Assignment 1

- **What was the problem with the Google flu detection algorithm?**

In comparison to the Centers for Disease Control and Prevention, which estimates influenza-like illness based on surveillance reports from laboratories across the country, Google Flu's algorithm predicted more than double the proportion of doctor visits. The two factors contributing to Google Flu Trends' (GFT's) errors were data hubris and algorithm dynamics.

- **What is big data hubris?**

In Big data hubris, the implicit assumption is that big data is more of a substitute than a complement to traditional data collection and analysis. Big data provides tremendous scientific opportunities. The quantity of data does not mean that one can ignore foundational issues such as the validity and reliability of measurements and constructs, and the dependence between data.

- **What approach could have been used to improve the Google flu detection algorithm?**

The Google Flu detection algorithm can be enhanced by combining it with other near-real-time health data. We can substantially improve on the performance of either Google Flu Trends (GFT) or the Centers for Disease Control and Prevention (CDC) alone by combining GFT and lagged CDC data, as well as dynamically recalibrating GFT. By incorporating this information, GFT had the potential to largely heal itself.

- **What is “algorithm dynamics?”**

Algorithmic dynamics refers to the changes made by engineers to improve a commercial service and by consumers in order to make that service more efficient.

- **What aspect of algorithm dynamics impacted the Google flu detection algorithm?**

It is important to note that the Google search algorithm is something that is not static. Considering algorithm dynamics affecting Google's search algorithm, it's quite likely that Google flu trends(GFT) was an unstable reflection of flu prevalence. Several changes to Google's search algorithm and user behavior are likely to have an effect on GFT's tracking. There is a sense of misdirection in the GFT's search terms examples.

There are multiple challenges to replicating GFT’s original algorithm. It is most likely that GFT's error was caused by a panic spurred by the media during the last flu season. It is unclear why GFT has missed high by wide margins for more than two years, even if this is a factor.