Data Visualization in Biomedical Research

Meridith Blevins, MS
Department of Biostatistics
Vanderbilt Institute for Global Health
Experts in data visualization

Edward Tufte

William S. Cleveland
_Visualizing data_. 1993.
A four step recipe for creating a graphic
Step 1: Ingredients

Identify Key Variables

- There are different types of graphs to support the different scales of measurement
  - Nominal Scale
  - Ordinal Scale
  - Interval Scale
Step 2: Palette

Determine Role of Visualization

- There are different roles for data visualization
  - Presentation vs. Exploratory (then stop here)
  - Summary/Descriptive
  - Distribution
  - Comparison
  - Association
- Answer these questions
  - What to Whom, How, and Why?
Step 3: The Dish

Choice of Graphical Form

- Based on the Key Variables (or ingredients) and Role of Visualization (or palette)
  - 2D plot (most common)
  - Multiple 2D plots
  - Interactive plot
  - Multiple plots stringed together like a story line
- We will go through the different graphical forms in our R examples
Step 4: Add Garnish

• Think about improvements that could be made to improve clarity and focus
  ○ Scales
  ○ Sorting and Ordering
  ○ Overlays
  ○ Positioning in text
  ○ Size, Frames and Aspect Ratio
  ○ Color
“Ink-to-Information”

= data-ink ÷
  total ink used to print the graphic
= proportion of graphic’s ink devoted to
  the non-redundant display of data
= 1 – proportion of graphic that can be erased

- Also called data-ink ratio
“Ink-to-Information”

- $\text{Ink-to-Information} = \frac{\text{data-ink}}{\text{total ink used to print the graphic}}$
- $\text{Ink-to-Information} = \text{proportion of graphic’s ink devoted to the non-redundant display of data}$
- $\text{Ink-to-Information} = 1 - \text{proportion of graphic that can be erased}$

(source: The Visual Display of Quantitative Information by Edward R. Tufte)
Choosing Colors

colorbrewer2.org
Choosing Colors
Designing for Colorblind Audience

http://www.color-blindness.com/coblis-color-blindness-simulator/

- Protanopia (red-blind)
- Deuteranopia (green-blind)
- Tritanopia (blue-blind)
- Protanomaly (red-weak)
- Deuteranomaly (green-weak)
- Tritanomaly (blue-weak)
- Achromatopsia (monochromacy)
- Blue Cone Monochromacy

In USA, estimated:
♂ 7%
♀ 0.4%
Bad Practices

- How much ink for 5 numbers?
- People are bad at judging the relative magnitude of angles
- If you twist the rotation of the pie, you can cause observers to systematically misjudge the proportions
- The third dimension makes judgment worse
3D representation of 2D data

- 3D columns can cross gridlines
- Display of zero values would be misleading
- Difficult to make fair comparisons

Figure 2.2. Church attendance (DDB Life Style Survey 1975–1998)
Dynamite Plots

- You want to capture measure of central tendency and variability, but not like this...
Regression to the mean
Graphics for Grant Applications

- **Planning (Gantt Chart)**

<table>
<thead>
<tr>
<th>Task Name</th>
<th>Sep 9, '01</th>
<th>Sep 16, '01</th>
<th>Sep 23, '01</th>
<th>Sep 30, '01</th>
<th>Oct 7, '01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milestone 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milestone 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Significance through Ngrams** ([click here](#))

- **Sample size and power analyses**
  - Precision
  - Correlation
  - Two group comparison
  - Cluster randomized trial

- **Make sure they are small with large, concise fonts and good resolution. Must stand on their own.**
Study Flow Diagrams

- CONSORT Flow Diagram
- STARD Flow Diagram
- STROBE Checklist and Flow Diagram (if applicable)
Dialect Map

Take the quiz: http://www.nytimes.com/interactive/2013/12/20/sunday-review/dialect-quiz-map.html

What is your generic term for a sweetened carbonated beverage?

Joshua Katz, Department of Statistics, NC State University
Based on data from the Harvard Dialect Survey by Bert Vaux & Scott Golder
Babies Dataset

The data provided here are part of the Child Health and Development Studies (CHDS)— a comprehensive investigation of all pregnancies that occurred between 1960 and 1967 among women in the Kaiser Foundation Health Plan (a prepaid medical care program). The women in the study were all those enrolled in the Kaiser Plan who had obtained prenatal care in the San Francisco–East Bay area and who delivered at any of the Kaiser hospitals in Northern California. The R dataset has a subset of this information collected for 1236 babies.

Baby boys born during one year of the study who lived at least 28 days and who were single births (i.e., not one of a twin or triplet).
<table>
<thead>
<tr>
<th>Name</th>
<th>Role</th>
<th>Label</th>
<th>Units</th>
<th>Type</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP</td>
<td>Predictor</td>
<td>Treatment</td>
<td></td>
<td>Binary</td>
<td>1 = Placebo; 2 = Treatment</td>
</tr>
<tr>
<td>AGE</td>
<td>Predictor</td>
<td>Age</td>
<td>Years</td>
<td>Continuous</td>
<td>18 - 75</td>
</tr>
<tr>
<td>SEX</td>
<td>Predictor</td>
<td>Gender</td>
<td></td>
<td>Binary</td>
<td>1 = Female; 2 = Male</td>
</tr>
<tr>
<td>HT</td>
<td>Predictor</td>
<td>Height</td>
<td>in.</td>
<td>Continuous</td>
<td>48 - 96</td>
</tr>
<tr>
<td>WT</td>
<td>Predictor</td>
<td>Weight</td>
<td>lbs.</td>
<td>Continuous</td>
<td>75 - 350</td>
</tr>
<tr>
<td>HCT</td>
<td>Predictor</td>
<td>Heart rate</td>
<td>beats/min.</td>
<td>Continuous</td>
<td>30 - 50</td>
</tr>
<tr>
<td>BPSYS</td>
<td>Predictor</td>
<td>Systolic BP</td>
<td>mmHg</td>
<td>Continuous</td>
<td>100 - 160</td>
</tr>
<tr>
<td>BPDIAS</td>
<td>Predictor</td>
<td>Diastolic BP</td>
<td>mmHg</td>
<td>Continuous</td>
<td>80 - 150</td>
</tr>
<tr>
<td>STAGE</td>
<td>Predictor</td>
<td>WHO stage</td>
<td></td>
<td>Discrete numeric</td>
<td>1 - 4</td>
</tr>
<tr>
<td>RACE</td>
<td>Predictor</td>
<td>Race</td>
<td></td>
<td>Categorical</td>
<td>1 = White; 2 = Black; 3 = Other</td>
</tr>
<tr>
<td>DATE1</td>
<td>Additional</td>
<td>Date of last visit</td>
<td></td>
<td></td>
<td>mm/dd/yyyy</td>
</tr>
<tr>
<td>COMPLIC</td>
<td>Outcome</td>
<td>Complications?</td>
<td></td>
<td>Binary</td>
<td>0 = No; 1 = Yes</td>
</tr>
</tbody>
</table>
Nomogram

- Text
One Continuous Variable

- Histogram
- Boxplot
- Violin Plot
- Stripchart
- Density plot
One categorical variable

- Text
Layering to improve information

- = data-ink ÷