Information Visualization
PERCEPTION and COLOR

Tobias Isenberg
tobias.isenberg@inria.fr
Recap

In Lecture 1 you learned about the basic components of visualization:

- marks and visual variables
Summary

- You know the main building blocks are **marks**
- Marks are modified by **visual variables**
- Visual variables have **specific characteristics**
- These influence how the data will be **perceived**

Today you will

- Learn details about the **perception of color** and a few **other visual encodings**
- See that the vision system is **quicker and better** at detecting certain visual encodings
WHAT IS COLOR?
Let’s do an experiment ...
Let’s do an experiment ...
What is Color?

- color is a **human reaction** to light (change)
What is Color?

Physical World
- Lights, surfaces, objects

Visual System
- Eye, optic nerve, visual cortex

Mental Models
- Red, green, brown
- Bright, light, dark, vivid, colorful, dull

Color Models
- RGB, CMYK, CIE XYZ, CIE Lab
- HSV/HSB, ...

“Yellow”
Physical World

Light is radiation in a range of wavelengths: 370–730nm

Light of a single wavelength is *monochromatic*
Most colors are not monochromatic
You do not see the spectrum of light

- Eyes make limited measurements
- Eyes physically adapt to circumstance
- You brain adapts in various ways
- Weird stuff happens
Rods
No color (sort of)
All over the retina
More sensitive

Cones
Three different kinds of “color receptors”
Mostly in the center
Less Sensitive

Simple Anatomy of the Retina, Helga Kolb
Cone response

LMS (Long, Middle, Short) cones
Sensitive to different wavelengths

A Field Guide to Digital Color, Maureen Stone
Cone response

Input Stimulus

Cone Response Curves

Product \rightarrow Response

Integrate

A Field Guide to Digital Color, Maureen Stone
Visual System ➔ Color Models

Product ➔ Response
Integrate

Graph showing different curves and a bar chart with labels S, M, L.

Question mark on a yellow background.
Two Principles of Color Perception

• **trichromacy**: representation of all spectral distributions possible with **three values** without information loss (w.r.t. the visual system) → essential for CS!

• **metamerism**: different spectra exist that produce the same trichromatic response
XYZ Color Model

• definition of three primary colors: X, Y, Z
  – color-matching functions are non-negative
  – Y follows the standard human response to luminance, i.e., the Y value represents perceived brightness
  – can represent all perceivable colors

• mathematically derived from color matching experiments

Stone 2005
XYZ CIE Color Space

- plotting XYZ space in 3D
- all colors that are perceivable by humans form a deformed cone
- $X$, $Y$, and $Z$-axes are outside this cone

Foley et al. 1990
CIE Chromaticity Diagram

- projection of XYZ space onto $X+Y+Z = 1$ (to factor out a color’s brightness):
  $$x = X/(X+Y+Z) \quad y = Y/(X+Y+Z)$$

- monochromatic colors on upper edge

Foley et al. 1990
Color Gamut

- color gamut: the area of colors in the CIE chromaticity diagram that can be created by adding together colors from the base colors
- if two colors are added, resulting color lies on straight line between them
- RGB shape: triangle

http://www.techmind.org/
Other Color Models: RGB & CMYK

- (physical) color mixing depends on color production process
  - light emission: additive mixing (CRTs etc.): **RGB model**
  - light absorption: subtractive mixing (printing process): **CMY(K) model**
RGB and XYZ

- RGB to XYZ conversion
  
  ![RGB intensity values](image1)
  ![XYZ tristimulus values](image2)
  ![CIE XYZ to CIE xy](image3)

  - RGB space: distorted cube
  - black: origin of XYZ and projection center
  - RGB projected to triangle

  ![Stone 2005](image4)
  ![Foley et al. 1990](image5)

  \[
  x = \frac{X}{X+Y+Z} \\
  y = \frac{Y}{X+Y+Z}
  \]

  http://www.techmind.org/
Can RGB Represent All Visible Colors?

- no, because all colors form horseshoe shape in CIE chromaticity diagram and RGB gamut is triangular

- But my shiny new 30” LCD is state-of-the-art, it can surely show all colors!

  → Let’s see a color that it cannot show …
Let’s see REAL cyan …
Let’s see REAL cyan ...
Let’s see REAL cyan ...
Visual System ➔ Color Perception

Edward H. Adelson

Slide adapted from Stone & Zellweger
Visual System ➔ Color Perception

Edward H. Adelson
Slide adapted from Stone & Zellweger
Visual System ➔ Color Perception

Josef Albers
Simultaneous Contrast
Simultaneous Contrast

Josef Albers
Simultaneous Contrast
Bezold Effect

[Image of two green patterns with geometric shapes]
Crispening

Perceived difference depends on background

From Fairchild, *Color Appearance Models*
Spreading

Spatial frequency
  – The paint chip problem
  – Small text, lines, glyphs
  – Image colors

Adjacent colors blend

Redrawn from *Foundations of Vision*
© Brian Wandell, Stanford University
What color is this?
Color Perception ➔ Color Naming

“Yellow”
Color Perception → Color Naming

What color is this?
Color Perception → Color Naming

“Blue”
What color is this?
Color Perception → Color Naming

“Teal ?”
Color Perception ➔ Color Naming

“Teal ?”

“Turquoise ?” “Blue-Green ?” “Sarcelle ?”
Color according to gender?

Color names if you're a girl...

Maraschino
Cayenne
Maroon
Plum
Eggplant
Grape
Orchid
Lavender
Carnation
Strawberry
Bubblegum
Magenta
Salmon
Tangerine
Cantaloupe
Banana
Lemon
Honeydew
Lime
Spring
Clover
Fern
Moss
Flora
Sea Foam
Spindrift
Teal
Sky
Turquoise

Color names if you're a guy...

Red
Purple
Pink
Orange
Yellow
Green
Blue

Doghouse Diaries
"We take no as an answer."
Color according to XKCD

A crowdsourced color-labeling game
~5 million colors
~222,500 user sessions

http://blog.xkcd.com/2010/05/03/color-survey-results/
Color according to XKCD

*Actual color names* if you’re a girl ...  *Actual color names* if you’re a guy ...

- red
- magenta
- purple
- blue
- pink
- hot pink
- hot pink
- salmon
- orange
- yellow
- light green
- lime green
- neon green
- green
- teal
- blue
We associate and group colors together, often using the name we assign to the colors.
We associate and group colors together, often using the name we assign to the colors.
Color Naming

We associate and group colors together, often using the name we assign to the colors.
Are there natural boundaries?

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).
Basic Color Terms

• Brent Berlin & Paul Kay 1969
• let’s look at two specific places
World Color Survey

Surveyed 2616 speakers of 110 languages using 330 different color chips
Results from WCS (Mexico)
Results from WCS (South Pacific)
Universal (?) Basic Color Terms

Basic color terms recur across languages

- White
- Red
- Pink
- Grey
- Yellow
- Brown
- Black
- Green
- Orange
- Blue
- Purple
Evolution of Basic Color Terms

Proposed universal evolution of color names across languages.
COLOR FOR VISUALIZATION
Why are color choices important?

Example: The Rainbow Color Scale
- Represent data by varying hue across (approximately) the full range of visible wavelengths
- One of the most common color scales in use today

And it’s (usually) a huge mistake!
General Bathymetric Chart of the Ocean

Every color mark signals:
longitude, latitude, sea/land, depth/altitude
General Bathymetric Chart of the Ocean

Every color mark signals:
longitude, latitude, sea/land, depth/altitude

Where is the land?
Where is the sea the deepest?
General Bathymetric Chart of the Ocean

Every color mark signals: longitude, latitude, sea/land, depth/altitude

Where is the land?

Where is the sea the deepest?
General Bathymetric Chart of the Ocean

Now describe what kind of color scale was possibly used here.
General Bathymetric Chart of the Ocean

Now describe what kind of color scale was possibly used here.
Perceptual Ordering

Rainbow Color Scale
- Is ordered by wavelength
- Is **not** perceptually ordered

Gray Scale
- Increases luminance (value) from dark to light
- Is perceptually ordered

From: Rainbow Color Map (Still) Considered Harmful, CG&A
Color Scale Luminance

Rainbow Color Scale

- The visual system perceives high spatial frequencies through changes in luminance
- Is isoluminant (for large portions), changes only appear at color boundaries
- Obscures small details in the data

From: Rainbow Color Map (Still) Considered Harmful, CG&A 07
Color Scale Transitions

Rainbow color scale

– appears separated into bands of almost constant hue
– sharp transitions between hues are perceived as sharp transitions in the data

From: Rainbow Color Map (Still) Considered Harmful, CG&A
HOW TO PICK COLORS
A Few General Rules

- Always have **high luminance contrast** between foreground and background
- Use **only a few distinct colors**

> 12 colors will likely not work
~5 colors recommended

From Ware, Information Visualization
Using Color to Label
(For groups, categories, highlights, etc.)

Colors should be distinctive and named
“Blue”  “Blue-er?”  “Other Blue???”

Use cultural conventions & appreciate symbolism

<table>
<thead>
<tr>
<th>Fruits</th>
<th>Brands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple</td>
<td>Apple</td>
</tr>
<tr>
<td>Banana</td>
<td>AT&amp;T</td>
</tr>
<tr>
<td>Blueberry</td>
<td>Home Depot</td>
</tr>
<tr>
<td>Cherry</td>
<td>Kodak</td>
</tr>
<tr>
<td>Grape</td>
<td>Starbucks</td>
</tr>
</tbody>
</table>

Lin et al. (2013) Selecting Semantically-Resonant Colors for Data Visualization

Beware of bad interactions (red/blue etc.)
Using Color for Scales
(For ordinal or quantitative data)

Use a scale that varies lightness in addition to color.
Shades of gray or shades of a single color are easiest.

For diverging scales, use a lighter, desaturated value for the critical mid-point and darker hues for the ends.
Highly recommended!

Designed originally for maps but will also work well for other types of visualizations.

http://colorbrewer2.org/
ColorBrewer

(RGB)

127, 201, 127
190, 174, 212
253, 192, 134
255, 255, 153
56, 108, 176

(Hex)

0x7FC97F
0xBEAED4
0xFDC086
0xFFF99
0x386CB0
Every ColorBrewer Scale

For CSS and JavaScript (by Mike Bostock)

http://bl.ocks.org/mbostock/5577023
ONE WARNING ABOUT RED-GREEN

If you use red-green 7% of the viewers may not see anything.

The following slides on the topic are adapted from Tobias Isenberg's
Color Vision Deficiency

~ 7% of male population color-deficient

Mostly red-green color deficiency (Deuteranopia or Protanopia) - but other forms as well.
Color Deficiency Test (Ishihara Test)
Color Deficiency Test
Color Deficiency

YELLOW
Color Deficiency
Examples from VIS/InfoVis 2004
Examples from VIS/InfoVis 2004
Better: Red-Blue Contrast
Better: Red-Blue Contrast
Check Your Visualizations!

When possible, avoid red-green color contrasts for visualization purposes.

To test your visualizations, use proofing modes in PhotoShop and GIMP, or try VisCheck http://www.vischeck.com/
Color Resources

Maureen Stone’s Resources
*A Field Guide to Digital Color*
http://www.stonesc.com

Cindy Brewer’s *ColorBrewer*
http://colorbrewer2.org
For CSS and JavaScript
http://bl.ocks.org/mbostock/5577023

Community Palette Sharing
http://www.colourlovers.com
http://kuler.adobe.com
(Fun) Color Resources!

**Wired** “The Crayola-fication of the World”  
by Aatish Bhatia  

**RadioLab** “Colors”  
WNYC Podcast  
PERCEPTION OF OTHER VISUAL ENCODINGS
Perception of Visual Encodings

There are **lots** of possible visual encodings

Their **effectiveness** is related to how they are handled by our perceptual system
Elementary Graphical Perception Tasks

William S. Cleveland (1980s)

Performed controlled experiments to determine how effectively people could judge changes in visual features. Focus on quantitative information.

Variables used: angle, area (size), color hue, color saturation, density (value), length, position, slope, volume.

Figure 3. Graphs from position–angle experiment.
Elementary Graphical Perception Tasks
William S. Cleveland (1980s)
Color Value

• What percentage in value is the right from the left (=100%)?

66%
Color Value

- What percentage in value is the right from the left (=100%)?

60%
Area

- What percentage in size is the right from the left (=100%)?

52%
Area

- What percentage in size is the right from the left (=100%)?

36%
Volume

- What percentage in size is the right from the left (=100%)?

40%
Why people so bad at this?

Relationship between stimulus and perception isn’t always linear!

Stevens’ power law describes a relationship between a physical stimulus \((S)\) and its perceived intensity or strength \((P)\).

[Graph from Wilkinson 99, based on Stevens 61]
Perception

People tend to **correctly estimate lengths**

They tend to **underestimate areas and volumes.**

When asked to pick a circle **2 times** the size, people tend to pick a circle **~1.8 times** larger.

This tendency **gets worse** as area grows.

**Volume is even worse!**

Circles drawn by absolute scaling

Circles drawn by apparent scaling (Flannery)

$S = 0.98A^{0.87}$ [from Flannery 71]
Area

• What percentage in size is the red from the blue (=100%)?

no idea – this is very difficult
What percentage in length is the right from the left (\(=100\%\))? 

75\%
Length / Position

• What percentage in length is the right from the left (=100%)?

25%
Effectiveness of Data Encodings

<table>
<thead>
<tr>
<th>Quantitative</th>
<th>Ordinal</th>
<th>Nominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Position</td>
<td>Position</td>
</tr>
<tr>
<td>Length</td>
<td>Density</td>
<td>Color Hue</td>
</tr>
<tr>
<td>Angle</td>
<td>Color Saturation</td>
<td>Texture</td>
</tr>
<tr>
<td>Slope</td>
<td>Color Hue</td>
<td>Connection</td>
</tr>
<tr>
<td>Area</td>
<td>Texture</td>
<td>Containment</td>
</tr>
<tr>
<td>Volume</td>
<td>Connection</td>
<td>Density</td>
</tr>
<tr>
<td>Density</td>
<td>Containment</td>
<td>Color Saturation</td>
</tr>
<tr>
<td>Color Saturation</td>
<td>Length</td>
<td>Shape</td>
</tr>
<tr>
<td>Color Hue</td>
<td>Angle</td>
<td>Length</td>
</tr>
<tr>
<td>Texture</td>
<td>Slope</td>
<td>Angle</td>
</tr>
<tr>
<td>Connection</td>
<td>Area</td>
<td>Slope</td>
</tr>
<tr>
<td>Containment</td>
<td>Volume</td>
<td>Area</td>
</tr>
<tr>
<td>Shape</td>
<td>Shape</td>
<td>Volume</td>
</tr>
</tbody>
</table>

Mackinlay 1986
Elementary Graphical Perception Tasks

William S. Cleveland (1980s)

also beware of the physical presentation:
PREATTENTIVE PROCESSING
How many 3’s do you see?

1281768756138976546984506985604982826762
9809858458224509856458945098450980943585
9091030209905959595772564675050678904567
8845789809821677654876364908560912949686

From: Ware, Information Visualization using Vision to Think
How about now?
Preattentive Processing

- Some stimuli can be perceived **without** the need for focused attention
- Generally within **200-250 ms**
- Seems to be done **in parallel** by the low-level vision system

**Visual encoding has a big impact on this!**
Visual encodings influence **preattentive** processing

**DETERMINE IF A RED CIRCLE IS PRESENT**
Yes, can be done preattentively

From: Healey, Perception in Visualization
Shape

Yes, can be done preattentively

From: Healey, Perception in Visualization
Hue and Shape

Cannot be done preattentively due to the **conjunction** of shape and hue
→ need to search

From: Healey, Perception in Visualization
Preattentive visual features (some)

- orientation
- length, width
- closure
- size
- curvature
- density, contrast
- number, estimation
- hue
- intensity
- direction of motion
- velocity of motion
- 3D orientation
- artistic properties

From: Healey, Perception in Visualization
When designing visualizations, try to use pre-attentive features to support the most important tasks.

From: Healey, Perception in Visualization
Preattentive visual features (some)

Avoid conjunctions that inhibit preattentive recognition.

(Most conjunctions are require search.)

From: Healey, Perception in Visualization
Applying what we know to

ASSESS VISUAL REPRESENTATIONS
Let’s evaluate…

<table>
<thead>
<tr>
<th>Car / Nation</th>
<th>USA</th>
<th>Japan</th>
<th>Germany</th>
<th>France</th>
<th>Sweden</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accord</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMC Pacer</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audi 5000</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMW 320i</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Champ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Chevy Nova</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saab 9000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

What kind of data are we looking at?

Nations: **Nominal**
Cars: **Nominal**
(Nation,Car): **Nominal**
Let’s evaluate...

Problem:
Length of bar suggests an order or quantity (e.g. Swedish cars are better)
Let's evaluate...

Better!
Let’s evaluate…

**Banks: Market Cap**

- Market Value as of January 20th, 2009, $Bn
- Market Value as of Q2 2007, $Bn

**Market Capitalization =**
What would it cost to buy all of a company’s stock at the current price.

Compares 15 major banks on two dates:
- **January 20th, 2009**
- **Q2 2007** (before banking crisis hit)

Source: Bloomberg, Jan 20th 2009
Let's evaluate...

Banks: Market Cap

- Market Value as of January 20th 2009, $Bn
- Market Value as of Q2 2007, $Bn

Source: Bloomberg, Jan 20th 2009
Problems here?

We are not good at comparing areas. (And the areas here are actually misleading!)

\[
\frac{85}{165} = \approx 50\% 
\]

But this is actually the ratio of the radii, not the areas!

A bar chart would be better.
Problem here?

- There is likely a **bug or error** in the data
- Pie slices are difficult to compare by **area** or by **angle**
- **Similar colors** are difficult to distinguish
- **Perspective distortion** adds to the problem
Similarly...3D bar charts are not recommended

These are much easier to read & compare!
Problem here?

NEWSPAPER MAGAZINE STAFF SIZE OVER TIME
Time and Newsweek select years 1983 - 2005

NUMBER OF CORRESPONDENTS IN BUREAUS OVER TIME
Time and Newsweek, select years 1983 - 2005

NEWSPAPER MAGAZINE BUREAUS OVER TIME
Time and Newsweek select years 1983 - 2005

Journalism.org (via Stephen Few)
Length Comparison

At first glance:
- A huge overall decline
- In 2003, Newsweek is 50% of Time

If we add a proper baseline at 0:
- The downward trend is less severe
- 2003: Newsweek is ~80% of Time
Moreover...

NEWS MAGAZINE STAFF SIZE OVER TIME
Time and Newsweek select years 1983 - 2005

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Staff</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>360</td>
</tr>
<tr>
<td>1993</td>
<td>320</td>
</tr>
<tr>
<td>2003</td>
<td>240</td>
</tr>
<tr>
<td>2004</td>
<td>220</td>
</tr>
<tr>
<td>2005</td>
<td>200</td>
</tr>
</tbody>
</table>

- 10 years each
- 1 year each
Redesign (by Stephen Few)

Time Magazine’s vs. Newsweek Magazine’s Size Over Time

- **Staff Count**
  - Time Magazine
  - Newsweek

- **Correspondent Count**
  - Time Magazine
  - Newsweek

- **Bureau Count**
  - Time Magazine
  - Newsweek


Note: A dashed line connecting two points indicates that there are years between the points for which values were not available. If the values were available, the shape of the lines might vary significantly.
A few more (classic) guidelines!

**Good reference:** *How to Lie with Statistics*, by Darrell Huff (1954)
Chart Rules

• Provide a proper baseline

A 10% increase. Good!

Already looks more impressive

Wow!
Chart Rules

• Provide a **proper baseline** & **label your axes**
Chart Rules

- Provide a **proper baseline** & **label your axes**
- Avoid **eye-candy**

The same data with eye-candy & no numbers ... but at least it tells the same general story.

**Actual data**

Impressive, but a lie!
Chart Rules

- Provide a **proper baseline & label your axes**
- Avoid **eye-candy**
- Avoid **area comparisons** whenever possible
Chart Rules

- Provide a **proper baseline** & **label your axes**
- Avoid **eye-candy**
- Avoid **area comparisons** whenever possible
- Provide **legends**
- **Grids help** – but make them subtle
  (about 20% opacity – **no black lines**)

![Diagram showing processes and visualizations from 1995 to 2010](chart).
Many more useful guidelines!
Today you learned

Details about the perception of color and a few other visual variables

Saw that the vision system is quicker and better at detecting certain visual variables

Learned how to critique visualizations
For Your Projects

Apply what you learned about color

Use **color judiciously**

Pick good colors based on the **data and task**
(e.g. Don’t use a rainbow color scale unless you have a **very** good reason)

Respect the **color blind**

Consider perception when choosing encodings

Choose visual representations that **support your task** and don’t **misrepresent the data**