CSE 412 - Intro to Data Visualization

Visual Encoding Design

Jane Hoffswell  University of Washington
Guest Lecture: Narrative Visualization

This Friday Jan. 22 - Guest: Matt Conlen (UW)
https://mathisonian.com/

Image: "Narrative Visualization: Telling Stories with Data." Segel & Heer. InfoVis 2010
A Design Space of Visual Encodings
Mapping Data to Visual Variables

Assign **data fields** (e.g., with $N, O, Q$ types) to **visual channels** ($x, y, color, shape, size, \ldots$) for a chosen **graphical mark** type ($point, bar, line, \ldots$). Additional concerns include choosing appropriate **encoding parameters** ($log$ scale, sorting, \ldots) and **data transformations** ($bin$, group, aggregate, \ldots). These options define a large combinatorial space, containing both useful and questionable charts!
1D: Nominal

Raw

Aggregate (Count)
Expressive?

Raw

Aggregate (Count)
1D: Quantitative

Raw

Aggregate (Count)
Expressive?

**Raw**

**Aggregate (Count)**

- Miles_per_Gallon
- BIN(Miles_per_Gallon)
- COUNT
  - 20
  - 40
  - 60
  - 80
Effective?

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**

<table>
<thead>
<tr>
<th>Origin</th>
<th>Cylinders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>3</td>
</tr>
<tr>
<td>Japan</td>
<td>4</td>
</tr>
<tr>
<td>USA</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
</tbody>
</table>

**Aggregate (Count)**

<table>
<thead>
<tr>
<th>Origin</th>
<th>Cylinders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>3</td>
</tr>
<tr>
<td>Japan</td>
<td>4</td>
</tr>
<tr>
<td>USA</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
</tbody>
</table>

COUNT

<table>
<thead>
<tr>
<th>Cylinders</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
</tr>
<tr>
<td>6</td>
<td>80</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
</tr>
</tbody>
</table>
2D: Quantitative x Quantitative

Raw

Aggregate (Count)
2D: Nominal x Quantitative

Raw

Aggregated (Mean)
Treemap

Bubble Chart

Beeswarm Plot

Origin
- Europe
- Japan
- USA
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?*

[Bertin]
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

**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

[Mackinlay 86]
# Effectiveness Rankings

Mackinlay 86

<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 (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>Volume</td>
<td>Connection</td>
<td>Density (Value)</td>
</tr>
<tr>
<td>Density (Value)</td>
<td>Containment</td>
<td>Color Sat</td>
</tr>
<tr>
<td>Color Sat</td>
<td>Length</td>
<td>Color Hue</td>
</tr>
<tr>
<td>Color Hue</td>
<td>Angle</td>
<td>Shape</td>
</tr>
<tr>
<td>Texture</td>
<td>Slope</td>
<td>Length</td>
</tr>
<tr>
<td>Connection</td>
<td>Area (Size)</td>
<td>Angle</td>
</tr>
<tr>
<td>Containment</td>
<td>Volume</td>
<td>Slope</td>
</tr>
<tr>
<td>Shape</td>
<td>Shape</td>
<td>Area</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Volume</td>
</tr>
</tbody>
</table>
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

[Mackinlay 86]
Color Encoding (Choropleth Map)
# Effectiveness Rankings

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

Color Encoding

Position Encoding
Artery Visualization [Borkin et al '11]

Rainbow Palette

2D

62%

39%

3D

Diverging Palette

92%

71%
Effectiveness Rankings

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

**DENSITY (VALUE)**
- Position
- Density (Value)
- Color Sat
- Color Hue
- Texture

**COLOR SAT**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**COLOR HUE**
- Position
- Color Hue
- Texture
- Connection
- Containment

**TEXTURE**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**SHAPE**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**VOLUME**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**AREA (SIZE)**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**SLOPE**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**ANGLE**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**CONTAINMENT**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**SHAPE**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**VOLUME**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**CONTAINMENT**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**SHAPE**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**VOLUME**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**CONTAINMENT**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**SHAPE**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**VOLUME**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**CONTAINMENT**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**SHAPE**
- Position
- Color Sat
- Color Hue
- Texture
- Connection

**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, Jan 25
Quiz Section: Vega-Lite

Tomorrow, Thursday January 21st
Hands-on experience with Vega-Lite
Come prepared with questions!

Vega-Lite – A Grammar of Interactive Graphics
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?

Compare Proportions (Q-Ratio)

Compare Relative Position (Q-Interval)

Violates Expressiveness Principle!
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 multiplication into addition.

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

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)
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 $45^\circ$

Method: optimize the aspect ratio such that the average absolute angle of all segments is $45^\circ$
Alternative: Minimize Arc Length while holding area constant [Talbot et al. 2011]

Straight line -> 45°

Ellipse -> Circle
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]

CO$_2$ 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!