Problem Set 2

Lilly Kirby-Rivera

February 6, 2024

Data Science Toolbox

- Measurement: Decide what metric(s) you will use to quantify your data

- Statistical Programming Languages

    - Main languages for this class: R, Python, and Julia
    - Others: Stata, SAS, SPSS, Matlab, and JavaScript

- Web Scraping: Through use of APIs or parsing HTML codes

- Handling large data sets

    - Resilient Distributed Datasets (RDDs): Helps break data into manageable bits
    - SQL: Popular database language that can help you manage large data sets

- Visualization: A sense general expectations for you data and understanding what it tells you

    - ggplot2 (R)
    - matplotlib (Python)
    - Plots.jl (Julia)
    - Tableau

- Modeling: Using your data to answer research questions