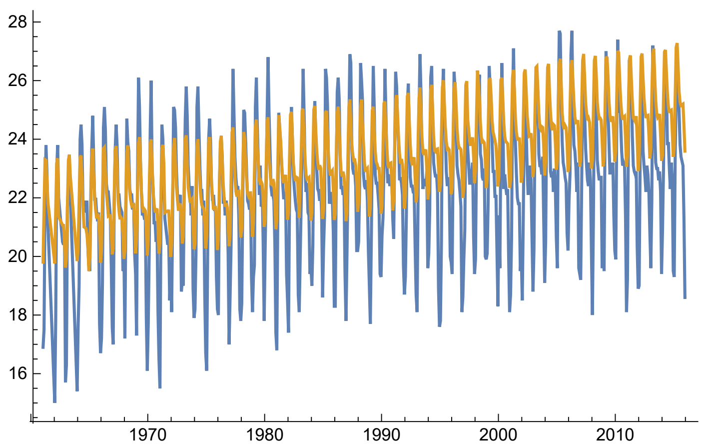
Final Model Fit to City by Matthew Gall and Maria

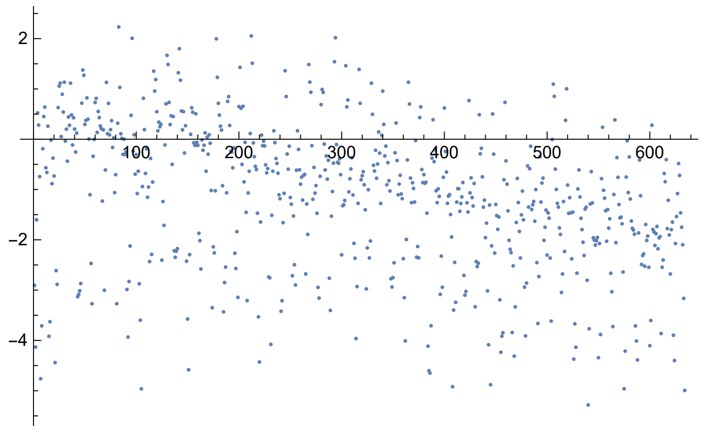
Min Temp vs. Predicted Min

```
minmodel = ListLinePlot[{realMin, predictedMinimum}]
```



Residuals

```
ListPlot[mindata - predictedMin]
```



R^2

```
Apply[Plus, Map[f, predictedMin - Mean[mindata]]] /
  (Apply[Plus, Map[f, mindata - Mean[mindata]]])
```

0.704254

```
finalmin1 = LinearModelFit[mindata1, {
    x,
    Sin[2 * π * x], Cos[2 * π * x],
    Sin[2 * π * x / (1 / 2)], Cos[2 * π * x / (1 / 2)],
    Sin[2 * π * x / (1 / 3)], Cos[2 * π * x / (1 / 3)],
    Sin[2 * π * x / (13)], Cos[2 * π * x / (13)],
    Sin[2 * π * x / (20)], Cos[2 * π * x / (20)],
    latitude,
    latitude * longitude,
    longitude,
    longitude^2,
    latitude^2,
    elev,
    enso,
    sst

   }, {latitude, longitude, elev, enso, sst, x}]
finalmin1["ParameterTable"]
finalmin1["AdjustedRSquared"]
```

LinearModelFit: The rank of the design matrix 14 is less than the number of terms 20 in the model. The model and results based upon it may contain significant numerical error.

FittedModel[$2.42943 + 0.0166399\,\text{elev} - 0.00899611\,\text{enso} + 0.234458\,\text{latitude} + 0.0226269\,\text{latitude}^2 + \ll 16 \gg + 0.0447066\,\text{Sin}\left[\frac{2\pi x}{13}\right] + 1.30183\,\text{Sin}[2\pi x] - 0.0777011\,\text{Sin}[4\pi x] - 0.223193\,\text{Sin}[6\pi x]$]

	Estimate	Standard Error	t-Statistic	P-Value
1	2.42943	0.308503	7.87489	1.55238×10^{-14}
x	0.0421333	0.00244945	17.2012	2.04066×10^{-54}
$\text{Sin}[2\pi x]$	1.30183	0.136718	9.52194	3.84456×10^{-20}
$\text{Cos}[2\pi x]$	−2.01516	0.0834414	−24.1507	4.24219×10^{-91}
$\text{Sin}[4\pi x]$	−0.0777011	0.0713563	−1.08892	0.276618
$\text{Cos}[4\pi x]$	−1.69819	0.0497867	−34.1094	7.57022×10^{-144}
$\text{Sin}[6\pi x]$	−0.223193	0.0479144	−4.65816	3.91393×10^{-6}
$\text{Cos}[6\pi x]$	−0.2911	0.0478778	−6.08006	2.11156×10^{-9}
$\text{Sin}\left[\frac{2\pi x}{13}\right]$	0.0447066	0.0494492	0.904092	0.366301
$\text{Cos}\left[\frac{2\pi x}{13}\right]$	−0.0944232	0.0512014	−1.84415	0.0656423
$\text{Sin}\left[\frac{\pi x}{10}\right]$	0.0681308	0.0505443	1.34794	0.178175
$\text{Cos}\left[\frac{\pi x}{10}\right]$	−0.129169	0.0503085	−2.56753	0.0104781
latitude	0.234458	0.0297728	7.87489	1.55238×10^{-14}
latitude longitude	0.49974	0.0634599	7.87489	1.55238×10^{-14}
longitude	5.17826	0.657565	7.87489	1.55238×10^{-14}
longitude^2	11.0373	1.40158	7.87489	1.55238×10^{-14}
latitude^2	0.0226269	0.0028733	7.87489	1.55238×10^{-14}
elev	0.0166399	0.00211304	7.87489	1.55238×10^{-14}
enso	−0.00899611	0.00333179	−2.70008	0.00712336
sst	0.146348	0.0793749	1.84376	0.0657

```
0.875007
```

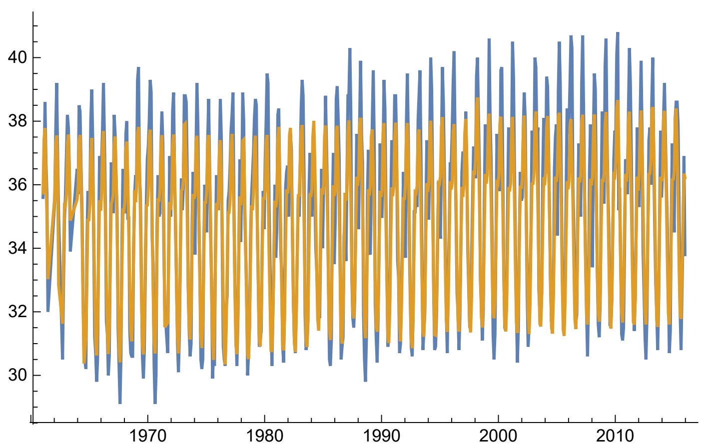
```
finalmin1["ParameterConfidenceIntervals"]
```

```
{{1.82358, 3.03528}, {0.037323, 0.0469436}, {1.03333, 1.57032}, {-2.17903, -1.8513},
 {-0.217833, 0.0624309}, {-1.79597, -1.60042}, {-0.317289, -0.129097},
 {-0.385124, -0.197076}, {-0.0524034, 0.141817}, {-0.194974, 0.00612784},
 {-0.03113, 0.167391}, {-0.227966, -0.0303711}, {0.175989, 0.292927},
 {0.375115, 0.624364}, {3.88691, 6.46961}, {8.28482, 13.7898}, {0.0169842, 0.0282696},
 {0.0124903, 0.0207896}, {-0.0155392, -0.00245302}, {-0.00953105, 0.302228}}
```

The first thing that I did was to examine the data points and make sure that we had good data for our minimum data and for other parameters such as latitude, longitude, elevation, sst, and enso. I had to clean up the sst and enso data to make sure the enso and sst data points matched up with the decimal years that we had in our model and remove the outliers. After doing this I felt good about the data that i had received from the data quality team. I first ran the actual data against the model that was produced for all of Togo. As you can see from above I overlaid model and actual data. the yellow oscillations are the model and the blue oscillations are the actual data points from Mango. There are a couple of take-aways that I had from this model. One major take-away from the picture is just that the model for all of Togo did not capture the extreme temperatures that occurred in the Mango data. The oscillations in the model are much smaller than the oscillations that occurred in the actual data. This was something that i had expected to see after we looked at the Mathematica code that you showed us with the manipulate command. Looking at the manipulate for the min model it showed how the data points for mango were not captured by the model. Even though the model didn’t have a large enough amplitude to capture the oscillations of the actual data it did capture the trend of the data. Looking at the overlay of the two data sets the model shows the oscillations of the data accurately. The model shows the upward trend of the data however the model increases a little more rapidly than the data does. I was concerned with the residuals though that were produced from the model. There is definitely a linear decrease in the residuals over time. At the beginning of the data the residuals are mostly positive and at the end of the data set the residuals are almost all negative. These residuals are definitely not identically and independently distributed. If we had more time on for this project I would go back and see how we can change the model to better fit the data. Another analysis of the data that we can look at is the R^2. The R^2 for this particular model was actually not as bad as i thought it would be. The R^2 was 0.704254. This means our model explains 70.4254 percent of the variance in the data. For the final analysis of the model I used LinearModelFit with my data and the same periods as the overall model. I then looked for significance of each predictor and if the coefficients of the overall model were contained in the confidence intervals for the model that I created. It turns out with the new model that the 13 year period is not significant with my data and neither is the sst. The sst not being significant is kind of surprising since in the overall model the sst seemed to have a big significance in capturing the data. Looking at the confidence intervals for the model I found that the Cos (2 Pi x /(1/3)), Cos (2 Pi x /(13), Sin (2 Pi x /(13), and sst had the only coefficients in the overall model that were with the confidence intervals of the new model I created. This is a little concerning. I am not shocked though because of the fact that the oscillations of the model are not matching the amplitude of the actual data of Mango.

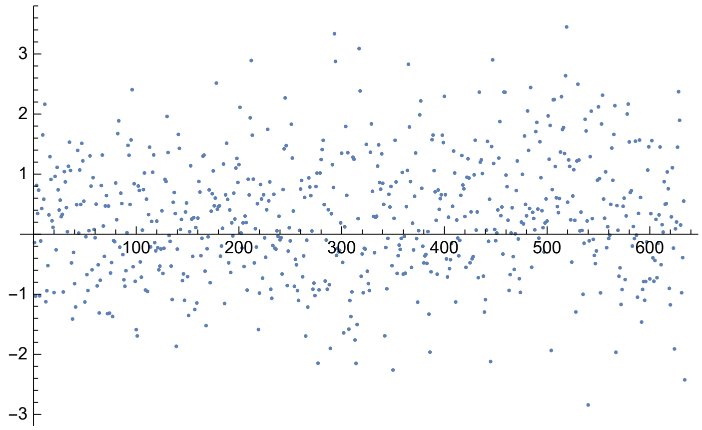
Max Temp vs. Predicted Max

```
maxmodel = ListLinePlot[{realMax, predictedMaximum}]
```



Residuals

```
ListPlot[maxdata - predictedMax]
```



R^2

```
Apply[Plus, Map[f, predictedMax - Mean[maxdata]]] /
  (Apply[Plus, Map[f, maxdata - Mean[maxdata]]])
```

```
0.653229
```

```
finalmax1 = LinearModelFit[maxdata1, {
    x,
    Sin[2 * π * x], Cos[2 * π * x],
    Sin[2 * π * x / (1 / 2)], Cos[2 * π * x / (1 / 2)],
    Sin[2 * π * x / (1 / 3)], Cos[2 * π * x / (1 / 3)],
    Sin[2 * π * x / (1 / 4)], Cos[2 * π * x / (1 / 4)],
    Sin[2 * π * x / (13)], Cos[2 * π * x / (13)],
    latitude,
    latitude * longitude,
    longitude,
    longitude^2,
    latitude^2,
    elev,
    enso,
    sst

   }, {latitude, longitude, elev, enso, sst, x}]
finalmax1["ParameterTable"]
finalmax1["AdjustedRSquared"]
```

LinearModelFit: The rank of the design matrix 14 is less than the number of terms 20 in the model. The model and results based upon it may contain significant numerical error.

FittedModel[$3.11197 + 0.0213149\,\text{elev} - 0.00741464\,\text{enso} + 0.300328\,\text{latitude} + 0.0289839\,\text{latitude}^2 + \ll 16 \gg + 1.78278\,\text{Sin}[2\pi x] - 0.543635\,\text{Sin}[4\pi x] - 0.323019\,\text{Sin}[6\pi x] - 0.0348722\,\text{Sin}[8\pi x]$]

	Estimate	Standard Error	t-Statistic	P-Value
1	3.11197	0.299751	10.3819	2.25712×10^{-23}
x	0.021159	0.00238484	8.8723	7.78925×10^{-18}
Sin[2 π x]	1.78278	0.133201	13.3841	4.97176×10^{-36}
Cos[2 π x]	1.96244	0.0814535	24.0928	8.70048×10^{-91}
Sin[4 π x]	−0.543635	0.0700027	−7.76592	3.41124×10^{-14}
Cos[4 π x]	−1.02143	0.0492963	−20.7203	1.02766×10^{-72}
Sin[6 π x]	−0.323019	0.0475253	−6.79679	2.53703×10^{-11}
Cos[6 π x]	−0.654322	0.0474803	−13.7809	7.71209×10^{-38}
Sin[8 π x]	−0.0348722	0.0475173	−0.733884	0.4633
Cos[8 π x]	−0.232375	0.0476099	−4.88082	1.34838×10^{-6}
$\text{Sin}[\frac{2\pi x}{13}]$	−0.0118528	0.0482017	−0.2459	0.805842
$\text{Cos}[\frac{2\pi x}{13}]$	−0.0444647	0.0494053	−0.899998	0.368474
latitude	0.300328	0.0289282	10.3819	2.25712×10^{-23}
latitude longitude	0.64014	0.0616595	10.3819	2.25712×10^{-23}
longitude	6.63307	0.63891	10.3819	2.25712×10^{-23}
longitude^2	14.1382	1.36182	10.3819	2.25712×10^{-23}
latitude^2	0.0289839	0.00279178	10.3819	2.25712×10^{-23}
elev	0.0213149	0.00205309	10.3819	2.25712×10^{-23}
enso	−0.00741464	0.00329241	−2.25204	0.0246722
sst	0.457821	0.0770535	5.9416	4.73059×10^{-9}

0.910381

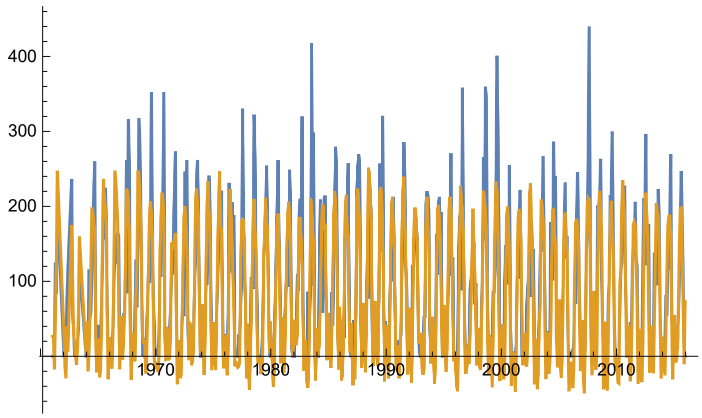
```
finalmax1["ParameterConfidenceIntervals"]
```

```
{{2.52331, 3.70063}, {0.0164756, 0.0258425}, {1.5212, 2.04437}, {1.80248, 2.1224},
 {-0.681109, -0.406161}, {-1.11824, -0.924622}, {-0.416351, -0.229688},
 {-0.747565, -0.561079}, {-0.128188, 0.058444}, {-0.325873, -0.138877},
 {-0.106513, 0.0828074}, {-0.141489, 0.0525592}, {0.243518, 0.357138},
 {0.519051, 0.761229}, {5.37836, 7.88779}, {11.4638, 16.8126}, {0.0235013, 0.0344665},
 {0.017283, 0.0253468}, {-0.0138804, -0.000948892}, {0.306501, 0.609142}}
```

After running the overall model versus the actual data of the maximum average temperatures of Mango I am pleased with the results that we got. First just looking at the graphical fit of the model versus the actual data the fit looked pretty good. Unlike the minimum model the maximum model does a good job of capturing the magnitude of the oscillations of the data. Now the amplitude of oscillations of the actual data are a little larger than the model however it is not as bad as the minimum model. Also the maximum model accurately shows the trend of the maximum temperature increasing over time. This was also what I had expected after we had looked at the Mathematica code of the manipulate command in class. Looking at the residuals initially I was pleased with them. There are no extreme residual values. Most of them fall between the values of -3 to 3. There is also no periodic behavior in the residuals also. However with closer examination you can see that most of the residuals are positive. The residuals are not extreme positive numbers however there seems to be a pattern in them where a majority of the residuals are positive. Also for the maximum model we considered the R^2 to see how well the model represented the actual data. Our R^2 for the maximum model was 0.653229. Also for the maximum model.For the final analysis of the model I used LinearModelFit with my data and the same periods as the overall model. I then looked for significance of each predictor and if the coefficients of the overall model were contained in the confidence intervals for the model that I created. For the maximum model the only predictor that was not significant was the 13 year issue. Looking at the confidence intervals I found that Sin(2 Pi x/ (1/4)), Cos(2 Pi x/ (1/4)),Sin(2 Pi x/ (13)), latitude* longitude, enso, and, sst had the only coefficients in the overall model that were with the confidence intervals of the new model I created for the maximum data. This is more than there was for the minimum model which is good and shows that the overall model is doing a better job predicting maximum temperature than minimum temperature in Mango. Considering all of this analysis and that the model was meant for 10 different cities i am very pleased with the model would feel comfortable keeping the model for maximum temperature.

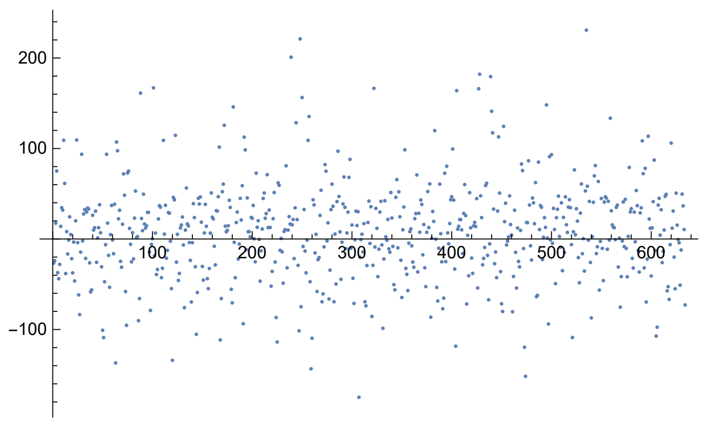
Rainfall vs. Predicted Rainfall

```
rainmodel = ListLinePlot[{realRain, predictedRainfall}]
```



Residuals

```
ListPlot[rainfall - predictedRain]
```



R^2

```
Apply[Plus, Map[f, predictedRain - Mean[rainfall]]] /
  (Apply[Plus, Map[f, rainfall - Mean[rainfall]]])
```

```
0.88462
```

```
finalrainfall1 = LinearModelFit[rainfall1, {
    x,
    Sin[2 * π * x], Cos[2 * π * x],
    Sin[2 * π * x / (1 / 2)], Cos[2 * π * x / (1 / 2)],
    Sin[2 * π * x / (1 / 3)], Cos[2 * π * x / (1 / 3)],
    Sin[2 * π * x / (1 / 4)], Cos[2 * π * x / (1 / 4)],
    latitude,
    latitude * longitude,
    longitude,
    longitude^2,
    latitude^2,
    elev,
    enso,
    sst

  }, {latitude, longitude, elev, enso, sst, x}]
finalrainfall1["ParameterTable"]
finalrainfall1["AdjustedRSquared"]
```

LinearModelFit: The rank of the design matrix 12 is less than the number of terms 18 in the model. The model and results based upon it may contain significant numerical error.

FittedModel[1.98717 + 0.0136108 elev + 0.227271 enso + 0.191777 latitude + 0.0185078 $latitude^2$ + ≪11≫ + 0.813919 Cos[8 π x] − 52.2381 Sin[2 π x] + 33.9119 Sin[4 π x] + 1.34203 Sin[6 π x] − 5.14994 Sin[8 π x]]

	Estimate	Standard Error	t-Statistic	P-Value
1	1.98717	15.6916	0.12664	0.899267
x	−0.123789	0.123993	−0.998357	0.318498
Sin[2 π x]	−52.2381	6.9732	−7.49126	2.3778×10^{-13}
Cos[2 π x]	−101.866	4.26506	−23.8838	9.7453×10^{-90}
Sin[4 π x]	33.9119	3.6673	9.2471	3.72993×10^{-19}
Cos[4 π x]	−3.50356	2.58485	−1.35542	0.17578
Sin[6 π x]	1.34203	2.49204	0.538527	0.590408
Cos[6 π x]	20.1449	2.48979	8.09098	3.1606×10^{-15}
Sin[8 π x]	−5.14994	2.49167	−2.06686	0.0391642
Cos[8 π x]	0.813919	2.49658	0.326013	0.744525
latitude	0.191777	1.51435	0.12664	0.899267
latitude longitude	0.408766	3.22779	0.12664	0.899267
longitude	4.23559	33.4461	0.12664	0.899267
$longitude^2$	9.02804	71.2892	0.12664	0.899267
$latitude^2$	0.0185078	0.146146	0.12664	0.899267
elev	0.0136108	0.107476	0.12664	0.899267
enso	0.227271	0.165906	1.36988	0.171225
sst	2.78937	4.03422	0.691427	0.489558

```
0.772566
```

```
finalrainfall1["ParameterConfidenceIntervals"]
```

```
{{-28.8282, 32.8026}, {-0.367289, 0.119711}, {-65.9322, -38.5439},
 {-110.242, -93.49}, {26.71, 41.1139}, {-8.57976, 1.57263}, {-3.5519, 6.23596},
 {15.2553, 25.0344}, {-10.0431, -0.256746}, {-4.08892, 5.71676}, {-2.78214, 3.16569},
 {-5.93004, 6.74757}, {-61.4465, 69.9177}, {-130.971, 149.027}, {-0.268497, 0.305512},
 {-0.197454, 0.224675}, {-0.0985394, 0.55308}, {-5.13312, 10.7119}}
```

My final analysis was for the rainfall model. For this I compared the overall rainfall model for all of Togo and the actual rainfall data for Mango. Graphically I was pleased with the results that I got. The model seemed to capture the overall trend and oscillations of the data. However it seems that the amplitude of oscillation of the actual Mango rainfall are larger than the amplitude of oscillations of the overall model. This then lead me to look at the residual graph for the model and data. Residuals seem to be independent and identically distributed of each other. There seems to be no periodic behavior or pattern of the residuals. Also there are not really any extreme residual values. Also to anyalse the model and its fit I calculated the R^2 for the model and data. I was very pleased with the result. The R^2 was 0.88462. This gave me confidence that the overall Togo model was doing a good job predicting the trends of rainfall in Mango which is what we were looking for. For the final analysis of the model I used LinearModelFit with my data and the same periods as the overall rainfall model. I looked for the significance of each of the terms in the new model and if the coefficients from the overall model were contained in the confidence intervals in the new model. When looking at the new model with Mango's data almost none of the periods or predictors that were significant in the overall model were significant in this model. the only predictors that were significant were periods of one year, half a year, and one third of a year. looking at the confidence intervals I found the decimal year, Sin (2 Pi x/ (1/4)), longitude ^2, elevation, and enso were the only predictors whose coefficients in the overall model were contained in the new model I created. I wasn't sure what to do with this since we did have a really good residual graph and R^2 for this model. I was pleased with how this model fit the data for Mango.

Summary for the paper

When comparing the overall models for minimum temperature, maximum temperature, and rainfall we were pleased to see the trend of the data for Mango being captured by the models. Although the models were not perfect fits for Mango the models did a good job describing the data and we feel would give an accurate prediction of temperatures and rainfall into the future for mango. For all there models the one aspect that they seemed to be lacking when it came to the Mango data was that the models didn't seem to quite capture the amplitude of oscillations in the data.The data had larger amplitudes of oscillations than the model did. A possible answer to this is that more periods of higher amplitude are needed in the model. the rainfall model fit its data the best of of all three of the models when we consider the residuals and R^2 for the models and data. The models seem to be acceptable fits for the data however they are not the best possible fits for the Mango data.