This was a very interesting experience in not only the process of acquiring the data but also in finding the correct sources for data. As mentioned in my previous milestones, I found it necessary to choose new data sources for both the Web scraping and API portions of this project. Additionally, finding a way to finally join the API section to the other two sections of data was a challenge prior to choosing to randomly place each instance in one of the nine different regions from the climate data.

To successfully merge all three data sets, I first took the census data and connected the various states found in the Locations column to their associated regions according to the information from the website <https://www.ncdc.noaa.gov/>:

* Northeast: Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont
* Upper Midwest: Iowa, Michigan, Minnesota, Wisconsin
* Ohio Valley: Illinois, Indiana, Kentucky, Missouri, Ohio, Tennessee, West Virginia
* Southeast: Alabama, Florida, Georgia, North Carolina, South Carolina, Virginia
* Northern Rockies: Montana, Nebraska, North Dakota, South Dakota, Wyoming
* South: Arkansas, Kansas, Louisiana, Mississippi, Oklahoma, Texas
* Southwest: Arizona, Colorado, New Mexico, Utah
* Northwest: Idaho, Oregon, Washington
* West: California, Nevada

After adding the Region column to the census data, I merged both the climate and census data sets using an outer join. After splitting the Steam game data randomly into each region and using another outer join, the resulting data set had 108 columns and over 6.5 million instances. Reviewing the columns for the census data I chose to remove most of them in order to reduce the size of the final data set and make it more closely match what I desired for my analysis. The final cleaned data set has 28 columns by 6830663 rows.