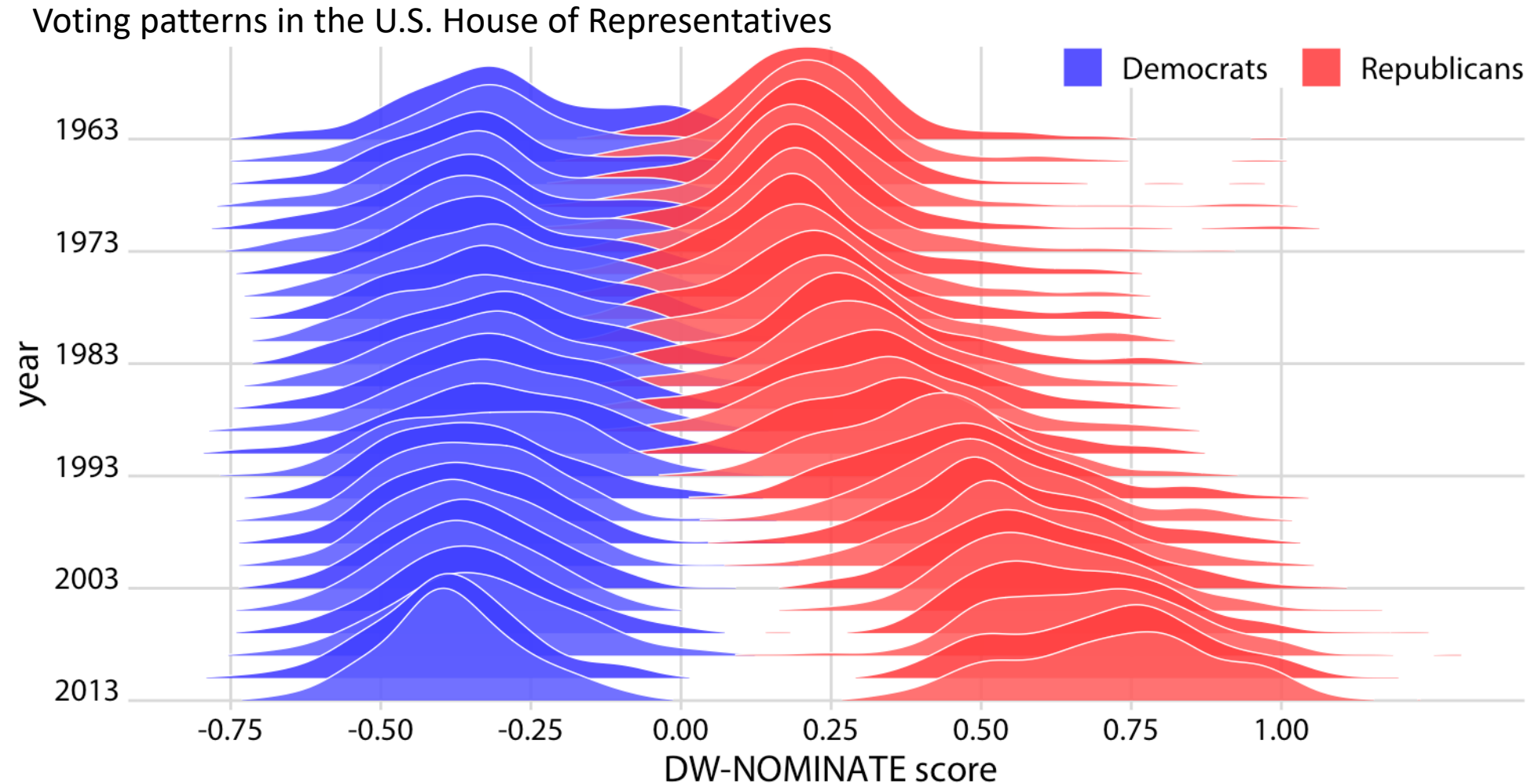
Ridgeline plot

Allow for easy comparison of density shapes and height across groups



Ridgeline plot

Especially effective at comparing two (or more) trends over time

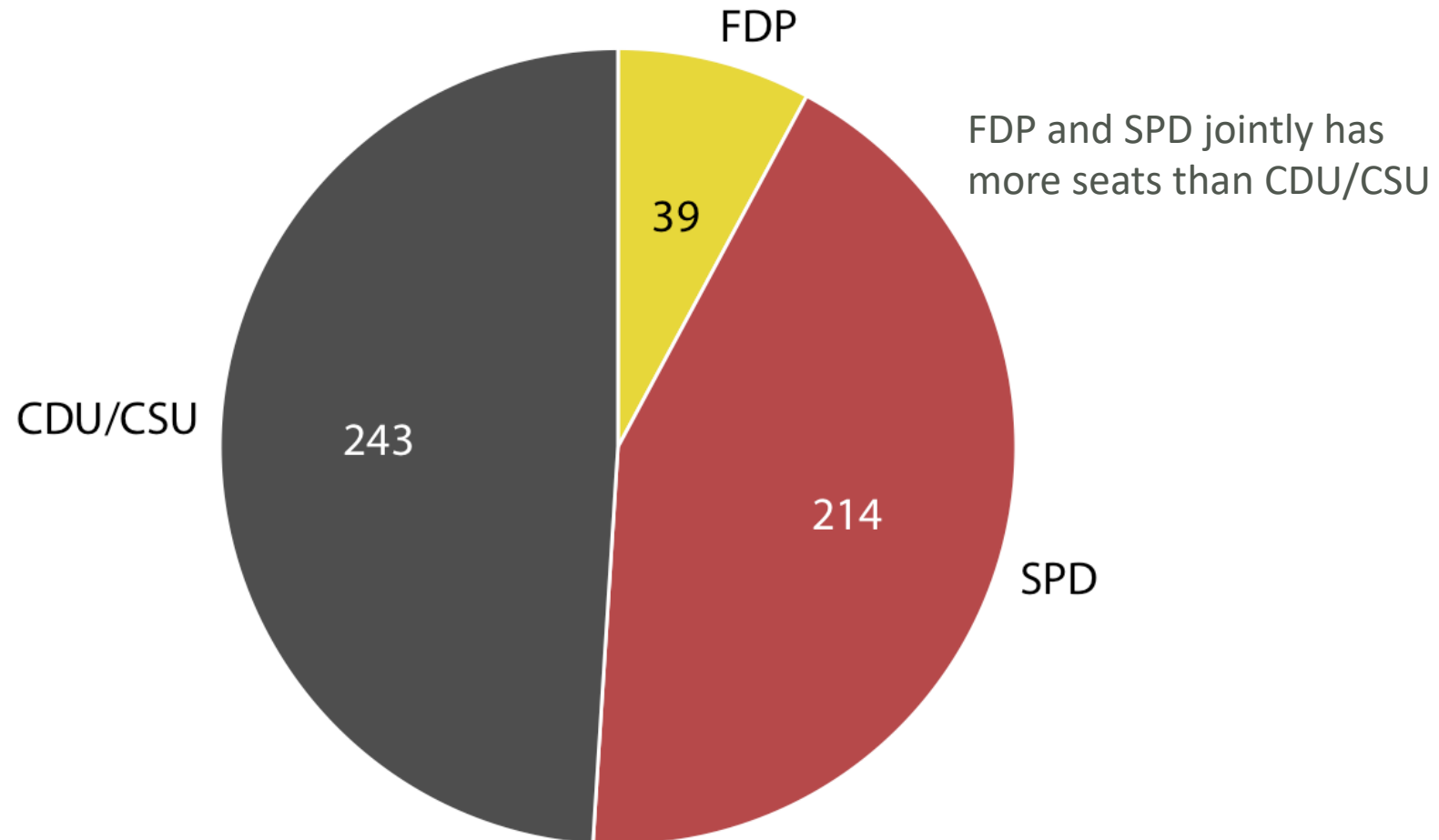


Visualizing proportions

Pie charts

Effective for visualizing data as proportion of a whole for few categories

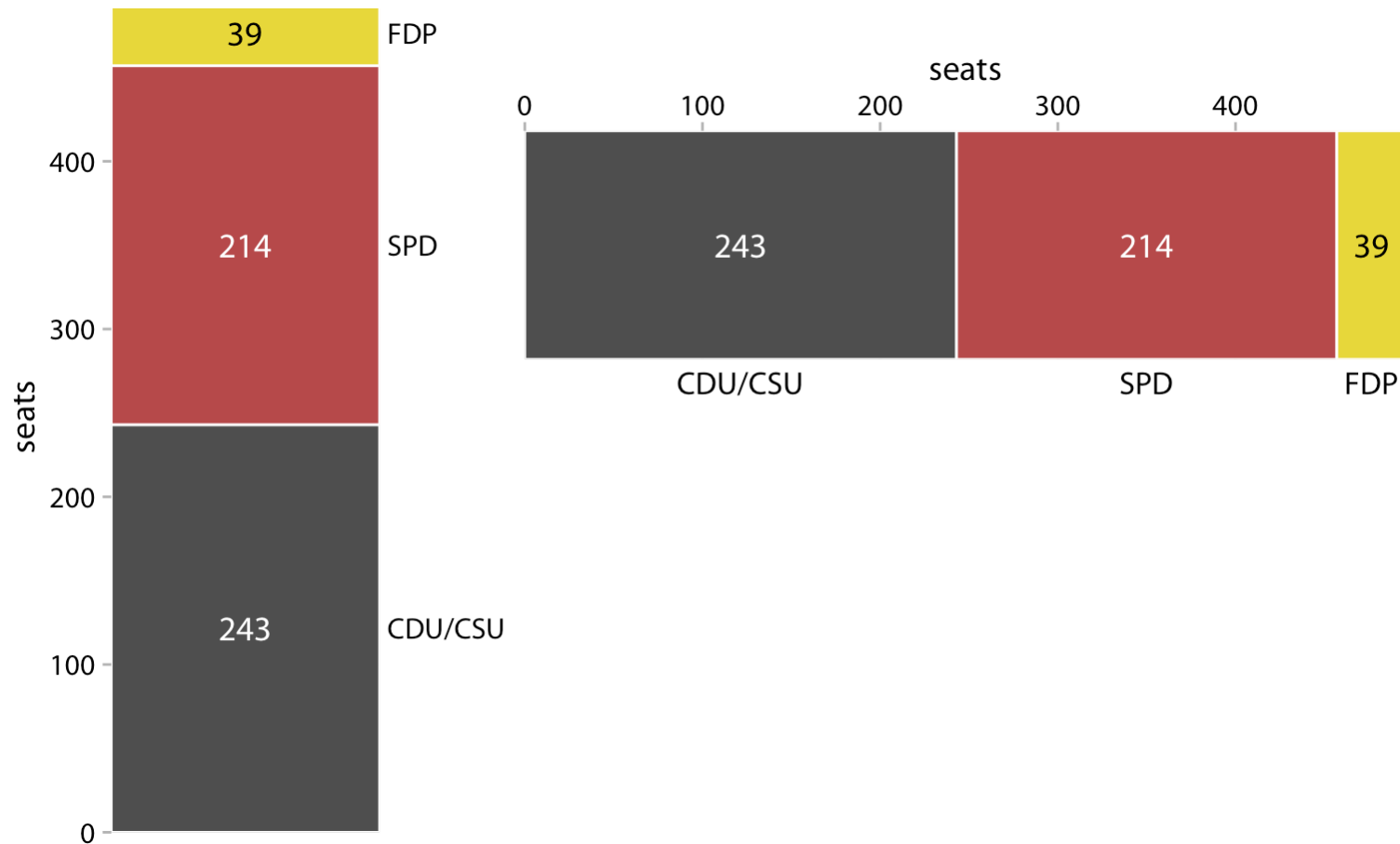
Party composition of the 8th German Bundestag, 1976–1980



Stacked bar plots

Work better for multiple side-by-side comparisons by condition or time

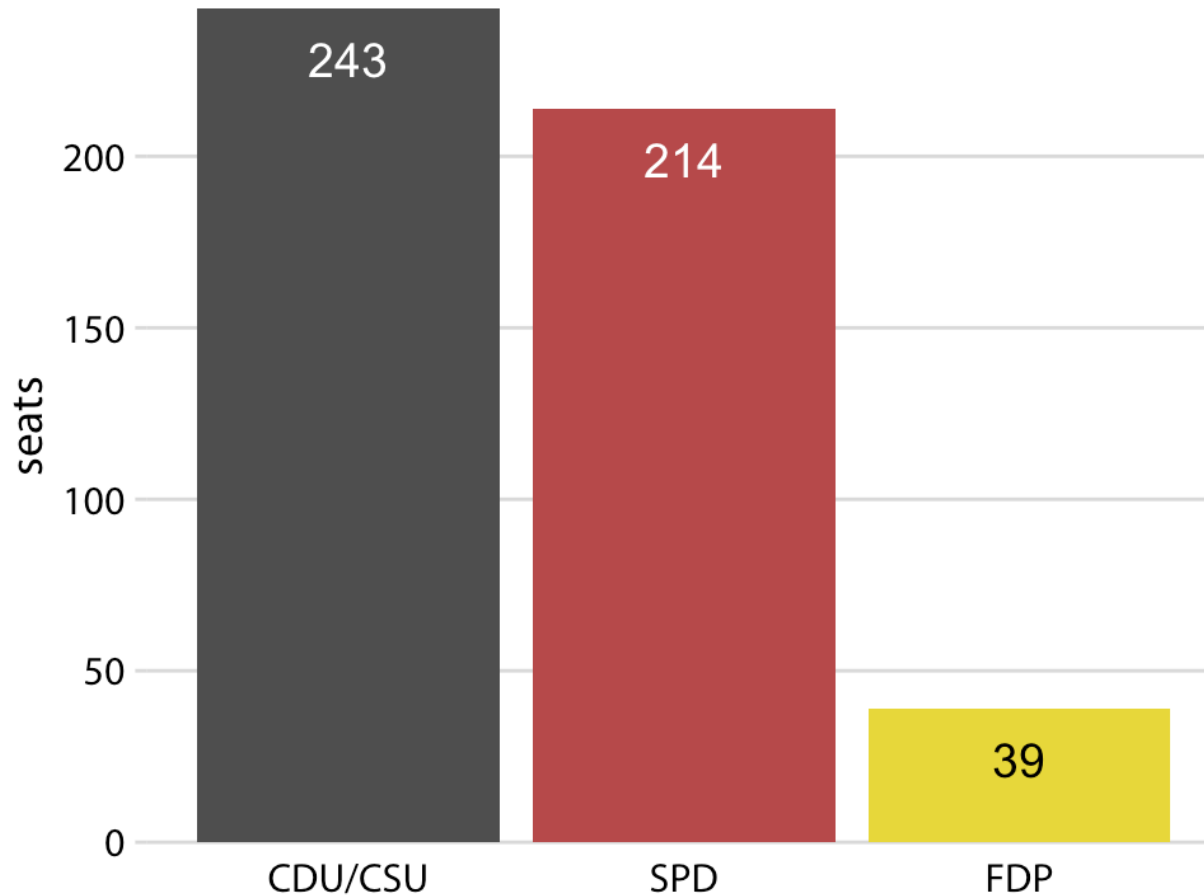
Party composition of the 8th German Bundestag, 1976–1980



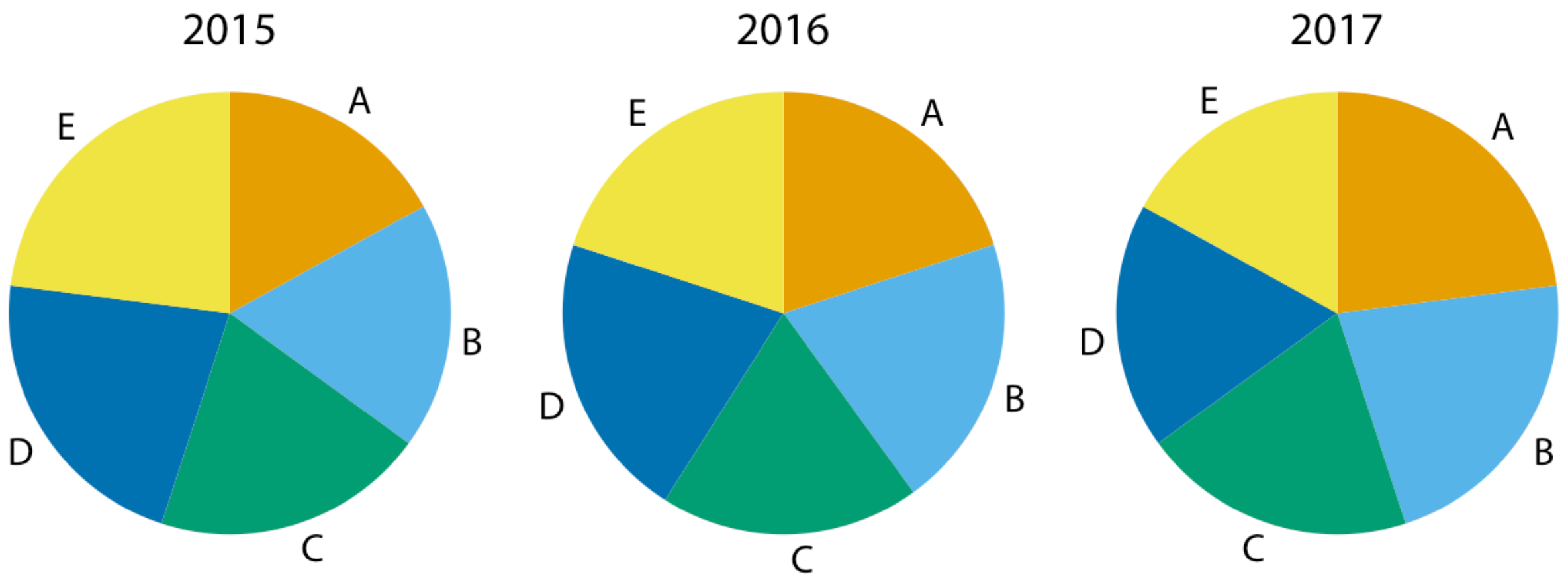
Bar plots

Better when we want to directly compare individual fractions

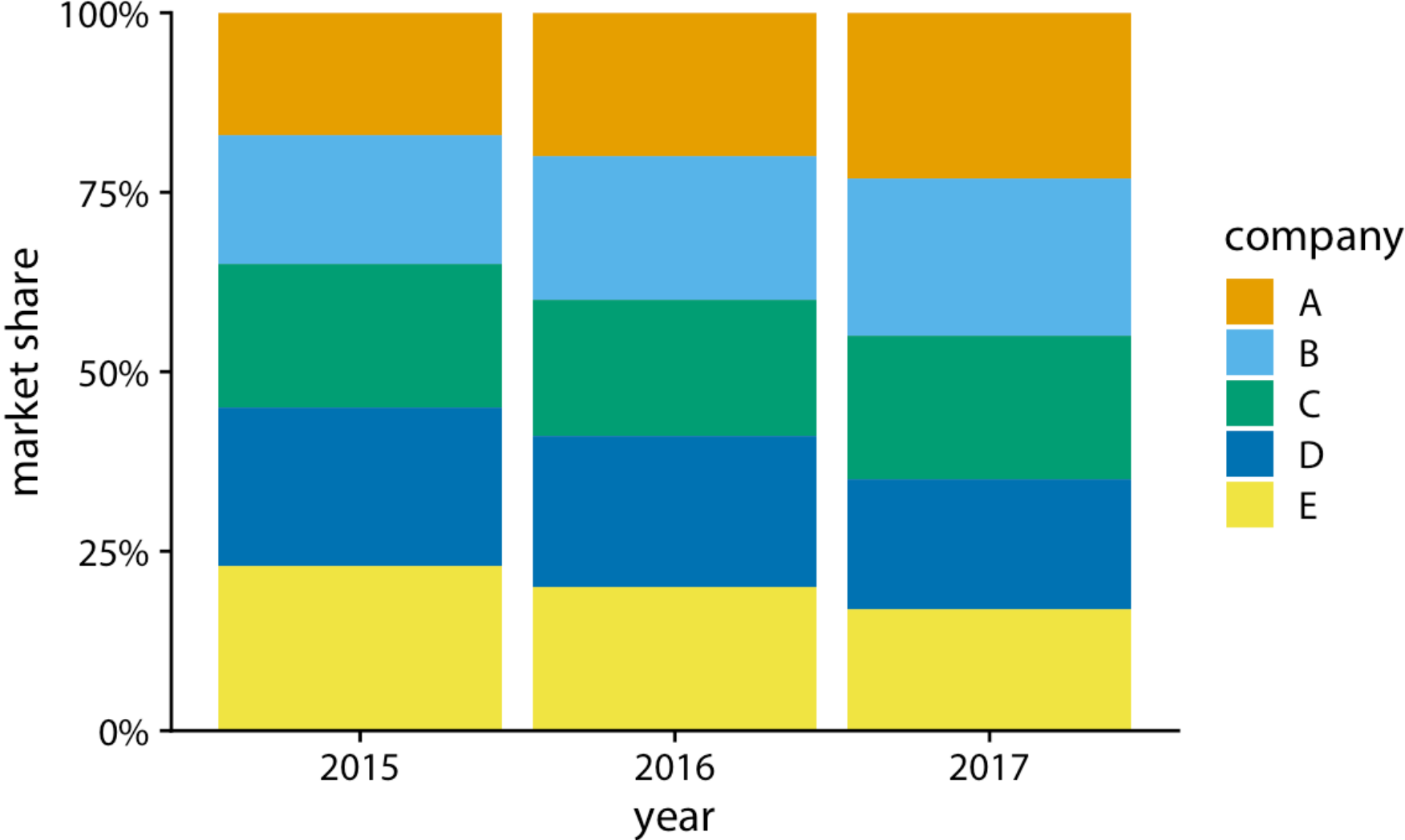
Party composition of the 8th German Bundestag, 1976–1980



Market share of five hypothetical companies, A–E, for the years 2015–2017



Market share of five hypothetical companies, A–E, for the years 2015–2017



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

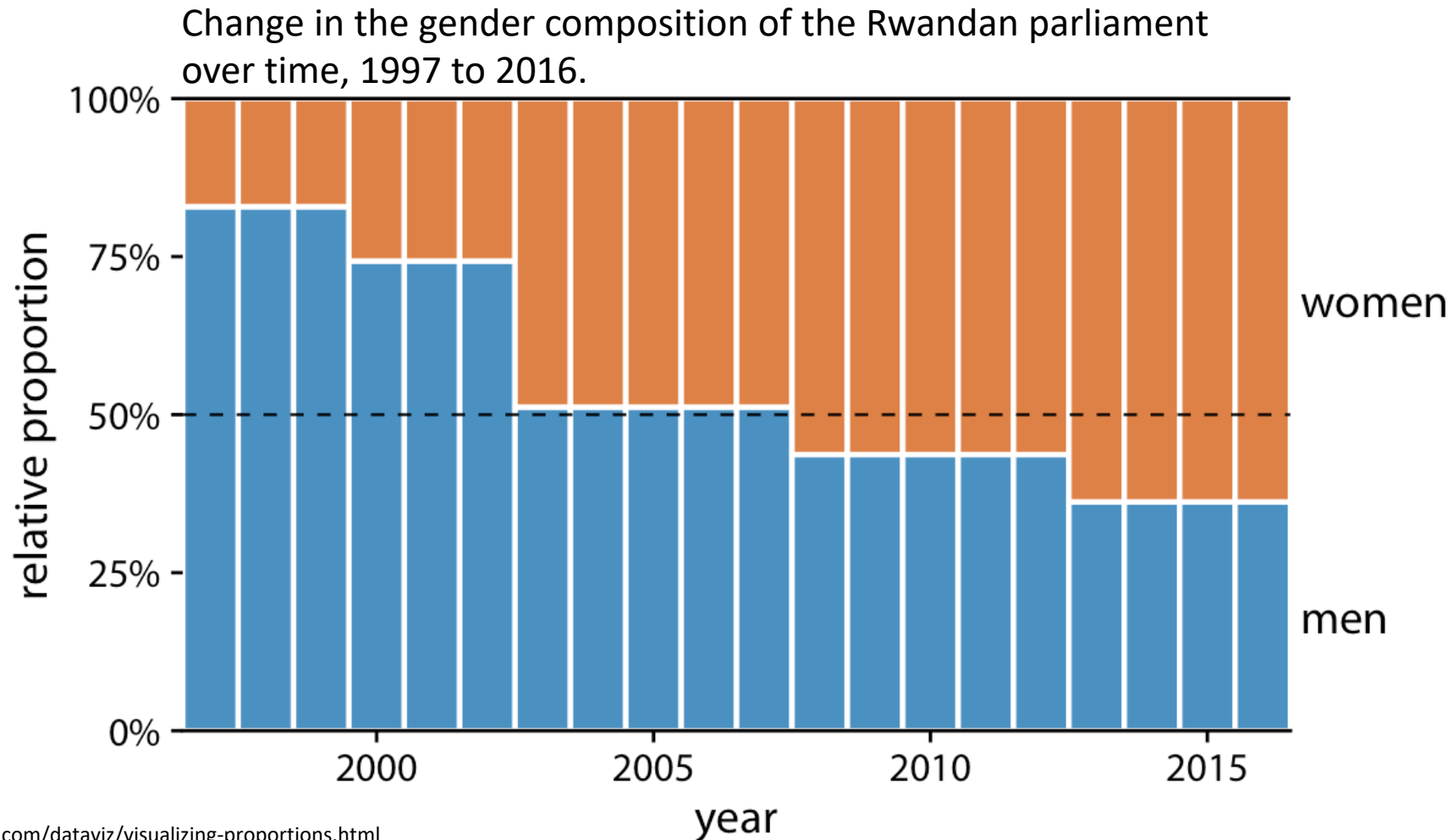
Market share of five hypothetical companies, A–E, for the years 2015–2017



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

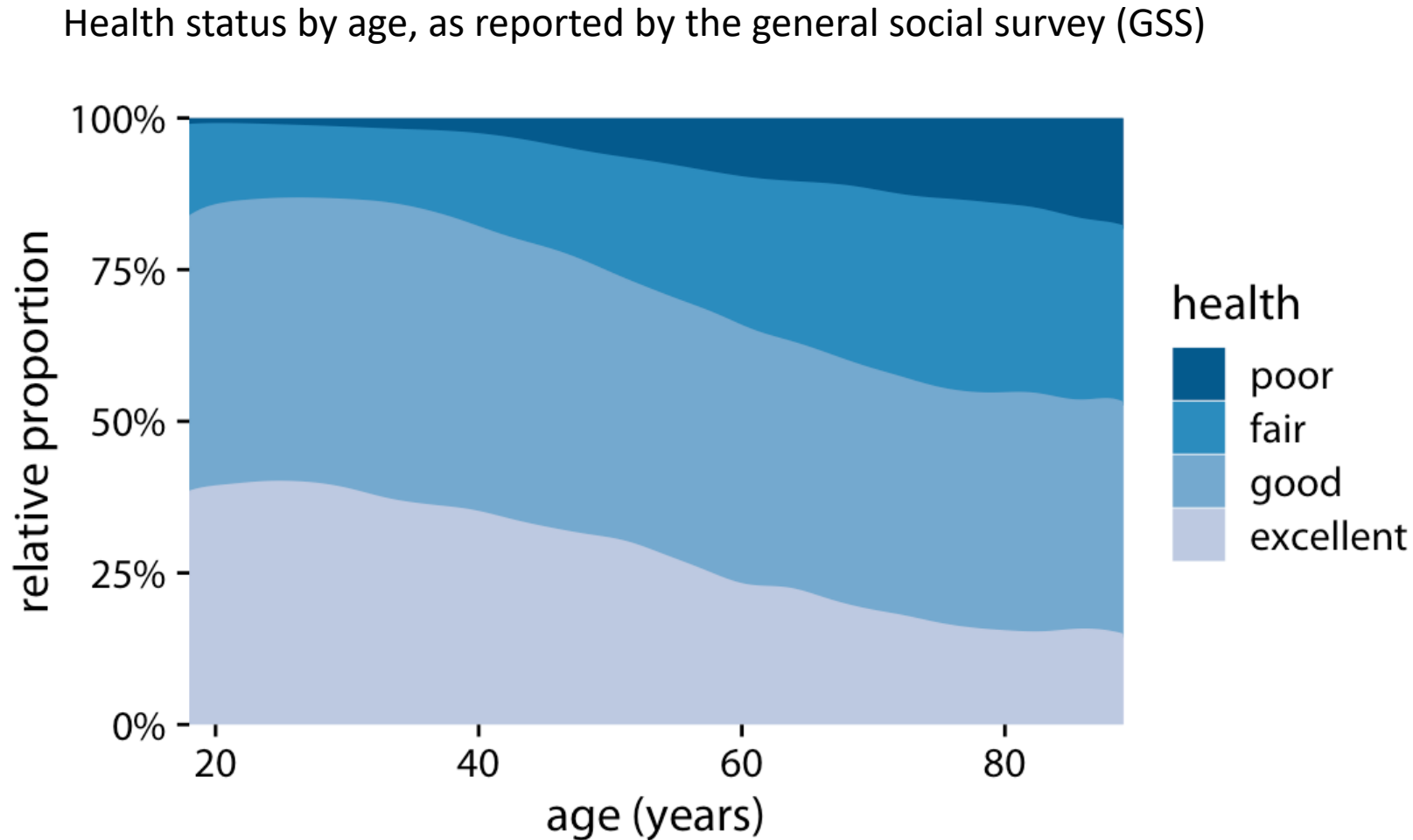
Stacked bar plots

Work better for multiple side-by-side comparisons by condition or time when there are only two categories



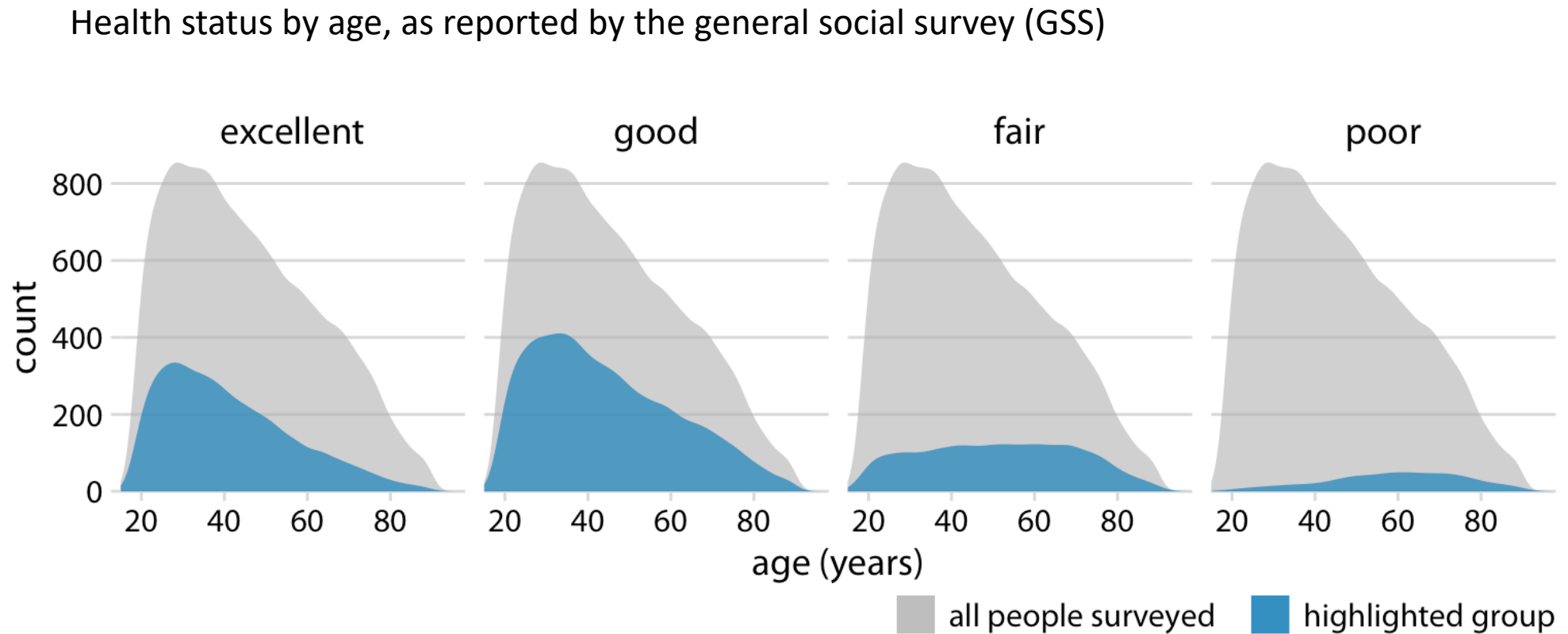
Stacked density plots

Show how proportions change in response to a continuous variable



Small multiples density plots

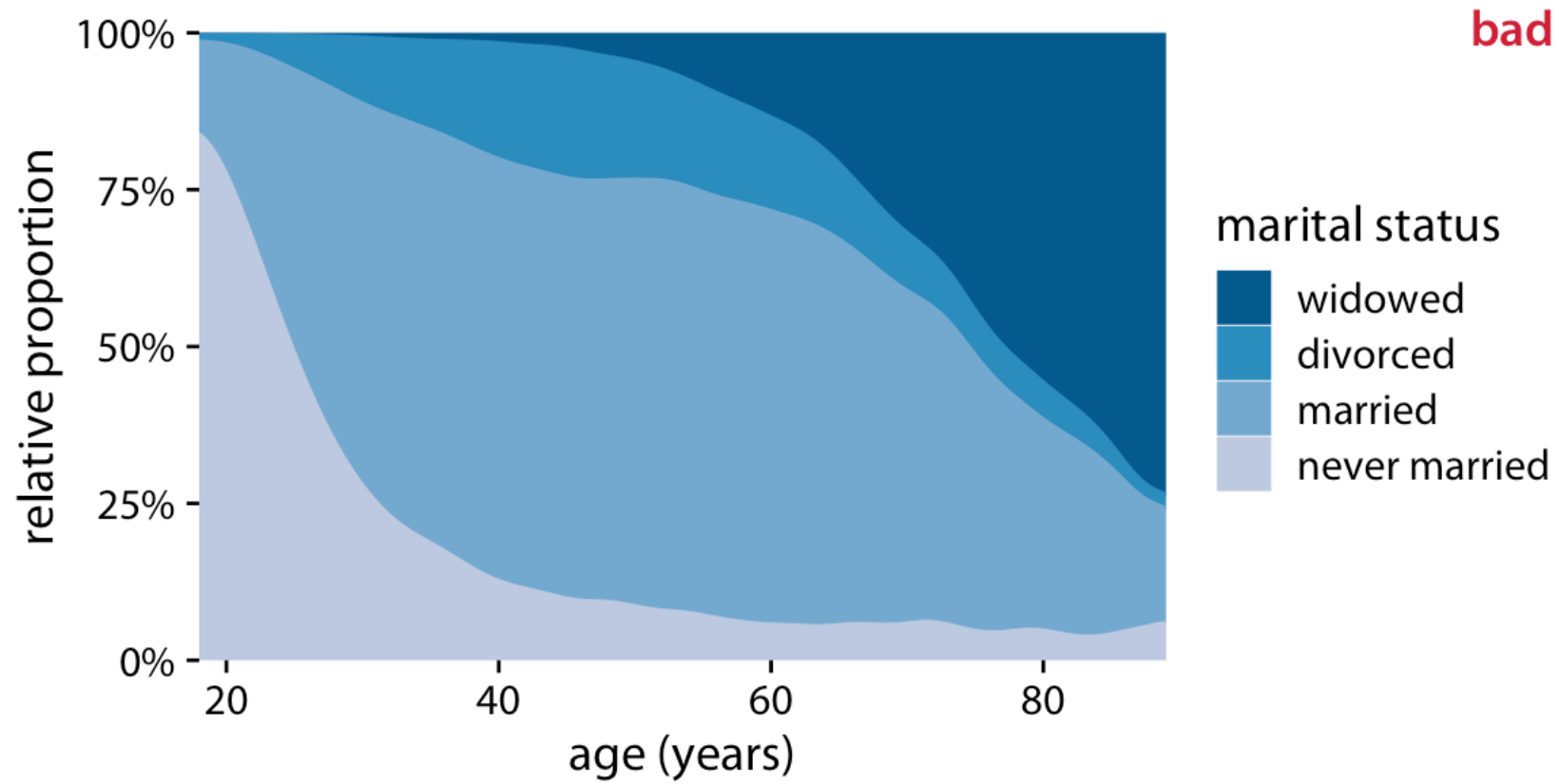
Show how proportions change in response to a continuous variable



Stacked density plots

Show how proportions change in response to a continuous variable

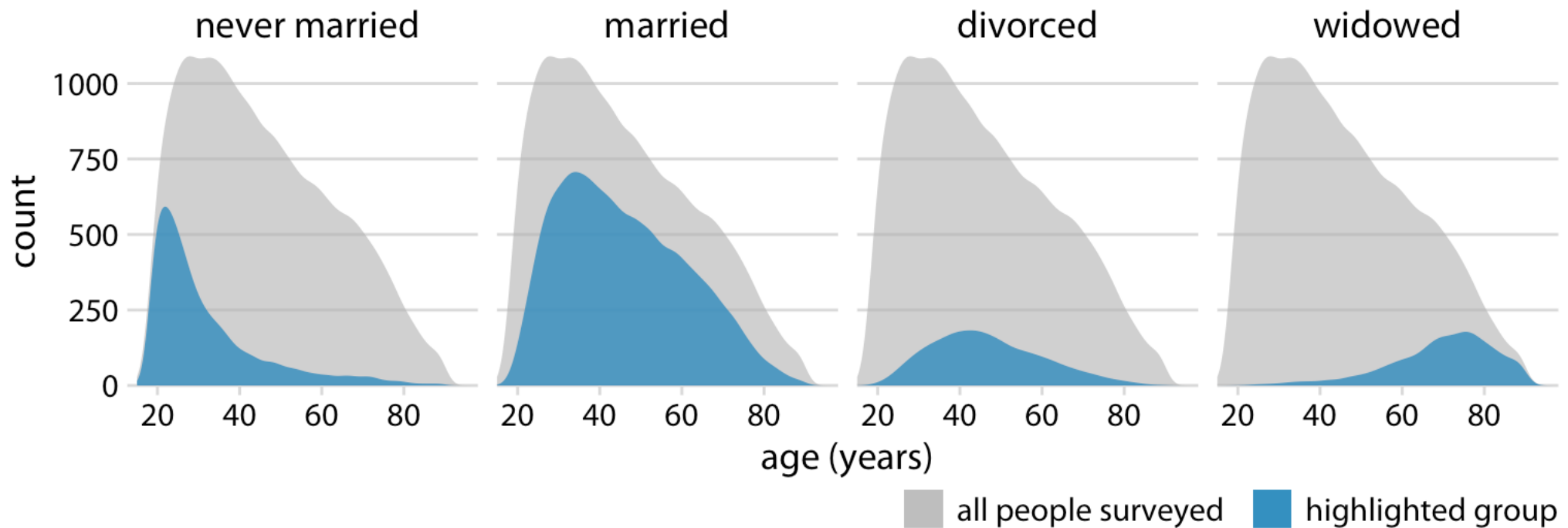
Marital status by age, as reported by the general social survey (GSS)



Small multiples density plots

Makes it harder to determine relative proportions

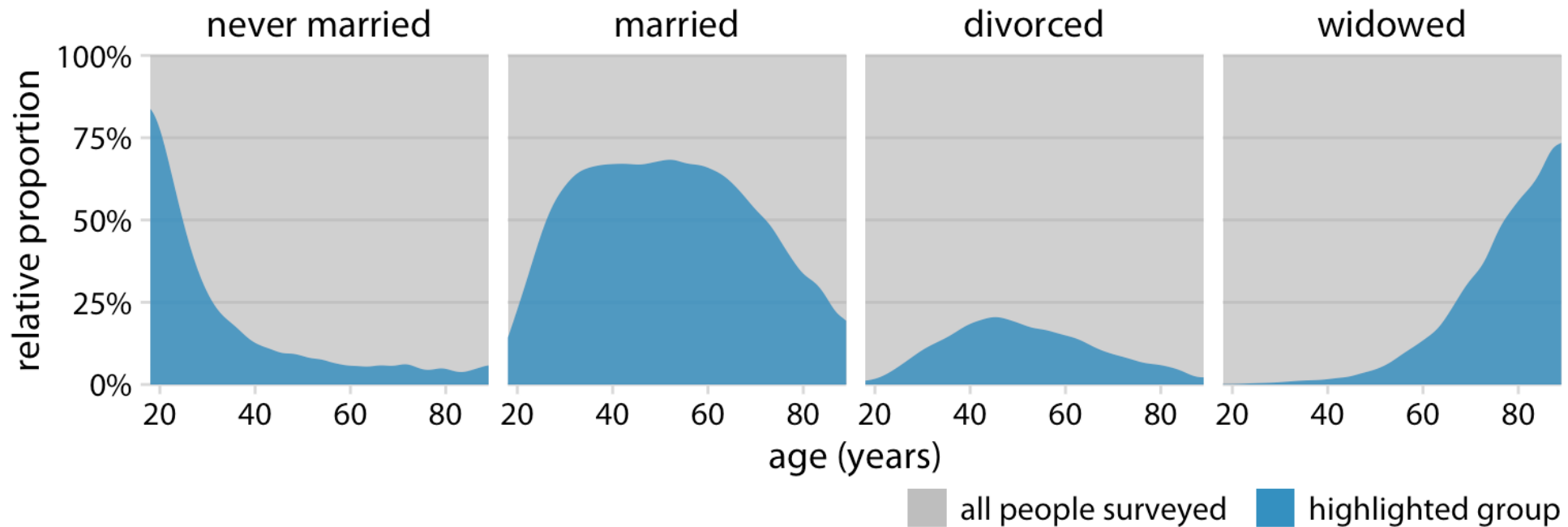
Marital status by age, as reported by the general social survey (GSS)



Small multiples density plots

Using a relative proportion makes proportion comparisons easy

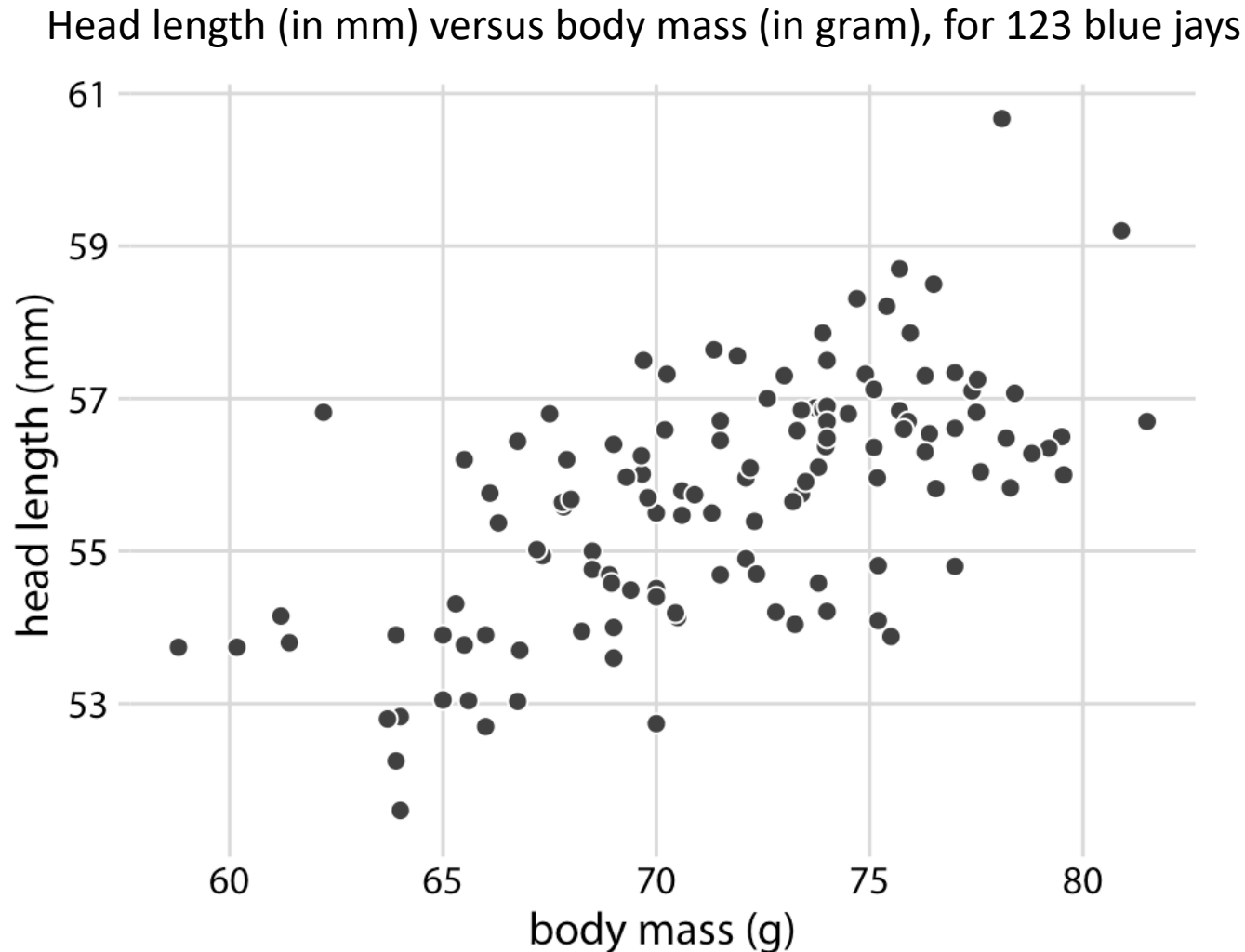
Marital status by age, as reported by the general social survey (GSS)



Visualizing associations between continuous variables

Scatter plot

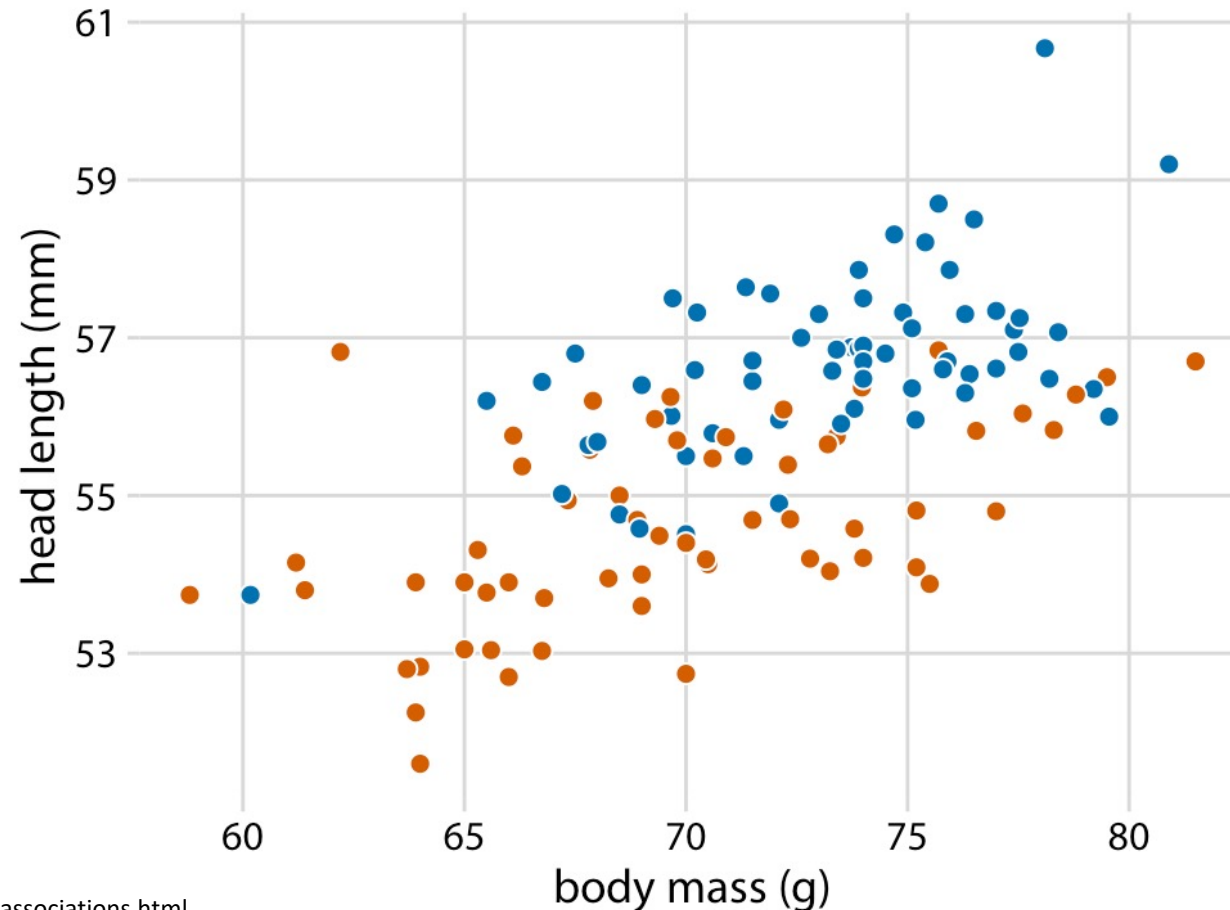
Effective for showing overall relationship between two variables



Scatter plot

Effective for showing overall relationship between two variables

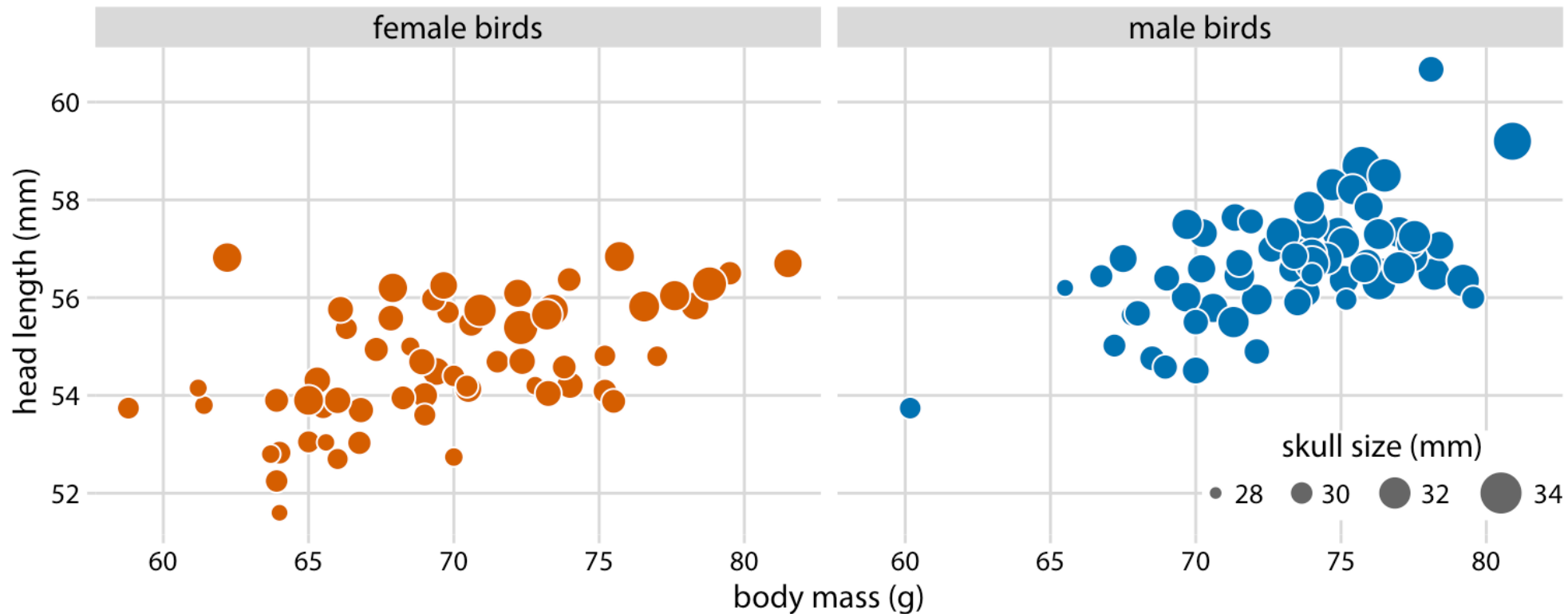
Head length (in mm) versus body mass (in gram), for 123 blue jays.
The birds' sex is indicated by color. ● female birds ● male birds



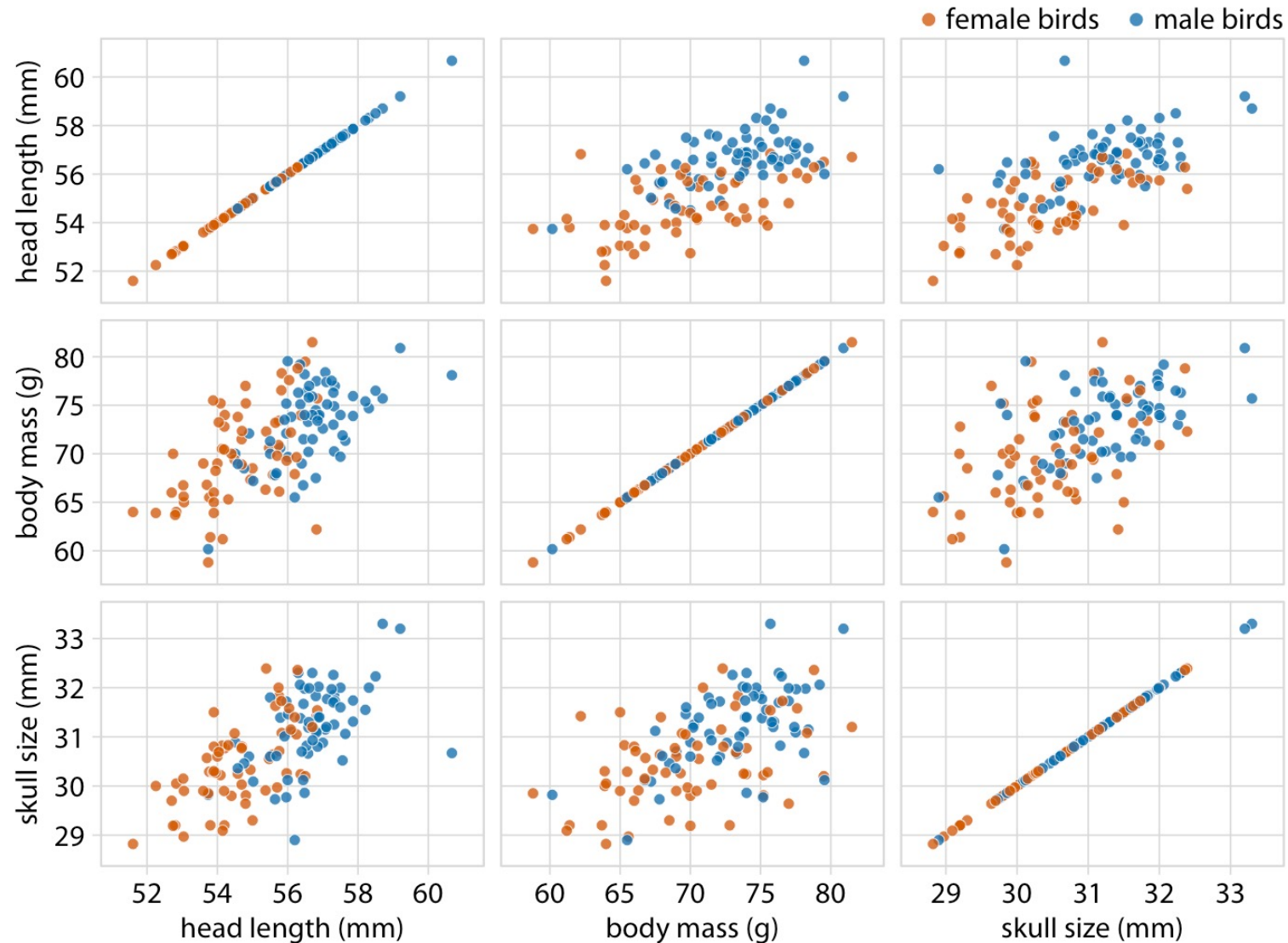
Bubble plot

Effective for showing encoding a third quantitative variable

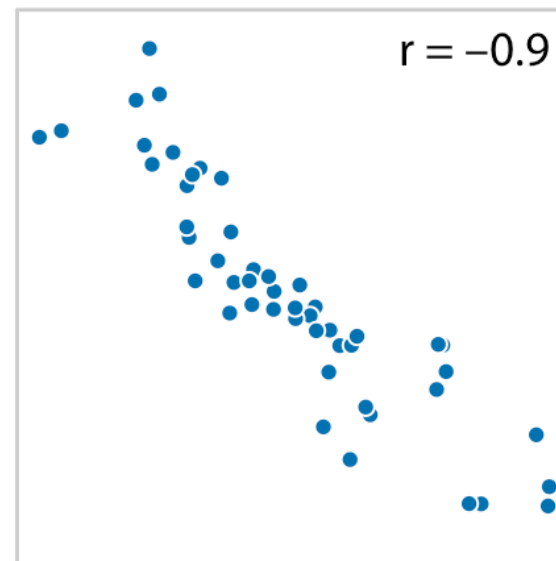
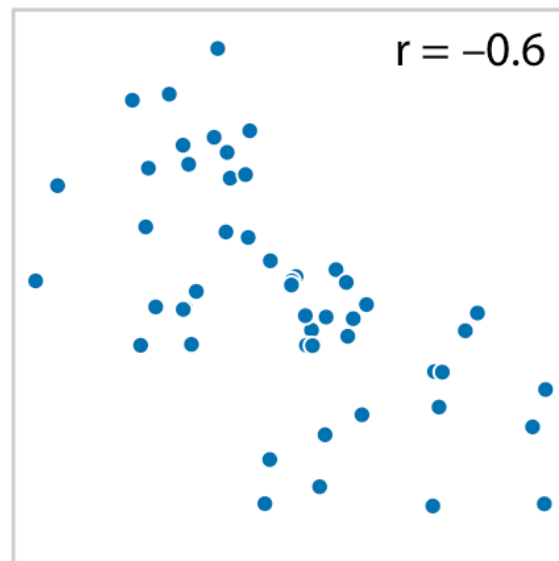
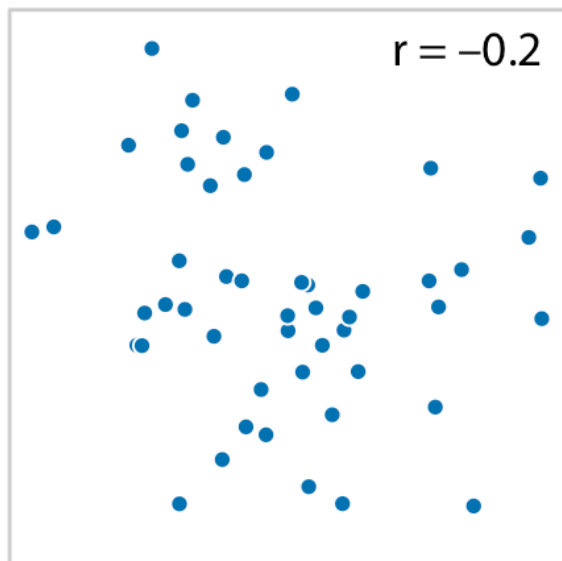
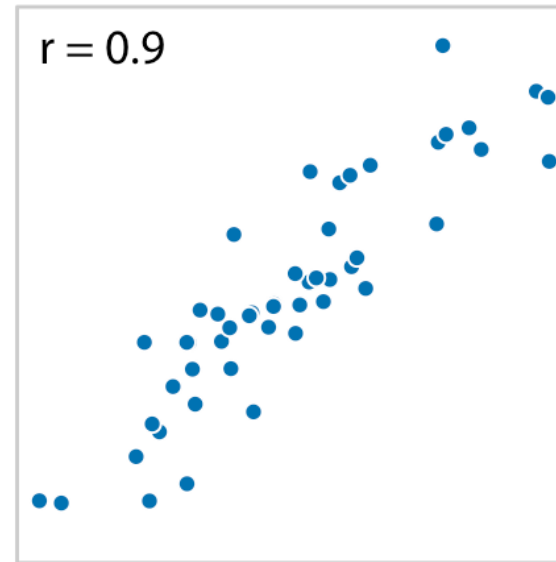
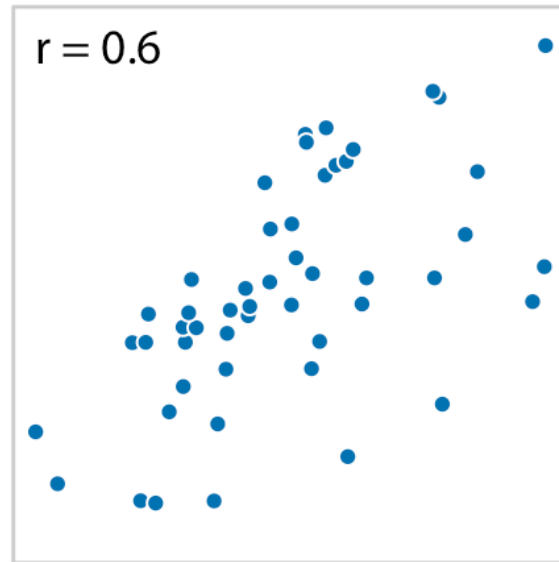
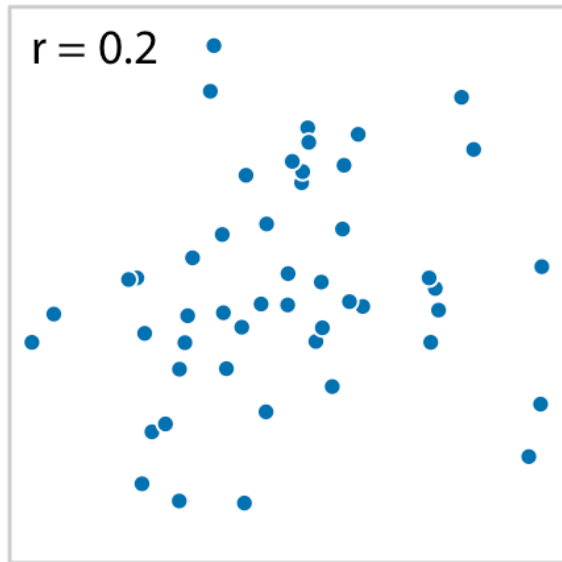
Head length (in mm) versus body mass (in gram), for 123 blue jays.
The birds' sex is indicated by color, and the birds' skull size by symbol size.



All-against-all scatter plot matrix



Summarize data with correlation coefficient (r)



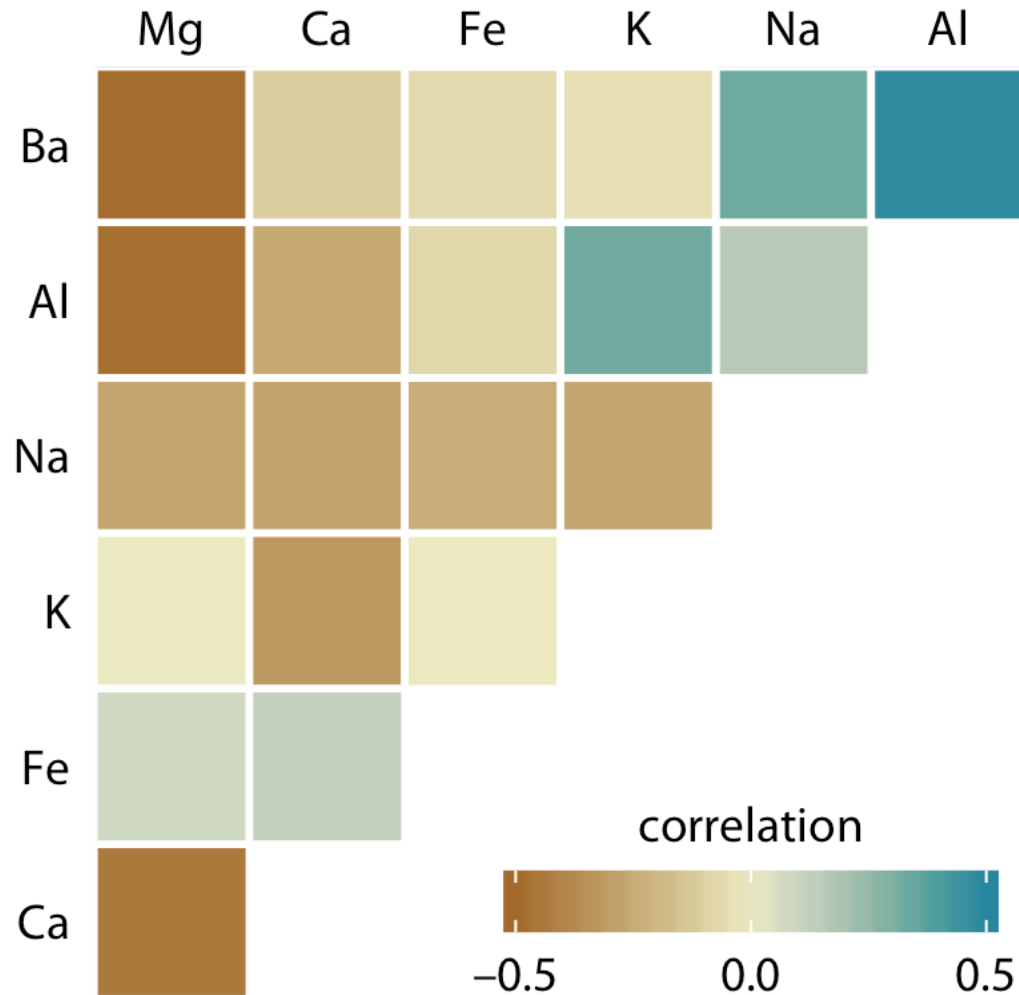
$r = 0$, no association

$r = 1$, perfect positive association

$r = -1$, perfect negative association

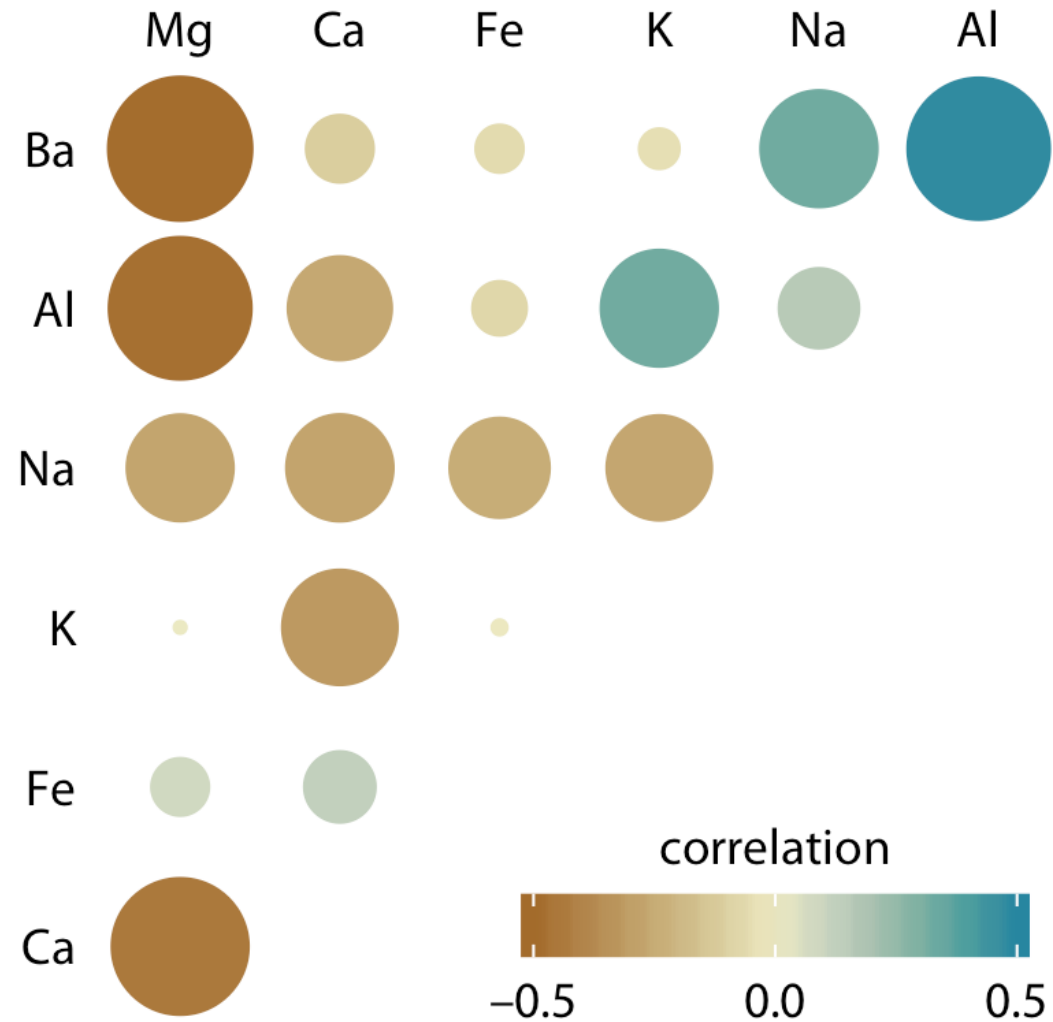
Correlogram

Correlations in mineral content for 214 samples of glass fragments obtained during forensic work.



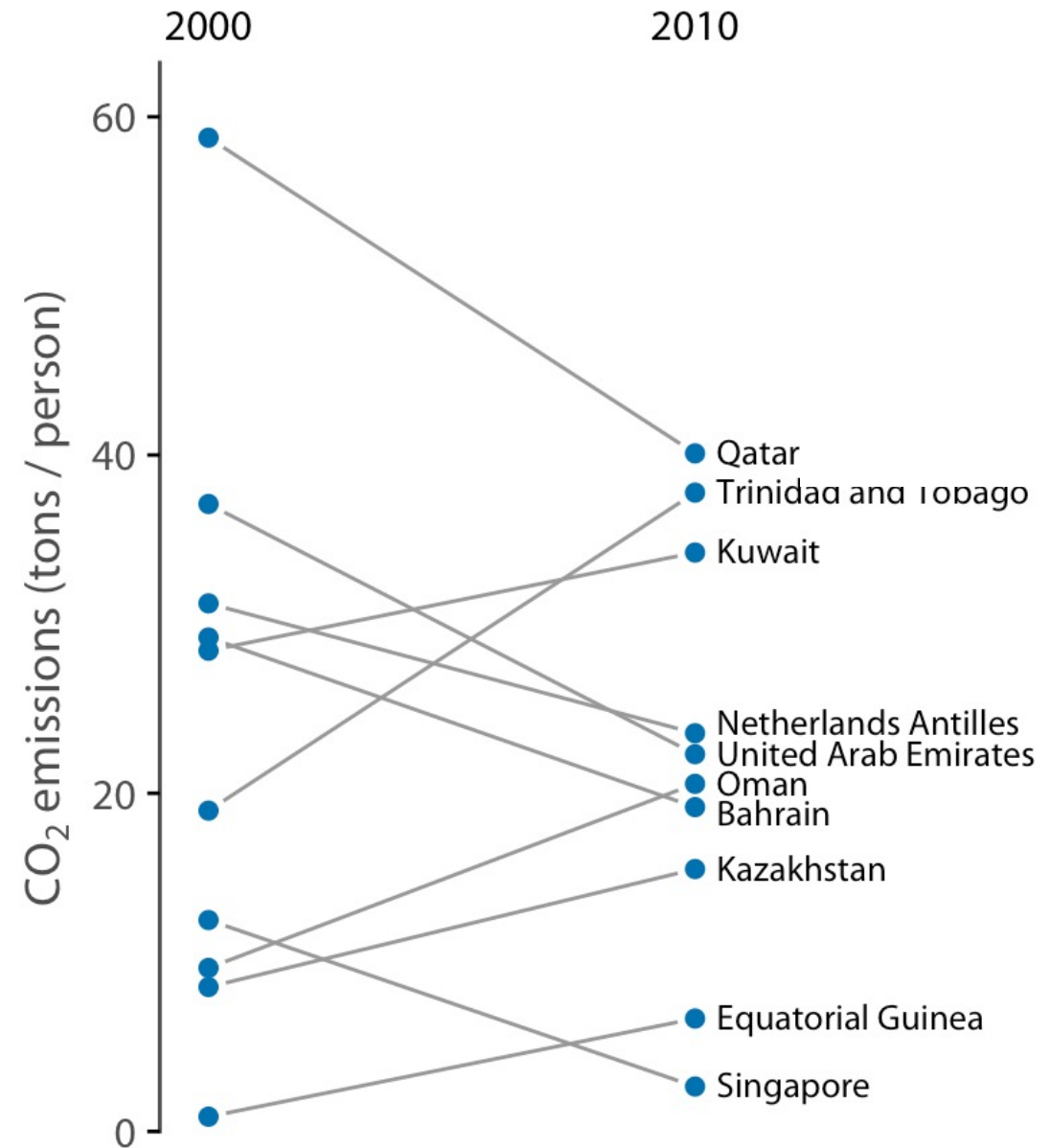
Correlogram

Correlations in mineral content for 214 samples of glass fragments obtained during forensic work.



Slopegraphs

Compare multiple measurements over time



Carbon dioxide (CO₂) emissions per person in 2000 and 2010, for the ten countries with the largest difference between these two years.

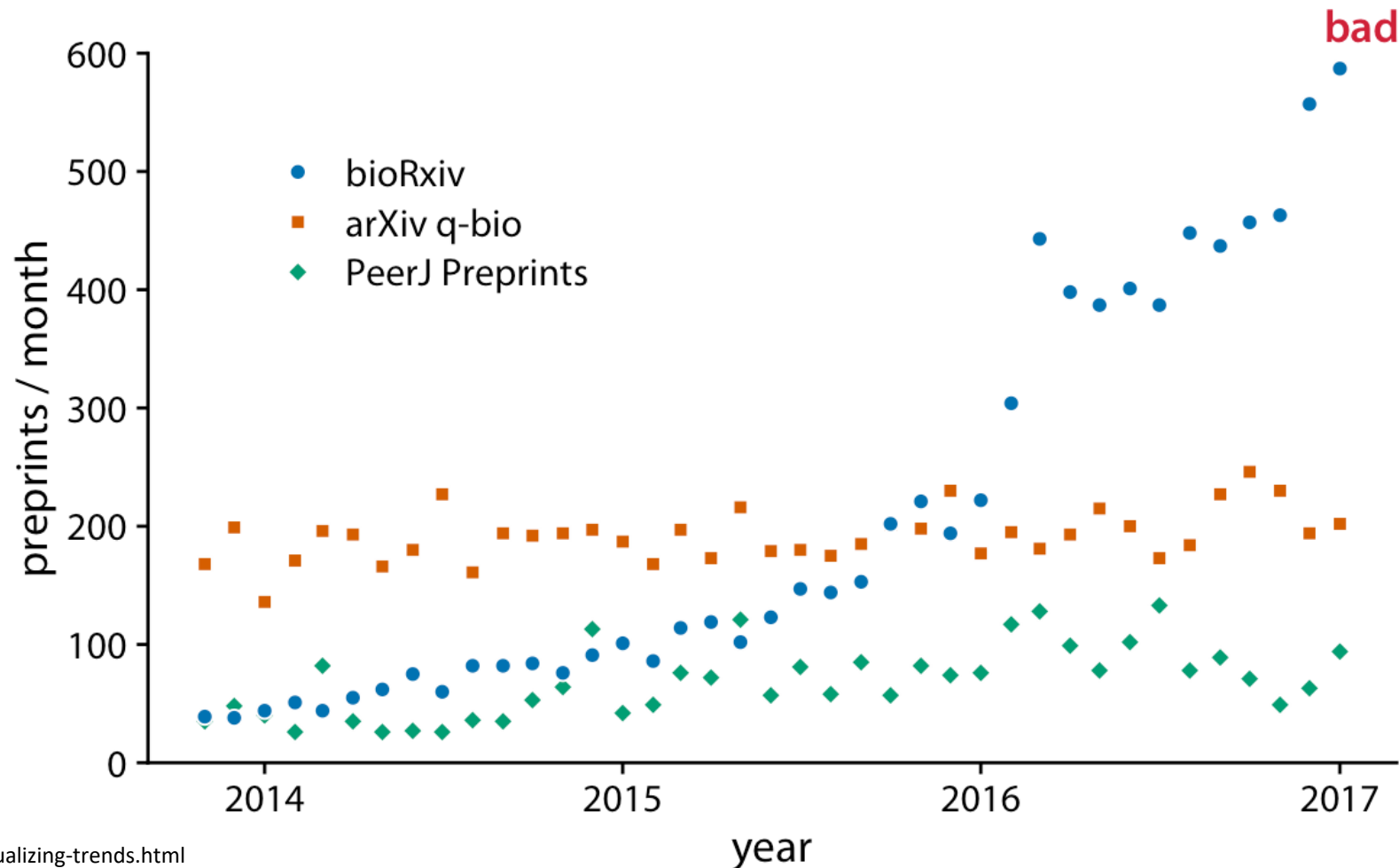
Visualizing time series

Visualize data with quantitative data with one independent variable (like time) that imposes an inherent ordering

Time series

Compare multiple measurements over time

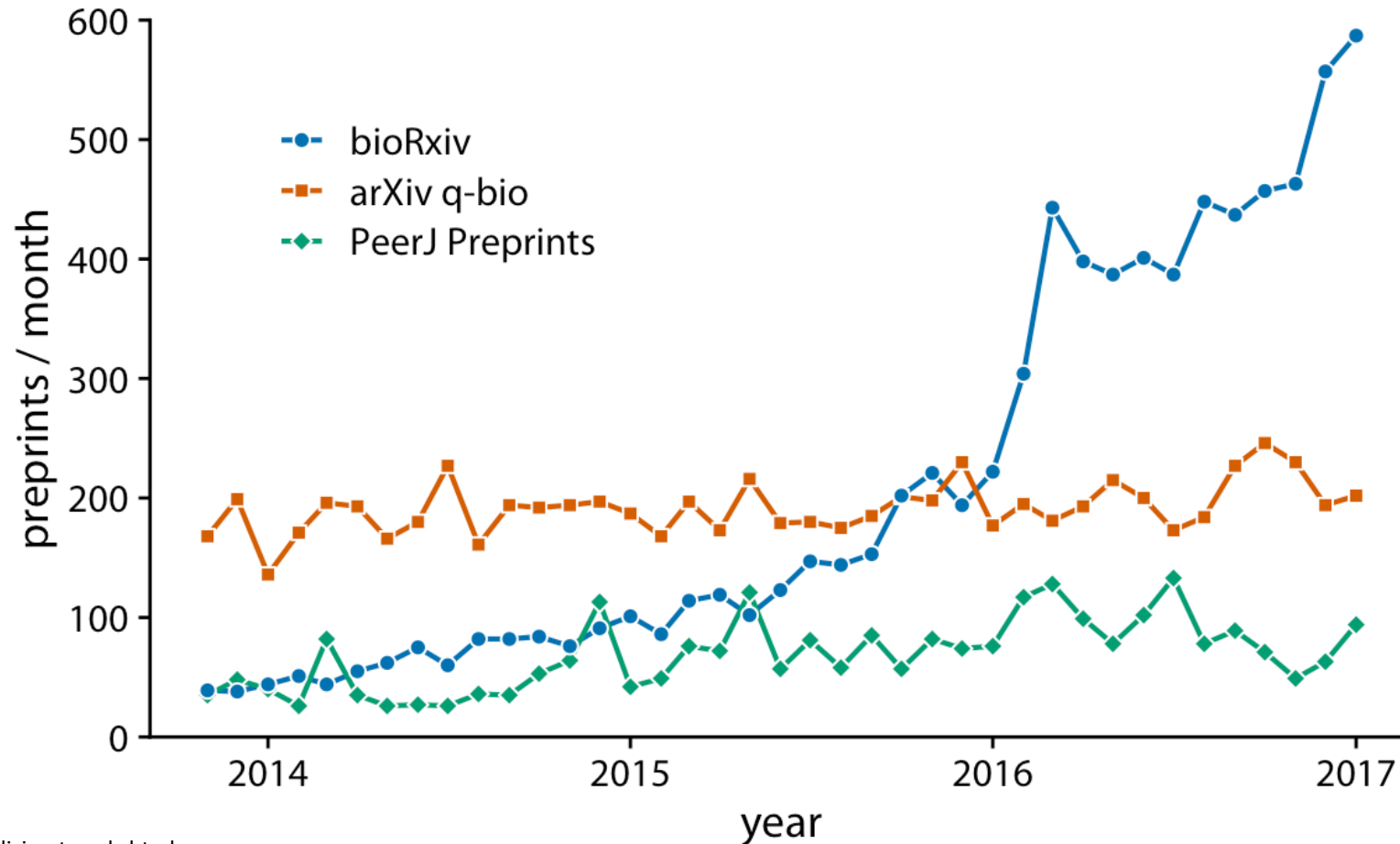
Monthly submissions to three preprint servers covering biomedical research: bioRxiv, the q-bio section of arXiv, and PeerJ Preprints.



Time series

Compare multiple measurements over time

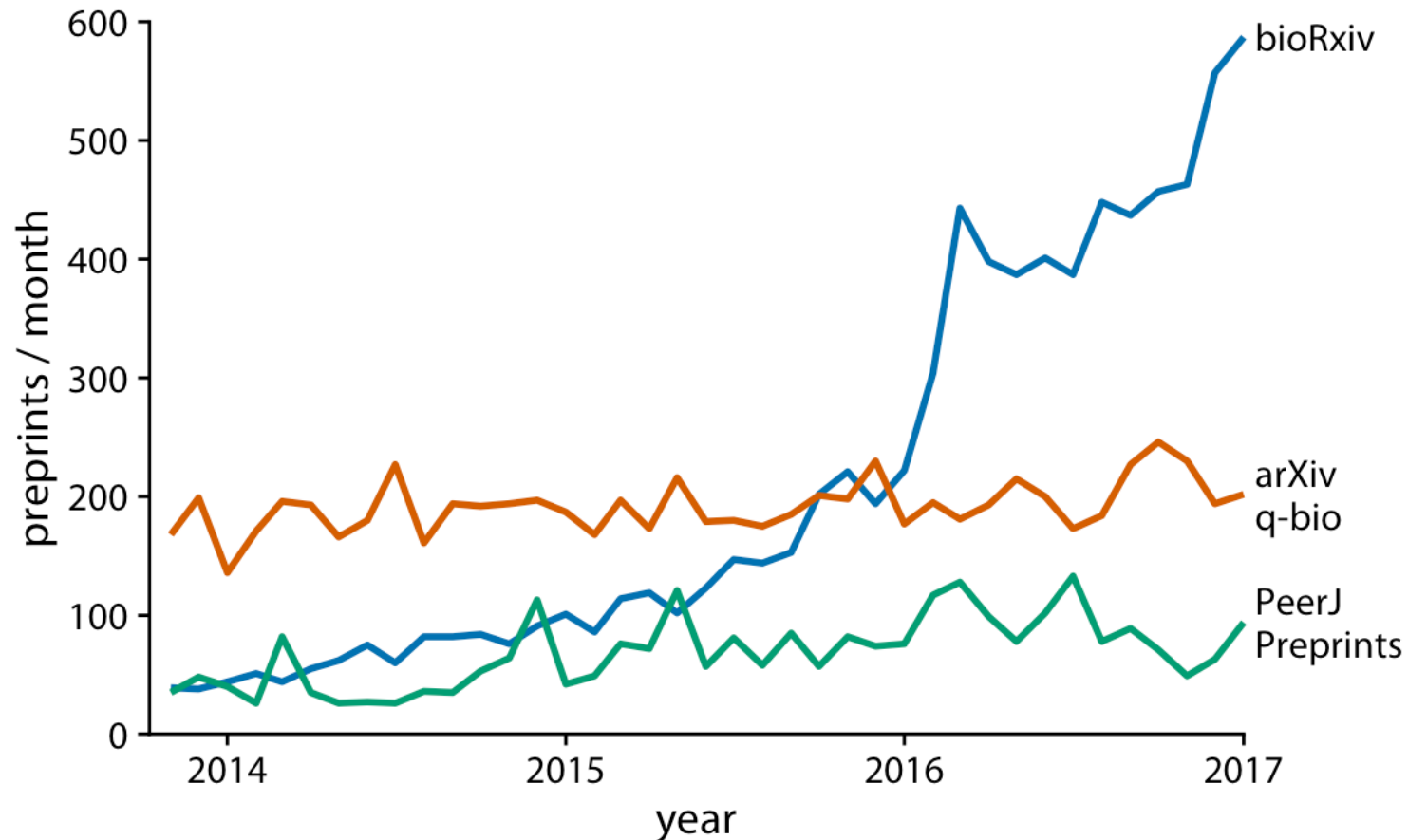
Monthly submissions to three preprint servers covering biomedical research: bioRxiv, the q-bio section of arXiv, and PeerJ Preprints.



Time series

Compare multiple measurements over time

Monthly submissions to three preprint servers covering biomedical research: bioRxiv, the q-bio section of arXiv, and PeerJ Preprints.



Visualizing trends

Smoothed time series, effective for highlighting overall trend

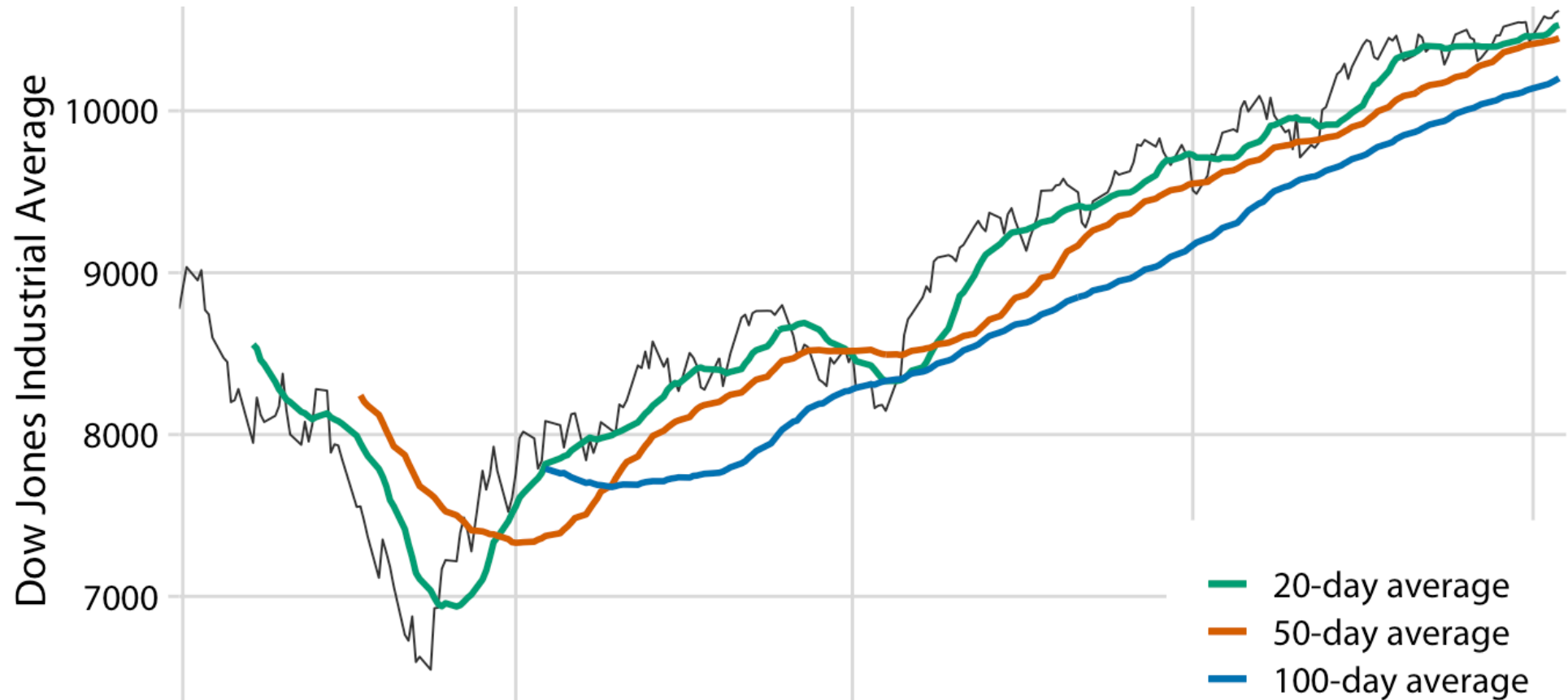
Daily closing values of the Dow Jones Industrial Average for the year 2009.



Moving average

Compute a sequence of averaged values

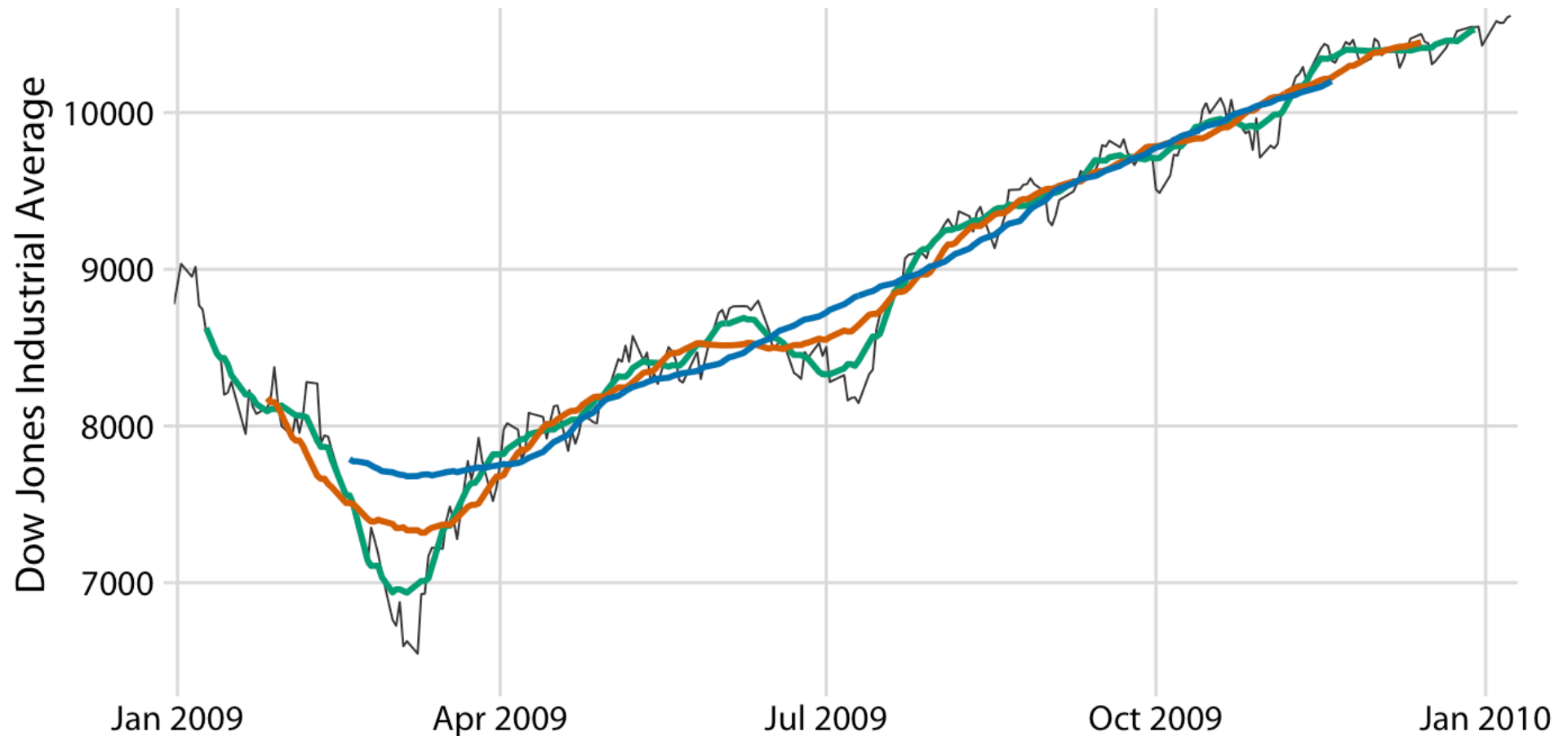
Daily closing values of the Dow Jones Industrial Average for the year 2009.



Moving average

Plot the average at the center of the time window

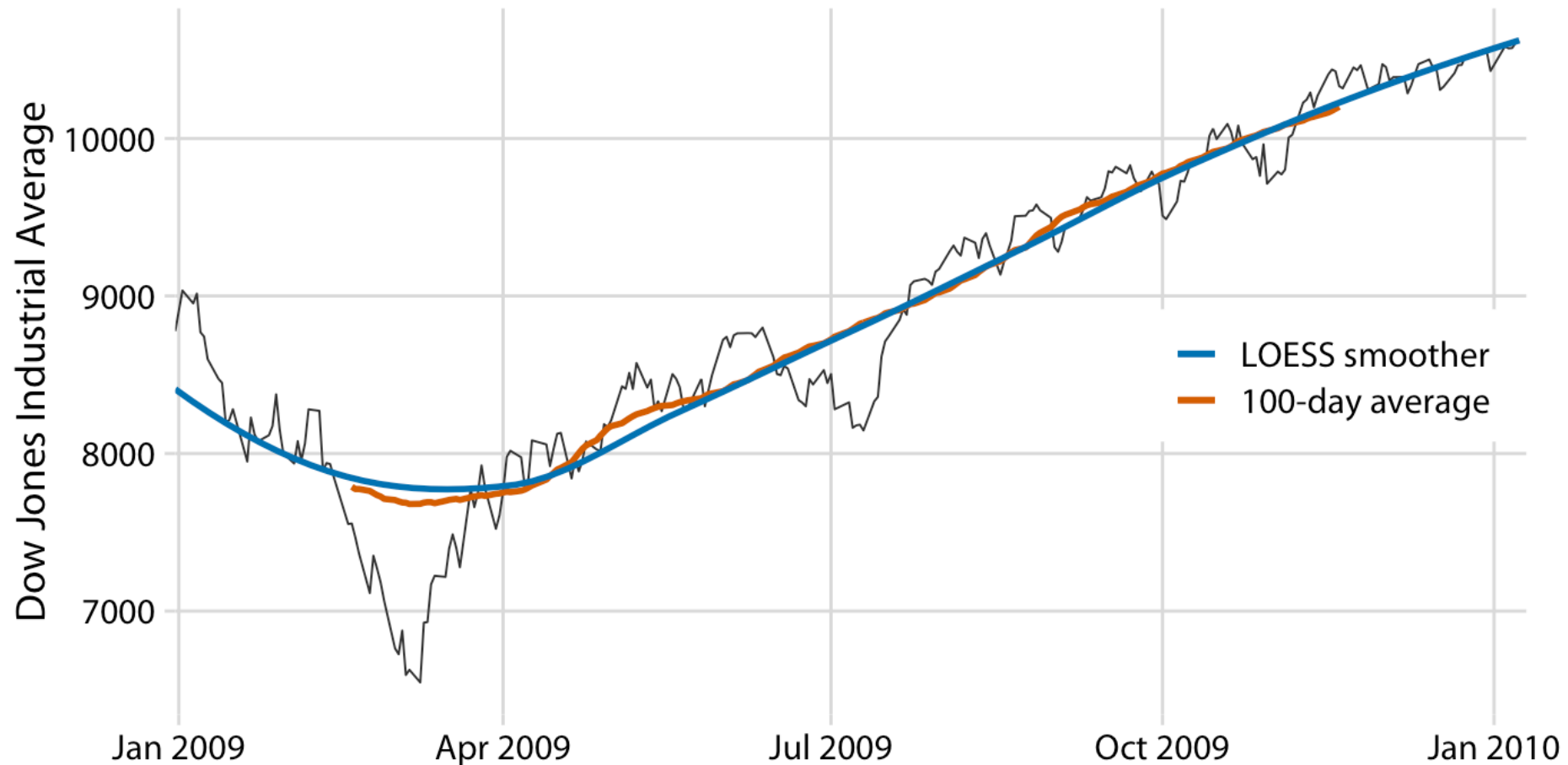
Daily closing values of the Dow Jones Industrial Average for the year 2009.



LOESS Smoother

Locally estimated scatterplot smoothing

Daily closing values of the Dow Jones Industrial Average for the year 2009.



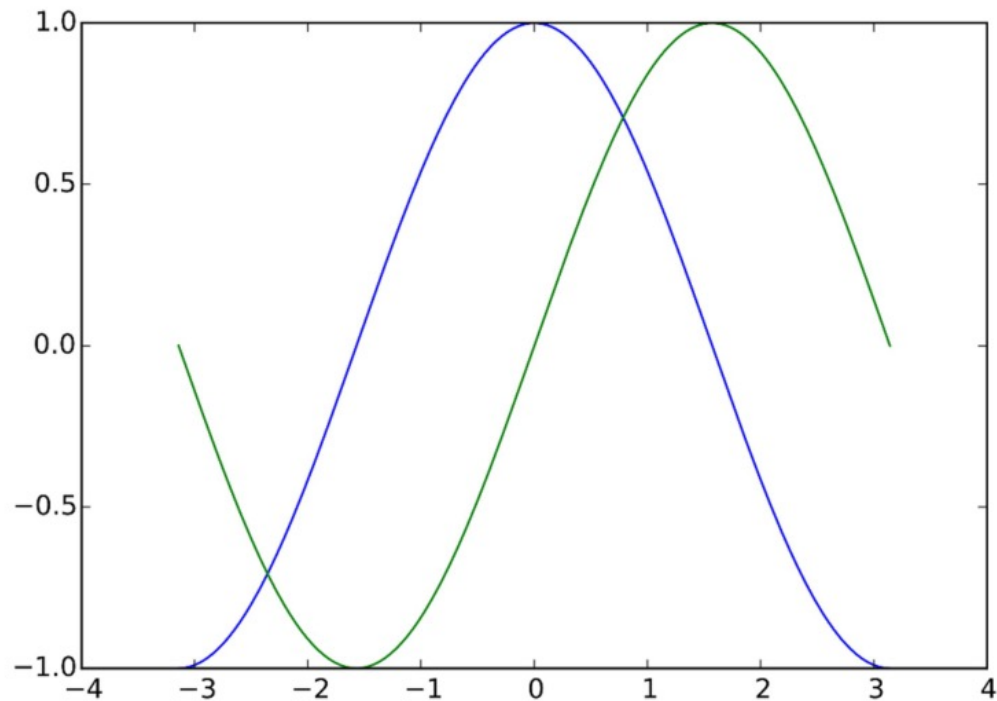
Design tips for effective visualizations

Don't trust the defaults

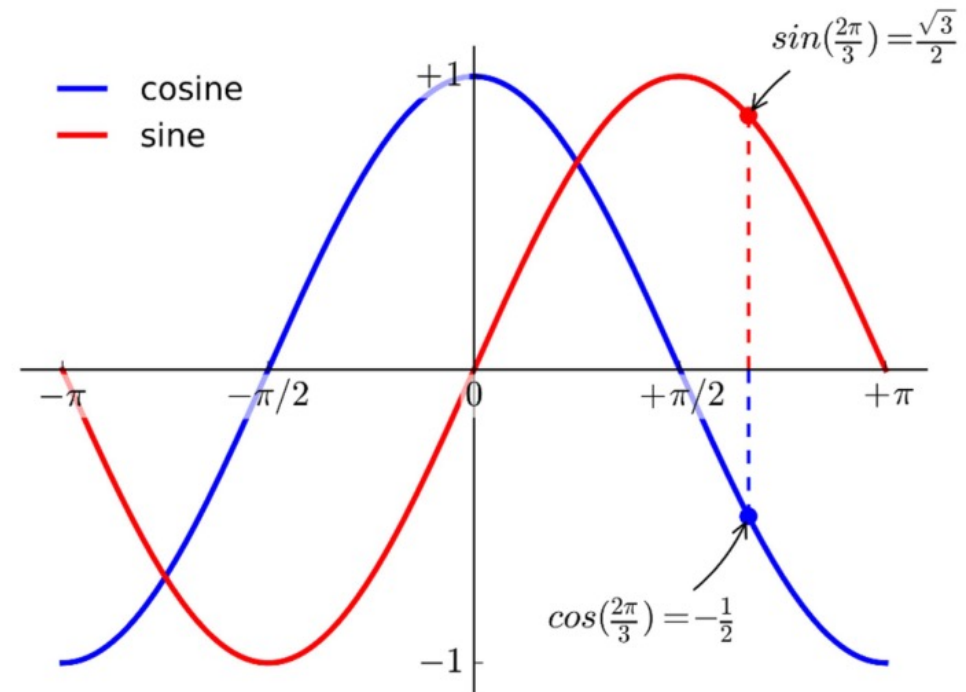
No software is perfect.

If you can tell what software created your data visualization, you likely need to **rework** your visualization

Don't trust the defaults



matplotlib defaults



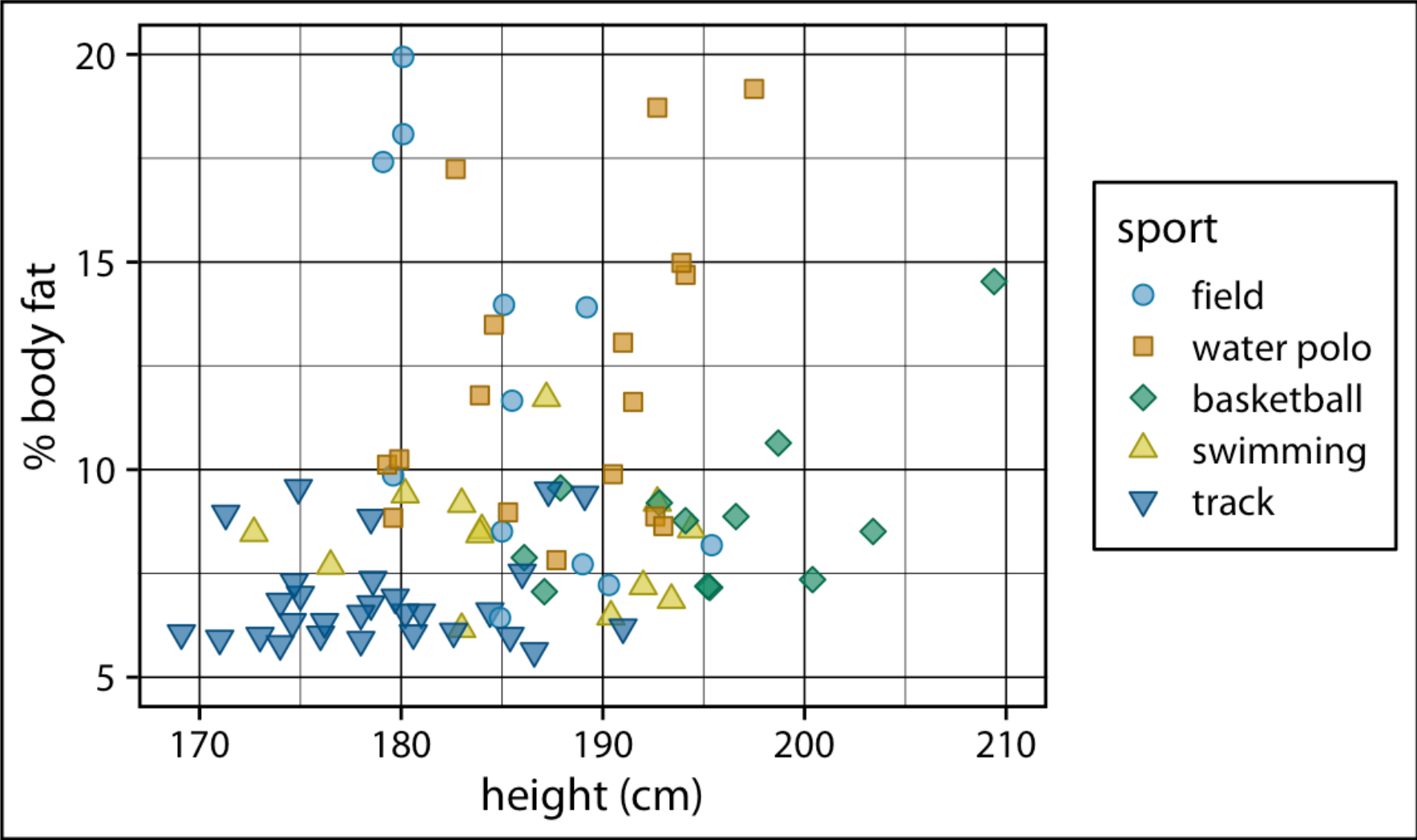
Visually improved plot

Keep it simple

Eliminate visual clutter. If a component of a visualization doesn't encode information, remove it!

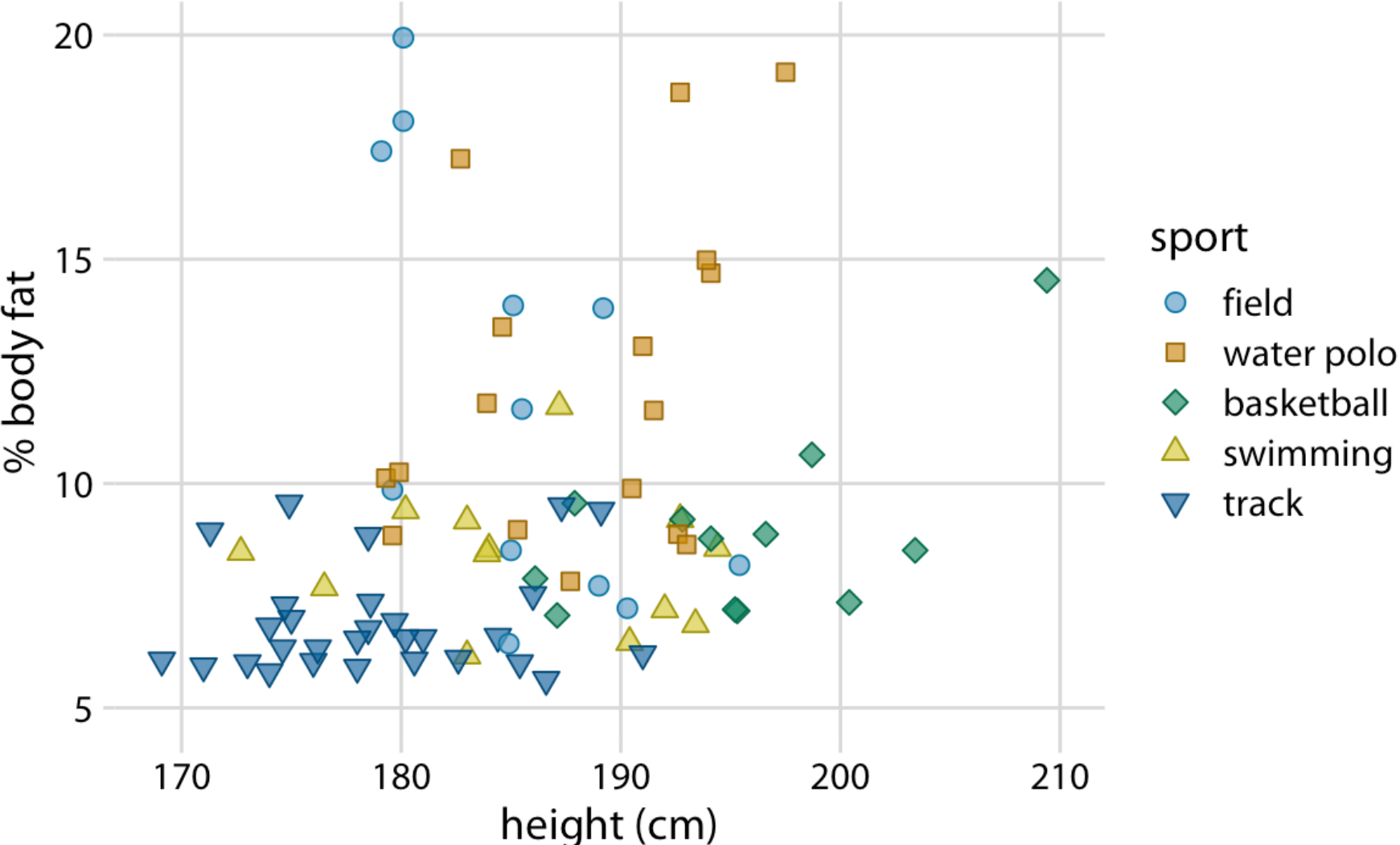
Remove non-encoding visual elements

ugly

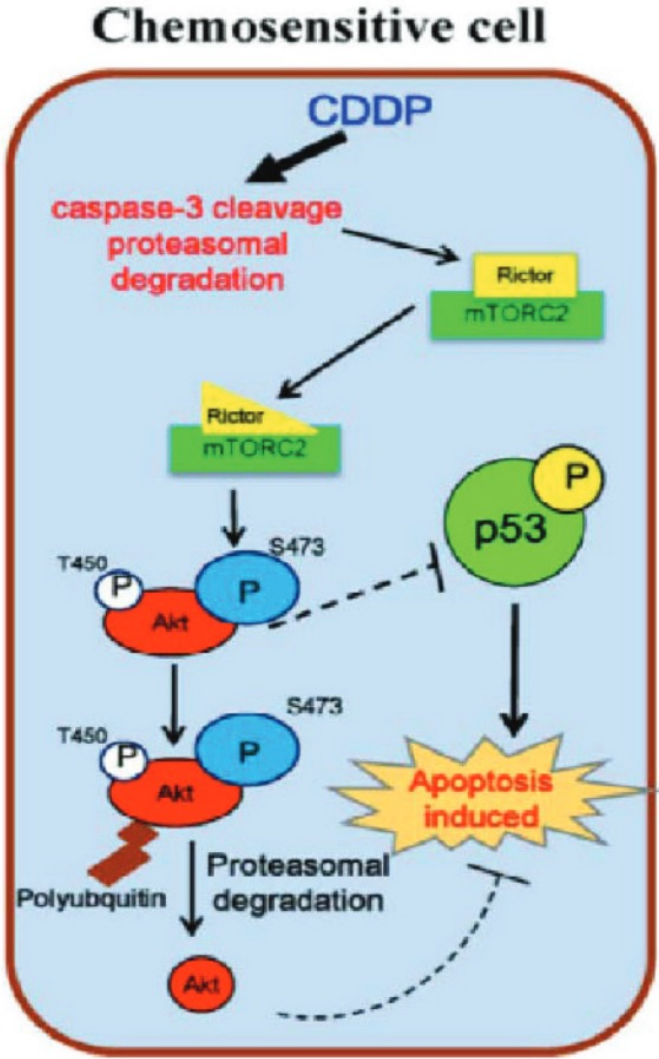


Source: <https://clauswilke.com/dataviz/balance-data-context.html>

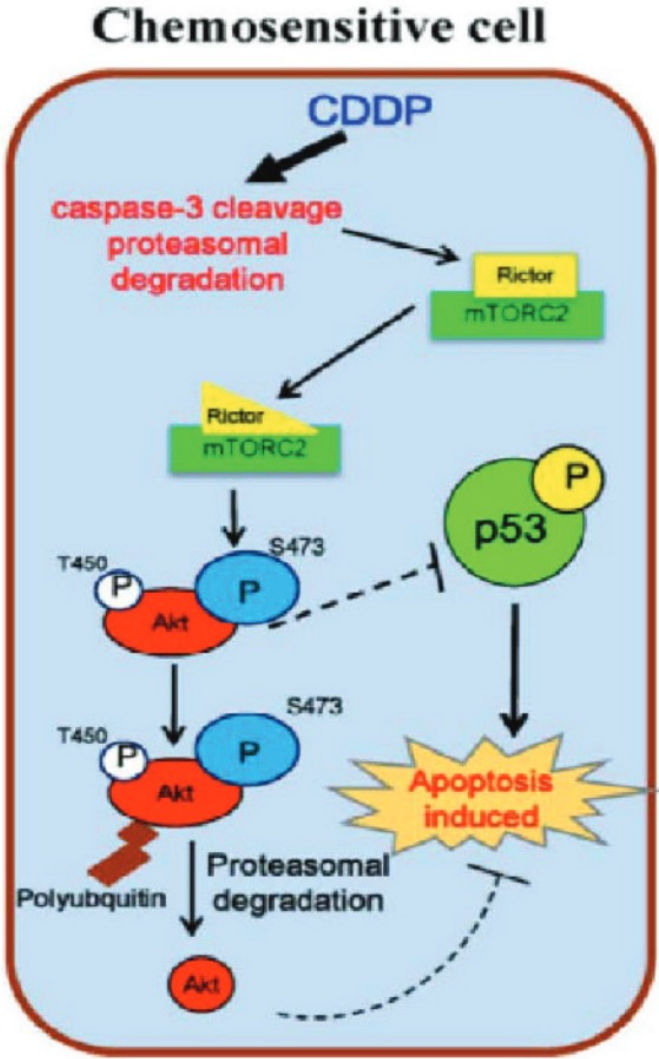
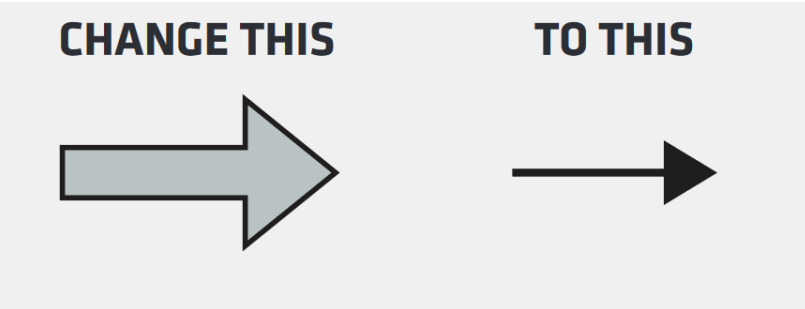
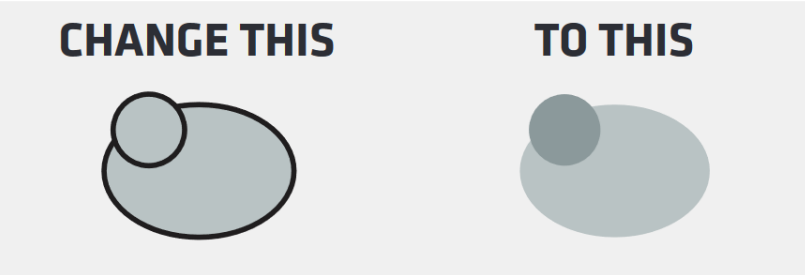
Remove non-encoding visual elements



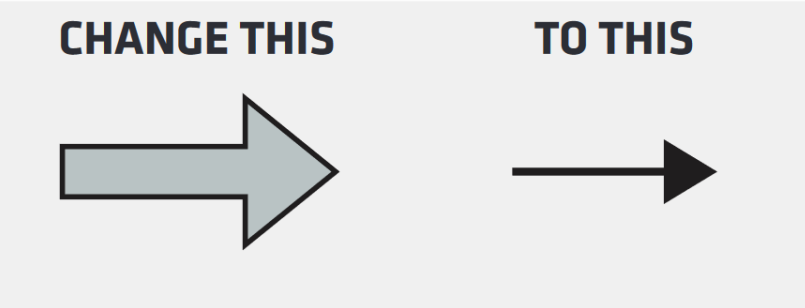
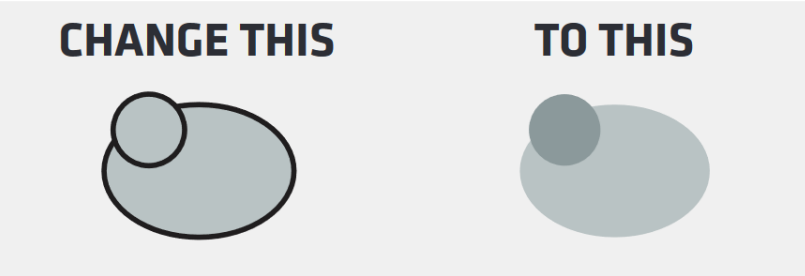
Resist decoration



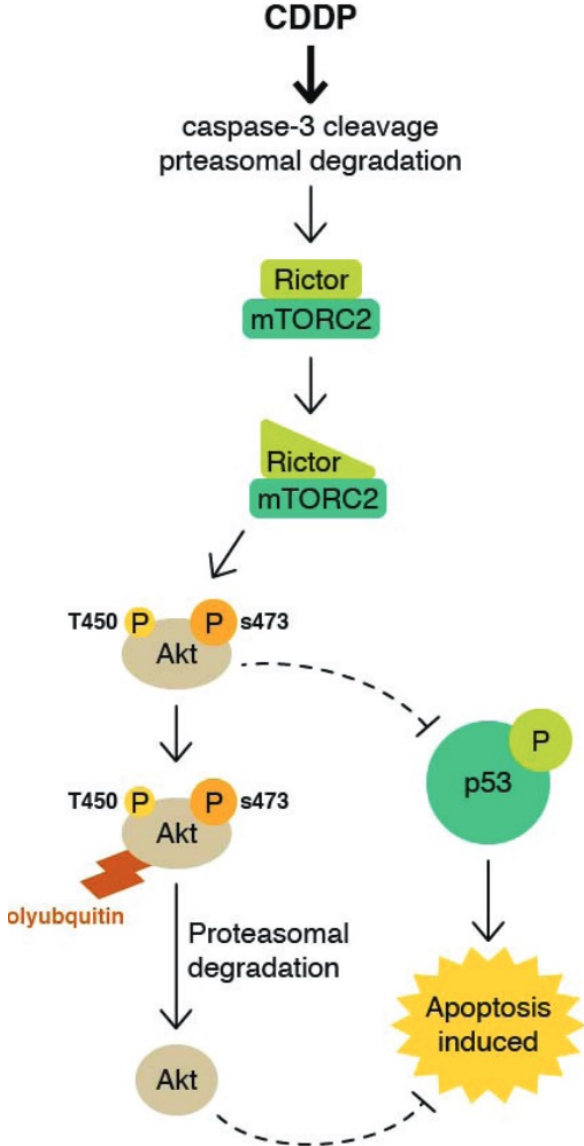
Resist decoration



Resist decoration

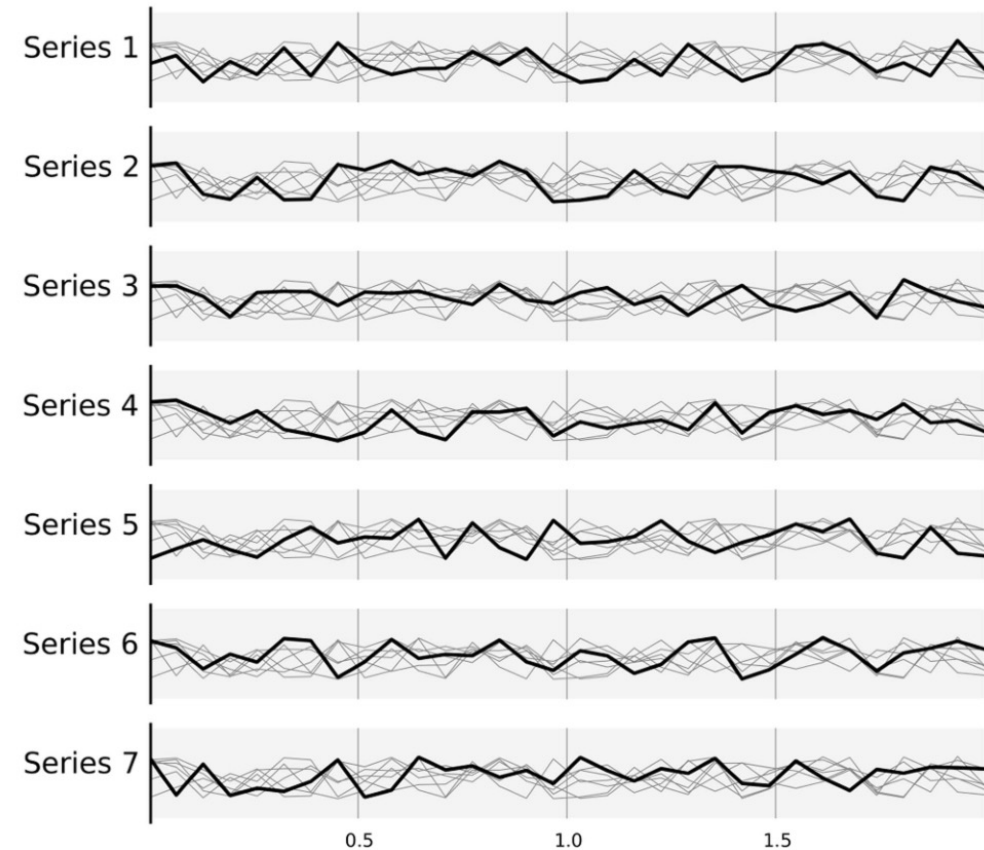
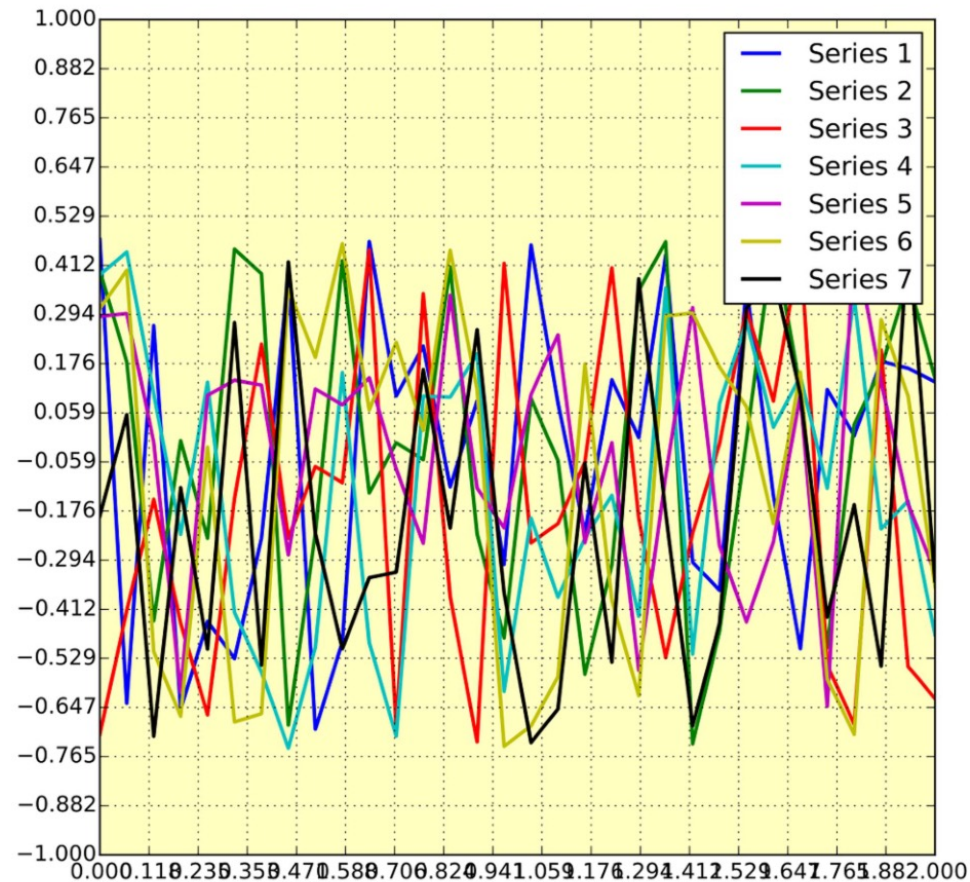


Chemosensitive Cell



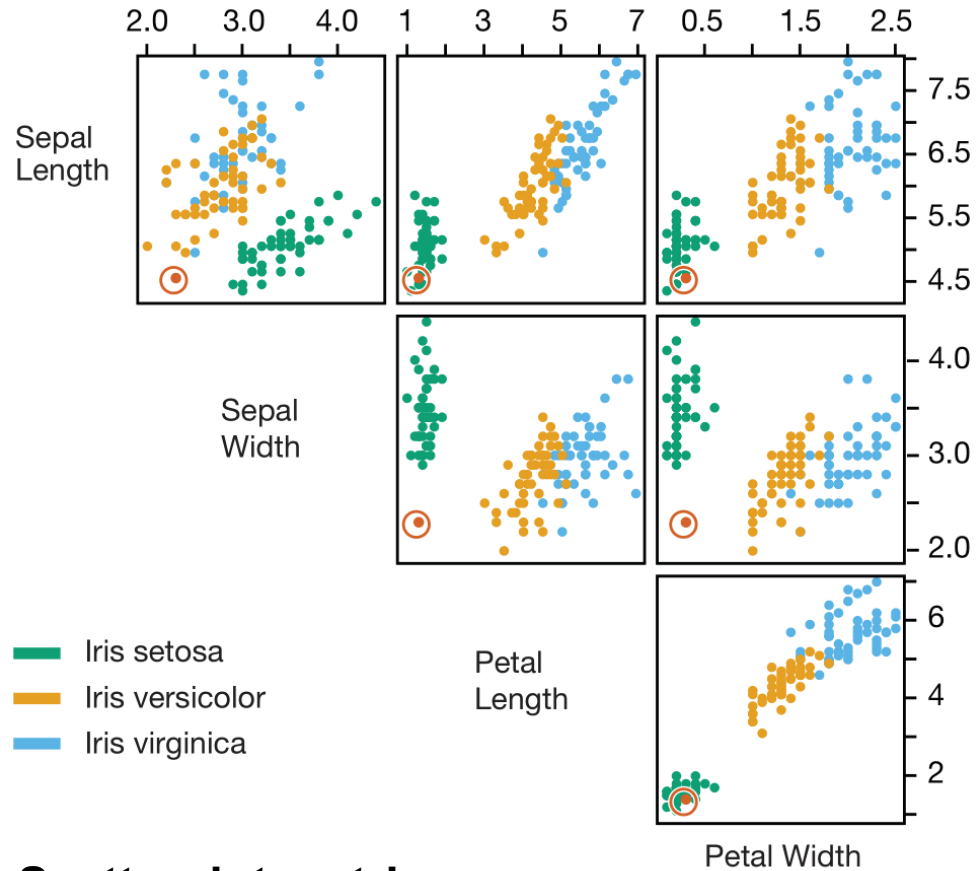
Small multiples

With more than 6 categorical variables, use small multiples

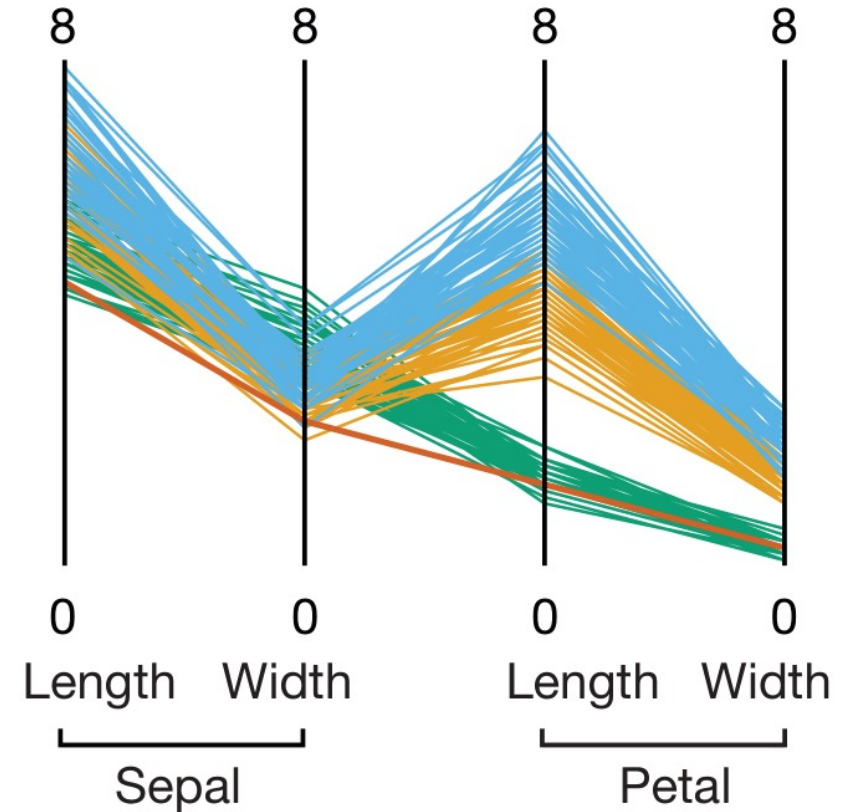


Small multiples

Encoding multivariate data



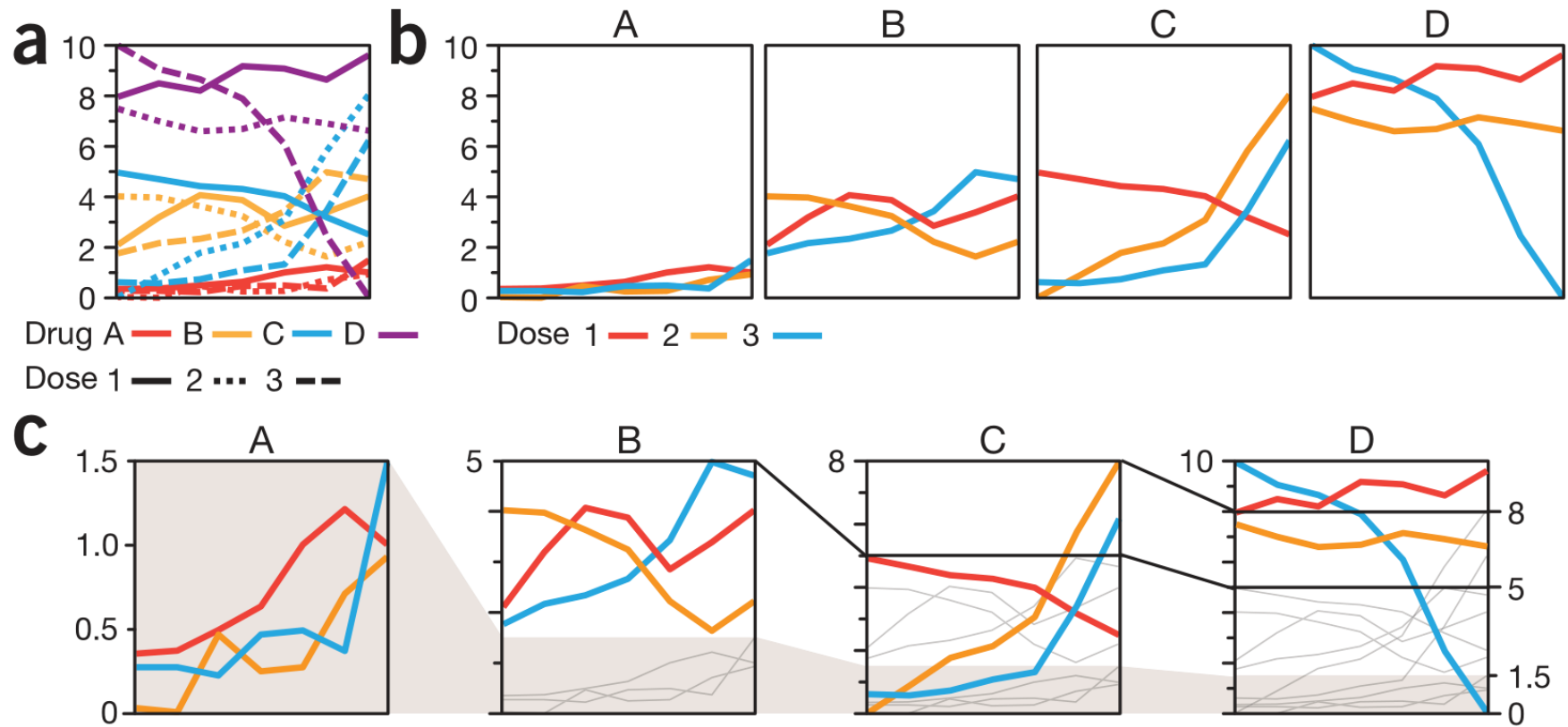
**Scatterplot matrix
(facet)**



Parallel coordinate plot

Small multiples

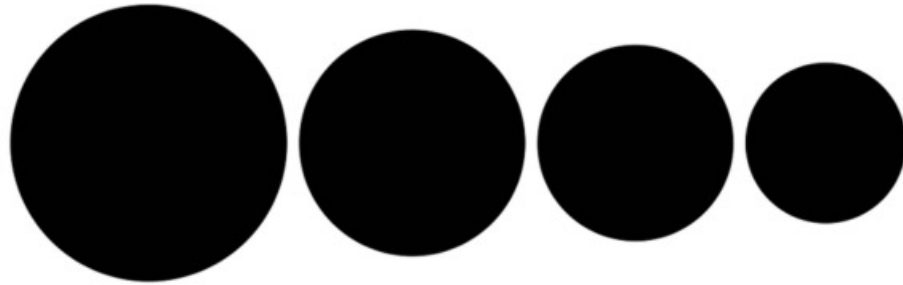
Encoding multivariate data



Use accurate encodings

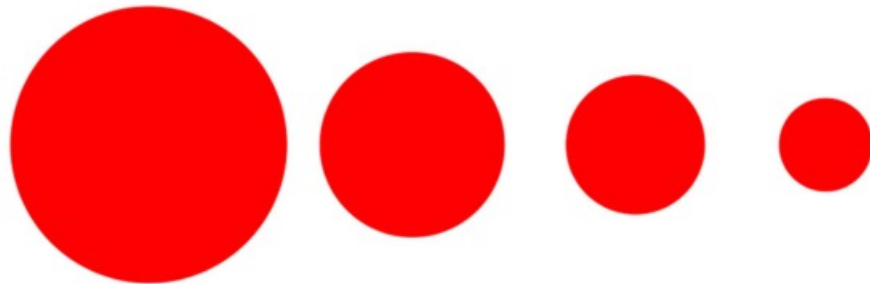
Be certain that your encoding is accurately representing your data

Don't be misleading



Relative size using disc area

Relative size using disc radius



Relative size using full range

Relative size using partial range

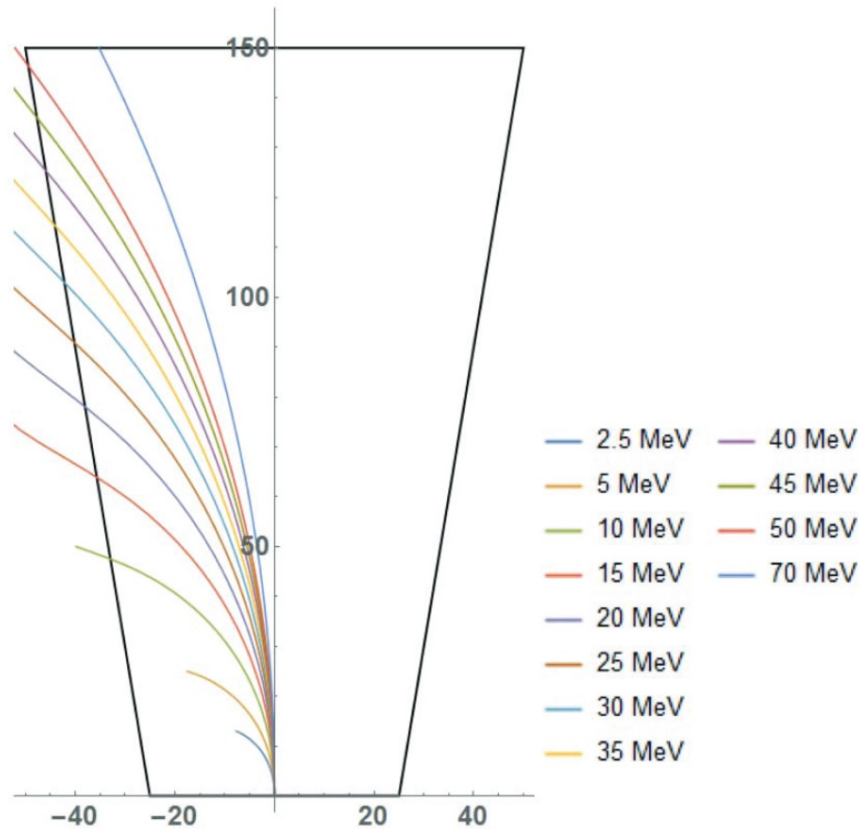


Lose the legend

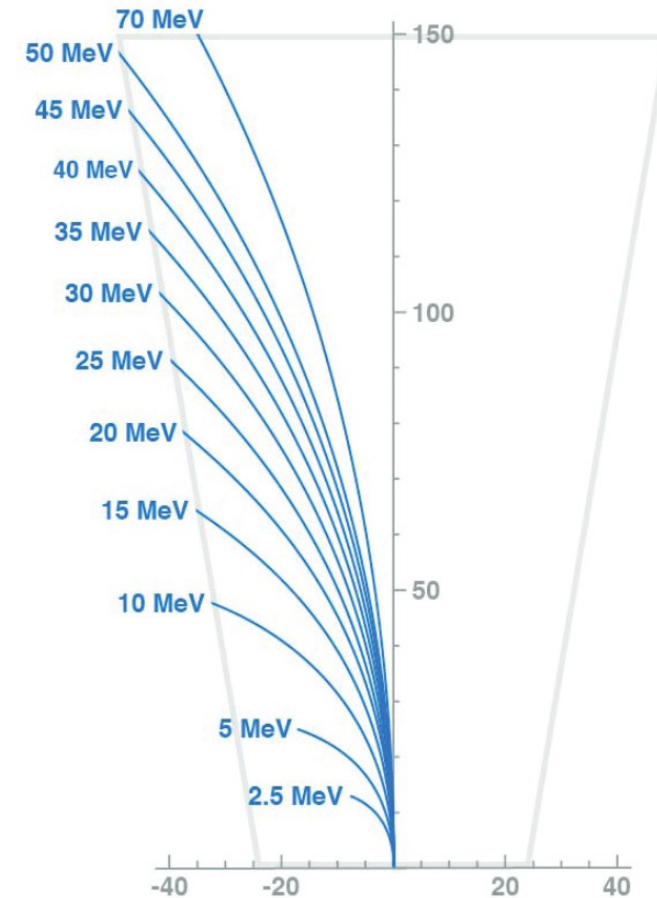
Design your visualization so that it is **self-explanatory** and does not require a separate legend

Use direct labeling if possible

Computing Electron Paths for Various Energies

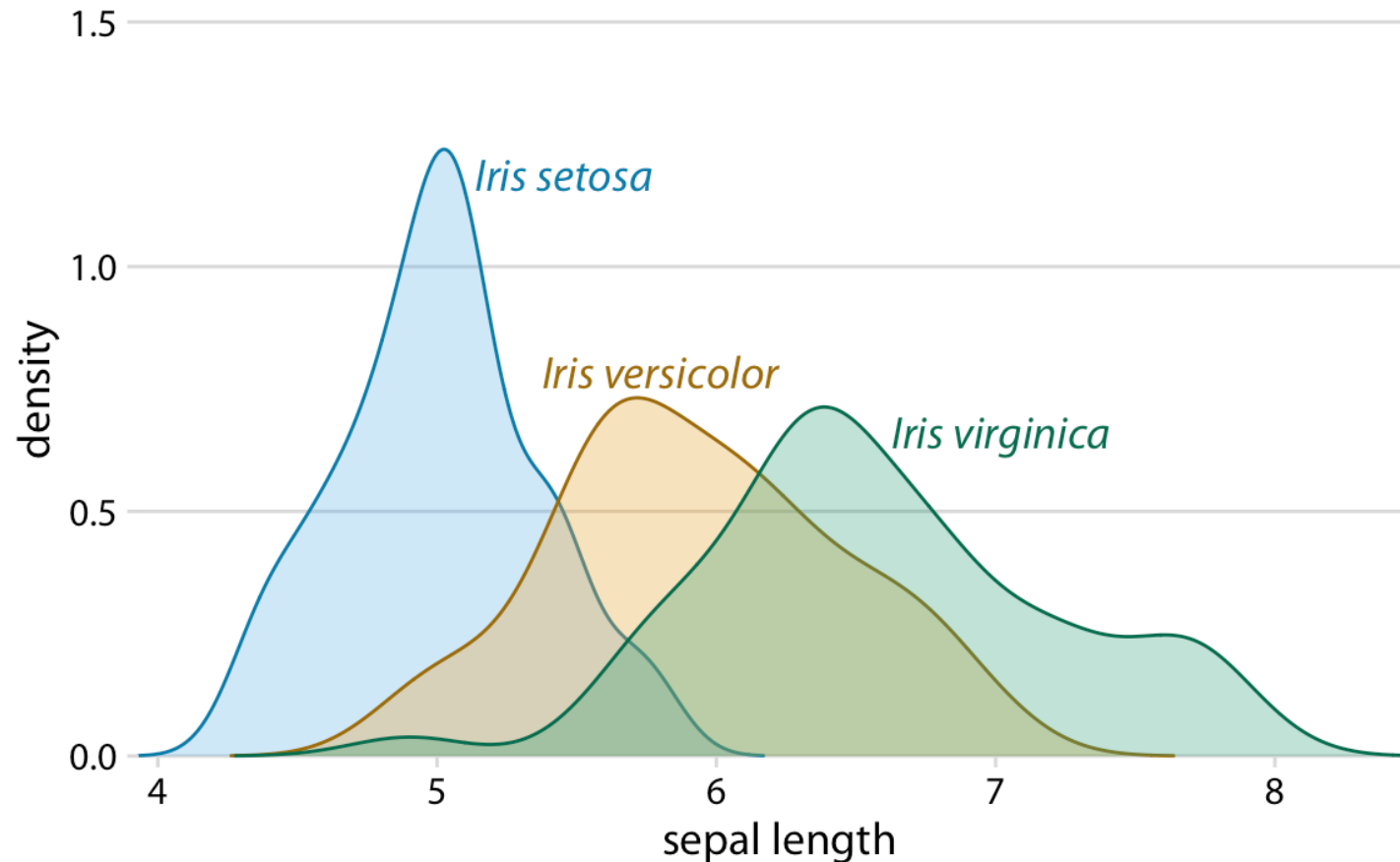


Computing Electron Paths for Various Energies



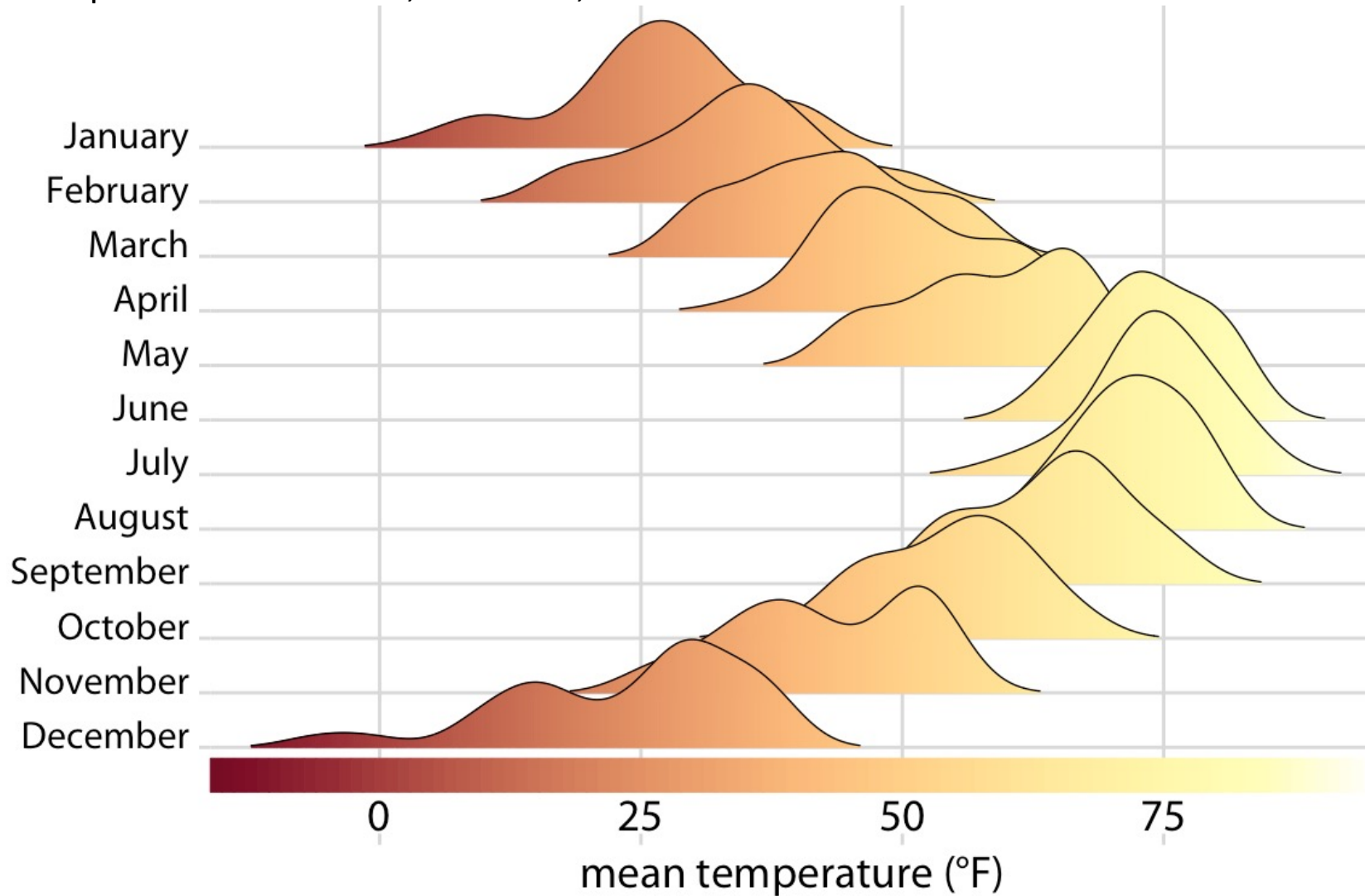
Use direct labeling if possible

Density estimates of the sepal lengths of three different iris species.



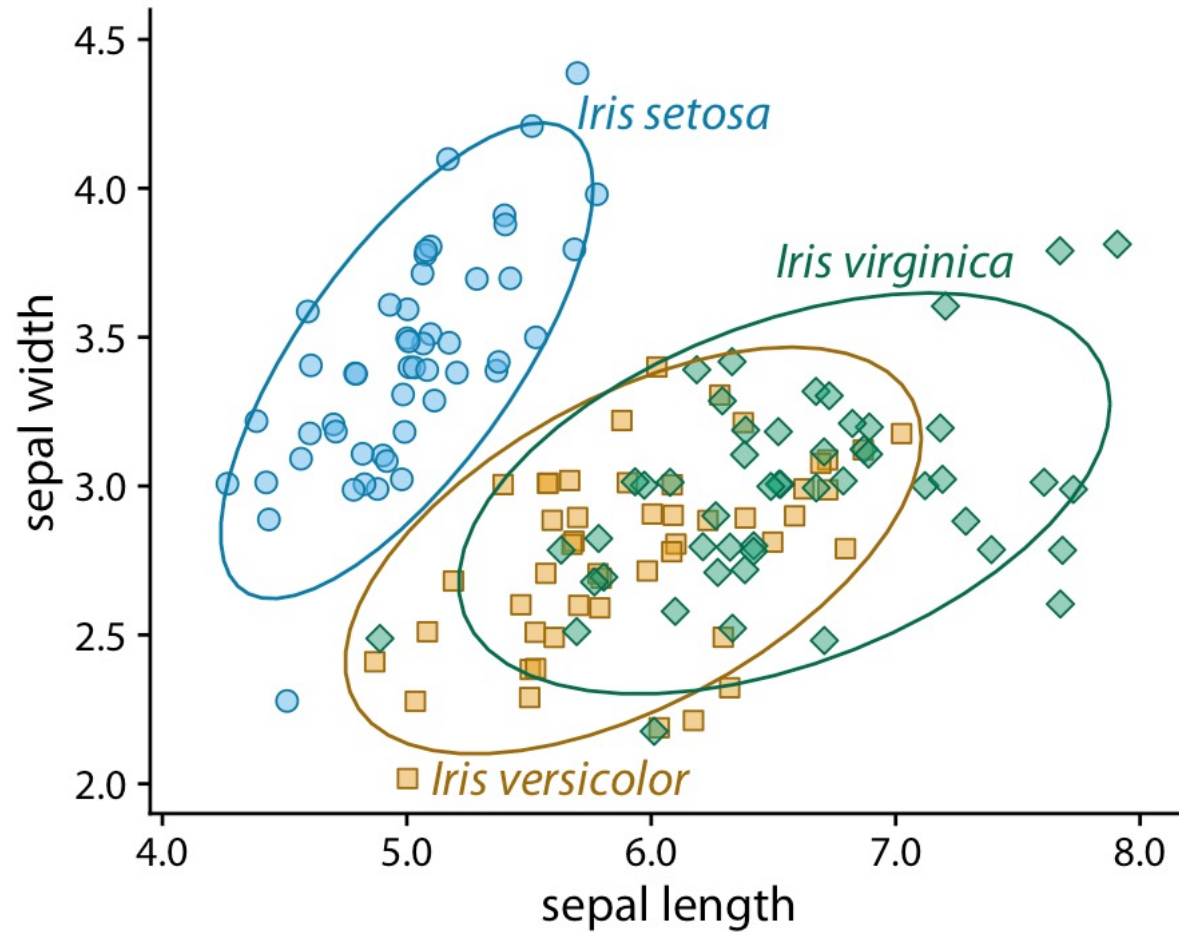
Integrated legends

Temperatures in Lincoln, Nebraska, in 2016.



Use direct labeling if possible

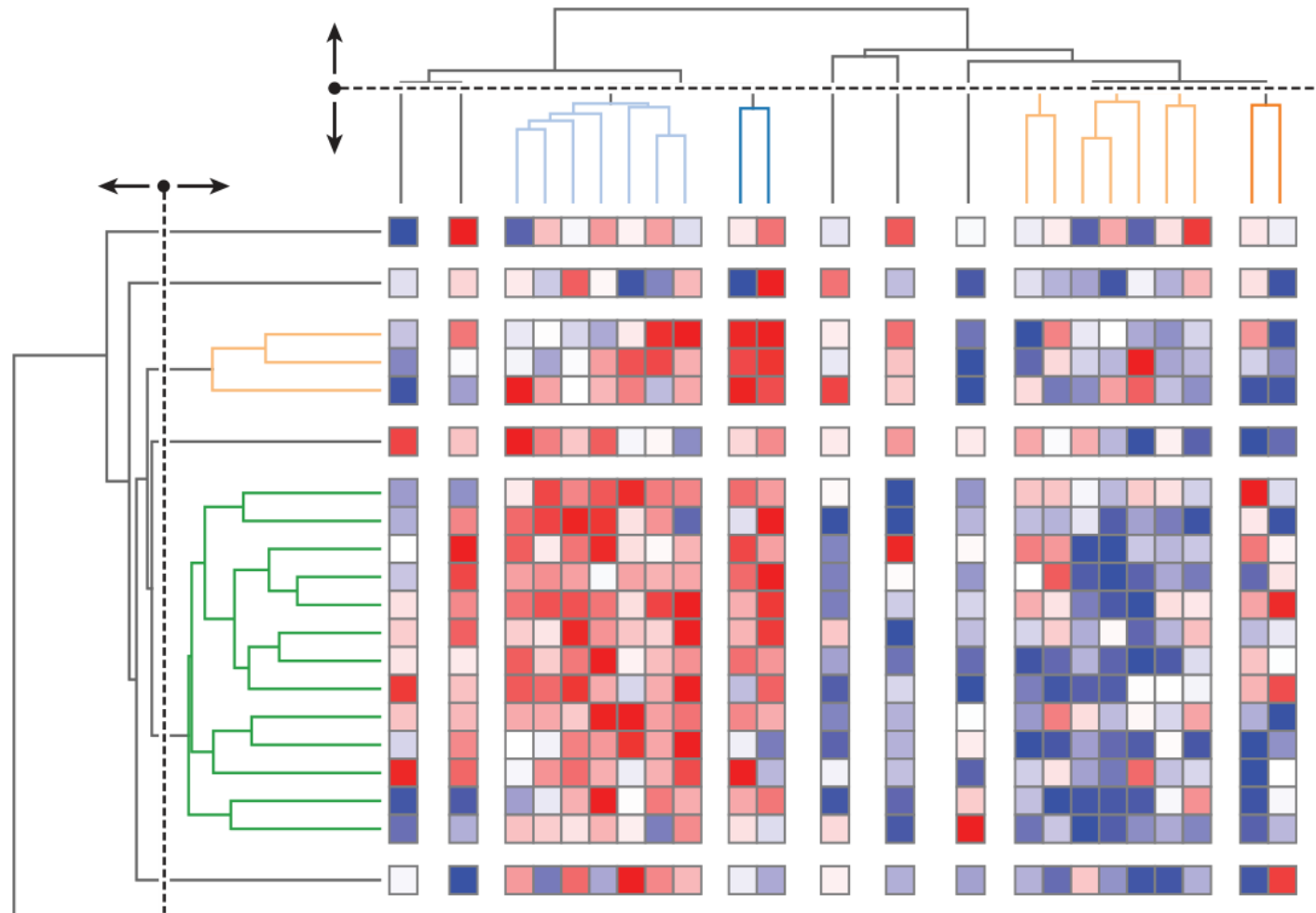
Sepal width versus sepal length for three different iris species



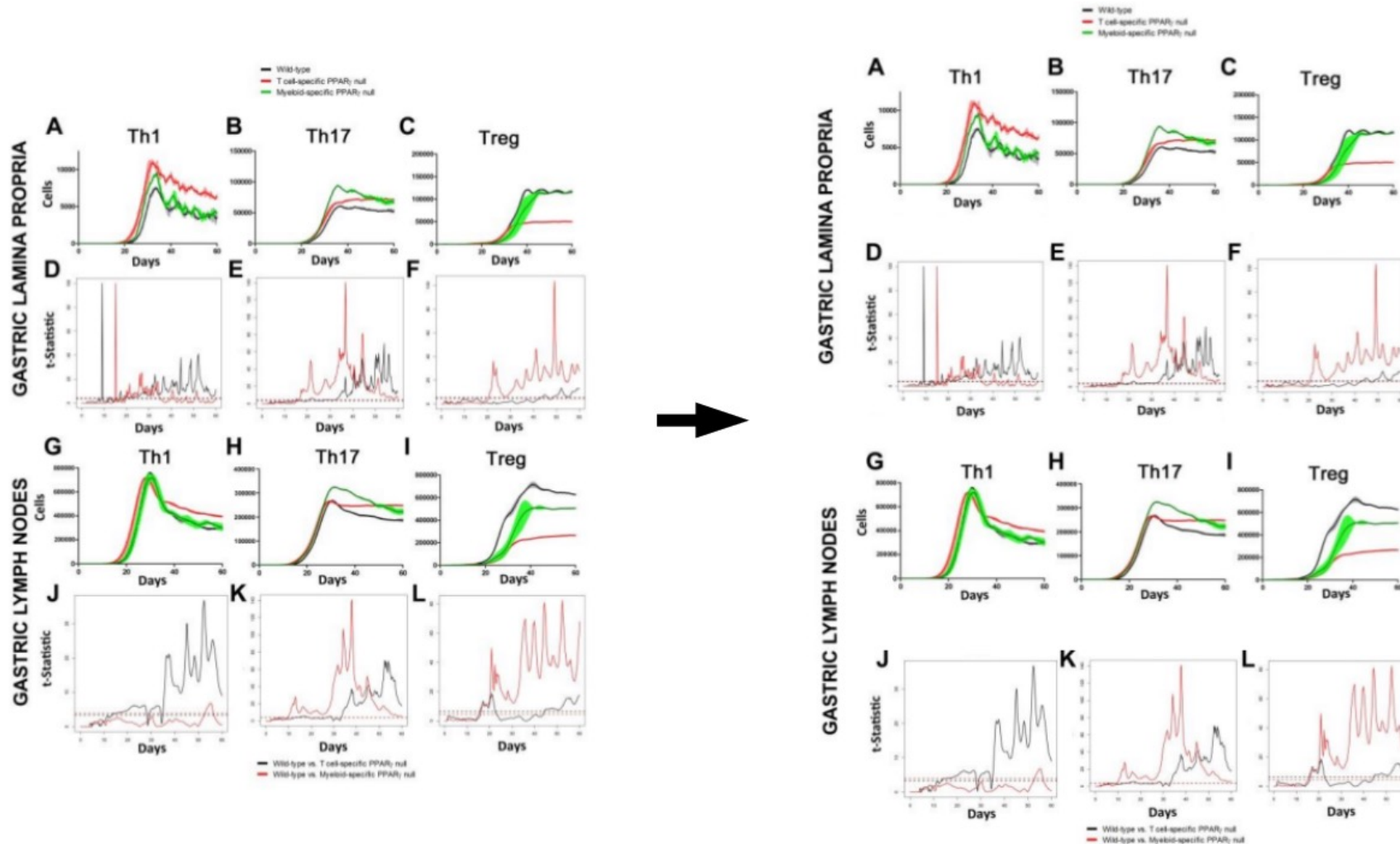
Use whitespace

Use whitespace to create natural groupings in your visualization

Highlight groups within your data with whitespace



Highlight groups within your data with whitespace



Axes, ticks, and gridlines

Recommended practices

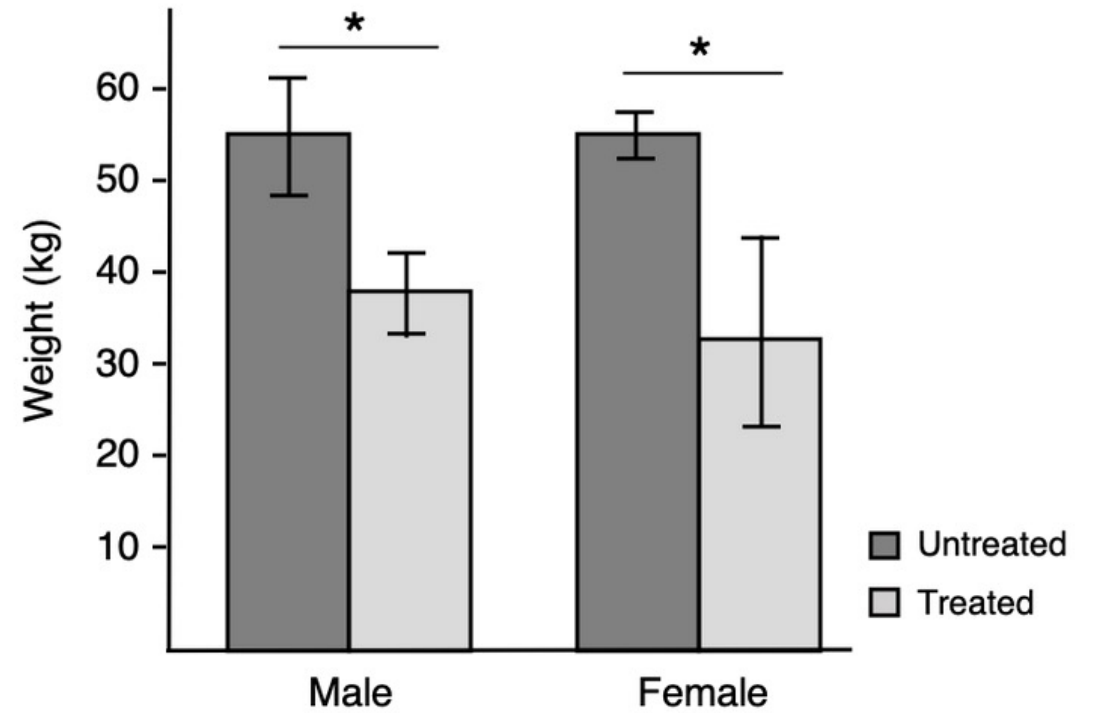
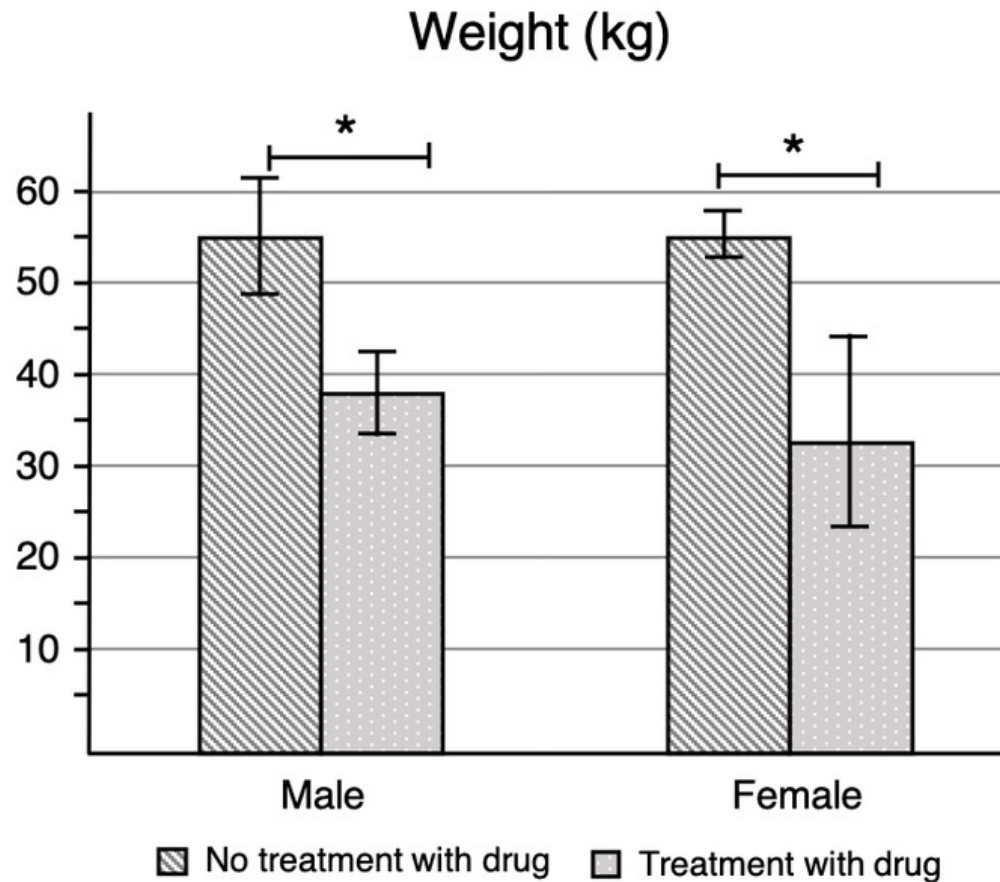
Gridlines

Gridlines should enhance readability



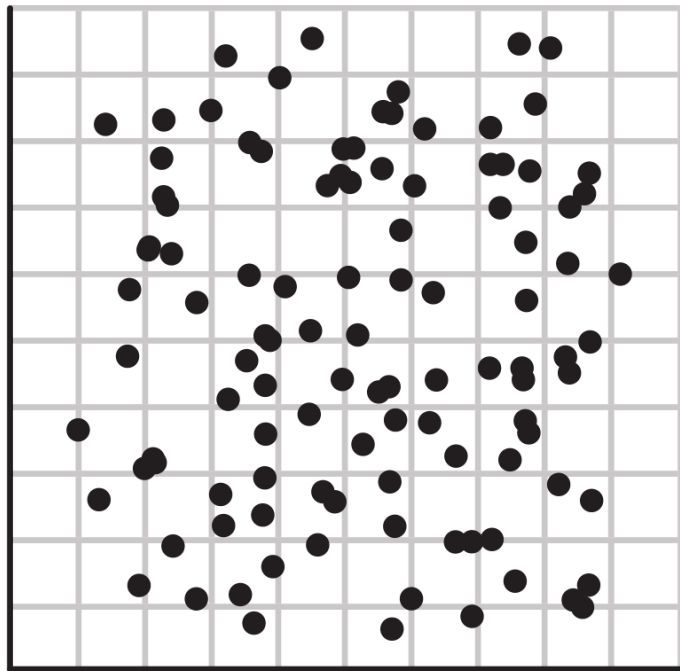
Gridlines

Gridlines should enhance readability



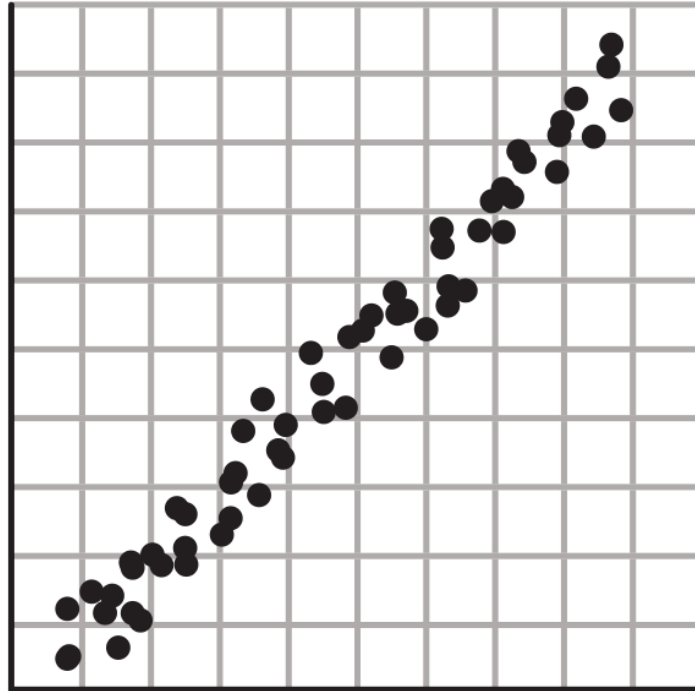
Axes, ticks, and grids

Scale grid opacity to data density



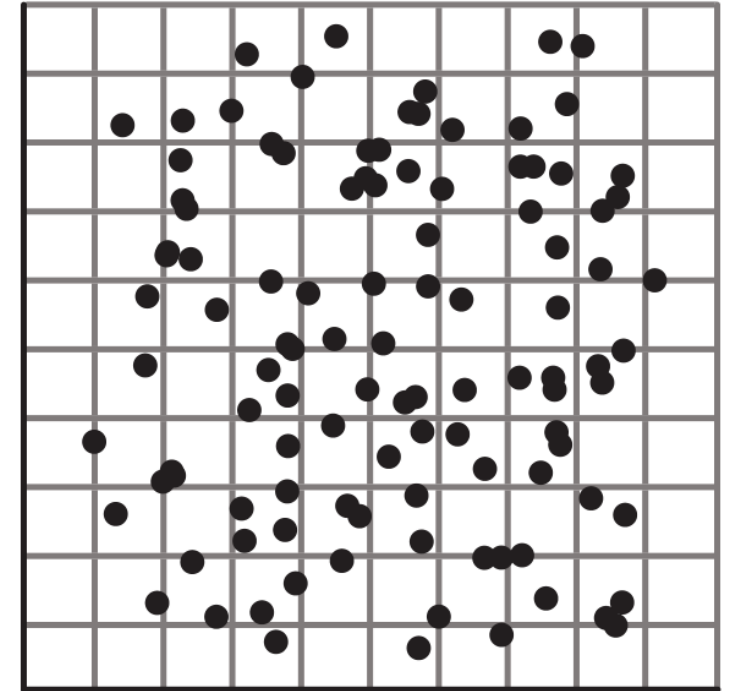
15%

Lightest usable



25%

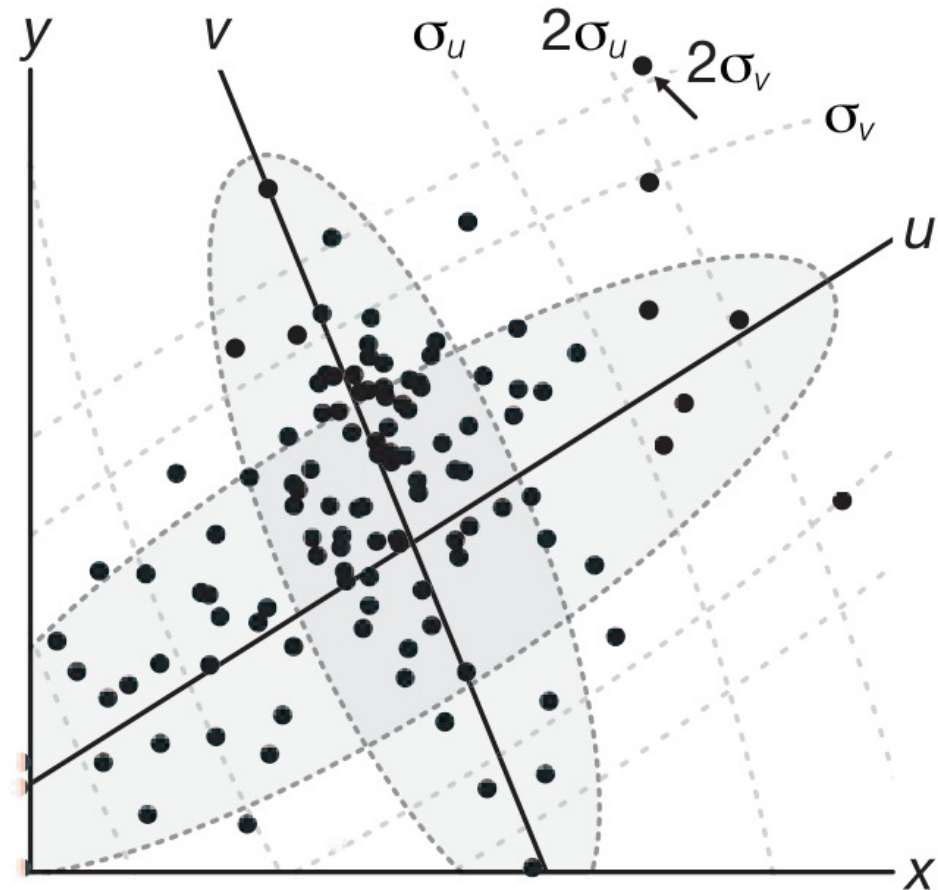
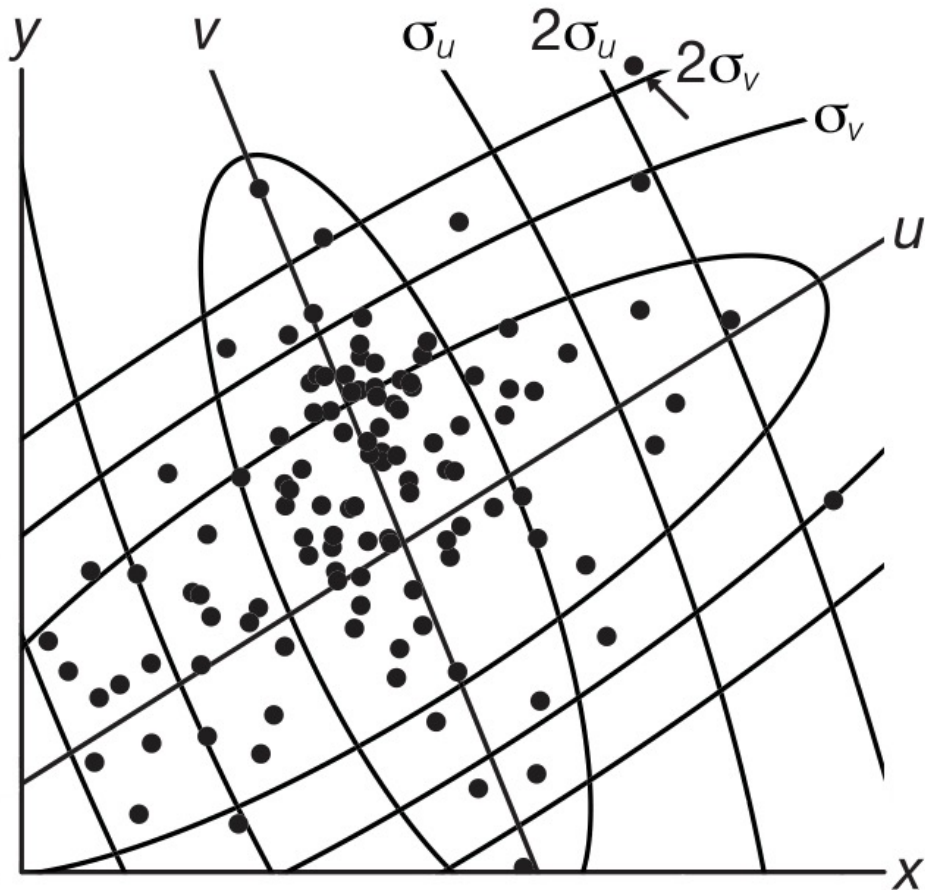
Darkest usable



45%

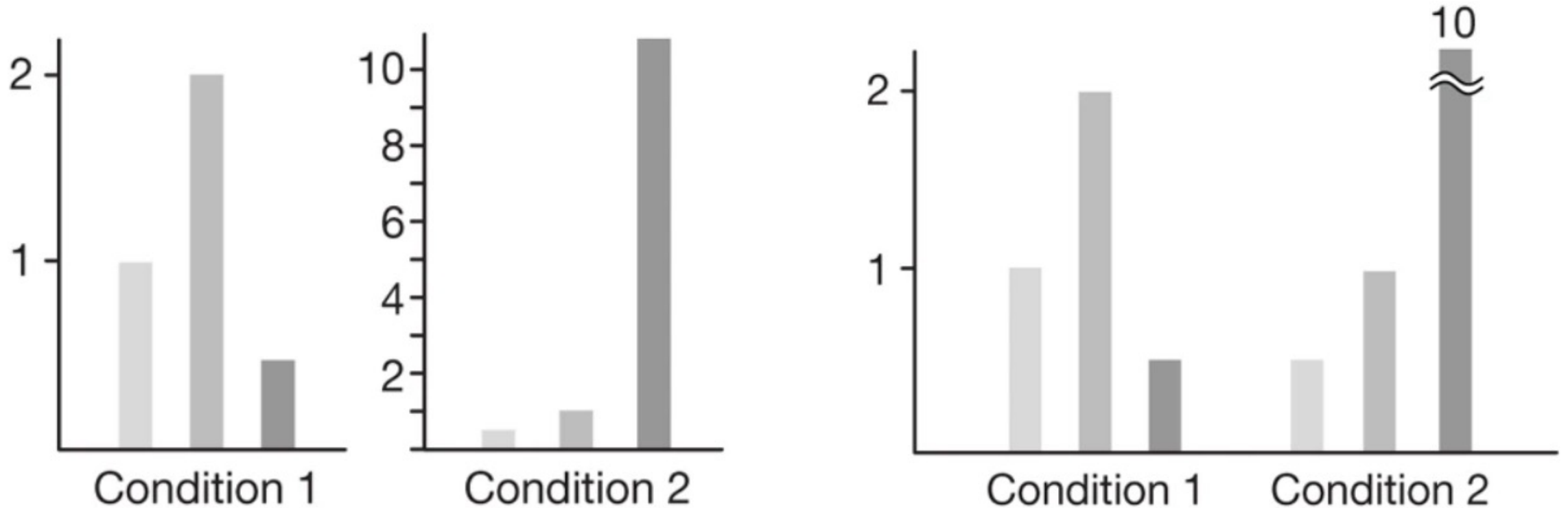
Axes, ticks, and grids

Make navigational elements distinct



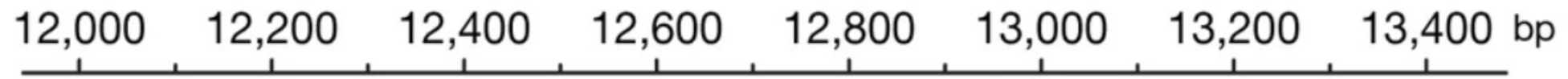
Axes, ticks, and grids

If absolute differences are important, **maintain scaling**



Axes, ticks, and grids

Use the appropriate unit in your labels



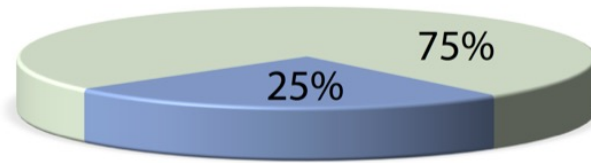
Avoid 3D

Three-dimensional data visualizations are rarely justified

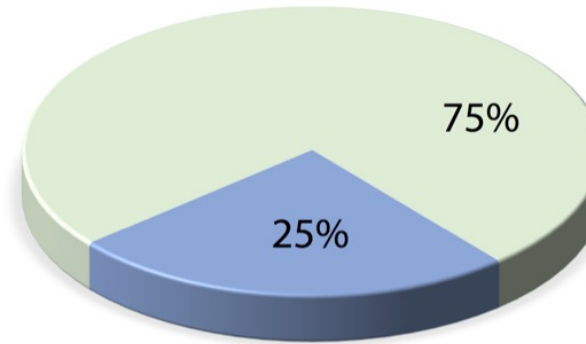
Avoid unjustified 3D

If the third dimension does not encode information, 3D is not justified

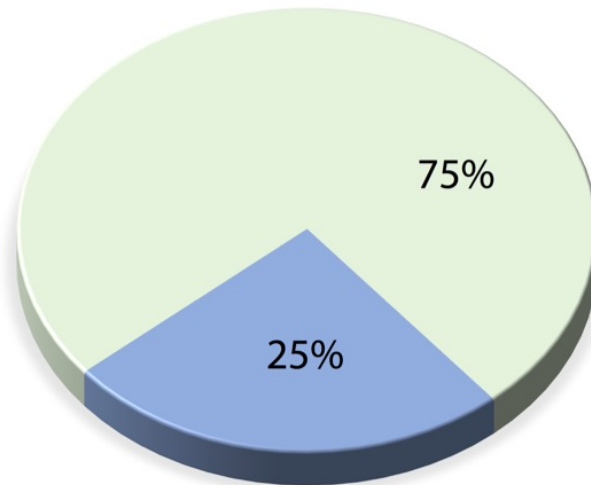
a



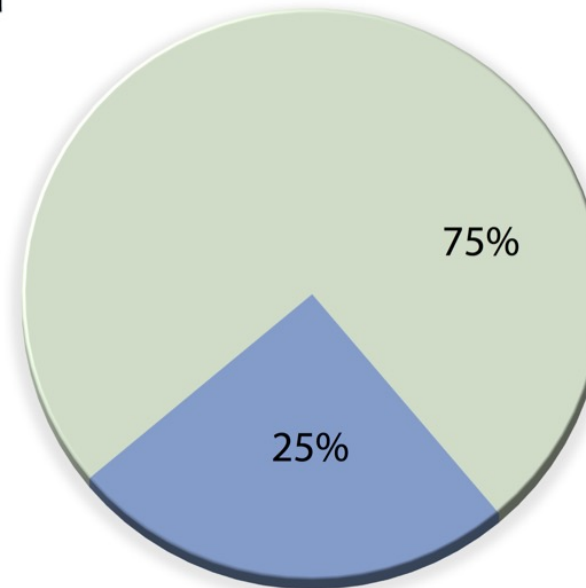
b



c

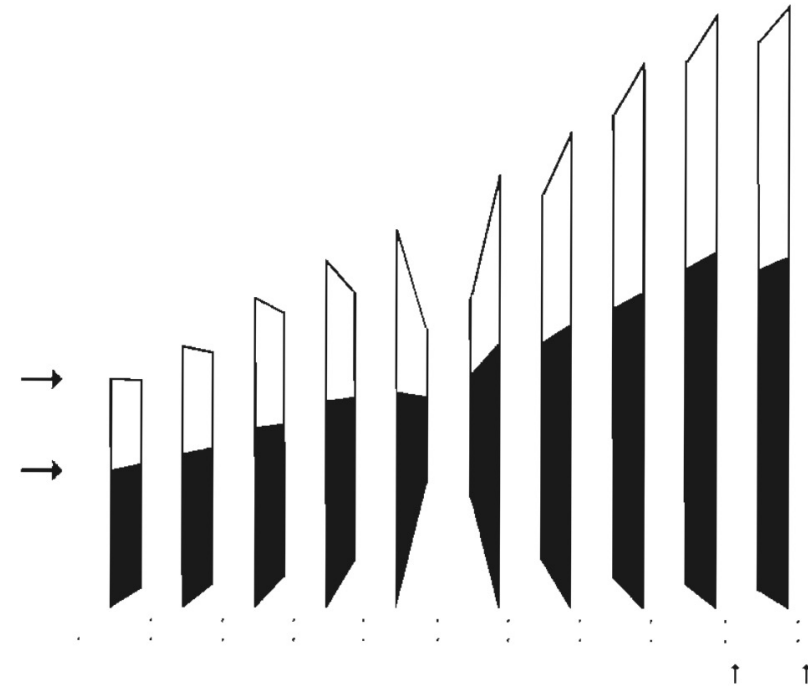
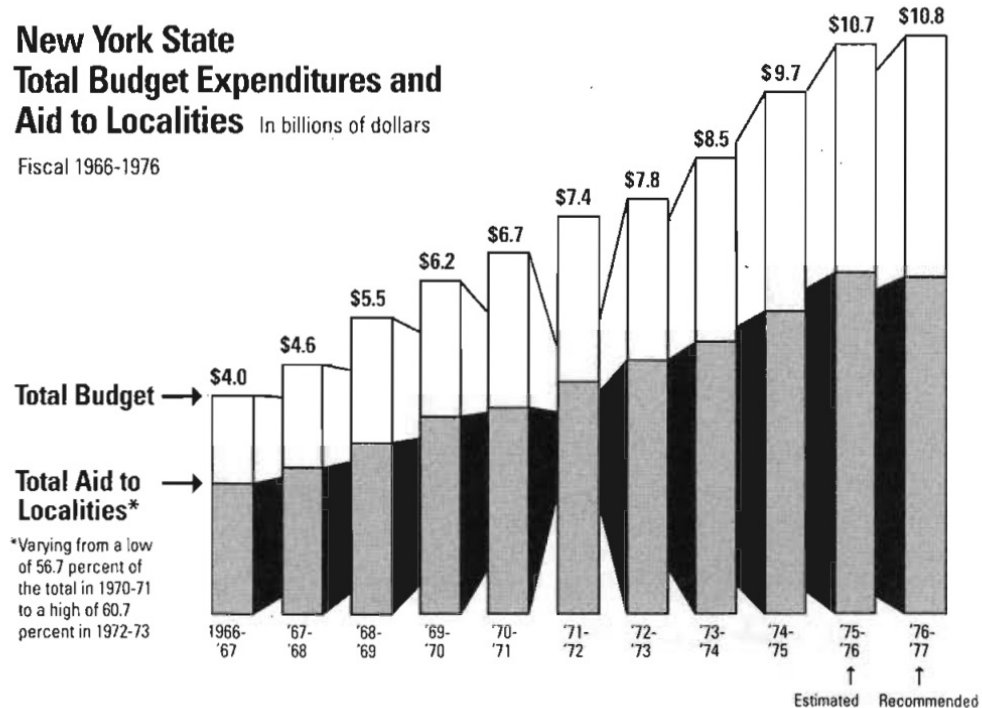


d



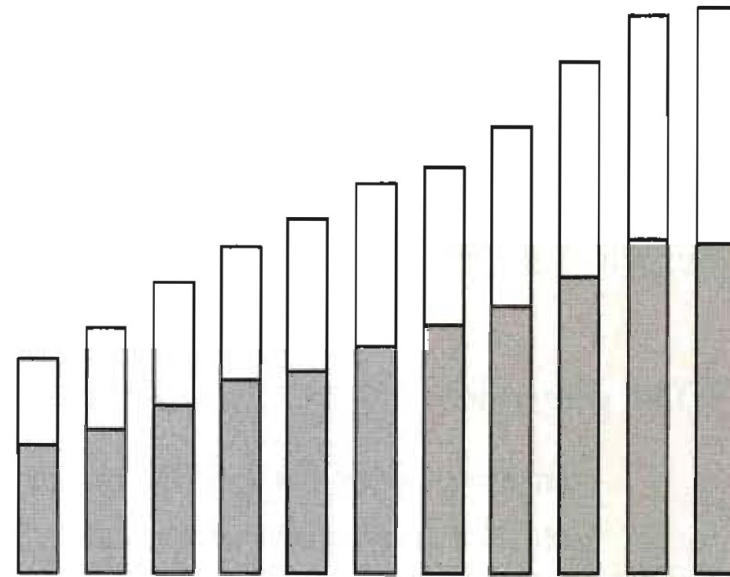
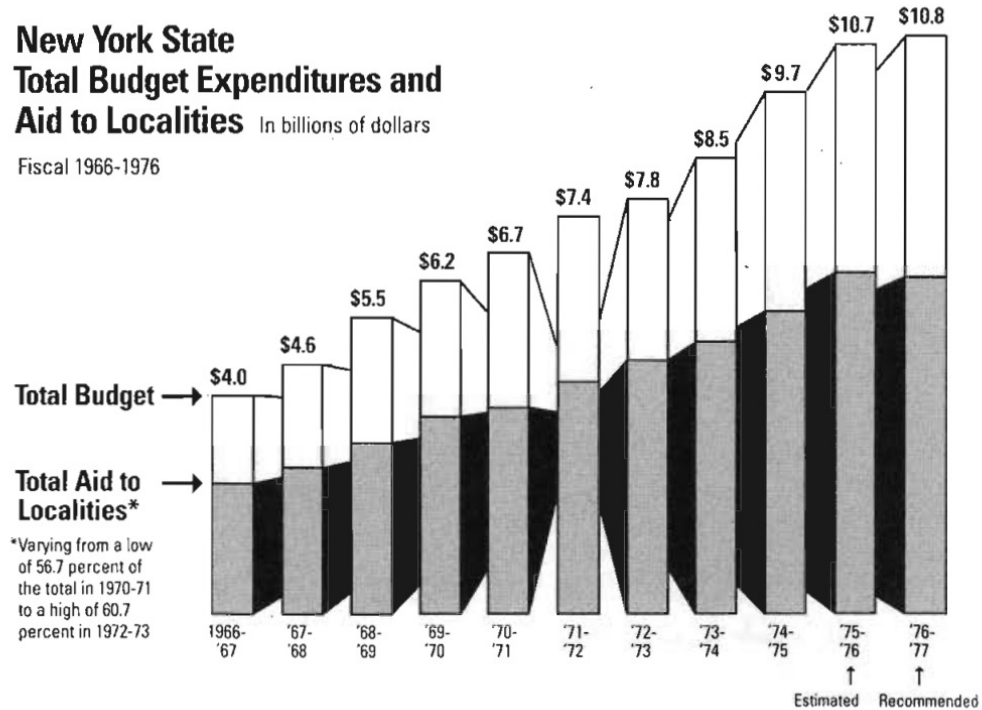
Avoid unjustified 3D

Use of 3D can bias interpretation



Avoid unjustified 3D

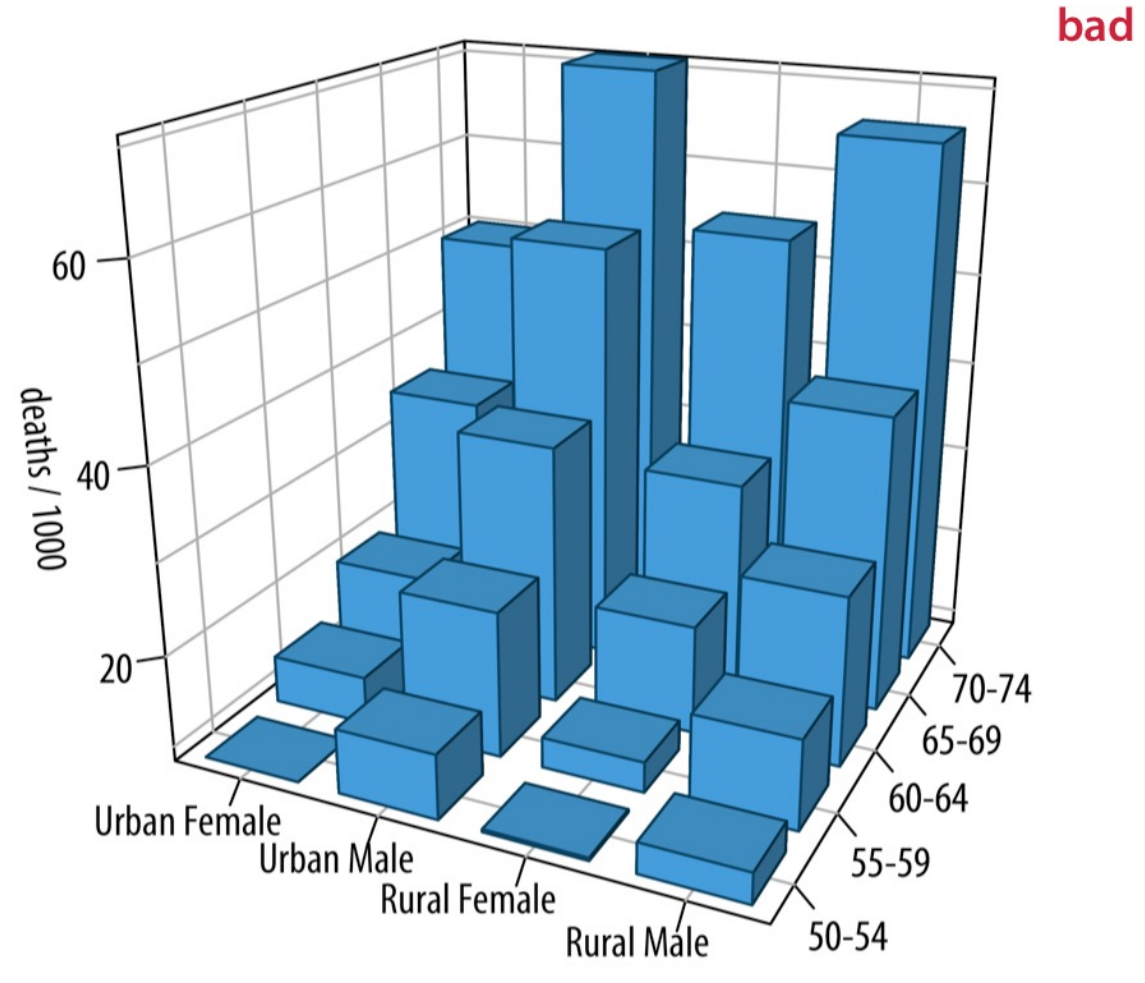
Use of 3D can bias interpretation



Avoid unjustified 3D

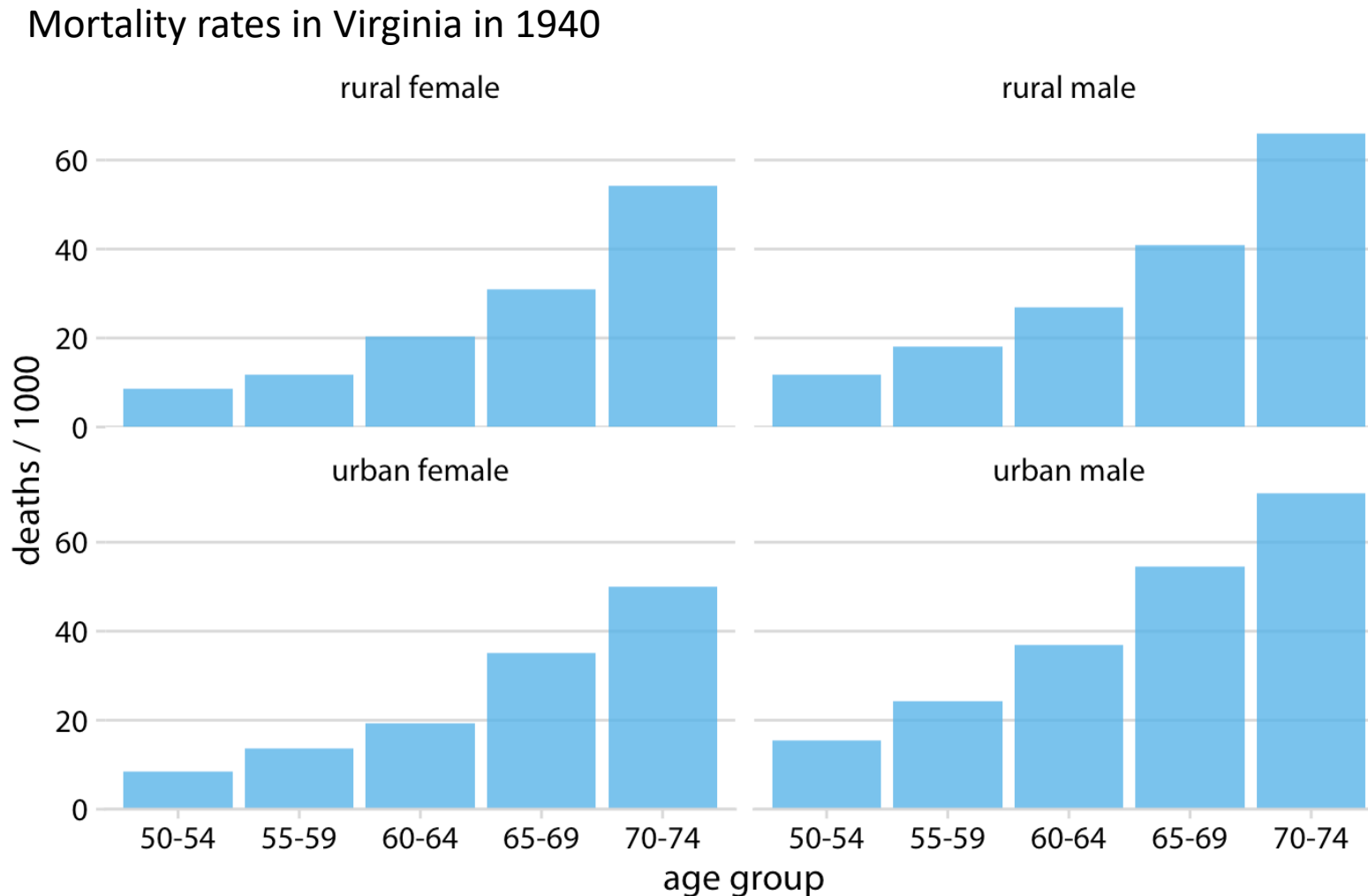
Use of 3D can *impair* interpretation

Mortality rates in Virginia in 1940



Avoid unjustified 3D

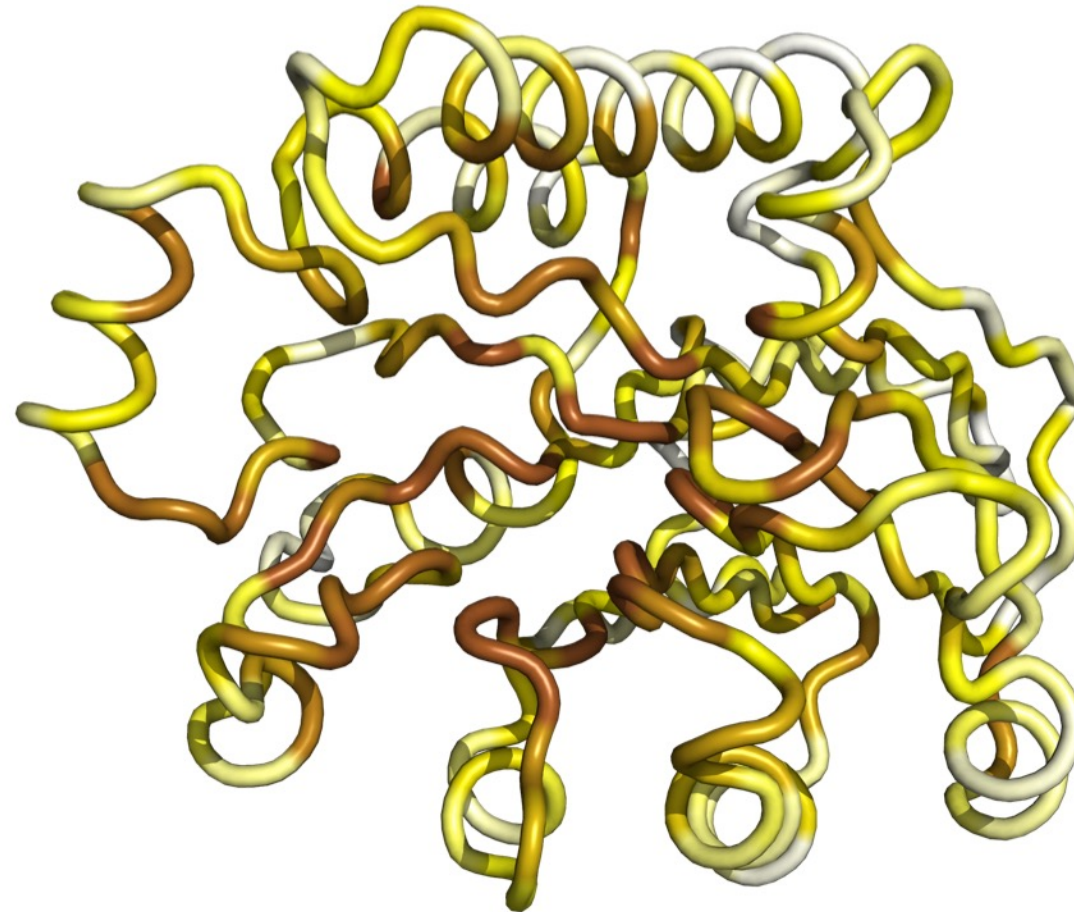
Use of 3D can *impair* interpretation



Appropriate 3D visualization



Appropriate 3D visualization



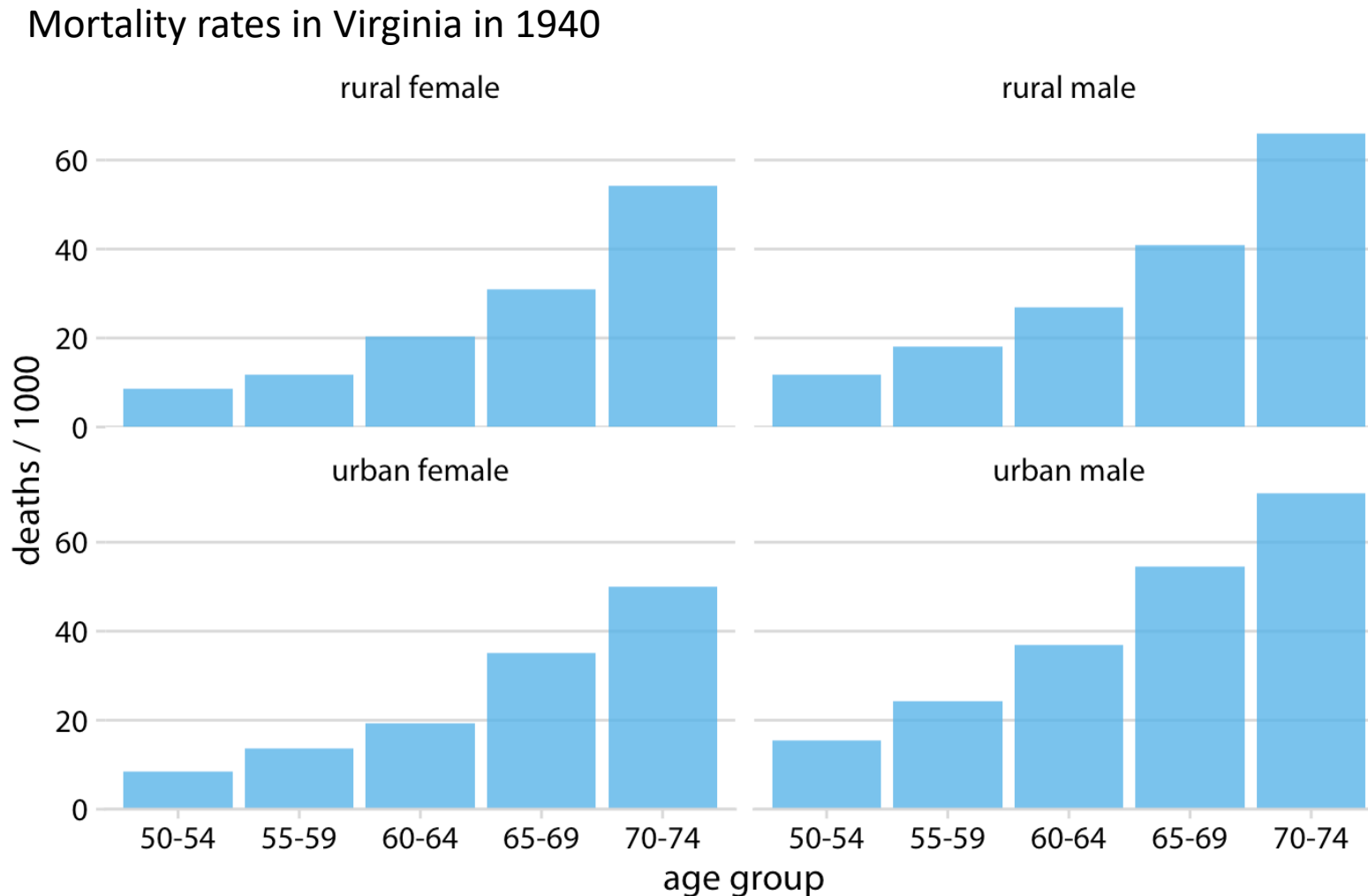
sequence conservation



Patterns of evolutionary variation in a protein.

Avoid unjustified 3D

Use of 3D can *impair* interpretation



Use color carefully

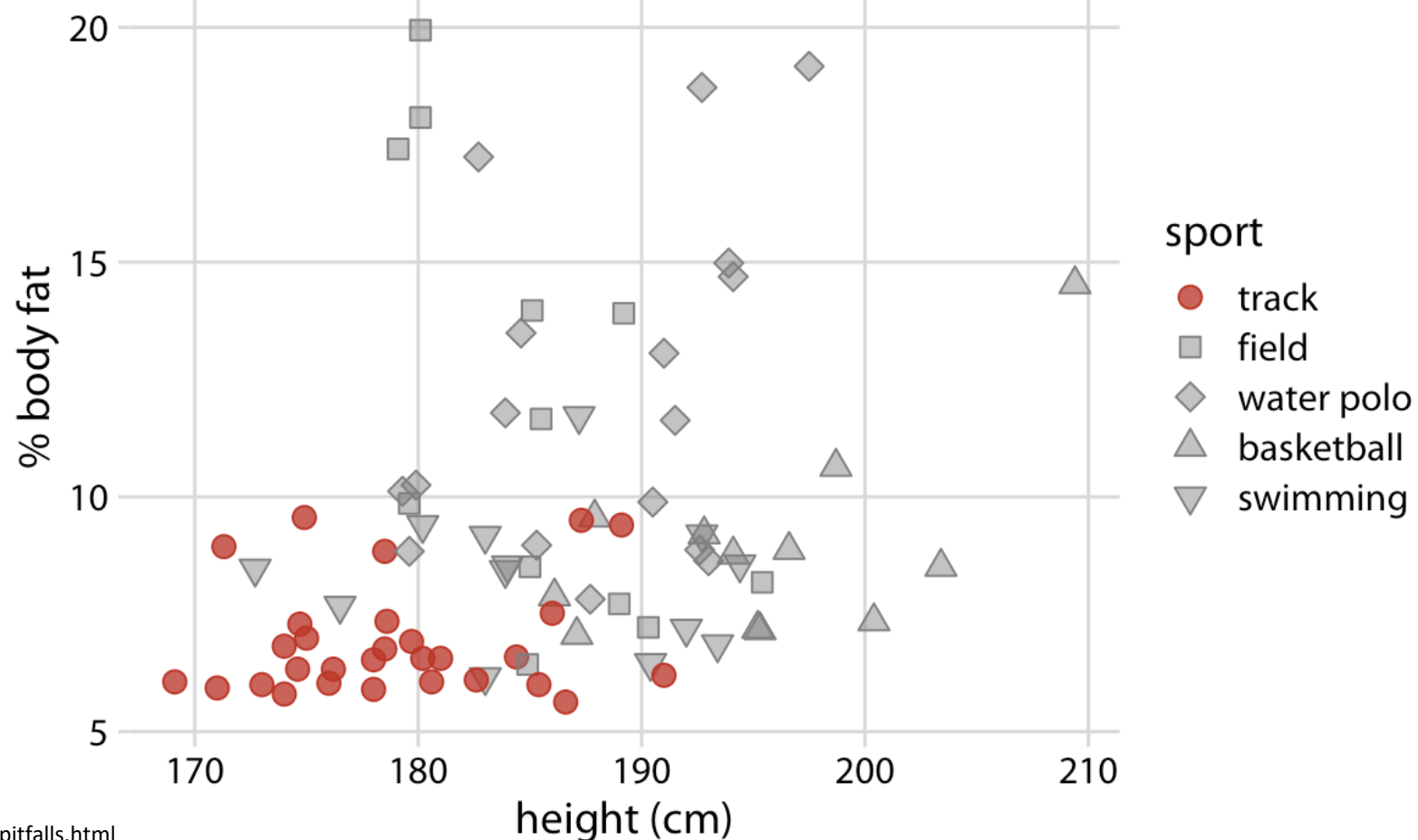
Color choices are key in creating effective data visualizations

Get it Right in Black and White

Use color to highlight important attributes

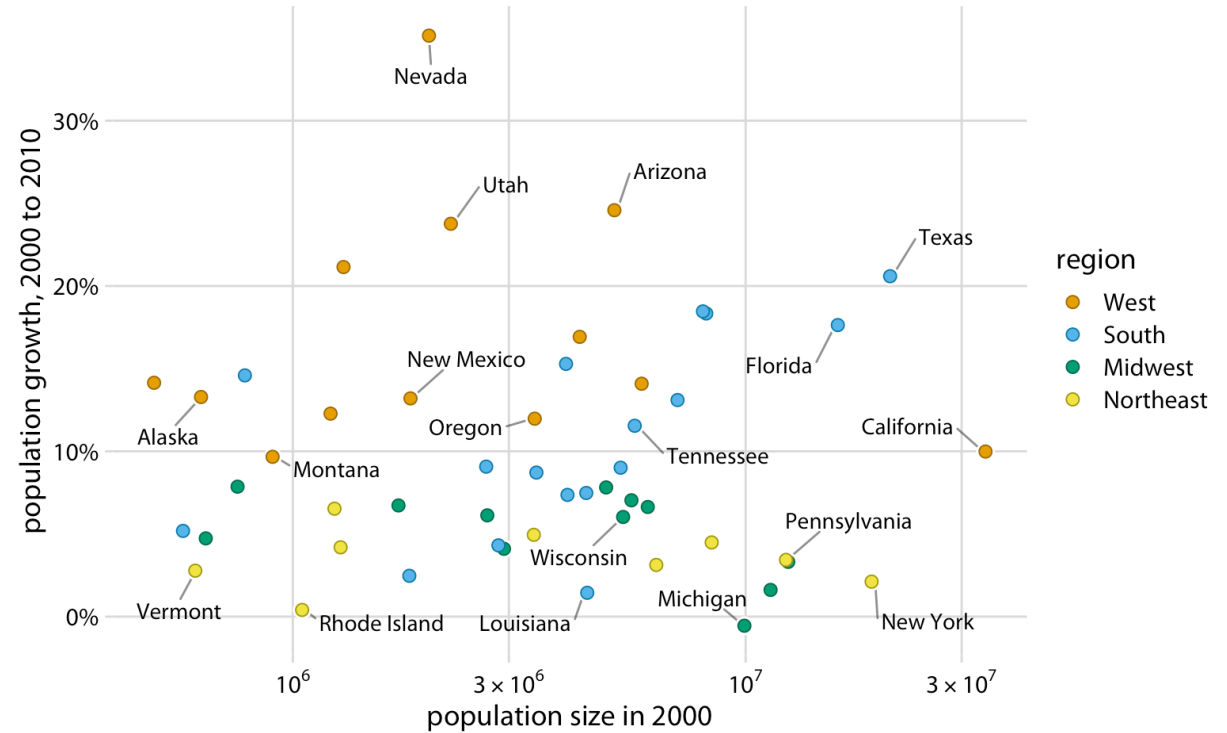
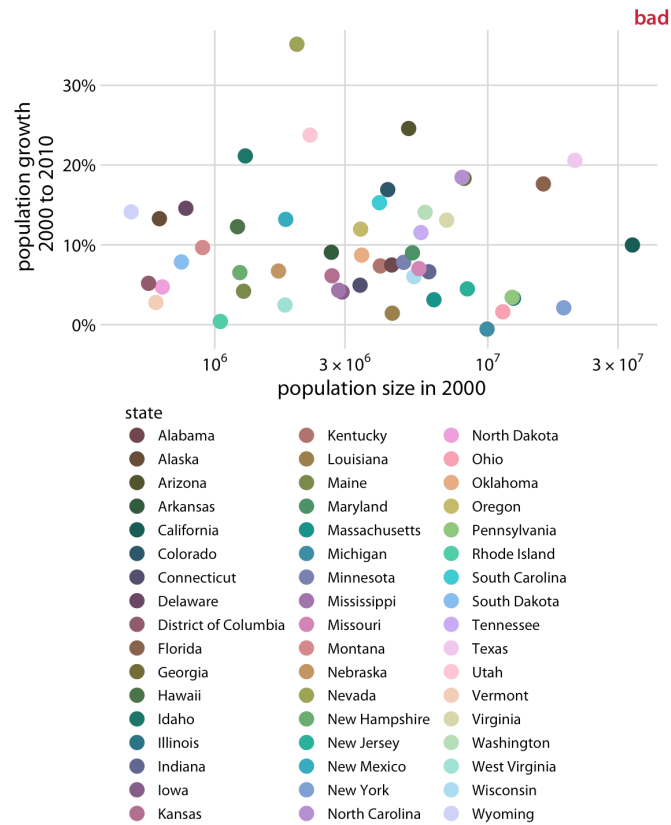
Minimal color use can draw attention to key features

Track athletes are among the shortest and leanest of male professional athletes participating in popular sports.



Don't overdo the use of color

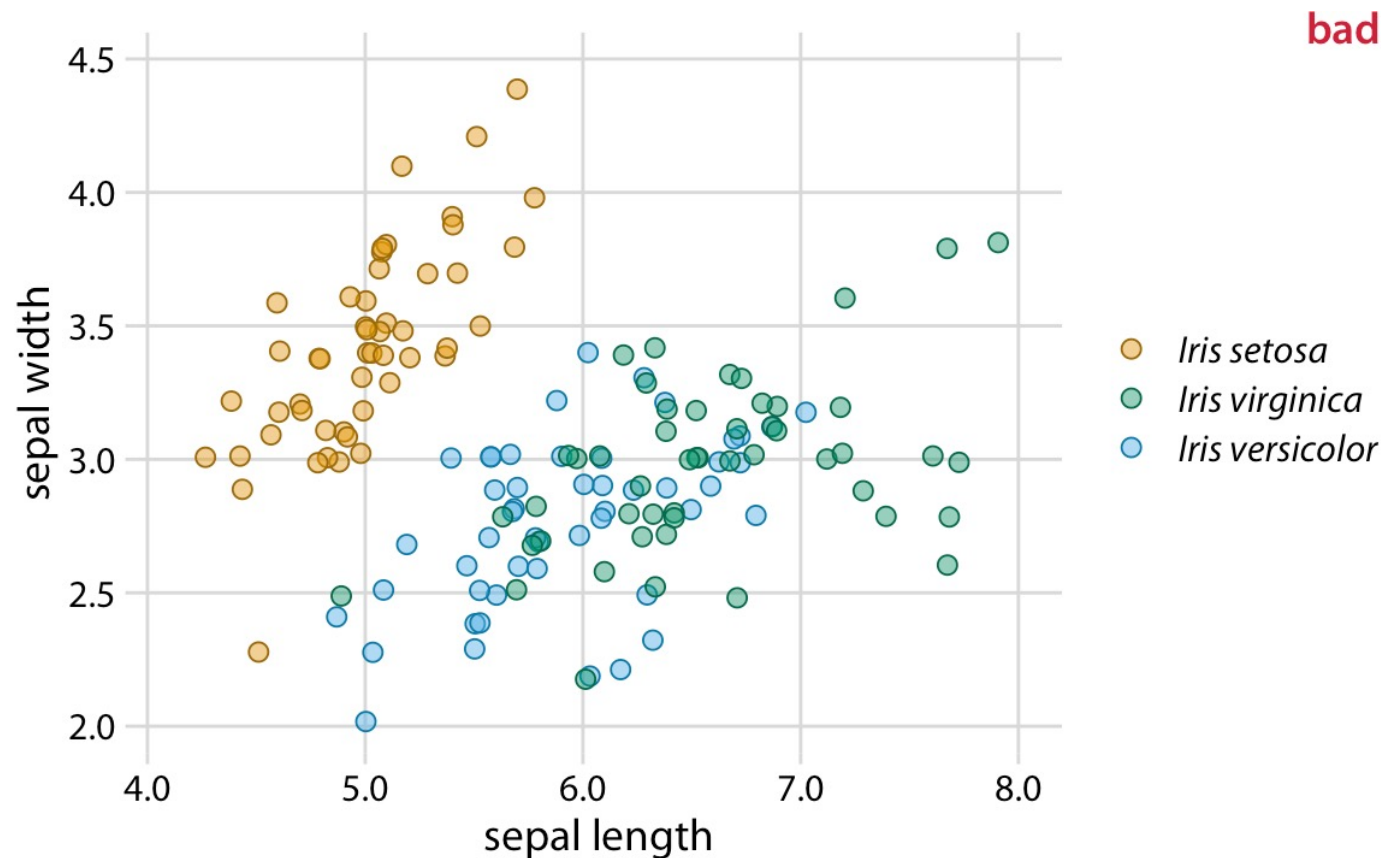
Qualitative color scales work best when there are **three to five** different categories that need to be colored



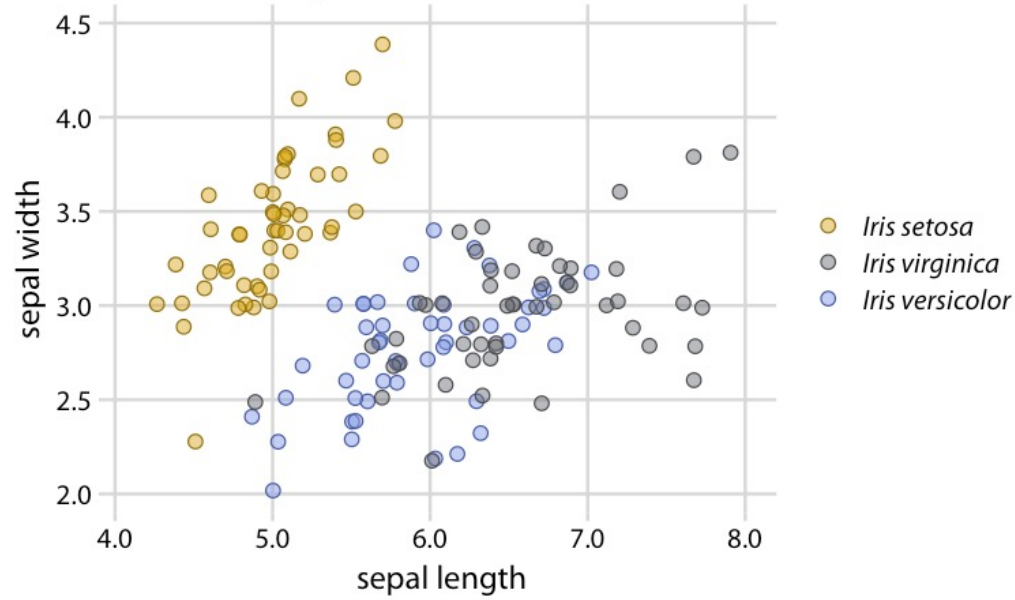
Redundant coding

Visualizing data with multiple encodings will aid in interpretation

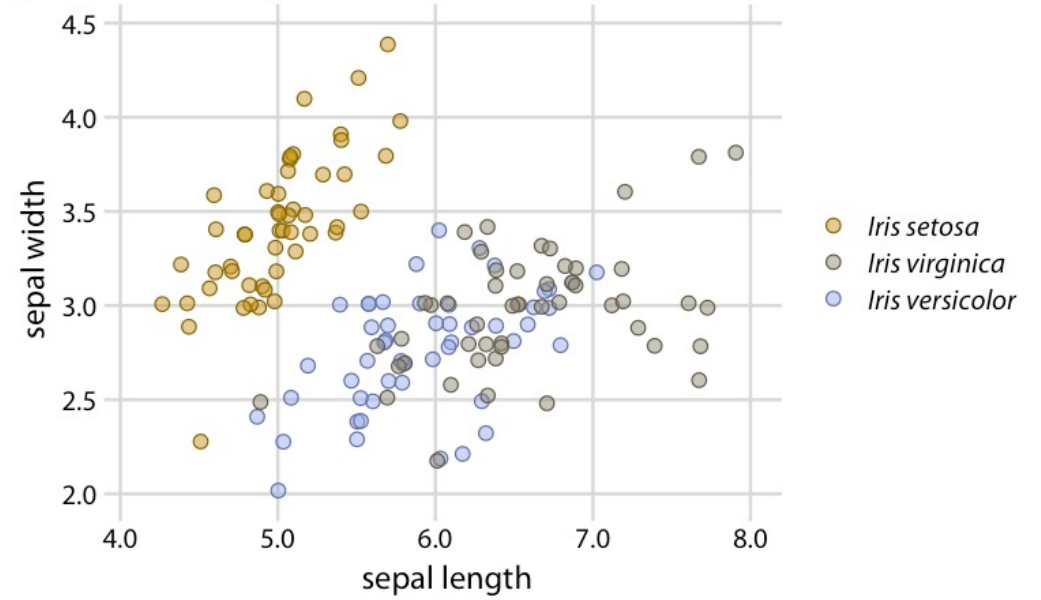
Sepal width versus sepal length for three different iris species.



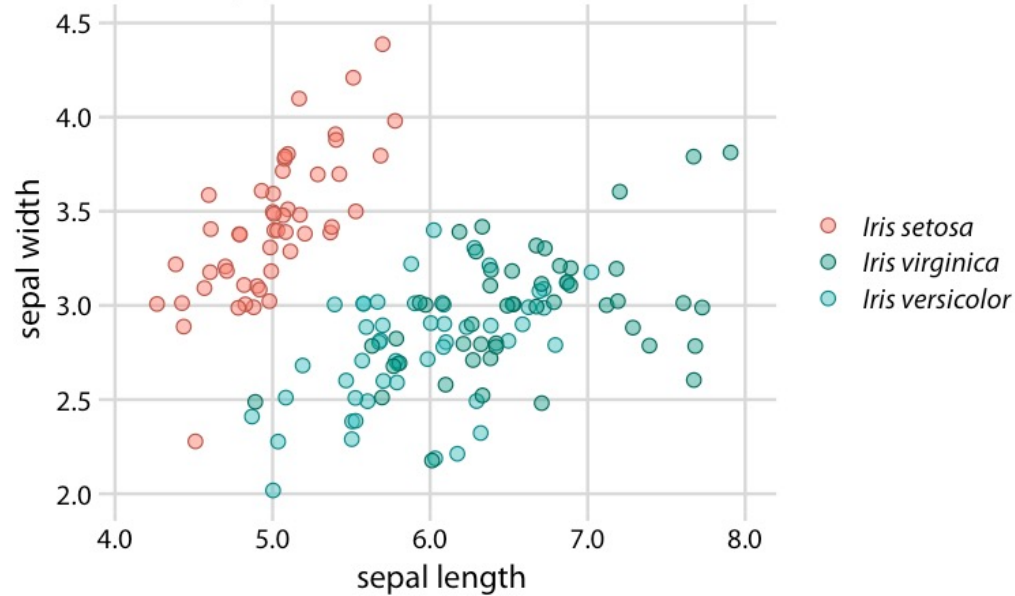
deuteranomaly



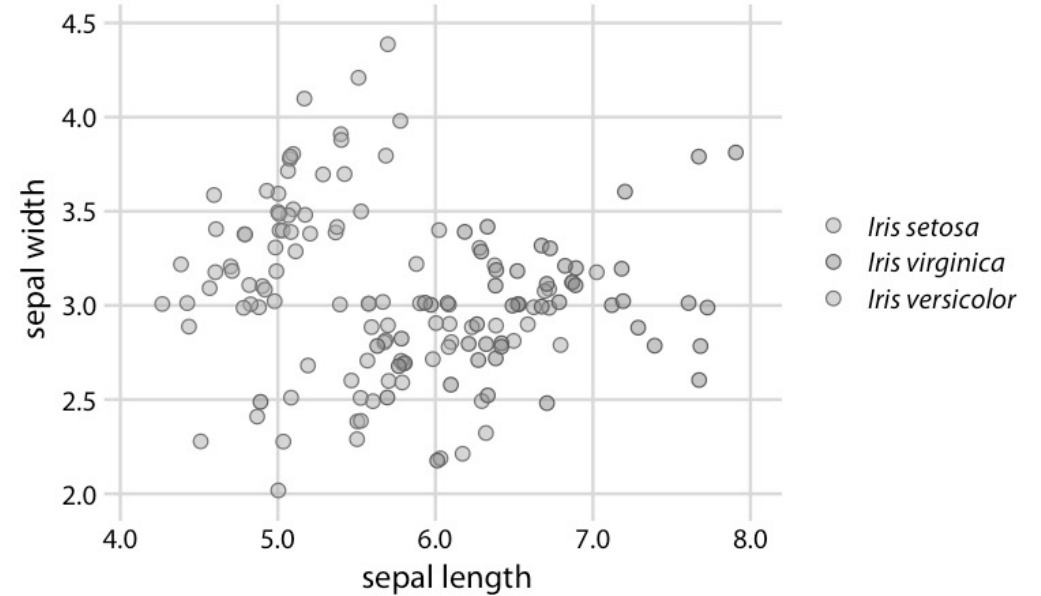
protanomaly



tritanomaly



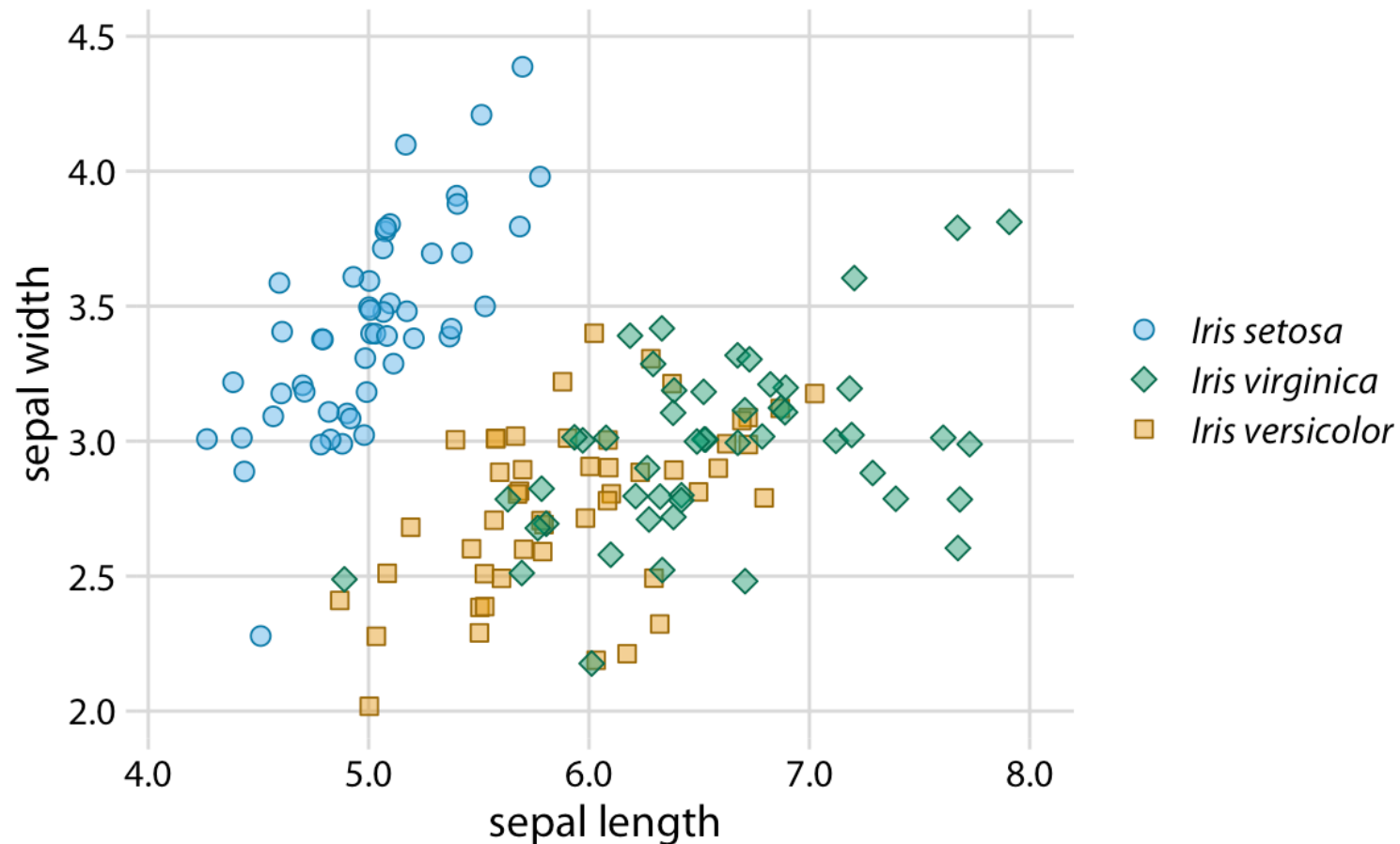
desaturated



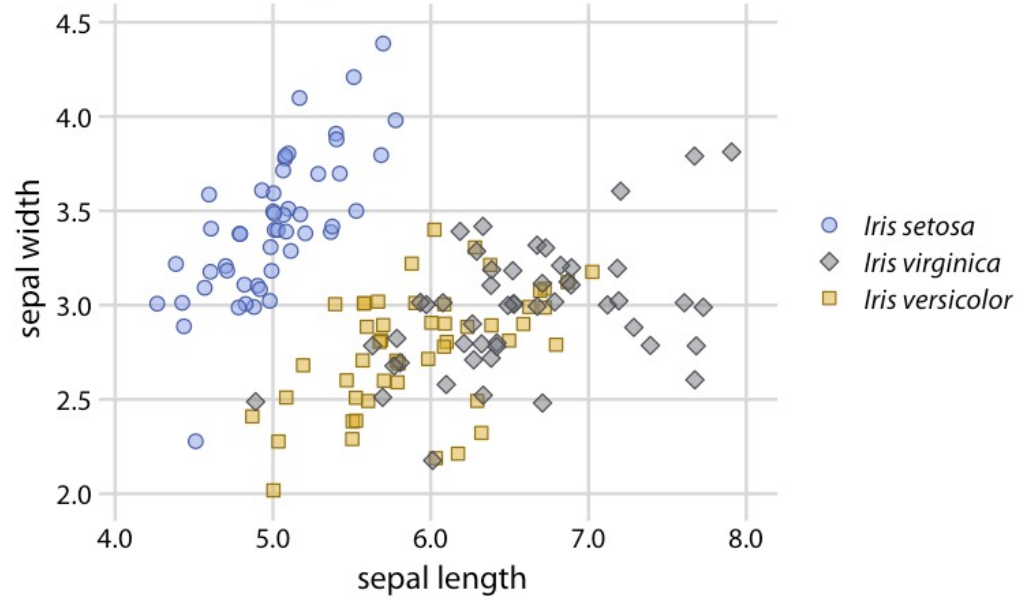
Redundant coding

Visualizing data with multiple encodings will aid in interpretation

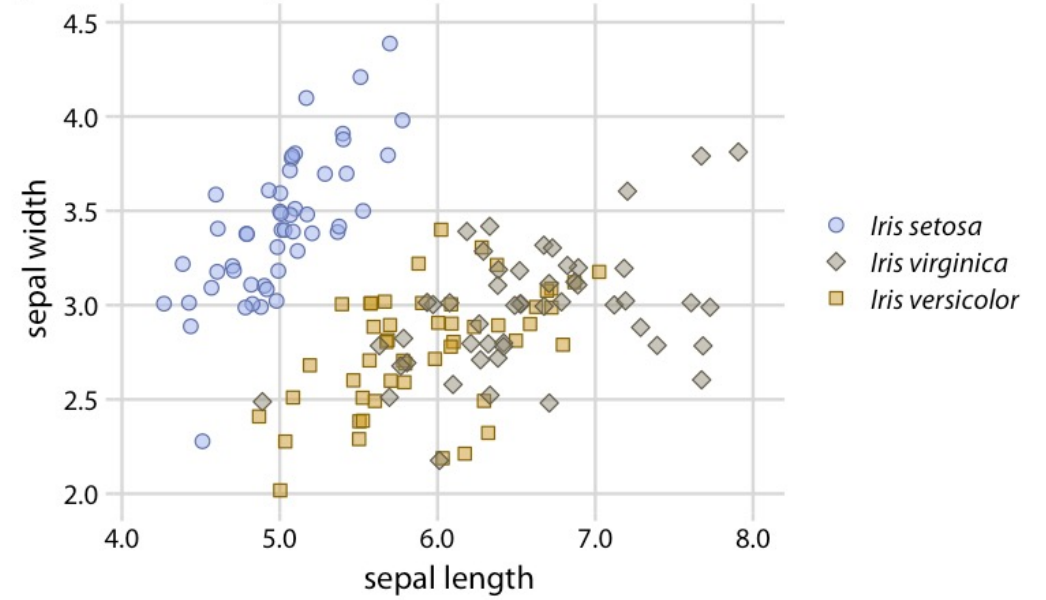
Sepal width versus sepal length for three different iris species.



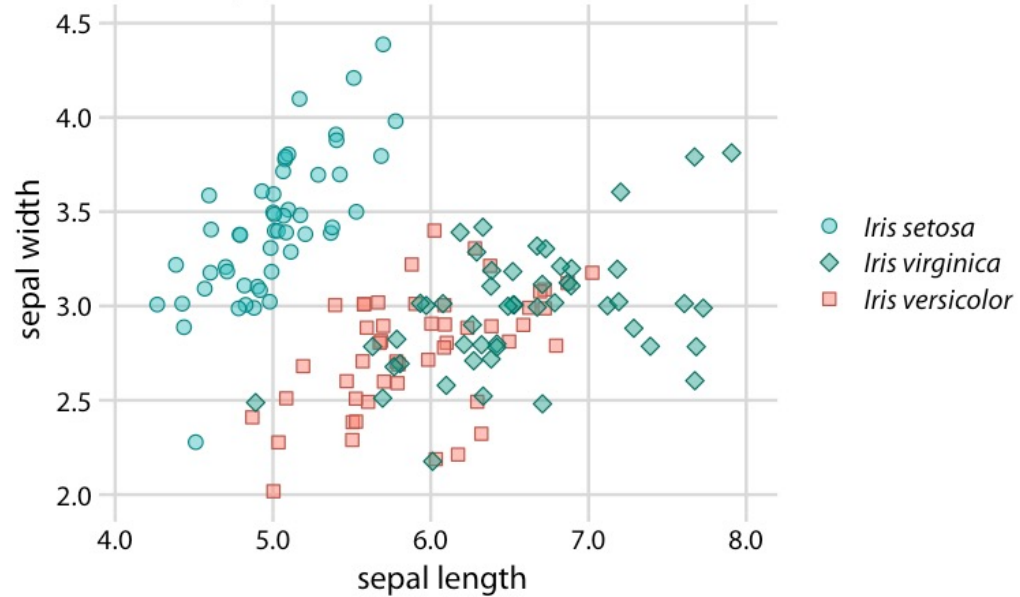
deuteranomaly



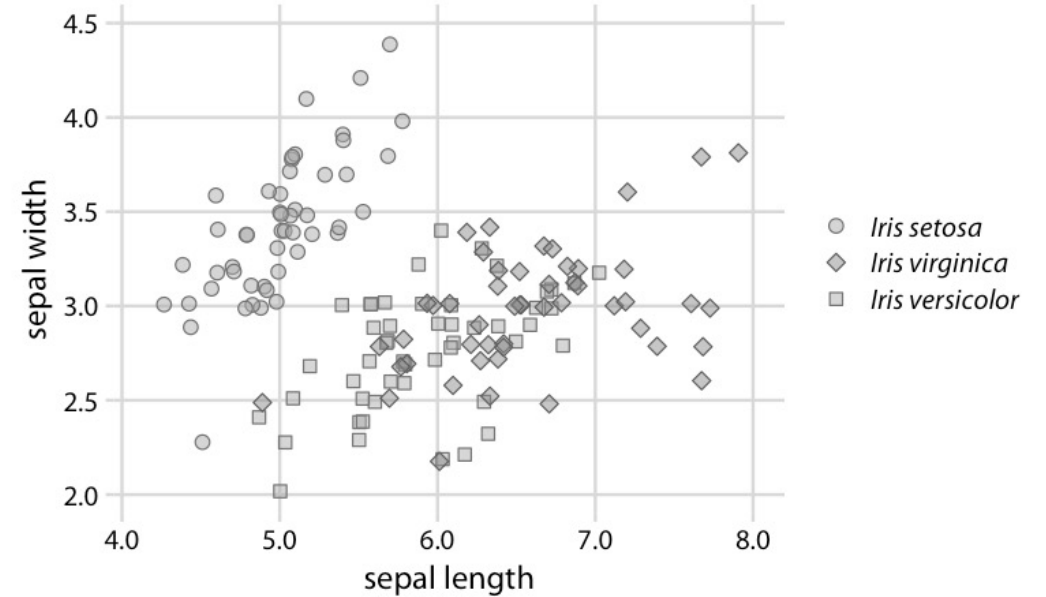
protanomaly



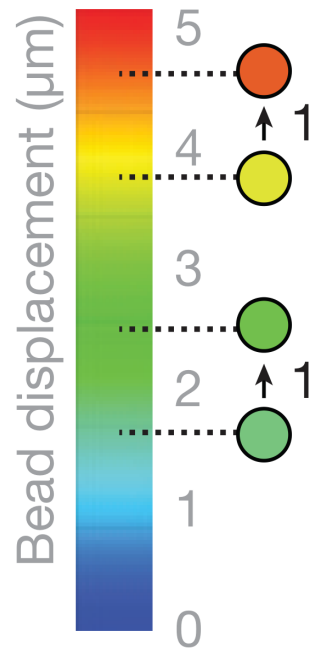
tritanomaly



desaturated

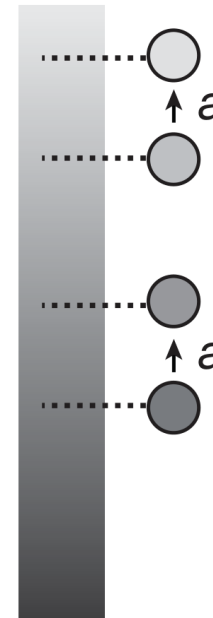


Color can misrepresent data



Avoid the rainbow

Shifts in color hue do not match unit changes in value



Gradation from 10-90% black produces even transitions

Color can misrepresent data

Standard Rainbow Colormap
Perceptually Nonlinear



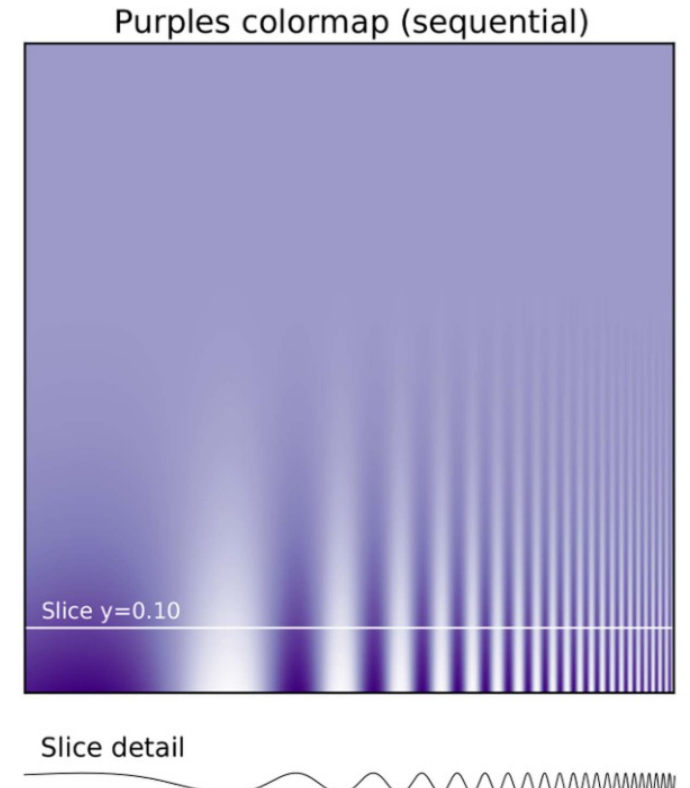
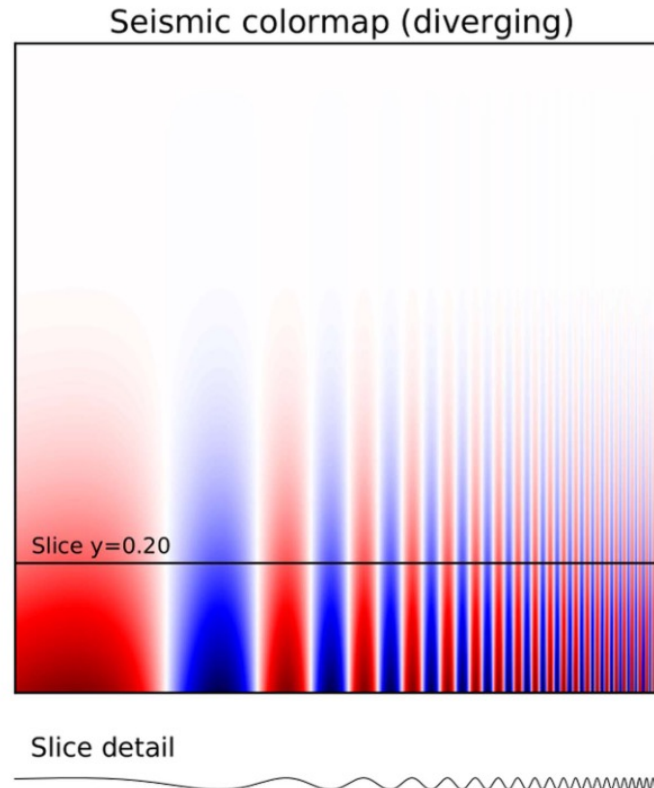
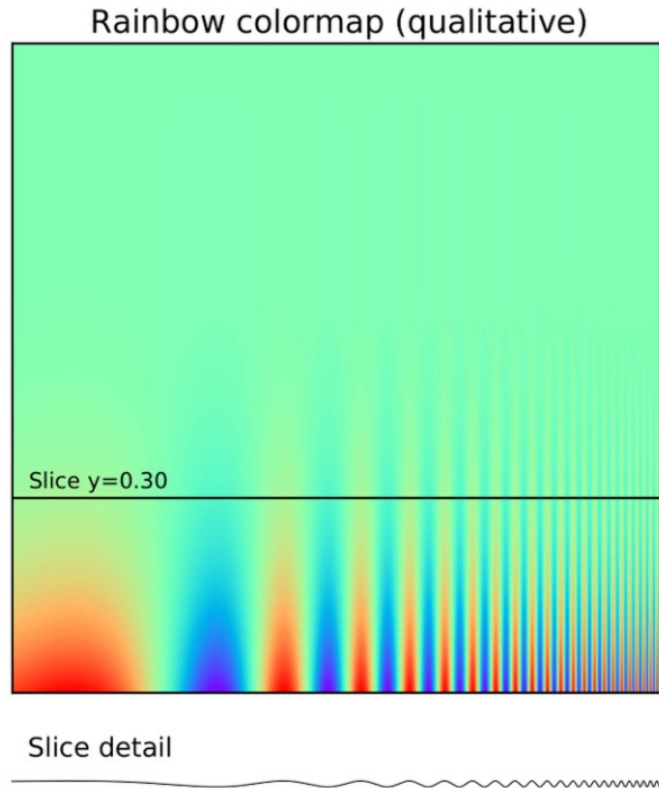
Corrected Rainbow Colormap
Perceptually Linear



Segmented Rainbow Colormap
Categorical Data



Color can misrepresent data

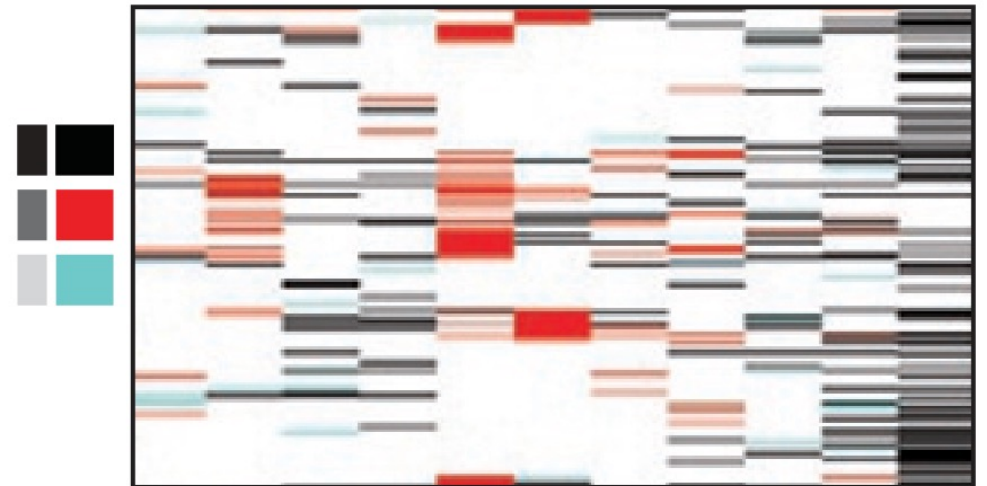


Color can misrepresent data



Use uniform colormaps

Color scales with sharp transitions can exaggerate data ranges



When colors have uneven saturation, data can be underrepresented

Use the appropriate colormap

QUALITATIVE

set1



set2



pastel2



dark2



SEQUENTIAL

blues



greens



reds



ylorbr



DIVERGING

spectral



rdylbu



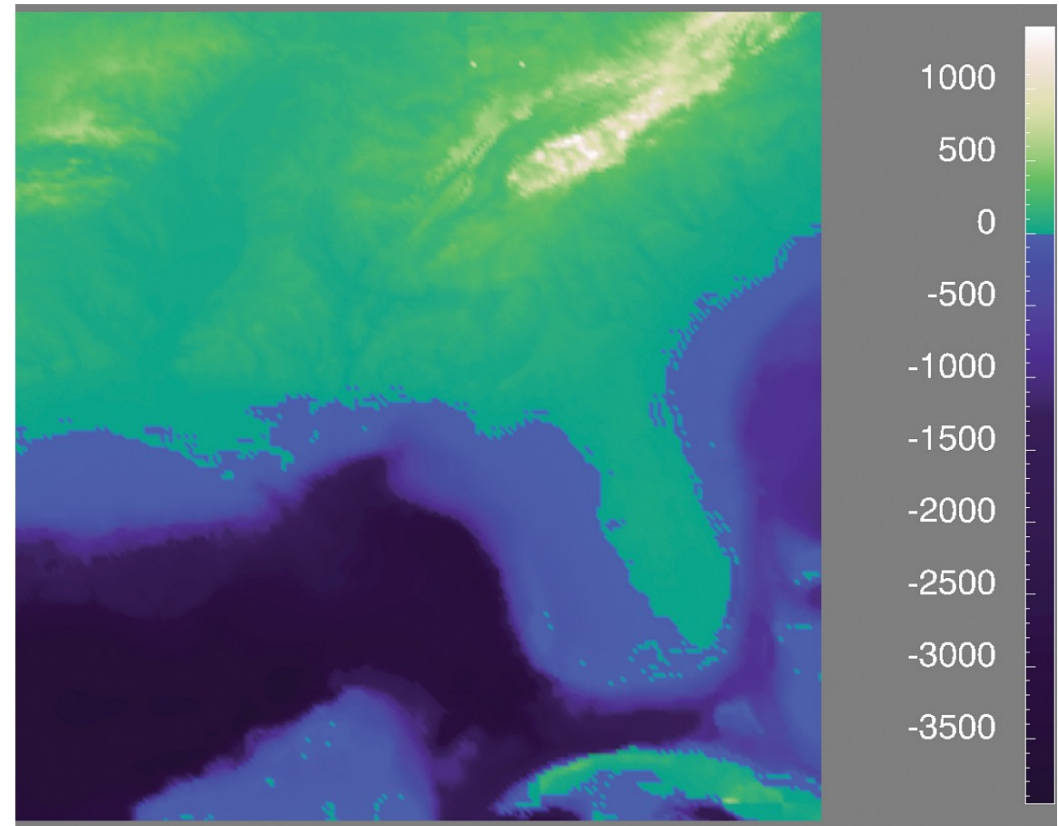
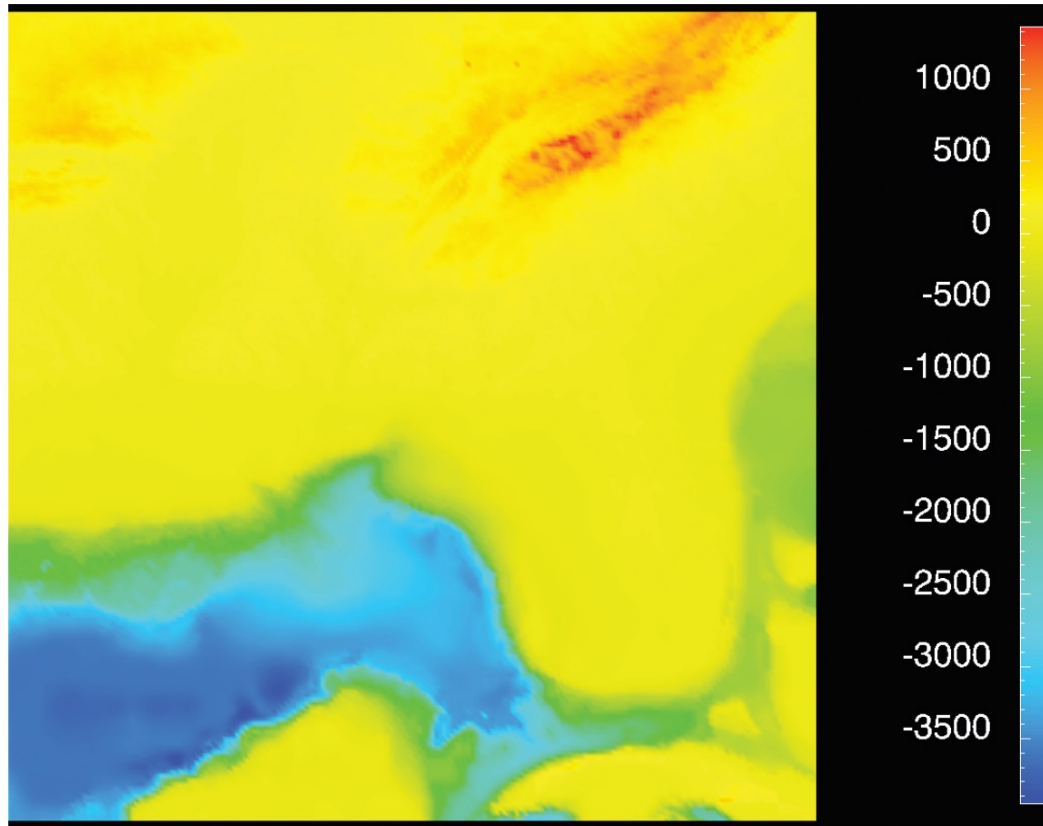
rdylgn



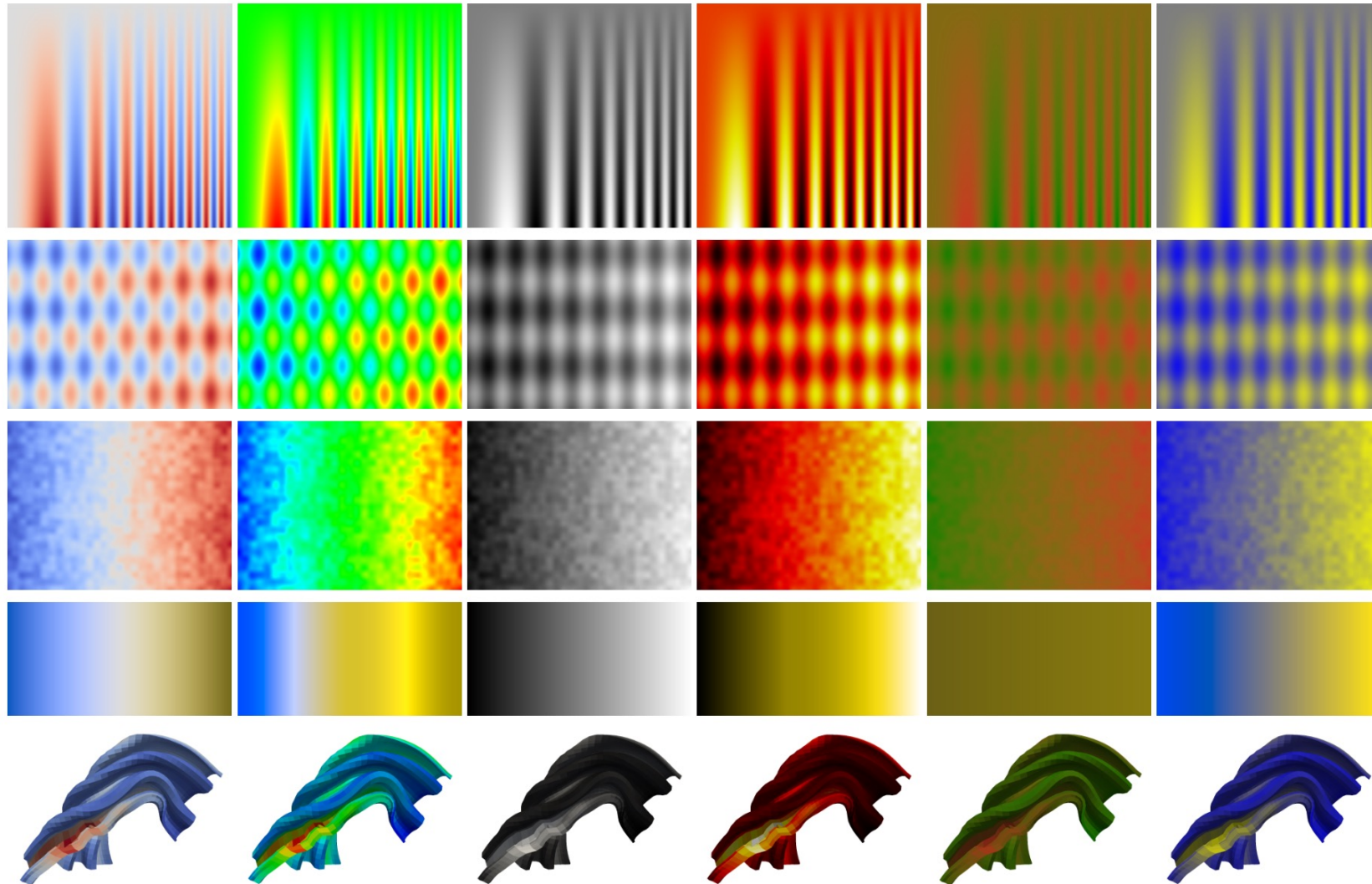
piyg



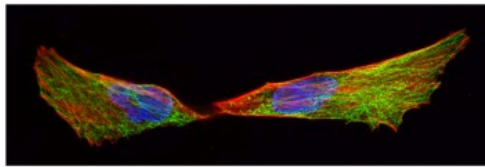
Use the appropriate colormap



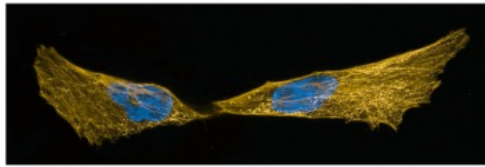
Use color effectively



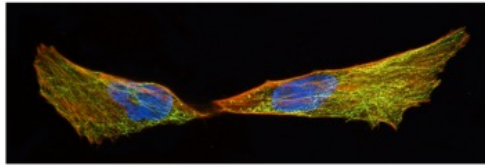
Be considerate of colorblindness



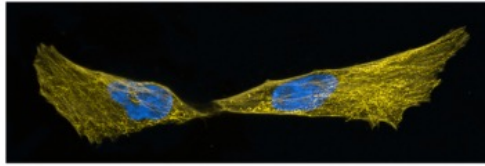
Wild-type photoreceptors



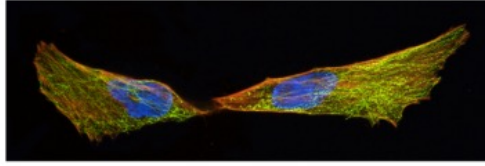
Deuteranopia (no green)



Deuteranomaly (reduced green)



Protanopia (no red)



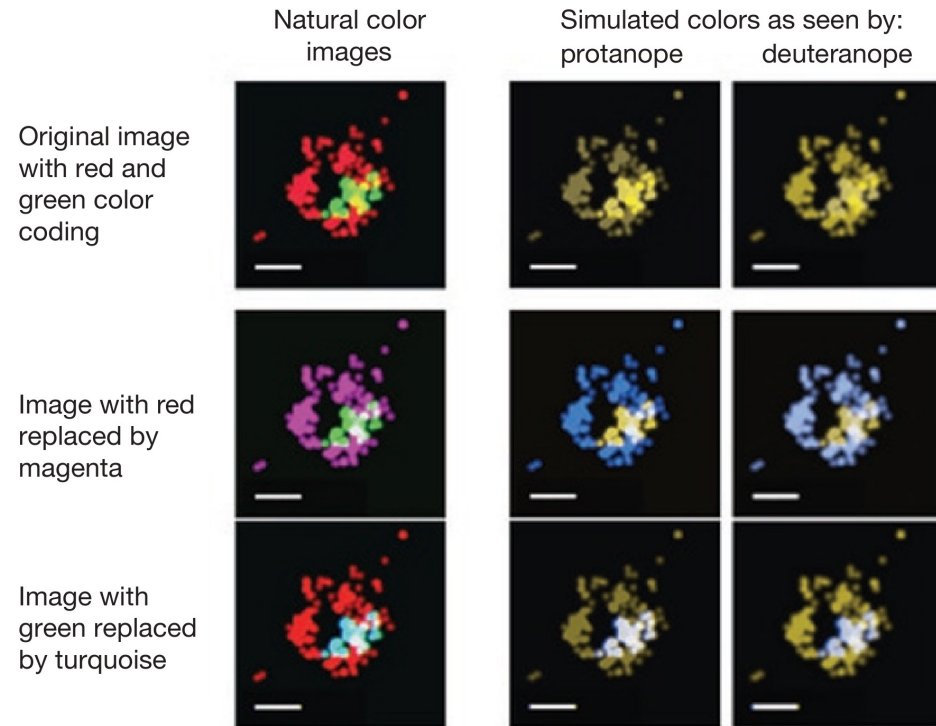
Protanomaly (reduced red)

Your next reviewer may be colorblind

Up to **8%** of males and **.5%** of females have some form of color blindness

For three reviewers you have (at worst) an over **22%** chance of being assigned a colorblind reviewer.

Choose a colorblind friendly palette



Avoid red and green color combinations

Color	Color name	RGB (1–255)	CMYK (%)	P	D
	Black	0, 0, 0	0, 0, 0, 100		
	Orange	230, 159, 0	0, 50, 100, 0		
	Sky blue	86, 180, 233	80, 0, 0, 0		
	Bluish green	0, 158, 115	97, 0, 75, 0		
	Yellow	240, 228, 66	10, 5, 90, 0		
	Blue	0, 114, 178	100, 50, 0, 0		
	Vermillion	213, 94, 0	0, 80, 100, 0		
	Reddish purple	204, 121, 167	10, 70, 0, 0		

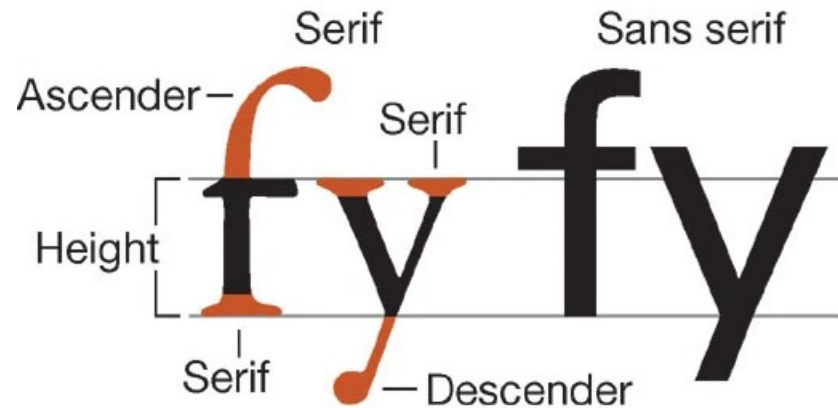
Use a colorblind friendly color palette

Refine the text style

Typography matters.

Typography

Use one font-family

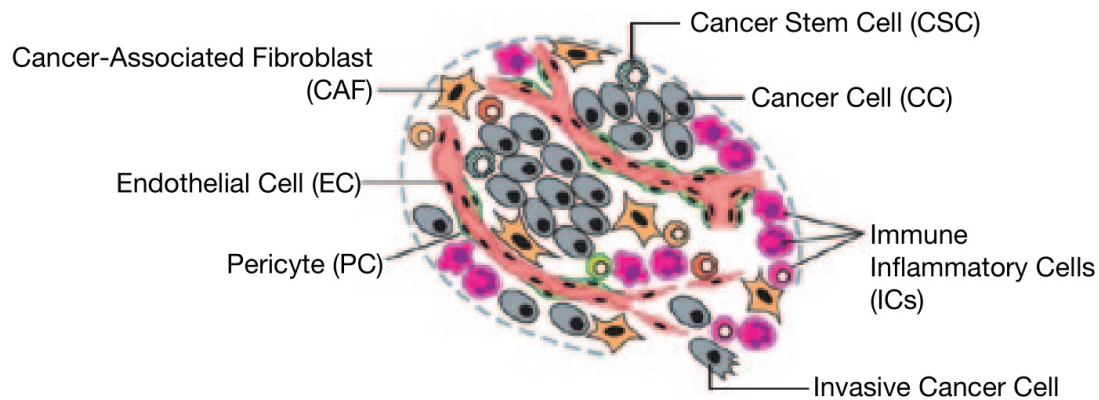


Use **Serif** for text, **Sans serif** for figures

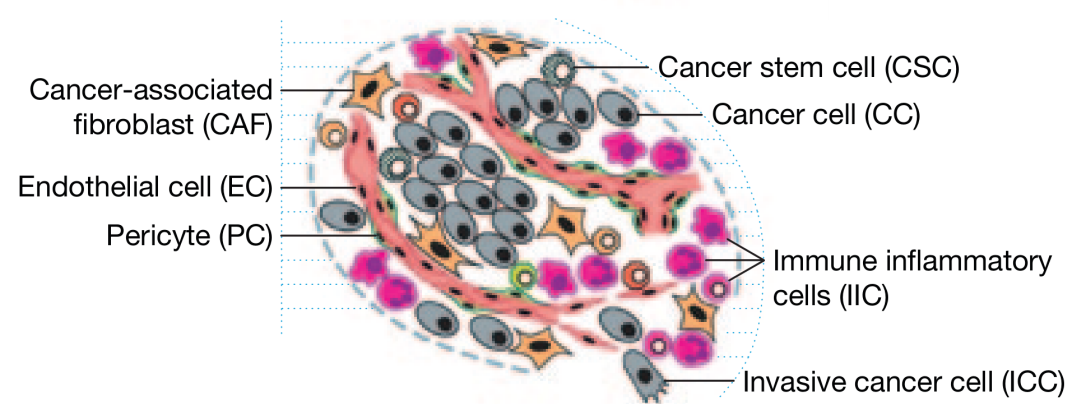
Baskerville
Helvetica
Palatino
Times New Roman

Arrows and Labels

Align figure callout lines and labels



Unaligned figure



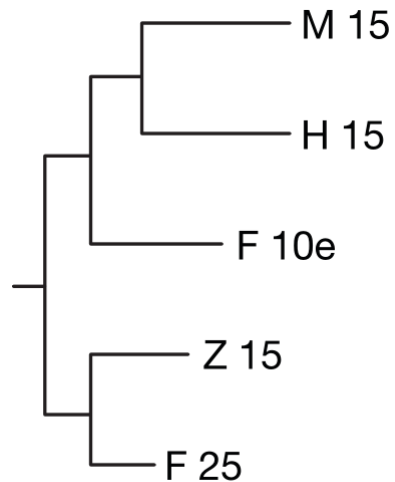
Callout labels are aligned when feasible

Use horizontal callout lines or fixed angle lines (30 or 45°) when not possible.

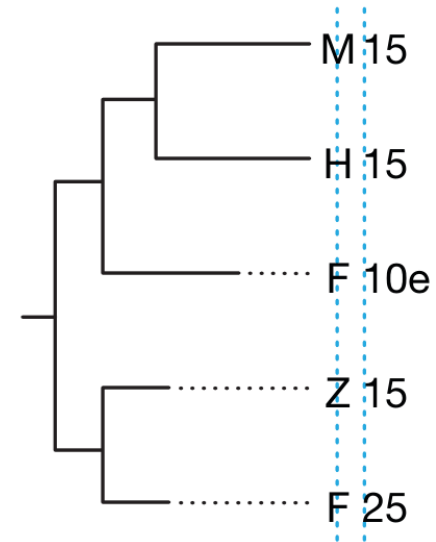
Align labels when necessary, otherwise follow the curve of the schematic.

Arrows and Labels

Align figure callout lines and labels



Unaligned figure



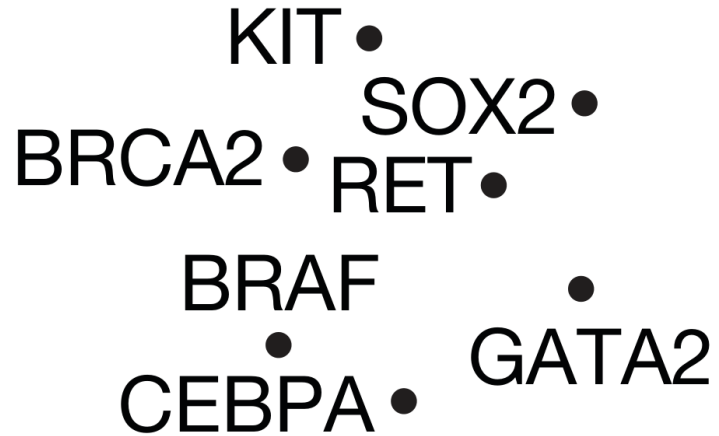
Callout labels are aligned when feasible

Use horizontal callout lines or fixed angle lines (30 or 45°) when not possible.

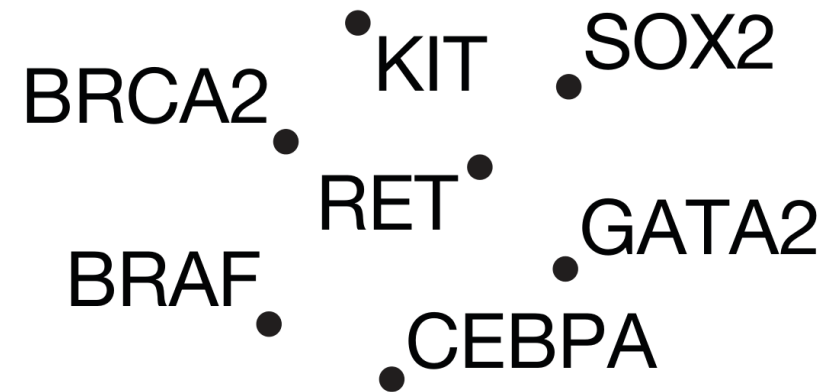
Align labels when necessary, otherwise follow the curve of the schematic.

Arrows and Labels

Position labels consistently



Ambiguous label placement

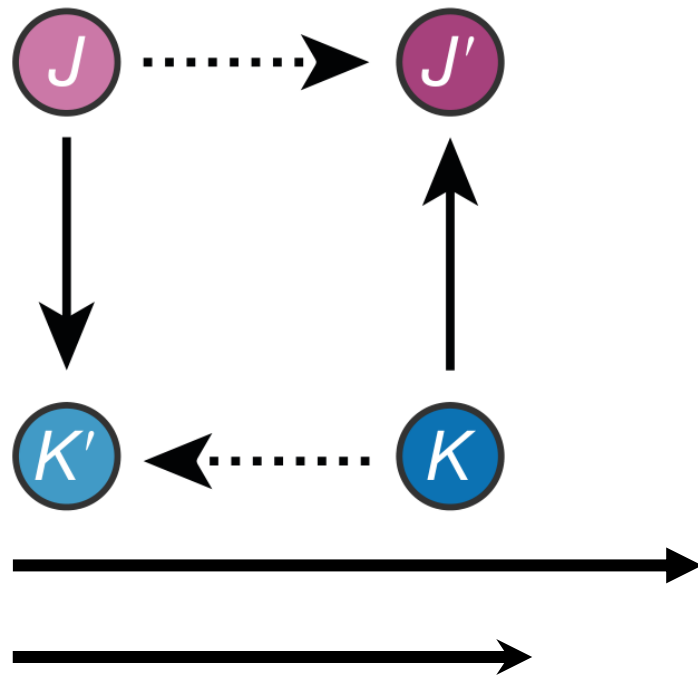


Good label placement

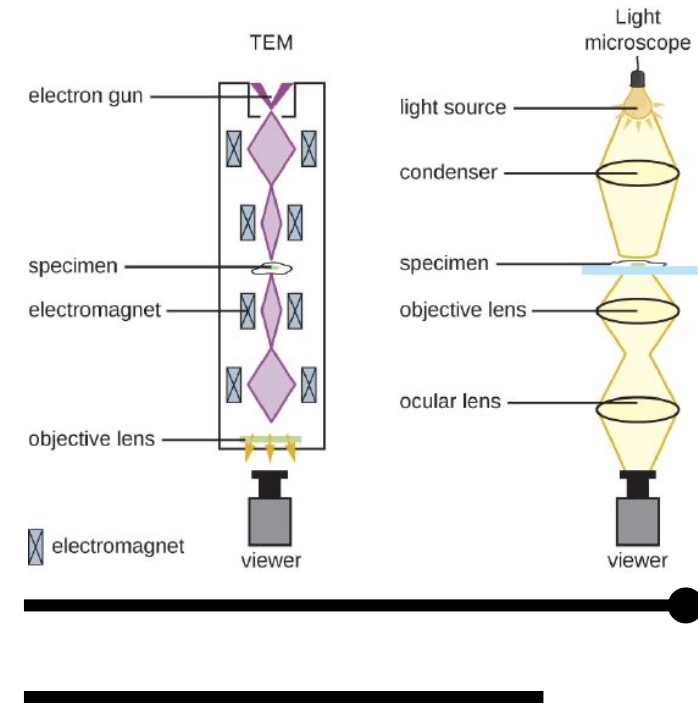
Distance and alignment of labels should be fixed

Arrows and Labels

Use arrows sparingly and effectively



Use arrows for the **functional** inter-relatedness of parts



Use lines for the **spatial** inter-relatedness of parts

Challenge

Redesign a data visualization

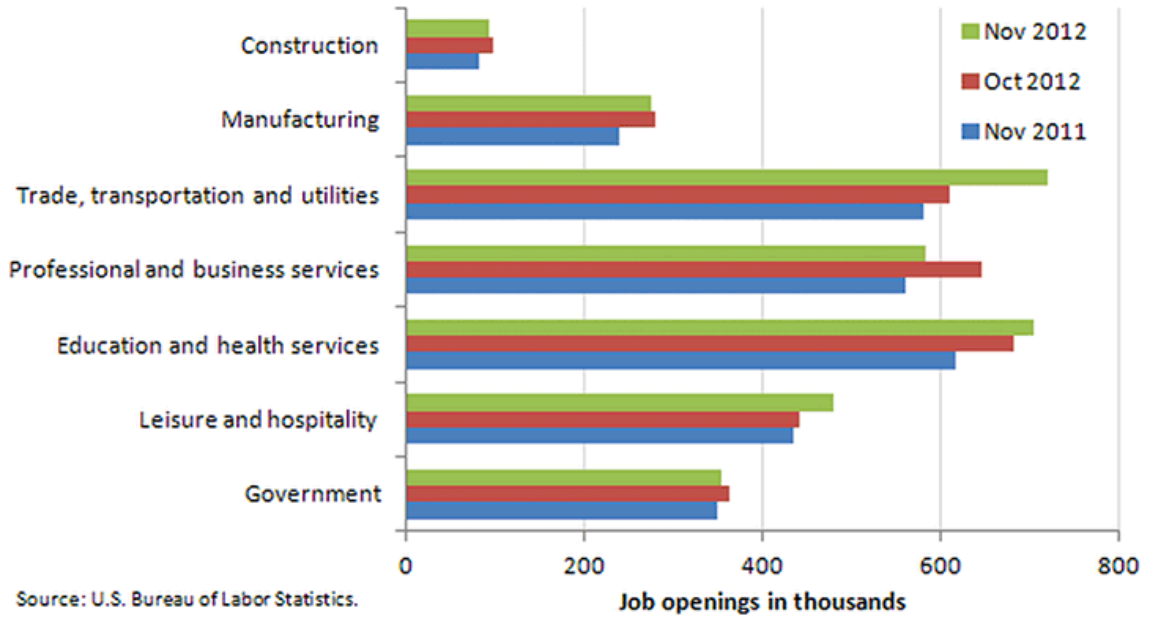
Job openings in November 2012

JANUARY 11, 2013

There were 3.7 million job openings on the last business day of November 2012, unchanged from October 2012. In November 2011 there were 3.3 million job openings.

CHART IMAGE CHART DATA

Job openings by industry, November 2011, October 2012 and November 2012, seasonally adjusted

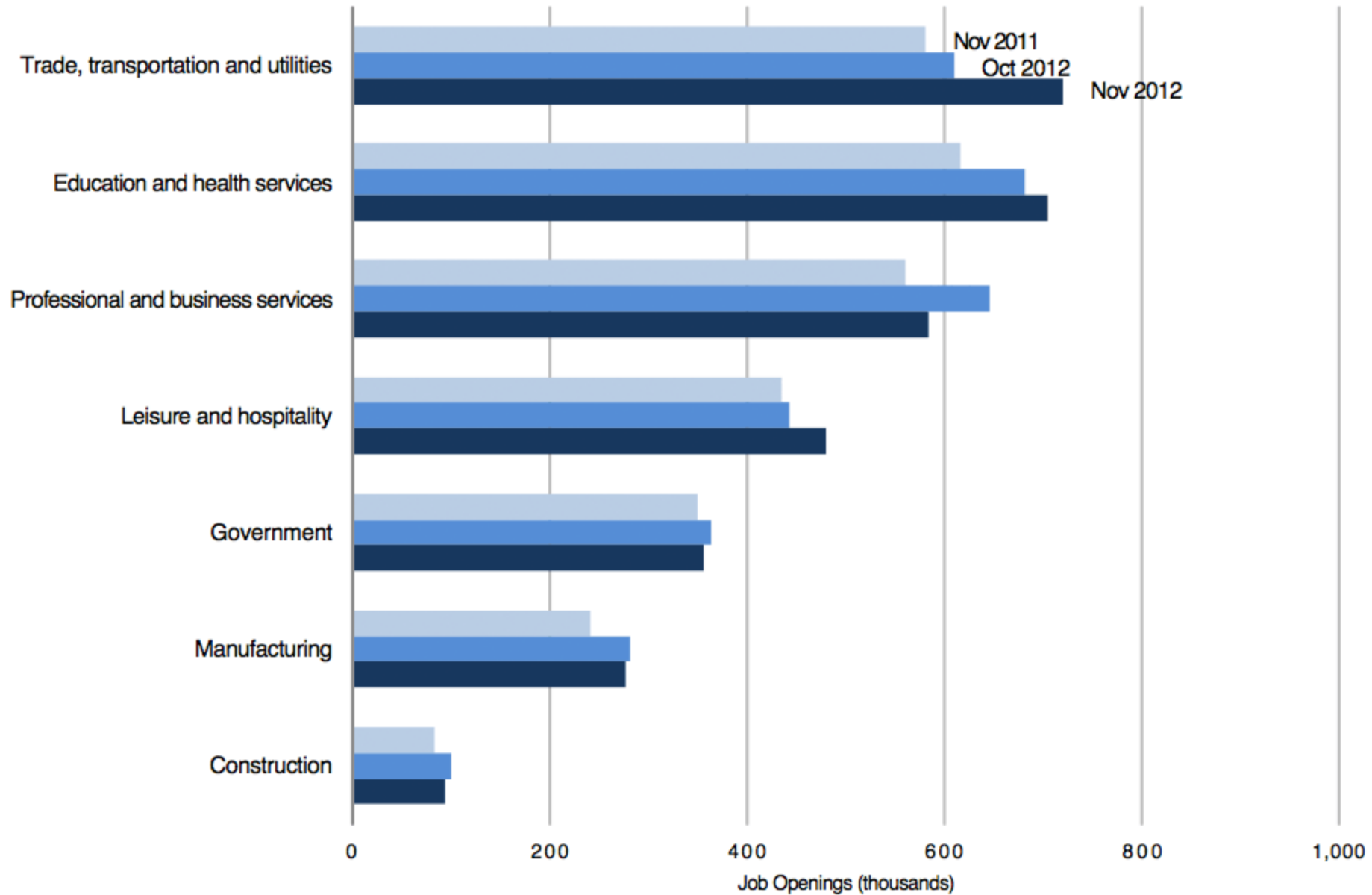


From November 2011 to November 2012, job openings increased most in retail trade (144,000, within the trade, transportation and utilities industry) and health care and social assistance (91,000, within the education and health services industry).

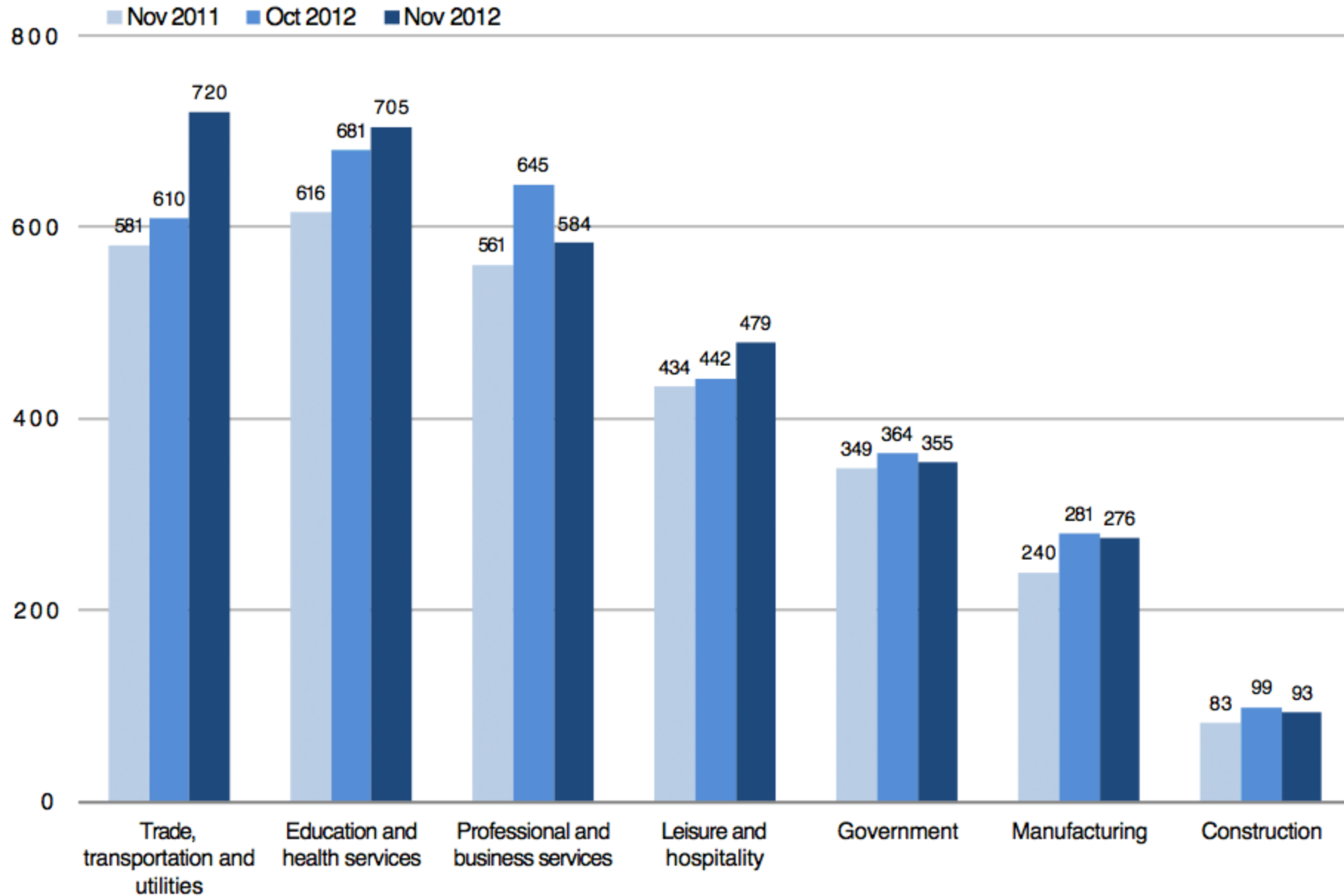
Government job openings increased the least, by 6,000.

These data are from the [Job Openings and Labor Turnover Survey](#). Data for the most recent month are preliminary and subject to revision. For additional information, see [Job Openings and Labor Turnover — November 2012](#) (HTML) (PDF), news release USDL-13-0015. More charts featuring data on job openings, hires, and employment separations can be found in [Job Openings and Labor Turnover Survey Highlights: November 2012](#) (PDF).

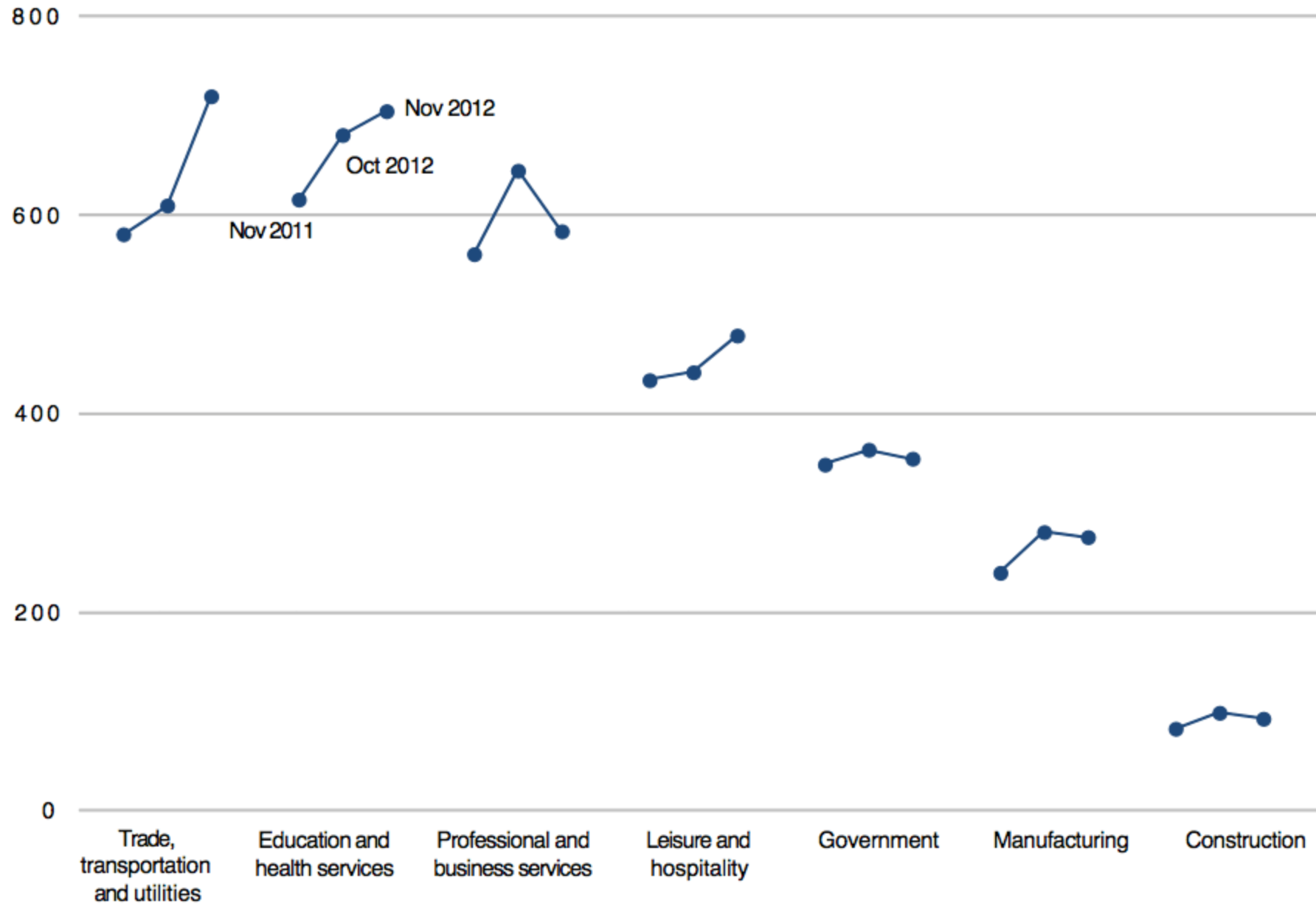
Job openings by industry, November 2011, October 2012 and November 2012, seasonally adjusted



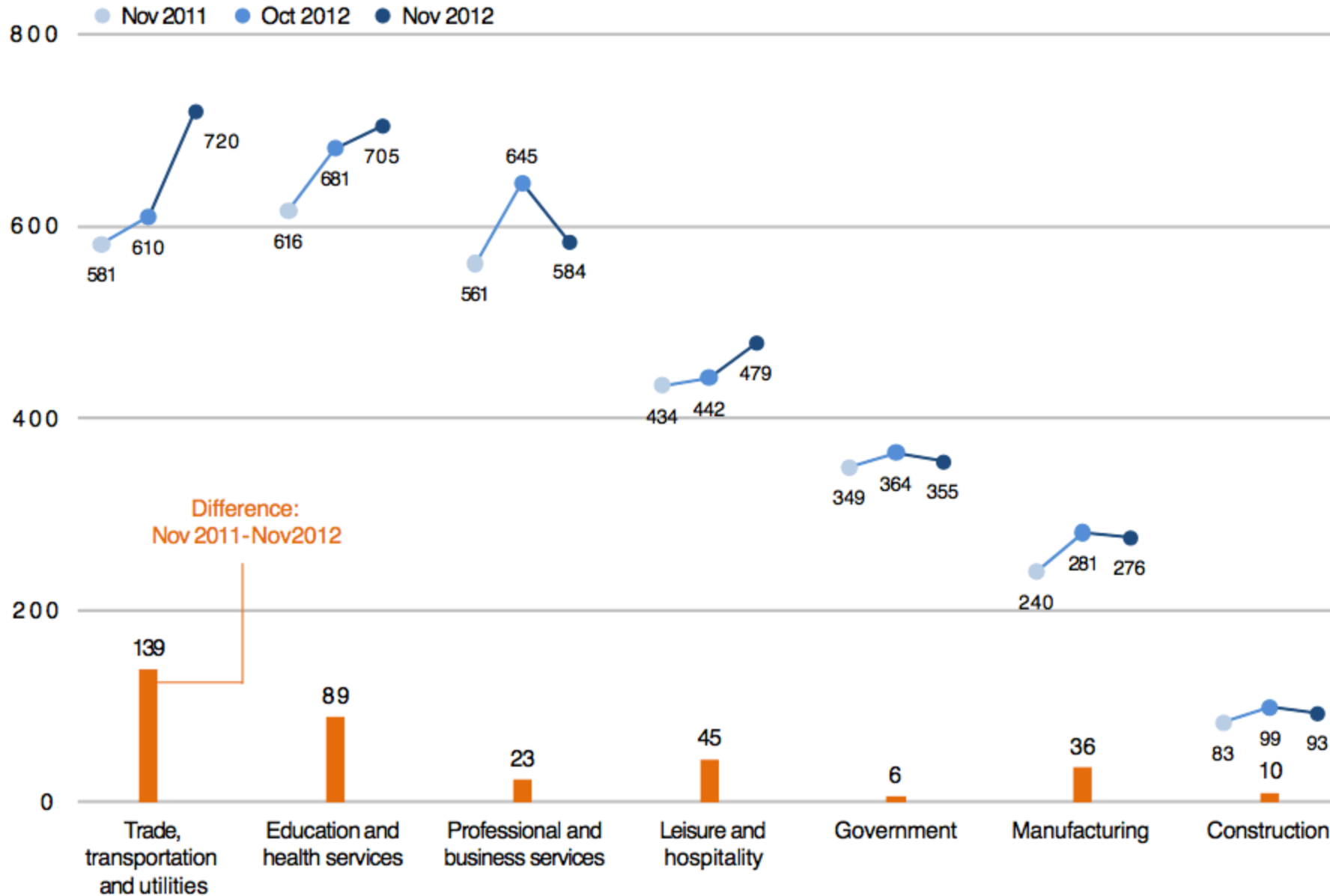
Job openings by industry, November 2011, October 2012 and November 2012, seasonally adjusted
(Thousands of jobs)



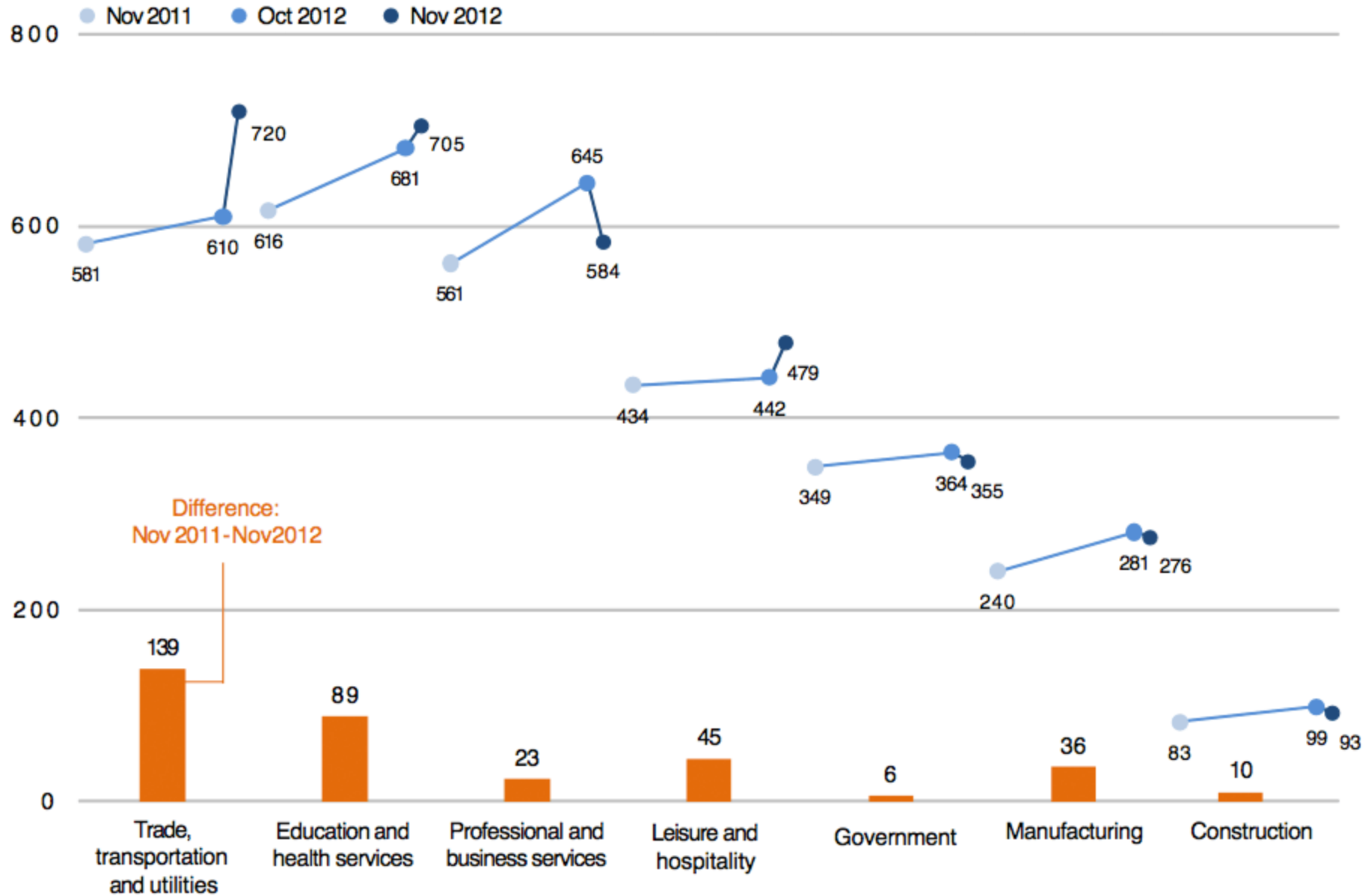
Job openings by industry, November 2011, October 2012 and November 2012, seasonally adjusted
 (Thousands of jobs)



Job openings by industry, November 2011, October 2012 and November 2012, seasonally adjusted
 (Thousands of jobs)



Job openings by industry, November 2011, October 2012 and November 2012, seasonally adjusted
(Thousands of jobs)



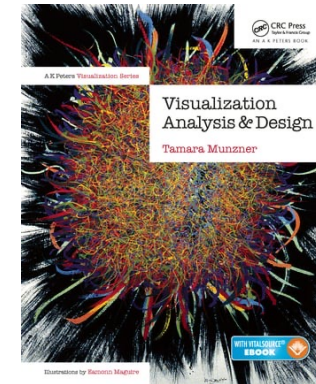
Design resources

Recommended reading



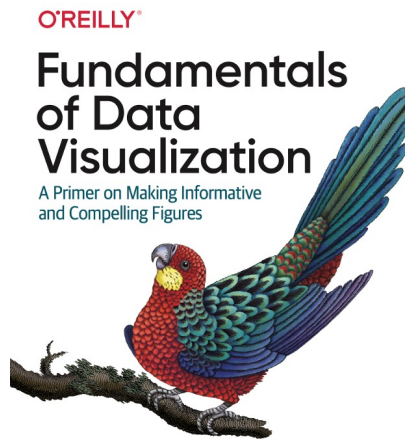
Points of View

Bang Wong and Martin Krywinski



Visualization Analysis and Design

Tamara Munzner



Claus O. Wilke



Fundamentals of Data Visualization

Claus Wilke



Color Map Advice for Scientific Visualization

Kenneth Moreland

Checklist

is your figure effective?

- The figure is **self contained**: understandable without additional information
- Every element is **labelled** or explained in the caption, including x and y units
- x and y axis: **scales** show appropriate variation of the data, or are comparable
- Readability** and **contrast** are appropriate
- Every use of **colour** has a reason
- The figure works in **grayscale** (except for very complex figures)
- If there are **groupings**, they help understand the message without manipulating
- There are no channel **inconsistencies** within the figure
- It is as **simple** as possible: i.e. no decorations, every piece that could be eliminated without losing information has been eliminated
- Has been **validated** with other people...

Thank you!

any questions?



I HAVE NO IDEA HOW
TO CHANGE EXCEL
GRAPH COLOURS



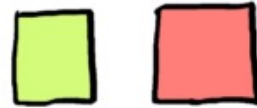
I CRAVE BLANDNESS
IN ALL THINGS



I THINK GRAY
SCALE IS TOO ARTSY



I WANT PEOPLE TO SEE
MY GRAPHS FROM SPACE



I HATE COLOUR-
BLIND PEOPLE



OMG UNICORNS!

WHAT YOUR GRAPH COLOUR PALLET SAYS ABOUT YOU
ERRANTSCIENCE.COM