

PS2

William Townsend

February 6, 2022

1 List of Data Scientist Tools

1.1 Measurement

This is how policies are constructed.

1.2 Statistical Programming Languages

Data scientists use a plethora of different statistical programming languages such as R, Python, and Julia. Other languages that can be used for statistical work includes Stata, SAS, SPSS, Matlab, and JavaScript.

1.3 Web Scraping

Web scraping entails using an API to download data. It also can involve downloading HTML files and parsing their text to extract data.

1.4 Resilient Distributed Data Sets

A person needs a cluster of computers and software such as Hadoop or Spark to use RDDs. Using RDDs allows people to subset data, create summary statistics, and withstand any disruptions in computing clustering.

1.5 Structured Query Language

Allows people to easily subset, merge, and perform other common data transformations.

1.6 Visualization

Allows people to see in multiple dimensions what data looks like, spot outliers, and perform sanity checks on data. Examples are ggplot2, matplotlib, plots.jl, and tableau.

1.7 Modeling

This allows people to use the data to test theories, predict behavior and explain behavior.