1) We certainly agreed with the first group who did this city about the minimum temperature increasing. However, we disagree about maximum temperature decreasing. We do not believe we can that maximum temperature is increasing or decreasing. I believe that the first group should have provided some statistics to support their conclusions.

2) Dapaong is located in northern Togo. Its population is 58, 071. Its elevation is 1082 feet. It is in the Savanas region.

3) Looking over monthly minimum and maximum temperatures of Dapaong over the last 50 years, it is difficult to find any sort of trend. After taking a closer look at the minimum monthly temperatures, it appears that November, December, and January had significantly lower temperatures before the year 1980. I think it explains why the first group's model displays a significant upward trend for minimum yearly data.
When comparing the yearly means to the monthly means, I could only find 2 mismatches. The years 1978 and 1979 contain separate values in the maximum temperature datasets. Both temperatures are only off by about 1 degree Celsius (21.1 vs 22.1 and 21.2 vs 22.1). This could have been a typo.

Minimum Data

```
mindata = Import["C:\\Users\\Maria\\Documents\\NKU Spring 2018\\MAT 375\\MIN.xlsx"];

Years = Table[i, {i, 1, 660}];

mindataclean = Flatten[mindata];
```

4) Determine any major outliers by finding "outer fences" for dataset.

```
Quartiles[Sort[mindataclean]]

{21.2262, 22.1097, 23.6367}

(*Find interquartile range *)

InterquartileRange[Sort[mindataclean]]

2.41048

(*Determine the "outer fences"*)

23.63666666666667` + (2.410483870967745` * 3)

30.8681

21.226182795698925` - (3 * 2.410483870967745`)

13.9947
```

4) We need to eliminate any values that are above 30.8681 and any data values that are below 13.9947. So, we will eliminate the following data points: (570, 30.9), (66, 34.5), and (291, 34.2). We will run analysis on both the original dataset and the dataset with the outliers removed.

```
datatogether = Table[{Years[[i]], mindataclean[[i]]}, {i, 660}];

outlierremoved = Delete[datatogether, 570];

outlierremoved2 = Delete[outlierremoved, 66];

outlierremoved3 = Delete[outlierremoved2, 291];
```

5) We believe that there is a significant increase in minimum temperature over time. The coefficient on the x term is positive, which would indicate that the temperature is increasing over time. Also, the t-statistic for the x term was 11.6531, which means that we are over 11 standard deviations away from zero. The p-value was very small: 1.21288×10^{-28}. This supports the hypothesis that the x term does not equal zero. Using this evidence, we are able to state that there is a significant increase in temperature over time.

6) We believe that the best model is the linear model:
$20.9413 + 0.0039341\,x - 1.56935\,\text{Cos}\left[\frac{\pi x}{6}\right] + 1.1958\,\text{Sin}\left[\frac{\pi x}{6}\right]$. The linear model shows a significant increase over time and has a positive slope of .0038. Also, the confidence intervals did not include zero for any of the terms, which indicates that the none of the terms equal zero. This was further supported by the small p-values and the large t-statistics. Also, the residuals did not show any pattern or “smiley faces.” The R^2 value was a bit disappointing--it was only around .46 for the original dataset and around .48 for the dataset with the outlier removed. We would like to have seen a higher R^2 value so that we could know that the variation in our data was properly represented. Because the R^2 was higher for the dataset with the outliers removed, we would certainly use that dataset.

We also considered using higher order models. We looked at the quadratic and the cubic. We rejected the quadratic model because the confidence intervals included zero for the x2 term (which would mean that the x2 term probably equals zero. This was further supported by the relatively large p-value and relatively small t-statistic for the x2 term. We rejected the cubic model because the confidence intervals included zero for the x3, x2, and x terms. The hypothesis that those three terms would equal zero was further supported by the relatively high p-values and relatively small t-statistics for those three terms.

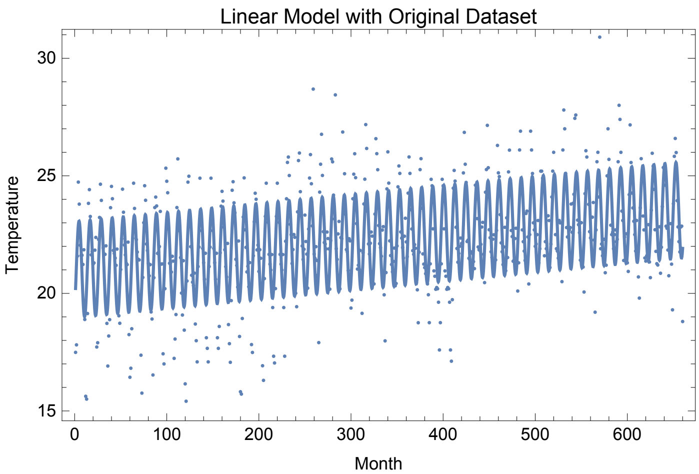
Minimum Linear Model with Sin and Cos (Original Dataset)

7) Graph of data

```
line = LinearModelFit[datatogether, {Sin[2 * π * x / 12], Cos[2 * π * x / 12], x}, x]
```

FittedModel[$21.0114 + 0.00385149\,x - 1.63913\,\mathrm{Cos}\left[\frac{\pi x}{6}\right] + 1.22249\,\mathrm{Sin}\left[\frac{\pi x}{6}\right]$]

```
Show[ListPlot[datatogether], Plot[line[x], {x, 1, 660}], Frame → True,
 PlotLabel → "Linear Model with Original Dataset", FrameLabel → {"Month", "Temperature"}]
```

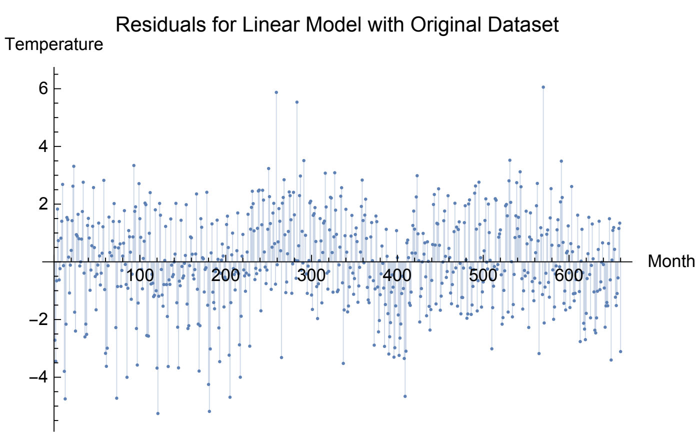



```
line["ParameterTable"]
```

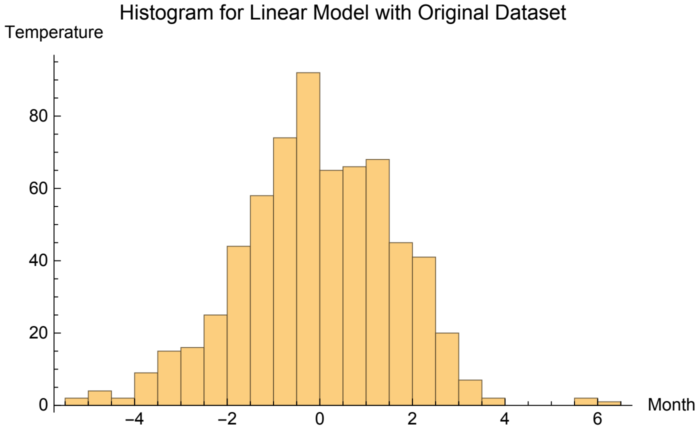
	Estimate	Standard Error	t-Statistic	P-Value
1	21.0114	0.137851	152.42	$2.519171753867 \times 10^{-514}$
$\mathrm{Sin}\left[\frac{\pi x}{6}\right]$	1.22249	0.0973668	12.5555	1.47779×10^{-32}
$\mathrm{Cos}\left[\frac{\pi x}{6}\right]$	−1.63913	0.0973581	−16.8361	3.92198×10^{-53}
x	0.00385149	0.000361365	10.6582	1.42953×10^{-24}

```
ListPlot[line["FitResiduals"], Filling → Axis,
 Axes → True, AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Residuals for Linear Model with Original Dataset"]
```

8) Graphs of Residuals



```
Histogram[line["FitResiduals"], AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Histogram for Linear Model with Original Dataset"]
```



From the list plot of the residuals, I do not see a super prominent pattern or smiley face. From the histogram, the distribution looks somewhat normal. There is a “shrug” on the left side of the histogram. Overall, the residuals look pretty good.

```
line["ParameterConfidenceIntervals"]
```

```
{{20.7407, 21.282}, {1.0313, 1.41368}, {-1.8303, -1.44796}, {0.00314192, 0.00456106}}
```

The confidence intervals do not contain zero which means that the terms probably do not equal zero. The p-values are very small for all of the terms, and the t-statistics are large for all the terms. This supports the statement that none of the terms equal zero.

```
line["RSquared"]
```

```
0.455921
```

This R^2 value is smaller than we would like to see. The R^2 value represents the amount of variation in the data that the model explains (fraction of total variation). We would like our R^2 value to be large so that most of the variation in the data is explained by the model. Since this R^2 value is small, it means that our model does not do a good job of representing the variation in the data.

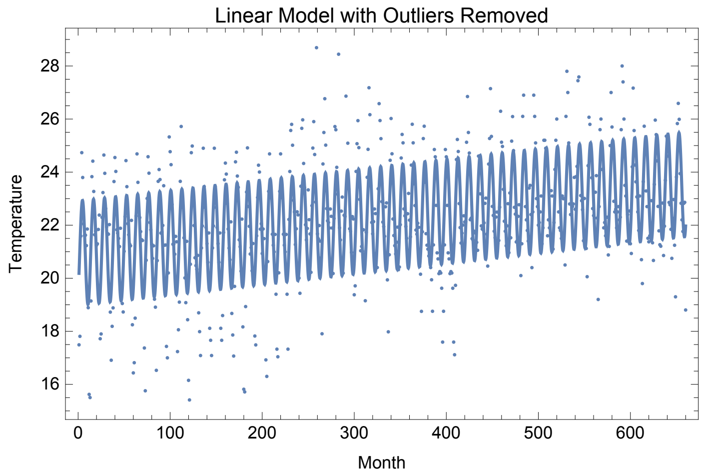
Minimum Linear Model with Sin and Cos (Outlier Removed)

7) Graph of data

```
lineo = LinearModelFit[outlierremoved3, {Sin[2*π*x/12], Cos[2*π*x/12], x}, x]
```

FittedModel[$20.9413 + 0.0039341\,x - 1.56935\,\mathrm{Cos}\left[\frac{\pi x}{6}\right] + 1.1958\,\mathrm{Sin}\left[\frac{\pi x}{6}\right]$]

```
Show[ListPlot[outlierremoved3], Plot[lineo[x], {x, 1, 660}], Frame → True,
 FrameLabel → {"Month", "Temperature"}, PlotLabel → "Linear Model with Outliers Removed"]
```

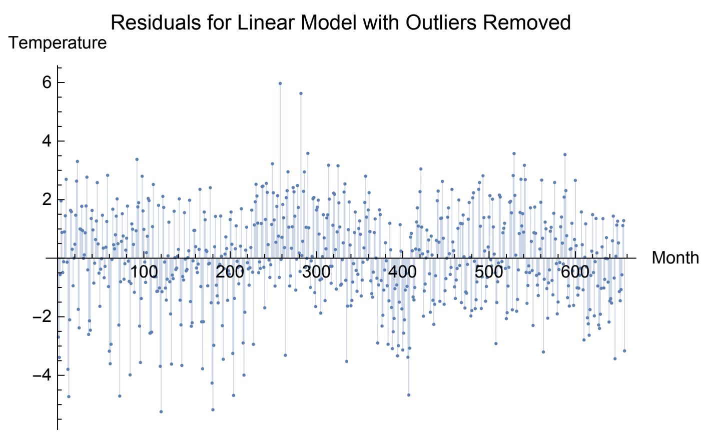



```
lineo["ParameterTable"]
```

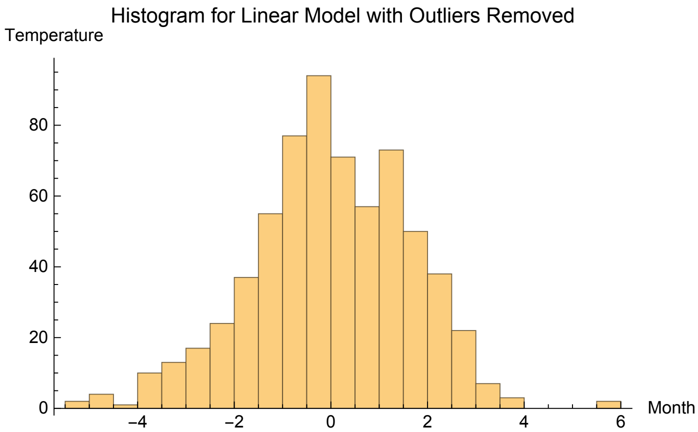
	Estimate	Standard Error	t-Statistic	P-Value
1	20.9413	0.1288	162.588	$1.852227049373 \times 10^{-530}$
$\text{Sin}\left[\frac{\pi x}{6}\right]$	1.1958	0.090823	13.1663	2.79178×10^{-35}
$\text{Cos}\left[\frac{\pi x}{6}\right]$	−1.56935	0.0910239	−17.2411	3.54407×10^{-55}
x	0.0039341	0.000337603	11.6531	1.21288×10^{-28}

8) Graphs of Residuals

```
ListPlot[lineo["FitResiduals"], Filling → Axis,
 Axes → True, AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Residuals for Linear Model with Outliers Removed"]
```



```
Histogram[lineo["FitResiduals"], AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Histogram for Linear Model with Outliers Removed"]
```



This list plot from the residuals with the outliers removed are similar to the list plot from the residuals with the original dataset. There is not a prominent pattern to the residuals. The histogram for the data with the outliers removed looks better than the histogram with the original data set. This histogram looks more normal than the previous histogram.

```
lineo["ParameterConfidenceIntervals"]
```

```
{{20.6884, 21.1942}, {1.01746, 1.37414}, {-1.74809, -1.39062}, {0.00327118, 0.00459702}}
```

The confidence intervals do not contain zero which means that the terms probably do not equal zero. The p-values are very small for all of the terms, and the t-statistics are large for all the terms. This supports the statement that none of the terms equal zero.

```
lineo["RSquared"]
```

```
0.479021
```

This R^2 value larger than the R^2 value for the original data set. This means that there is less variation in this data, which makes sense since we have removed the outliers. This R^2 value is pretty small. The model still does not do a good job of representing the variation in the data.

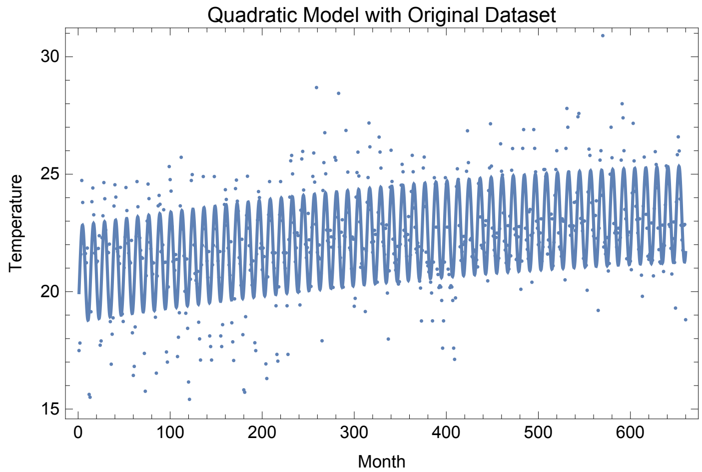
Minimum Quadratic Model with Sin and Cos (Original Dataset)

7) Graph of data

```
quad = LinearModelFit[datatogether, {Sin[2 * π * x / 12], Cos[2 * π * x / 12], x^2, x}, x]
```

FittedModel[$20.7593 + 0.00613582\,x - 3.45588\times 10^{-6}\,x^2 - 1.63908\,\text{Cos}\left[\frac{\pi x}{6}\right] + 1.2225\,\text{Sin}\left[\frac{\pi x}{6}\right]$]

```
Show[ListPlot[datatogether], Plot[quad[x], {x, 1, 660}],
 Frame → True, FrameLabel → {"Month", "Temperature"},
 PlotLabel → "Quadratic Model with Original Dataset"]
```

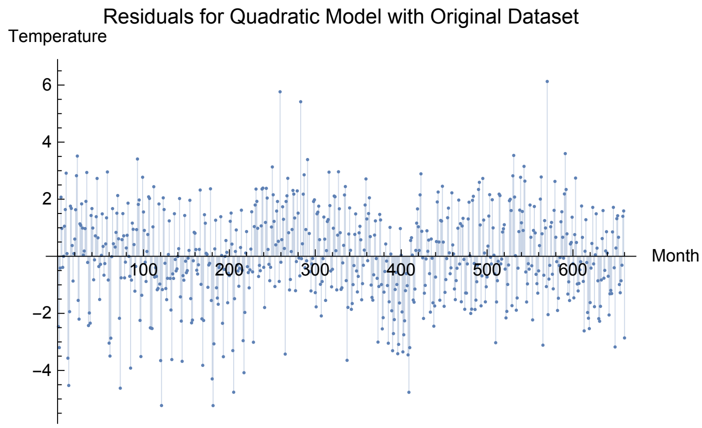



```
quad["ParameterTable"]
```

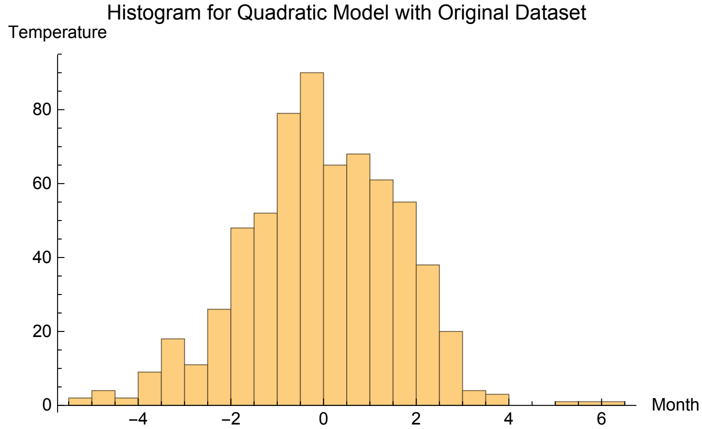
	Estimate	Standard Error	t-Statistic	P-Value
1	20.7593	0.206899	100.336	$8.08730359121 \times 10^{-400}$
$\text{Sin}\left[\frac{\pi x}{6}\right]$	1.2225	0.0972436	12.5715	1.26886×10^{-32}
$\text{Cos}\left[\frac{\pi x}{6}\right]$	−1.63908	0.0972349	−16.8569	3.1577×10^{-53}
x^2	-3.45588×10^{-6}	2.11766×10^{-6}	−1.63193	0.103174
x	0.00613582	0.00144555	4.24463	0.0000250592

8) Graphs of Residuals

```
ListPlot[quad["FitResiduals"], Filling → Axis,
 Axes → True, AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Residuals for Quadratic Model with Original Dataset"]
```



```
Histogram[quad["FitResiduals"], AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Histogram for Quadratic Model with Original Dataset"]
```



The residuals look pretty random and independent. I do not see any prominent pattern or "smiley faces" in the residuals. From the histogram, the residuals look close to a normal distribution. There is a "shrug" on the left side, which is cause for some concern.

```
quad["ParameterConfidenceIntervals"]
```

$\{\{20.3531, 21.1656\}, \{1.03155, 1.41345\}, \{-1.83001, -1.44815\}, \{-7.61409\times10^{-6}, 7.02337\times10^{-7}\}, \{0.00329735, 0.00897429\}\}$

The confidence interval contains 0 for the x^2 term, which means that the x^2 probably equals zero. Since the p-value for x^2 is around .103 (which is fairly large), the probability that the x^2 is zero is fairly high. The statement that the x^2 term equals zero is further supported by the fairly small t-statistic, which is -1.63. The evidence points to the conclusion that the x^2 probably equals zero.

```
quad["RSquared"]
```

0.458124

This R^2 value is pretty small. We would like to a larger R^2 value.

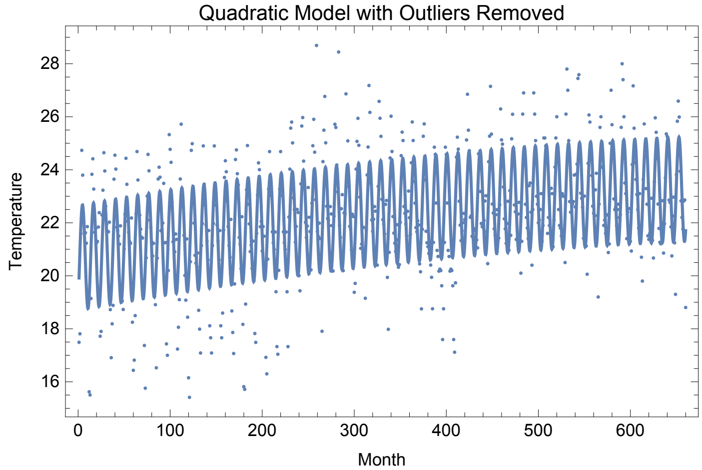
Minimum Quadratic Model with Sin and Cos (Outlier Removed)

7) Graph of data

```
quado = LinearModelFit[outlierremoved3, {Sin[2 * π * x / 12], Cos[2 * π * x / 12], x^2, x}, x]
```

FittedModel[$20.6704 + 0.00638788\,x - 3.71161\times10^{-6}\,x^2 - 1.56888\,\mathrm{Cos}\left[\frac{\pi x}{6}\right] + 1.19616\,\mathrm{Sin}\left[\frac{\pi x}{6}\right]$]

```
Show[ListPlot[outlierremoved3], Plot[quado[x], {x, 1, 660}],
 Frame → True, FrameLabel → {"Month", "Temperature"},
 PlotLabel → "Quadratic Model with Outliers Removed"]
```

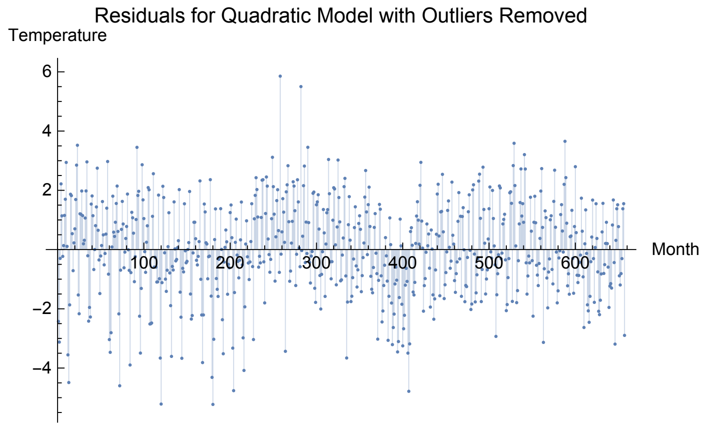



```
quado["ParameterTable"]
```

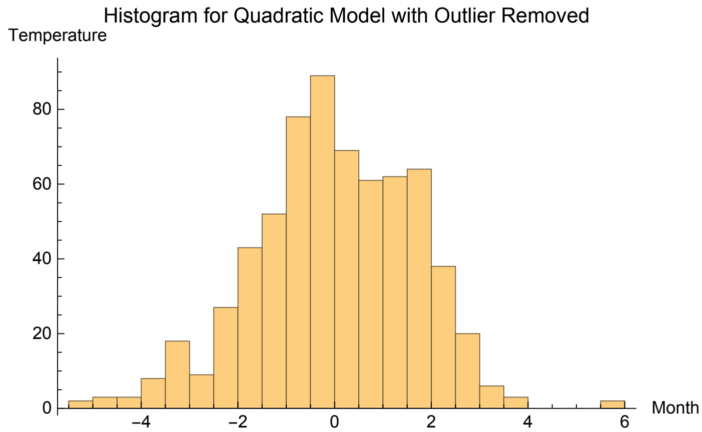
	Estimate	Standard Error	t-Statistic	P-Value
1	20.6704	0.193199	106.99	$9.47794518976\times10^{-416}$
$\text{Sin}\left[\frac{\pi x}{6}\right]$	1.19616	0.0906478	13.1956	2.08021×10^{-35}
$\text{Cos}\left[\frac{\pi x}{6}\right]$	−1.56888	0.0908485	−17.2692	2.62001×10^{-55}
x^2	-3.71161×10^{-6}	1.97575×10^{-6}	−1.87858	0.0607477
x	0.00638788	0.00134895	4.73545	2.68332×10^{-6}

8) Graphs of Residuals

```
ListPlot[quado["FitResiduals"], Filling → Axis,
 Axes → True, AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Residuals for Quadratic Model with Outliers Removed"]
```



```
Histogram[quado["FitResiduals"], AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Histogram for Quadratic Model with Outlier Removed"]
```



These residuals look pretty similar to the residuals for the original dataset. They are fairly random and does not seem to have a “smiley face.” Both histograms look pretty normal--the histogram for the original data set is a bit “smoother.”

```
quado["ParameterConfidenceIntervals"]
```

```
{{20.291, 21.0498}, {1.01816, 1.37415}, {-1.74727, -1.39049},
 {-7.59121 × 10^-6, 1.67995 × 10^-7}, {0.00373907, 0.00903669}}
```

Similar to the confidence intervals with the original data set, these confidence intervals contain 0 for the x^2 term, which means that the x^2 probably equals zero. Since the p-value for x^2 is around .06, the probability that the x^2 equals zero is fairly high. The statement that the x^2 equals zero is further supported by the fairly small t-statistic, which is -1.89. The evidence points to the conclusion that the x^2 probably equals zero.

```
quado["RSquared"]
```

```
0.481826
```

This R^2 value higher than the R^2 value of the original data set. This makes sense since we have removed the outliers. We would still like to see a higher R^2 value.

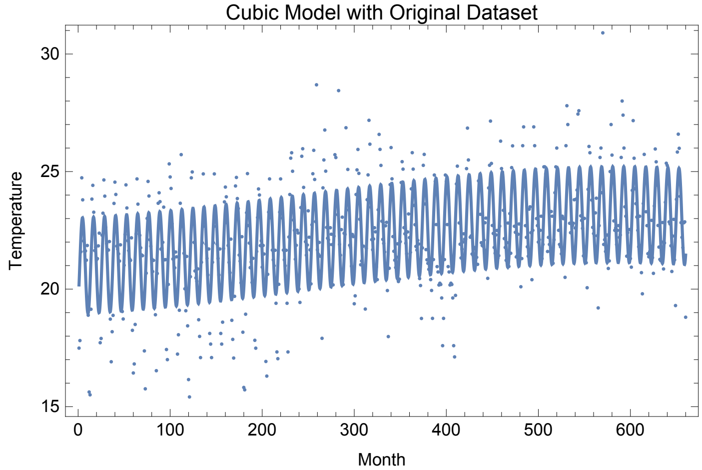
Minimum Cubic Model with Sin and Cos (Original Dataset)

7) Graph of data

```
cub = LinearModelFit[datatogether, {Sin[2 * π * x / 12], Cos[2 * π * x / 12], x^3, x^2, x}, x]
```

FittedModel[$20.983 + 0.00209015\,x + 0.0000118334\,x^2 - 1.54203 \times 10^{-8}\,x^3 - 1.63841\,\mathrm{Cos}\left[\frac{\pi x}{6}\right] + 1.21999\,\mathrm{Sin}\left[\frac{\pi x}{6}\right]$]

```
Show[ListPlot[datatogether], Plot[cub[x], {x, 1, 660}], Frame → True,
 PlotLabel → "Cubic Model with Original Dataset", FrameLabel → {"Month", "Temperature"}]
```

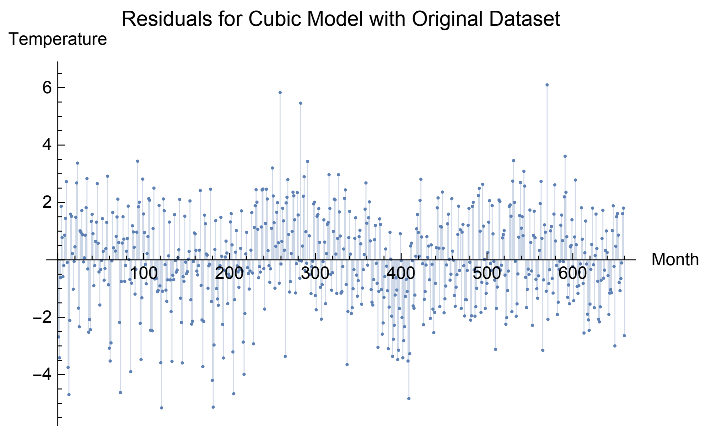



```
cub["ParameterTable"]
```

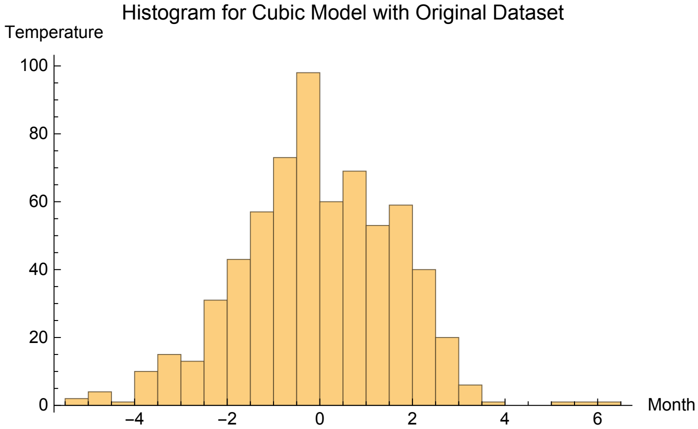
	Estimate	Standard Error	t-Statistic	P-Value
1	20.983	0.276547	75.8751	$2.216373263263 \times 10^{-326}$
$\text{Sin}\left[\frac{\pi x}{6}\right]$	1.21999	0.0972294	12.5476	1.63699×10^{-32}
$\text{Cos}\left[\frac{\pi x}{6}\right]$	−1.63841	0.0972005	−16.8559	3.28725×10^{-53}
x^3	-1.54203×10^{-8}	1.26532×10^{-8}	−1.21868	0.223404
x^2	0.0000118334	0.000012723	0.930073	0.352676
x	0.00209015	0.00362057	0.5773	0.563936

8) Graphs of Residuals

```
ListPlot[cub["FitResiduals"], Filling → Axis,
 Axes → True, AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Residuals for Cubic Model with Original Dataset"]
```



```
Histogram[cub["FitResiduals"], AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Histogram for Cubic Model with Original Dataset"]
```



The residuals look independent and random. I do not see any pattern or "smiley faces". The histogram looks pretty normal.

```
cub["ParameterConfidenceIntervals"]
```

```
{{20.44, 21.5261}, {1.02908, 1.41091}, {-1.82927, -1.44754},
 {-4.02662 × 10^-8, 9.42559 × 10^-9}, {-0.0000131496, 0.0000368163}, {-0.00501919, 0.00919949}}
```

The confidence intervals contain zero for the x^3, x^2, and x terms. This means that these terms probably equal zero. Also, the p-values for x^3, x^2, and x are pretty high: .22, .35, and .56 (respectively). This says that the probability that these three terms equal zero is high. This is further supported by the t-statistic for these three terms, which are small. All the evidence supports the statement that says that the x^3, x^2, and x terms equal zero.

```
cub["RSquared"]
```

```
0.459352
```

This R^2 value is pretty small. Since R^2 represents the total amount of variation that the model explains, we want the R^2 to be high. The low R^2 indicates that this model does not do a good job of representing the variation in our data.

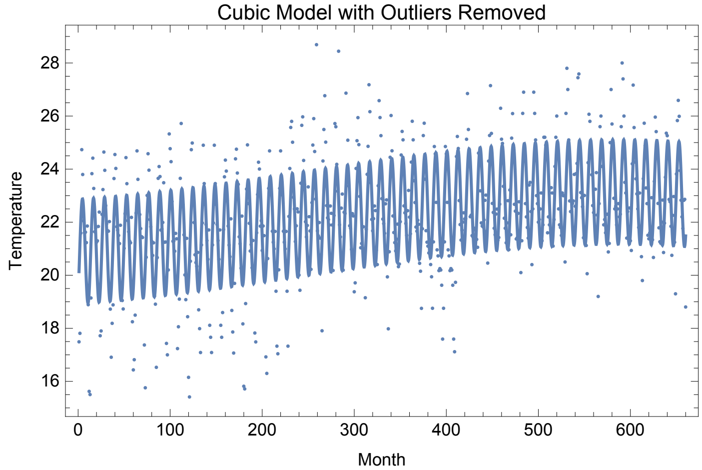
Minimum Cubic Model with Sin and Cos (Outlier Removed)

7) Graph of data

```
cubo = LinearModelFit[outlierremoved3, {Sin[2 * π * x / 12], Cos[2 * π * x / 12], x^3, x^2, x}, x]
```

FittedModel[$20.8952 + 0.00232436\,x + 0.0000116399\,x^2 - 1.54793 \times 10^{-8}\,x^3 - 1.56829\,\mathrm{Cos}\left[\frac{\pi x}{6}\right] + 1.19353\,\mathrm{Sin}\left[\frac{\pi x}{6}\right]$]

```
Show[ListPlot[outlierremoved3], Plot[cubo[x], {x, 1, 660}], Frame → True,
 PlotLabel → "Cubic Model with Outliers Removed", FrameLabel → {"Month", "Temperature"}]
```

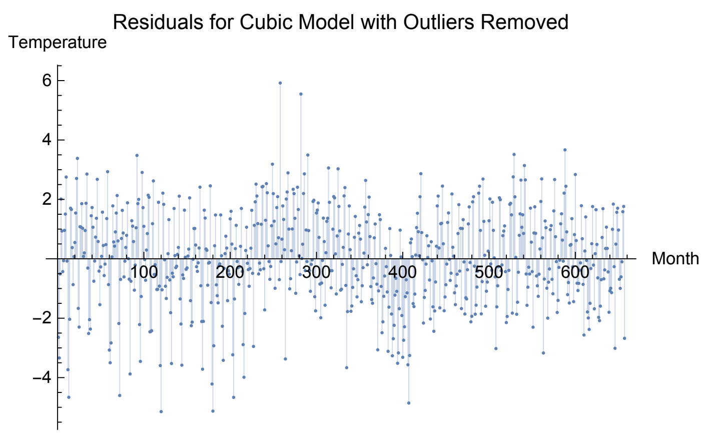



```
cubo["ParameterTable"]
```

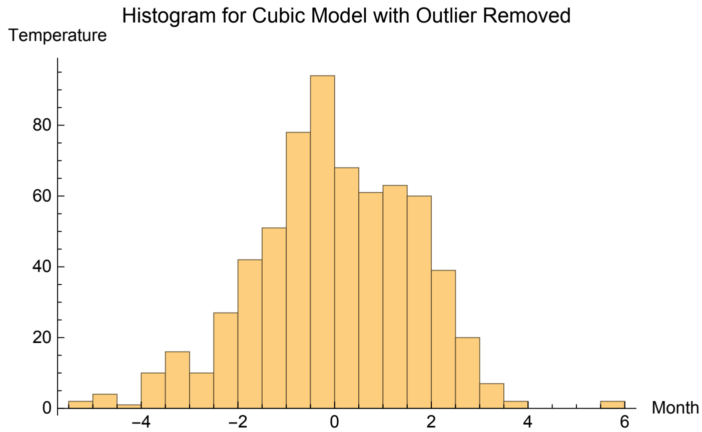
	Estimate	Standard Error	t-Statistic	P-Value
1	20.8952	0.258018	80.9837	$3.896964229785 \times 10^{-342}$
$\sin\left[\frac{\pi x}{6}\right]$	1.19353	0.0906194	13.1708	2.73213×10^{-35}
$\cos\left[\frac{\pi x}{6}\right]$	−1.56829	0.0907991	−17.272	2.6133×10^{-55}
x^3	-1.54793×10^{-8}	1.1783×10^{-8}	−1.3137	0.18941
x^2	0.0000116399	0.0000118514	0.982157	0.326388
x	0.00232436	0.00337423	0.688856	0.491159

8) Graphs of Residuals

```
ListPlot[cubo["FitResiduals"], Filling → Axis,
 Axes → True, AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Residuals for Cubic Model with Outliers Removed"]
```



```
Histogram[cubo["FitResiduals"], AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Histogram for Cubic Model with Outlier Removed"]
```



Like the previous residuals, these look pretty independent. The histogram for the residuals of the original dataset looks more normal than this histogram.

```
cubo["ParameterConfidenceIntervals"]
```

$\{\{20.3886, 21.4019\}, \{1.01559, 1.37148\}, \{-1.74658, -1.38999\},$
$\{-3.86165\times10^{-8}, 7.65793\times10^{-9}\}, \{-0.0000116316, 0.0000349115\}, \{-0.00430133, 0.00895005\}\}$

Not surprisingly, the confidence intervals contain zero for the x^3, x^2, and x terms, which means that these terms probably equal zero. Also, the p-values for x^3, x^2, and x are pretty high: .19, .33, and .49 (respectively). This says that the probability that these three terms equal zero is high. This is further supported by the t-statistic for these three terms, which are small. All the evidence supports the statement that says that the x^3, x^2, and x terms equal zero.

```
cubo["RSquared"]
```

0.483196

This R^2 value is higher than the R^2 for the original dataset. This R^2 value is still pretty small, which means that the model does not do a good job of representing the variation in the data.

Maximum Data

```
maxdata = Import["C:\\Users\\Maria\\Documents\\NKU Spring 2018\\MAT 375\\MAX.xlsx"];
```

```
Years1 = Table[i, {i, 1, 660}];
```

```
maxdataclean = Flatten[maxdata];
```

```
maxdatatogether = Table[{Years1[[i]], maxdataclean[[i]]}, {i, 660}];
```

4) Determine any major outliers by finding “outer fences” for dataset.

```
Quartiles[Sort[maxdataclean]]
```

{31.0871, 33.8435, 35.8067}

```
InterquartileRange[Sort[maxdataclean]]
```

4.71957

```
31.087096774193547` - (3 * 4.7195698924731175`)
```

16.9284

```
35.806666666666665` + (3 * 4.7195698924731175`)
```

49.9654

4) We will remove any data points that are below 4.72 and any data points that are above 50.97. The data point (411,0) is certainly an outlier that we will remove. In the spirit of being **ruthless**, we considered removing the following data points: (418, 22.0), (430, 21.8), (466, 22.0), and (478, 22.3). But, we since 22.0, 21.8, 22.0, and 22.3 are above 16.9284, we decided to keep those values.

```
outlierremoved1 = Delete[maxdatatogether, 411];
```

5) We do not believe that there is a significant increase in temperature for the maximum data. The coefficient on the x^2 term is negative and the coefficient on the x^3 term is positive, but is extremely small. Also, by eyeballing the data graph with the model, it does not appear to be increasing.

6) We believe that the best model is the cubic model:
$33.5259 + 0.00919088\, x - 0.0000431869\, x^2 + 4.66797 \times 10^{-8}\, x^3 + 1.08524\, \text{Cos}\left[\frac{\pi x}{6}\right] + 3.17229\, \text{Sin}\left[\frac{\pi x}{6}\right]$. We chose this model because the confidence intervals did not include zero for any of the terms, which implies that none of the terms equal zero. This was further supported by fairly small p-values and fairly large t-statistics. The R^2 value was the highest for the cubic model. For the original dataset, the R^2 value was .5076 and for the dataset with the outlier removed, the R^2 value was .6359. Of course, we would use the dataset with the outlier removed. There was no pattern to the residuals for the cubic model.

We considered the linear and the quadratic model as well. We rejected the quadratic model because the confidence intervals included zero for the x^2 term (which would mean that that the x^2 term probably equals zero). We chose the cubic model of the linear model because the cubic model had a higher R^2 value.

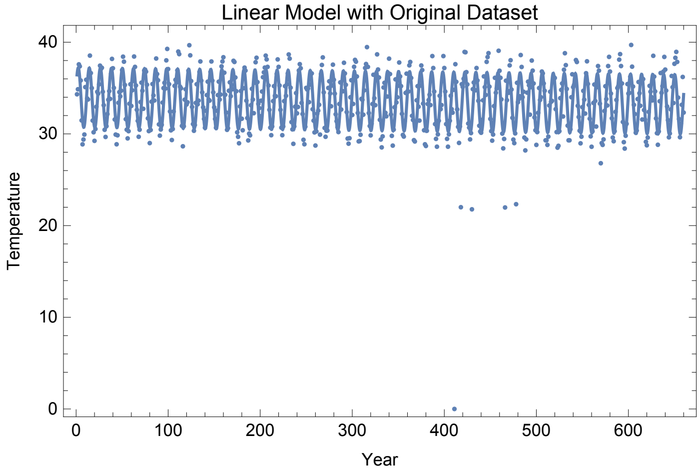
Maximum Linear Model with Sin and Cos (Original Dataset)

7) Graph of data

```
line1 = LinearModelFit[maxdatatogether, {Sin[2 * π * x / 12], Cos[2 * π * x / 12], x}, x]
```

FittedModel[$33.9646 - 0.00113761\,x + 1.08745\,\text{Cos}\left[\frac{\pi x}{6}\right] + 3.05387\,\text{Sin}\left[\frac{\pi x}{6}\right]$]

```
Show[ListPlot[maxdatatogether], Plot[line1[x], {x, 1, 660}], Frame → True,
 PlotLabel → "Linear Model with Original Dataset", FrameLabel → {"Year", "Temperature"}]
```

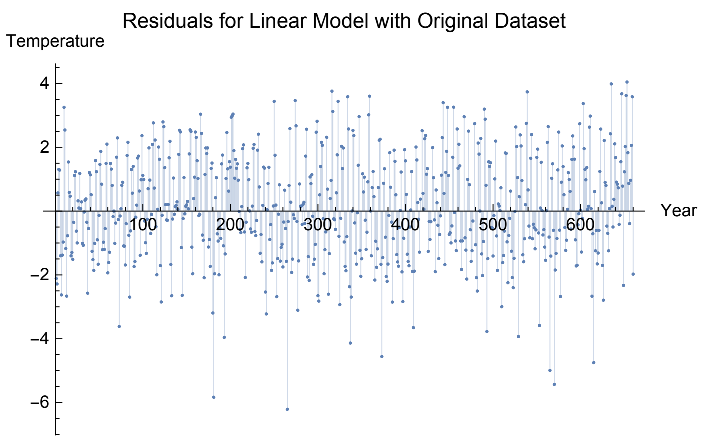



```
line1["ParameterTable"]
```

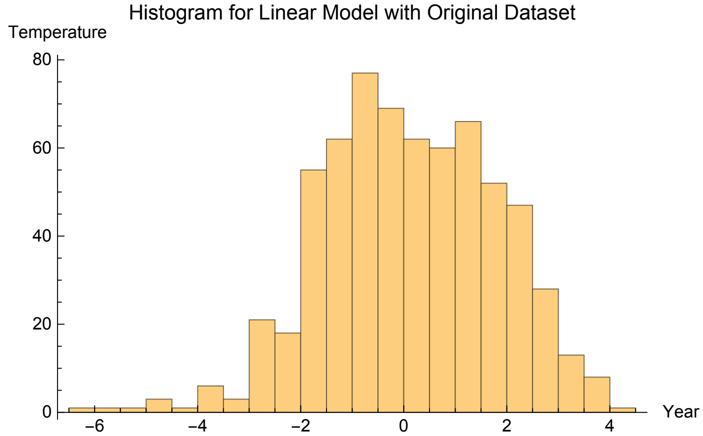
	Estimate	Standard Error	t-Statistic	P-Value
1	33.9646	0.181291	187.349	$8.81662311934 \times 10^{-572}$
$\text{Sin}\left[\frac{\pi x}{6}\right]$	3.05387	0.128049	23.8493	5.24033×10^{-91}
$\text{Cos}\left[\frac{\pi x}{6}\right]$	1.08745	0.128037	8.49321	1.3445×10^{-16}
x	−0.00113761	0.000475237	−2.39377	0.0169561

8) Graphs of Residuals

```
ListPlot[line1["FitResiduals"], Filling → Axis,
 Axes → True, AxesLabel → {"Year", "Temperature"},
 PlotLabel → "Residuals for Linear Model with Original Dataset"]
```



```
Histogram[line1["FitResiduals"], AxesLabel → {"Year", "Temperature"},
 PlotLabel → "Histogram for Linear Model with Original Dataset"]
```



From the list-plot of the residuals, I do not see a super prominent pattern or smiley face. From the histogram, the distribution is not as normal as we would like to see.

```
line1["ParameterConfidenceIntervals"]
```

```
{{33.6086, 34.3205}, {2.80244, 3.30531}, {0.836035, 1.33886}, {-0.00207078, -0.000204439}}
```

The confidence intervals do not contain zero which means that the terms probably do not equal zero. The p-values are less than .05 for all the terms. The t-statistics are pretty large--the only concern is the t-statistic for x which is -2.39. We would like for this to be larger, but the probability that the x does not equal zero is still pretty high.

```
line1["RSquared"]
```

```
0.497009
```

This R^2 value is smaller than we would like to see. This means that our model is not representing the variation in our data very well.

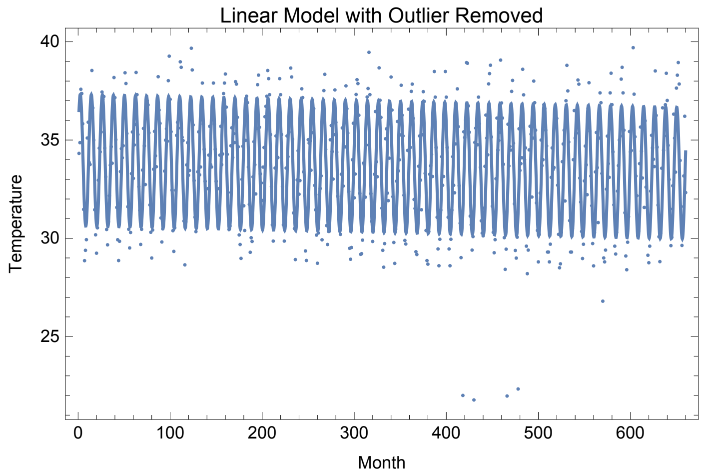
Maximum Linear Model with Sin and Cos (Outlier Removed)

7) Graph of data

```
lineo1 = LinearModelFit[outlierremoved1, {Sin[2 * π * x / 12], Cos[2 * π * x / 12], x}, x]
```

FittedModel[$33.9775 - 0.00100845\,x + 1.08732\,\mathrm{Cos}\left[\frac{\pi x}{6}\right] + 3.16565\,\mathrm{Sin}\left[\frac{\pi x}{6}\right]$]

```
Show[ListPlot[outlierremoved1], Plot[lineo1[x], {x, 1, 660}], Frame → True,
 PlotLabel → "Linear Model with Outlier Removed", FrameLabel → {"Month", "Temperature"}]
```

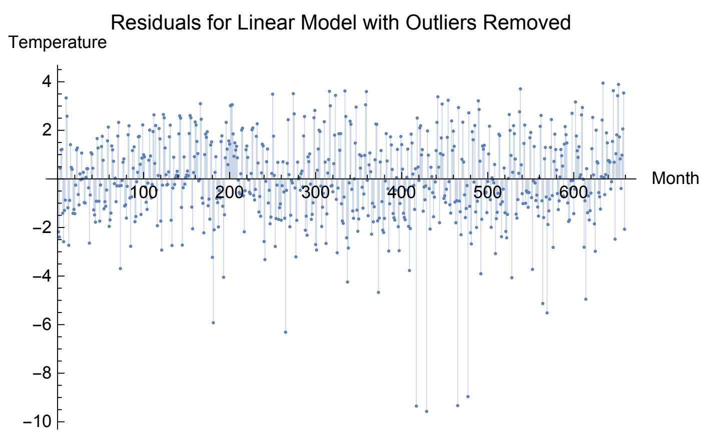



```
lineo1["ParameterTable"]
```

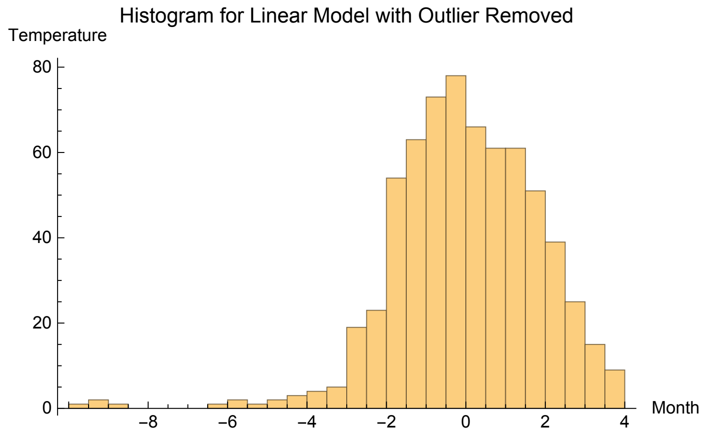
	Estimate	Standard Error	t-Statistic	P-Value
1	33.9775	0.143056	237.512	$1.258018528274 \times 10^{-637}$
$\text{Sin}\left[\frac{\pi x}{6}\right]$	3.16565	0.101197	31.2821	4.17819×10^{-132}
$\text{Cos}\left[\frac{\pi x}{6}\right]$	1.08732	0.101033	10.762	5.53766×10^{-25}
x	−0.00100845	0.000375061	−2.68876	0.00735419

8) Graphs of Residuals

```
ListPlot[lineo1["FitResiduals"], Filling → Axis,
 Axes → True, AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Residuals for Linear Model with Outliers Removed"]
```



```
Histogram[lineo1["FitResiduals"], AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Histogram for Linear Model with Outlier Removed"]
```



Like the previous residual, there is not a prominent pattern. The histogram looks more normal than the histogram for the original data set.

```
lineo1["ParameterConfidenceIntervals"]
```

```
{{33.6966, 34.2584}, {2.96694, 3.36436}, {0.888931, 1.28571}, {-0.00174492, -0.000271983}}
```

The confidence intervals do not contain zero which means that the terms probably do not equal zero. The p-value and the t-statistic look better for this data set. The p-value got smaller and the t-statistic got larger. So, the probability that the x term equals zero is higher for this dataset than for the original dataset.

```
lineo1["RSquared"]
```

```
0.627652
```

This R^2 value is significantly higher than the R^2 value for the original dataset. This is the best R^2 value we have had yet! It still is not as high as we would like to see.

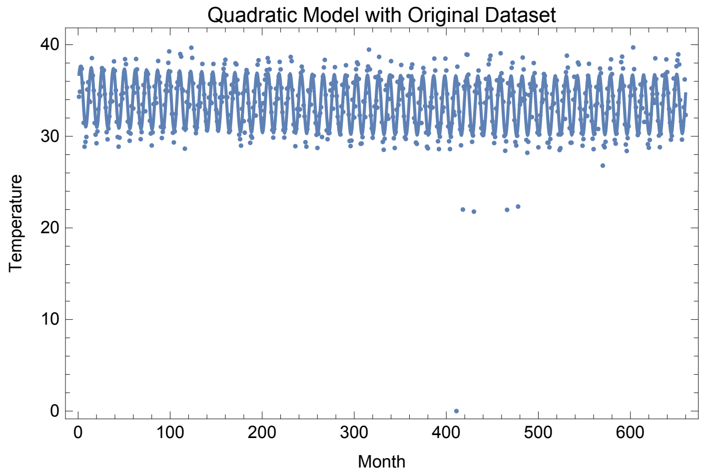
Maximum Quadratic Model with Sin and Cos (Original Dataset)

7) Graph of data

```
quad1 = LinearModelFit[maxdatatogether, {Sin[2 * π * x / 12], Cos[2 * π * x / 12], x^2, x}, x]
```

FittedModel[$34.3042 - 0.00421609\,x + 4.6573\times10^{-6}\,x^2 + 1.08738\,\mathrm{Cos}\left[\frac{\pi x}{6}\right] + 3.05386\,\mathrm{Sin}\left[\frac{\pi x}{6}\right]$]

```
Show[ListPlot[maxdatatogether], Plot[quad1[x], {x, 1, 660}],
 Frame → True, PlotLabel → "Quadratic Model with Original Dataset",
 FrameLabel → {"Month", "Temperature"}]
```

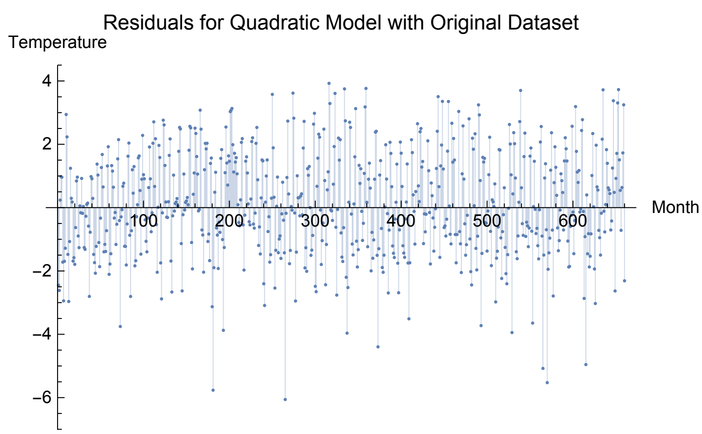



```
quad1["ParameterTable"]
```

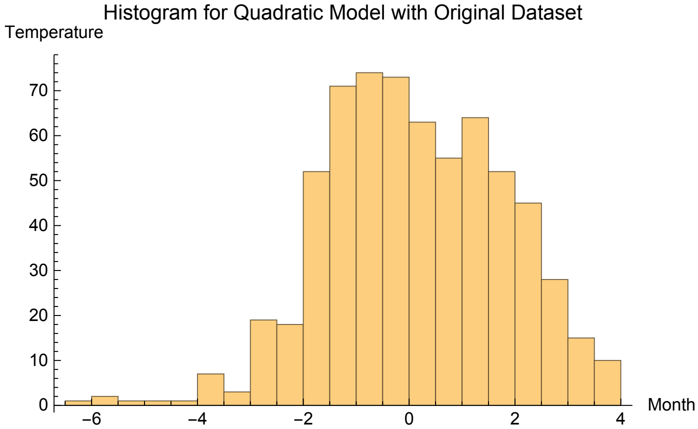
	Estimate	Standard Error	t-Statistic	P-Value
1	34.3042	0.272068	126.087	$1.380899963598 \times 10^{-461}$
$\text{Sin}\left[\frac{\pi x}{6}\right]$	3.05386	0.127874	23.8818	3.74745×10^{-91}
$\text{Cos}\left[\frac{\pi x}{6}\right]$	1.08738	0.127862	8.50433	1.23674×10^{-16}
x^2	4.6573×10^{-6}	2.78468×10^{-6}	1.67247	0.0949084
x	−0.00421609	0.00190087	−2.21798	0.0268987

8) Graphs of Residuals

```
ListPlot[quad1["FitResiduals"], Filling → Axis,
 Axes → True, AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Residuals for Quadratic Model with Original Dataset"]
```



```
Histogram[quad1["FitResiduals"], AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Histogram for Quadratic Model with Original Dataset"]
```



The residuals look look random and independent. The histogram is not as normal as we would like to see. There is a slight “shrug” on the right side.

```
quad1["ParameterConfidenceIntervals"]
```

{{33.77, 34.8385}, {2.80276, 3.30495}, {0.836314, 1.33845},
 {-8.10671×10^{-7}, 0.0000101253}, {-0.00794862, -0.000483551}}

The confidence interval contains 0 for the x^2 term, which means that the x^2 probably equals zero. Since the p-value for x^2 is around .09 (which is larger than .05), the probability that the x^2 is zero is fairly high. The statement that the x^2 term equals zero is further supported by the fairly small t-statistic, which is 1.67. The evidence points to the conclusion that the x^2 probably equals zero.

```
quad1["RSquared"]
```

0.499147

This R^2 value is pretty small. We would like to see a larger R^2 value.

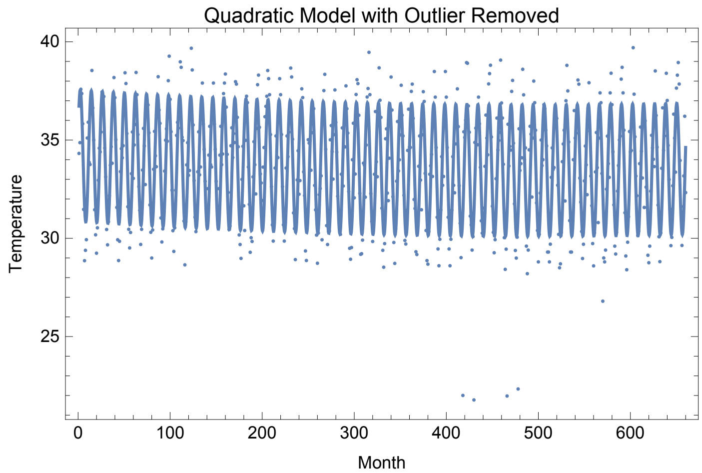
Maximum Quadratic Model with Sin and Cos (Outlier Removed)

7) Graph of data

```
quado1 = LinearModelFit[outlierremoved1, {Sin[2 * π * x / 12], Cos[2 * π * x / 12], x^2, x}, x]
```

FittedModel[$34.2026 - 0.00304918\,x + 3.08685 \times 10^{-6}\,x^2 + 1.08728\,\mathrm{Cos}\left[\frac{\pi x}{6}\right] + 3.16536\,\mathrm{Sin}\left[\frac{\pi x}{6}\right]$]

```
Show[ListPlot[outlierremoved1], Plot[quado1[x], {x, 1, 660}], Frame → True,
 PlotLabel → "Quadratic Model with Outlier Removed", FrameLabel → {"Month", "Temperature"}]
```

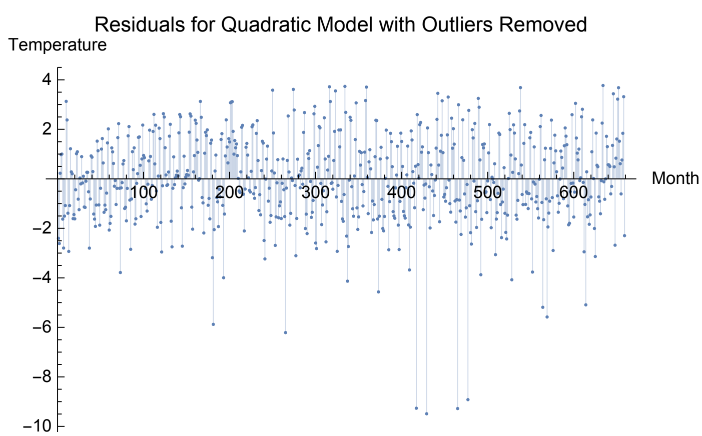



```
quado1["ParameterTable"]
```

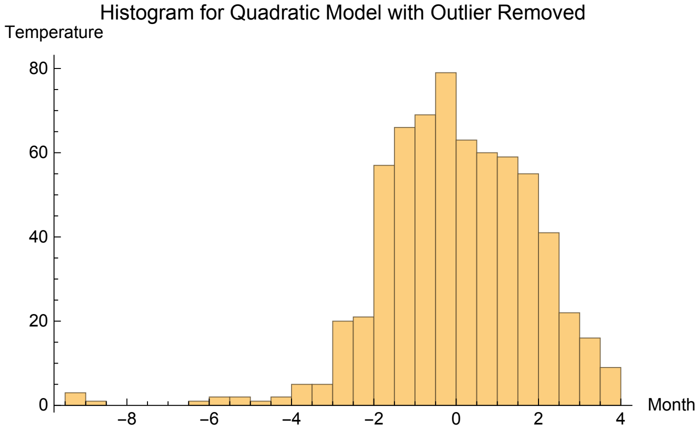
	Estimate	Standard Error	t-Statistic	P-Value
1	34.2026	0.214882	159.169	$3.615500972039 \times 10^{-525}$
$\text{Sin}\left[\frac{\pi x}{6}\right]$	3.16536	0.101123	31.3022	3.80044×10^{-132}
$\text{Cos}\left[\frac{\pi x}{6}\right]$	1.08728	0.100958	10.7695	5.19934×10^{-25}
x^2	3.08685×10^{-6}	2.20016×10^{-6}	1.40301	0.161088
x	−0.00304918	0.00150205	−2.03002	0.042759

8) Graphs of Residuals

```
ListPlot[quado1["FitResiduals"], Filling → Axis,
 Axes → True, AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Residuals for Quadratic Model with Outliers Removed"]
```



```
Histogram[quado1["FitResiduals"], AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Histogram for Quadratic Model with Outlier Removed"]
```



The residuals look random. The histogram is not as normal as we would like. There is a “shrug” on the left side of the histogram. It does look more normal than the previous histogram.

```
quado1["ParameterConfidenceIntervals"]
```

$\{\{33.7807, 34.6246\}, \{2.9668, 3.36392\}, \{0.889034, 1.28552\}, \{-1.23338\times10^{-6}, 7.40708\times10^{-6}\}, \{-0.0059986, -0.000099766\}\}$

The confidence interval contains 0 for the x^2 term, which means that the x^2 probably equals zero. Since the p-value for x^2 is around .16, the probability that the x^2 is zero is fairly high. The statement that the x^2 term equals zero is further supported by the fairly small t-statistic, which is 1.4. The evidence points to the conclusion that the x^2 probably equals zero. The confidence interval for the x term does not include zero, but the p-value and the t-statistic for the x term indicate that the probability that the x term equals zero is somewhat high.

```
quado1["RSquared"]
```

0.62877

This R^2 value is significantly higher than the R^2 value for the original dataset.

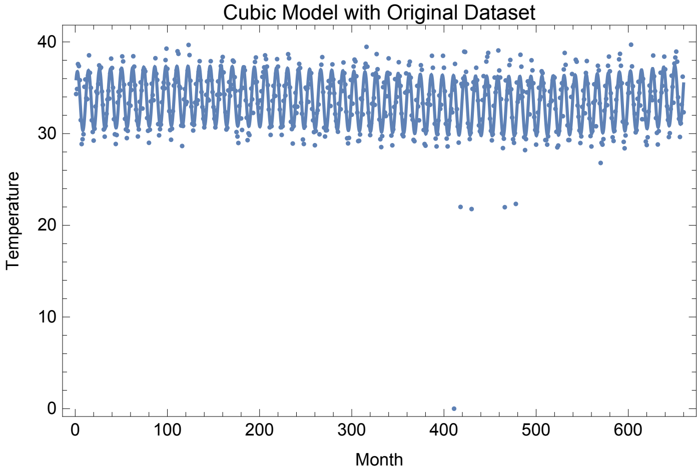
Maximum Cubic Model with Sin and Cos (Original Dataset)

7) Graph of data

```
cub1 = LinearModelFit[maxdatatogether, {Sin[2 * π * x / 12], Cos[2 * π * x / 12], x^3, x^2, x}, x]
```

FittedModel[$33.5028 + 0.0102758\,x - 0.0000501098\,x^2 + 5.52366\times10^{-8}\,x^3 + 1.08498\,\mathrm{Cos}\left[\frac{\pi x}{6}\right] + 3.06283\,\mathrm{Sin}\left[\frac{\pi x}{6}\right]$]

```
Show[ListPlot[maxdatatogether], Plot[cub1[x], {x, 1, 660}], Frame → True,
 PlotLabel → "Cubic Model with Original Dataset", FrameLabel → {"Month", "Temperature"}]
```

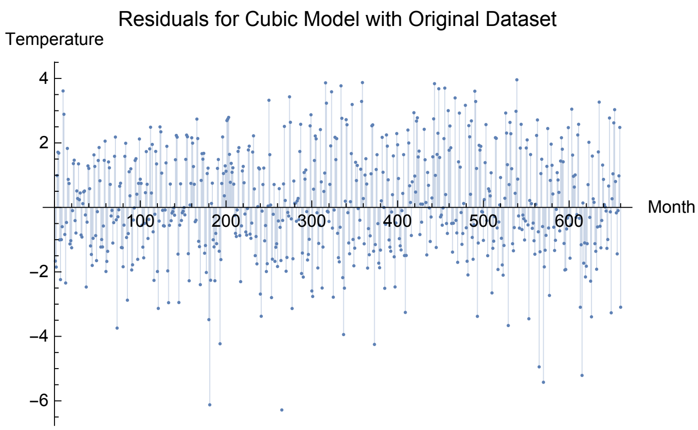



```
cub1["ParameterTable"]
```

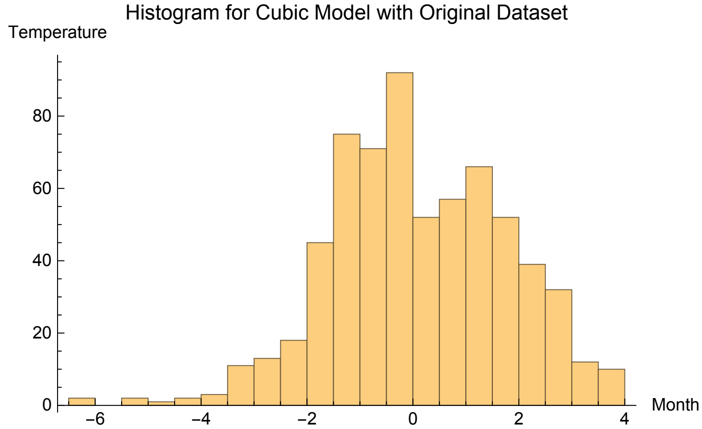
	Estimate	Standard Error	t-Statistic	P-Value
1	33.5028	0.360994	92.8072	$1.027904336503\times10^{-378}$
$\mathrm{Sin}[\frac{\pi x}{6}]$	3.06283	0.126919	24.1321	1.65714×10^{-92}
$\mathrm{Cos}[\frac{\pi x}{6}]$	1.08498	0.126882	8.55109	8.62045×10^{-17}
x^3	5.52366×10^{-8}	1.6517×10^{-8}	3.34422	0.000872271
x^2	−0.0000501098	0.0000166081	−3.01718	0.00265046
x	0.0102758	0.00472615	2.17424	0.0300447

8) Graphs of Residuals

```
ListPlot[cub1["FitResiduals"], Filling → Axis,
 Axes → True, AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Residuals for Cubic Model with Original Dataset"]
```



```
Histogram[cub1["FitResiduals"], AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Histogram for Cubic Model with Original Dataset"]
```



The residuals look independent. The histogram looks more normal on the left side than on the right side.

```
cub1["ParameterConfidenceIntervals"]
```

```
{{32.794, 34.2116}, {2.81361, 3.31205}, {0.835833, 1.33412},
 {2.28038 × 10^-8, 8.76694 × 10^-8}, {-0.0000827215, -0.0000174981}, {0.00099554, 0.019556}}
```

The confidence intervals do not contain zero for any of the terms. Also, the p-values are pretty small and the t-statistics are pretty large. So, the probability that any term equals zero is pretty low.

```
cub1["RSquared"]
```

```
0.507568
```

This R^2 value is pretty small. Since R^2 represents the total amount of variation that the model explains, we want the R^2 to be high. The low R^2 indicates that this model does not do a good job of representing the variation in our data.

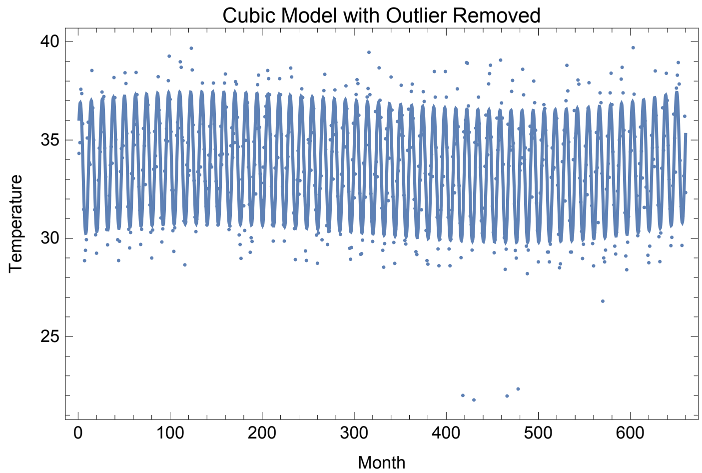
Maximum Cubic Model with Sin and Cos (Outlier Removed)

7) Graph of data

```
cubo1 = LinearModelFit[outlierremoved1, {Sin[2 * π * x / 12], Cos[2 * π * x / 12], x^3, x^2, x}, x]
```

FittedModel[$33.5259 + 0.00919088\,x - 0.0000431869\,x^2 + 4.66797 \times 10^{-8}\,x^3 + 1.08524\,\mathrm{Cos}\left[\frac{\pi x}{6}\right] + 3.17229\,\mathrm{Sin}\left[\frac{\pi x}{6}\right]$]

```
Show[ListPlot[outlierremoved1], Plot[cubo1[x], {x, 1, 660}], Frame → True,
 PlotLabel → "Cubic Model with Outlier Removed", FrameLabel → {"Month", "Temperature"}]
```

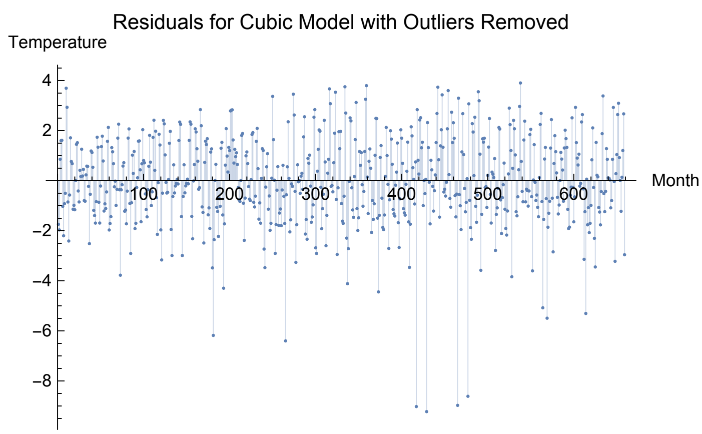



```
cubo1["ParameterTable"]
```

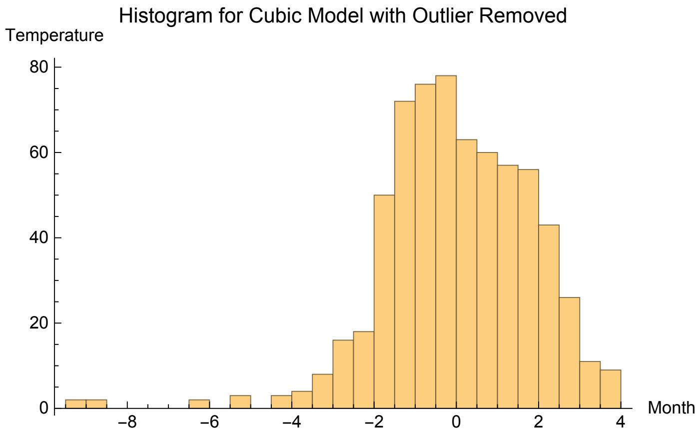
	Estimate	Standard Error	t-Statistic	P-Value
1	33.5259	0.284683	117.766	$4.713104664250 \times 10^{-442}$
$\text{Sin}\left[\frac{\pi x}{6}\right]$	3.17229	0.100239	31.6473	5.92546×10^{-134}
$\text{Cos}\left[\frac{\pi x}{6}\right]$	1.08524	0.100059	10.846	2.58041×10^{-25}
x^3	4.66797×10^{-8}	1.30324×10^{-8}	3.58182	0.000366575
x^2	−0.0000431869	0.0000131018	−3.29626	0.00103288
x	0.00919088	0.00372744	2.46573	0.0139294

8) Graphs of Residuals

```
ListPlot[cubo1["FitResiduals"], Filling → Axis,
 Axes → True, AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Residuals for Cubic Model with Outliers Removed"]
```



```
Histogram[cubo1["FitResiduals"], AxesLabel → {"Month", "Temperature"},
 PlotLabel → "Histogram for Cubic Model with Outlier Removed"]
```




```
cubo1["ParameterConfidenceIntervals"]
```

$\{\{32.9669, 34.0849\}, \{2.97546, 3.36912\}, \{0.888767, 1.28172\}, \{2.10892\times10^{-8}, 7.22703\times10^{-8}\}, \{-0.0000689136, -0.0000174602\}, \{0.00187165, 0.0165101\}\}$

None of the confidence intervals include zero for any of the terms. Also, the p-values have gotten smaller with this dataset than with the previous dataset and the t-statistics have gotten larger. This indicates that the none of the terms probably equal zero.

```
cubo1["RSquared"]
```

0.635923

This R^2 value is significantly larger than the R^2 for the original dataset. The model represents the variation in this dataset better than in the other dataset. Again, this makes sense since we have removed the outliers in this dataset.

9) We could always use finer data to strengthen our observations. Daily temperature data would give us a closer look at the temperature trends. Hourly temperature data would be even better for the same reasons. Having data that goes further back in time would help us have a better perspective of the data. However, it may not be possible to obtain this, and the measurements become more unreliable. Other kinds of data to consider would be overcast data, humidity levels, CO2 levels, and precipitation.

The accuracy of the data could be improved as well. There are two occurrences in the maximum data where two consecutive years have identical data for each month (1972-1973 & 1993-1994). The same problem also occurred in the minimum data. The same years (1972-1973 & 1993-1994) again have the exact data for each month. It is nearly impossible that the data for each month in both the years would be the same. The fact that this issue has occurred twice in both the minimum and maximum data for the same years makes us question how reliable the data is. If we are using data that is questionable from the start, our models will probably not be accurate or representative of what actually happened. There was also a 0 degrees Celsius temperature recorded in March 1995 in the monthly maximum dataset, which is clearly not accurate. If there are at least 4 rows that are obviously inaccurate, there could also be a mess of other errors and inaccuracies in the data.

In both minimum and maximum datasets respectively, some data values have one or two decimal places while others have up to 13 places, which causes the data overall to be inconsistent. This lack of precision can cause models to become less accurate, and it also makes us curious about how the temperatures were measured and recorded. We also wonder if the temperatures were taken by the same person each time. If the temperatures were recorded by different people, there is a possibility that the temperatures were measured and recorded differently each time. This may have caused some issues in the data, such as the precision and accuracy.