

Problem Set 2

Mason Ross Hayes

January 2020

1 Main Tools of a Data Scientist

- Measurement
- Programming languages
 - R
 - Python
 - Julia
- Web scraping
 - APIs
 - Parsing
- Handling large data sets
 - RDDs
 - * Hadoop
 - * Spark
 - SQL
- Visualization
 - ggplot2 (R)
 - matplotlib (Python)
 - Plots.jl (Julia)
- Modeling