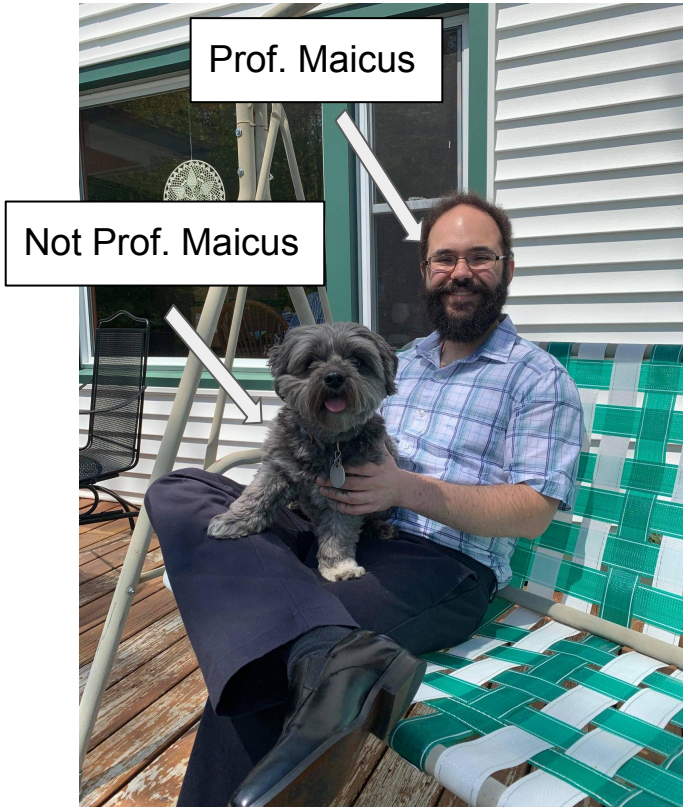
CSCI 4550/6550
Interactive Visualization
Summer 2021

https://emaicus.com/courses/u21/csci4550/landing_page

Prof. Evan Maicus
maicue@rpi.edu
https://rensselaer.webex.com/meet/maicue

Prof. Evan Maicus

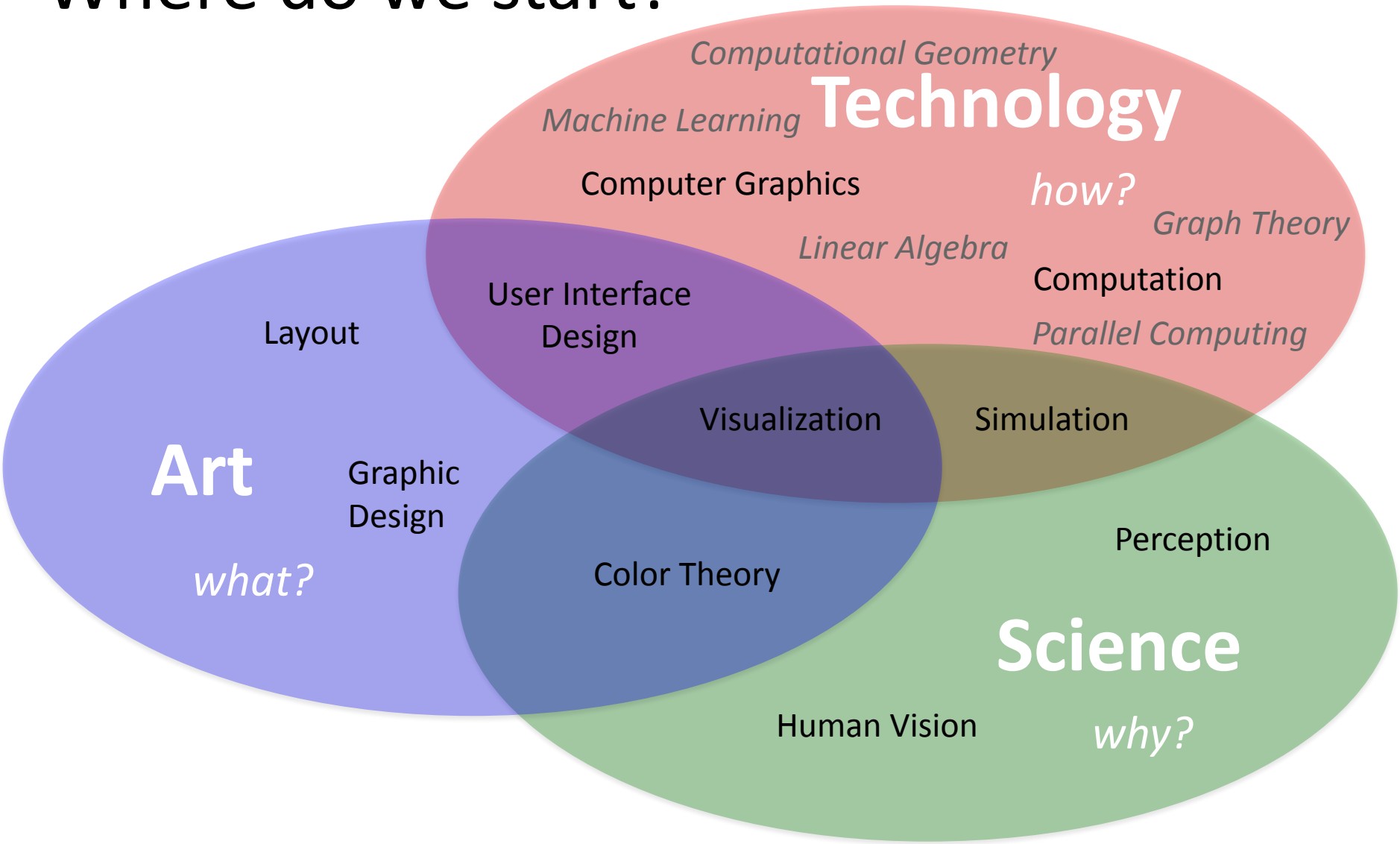
- Researcher in automated code assessment
- 3rd most commits to Submitty
- Presented at GSOC 2019 Munich Conference
- Also likes talking about 3D printing, fantasy literature, and cooking



Today

- Motivational Examples of Visualization Process
- Class Website & Syllabus
- Break!
- How do we Make Visualizations?
- Readings for Friday
- Homework 1 for Friday
- Criteria for a “good” Visualization

“Introduction” to Visualization: Where do we start?

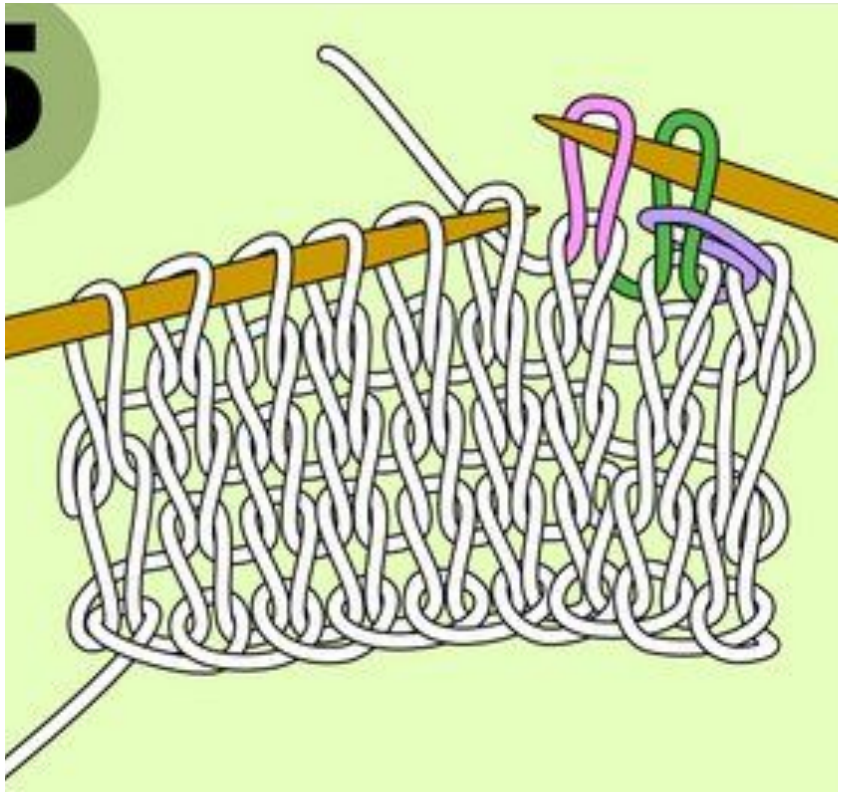


The Visualization Process

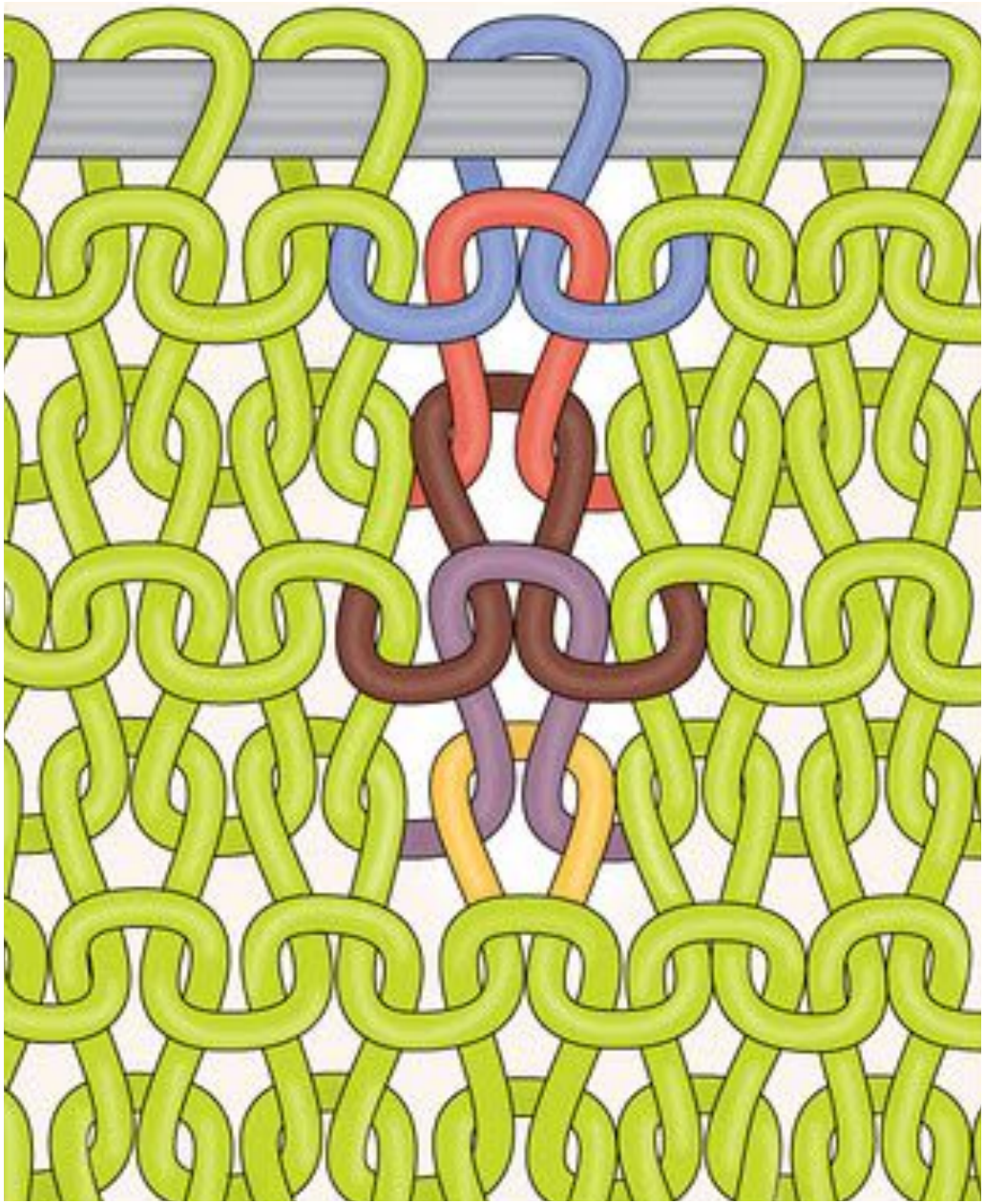
- Motivation & Problem Definition
- Visualization Design
- Data Collection
- Visualization Execution
- Analysis & Validation
- Visualization Revision
- Presentation

The Visualization Process

- Motivation & Problem Definition
  - e.g., audience, purpose, goals, interdisciplinary collaboration
- Visualization Design
- Data Collection
- Visualization Execution
- Analysis & Validation
- Visualization Revision
- Presentation



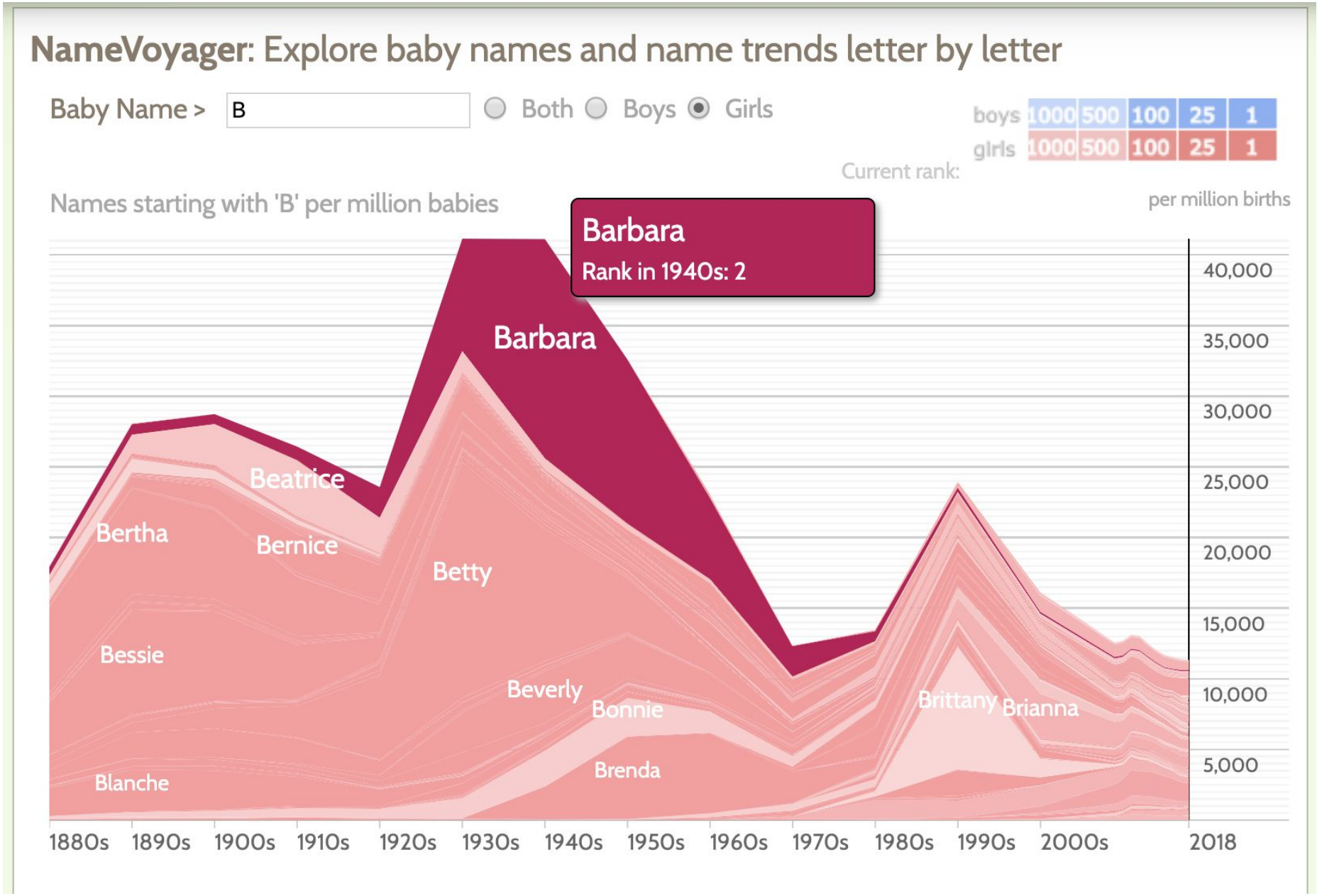
http://techknitting.blogspot.com/

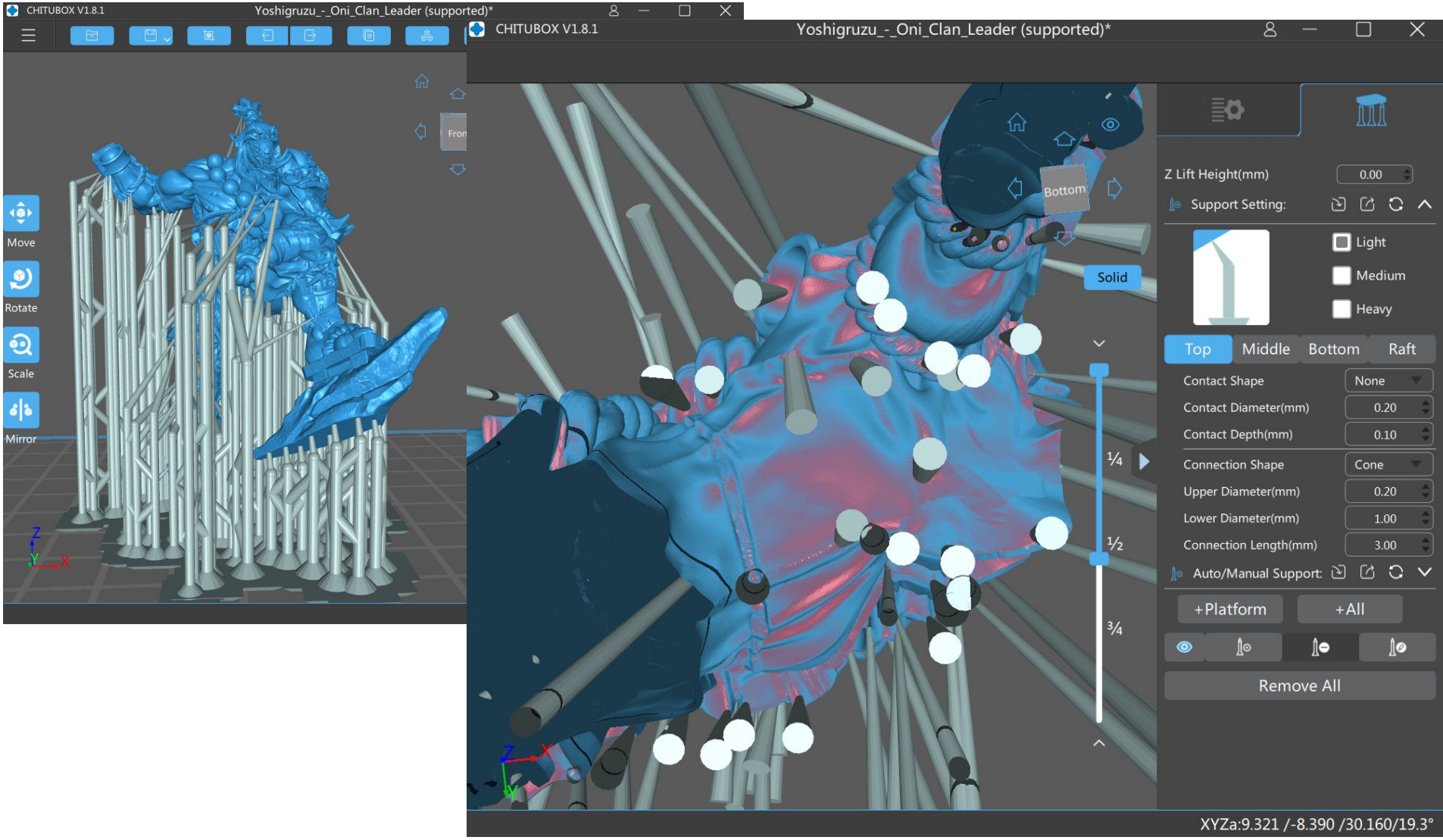


The Visualization Process

- Motivation & Problem Definition
- Visualization Design
  - e.g., media, color, organization, layout, static vs. dynamic, creativity
- Data Collection
- Visualization Execution
- Analysis & Validation
- Visualization Revision
- Presentation

http://www.babynamewizard.com/voyager



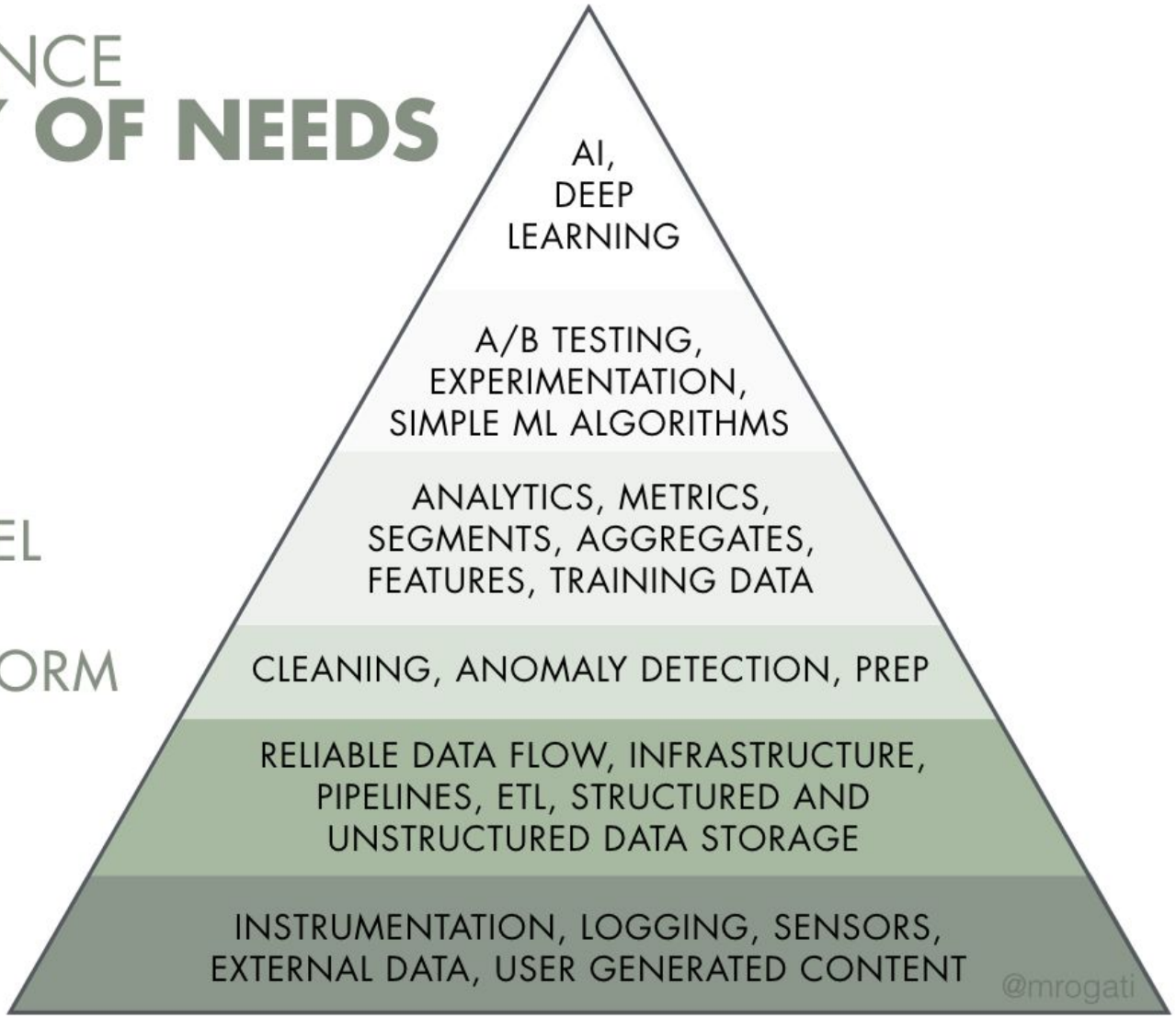


https://www.chitubox.com/en

The Visualization Process

- Motivation & Problem Definition
- Visualization Design
- Data Collection
  - e.g., data structures, file formats, parsing, performance & efficiency, databases, very large datasets, interdisciplinary collaboration
- Visualization Execution
- Analysis & Validation
- Visualization Revision
- Presentation

THE DATA SCIENCE HIERARCHY OF NEEDS



LEARN/OPTIMIZE

AGGREGATE/LABEL

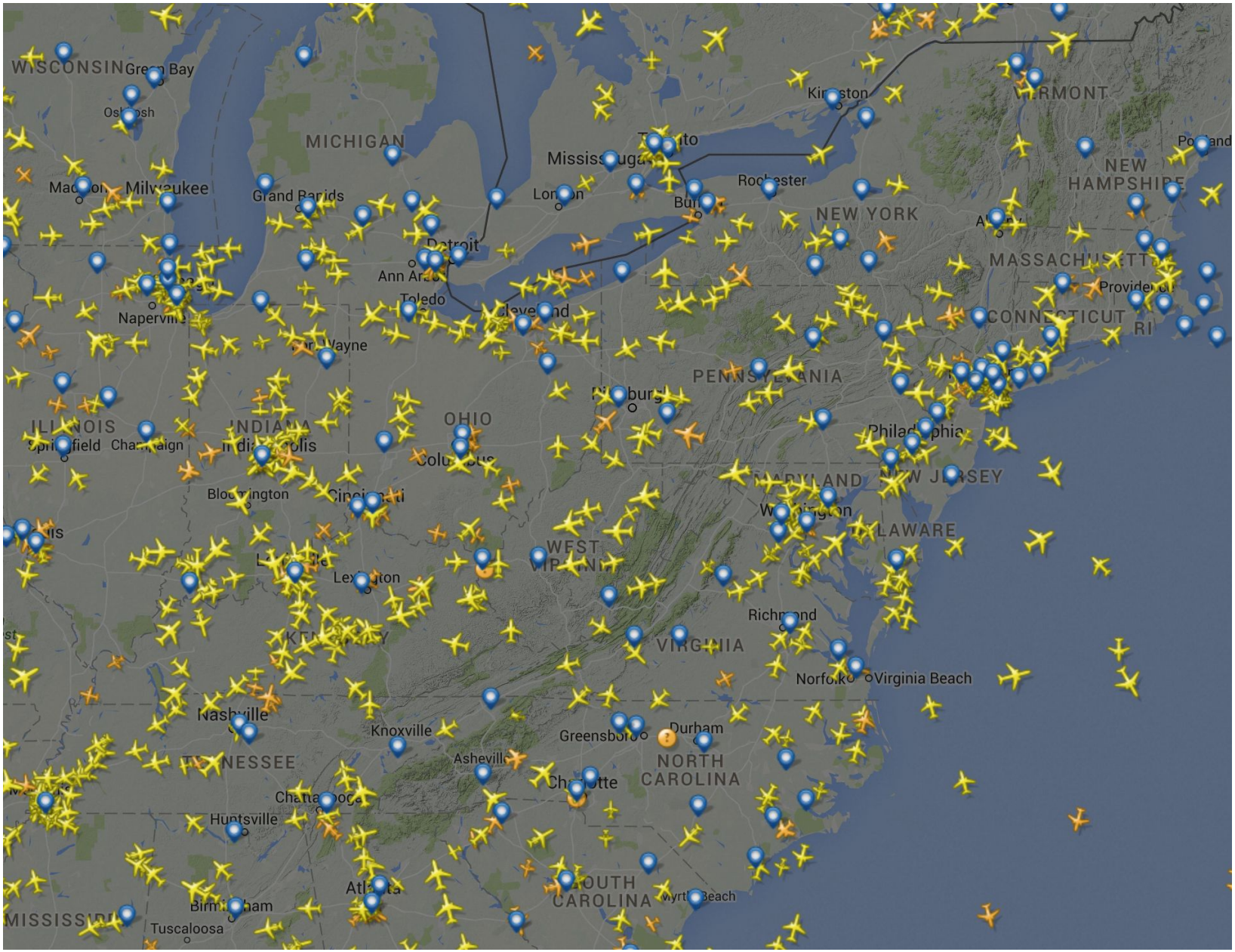
EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007

www.flightradar24.com

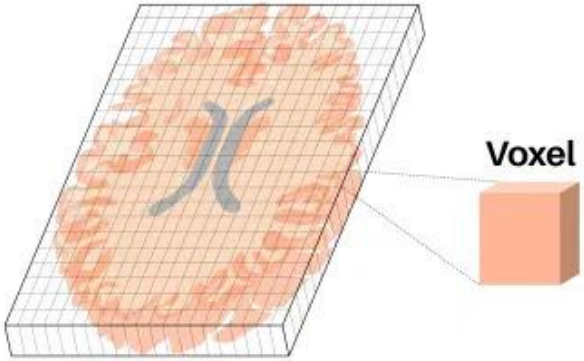


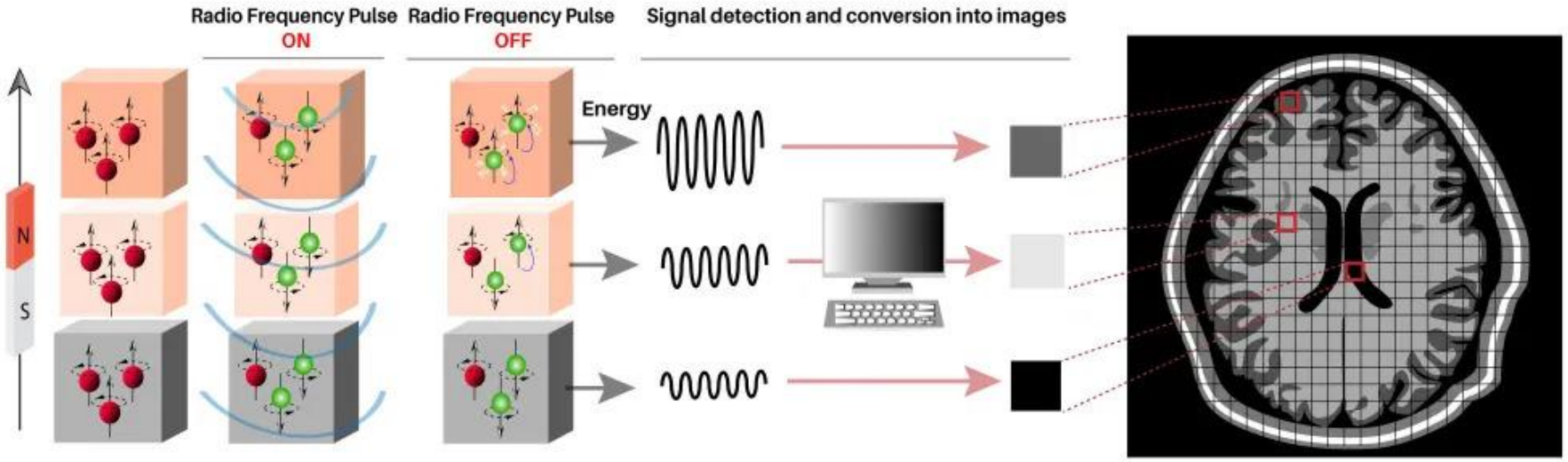

Iceland
Sweden
Norway
Finland
Baltic Sea
Estonia
Latvia
North Sea
Lithuania
Denmark
United Kingdom
Belarus
Poland
Prague
Ukraine
Moldova
Kazakhstan
France
Italy
Rome
Black Sea
Bulgaria
Georgia
Caspian Sea
Azerbaijan
Portugal
Madrid
Spain
Tyrrhenian Sea
Turkey
Turkmenistan
Syria
Tunisia
Mediterranean Sea
Lebanon
Iraq
Afghanistan
Morocco
Jordan
Algeria
Libya
Egypt
Western Sahara

Norway
Finland
Baltic Sea
Estonia
Latvia
Lithuania
North Sea
United Kingdom
Belarus
Ukraine
Moldova
France
Croatia
Black Sea
Greece
Spain
Turkmenistan
Syria
Iraq
Iran
Jordan
Tunisia
Morocco
Algeria
Libya
Egypt

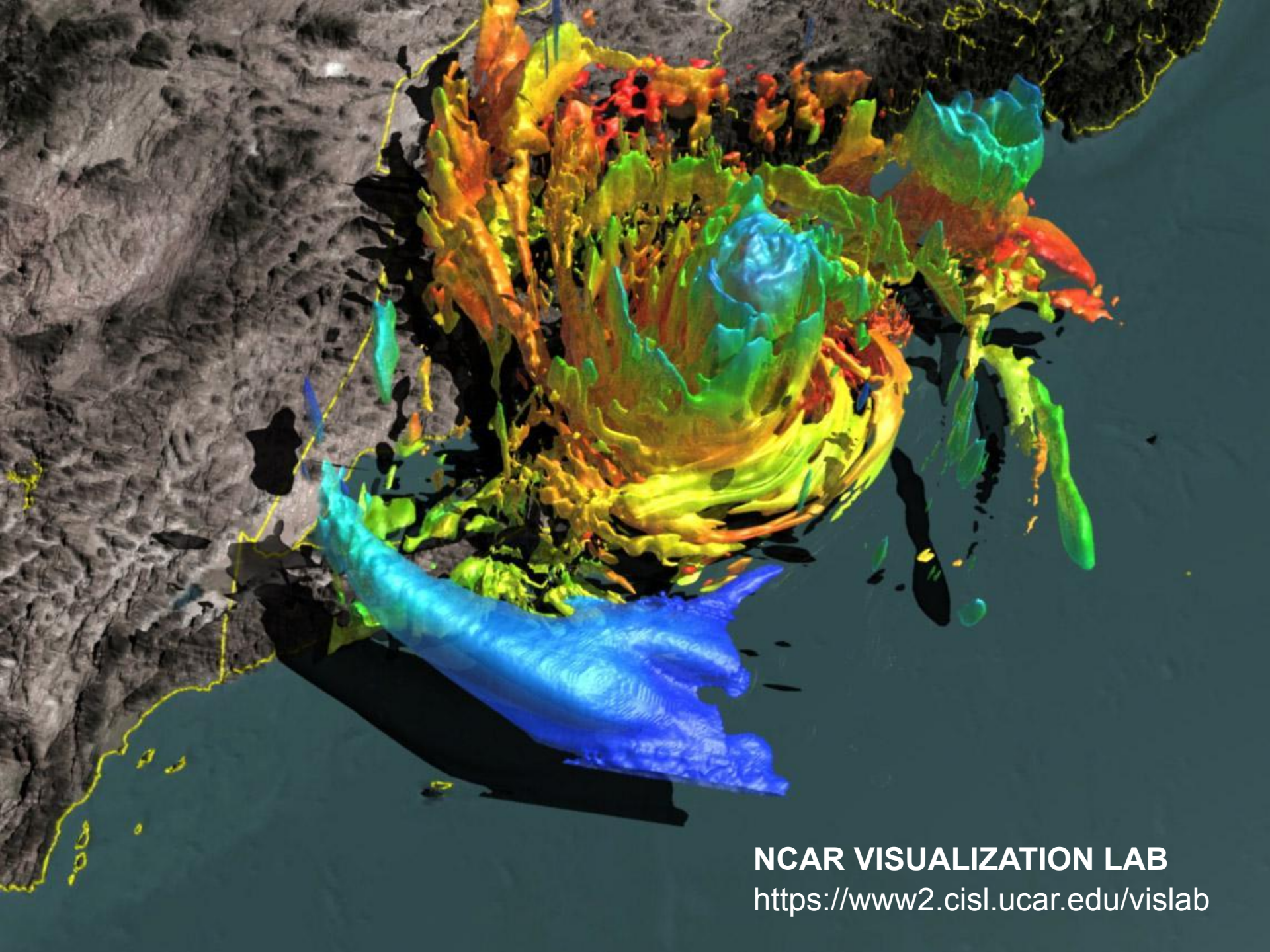
The Visualization Process

- Motivation & Problem Definition
- Visualization Design
- Data Collection
- Visualization Execution
  - e.g., data structures, implementation details, visualization tookits/environments (VTK, OpenGL, d3.js, etc.), performance & efficiency
- Analysis & Validation
- Visualization Revision
- Presentation





https://knowingneurons.com/2017/09/27/mri-voxels/


NCAR VISUALIZATION LAB
https://www2.cisl.ucar.edu/vislab

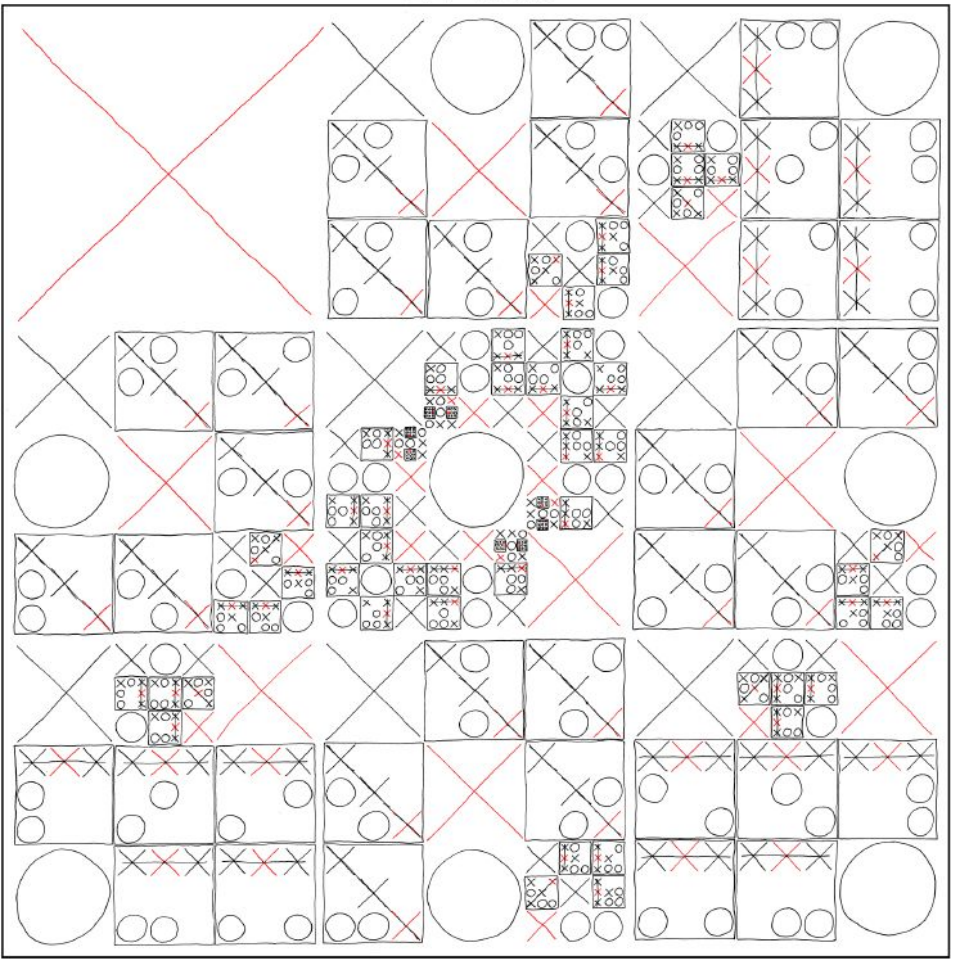
The Visualization Process

- Motivation & Problem Definition
- Visualization Design
- Data Collection
- Visualization Execution
- Analysis & Validation
  - e.g., debugging, drawing conclusions from data, accuracy, precision, interpretation, useability
- Visualization Revision
- Presentation

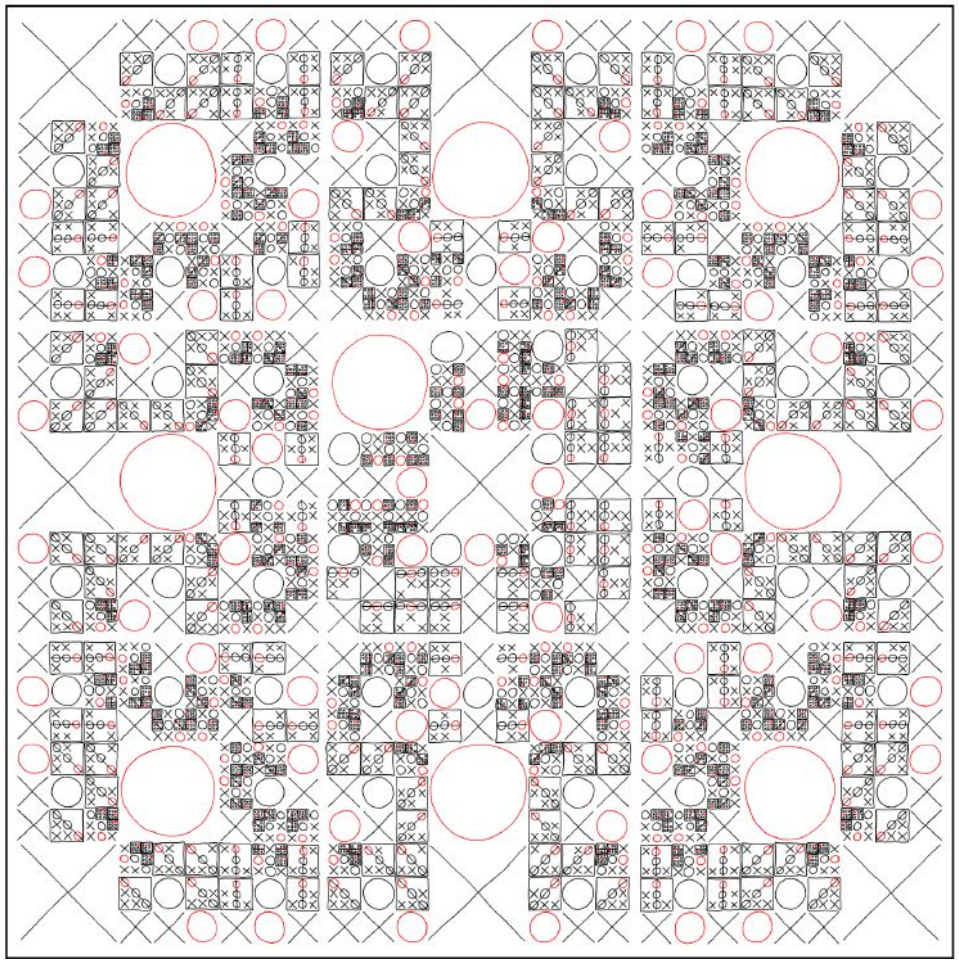
COMPLETE MAP OF OPTIMAL TIC-TAC-TOE MOVES

YOUR MOVE IS GIVEN BY THE POSITION OF THE LARGEST RED SYMBOL ON THE GRID. WHEN YOUR OPPONENT PICKS A MOVE, ZOOM IN ON THE REGION OF THE GRID WHERE THEY WENT. REPEAT.

MAP FOR X:



MAP FOR O:



http://xkcd.com/832/

The Visualization Process

- Motivation & Problem Definition
- Visualization Design
- Data Collection
- Visualization Execution
- Analysis & Validation
- Visualization Revision
  - e.g., prototype & revise, iterated design, comparing before & after, solicit user feedback, formal user studies
- Presentation

2019 Agency Goal Progress

Service	2019 Goals	Q1 Progress	Q2 Progress	Q3 Progress	Q4 Progress	YTD Progress
# of Participants **Served** • *Defined as receiving at least one service*	4,000	1,800	1,600			3,400
# of Participants **Gaining Skills** for education or employment • *Defined as skills gained through training, workshop or work readiness*	2,000	500	400			900
Financial Wellness						
• Financial Literacy classes	600	100	150			250
• Financial Coaching	500	100	120			220
Career Training & Education						
• Training Enrollment/ Completion (65%)	1,000/ 500	200/150	250/200			450/350
• Credential completion	250	80	60			140
Placements • Job Placements *(includes Internships/Trainees)*	3,000	400	500			900
Increase **Average Wage** of people we serve *(does not include internship/trainee wages)*	$15.00	$14.80	$14.60			$14.70
Increase **Job Retention** of those we serve						
• 3 month	65%	70%	80%			75%
• 6 month	65%	60%	70%			65%
• 12 month	65%	50%	60%			55%
Transitional Employment						
• # of Trainee jobs	1,000	250	200			450
• # of Trainees that move to external jobs	500	100	120			220
Community Engagement • *Defined as attendees at community events, job fairs and information Session*	---	600	400			1,000

https://depictdatastudio.com/how-to-use-an-iterative-process-to-hone-the-perfect-data-visualization/

https://depictdatastudio.com/how-to-use-an-iterative-process-to-hone-the-perfect-data-visualization/

2019 Agency Goal Progress

Goal Area	Specific Goal	2019 Goal	Q1 Progress	Q2 Progress	Q3 Progress	Q4 Progress	YTD Results	YTD Progress	On Track?
Services	Participants **Served** *with at least one service*	4000	1800	1600			3400	85%	
Skill Gains	Participants **Gaining Skills** *via training or workshop*	2000	500	400			900	45%	
Financial Wellness	Participants in **Financial Literacy** classes	600	100	150			250	42%	
	Participants in **Financial Coaching**	500	100	120			220	44%	
Career Training and Education	Participants with **Training Enrollments**	1000	200	250			450	45%	
	Participants with **Training Completions**	500	150	200			350	70%	
	Participants with **Credentials Attained**	250	80	60			140	56%	
Placements	**Job Placements** *including interns and trainees*	3000	400	500			900	30%	
Wages	**Average Wage** *not including intern/trainee wages*	$15.00	$14.80	$14.60			$14.70		
Job Retention	**3 Month Retention**	65%	70%	80%			75%		
	6 Month Retention	65%	60%	70%			65%		
	12 Month Retention	65%	50%	60%			55%		
Transitional Employment	**Trainee Jobs**	1000	250	200			450	45%	
	Trainees moved to external jobs	500	100	120			220	44%	
Community Engagement	**People engaged at community events**, job fairs, and information sessions	---	600	400			1000		

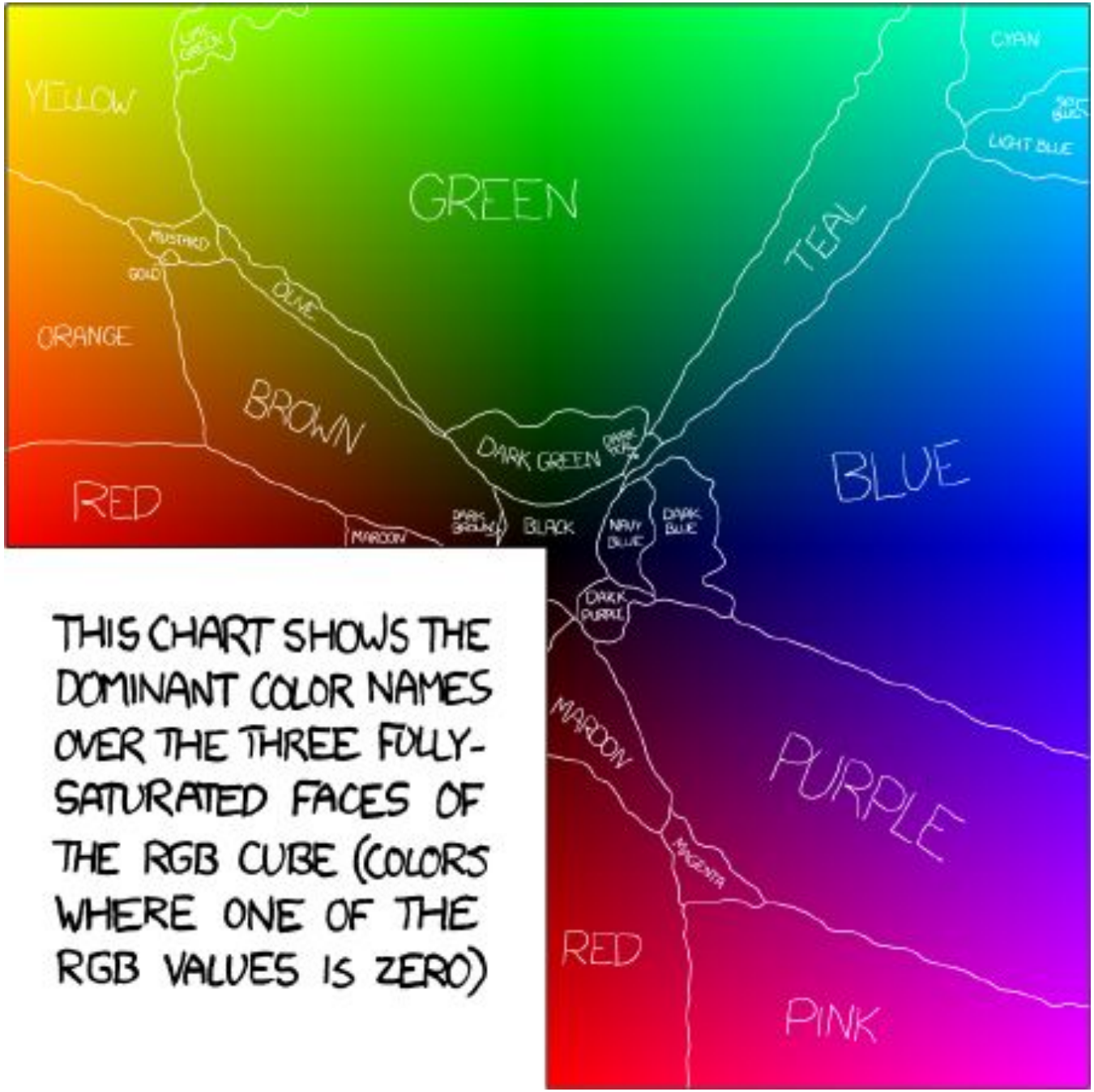
http://www.heraldsun.com.au/ Getty

http://www.digitalglobe.com/sites/default/files/italy_giglio_jan17_2012_0.jpg

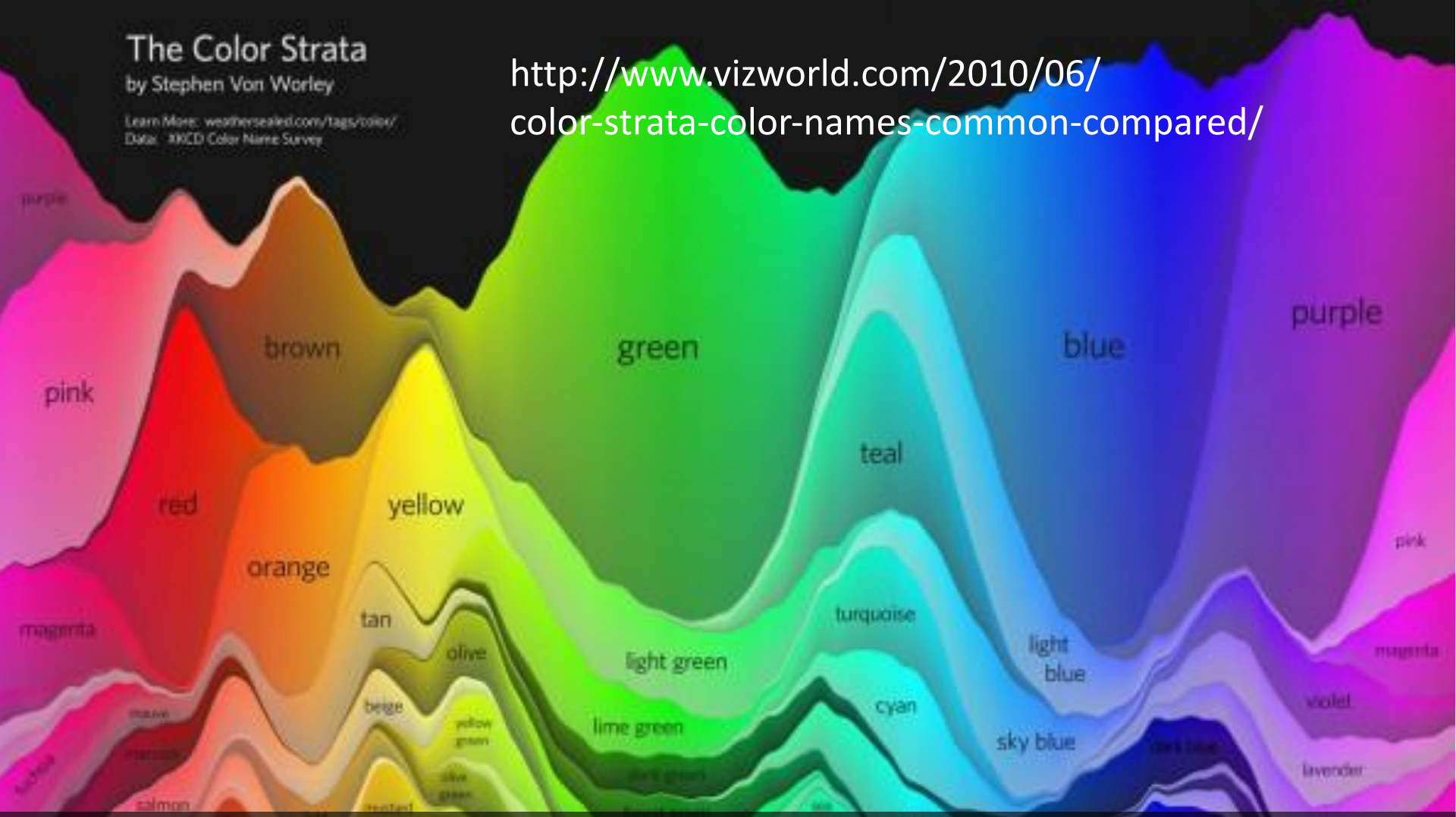
From somewhere on Facebook….
http://www.facebook.com/babayoff

The Visualization Process

- Motivation & Problem Definition
- Visualization Design
- Data Collection
- Visualization Execution
- Analysis & Validation
- Visualization Revision
- Presentation
  - e.g., mixed media, descriptive titles/labels, concise and complete captions/companion text, elevator pitch, documentation



https://blog.xkcd.com/2010/05/03/color-survey-results/



"The Color Strata includes the 200 most common color names (excluding black-white-grayish tones), organized by hue horizontally and relative usage vertically, stacked by overall popularity, shaded representatively, and labeled where possible. Besides filtering spam, ignoring cruft, normalizing grey to gray, and correcting the most egregious misspellings (here's looking at you, fuchsia), the results are otherwise unadulterated. As such, similar color names, like sea green, seafoam green, and seafoam, each appear separately. They're synonymous… or are they?"

Today

- Motivational Examples of Visualization Process
- Class Website & Syllabus
- Break!
- How do we Make Visualizations?
- Readings for Friday
- Homework 1 for Friday
- Criteria for a “good” Visualization

Website, Syllabus, & Course Grades

Syllabus

Interactive Visualization, Summer 2021

Course Overview

Visualizing data is a key step in understanding many problems. This course is designed to introduce students to methods of visualizing many different types of data, such as images, numerical/statistical data, 3D surfaces, flow fields, and medical data. We will both use existing visualization software and program custom visualizations using JavaScript. Course activities include discussion of recent and classic research papers, weekly homework assignments, in-class critiques of visualization artifacts, and a final project to explore creative uses of these techniques.

Prerequisites

CSCI 1200 Data Structures and CSCI 2300 Intro to Algorithms or CSCI 2600 Principles of Software or permission of instructor. C++ and sufficient prior programming experience is required.

Learning Outcomes

Students who have successfully completed this course will be able to:

- Analyze, interpret, and evaluate a specific visualization example and discuss how the visualization might be improved for more accurate interpretation or communication of patterns in the data.
- Select or design an effective visualization strategy for a variety of different types of data.
- Create a visualization of a new dataset using available open-source visualization resources.
- Use visualization to communicate results of experiments and research in their field of study.
- Incorporate visualization for debugging and improved program development or experimental data analysis in their field of study.

Course Grades

https://emaicus.com/courses/u21/csci4550/syllabus

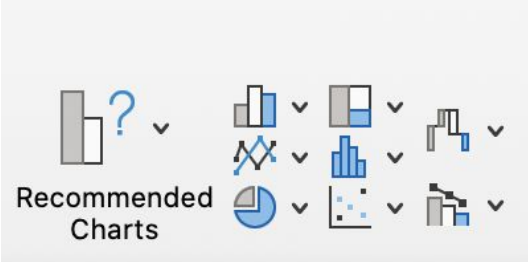
Today

- Motivational Examples of Visualization Process
- Class Website & Syllabus
- Break!
- How do we Make Visualizations?
- Readings for Friday
- Homework 1 for Friday
- Criteria for a “good” Visualization

Today

- Motivational Examples of Visualization Process
- Class Website & Syllabus
- Break!
- How do we Make Visualizations?
- Readings for Friday
- Homework 1 for Friday
- Criteria for a “good” Visualization

Google Sheets


Recommended
Charts

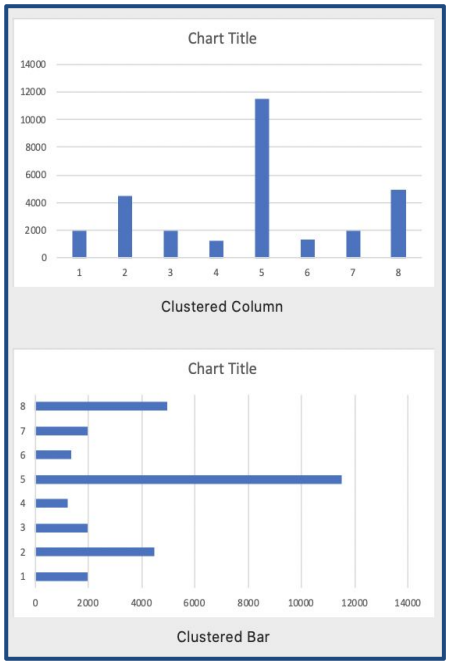


Chart Title
Clustered Column
Chart Title
Clustered Bar

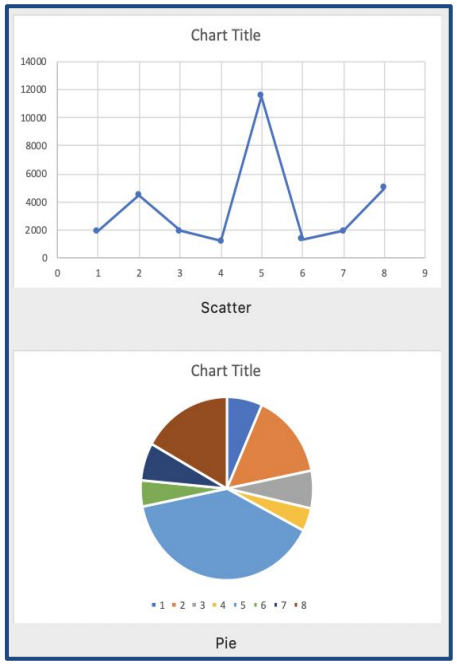
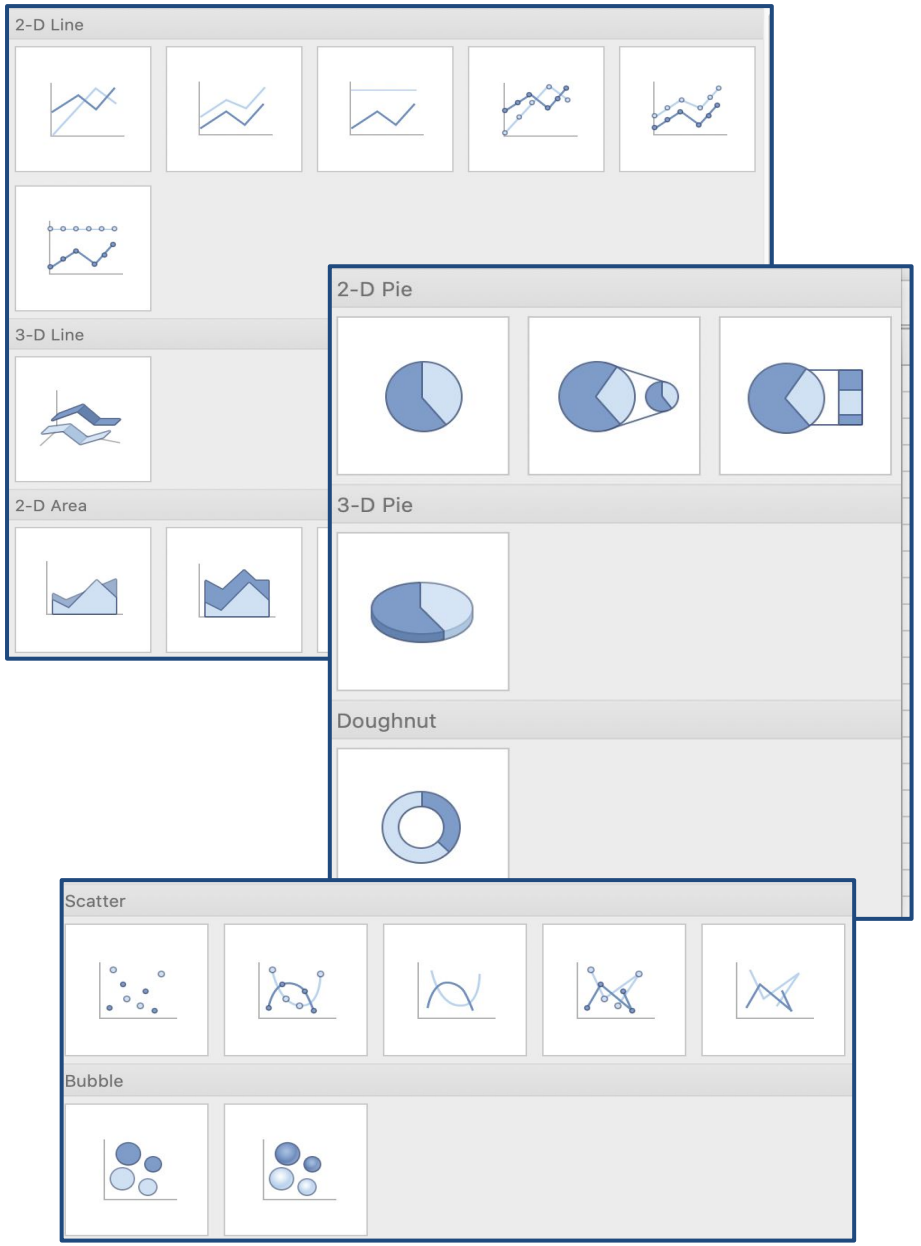
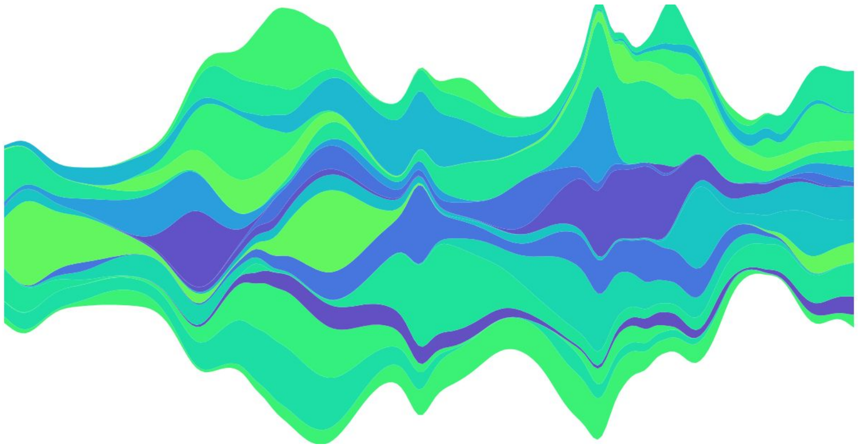


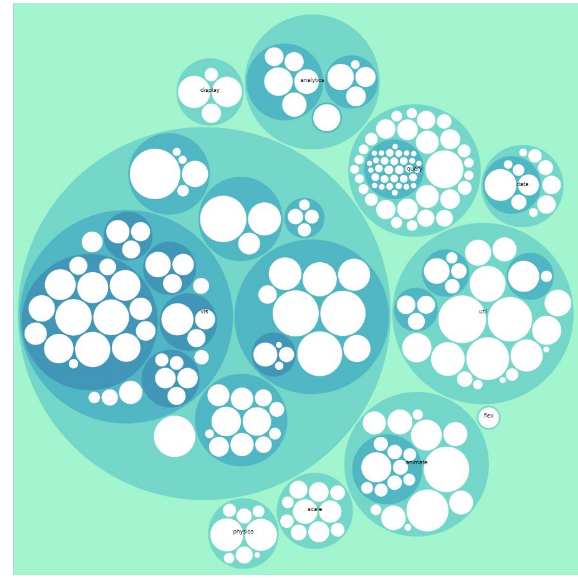
Chart Title
Scatter
Chart Title
Pie

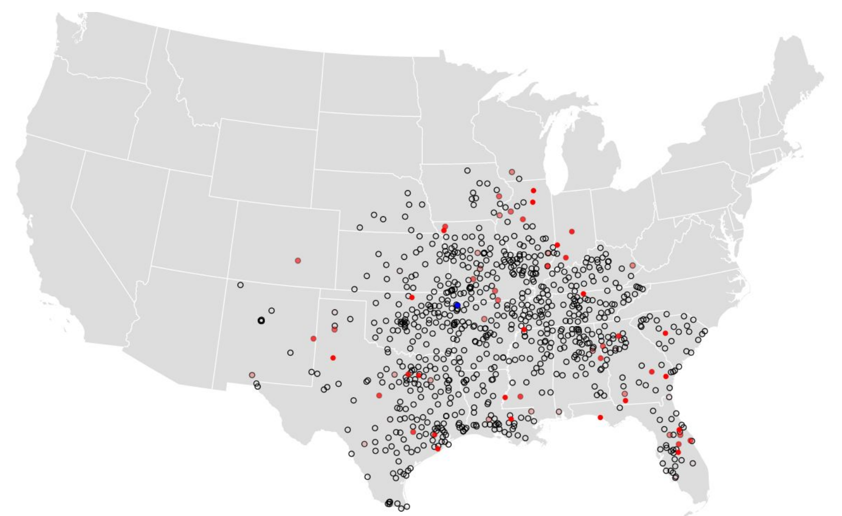

2-D Line
3-D Line
2-D Area
2-D Pie
3-D Pie
Doughnut
Scatter
Bubble

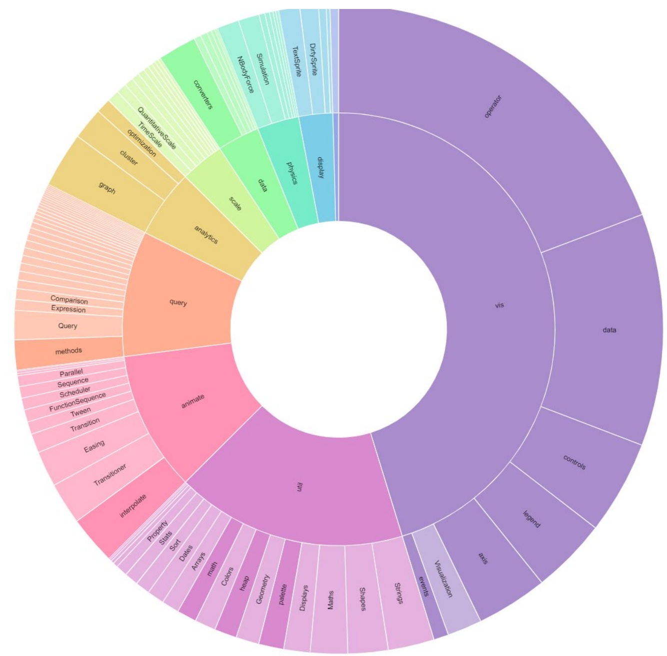
Excel

d3.js

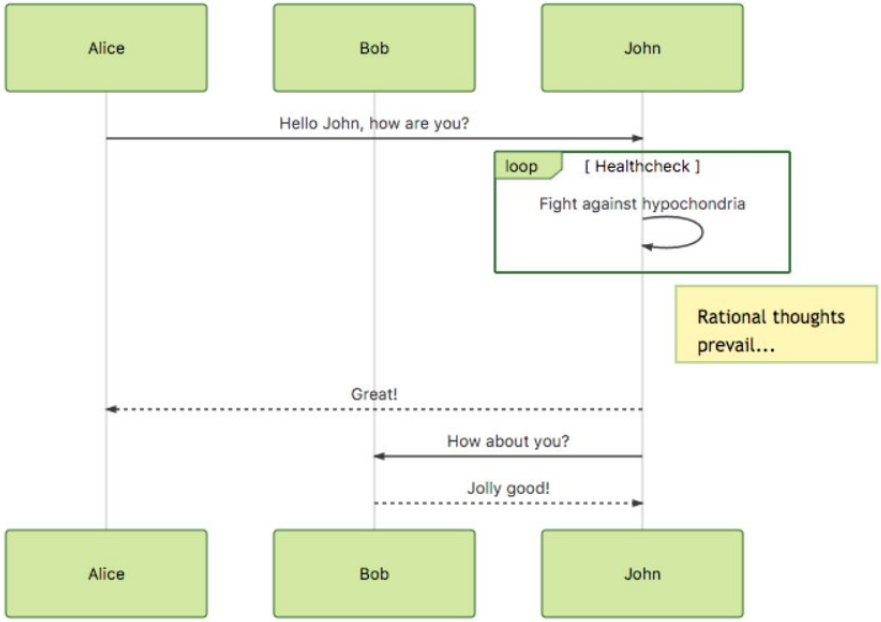


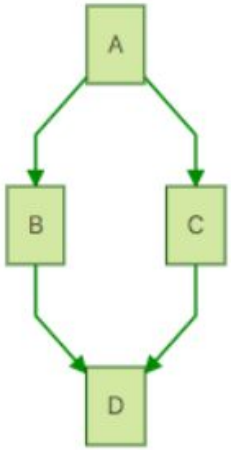

display
analytics
Query
data
vis
util
flex
animate
scale
physics

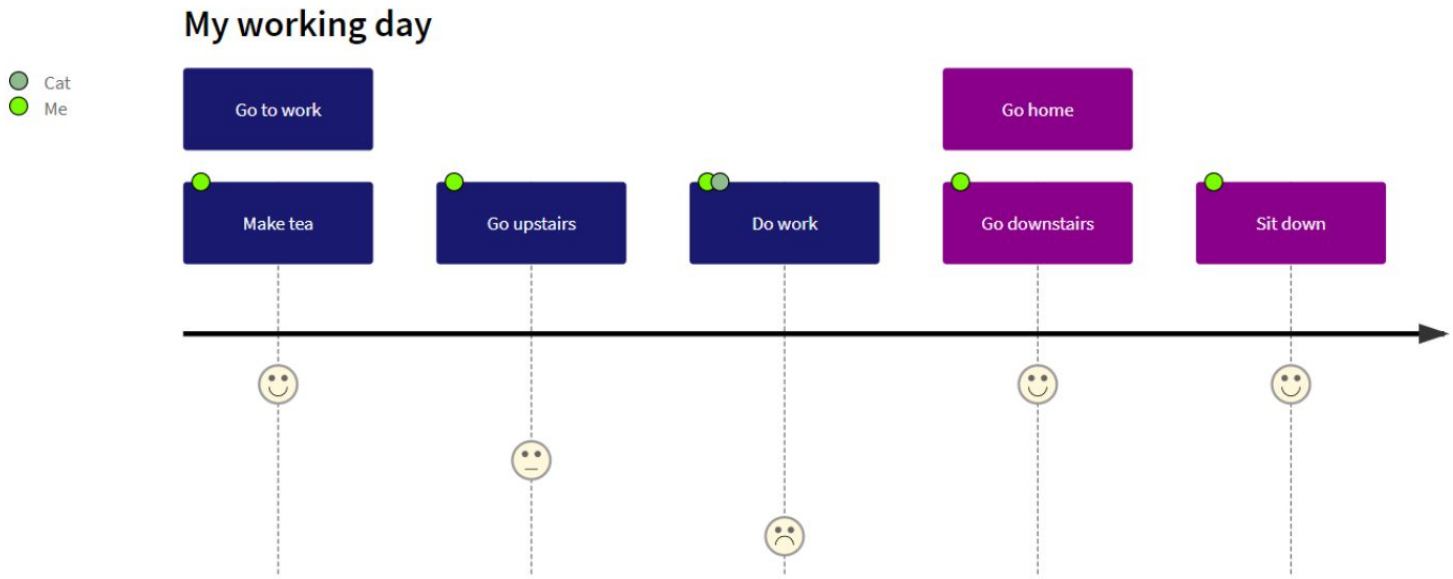



operator
vis
data
controls
legend
axis
Visualization
events
util
Strings
Shapes
Maths
Displays
palette
Geometry
heap
Colors
math
Arrays
Dates
Sort
Stats
Property
interpolate
Transitioner
Easing
Transition
Tween
FunctionSequence
Scheduler
Sequence
Parallel
animate
methods
Query
Expression
Comparison
query
graph
cluster
optimization
TimeScale
QuantitativeScale
analytics
scale
data
converters
NBodyForce
Simulation
physics
TextSprite
DirtySprite
display

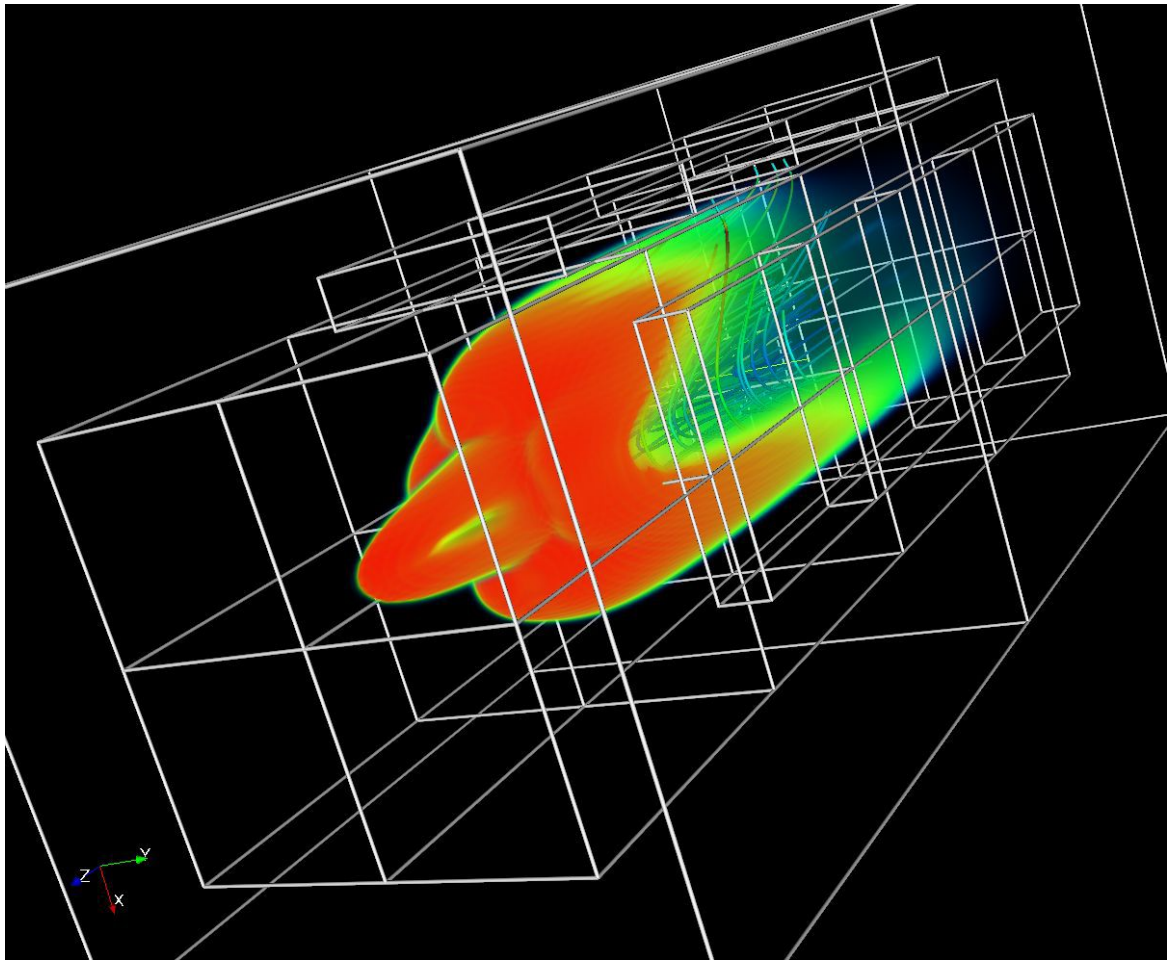
Mermaid.js


Alice
Bob
John
Hello John, how are you?
loop
[Healthcheck]
Fight against hypochondria
Rational thoughts prevail...
Great!
How about you?
Jolly good!
Alice
Bob
John


A
B
C
D


My working day
Cat
Me
Go to work
Make tea
Go upstairs
Do work
Go home
Go downstairs
Sit down

Paraview



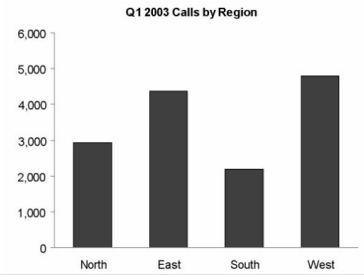
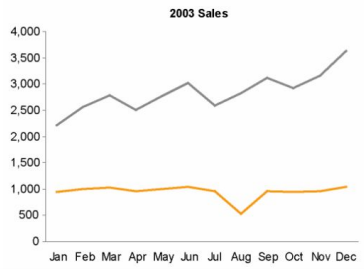
Today

- Motivational Examples of Visualization Process
- Class Website & Syllabus
- Break!
- How do we Make Visualizations?
- Readings for Friday
- Homework 1 for Friday
- Criteria for a “good” Visualization

Reading 1 (Choose 1)

Post a comment/question on LMS by Thursday @8pm

- "Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message", Stephen Few, Intelligent Enterprise, 2004

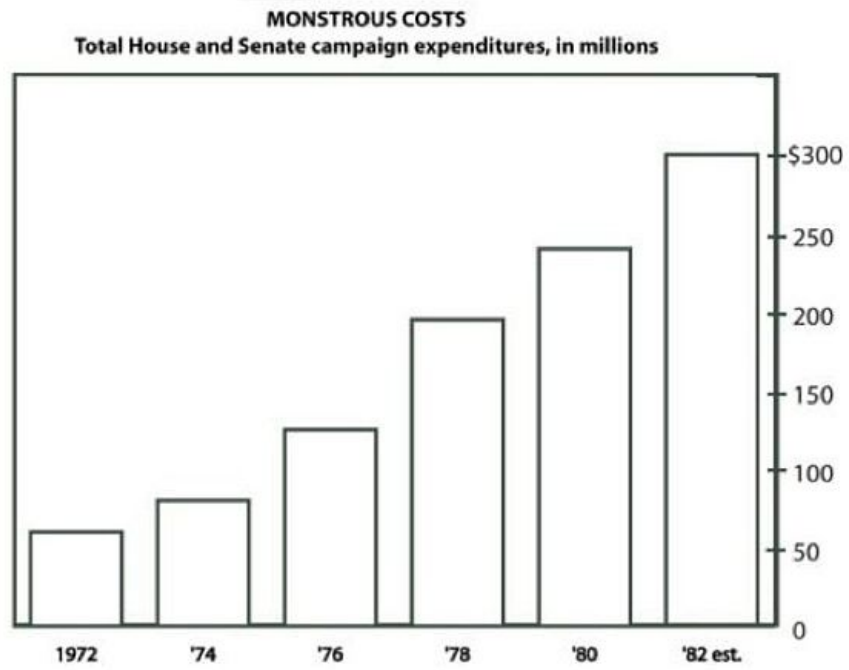
Type/Description	Encoding Methods	Example
Nominal Comparison A simple comparison of the categorical subdivisions of one or more measures in no particular order	• Bars only (horizontal or vertical)	
Time Series Multiple instances of one or more measures taken at equidistant points in time	• Lines to emphasize overall pattern • Bars to emphasize individual values • Points connected by lines to slightly emphasize individual values while still highlighting the overall pattern • Always place time on the horizontal axis	
Ranking Categorical subdivisions of a measure ordered by size (either descending or ascending)	• Bars only (horizontal or vertical) • To highlight high values, sort in descending order • To highlight low values, sort in ascending order	

- http://www.perceptualedge.com/articles/ie/the_right_graph.pdf

Reading 2 (Choose 1)

Post a comment/question on LMS by Thursday @8pm

- "Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts" Bateman et al., CHI 2010.



http://eagereyes.org/criticism/chart-junk-considered-useful-after-all

Today

- Motivational Examples of Visualization Process
- Class Website & Syllabus
- Break!
- How do we Make Visualizations?
- Readings for Friday
- Homework 1 for Thursday
- Criteria for a “good” Visualization

Homework Assignment 1: Inspirational Visualization Images

- Find two example visualization *images*:
  - one great visualization
  - one example that needs revision to be effective
- For each example write a paragraph or two describing:
  - the author, context, audience, original media format and purpose of the visualization
  - your analysis of the positive and negative aspects of each example and how it could be improved, and
  - any generalizations you can make about what makes for a compelling, high-quality visualization
- Upload your assignment to Submitty by 11:59pm on Thursday. And post one of the images on the forum...

Today

- Motivational Examples of Visualization Process
- Class Website & Syllabus
- Reading for Today
- Reading for Friday
- Homework 1 for Friday
- Criteria for a “good” Visualization

Is this a Visualization?

“Been wondering for years where it is cats put their feet when they settle down into this pose”
“whoa, so that’s how they do it!”

From somewhere on Facebook

Criteria for label “(good) Visualization”

- Cat == visualization
- Easy to understand/intuitive
- Aesthetic appeal
- Clean and well formatted
- Accurate
- Clearly see why the visualization work
- Quantifiable data, clear and accurate
- Effectively present key info
- Good perspective
- “Un-misinterpretable”
- Accurate and clear titles and labels.

Criteria for label “(good) Visualization”

- From the S20 course:
- Needs to have numbers/be based on data -- or maybe not?
  - Not all visualization have quantitative data, coordinates
  - Flow charts are visualization
- Puts an image to something
- Should not have extra stuff, should be simplified to show the point/purpose (extract), don’t show unnecessary context
- Not just an observation, should be an abstraction of the information
- Add something, serve a purpose (if we had overlaid/augmented with outlines, or did a cross section), a table of data isn’t a visualization, but a graph of data is
- Convey information by showing view that is not normally seen
- If this isn’t, what is???
- If you didn’t work hard to produce it, it is not a visualization
- Need enough data to make comparison, help people make conclusion/model
- Should have companion text/explanation, should be labeled
- Good use of image, hook to get you to read the text
- Would be better if it were animated, multi-frame

Criteria for label “(good) Visualization”

- From F18 course:
- Some image, with text, describes purpose
  - Maybe animation/interaction can substitute for needed text
  - Or maybe not even text required (could rely on context/convention/intuition/human experience)
  - Text should be concise
  - Maybe pictogram instead of text (symbols for good vs bad)
- Clearly convey intended information, be clear
  - If text is needed… its there, if not then don’t
- Address question or concern
  - Purpose should be clear
- Be accurate, don’t misrepresent data
- Has added value: more intuitive than text (a reason why not just text), information that is not easily available (or ok if its a shift in perspective a new way of looking at info)
- Fit into the argument of paper (the context)
- Understand the audience
- Pleasing to look at
- Avoid superfluous information
- Should cite sources! (available, but off to the side)
  - Give credit to author
  - Verify the credibility

Under-Dogs: I Photograph Dogs from Underneath, Andrius Burba, https://www.boredpanda.com/under-dogs-i-photograph-dogs-from-underneath/

Criteria for label “(good) Visualization”

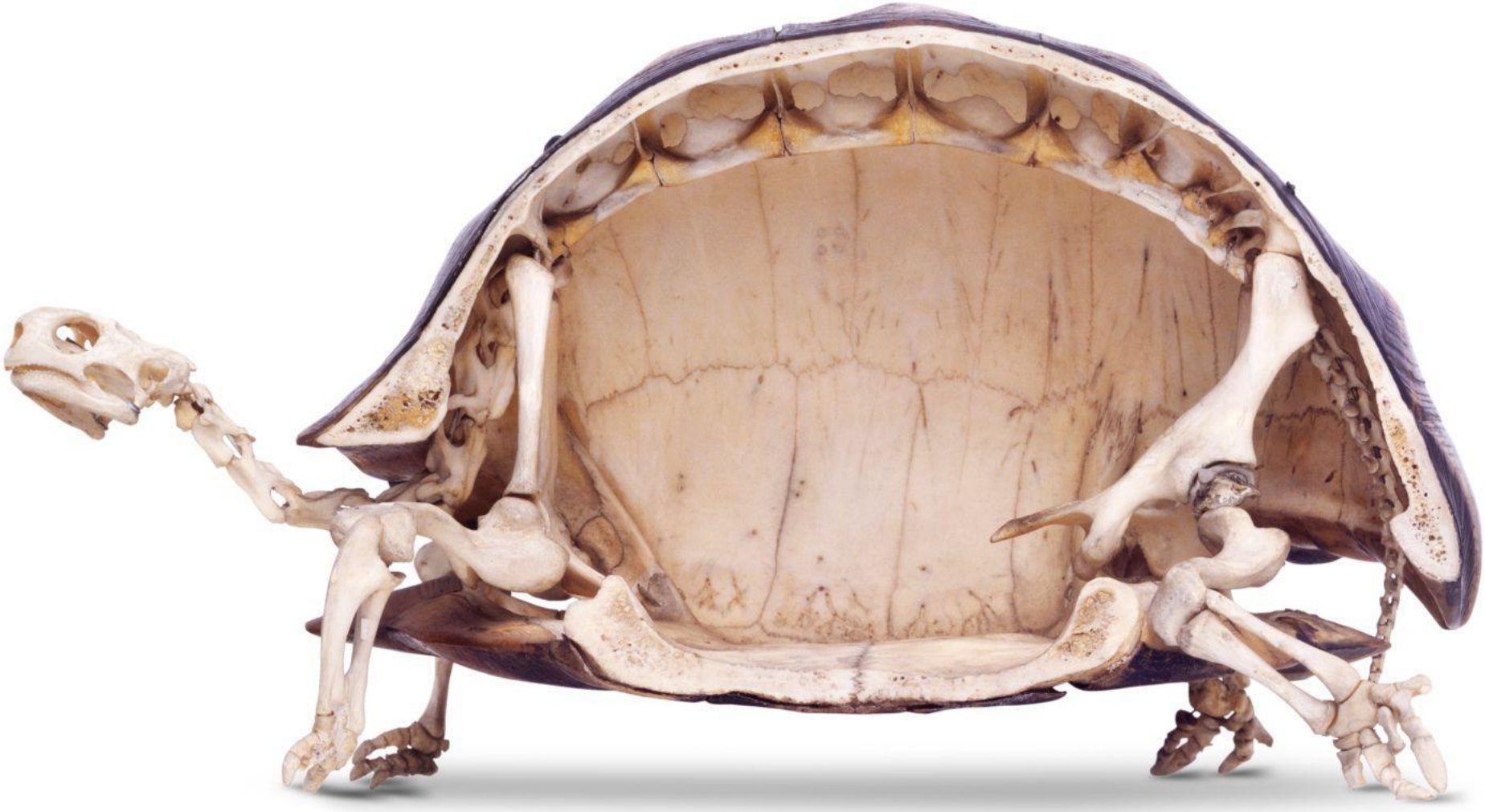
- From S16 course:
  - Author choice
  - Not innately visual, the author transformed it to be visual
  - Clearly (perfectly uncluttered) data
  - Defined metrics (science)
  - Highlight important aspects of the dataset
  - Intention/purpose?
  - Need more than one datapoint, need to show a trend, want to generalize? Do we need time? Maybe not.
    - Comparisons can be very valuable. Sufficient quantity of data to draw conclusions. Other similar datapoints, or datapoints for comparison.
  - Needs to be interactive! Want to move the cat around, virtual reality cat!

Criteria for label “(good) Visualization”

- From the F14 class…
  - Reveals something you didn’t know (about cats)
  - Needs to have an X & Y axis (not really?)
    instead… Needs to exist in a space
  - Should be clear in meaning & purpose
  - Just enough information and no more
  - Intentional
  - If the thing can be quantified, must indicate precision & accuracy
  - Shouldn’t be trying to mislead you
    but ok to have busy-ness to express the complexity
  - Be a scientist, have a hypothesis but look at the data with fresh eyes – don’t bias your conclusions, allow for interpretation

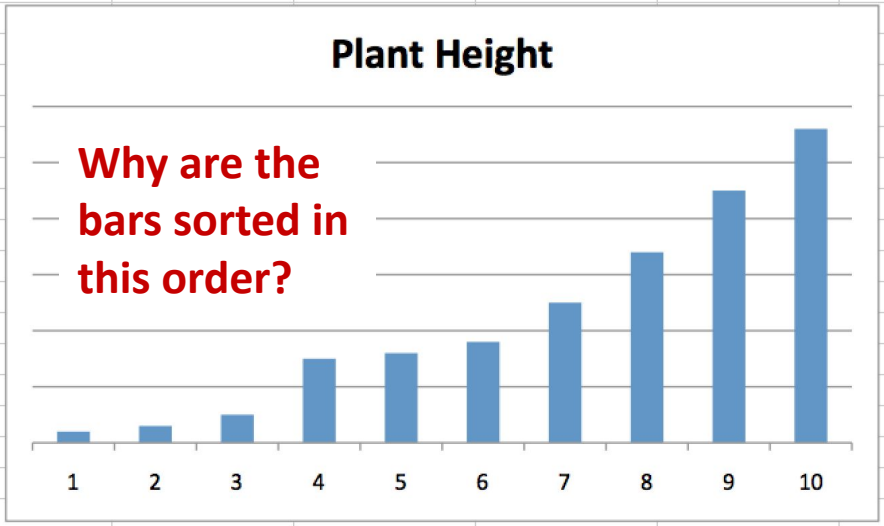
"This is why turtle cannot come out from its shell"

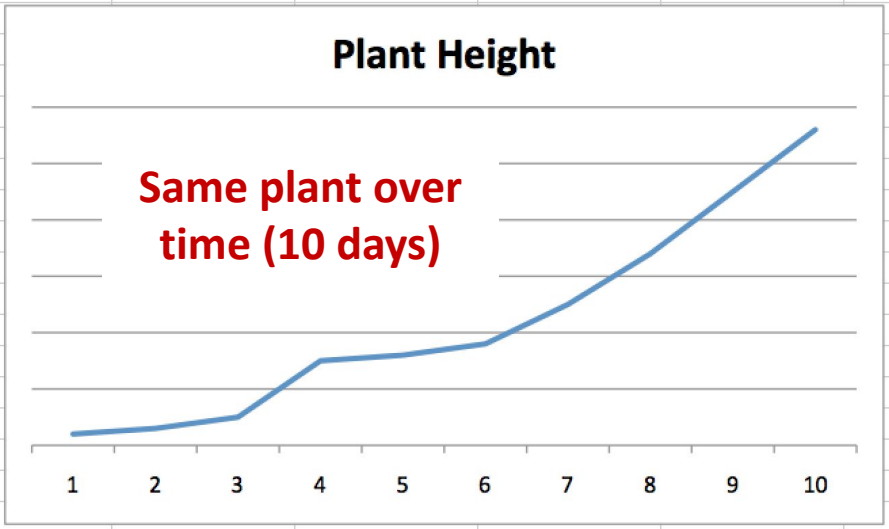
https://www.reddit.com/r/pics/comments/7srqkj/this_is_why_turtle_cannot_come_out_from_its_shell/

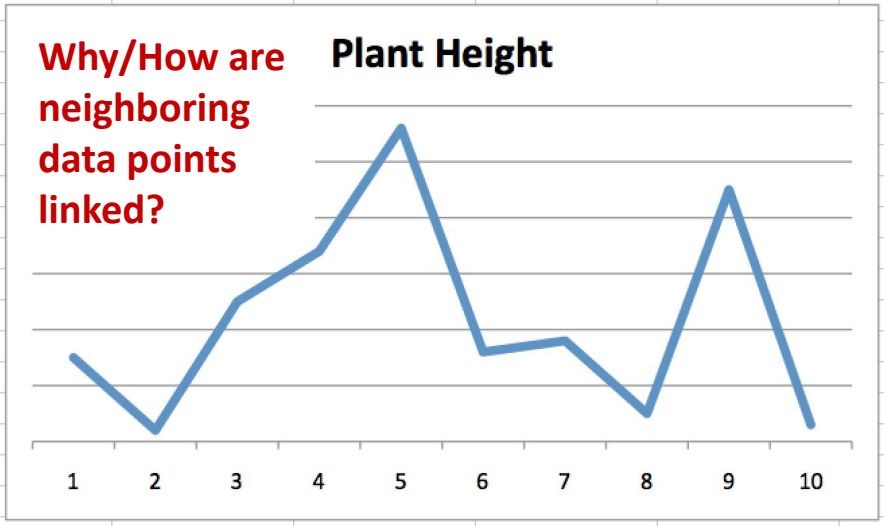


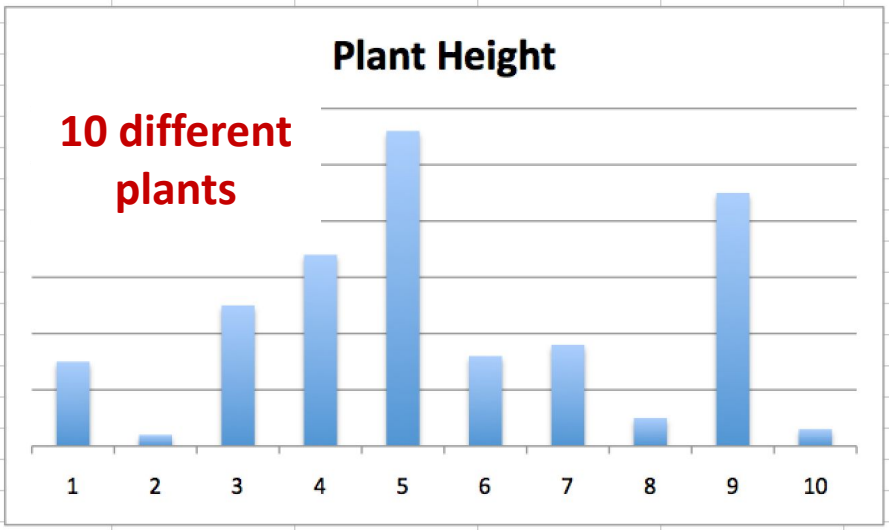
https://imgur.com/gallery/BuyO47o

What I learned in 7th grade Science Fair: Presenting Scientific Results









Variable Types

- **Categorical/Nominal:**
  - Has categories
  - No agreed upon ordering ordering
  - For example: hair color (red, brown, black, grey…)
- **Ordinal Data:**
  - Has categories
  - There *is* an agreed upon ordering
  - Spans may be unequal
  - Likert Scale: (strongly like, like, neutral, dislike, strongly dislike)
- **Interval:**
  - Has categories
  - There *is* an agreed upon ordering
  - Spans are equal
  - E.g., height

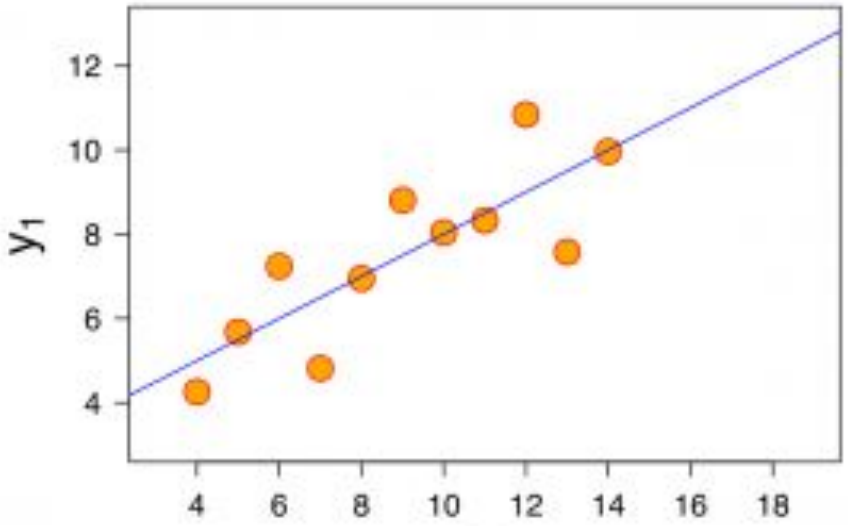
How not to… Ordinal Data

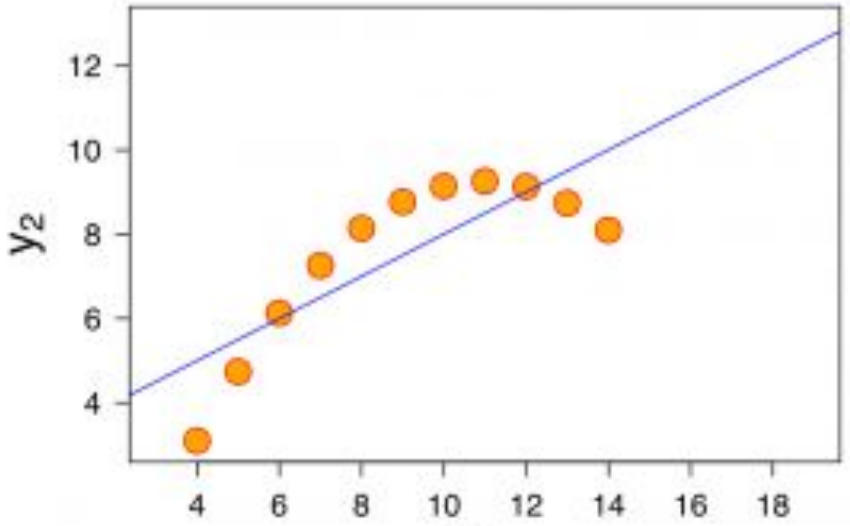
It's mid-April… how much $ is remaining in the travel budget?

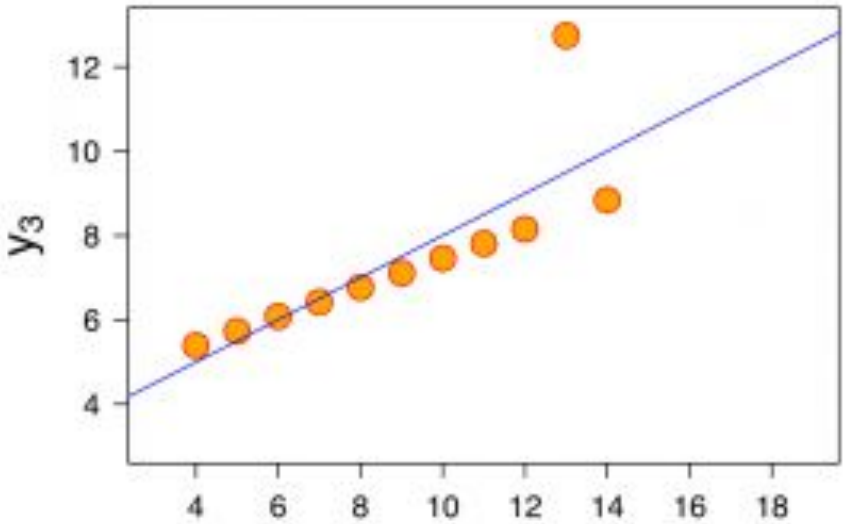
April '18	0
August '18	1200
December '18	1950
February '19	1950
January '19	1350
July '18	0
June '18	0
March '19	4950
May '18	0
November '18	1950
October '18	4450
September '18	1950
Expenditures	19750
Total Grant	25000

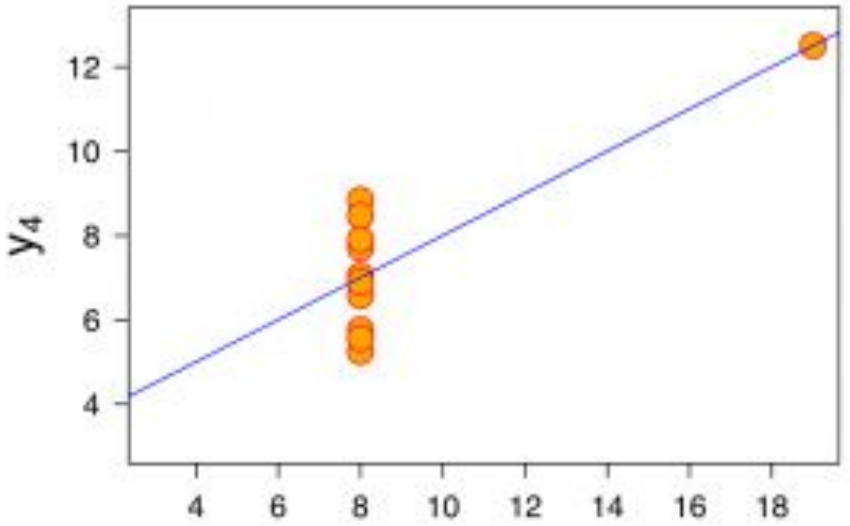
	student pay	travel	total
July '18	0		0
August '18	1200		1950
September '18	1950		1950
October '18	1950	2500	4450
November '18	1950		1950
December '18	1950		1200
Fall 2018 Total	9000	2500	11500
January '19	1350		1350
February '19	1950		1950
March '19	1950	3000	4950
April '19	?		
May '19	?		
June '19	?	?	
Spring 2019 Total	5250	3000	8250
Remaining Budget	3750	1500	5250
Total	18000	7000	25000

These 4 data sets while quite different. Coincidentally all have the same mean, variance, correlation, and regression.









Francis Anscombe http://en.wikipedia.org/wiki/File:Anscombe%27s_quartet_3.svg

Spark Lines – intense word sized graphics

- Term coined by Edward Tufte in book *Beautiful Evidence*

- Typical Data = word & number
  + Over time!
  + Quantified (last measurement)
  + Range of what's normal

Spark Lines – intense word sized graphics

- High resolution
- Integrated with prose
- Multiple spark lines compared to each other

	1999.1.1	65 months	2004.4.28	low	high
Euro foreign exchange $	1.1608		1.1907	.8252	1.2858
Euro foreign exchange ¥	121.32		130.17	89.30	140.31
Euro foreign exchange £	0.7111		0.6665	.5711	0.7235

Win/Loss over sports season

http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0001OR&topic_id=1