

2016

Data, Data, Data

Mary Whisner

University of Washington School of Law, whisner@uw.edu

Follow this and additional works at: <https://digitalcommons.law.uw.edu/librarians-articles>

 Part of the [Legal Writing and Research Commons](#), and the [Statistics and Probability Commons](#)

Recommended Citation

Mary Whisner, *Data, Data, Data*, 108 *LAW LIBR. J.* 313 (2016), <https://digitalcommons.law.uw.edu/librarians-articles/18>

This Article is brought to you for free and open access by the Librarians' Publications at UW Law Digital Commons. It has been accepted for inclusion in Librarians' Articles by an authorized administrator of UW Law Digital Commons. For more information, please contact cnyberg@uw.edu.

Data, Data, Data*

Mary Whisner**

The legal profession often requires extensive data for everything from simple statistical questions to large-scale empirical research projects. Ms. Whisner discusses some of her favorite sources for finding and evaluating statistics.

I have no data yet. It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.—Sherlock Holmes¹

“Data! data! data!” he cried impatiently. “I can’t make bricks without clay.”—Sherlock Holmes²

For the rational study of the law the black-letter man may be the man of the present, but the man of the future is the man of statistics and the master of economics.—Oliver Wendell Holmes, Jr.³

¶1 I’ve heard law students and lawyers joke that they chose law because they’re not good at math. But, like both Holmeses, legal professionals often hunger for data. Numbers help us understand the world and can inform policy arguments.⁴ And so

* © Mary Whisner, 2016.

** Reference Librarian, Marian Gould Gallagher Law Library, University of Washington School of Law, Seattle, Washington.

1. Arthur Conan Doyle, *A Scandal in Bohemia*, STRAND MAG., July 1891, at 61, 63.

2. Arthur Conan Doyle, *The Adventure of the Copper Beeches*, in 1 THE COMPLETE SHERLOCK HOLMES 377, 383 (2004).

3. Oliver Wendell Holmes, Jr., *The Path of the Law*, 10 HARV. L. REV. 457, 469 (1897).

4. One surprising twist on the influence of statistics comes from a famous experiment that gave students \$5 for their participation and then asked them how much they’d like to donate to a charity fighting hunger in Africa. Students who were given statistics about the scope of the problem gave significantly less than those who were told a story about one poor child; those given statistics *and* the individual story also gave less. It was almost as if the statistics turned off their compassion. Deborah A. Small, George Loewenstein & Paul Slovic, *Sympathy and Callousness: The Impact of Deliberative Thought on Donations to Identifiable and Statistical Victims*, 102 ORG. BEHAV. & HUM. DECISION PROCESSES 143 (2007); see also Deborah A. Small & George Loewenstein, *Helping a Victim or Helping the Victim: Altruism and Identifiability*, 26 J. RISK & UNCERTAINTY 5 (2003).

I assumed that this experiment is “famous” because I’ve read about it in two different books: DAN ARIELY, *THE UPSIDE OF IRRATIONALITY: THE UNEXPECTED BENEFITS OF DEFYING LOGIC AT WORK AND AT HOME* 239–41, 245–48 (2010); CHIP HEATH & DAN HEATH, *MADE TO STICK: WHY SOME IDEAS SURVIVE AND OTHERS DIE* 164–67 (2007). To quantify the experiment’s fame, I ran the article through Web of Science and found 173 citing references. Search conducted Jan. 26, 2016: CITED AUTHOR: (small d*) AND CITED WORK: (ORGAN BEHAV HUM DEC) AND CITED YEAR: (2007). Fame among scholarly authors is one thing. The books aimed at nonspecialists surely reach more people overall. Browsing the Amazon Best Sellers Rank for *Made to Stick*, for example, I see that the hardback

we reference librarians are often asked to find statistics. I am not talking here about huge datasets needed for some types of empirical legal research.⁵ I just want to discuss the garden-variety, apparently simple questions we get, like “How many African American men are in prison?” or “How much money is spent on global health?” I’d like to share some of my favorite sources and reflect on some of the challenges of finding and evaluating statistics.

Legal Profession

¶2 Recently I was asked whether it’s true that most lawyers are in solo or small-firm practice. I was happy to have a good source handy: the American Bar Association’s Lawyer Demographics page (a one-page fact sheet).⁶ The page gives the percentage of lawyers in private practice who practice in different sizes of firm. I did just a little arithmetic to give the percentage of lawyers overall:

- 75% of lawyers are in private practice.
- 49% of lawyers who are in private practice are in solo practice. Therefore, 36.8% of all lawyers are in solo practice.
- 14% of lawyers who are in private practice are in firms of 2–5 lawyers. Therefore, 10.5% of all lawyers are in firms of 2–5 lawyers.
- 6% of lawyers who are in private practice are in firms of 6–10 lawyers. Therefore, 4.5% of all lawyers are in firms of 6–10 lawyers.
- 69% of lawyers who are in private practice are solo or in firms of 10 lawyers or fewer. Therefore, 51.8% of all lawyers are in small or solo practice.

I was careful to point out that the numbers were from 2005 and that it was possible that proportions had changed since the Great Recession⁷ and the downturn in the market for lawyers.⁸ But I also noted that the percentage of practitioners in solo

edition is #6 in Books > Business & Money > Skills > Communications. That is consistent with my own experience (I read it and I’ve seen it in airport bookstores). *Made to Stick* was even discussed in this journal: Jean M. Holcomb, *Got Ideas?*, 100 LAW LIBR. J. 587, 2008 LAW LIBR. J. 28.

The attempt to quantify fame illustrates the point: sometimes anecdotal information (“I’ve read about this experiment twice, so it’s famous”) is more meaningful than statistics (e.g., citation counts or best seller lists). Using it can also be quicker and easier than carefully compiling numbers.

5. See Allison C. Reeve & Travis Weller, *Empirical Legal Research Support Services: A Survey of Academic Law Libraries*, 107 LAW LIBR. J. 399, 2015 LAW LIBR. J. 19.

6. *Lawyer Demographics Year 2015*, AM. BAR ASS’N, http://www.americanbar.org/content/dam/aba/administrative/market_research/lawyer-demographics-tables-2015.authcheckdam.pdf [<https://perma.cc/5MVK-GKC6>]. The person who asked the question had searched on her own and found Joseph D. Lawson, *What About the Majority? Considering the Legal Research Practices of Solo and Small Firm Attorneys*, 106 LAW LIBR. J. 377, 2014 LAW LIBR. J. 21. Lawson cites the 2013 version of this fact sheet in his third footnote, but the researcher had not mined the article’s footnotes. Bright, competent people can get close to the information they want and still miss it—one reason there is still a role for reference librarians.

7. Robert Rich, *The Great Recession of 2007–09*, FED. RESERVE HISTORY (Nov. 22, 2013), <http://www.federalreservehistory.org/Events/DetailView/58> [<https://perma.cc/724H-EXCA>] (“The Great Recession began in December 2007 and ended in June 2009, which makes it the longest recession since World War II.”).

8. See, e.g., Nicole Black, *The Myth of the Upper-Middle-Class Lawyer*, GPSOLO, Oct. 2012, at 56, 58 (“[T]he legal job market has declined significantly since the recession hit in 2008. According to seasonally adjusted data from the U.S. Bureau of Labor Statistics, the legal services sector has lost some 54,000 jobs since May 2008.”).

practice has held steady for a long time—49% in 1980, 45% in 1991, 48% in 2000, and 49% in 2005—despite ups and downs in the economy.

¶3 I've used this fact sheet (or its previous editions) for years. Even though I'm aware of it, it sometimes takes me a couple of clicks to remember where it is. In case you need this path: From the ABA's home page, go to "Resources for Lawyers," then choose "Legal Profession Statistics,"⁹ where you'll find this fact sheet and many other valuable sources. Also worth looking at is the Diversity Resources page,¹⁰ which links to the *National Database on Judicial Diversity in State Courts*.¹¹

¶4 Because I thought the requestor would be especially interested in our state, I made a table with the corresponding numbers from *The Lawyer Statistical Report*, which has detailed state-by-state data.¹² I love this book of data, produced by the American Bar Foundation, and wish only that they could crunch the numbers faster so we could have more current information.

¶5 For more on the legal profession, check out NALP's research projects and publications.¹³ NALP started out as the National Association for Law Placement but more often goes by its acronym. Its researchers are following trends in the profession, such as diversity, recruitment, and professional development. Many reports are free online; more substantial reports can be purchased from the NALP online bookstore. But that's not all: there's a separate website for the NALP Foundation,¹⁴ which funds a lot of research, notably the "After the JD" project, an ambitious longitudinal study of 5000 law graduates during the first ten years of their careers.¹⁵

Bibliometrics

¶6 Some data requests are bibliometric—that is, they call for information about publication or citation patterns. Over the years, one of our professors has asked us

9. *Legal Profession Statistics*, AM. BAR ASS'N, http://www.americanbar.org/resources_for_lawyers/profession_statistics.html (last visited Jan. 23, 2016).

10. *Diversity Resources*, AM. BAR ASS'N, <http://www.americanbar.org/groups/judicial/resources/diversity.html> (last visited Jan. 23, 2016).

11. *National Database on Judicial Diversity in State Courts*, AM. BAR ASS'N, <http://apps.americanbar.org/abanet/jd/display/national.cfm> (last visited Jan. 23, 2016). For federal courts, use the Federal Judicial Center's *Biographical Directory of Federal Judges, 1789–Present*, <http://www.fjc.gov/history/home.nsf/page/judges.html> (last visited Jan. 23, 2016), which allows you to sort by race or ethnicity (among other variables). Or use the lists that have already been generated: *Diversity on the Bench*, FED. JUD. CTR., http://www.fjc.gov/history/home.nsf/page/judges_diversity.html (last visited Jan. 23, 2016).

12. CLARA N. CARSON & JEEYUN PARK, *THE LAWYER STATISTICAL REPORT: THE U.S. LEGAL PROFESSION IN 2005* (2012).

13. NALP, <http://www.nalp.org/research> (last visited Jan. 23, 2016) ("NALP is the premier resource for information on legal employment and recruiting. Analysis of data sources such as the Employment Report and Salary Survey, the Associate Salary Survey, the NALP Directory of Legal Employers, and others allows NALP to provide comprehensive information on a variety of topics.").

14. NALP FOUND., <http://www.nalpfoundation.org/> (last visited Jan. 23, 2016). For what it's worth, I didn't notice the separate Foundation website—with its separate bookstore—until I'd been looking around the NALP website for some time.

15. So far there are three major reports: NALP FOUND. & AM. BAR FOUND., *AFTER THE JD: FIRST RESULTS OF A NATIONAL STUDY OF LEGAL CAREERS* (2004), <http://www.nalpfoundation.org/uploads/50-AJD1.pdf> [<https://perma.cc/67V7-AMWP>]; NALP FOUND. & AM. BAR FOUND., *AFTER THE JD II: SECOND RESULTS FROM A NATIONAL STUDY OF LEGAL CAREERS* (2009); NALP FOUND. & AM. BAR FOUND., *AFTER THE JD III: THIRD RESULTS FROM A NATIONAL STUDY OF LEGAL CAREERS* (2014); see also *After the JD*, AM. BAR FOUND., <http://www.americanbarfoundation.org/research/project/44> [<https://perma.cc/NY78-YZV9>].

to give him counts of the number of cases citing a particular statute or law review articles using a given word in the title.¹⁶ Another asked for a decade-by-decade count of “cost-benefit” or “benefit-cost” in law review titles. Running searches like these is not hard, but we should be careful in explaining the results. For example, my colleague who did the “cost-benefit” search explained to the professor that an article published in November 1949 can be retrieved in a HeinOnline search limited to 1950–1959, because the journal date is coded as 1949–1950. She also noted that the HeinOnline searches picked up articles from a few foreign journals, but coverage of foreign journals is not comprehensive. Still, the numbers she found gave a sense of the snowballing interest in cost-benefit analysis in law journals, from 36 articles in the 1950s, to 383 in the 1960s, to 3126 in the 1970s and so on to 11,999 in 2000–2009.¹⁷

¶7 Of course a category of bibliometric data that academics are very interested in is citation patterns—in particular, citations to their own work. As individuals scramble for tenure and merit raises and law schools scramble for boosts in their rankings, people are looking for metrics—something more easily quantifiable than “This article was interesting, original, and useful.” And they turn to citation counts and downloads.

¶8 The numbers are illustrative, but often need explaining. For instance, one professor wondered why Westlaw showed several citations to one of her papers while SSRN showed none. Why did SSRN miss them? I explained that SSRN’s algorithm for counting citing references looks only at references that are in a list at the end of a paper—a format that is common in many disciplines, but not law, which uses footnotes. SSRN’s system also counts only citing references in papers that are posted on SSRN, and authors who cite you might not post there. KeyCite and Shepard’s reflect only citing references that appear in journals, cases, and treatises that are on Westlaw and LexisNexis, respectively, and so will miss citations in other works. HeinOnline Author Profiles pick up citations in works that are on HeinOnline, but they can undercount if the author has published under different names (e.g., with and without a middle initial, or “Bill” and “William”).¹⁸ If an author is cited by papers outside law—for example, philosophy, political science, or public health—then those citations won’t be picked up by Westlaw and LexisNexis. Citations to coauthored works might be missed when a second or third author is subsumed under “et al.”¹⁹ And so on. Because of all these complexities, when we have been asked to look for works citing a professor who is up for tenure, we try multiple sources: Web of Science, KeyCite, Shepard’s, Google Scholar, Google Books, SSRN. Trying to be thorough (albeit short of absolutely comprehensive) takes a lot of work. For some purposes, it is better to use a simple (but incomplete) measure. For instance, it is easy to find out which of our professors who post on SSRN have the most downloads, and that can

16. Sometimes the results make their way into one of his footnotes. See, e.g., WILLIAM H. RODGERS, 3 ENVIRONMENTAL LAW § 5:28 n.43, Westlaw (database updated Dec. 2015) (“A computer search in January of 1987 yields more than 900 decisions containing references to ‘pesticides’ and related subjects.”); Anna T. Moritz et al., *Biodiversity Baking and Boiling: Endangered Species Act Turning Down the Heat*, 44 TULSA L. REV. 205, 213 n.75 (2008) (“A Westlaw keyword search of ‘polar bear’ and ‘Endangered Species Act’ within the Journals and Law Reviews database returns 90 articles since March 2006.”).

17. HeinOnline searches performed by Cheryl Nyberg on Oct. 22, 2013.

18. If an author writes to Hein, the staff can unite the records for the different names.

19. See THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION R. 15.1(b), at 150 (20th ed. 2015).

be informative, even if we are aware that other readers are getting articles from other platforms (or even from print journals!).

Statistical Gaps

¶9 Sometimes the statistics a researcher wants just haven't been compiled. In the last two years, public outcry has called attention to the deaths of people of color at the hands of (or while in the custody of) police.²⁰ These deaths “raised two closely related questions: How many people are killed by police every year, and how many people are shot by police each year? People who wanted to know—from reporters to FBI Director James Comey—seemed surprised to learn that there are no reliable statistics to answer [those questions].”²¹ Journalists and nonprofits have moved to gather better data. The *Guardian*, a British newspaper with strong coverage of the United States and elsewhere, created *The Counted*, a database of people killed by the police in 2015 (and going forward); it combines *Guardian* reporting with “verified crowdsourced information.”²² At the same time, the *Washington Post* began counting fatal shootings by police.²³ The project revealed how inadequate the government statistics were:

The *Post* documented well more than twice as many fatal shootings this year as the average annual tally reported by the FBI over the past decade. The FBI and the federal Bureau of Justice Statistics now acknowledge that their data collection has been deeply flawed. FBI Director James B. Comey called his agency's database “unacceptable.”

Both agencies have launched efforts to create new systems for documenting fatalities.²⁴

20. See, e.g., David A. Graham, *A Year After Eric Garner's Death, Has Anything Changed?*, ATLANTIC (July 17, 2015), <http://www.theatlantic.com/politics/archive/2015/07/eric-garner-anniversary/398837/> [<https://perma.cc/V7YU-S5ZE>] (listing Eric Garner, Michael Brown, Tamir Rice, Walter Scott, and Freddie Gray, and adding Sandra Bland, who had that week been found dead in her jail cell). For a personal reflection on these events, see Ronald E. Wheeler, Jr., *Michael Brown, Eric Garner, and Law Librarianship*, 107 LAW LIBR. J. 467, 2015 LAW LIBR. J. 22.

21. Graham, *supra* note 20.

22. *About the Project*, GUARDIAN, <http://www.theguardian.com/us-news/ng-interactive/2015/jun/01/about-the-counted> [<https://perma.cc/D6YZ-TU7H>].

23. . . . *People Shot Dead by Police This Year*, WASH. POST, <https://www.washingtonpost.com/graphics/national/police-shootings/> (last visited Jan. 24, 2016) (continually updated infographic). The *Post* was recently awarded a Pulitzer Prize for its coverage. *The Washington Post Staff*, THE PULITZER PRIZES, <http://www.pulitzer.org/winners/washington-post-staff> [<https://perma.cc/4D3A-XG6K>].

24. *A Year of Reckoning: Police Fatally Shoot Nearly 1,000*, WASH. POST (Dec. 26, 2015), <http://www.washingtonpost.com/sf/investigative/2015/12/26/a-year-of-reckoning-police-fatally-shoot-nearly-1000/> [<https://perma.cc/6KLL-L7XQ>]. See Press Release, U.S. Dep't of Justice, Ofc. of Pub. Affairs, Attorney General Lynch: Use-of-Force Data Is Vital for Transparency and Accountability (Oct. 5, 2015), <http://www.justice.gov/opa/pr/attorney-general-lynch-use-force-data-vital-transparency-and-accountability> [<https://perma.cc/G7KU-7S7Z>]; FBI Nat'l Press Ofc., FBI Releases 2014 Crime Statistics from the National Incident-Based Reporting System (Dec. 14, 2015), <https://www.fbi.gov/news/pressrel/press-releases/fbi-releases-2014-crime-statistics-from-the-national-incident-based-reporting-system> (“Director Comey said, ‘. . . without information, every single conversation in this country about policing and reform and justice is uninformed, and that is a very bad place to be.’”).

As early as 2009, the Cato Institute had begun collecting information about police misconduct, including nonfatal incidents. *National Police Misconduct Reporting Project*, CATO INST., <http://www.policemisconduct.net/> (last visited Jan. 24, 2016). The set of misconduct incidents is larger than the sets of fatalities compiled by the *Guardian* and the *Post*. But note that not all fatalities are misconduct.

Expertise and Hidden Assumptions

¶10 Few of us—reference librarians or legal researchers—have serious backgrounds in statistical analysis. And so we often must rely on the expertise of the demographers, economists, sociologists, and others who compile and disseminate statistics; at the same time, we should maintain a healthy skepticism. For instance, we can accept that the professional statisticians at the Bureau of Justice Statistics know their business, but also be aware that, because of limited reporting from local law enforcement and other factors, their statistics are necessarily incomplete: they have only the data they have.²⁵ As we provide our patrons with statistics (or help them look themselves), we should be aware of how the statistics might be incomplete or somehow skewed. Are all law enforcement agencies participating in a data collection project? Do they follow the same definitions when they categorize their data? How big was the sample used in a survey? How was it generated? How many people responded?²⁶

¶11 We might also try to think about what was asked and why, understanding that the statistical reports that are presented as neutral are built on definitions and assumptions that are far from neutral. One vivid way to see this is by looking at the history of race in the census, from the constitutional reckoning that counted enslaved people (at the discounted rate of 3/5) in determining congressional representation²⁷ to today. Races have always been counted in the United States (but not everywhere²⁸), but they have been counted in different ways, and the counting has shaped policy and society.

Education and Current Awareness

¶12 If you don't feel prepared to be an intelligent consumer of statistics, you might read about the topic. Don't worry: it doesn't take higher math to improve your statistical literacy. For instance, generations of readers have benefited from Darrell Huff's very accessible book, *How to Lie with Statistics*.²⁹ More recently, Joel

25. Another significant limitation worth noting is the underreporting of many crimes, such as rape, child abuse, and domestic violence.

26. For the broader issue of interpreting popular reports of health studies, see Jeff Leek, *Finally, a Formula for Decoding Health News*, FIVETHIRTYEIGHT (Mar. 17, 2014), <http://fivethirtyeight.com/features/a-formula-for-decoding-health-news/> [https://perma.cc/QJ4E-ZE8N] (suggesting that, in addition to other factors, readers consider whether a study was a randomized, controlled trial and whether it had at least hundreds of patients).

27. U.S. CONST. art. I, § 3.

28. KENNETH PREWITT, WHAT IS YOUR RACE? THE CENSUS AND OUR FLAWED EFFORTS TO CLASSIFY AMERICANS 19–20 (2013):

Ethnoracial counting is widespread but not universal; it is a choice, and we want to understand what goes into the choosing. . . . *Race, ethnicity, nationality, ancestry, indigenous people, and tribal status* are among the various terms that appear in censuses that measure group membership based on descent. The actual term *race* itself is not common; it is found in only thirteen censuses around the world—even of them former slaveholding countries.

As well as being an academic, Prewitt served as director of the United States Bureau of the Census from October 1998 to January 2001. I highly recommend this book.

29. DARRELL HUFF, HOW TO LIE WITH STATISTICS (1954), and various reprint editions. See also DARRELL HUFF, HOW TO TAKE A CHANCE (1959) (probability). I read these when I was in high school—so I can vouch for their accessibility—and their basic lessons have stayed with me.

Best, a professor of sociology and criminal justice, has published several works for the general reader.³⁰ (I'm sure there are other books in this vein: these are just ones I can recommend from my own reading.)

¶13 Statistics can also become part of current awareness reading (or skimming). Last fall I subscribed to *The Economics Daily*³¹ from the Bureau of Labor Statistics and enjoy taking a quick look at it in my e-mail. The editors mix the topics, for instance, highlighting employment in breweries³² or pet-related industries³³ in addition to the Producer Price Index, the unemployment rate, and other standard measures you might expect. I was happy to pass along to our career services staff the prediction that legal employment will increase 5.1% by 2024.³⁴

¶14 Some other interesting statistical sources you might follow, with their Twitter handles, include:

- Brookings Institute, <http://www.brookings.edu/> (@Brookingsinst)
- Bureau of Justice Statistics, <http://www.bjs.gov/> (@BJSgov)
- FiveThirtyEight, <http://fivethirtyeight.com/> (@FiveThirtyEight)
- PewResearch Center FactTank, <http://www.pewresearch.org/fact-tank/> (@FactTank)
- ProPublica, <https://www.propublica.org/> (@ProPublic). See the Data page: <https://www.propublica.org/data/>
- TRACfed, <http://tracfed.syr.edu/> (@TRACReports)
- United States Census Bureau, <http://census.gov/> (@uscensusbureau)

Conclusion

¶15 Even if we aren't primarily "numbers people," we use and help others find statistics. Whether it's an infographic a friend posts on Facebook or a claim made by a presidential candidate, we should be able to evaluate it. Does the data come from a reliable source? Is it presented in its proper context? Are there potential gaps? Is it consistent with data from other sources? With a little self-education, we can enrich our own understanding and better serve our users.

30. JOEL BEST, DAMNED LIES AND STATISTICS: UNTANGLING NUMBERS FROM THE MEDIA, POLITICIANS, AND ACTIVISTS (2d ed. 2012); JOEL BEST, MORE DAMNED LIES AND STATISTICS: HOW NUMBERS CONFUSE PUBLIC ISSUES (2004); JOEL BEST, STAT-SPOTTING: A FIELD GUIDE TO IDENTIFYING DUBIOUS DATA (updated and expanded ed. 2013).

31. *TED: The Economics Daily*, U.S. BUREAU OF LABOR STATS., <http://www.bls.gov/opub/ted/> (last visited Mar. 29, 2016).

32. *Prost! Celebrating Oktoberfest with a Look at Brewery Employment*, *TED: THE ECONOMICS DAILY* (Sept. 30, 2015), <http://www.bls.gov/opub/ted/2015/celebrating-oktoberfest-with-a-look-at-brewery-employment.htm> [<https://perma.cc/VH92-KYWY>].

33. *Employment and Wages in Pet-Related Industries*, *TED: THE ECONOMICS DAILY* (Dec. 4, 2015), <http://www.bls.gov/opub/ted/2015/employment-and-wages-in-pet-related-industries.htm> [<https://perma.cc/ZMQ9-SZF3>].

34. *Employment Projections by Occupational Group 2014–24*, *TED: THE ECONOMICS DAILY* (Dec. 30, 2015), <http://www.bls.gov/opub/ted/2015/employment-projections-for-occupational-groups-2014-to-2024.htm> [<https://perma.cc/X5MS-QGUZ>].