Title
Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are. A Book Review.

Permalink
https://escholarship.org/uc/item/8jp5b1kt

Journal
InterActions: UCLA Journal of Education and Information Studies, 14(2)

Author
Cho, Richard Minkyu

Publication Date
2018

DOI
10.5070/D4142037710

Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are. By Seth Stephens-Davidowitz. New York: Dey St., 2017. Pp. v + 338. $27.99 (hardcover).

Everyone lies now and then. We lie to others, on surveys, and, tragically, we tell the biggest lies to ourselves. But to what extent do we lie?

Seth Stephens-Davidowitz, the author of *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*, argues that we now have the means to answer this question. He calls the Google search box a "digital truth serum," a confessional in which secret yearnings and personal sins are revealed. Conversational in tone yet articulately written, this book is a first-class primer on Big Data. Stephens-Davidowitz mines his ten years of experience as an academic data researcher and data scientist at Google to reveal what the Google search box can do, what it cannot do, and what it has revealed so far about human nature, all with sparkling wit and entertaining anecdotes. Although he is a steadfast supporter of the positive potential applications of Big Data, saying that "this book is going to show how Big Data is best used and explain in detail why it can be so powerful," Stephens-Davidowitz's humility and healthy skepticism toward even his own data and discipline is his book's auxiliary charm.

In his introduction to the book, renowned psychologist and author Steven Pinker writes that Stephens-Davidowitz introduces "a whole new way of studying the mind," and that his studies grant "an unprecedented peek into people's psyches." Stephens-Davidowitz's main tool for this task is Google Trends, a public web facility that informs users how frequently any word or phrase has been searched in different locations at different times. His main strength lies in his creative application of this new method. He looks into fields "where old methods (of research) are lousy," and with data in galore in his grasp—the kind of data he claims that Foucault and Freud would drool over—he probes a number of social and psychological questions such as:

- How many men are gay?
- Is the media biased?
- Are Freudian slips real?
- Who cheats on their taxes?
- Does it matter where you go to college?
- What should you talk about on a first date if you want a second?

His findings overturn many of our intuitive conclusions of the past. For example, he disproves the accusation that social media is in fact merely an echo chamber as

it's often portrayed, and offers substantial, convincing evidence that newspapers are less political than popularly assumed.

As expected, the author talks at length about Big Data, but always with relevant (and entertaining) examples. He focuses on four reasons for the revolutionary power of Big Data: the information it provides is honest, qualitatively new, allows us to zoom in on small subsets of people, and facilitates many causal experiments. In Part II, Stephens-Davidowitz recounts his studies and conversations with people who use Big Data in their fields, touching upon a broad range of topics including sex, hate and prejudice, internet, child abuse, abortion, Facebook friends, and much more. With such an enormous quantity of data to draw from, he can zoom in and study the demographics and social dynamics not just of nations and cities, but individual counties and even neighborhoods. For example, he analyzes the cities, counties, and neighborhoods where a notably high number of successful people grew up and concludes that encouraging immigration, subsidizing universities, and supporting the arts are the three main factors that produce competent adults. Another example studies the so-called "Doppelganger searches" made possible by the emergence of Big Data. These searches have been successfully used to predict how well baseball players perform, as well as to help doctors treat patients with similar symptoms and conditions effectively.

Part III is titled *Big Data: Handle with Care*. As mentioned previously, the author emphasizes what Big Data cannot do as much as what it can. He introduces the phrase "the curse of dimensionality," explaining that newer datasets frequently give us exponentially more variables than traditional sources. To find what the numbers miss is just as important as determining what they reveal, and he cautions future data scientists not to fall in love with their results. He ends this section by discussing the ethical implications of using Big Data, such as when casinos gather data to assign "pain points" to each gambler in order to maximize the casinos' profit.

Everybody Lies is timely and includes examples drawn from the 2016 United States presidential election. Geared toward those curious about the idiosyncrasies of human nature, this book will enlighten and entertain novices of data science as well as academics, librarians, and executives who would like to incorporate data science into their work. To its credit, the book references a number of other popular books in the field, such as books by Malcolm Gladwell, Nate Silver, Peter Thiel, Scott Ganu, and retrospectively analyses the conclusions derived by these authors. The author predicts that Big Data will encroach into every aspect of our society. But he ends the book on a positive note by asserting that Big Data, "is poised to improve our lives."