Data Science for Everyone – Visualization Pt 2

Dr. Ab Mosca (they/them)

Slides based off slides courtesy of Jordan Crouser (https://jcrouser.github.io/)

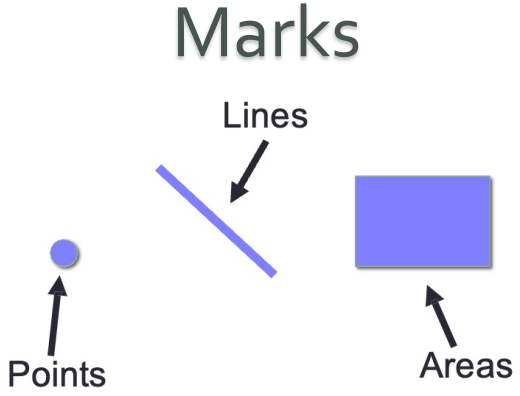
Plan for Today

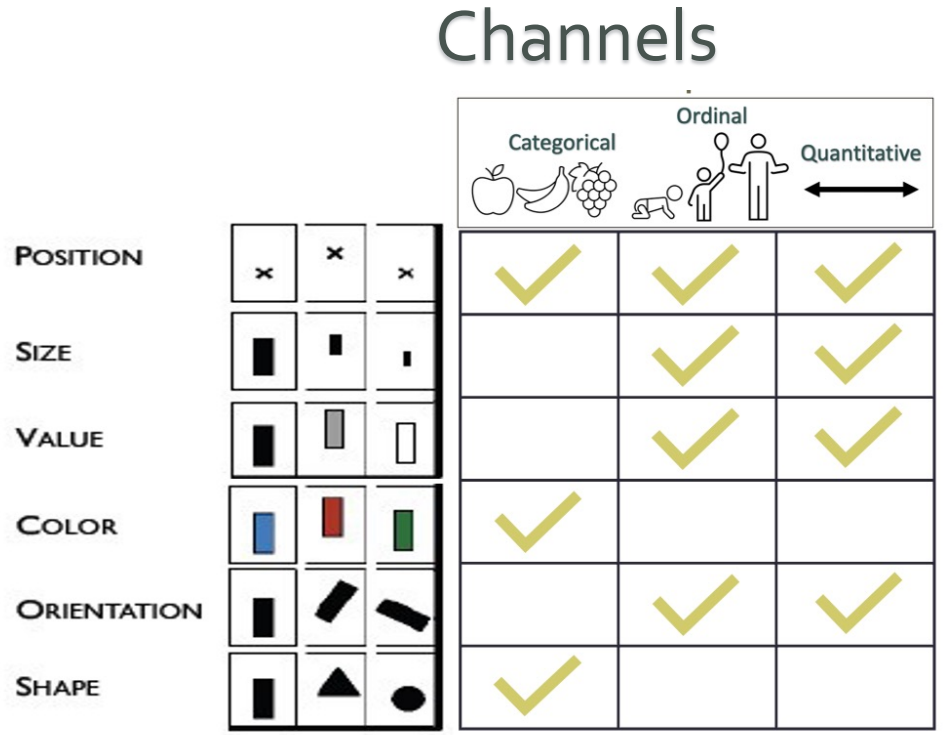
- Visualization Guiding Principles
- Visualization Ethics

Visualization Guiding Principles

Recall

- Visualizations (i.e. visual encodings) are made up of **marks** and **channels**
- We select marks and channels based on goals, data, and **other principles**



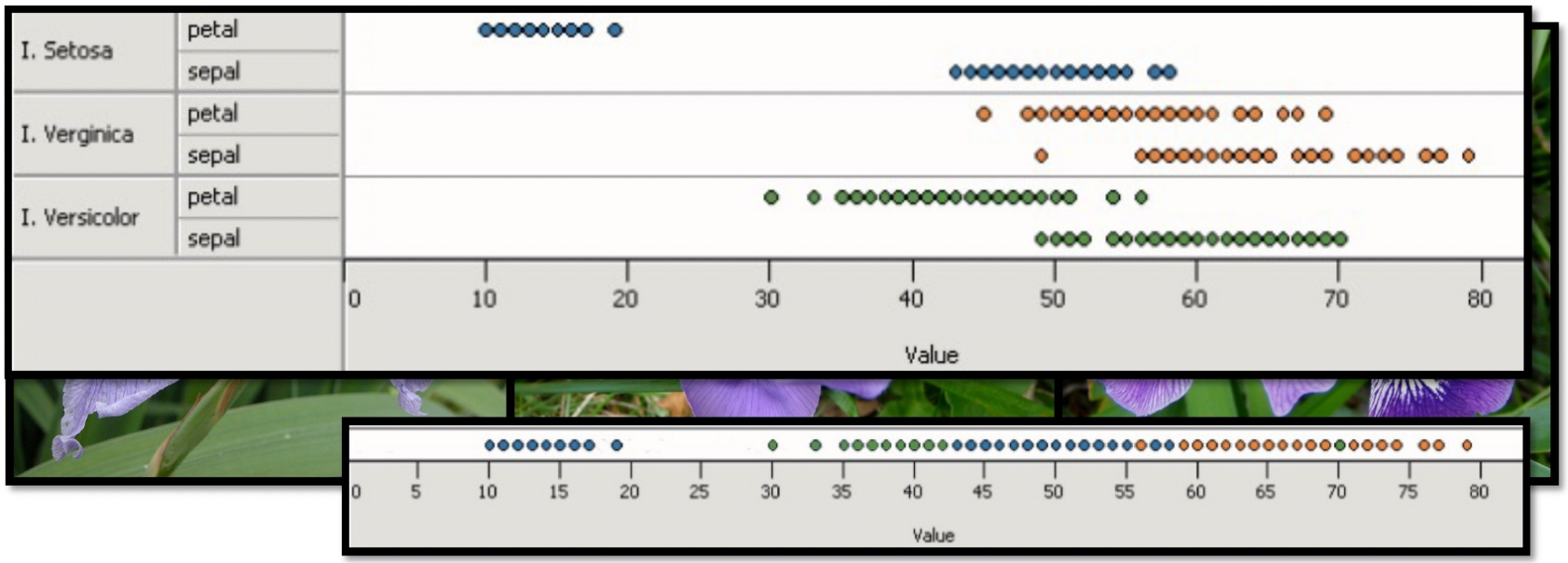


Jacques Bertin, *Semiologie Graphique* (Semiology of Graphics), 1967.

Mapping Data → Visuals

Principle 1: Expressiveness

*Encode **all the facts** and only the facts*



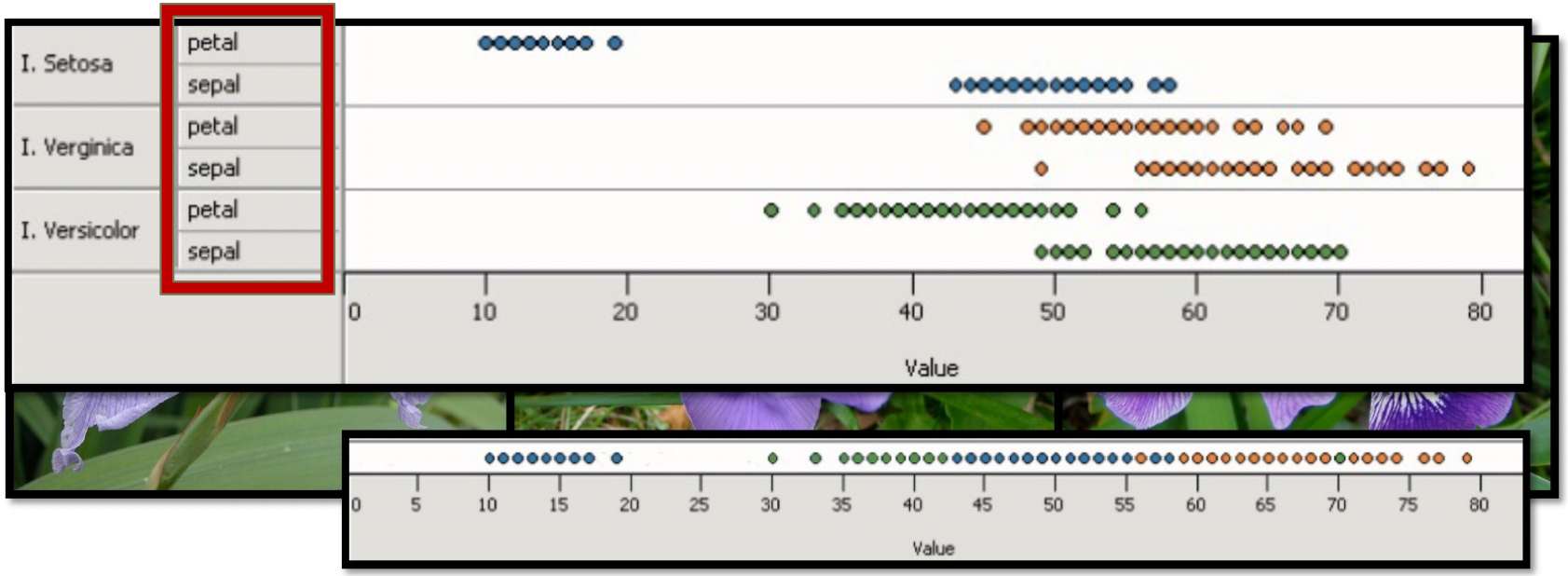
What data is in the top chart and not in the bottom chart?

Fig. Courtesy of M Krzywinski

Mapping Data → Visuals

Principle 1: Expressiveness

*Encode **all the facts** and only the facts*



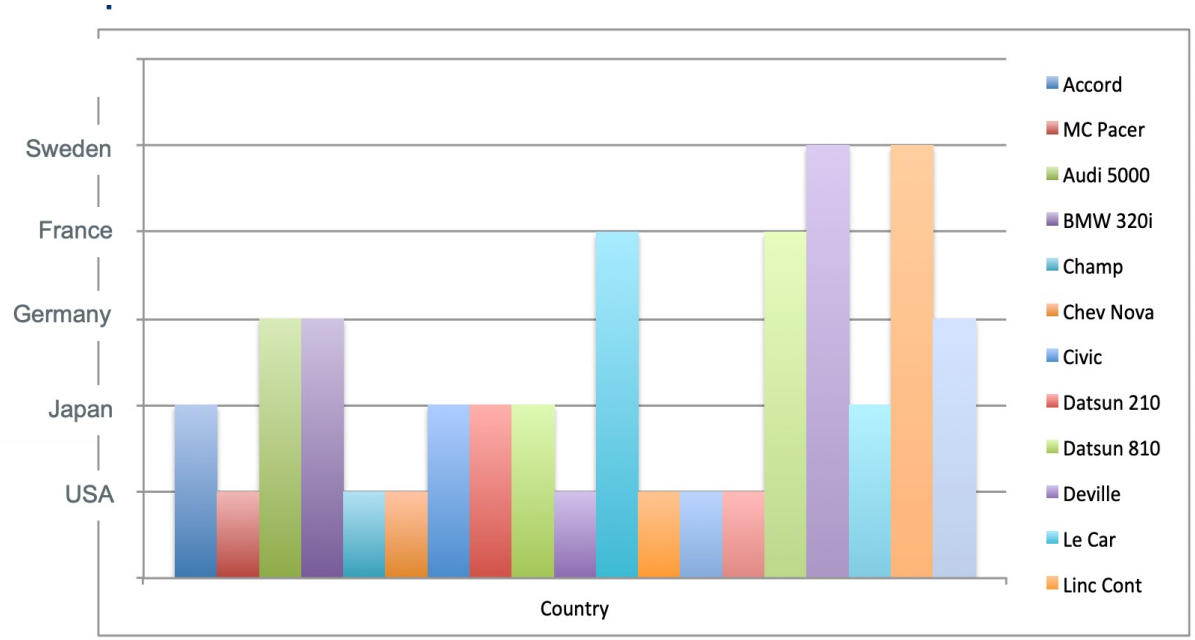
What data is in the top chart and not in the bottom chart?

Fig. Courtesy of M Krzywinski

Mapping Data → Visuals

Principle 1: Expressiveness

*Encode all the facts and **only the facts***

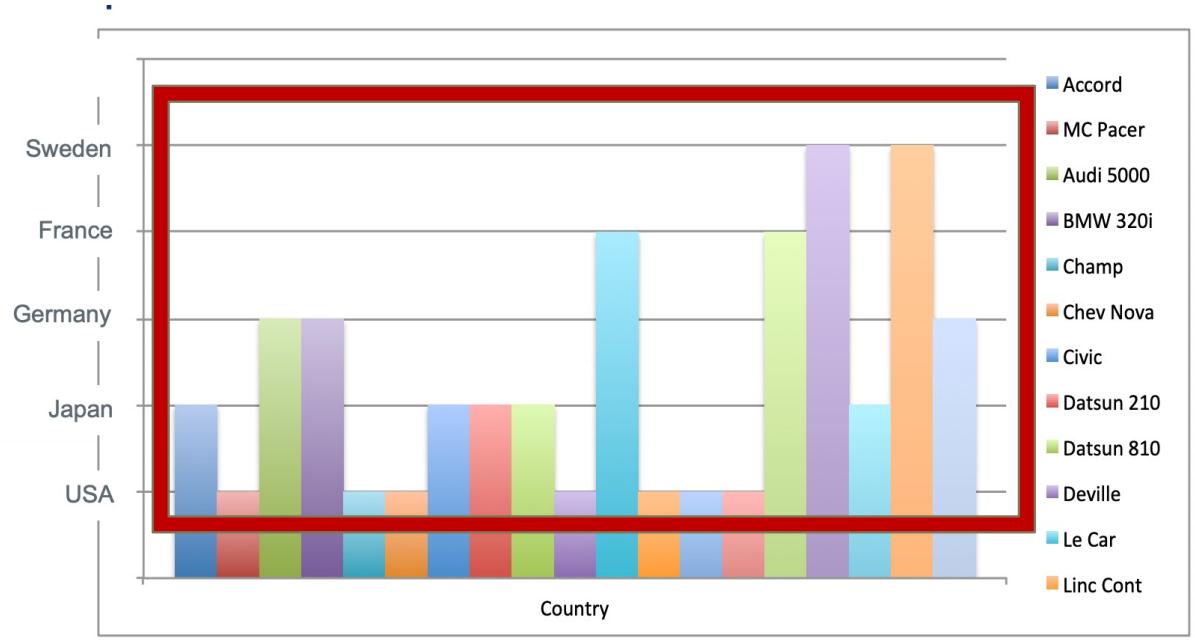


What "extra" data is included in this visualization?

Mapping Data → Visuals

Principle 1: Expressiveness

*Encode all the facts and **only the facts***

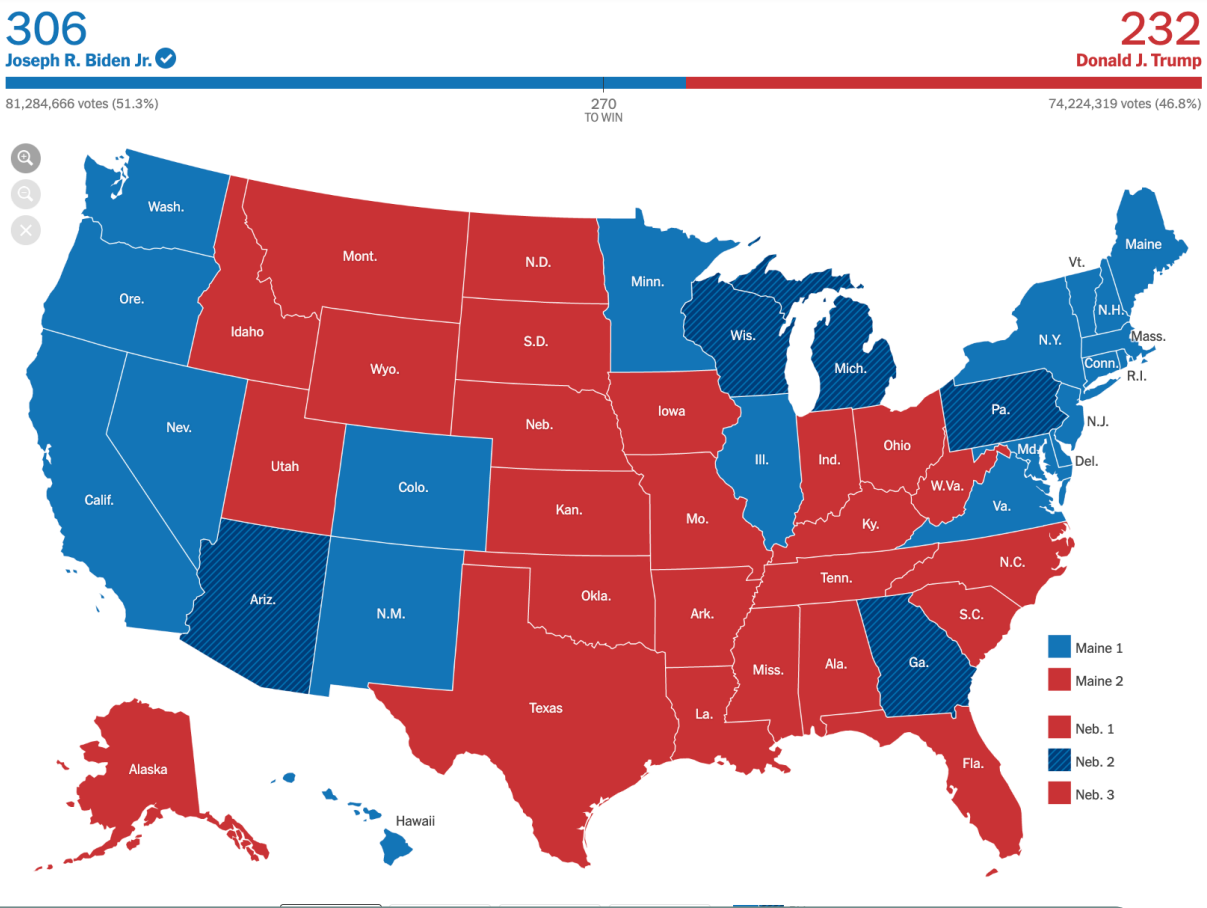


What "extra facts" are included in this visualization?

Mapping Data → Visuals

Principle 1: Expressiveness

Encode all the facts and only the facts



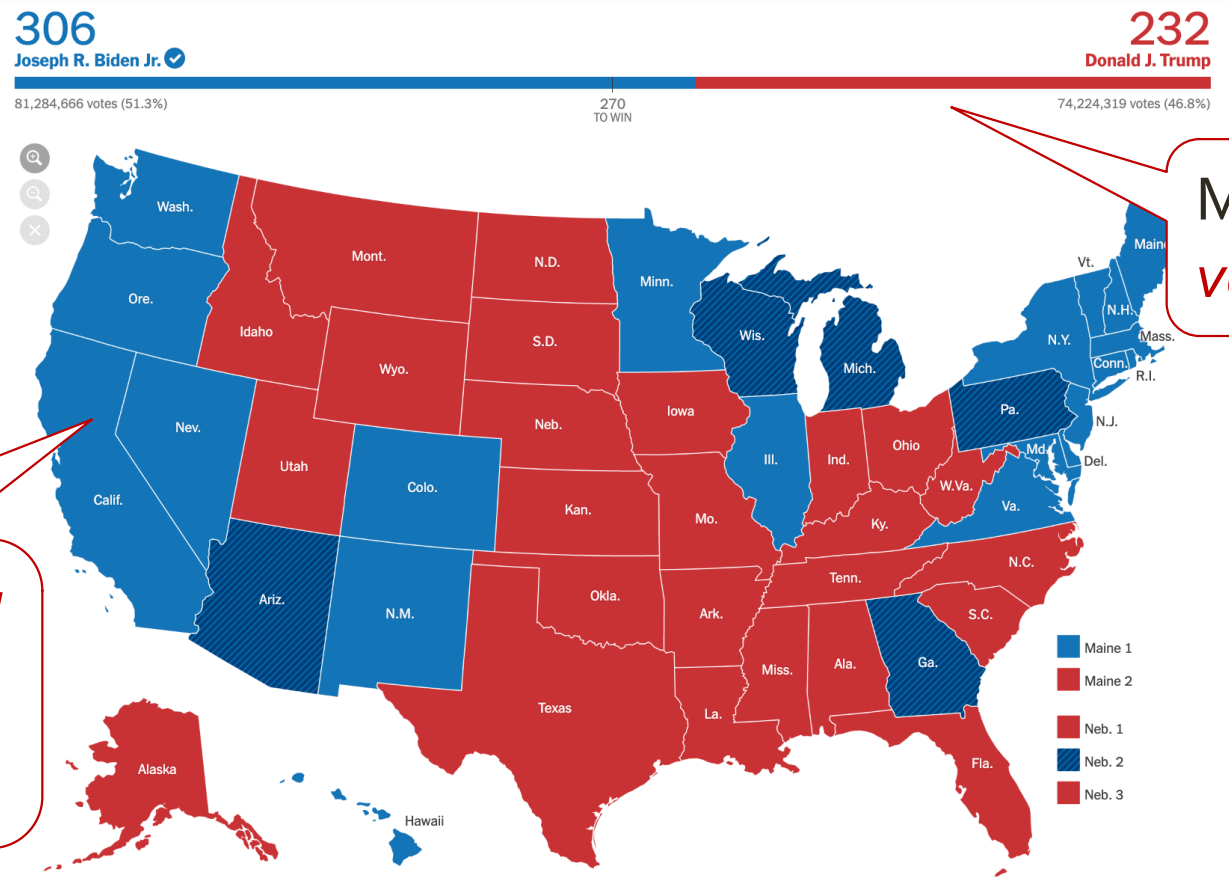
What is wrong with this visualization?

https://www.nytimes.com/interactive/2020/11/03/us/elections/results-president.html

Mapping Data → Visuals

Principle 1: Expressiveness

Encode all the facts and only the facts



Most *electoral votes* win

Colors highlight *land area* per state, not electoral votes

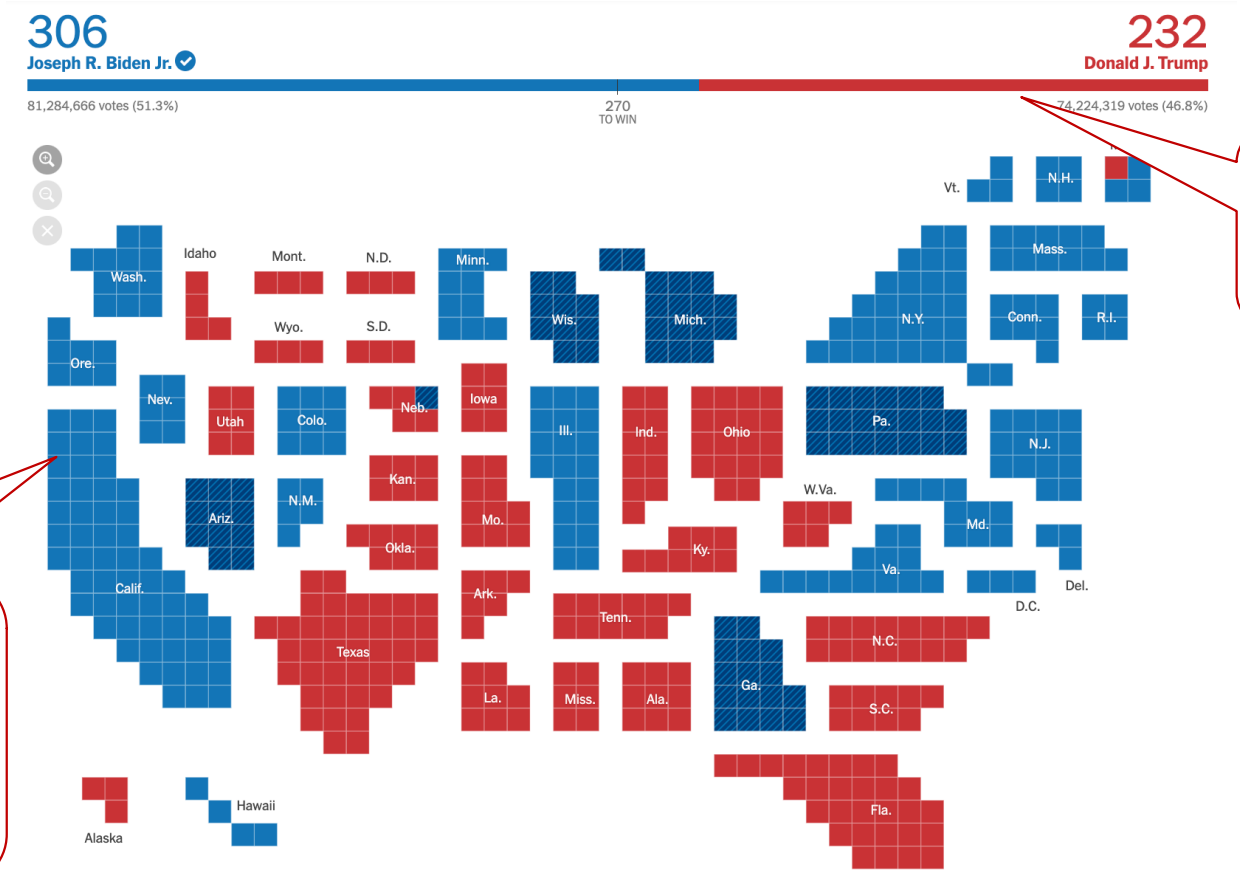
What is wrong with this visualization?

https://www.nytimes.com/interactive/2020/11/03/us/elections/results-president.html

Mapping Data → Visuals

Principle 1: Expressiveness

Encode all the facts and only the facts



Most *electoral votes* win

Colors highlight *electoral votes* per state

https://www.nytimes.com/interactive/2020/11/03/us/elections/results-president.html

Mapping Data → Visuals

Principle 2: Effectiveness

Most effective channels should be used for most important data

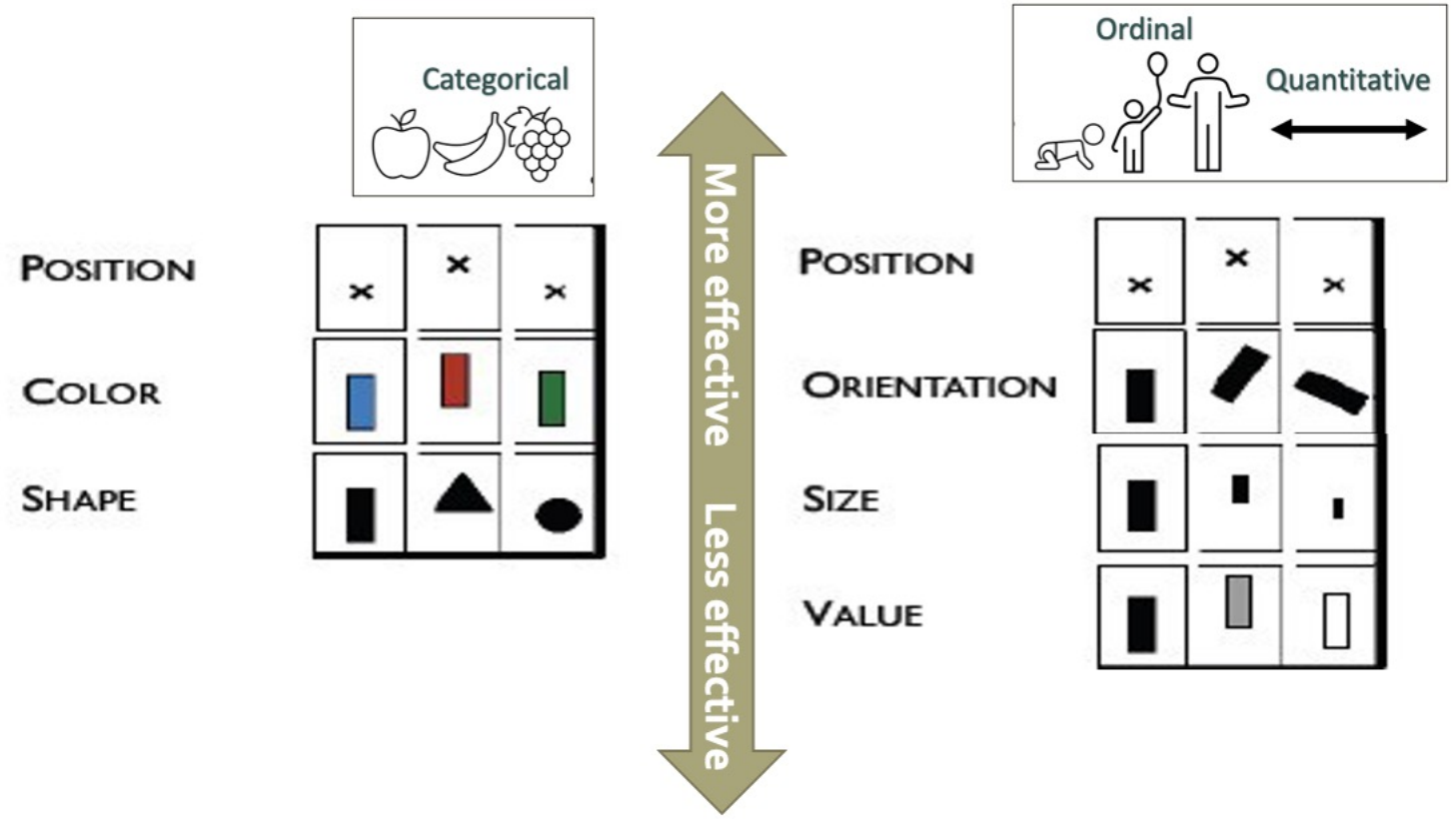
Effectiveness = Based on a compilation of research, how well a channel supports:

- Accuracy
- Discriminability
- Separability
- Visual popout
- Grouping

Mapping Data → Visuals

Principle 2: Effectiveness

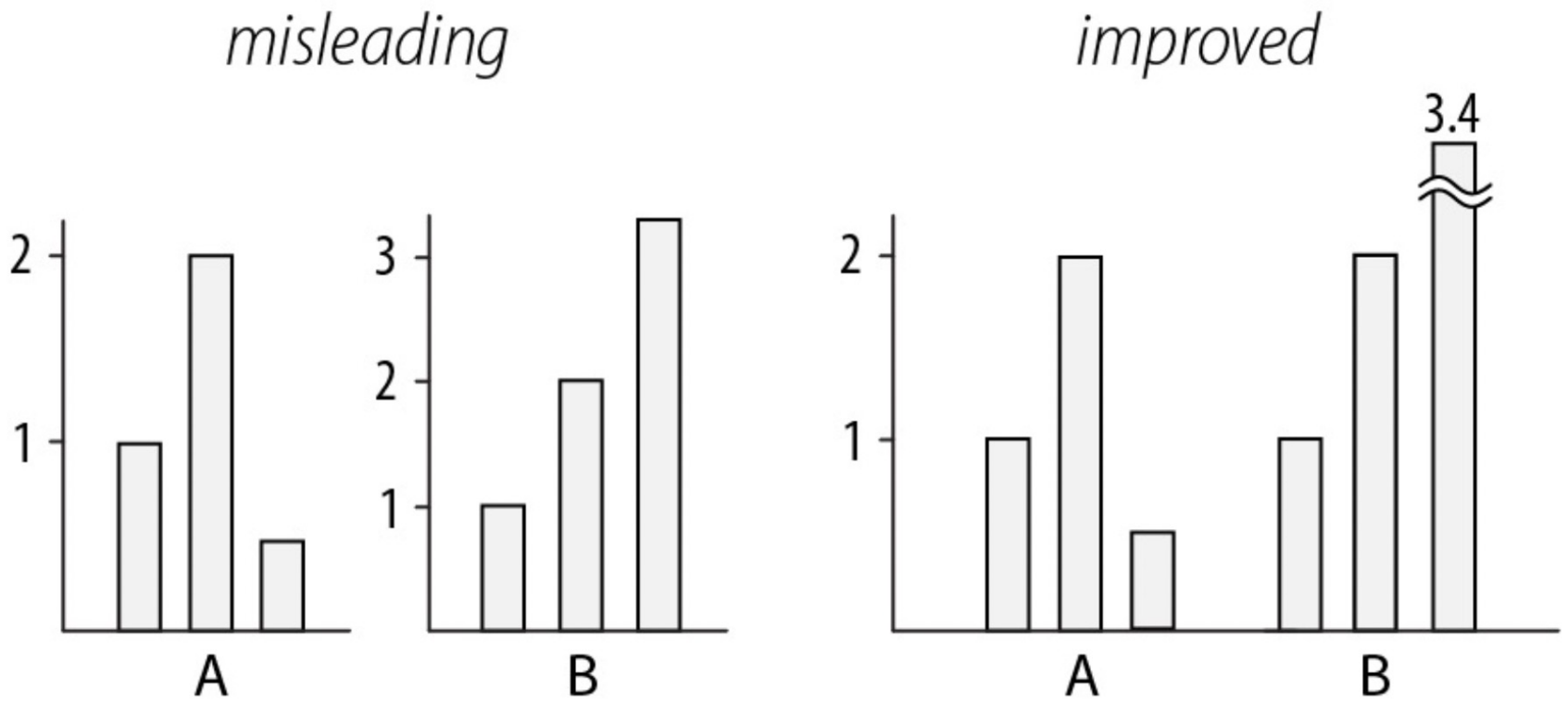
Most effective channels should be used for most important data



Mapping Data → Visuals

Principle 3: Consistency

Use consistent axes for comparisons



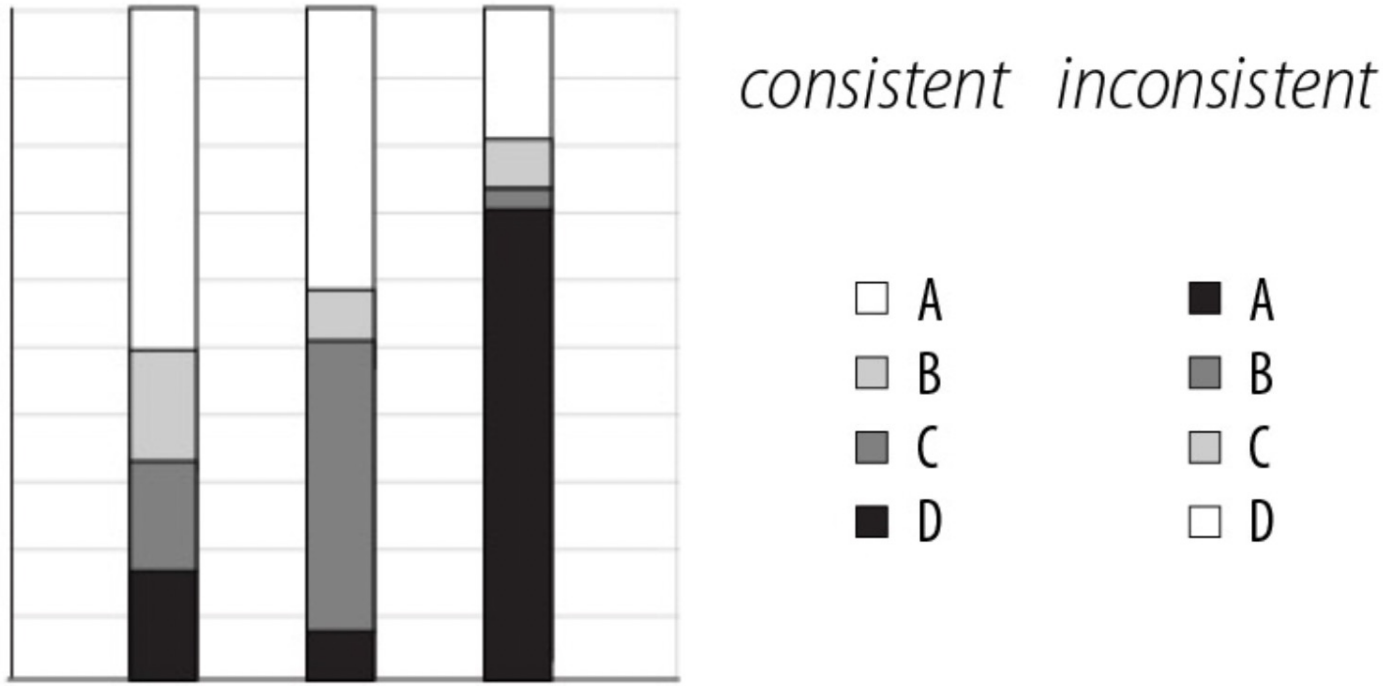
Raina SZ, et al. (2005) Evolution of base-substitution gradients in primate mitochondrial genomes. Genome Res 15: 665-673.

M. Krzwinski, behind every great visualization is a design principle, 2012

Mapping Data → Visuals

Principle 3: Consistency

Order legend items according to appearance

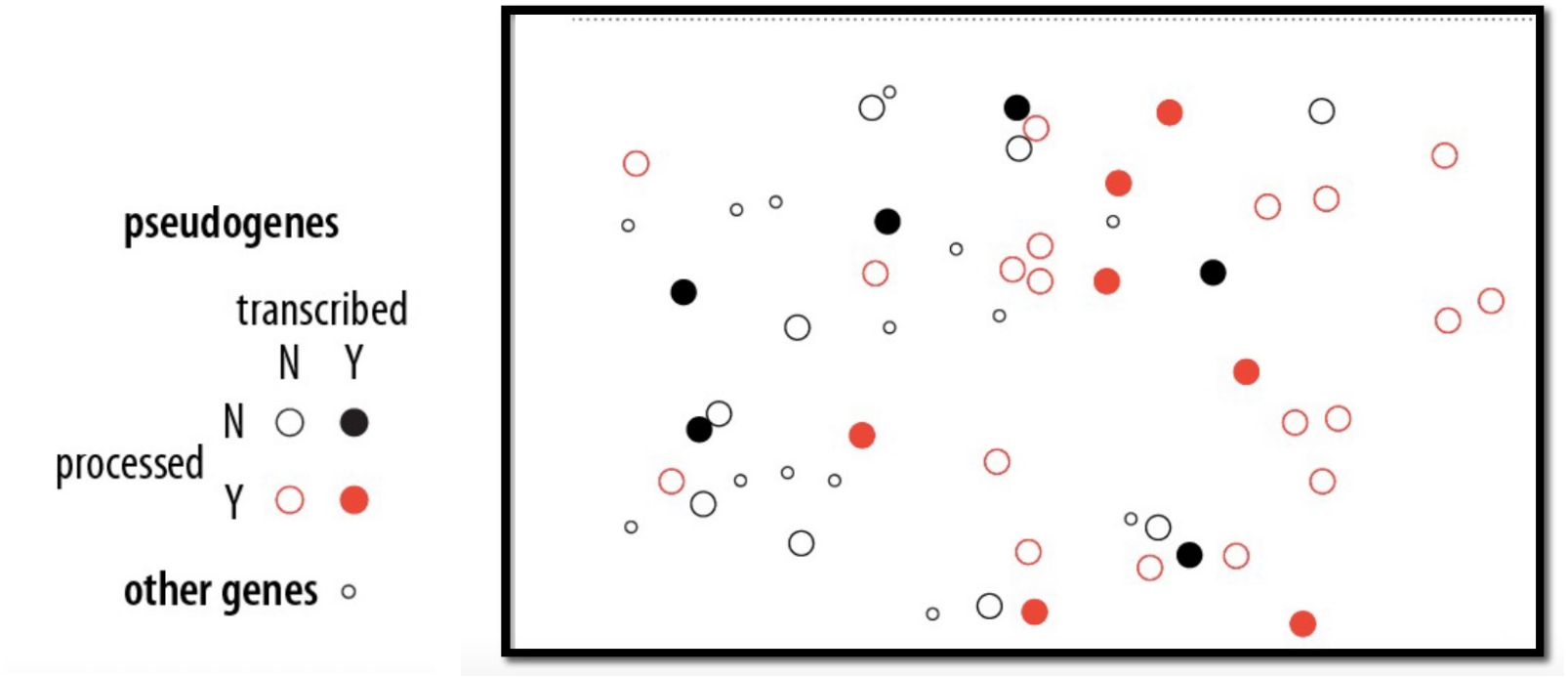


M. Krzwinski, behind every great visualization is a design principle, 2012

Mapping Data → Visuals

Principle 3: Separability

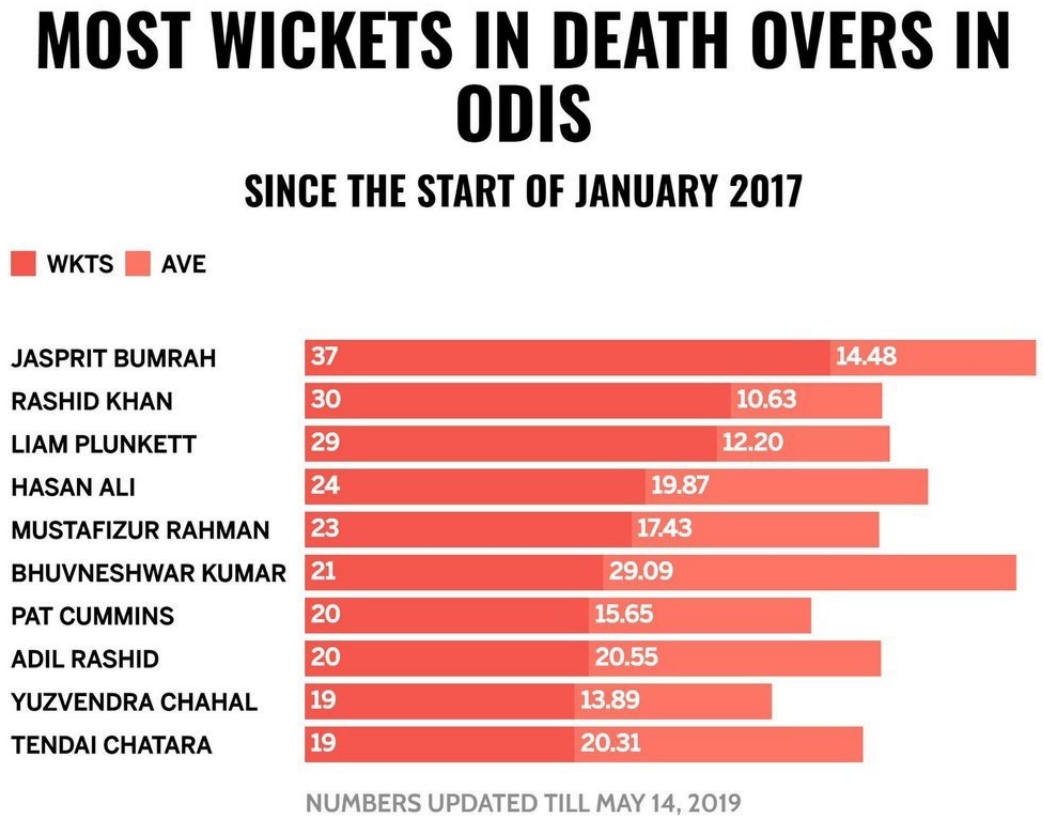
Avoid visually similar encodings for independent variables



Mapping Data → Visuals

Principle 3: Separability

Avoid visually similar encodings for independent variables

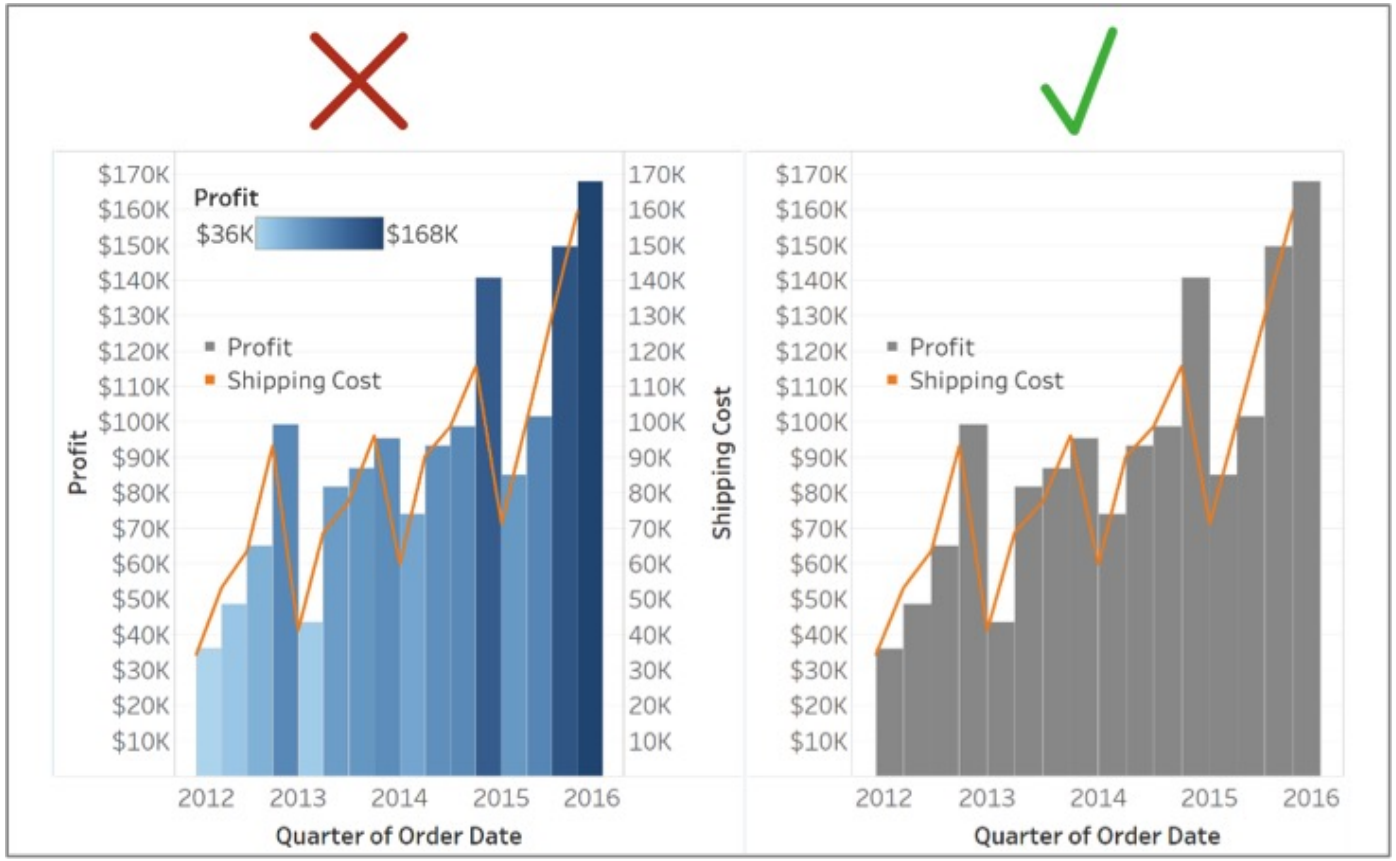


https://www.syntaxtechs.com/blog/data-visualization-examples

Mapping Data → Visuals

Principle 4: Simplicity

Avoid double encoding data

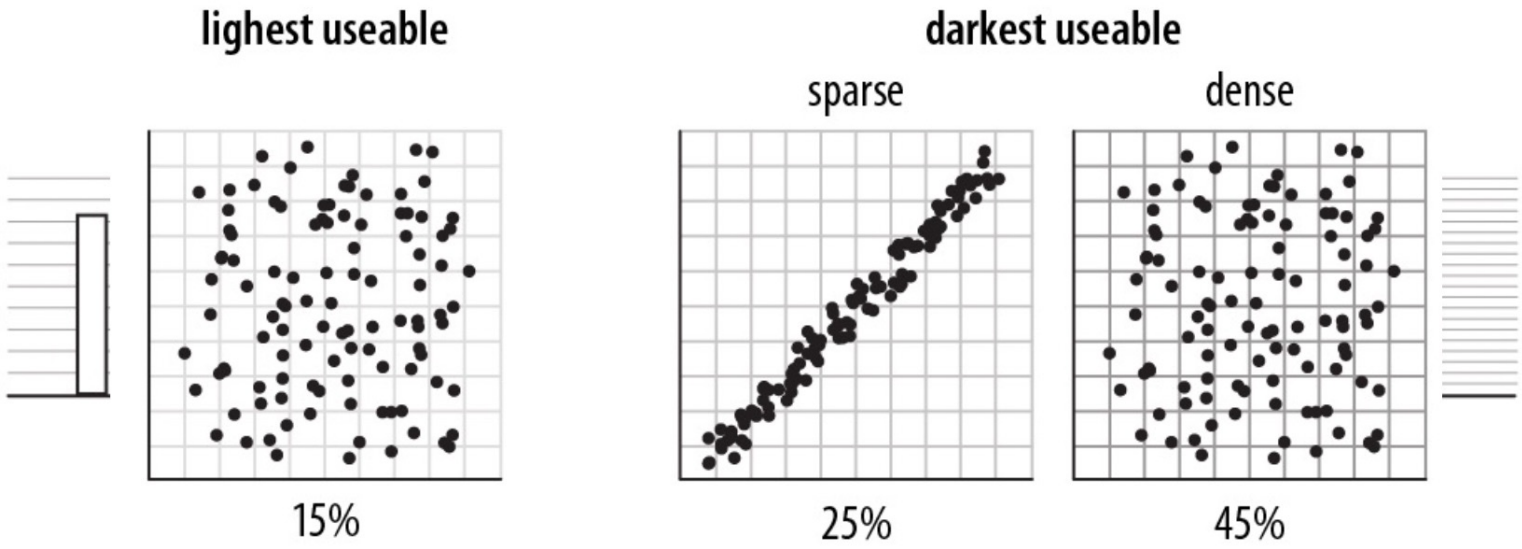


https://www.teknionusa.com/blog/the-10-commandments-of-visual-analytics-in-tableau

Mapping Data → Visuals

Principle 4: Simplicity

Navigational aids should not compete with data



Heer J, Bostock M (2010) Crowdsourcing graphical perception: using mechanical turk to assess visualization design. Proceedings of the 28th international conference on Human factors in computing systems. Atlanta, Georgia, USA: ACM. pp. 203-212.

Responsible Creation and Consumption of Vis

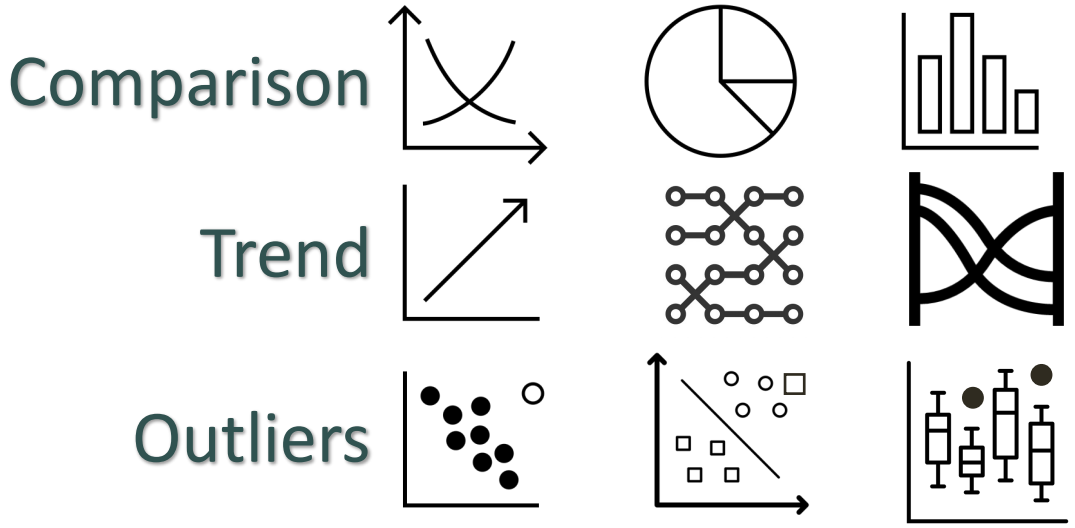
How to visualize data?

1. Goal

2. Data Types

How to visualize data?

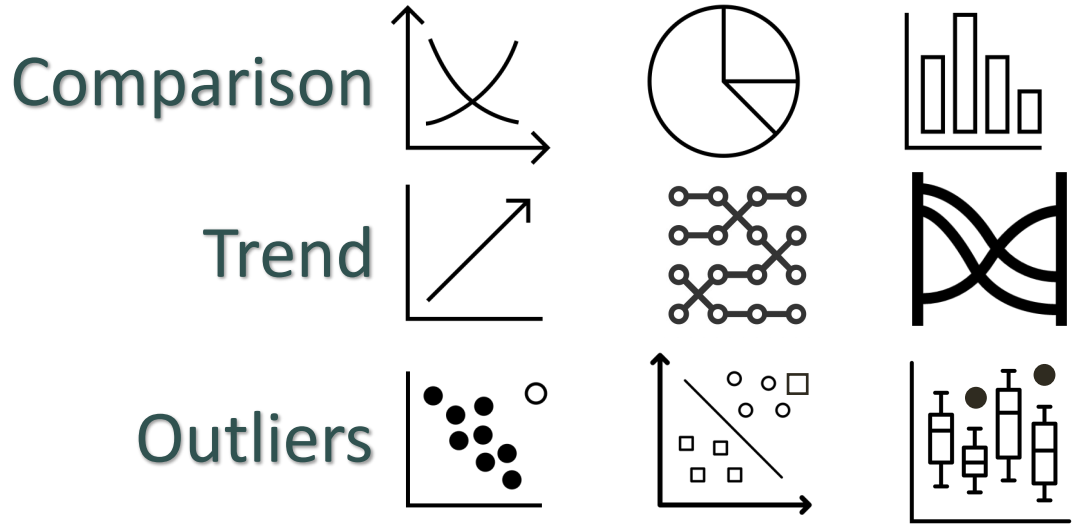
1. Goal → What do you want to communicate or facilitate?



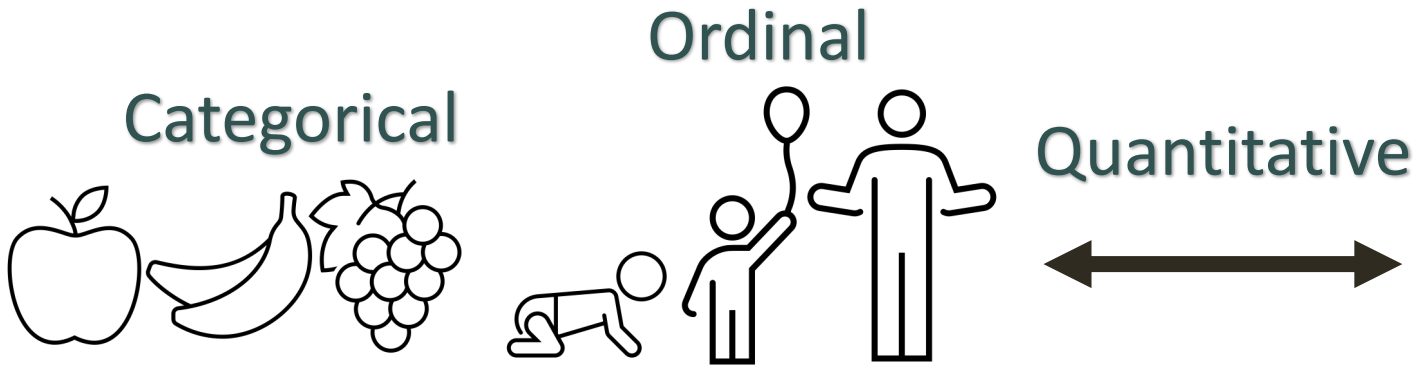
2. Data Types

How to visualize data?

1. **Goal** → What do you want to communicate or facilitate?

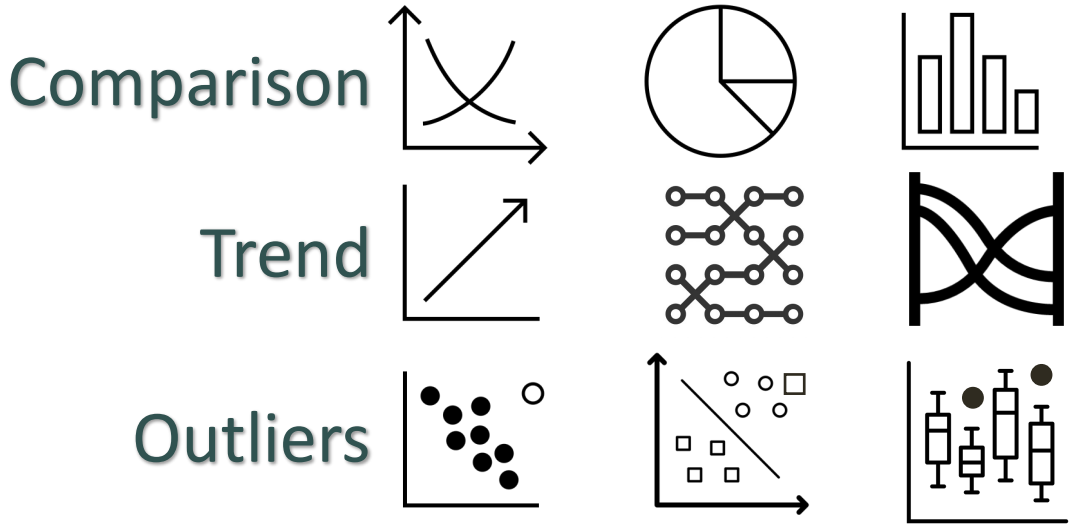


2. **Data Types** → What kinds of data do you need to show?



How to visualize data?

1. **Goal** → What do you want to communicate or facilitate?

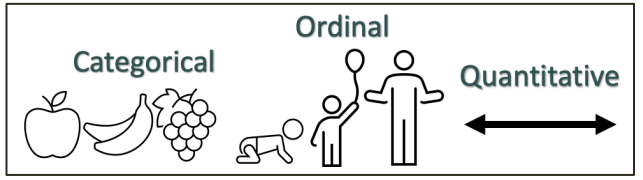


2. **Data Types** → What kinds of data do you need to show?

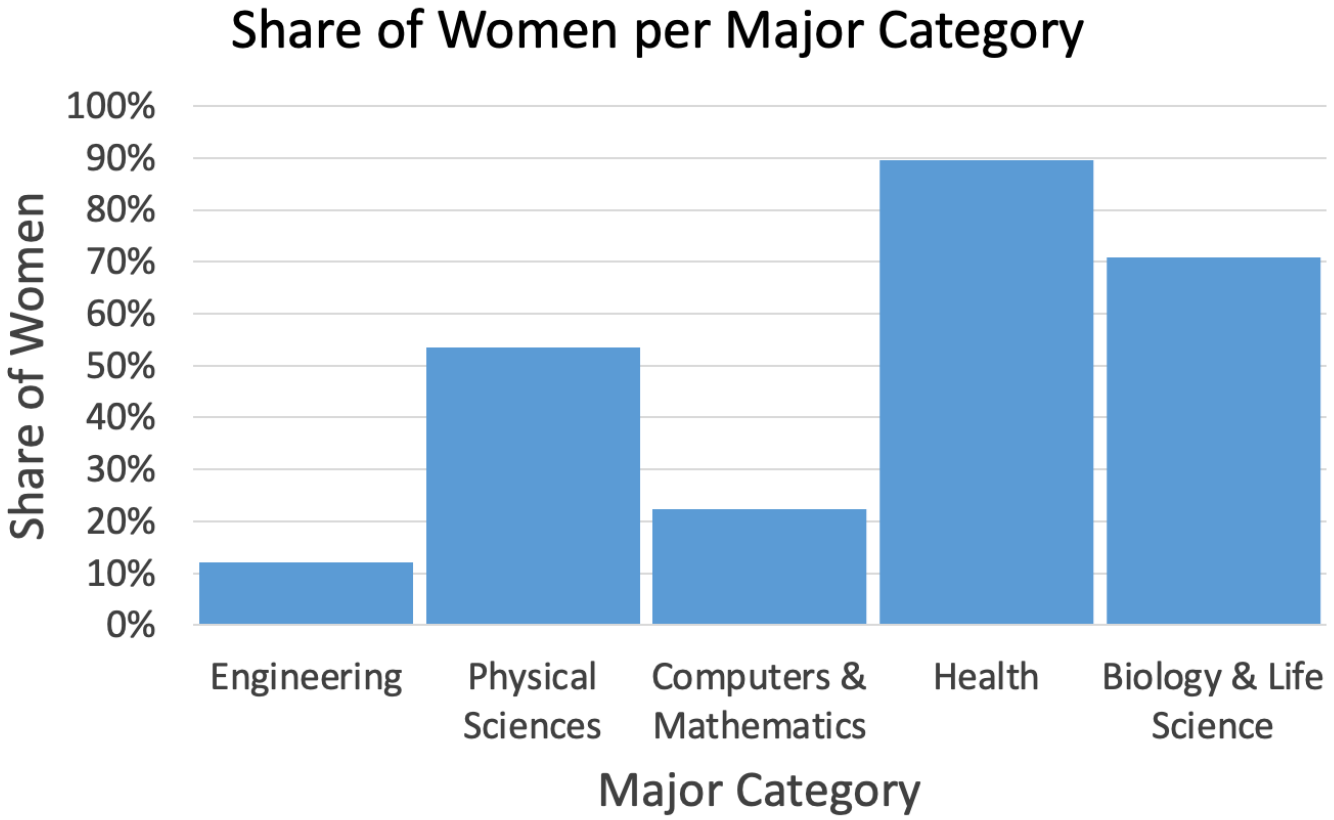
Rank	Major_category	Total	Men	Women	Share_ women	Median_ earnings
1	Engineering	2339	2057	282	12%	110000
7	Physical Sciences	1792	832	960	54%	62000
19	Computers & Mathematics	128319	99743	28576	22%	53000
27	Health	209394	21773	187621	90%	48000
36	Biology & Life Science	1762	515	1247	71%	45000

Bar charts

1. Goal → Comparison
2. Data Types → Categorical or Ordinal vs. Quantitative



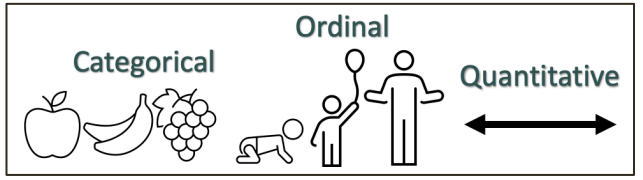
Rank	Major_category	Total	Men	Women	Share_ women	Median_ earnings
1	Engineering	2339	2057	282	12%	110000
7	Physical Sciences	1792	832	960	54%	62000
19	Computers & Mathematics	128319	99743	28576	22%	53000
27	Health	209394	21773	187621	90%	48000
36	Biology & Life Science	1762	515	1247	71%	45000



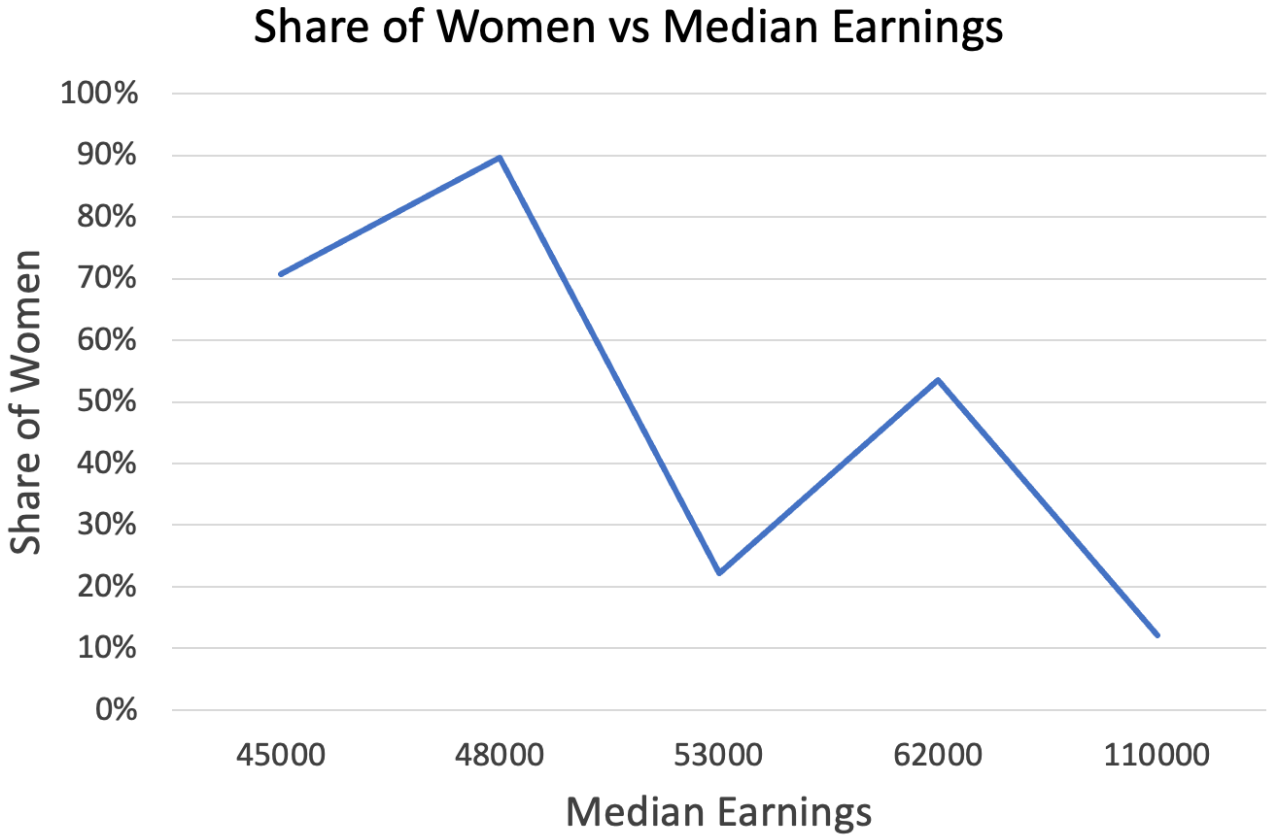
Data:
https://github.com/fivethirtyeight/data/blob/master/college-majors/women-stem.csv

Line charts

1. Goal → Trend
2. Data Types → Ordinal or Quantitative vs. Quantitative



Rank	Major_category	Total	Men	Women	Share_ women	Median_ earnings
1	Engineering	2339	2057	282	12%	110000
7	Physical Sciences	1792	832	960	54%	62000
19	Computers & Mathematics	128319	99743	28576	22%	53000
27	Health	209394	21773	187621	90%	48000
36	Biology & Life Science	1762	515	1247	71%	45000



Data:
https://github.com/fivethirtyeight/data/blob/master/college-majors/women-stem.csv

Avoiding Bias and Trickery

How do we avoid bias & trickery?

Inspect the data

→ Source?

→ Biases?

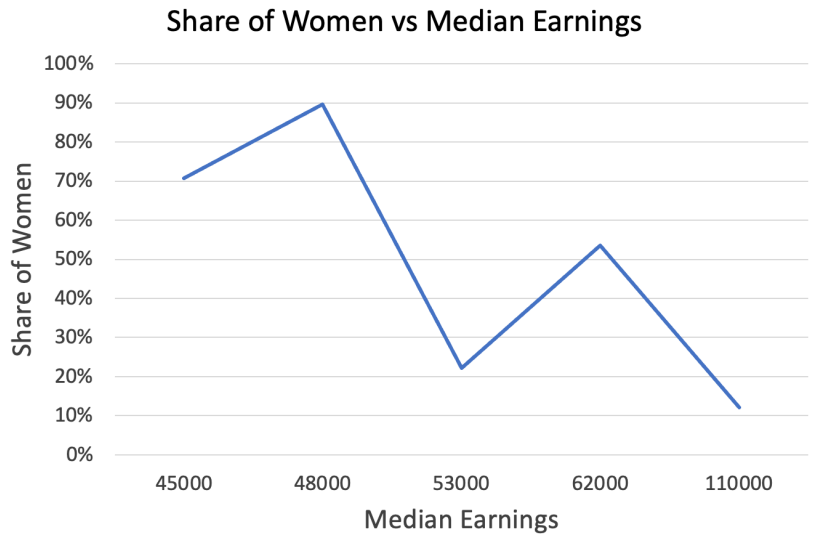
Rank	Major_category	Total	Men	Women	Share_ women	Median_ earnings
1	Engineering	2339	2057	282	12%	110000
7	Physical Sciences	1792	832	960	54%	62000
19	Computers & Mathematics	128319	99743	28576	22%	53000
27	Health	209394	21773	187621	90%	48000
36	Biology & Life Science	1762	515	1247	71%	45000

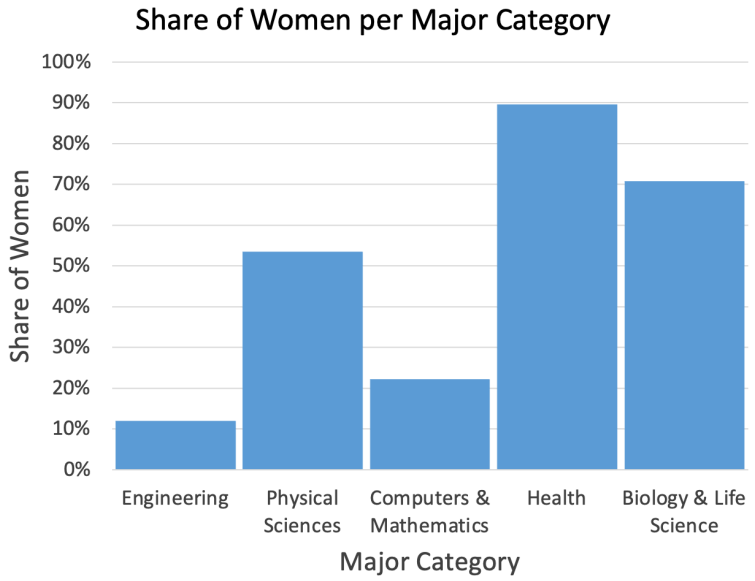
Data:
https://github.com/fivethirtyeight/data/blob/master/college-majors/women-stem.csv

How do we avoid bias & trickery?

Design Contentiously & Read Critically

→ What's shown vs not?





Data:

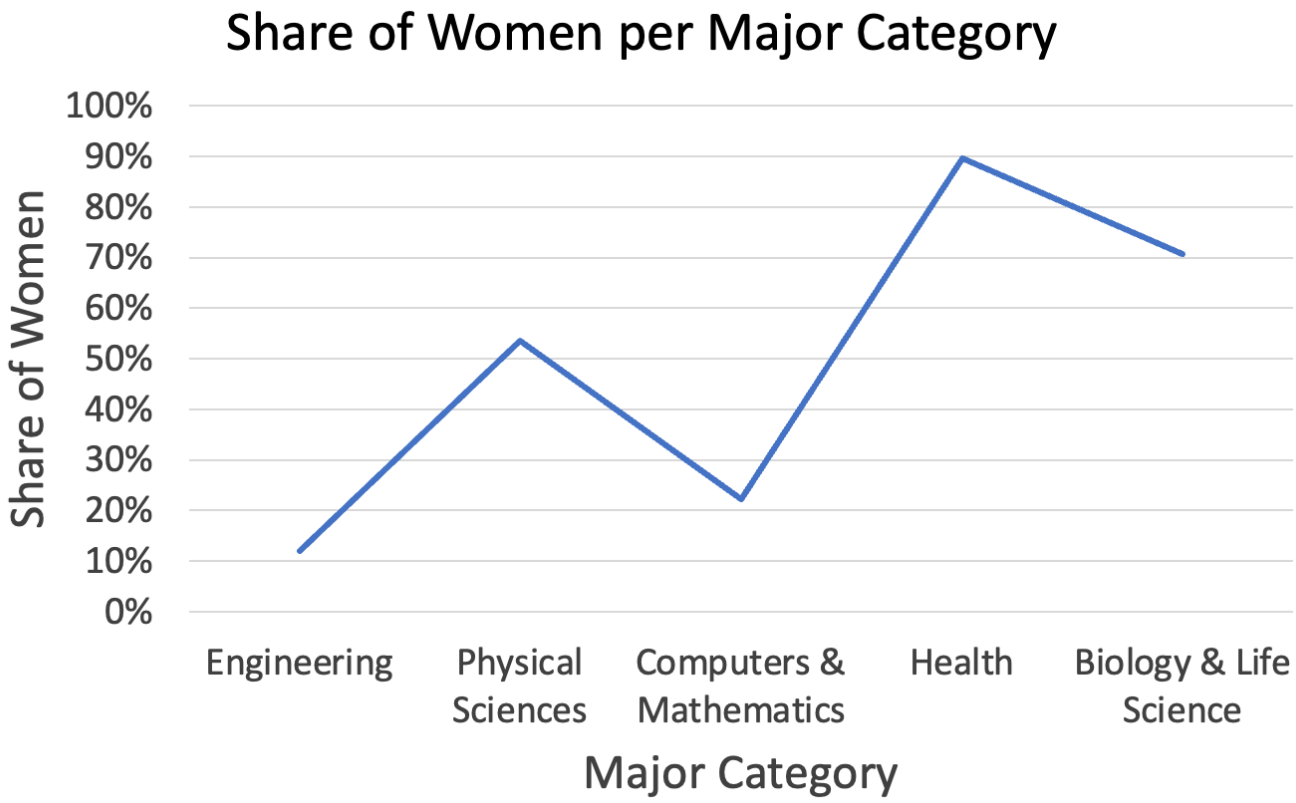
https://github.com/fivethirtyeight/data/blob/master/college-majors/women-stem.csv

How do we avoid bias & trickery?

Design Contentiously & Read Critically

→ Goal

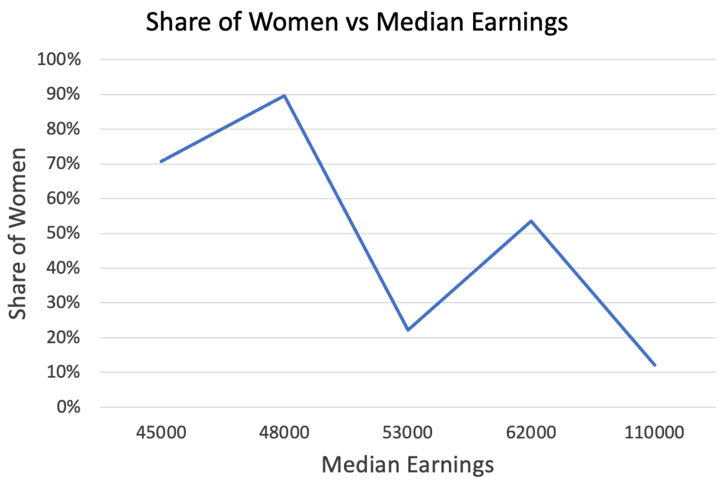
→ Data types

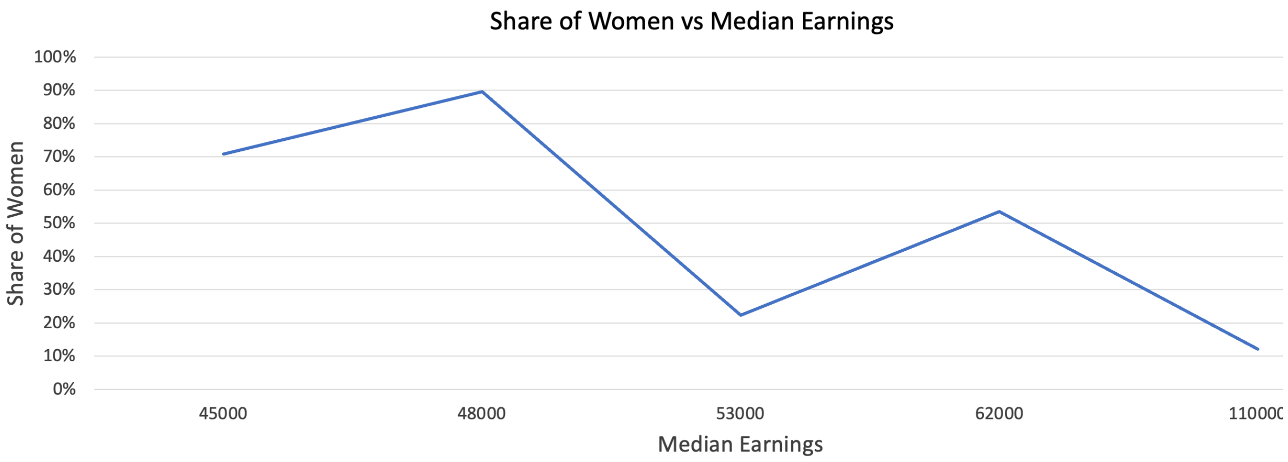


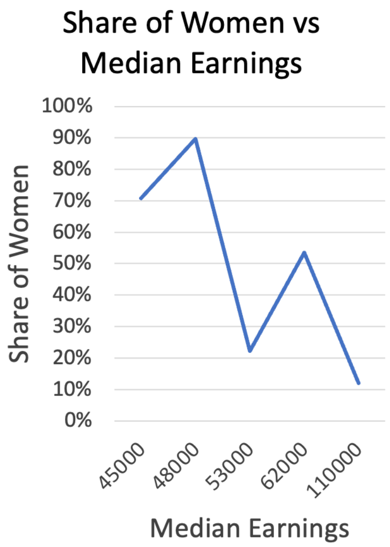
Data:
https://github.com/fivethirtyeight/data/blob/master/college-majors/women-stem.csv

How do we avoid bias & trickery?

Design Contentiously & Read Critically
→ Aspect ratio





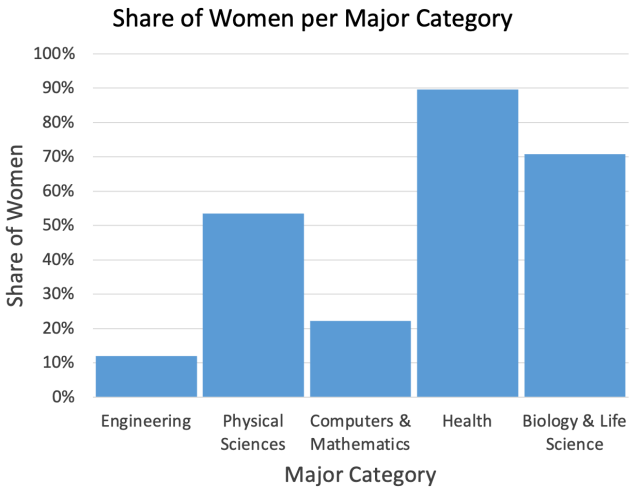


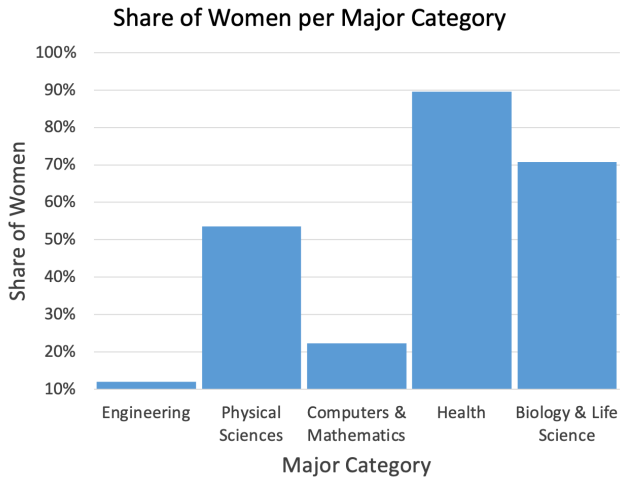
Data:
https://github.com/fivethirtyeight/data/blob/master/college-majors/women-stem.csv

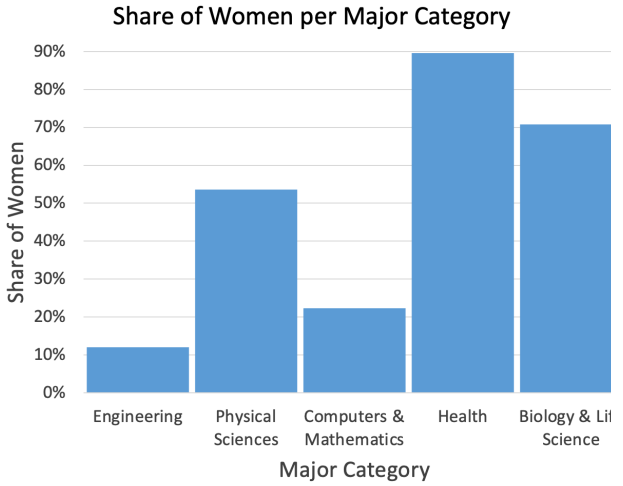
How do we avoid bias & trickery?

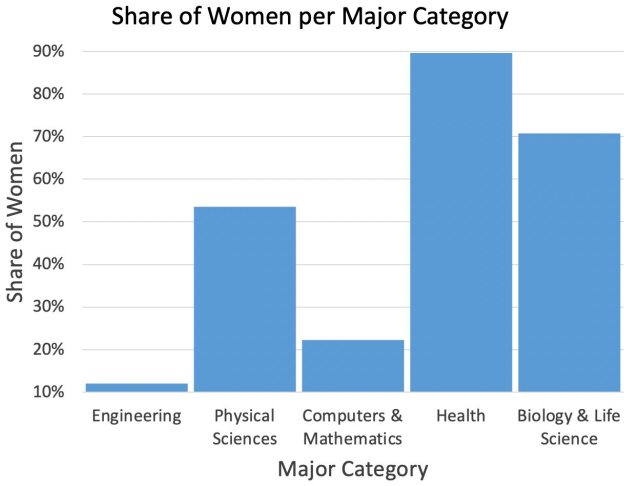
Design Contentiously & Read Critically

→ Axes





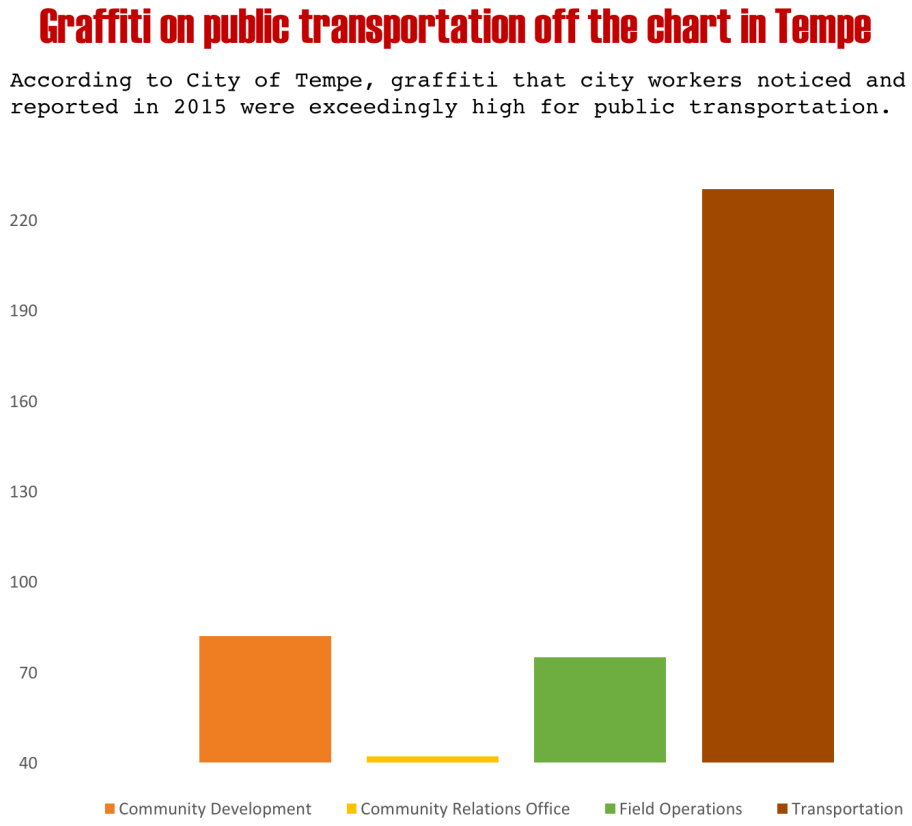




Data:
https://github.com/fivethirtyeight/data/blob/master/college-majors/women-stem.csv

Let’s practice

Take a critical look at this chart. Notice anything?



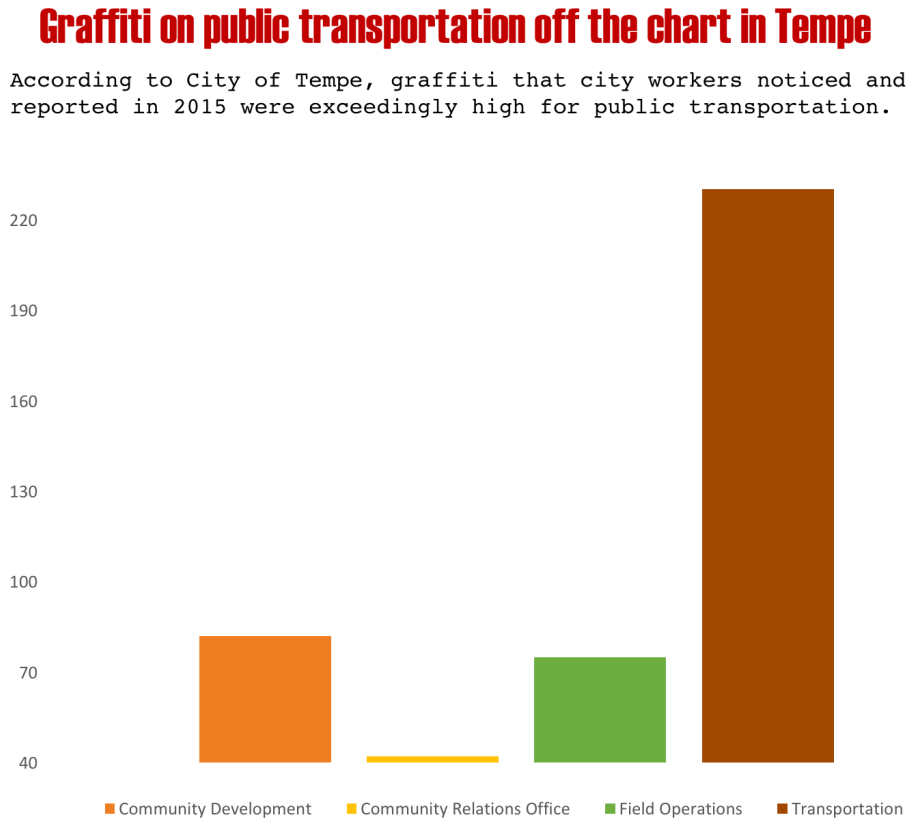
Source: City of Tempe, 2015

https://www.vislies.org/2017/gallery/

Let’s practice

Take a critical look at this chart. Notice anything?

Re-design the chart (you can add more data if you want)

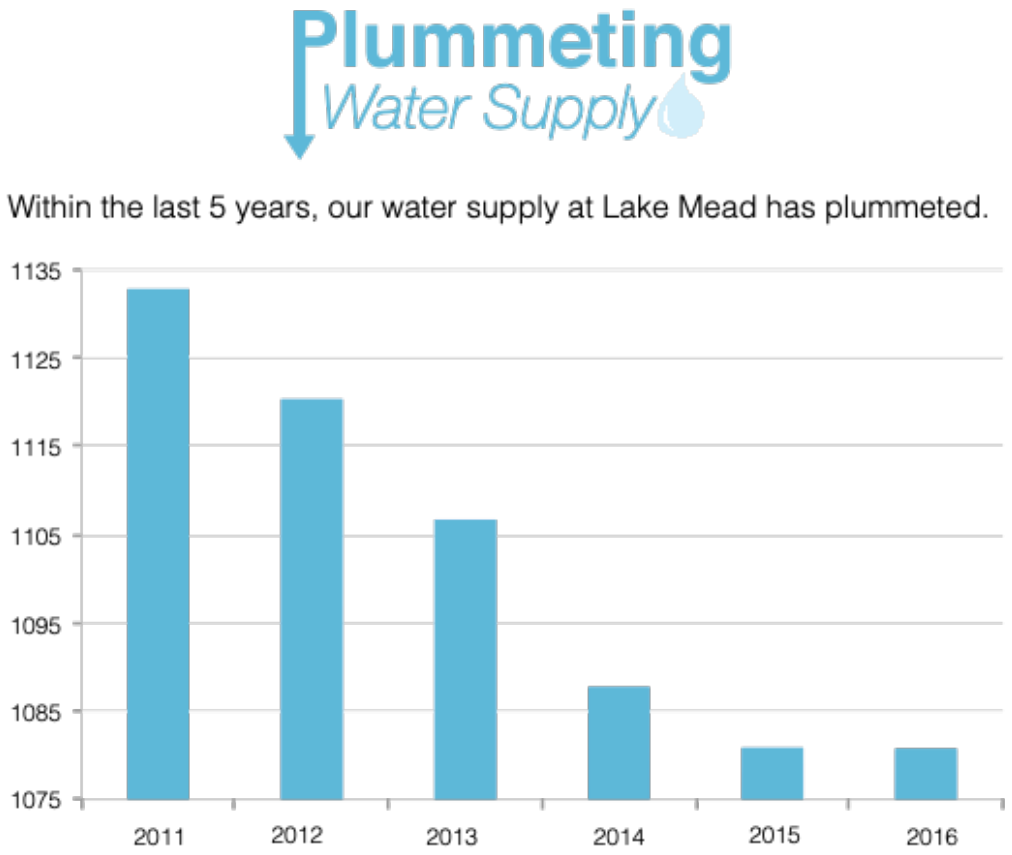


Source: City of Tempe, 2015

https://www.vislies.org/2017/gallery/

Let’s practice

Take a critical look at this chart. Notice anything?

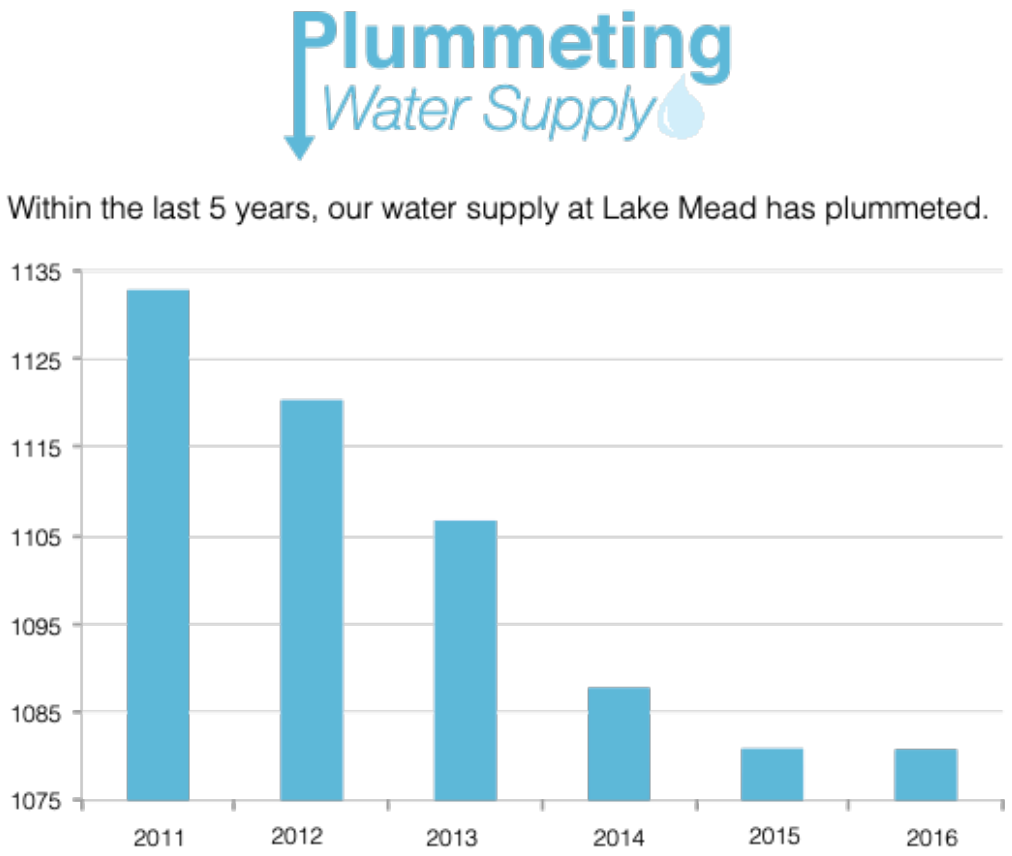


https://www.vislies.org/2017/gallery/

Let’s practice

Take a critical look at this chart. Notice anything?

Re-design the chart (you can add more data if you want)



https://www.vislies.org/2017/gallery/