Learning Objectives

- Learn different statistical data graphs
  - Line graph, Bar Graph, Scatterplot, Trellis, Crosstab, Stacked bars, Dotplot, Radar graph, Box plot, Pareto chart, Bump chart, Histogram, Frequency plot, Strip plot, Steam-and-leaf plot, Heatmap
- Learn type of data and analytic goal each technique best applies to
- Develop skill at choosing graph(s) to display different types of data and data sets
- Learn approaches to address overplotting
- Understand concept of “banking to 45°”
Sources Used

Few’s Selection & Design Process

- Determine your message and identify your data
- Determine if a table, or graph, or both is needed to communicate your message
- Determine the best means to encode the values
- Determine where to display each variable
- Determine the best design for the remaining objects
  - Determine the range of the quantitative scale
  - If a legend is required, determine where to place it
  - Determine the best location for the quantitative scale
  - Determine if grid lines are required
  - Determine what descriptive text is needed
- Determine if particular data should be featured and how

S Few
“Effectively Communicating Numbers”
http://www.perceptualedge.com/articles/Whitespapers/Communicating_Numbers.pdf

Some examples...
Points, Lines, Bars, Boxes

- Points
  - Useful in scatterplots for 2-values
  - Can replace bars when scale doesn’t start at 0
- Lines
  - Connect values in a series
  - Show changes, trends, patterns
  - Not for a set of nominal or ordinal values
- Bars
  - Emphasizes individual values
  - Good for comparing individual values
- Boxes
  - Shows a distribution of values

Vertical vs. Horizontal Bars

- Horizontal can be good if long labels or many items
Multiple Bars

- Can be used to encode another variable

Examples
Goal

You want to present quantitative sales performance data for the 4 regions of your company for the four quarters of the year.

Design
Before

After
Surveys

• Who hasn’t completed one?
• Class slides: internal & external

Upcoming Examples

• Page references are from *Now You See It*
Add Reference Lines

More Reference Lines
Trellis Display

Typically varies on one variable

Distribute different values of that variable across views

Crosstab

Varies across more than one variable
Concurrent Views

- He calls such things *faceted analytical displays*
  - Sometimes that term is used in other ways in infovis
- As opposed to *dashboards*
  - They are for monitoring, not analysis

Overplotting

Too many data points
Overplotting Solutions

- Reducing size of data objects
- Removing all fill color from data objects
- Changing the shape of data objects
- Jittering data objects
- Making data objects transparent
- Encoding the density of values
- Reducing the number of values
  - Aggregating the data
  - Filtering the data
  - Breaking the data into a series of separate graphs
  - Statistically sampling the data

Quantitative Data

- Fundamental visualization techniques
**Time Series Data**

- Patterns to be shown
  - Trend
  - Variability
  - Rate of change
  - Co-variation
  - Cycles
  - Exceptions

**Time Series Visualizations**

- Effective visualization techniques include...
Line Graphs

When to use:
When quantitative values change during a continuous period of time

Bar Graphs

When to use:
When you want to support the comparison of individual values
**Dot Plots**

When to use:
When analyzing values that are spaced at irregular intervals of time

p. 153

**Radar Graphs**

When to use:
When you want to represent data across the cyclical nature of time

p. 154
Heatmaps

When to use:
When you want to display a large quantity of cyclical data (too much for radar)

Box Plots

When to use:
You want to show how values are distributed across a range and how that distribution changes over time
Animated Scatterplots

When to use:
To compare how two quantitative variables change over time
p. 159

Banking to 45°

Same diagram, just drawn at different aspect ratios
People interpret the diagrams better when lines are around 45°, not too flat, not too steep
p. 171
Question

Which is increasing at a faster rate, hardware sales or software sales? Log scale shows this

Both at same rate, 10%

Patterns

Daily sales Average per day

p. 176
Cycle Plot

Combines visualizations from two prior graphs

A Story

How much wine of different varieties is produced?
Stacked Bars

https://priceonomics.com/ranking-the-most-and-least-diverse-colleges-in/

Pareto Chart

Shows individual contributors and increasing total

80/20 rule – 80% of effect comes from 20%

p. 194
**Bump Chart**

Shows how ranking relationships change over time

---

**Deviation Analysis**

Do you show the two values in question or the difference of the two?
Distribution Analysis Views

- Histogram
- Frequency polygon
- Strip plot
- Stem-and-leaf plot
Stem-and-leaf Plot

Comparisons

Note how first one’s curve is smooth (not such a noticeable difference). Second one is more noticeable. Same data.
Bleah. How can we clean this up?

Crosstab
Color Choice in Heatmaps

Argues that black should not be used as a middle value because of its saliency (visual prominence)

Some people are red-green color blind too

p. 285-7

Fun Examples

Critique It

Tracking the plunge

The sharp decline in home values has left many homeowners "under water"—owing more on their homes than their property is worth. This means that, in spite of low interest rates that would normally lead people to refinance their mortgages, many homeowners no longer have sufficient equity to qualify for refinancing.
Learning Objectives

- Learn different statistical data graphs
  - Line graph, Bar Graph, Scatterplot, Trellis, Crosstab, Stacked bars, Dotplot, Radar graph, Box plot, Pareto chart, Bump chart, Histogram, Frequency plot, Strip plot, Steam-and-leaf plot, Heatmap
- Learn type of data and analytic goal each technique best applies to
- Develop skill at choosing graph(s) to display different types of data and data sets
- Learn approaches to address overplotting
- Understand concept of “banking to 45°”

HW 2

- Table and graph design
- Given two (Excel) data sets, design a table and graph for the data, respectively
- Due next Wednesday
- Submit 2 copies
Project

- Everyone’s description on wiki?
- 3-4 person teams
  - Self-select or instructor-paired
- Milestones
  - Teams & topics 2 weeks from Wednesday
  - Email me to get paired up
- Topic ideas

Project Topic Ideas

- Check out t-square wiki page
- Two nice submitted projects
- Think about your personal interests
- Think investigative journalism

- Want projects that we could show off externally and would be of broad interest
Vis Lab machines available for use

Fall 2016
CS 7450
63
Reading

- S. Few, “Effectively Communicating Numbers” web article
- 538 Best of ‘15 webpage
Upcoming

- No Class – Labor Day

- Visual Perception