EM622 Data Analysis and Visualization Techniques for Decision-Making

From Data Visualization to Data Analytics

STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY

Agenda

- Why visualization?
- Graphs that tell you something- Pizza place geography
- Brief history of data visualization
- Beautiful examples
- Trends & tools
- Course overview

Why visualization?

Data visualization is the creation and study of the visual representation of data, meaning "information that has been abstracted in some schematic form, including attributes or variables for the units of information" - Michael Friendly (2008). "Milestones in the history of thematic cartography, statistical graphics, and data visualization"

- *"The great fun of information visualization is that it gives you answers to questions you didn't know you had"* - Prof.Ben Schneiderman, University of Maryland;
- *"The great use of data visualization is that it gives you clues to questions you didn't know you had"* - Saaien Tist

Why visualization?

Direct Navigation:
Answer Questions of Interests

1. • Look for patterns
2. • Look for expectations
3. • Similar to querying a database

"INFORMS Data Exploration Workshop Workbook"- Shmueli & Hardoon 2013

Why visualization?

"INFORMS Data Exploration Workshop Workbook"- Shmueli & Hardoon 2013

Why visualization?

Uses of Data Visualization:

- Evaluate data quality and data issues;
- Answer questions of interest;
- Create 'data stories' to identify opportunities and challenges;
- Communicate 'data stories' to an audience (data presentation).
- "How to tell a story with data?" by Nathan Yau:

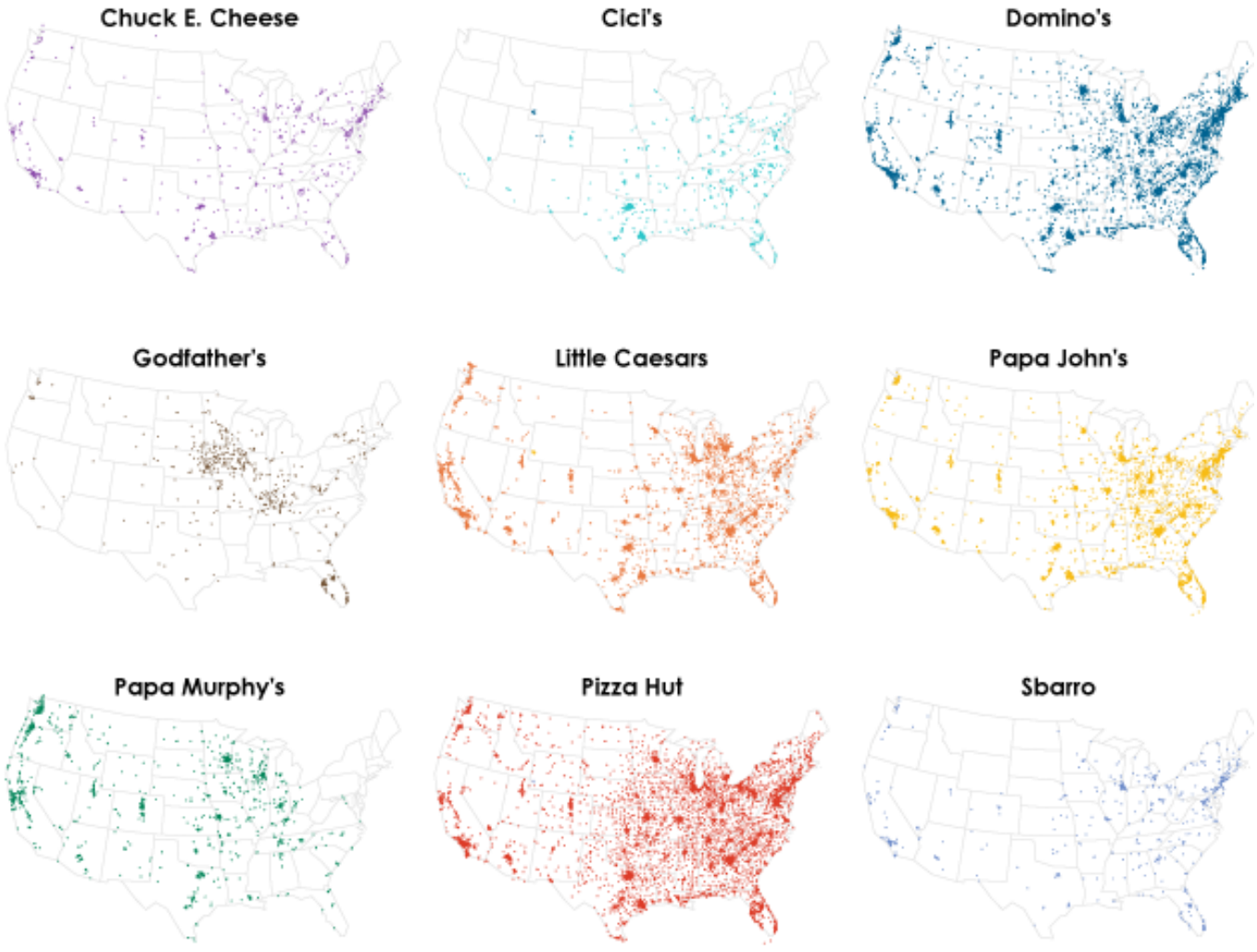
  `http://www.youtube.com/watch?v=mkEXx7sDXAI`

Graphs that tell you something- Pizza place geography

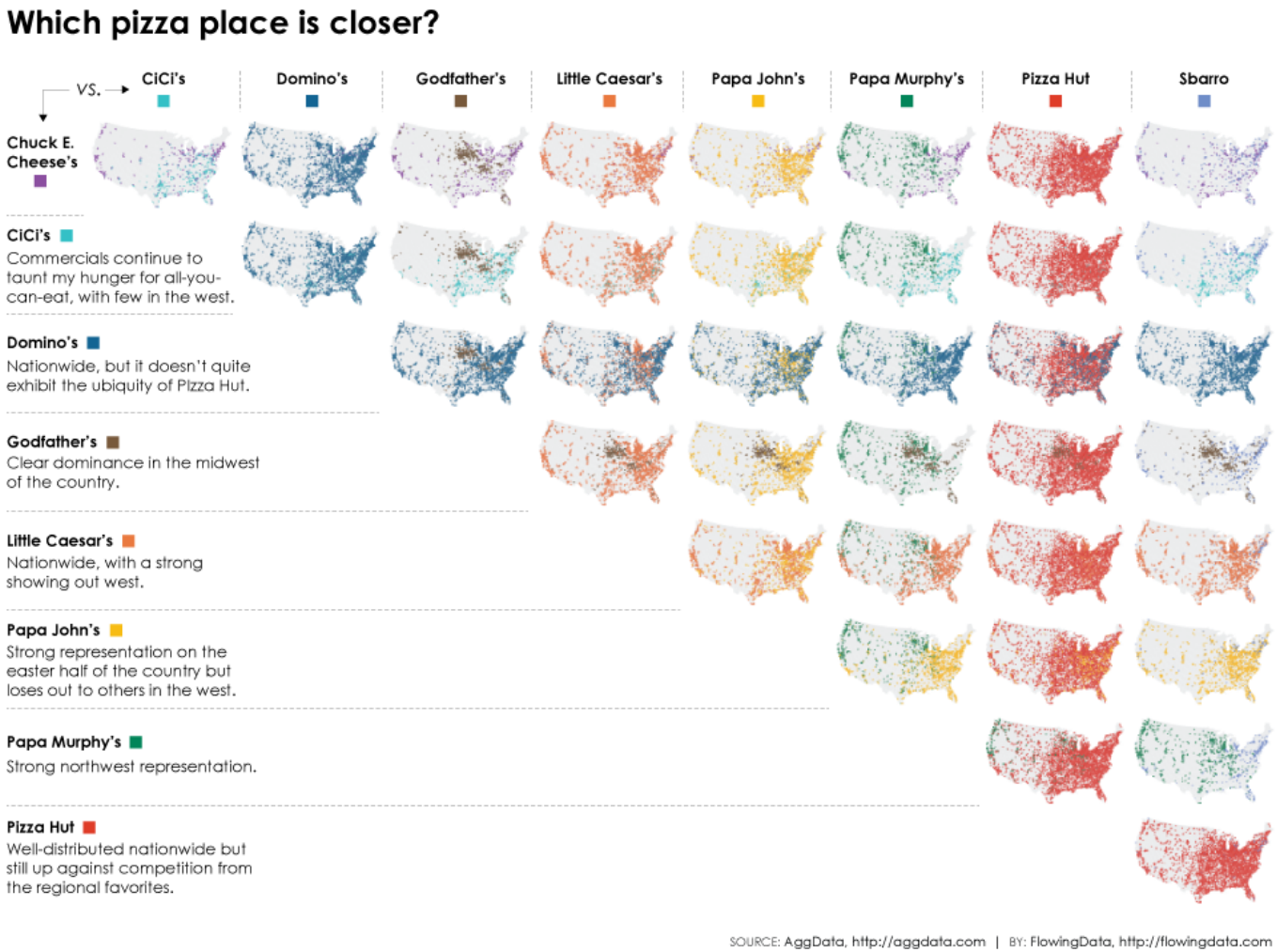
- Taken from FlowingData (http://flowingdata.com) by Nathan Yau.
- Data source: AggData(http://aggdata.com).
- Pizza chains are everywhere, but some are more popular than the others in a certain region.
- Given the store locations of Chunk E. Cheese's, CiCi's, Domino's, Godfather's, Little Caesar's, Papa John's, Papa Murphy's, Pizza hut and Sbarro, what can you tell?

Graphs that tell you something- Pizza place geography

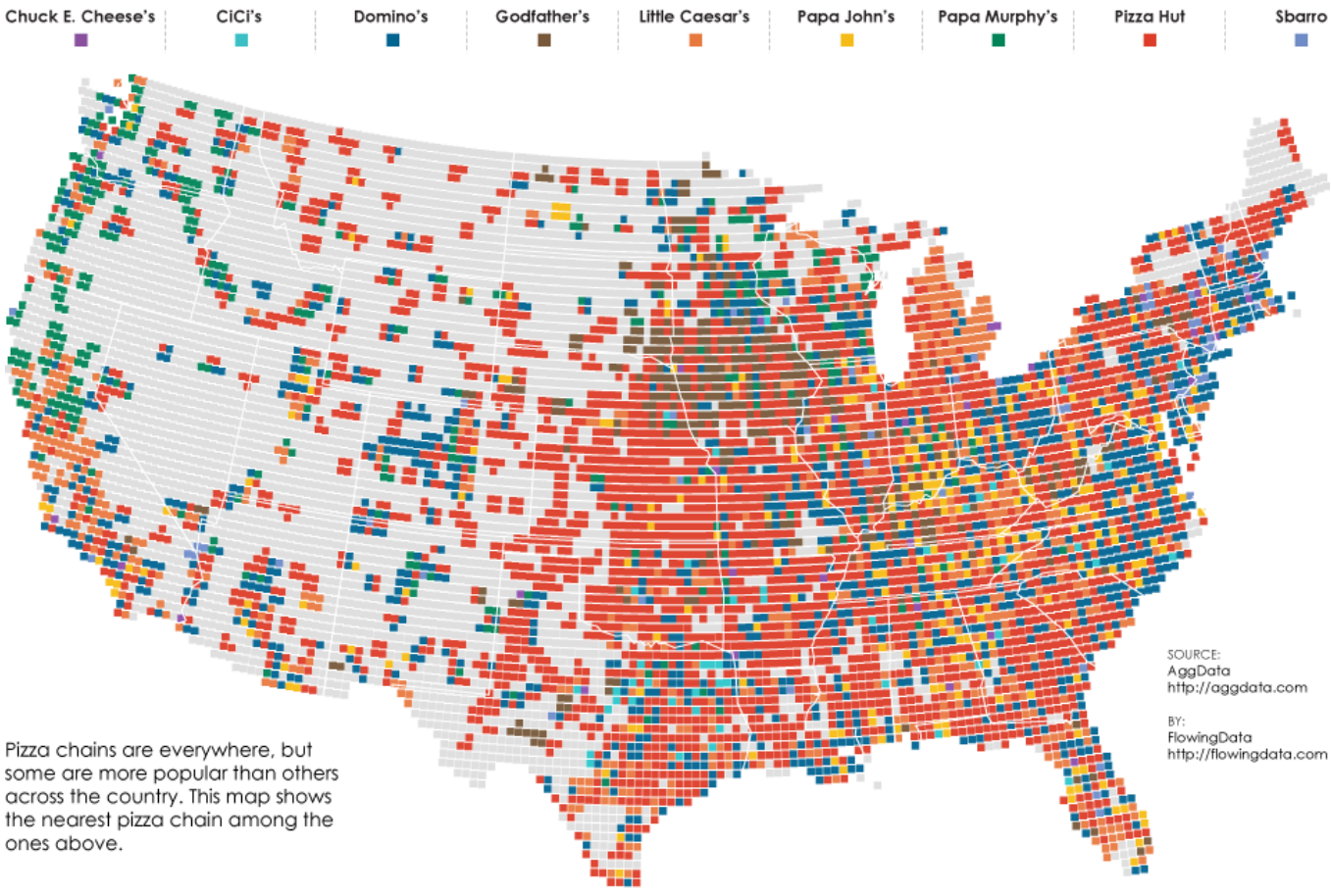
- Which restaurant is dominant nationally/regionally?



Graphs that tell you something- Pizza place geography



Graphs that tell you something- Pizza place geography



Brief history of data visualization

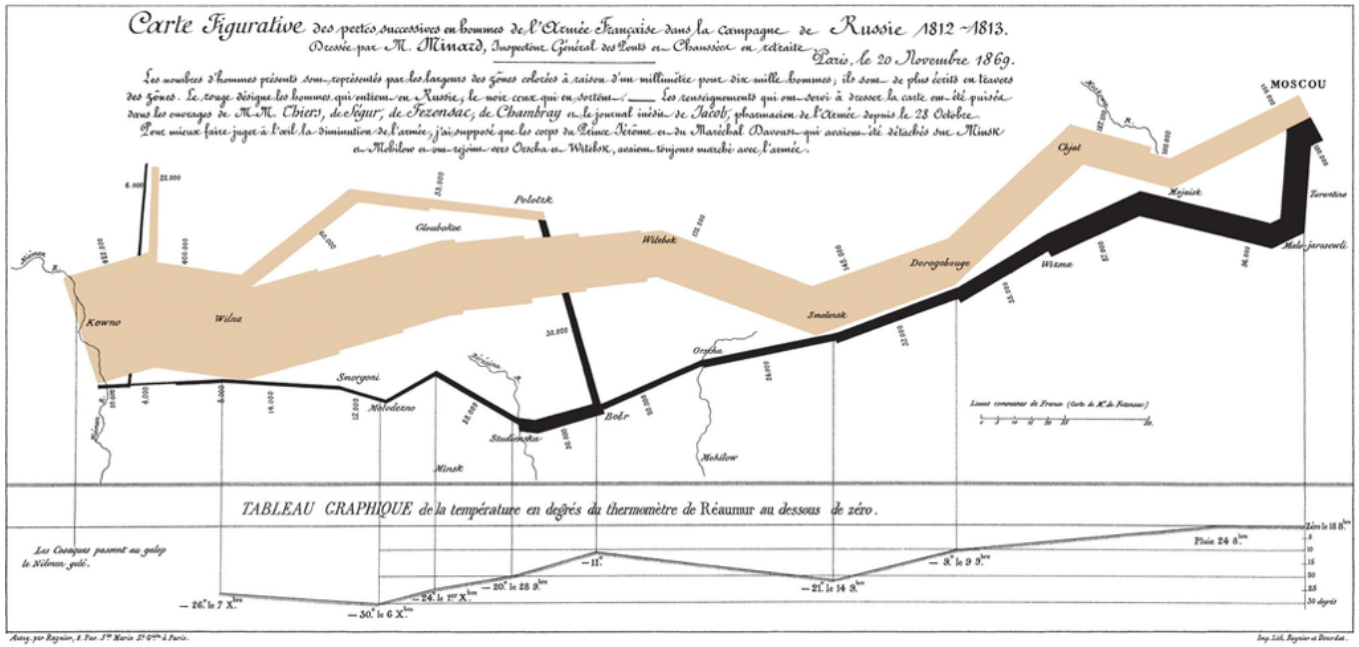
- The graphic portrayal of quantitative information has deep roots. These roots reach into histories of thematic cartography, and statistical graphics.
- The earliest seeds arose in geometric diagrams and in the making of maps to aid in navigation and exploration.
- Diagrams were used to illustrate mathematical proofs and functions; various graphic forms were invented to make the properties of empirical numbers(trends,tendencies, and distributions- more easily communicated, or accessible to visual inspection.
- Maps, diagrams and graphs were initially hand drawn, piece-by-piece -> etched on copper-plate and manually colored -> lithography and photo-etching-> computer software

Source: "Milestones in the history of thematic cartography, statistical graphics, and data visualization" by Michael Friendly: http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf

Beautiful Examples

Charles Minard's flow map of Napoleon's March(1869)

- The French engineer, Charles Minard (1781-1870), illustrated the disastrous result of Napoleon's failed Russian campaign of 1812.
- The brown line (on the top) shows the troop levels as they begin their march towards Moscow.The black line shows their retreat.



Beautiful Examples

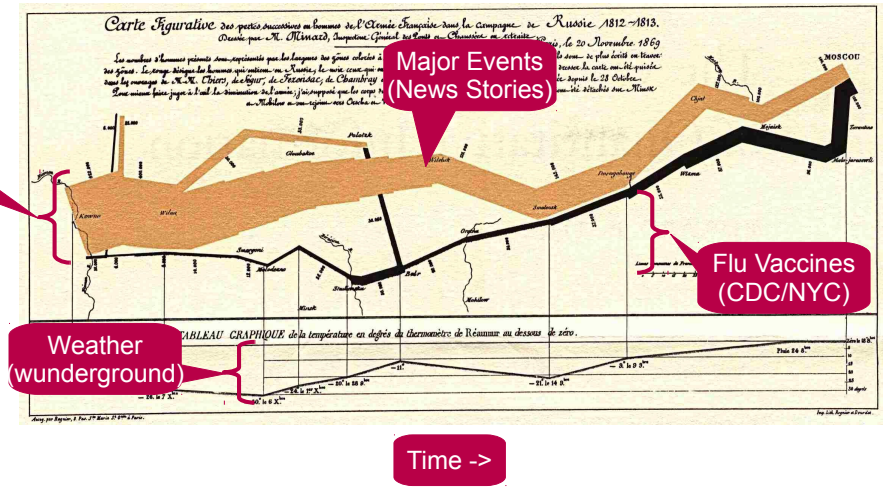
Principles for designing the visualization(Notes from Edward Tufte's training):

1. Show comparisons, contrasts, differences.
2. Show causality, mechanism, explanation, systematic structure.
3. Show multivariate data; that is, show more than 1 or 2 variables.
4. Completely integrate words, numbers, images, diagrams.
5. Documentation. The credibility of an evidence presentation depends significantly on the quality and integrity of the authors and their data sources.(Who/When/Where)
6. Content counts most of all.

Challenge 1

Based on Minard's flow map of Napoleon's March, think about a new example nowadays that may use principles demonstrated in the flow map.

HOW DO YOU TRACK THE SPREAD OF FLU THROUGHOUT NYC OVER LONG PERIODS OF TIME WHILE SHOWING THE EFFECTS OF VACCINES, MAJOR EVENTS, AND THE WEATHER?

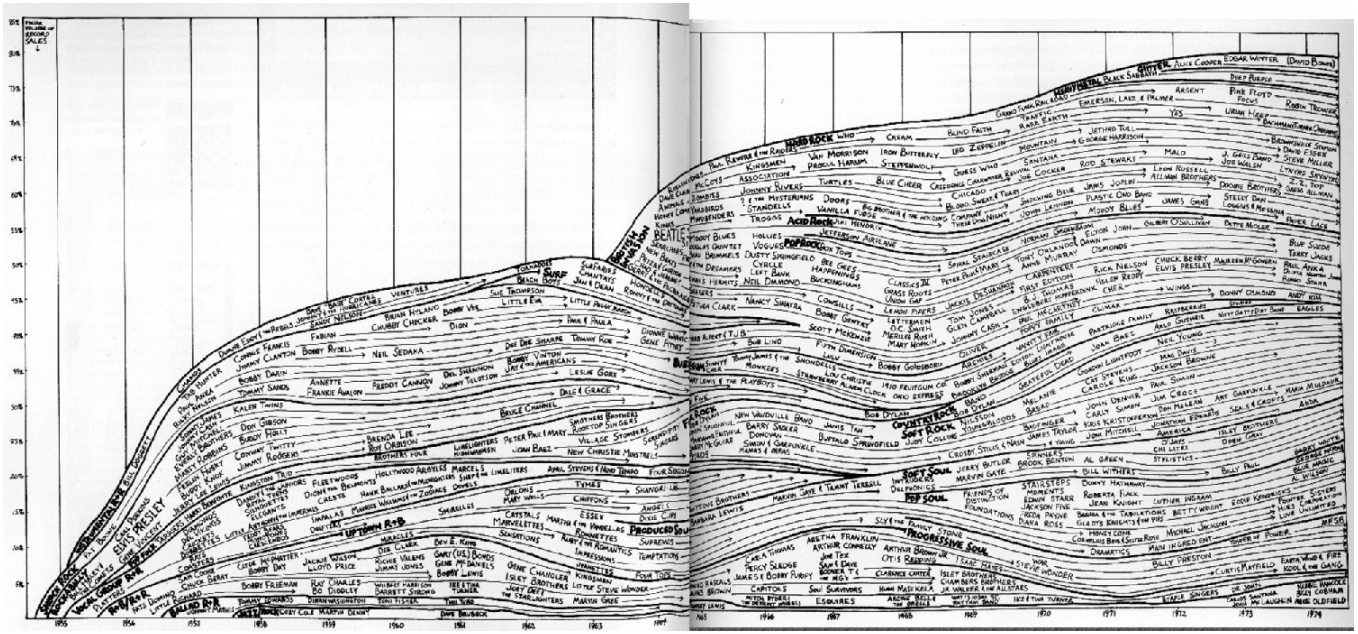


FIVE LAYERS OF 19TH CENTURY INFORMATIONAL GRAPHIC AWESOMENESS, THATS HOW!

Beautiful Examples

Rock and Roll History

- This graphic was originally designed by Steve Chappel and Reebe Garofalo in Rock 'N' Roll is Here to Pay: The History and Politics of the Music Industry (1977)



Beautiful Examples

- Hans Rosling's motion bubble chart
- Rosling's lectures use huge quantities of public data to reveal the story of the world's past, present and future development. Now he tells the story of the world in 200 countries over 200 years using 120,000 numbers - in just four minutes. http://www.youtube.com/watch?v=jbkSRLYSojo
- Play with the similar motion bubble chart on world population for recent 50 years developed by Google :

  https://code.google.com/p/google-motion-charts-with-r/
- Economic infographics on US debt 2013 in 100 dollar bills:

  http://demonocracy.info/

Trend & Tools

Emerging industry:biochemical, financial industry, etc.

- FINVIS-Stock screener for investors and traders, financial visualizations.
- http://finviz.com/map.ashx?t=sec

Popular trend: interactive infographics with audience or analyst

- "Where the Heat and the Thunder Hit Their Shots"(NYT):

  http://www.nytimes.com/interactive/2012/06/11/sports/basketball/nba-shot-analysis.html
- How to get the data? NBA player tracking:

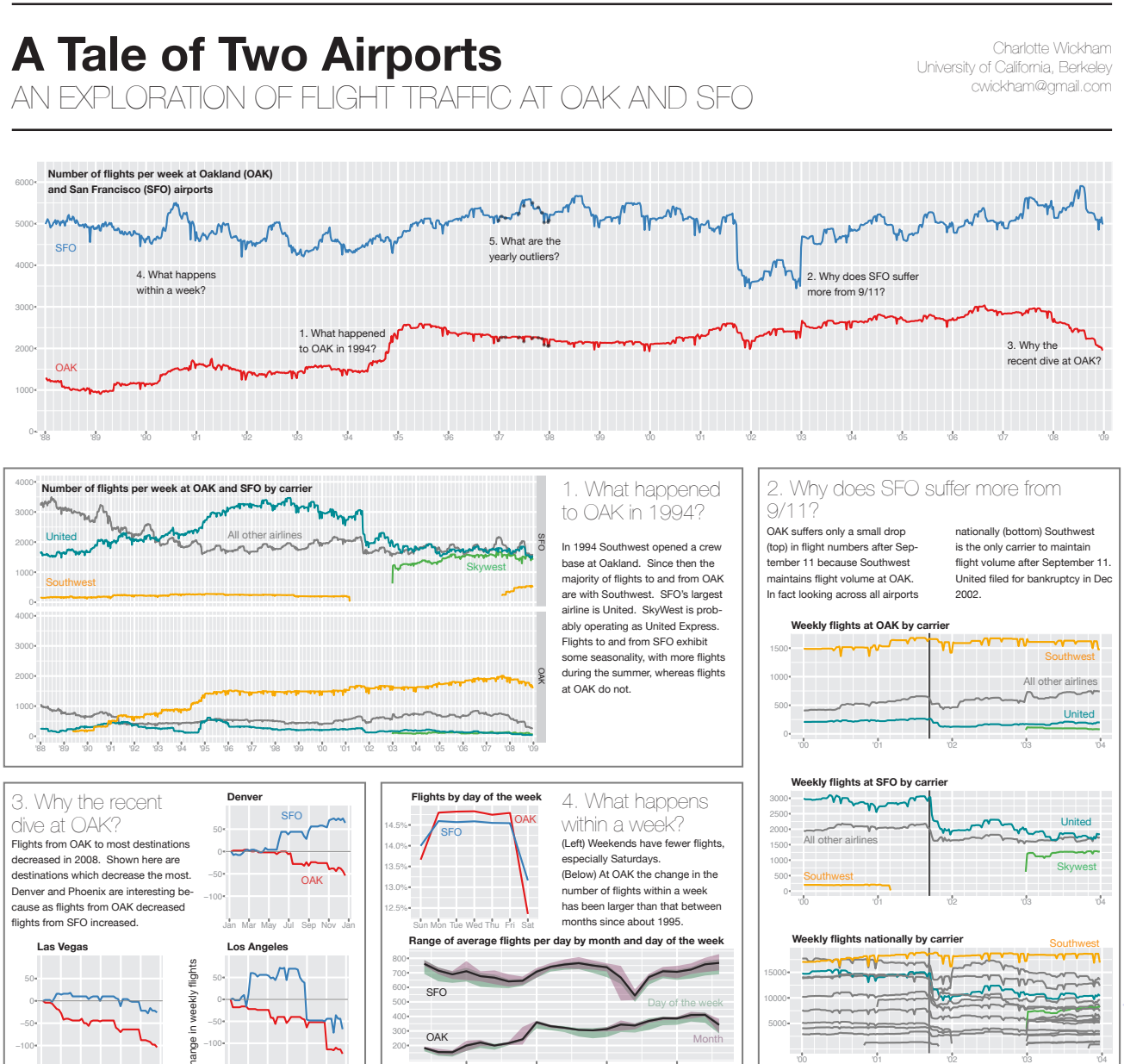
  http://stats.nba.com/playerTracking.html
- Microsoft Research prototype: Draw your graphics on a touch screen and create your datavis in real time:http://www.datastorytelling.tv/

  Microsoft-Research-prototype-Draw-your-graphics-on-a-touch-screen-and-create-your-datavis-in-real-time_

  a47.html

Trend & Tools

Popular tools: R



Trend & Tools

- Popular tools: Spotfire Analytics
- A business intelligence software that designs, develops and distributes in-memory analytics with interactive visualization;
- Fastest To Actionable Insight;
- Visibility Into the Unknown;
- Self-Service Data Discovery;
- Universal Adaptability;
- Feature demo `http://www.youtube.com/watch?v=KoUswm18ZfQ`

Trend & Tools

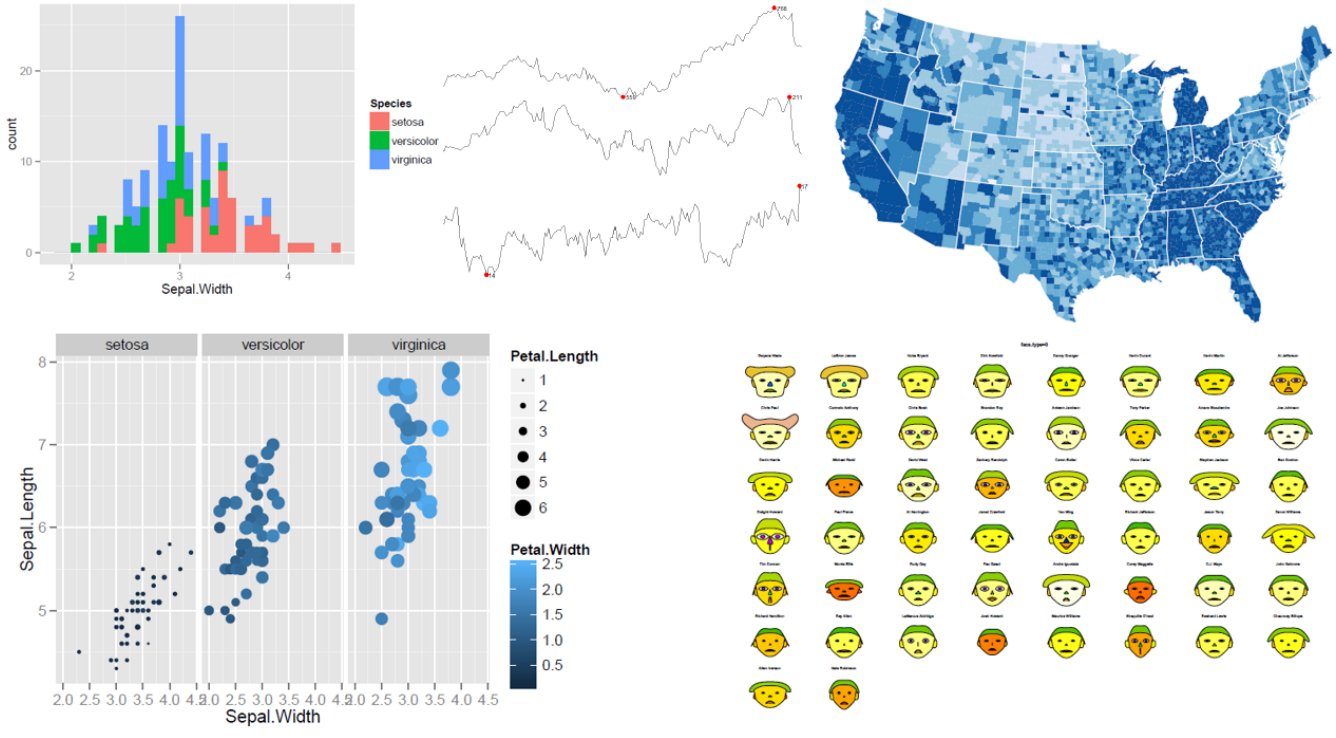
- Popular tools: Google APIs
- Google Fusion Table API can insert, update, delete and query data programmatically;
- Visualize the table instantly on a map or as a chart;
- Can be embed in a web page or blog post;
- Example: Qing's project on visualizing Sandy's impact in NJ.

Trend & Tools

- Popular tools: D3.js
- A JavaScript library for manipulating documents based on data allowing highly efficient exploration on large multi-dimensional database.
- D3 helps you bring data to life using HTML, SVG and CSS.
- Example: Where Are House Hunters Searching?

`http://trends.truliablog.com/vis/metro-movers/`

Course Overview



Resources:

1. Data visualization course material offered by Hadley Wickham: http://courses.had.co.nz/
2. "A Tour through the Visualization Zoo" by Jeffrey Heer, Michael Bostock, Vadim Ogievetsky: http://queue.acm.org/detail.cfm?id=1805128
3. "Storytelling with maps" from Esri: http://storymaps.esri.com/home/
4. "Visualize This: The FlowingData Guide to Design, Visualization, and Statistics": http://www.amazon.com/gp/product/0470944889/?tag=flowingdata-20
5. Visual Exploration of Big Spatio-Temporal Urban Data:A Study of New York City Taxi Trips http://vgc.poly.edu/projects/taxivis/#video

What have we learned?

1. The advantages of data visualization.
2. How to develop a data story.
3. Current trends in visualization and emerging industries.
4. Popular tools in developing nice visualization.

Homework

Read material in the "Resources" slide, and prepare for discussions:

1. Recall Prof. Rosling's motion bubble chart, can you tell us a story from your own perspective?(by region/country/time)
2. Recall the NBA heatmap from New York Time, summarize your takeaways from this chart that might be used to create other visualizations?