

INFM 603: Information Technology and Organizational Context

# Session 9: Visualization



Jimmy Lin  
The iSchool  
University of Maryland

Thursday, November 1, 2012

# Visualization

- Why visualizations?
  - Human perceptual skills are remarkable
  - Human image storage storage is fast and vast
- What are visualizations for?
  - A good visualization tells a story
  - A great visualization delivers insight
- Science as well as an art...

# Raw Ingredients

- Types of data
  - Categorical data
  - Ordinal data
  - Continuous data
- Ways of representing data
  - Height
  - Width
  - Thickness
  - Area
  - Angles
  - Size
  - Shape
  - Color



Panda Viewing  
Sun - 5:30pm



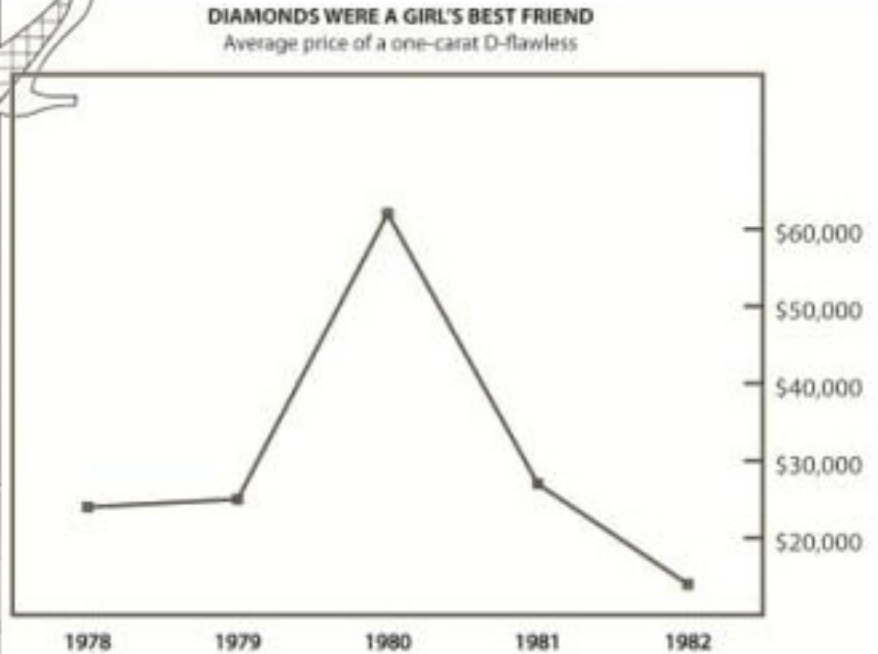
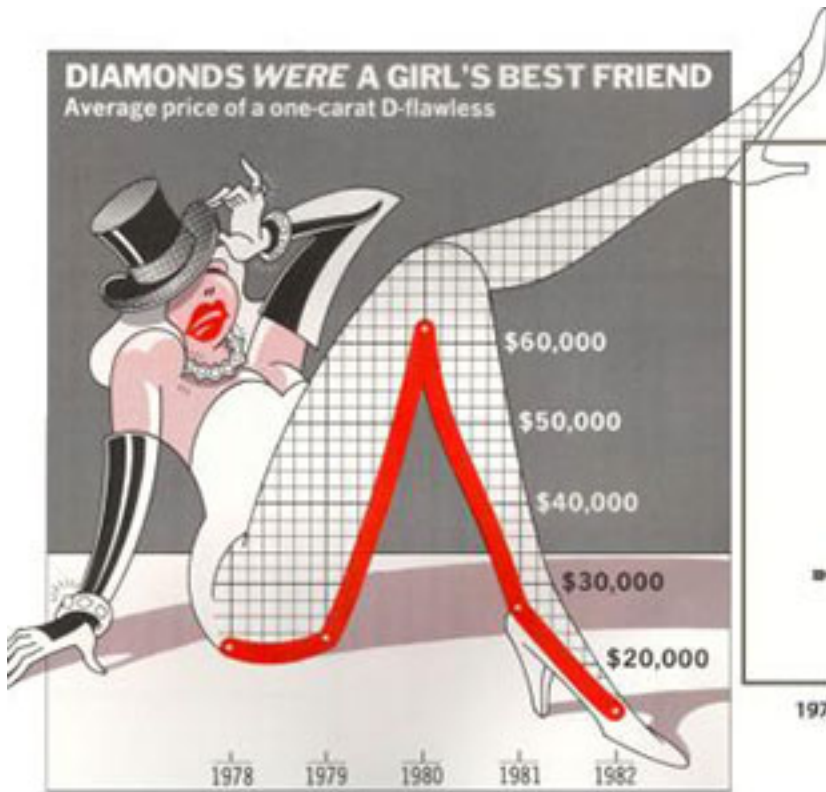


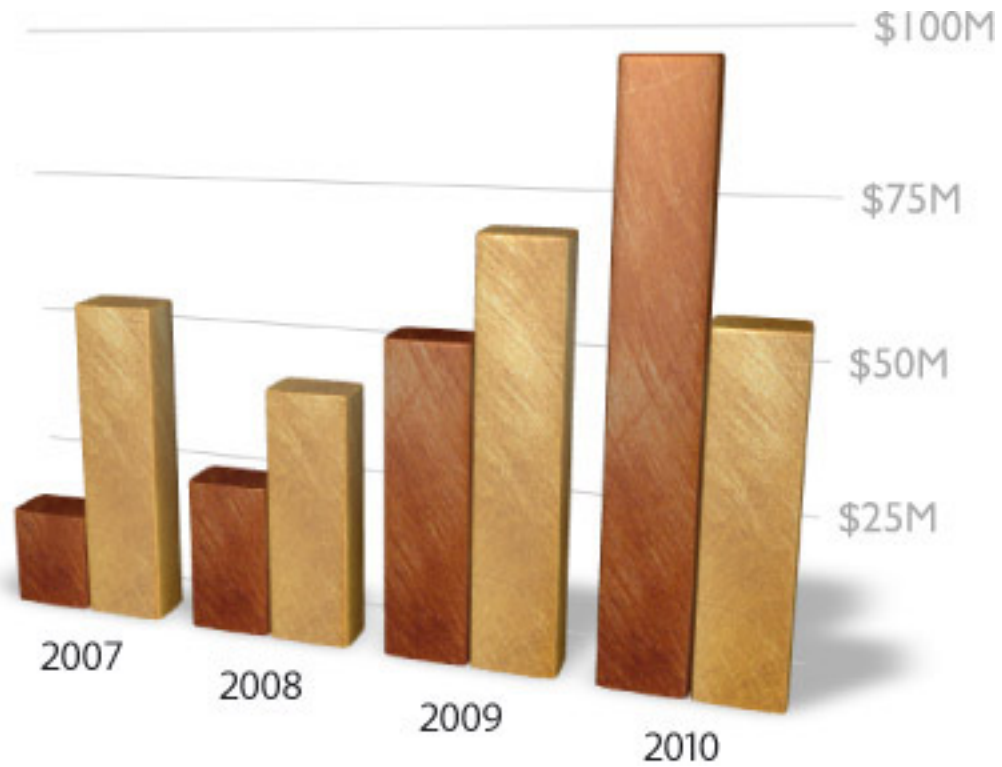
# Static vs. Interactive

- Interactive visualizations allow users to “explore” the data
  - Overview
  - Zoom and filter
  - Details-on-demand
- Brushing and linking is a powerful technique

# Info Porn

- Informative visualizations vs. info porn
- Beware of infographics and chart junk

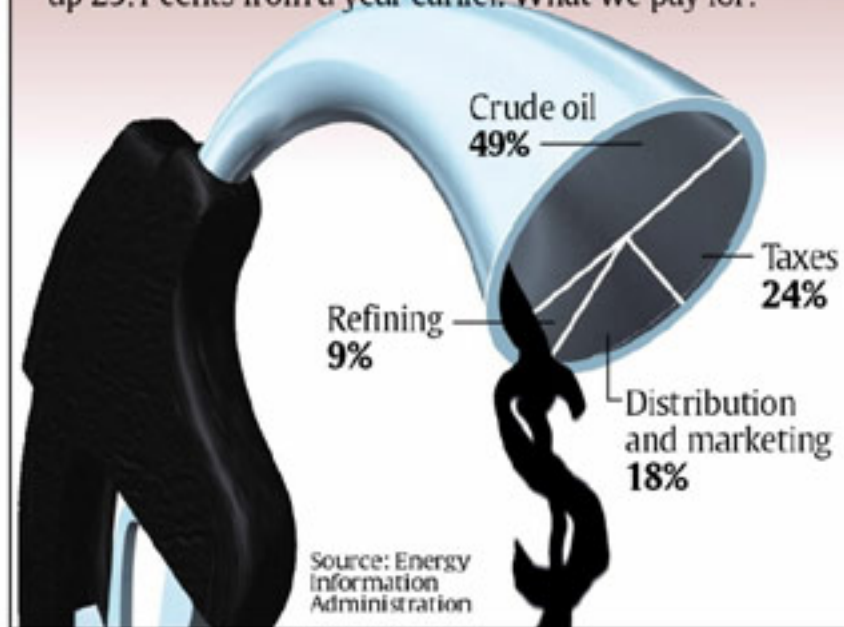




## USA TODAY Snapshots®

### What gas bucks buy

The average price of regular gas climbed to \$1.85 per gallon as of Jan. 24, up 3.4 cents from a week earlier and up 23.1 cents from a year earlier. What we pay for:



By Shannon Reilly and Marcy E. Mullins, USA TODAY



# TRACK & FIELD: THE BEST RESULTS

Technical innovations, rule changes and the professionalization of sports brought new records in the 20th century. How much longer can athletes keep improving?

## 100 METERS

**+19%**  
**+11%**

THE LAST RECORD WAS SET IN 1993 BY FLORENCE GRIFFITH-JOYNER

USAIN BOLT PULVERIZED RECORDS IN 2008. WILL HIS RECORD EVENTUALLY BE BEATEN?

140's  
130's  
120's  
110's  
100's  
90's

### RACE TIME

THE GRAPHIC SHOWS THE EVOLUTION OF THE 10 BEST 100M PERFORMANCES



## HIGH JUMP

**+30%**

PERFORMANCE HAS LEVELLED SINCE THE 1990s

2.45m  
JAVIER SOTOMAYOR  
CUB  
1993

## POLE VAULT

**+71%**

PERFORMANCES HAVE DECLINED SINCE 2000

6.14m  
SERGEY BUBKA  
UKR  
1984

13.69s  
FLORENCE GRIFFITH-JOYNER  
USA  
1988

9.95s  
JIM RINGS  
USA  
1968

9.85s  
BEN JOHNSON  
USA  
1988

9.79s  
MAURICE GREEN  
USA  
1993

9.74s  
ASAFU POWELL  
JAM  
2007

9.58s  
USAIN BOLT  
JAM  
2008

1984-1988 Sergey Bubka set 2 world records

1988 The first fiberglass pole vault

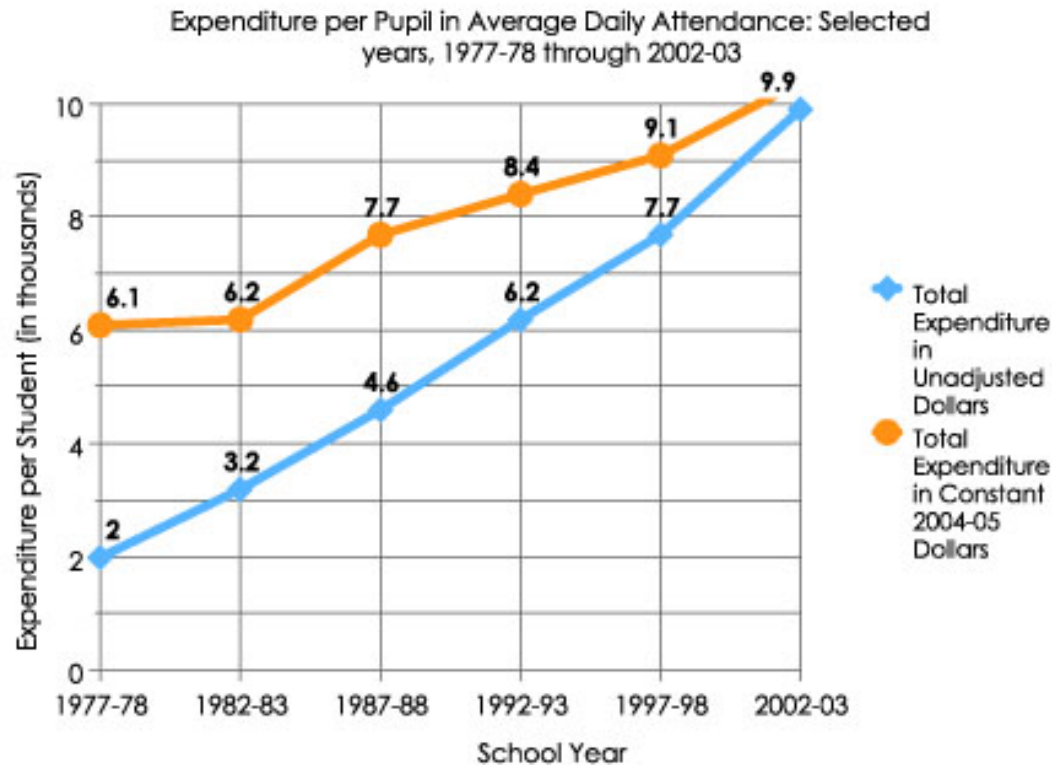
1993 The first pole vault over 6m

2008 The first 100m gold medalist who broke the 10s barrier

[http://junkcharts.typepad.com/junk\\_charts/music/](http://junkcharts.typepad.com/junk_charts/music/)  
Compare: New York Times interactive visualization

# Line Graphs (+ Variants)

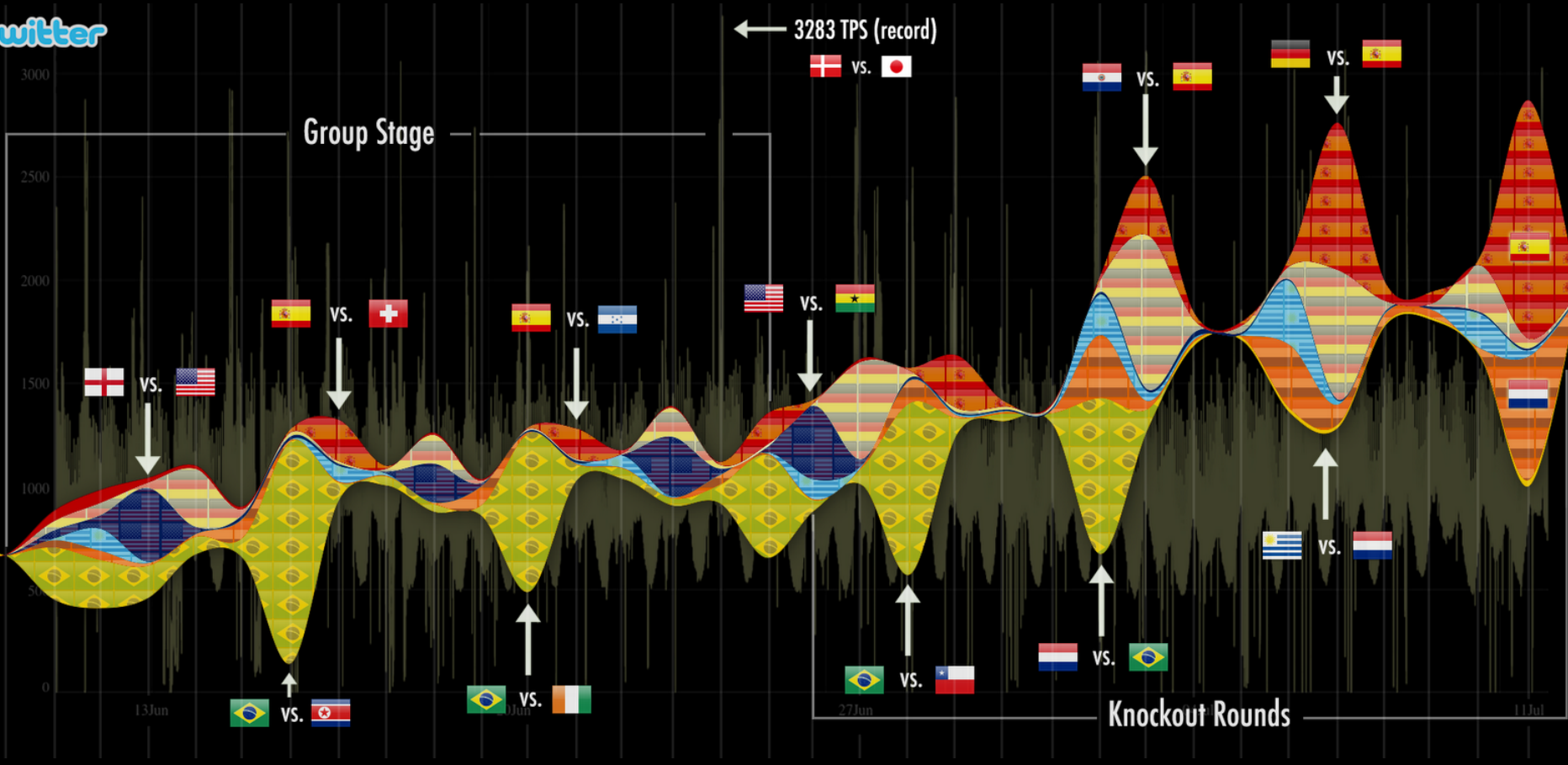
- Suitable for what types of data?
- What makes a good line graph?



The NCES Common Core of Data (CCD) 2004-2005

# Stacked Graphs

- Multiple time-series data “stacked” on top of each other
  - Useful for showing aggregates
- Sometimes called “stream graphs”

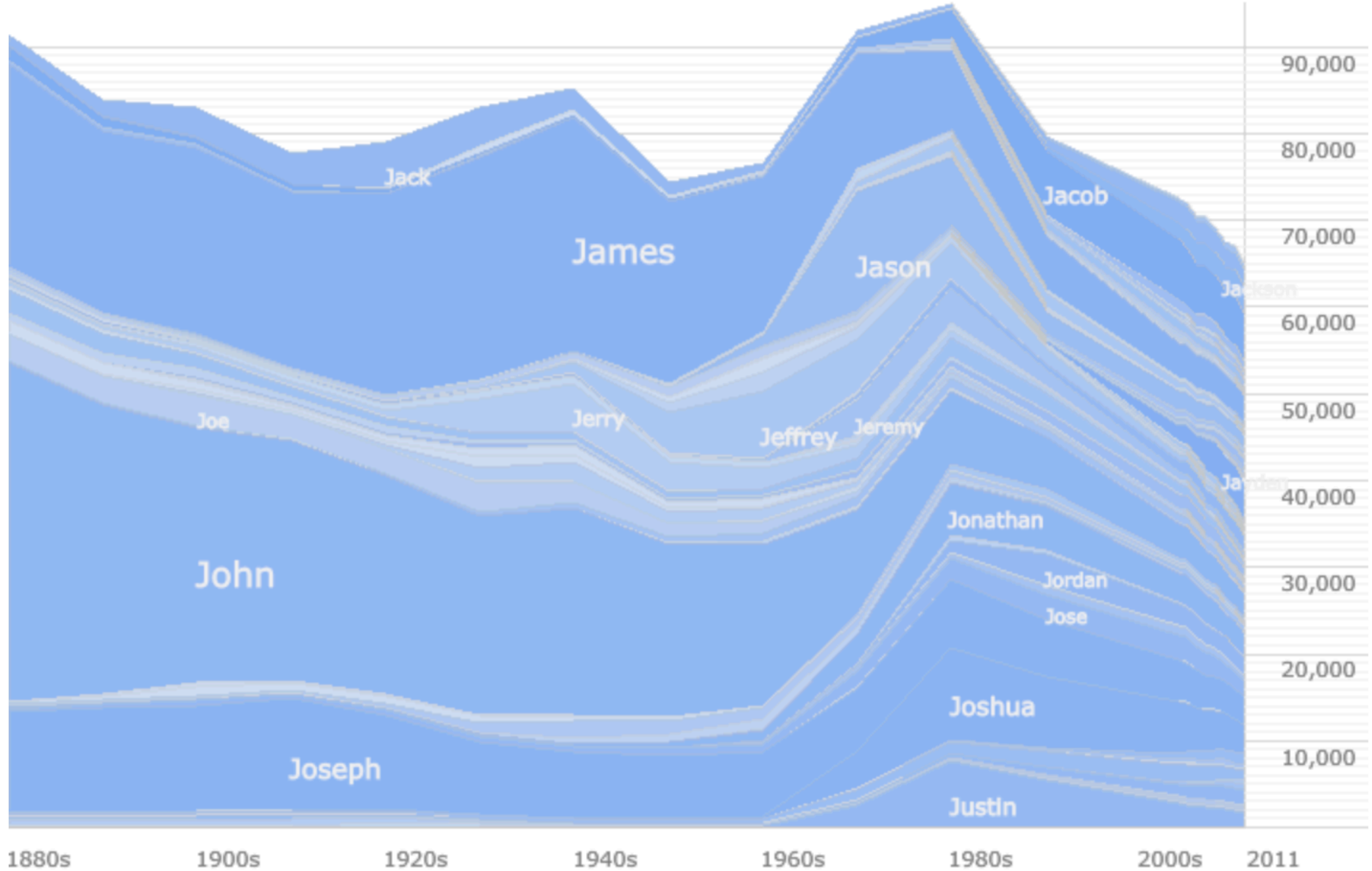


Baby Name >

Both  Boys  Girls

2010 rank: boys	1000	500	100	25	1
girls	1000	500	100	25	1

Names starting with 'J' per million babies

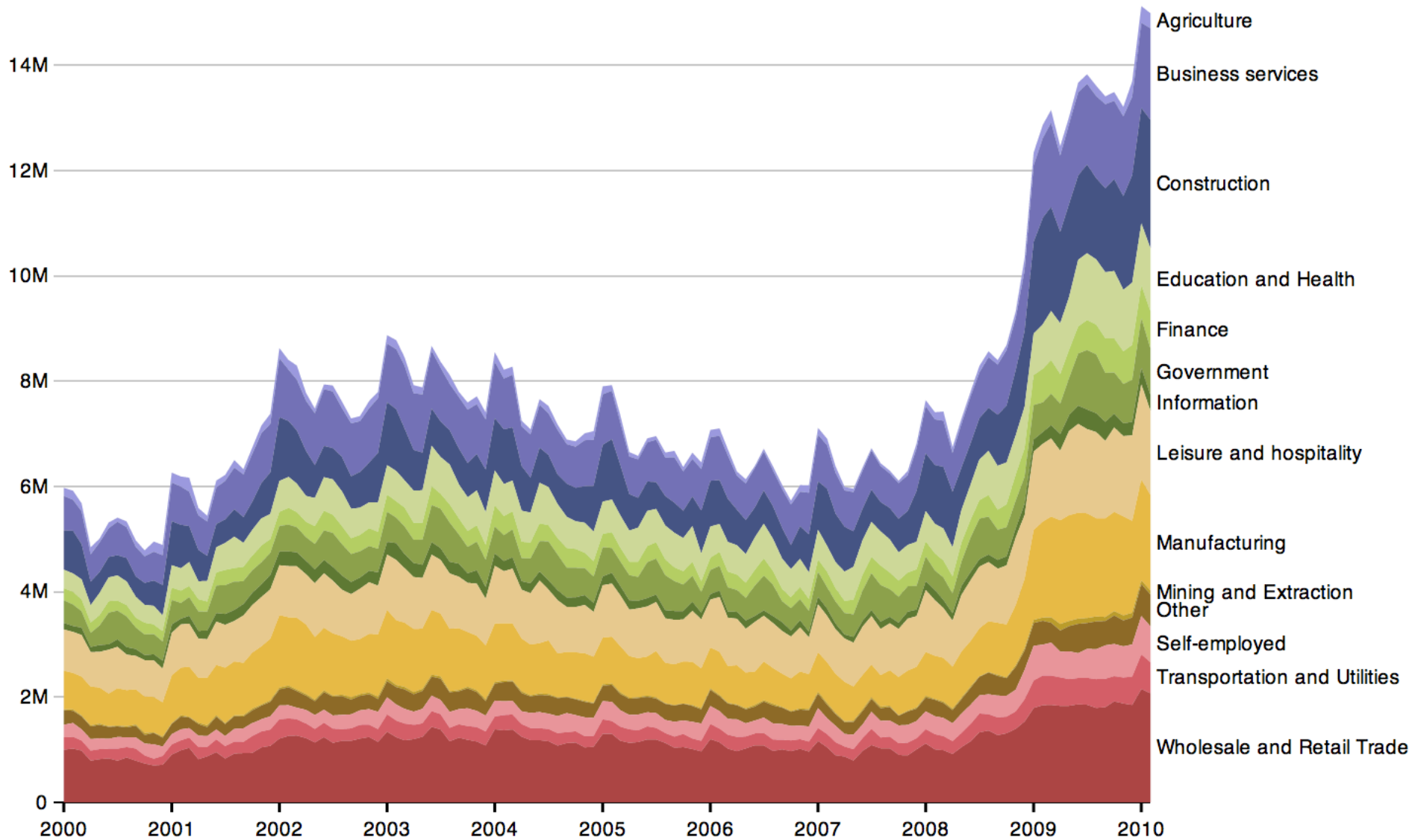


<http://www.babynamewizard.com/voyager#>



# Stacked Graph of Unemployed U.S. Workers by Industry

View:

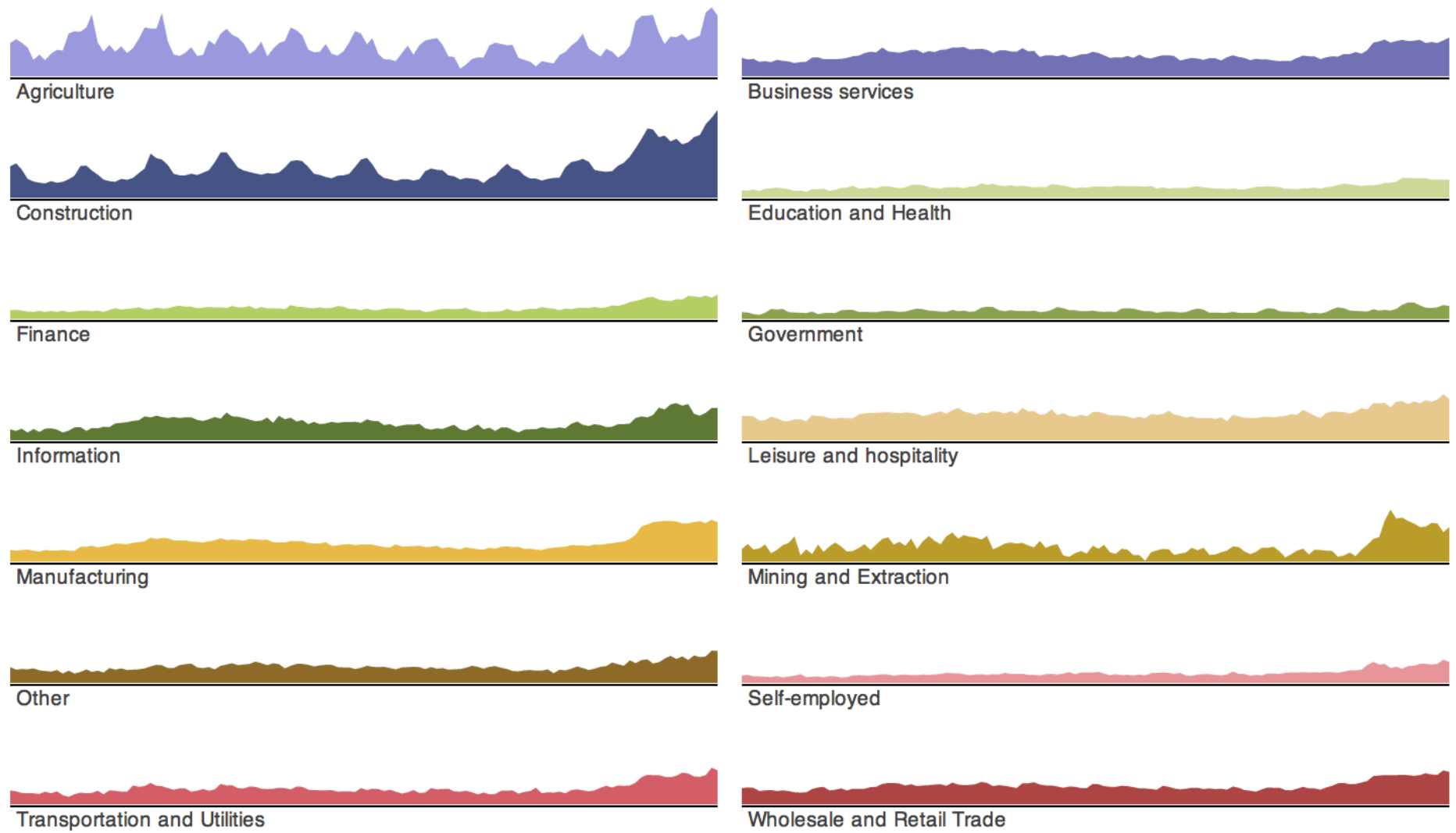


Total counts of unemployed persons per industry, 2000-2010.

Source: [U.S. Bureau of Labor Statistics](http://www.bls.gov)

<http://hci.stanford.edu/jheer/files/zoo/>

# Unemployment Rate of U.S. Workers by Industry, 2000-2010



Unemployment rates of U.S. workers per industry.

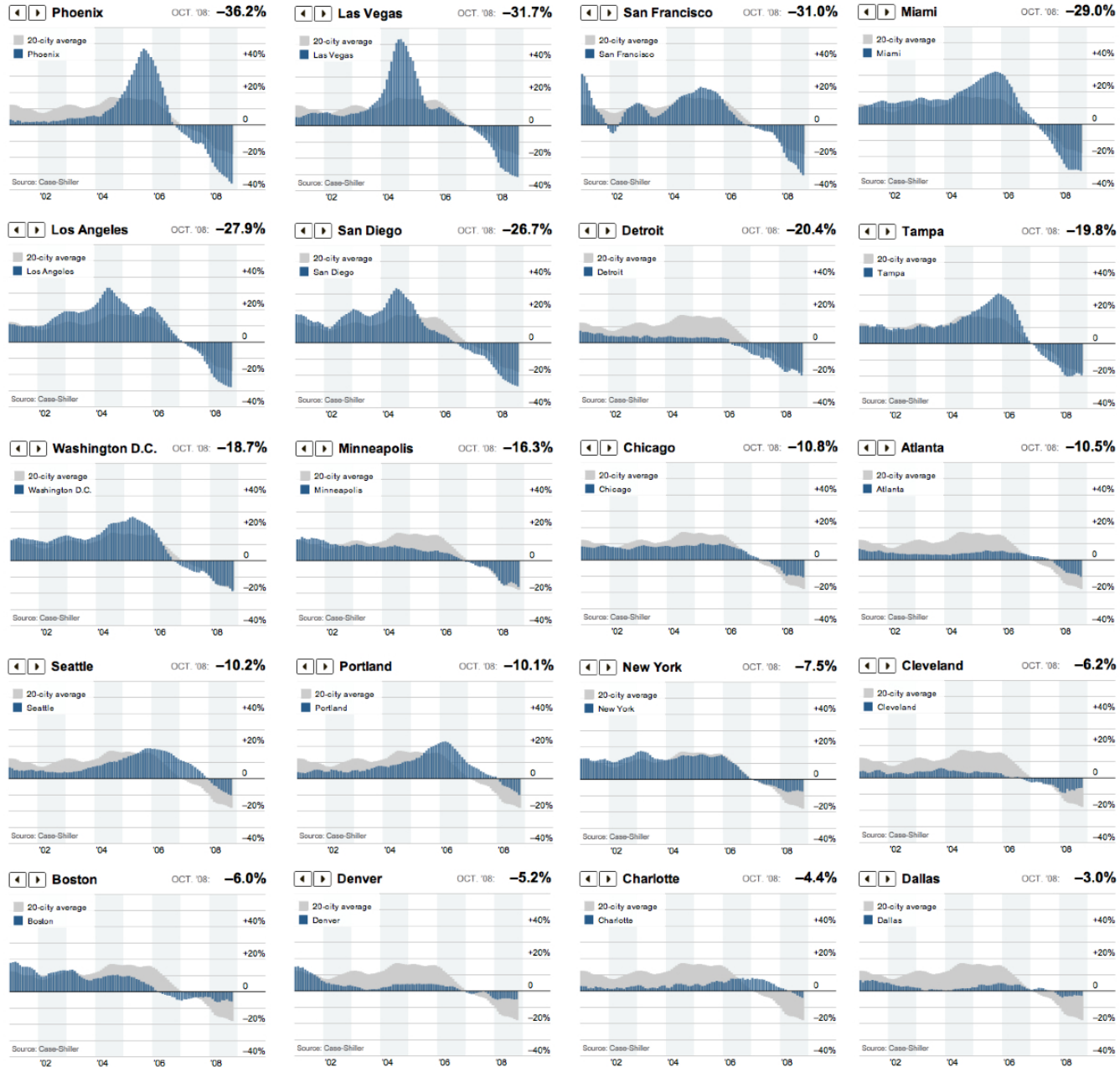
Source: [U.S. Bureau of Labor Statistics](http://www.bls.gov)

# Small Multiples

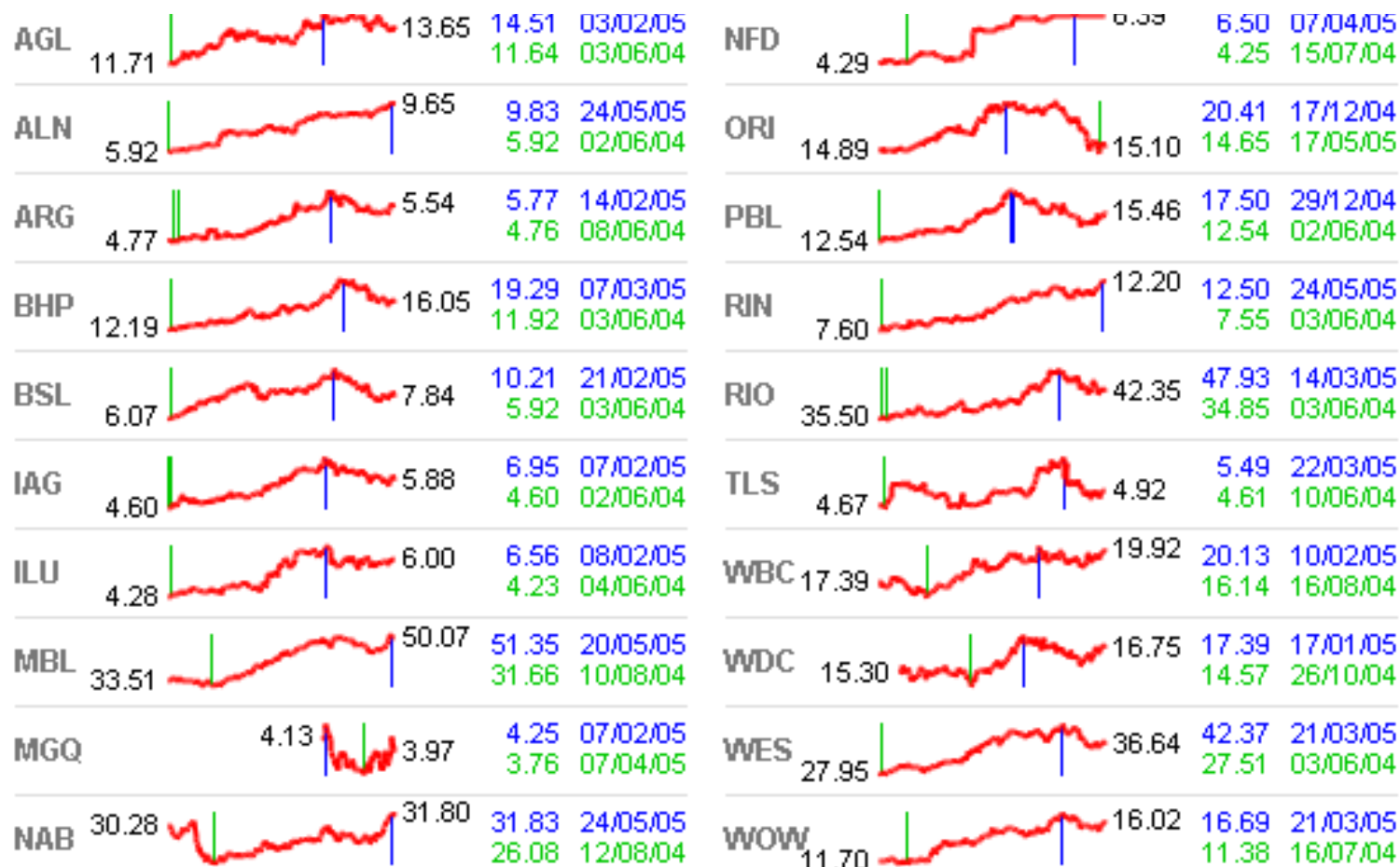
- Multiple small graphs arranged in a grid
- Advantages and disadvantages?

# Change in Home Prices (year over year)

From New York Times Economix blog



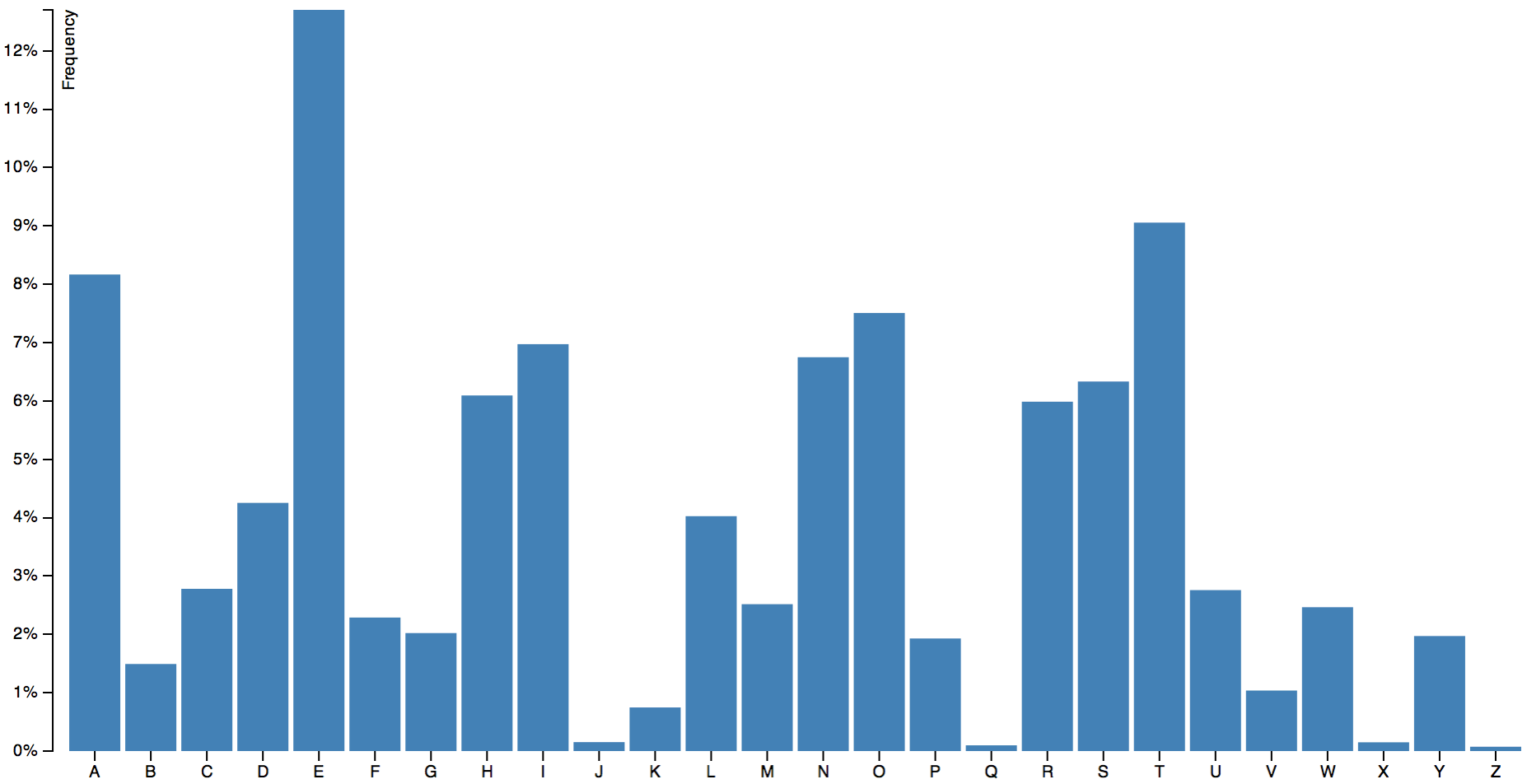
# Spark Lines

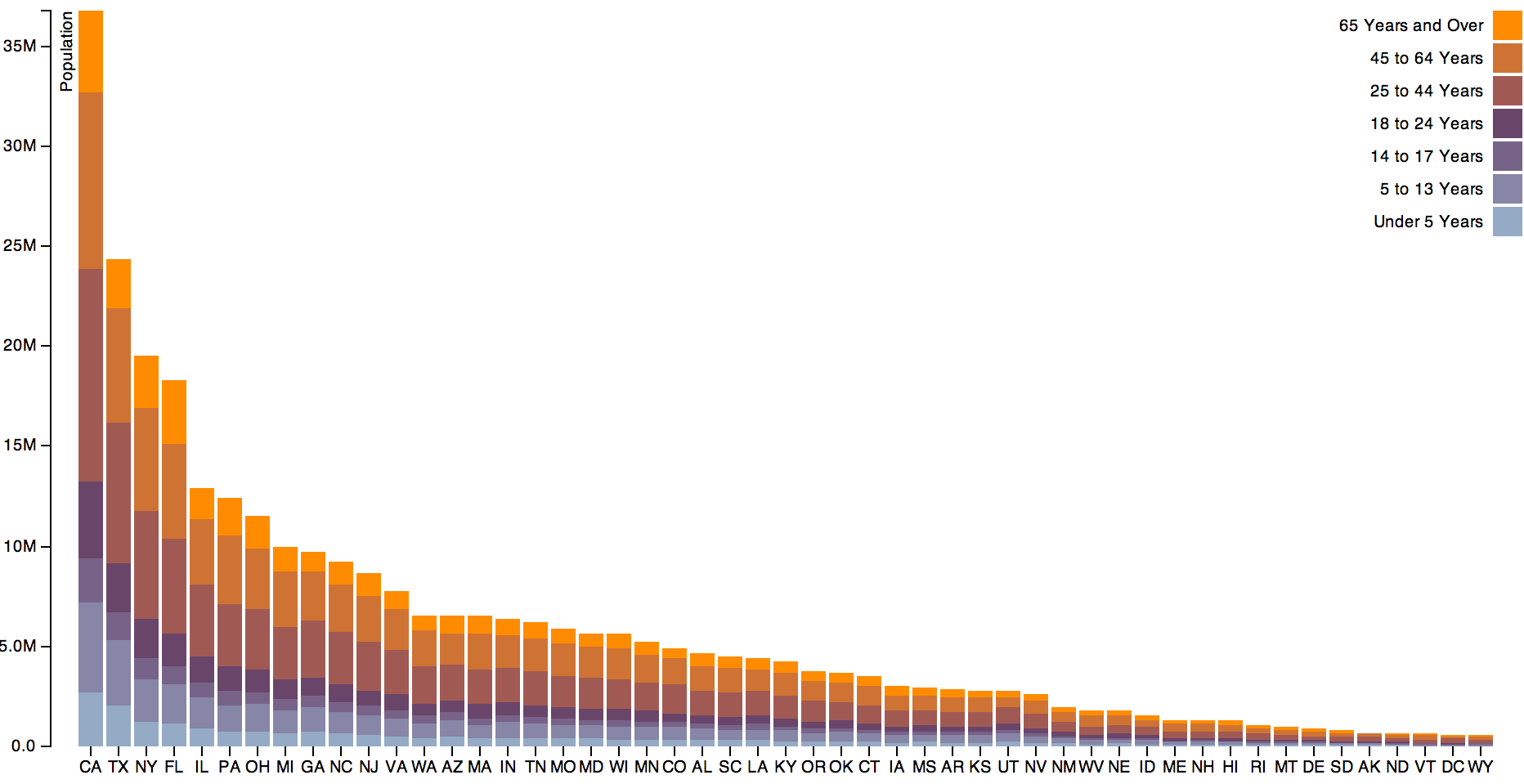




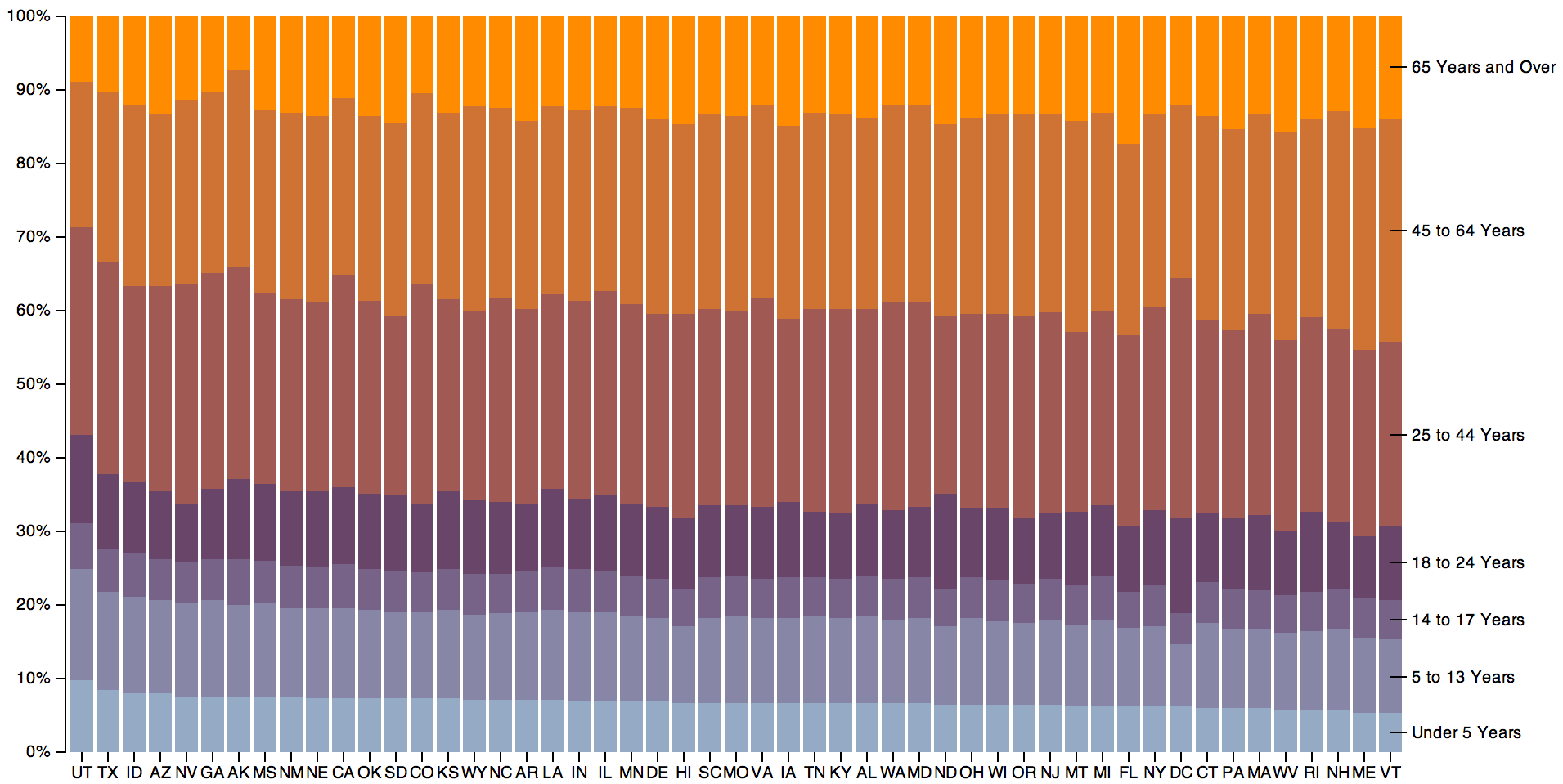
# Bar Graphs (+ Variants)

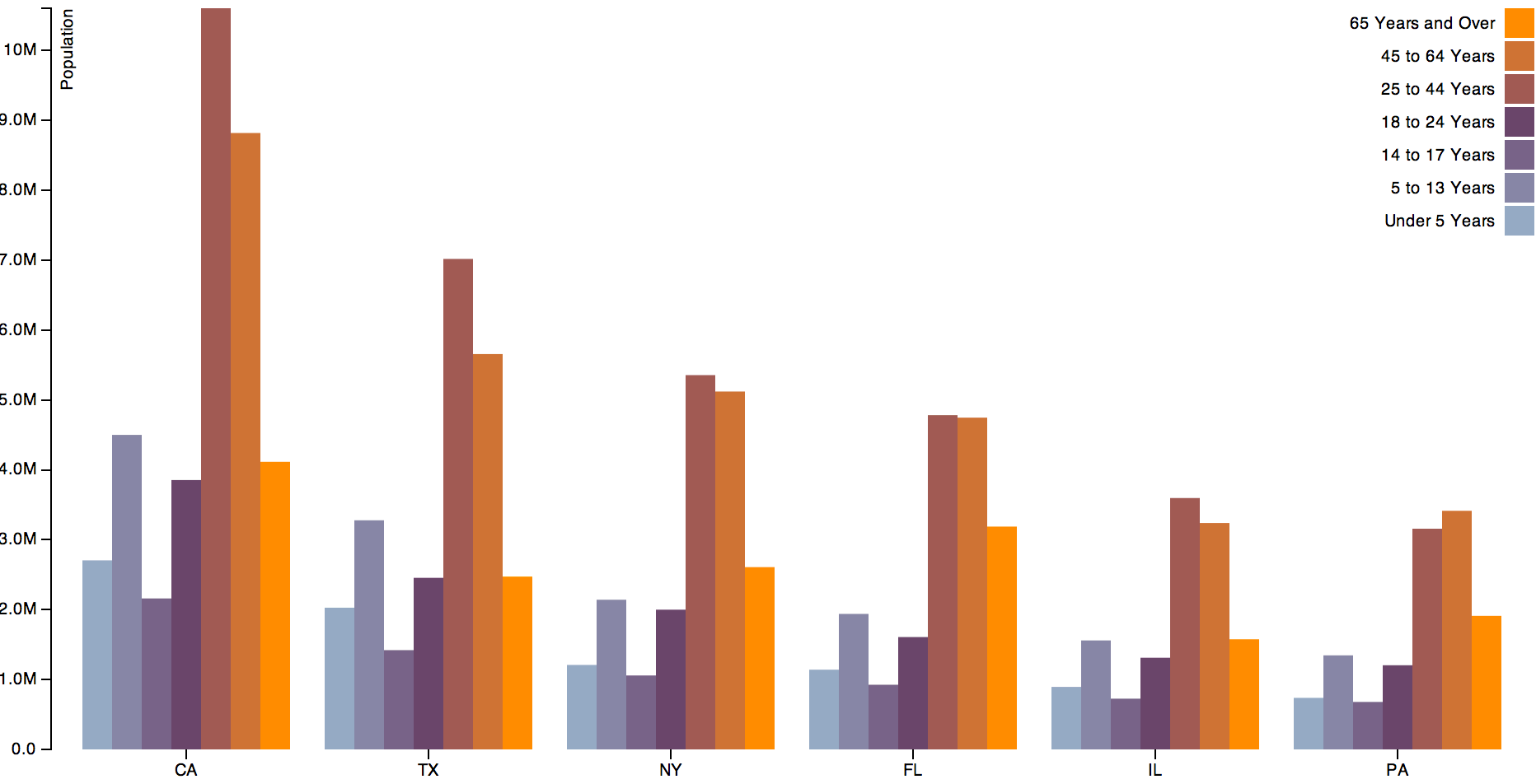
- Suitable for what types of data?
- Following examples from d3.js





CA TX NY FL IL PA OH MI GA NC NJ VA WA AZ MA IN TN MO MD WI MN CO AL SC LA KY OR OK CT IA MS AR KS UT NV NM WV NE ID ME NH HI RI MT DE SD AK ND VT DC WY

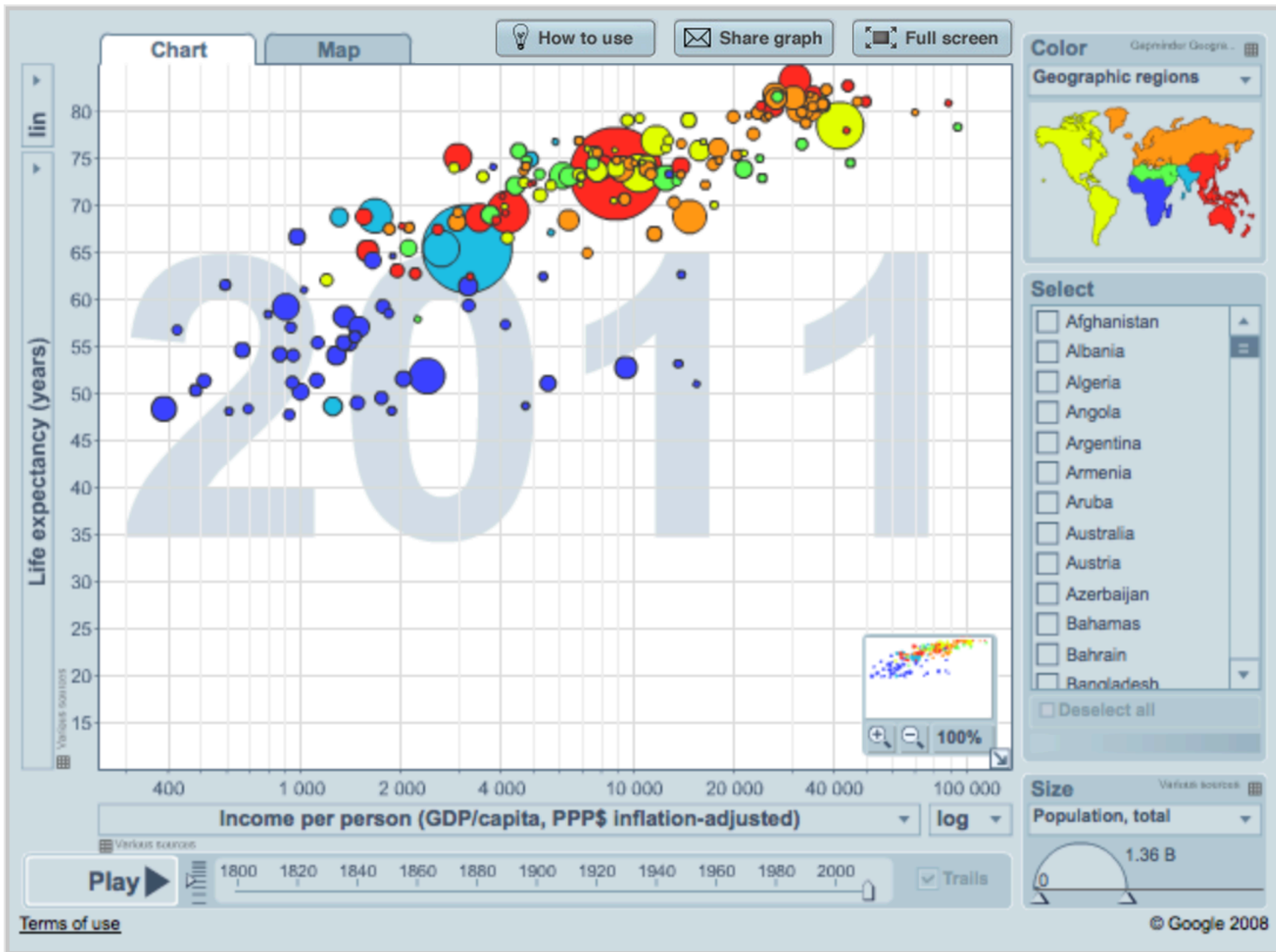






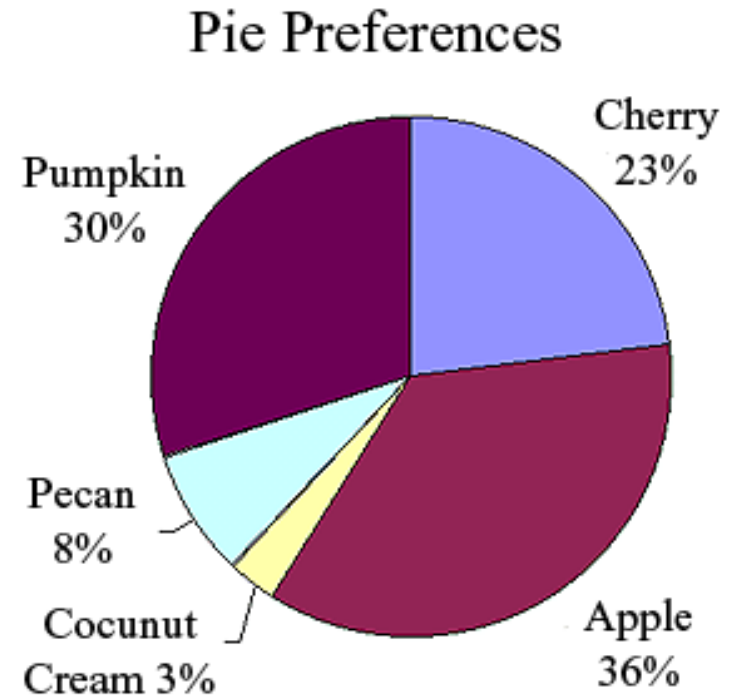
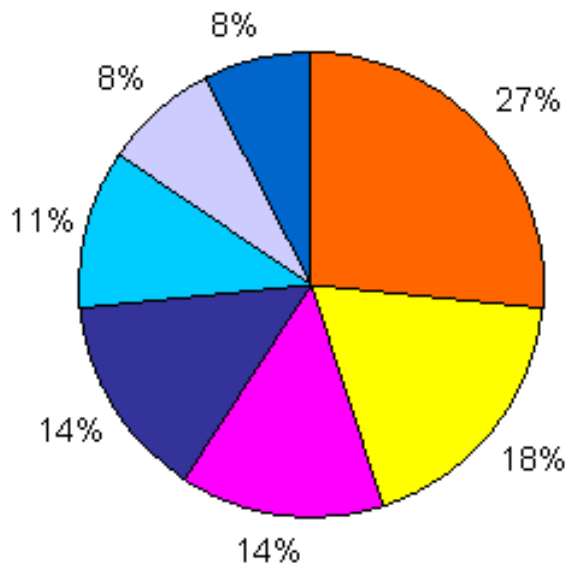
# Scatterplots

- Suitable for what types of data?
- Note availability of other ways to encode data at points

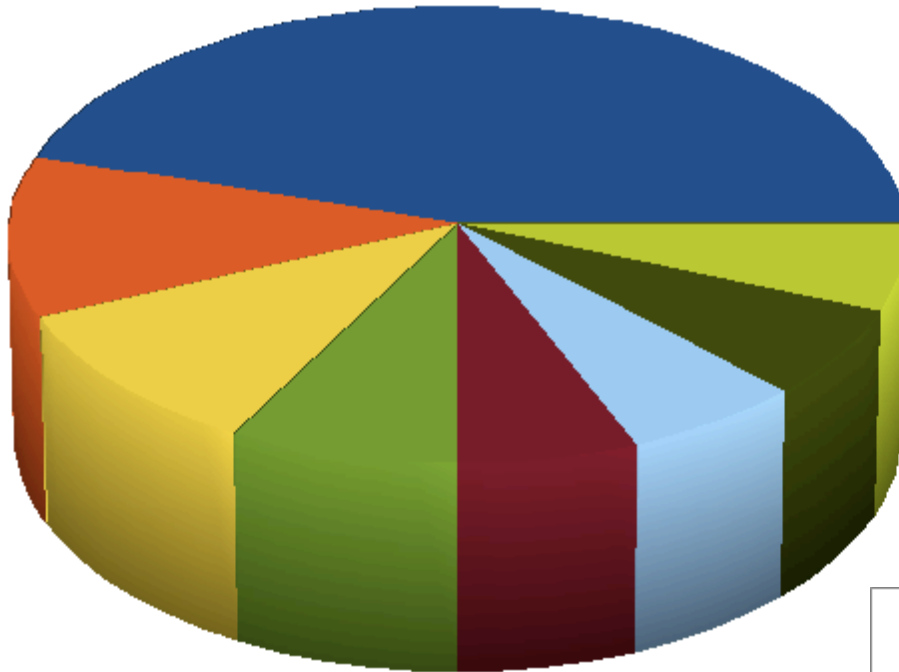


# Pie Charts (+ Variants)

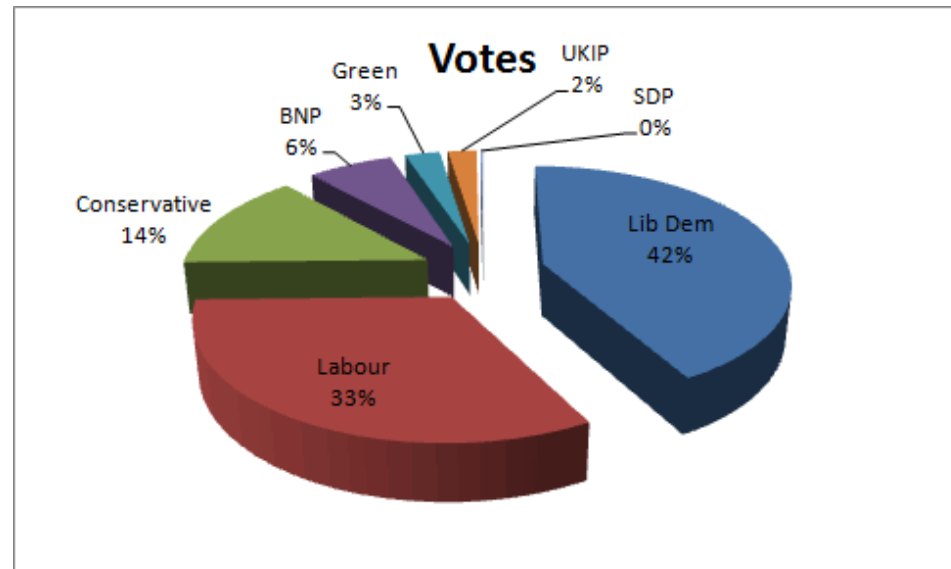
- Suitable for what types of data?
- Different ways for labeling



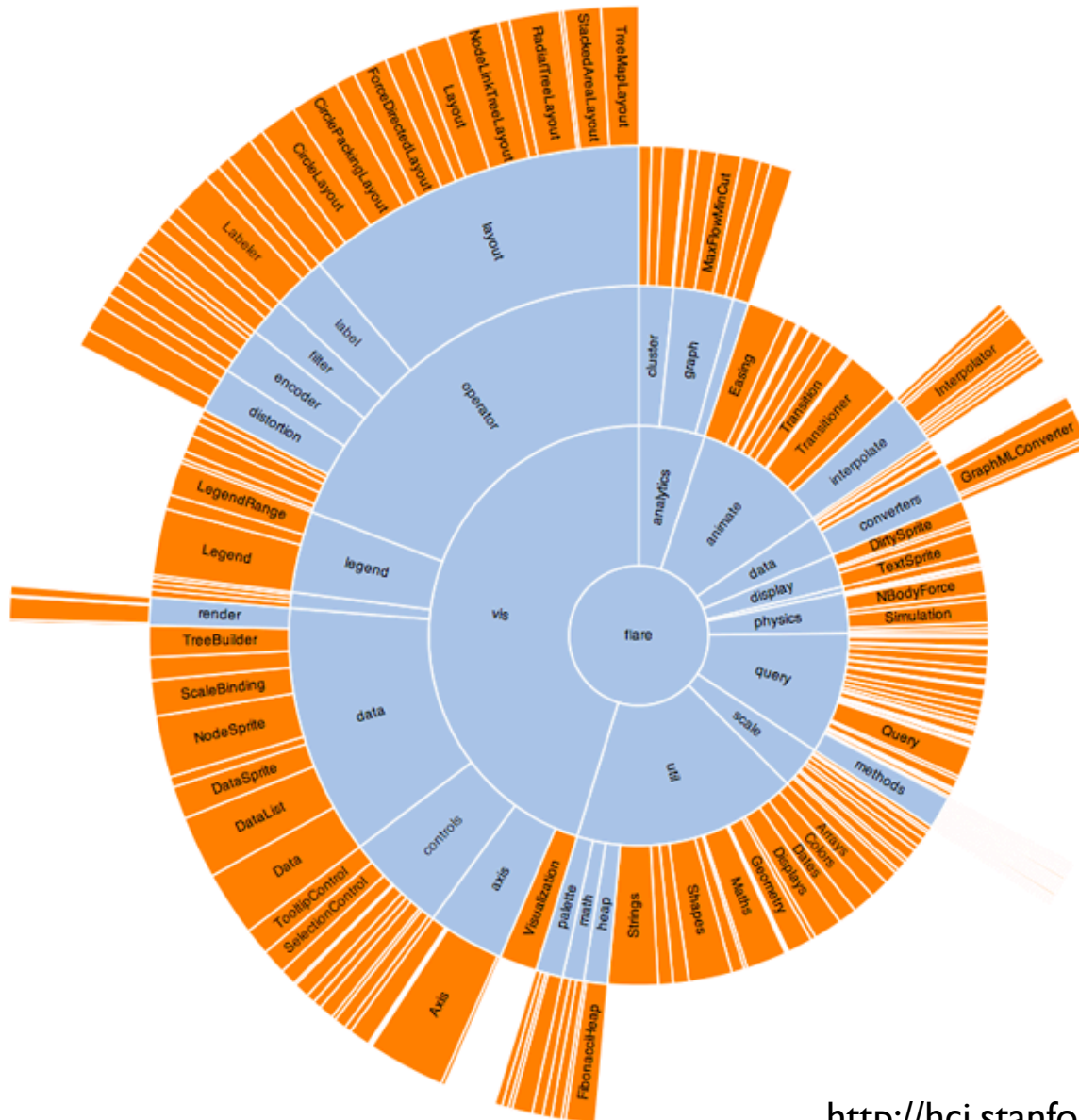
# More Chart Junk



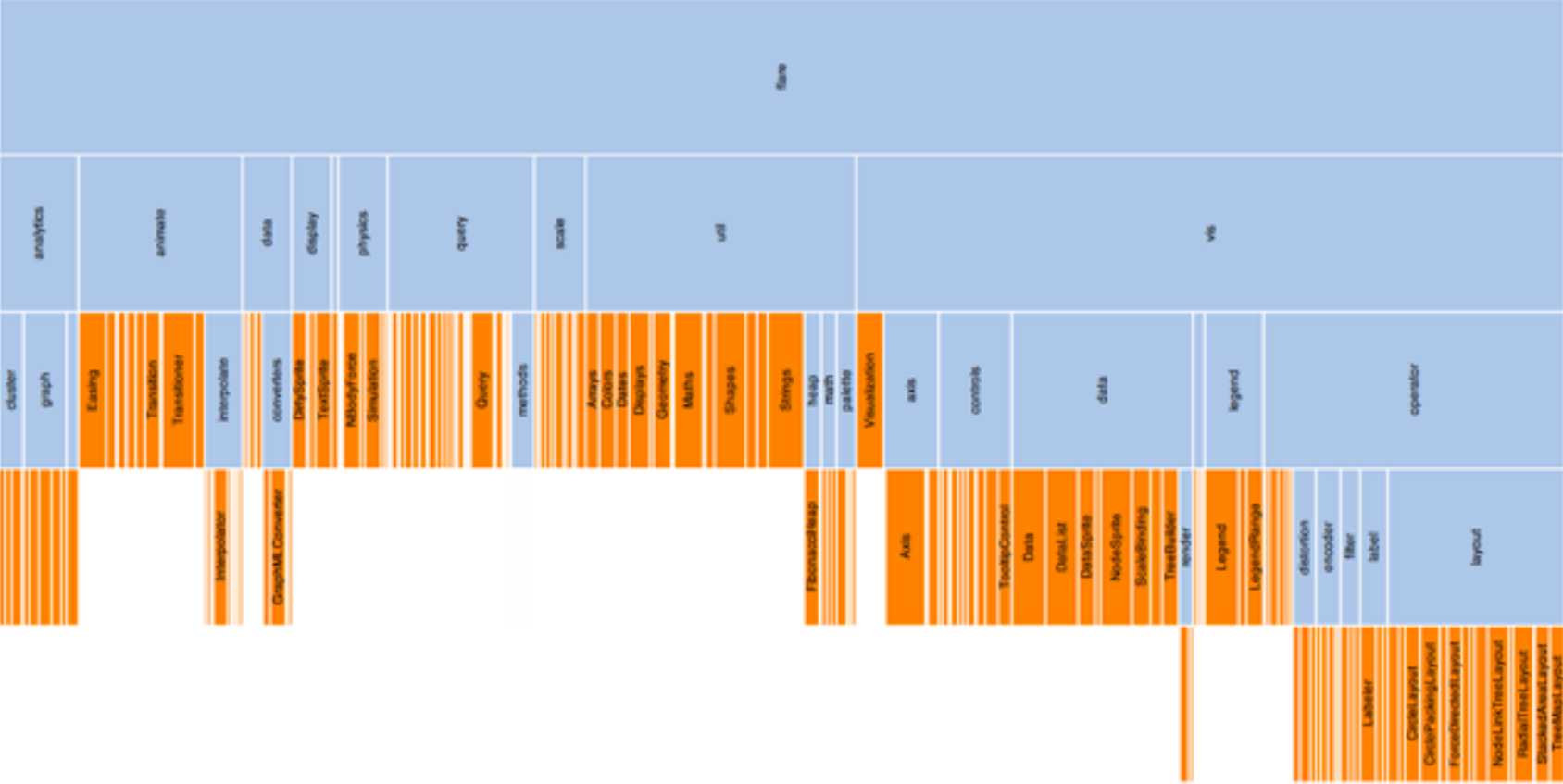
- MySpace
- Facebook
- Orkut
- hi5
- Xanga
- Live Spaces
- Classmates.com
- Bebo



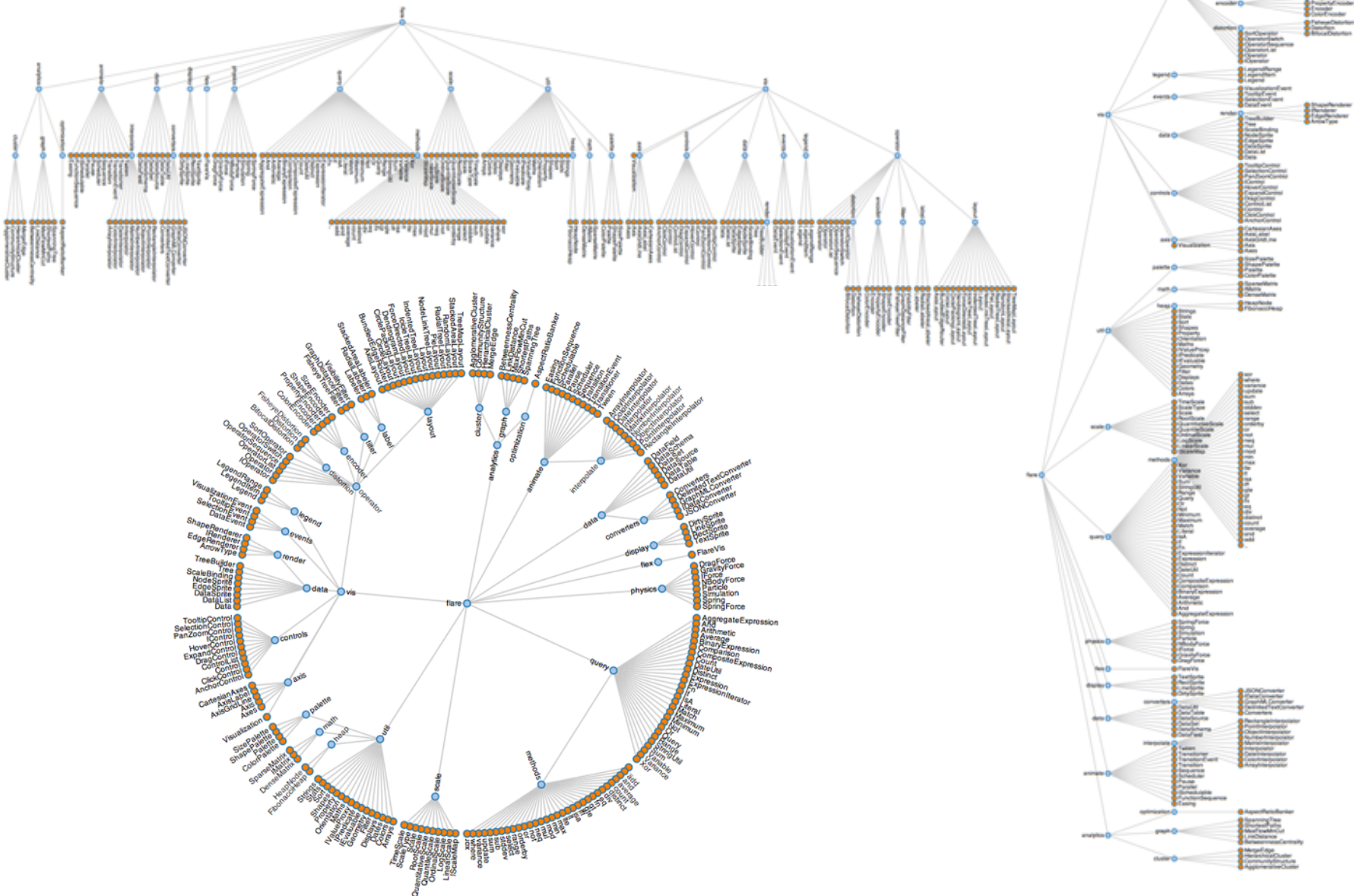
# Sunburst Trees



# Icicle Diagrams



# Visualizing Hierarchies

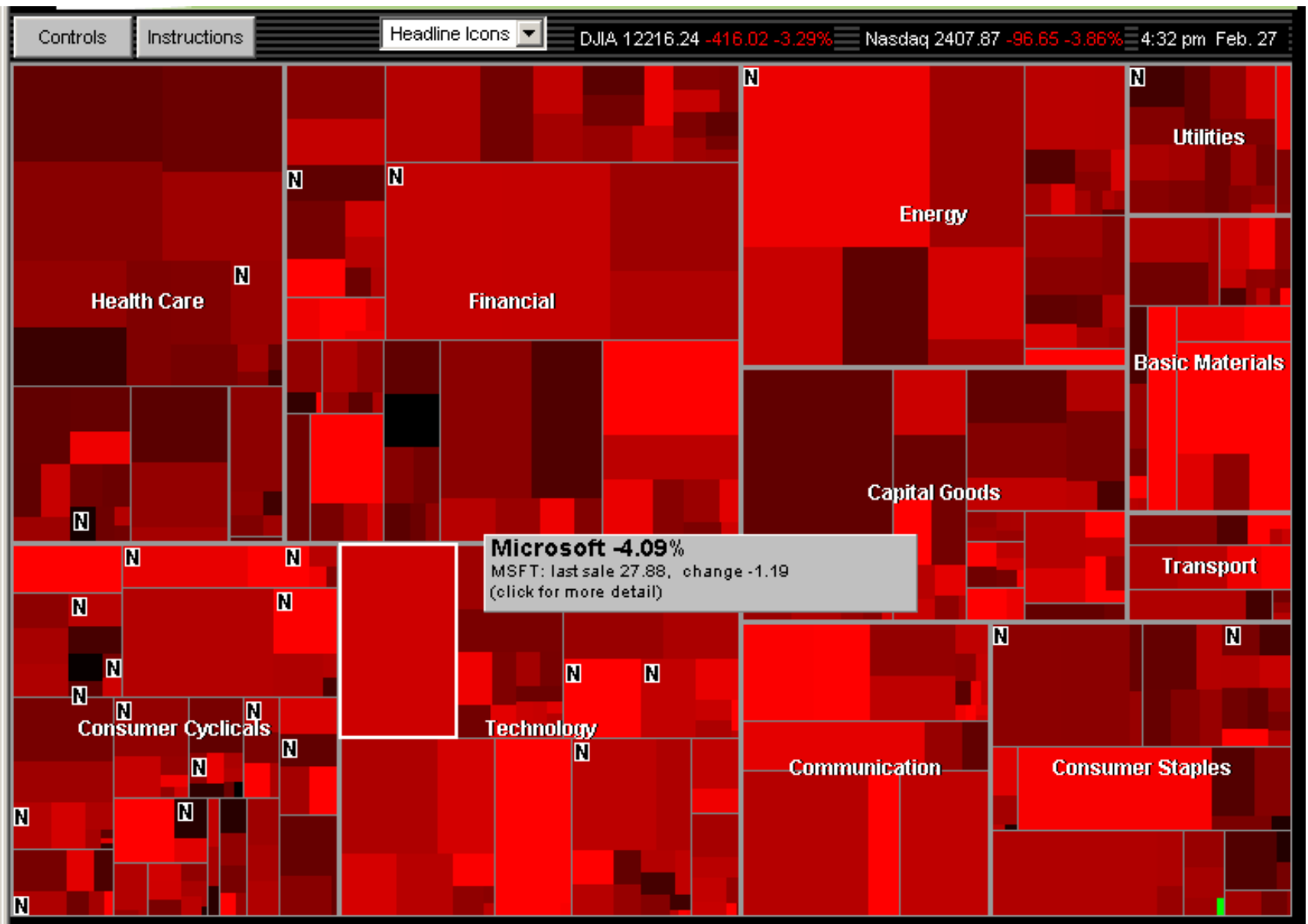


# TreeMaps

- Useful for representing hierarchical relationships
- Two mechanisms for coding information:
  - Size and Color
  - Physical location of boxes not usually meaningful
- Requires learning to read

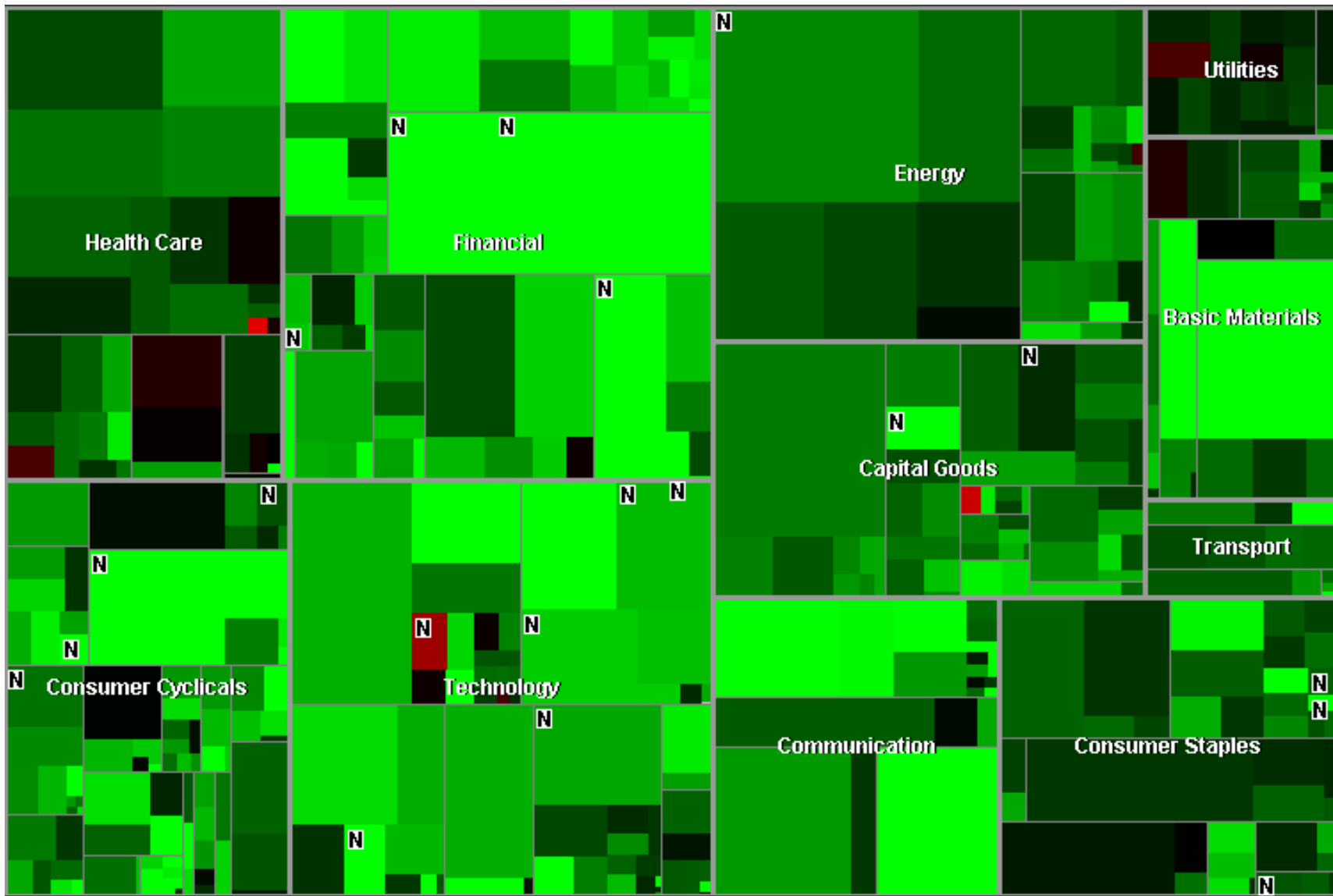






Market mixed, February 8, 2008:

Energy & Technology up Financial & Health Care down

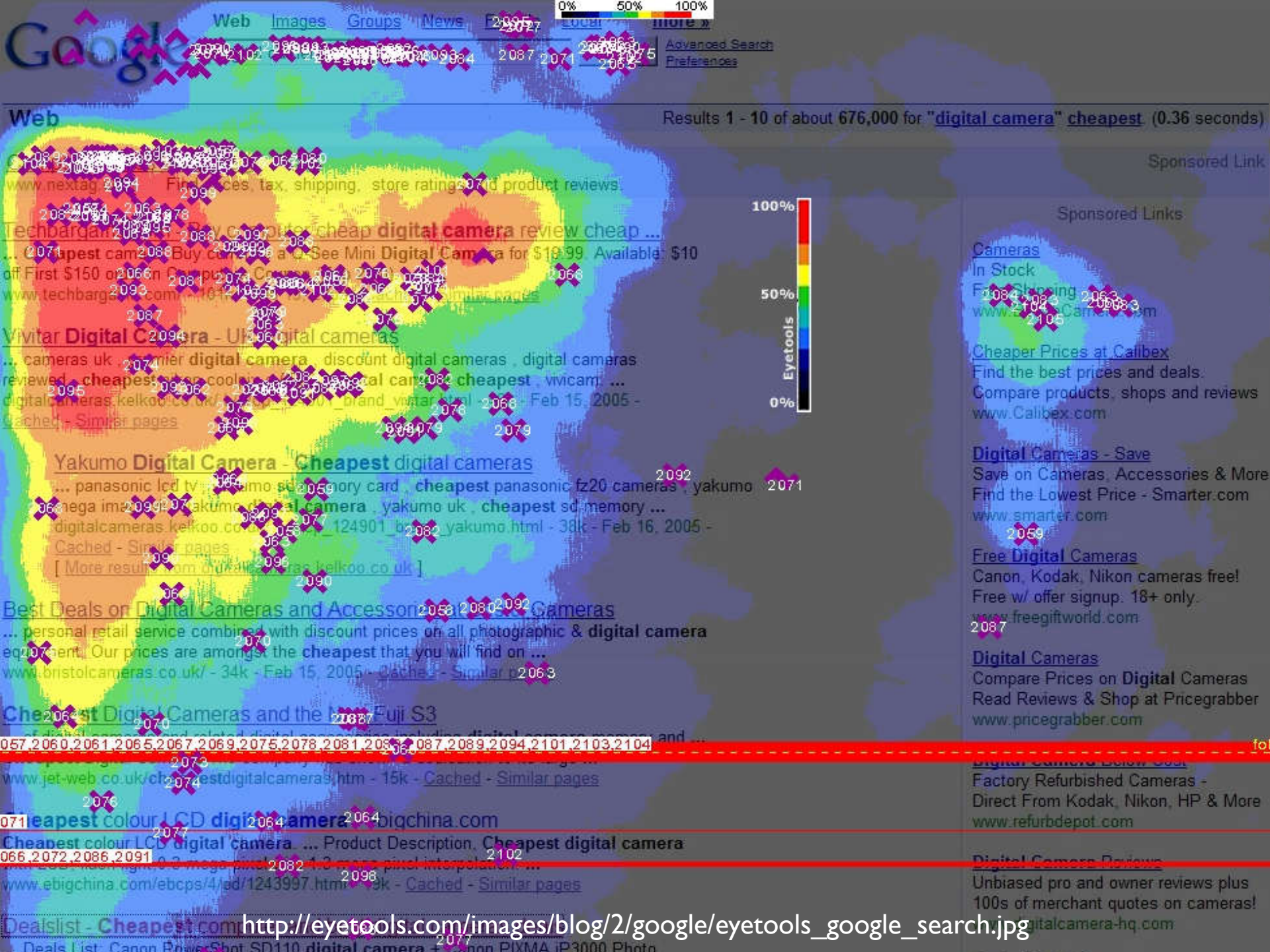


Market rises 319 points, November 13, 2007:

Five exceptions

# Heatmaps

- Data represented in a matrix
- Matrix values are color coded

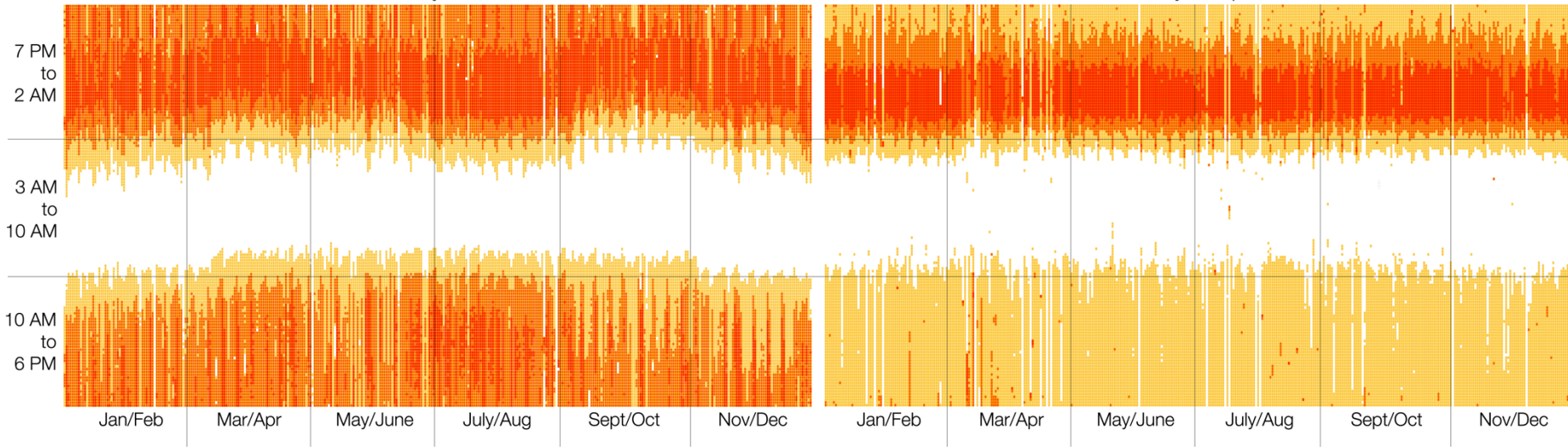




Low Tweet Volume  High Tweet Volume

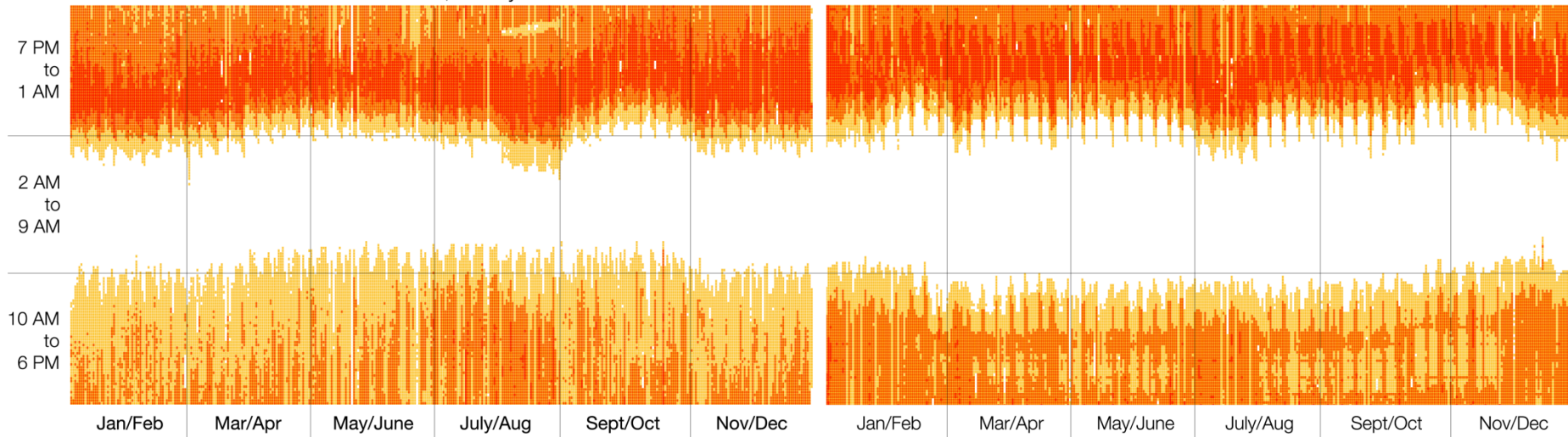
New York City, United States

Tokyo, Japan



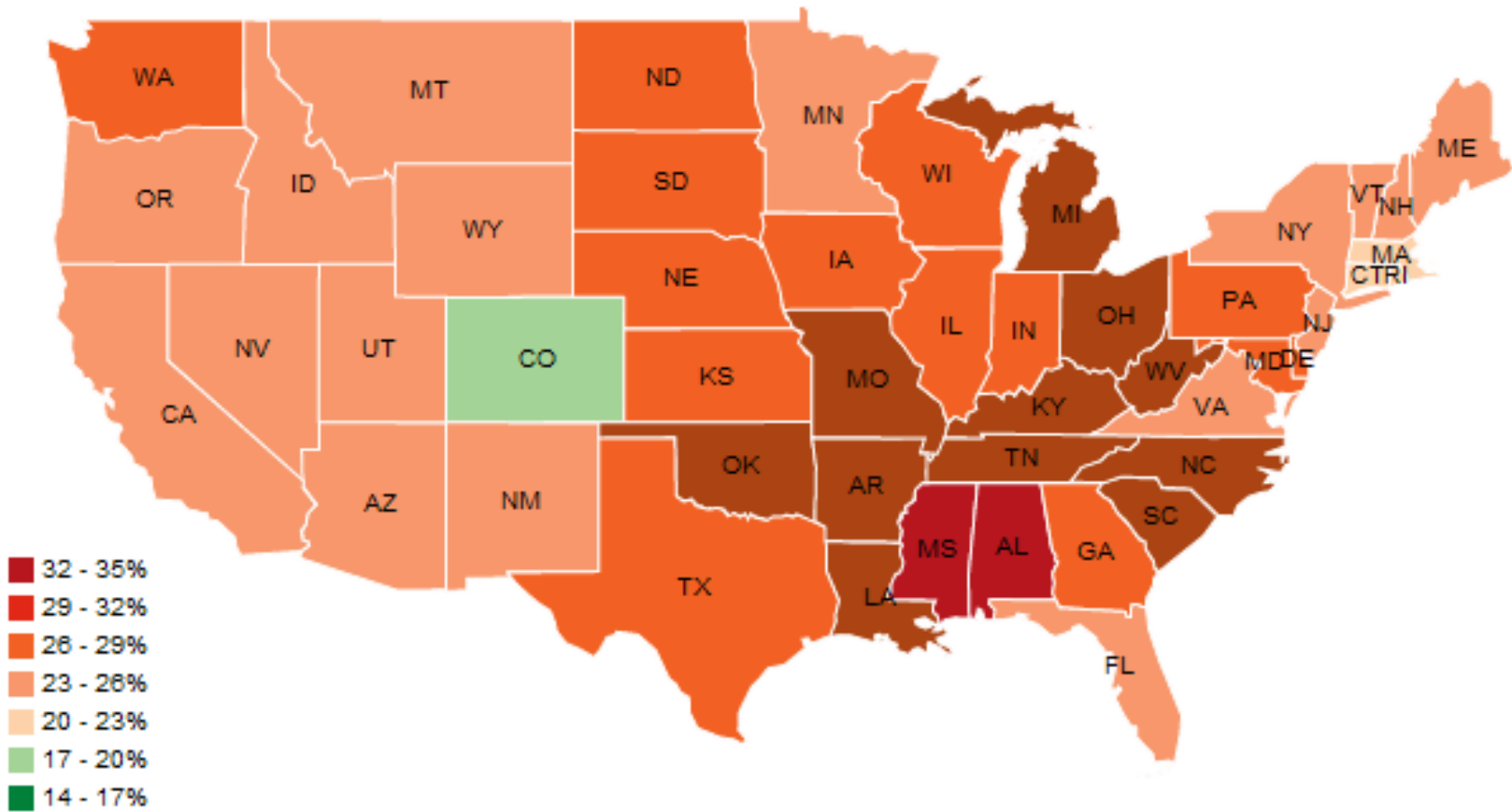
Istanbul, Turkey

São Paulo, Brazil

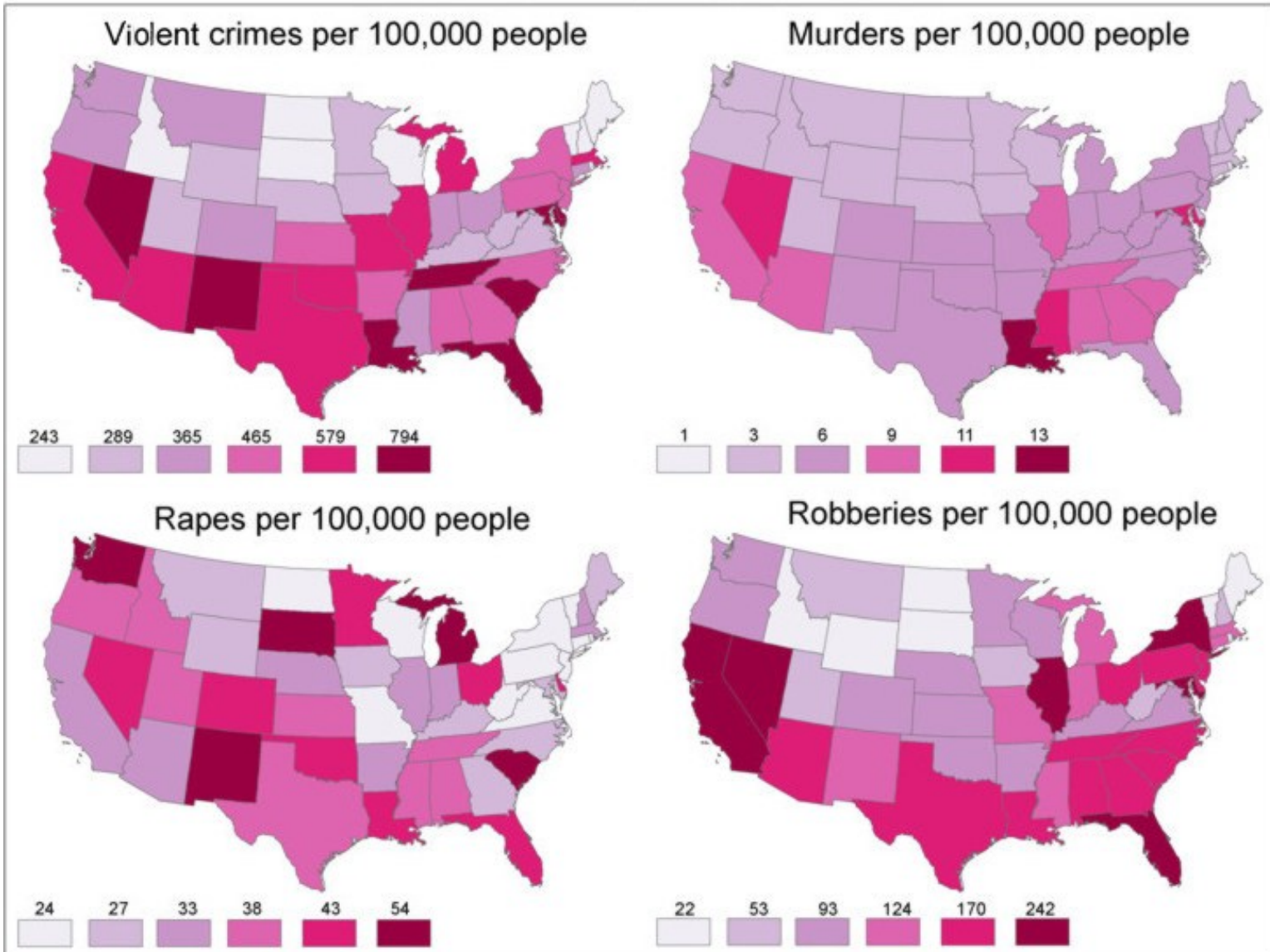


Day of year (2011)

# Choropleth Maps



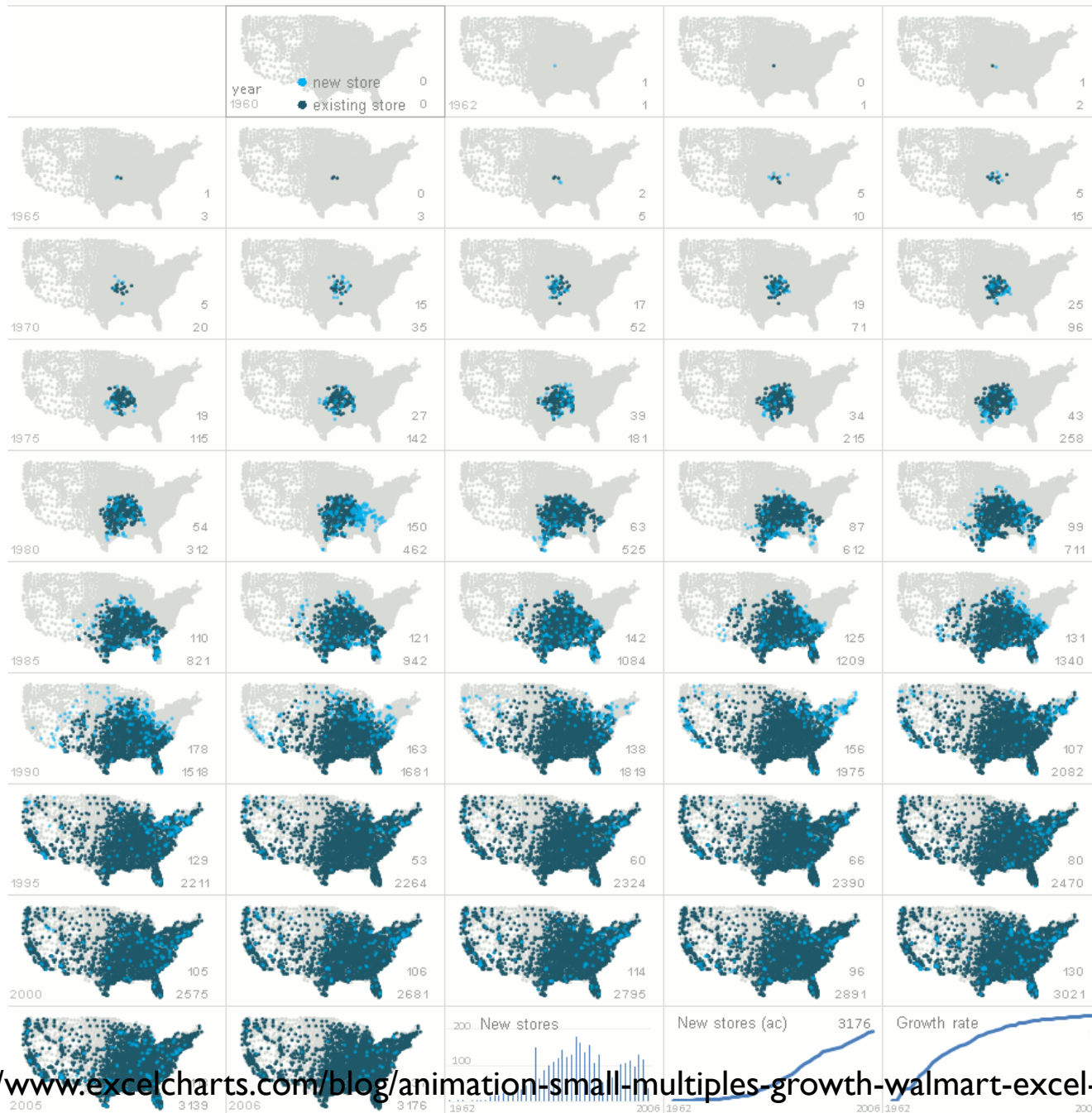
# Crime Rates in the US - 2003



Source: *Crime in the United States, 2003*, FBI, Uniform Crime Reports



# Growth of Walmart

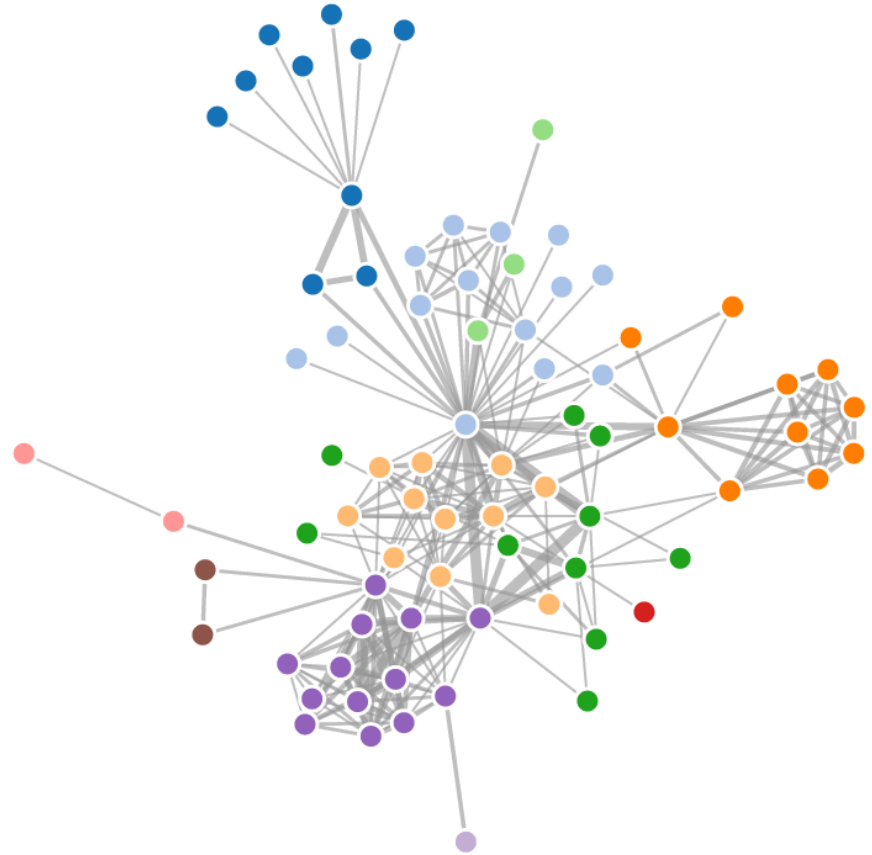


# Small Multiples vs. Animations

- Advantages and disadvantages?

# Visualizing Networks

- Types of networks: social networks, interaction networks, biological networks, computer networks, etc.
- Standard algorithm uses force-directed layout







Panda Viewing  
Sun - 5:30pm

