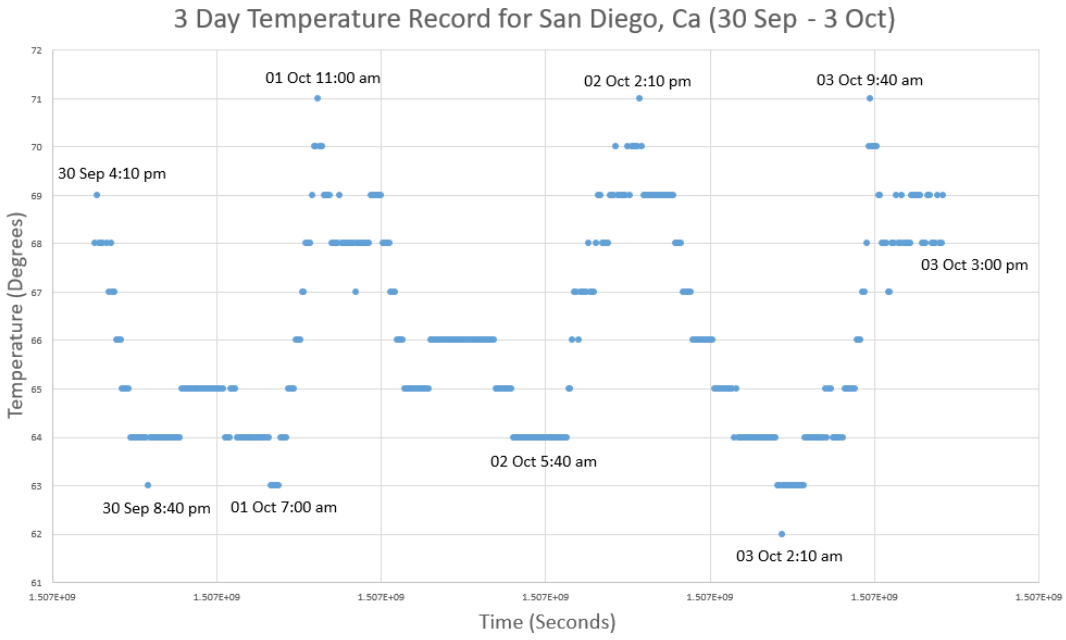
Assignment 4 - Show Me the Data. Go get it too!

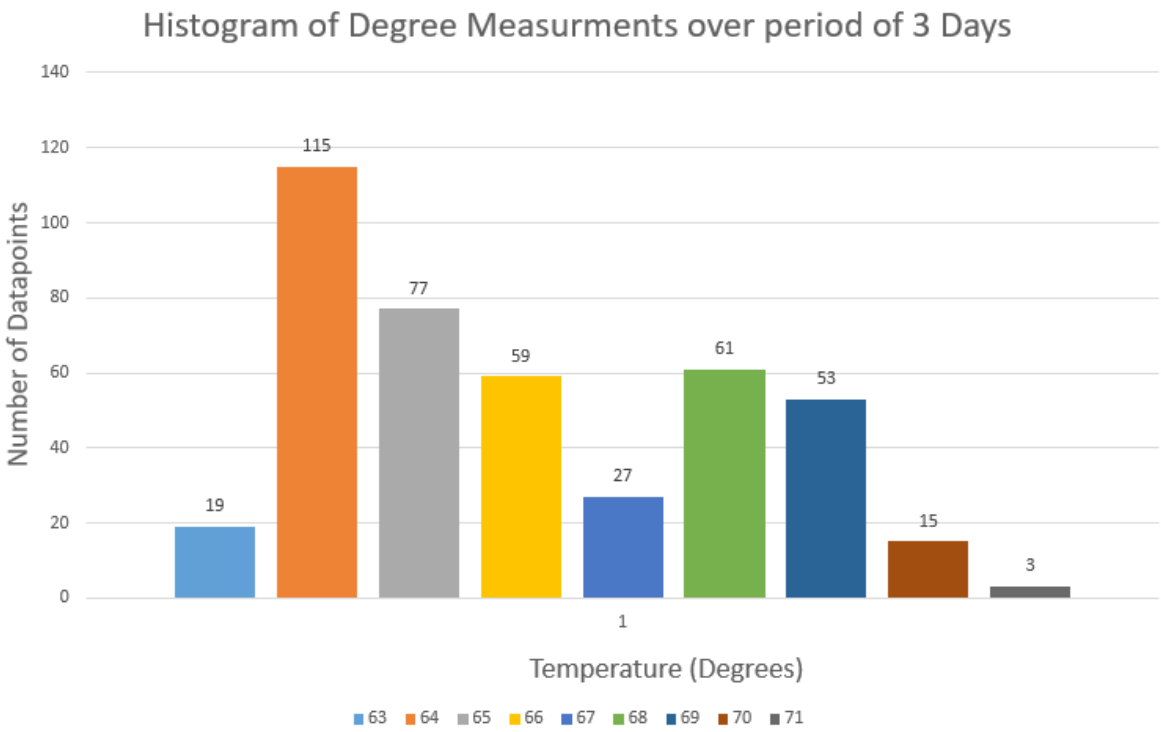
Austin Derbique A01967241 CS5665 10/3/17

1. While looking at the request placed In “GetNDFDWeatherForPlace.py” file, it appears as if the call made from the National Weather Service website gives back a result in the local time of the location queried. While parsing the data, the local time of the machine running the system is used to process the data. This becomes an issue as the time differences between the location of the query and the local time grow. To fix this issue, both times must be operating on the same time zone to correctly display the right image. This can be done by either shifting the current time to match the target time or vice versa.
2. This did not work for several zipcodes I tried including Rapid City, South Dakota (zipcode 57701). Instead, I chose San Diego, California (zipcode 92129). The XML link is: http://www.wrh.noaa.gov/mesowest/getobextXml.php?sid=DMHSD&num=72

3. I used Element Tree. This is the XML parsing package used in the example code. The reason I chose this package is for simplicity of use. Saved in files for this problem are parse.py, data.txt, and data.csv.
4. The graph for Temperature over Time can be found at graph.png



5. The source code for this can be found as histogram.py with output files as histogram.csv and histogram.png



6. This file is named as getairtemp.py
7. These files are stored in the directory air_temp_files. The source code file for this is called getairtemp.py. My approach for this problem is to have a loop from 0 to 191 that calls the XML rquest and saves to a file. After the file is saved, there is a time.sleep(900) function called which pauses the program for 15 minutes. Once the 15 minutes is up, the loop will iterate to the next value, all the way to 192 iterations.
8. The combined unzipped size for all 192 files is 2.2MB. The average size per file is 11.45 kilobytes.
9. It appears that the 15 minute collection interval is sufficient. I believe it could even be increased to a longer time in between data collection. The reason for this is because the values do not appear to be changing very often which means we're collecting a lot more data that does not necessarily bring more value to our predictions. That being said, it is relatively cheap to record data as these files are small and easy to parse.