Slaying the Data Dragon

AASL 2013: 11/16/2013, 8:00 – 9:15AM

Debbie Abilock
NoodleTools, Inc.
debbie@noodletools.com
@noodletools
Resources for Math and Visual Literacy here:
http://www.noodletools.com/debbie/literacies/

Kristin Fontichiaro
University of Michigan School of Information
font@umich.edu
@activelearning

Slides: http://bit.ly/fontblog

"dragon-fire" (cropped) by JAVIER DEVILMAN on Flickr. CC-BY
http://flickr.com/photos/98414686@N05/9316434083/

Today is about big data as an emerging, important kind of resource for librarians and learners.

(It's not about how to use student test data. But that's part of the big data story, just not the focus of our talk.)

"MS-DOS Blue Horizontal" by Kjetil Korslien on Flickr. CC-BY-NC. http://flickr.com/photos/kjetikor/8484119632/

How We Came to Care About Big Data

Cheaper by the Dozen

Schenke, Jim. *June 16, 2006: Films from Real-life* Cheaper by the Dozen *Parents Preserved, Released. Purdue University*. Purdue U, 16 June 2006. Web. 21 Oct. 2013. <http://www.purdue.edu/uns/html4ever/2006/060616.Gilbreth.film.html>.

Doubts About Data

The Wright brothers had designed two gliders using the accepted lift and drag equations, Otto Lilienthal's aerodynamic data, and Smeaton's coefficient. Neither glider produced the lift those calculations predicted. Wilbur and Orville felt it was time to perform their own aerodynamic research.

The Wrights examined all the terms in the lift and drag equations. Some values—weight of the craft, wind speed, and wing surface area—could be directly measured, so the Wrights were confident of their accuracy. But the coefficients of lift and drag and Smeaton's coefficient were drawn from the work of others. The brothers focused on these as the possible source of their gliders' poor lift performance.

The Wrights' Bicycle Apparatus

In the fall of 1901, the Wrights began to test the accuracy of the equations they had been using. Tapping their familiarity with bicycles, they created a device to compare the forces acting on two objects with different shapes.

This is how it worked:

- They mounted a free-spinning bicycle wheel horizontally on the front fork of a bicycle.
- To the wheel they attached two objects, as shown here: a flat plate (left) and a model wing (right) patterned after Lilienthal's.
- When they rode the bicycle, the air flowing past the two objects caused the wheel to rotate into an equilibrium position that depended on where the forces on the two objects balanced.

Then, data recording and calculations:

- They recorded the angle of the wing to the airflow at which the air pressure on the wing shape balanced the air pressure on the plate. This is called the *angle of attack*.
- Then they calculated what Lilienthal's data indicated the angle *should* have been.

The Results

The model wing required more than three times the angle of attack than Lilienthal's data predicted. The Wrights concluded that either Lilienthal's data or Smeaton's coefficient—or both—were in error.

Courtesy of The Henry Ford, Dearborn, Mich.

Picture by © Debbie Abilock of *Doubts about Data*. Chart. Washington: Smithsonian. The Wright Brothers: The Invention of the Aerial Age. Smithsonian National Air and Space Museum, Washington. © Debbie Abilock

Understand what the numbers mean…

Study size: Sample size of 10 (N=10)

Results: Objective scoring by blinded investigators indicated that the treatment resulted in improvement in all (100%) of the students. Of students showing overall improvement, 78% were graded as having either excellent or moderate achievement.

What is Big Data? – Kristin's Take

- Mechanical repositories of data
- Huge in scope – more than humans could process
- Computers ingests a lot; then mechanically looks for correlations
- Unearths correlations that are useful, even if you don't know why they are useful
- Cannot evaluate "good" vs. "bad"
- Might turn up connections without explanations (we can hypothesize)

What is Big Data? – Debbie's Take

- **Technology** (built by …?...) ingests vast amounts of data (not a sampling of data)
- **Measure** what was not measured before (but messy)
- **Correlations**, not causations (you figure "why")
- **Drives decision-making** (vs. experience, intuition)
- "Extract new insights or create new forms of value in ways that **change** markets, organizations and the relationship between citizens and governments, and more" (6).

Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton, 2013. Print.

Tradeoffs: Both have biases
Random (clean) vs. Comprehensive (messy)

- Random samples: "**Systematic biases** in the way the data is collected can lead to the extrapolated results being very wrong" (23).
- Suck in as much as feasible (28) a "fishing expedition" (29),"collected passively" (30) and calculate to account for **messy** (40) but correlation increases likelihood of **errors** (33) and maybe you **missed** important sources of data
  - So iterative gathering, e.g., Google translate (39)

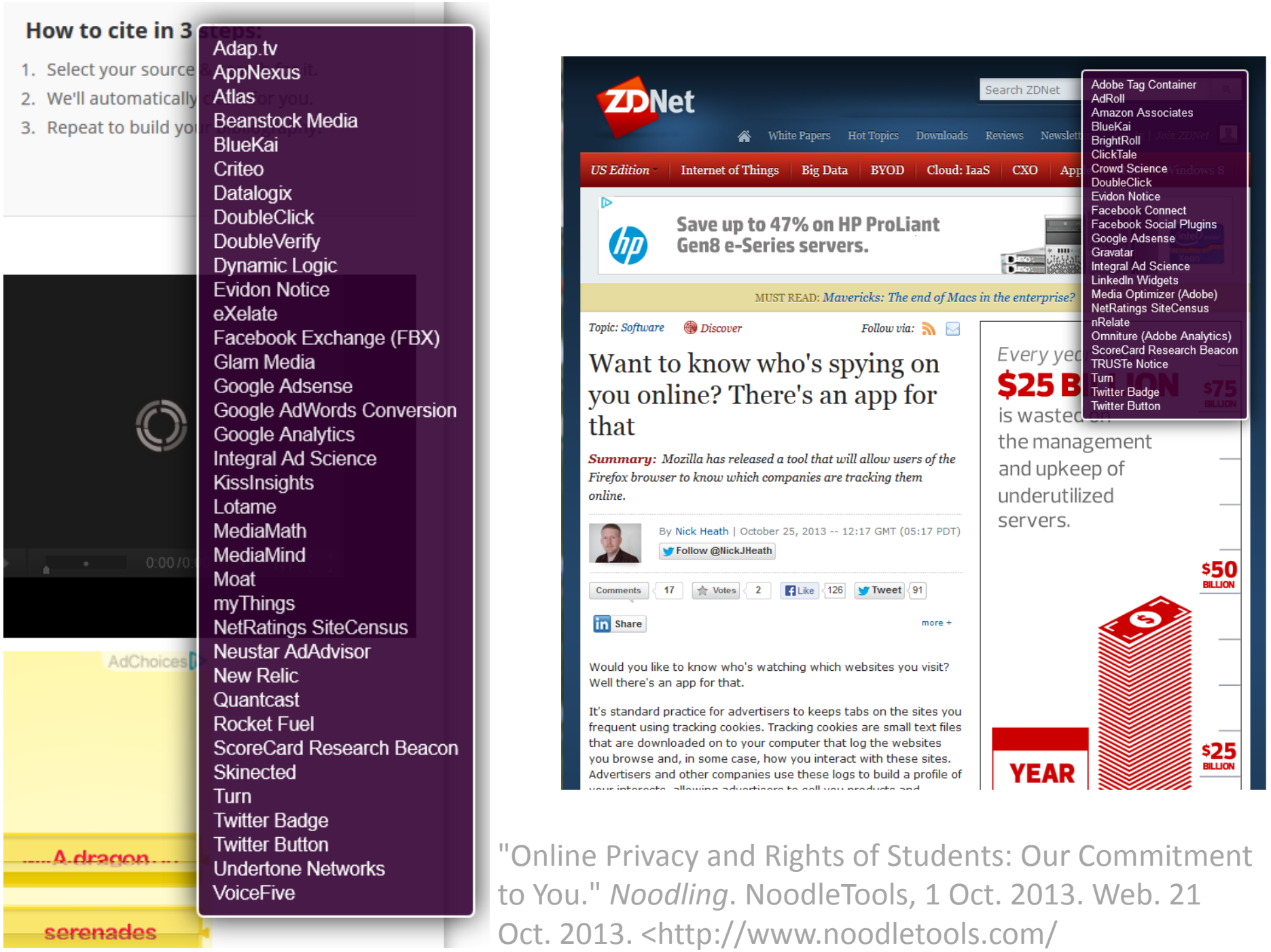
Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton, 2013. Print.

Big data is growing (in 2013)

- Stored information in the world is currently estimated to be around **1200 exabytes**)
- Less than 2% (60 billion gigabytes) is non-digital
- NSA's Utah facility could hold up to **12 exabytes**

Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton, 2013. Print.
Hill, Kashmir. "Blueprints of NSA's Ridiculously Expensive Data Center in Utah Suggest It Holds Less Info than Thought." *Tech*. Forbes, 24 July 2013. Web. 18 Nov. 2013. <http://www.forbes.com/sites/kashmirhill/2013/07/24/blueprints-of-nsa-data-center-in-utah-suggest-its-storage-capacity-is-less-impressive-than-thought/>.

acxiom

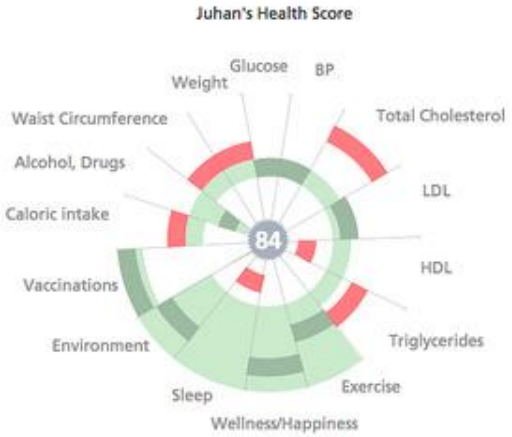


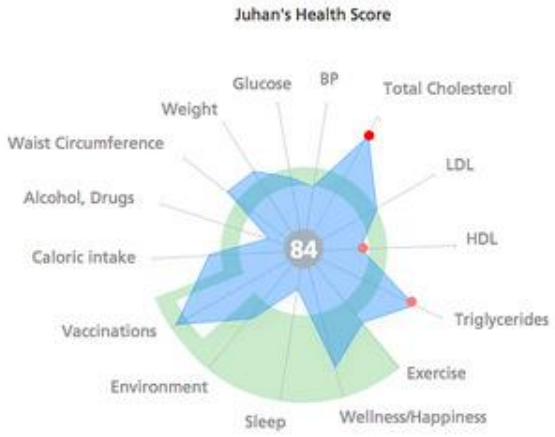
"Online Privacy and Rights of Students: Our Commitment to You." *Noodling*. NoodleTools, 1 Oct. 2013. Web. 21 Oct. 2013. <http://www.noodletools.com/

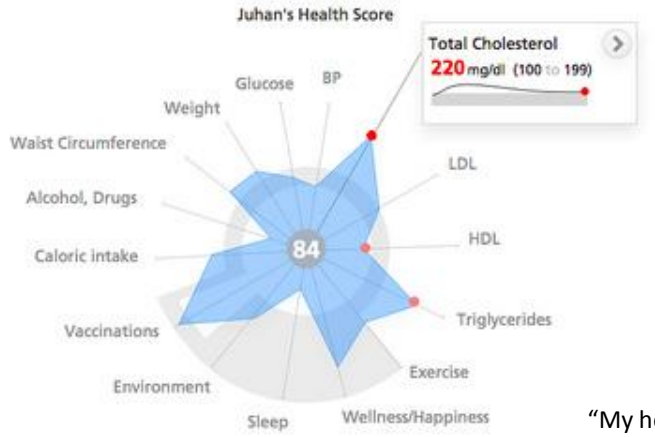
Quantified Self: Personal Data Aggregators

HRECORD
HEALTHSCORE SPIDER IDEAS V4
VISUAL TARGETS

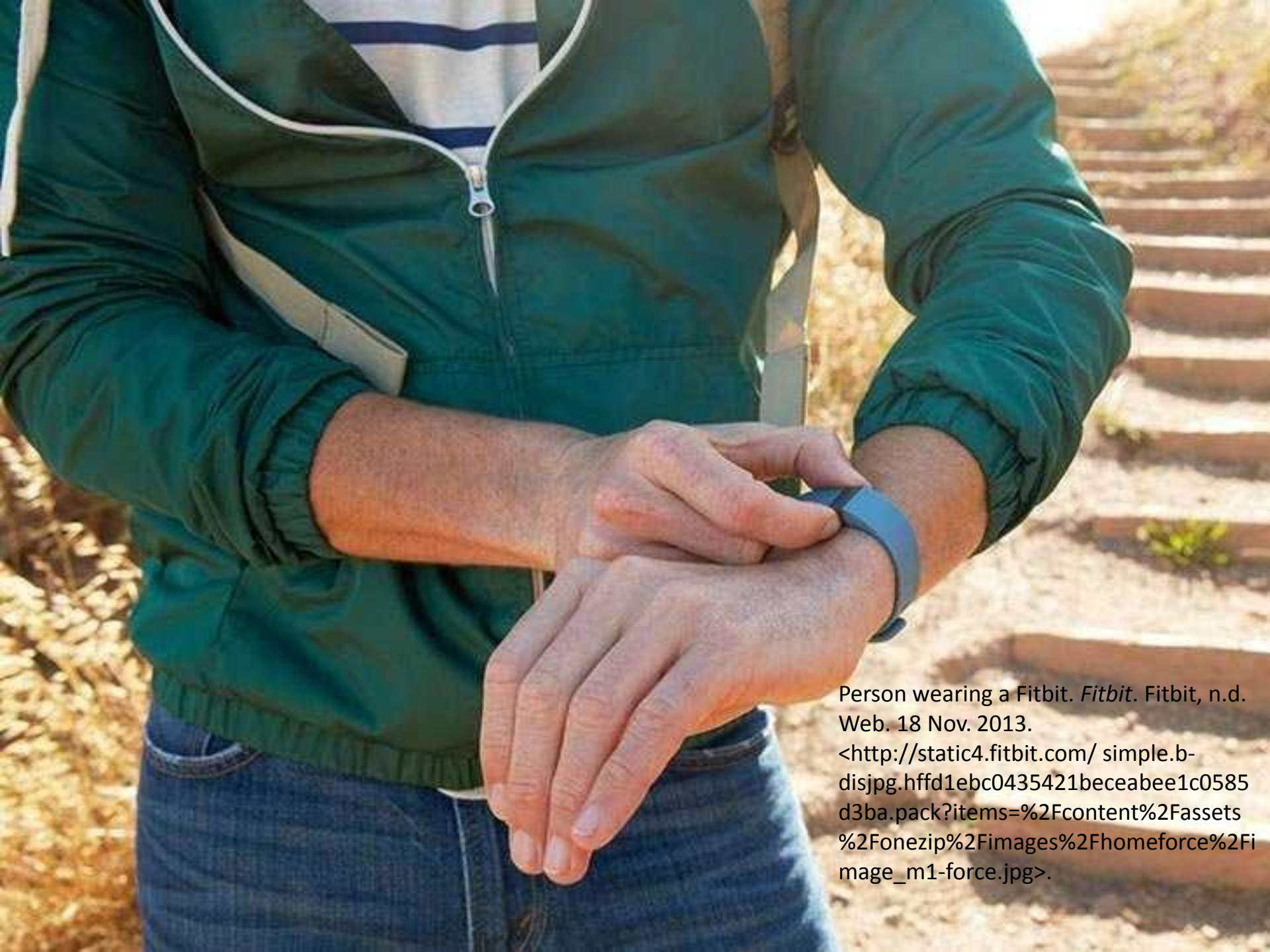
JUHAN SONIN
27.MAY.10







"My health graph v2" by Juhan Sonin on Flickr. CC-BY.
http://flickr.com/photos/juhansonin/4646203016/



Person wearing a Fitbit. *Fitbit*. Fitbit, n.d. Web. 18 Nov. 2013. <http://static4.fitbit.com/ simple.b-disjpg.hffd1ebc0435421beceabee1c0585d3ba.pack?items=%2Fcontent%2Fassets%2Fonezip%2Fimages%2Fhomeforce%2Fimage_m1-force.jpg>.

Geographic Information

Source: IT Enterprise System - GIS

URL: http://bit.ly/LZ2cI

4 | 2 | 0

Tweet | Like | +1

Palo Alto Park Locations

Share | Actions

Palo Alto Trees

Share | Actions

Palo Alto Pavement

Share

Palo Alto Paths and Trails

Palo Alto Zoning

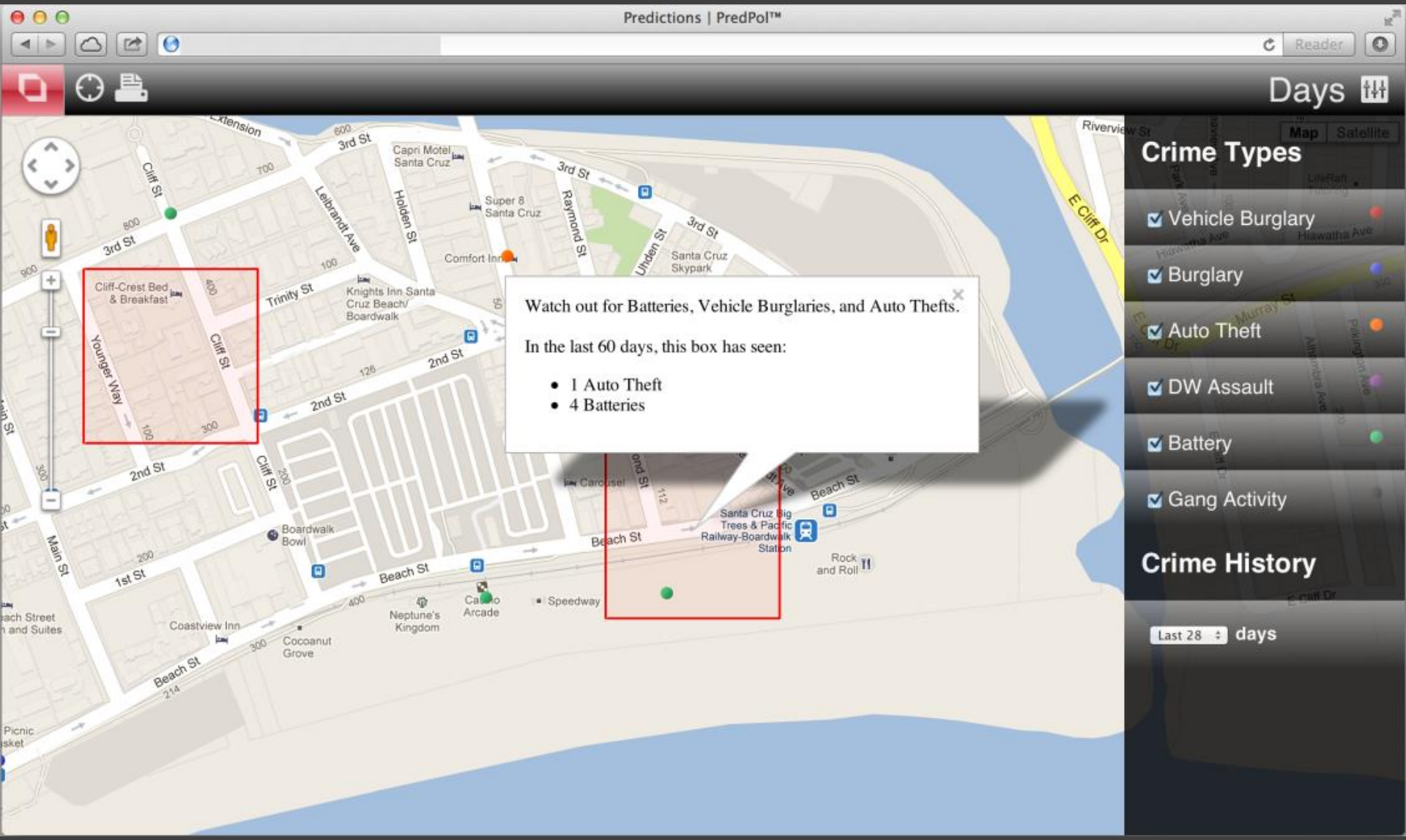
"I have a bad feeling about this" by Rob Ireton on Flickr. CC-BY-NC-SA.
http://flickr.com/photos/aoisakana/3224021222/

Predictive Modeling

A statistical model of future behavior

Hill, Jonah, perf. *Moneyball*. Columbia, 2011. Film.

Spielberg, Steven, dir. *Minority Report*. 20th Century Fox & Dreamworks, 2002 .Film.


Predictions | PredPol™
Reader
Days
Watch out for Batteries, Vehicle Burglaries, and Auto Thefts.
In the last 60 days, this box has seen:
1 Auto Theft
4 Batteries
Map
Satellite
Crime Types
Vehicle Burglary
Burglary
Auto Theft
DW Assault
Battery
Gang Activity
Crime History
Last 28
days
3rd St
2nd St
1st St
Beach St
Cliff St
Main St
Trinity St
Younger Way
Leibrandt Ave
Holden St
Raymond St
Unden St
E Cliff Dr
Riverview St
Capri Motel Santa Cruz
Super 8 Santa Cruz
Comfort Inn
Santa Cruz Skypark
Cliff-Crest Bed & Breakfast
Knights Inn Santa Cruz Beach/ Boardwalk
Carousel
Boardwalk Bowl
Santa Cruz Big Trees & Pacific Railway-Boardwalk Station
Rock and Roll
Neptune's Kingdom
Cocoanut Grove
Coastview Inn
Speedway

Example: Target Can Predict Pregnancy

Forbes ▾ | The Obamacare Exchange Scorecard: Around 100,000 Enrollees And Five... *+53 comments in last 24 hours* ▾ | Searc

Kashmir Hill
Forbes Staff

FOLLOW

Welcome to The Not-So Private Parts where technology & privacy collide
full bio →

Share

TECH 2/16/2012 @ 11:02AM | 2,219,161 views

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

325 comments, 170 called-out + Comment Now + Follow Comments

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy. Target, for example, has figured out how to data-mine its way into your womb, to figure out whether you have a baby on the way long before you need to start buying diapers.

Charles Duhigg outlines in the New York Times how Target tries to hook parents-to-be at that crucial moment before they turn into rampant — and loyal — buyers of all things pastel, plastic, and miniature. He talked to Target statistician Andrew Pole — before Target freaked out and cut off all communications — about the clues to a customer's impending bundle of joy.

CONFERENCES AND MORE

Hill, Kashmir. "How Target Figured out a Teen Girl Was Pregnant before Her Father Did." *Tech*. Forbes, 16 Feb. 2012. Web. 18 Nov. 2013. <http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>. /

Algorithmic Regulation

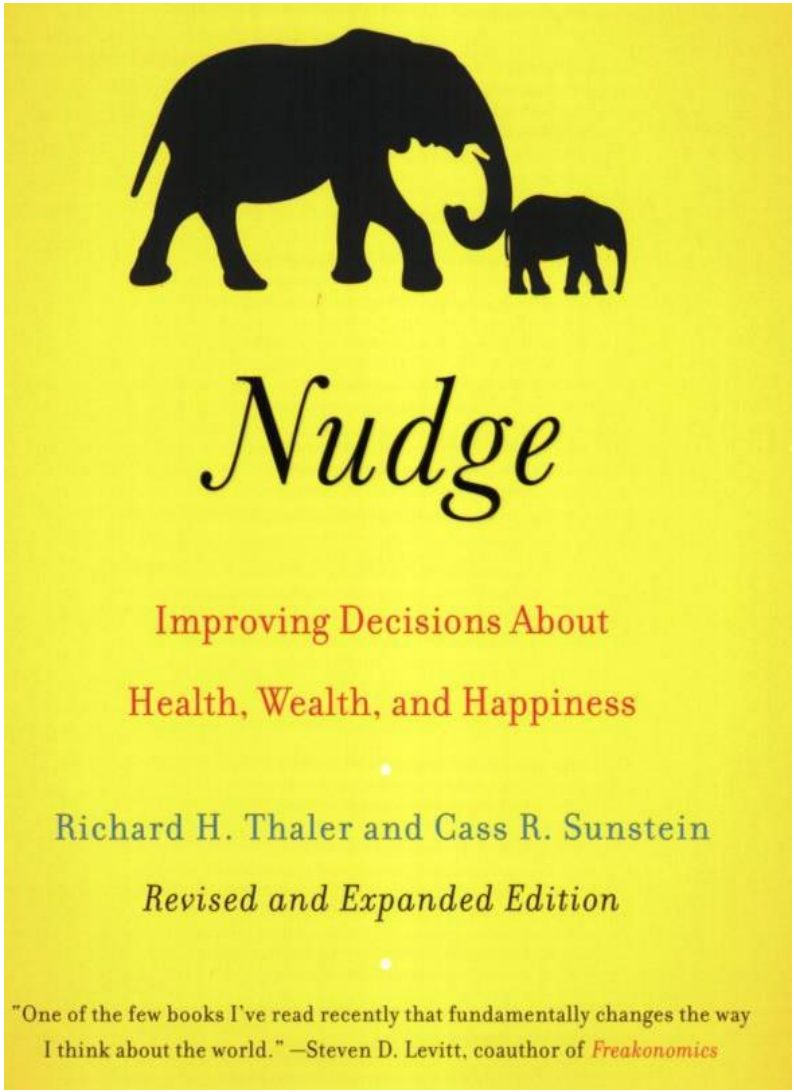
"…solve public problems without having to explain or justify themselves to citizens…with a perfect, highly personalized, irresistible nudge"

Morozov, Evgeny. "The Real Privacy Problem." *MIT Technology Review* Nov.-Dec. 2013: n. pag. *MIT Technology Review*. Web. 18 Nov. 2013. <http://www.technologyreview.com/featuredstory/520426/the-real-privacy-problem/>

"Chicago at Night – From a Distance" by Kevin Dooley on Flickr. CC-BY. http://flickr.com/photos/kdoo/2600517438/

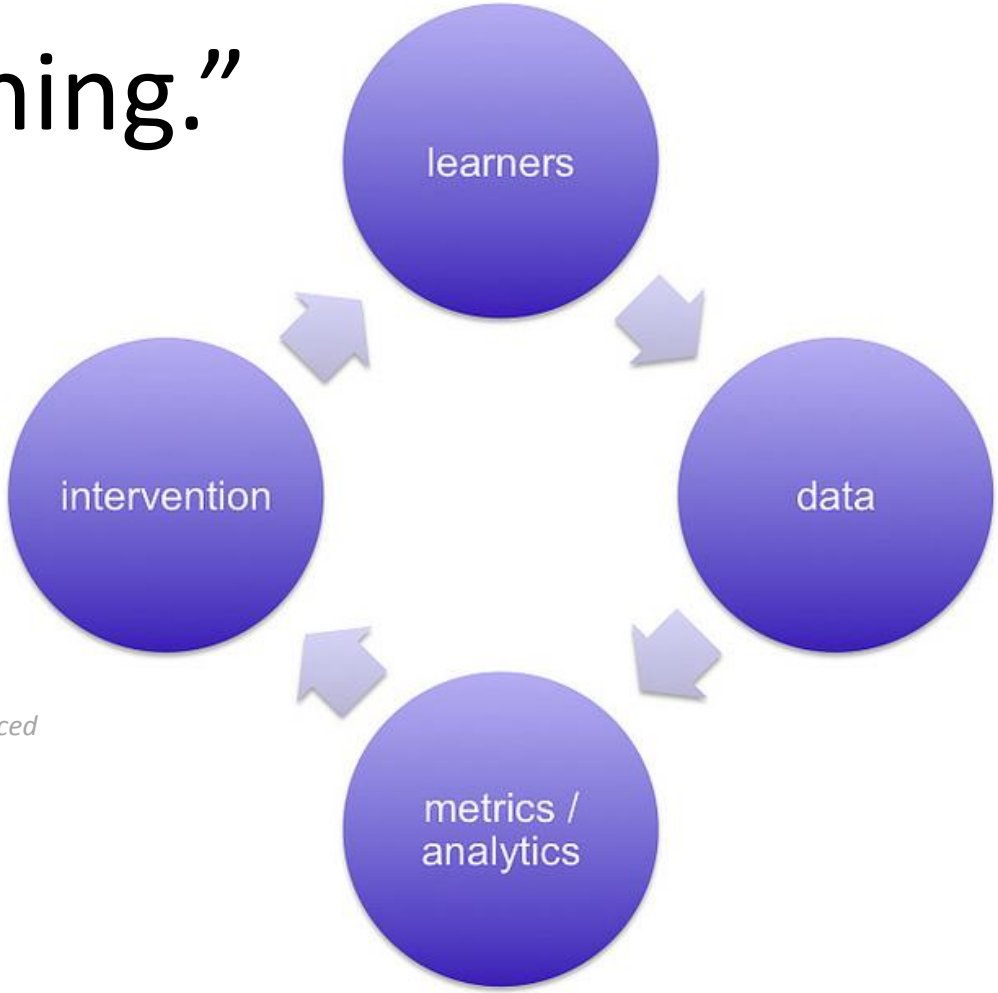
“Nudge”
Bad, evil, paternalistic…or good?

Suzuki, Miwa. "Japan 'Diet Glasses' Fool Wearers into Eating Less." *Google*. Google, 4 June 2012. Web. 18 Nov. 2013. <http://www.google.com/hostednews/afp/article/ALeqM5hWiomAbvJKjULCAP5c0wJwBk9TpQ?docId=CNG.2ef7d92f5a6586fb67dc4c456ff94dae.4b1>.



Learning Analytics

"The collection, analysis and use of data patterns to optimize conditions for improving learning."



Graduate School of Education. *Learning Analytics and Technology Enhanced Assessment (TEA)*. Discussion Paper no. 4. *University of Bristol*. U of Bristol, 2012. Web. 18 Nov. 2013. <http://www.bris.ac.uk/education/research/sites/tea/publications/learninganalytics.pdf>."

The Learning Analytics Cycle" by Doug Clow on Flickr. CC-BY. http://flickr.com/photos/dougclow/5484184568/

Dipping Your Toe in Big Data

- How can your students take advantage of Big Data?
- What Big Data repositories can they access for research?
- How can students participate in as contributors in Big Data projects?

"But it's too cold to go in!" by Squiggle on Flickr. CC-BY-NC-SA.
http://flickr.com/photos/stephenr/2828804233/

google.org Flu Trends

Language: English (United States)

Google.org home

Dengue Trends

Flu Trends

Home

Select country/region

How does this work?

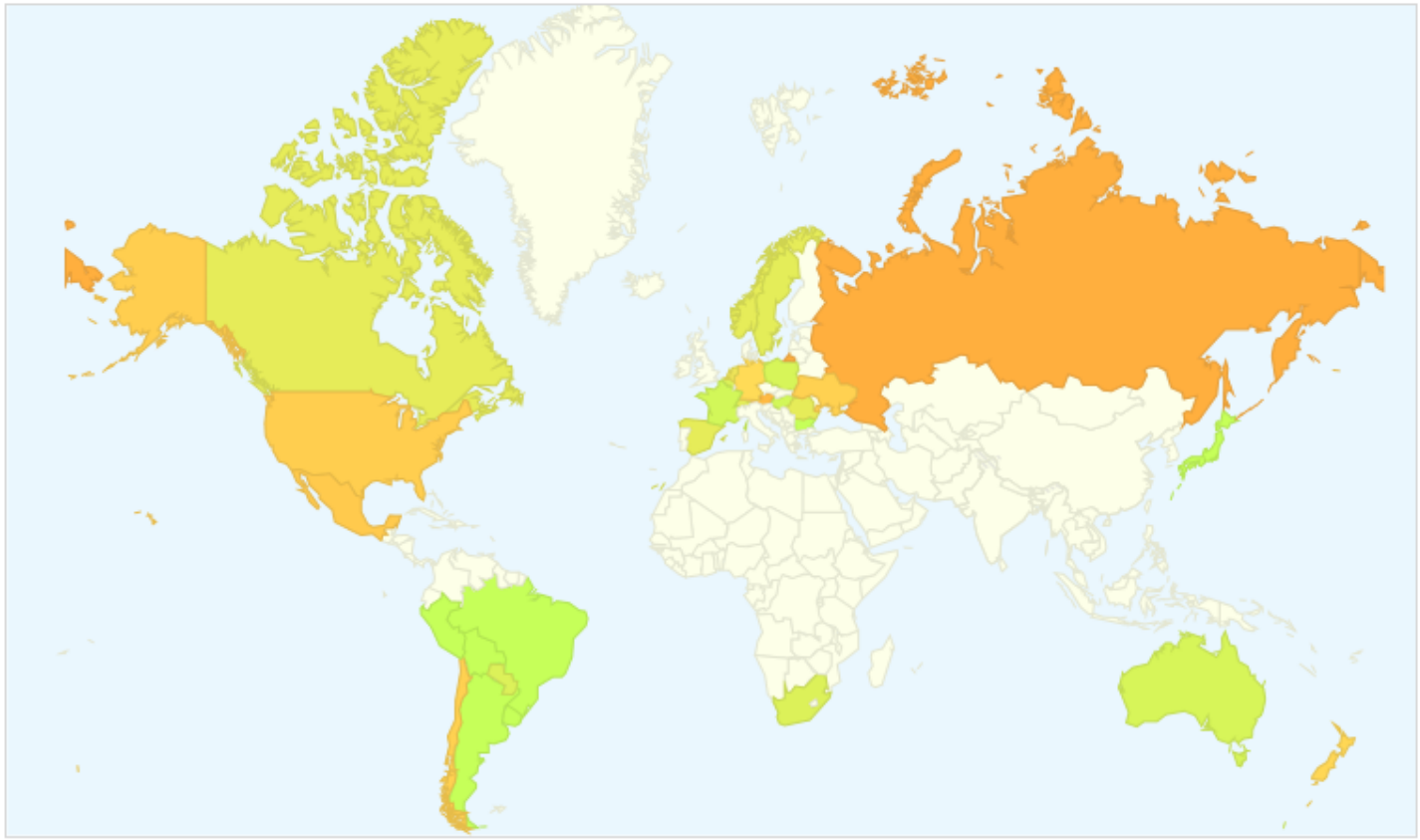
FAQ

Flu activity

- Intense
- High
- Moderate
- Low
- Minimal

Explore flu trends around the world

We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. Learn more »



Download world flu activity data - Animated flu trends for Google Earth - Compare flu trends across regions in Public Data Explorer

BUT Software algorithms may fail when human behavior changes

HealthMap

Global Health, Local Knowledge

What is HealthMap?

Global | Local | News | About | Projects | Mobile | Donate | Feedback | English

Search disease or location GO Advanced Search

Display Options:

Activity Index: Low High — Country level, Province or local level

Alerts from past week | Avian Influenza H7N9 | Coronavirus Surveillance

1430 Alerts Now Showing

Date: 29 Mar 2013 - 6 Apr 2013 | Diseases: all | Places: all | Sources: all | Categories: all

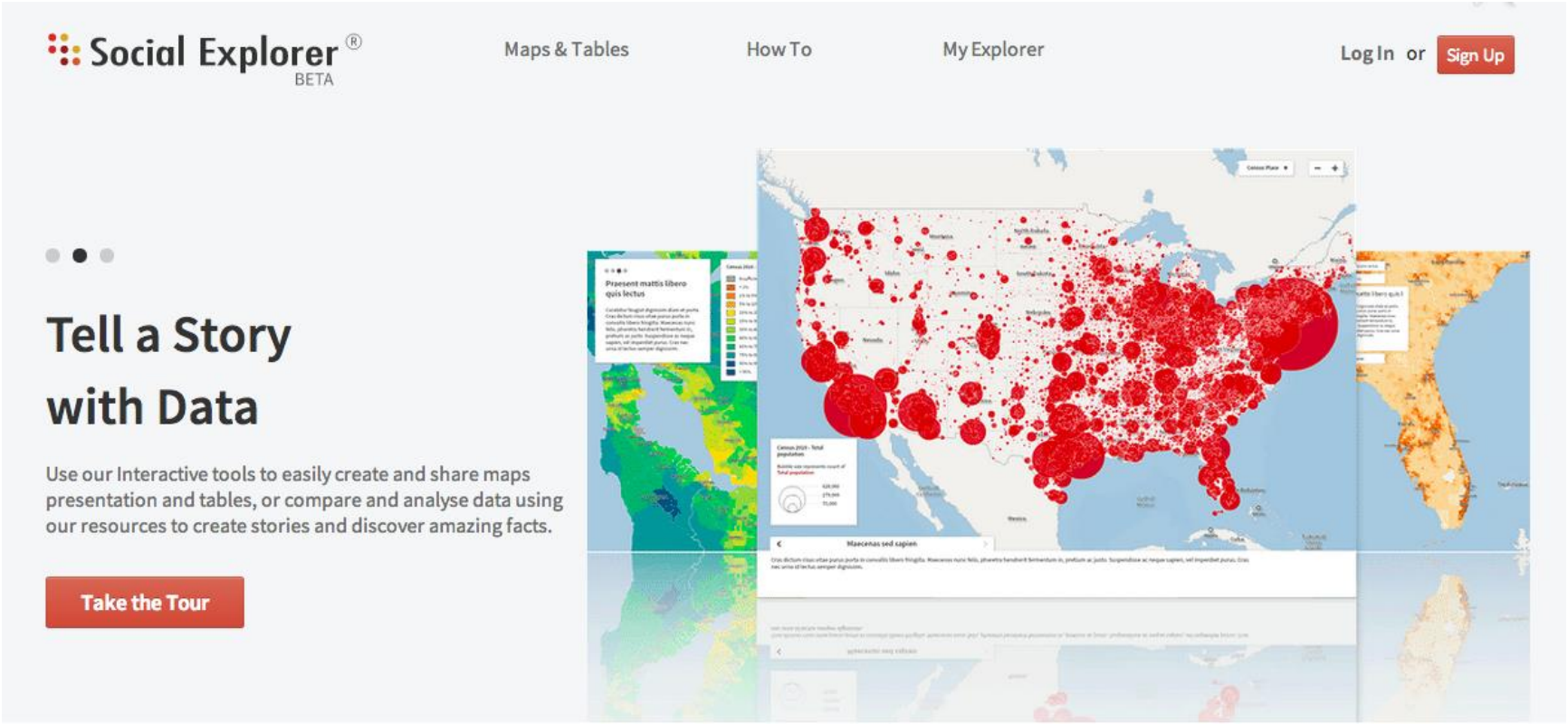
Source	Date	Summary	Disease	Location	Species	Cases	Deaths	Significance
G	5 Apr 2013	Beijing temporarily enhance the epidemic prevention and control level ...	Avian Influenza H7N9	Beijing, China	Humans			1 vote
G	5 Apr 2013	Beijing temporarily enhance the epidemic prevention and control level ...	Avian Influenza	Beijing, China	Humans			1 vote
G	5 Apr 2013	Hong Kong Hospital Authority experts in Shanghai about avian flu - ...	Avian Influenza	Hong Kong	Humans			1 vote
G	5 Apr 2013	Hong Kong Hospital Authority experts in Shanghai about avian flu - ...	Avian Influenza	Shanghai, China	Humans			1 vote
G	5 Apr 2013	Undersecretary of Health visited patient suspected of ... - The ...	Meningitis	Chile	Humans			1 vote

View all

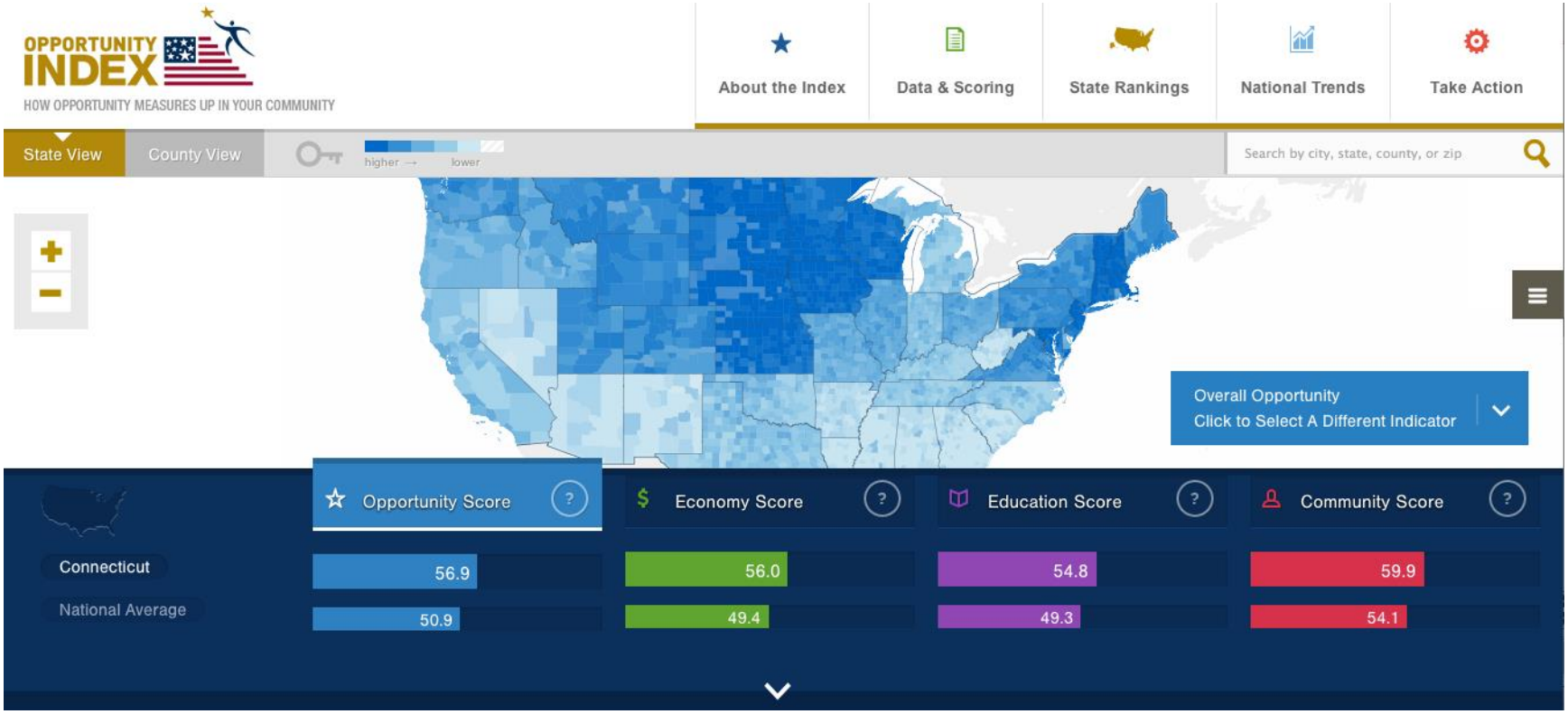
Last update: 5 Apr 2013 02:00 | Next update: 5 Apr 2013 03:00

Boston Children's Hospital

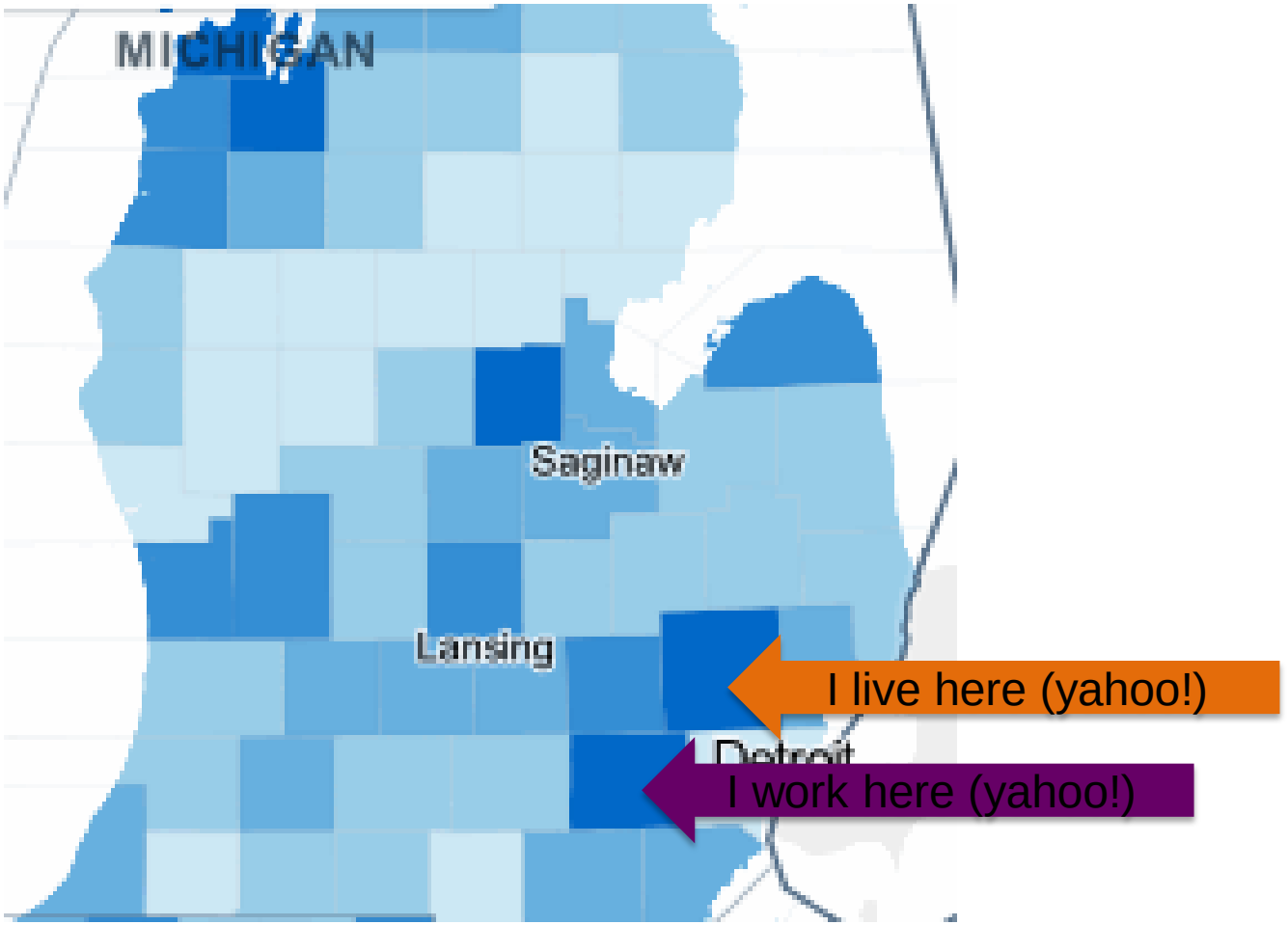
Socialexplorer.com:
Historical census data visualized



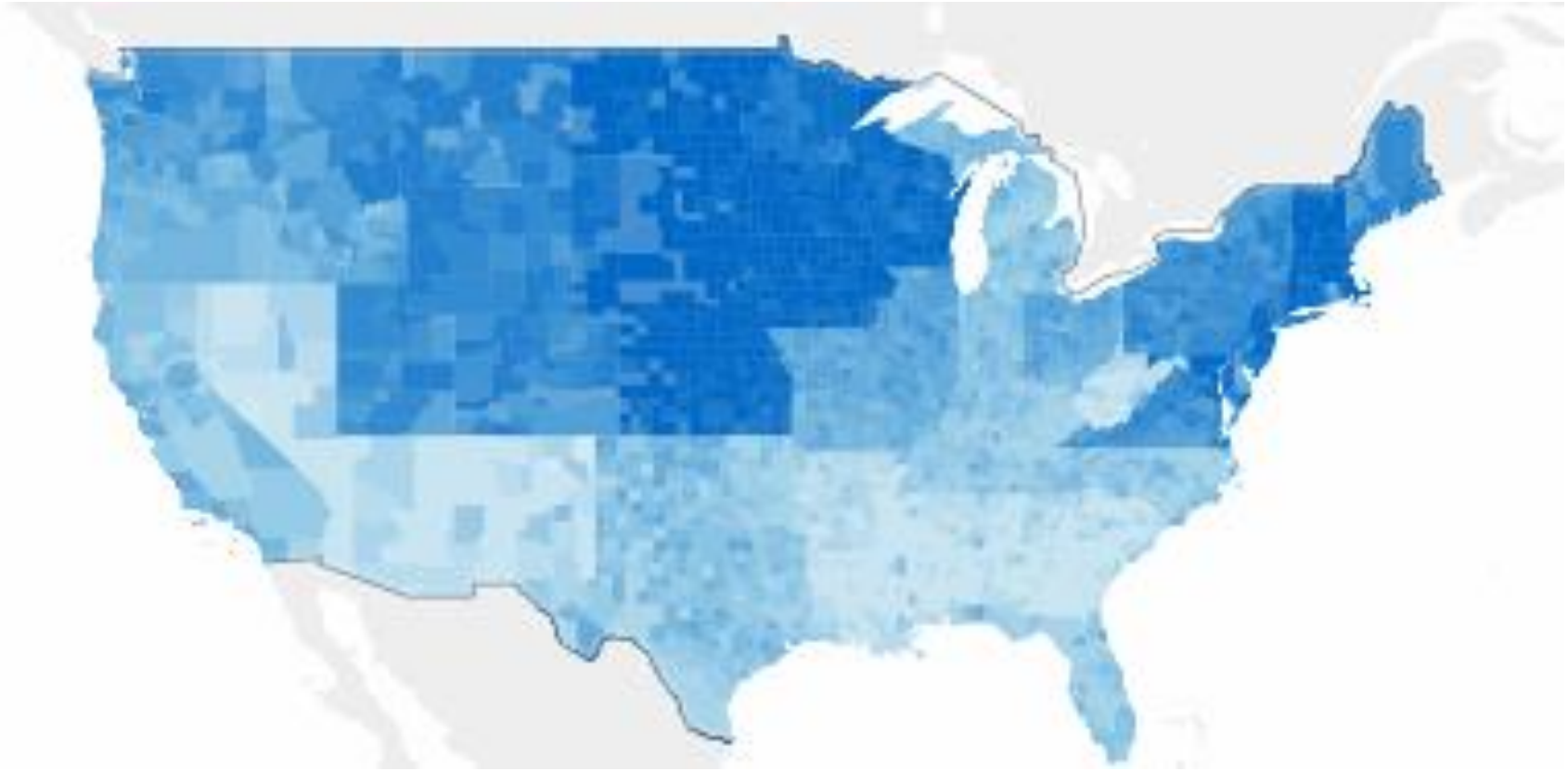
Opportunityindex.org



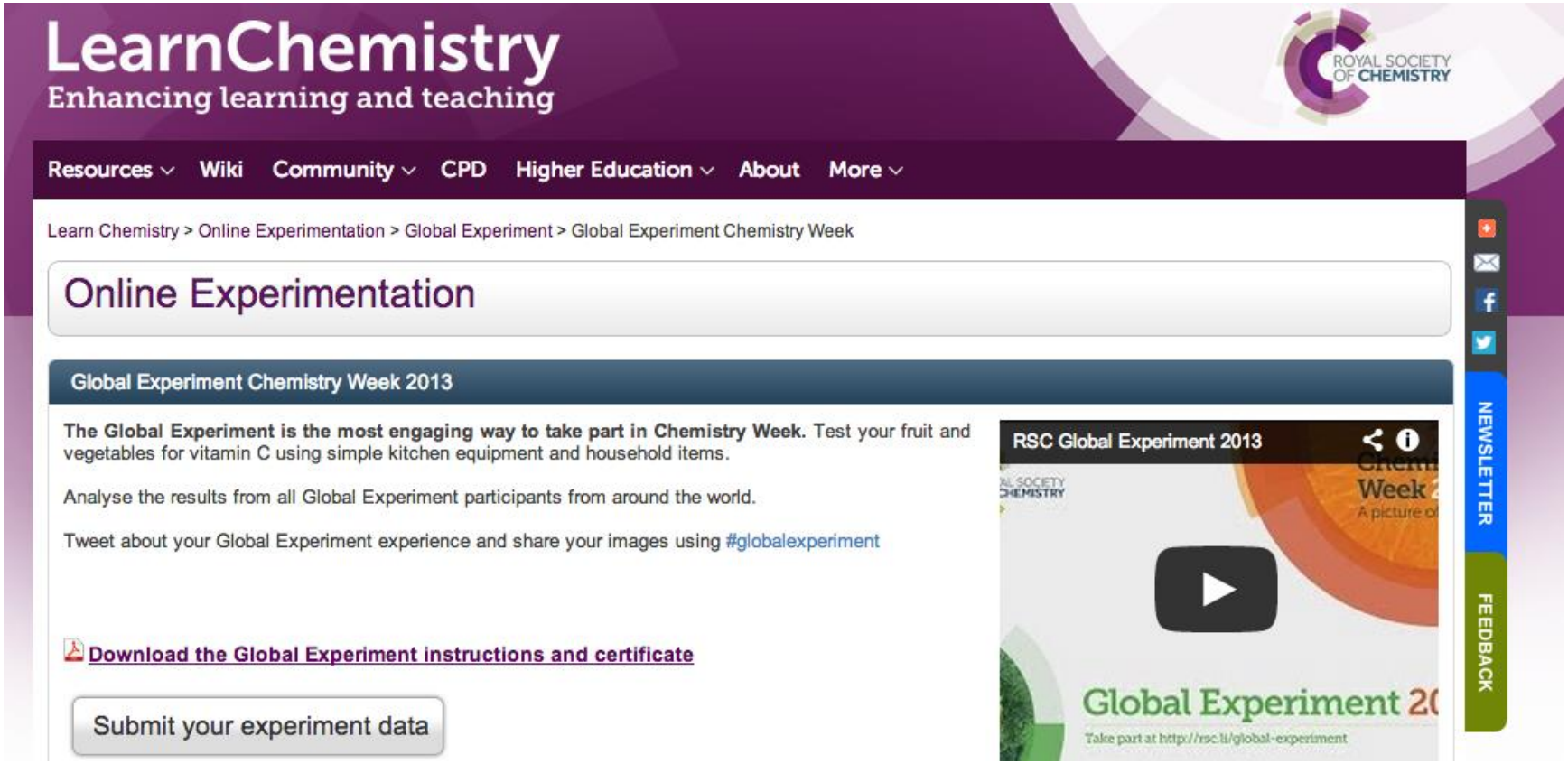
Opportunityindex.org



Opportunityindex.org:
Bigger Sample Size … Bigger Picture



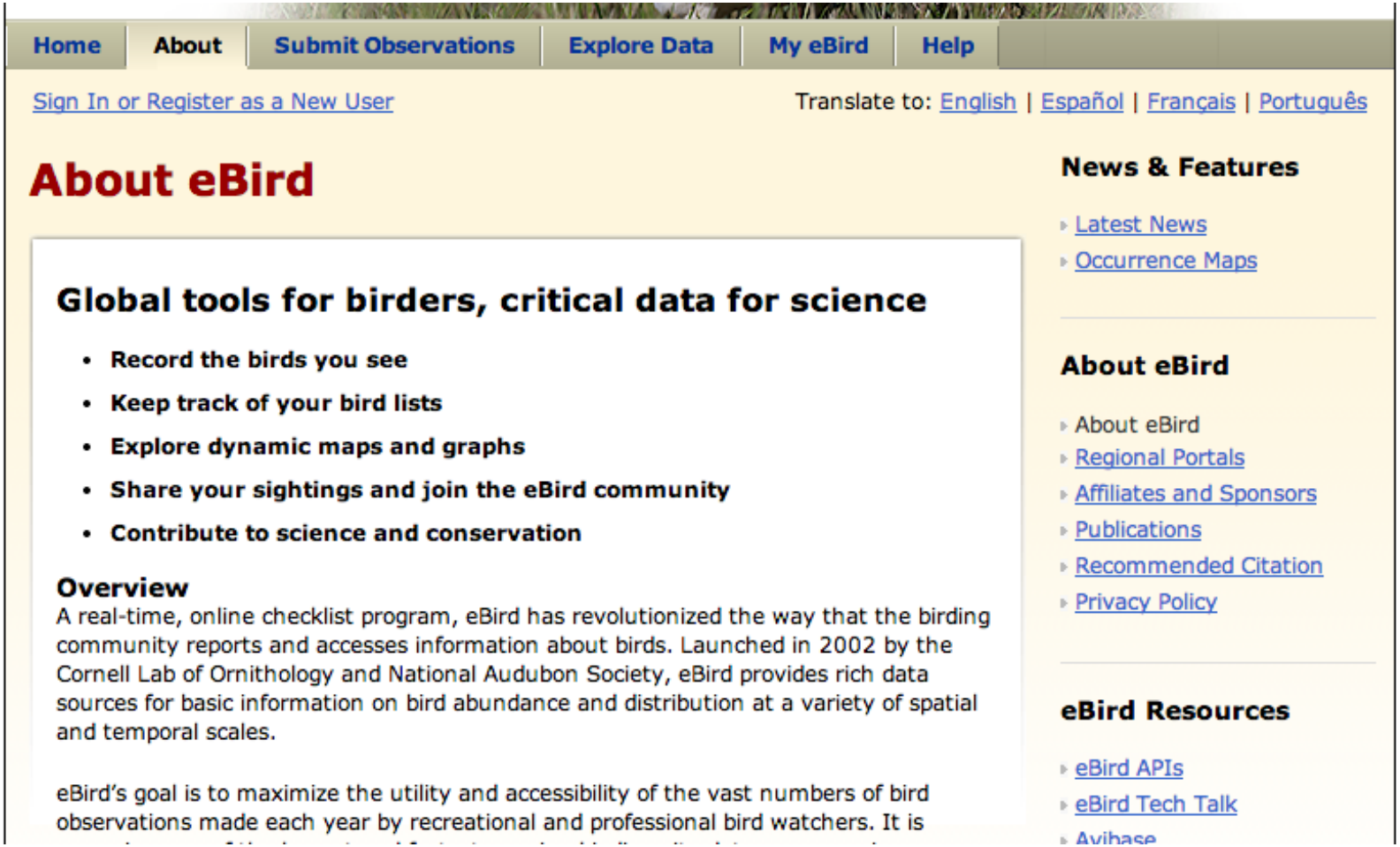
Royal Society of Chemistry: Vitamin C "Global Experiment"



http://www.rsc.org/learn-chemistry/collections/online-experimentation/global-experiment/chemistry-week-2013

Photosynq.org (work in progress)

eBird.org


Home
About
Submit Observations
Explore Data
My eBird
Help
Sign In or Register as a New User
Translate to: English | Español | Français | Português
About eBird
Global tools for birders, critical data for science
Record the birds you see
Keep track of your bird lists
Explore dynamic maps and graphs
Share your sightings and join the eBird community
Contribute to science and conservation
Overview
A real-time, online checklist program, eBird has revolutionized the way that the birding community reports and accesses information about birds. Launched in 2002 by the Cornell Lab of Ornithology and National Audubon Society, eBird provides rich data sources for basic information on bird abundance and distribution at a variety of spatial and temporal scales.
eBird's goal is to maximize the utility and accessibility of the vast numbers of bird observations made each year by recreational and professional bird watchers. It is
News & Features
Latest News
Occurrence Maps
About eBird
About eBird
Regional Portals
Affiliates and Sponsors
Publications
Recommended Citation
Privacy Policy
eBird Resources
eBird APIs
eBird Tech Talk
Avibase

GalaxyZoo.org

Few have witnessed what you're about to see

Experience a privileged glimpse of the distant universe as observed by the SDSS, the Hubble Space Telescope, and UKIRT

GALAXY ZOO QUENCH

We are trying something new! Come help us understand a very specific type of galaxy and experience science from start to end. Take part

Classify Galaxies

To understand how galaxies formed we need your help to classify them according to their shapes. If you're quick, you may even be the first person to see the galaxies you're asked to classify.

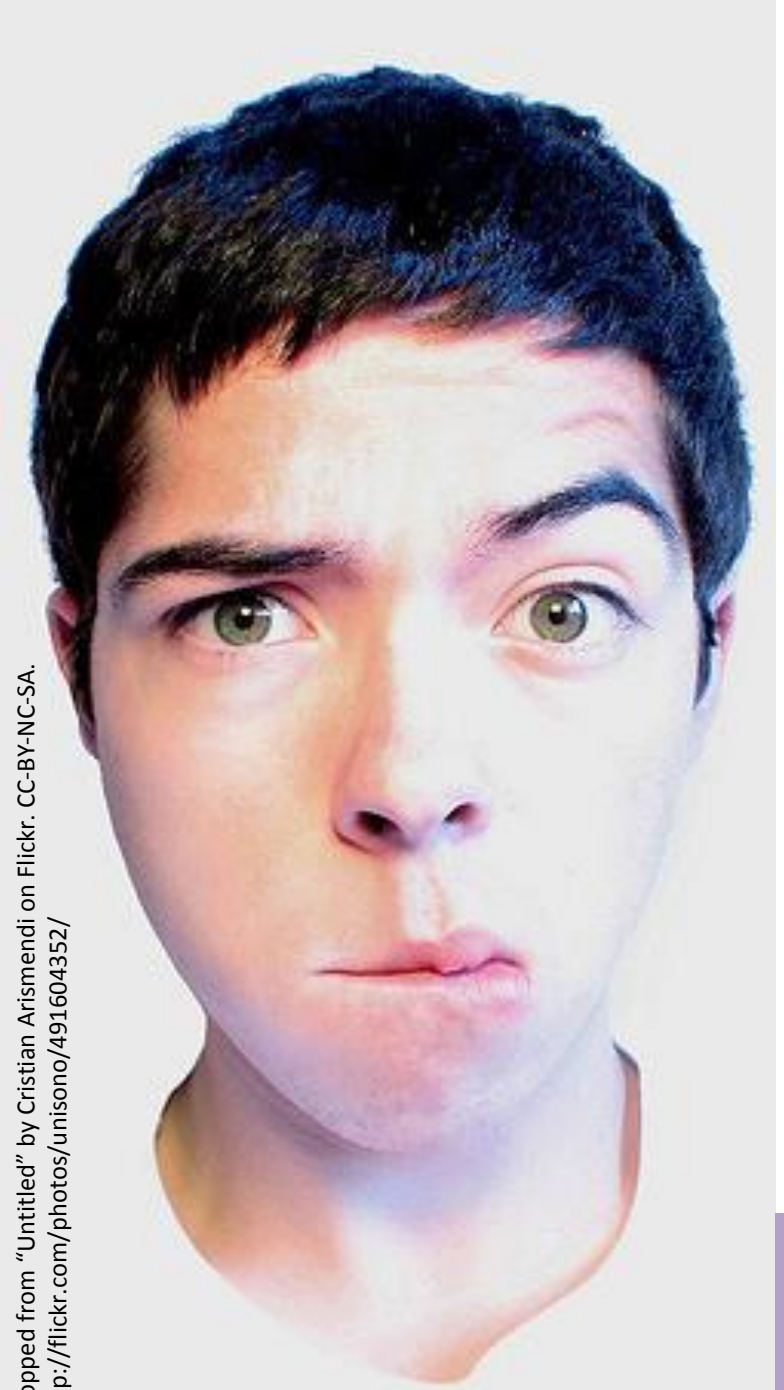
Begin Classifying

How Do Galaxies Form?

Roughly one hundred billion galaxies are scattered throughout our observable Universe, each a glorious system that might contain billions of stars. Many are remarkably

History of Galaxy Zoo

The launch of this new version of Galaxy Zoo, the 4th, comes just a few weeks after the site's 5th birthday. It all started back in July 2007, with a data set made up of a million



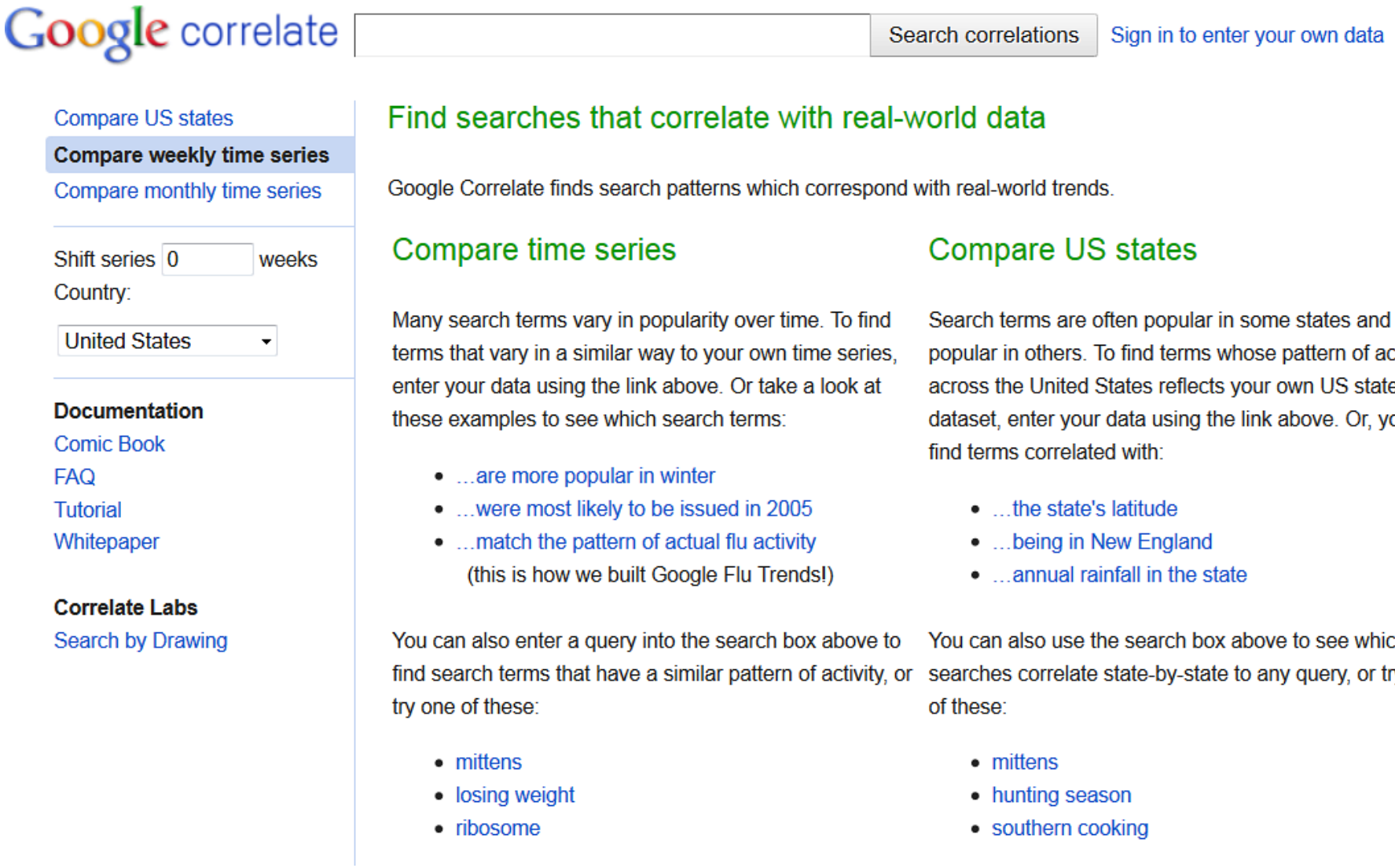
Cropped from "Untitled" by Cristian Arismendi on Flickr. CC-BY-NC-SA.
http://flickr.com/photos/unisono/491604352/

Big Data Isn't Necessarily Neutral Data

- Kowtowing to data is hot
- InBloom
- Amazon recommender

Big Data can be wrong …
and your biases can show up!

Try Correlations



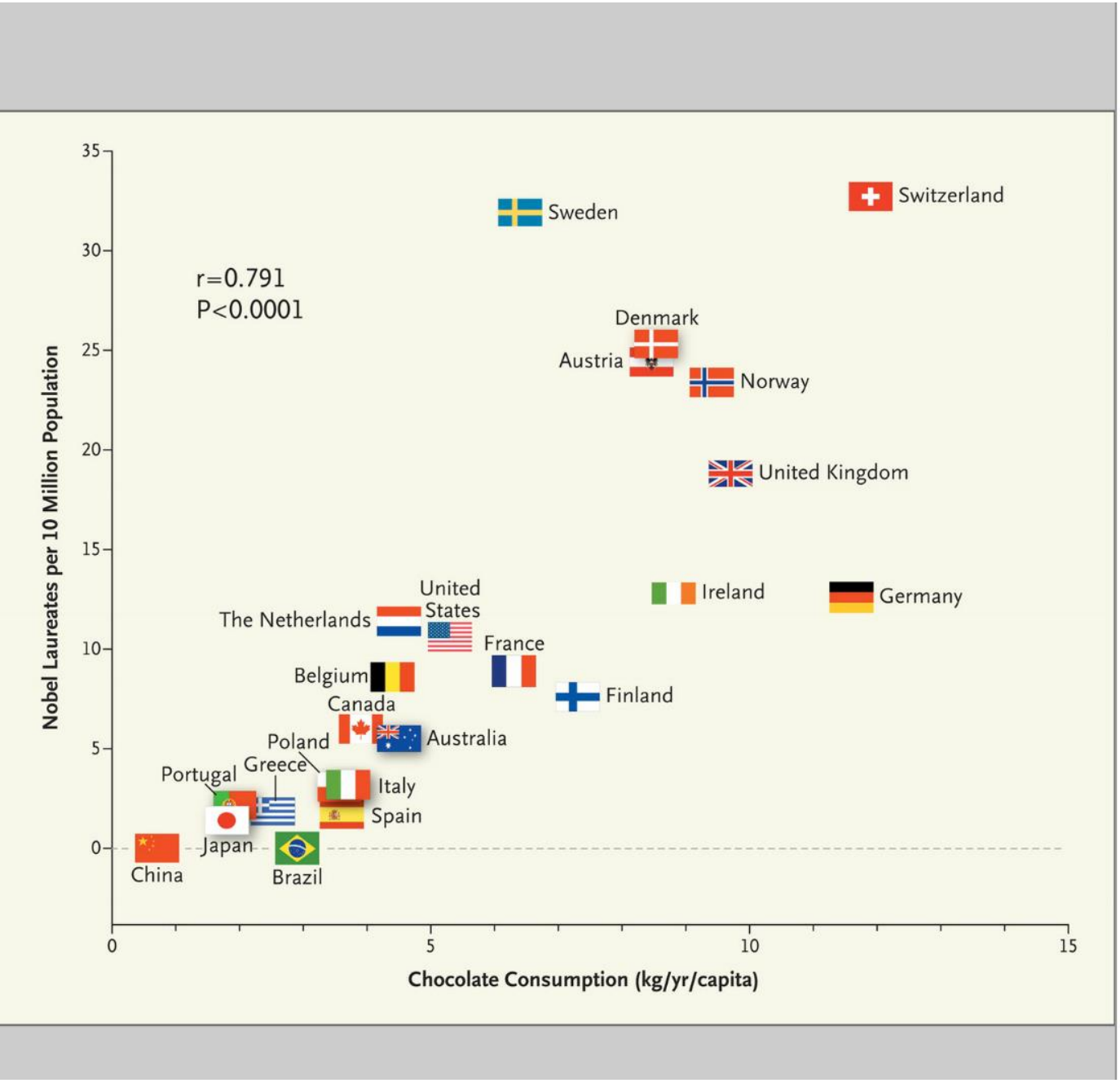
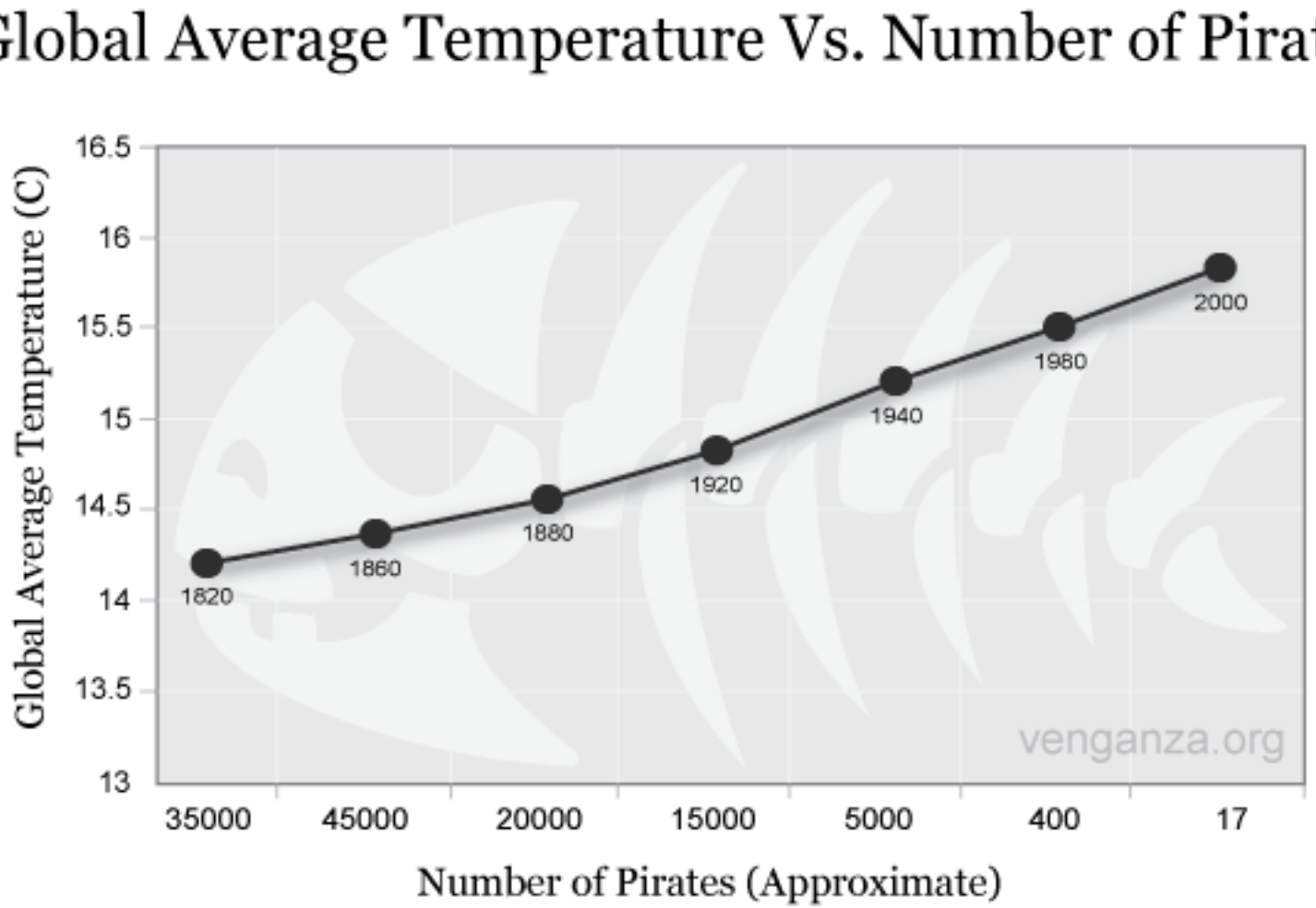


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

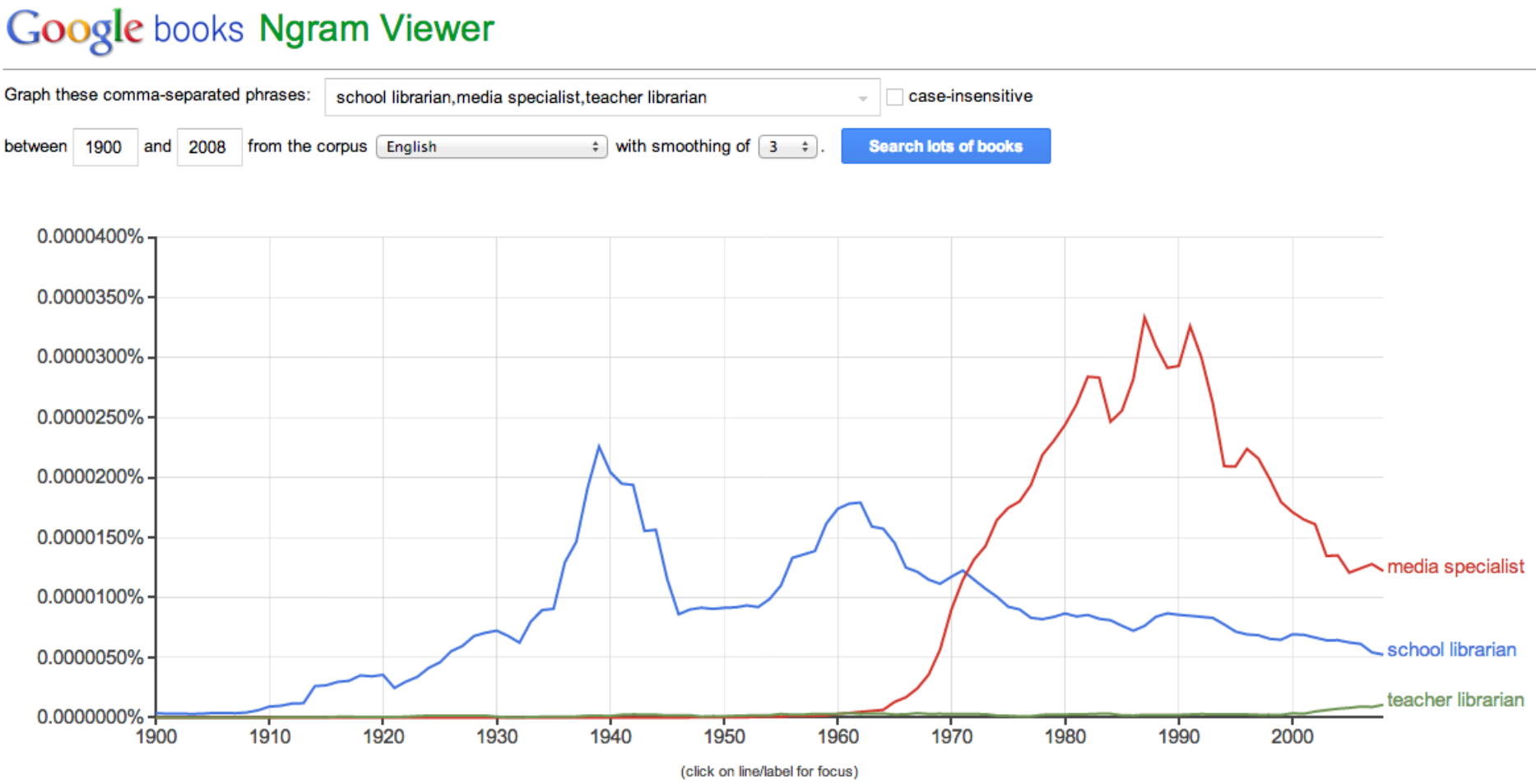
Messerli, Hans. "Chocolate Consumption, Cognitive Function, and Nobel Laureates." *New England Journal of Medicine* 16.367 (2012): 1562-64. *New England Journal of Medicine*. Web. 18 Nov. 2013. <http://www.nejm.org/doi/full/10.1056/NEJMon1211064>.

Really?



Henderson, Bobby. "Open Letter to Kansas School Board." *Church of the Flying Spaghetti Monster*. N.p., n.d. Web. 18 Nov. 2013. <http://www.venganza.org/about/open-letter/>.

https://books.google.com/ngrams



Cautions

- Your first, quick correlation may be wrong
- Don’t base decisions on averages

$$\mathcal{A}\ \mathcal{A}\ \mathcal{A}\ \mathcal{A}\ \mathcal{F} = 80\%$$

Parry, Marc. "What Sports Can Teach about the ‘Cautionary Side’ of Big Data." *Wired Campus*. Chronicle of Higher Education, 31 Oct. 2013. Web. 31 Oct. 2013. <http://chronicle.com/blogs/wiredcampus/ what-sports-can-teach-about-the-cautionary-side-of-big-data/47937>.

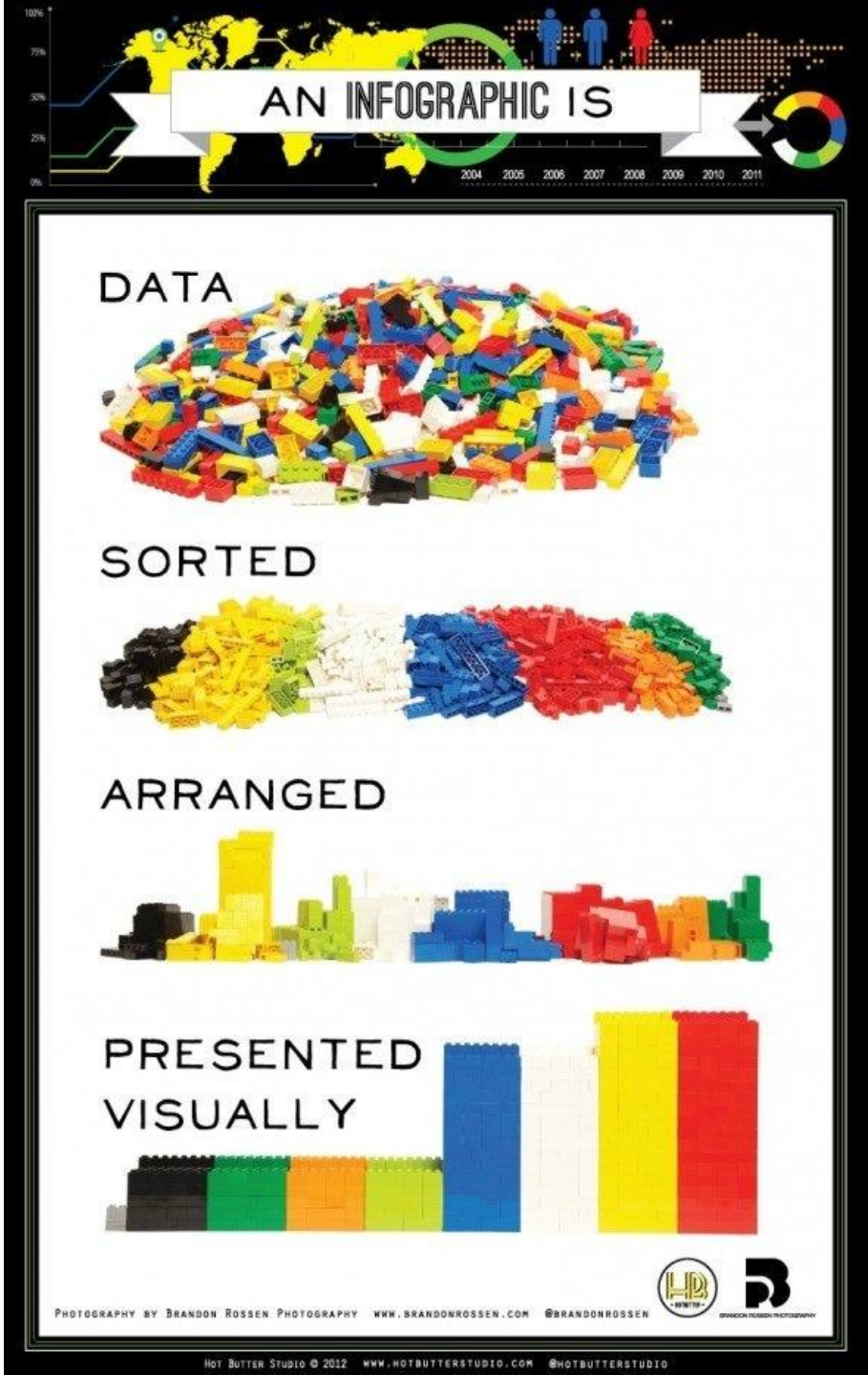
Visual Literacy Competency Standards for Higher Education

"Visual literacy is a set of abilities that enables an individual to effectively find, interpret, evaluate, use, and create images and visual media. Visual literacy skills equip a learner to understand and analyze the contextual, cultural, ethical, aesthetic, intellectual, and technical components involved in the production and use of visual materials. A visually literate individual is both a critical consumer of visual media and a competent contributor to a body of shared knowledge and culture."

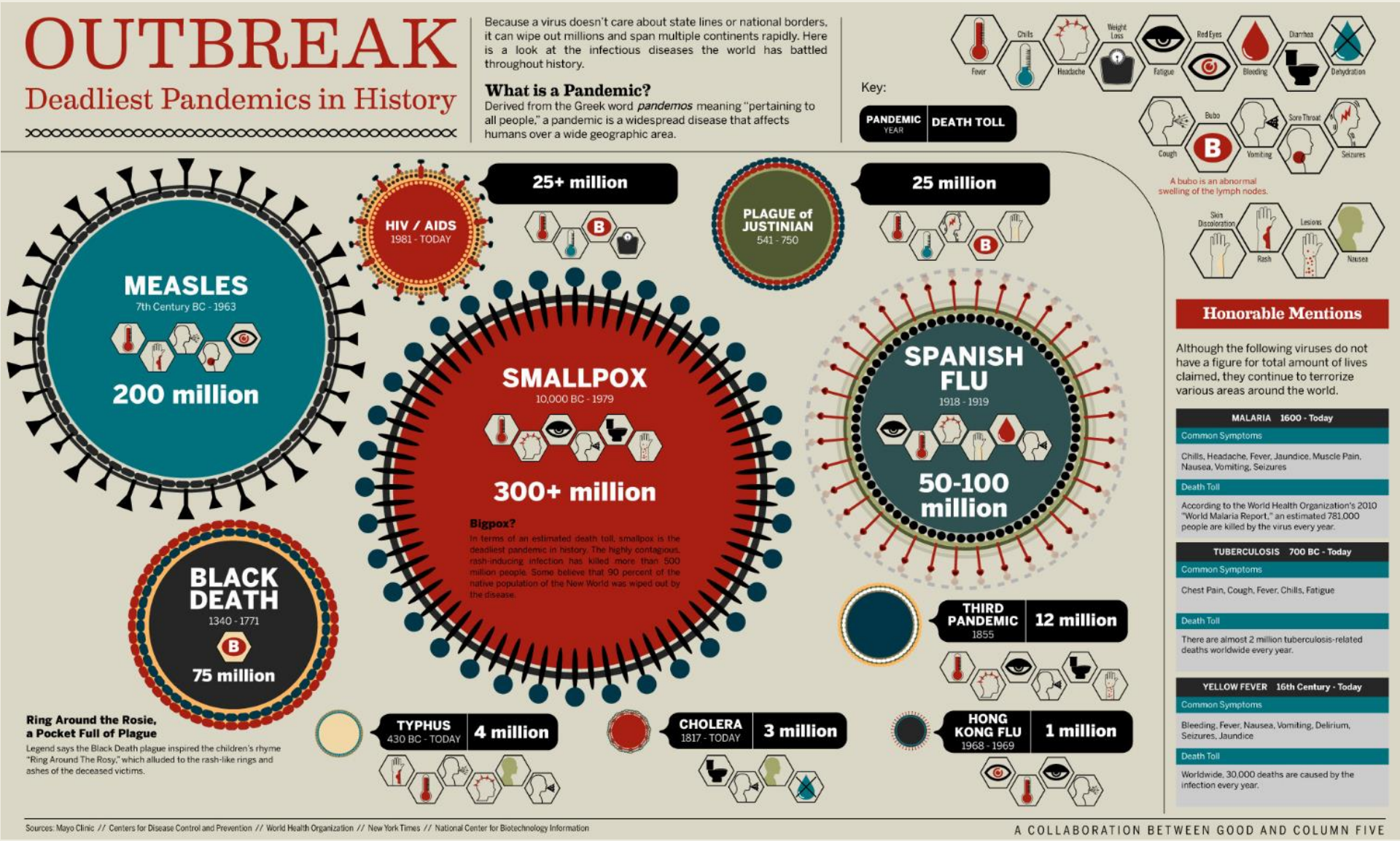
Hattwig, Denise, et al. "ACRL Visual Literacy Competency Standards for Higher Education." *American Library Assn.* ALA, n.d. Web. 18 Nov. 2013. <http://www.ala.org/acrl/standards/visualliteracy>.

Teach students how to visualize data

- Arrange data spatially
- To help us understand
- Something useful
- That we couldn't see from the numbers alone

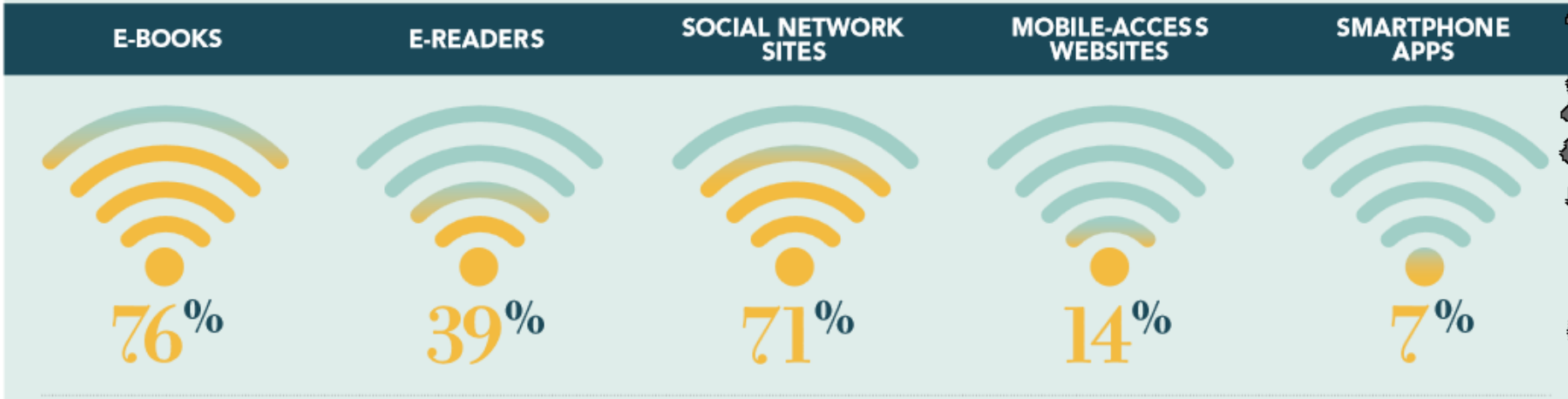


"What Is an Infographic?" Infographic. *HotButter*. Hot Butter Studio, 2012. Web. 18 Nov. 2013. <http://www.hotbutterstudio.com/#/alps/>.



"Outbreak: The Deadliest Pandemics in History." Infographic. *GOOD*. GOOD Worldwide, 11 Aug. 2011. Web. 29 Dec. 2012. <http://awesome.good.is/transparency/web/1108/deadliest-pandemics/flash.html>.

FIGURE A-2: U.S. PUBLIC LIBRARIES: TRENDS IN E-BOOKS AND E-ACCESS



Percentage of libraries that offer services, FY2012

Museum and Library Services (IMLS) released *Building Digital Communities*[2] designed to raise awareness about the access essential to digital communities and identify goals and strategies for focus in digital inclusion efforts (e.g., economic and workforce development, education and civic engagement). Last year, the Department of Commerce's National Telecommunications and In-

and the general public, and for marketing purposes.

- Almost half of public libraries (45.6 percent) report using communication tools (e.g., Blogger, WordPress, Vox, Twitter) to reach the public.
- Over a third (37.3 percent) report using photography sites (e.g., Flickr, Zoomr).

enced a notable uptick in smartphone market penetration over the last year.

David Lee King, digital branch and services manager, Topeka and Shawnee County (KS) Public Library, reported in his April 5, 2012 blog post that in March 2012, slightly more than 11 percent of their library website visits were via a mobile device. Referencing the Pew smartphone surve

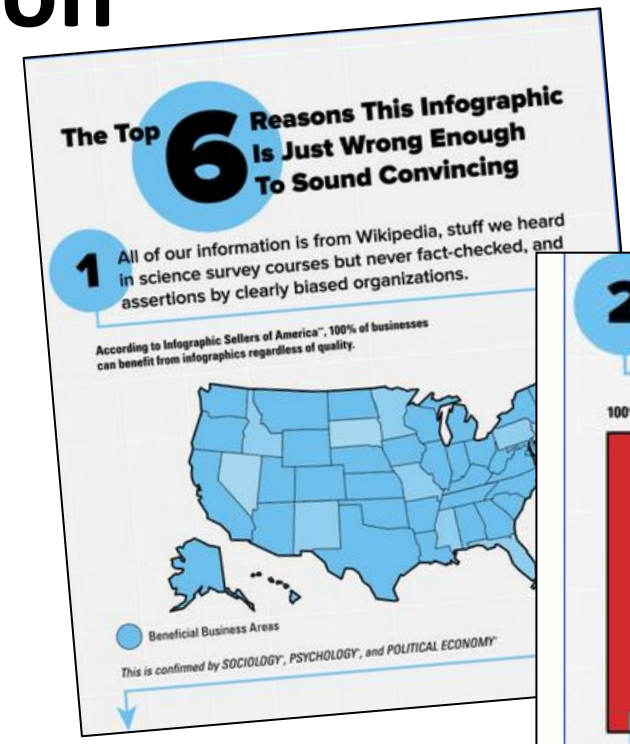
Hoffman, Judy, John Carlo Bertot, and Denise M. Davis. "Executive Summary: Public Library Funding & Technology Access Study 2011–2012." *Digital Supplement: Libraries Connect Communities*. Spec. issue of *American Libraries* Summer 2012: 7-9. *ALA*. Web. 31 Oct. 2013. <http://ebooknoir.files.wordpress.com/2012/06/digitalsupplement.pdf>.

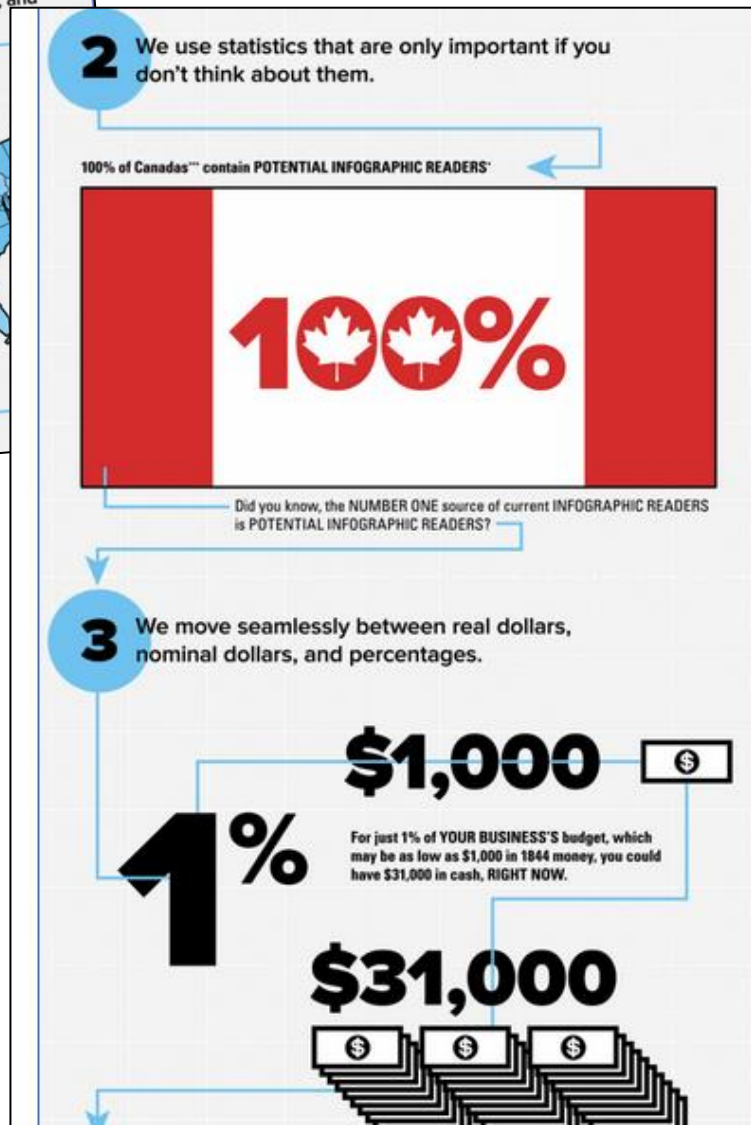
Teach students which chart to use

- **Line charts** analyze trends, patterns over time
- **Bar charts** compare categories of data within a time period
- **Scatter plots** show how two attributes vary together
- **Histograms** view and compare distributions

The visualization has to be:

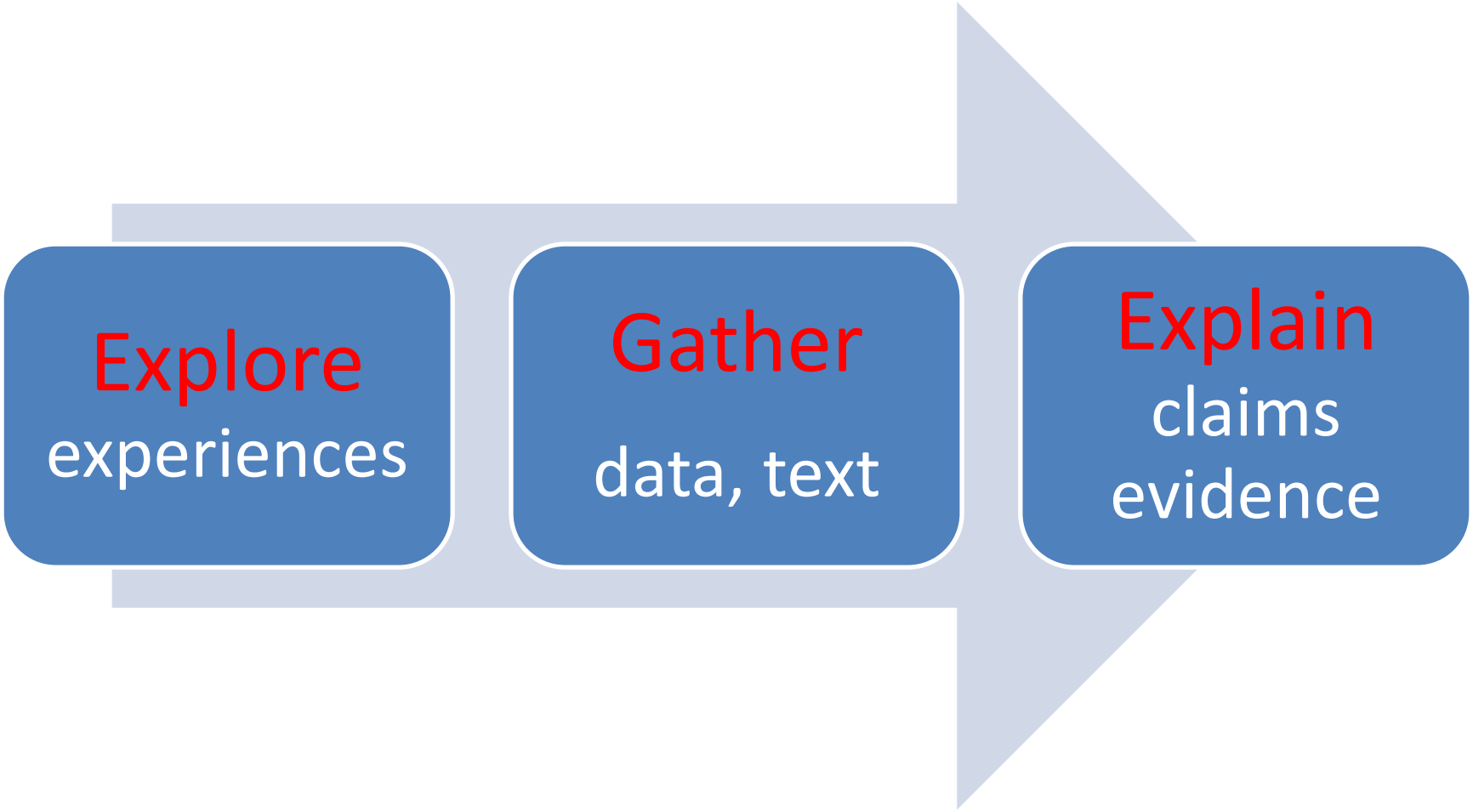
- Accurate
- Thorough
- Simple
- Truthful
- Clear
  ("Chart junk" – Tufte)





McNulty, Johnny. "This Is the Most Perfect Infographic Ever Created. It's about Why Infographics Suck." *Pics & Posts*. Someecards, 12 Nov. 2013. Web. 18 Nov. 2013. <http://happyplace.someecards.com/27460/this-is-the-most-perfect-infographic-ever-created>.

Data can be used as evidence (Common Core, Next Gen Science)



Argument in writing (CER+A)

- Claim or thesis
- Backed up with evidence
- Reasoning that supports idea / claim

+ Alternatives (Rebuttal)

Teaching with data

- Using real-time or archived published data
- Collect their own data
- Contribute data to larger sets
- Evaluate data
- Create products and papers using data
- Others…?

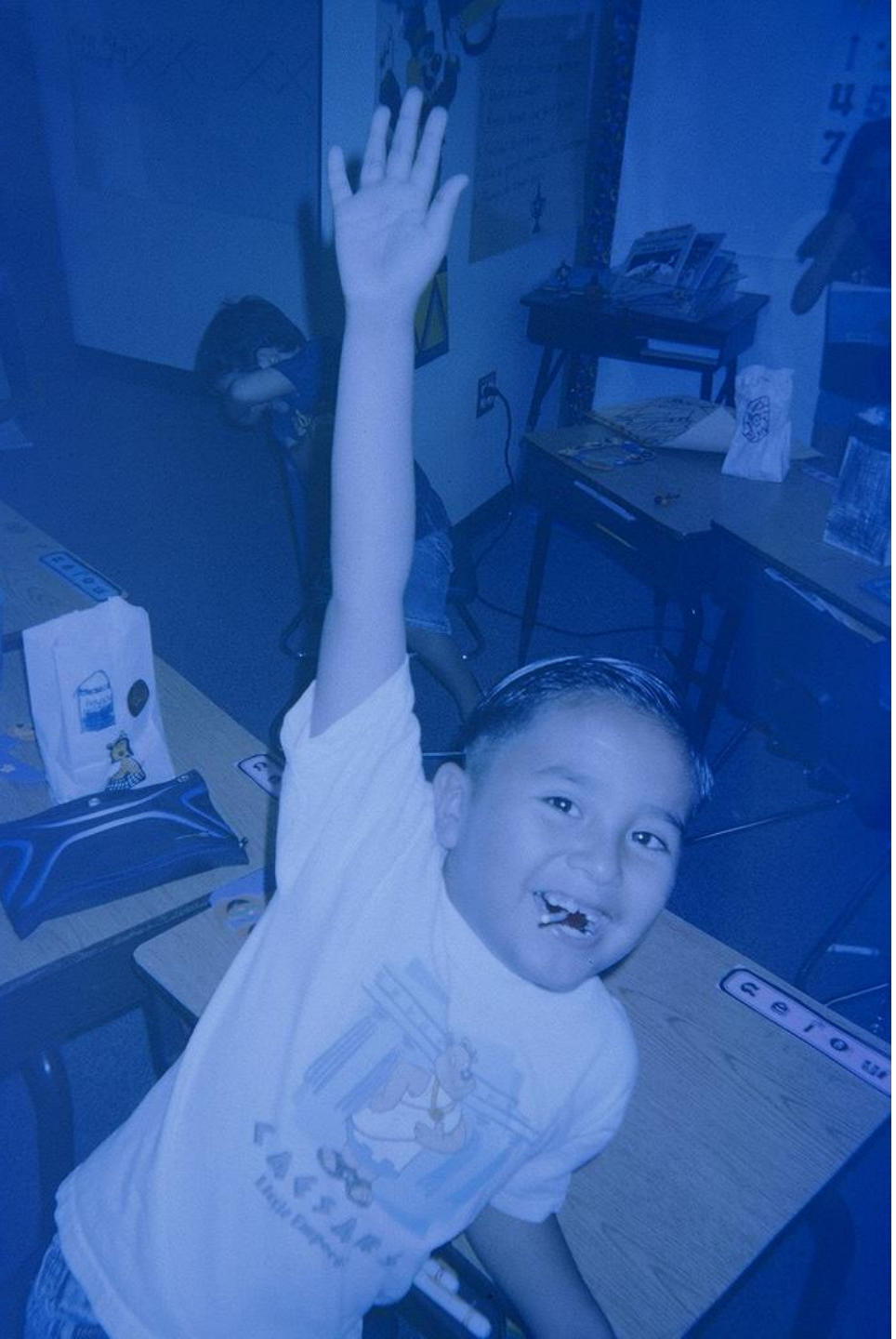
"I want your content – Uncle Sam" by thomas lapperre on Flickr. CC-BY.
www.bloei.se. http://flickr.com/photos/107790207@N08/10688617585/

Exploration Time!

- Opportunityindex.org
- Socialexplorer.com
- Galaxyzoo.org
- eBird.org
- Books.google.com/ngrams
- google.org/flutrends/us/
- google.com/trends/correlate

Debrief and Discussion

"'So I says to him, 'THIS is how you do a karate chop'" byAislinn Ritchie on Flickr. CC-BY-SA. http://flickr.com/photos/richteabiscuit/600780410/



Questions?

Debbie Abilock
debbie@noodletools.com

Kristin Fontichiaro
font@umich.edu
@activelearning

Slides
http://bit.ly/fontblog
http://www.noodletools.com/debbie/literacies

"Hey, teacher! Inow the answer!" by Kevin Dooley on Flickr. CC-BY.
http://flickr.com/photos/pagedooley/3364591795/