



Carleton
UNIVERSITY

Canada's Capital University

**What are Data?
What are Statistics?
Why do they matter to me?**

Jane Fry

May 9, 2017



- **Definitions**
- **Dataphobia**
- **Data literacy**
- **Research data management (RDM)**
- **Why does this matter to you?**

- **Why important?**

- | To support arguments, hypotheses and theories
- | Provide snapshots in time
- | Show changes over time
- | Mentioned in the news often

■ What are data?

- | Interdisciplinary
- | “Facts or figures from which conclusions can be drawn.”
(Statistics Canada)
- | “Characteristics or information, usually numerical, that are collected through observation.” (OECD)
- | “Facts, measurements, recordings, records, or observations about the world collected by scientists and others, with a minimum of contextual interpretation. Data may be any format or medium taking the form of writings, notes, numbers, symbols, text, images, films, video, sound recordings, pictorial reproductions, drawings, designs or other graphical representations, procedural manuals, forms, diagrams, work flow charts, equipment descriptions, data files, data processing algorithms, or statistical records.” (RDC)

- **Why collect data**
 - | To provide information
 - | To help make informed decisions

- **Different types of data**
 - | Numerical
 - | Interviews
 - | Radio signals
 - | Digitized images
 - | ...

■ What are statistics?

- | Interdisciplinary
- | Numeric facts and figures created from data
- | “Numerical data relating to an aggregate of individuals; the science of collecting, analysing and interpreting such data.” (OECD)
- | “Statistics represent a common method of presenting information. In general, statistics relate to numerical data, and can refer to the science of dealing with the numerical data itself. Above all, statistics aim to provide useful information by means of numbers. Therefore, a good definition of statistics is “a type of information obtained through mathematical operations on numerical data.”” (Statistics Canada)

- **Statistics**
 - | Opposing view points
 - *Elitist vs validation**

- **Stats are part of our everyday lives**

- **Statistics anxiety***
 - | “Has been recognized as a deterrent to students finishing their degrees”
 - | Can be slightly lessened
 - *If an instructor is there to help immediately*
 - *By encouraging students to talk about it*
 - | Important because
 - *“affects ability to understand research articles, data analysis, and interpretation of analyses”*



- ***“In theory, statistics should help settle arguments. They ought to provide stable reference points that everyone – no matter what their politics – can agree on. Yet in recent years, divergent levels of trust in statistics has become one of the key schisms that have opened up in western liberal democracies. Shortly before the November presidential election, a study in the US discovered that 68% of Trump supporters distrusted the economic data published by the federal government. In the UK, a research project by Cambridge University and YouGov looking at conspiracy theories discovered that 55% of the population believes that the government “is hiding the truth about the number of immigrants living here”.*”**

- ***“Critical thinking is a key skill in media and information literacy, and the mission of libraries is to educate and advocate its importance.”****

- **2 parts**
 - | Understand
 - | Communicate

*Source: IFLA 2017

■ Be a better statistical consumer

| UK House of Commons Library “[Statistical Literacy Guide](#)”

- *To better understand stats used in press stories, debates, reports, books, ...*
- *How to make basic calculations*
- *Basic concepts*
- *How to read charts*
- *How to spot spin and inappropriate use of stats*
- *Some more advanced stats*

- **Reading charts and graphs***
 - | Need to know
 - *What is being measured*
 - *What is the unit of measurement*
 - *What is the relationship between the two*
 - *What is the sample size*
 - *Comparing apples to oranges*
 - *Do the numbers add up (e.g. the pie chart)*
 - *Do the x and y axes start at 0 (where applicable)*

*Source: J. Darragh (n.d.)



Prisoners on death row, 1954-2006

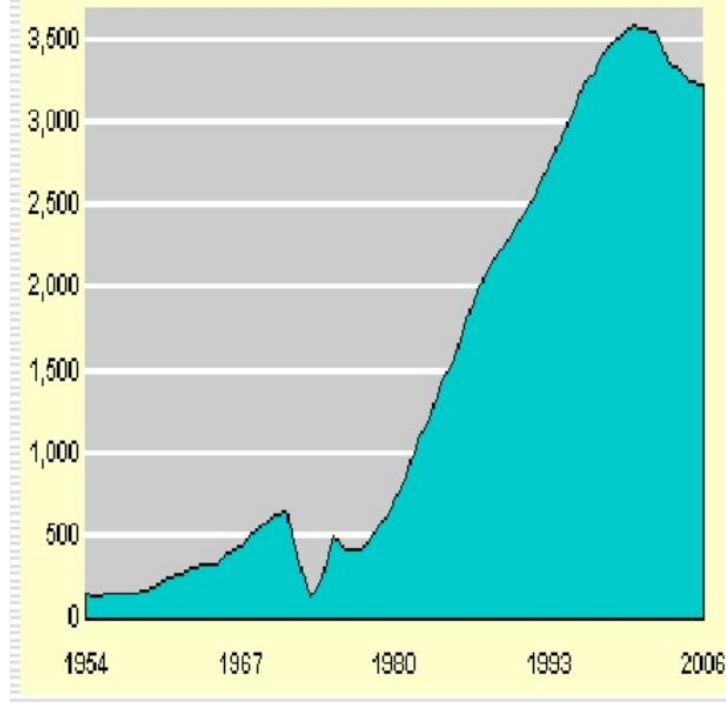


Chart source: Bureau of justice statistics prisoners under sentence of death trends chart. Retrieved 10/16/2008, 2008, from <http://www.ojp.usdoj.gov/bjs/glance/dr.htm>

*Source: J. Darragh (n.d.)



- Think like a statistician – without the math
 - | Pay attention to detail
 - | See the big picture
 - | No agendas
 - | Look outside the data
 - | Ask why

HOW TO SPOT FAKE NEWS



CONSIDER THE SOURCE

Click away from the story to investigate the site, its mission and its contact info.



READ BEYOND

Headlines can be outrageous in an effort to get clicks. What's the whole story?



CHECK THE AUTHOR

Do a quick search on the author. Are they credible? Are they real?



SUPPORTING SOURCES?

Click on those links. Determine if the info given actually supports the story.



CHECK THE DATE

Reposting old news stories doesn't mean they're relevant to current events.



IS IT A JOKE?

If it is too outlandish, it might be satire. Research the site and author to be sure.



CHECK YOUR BIASES

Consider if your own beliefs could affect your judgement.



ASK THE EXPERTS

Ask a librarian, or consult a fact-checking site.

- **Consider the source**
 - | Check out the website
 - | Check out the contact info

Source: <http://www.factcheck.org/2016/11/how-to-spot-fake-news/>



- **Go one step further**
 - | Don't just read the headlines
 - | Read the whole story
 - | Check out the links
 - | Check out the references

Source: <http://www.factcheck.org/2016/11/how-to-spot-fake-news/>



- **Check the Author**
 - | Are they credible
 - | Are they real



- **What are the supporting sources**
 - | Do the links work
 - | Do they support the data

Source: <http://www.factcheck.org/2016/11/how-to-spot-fake-news/>

- **Check the date**
 - | How recent is the data

Source: <http://www.factcheck.org/2016/11/how-to-spot-fake-news/>



- **Is it a joke?**
 - | Is it satire?
 - | Is it too outlandish?
 - | Go beyond
 - | Research the site
 - | Research the author

Source: <http://www.factcheck.org/2016/11/how-to-spot-fake-news/>



- **Check your biases**
 - | Very important as a librarian
 - | What are your beliefs
 - | Are they affecting your judgement

Source: <http://www.factcheck.org/2016/11/how-to-spot-fake-news/>

■ Ask the experts

- | The subject specialist
- | The data librarian
- | A fact-checking site
 - *FactCheck.org*
 - *Snopes.com*
 - *The Washington Post Fact Checker*
 - *PolitiFact.com*



- **Very important**
 - | All data should be cited
 - | Check out the citation

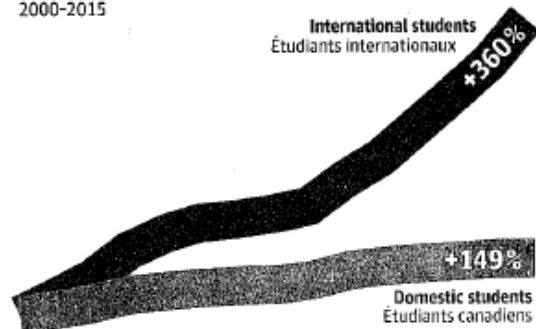
Student stats & facts

Chiffres et données sur la population étudiante postsecondaire

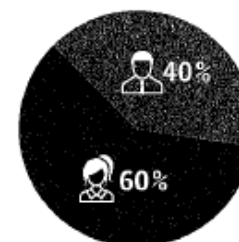
Average undergraduate tuition fees by province
Frais de scolarité moyens au 1^{er} cycle par province
2016-2017



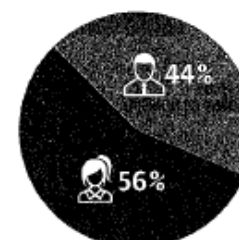
Domestic vs international student enrolment
Effectifs étudiants canadiens par rapport aux effectifs internationaux
2000-2015



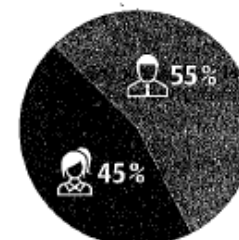
Degrees granted by gender
Diplômes décernés selon le sexe
2014-2015



Undergraduate / 1^{er} cycle



Masters / 2^e cycle



Doctorates / Doctorats

- **aka RDM**
- **aka data management**
- **aka DM**

- **What is it?**
- **Includes**
 - | Sound practices
 - | Data curation
 - | Data stewardship

- **Research data**

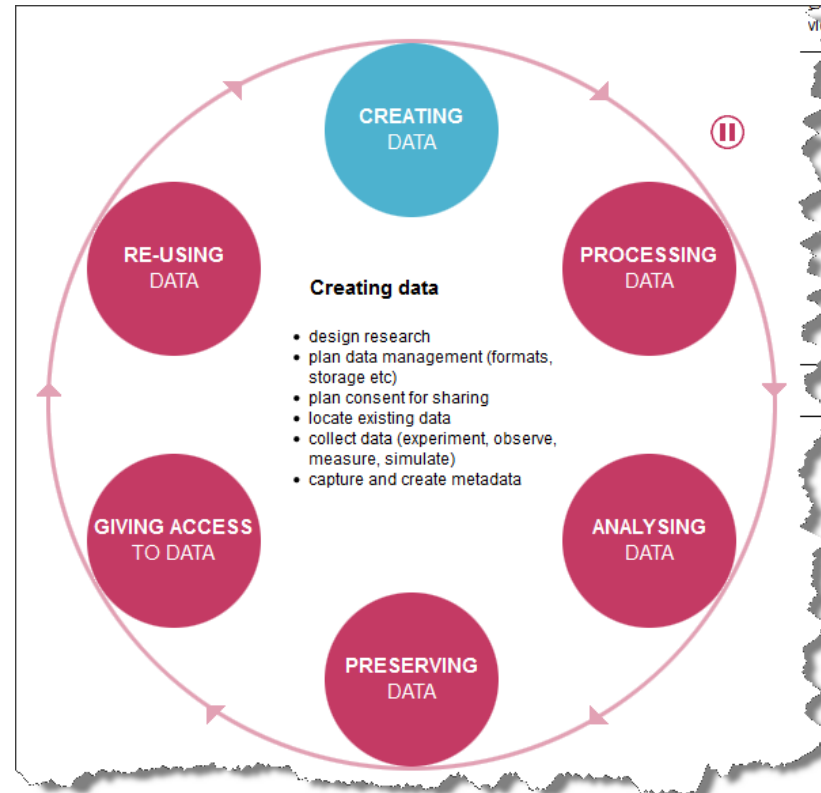
- | What are they important?
- | What is sharing?

- **Benefits of RDM**

- | Confirmation of original findings
- | Further research
- | Planning follow-up studies
- | Bonus ...

■ Metadata

- | The key to research data re-use
- | Standardized and structured
- | Explains *everything* about the dataset
 - *Purpose*
 - *Time frame*
 - *Geographic location*
 - *PI(s)*
 - *Questions*
 - ...
- | Who is the best person?



- **Have to remain current with your job**
 - | Check out ACRL bi-annual reports
- **ACRL Top Trends (2016)**
 - | Research data services
 - *Informational and consultative*
 - | Data policies and data mgmt plans
 - | Professional development for librarians providing data services
- **ACRL Top Trends (2014)**
 - | Data
 - *“provide data mgmt support through the research process”*
 - *Help researchers to share, analyze and re-use data*
 - *Which repository to use*
 - *Which journal requires underlying data to be published*

- **ACRL Environmental Scan (2017)**
 - | Evidence-based decision making in academic libraries
 - | Information literacy issues
 - | Curating research data

- **ARL Annual Salary Survey, 2014-2015**
 - | [Salary Trends in Canadian ARL University Libraries](#)

- **“ARL Academic Law Library Statistics: 2014-2015”**
 - | Library expenditures
 - *Collections and collection expenditures*
 - *Salary expenditures*
 - *Volumes in Library*
 - *Titles Held*
 - *Total staff*

COLLECTIONS AND COLLECTION EXPENDITURES

	Notes	Titles Held 1	Volumes In Library 2	Electronic Books 4	One-time resource purchases 7a	Ongoing resource purchases 7b	Collection Support 7c	Total Library Materials 7
BRITISH COLUMBIA	b+	112,048	256,360	22,167	66,288	788,863	4,076	859,227
TORONTO	+	.	236,987	10,202	76,345	937,763	8,577	1,022,684
WESTERN	b+	59,688	113,087	.	105,672	629,391	7,080	742,123
YORK	+	356,519	852,254	117,211	228,301	490,171	11,038	729,510

COLLECTIONS AND COLLECTION EXPENDITURES Summary Data

	Titles Held 1	Volumes In Library 2	Electronic Books 4	One-time resource purchases 7a	Ongoing resource purchases 7b	Collection Support 7c	Total Library Materials 7
ARL 2014-15 (Cont'd)							
Mean	328,535	543,032	154,788	214,162	1,125,406	57,506	1,394,361
Median	230,259	461,436	68,929	105,672	1,026,691	25,117	1,253,729
High	1,315,828	1,971,068	1,224,781	1,071,017	3,013,280	762,249	4,609,539
Low	43,162	100,235	1	3,513	155,467	850	384,914
Sum	22,997,431	39,641,335	9,442,049	15,633,813	82,161,233	3,795,381	103,182,702
n	70	73	61	73	73	66	74

Why? (cont'd)

- **Fulfilling your job description**
 - | Finding data to support your researchers
 - | Conducting your own research (on sabbatical?)
- **Carleton U Legal Studies Librarian**
 - | “Strong knowledge of reference techniques and bibliography, including both print and electronic information resources, particularly relevant to Legal Studies, Criminology & Criminal Justice and Human Rights.” (2012)
- **U of Saskatchewan Law Librarian (2017)**
 - | “... providing in-depth research support”
 - | “... teaching information literacy skills”



■ Conferences

| CALL/ACBD 2017

- *Poster and Infographics Session*

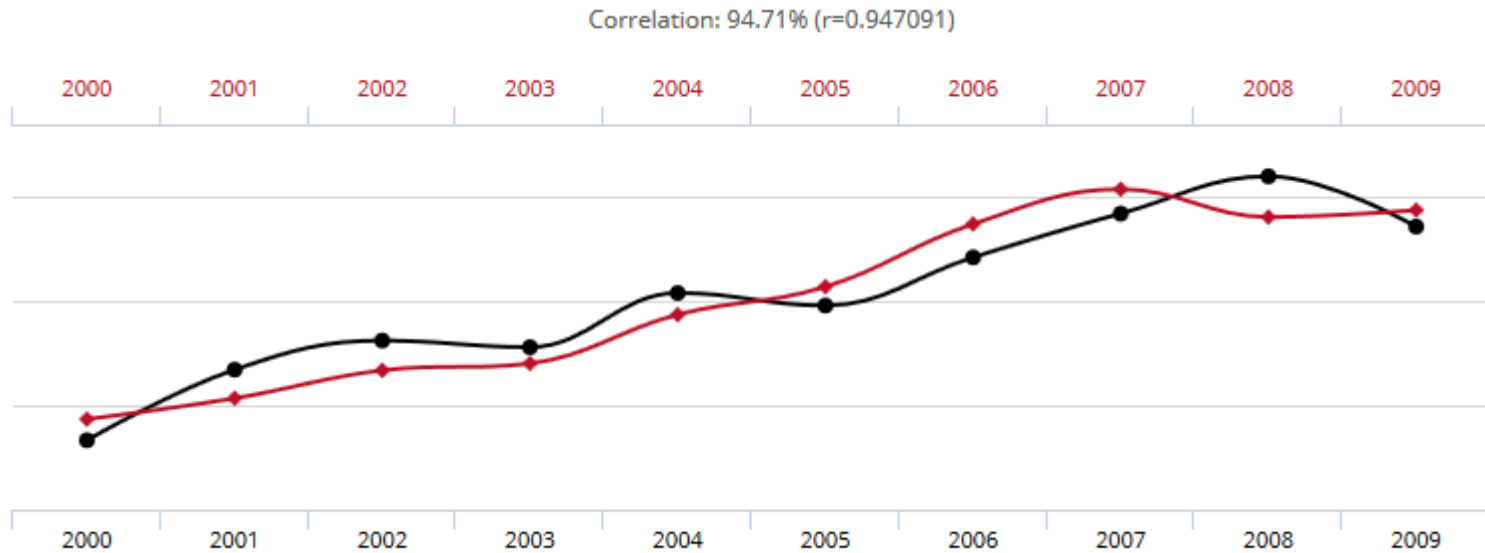
| Professional organizations

- *CALL/ACBD Compensation surveys (salary and benefits)*
 - [2013](#)
 - [2006](#)
 - [2003](#)



- **Things to remember**
 - | Sponsorship
 - | Who collected the data
 - | Does the author use anecdotal evidence when data is available

- **Think!**

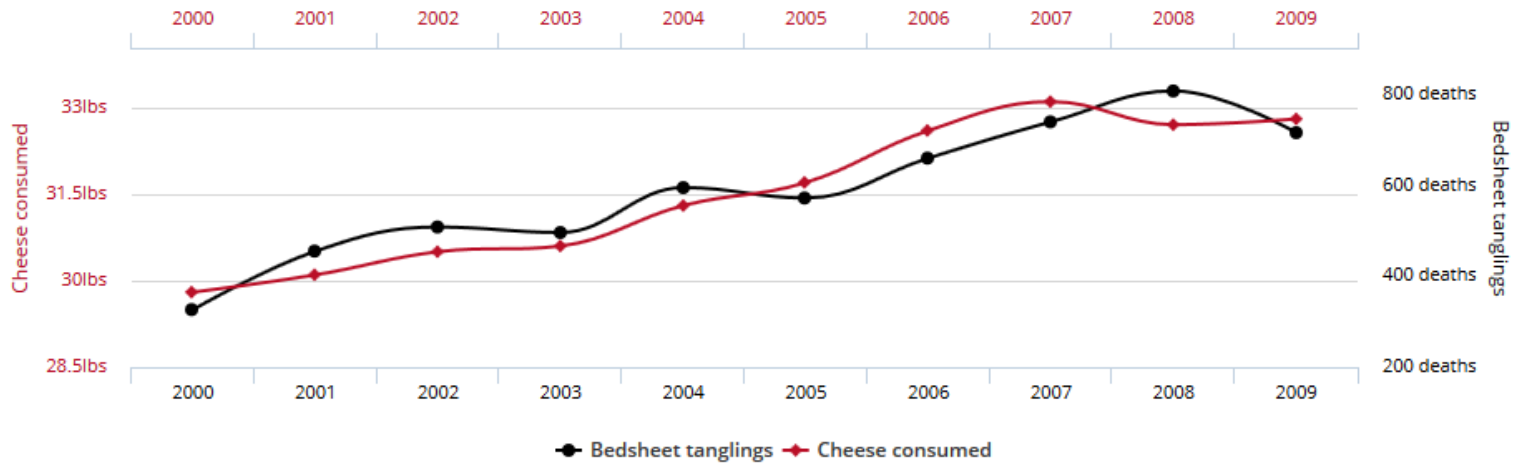




Per capita cheese consumption correlates with

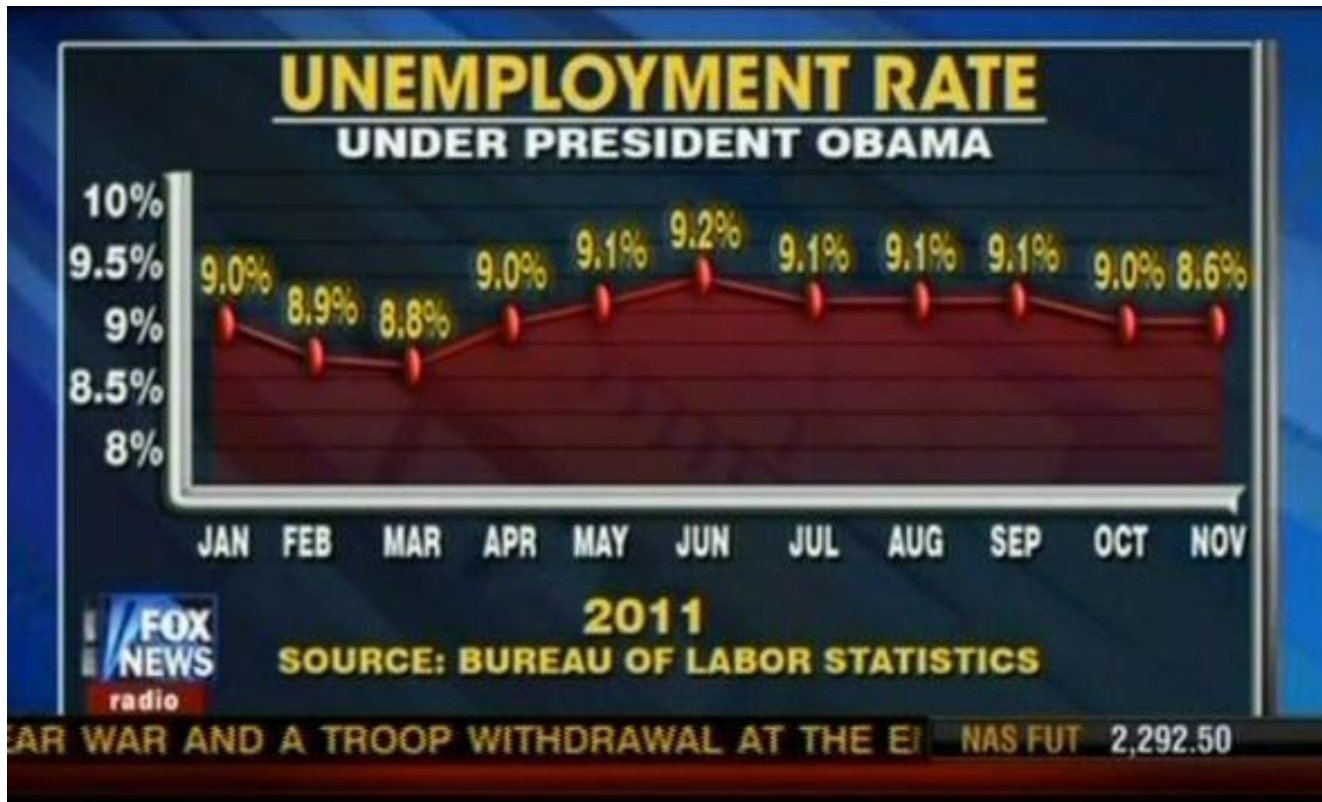
Number of people who died by becoming tangled in their bedsheets

Correlation: 94.71% (r=0.947091)

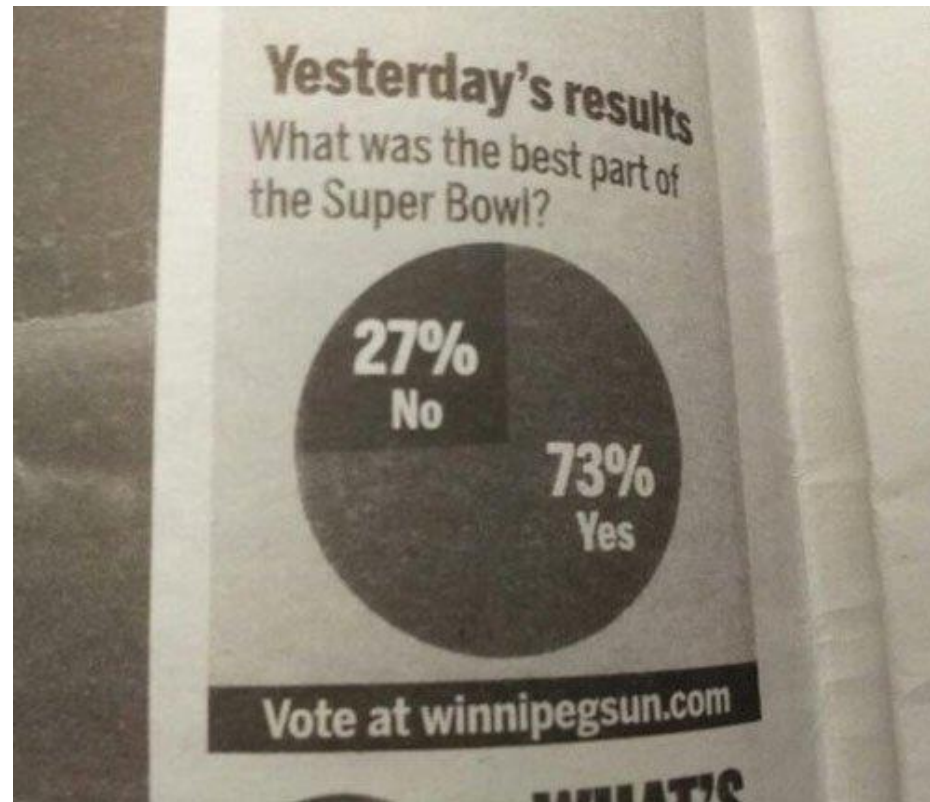


Data sources: U.S. Department of Agriculture and Centers for Disease Control & Prevention

tylervigen.com

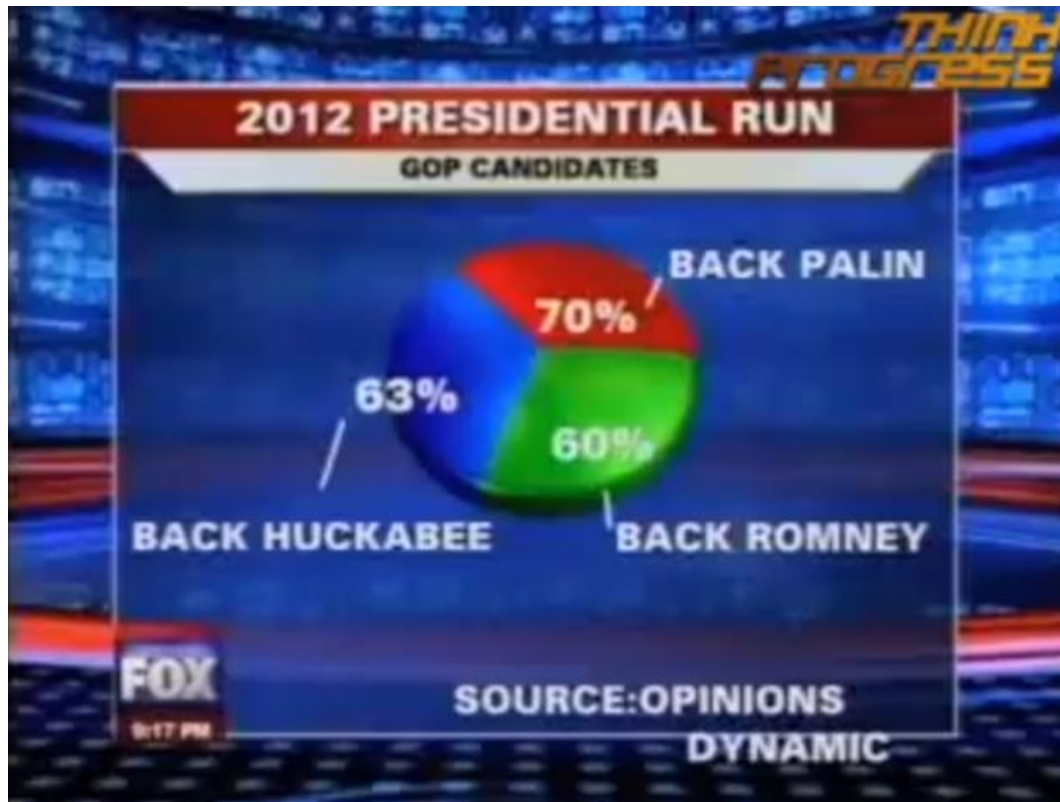


Source: Fox News (December 12, 2011)
<http://flowingdata.com/category/statistics/mistaken-data/>



Source: Winnipeg Sun (2013)

<http://flowingdata.com/category/statistics/mistaken-data/>



Source: Fox News (2009)

<http://flowingdata.com/category/statistics/mistaken-data/>



■ Reliability and Reproducibility

- | Political Persuasion and Attitude Change Study: The Los Angeles Longitudinal Field Experiments, 2013-2014
- | Principal Investigator: Michael J. LaCour



Irregularities in LaCour (2014)

David Broockman, Assistant Professor, Stanford GSB (as of July 1),
dbroockman@stanford.edu

Joshua Kalla, Graduate Student, UC Berkeley, kalla@berkeley.edu

Peter Aronow, Assistant Professor, Yale University, peter.aronow@yale.edu

May 19, 2015

Summary

We report a number of irregularities in the replication dataset posted for LaCour and Green (*Science*, “When contact changes minds: An experiment on transmission of support for gay equality,” 2014) that jointly suggest the dataset (LaCour 2014) was not collected as described. These irregularities include baseline outcome data that is statistically indistinguishable from a national survey and over-time changes that are unusually small and indistinguishable from perfectly normally distributed noise. Other elements of the dataset are inconsistent with patterns typical in randomized experiments and survey responses and/or inconsistent with the claimed design of the study. A straightforward procedure may generate these anomalies nearly exactly: for both studies reported in the paper, a random sample of the 2012 Cooperative Campaign Analysis Project (CCAP) form the baseline data and normally distributed noise are added to simulate follow-up waves.

Timeline of Disclosure

- January - April, 2015. Broockman and Kalla were impressed by LaCour and Green (2014) and wanted to extend the article’s methodological and substantive discoveries. We began to plan an extension. We sought to form our priors about several design parameters based on the patterns in the original data on which the paper was based, LaCour (2014). As we examined the study’s data in planning our own studies, two features surprised us: voters’ survey responses exhibit much higher test-retest reliabilities than we have observed in any other panel survey data, and the response and reinterview rates of the panel survey were significantly higher than we expected. We set aside our doubts about the study and awaited the launch of our pilot extension to see if we could manage the same parameters. LaCour and Green were both responsive to requests for advice about design details when queried.
- May 6, 2015. Broockman and Kalla launch a pilot of the extension study.

More details about our questions about the data are as follows: [https://www.stanford.edu/~dbroockman/la_cour_questions.pdf](#)



When contact changes minds: An experiment on transmission of support for gay equality

Michael J. LaCour¹, Donald P. Green²

+ Author Affiliations

Science 12 Dec 2014:

Vol. 346, Issue 6215, pp. 1366-1369

DOI: 10.1126/science.1256151

Article

Figures & Data

Info & Metrics

eLetters

 PDF

You are currently viewing the abstract.

View Full Text



This article has been retracted. Please see:

[Is retracted by - June 05, 2015](#)



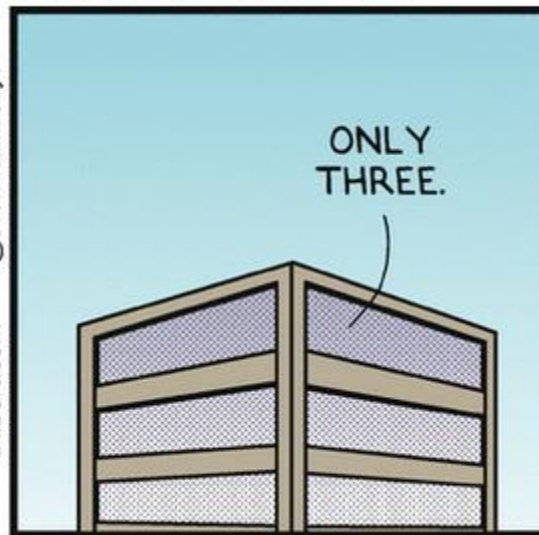
- **We now know**
 - | Definition of data and statistics
 - | Dataphobia
 - | Data literacy
 - | Research data management
 - | Why all this matters to you



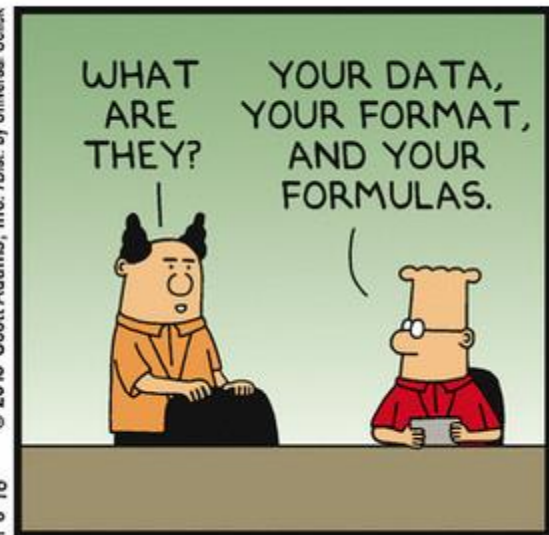
Wednesday January 06, 2016 *Three Problems With Spreadsheet*



Dilbert.com @ScottAdamsSays



© 2016 Scott Adams, Inc. /Dist. by Universal Uclick
1-6-16





- ACRL Research Planning and Review Committee (March 2017). “Environmental Scan”.
- ACRL Research Planning and Review Committee (June 2016). “[2016 top trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education](#)”.
- ACRL Research Planning and Review Committee (June 2014). “[Top trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education](#)”.
- ARL. “[ARL Annual Salary Survey 2015-2015](#)”.
- CAUT (April 2017). “bulletin: Canada’s voice for academics”. 64(4). Ottawa, ON.
- Darragh, Jen (n.d.). “[Building statistical literacy: reading charts and graphs.](#)”
- Davies, William (January 19, 2017). “[How statistics lost their power – and why we should fear what comes next](#)”. The Guardian.

- House of Commons Library (January 22, 2009). “[Statistical Literacy Guide](#)”. UK Parliament.
- Kiely, E. & L. Robertson (November 18, 2016). “[How to Spot Fake News](#)”. FactCheck.org
- IFLA (2017). “[How to spot fake news](#)”.
- Morris, S. & G. Roebuck (2017). “[ARL Academic Law Library Statistics: 2014-2015.](#)”
- OECD. “[Glossary of statistical terms](#)”.
- Research Data Canada. “[Original RDC Glossary](#)”
- Statistics Canada. “[Statistics: Power from data! Glossary](#)”.
- UK Data Archive. “[Research data lifecycle](#)”.
- Williams, Amanda S. (2010). “[Statistics anxiety and instructor immediacy](#)”. Journal of Statistics Education v18(2).
- Yau, Nathan (n.d.). “[Think like a Statistician – without the math](#)”.



- **Best, J. (2004). More damned lies and statistics. Berkeley, California: University of California Press.**
- **CARL. “[Portage: Shared stewardship of research data](#)”.**
- **Carleton University. “[Research data Management](#)”.**
- **Johnson, J.H. & Mike Gluck (2016). Every data: The misinformation hidden in the *little data* you consume every day. Bibliomotion, Inc. Brookline, MA**
- **Statistics Canada (2007). “[Finding and using statistics](#)”.**
- **York University Libraries. “[Data & statistical literacy: A research guide.](#)”**

Contact Information

Jane Fry
Data Services Librarian
Rm 122
MacOdrum Library
613.520.2600 x1121
jane.fry@Carleton.ca

<http://www.library.carleton.ca/find/data>