Data Visualization (DSC 530/CIS 602-02)

Interactive Visualization Guidelines

Dr. David Koop
Light Reflection & Absorption

[Diagram showing light reflection and absorption on orange, apple, and strawberries with wavelength scale (400, 500, 600, 700 nm)]

[via M. Meyer]
Color != Wavelength

[Diagram showing the relationship between wavelength and energy density, with peaks for yellow and brown colors.]
Human Color Perception

• Humans **do not** detect individual wavelengths of light
• Use **rods** and **cones** to detect light
  - rods capture intensity
  - cones capture color

Opponent Process Theory

[Machado et. al, 2009]
Color Blindness

Normal

700  650  600  550  500  450  400

Protanopia

700  650  600  550  500  450  400

Deuteranopia

700  650  600  550  500  450  400

Tritanopia

700  650  600  550  500  450  400

[via M. Meyer]
Color Spaces, Gamut, and Model

• Color **space**: the organization of all colors in space
• Color **gamut**: a subset of colors
• A color **model** is a representation of color using some basis
  - RGB uses three numbers (red, blue, green) to represent color
  - Hue-Saturation-Lightness (HSL) is more intuitive and useful
Simultaneous Contrast
What colors?
What colors?

Red, yellow, blue

Purple, orange do not exist!

[A. Kitaoka]
Colormap

- A colormap specifies a mapping between colors and data values
- Colormap should follow the expressiveness principle
- Types of colormaps:

  **Binary**
  - y
  - n

  **Diverging**
  - -1
  - 0
  - +1

  **Categorical**
  - T
  - F
  - A

  **Sequential**
  - 3
  - 2
  - 1
Categorical vs. Ordered

- Hue has no implicit ordering: use for categorical data
- Saturation and luminance do: use for ordered data

[Luminance]

[Saturation]

[Hue]

[Munzner (ill. Maguire), 2014]
Categorical Colormaps

[link to colorbrewer2.org]
Continuous Sequential Colormap

US EPA Regional Oxidant Model -- Midwest Ozone (ppbv): June 26, 1987, 18:00

[Bergman et al., 1995]
Segmented Sequential Colormap

US EPA Regional Oxidant Model -- Midwest
Ozone (ppbv): June 26, 1987, 18:00

[Bergman et al., 1995]
Project Proposal

• Due **Tomorrow**
• Identify dataset
  - Potential data sources:
    https://github.com/caesar0301/awesome-public-datasets
• Understand domain
• Decide on tasks
  - Does this require a visualization, or could I write code for it?
• Start brainstorming on visualization and interaction design
  - **Do not** create three separate visualizations
  - You will create **one** visualization, but it may have **multiple views**
    that are tied together via **interaction**
Assignment 3

- Soccer data
  - Draw two choropleth maps
  - Use the same function for both!
  - Draw a teammate graph using force-directed layout
  - Use d3.queue to load data, code provided
Not only does the rainbow color map confuse viewers through its lack of perceptual ordering and obscure data through its inability to present small details, but it actively misleads the viewer by introducing artifacts to the visualization. The rainbow color map appears as if it's separated into bands of almost constant hue, with sharp transitions between hues. Viewers perceive these sharp transitions as sharp transitions in the data, even when this is not the case (see Figure 3). When combined with the lack of perceptual ordering, viewers face a daunting task when trying to correctly interpret the data via the rainbow color map. The goal of visualization is to present data so that viewers can quickly and accurately learn about the underlying data. The rainbow color map does a great deal to hinder this learning process by introducing confusing artifacts in some locations and reducing detail in others.

Prevalence of the rainbow color map

Although researchers have well documented these deficiencies, the visualization community still widely uses the rainbow color map. We present the findings of two surveys illustrating this prevalence. The first survey looks at papers in the IEEE Visualization Conference proceedings; the second considers visualization toolkits. IEEE Visualization proceedings

We searched the IEEE Visualization conference proceedings from 2001 through 2005 for papers that displayed data using a pseudocolor map. We included visualizations in which the rainbow color map was applied to surfaces, such as isosurfaces and streamlines. We excluded volume renderings as the literature does not address the relative merits of the rainbow color map when used for a color transfer function (although it seems clear that the same objections would apply). We did not count visualizations that used a banded version of the rainbow color map because explicit banding can be a useful visualization technique. We only included scalar data visualizations, excluding techniques such as mapping vector components to RGB—which is common with diffusion tensor MRI images. Such visualizations can appear at first glance to use the rainbow color map, but they are in fact using a different technique (see Rheingans for a discussion of the hazards of encoding multiple values into a pseudocolor map).

Results

Table 1 (next page) presents statistics from the 2001 through 2005 IEEE Visualization Conference proceedings. The table gives percentages of papers implementing pseudocoloring to display data using the rainbow color map. We've included all papers that include at least one use of the rainbow color map. The results are alarming:

- Each year between 40 and 59 percent of all papers using pseudocoloring used a rainbow color map.

[Ordering Color?]

[Borland & Taylor, 2007]
Rainbow Colormap

[Bergman et al., 1995]
Artifacts from Rainbow Colormaps

[Borland & Taylor, 2007]
Two-Hue Colormap

[Bergman et al., 1995]
Rainbow Colormap

Obama & Romney Tweets (2012 Election)

[A. C. Robinson, 2012]
Single-Hue Colormap

Obama & Romney Tweets (2012 Election)

[A. C. Robinson, 2012]
Isoluminant Rainbow Colormap

Original

Isoluminant

[Kindlmann et al., 2002]
d3-scale-chromatic

- https://github.com/d3/d3-scale-chromatic
- D3's built-in color scales
- Derived from ColorBrewer
- Sequential and diverging scales created using interpolation
"Get It Right in Black and White" - M. Stone

Matlab jet colormap

[S. Eddins (Matlab Blog), 2014]
"Get It Right in Black and White" - M. Stone

Matlab jet colormap (B&W)

[S. Eddins (Matlab Blog), 2014]
"Get It Right in Black and White" - M. Stone

Matlab parula colormap

[S. Eddins (Matlab Blog), 2014]
"Get It Right in Black and White" - M. Stone

Matlab parula colormap (B&W)

[S. Eddins (Matlab Blog), 2014]
Bivariate Colormaps

[Muñzner (ill. Maguire), 2014]
Guidelines for Visualization Design
Tufte: "The da Vinci of Data" —NYTimes

[https://www.edwardtufte.com/tufte/, 2017]
Bad: Data magnitude $\not\leftrightarrow$ Mark magnitude
Good: Data magnitude <=> Mark magnitude

[Flowing Data, 2012]
Starting Scales at Zero?
Cherry-picking data

[Fox News via Media Matters, 2012]
Show all the data

[AAA via Media Matters, 2012]
Tufte's Lie Factor

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.

Fuel Economy Standards for Autos
Set by Congress and supplemented by the Transportation Department. In miles per gallon.

This line, representing 27.5 miles per gallon in 1985.

[NYTimes via Tufte, 1991]
Tufte's Lie Factor

- Size of effect = (2nd value - 1st value) / (1st value)
- Lie factor = (size of effect in graphic) / (size of effect in data)
- In the graphic:

\[
\text{Lie Factor} = \frac{5.3 - 0.6}{0.6} = 14.8
\]
(Some of) Tufte's Integrity Principles

- Show data variation, not design variation
- Clear, detailed, and thorough labeling and appropriate scales
- Size of the graphic effect should be directly proportional to the numerical quantities ("lie factor")
Avoid Chartjunk
Avoid Chartjunk
Avoid Chartjunk
Avoid Chartjunk?

Memorable

Low Quality Description

Forgettable

[M. Borkin et al., InfoVis 2015]
Data-to-Ink Ratio (Also Unjustified 3D)
Maximize Data-to-Ink Ratio

![Bar chart showing the distribution of males and females across income levels: 0-$24,999, $25,000+, 0-$24,999, and $25,000+.](image)

[via A. Lex]
No Unjustified 3D
No Unjustified 3D

- Occlusion hides information
- Perspective distortion dangers
- Tilted text isn't legible

- Can help with shape perception
Eyes Beat Memory

• Reduce cognitive load (using up working memory)
• Animation versus side-by-side views
• Change blindness
Multiple Views

- **Juxtapose and Coordinate Multiple Side-by-Side Views**

  - **Share Encoding: Same/Different**
    - *Linked Highlighting*

  - **Share Data: All/Subset/None**

  - **Share Navigation**

[Munzner (ill. Maguire), 2014]
Multiform

[Improvise, Weaver, 2004]
Small Multiples

- Same encoding, but different data in each view (e.g. SPLOM)

[http://bl.ocks.org/mbostock/4063663]
Brushing

[http://bl.ocks.org/mbostock/4063663]
The InfoVis Mantra

- [Schneiderman, 1996]
- Overview First
- Zoom and Filter
- Details on Demand
Overview-Detail View
Aggregation: Boxplots

• **Show distribution**
• Single value (e.g. mean, max, min, quartiles) doesn't convey everything
• Created by John Tukey who grew up in New Bedford!
• **Show spread and skew** of data
• Best for **unimodal** data
• Variations like vase plot for multimodal data
• Aggregation here involves many different marks
Boxplot Example

(a) Overall Activity
(b) Structural Activity
(c) Parameter Activity
(d) Layout Activity

[L. Lins et al., 2008]
Responsiveness Required

- Delays are perceived by users
- Visual feedback
  - Show the user they did something (highlighting, etc)
  - Interaction should happen quick!
- Latency: mouse click versus mouse hover
- Popup versus detail displays
Function First, Form Next