

Effective data visualization

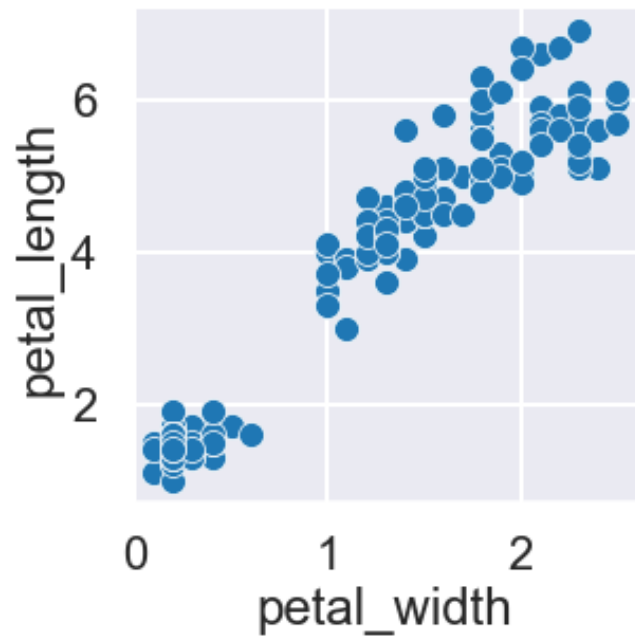
Oliver W. Layton

CS251: Data Analysis and Visualization

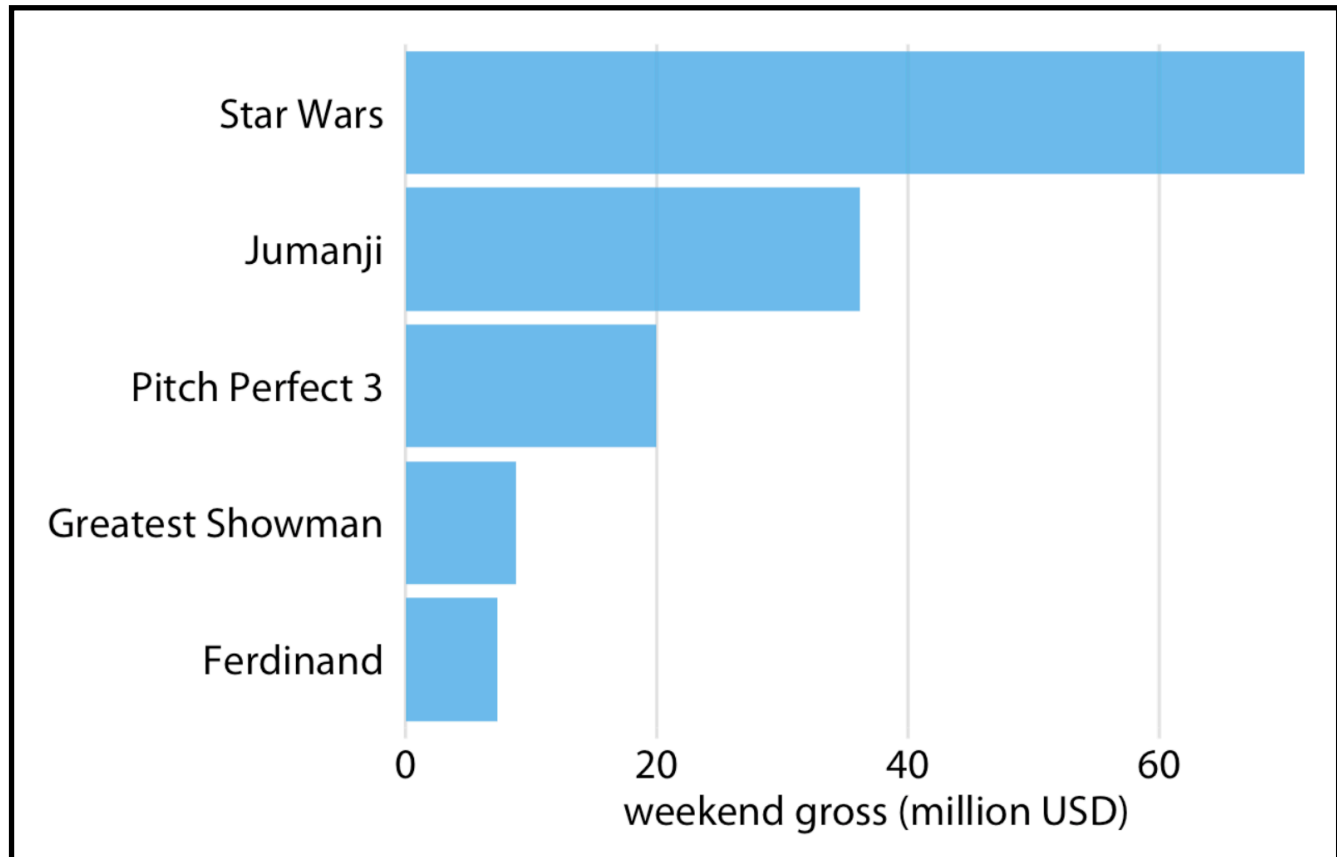
Lecture 2

Spring 2019

Some common data visualizations

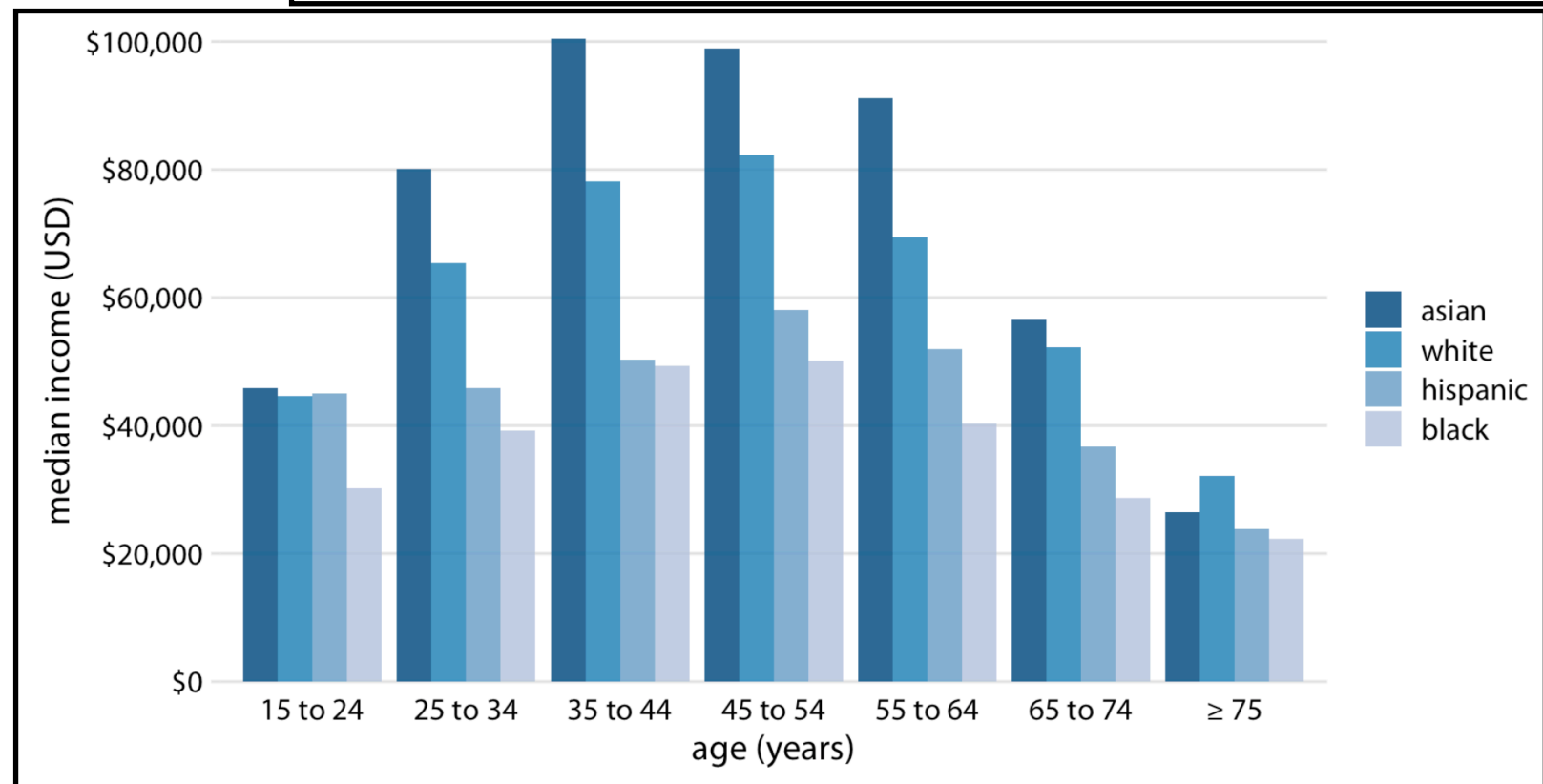


Scatter plot



Bar plot

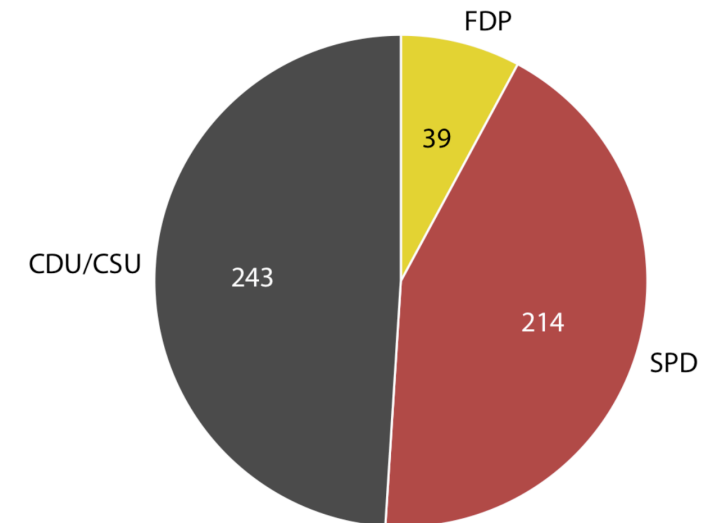
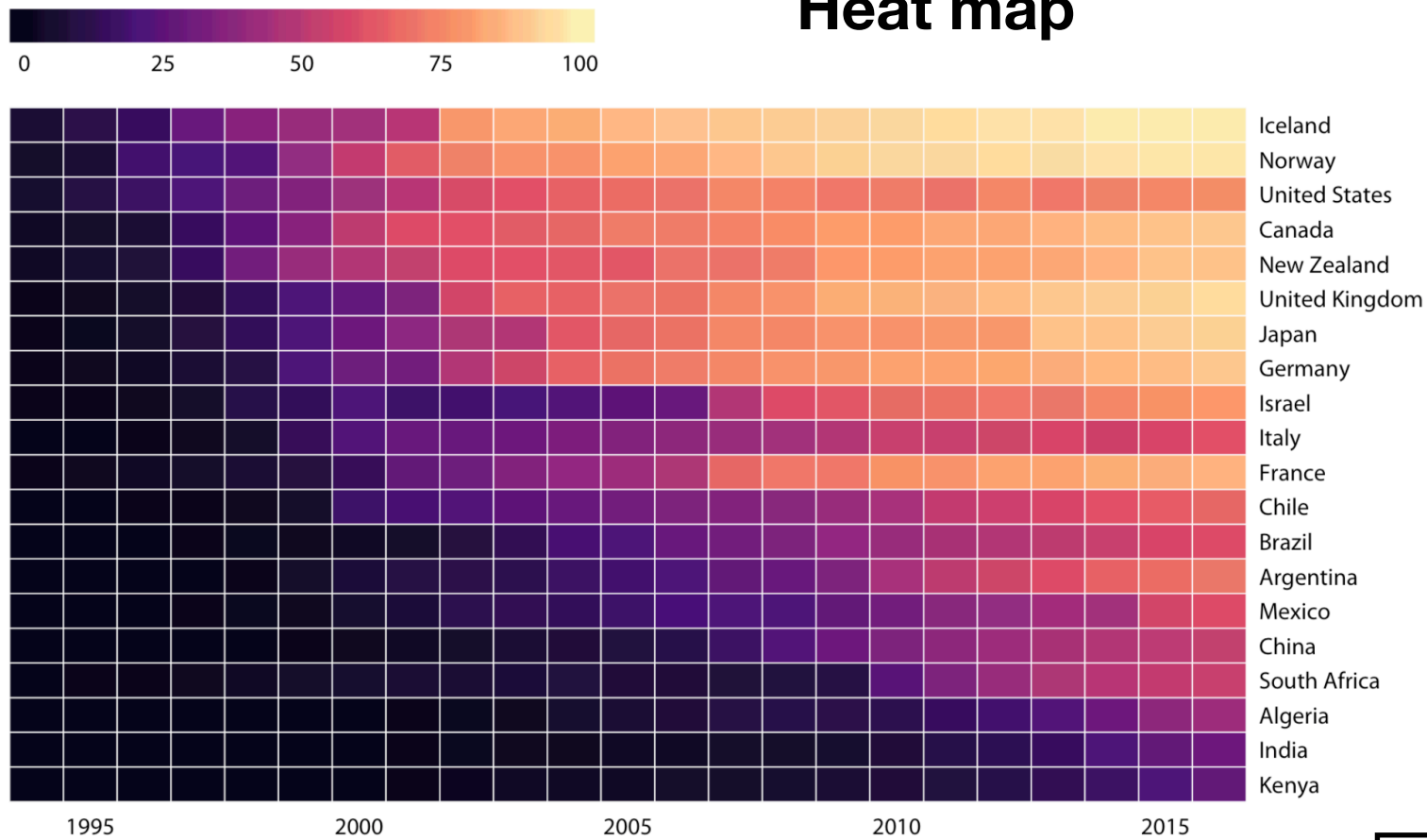
Side-by-side Bar plot



Some common data visualizations

internet users / 100 people

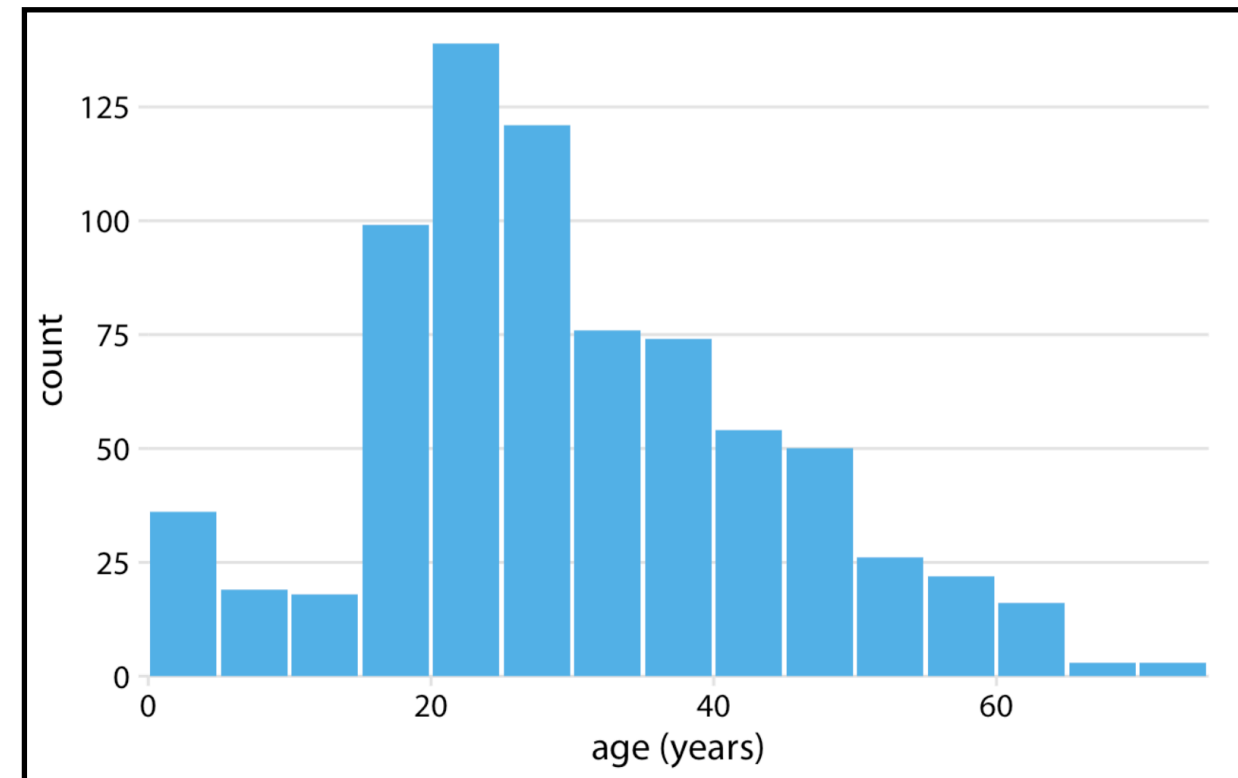
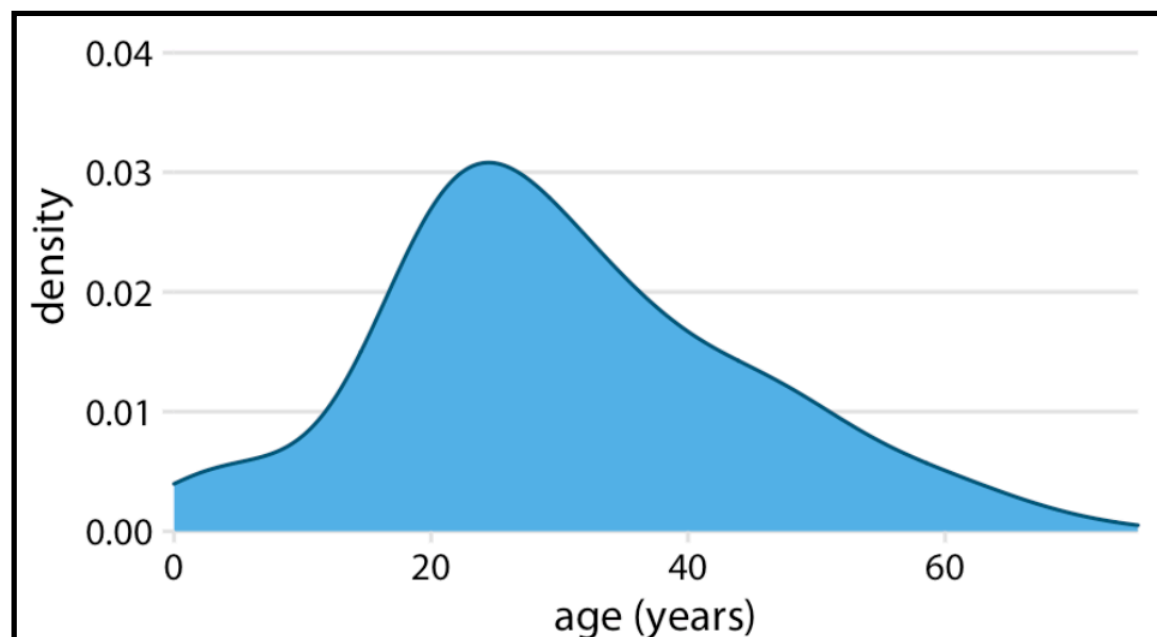
Heat map



Pie Chart

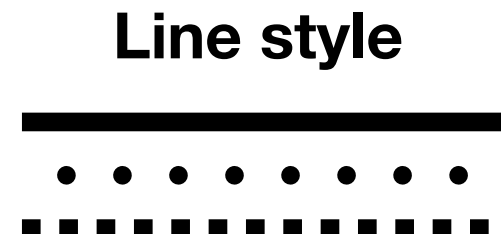
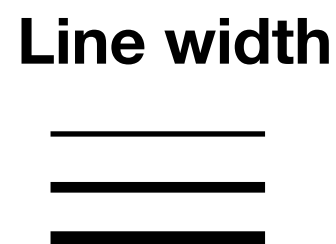
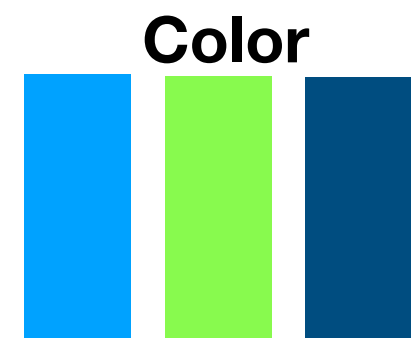
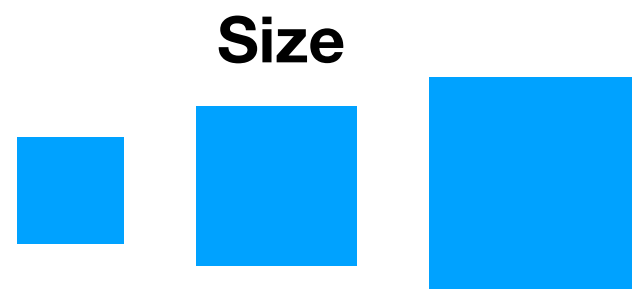
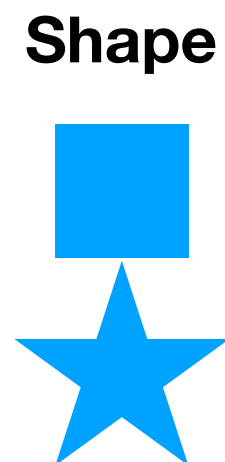
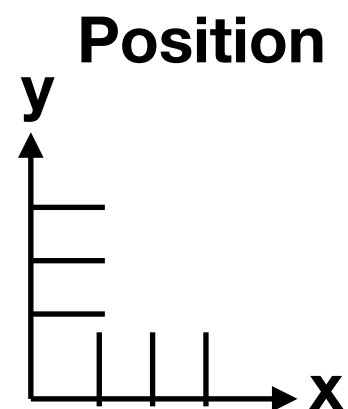
Histogram

Density plot



Aesthetics

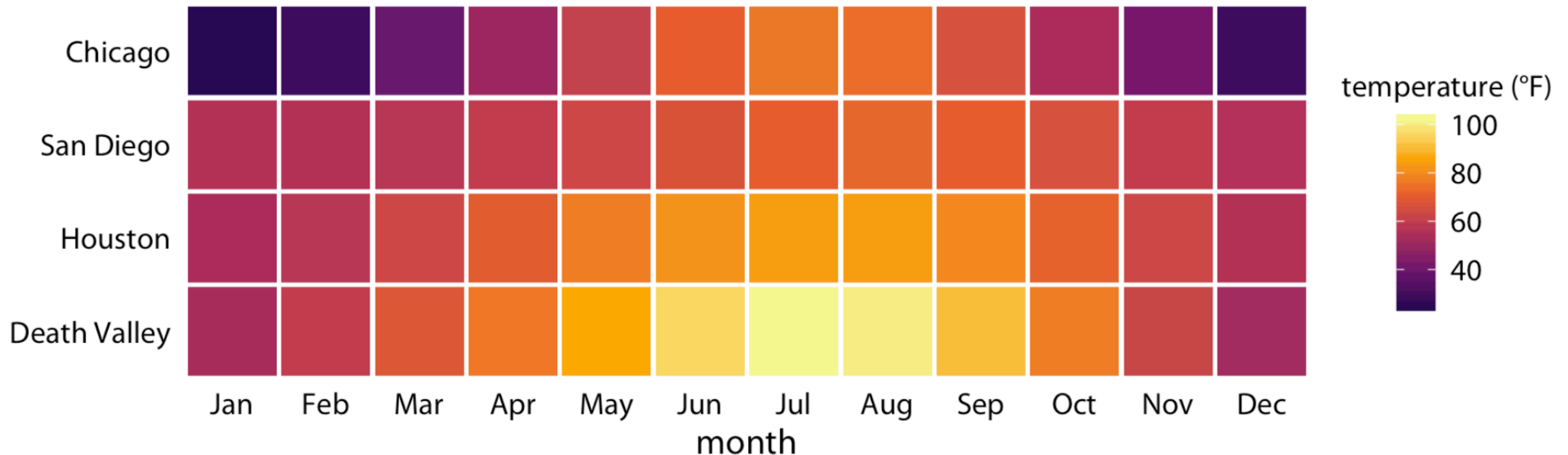
- Data visualizations map data values 1-to-1 onto quantifiable features in the graphic called **aesthetics**.
- Generalization of a “plot axis”.
- These are the visual “levers” we can pull to display one data feature to human observers: How we show someone there is variation along some data dimension.



Scales

- **Scales** define the quantitative mapping *data values* $\longleftrightarrow$ *aesthetics* (i.e. specific implementations for aesthetics).
- Example 1: continuous quantitative variable (x-axis) | continuous quantitative variable (y-axis)
- Example 2: Unordered factor (ind var) | discrete quantitative variable (dep var)
- Example 3: Assign feature where values are bounded $[-1, +1]$ to a color scale (specific data-to-RGB value mapping) (*Heat map*).
- Valid scales objectively map 1-to-1 to data.
- In a scatterplot, if we fix color and size, how many scales can we use to visualize data?
 - 3!
- Aesthetics and scales matter for data science! Plotting libraries like *ggplot2* in R require you to specify both.

Example



Monthly normal mean temperatures for four locations in the US. Data source: NOAA

What are the aesthetics? Position, color

How many scales are there? 3

Does the color scale map 1-to-1 to data? Yes

What are the scales? X-axis position (month), Y-axis position (city), color (temp)

Why are the color squares so large?

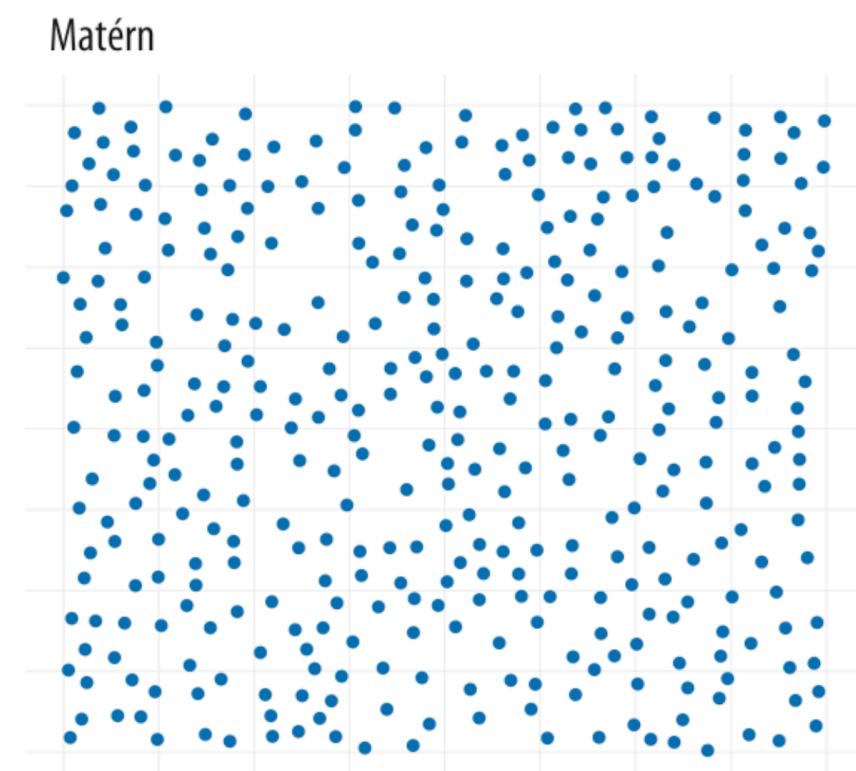
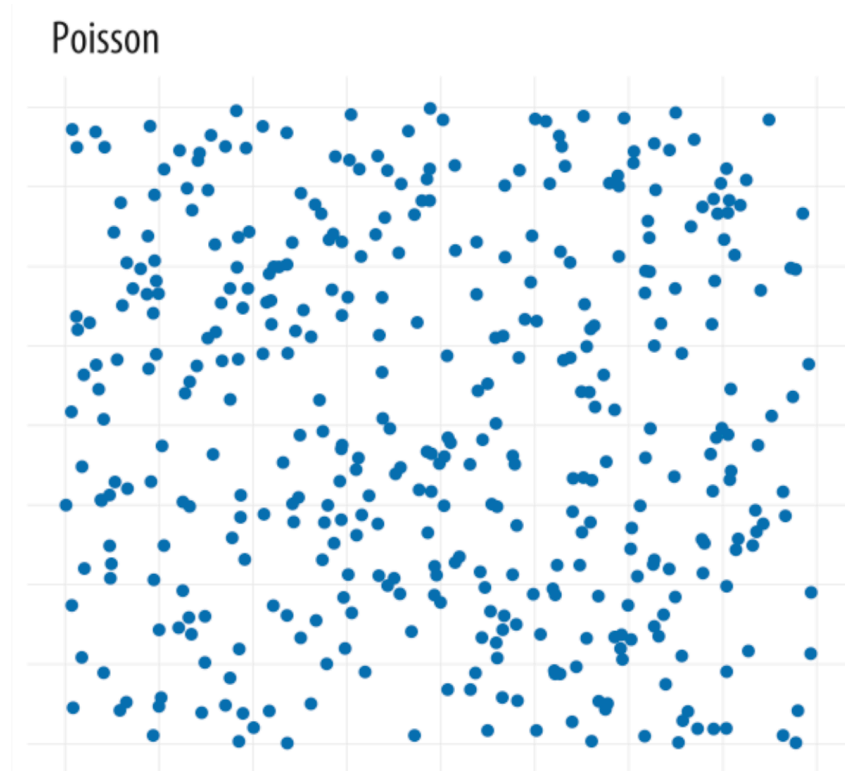
Human color perception degrades when the area becomes too small

(more on this later)

Wilke (2019)

Data visualization can mislead us

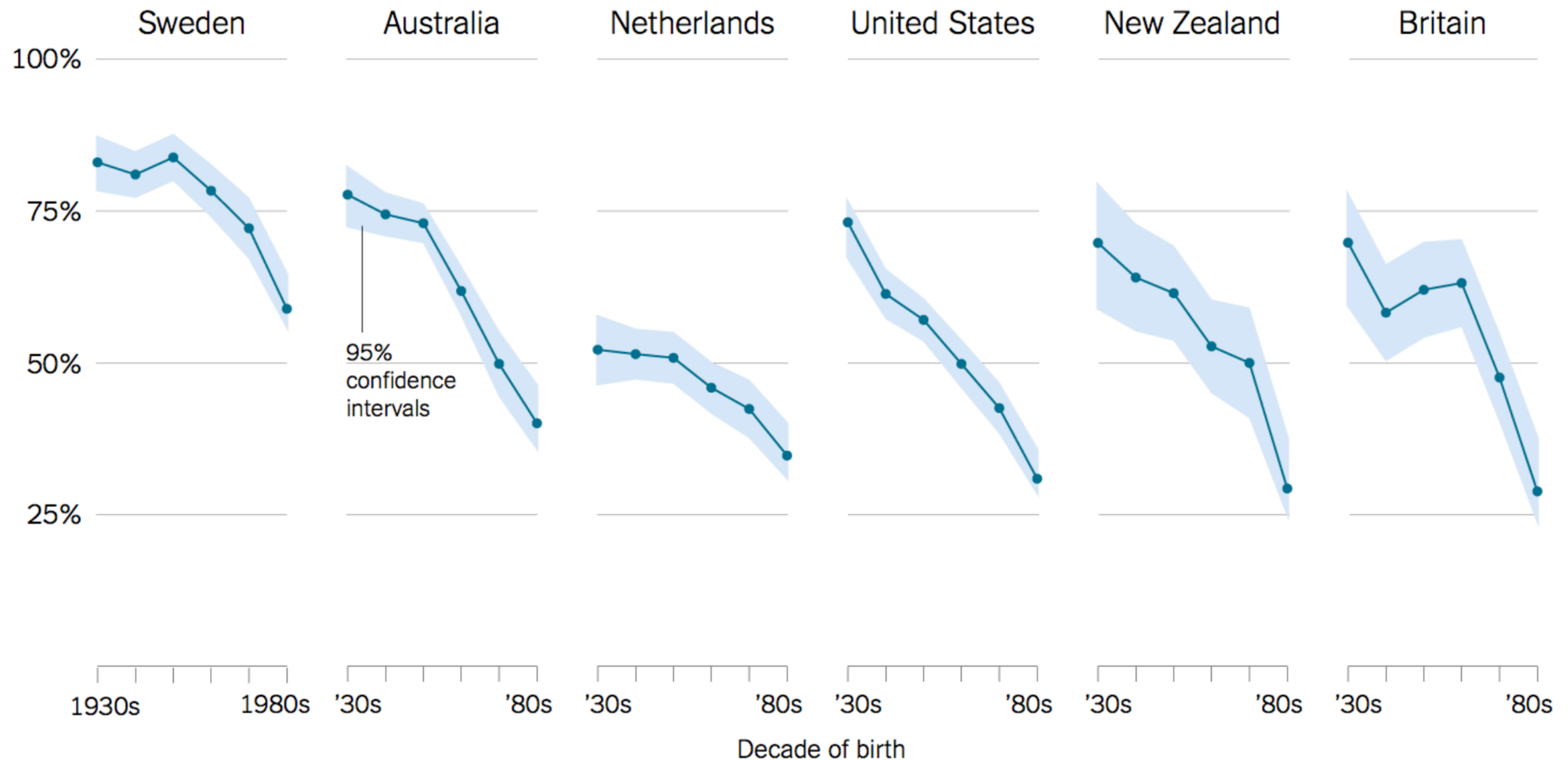
- Ultimately humans interpret visualizations...which can be a problem. Our perception *limits* and *influences* the effectiveness of visualizations.
- We can easily fool ourselves and others when plotting data. Which pattern below is randomly generated?
 - They both are! People see “clusters” in the left, the right looks more “random”. **People see patterns even when there are none!**



1) What are these data showing?

2) What looks potentially misleading?

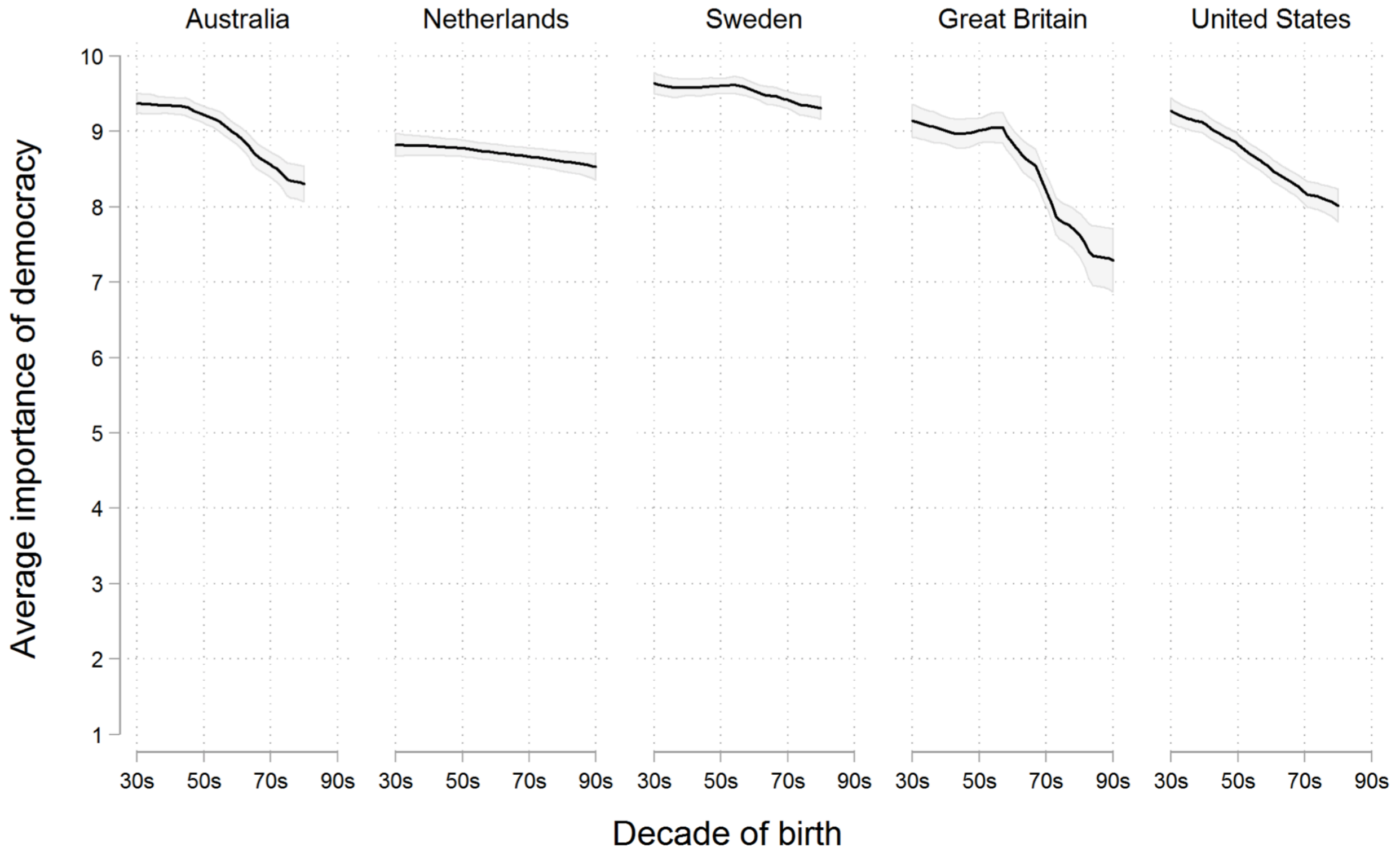
Percentage of people who say it is “essential” to live in a democracy



Issues

1. Connected lines suggest a *trend over time*, but people of different ages (x axis) were polled at the same time!
2. Deception through cherry-picking: Polling question wasn't binary (essential or not), but rating on 1-10 point scale. Author only plotted 10s!

Average democracy data

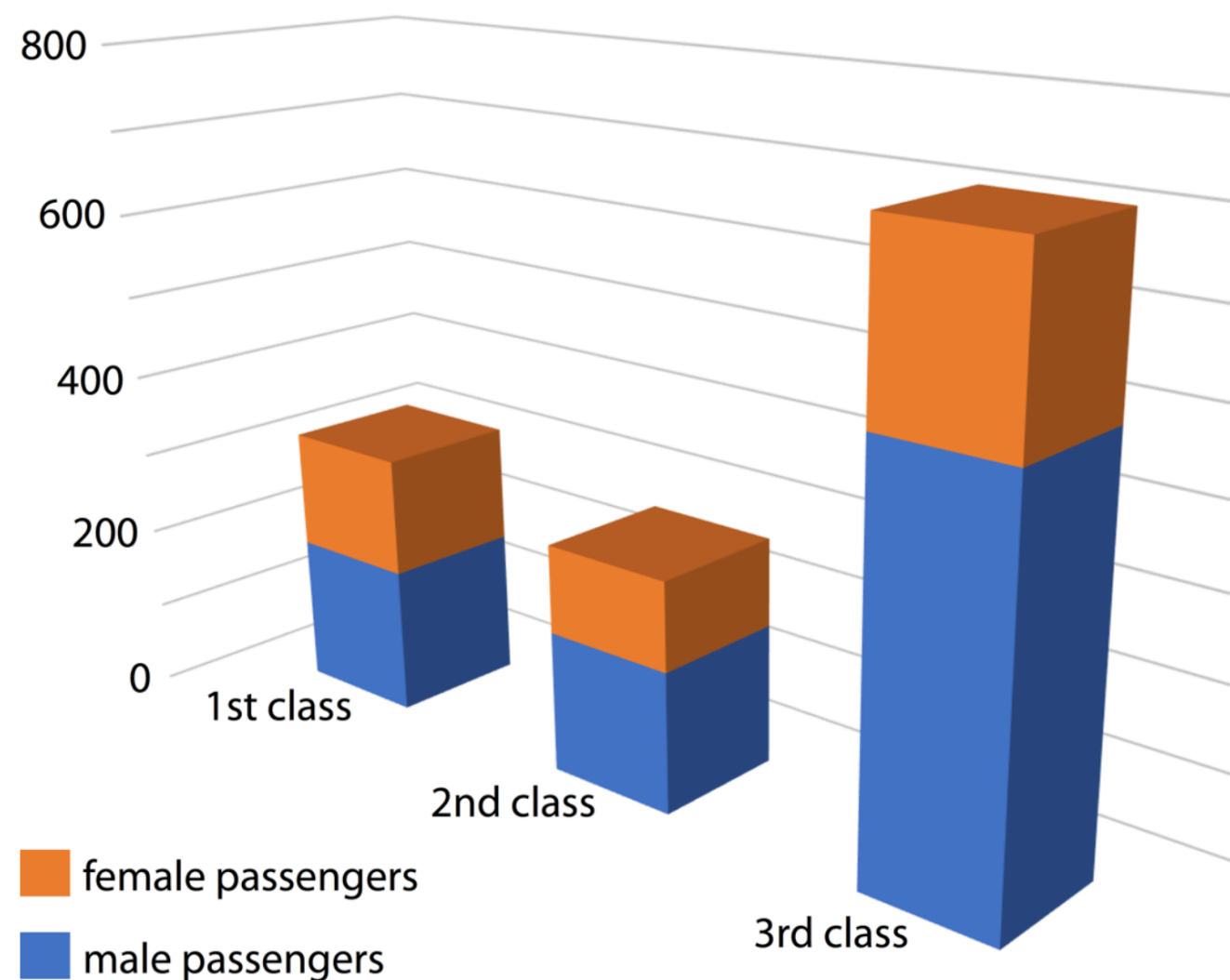


Visualization and human visual perception

3D Plots

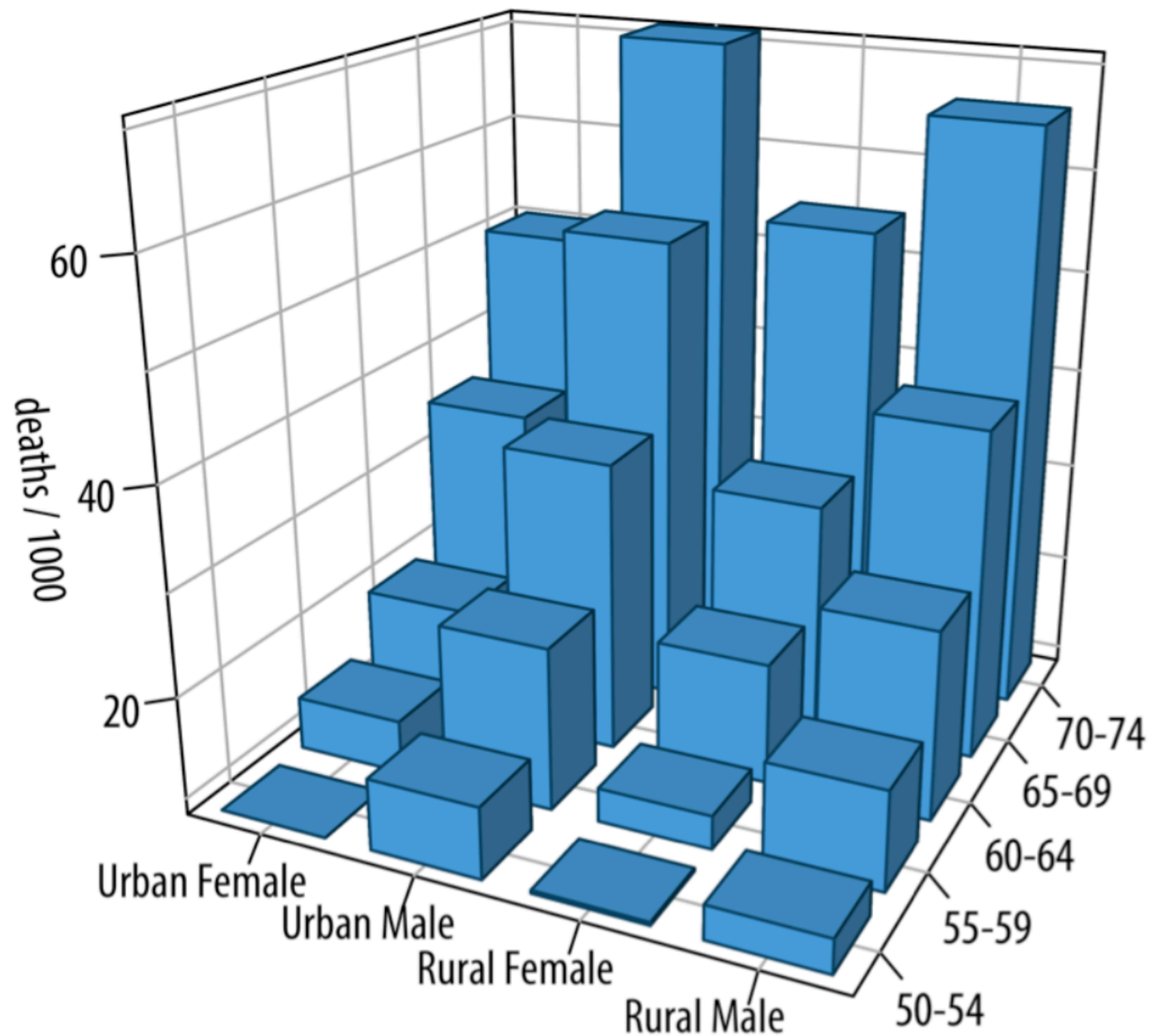
Our perception matters when visualizing and presenting data

What is the value of 1st class? **322!**



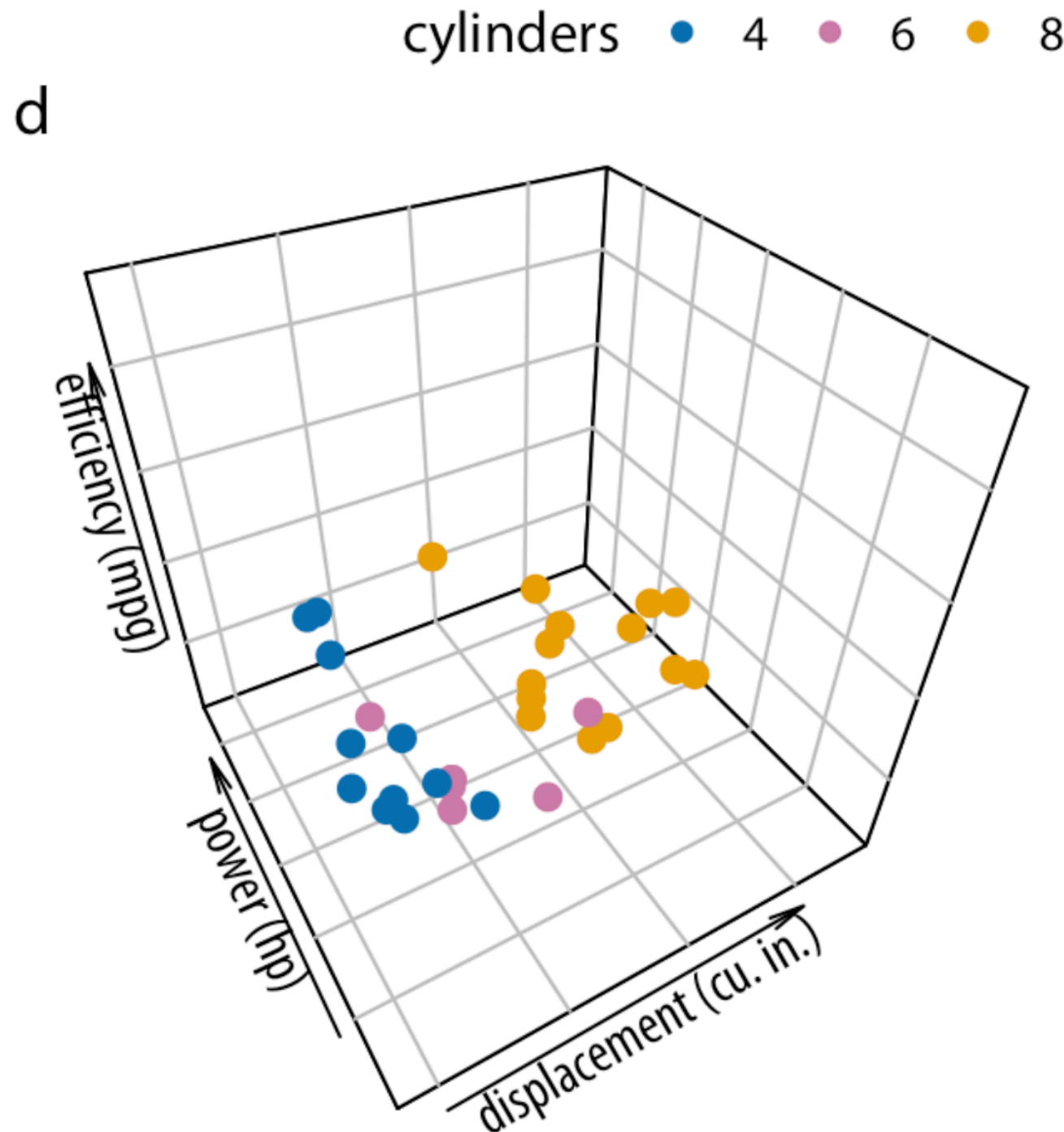
Involves 3 different projections: 3D model space → 2D image screen → brain attempts to invert 2D back to 3D

Yikes!



Don't use static 3D plots

3D perception in a 2D image critically depends on reference information and perspective.



Don't use static 3D plots

3D perception in a 2D image critically depends on reference information and perspective.

d
cylinders ● 4 ● 6 ● 8

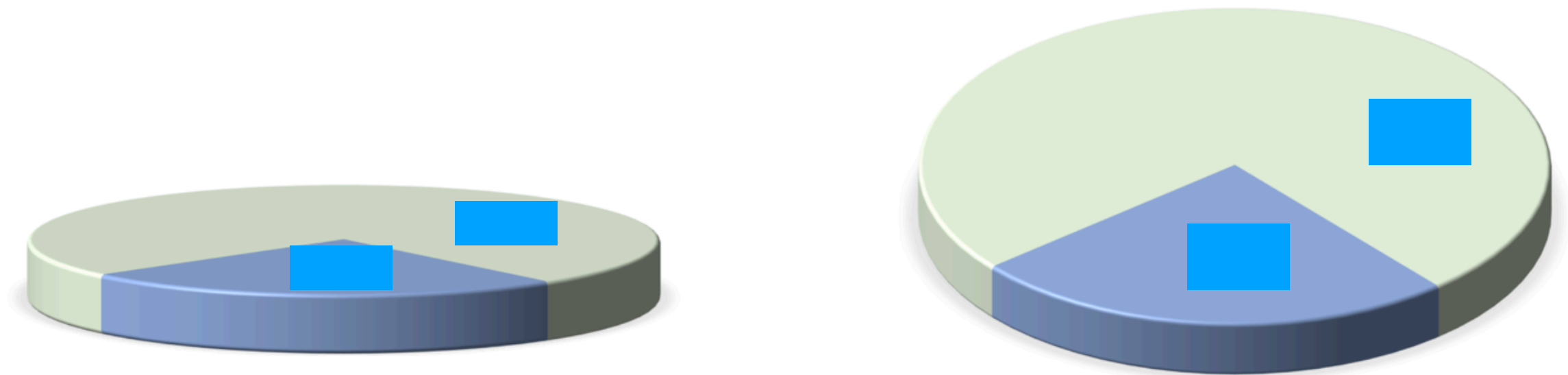


Remove/play with reference frame, consistent interpretation becomes impossible!

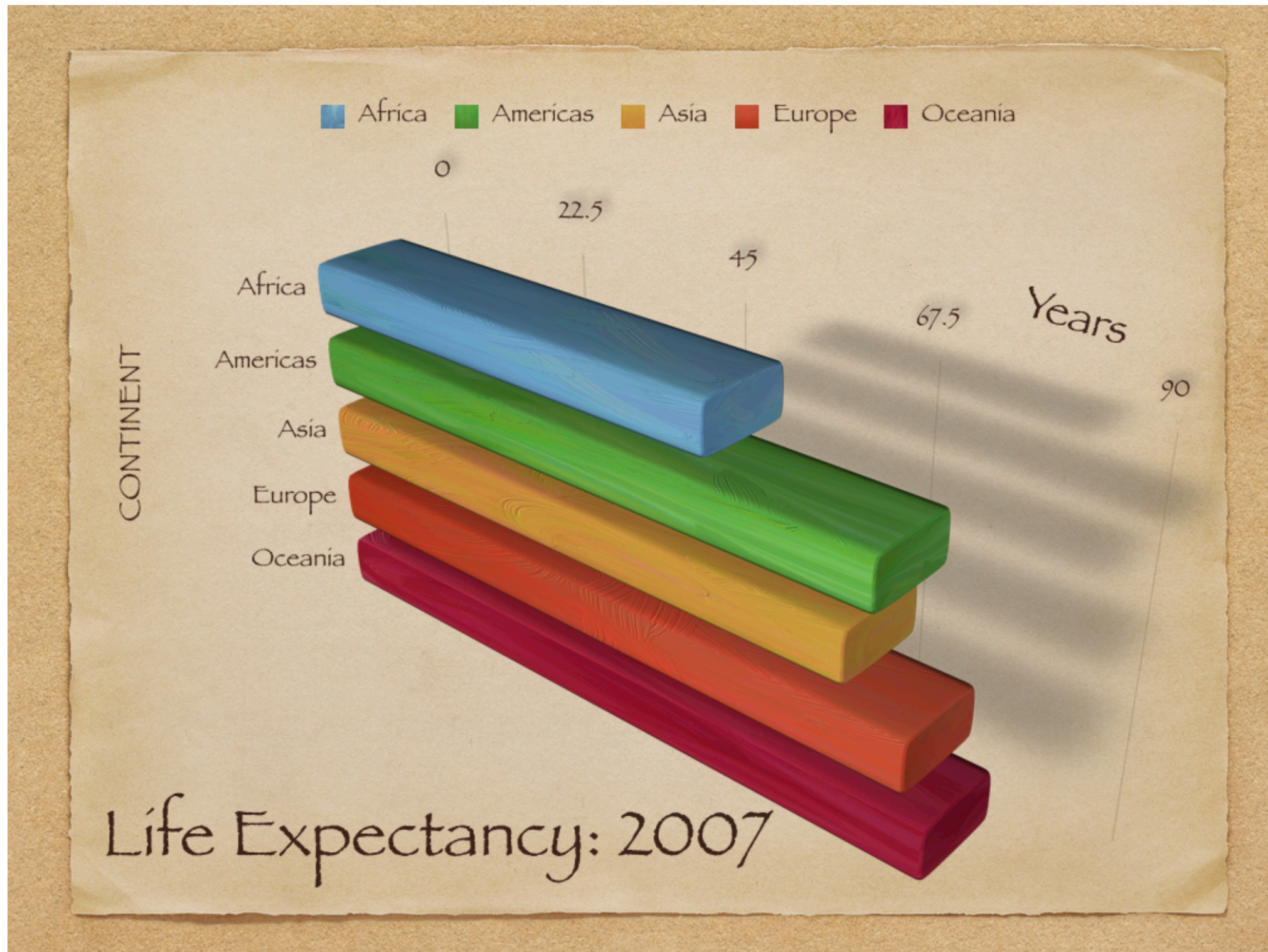
Wilke (2019)

Don't use static 3D plots

Which blue-ish slice is bigger?

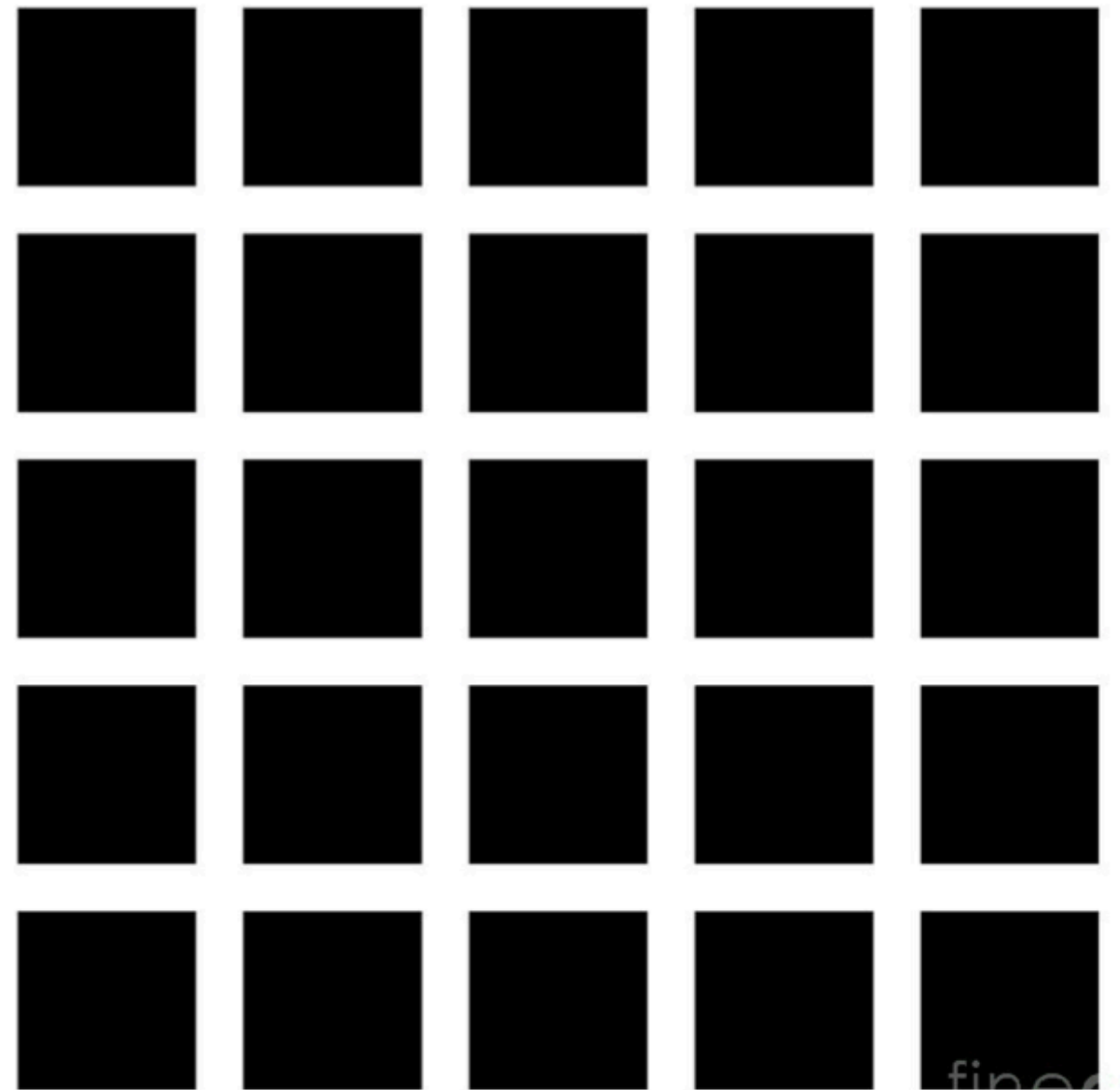


What factors hinder data interpretation?



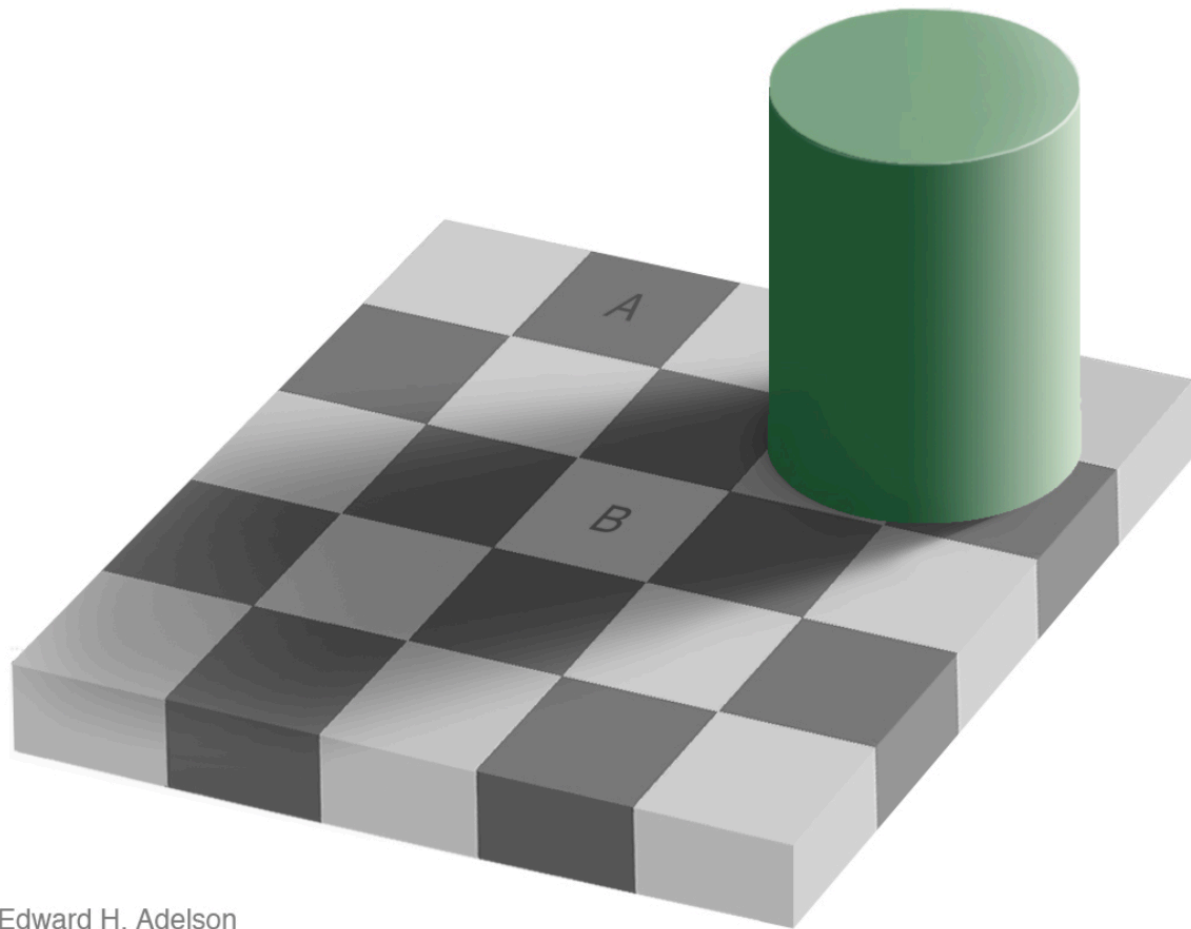
We don't see objective reality

- Perception is NOT about the brain neatly “piping” inputs from eye.
- Structure affects function.
- Our brain makes it easier to perceive certain visualizations better than others.
- Visual illusions clue us in about “artifacts” in the brain’s processing mechanisms.

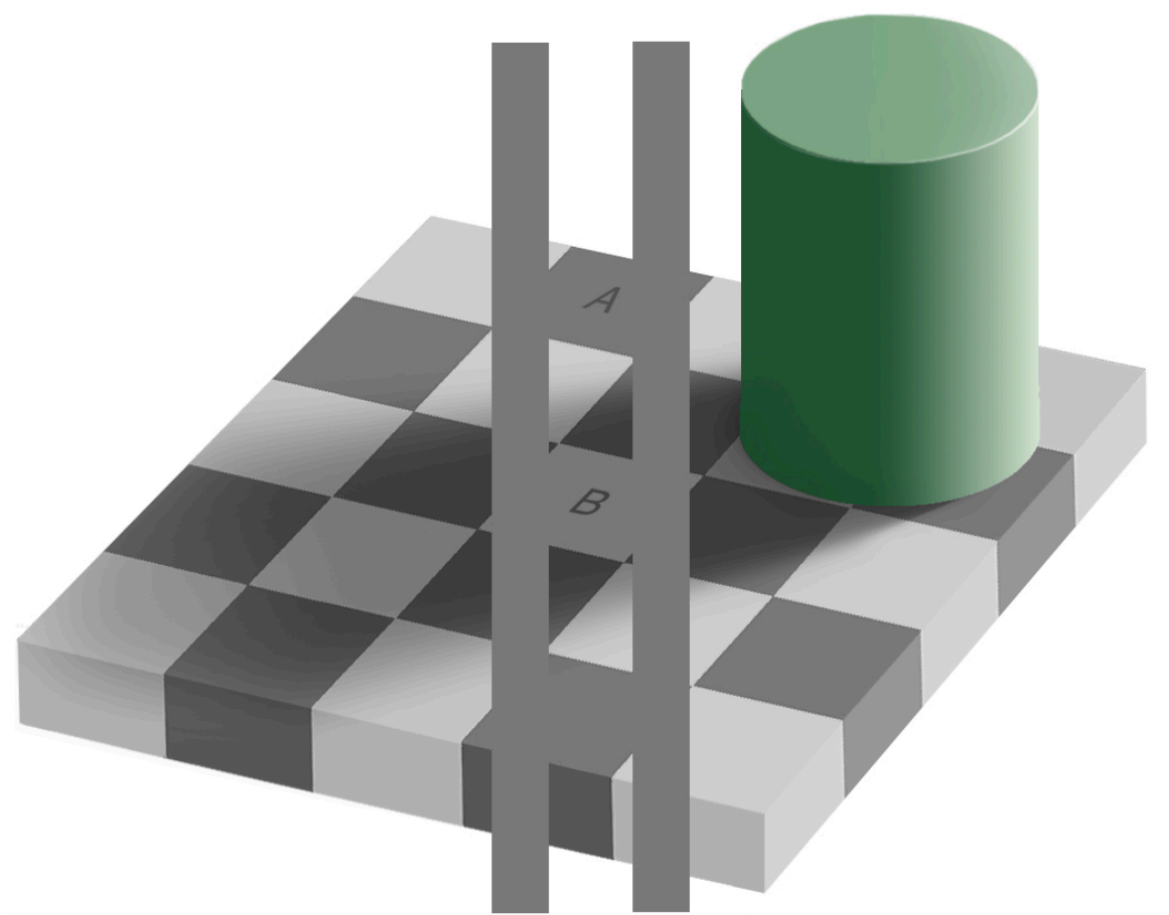


Hermann Grid

Checker-shadow Illusion

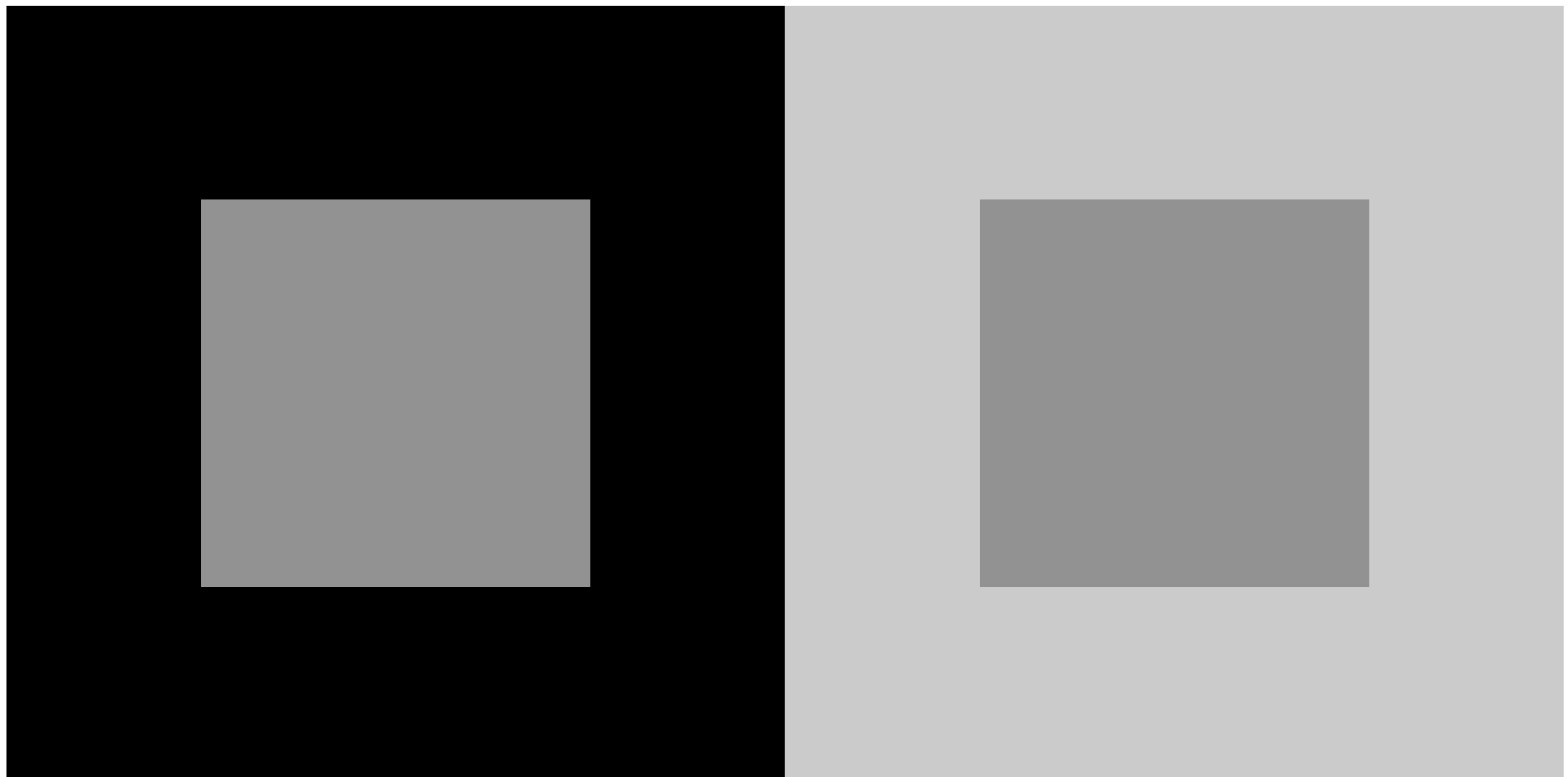


Edward H. Adelson

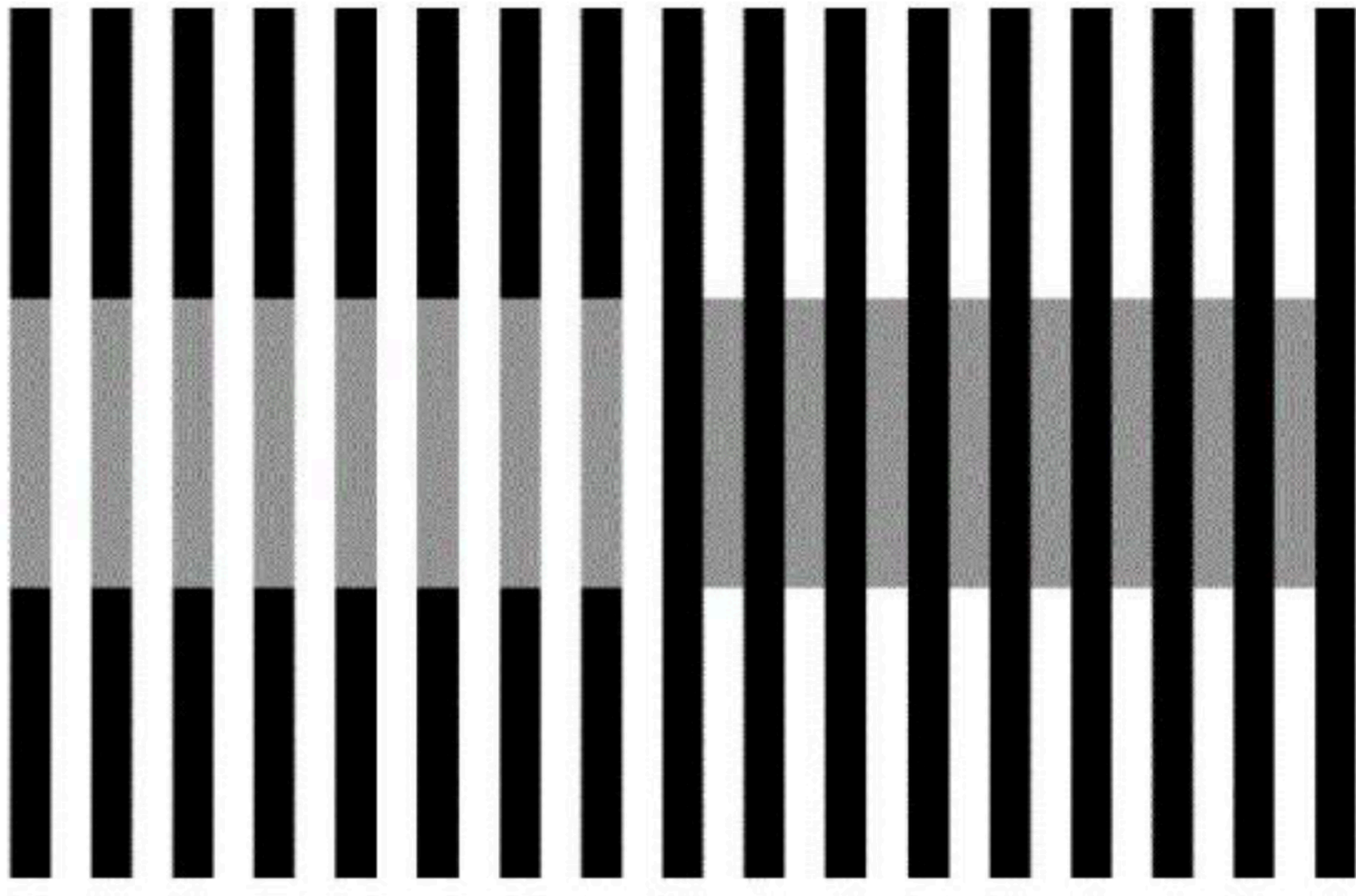


A = B!

Brightness contrast



Simultaneous Contrast



Perception matters when designing color bars



Mach bands

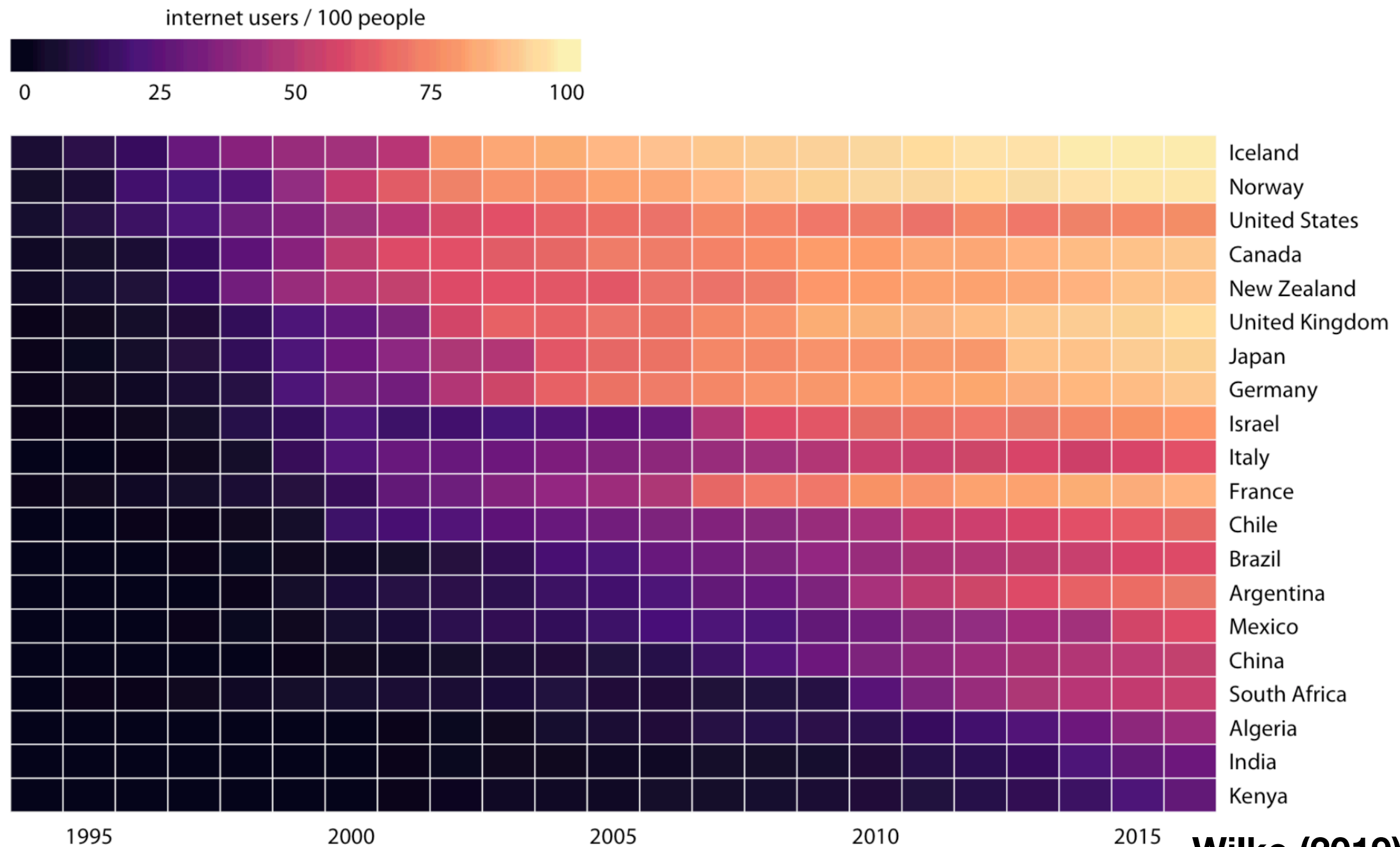
Gray bars that share a boundary perceptually increase their contrast.

You should also see narrow bright “strips” at the boundaries. *These are illusory!*

Happens because visual system actively attempts to “normalize” the gray values to make perception easier.



Separate adjacent color values when possible



Perceptual uniformity

- Whether your color scale data is continuous or discrete, you want equidistant data points to appear equally different in color (**perceptual uniformity**).
- 0.0 - 0.1 should look like 0.9 - 1.0!
- Popular rainbow scale is NOT perceptually uniform!
- ColorBrewer scales tend to do a good job.

rainbow scale

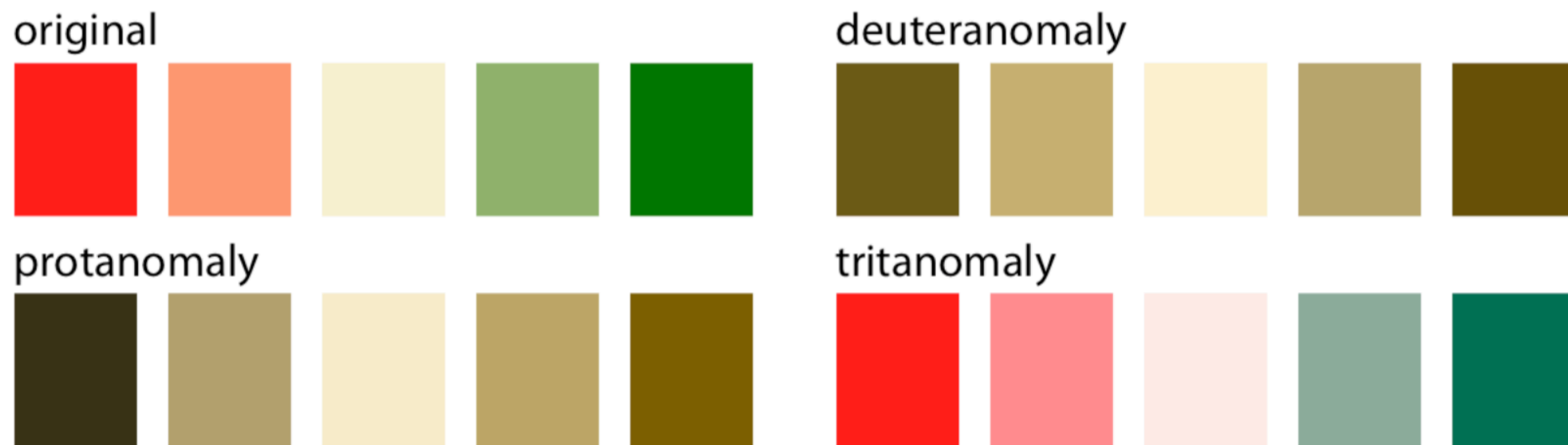


rainbow converted to grayscale



Colorblindness

- 8% of men (including myself!) and 0.5% of women are colorblind. Problematic when designing color scales.
- Most common: red-green (deuteranopia and protanopia). Also blue-yellow (tritanopia).
- Most people find that colors between green/red exhibit high contrast... I do not!
- Example: Qualitative categorical color scales:



Use colorblind compatible palettes

- Be careful when selecting color scales in ad hoc ways!
- Fortunately, there are many well-studied solutions.
- Example: Okabe & Ito (2008) color palette. Works for all colorblind types. Here are the RGB hex values:



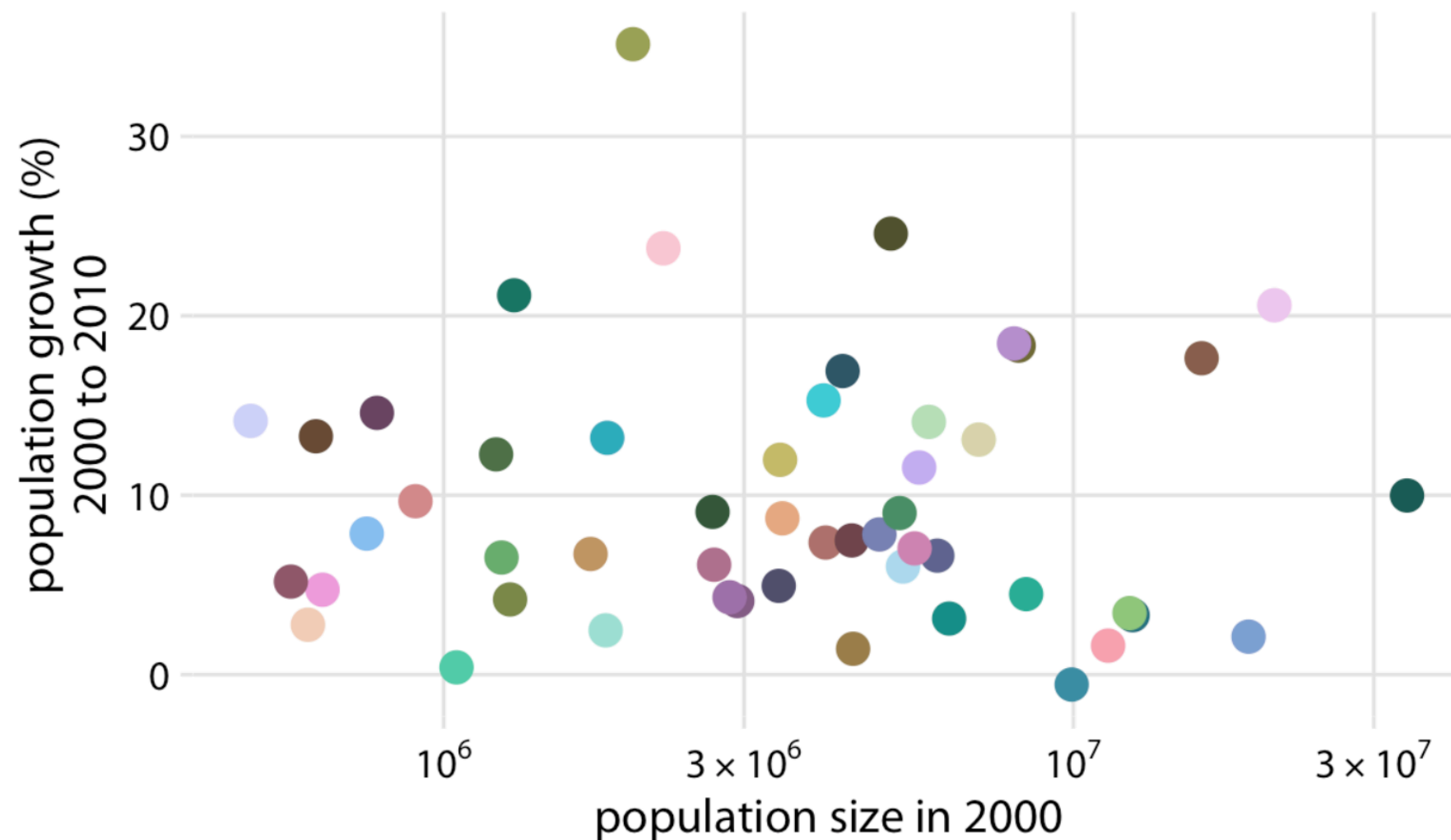
- Another good choice: ColorBrewer PiYG (Pink to yellow-green):



Visualization and human visual perception

Misperceiving scales

Categorical color scales most effective with no more than ~3-5 levels



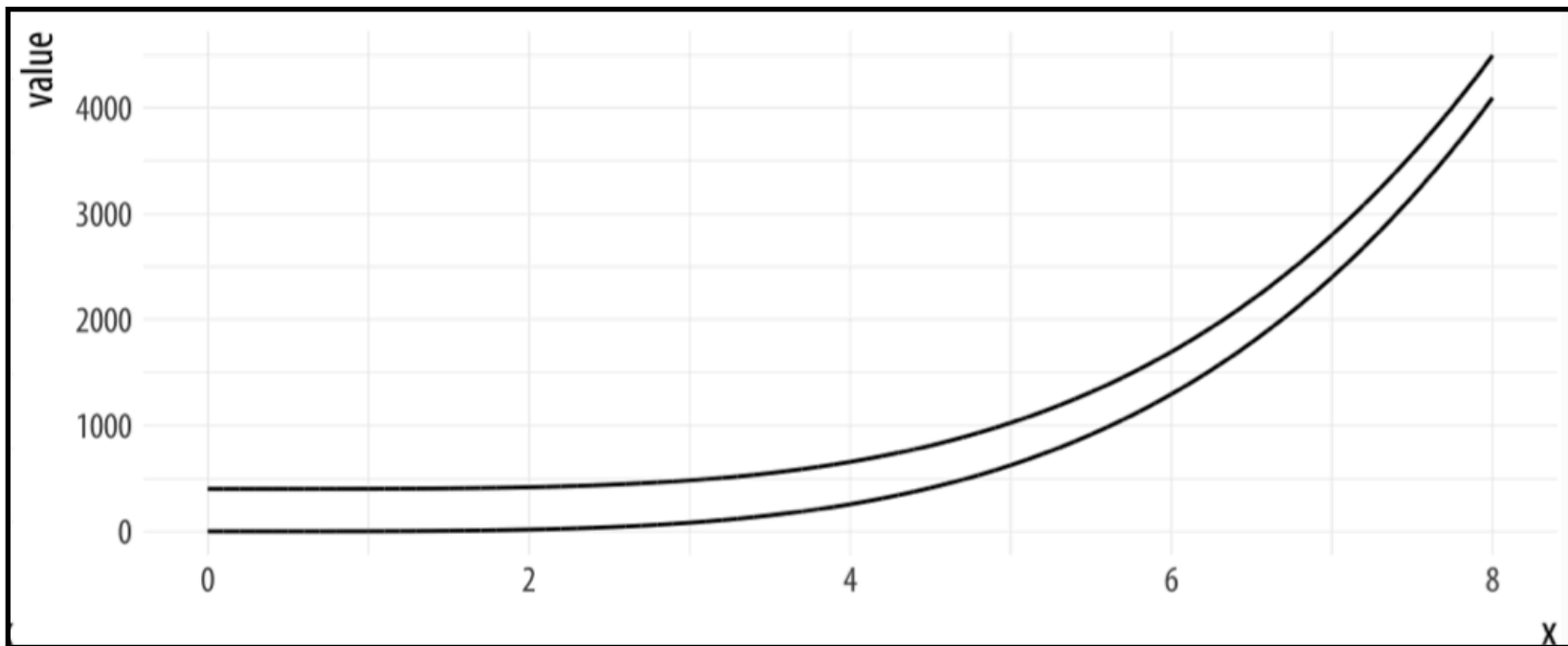
state

Alabama
Alaska
Arizona
Arkansas
California
Colorado

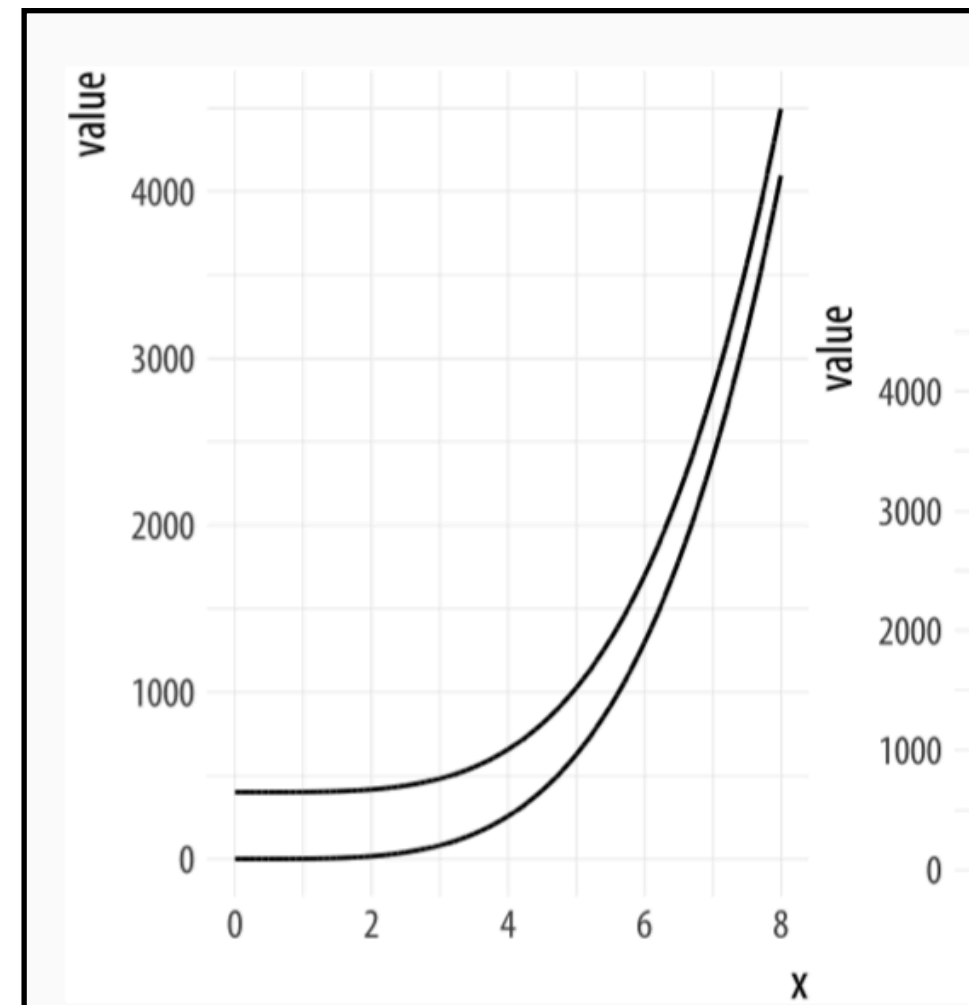
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan

North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island

Aspect Ratio



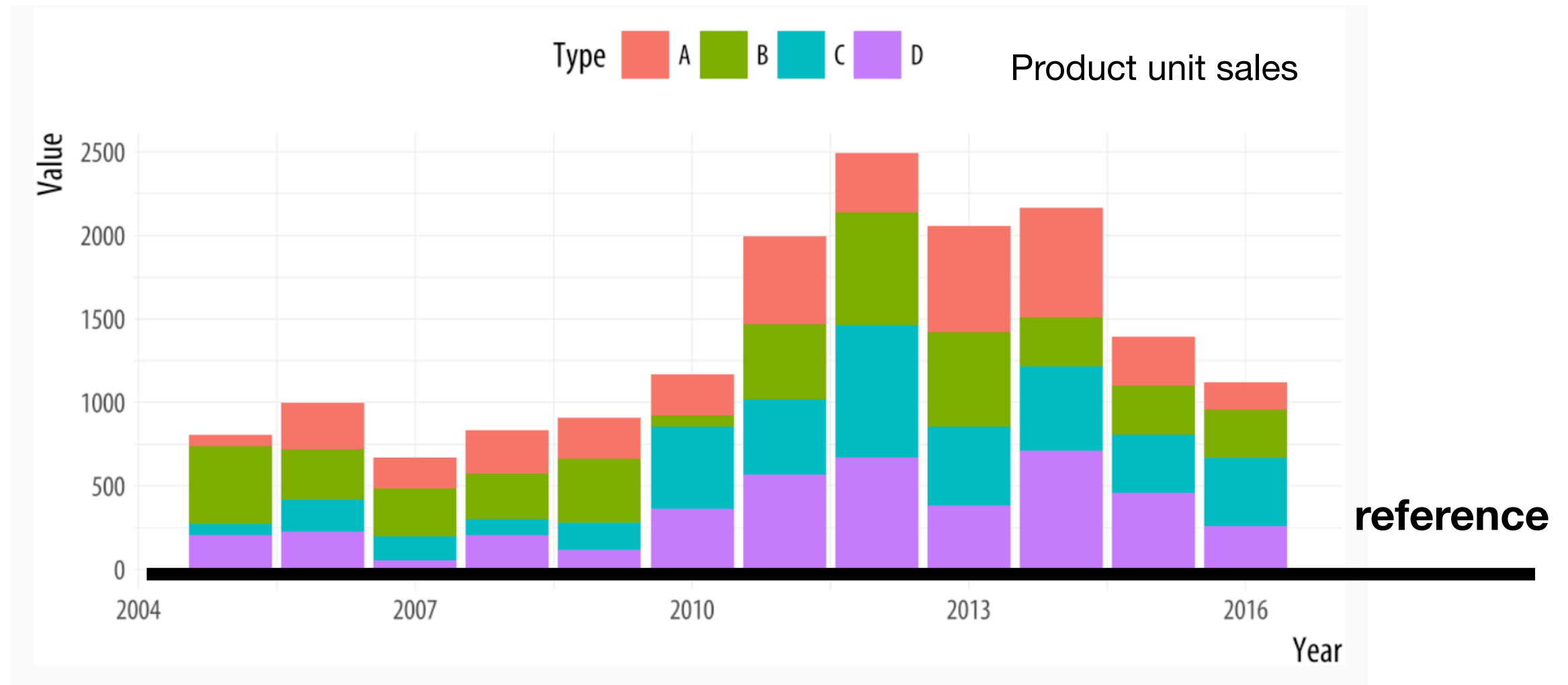
Small aspect ratio



Aspect ratio (unit-less): figure width (px) / height (px). Often expressed as a fraction, where the usual / is replaced with a colon. Example: 16:9

Different than **scaling**: Multiplying data by a integer or floating point number to change units (e.g. lbs to kg).

Human viewers struggle when there isn't a common reference



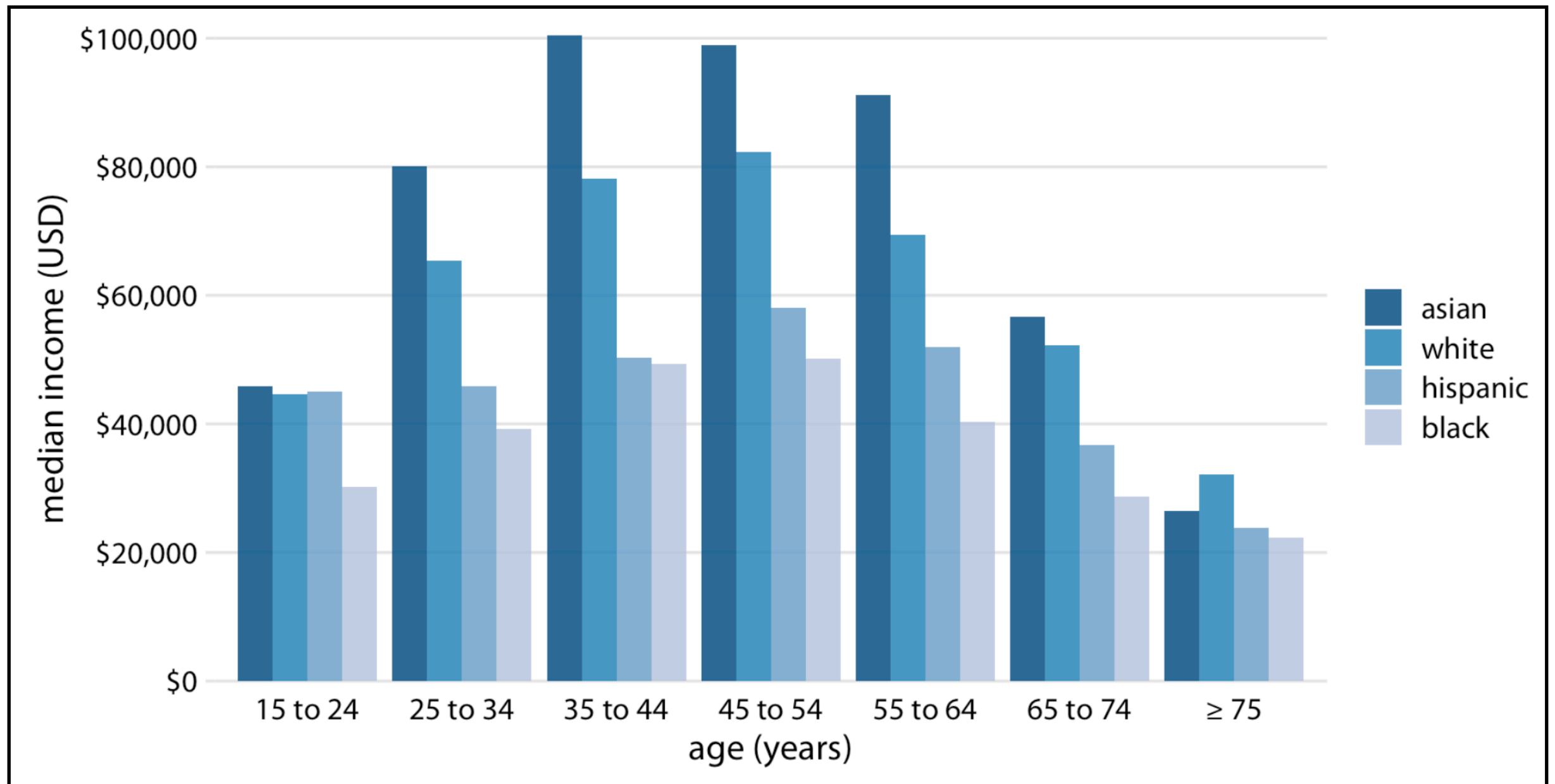
Are blue sales units larger in 2016 compared to 2013?

Are gray sales units larger in 2016 compared to 2013? 2011?

**Humans more readily perceive relative quantities
when they share a common baseline (reference)**

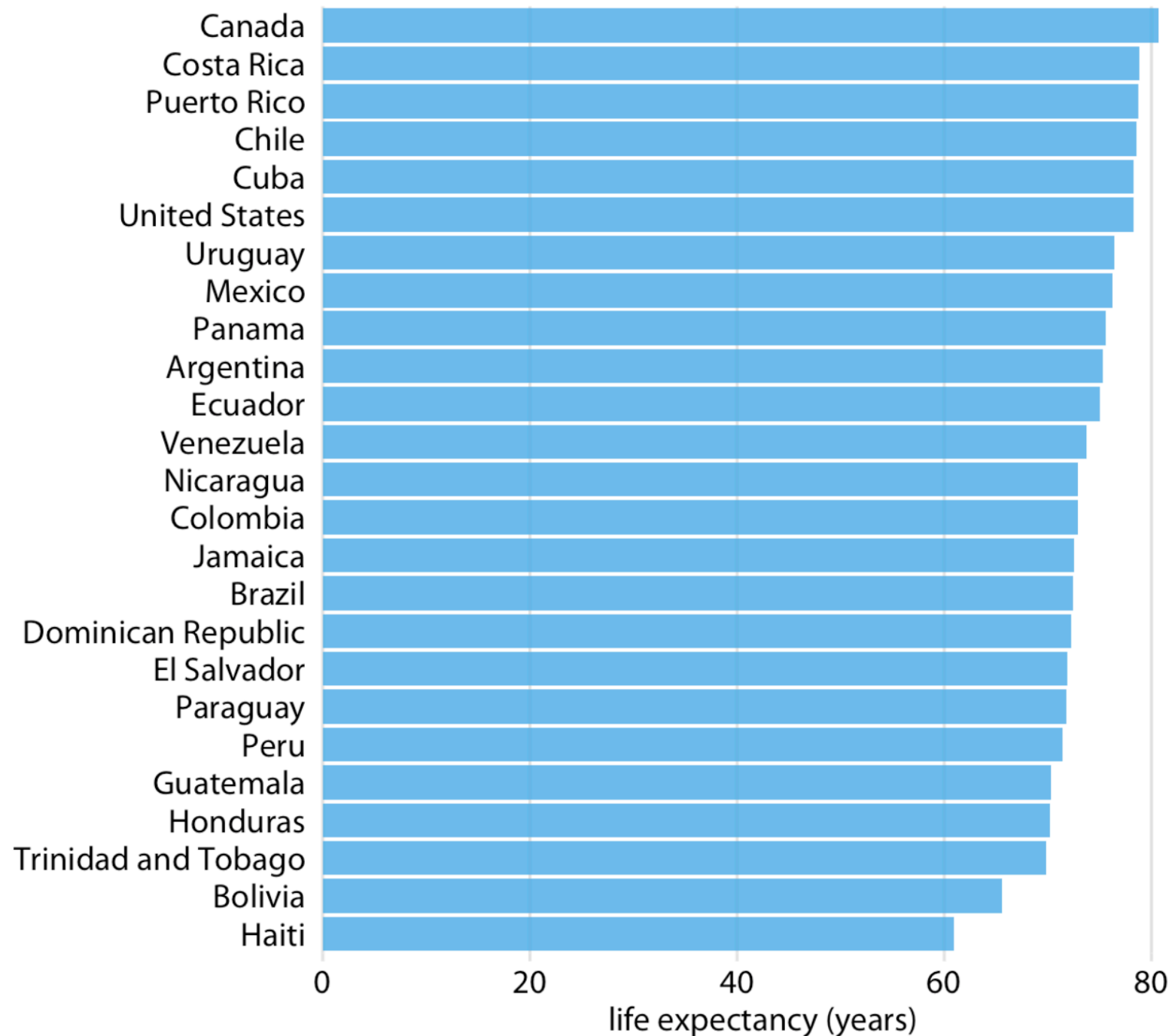
Healy (2018)

Fix: Side-by-Side bar plot



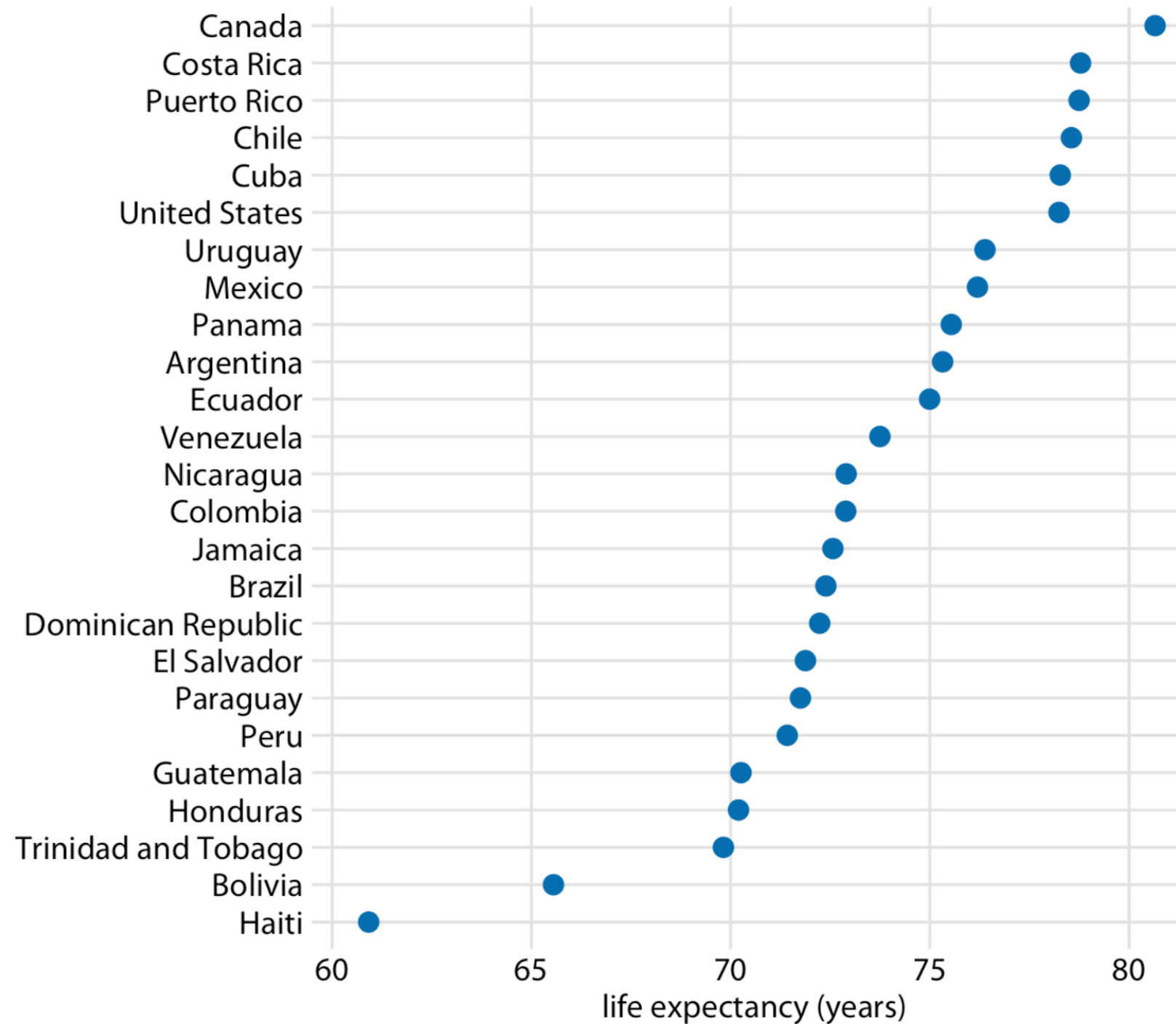
Effective data visualizations should communicate a message

Problem: Perceptual length compression



Source: Gapminder project. Wilke (2019)

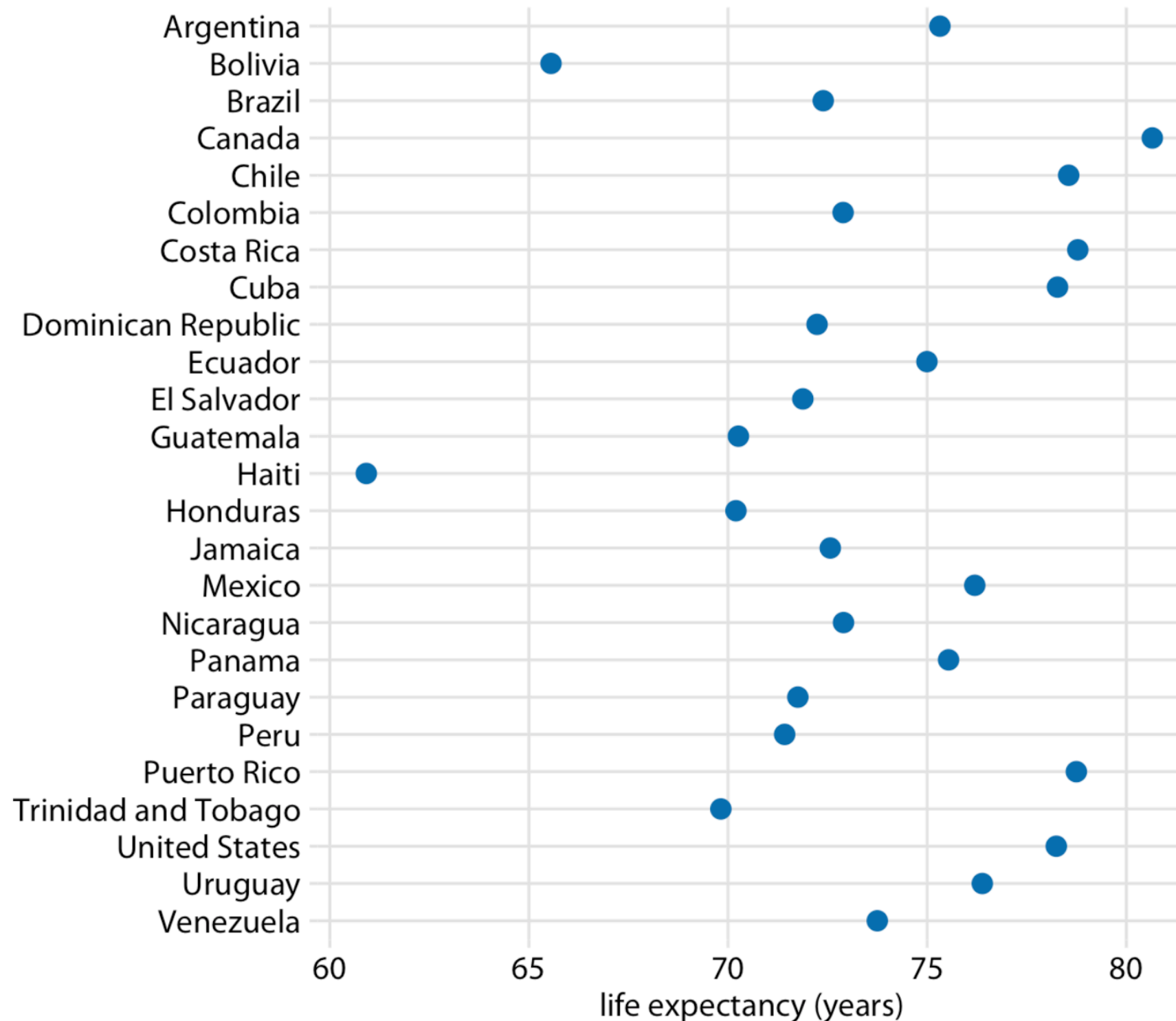
Fix: Dot plot



Also note the change in x scale.

Source: Gapminder project. Wilke (2019)

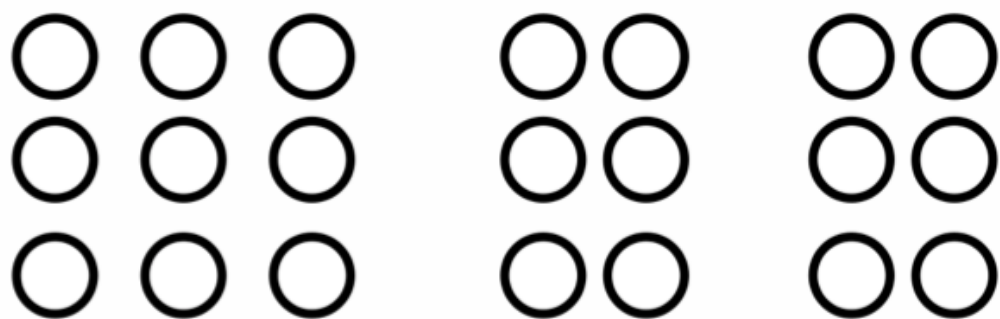
Is this version effective?



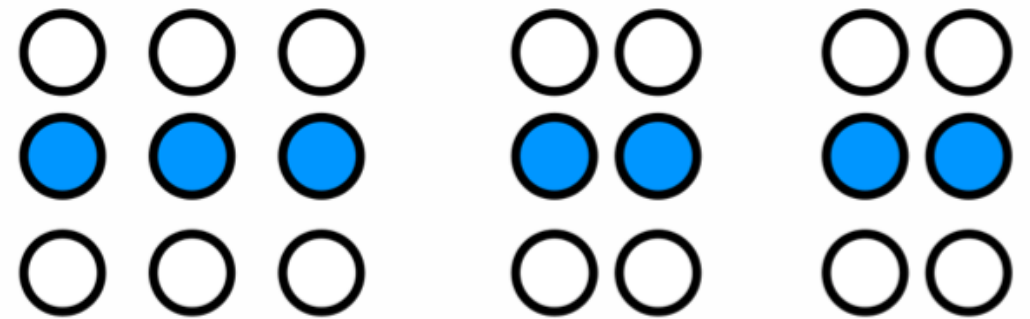
Source: Gapminder project. Wilke (2019)

Gestalt Rules

- **Gestalt Rules:** Competing factors that influence how we **group** collections of visual elements.
- Gestalt means “whole” in German. (*Gestalt rules emerged from psychophysical experiments performed by mostly German psychologists in the early 1900s*).



Proximity



Similarity

Next week

Designing a GUI in Python that
we will use to plot data.