

# Problem Set 2

Matt Mullins

24 January 2018

## 1 Tools of a Data Scientist

Data scientists have many tools that they can use to clean and manipulate data in order to uncover new insights. The key tools (big 4) they use include:

- 1. Data Collection - In this step of the process the scientist wants to ensure they are collecting data that supports what they actually are wanting to measure. At this stage they should also determine how to quantify what "best" is in terms of data and results.
- 2. Data Storage and Manipulation - Here the data scientist is cleaning and organizing data so that it will prove easier to work with.
- 3. Visualization - One way this can be done is by viewing the summary of statistics from the data including mean, median, etc. This will help the scientist understand how the data actually looks compared to how it was hypothesized to look.
- 4. Statistical Programming and Modeling - Using any methods described in the 5 major tribes of machine learning, data scientists can use machine learning and advanced algorithms to uncover new and more accurate insights from the data that they are attempting to study.