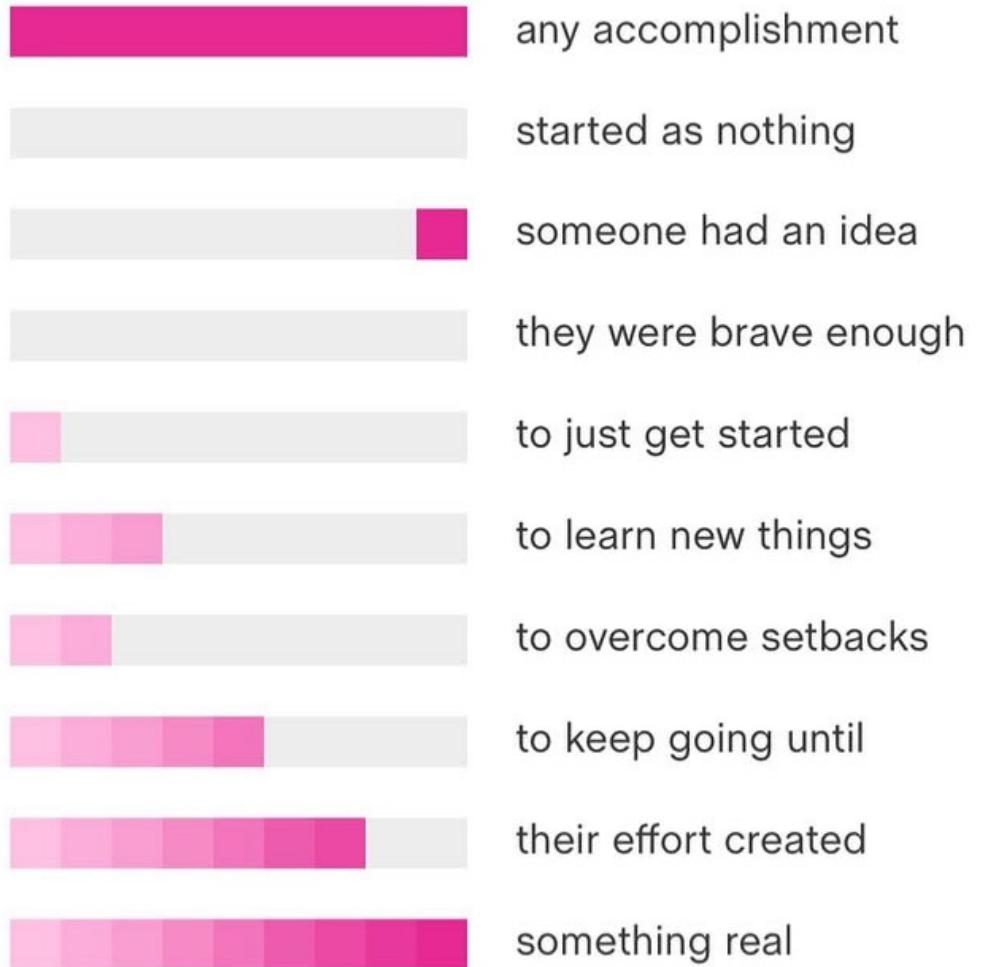


Visualization for Data Science

The Case for Visualizations

While you wait go to Canvas and log in/sign up for

1. iClicker Cloud
2. Ed Discussion
3. PraireLearn
4. UBC Github <https://activate1.github.ubc.ca/>
5. Then access the course website <https://tinyurl.com/dsci320>



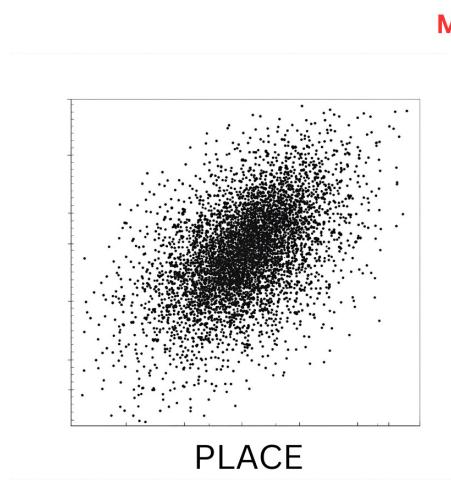
Introductions

Getting to know your table mates

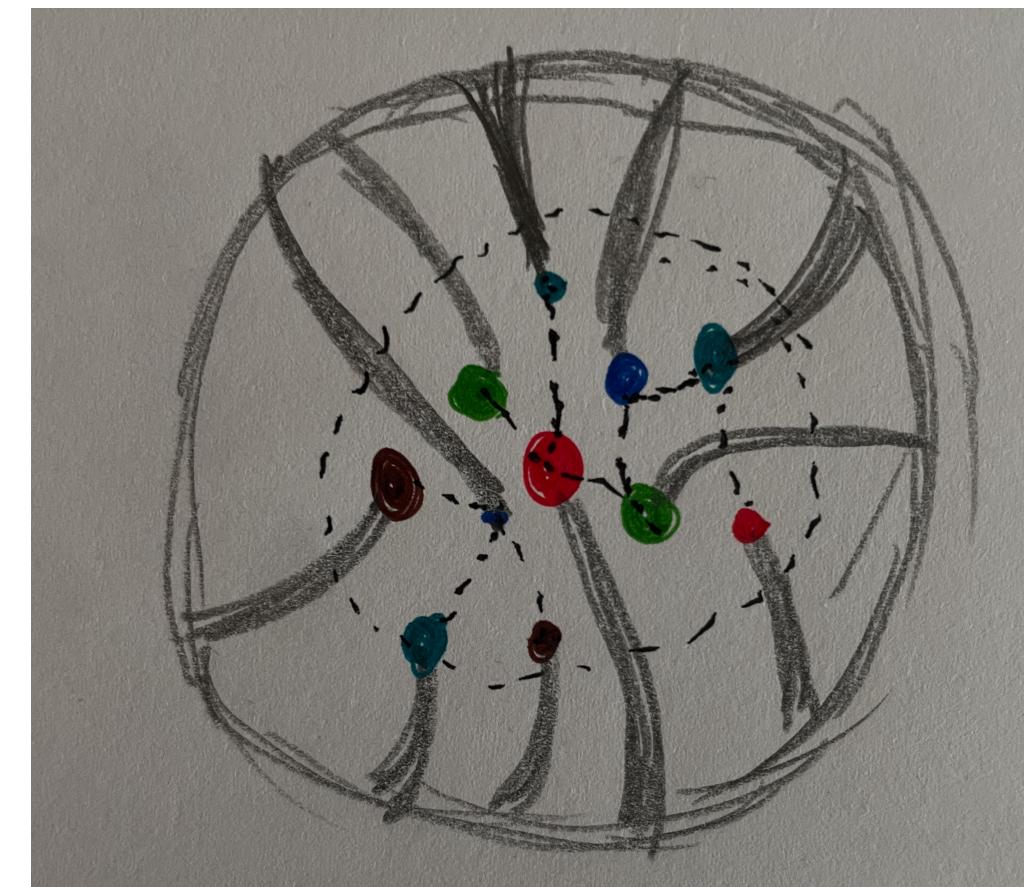
Draw a chart type that represents you and explain it in one sentence

3 minutes to draw

2 minutes to share with at least one person on your table



I'm a scatter plot
because
I'm all over the
place.



I connect people
through data, stories,
code and songs, I'm a
chord bubble diagram

[HTML] Beyond simple charts: Design of visualizations for big health data

[O Ola](#), [K Sedig](#) - Online journal of public health informatics, 2016 - ncbi.nlm.nih.gov

Health data is often big data due to its high volume, low veracity, great variety, and high velocity.

Big health data has the potential to improve productivity, eliminate waste, and support a ...

[☆ Save](#) [99 Cite](#) [Cited by 50](#) [Related articles](#) [All 10 versions](#)

[Evidence for teaching practices that broaden participation for women in computing](#)

..., [HH Hu](#), [M Kallia](#), [F McNeill](#), [O Ola](#)... - Proceedings of the ..., 2021 - dl.acm.org

Computing has, for many years, been one of the least demographically diverse STEM fields, particularly in terms of women's participation [12, 36]. The last decade has seen a ...

[☆ Save](#) [99 Cite](#) [Cited by 13](#) [Related articles](#) [All 6 versions](#)

[HTML] Understanding discussions of health issues on twitter: a visual analytic study

[O Ola](#), [K Sedig](#) - Online Journal of Public Health Informatics, 2020 - ncbi.nlm.nih.gov

Social media allows for the exploration of online discussions of health issues outside of traditional health spaces. Twitter is one of the largest social media platforms that allows users to ...

[☆ Save](#) [99 Cite](#) [Cited by 8](#) [Related articles](#) [All 5 versions](#)

Music, Stories, and Progress Clickers: Experiences Improving Classroom Climate with "Small" Socio-emotional Activities

[O Ola](#) - Proceedings of the 24th Australasian Computing ..., 2022 - dl.acm.org

Online learning is typically viewed as demotivating, and, for some students, isolating. These same characterizations have been used to describe large undergraduate courses in which ...

[☆ Save](#) [99 Cite](#) [Cited by 1](#) [Related articles](#)

Discourse with visual health data: Design of human-data interaction

[O Ola](#), [K Sedig](#) - Multimodal Technologies and Interaction, 2018 - mdpi.com

... Kamran Sedig and **Oluwakemi Ola** contributed to the design of the tool and the study; **Oluwakemi Ola** implemented the tool, performed the experiment, and analyzed the data. Both ...

[☆ Save](#) [99 Cite](#) [Cited by 15](#) [Related articles](#) [All 5 versions](#) [»](#)

[HTML] Beyond information access: Support for complex cognitive activities in public health informatics tools

[K Sedig](#), [P Parsons](#), [M Dittmer](#), [O Ola](#) - Online Journal of Public ..., 2012 - ncbi.nlm.nih.gov

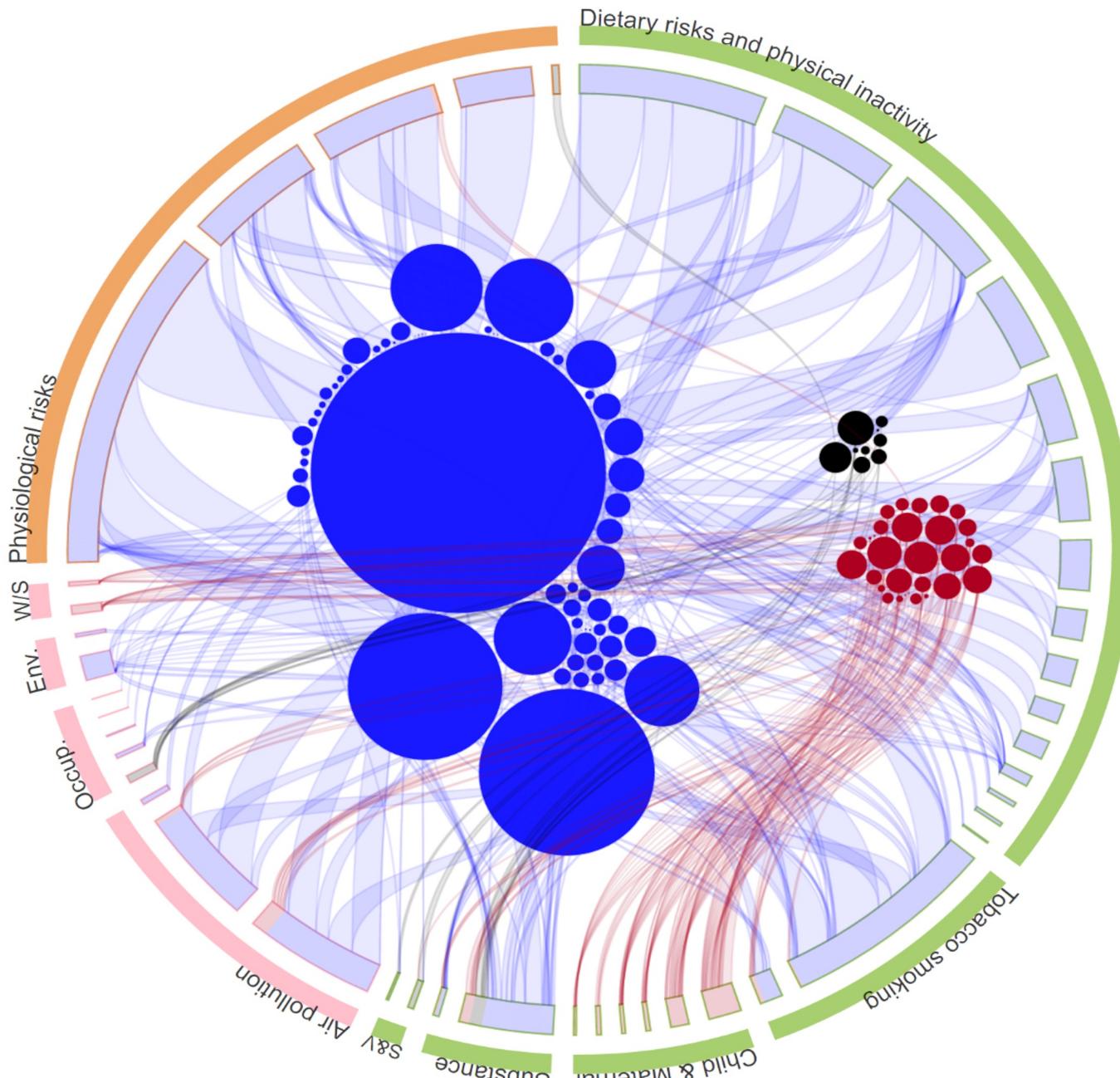
Public health professionals work with a variety of information sources to carry out their everyday activities. In recent years, interactive computational tools have become deeply ...

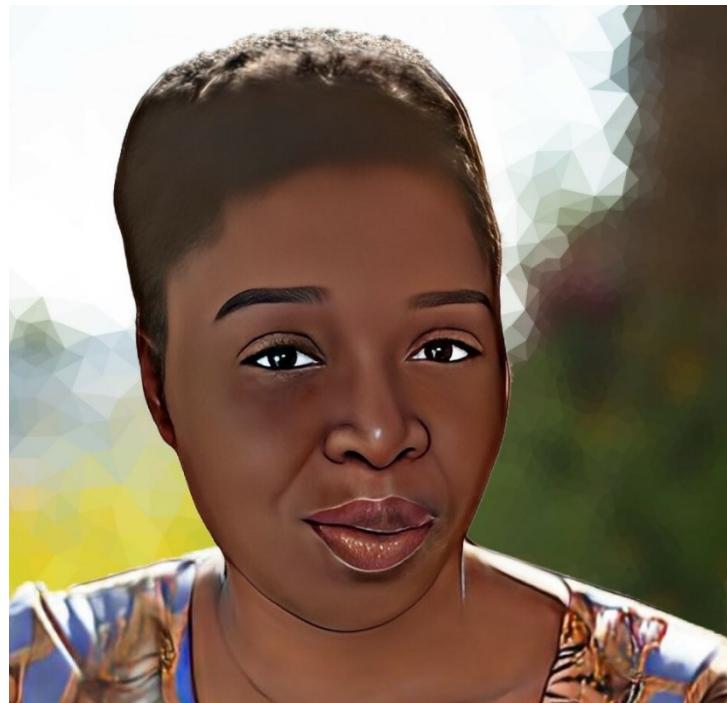
[☆ Save](#) [99 Cite](#) [Cited by 19](#) [Related articles](#) [All 9 versions](#)

Health literacy for the general public: Making a case for non-trivial visualizations

[O Ola](#), [K Sedig](#) - Informatics, 2017 - mdpi.com

Health literacy is concerned with the degree to which individuals can access and understand

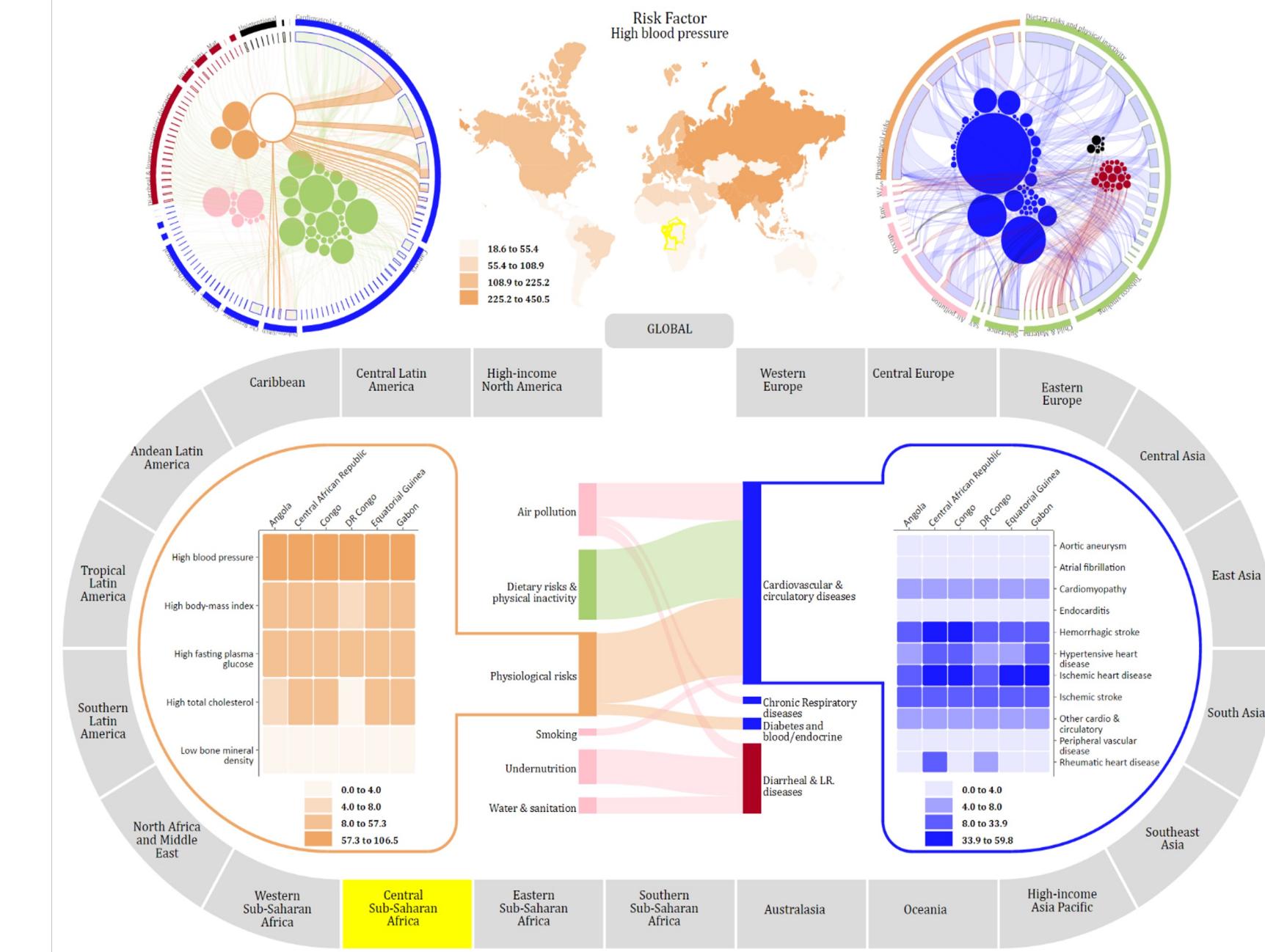


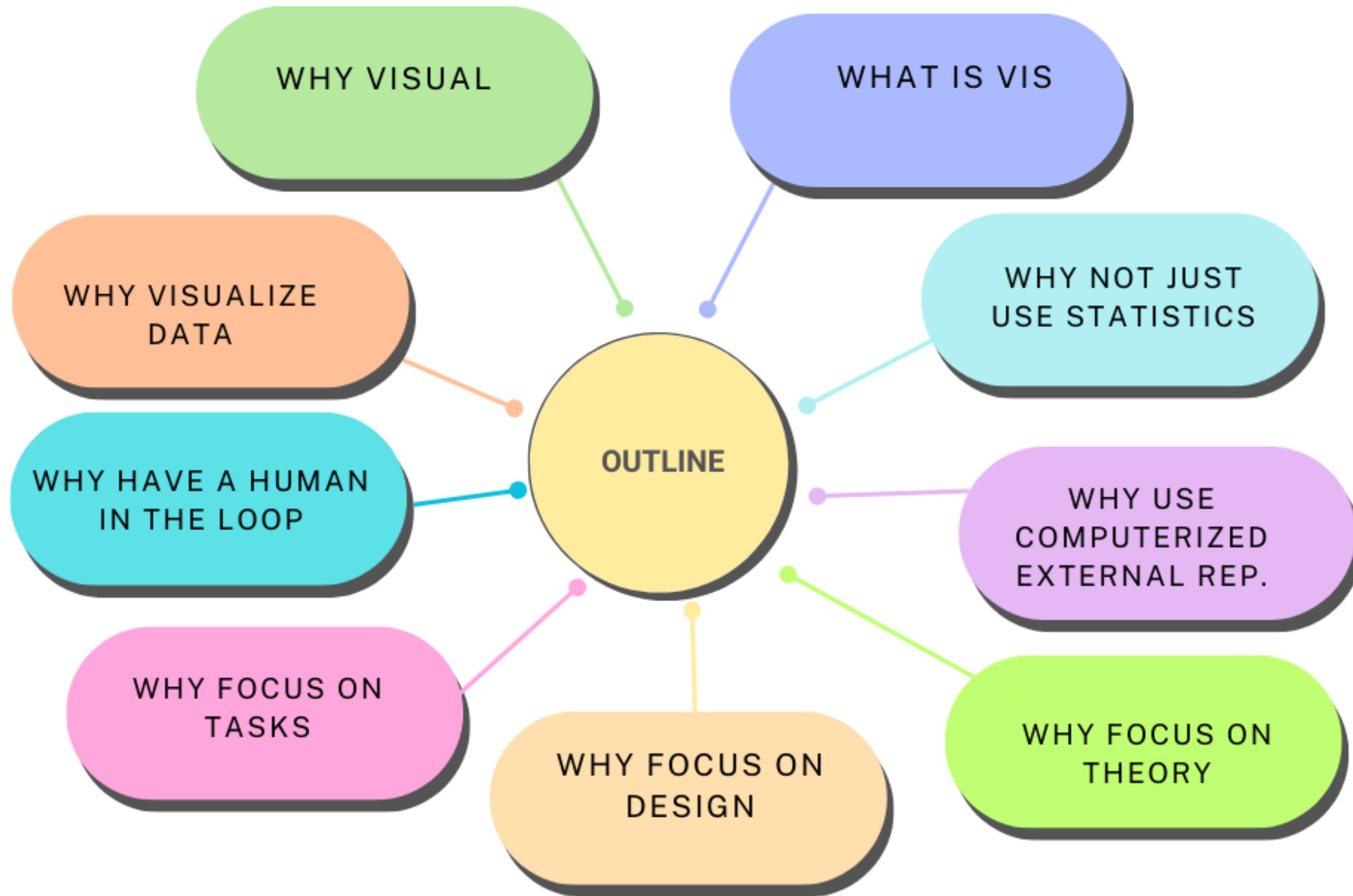


Dr. Oluwakemi Olamudzengi [Dr. K.]
Associate Professor of Teaching
Computer Science Department

Teaching
Computational Thinking (100)
Computation, Programs & Programming (110)
Introduction to Computer Systems (213)
Visualization for Data Science (DSCI 320)

Research Focus
Socio-emotional Learning
CS Pedagogy
Information Visualization

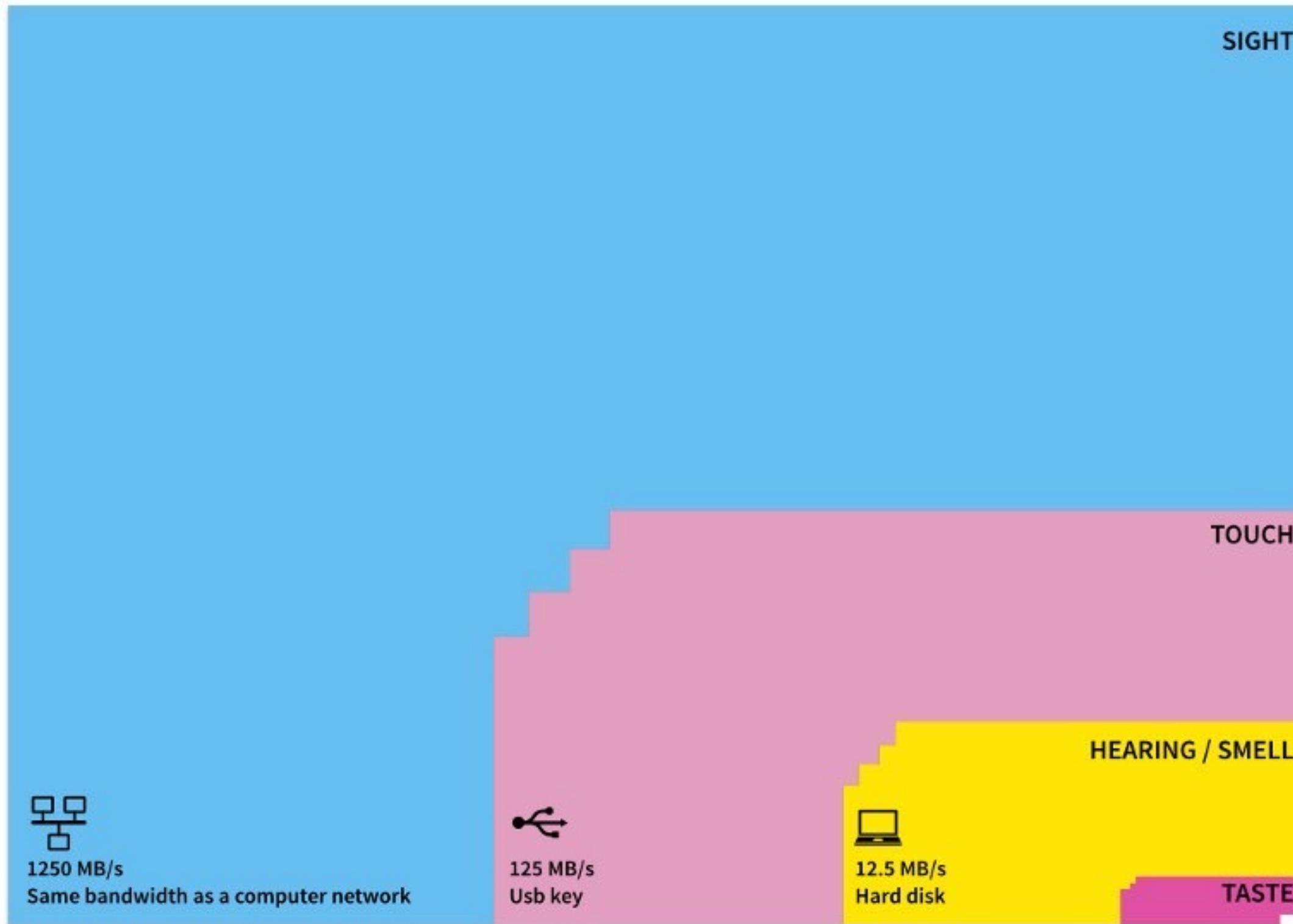




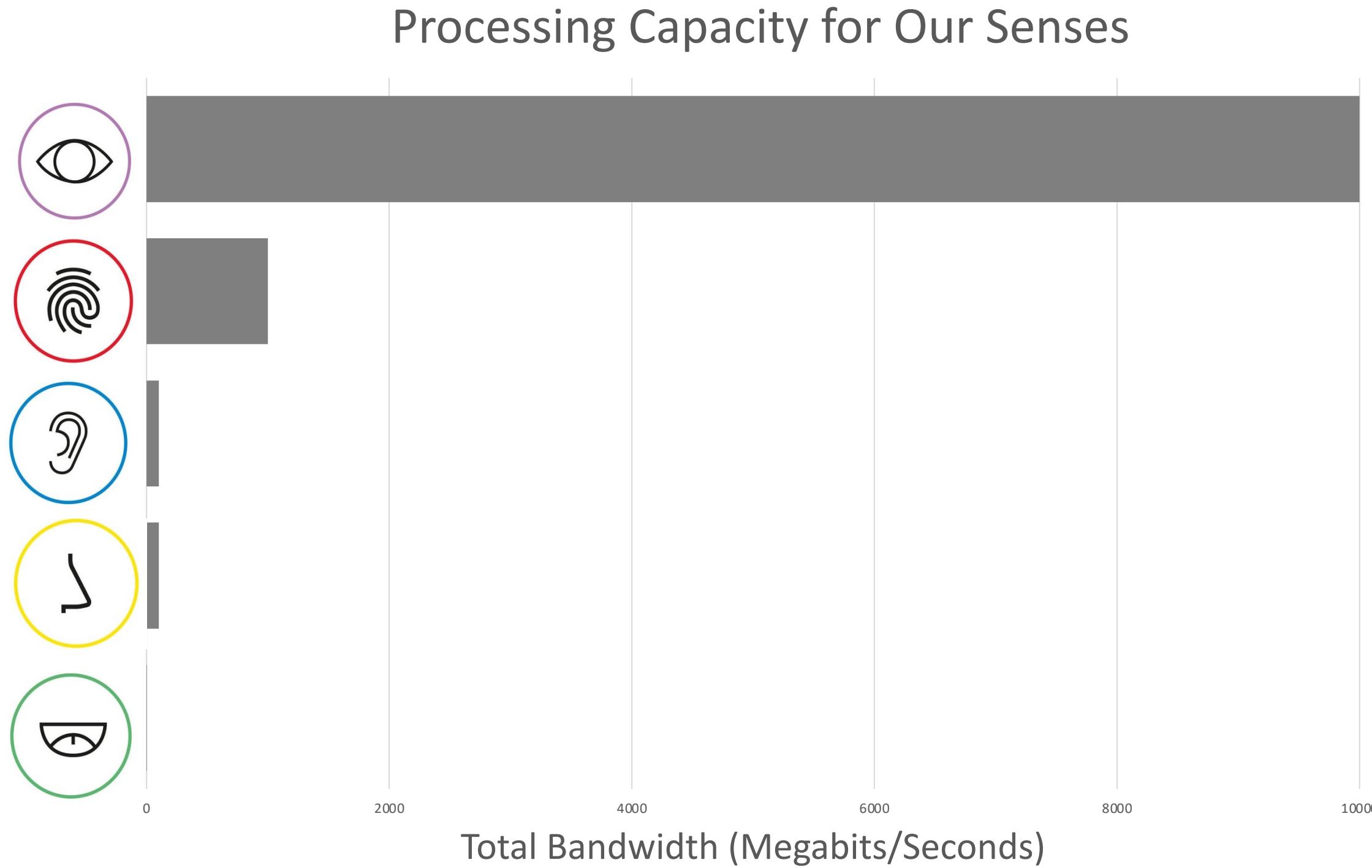
Why visual?

Why visual?

Bandwidth of our Senses by Tor Nørretranders



Why visual?



Why depend on vision?

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

- 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

They are scattered through all 50 states, the District of Columbia and Puerto Rico — 623 in Utah, 1,114 in Kansas, 101 way out in Alaska. They are clustered by the thousands in large Southern cities like Dallas, Atlanta and Memphis, and huddled in handfuls in unlikely hamlets like Shell Knob, Mo. (pop. 1,393) and Fountain Run, Ky. (pop. 236).

Evacuees fled Hurricane Katrina and the floods that followed in caravans of cars and fleets of buses, on helicopters and chartered planes, by boat and, a few, on foot. A month after the storm, a map

emerges of where they landed, based on ZIP codes from which applications for aid were submitted to the Federal Emergency Management Agency as of Sept. 23.

Of 1,356,704 applications, 86 percent came from Louisiana, Mississippi, Texas and Alabama. But 35,539 families were more than 1,000 miles from the Gulf — among the farthest: one in Nome, Alaska, 3,931 miles from the French Quarter and another in Lihue, Hawaii, 4,279 miles away.

Residents of New Orleans, a city that was two-thirds black, seem to have flocked to the nation's African-American population

centers. On average, the applicants came from counties where blacks were 28 percent of the population, more than twice the national average.

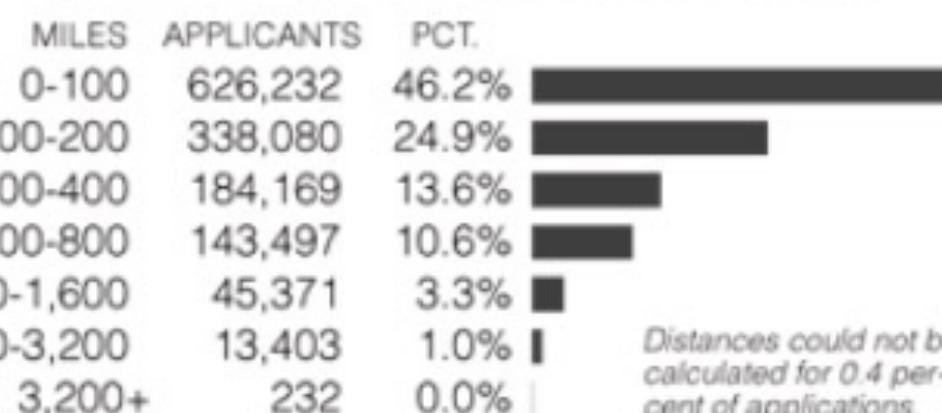
Baton Rouge, La., appears to be temporary home to 10 percent of evacuees, Houston 6.25 percent. But after the top 18 hubs, applicants are spread like the wind that whipped through their old neighborhoods: none of the other 900-plus metropolitan areas has even 1 percent of the total.

Some 4,000 ZIP codes — among them Pocahontas, Miss.; Promise City, Iowa; and Hope, Mich. — had just one applicant.

Applications by state

Louisiana	523,149	38.6%
Mississippi	383,840	28.3%
Texas	156,895	11.6%
Alabama	109,469	8.1%
Georgia	35,342	2.6%
Florida	31,005	2.3%
Tennessee	15,529	1.1%
Arkansas	11,027	0.8%
California	10,953	0.8%
Illinois	6,430	0.5%
Others	73,065	5.4%

Applications by distance from New Orleans

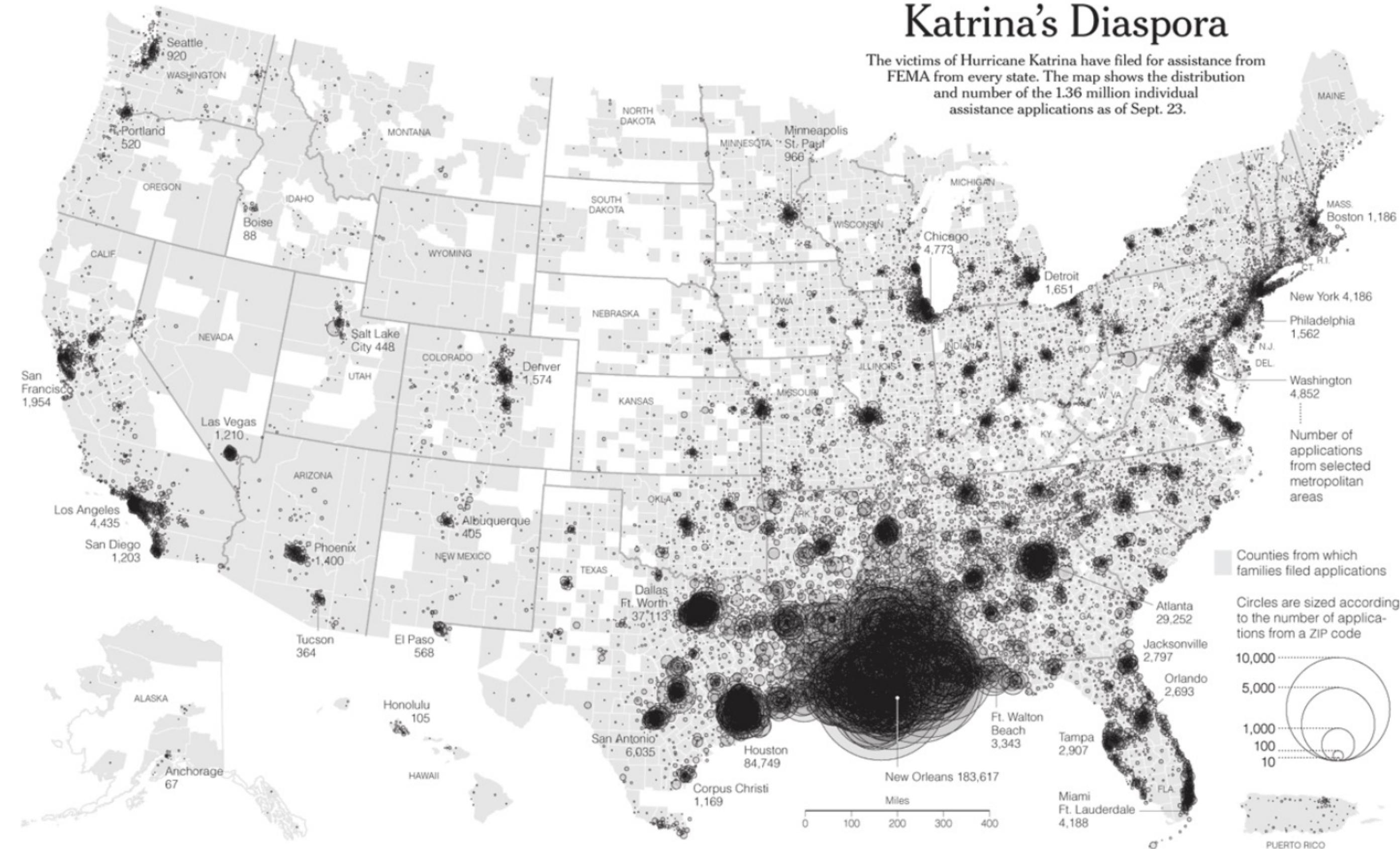


Distances could not be calculated for 0.4 percent of applications.

Sources: FEMA; Census Bureau; Queens College Sociology Department
Matthew Ericson, Archie Tse and Jodi Wilgoren/The New York Times

Katrina's Diaspora

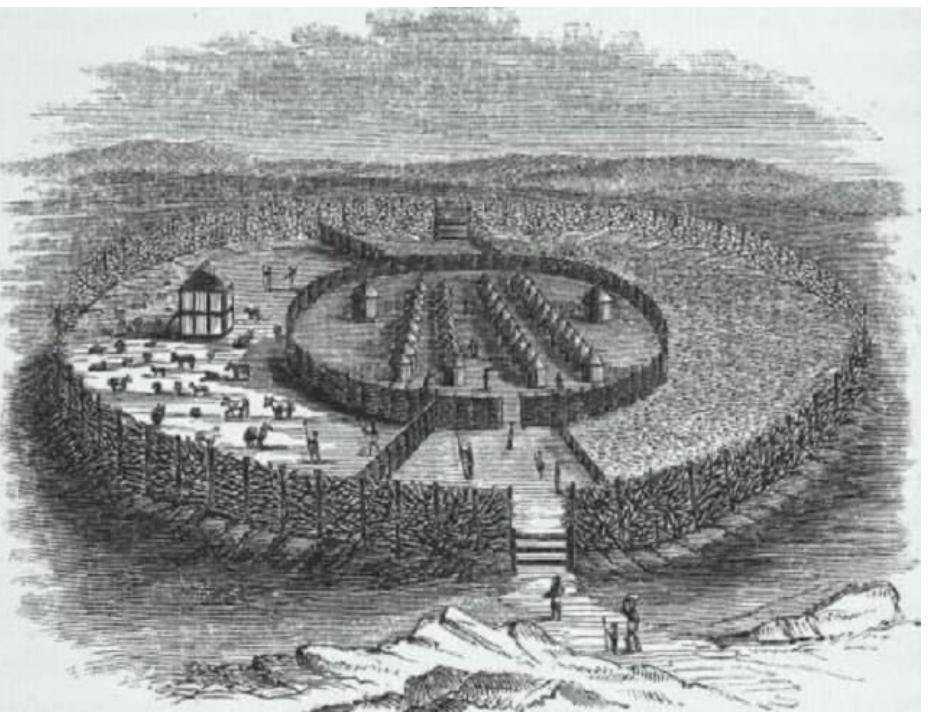
The victims of Hurricane Katrina have filed for assistance from FEMA from every state. The map shows the distribution and number of the 1.36 million individual assistance applications as of Sept. 23.



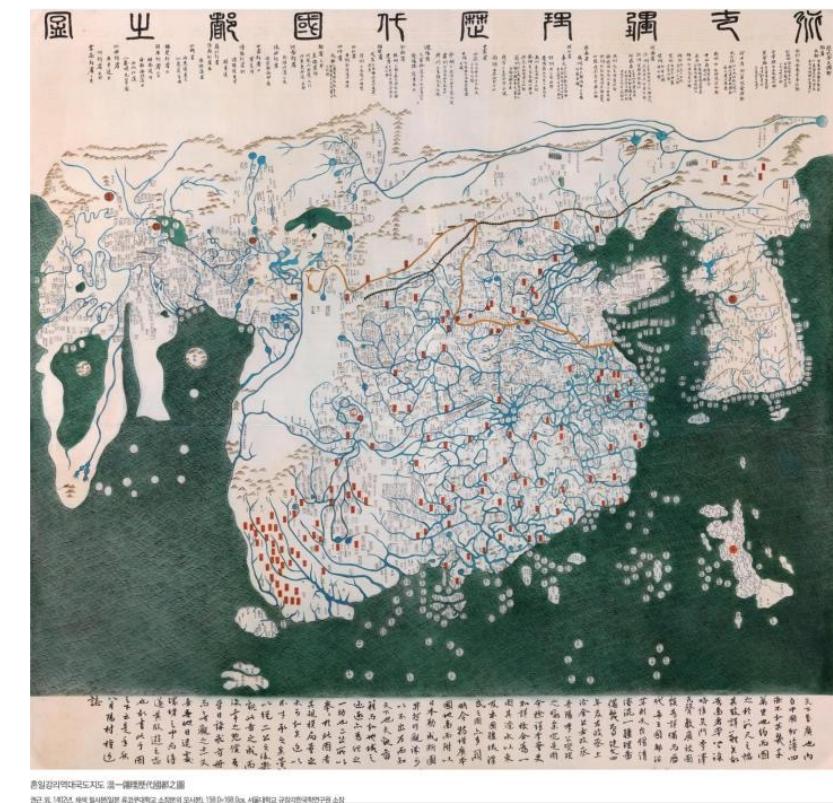
What is visualization?



PTOLEMY (c. 150)

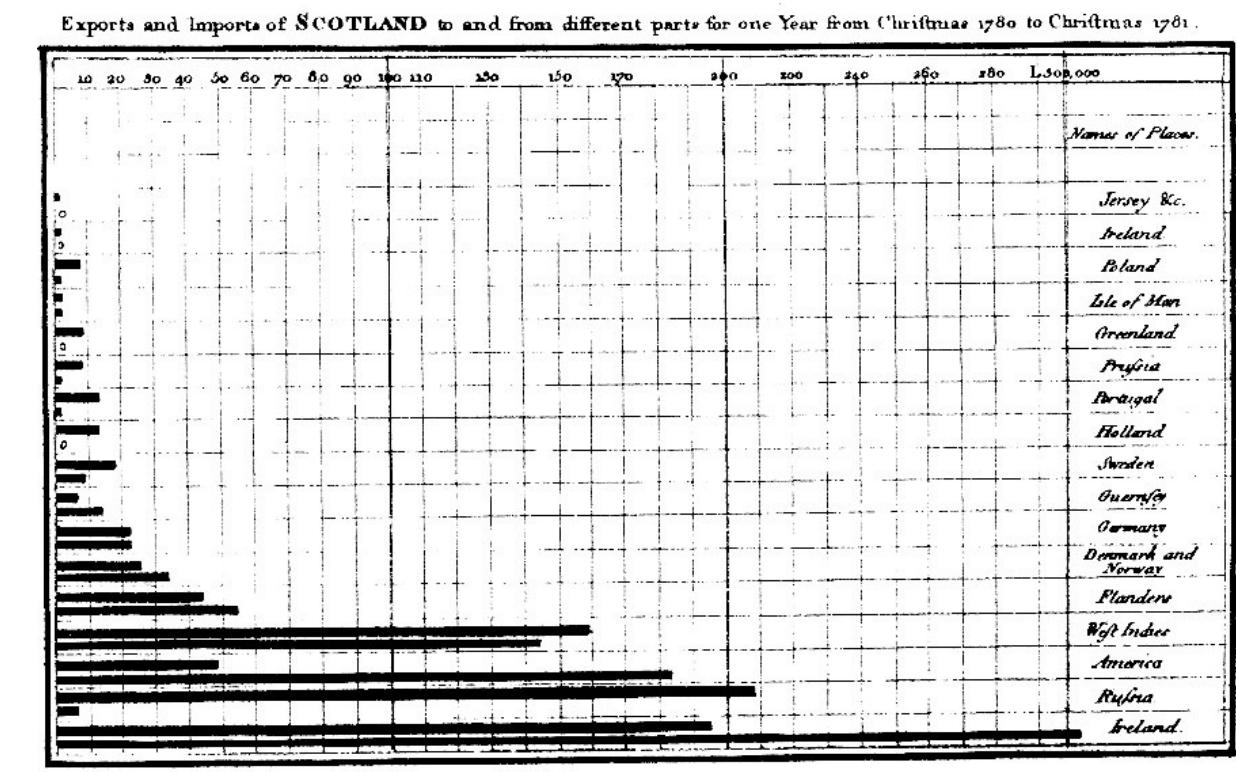


BENIN KINGDOM (c. 1600)

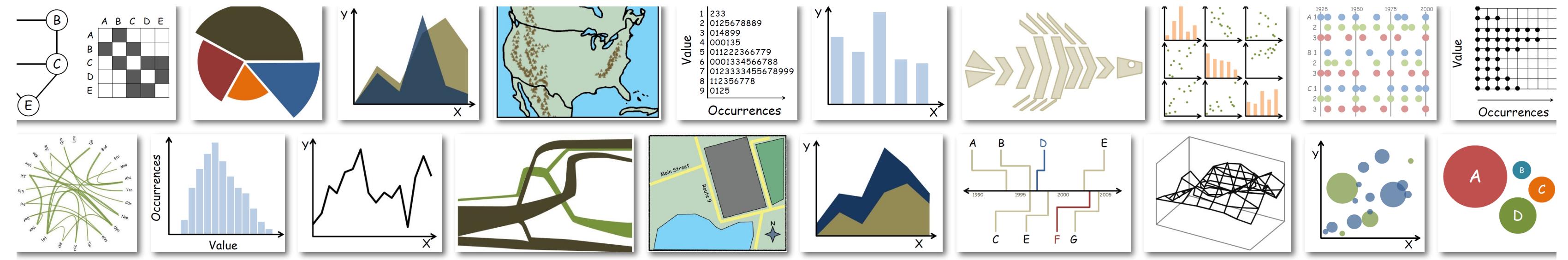


KANGNIDO (c. 1402)

WILLIAM
PLAYFAIR (c. 1786)



The upright divisions are Ten Thousand Pounds each. The Black Lines are Exports the Ribbed Lines Imports.
Published in the Act of Parliament for the year of 1786 by W. Playfair.

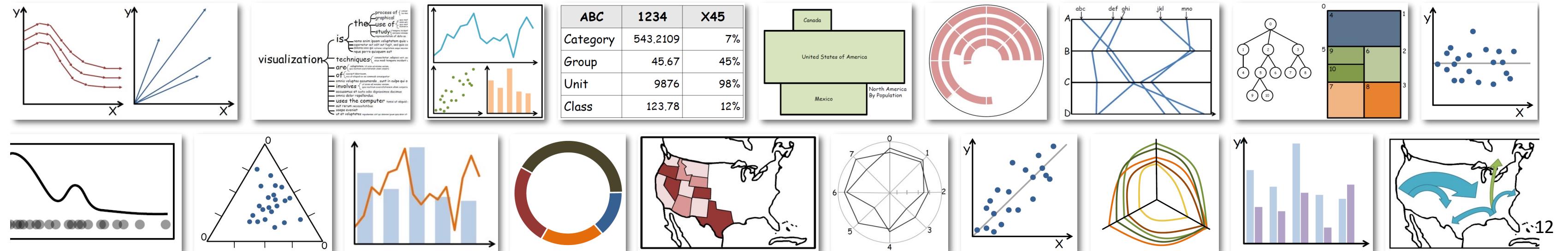


static or interactive

abstract or spatial

The visual representation of data to enhance cognitive tasks

problem solving, decision making, planning, sensemaking, storytelling



Why visualize data?

Why viz. data?

Data is Invisible

- Data is an observed phenomenon recorded so it can be stored, processed, and communicated.
- Raw data has no physical form until we represent it.
- We usually represent data as text, numbers, images, or symbols.

Data Needs Context

- On its own, data has no meaning.
- Meaning comes from context and relationships (e.g., comparing values, linking to questions, or connecting to the real world).

Visualization provides context by giving form to data, making patterns and meaning visible.

5

Why viz. data?

Pattern Recognition

Humans excel at spotting visual patterns

We can see trends, outliers, clusters instantly

What takes minutes to calculate, we see in seconds

Why viz. data?

Pattern Recognition

Humans excel at spotting visual patterns

We can see trends, outliers, clusters instantly

What takes minutes to calculate, we see in seconds

Why viz. data?

Memory and Communication

Visual information sticks in memory longer

Teams can literally 'see' the same thing

Externalized representations are easier to manipulate

Why viz. data?

Scale and Complexity

Modern datasets are too large for spreadsheets

Visualization handles millions of data points simultaneously



**Why not just use
statistics?**

Why not just use statistics?

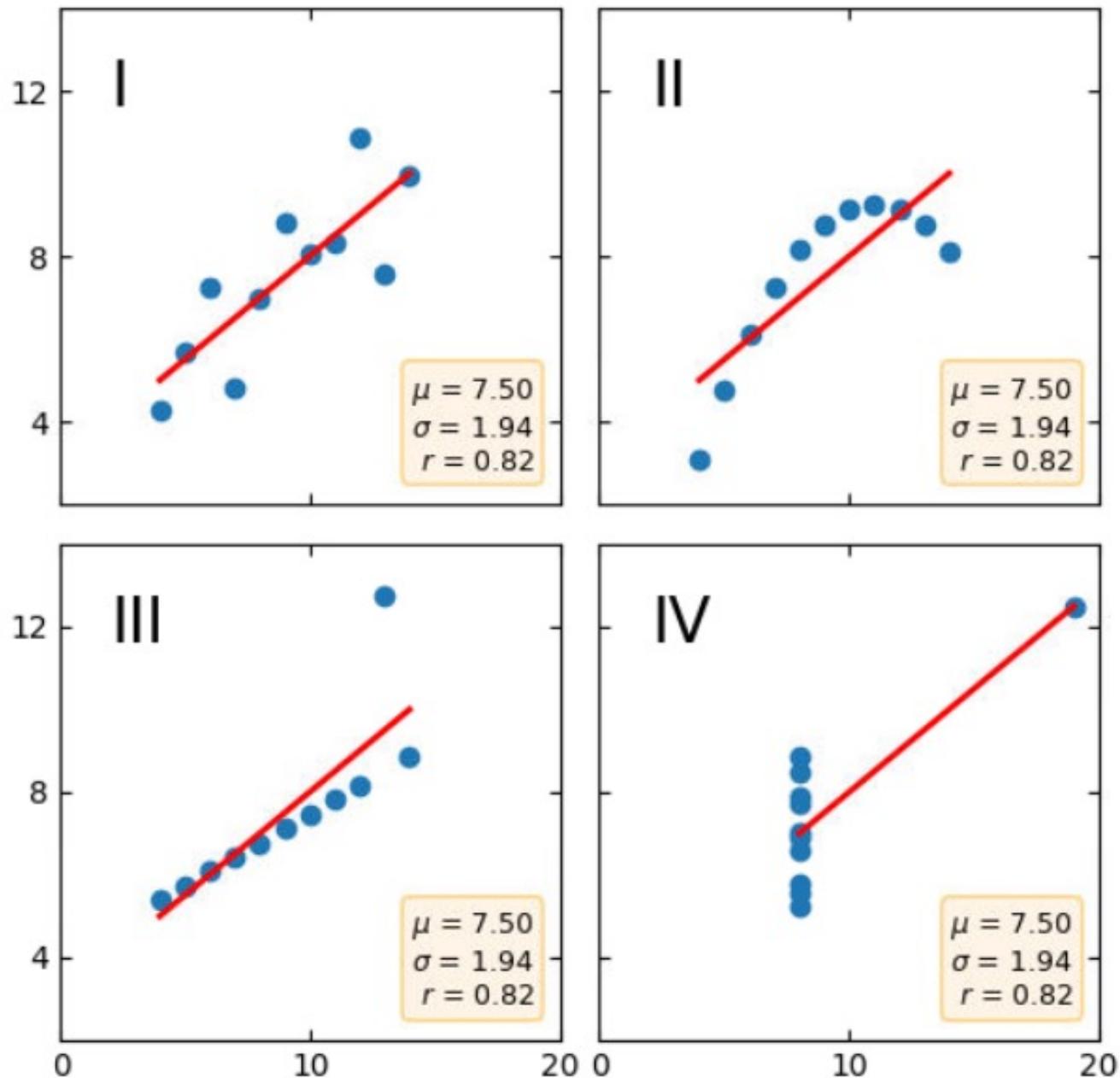
Anscombe's Quartet

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

These four datasets have the same means, variance, R-squared, correlations, and linear regression lines

Why not just use statistics?

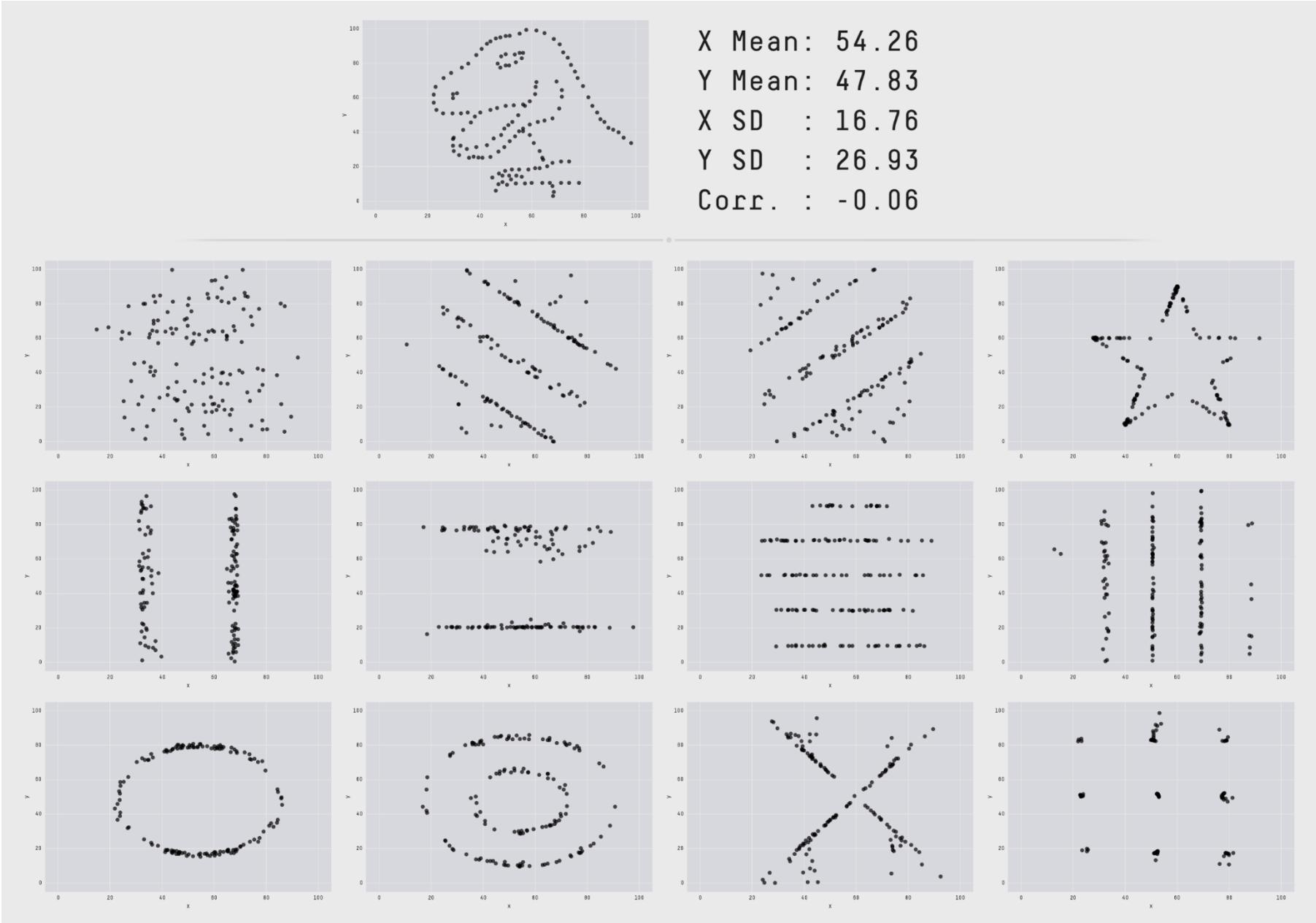
Anscombe's Quartet



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

Why not just use statistics?

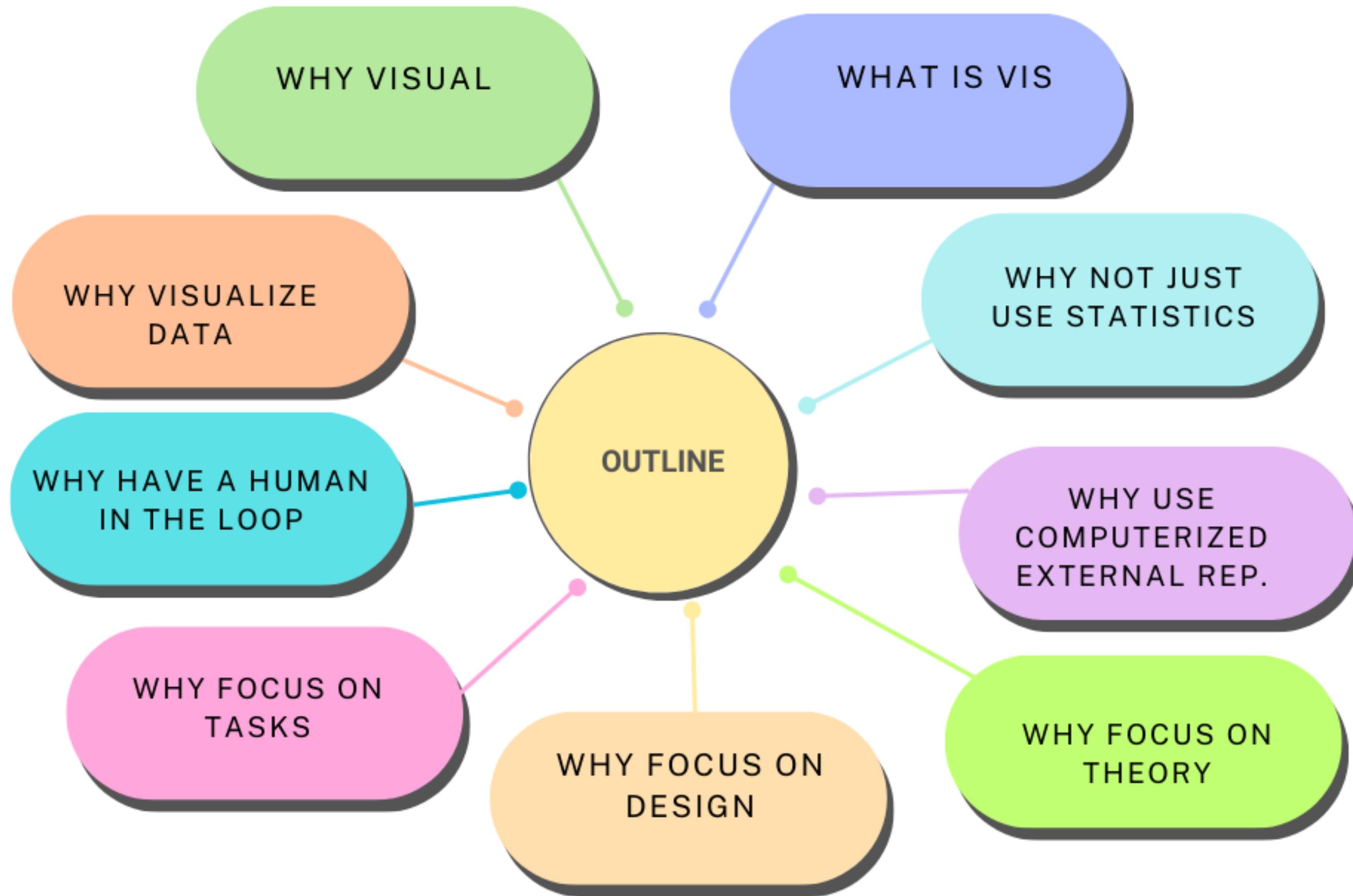
Anscombe's Quartet



<https://www.youtube.com/watch?v=DbJyPELmhJc>

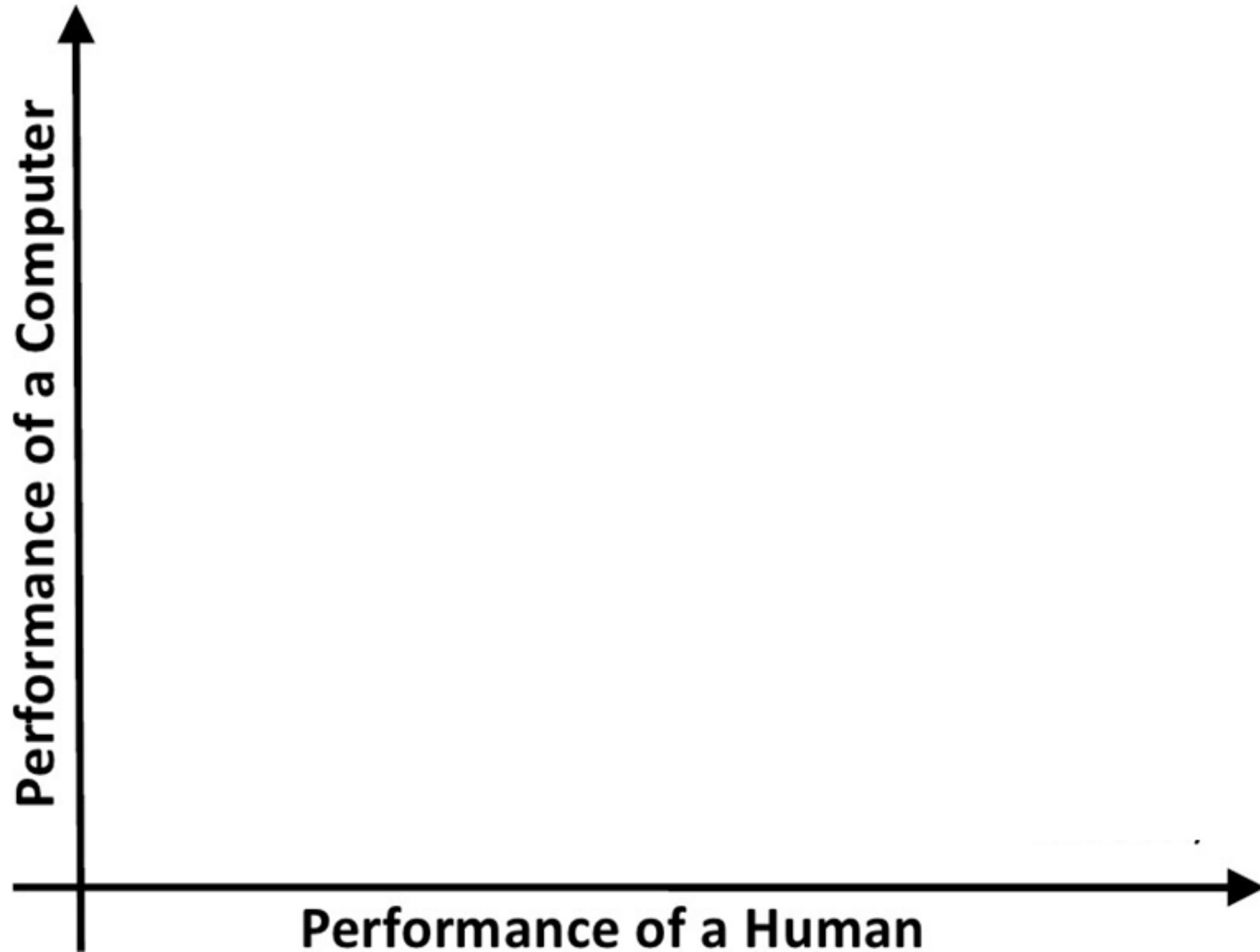
<https://www.autodesk.com/research/publications/same-stats-different-graphs>

<https://www.youtube.com/watch?v=RbHCeANCbW0>



**Why have a human in
the loop?**

Humans vs. Computers



Why have a human in the loop?

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

- many analysis problems ill-specified
 - don't know exactly what questions to ask in advance
- possibilities
 - long-term use for end users (ex: exploratory analysis of scientific data)
 - presentation of known results (ex: New York Times Upshot)
 - stepping stone to assess requirements before developing models
 - help automatic solution developers refine & determine parameters
 - help end users of automatic solutions verify, build trust

**Why use computerized
external representations?**

Why use computerized external representation?

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

- External representation: replace cognition with perception
- Interaction – allows us to interact with the data
- Integration – integrate with algorithms
- Make visualization part of a data analysis pipeline
- Efficiency- re-use charts/methods for different datasets
- Quality- precise data driven rendering
- Storytelling – use time

Spend 5 minutes exploring the visualization at the url

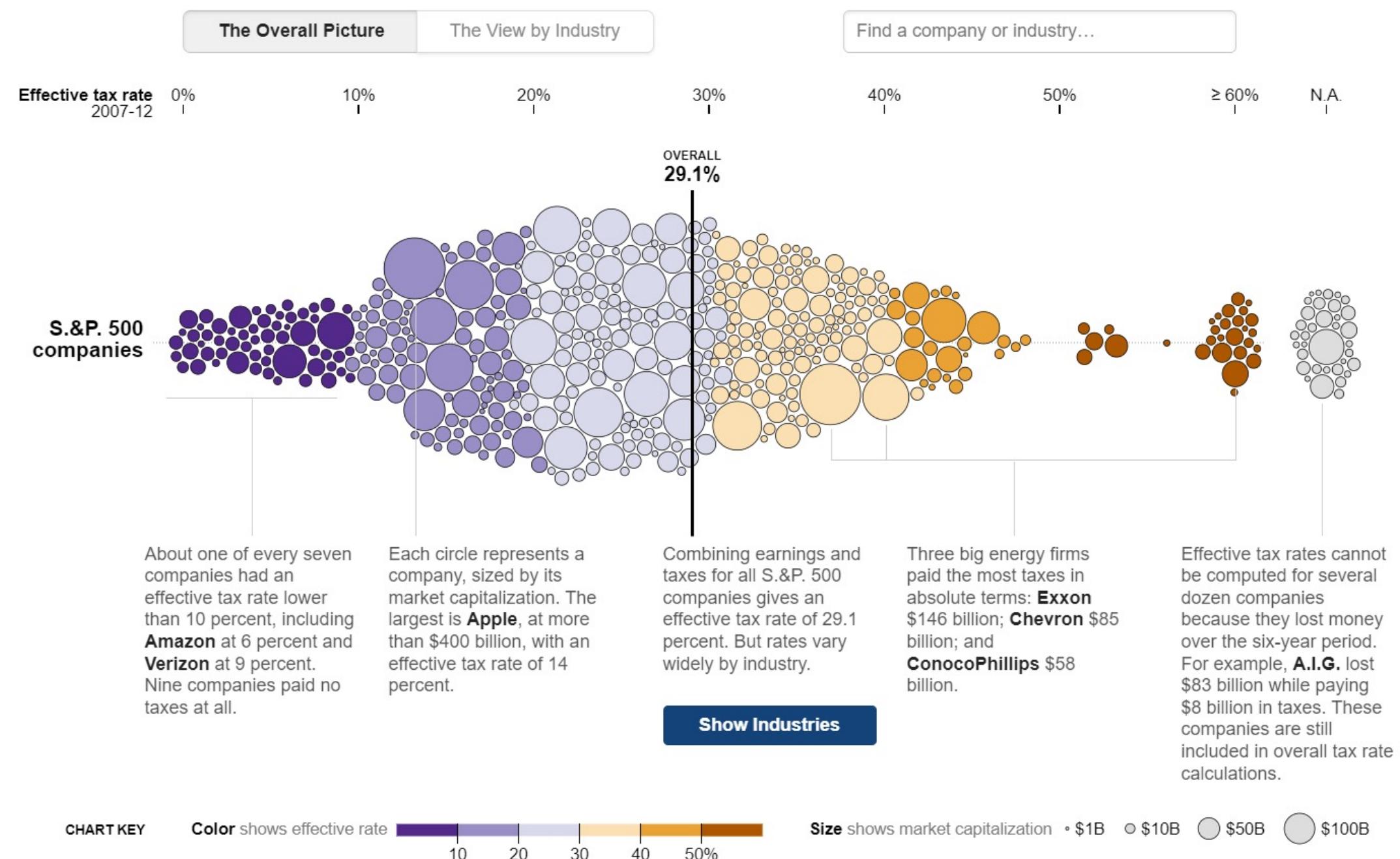
Then using the sheets on your table, describe and critique the visualization

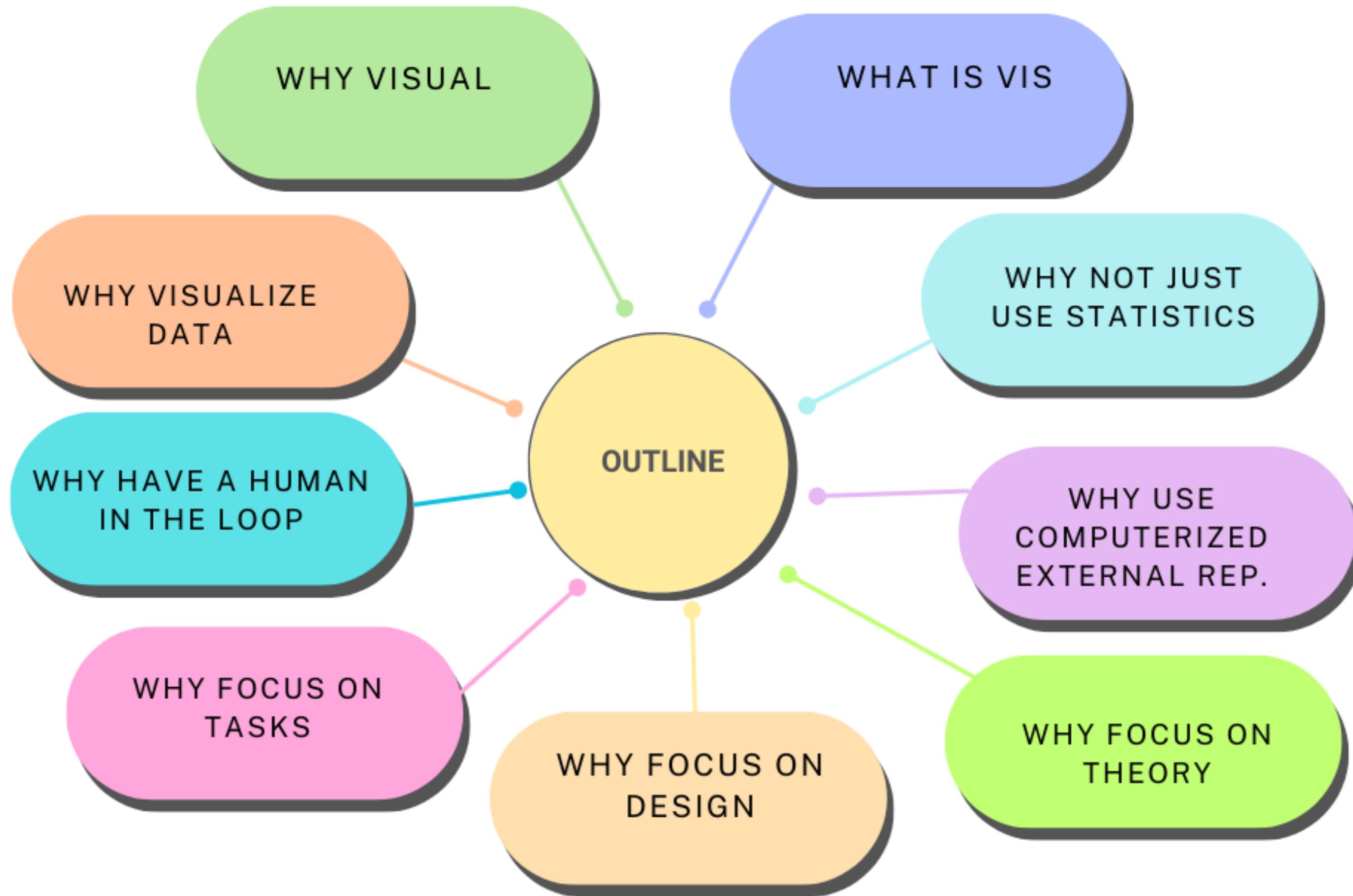
<https://tinyurl.com/dsci320-lec1>

When you are done, take a picture, go to PL and load your image to the class activity for today

Across U.S. Companies, Tax Rates Vary Greatly

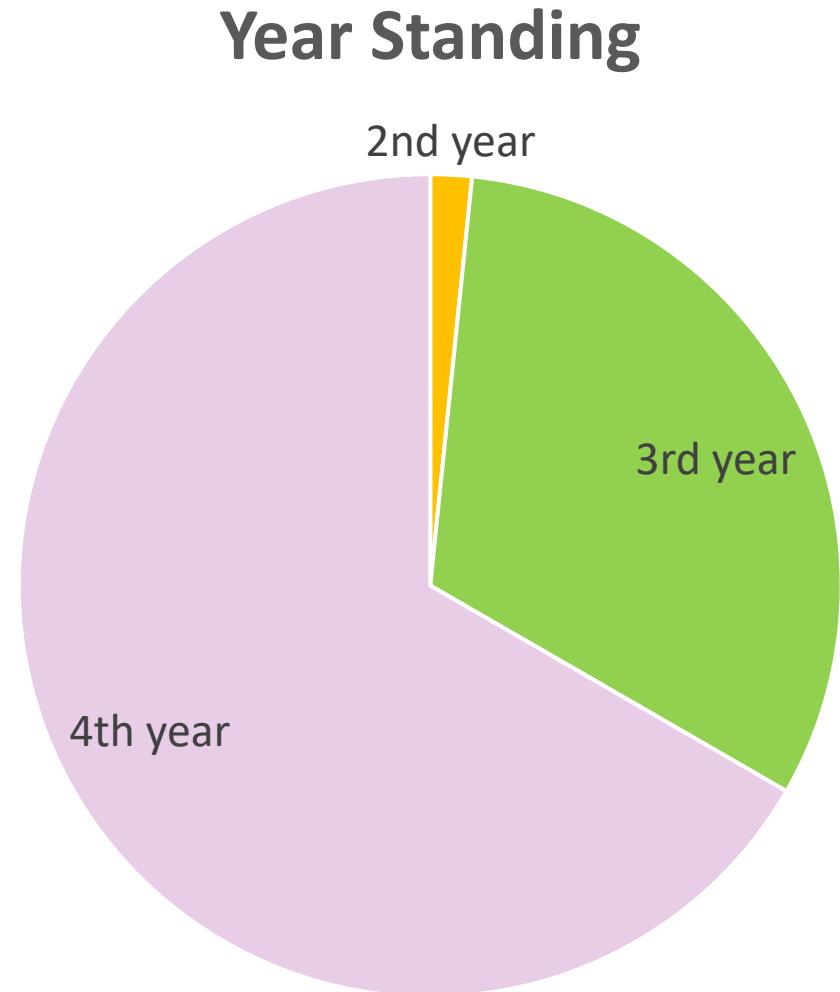
Last week, in a Congressional hearing, Apple got grilled for its low-tax strategy. But not every business can copy that approach. Here is a look at what S.&P. 500 companies paid in corporate income taxes — federal, state, local and foreign — from 2007 to 2012, according to S&P Capital IQ. [Related Article »](#)





**Why focus on tasks,
design and theory**

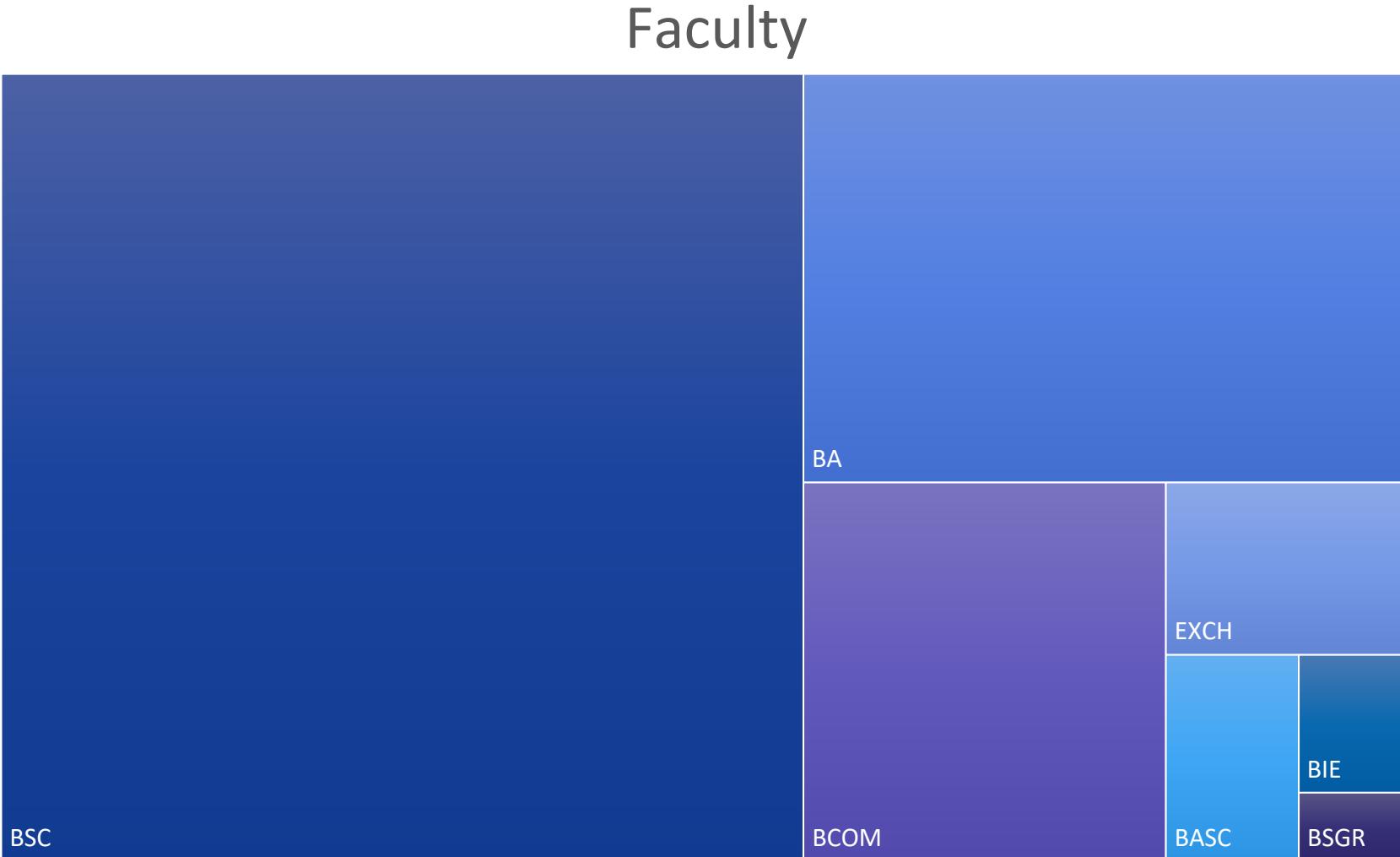
Who is in the room?



What percentage of students in our course
are in second year?

What year standing is most common in our
class?

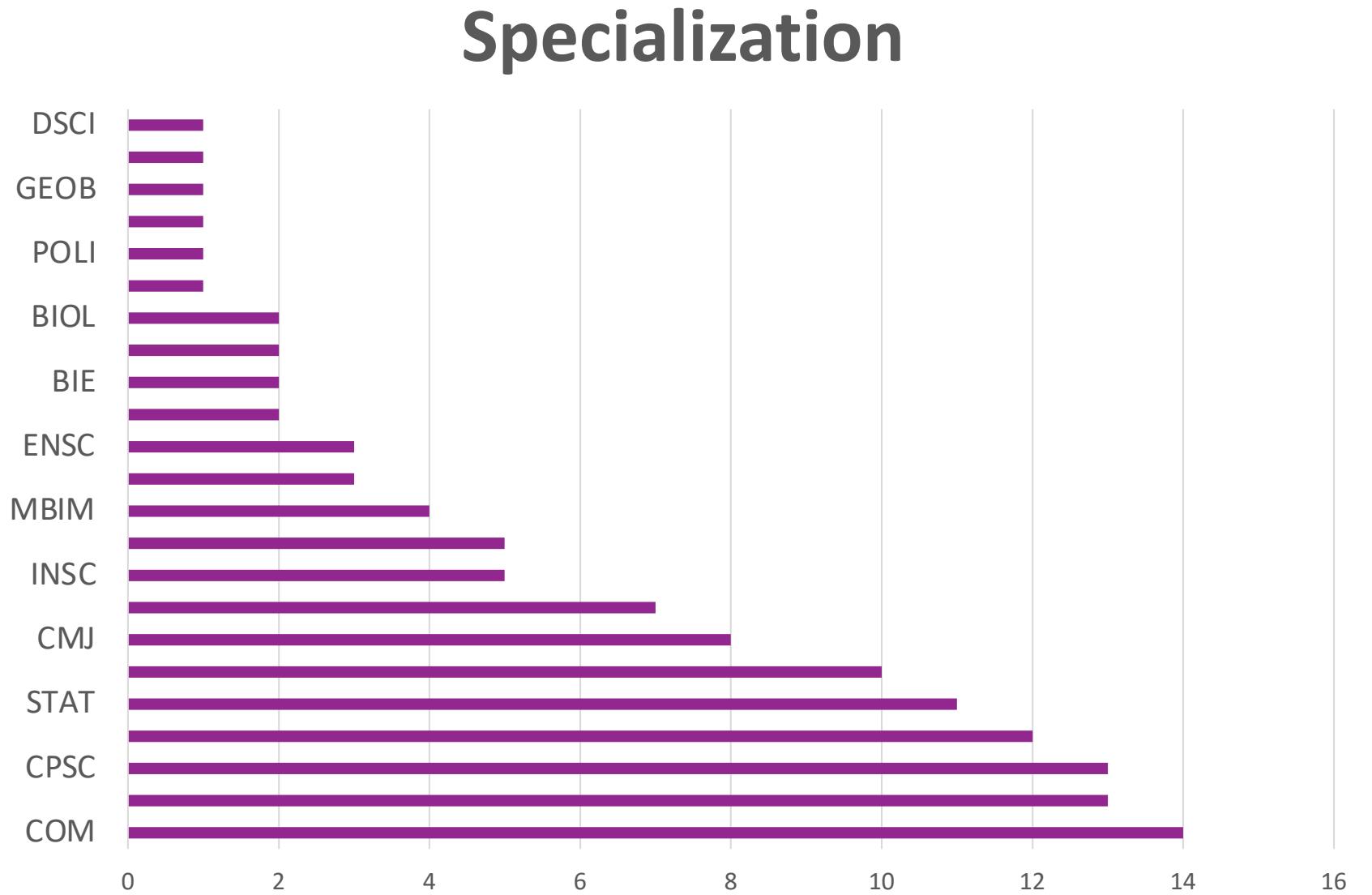
Who is in the room?



How many faculties are represented?

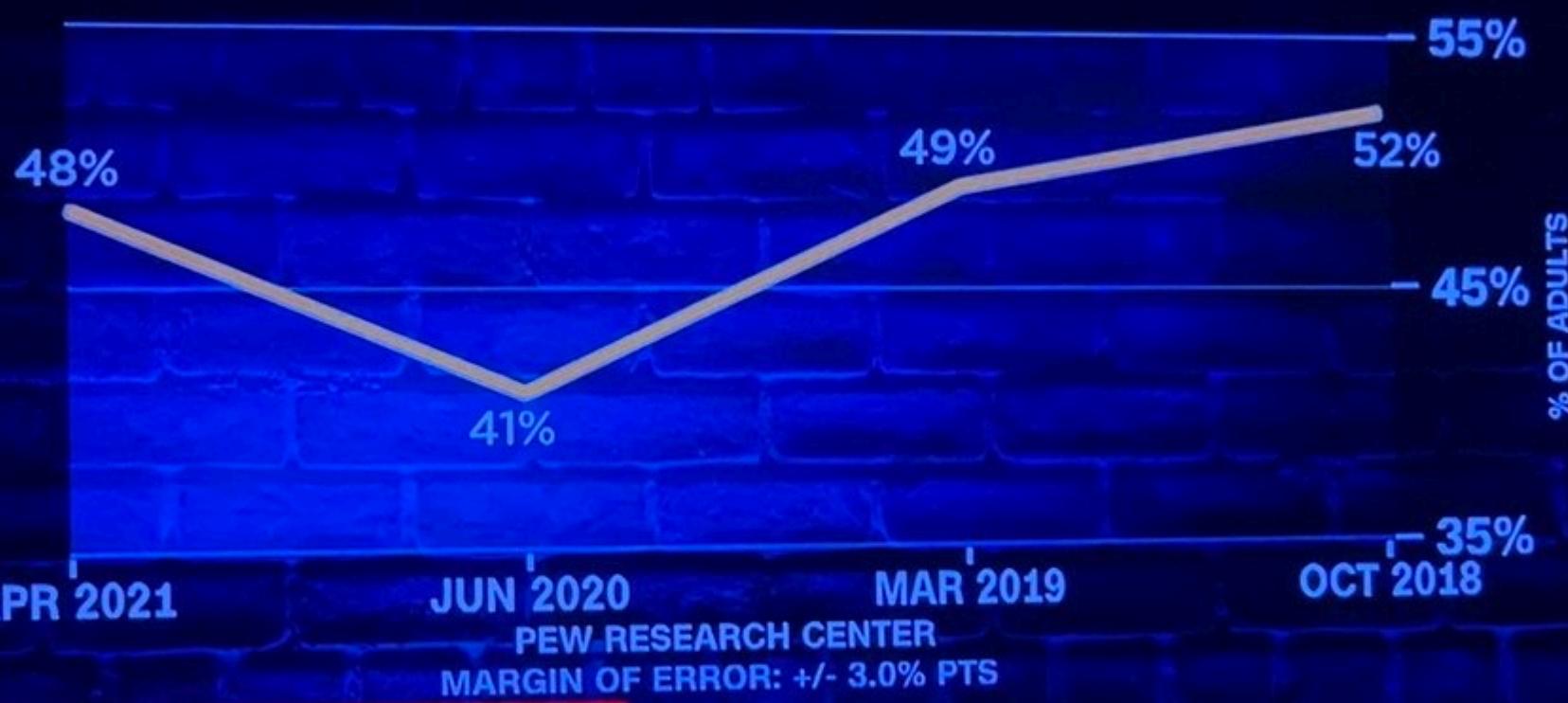
How many students are exchange students?

Who is in the room?



What proportion of our class
comes from INSC?

VIOLENT CRIME IS A VERY BIG PROBLEM ADULTS



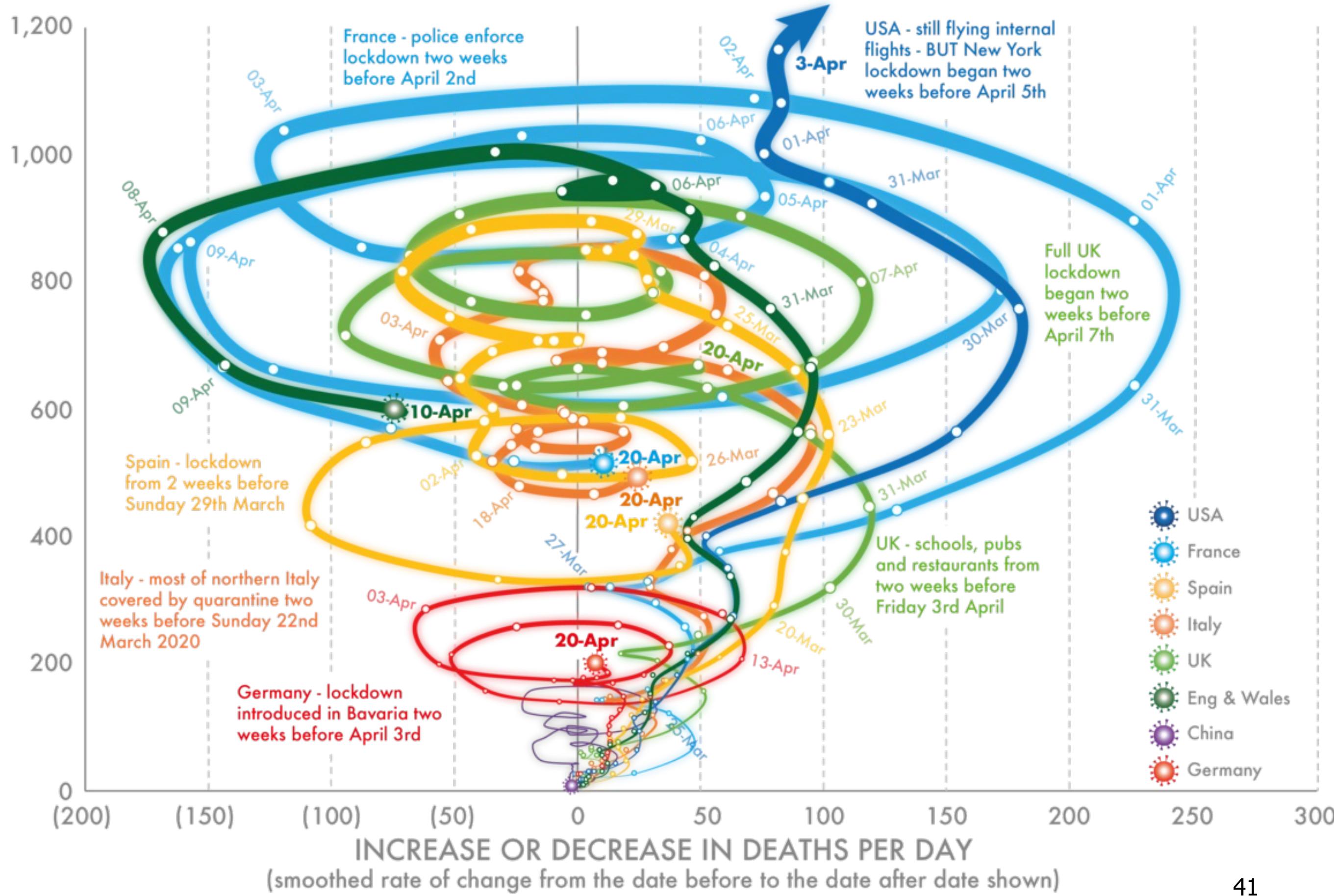
THE WIZARD OF ODDS

WHITE HOUSE PREPARES TO ADDRESS SURGE IN VIOLENT CRIME

"I LESS EAGER TO GET THE SHOT," JEFF ZIENTS SAYS ► BIDEN WANTED 70 CUOMO PRIME TIME

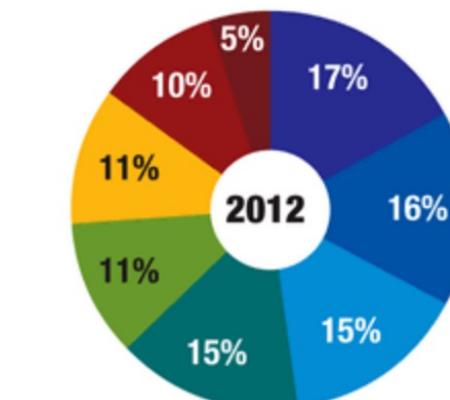
LIVE
CNN
6:32 PM PT

AVERAGE NUMBER OF DEATHS PER DAY on that date, the day before and the day after

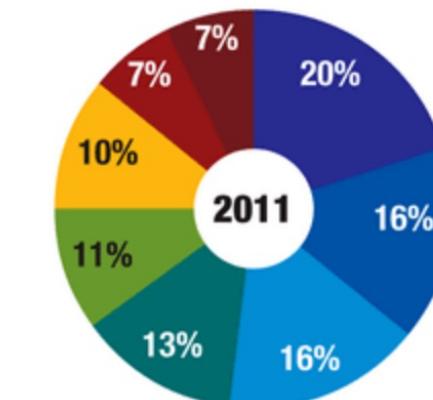




How many nights do you typically stay away from home due to work?



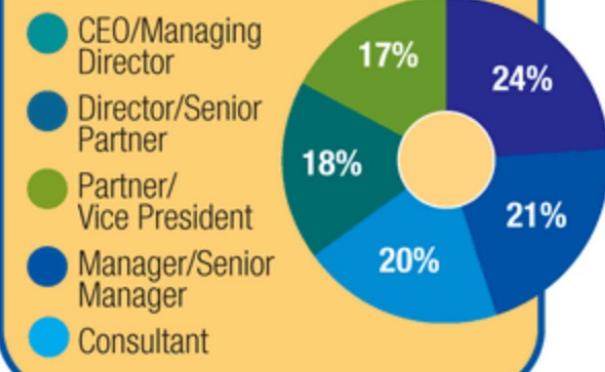
- Fewer than 25
- 26-50 days
- 51-75 days
- 76-100 days
- 101-125 days
- 126-150 days
- 151-200 days
- 200+ days



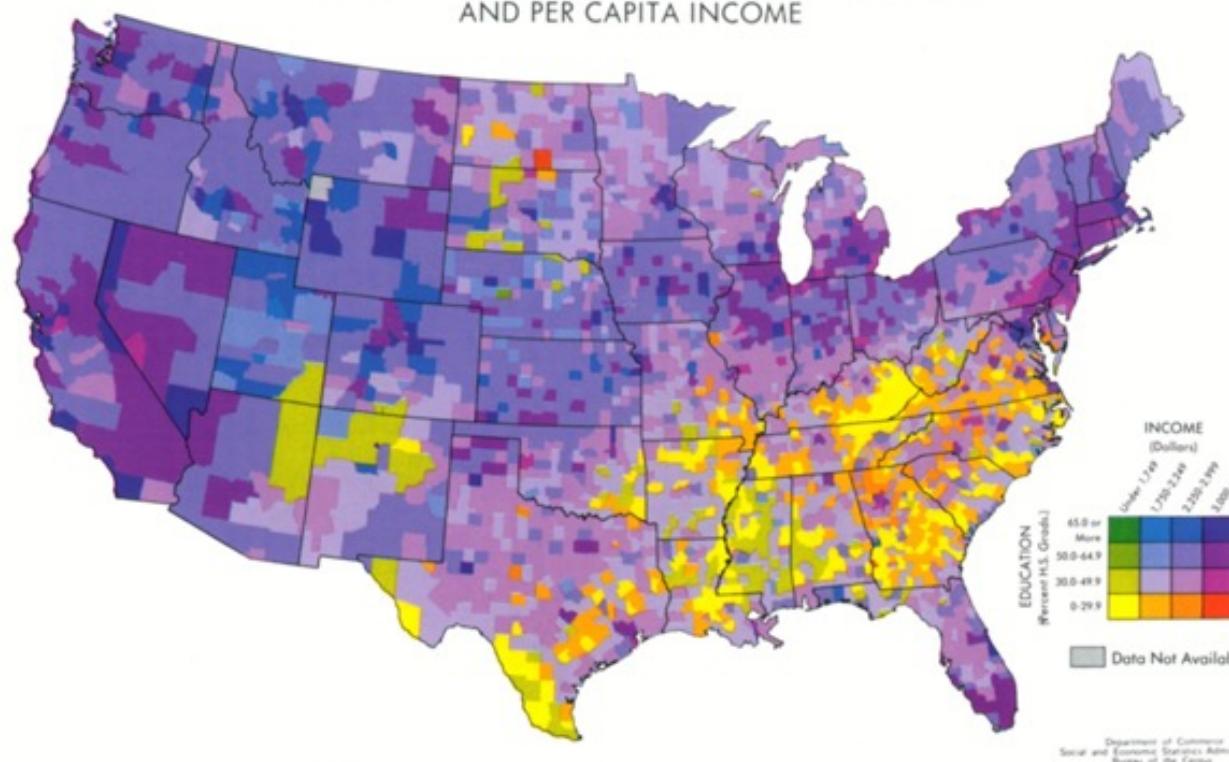
- Fewer than 25
- 26-50 days
- 51-75 days
- 76-100 days
- 101-125 days
- 126-150 days
- 151-200 days
- 200+ days

Who Took The Survey?

Overall, 170 consultants took our annual Best Places to Stay survey in the spring of 2012.



INTERRELATIONSHIP OF EDUCATIONAL ATTAINMENT
AND PER CAPITA INCOME



<https://connorrothschild.github.io/v2/post/color-in-data-vis/>

<https://cartographicperspectives.org/index.php/journal/article/view/1538/1819>



AROUND
the
WORLD

GIANT HORNETS KILL 42 PEOPLE IN CHINA

LIVE
CNN

RIGHT NOW

CHICAGO



71°

DETROIT



66°

HOUSTON



77°

EALTH EVALUATION IN MIRIAM CAREY'S HOME AFTER DC C

CNN.com



Ege Erdil ✅
@EgeErdil2

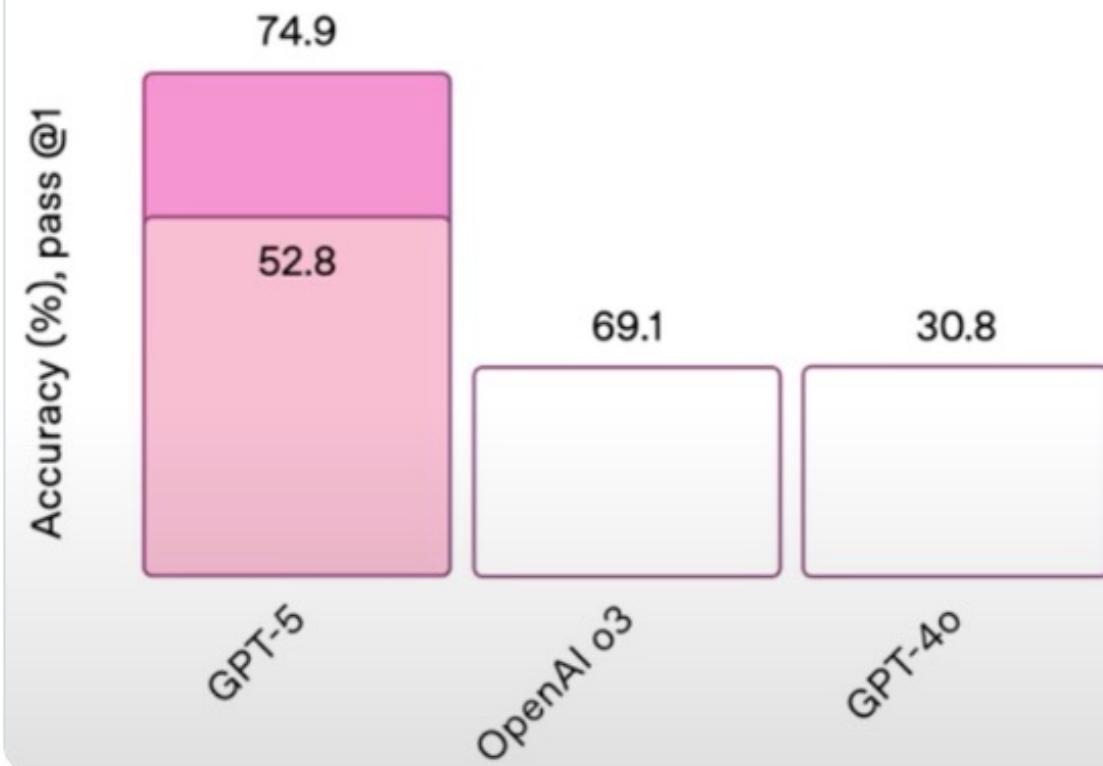
...

this screenshot from GPT-5 livestream has to be among the worst chart crimes of the century

SWE-bench Verified

Software engineering

● Without thinking ● With thinking



10:16 AM · Aug 7, 2025 · 812K Views

89

250

2.1K

186

↑

Read 89 replies

Altair Basics

Create a Chart Object

```
alt.Chart(data).mark_bar().encode(  
    channel_1 = 'column1',  
    channel_2 = 'column2',  
)
```

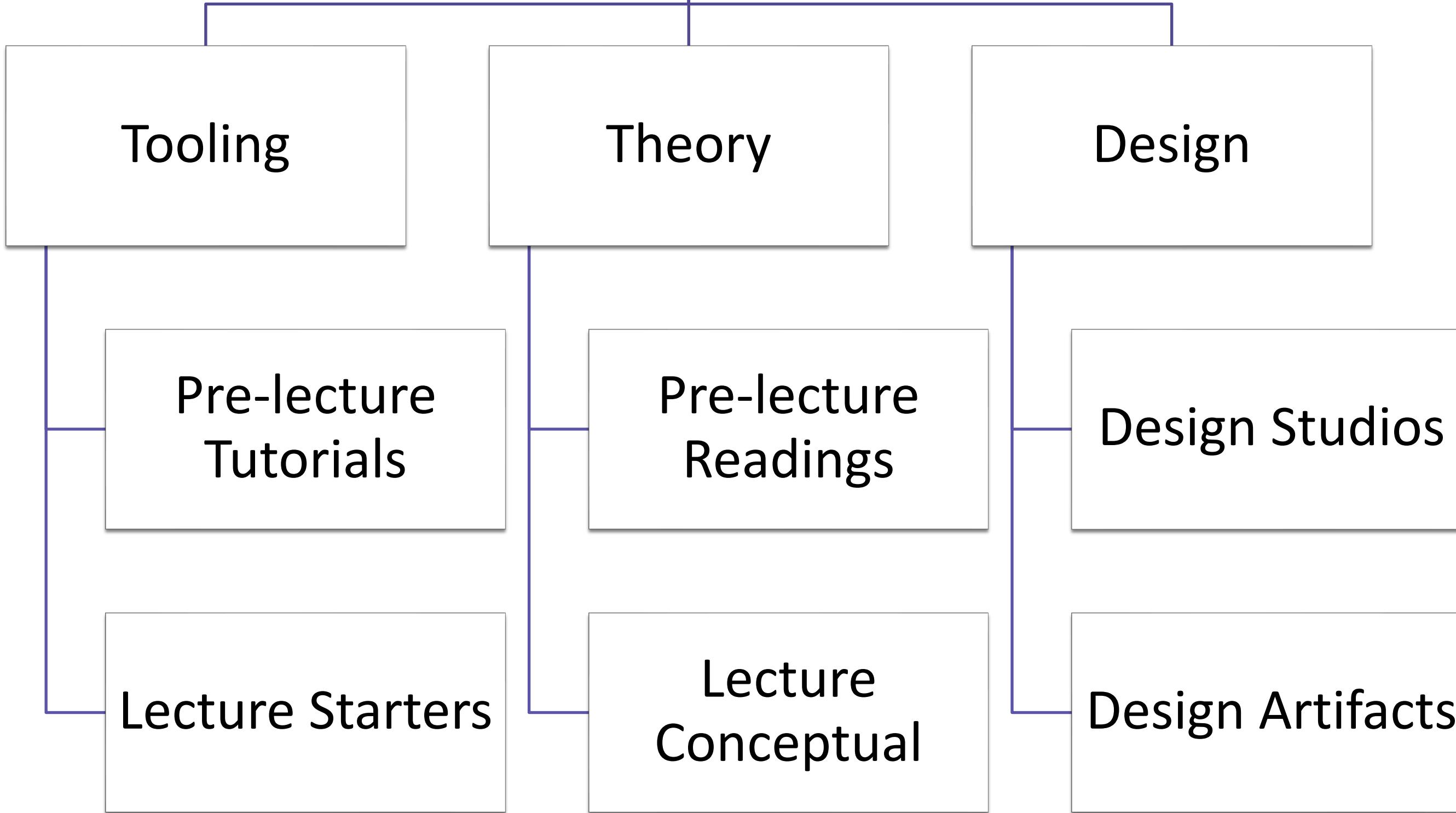
Attach data to the Chart Object

Specify the mark type

Specify each channel and what data it encodes

**“By the end of this course, my goal is
for you to have a toolkit to design,
create, critique, and interpret interactive
visualizations.**

Viz 4 DSCI



Weekly Rhythm

Knowledge Building

- Prep activities (readings, videos, coding) ~90 mins/week before lecture
- Lectures: interactive, semi-flipped, with clicker Qs + skill practice
- Weekly quizzes (50 mins, CBTF) track progress on programming + theory

Explore & Create

- Design Studios: hands-on workshops for sketching, critique, design thinking
- Design Artifacts: sketches, wireframes, reviews — showcase growth (2–4 hrs follow-up)

Capstone Journey

- Project Labs: team-based sessions → prototypes & real datasets
- Project Milestones: final project (dashboard + presentation) integrating design, theory & programming

In the first two weeks expect to spend more time on the programming aspects of the course

Assessment of Learning

Activity	Weight (%)	Details
Quizzes	30	Each worth 6% (drop lowest two). Proctored and must be completed in CBTF
Design Challenges	10	Worth 3%, 3%, and 4%; started in lab and completed the following week
Project	30	4 main deliverables, must work in a 3-person group. Project labs are worth 5%
Engagement	0 - 5	Can be used to replace up to 5% of your final exam grade
Final Exam	25 - 30	Scheduled by Registrar: comprehensive assessment

Passing Criteria

All students must satisfy ALL conditions to pass the course:

- Pass the Final Exam with a grade of at least 50%,
- Pass the Project with a grade of at least 50%.

If a student does not satisfy these requirements, the student will be assigned the **lower** of their earned course grade or, a maximum overall grade of 45 in the course.

Get Stepping

- Read through the Syllabus
- Buy the required supplies
 - At least five different colors of pens—e.g., Pilot G2 packs 31128 or pencils
 - Plenty of white or near-white drawing paper
- Work through the JumpStart
 - Setup
 - Introduction to JupyterLab
 - Python Crash Course
- Work through the first tutorial