Introduction to Data Visualization

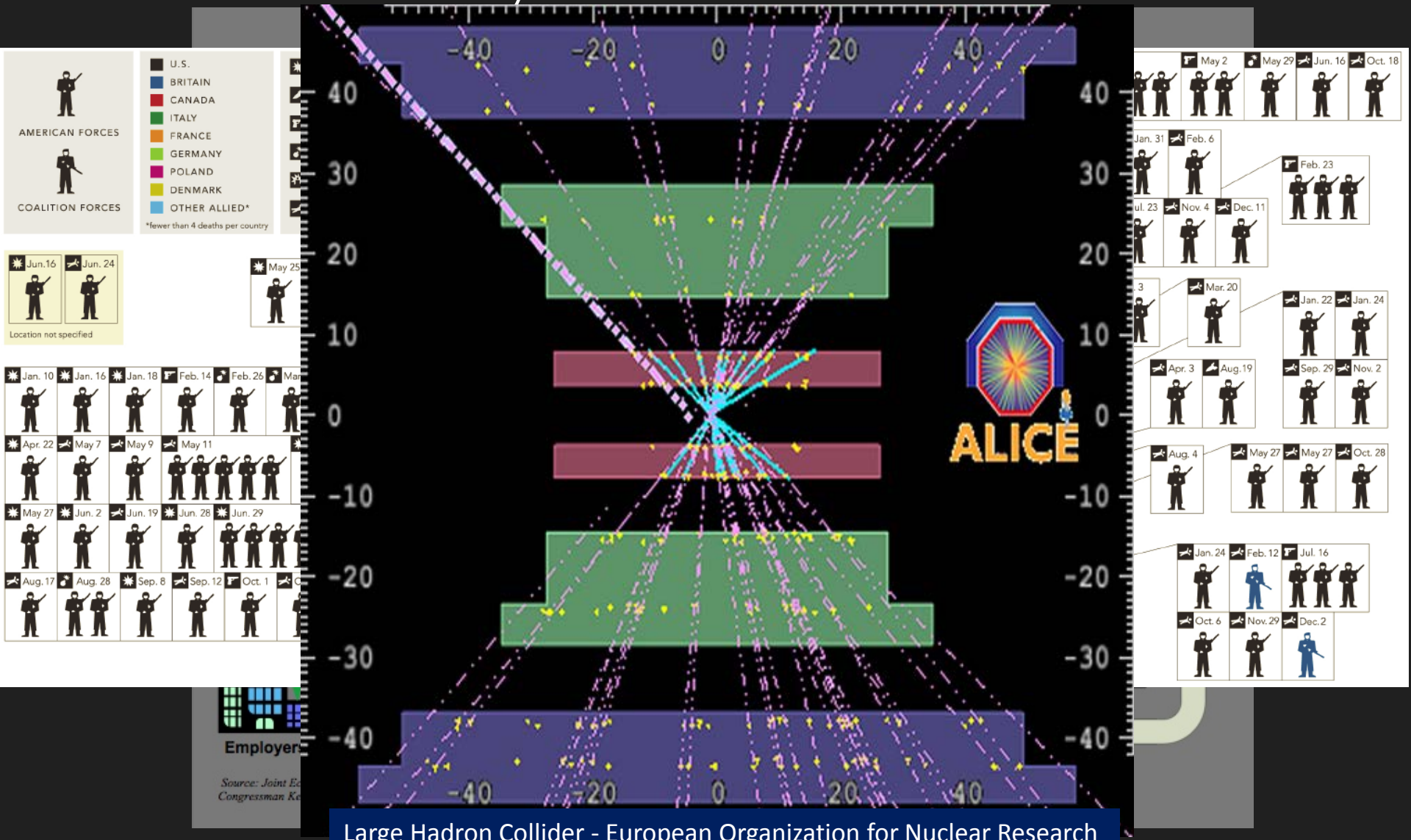
Alark Joshi

Introduction

- Ph.D. in Computer Science with an emphasis on Data Visualization - University of Maryland
- Postdoctoral Fellow - Yale University

- Conduct research on developing effective visualizations
  - Neurosurgical applications
  - Atmospheric Physics
  - Data Visualization on Tablets

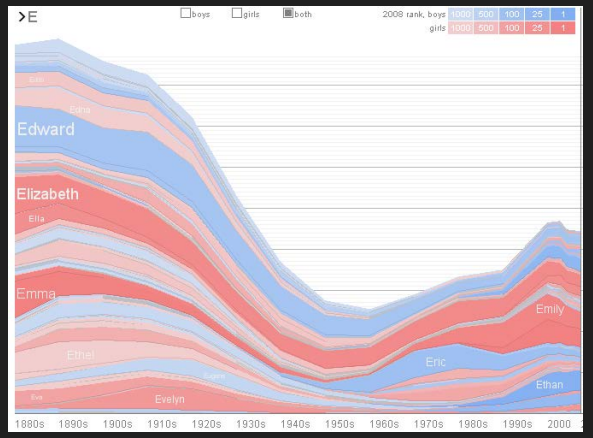
Why are we here?



Large Hadron Collider - European Organization for Nuclear Research

Image Credits: NYTimes, Infosthetics, FlowingData

Why are we here?



- Baby Name Wizard
  - http://www.babynamewizard.com/voyager
- Origin of Species – Edits
  - http://benfry.com/traces/
- Netflix Queues
  - http://www.nytimes.com/interactive/2010/01/10/nyregion/20100110-netflix-map.html?ref=nyregion
- Unemployment Visualization (NYTimes)
  - http://www.nytimes.com/interactive/2009/11/06/business/economy/unemployment-lines.html

Goals

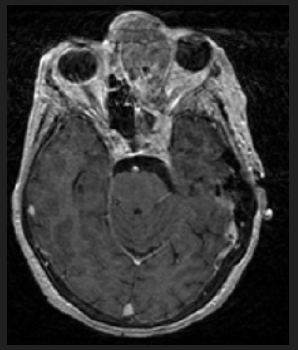
- Understand what makes a visualization effective
- Critically evaluate a visual representation of data by looking at various examples in media (newspapers, television and so on)
- Gain hands-on experience with visualization tools (Tableau, Many Eyes)
- Incorporate visualization principles to build an interactive visualization of your own data

Data Scientist

- Professionals responsible for filtering out the noise and analyzing essential information
- Integral part of competitive intelligence, a newly emerging field that encompasses data analysis to help businesses gain a competitive edge
- A shortfall of about 140,000 to 190,000 individuals with analytical expertise is projected by 2018
- Glassdoor.com shows average data scientist salaries ranging from $60,000 to $115,000

Slide credits: Data Scientist: The Hottest Job You Haven't Heard Of – AOL Jobs

What is Data Visualization?



- Visual Representation of Data
- For exploration, discovery, insight, ..
- Interactive component provides more insight as compared to a static image

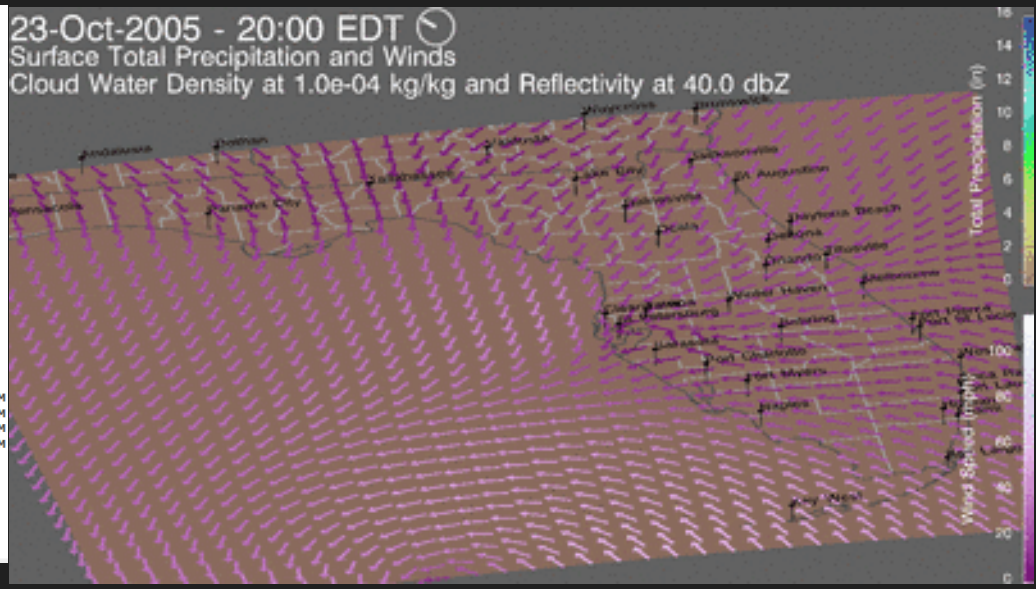


Image credits: Yahoo Finance, IBM

In-class Reading

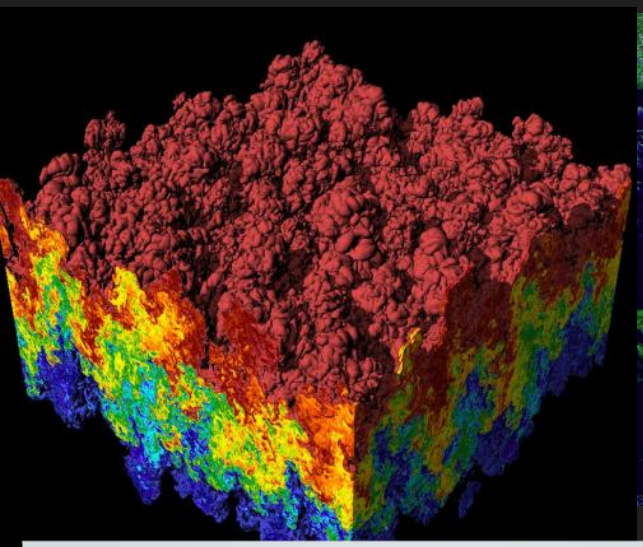
- Seven things you need to know about Data Visualization (5 mins)

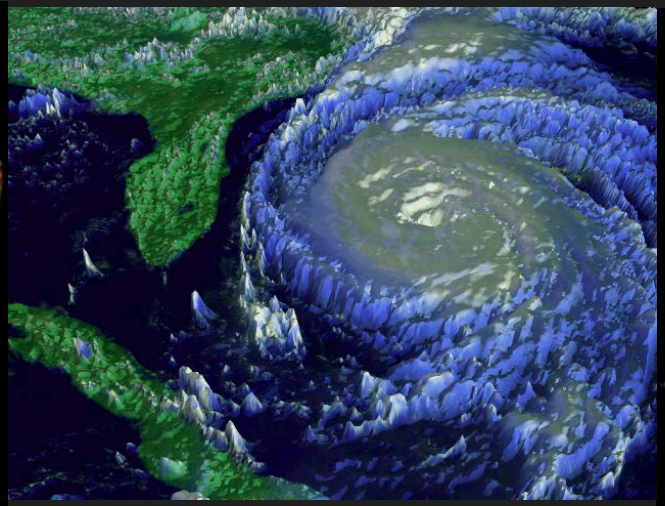
- What did you find out about data visualization that you did not know?

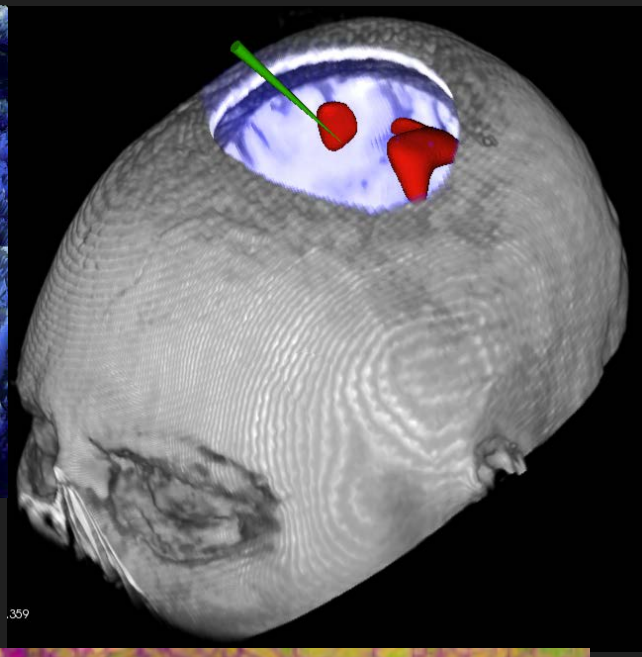
Types of Data Visualization

- Scientific Visualization –
  - Structural Data – Seismic, Medical, ..
- Information Visualization
  - No inherent structure – News, stock market, top grossing movies, facebook connections
- Visual Analytics
  - Use visualization to understand and synthesize large amounts of multimodal data – audio, video, text, images, networks of people ..

Scientific Visualization

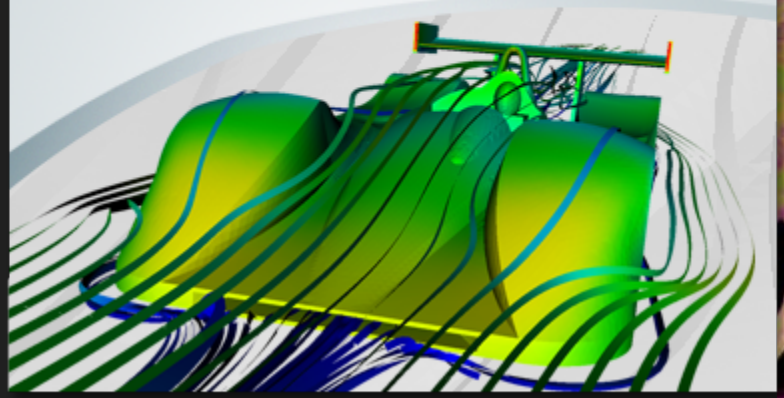


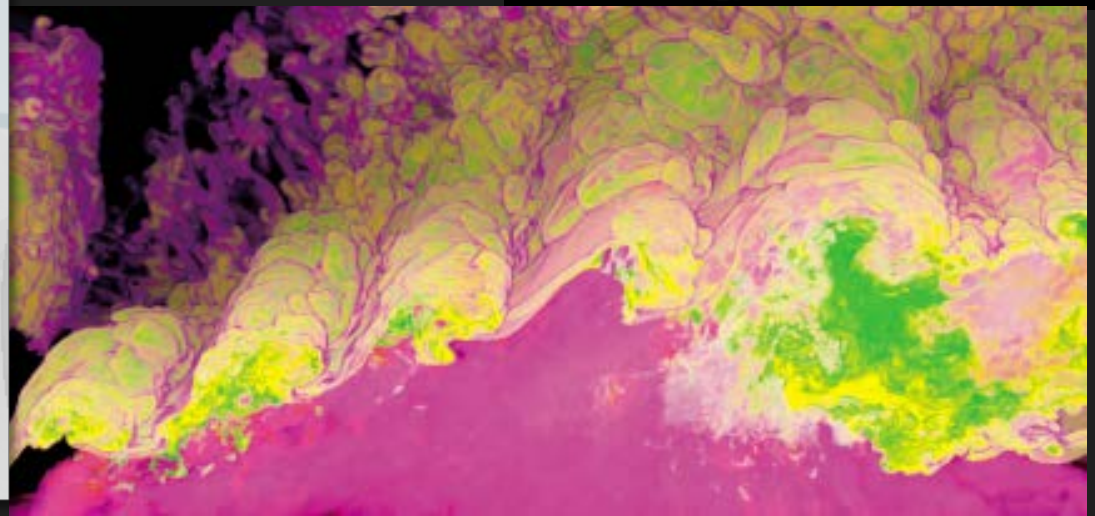




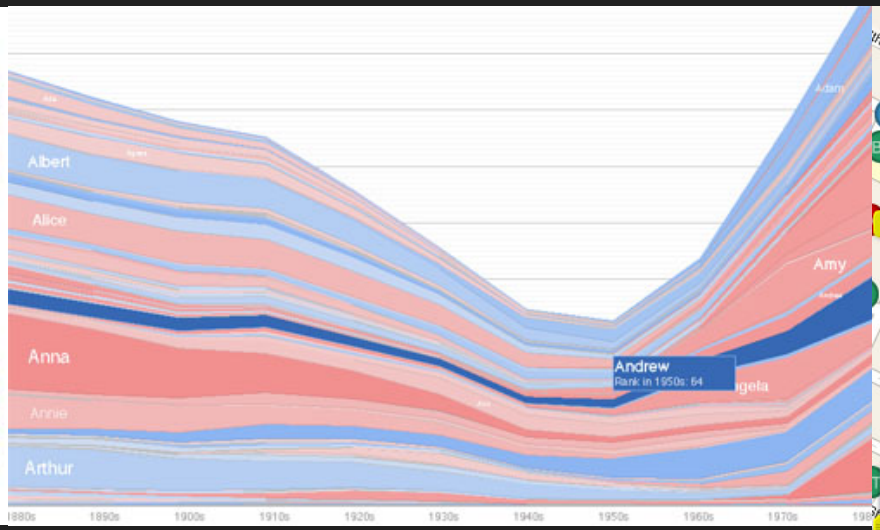
ParaView

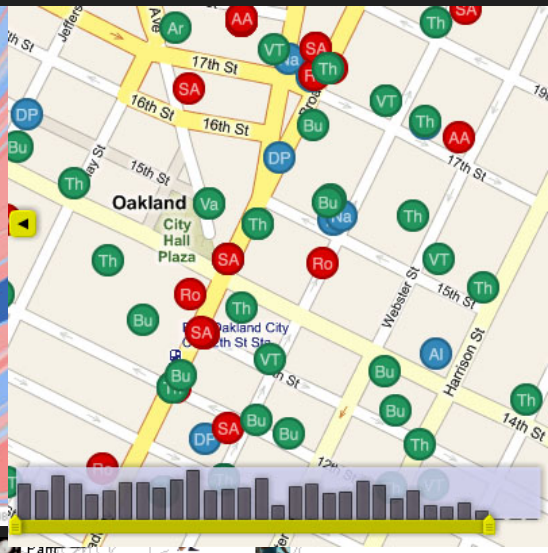
Visualize data sets of size, from small to very large on desktop computers or high-performance clusters, using this open-source, multi-platform application.

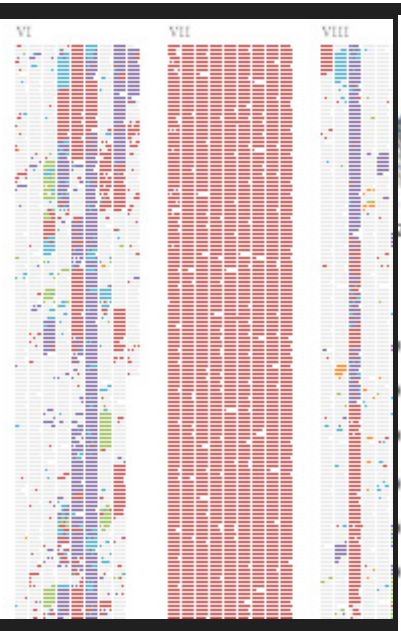


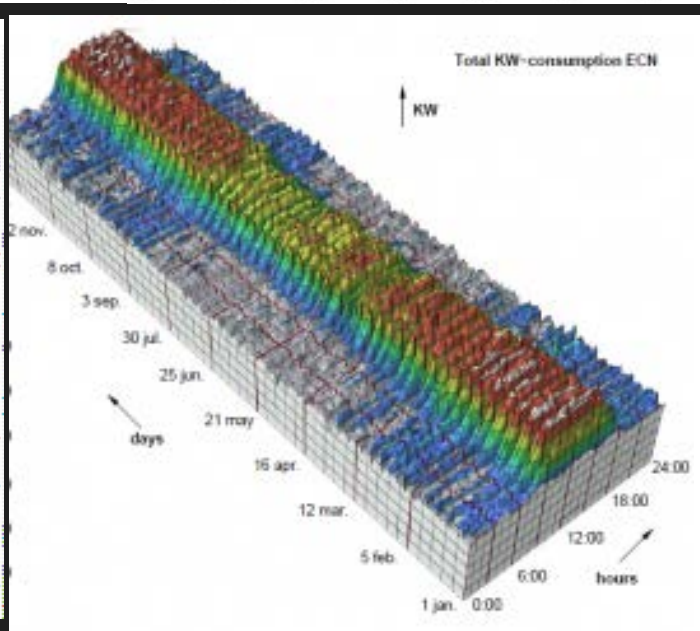


Information Visualization



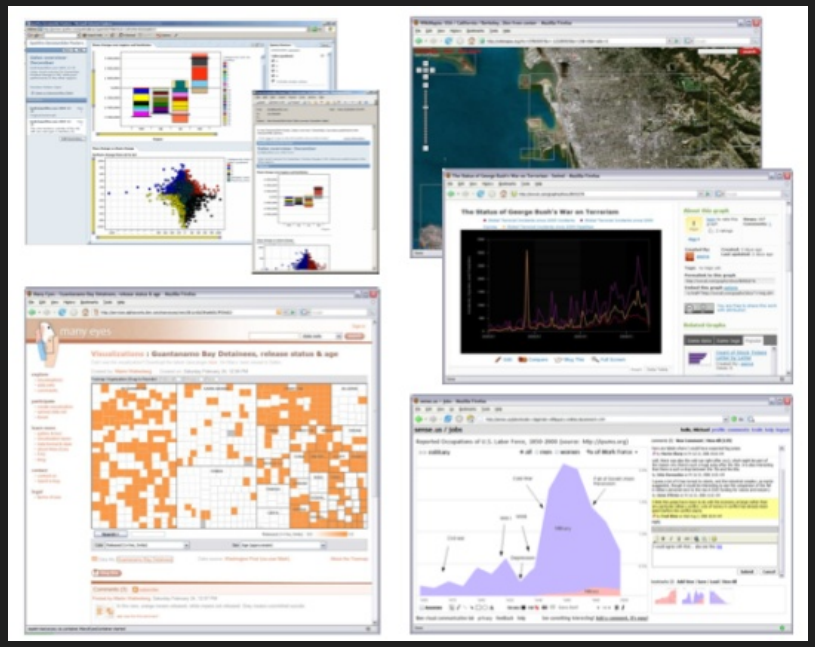


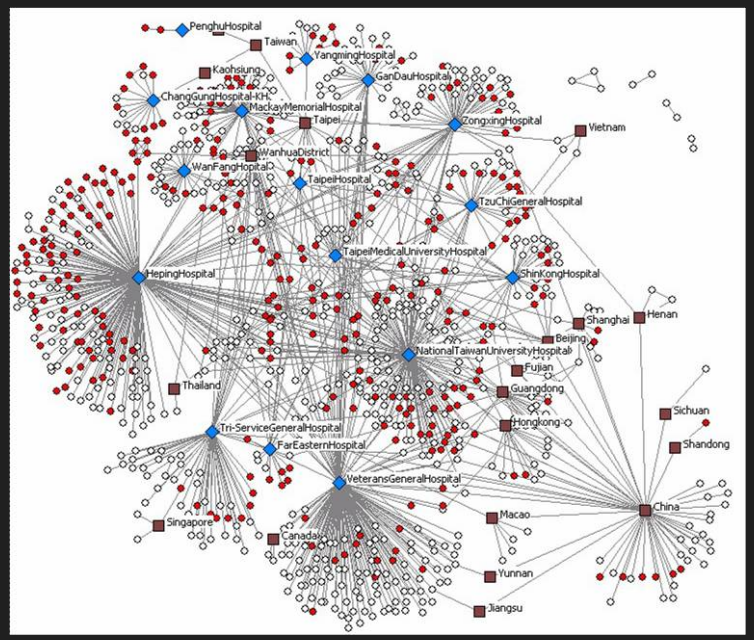




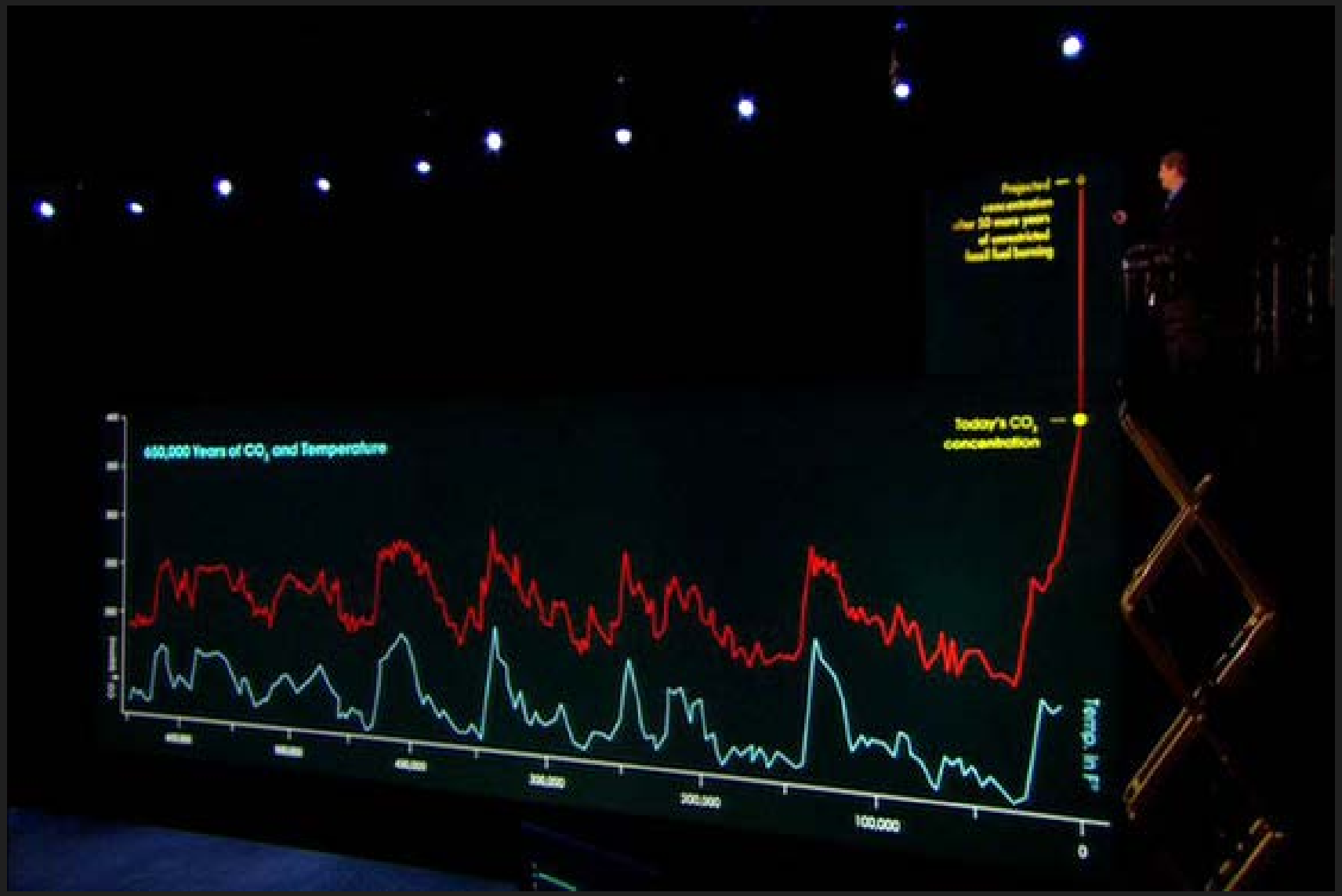
Visual Analytics

- Integration of interactive visualization with analysis techniques to answer a growing range of questions in science, business, and analysis.
- Making sense of multimodal data -audio clips, video, photographs, transcripts, …

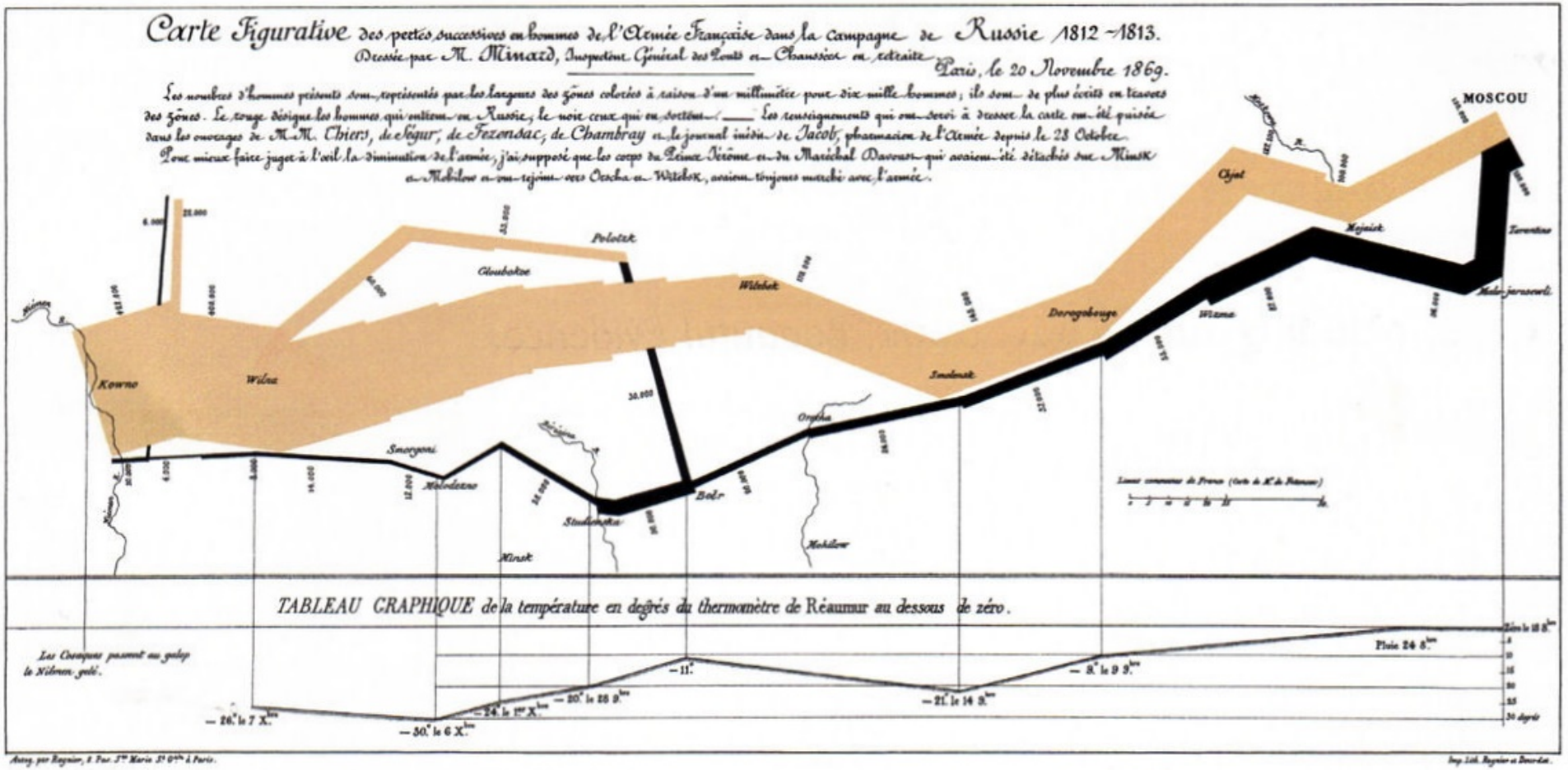




Al Gore – An Inconvenient Truth



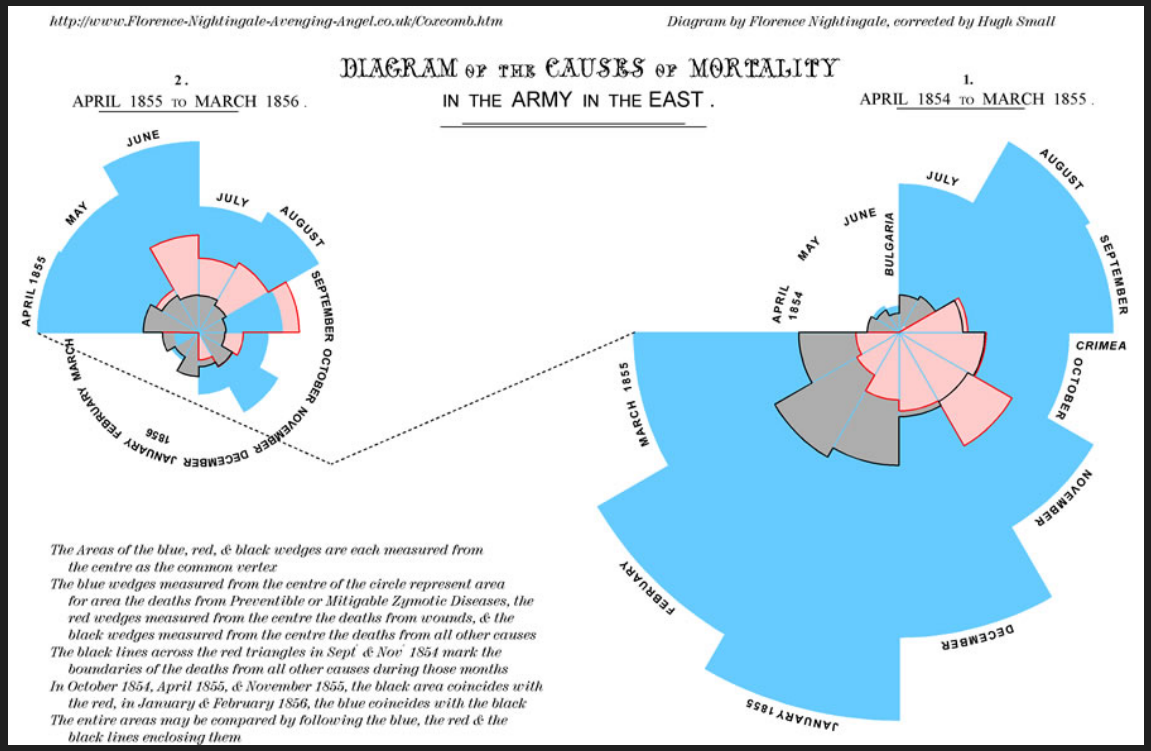
Visualization of Napoleon's Army



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. Napoleon's March poster $14 postpaid; English/French version $18 postpaid.

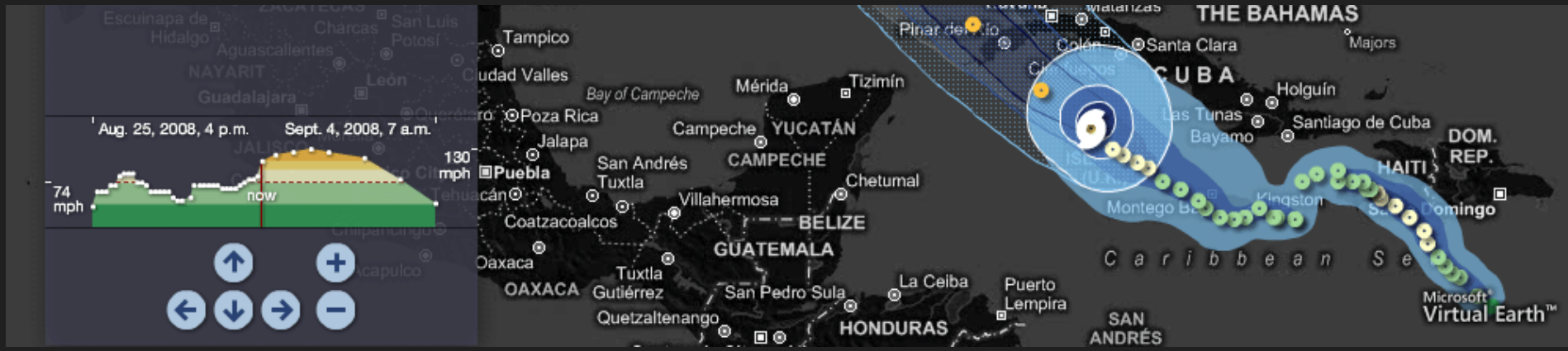
Impact of Visualization

- Huge impact on policy, planning and disaster avoidance.
- Florence Nightingale's visualization of casualties during the Crimean War



Impact of Visualization

- Hurricane Visualization for the common man

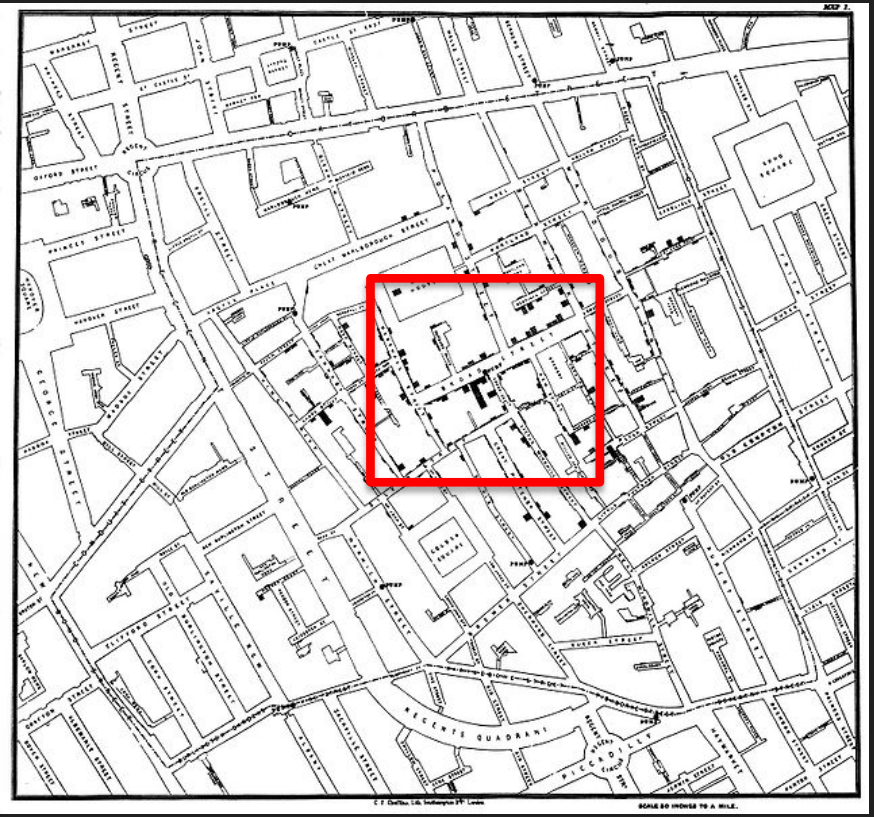


Demo:
http://www.msnbc.msn.com/id/26295161?preferredName=Gustav

Image Credits: Stamen Design

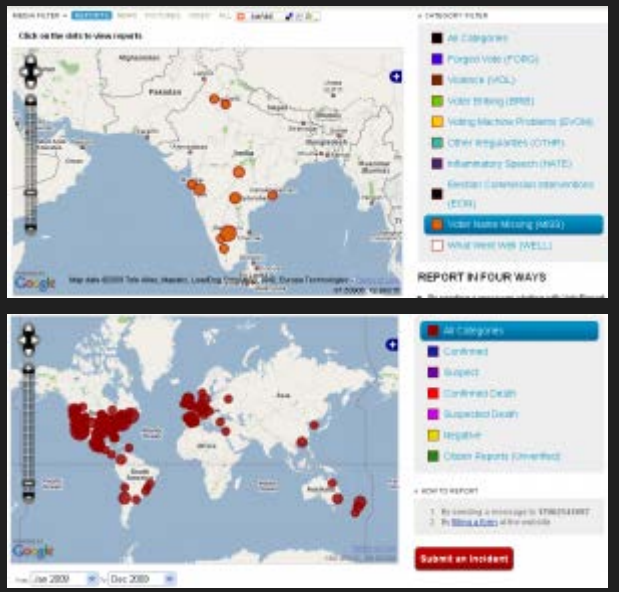
Impact of Visualization

- John Snow's Cholera Map
- Snow used a spot map to illustrate how cases of cholera clustered around the pump



Ushahidi

- “testimony” in Swahili
- Developed to map user reported violence in Kenya after the post-election fallout at the beginning of 2008
- Adapted and used by
  - votereport.in and
  - swineflu.ushahidi.com.



Good data representation principles

- Breakout into groups of two and identify five good data visualization principles
  - 5 minutes

List of principles

- Integrate and distill the principles
- “Everything should be made as **simple** as possible, but not **simpler**.” ~Albert Einstein

Contact Information

- Email: apjoshi@usfca.edu
- Webpage: http://cs.usfca.edu/~apjoshi/cs686/
- Office location: HR 510-i
- Weekly office hours:
  - Monday – 1:30-3pm
  - Wednesday – 11am-12pm
  - Friday Skype office hours – 10am-11am
    - Skype id - alark.usf

Resources

- Required books
  - Getting Started with Processing by Casey Reas and Ben Fry. (Required)
  - Interactive Data Visualization for the Web by Scott Murray. (Required)
- Research Papers
- Websites/Blogs
- Relevant book chapters that will be provided well in advance

Course Details

- Assignments - 40%
- Reading Response (blog posts) - 15%
- Tool/Library presentation -5%
- In-class participation – 10%
- Final Project - 30%
  - Client interview + Project Proposal – 5%
  - Annotated bibliography – 5%
  - Alpha release – 5%
  - Beta release – 5%
  - Final Project Presentation, Report (8-pages), Source Code and Client Testimonial – 10%

Assignments

- A1 - Introduction to Data Visualization through Tableau
- A2 - Data exploration through Visualization in R
- A3 - Multivariate Visualization in Processing
- A4 - Web-based Visualization in D3

Graphic Design IQ Test

- http://www.perceptualedge.com/files/GraphDesignIQ.html

Activities for next class

- Create a blog (wordpress.com or any hosting service of your choice that allows comments) and email me the link

Next class

- Design Principles
- Graphical Integrity

For next week

- Look through some of the popular media (websites/newspapers etc.) and post an image of a good and a bad visualization on your blog before next class
- Watch Hans Rosling's TED talk and read assigned paper and post your reaction on your blog by next week