

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

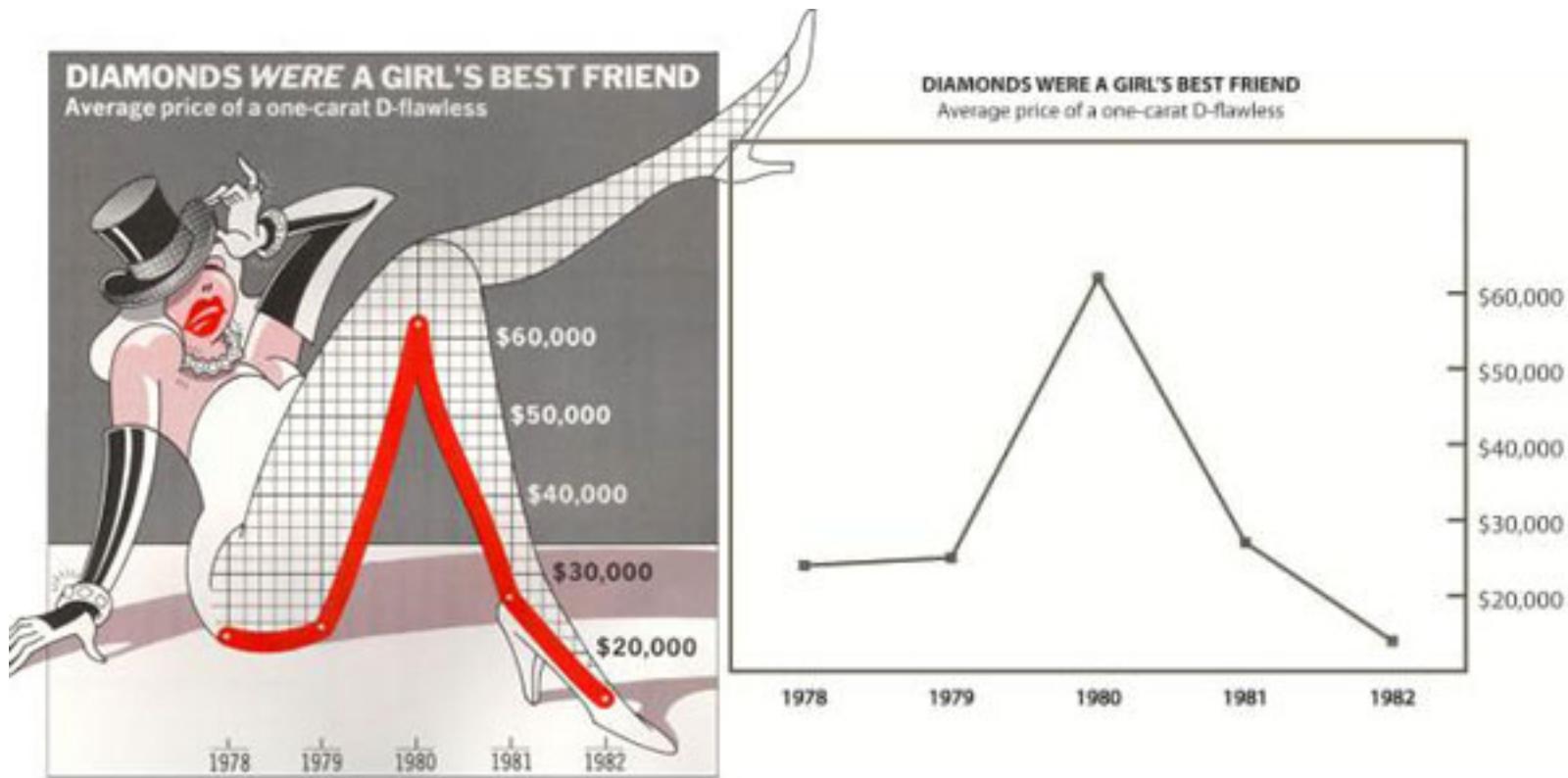


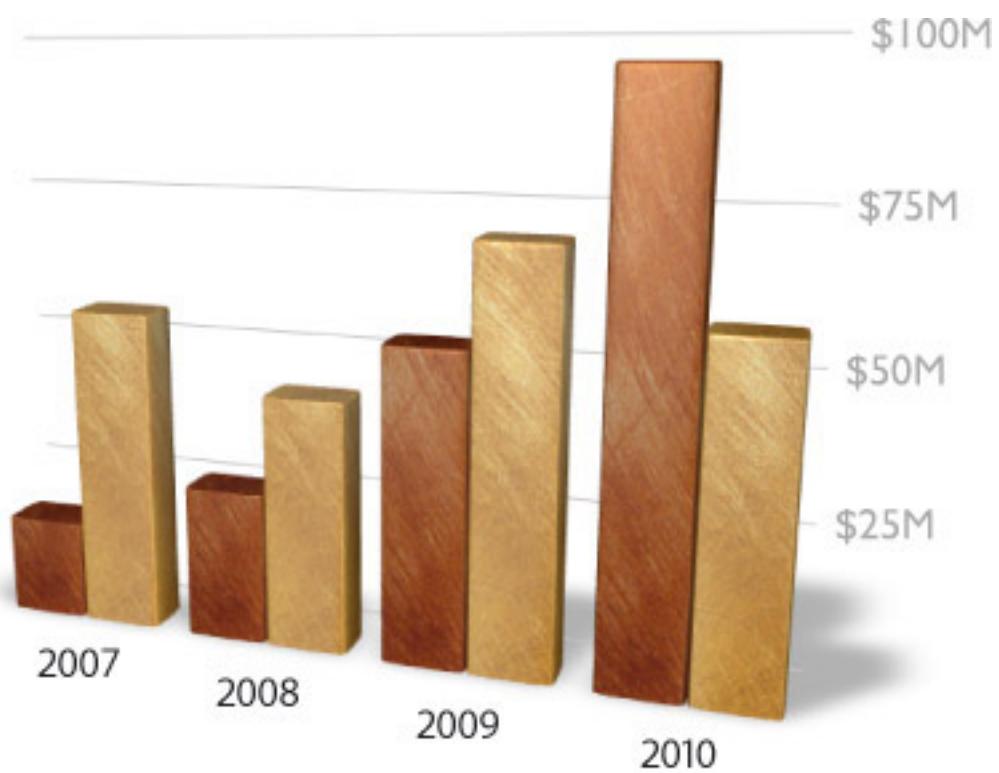
Static vs. Interactive

- 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

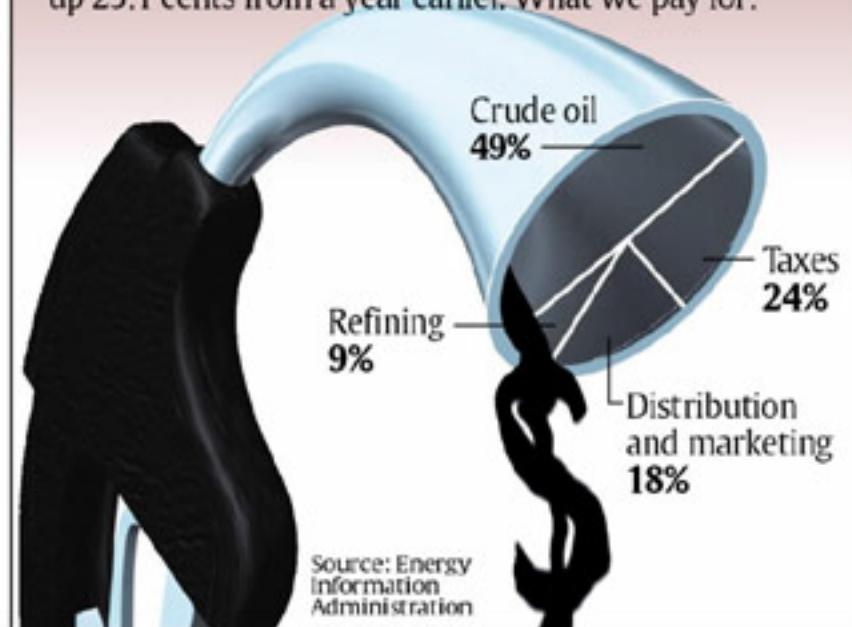




USA TODAY Snapshots®

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

2.40 m
2.35 m
2.30 m
2.25 m
2.20 m
2.15 m
2.10 m
2.05 m
2.00 m
1.95 m
1.90 m
1.85 m
1.80 m
1.75 m
1.70 m
1.65 m
1.60 m
1.55 m
1.50 m
1.45 m
1.40 m
1.35 m
1.30 m
1.25 m
1.20 m
1.15 m
1.10 m
1.05 m
1.00 m
0.95 m
0.90 m
0.85 m
0.80 m
0.75 m
0.70 m
0.65 m
0.60 m
0.55 m
0.50 m
0.45 m
0.40 m
0.35 m
0.30 m
0.25 m
0.20 m
0.15 m
0.10 m
0.05 m
0.00 m

POLE VAULT
+71%
Performance history

PERFORMANCES HAVE DECLINED SINCE 2000.

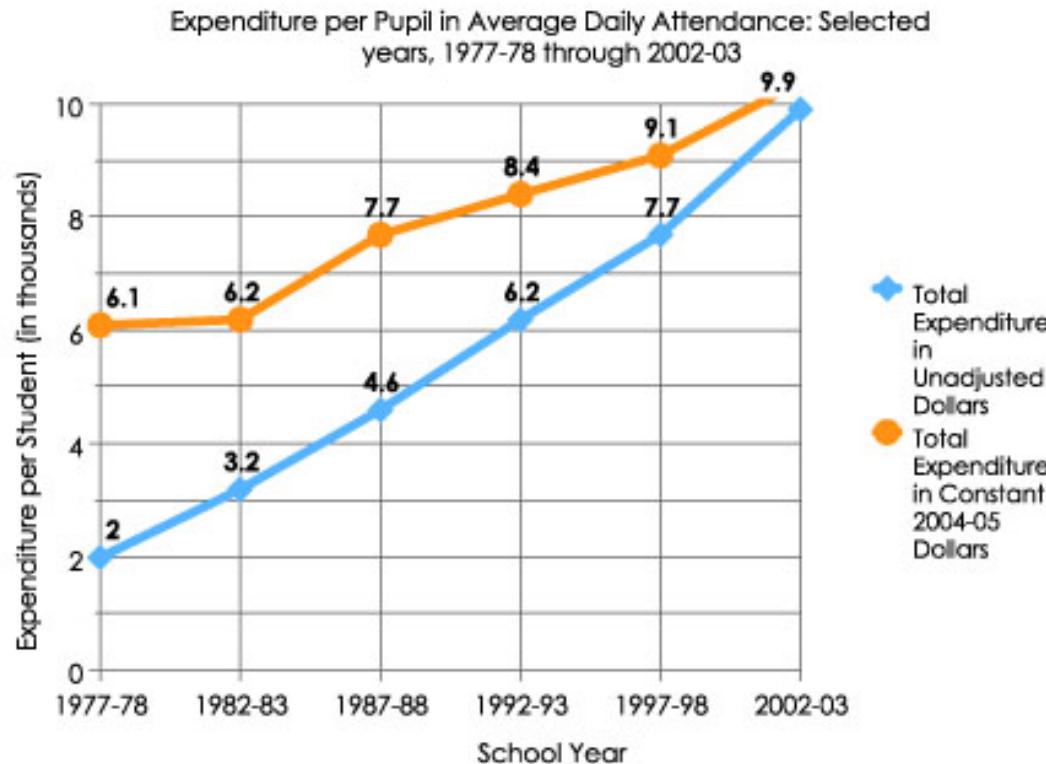
6.14 m
SERGEY BUBKA UKR 1994

6.10 m
6.00 m
5.90 m
5.80 m
5.70 m
5.60 m
5.50 m
5.40 m
5.30 m
5.20 m
5.10 m
5.00 m
4.90 m
4.80 m
4.70 m
4.60 m
4.50 m
4.40 m
4.30 m
4.20 m
4.10 m
4.00 m
3.90 m
3.80 m
3.70 m
3.60 m
3.50 m
3.40 m
3.30 m
3.20 m
3.10 m
3.00 m
2.90 m
2.80 m
2.70 m
2.60 m
2.50 m
2.40 m
2.30 m
2.20 m
2.10 m
2.00 m
1.90 m
1.80 m
1.70 m
1.60 m
1.50 m
1.40 m
1.30 m
1.20 m
1.10 m
1.00 m
0.90 m
0.80 m
0.70 m
0.60 m
0.50 m
0.40 m
0.30 m
0.20 m
0.10 m
0.00 m

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

Line Graphs (+ Variants)

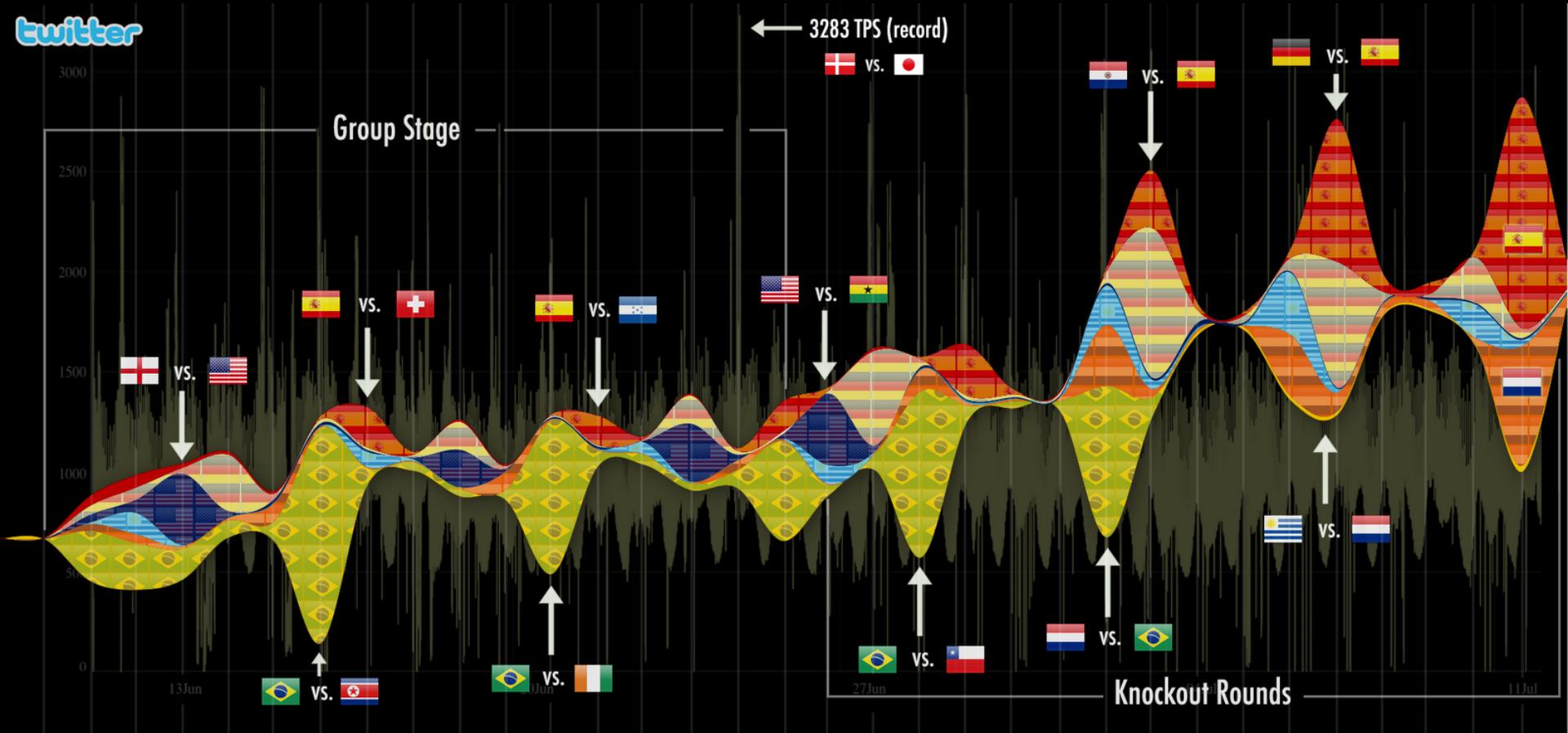
- Suitable for what types of data?
- What makes a good line graph?



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”



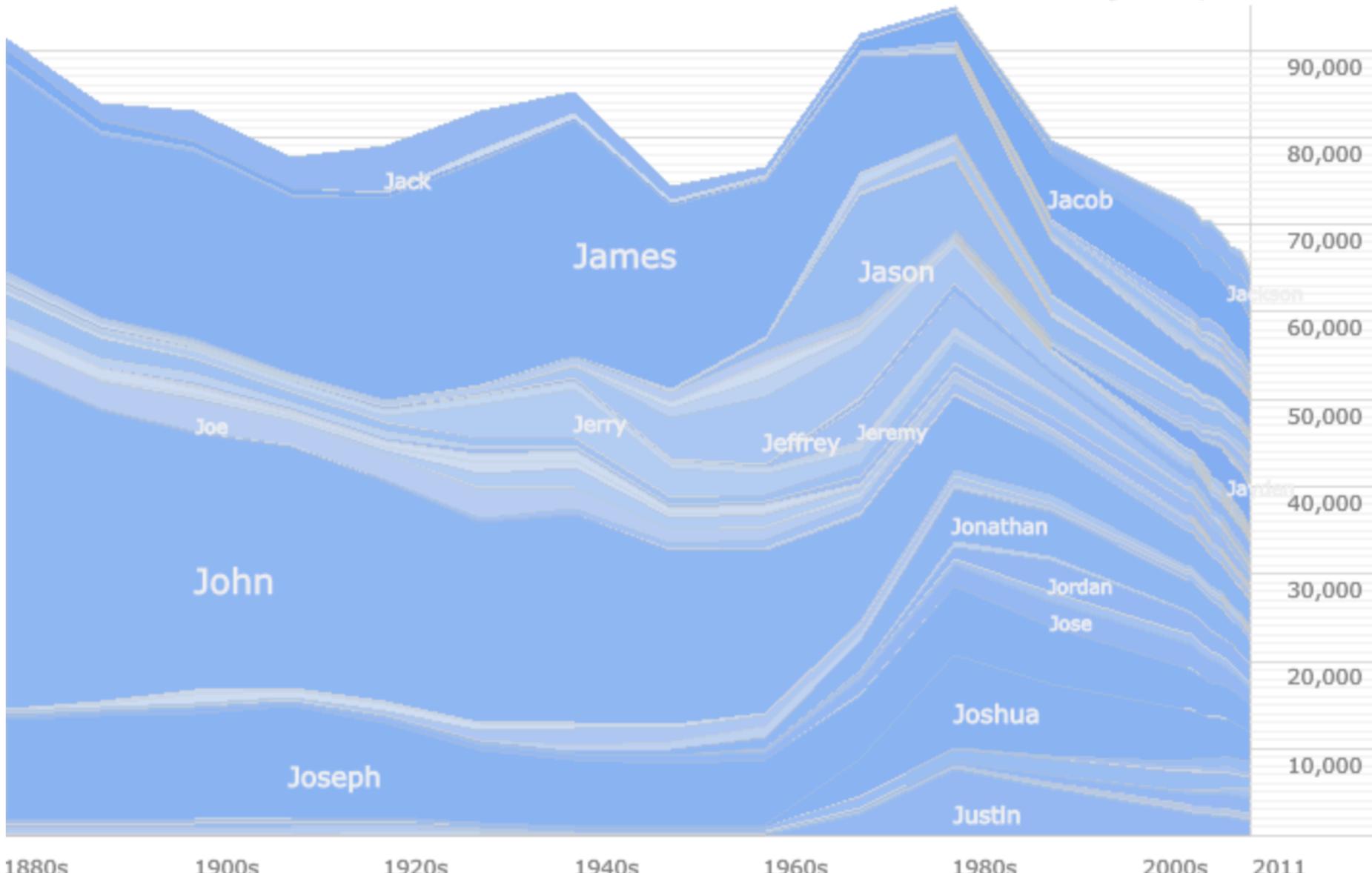
Baby Name > j



Both Boys Girls

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

Names starting with 'J' per million babies



1880s

1900s

1920s

1940s

1960s

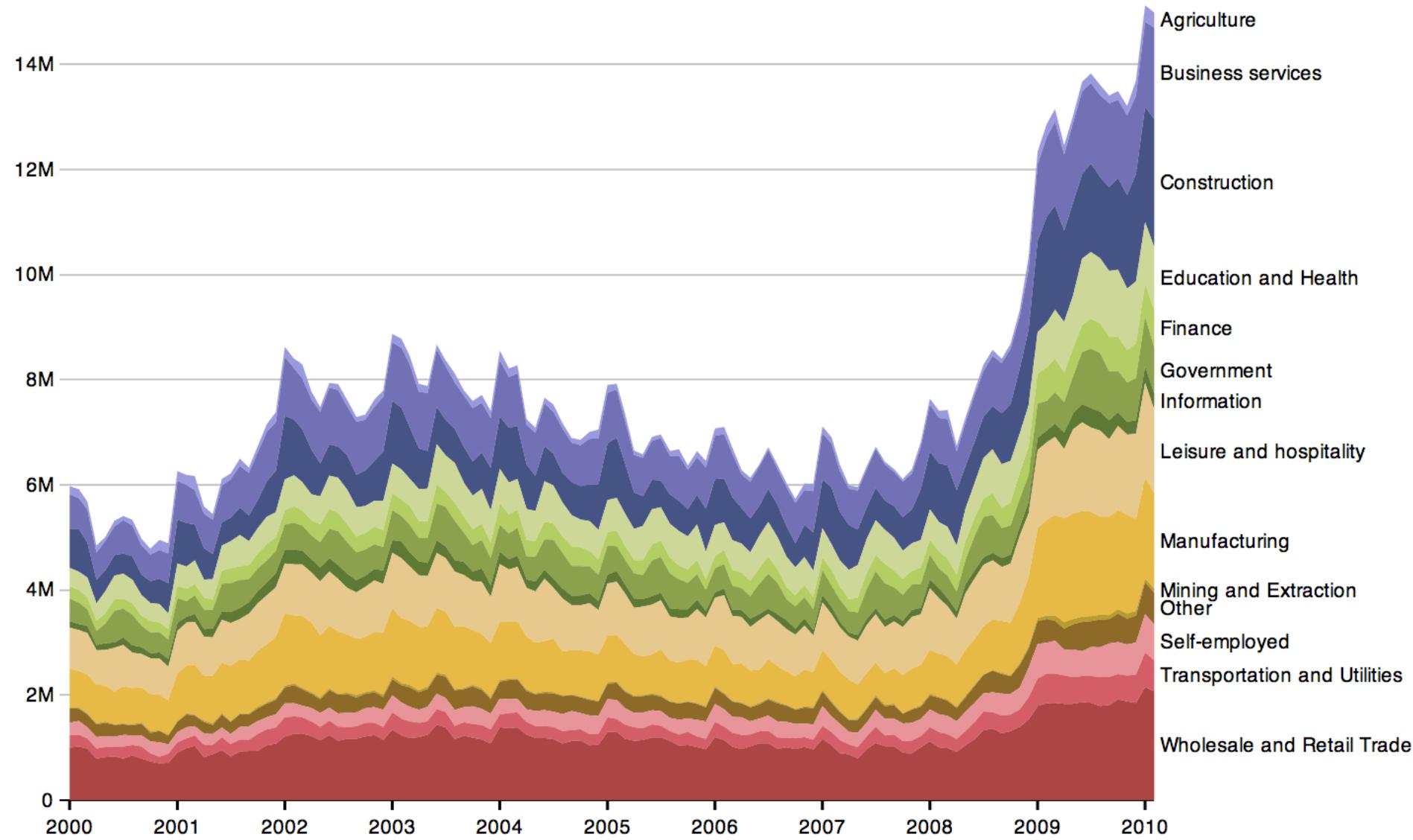
1980s

2000s

2011

Stacked Graph of Unemployed U.S. Workers by Industry

View: Counts (Zero-Baseline) ▾

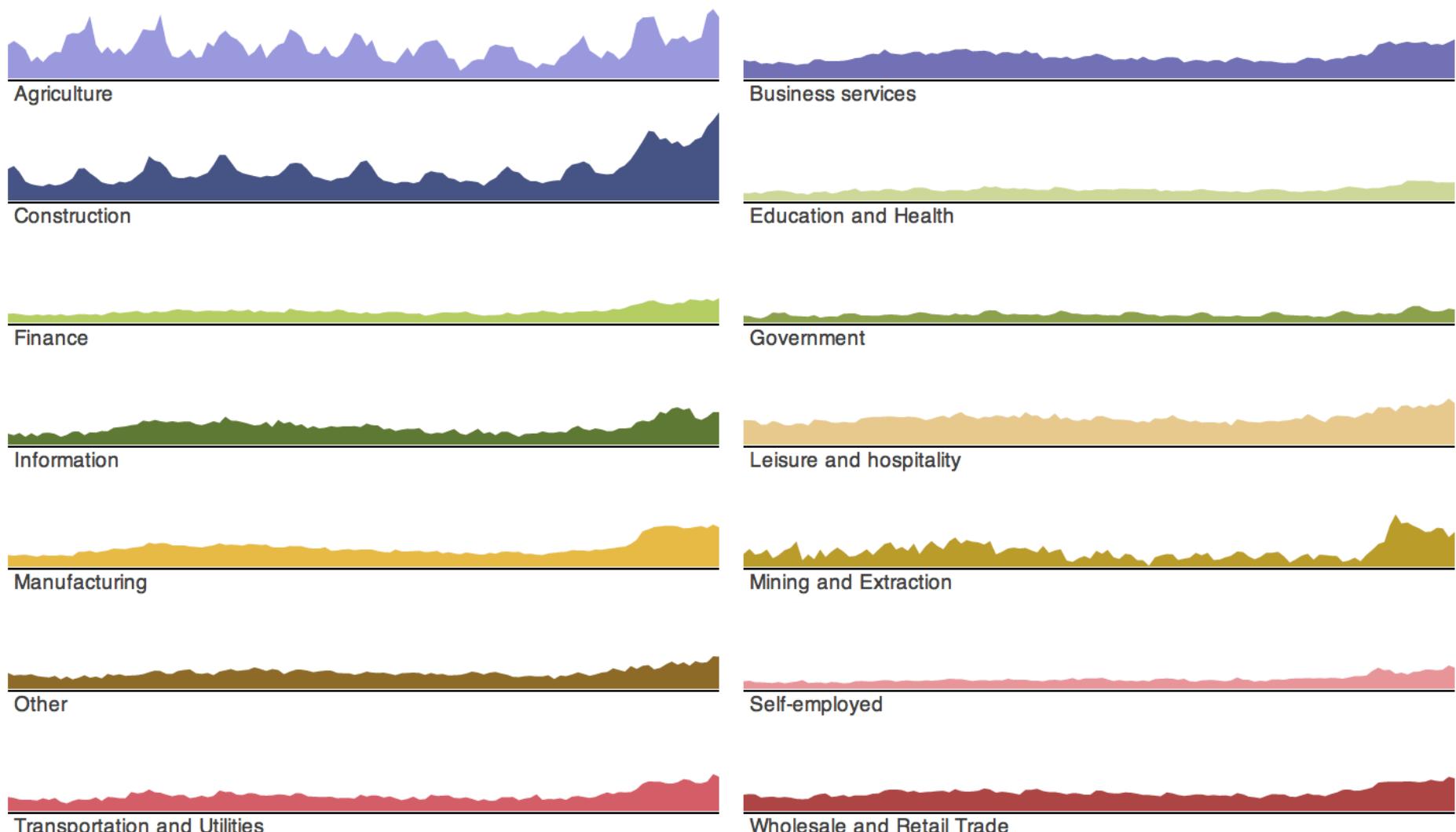


Total counts of unemployed persons per industry, 2000-2010.

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.

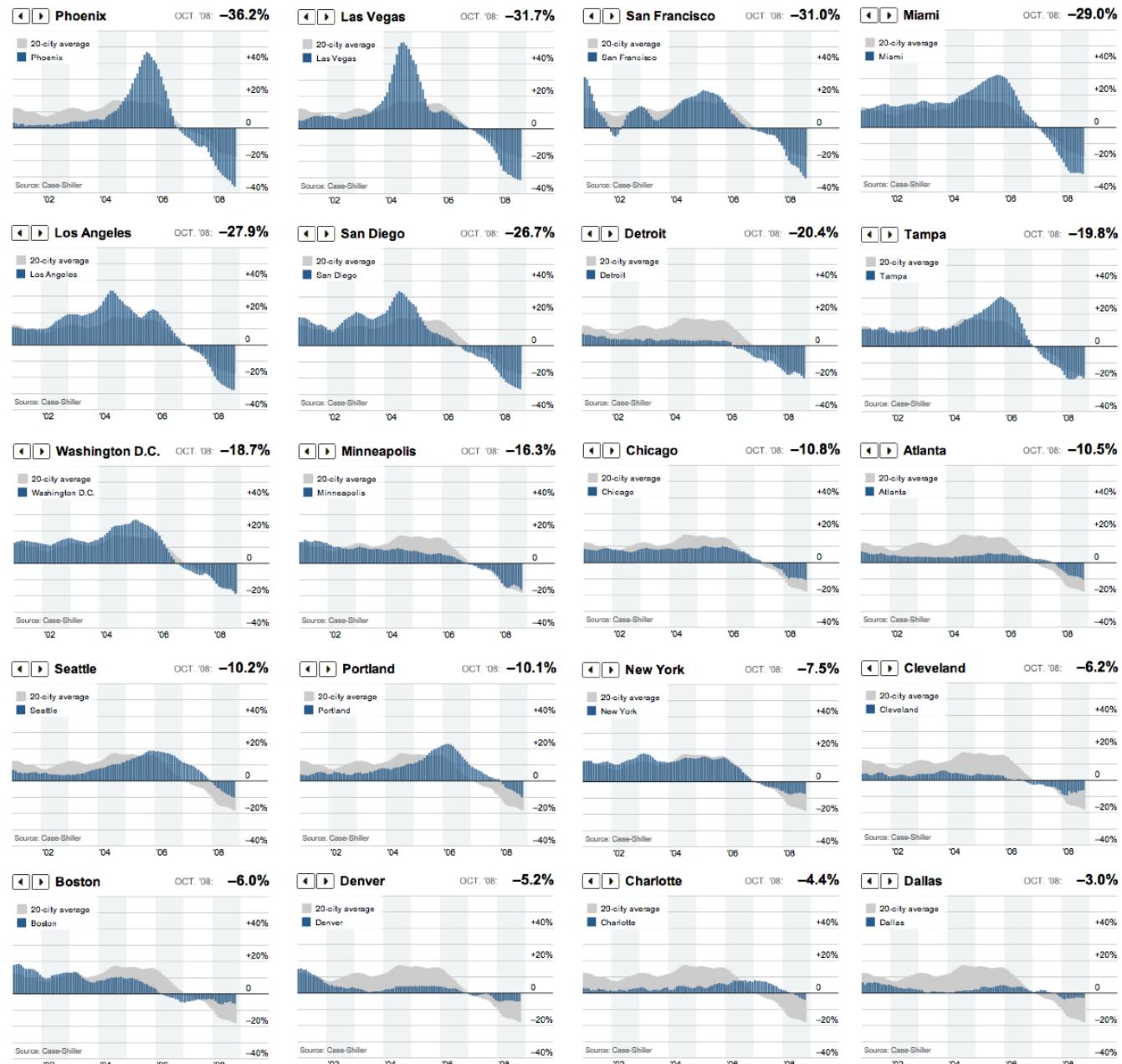
Source: [U.S. Bureau of Labor Statistics](#)

Small Multiples

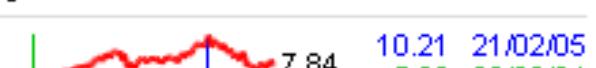
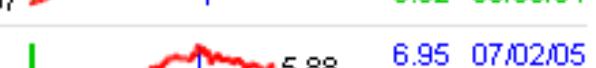
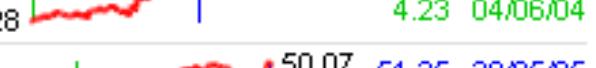
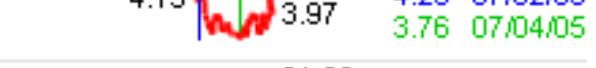
- Multiple small graphs arranged in a grid
- Advantages and disadvantages?

Change in Home Prices (year over year)

From New York Times Economix blog

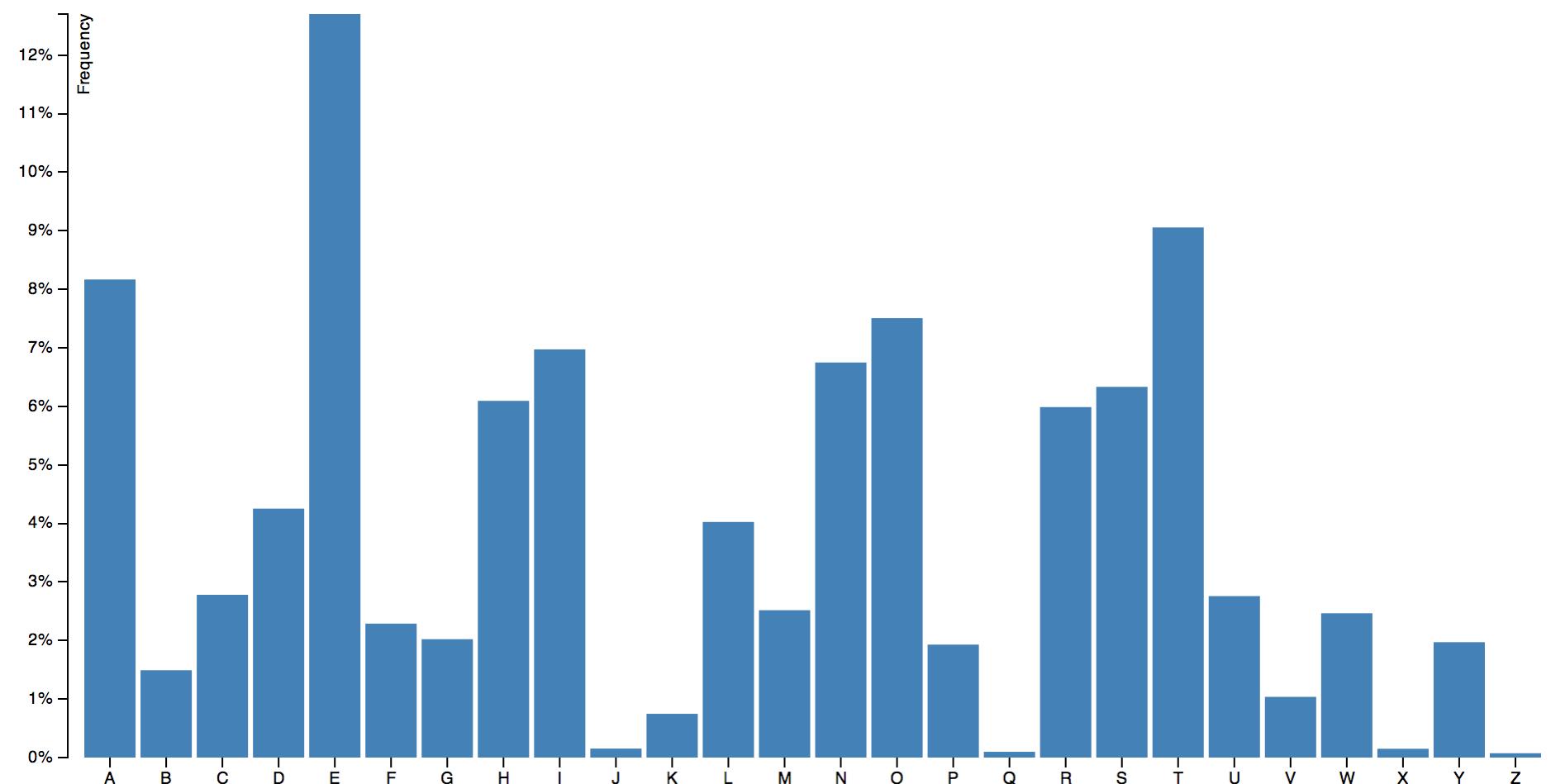


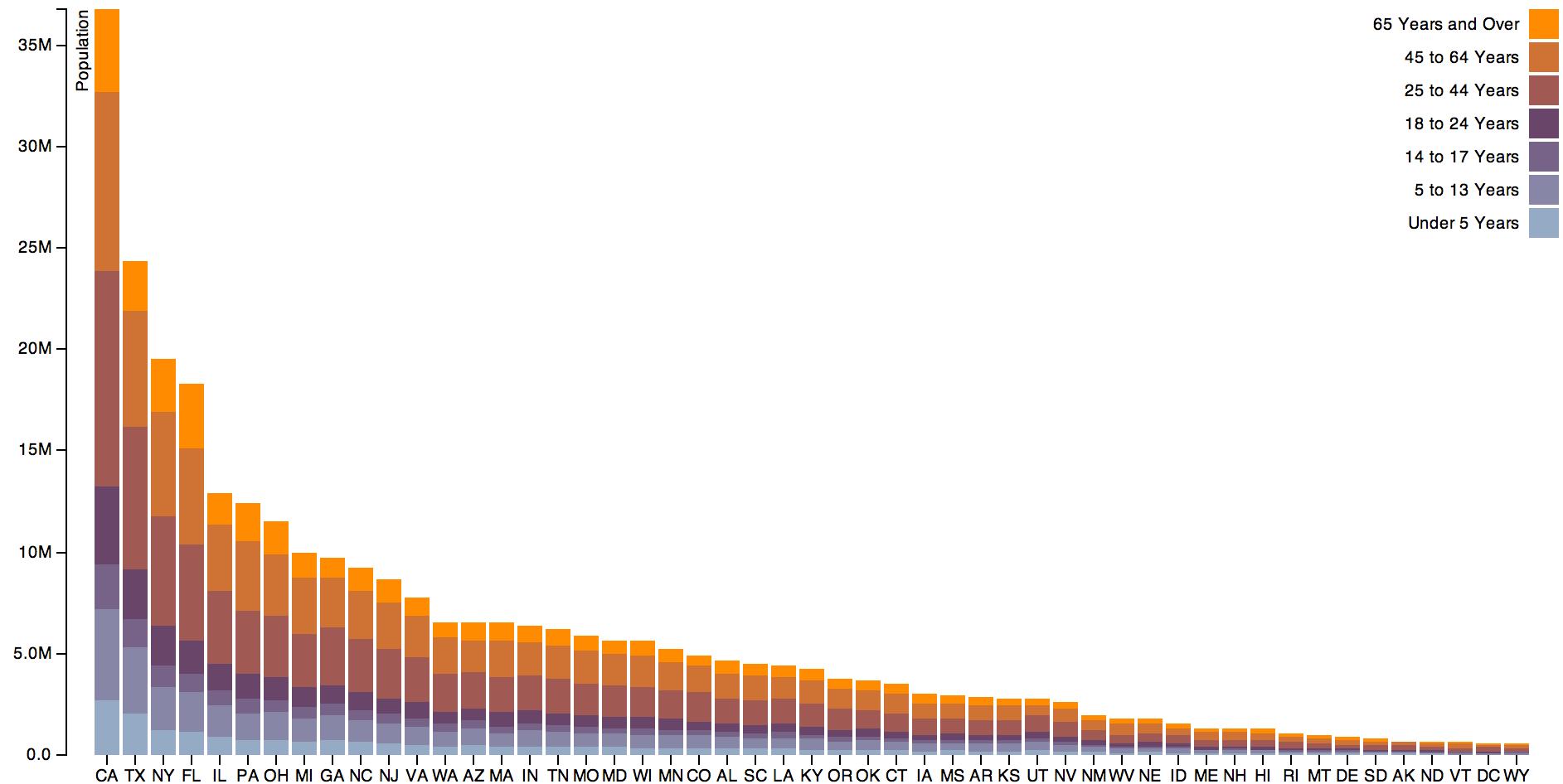
Spark Lines

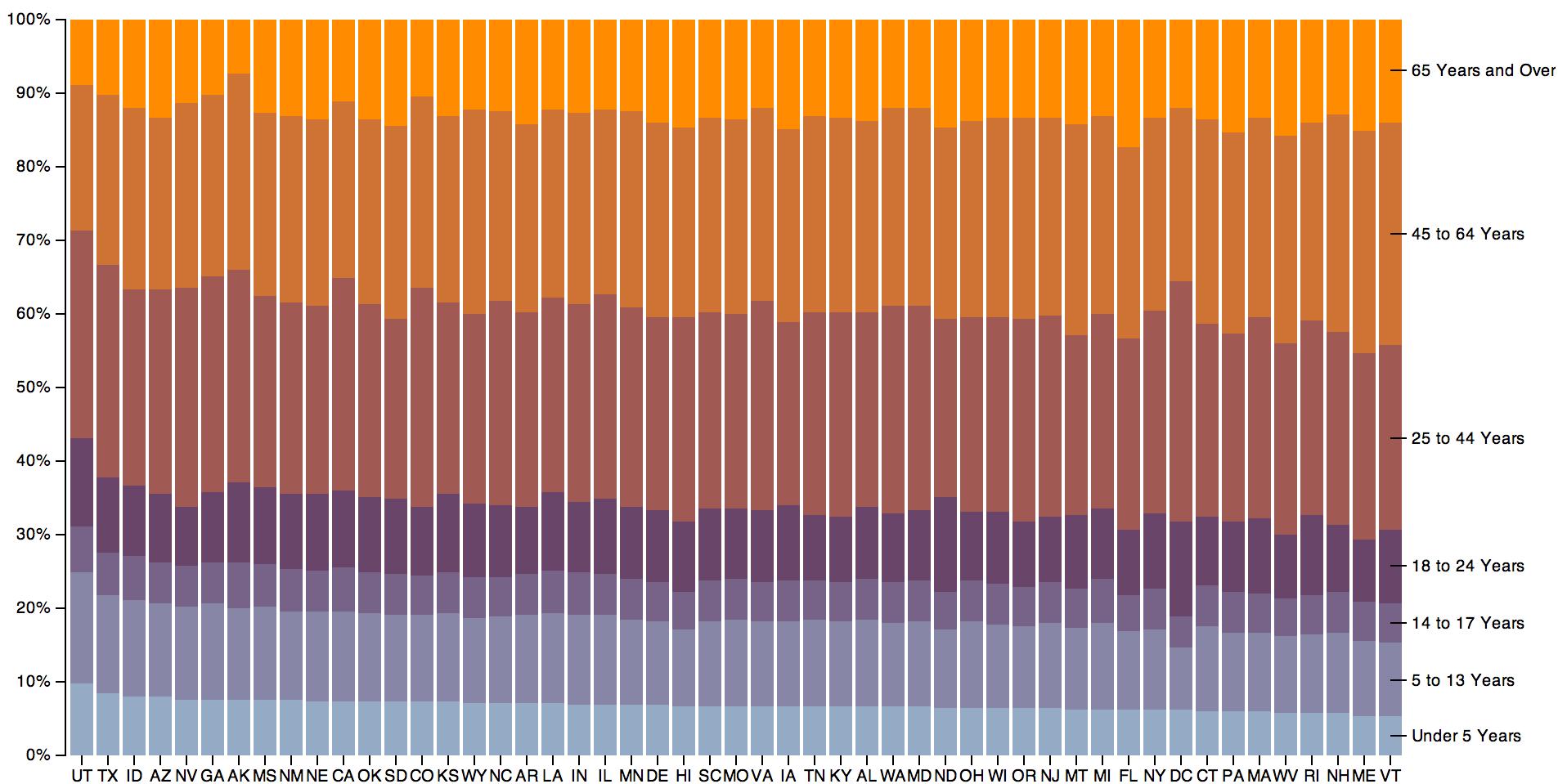
AGL	11.71		13.65	14.51 11.64	03/02/05 03/06/04	NFD	4.29		0.03	6.50 4.25	07/04/05 15/07/04
ALN	5.92		9.65	9.83 5.92	24/05/05 02/06/04	ORI	14.89		15.10	20.41 14.65	17/12/04 17/05/05
ARG	4.77		5.54	5.77 4.76	14/02/05 08/06/04	PBL	12.54		15.46	17.50 12.54	29/12/04 02/06/04
BHP	12.19		16.05	19.29 11.92	07/03/05 03/06/04	RIN	7.60		12.20	12.50 7.55	24/05/05 03/06/04
BSL	6.07		7.84	10.21 5.92	21/02/05 03/06/04	RIO	35.50		42.35	47.93 34.85	14/03/05 03/06/04
IAG	4.60		5.88	6.95 4.60	07/02/05 02/06/04	TLS	4.67		4.92	5.49 4.61	22/03/05 10/06/04
ILU	4.28		6.00	6.56 4.23	08/02/05 04/06/04	WBC	17.39		19.92	20.13 16.14	10/02/05 16/08/04
MBL	33.51		50.07	51.35 31.66	20/05/05 10/08/04	WDC	15.30		16.75	17.39 14.57	17/01/05 26/10/04
MGQ			4.13 3.97	4.25 3.76	07/02/05 07/04/05	WES	27.95		36.64	42.37 27.51	21/03/05 03/06/04
NAB	30.28		31.80	31.83 26.08	24/05/05 12/08/04	WOW	11.70		16.02	16.69 11.38	21/03/05 16/07/04

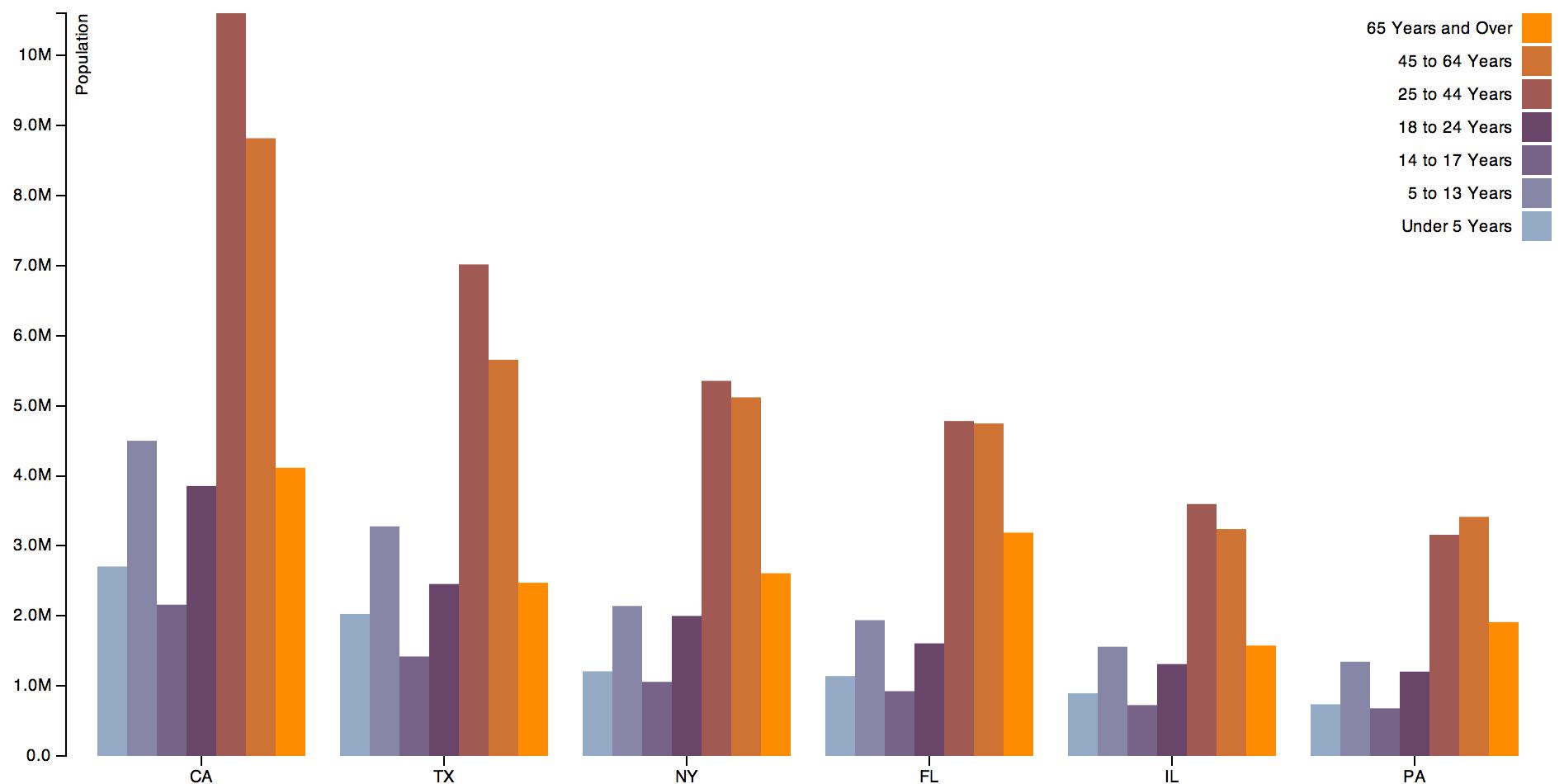
Bar Graphs (+ Variants)

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



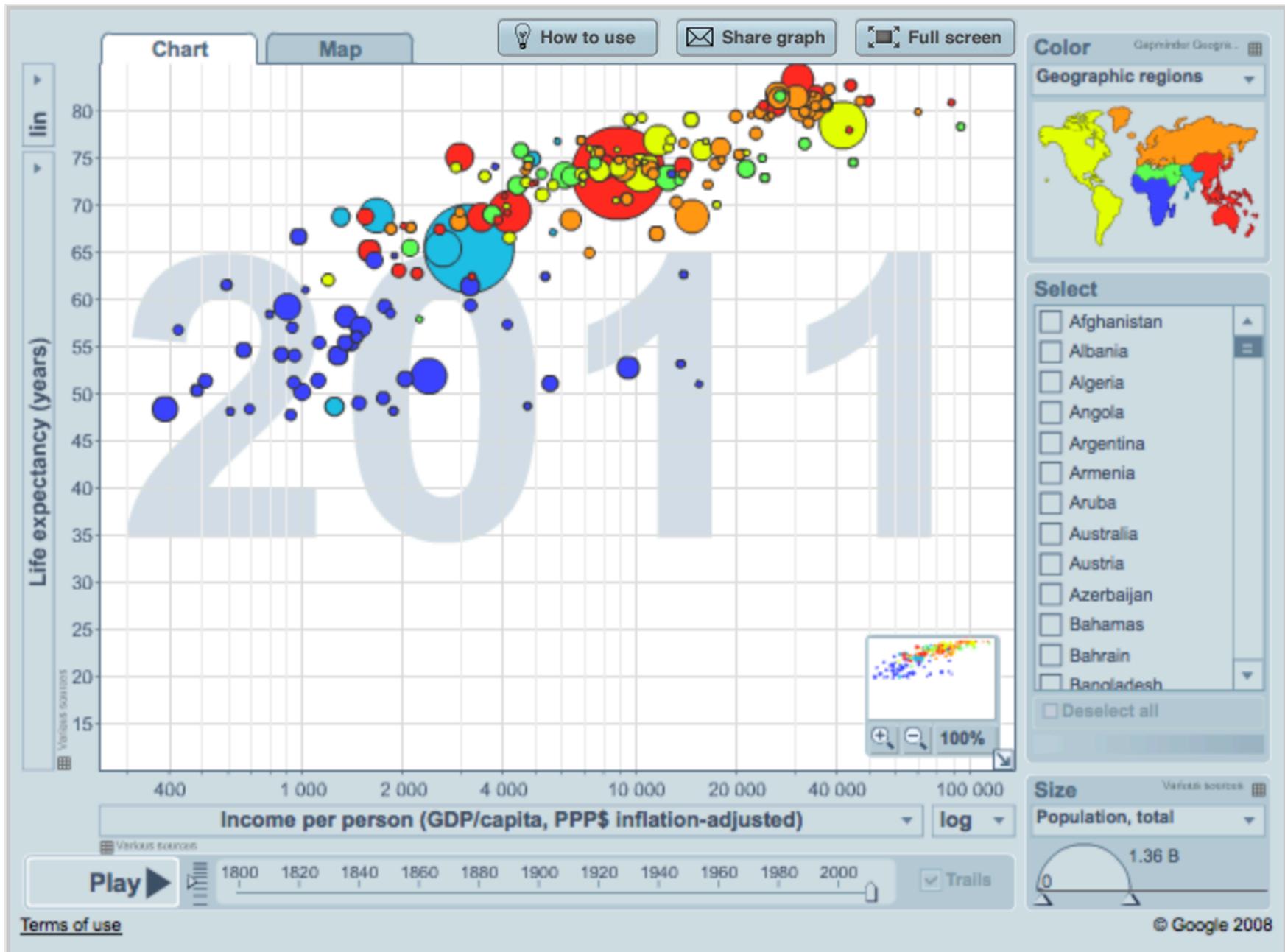






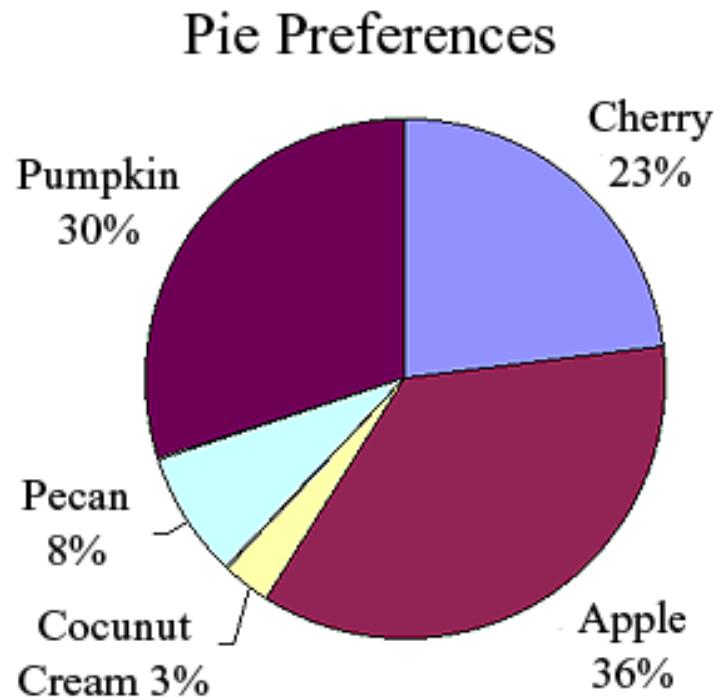
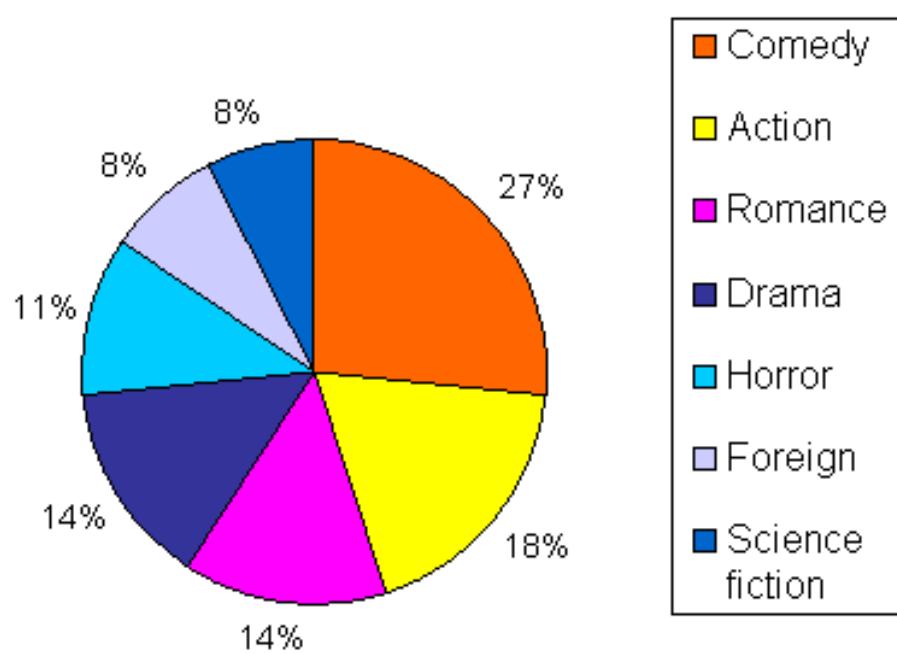
Scatterplots

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

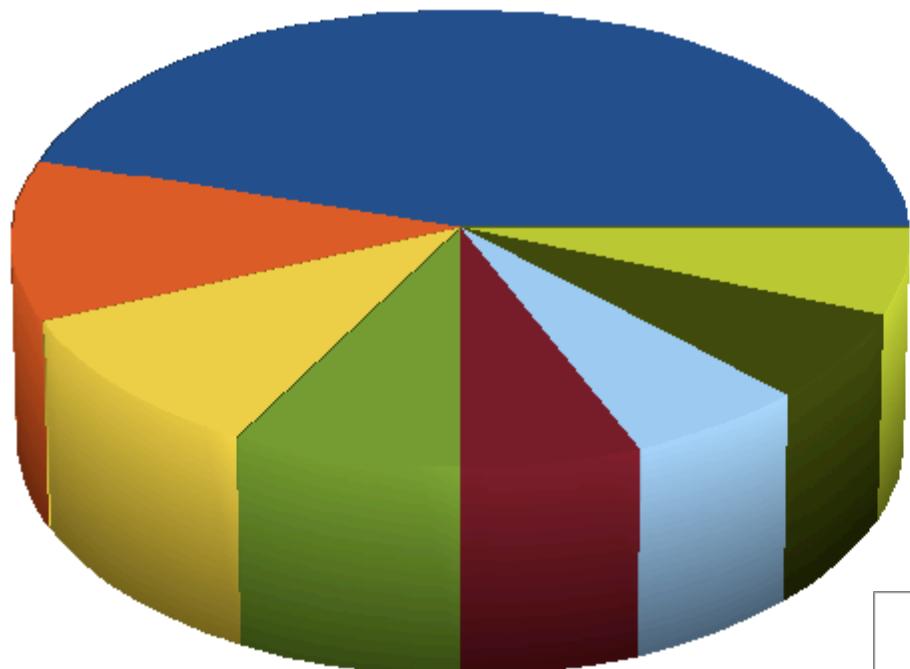


Pie Charts (+ Variants)

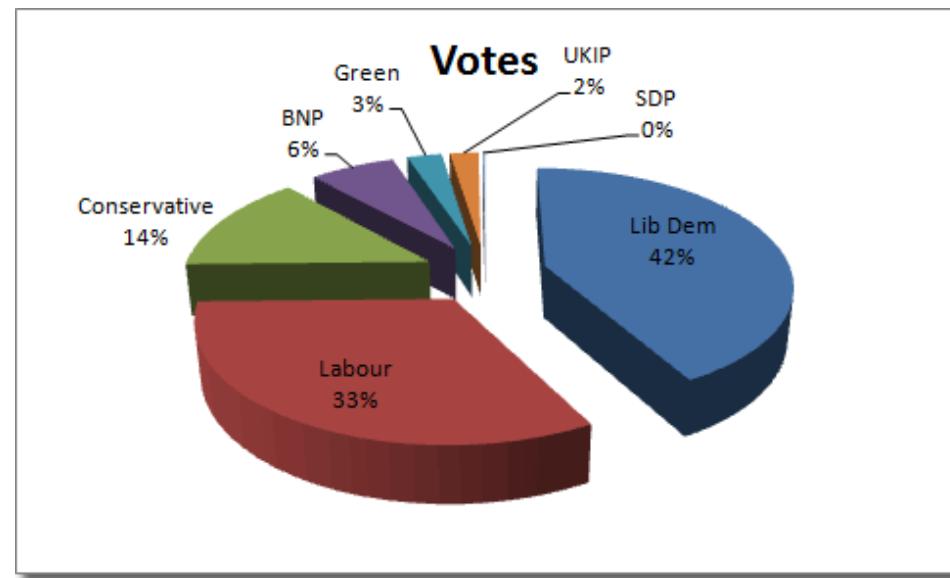
- Suitable for what types of data?
- Different ways for labeling



More Chart Junk



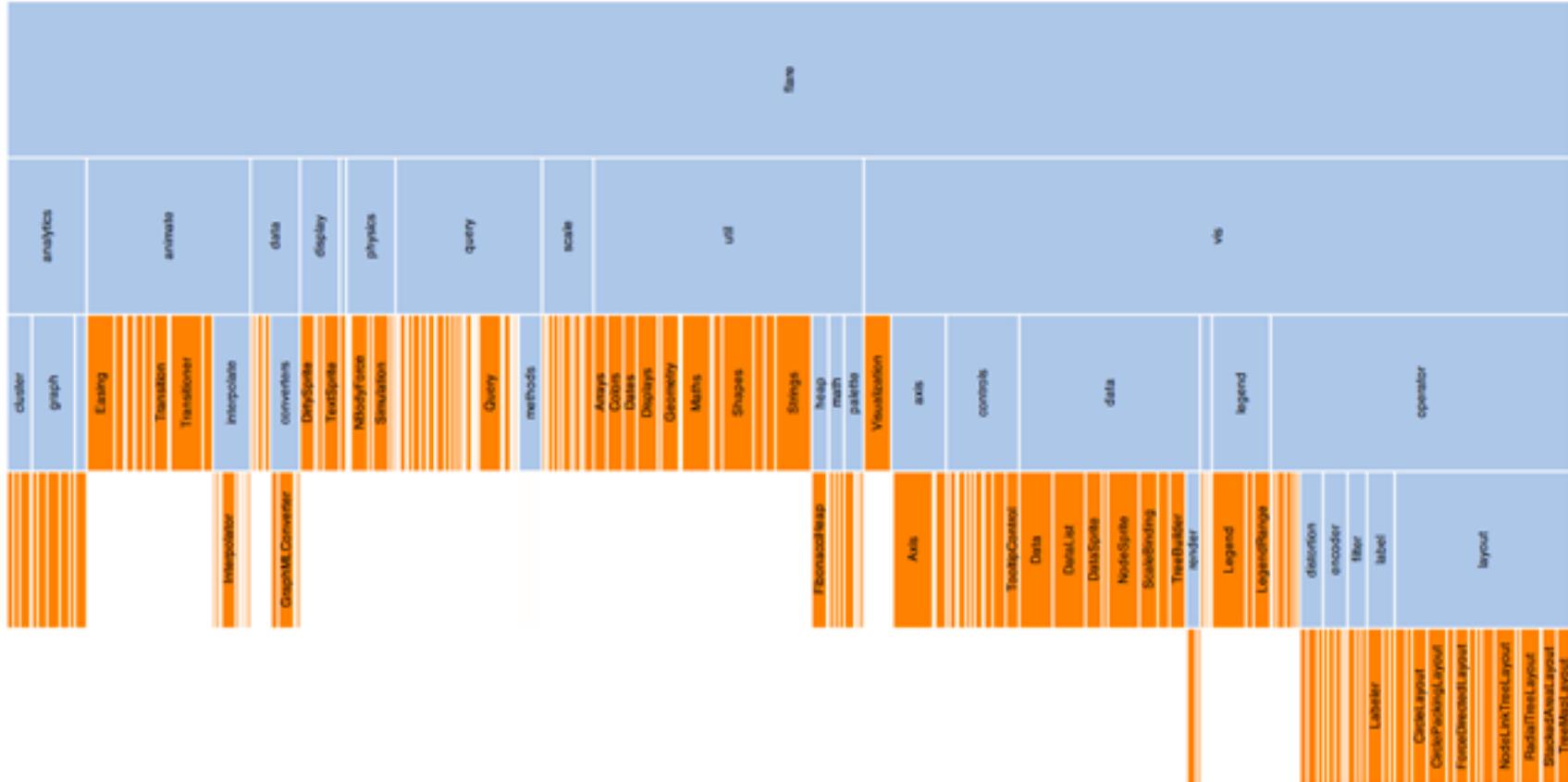
■ MySpace
■ Facebook
■ Orkut
■ hi5
■ Xanga
■ Live Spaces
■ Classmates.com
■ Bebo



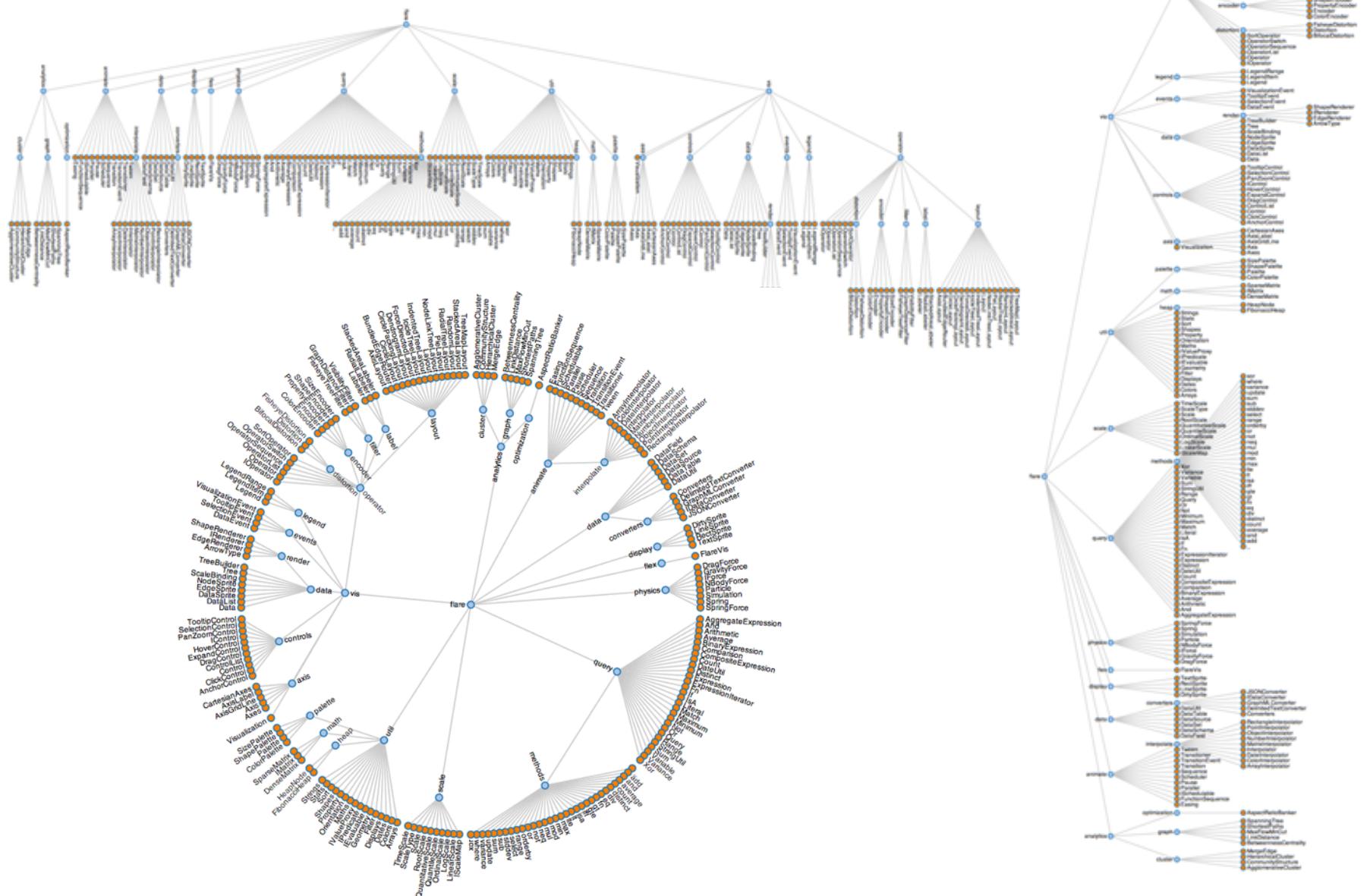
Sunburst Trees



Icicle Diagrams

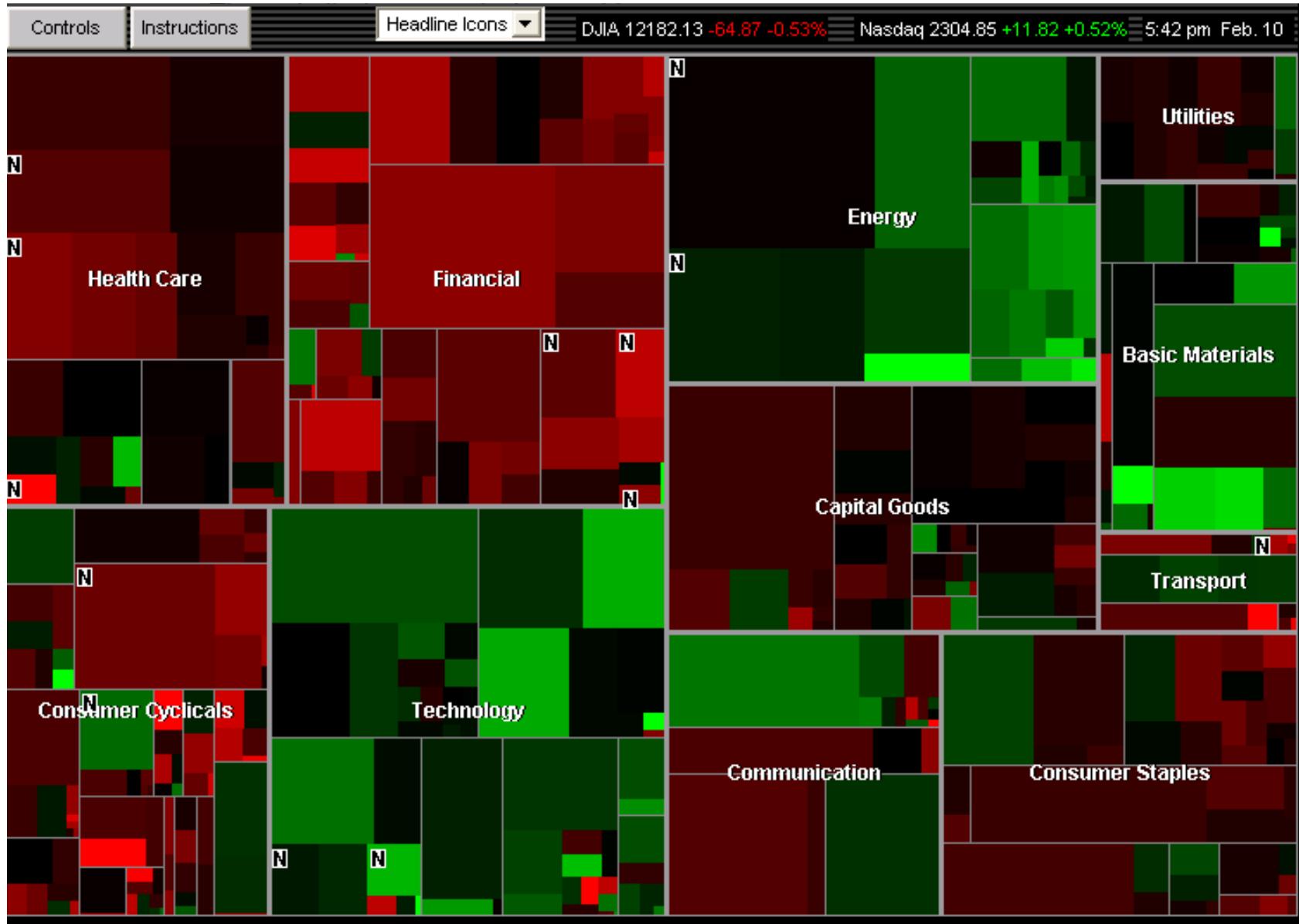


Visualizing Hierarchies



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

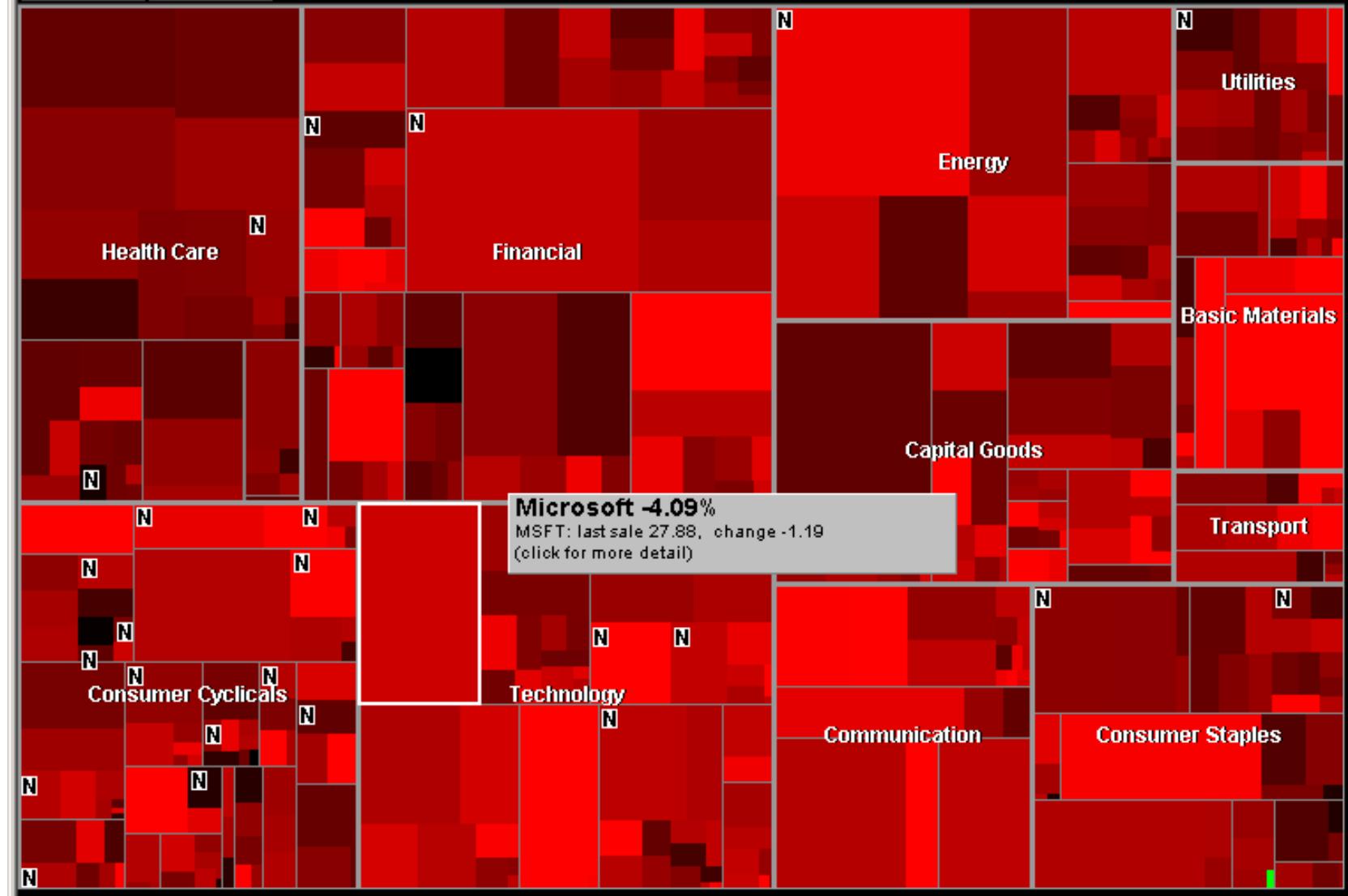
Instructions

Headline Icons

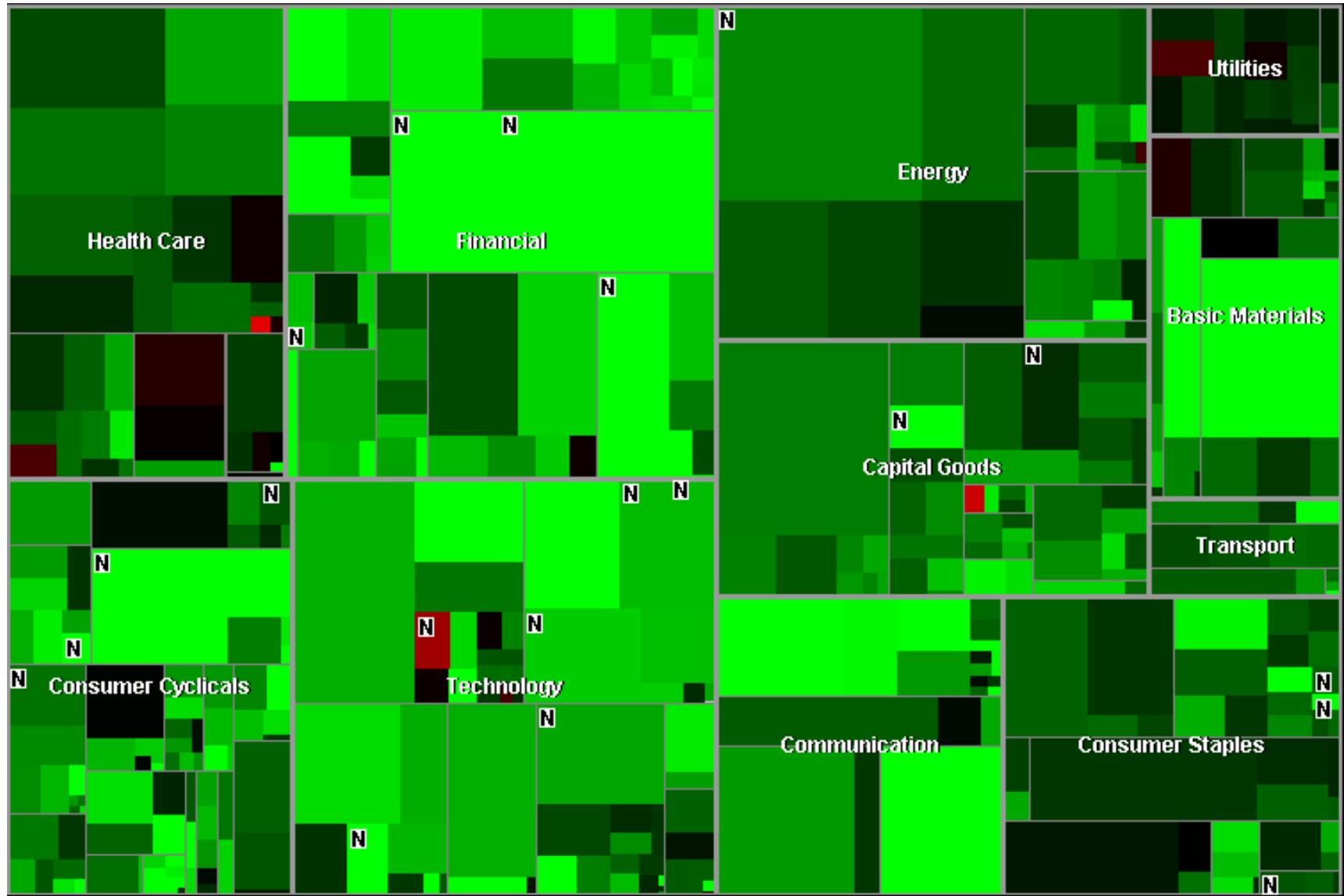
DJIA 12216.24 -416.02 -3.29%

Nasdaq 2407.87 -96.65 -3.86%

4:32 pm Feb. 27



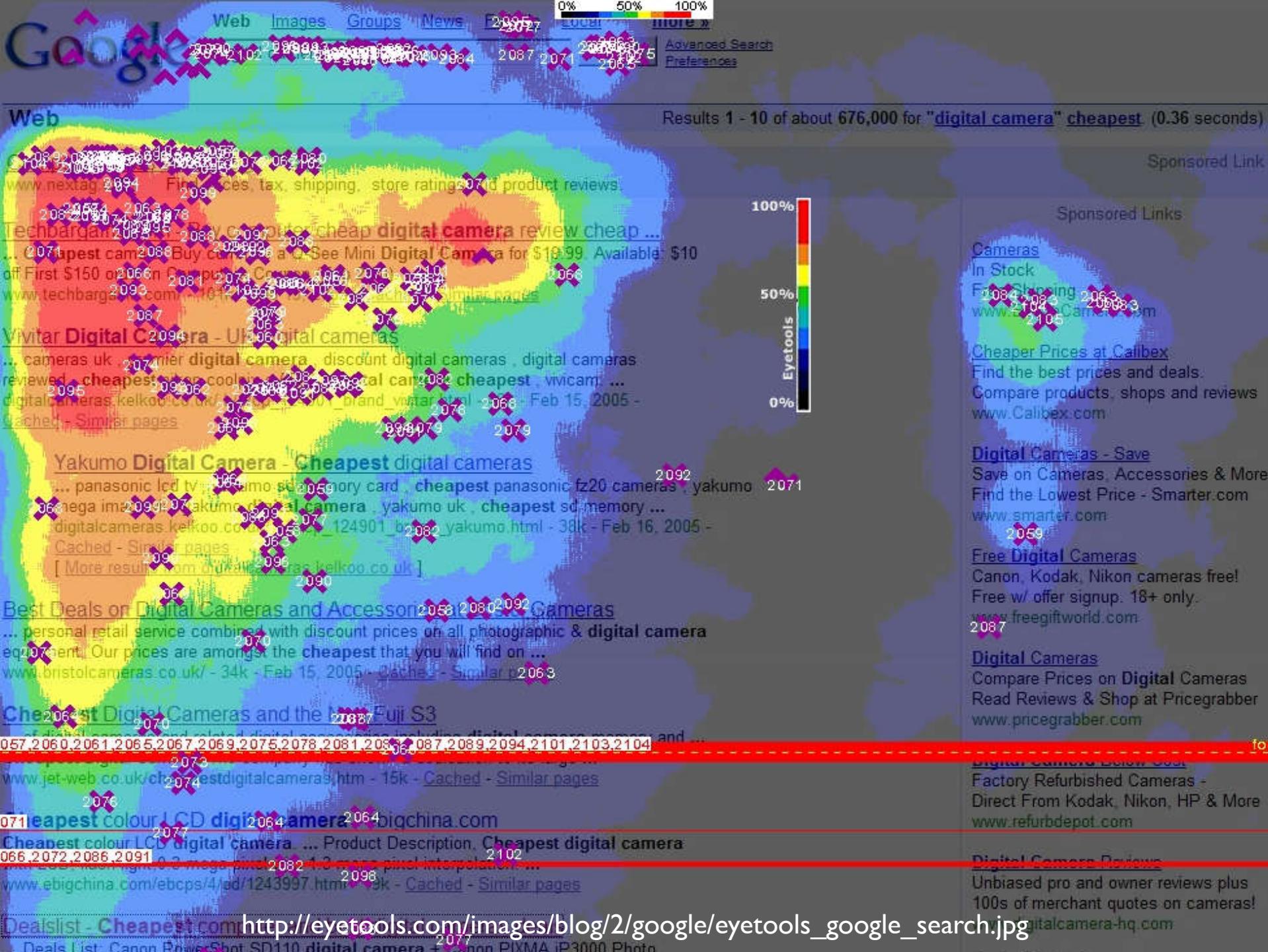
Market mixed, February 8, 2008:
Energy & Technology up Financial & Health Care down



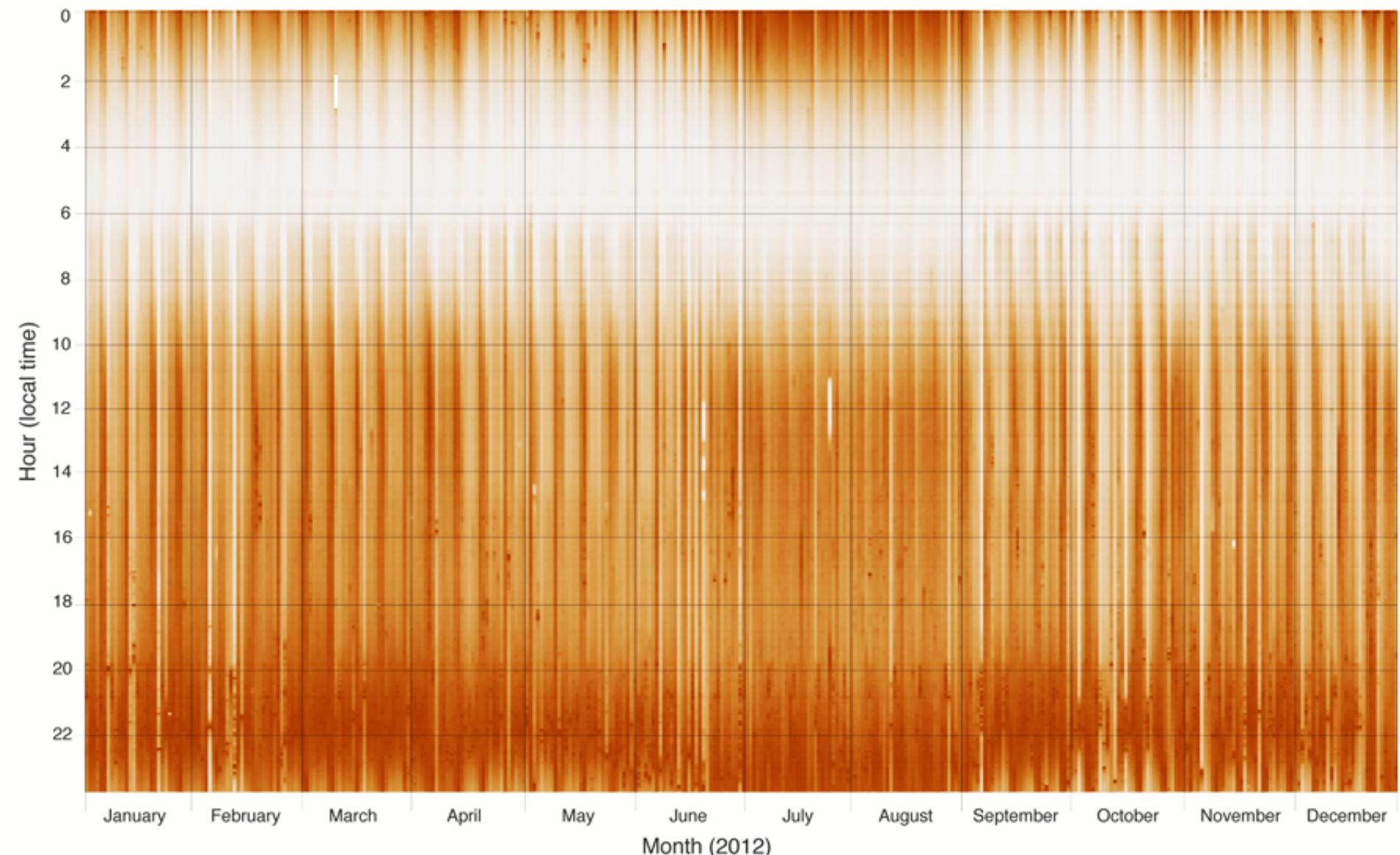
Market rises 319 points, November 13, 2007:
Five exceptions

Heatmaps

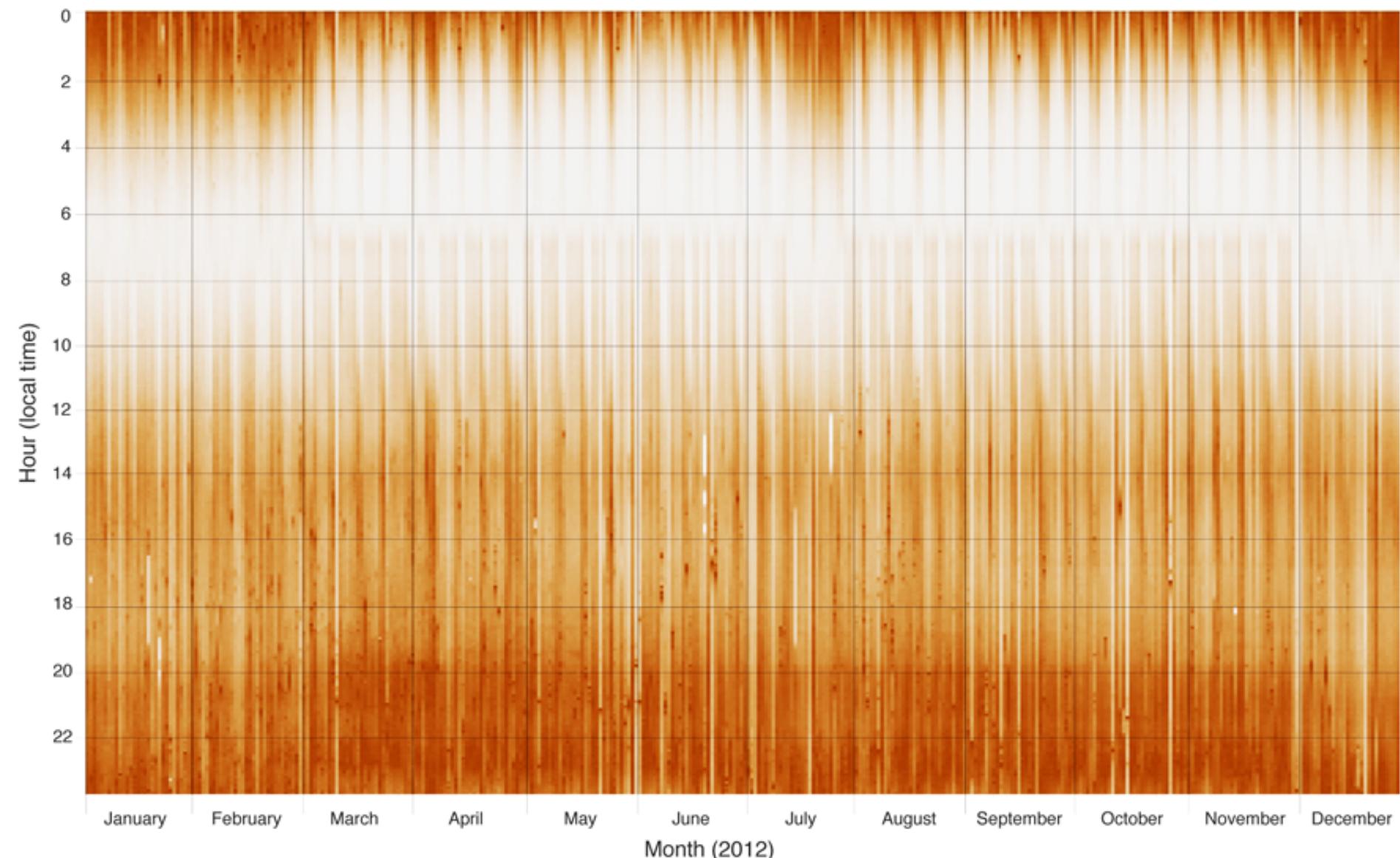
- Data represented in a matrix
- Matrix values are color coded



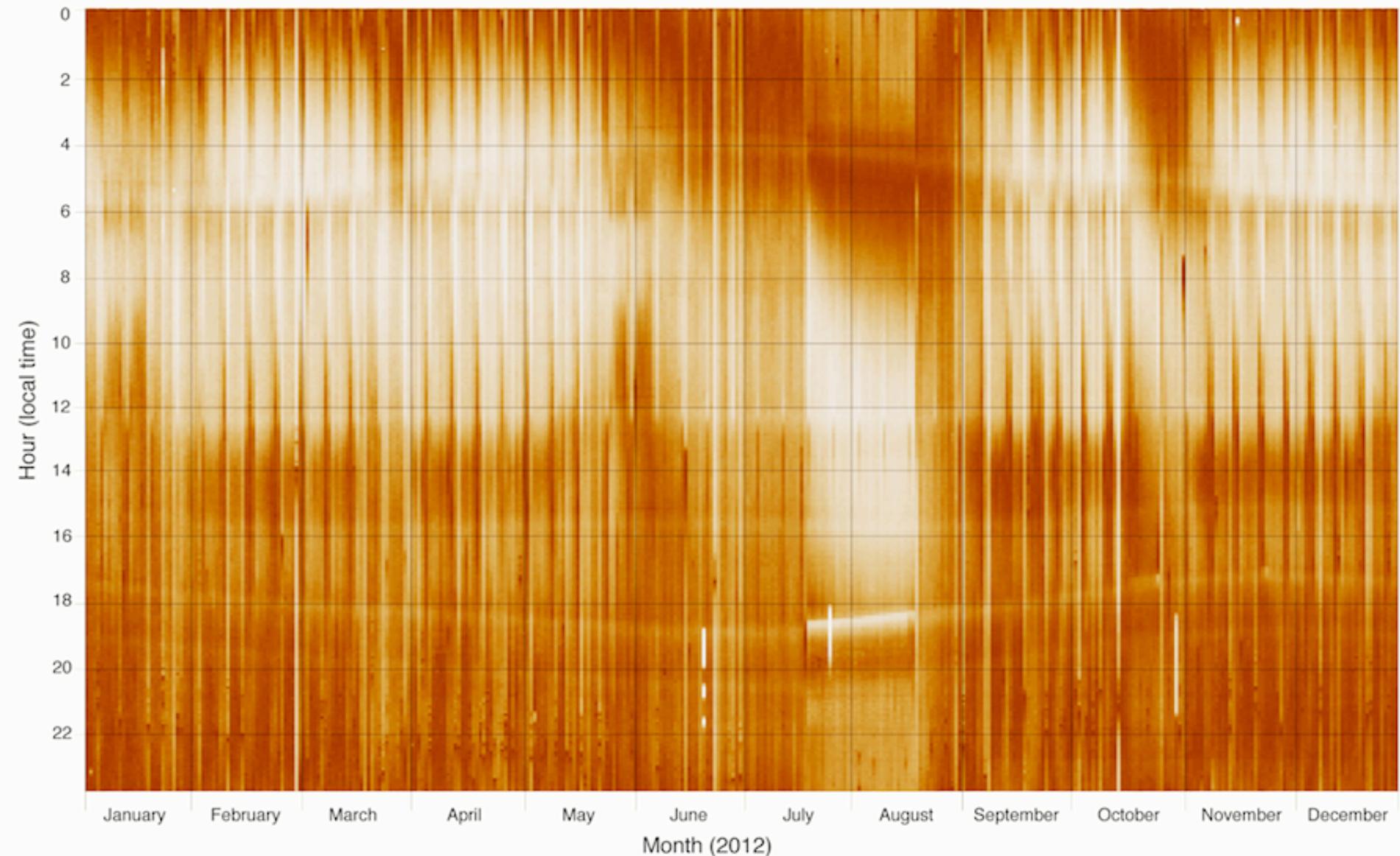
Tweet Activity (New York City, New York, US)



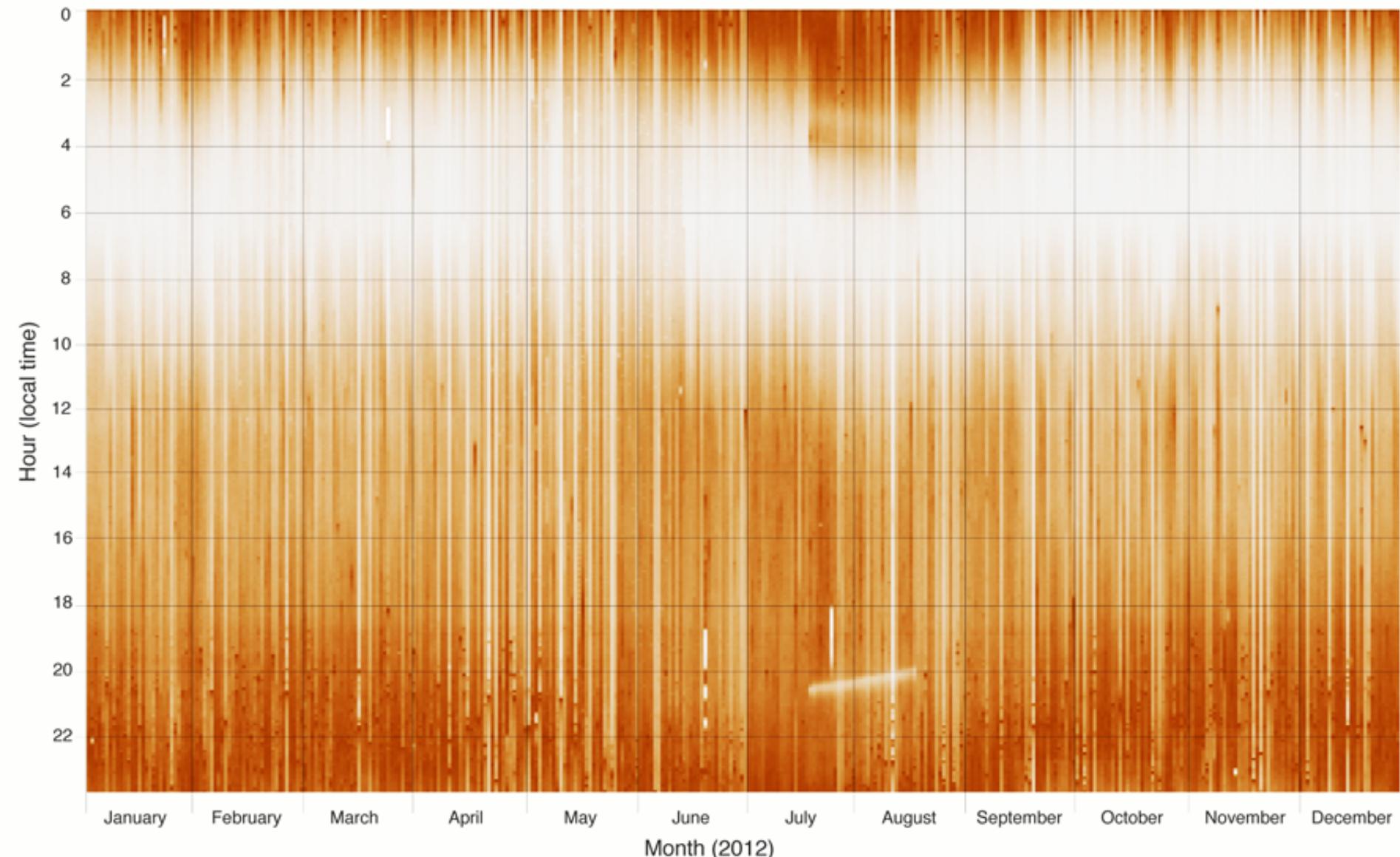
Tweet Activity (Buenos Aires, Argentina)



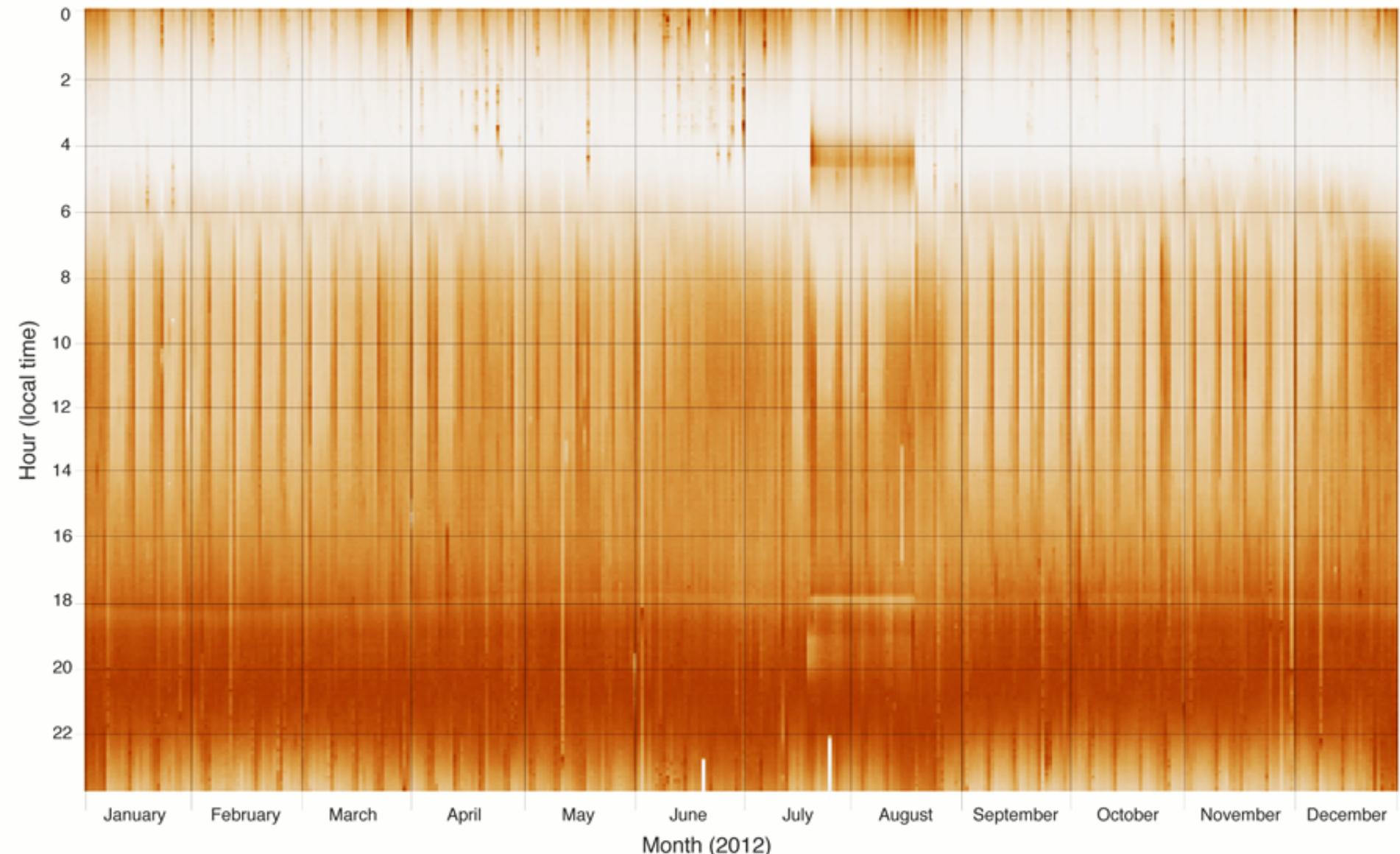
Tweet Activity (Riyadh, Saudi Arabia)



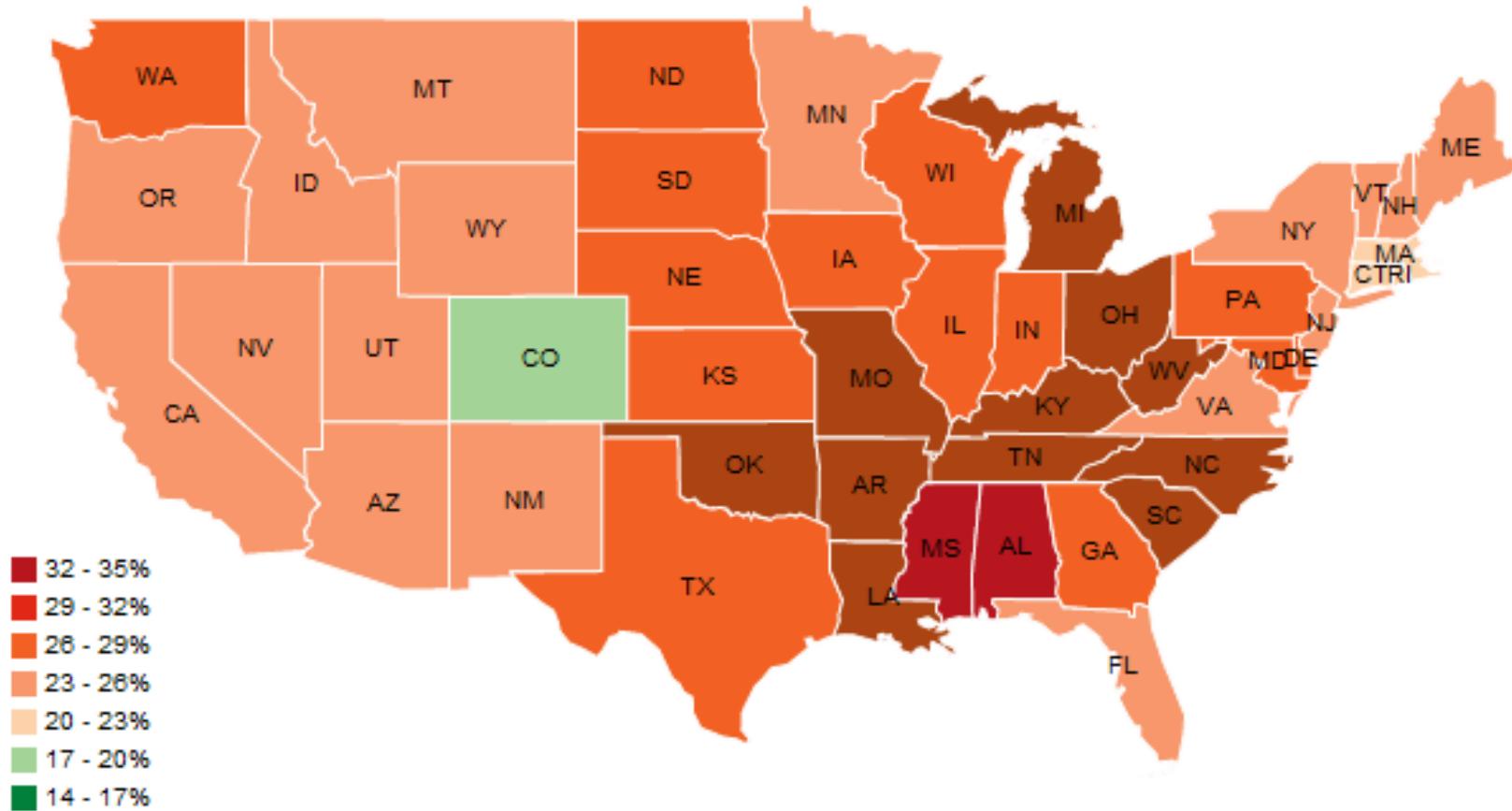
Tweet Activity (Istanbul, Turkey)



Tweet Activity (Jakarta, Indonesia)

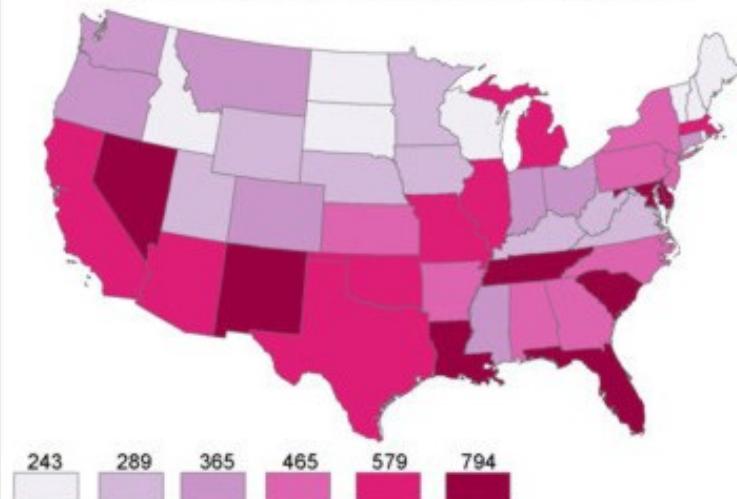


Choropleth Maps

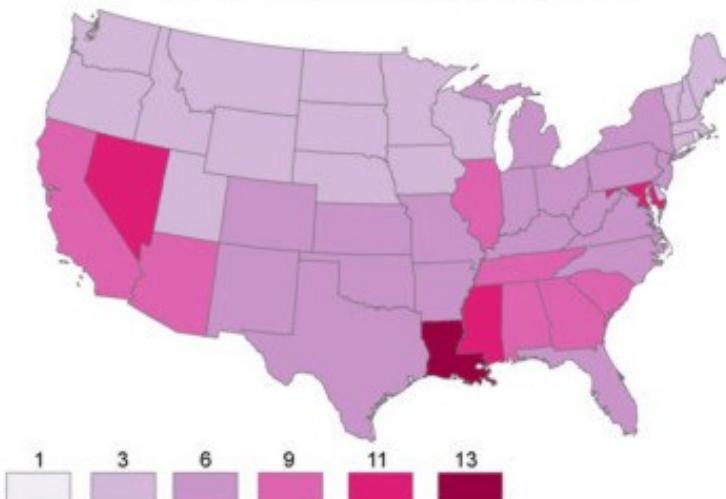


Crime Rates in the US - 2003

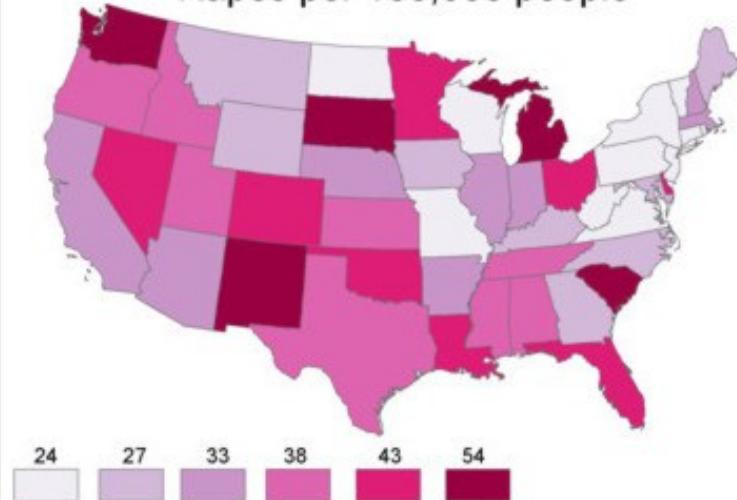
Violent crimes per 100,000 people



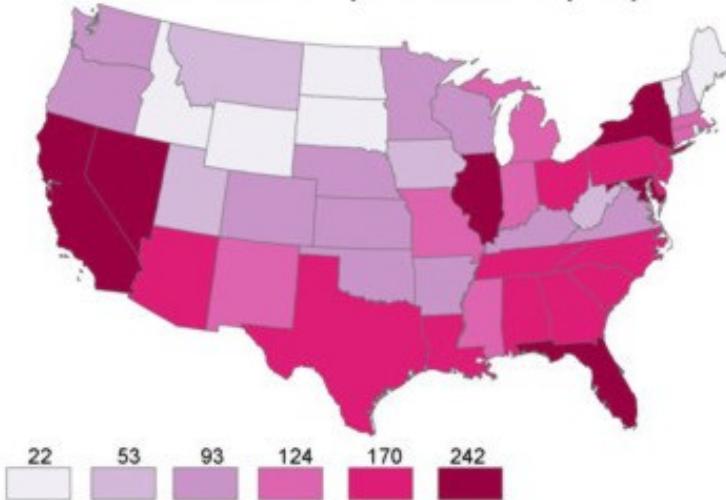
Murders per 100,000 people



Rapes per 100,000 people

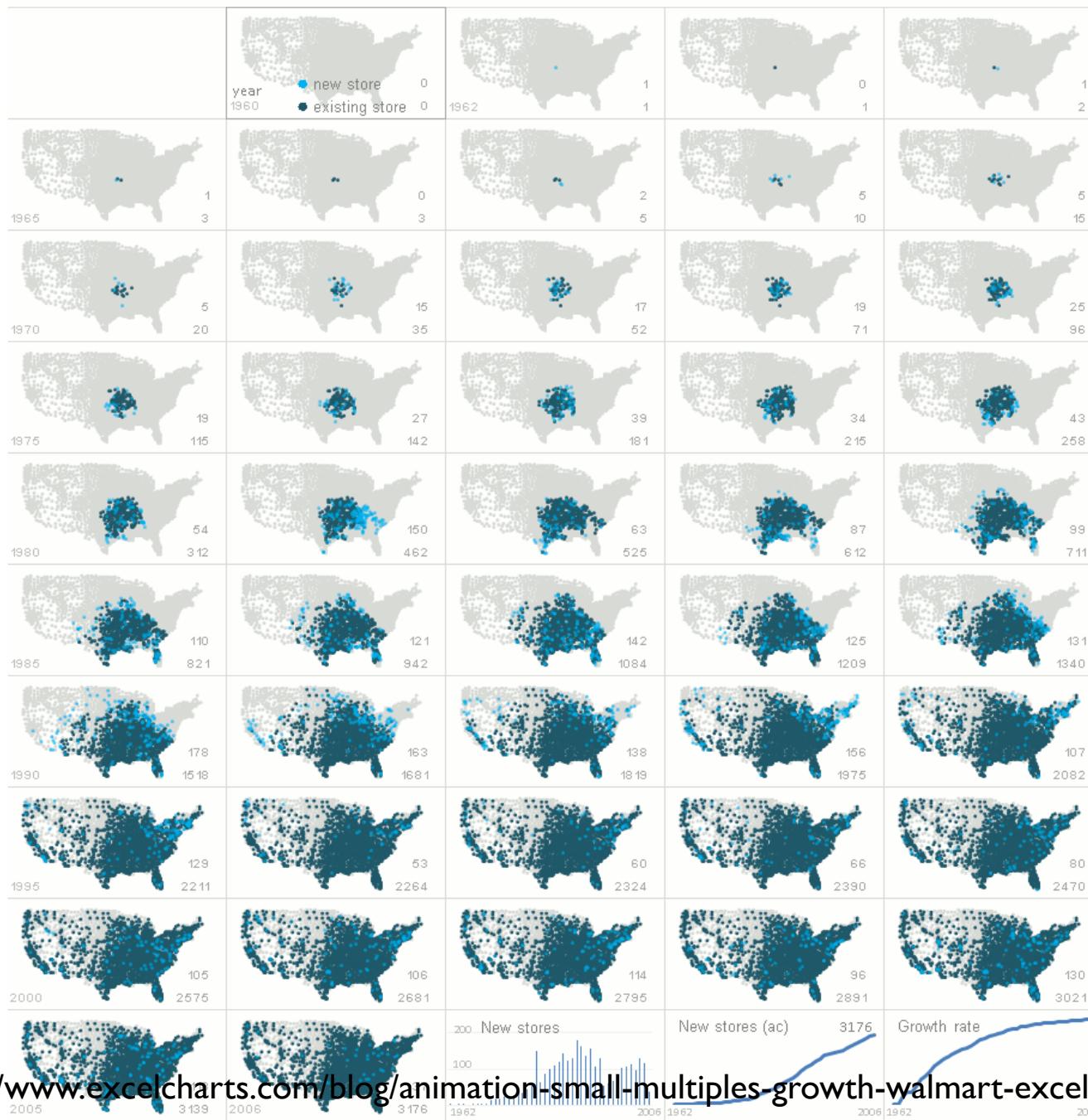


Robberies per 100,000 people



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

Growth of Walmart

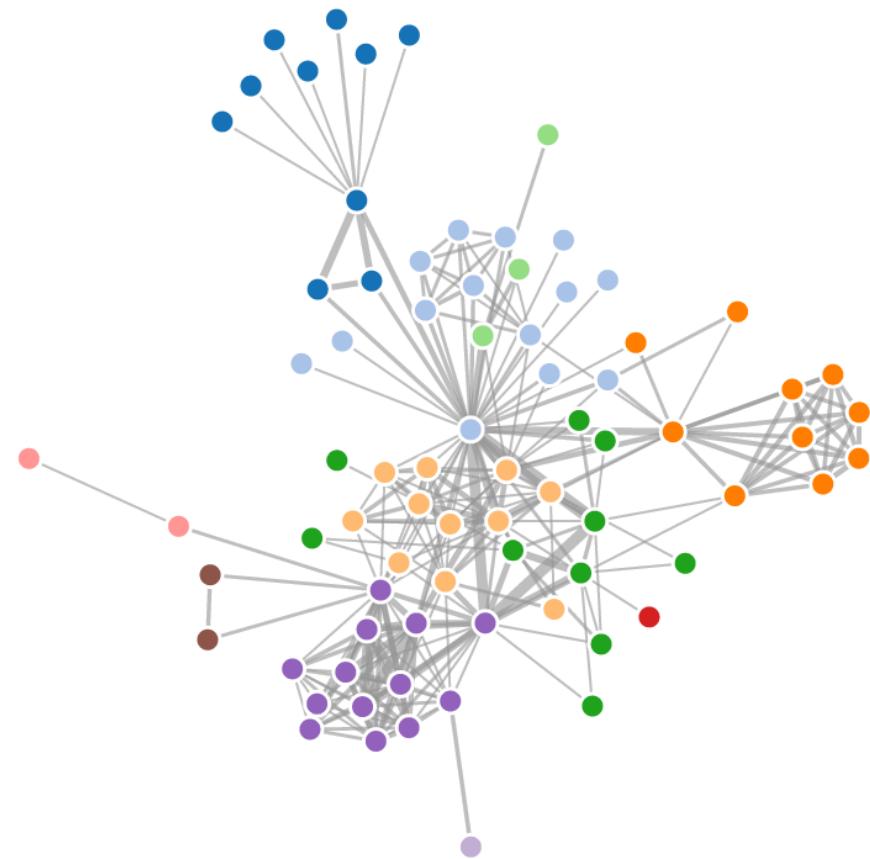


Small Multiples vs. Animations

- 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

