Interviewing Data - The art of interpretation in analytics

Anne L. Washington
George Mason University

Abstract
Algorithms and statistical models produce consistent results with confidence, yet they do so with data that are subject to change. Furthermore, the underlying digital traces created within specifically designed platforms are rarely transparent. The emerging field that incorporates analytics, predictive behavior, big data, and data science is still contesting its methodological boundaries. How can we use existing research tools to validate the reliability of data? This paper explores alternatives to statistical validity by situating analytics as a form of naturalistic inquiry. A naturalistic research model, which has no assumption of an objective truth, places greater emphasis on logical reasoning and researcher reflectivity. "Interviewing data," based on journalistic practices, is introduced as a tool to convey the reliability of data. The misleading 2013 flu prediction illustrates this approach and is discussed within the context of ethics and accountability in data science.
Keywords: data analytics; naturalistic inquiry; big data; ethics; flu prediction; validity
doi: 10.9776/16256

Contact: washingtona@acm.org

1 Introduction

Scholars continue to contest the methods used in the emerging field that encompasses analytics, big data, business intelligence, computational prediction, and data science. The standard methods for analytics projects that use large data sets are statistical modeling and algorithms. Quantitative methods have been reliably consistent for structured quantitative data, yet the many unstructured data sets that feed analytics are different. Their underlying observations are subject to change in ways the researcher cannot determine. The need to interpret and contextualize data is an essential but rarely discussed aspect of analytics. Google Flu Trends (Butler, 2013; Ginsberg, 2009) illustrates the role of interpretation in data analytics. The increasingly skewed flu predictions were due, in part, to the researchers' failure to interpret the reliability of their data sources.

Interpretation is central to naturalistic forms of inquiry. Naturalistic inquiry (Lincoln & Guba, 1985) stems from ethnography, field work or observations of uncontrolled environments where an objective truth (Maxwell, 2005) cannot be assumed. Rubin and Rubin (2005), in "Qualitative Interviewing: The Art of Hearing Data," describe naturalistic inquiry as the generation of data through encounters between participant and researcher. Considering analytics from a naturalistic perspective is in direct opposition to its more common association with quantitative scientific methods (Dhar, 2013; Goes, 2014; Lin, 2015). The most significant difference is that the naturalistic researcher generates data and serves as the sole research instrument. However, both analytics and naturalistic researchers select from a range of found data they happen to encounter. Additionally, the reduction of large collections of words is difficult to convey to critics, whether it is from thick description or big data.

To emphasize both the interpretive responsibility of the analyst and the contextual aspects of the data, "interviewing data" is introduced as an assessment tool. It is the responsibility of the person using the data to identify inherent preferences and biases (Willis, 2014). Interviewing data is described as a series of journalistic questions that can be applied to understanding data. This approach is suitable for researchers with intimate access to the data set who need to convey its reliability.

This essay continues a strain of research that calls for increased attention to agency in big data analytics (Fricke, 2014; Gillespie, 2012; Lagoze, 2014). Like naturalistic researchers, data analysts play a role in making decisions that impact the outcome, but often without accountability. Intellectual debate and scientific skepticism are hindered if those without access to the data have no way to evaluate the results. Interviewing data is one step towards challenging the assumption that all found data can be used without interpretation.

2 Related Work

Naturalistic inquiry, as described by Lincoln and Guba (1985), takes place in environments that the researcher does not control. The observations for naturalistic inquiry emerge from interactions between the participants and the researcher. The resulting documentation of these observations generates qualitative data, such as images and words (DeWalt & DeWalt, 2002). The researcher is the essential instrument for gathering observations and is vital to the claims made from the observations (Bisel, 2014; LeCompte, 1999). This is in contrast to predetermined propositions about controlled observations whose results are quantified as numbers and analyzed with statistics.

2.1 Validity

Validity verifies research projects and confirms that the results can be applied within the given context. More importantly, validity encourages trust in the results by indicating how the research might be wrong. Validity assessments provide enough information to determine whether the data and methods support the assertions in the results. Validity is considered here as the verification of the results through evidence and reasoning. These strategies focus on methodological validity, not the validity of the analysis (Maxwell, 2005; Silverman, 2000). Positivist research relies on strict research design methods to achieve statistical validity (Lee, 1989). Validity in naturalistic inquiry relies on researchers' abilities to be reflective and use themselves as instruments of observation (Maxwell, 2005).

Validity for statistical and experimental research is described in terms of reliability, including objectivity as well as threats to internal and external validity (Campbell & Stanley, 1966). Threats to internal validity include inconsistent procedures and methods. Threats to external validity include inaccurate assumptions or inferences. Reliability is confirmation that data are observed under uniform conditions. Objectivity is confirmation that the researcher is neutral, often by using instruments that remove human perception. Interpretive naturalistic research cannot use identical strategies, but it can use similar ones.

Lincoln and Guba (1985) offer a set of four alternative validity checks for naturalistic inquiry that match internal threats, external threats, reliability, and objectivity. Naturalistic research is valid when it is credible, dependable research that can be confirmed in the current context and extended to others. First, the credibility of the researcher and her epistemology establishes internal validity. Second, the ability to transfer the findings to another situation establishes external validity and generalizable results. Third, reliability verifies that future researchers can repeat the work dependably by following the same inquiry, strategy, and methods. Fourth, objectivity confirms a lack of systematic bias in both source and analysis. The four checks for naturalistic validity are credibility, transferability, dependability, and confirmability. A naturalistic approach emphasizes the interpretive role of the researcher.

Statistical Validity	Naturalistic Validity	Consideration
Internal threats	Credibility	Is the research consistent?
External threats	Transferability	Are the results applicable elsewhere?
Reliability	Dependability	Can the research be repeated?
Objectivity	Confirmability	Is there systematic bias?

Table 1. A comparison of validity assessments based on Lincoln and Guba (1985, p. 300).

2.2 Analytics as Naturalistic Inquiry

Analytics researchers face questions of validity similar to those naturalistic researchers confront. Streaming sensor information and unstructured text from multiple sources often contain personally identifiable information and would not be shared openly because of privacy concerns. Moreover, these sources are often owned by commercial companies that are unlikely to open proprietary data. Algorithms with little transparency reduce unstructured text into baskets, categories, and associations (Provost, 2013). Big data researchers have the added challenge of never stepping into the same river twice. A Deluzian mapping suggests that multiple states transform socially constructed data over time with multiple individual journeys (Mazzei, 2010; Ruppert, 2009). While algorithms may be tailored for specific platforms, the people using these platforms may begin to flow data in new ways at any moment.

There is still a lively debate over whether applying the scientific method to digital traces can define social worlds, much less society (Couldry & Powell, 2014; Ruppert, 2011). Ethicists question decisions hidden in algorithms and other software that contain formalized procedures (Diakopolous, 2014), but few have considered the important role of validating data. A first step towards creating trustworthy data is to insert the data analyst into the conversation.

The reduction of both thick description (Miles & Huberman, 1994) and big data (Bisel, 2014) is difficult to convey to those without intimate knowledge of the material. Both types of researchers attempt to select sources that best represent traces of the communities they are observing. By focusing on the uncontrolled nature of the data sources, analytics can be viewed as a form of naturalistic inquiry. This opens up analytics to alternative validity checks.

3 Interviewing Data

Interviewing data, according to one *New York Times* data journalist, is like interviewing people, but without any social context. Derek Willis (2011) argues that all data are flawed data in some way, and journalists should approach their digital sources the same way they would approach an interview. Interviews between researchers and participants are central to naturalistic modes of inquiry. Interviewers choose interviewees based on identified characteristics that suggest that they would have a good story to share.

Rubin and Rubin (2005) argue that the interview anticipates the form of the research questions, not the content of the answers. For instance, a psychology research question might investigate how people react, while a policy research question might investigate how institutions respond. Like an interviewer, data analysts could situate their sources before deciding whether they are appropriate to address their research questions. This perspective emphasizes the responsibility of the data analyst to establish the validity of sources to begin a logical chain from data to results.

Question	Asks
When	What historic or periodic context is represented?
Why	Why was the data created? What was the original purpose?
How	How were data produced and with what platform constraints?
Who	What populations are represented?

Table 2. A data interview is a series of journalistic questions used to interpret data sources.

Interviewing data is like interviewing any other source (Myers, 2007; Rubin, 2005). The rubric of basic journalistic questions – when, why, how, and who – can be used to interview data. These questions serve as the initial step in building a case to support the reasoning (Mason, 2002) of the results.

"When" questions consider the historic context of data creation. Questioning when something occurs captures an understanding of trends across time. Borgman (2015) asks "when are data" to discuss data needs in the humanities. Early modern scholars might need to understand the difference between a 1780 and a 1980 interpretation of Shakespeare. A contemporary scholar might be more likely to include feminist criticism than an earlier one. Analytics researchers might consider current technology trends, such as mobile phone use, that impact their data.

"Why" questions address the original purpose for data. Questioning why digital traces are created addresses their original purpose and intent. It also provides the appropriate scope to the data and therefore to the inferences. It is particularly vital when joining data from multiple sources. Combining census data created for reporting on populations with other sources may lead to unexpected results (Krogstad, 2014; McMillen, 2004). Even when digital traces match, there may be ethical concerns about reuse. The DNA from a routine medical exam at a public university identified the patient's father in a crime investigation (Nakashima, 2008). Understanding why the data were created in particular ways uncovers the reasoning and assumptions underneath the traces.

"How" questions ask how the data are constructed and under what constraints. Questioning how uncovers the processes, procedures, and practices that build the data product. The comforting precision of a number might mask its provisional or conditional meaning (Lampland, 2010). Bowker and Star (1999) document how nurses intentionally provide sparse data about patients in order to manage the level of attention given by the hierarchy in a hospital. Analytics researchers might consider platform constraints, such as knowing that submissions are limited to 140 characters. In addition, how questions consider algorithmic transformations of the data as it is aggregated and categorized across multiple actors.

"Who" questions ask about the people or populations involved. These questions consider how the people who needed, created, or requested the data might have impacted what is available. Those interpreting the data must consider whether the population is using words literally or to convey other meanings. For example, young people use words with encoded meanings to deflect adult attention and still share with their friends on social media platforms (boyd, 2011). Despite Wikipedia policy, many editors do not disclose that they are freelance writers paid to edit pages (Pinsker, 2015). Analytics researchers need to address whether or how they have accounted for bots, paid participants, or other issues about how accurately the population is represented.

4 Google Flu Trends

Predicting flu outbreaks was a model for the potential of big data analytics. Organizations responsible for providing vaccines anticipate public health outbreaks. This is a complicated calculation that requires both anticipating possible problems while not scaring the public. The Centers for Disease Control and Prevention (CDC), a US federal agency, has tracked public health trends using reports from labs, physicians, and hospital records as the basis of decisions (Helft, 2008). In 1976, a flu epidemic was declared in the United States but did not occur, generating concern about the process of making these decisions (Agryis, 1979; Neustadt & Fineberg, 1983).

In 2008, the CDC began to work with computer scientists from a popular search engine company to identify search trends. Search trends analyze query logs to identify time-series patterns. Searches about cold remedies, symptoms, and other topics were arranged by geographic area and modeled as trends. The Internet search engine Google began to publish experiments about predicting behavior from search queries (Matias, Efron, & Shimshoni, 2009; Varian & Choi, 2009). When a new form of influenza spread in 2009, the search trend algorithm was able to match and anticipate CDC trends (Ginsberg et al., 2009).

Google Flu Trends was heralded as preventing a rash public health response to a “swine flu epidemic." The flu trends model was able to predict the outbreak of influenza up to two weeks before the traditional public health data were released (Christakis & Fowler, 2010). In addition to mirroring results in the United States, the model used World Health Organization data to monitor the spread of flu in countries worldwide. The model could identify the outbreak of flu at a regional, city, or other smaller geopolitical level. From 2008-2012, it represented the ideal public service that big data could offer. By 2013, though, Google Flu Trends began to lose its predictive reliability. It over predicted flu outbreaks. Aside from potential problems with the algorithms (Butler, 2013; Lazar, 2014; Madrigal, 2014), the changing nature of the data might have impacted the results.

What if the data were interviewed like a human source to see if they represented what the analysts intended? A data interview incorporates typical journalistic questions that ask when, why, who, and how.

- WHEN - Lazar (2014) suggests that an early version detected flu symptoms as well as determining whether it was winter. Google never releases queries, so it is not possible to confirm the exact words. This perspective caused them to miss a non-seasonal flu outbreak in 2009. Google Flu Trends realized that seasonality was critical and began to consider how to incorporate it into later calculations.
- WHY - Search queries show an interest in a topic, but searchers have many reasons why they might initiate a search. The assumption underlying the project was that flu-related queries indicated that the searcher was sick. Butler (2013) suggested that queries were in response to increased media reports about the spread of the flu. The revised assumption is that people completed searches because they heard news reports but did not personally have or know someone with flu symptoms. Considering a range of plausible reasons about why data are created could lead to broader perspectives about the quality of the results.
- WHO - Asking why the data were created leads to whom the data were created for. The searcher may be someone consuming media instead of a sick person or someone who cares for a sick person. In this example it was not reasonable to equate searching to representing. In contrast, the CDC data is based on doctors who describe patients and therefore is more strictly limited to those who may be sick.
- HOW - The algorithms processing the queries were constantly changing (Butler, 2013; Lazar, 2014). Although Google Flu Trends might have access to the changing dynamics of Google searches, they did not choose to use this information in the project. In addition, Google products evolved over this period. Mail, news, and search became more integrated (Treese, 2009). The platform might have encouraged behavior that introduced a systematic bias.

Interviewing data is suggested as a method to share the reasoning for data selection. It is a way to show due diligence without revealing details when it is not possible to share the original data. Furthermore, this method provides outsiders a way to assess the methods and interrogate the results. Data interviews could be held before an algorithm is run to identify whether the results are likely to be reliable. For projects like Google Flu Trends, which ran for many years, data interviews could also document the changing context of their source observations. Conversely, interviewing data after running the analysis may be useful as a confirmatory tool or to investigate failures.

5 Discussion

Analytics scholars need more tools to justify their assertions. Like naturalistic inquiry, their assertions are grounded in data (Mason, 2002), but analytics lacks the conceptual and practical tools to demonstrate methodological processes. What is missing in many studies is logical reasoning about data sources. Big data can be too large to reasonably publish. Data used in social media research may not be open due to commercial and privacy restrictions. Instead of presenting data sources as a black box, interviewing data is suggested as a path towards more research transparency. Without demonstrating the reliability of the evidence, analytics results are easy to question.

Asking questions is one path towards establishing reliability in naturalistic research (Goia, 2013). Van Maanen (1979) argues that naturalistic research requires both a first order insider perspective and a second order outsider perspective. The researcher demonstrates a lack of systematic bias by acknowledging the difference between the researcher’s and the participant's perspectives. When analytics researchers do not offer any reflective consideration of these critical internal decisions, at best they confuse the first order and second order perspective. More troubling is that they might appropriate the participant's voice as their own.

Analytics researchers exacerbate concerns over ethics by relying solely on statistical validity. This leaves little room for informed debate. Some commercial companies generated controversy when they released their analytics experiments to the public. The Facebook emotional contagion study and the misleading matches on OK Cupid (Meyer, 2015) both presented evidence of theoretical and statistical validity to justify their findings. The size and scale of the research and specific statistical tests were not enough to substantiate the experimental design nor the results to the public. Considering these projects as a naturalistic form of inquiry might have encouraged conversations and reflections within the organizations. Subsequently, they might have expressed more empathy towards the participants who unwittingly took part in these experiments. While many have expressed concern over the power of algorithms in analytics, there is equal need to investigate the reliability and validation of data sources. More information about decisions that inform analytics projects could lead to more fruitful conversations about methodological choices.

To move beyond assumptions that statistics and scale are sufficient justification for results, this essay situates analytics as a form of naturalistic inquiry that requires interpretation. Information schools have an opportunity to meet the demand for a more reflective approach to analytics. Although large-scale data sets are relatively new, library and information science has been considering how to evaluate sources for centuries. Many data science programs are housed in business administration or computer science departments. These departments tend to extend existing curriculum on logic and statistics under the name of data science (Howe, 2014). However, few have required classes on ethics, data engineering, or data management. Information schools are well positioned to explore a more nuanced approach to the art of interpreting data.

6 References

Argyris, C. (1979). Using Qualitative Data to Test Theories The Swine Flu Affair: Decision Making on a Slippery Disease. Administrative Science Quarterly, 24(4), 672–679.

Bisel, R. S., Barge, J. K., Dougherty, D. S., Lucas, K., & Tracy, S. J. (2014). A Round-Table Discussion of “Big” Data in Qualitative Organizational Communication Research. Management Communication Quarterly, 28(4), 625–649. http://doi.org/10.1177/0893318914549952

Borgman, C. L. (2015). Big data, little data, no data: scholarship in the networked world. MIT Press.

Bowker, G. & Star, S. L. (1999). Sorting things out: Classification and its consequences. Cambridge: MIT Press.

boyd, danah. (2011). Dear Voyeur, meet Flâneur... Sincerely, Social Media. Surveillance & Society, 8(4), 505–507.

Butler, Declan (13 February 2013). "When Google got flu wrong". Nature 494: 155–156. doi:10.1038/494155a

Campbell, D. T., & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research (2nd ed.). Chicago: Rand McNally. ISBN: 9780395307878 http://lccn.loc.gov/66025213

Christakis, N. A., & Fowler, J. H. (2010). Social network sensors for early detection of contagious outbreaks. PloS one, 5(9), e12948.

Couldry, N., & Powell, A. (2014). Big Data from the bottom up. Big Data & Society, 1(2), 2053951714539277. http://doi.org/10.1177/2053951714539277

Dhar, V. (2013). Data Science and Prediction. Communications of the ACM, 56(12), 64–73.

Diakopoulos, N. (2014). Algorithmic Accountability. Digital Journalism, 0(0), 1–18. doi:10.1080/21670811.2014.976411
Frické, M. (2014). Big data and its epistemology. Journal of the Association for Information Science and Technology. http://doi.org/10.1002/asi.23212
Gillespie, T. (2012). Can an Algorithm be Wrong? Limn, 1(2). http://doi.org/http://escholarship.org/uc/item/0jk9k4hj
Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. Nature, 457(1012-1014).
Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking Qualitative Rigor in Inductive Research Notes on the Gioia Methodology. Organizational Research Methods, 16(1), 15–31.
Goes, P. (2014). Editor's Comments: Big Data and IS Research. MIS Quarterly, 38(3), iii–viii.
Helft, M. (2008 November 12). Google uses searches to track flu's spread. The New York Times. Sect A 1.
Howe, Franklin, Freire, Frew, Kraska, and Ramakrishnan. 2014. Should we all be teaching "intro to data science" instead of "intro to databases"?. In Proceedings of the 2014 ACM SIGMOD international conference on Management of data (SIGMOD '14). 917-918. http://doi.acm.org/10.1145/2588555.2600092
Krogstad, J. M. (2014, March 24). Census Bureau explores new Middle East/North Africa ethnic category. Retrieved from http://www.pewresearch.org/fact-tank/2014/03/24/census-bureau-explores-new-middle-eastnorth-africa-ethnic-category/
Lagoze, C. (2014). Big Data, data integrity, and the fracturing of the control zone. Big Data & Society, 1(2), 2053951714558281. http://doi.org/10.1177/2053951714558281
Lampland, M. (2010). False Numbers as Formalizing Practices. Social Studies of Science, 40(3), 377–404.
Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. Science, 343(6176), 1203–1205. http://doi.org/10.1126/science.1248506
LeCompte, M. D., & Schensul, J. J. (1999). Analyzing & interpreting ethnographic data. Walnut Creek, Calif.: AltaMira Press.
Lee, Allen S. (1989) A Scientific Methodology for Mis Case Studies . MIS Quarterly 1989/03, Vol.13 (1)
Lin, J. (2015). On Building Better Mousetraps and Understanding the Human Condition Reflections on Big Data in the Social Sciences. The ANNALS of the American Academy of Political and Social Science, 659(1), 33–47. http://doi.org/10.1177/0002716215569174
Madrigal, Alexis C. (2014 March 27). In Defense of Google Flu Trends. The Atlantic . Retrieved from: http://www.theatlantic.com/technology/archive/2014/03/in-defense-of-google-flu-trends/359688/
Mason, J. (2002). Making Convincing Arguments with Qualitative Data. In Qualitative researching (Vol. 2nd, p. viii, 223 p.). London; Thousand Oaks, Calif.: Sage Publications.
Matias, Y., Efron, N., & Shimshoni, Y. (2009). On the predictability of Search Trends: Google Research Blog. Retrieved from http://googleresearch.blogspot.com/2009/04/predicting-present-with-google-trends.html
Maxwell, J. A. (2005). Qualitative research design : an interactive approach (Vol. 2nd). Thousand Oaks, CA: Sage Publications.
Mazzei, L. A., & McCoy, K. (2010). Thinking with Deleuze in qualitative research. International Journal of Qualitative Studies in Education, 23(5), 503–509. http://doi.org/10.1080/09518398.2010.500634
McMillen, D. (2004). Privacy, confidentiality, and data sharing: Issues and distinctions [discussion]. Government Information Quarterly, 21(3), 359-382.
Meyer, M. (2015) Two Cheers for Corporate Experimentation: The A/B Illusion and the Virtues of Data-Driven Innovation. 13 Colorado Tech. L.J. 273. http://ssrn.com/abstract=2605132
Miles, M. B. (1979). Qualitative Data as an Attractive Nuisance: The Problem of Analysis. Administrative Science Quarterly, 24(4), 590–601.
Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. Information and Organization, 17(1), 2-26.
Nakashima, E. (2008, April 21). From DNA of Family, a Tool to Make Arrests Privacy Advocates Say the Emerging Practice Turns Relatives Into Genetic Informants. Washington Post.
Neustadt & Fineberg. (1983) The Epidemic that Never Was: The Swine Flu Affair, Vintage Books
Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. Big Data. 1 (1), 51-59.
Rubin, H. J., & Rubin, I. (2005). Qualitative interviewing: the art of hearing data (Vol. 2nd). Thousand Oaks, Calif.: Sage Publications.

Ruppert, E. (2009). Becoming Peoples. Journal of Cultural Economy, 2(1-2), 11–31. http://doi.org/10.1080/17530350903063909

Ruppert, E. (2011). Population Objects: Interpassive Subjects. Sociology, 45(2), 218–233. http://doi.org/10.1177/0038038510394027

Silverman, D. (2000). Doing qualitative research : a practical handbook. London ; Thousand Oaks, Calif.: Sage.

Star, S. L., & Bowker, G. C. (2007). Enacting silence: Residual categories as a challenge for ethics, information systems, and communication. Ethics and Information Technology, 9(4), 273–280.

Treese, W. (2009). Is Google too big to fail? netWorker, 13(2), 11. http://doi.org/10.1145/1540392.1540396

Varian, H. R., & Choi, H. (2009). Predicting the Present with Google Trends. Retrieved from http://googleresearch.blogspot.com/2009/04/predicting-present-with-google-trends.html

Willis, D. (2011) Interviewing Data. The Scoop. Retrieved from http://thescoop.org/archives/2011/05/01/interviewing-data/

Willis, D. (2014) Dealing With Data: Be Very, Very Skeptical. Journalism Interactive 2014 conference. Retrieved from http://ajr.org/2014/04/05/dealing-data-skeptical/