**Data Bootcamp Project 1 – Climate Study**

**Team Members**

* Will Davis
* Mohamed Mursal
* Thaimu Sesay
* Jessica Snowden

**Project Ideas**

* Question: How much has the average high temperature for the 5 largest cities in the USA changed over the past 20 years?
* Question: How much has the average annual precipitation for the 5 largest cities in the USA changed over the past 20 years?
* Question: Is there a relationship between average high temperature and average annual precipitation?

**Data Exploration**

* What are the largest 5 cities?
  + Use Census data to determine the 5 largest cities
    - Find .csv or API with population by city
    - Sort descending by population
    - Drop all but top 5
* What is the annual high temperature for each city?
  + Using OpenWeatherAPI, pull daily high temperature by city
  + Compress data into measure of central tendency
    - Average by month?
    - Average by year?
    - Average of the highest values? (set cut-off for highest values/year by using box/whisker analysis)?
* What is the annual precipitation by location?
  + Using OpenWeatherAPI, pull daily precipitation by city
  + Compress data into measure of central tendency
    - Sum by year?
* Few notes
  + 20 years of data = 7,305 daily values
  + We can compress/reduce this data to 20 rows by choosing a way to represent each year with a single value of temperature and precipitation

**Proposed Analysis**

* Map showing 5 largest cities
* Box-Whisker plot to show how we decided how to reduce daily temperatures into a single representative value for each year
* Line graph showing average temperature by year for all 5 cities (on same plot)
* Line graph showing annual precipitation by year for all 5 cities (on same plot)
* Scatter plots showing relationship between average temperature (x-axis) and annual precipitation (y-axis) for each city
  + Plot regression lines on each scatter plot