

INFM 603: Information Technology and Organizational Context

Session 9: Visualization



**Jimmy Lin
The iSchool
University of Maryland**

Thursday, October 31, 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
 - Numerical data
- Ways of representing data
 - Height
 - Width
 - Thickness
 - Area
 - Angles
 - Size
 - Shape
 - Color



Panda Viewing
9am - 5:30pm

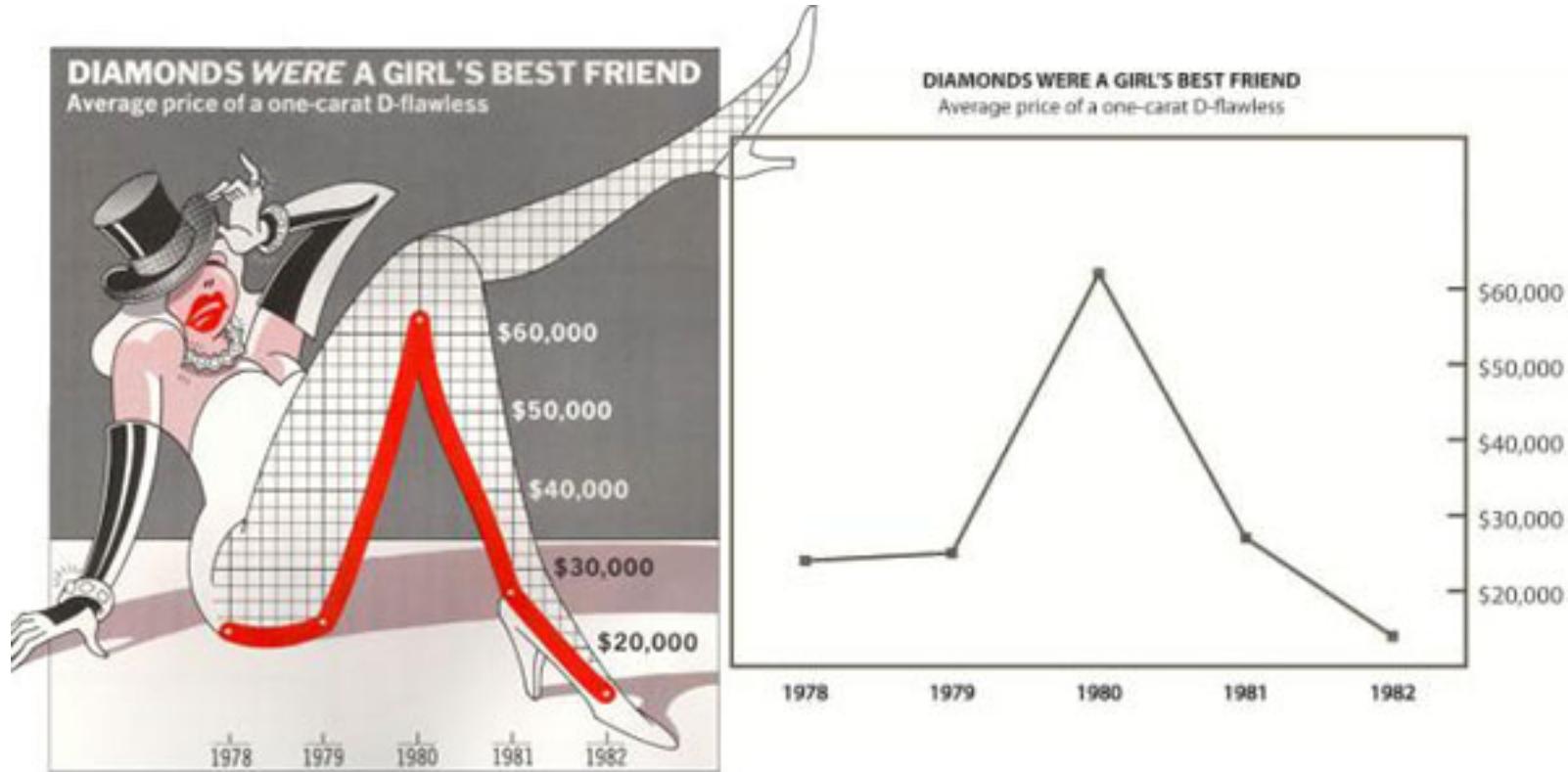


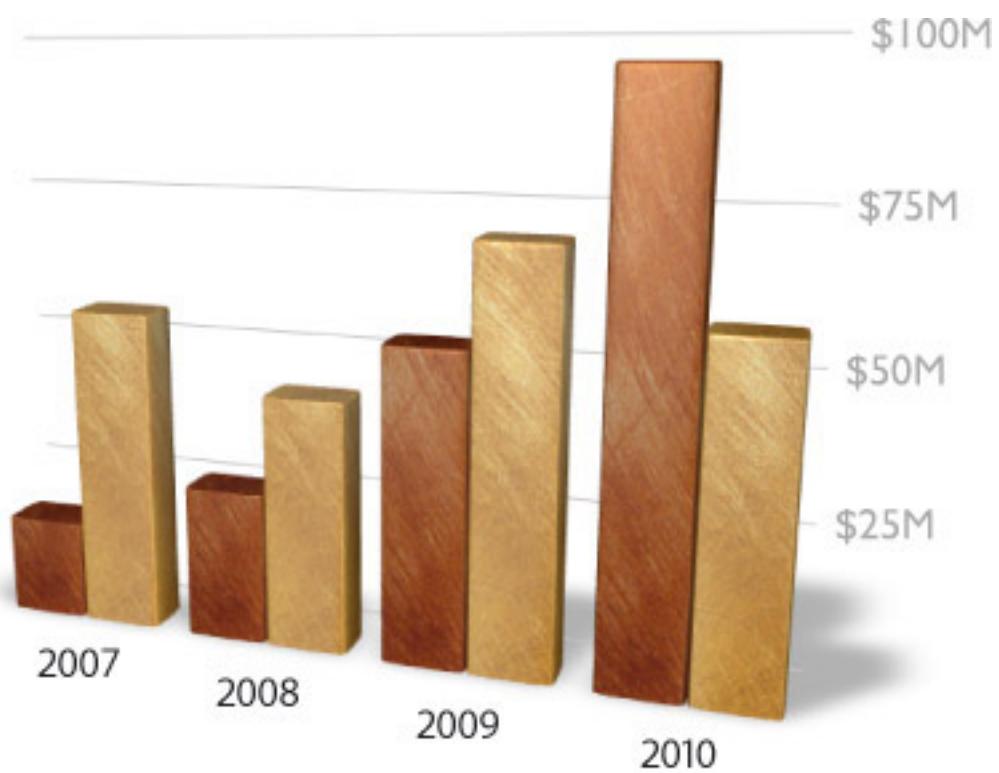
Static vs. Interactive

- 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

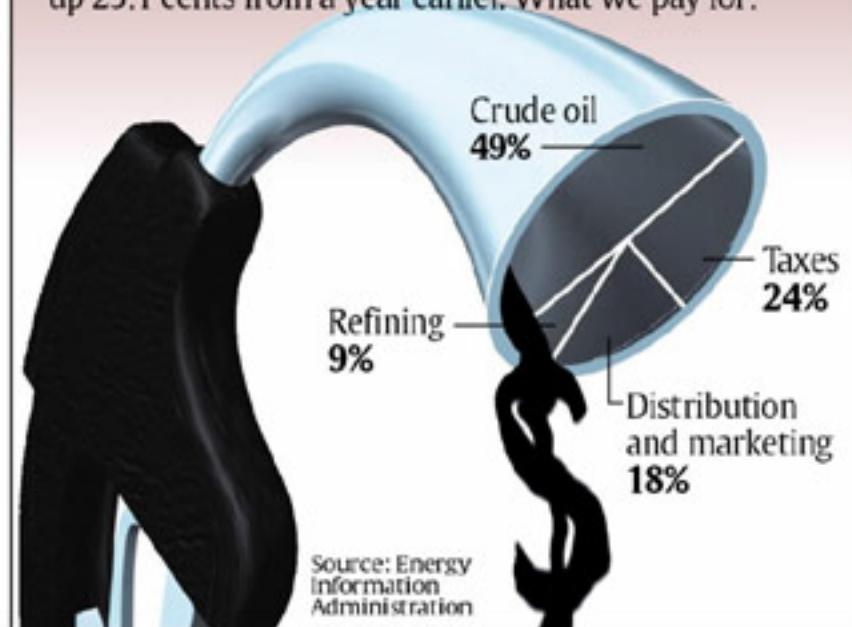




USA TODAY Snapshots®

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%
IMPROVEMENT FACTORY

THE LAST RECORD WAS SET IN 1998 BY FLORENCE GRIFITH-JONES. USAIN BOLT PULVERIZED RECORDS IN 2008. WILL HIS RECORD EVENTUALLY BE BEATEN?

♂ +11%
IMPROVEMENT FACTORY

RACE TIME
THE GRAPH SHOWS THE EVOLUTION OF THE 10 BEST YEARLY PERFORMANCES

MARATHON
+20%
IMPROVEMENT FACTORY

WILL THE 2-HOUR LIMIT BE BROKEN SOMEDAY?

2010
HALF
GERMANY
KEN 2011

HIGH JUMP
+30%
IMPROVEMENT FACTORY

PERFORMANCE HAS LEVELED SINCE THE 1990s.

2.45 m
JAVIER SOTOMAYOR
CUBA 1993

2.40 m
GARRETTE COOPER
USA 1996

2.35 m
FLORENCE GRIFFITH-JONES
GBR 1938

2.30 m
DONALD LIPFINGOTT
USA 1912

2.25 m
JIM HINES
USA 1968

2.20 m
BERNIE JOHNSON
USA 1983

2.15 m
MAURICE BENJAMIN
GBR 1956

2.10 m
USAIN BOLT
JAM 2009

POLE VAULT
+71%
IMPROVEMENT FACTORY

PERFORMANCES HAVE DECLINED SINCE 2000.

6.14 m
SERGEY BUBKA
UKR 1994

6.10 m
MARK MCGRATH
IRE 2000

6.06 m
SERGEY BUBKA
UKR 1998

6.02 m
SERGEY BUBKA
UKR 2000

5.98 m
SERGEY BUBKA
UKR 2002

5.94 m
SERGEY BUBKA
UKR 2004

5.90 m
SERGEY BUBKA
UKR 2006

5.86 m
SERGEY BUBKA
UKR 2008

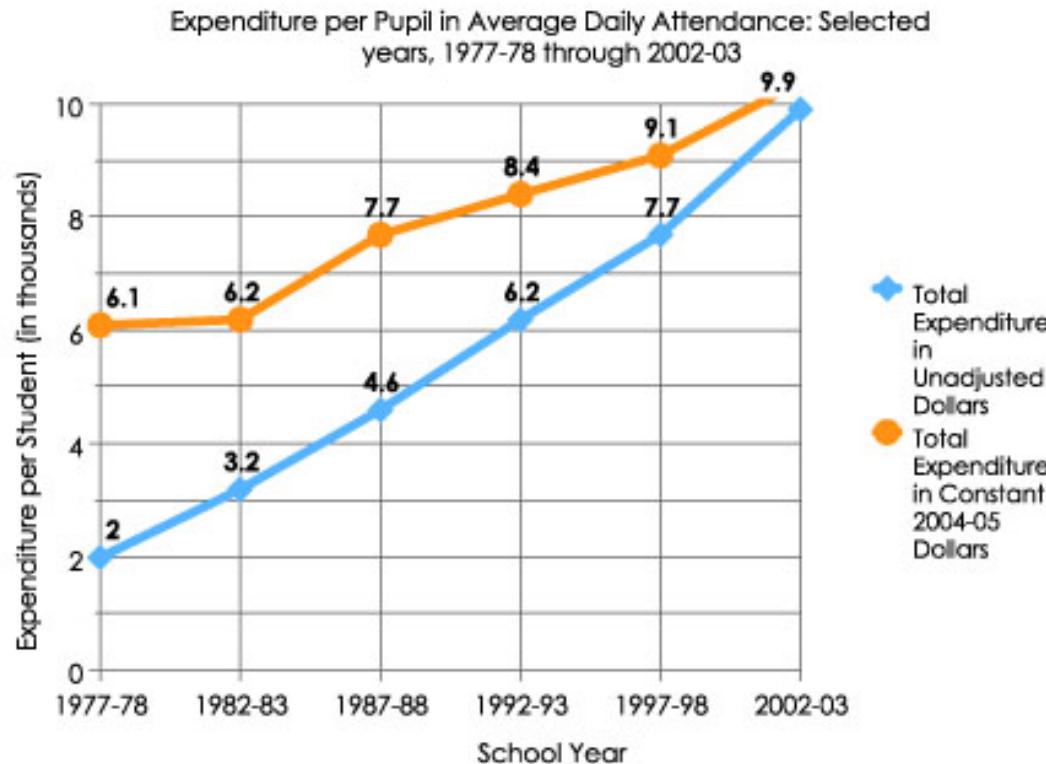
5.82 m
SERGEY BUBKA
UKR 2010

http://junkcharts.typepad.com/junk_charts/music/

Compare: [New York Times interactive visualization](#)

Line Graphs (+ Variants)

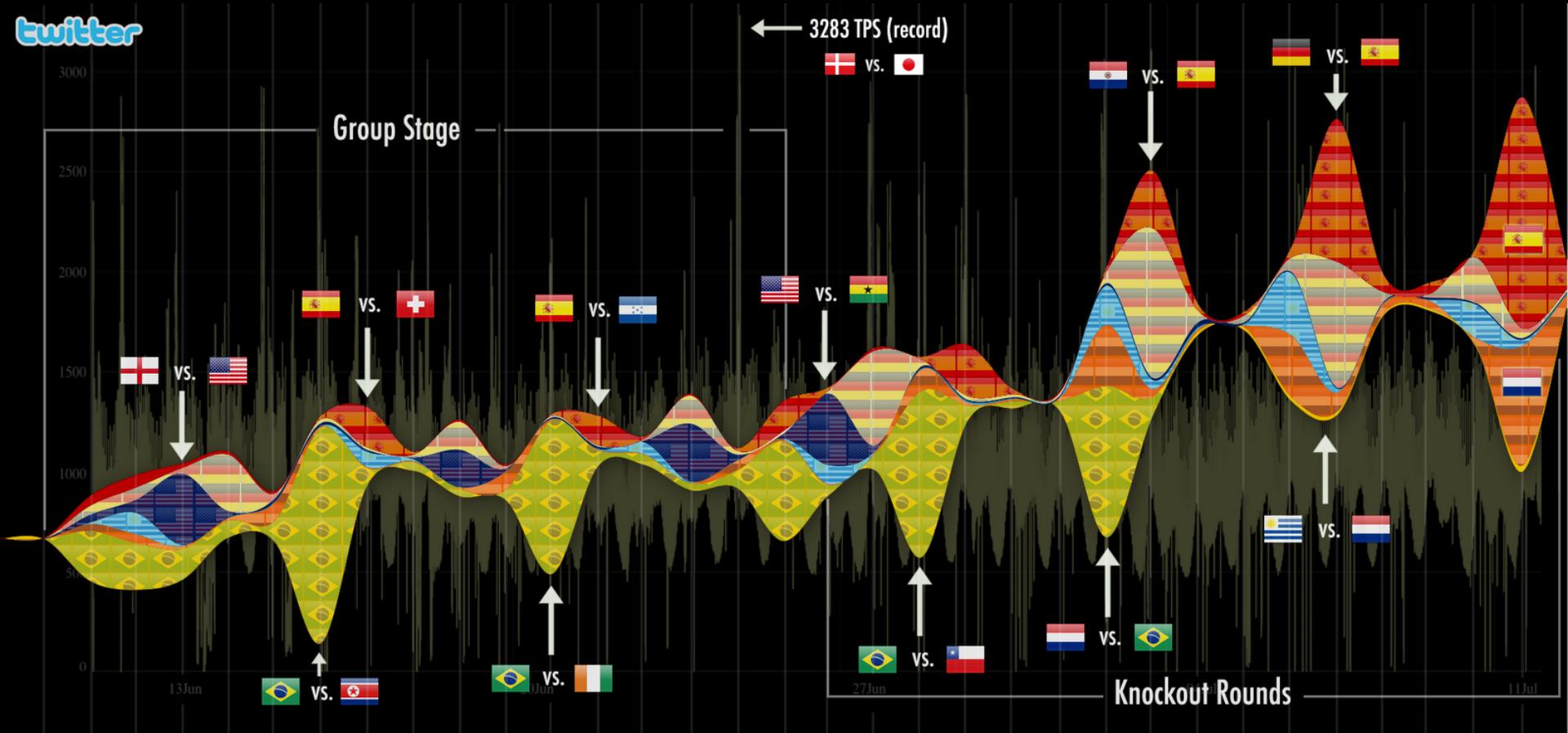
- Suitable for what types of data?
- What makes a good line graph?



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”



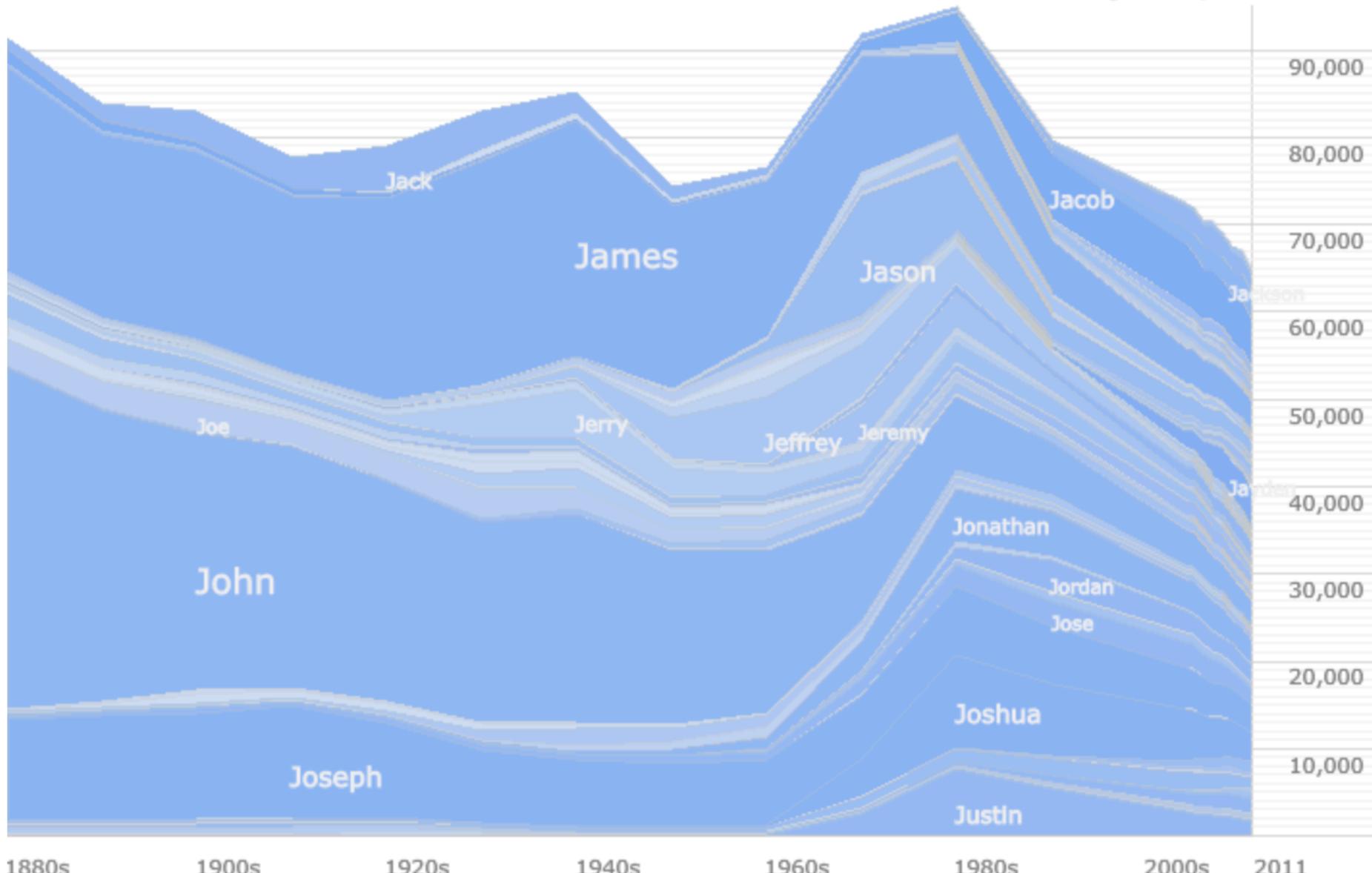
Baby Name > j



Both Boys Girls

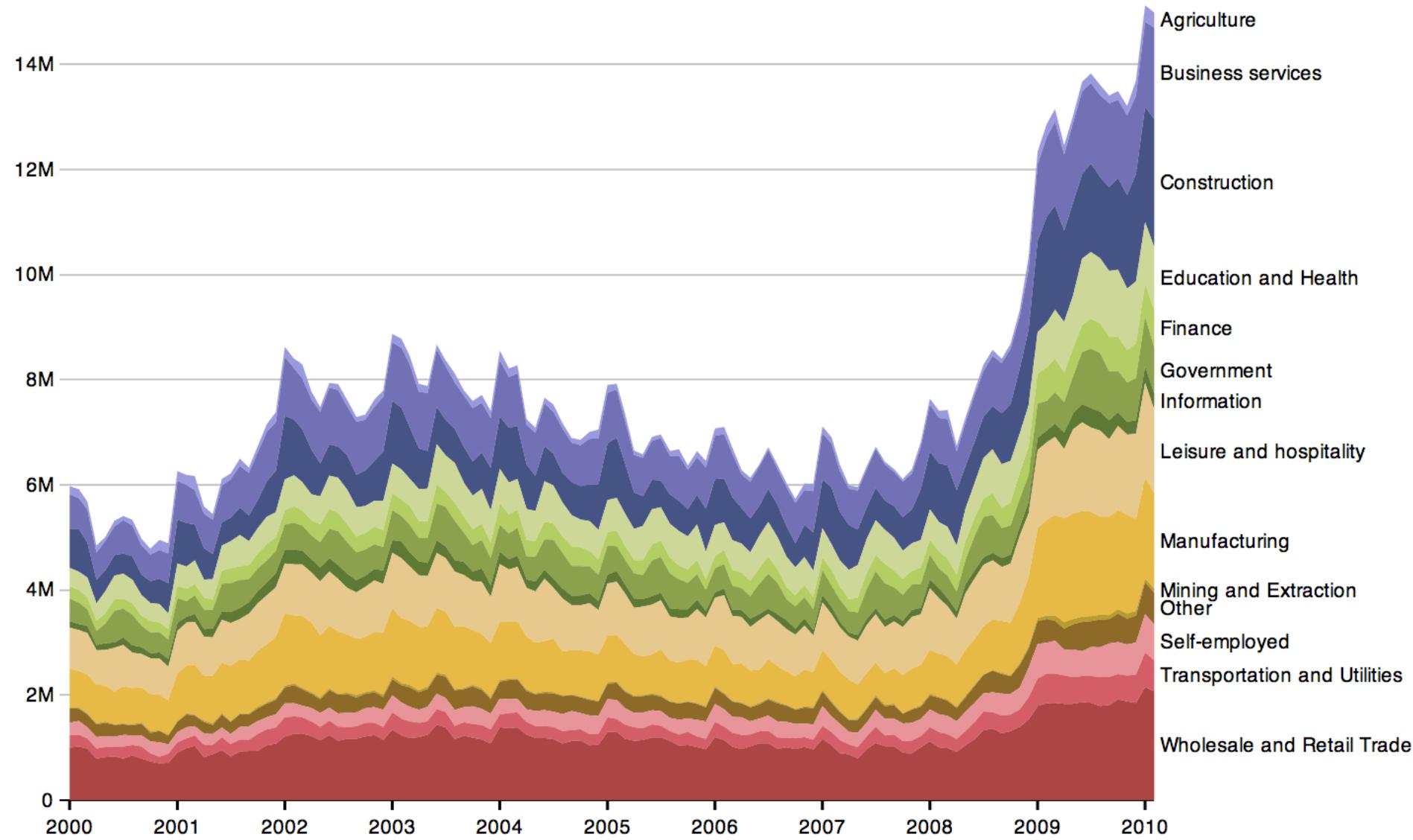
2010 rank: boys 1000 500 100 25 1
girls 1000 500 100 25 1

Names starting with 'J' per million babies



Stacked Graph of Unemployed U.S. Workers by Industry

View: Counts (Zero-Baseline) ▾

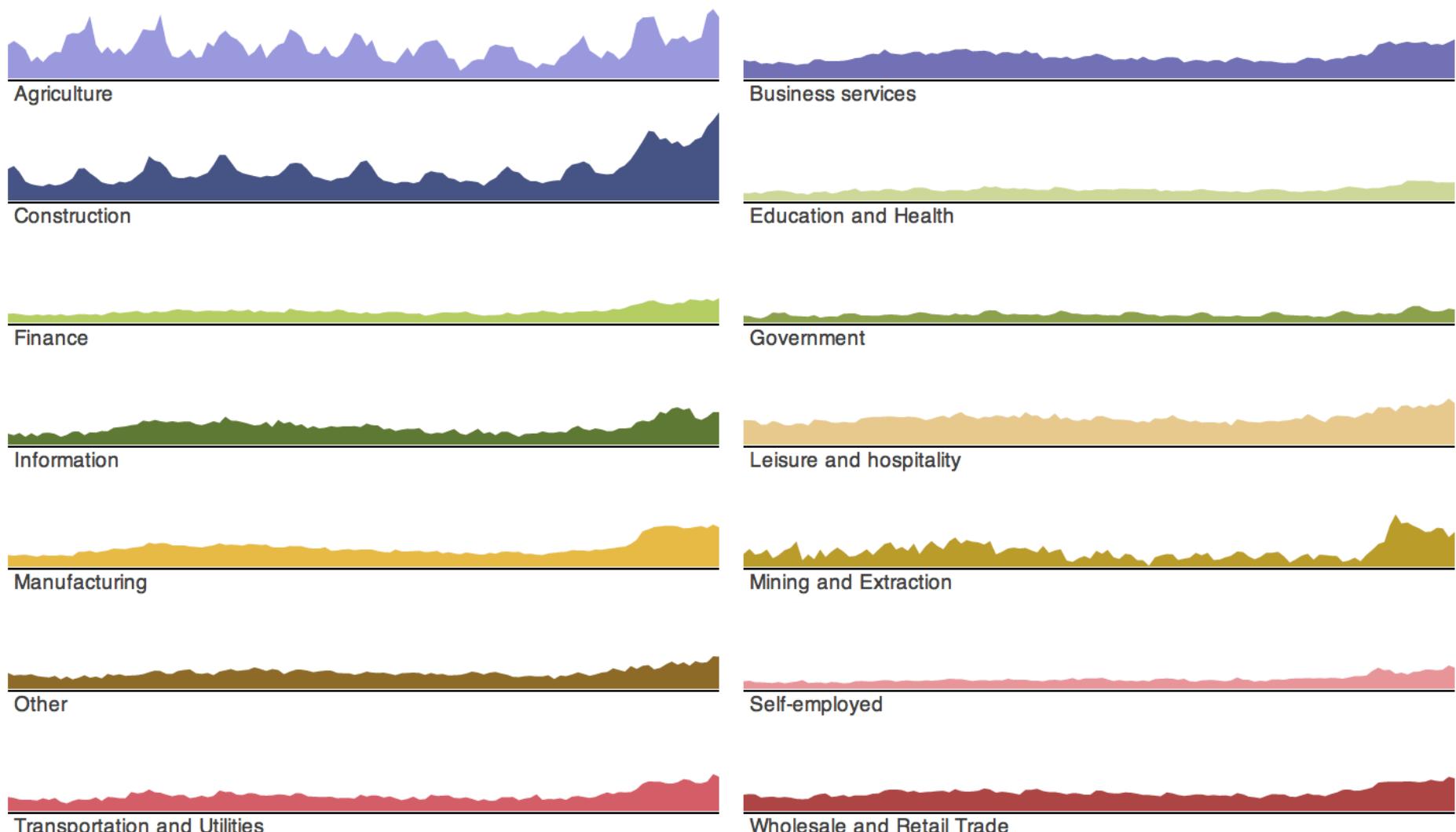


Total counts of unemployed persons per industry, 2000-2010.

Source: [U.S. Bureau of Labor Statistics](#)

<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.

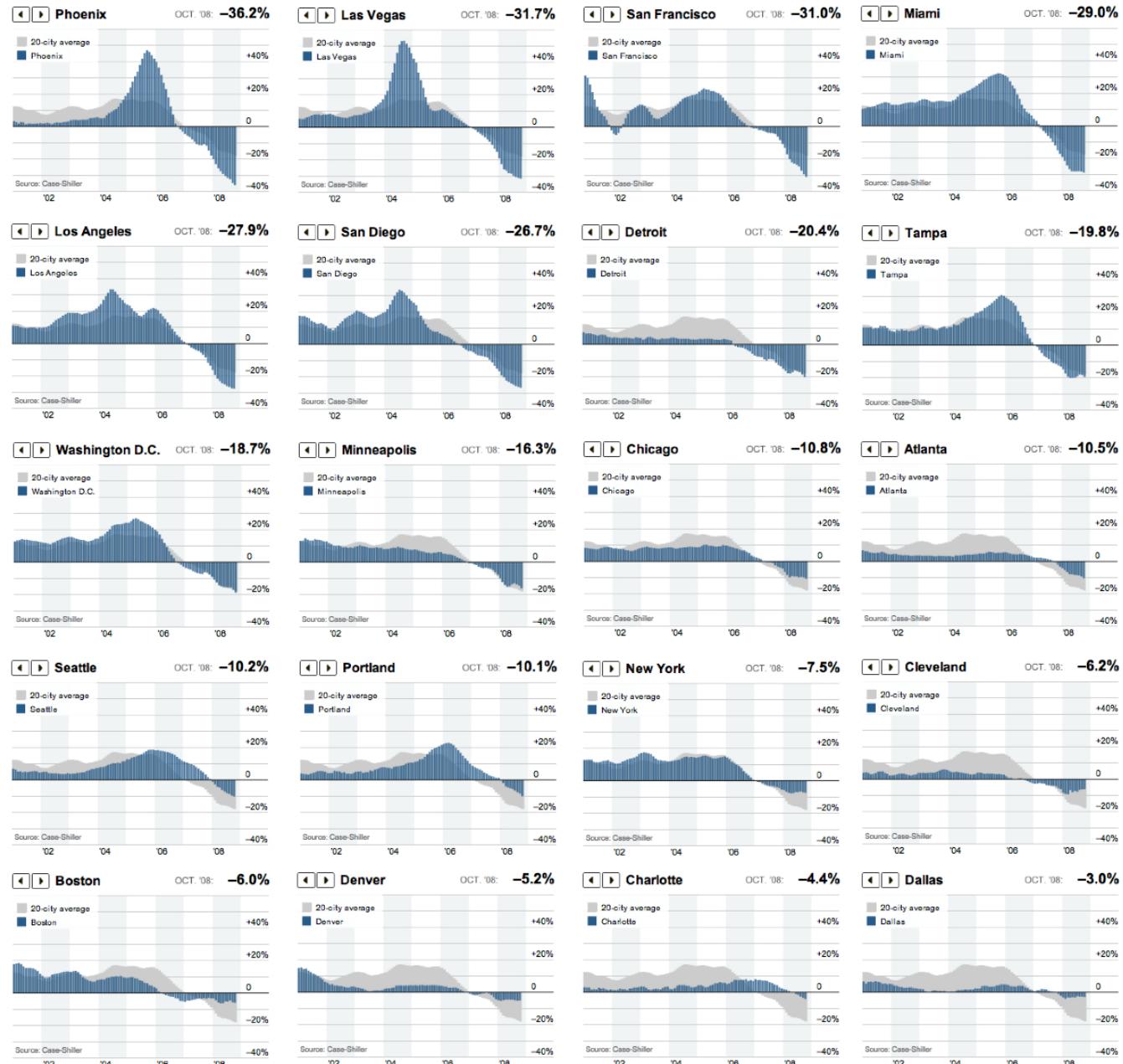
Source: [U.S. Bureau of Labor Statistics](#)

Small Multiples

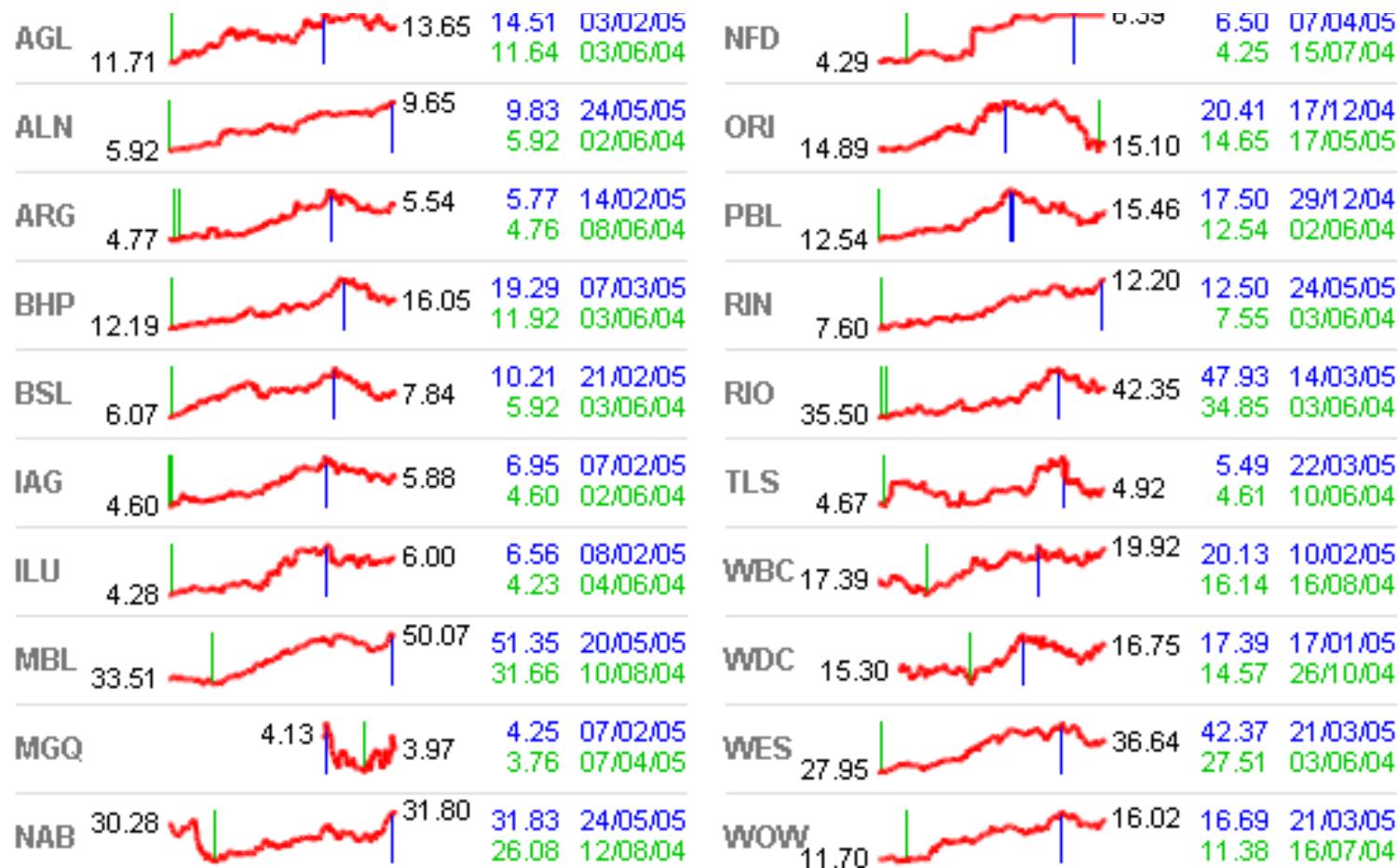
- Multiple small graphs arranged in a grid
- Advantages and disadvantages?

Change in Home Prices (year over year)

From New York Times Economix blog

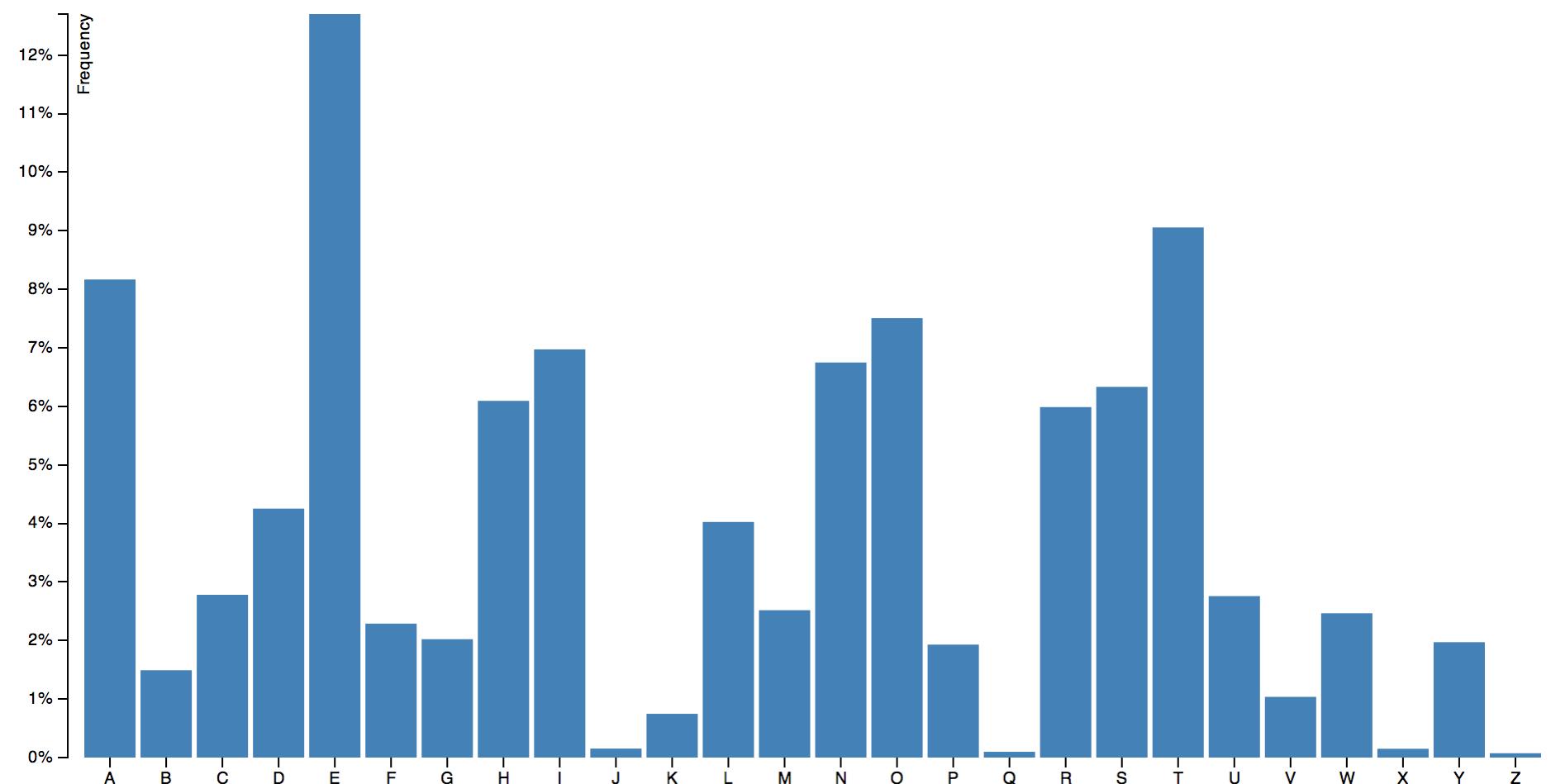


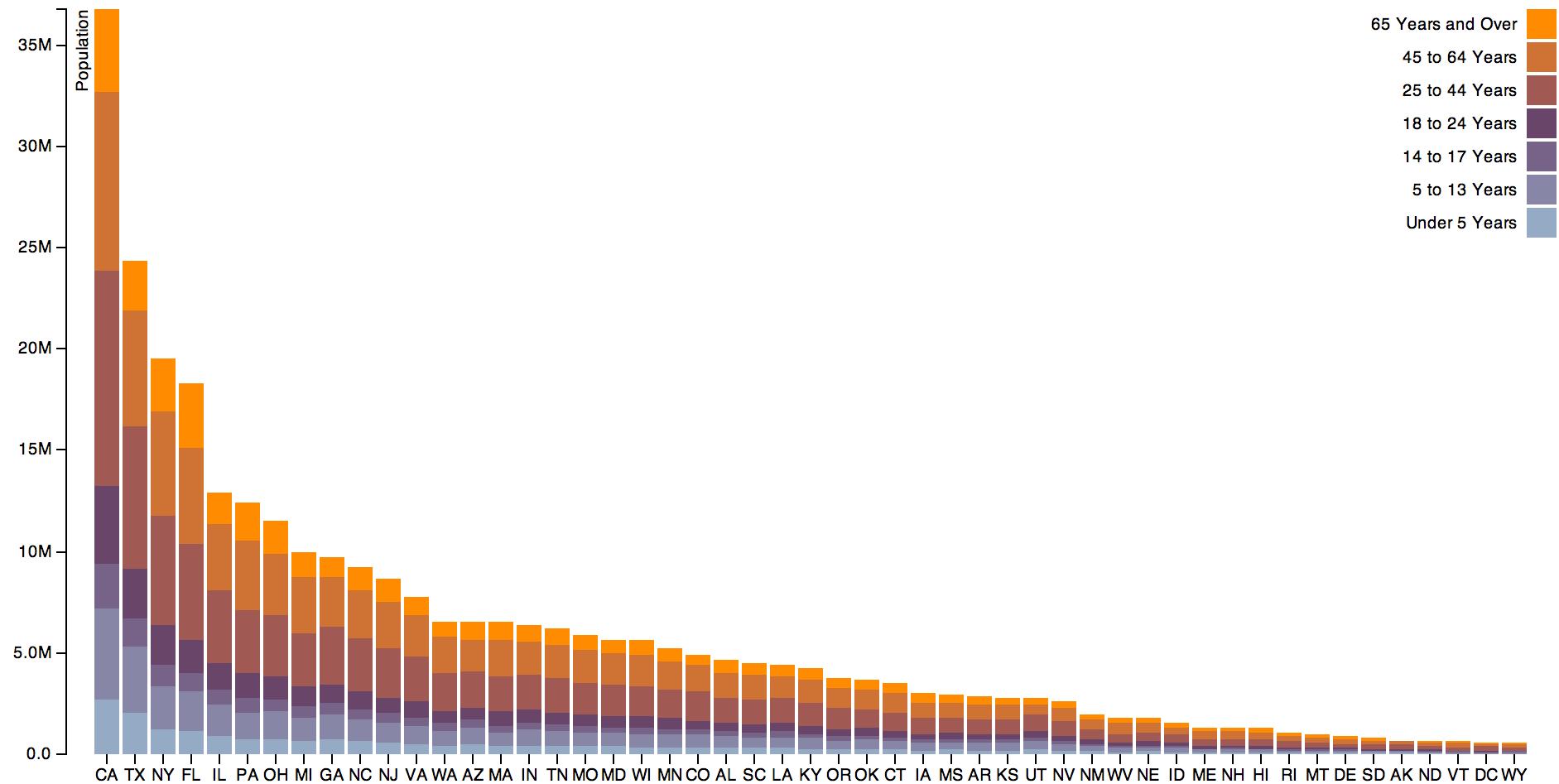
Spark Lines

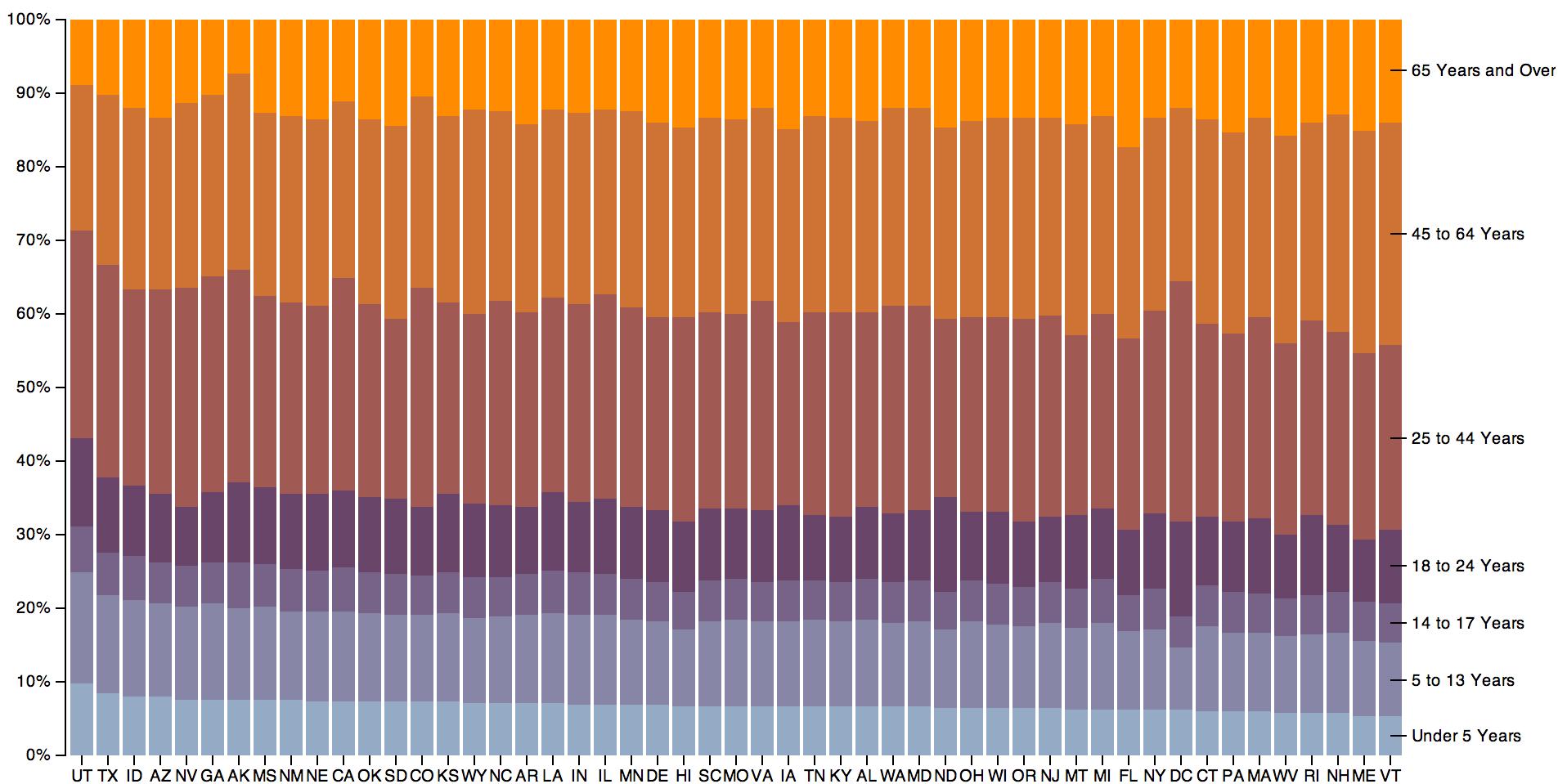


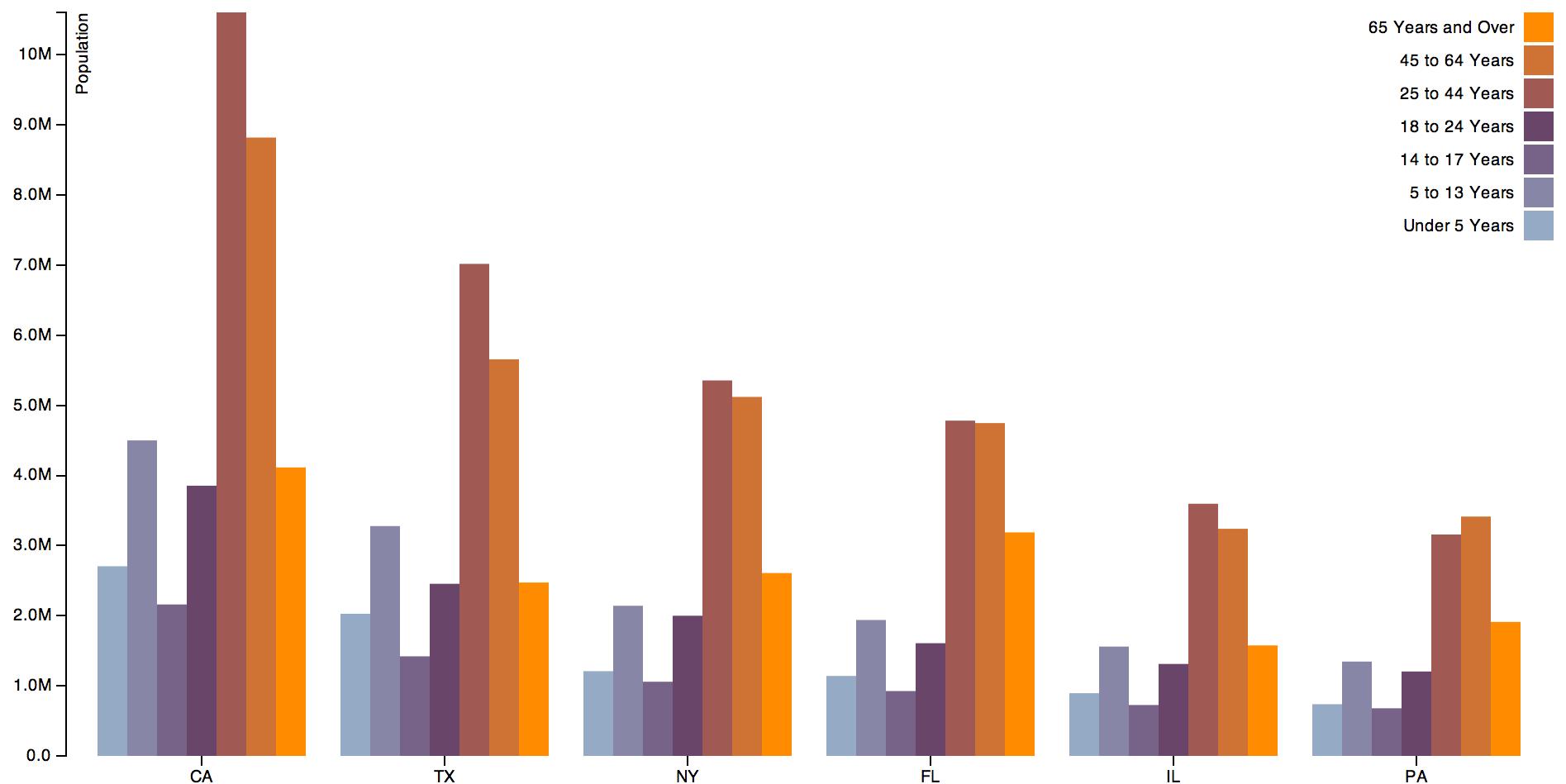
Bar Graphs (+ Variants)

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



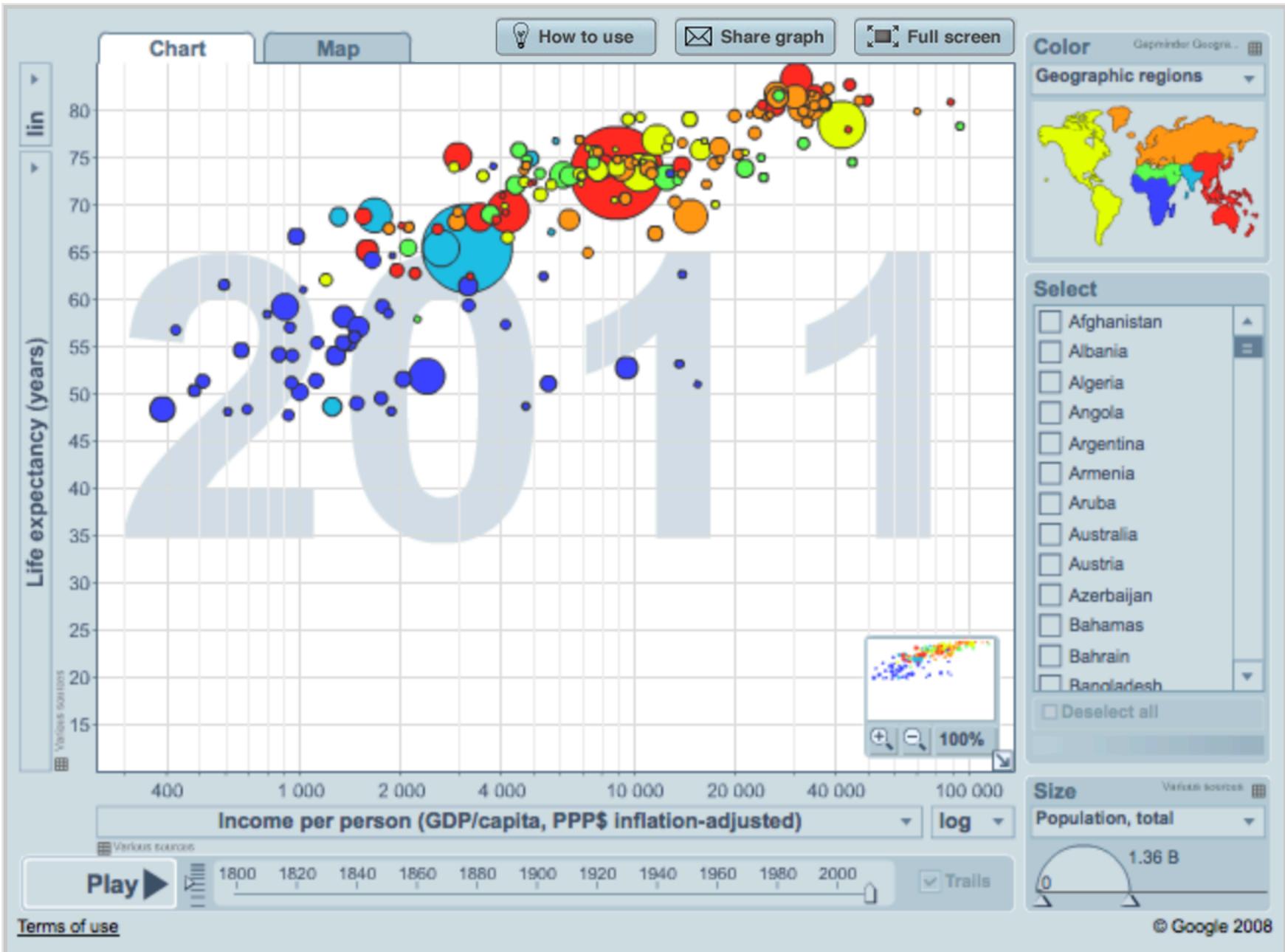






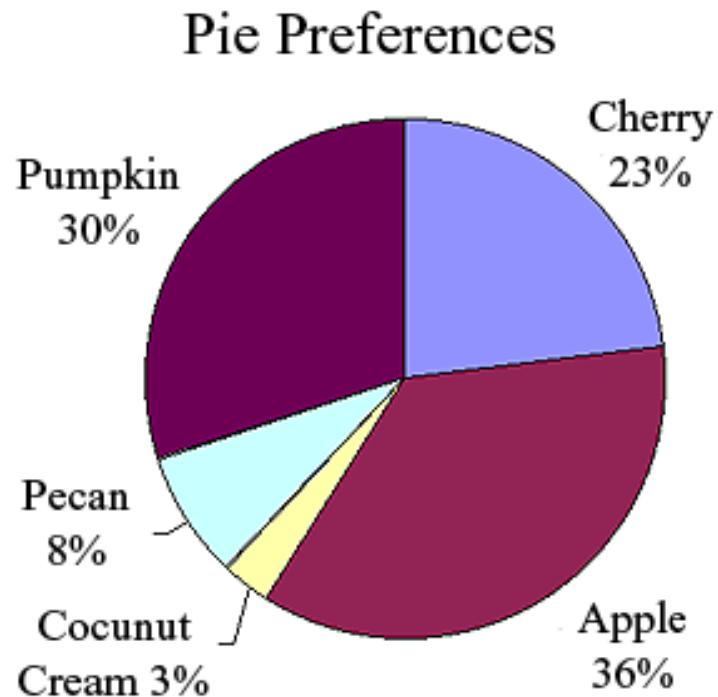
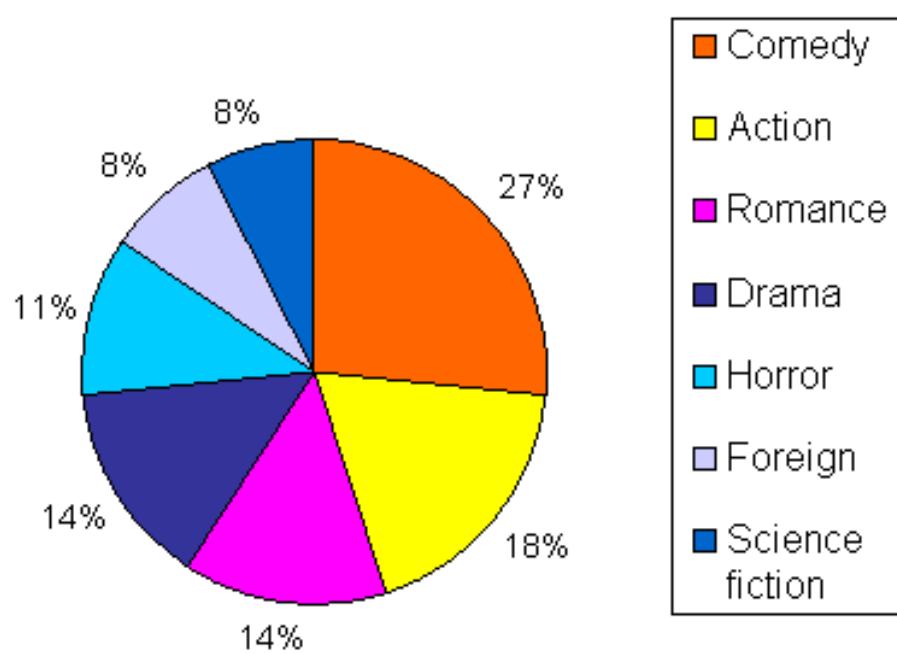
Scatterplots

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

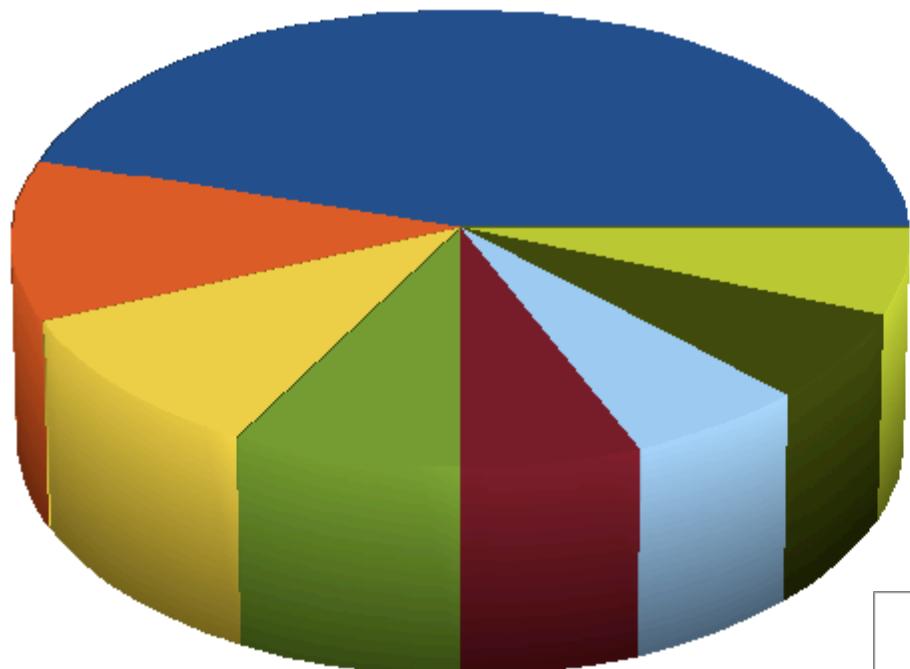


Pie Charts (+ Variants)

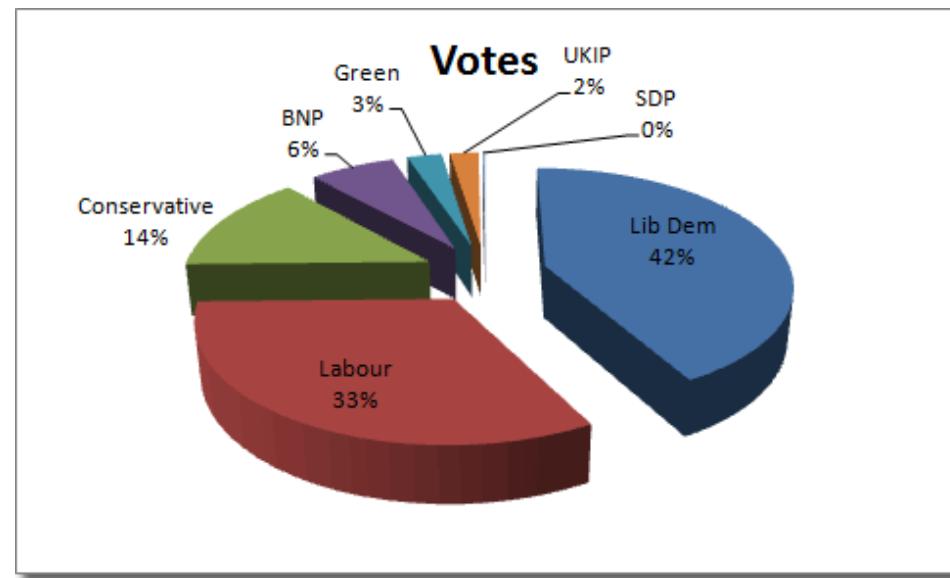
- Suitable for what types of data?
- Different ways for labeling



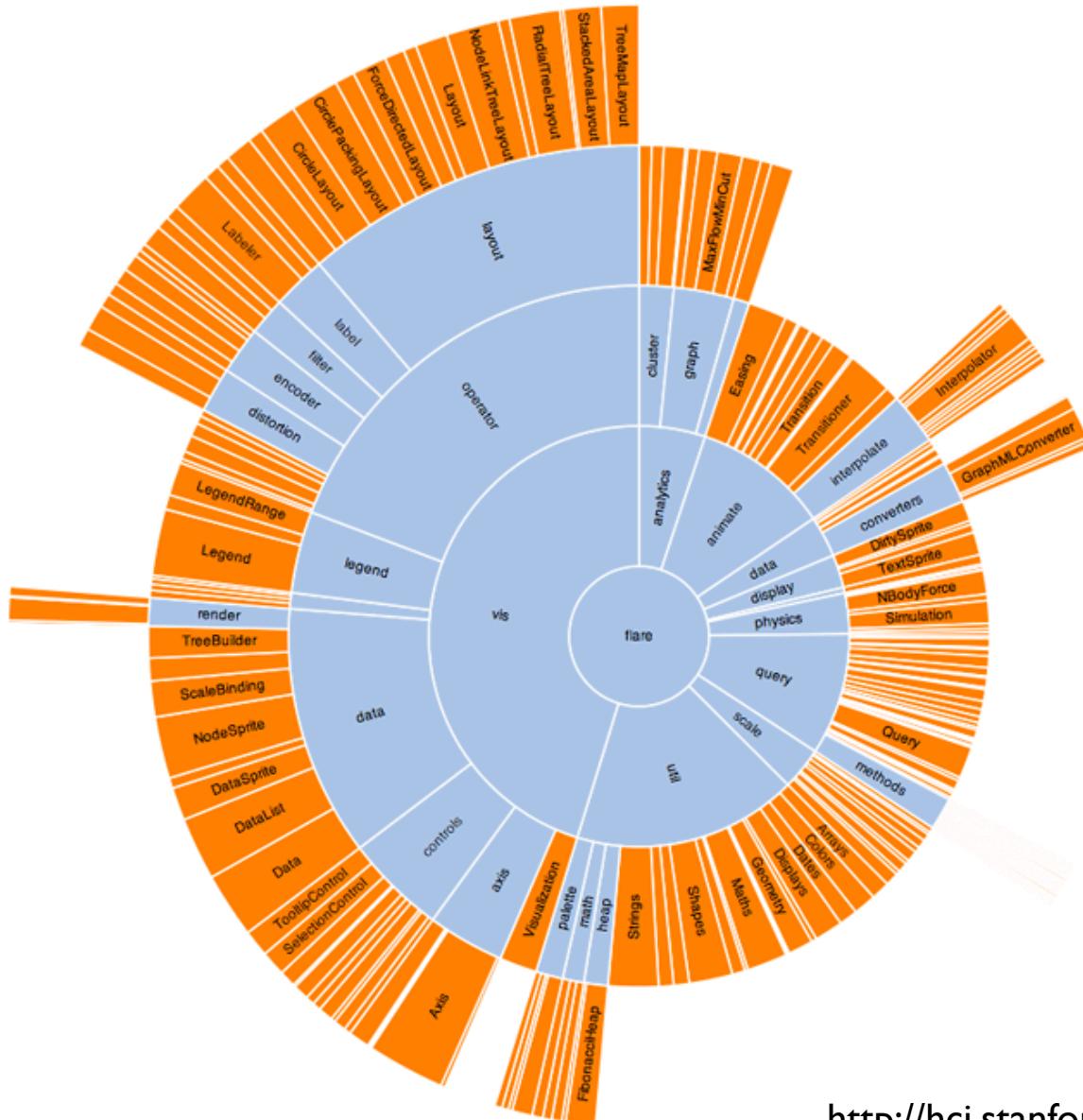
More Chart Junk



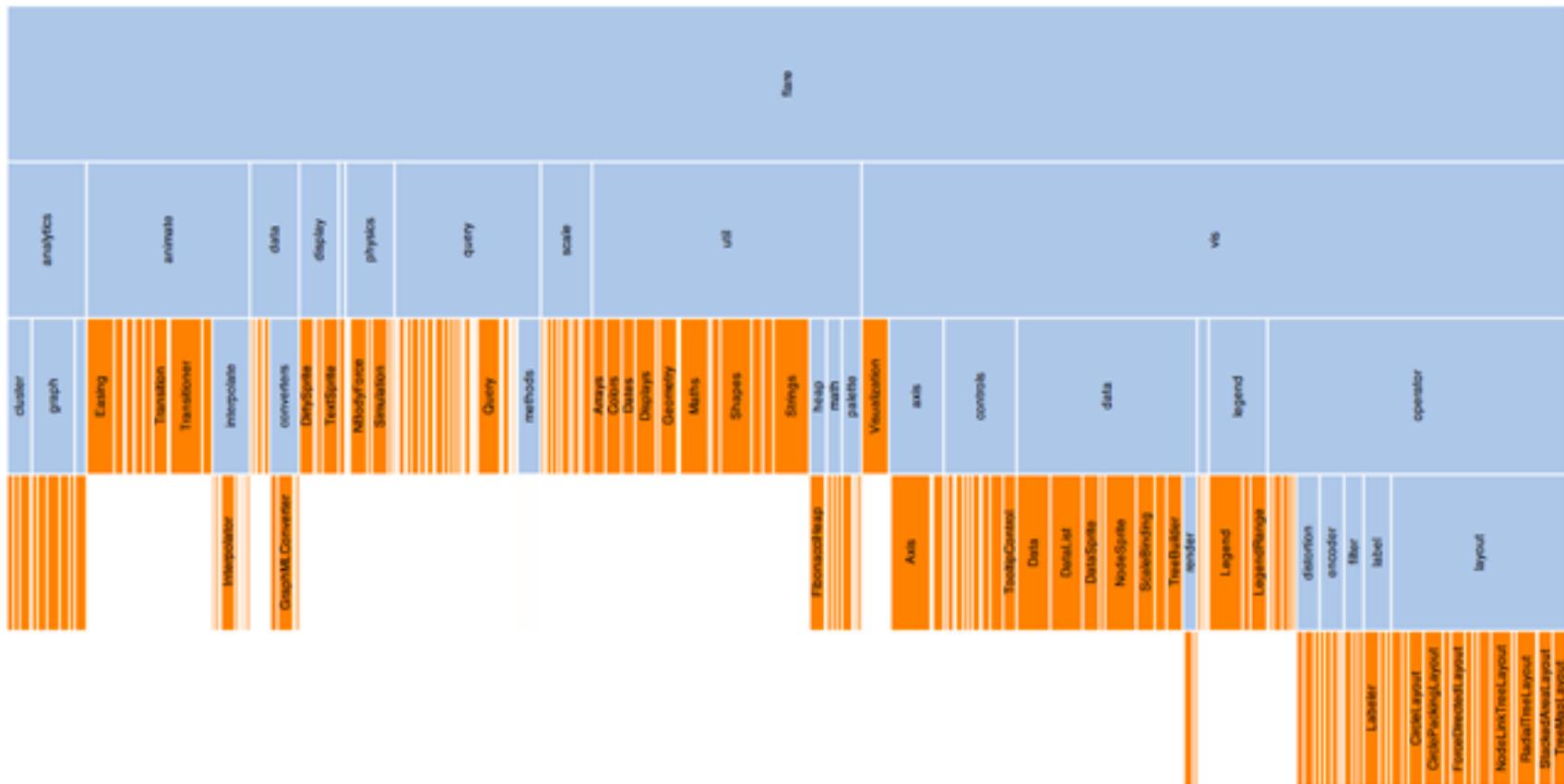
■ MySpace
■ Facebook
■ Orkut
■ hi5
■ Xanga
■ Live Spaces
■ Classmates.com
■ Bebo



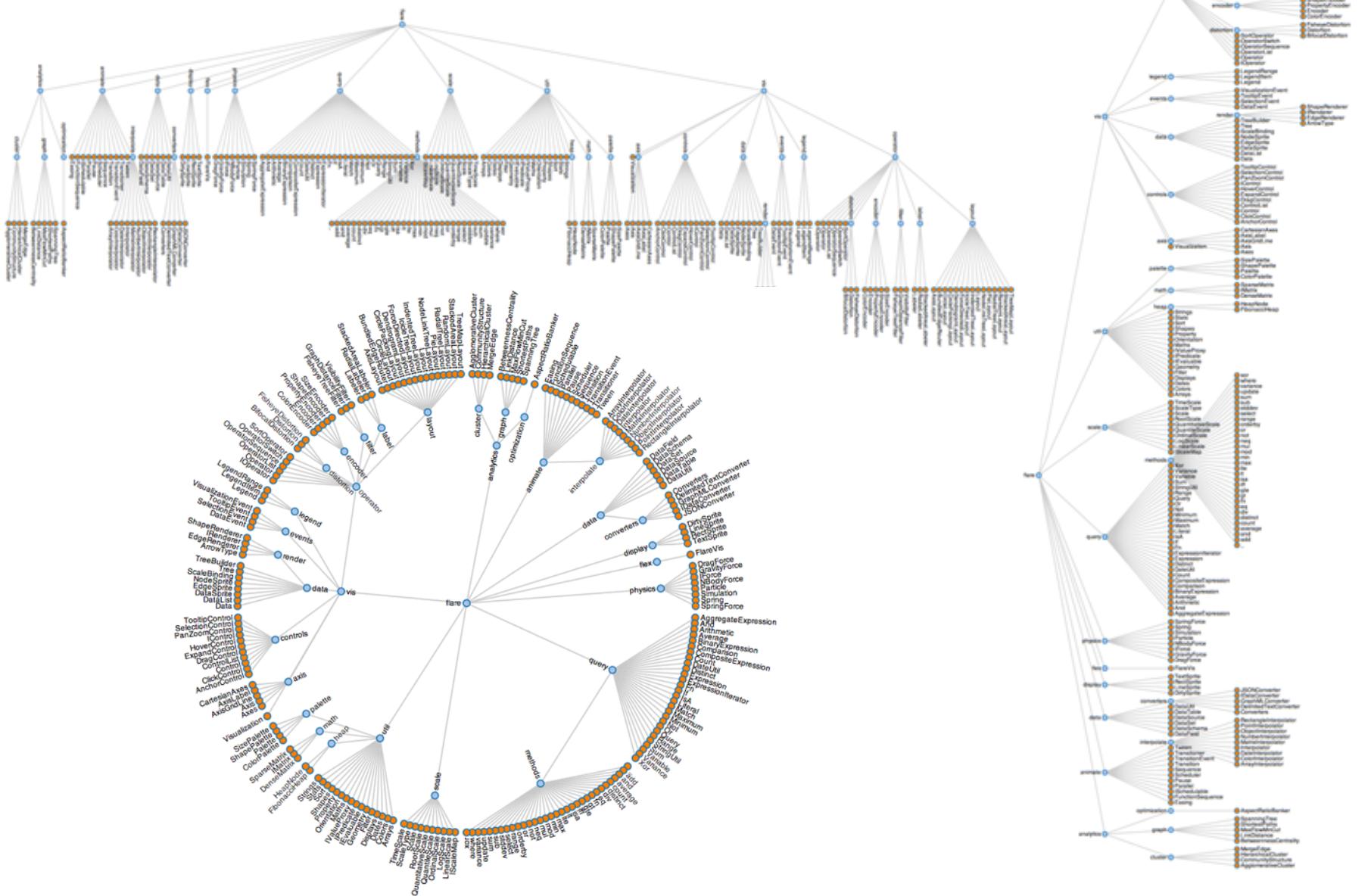
Sunburst Trees



Icicle Diagrams

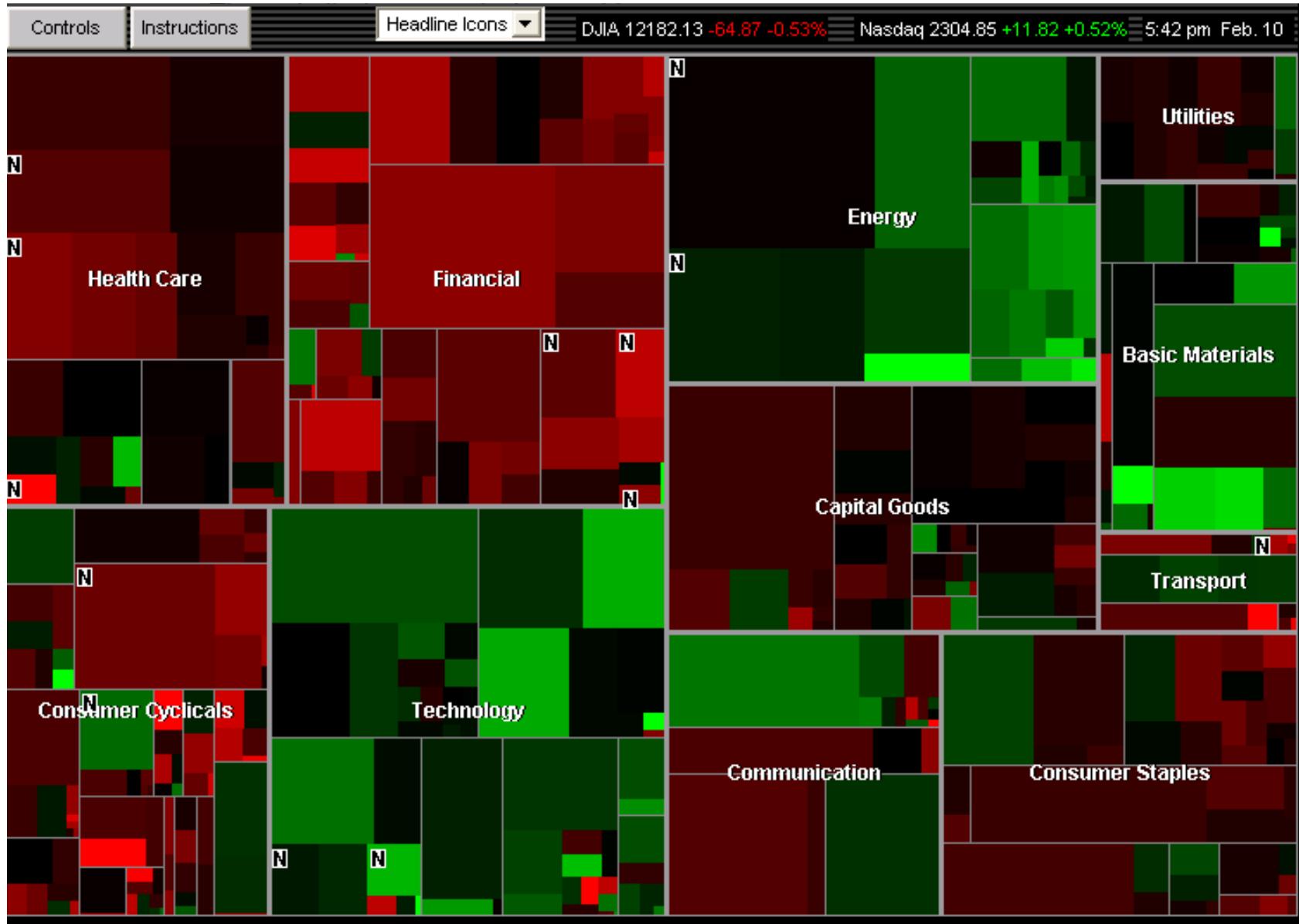


Visualizing Hierarchies



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



Controls

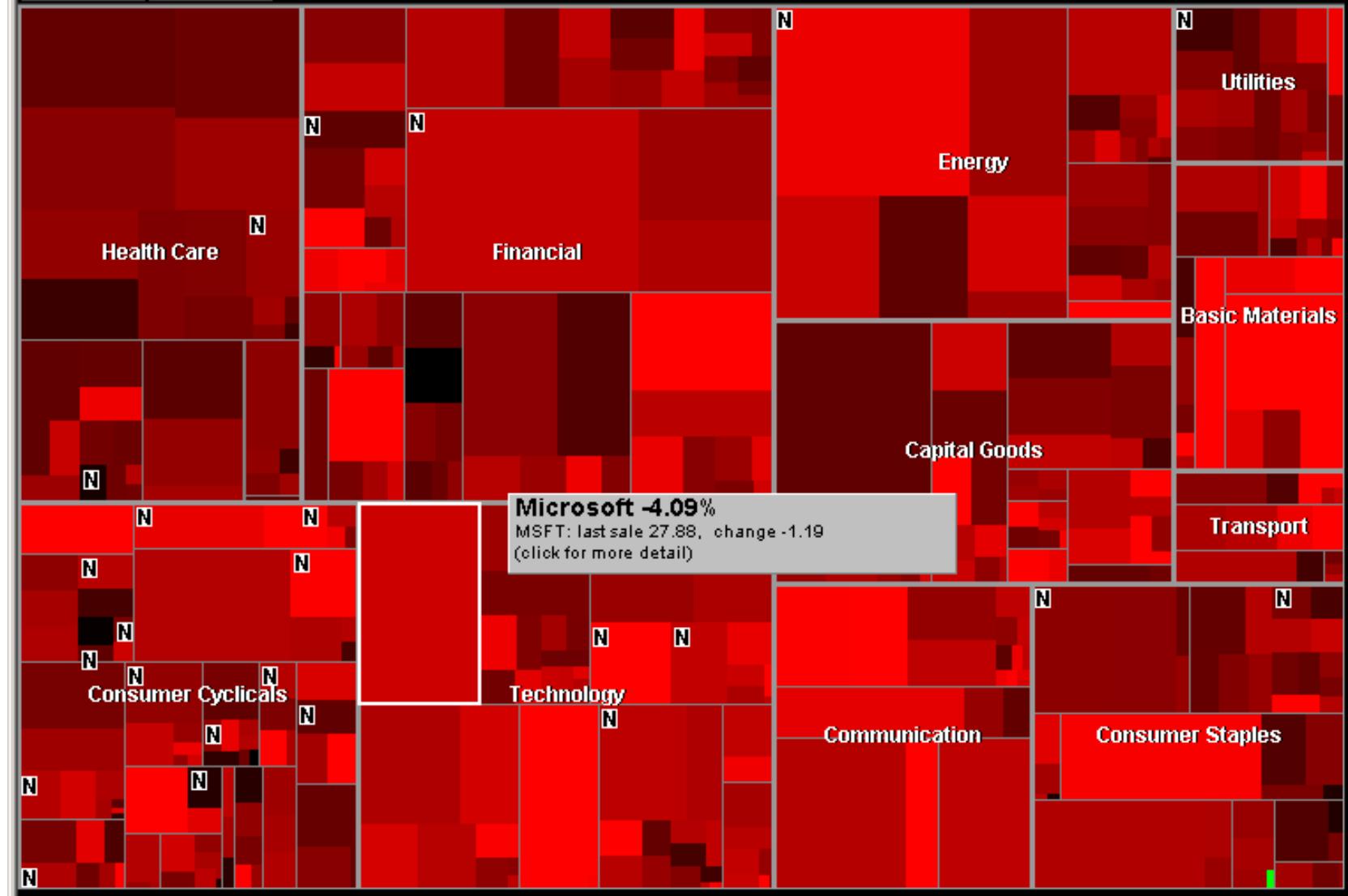
Instructions

Headline Icons

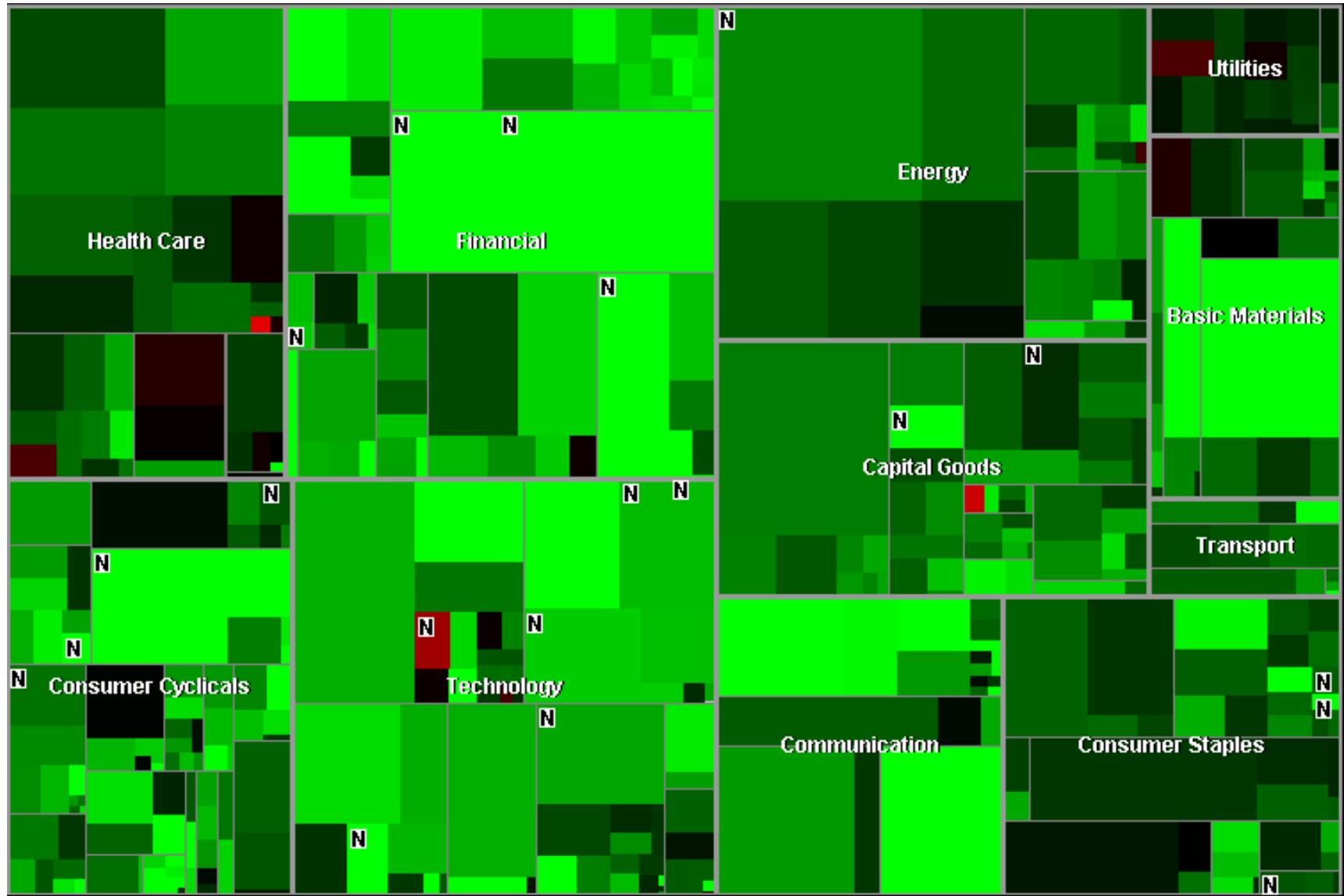
DJIA 12216.24 -416.02 -3.29%

Nasdaq 2407.87 -96.65 -3.86%

4:32 pm Feb. 27



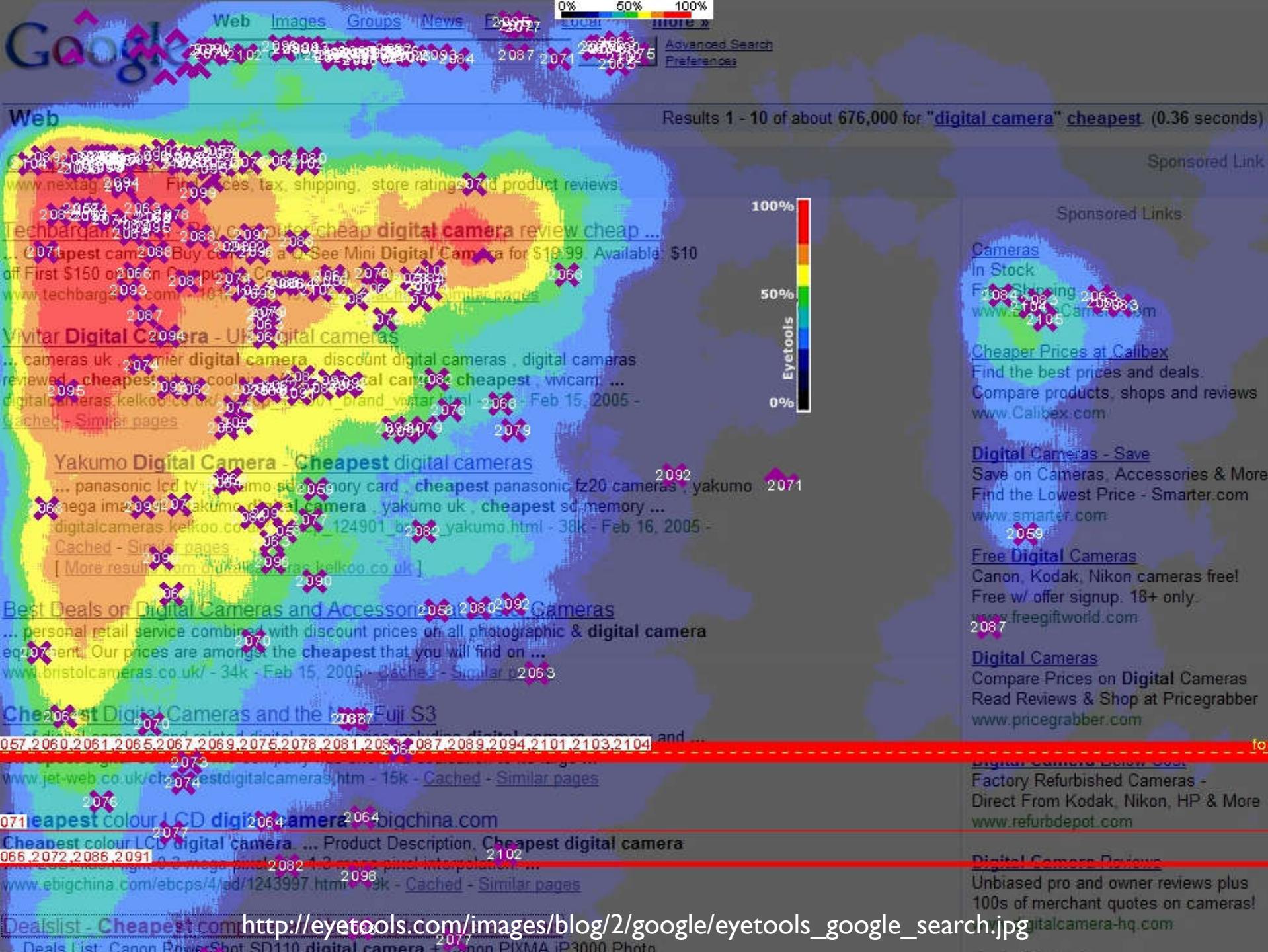
Market mixed, February 8, 2008:
Energy & Technology up Financial & Health Care down



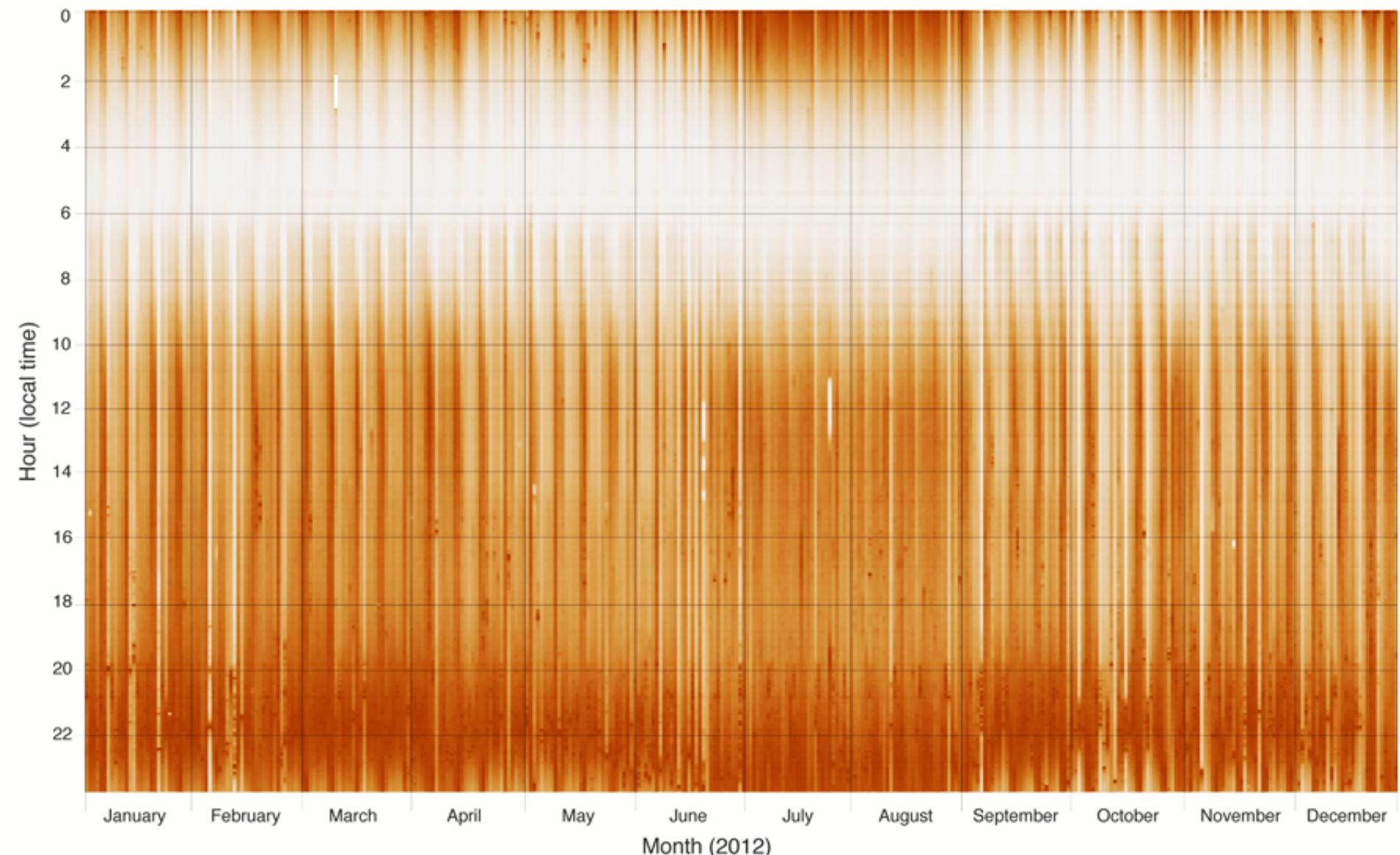
Market rises 319 points, November 13, 2007:
Five exceptions

Heatmaps

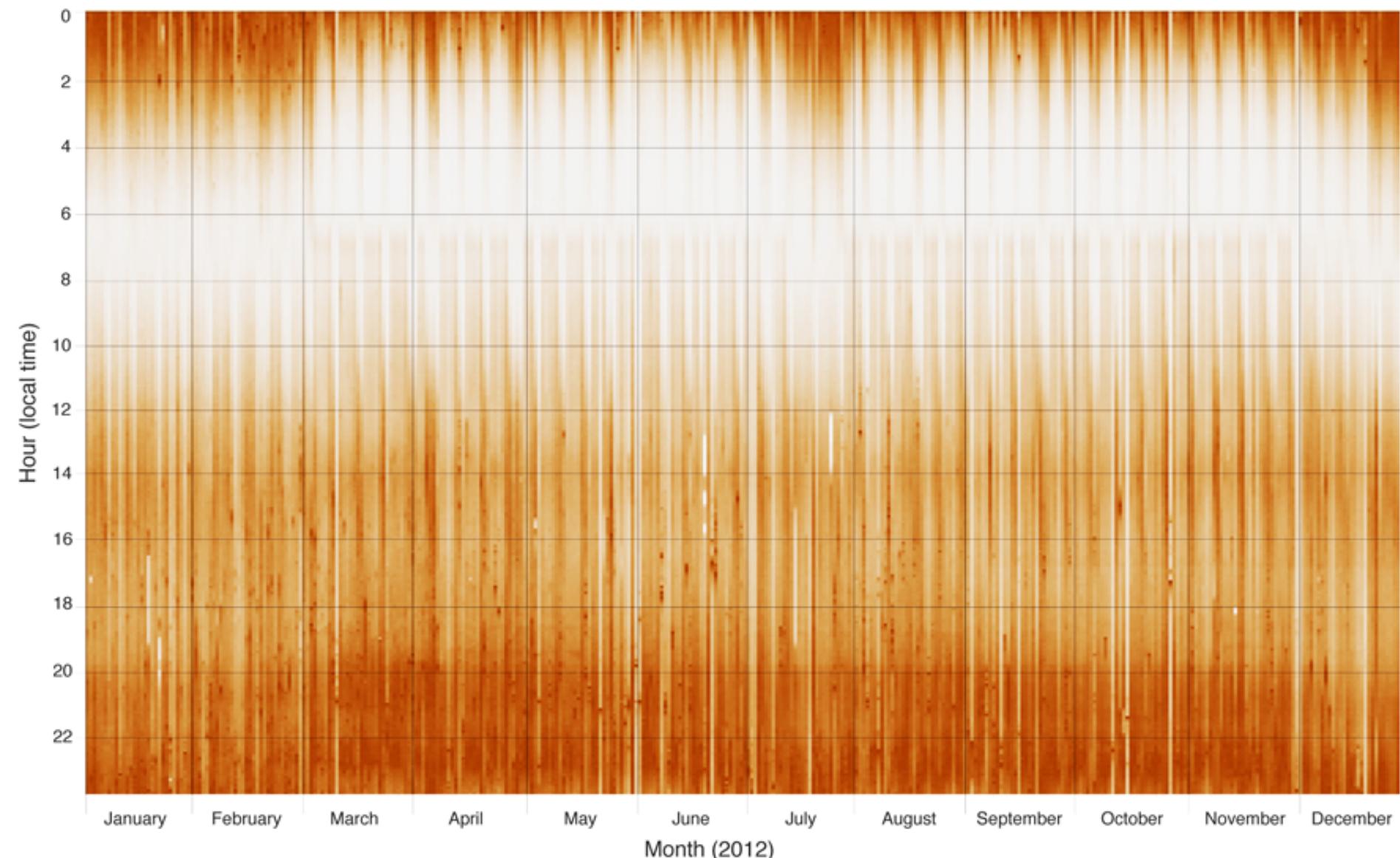
- Data represented in a matrix
- Matrix values are color coded



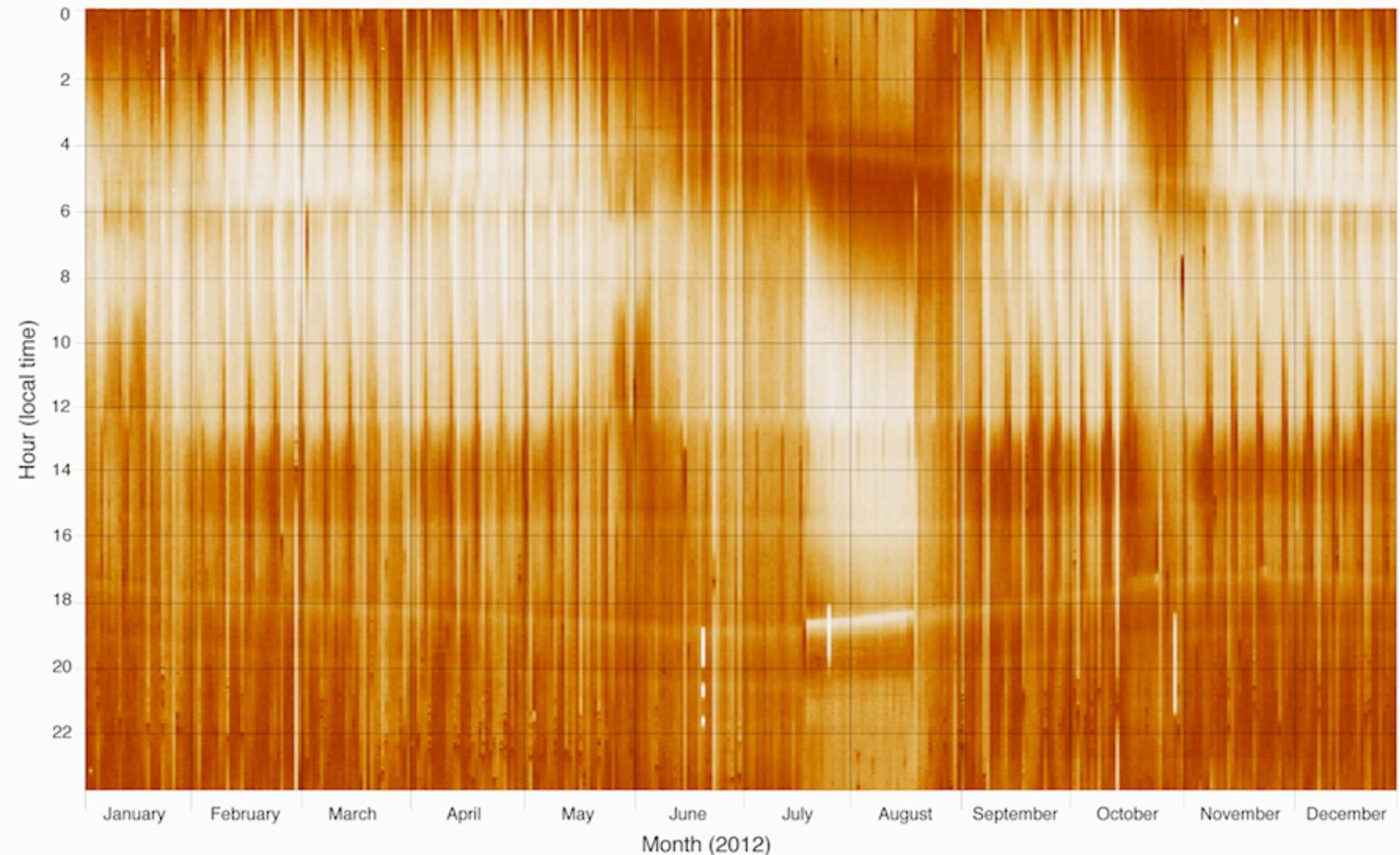
Tweet Activity (New York City, New York, US)



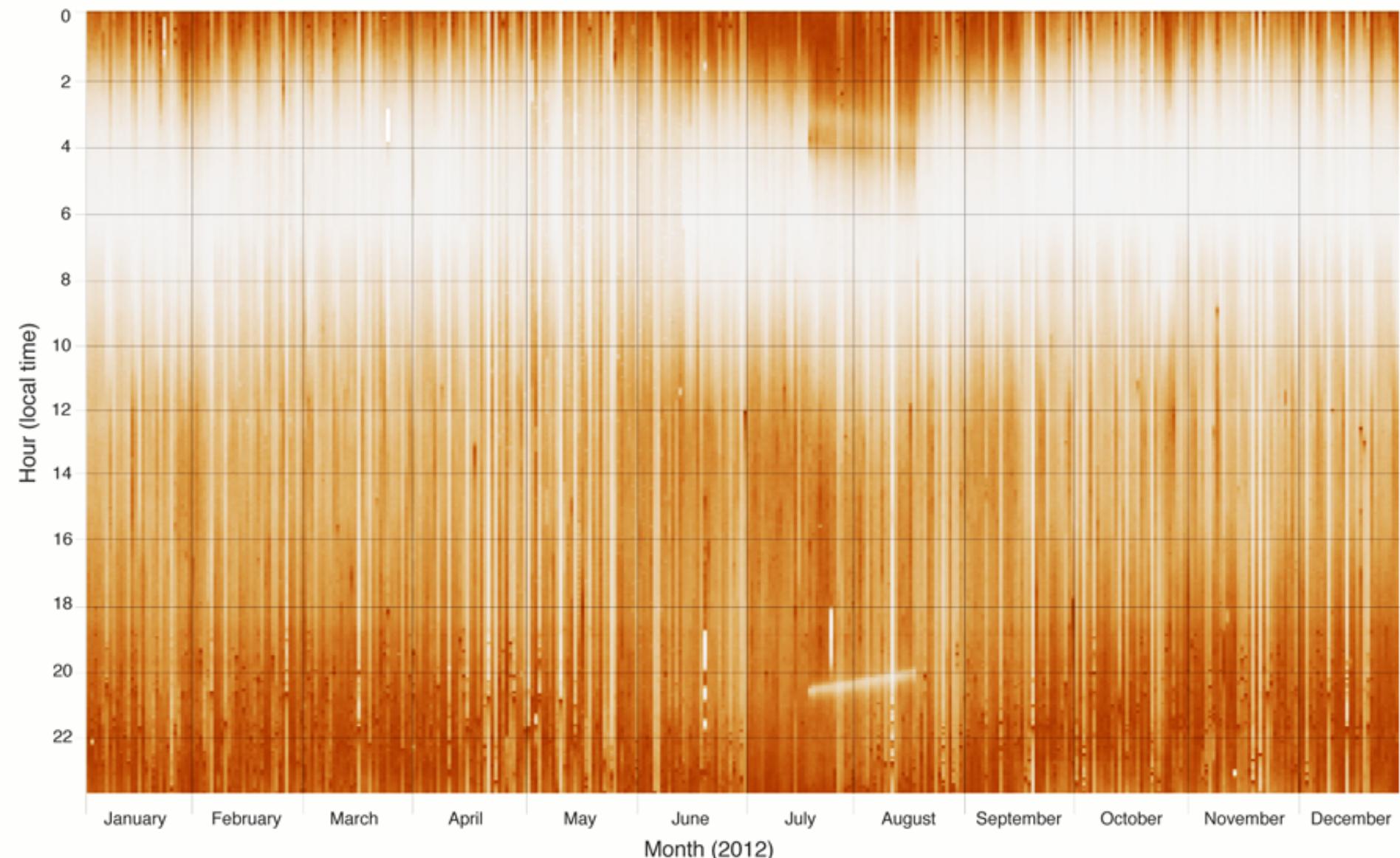
Tweet Activity (Buenos Aires, Argentina)



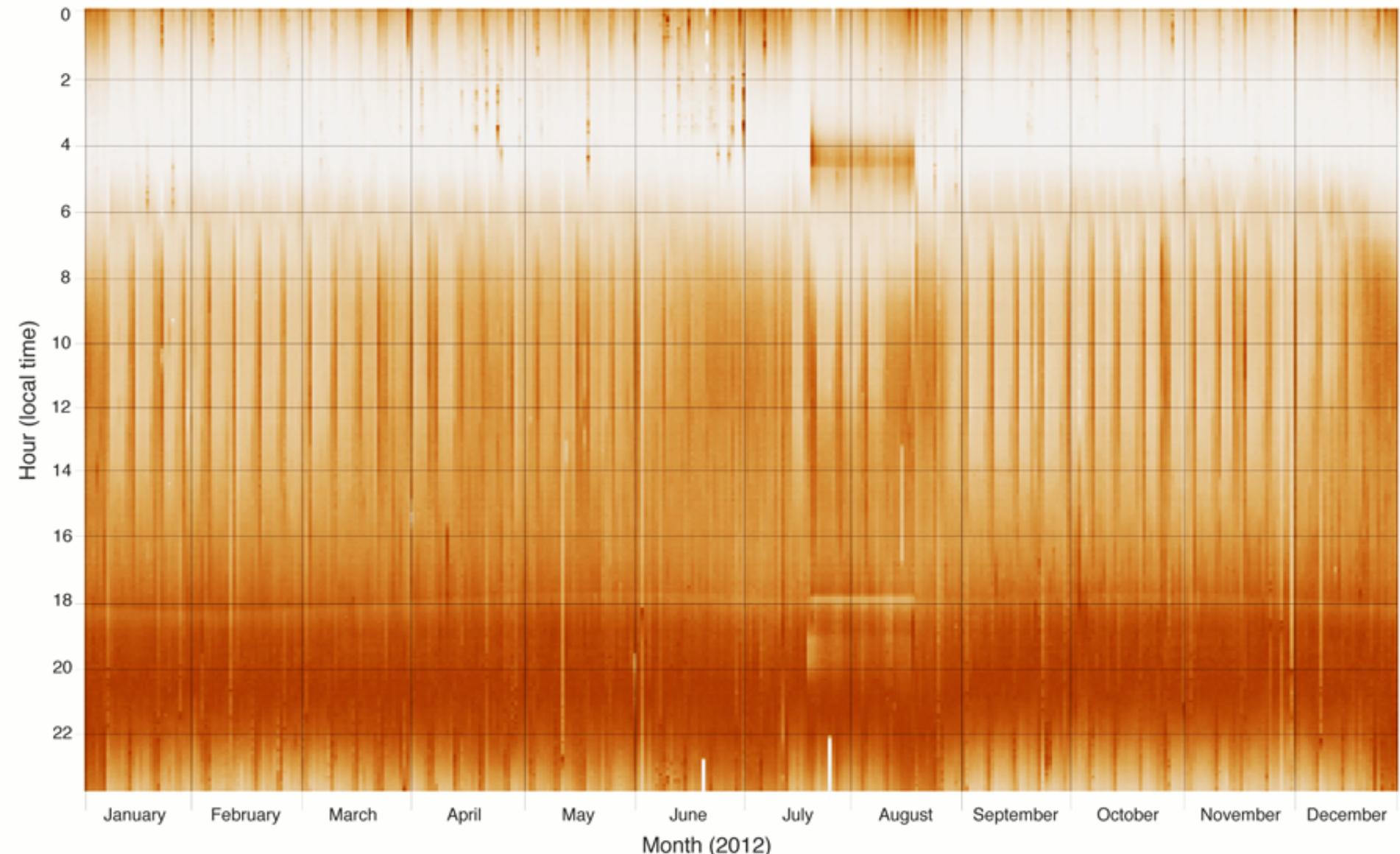
Tweet Activity (Riyadh, Saudi Arabia)



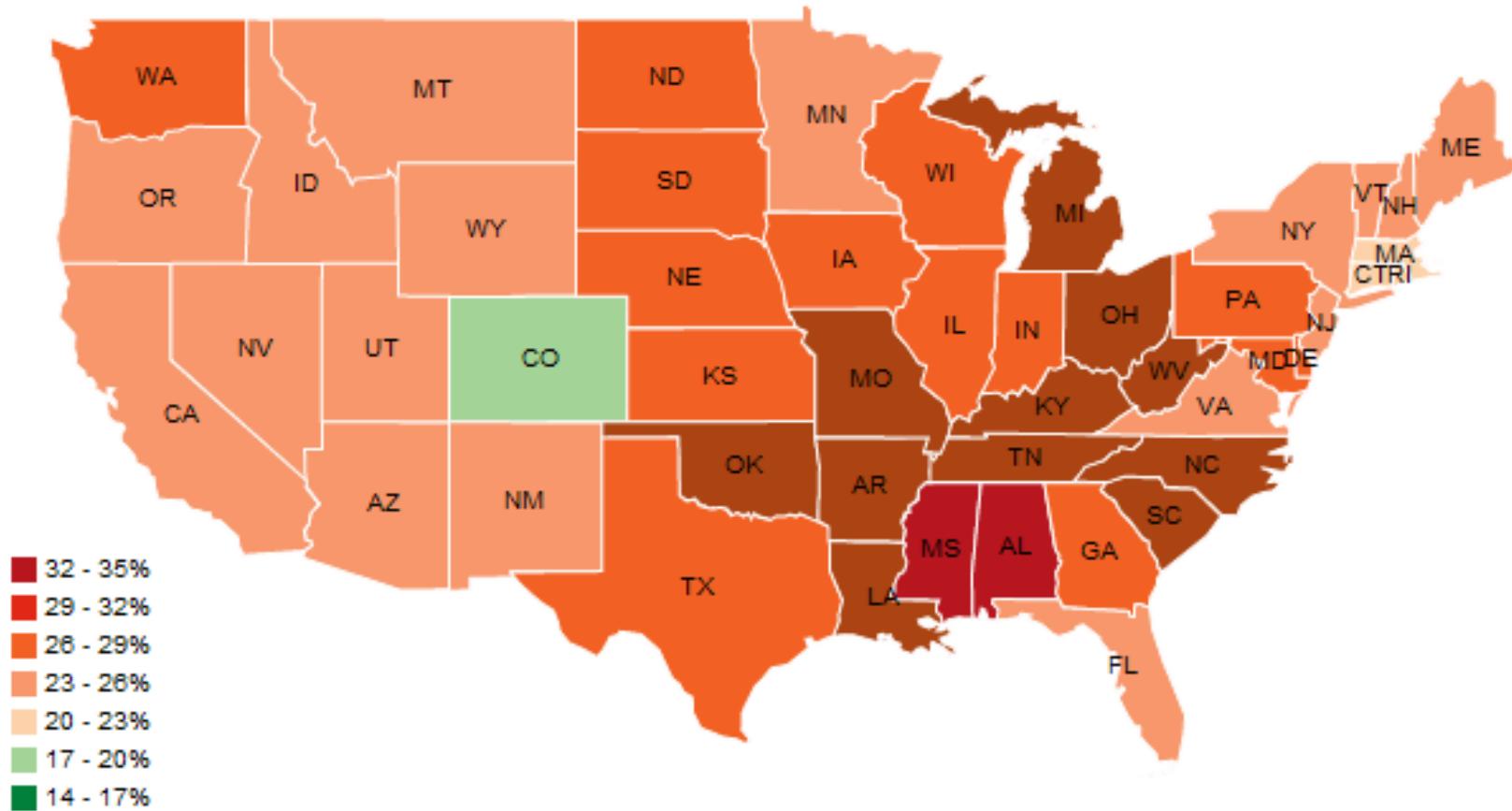
Tweet Activity (Istanbul, Turkey)



Tweet Activity (Jakarta, Indonesia)

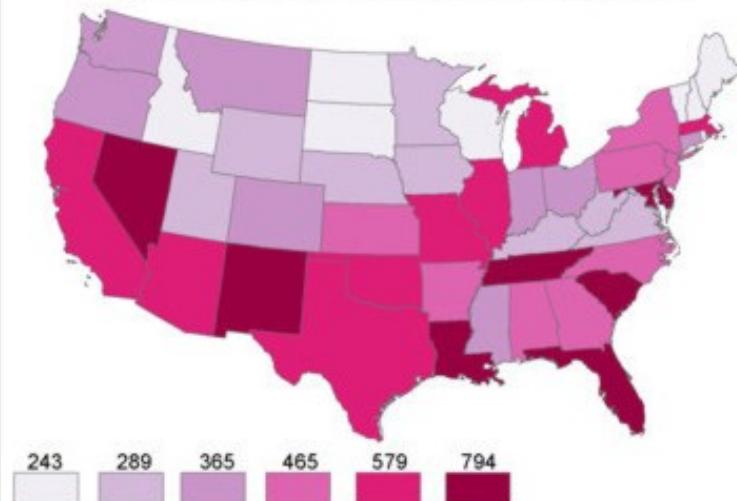


Choropleth Maps

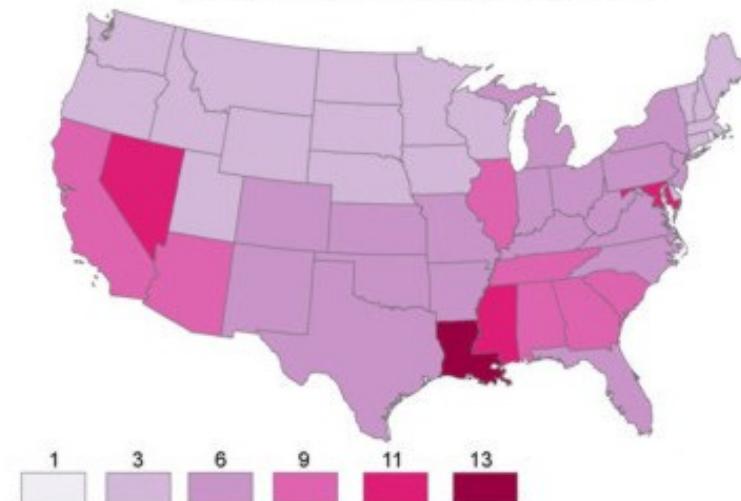


Crime Rates in the US - 2003

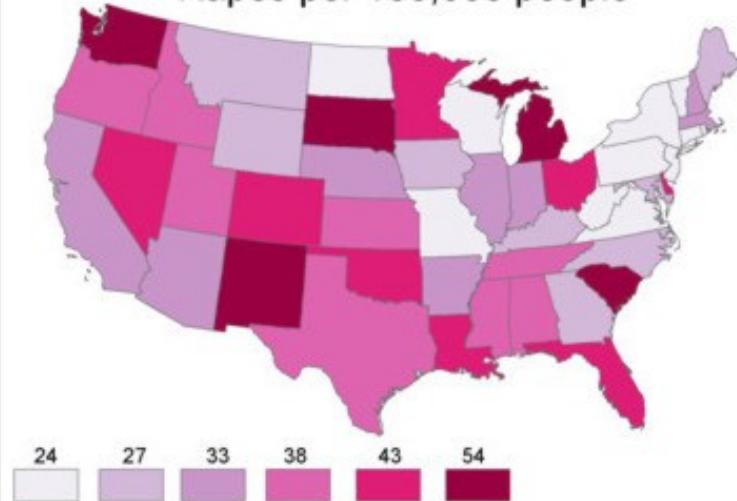
Violent crimes per 100,000 people



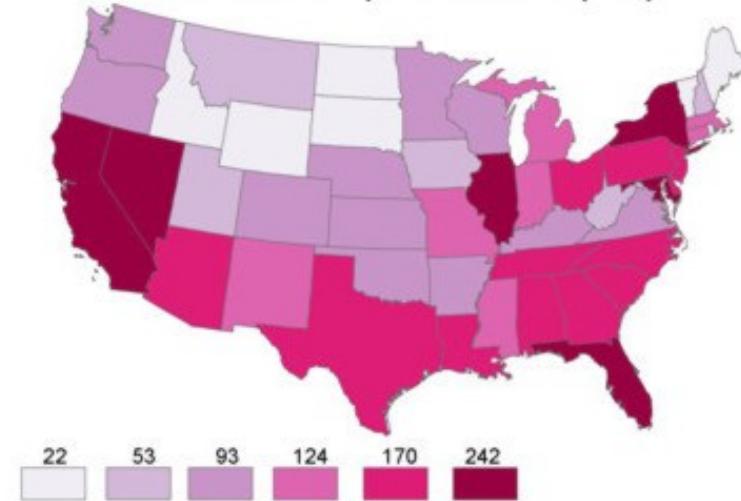
Murders per 100,000 people



Rapes per 100,000 people

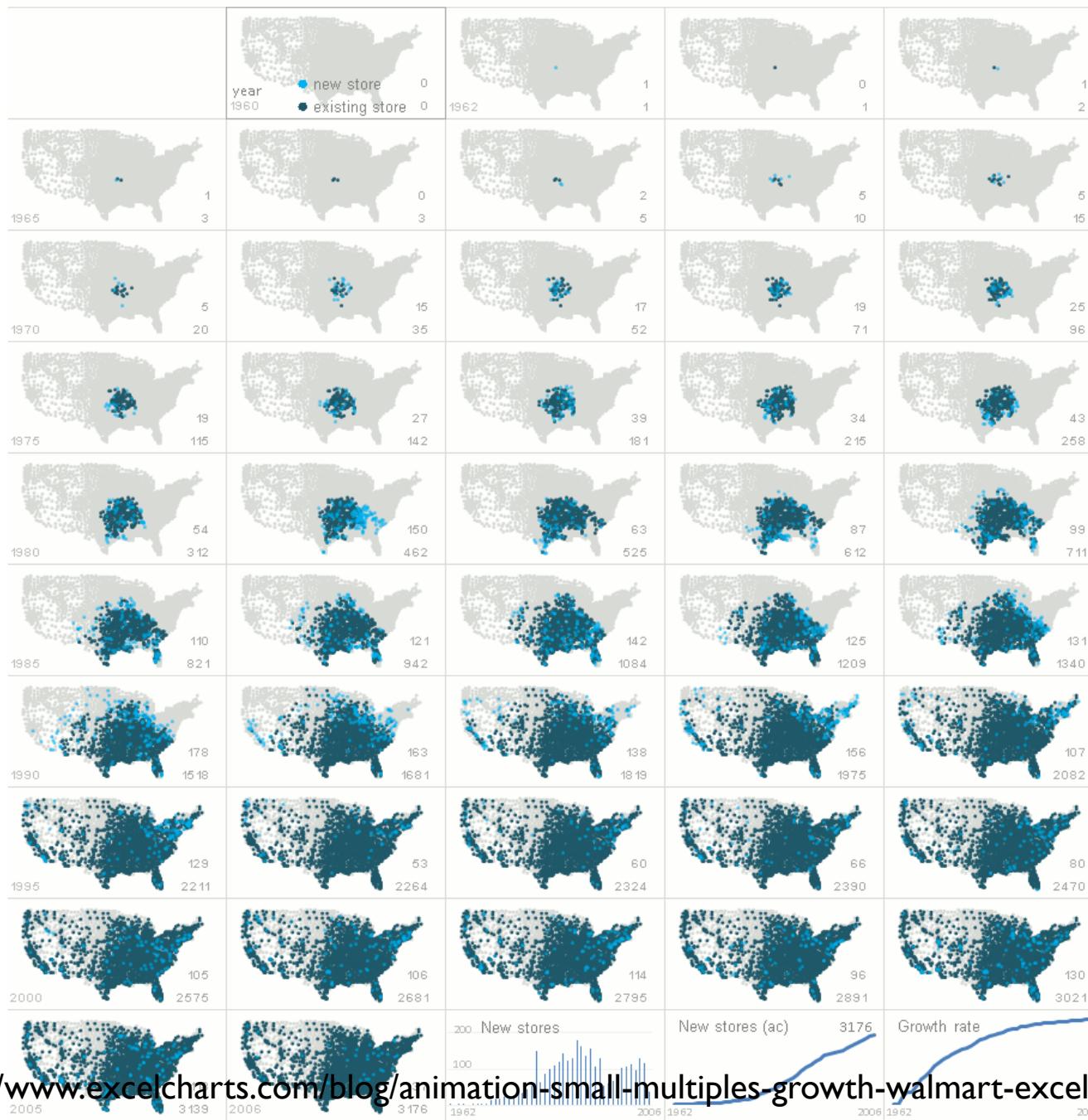


Robberies per 100,000 people



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

Growth of Walmart

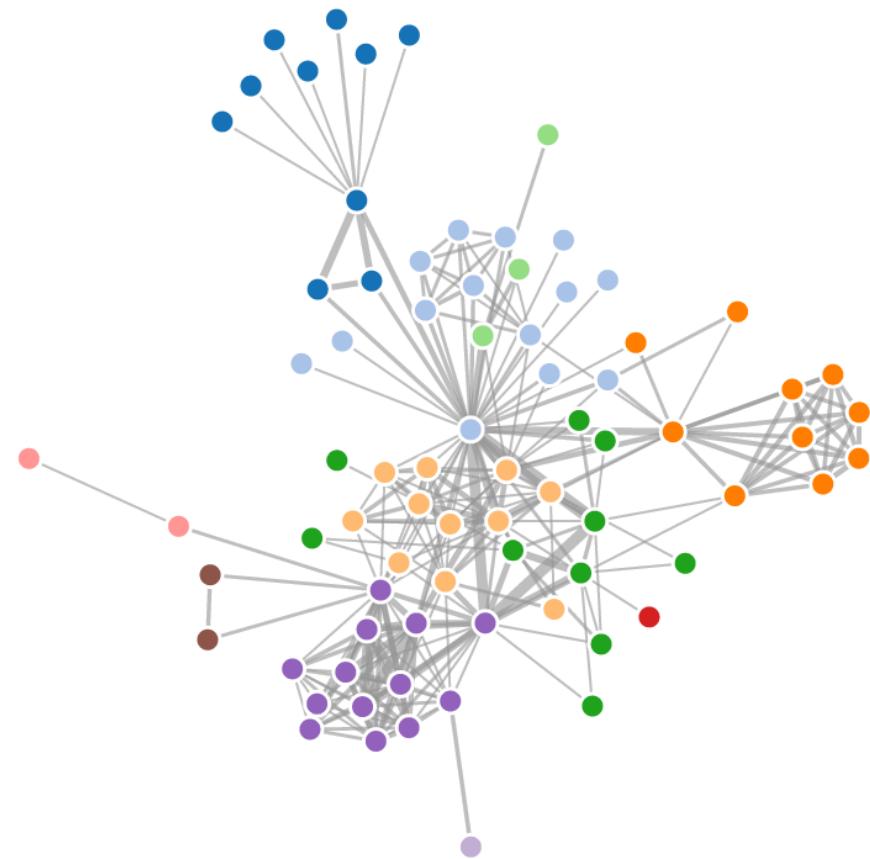


Small Multiples vs. Animations

- 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
9am - 5:30pm

