

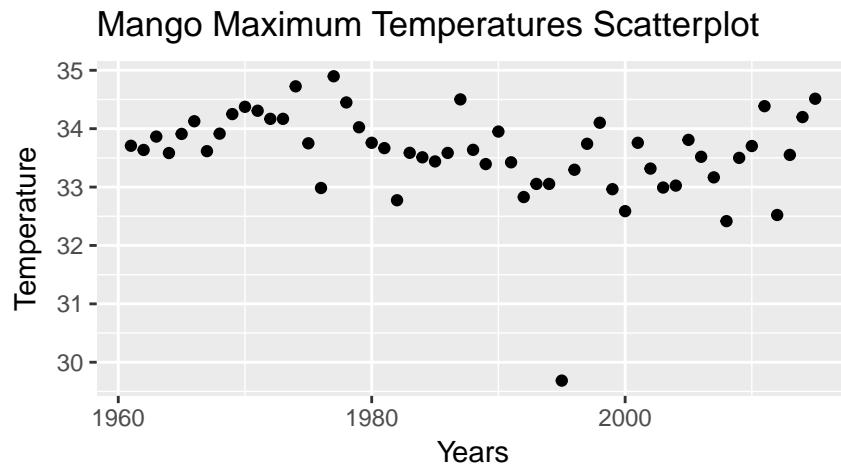
# Mango Temperature

Allison Zembrodt and Austin Hardesty

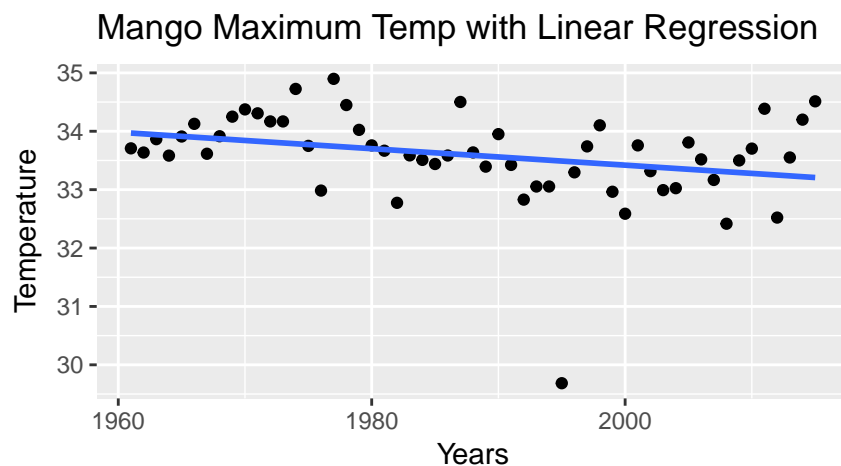
1/26/2018

In this report, we will be using linear regression to determine if the city of Mango demonstrates significant increases or decreases in temperature over time. To begin, we read in the temperature data files and stored them in two separate variables, MangoMax for the maximum temperatures and MangoMin for the minimum temperatures. The first half of this report, we will discuss the maximum temperatures and the results of our linear regression and residuals. The second half will focus on the minimum temperatures. The first graph in this report, is a simple scatter plot of the annual maximum temperatures.

## Maximum Temperatures



From just looking at the points, the graph seems to be fairly consistent, with maybe a slight downward trend. There is one clear outlier of ~29 degrees in the year 1995. Other than that point, all other temperatures seem to fluctuate between 32.5 and 35 degrees celsius, with a mean of 33.6. To investigate the possible downward trend further, we ran a linear regression and added it to our scatter plot.



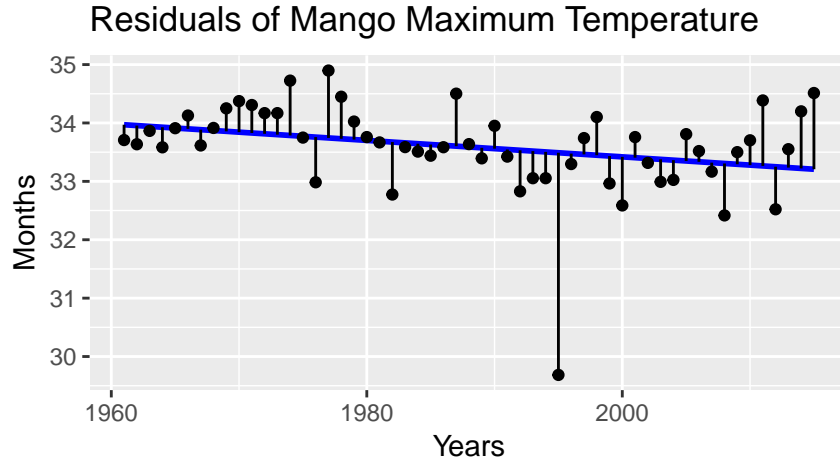
The r-squared value from the linear regression is only .0849. Since this is so low, we are not confident in this model. We graphed the residuals in the following plot, just to see the variations in the points. At this point,

Table 1: Maximum Temperature Linear Regression Analysis

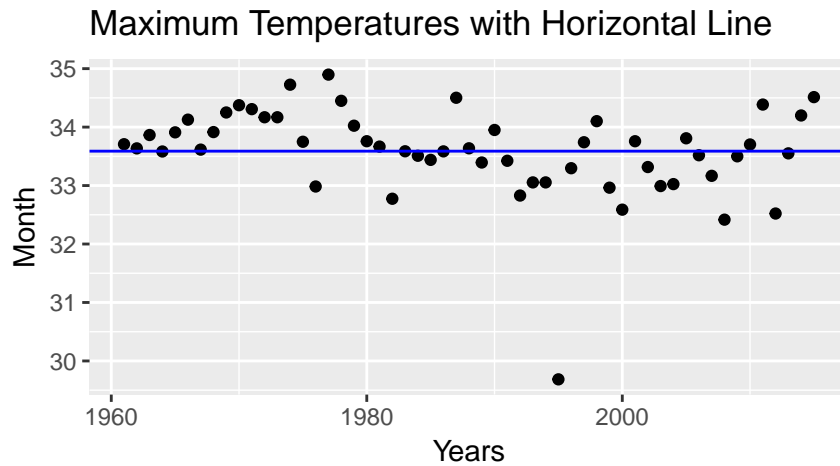
<i>Dependent variable:</i>	
MOY	
ANNEES	-0.014** (0.006)
Constant	61.628*** (12.645)
Observations	55
R <sup>2</sup>	0.085
Adjusted R <sup>2</sup>	0.068
Residual Std. Error	0.749 (df = 53)
F Statistic	4.917** (df = 1; 53)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

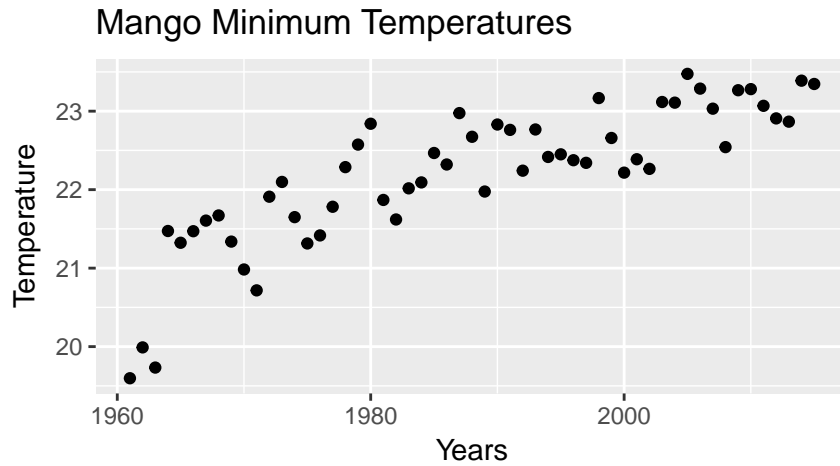
it is clear that linear regression is not the best option.



From the summary of the linear regression and further examination of a graph of the residuals, we do not feel comfortable saying that there is evidence that the temperature decreased. Since the r squared value is so low, we do not feel like there is evidence that the coefficient is different than 0. Therefore, we do not believe that the maximum temperature demonstrates a significant decrease. The next graph, is of what we feel is the best model for the maximum temperatures, a horizontal line through the mean.



# Minimum Temperatures



Just by looking at the minimum temperature points, it is clear that there is some type of upward slope. To further investigate, we added a linear regression line to the scatterplot.

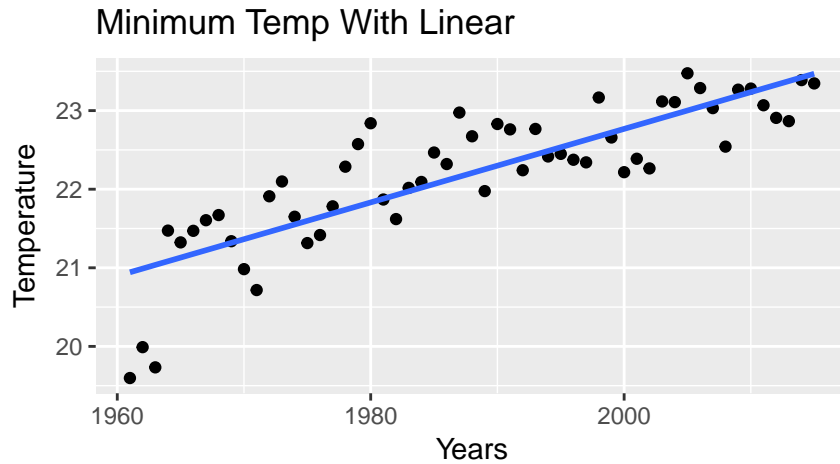
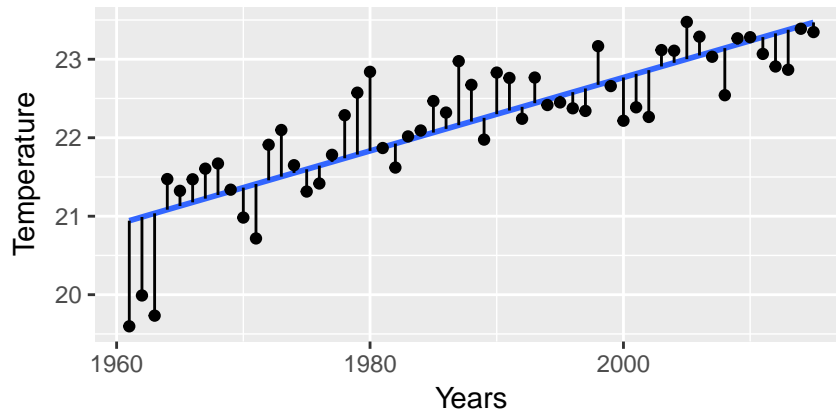


Table 2: Minimum Linear Regression Analysis

<i>Dependent variable:</i>	
MOY	
ANNEES	0.047*** (0.004)
Constant	-70.805*** (8.252)
Observations	55
R <sup>2</sup>	0.706
Adjusted R <sup>2</sup>	0.700
Residual Std. Error	0.489 (df = 53)
F Statistic	127.043*** (df = 1; 53)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

### Residuals of Linear Regression Mango Minimum 1



From the summary of the linear regression, we can see that our r-squared value is about .70. The residuals from the linear regression seem to fall evenly above and below the regression line, but not that many are right on the line. While the linear regression is good, we would like to see if a quadratic model would fit better.

### Quadratic Regression Minimum

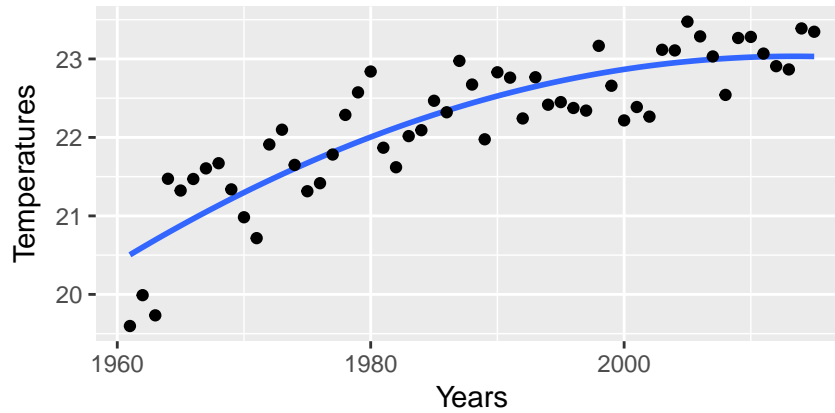
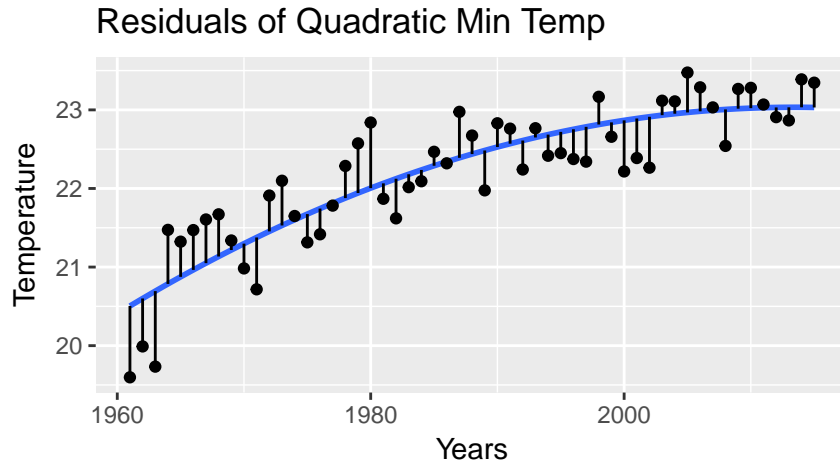


Table 3: Min Temp Quadratic Regression Analysis

<i>Dependent variable:</i>	
MOY	
ANNEES	3.701*** (1.059)
ANNEES2	-0.001*** (0.0003)
Constant	-3,702.762*** (1,052.703)
Observations	55
R <sup>2</sup>	0.760
Adjusted R <sup>2</sup>	0.751
Residual Std. Error	0.445 (df = 52)
F Statistic	82.542*** (df = 2; 52)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

In the quadratic model, the r-squared value did improve to .76, so we feel better about this model. We believe it is the best model so far. To be sure, we graphed the residuals in the following plot.



### Conclusion

Based on the data we were given, we believe that there is enough evidence to conclude there is a significant increase in the minimum temperatures, but there is no significant difference in the maximum temperatures. While the maximum temperature seemed at first to have a negative slope, upon further examination and an  $r$  squared value of .07, we have determined the best model for the maximum temperature is a horizontal line through the mean. For the minimum temperatures, we tried both a linear regression and a quadratic. The linear gave us an  $r$  squared of .70, but the quadratic gave us an improved  $r$  squared of .76. It is for this reason that we believe that our current best model for the City of Mango's minimum temperatures is the quadratic.

### Additional Information Moving Forward

First, since there was an outlier of ~29 Maximum temperature in 1995, we would like to know if this was a mistake, or if something odd happened that year. The point is 4 degrees below the mean and seems out of place. Is this human error? Or was 1995 a very cold year for Mango? Next, we think that it would be very helpful to have the temperature data broken down the averages of the months or seasons. It would be interesting to look and see if specific months are increasing more than others. Or if the winters are getting warmer, but the summers are staying the same. It would also be nice to have some geographic data, like sea level or miles from the coast. Lastly, we would like to know more about the data collection. Was it all collected by the same person? Were the temperatures collected at the exact same geographic location every time?