Jessica Dorman

Module 1: Assignment 1

In 2008 researchers from Google claimed that they could produce estimates of flu prevalence two weeks earlier than the CDC based on peoples searches of flu- like symptoms. If accurate Google could turn the digital refuse of peoples searches to potentially life- saving insights.

The Google Flu Trend (GFT) failed miserably missing the peak of the 2013 flu season by 140 percent. Google quietly killed the program turning it into the poster child of the foibles of big data. Research teams from several universities compared Googles findings to that of the CDC and determined that GFT had begun to perform worse, requiring several revisions after two to three years.

When determining the value of big data we must focus on the methodology used. Google is the powerhouse of big data giving them the responsibility to use the data din the publics best interest. Googles algorithm was very vulnerable to seasonal terms unrelated to the flu. As with any prediction there were bound to be searches strongly correlated simply by chance, unlikely to driven by actual flu cases. Google also introduced their suggested search feature which could have thrown of tracking.

If done right the GFT could serve as a model for collaboration around big data for the public good. If Google does not work with organizations such as the CDC to improve their methodology and continually update the fit of the data to flu prevalence with value of the data will rapidly decline.

While this prediction of flu prevalence did not succeed we can see the impact something as simple as data around peoples searches could have on a variety of public issues.

<https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/>