PROJECT OVERVIEW, DATA VIZ

Jack Hester, MPH

CEPC 0904

Summer 2022

CORRECTIONS/NOTES

- Pre-recorded video had wrong % expected within SD ranges
- Boxplot summary slide may have been a bit confusing
- Both of these things addressed in Slack
- For small sample sizes (n<30) best practice is to use t over z test, BUT…
  - It's often an even better bet to use nonparametric tests, especially if the distribution is a little wack
  - (Question on the quiz is correct but I might throw it out)

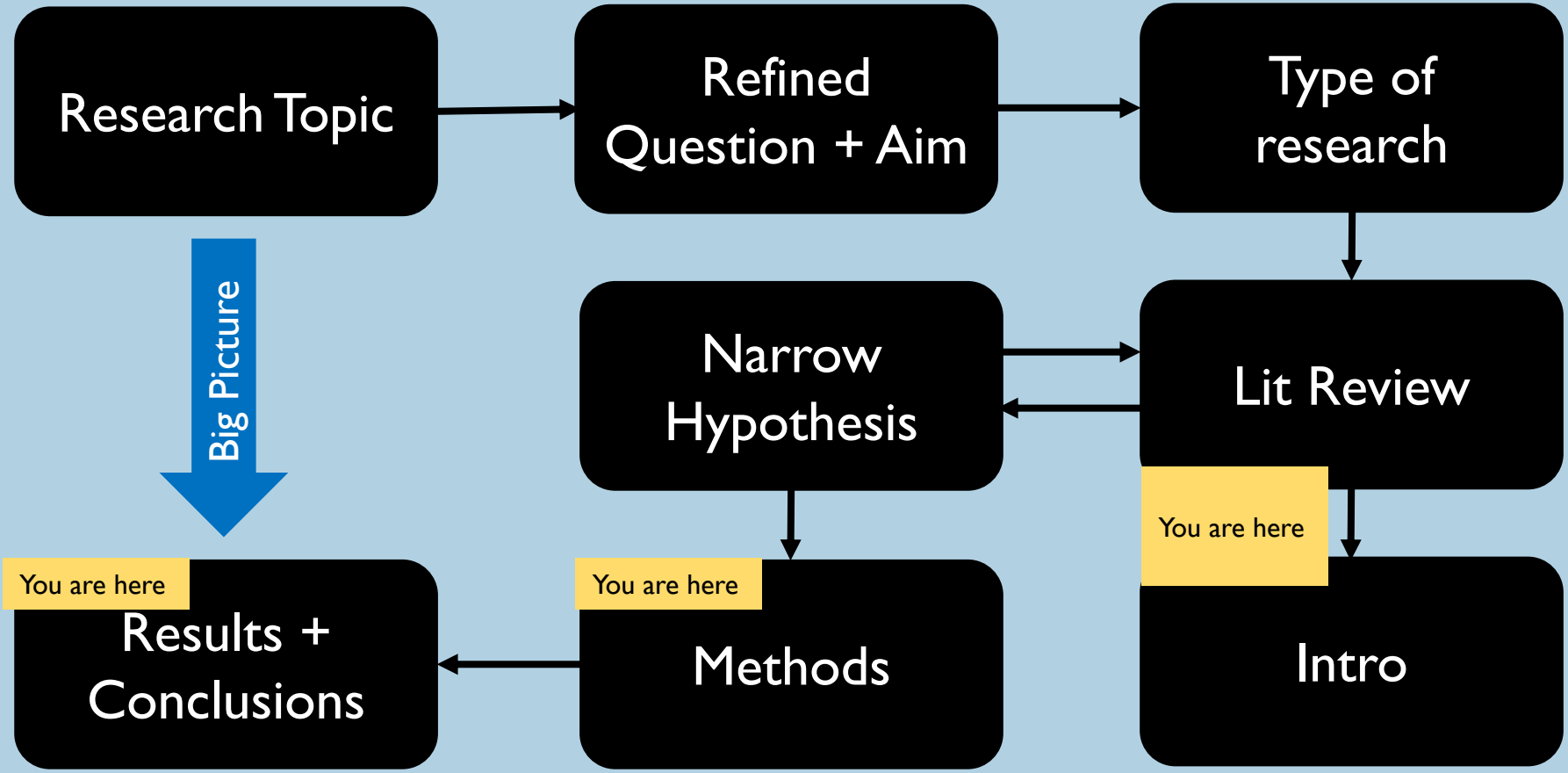
COURSE PROGRESS

- HW 3 should be graded, HW 5 asap
- If HW 1-3, quiz 1-2, or Datacamp are ungraded, let me know ASAP
- If you're behind, let me know what I can do to help!
  - Recommend prioritizing HW assignments relating to project, getting project done, filling in the rest at the end
  - Quizzes have a finite due date/window in which you can take them, so I will need to re-open if you missed one

COURSE PROGRESS CONT.

- Lots due this week/next don't forgor 💀 anything!
- All remaining HW (6,7,8) project-based
- Quiz 4 end of week
  - Open note/everything, no note submission
- Project check-ins starting tomorrow (look for announcement)
- Projects due next week, presentations!

RESEARCH PROJECT – BIG PICTURE



RECAP: WHAT IS THE FINAL PROJECT?

- Peer-review-publication-like paper
- Your own research (analysis, lit review, etc.)
- Clear sections
- Final presentation next week!
- Due *no later than* August 5 (unless extension previously granted)

MORE INFORMATION

- https://cepc0904-22.jackhester.com/project-details.html

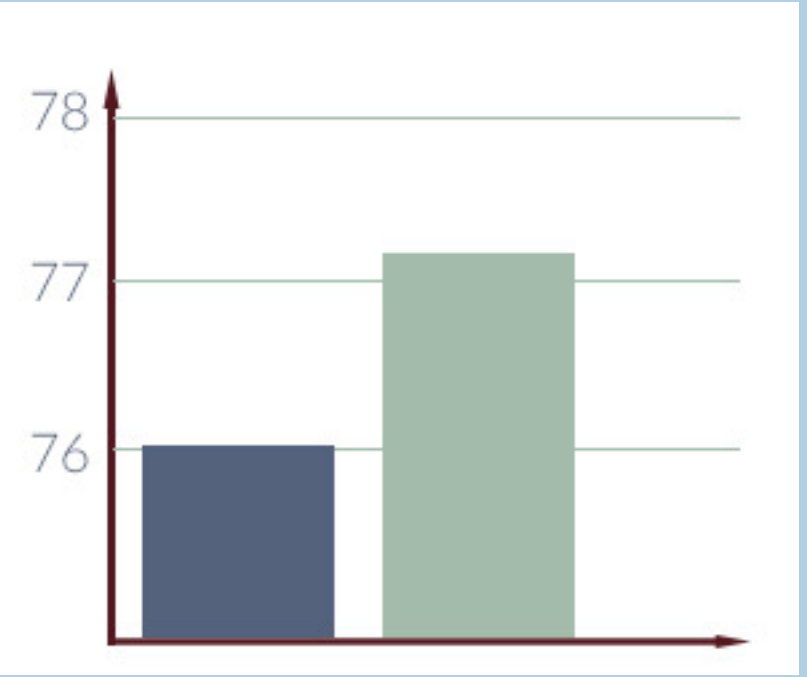
DATA VISUALIZATION

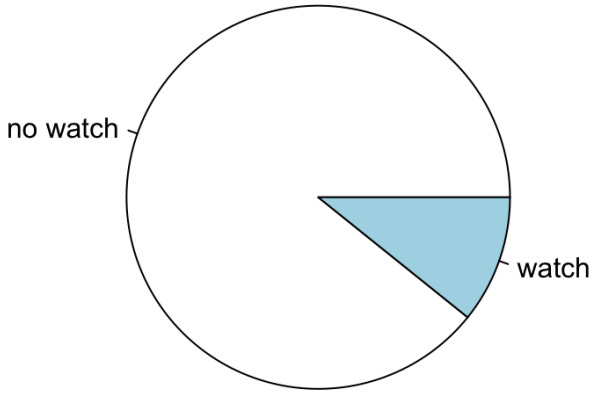
WHY DOES DATA VIZ MATTER?

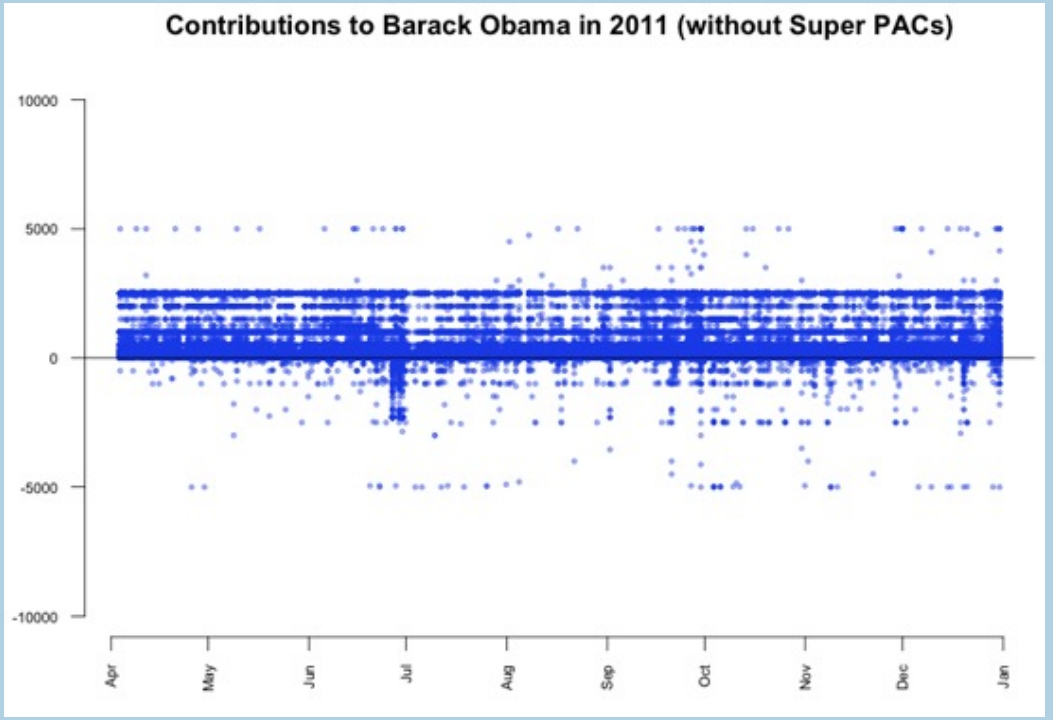
- Data understanding
- Data accessibility
- Data retention
- You're telling a story!

EXAMPLES

BAD 😳

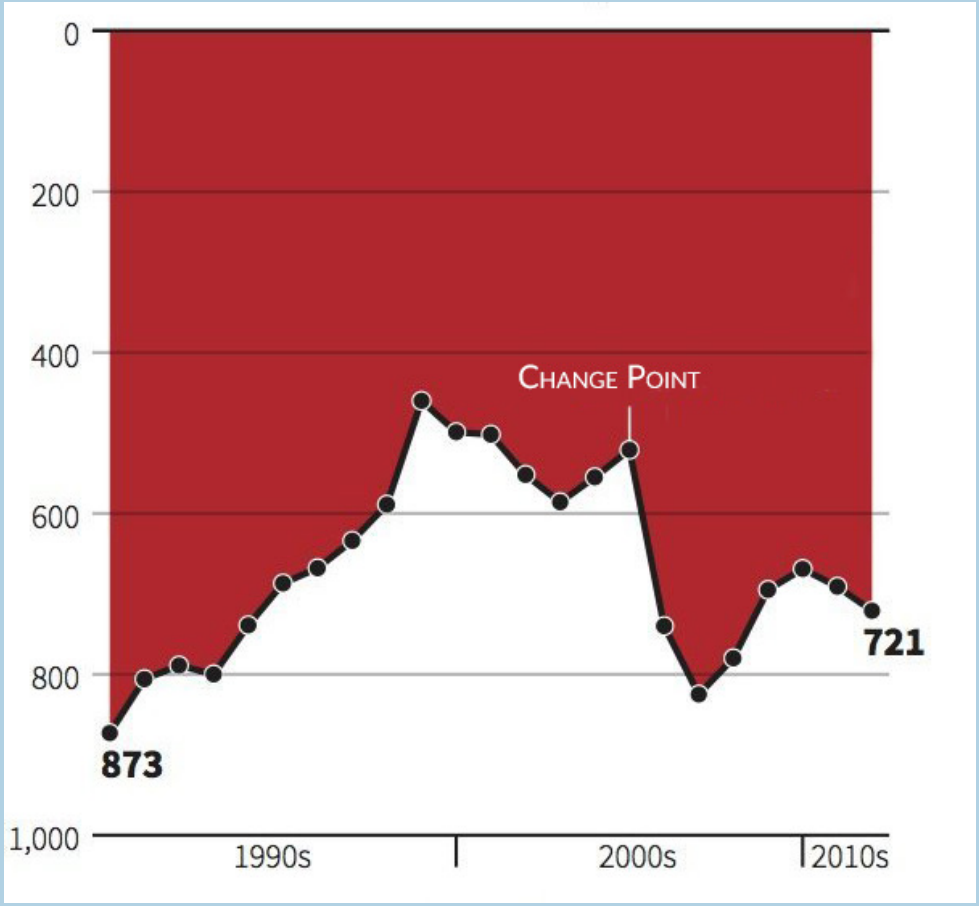






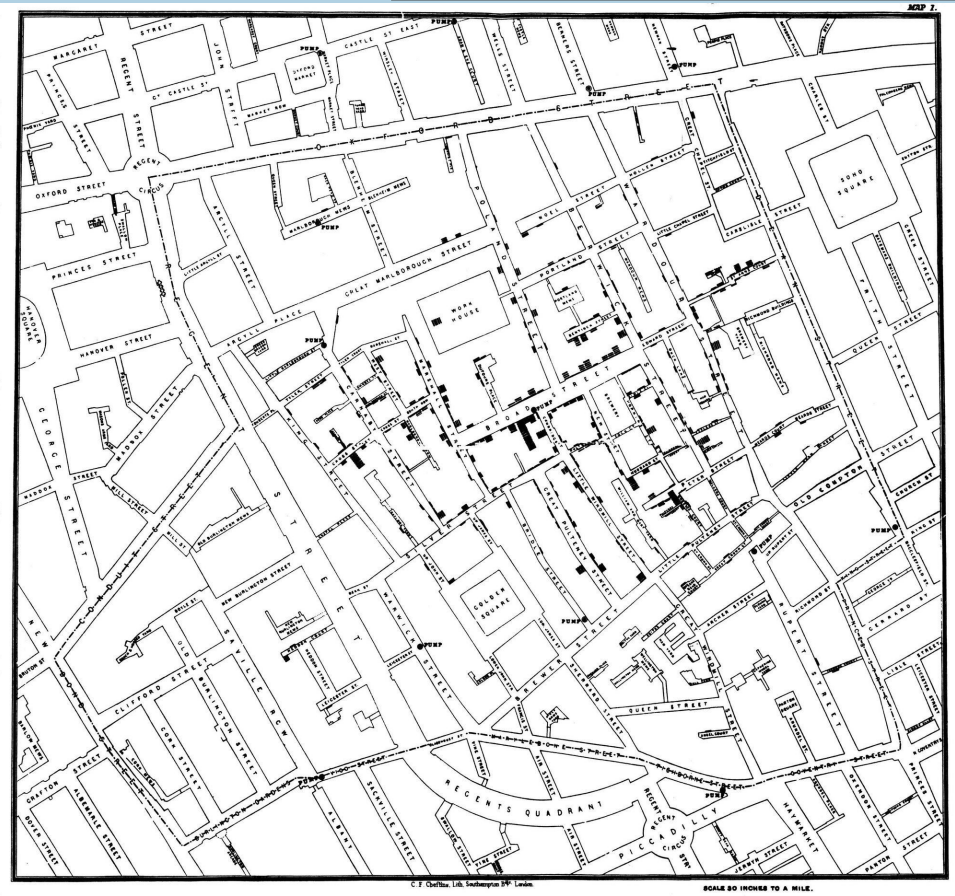
RIP those amputated axes 🙏🕊️

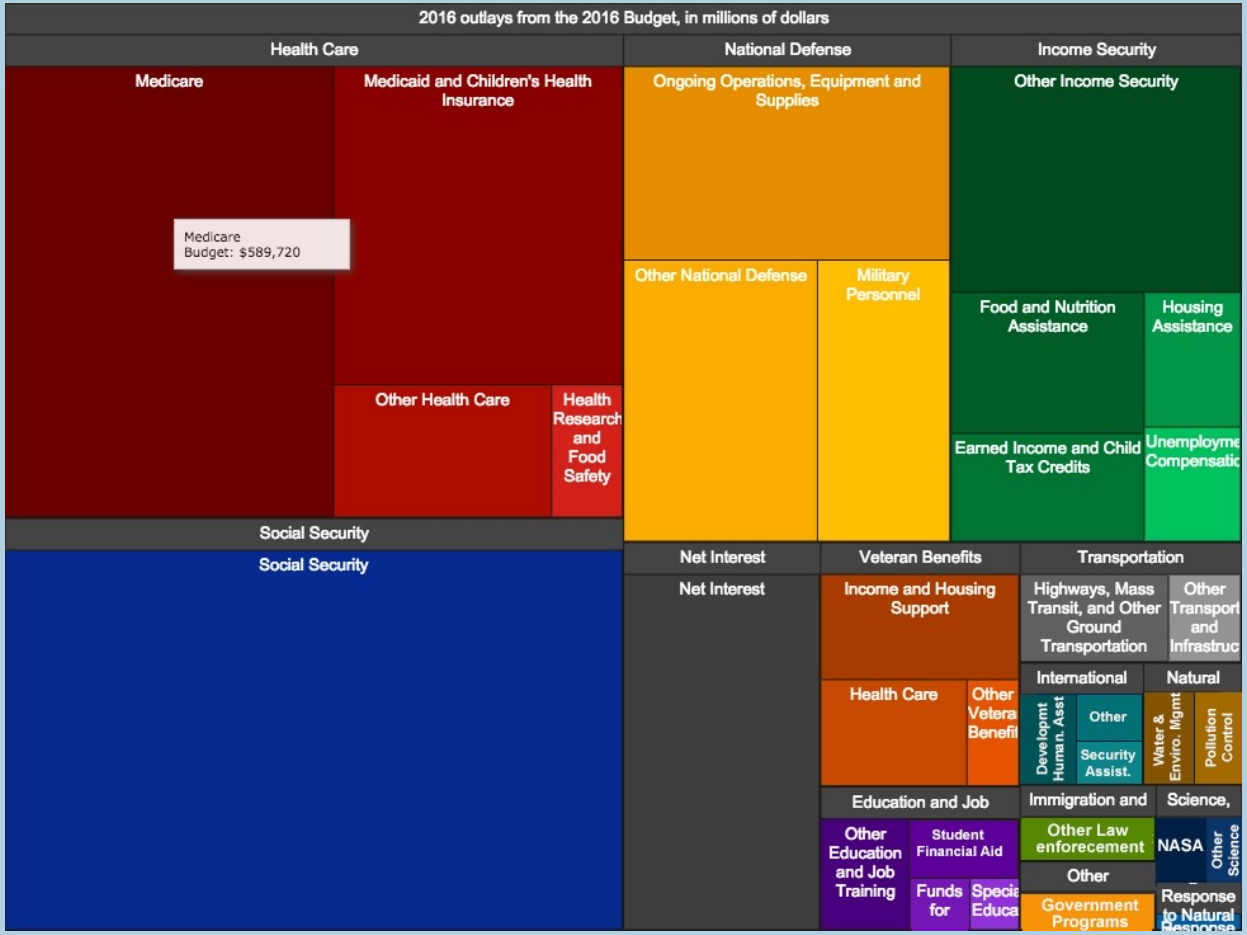
WORSE!


0
200
400
600
800
1,000
CHANGE POINT
873
721
1990s
2000s
2010s

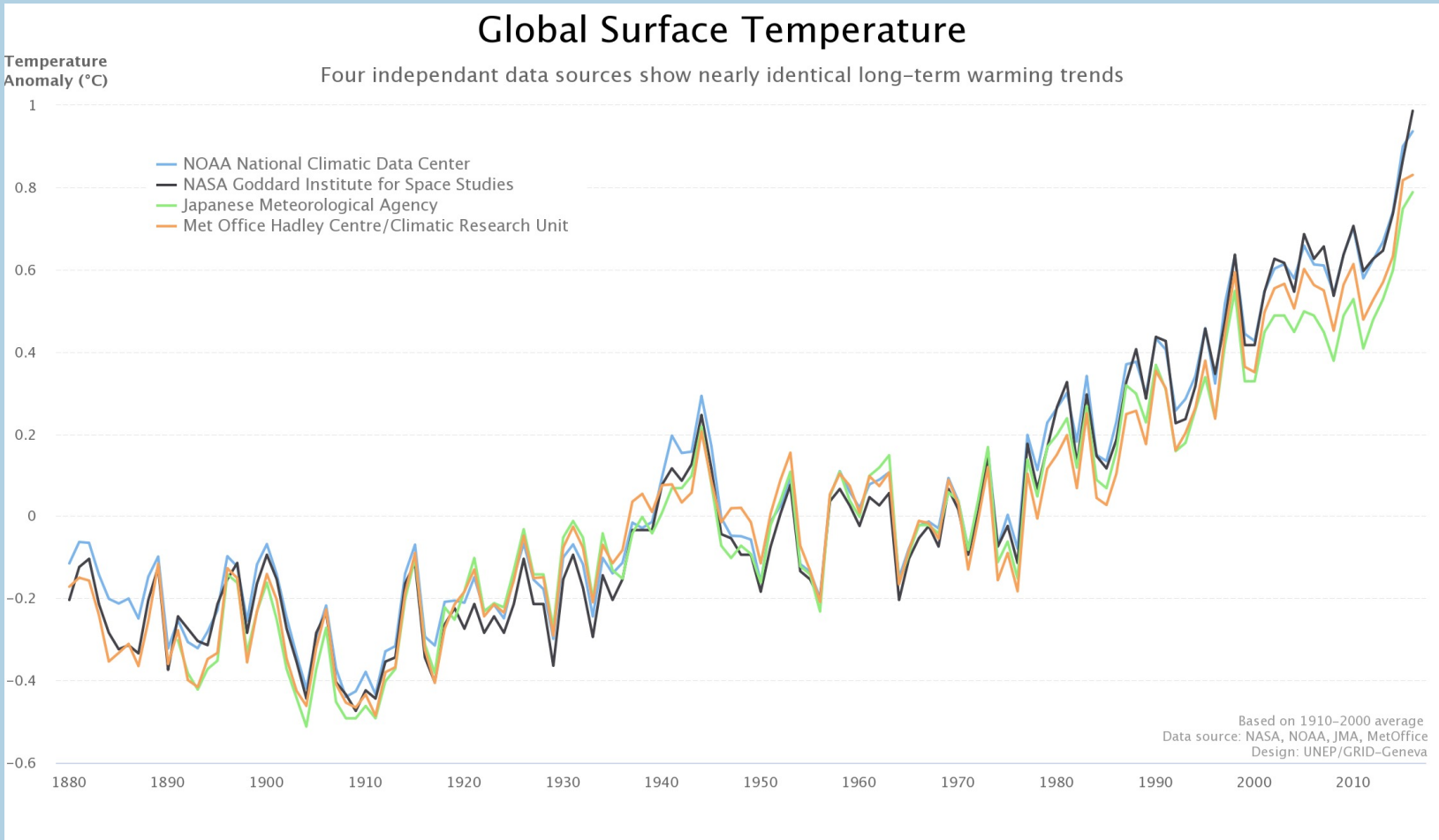
TECH STOCK THROWBACK?
NETFLIX (JUN 2011-NOV 2013)
TESLA (JAN 2019-TODAY)
50
30
10
0
360
260
160

GOOD ON PURRR





ALSO PRETTY GOOD



THIS ONE GOES INCREDIBLY HARD

https://www.youtube.com/watch?v=cDDWvj_q-o8

RELEVANT TO THIS CLASS

- Histograms
- Bar charts
- Scatterplots
- Tables
- Word clouds

ABSOLUTE ESSENTIALS

- Keep it simple
- Value added over text
- Label everything with titles, keys, legends, text, etc.
  - But keep text to a minimum
- Be honest
  - Tell the story, but don’t tell a misleading one
- Make it big enough to be readable

TABLE EXAMPLE

Table 1 Properties of each study including length of study, N, ecological momentary assessments (EMAs) collected and passive data streams collected

Authors	Year	New data set	Focus	N	Duration	EMA type	EMA	Accelerometer	App usage	Battery/ charge	Browser history	Bluetooth	Call logs	Camera events	Screen	Location	Light sensor	Microphone	SMS/Email	Lock/unlock
Wang *et al*[8]	2017	✓	Mental health	48	10 weeks	Survey	8/day (avg)	✓				✓				✓	✓	✓		
Chow *et al*[14]	2017	✓	Social anxiety	63	2 weeks	Survey	6/day (avg)									✓				
Wang *et al*[23]	2018	✓	Depression	83	9 weeks	PHQ-4	1/week	✓								✓	✓	✓		✓
Sano *et al*[15]	2015	✓	GPA, sleep, stress	66	1 month	Survey	2/day	✓					✓		✓	✓			✓	
Boukhechba *et al*[16]	2018	✓	Mental health	72	2 weeks	Survey	7/day	✓					✓			✓			✓	
Boukhechba *et al*[31]	2018	✓	Social anxiety	228	2 weeks	N/A	N/A									✓				
DaSilva *et al*[30]	2019	✓	Stress	72	43 days	MPSM	1/day	✓							✓	✓	✓	✓		✓
Xu *et al*[24]	2019	✓	Depression	350	106 days	N/A	N/A	✓		✓		✓	✓		✓	✓				✓
Abdullah *et al*[17]	2014	✓	Sleep	9	97 days	Sleep journal	1/day		✓		✓		✓		✓	✓			✓	
Doryab *et al*[25]	2019	✓	Loneliness	160	16 weeks	N/A	N/A	✓				✓	✓		✓	✓		✓	✓	✓
Jacques *et al*[18]	2015		Happiness	68	1 month	survey	2/day	✓					✓		✓	✓			✓	
Vhaduri *et al*[12]	2016		Activity	195	Three semesters	N/A	N/A									✓				
Lane *et al*[13]	2014	✓	Sleep, physical activity, social activity	27	19 days	N/A	N/A	✓										✓		
Gong *et al*[32]	2019	✓	Social anxiety	52	2 weeks	N/A	N/A	✓					✓			✓			✓	
Huang *et al*[28]	2016	✓	Social anxiety	18	10 days	N/A	N/A									✓				
Boukhechba *et al*[33]	2017	✓	Social anxiety	54	2 weeks	N/A	N/A						✓			✓			✓	
Asselbergs *et al*[34]	2016	✓	Mood	27	35.5 days (avg)	eMate	5/day	✓	✓				✓	✓	✓				✓	
Wang *et al*[27]	2015		GPA	48	10 weeks	survey	8/day (avg)	✓	✓			✓				✓	✓	✓		
Jacobson *et al*[19]	2020	✓	Social anxiety	59	2 weeks	N/A	N/A	✓					✓						✓	
Marquet *et al*[26]	2018	✓	Physical activity	74	1 week	survey	3/day	✓												
Ben-Zeev *et al*[29]	2015	✓	Mental health	47	10 weeks	Survey	5/week	✓							✓	✓	✓	✓		✓
Singh *et al*[20]	2016	✓	Cooperation	54	10 weeks	N/A	N/A						✓			✓			✓	
Saeb *et al*[10]	2016		Depression	48	10 weeks	N/A	N/A									✓				
Lee *et al*[21]	2016	✓	Problematic phone usage	95	27 days	N/A	N/A		✓		✓		✓		✓				✓	✓
Stütz *et al*[22]	2015	✓	Stress	15	2 weeks	Survey	7/day	✓	✓				✓		✓		✓	✓	✓	

GPA, Grade point average; MPSM, Mobile photographic stress meter; N, Number of participants in the study; PHQ, Patient health questionnaire; SMS, Short message service.

TABLE PAPER INFO

- Other good examples in there too!
- https://canvas.brown.edu/courses/1088637/files?preview=68382733

POTENTIALLY USEFUL SOFTWARE

- R has ggplot and other data viz packages!
- https://jackhester.com/documents/Data-Vis-Software-List-F20.pdf
- Basic plots can be done easily in google sheets/excel

DONE WITH CLASS!

- Questions on
  - Content
  - R HW
  - Project
  - Quiz
  - Other assignments
- Concerns?
- Other Comments?

WHAT I'D LIKE

- Meetings!
- Feedback on the course/concerns as you finish!!
- Final paper drafts over the weekend!!!!!!!!!