“Introduction” to Visualization: Where do we start?

- Art
  - Graphic Design
  - User Interface Design
  - Layout

- Technology
  - Computational Geometry
  - Machine Learning
  - Computer Graphics
  - Linear Algebra
  - Computation
  - Parallel Computing
  - Graph Theory
  - Simulation

- Science
  - Perception
  - Human Vision
  - Color Theory

- what?

- why?

- how?
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
NameVoyager: Explore baby names and name trends letter by letter

Baby Name: B

Current rank:
- Boys: 1,000, 500, 100, 25, 1
- Girls: 1,000, 500, 100, 25, 1

Names starting with 'B' per million babies

- Barbara
  - Rank in 1940s: 2

Time periods:
- 1880s
- 1890s
- 1900s
- 1910s
- 1920s
- 1930s
- 1940s
- 1950s
- 1960s
- 1970s
- 1980s
- 1990s
- 2000s
- 2018
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 toolkits/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
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

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/
“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
• Reading for Today
• Reading for Friday
• Homework 1 for Thursday
• Criteria for a “good” Visualization
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/S20/
“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.
  – 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
Today

• Motivational Examples of Visualization Process
• Class Website & Syllabus
• **Reading for Today**
  – "Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message"
  – Spark Lines
• **Reading for Friday**
• Homework 1 for Thursday
• Criteria for a “good” Visualization
Reading for Today:

• "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.

• *Normally we’ll have a student discussant… but I’ll do today’s!*
"Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message", Stephen Few, Intelligent Enterprise, 2004

• Paper Summary / Contributions (5 mins max!)
  – Learn conventions/patterns, applies to almost all business data (not necessarily all scientific data)
  – Avoid viewer confusion / mis-information
  – Definition: Categorical vs. quantitative
  – Definition: Nominal (order free), ordinal (ordered), interval (e.g. histogram)
  – Available visual attributes for encoding data: location, size, shape, orientation, color
  
  choose wisely!!
<table>
<thead>
<tr>
<th>Type/Description</th>
<th>Encoding Methods</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal Comparison</td>
<td>• Bars only (horizontal or vertical)</td>
<td><img src="image" alt="Q1 2003 Calls by Region" /></td>
</tr>
<tr>
<td>A simple comparison of the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>categorical subdivisions of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>one or more measures in no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>particular order</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time Series</td>
<td>• Lines to emphasize overall pattern</td>
<td><img src="image" alt="2003 Sales" /></td>
</tr>
<tr>
<td>Multiple instances of one or</td>
<td>• Bars to emphasize individual values</td>
<td></td>
</tr>
<tr>
<td>more measures taken at equidistant points in time</td>
<td>• Points connected by lines to slightly emphasize individual values while still highlighting the overall pattern</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Always place time on the horizontal axis</td>
<td></td>
</tr>
<tr>
<td>Ranking</td>
<td>• Bars only (horizontal or vertical)</td>
<td><img src="image" alt="Headcount" /></td>
</tr>
<tr>
<td>Categorical subdivisions of a</td>
<td>• To highlight high values, sort in descending order</td>
<td></td>
</tr>
<tr>
<td>measure ordered by size (either</td>
<td>• To highlight low values, sort in ascending order</td>
<td></td>
</tr>
<tr>
<td>descending or ascending)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part-to-Whole</td>
<td>• Bars only (horizontal or vertical)</td>
<td><img src="image" alt="Regional % of Total Expenses" /></td>
</tr>
<tr>
<td>Measures of individual categorical</td>
<td>• Use stacked bars only when you must display measures of the whole as well as the parts</td>
<td></td>
</tr>
<tr>
<td>subdivisions as ratios to the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>whole</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

"Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message", Stephen Few, Intelligent Enterprise, 2004
Deviations
Categorical subdivisions of a measure compared to a reference measure, expressed as the differences between them
- Lines to emphasize the overall pattern only when displaying deviation and time-series relationships together
- Points connected by lines to slightly emphasize individual data points while also highlighting the overall pattern when displaying deviation and time-series relationships together
- Bars to emphasize individual values, but limit to vertical bars when a time-series relationship is included
- Always include a reference line to compare the measures of deviation against

Frequency Distribution
Counts of something per categorical subdivisions (intervals) of a quantitative range
- Vertical bars to emphasize individual values (called a histogram)
- Lines to emphasize the overall pattern (called a frequency polygon)

Correlation
Comparisons of two paired sets of measures to determine if as one set goes up the other set goes either up or down in a corresponding manner, and if so, how strongly
- Points and a trend line in the form of a scatter plot
- Bars may be used, arranged as a paired bar graph or a correlation bar graph, if scatter plots are unfamiliar
- (Note: For descriptions of these graphs, see my book Show Me the Numbers.)
“Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message”, Stephen Few, Intelligent Enterprise, 2004

• Key Points from Submitty Discussion (~15+ mins)
  – This is a nice reference / cheat sheet
  – Information was obvious / intuitive
  – These are all the charts you can make in Excel / Google Sheets
  – Are there similar rules for 3D, interaction, color, etc??
  – Including “bad” examples might have been educational
  – What about creativity? (Do we always have to follow the rules?)
  – (Un)intentional data misrepresentation... *Ethics!*
  – What about pie charts?
  – What about spatial/map data?
  – What about non-scientific data? “information visualization”
  – Writing: This is not a normal research paper (e.g., missing references to other papers)
Recommended Charts

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

Why/How are neighboring data points linked?

Why are the bars sorted in this order?

Same plant over time (10 days)

10 different plants
How not to... Ordinal Data

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

<table>
<thead>
<tr>
<th></th>
<th>student pay</th>
<th>travel</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>July '18</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>August '18</td>
<td>1200</td>
<td></td>
<td>1950</td>
</tr>
<tr>
<td>September '18</td>
<td>1950</td>
<td></td>
<td>1950</td>
</tr>
<tr>
<td>October '18</td>
<td>1950</td>
<td>2500</td>
<td>4450</td>
</tr>
<tr>
<td>November '18</td>
<td>1950</td>
<td></td>
<td>1950</td>
</tr>
<tr>
<td>December '18</td>
<td>1950</td>
<td></td>
<td>1200</td>
</tr>
<tr>
<td>Fall 2018 Total</td>
<td>9000</td>
<td>2500</td>
<td>11500</td>
</tr>
<tr>
<td>January '19</td>
<td>1350</td>
<td></td>
<td>1350</td>
</tr>
<tr>
<td>February '19</td>
<td>1950</td>
<td></td>
<td>1950</td>
</tr>
<tr>
<td>March '19</td>
<td>1950</td>
<td>3000</td>
<td>4950</td>
</tr>
<tr>
<td>April '19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May '19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>June '19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spring 2019 Total</td>
<td>5250</td>
<td>3000</td>
<td>8250</td>
</tr>
<tr>
<td>Remaining Budget</td>
<td>3750</td>
<td>1500</td>
<td>5250</td>
</tr>
<tr>
<td>Total</td>
<td>18000</td>
<td>7000</td>
<td>25000</td>
</tr>
</tbody>
</table>
These 4 data sets while quite different. Coincidentally all have the same mean, variance, correlation, and regression.
Spark Lines – intense word size 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 size graphics

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

Win/Loss over sports season

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

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

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

Article discussed here: http://eagereyes.org/criticism/chart-junk-considered-useful-after-all
Today

• Motivational Examples of Visualization Process
• Class Website & Syllabus
• Reading for Today
• Reading 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 Thursday
• Criteria for a “good” 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”

• 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,
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/