PS 2

Mengyang Davila

February 7, 2023

Main tools of a data scientist:

- Measurement

- Statistical programming languages

  - Mathematical and statistical programming languages: R, Python, and Julia

  - Other languages: Stata, SAS, SPSS, Matlab, and JavaScript

- Visualization tools

  - R: ggplot2

  - Python: matplotlib

  - Julia: Plots.jl

- Big Data management software

  - Resilient Distributed Datasets (RDDs)

  - SQL (Structured Query Language)

- Data collection tools

  - Web scraping

    * Using an API to download data

    * Downloading HTML files and parsing text to extract data