

Statistical Graphs & Charts



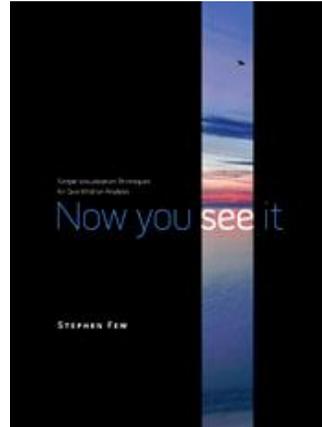
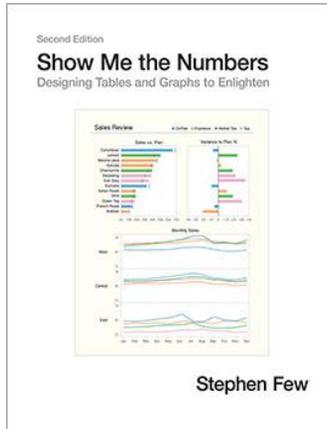
CS 7450 - Information Visualization
August 31, 2016
John Stasko

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°”

Sources Used



Fall 2016

CS 7450

3

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...

Fall 2016

CS 7450

4

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

Fall 2016

CS 7450

5

Vertical vs. Horizontal Bars



- Horizontal can be good if long labels or many items

Fall 2016

CS 7450

6

Multiple Bars



- Can be used to encode another variable

Fall 2016

CS 7450

7



Examples

Fall 2016

CS 7450

8

Goal



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

Fall 2016

CS 7450

9

Design



Fall 2016

CS 7450

10

Before



Fall 2016

CS 7450

11

After



Fall 2016

CS 7450

12

Before



Fall 2016

CS 7450

13

After



Fall 2016

CS 7450

14

Before



Fall 2016

CS 7450

15

After



Fall 2016

CS 7450

16

Surveys



- Who hasn't completed one?
- Class slides: internal & external

Fall 2016

CS 7450

17

Upcoming Examples



- Page references are from *Now You See It*

Fall 2016

CS 7450

18

Add Reference Lines



p. 96

Fall 2015

CS 7450

19

More Reference Lines



p. 97

Fall 2015

CS 7450

20

Trellis Display



Typically varies on
one variable

Distribute different
values of that
variable across views

p. 100

Fall 2016

CS 7450

21

Crosstab



Varies across more
than one variable

p. 102

Fall 2015

CS 7450

22

Crosstab



p. 103

Fall 2015

CS 7450

23

Multiple Concurrent Views



Vintage
infovis

Fall 2015

CS 7450

p. 107

24

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

Fall 2015

CS 7450

25

Overplotting



Too many data points

p. 118

Fall 2015

CS 7450

26

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 2015

CS 7450

27

Quantitative Data



- Fundamental visualization techniques

Fall 2015

CS 7450

28

Time Series Data



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

Fall 2015

CS 7450

29

Time Series Visualizations



- Effective visualization techniques include...

Fall 2015

CS 7450

30

Line Graphs



When to use:

When quantitative values change
during a continuous period of time

p. 151

Fall 2015

CS 7450

31

Bar Graphs



When to use:

When you want to support the
comparison of individual values

p. 152

Fall 2015

CS 7450

32

Dot Plots



When to use:

When analyzing values that are spaced at irregular intervals of time

p. 153

Fall 2015

CS 7450

33

Radar Graphs



When to use:

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

p. 154

Fall 2015

CS 7450

34

Heatmaps



When to use:

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

p. 157

Fall 2015

CS 7450

35

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

Fall 2015

CS 7450

36

Animated Scatterplots



When to use:

To compare how two quantitative variables change over time

Fall 2015

CS 7450

p. 159

37

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

Fall 2015

CS 7450

p. 171

38

Question



Which is increasing at a faster rate, hardware sales or software sales?

Log scale shows this

Both at same rate, 10%

Fall 2015

CS 7450

p. 172
39

Patterns



Daily sales

Average per day

Fall 2015

CS 7450

p. 176
40

Cycle Plot



Combines visualizations
from two prior graphs

p. 177

Fall 2015

CS 7450

41

A Story

How much wine of
different varieties is produced?



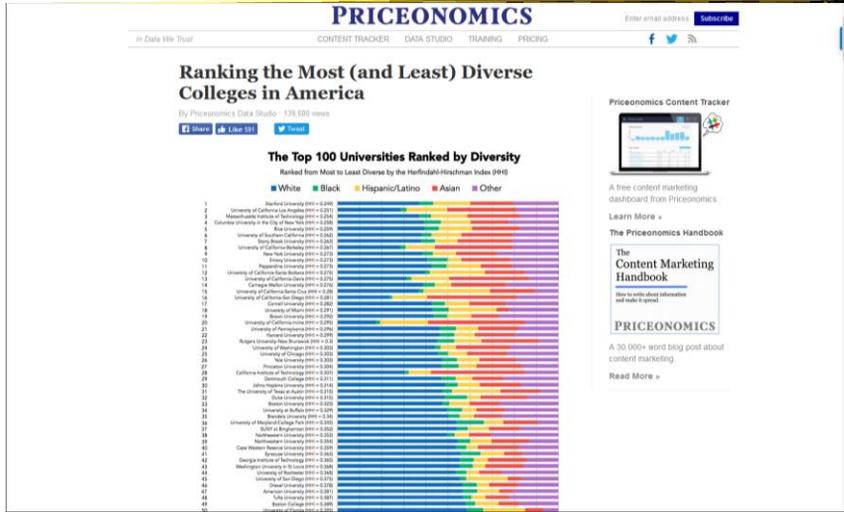
p. 191-2

Fall 2015

CS 7450

42

Stacked Bars



https://priceonomics.com/ranking-the-most-and-least-diverse-colleges-in/
 Fall 2016 CS 7450 43

Pareto Chart

Shows individual contributors and increasing total

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

Fall 2015 CS 7450

p. 194 44

Bump Chart



Shows how ranking relationships change over time

p. 201

Fall 2015

CS 7450

45

Deviation Analysis



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

p. 203

Fall 2015

CS 7450

46

Distribution Analysis Views



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

Fall 2015

CS 7450

47

Histogram



Fall 2015

CS 7450

p. 225

48

Frequency Plot



Fall 2015

CS 7450

p. 226
49

Strip Plot



Fall 2015

CS 7450

p. 227

50

Stem-and-leaf Plot



p. 228

Fall 2015

CS 7450

51

Comparisons



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

p. 234

Fall 2015

CS 7450

52

Correlation Analysis



Bleah. How can we clean this up?

p. 276

Fall 2015

CS 7450

53

Crosstab



p. 277

Fall 2015

CS 7450

54

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

Fall 2015

CS 7450

55

Fun Examples

22.



FiveThirtyEight
Politics Sports Science & Health Economics Culture

DEC 31, 2015 AT 7:01 AM

Our 47 Weirdest Charts From 2015

By [Andrei Scheinkman](#)
Filed under [2015 Year In Review](#)

We made more than 1,500 charts in 2015 at FiveThirtyEight. Many were bar charts, line charts and scatterplots — but not all. Here are some of the more unusual graphics we published.

- ### The Republicans' Five-Ring Circus

A graphic conception of the GOP field

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

Fall 2016

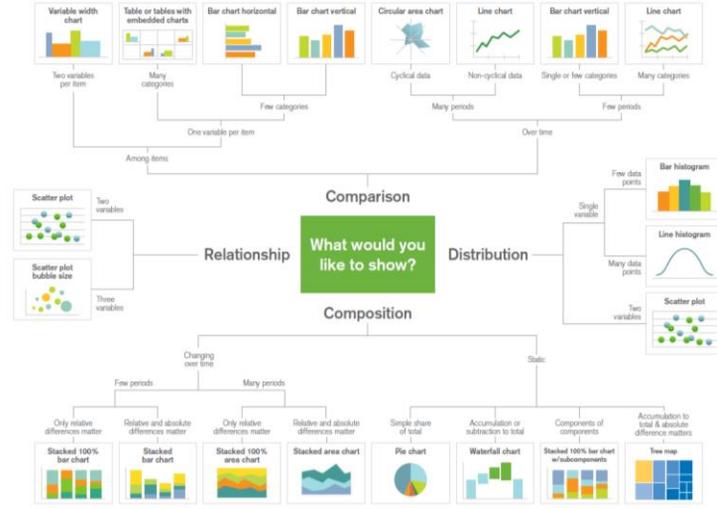
CS 7450

56

From QlikView



Visualizations

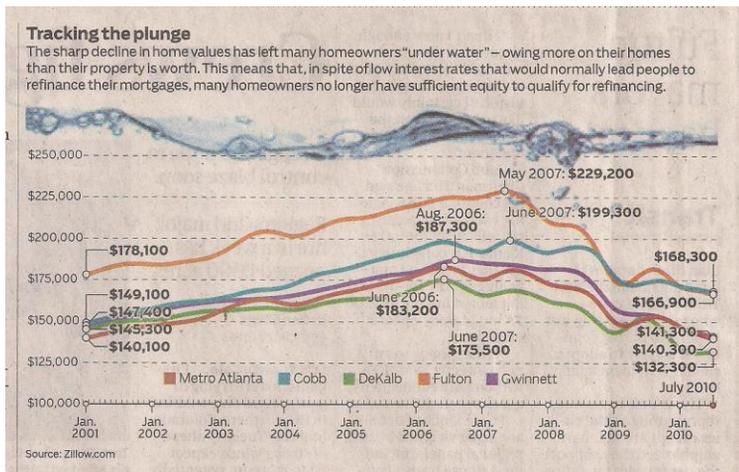


Fall 2015

CS 7450

57

Critique It



AJC, July 2010

Fall 2015

CS 7450

58

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°”

Fall 2016

CS 7450

59

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

Fall 2016

CS 7450

60

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

Fall 2015

CS 7450

61

Project Topic Ideas

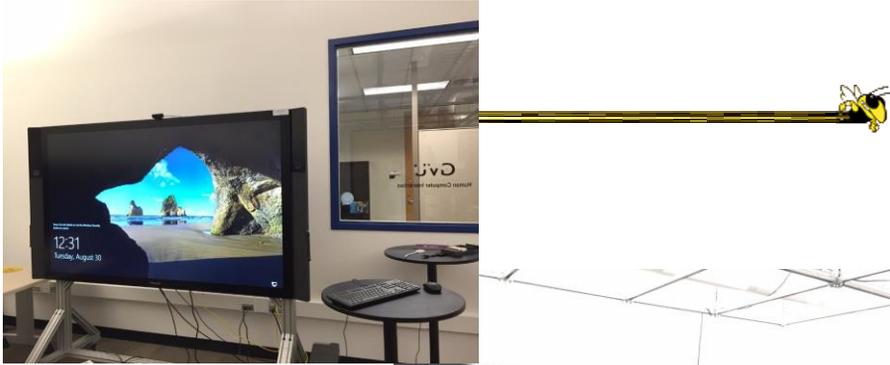


- Check out t-square wiki page
 - Two nice submitted projects
 - Think about your personal interests
 - Think investigative journalism
-
- Want projects that we could show off externally and would be of broad interest

Fall 2016

CS 7450

62



Vis Lab machines
available for use



Fall 2016

CS 7450

63



Fall 2016

CS 7450

64



Fall 2016

CS 7450

65

Reading

- S. Few, "Effectively Communicating Numbers" web article
- 538 Best of '15 webpage

Fall 2016

CS 7450

66

Upcoming



- **No Class** – Labor Day
- Visual Perception