```
Laura Wilke
A09240216
```

Alabama Florida Border Weather Analysis

This is a report on the analysis of historial weather data in the Alabama Florida Border region of the United States. The historical data is from NOAA (https://www.ncdc.noaa.gov/ (https://www.ncdc.noaa.gov/)) and was downloaded from (ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/).

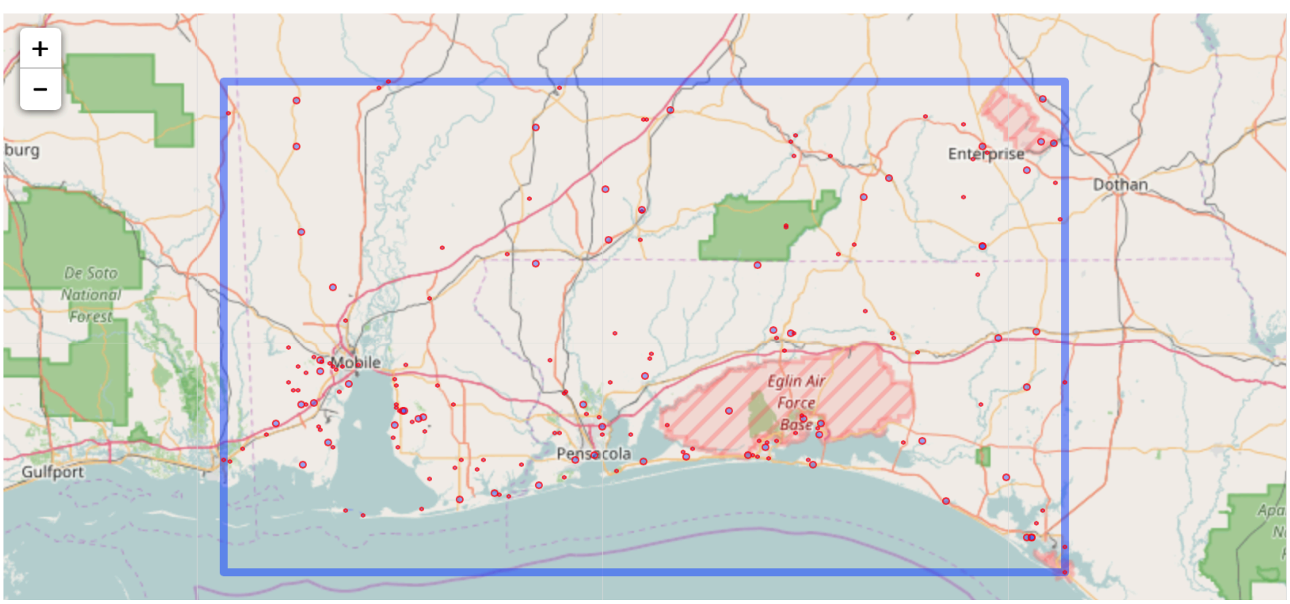
For this analysis, the focus is on six different weather measurements. The measurements are as follows:

- **TMIN, TMAX:** the daily minimum and maximum temperature (degrees F).
- **TOBS:** The average temperature for each day (degrees F).
- **PRCP:** Daily Percipitation (in mm).
- **SNOW:** Daily snowfall (in mm).
- **SNWD:** The depth of accumulated snow (in mm).

The region in particular that the analysis is focused on is within a box with the following coordinates:

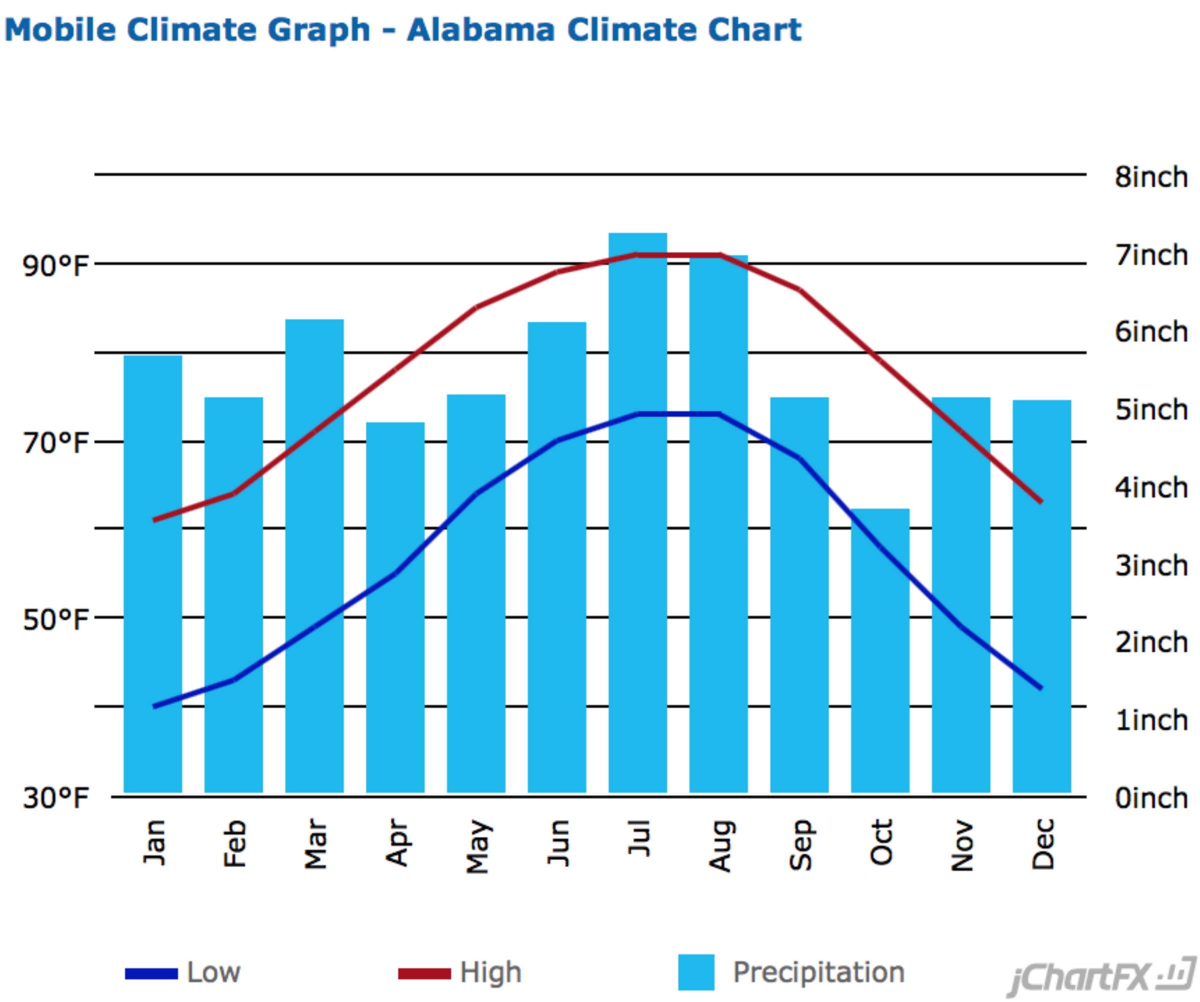
- **min latitude:** 30.066700000000001
- **max latitude:** 31.525300000000001
- **min longitude:** -88.5
- **max longitude:** -85.582400000000007

Within this region, there are 177 weather stations that the six weather measurements have been collected from. On the map below, the area within the box, is the region that is being analyzed. All of the circles represent one of the 177 weather stations.



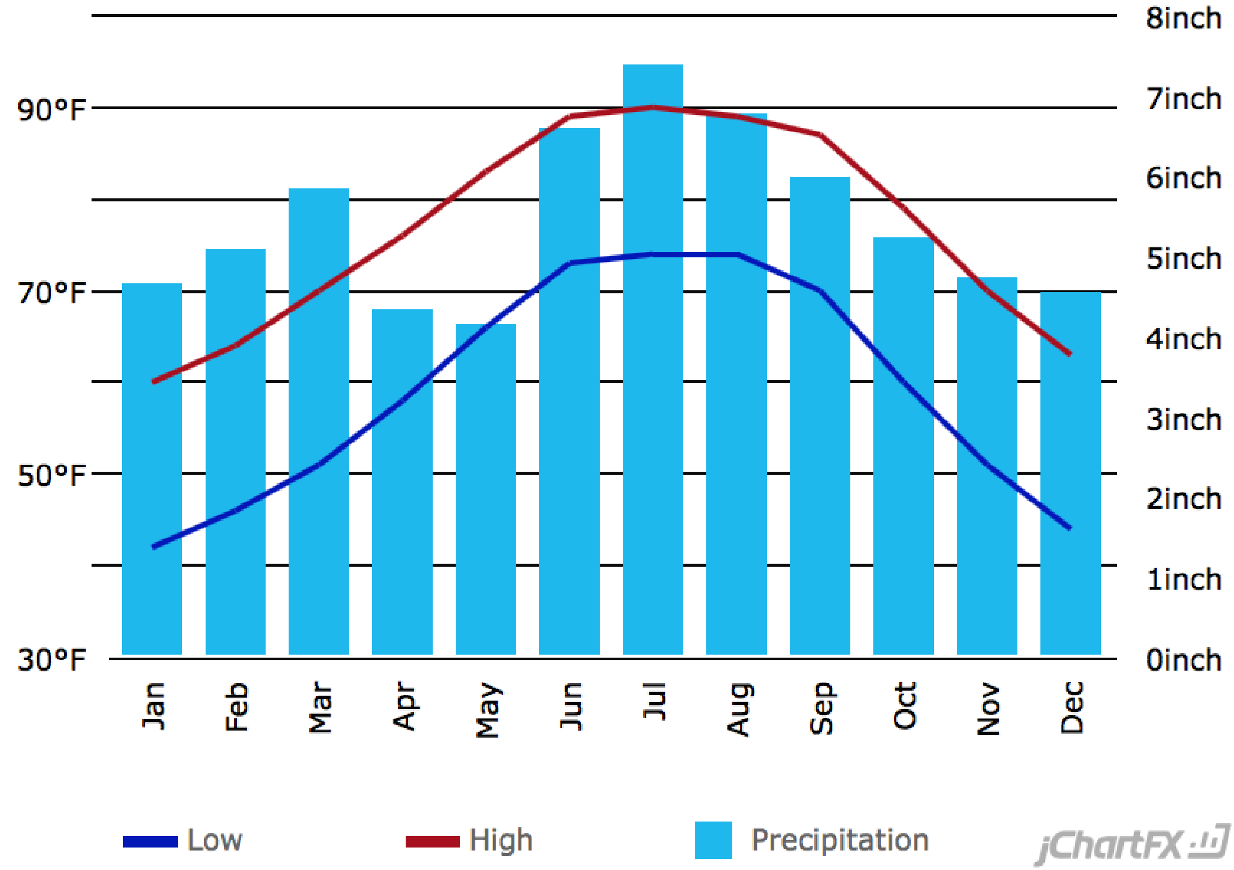
General Weather Trends

In order to validate that the general weather trends analysis, US Climate data obatined from US Climate Data (http://www.usclimatedata.com/climate) will be used to verify the trends. For this verification, Mobile, AL and Pensacola, FL (two cities within the region) weather data will be used.



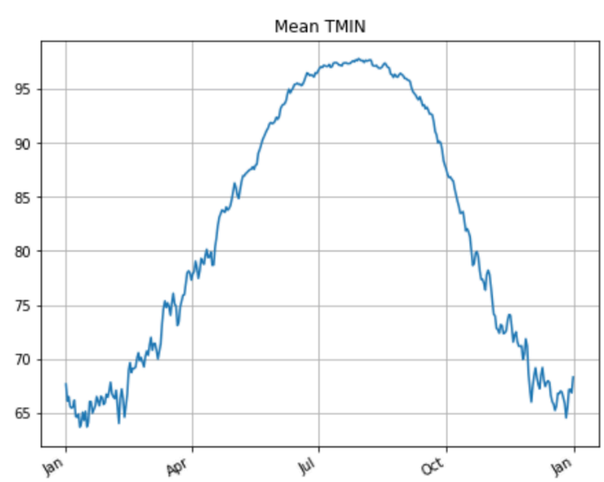
These graphs indicate the average min and max temp per month and the average percipitation.

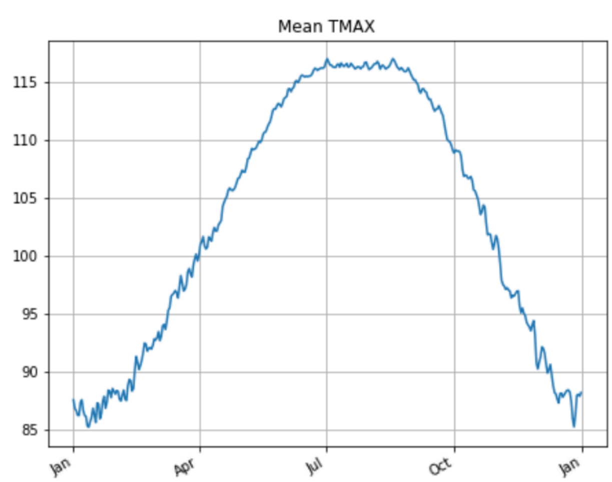
Pensacola Climate Graph - Florida Climate Chart



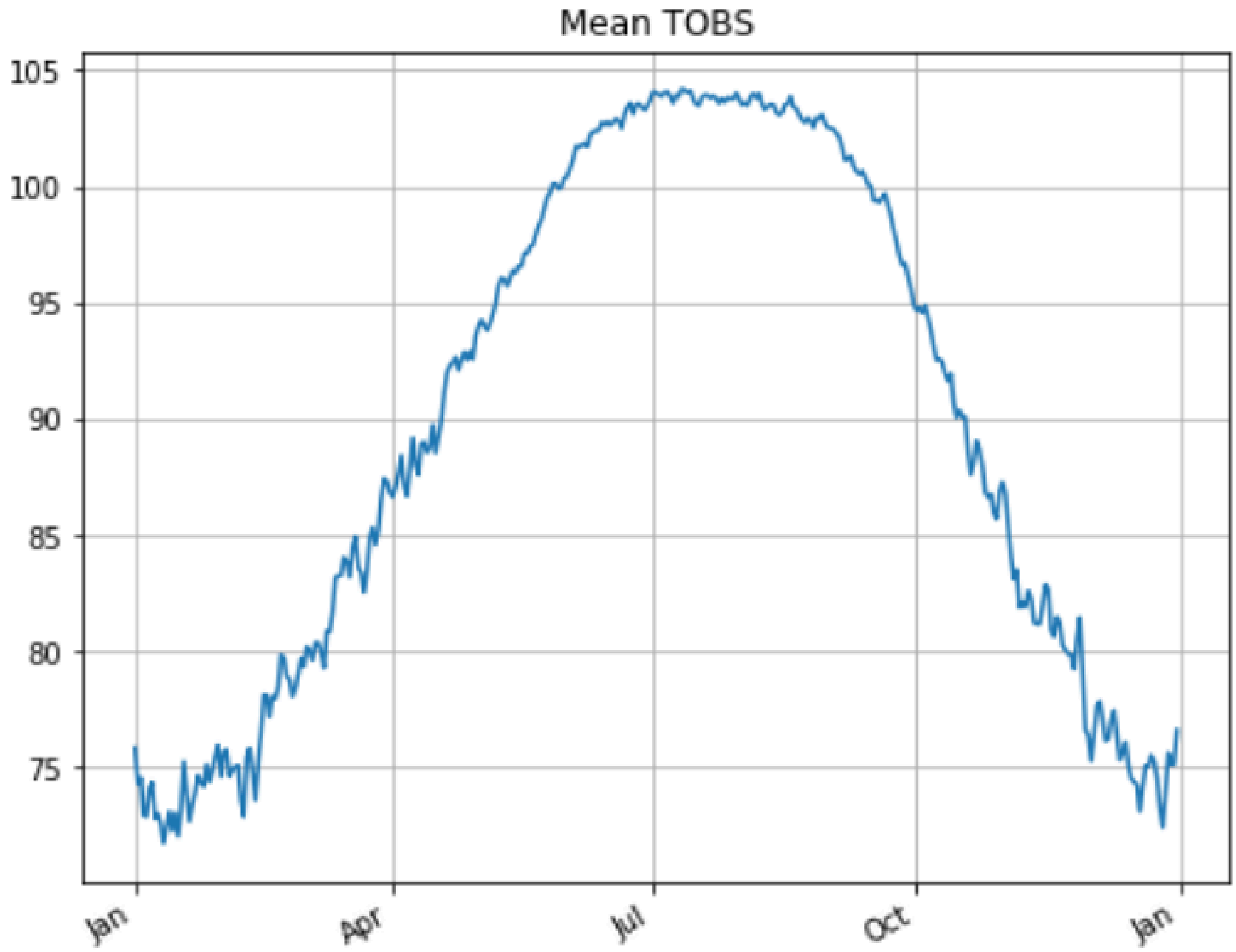
Temperature

The average TMIN and TMAX follow the trend of the temperature increasing in the summer and decreasing in the winter. The chart below depicts the average daily min temperature and max temperature throughout the year.





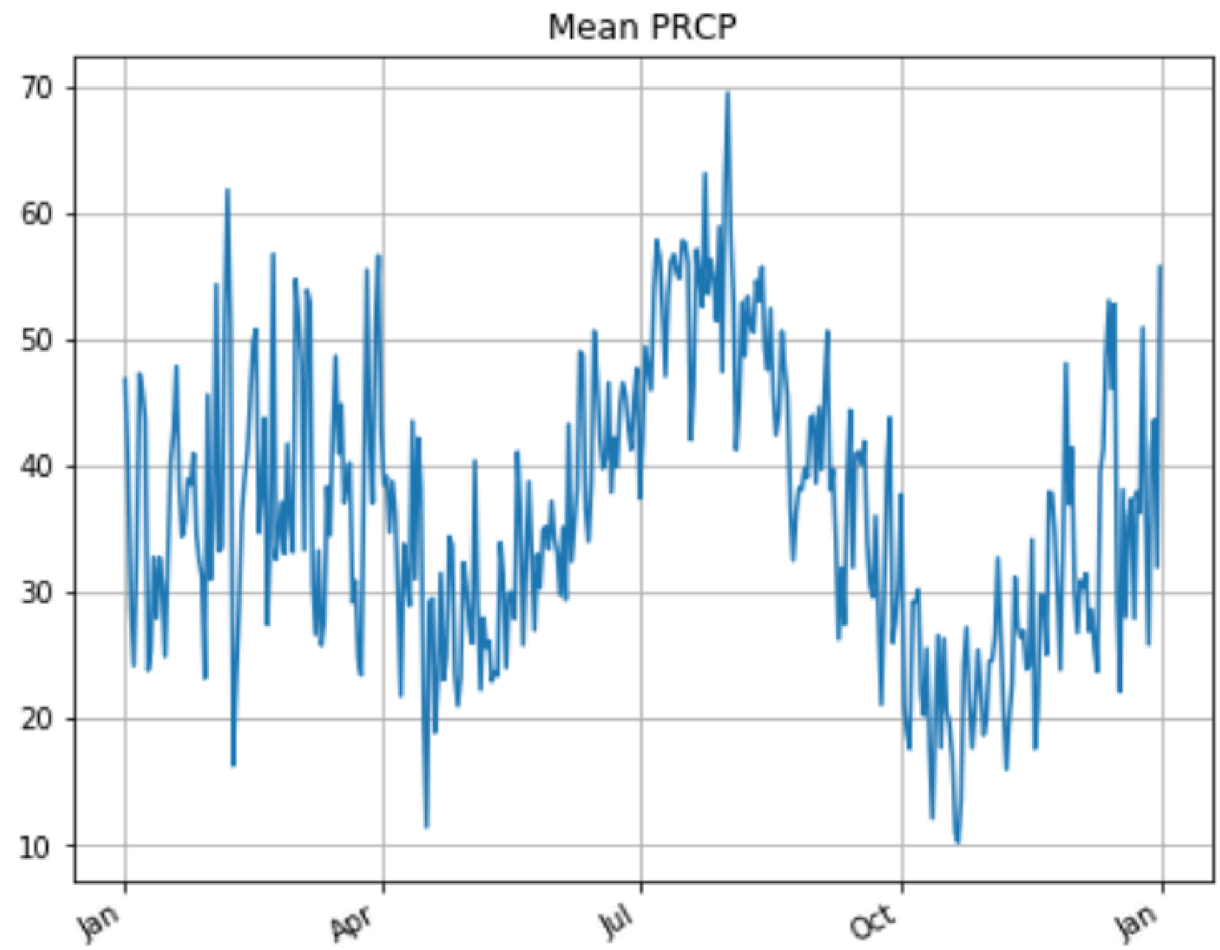
The daily average temperature is depecited below throughout the year.



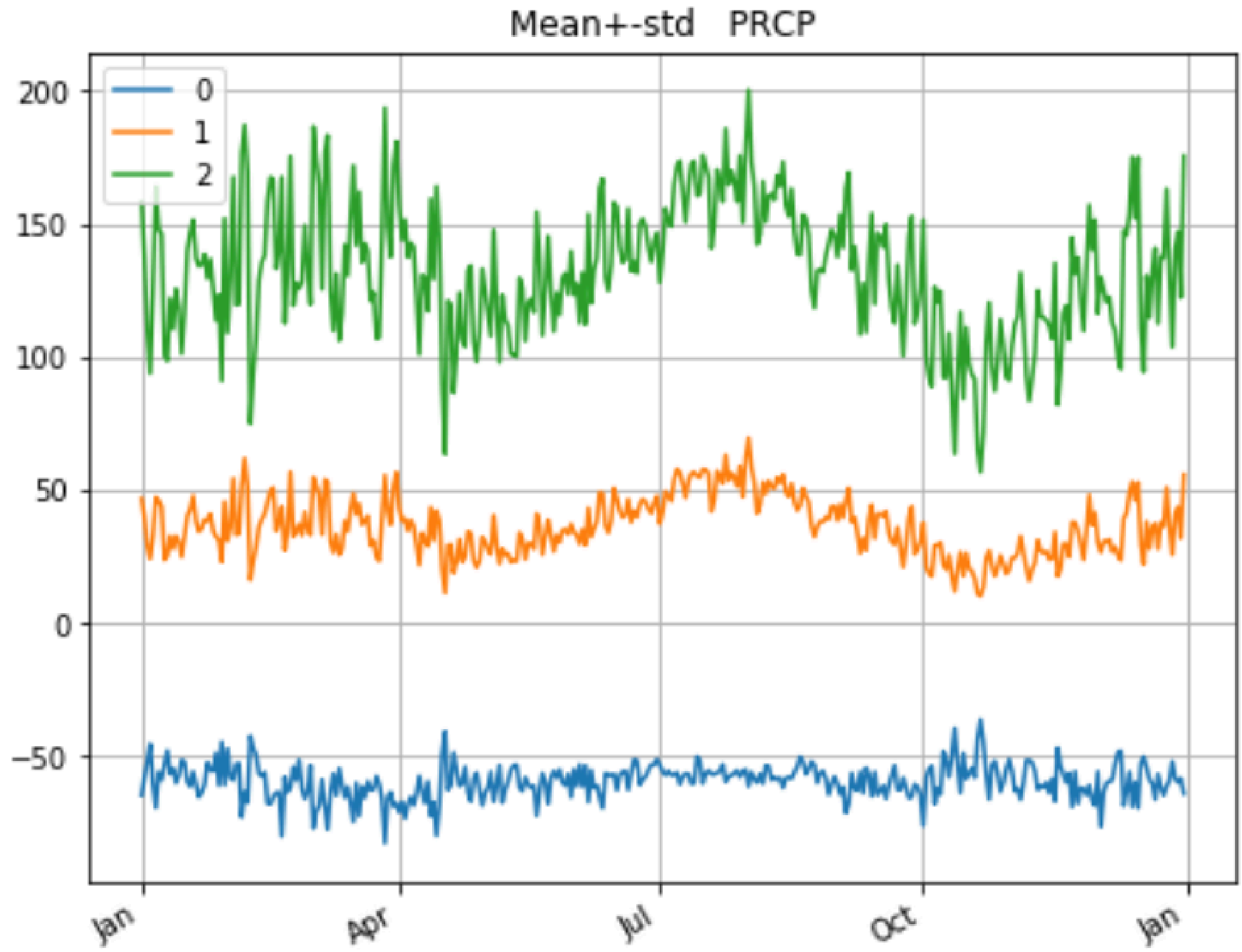
All three of these graphs indicate that the temperature in this region is warmer than the US Climate temperature data. Later on, there will be a deep dive into the temperature data in order to account for this variation between the two data sets.

Precipitation

According to the analysis, the average daily rainfall is 36.41 mm or 1.43 inches.



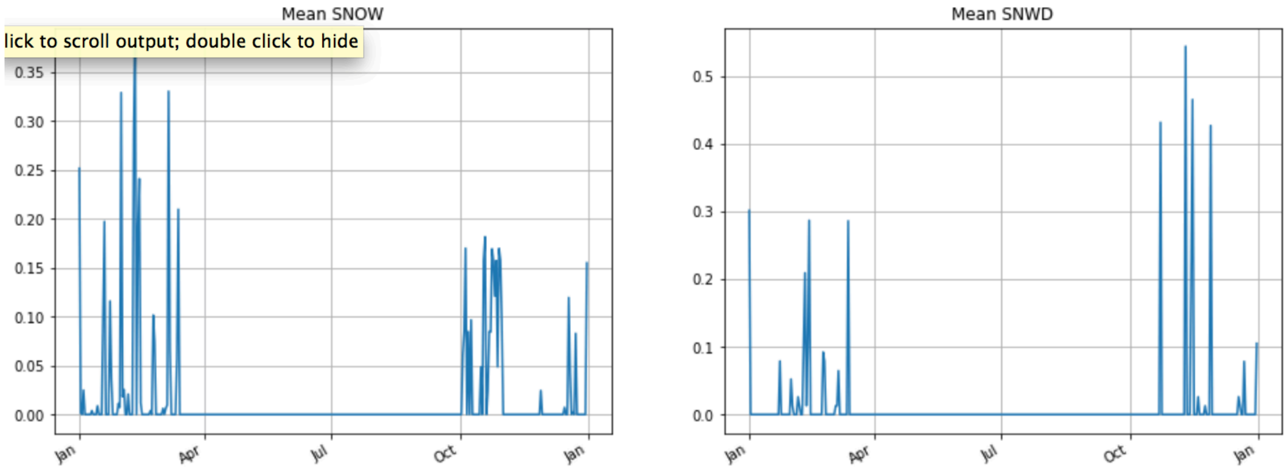
The image above depicts the average daily rainfall throughout the year. April, May, and June are the monthes where there is a drop in the average daily rainfall. In order to fully capture the amount of daily rainfall this region can experience, the next image depicts the full range of possible daily rainfall per month.



This image depicts the mean (orange), mean+standard deviation (green), and mean-standard deviation (blue). This image gives a better representation of the possible rainfall per month throughout the year. In the summer monthes, the rain fall can get up to around 7 inches/day. It captures the total range of the possible rainfall and better correlates with the US Climate data for the amount of precipitation in Mobile and Pensacola.

Snow and Snow Depth

As one would expect, there is very little snow in this region of the United States. The following image shows the average amount of snow in mm and the average snow depth throughout the year.



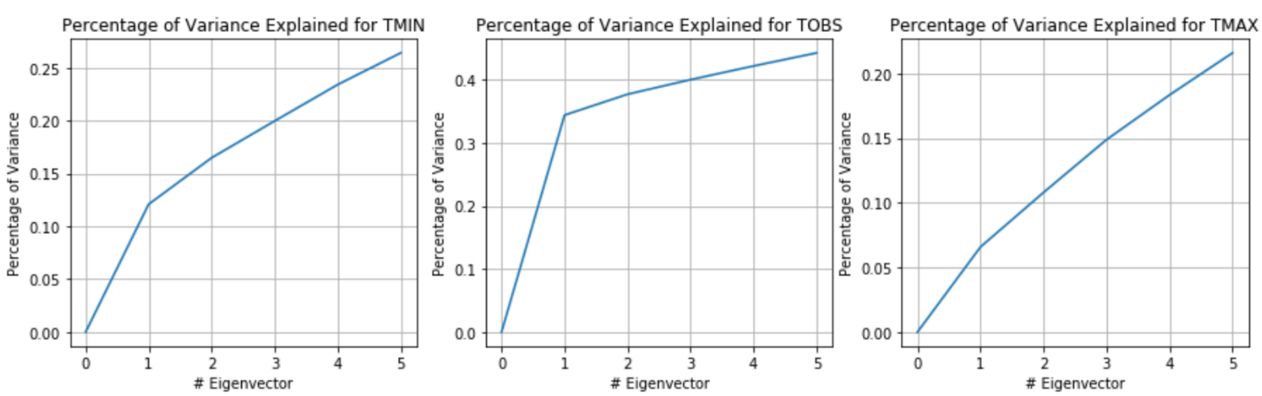
There is no snow between April to October. If there is snowfall in the winter months, there is very little of it and it does not stay on the ground for very long.

PCA Analysis

For each of the six measurements, the percentage of variance explained as a function of the number of eigen-vectors used was calculated.

Temperature

For the first 5 eigen-vectors for TMIN, TMAX, and TOBS, the percentage of variance that was explained was 27%, 22%, and 47%, respectfully. For TMIN, the first eigen-vector captures 12% of the variance and for TMAX 7% of the variance is explained by the first eigen-vector. For TOBS, the first eigen-vector captures 35% of the variance.

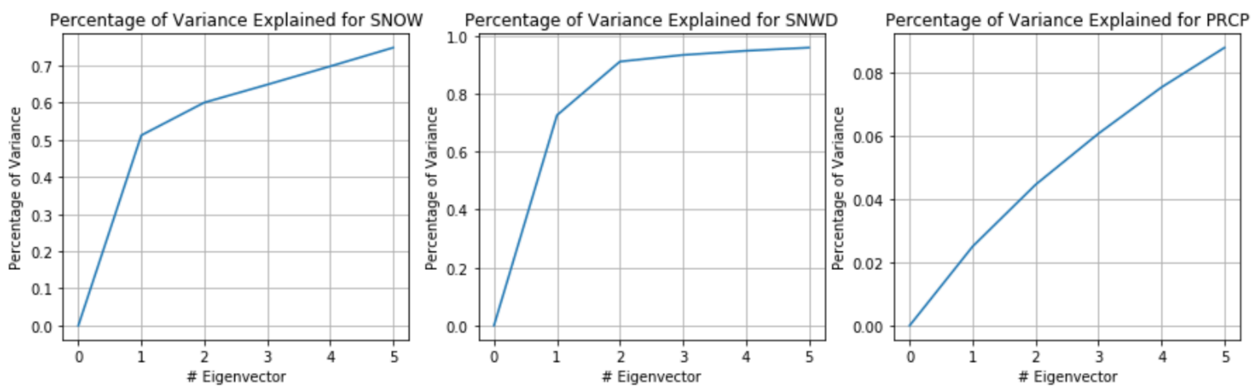


Percipitation

For the first 5 eigen-vectors of PRCP only 8.5% of the variance is explained and the first eigen-vector only captures 2.5% which is extremly low (illustrated on the graph below). This implies that the percipitation fluxuates throughout the year since there is a lot of noise.

Snow and Snow Depth

The first 5 eigen-vectors of SNOW and SNWD capture 75% and 98% of the variance, respectfully. The second eigen-vector for SNOW captures 60% of the variance and for SNWD the second eigen-vector captures 90%. Due to the fact that it rarely snows in this region, it makes sense that there isn't much noise and that the first two eigen-vectors capture a significant amount of the variance.

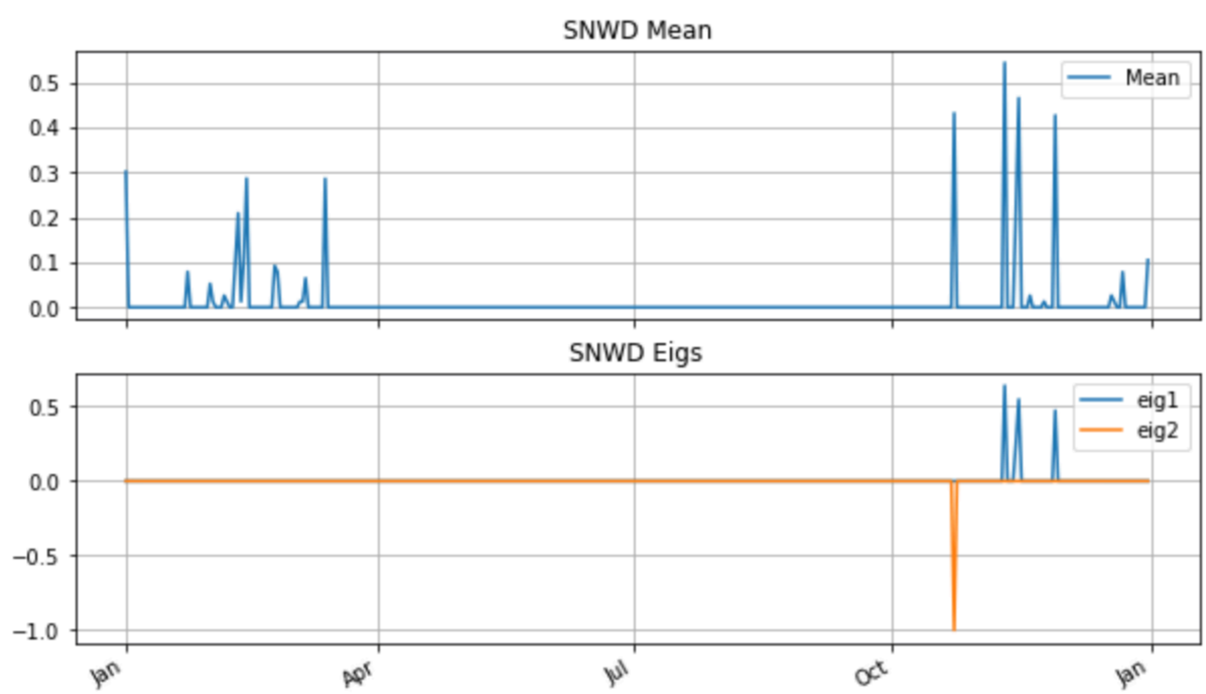


Snow Depth

Eigen-Decomposition

Due to the fact that the first two eigen-vectors explain 90% of the variance, we will analyze the eigen-decomposition for snow depth.

The graphs below show the mean of the snow depth (same as above) and the top two eigen-vectors.



We can observe that snow season is from October to March timeframe and that snow depth spikes in the October to December timeframe.

Next we look at the eigen-functions. The first eigen function (eig1) has a similar shape to the mean towards the end of the year. The main difference is that it is zero for most of the year, where as the mean is only zero in the April to October timeframe. The interpretation of eig1 is that it is the overall amount of snow above the mean.

eig 2 has one negative spike in October which is similar to the spike in the mean snow depth in October, elsewhere the eigen-function is zero. This can be intpreted as typically snow does not stay on the ground very long in October.

These two eigen-functions only account for the second half of the mean chart but they capture 90% of the overall variance. This would suggest that the amount of snow the stays on the ground fluxuated the most at the end of the year. The last 10% of the variance is explained in the other eigen-vectors, which describe how the snow depth fluxuates in the first taks of the year.

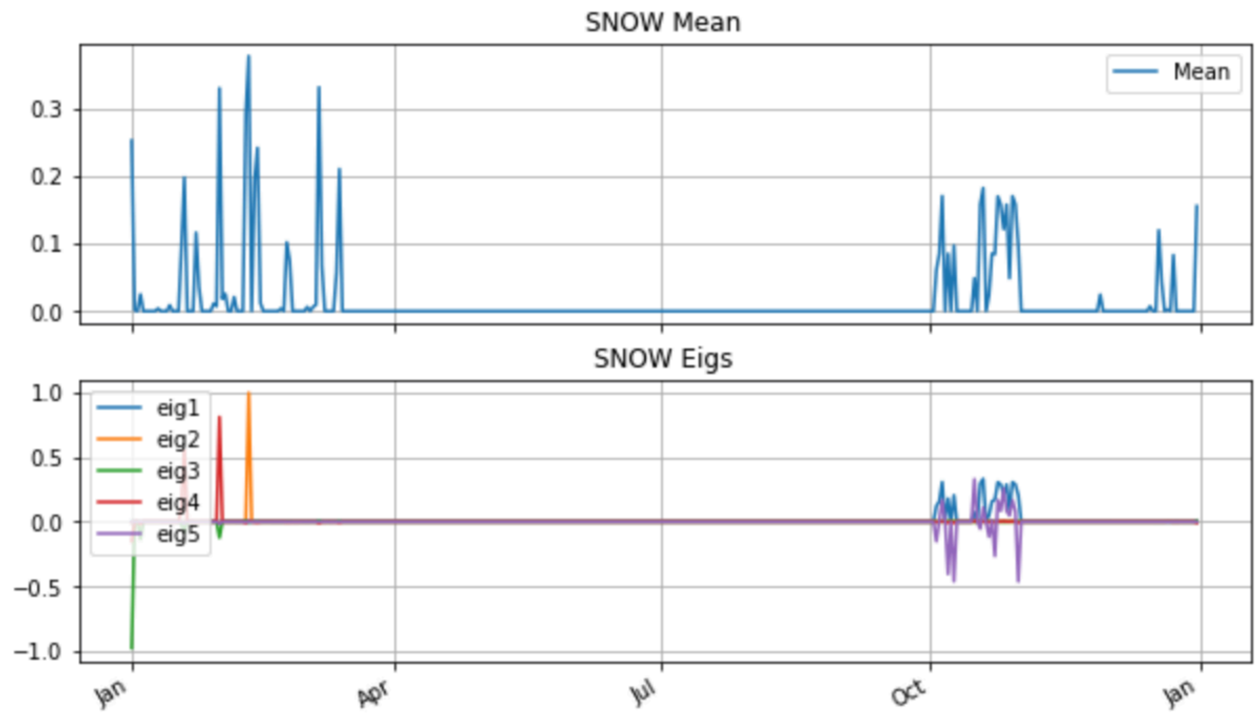
Root Mean Squared (RMS)

Since there is very little snow, it is interesting to ask if the snow depth is dependent on location or on the year. In order to do this, RMS was computed to estimate the effect of location verus the effect of time on the snow depth. The total RMS is 103.88. After subtracting the mean by station, RMS is 73.49 and after subtracting the mean by year, RMS is 84.86. This implies that the snow depth is temporal.

Snow

Eigen-Decomposition

In order for there to be snow depth, there first needs to be snow. As stated above, the first five eigen-vectors captures 75% of the variance. The graph below shows the mean of snow and the first five eigen-vectors.



The first eigen-function (eig 1) has a similart shape as the mean in the October to November timeframe which is similar to the SNWD eig 1. It illustrates the amount of snow above the mean. The next four eigen-functions capture more the variance and can be inturpretted as follows:

- eig 2: Captures the amount of snow there is above the mean in February.
- eig 3: Shows the amount of snow there is below the mean in Janurary to February.
- eig 4: Captures the amount of snow above and below the mean in Janurary to February.
- eig 5: Illustrates the amount of snow above and below the mean in the October to November timeframe.

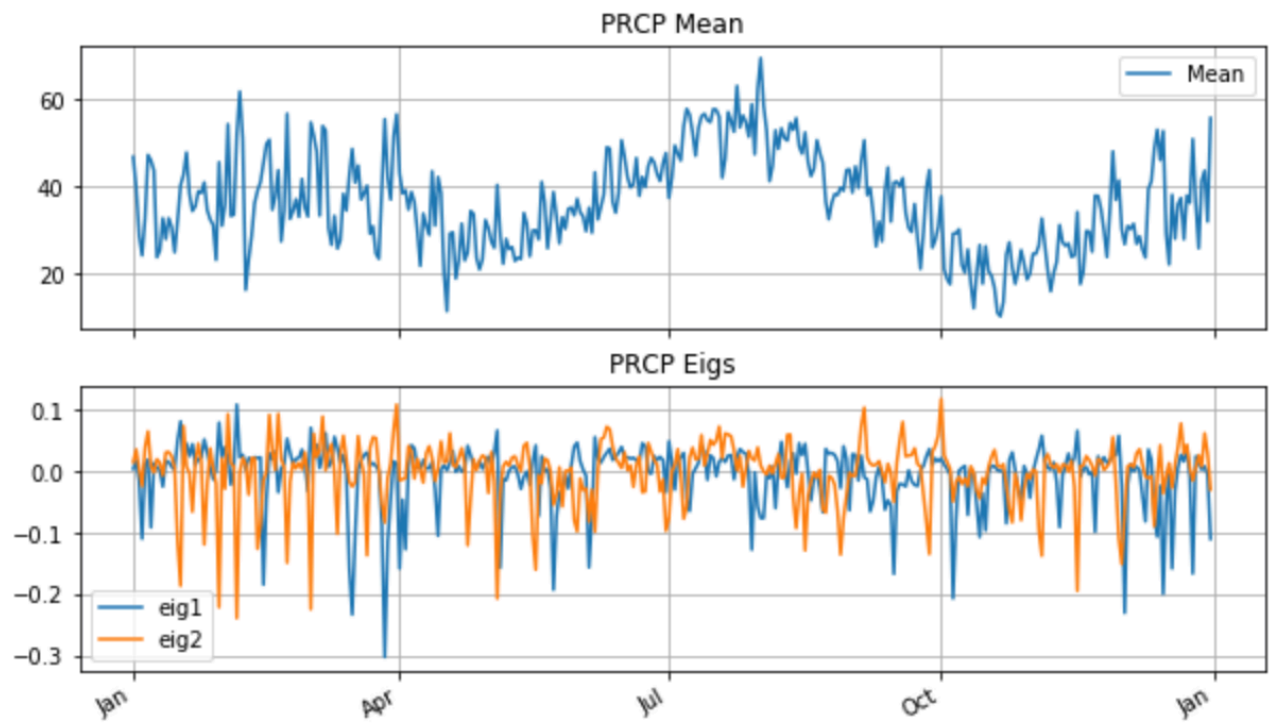
Root Mean Squared (RMS)

Since we have determined the snow depth is temporal utilizing RMS, we will determine if snow is also temporal. The total RMS is 28.76, when mean by station is subtracted RMS is 23.02, and when mean by year is subtacted RMS is 23.57. This implies that snow is temporal, but this is not a strong statement since the RMS values are extermely similar. One can conclude that other weather conditions throughout the year contibuted to the amount of snow depth since the amount of snow is neither temporal or spacial.

Percipitation

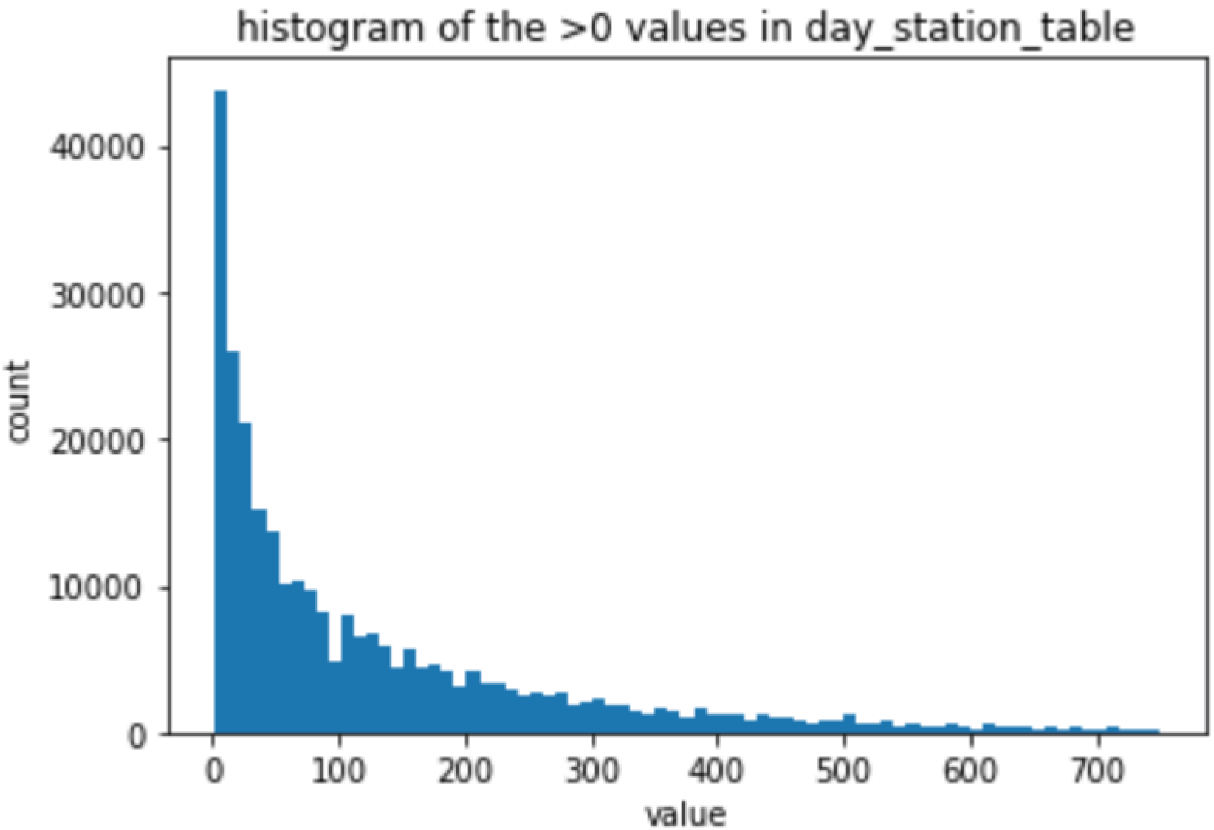
Eigen-Decomposition

There is a large amount of variance in the amount of percipitation per year in this region. Unlike Snow and Snow Depth where the first few eigen-vectors capture a majority of the variance, for percipitation, the first 10 eigen-vectors only capture just over 12% of the variance. The next image depicts the mean of percipitation throughout the year and the first two eigen-vectors. The eigen vectors illustrate the amout of percipitation above and below the mean. There is a lot of fluxuation throughout the year for the amount of rainfall.

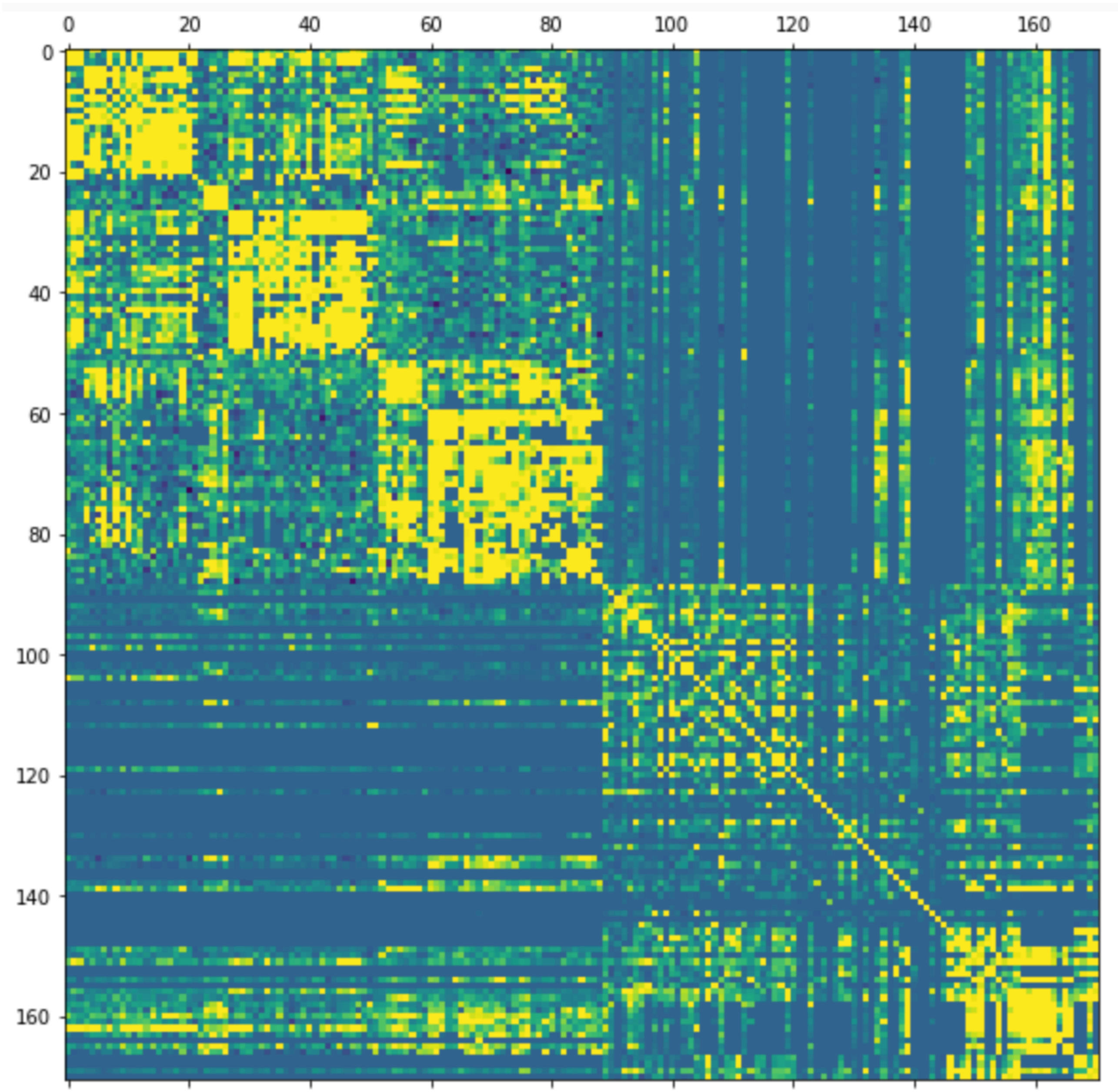


Analyzing Residuals

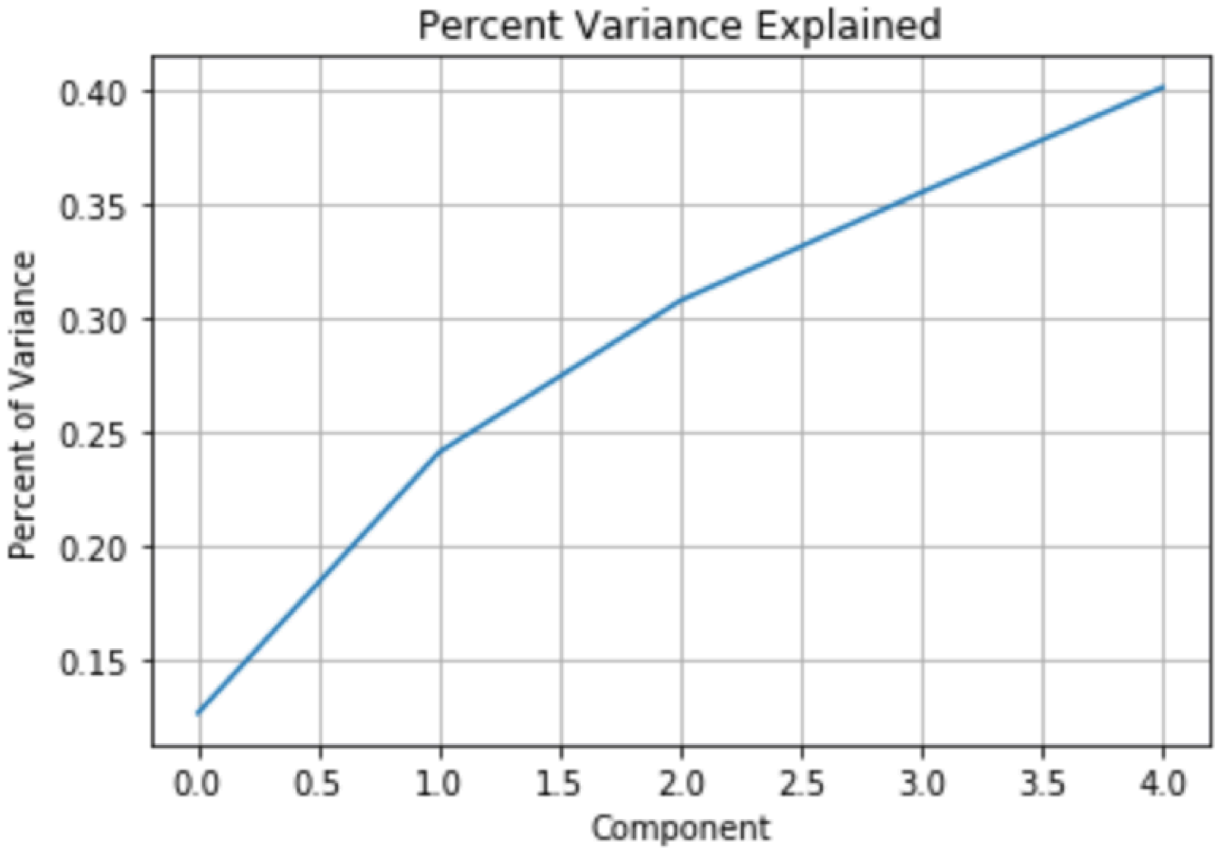
The PRCP data spans 44895 days, from 1890 to 2012. Throughout this timeperiod, there is bound to be missing data due to stations being closed down and newly created. There are 171 stations that have PRCP data for this region. The histogram below indicates the number of nonnan entries within the PRCP data set.



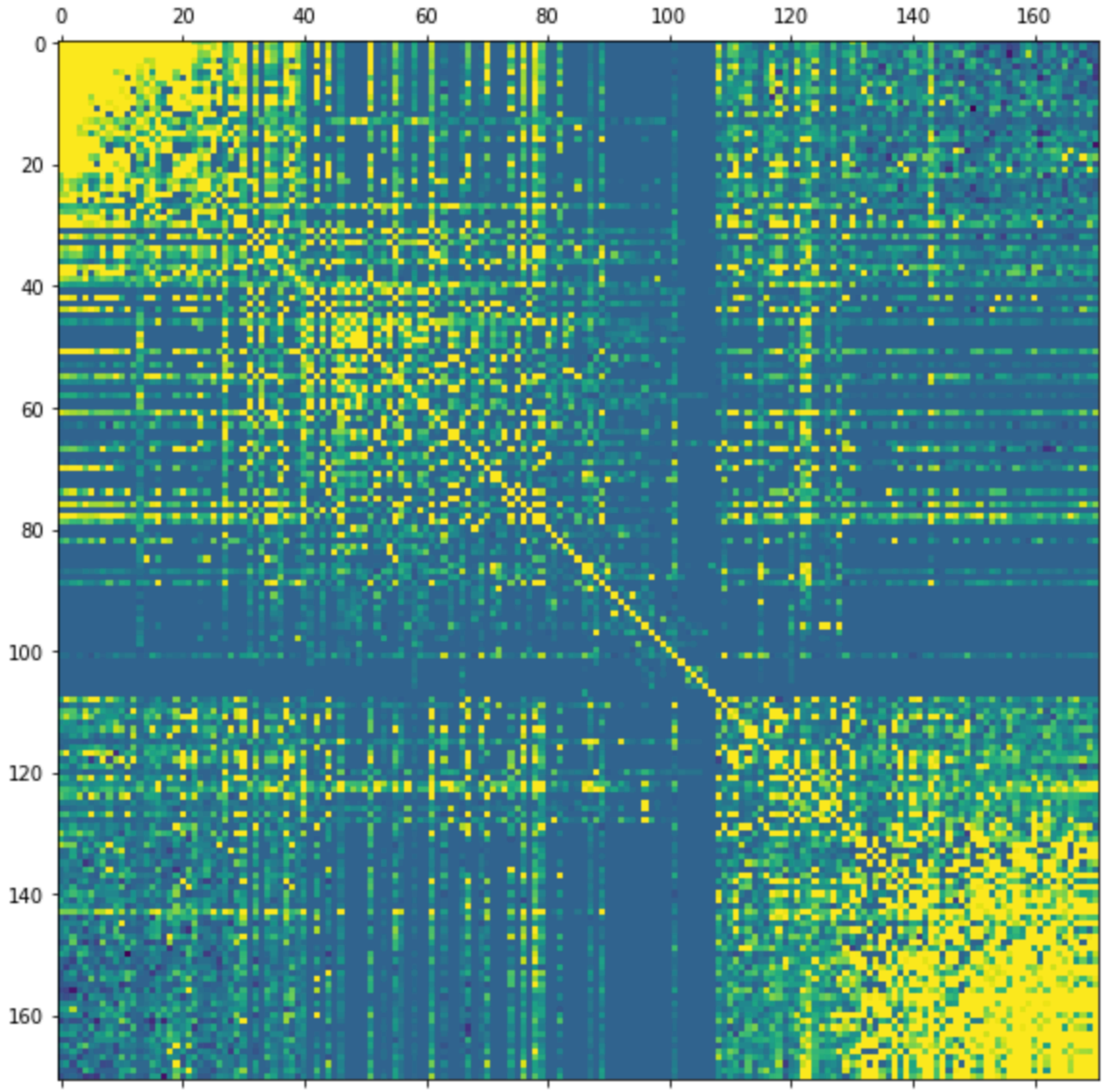
RMS was performed on the PRCP data as well. The total RMS=104.14. RMS minus the station mean is 97.53 and the RMS minus the year mean is 73.54. This implies that percipitation is spatial, unlike snow depth. The graph below indicates the probability that two stations will experience rain on the same day. Blue indicates low probability and green indicates high probability.



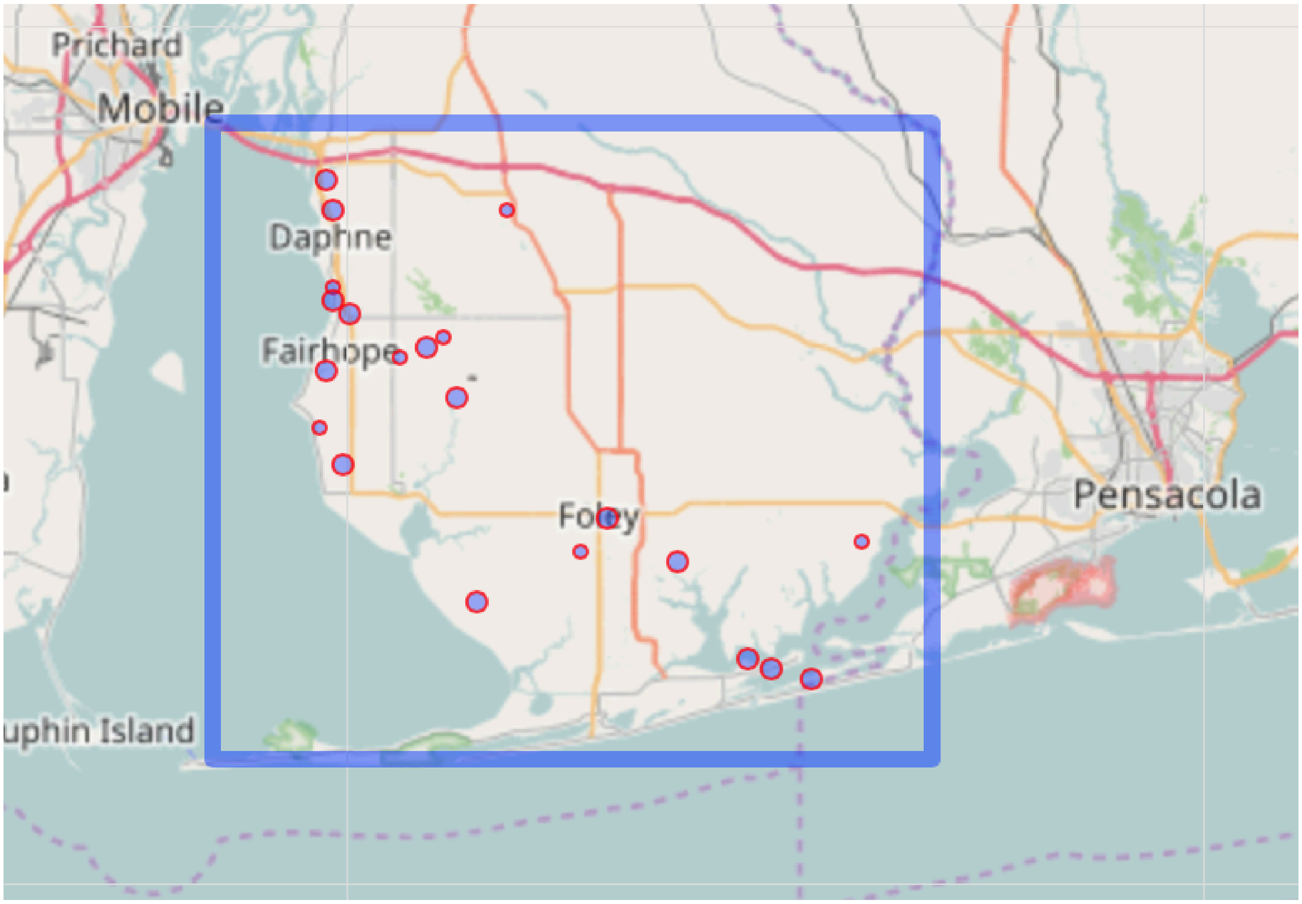
In order to investigate further into the probability of two stations experiencing rain on the same day, PCA analysis was performed. For the PCA analysis 5 components were chosen. The PCA was done on the maximum covariance data which is depicted above.The graph below shows the percentage of variance that is explained by each component.



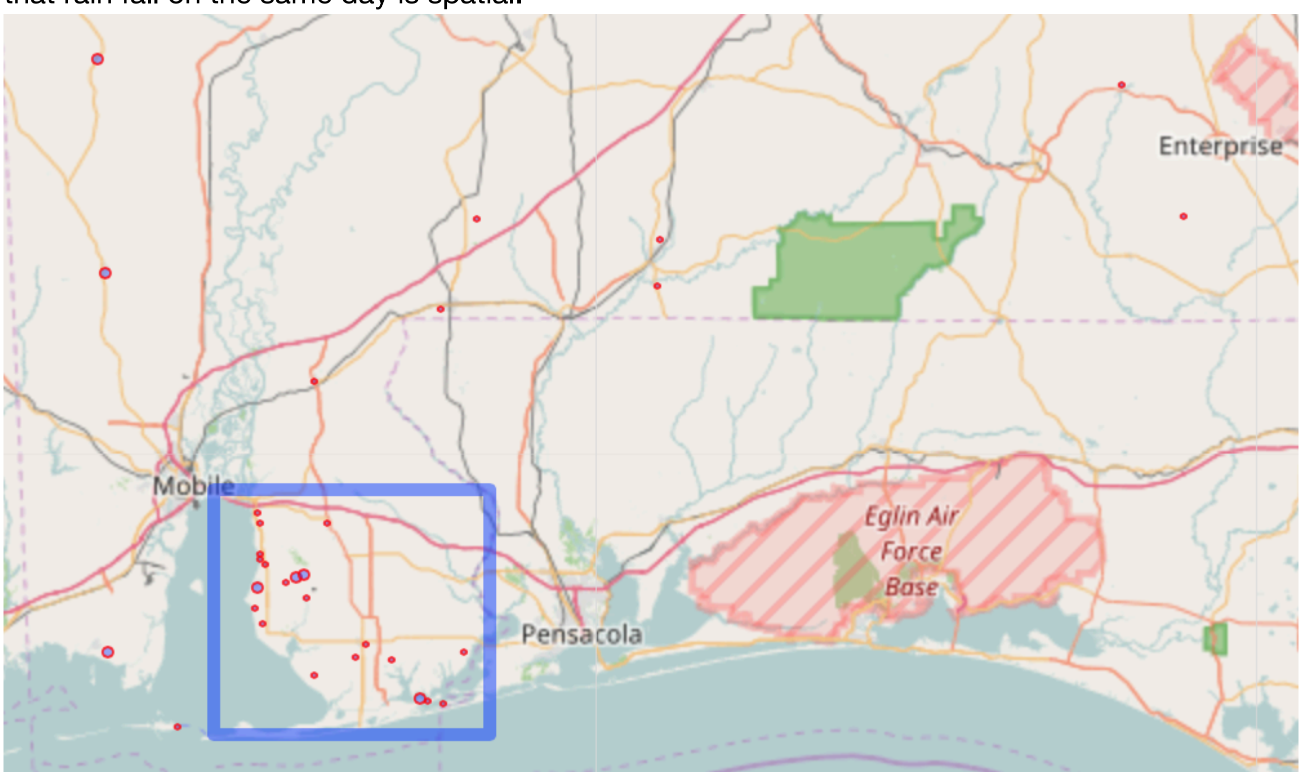
When the components are ordered and the covariance between two stations is graphed, we get the following image. This image is different than the one above due to the fact that the PCA analysis is capturing more of the variance than above.



This image indicates that stations 0-20 have a high probability of raining on the same day. In order to see if stations 0-20 are located in the same area the next image was created.



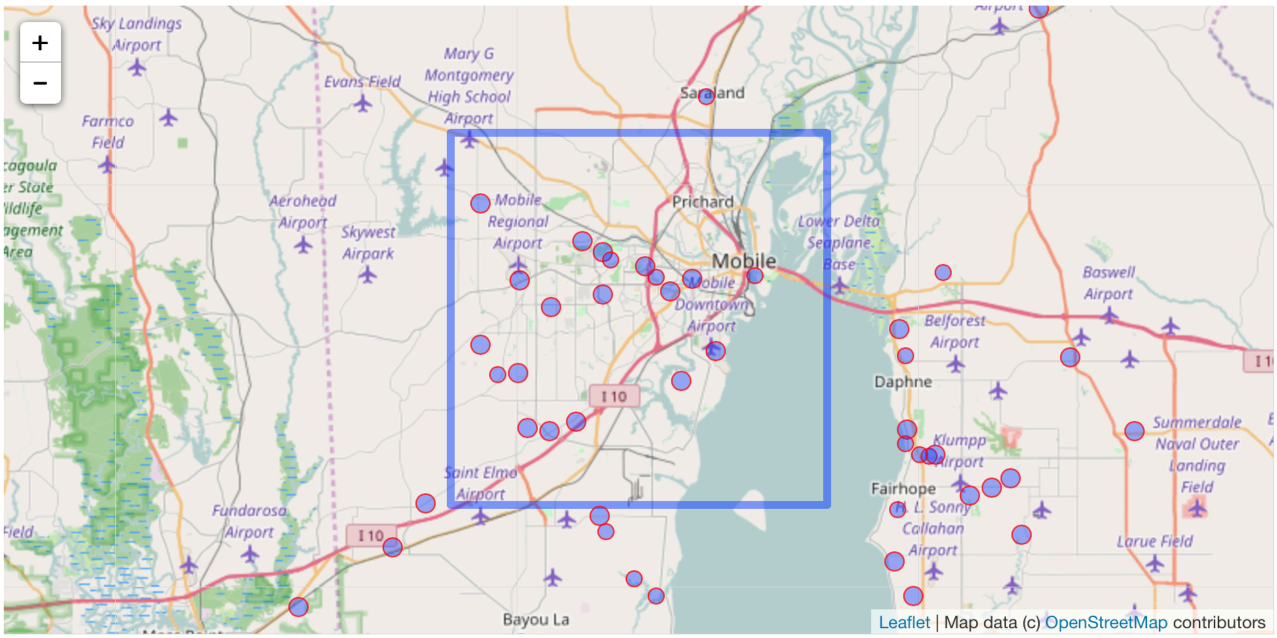
As one can tell the stations are in the same area, supporting the fact that rainfail is spatial. In order to further confirm this, let's look at the location stations where there is little percentage they will experience rain on the same day. The location of stations 90-100 were examined. The image below shows the 0-20 stations in the box and all the circles outside the box are the 90-100 stations. The 90-100 stations are some distance away from the first 20 stations supporting that rain fall on the same day is spatial.



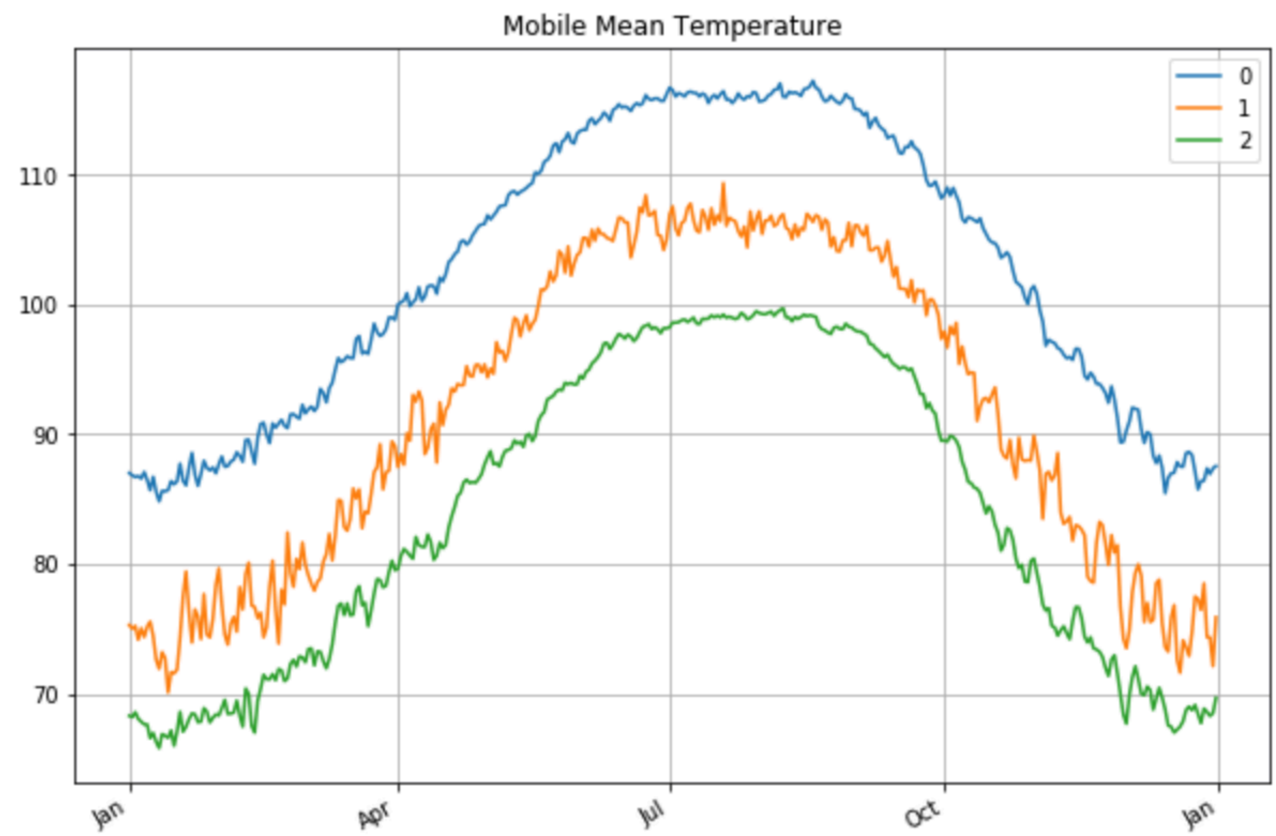
Temperature

Mobile

A box has been created within the region, where Mobile, AL is at the center. This box contains 20 weather stations that will make up Mobile's temperature analysis, the center of Mobile is at the center of the box. The region is depicted below.



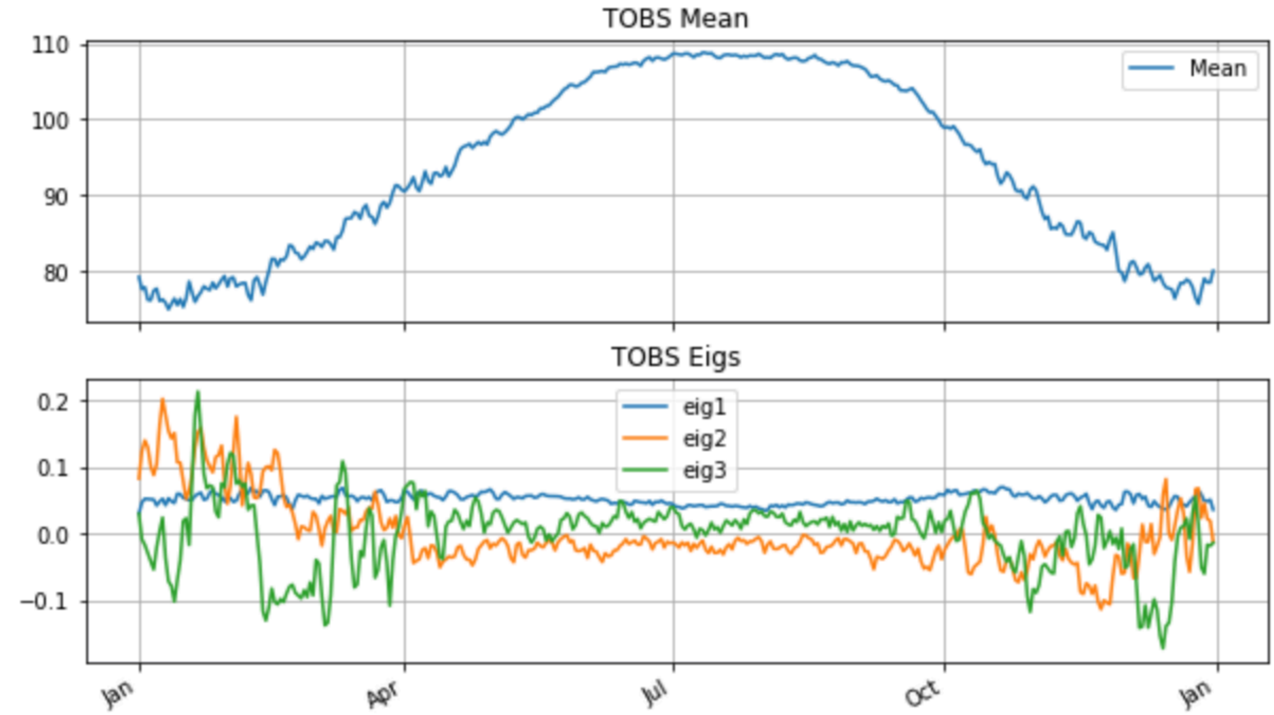
The average TMIN (green), TOBS (orange), and TMAX (blue) were calculated for Moible. Below shows the results of those calculations.



These temperatures do not match the US Climate Data. The average daily temperature within this data set is 91 degrees but the US Climate Data indicates that 91 degress is a maximum. The dataset that is being examined spans 1948-2012. The US Climate data for Mobile spans 1981-2010 which is a shorter timeframe.

TOBS

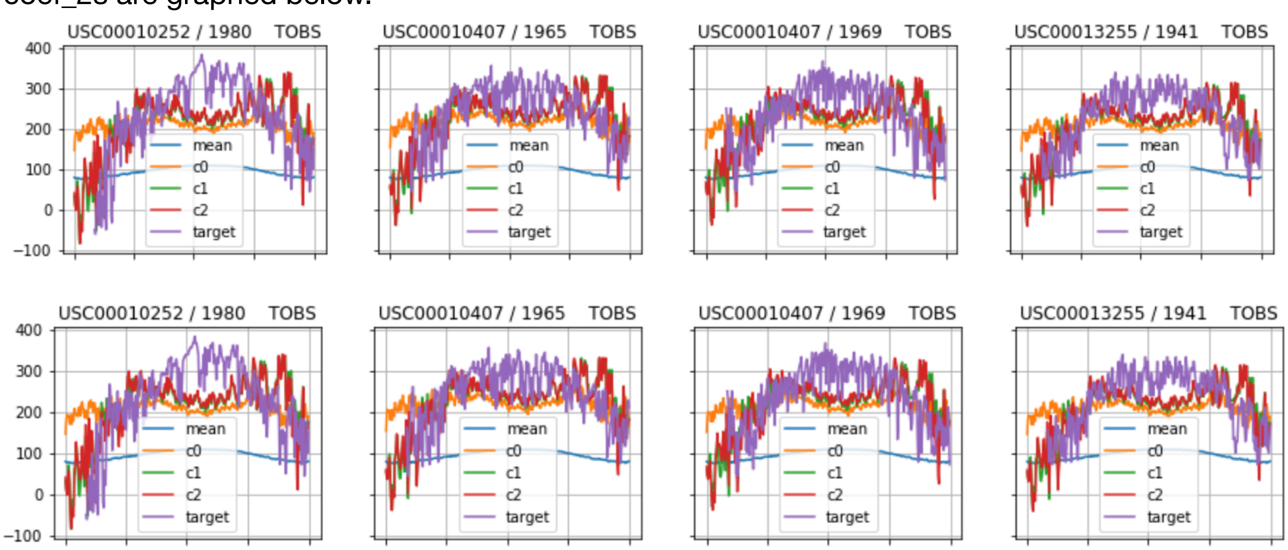
In order to look closer at temperature, we will focus on TOBS since it is a average temperature each day. The figure below shows the mean and the first three eigen-vectors of TOBS.



The first three eigen-vectors only account for 40% of the variance for TOBS. All of the eigen-vectors show the difference above or below the mean the average daily temperature fluxuates. In the summer, the first three eigen-vectors are below the mean. This means that there is some variance in the temperature in the summer. This can account for the differences between the TOBS tempature for the region and the high temperatures for Mobile and Pensacola reported by US Climate.

Analyzing Residuals

The residual_mean, residual_1, residual_2, and residual_3 were calculated for the TOBS dataset. With a positive and negative coeficients, the reconstruction of TOBS is pretty close to the target. In order to get the temperature in Celsius, divide by 10. The four most negative and positive coef_2s are graphed below.



Consistently, the temperature is in the summer is less than the target value with a postive and negative coeficient.

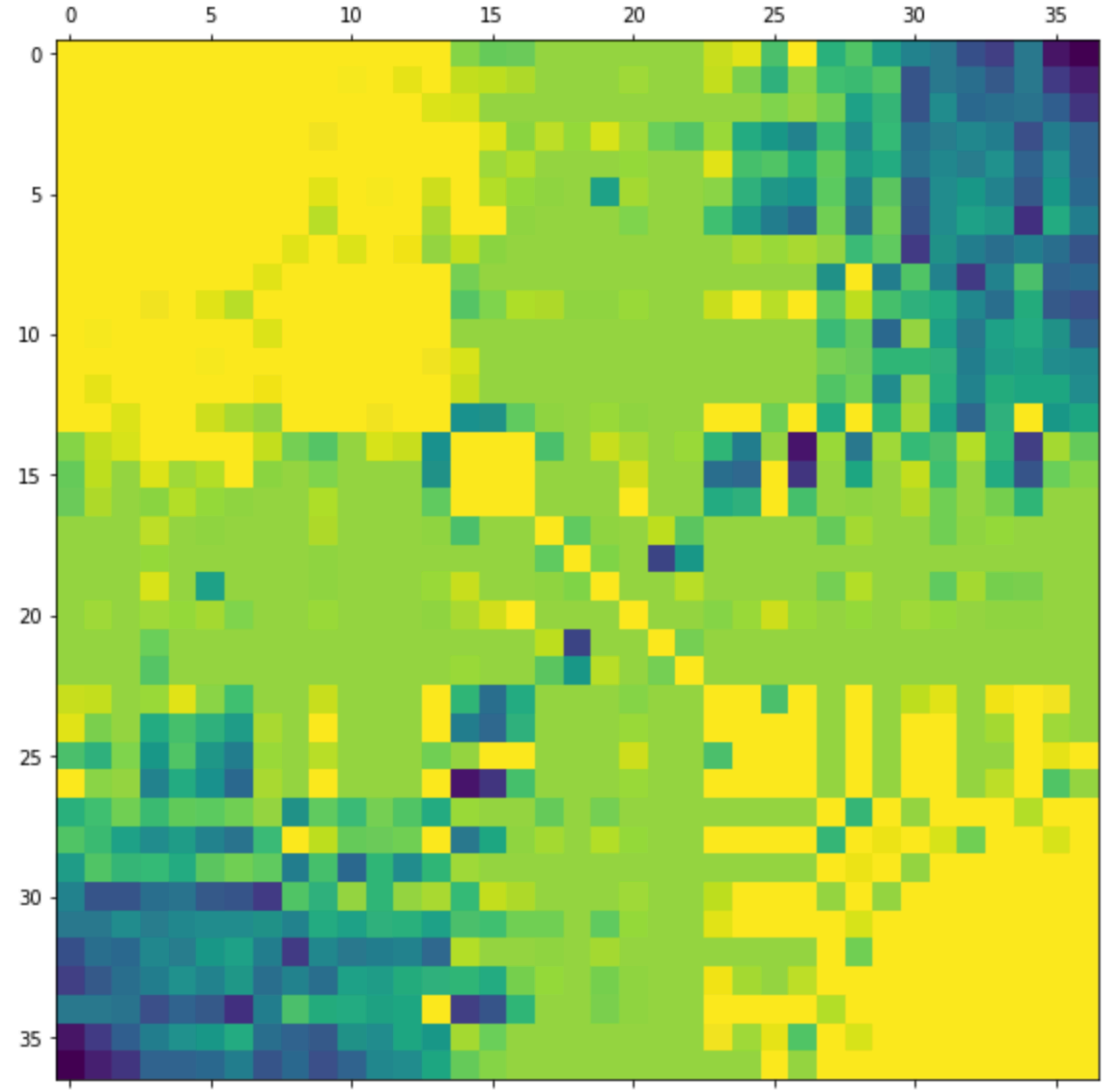
The best reconstruction is pictured below.

The residual normalized norm after mean is 0.299237955917 and the residual normalized norm after mean + top eigs is [0.36746426 0.36913429 0.31848336].

PCA

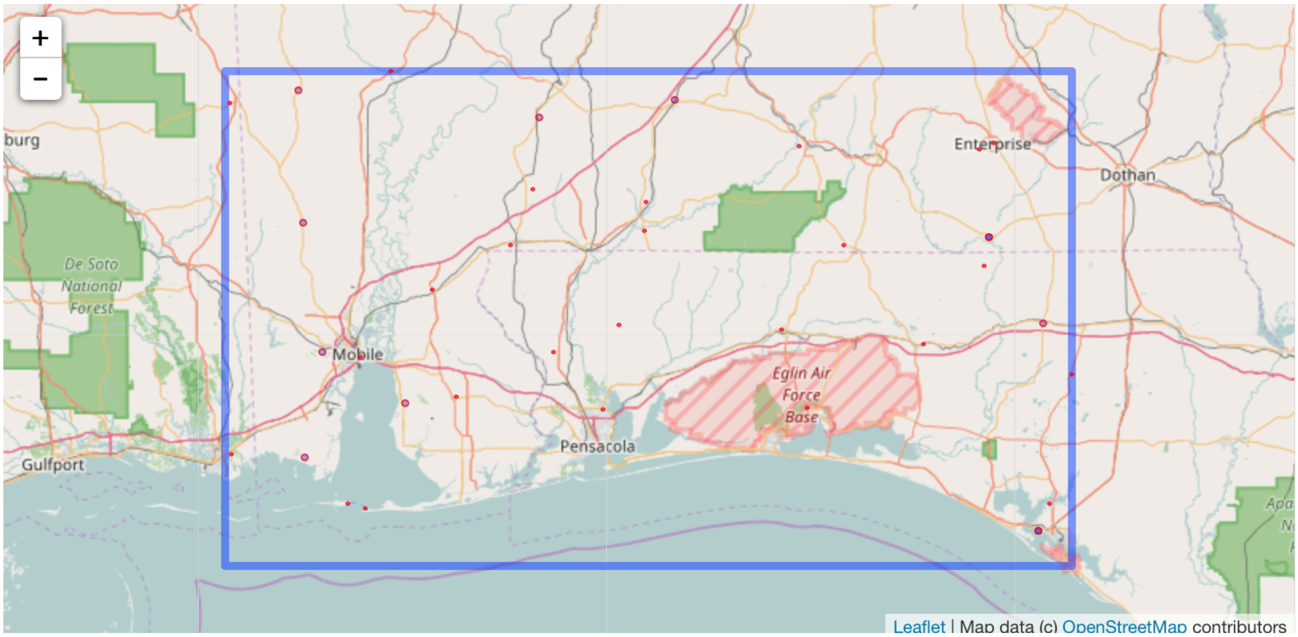
After looking at the residuals for the data, the TOBS data was narrowed down to the stations where the residual mean is less than 1. This data was used for the PCA.

For PCA, 5 components were chosen to explain 80% of the variance within the TOBS dataset. When looking at the covariance matrix below, it is clear that the average daily temperature is highly coorelated between two stations.

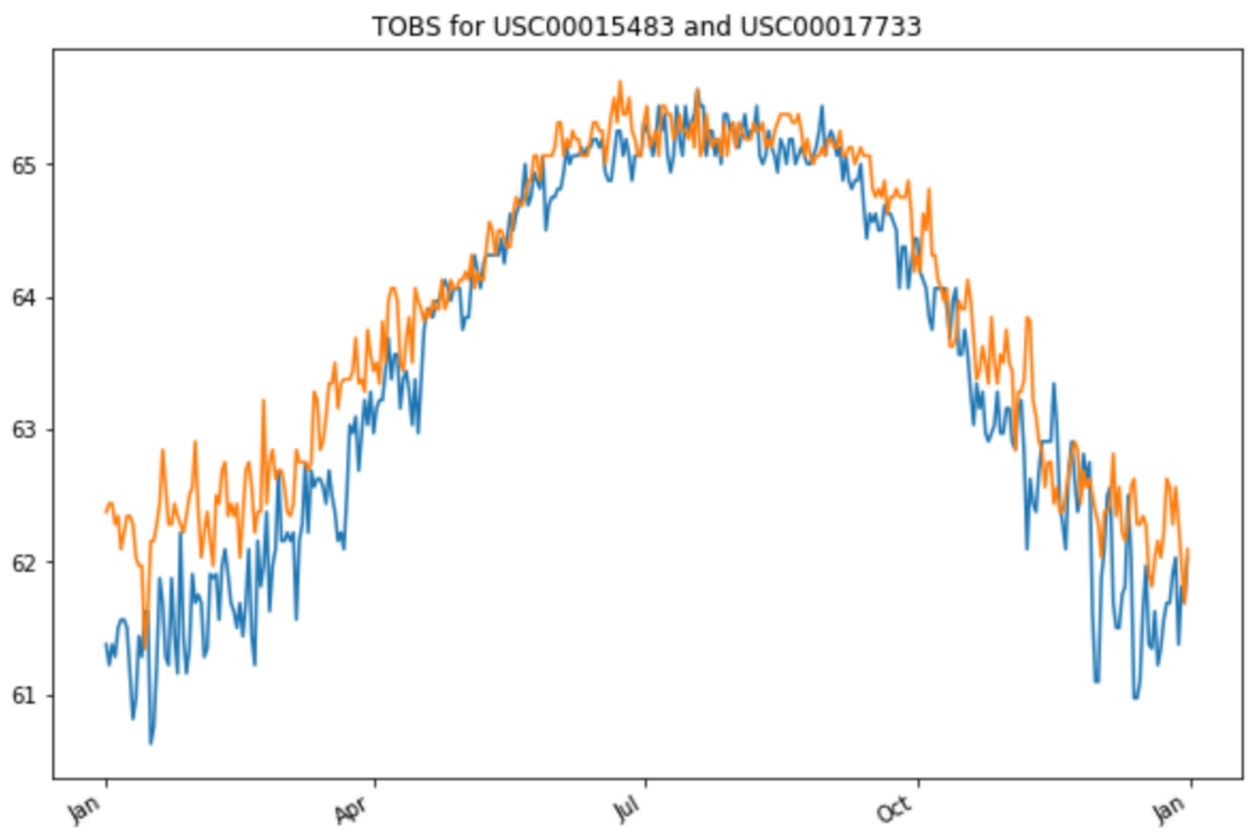


In order to verify this, we look at the RMS. The total RMS is 197.12 and the RMS by removing the mean by station is 79.61. The RMS by removing the mean by day is 34.43 which impies that stations within the are area are more likely to have the same temperature. This supports the coveriance matrix.

The map below shows the stations that the TOBS data is from for PCA. Even though the stations are not geographically close together, the probability that the temparature will be the same is high. This means that the temperatures across the region are pretty consistent.



There are only two stations that would contribte to Mobile's temperature report and one that would contribute to Pensacola's. This along with the longer timeframe of capturing the data could contribute to the differences between the data that is currently being analyzed and the US climate data. Below shows the average daily tempterature (TOBS) for the two Mobile stations.



This is more consistent with the average temperature for Mobile. Utilizing the data where the residuals are less than 1, is a better representation of the data.

Conclusion

This region of the United States has distinct weather. There is very little snow and hence snow depth in this region. This finding makes sense due to the location of the region, on the Gulf. The temperature throughout the year is too warm for it to snow. There is a significant amount of rain throughout the year. The percipitation data has a significant amount of noise. The temperature of the region gets warmer in the summer which is reflected in the data. There is a high probability that two stations will have the temperature. Those stations are dispersed throughout the region implying that the temperature is the same throughout the region.