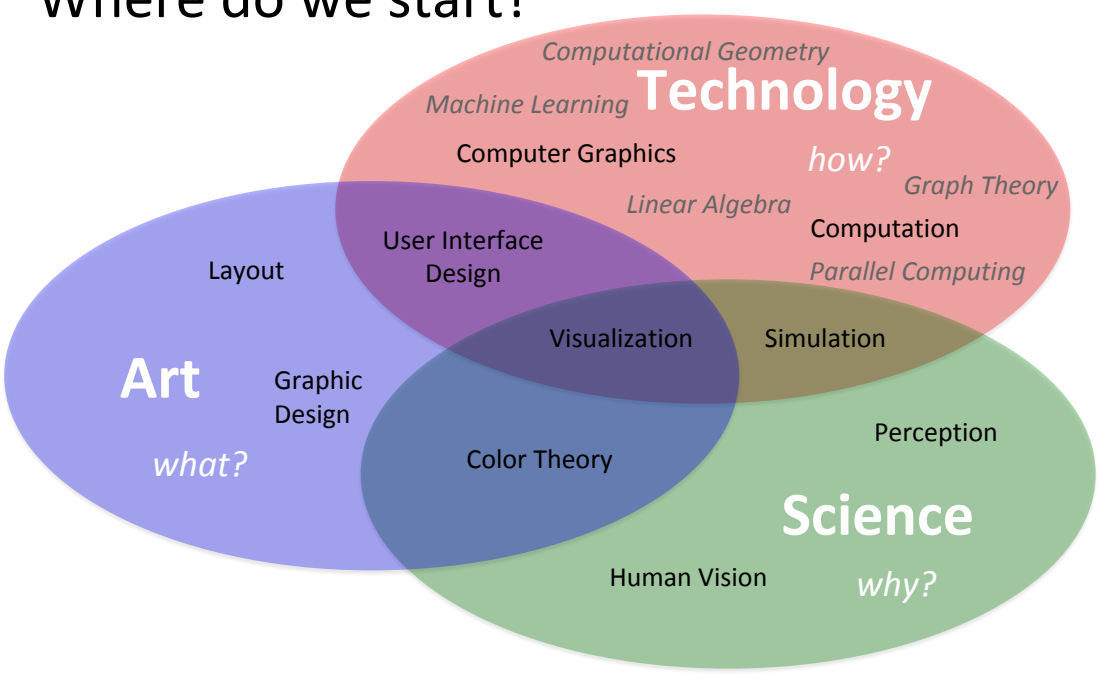
CSCI 4550/6550
Interactive Visualization
Spring 2018

http://www.cs.rpi.edu/~cutler/classes/visualization/S18/
Barb Cutler
cutler@cs.rpi.edu
Lally 302

"Introduction" to Visualization: Where do we start?

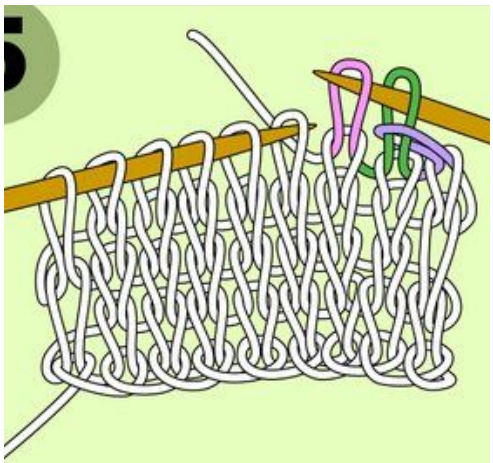


The Visualization Process

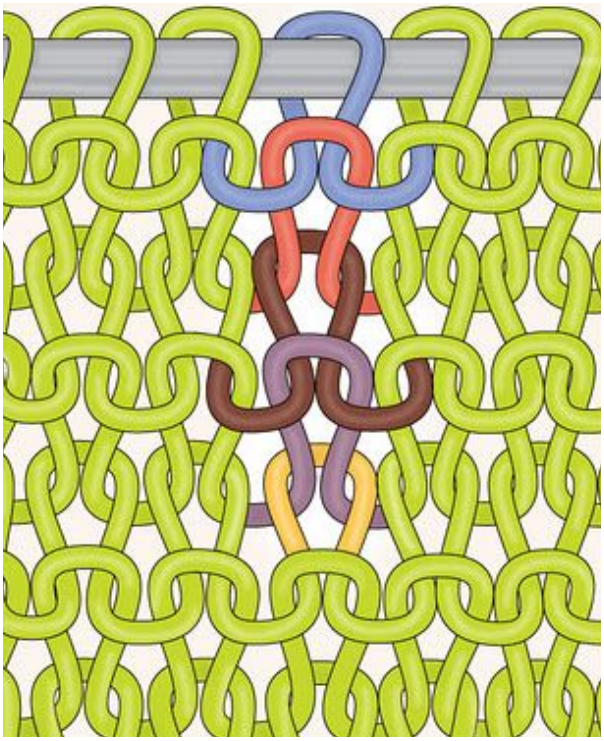
- Motivation & Problem Definition
- Visualization Design
- Data Collection
- Visualization Execution
- Analysis & Validation
- Visualization Revision
- Presentation

The Visualization Process

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



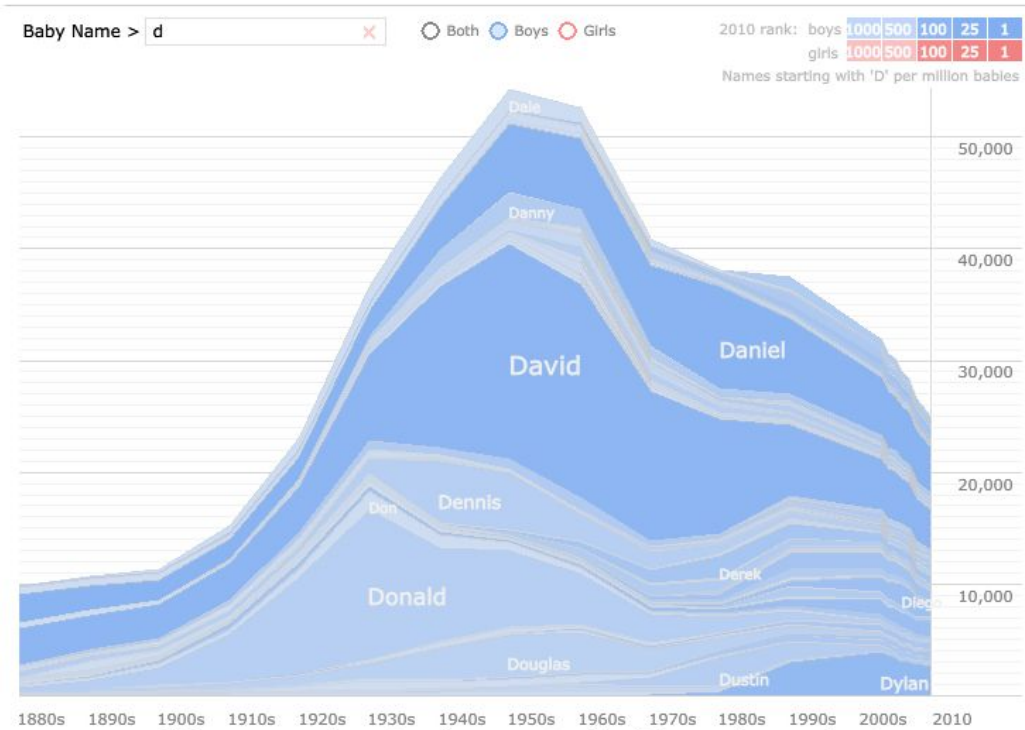
http://techknitting.blogspot.com/



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

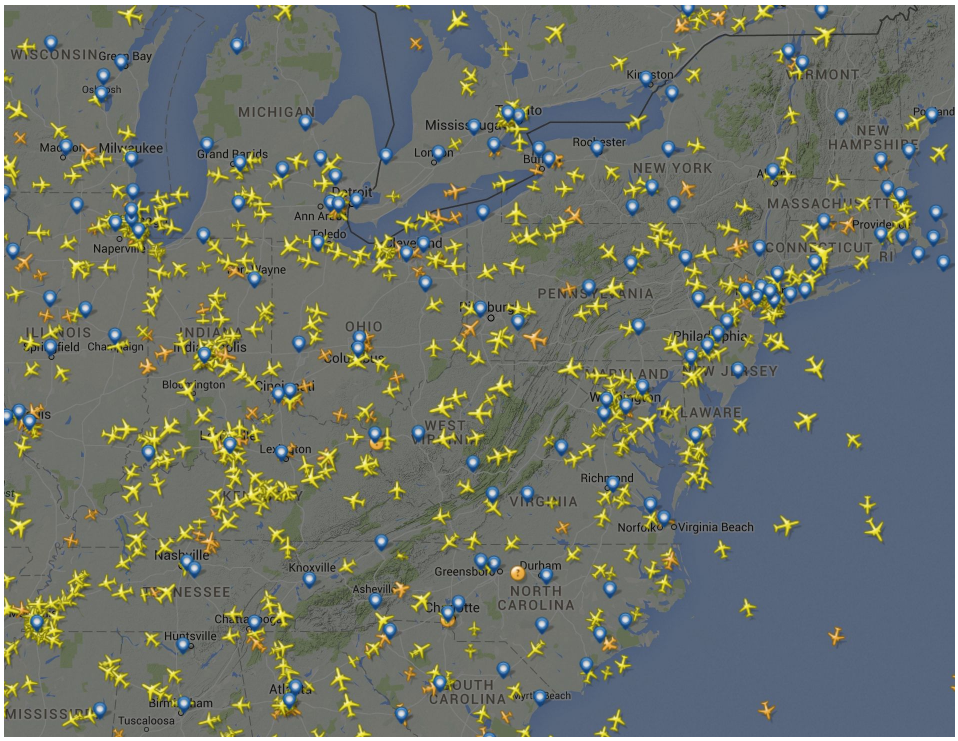
http://www.babynamewizard.com/



The Visualization Process

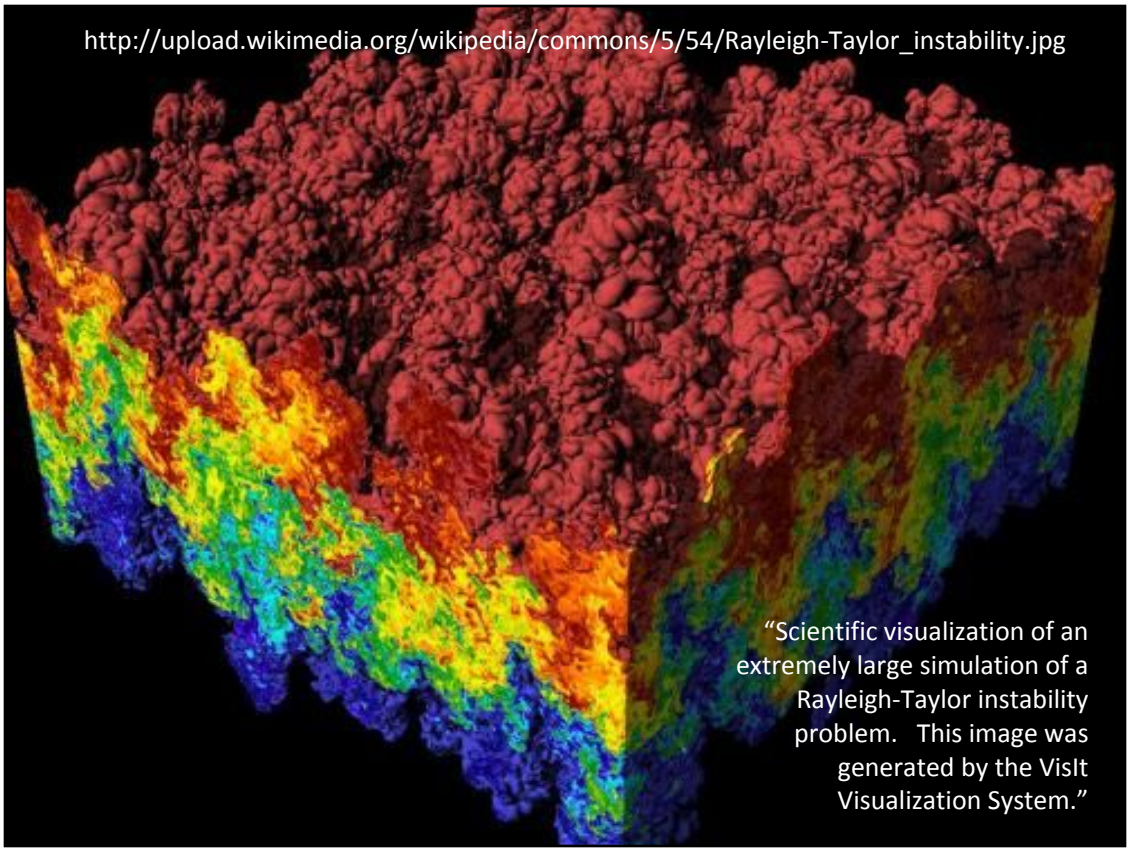
- Motivation & Problem Definition
- Visualization Design
- Data Collection
  - e.g., data structures, file formats, parsing, performance & efficiency, databases, very large datasets, interdisciplinary collaboration
- Visualization Execution
- Analysis & Validation
- Visualization Revision
- Presentation

www.flightradar24.com



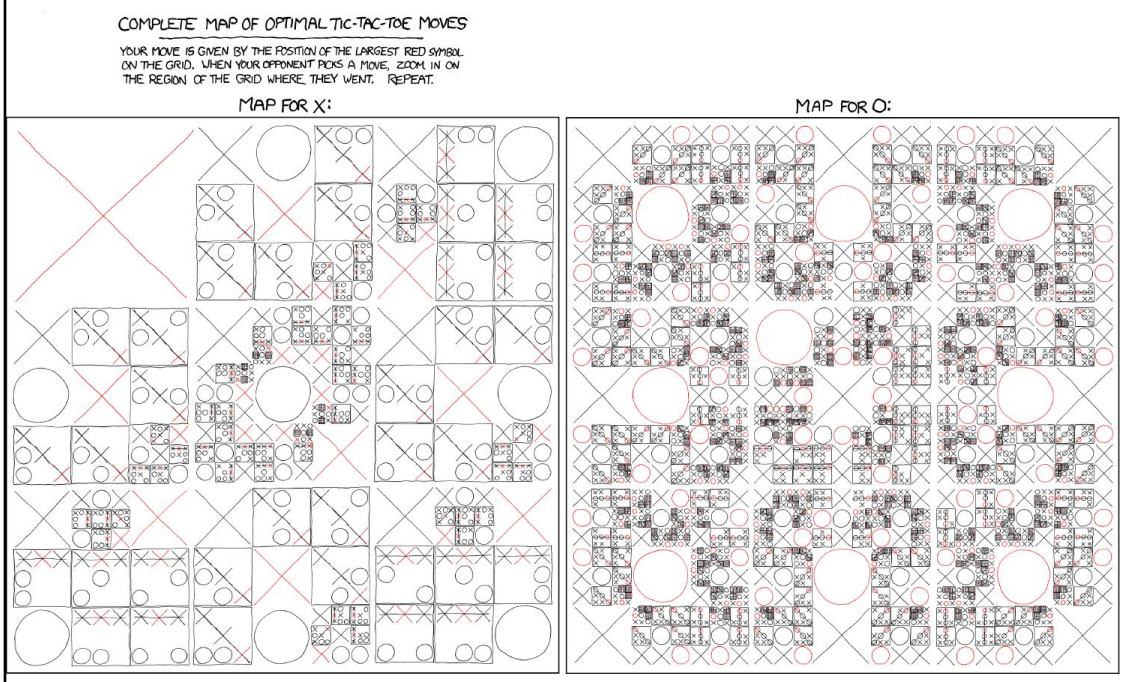
The Visualization Process

- 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



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



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

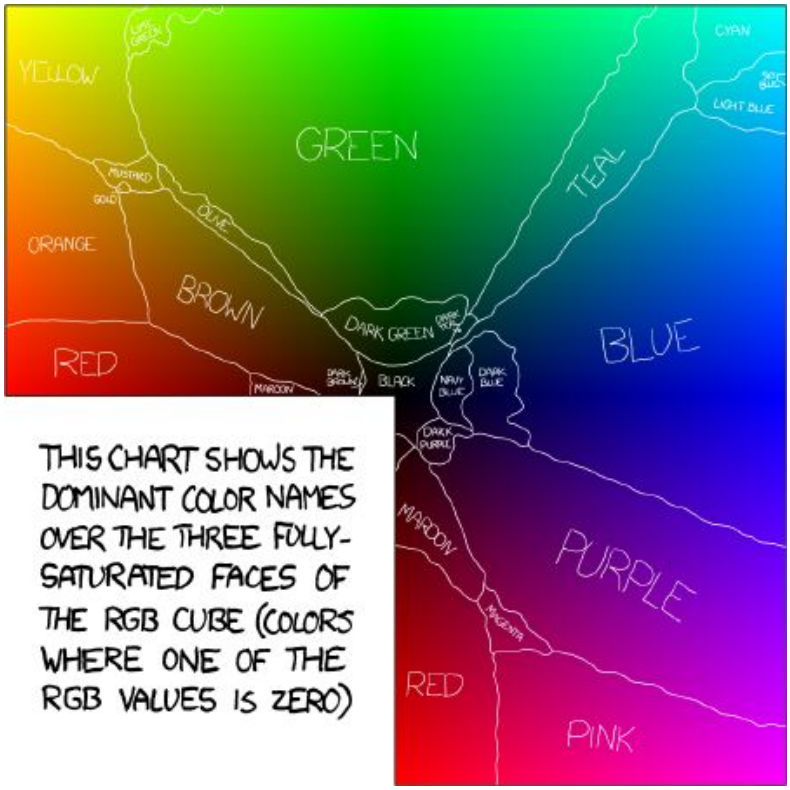
http://www.heraldsun.com.au/ Getty

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

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

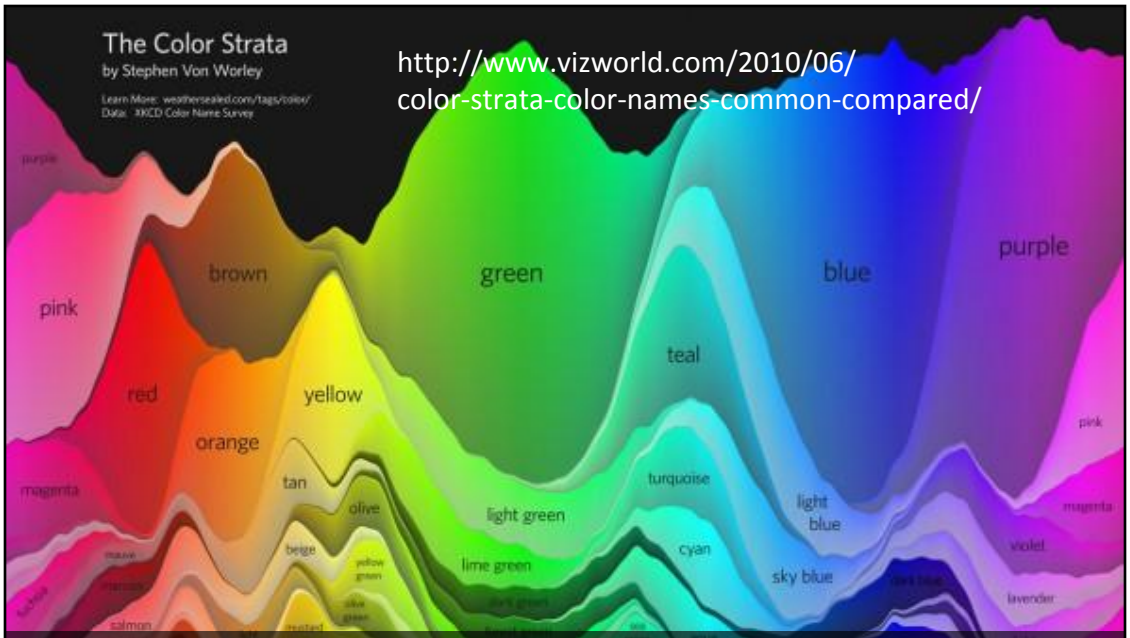
The Visualization Process

- 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



THIS CHART SHOWS THE DOMINANT COLOR NAMES OVER THE THREE FULLY-SATURATED FACES OF THE RGB CUBE (COLORS WHERE ONE OF THE RGB VALUES IS ZERO)

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



http://www.vizworld.com/2010/06/color-strata-color-names-common-compared/

"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
- Social Network Data Collection
- Class Website & Syllabus
- Reading for Friday
- Homework 1 for Thursday
- Criteria for a “good” Visualization
- Computational Geometry: Closest Pair of Points

Exercise: Social Network Data Collection

- Grab a sheet of paper & a pen/pencil.
- In a moment we will go around the room for introductions.
- Keep a running tally (count only, names not necessary) for each of the 6 categories below.

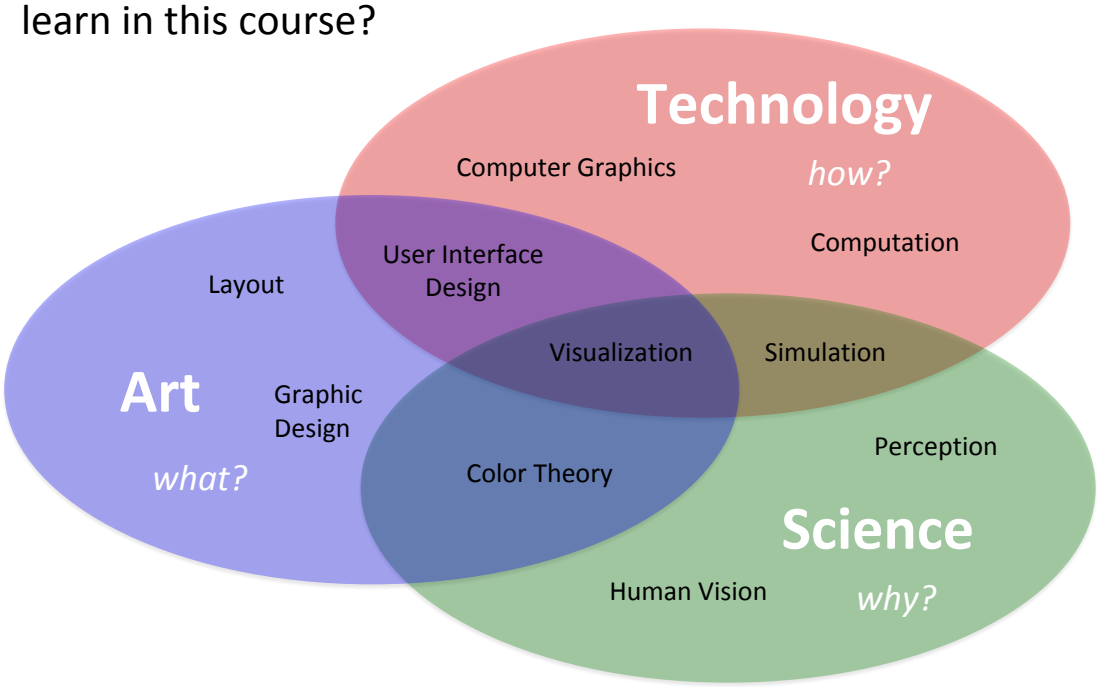
each other person in the class should be in exactly one category					
(choose the leftmost column if a person is in multiple categories)					
knew before RPI	share(d) dormroom or apartment	met in a social (non course) activity at RPI	met in CSCI 1200 Data Structures	met at RPI through a course	just met today!

Social Network Hypotheses

- But before we do that, make a few hypotheses about the results.
  - Your totals
  - The class average
  - Outliers: Do they tell an interesting story?

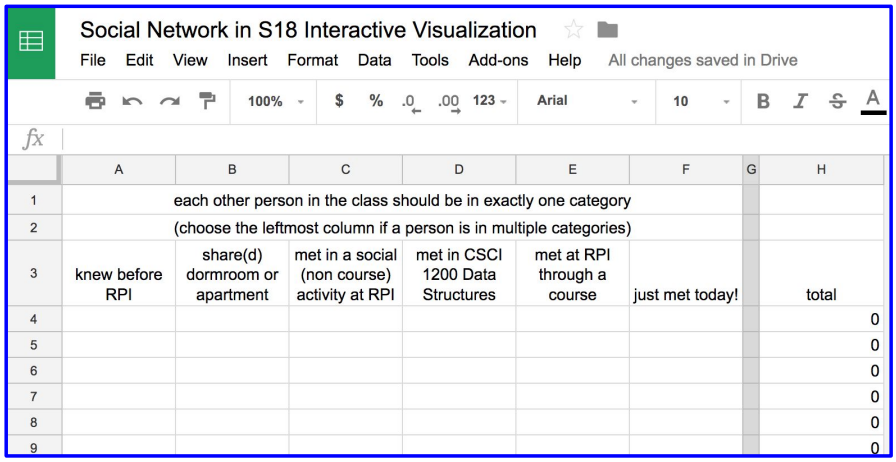
each other person in the class should be in exactly one category					
(choose the leftmost column if a person is in multiple categories)					
knew before RPI	share(d) dormroom or apartment	met in a social (non course) activity at RPI	met in CSCI 1200 Data Structures	met at RPI through a course	just met today!

And you? Major/Research Area? Skills & Strengths? Related hobbies? Career direction? What do you hope to learn in this course?



Exercise: Social Network Data Collection

- I'll share this google drive link on our Submitty forum
- Everyone will (anonymously*) enter their data as a row



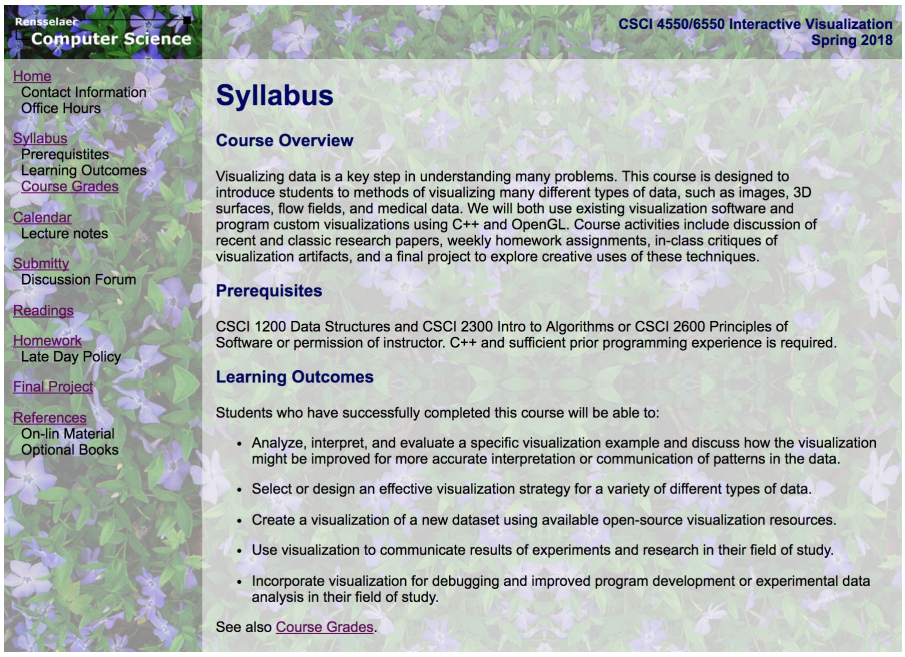
	A	B	C	D	E	F	G	H
1	each other person in the class should be in exactly one category							
2	(choose the leftmost column if a person is in multiple categories)							
3	knew before RPI	share(d) dormroom or apartment	met in a social (non course) activity at RPI	met in CSCI 1200 Data Structures	met at RPI through a course	just met today!		total
4								0
5								0
6								0
7								0
8								0
9								0

**Is anonymity guaranteed?*

Today

- Motivational Examples of Visualization Process
- Social Network Data Collection
- Class Website & Syllabus
- Reading for Friday
- Homework 1 for Thursday
- Criteria for a "good" Visualization
- Computational Geometry: Closest Pair of Points

Website, Syllabus, & Course Grades



Rensselaer Computer Science

CSCI 4550/6550 Interactive Visualization
Spring 2018

Home
Contact Information
Office Hours

Syllabus
Prerequisitites
Learning Outcomes
Course Grades

Calendar
Lecture notes

Submitty
Discussion Forum

Readings

Homework
Late Day Policy

Final Project

References
On-lin Material
Optional Books

Syllabus

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, 3D surfaces, flow fields, and medical data. We will both use existing visualization software and program custom visualizations using C++ and OpenGL. 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.

See also Course Grades.

http://www.cs.rpi.edu/~cutler/classes/visualization/S18/

"Rules" for the course

- As class participation is 10% of your grade:
  - *Using laptops during class is strongly discouraged*
    - If you're using your laptop you need to participate twice as much as everyone else because I'm going to assume you're doing something else.
    - If you are likely to be distracted by your laptop (email, homework, web-surfing, games), please close the lid
  - Use of laptops for reference during paper discussion is allowed
- Sit in a different seat, next to different people, each lecture
  - To facilitate mixing for feedback and brainstorming during in-class exercises

Reading for Friday:

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

- Everyone must post a non-trivial comment or question on the reading (~200 words) to the Submitty discussion forum by Friday @ 10:00am.

- Volunteer to lead the discussion (present ~5 minute summary of online discussion) during lecture on Friday?

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. We may also discuss via Submitty forum...

Today

- Motivational Examples of Visualization Process
- Social Network Data Collection
- Class Website & Syllabus
- Reading for Friday
- Homework 1 for Thursday
- Criteria for a “good” Visualization
- Computational Geometry: Closest Pair of Points

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”

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”

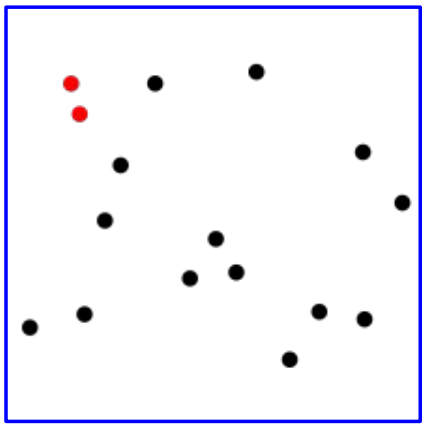
- From the F14 class…
  - Reveals something you didn’t know (about cats)
  - Needs to have an X & Y axis (not really?)
    instead… Needs to exist in a space
  - Should be clear in meaning & purpose
  - Just enough information and no more
  - Intentional
  - If the thing can be quantified, must indicate 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

Today

- Motivational Examples of Visualization Process
- Social Network Data Collection
- Class Website & Syllabus
- Reading for Friday
- Homework 1 for Thursday
- Criteria for a “good” Visualization
- Computational Geometry: Closest Pair of Points

Closest Pair of Points Problem

- Given *n* points,
  find the two points that have the smallest distance between each other.



- Applications?
  - collision detection
  - merging similar data points (data size reduction)
  - ...

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

Brute Force Algorithm

```
minDist = infinity
for i = 1 to length(P) - 1
 for j = i + 1 to length(P)
  let p = P[i], q = P[j]
  if dist(p, q) < minDist:
   minDist = dist(p, q)
   closestPair = (p, q)
return closestPair
```

Analysis? For *n* points? $O(n^2)$

Divide & Conquer

- Sort points by one of the axes
  - Find middle point,
  - Split points into two equal sized groups
  - & Recurse…
- When merging:
  - Let *d* be the smallest distance of the 2 halves
  - Consider points within *d* of the split point
    - Determine if any pairs spanning the split are closer than *d*
- Analysis? $O(n \log n)$ *need to show that only a small number of spanning pairs need to be checked*

Can we do better?

- Does it work?
  - In 2D? In 3D? Higher dimensions?
- Can we do better?
- What about dynamic data?
  - Applications?