From “Skip The Numbers” To “Great Stuff”: A Data Education Project

BEYOND THE NUMBERS, FEDERAL RESERVE BANK OF ST. LOUIS, 11/8/2018, 4:15 - 5:15pm

Kristin Fontichiaro
U. of Michigan School of Information

Wendy Stephens
Jacksonville St. University

Slides: http://bit.ly/btn-data-lit

Today we'll talk about …

- The impetus, structure, and deliverables of our project
- The learnings reported by our "experts"
- Current findings and implications for your practice

slides: bit.ly/btn-data-lit

Hello!

Who are you?

Survey: http://bit.ly/infosurvey18

Hello!

We're Kristin Fontichiaro and Wendy Stephens.

Supporting Librarians in Adding Data Literacy Skills to Information Literacy Instruction

Made possible in part by the Institute of Museum and Library Services RE-00-15-0113-15

Core Personnel

Kristin Fontichiaro, PI, UMich School of Information (UMSI)
Jo Angela Oehrli, Co-PI, UMich Library
Amy Lennex, Project Manager
Tyler Hoff, Project Assistant, UMIS
Kelly Hovinga, Project Assistant, UMSI
Martha Stuit Project Assistant, UMSI

Curriculum Experts

Debbie Abilock, NoodleTools
Susan D. Ballard, National Center for Digital Equity
Tasha Bergson-Michelson, Castilleja School
Jennifer Colby, Huron High School
Jole Seroff, Castilleja School
Susan Smith, Harker School
Wendy Steadman Stephens, Jacksonville State University
Connie Williams, ret. from Petaluma High School

Data Experts

Jacob Carlson, UMich Library
Lynette Hoelter, ICPSR
Justin Joque, UMich Library

Special Contributors/Guest Presenters

Tuvya Bergson-Michelson, Lick-Wilmerding School
Catherine d'Ignazio, Emerson College
Charissa Jefferson, Cal. State Northridge
Diego Mendez-Carbajo, Illinois Wesleyan
Katrina Stierholz, FRED/Federal Reserve Bank of St. Louis
Justin Schell, UMich Library
Tierney Steelberg, Guilford College
Samantha Viotty, formerly Emerson College, currently Obama Foundation
Andrew Whitehead, Assn. of Religion Data Archives / Clemson University

INSTITUTE of Museum and Library SERVICES

DATA LITERACY

The ability to "read" and "write with" data

slides: bit.ly/btn-data-lit

Project Overview

Turn and Talk:

What do you tell your students/learners about how to read a scholarly article?

Turn and Talk:

What do you tell your students/learners about how to read a scholarly article?

Share out …

1. “I just tell them to read the text and skip the numbers.”

2. Bad Infographics.

3. Belief that 2016 would mirror 2012 election, with campaigns rich with data and stats and with microtargeting of voters. Were high schoolers ready to be voters?

4. Emergence of Big Data and automated, non-human, algorithmic decision-making

5. Growing focus on research data management / data repositories / data information literacy at U-M Library

With that in mind, let's look back at our survey results

Data/stats comprehension

Big Data / Citizen Science

Data in arguments

Ethical data use

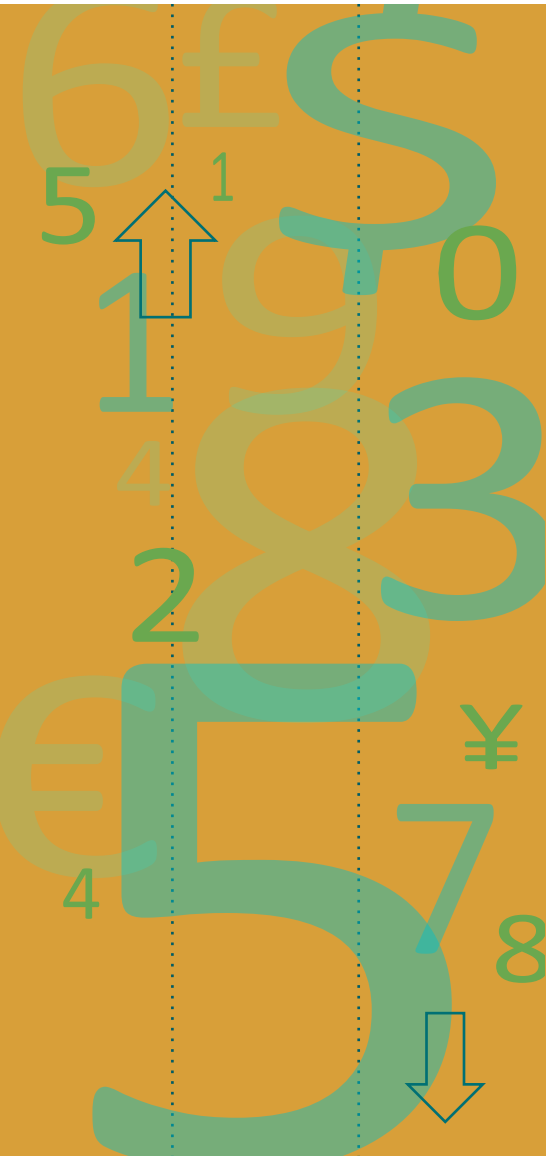
Data visualization

Personal data management

We planned	We ended up with
2 virtual conferences	3 virtual conferences
One short “rules of thumb” book	2 books, totalling nearly 700 pages
Project evaluation	Project evaluation (in process)
	(8-book series for middle-grade readers)

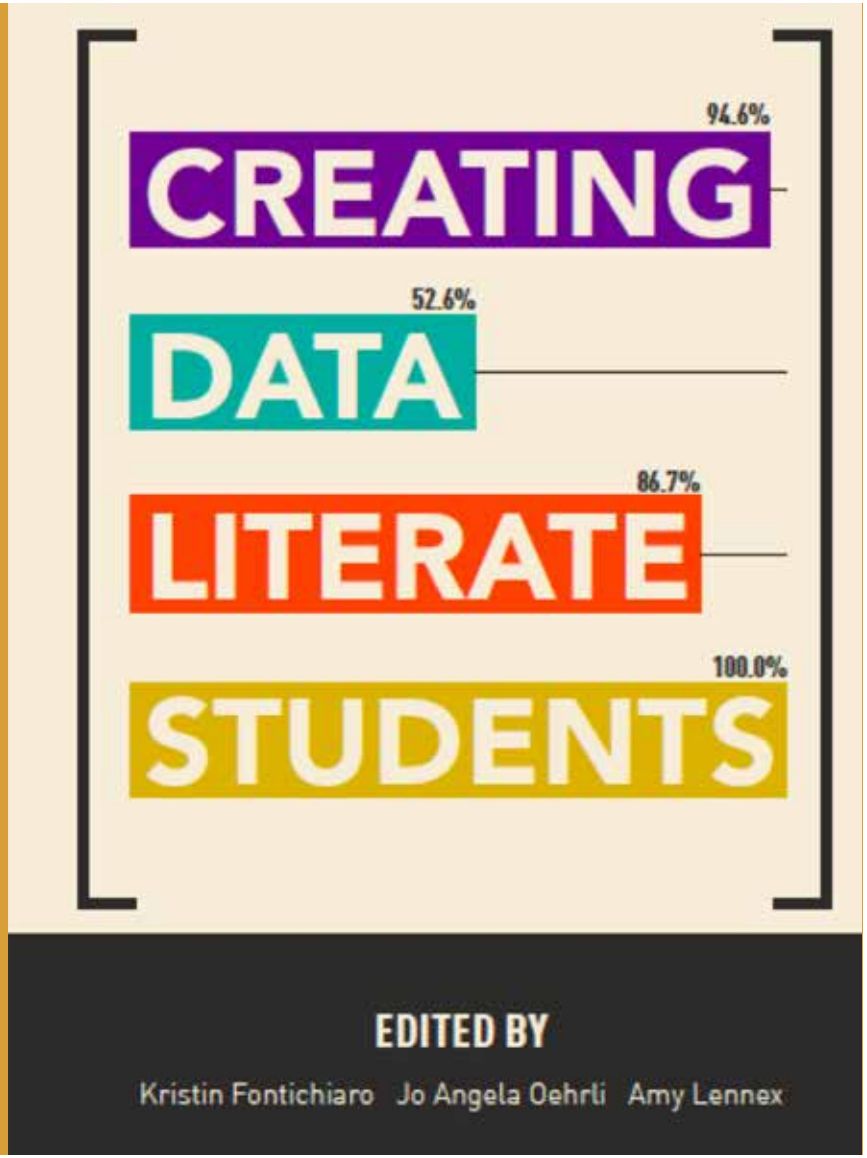
4T Virtual Conference on Data Literacy

- Two-day event for each of 3 years
- Free & online
- 2016 & 2017 focused on one of 3 corresponding annual themes
- 2018 open topics (co-sponsor ICPSR)
- Technically focused on high school librarians and educators, but ~ 2/3 of population over 3 years were *not* in this group



Introduction to Statistical Literacy / Lynette Hoelter
Statistical Storytelling: The Language of Data / Tasha Bergson-Michelson
Using Data in the Research Process / Jole Seroff
Real world data fluency: How to use raw data / Wendy Steadman Stephens
Manipulating data in spreadsheets / Martha Stuit
Making Sense of Data Visualization / Justin Joque
Data presentation: Showcasing your data with charts and graphs / Tierney Steelberg
Deconstructing data visualizations: What every teen should know / Susan Smith
Designing your infographic: Getting to design / Connie Williams
Using data visualizations in the content area / Jennifer Colby
Teaching Data Contexts: An Instructional Lens / Debbie Abilock
Diving Lessons: Taking the Data Literacy Plunge Through Action Research / Susan D. Ballard

dataliteracy.si.umich.edu/books/



Part I:

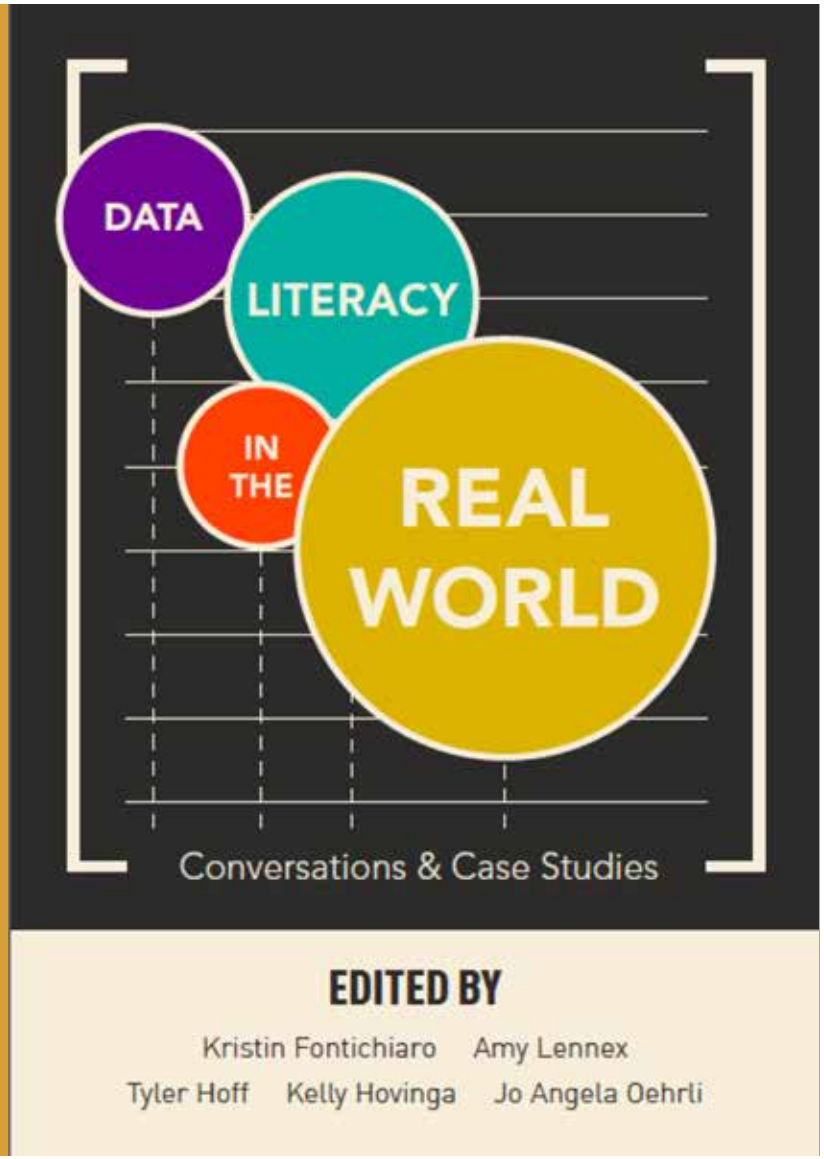
"PD in a box." Discussion questions and activities based on archived sessions from 2016 & 2017 4T Virtual Conference on Data Literacy

Part II:

45+ Case Studies drawn from current events:

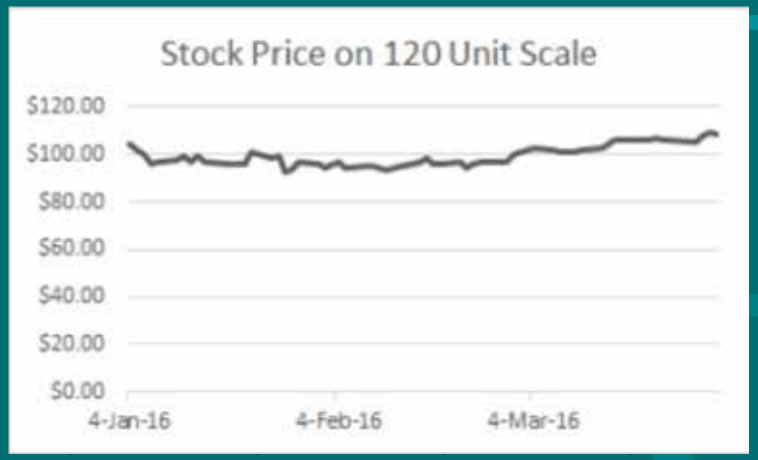
- Cambridge Analytica, FitBit, predictive policing, racist policies and data, citizen science projects, ethical data use, use of security cameras in special ed. Classrooms, K-12 student data privacy, Amazon Echo Look, etc.

dataliteracy.si.umich.edu/books/

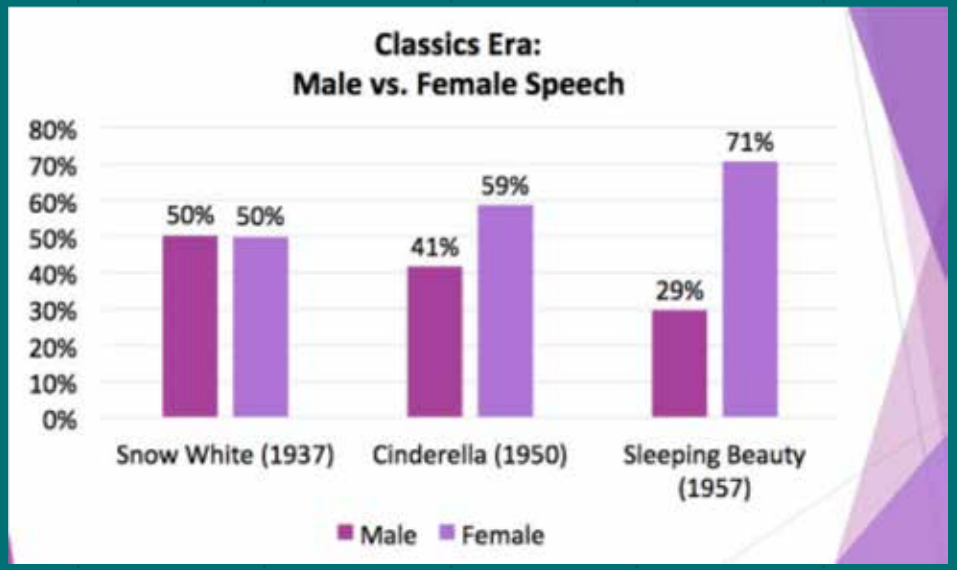


2.
What the curriculum team learned

Thinking about numeracy



Variables and ambiguities



Interrogating the data

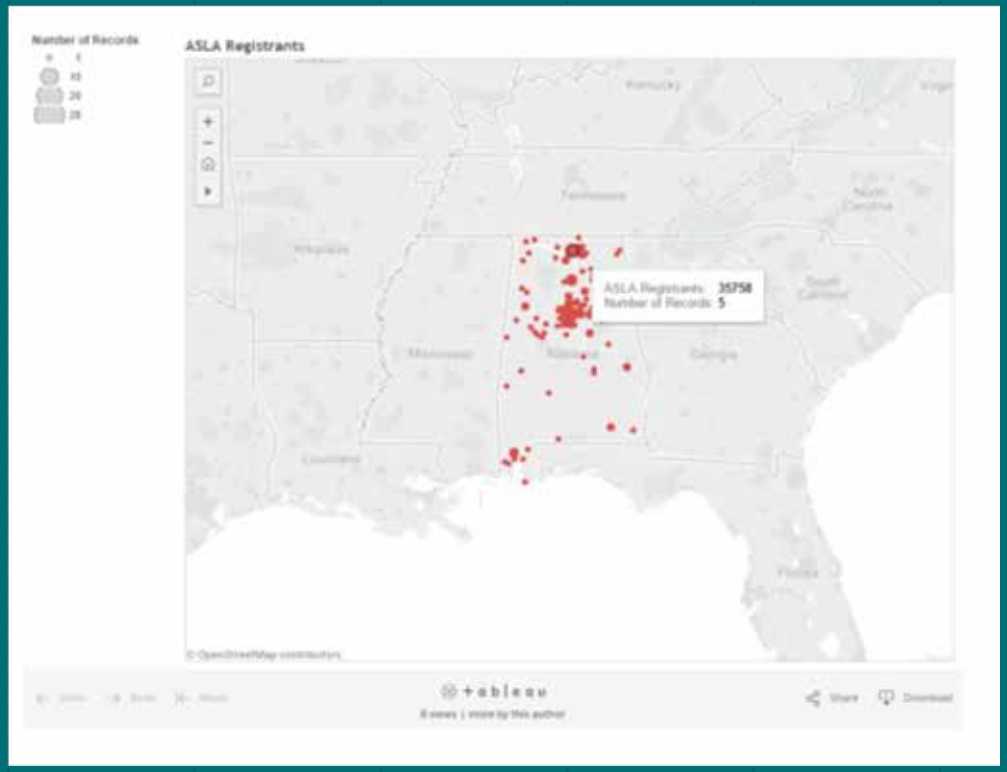
“47 percent pay no taxes”

“Majority think nuclear power safest”

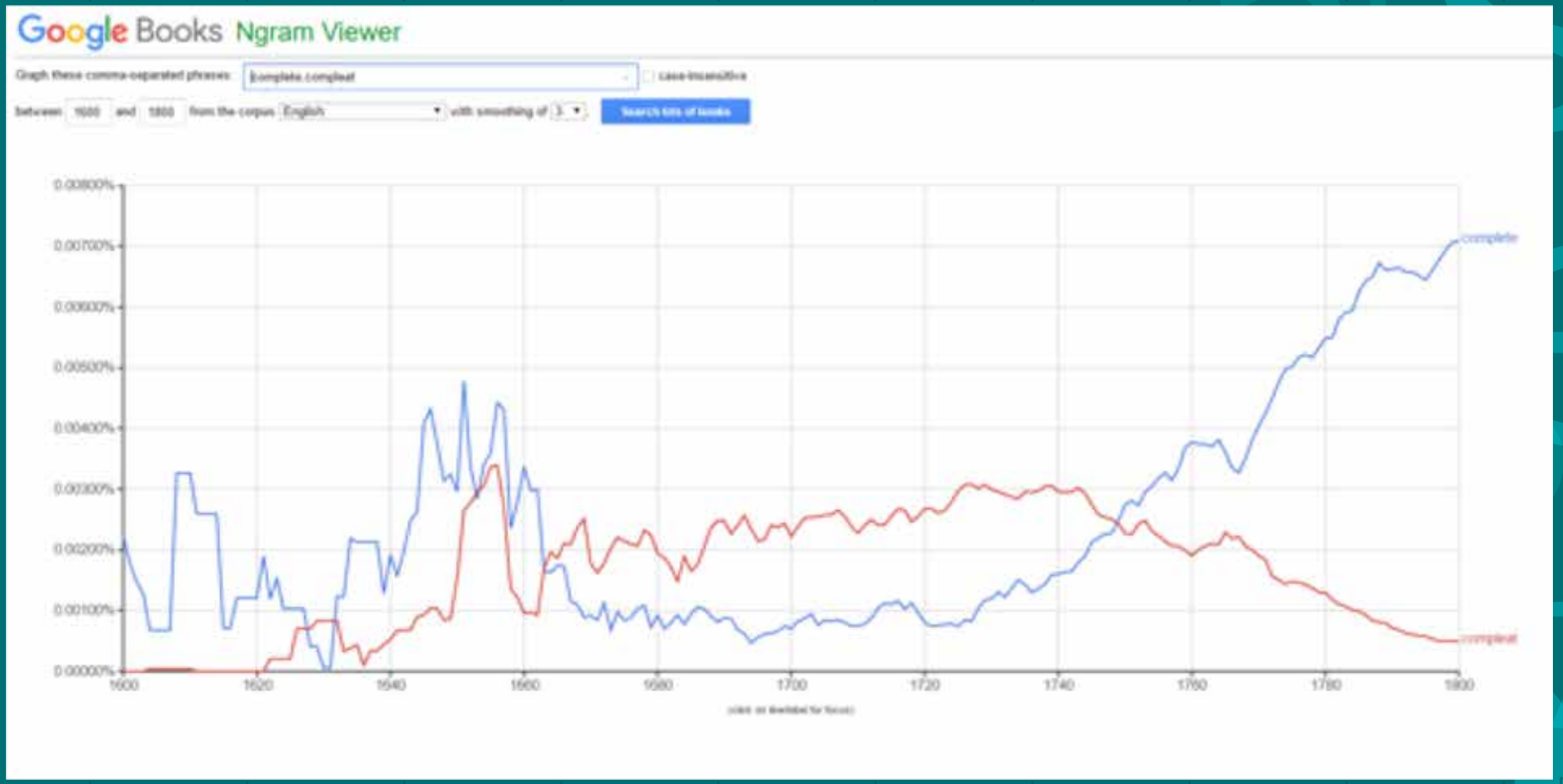
“Twenty percent support it”

Cut-and-paste without context

+ableau+public

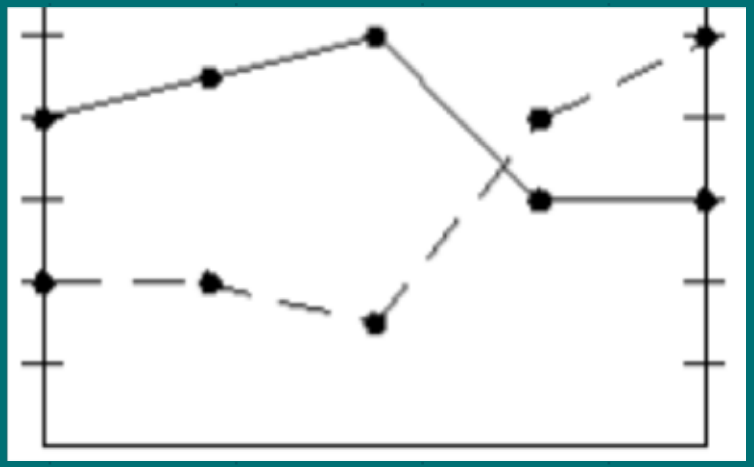


Making responsible use of data

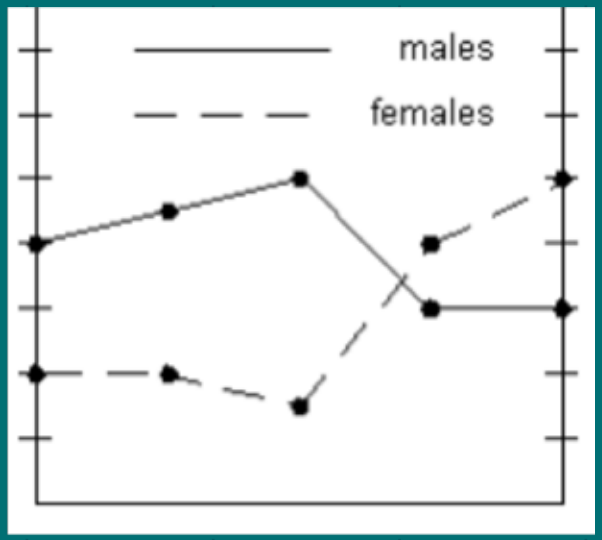

Google Books Ngram Viewer
Graph these comma-separated phrases: |complete,compleat
case-insensitive
between 1600 and 1800 from the corpus English with smoothing of 3
Search lots of books
0.00800%
0.00700%
0.00600%
0.00500%
0.00400%
0.00300%
0.00200%
0.00100%
0.00000%
1600
1620
1640
1660
1680
1700
1720
1740
1760
1780
1800
complete
compleat

Data fluency

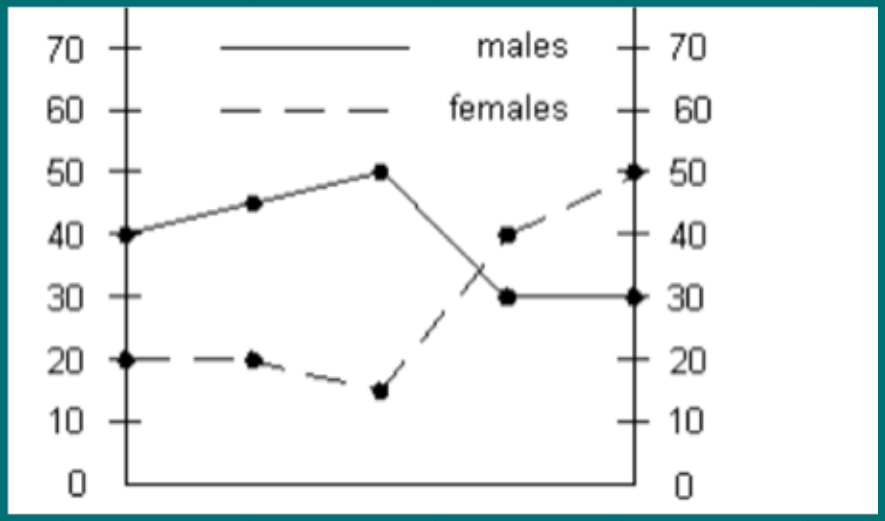
- Thinking computationally
- Finding existing data sets
- Traveling backward from news' accounts and soundbites
- Are the parameters explicit?
- Making responsible use of data



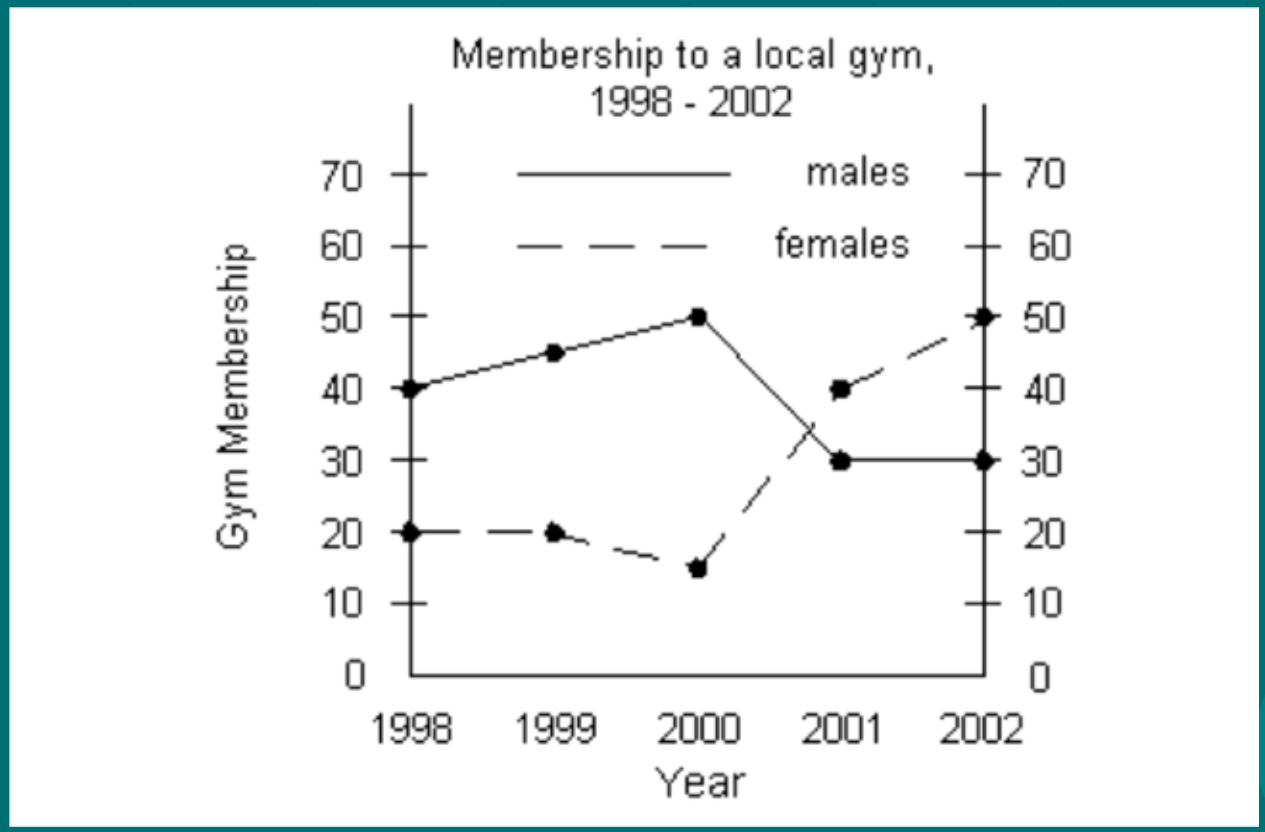
Trends



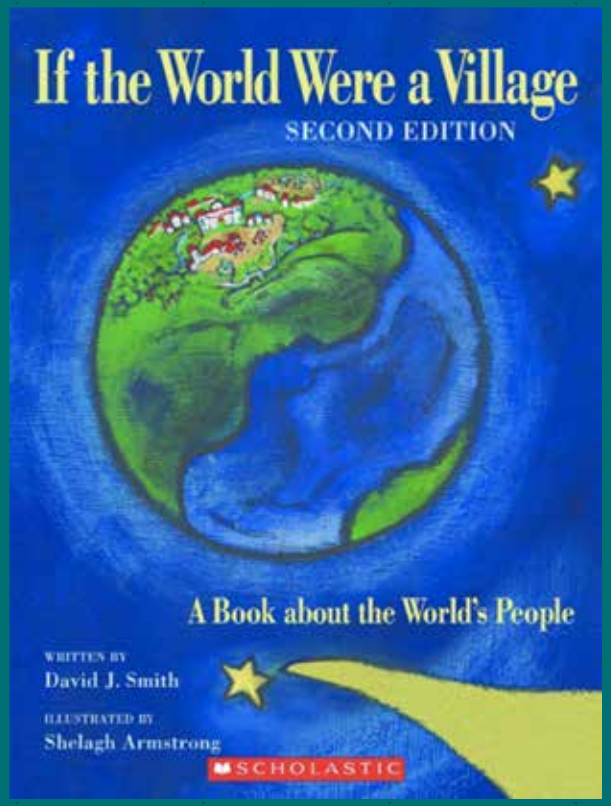
Labels

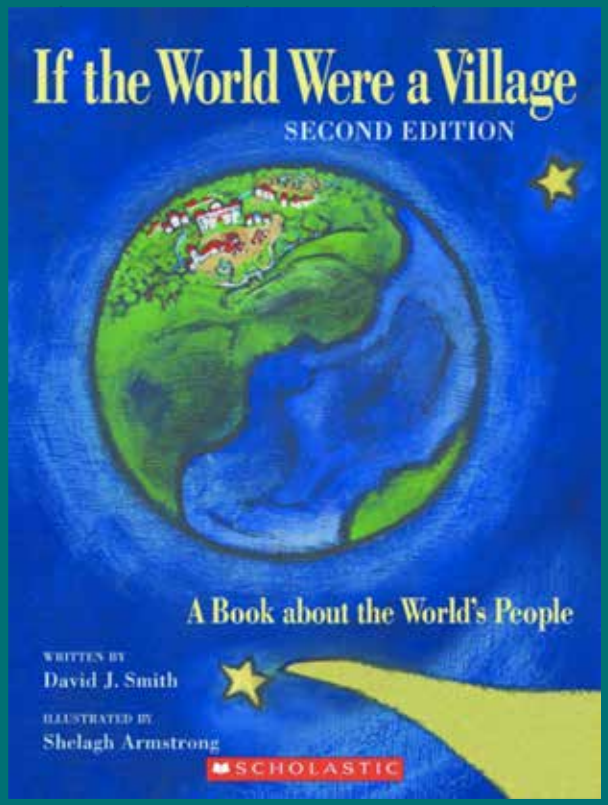


Values


Membership to a local gym,
1998 - 2002
Gym Membership
70
60
50
40
30
20
10
0
males
females
1998
1999
2000
2001
2002
Year

Key


If the World Were a Village
SECOND EDITION
A Book about the World's People
WRITTEN BY
David J. Smith
ILLUSTRATED BY
Shelagh Armstrong
SCHOLASTIC



Of the 100 people in the global village

61 are from Asia

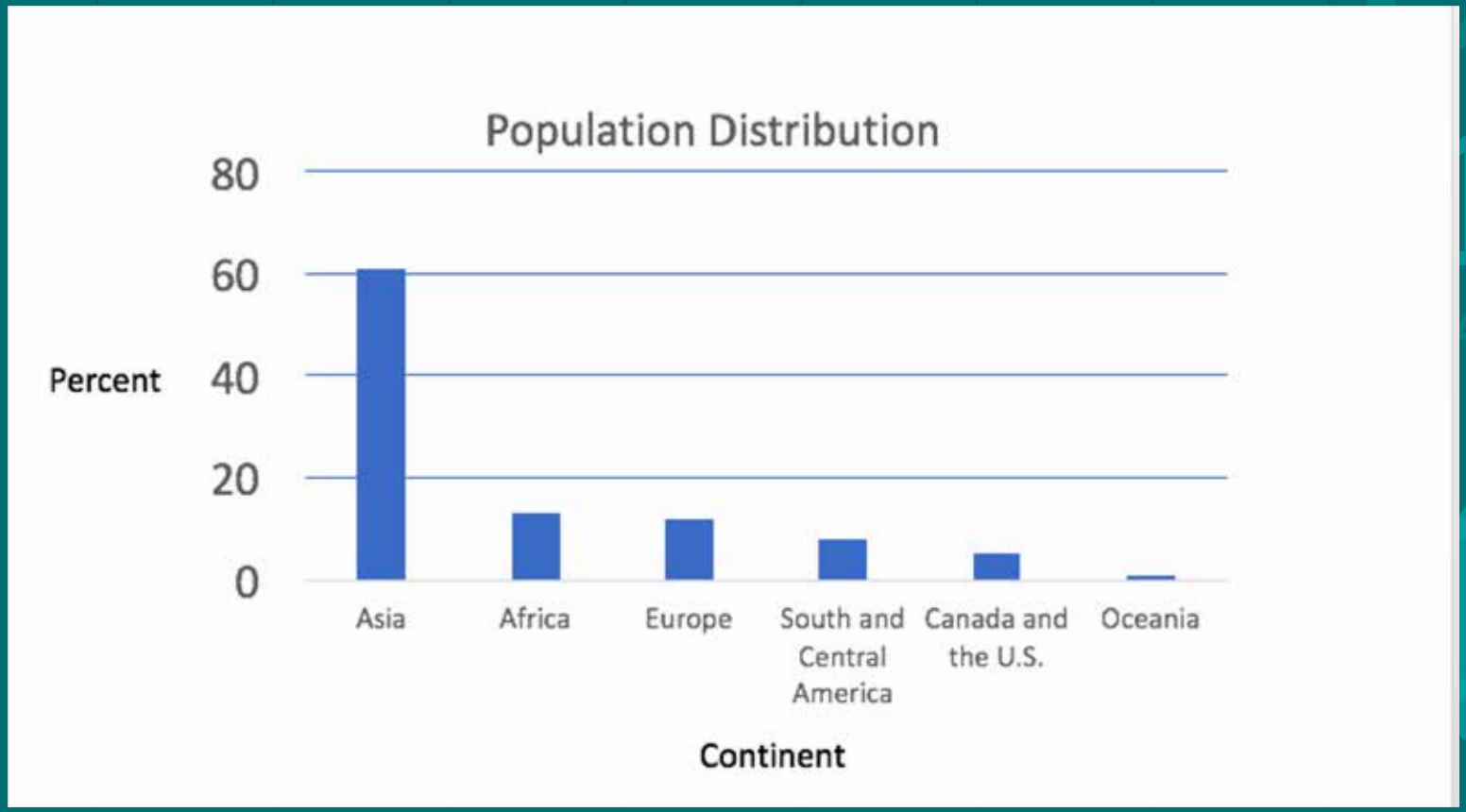
13 are from Africa

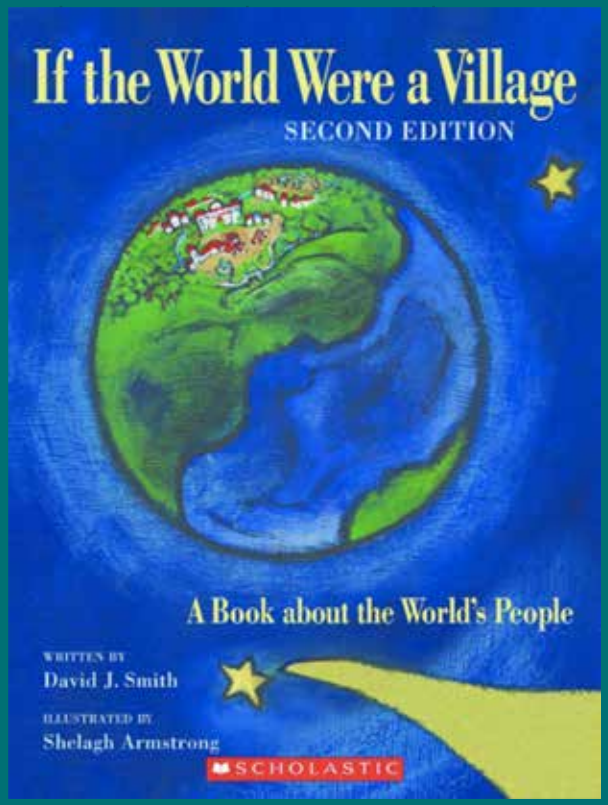
12 are from Europe

8 are from South and Central America

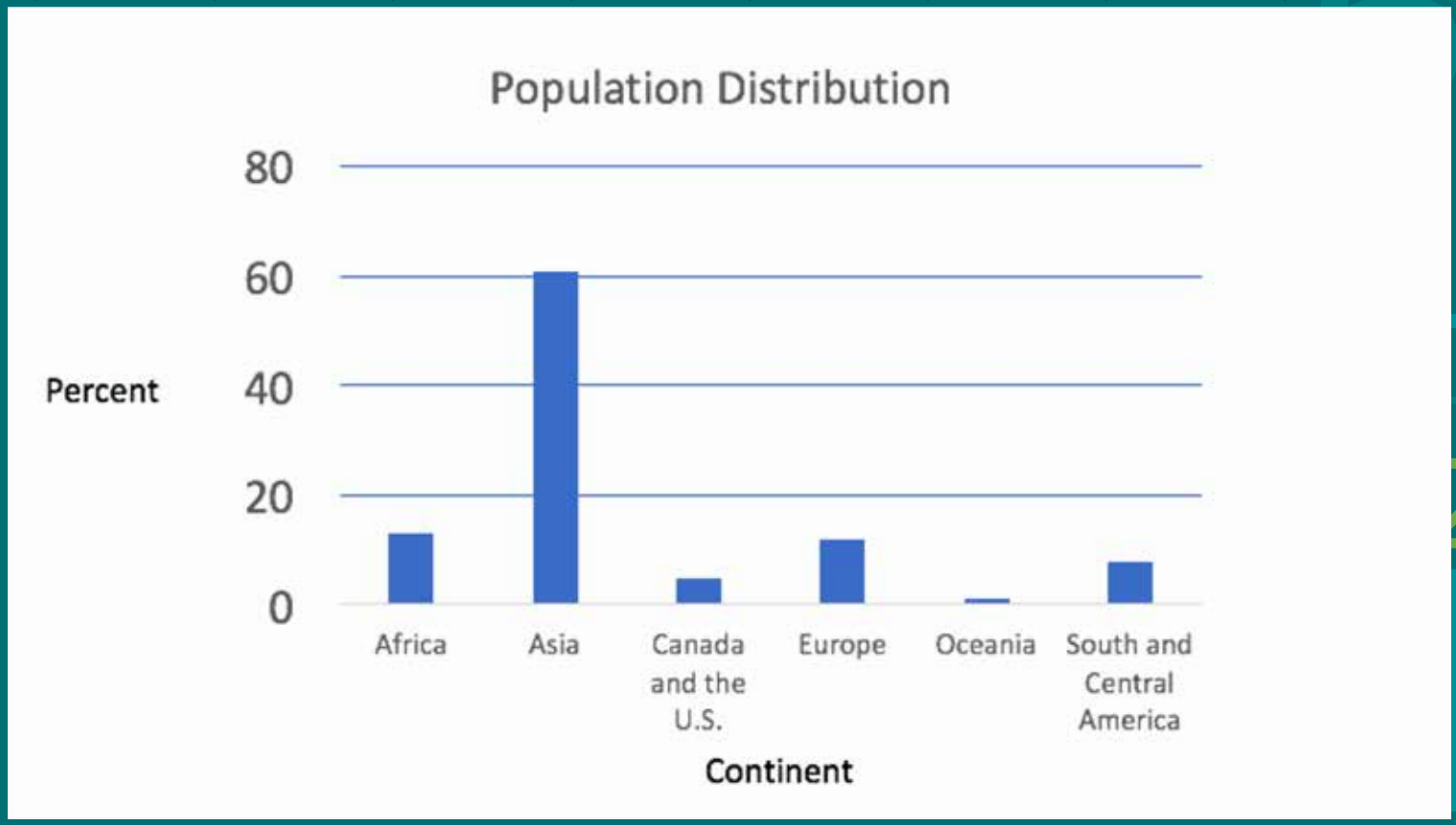
5 are from Canada and the U.S.

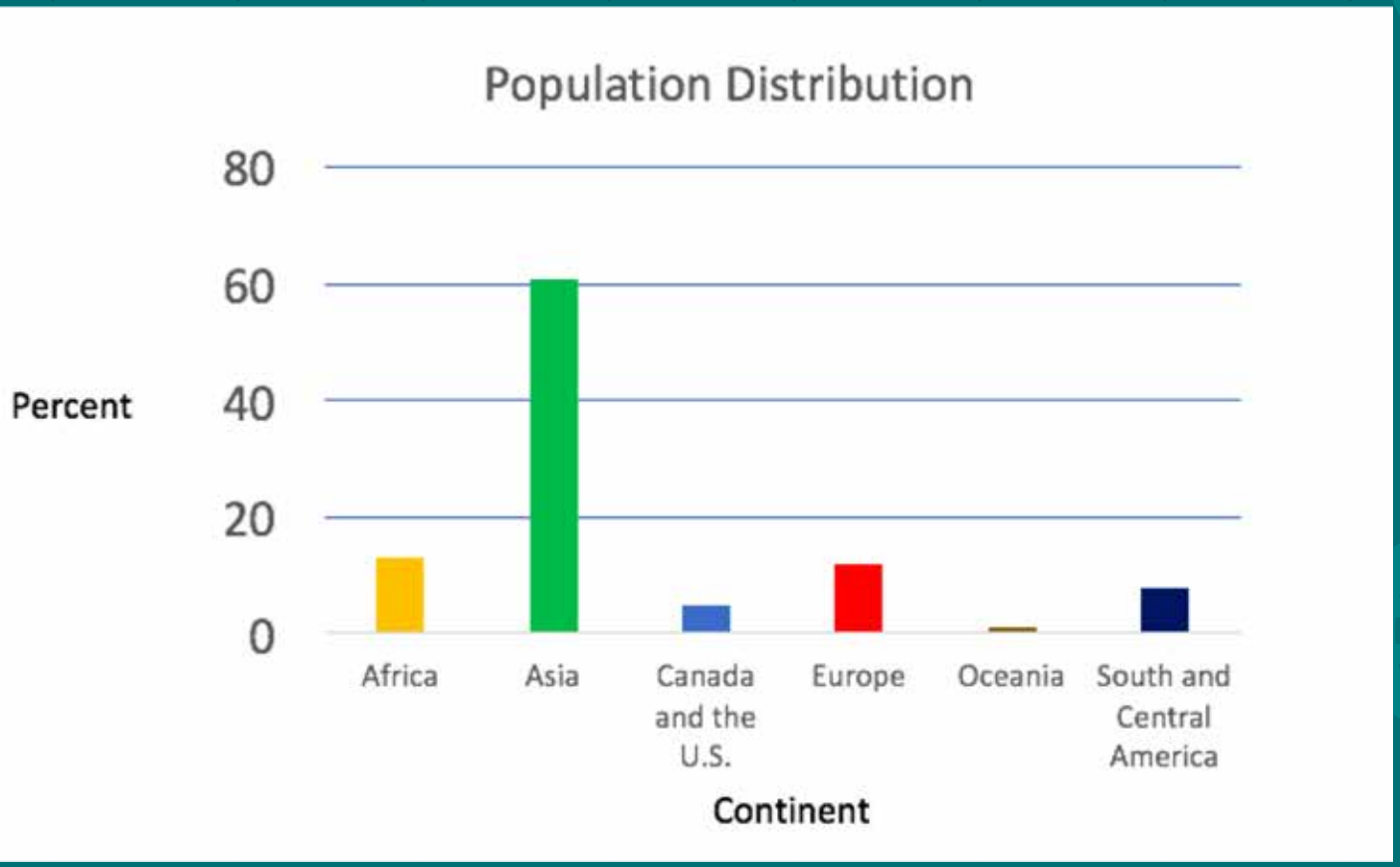
1 is from Oceania

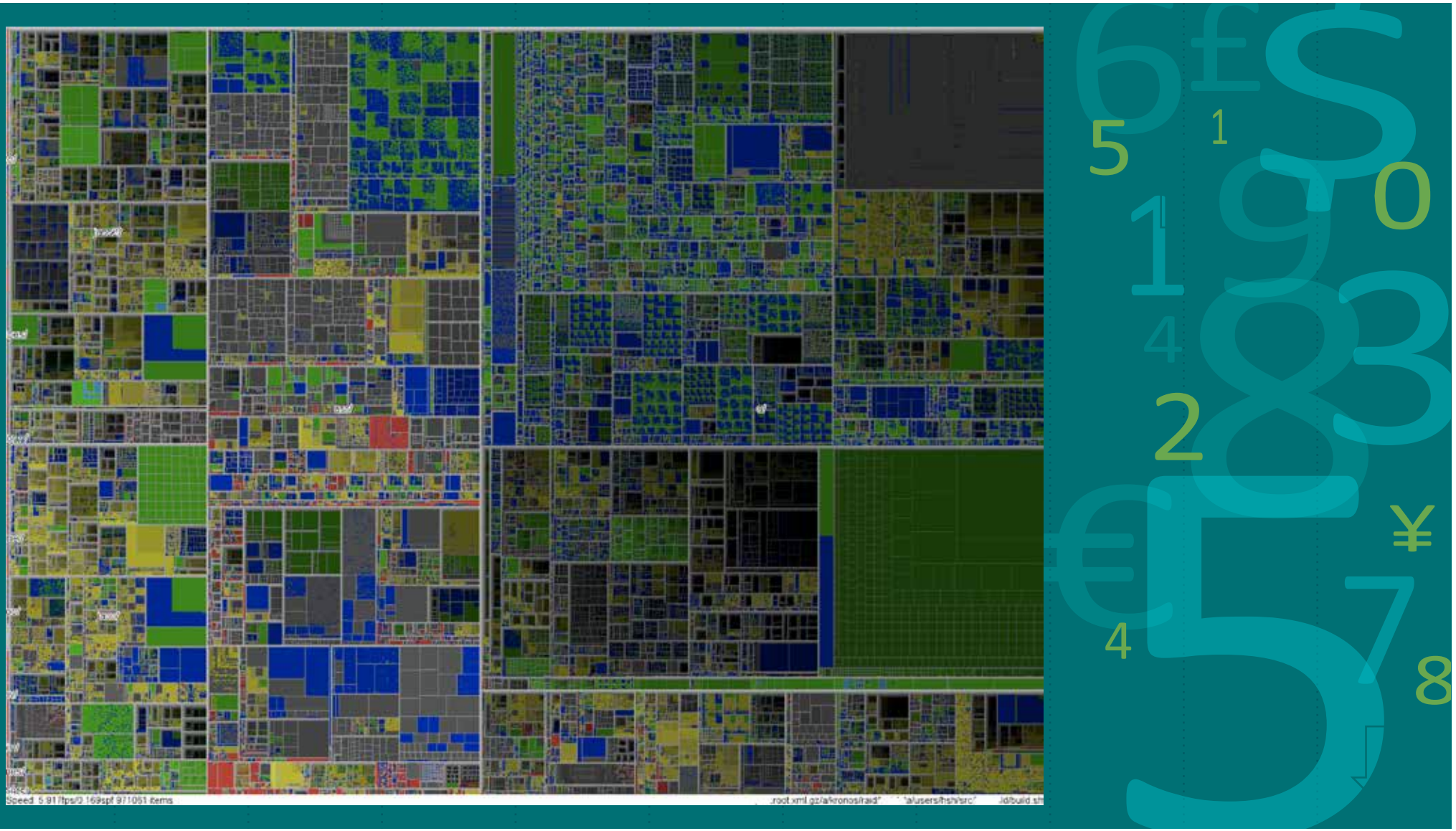

Population Distribution
80
60
40
20
0
Percent
Asia
Africa
Europe
South and Central America
Canada and the U.S.
Oceania
Continent



Of the 100 people in the global village
13 are from Africa
61 are from Asia
5 are from Canada and the U.S.
12 are from Europe
1 is from Oceania
8 are from South and Central America


Population Distribution
80
60
40
20
0
Percent
Africa
Asia
Canada and the U.S.
Europe
Oceania
South and Central America
Continent


Population Distribution
Percent
80
60
40
20
0
Africa
Asia
Canada and the U.S.
Europe
Oceania
South and Central America
Continent


Speed: 5.91 fps/0.169spf 971051 items
/root.xml.gz/a/kronos/raid"
"a/users/hsh/src/"
Jd/build.sh

3.
What we've learned so far from a project perspective

A little data lit goes a long way.

Fontichiaro, Oehrli, & Hoff. ALA Annual 2017. Design by Jo Angela Oehrli.

REAL STRATEGIES TO ADDRESS FAKE NEWS

LIBRARIANS, DATA LITERACY AND THE POST-TRUTH WORLD

IN THE NEWS

What is Fake News?

The term "fake news" has been used in a lot of situations. For the purpose of our work, fake news can be any of these 3 things.

* Fake News: Manipulative/made up content. Many times this material is created online by content farms and shared via social media
* Fake News: Authentic news stories labeled as fake by some politicians who think it is unbearable to themselves
* Bad information: Sloppy reporting, clickbaity headlines, and/or news created by well-intentioned folks who don't know any better

About the Project.

In what some are calling a "post-truth world," students seem to focus on numbers a lot. Students believe that if a number is connected to information, it has to be a fact. But numbers are manipulated all of the time. We want students to have a tool kit of questions that they can use to question the data that is put there.

Supporting Librarians in Adding Data Literacy Skills to Information Literacy Instruction is a two-year project running from October 2015 through September 2017. It is designed to develop the data and statistical literacy skills in high school librarians so they can better support critical comprehension skills in their students. You can find more information at http://datalit.eracy.umich.edu.

This project was made possible in part by the Institute of Museum and Library Services RE-00-15-0013-15.

What is Being Counted?

When you look at research being presented as news, make sure you know who was studied, how they were selected, and what data points were being counted.

For example, when unemployment numbers are reported in the news, consider who was counted as unemployed. Does that include retired people? Children under 16? People who aren't looking for jobs any more? To see how the U.S. calculates unemployment, see the Bureau of Labor Statistics. For international guidelines, consult the International Labour Organization's standards.

SPECIAL COVERAGE

MEAN | MEDIAN | MODE

Don't Settle for Average!!

"Average" is sometimes used as a synonym for three distinct mathematical calculations: mean, median, and mode. Knowing the difference can help you ask, "How was this average calculated?"

Mean = Adding up all numbers and dividing by the quantity of numbers
Median = Middle Number
Mode = Most Common Number

When considering statistics with major outliers, the mean may skew way outside the majority of numbers. Think about calculating the mean of Bill Gates' net worth with that of every U.S. librarian. His wealth is so massive it would reveal the "average" way too high! Calculating the median would give a much more realistic number. Mode could help us know where there are clusters of identical net worth, which might help us identify trends for further inquiry.

Causation ≠ Correlation.

Beware of headlines saying one thing causes another. That degree of certainty is very difficult to achieve. It's more likely correlation (two variables moving in sync with one another, but without knowing for certain that there isn't coincidence or other factors involved).

For example, just because people in a certain geographic area searched online for information about flu symptoms, doesn't mean that Google can predict flu trends. There could be other factors involved.

RECOMMENDED

Are You Comparing Apples to Apples?

Is the comparison using the same units of measurement? Time periods? Is it fair to compare these two (or many) things?

For example, is it fair to compare the 2009 video streaming viewership of the inauguration ceremony to the 2017 viewership? Video streaming is more prevalent and easier in 2017 than it was in 2009. When comparing 2.3 million streaming CNN viewers in 2017 to 1.3 million in 2009, you are not comparing apples to apples.

Where is this Data Coming From?

Is the data coming from a reliable source? Dig deeply to find out.

Consider the chocolate milk research that came out in 2016. Chocolate milk was supposed to cure concussions! There was a study conducted at the University of Maryland, a respected university. It turns out that the study was paid for by the dairy industry, the study wasn't peer reviewed, and there are some serious questions about the design of the study. Yet the sponsors of the study used it to advertise their product and many news releases were issued! Researchers are supposed to declare any potential conflicts of interest. And this researcher didn't.

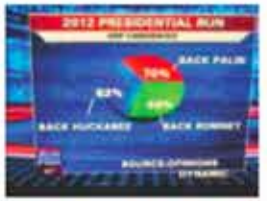
When was the data collected?

Is the data you're looking at still relevant? Is it fair to use the data you're reading in another situation? If someone is discussing the cost of college today, make sure they have current data!

IN THE NEWS

Pie Charts Should Add Up To 100%.

Sometimes they don't...



POPULAR

If It's Too Good to Be True, It Probably Is!

If what you're reading radically validates what you are feeling, you should question it!

So... if you really want to believe that Coca-Cola is good for you so that you can keep drinking it by the gallon, and then you read that it cures cancer, take a moment to read closer!

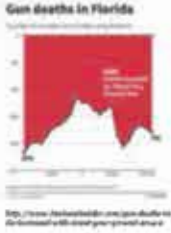
MOST RECENT

Be Aware of How Color, Font, and Design Elements May Be Swaying Your Reaction to Graphical Content.

Is there a lot of red in a visualization? How does that make you feel? Is the font in the image the same as the font on warning signs? Graphic design elements can cause a knee-jerk reaction that the data itself may not support. Explore emotional elements and check yourself to see if they are supported by the data.

Keep an Eye on the Y-Axis.

The y-axis is the vertical axis, usually labeled on the left side of a graph. Look at the intervals on the y-axis. Are they at the same interval throughout the graph? Does it start with zero? Are the numbers increasing or decreasing? How does that impact your reading of the graph?



Is That a Big Number? Compared to What?

Keep some ready reference numbers in your toolkit and be ready to compare any number you encounter in the news to what you already know. Those numbers are known as statistical benchmarks. For example,

* U.S. Population: 325 million (U.S. Census 2017)
* World population: Just over 7.3 billion (U.S. Census 2017)
* Number of U.S. births each year: Just under 4 million (NCHS 2014)

Know the statistical benchmarks that impact your world. Consider that if jobs in the energy industry are an important aspect of the news, you should know that the coal industry employs approximately 10,000 people and that the number is declining. Alternatively, the solar industry employs 260,000 people ... and it's rising. (https://www.thesolarfoundation.org/national/).

Is the Evidence Biased?

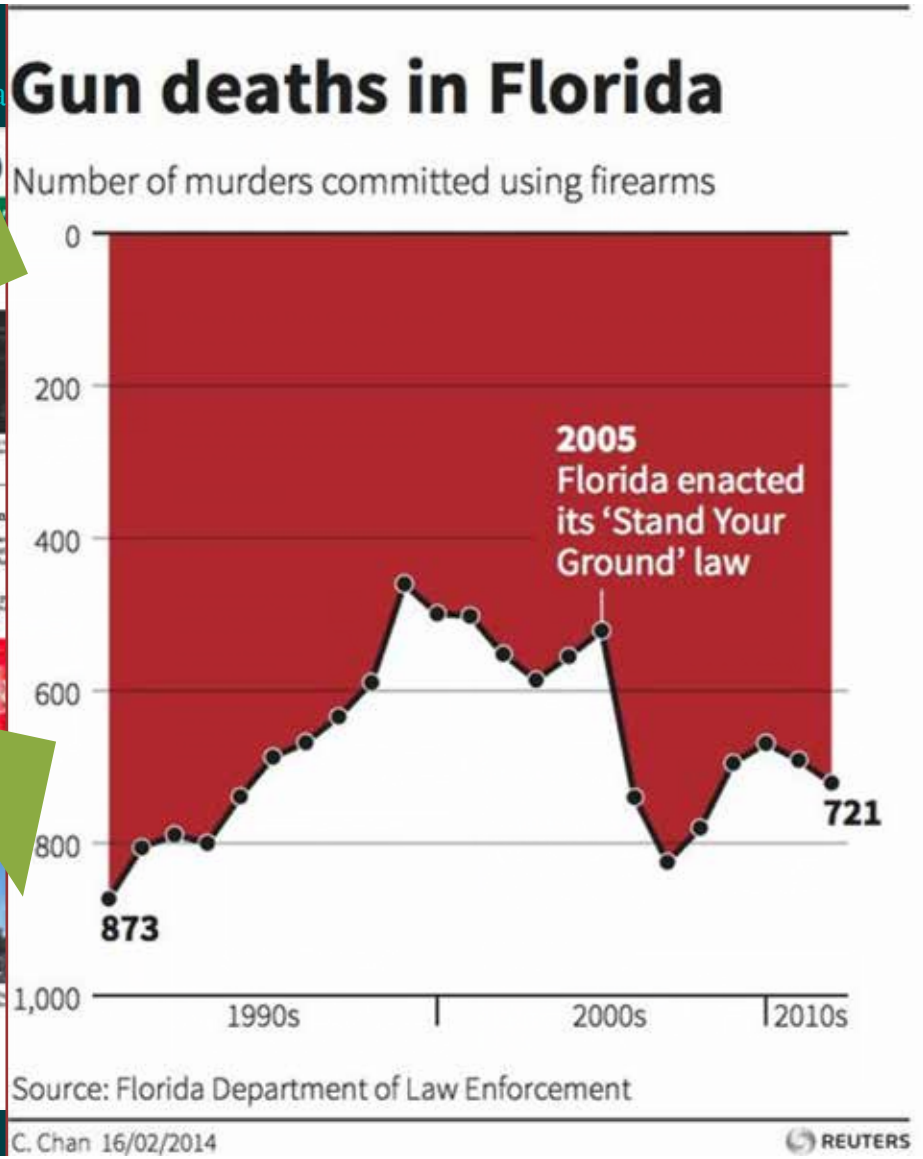
Big Data is used to make a lot of claims. But Big Data algorithms can be based on biased data. For example, software used to predict criminality has been proven to be biased against African Americans. Don't assume that Big Data is infallible – there could be implicit bias hiding beneath its algorithms.

SPONSORED BY

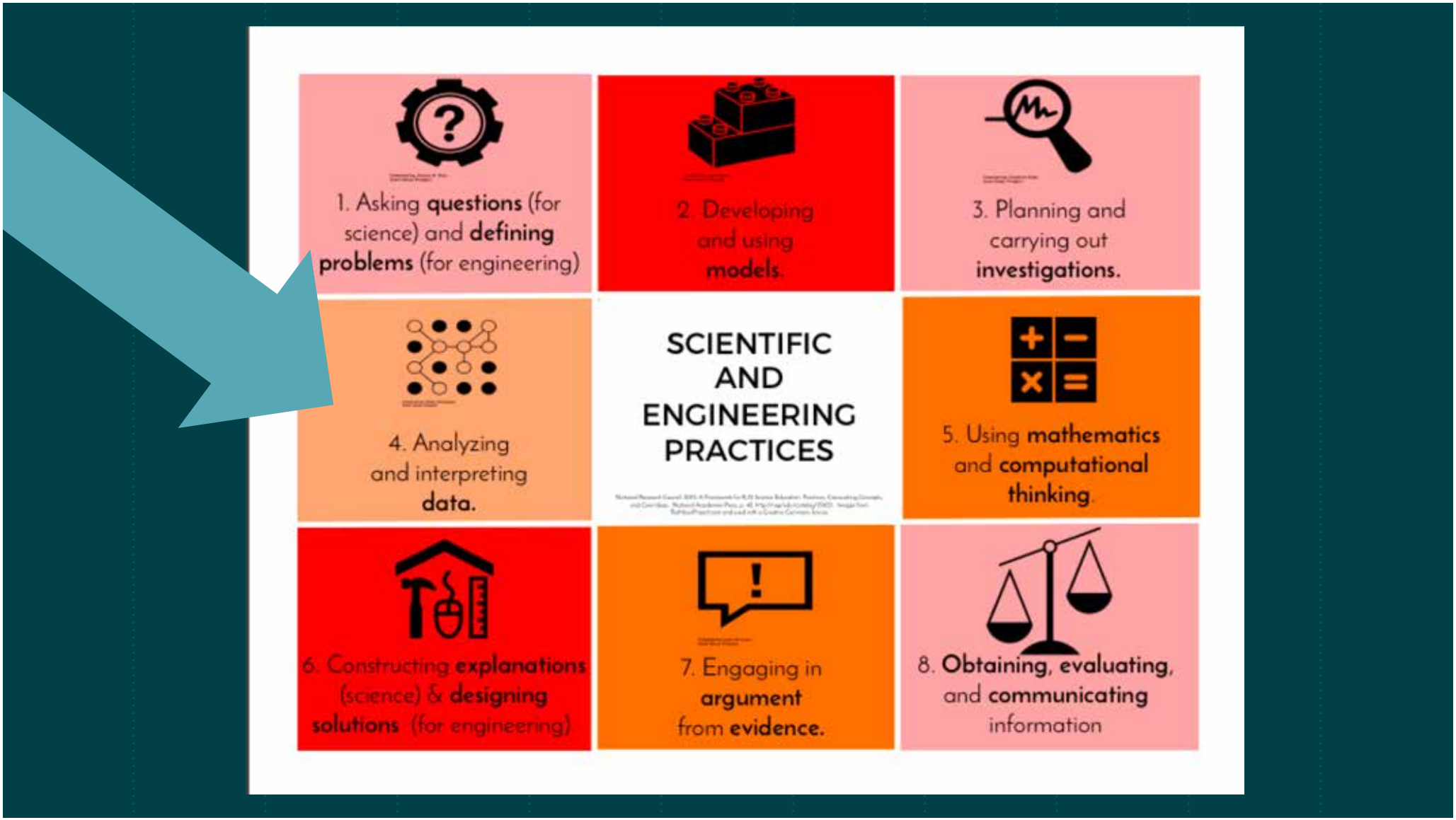
Kristin Fontichiaro, Tyler Hoff, and Jo Angela Oehrli

Fontichiaro, Oehrli, & Hoff. ALA Annual 2017. Design by Jo Angela

https://www.wonkette.com/reuters-cool-upside-down-chart-makes-stand-your-ground-look-awesome



Almost everybody is concerned about data literacy.


1. Asking **questions** (for science) and **defining problems** (for engineering)
2. Developing and using **models**.
3. Planning and carrying out **investigations.**
4. Analyzing and interpreting **data.**
SCIENTIFIC AND ENGINEERING PRACTICES
5. Using **mathematics** and **computational thinking**.
6. Constructing **explanations** (science) & **designing solutions** (for engineering)
7. Engaging in **argument** from **evidence.**
8. **Obtaining, evaluating,** and **communicating** information

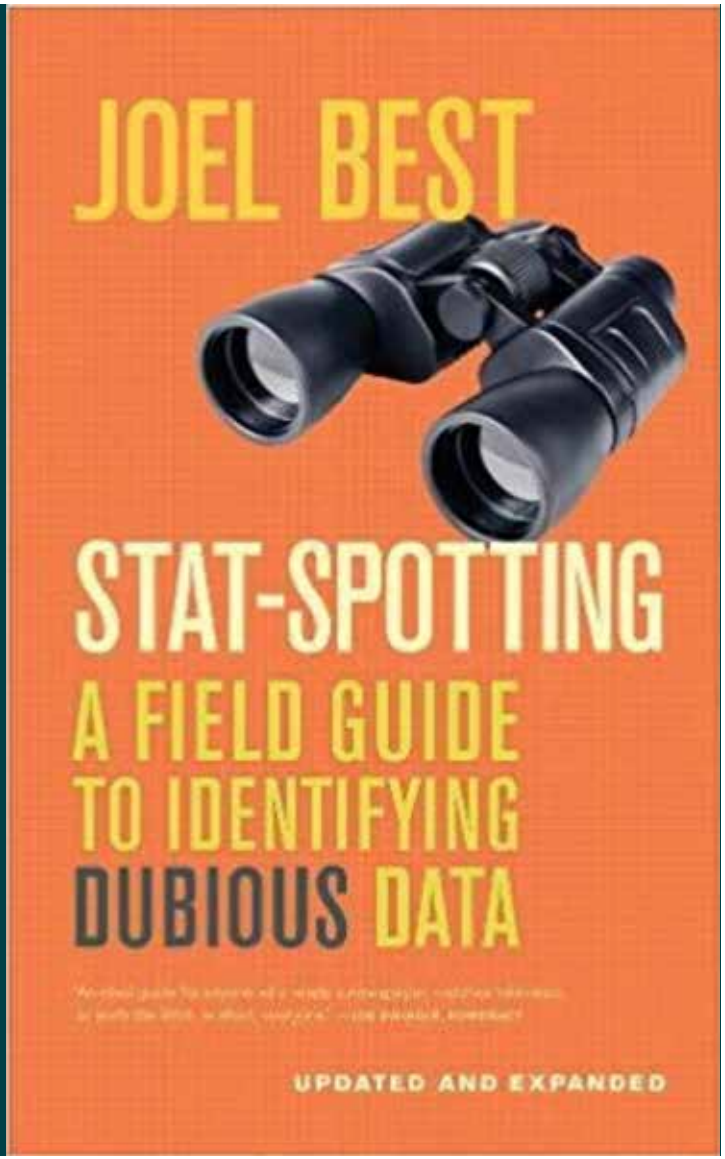
Almost anyone can benefit from data literacy.

"The Data Geek series supports the new curriculum standards that focus on understanding, interpreting, and gathering data. Information in each book is designed to help readers explore all kinds of data and data sources in order to objectively understand data in the 21st century. Readers are encouraged to think critically about the ways data is used in their lives and in the media … Grades 4-7."

https://cherrylakepublishing.com/shop/show/50829

Statistical benchmarks offer a foundation for meaningful comparison.

See also: “compared to what?” and, “Is that big number?”



Data literacy is a prerequisite to larger areas of study like data science, data crunching, or lab-based research.

Entering a data literacy conversation via faculty/student perceptions of pain points is effective.

Variation: framing data lit around existing curriculum.

Addie Matteson
@queenaddie

All of today's session were very interesting (and maybe a little scary). My school already puts a lot of time and resources into teaching our kiddos about being safe on the internet, but I don't think we specifically address how data about them is being accumulated. After today, I feel like that's an inexcusable oversight.
Great stuff. Can't WAIT to hear from Tuvya tomorrow!

Jul 12

Used with permission

Master of Applied Data Science

Coming in Fall 2019

M | UMSI

http://bit.ly/umsi-mads

Takeaways:

1. The need for quality data literacy education is everywhere (e.g., K-12, university departments, not-for-profits), and a little goes a long way.

2. Look for "pain points" with faculty that you can solve with data literacy education (e.g., data viz).

3. Recognize that before a student or librarian can tackle datasets or scholarly articles, they may need guidance in data lit principles (can our deliverables help?).

Kristin Fontichiaro
University of Michigan
School of Information
font@umich.edu
@activelearning
http://fontichiaro.com/blog

Wendy Steadman Stephens
Jacksonville State University
wstephens@jsu.edu
@wsstephens

Project: dataliteracy.si.umich.edu

Today's slides: bit.ly/btn-data-lit