

# Introduction to Visualization

SSEP 2022 Morning Day 4

Dr. Ab Mosca (they/them)

Why?

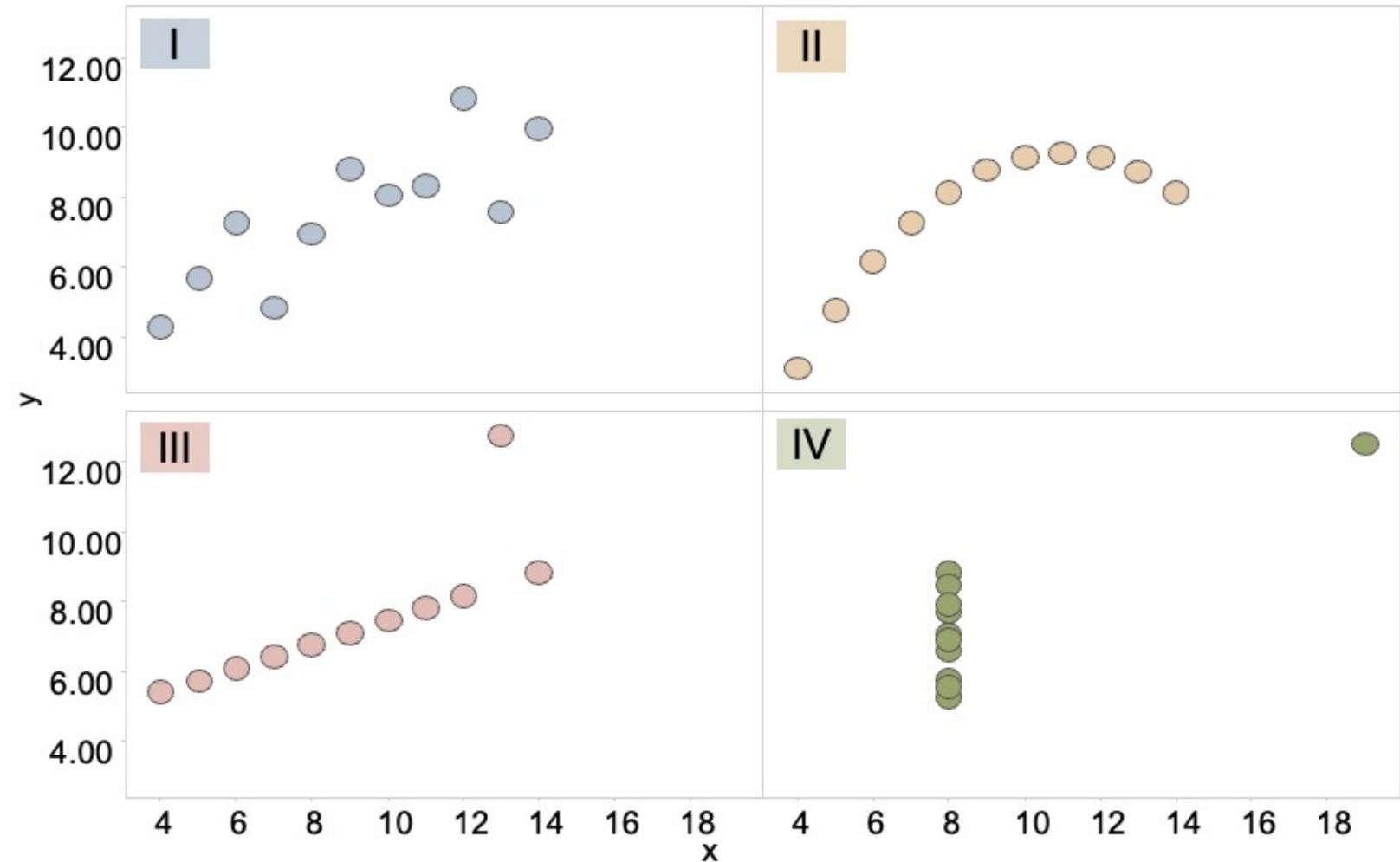
# Why visualize data?

I		II		III		IV	
x	y	x	y	x	y	x	y
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

# Why visualize data?

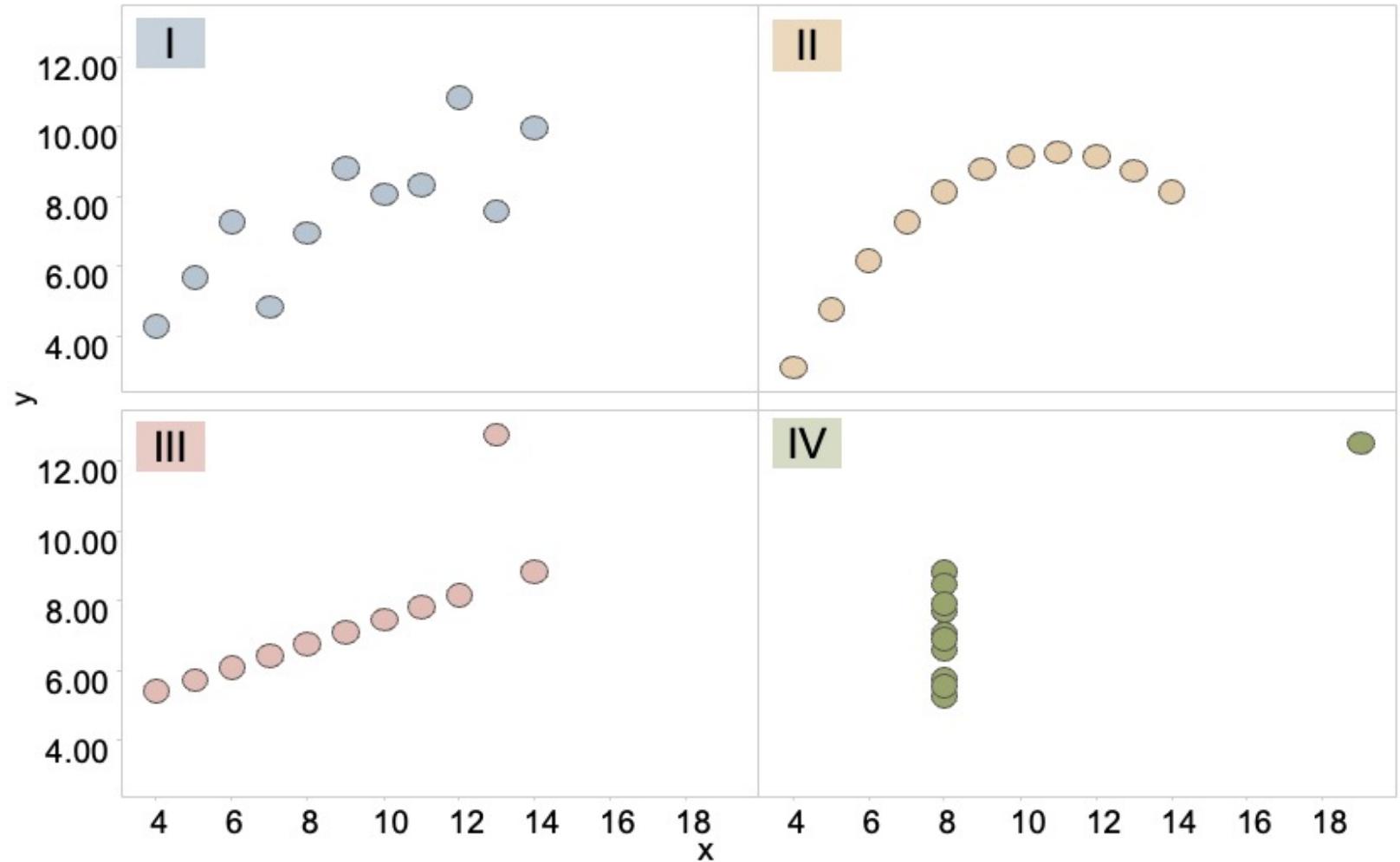
	I	II	III	IV					
	x	y	x	y	x	y	x	y	
Data	10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58	
	8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76	
	13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71	
	9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84	
	11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47	
	14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04	
	6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25	
	4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50	
	12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56	
	7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91	
Summary	5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89	
	Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
	Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816		

# Why visualize data?



# Why visualize data?

Reveal  
Patterns



# Why visualize data?

On the next slide, count  
the number of “t”s as  
fast as you can

# Why visualize data?

q t f j n i x i g j u n a s b b t g  
r k c l b v t x j x z x m x r g k  
l x q h m z y w t e y j w n o p

# Why visualize data?

q t f j n i x i g j u n a s b b t g  
r k c l b v t x j x z x m x r g k  
l x q h m z y w t e y j w n o p

# Why visualize data?

On the next slide, count  
the number of “j”s as  
fast as you can

# Why visualize data?

q t f **j** n i x i g **j** u n a s b b t g  
r k c l b v t x **j** x z x m x r g k  
l x q h m z y w t e y **j** w n o p

# Why visualize data?

Leverage Human  
Perception

q t f j n i x i g j u n a s b b t g  
r k c l b v t x j x z x m x r g k  
l x q h m z y w t e y j w n o p

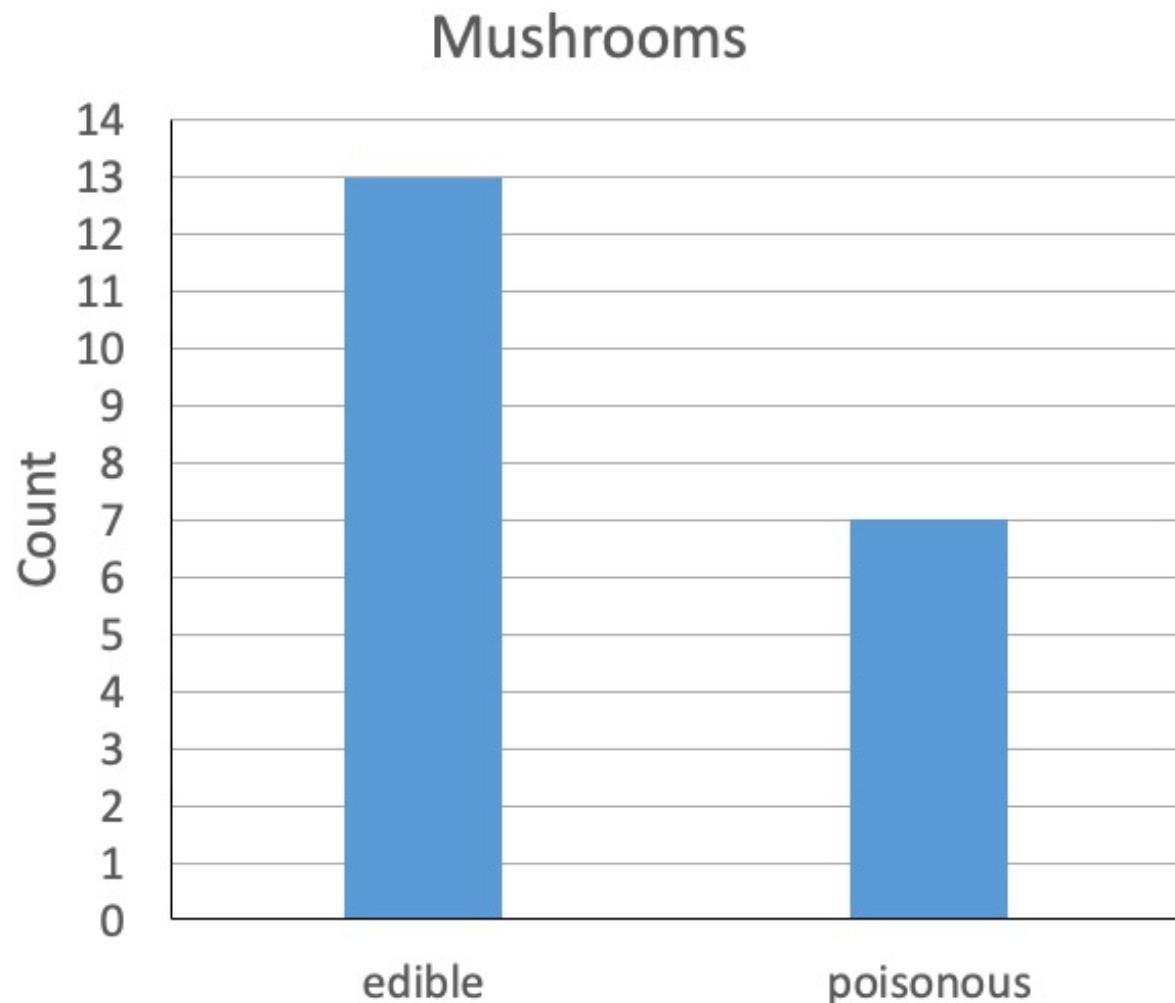
# Why visualize data?

How many mushrooms are poisonous?

class	cap-shape	cap-surface	cap-color	bruises	odor
poisonous	convex	smooth	brown	yes	pungent
edible	convex	smooth	yellow	yes	almond
edible	bell	smooth	white	yes	anise
poisonous	convex	scaly	white	yes	pungent
edible	convex	smooth	gray	no	none
edible	convex	scaly	yellow	yes	almond
edible	bell	smooth	white	yes	almond
edible	bell	scaly	white	yes	anise
poisonous	convex	scaly	white	yes	pungent
edible	bell	smooth	yellow	yes	almond
edible	convex	scaly	yellow	yes	anise
edible	convex	scaly	yellow	yes	almond
edible	bell	smooth	yellow	yes	almond
poisonous	convex	scaly	white	yes	pungent
edible	convex	fibrous	brown	no	none
edible	sunken	fibrous	gray	no	none
edible	flat	fibrous	white	no	none
poisonous	convex	smooth	brown	yes	pungent
poisonous	convex	scaly	white	yes	pungent
poisonous	convex	smooth	brown	yes	pungent

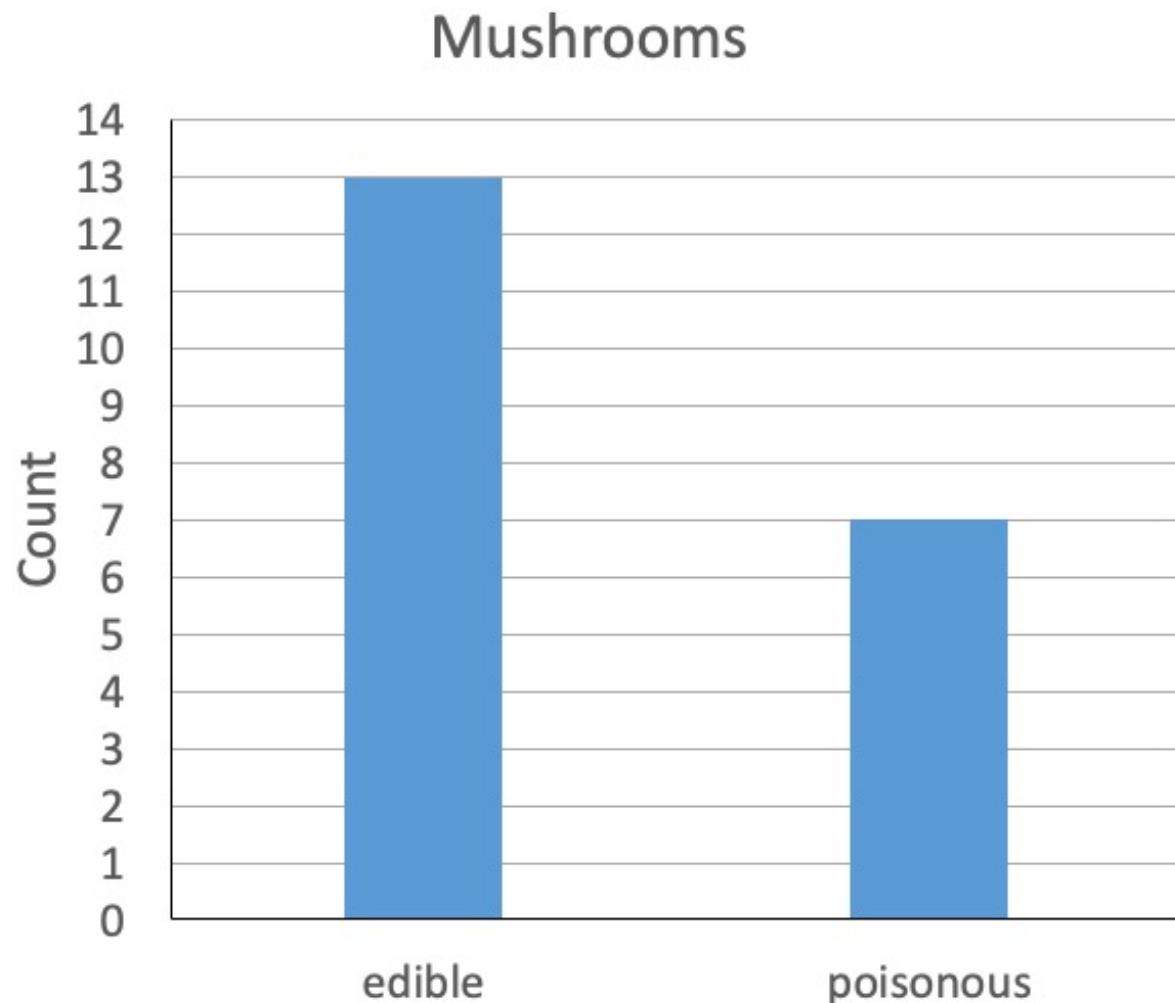
# Why visualize data?

How many mushrooms are poisonous?



# Why visualize data?

Reduce Cognitive Load



Can visualization actually make a difference?

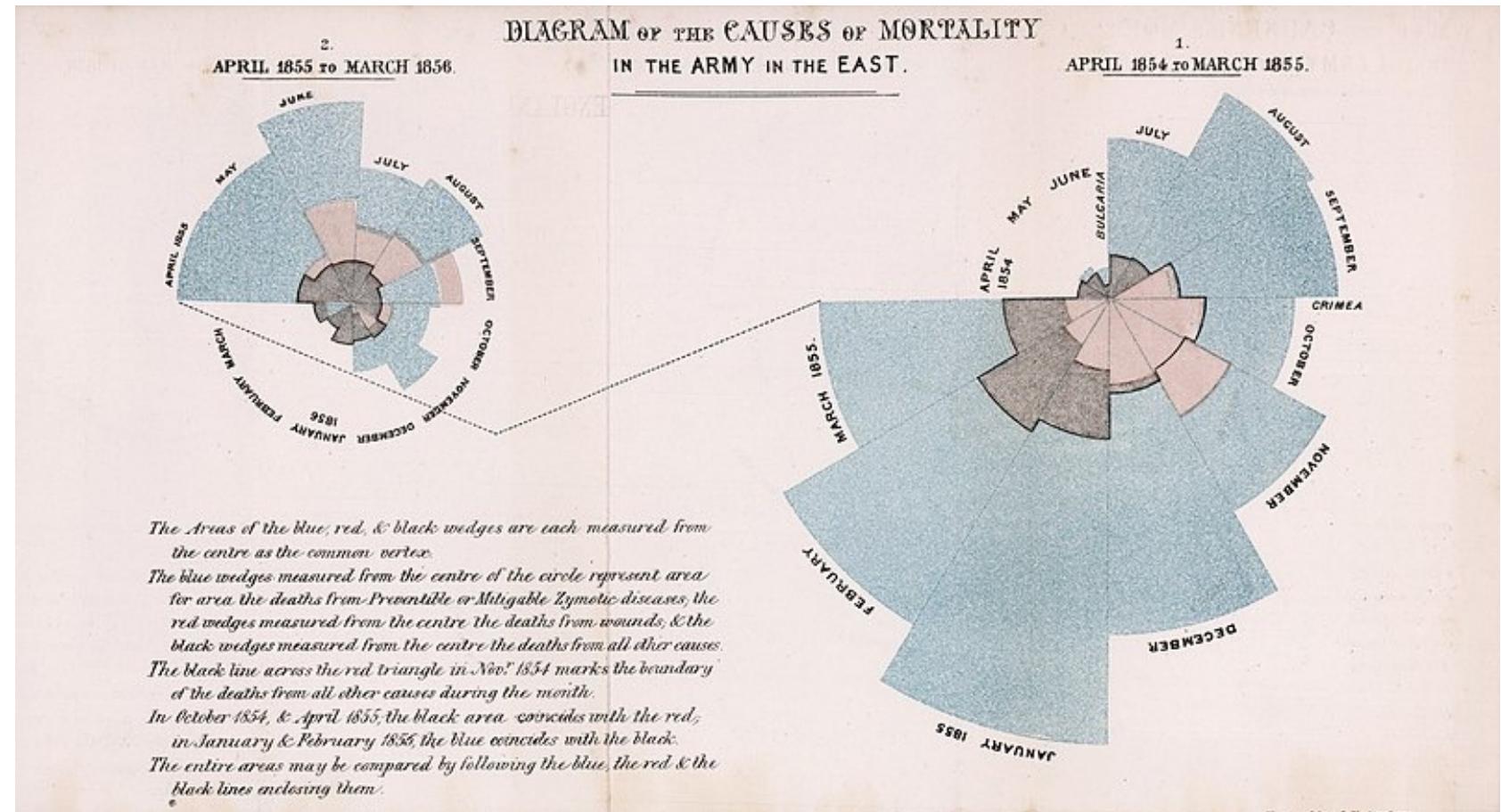
# Does it really work?

## John Snow's Cholera Map



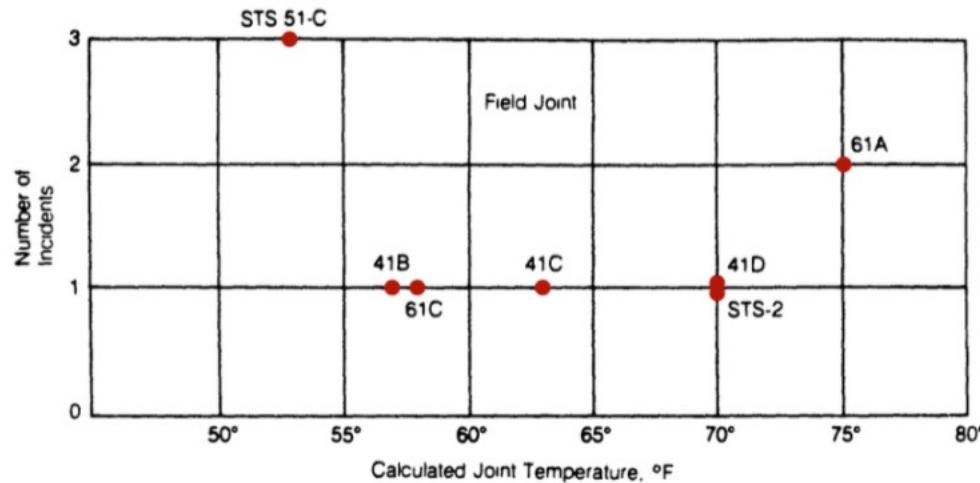
# Does it really work?

## Florence Nightingale's Rose



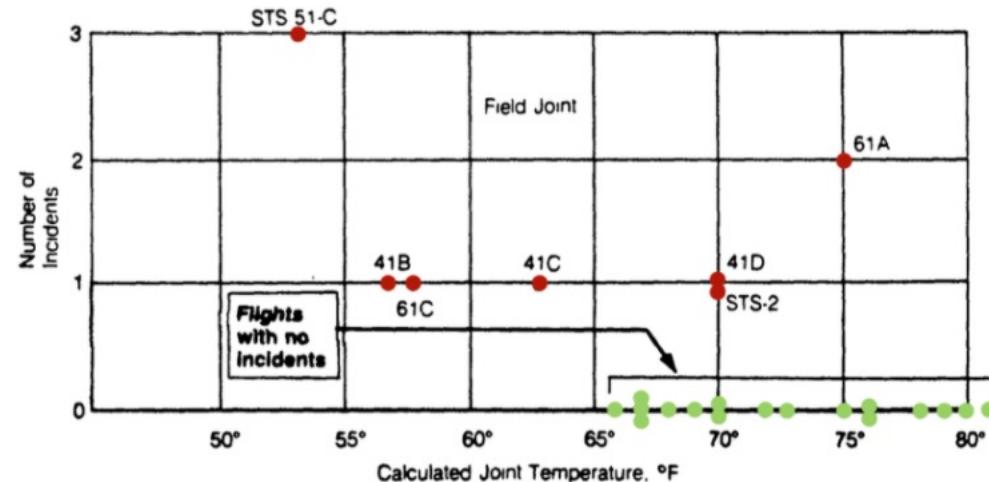
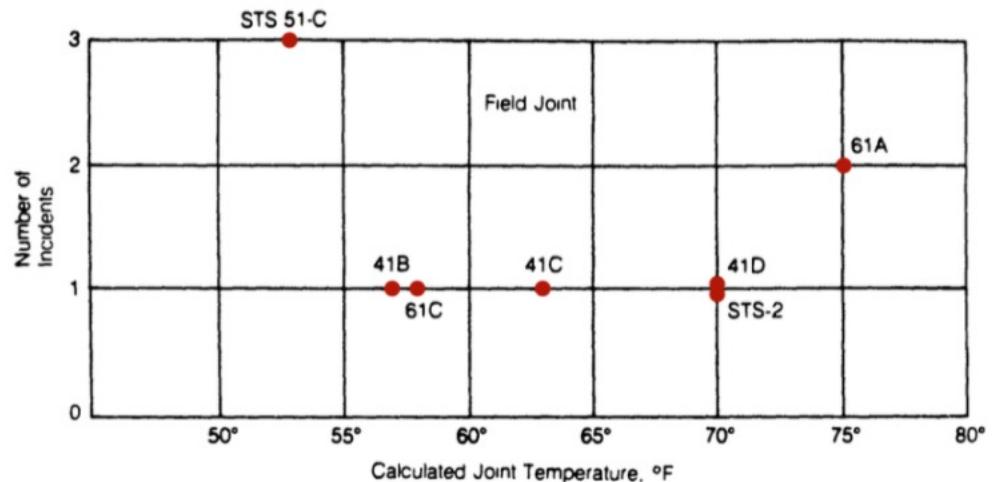
# Does it really work?

## Challenger Shuttle



# Does it really work?

## Challenger Shuttle



How?

# How to visualize data?

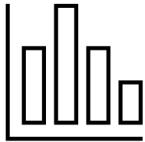
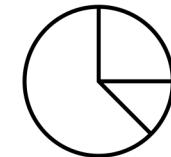
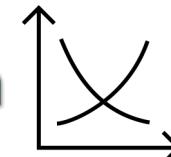
1. Goal

2. Data Types

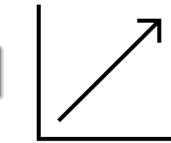
# How to visualize data?

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

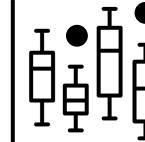
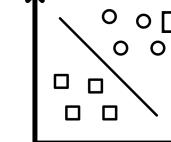
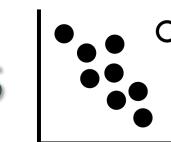
Comparison



Trend



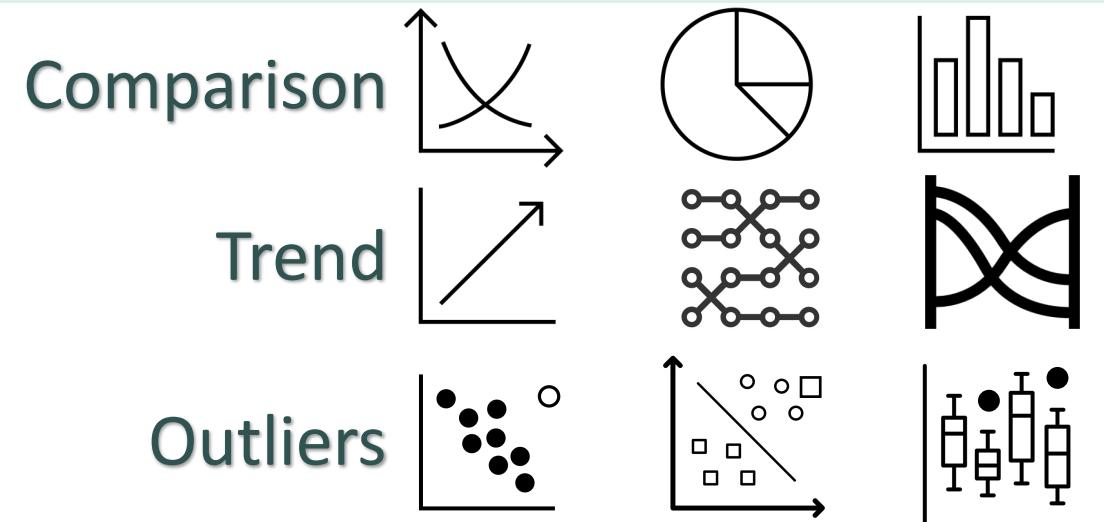
Outliers



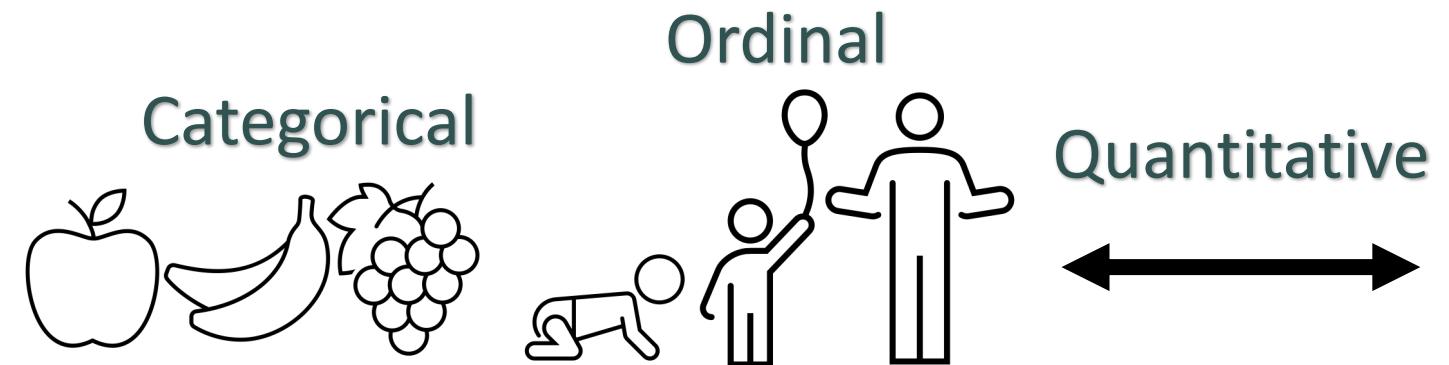
2. Data Types

# How to visualize data?

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

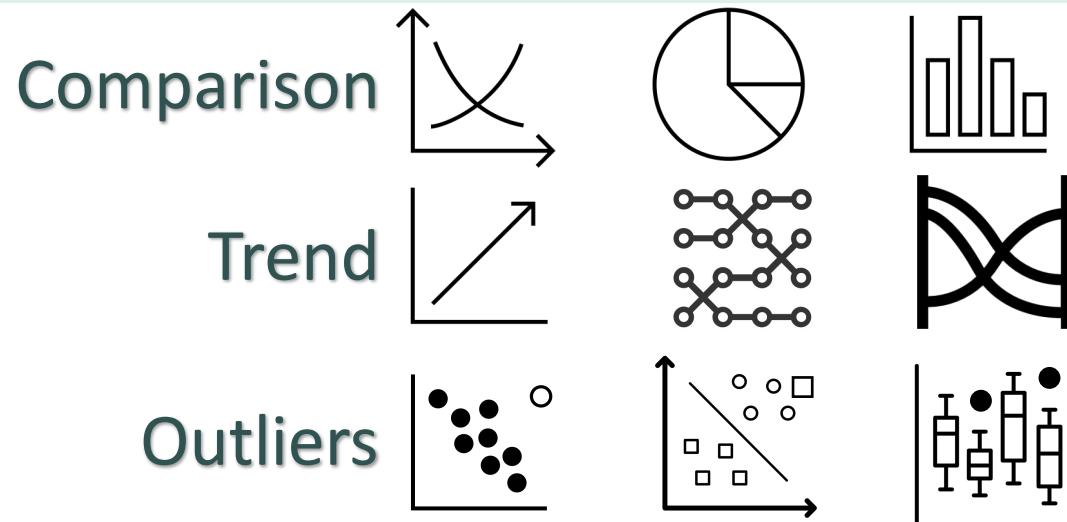


2. Data Types → What kinds of data do you need to show?



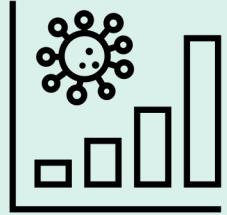
# How to visualize data?

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



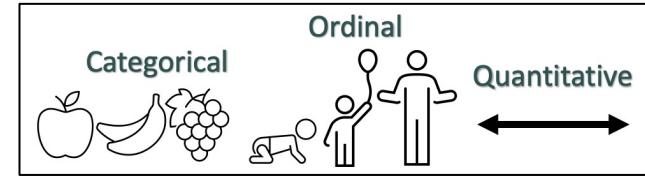
2. Data Types → What kinds of data do you need to show?

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

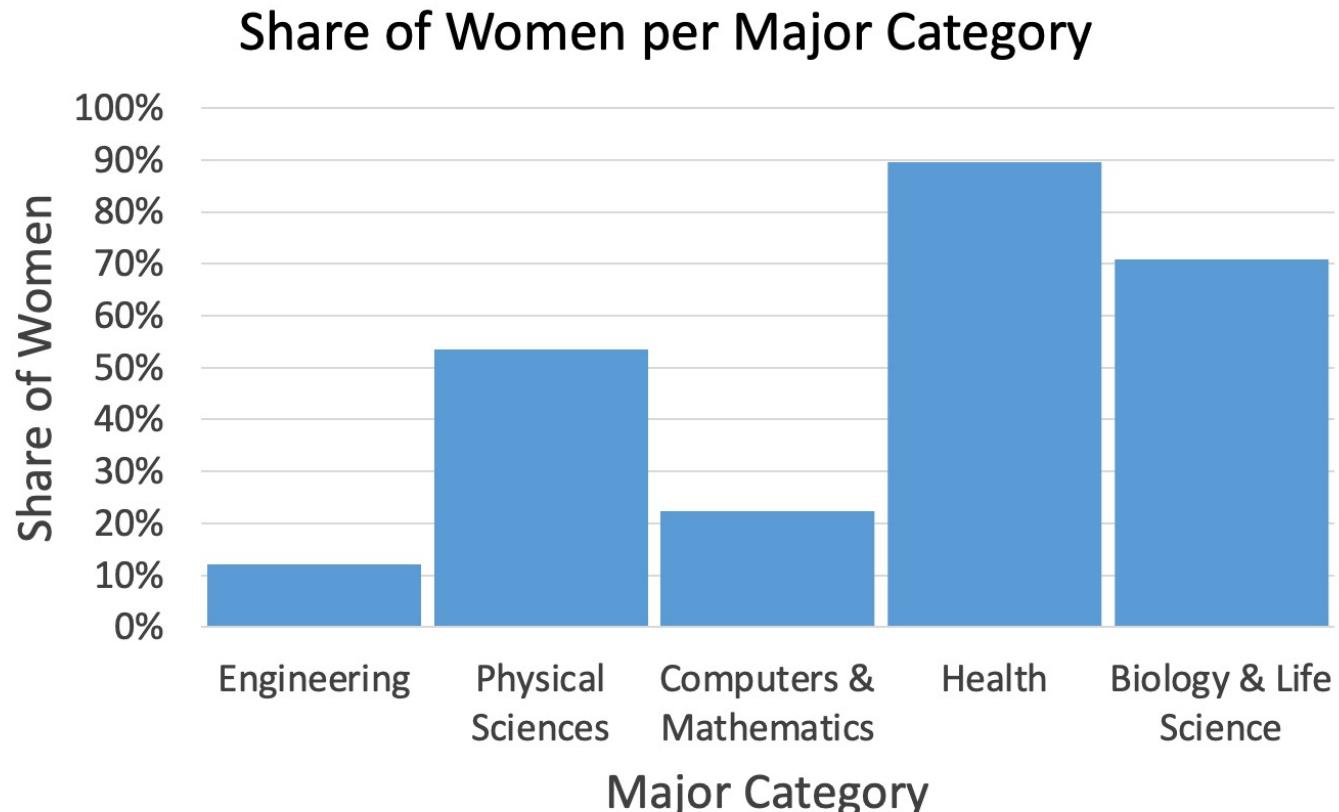


# Bar charts

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

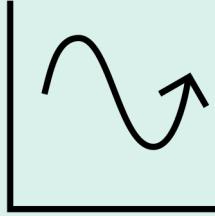


Rank	Major_category	Total	Men	Women	Share_women	Median_earnings
1	Engineering	2339	2057	282	12%	110000
7	Physical Sciences	1792	832	960	54%	62000
19	Computers & Mathematics	128319	99743	28576	22%	53000
27	Health	209394	21773	187621	90%	48000
36	Biology & Life Science	1762	515	1247	71%	45000



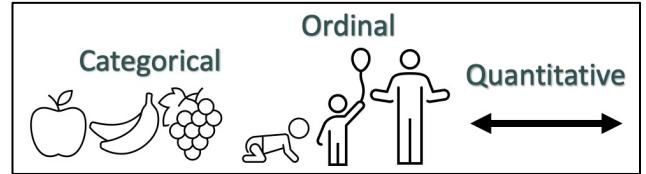
Data:

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



# Line charts

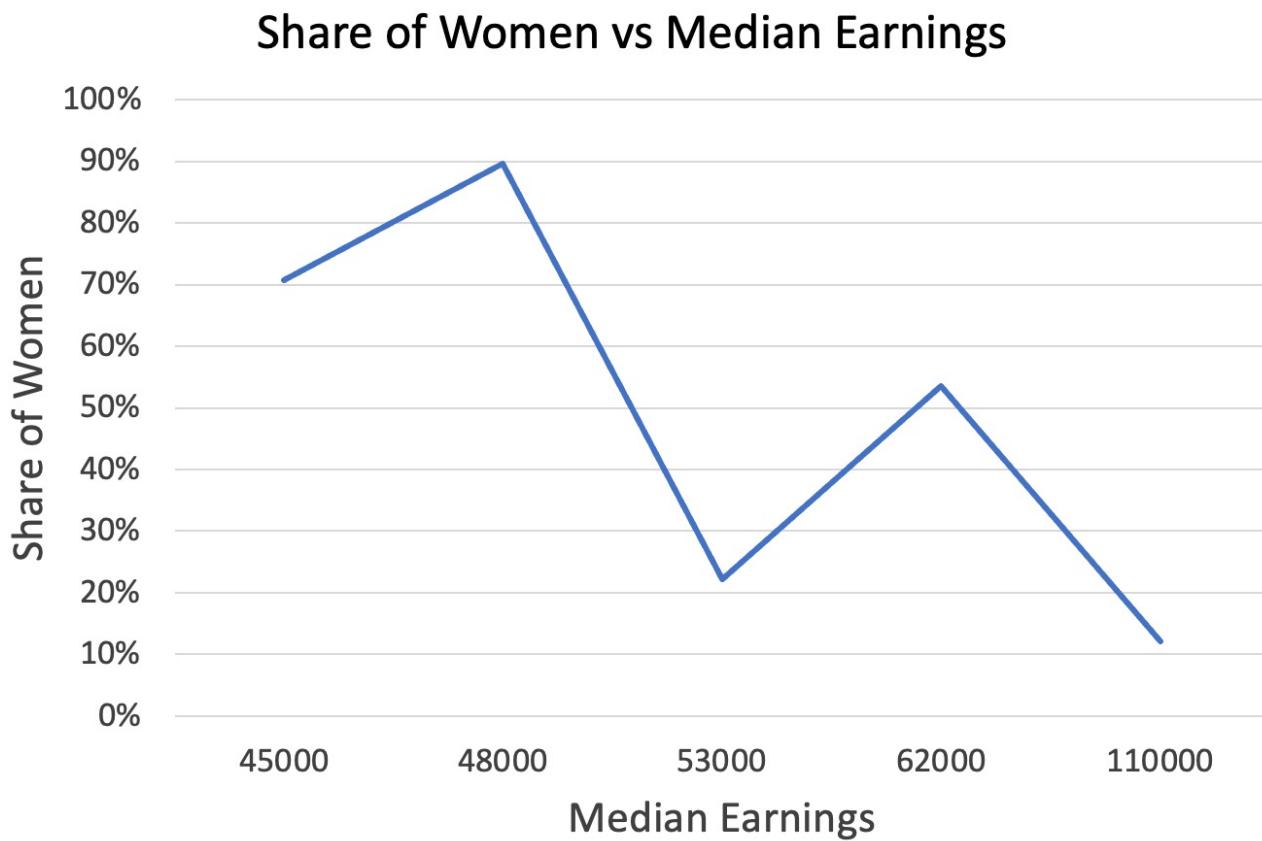
1. Goal → Trend
2. Data Types → Ordinal or Quantitative vs. Quantitative



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?

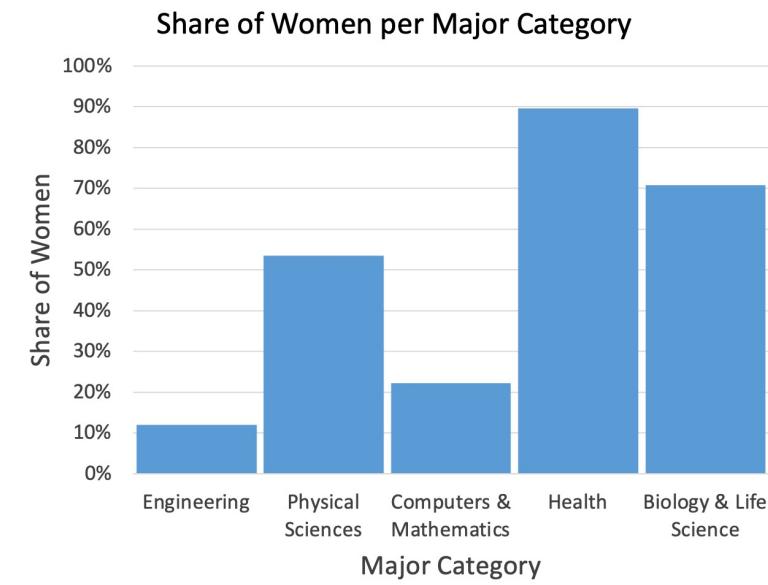
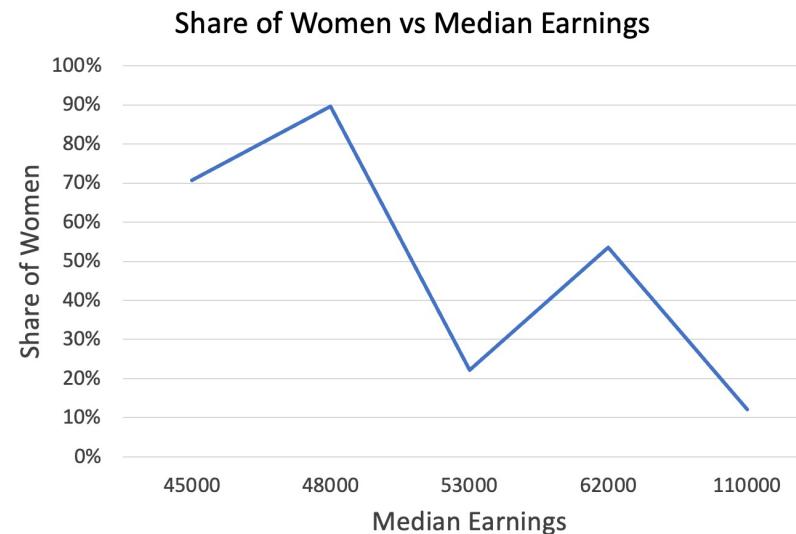
Rank	Major_category	Total	Men	Women	Share_women	Median_earnings
1	Engineering	2339	2057	282	12%	110000
7	Physical Sciences	1792	832	960	54%	62000
19	Computers & Mathematics	128319	99743	28576	22%	53000
27	Health	209394	21773	187621	90%	48000
36	Biology & Life Science	1762	515	1247	71%	45000

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?



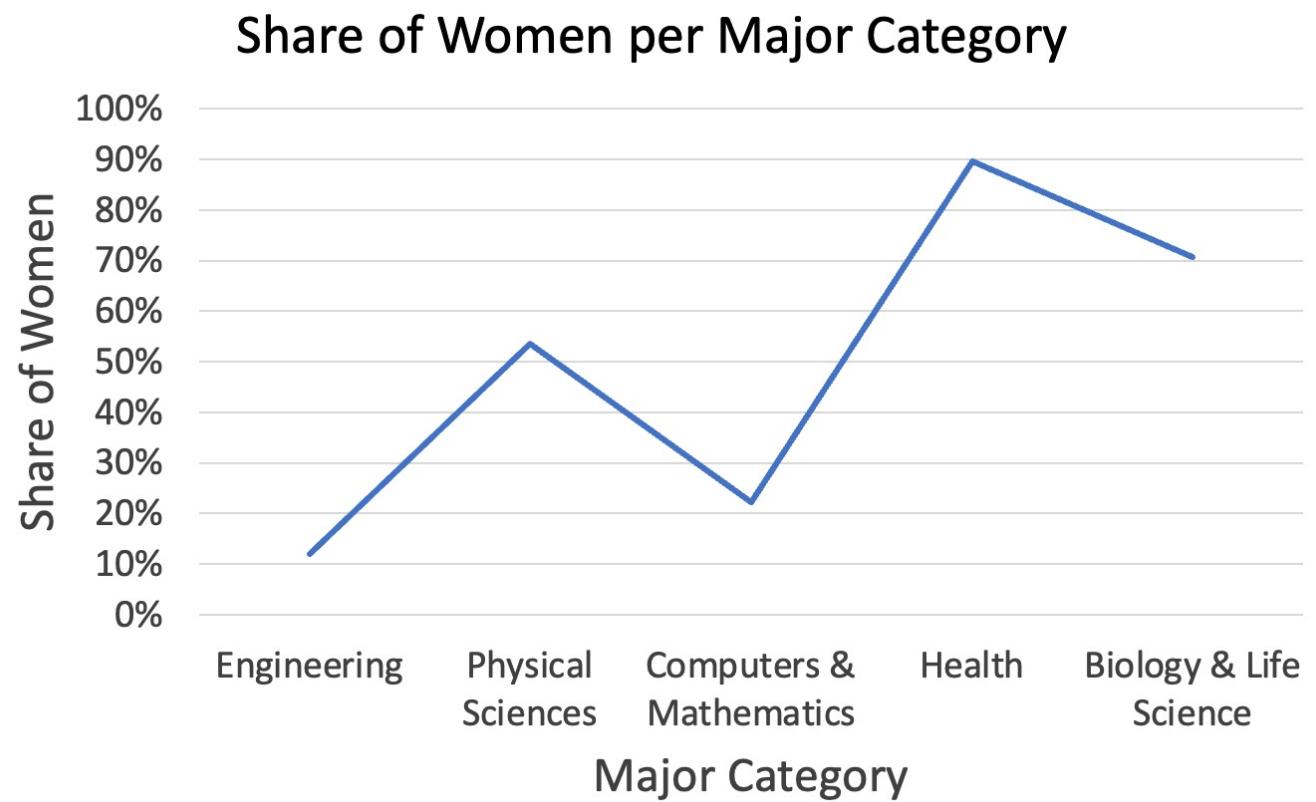
**Data:**

<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

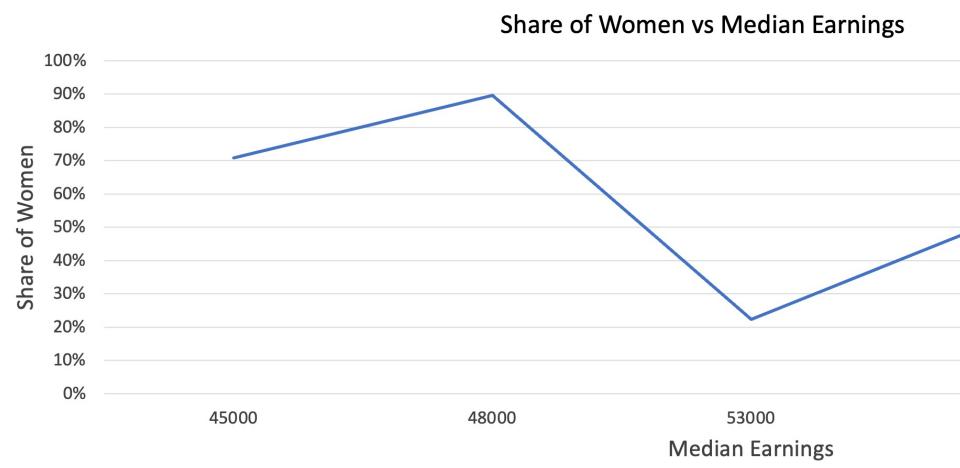
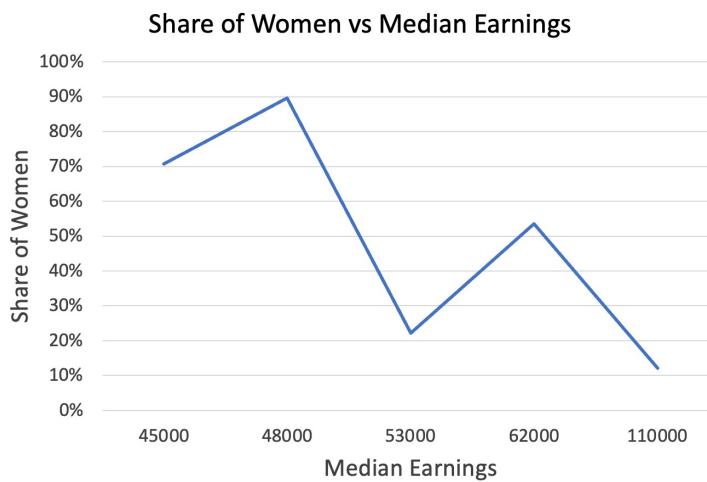


**Data:**

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

# How do we avoid bias & trickery?

**Design Contentiously & Read Critically**  
→ Aspect ratio

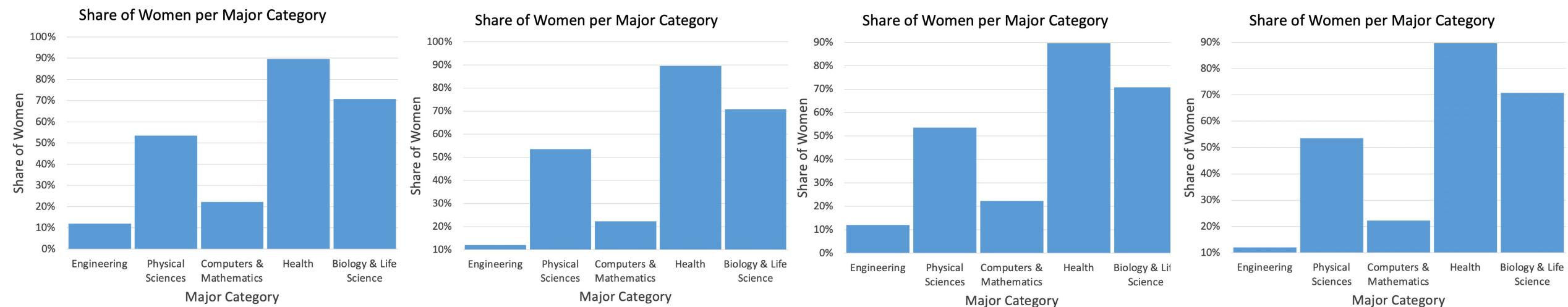


**Data:**

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

# How do we avoid bias & trickery?

Design Contentiously & Read Critically  
→ Axes



**Data:**

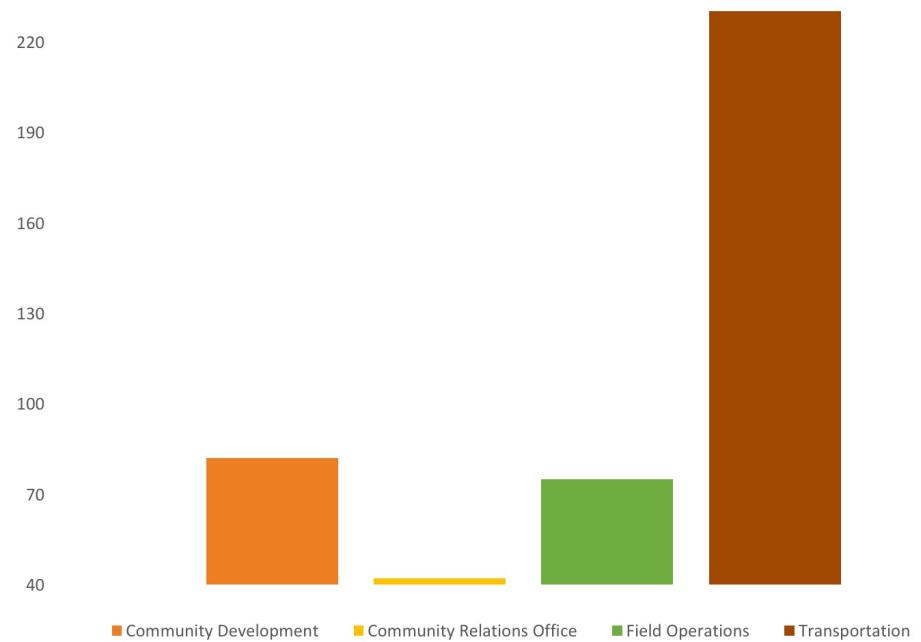
<https://github.com/fivethirtyeight/data/blob/master/college-majors/women-stem.csv>

# Let's practice

Take a critical look at  
this chart. Notice  
anything?

## Graffiti on public transportation off the chart in Tempe

According to City of Tempe, graffiti that city workers noticed and reported in 2015 were exceedingly high for public transportation.



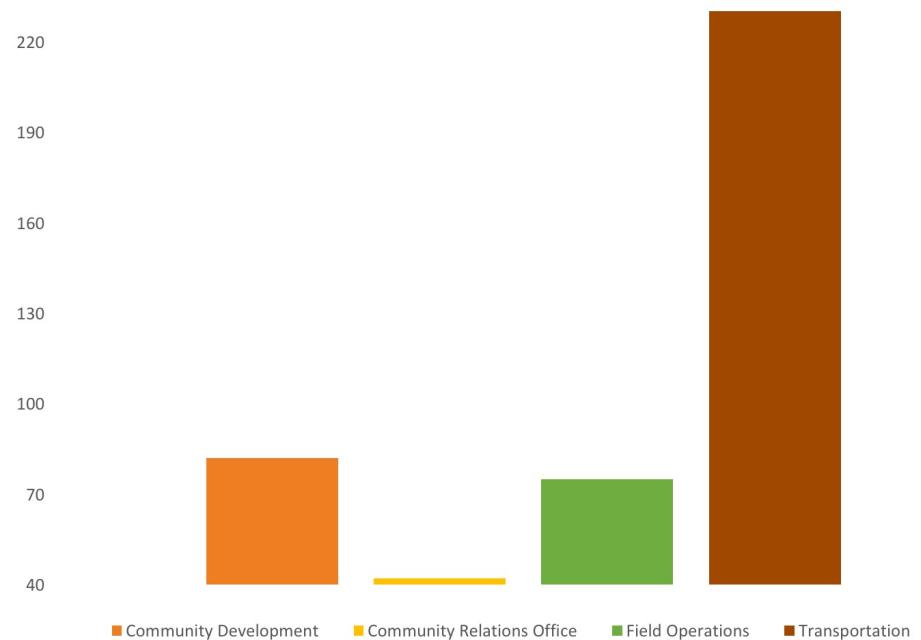
# Let's practice

Take a critical look at  
this chart. Notice  
anything?

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

## Graffiti on public transportation off the chart in Tempe

According to City of Tempe, graffiti that city workers noticed and reported in 2015 were exceedingly high for public transportation.

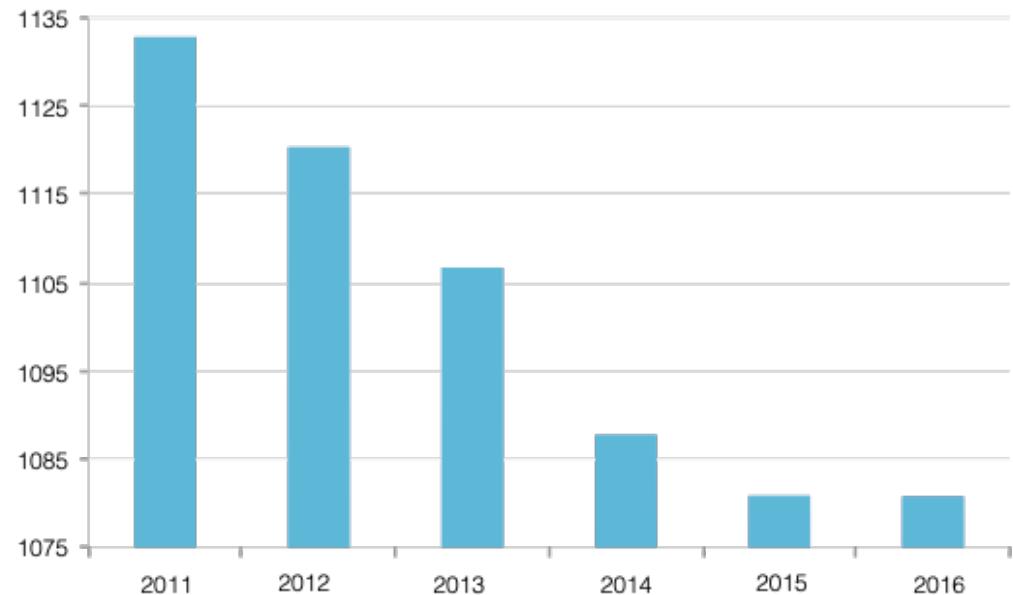


# Let's practice

Take a critical look at  
this chart. Notice  
anything?

Plummeting  
Water Supply

Within the last 5 years, our water supply at Lake Mead has plummeted.



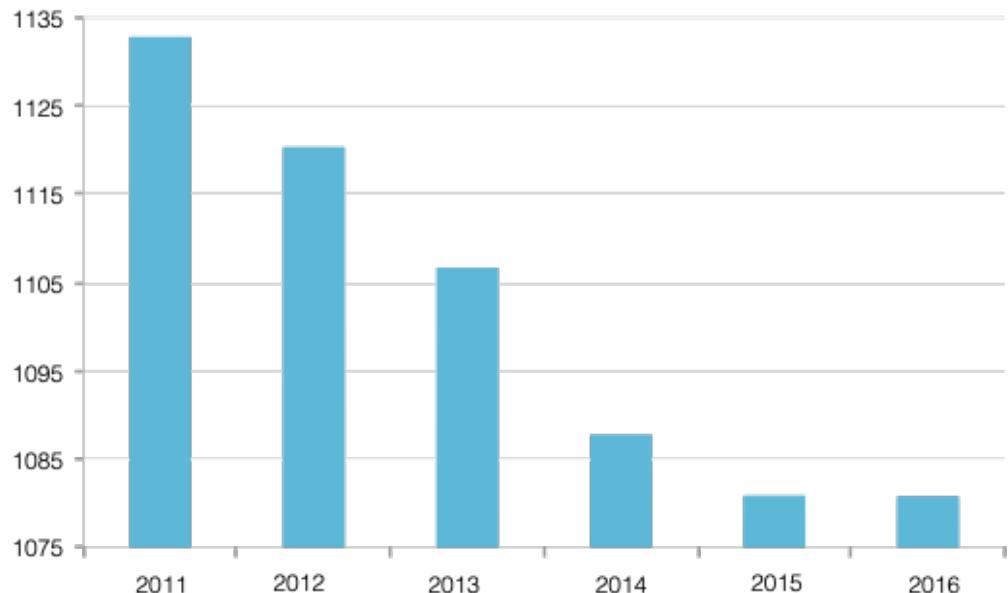
# Let's practice

Take a critical look at this chart. Notice anything?

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

Plummeting  
Water Supply 

Within the last 5 years, our water supply at Lake Mead has plummeted.



# Let's practice

Pick a dataset from the options below.

Design a chart using this data that conveys a point untruthfully.

**Data:** <https://www.tableau.com/learn/articles/free-public-data-sets>

# Let's practice

Now, trade with the person next to you.

Can you identify the "lie" in your partner's chart? How would you re-design the chart to be truthful?

**Data:** <https://www.tableau.com/learn/articles/free-public-data-sets>

Let's take a break! Stretch,  
use the restroom, talk to  
your friends, etc.. We'll  
start again in 10 minutes





# Formalizing Data Visualizations

# Data

- Remember...

country	year	cases	population
Afghanistan	2000	245	127071
Afghanistan	2000	2666	2095360
Brazil	1999	31737	17206362
Brazil	2000	80488	17404898
China	1999	212258	1272115272
China	2000	21166	128028583

## variables

country	year	cases	population
Yemen	1988	743	188817
Yemen	2000	2000	200000
Burkina Faso	1988	67767	1720000
Burkina Faso	2000	66700	1715010
China	1988	212200	12720100
China	2000	210700	12600120

## observations

## values

# Data → Visuals

- Remember...

country	year	cases	population
Afghanistan	1999	745	1837071
Afghanistan	2000	2666	20495360
Brazil	1999	37737	172006362
Brazil	2000	80488	17404898
China	1999	212258	1272915272
China	2000	213766	128042583

variables

country	year	cases	population
China	1999	745	1837071
China	2000	2666	20495360
Brazil	1999	37737	172006362
Brazil	2000	80488	17404898
Afghanistan	1999	212258	1272915272
Afghanistan	2000	213766	128042583

observations

country	year	cases	population
Afghanistan	99	745	1837071
China	99	212258	1272915272
China	00	213766	128042583

values

- Big idea behind visualization

- Data have dimensions
- Visualizations have dimensions, too
- To build good visualizations, we need to **map data dimensions to visual dimensions** in a principled way

# Data → Visuals

- Remember...

country	year	cases	population
Afghanistan	1999	745	1837071
Afghanistan	2000	2666	20495360
Brazil	1999	37737	172006362
Brazil	2000	80488	17404898
China	1999	212258	1272915272
China	2000	213766	128042583

variables

country	year	cases	population
China	1999	745	1837071
China	2000	2666	20495360
Brazil	1999	37737	172006362
Brazil	2000	80488	17404898
Afghanistan	1999	212258	1272915272
Afghanistan	2000	213766	128042583

observations

country	year	cases	population
Afghanistan	1999	745	1837071
China	1999	212258	1272915272
China	2000	213766	128042583

values

- Big idea behind visualization

- Data have dimensions
- Visualizations have dimensions, too
- To build good visualizations, we need to  
**map data dimensions to visual dimensions in a principled way**

# Data → Visuals

- Remember...

country	year	cases	population
Afghanistan	1999	745	1837071
Afghanistan	2000	2666	20495360
Brazil	1999	37737	172006362
Brazil	2000	80488	17404898
China	1999	212258	1272915272
China	2000	213766	128042583

variables

country	year	cases	population
China	1999	745	1837071
China	2000	2666	20495360
Brazil	1999	37737	172006362
Brazil	2000	80488	17404898
Afghanistan	1999	212258	1272915272
Afghanistan	2000	213766	128042583

observations

country	year	cases	population
Afghanistan	1999	745	1837071
China	1999	212258	1272915272
China	2000	213766	128042583

values

- Big idea behind visualization

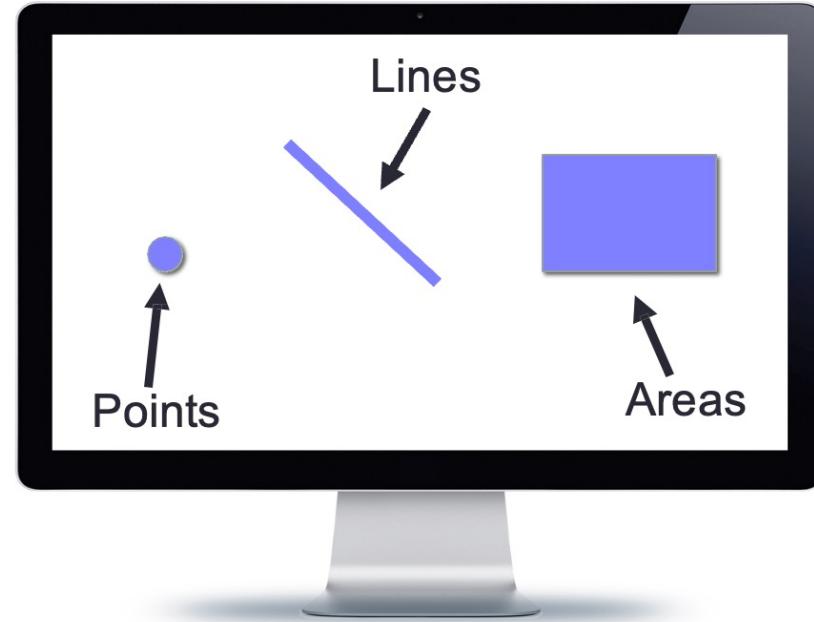
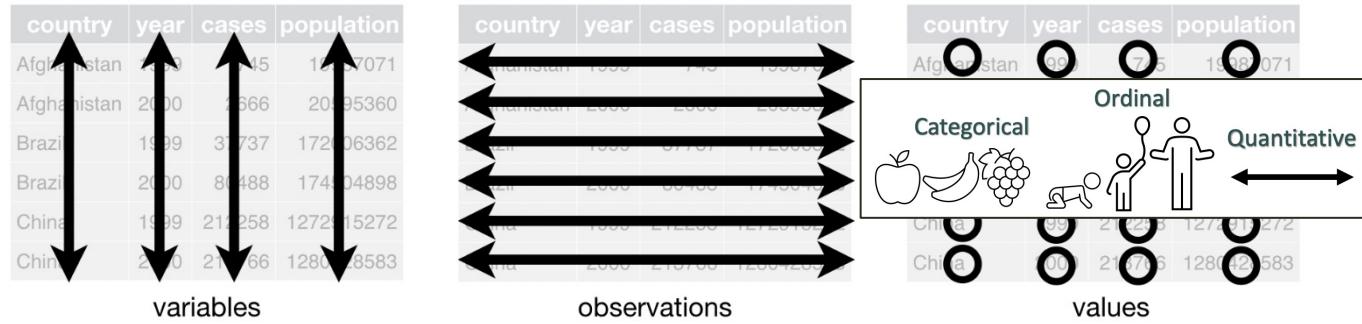
- Data have dimensions
- Visualizations have dimensions, too
- To build good visualizations, we need to  
**map data dimensions to visual dimensions in a principled way**

Data → Visuals

# Data

## Visuals

- Marks
  - The “ink”



Data → Visuals

## Data

country	year	cases	population
Afghanistan	1993	145	1837071
Afghanistan	2000	1666	2035360
Brazil	1999	3737	17206362
Brazil	2000	8488	174304898
China	1999	21258	127215272
China	2000	2166	128028583

variables

country	year	cases	population
Afghanistan	1993	145	1837071
Afghanistan	2000	1666	2035360
Brazil	1999	3737	17206362
Brazil	2000	8488	174304898
China	1999	21258	127215272
China	2000	2166	128028583

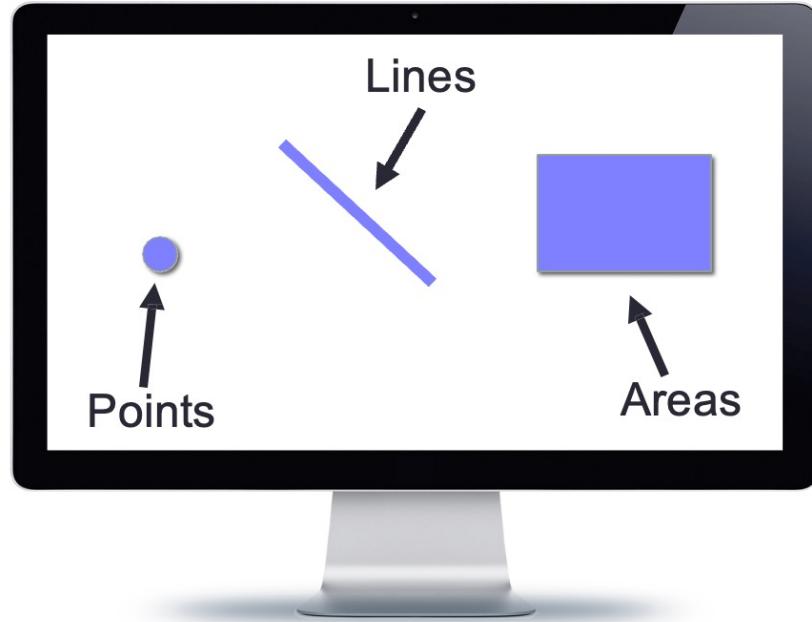
observations

country	year	cases	population
Afghanistan	1993	145	1837071
Afghanistan	2000	1666	2035360
Brazil	1999	3737	17206362
Brazil	2000	8488	174304898
China	1999	21258	127215272
China	2000	2166	128028583

values

## Visuals

- Marks
  - The “ink”

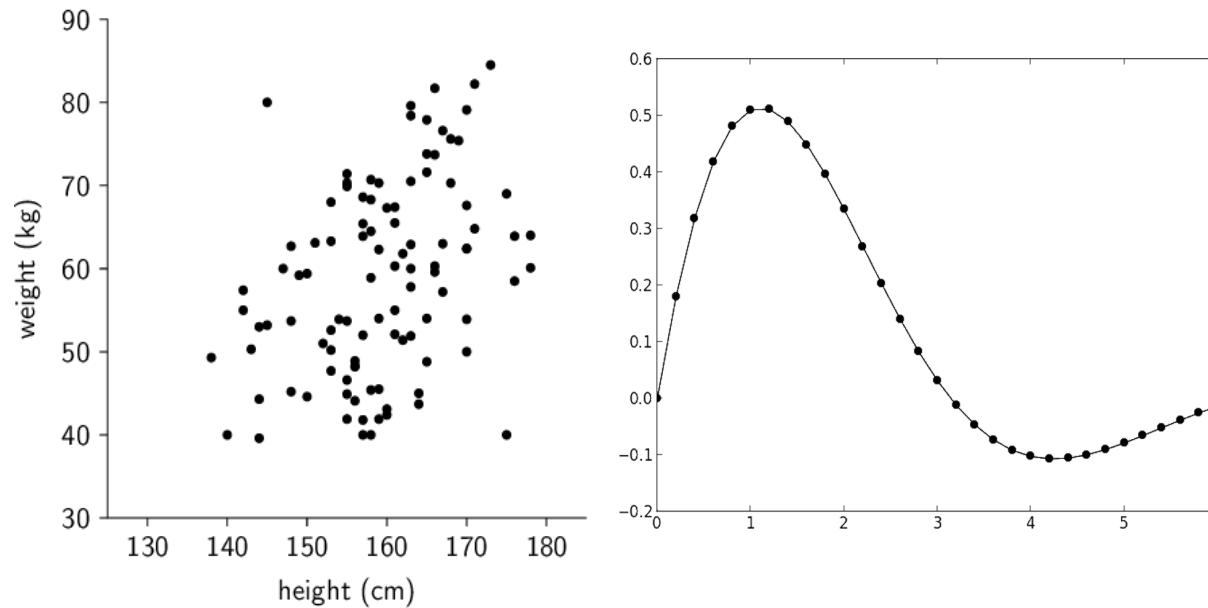


- Channels or dimensions
  - *How the marks show up on the page*

# Visual Channels / Dimensions

## Position

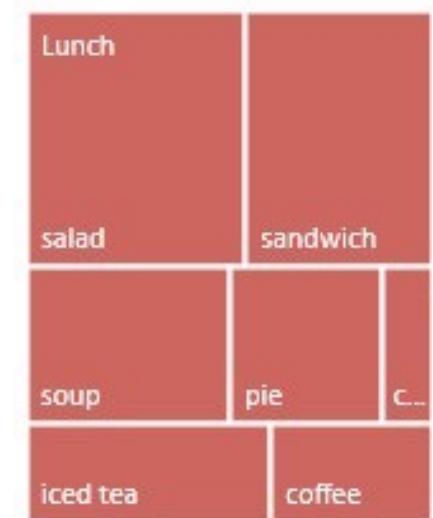
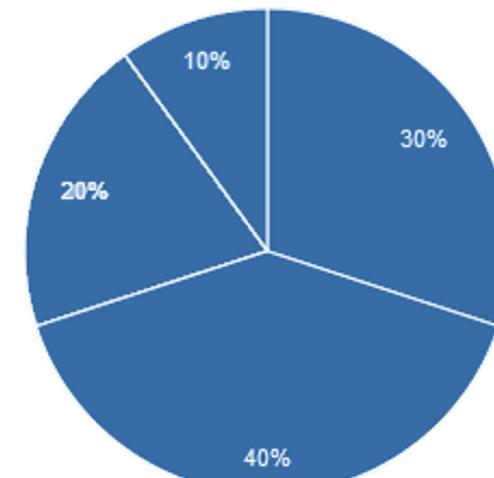
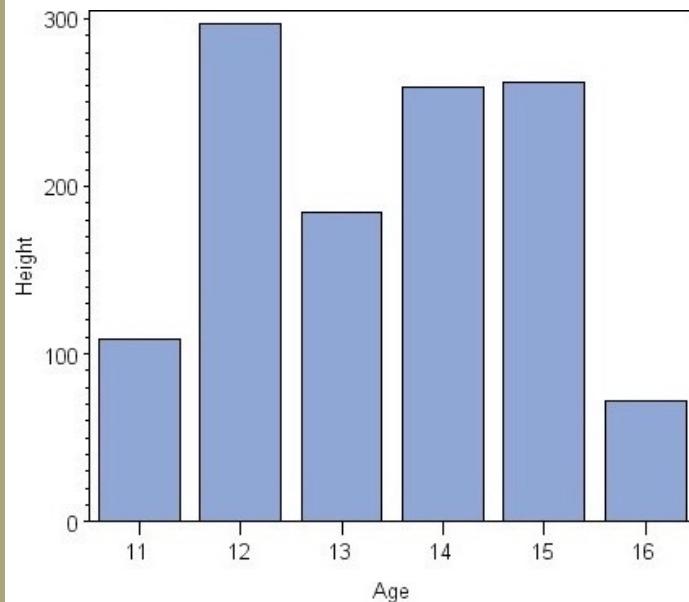
- Encode information using *where* mark is drawn
- Ex.



# Visual Channels / Dimensions

## Size

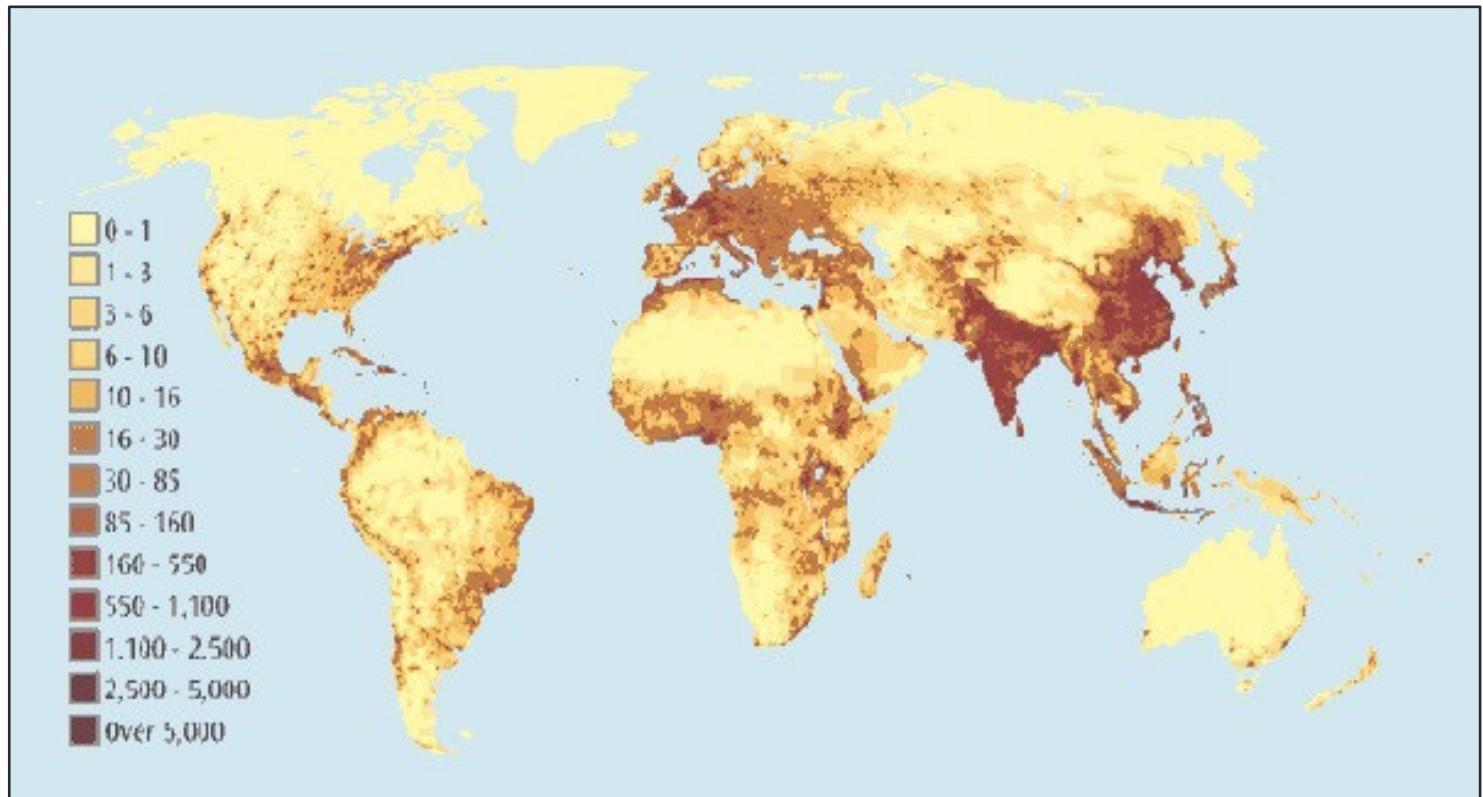
- Encode information using *how big* mark is drawn
- Ex.



# Visual Channels / Dimensions

## *Value*

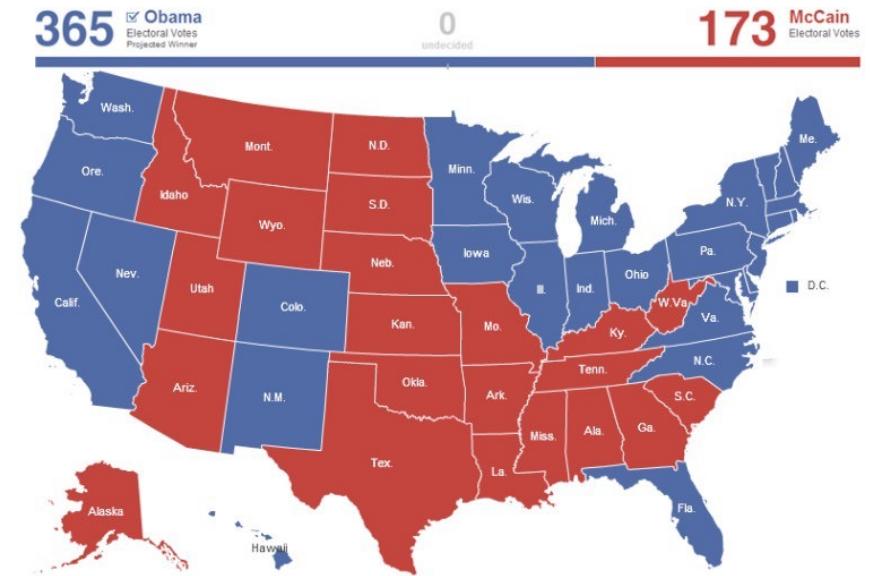
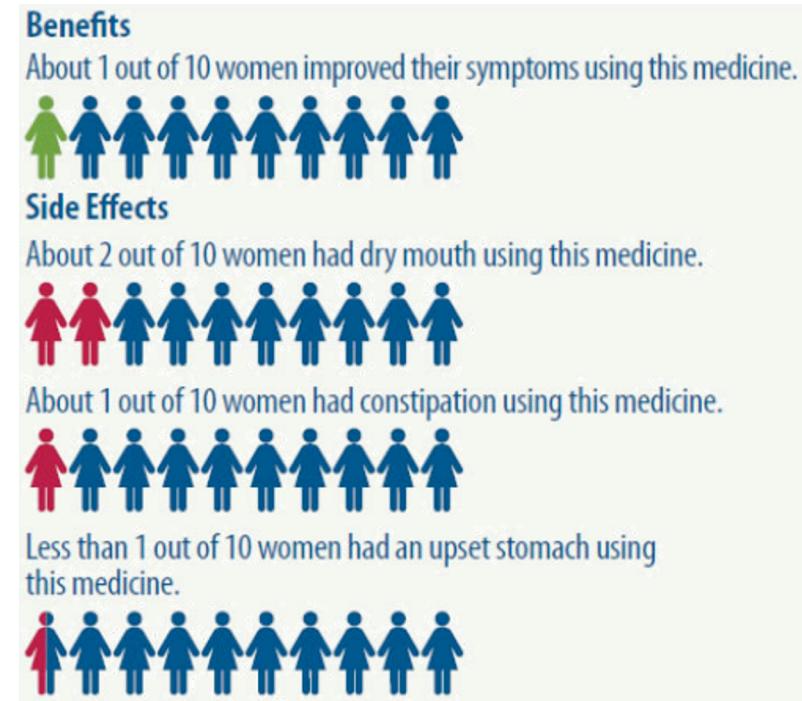
- Encode information using *how dark* mark is drawn
- Ex.



# Visual Channels / Dimensions

## Color

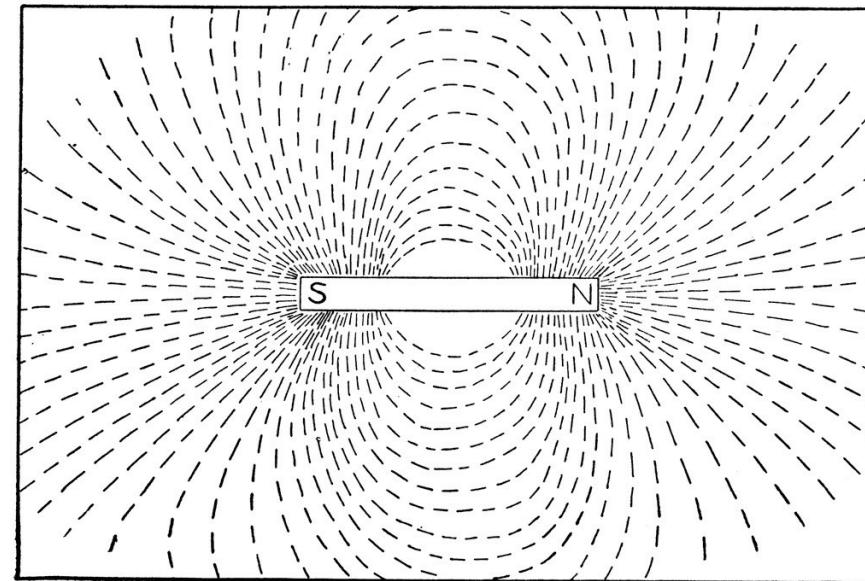
- Encode information using *hue* of mark
- Ex.



# Visual Channels / Dimensions

## Orientation

- Encode information using how mark is *rotated*
- Ex.



# Visual Channels / Dimensions

## *Shape*

- Encode information using how mark is *shaped*
- Ex.

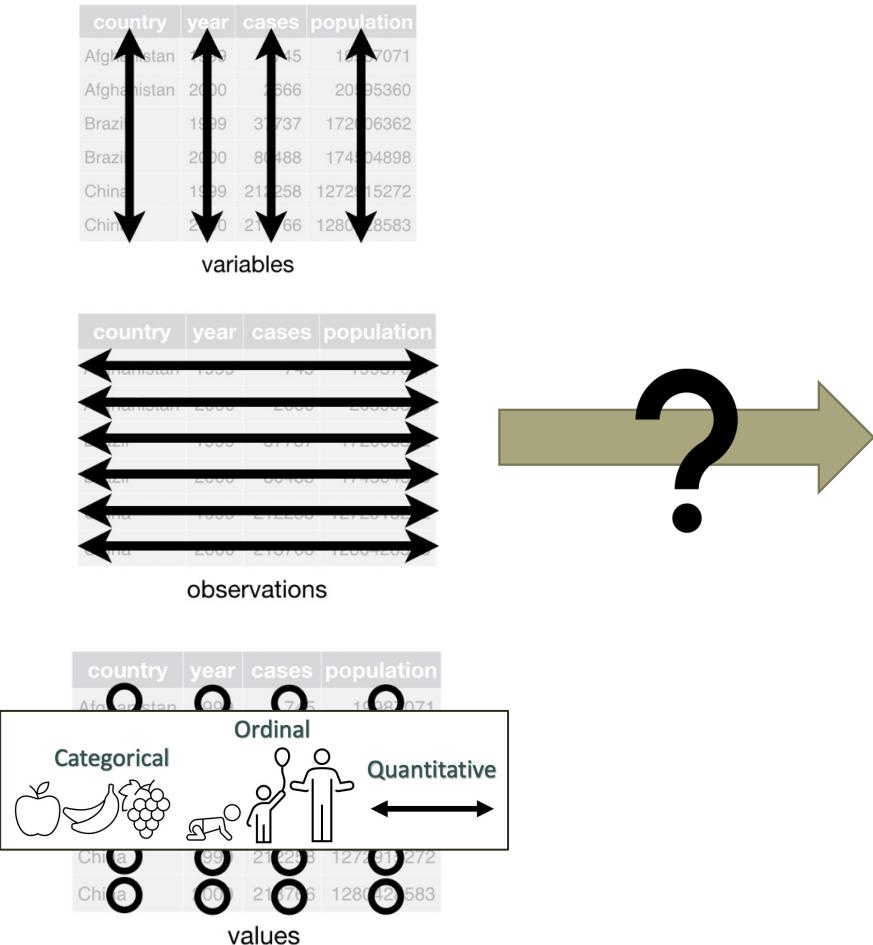


Data → Visuals

- Remember... **Big idea behind visualization**
  - Map data dimensions to visual dimensions in a principled way

# Data → Visuals

- Remember... **Big idea behind visualization**
  - Map data dimensions to visual dimensions in a principled way



POSITION

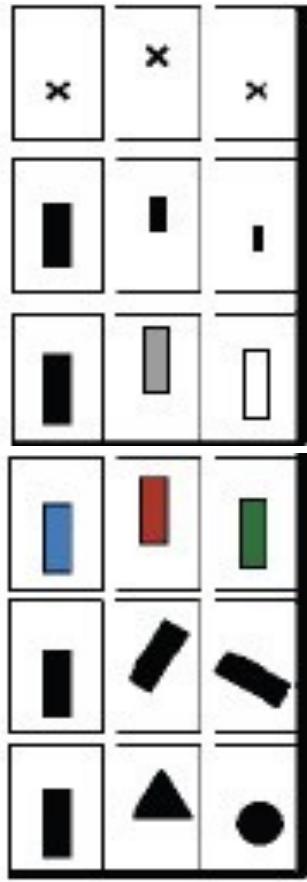
SIZE

VALUE

COLOR

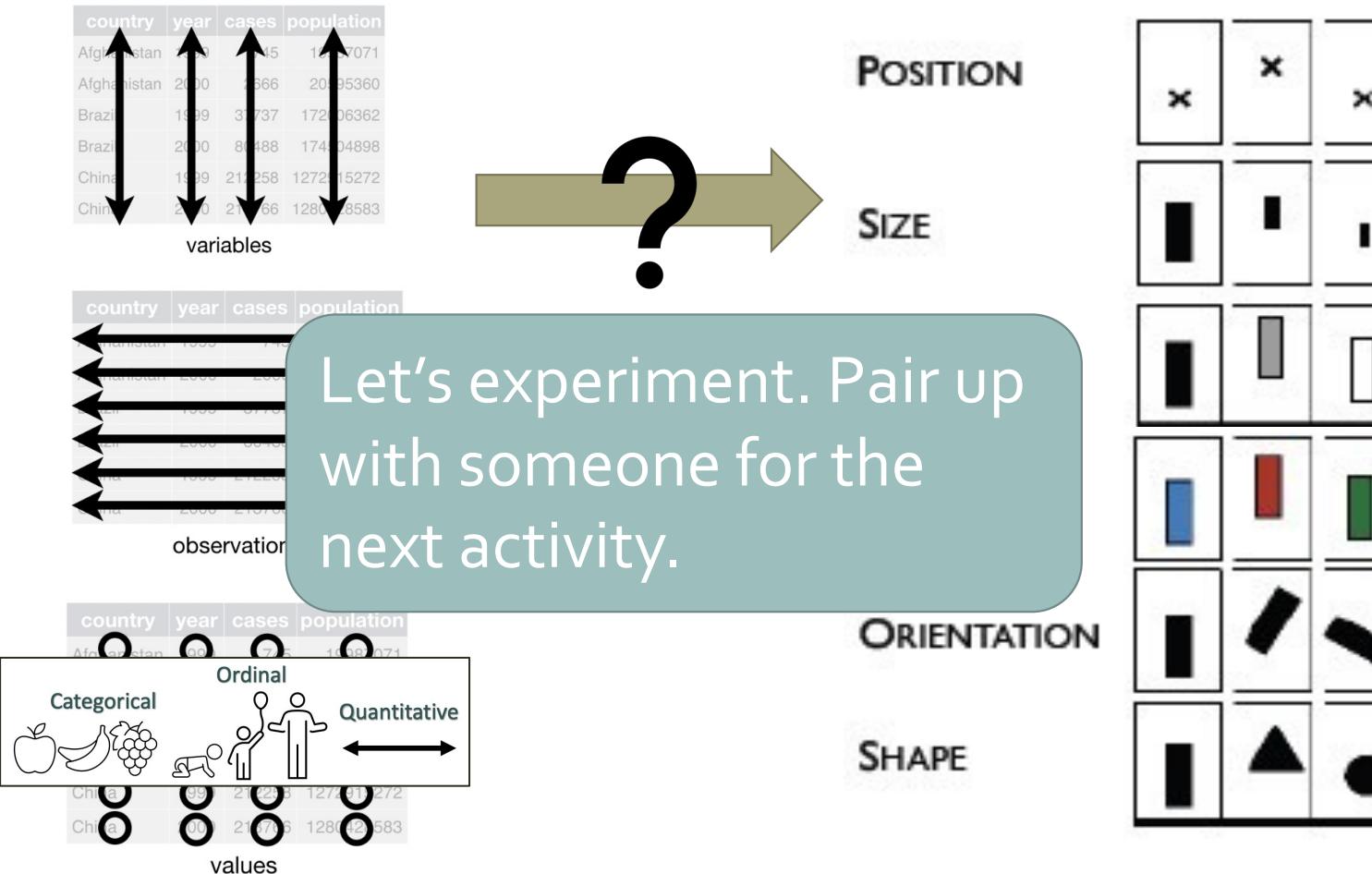
ORIENTATION

SHAPE



# Data → Visuals

- Remember... **Big idea behind visualization**
  - Map data dimensions to visual dimensions in a principled way



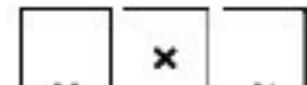
Data → Visuals



Work with your partner to represent apple, banana, grapes, with each visual channel.

Be prepared to share with the class.

POSITION



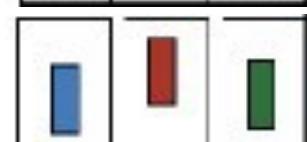
SIZE



VALUE



COLOR



ORIENTATION



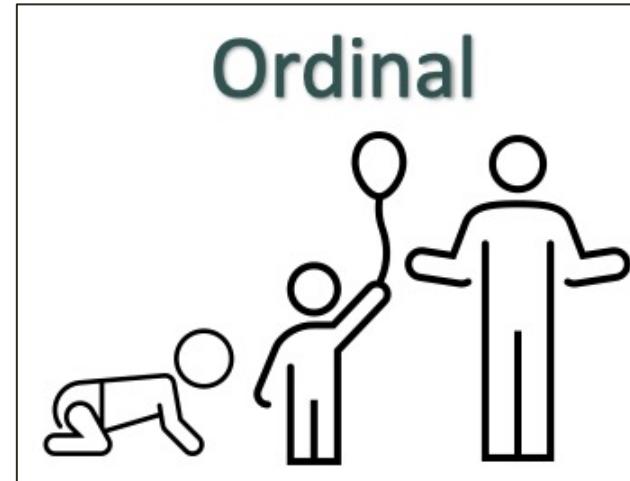
SHAPE



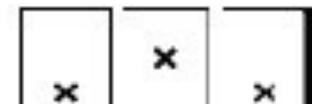
Data → Visuals

Work with your partner to represent baby, toddler, adult, with each visual channel.

Be prepared to share with the class.



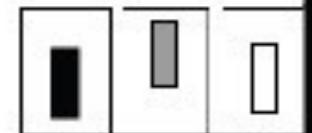
POSITION



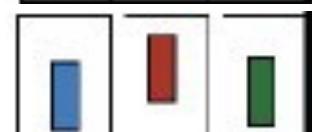
SIZE



VALUE



COLOR



ORIENTATION



SHAPE



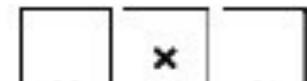
Data → Visuals

Work with your partner to represent 1, 2, 3, with each visual channel.

Be prepared to share with the class.



POSITION



SIZE



VALUE



COLOR



ORIENTATION



SHAPE



# Data → Visuals

- Remember... **Big idea behind visualization**
  - Map data dimensions to visual dimensions in a principled way
  - Not all visual dimensions can represent all data types

	Categorical	Ordinal	Quantitative
POSITION	x x x		
SIZE	■ ■ ■	▢ ▢ ▢	▪ ▪ ▪
VALUE	■ ■ ■	▢ ▢ ▢	▢ ▢ ▢
COLOR	■ ■ ■	▢ ▢ ▢	▢ ▢ ▢
ORIENTATION	■ ■ ■	▢ ▢ ▢	▢ ▢ ▢
SHAPE	■ ■ ■	▢ ▢ ▢	▢ ▢ ▢

Legend: Categorical icons (apple, banana, grapes), Ordinal icons (children), Quantitative icons (adult).

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

## Try it out!

- Work with 1 – 2 other people. Be prepared to share your work with the class.
- Find a data visualization you think is interesting
  - Some ideas: NYTimes, VisualisingData.com, Visual.ly
  - Remember to cite your source!
- Identify the following:
  - What is the data that's being visualized? Where did it come from?
  - Which data dimensions are mapped to which visual dimensions?
  - How does this shape your understanding of the data?
  - If you liked the visualization: what is it doing well?
  - If you disliked the visualization: what would you change?