Lizzy Compton
Dr. Long
MAT 375
4/29/2018

Atakpame Final Report

Set-up

Preparing the Data

To set up the data that I would be using for the final models, I first had to create a dataset that included all the variables, meaning I needed to add in the ENSO and SST data to the Complete_Clean data. To do this, I merged the files by decimal year. Next I subtracted 1961 from the decimal year so that our data would begin at year 0. Lastly I eliminated any outliers and split the data into three datasets - Min, Max, and Rainfall.

```
#Merge the files together based on Decimal year
FinalTogo <- merge(FinalTogo, EnsoSST)
FinalTogo$DecYear <- FinalTogo$DecYear -1961
#Make Min and Max Columns
Min <- FinalTogo %>% filter(MinOutlier=="No")
Max <- FinalTogo %>% filter(MaxOutlier == "No")
Rainfalldata <- FinalTogo %>% filter(MaxOutlier== 'No') %>% filter(MinOutlier== 'No')
Rainfalldata= Rainfalldata[complete.cases(Rainfalldata$Rainfall),]
```

Creating Predicted Values

In order to make creating the graphs of the residuals easier in the future, I created a predicted value for all of our data points using the built in linear model predict function within R. I then created a new column and added these values into the three datasets.

```
summary(rainfallLinear)
#add a predicted column to each of the files
Min = cbind(
  Min,
  minPredict= predict(minlinear, newdata = Min))
Max= cbind(
  Max,
  maxPredict= predict(maxlinear, newdata = Max))
Rainfalldata = cbind(
  Rainfalldata,
  rainPredict = predict(rainfallLinear, newdata = Rainfalldata))
```

Filtering Data for Atakpame

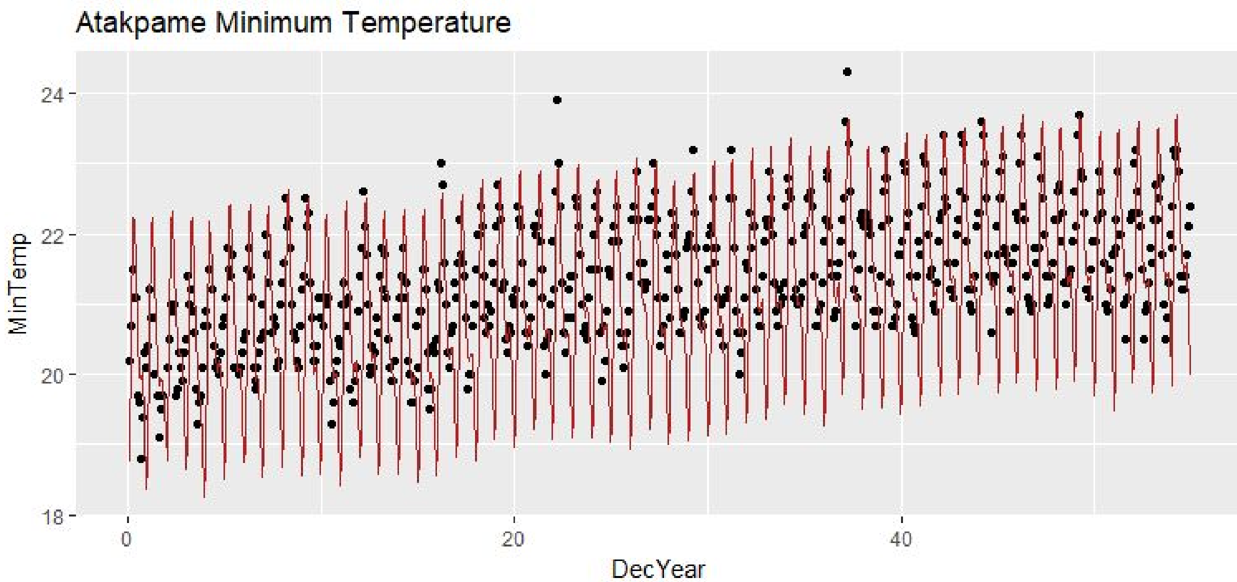
Lastly, since I just wanted to look at the data for Atakpame, I used the “filter” function to only select data points in Atakpame within each of the three datasets and created new datasets named AMin, AMax, and ARain.

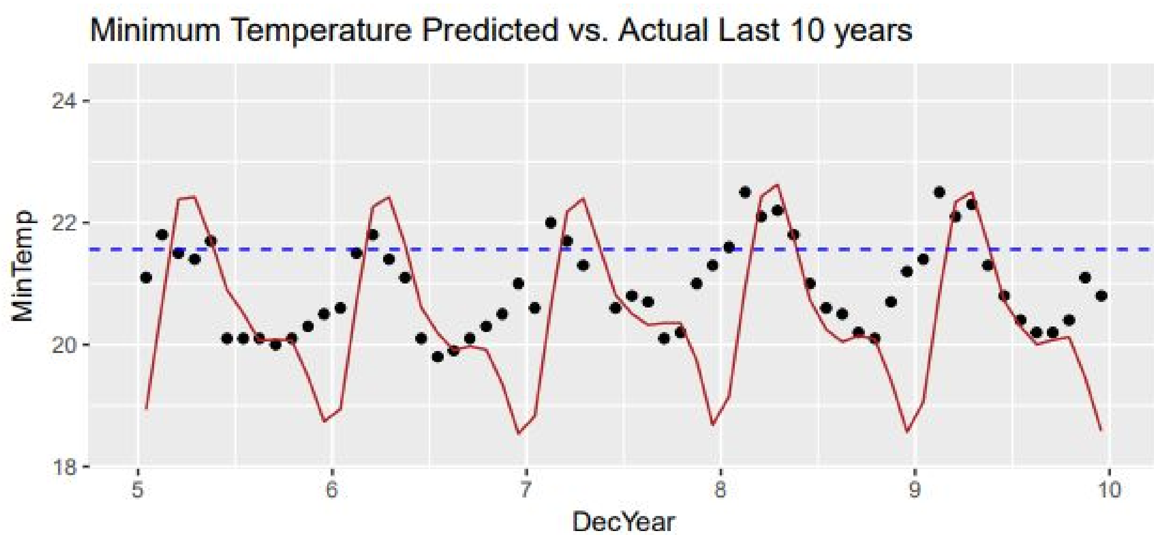
```
#Filter each of the files to just our city
AMin = Min %>% filter(location =='Atakpame')
AMax = Max %>% filter(location =="Atakpame")
ARain = Rainfalldata %>% filter(location == "Atakpame")
```

Minimum Temperature Model

Analysis

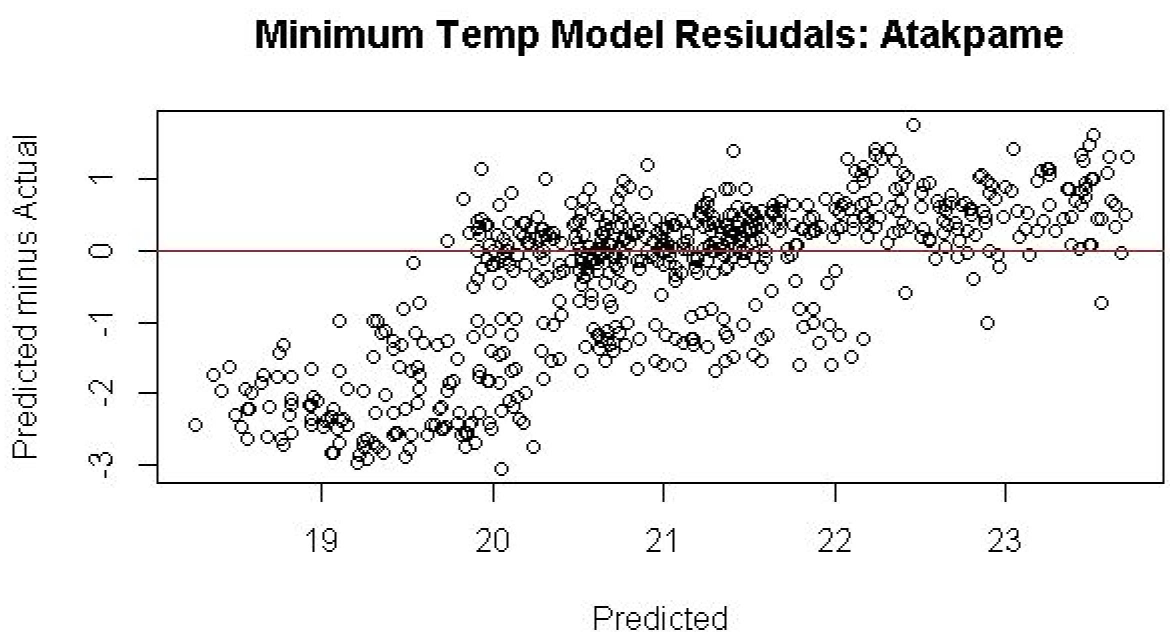
Looking at the plot of the minimum predicted values vs. the actual values for Atakpame, it appears that the red lines (predicted) and the black dots (actual) line up together and that the model fits the overall upward trend of the data. However, when you zoom in on the graph and look at a ten year period of the model and focus on later half of the period, you can clearly see that the model does not fit the oscillations of the data.

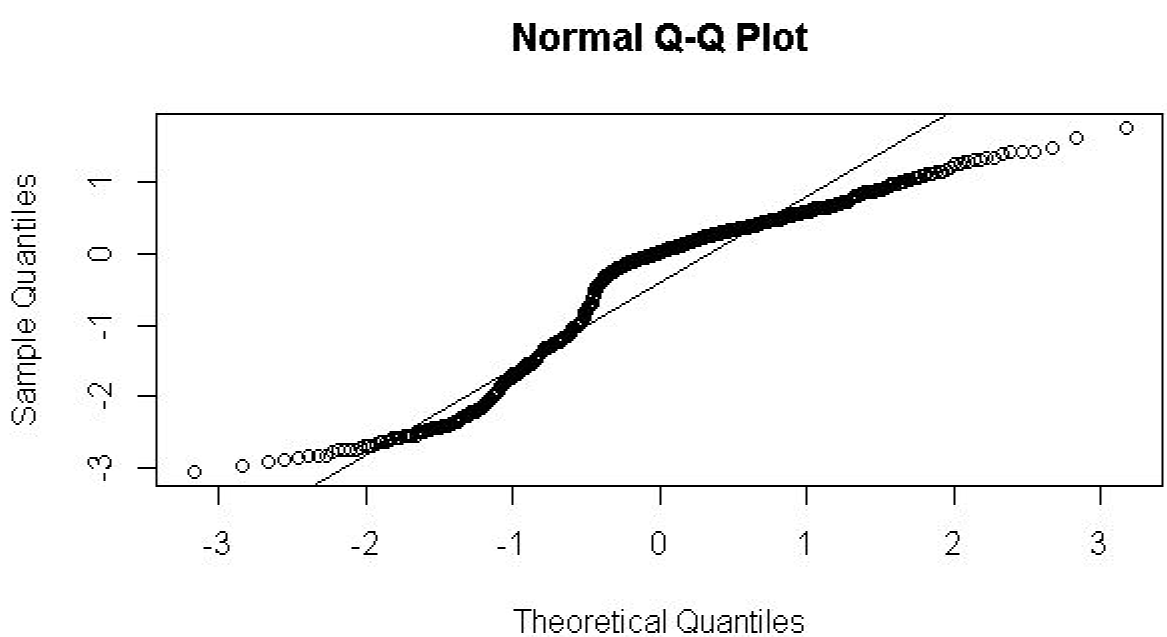




Residuals

The residuals for this model are quite concerning. There is a definitive pattern among the residuals and they appear to have an upward trend. Additionally, the QQ-plot for the residuals strays from the normal line multiple times and has an S shape. This could be attributed to the fact that it appears that our model fits northern cities better than southern cities like we discussed in class since Atakpame is located in towards the southern part of the country.

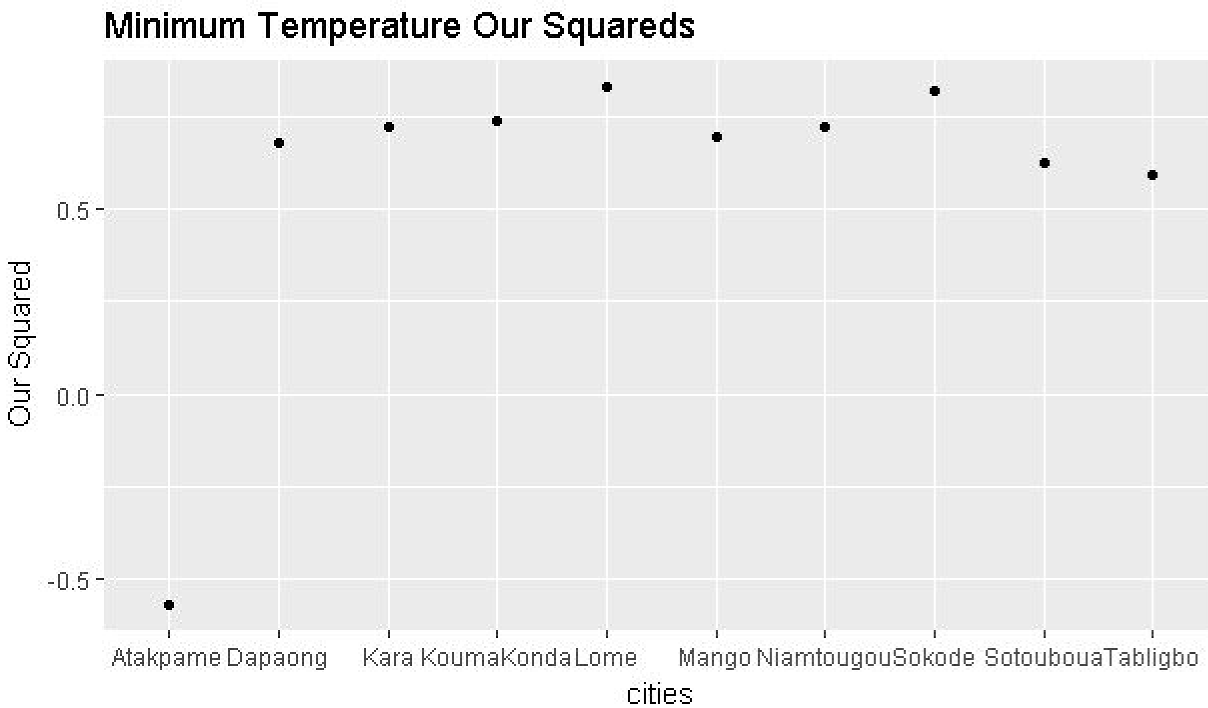




“Our”-Squared Value

$$R^2 = 1 - \frac{SSE}{SS_{Tot}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

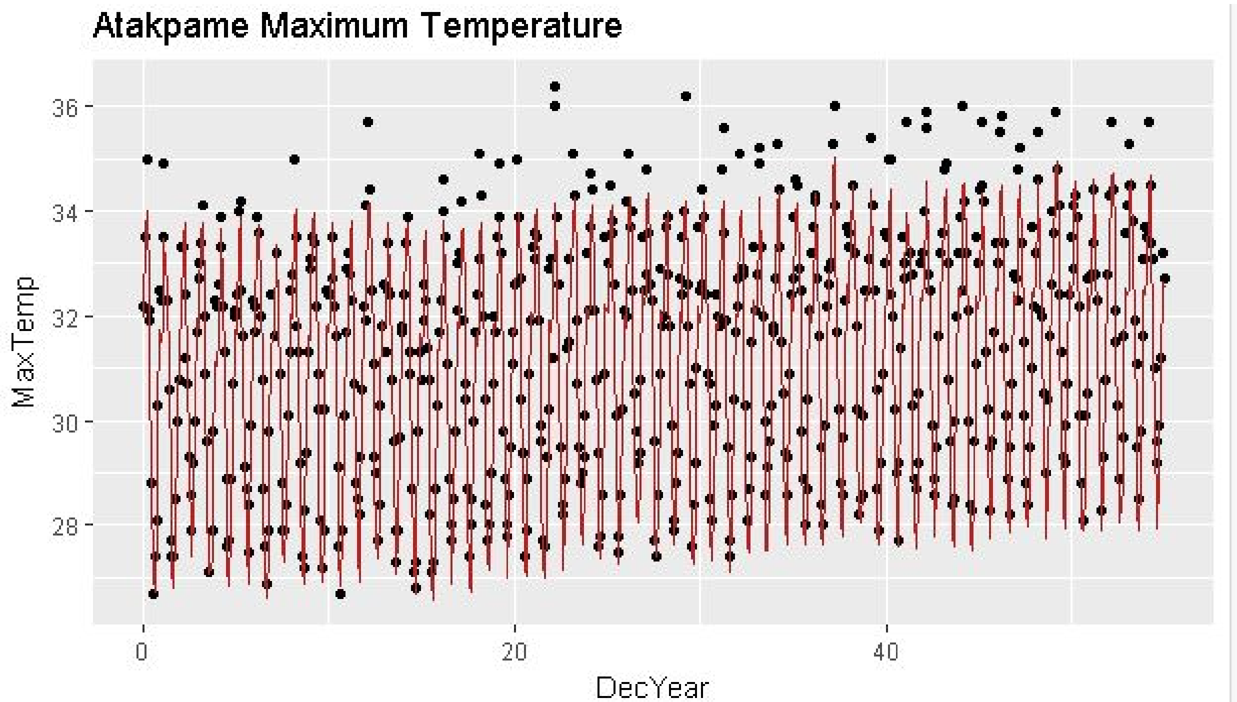
This is the equation that we came up with in class and it is how we will evaluate the model. The value for the minimum model is -0.5714952. Since our function allows us to have negative values, this is not an entirely bazar number however it does indicate that the model does not fit the data well. A negative value indicates that mean better predicts our city data than the model does. When we compare this value to the values from the other cities, we see that Atakpame is the only city with a negative “our”-squared value and that all others are above a .50. Thus Atakpame falls outside the norm of the model.

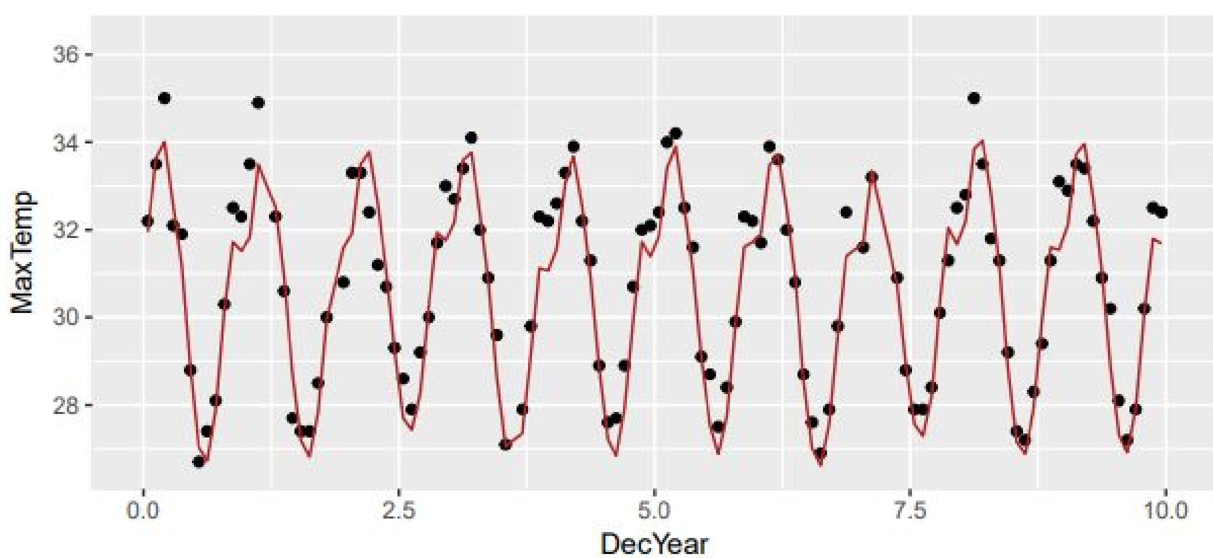


Maximum Temperature Model

Analysis

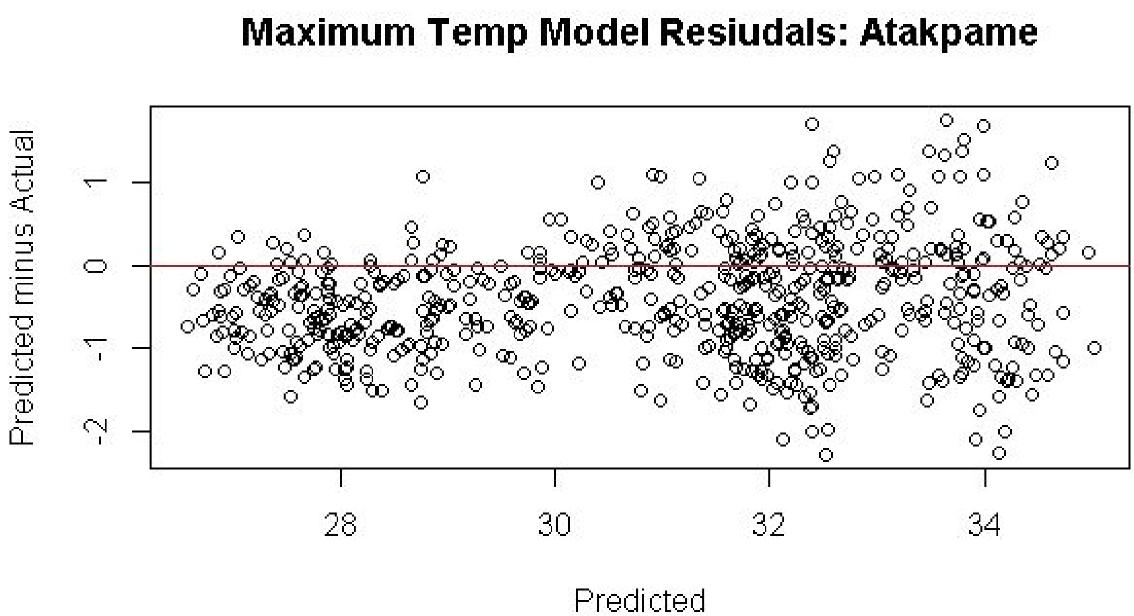
Looking at the plot of the maximum predicted values vs. the actual values for Atakpame, it appears that the model fits the data fairly well with a few outliers at the top of the model. Looking more closely at the last ten years and zooming in on the plot, we can see that the model does a good job of capturing the oscillations in the data. Both the overall trend of the data and the yearly oscillations are fit well with this model.

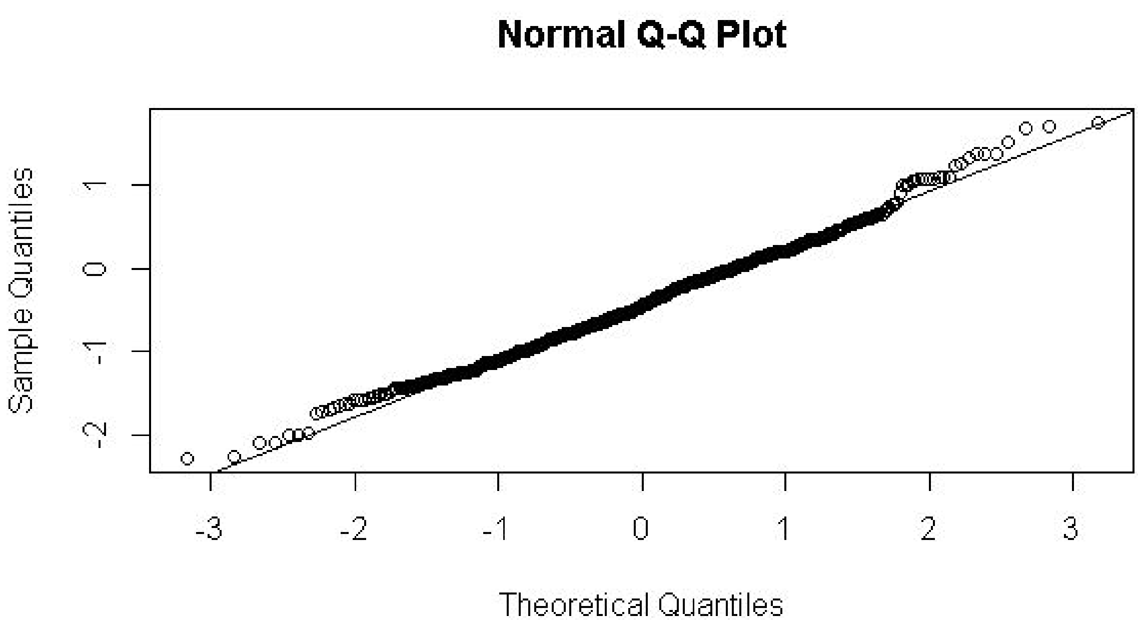




Residuals

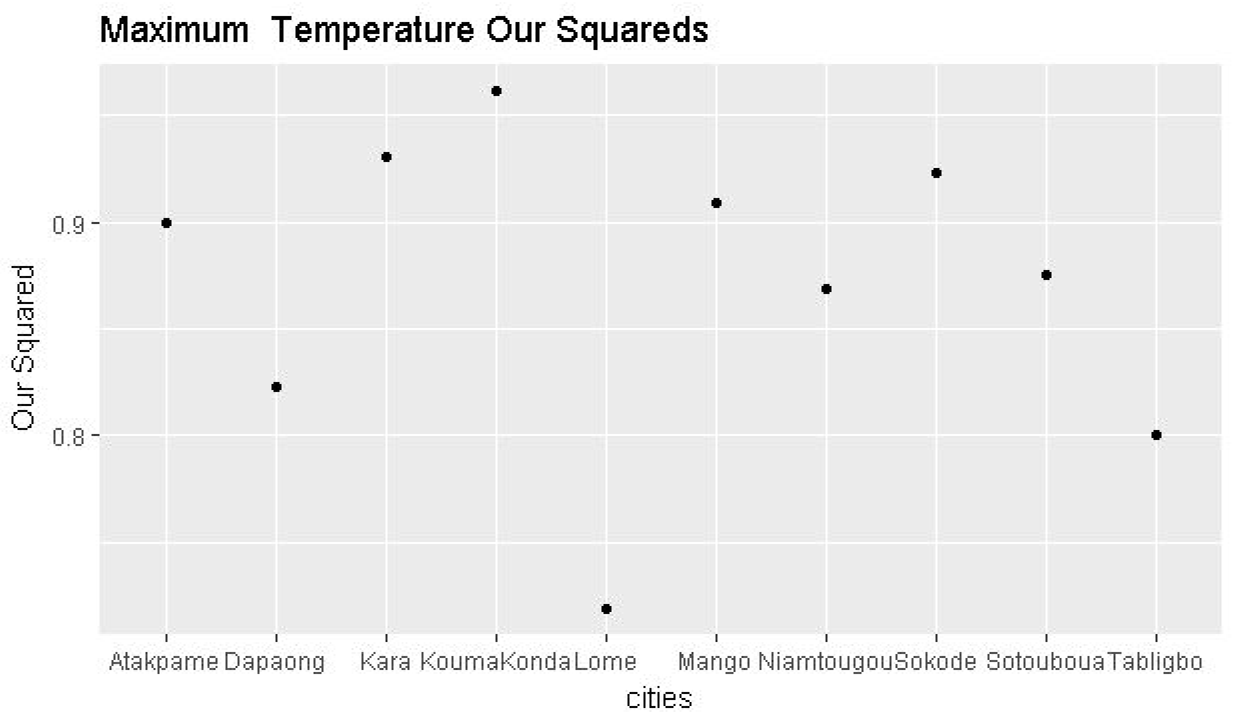
The residuals for the maximum model are much better than the minimum model. There is no clear pattern to the residuals and overall they appear be independent and identically distributed. There might be a slight microphoning effect where the residuals are more spread out and farther from zero towards the end of the graph but that doesn't appear to be a big concern. Additionally, the QQ-plot for the residuals follows the normal line with only slight departures at the beginning and end of the graph. Overall there is not much concern with the residuals for the maximum model.





“Our”-Squared Value

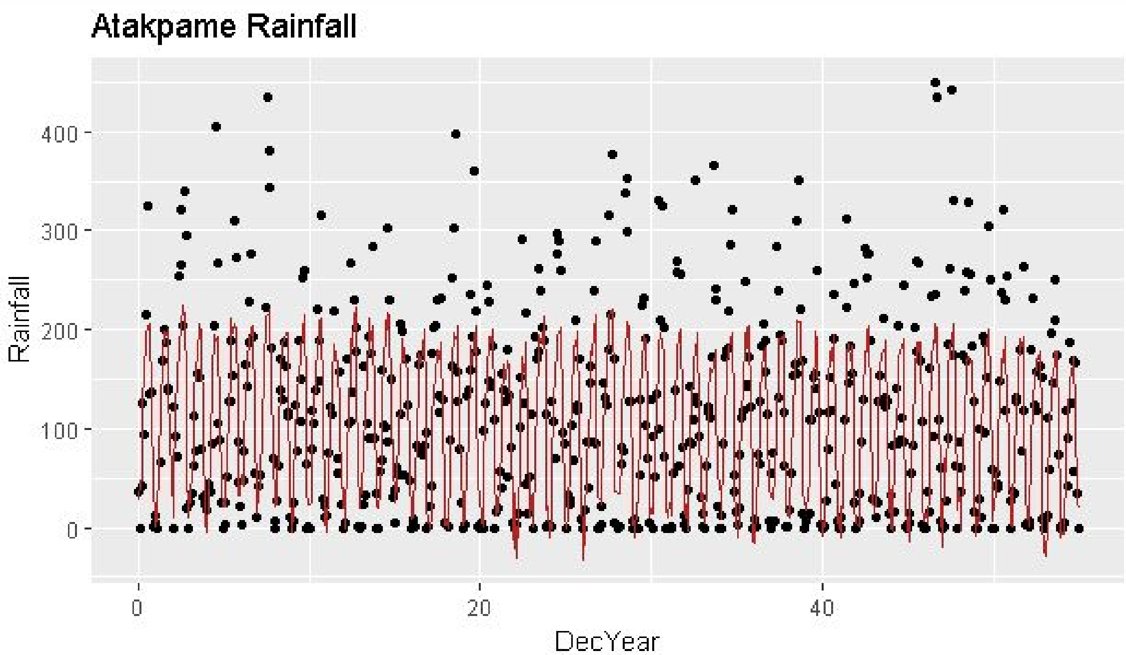
Again we will use the “our” squared equation from class to evaluate the model. The value for the maximum model is 0.8994645. This indicates the 89.94% of the variation within the data is accounted for by the model. When we compare this value to the values from the other cities, we see that Atakpame is fairly central among the other cities. Additionally, since all the city values are above .7, we see that the maximum model is a good fit for all of the cities.

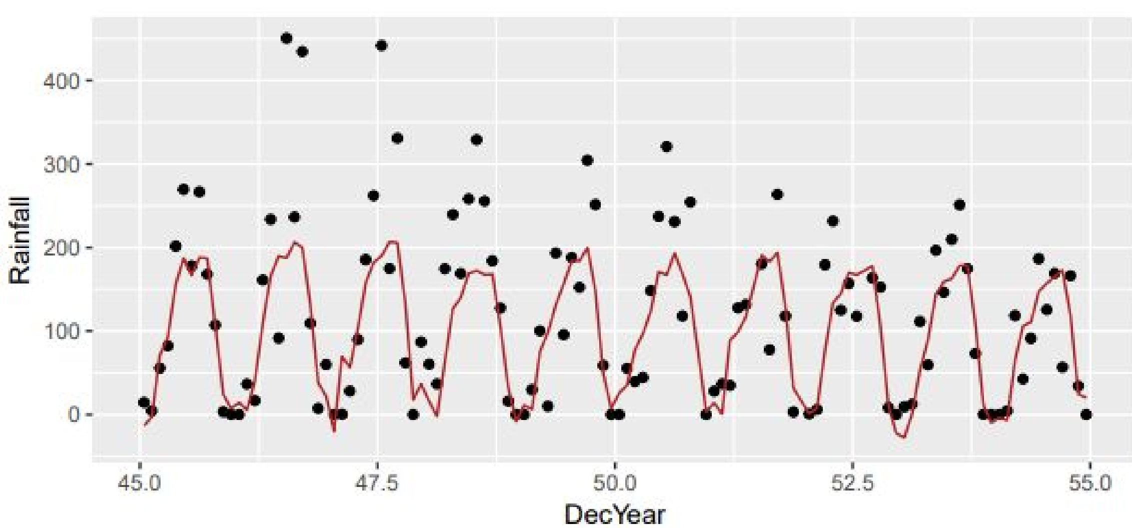


Rainfall Model

Analysis

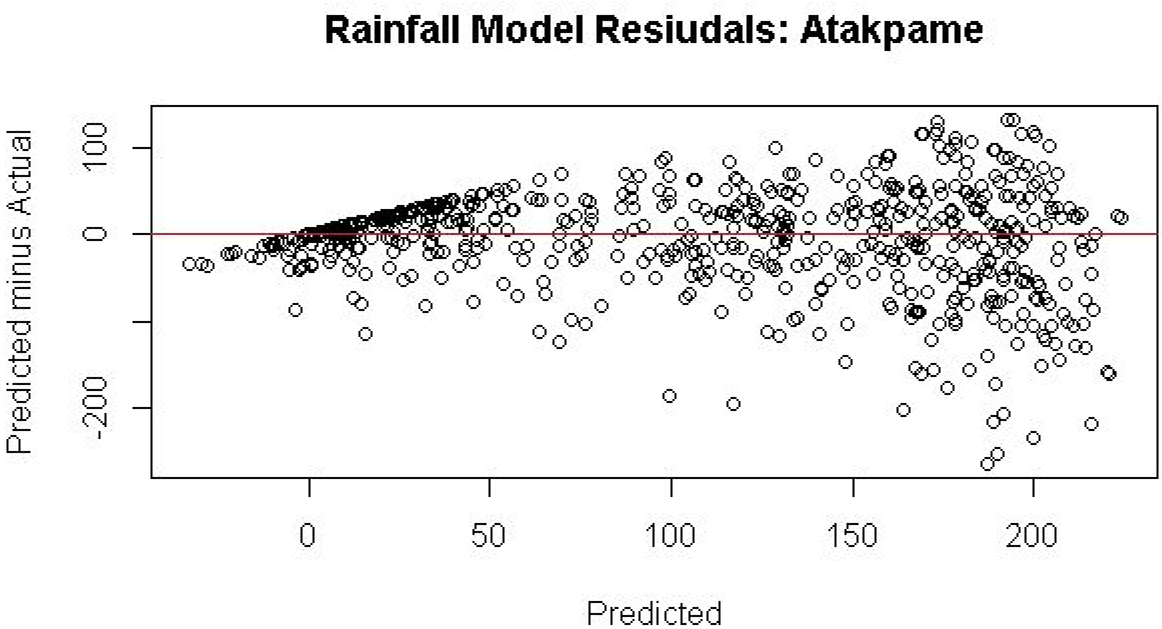
Looking at the plot of the rainfall predicted values vs. the actual values for Atakpame, it appears that the higher rainfall points are not captured by the model. Looking more closely at the last ten years of the data, we can see that the model does an ok job of fitting the oscillations in the data but again the larger values are not captured. Since there is a lot of variation within our data it makes sense that the model struggles to fit the data.

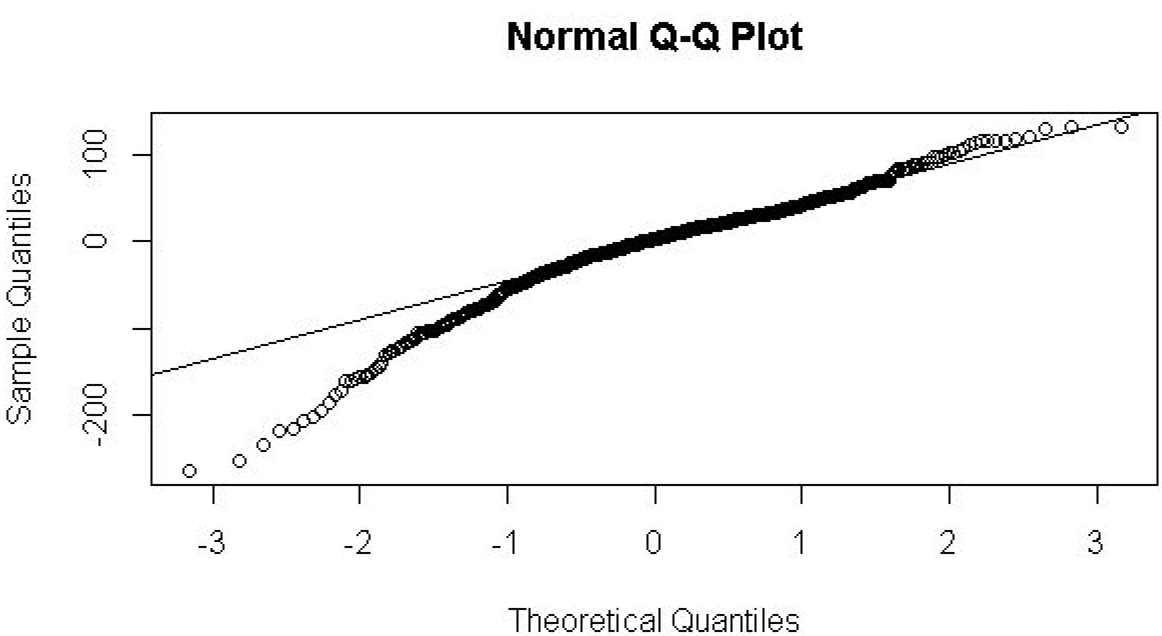




Residuals

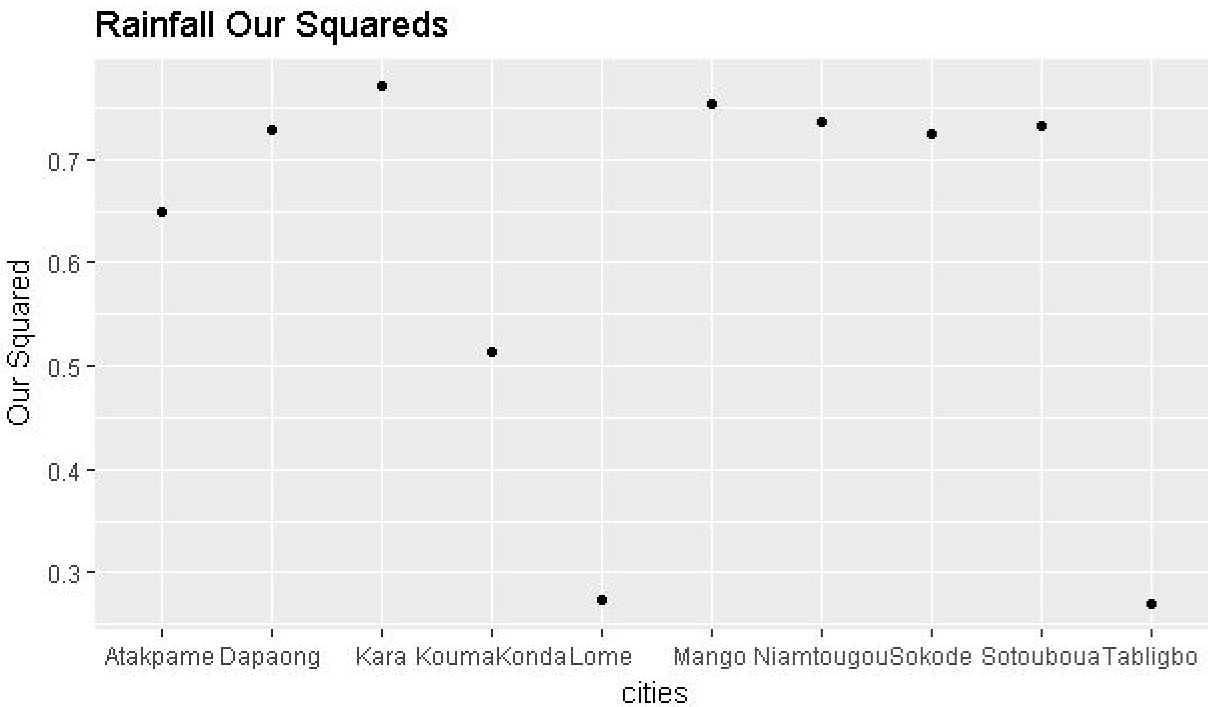
The residuals for the rainfall model are not normal but that makes sense given how many zero values there were in the data.. There is a clear microphoning effect within the residual plot as the residuals are closer to zero at the beginning and then more spread out towards the end of the graph. Additionally, the QQ-plot for the residuals departs from the normal line in the beginning and then gets better at the end of the graph. Overall though the residuals are concerning.





"Our"-Squared Value

Again we will use the "our" squared equation to evaluate the model. The value for the rainfall model is 0.6500426. When we compare this value to the values from the other cities, we see that Atakpame falls in the middle of the other cities. Since all of the city values are between .3 and .8 the model could definitely be improved upon but considering the variation in the data, it does fairly well.



Conclusion

The minimum model did not fit the data for Atakpame well at all. While the model captures the overall increasing trend of the data, it does not fit the yearly oscillations. Additionally, from the "our" squared value we see that the mean model would do a better job of predicting values than the model. The maximum model did much better. It had a high "our" squared value of .8994 and this was consistent with all of the other cities which all had "our" squared values above .7 indicating that the maximum model was a good fit. Lastly the rainfall model did an average job of modeling the data but was worse when it came to modeling the extreme data points. This was to be expected as there was a lot of variation in the model as well as a lot of zero values.