Loreen Henry

EPPS 6356 Data Visualization (Fall 2021)

The Parable of Google Flu: Traps in Big Data Analysis

September 20, 2021

Google Flu Trends (GFT), the flu tracking system, in February 2013 predicted that there would be more than double the incidence of doctor visits for flu-like symptoms for that year. The science journal, Nature, reported that GFT's creators were estimating above that of the Centers For Disease Control (CDC) which used reports from laboratories across the United States to create its estimates. Many people wondered how such an over-ambitious estimate could have occurred. The authors examined their thesis that often big data can lead to inaccurate outcomes. Coupled with algorithm dynamics, the changes made by engineers to improve the commercial service and by consumers in using that service, there can be unexpected interpretation in the analysis of the big data. At such time the analytics becomes flawed, and the outcome cannot be trusted.

Researchers and data scientists need to cognizant that there are many errors that can occur when big data is prepared. They should be attentive to the fact that they are trusted to not manipulate data or findings for their own benefit or to influence the population. The "Parable of Google Flu" which is about overfitting and how such errors can impact people's confidence in data and its use. Technology trends today and the future of big data needs to be cognizant of how there can be false analysis of what the numbers are revealing. This can impact human lives in numerous ways.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. “The Parable of Google Flu: Traps in Big Data Analysis.” *Science* 343 (March): 1203-1205. AAAS (American Association for the Advancement of Science) Journals (September 20, 2021).