

# Problem Set 2

savannahjsimpson

February 2024

## 1 Main Tools of Data Scientists

- Measurements allow data scientists to construct insights and policies by allowing them to track what they are meant to accomplish. Measurements can take on various forms depending on what they are trying to capture. They could be continuous or discrete.
- Statistical programming languages are specialized sets of commands and characters that allow data scientists to easily manipulate data. There are many different languages such as R, Python, Julia, etc. Many are similar, but there is an appropriate language for each user and their needs.
- Web scraping provides a useful way for data scientists to pull data from the internet. This can be accomplished in almost any programming language.
- Resilient distributed datasets (RDDS) serve as a tool to better tackle large data sets. They require a cluster of computers and software such as Hadoop or Spark.
- Visualization allows data scientists to present data by creating models that are easily understood by the average user. Visualization is key in communicating data patterns to professionals who are not as familiar with data.