Data Literacy

Ryan Womack

Data Librarian, Rutgers University, https://ryanwomack.com

National Statistics Office of Mongolia, Ulaanbaatar, Mongolia, May 9, 2017

About Me

- Data Librarian
- Masters in Economics, Library Science, and Statistics
    - working in the intersection of these fields
    - help people find and use data
    - manage data for the Libraries

- interaction with advanced scholarship and research

About Rutgers

- Rutgers, The State University of New Jersey
- New Jersey, 8.8 million people, between New York and Philadelphia
- Rutgers founded in 1766, over 250 years old
- Carnegie Classification: Research - Very High Research Activity, the highest classification (R1)
- more than 100 major programs of study
- over 68,000 students (7,500 international from over 120 countries)
- Top 100 university in the world (Times Higher Education, Shanghai Ranking, CWUR, and others)

About Rutgers

NEW YORK

MASSACHUSETTS

Boston

CONNECTICUT

RHODE ISLAND

PENNSYLVANIA

Atlantic Ocean

Rutgers University–Newark

New York

Rutgers University–New Brunswick

Philadelphia

Trenton

Rutgers University–Camden

MARYLAND

NEW JERSEY

Baltimore

Washington, D.C.

DELAWARE

- New Brunswick to Newark: 28 mi (45 km)
- New Brunswick to New York City: 41 mi (70 km)
- New Brunswick to Camden: 62 mi (100 km)
- New Brunswick to Philadelphia: 66 mi (106 km)
- Newark to Camden: 83 mi (134 km)
- Newark to New York City: 13 mi (21 km)
- Camden to Philadelphia: 4 mi (6.4 km)

About IASSIST

- IASSIST is the International Association for Social Science Information Services and Technology
- A community of over 300 data professionals working in information technology, libraries, data services, research & higher education, government, non-profit and private research sector
- IASSIST seeks to
    - Foster and promote a network of excellence for data service delivery
    - Advance infrastructure in the social sciences and beyond
    - Provide opportunities for collegial exchange of sound professional practices

- Annual conference is a forum for presentation and networking (US, Canada, Europe, next Asia?)
- IASSIST is seeking increased membership from Asia

Outline

Data Literacy

Ryan Womack

Introduction

Statistical
Literacy

Data Literacy

Competencies

Role of NSOs

In Practice

- Introduction
- Statistical Literacy
- Data Literacy
- Competencies
- Role of NSOs and Academia
- In Practice

Literacy: What do we mean?

- **Literate**: "able to read and write"
- Literate: "having knowledge or competence"
- An old concept that has been extended into many new areas

Multiple Literacies

- "Literacy" borrows the force of our original reading experience
    - the critical skills to understand, interpret, and create in that domain

- *Media Literacy*, *Digital Literacy*, *Social/Emotional Literacy* (Digital Age Literacies)
- *Scientific Literacy*
- *Numeracy*, *Quantitative Literacy* - in the domain of mathematics

Which Literacy?

We will explore:

- Statistical Literacy
- Information Literacy
- Data Information Literacy
- Data Literacy –> our final goal

Statistics

Two meanings:

- *numerical* - a collection of quantitative data
- *conceptual* - a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data

Statistical Literacy

Data Literacy

Ryan Womack

Introduction

Statistical
Literacy

Data Literacy

Competencies

Role of NSOs

In Practice

Statistical literacy deals with both numerical and conceptual meanings.

- An old concept: see Walker, "Statistical Literacy in the Social Sciences", *The American Statistician*, 1951
- The original concept focused on limited examples, basic math
- Numeracy and Quantitative Literacy are similar - focus on basic math skills
- Statistical Literacy requires Quantitative Literacy
- But emphasis is on using the statistical concepts in practice (with statistical sources)

Statistical Literacy, example

Question from ARTIST:

Statisticians often use the term "robust" when referring to a statistical measure or test. Which of the following is an example of a robust measure of center?

1. Mean, because it is the most commonly used measure of center.

2. Median, because it is resistant to outliers.

3. Mode, because it can be used with observational or quantitative data.

Statistical Literacy, example

The prior example is addressed to a statistical audience, but we can illustrate with examples that would work for any level of audience.

- Dataset: 1,2,2,2,3
- What is Mean, Median, Mode?
- Replace 3 with 18
- What is "robustness"? Develop intuitive understanding.

Statistical Literacy, defined

- Gal, "Adults' Statistical Literacy: Meanings, Components, Responsibilities." *International Statistical Review,* 2002

"statistically literate behavior depends on the availability and joint activation of multiple knowledge bases"

- Literacy (reading, but also including handling graphs, charts, and tables, and other forms of textual evidence)

- Statistical

- Mathematical

- Context

- Critical Skills

these are combined with *Beliefs and Attitudes* and *Critical Stance*

=> this adds up to *Statistical Literacy*

Statistical Literacy, in practice

Everyday statistical literacy

- Knowing things like mean, median, mode
- Understanding that data points can be distributed in different ways
- Being able to interpret a graph or chart
- Knowing to look for bias, critically analyze

Statistical Knowledge Base

Data Literacy

Ryan Womack

Introduction
Statistical
Literacy
Data Literacy
Competencies
Role of NSOs
In Practice

Statistical Knowledge Base

1 Knowing why data are needed and how data can be produced

2 Familiarity with basic terms and ideas related to descriptive statistics

3 Familiarity with basic terms and ideas related to graphical and tabular displays

4 Understanding basic notions of probability

5 Knowing how statistical conclusions or inferences are reached

quoted from Gal, 2002

more on Statistical Knowledge Base

Data Literacy

Ryan Womack

Introduction

Statistical
Literacy

Data Literacy

Competencies

Role of NSOs

In Practice

1. When it can be concluded that a relationship is one of cause and effect, and when it cannot, including the difference between randomized experiments and observational studies.

2. The difference between statistical significance and practi- cal importance, especially when using large sample sizes.

3. The difference between finding "no effect" or "no difference" and finding no statistically significant effect or difference, especially when using small sample sizes.

4. Common sources of bias in surveys and experiments, such as poor wording of questions, volunteer response, and socially desirable answers.

5. The idea that coincidences and seemingly very improbable events are not uncommon because there are so many possibilities.

6. "Confusion of the inverse" in which a conditional probability in one direction is confused with the conditional probability in the other direction.

7. Understanding that variability is natural, and that "normal" is not the same as "average."

quoted from Utts, "What Educated Citizens Should Know about Statistics and Probability", *The American Statistician*, 2003

more on Statistical Knowledge Base

Data Literacy

Ryan Womack

Introduction
Statistical
Literacy
Data Literacy
Competencies
Role of NSOs
In Practice

"Watson, in turn, identified three abilities that determine statistical competence. In increasing order of complexity, they are: a) a basic understanding of probabilistic and statistical terminology and the ability to perform analytical and statistical calculations; b) the ability to interpret probabilistic and statistical language and concepts when they are embedded in social media contexts; c) the ability to critically evaluate statistical claims related to sampling, the distribution of raw data, appropriate use of statistics, graphs, causal claims made, and probabilistic statements (Watson 1997)"
quoted from Prado and Marzal, "Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents", *Libri*, 2013

Statistical Thinking

Data Literacy

Ryan Womack

Introduction

Statistical
Literacy

Data Literacy

Competencies

Role of NSOs

In Practice

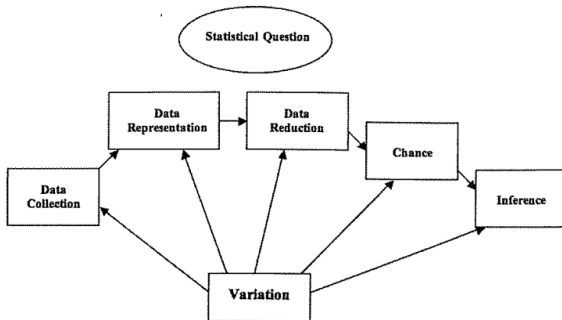J.M. Watson / Foundations for improving statistical literacy

Fig. 1. Relationships and components in statistical thinking.

Watson, "Foundations for improving statistical literacy",
Statistical Journal of the IAOS, 2011

How to measure statistical literacy

Data Literacy

Ryan Womack

Introduction

Statistical Literacy

Data Literacy

Competencies

Role of NSOs

In Practice

- *Consistent non-critical* - Appropriate but non-critical engagement with context, multiple aspects of terminology usage
- *Critical* - Critical, questioning engagement in contexts that do not involve proportional reasoning, but which do involve appropriate use of terminology
- *Critical mathematical* - Critical, questioning engagement with context, using proportional reasoning particularly in chance contexts, showing appreciation of the need for uncertainty in making predictions, and interpreting subtle aspects of language
- see Tarran, "How to measure statistical literacy", *Significance*, 2017

Information Literacy

- **Information Literacy** is a major concept in libraries and elsewhere in education
- Interpreting textual material
- Understanding the information environment
- Many library educational standards have been defined around this concept
- For librarians, data literacy can be viewed as an offshoot or subset of information literacy
- see Scheld, "Information Literacy, Statistical Literacy and Data Literacy", *IASSIST Quarterly*, Summer/Fall 2004

Research Data Management

Libraries and others have become increasingly involved in
Research Data Management

- supporting the data lifecycle for researchers
- creating, analyzing, sharing, and reusing data
- less emphasis on statistical skills and understanding
- MacMillan, "Data Sharing and Discovery: What Librarians need to know" *The Journal of Academic Librarianship*, 2014

Data *Information* Literacy

A hybrid of data management and information literacy, this terminology is new and not as widely used.

- How to Inventory, Store, and Backup Your Data
- How to Create Data That You (and Others) Can Understand
- How to Navigate Rights and Ownership of Your Research Data
- How to Share Your Data and Ethically Reuse Data Created by Others
- How to Digitally Preserve Your Data for the Future
    - see Johnston and Jeffries, "Steal This Idea: A library instructors' guide to educating students in data management skills", *College and Research Libraries News*, September 2014

Data Explosion

Larger worldwide trend of interest in data

- Big Data
- Data Science
- Data Analytics
- driving a different kind of focus on data vs. statistics

Data Science Skills

- more emphasis on large samples, multivariate description, and data visualization
- understanding and exploring data vs. applying techniques to sample datasets
- students need to learn to think statistically and to develop an aesthetic for data handling and modelling based on solving practical problems
- see Ridgway, "Implications of the Data Revolution for Statistics Education", *International Statistical Review*, 2016

Data Literacy, first steps

- Data literacy, unlike research data management or data information literacy, places focus on the understanding and analysis of data
- As in all "literacy", competency and knowledge (of data) are important
- Statistical knowledge is required
- A greater mixture of data exploration, data cleaning, "data wrangling"
- Making sense of disparate data through visualization and other methods (e.g., machine learning)

Data Literacy, a definition

- An emerging area
  - "there is no standard definition for data literacy nor is there a definitive list of competencies" [Dechman and Syms, "Working Together to Maximize the Utilization of Open Data Across Social Science and Professional Disciplines", *Behavioral & Social Sciences Librarian*, 2014]

- But here is a reasonable start
  - "Data literacy can be defined, then, as the component of information literacy that enables individuals to access, interpret, critically assess, manage, handle and ethically use data" [Prado and Marzal, "Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents", *Libri*, 2013]

Data Literacy and Librarians

- Viewed in this way, librarians have traditionally trained users in how to access, interpret, handle, and ethically use information
- So extending those roles to data via data literacy seems more natual for librarians than the methodologies emphasized in statistical literacy
- Open Data and online analysis broaden the applicability of data literacy to a wide audience, increasing the potential for librarians to play a role.
- Data literacy can be applied to many disciplinary contexts, not just statistics. Social justice, urban planning, and other fields.
- Just as quantitative literacy is required for statistical literacy, statistical literacy is required for data literacy. It is necessary, but not sufficient.

Ingredients of Competencies

Data Literacy

Ryan Womack

Introduction

Statistical
Literacy

Data Literacy

Competencies

Role of NSOs

In Practice

Beyond statistical literacy...

- Ethics
- Real-world problem-solving context
- Ask questions from data
- Develop hypotheses and identify potential sources of data
- Collect or acquire data
- Analyse and create explanations from data
- Evaluate the validity of explanations based on data and formulate new questions
- more...?

Data Literacy Competencies, a working definition

- **Understanding Data** (Role of Data in Society)
- **Finding and/or Obtaining Data** (Sources, Search, Assessment)
- **Reading, Interpreting and Evaluating Data** (Statistical Literacy...)
- **Managing Data** (Metadata, Databases, practices)
- **Using Data** (Handling, Synthesis, Ethics)
- from Prado and Marzal, 2013

How to Develop Skills

- Training should be adapted to the level of the user
- Make it Personal
- Keep it Fun
- See MacCuirc, "You Don't Teach, Students Learn: A Report on a Project on Statistical Literacy in Ireland", *Austrian Journal of Statistics*, 2015
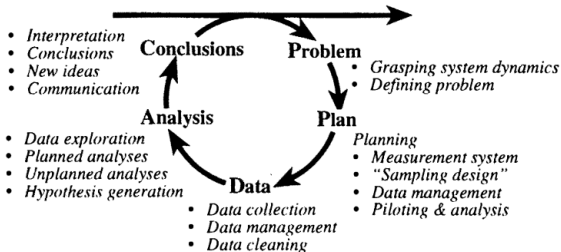
Types of Data Literate Citizen

- **Communicators** – make sense of and tell stories about data for others
- **Readers** – need skills to interpret data as part of their every day life
- **Makers** – need skills to integrate data real-world problem-solving
- **Scientists** – need strong technical data skills with domain expertise
- see Wolff, et. al. "Creating an Understanding of Data Literacy for a Data-driven Society", *Journal of Community Informatics*, 2017

PPDAC

Data Literacy

Ryan Womack

Introduction

Statistical
Literacy

Data Literacy

Competencies

Role of NSOs

In Practice

What kind of assignment?
Problem, Plan, Data, Analysis, Conclusions

Wild, Pfannkuch, "Statistical Thinking in Empirical Inquiry", *International Statistical Review*, 1999

National Statistical Offices

Data Literacy

Ryan Womack

Introduction

Statistical
Literacy

Data Literacy

Competencies

Role of NSOs

In Practice

- Most nations have a lead statistical agency, like the National Statistical Office of Mongolia, Census Bureau in the US, German Destatis (Federal Statistics Office), Russian Federal State Statistics Service
- We will refer to these in general as NSO
- NSO has a key role in the distribution of statistical information, as the official source for many statistics
- This role goes a long way back...
- See Lancaster, "How statistical literacy, official statistics and self-directed learning shaped social enquiry in the 19th and early 20th centuries," *Statistical Journal of the IAOS*, 2011

NSO educating for Statistical Literacy

Data Literacy

Ryan Womack

Introduction
Statistical
Literacy
Data Literacy
Competencies
Role of NSOs
In Practice

Some literature on this:

- Gal and Murray, "Responding to diversity in users' statistical literacy and information needs: Institutional and educational implications", *Statistical Journal of the IAOS,* 2011

- Gal, "Teaching for Statistical Literacy and Services of Statistics Agencies", *The American Statistician*, 2003

- Townsend, "The national statistical agency as educator", *Statistical Journal of the IAOS,* 2011

NSO, recommended actions

- Encourage statistical education and outreach
- Publications should allow for customization by audience and statistical literacy level [our types of data literate citizens]
- Release Open Data to allow for discovery, sharing, and reuse
- Assess materials for comprehensibility and produce documents according to standards
- see Scheld, "Statistical literacy: A new mission for data producers Milo Schield" *Statistical Journal of the IAOS*, 2011

- Statistics Canada has an
    - Community Data Program
    - Data Liberation Initiative to expose students to microdata
    - Research Data Centres

- New Zealand provides training for a National Certificate of Official Statistics (18-month part time course), increasing statistical capacity of government workers, and sponsors Academic Honors papers.

Data Producers, Consumers, and Communicators

Data Literacy

Ryan Womack

Introduction
Statistical
Literacy
Data Literacy
Competencies
Role of NSOs
In Practice

- NSOs are data producers, and are often thought of as making products for data consumers
- Data communicators are an important intermediary category
- Interaction with data communicators to provide training and collect feedback will result in increased data literacy.
- Communicators include academics, librarians, journalists
- see Rumsey, "Discussion: Statistical Literacy: Implications for Teaching, Research, and Practice" and Harraway and Forbes, "Partnership between national statistics offices and academics to increase official statistical literacy", *Statistical Journal of the IAOS*, 2013

What to do

Data Literacy

Ryan Womack

Introduction
Statistical
Literacy
Data Literacy
Competencies
Role of NSOs
In Practice

Remember PPDAC - Problem, Plan, Data, Analysis, Conclusions

- Any tools that provide interactive, personalized exploration and analysis of data can be used for outreach and training in data literacy
- Tishkovskaya and Lancaster, "Statistical education in the 21st Century: a review of challenges, teaching innovations and strategies for reform", *Journal of Statistics Education*, 2012

IPUMS

Data Literacy

Ryan Womack

Introduction

Statistical
Literacy

Data Literacy

Competencies

Role of NSOs

In Practice

- Microdata to develop digital literacy
- IPUMS offers a uniform solution for internationally comparable data
- endorsed by 85 official statistical agences
- constantly expanding (Mongolia 2000 and 2009)
- Meier and Lam, "Creating Statistically Literate Global Citizens: The Use of IPUMS- International Integrated Census Microdata in Teaching", *Statistical Journal of the IAOS*, 2011

International Statistical Literacy Project

- The International Statistical Literacy Project (ISLP) is "the only international program dedicated exclusively to the promotion of statistical literacy"
- Now 70 country coordinators
- Links to resources such as Comparing Populations
- Other statistical and data organizations like the American Statistical Association and IASSIST have a role to play

ARTIST

Data Literacy

Ryan Womack

Introduction

Statistical Literacy

Data Literacy

Competencies

Role of NSOs

In Practice

- ARTIST (Assessment Resource Tools for Improving Statistical Thinking)
- Allows the construction of tests and assessments to meet specific goals in
  - Statistical Literacy
  - Statistical Reasoning
  - Statistical Thinking

Teaching with Data

Data Literacy

Ryan Womack

Introduction

Statistical
Literacy

Data Literacy

Competencies

Role of NSOs

In Practice

- **Teaching with Data** identifies specific learning modules matched to subjects
- Social justice topics and data
- Lesser, "Critical Values and Transforming Data: Teaching Statistics with Social Justice", *Journal of Statistics Education*, 2007
- Data Murals: teaching data with art
- Bhargava, et. al. "Data Murals: using the arts to build data literacy". *The Journal of Community Informatics*, 2016
- Open data allows data relevant to everyday lives to enter into a usable format. Users with different levels of data literacy can create, analyze, and consume personalized open data

US Data resources

- Data.gov - raw release of data
- American Factfinder- first stop for US Census Data (demographics)
- Census Data Visualization Gallery
- Drought

Other Data resources

- Just for fun, The website TylerVigen offers a new spurious correlation every day, by using open data sources and simple data-mining techniques.Social Explorer
- Data Journalism Handbook
- Social Explorer - commercial database to provide improved access to Census data
- Open data allows data relevant to everyday lives to enter into a usable format. Users with different levels of data literacy can create, analyze, and consume personalized open data

Data Archives

- General data archives are an essential tools for basic research using data
- ICPSR (Interuniversity Consortium for Political and Social Research) - largest social science data archive in the world
- UKDA (United Kingdom Data Archive)
- GESIS (Germany's Data Archive for the Social Sciences)
- Social Science Japan Data Archive and other national archives

Open Data in Mongolia

- Chinggis Khan - First Census in 1206
- Mongolia ranks 3rd in Openness for Government Data (of 125 developing countries) using OpenDataWatch information.
- Open Data Initiative of Mongolia
- 1212.mn for sharing and education
- A good foundation for data literacy!

Conclusion

- Data Literacy has been recognized as an essential component of education
- Initial development of official statistics and the discipline of statistics reflects the needs of the industrial age
- Now the information age is upon us
- Why Data Literacy?
- Data literacy lets us seek our own answers in the mass of big data, empowering future development

Төгсгөл

- Асуулт байна уу?
- Маш их баярлалаа!
- Намайг дагаарай...
    - https://youtube.com/librarianwomack
    - https://www.linkedin.com/in/ryanwomack
    - https://twitter.com/ryandata
    - https://ryandata.wordpress.com