**The Parable of Google Flu**

This article is about some pitfalls in big data analyses that Google fell in. it talks about when Google over predicted the proportion of doctor visits for influenza-like illness. This prediction was noted to be about twice the figure for the Center for Disease and Control Prevention (CDC).

The mistake came from what the article calls big data hubris and algorithm dynamics. The articles talks about the fact that most people treat big data as a substitute instead of a supplement to traditional methods of data collection. The problem is that some big data are not necessarily collected with reliable and dependable instruments of measurement. In trying to curate search terms to fit 1152 data points, the GFT ended up detecting not just flu-like terms, but also winter-like terms that have seasonal influences. These seasonal observations essentially threw off the predictions by overestimating the possible number of cases. Also, the serial correlation of errors meant that GFT was not infact dealing with randomly distributed errors.

The algorithm dynamics deal with the fact that the instrumentation may not actually be capturing what you said it’s capturing. Google was constant changing its own search algorithms, forgetting that the new algorithms may measure the data they are trying to gather differently. Media activity also could have thrown the normal levels of flu search overboard through panic and fear warnings.

What I learnt from this paper is that as researchers, we need to be careful not to forget the importance of traditional methods of data collection. We need to also be sure to have properly defined scopes of data collection before we begin. And finally, we need to make sure the instruments we are using are consistent across time and space such that they will measure the same thing at the same rate and accuracy over and over again.