



Visualization Methods and Tools

Jay Urbain, Ph.D.

Electrical Engineering and Computer Science Department
Milwaukee School of Engineering

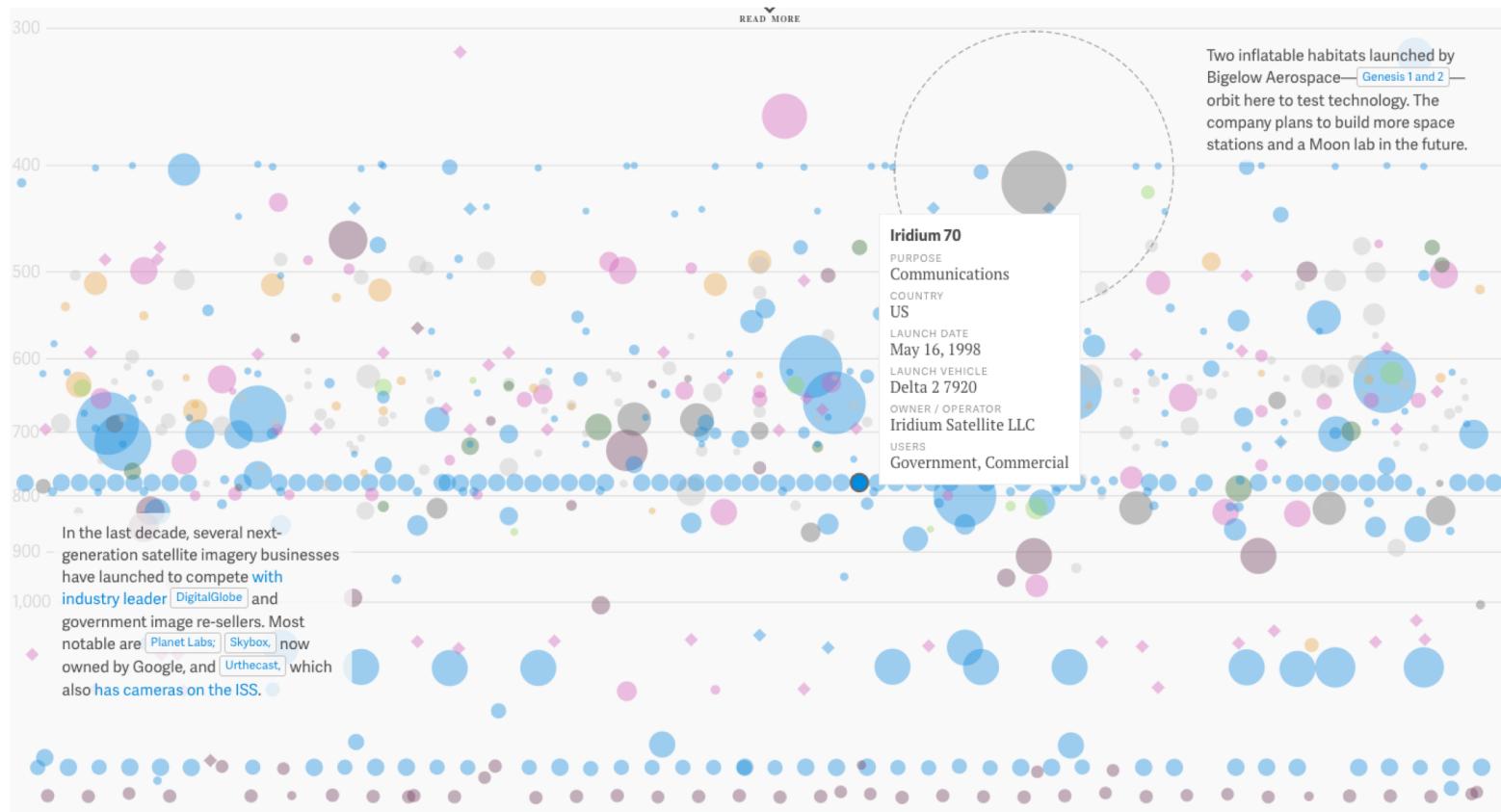
Credits: Erick Burger, Allison Stadd

Data Visualization

- If a picture is worth a thousand words, a data visualization is worth at least a million.
- One of the most impactful ways data scientists can communicate their findings is through data visualizations.
- Data visualizations manipulate complex pools of data to visually display the data's patterns, trends, and correlations.
- The best data visualizations are storytelling tools that spark discussion and elicit calls to action.

Every Satellite Orbiting Earth

<http://qz.com/296941/interactive-graphic-every-active-satellite-orbiting-earth/>



- Interactive graph, built using a [database from the Union of Concerned Scientists](#), displays the trajectories of the 1,300 active satellites orbiting the Earth as you read this. Each satellite is represented by a circular icon, color-coded by country and sized according to launch mass. Scroll through the visualization to explore each satellite's path, individually and in aggregate. You'll also learn what kinds of satellites provide broadband Internet, GPS, and Sirius XM.

Simpson's Paradox

<https://flowingdata.com/2013/09/19/a-visual-explanation-of-simpsons-paradox/>



- The Visualizing Urban Data Idealab (VUDlab) out of the University of California-Berkeley put together this visual look at data that disproves the claim in a 1973 suit that charged the school with sex discrimination. Though the graduate schools had accepted 44% of male applicants but only 35% of female applicants, researchers later uncovered that if the data were properly pooled, there was actually a small but statistically significant bias in favor of women. That's called a Simpson's Paradox. The interactive graphs in the data visualization let you combine and separate different segments of the data to see what exactly went down back in 1973.

Simpson's Paradox

- Lisa and Bart, each edit articles for two weeks.

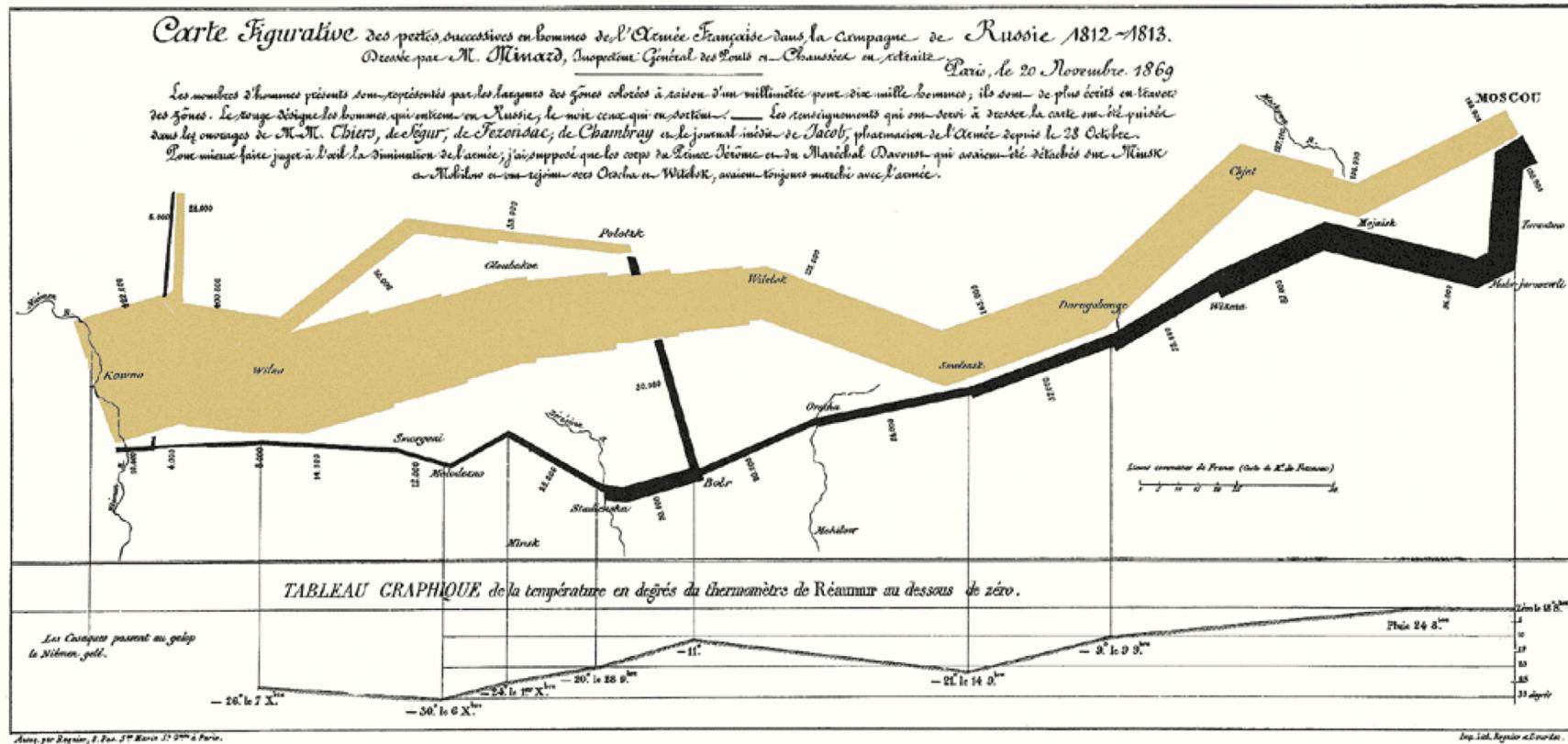
Editor \ Period	Week 1	Week 2	Total
Lisa	0/1	3/4	3/5
Bart	1/4	1/1	2/5

- Both times Bart improved a higher percentage of articles than Lisa, but the actual number of articles each edited (the bottom number of their ratios, also known as the *sample size*) were not the same

Editor \ Period	Week 1 quantity	Week 2 quantity	Total quantity <i>and</i> weighted quality
Lisa	0%	75%	60%
Bart	25%	100%	40%

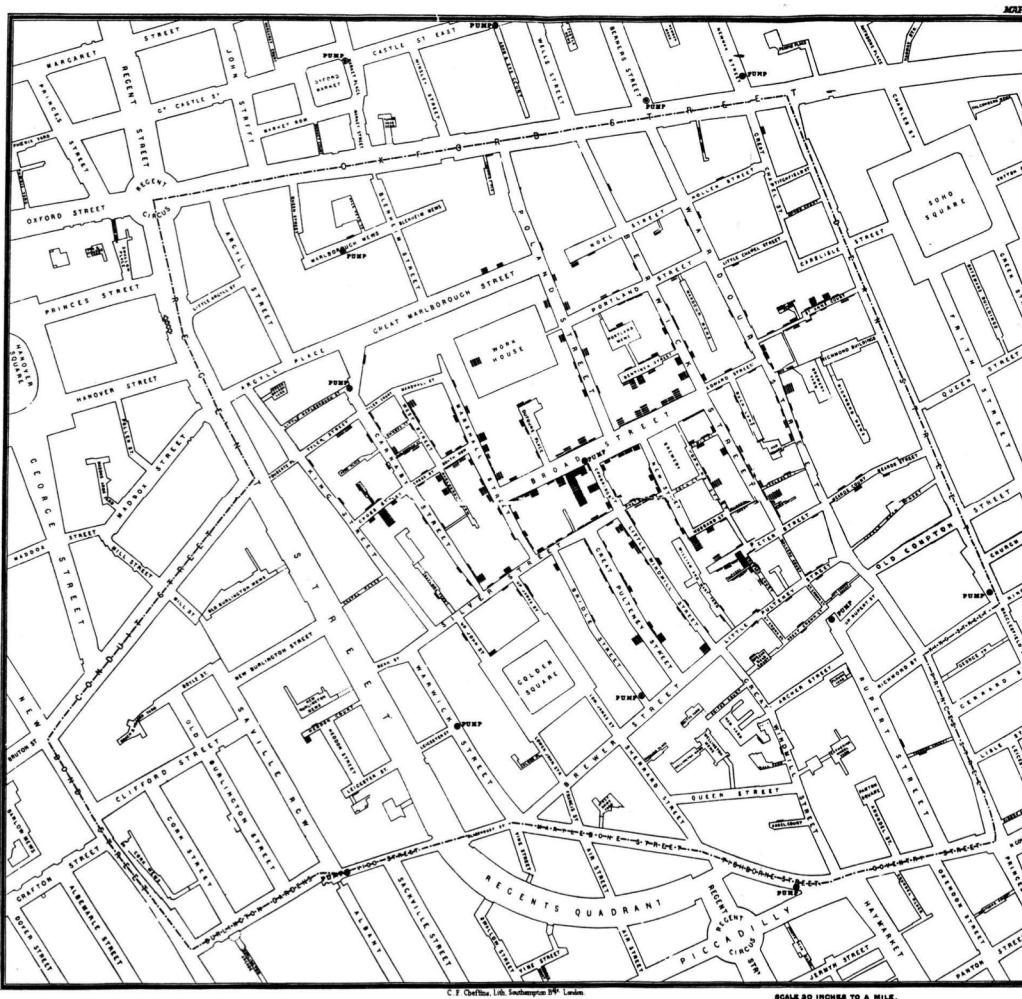
Charles Minard's Visualization of Napoleon's 1812 March

<http://www.edwardtufte.com/tufte/minard>



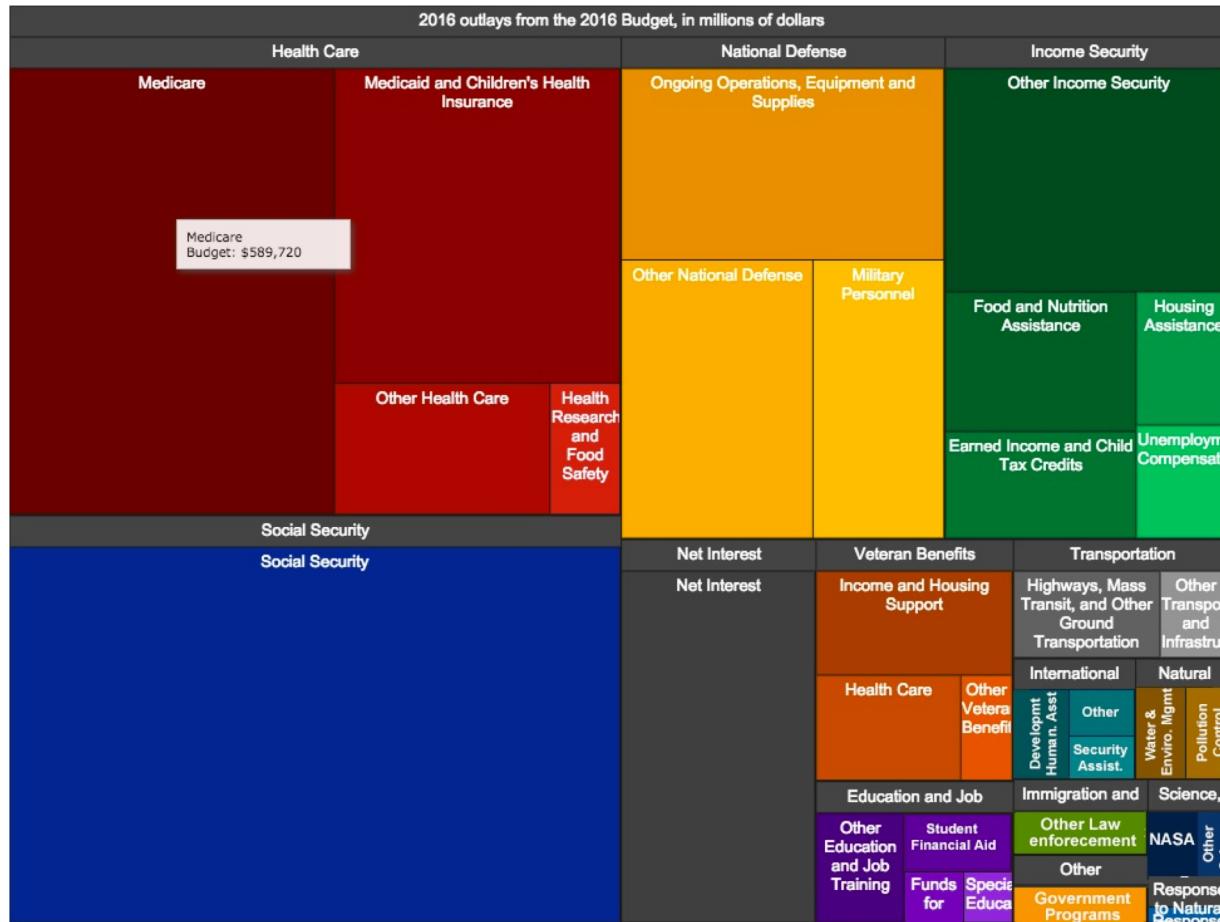
- Classic lithograph dates back to 1869, displaying the number of men in Napoleon's 1812 Russian army, their movements, and the temperatures they encountered along their way. It's been called one of the “best statistical drawings ever created.” The work is an important reminder that the fundamentals of data visualization lie in a nuanced understanding of the many dimensions of data. Tools like D3.js and HTML are no good without a firm grasp of your dataset and sharp communication skills.

1854 Broad Street Cholera Outbreak Map



- The John Snow Cholera map (not the Game of Thrones' Jon Snow) is essentially an early dot map visualization. It uses small bar graphs on city blocks to mark the number of cholera deaths at each household in a London neighborhood. The concentration and length of these bars show a specific collection of city blocks in an attempt to discover why the trend of deaths is higher than elsewhere. The finding: the households that suffered the most from cholera were all using the same well for drinking water.

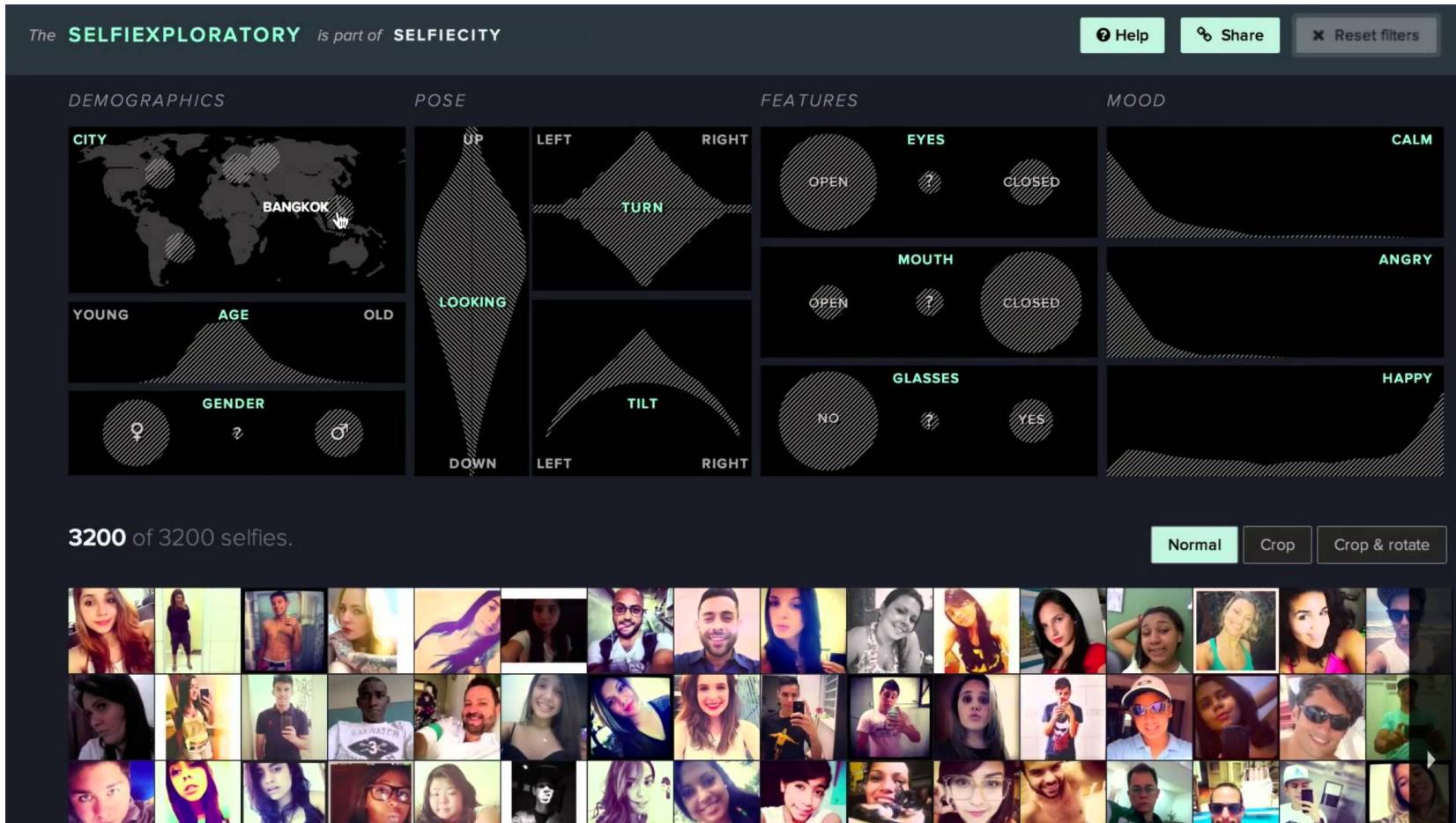
Interactive Government Budget



- All governments, and particularly the USA, have notoriously obscure and tough to understand government budgets. This treemap, created by The White House during Barack Obama's presidency, visually broke down the United State's 2016 budget to put government programs in context.

Selfiecity

<http://selfiecity.net/>



- Selfiecity presents a wide view of selfie data in the context of a transnational phenomenon. 120,000 selfies from around the world are analyzed to study how people take selfies. What's incredible here is just how comprehensive the study is and how seriously it slices and dices every aspect of selfies. We can find trends in everything from head tilt or pose trends by city to smile frequency by age group and gender.

Hans Rosling's 200 Countries, 200 Years, 4 Minutes

https://www.youtube.com/watch?feature=player_embedded&v=jbkSRLYSOjo



- Global health data expert Hans Rosling's famous statistical documentary *The Joy of Stats* aired on BBC in 2010, but it's still turning heads. One segment in particular is pretty mind-blowing. In "200 Countries, 200 Years, 4 Minutes," Rosling uses augmented reality to explore public health data in 200 countries over 200 years using 120,000 numbers, in just four minutes.

Renting vs. Buying

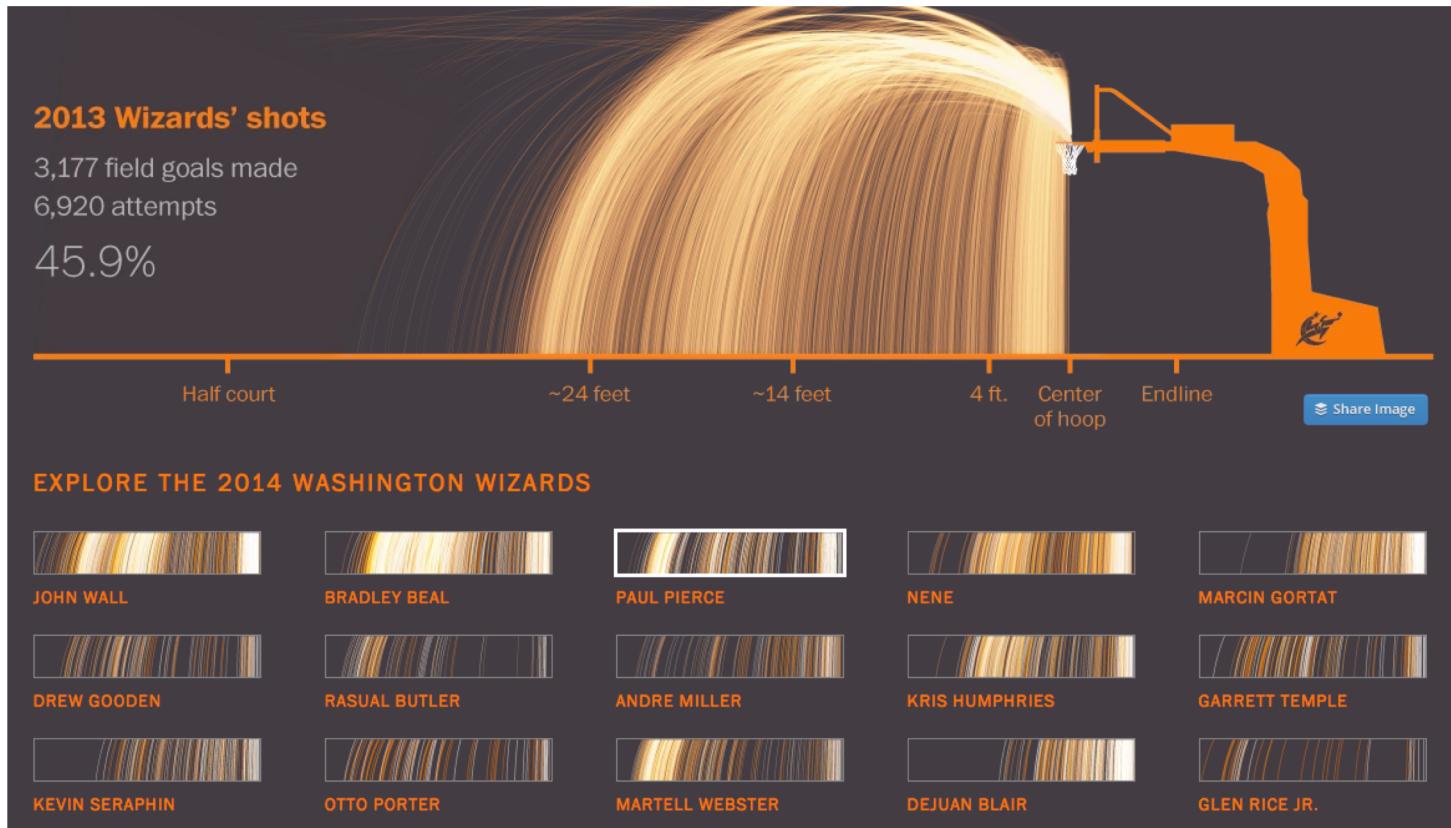
<http://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html>



- Mike Bostock, *New York Times* graphics department editor and inventor of D3.js, built a complex interactive data calculator that offers a cost/benefit analysis for prospective homebuyers. Along with his colleagues Shan Charter and Archie Tse, Bostock tapped into everything from home price and mortgage-interest tax deduction to property tax rate and inflation to help you determine whether to rent or buy a home.

Washington Wizards' Shooting Stars

<http://www.washingtonpost.com/wp-srv/special/sports/wizards-shooting-stars/>



- This detailed data visualization demonstrates D.C.'s basketball team's shooting success during the 2013 season. Using stats released by the NBA, the visualization lets you examine data for each of 15 players. See how successful each person was at a variety of types of shots from a range of spots on the court, compared with others in the league.

Global Carbon Emissions – interactive

<https://www.theguardian.com/environment/ng-interactive/2014/dec/01/carbon-emissions-past-present-and-future->



- This data visualization, based on data from the World Resource Institute's [Climate Analysis Indicators Tool](#) and the Intergovernmental Panel on Climate Change, shows how national CO₂ emissions have transformed over the last 150 years and what the future might hold. Explore emissions by country for a range of different scenarios.

Interactive Facet Search

Jay Urbain, MSOE, CTSI/MCW

New Non-Time Series Dashboard

SEARCH QUERY TOTAL HITS 492

CONCEPTS

- CPT:NUR725 (492)
- CLARITY_MED_ID:68840 (492)
- CPT:IVT3 (234)
- CPT:NUR335 (220)
- CPT:NUR185 (197)
- CPT:9999Z.284 (178)
- CPT:PRO47 (164)
- CLARITY_MED_ID:43691 (136)
- CLARITY_MED_ID:131640 (136)
- ICD10:Z12.11 (111)

CUI CODING

- SNOMEDCT_US:404684003 (492)
- SNOMEDCT_US:34402009 (492)
- SNOMEDCT_US:313268005 (492)
- SNOMEDCT_US:29303009 (492)
- SNOMEDCT_US:281004000 (492)
- SNOMEDCT_US:264598005 (492)
- SNOMEDCT_US:263977005 (492)
- SNOMEDCT_US:252465000 (492)
- SNOMEDCT_US:247347003 (492)
- SNOMEDCT_US:104847001 (492)

AGE CAT

- 50-59 (492)

SEX

- f (492)

RESULTS

Fields age

patient_num ▾
31155524
19319125

View: Table / JSON / Raw

Field	Action	Value
version	Q O #	1602556165364908000
age	Q O #	50
age_cat	Q O #	50-59
concepts	Q O #	ICD10:Z87.891,ICD10:K64.4,ICD10:K44.9,CPT:00740,CLARITY_MED_ID:68840
cui_begin	Q O #	0,12,38,89,215,222,256,275,275,321,321,332,347,347,354,458,462,47
cui_coding_scheme	Q O #	SNOMEDCT_US:73761001,SNOMEDCT_US:71388002,SNOMEDCT_US:71388

CUI TEXT

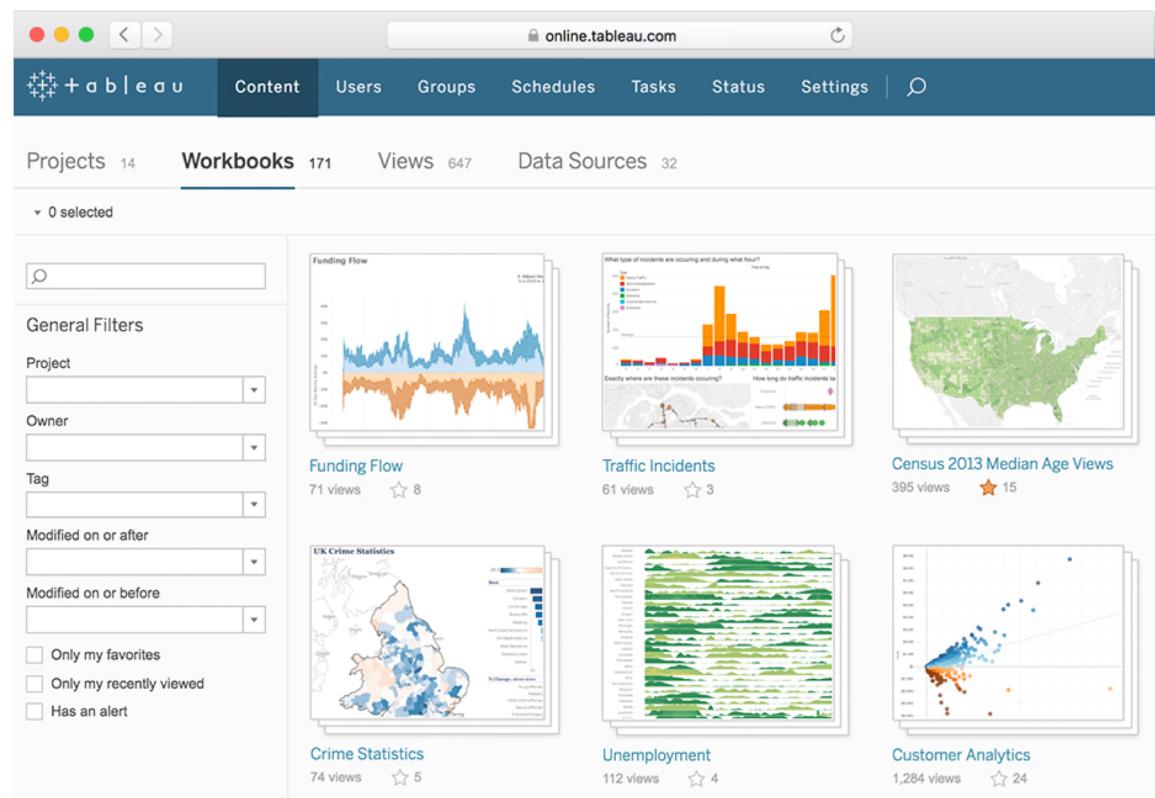
Mucous Membrane
Actual Discomfort

TAGCLOUD

TABLEAU

<https://www.tableau.com/>

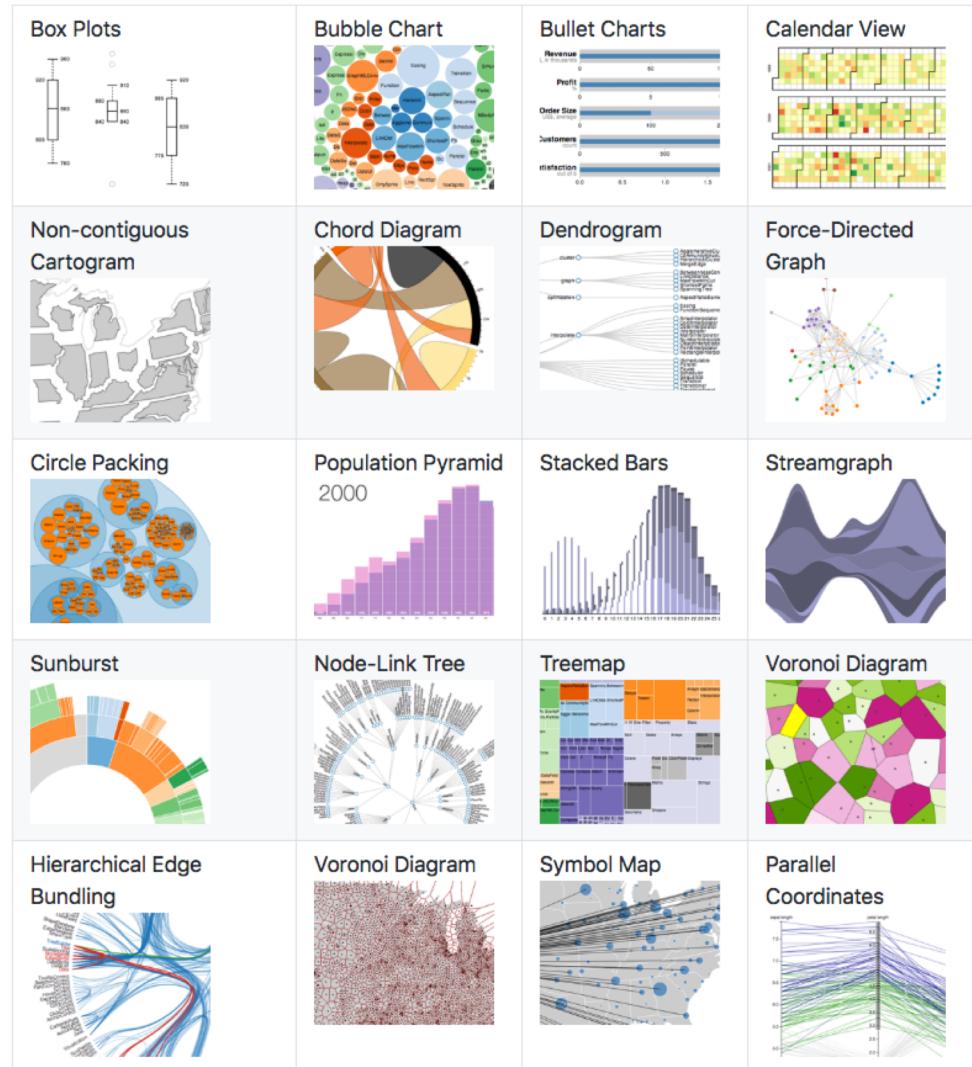
- One of the major tools in this category .
- Tableau is famous for this drag and drop features in user Interface .
- Free for some basic versions .
- Supports multi format data like xls, csv, xml, json , sql data base connections etc .



D3.js

<https://d3js.org/>

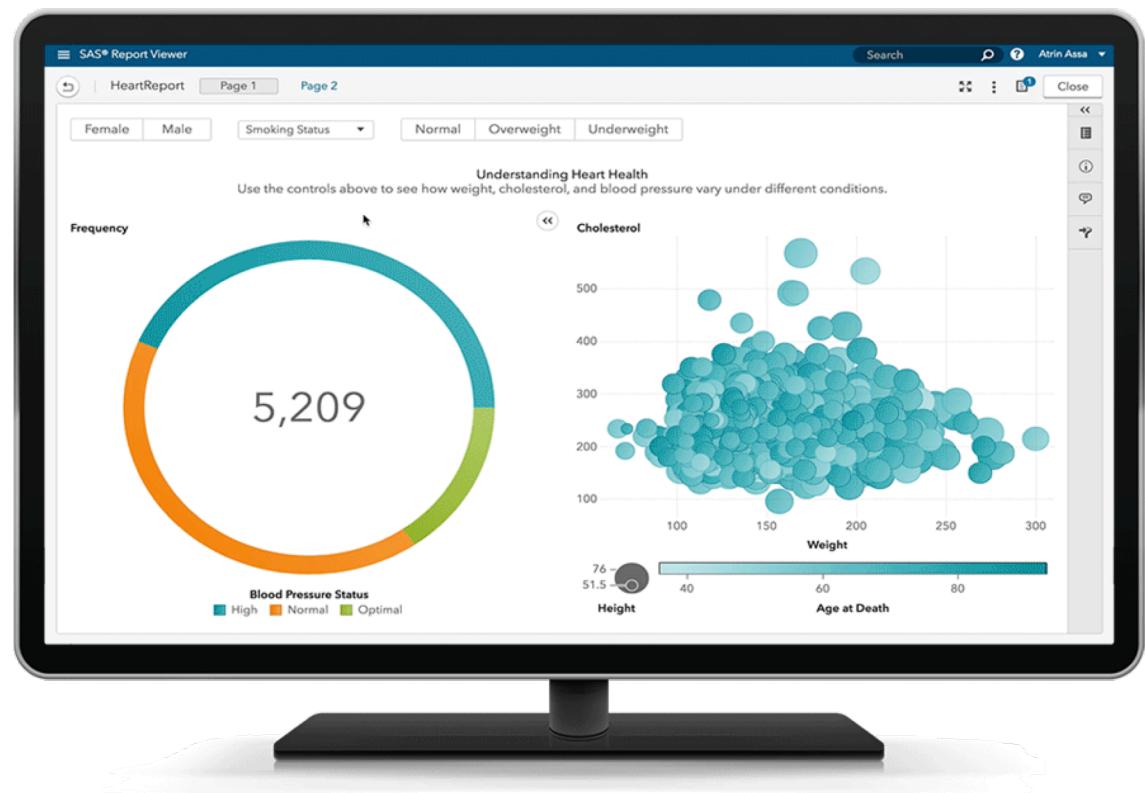
- D3 is java script library .
- Open source.
- Can bind to arbitrary data with Document object Model .
- Nice tutorials on D3.js .



SAS Visual Analytics

https://www.sas.com/en_in/software/business-intelligence/visual-analytics.html

- SAS VA is not only a data visualization tool but it is capable of **predictive modeling and forecasting** .



Qlik View

<https://help.qlik.com/en-US/qlikview/12.1/Content/Home.htm>

Qlik Sense

http://www.qlik.com/us/lp/ppc/qlik-sense-desktop/brand?sourceID1=google&Campaign_Type=Brand&KW=qlik&k_clickid=9e7c003b-78f9-4604-9fb6-8cc4d0a3bf28&gclid=CjwKEAjwqIfLBRCK6vH_rJq7yD0SJACG18frPT8HT4xZL33YdWem2B-YeyzzVrNEQU_i9gudDekpXxoCUFfw_wcB

- Qlik view is a BI tool for decision making .
 - Qlik view is a more user-friendly version with drag and drop.
 - Pay.



Tutorial

Option 1: D3.js

- Complete the introductory tutorial, and two additional tutorials of your choosing.
- <https://github.com/d3/d3/wiki/Tutorials>
- Submit screen shots of each tutorial to Blackboard.

Option 2: Tableau

- Sign-up for Tableau trial, complete the getting started section.
- <https://www.tableau.com/>
- Submit screen shot showing completing of getting started section to Blackboard.