Visualization Design & Memorable Chart Junk
Reminder: “Rules” for the course

• Sit in a different seat, next to different people, each lecture

• Close your laptop/phone
  – Except to reference the paper we’re discussing
  – Participate!
  – Talk to me after class if you really need/want to use your laptop to take notes
Today

• “Good” Design (30 min)
  – Photography tips
  – Principles of Effective Website Design
  – Principles of Good User Interface Design
  – Examples of Good (Bad) Visualization Design

• Today’s Reading (30 min): “Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts”

• Worksheet: Time-based Data & Simple Charts (20 min)

• Graph Drawing Preview: Terminology & Goals (15 min)

• Readings for Tuesday: Graph Drawing
1. Get down on their level
2. Use a plain background
3. Use flash outdoors
4. Move in close
5. Take some vertical pictures
6. Lock the focus
7. Move it from the middle
8. Know your flash’s range
9. Watch the light
10. Be a picture director
“Canonical” Viewpoints

• From Dictionary.com:
  – authorized; recognized; accepted
  – the body of rules, principles, or standards accepted as axiomatic and universally binding in a field of study or art: the neoclassical canon
  – a fundamental principle or general rule: the canons of good behavior
  – a standard; criterion: the canons of taste

“What object attributes determine canonical views?” Blanz, Tarr, & Bulthoff, Perception 1999
Suppose you were making a brochure and you tried to give your customers the best possible impression of the objects shown on the static page. Which views would you choose?

“What object attributes determine canonical views?” Blanz, Tarr, & Bulthoff, Perception 1999
• Salience and significance of the features
• Stability of viewpoint to small transformations
• Minimize number of occluded features
• Familiarity, Functionality, Aesthetic criteria

“What object attributes determine canonical views?” Blanz, Tarr, & Bulthoff, Perception 1999
Rule of Thirds


– align subject with guide lines and intersection points, discourage placement of the subject at the center
– placing the horizon on the top or bottom line, avoid dividing picture in half
Today

• “Good” Design (30 min)
  – Photography tips
  – Principles of Effective Website Design
  – Principles of Good User Interface Design
  – Examples of Good (Bad) Visualization Design

• Today’s Reading (30 min): “Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts”

• Worksheet: Time-based Data & Simple Charts (20 min)

• Graph Drawing Preview: Terminology & Goals (15 min)

• Readings for Tuesday: Graph Drawing
Principles of Effective Website Design

• Guiding the eye (position, color, contrast, size, design elements)
• Spacing, padding, white/empty space, reduce cognitive load
• Navigation/orientation
• Typography (font, size, color, paragraphs)
• Usability/standards/conventions be obvious, “Don’t make users think”
• Consistency
• Alignment, polished, simplicity
• Effective writing
• Clarity, sharpness, contrast, exaggeration

http://psd.tutsplus.com/tutorials/designing-tutorials/9-essential-principles-for-good-web-design/
http://uxdesign.smashingmagazine.com/2008/01/31/10-principles-of-effective-web-design/
Principles of Good User Interface Design

- **Consistency and standards**
  - Match real world: words, phrases and concepts familiar to the user, real-world conventions, natural and logical order, coherency

- **Flexibility and efficiency of use:** cater/tailor to both inexperienced and experienced users
  - Know your user, user testing, listen to the user

- **User control and freedom:** clearly marked "emergency exit" to leave the unwanted, support undo and redo

- **Aesthetic and minimalist design:** every extra unit of information competes with and diminishes visibility of relevant information
  - System status: keep users informed

- **Recognize, diagnose, and recover from errors**
  - Error prevention: good error messages, eliminate error-prone conditions, confirmation option

- **Help and documentation**
  - Recognition rather than recall: information/instructions should be visible or easily retrievable

http://www.useit.com/papers/heuristic/heuristic_list.html
Today

• “Good” Design (30 min)
  – Photography tips
  – Principles of Effective Website Design
  – Principles of Good User Interface Design
  – Examples of Good (Bad) Visualization Design
• Today’s Reading (30 min): “Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts”
• Worksheet: Time-based Data & Simple Charts (20 min)
• Graph Drawing Preview: Terminology & Goals (15 min)
• Readings for Tuesday: Graph Drawing
Mushroom is the UK’s most liked pizza topping

Generally speaking, which of the following toppings do you like on a pizza? Select as many as you like

- 42% Sweetcorn
- 60% Peppers
- 49% Bacon
- 56% Pepperoni
- 26% Spinach
- 33% Olives
- 42% Pineapple
- 51% Tomato (as a topping)
- 61% Ham
- 65% Mushrooms

Other items not depicted include: onions (62%), chicken (56%), beef (36%), chillies (31%), jalapeños (30%), pork (25%), tuna (22%), anchovies (18%). 2% of people say they only like Margherita pizzas

February 26-28, 2017
How old are national flags?

This timeline is based on the date in which each nation adopted their current national flag and is sorted by a 20 year interval. As you can see only four of the current national flags are dated before the 19th century.
Wide partisan, demographic differences in views of U.S. decision to conduct airstrike that killed Soleimani

% who say U.S. decision to conduct the airstrike that killed Iranian Gen. Soleimani was the ...

<table>
<thead>
<tr>
<th></th>
<th>Wrong Decision</th>
<th>Right decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>43</td>
<td>48</td>
</tr>
<tr>
<td>Men</td>
<td>34</td>
<td>58</td>
</tr>
<tr>
<td>Women</td>
<td>52</td>
<td>37</td>
</tr>
<tr>
<td>White</td>
<td>33</td>
<td>58</td>
</tr>
<tr>
<td>Black</td>
<td>71</td>
<td>17</td>
</tr>
<tr>
<td>Hispanic</td>
<td>61</td>
<td>32</td>
</tr>
<tr>
<td>Ages 18-29</td>
<td>51</td>
<td>40</td>
</tr>
<tr>
<td>30-49</td>
<td>48</td>
<td>44</td>
</tr>
<tr>
<td>50-64</td>
<td>36</td>
<td>56</td>
</tr>
<tr>
<td>65+</td>
<td>37</td>
<td>52</td>
</tr>
<tr>
<td>Postgrad</td>
<td>56</td>
<td>31</td>
</tr>
<tr>
<td>College grad</td>
<td>48</td>
<td>47</td>
</tr>
<tr>
<td>Some college</td>
<td>38</td>
<td>51</td>
</tr>
<tr>
<td>HS or less</td>
<td>41</td>
<td>51</td>
</tr>
<tr>
<td>Rep/Lean Rep</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conserv</td>
<td>11</td>
<td>84</td>
</tr>
<tr>
<td>Mod/Lib</td>
<td>22</td>
<td>72</td>
</tr>
<tr>
<td>Dem/Lean Dem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cons/Mod</td>
<td>73</td>
<td>17</td>
</tr>
<tr>
<td>Liberal</td>
<td>81</td>
<td>10</td>
</tr>
</tbody>
</table>

Notes: Don’t know responses not shown. Whites and blacks include only those who are not Hispanic; Hispanics are of any race.

PEW RESEARCH CENTER

Unbiased reporting
Good data-ink ratio
Figure 2. Parent Response to “New vaccines are safe for my child(ren).”

8% Strongly Disagree/Disagree + 49% Neutral = 57%
43% Strongly Agree/Agree

Source: C.S. Mott Children’s Hospital National Poll on Children’s Health, March 2007

Over-simplication?
Manipulative visualization?
Confusing / unnecessary colors/gradient
Simple, intuitive, effective
Educational for all ages
Compare players within team
Compare players on different teams
Different team choices/strategies

http://cahaber.me/#/basketball
Effective non-pie shaped pie chart!
Are all datasets relevant to plot on a world map?
Allows us to clearly see significant variations in distributions of rivers AND use (overuse?) of technology to control nature
If you are admitted to a hospital, you have a 5% chance of contracting an HAI.

1.7 million people per year get an infection during a hospital stay.

9.4% of total inpatient costs are HAI-related.

More than ⅓ of HAIs affect people with Medicare or Medicaid.

Stereotypical Infographic

http://www3.gehealthcare.com/en/Products/Categories/Healthcare_IT/Quality_Management#tabs/tab1900328377C74CAC8AD7E8D4A2072591
Healthy Eating Plate

- Use healthy oils (like olive and canola oil) for cooking, on salad, and at the table. Limit butter. Avoid trans fat.
- Drink water, tea, or coffee (with little or no sugar). Limit milk/dairy (1-2 servings/day) and juice (1 small glass/day). Avoid sugary drinks.
- The more veggies – and the greater the variety – the better. Potatoes and French fries don’t count.
- Eat plenty of fruits of all colors.
- Eat a variety of whole grains (like whole-wheat bread, whole-grain pasta, and brown rice). Limit refined grains (like white rice and white bread).
- Choose fish, poultry, beans, and nuts; limit red meat and cheese; avoid bacon, cold cuts, and other processed meats.

Stay active!

© Harvard University

Harvard T.H. Chan School of Public Health
The Nutrition Source
www.hsph.harvard.edu/nutritionsource

Harvard Medical School
Harvard Health Publications
www.health.harvard.edu

Vast improvement on original food pyramid
• **Scientific Visualization (SciVis)**
  – really large quantities of data
  – data usually has inherent structure
  – often has a spatial and/or temporal component (coordinate system)
  – often appropriate to use of 3D visualization techniques
  – such as medical, hurricane, computational fluid dynamics (CFD) data

• **Information Visualization (InfoVis)**
  – smaller datasets
  – data that does not have an inherent structure (may not have coordinate system)
  – financial stock market data, demographic census data, genetic data, etc.

• **Visual Analytics**
  – involves a cycle of rapidly creating visualizations to answer questions and generate new questions about a dataset

• **Infographics** are typically in the realm of InfoVis, and often they show the results of the visual analytics process, but SciVis is not really a part of most infographics.

• Annual IEEE Visualization and IEEE InfoVis (Information Visualization) conferences are two separate entities. The set of people organizing, attending and involved one conference is almost disjoint from the other set.

An Interactive Visualization!
(interactive data collection)
Visualization Challenge: very large datasets
Visualization Challenge: representing different scales

### Radiation Dose Chart

This is a chart of the effects of ionizing radiation from various sources. The unit for absorbed dose is "sievert" (Sv), and measures the effects of radiation on the cells of the body. The amount (all at once) will make you sick, and too many x-rays will kill you, but we safely absorb small amounts of natural radiation daily. Note: the same unit of effective dose in a shorter time will generally cause more damage, but your cumulative lifetime dose does not gain in intensity like cancer risk.

#### Dose Levels

- **Total dose**: All doses in a single exposure (1 Sv)
- **One-year dose**: Highest dose to one organ or tissue (1 Sv)
- **Equivalent dose**: Dose equivalent per unit of radiation (1 Sv)
- **Effective dose**: Dose equivalent per unit of radiation (1 Sv)

#### Examples of Radiation Sources

- **Nuclear power plant** (1 Sv/year)
- **Cosmic rays** (1 Sv/year)
- **Medical X-rays** (1 Sv/year)
- **Medical CT scans** (1 Sv/year)
- **Nuclear weapons** (1 Sv/year)

#### Radiation Dose Comparison

- **Lowest radiation dose**: Lowest one-year dose, usually assessed in the context of medical procedures (1 mSv)
- **Highest radiation dose**: Highest one-year dose, usually associated with nuclear accidents (1 Sv)

#### Radiation Effects

- **Cell death**: Causes radiation sickness or death (10 Sv)
- **Immuno depressive effects**: Causes immune system depression (500 mSv)
- **Cancer risk**: Causes cancer (50 mSv)

#### Radiation Protection

- **Personal protective equipment**: Radiation protective clothing and shielding (10 mSv)
- **Medical protection**: Medical interventions to reduce radiation exposure (1 mSv)

#### Radiation Monitoring

- **Radiation detectors**: Monitor radiation levels in the environment (1 mSv)
- **Medical monitoring**: Monitor radiation levels in the body (1 mSv)

### Sources

- [xkcd.com/radiation/](https://xkcd.com/radiation/)

By Randall Munroe, with help from Ellen, Senior Reactor Operator at the Redstone Test Reactor, who suggested the idea and provided a list of sources. It’s safe to add these to your list of ‘must-reads’! It’s for general education only. If you’re doing radiation safety procedures on an internet PDU large and things go wrong, you have no clue how bad it is. Good luck!”
### Periodic Table of Elements

<table>
<thead>
<tr>
<th>Element</th>
<th>Atomic Number</th>
<th>Symbol</th>
<th>Mass Number</th>
<th>Period</th>
<th>Group</th>
<th>State of Matter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydrogen</td>
<td>1</td>
<td>H</td>
<td>1.00794</td>
<td>1</td>
<td>1</td>
<td>Gas</td>
</tr>
<tr>
<td>Helium</td>
<td>2</td>
<td>He</td>
<td>4.00302</td>
<td>1</td>
<td>18</td>
<td>Gas</td>
</tr>
<tr>
<td>Lithium</td>
<td>3</td>
<td>Li</td>
<td>6.941</td>
<td>2</td>
<td>1</td>
<td>Solid</td>
</tr>
<tr>
<td>Beryllium</td>
<td>4</td>
<td>Be</td>
<td>9.012182</td>
<td>2</td>
<td>2</td>
<td>Solid</td>
</tr>
<tr>
<td>Sodium</td>
<td>11</td>
<td>Na</td>
<td>22.98976926</td>
<td>3</td>
<td>1</td>
<td>Solid</td>
</tr>
<tr>
<td>Magnesium</td>
<td>12</td>
<td>Mg</td>
<td>24.305</td>
<td>3</td>
<td>2</td>
<td>Solid</td>
</tr>
<tr>
<td>Potassium</td>
<td>19</td>
<td>K</td>
<td>39.0983</td>
<td>4</td>
<td>1</td>
<td>Solid</td>
</tr>
<tr>
<td>Rubidium</td>
<td>37</td>
<td>Rb</td>
<td>85.467</td>
<td>5</td>
<td>1</td>
<td>Solid</td>
</tr>
<tr>
<td>Caesium</td>
<td>55</td>
<td>Cs</td>
<td>132.9055</td>
<td>6</td>
<td>1</td>
<td>Solid</td>
</tr>
<tr>
<td>Barium</td>
<td>56</td>
<td>Ba</td>
<td>137.327</td>
<td>6</td>
<td>2</td>
<td>Solid</td>
</tr>
<tr>
<td>Francium</td>
<td>87</td>
<td>Fr</td>
<td>223</td>
<td>7</td>
<td>7</td>
<td>Solid</td>
</tr>
</tbody>
</table>

For elements with no stable isotopes, the mass number of the isotope with the longest half-life is in parentheses.

---

Theodor Benfey, “periodic snail”, 1964

Better proximity for related elements (but more difficult to embed numerical data)
<table>
<thead>
<tr>
<th>chemical elements</th>
<th>us states</th>
</tr>
</thead>
<tbody>
<tr>
<td>H He Li Be B C N O F Na Mg Si P S Cl K Ti V Cr Fe Ni Cu Zn Ge As Se Br Kr Rb Sr Y Zr Nb Tc Ru Rn Pd Ag Cd Sn Sb Te I Xe Cs Ba Ce Pr Pm Sm Eu Gd Tb Dy Ho Er Tm Yb Lu Hf Ta W Re Os Ir Pt Au Hg Tl Pb Bi Po At Rn Fr Ra Ac Th U Np Pu Am Cm Bk Lv Uup Uut Cn Rg Ds Hs Bh Sg Db Rf Lr No Fm Es Cf Uus Uuo</td>
<td>Ak Az Ct De Ne Al Ar Ca Sc Mn Co Ga Mo In Pa La Nd Mt Md Fl Hi Id Il Ia Ks Ky Me Ma Mi Ms Nv Nh Nj Nm Ny Nc Oh Ok Or Ri Sd Tn Tx Ut Vt Va Wa Wv Wi Wy</td>
</tr>
</tbody>
</table>

chicanaspice
this information is so satisfying but idk what to do with it
Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812-1813.
Charles Joseph Minard, 1869.

The numbers of men present are represented by the widths of the colored zones at a rate of one millimeter for every ten thousand men; they are further written across the zones. The red designates the men who enter Russia, the black those who leave it. — The information which has served to draw up the map has been extracted from the works of M.M. Thiers, de Ségur, de Fezensac, de Chambray and the unpublished diary of Jacob, the pharmacist of the Army since October 28th.

In order to better judge with the eye the diminution of the army, I have assumed that the troops of Prince Jérôme and of Marshal Davout, who had been detached at Minsk and Mogilev and have rejoined near Orsha and Vitebsk, had always marched with the army.
These charts show movie character interactions. The horizontal axis is time. The vertical grouping of the lines indicates which characters are together at a given time.

Lord of the Rings

Star Wars (Original Trilogy)

Jurassic Park

12 Angry Men

Primer

http://imgs.xkcd.com/comics/movie_narrative_charts_large.png
Today

• “Good” Design (30 min)
  – Photography tips
  – Principles of Effective Website Design
  – Principles of Good User Interface Design
  – Examples of Good (Bad) Visualization Design

• Today’s Reading (30 min): “Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts”

• Worksheet: Time-based Data & Simple Charts (20 min)

• Graph Drawing Preview: Terminology & Goals (15 min)

• Readings for Tuesday: Graph Drawing
Reading for Today

• “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
What is “Chart Junk”?

• Extraneous elements in a chart or visualization
• Does not represent data
• Data-to-ink ratio (aim to convey more data with less ink)
• According to Edward Tufte: It’s not just unnecessary, it’s harmful (distracting)
• According to Nigel Holmes: Visualization should engage the reader’s interest
Study Design

• Compare embellished charts to plain ones
• Measured:
  – interpretation accuracy
    *was no worse for embellished charts*
  – long-term recall (2-3 weeks later)
    *was better for embellished charts, topic & details of the chart were more memorable*
• Prior work:
  – Higher data-to-ink ➔ faster response & greater accuracy
    [Gilan & Richman]
  – Other work shows a somewhat weak correlation between data-to-ink and interpretability or aesthetics
• Author’s caution:
  – Not an endorsement of chart junk
  – Embellishments can lead to bias!

“Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts” Bateman, Mandryk, Gutwin, Genest, McDine, & Brooks, CHI 2010
• Relatively small sample pool
  – 10 tested with ~5 min recall
  – 10 tested with 2-3 week recall
• Found no difference in time to read & describe embellished vs. non-embellished
• Participants preferred the embellished charts and found them more attractive
• Viewing time was unlimited for this study
  – Participants ended up spending the ~same amount of time on embellished vs. non-embellished
  – Effect of limiting time not measured

• Chart junk for these examples was tightly coupled with subject & details of chart
  – Quote from Holmes: “I think [Tufte] missed the point of much that I was trying to do: TIME magazine charts were aimed at lay readers, not unintelligent ones, but busy ones. I knew they’d get the point quicker if they were somehow attracted to the graphic.”

• What about charts from paper on last slide? What was their point? How good is your recall? Will your recall them in 2-3 weeks? Why didn’t the authors use embellishment?
Today

• “Good” Design (30 min)
  – Photography tips
  – Principles of Effective Website Design
  – Principles of Good User Interface Design
  – Examples of Good (Bad) Visualization Design

• Today’s Reading (30 min): “Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts”

• Worksheet: Time-based Data & Simple Charts (20 min)

• Graph Drawing Preview: Terminology & Goals (15 min)

• Readings for Tuesday: Graph Drawing
Pair Worksheet (~20 minutes)

• Meet new people
  – *Work with someone you did not know before this class*
  – *Work with a different partner every time*

• 1 worksheet per team of 2
  – *Sketch & brainstorm on the page*
  – *(I’ll scan & upload to Submitty for your review)*

• Use color! Be creative!
Homework Assignment 2: Time-Based Datasets

• Team of 2 or 3
• Obtain an interesting time-based dataset
  – Should be collectable* from online sources, and
  – Require a modest effort to prepare*
    * = you’ll submit your scripts/code to document
• Use Excel or Google Sheets
  – Create a variety (one of each?!) of the charts following the guidelines from "Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message”
  – Excellent labels and captions for each.
• Upload your assignment to Submitty by 11:59pm on Thursday. And post one of the images on the forum...
Tools for scraping data from the web

- copy-paste
- wget
- grep / sed / awk / sort / uniq
- Favorite programming language to parse/strip out unnecessary html formatting
- Save as .csv (comma separated value) files to upload to Excel / Google Sheets
- Python has lots of packages for parsing (e.g., json format)
- Selenium for automated browsing of websites
Today

• “Good” Design (30 min)
  – Photography tips
  – Principles of Effective Website Design
  – Principles of Good User Interface Design
  – Examples of Good (Bad) Visualization Design
• Today’s Reading (30 min): “Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts”
• Worksheet: Time-based Data & Simple Charts (20 min)
• Graph Drawing Preview: Terminology & Goals (15 min)
• Readings for Tuesday: Graph Drawing
Graph Terminology I

• Directed / Undirected Edges
• Tree (no cycles) vs. Graph (cycles allowed)
  – Cycle: A path along edges through the graph starting & ending at the same vertex.
  – Cycle Variants: closed walk, simple cycle, directed cycle, …
• Valence (a.k.a. Degree) of a Vertex: # of edges incident on the vertex
• Regular: Each vertex has same valence, a 3-regular graph is also called cubic
Graph Terminology II

- Polygon: 2D flat or on a sphere, with straight or great circle edges
- Polyhedron: 3D solid formed by flat faces
- Polytope: flat sides in any dimension
- Bipartite: vertices can be split into two groups, A & B. No edge connects a vertex in A to another vertex in A. Same for B.
- Clique – subset of vertices in an undirected graph with an edge connecting every pair of vertices in the subset.
Graph Terminology III

• Upward Drawing (of a tree) – no child is drawn with vertically above (higher y value than) its parent.
• Plane Graph – A 2D drawing of the graph where no edges cross (touching at the endpoint vertices they share is ok)
• Planar Graph – A graph for which a Plane Graph exists.
• Euler’s Theorem for planar graphs: For a plane graph with $n$ vertices, $m$ edges and $f$ faces, we have $n - m + f = 2$. 
Benjamin B. (from Spring ‘12)
Graph Drawing Goals

- Automated!
- Can read all of the labels
- Can follow the line and see exactly which 2 vertices it connects
- Aesthetically pleasing
- Layout should display as much symmetry as possible
- Crossing free or minimal-crossing layout
- All edge lengths are approximately equal
- Even vertex distribution
- Distance between nodes in final layout should be as close as possible to “graph distance” (# of edges on shortest path between those nodes)
Graph Drawing Questions

- What is the metric of success for each of our goals?
- Can we guarantee to find a solution? The optimal or best solution?
- Can we use randomness? Does it help?
- How expensive/slow are the different algorithms to draw graphs?
- How does it scale with more nodes/edges?
  - Does it lose effectiveness in meeting our goals?
  - How is the running time affected?
- How do we label the nodes/edges with color/words/images?
Today

• “Good” Design (30 min)
  – Photography tips
  – Principles of Effective Website Design
  – Principles of Good User Interface Design
  – Examples of Good (Bad) Visualization Design
• Today’s Reading (30 min): “Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts”
• Worksheet: Time-based Data & Simple Charts (20 min)
• Graph Drawing Preview: Terminology & Goals (15 min)
• Readings for Tuesday: Graph Drawing
Reading for Tuesday (pick one)

“Improved force-directed layouts”, Gansner and North, Graph Drawing, 1999.
Reading for Tuesday \textit{(pick one)}