

**INFM 603: Information Technology and Organizational Context**

# **Session 9: Visualization**



**Jimmy Lin  
The iSchool  
University of Maryland**

**Thursday, November 6, 2014**

# 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

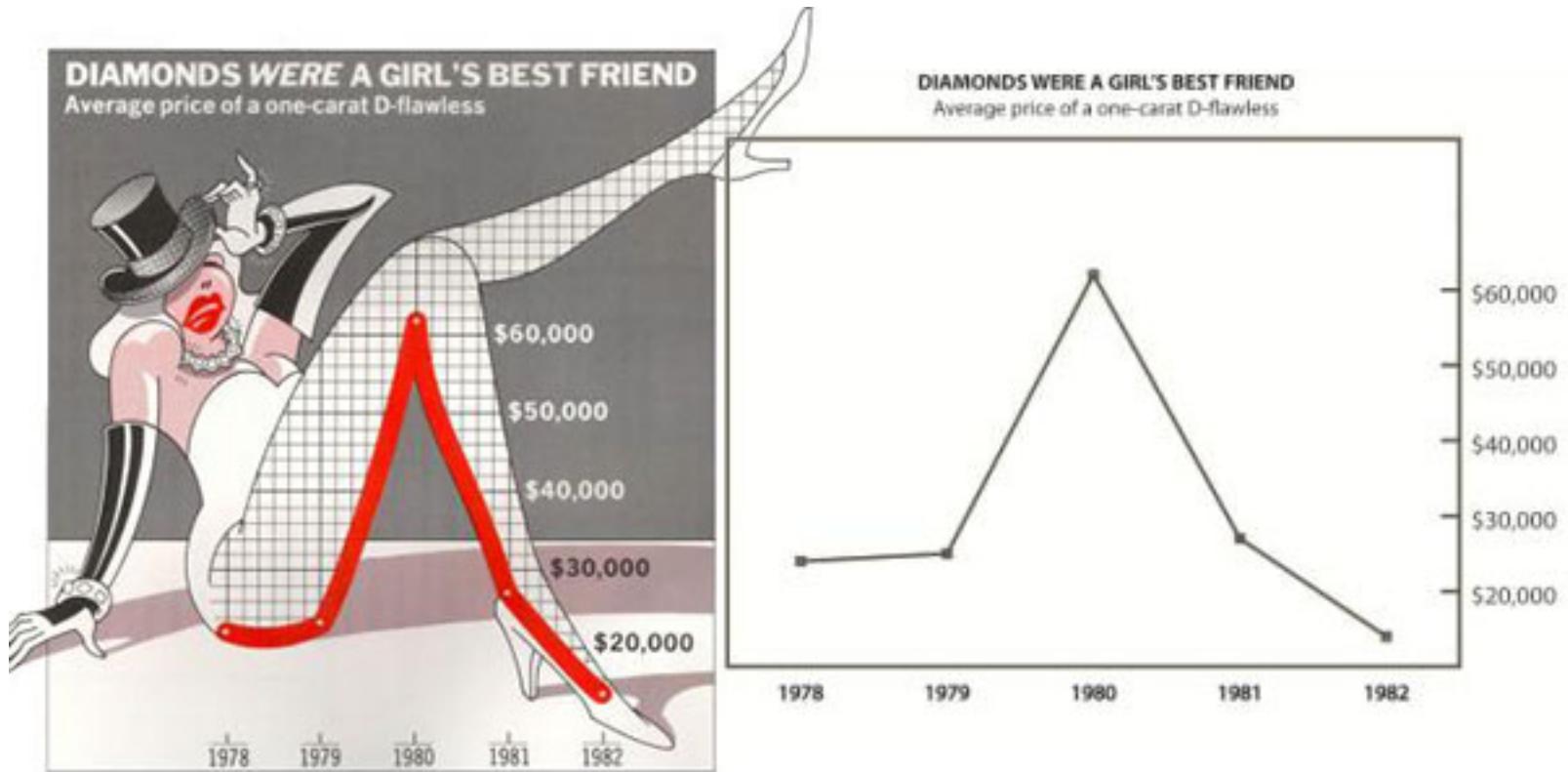


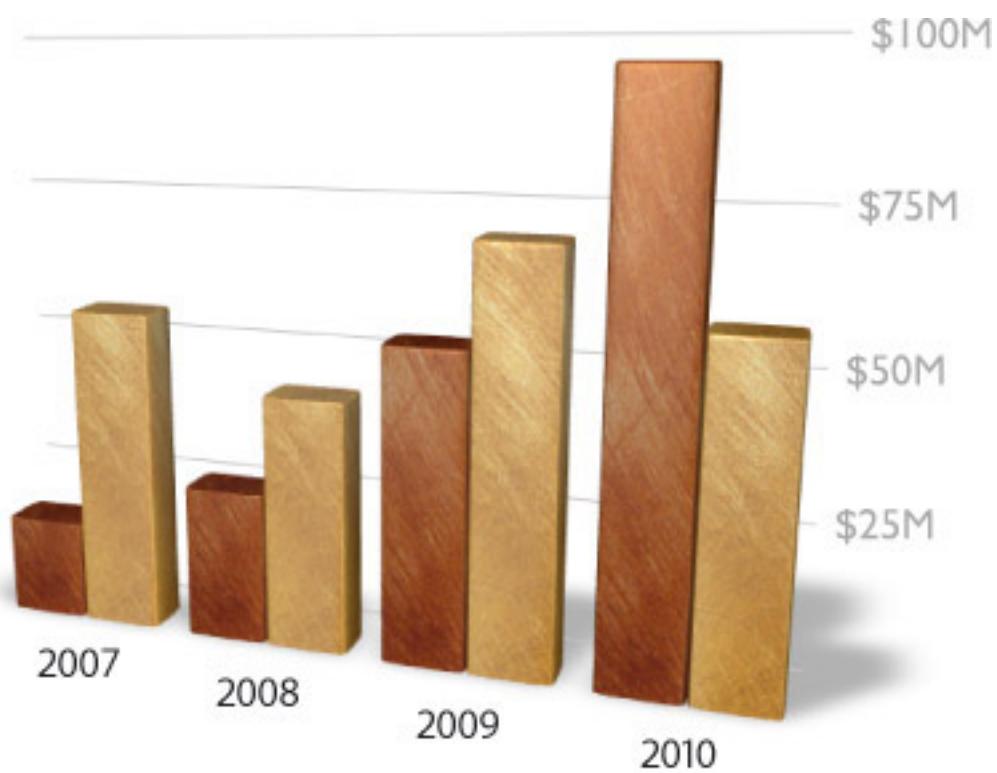
# Static vs. Interactive

- 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



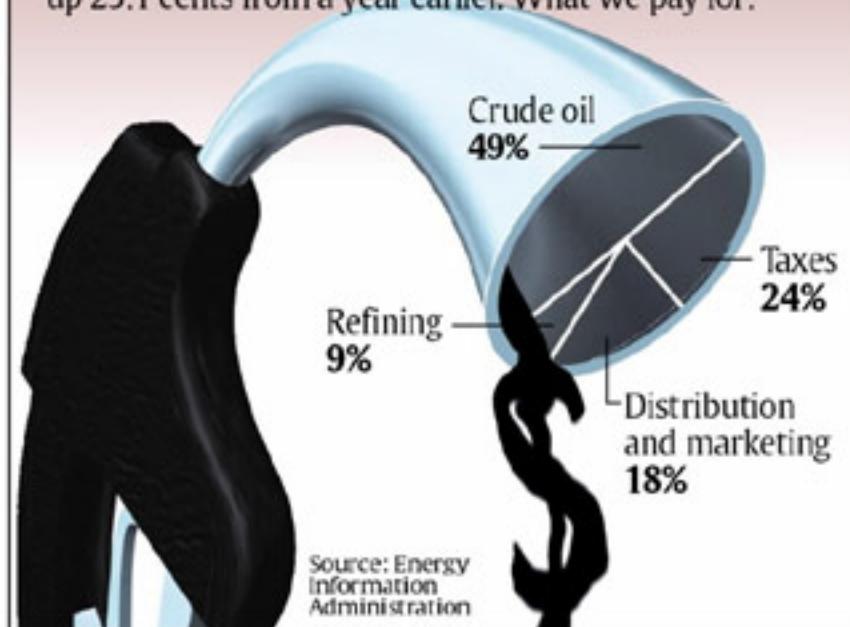


---

## 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%  
Performance history

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

♂ +11%  
Performance history

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

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

13.0 s  
12.5 s  
12.0 s  
11.5 s  
11.0 s  
10.5 s  
10.0 s  
9.5 s

**MARATHON**  
♂ +20%  
Performance history

WILL THE 2-HOUR LIMIT BE BROKEN SOMEDAY?

2010  
HALF  
GERMANY  
KEN 2011

**HIGH JUMP**  
+30%  
Performance history

PERFORMANCE HAS LEVELED SINCE THE 1990s.  
2.45 m JAVIER SOTOMAYOR CUBA 1993

1918: Hubbard's accidentally does a backflip over the bar.  
1920: US athlete Frank Wykoff sets a record of 2.34 m.  
1924: US athlete Harry Edwards sets a record of 2.34 m.  
1928: US athlete Eddie Tolan sets a record of 2.34 m.

1932: Florence Griffith-Joyner USA 10.60 s  
1936: Jim Hines USA 9.05 s  
1968: Maurice Greene USA 9.79 s  
1993: Usain Bolt JAM 9.74 s  
2007: Asafa Powell JAM 9.74 s  
2009: Usain Bolt JAM 9.58 s

2010: The 2009 edition of the IAAF World Championships in Berlin was proven wrong by Usain Bolt.

1972–1979: Athletes from the United States and West Germany dominate the men's high jump. Soviet Union's Valeriy Burev sets a record of 2.40 m.

1980–1985: Athletes from the Soviet Union and East Germany dominate the women's high jump. Soviet Union's Galina Kotsenko sets a record of 2.29 m.

1986–1991: Athletes from the United States and Soviet Union dominate the men's high jump. Soviet Union's Sergey Bubka sets a record of 2.35 m.

1992–1997: Athletes from the United States and Cuba dominate the women's high jump. Cuban Yarisley Silva sets a record of 2.29 m.

1998–2003: Athletes from the United States and Cuba dominate the men's high jump. Cuban Yarisley Silva sets a record of 2.35 m.

2004–2009: Athletes from the United States and Cuba dominate the women's high jump. Cuban Yarisley Silva sets a record of 2.35 m.

2010–2011: Athletes from the United States and Cuba dominate the men's high jump. Cuban Yarisley Silva sets a record of 2.35 m.

**POLE VAULT**  
+71%  
Performance history

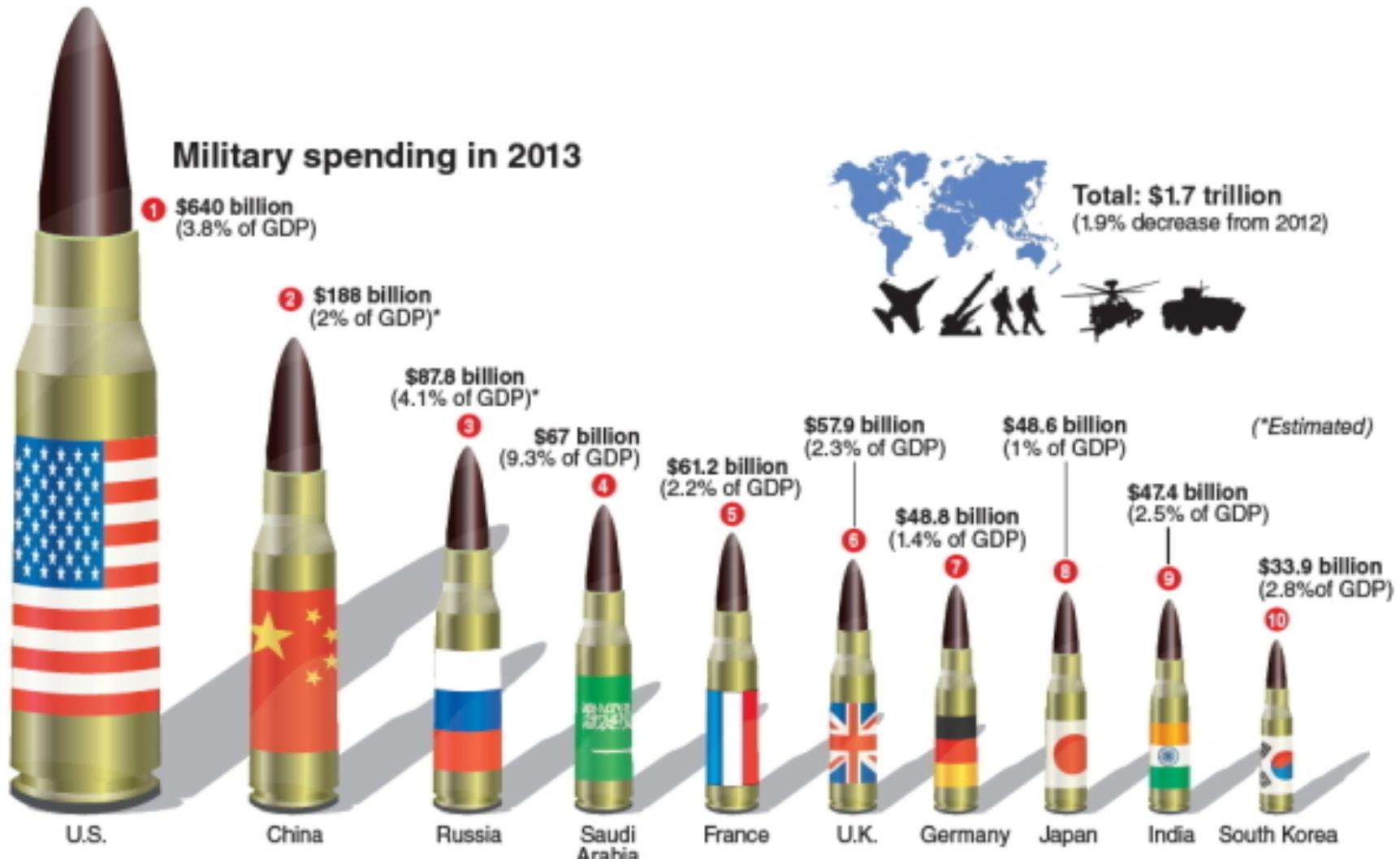
PERFORMANCES HAVE DECLINED SINCE 2000

6.14 m SERGEY BUBKA UKR 1994

1968: USSR's Sergey Bubka sets a record of 5.90 m.  
1972: USSR's Sergey Bubka sets a record of 6.00 m.  
1976: USSR's Sergey Bubka sets a record of 6.14 m.  
1980: USSR's Sergey Bubka sets a record of 6.20 m.  
1984: USSR's Sergey Bubka sets a record of 6.28 m.  
1988: USSR's Sergey Bubka sets a record of 6.34 m.  
1992: USSR's Sergey Bubka sets a record of 6.40 m.  
1996: USA's Sergey Bubka sets a record of 6.40 m.  
2000: USA's Sergey Bubka sets a record of 6.40 m.  
2004: USA's Sergey Bubka sets a record of 6.40 m.  
2008: USA's Sergey Bubka sets a record of 6.40 m.  
2012: USA's Sergey Bubka sets a record of 6.40 m.

[http://junkcharts.typepad.com/junk\\_charts/music/](http://junkcharts.typepad.com/junk_charts/music/)

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

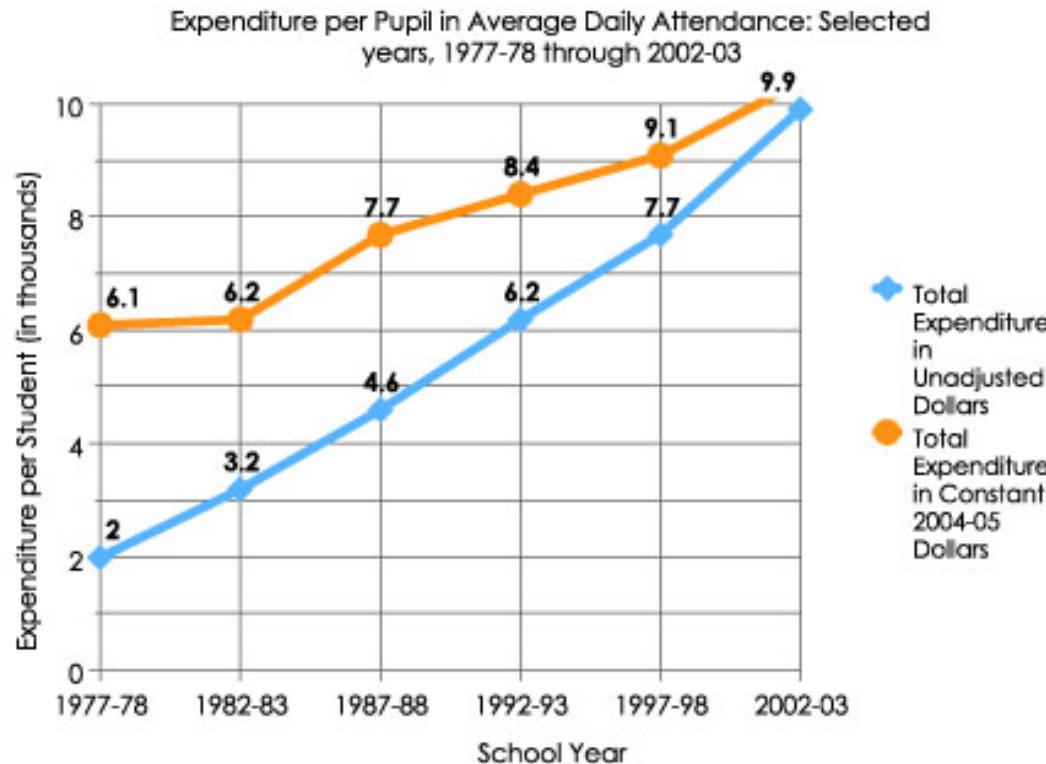


Source: Stockholm International Peace Research Institute

Graphic by Han Chang-duk

# Line Graphs (+ Variants)

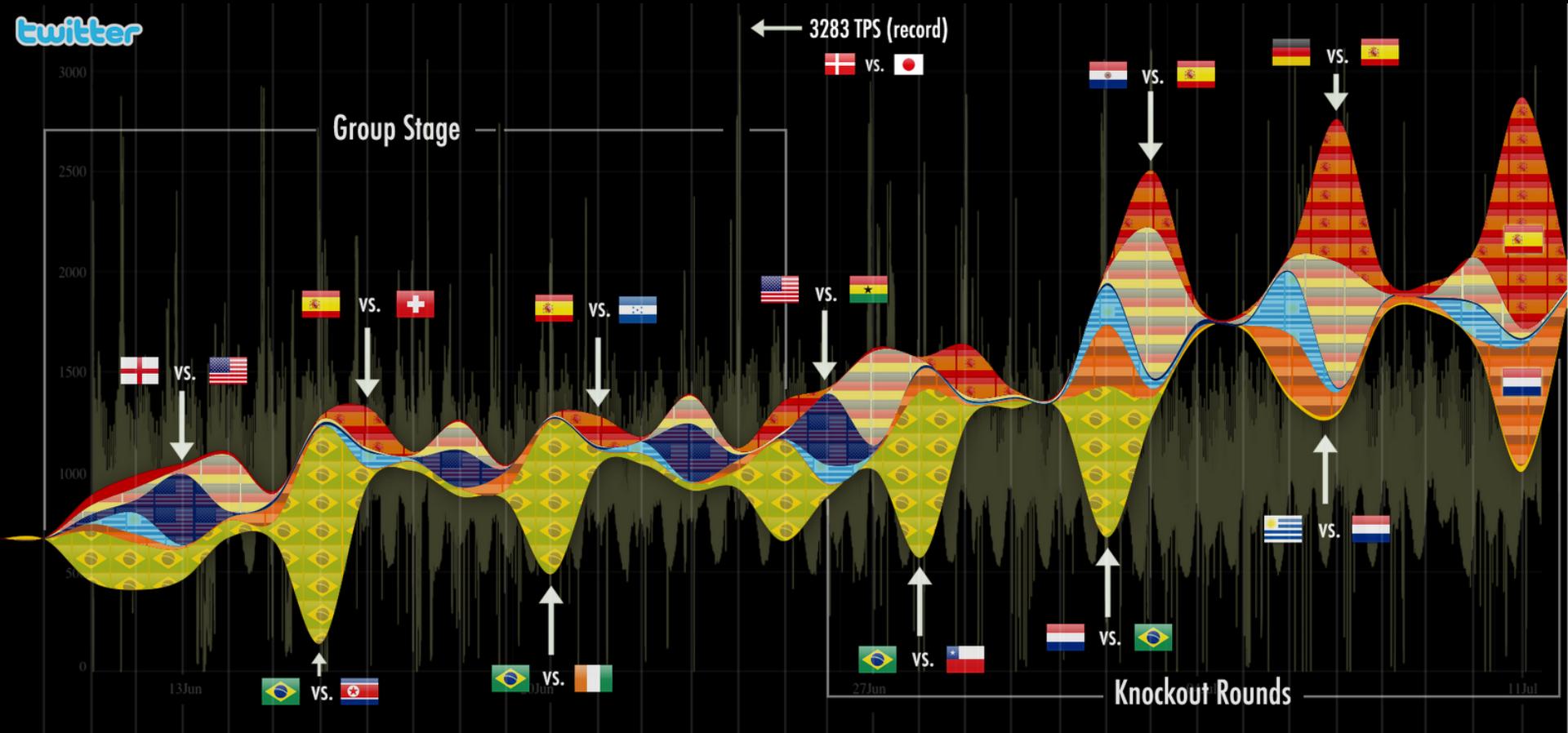
- Suitable for what types of data?
- What makes a good line graph?



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”



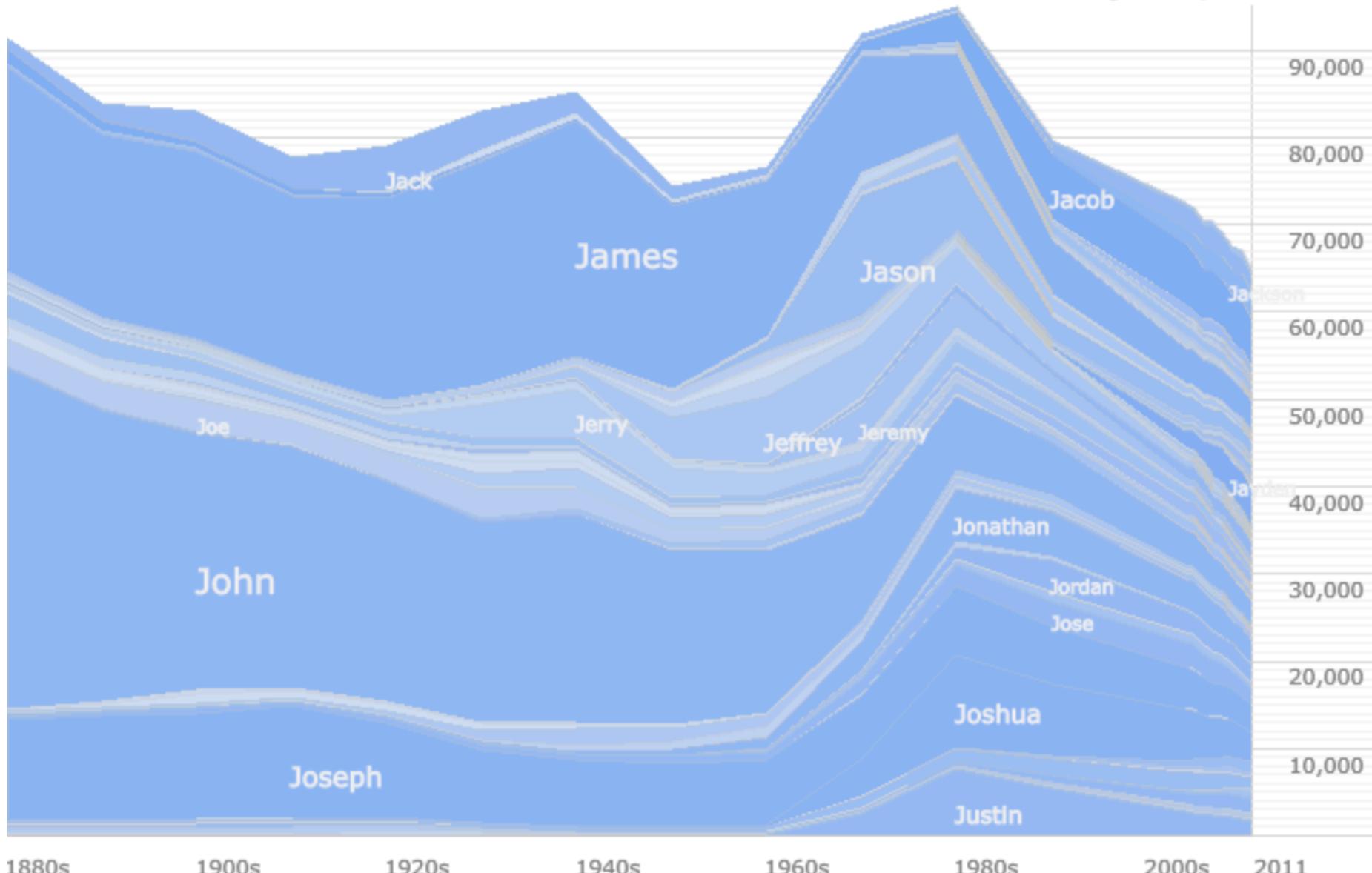
Baby Name &gt; j



Both Boys Girls

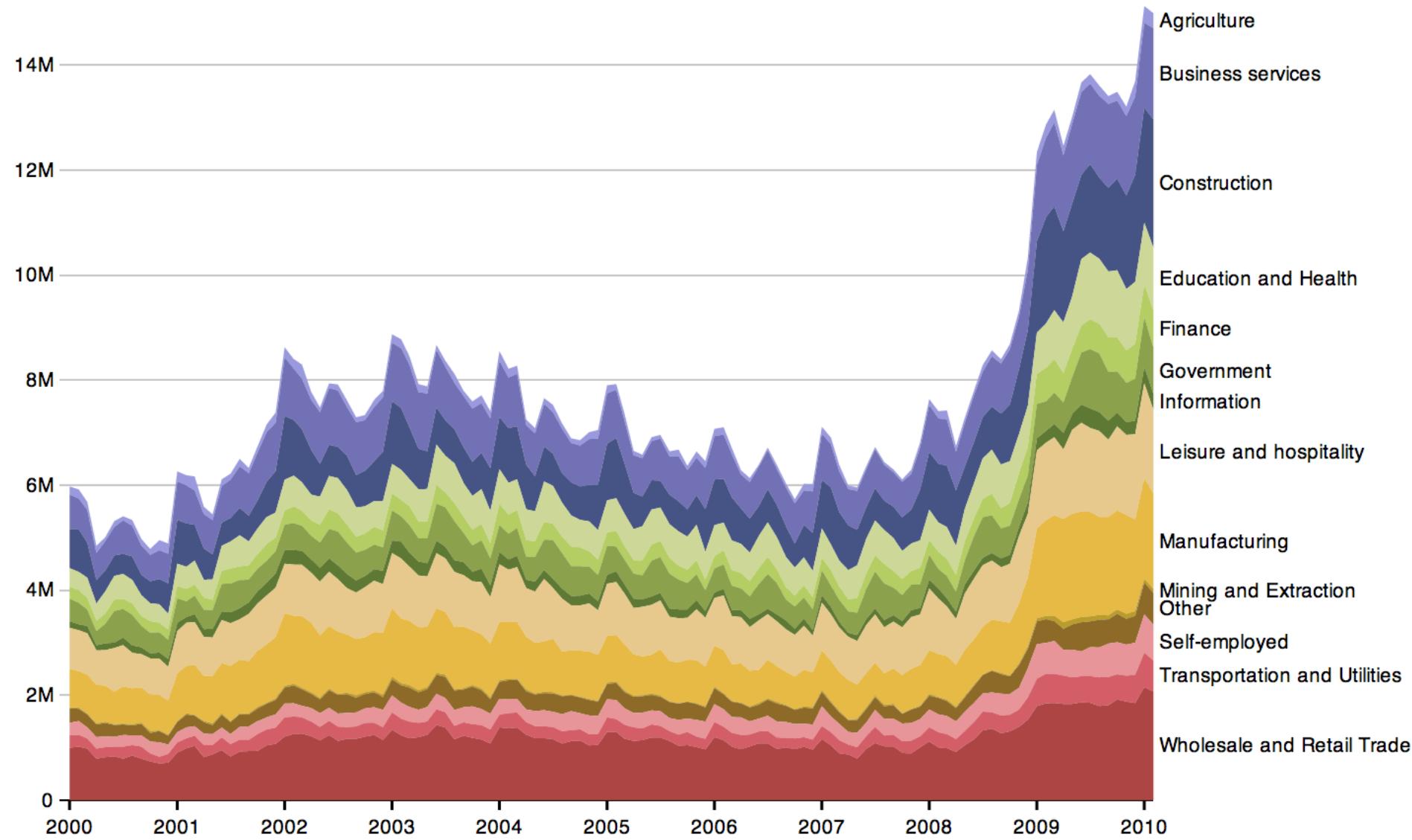
2010 rank: boys 1000 500 100 25 1  
girls 1000 500 100 25 1

Names starting with 'J' per million babies



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

View: Counts (Zero-Baseline) ▾

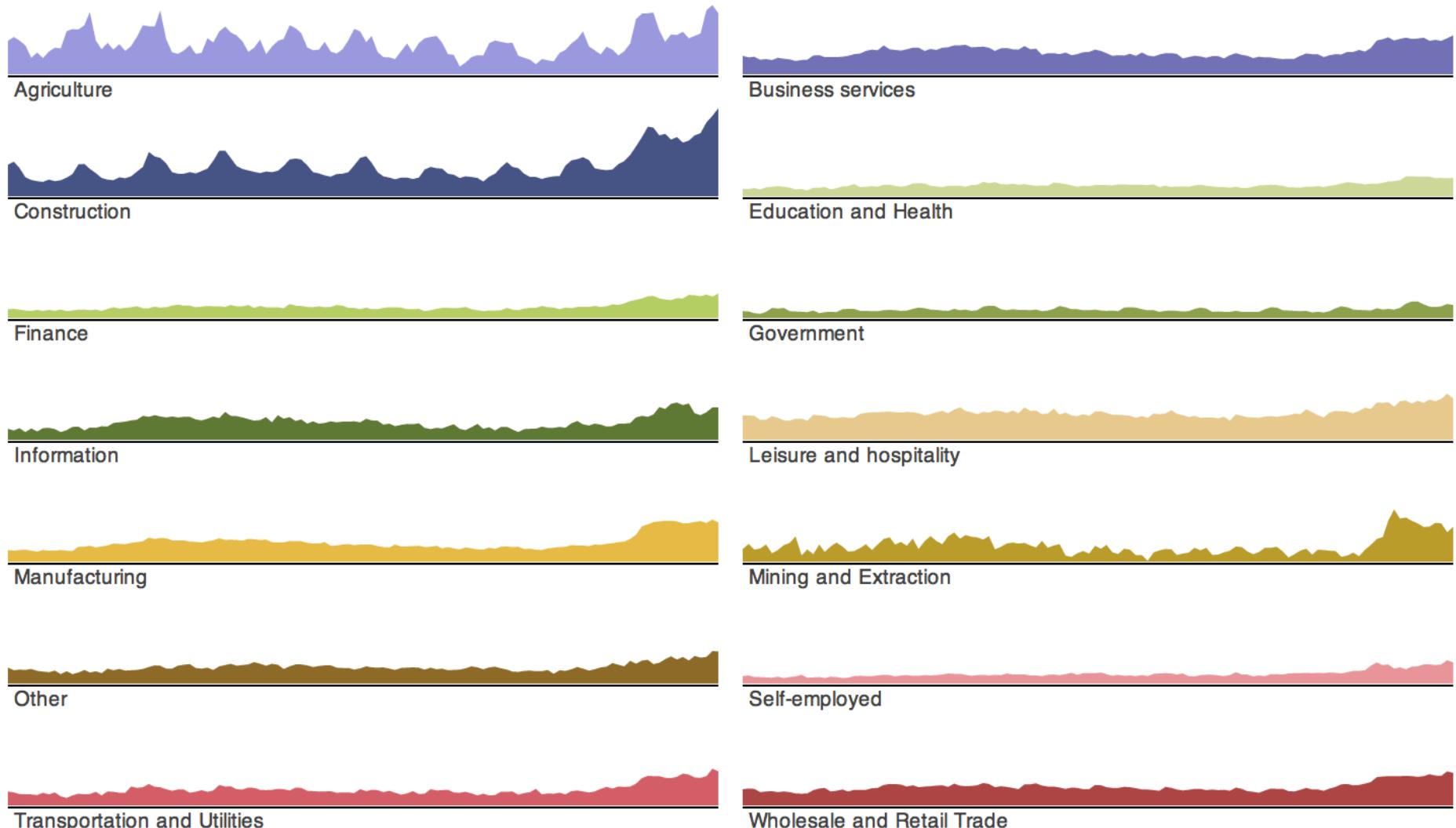


Total counts of unemployed persons per industry, 2000-2010.

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.

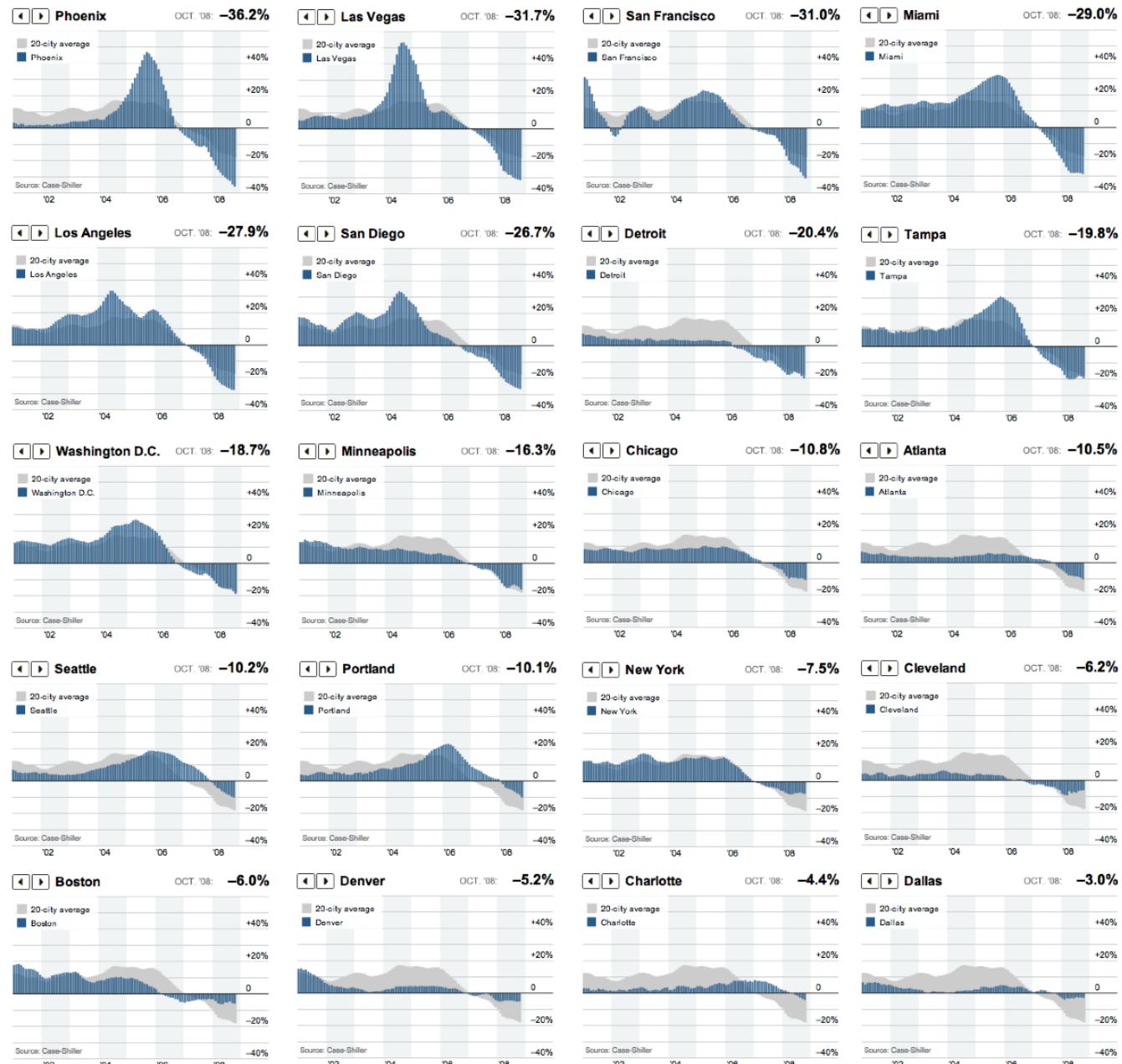
Source: [U.S. Bureau of Labor Statistics](#)

# Small Multiples

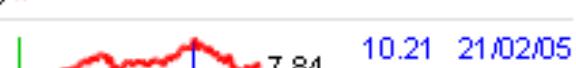
- Multiple small graphs arranged in a grid
- Advantages and disadvantages?

# Change in Home Prices (year over year)

From New York Times Economix blog

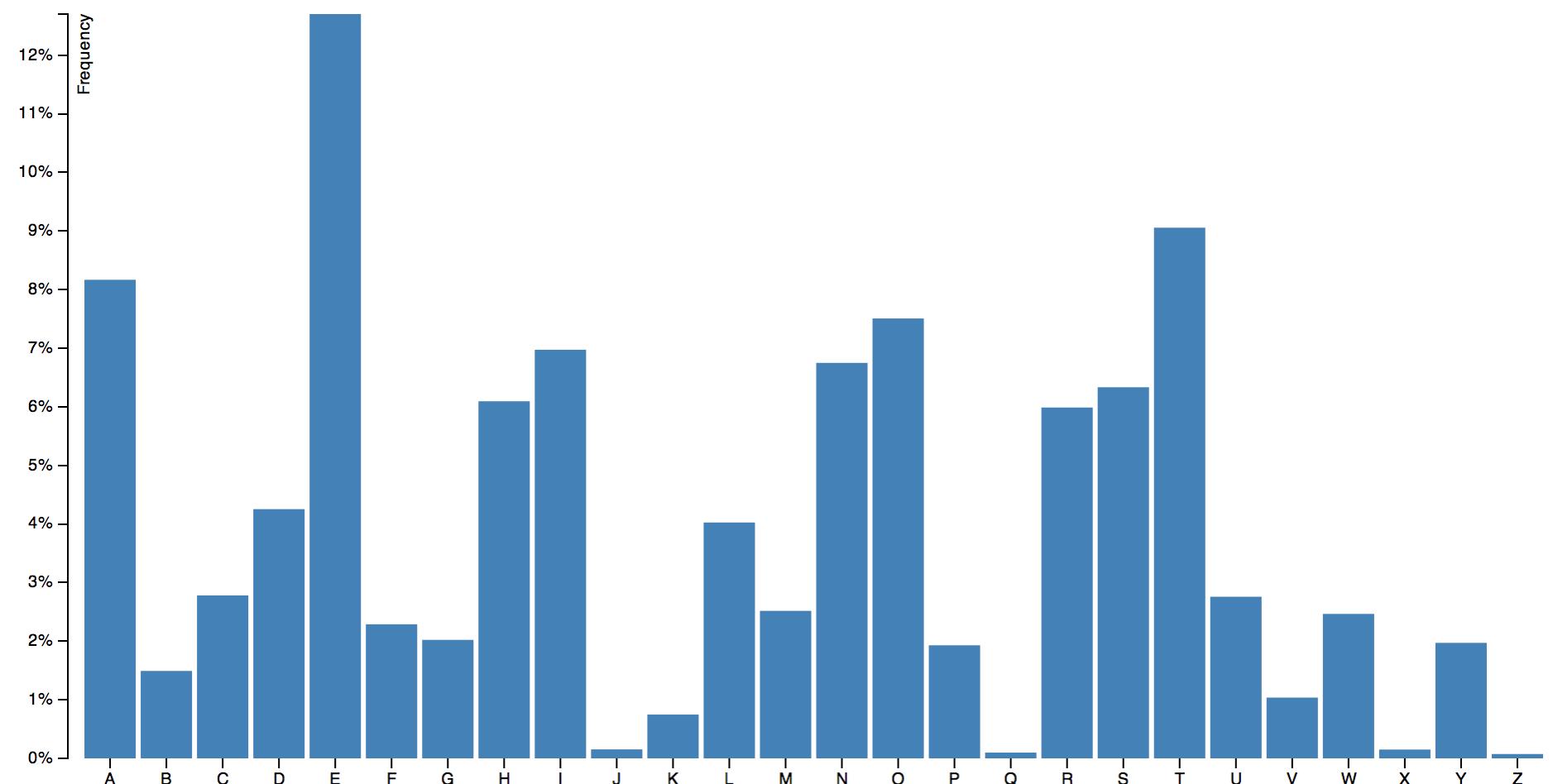


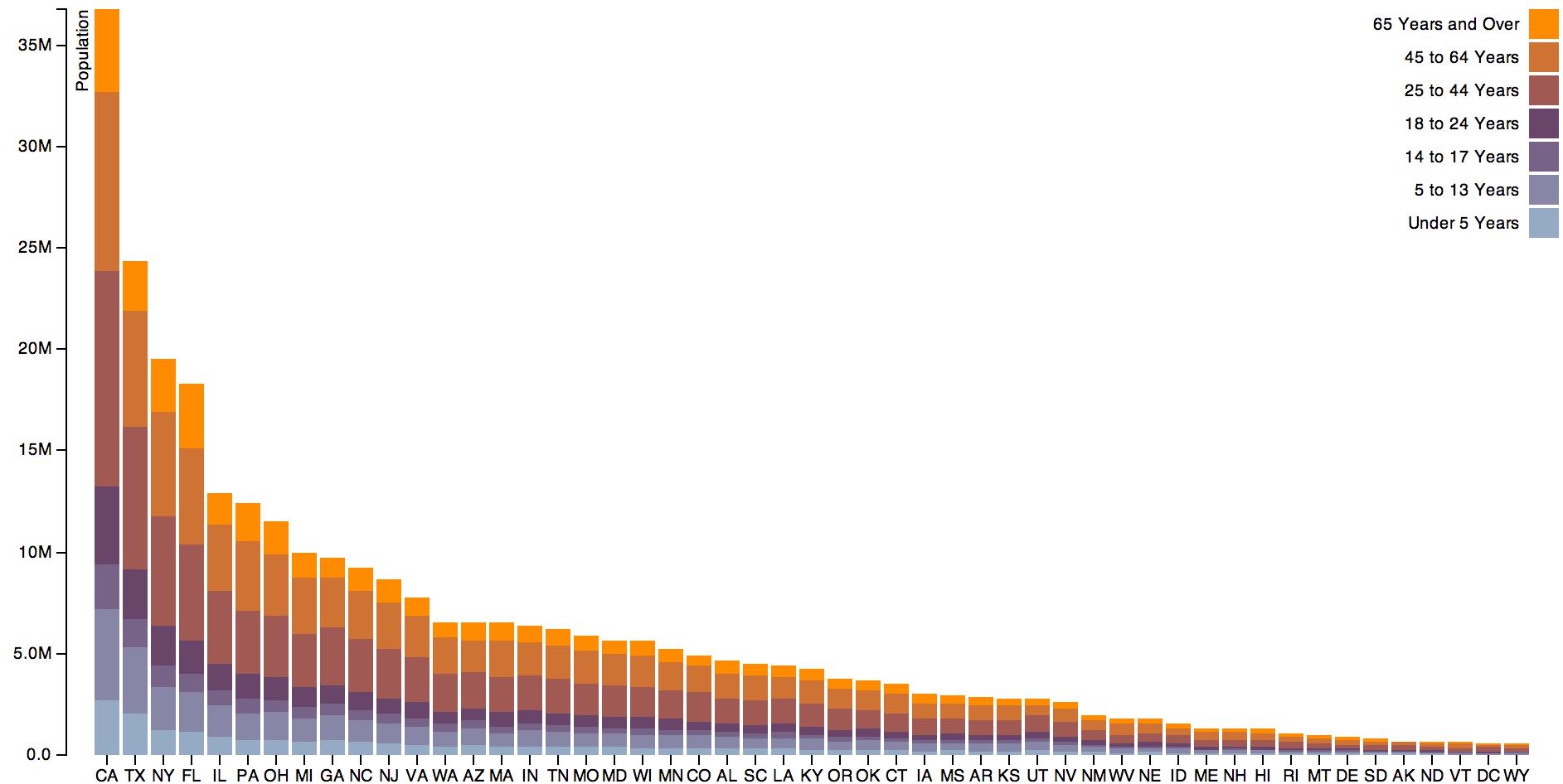
# Spark Lines

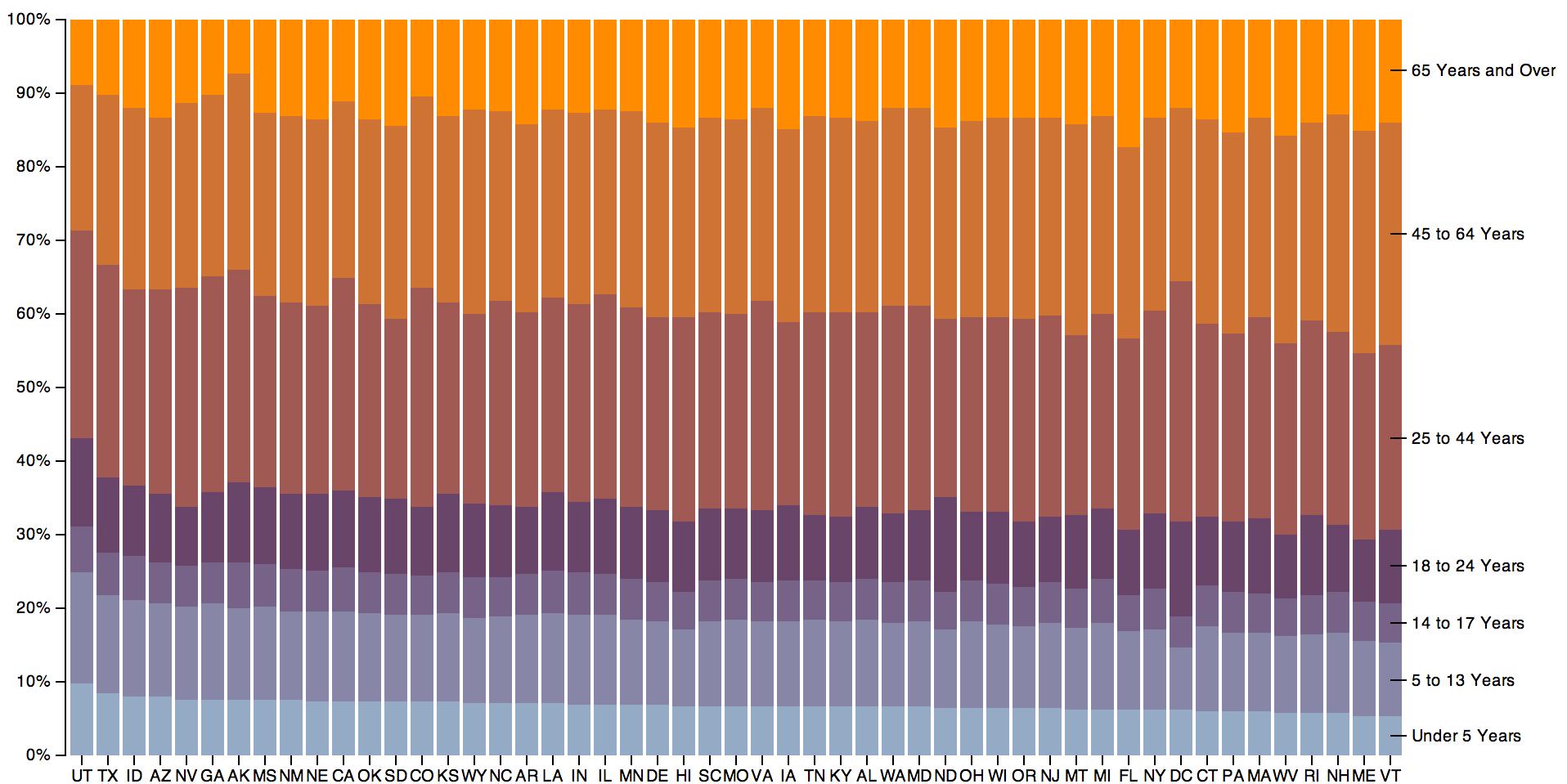
AGL	11.71		13.65	14.51 11.64	03/02/05 03/06/04
ALN	5.92		9.65	9.83 5.92	24/05/05 02/06/04
ARG	4.77		5.54	5.77 4.76	14/02/05 08/06/04
BHP	12.19		16.05	19.29 11.92	07/03/05 03/06/04
BSL	6.07		7.84	10.21 5.92	21/02/05 03/06/04
IAG	4.60		5.88	6.95 4.60	07/02/05 02/06/04
ILU	4.28		6.00	6.56 4.23	08/02/05 04/06/04
MBL	33.51		50.07	51.35 31.66	20/05/05 10/08/04
MGQ			4.13	4.25 3.76	07/02/05 07/04/05
NAB	30.28		31.80	31.83 26.08	24/05/05 12/08/04
NFD	4.29		0.00	6.50 4.25	07/04/05 15/07/04
ORI	14.89		15.10	20.41 14.65	17/12/04 17/05/05
PBL	12.54		15.46	17.50 12.54	29/12/04 02/06/04
RIN	7.60		12.20	12.50 7.55	24/05/05 03/06/04
RIO	35.50		42.35	47.93 34.85	14/03/05 03/06/04
TLS	4.67		4.92	5.49 4.61	22/03/05 10/06/04
WBC	17.39		19.92	20.13 16.14	10/02/05 16/08/04
WDC	15.30		16.75	17.39 14.57	17/01/05 26/10/04
WES	27.95		36.64	42.37 27.51	21/03/05 03/06/04
WOW	11.70		16.02	16.69 11.38	21/03/05 16/07/04

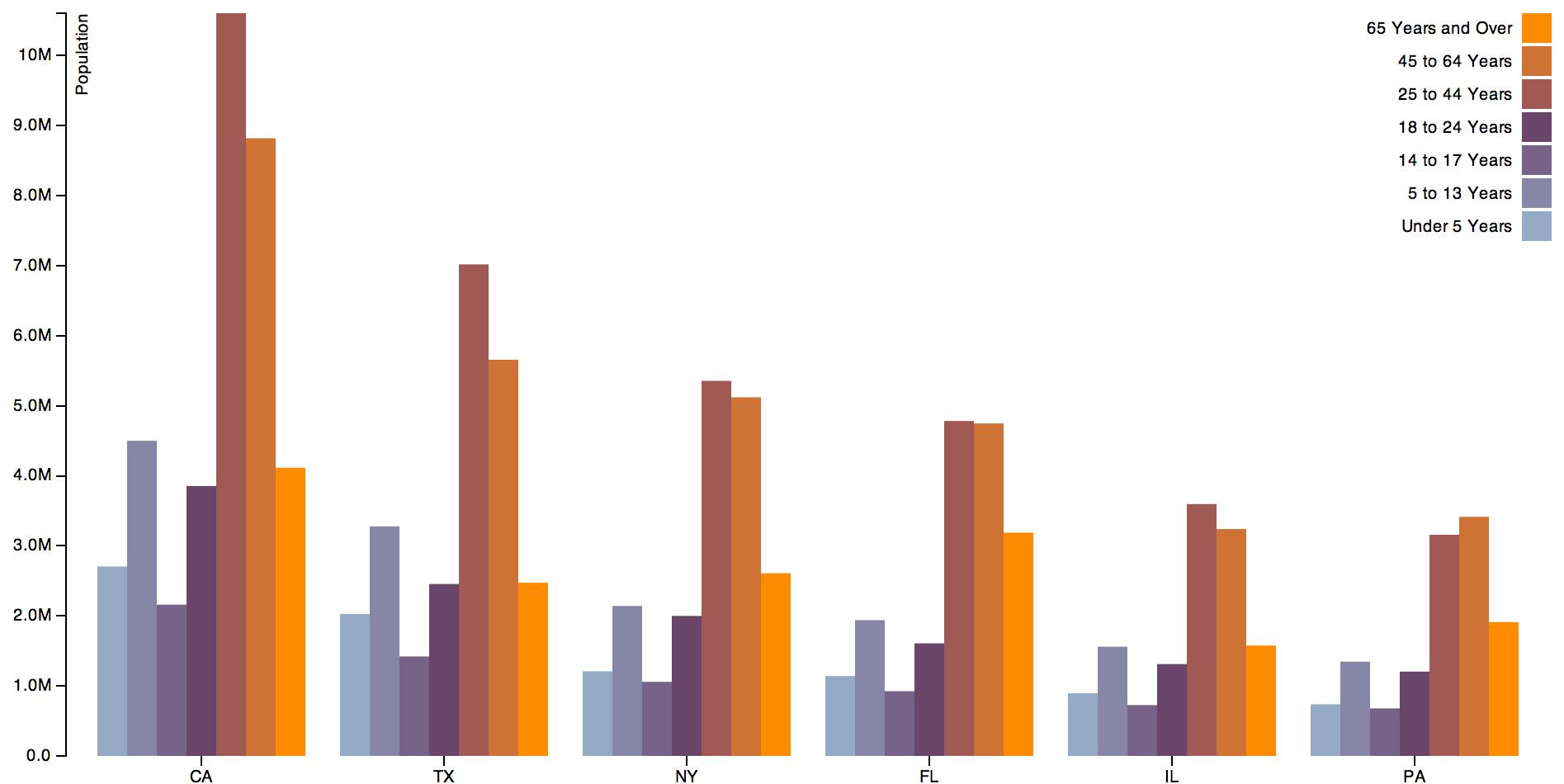
# Bar Graphs (+ Variants)

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



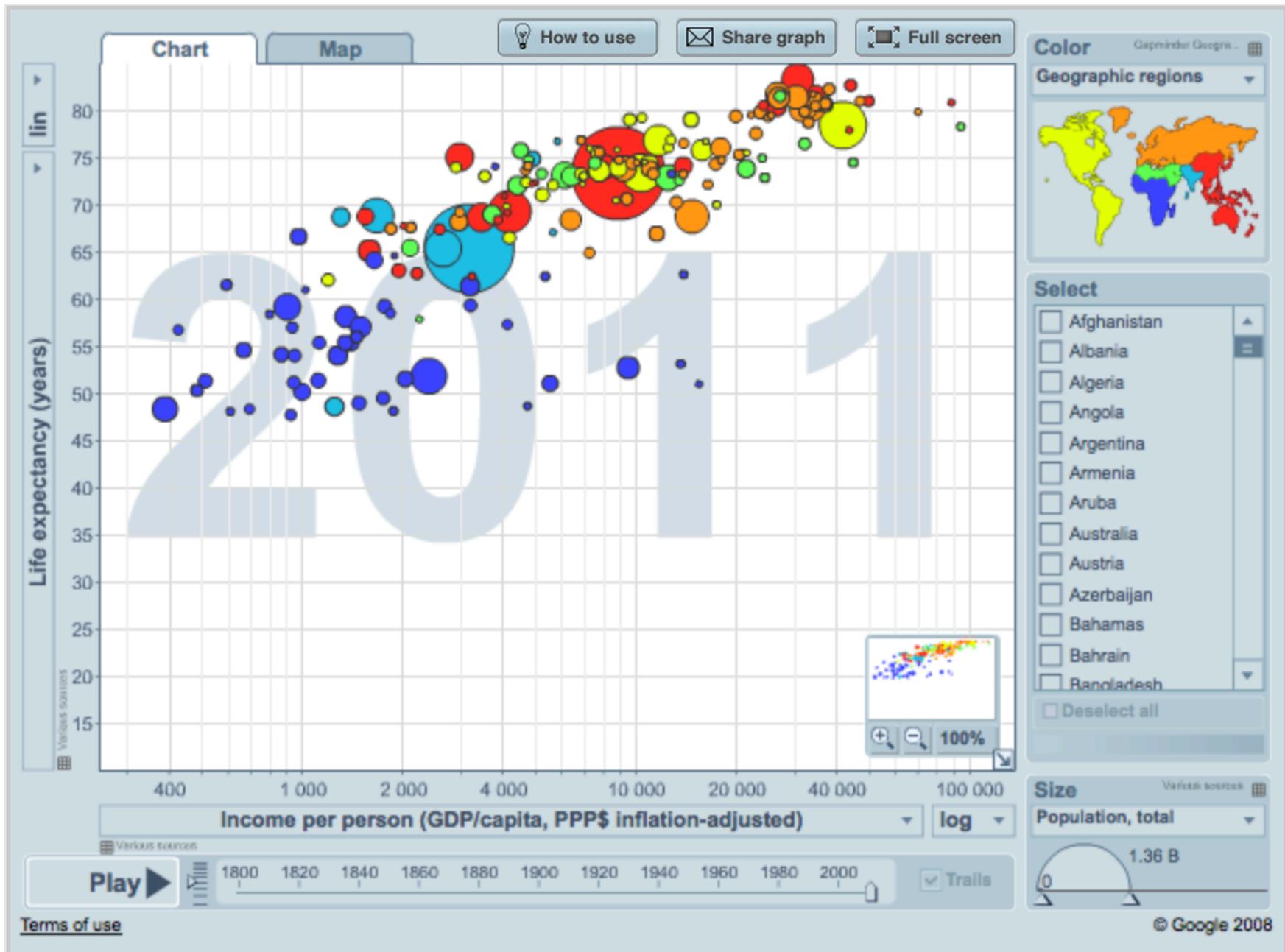






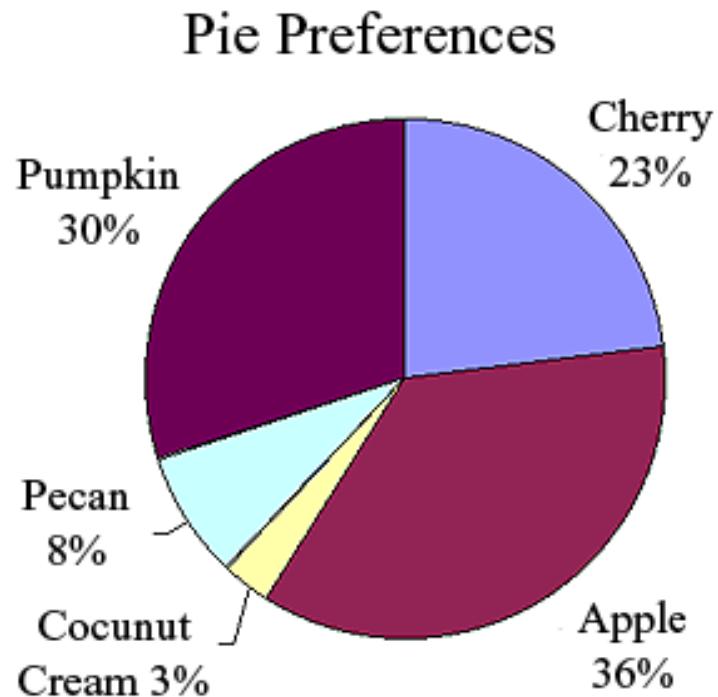
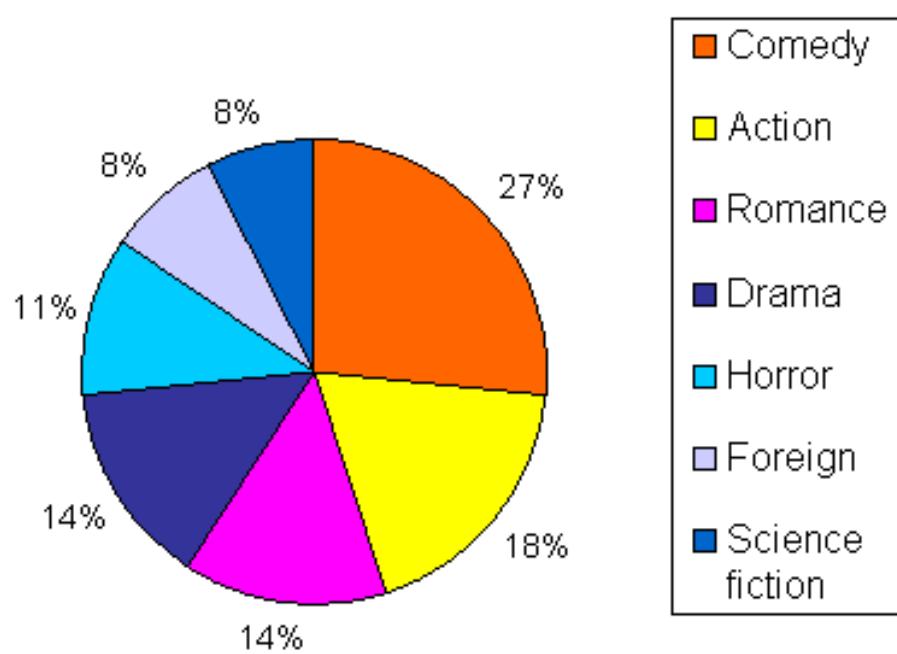
# Scatterplots

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

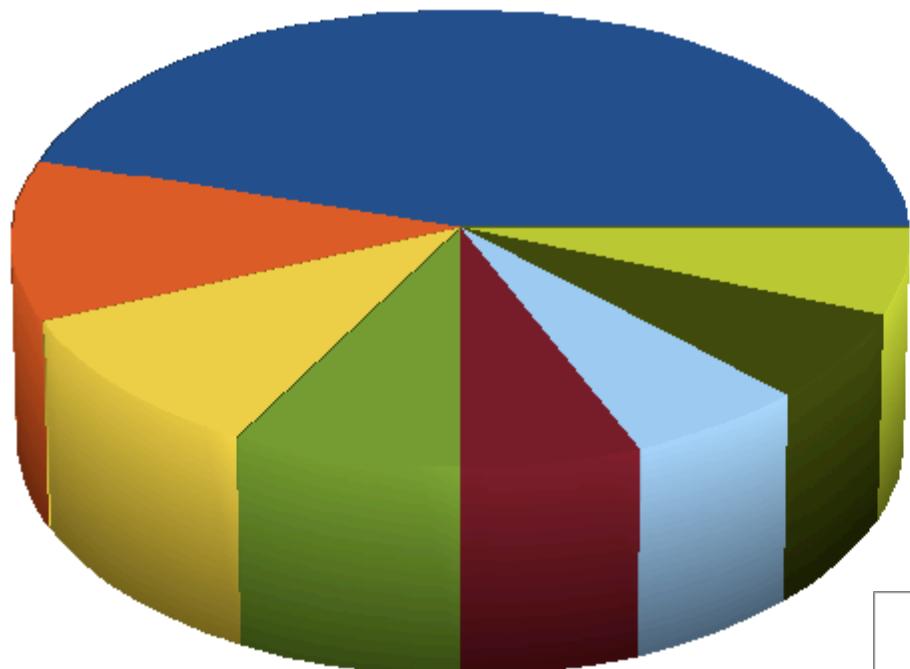


# Pie Charts (+ Variants)

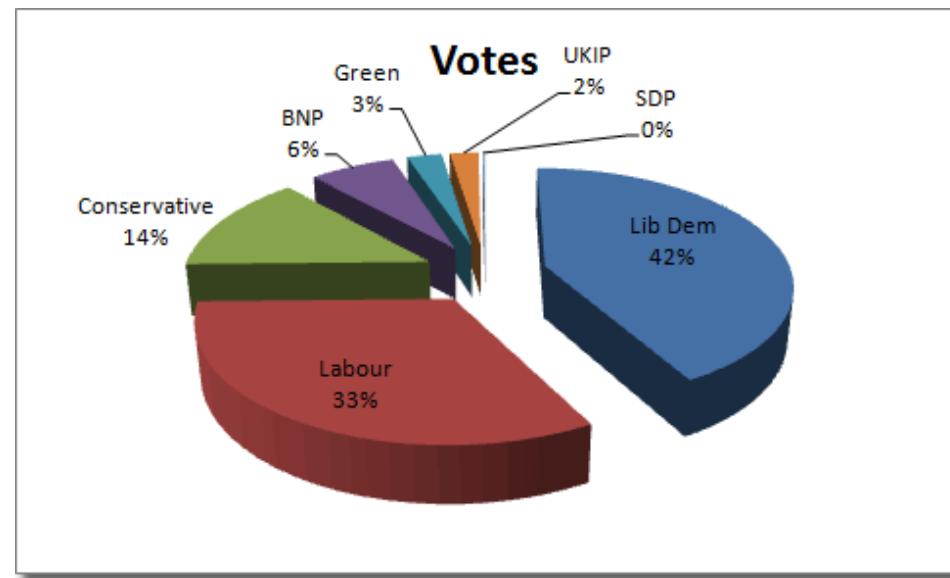
- Suitable for what types of data?
- Different ways for labeling



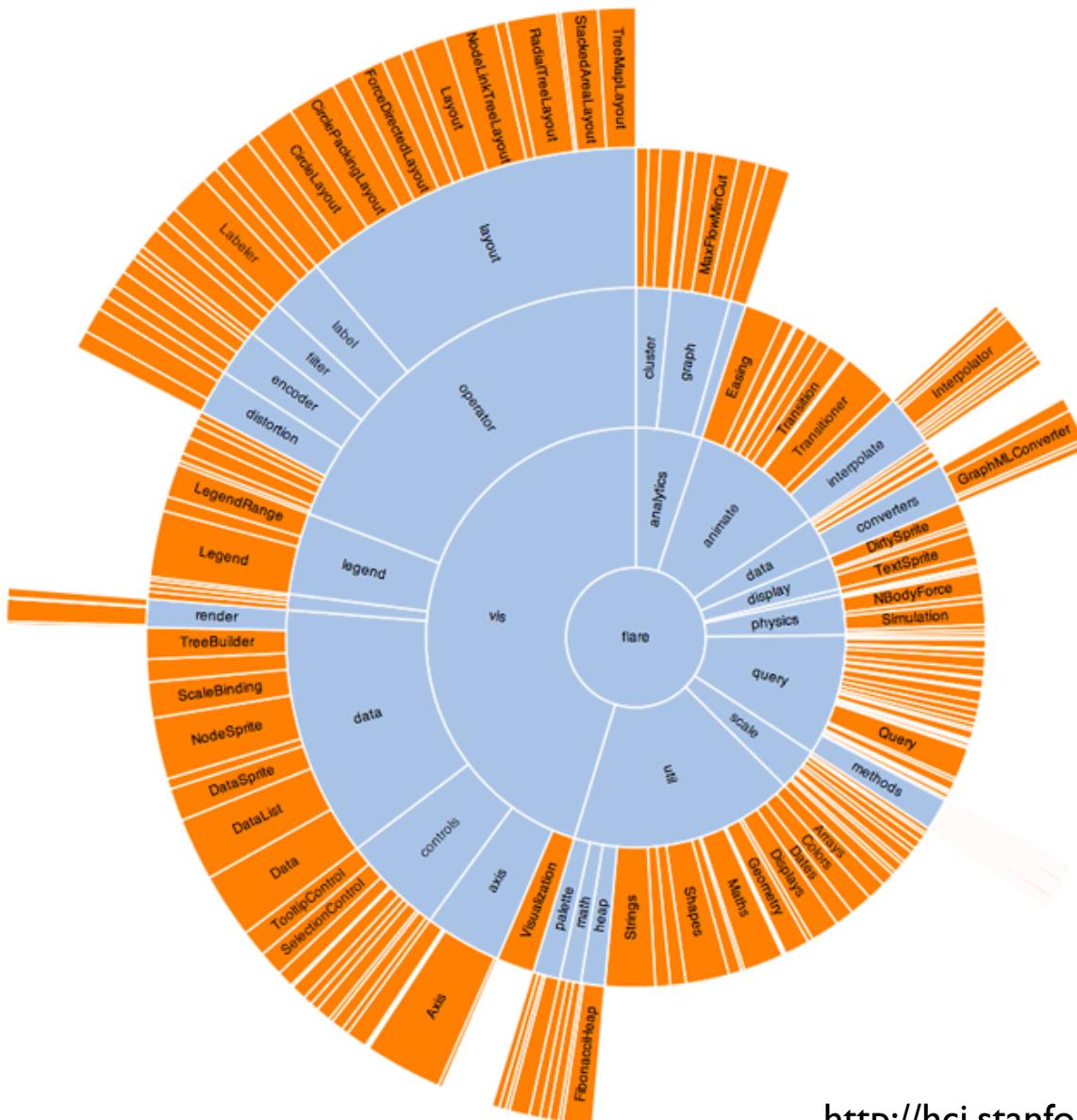
# More Chart Junk



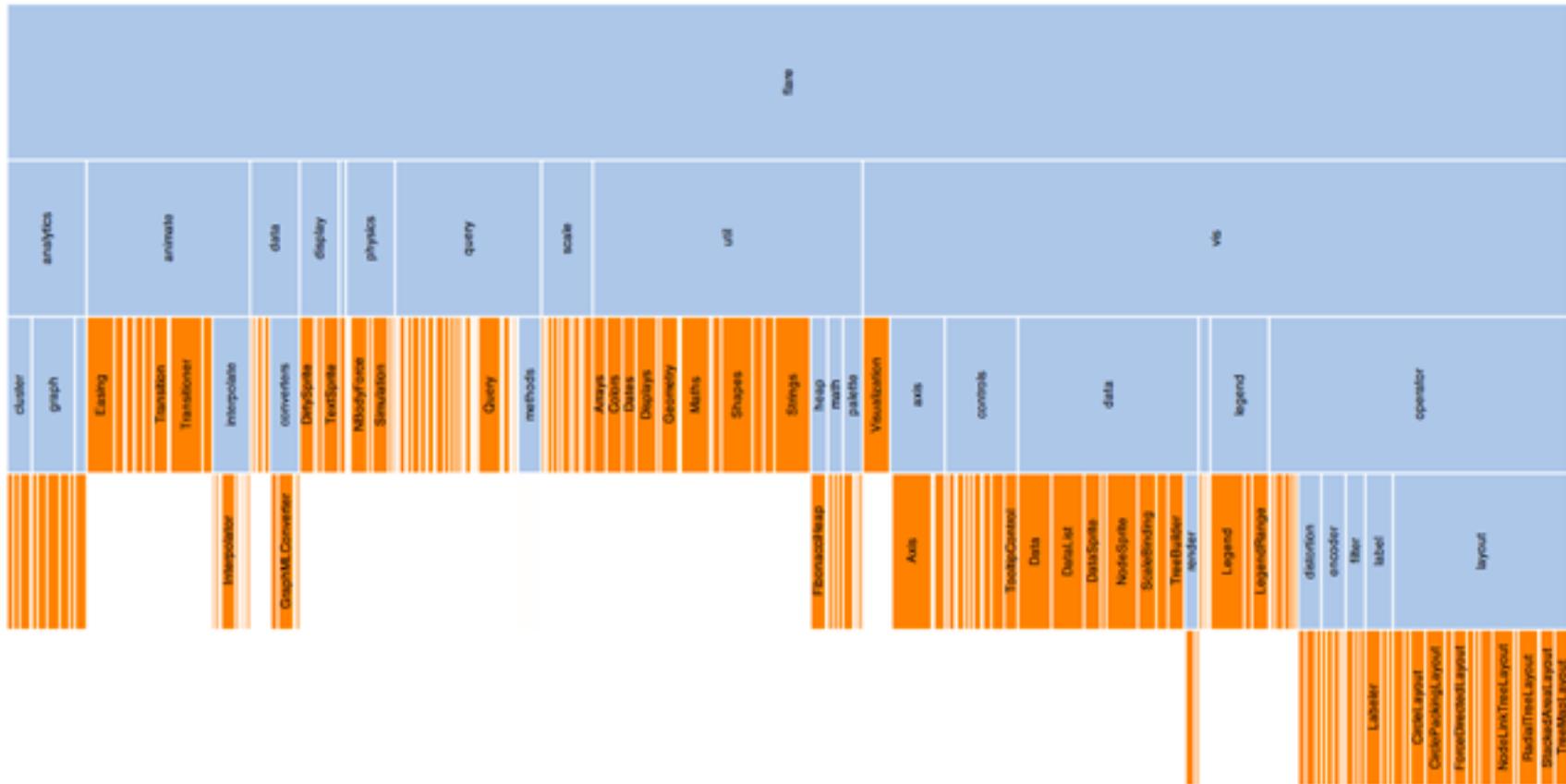
- MySpace
- Facebook
- Orkut
- hi5
- Xanga
- Live Spaces
- Classmates.com
- Bebo



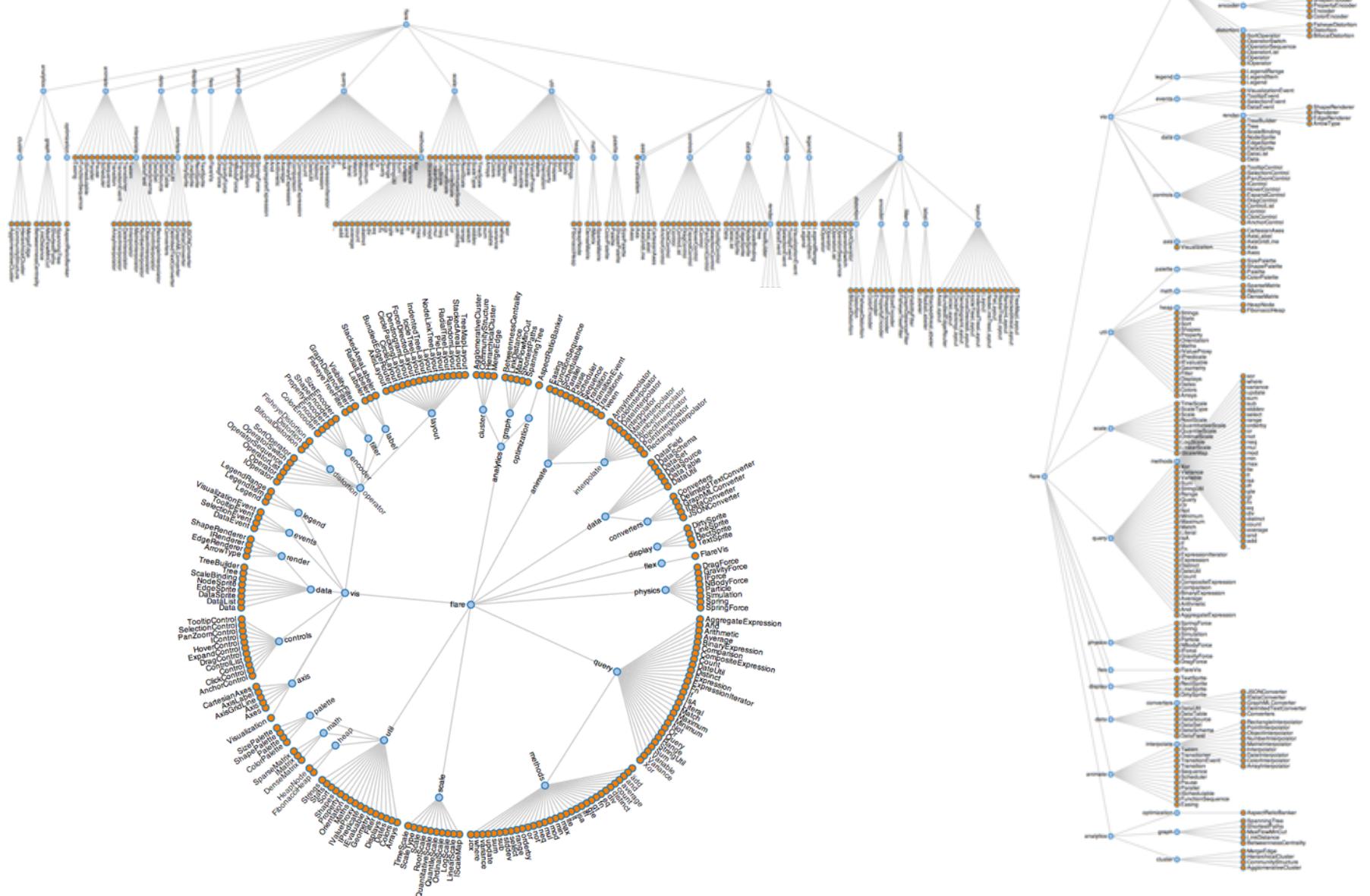
# Sunburst Trees



# 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

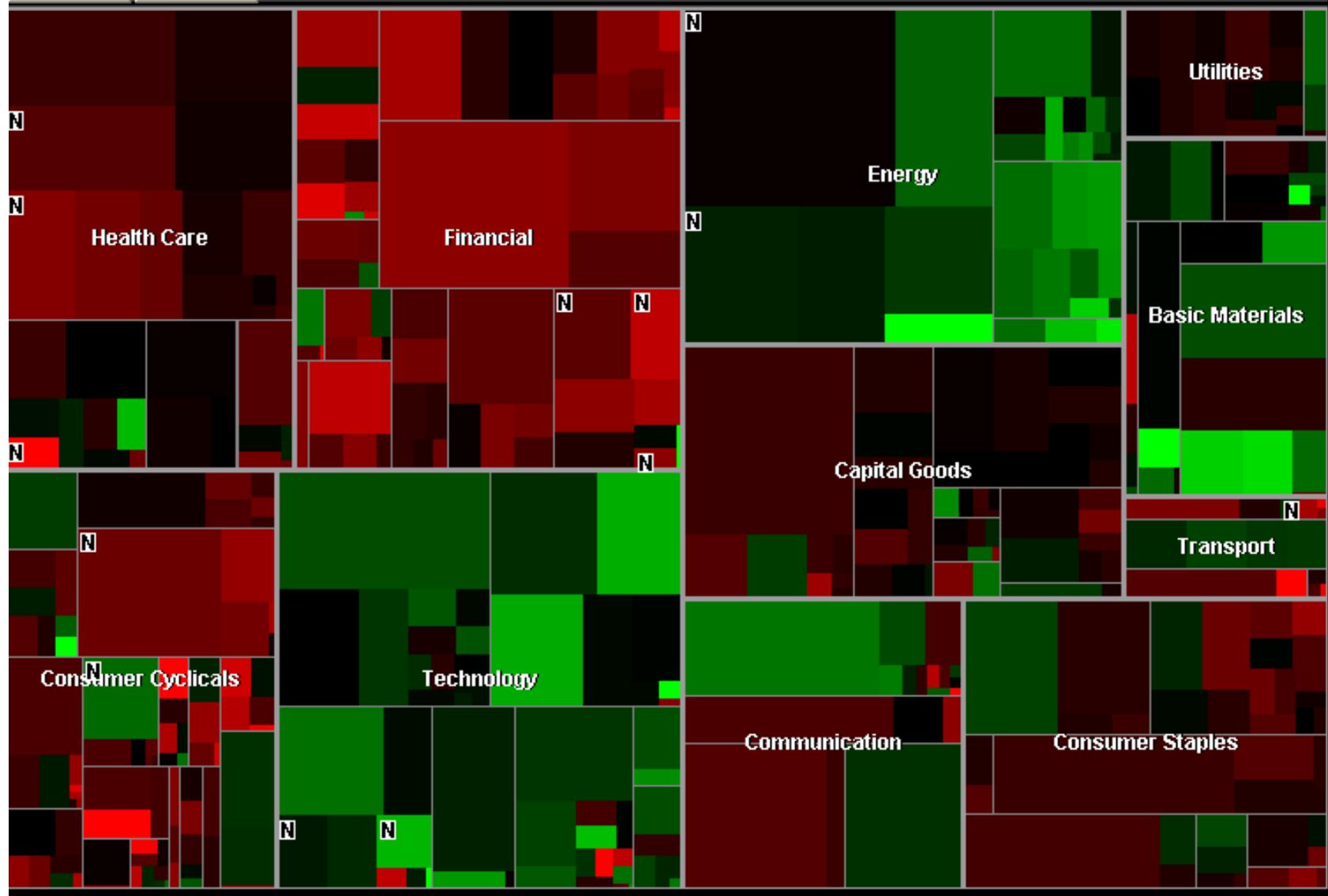
Controls Instructions

Headline Icons ▾

DJIA 12182.13 -64.87 -0.53%

Nasdaq 2304.85 +11.82 +0.52%

5:42 pm Feb. 10



Controls

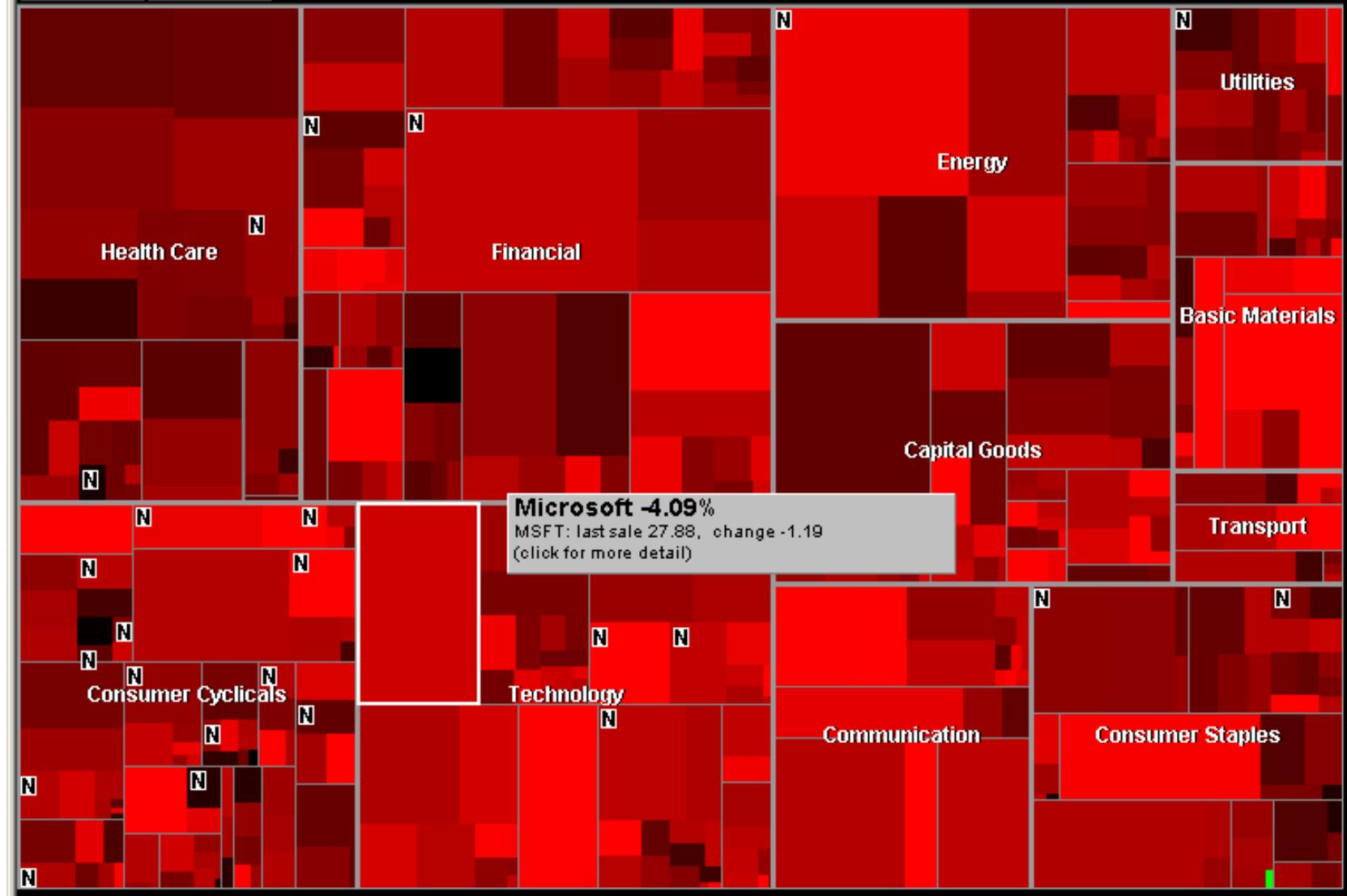
Instructions

Headline Icons

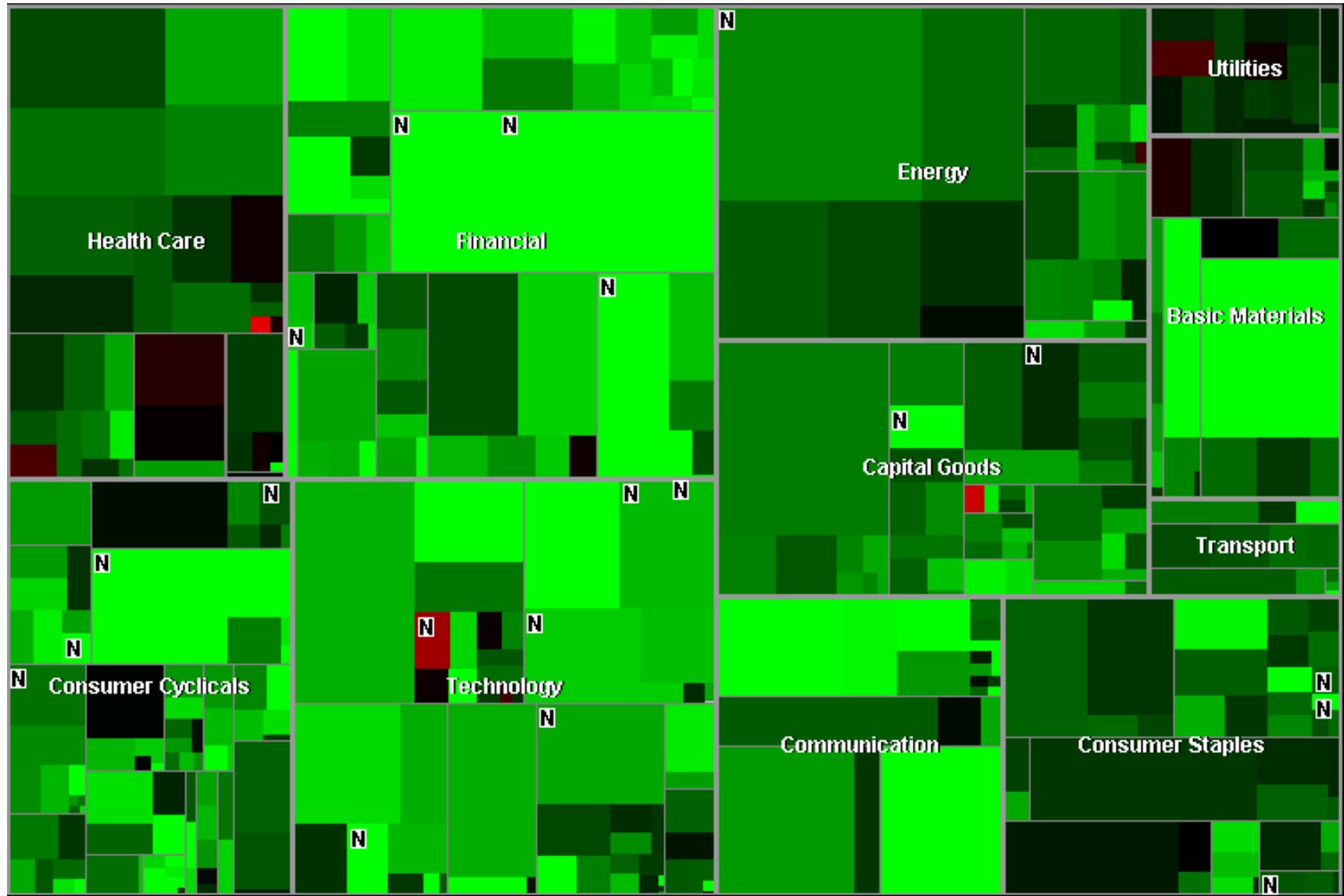
DJIA 12216.24 -416.02 -3.29%

Nasdaq 2407.87 -96.65 -3.86%

4:32 pm Feb. 27



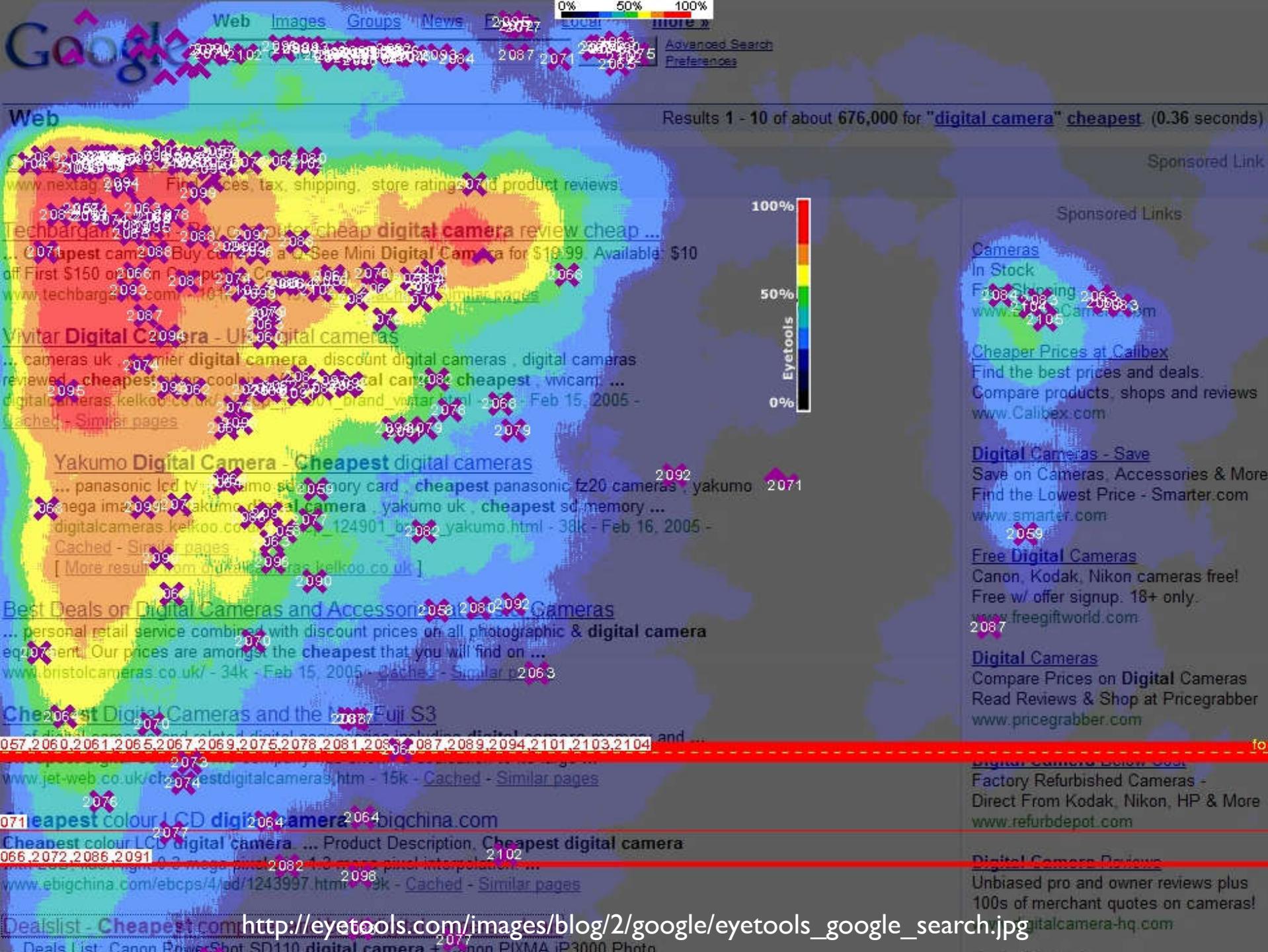
Market mixed, February 8, 2008:  
Energy & Technology up Financial & Health Care down



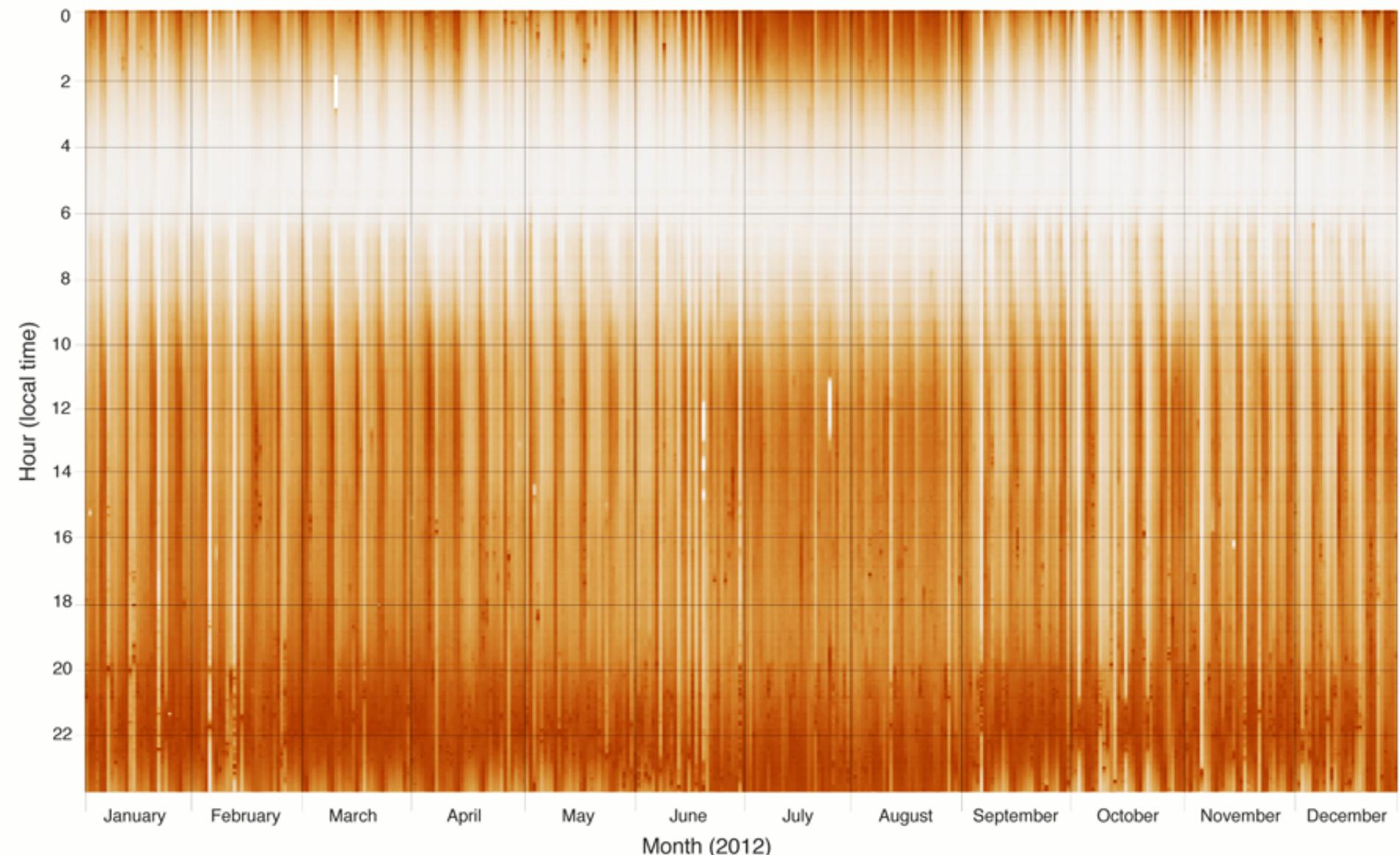
Market rises 319 points, November 13, 2007:  
Five exceptions

# Heatmaps

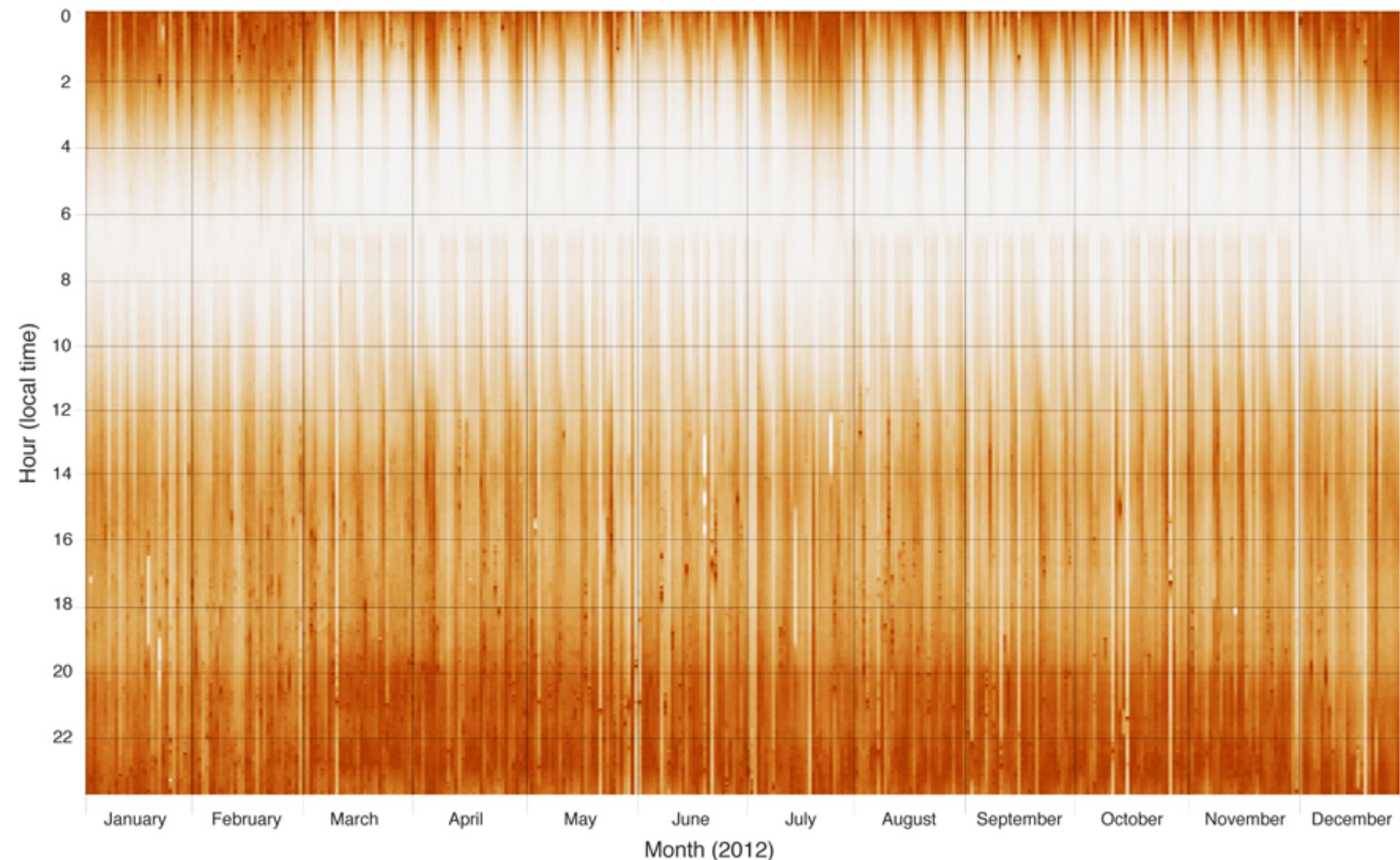
- Data represented in a matrix
- Matrix values are color coded



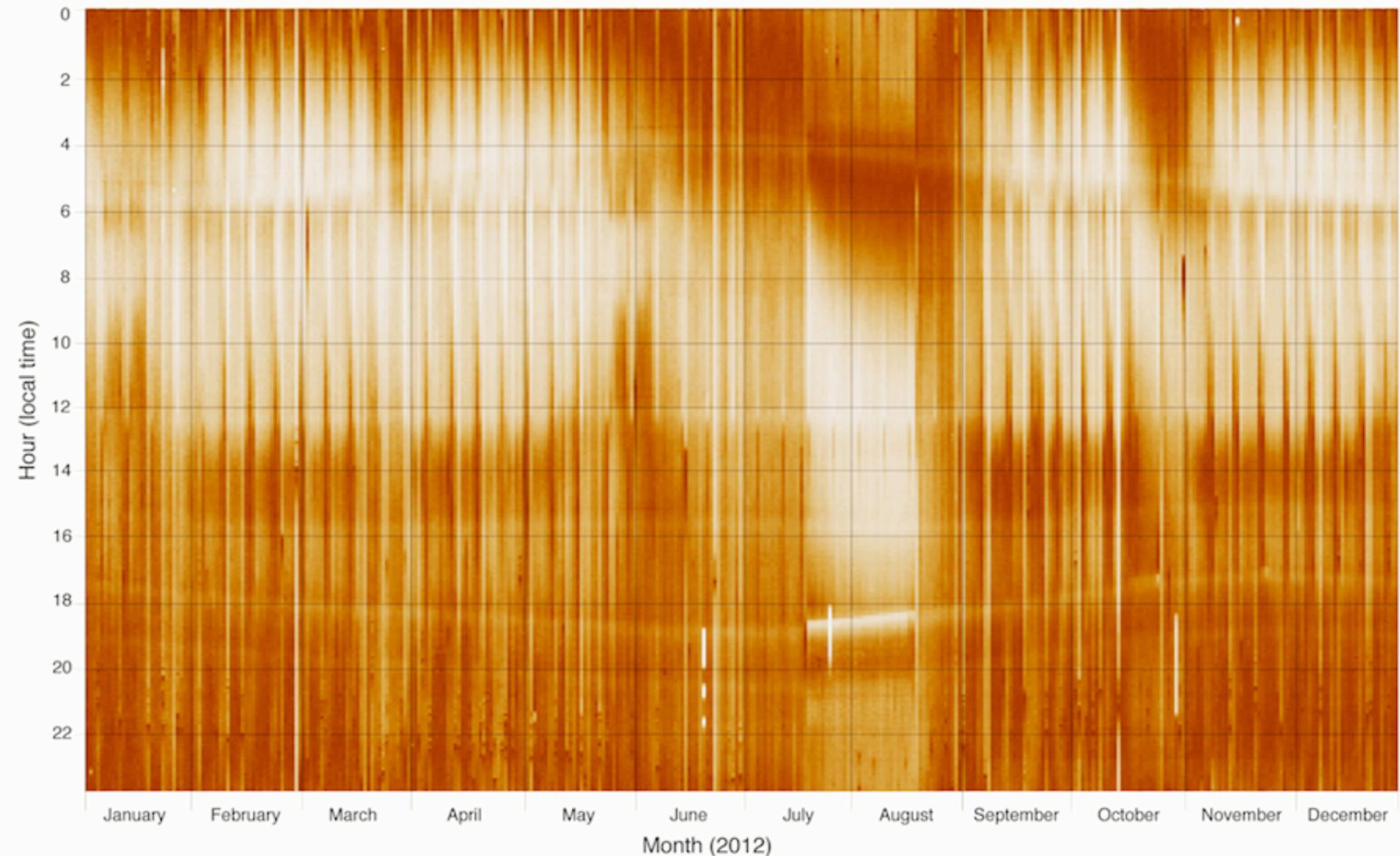
### Tweet Activity (New York City, New York, US)



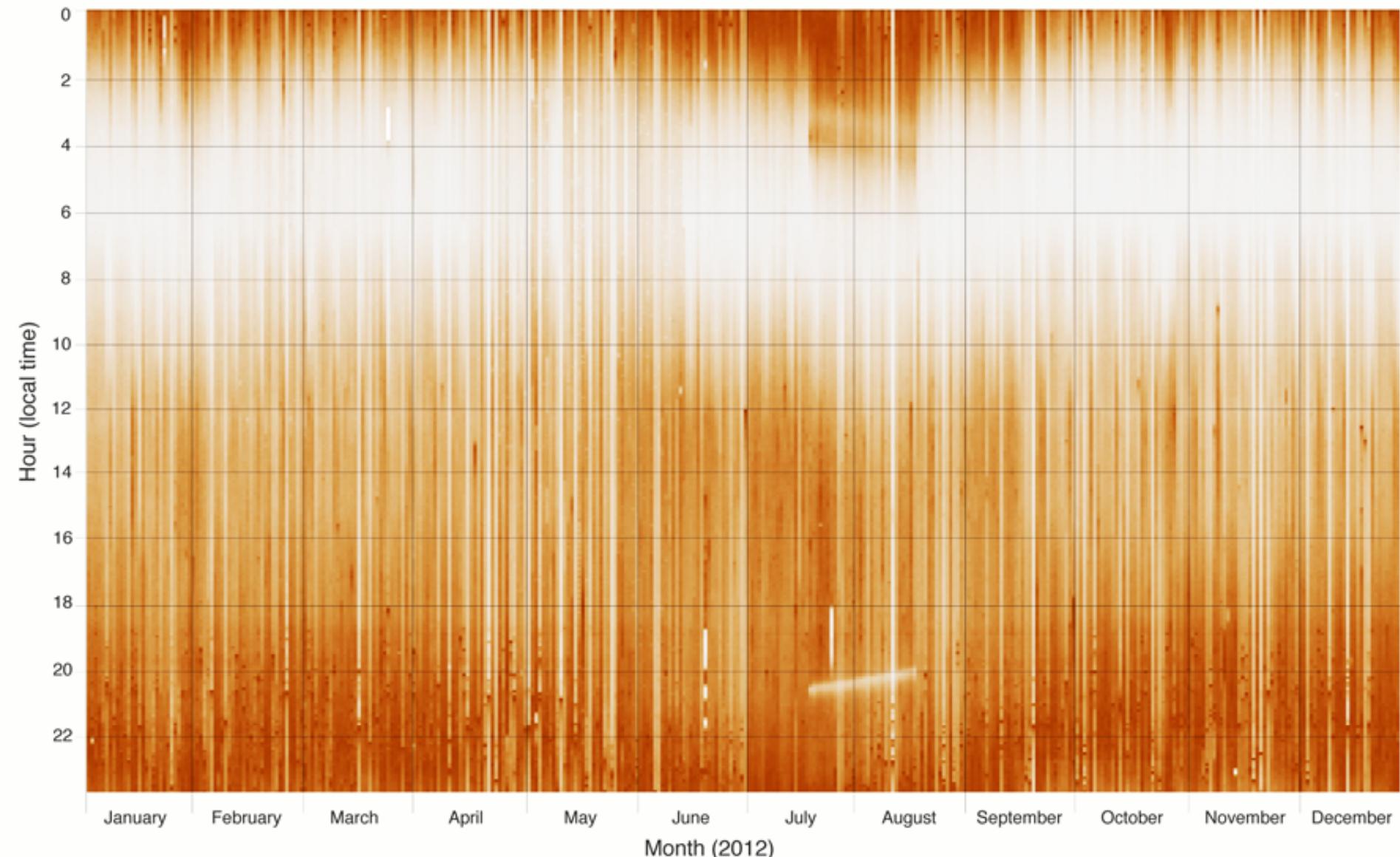
Tweet Activity (Buenos Aires, Argentina)



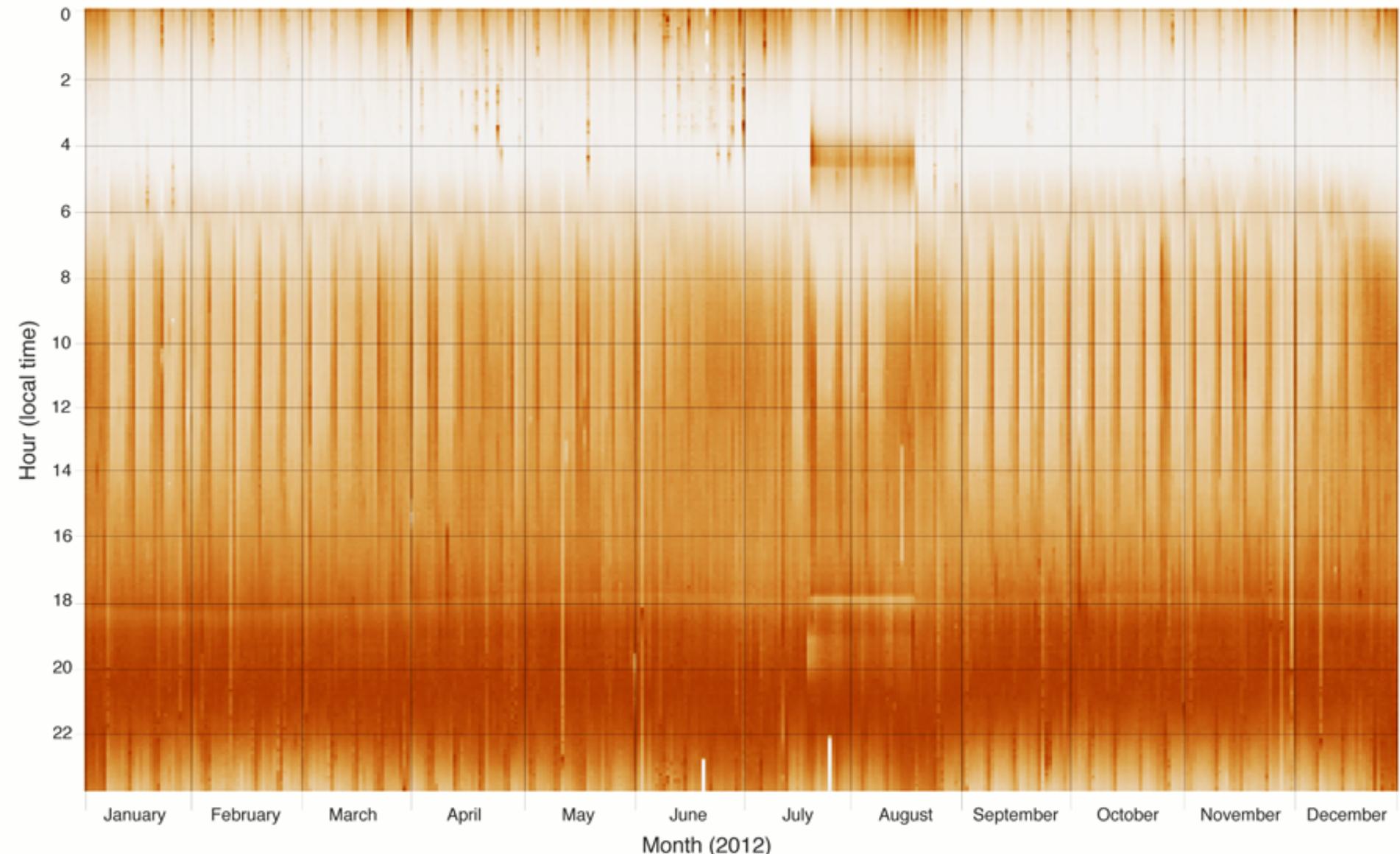
Tweet Activity (Riyadh, Saudi Arabia)



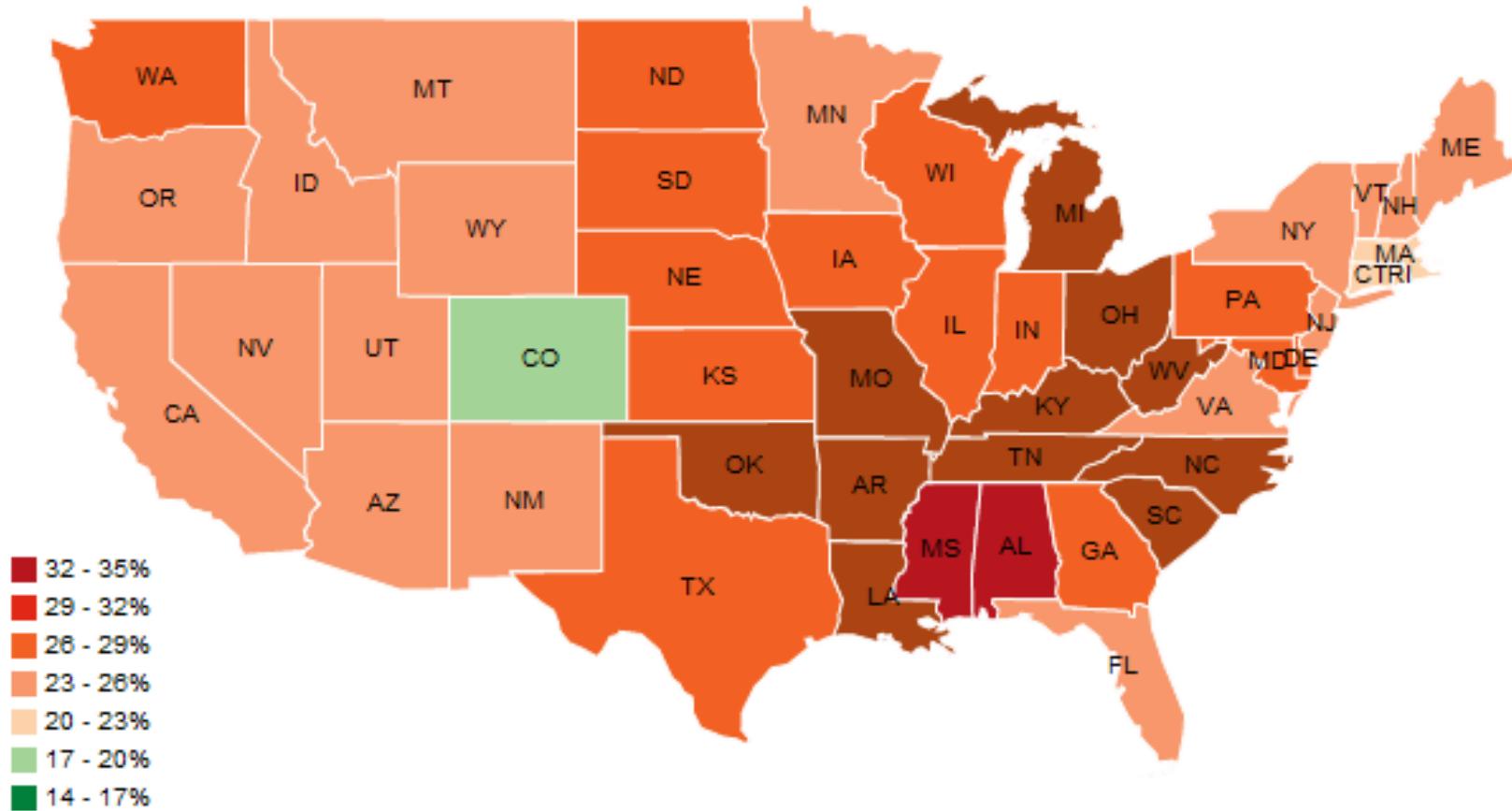
Tweet Activity (Istanbul, Turkey)



Tweet Activity (Jakarta, Indonesia)

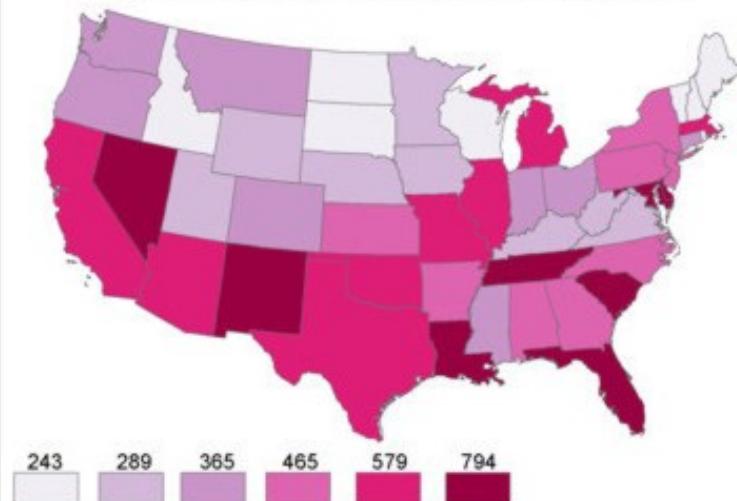


# Choropleth Maps

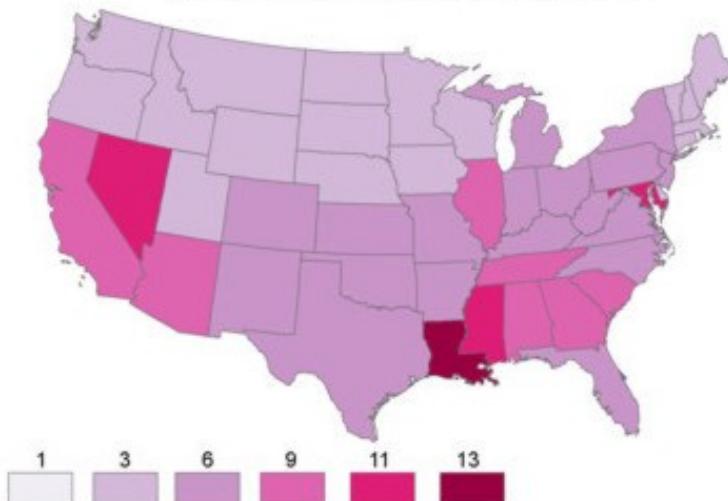


## *Crime Rates in the US - 2003*

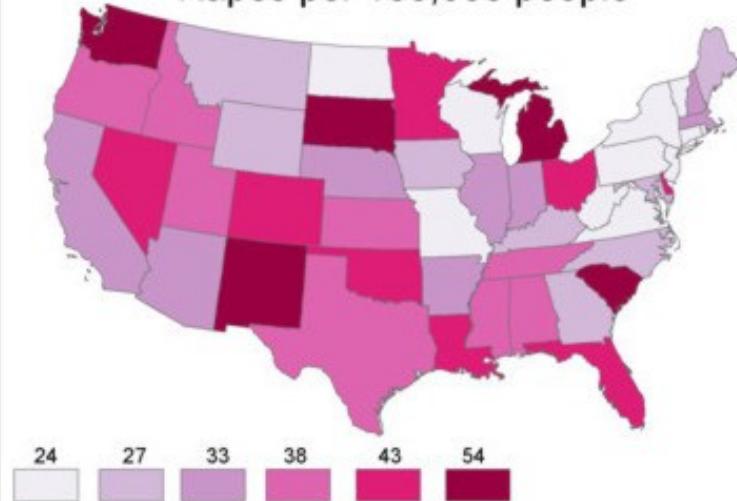
Violent crimes per 100,000 people



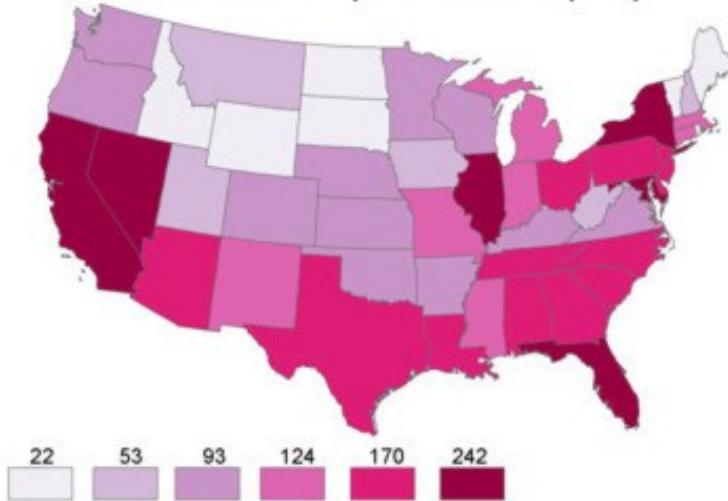
Murders per 100,000 people



Rapes per 100,000 people

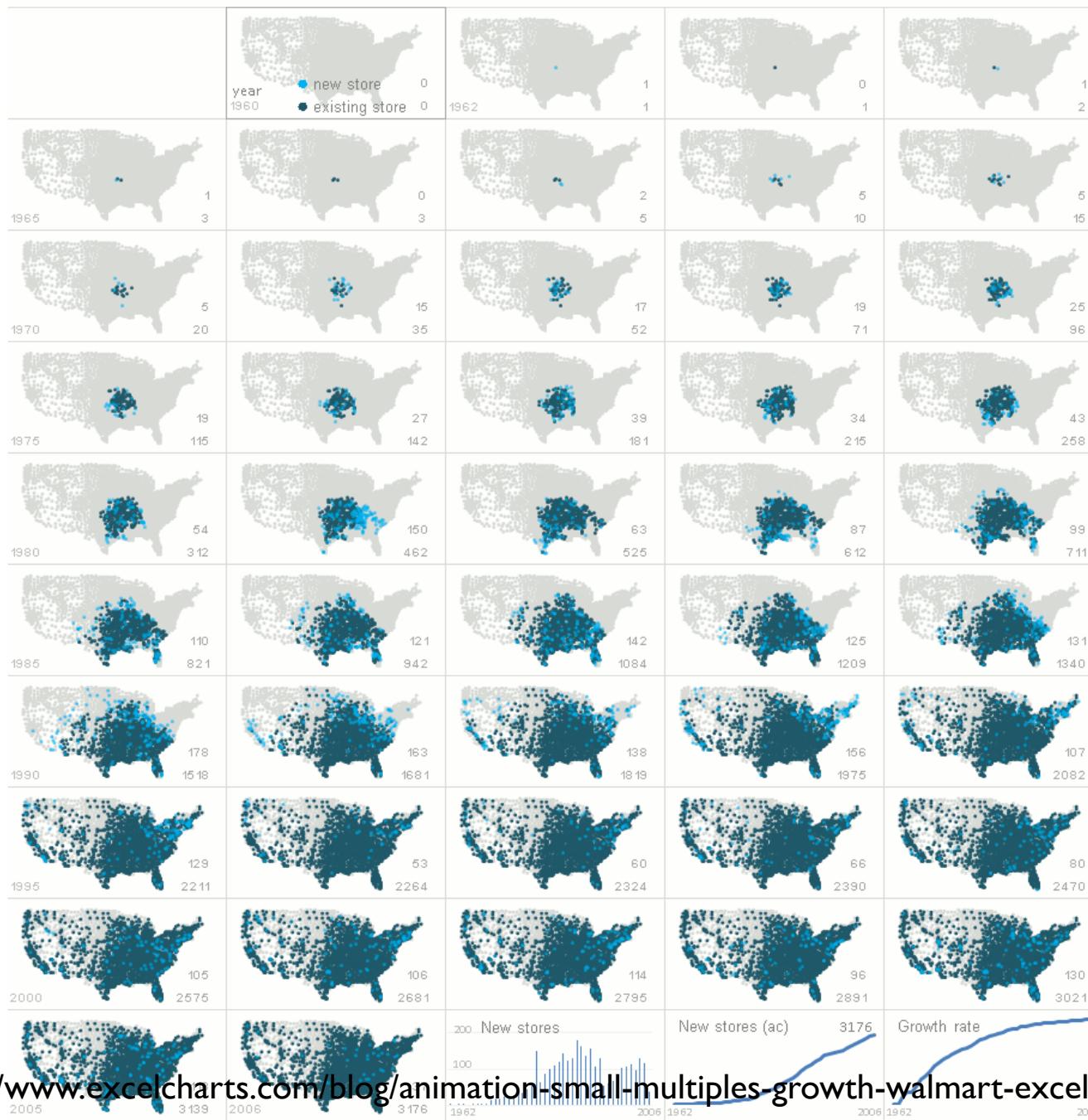


Robberies per 100,000 people



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

## Growth of Walmart

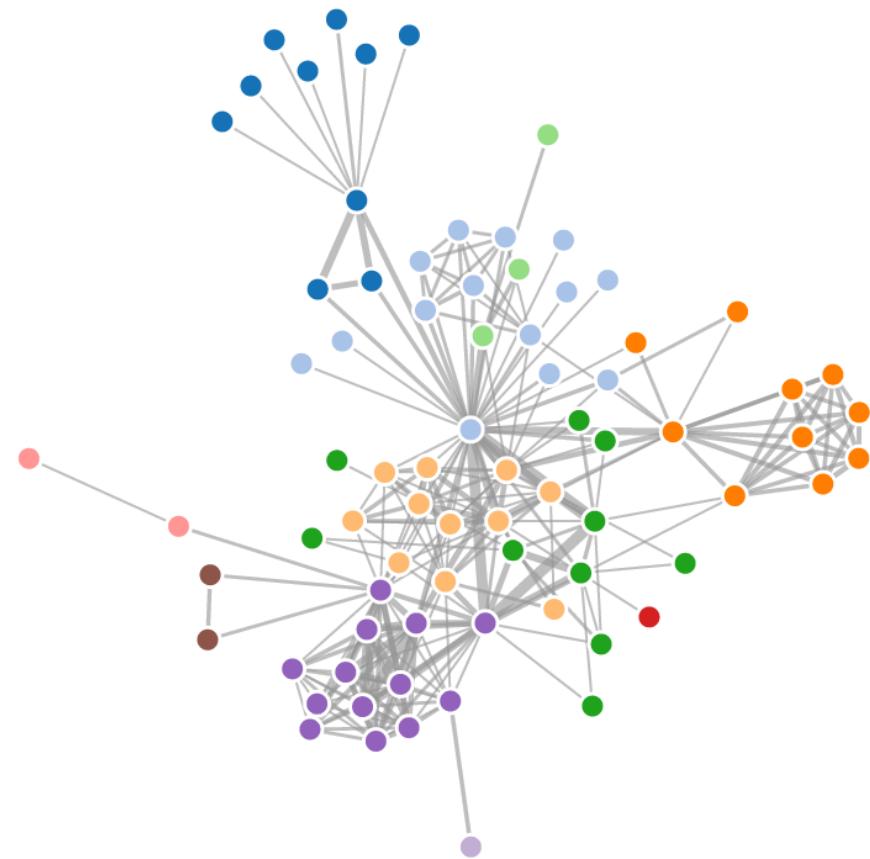


# **Small Multiples vs. Animations**

- 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

