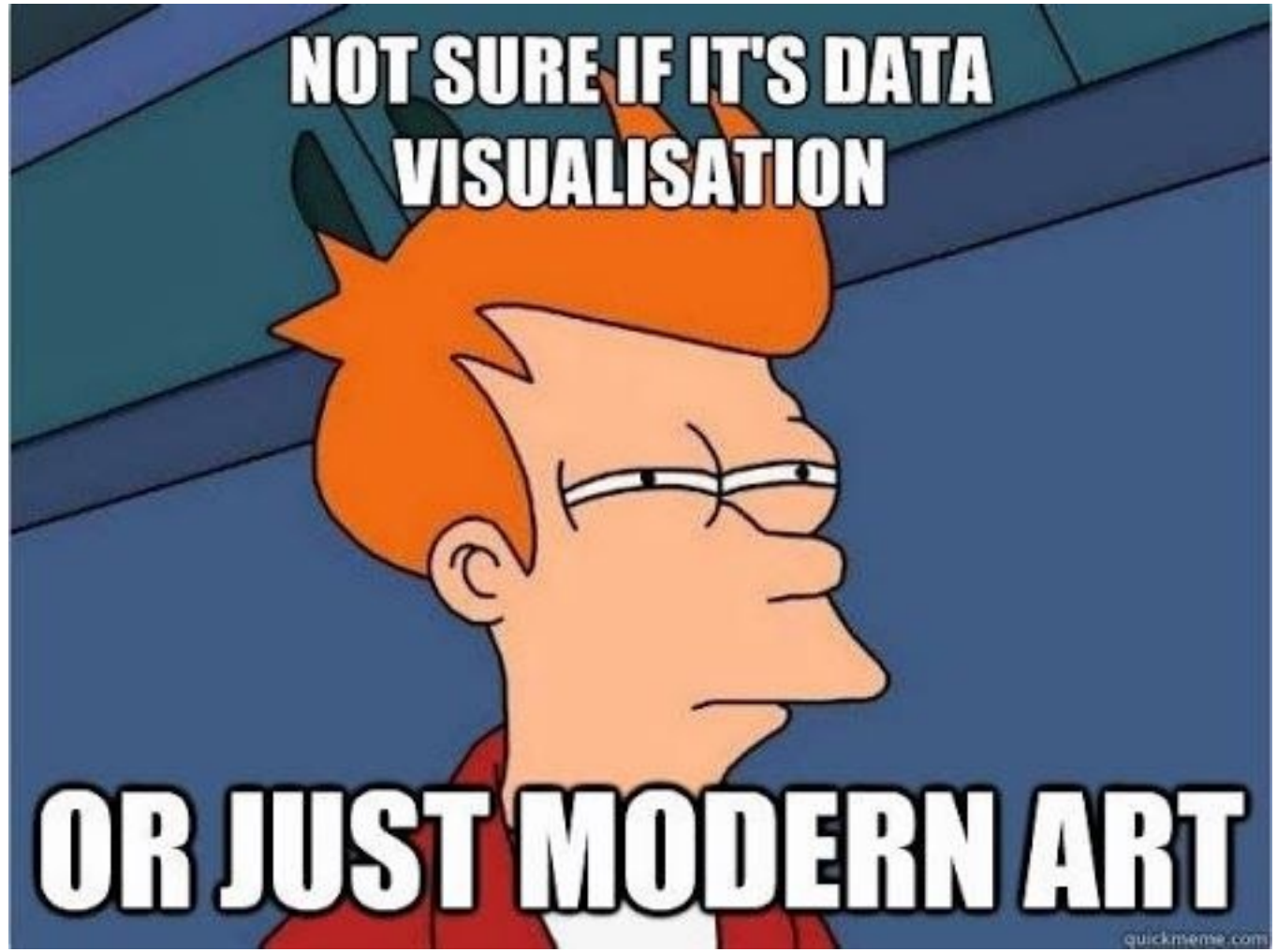

NOT SURE IF IT'S DATA VISUALISATION
OR JUST MODERN ART
quickmeme.com

Introduction to Data Visualization

329E Data Visualization

What *is* Information Visualization?

- The use of computer-supported, interactive, visual representations of data to amplify cognition

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- Visualization is the process that transforms (abstract) data into interactive graphical representations for the purpose of exploration, confirmation, or presentation.

Why Create Visualizations?

Exploration – Take new perspectives on data, analyze complex patterns and reveal unseen relationships

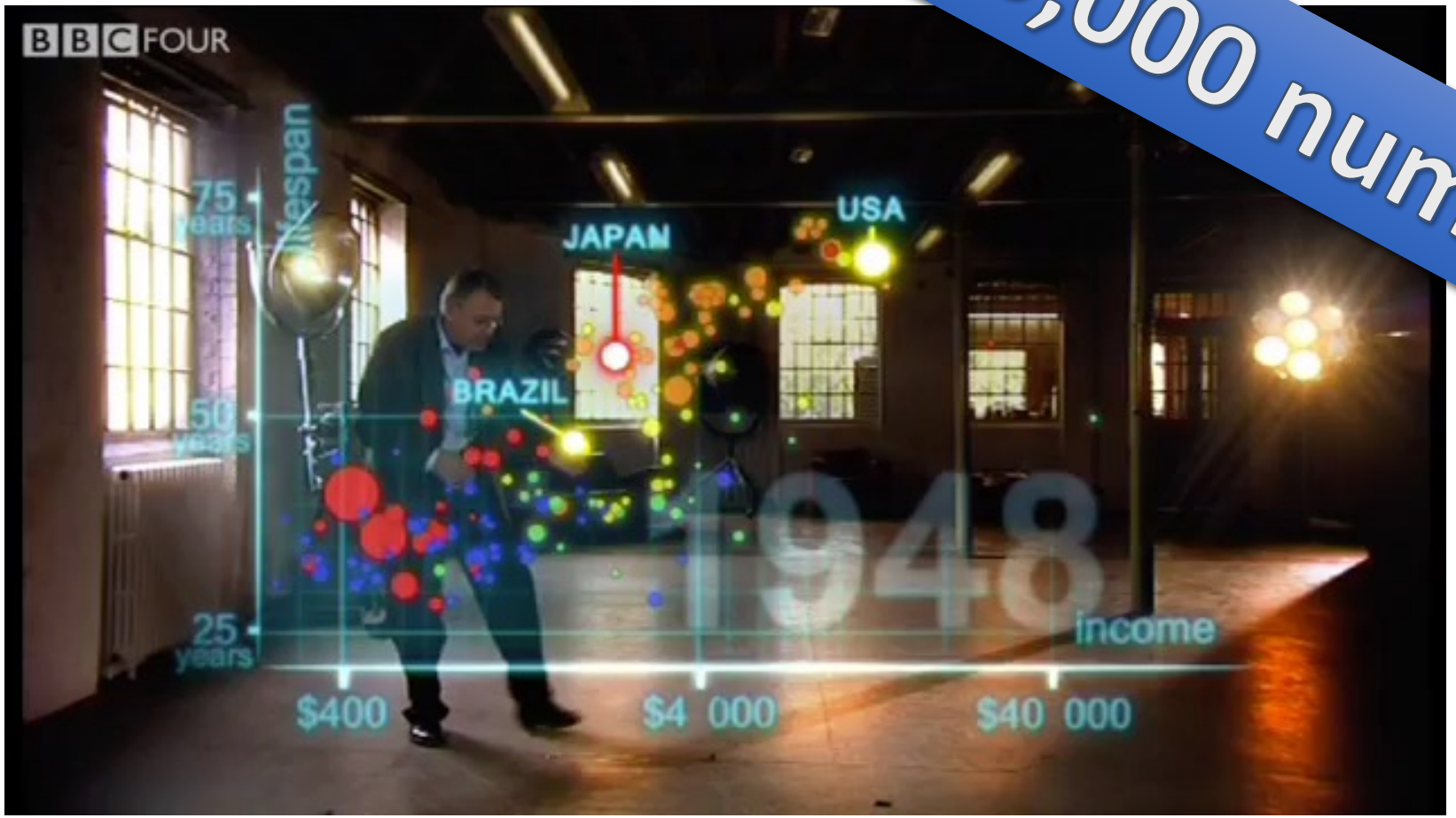
- Expand memory
- Find/reveal patterns
- Generate and confirm hypotheses
- Inspire
- See data in context

Explanation –Communicate insights, inspire new thinking about a problem, and inform decisions

- Analyze data to support reasoning
- Answer questions
- Communicate ideas to others
- Make decisions
- Record information
- Support computational analysis
- Tell a story

The Joy of Stats

120,000 numbers!



Hans Rosling's 200 Countries, 200 Years, 4 Minutes – BBC, The Joy of Stats (2010)
https://youtu.be/jbkSRLYSojo

Visualization is the Confluence of

- Cartography
- Statistics
- Perceptual psychology
- Computer graphics
- Human-computer interaction
- Graphic design

In industry this spans roles across the organization

Visualization in the Data Science Process

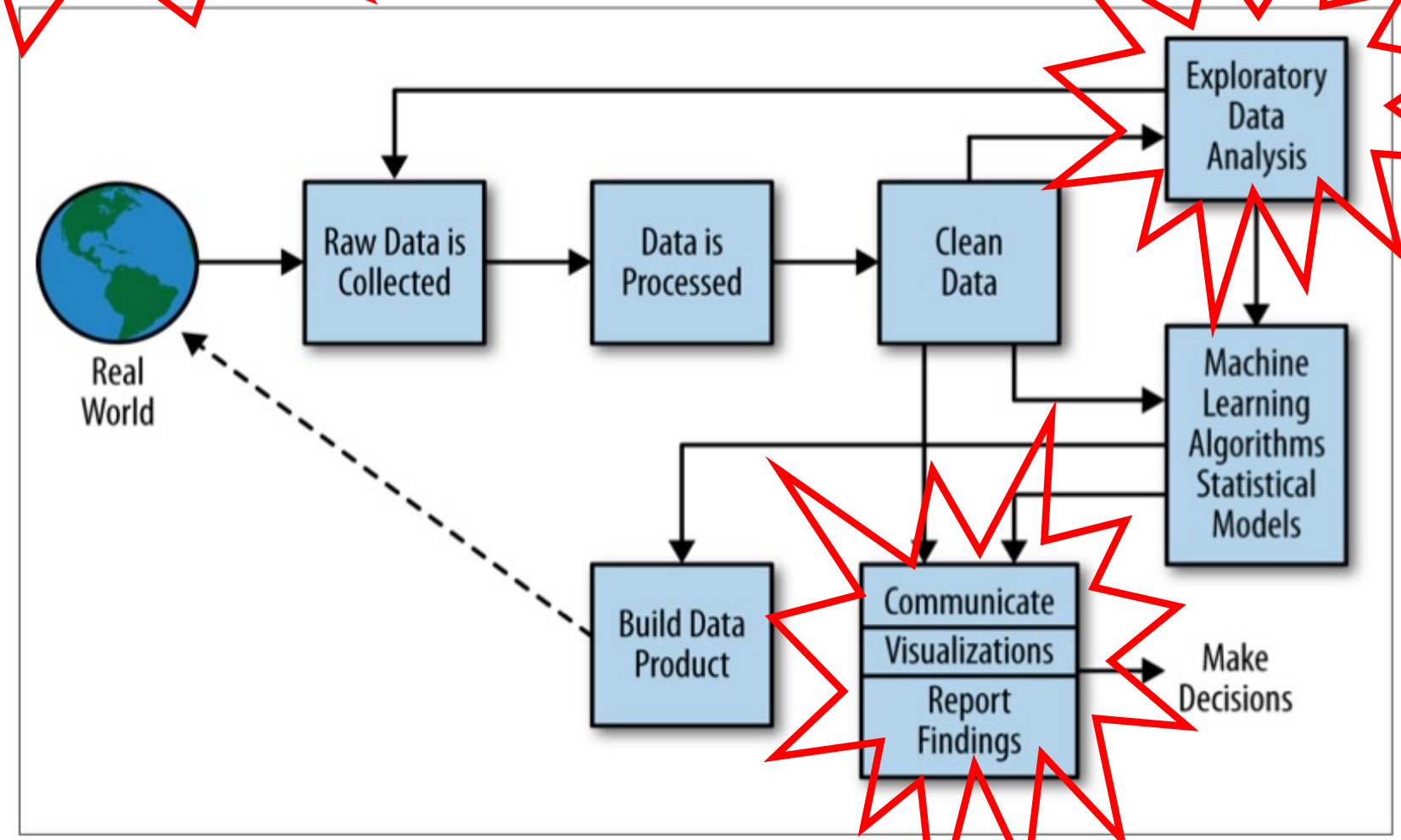


Figure 2-2. The data science process

- Rachel Schutt and Cathy O'Neil, *Doing Data Science*, O'Reilly Media, Inc., 2014.

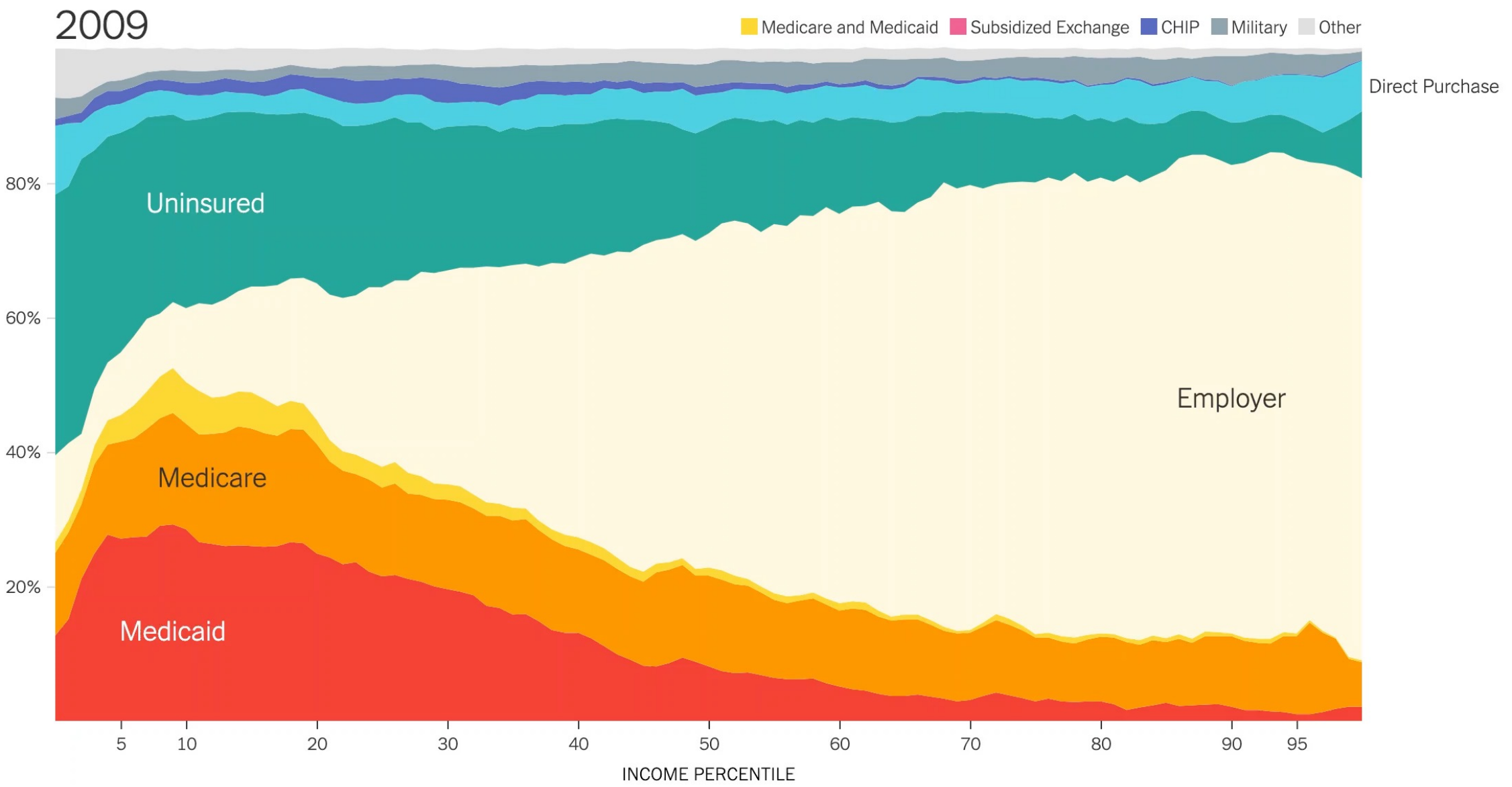
Visualization Motivation

- **Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.**

- Human in the loop needs the details & no trusted automatic solution exists
  - Doesn't know exactly what questions to ask in advance
  - Exploratory data analysis
    - ***Speed up*** through human-in-the-loop visual data analysis
  - Present known results to others
  - Steppingstone towards automation
    - Before model creation to provide understanding
    - During algorithm creation to refine, debug, set parameters
    - Before or during deployment to build trust and monitor

Why Use an External Representation?

- External representations → replace cognition with perception



Note: The income measure used here is **disposable income**, which is defined as market income plus cash transfers from the government and noncash benefits like food stamps and public housing. It excludes taxes, out-of-pocket medical spending, insurance premiums and other expenses. There have been small changes in the census questionnaire between the two survey dates. Source: Current Population Survey, analysis by Matt Bruenig

https://www.nytimes.com/2020/12/29/upshot/obamacare-recession-big-test.html

Why Depend on Vision?

- Human visual system is high-bandwidth channel to brain
  - Overview possible due to background processing
    - Subjective experience of seeing everything simultaneously
    - Significant processing occurs in parallel and pre-attentively
- Sound: lower bandwidth and different semantics
  - Overview not supported
    - Subjective experience of sequential stream
- Touch/haptics: impoverished record/replay capacity
  - Only very low-bandwidth communication thus far
- Taste, smell: no viable record/replay devices

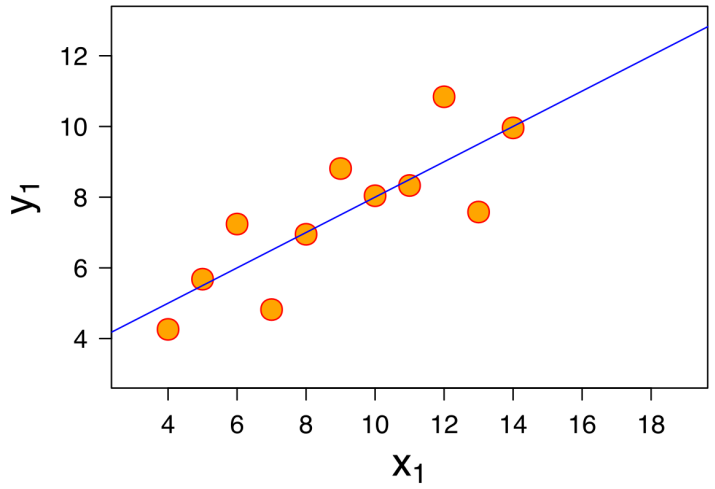
Why represent all the data?

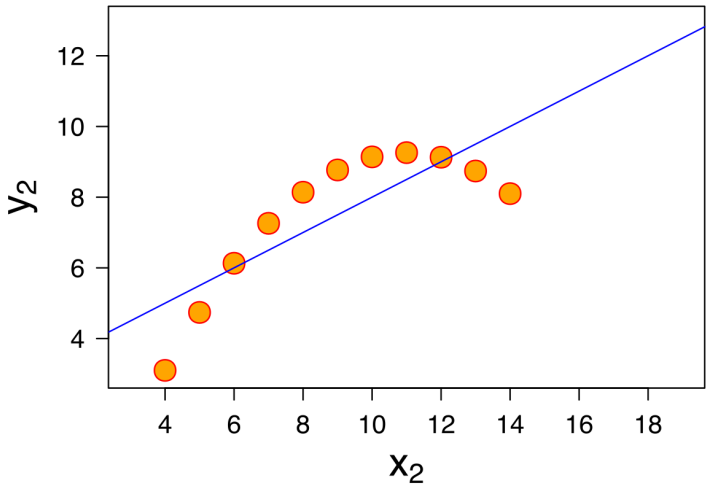
- Summaries lose information, details matter
  - Confirm expected and find unexpected patterns
- Assess validity of statistical model

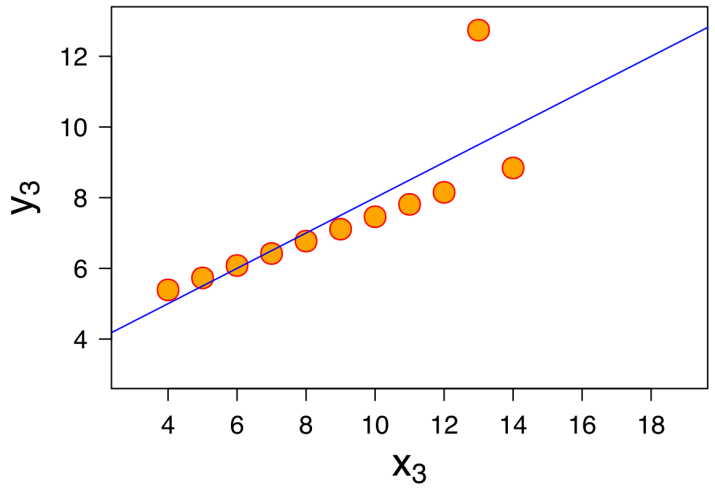
Anscombe's Quartet

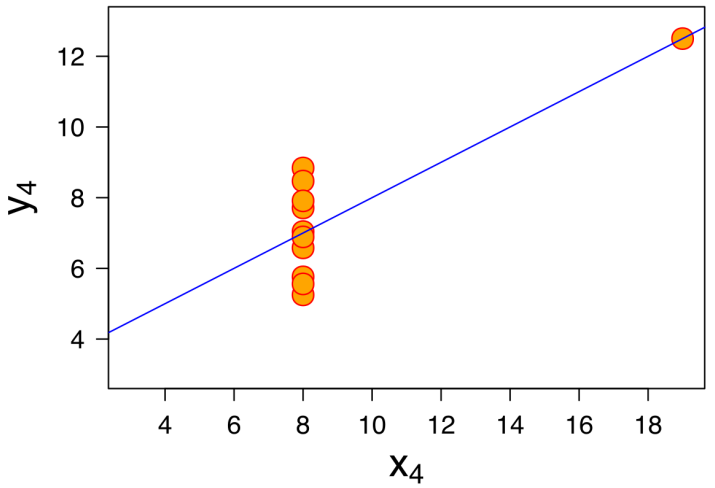
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

All four series have identical:
Mean x: 9 y: 7.50
Variance x: 11 y: 4.122
Correlation x – y: 0.816
Linear regression: y = 3.00 + 0.500x









Why focus on tasks and effectiveness?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- Effectiveness requires match between data/task and representation
  - set of representations is huge
  - many are ineffective mismatch for specific data/task combo
  - increases chance of finding good solutions if you understand full space of possibilities
- What counts as effective?
  - novel: enable entirely new kinds of analysis
  - faster: speed up existing workflows
- How to validate effectiveness
  - many methods, must pick appropriate one for your context

What resource limitations are we faced with?

Vis designers must consider three very different kinds of resource limitations: those of computers, of humans, and of displays.

- Computational limits
  - processing time
  - system memory
- Human limits
  - human attention, cognition, and memory
- Display limits
  - pixels are precious resource, the most constrained resource
  - **information density**: ratio of space used to encode info vs unused whitespace
    - tradeoff between clutter and wasting space, find sweet spot between dense and sparse

Why Does Visualization Work?

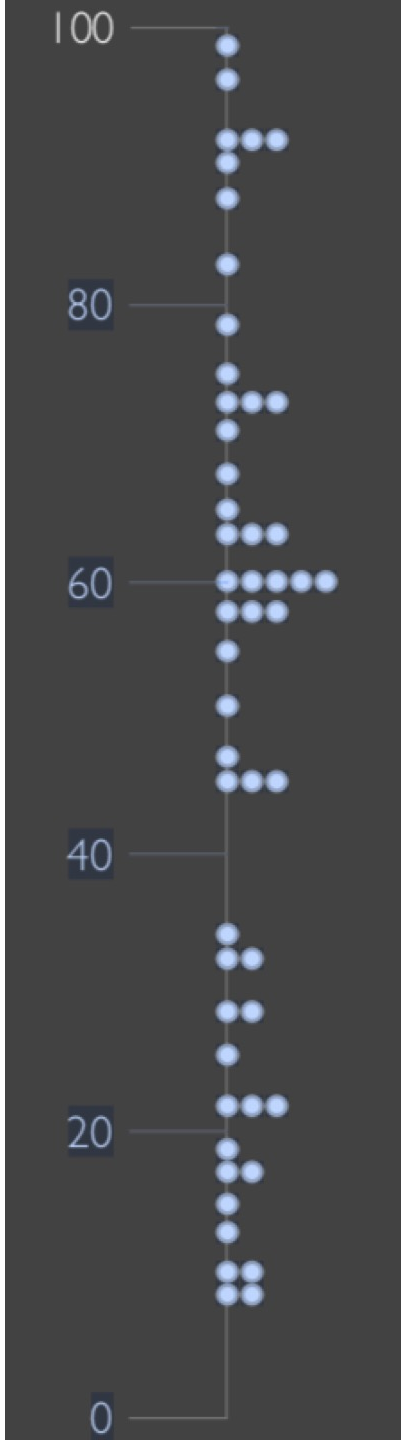
- Power of Perception … what number appears most often?

15 19 60 33 11 75 57 34 79 18 51 92 73 22 13 71 60
22 17 10 68 73 18 55 65 46 29 60 73 22 46 92 97 10
58 46 57 17 83 26 99 33 88 92 60 91 29 57 96 12
47

Why Does Visualization Work?

- Power of Perception … what number appears most often?

15 19 60 33 11 75 57 34 79 18 51 92 73 22 13 71 60
22 17 10 68 73 18 55 65 46 29 60 73 22 46 92 97 10
58 46 57 17 83 26 99 33 88 92 60 91 29 57 96 12
47



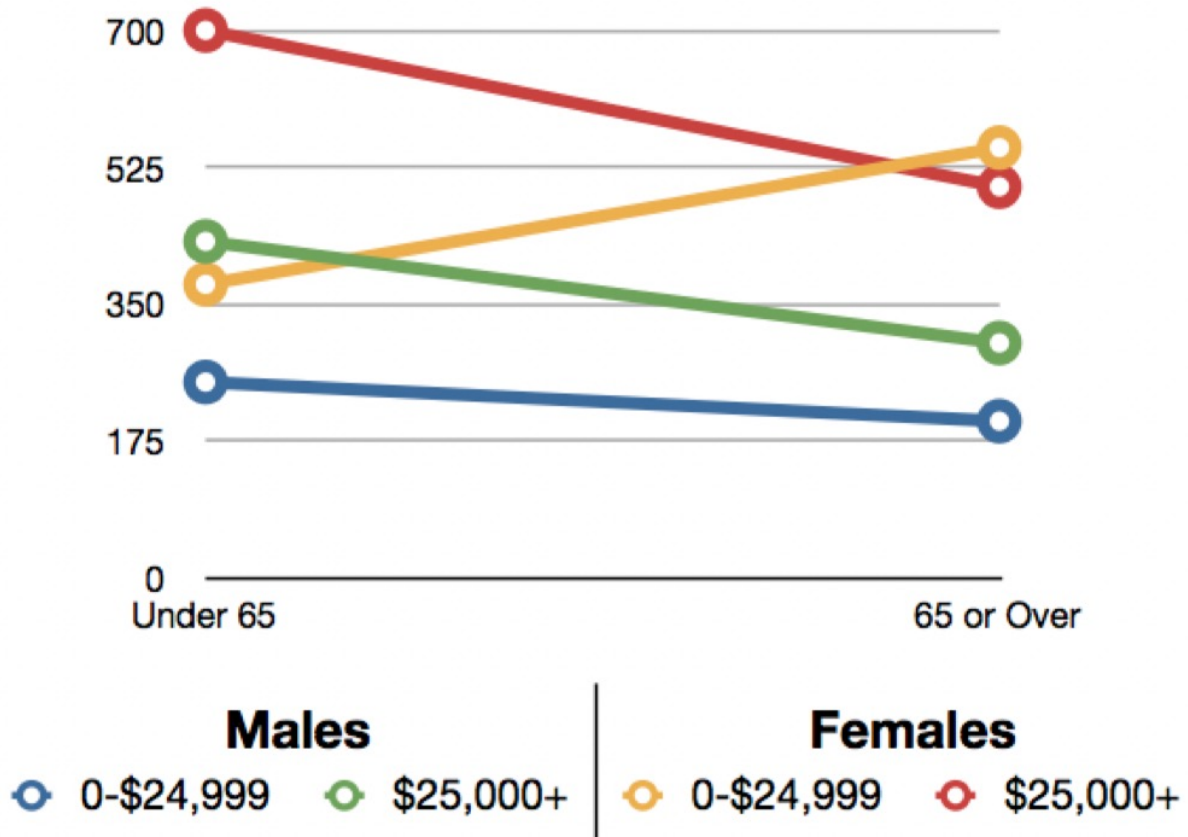
Example

- Which gender and income level shows a different effect of age on triglyceride levels?

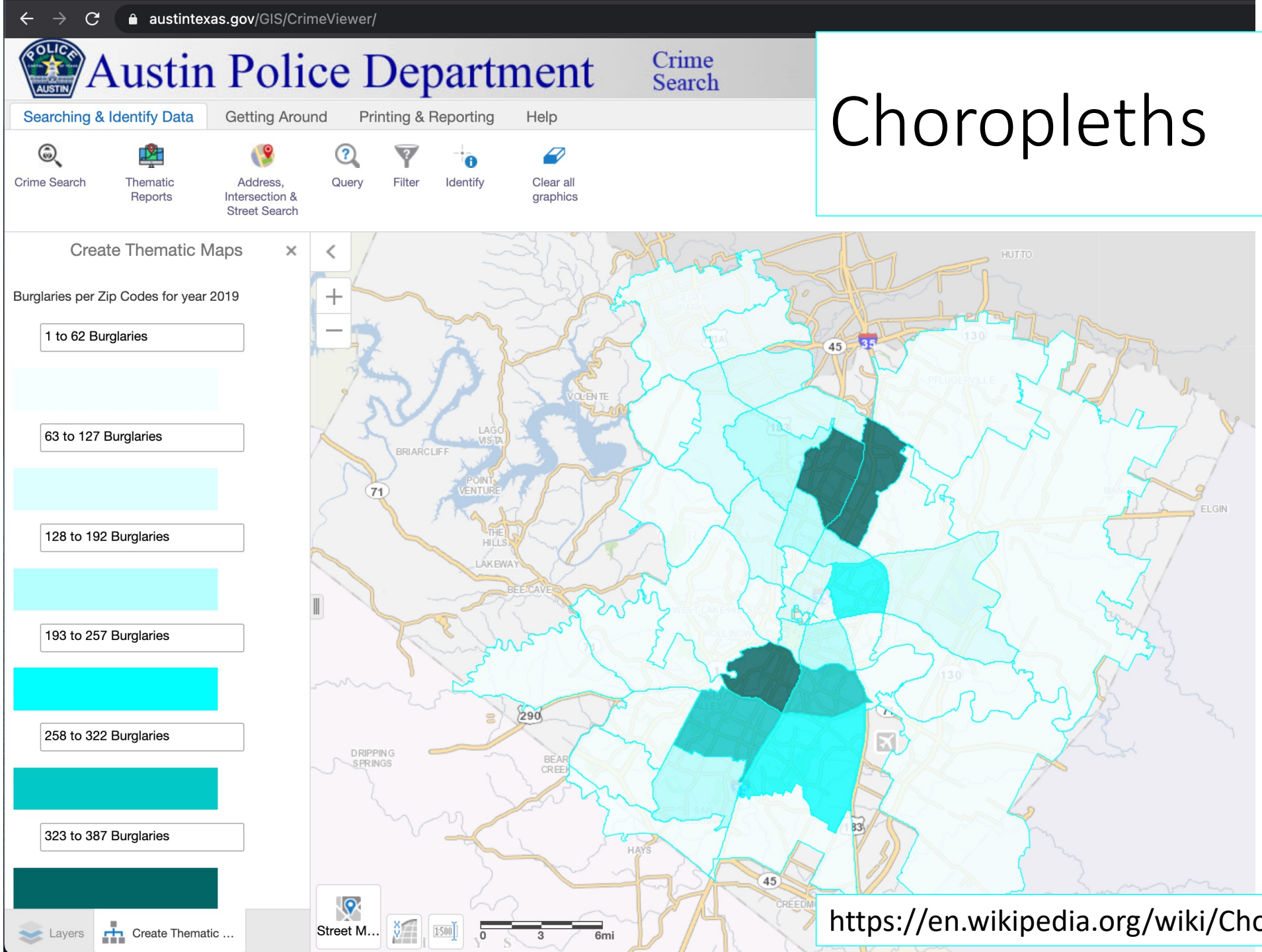
	Males		Females	
Income Group	Under 65	65 or Over	Under 65	65 or Over
0-$24,999	250	200	375	550
$25,000+	430	300	700	500

Example

- Which gender and income level shows a different effect of age on triglyceride levels?



Choropleths



https://en.wikipedia.org/wiki/Choropleth_map

Course Mechanics

Learning Goals:

- *Evaluate* and *critique* visualization designs
- *Design* and *implement* interactive data visualizations
- *Gain* an overview of techniques and grammar
- *Develop* a substantial visualization project
- *Collaborate* with your colleagues
- *Navigate* and *understand* technical documentation

Course Components

- Lectures in class twice a week (MW 1:00pm – 1:50pm or 2:00pm – 2:50pm) – Only Zoom Lectures will be recorded
- Quizzes will be based on the reading and lectures.
- Lab & Discussion Group Activities
  - Synchronous (in person) once a week (F 1:00pm – 1:50pm or 2:00pm – 2:50pm)
  - Attendance is mandatory, if you can't make it, or had an unexpected conflict arise you must contact me
  - There will be a group activity
- Homework help practice specific skills (individual)
- Final Project gives you a chance to go through a complete visualization project

Grade Breakdown

- Quizzes – 20%
- Group Lab Activities - 25%
- Homework Assignments – 20%
- Final Project – 35%

Final Project

- At the core of the course is your project. The goal of the project is to design an interactive visualization that answer questions you have about some topic of your own choosing. You will acquire the data, design your visualizations, implement them, and critically evaluate the results. The path to a good visualization is going to involve mistakes and wrong turns. It is therefore important to recognize that mistakes are valuable in finding the path to a solution, to broadly explore the design space, and to iterate designs to improve possible solutions. The project has an intermediate milestone that will allow you to get feedback and to iterate.

Final Project

- Visualization research project on a topic of your choice (from your major, or from any number of open data sources)
- Work in teams of 2-3 (must be in your section)
- Deliverables:
  - Design and Pitch Idea
  - Design Feedback to other teams
  - Weekly Journal Entries
  - Project write-up in form of an 8 page research paper
  - Oral presentation (10 mins) with slides to class (last week)
  - Oral presentation feedback to other teams
  - Intra-team feedback

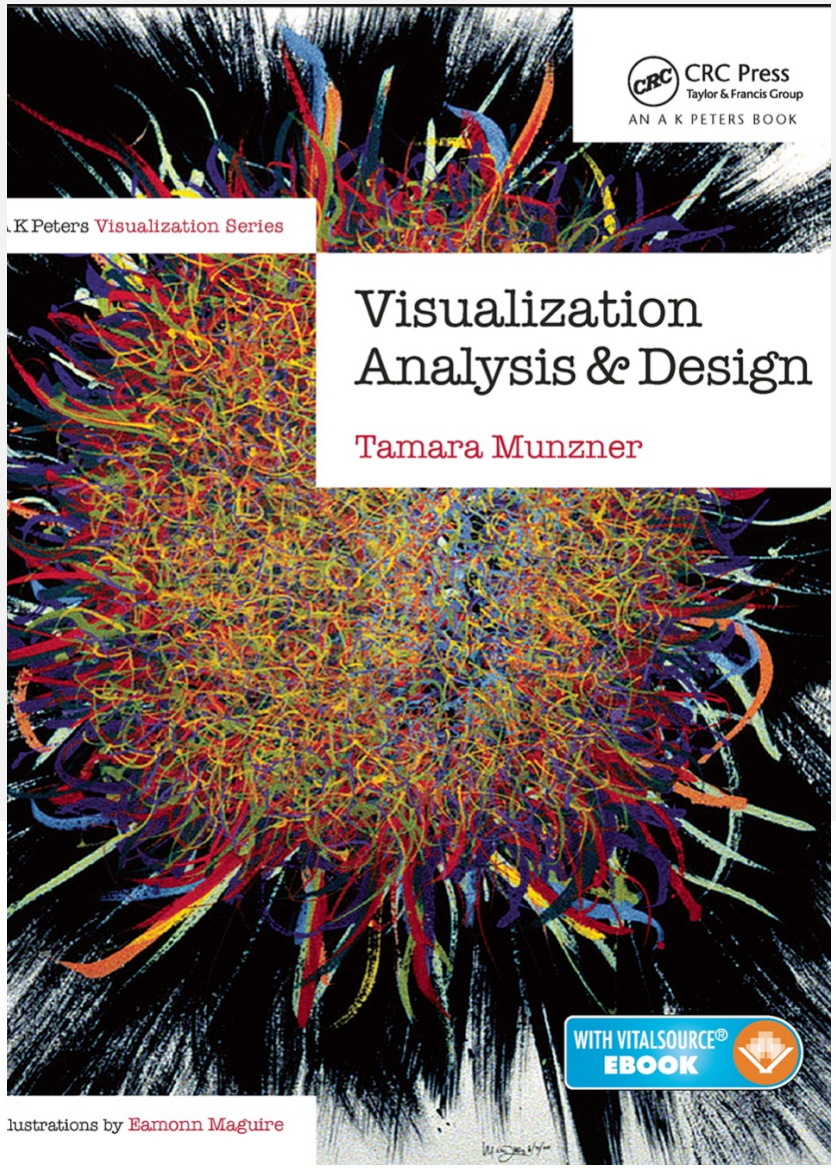
Regrading Policy

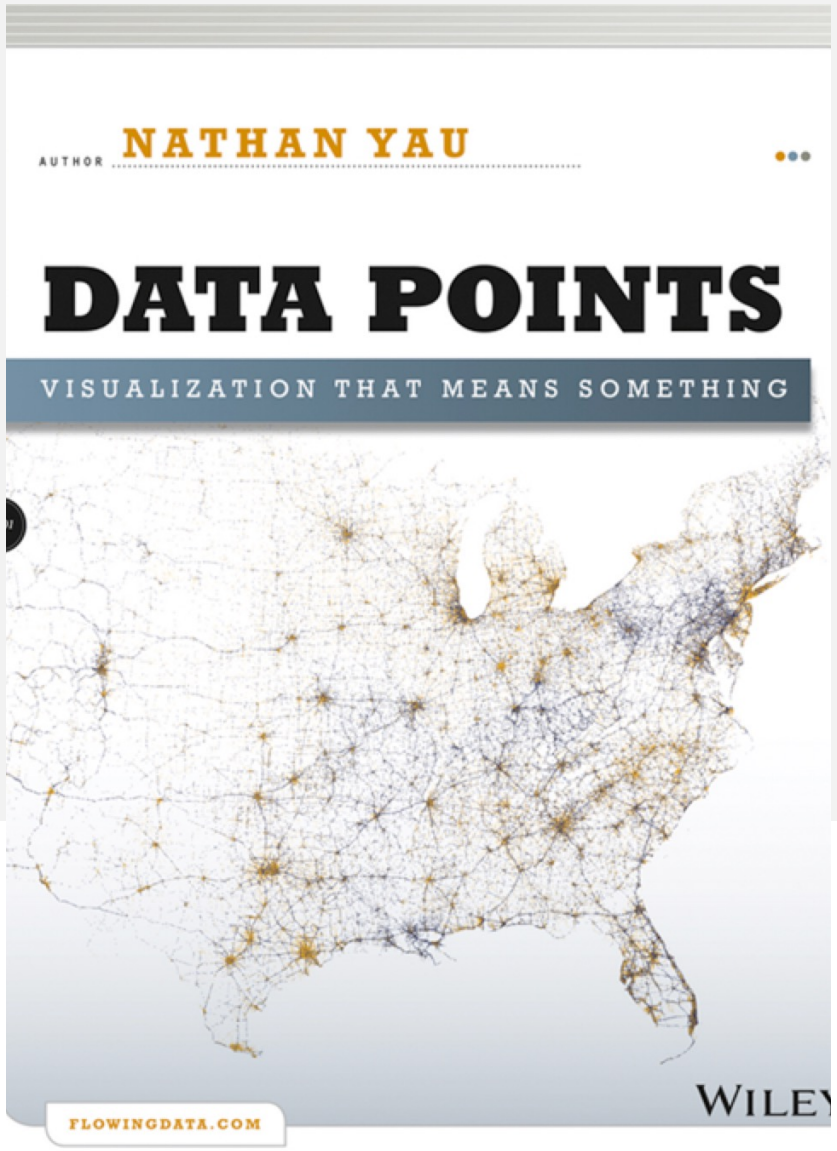
- It is very important to us that all assignments are properly graded. If you believe there is an error in your assignment grading, please submit an explanation via email to the TA **within 7 days** of receiving the grade and copy me. No regrade requests will be accepted orally, and no regrade requests will be accepted more than 7 days after you receive the grade for the assignment.

GRACE POLICY: Homework Time-bank options

- You may have a two-day grace period for one homework assignment, OR
- You may have 2 one-day extensions for two different homework assignments

- Homework will not be accepted more than 2 days late without speaking to me regarding your extenuating circumstances
- If you have exhausted your grace days, you may still turn in the homework up to two days late with a 10% deduction per late day.





Books – Both available at UT Library as Digital Loan

Other TODOs

- Reading Munzner Chapter 1
- Reading Yau Chapter 2
- Update your Zoom (if you haven’t already) so you can join your group
- Familiarize yourself with Canvas