Lecture 1: Introduction to Visualization

“Introduction” to Visualization

- Art: what?
  - Graphic Design
  - User Interface Design
  - Visualization
- Technology: how?
  - Computer Graphics
  - Machine Learning
  - Computational Geometry
  - Linear Algebra
  - Graph Theory
- Science: why?
  - Perception
  - Human Vision
  - Simulation
- Graphic Design
  - Human Vision
  - Simulation
Today

- The Visualization Process
  - Motivation & Problem Definition
  - Visualization Design
  - Data Collection
  - Visualization Execution
  - Analysis & Validation
  - Visualization Revision
  - Presentation
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph… ”
Target audience is knitters not people who wear sweaters.

http://techknitting.blogspot.com/

Today

- The Visualization Process
  - Motivation & Problem Definition
  - Visualization Design
    - e.g., media, color, organization, layout,
    - static vs. dynamic, interactive, art & creativity!
  - Data Collection
  - Visualization Execution
  - Analysis & Validation
  - Visualization Revision
  - Presentation
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph… ”
Today

- 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
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph…”
Collecting the data can be a lot of work!
Today

● The Visualization Process
  ○ Motivation & Problem Definition
  ○ Visualization Design
  ○ Data Collection
  ○ Visualization Execution
e.g., data structures, implementation details, visualization toolkits (VTK, OpenGL, d3.js, etc.), performance & efficiency
  ○ Analysis & Validation
  ○ Visualization Revision
  ○ Presentation
● Website & Syllabus, Homework 1, etc.
● Reading: “ Eenie, Meenie, Minie, Moe: Selecting the Right Graph… ”

Learn about and how to use 3rd party tools!

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

- 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
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph… ”

Make sure the finished visualization is accurate!

http://xkcd.com/832/
Today

- 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
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph…”


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

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

- 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
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph…”

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?

Captions should give context and guide the viewer to conclusions can be drawn from the data.
Today

- The Visualization Process
  - Motivation & Problem Definition
  - Visualization Design
  - Data Collection
  - Visualization Execution
  - Analysis & Validation
  - Visualization Revision
  - Presentation
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph…”

Website, Syllabus, etc.

“Rules” for the course

- As class participation is 5% 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
  - Work with a different person for each in-class “worksheet”

Today

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

- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph…”
Paper for Today & Discussant

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

- Normally…
  - Everyone would have made a Submitty Forum post (~200 words) before 10am with a non-trivial *comment* or *question*
  - We’d have one or two student *discussants*… but I’ll do that today
  - **Summarize Paper (< 5 min)** everyone already read the paper!
  - **Summarize Discussion on Submitty Forum** & Moderate In-Class Discussion (~15 min)

"Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message", Stephen Few, 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 (what) vs. Quantitative (how much)
  - Definition: Nominal (order free), Ordinal (ordered), Interval (e.g. histogram)
  - Available visual attributes for encoding data: location, size, shape, orientation, color – *choose wisely*!!
    - Points (scatter plots) & lines (show trends) & bars (emphasize values for categorical data)
    - Size & color less effective for encoding quantitative values
<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="image1" alt="Graph" /></td>
</tr>
<tr>
<td>Time Series</td>
<td>• Lines to emphasize overall pattern</td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>• Bars to emphasize individual values</td>
<td></td>
</tr>
<tr>
<td></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="image3" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>• To highlight low values, sort in ascending order</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• To highlight low values, sort in descending order</td>
<td></td>
</tr>
<tr>
<td>Part-to-Whole</td>
<td>• Bars only (horizontal or vertical)</td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>• Use stacked bars only when you must display measures of the whole as well as the parts</td>
<td></td>
</tr>
</tbody>
</table>

**Deviation**
- Categorical subdivisions of 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, 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
  - Including “bad” examples might have been educational
  - What about creativity? (Do we always have to follow the rules?)
  - (Un)intentional data misrepresentation... *Ethics are important!*
  - What about pie charts? What about spatial/map data?
  - Are there similar rules for 3D, interaction, color, etc??
  - What about non-scientific data? “information visualization”
  - Writing / Communication: This is *not* a normal research paper (e.g., missing references to other papers)
What I learned in 7th grade Science Fair:

1. Always label your axes!
2. Choose the correct chart!

Confusion:
Why are the bars sorted in this order??

Confusion:
Why/How are neighboring data points linked?

Use Case:
Same plant over time (10 days)

Use Case:
10 different plants
How not to sort Ordinal Data

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

<table>
<thead>
<tr>
<th>Month</th>
<th>Expenditures</th>
<th>Total Grant</th>
</tr>
</thead>
<tbody>
<tr>
<td>April '18</td>
<td>19750</td>
<td>25000</td>
</tr>
<tr>
<td>August '18</td>
<td>1200</td>
<td>0</td>
</tr>
<tr>
<td>December '18</td>
<td>1950</td>
<td>0</td>
</tr>
<tr>
<td>February '19</td>
<td>1950</td>
<td>0</td>
</tr>
<tr>
<td>January '19</td>
<td>1350</td>
<td>0</td>
</tr>
<tr>
<td>July '18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>June '18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>March '19</td>
<td>4950</td>
<td>0</td>
</tr>
<tr>
<td>May '18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>November '18</td>
<td>1950</td>
<td>0</td>
</tr>
<tr>
<td>October '18</td>
<td>4450</td>
<td>0</td>
</tr>
<tr>
<td>September '18</td>
<td>1950</td>
<td>0</td>
</tr>
</tbody>
</table>

Proper chronological order is actually useful

<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>1200</td>
<td>1950</td>
</tr>
<tr>
<td>August '18</td>
<td>1950</td>
<td>0</td>
<td>1950</td>
</tr>
<tr>
<td>September '18</td>
<td>1950</td>
<td>0</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>0</td>
<td>1950</td>
</tr>
<tr>
<td>December '18</td>
<td>1950</td>
<td>0</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>0</td>
<td>1350</td>
</tr>
<tr>
<td>February '19</td>
<td>1950</td>
<td>0</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>

Confusion: Don’t sort months alphabetically!

These 4 data sets are quite different

But coincidentally all have the same:
- mean,
- variance,
- correlation,
- and
- regression

Sparklines – intense word-size graphics

- Term coined by Edward Tufte in book *Beautiful Evidence*
- About the size of a word
- Can be embedded within a paragraph of text

The Phillies 2017 season as a sparkline: 

![Sparkline example](image)

Compare that to the 2017 Washington Nationals season: 

![Sparkline example](image)

The one element missing from my implementation is an indicator for home games.

- Quantified
- Variation over time
- Highlight current/last measurement
- Range of what’s normal

Reading for Friday


*Post a comment/question on Submitty Discussion Forum by Friday @10am*

- Article discussed here: [http://eagereyes.org/criticism/chart-junk-considered-useful-after-all](http://eagereyes.org/criticism/chart-junk-considered-useful-after-all)
Homework 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 (full bibliographic citation), 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 Submitty discussion forum.

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!”

https://www.deviantart.com/theninthwavetnw/art/Upo-375998592
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, don’t show unnecessary context
• Not just an observation, should be an abstraction of the information
• Adds something, serves a purpose (augmented or cross section), a table of data isn’t a visualization, but a graph of data is
• Convey information by showing a 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 data to make comparison, help 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 Fall 2018

- 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

Criteria for label “(good) Visualization”  

From Spring 2016

- 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”

- 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

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

https://imgur.com/gallery/Buy047o