Data Science
Student: Ruichun Liu
Problem Set 2: Data science toolbox

1. **Measurement**
   In class, we first discussed what measurement is. Measurement is how "insights" or "policies" are constructed, because one cannot construct a policy without first measuring what the policy is meant to accomplish.

2. **Statistical Programming Languages**

   - R
   - Python
   - Julia
   - Stata
   - SAS
   - SPSS
   - Matlab
   - JavaScript
   - C
   - C++
   - Fortran

3. **Web Scraping**

   (3.1) How does web scraping work?
   - Using an application program interface (API) to download data
   - Downloading HTML files and parsing their text to extract data

   (3.2) Web scraping in the three data science languages.
   - R
   - Python

4. **Handling large data sets**

   (4.1) RDDs
   (4.2) SQL

5. **Visualization**

   (5.1) ggplot2 (R)
   (5.2) matplotlib (Python)
   (5.3) Plots.jl (Julia)
   (5.4) Tableau

6. **Modeling**

   - Use the data to test theories.
   - Use the data to predict behavior.
   - Use the data to explain behavior.