YOUR KINDLE NOTES FOR:

Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are

by Seth Stephens-Davidowitz

16 Highlights

Highlight (Yellow) | Location 74

And though I like to think that nothing can shock me, I was shocked aplenty by what the internet reveals about human sexuality—including the discovery that every month a certain number of women search for "humping stuffed animals."

Highlight (Yellow) | Page 5

The power in Google data is that people tell the giant search engine things they might not tell anyone else.

Highlight (Yellow) | Page 12

Search data revealed that we lived in a very different society from the one academics and journalists, relying on polls, thought that we lived in. It revealed a nasty, scary, and widespread rage that was waiting for a candidate to give voice to it.

Highlight (Yellow) | Page 13

Silver found that the single factor that best correlated with Donald Trump's support in the Republican primaries was that measure I had discovered four years earlier. Areas that supported Trump in the largest numbers were those that made the most Google searches for "nigger."

Highlight (Yellow) | Page 14

I am now convinced that Google searches are the most important dataset ever collected on the human psyche.

Highlight (Yellow) | Page 54

Big Data allows us to finally see what people really want and really do, not what they say they want and say they do. Providing honest data is the second power of Big Data.

Highlight (Yellow) | Page 54

Allowing us to zoom in on small subsets of people is the third power of Big Data.

Highlight (Yellow) | Page 54

Allowing us to do many causal experiments is the fourth power of Big Data.

Highlight (Yellow) | Page 58

Can we tell, simply by what people are Googling, how many people are unemployed, and can we do so well before the government collates its survey results? One day, I put the United States unemployment rate from 2004 through 2011 into Google Correlate. Of the trillions of Google searches during that time, what do you think turned out to be most tightly connected to unemployment? You might imagine "unemployment office"—or something similar. That was high but not at the very top. "New jobs"? Also high but also not at the very top. The highest during the period I searched—and these terms do shift—was "Slutload." That's right, the most frequent search was for a pornographic site. This may seem strange at first blush, but unemployed people presumably have a lot of time on their hands. Many are stuck at home, alone and bored. Another of the highly correlated searches—this one in the PG realm—is "Spider Solitaire." Again, not surprising for a group of people who presumably have a lot of time on their hands.

Highlight (Yellow) | Page 71

First, and perhaps most important, if you are going to try to use new data to revolutionize a field, it is best to go into a field where old methods are lousy. The pedigree-obsessed horse agents whom Seder beat left plenty of room for improvement. So did the word-count-obsessed search engines that Google beat. One weakness of Google's attempt to predict influenza using search data is that you can already predict influenza very well just using last week's data and a simple seasonal adjustment. There is still debate about how much search data adds to that simple, powerful model. In my opinion, Google searches have more promise measuring health conditions for which existing data is weaker and therefore something like Google STD may prove more valuable in the long haul than Google Flu.

Highlight (Yellow) | Page 78

For instance, Google Ngrams can teach us how national identity formed. One fascinating example is presented in Aiden and Michel's book, Uncharted. First, a quick question. Do you think the United States is currently a united or a divided country? If you are like most people, you would say the United States is divided these days due to the high level of political polarization. You might even say the country is about as divided as it has ever been. America, after all, is now color-coded: red states are Republican; blue states are Democratic. But, in Uncharted, Aiden and Michel note one fascinating data point that reveals just how much more divided the United States once was. The data point is the language people use to talk about the country. Note the words I used in the previous paragraph when I discussed how divided the country is. I wrote, "The United States is divided." I referred to the United States as a singular noun. This is natural; it is proper grammar and standard usage. I am sure you didn't even notice. However, Americans didn't always speak this way. In the early days of the country, Americans referred to the United States using the plural form. For example, John Adams, in his 1799 State of the Union address, referred to "the United States in their treaties with his Britanic Majesty." If my book were written in 1800, I would have said, "The United States are divided." This little usage difference has long been a fascination for historians, since it suggests there was a point when America stopped thinking of itself as a collection of states and started thinking of itself as one nation.

Highlight (Yellow) | Page 105

People lie to friends. They lie to bosses. They lie to kids. They lie to parents. They lie to doctors. They lie to husbands. They lie to wives. They lie to themselves. And they damn sure lie to surveys. Here's my brief survey for you: Have you ever cheated on an exam? _____ Have you ever fantasized about killing someone? _____ Were you tempted to lie? Many people underreport embarrassing behaviors and thoughts on surveys. They want to look good, even though most surveys are anonymous. This is called social desirability bias.

Highlight (Yellow) | Page 162

When we lecture angry people, the search data implies that their fury can grow. But subtly provoking people's curiosity, giving new information, and offering new images of the group that is stoking their rage may turn their thoughts in different, more positive directions.

Highlight (Yellow) | Page 192

Key to their study was the fact that on some weekends, the most popular movie was a violent one—Hannibal or Dawn of the Dead, for example—while on other weekends, the most popular movie was nonviolent, such as Runaway Bride or Toy Story. The economists could see exactly how many murders, rapes, and assaults were committed on weekends when a prominent violent movie was released and compare that to the number of murders, rapes, and assaults there were on weekends when a prominent peaceful movie was released. So what did they find? When a violent movie was shown, did crime rise, as some experiments suggest? Or did it stay the same? On weekends with a popular violent movie, the economists found, crime dropped. You read that right. On weekends with a popular violent movie, when millions of Americans were exposed to images of men killing other men, crime dropped—significantly.

Highlight (Yellow) | Page 207

The media bombard us with correlation-based studies seemingly every day. For example, we have been told that those of us who drink a moderate amount of alcohol tend to be in better health. That is a correlation. Does this mean drinking a moderate amount will improve one's health—a causation? Perhaps not. It could be that good health causes people to drink a moderate amount. Social scientists call this reverse causation. Or it could be that there is an independent factor that causes both moderate drinking and good health. Perhaps spending a lot of time with friends leads to both moderate alcohol consumption and good health. Social scientists call this omitted-variable bias. How, then, can we more accurately establish causality?

Highlight (Yellow) | Page 217

The lesson of A/B testing, to a large degree, is to be wary of general lessons. Clark Benson is the CEO of ranker.com, a news and entertainment site that relies heavily on A/B testing to choose headlines and site design. "At the end of the day, you can't assume anything," Benson says. "Test literally everything." Testing fills in gaps in our understanding of human nature. These gaps will always exist. If we knew, based on our life experience, what the answer would be, testing would not be of value. But we don't, so it is. Another reason A/B testing is so important is that seemingly small changes can have big effects. As Benson puts it, "I'm constantly amazed with minor, minor factors having outsized value in testing."