



Article from

***Predictive Analytics and Futurism***

December 2019

# An Ever-Welcome Warning Against Big Data Hype: A Review of *Big Data, Big Dupe* by Stephen Few

By Mary Pat Campbell

Recently, I had the opportunity to test out a tool intended to provide more useful information about individuals, which I could use for my job. The tool was supposed to have used natural language processing and various artificial intelligence algorithms to appropriately structure and cluster results, so that what one received would be more useful than a simple internet search.

Of course, the first person I searched for was myself.

The result? Underwhelming, especially with the amount of completely wrong information.

My employment dates? Completely wrong for every single position mentioned, of the ones where I actually worked. In addition, it listed organizations and companies with which I've had no associations, ever. Now, of course, other Mary Pat Campbells exist, but the point of these types of tools is that they are supposed to automatically distinguish between the actuary Mary Pat Campbell and the psychotherapist Mary Pat Campbell and all the other Mary Pat Campbells.

I have the daily experience of Amazon ads being served up to me on every web browser page, trying to sell to me the book I just bought. Amazon has 20 years of my browsing and buying data—and this is what they do with it. I have political ads served to me on Twitter that are completely out of line with my politics, even though I have put my political labels right in my profile, and what I tweet and retweet gives it all away. For some reason, JCPenney sends me ads on Facebook indicating it thinks I want to buy a top hat. I have yet to buy a top hat.

All of these have been courtesy of big data, which not only has been touted as some paradigm-changing aspect of the business

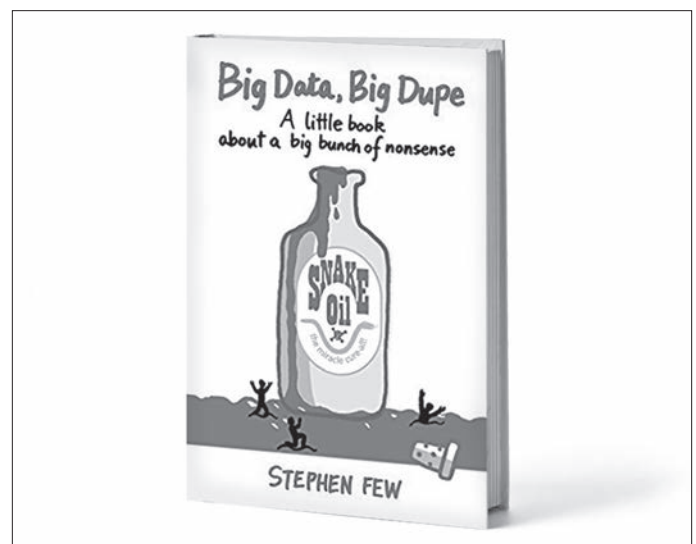
environment, but now has been cited as a danger to democracy in Congressional investigations. Are these benefits, and dangers, over-dramatized?

Given I have been deeply unimpressed with the results of big data as it has applied to me, I have obviously bought into the central premise of Stephen Few's book *Big Data, Big Dupe*.<sup>1</sup> Few argues that big data is an ill-defined marketing buzzterm, and he rebuts the hype some have published about this supposedly new phenomenon of the past decade.

The book comes with the subtitle "A little book about a big bunch of nonsense," which is absolutely true—the book is fewer than 100 pages, and there really isn't much in the bulk of the book other than debunking the razzmatazz surrounding big data. He especially likes beating up on the concept of the three V's: volume, velocity and variety. Most of his critique can be found in its original form in blog posts and newsletters at his site.<sup>2</sup> The original entries range from 2009 to 2016, and the book itself was published in 2018.

In particular, the argument regarding both the definition of big data and what qualities are supposed to make it different from the data landscape before takes the form of back-and-forth writing between the proponents of big data as a concept and Few.<sup>3</sup> Some of that is reproduced in the book, and I found it tedious.

Few's book has the problem of most business books: The bulk of the book is of little value if you don't need to be convinced of the premise, and the most important message is in the appendix or epilogue. You can follow the links provided and read (for free) what Few originally wrote before being professionally edited, if you need convincing that big data often doesn't deliver what the proponents say. I am unsure that actuaries need this warning,



given Actuarial Standard of Practice No. 23 on Data Quality.<sup>4</sup> Actuaries have always been alive to the possibilities that data may be of very poor quality, and that simply adding more data does not necessarily make the result better.

I have heard horror stories from others where generalist data scientists bring back analyses on books of business that were completely useless. . . .

Haranguing against a business fad, while always a popular pastime, is not useful if you're trying to accomplish something. That's where the real value of the book comes in: the epilogue, in which Few advocates an approach to effective data use. In January 2013, he wrote a blog post in which he proposed three S's in response to the three V's: small, slow and sure, which forms the heart of the book's epilogue.<sup>5</sup>

Using Few's own words:

**Small.** "As data increases in volume, we should keep in mind that only a relatively small amount is useful."

**Slow.** "No matter how fast data is generated and transmitted, the act of data sense-making, which must precede its use, is necessarily a slow process. We must take time to understand information and act upon it wisely."

**Sure:** "Just because you can collect data about something doesn't mean you should. In fact, given all the data that you've already collected, wouldn't it make sense to spend more time making use of it rather than getting wrapped up in the acquisition of more? When you recognize an opportunity to do something useful with data, that's when it becomes sure."<sup>6</sup>

In supporting this "slow data movement," Few provides nine items at the very end of the book for organizations and people to consider when they want to make effective use of their data. These nine items tie to the small, slow and sure items. Few focuses on "data-sensemaking," the entire point of using data in the first place. All the fancy statistical tools and huge volume of data will be useless if we don't understand what we're looking at.

Of the nine items, most of which should be extremely evident to actuaries (example: "3. Information is useful only if it is understood."<sup>7</sup>), the last one has the most impact to us professionally:

"9. Skilled data sensemakers combine general analytical skills with specific domain knowledge; one without the other is not enough."<sup>8</sup>

I have heard horror stories from others where generalist data scientists bring back analyses on books of business that were completely useless because the analysts didn't understand the nature of the products ("Profitability is driven by policy year!" "Great, we'll go back in time and write more 2015 business"). The situation I've most run into with respect to insurance products and data science is having to operate in a restrictive environment compared to attempting to serve up ads online. Some data, even if available, cannot be legally used. Other data are questionable to use in an industry where reputation with customers is extremely important.

This book may be useful in providing warning to those who do not realize the danger of indiscriminately trying all the computing toys available. At best, some time and effort are wasted if care is not taken. At worst, pricing, underwriting and more could get out of whack if one does not have appropriate knowledge about both the tools and the data the tools are being applied to.

As noted, the book is small, and this reads extremely quickly. If you know of people who need to be warned about what big data can do for their companies, this book would be a good way to warn them—and, as noted above, what they should be doing instead is given at the end. Go out and make sense of your data!

And to JCPenney: No, I don't want a top hat.



Mary Pat Campbell, FSA, MAAA, PRM, is vice president, insurance research at Conning in Hartford, Conn. She can be reached at [marypat.campbell@gmail.com](mailto:marypat.campbell@gmail.com).

## ENDNOTES

- 1 Few, Stephen. 2018. *Big Data, Big Dupe: A Little Book About a Big Bunch of Nonsense*. El Dorado Hills, CA: Analytics Press.
- 2 <http://www.perceptualedge.com/>.
- 3 For an example, see Few, Stephen. Big Mouths on Big Data. *Visual Business Intelligence* (blog). April 30, 2014. <http://www.perceptualedge.com/blog/?p=1891>.
- 4 Actuarial Standards Board (ASB). 2016. Actuarial Stand of Practice (ASOP) No. 23: Data Quality. Revised edition. Doc. No. 185. <http://www.actuarialstandardsboard.org/asops/data-quality/>.
- 5 Few, Stephen. The Slow Data Movement: My Hope for 2013. *Visual Business Intelligence* (blog). Jan. 2, 2013. <http://www.perceptualedge.com/blog/?p=1460>.
- 6 Ibid.
- 7 Few, *Big Data, Big Dupe*.
- 8 Ibid.