Problem Set 2

Trevor Jones

February 7, 2023

1 Tools of a data scientist

1. Measurement

2. Statistical Programming Languages

   (a) R
   (b) Python
   (c) Julia + many others

3. Visualization Tools

   (a) ggplot2 (R)
   (b) matplotlib (Python)
   (c) Plots.jl (Julia)
   (d) Tableau

4. Big Data Management Software

   (a) Resilient Distributed Datasets (RDDs)
   (b) SQL (Structured Query Language)

5. Data Collection Tools

6. Modeling