Data Visualization

or, how data presentation is a communicative choice

Hannah Ringler, PhD
Humanities Department
October 6, 2023

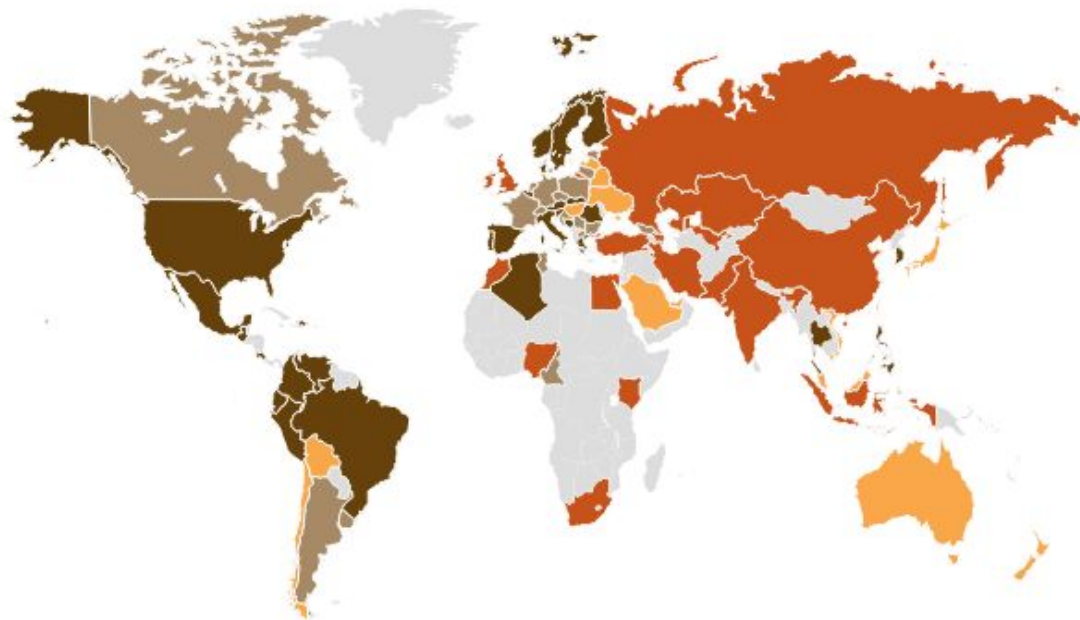ILLINOIS INSTITUTE OF TECHNOLOGY

Which sounds better?

1 in 50 pregnancies by older couples will result in fetal abnormalities.

There is a 97% chance that babies born to older couples will be healthy.

Tea or coffee

Tea and coffee consumption
% of total, 2012
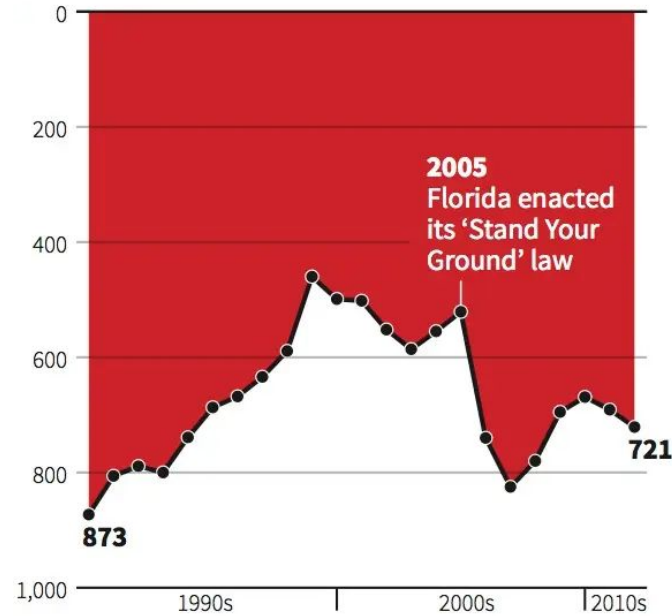
Zoom to ▼

Legend: Tea (75-100%) | Tea (50-75%) | Coffee (50-75%) | Coffee (75-100%) | No data

Gun deaths in Florida
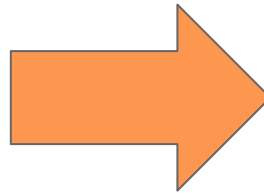
Number of murders committed using firearms

2005
Florida enacted its 'Stand Your Ground' law

873

721

1990s   2000s   2010s

Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

Table format: XY		X	Group A		
		minutes	Test group A		
	⊗	X	A:Y1	A:Y2	A:Y3
1	Title	0	0.0	0.0	0.0
2	Title	2	391.0	384.0	543.0
3	Title	4	562.0	478.0	584.0
4	Title	6	746.0	798.0	715.0
5	Title	8	823.0	754.0	669.0
6	Title	10	736.0	846.0	742.0
7	Title	12	832.0	855.0	799.0
8	Title	14	923.0	750.0	816.0
9	Title	16	801.0	854.0	826.0
10	Title	18	811.0	795.0	864.0
11	Title	20	942.0	831.0	938.0

40%
75%
90%

70

Visualizing data strategically helps both **you** and **your audience**.

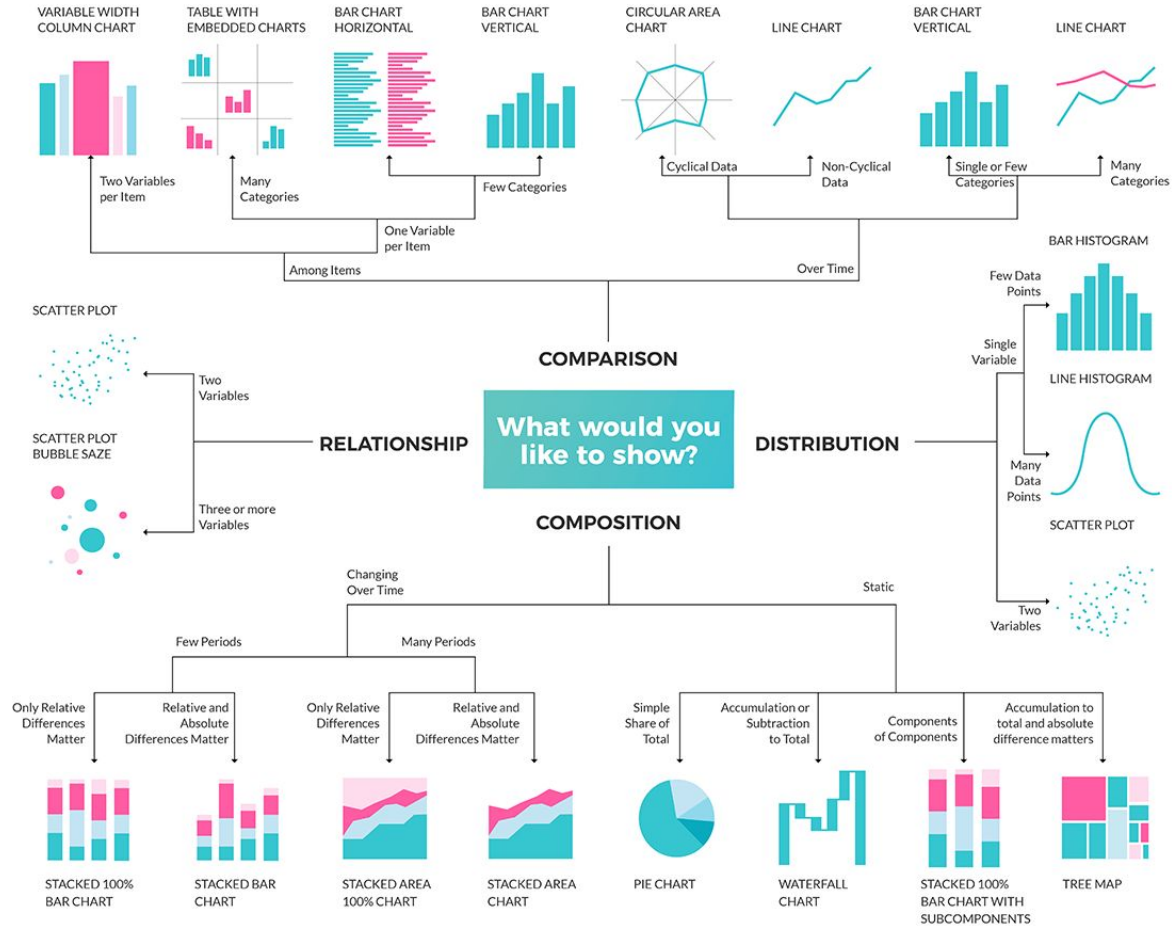| Table format: XY		X	Group A		
		minutes	Test group A		
	✕	X	A:Y1	A:Y2	A:Y3
1	Title	0	0.0	0.0	0.0
2	Title	2	391.0	384.0	543.0
3	Title	4	562.0	478.0	584.0
4	Title	6	746.0	798.0	715.0
5	Title	8	823.0	754.0	669.0
6	Title	10	736.0	846.0	742.0
7	Title	12	832.0	855.0	799.0
8	Title	14	923.0	750.0	816.0
9	Title	16	801.0	854.0	826.0
10	Title	18	811.0	795.0	864.0
11	Title	20	942.0	831.0	938.0

Better — **Data representation** → Worse

Tells useful story(ies)
Easy to read
Accurate
Minimal/no negative implications

Tells useless story(ies)
Confusing
Misleading
Negative implications

How do you create a good data viz?

How do you create a good data viz?

Start with your data story.

VARIABLE WIDTH COLUMN CHART — Two Variables per Item

TABLE WITH EMBEDDED CHARTS — Many Categories

BAR CHART HORIZONTAL — One Variable per Item

BAR CHART VERTICAL — Few Categories

Among Items

CIRCULAR AREA CHART — Cyclical Data

LINE CHART — Non-Cyclical Data

BAR CHART VERTICAL — Single or Few Categories

LINE CHART — Many Categories

Over Time

COMPARISON

SCATTER PLOT — Two Variables

SCATTER PLOT BUBBLE SAZE — Three or more Variables

RELATIONSHIP

What would you like to show?

DISTRIBUTION

BAR HISTOGRAM — Few Data Points

Single Variable

LINE HISTOGRAM

Many Data Points

SCATTER PLOT

Two Variables

COMPOSITION

Changing Over Time

Few Periods — Only Relative Differences Matter — STACKED 100% BAR CHART

Relative and Absolute Differences Matter — STACKED BAR CHART

Many Periods — Only Relative Differences Matter — STACKED AREA 100% CHART

Relative and Absolute Differences Matter — STACKED AREA CHART

Static

Simple Share of Total — PIE CHART

Accumulation or Subtraction to Total — WATERFALL CHART

Components of Components — STACKED 100% BAR CHART WITH SUBCOMPONENTS

Accumulation to total and absolute difference matters — TREE MAP

(Microsoft Education)

We'll focus on 5 main types of data stories:

X is more/less/the same as Y

X is changing over time
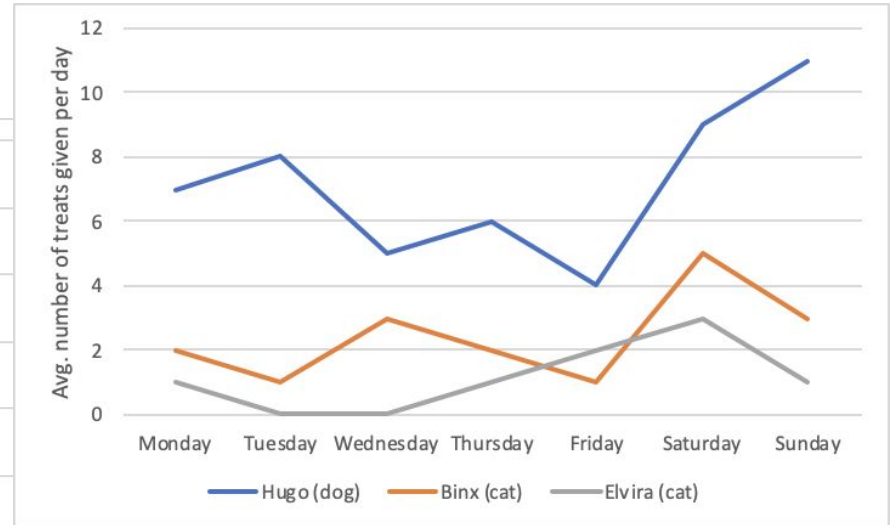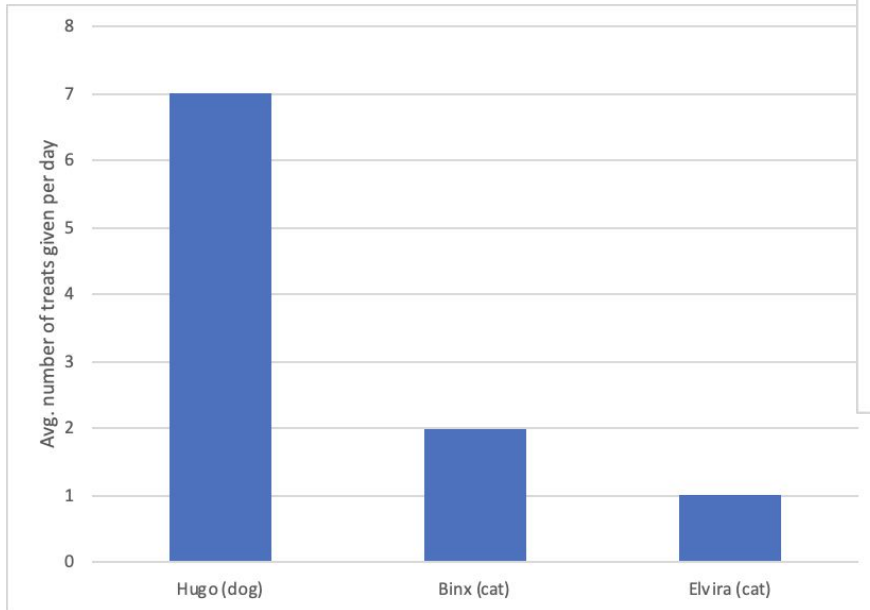
X is changing in a similar/opposite direction to Y

X is a large/small/average part of a whole
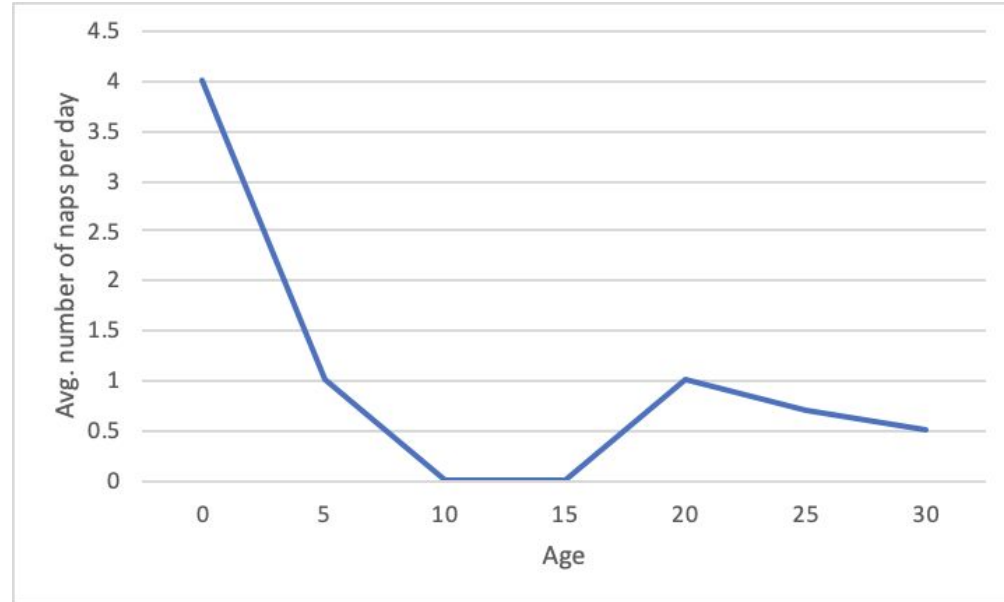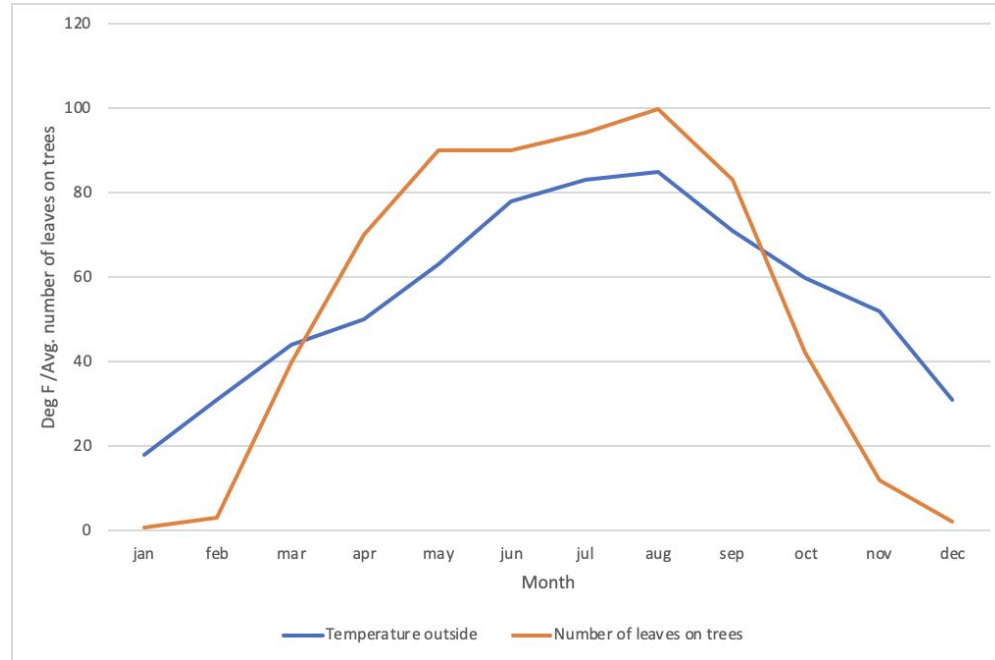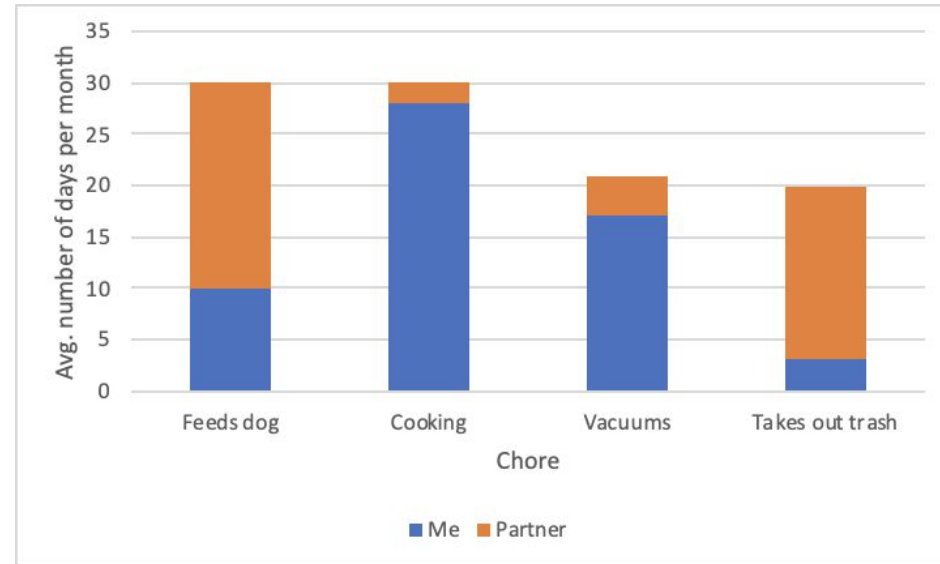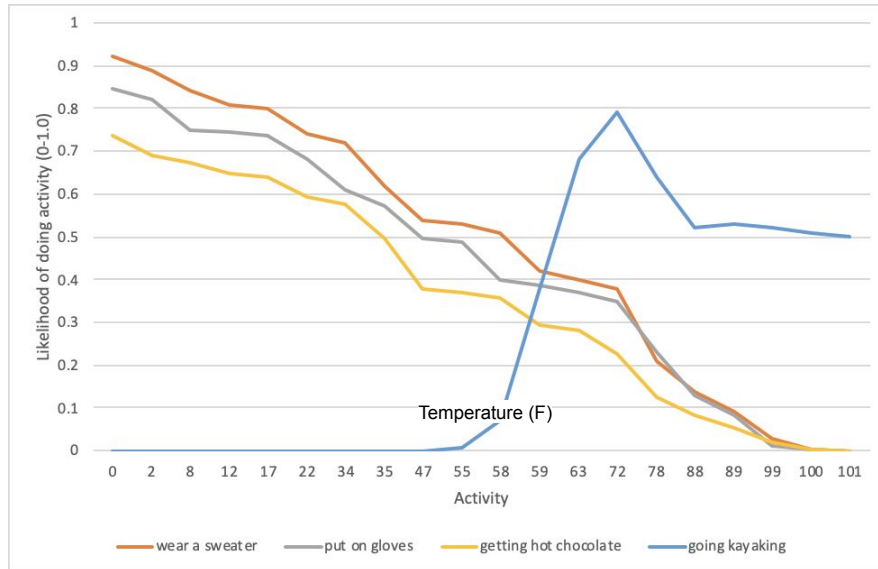
X goes against the trend

The remainder of the information in this presentation is adapted from work by Joanna Wolfe

Bar graph, line graph, or table

Line graph

Line graph

Pie chart or stacked bar graph

Who cooks dinner

- Me
- Partner
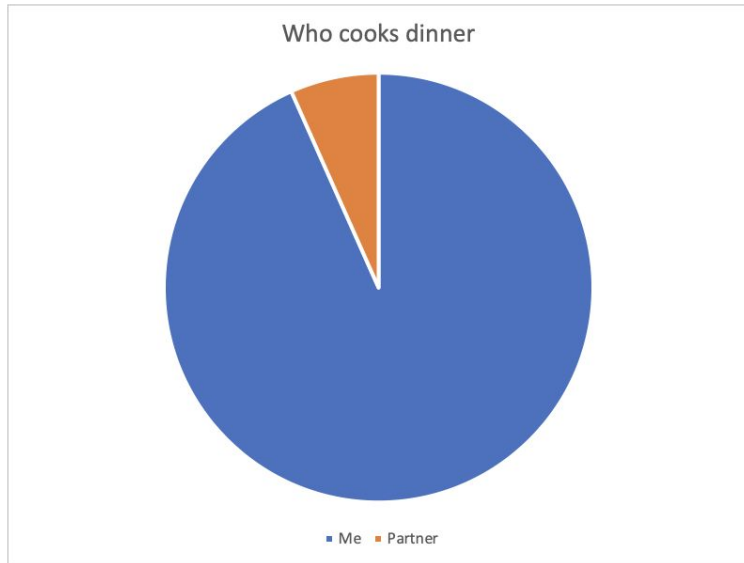
Line graph, scatter plot, or table
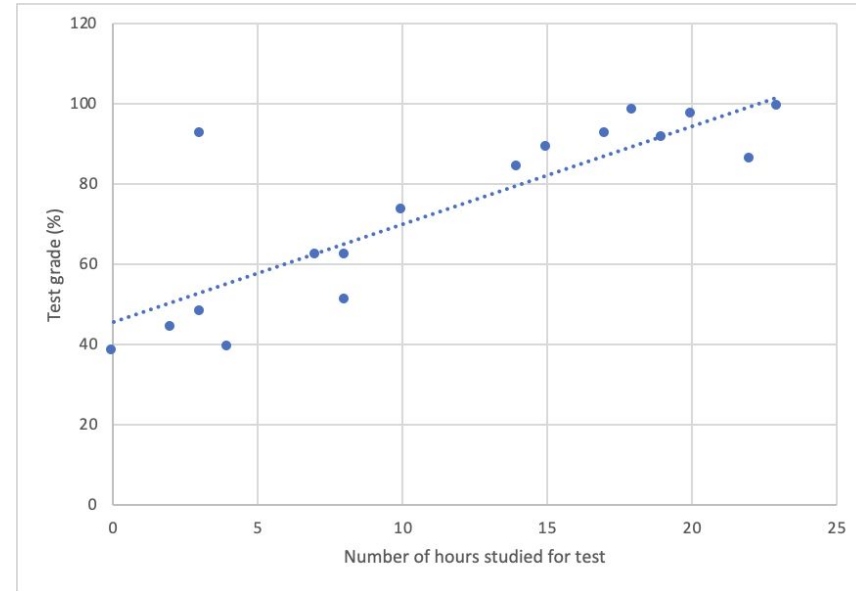
In sum:

Type of graph	Stories it can tell	Features
Bar chart	X is more/less/the same as Y X is more than Y is made up of Z parts (stacked bar)	+ Good for many variables - Stacked bar graphs are hard to read with more than 2 features
Line graph	X is more/less/the same as Y X is changing over time X goes against the trend	+ Most frequently used for change over time
Pie chart	X is a large/small/average part of the whole	- Hard to show minute differences - Sometimes seen as imprecise
Scatter plot	X goes against the trend	+ Trend lines tend to help pull out patterns
Table	X is more/less/the same as Y X goes against the trend	+ Good for conveying precise values & multiple types of data

6 basic principles of data visualization:

1. sort data to emphasize your story

2. group data to foreground one story over another

3. reduce non-data ink

4. minimize eye movement

5. use contrast to emphasize your story

6. be consistent in how you report numbers

1. Sort data to emphasize your story

State	Homicides per 100,000
Northern states	
Maine	1.6
Vermont	1.6
Connecticut	2.4
Rhode Island	2.4
New York	3.1
Western states	
Oregon	2.0
Washington	2.5
Colorado	2.8
California	4.4
Southern states	
Arkansas	5.6
Georgia	5.7
Florida	5.8
South Carolina	6.4
Louisiana	10.3

Table 1: 2014 Homicide rates of select US states per 100,000 residents

3. Reduce non-data ink

State	Homicides per 100,000
Northern states	
Maine	1.6
Vermont	1.6
Connecticut	2.4
Rhode Island	2.4
New York	3.1
Western states	
Oregon	2
Washington	2.5
Colorado	2.8
California	4
Southern states	
Arkansas	5.6
Georgia	5.7
Florida	5.8
South Carolina	6.4
Louisiana	10.3

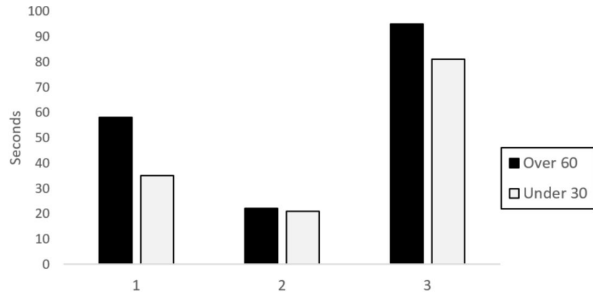| *State*	*Homicides per 100,000*
Northern states	
Maine	1.6
Vermont	1.6
Connecticut	2.4
Rhode Island	2.4
New York	3.1
Western states	
Oregon	2.0
Washington	2.5
Colorado	2.8
California	4.4
Southern states	
Arkansas	5.6
Georgia	5.7
Florida	5.8
South Carolina	6.4
Louisiana	10.3

Figure 4.5a: Times to complete three different tasks by adults over 60 years and under 30 years. Task 1 = Log in; Task 2 = Search for product; Task 3 = Complete order form.
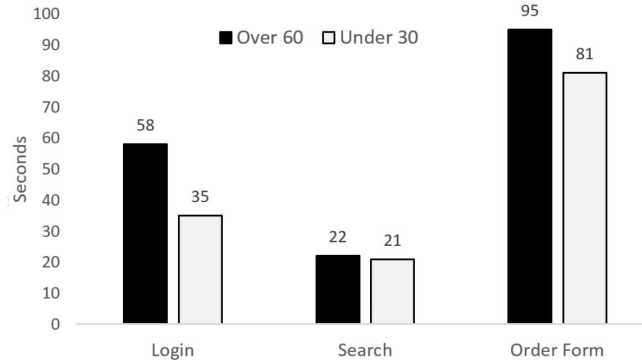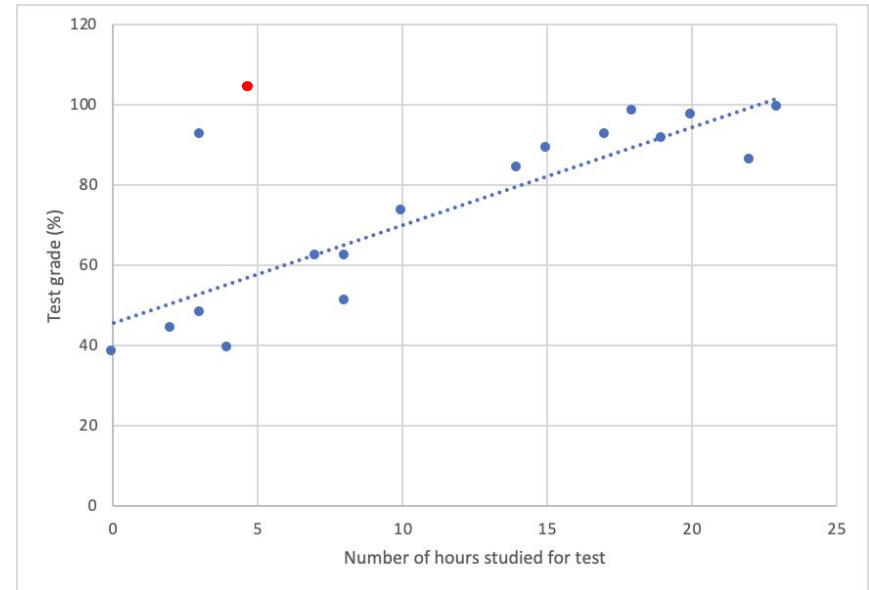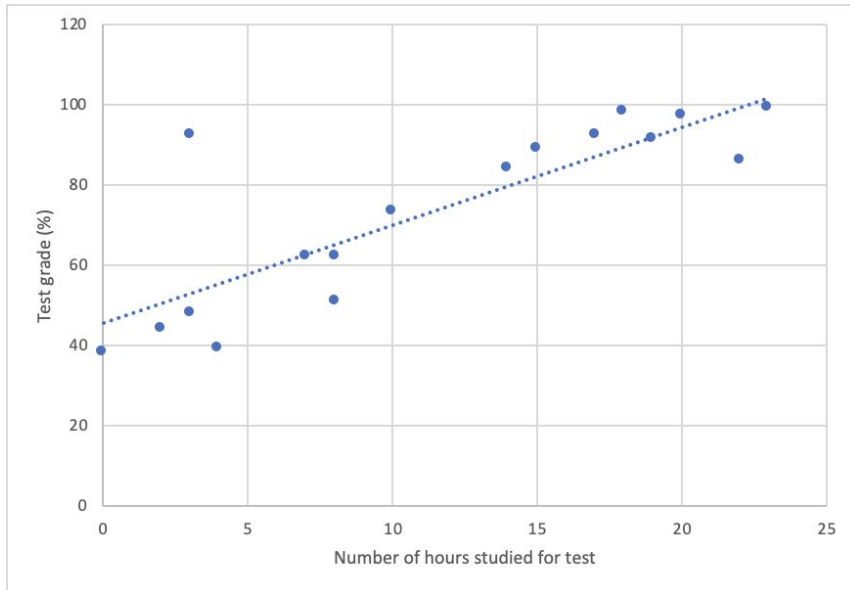

Figure 4.5b: Times to complete three different tasks by adults over 60 years under 30 years.

5. Use contrast to emphasize your story

6. Be consistent in how you report numbers

State	Homicides per 100,000
Northern states	
Maine	1.65
Vermont	1.6
Connecticut	2
Rhode Island	2.4
New York	3.19
Western states	
Oregon	2.1
Washington	2.542
Colorado	2.83
California	4
Southern states	
Arkansas	5.63
Georgia	5.7
Florida	6
South Carolina	6.43122
Louisiana	10.32

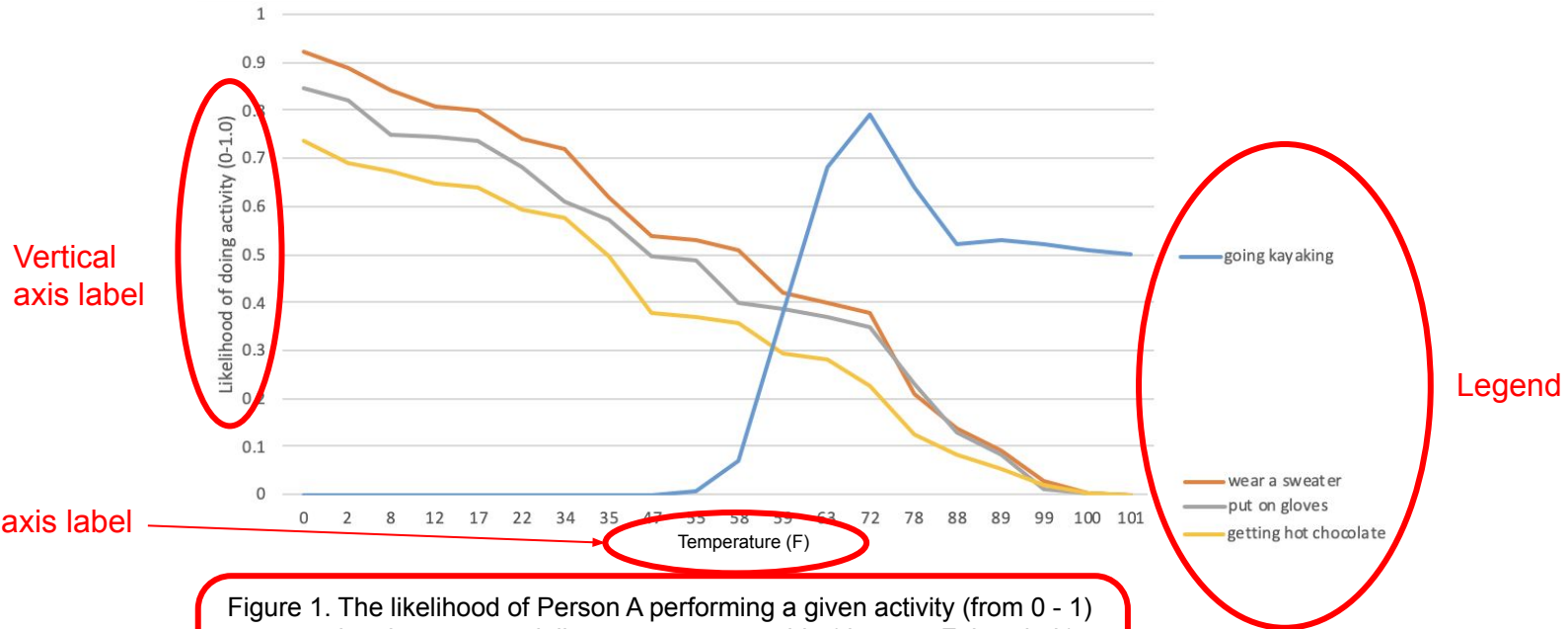| State	Homicides per 100,000
Northern states	
Maine	1.6
Vermont	1.6
Connecticut	2.4
Rhode Island	2.4
New York	3.1
Western states	
Oregon	2.1
Washington	2.5
Colorado	2.8
California	4.0
Southern states	
Arkansas	5.6
Georgia	5.7
Florida	5.8
South Carolina	6.4
Louisiana	10.3

Captions, labels, and legends

Figure 1. The likelihood of Person A performing a given activity (from 0 - 1) compared to the average daily temperature outside (degrees Fahrenheit). The likelihood of going kayaking is negatively correlated with winter activities including wearing a sweater, putting on gloves, and getting hot chocolate.

Want to learn more about communicating via data visualization and statistics?

- Cairo, Alberto. *How charts lie: Getting smarter about visual information*. WW Norton & Company, 2019.

- Huff, Darrell. *How to lie with statistics*. Penguin UK, 2023.

- https://www.storytellingwithdata.com/