W03D1

Data Visualization

Instructor: Eric Elmoznino

Adapted from material in Fundamentals of Data Visualization

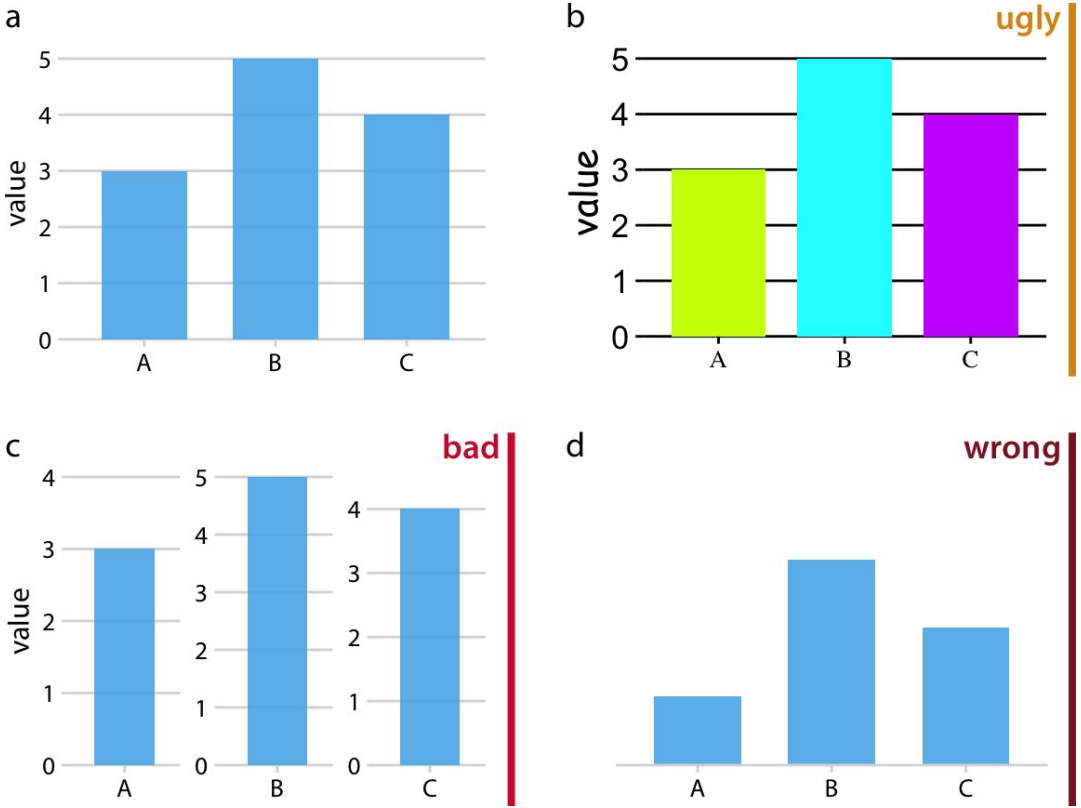
Outline for today

- Motivation
- Types of visualizations
- Design principles and graph critique
- Break (10 mins)
- Matplotlib and Seaborn demo

Motivation: why not just numbers?

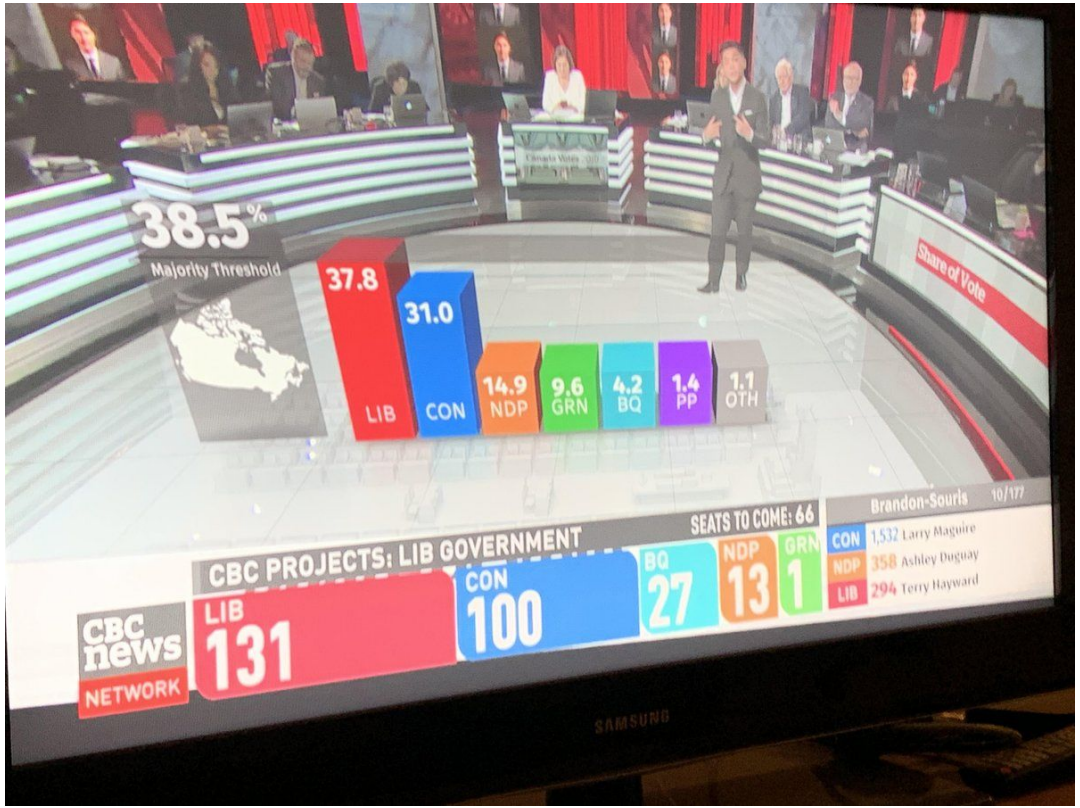
- Humans have evolved to have excellent vision
  - Size and proportion
  - Space and spatial relationships
  - Colour
  - Approximate quantity
- Humans ***have not*** evolved to read numbers or tables
- In scientific publications and presentations, most people only look at figures

Motivation: what can go wrong?



- **ugly**: A figure that has aesthetic problems but otherwise is clear and informative.
- **bad**: A figure that has problems related to perception; it may be unclear, confusing, overly complicated, or deceiving.
- **wrong**: A figure that has problems related to mathematics; it is objectively incorrect.

Motivation: what can go wrong?



Types of visualizations

- **Visualizing amounts**: bar plots, grouped/stacked bars, dot plots, heatmaps
- **Visualizing distributions**: histograms, density plots, boxplots
- **Visualizing proportions**: bars, mosaic plots, tree maps
- **Visualizing x-y relationships**: scatterplot, bubble chart
- **Visualizing geospatial data**: choropleth, cartogram
- **Multi-panel figures**
- Many more: https://clauswilke.com/dataviz/directory-of-visualizations.html

Types of visualizations: activity (15 mins)

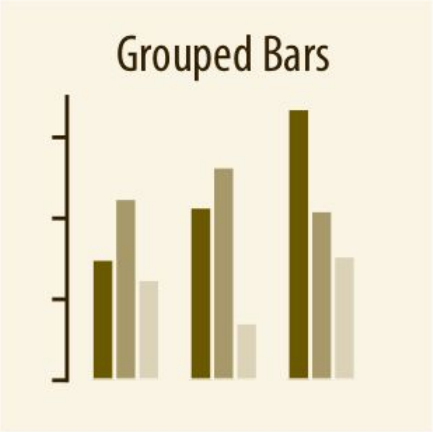
- 4 groups: amounts, distributions, proportions, x-y relationships
- Pick **1 simple** chart and **1 complex** chart and show us:
  - Example use-case
  - What would be the x/y axis
  - What other encodings (color, size) are used / can be used
  - Can they be made in Python? Find me some code samples
- Present (no need to prep slides, walk through it while on the website)
- Zoom breakout groups: I'll activate and assign topics via Slack
- Directory of Visualizations: https://serialmentor.com/dataviz/directory-of-visualizations.html
- Don’t be afraid to Google the chart types to understand them better

Types of visualizations: amounts

Simple
Population of countries

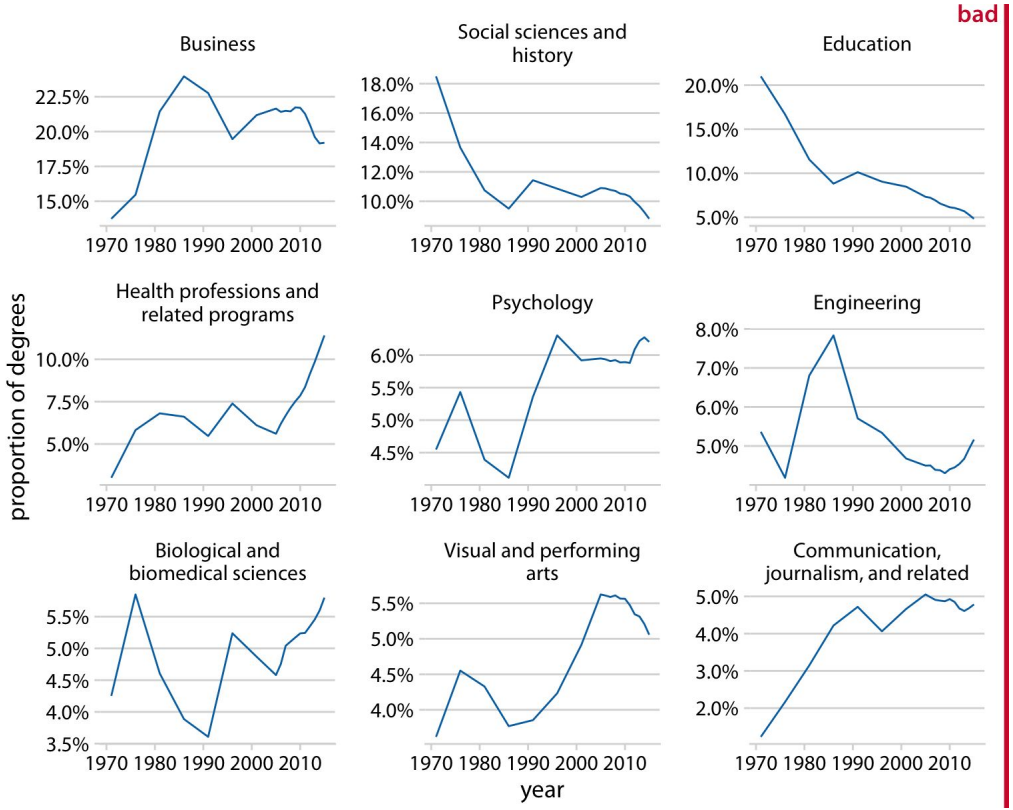
x-axis: country
y-axis: population
color: emphasis on specific country

Complex
Sales by business unit by continent

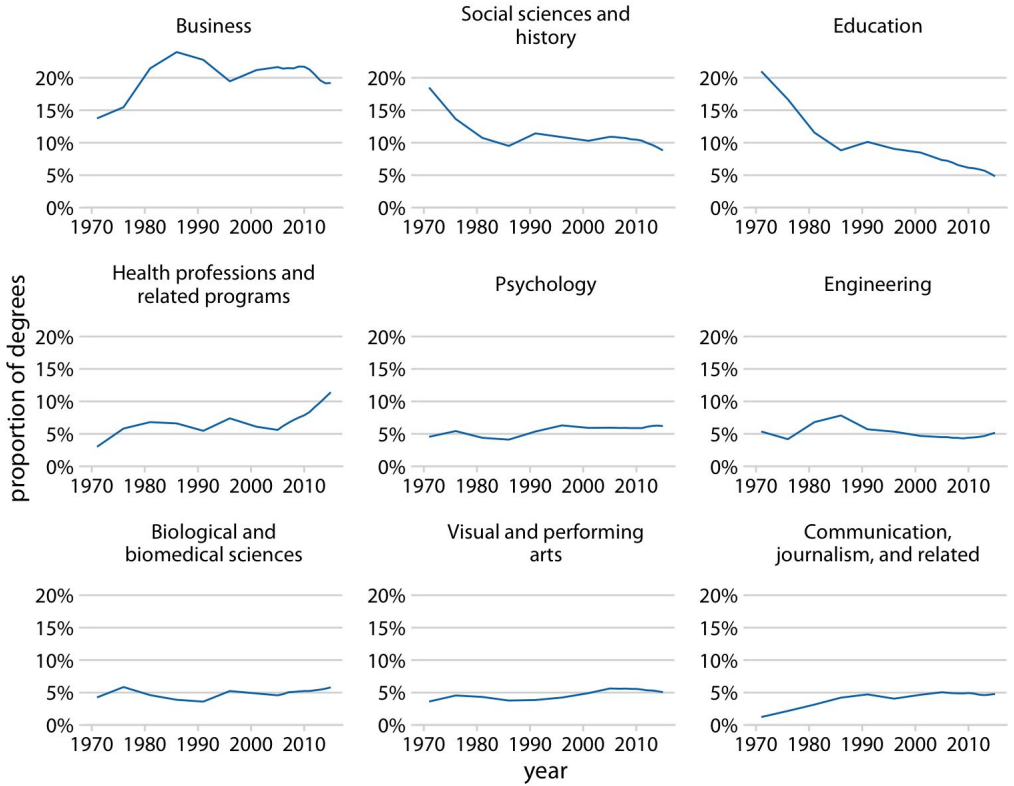


x-axis: continent
y-axis: sales in USD
color: business unit

Types of visualizations: multi-panel figures



Types of visualizations: multi-panel figures



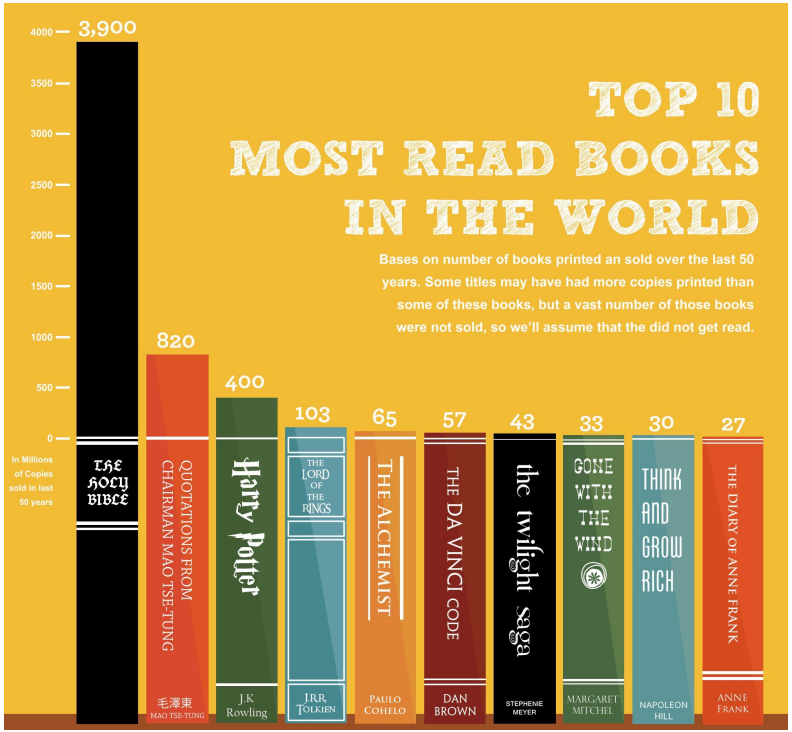
Design principles

1. Principle of proportional ink
2. Picking colours that have meaning
3. Use encodings to your advantage
4. Label clearly

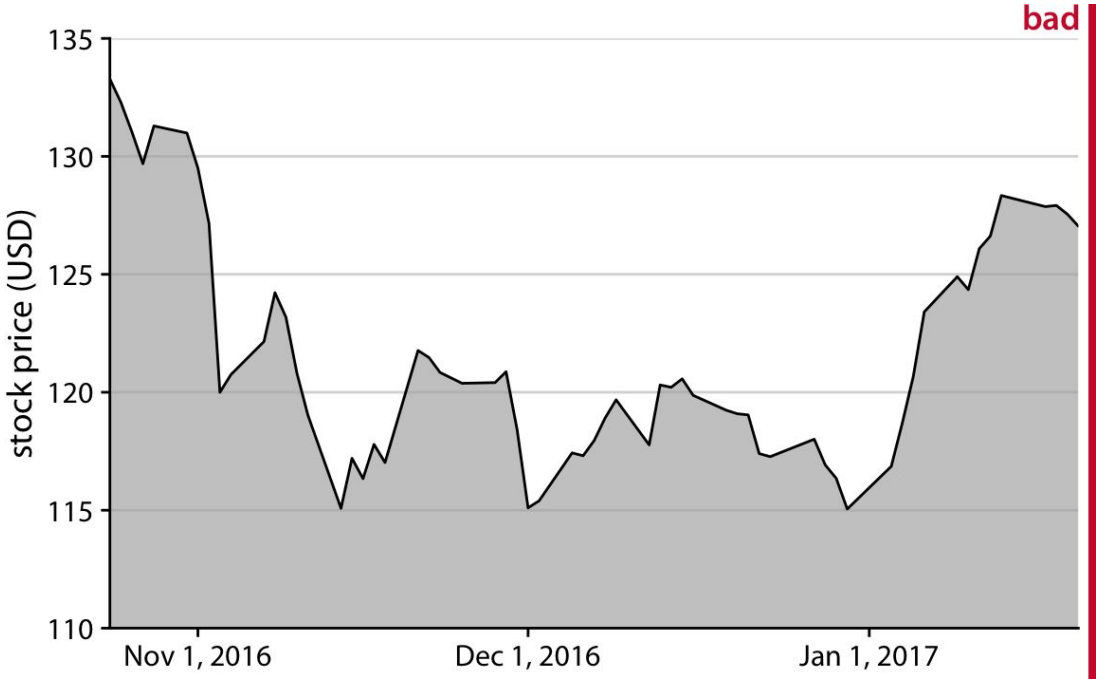
Design principles
1. The principle of proportional ink

The principle of proportional ink: The sizes of shaded areas in a visualization need to be proportional to the data values they represent.

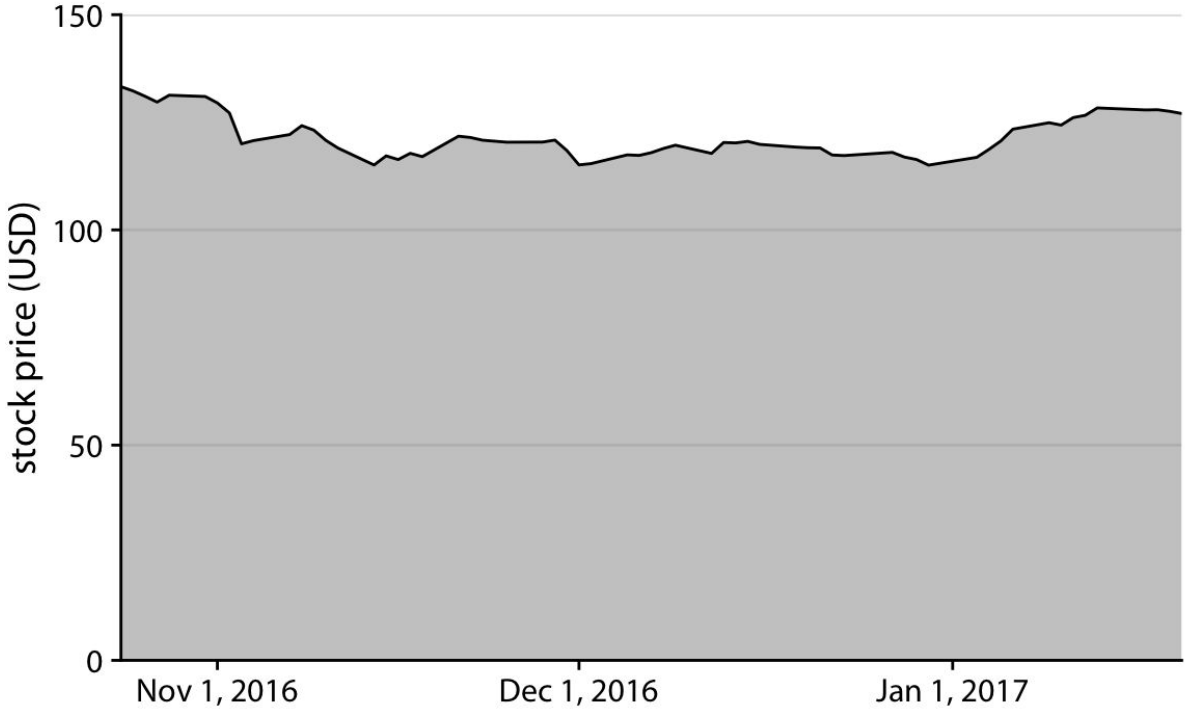
Design principles
1. The principle of proportional ink



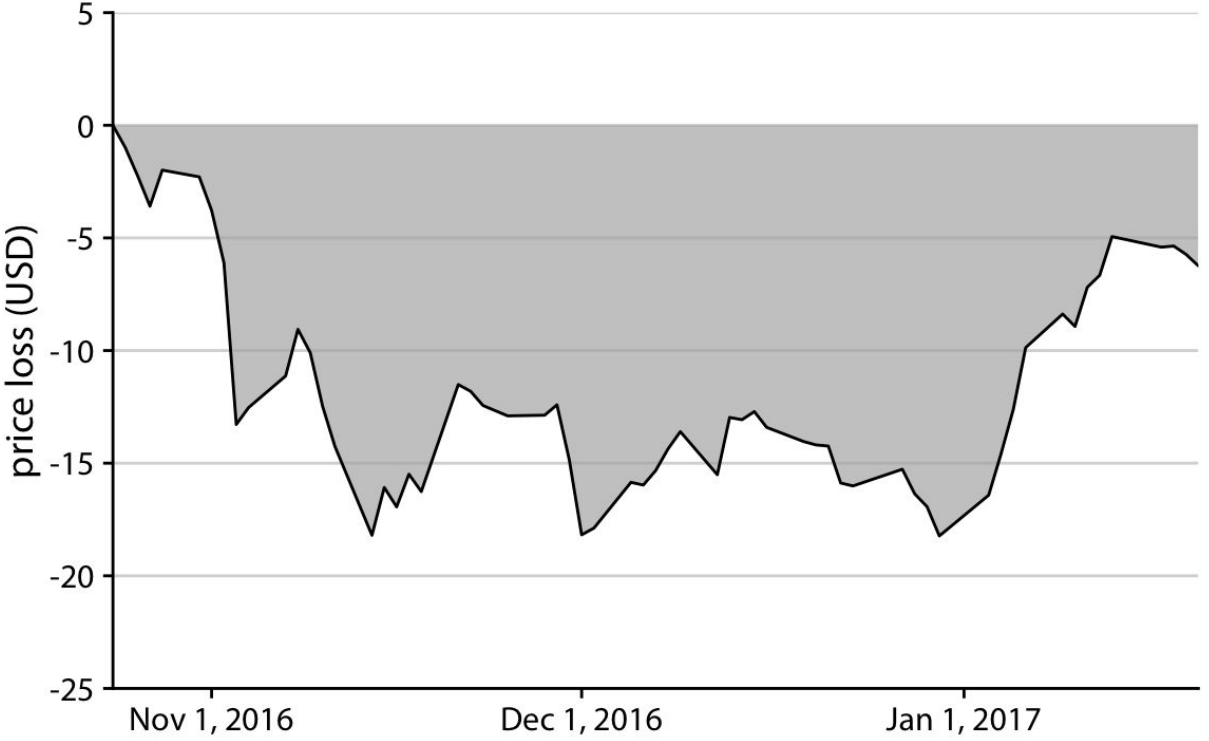
Design principles
1. The principle of proportional ink



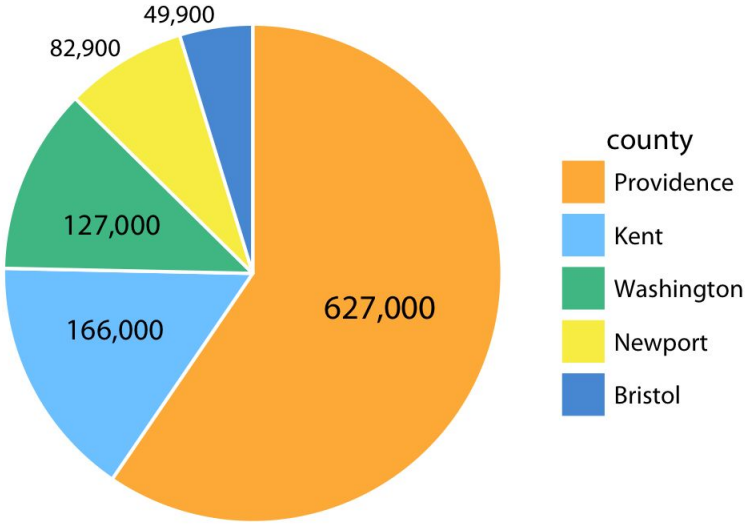
Design principles
1. The principle of proportional ink

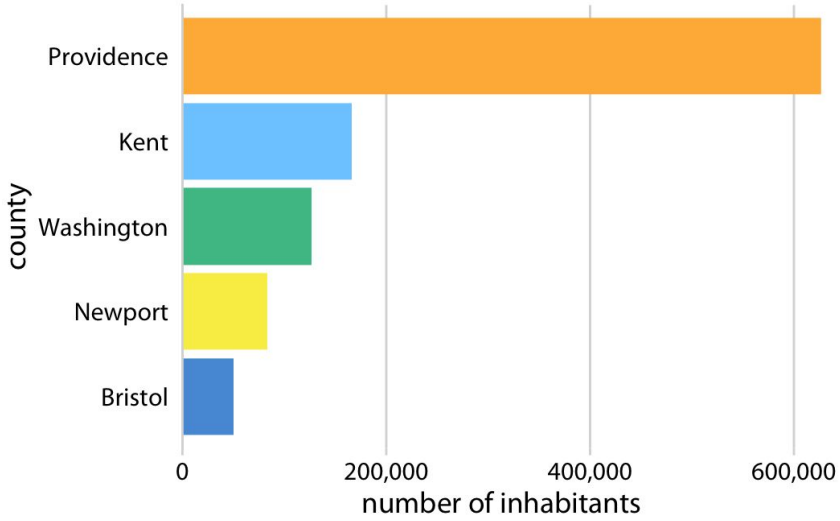


Design principles
1. The principle of proportional ink



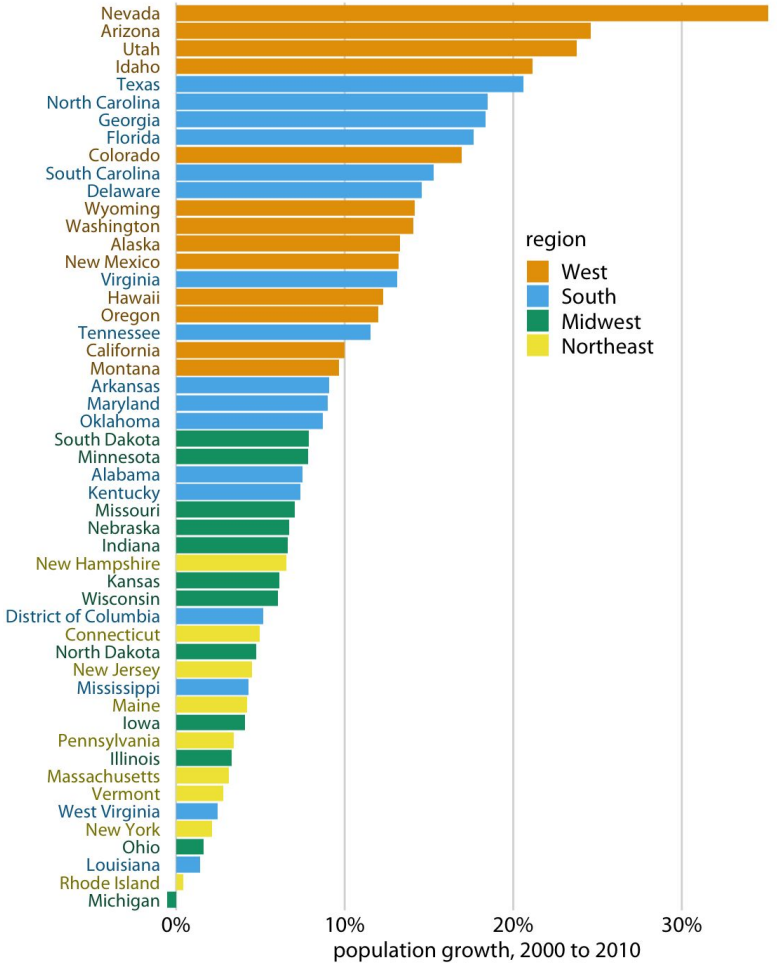
Design principles
1. The principle of proportional ink





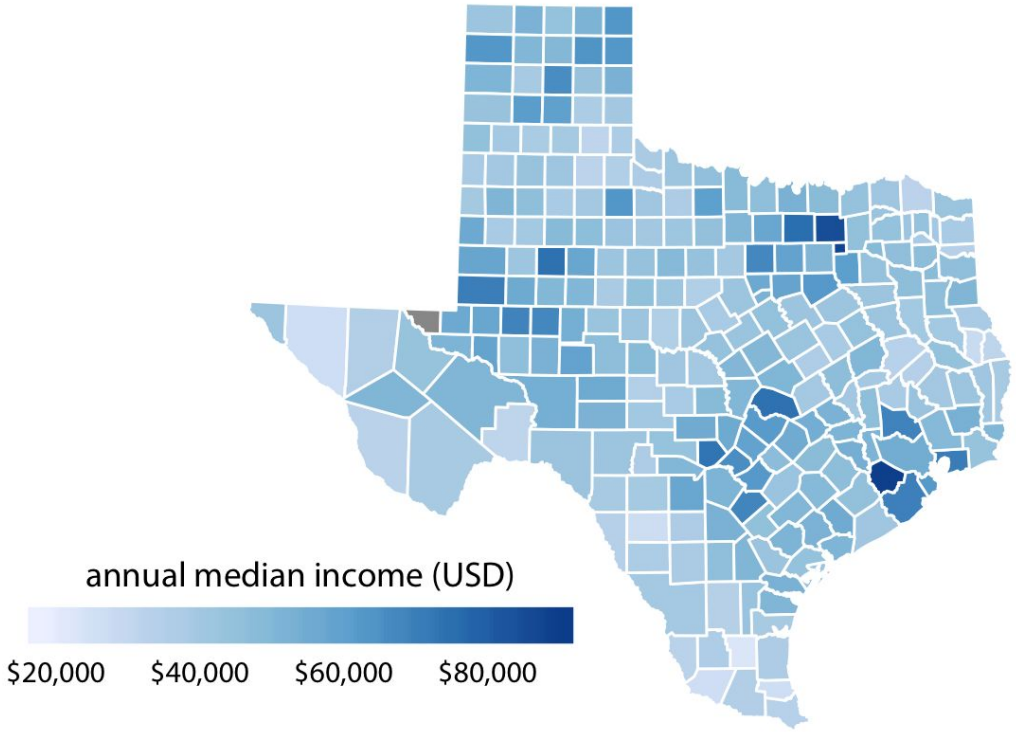
Design principles
2. Using colour

Using color to represent discrete items or groups that don't have an intrinsic order



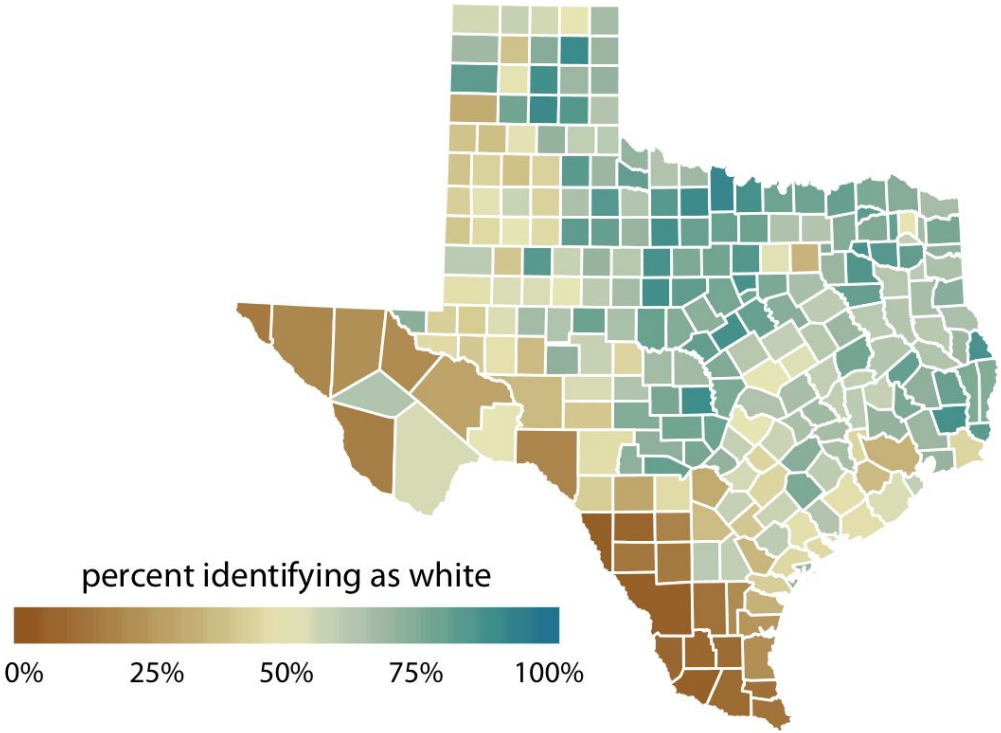
Design principles
2. Using colour

Using color to represent data values



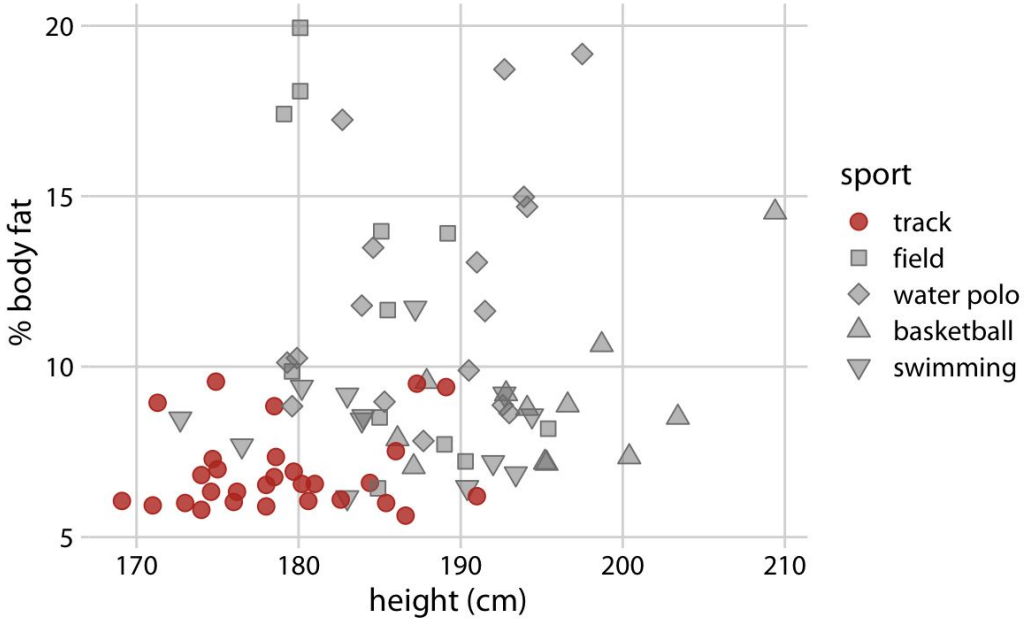
Design principles
2. Using colour

Using color to represent data values on a divergent scale



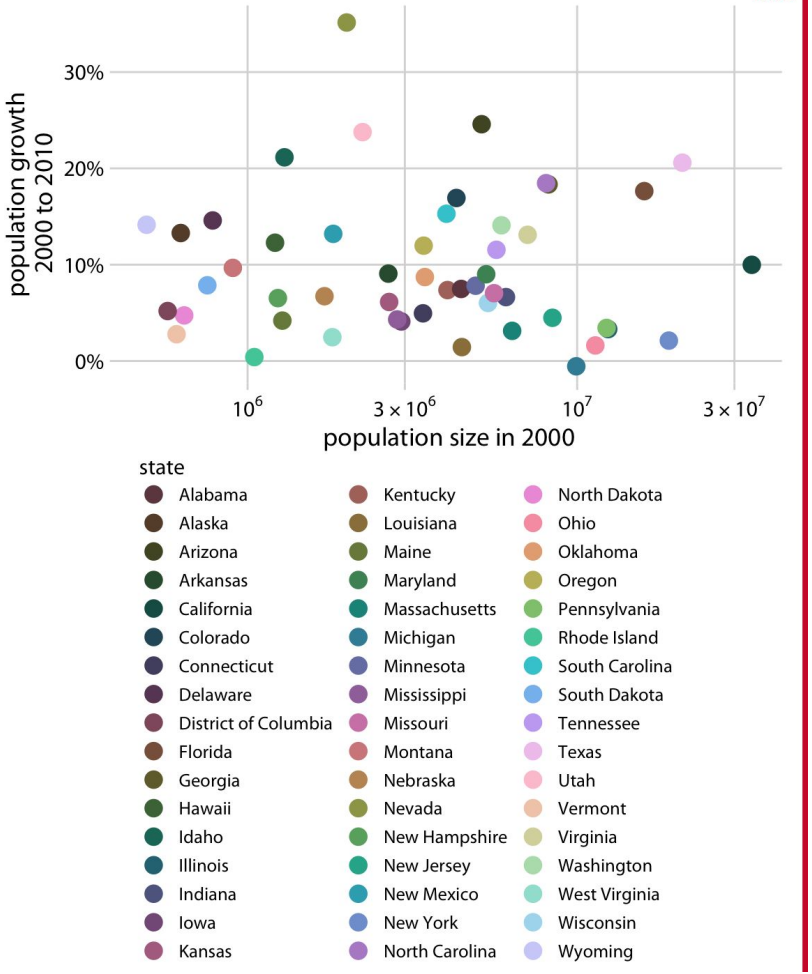
Design principles
2. Using colour

Using color to highlight a particular group

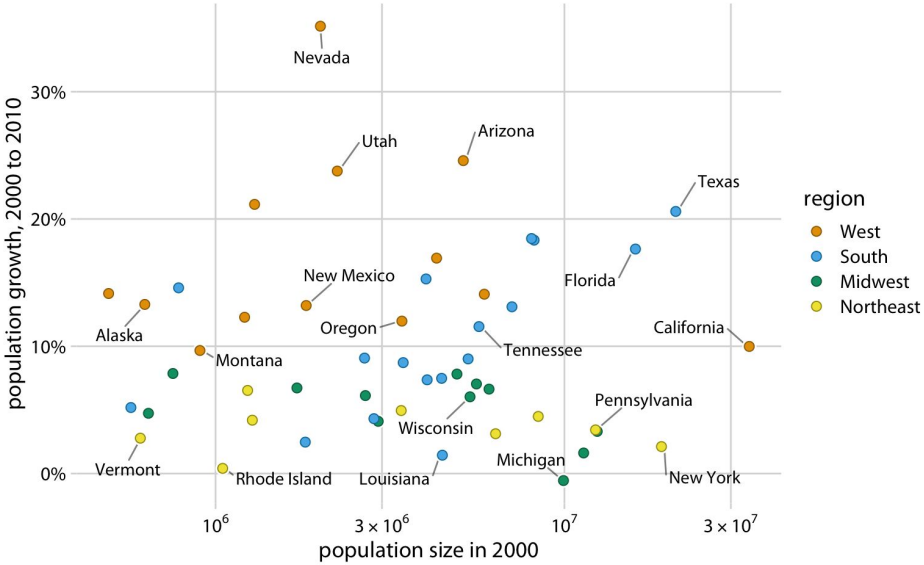


Design principles
2. Using colour

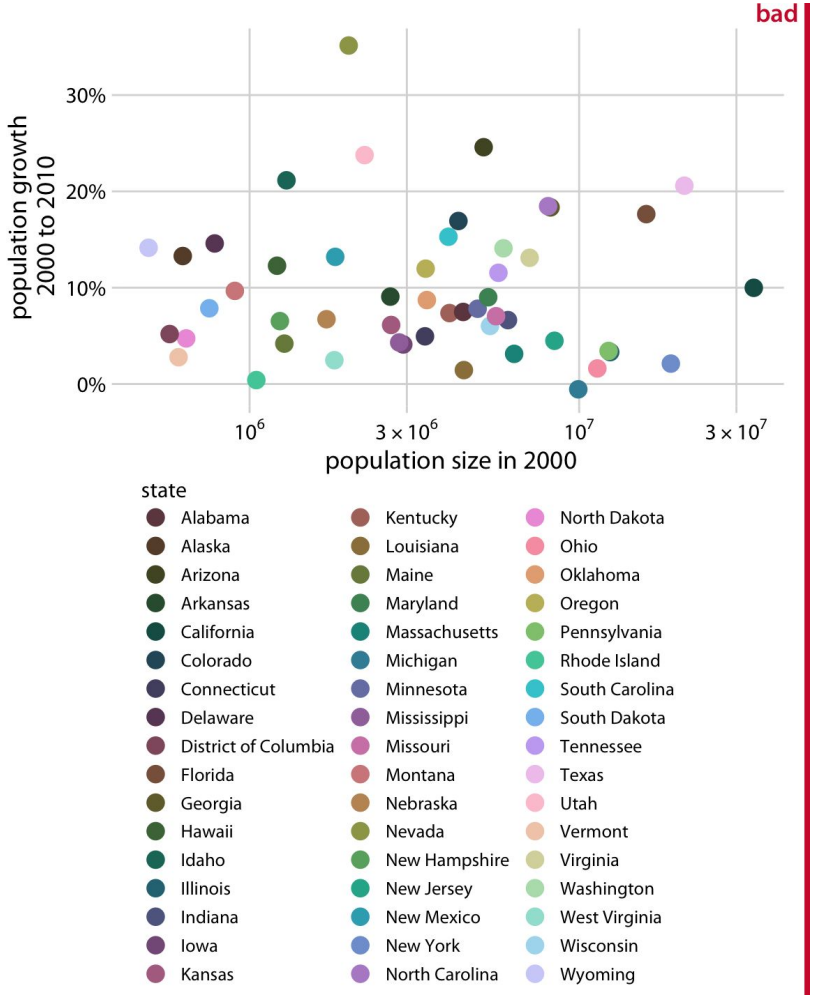
bad



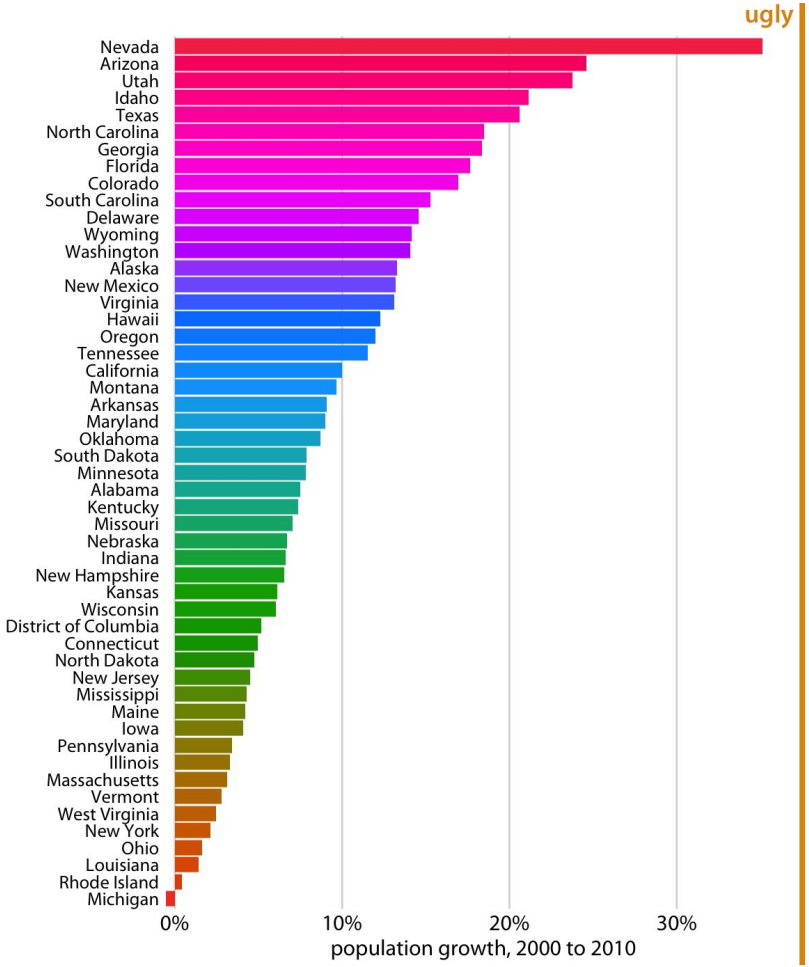
Design principles
2. Using colour



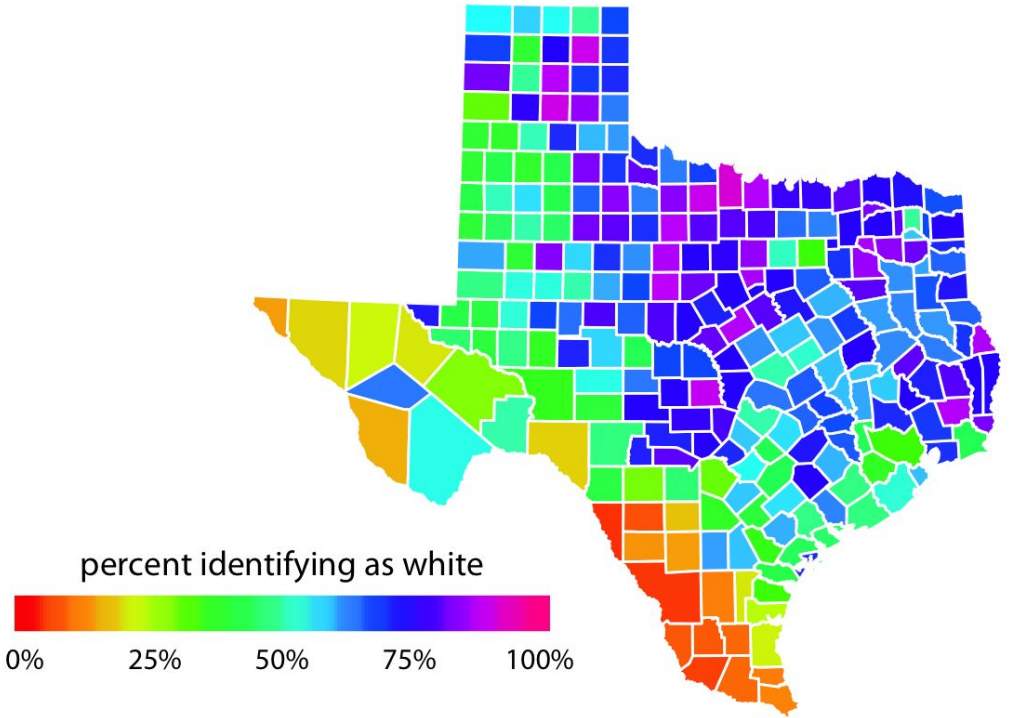
bad



Design principles
2. Using colour



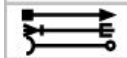
Design principles
2. Using colour



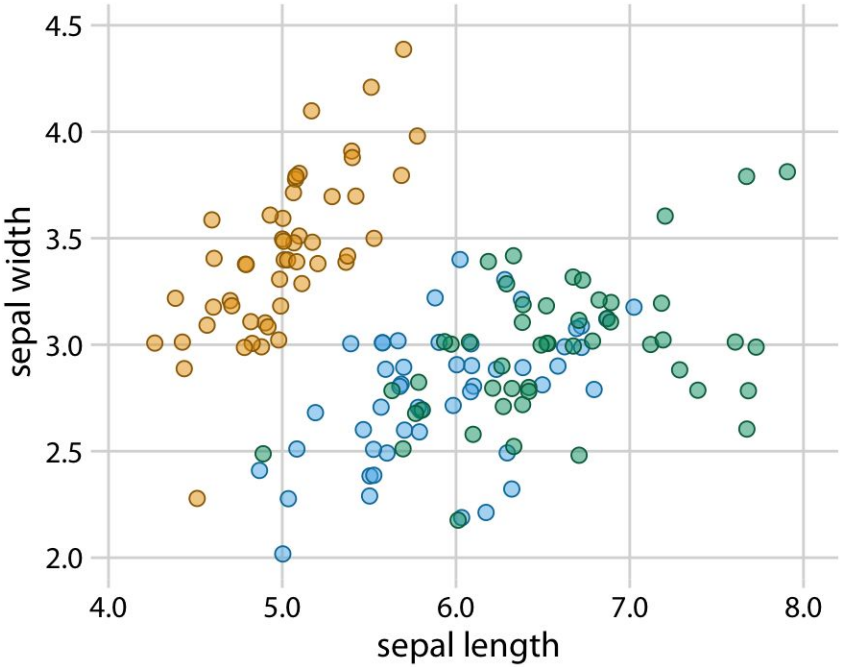
bad

Design principles
3. Using encodings

Pick the right encoding for your purpose

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	yes	infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium	Good	Good		
	pattern density	yes	few	Good	Good		
	weight, boldness	yes	few		Good		
	saturation, brightness	yes	few		Good		
	color	no	few (<20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	infinite			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	yes	few		Good		

Design principles
3. Using encodings

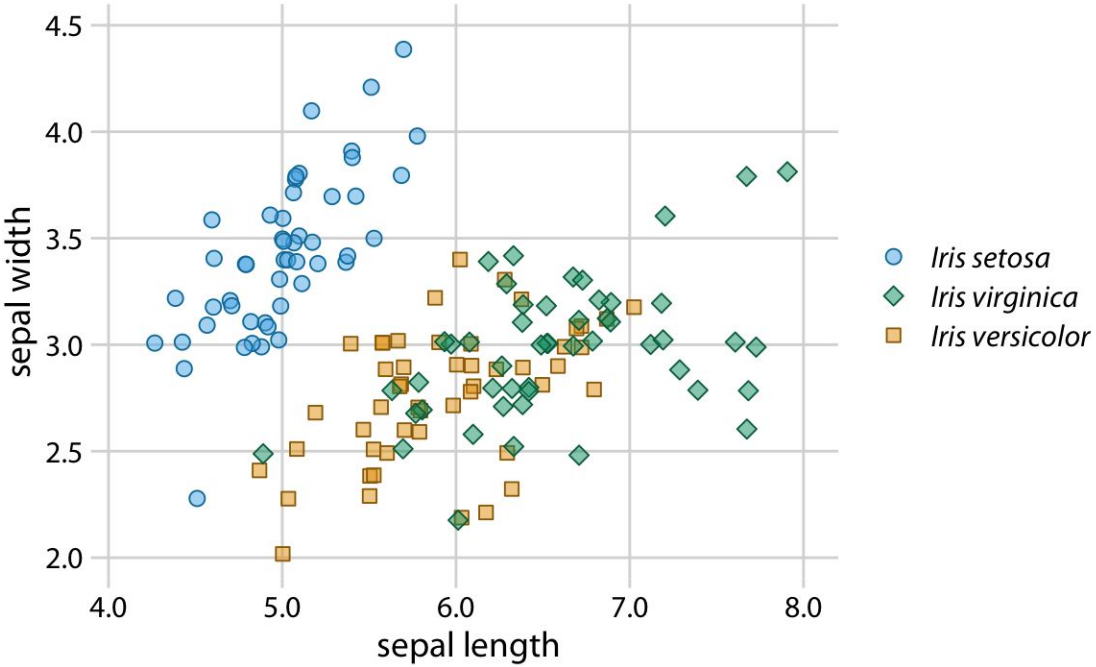


bad

Design principles
3. Using encodings

Redundant coding:

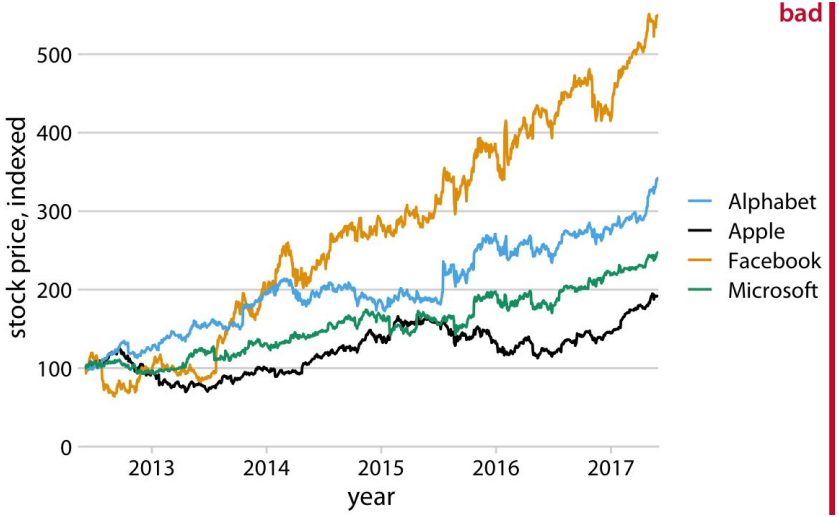
- In cases of clutter
- Color-vision deficiency
- Grayscale (printing)

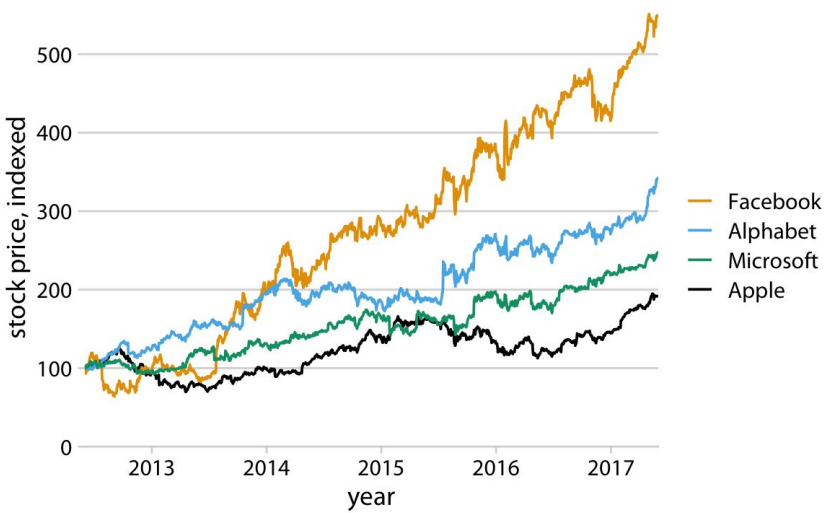


Design principles
3. Using encodings

Redundant coding:
- In cases of clutter
- Color-vision deficiency
- Grayscale (printing)

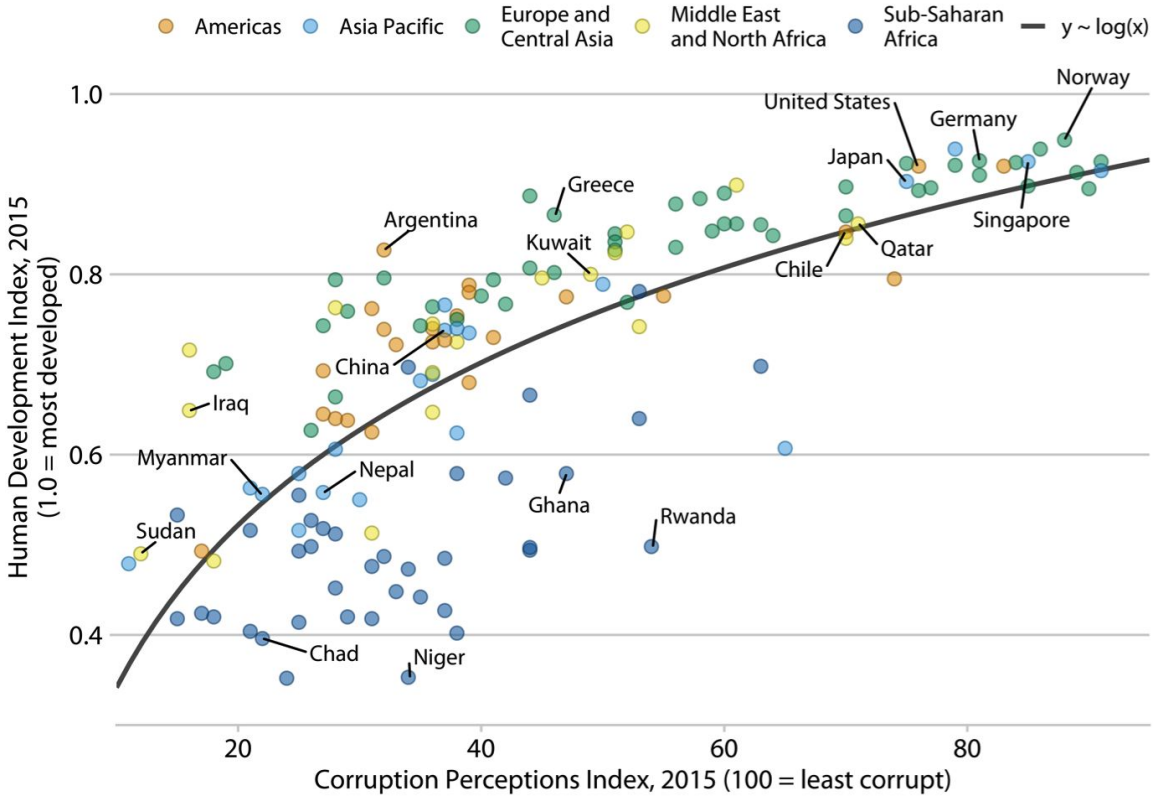
bad





Design principles
4. Label clearly

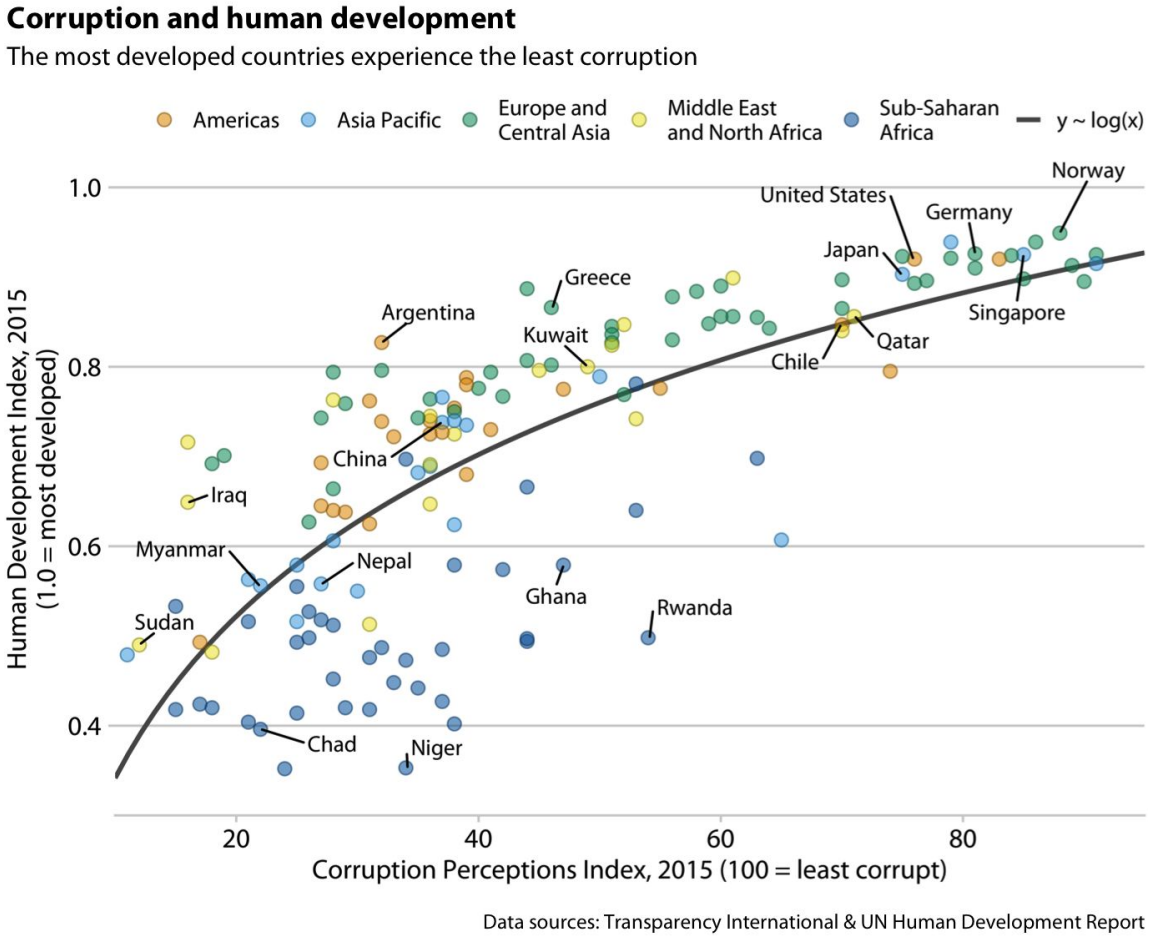
What does this figure show?



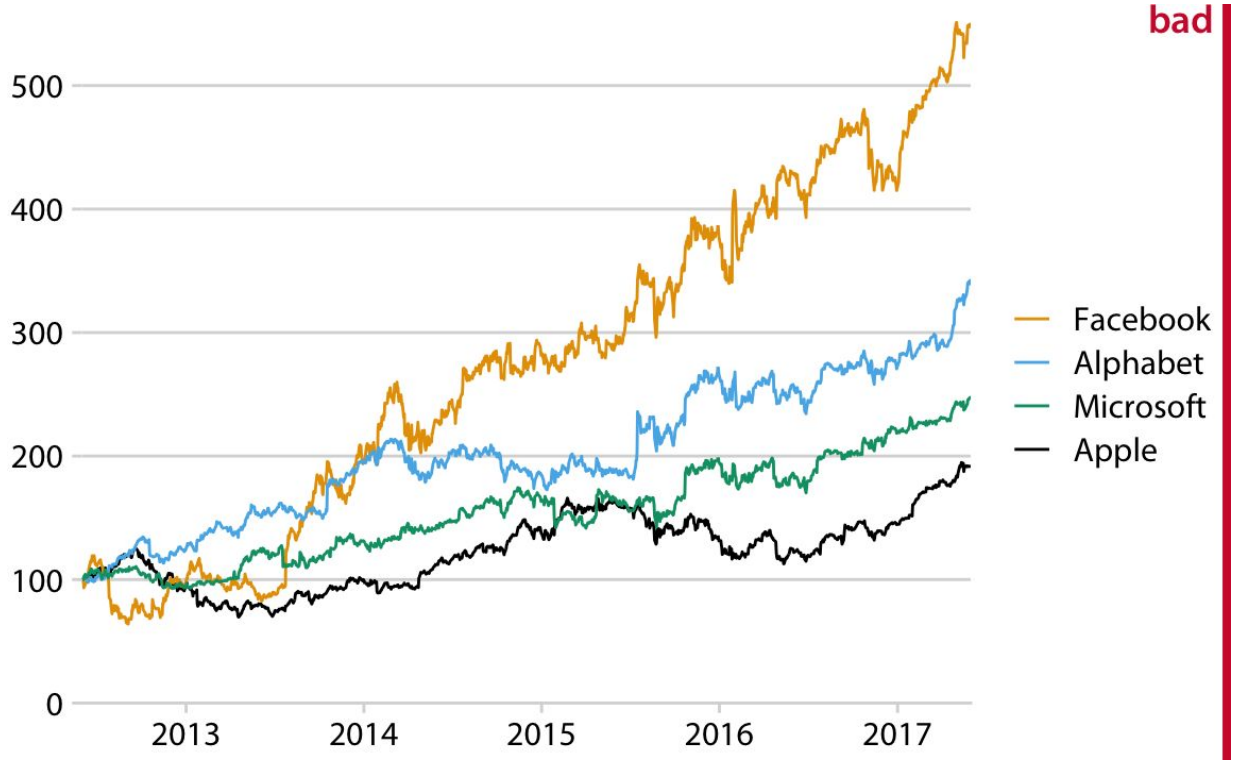
Design principles
4. Label clearly

What does this figure show?

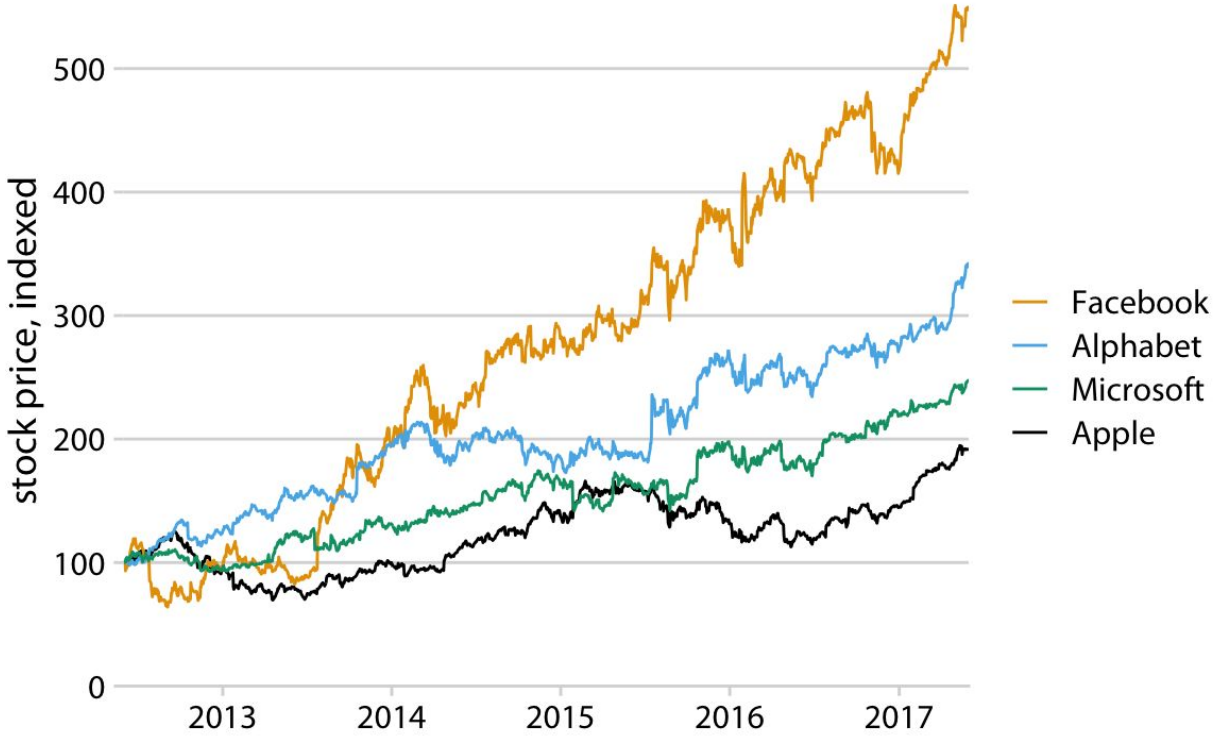
A title can help give context (especially when there is no figure caption).



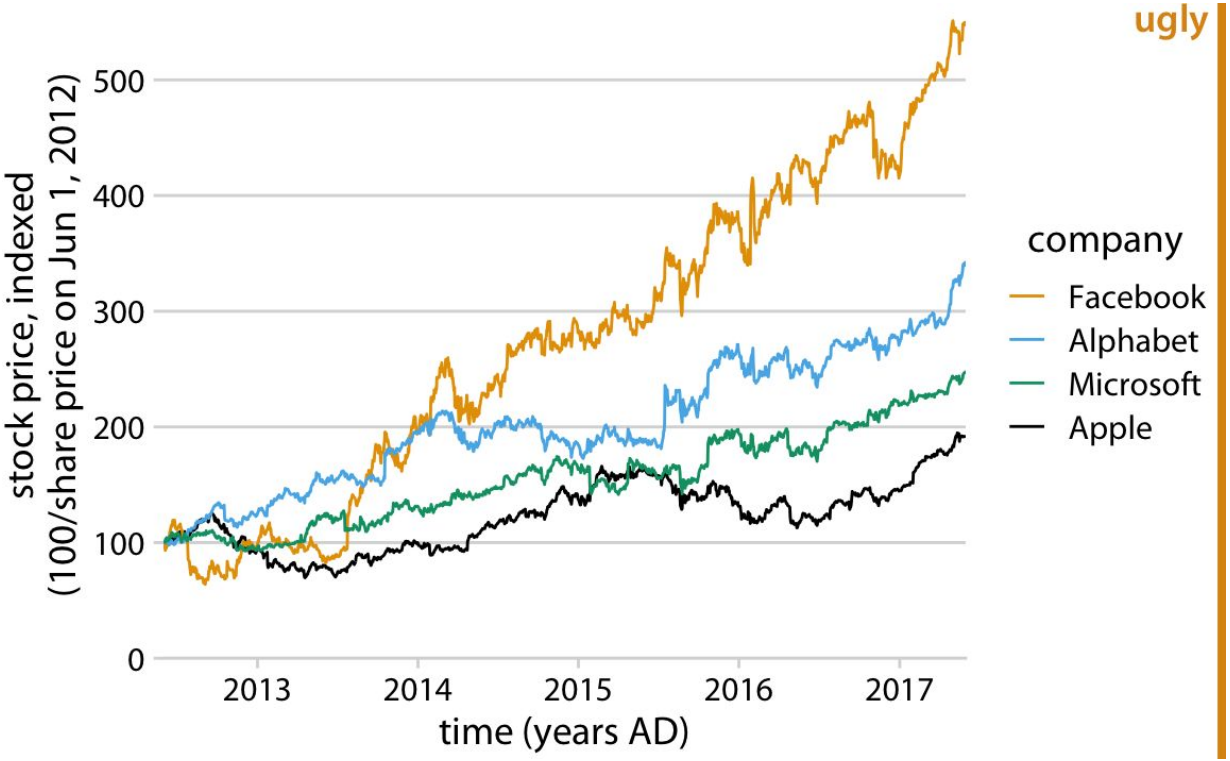
Design principles
4. Label clearly



Design principles
4. Label clearly



Design principles
4. Label clearly



For more on data visualization

- *Fundamentals of Data Visualization*

Demo

- https://github.com/EricElmoznino/lighthouse_visualization_tutorial