SICSS

Data Storytelling and Visulization

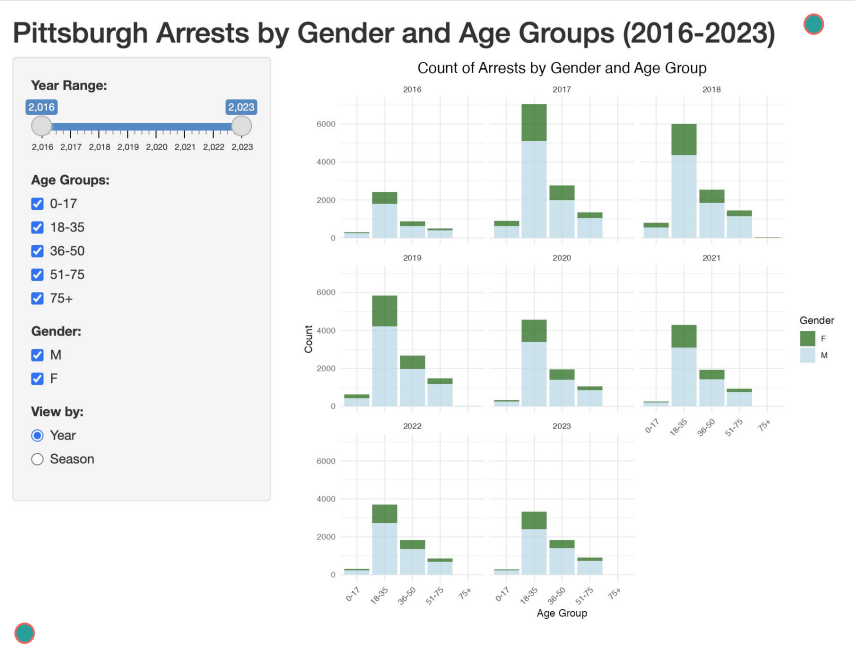
2025

Dr. Emma Slayton
Data Education Librarian
CMU Libraries

What We Will Cover

01. Introduction to Data Storytelling and Visualization Best Practices

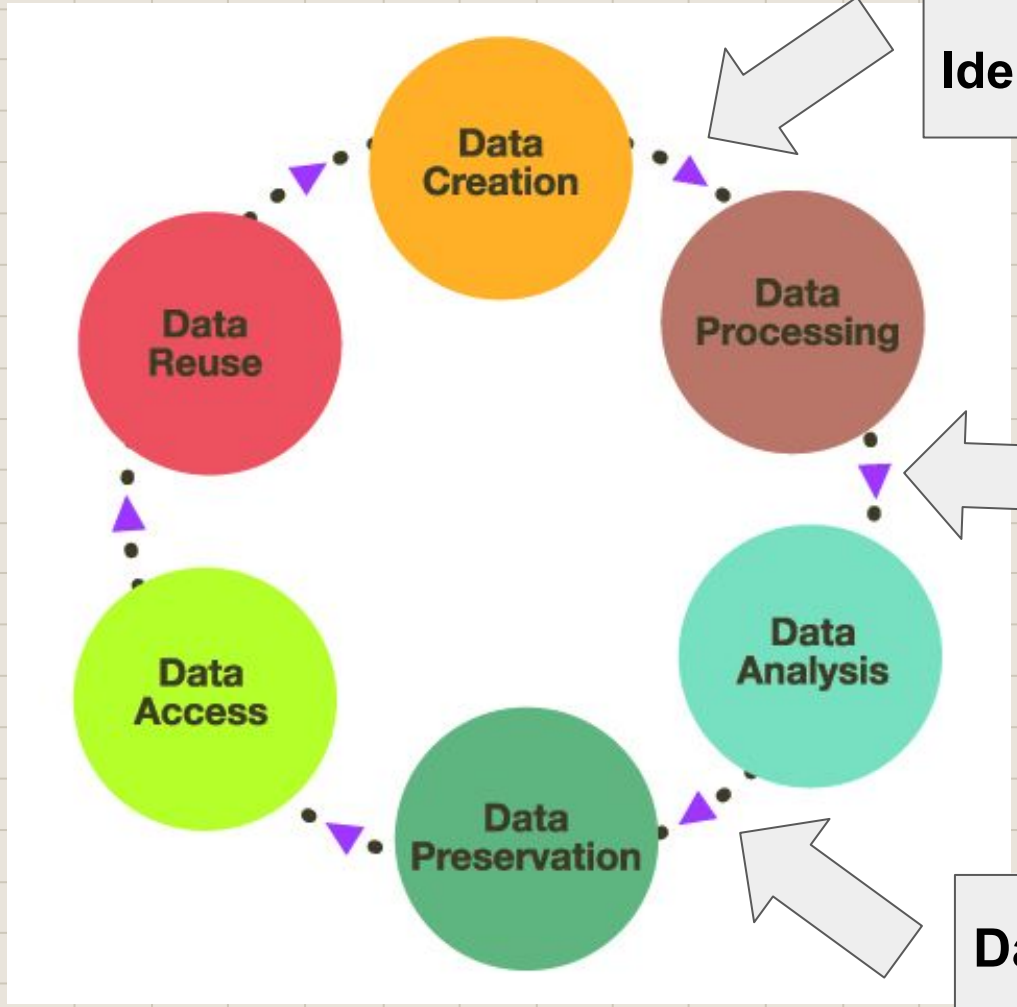
02. R Shiny Workshop



Data Storytelling

TIME: 2 hours

NEWS



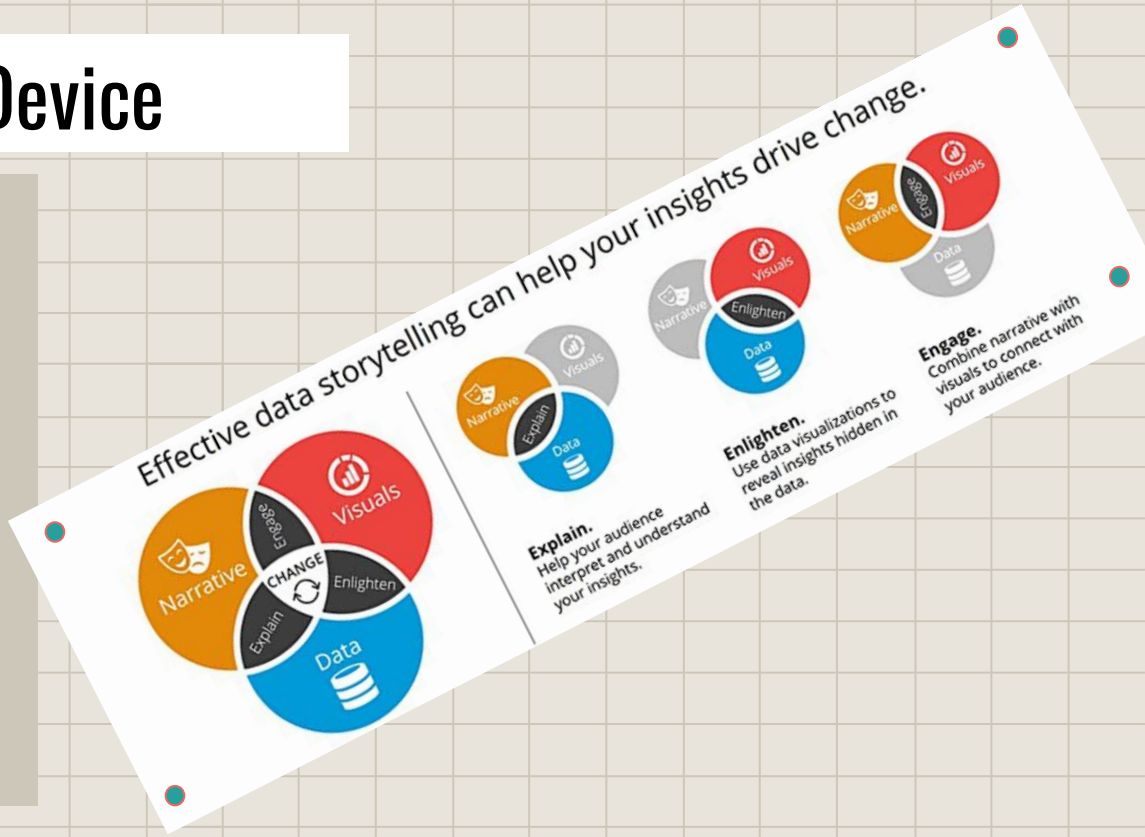
Identifying Story

Outlining Data Story

Data Storytelling

Why Use Stories as a Device

- People remember stories better than data, helps us recall and retell information
- Details of information in context, helps put data in order
- Repetition within a story frame, helps to keep keep points clear and memorable



What is your mission statement?

Identifying a research question that aligns with the data can go a long way to impacting data storytelling!

Age of individual can impact likely of crime reports

What is your conext?

What does your audience know about the subject? What do they need to know about the data?

The WPRDC provides access to police reporting on arrest data. This data was collected between 2016 and 2023.

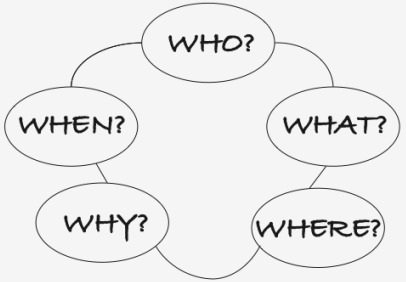
Who / What is your protagonist?

What is the core data variable/s that can best tell your story? What kinds of graphs do we need to tell those stories?

As our question deals with Age, we know we need to identify any demographic data that deals with that value.

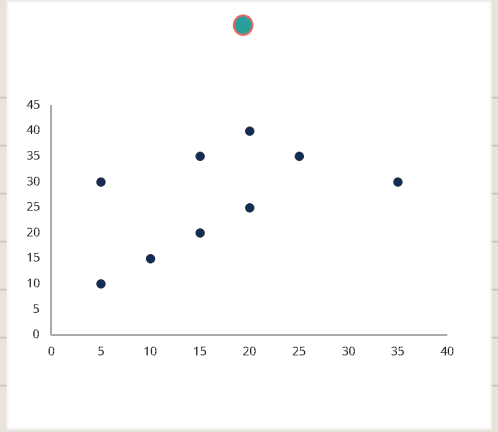
Understanding Data and Communicating Data

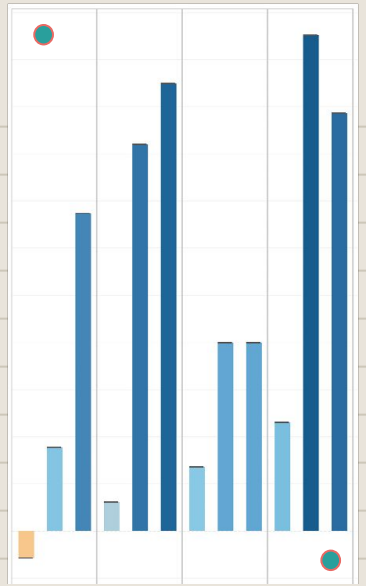
- **Making data a pivotal point in an overarching story**
  - Don’t just show data, use data
- **Break out of default assumptions**
  - What does your audience know?
  - What can they learn based on what they know?
- **What do you need to share, to share the point**
  - What data elements are essential, which distract
  - How can we showcase this information

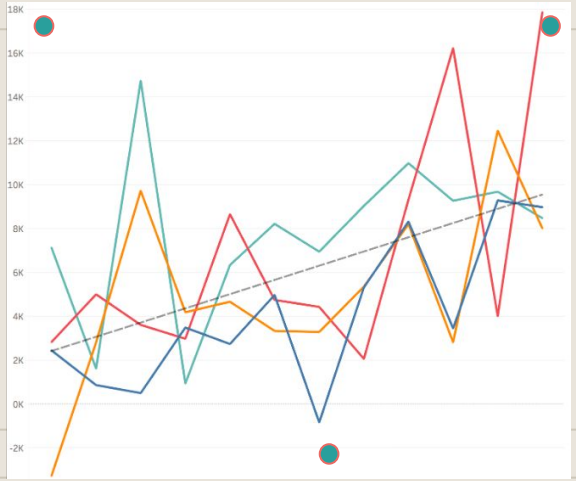


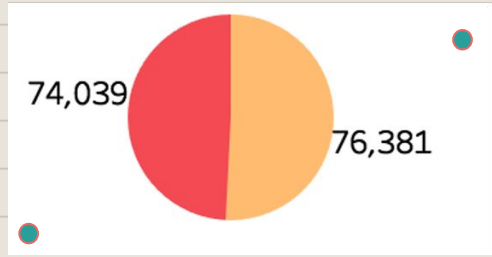
Data visualization

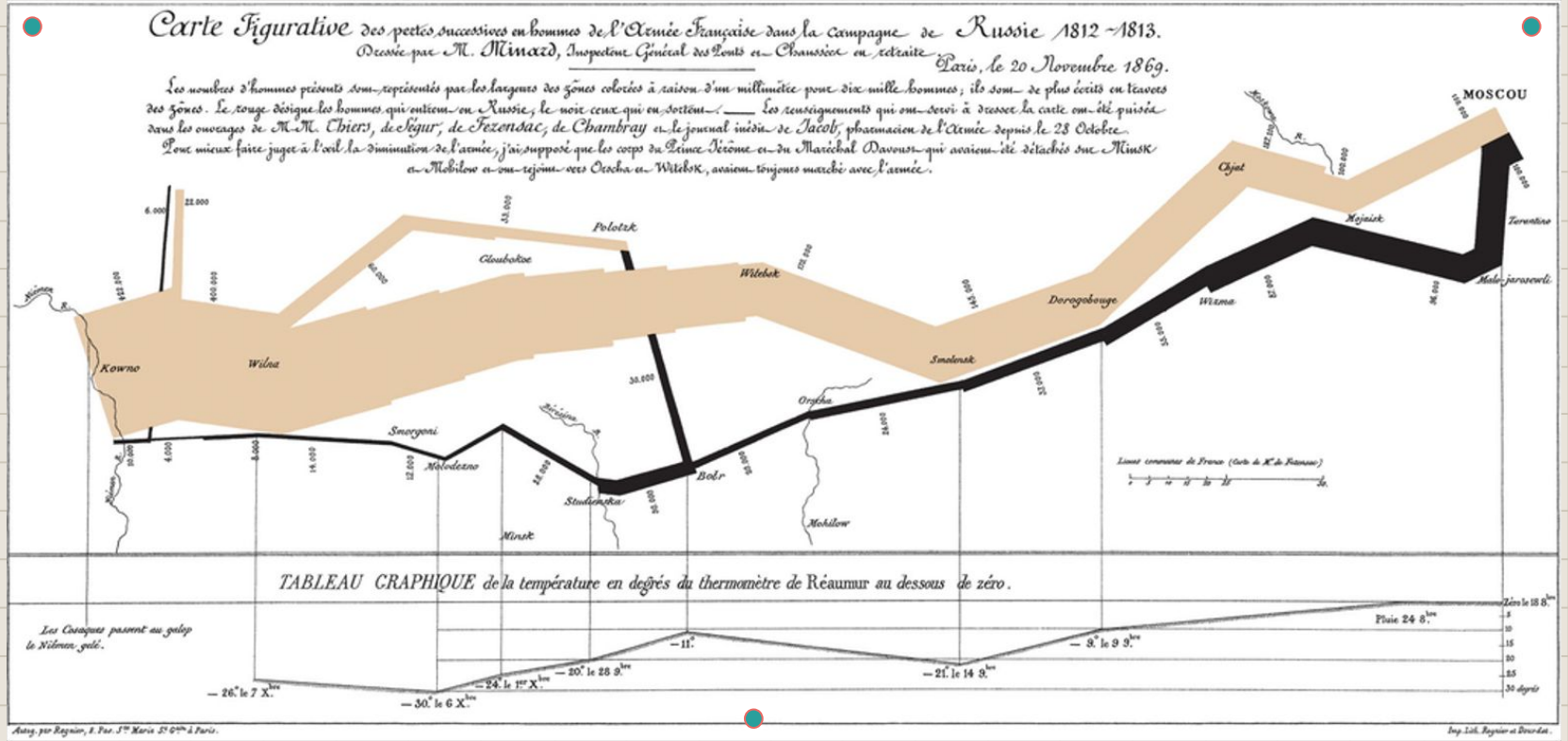
is the process of presenting information through visual means



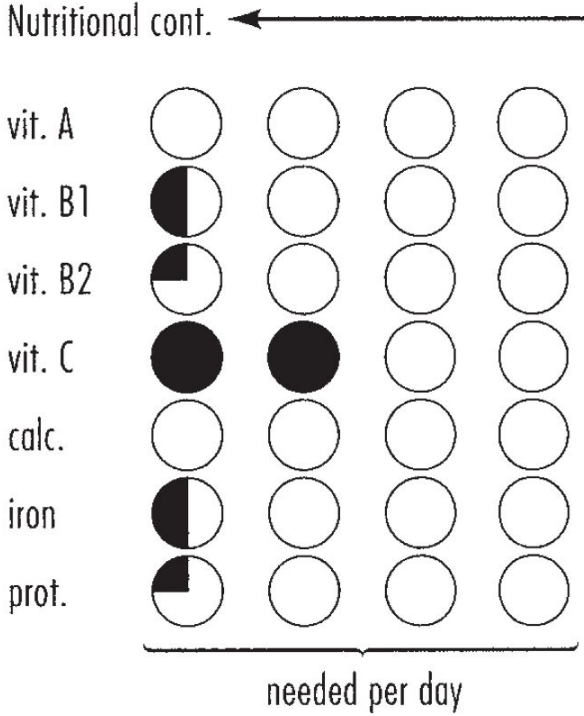


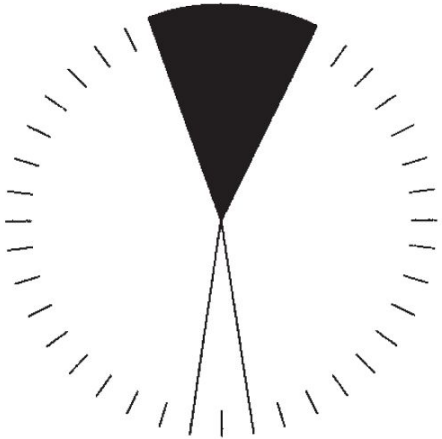





Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.
Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.
Paris, le 20 Novembre 1869.
Les nombres d'hommes présents sont représentés par les largeurs des zônes colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zônes. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M.M. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.
Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mohilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.
MOSCOU
Kowno
Wilna
Gloubokoe
Polotzk
Witebsk
Smolensk
Dorogobouge
Wizma
Chjat
Mojaisk
Terentino
Malo-jarosewli
Smorgoni
Molodezno
Studienska
Bobr
Orscha
Mohilow
Minsk
Niémen R.
Bérésina R.
Moskowa R.
Lieues communes de France (Carte de M. de Fezensac)
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.
Les Cosaques passent au galop le Niémen gelé.
— 26.° le 7 X.bre
— 30.° le 6 X.bre
— 24.° le 1.er X.bre
— 20.° le 28 9.bre
— 11.°
— 21.° le 14 9.bre
— 9.° le 9 9.bre
Pluie 24 8.bre
Zéro le 18 8.bre
30 degrés
Autog. par Regnier, 8. Pas. S.te Marie S.t G.ain à Paris.
Imp. Lith. Regnier et Dourdet.

Nutritional Information per Serving





SERVING SIZE	ONE CUP
SERVINGS PER CONTAINER	8

protein	4 g
fat	0 g
carbohy.	38 g

170 kilocalorie

Nutrition Facts

Servings Per Container 1

Serving Size **1 bottle (414 mL)**

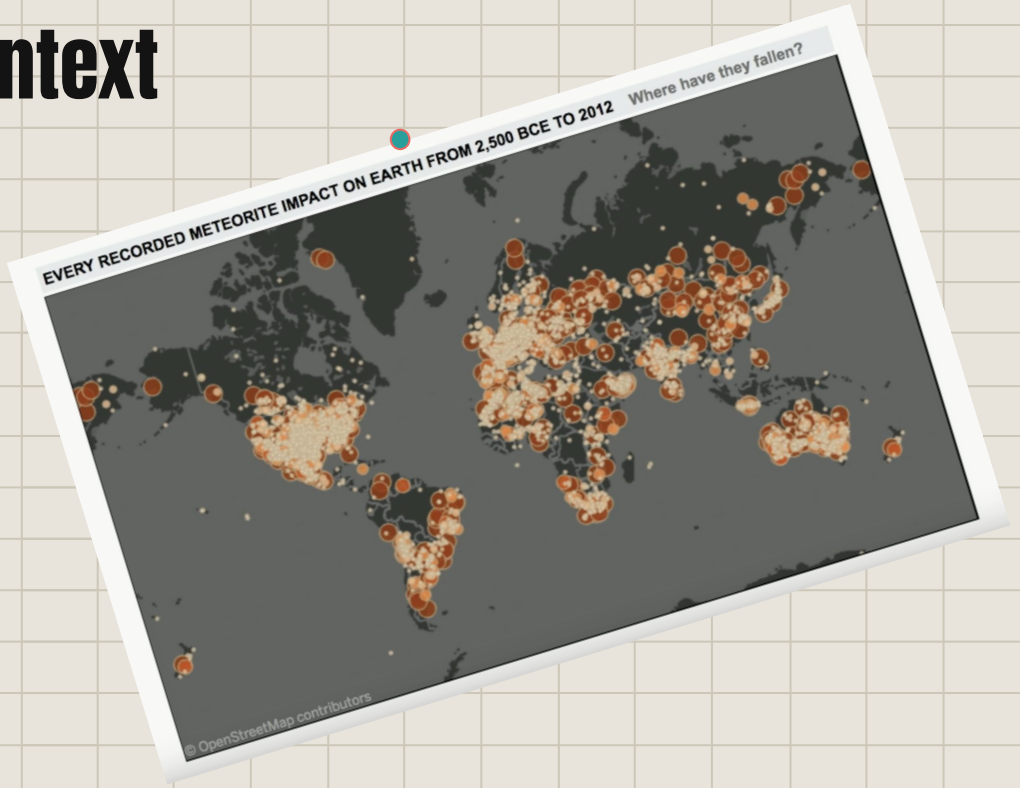
Amount Per Serving

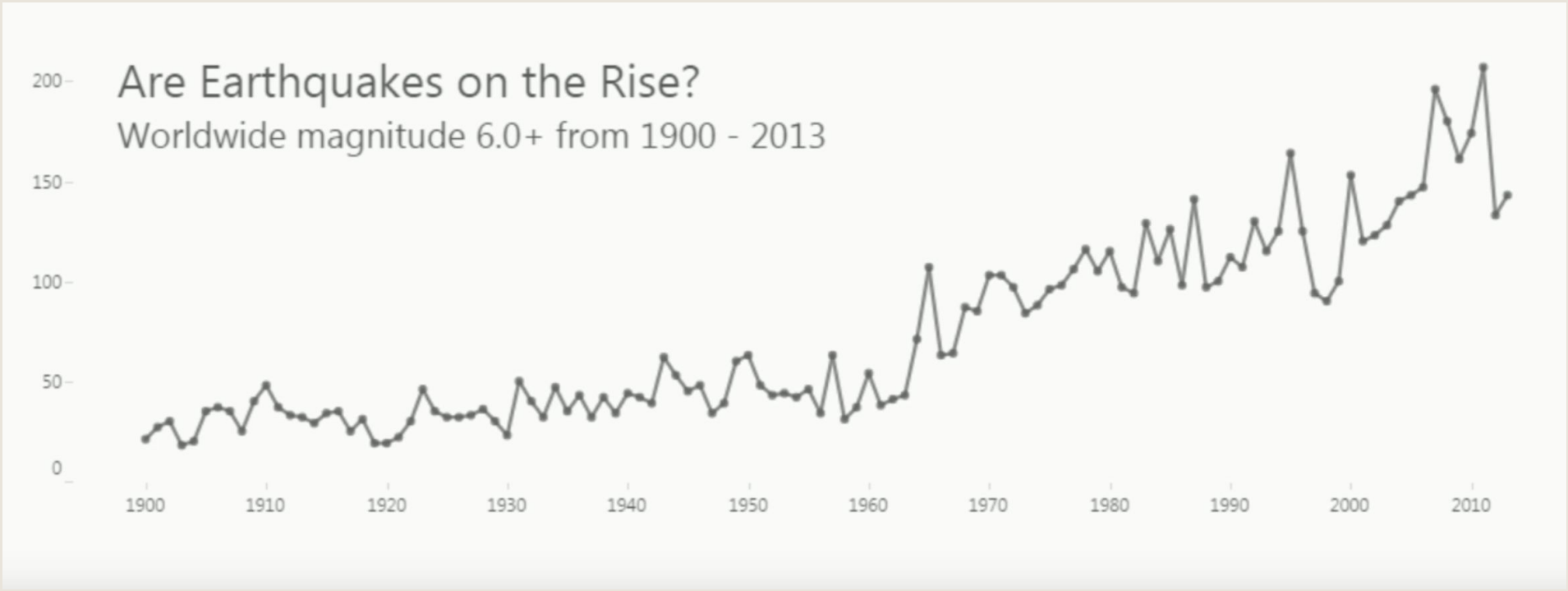
Calories	**400**
	% Daily Value*
Total Fat 21g	**32%**
Saturated Fat 2g	**10%**
Polyunsaturated Fat 2.5g	
Monounsaturated Fat 16g	
Trans Fat 0g	
Cholesterol 0mg	**0%**
Sodium 300mg	**13%**
Total Carbohydrate 37g	**12%**
Dietary Fiber 3g	**12%**
Soluble Fiber 1g	
Total Sugars 9g	
Includes 9g Added Sugars	**18%**
Protein 20g	

Think about provided context

What would it hurt your audience to not know about the data you are visualizing before it is in a graph?

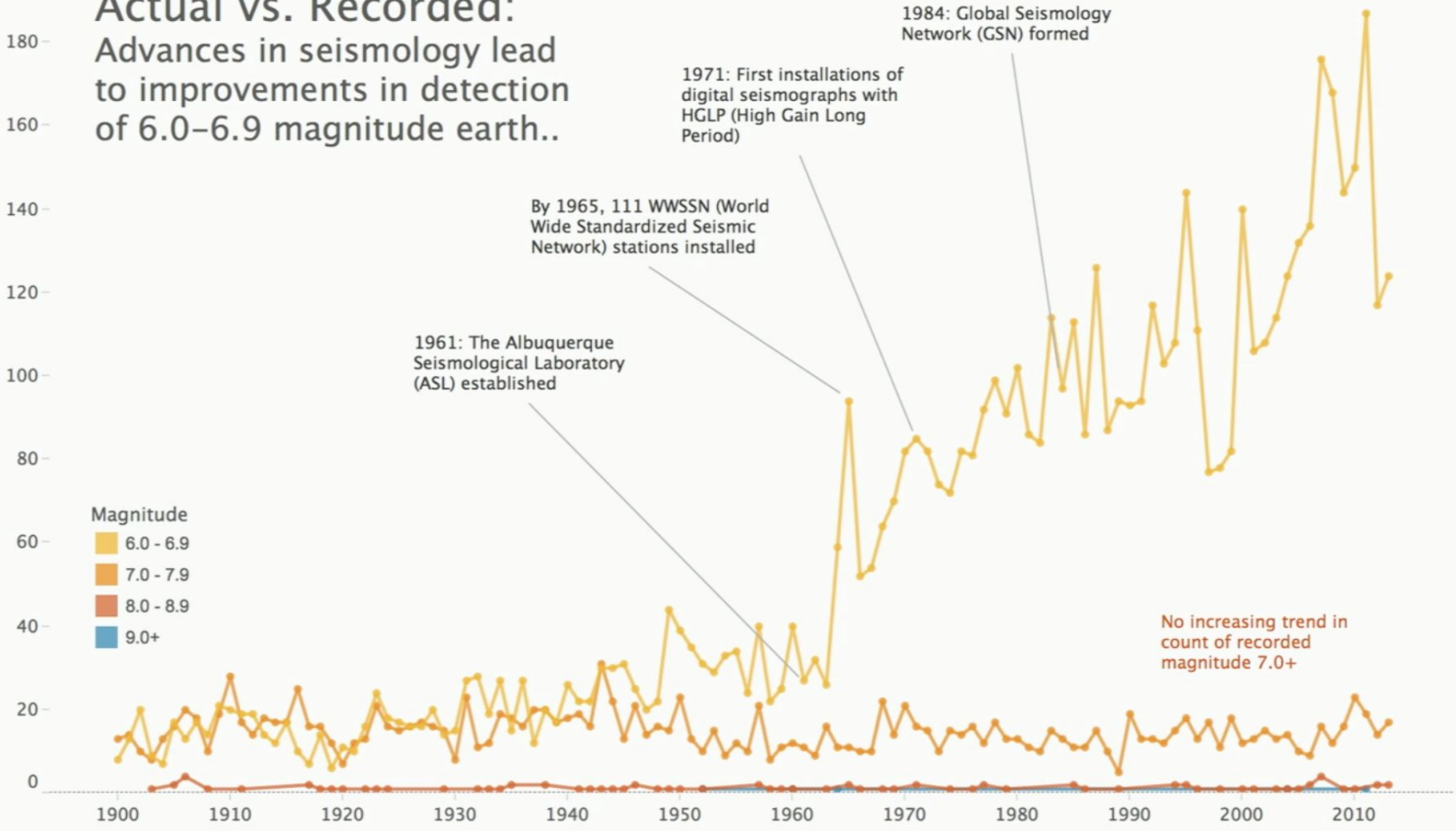
What clues can we give to our audience?




Are Earthquakes on the Rise?
Worldwide magnitude 6.0+ from 1900 - 2013
200
150
100
50
0
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010

Actual vs. Recorded:

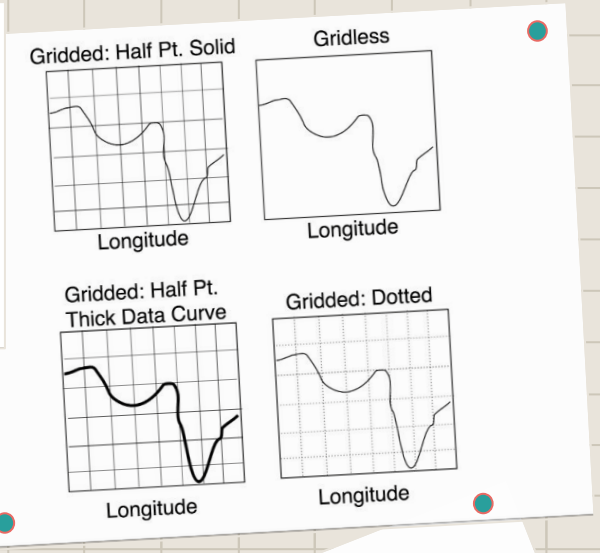
Advances in seismology lead to improvements in detection of 6.0-6.9 magnitude earth..



Sources | Data: http://earthquake.usgs.gov/earthquakes/search/, Dates: http://pubs.usgs.gov/fs/2011/3065/pdf/FS11-3065.pdf

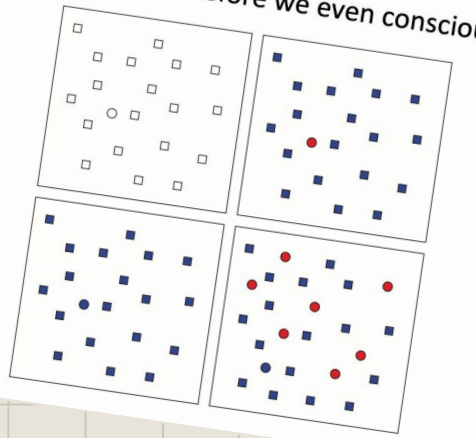
Tips and Tricks

$$\begin{aligned} \textit{data-ink ratio} &= \frac{\textit{data-ink}}{\textit{total ink used to print the graphic}} \\ &= \textit{the proportion of a graphic's ink devoted to the non-redundant display of data-information} \\ &= 1.0 - \textit{proportion of a graphic that can be erased without loss of data-information} \end{aligned}$$

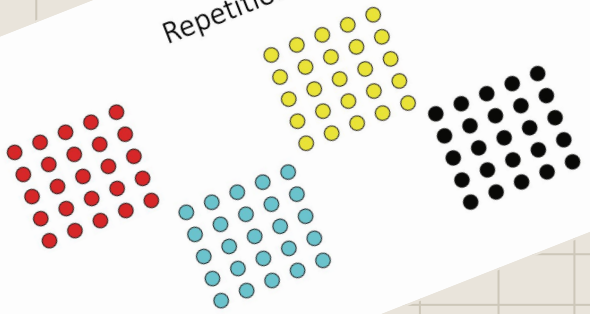


Pre-Perception in visualizations

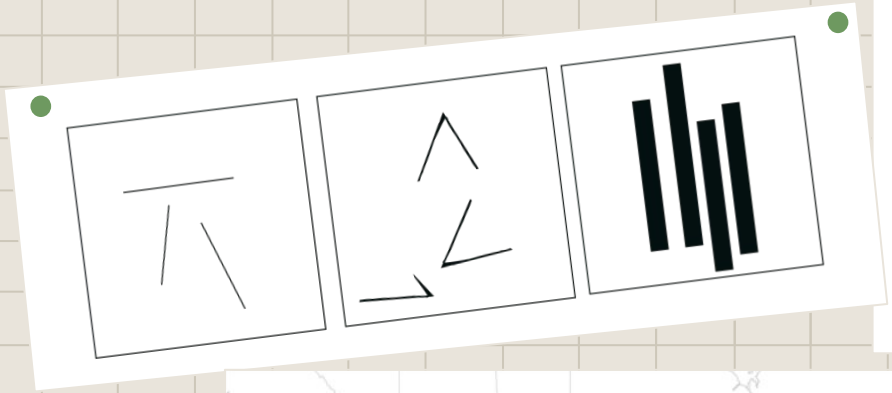
What we see and understand before we even consciously read a graph

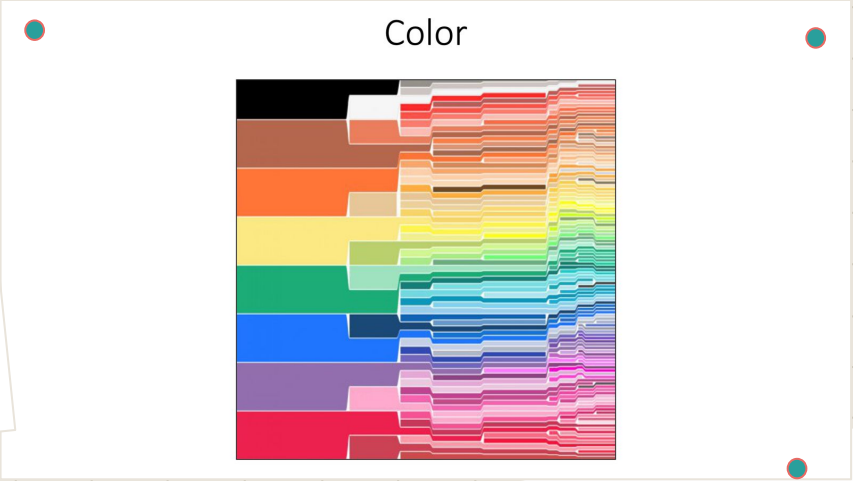


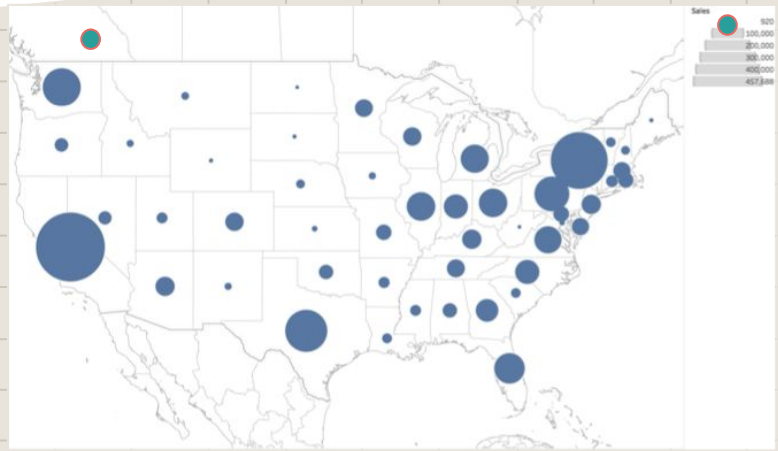
Repetition

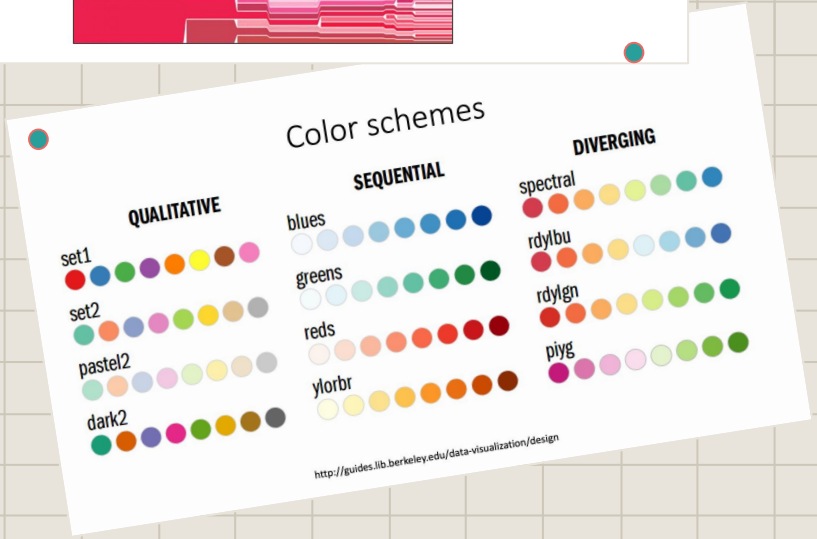


Tips and Tricks







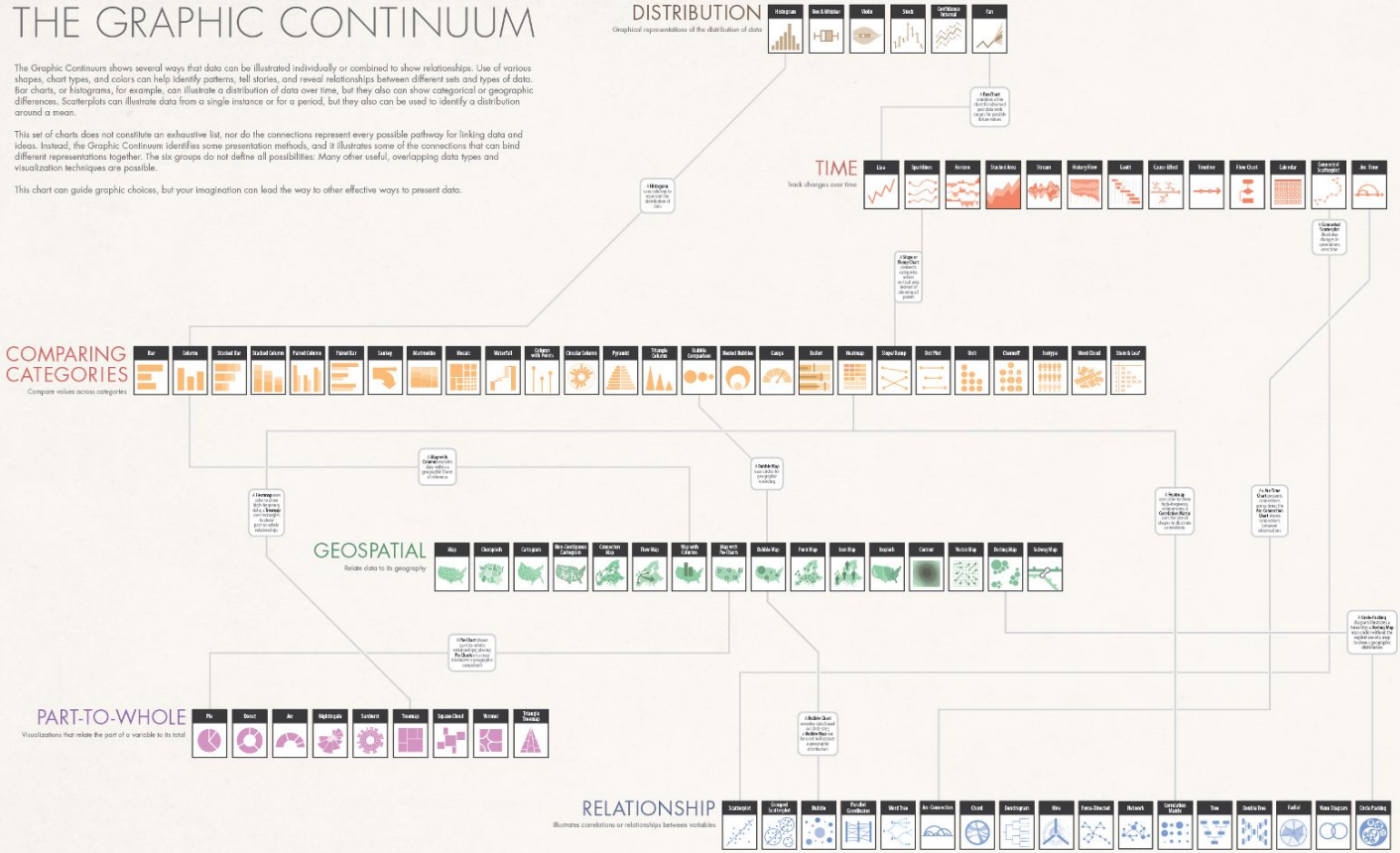


THE GRAPHIC CONTINUUM

The Graphic Continuum shows several ways that data can be illustrated individually or combined to show relationships. Use of various shapes, chart types, and colors can help identify patterns, tell stories, and reveal relationships between different sets and types of data. Bar charts, or histograms, for example, can illustrate a distribution of data over time, but they also can show categorical or geographic differences. Scatterplots can illustrate data from a single instance or for a period, but they also can be used to identify a distribution around a mean.

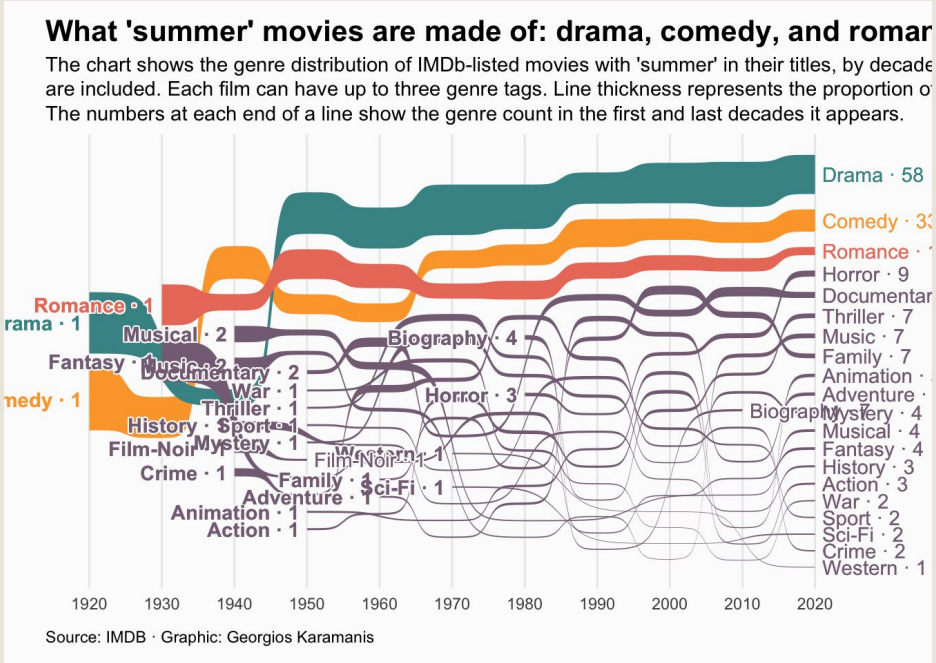
This set of charts does not constitute an exhaustive list, nor do the connections represent every possible pathway for linking data and ideas. Instead, the Graphic Continuum identifies some presentation methods, and it illustrates some of the connections that can bind different representations together. The six groups do not define all possibilities: Many other useful, overlapping data types and visualization techniques are possible.

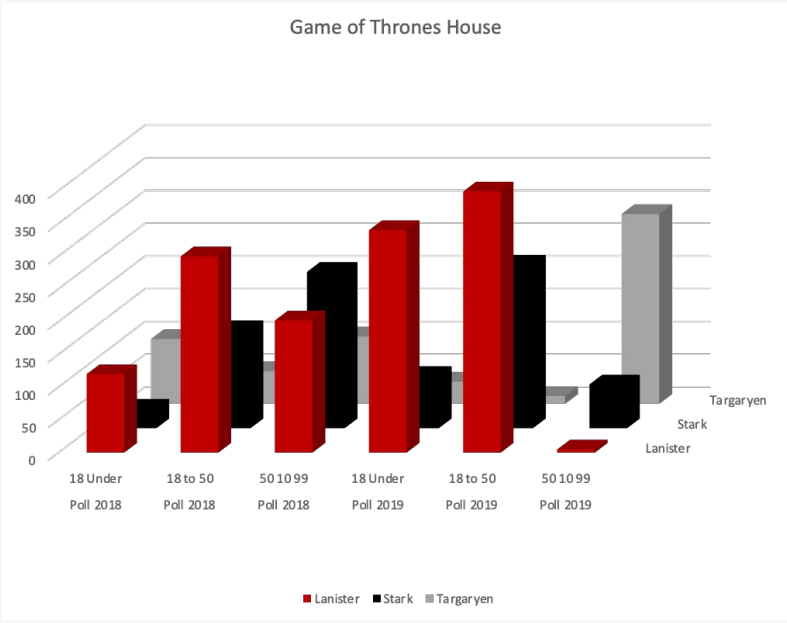
This chart can guide graphic choices, but your imagination can lead the way to other effective ways to present data.



© Jonathan Schwabish & Severino Ribecca

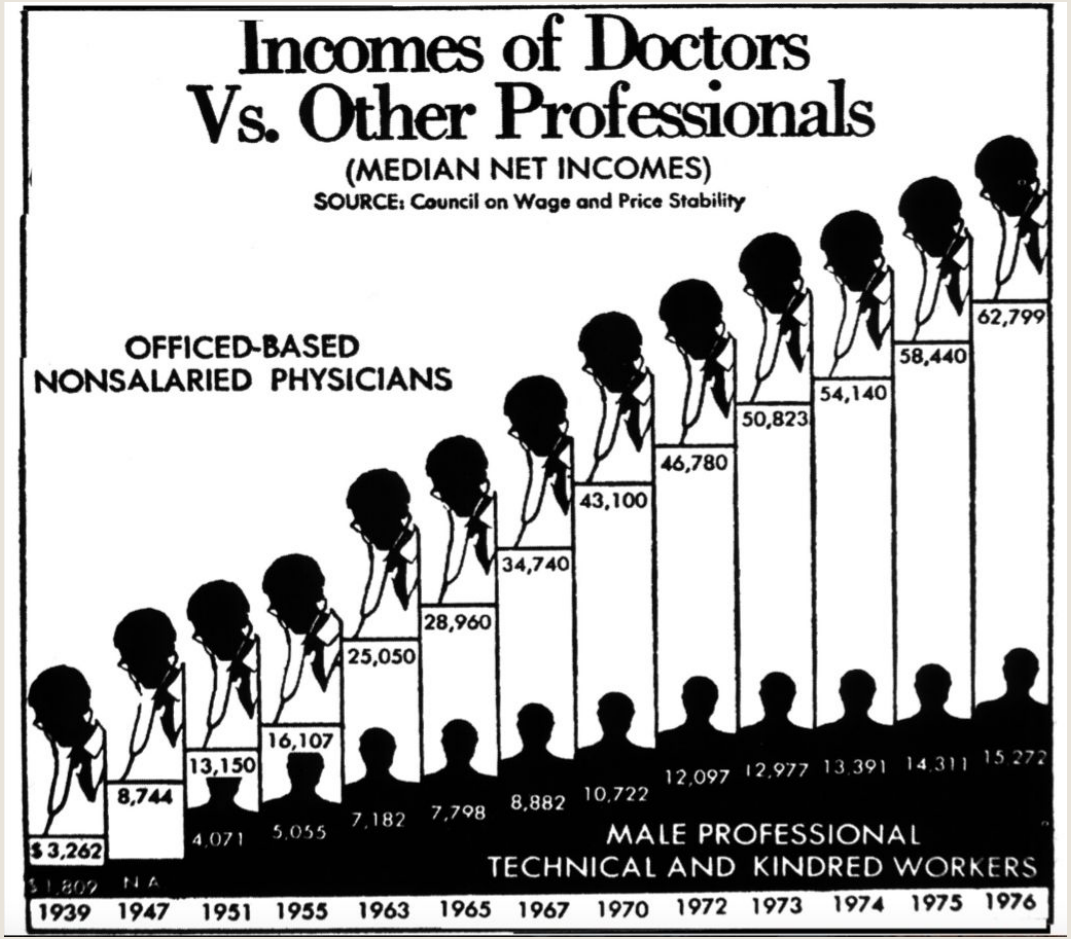
What elements do you see?





https://r-graph-gallery.com/best-r-chart-examples.html

Form Vs Function


Incomes of Doctors
Vs. Other Professionals
(MEDIAN NET INCOMES)
SOURCE: Council on Wage and Price Stability
OFFICED-BASED
NONSALARIED PHYSICIANS
$3,262
8,744
13,150
16,107
25,050
28,960
34,740
43,100
46,780
50,823
54,140
58,440
62,799
$1,802
N.A.
4,071
5,055
7,182
7,798
8,882
10,722
12,097
12,977
13,391
14,311
15,272
MALE PROFESSIONAL
TECHNICAL AND KINDRED WORKERS
1939
1947
1951
1955
1963
1965
1967
1970
1972
1973
1974
1975
1976

The Big Cheese

Papa John's founder John Schnatter has been the dominant presence at his company since its founding—even after turning over the CEO role to a rotating cast of executives

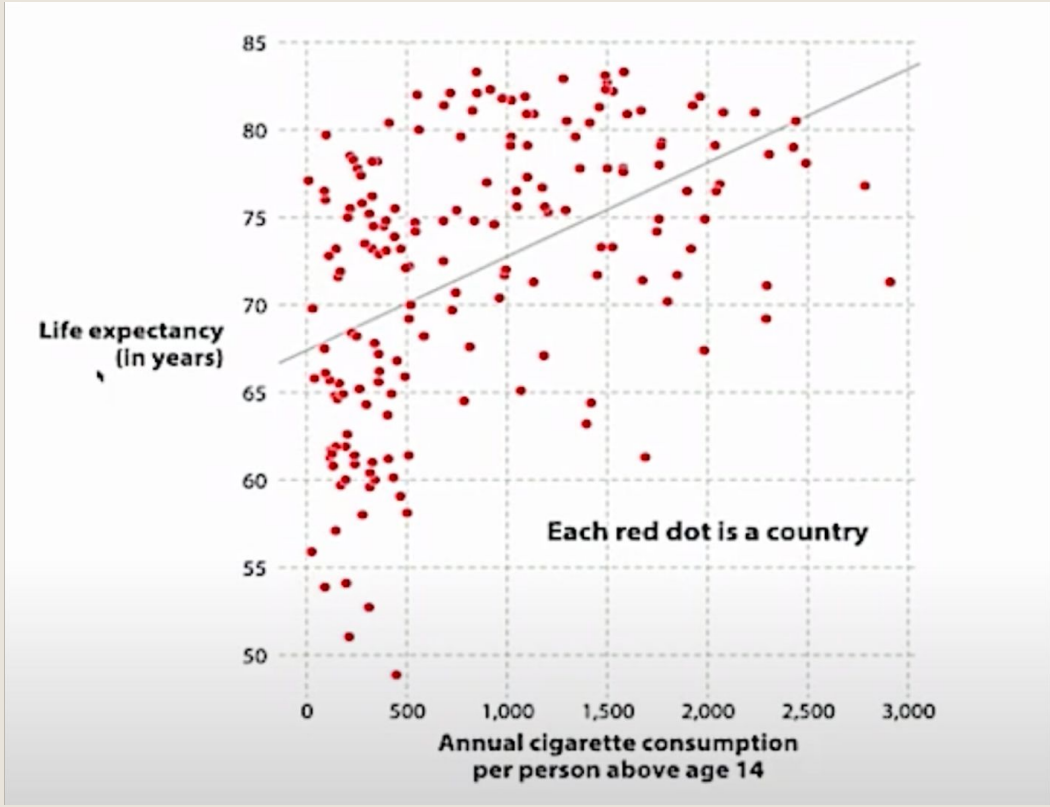
Steve Ritchie
January 2018–Present

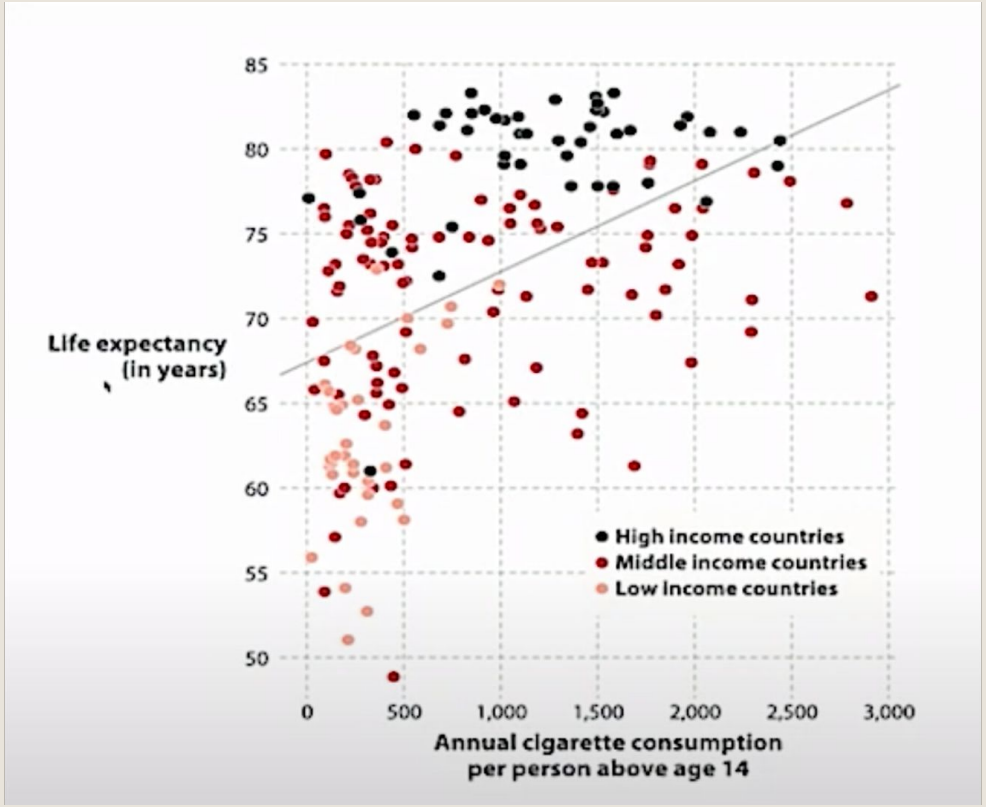
Jude Thompson
Co-CEO April 2010–April 2011

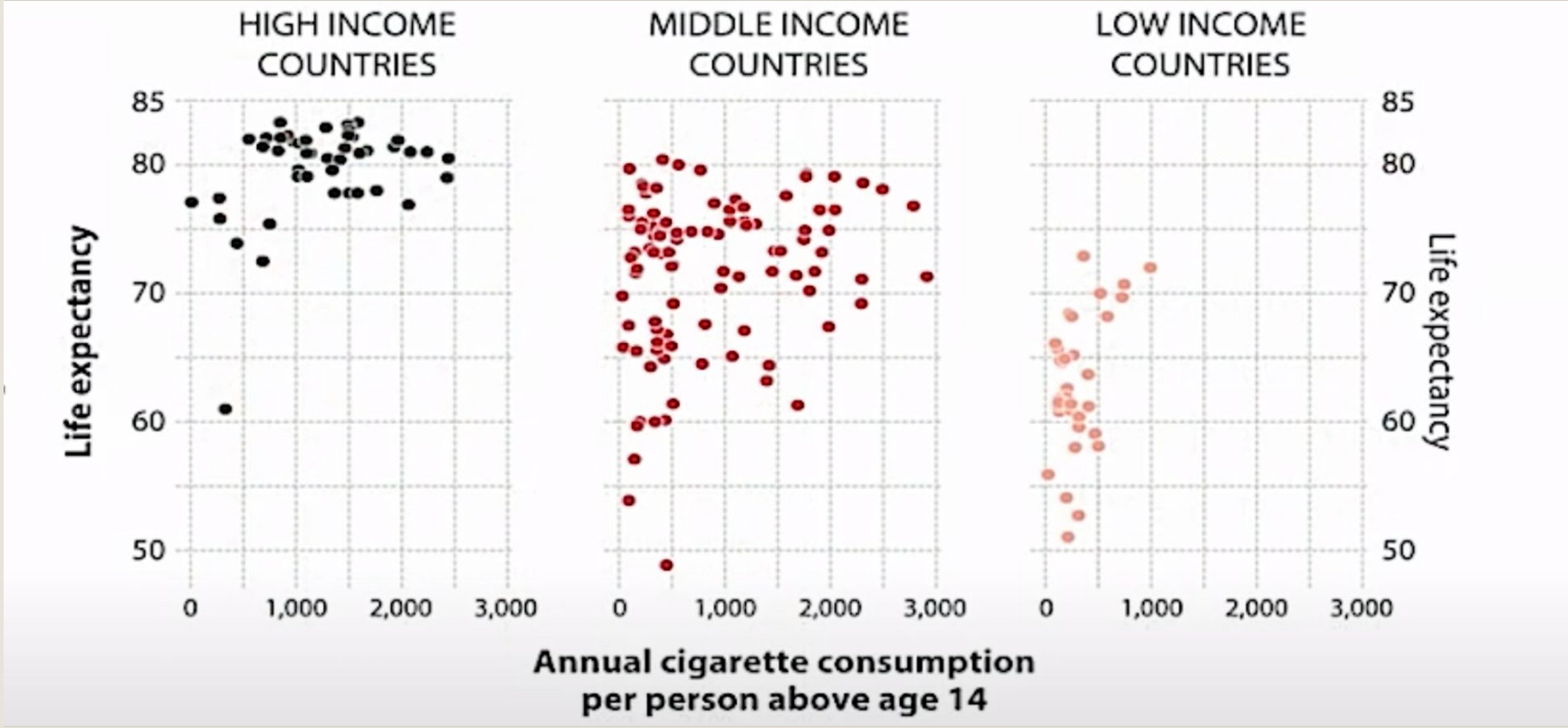
John Schnatter
December 2008–December 2017

Nigel Travis
January 2005–December 2008

John Schnatter
1990–January 2005

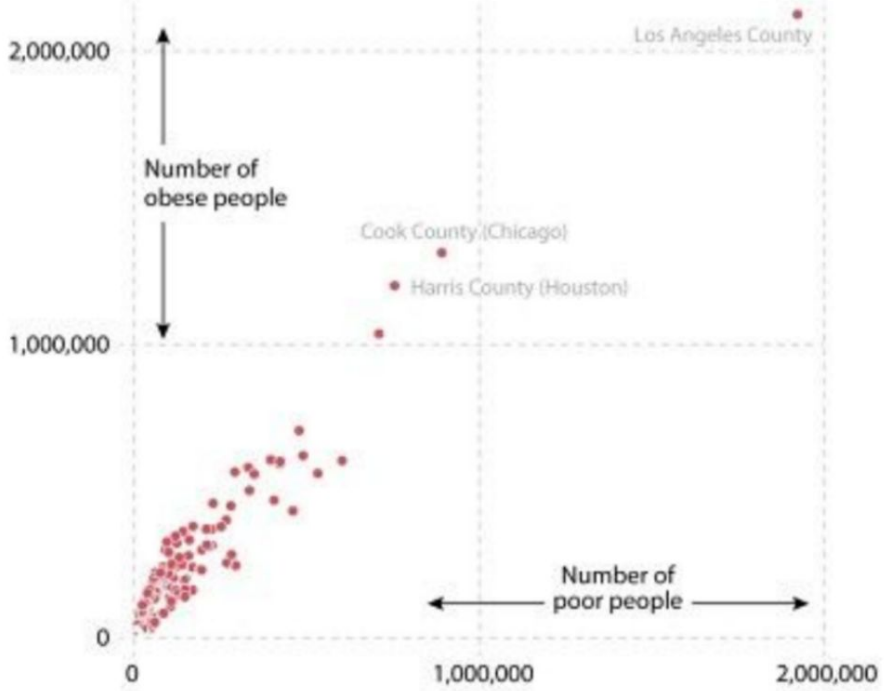

Life expectancy (in years)
85
80
75
70
65
60
55
50
0
500
1,000
1,500
2,000
2,500
3,000
Annual cigarette consumption per person above age 14
Each red dot is a country


Life expectancy (in years)
85
80
75
70
65
60
55
50
0
500
1,000
1,500
2,000
2,500
3,000
Annual cigarette consumption per person above age 14
High income countries
Middle income countries
Low income countries


HIGH INCOME COUNTRIES
MIDDLE INCOME COUNTRIES
LOW INCOME COUNTRIES
Life expectancy
85
80
70
60
50
0
1,000
2,000
3,000
Annual cigarette consumption per person above age 14

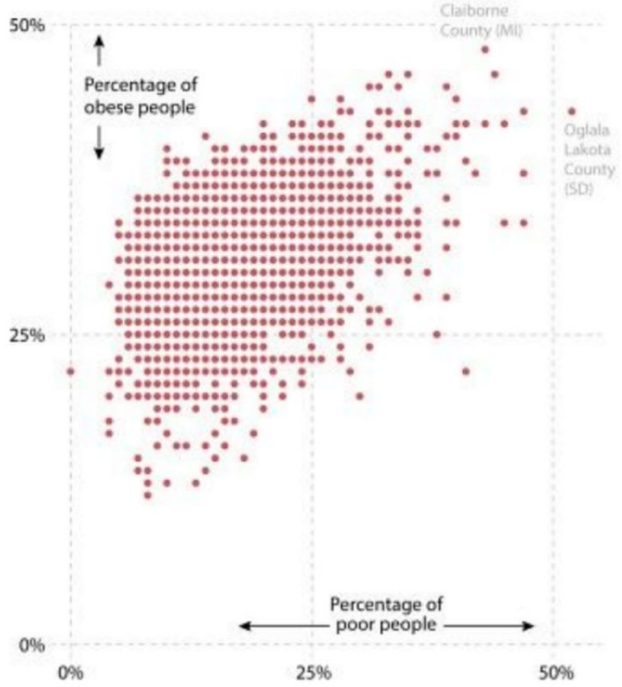
Poverty vs. obesity

(Each dot is a county.)



Poverty vs. obesity

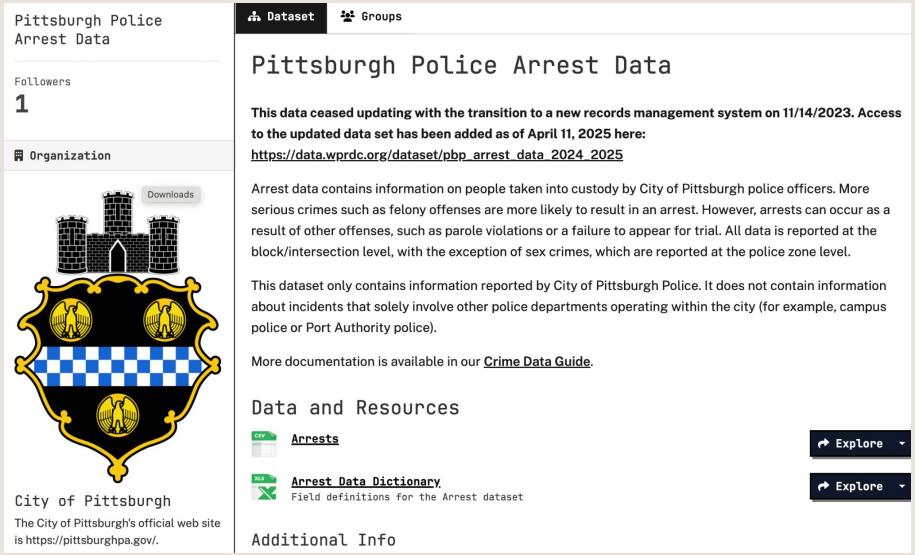
(Each dot is a county.)

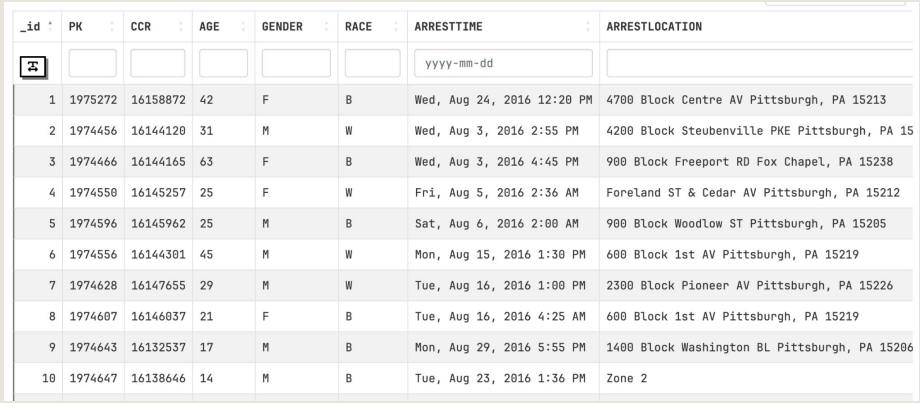


Let's Practice Graph Storytelling

In groups of **2 - 3** design either by hand or in ggplot graphs that focus on a story in this data set.

Include a **research question**, a **data protagonist**, and **needed context**!





_id	PK	CCR	AGE	GENDER	RACE	ARRESTTIME	ARRESTLOCATION
						yyyy-mm-dd	
1	1975272	16158872	42	F	B	Wed, Aug 24, 2016 12:20 PM	4700 Block Centre AV Pittsburgh, PA 15213
2	1974456	16144120	31	M	W	Wed, Aug 3, 2016 2:55 PM	4200 Block Steubenville PKE Pittsburgh, PA 15
3	1974466	16144165	63	F	B	Wed, Aug 3, 2016 4:45 PM	900 Block Freeport RD Fox Chapel, PA 15238
4	1974550	16145257	25	F	W	Fri, Aug 5, 2016 2:36 AM	Foreland ST & Cedar AV Pittsburgh, PA 15212
5	1974596	16145962	25	M	B	Sat, Aug 6, 2016 2:00 AM	900 Block Woodlow ST Pittsburgh, PA 15205
6	1974556	16144301	45	M	W	Mon, Aug 15, 2016 1:30 PM	600 Block 1st AV Pittsburgh, PA 15219
7	1974628	16147655	29	M	W	Tue, Aug 16, 2016 1:00 PM	2300 Block Pioneer AV Pittsburgh, PA 15226
8	1974607	16146037	21	F	B	Tue, Aug 16, 2016 4:25 AM	600 Block 1st AV Pittsburgh, PA 15219
9	1974643	16132537	17	M	B	Mon, Aug 29, 2016 5:55 PM	1400 Block Washington BL Pittsburgh, PA 15206
10	1974647	16138646	14	M	B	Tue, Aug 23, 2016 1:36 PM	Zone 2

https://shorturl.at/CF7yy

R SHINY

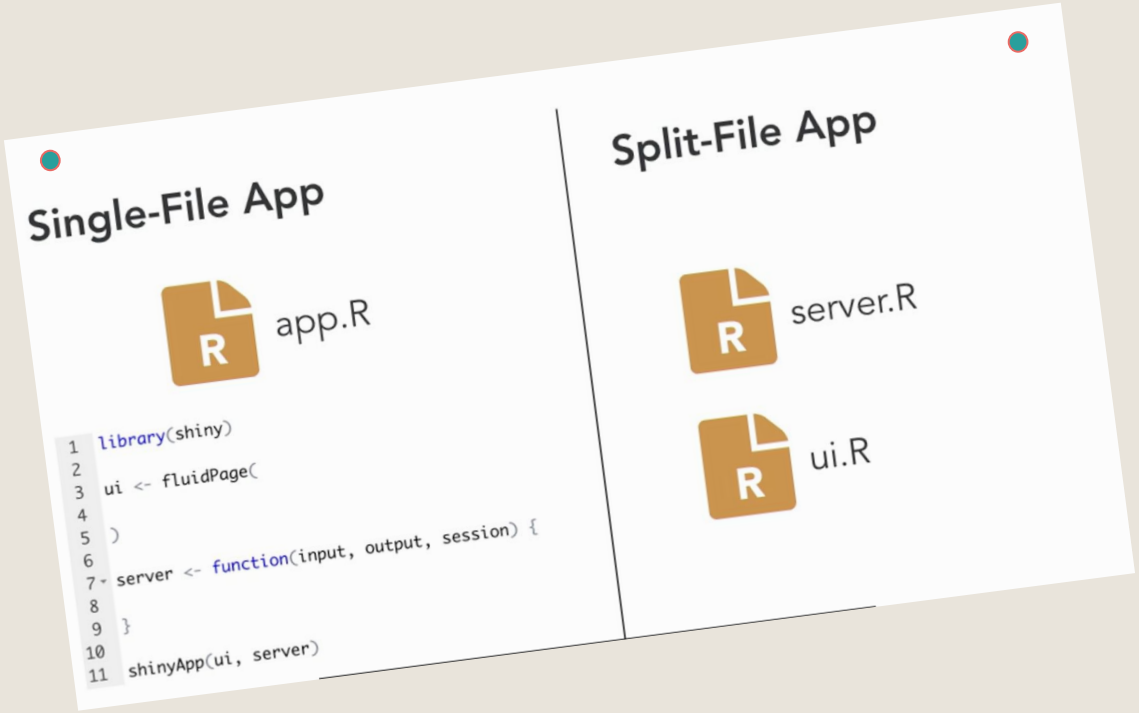
TIME: 1 hour

What is R Shiny?

Shiny is an R package that makes it easy to build interactive web applications (apps) straight from R

Shiny Apps can be hosted locally, making the easy to access for anyone with R!

R Shiny is dependent on two elements, the **ui** and the **server** code.



Wrap Up

Want Help? Contact me to attend a consultation:
eslayton@andrew.cmu.edu