Statistical Graphs & Charts

CS 4460 – Intro. to Information Visualization
August 30, 2017
John Stasko

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°"
- Just get better at applying and using the standard charts

Sources Used
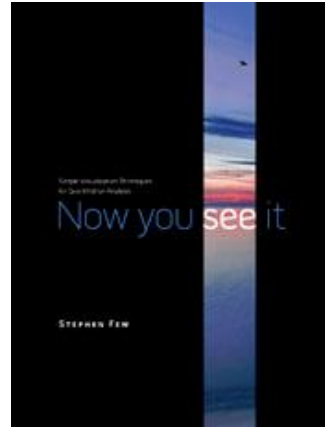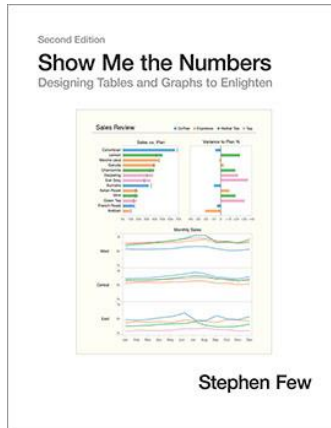
Few's Selection & Design Process

- 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"
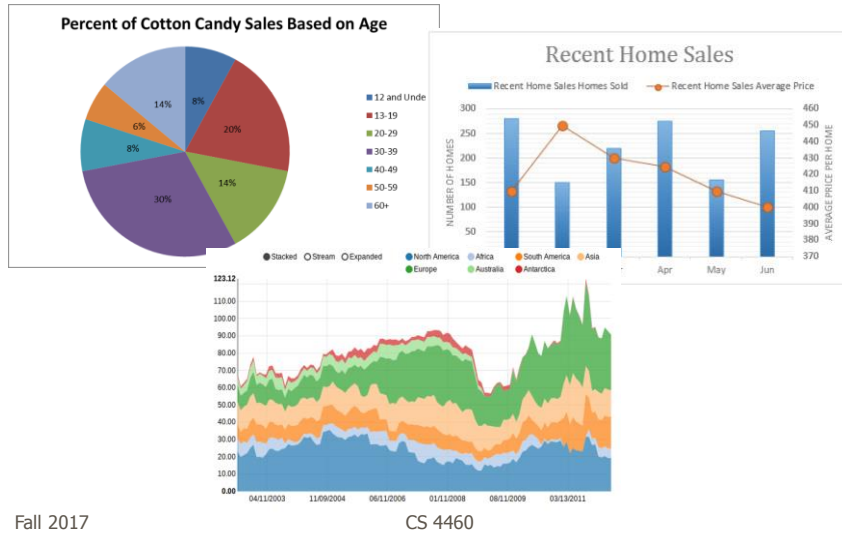http://www.perceptualedge.com/articles/Whitepapers/Communicating_Numbers.pdf

Some examples…

Let's See Some Examples

Vertical vs. Horizontal Bars

- Horizontal can be good if long labels or many items

Multiple Bars

- Can be used to encode another variable

Upcoming Examples

- Page references are from *Now You See It*

Add Reference Lines

p. 96

More Reference Lines

p. 97

Trellis Display

Typically varies on one variable

Distribute different values of that variable across views

p. 100

Crosstab

Varies across more than one variable

p. 102

Crosstab

p. 103

Overplotting

Too many data points

p. 118

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

Fall 2017   CS 4460   15

Time Series Data

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

Fall 2017   CS 4460   16

Time Series Visualizations

- Effective visualization techniques include…

Line Graphs

When to use:

When quantitative values change
during a continuous period of time

p. 151

Bar Graphs

When to use:

When you want to support the
comparison of individual values

p. 152

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)

p. 157

Box Plots

When to use:

You want to show how values are distributed
across a range and how that distribution
changes over time   p. 157

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%

p. 172

A Story

How much wine of
different varieties is produced?

p. 191-2

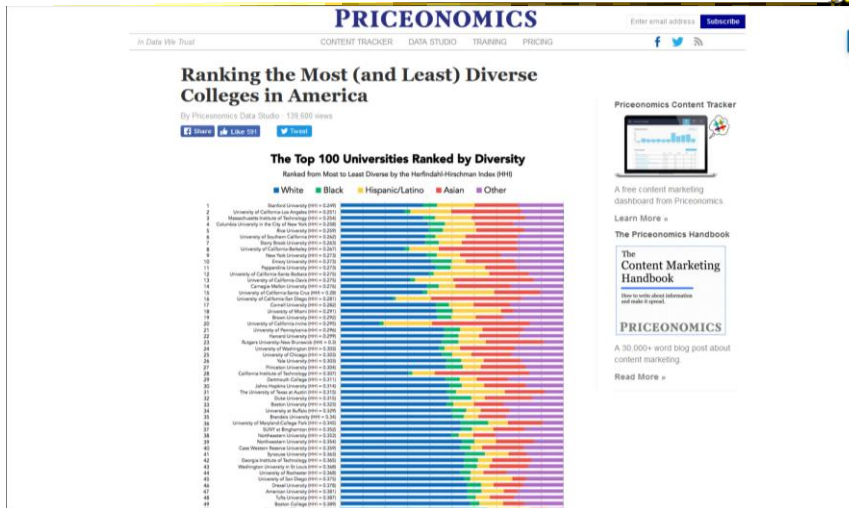# Stacked Bars

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

p. 201

Deviation Analysis

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

p. 203

Distribution Analysis Views

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

Histogram

p. 225

Frequency Plot

p. 226

Strip Plot

p. 227

Stem-and-leaf Plot

p. 228

Comparisons

Note how first one's curve is smooth (not such a noticeable difference). Second one is more noticeable. Same data.

p. 234

Correlation Analysis

Bleah. How can we clean this up?

p. 276

Crosstab

p. 277

Color Choice in Heatmaps

Argues that black should not be used as a middle value because of its saliency (visual prominence)
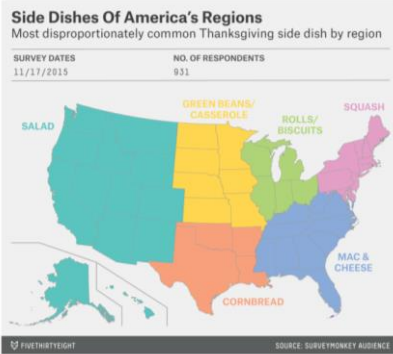
Some people are red-green color blind too

p. 285-7

Fun Examples

http://fivethirtyeight.com/features/our-47-weirdest-charts-from-2015/

From QlikView

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.
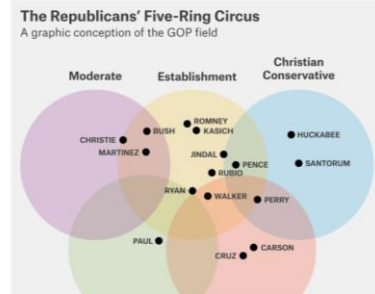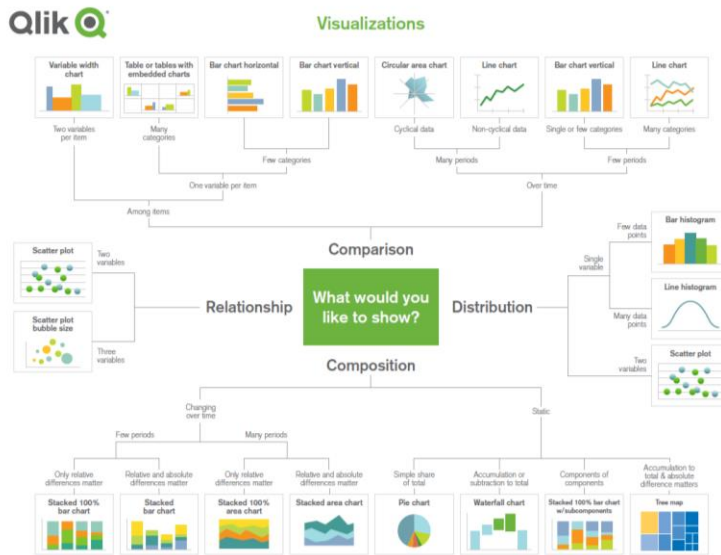
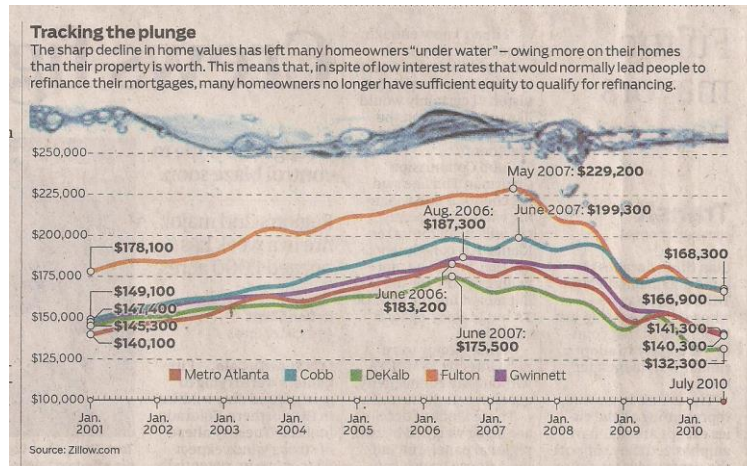AJC, July 2010

Fall 2017   CS 4460   43

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°"
- Just get better at applying and using the standard charts

Fall 2017   CS 4460   44

22

HW 1

- Questions?

- Remember to bring two hardcopies on Friday

Friday

- First lab of term
  - Prep: Read Murray 1st half chapter 3
  - Bring your laptop
  - Install the following on your laptop
    - sublime (or some other code editor/IDE)
    - Chrome (or some other browser)
    - python (if Mac or Linux, already there)
  - git clone or download starter code

Upcoming

- Lab 1 – HTML, CSS, DOM
  – Prep: Murray, chapter 3 up to Javascript

- No Class – Labor Day