PS2$_C$astiglione.tex

joe.castiglione.jr

January 27 2020

Main Tools of Data Science

1 Measurement

Measurement is how "insights" or "policies" are constricted.

2 Statistical Programming Languages

There are two different type of statistical programming languages **scripted** and **compiled**. Some examples are R, Python and Julia.

3 Web Scraping

Using the mass amount of publicaly stored data and taking it off to use.

4 Handling Large Data Sets

Data sets that are too big to fit on one hard drive are considered too big. There are many ways to use data that is too large, RDD and the SQL.

5 Modeling

Modeling allows you to test theories, predict behavior from data and find ways to explain it.

6 Vizualization

Allows you to see the data as multiple dimensions. Good way to spot outliers and perform "sanity checks" on the data.