PS2 McGuire

Joel McGuire

January 2018

1. Measurement. What is the question? Can we even answer it? What is the scope of the project? Can it be done on my macbook or does it need to be done on the cloud? Is this going to parametric or non-parametric? Are we looking for inference, causality, prediction or clustering?

2. Statistical Programming Languages. Python, R, Julia.

3. Web scraping.

    (a) API
    (b) Parsing HTML

4. Big Data

    (a) RDDS : Hadoop , Spark
    (b) Cloud computing / storage (AWS)
    (c) SQL

5. Visualization.

6. Modeling. Predict things, win acclaim.