

# **Information Visualization**

## **Part 1**

Dianna Xu  
Deepak Kumar

# BIG DATA

- Data intensive computing
  - capture
  - curation
  - storage
  - search
  - sharing
  - analysis
  - visualization



# Data Science



@chl  
Christian Langreiter

data scientist: statistician who lives in select regions of california or works at bitly

3 hours ago via web ☆ Favorite ↻ Retweet ↩ Reply



# Data Science



A screenshot of a tweet from the account **CS News Update** (@CSNewsUpdate), posted 17 hours ago. The tweet text reads: "EMC World: Tucci sees 'data science' overtaking computer science as top IT skill - V3.co.uk bit.ly/Jjfw0n". Below the text are interaction options: "Collapse", "Reply", "Retweet", and "Favorite". The tweet shows 4 retweets and 1 favorite. A row of five profile pictures is visible below the statistics. At the bottom, it says "9:54 PM - 21 May 12 via dlv.it - Details".

**CS News Update** @CSNewsUpdate 17h  
EMC World: Tucci sees 'data science' overtaking computer science as top IT skill - V3.co.uk bit.ly/Jjfw0n  
Collapse Reply Retweet Favorite

4 RETWEETS 1 FAVORITE

9:54 PM - 21 May 12 via dlv.it - Details

# Data Science

## Data science

---

From Wikipedia, the free encyclopedia

**Data science** defines a discipline that incorporates applying varying degrees of [statistics](#), [data visualizations](#), [computer programming](#), [data mining](#), [machine learning](#), and [database engineering](#) to solve complex [data](#) problems. A practitioner of Data Science is called a Data Scientist.

# What is *Data Science*?

- **The process of using data in the *wild***  
unstructured, unformatted, multiple sources,...
- **Involves**
  - Acquiring (finding and storing)
  - Analyzing
  - Discovering Patterns/Stories
  - Presenting results

# Data Science

**COMPUTER  
SCIENCE**

acquire  
parse

**MATHEMATICS,  
STATISTICS AND  
DATA MINING**

filter  
mine

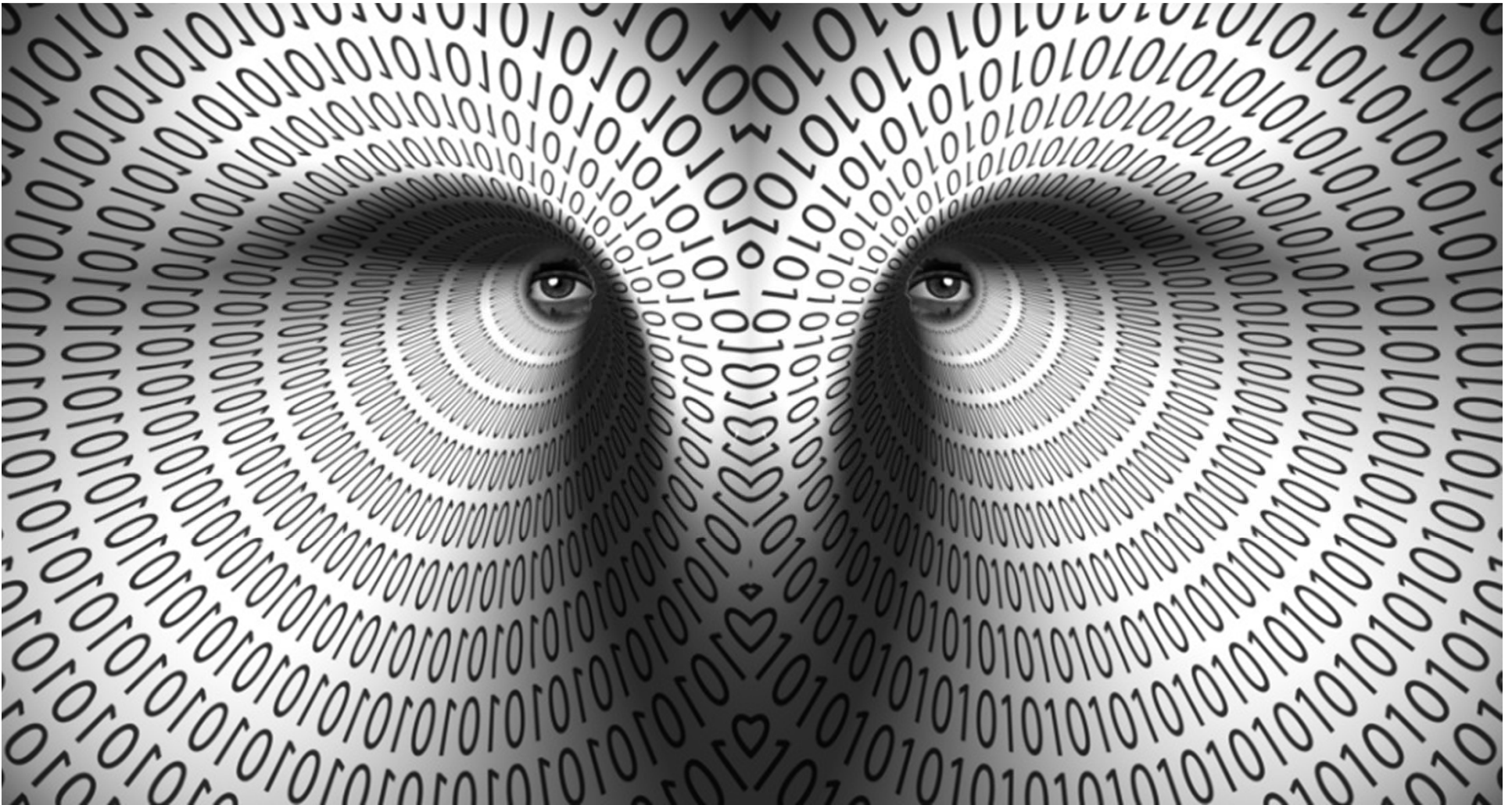
**GRAPHIC  
DESIGN**

represent  
refine

**INFOVIS AND  
HCI**

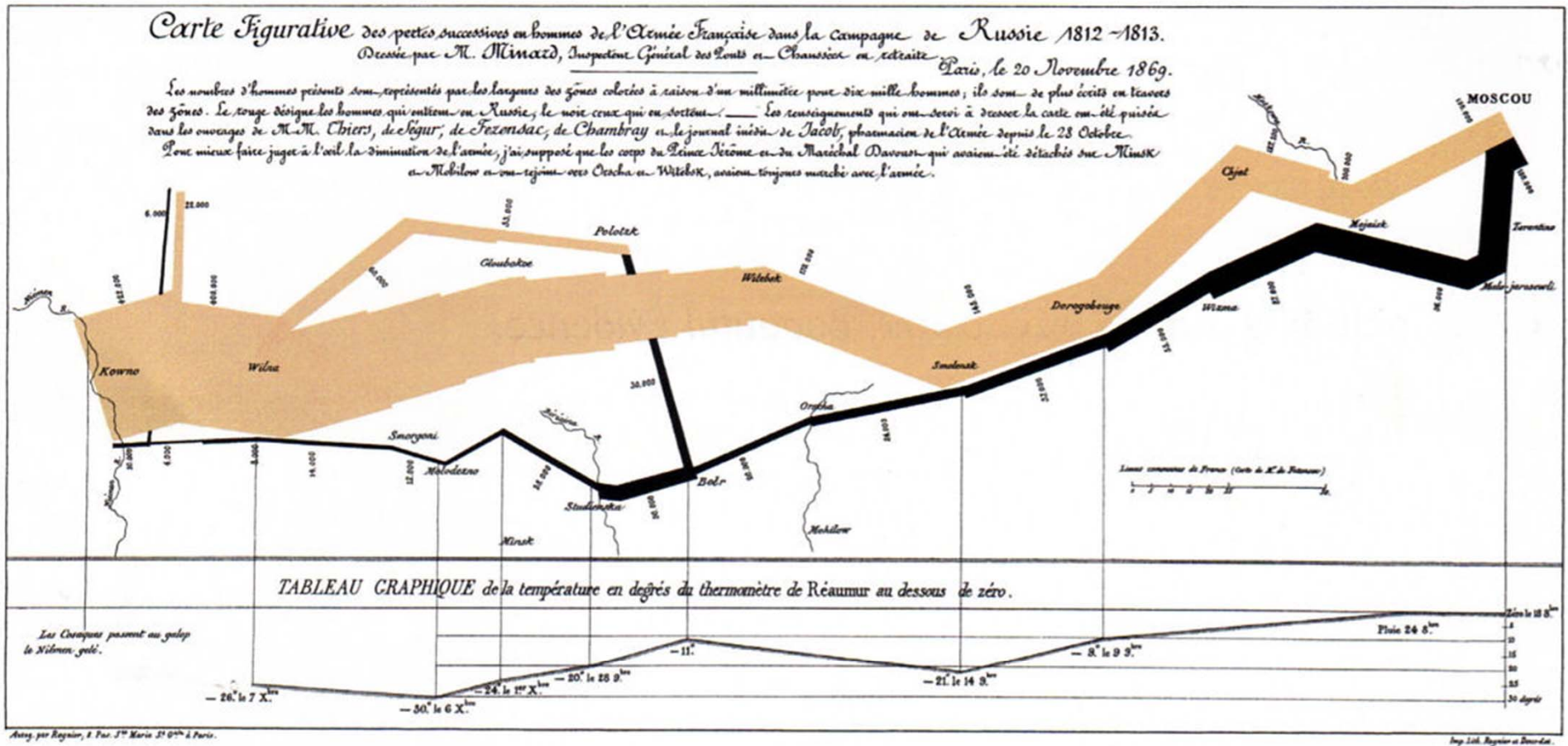
interact

# Visualizing Data





# Information Visualization



This map drawn by Charles Joseph Minard portrays the losses suffered by Napoleon's army in the Russian campaign of 1812. Beginning at the left on the Polish-Russian border near the Niemen, the thick band shows the size of the army (422,000 men) as it invaded Russia. The width of the band indicates the size of the army at each position. In September, the army reached Moscow with 100,000 men. The path of Napoleon's retreat from Moscow in the bitterly cold winter is depicted by the dark lower band, which is tied to temperature and time scales. The remains of the Grande Armée struggled out of Russia with 10,000 men. Minard's graphic tells a rich, coherent story with its multivariate data, far more enlightening than just a single number bouncing along over time. Six variables are plotted: the size of the army, its location on a two-dimensional surface, direction of the army's movement, and temperature on various dates during the retreat from Moscow. It may well be the best statistical graphic ever drawn.

# Big Data

- The amount of available data is quickly outpacing our ability to understand and use it in meaning full ways.
- The eventual solution to big data is likely really good black boxes.
- Black boxes only give answers if you know right questions to ask.

# Visualization and Explorative Data Analysis

- Visualization allows one to view things in global context.
- Increase the likelihood of spotting unexpected trends and perspectives.
- Allows new questions to be asked.
- The life cycle of explorative data analysis with visualization is likely an iterative process

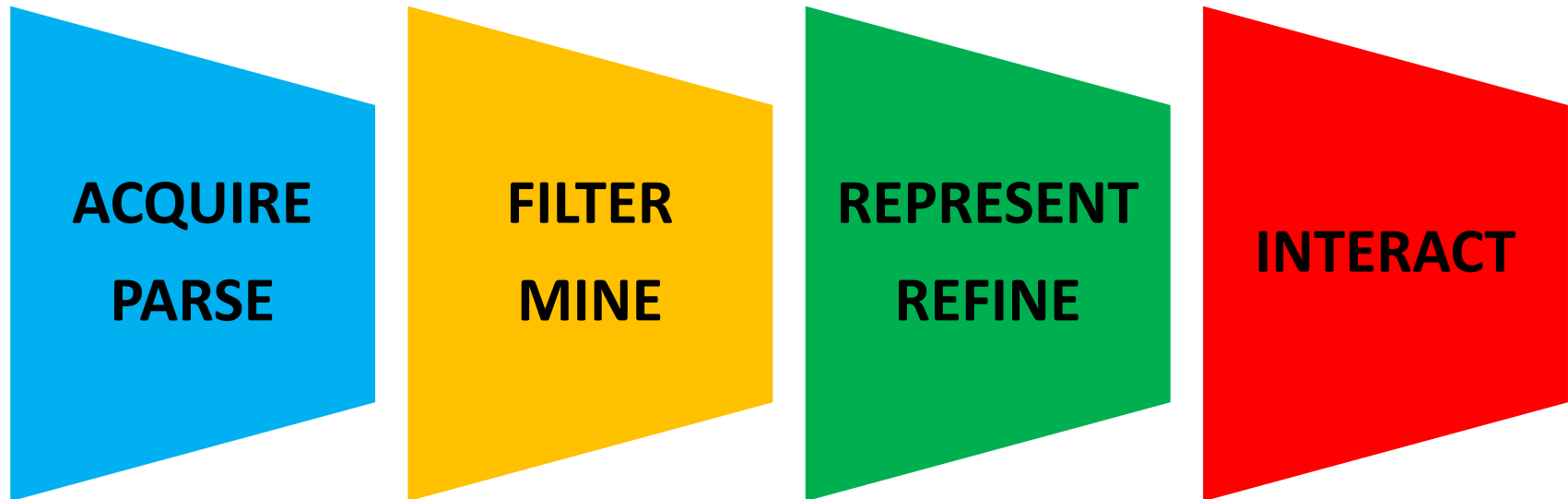
# Visualization Types

- Scientific visualization
  - numerical, high precision and has relative simple spatial relationships
  - accurate representation of numerical proportions and realistic rendering of physical properties
- Data visualization
  - social/economical data
  - highly categorical and have strong association with physical locations/coordinates on the map

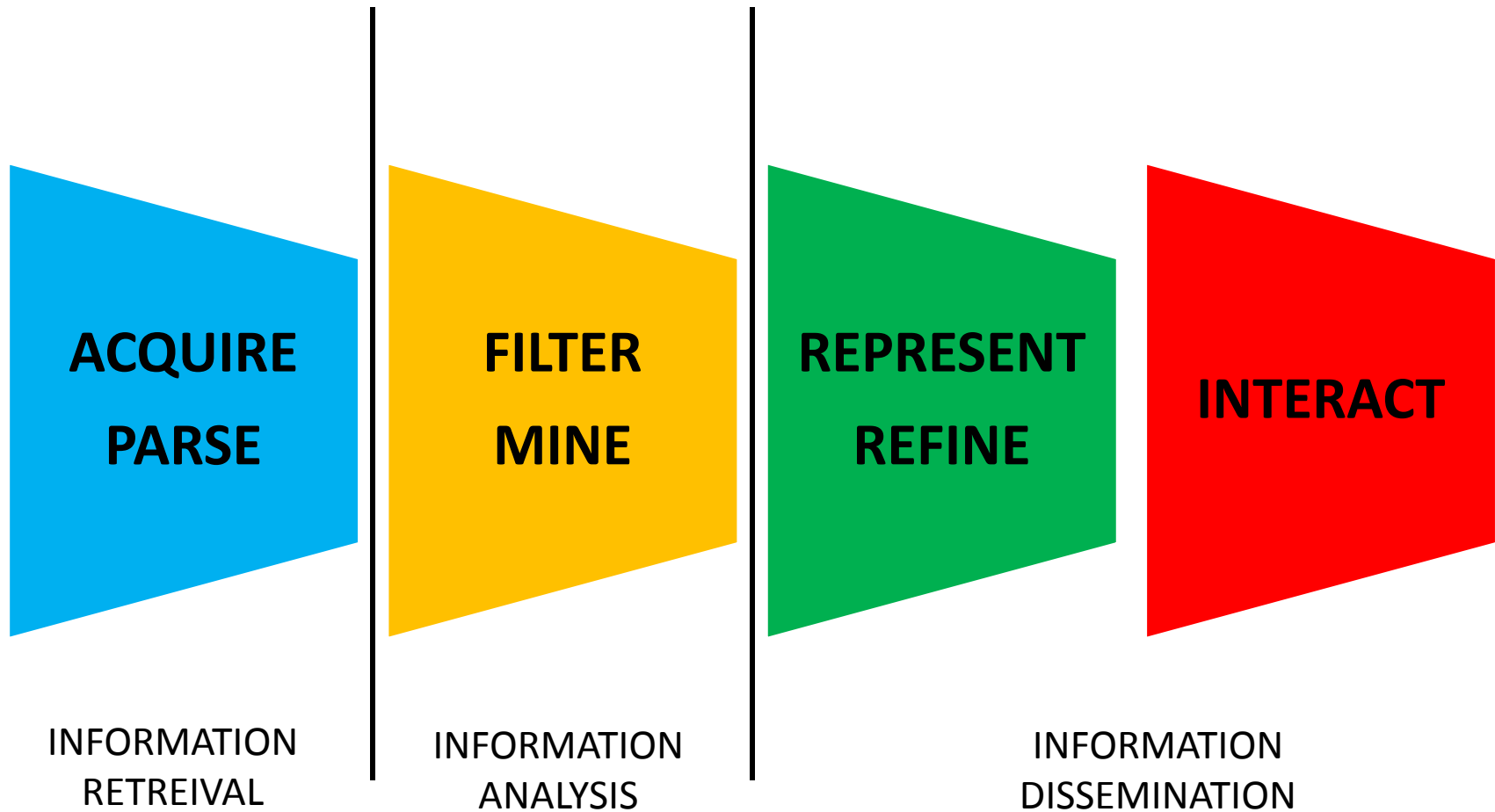
# Information Visualization

- Arbitrary or complex relationships with no clearly prescribed spatial representation choices
- Abstract representation techniques
- Communication is an important goal
- *infoviz* versus *data journalism*

# The Visualization Process



# The Visualization Process

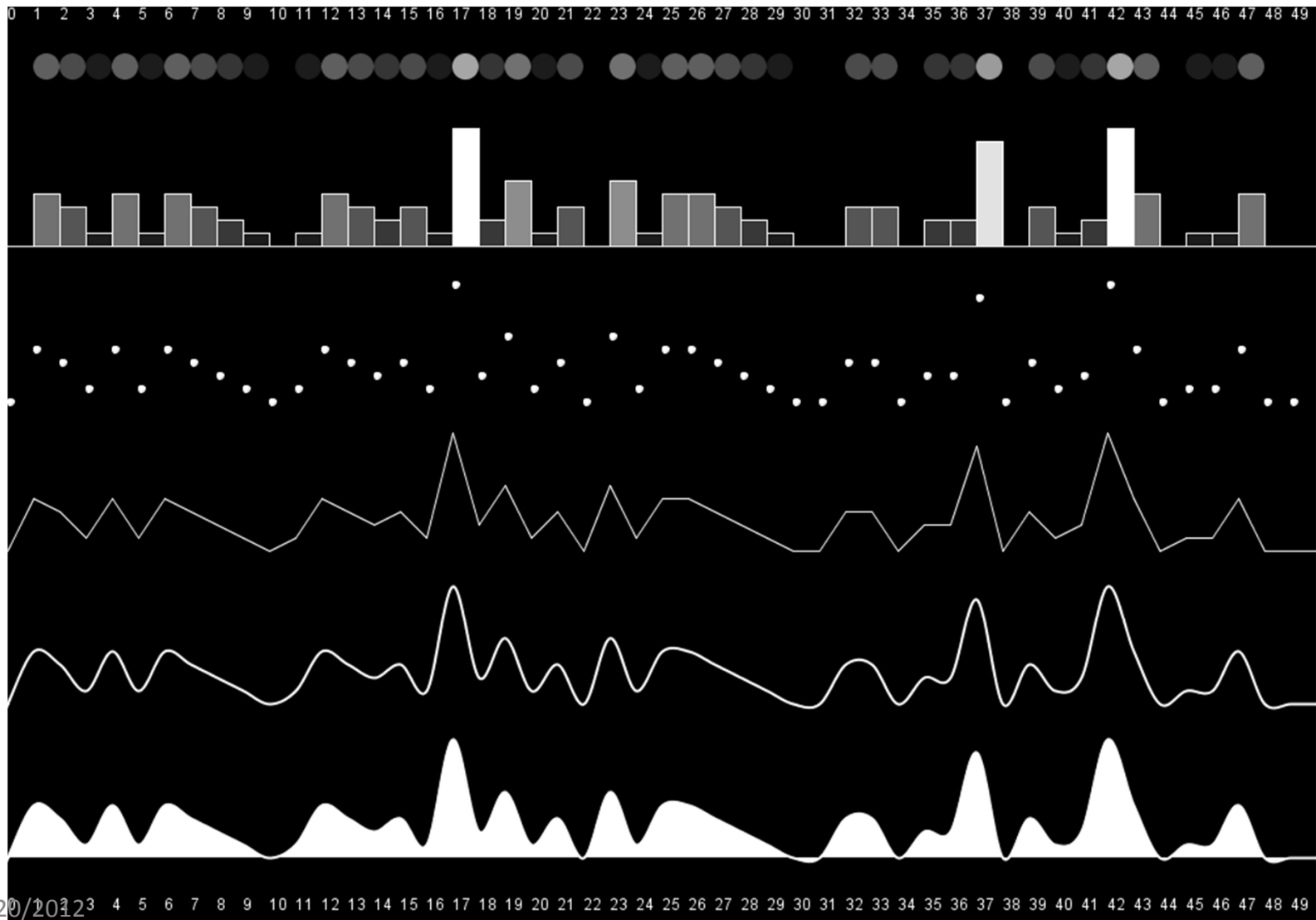


# Example: Mapping Numbers

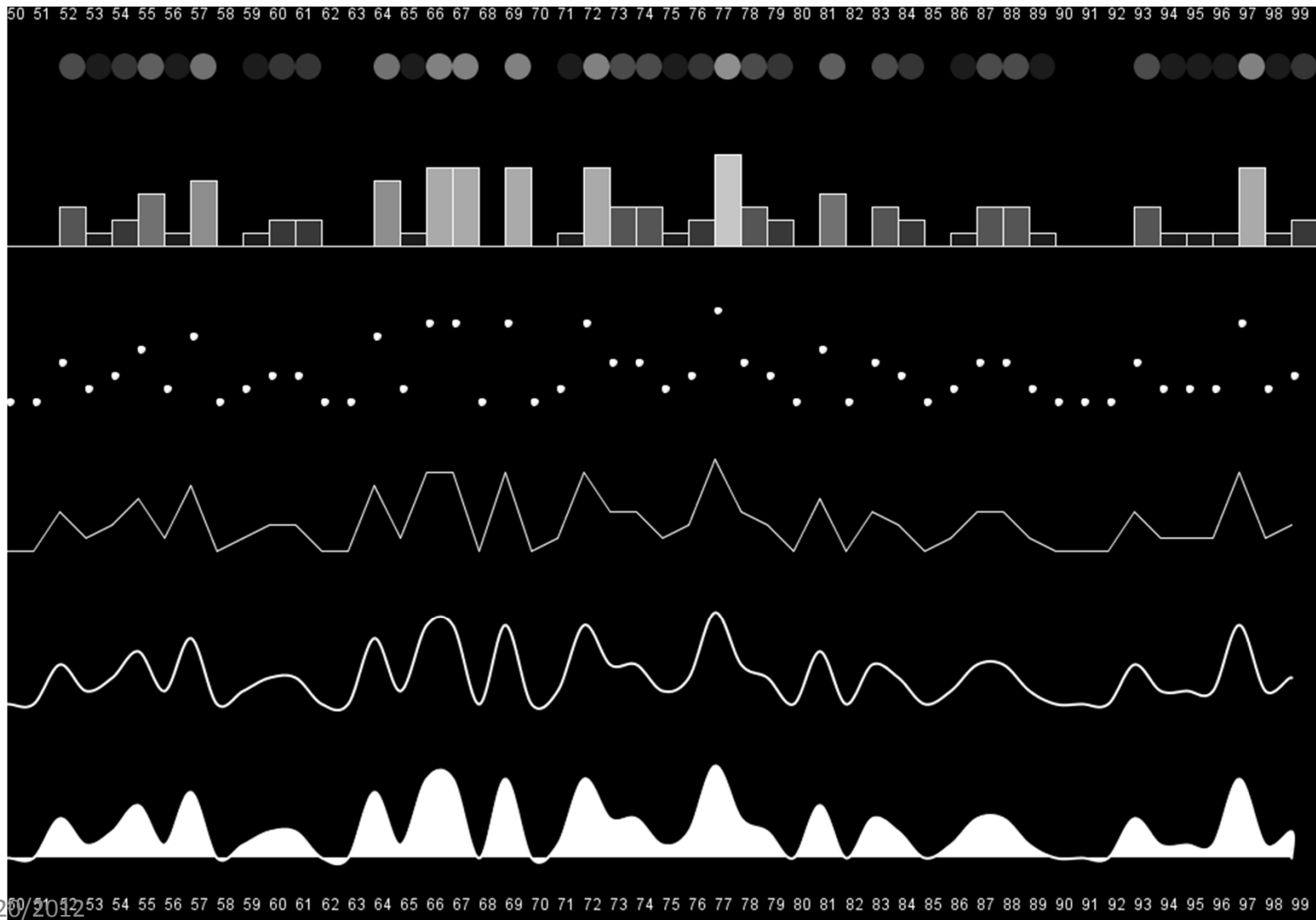
- How to represent numbers with visual primitives so that relative values of numbers can be quickly deduced from visual cues
- A simple data set
  - 255 integers valued 0-99
  - collected on twitter as random numbers



# Mapping Numbers



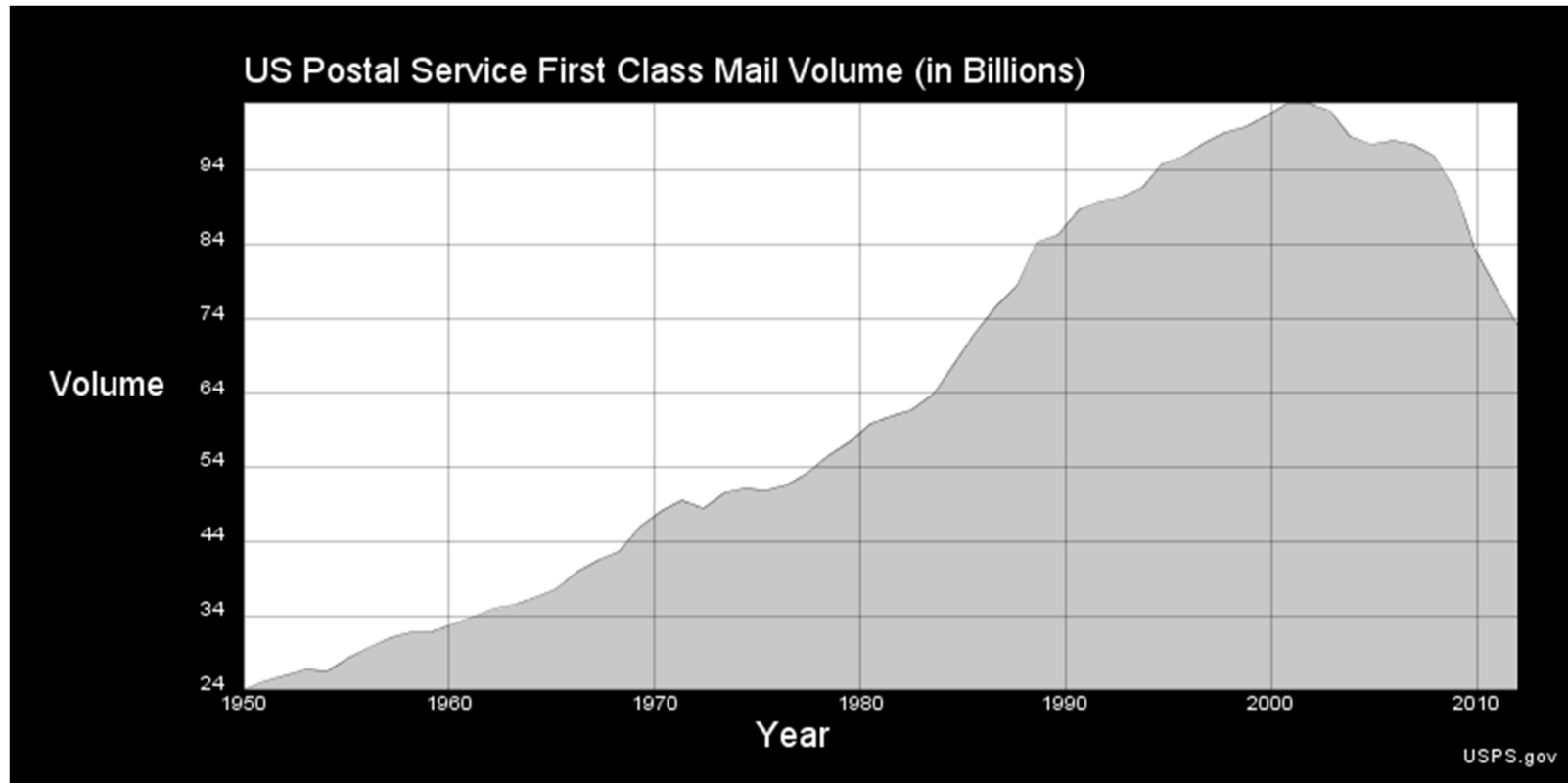
# Mapping Numbers



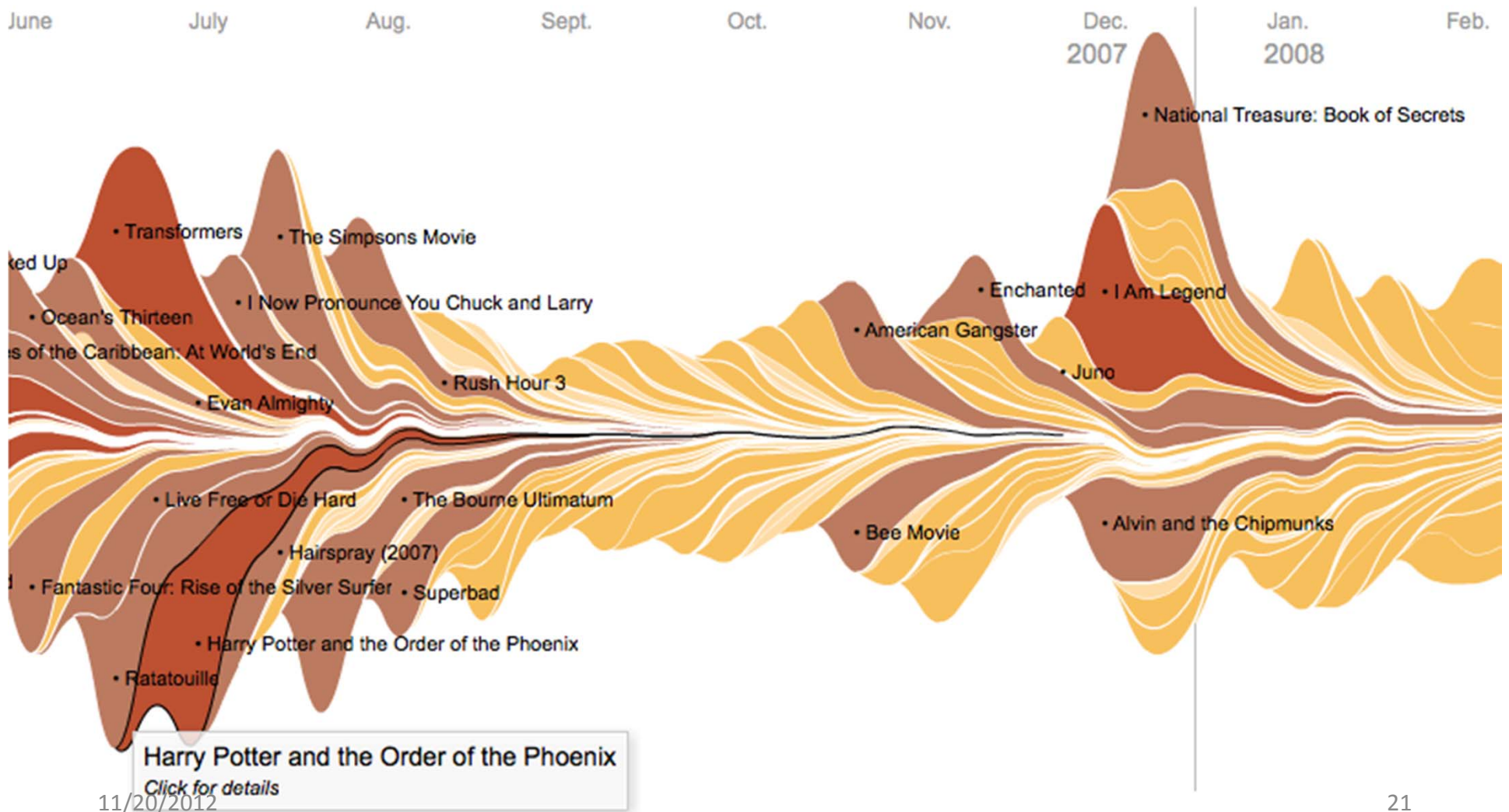
# Information Dimensions

- A fundamental challenge of visualization design is to add more information dimensions without introducing clutter
- Display is 2D
- Going 3D is not a solution

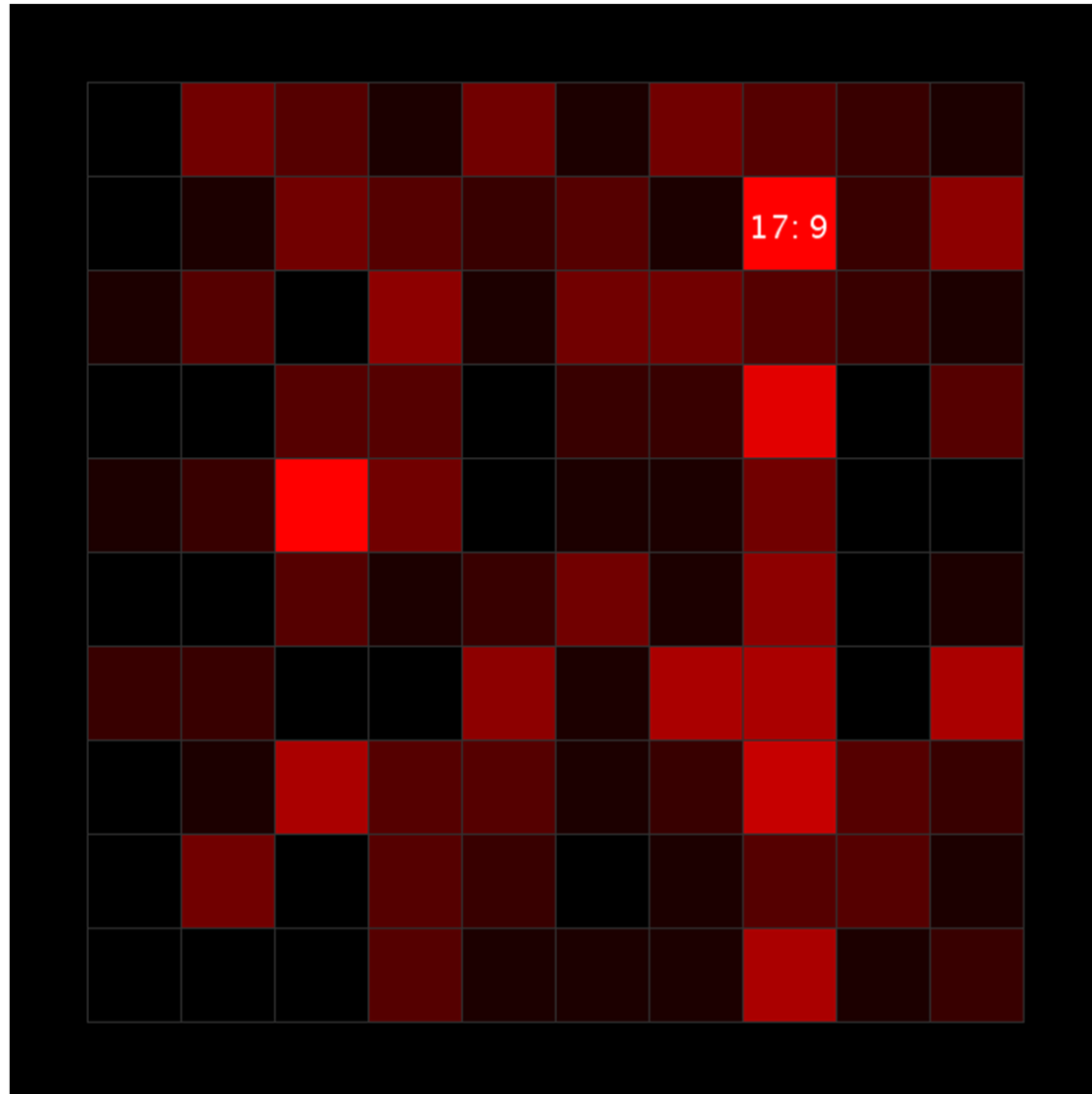
# 1. Time Series



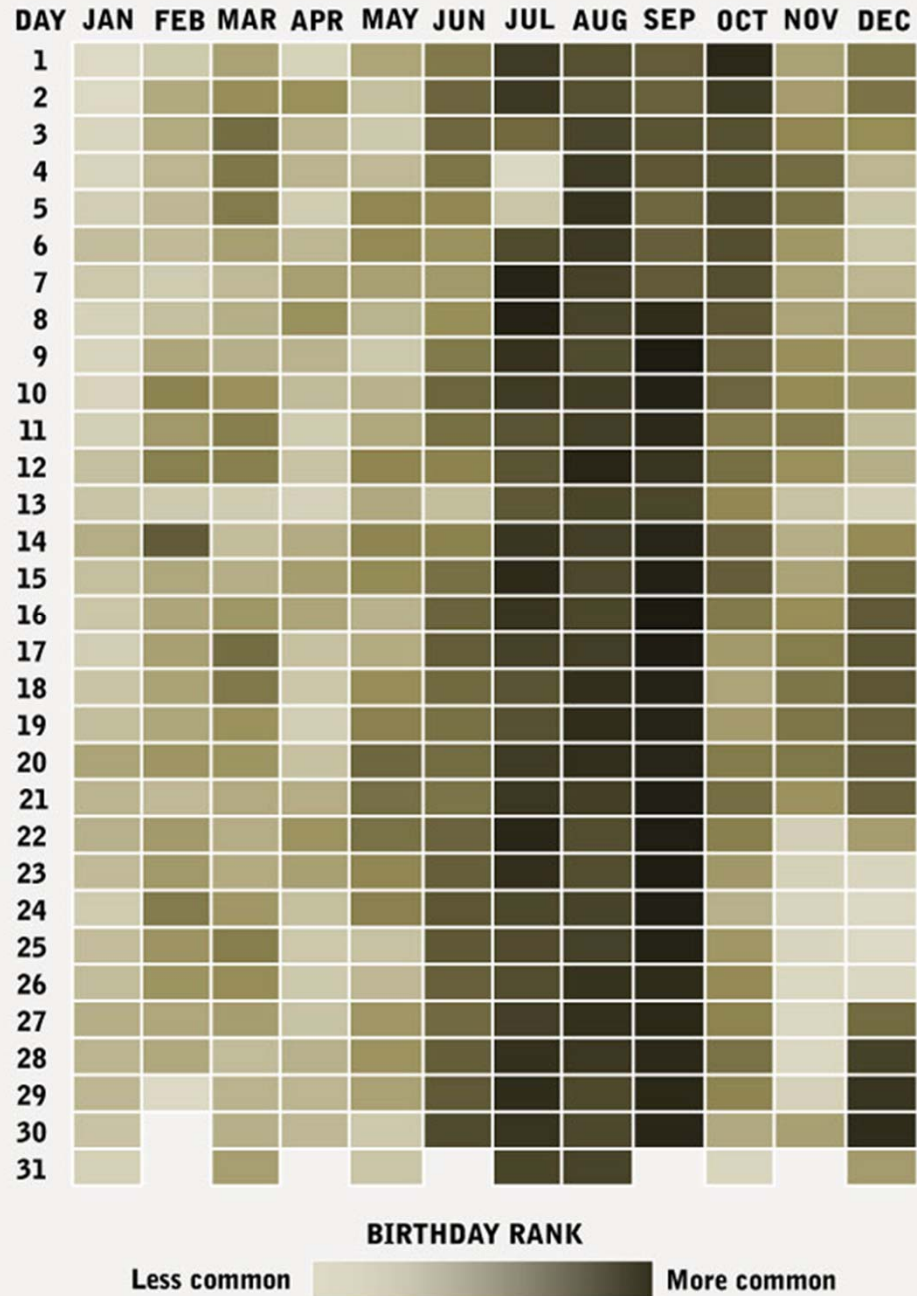
# 2. Stacked Graph



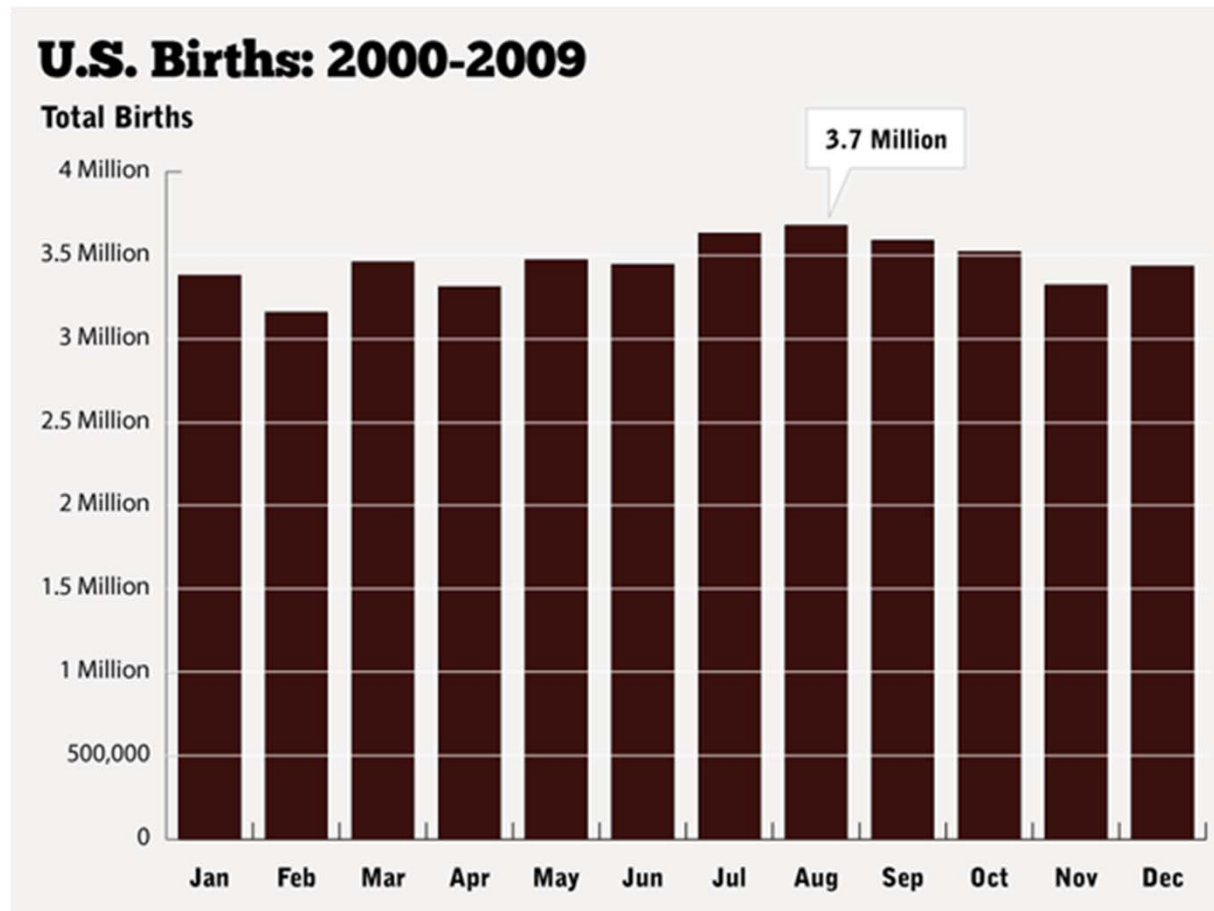
# 3. Heat Map



# Which Birth Dates Are Most Common?

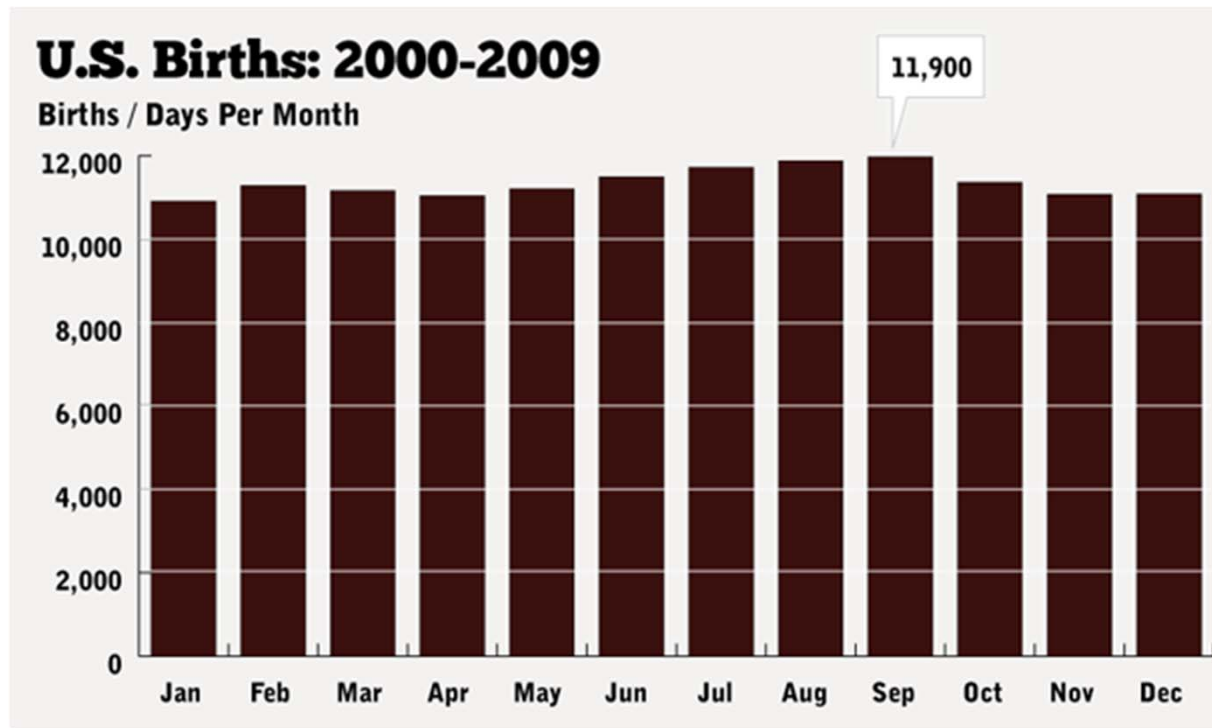


# Looking Deeper...



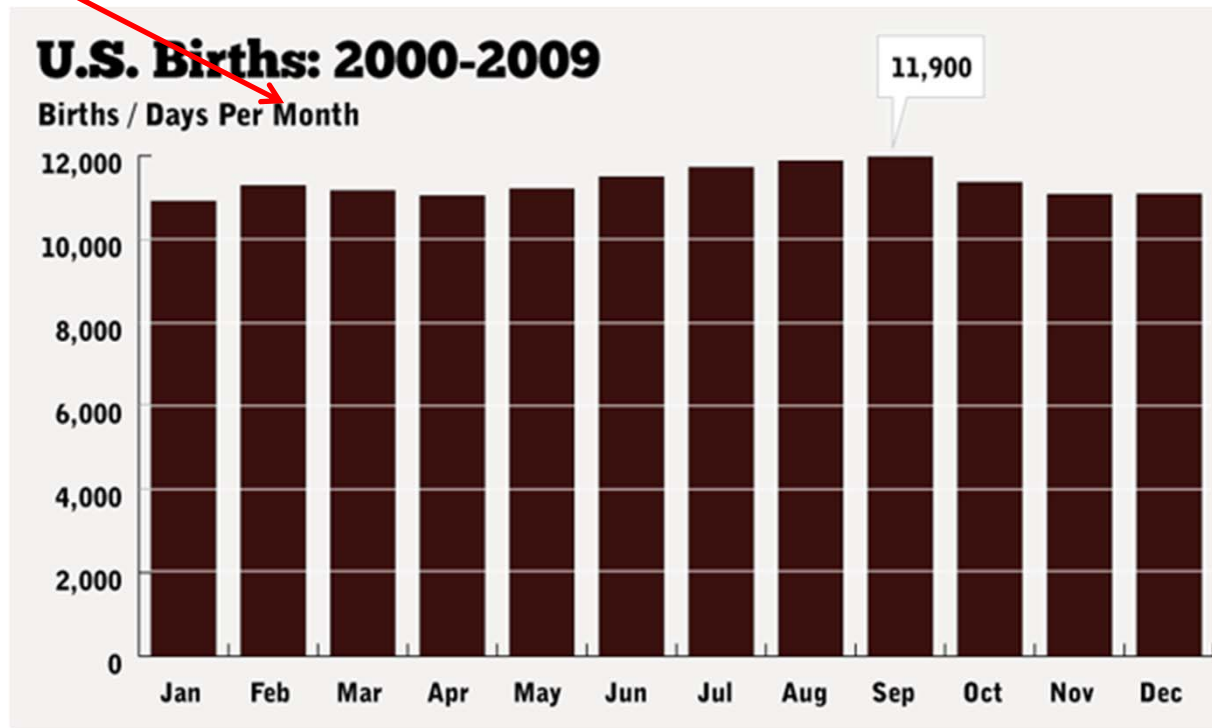


# Looking Deeper...

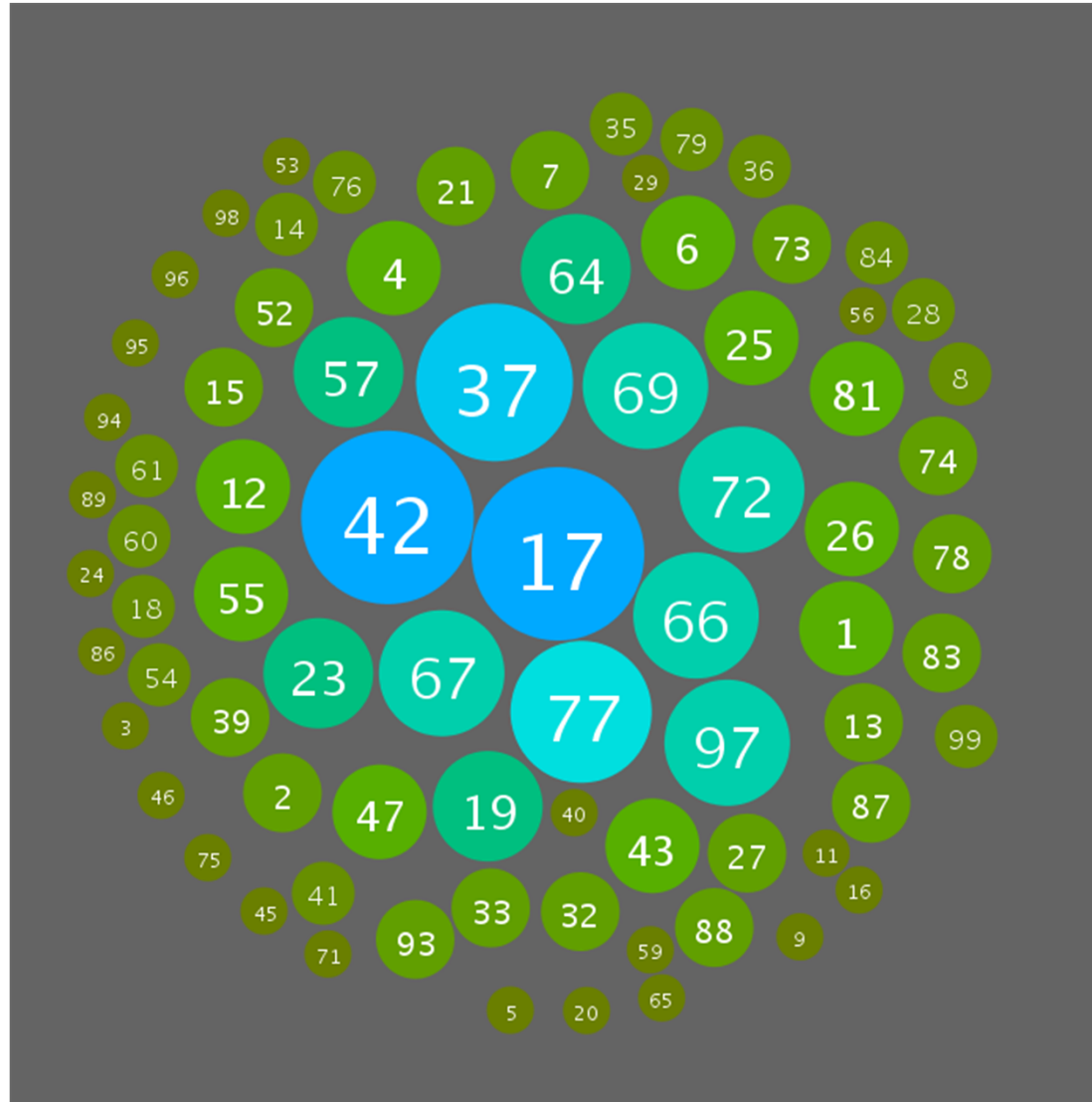


# Caution: There are errors here!

That should be  
Days Per Month  
Per Year!



# 4. Proportional Symbols



# 5. Word Clouds

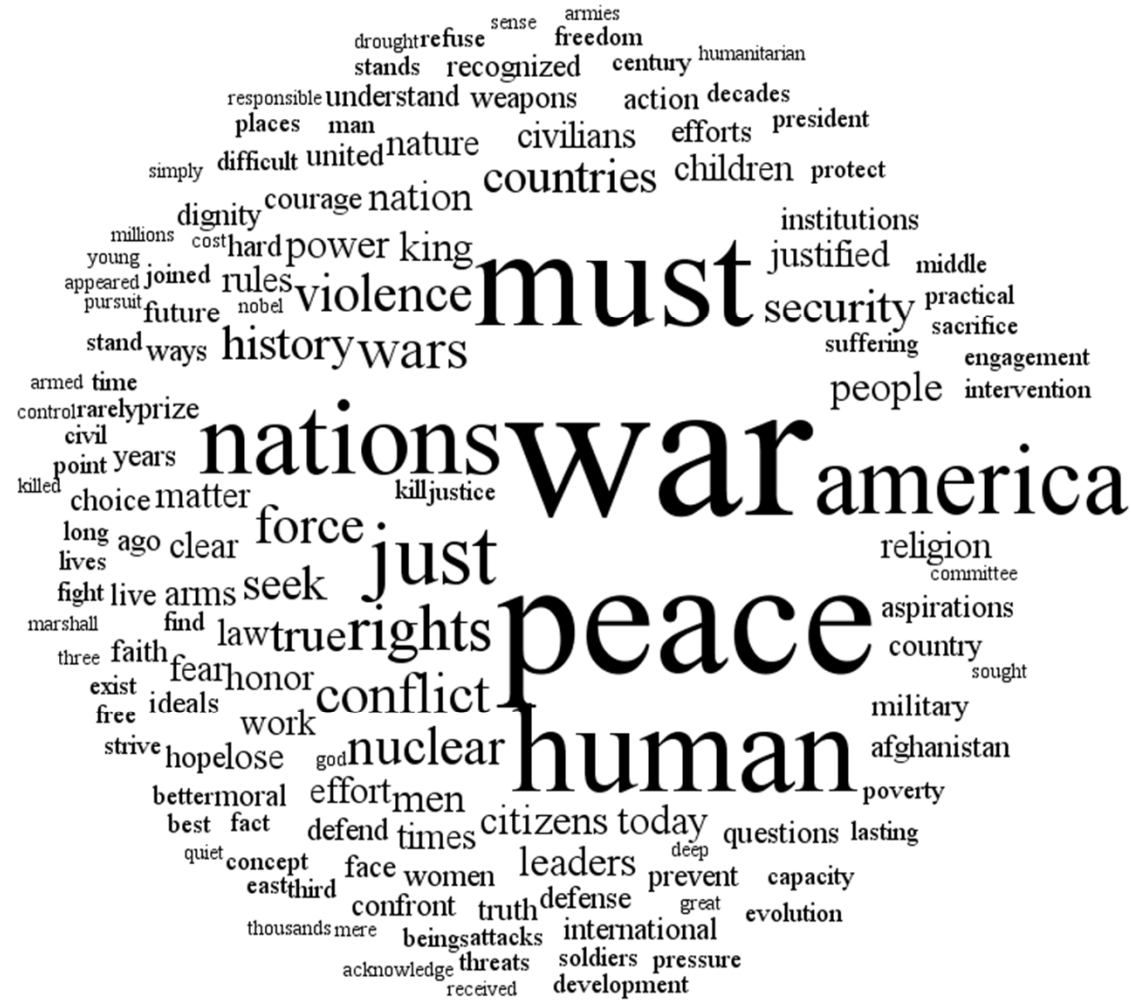




# Word Clouds: Spatial Packing

- Randomized placement allowing overlap
- Randomized greedy algorithm
  - fit the largest tiles first at randomly chosen location
  - if there is overlap, try again
- Exhaustive search
- Exhaustive search guided by space filling curve
  - Spiral, Peano, Hilbert, etc

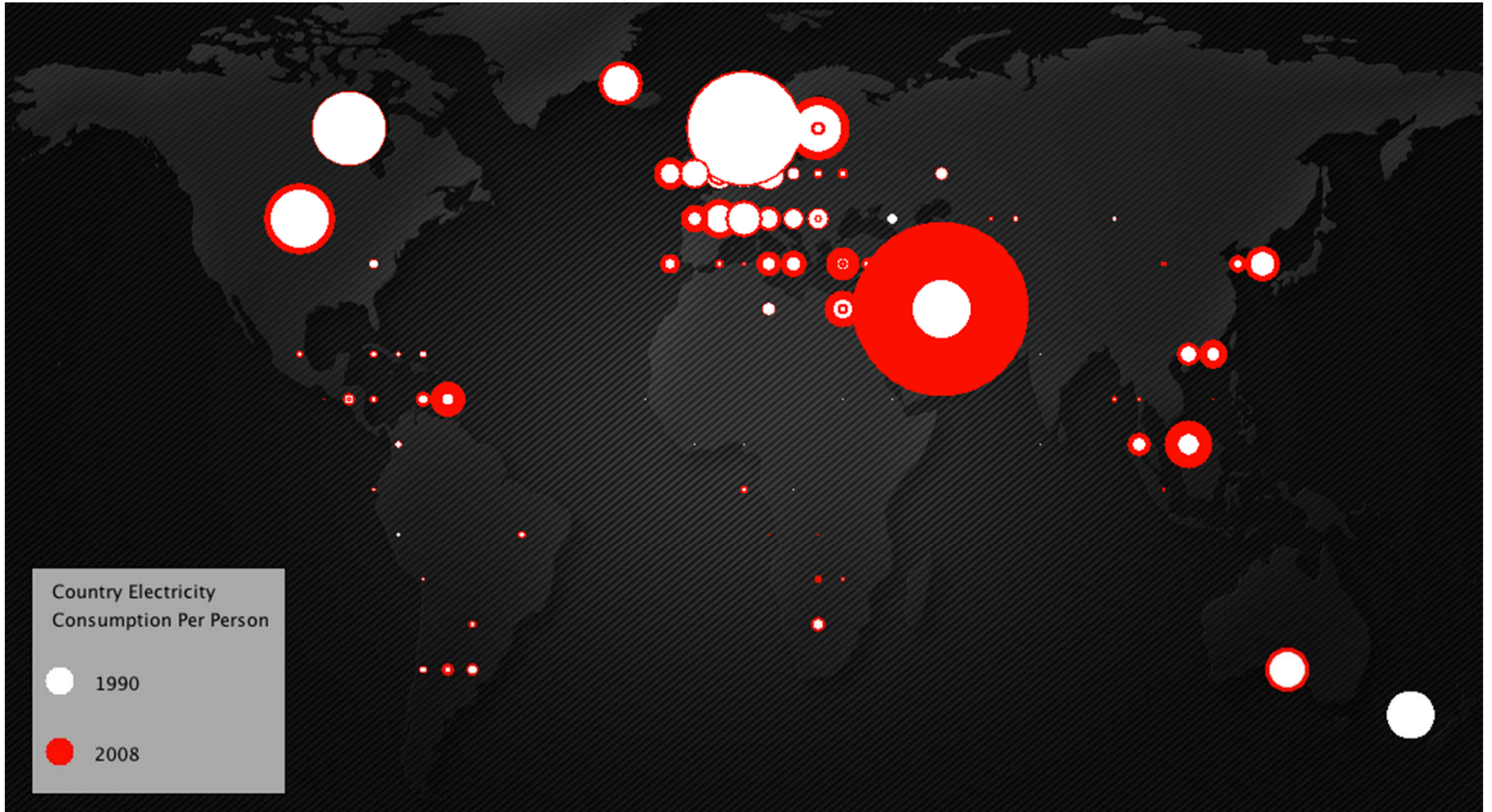
# Word Clouds: Spiral Packing







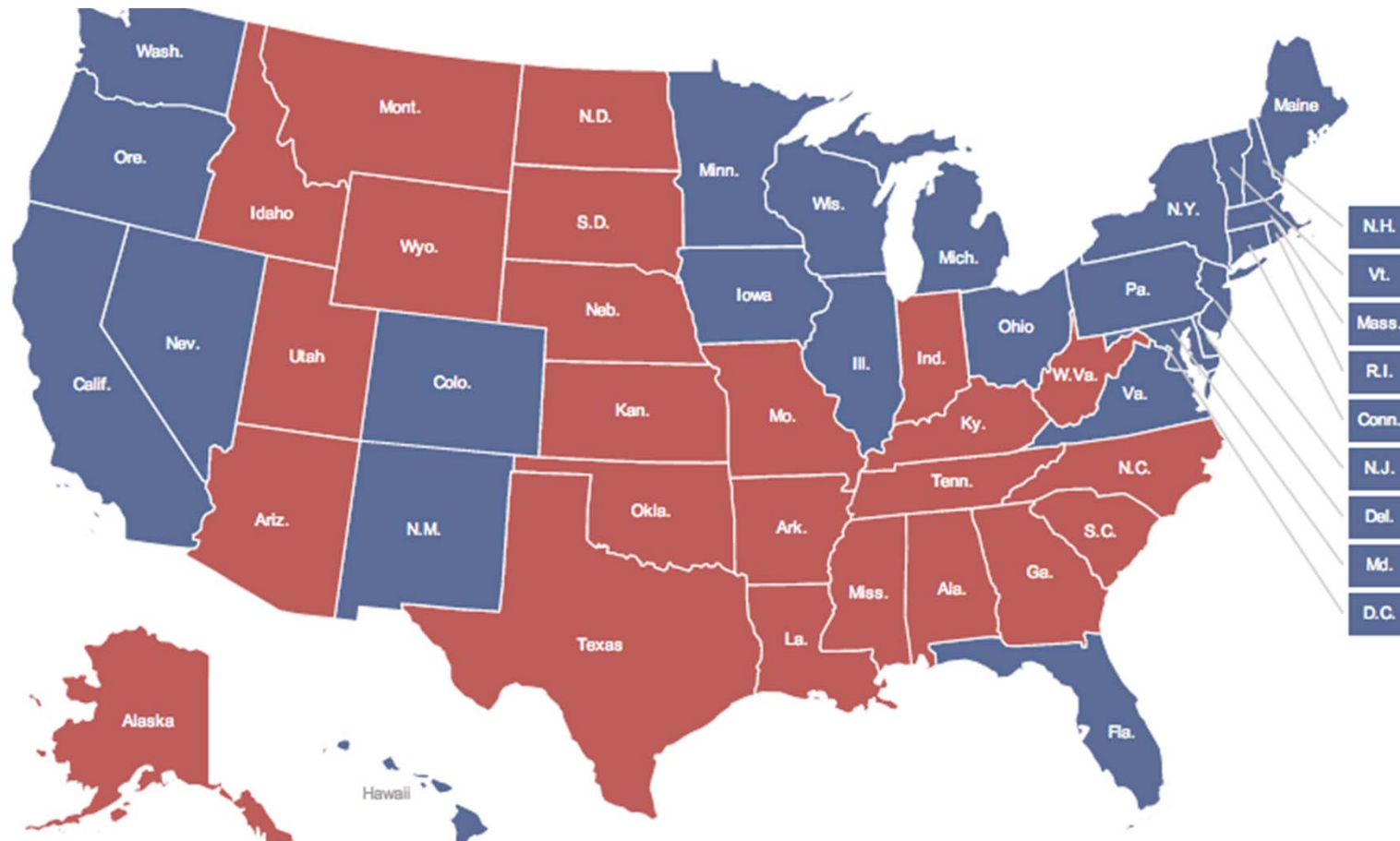
# 6. Map-based



# Coordinate Conversions

- Translate physical coordinates to drawing coordinates
- Geographical coordinates
- Modern world map employing a Universal Transverse Mercator (UTM) projection
- What if your data comes with names of countries or states?

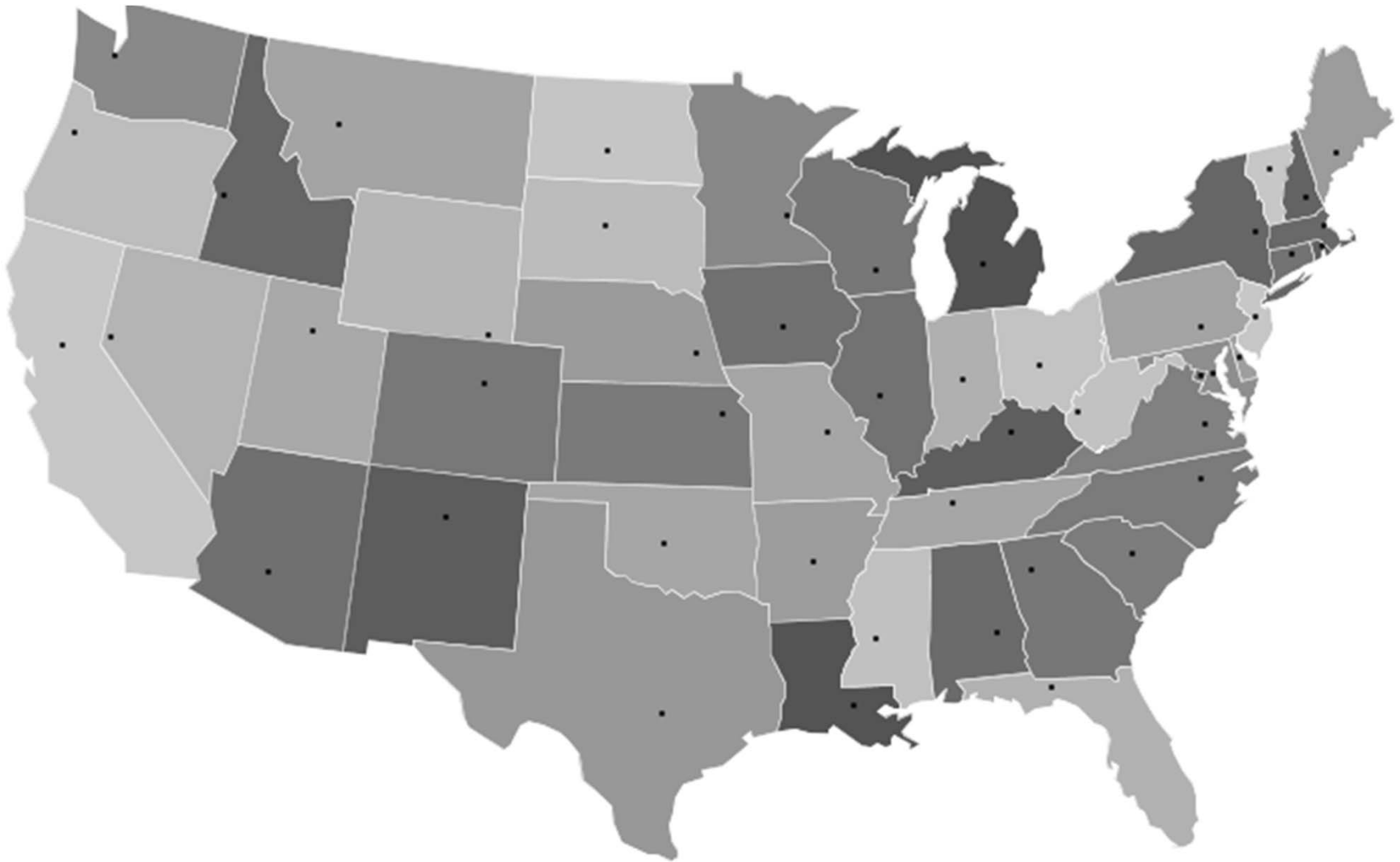
# 7. Choropleth Maps



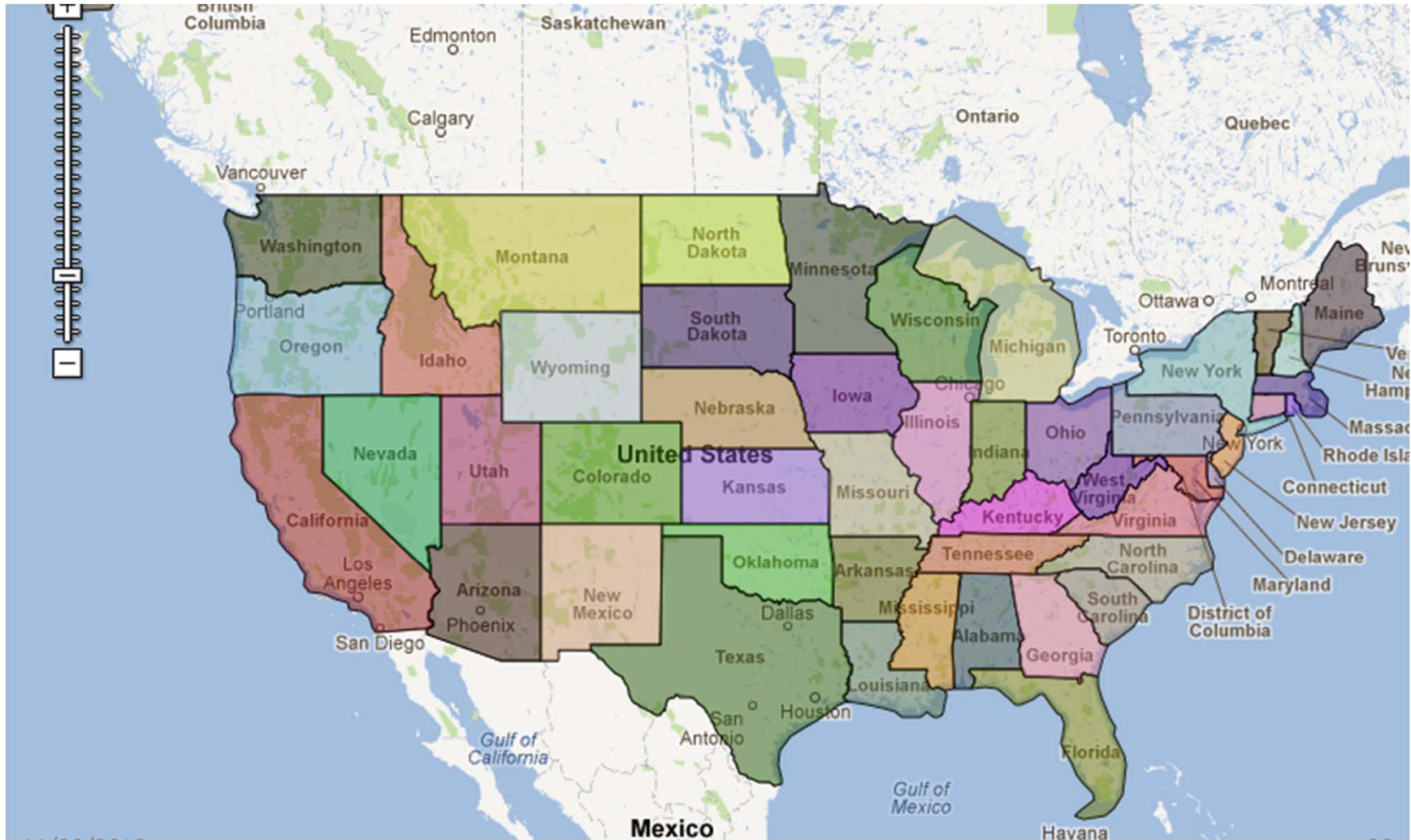
# Choropleth Maps

- Thematic maps require that you color an entire area (state) with a particular color
- Need polygonal outlines of a state
- A US political map of SVG format contains coordinates as an XML file
- Combine with Google Maps API to create interactive map overlay

# States as Polygons

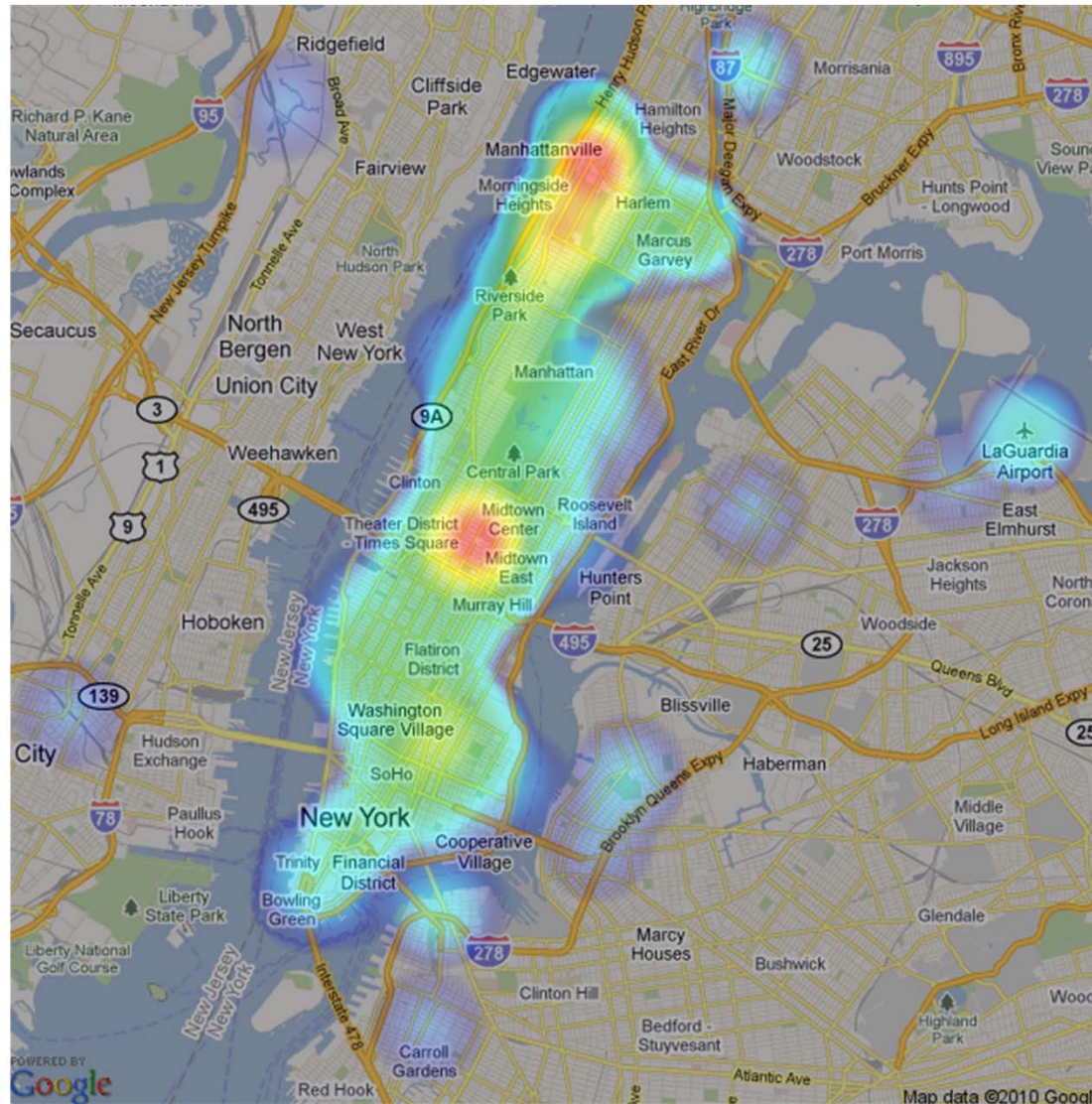


# Google Map Overlay



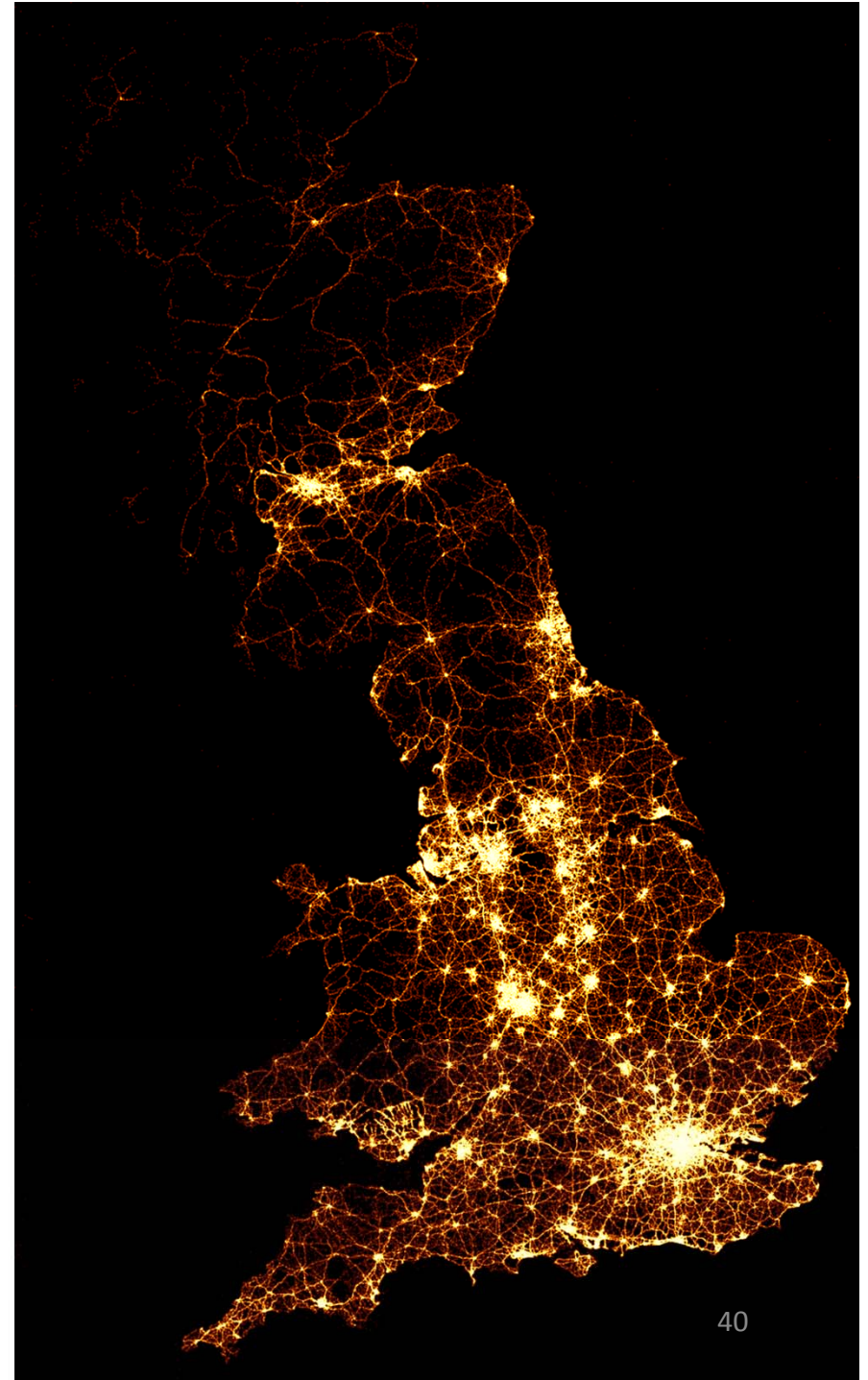


# Foursquares Check-ins



# UK Traffic Accidents 1999-2010

- Location of 2,396,750 road crashes.
- Each light point is an individual collision which resulted in a casualty.
- The intensity of brightness shows where collisions are more frequent.





# References

- Thorp, Jer: Your Random Numbers – Getting Started with Processing and Data Visualization
- By Jer | April 11, 2010. <http://blog.blprnt.com/blog/blprnt/your-random-numbers-getting-started-with-processing-and-data-visualization>
- The Data Deluge, *The Economist*, Feb 25th 2010.
- Edward Tufte, *Presenting Data and Information: A One-Day Course*. www.edwardtufte.com
- Ira Greenberg, Dianna Xu, Deepak Kumar, *Processing: Creative Coding and Generative Art*, FriendsOfEd, 2012, forthcoming.
- Nathan Yau, *Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics*, Wiley, 2011.
- Christian Langreiter, Tweet (@chl) at 6:10 AM on September 28, 2011.
- Hal Varian, *On How the Web Challenges Managers*, The McKinsey Quarterly, January 2009.
- National Science Foundation, Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA), Solicitation 12-499, 2012.
- MIT News, *MIT, Intel unveil new initiatives addressing 'big data'*, May 31, 2012.
- Wikipedia, *Data Science*. [http://en.wikipedia.org/wiki/Data\\_science](http://en.wikipedia.org/wiki/Data_science)
- Ben Fry, *Computational Information Design* (PhD Thesis), Massachusetts Institute of Technology, April 2004.
- David Smith, *Revolutions Blog* (<http://blog.revolutionanalytics.com/2011/09/data-science-a-literature-review.html>), September 2011.
- Tweet from CS News Update at 9:54PM on May 21, 2012.
- Matt Stiles, *How Common Is Your Birthday? Parts 1 & 2*. The Daily Viv Blog, May 12 & 18, 2012. Available at: <http://thedailyviz.com/2012/05/12/how-common-is-your-birthday/>
- Deepak Kumar. *Data Science Overtakes Computer Science*. ACM Inroads Magazine. Volume 3 Issue 3, September 2012. ACM New York.
- Illuminated Map Displays UK Traffic Casualties Posted by [Eugene](#) on December 7, 2011 at 1:00pm (<http://www.mymodernmet.com/profiles/blogs/illuminated-map-displays-uk-traffic-casualties>)
- Min Chen & Luciano Floridi, *An Analysis of Information in Visualization*, Synthese 2012 (to appear), Springer.