Visualization Techniques

SSEP 2022 Morning Day 5

Dr. Ab Mosca (they/them)

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**
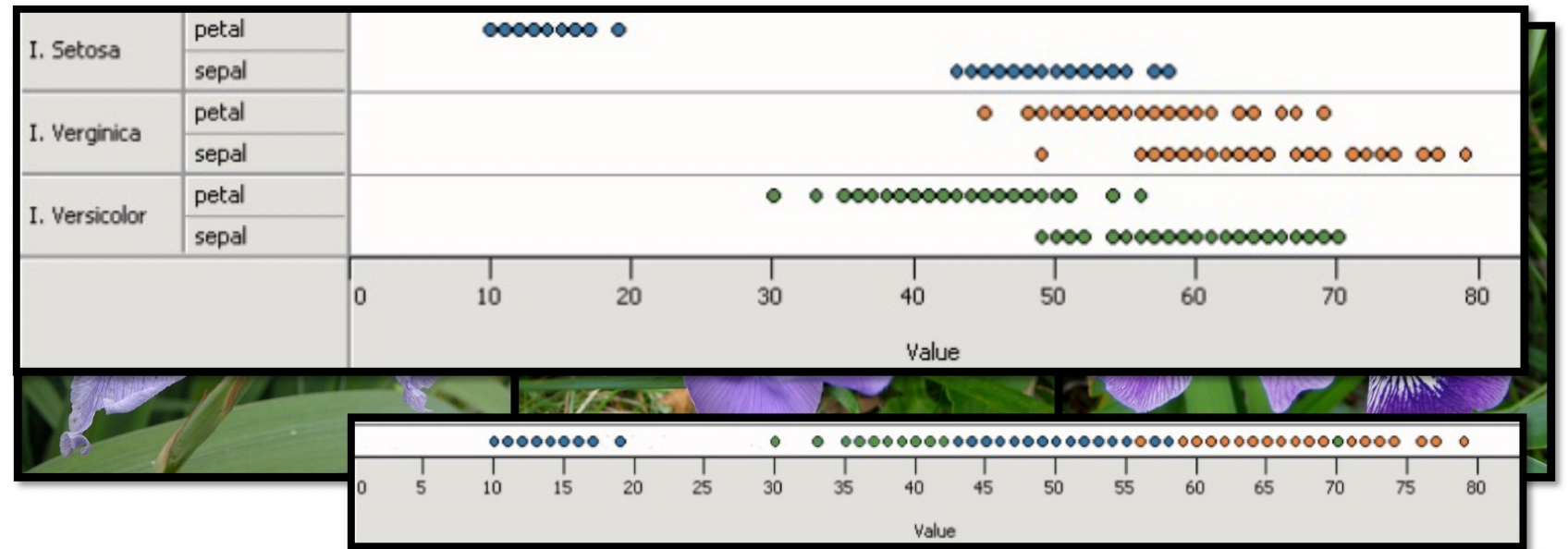
Channels

Marks

Lines

Points
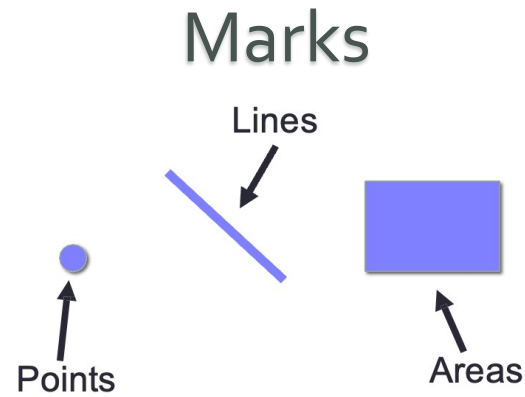
Areas

POSITION

SIZE

VALUE

COLOR

ORIENTATION

SHAPE

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

Mapping Data → Visuals

Principle 1: Expressiveness

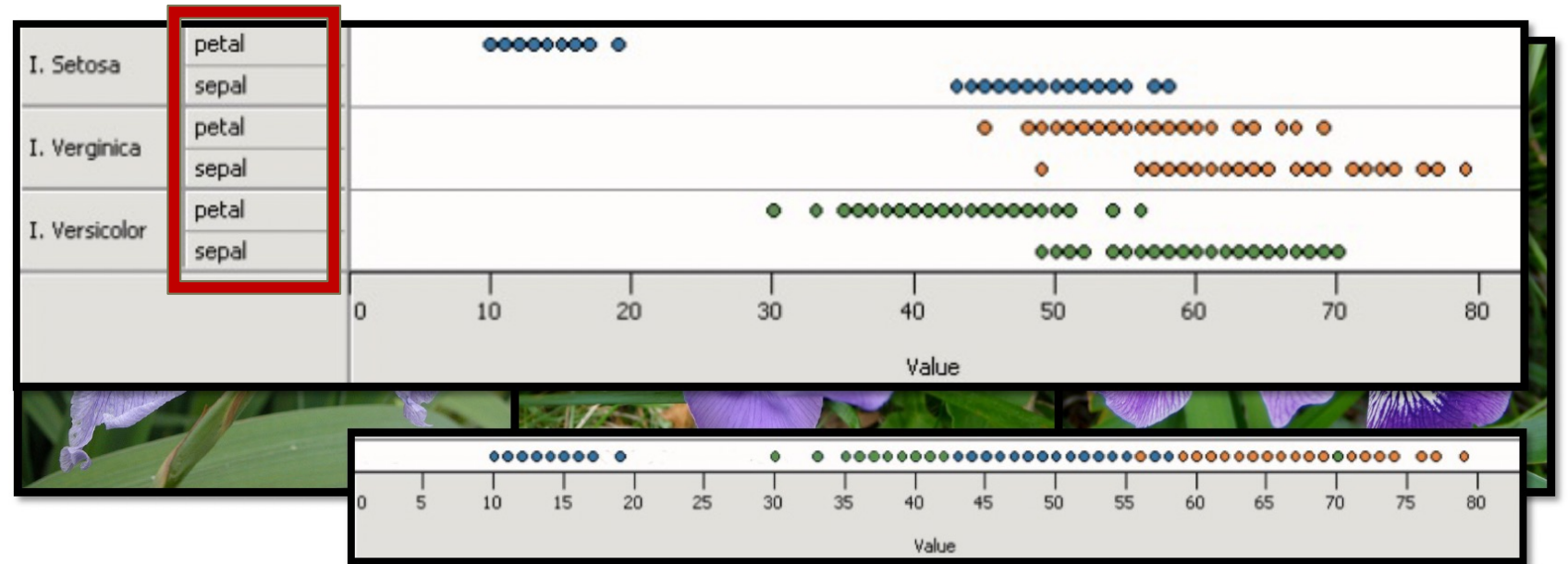*Encode **all the facts** and only the facts*

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

Mapping Data → Visuals

Principle 1: Expressiveness

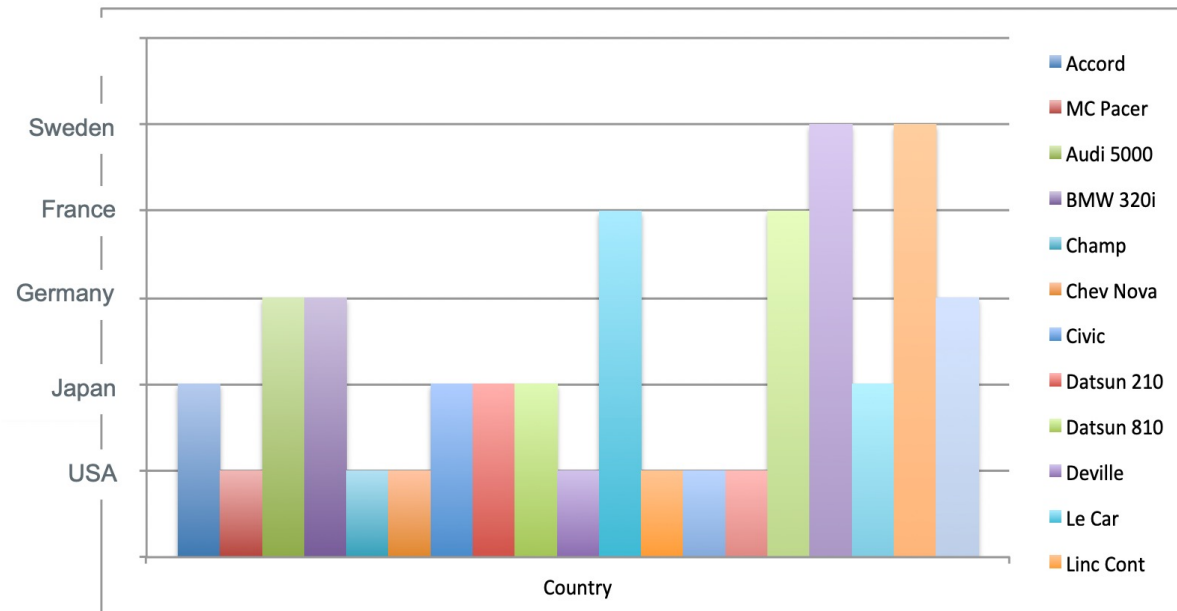*Encode **all the facts** and only the facts*

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

Fig. Courtesy of M Krzywinski

Mapping Data → Visuals

Principle 1: Expressiveness
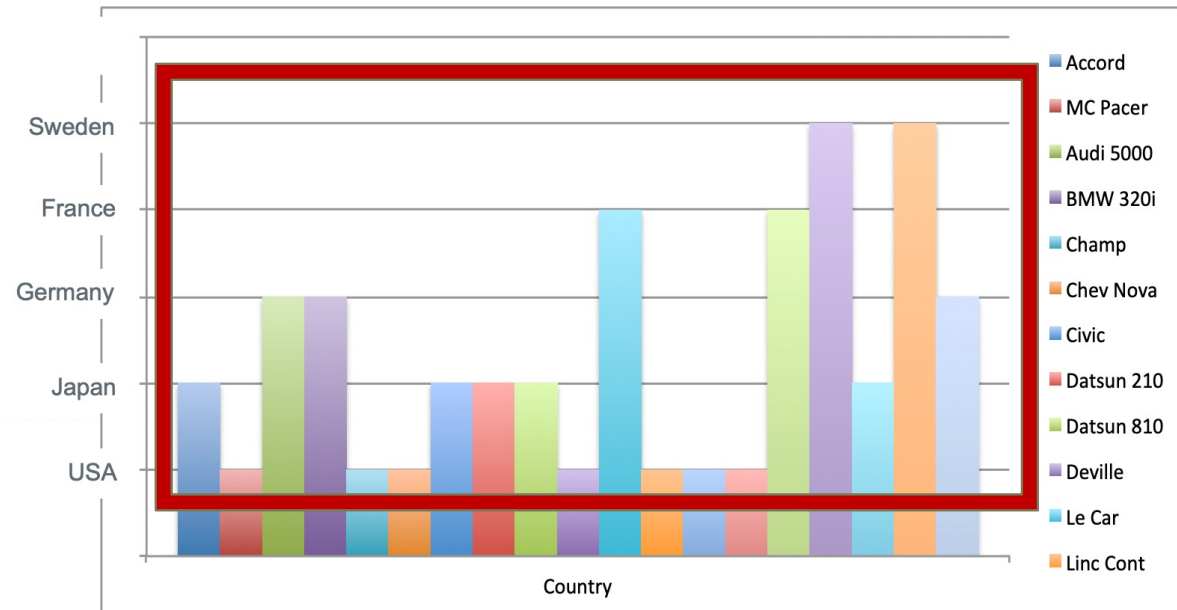
*Encode all the facts and **only the facts***

What "extra" data is included in this visualization?

Mapping Data → Visuals

Principle 1: Expressiveness
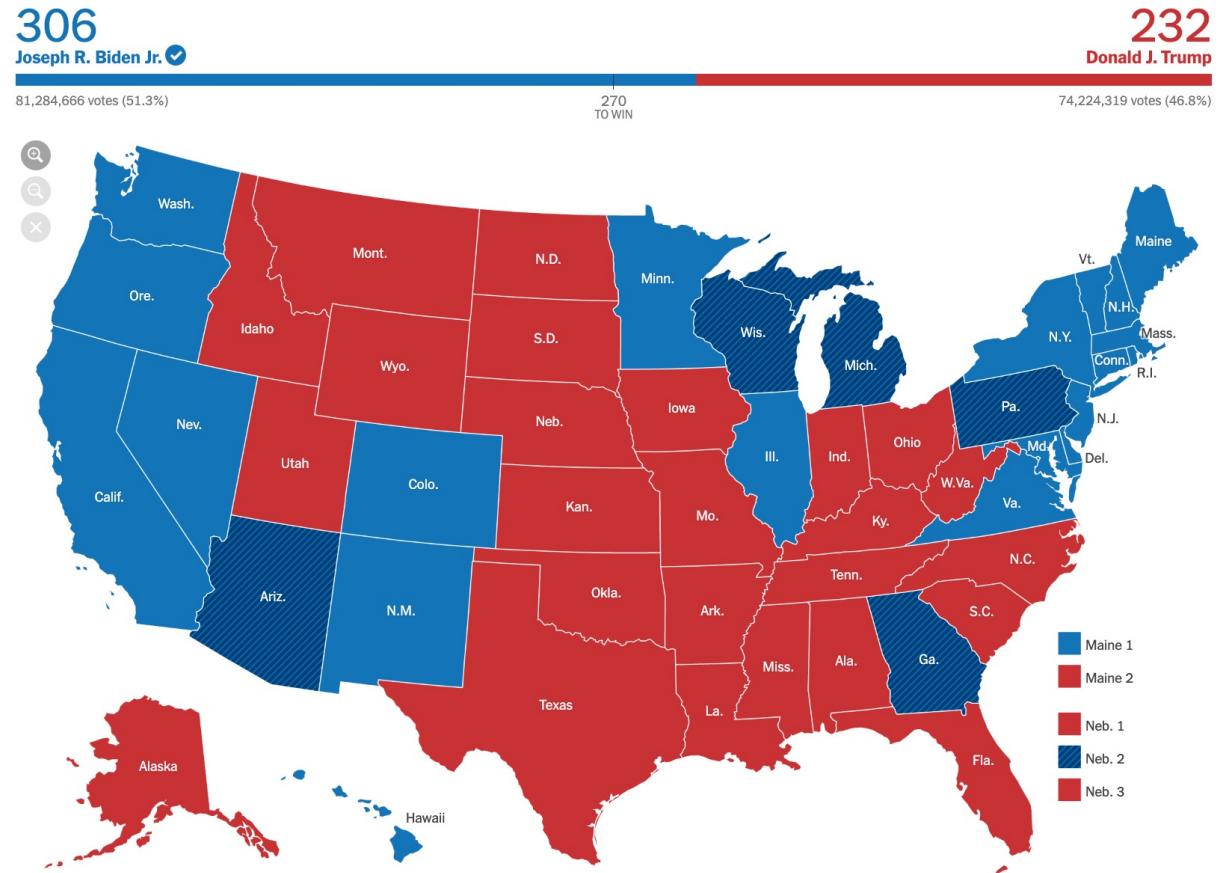
*Encode all the facts and **only the facts***

What "extra facts" are included in this visualization?

Principle 1: Expressiveness

Encode all the facts and only the facts

Mapping Data → Visuals

306
Joseph R. Biden Jr. ✓

232
Donald J. Trump

81,284,666 votes (51.3%)

270
TO WIN

74,224,319 votes (46.8%)

Maine 1
Maine 2
Neb. 1
Neb. 2
Neb. 3

What is wrong with this visualization?

Mapping Data → Visuals

Principle 1: Expressiveness

Encode all the facts and only the facts

Most *electoral votes* win

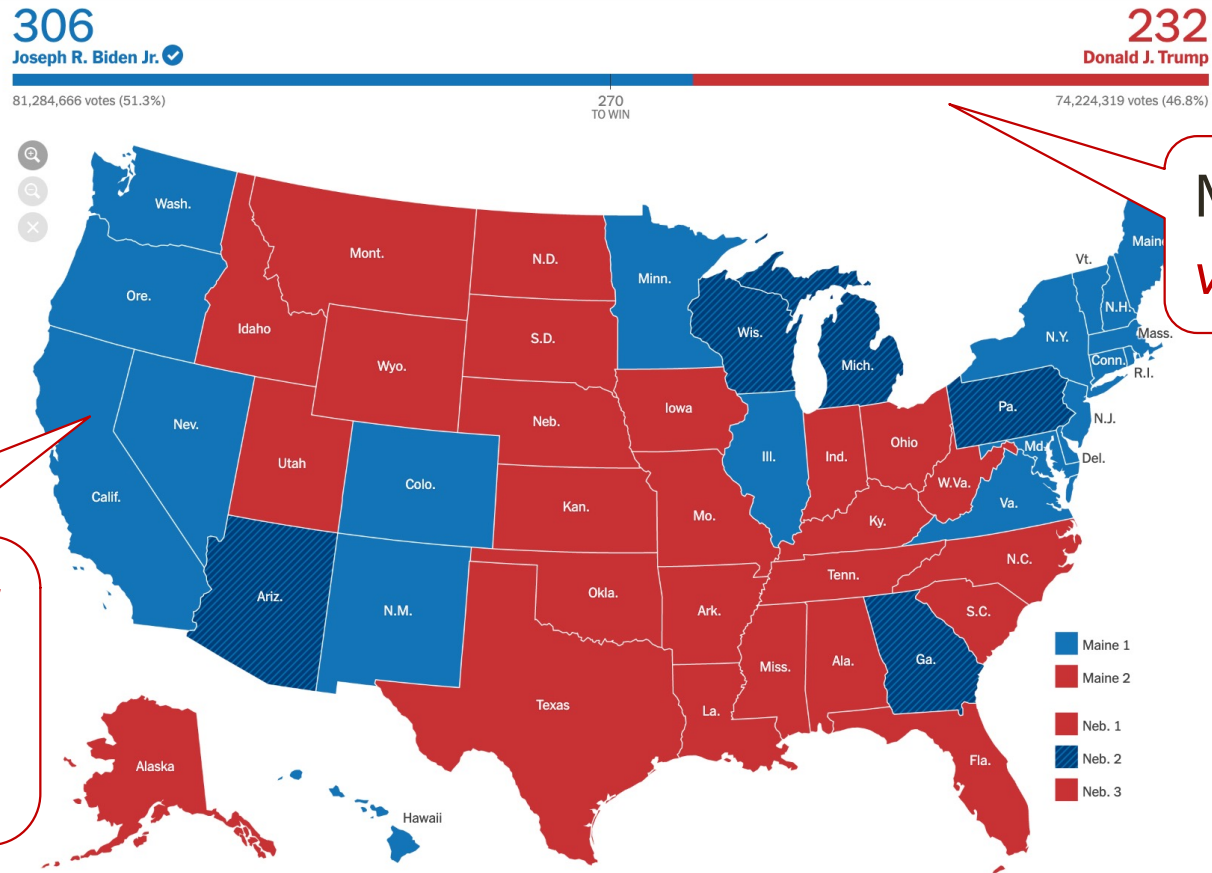Colors highlight *land area* per state, not electoral votes

What is wrong with this visualization?

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

Principle 1: Expressiveness

Encode all the facts and only the facts

Mapping Data → Visuals

306 Joseph R. Biden Jr. ✓
81,284,666 votes (51.3%)

270 TO WIN

232 Donald J. Trump
74,224,319 votes (46.8%)

Most *electoral votes* win

Colors highlight *electoral votes* per state
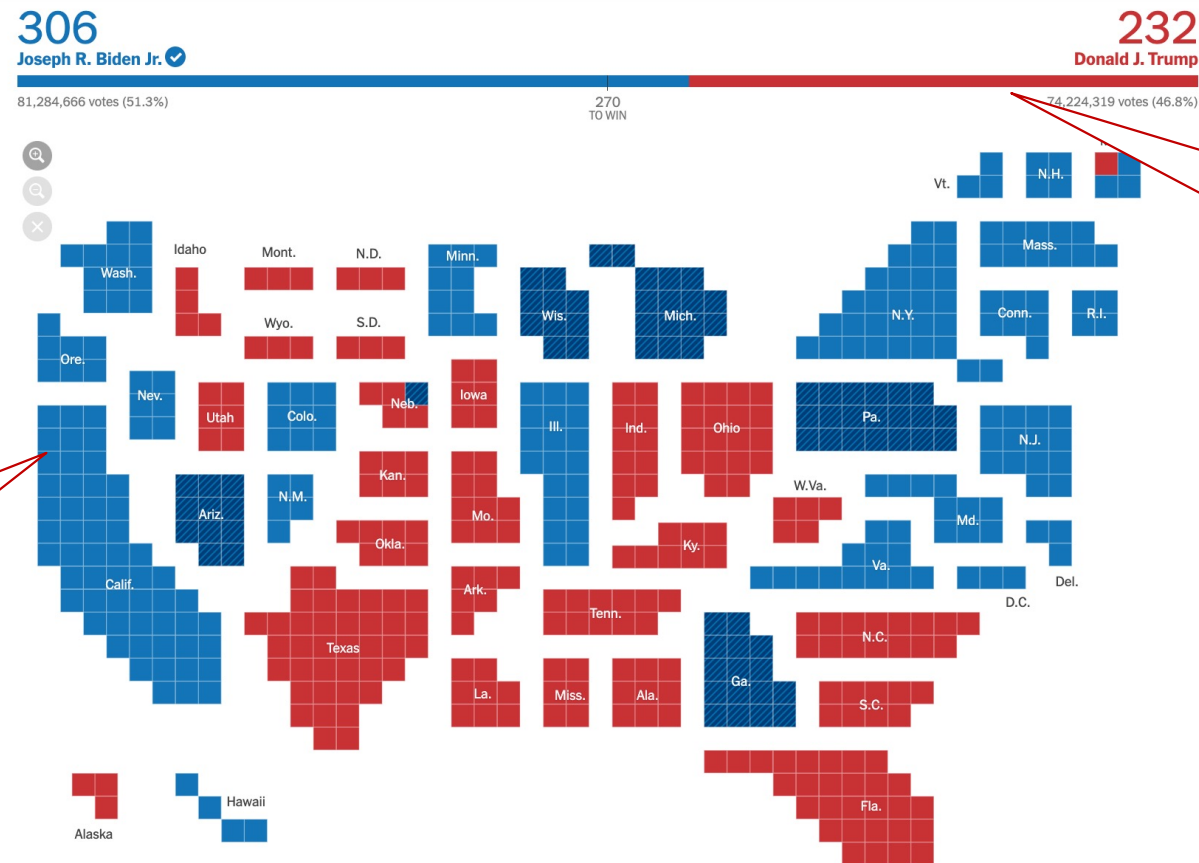
Mapping Data → Visuals

Principle 2: Effectiveness

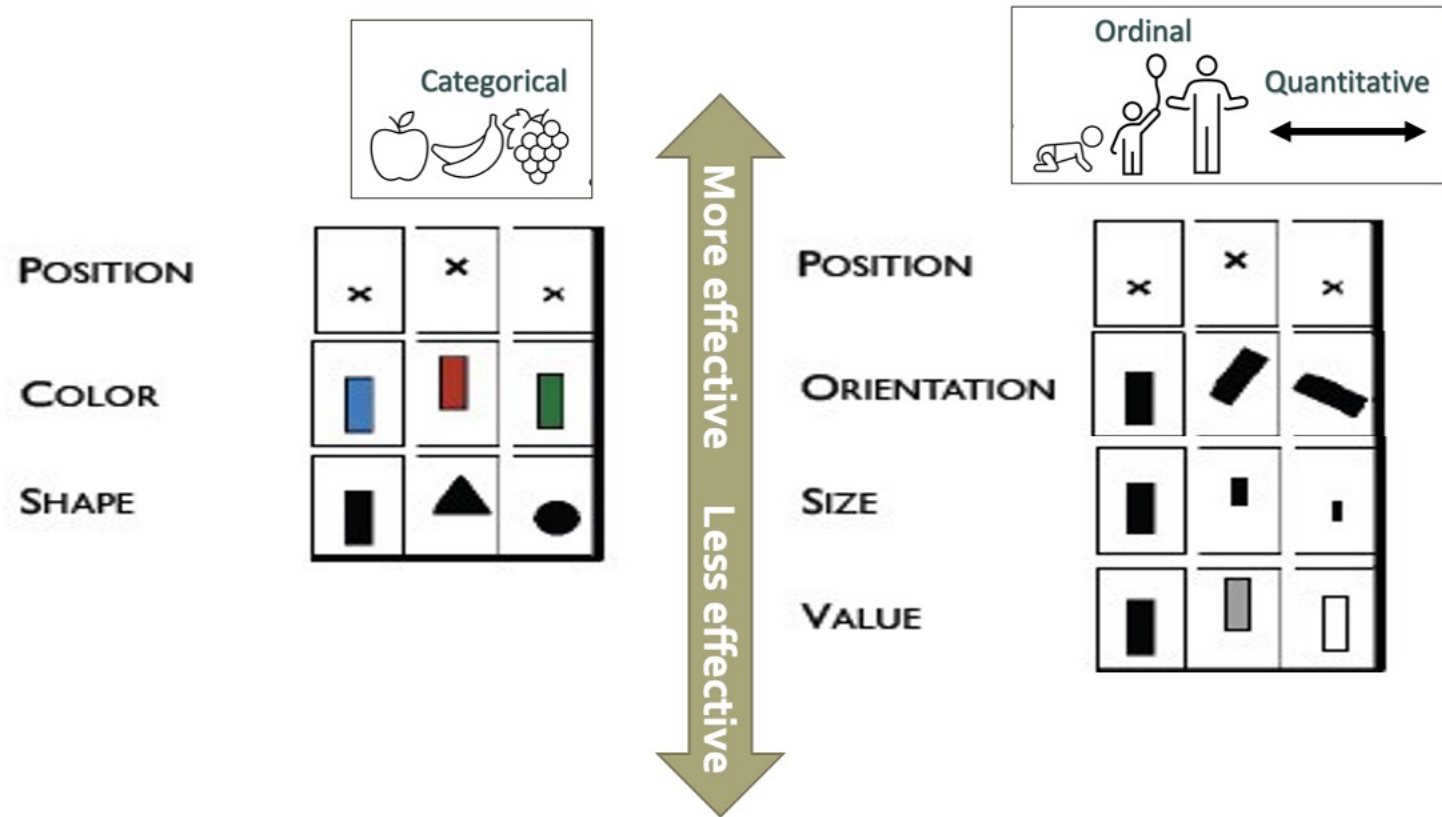*Most effective channels should be used for most important data*

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

- Accuracy
- Discriminability
- Separability
- Visual popout
- Grouping

Mapping Data → Visuals

Principle 2: Effectiveness

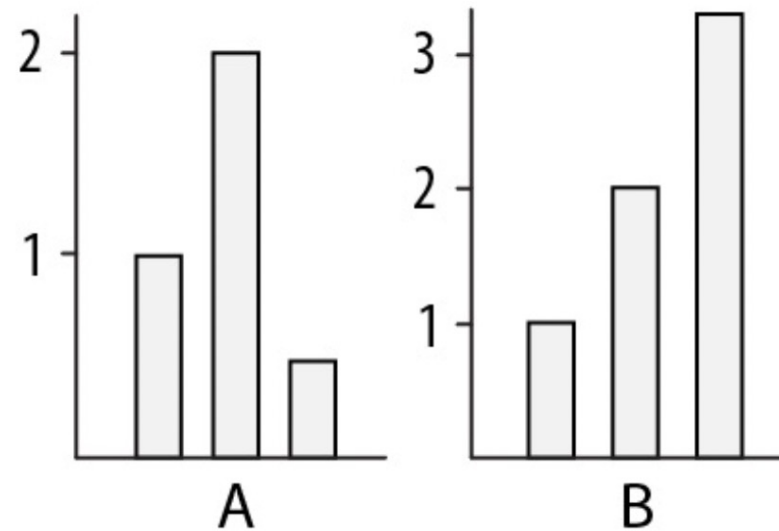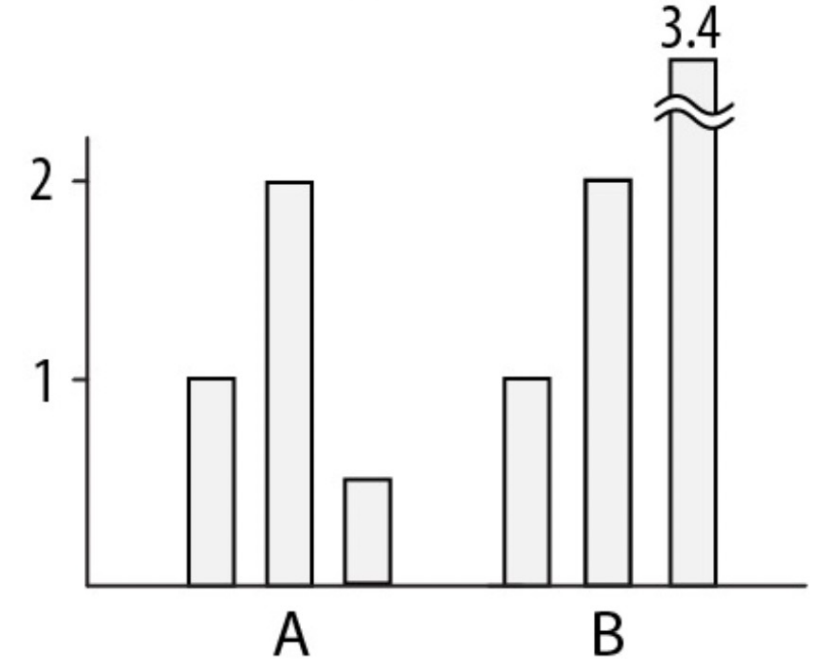*Most effective channels should be used for most important data*

Mapping Data → Visuals

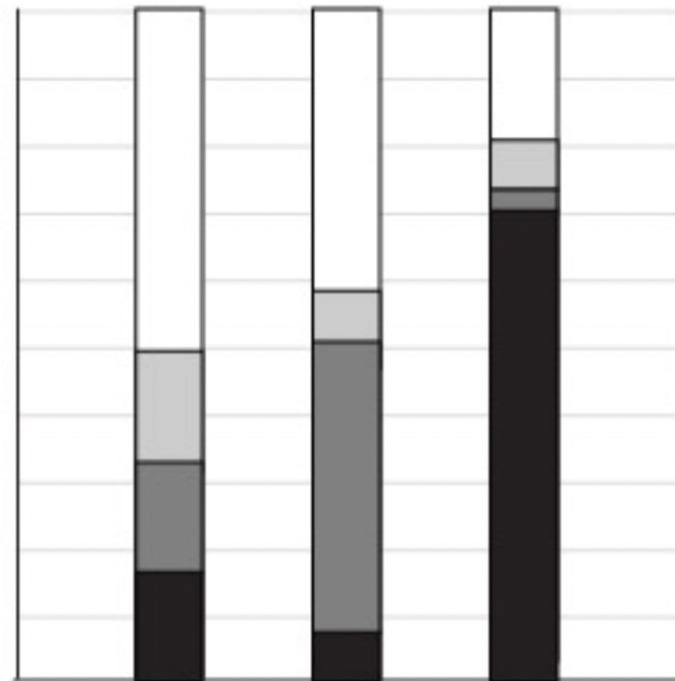## Principle 3: Consistency

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

consistent inconsistent

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

Mapping Data → Visuals

Principle 3: Separability

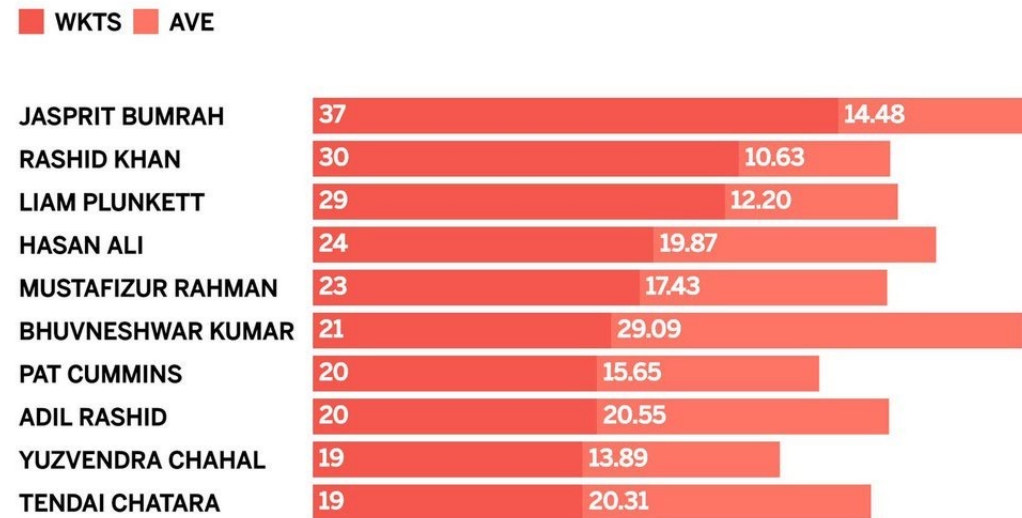*Avoid visually similar encodings for independent variables*
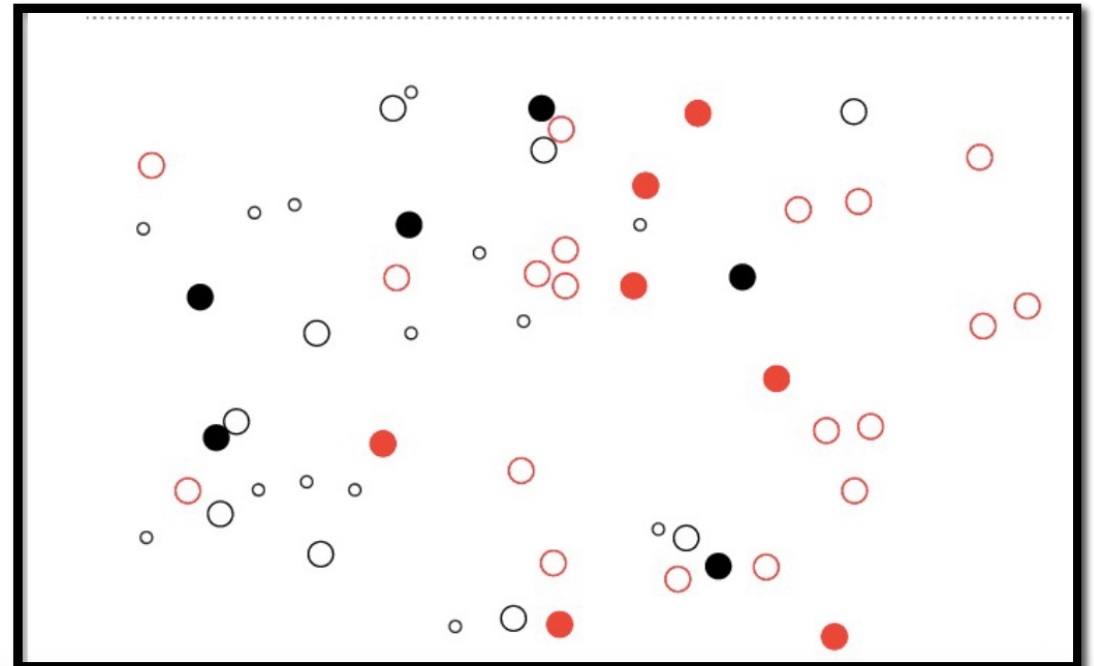
Mapping Data → Visuals

Principle 3: Separability

Avoid visually similar encodings for independent variables

MOST WICKETS IN DEATH OVERS IN ODIS

SINCE THE START OF JANUARY 2017

WKTS AVE

	WKTS	AVE
JASPRIT BUMRAH	37	14.48
RASHID KHAN	30	10.63
LIAM PLUNKETT	29	12.20
HASAN ALI	24	19.87
MUSTAFIZUR RAHMAN	23	17.43
BHUVNESHWAR KUMAR	21	29.09
PAT CUMMINS	20	15.65
ADIL RASHID	20	20.55
YUZVENDRA CHAHAL	19	13.89
TENDAI CHATARA	19	20.31

NUMBERS UPDATED TILL MAY 14, 2019

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

Mapping Data → Visuals
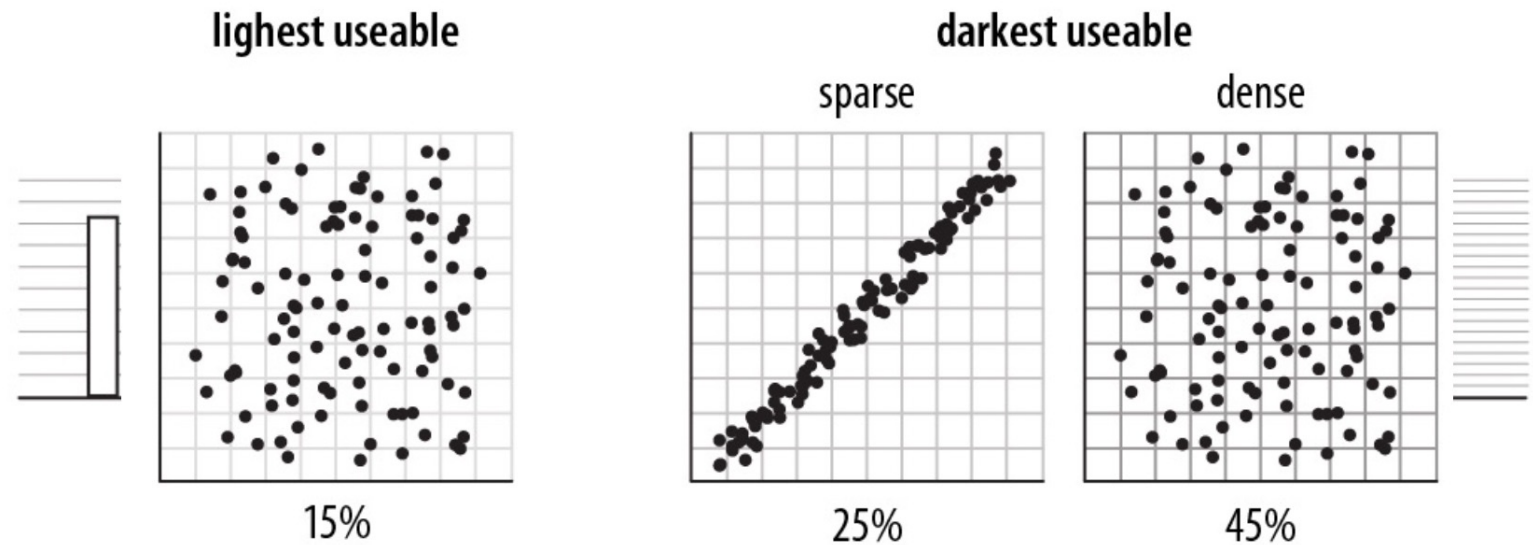
Principle 4: Simplicity

Avoid double encoding data
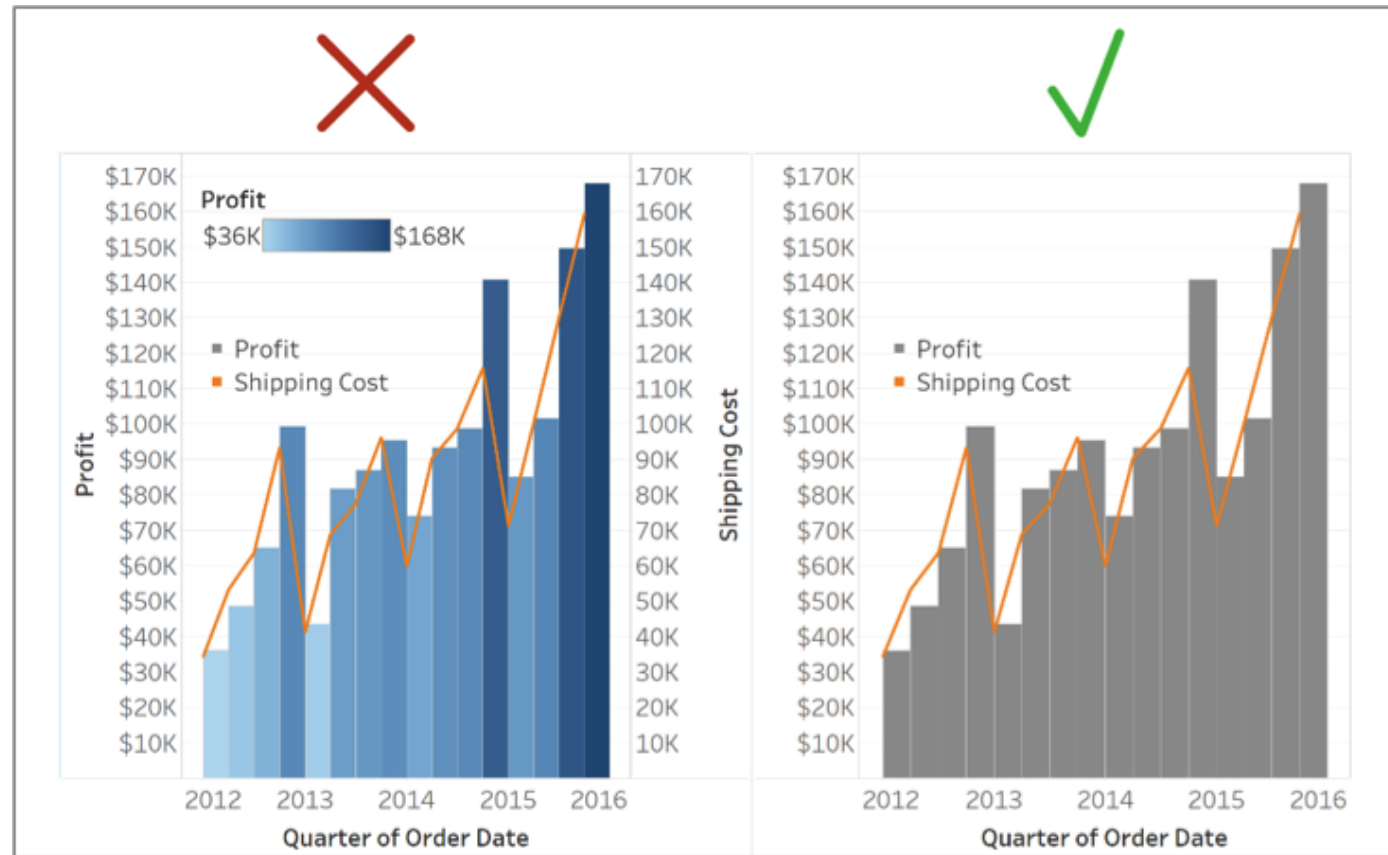
Mapping Data → Visuals
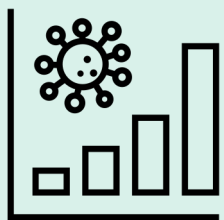
Principle 4: Simplicity

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.
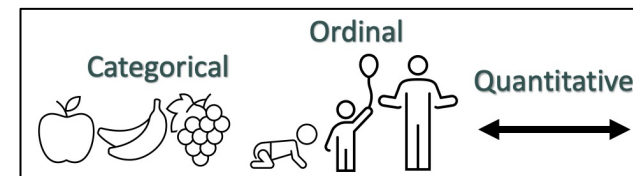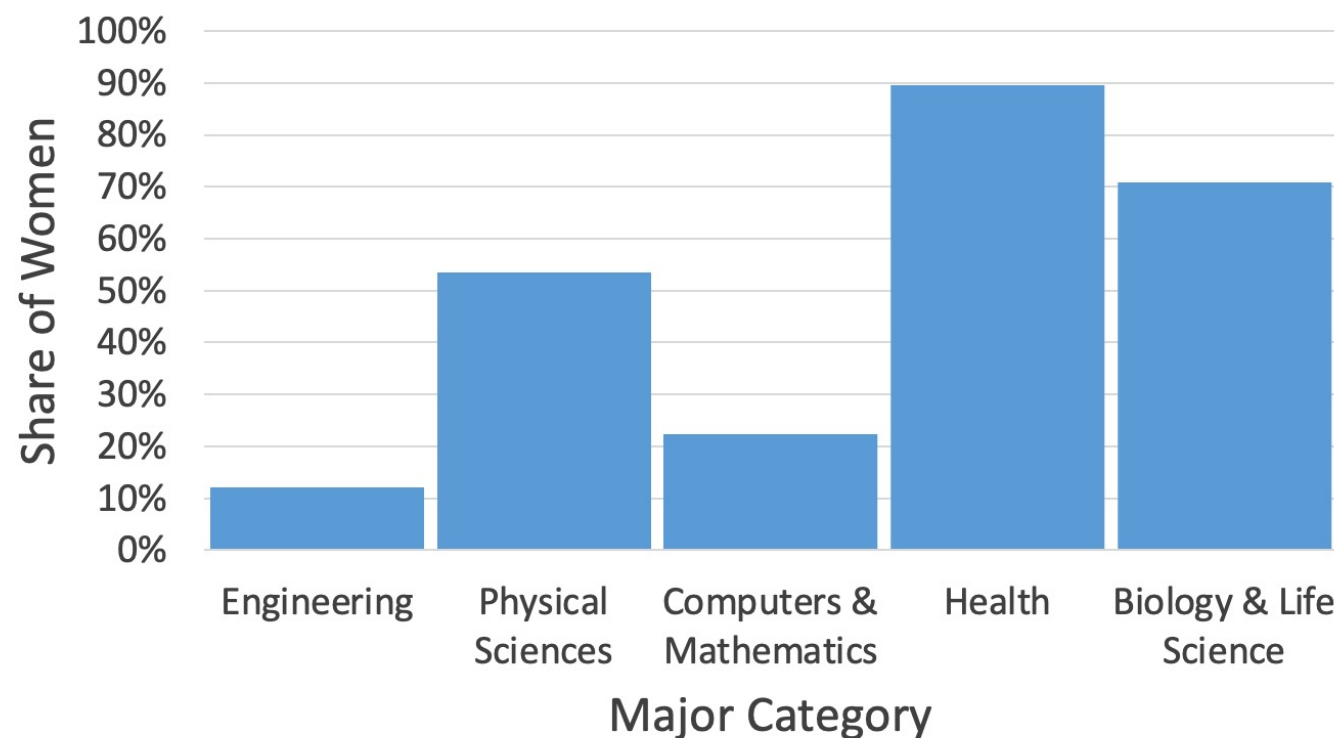
Common Visualizations

Bar charts

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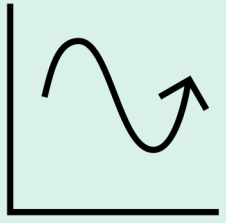2. Data Types → Categorical or Ordinal vs. Quantitative

Categorical   Ordinal   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

Share of Women per Major Category

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

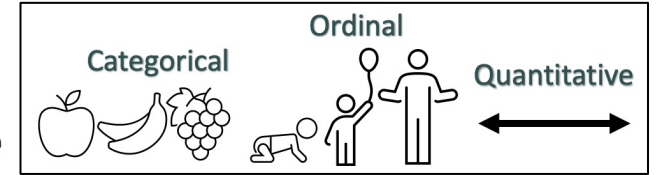**Share of Women vs Median Earnings**

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

Histogram

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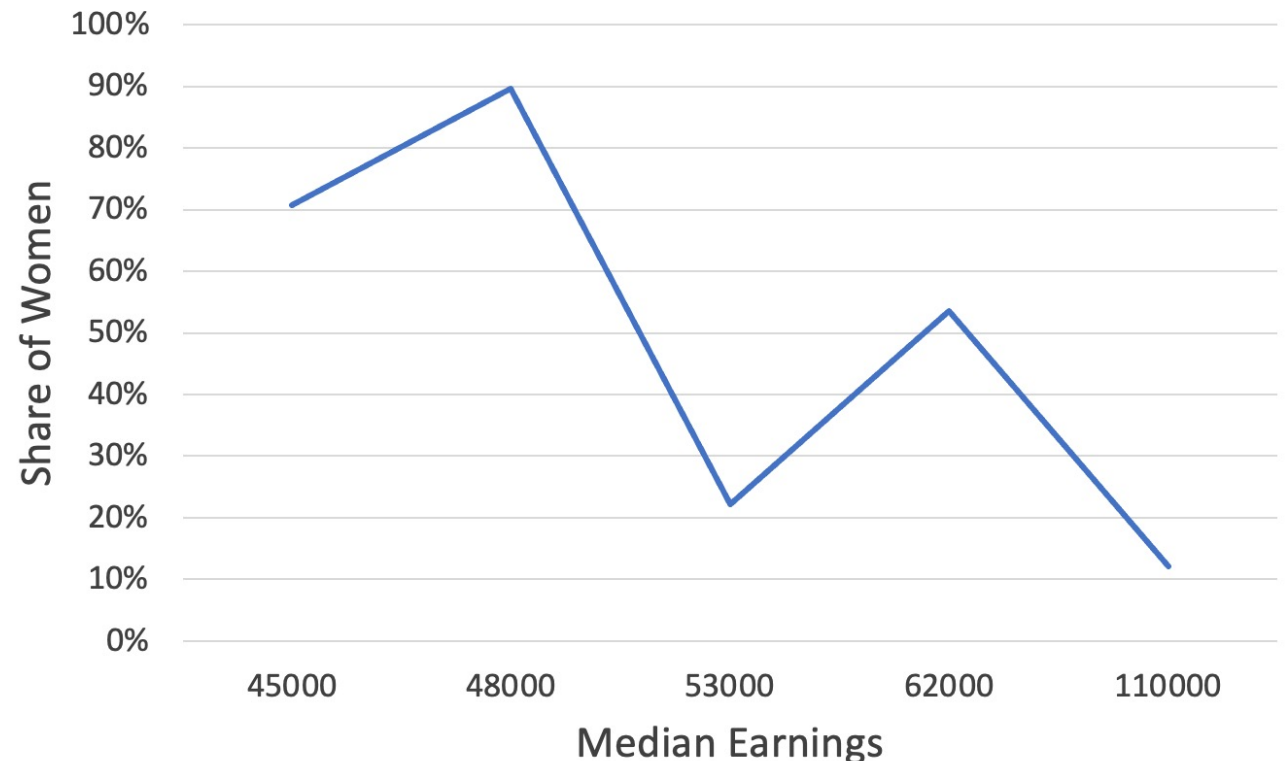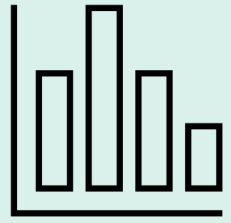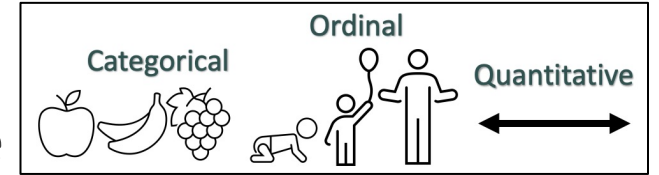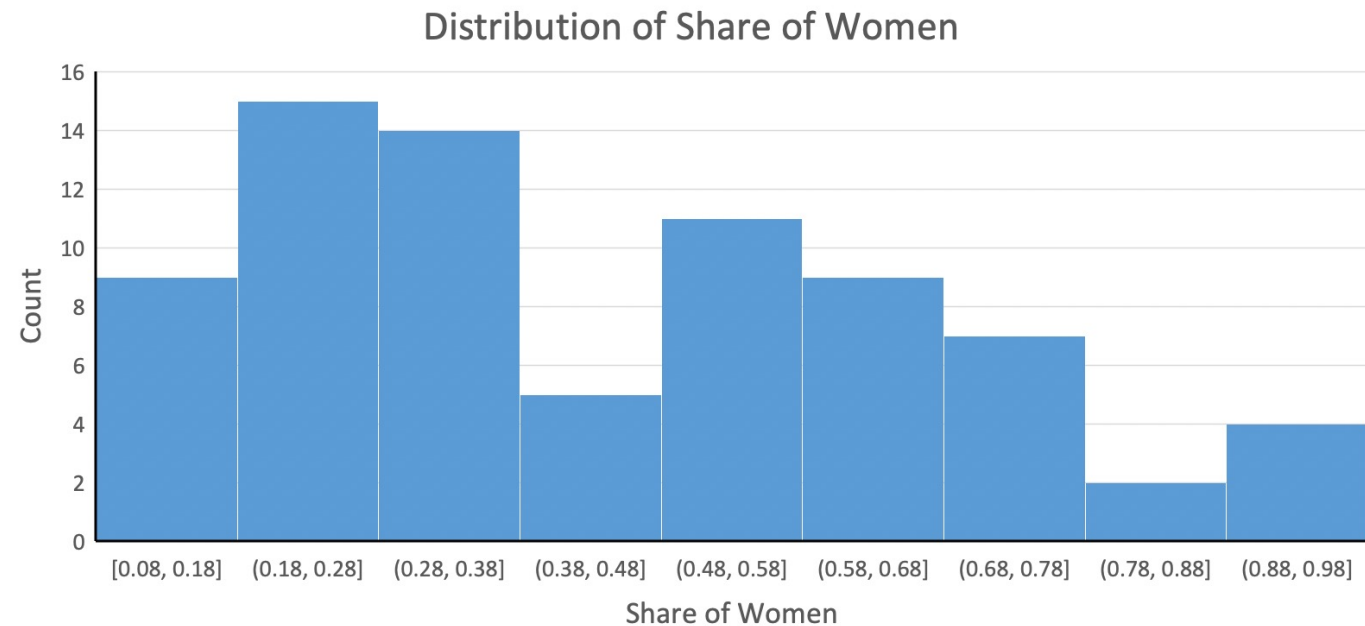2. Data Types → Ordinal or Quantitative vs. Quantitative

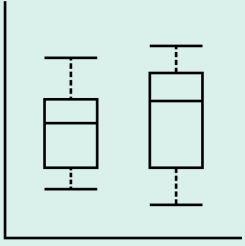| Rank	Major_category	Total	Men	Women	Share_women	Median_earnings
1	Engineering	2339	2057	282	12%	110000
7	Physical Sciences	1792	832	960	54%	62000
19	Computers & Mathematics	128319	99743	28576	22%	53000
27	Health	209394	21773	187621	90%	48000
36	Biology & Life Science	1762	515	1247	71%	45000

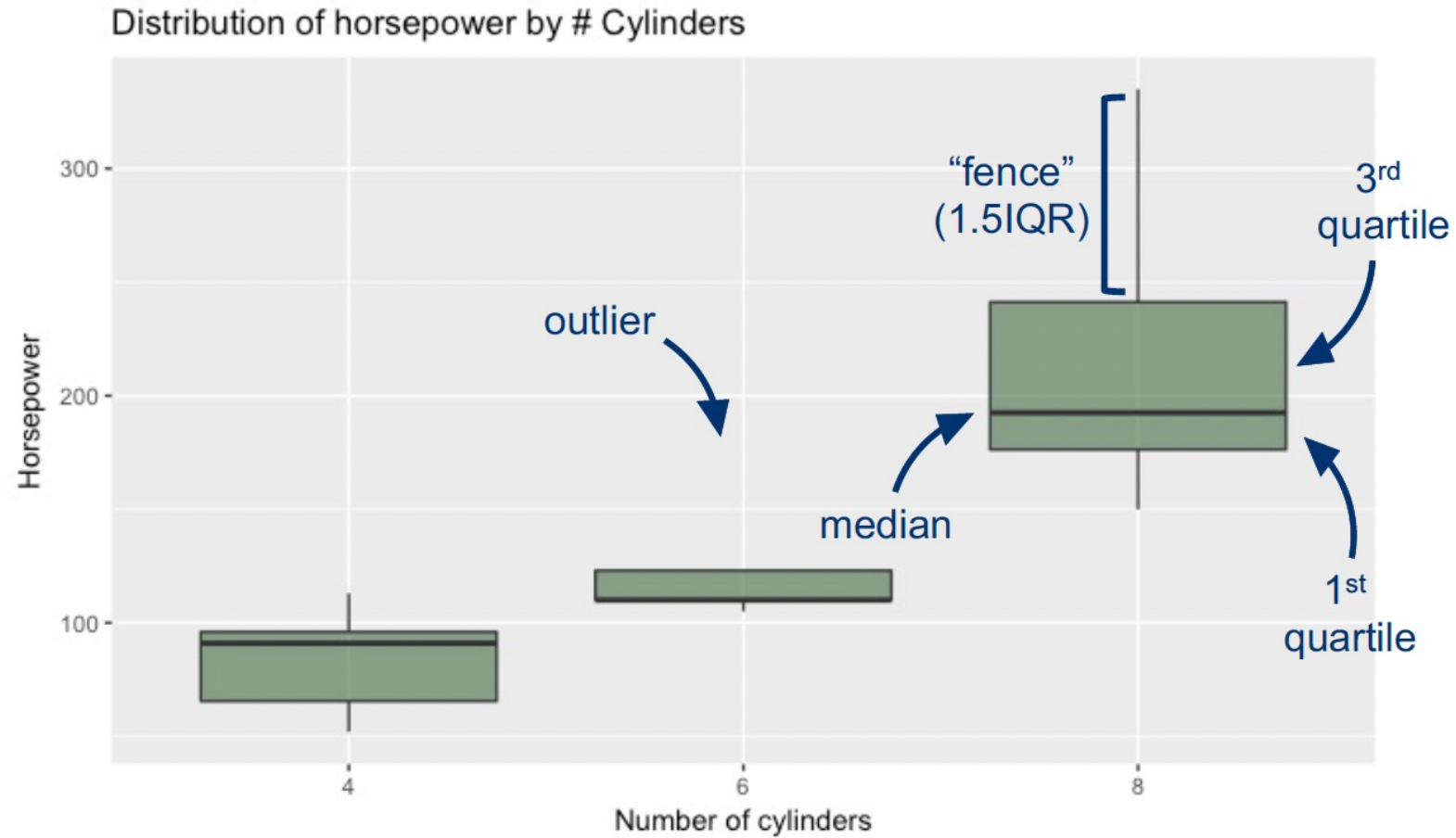### Distribution of Share of Women

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

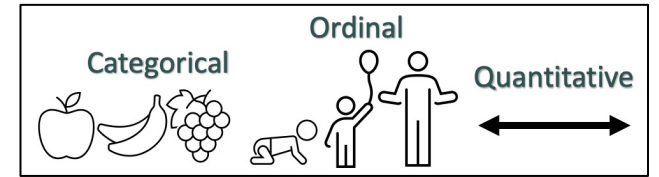# Boxplot

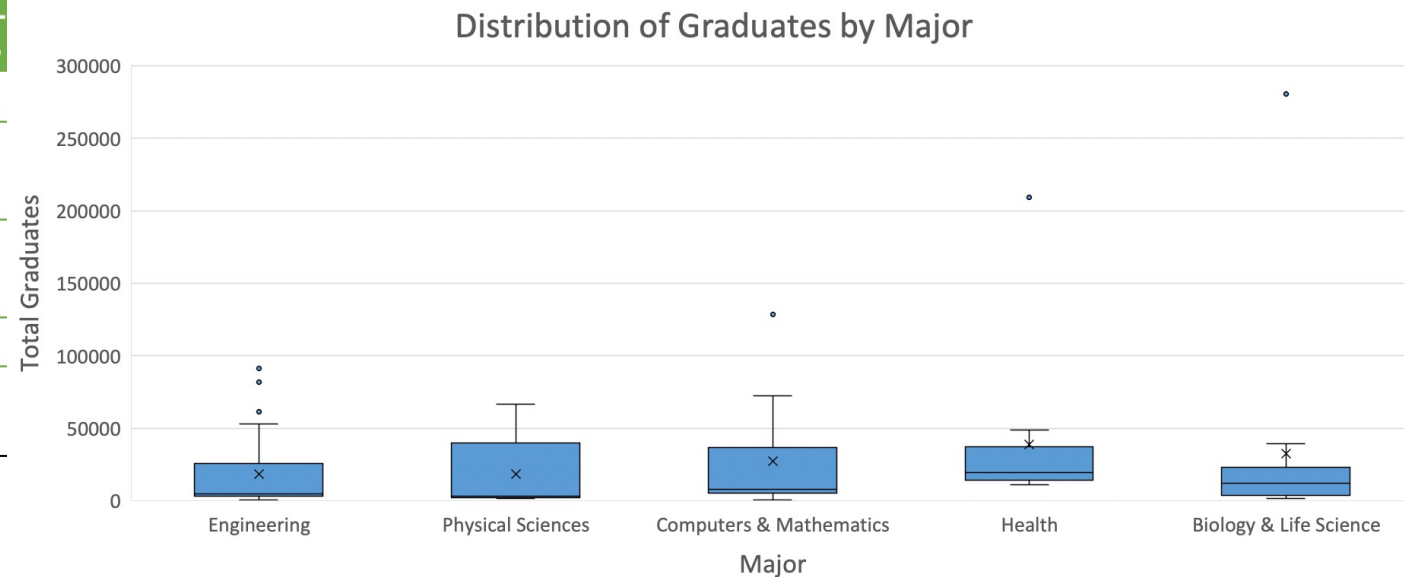Distribution of horsepower by # Cylinders

Boxplot

1. Goal → Distribution
2. Data Types → Ordinal or Categorical 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

Distribution of Graduates by Major

Data:

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

Common Visualizations

Break into groups of 2 – 3

Go to the Jamboard here: Visualization Examples

- Select one page of the Jamboard with a specific visualization to work on
- Add your names to that page so your classmates know it's taken
- Find your visualization here: https://datavizproject.com/
- On the Jamboard record the goal and data type(s) for that visualization
- Be prepared to share with the class
- If you finish early pick another visualization!