Example Visualizations

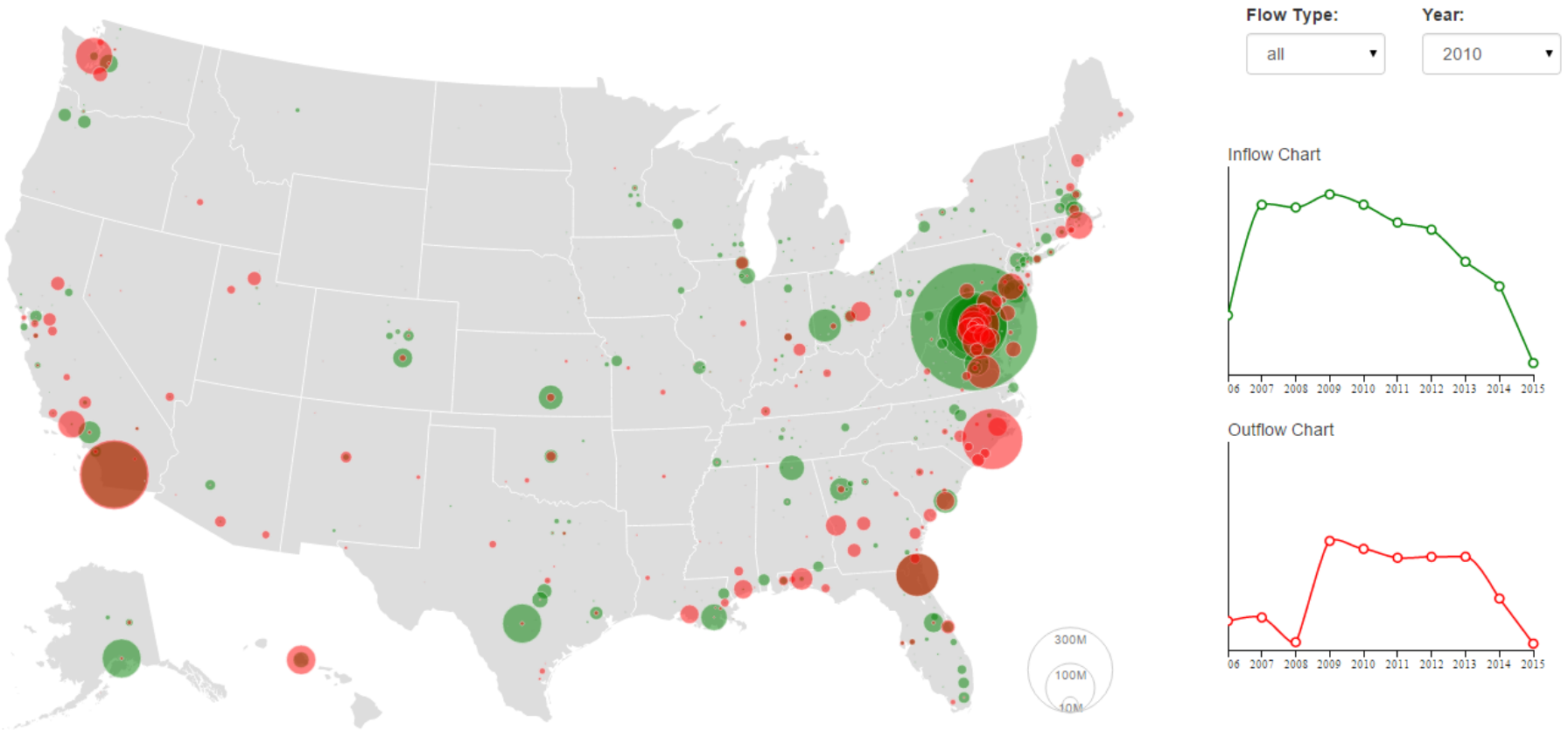
Professor Jeff Saltz

School of Information Studies
SYRACUSE UNIVERSITY

Learning Objectives

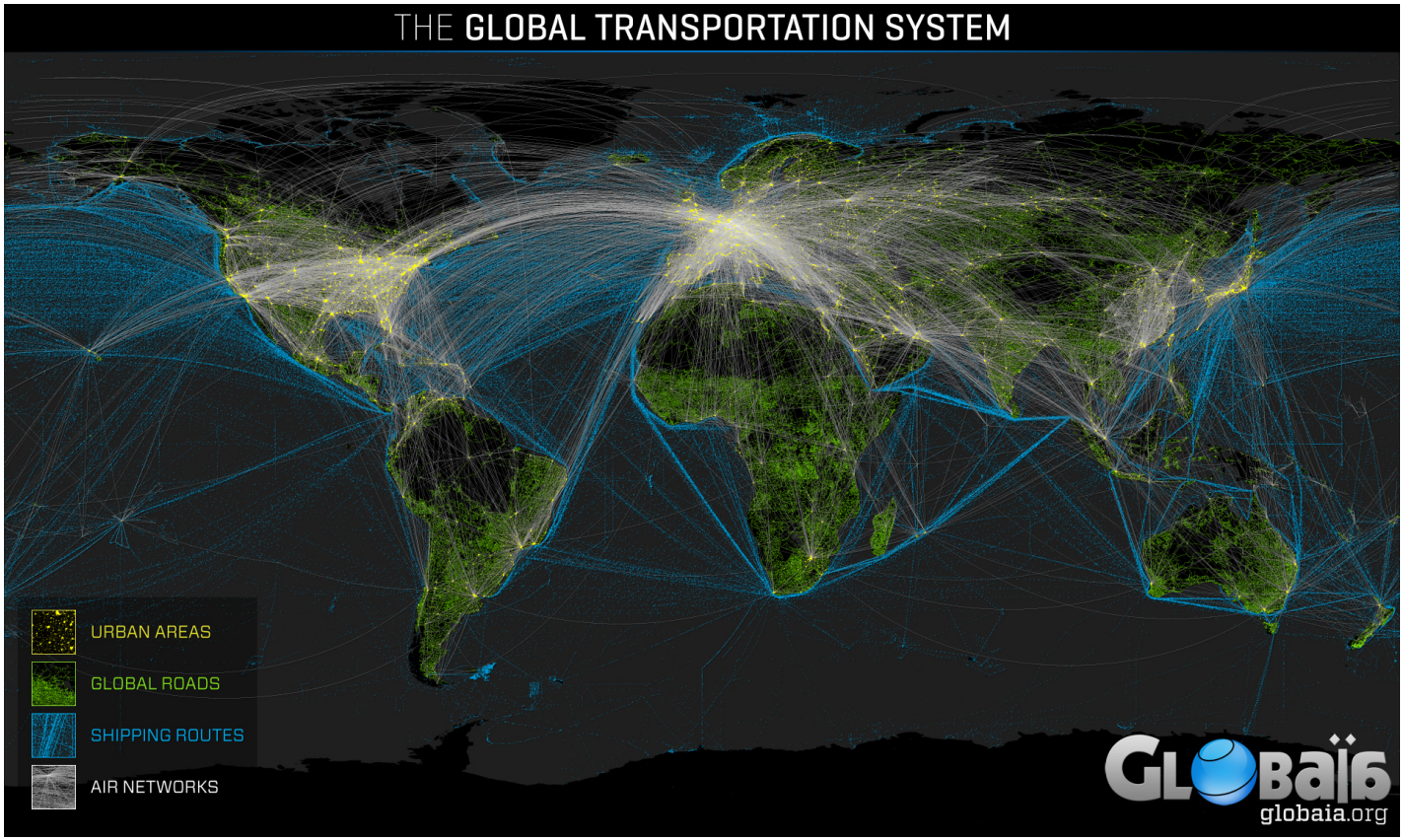
- Understand how visualization is used (via exploring some examples)
- Gain a basic understanding of how to represent numbers visually
- Gain knowledge of the principles of visualization
- Be able to generate basic visualizations using GGPLOT2 in R

Web Science and Digital Libraries Research Group



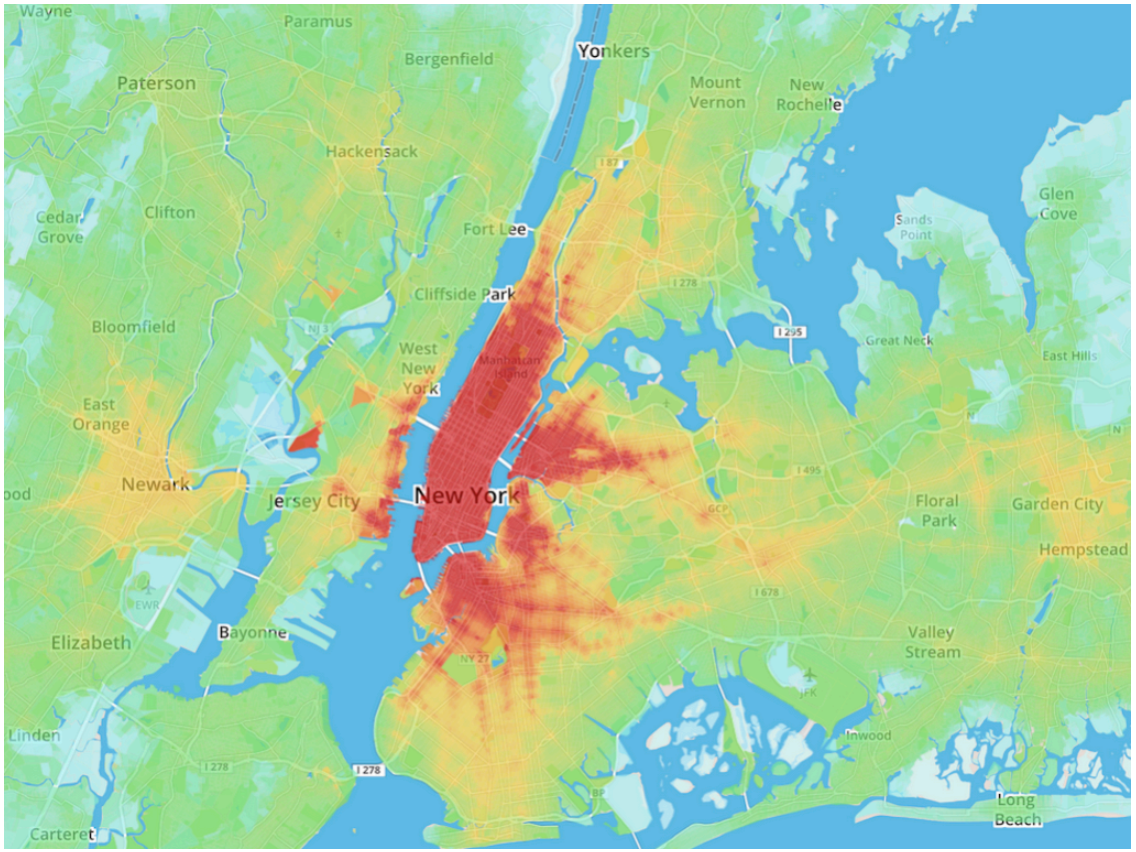
http://ws-dl.blogspot.com/

The Global Transportation System



http://globaia.org/portfolio/cartography-of-the-anthropocene/

America's 10 Best Cities for Commuting on Public Transit

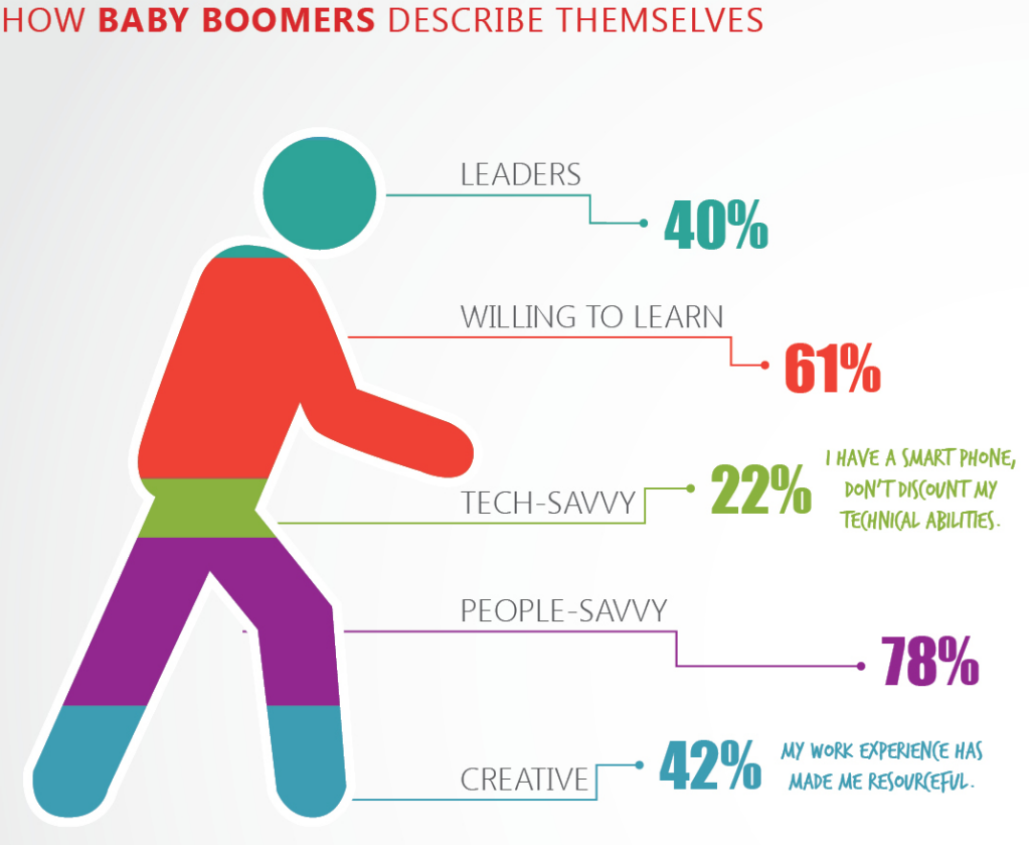


Jobs within 30 minutes
by transit, averaged 7 - 9 AM

- 0 - 1,000
- 1,000 - 2,500
- 2,500 - 5,000
- 5,000 - 7,500
- 7,500 - 10,000
- 10,000 - 25,000
- 25,000 - 50,000
- 50,000 - 75,000
- 75,000 - 100,000
- 100,000 - 250,000
- 250,000 - 500,000
- 500,000 - 750,000
- 750,000 - 1,000,000
- 1,000,000 +

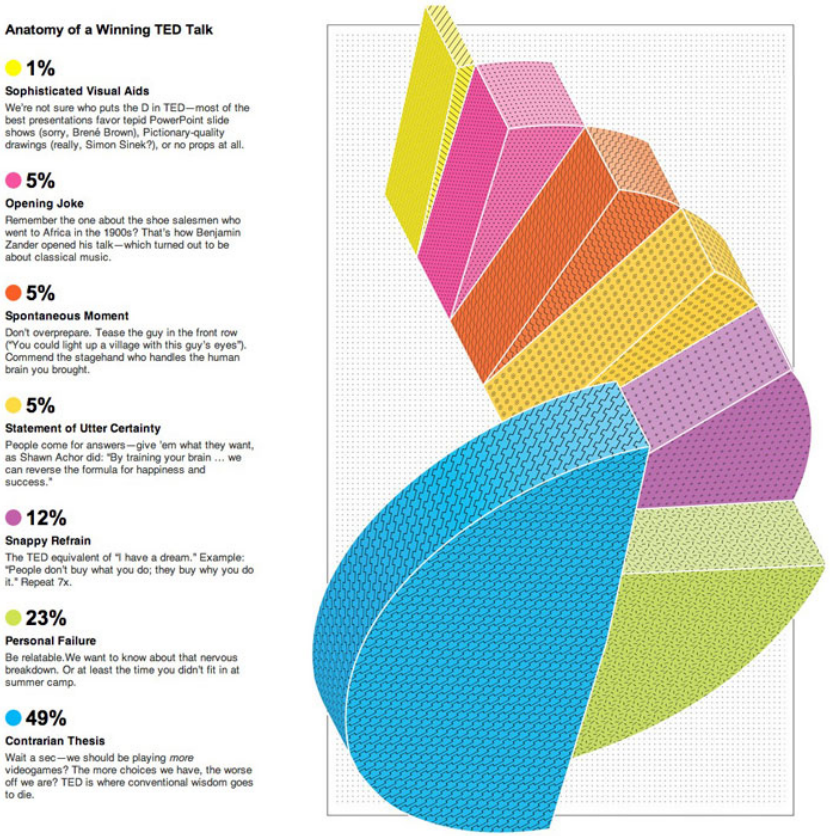
http://www.wired.com/2014/10/americas-10-best-cities-commuting-public-transit/

Data Man



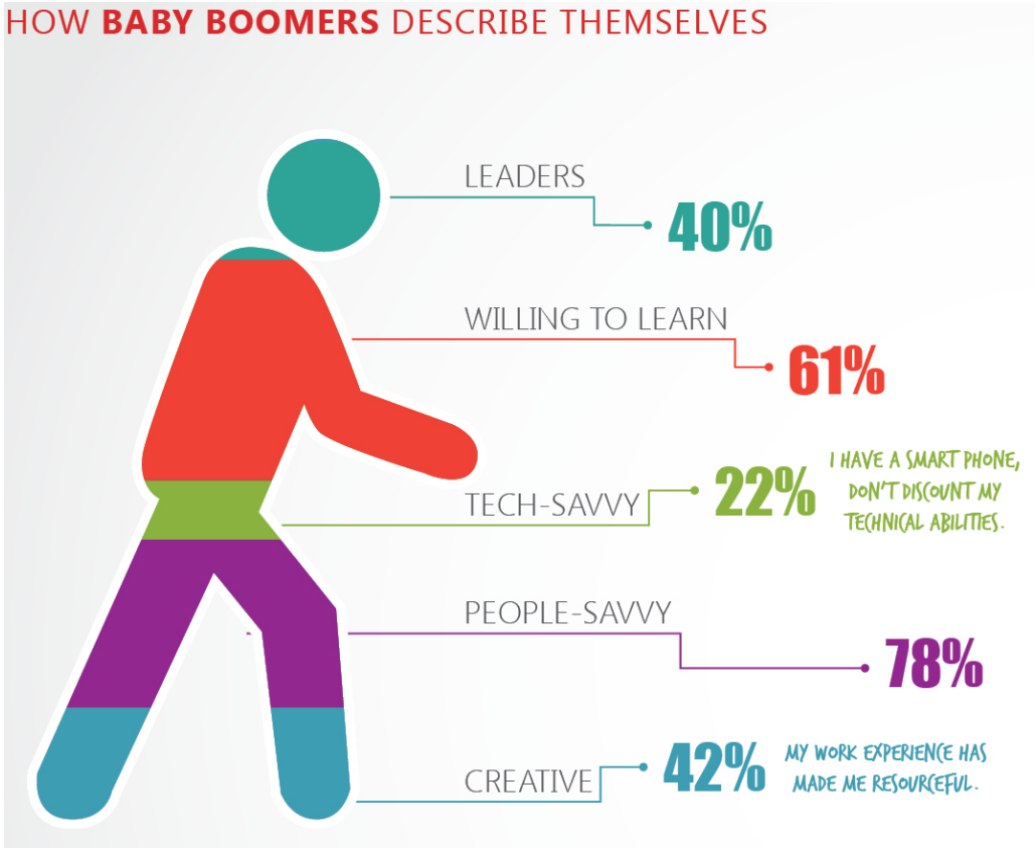
http://wpengine.com/blog/ugly-side-data-visualization/

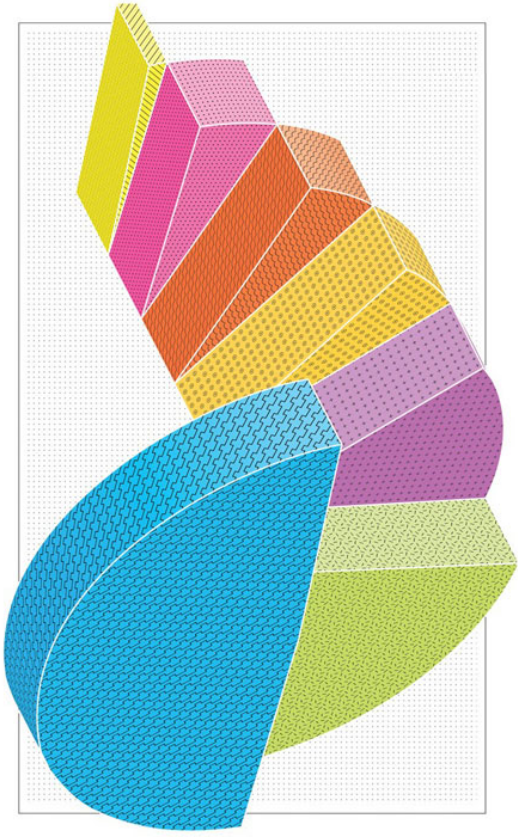
Spiral Graph



http://cdn.wpengine.com/wp-content/uploads/2014/03/Anatomy-of-a-winning-TED-Talk.jpg

Question: What is Good or bad about these charts?





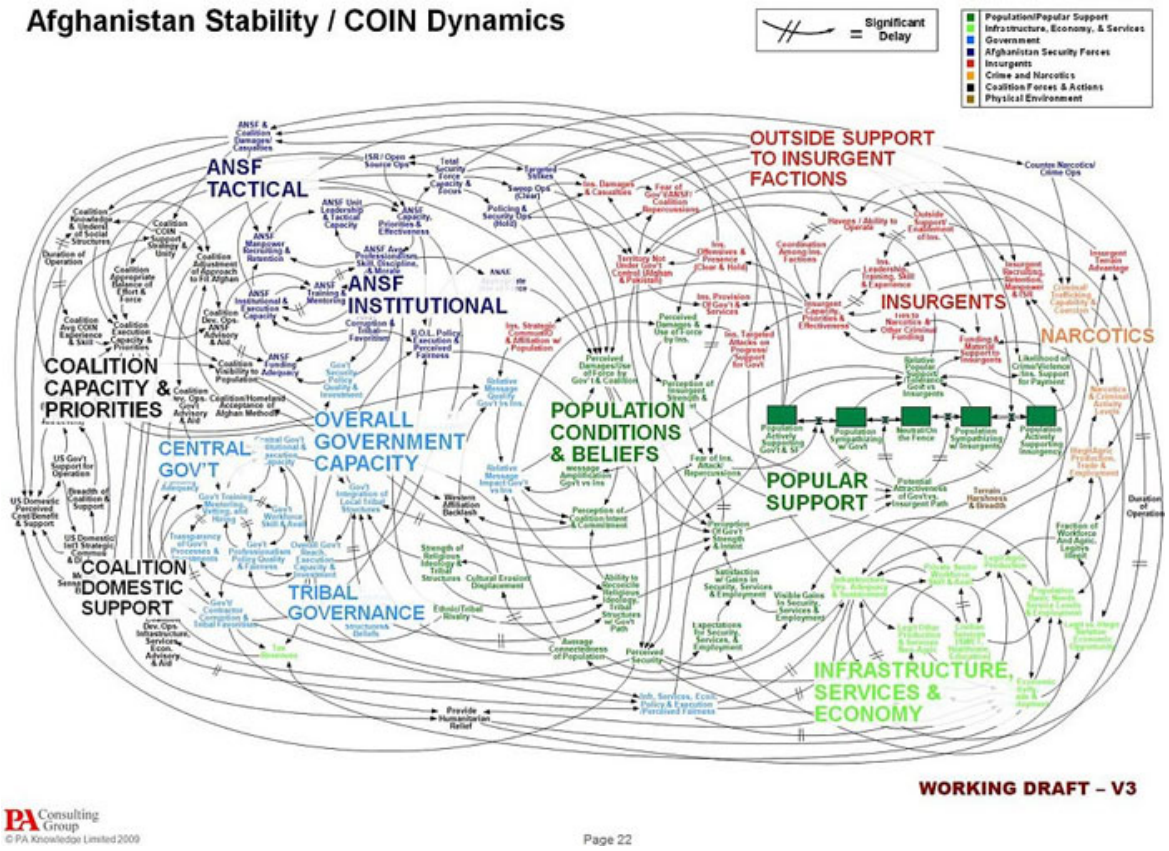
http://wpengine.com/blog/ugly-side-data-visualization/

Analyzing Visualizations

Professor Jeff Saltz

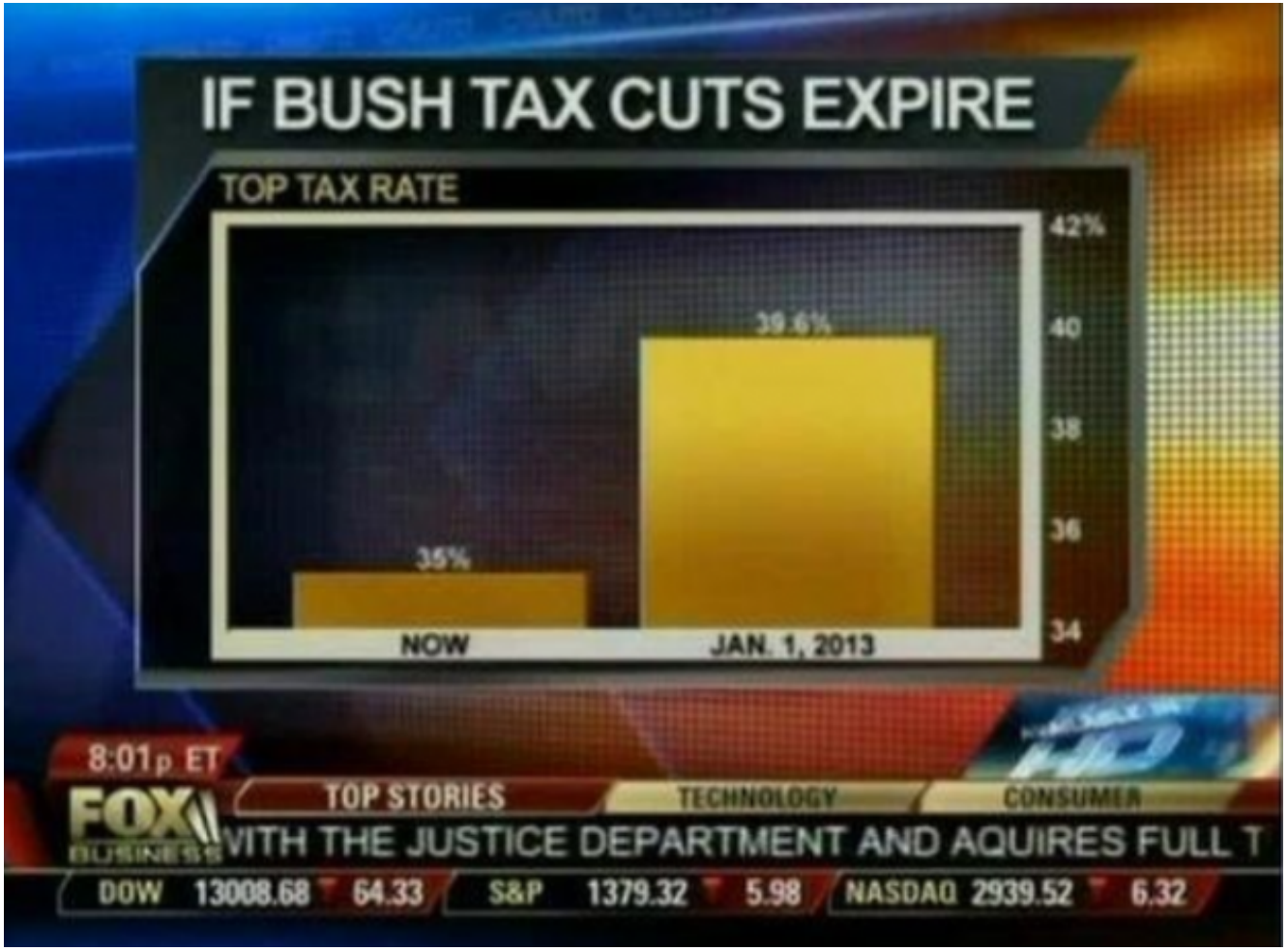
School of Information Studies
SYRACUSE UNIVERSITY

Too Much Information

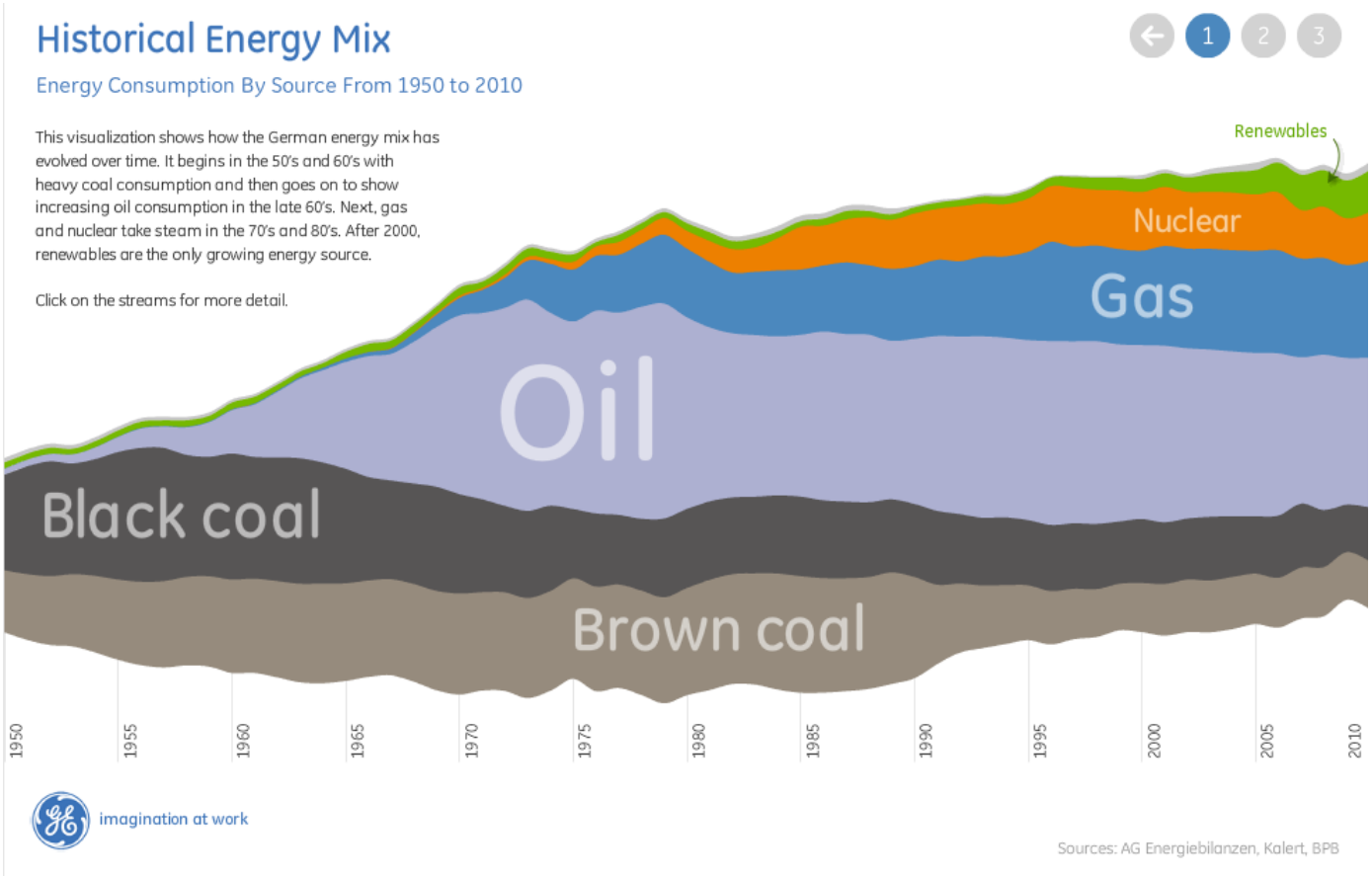


http://cdn.wpengine.com/wp-content/uploads/2014/03/Afghanistan-Stability-COIN-Dynamics.jpg

Simple Bar Chart

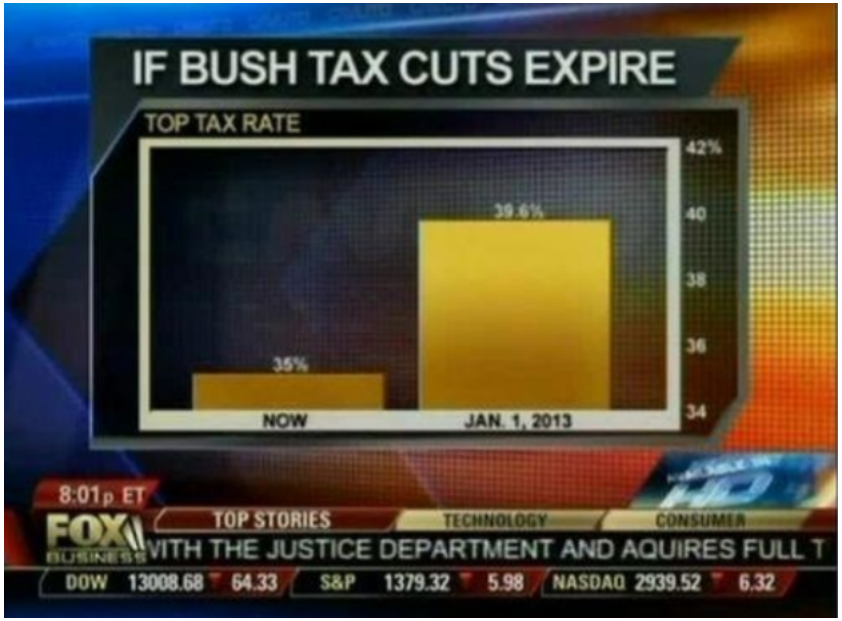


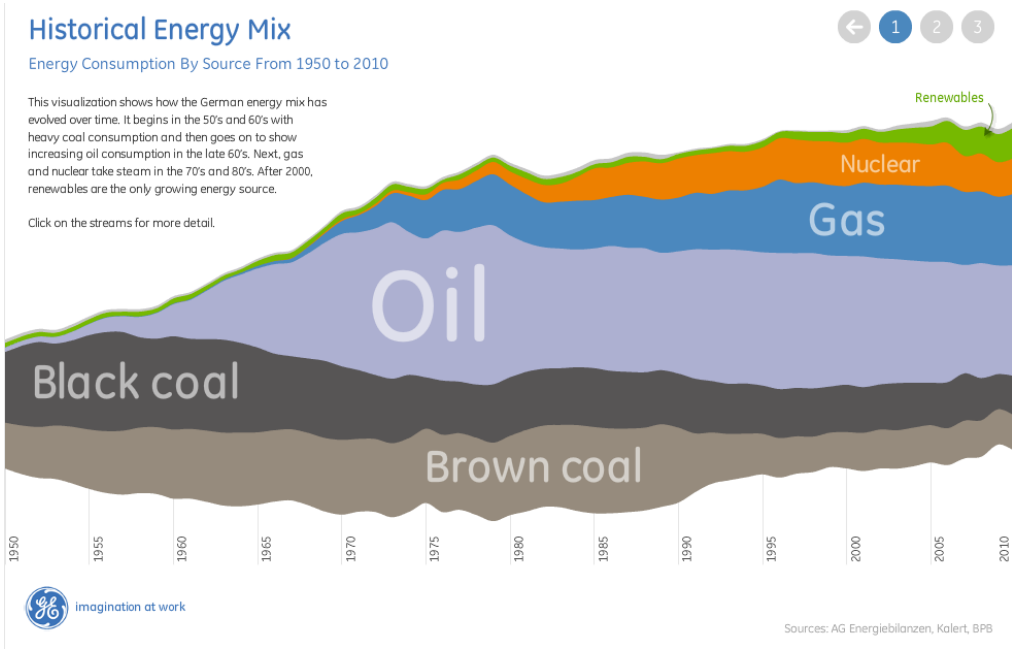
Oil Chart


Historical Energy Mix
Energy Consumption By Source From 1950 to 2010
This visualization shows how the German energy mix has evolved over time. It begins in the 50's and 60's with heavy coal consumption and then goes on to show increasing oil consumption in the late 60's. Next, gas and nuclear take steam in the 70's and 80's. After 2000, renewables are the only growing energy source.
Click on the streams for more detail.
1
2
3
Renewables
Nuclear
Gas
Oil
Black coal
Brown coal
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
imagination at work
Sources: AG Energiebilanzen, Kalert, BPB

Question

What is misleading or confusing about these two visualizations?





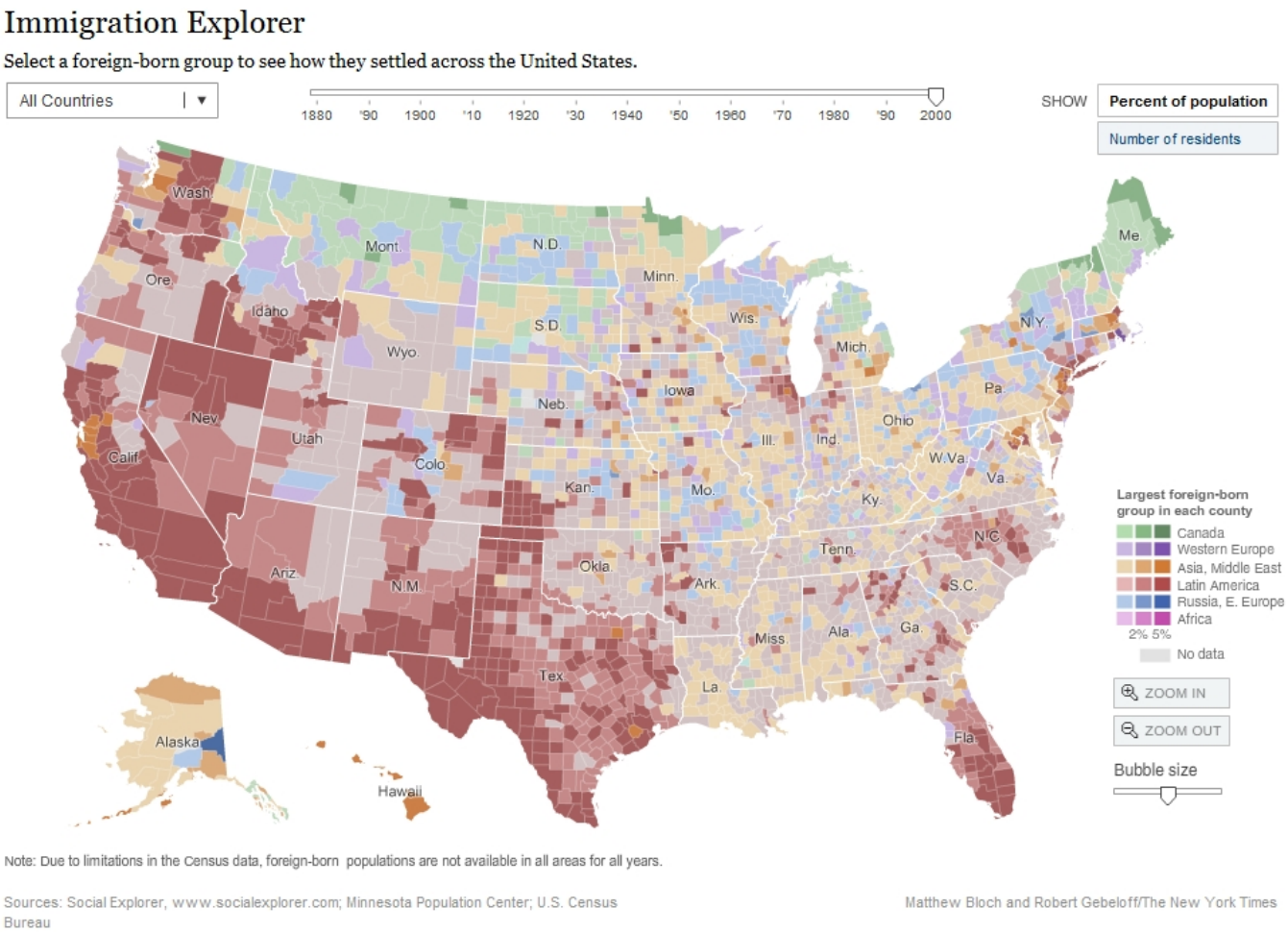
Play With Visualizations

Professor Jeff Saltz

School of Information Studies
SYRACUSE UNIVERSITY

Examples: Geo Data Mapping

[Demo]



Examples: Map Your Moves

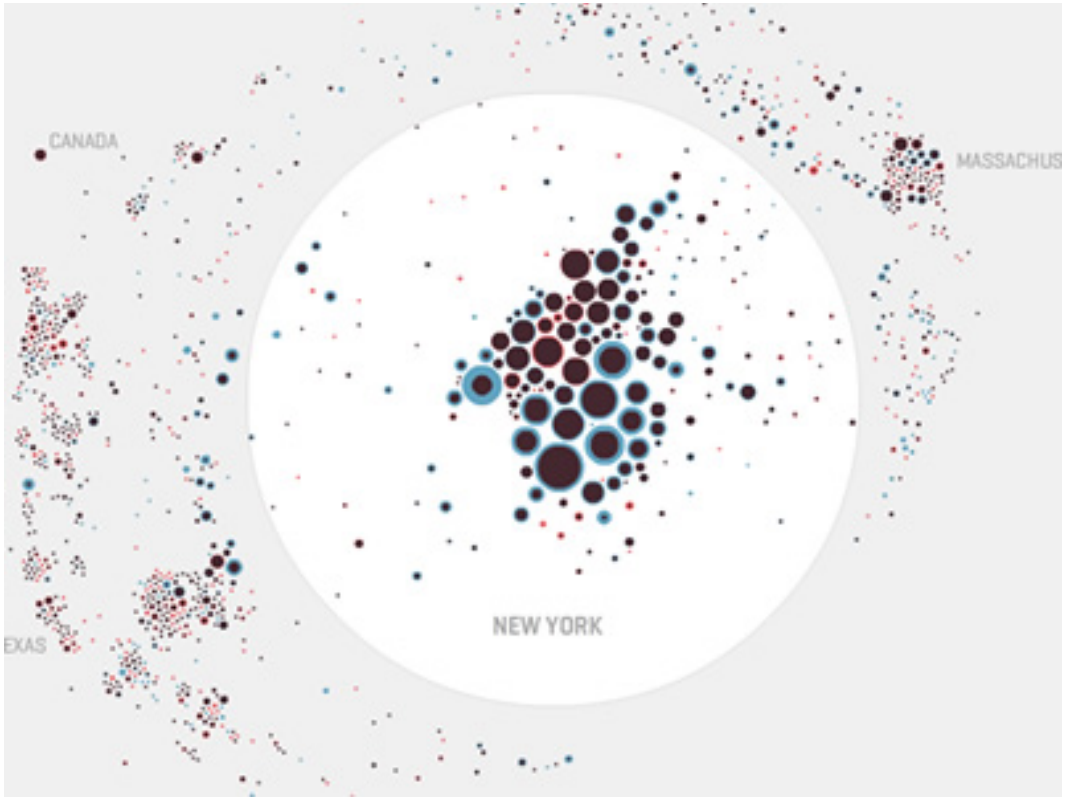
- Where New Yorkers move (10 years' data)

distorted map

circle = moves for
one zip code

red – out
blue – in

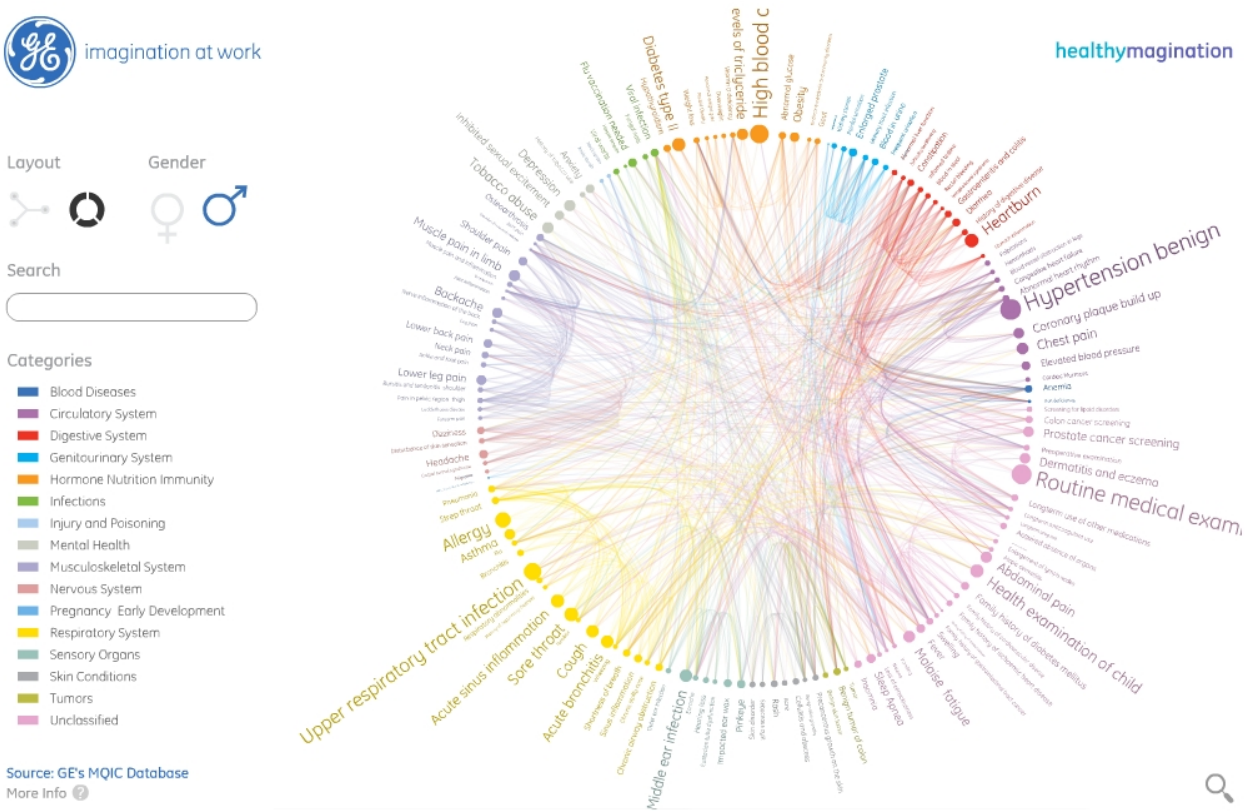
overlaid



Demo

Example: Circle Chart

Demo



Question: Which did you like best?

Why did you like it the best?

Could you suggest alternative visualizations?

Was the “interaction” helpful?

Introduction to Visualizations

Professor Jeff Saltz

School of Information Studies
SYRACUSE UNIVERSITY

What Is Visualization?

- “the communication of information using graphical representations” —Ward et al.
- “Transformation of the symbolic into the geometric” —McCormick et al.
- “…finding the artificial memory that best supports our natural means of perception” —Bertin

What Is Visualization? (cont.)

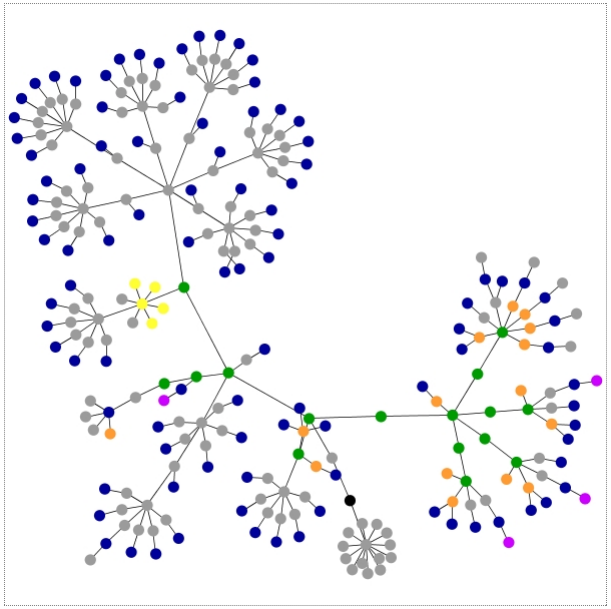
- *Information visualization* can be defined as the use of interactive visual representations of abstract data.

- Information visualization provides compact graphical presentations and user interfaces for interactively manipulating large numbers of items.

Information Visualization

- Information visualization: concerned with data that does not have a well-defined representation in 2D or 3D space (i.e., “abstract data”)

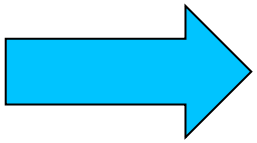
```
http://www.bu.edu/students/life/
http://www.bu.edu/students/life/housing/
http://www.bu.edu/students/life/dining-vending/
http://www.bu.edu/students/life/phone/
http://www.bu.edu/students/life/safety/
http://www.bu.edu/students/life/transportation/
http://www.bu.edu/students/life/activities/
http://www.bu.edu/students/life/bu-global-orientation-registration/
http://www.bu.edu/students/academics/
http://www.bu.edu/students/academics/link/
http://www.bu.edu/students/academics/admissions/
http://www.bu.edu/students/academics/registration/
http://www.bu.edu/students/academics/advising/
http://www.bu.edu/students/academics/grades/
http://www.bu.edu/students/academics/services/
http://www.bu.edu/students/academics/support/
http://www.bu.edu/students/health/
http://www.bu.edu/students/health/services/
http://www.bu.edu/students/health/counseling/
http://www.bu.edu/students/health/facilities/
http://www.bu.edu/students/health/clubsports/
http://www.bu.edu/students/health/varsitysports/
```

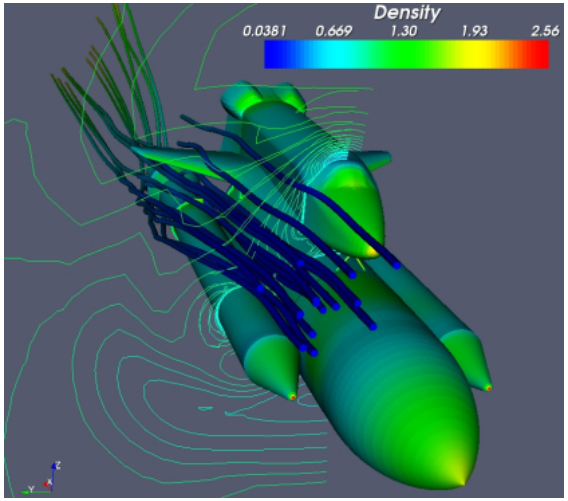


“Sci Viz” vs. “Info Viz”

- Visualization: converting raw data to a form that is viewable and understandable to humans
- Scientific visualization: specifically concerned with data that has a well-defined representation in 2D or 3D space (e.g., from simulation mesh or scanner)

```
0265640 132304 133732 032051 037334 024721 015013 052226 001662
0265660 025537 064663 054606 043244 074076 124153 135216 126614
0265700 144210 056426 044700 042650 165230 137037 003655 006254
0265720 134453 124327 176005 027034 107614 170774 073702 067274
0265740 072451 007735 147620 061064 157435 113057 155356 114603
0265760 107204 102316 171451 046040 120223 001774 030477 046673
0266000 171317 116055 155117 134444 167210 041405 147127 050505
0266020 004137 046472 124015 134360 173550 053517 044635 021135
0266040 070176 047705 113754 175477 105532 076515 177366 056333
0266060 041023 074017 127113 003214 037026 037640 066171 123424
0266100 067701 037406 140000 165341 072410 100032 125455 056646
0266120 006716 071402 055672 132571 105645 170073 050376 072117
0266140 024451 007424 114200 077733 024434 012546 172404 102345
0266160 040223 050170 055164 164634 047154 126525 112514 032315
0266200 016041 176055 042766 025015 176314 017234 110060 014515
0266220 117156 030746 154234 125001 151144 163706 136237 164376
0266240 137055 062276 161755 115466 005322 132567 073216 002655
0266260 171466 126161 117155 065763 016177 014460 112765 055527
0266300 003767 175367 104754 036436 172172 150750 043643 145410
0266320 072074 000007 040627 070652 173011 002151 125132 140214
0266340 060115 014356 015164 067027 120206 070242 033065 131334
0266360 170601 170106 040437 127277 124446 136631 041462 116321
0266400 020243 005602 004146 121574 124651 006634 071331 102070
0266420 157504 160307 166330 074251 024520 114433 167273 030635
0266440 133614 106171 144160 010652 007365 026416 160716 100413
0266460 026630 007210 000630 121224 076033 140764 000737 003276
0266500 114060 042647 104475 110537 066716 104754 075447 112254
0266520 030374 144251 077734 015157 002513 173526 035531 150003
0266540 146207 015135 024446 130101 072457 040764 165513 156412
0266560 166410 067251 156160 106406 136770 030516 064740 022032
0266600 142166 123707 175121 071170 076357 037233 031136 015232
0266620 075074 016744 044055 102230 110063 033350 052765 172463
```





*Adapted from The ParaView Tutorial, Moreland

Why Visualize?

- Vision is “highest bandwidth” sense
  - Fast and parallel
  - Preattentive processing
- Eye trained for pattern recognition
  - Scanning
  - Recognizing
  - Remembering images

Visualization Components

- Color
- Size
- Texture
- Proximity
- Annotation
- Interactivity
  - Selection/filtering
  - Zoom
  - Animation

GGPLOT

Professor Jeff Saltz

School of Information Studies
SYRACUSE UNIVERSITY

GGPLOT2

ggplot2 divides plot into three different fundamental parts:

Plot = Data + Aesthetics + Geometry

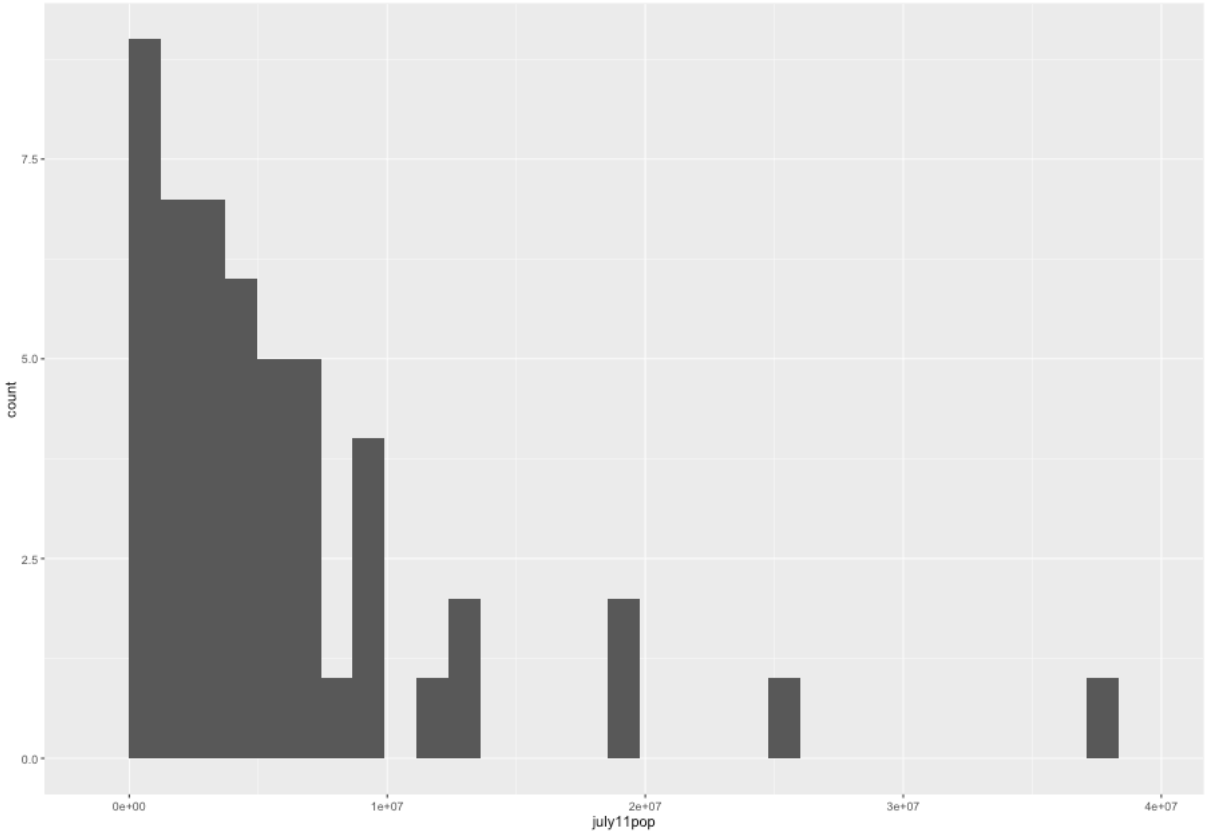
The principal components of every plot can be defined as follows:

- **Data** is a data frame.
- **Aesthetics** is used to indicate x and y variables. It can also be used to control the color, the size or the shape of points, the height of bars, etc.
- **Geometry** defines the type of graphics (histogram, box plot, line plot, density plot, dot plot, etc.)

GG stands for “grammar of graphics.”

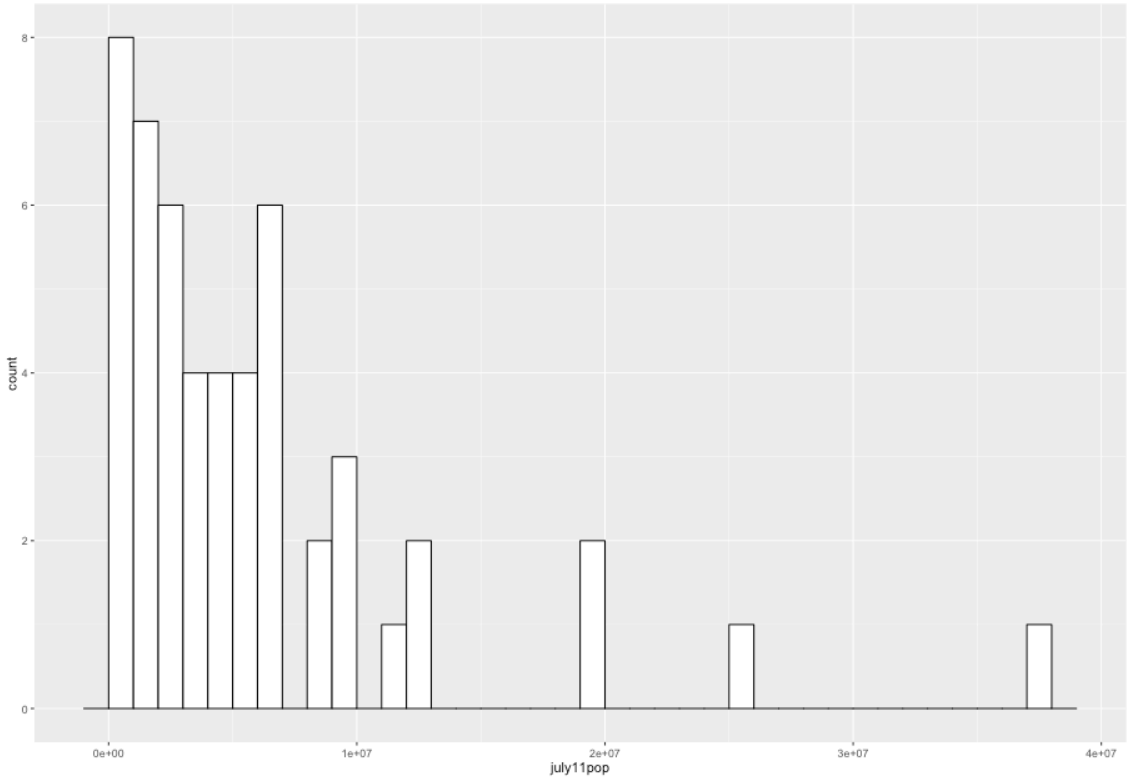
GGPLOT2: Histogram

```
ggplot(dfStates, aes(x=july11pop))      +
    geom_histogram(bins=30)
```



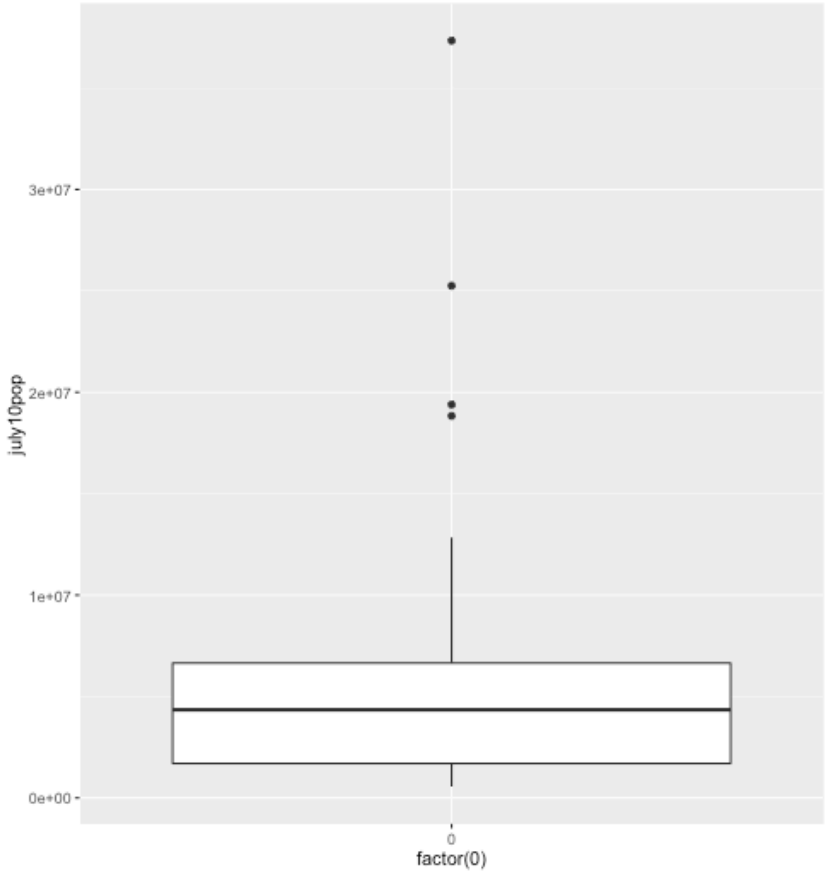
GGPLOT2: Histogram

```
ggplot(dfStates, aes(x=july11pop)) +
    geom_histogram(binwidth=1000000,
    color="black", fill="white")
```



GGPLOT2: Boxplot

```
ggplot(dfStates,aes(x=factor(0),july10pop))  +
    geom_boxplot()
```

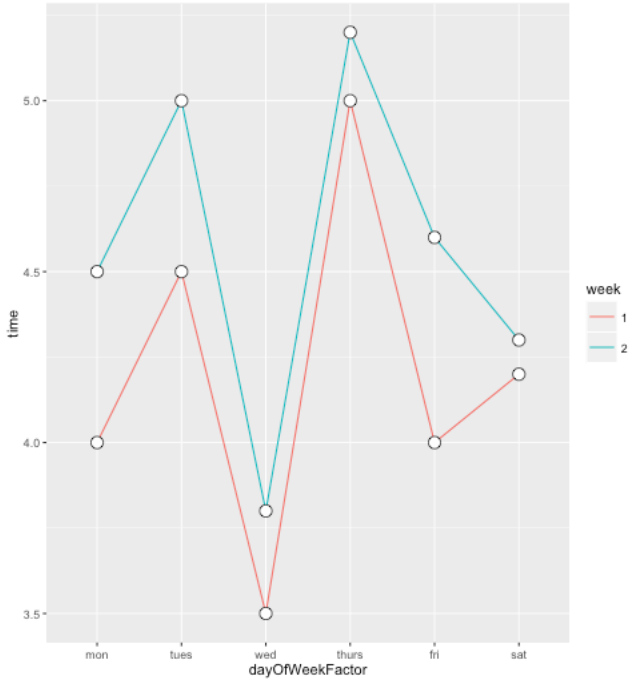


GGPLOT2: Timed Task

	dayOfWeek	time	week
1	mon	4.0	1
2	tues	4.5	1
3	wed	3.5	1
4	thurs	5.0	1
5	fri	4.0	1
6	sat	4.2	1
7	mon	4.5	2
8	tues	5.0	2
9	wed	3.8	2
10	thurs	5.2	2
11	fri	4.6	2
12	sat	4.3	2

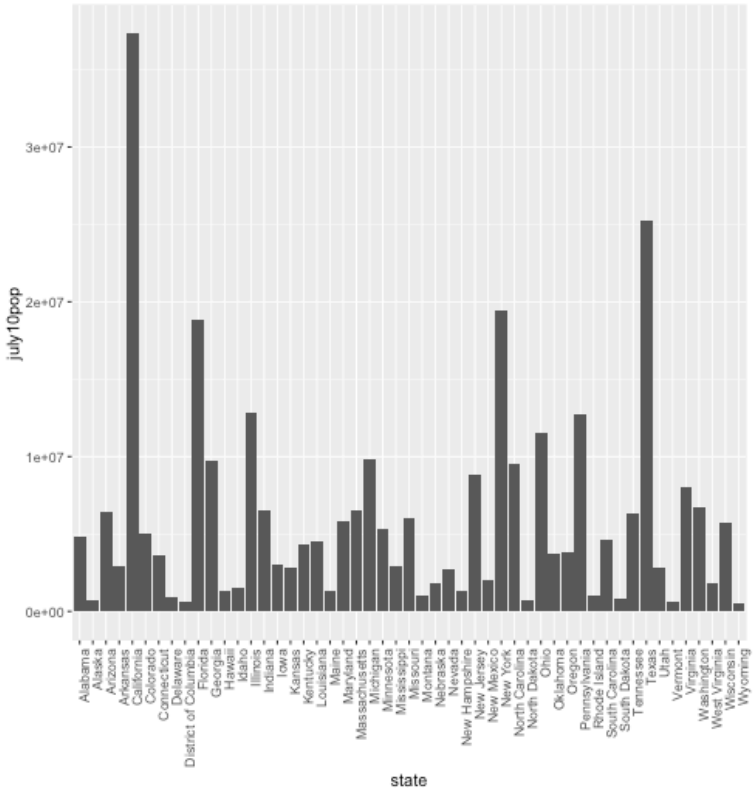
GGPLOT2: Lines

```
g <- ggplot(travel.df, aes(x=dayOfWeek,
          group=week, color=week))
        + geom_line(aes(y=time))

g <- g + geom_point(y=time, colour="black",
          size=4, shape=21, fill="white")

g <- g + ylab("time to NYC (in hours)") +
      ggtitle("compare weekly times")
g
```



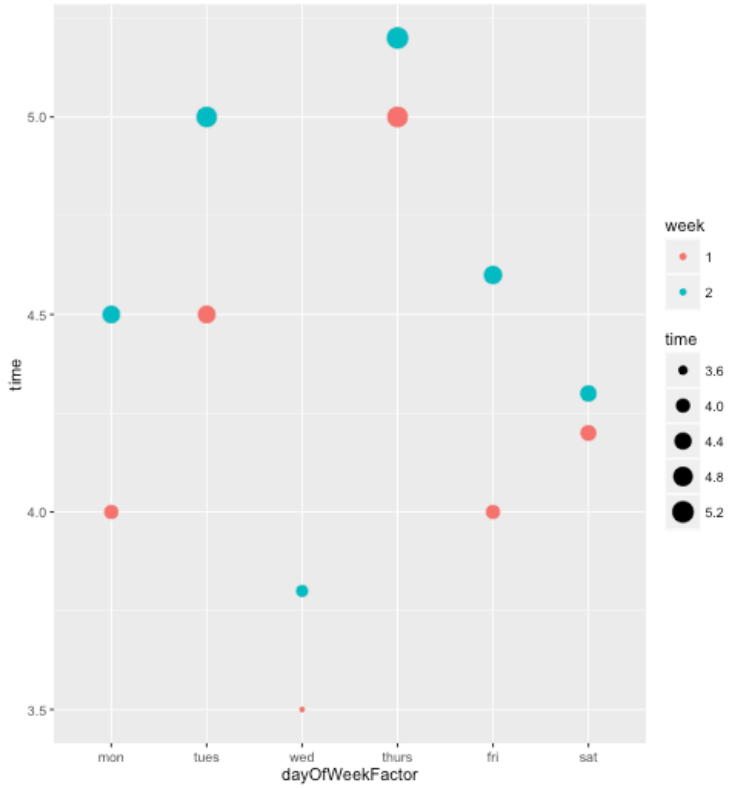
GGPLOT2: Bars

```
g <- ggplot(dfStates,
       aes(x=state, y=july10pop))
      + geom_bar(stat="identity")

g <- g + theme(axis.text.x =
      element_text(angle = 90, hjust = 1))

g
```



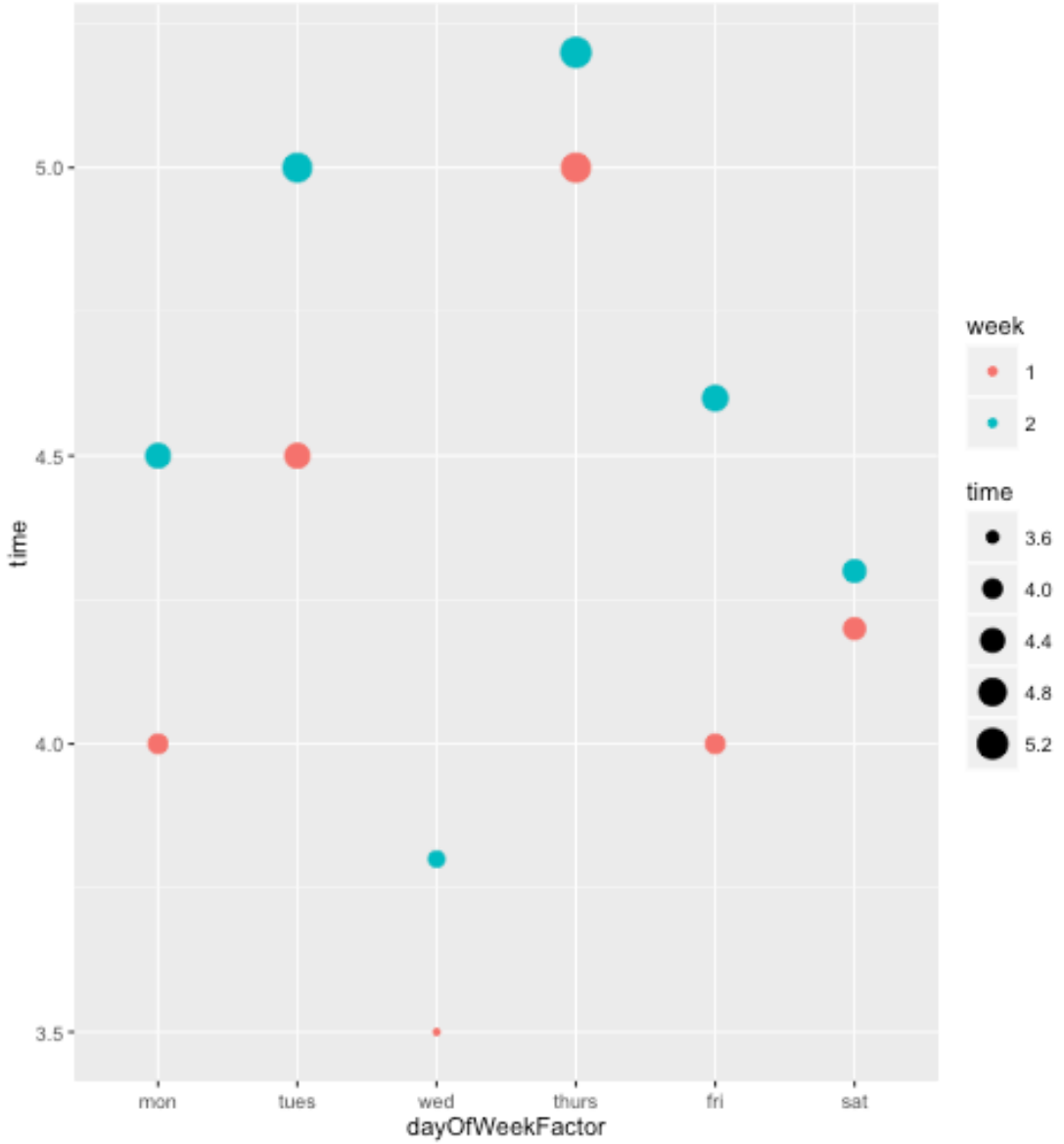
GGPLOT2: Scatter

```
ggplot(travel.df,
    aes(x=dayOfWeek, y=time))
     + geom_point(
          aes(size=time,
               color=week))
```



GGPLOT2: Scatter

```
ggplot(travel.df,  aes(x=dayOfWeek, y=tim
+ geom_point(aes(size=time,  color=week)
```



Ten Principles in Visualization

Professor Jeff Saltz

School of Information Studies
SYRACUSE UNIVERSITY

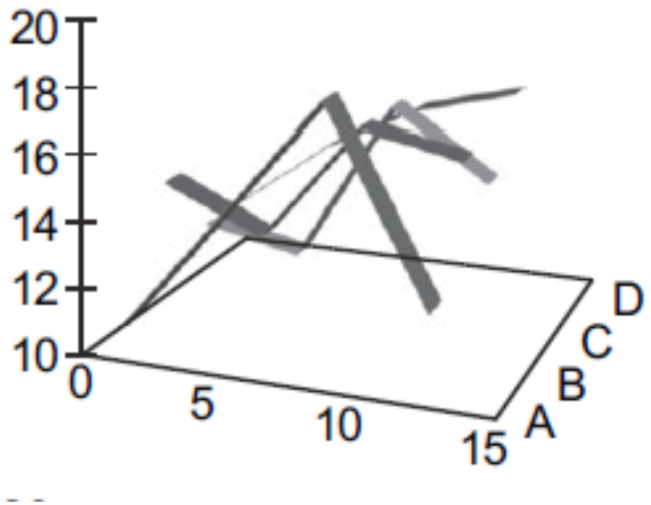
Ten Principles in Visualization

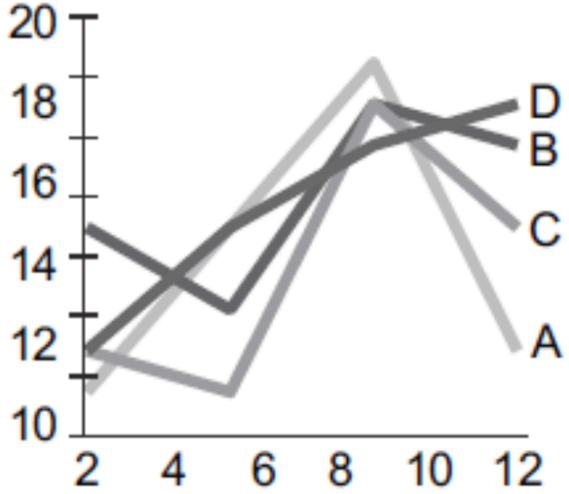
- Principles are fairly well known

→ Adapted from Kelleher and Wagener (2011)

1. Simplicity

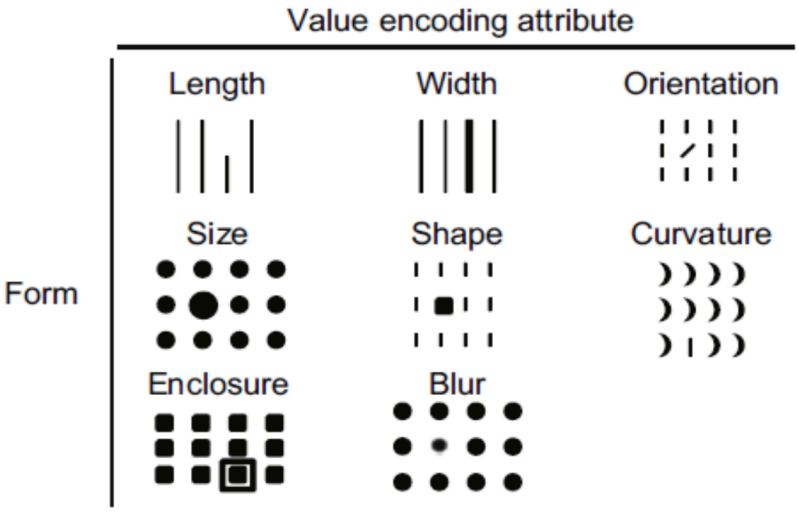
- Create the simplest graph that conveys the information you want to convey (Tufte).

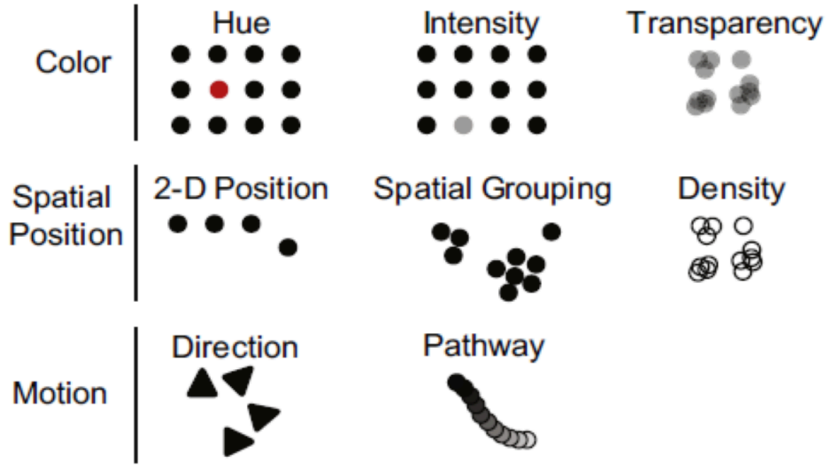




2. Encoding

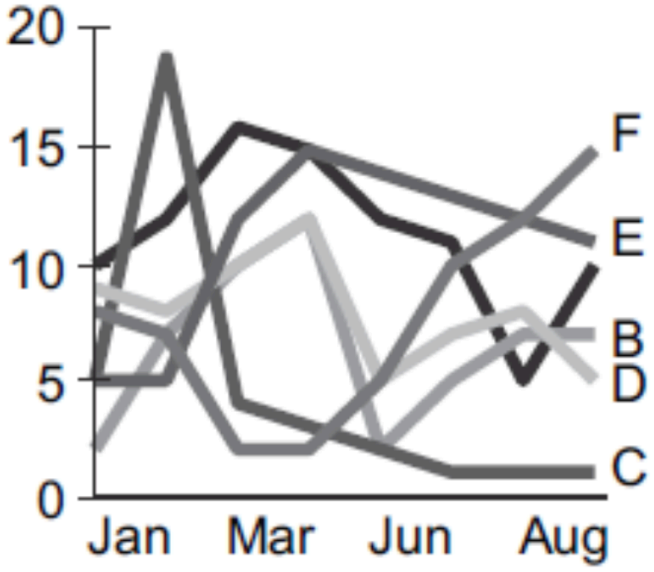
- Consider the type of encoding object and attribute used to create a lot (Chambers, 1983; Cleveland & McGill, 1984).

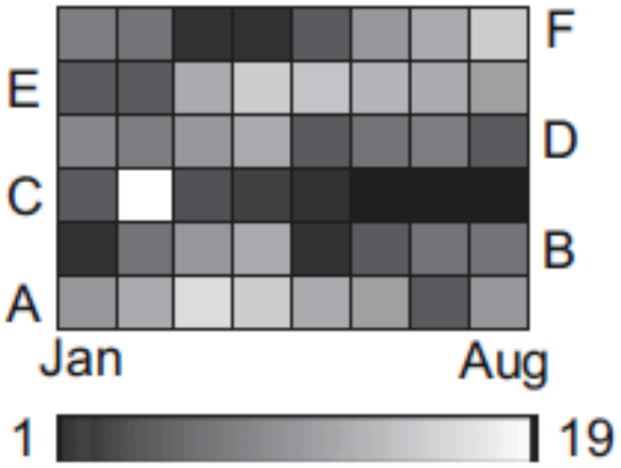




3. Patterns vs. Details

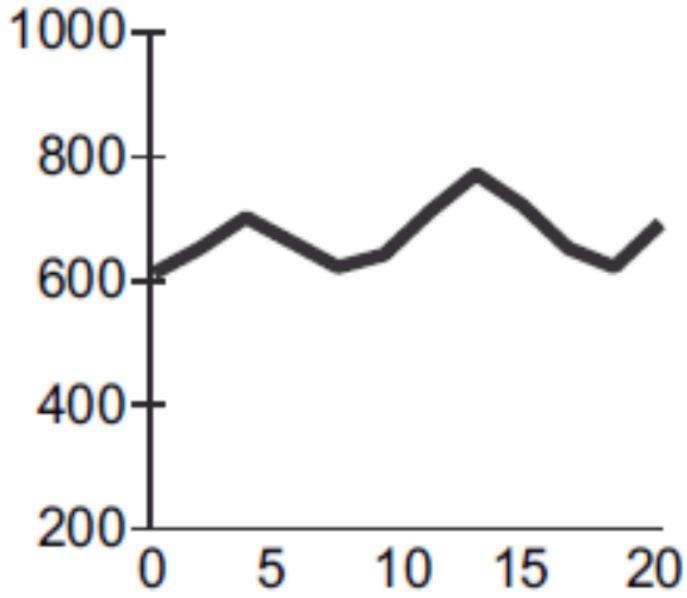
- Focus on visualizing patterns **or** on visualizing details (Few, 2004; Kosslyn & Chabris, 1992).

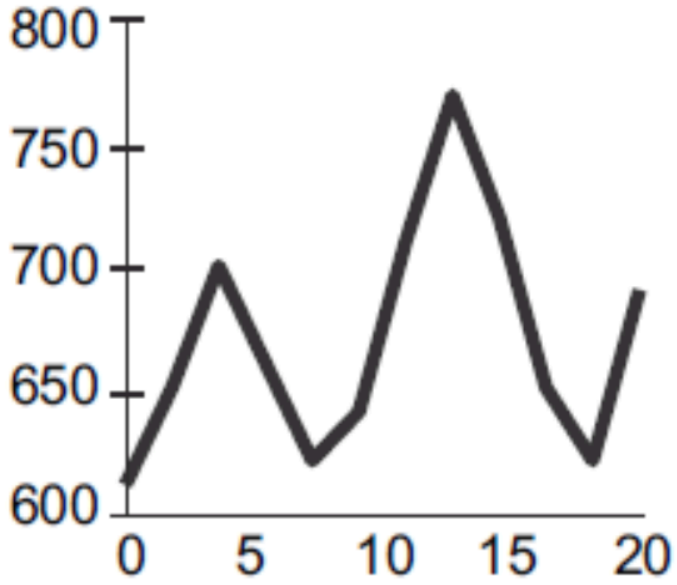




4. Ranges

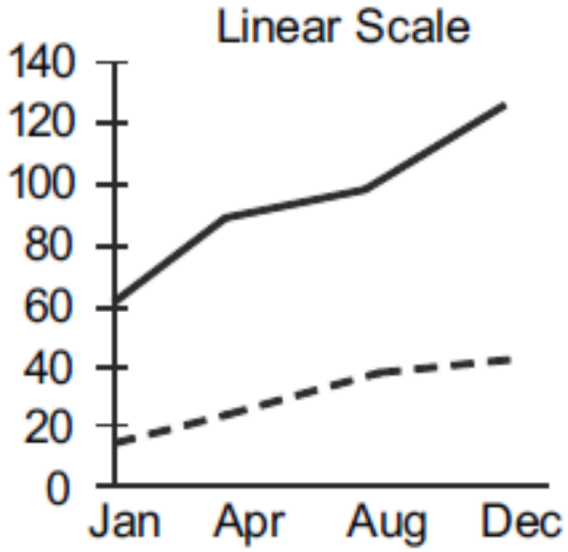
- Select meaningful axis ranges (Robbins,2005; Tufte, 2006; Strange, 2007)

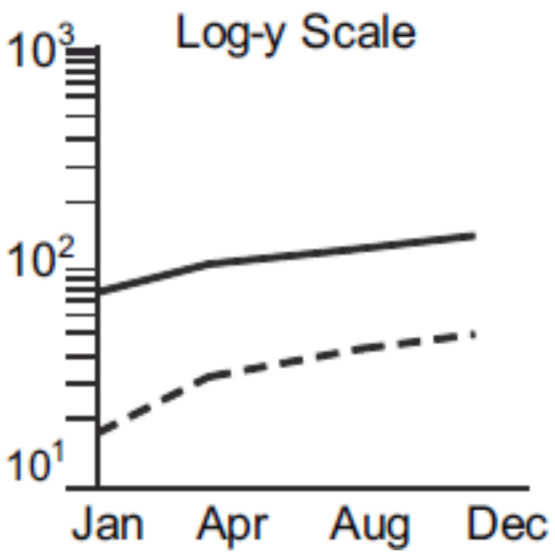




5. Transformations

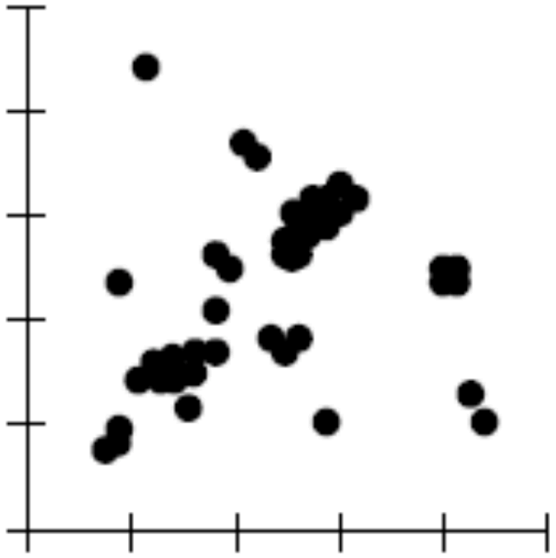
- Data transformations and carefully chosen aspect ratios can be used to emphasize rates of change for time-series data (Cleveland, 1994 [p. 66, 95, 103]).

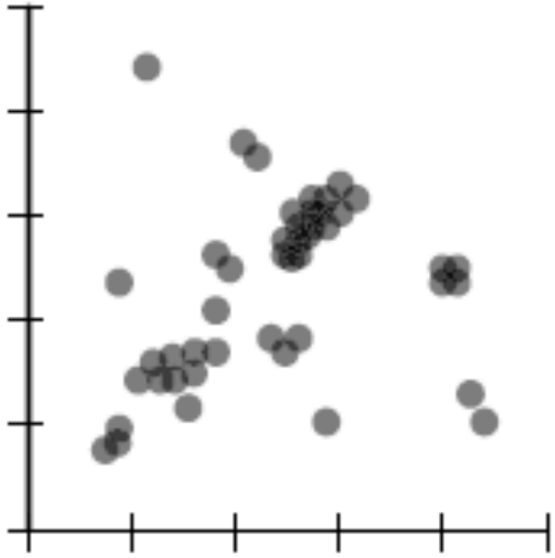




6. Show Density

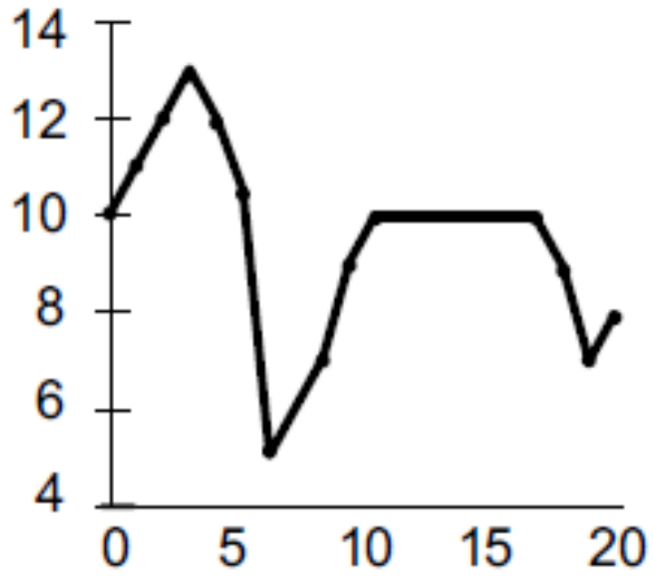
- Plot overlapping points in a way that **density difference become apparent** (Few, 2009; Cleveland, 1994).

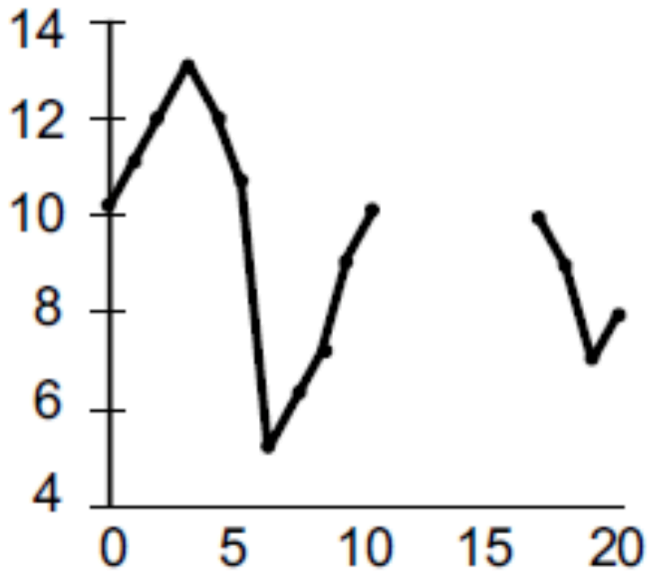




7. Connections

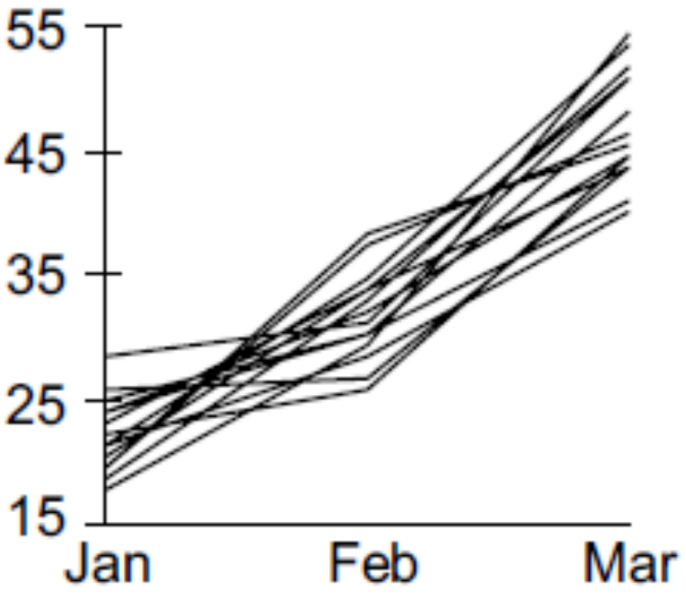
- Use lines when connecting sequential data in time-series plots (Strange, 2007).

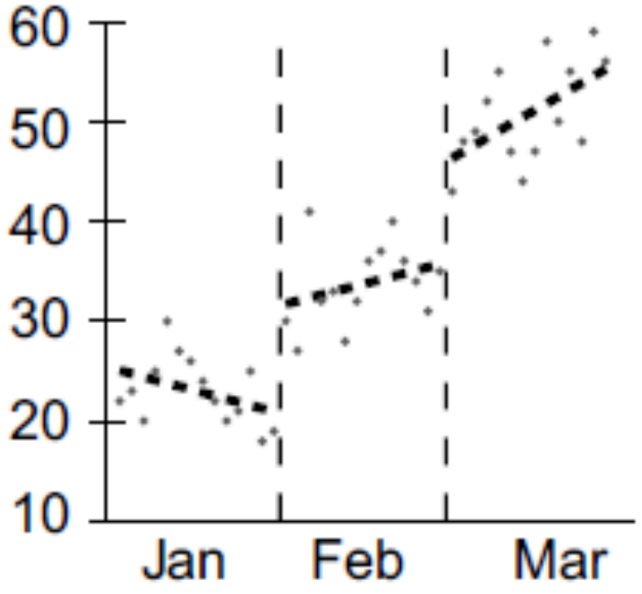




8. Aggregates

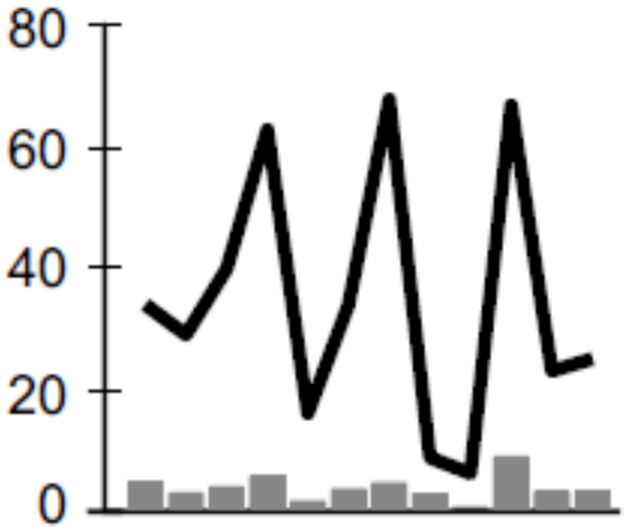
- Aggregate larger data sets in meaningful ways (Cleveland & Devlin, 1980; Chambers, 1983)

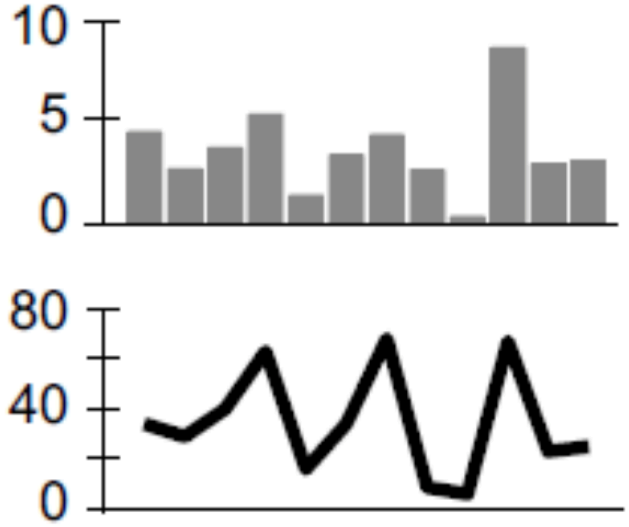




9. Comparison

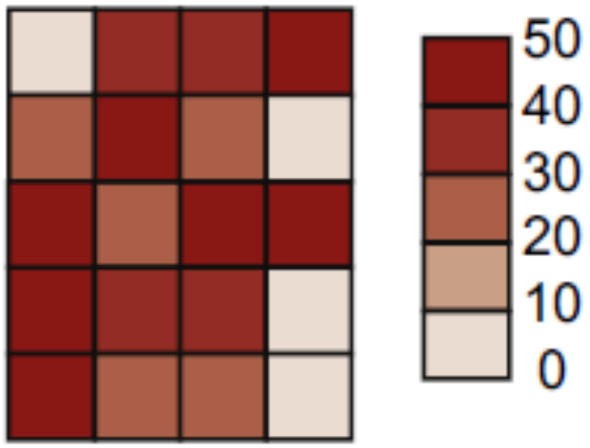
- **Keep axis ranges as similar as possible**
  → to compare variables
  (Cleveland, 1994; Few, 2009).

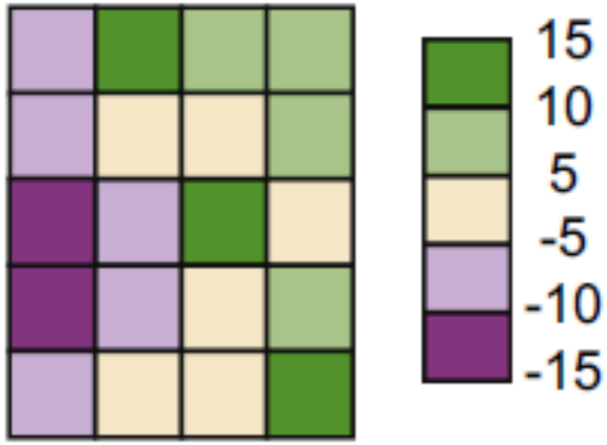




10. Colors

- Select an appropriate color scheme based on the type of data (Brewer, 1994; Harrower and Brewer, 2003).






SYRACUSE UNIVERSITY
SUOS
CULTORES
SCIENTIA
CORONAT