CSE 412 - Intro to Data Visualization

Visual Encoding Design

Jane Hoffswell  University of Washington
A Design Space of Visual Encodings
Mapping Data to Visual Variables

Assign **data fields** (e.g., with N, O, Q types) to **visual channels** \((x, y, \text{color}, \text{shape}, \text{size}, \ldots\)\) for a chosen **graphical mark** type \((\text{point}, \text{bar}, \text{line}, \ldots\)\). Additional concerns include choosing appropriate **encoding parameters** \((\text{log scale}, \text{sorting}, \ldots\)\) and **data transformations** \((\text{bin}, \text{group}, \text{aggregate}, \ldots\)\).

These options define a large combinatorial space, containing both useful and questionable charts!
Expressive?

Raw

Aggregate (Count)
1D: Quantitative

Raw

Aggregate (Count)
Expressive?

Raw

Aggregate (Count)
Raw (with Layout Algorithm)

Treemap

Bubble Chart

Aggregate (Distributions)

interquartile range  
(middle 50%)

low  median  high

Box Plot

Violin Plot
2D: Nominal x Nominal

Raw

Aggregate (Count)
2D: Quantitative x Quantitative

Raw

Aggregate (Count)
2D: Nominal x Quantitative

Raw

Aggregate (Mean)
3D and Higher

Two variables \([x,y]\)
Can map to 2D points.
Scatterplots, maps, ...

Third variable \([z]\)
Often use one of size, color, opacity, shape, etc. Or, one can further partition space.

What about 3D rendering?
Other Visual Encoding Channels?

Wind map

April 1, 2015
11:35 pm EST
(time of forecast download)

Top speed: 30.5 mph
Average: 10.2 mph
Encoding Effectiveness
Effectiveness Rankings [Mackinlay 86]

QUANTITATIVE
Position
Length
Angle
Slope
Area (Size)
Volume
Density (Value)
Color Sat
Color Hue
Texture
Connection
Containment
Shape

ORDINAL
Position
Density (Value)
Color Sat
Color Hue
Texture
Connection
Containment
Length
Angle
Slope
Area (Size)
Volume
Shape

NOMINAL
Position
Color Hue
Texture
Connection
Containment
Density (Value)
Color Sat
Shape
Length
Angle
Slope
Area
Volume
# Effectiveness Rankings

**QUANTITATIVE**

- Position
  - Length
  - Angle
  - Slope
  - Area (Size)
  - Volume
  - Density (Value)
  - Color Sat
  - Color Hue
  - Texture
  - Connection
  - Containment
  - Shape

**ORDINAL**

- Position
  - Density (Value)
  - Color Sat
  - Color Hue
  - Texture
  - Connection
  - Containment
  - Length
  - Angle
  - Slope
  - Area (Size)
  - Volume
  - Shape

**NOMINAL**

- Position
  - Color Hue
  - Texture
  - Connection
  - Containment
  - Density (Value)
  - Color Sat
  - Shape
  - Length
  - Angle
  - Slope
  - Area
  - Volume

Source: [Mackinlay 86]
Effectiveness Rankings

QUANTITATIVE
- Position
- Length
- Angle
- Slope
- Area (Size)
- Volume
- Density (Value)
- Color Sat
- Color Hue
- Texture
- Connection
- Containment
- Shape

ORDINAL
- Position
- Density (Value)
- Color Sat
- Texture
- Connection
- Containment
- Length
- Angle
- Slope
- Area (Size)
- Volume
- Shape

NOMINAL
- Position
- Color Hue
- Texture
- Connection
- Containment
- Density (Value)
- Color Sat
- Shape
- Length
- Angle
- Slope
- Area
- Volume
Color Encoding (Choropleth Map)
Effectiveness Rankings

**QUANTITATIVE**
- Position
- Length
- Angle
- Slope
- **Area (Size)**
- Volume
- **Density (Value)**
- Color Sat
- Color Hue
- Texture
- Connection
- Containment
- Shape

**ORDINAL**
- Position
- Density (Value)
- Color Sat
- Color Hue
- Texture
- Connection
- Containment
- Shape

**NOMINAL**
- Position
- Color Hue
- Texture
- Connection
- Containment
- Density (Value)
- Color Sat
- Shape
- Length
- Angle
- Slope
- Area
- Volume
Color Encoding (Choropleth Map)
Area Encoding (Symbol Map)
Gene Expression Time-Series [Meyer et al ‘11]

Color Encoding
# Effectiveness Rankings

<table>
<thead>
<tr>
<th>QUANTITATIVE</th>
<th>ORDINAL</th>
<th>NOMINAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Position</strong></td>
<td><strong>Position</strong></td>
<td><strong>Position</strong></td>
</tr>
<tr>
<td>Length</td>
<td>Density (Value)</td>
<td>Color Hue</td>
</tr>
<tr>
<td>Angle</td>
<td>Color Sat</td>
<td>Texture</td>
</tr>
<tr>
<td>Slope</td>
<td>Color Hue</td>
<td>Connection</td>
</tr>
<tr>
<td>Area (Size)</td>
<td>Texture</td>
<td>Containment</td>
</tr>
<tr>
<td><strong>Density (Value)</strong></td>
<td>Connection</td>
<td>Density (Value)</td>
</tr>
<tr>
<td>Color Sat</td>
<td>Containment</td>
<td>Color Sat</td>
</tr>
<tr>
<td><strong>Color Hue</strong></td>
<td>Length</td>
<td>Shape</td>
</tr>
<tr>
<td>Texture</td>
<td>Angle</td>
<td>Length</td>
</tr>
<tr>
<td>Connection</td>
<td>Slope</td>
<td>Angle</td>
</tr>
<tr>
<td>Containment</td>
<td>Area (Size)</td>
<td>Slope</td>
</tr>
<tr>
<td><strong>Shape</strong></td>
<td>Volume</td>
<td>Area</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Volume</td>
</tr>
</tbody>
</table>
Gene Expression Time-Series [Meyer et al. '11]

Color Encoding

Position Encoding

- g4
- g8
- g16
- g17
- g18
- g19
- g20

- s1
- s2
- s3
- s4
- s5
- s6
- s7
- s8
Artery Visualization [Borkin et al ’11]

Rainbow Palette

<table>
<thead>
<tr>
<th>2D</th>
<th>62%</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>39%</td>
</tr>
</tbody>
</table>

Diverging Palette

<table>
<thead>
<tr>
<th>2D</th>
<th>92%</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>71%</td>
</tr>
</tbody>
</table>
Effectiveness Rankings

**QUANTITATIVE**
- Position
- Length
- Angle
- Slope
- Area (Size)
- Volume

**Density (Value)**
- Connection
- Containment
- Color Sat
- Color Hue
- Texture

**Ordinal**
- Position
- Density (Value)
- Color Sat
- Color Hue
- Texture

**NOMINAL**
- Position
- Color Hue
- Texture
- Connection
- Containment
- Density (Value)
- Color Sat
- Shape
- Length
- Angle
- Slope
- Area
- Volume
Administrivia
A2: Exploratory Data Analysis

Use visualization software to form & answer questions

**First steps:**
Step 1: Pick domain & data
Step 2: Pose questions
Step 3: Profile the data
Iterate as needed

**Create visualizations**
Interact with data
Refine your questions

**Author a report**
Screenshots of most insightful views (8+)
Include titles and captions for each view

Due by 11:59pm Monday, Apr 19
Required Readings for Wed 4/14

Scales & Axes
Include Zero in Axis Scale?

Government payrolls in 1937 [How To Lie With Statistics. Huff]
Include Zero in Axis Scale?

Yearly CO$_2$ concentrations  [Cleveland 85]
Include Zero in Axis Scale?

Violates Expressiveness Principle!

Compare Proportions (Q-Ratio)

Compare Relative Position (Q-Interval)
Axis Tick Mark Selection

What are some properties of “good” tick marks?
Axis Tick Mark Selection

**Simplicity** - numbers are multiples of 10, 5, 2

**Coverage** - ticks near the ends of the data

**Density** - not too many, nor too few

**Legibility** - whitespace, horizontal text, size
How to Scale the Axis?
One Option: Clip Outliers
Clearly Mark Scale Breaks

Violates Expressiveness Principle!

Poor scale break [Cleveland 85]  Well-marked scale break [Cleveland 85]
Scale Break vs. Log Scale

[Cleveland 85]
Scale Break vs. Log Scale

Both increase visual resolution
Scale break: difficult to compare (cognitive – not perceptual – work)
Log scale: direct comparison of all data
Logarithms turn \textit{multiplication} into \textit{addition}.

\[
\log(x \cdot y) = \log(x) + \log(y)
\]

Equal steps on a log scale correspond to equal changes to a multiplicative scale factor.
Linear Scale vs. Log Scale

Linear Scale

Log Scale
Linear Scale vs. Log Scale

**Linear Scale**
Absolute change

**Log Scale**
Small fluctuations
Percent change
\[ d(10,30) > d(30,60) \]
When To Apply a Log Scale?

Address **data skew** (e.g., long tails, outliers)
Enables comparison within and across multiple orders of magnitude.

**Focus on multiplicative factors** (not additive)
Recall that the logarithm transforms \( \times \) to \(+\)!
Percentage change, not linear difference.

**Constraint:** **positive, non-zero values**
**Constraint:** **audience familiarity**?
Aspect Ratio
(width : height)
William S. Cleveland
The Elements of Graphing Data
William S. Cleveland
The Elements of Graphing Data
Banking to 45° [Cleveland]

To facilitate perception of trends, maximize the discriminability of line segment orientations.

Two line segments are maximally discriminable when their average absolute angle is 45°.

Method: optimize the aspect ratio such that the average absolute angle of all segments is 45°.
Alternative: Minimize Arc Length while holding area constant  [Talbot et al. 2011]
A Good Compromise

Arc-length banking produces aspect ratios in-between those produced by other methods.

[Talbot et al. 2011]
Trends may occur at different scales!

Apply banking to the original data or to fitted trend lines.

[Heer & Agrawala ’06]

**CO₂ Measurements**

William S. Cleveland

*Visualizing Data*
Visual Encoding Design

Use **expressive** and **effective** encodings

**Reduce** the problem space

Avoid **over-encoding**

Use **space** and **small multiples** intelligently

Use **interaction** to generate *relevant* views

Rarely does a single visualization answer all questions. Instead, the ability to generate appropriate visualizations quickly is critical!
About the design process...

Visualization draws upon both science and art! Principles like expressiveness & effectiveness are not hard-and-fast rules, but can assist us to guide the process and articulate alternatives. They can lead us to think more deeply about our design rationale and prompt us to reflect. It helps to know “the rules” in order to wisely bend (or break) them at the right times!