NYU

CS-GY 6313 B: Information Visualization

9/5/2024

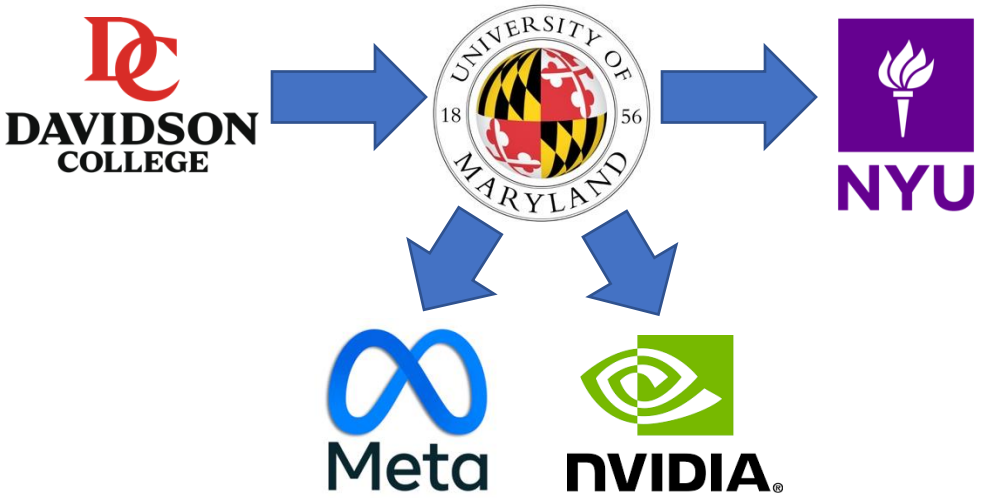
Who are we?

Instructor

Dr. Niall Williams
Research focuses on AR/VR, computer graphics, and applied human perception

"Niall" is pronounced the same as "Nile" (like the river)

nlw4415@nyu.edu



TA

Lokesh Shanmugam
"Hi everyone, I'm Lokesh, your TA and a second-year CS graduate student. I'm excited to work with you this semester! I will be posting my office hours soon, and in the meantime, feel free to reach out to me at ls6110@nyu.edu"

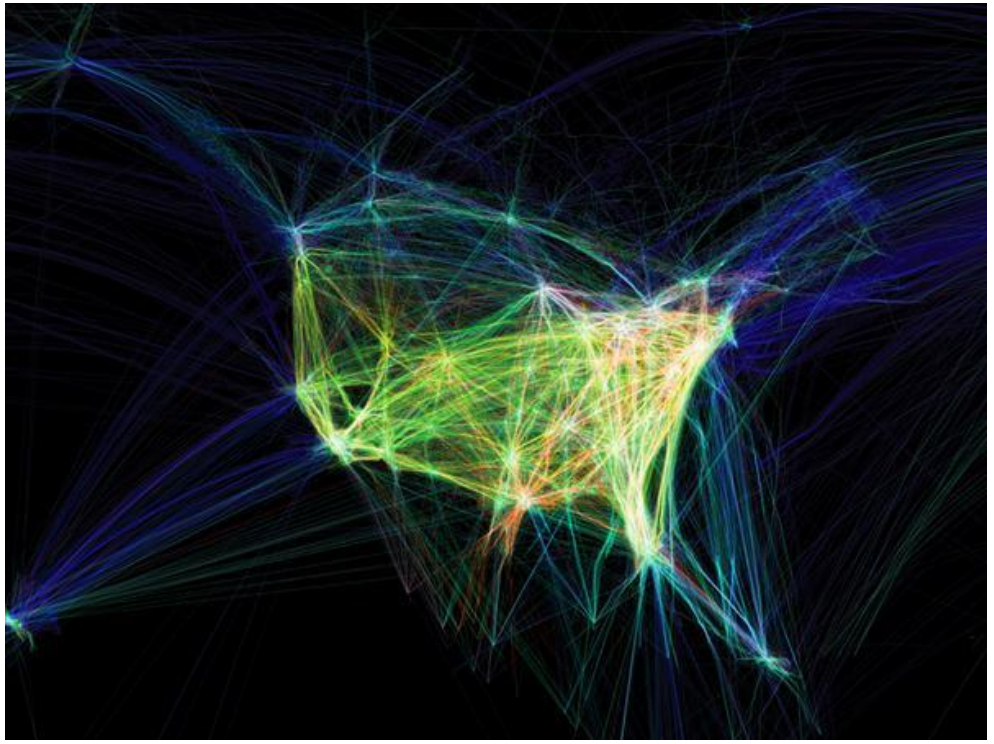
Who are you?

- Turn to your neighbor and discuss:
  - Your names
  - Your major/degree program
  - Why you are interested in taking this course
  - A hobby

This course is about

- How to generate visual representations of information (data) to communicate a message to a viewer
- Focus is on:
  - Tools for visualization generation
  - Underlying principles that make for a good visualization
  - Hands-on practice with visualization generation

What is this?



Aaron Koblin: Flight Patterns

This course is NOT about

- Data science/mining
- High-level seminar/overview (you will need to program)
- Web development
- Data processing/statistics (R, excel, etc.)

You can/should

- Interrupt me anytime with questions
- Ask technical/engineering questions to me/TA
- Discuss lecture content with peers
- Version control your code
- Request extension/absence with a reason

You CANNOT

- Collaborate with peers on assignments
  - Work that you submit must be your own

- Copy code/text from online/others/any resource – we will perform plagiarism check:
  - Acknowledge if you got help from somewhere/someone
  - Code and text you write must be your own

- **F** if you do any of these

Syllabus

Course Objectives

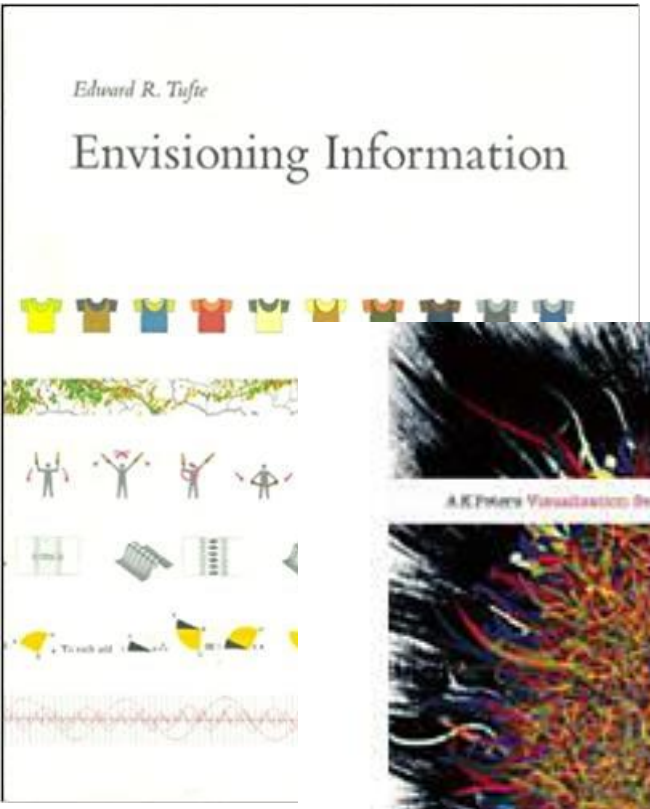
- How and why visualizations work
  - Visual perception system
  - Effective encoding channels for types of data
  - What kind of problems visualizations can solve
- How to create effective visualizations
  - Tools and programming fundamentals
  - Identify which aspects of a data set can/should be visualized
- Research skills
  - Reading and comprehending state-of-the-art research literature
  - Presenting your work in a clear, effective manner
- Hands-on experience
  - Assignments and final project

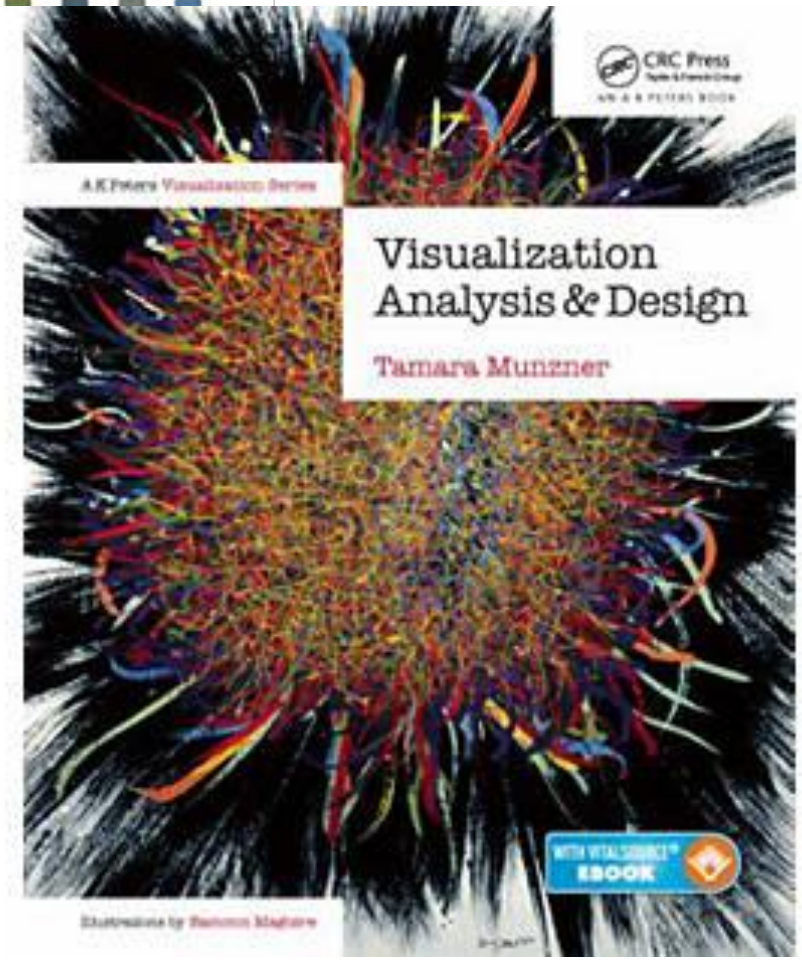
Prerequisites

- Programming
  - Python (matplotlib and Bokeh)
  - Javascript (use at your own risk!)
  - Data structures and algorithms

- Academic writing
  - LaTeX

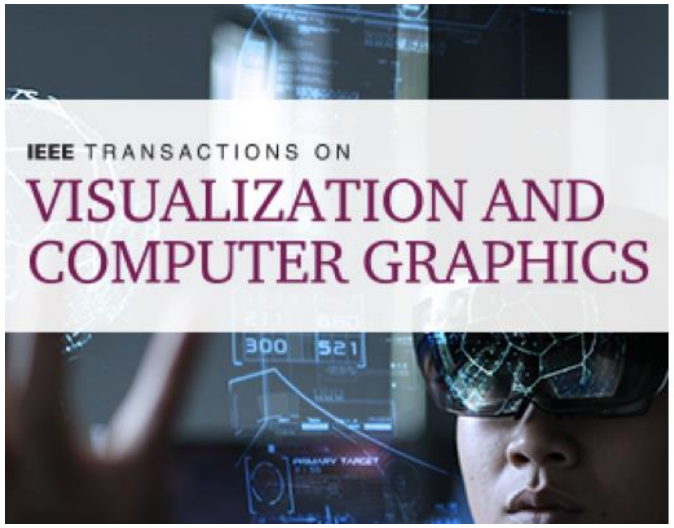
Materials





Google Scholar

VIS



bokeh

matplotlib

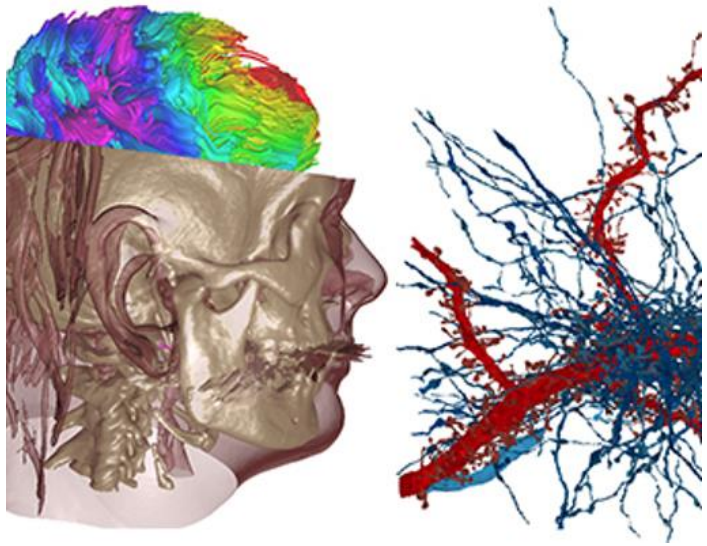
Assignments and grading

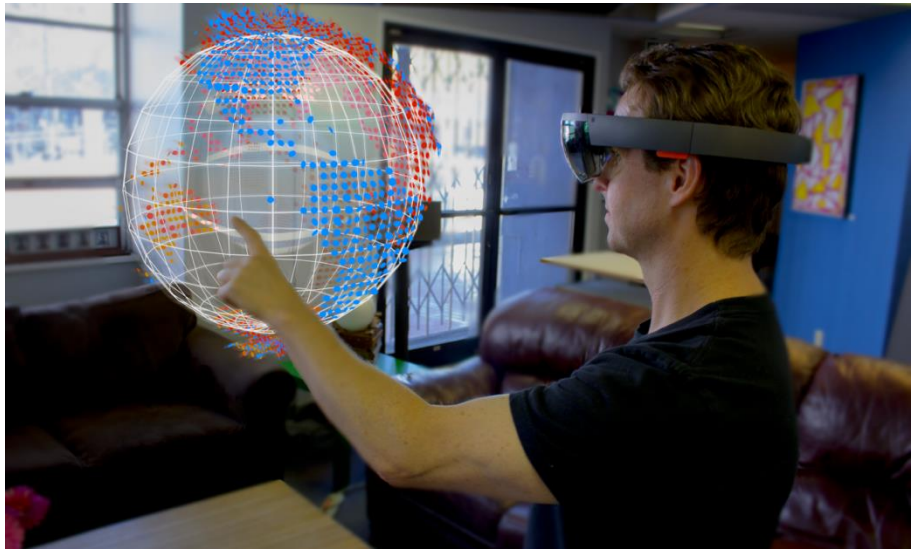
- 15% x 4 assignments (mini projects)
  - Work alone
  - Create visualization(s) and short written report
  - Someone presents their work at the start of a lecture (<5 mins)
  - Late policy:
    - 3 late credits that allow you to extend an assignment deadline by 24 hours per credit
    - A linear late penalty will be applied to assignments, up to 3 days
    - For example, if your assignment is 12 hours late, you will receive a 16.66% penalty (12/72 = 16.66)
    - Late assignments due to illness or unexpected events can be excused with doctor's notes or other forms of written indication
    - **Cannot be applied to the final project!**
- 10% Survey article + project proposal
  - Project topic of your choice
  - Work in groups of up to 3 people
- 30% Final project
  - Visualization(s), written report, and final presentation

Course Content

What is data visualization?

- Graphical representation of data or concepts
  - The point is to communicate a message





Visualization components

- Perceptually-grounded theories of color and graphics
- Scientific visualization
- Information visualization
- Interfaces

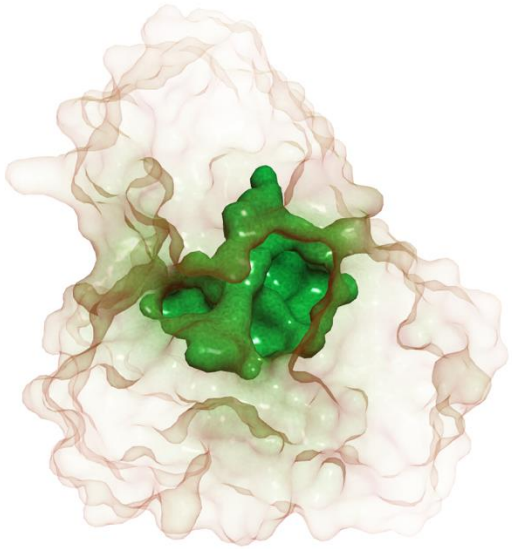
Scientific vs Information Visualization

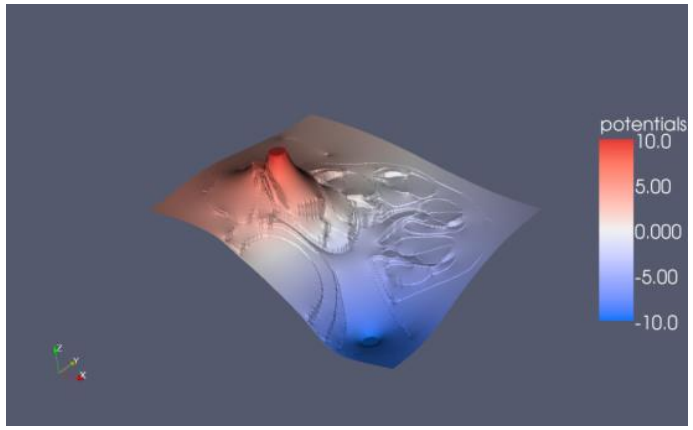
Scientific Visualization

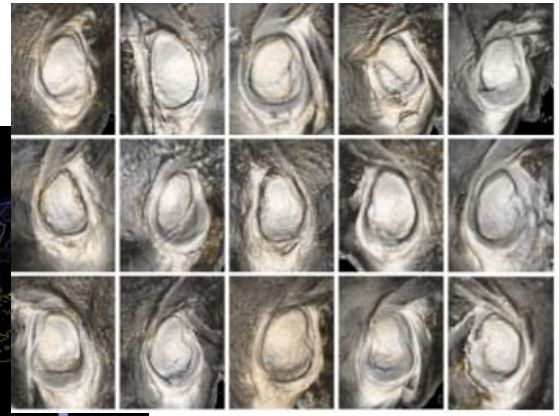
Information Visualization

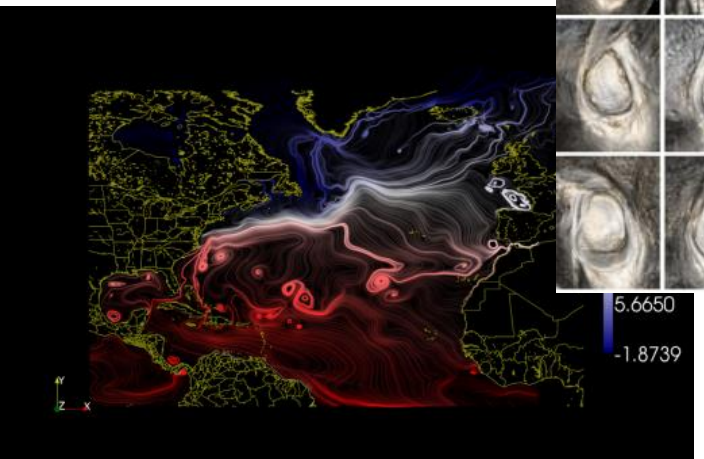
Scientific vs Information Visualization

Scientific Visualization

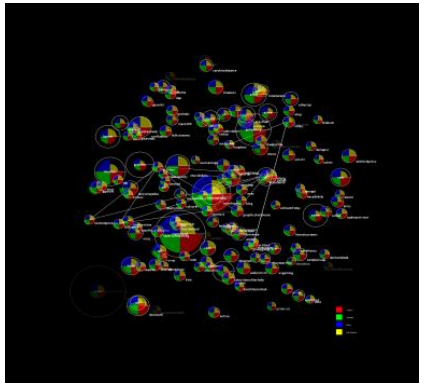




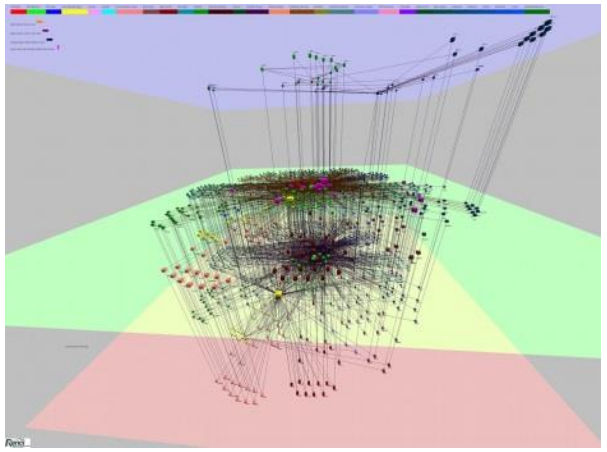




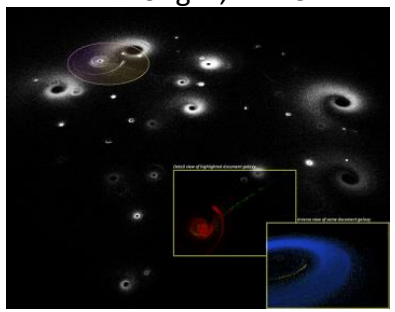
Information Visualization



Jeff Heard, RENCI

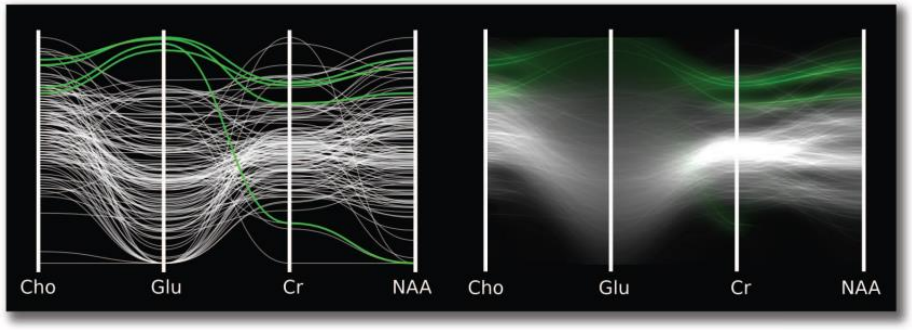


Hong Yi, RENCI



Jeff Heard, RENCI

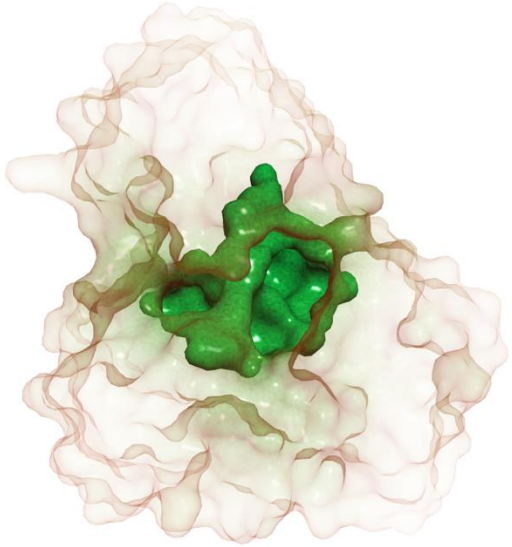
David Feng, UNC

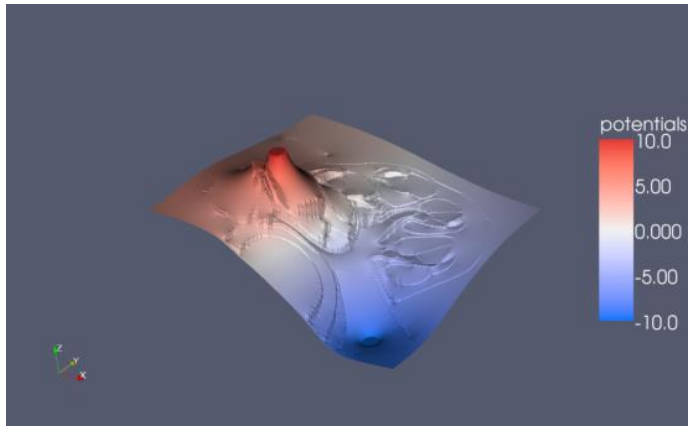


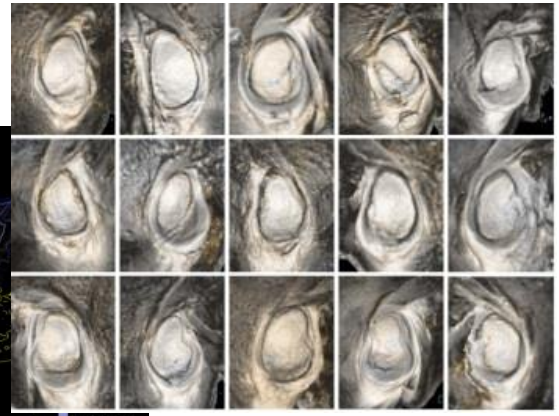
Scientific vs Information Visualization

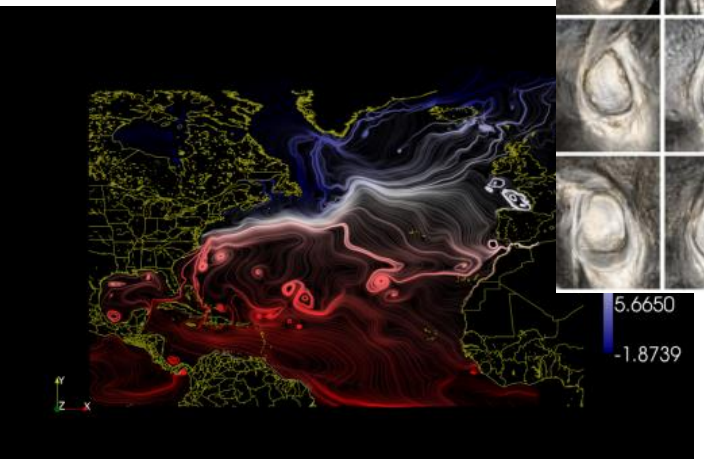
Scientific Visualization

- Spatially-embedded data



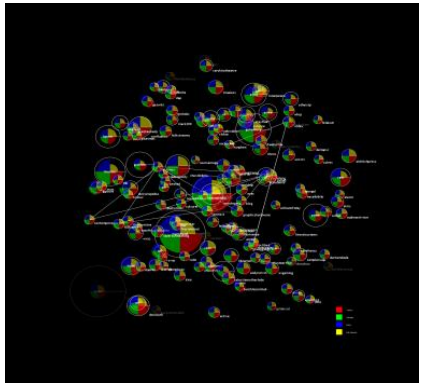




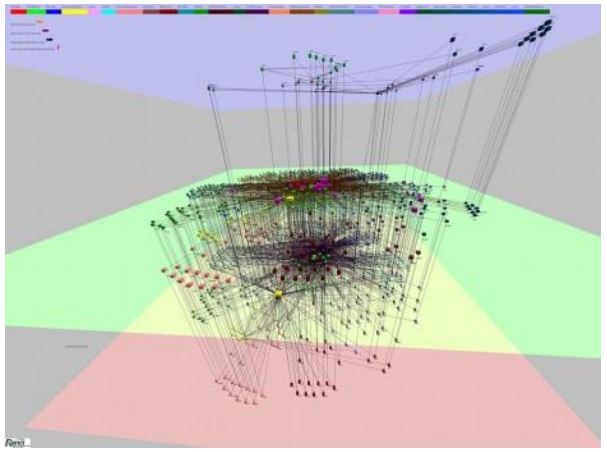


Information Visualization

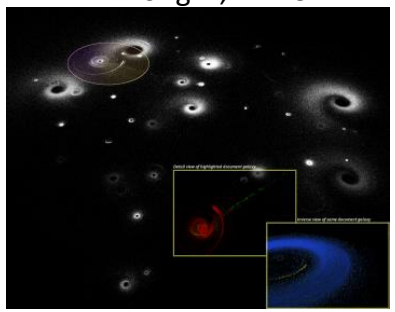
- Abstract data



Jeff Heard, RENCI

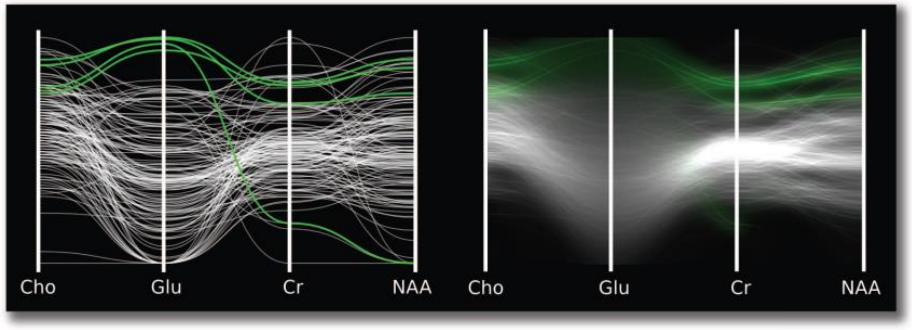


Hong Yi, RENCI



Jeff Heard, RENCI

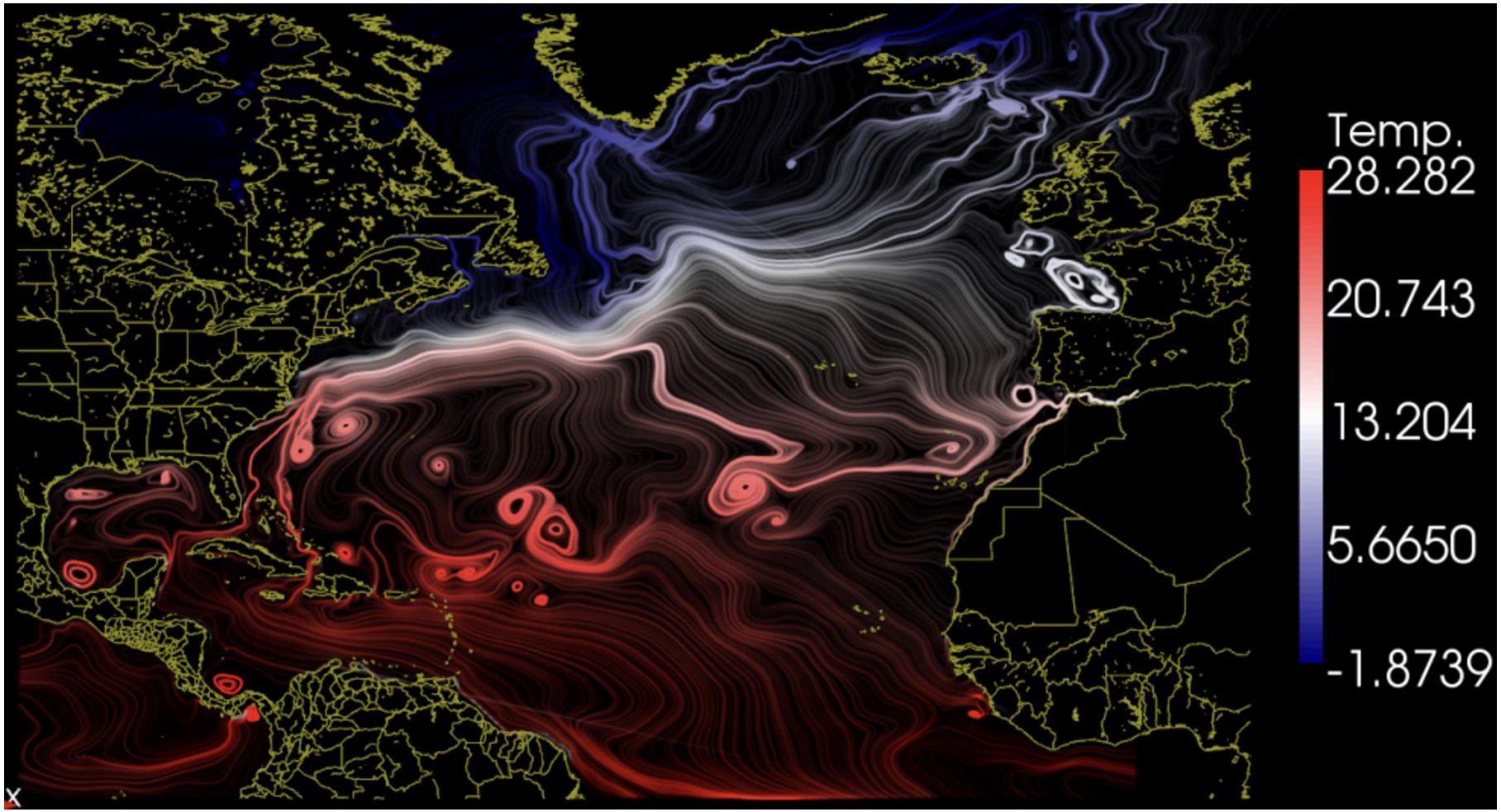
David Feng, UNC



Live “demo”

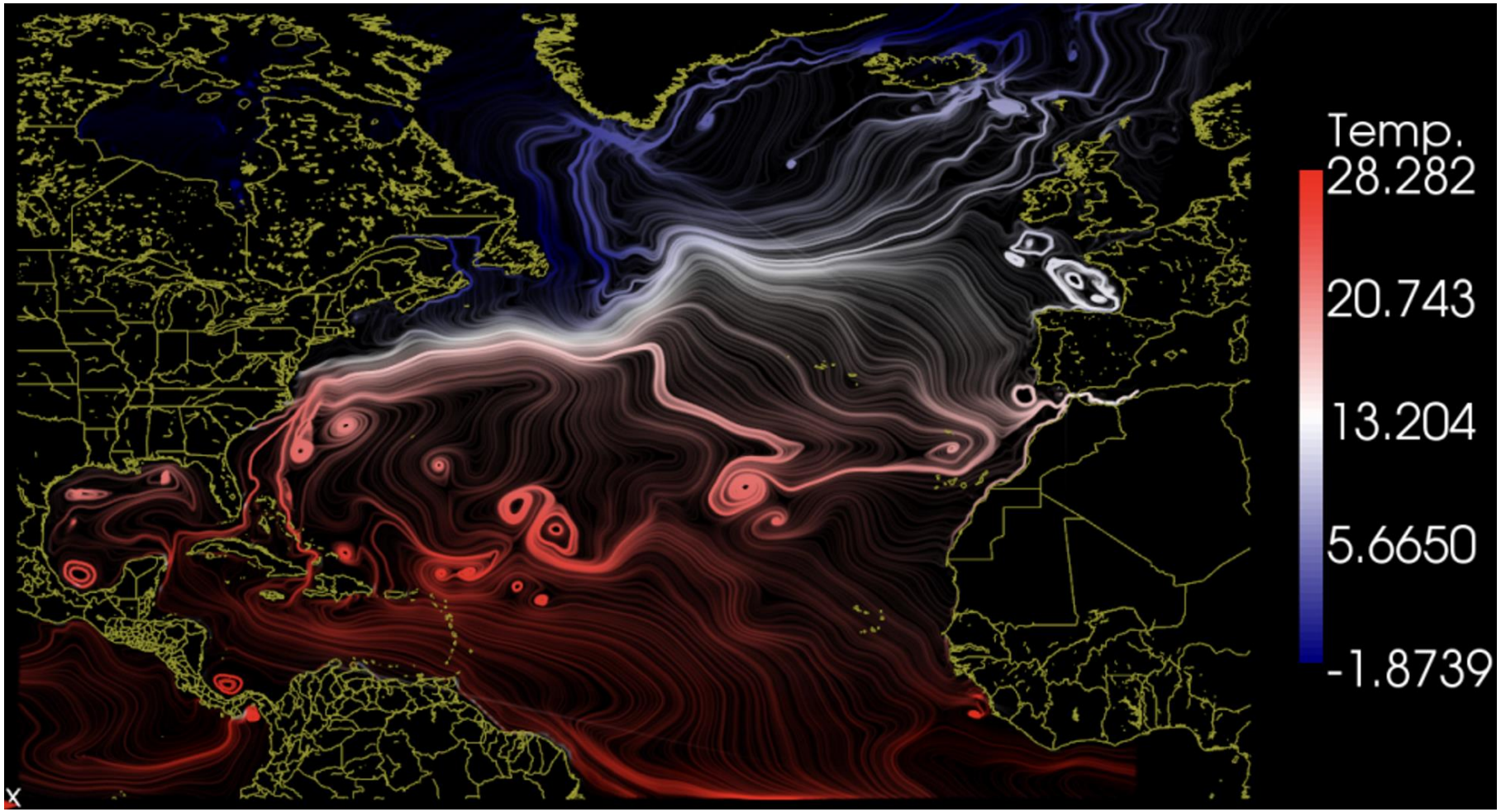
tinyurl.com/bdecrt33

Advantages of data viz



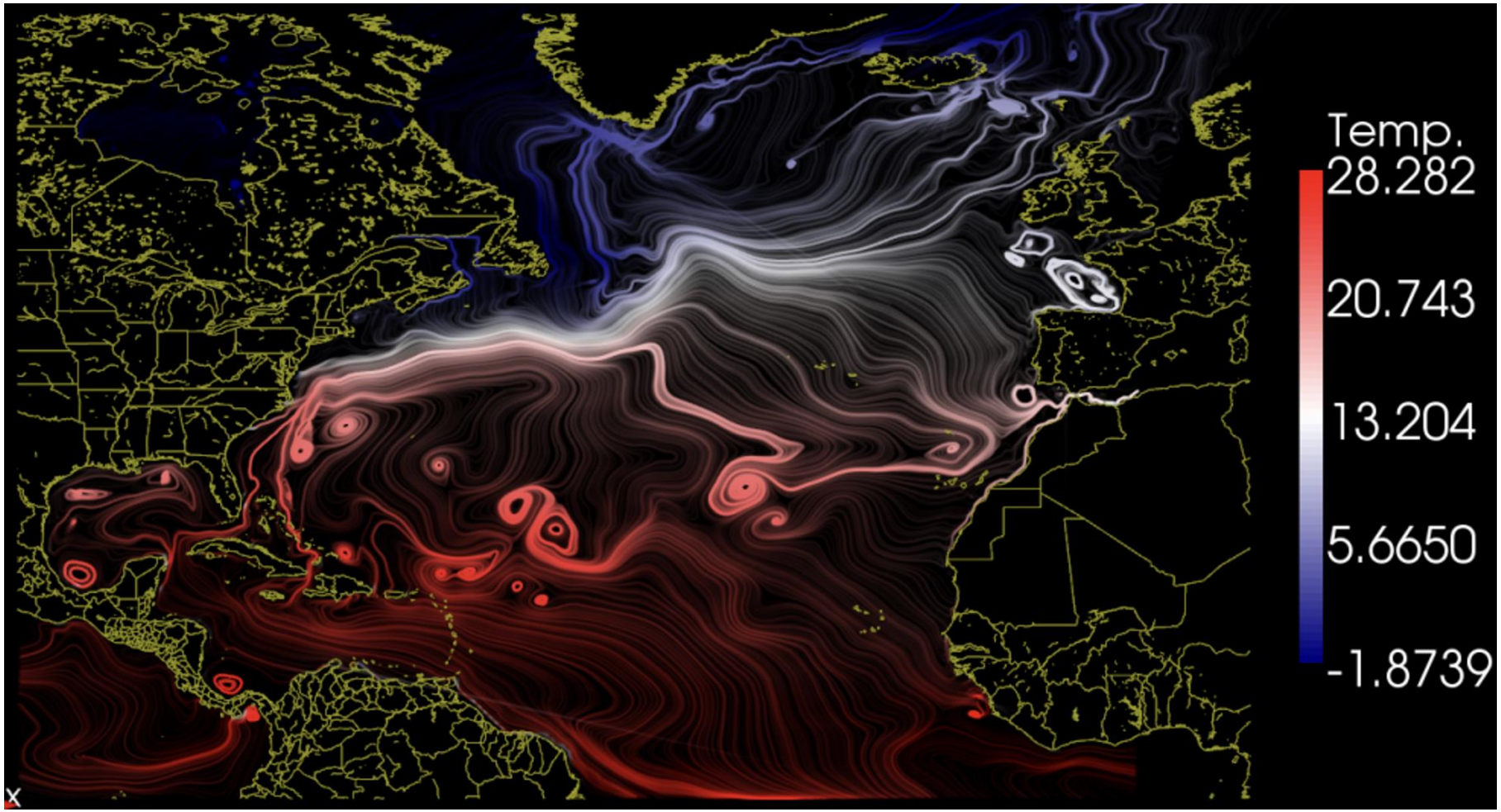
- Comprehend *huge amounts of data*

Advantages of data viz



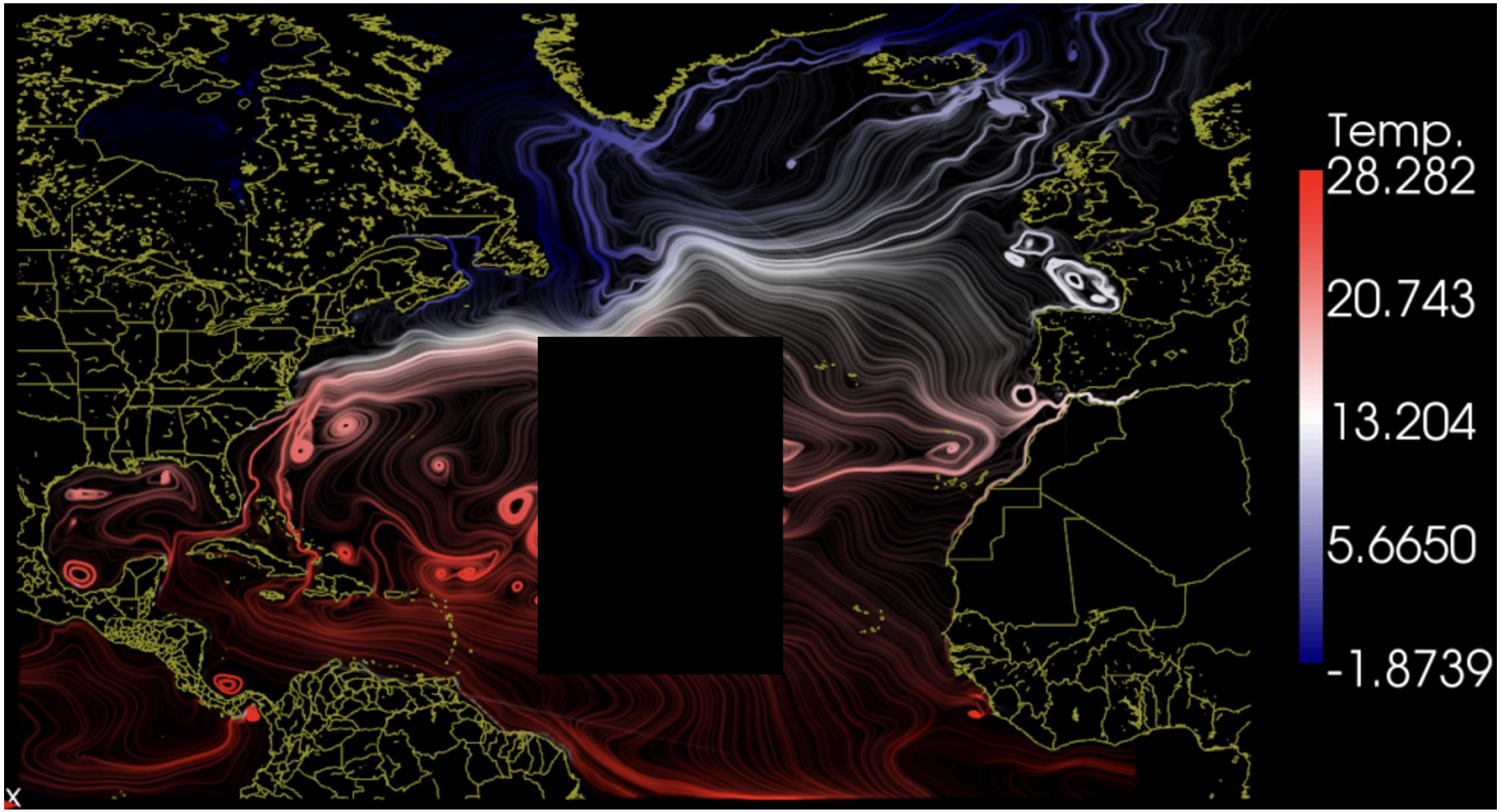
- Facilitates *simultaneous understanding* of both large-scale and small-scale features

Advantages of data viz



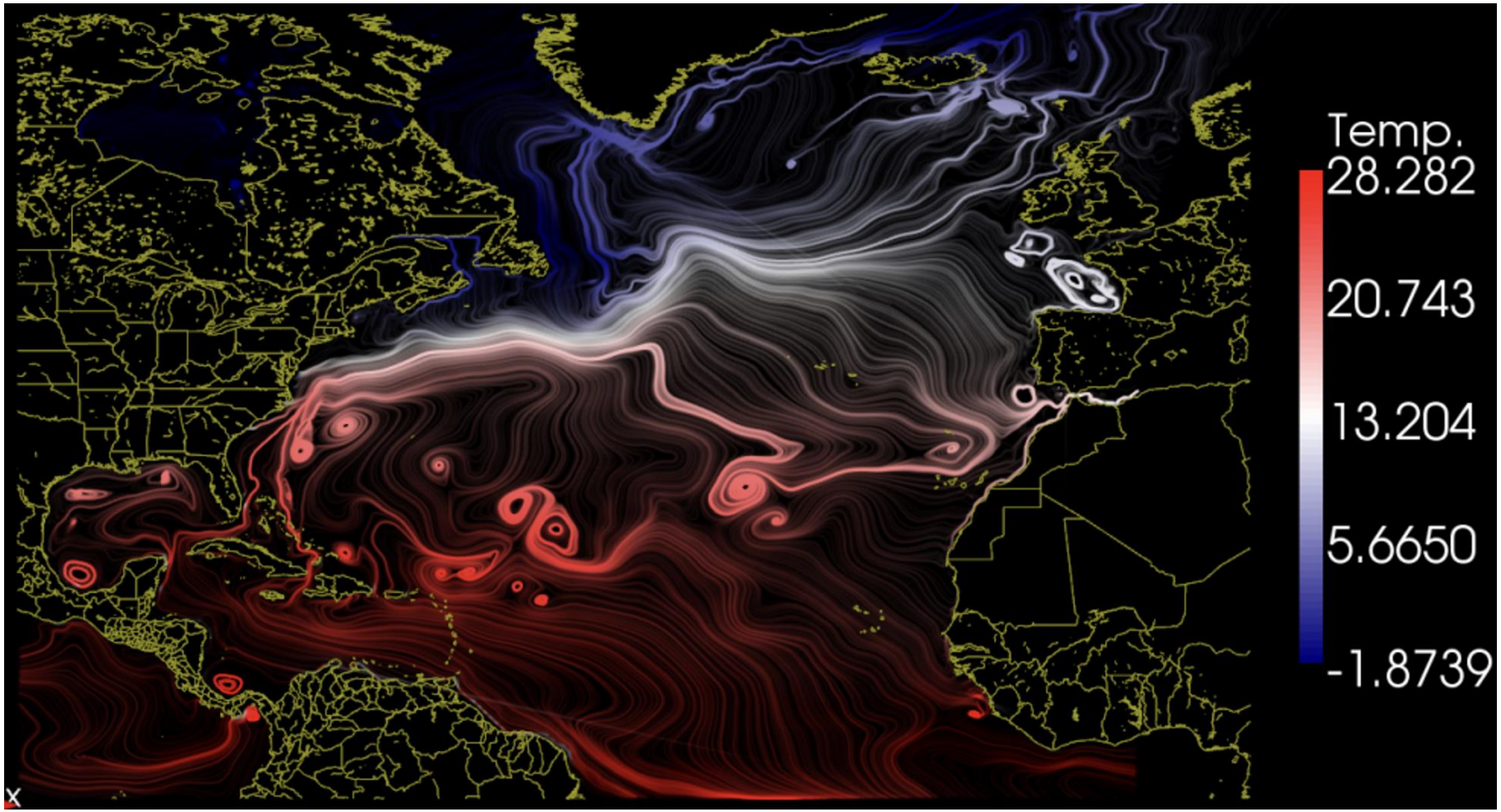
- Enables the *perception of emergent properties* that were not anticipated and *makes problems with the data collection/simulation apparent*

Advantages of data viz



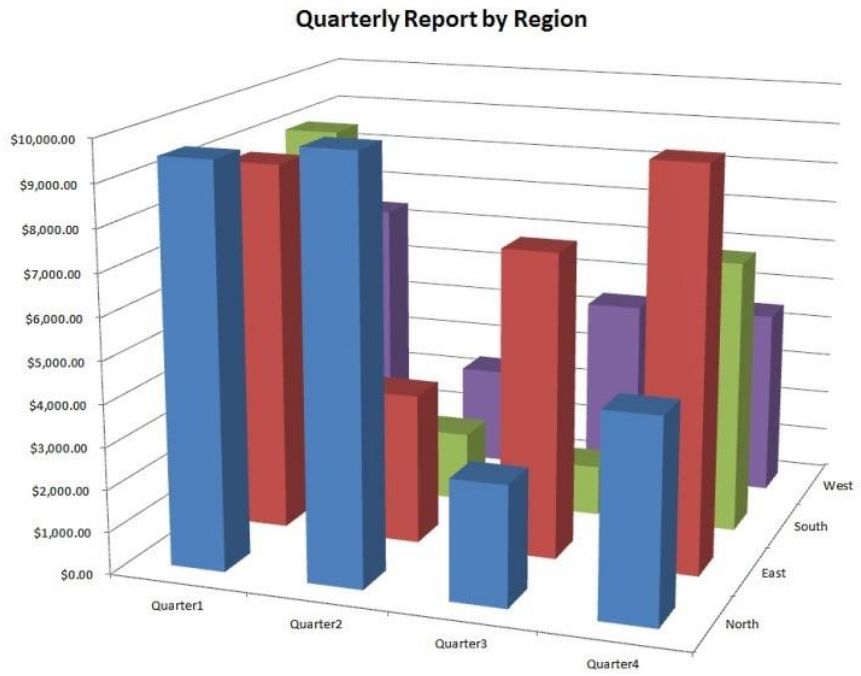
- Enables the *perception of emergent properties* that were not anticipated and *makes problems with the data collection/simulation apparent*

Advantages of data viz

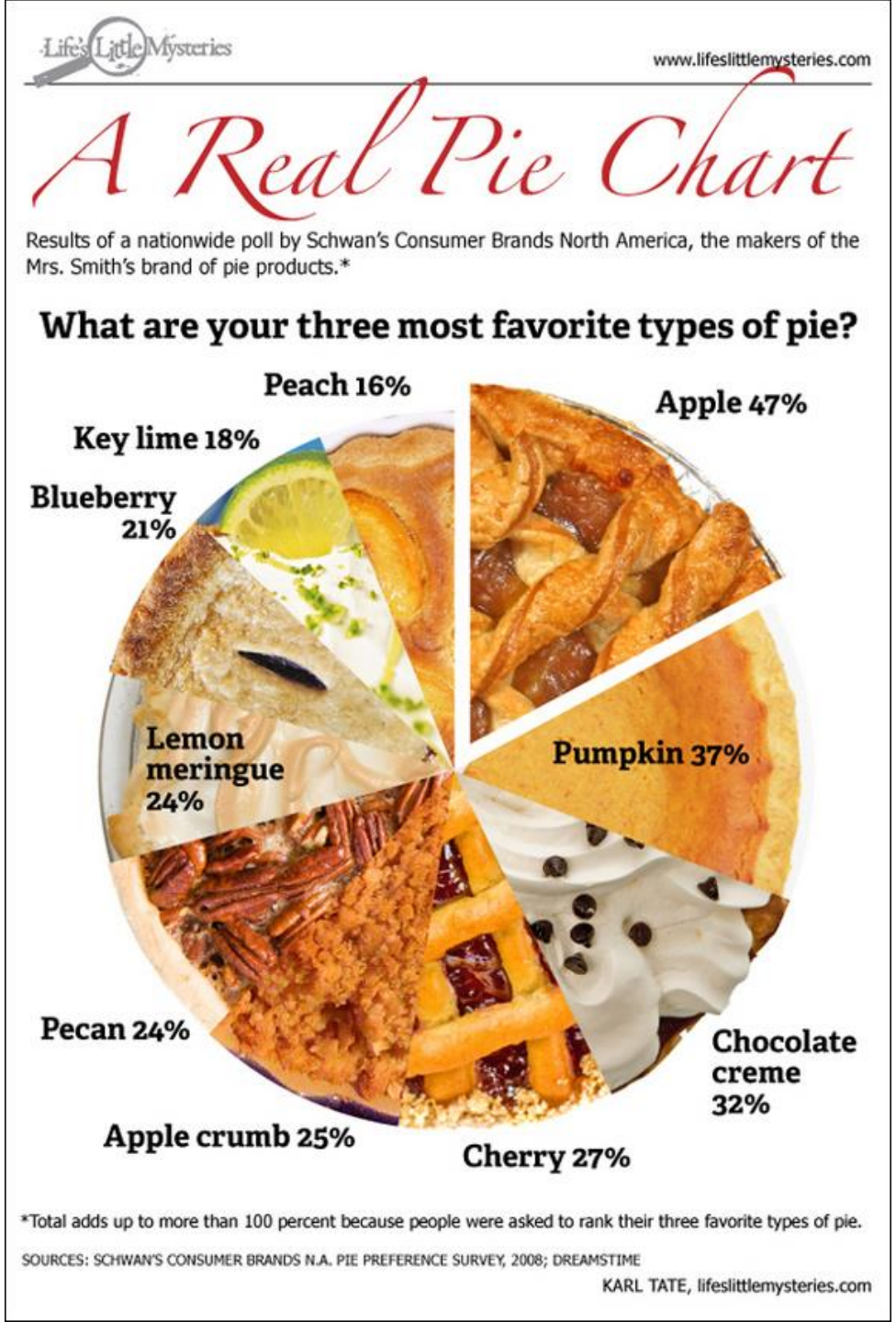


- Facilitates *hypothesis formation*

Very easy to do it poorly



viz.wtf for more



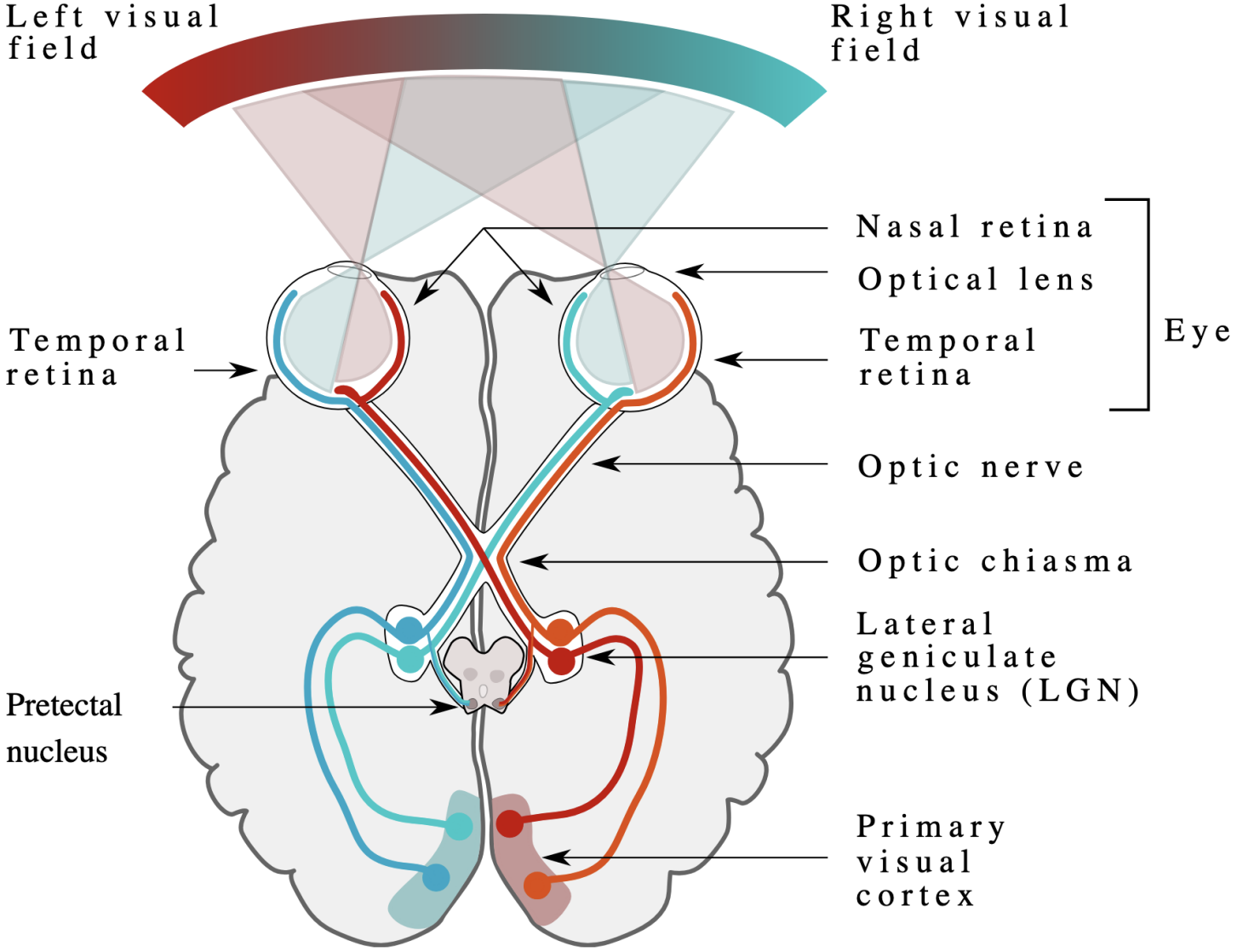
Considerations when creating visualizations

Considerations when creating visualizations

- Main question: How to transform data into something that people can understand for optimal decision making?

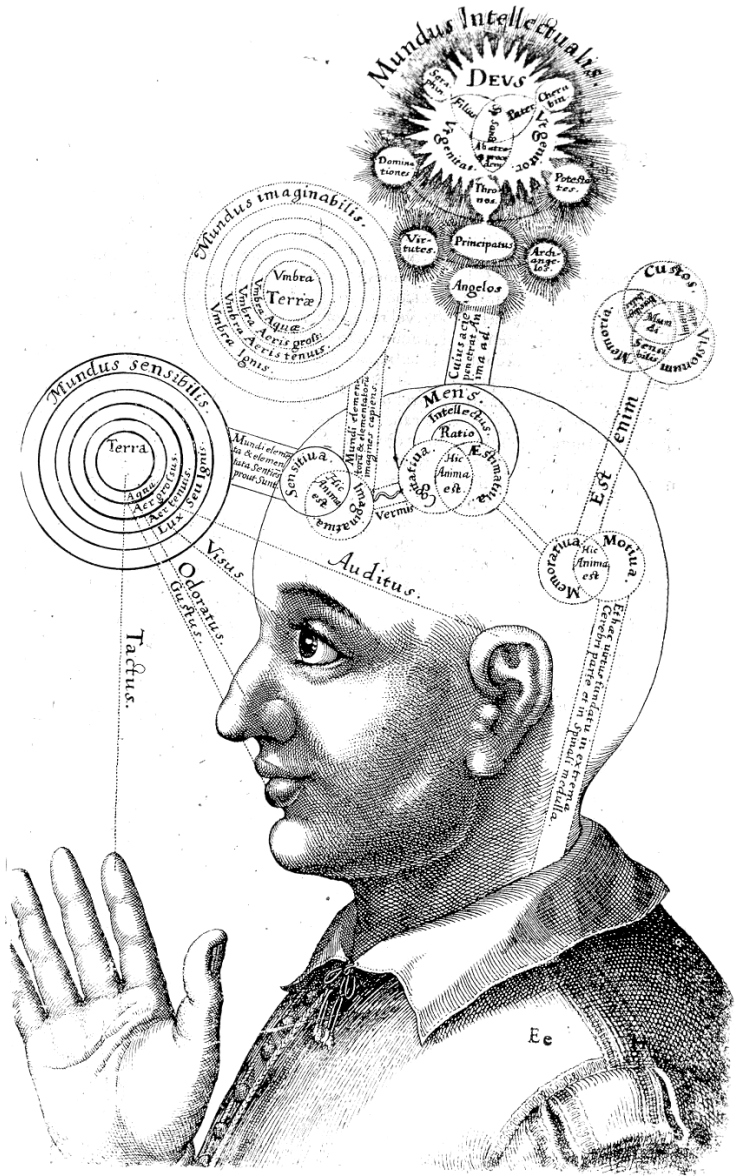
Considerations when creating visualizations

- Main question: How to transform data into something that people can understand for optimal decision making?
  - Human visual system: How people see



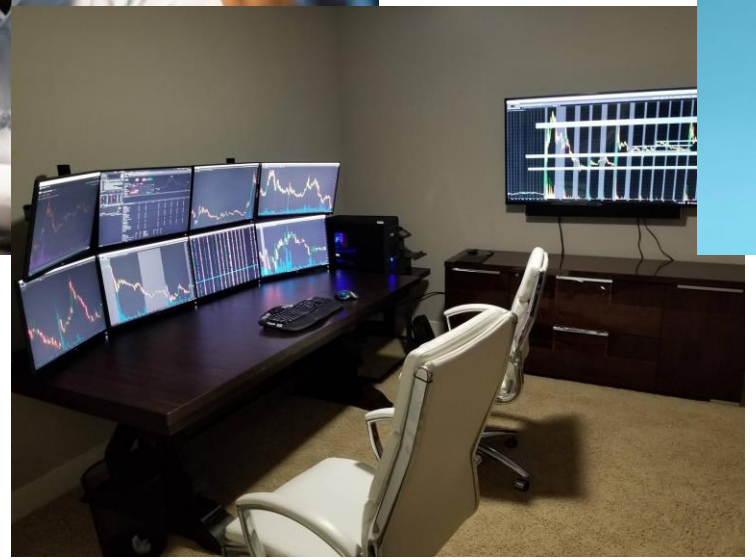
Considerations when creating visualizations

- Main question: How to transform data into something that people can understand for optimal decision making?
  - Human visual system: How people see
  - Cognitive system: How people think



Considerations when creating visualizations

- Main question: How to transform data into something that people can understand for optimal decision making?
  - Human visual system: How people see
  - Cognitive system: How people think
  - Display modality: How the viz will be displayed



Topics we will cover (approximately)

- Week 1: Introduction, syllabus

Topics we will cover (approximately)

- Week 1: Introduction, syllabus
- Week 2: Data types, viz techniques & tools

Python

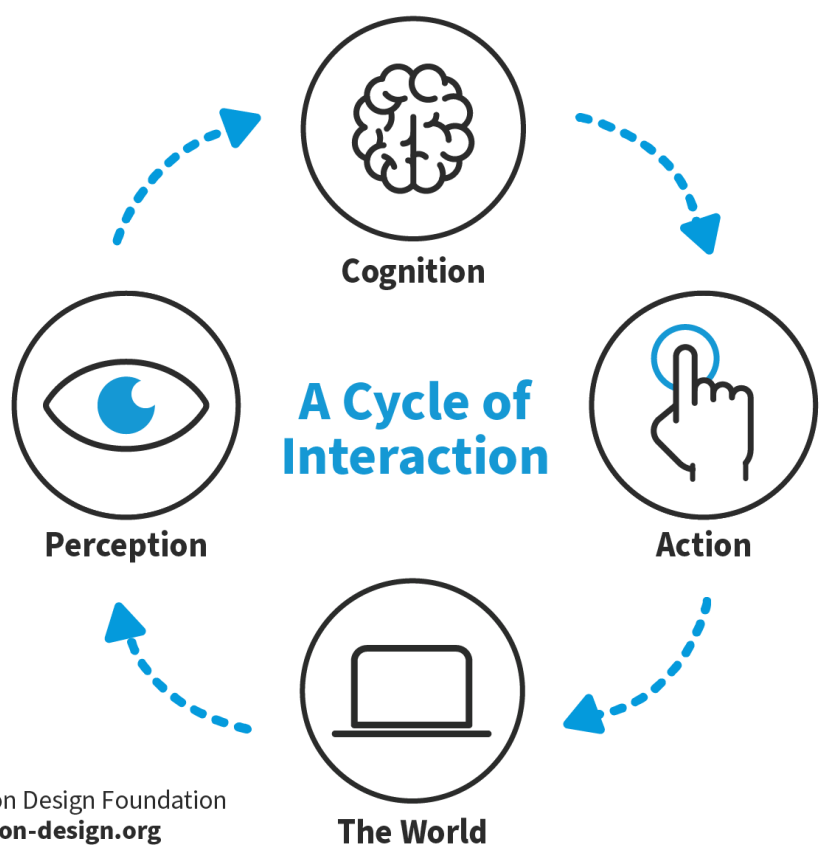
bokeh

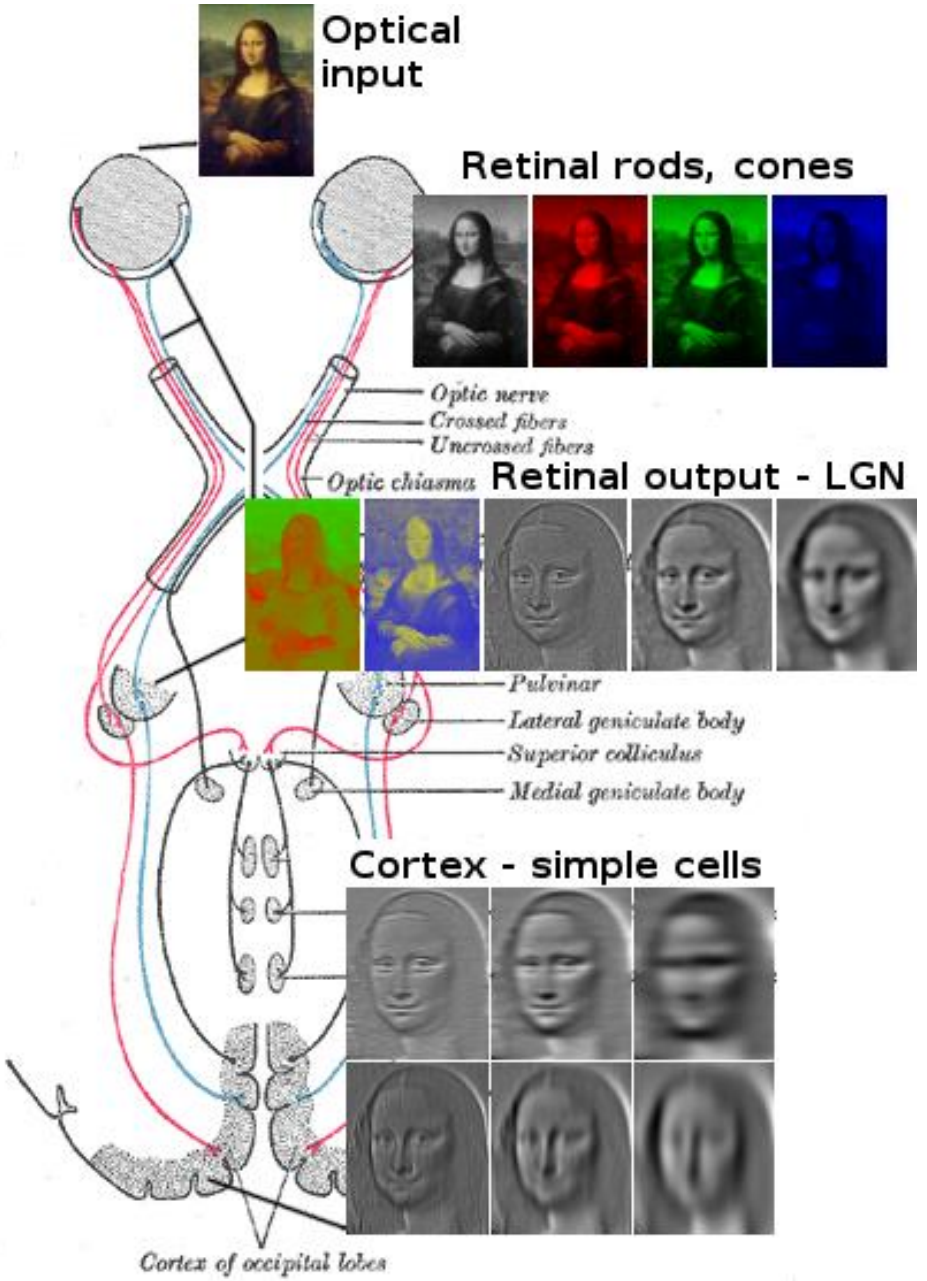
matplotlib

Javascript (use at your own risk)

Topics we will cover (approximately)

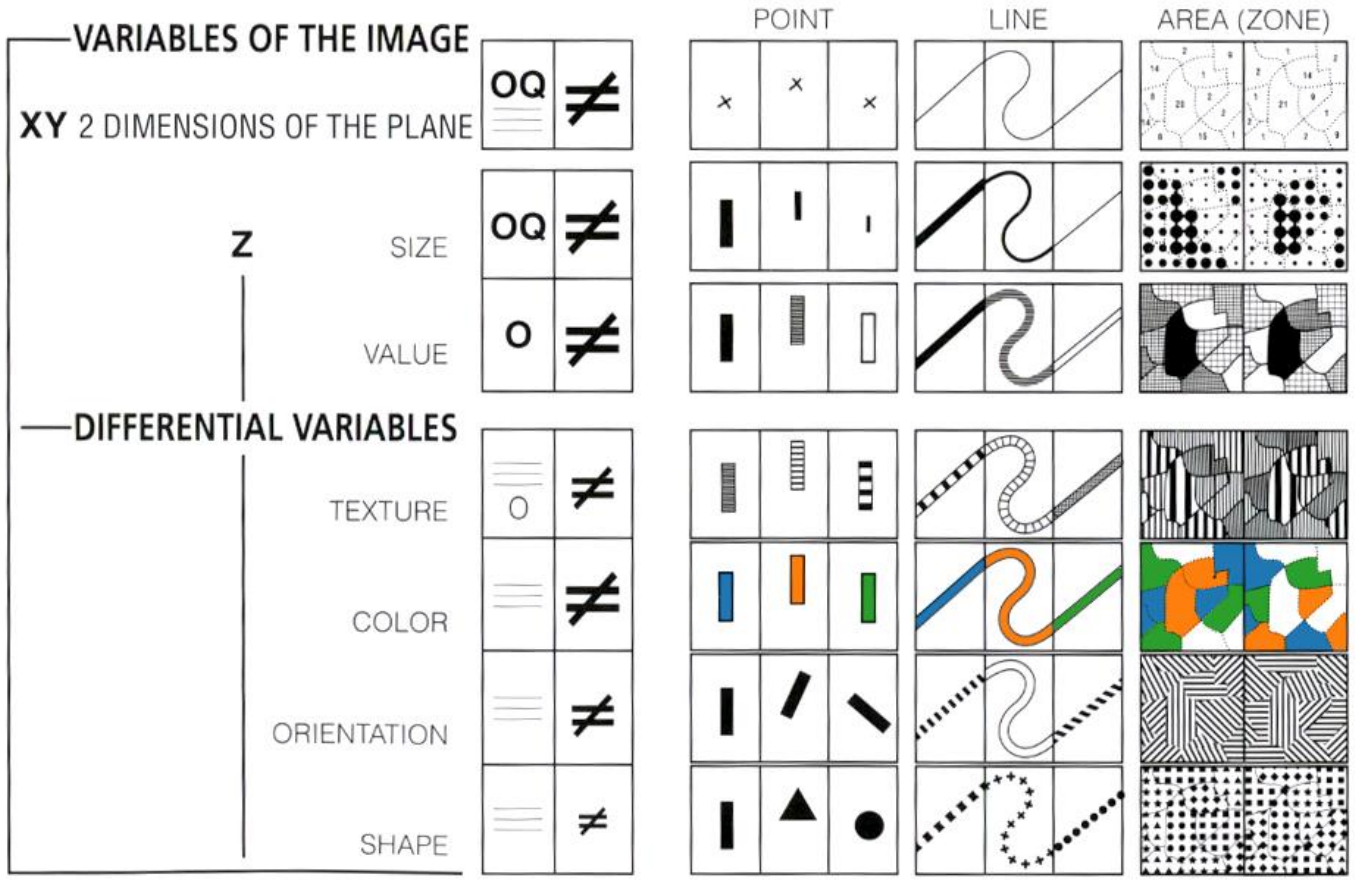
- Week 1: Introduction, syllabus
- Week 2: Data types, viz techniques & tools
- Week 3: The human visual system



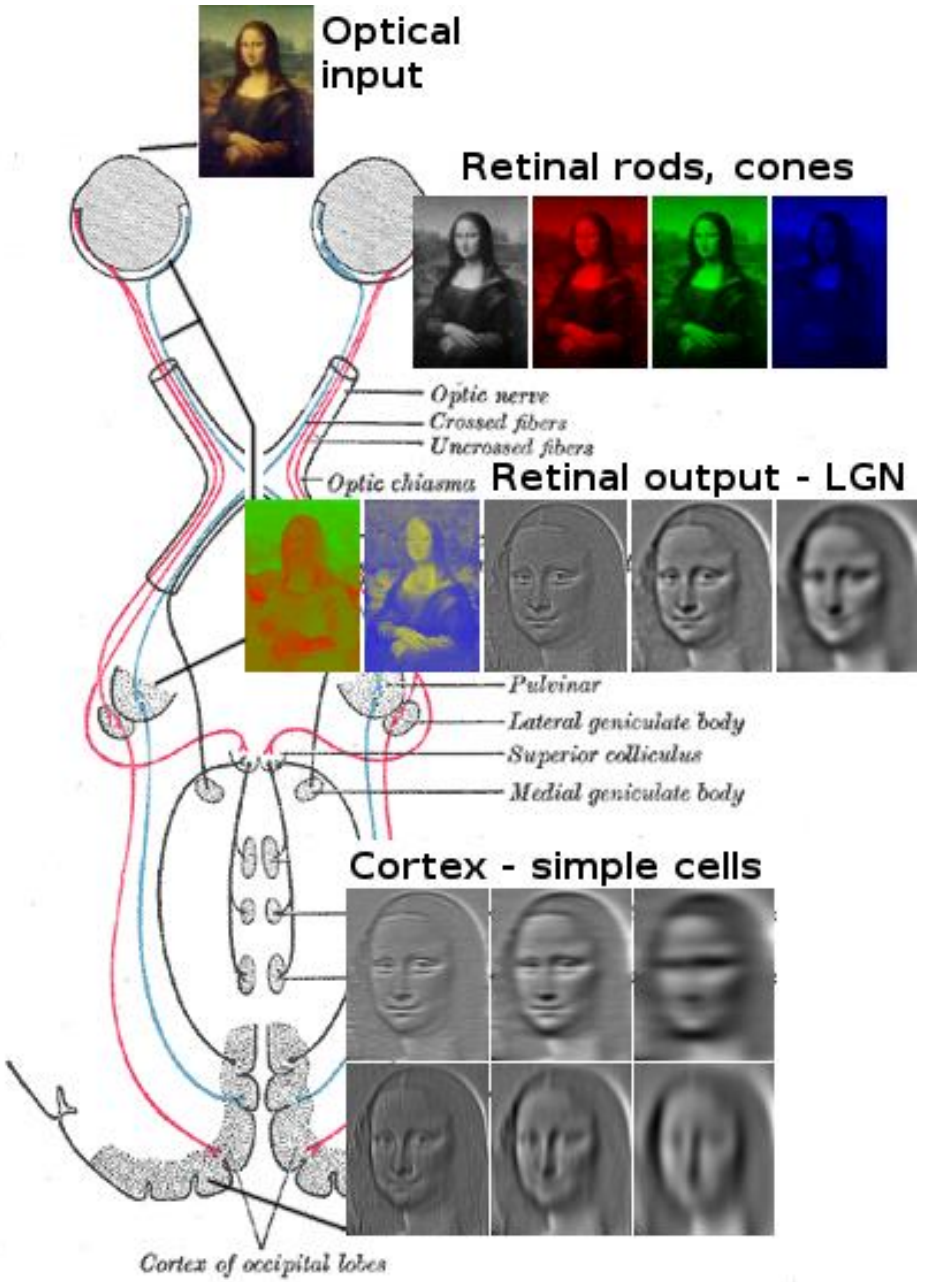


Topics we will cover (approximately)

- Week 1: Introduction, syllabus
- Week 2: Data types, viz techniques & tools
- Week 3: The human visual system
- Week 4: Visual encoding + perception

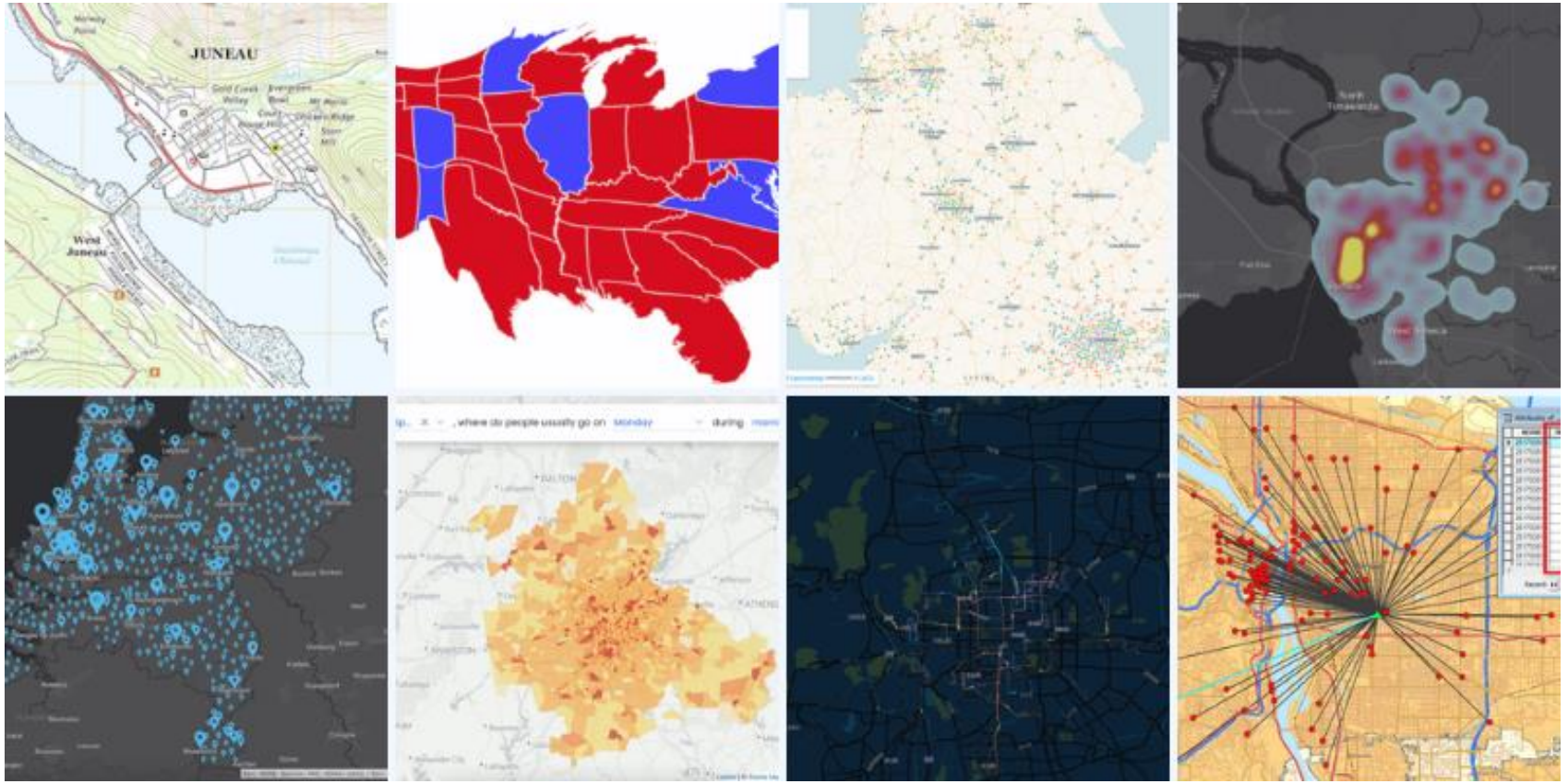


https://idl.uw.edu/visualization-curriculum/intro.html



Topics we will cover (approximately)

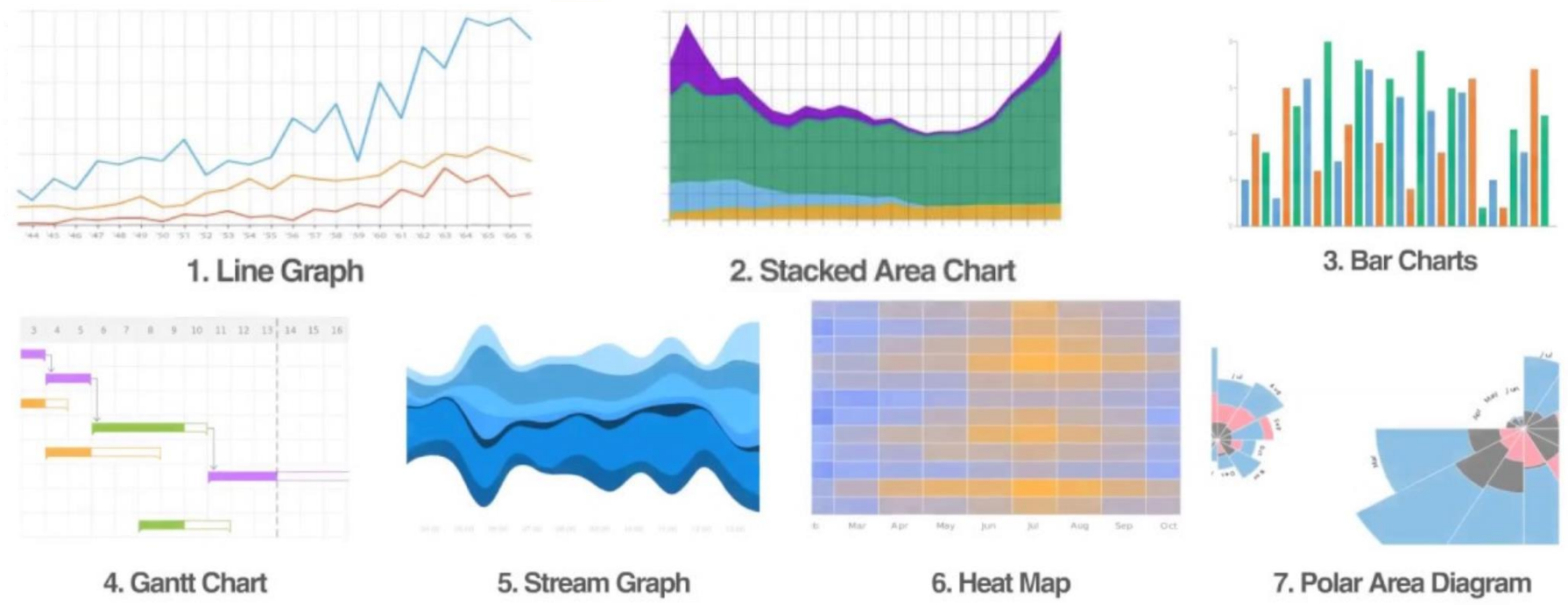
- Week 1: Introduction, syllabus
- Week 2: Data types, viz techniques & tools
- Week 3: The human visual system
- Week 4: Visual encoding + perception
- Week 5: 2D viz - spatial data



https://www.safegraph.com/guides/visualizing-geospatial-data

Topics we will cover (approximately)

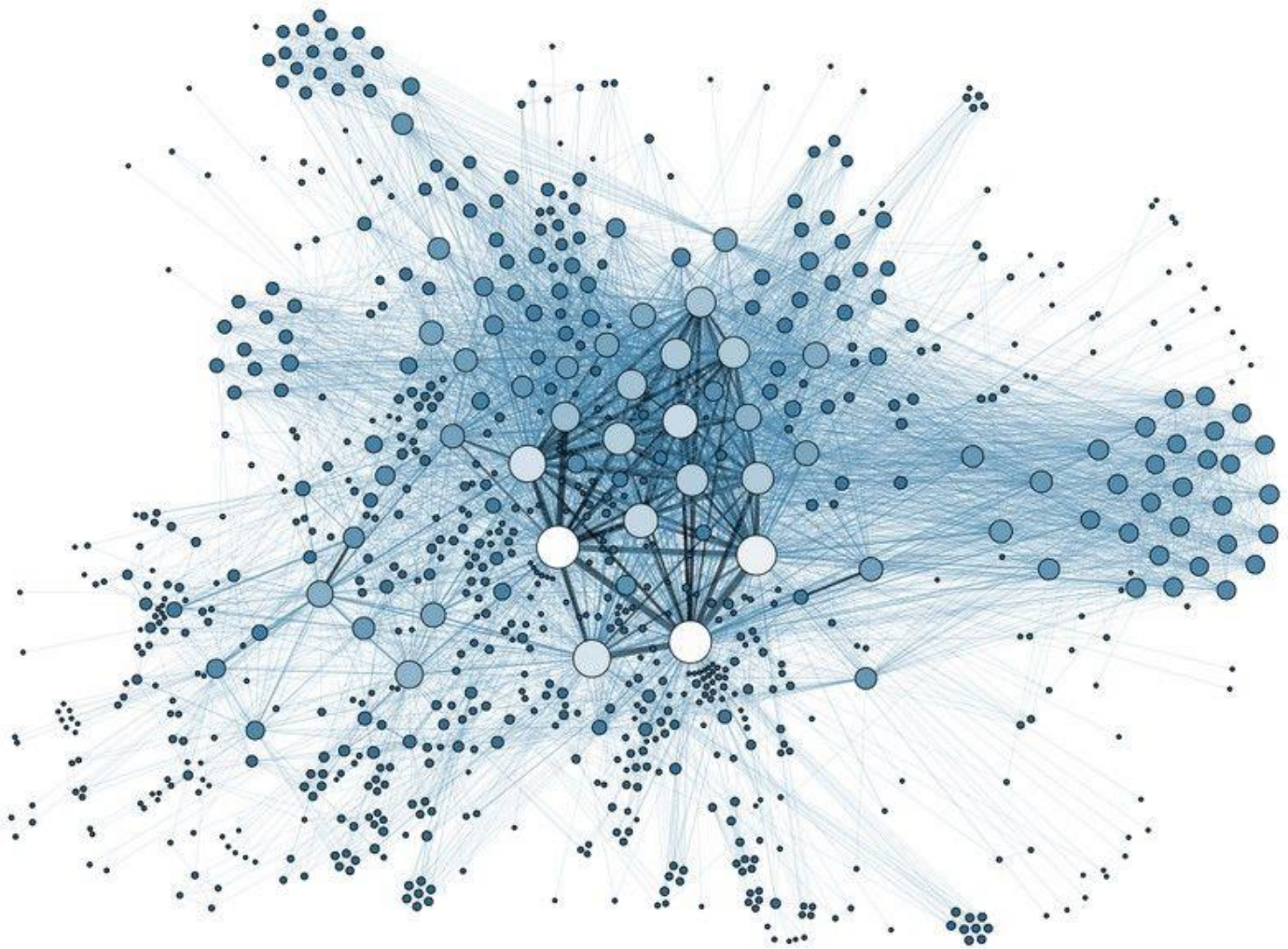
- Week 1: Introduction, syllabus
- Week 2: Data types, viz techniques & tools
- Week 3: The human visual system
- Week 4: Visual encoding + perception
- Week 5: 2D viz - spatial data
- Week 6: 2D viz - temporal data



https://humansofdata.atlan.com/2016/11/visualizing-time-series-data/

Topics we will cover (approximately)

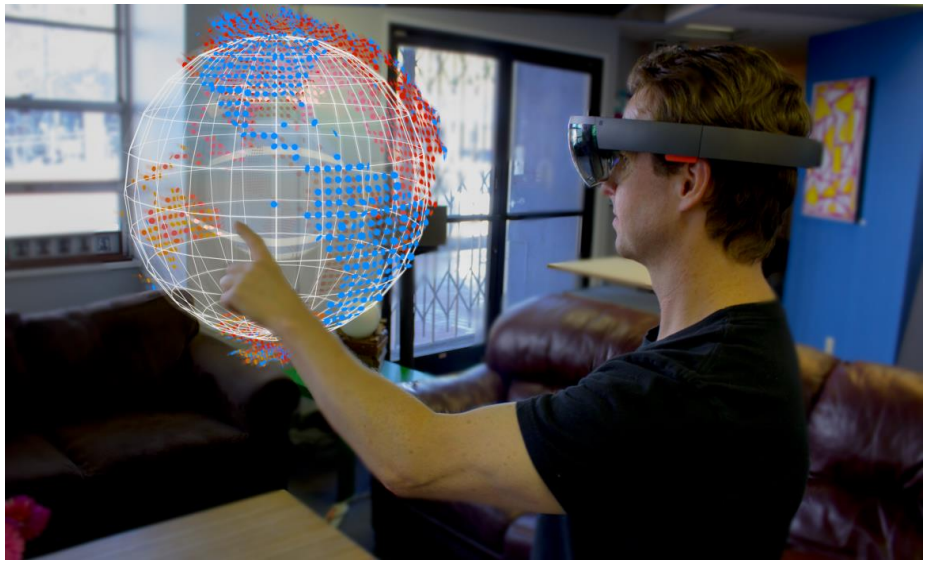
- Week 1: Introduction, syllabus
- Week 2: Data types, viz techniques & tools
- Week 3: The human visual system
- Week 4: Visual encoding + perception
- Week 5: 2D viz - spatial data
- Week 6: 2D viz - temporal data
- Week 7: 2D viz - network data

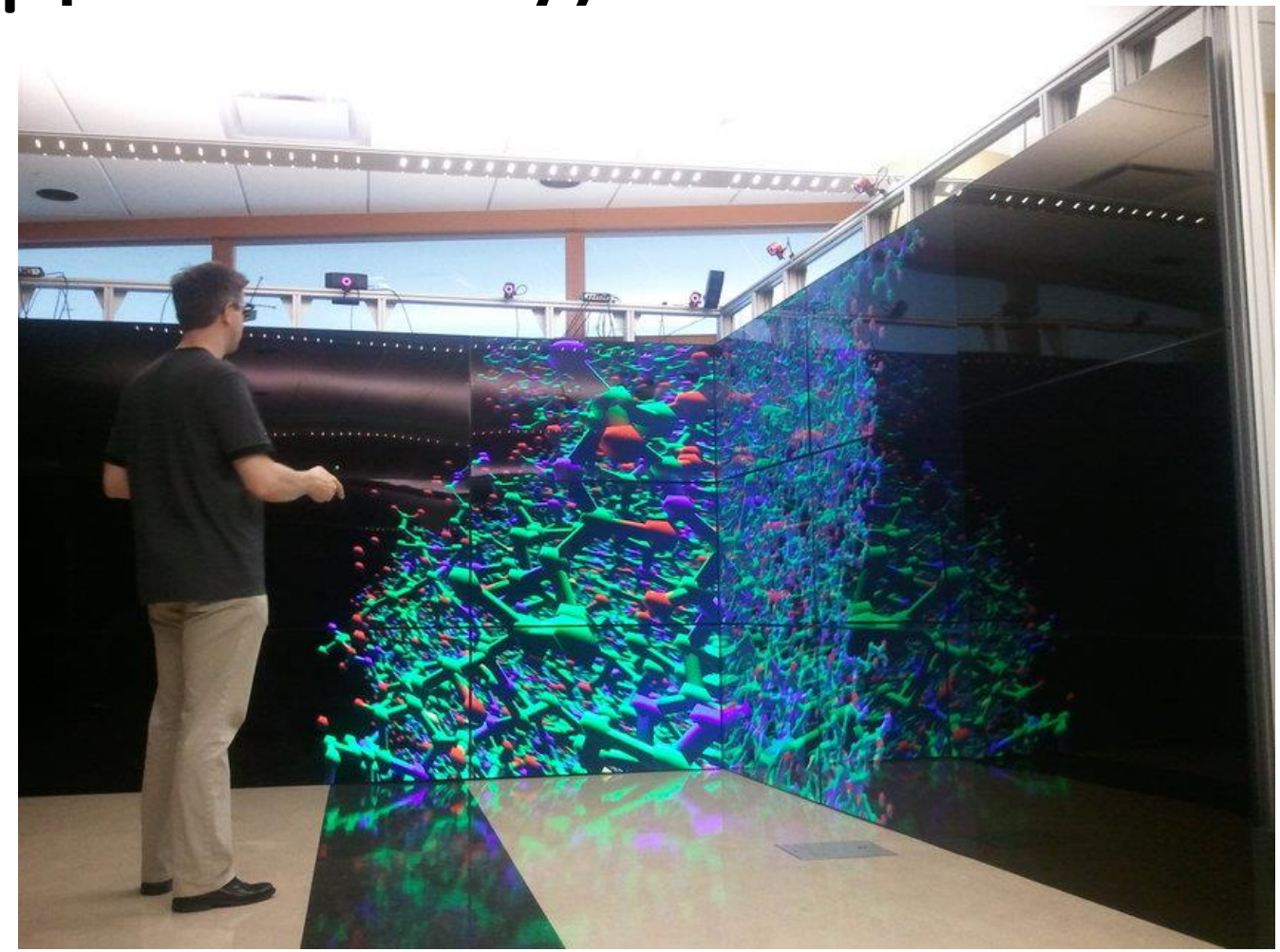


https://www.interaction-design.org/literature/article/how-to-display-complex-network-data-with-information-visualization

Topics we will cover (approximately)

- Week 1: Introduction, syllabus
- Week 2: Data types, viz techniques & tools
- Week 3: The human visual system
- Week 4: Visual encoding + perception
- Week 5: 2D viz - spatial data
- Week 6: 2D viz - temporal data
- Week 7: 2D viz - network data
- Week 8: 3D viz - projection & graphics

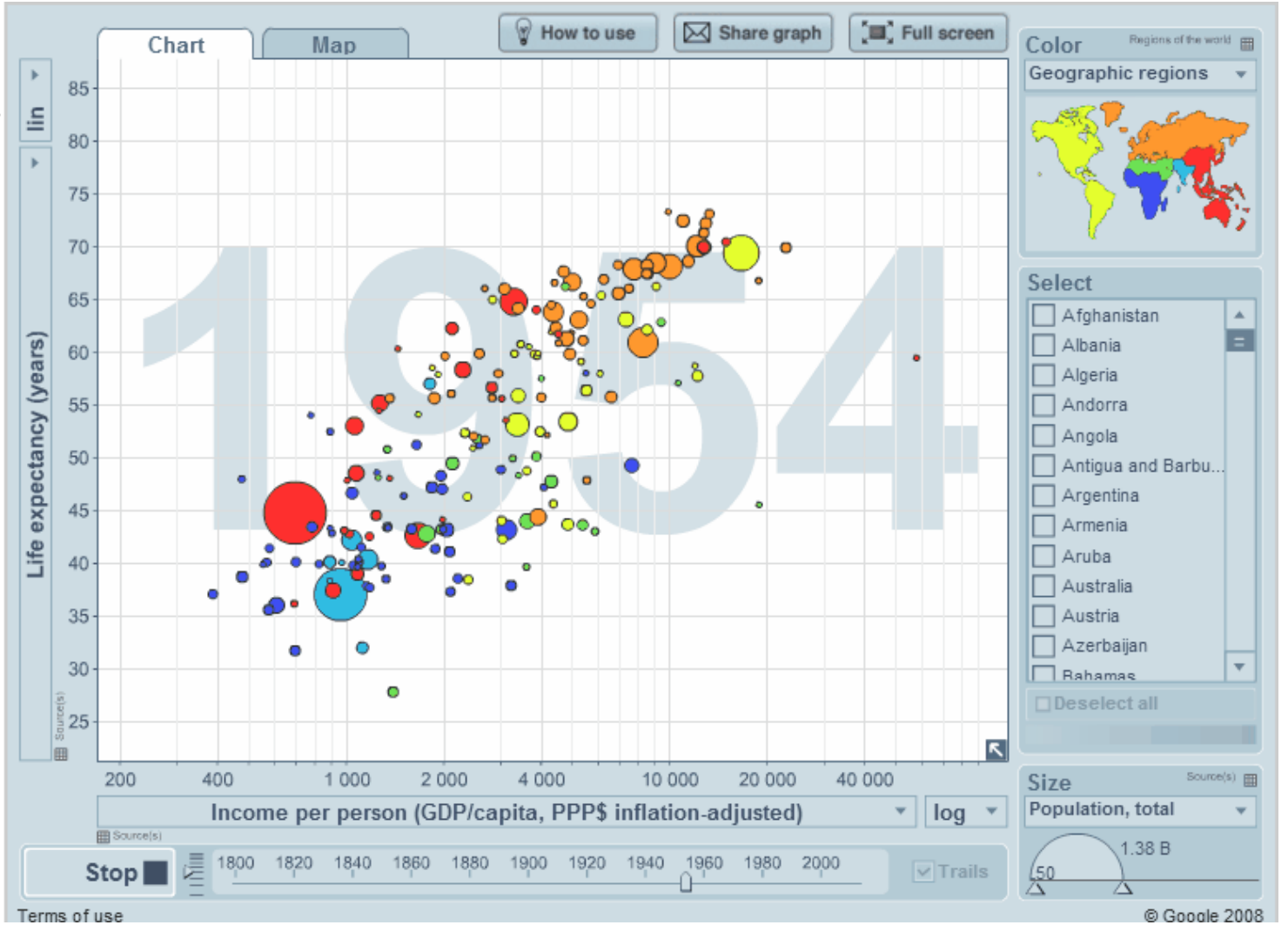




Wischgoll, Thomas, et al. "Display infrastructure for virtual environments." Electronic Imaging 30 (2017): 1-11.

Topics we will cover (approximately)

- Week 1: Introduction, syllabus
- Week 2: Data types, viz techniques & tools
- Week 3: The human visual system
- Week 4: Visual encoding + perception
- Week 5: 2D viz - spatial data
- Week 6: 2D viz - temporal data
- Week 7: 2D viz - network data
- Week 8: 3D viz - projection & graphics
- Week 9: Animation



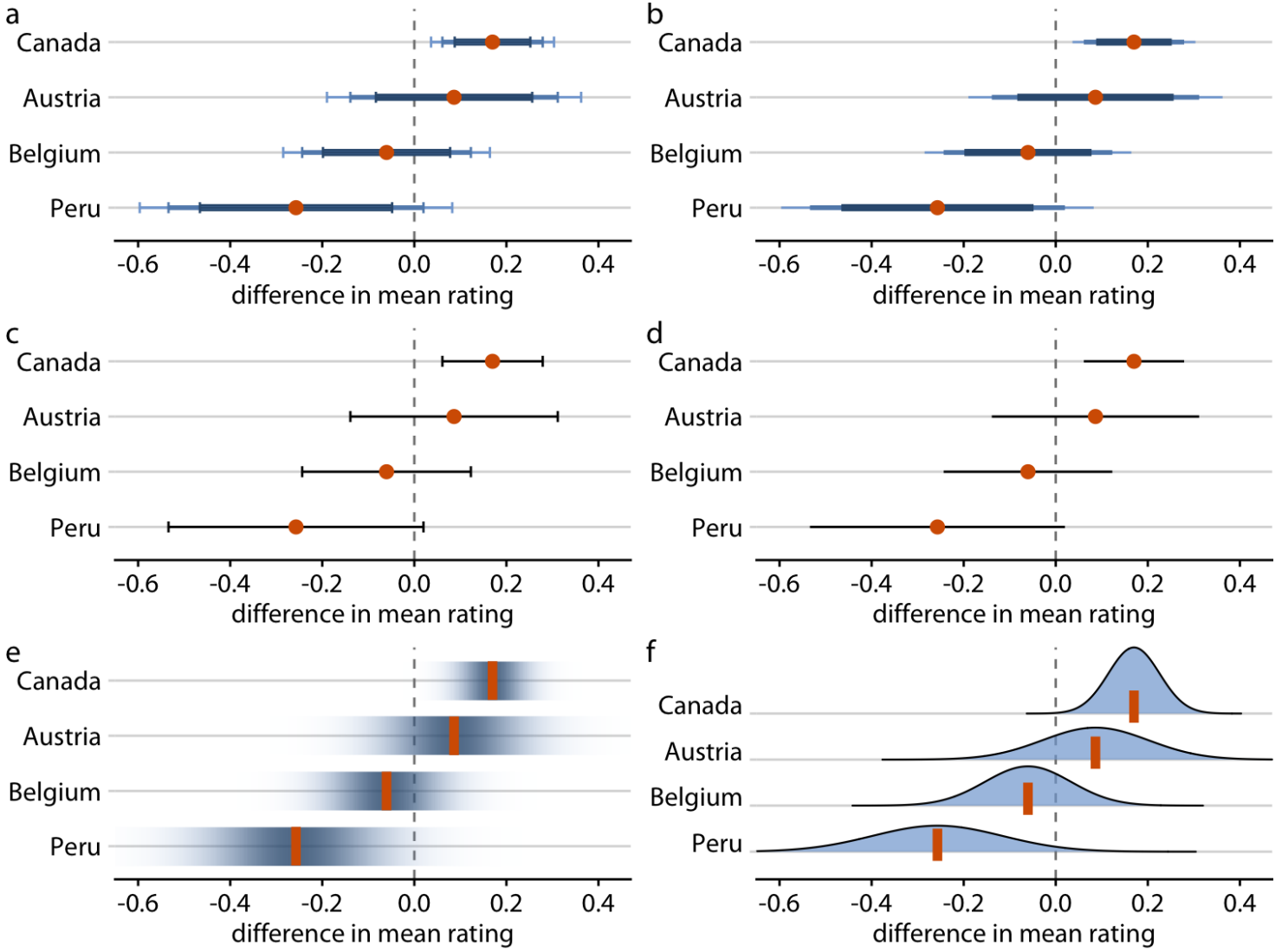
https://www.gapminder.org/videos/200-years-that-changed-the-world/

Topics we will cover (approximately)

- Week 1: Introduction, syllabus
- Week 2: Data types, viz techniques & tools
- Week 3: The human visual system
- Week 4: Visual encoding + perception
- Week 5: 2D viz - spatial data
- Week 6: 2D viz - temporal data
- Week 7: 2D viz - network data
- Week 8: 3D viz - projection & graphics
- Week 9: Animation
- Week 10: Final project proposals

Topics we will cover (approximately)

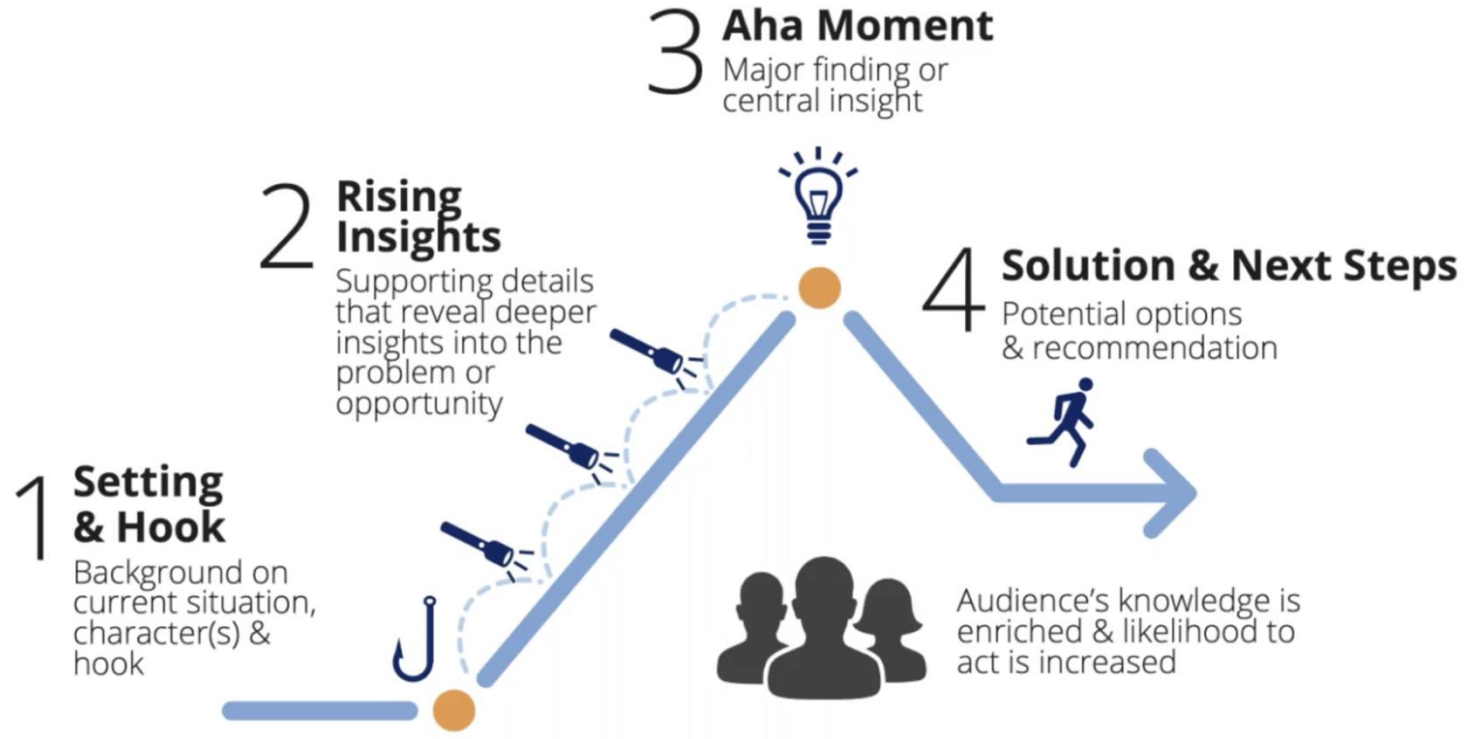
- Week 1: Introduction, syllabus
- Week 2: Data types, viz techniques & tools
- Week 3: The human visual system
- Week 4: Visual encoding + perception
- Week 5: 2D viz - spatial data
- Week 6: 2D viz - temporal data
- Week 7: 2D viz - network data
- Week 8: 3D viz - projection & graphics
- Week 9: Animation
- Week 10: Final project proposals
- Week 11: Uncertainty viz



https://clauswilke.com/dataviz/visualizing-uncertainty.html

Topics we will cover (approximately)

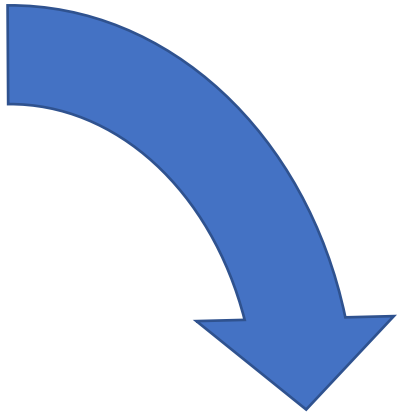
- Week 1: Introduction, syllabus
- Week 2: Data types, viz techniques & tools
- Week 3: The human visual system
- Week 4: Visual encoding + perception
- Week 5: 2D viz - spatial data
- Week 6: 2D viz - temporal data
- Week 7: 2D viz - network data
- Week 8: 3D viz - projection & graphics
- Week 9: Animation
- Week 10: Final project proposals
- Week 11: Uncertainty viz
- Week 12: Storytelling, ethics, and misinformation



https://www.datacamp.com/blog/telling-effective-data-stories-with-data-narrative-and-visuals

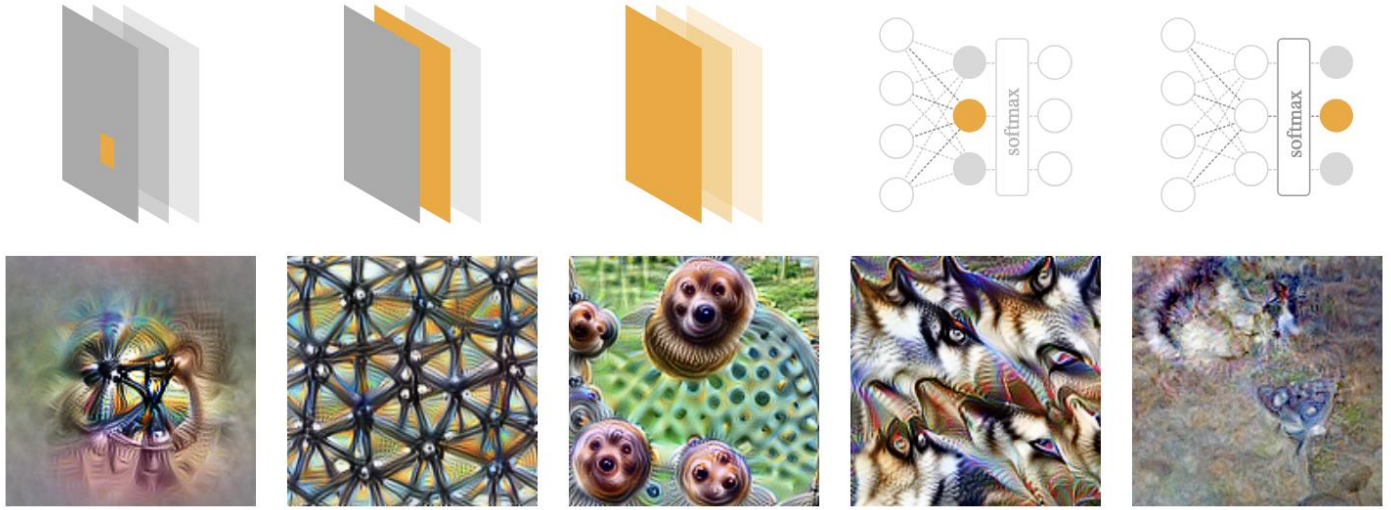
Topics we will cover (approximately)

- Week 1: Introduction, syllabus
- Week 2: Data types, viz techniques & tools
- Week 3: The human visual system
- Week 4: Visual encoding + perception
- Week 5: 2D viz - spatial data
- Week 6: 2D viz - temporal data
- Week 7: 2D viz - network data
- Week 8: 3D viz - projection & graphics
- Week 9: Animation
- Week 10: Final project proposals
- Week 11: Uncertainty viz
- Week 12: Storytelling, ethics, and misinformation
- Week 13: 🦃 Thanksgiving Break 🦃



Topics we will cover (approximately)

- Week 1: Introduction, syllabus
- Week 2: Data types, viz techniques & tools
- Week 3: The human visual system
- Week 4: Visual encoding + perception
- Week 5: 2D viz - spatial data
- Week 6: 2D viz - temporal data
- Week 7: 2D viz - network data
- Week 8: 3D viz - projection & graphics
- Week 9: Animation
- Week 10: Final project proposals
- Week 11: Uncertainty viz
- Week 12: Storytelling, ethics, and misinformation
- Week 13: 🦃 Thanksgiving Break 🦃
- Week 14: Modern topics: deep learning/XR/interaction



https://distill.pub/2017/feature-visualization/

Topics we will cover (approximately)

- Week 1: Introduction, syllabus
- Week 2: Data types, viz techniques & tools
- Week 3: The human visual system
- Week 4: Visual encoding + perception
- TODO: bump up uncertainty before the election/voting
- Week 5: 2D viz - spatial data
- Week 6: 2D viz - temporal data
- Week 7: 2D viz - network data
- Week 8: 3D viz - projection & graphics
- Week 9: Animation
- Week 10: Final project proposals
- Week 11: Uncertainty viz
- Week 12: Storytelling, ethics, and misinformation
- Week 13:  Thanksgiving Break 
- Week 14: Modern topics: deep, learning/XR/interaction
- Week 15: Final project presentations

Your work here!

Why?

Why have a user in the loop?

- We want to communicate a message
  - A message is useless without a recipient!

- Data viz allows users to answer/formulate questions
  - Questions that machines can't answer/formulate → Let the user explore with viz!
  - Use viz as a stepping-stone to developing a computer-based answer
  - Use viz as a "sanity check"

- Data viz can be used as a presentation tool

Why have a computer in the loop?

- Lots of data → need to process it all
  - Can’t draw all visualizations by hand
  - Different questions may require different processing
  - Data might change in real time

Why use an external representation?

- Renderings on a screen are external representations

- Allows you to:
  - Organize info by spatial organization
  - Offload internal cognition
  - Speed up recognition and search (leverage fast visual processing)

Why use human vision?

- What about other perceptual channels like auditory, haptic, taste, etc.?
- Visual perception is *fast* and *high bandwidth*
  - You can process more information visually than with other senses
- We have appropriate devices for vision
  - Computer displays are very advanced
  - Non-visual displays are clunky and expensive and still work poorly

Why show the data in detail?

- Why not just use summary statistics?

Why show the data in detail?

- Why not just use summary statistics?

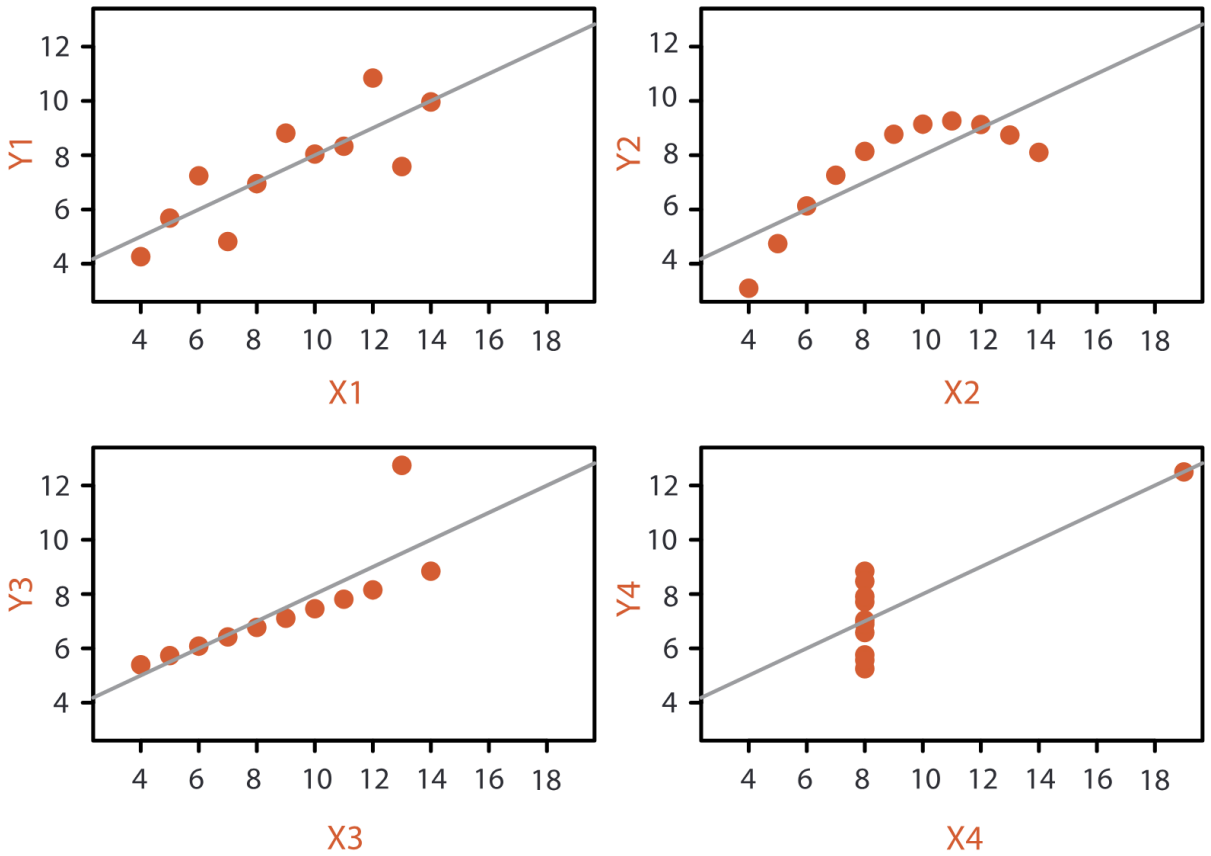
	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	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
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	

Anscombe's Quartet

Why show the data in detail?

- Why not just use summary statistics?

	1		2		3		4	
	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
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	



Anscombe's Quartet

Why show the data in detail?

- Why not just use summary statistics?
  - Specific questions likely require detailed insights into the data

Why is the viz design space so big?

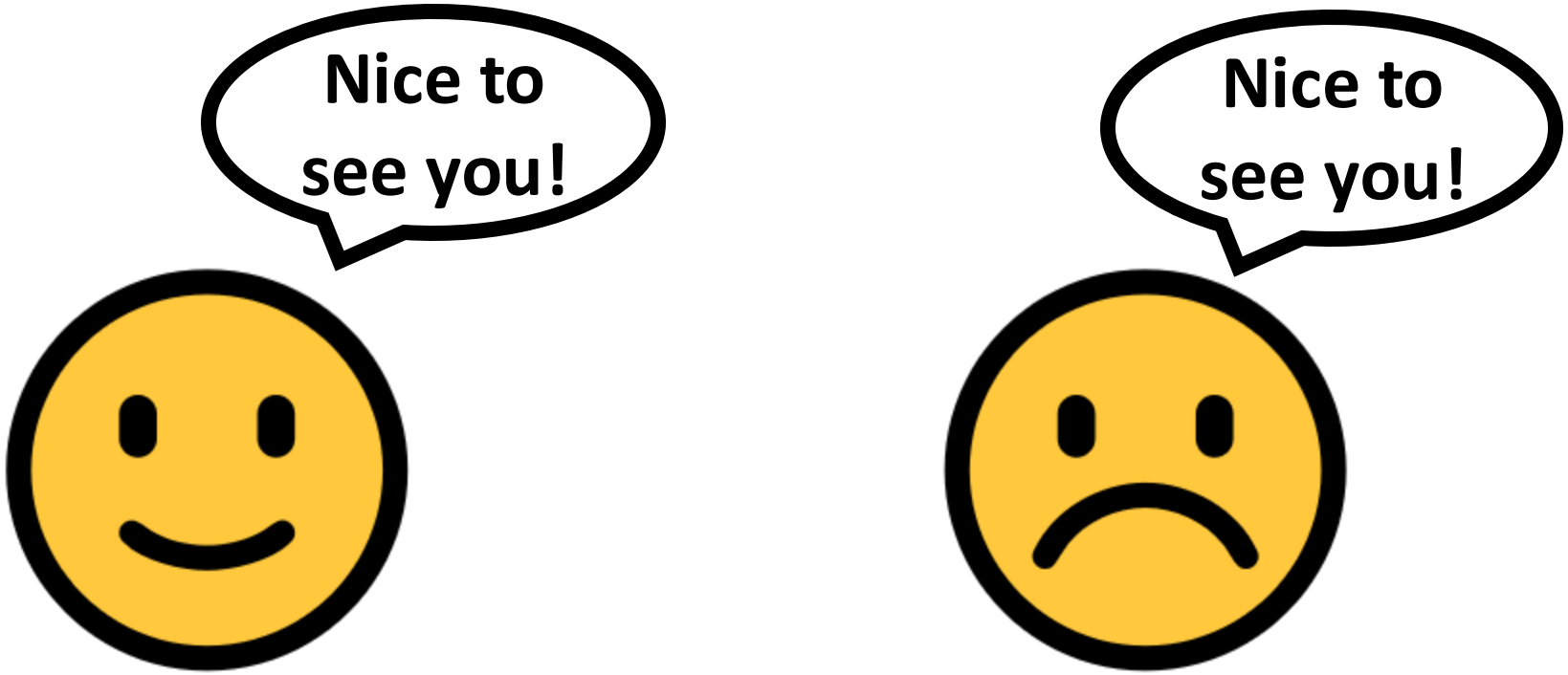
- In short: data is complex and there are many visual encodings
  - A single data type can be represented with many different visual representations
  - A single question can have many different data dimensions that answer it

Why consider different tasks?

- Remember: we want to communicate a message
- Different messages are best communicated in different ways
  - Choose the technique that is best for your message

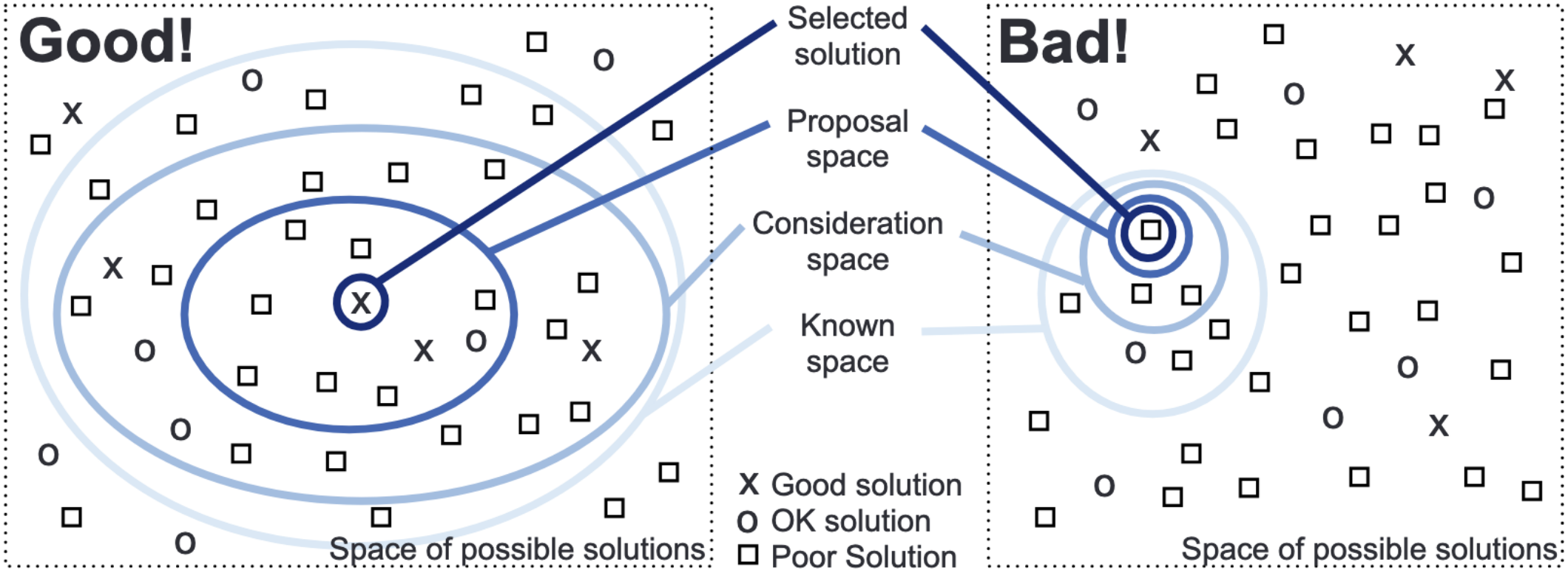
Why consider different tasks?

- Remember: we want to communicate a message
- Different messages are best communicated in different ways
  - Choose the technique that is best for your message



Why consider multiple techniques?

- There are multiple possible visualizations
  - Some will be bad
  - Some will be okay
  - Some will be good
- Consider many different options at once → more likely to consider a good design



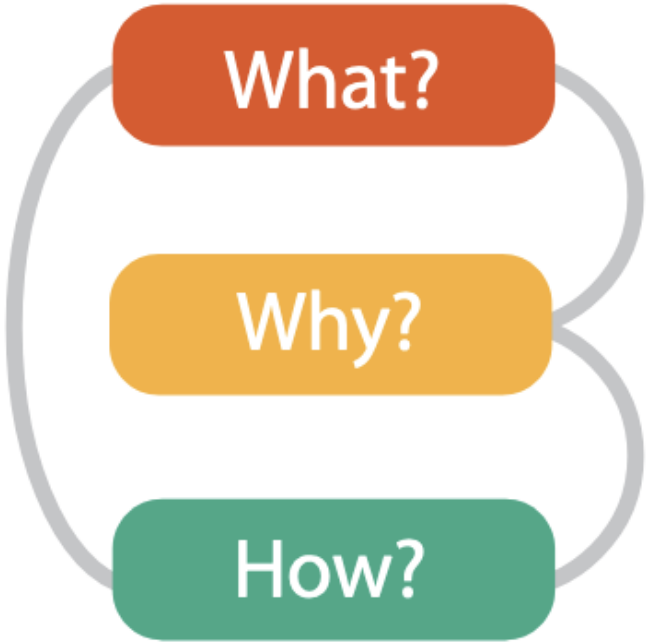
Munzner, Tamara. Visualization analysis and design. CRC press, 2014.

Why is validation difficult?

- Many reasonable ways to communicate one message
- Different representations resonate differently for different people
- Your audience's understanding will be biased by their experiences
  - Hard to predict a person's prior experiences
- Users will do unexpected things
  - Hard to predict every possible thing a user will do/think
- Many of our outcome measures are subjective
  - How do you measure "visually appealing"?

Why analyze existing visualizations?

- Short answer: most data sets are not *that* unique → there probably exists a similar data set with an associated visualization
  - Learn from others' mistakes and successes!



Why is the task being performed, what data is shown in the views, and how is the vis idiom constructed in terms of design choices.

Munzner, Tamara. Visualization analysis and design. CRC press, 2014.

Swearing map

- https://word.tips/countries-swearing-map/