

[BAD HESSian]

(<http://badhessian.org/>)

Big Data and the Survival of the Scientific Method

Posted on October 11, 2013 (<http://badhessian.org/2013/10/big-data-and-the-survival-of-the-scientific-method/>) by guest (<http://badhessian.org/author/guest/>) | 0 Comments (http://badhessian.org/2013/10/big-data-and-the-survival-of-the-scientific-method/#disqus_thread)

This is a guest post by Jen Schradie. Jen is a doctoral candidate in the Department of Sociology at the University of California-Berkeley and the Berkeley Center for New Media. She has a master's degree in sociology from UC Berkeley and a MPA from the Harvard Kennedy School. Using both statistical methods and qualitative fieldwork, her research is at the intersection of social media, social movements and social class. Her broad research agenda is to interrogate digital democracy claims in light of societal and structural differences. Before academia, she directed six documentary films on social movements confronting corporate power. You can find her at www.schradie.com (<http://www.schradie.com>) or @schradie (<https://twitter.com/schradie>) on Twitter.

Five years ago, Chris Anderson, editor-in-chief of *Wired Magazine*, wrote a provocative article entitled, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete" (http://www.wired.com/science/discoveries/magazine/16-07/pb_theory) (2008). He argued that hypothesis testing is no longer necessary with google's petabytes of data, which provides all of the answers to how society works. Correlation now "supercedes" causation:

This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.

An easy strawman, Anderson's piece generated a host of articles in academic journals decrying his claim. The overall consensus, to no surprise, was that the scientific method – i.e. hypothesis testing – is far from over. Most argued as Pigliucci (2009:534) articulated,

But, if we stop looking for models and hypotheses, are we still really doing science? Science, unlike advertising, is not about finding patterns—although that is certainly part of the process—it is about finding explanations for those patterns.

Other analysts focused on the debate around "correlation is not causation." Some critiqued Anderson in that correlation can lead you in the wrong direction with spurious noise. Others implicitly pointed to what Box (1976) articulated so well pre-Big Data – that science is an iterative process in which correlation is useful in that it can trigger research which uses

Follow Us!



([http://www.facebook.com](http://www.facebook.com/BadHessian)

/BadHessian)



([http://twitter.com](http://twitter.com/BadHessian)

/BadHessian)

Contributor

- Adam Slez (<http://badhessian.org/author/adam/>)
- Alex Hanna (<http://badhessian.org/author/admin/>)
- Benjamin Lind (<http://badhessian.org/author/benjamin/>)
- Dan Wang (<http://badhessian.org/author/dan/>)
- Matt Moehr (<http://badhessian.org/author/matt/>)
- Trey Causey (<http://badhessian.org/author/trey/>)
- Special Guests (<http://badhessian.org/author/guest/>)

Blogroll

- Andrew Gelman (<http://andrewgelman.com/>)
- Code and Culture (<http://codeandculture.wordpress.com/>)
- Crooked Timber (<http://crookedtimber.org/>)
- Dart-Throwing Chimp

<p>hypothesis testing.</p> <p>For sociologists, this issue may simply seem like fodder for an undergraduate lesson in the social scientific method. But Anderson's argument has not disappeared. Instead, it has grown exponentially, as has the term Big Data. In the process, a lot of confusion has arisen about what in the heck Big Data is and how it is used – are we talking a source of data, a data collection technique, an analytic tool, or perhaps data visualization? And how big is Big Data, anyway? Is it simply an accumulation of tweets on a hashtag or much larger? The answers to all of these questions is, well, kind of, sort of, yes.</p> <p>Digital technology has enabled researchers to access, store, analyze and report unprecedented massive amounts of data, often (though not always) from online sources. The trendy, and already clichéd term is not just data, but Big Data.</p> <p>Access to Big Data and information from the digital cloud, as well as research techniques using it, has become much easier. When Anderson wrote this article five years ago, conducting social network analysis required the use of clunky software and manually cleaning up databases. Twitter, for instance, was just emerging on the social media landscape. Mining Big Data from the digital cloud is increasingly prevalent in social science research, which can consist of writing computer code to scrape and download complex search results or using software to do so. Now, with a few mouse clicks, for instance, Twitter conversations are instantaneously displayed creating data for an article. Ok, it's not that simple, but it is so much more efficient than early forms of, say, UCINET (https://sites.google.com/site/ucinetsoftware/home), for instance. Research has never been so efficient.</p> <p>"Um, yeah," many of you may be thinking, "I know all of that." But I have found that many sociologists have a lot of questions about Big Data. Six months ago, a colleague contacted me to schedule a time to meet so that I could explain to her what Big Data was. Before we could connect, though, she had already started a job as a "data scientist." Now it's my turn to ask her what she's doing.</p> <p>But the question remains, has Big Data changed the scientific method? When I first read Anderson's article, I thought he was out to some digital utopian lunch. And I still do. In fact, on PBS' MediaShift, I blogged about consistent representativeness problems with using Big Data (http://www.pbs.org/mediashift/2013/07/big-data-not-big-enough-how-digital-divide-leaves-people-out/) scraped from the cloud because of digital inequality. And I am a proud card-carrying member of the scientific method society. However, is it possible that there could be something to correlation after all, that is more than a question of patterns (Gasp)? Big Data could actually be more of an anthropological source of data than we might realize, rather than the fetish that it is "right" and small data is "wrong."</p> <p>It might be useful to turn to the opposite end of the methodological spectrum: ethnography. Some ethnographers believe in grounded theory – in that we shouldn't go into a research site with preconceived theories or hypotheses to make sure we are open to any and all findings that emerge from the research site. Other ethnographers believe that researchers should start out with a theory and then constantly refine it with every observation or interview. Perhaps both of these approaches could be used with Big Data. With all of the methods we have of analyzing and visualizing Big Data so quickly, an anthropological approach could provide a rich and textured alternative path to along the data science (http://badhessian.org/2013/09/what-else-does-a-data-scientist-need/) road. As someone who uses multi-methods, I may be alienating both quantitative and qualitative researchers alike by being a participant observer of Big Data at times. However, I am not suggesting fishing expeditions in lieu of hypothesis testing nor any Anderson-esque junking of the scientific method. The numbers do not speak for themselves. Instead, it is our job as social scientists to understand the difference between the data, whatever its size, and the method, whatever that may be.</p> <p>References</p> <p>Anderson, Chris. 2008. "the end of theory: the data deluge makes the scientific method obsolete." <i>Wired Magazine</i>.</p>	<div data-bbox="1082 38 1348 1601"><p>(http://dartthrowingchimp.wordpress.com/)</p><ul style="list-style-type: none">• Exploring Possibility Space (http://exploringpossibilitiespace.blogspot.com/)• Gilad Lotan (http://giladlotan.com/blog/)• Jay Yonamine (http://jayyonamine.com/)• John Myles White (http://www.johnmyleswhite.com/)• Mobilizing Ideas (http://mobilizingideas.wordpress.com/)• OrgTheory (http://orgtheory.wordpress.com/)• Permutations (http://permut.wordpress.com/)• Phil Schrodtt (asecondmouse) (http://asecondmouse.wordpress.com/)• R-Bloggers (http://r-bloggers.com/)• Scatterplot (http://scatter.wordpress.com/)• Sean J. Taylor (http://seanjtaylor.com/)• Sociological Images (http://thesocietypages.org/g/socimages/)• Technosociology (http://technosociology.org/)• The Monkey Cage (http://www.washingtonpost.com/blogs/monkey-cage/)• Tom Slee (http://tomslee.net/)• Work in Progress (http://workinprogress.oowsection.org/)</div>	<div data-bbox="1042 1733 1308 1789"><h2>Categories</h2></div> <ul style="list-style-type: none">• Clusters (http://badhessian.org/category/clusters/)• Conferences (http://badhessian.org/category/conferences/)• Datasets (http://badhessian.org/category/datasets/)• Event data (http://badhessian.org/category/event-data/)
--	---	---

Box, G. E. P. 1976. "Science and Statistics." *In Journal of the American Statistical Association* (71).
Pigliucci, Massimo. 2009. "The End of Theory in Science." *EMBO reports* (10):534.

Posted in General (<http://badhessian.org/category/general/>).
Tagged big data (<http://badhessian.org/tag/big-data/>), method (<http://badhessian.org/tag/method/>).

0 Comments

 Login

Sort by Best

Share  Favorite 

Start the discussion...

ALSO ON BAD HESSIAN

WHAT'S THIS?

THE FINAL FOUR – Drag Race season 5, episode 11

2 comments • 11 months ago

Avatar **Alex Hanna** — Thanks hunty. It'll be a fun one.

Some comments on "Computer-aided content"

5 comments • a month ago

Avatar **Alex Hanna** — No but I've been using the Stanford Classifier for my dissertation.

Hacking GDELT

4 comments • 6 months ago

Avatar **Trent** — This is exciting news, Alex. Looking forward to San Francisco!

Converting approximating numerical words to

8 comments • 9 months ago

Avatar **Neal Caren** — Link Grammar is written in C with Python bindings available, so it triples

Next Post → (<http://badhessian.org/2013/11/python-and-pydata-conferences-are-important-for-the-future-of-social-science-research/>)
← Previous Post (<http://badhessian.org/2013/09/hacking-gdel/>)

- Event history analysis (<http://badhessian.org/category/event-history-analysis/>)
- Forecasting (<http://badhessian.org/category/forecasting-2/>)
- General (<http://badhessian.org/category/general/>)
- GLM (<http://badhessian.org/category/glm/>)
- Hardware (<http://badhessian.org/category/hardware/>)
- Java (<http://badhessian.org/category/java/>)
- Job Announcements (<http://badhessian.org/category/job-announcements/>)
- Logistic Regression (<http://badhessian.org/category/logistic-regression/>)
- Networks (<http://badhessian.org/category/networks/>)
- Perl (<http://badhessian.org/category/perl/>)
- Productivity (<http://badhessian.org/category/productivity/>)
- Python (<http://badhessian.org/category/python/>)
- Q&A (<http://badhessian.org/category/qa/>)
- R (<http://badhessian.org/category/r/>)
- Regression (<http://badhessian.org/category/regression/>)
- Simulation (<http://badhessian.org/category/simulation/>)
- Social (<http://badhessian.org/category/social/>)
- Social Movements (<http://badhessian.org/category/social-movements/>)
- Spatial Statistics (<http://badhessian.org/category/spatial-statistics/>)
- Statistics (<http://badhessian.org/category/statistics/>)
- Text analysis (<http://badhessian.org/category/text-analysis/>)
- Twitter

(<http://badhessian.org/category/twitter/>)

- Version control

(<http://badhessian.org/category/version-control/>)

Disclaimer

No university computers were used, let alone harmed, in the making of this site.

Meta

- Log in
(<http://badhessian.org/wp-login.php>)
- Entries RSS (Really Simple Syndication)
(<http://badhessian.org/feed/>)
- Comments RSS (Really Simple Syndication)
(<http://badhessian.disqus.com/latest.rss>)
- WordPress.org
(<http://wordpress.org/>)

© 2014 Bad Hessian (<http://badhessian.org/>), all rights reserved.

Proudly powered by WordPress (<http://wordpress.org/>)