Mini-Project 2: Niamtougou

Donna Odhiambo & Matthew Gall

Niamtougou is located in the Kara region of Togo and is composed of six villages: Niamtougou, Koka, Baga, Ténéga, Yaka, and Agbandé. While Togo is classified as a tropical, sub-Saharan country its length stretches it through six geographic regions, making its climate vary from tropical to savanna. The Kara region is in northern Togo, and thus Niamtougou has a tropical climate with an average elevation of 1535 ft.

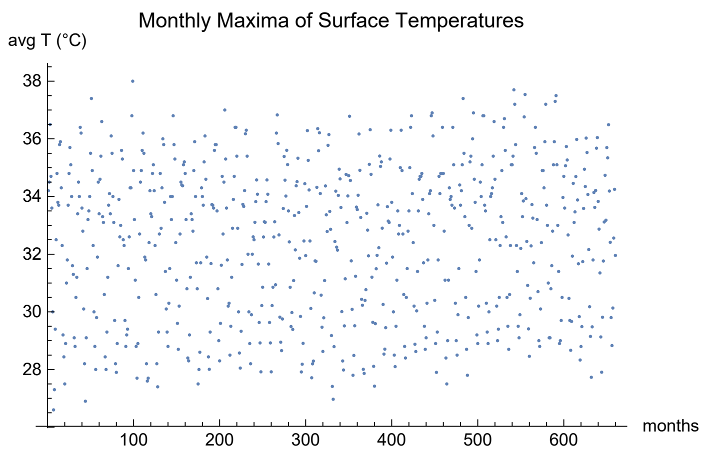
The purpose of this project is to analyze monthly surface temperature information from Niamtougou from years 1961-2015 and determine if the maximum and minimum temperature time-series data demonstrate significant increases over time. This question was explored in a preliminary fashion using yearly maximum and minimum temperature time-series by a previous group for Mini-Project 1. They found the data gave indeterminate results and lacked true discernible form. Those findings in mind, our group's first task was to corroborate the original group's yearly values with our own. Shown below are two matrices, the first containing comparisons of the minimum temperatures and the second containing comparisons of the maximum, under "1st data set" is the original group's data and under "2nd data set" is ours. The numbers from each data set that correspond to each other match, which gives us more confidence that we are receiving accurate and reliable data for the annual average minimum temperatures and annual average maximum temperatures for our city.

years	1st data set	2nd data set
1961.	20.85	20.85
1962.	20.8692	20.8692
1963.	21.2833	21.2833
1964.	21.2417	21.2417
1965.	20.9	20.9
1966.	19.9418	19.9418
1967.	20.65	20.65
1968.	20.7	20.7
1969.	21.1	21.1
1970.	21.0583	21.0583
1971.	20.6667	20.6667
1972.	20.8333	20.8333
1973.	21.15	21.15
1974.	20.7071	20.7071
1975.	20.075	20.075
1976.	20.3918	20.3918
1977.	21.3607	21.3607
1978.	21.1217	21.1217
1979.	21.2371	21.2371
1980.	21.5605	21.5605
1981.	20.9539	20.9539
1982.	20.9539	20.9539
1983.	20.9934	20.9934
1984.	20.55	20.55
1985.	20.7837	20.7837
1986.	20.5627	20.5627
1987.	21.1867	21.1867
1988.	21.0154	21.0154
1989.	20.6226	20.6226
1990.	21.1113	21.1113
1991.	20.9014	20.9014
1992.	20.5399	20.5399
1993.	20.6467	20.6467
1994.	20.5833	20.5833
1995.	20.9	20.9
1996.	21.05	21.05
1997.	21.1417	21.1417
1998.	21.7083	21.7083
1999.	20.85	20.85
2000.	20.7167	20.7167
2001.	20.925	20.925
2002.	21.1667	21.1667
2003.	20.0583	20.0583
2004.	20.9167	20.9167
2005.	21.3583	21.3583
2006.	21.275	21.275
2007.	21.0856	21.0856
2008.	20.775	20.775
2009.	21.2667	21.2667
2010.	21.7	21.7
2011.	21.0622	21.0622
2012.	20.0458	20.0458
2013.	20.4383	20.4383
2014.	20.6394	20.6394
2015.	20.8164	20.8164

years	1st data set	2nd data set
1961.	32.325	32.325
1962.	31.8782	31.8782
1963.	32.5	32.5
1964.	32.125	32.125
1965.	32.5417	32.5417
1966.	32.3615	32.3615
1967.	32.075	32.075
1968.	32.0167	32.0167
1969.	32.6167	32.6167
1970.	32.45	32.45
1971.	32.1333	32.1333
1972.	32.375	32.375
1973.	32.9583	32.9583
1974.	32.2643	32.2643
1975.	32.0167	32.0167
1976.	32.2332	32.2332
1977.	32.5849	32.5849
1978.	32.3917	32.3917
1979.	32.565	32.565
1980.	32.3483	32.3483
1981.	31.7086	31.7086
1982.	31.7086	31.7086
1983.	32.7656	32.7656
1984.	32.2443	32.2443
1985.	31.9346	31.9346
1986.	31.6985	31.6985
1987.	32.8213	32.8213
1988.	31.8567	31.8567
1989.	32.014	32.014
1990.	32.1795	32.1795
1991.	31.8627	31.8627
1992.	31.7726	31.7726
1993.	32.184	32.184
1994.	31.9833	31.9833
1995.	32.3333	32.3333
1996.	32.6611	32.6611
1997.	32.225	32.225
1998.	32.8417	32.8417
1999.	32.4083	32.4083
2000.	32.5083	32.5083
2001.	32.9917	32.9917
2002.	32.85	32.85
2003.	32.9167	32.9167
2004.	32.75	32.75
2005.	33.0167	33.0167
2006.	33.1583	33.1583
2007.	32.7522	32.7522
2008.	32.775	32.775
2009.	32.8	32.8
2010.	33.0806	33.0806
2011.	32.5269	32.5269
2012.	32.2006	32.2006
2013.	32.3522	32.3522
2014.	32.5372	32.5372
2015.	32.9033	32.9033

Maximum Values (1961-2015)

Out[3]=



In[4]:=
```
LM = LinearModelFit[max, {x, x^2, Sin[2π/12 x], Cos[2π/12 x]}, x, ConfidenceLevel → 0.95];
LM["ParameterTable"]
LM["ParameterConfidenceIntervals"]
```

Out[5]=

	Estimate	Standard Error	t-Statistic	P-Value
1	32.3438	0.126601	255.478	$3.787643986680 \times 10^{-658}$
x	−0.00119522	0.00088453	−1.35126	0.17708
x^2	3.11129×10^{-6}	1.29579×10^{-6}	2.40107	0.0166249
$\mathrm{Sin}[\frac{\pi x}{6}]$	2.86953	0.0595032	48.2247	1.07376×10^{-217}
$\mathrm{Cos}[\frac{\pi x}{6}]$	1.81988	0.0594979	30.5873	2.61645×10^{-128}

Out[6]= $\{\{32.0952, 32.5924\}, \{-0.00293208, 0.00054163\}, \{5.66881 \times 10^{-7}, 5.65569 \times 10^{-6}\}, \{2.75269, 2.98637\}, \{1.70305, 1.93671\}\}$

Our first model used x and x^2 terms just try to see what kind of a fit it would give considering this was the model the previous group used for the maximum data and it was somewhat successful for them as well as $\sin(\frac{2\pi}{12}x)$ and $\cos(\frac{2\pi}{12}x)$ terms to account for the periodicity we expected to see due to seasonal change. Examining the model plotted over the data, the fit seemed somewhat successful, however the parameter table showed the t-statistic for the x value was not extreme, the p-value was a little too high for our liking, and the confidence interval included the probability of the parameter being zero. We then attempted two more fits, one excluding the x^2 term and one excluding the x term. Shown below is LM1, the fit excluding the x^2 term.

In[7]:=
```
LM1 = LinearModelFit[max, {x, Sin[2 π x / 12], Cos[2 π x / 12]}, x, ConfidenceLevel → 0.95]
LM1["ParameterTable"]
LM1["AdjustedRSquared"]
LM1["ParameterConfidenceIntervals"]
ListPlot[LM1["FitResiduals"], PlotLabel → "Residuals for Max",
 Filling → Axis, AxesLabel → {"Fitted Value", "Residual"}]
Histogram[LM1["FitResiduals"], PlotLabel → "Histogram of Max Residuals",
 AxesLabel → {"Residual", "Frequency"}]
```

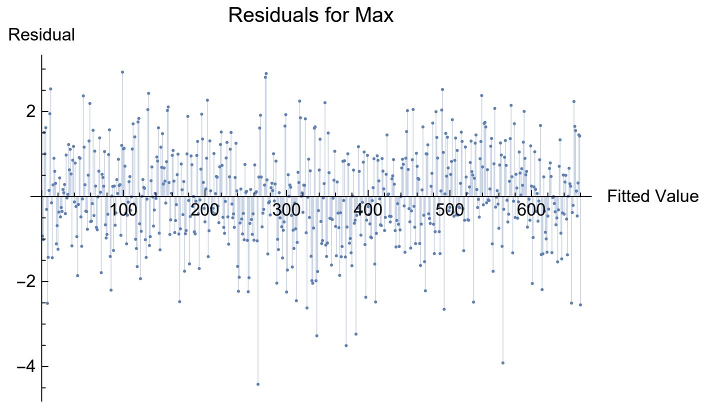
Out[7]= FittedModel[$32.1169 + 0.000861335\,x + 1.81993\,\mathrm{Cos}\left[\frac{\pi x}{6}\right] + 2.86954\,\mathrm{Sin}\left[\frac{\pi x}{6}\right]$]

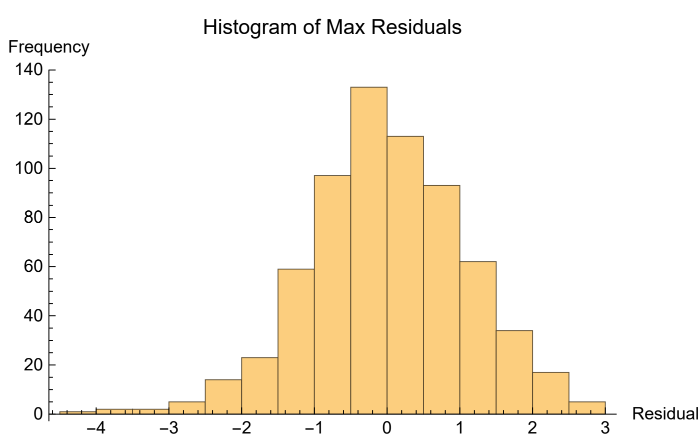
Out[8]=

	Estimate	Standard Error	t-Statistic	P-Value
1	32.1169	0.0845499	379.857	$3.645886231902 \times 10^{-771}$
x	0.000861335	0.00022164	3.88619	0.000112172
$\mathrm{Sin}\left[\frac{\pi x}{6}\right]$	2.86954	0.059719	48.0507	4.72185×10^{-217}
$\mathrm{Cos}\left[\frac{\pi x}{6}\right]$	1.81993	0.0597137	30.4775	9.00432×10^{-128}

Out[9]= 0.831243

Out[10]= {{31.9508, 32.2829}, {0.000426126, 0.00129654}, {2.75227, 2.9868}, {1.70267, 1.93718}}

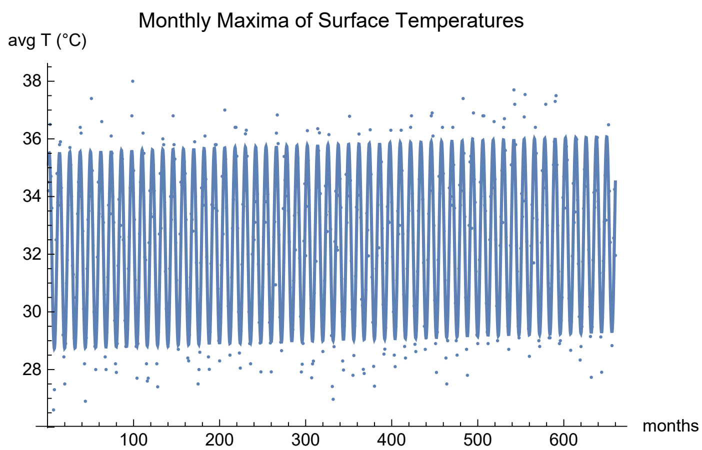
Out[11]=


Out[12]=


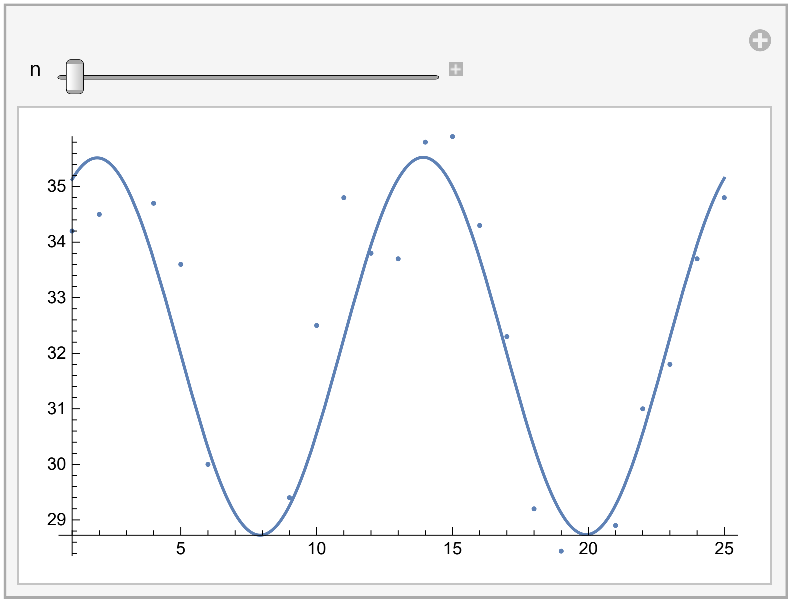
The parameter table for our new model with just x, sine, and cosine showed all of our parameters to be significant, since they had higher t-statistics, smaller p-values, and confidence intervals that excluded zero, so that was a good sign so we then plotted our residuals for the max and saw that they had no pattern and the residuals were not very extreme. Due to the sheer size of the dataset, a clearer picture of the residuals was generated using histogram format. The histogram showed the model was a good fit due to the normal distribution about 0 of the residuals. It also brought five data points (those left of -3) under suspicion, however five data points in a set of 660 was not enough to raise real alarm. The model with x^2, sine, and cosine gave very similar results concerning R^2 values and residuals, so we decided to keep it simple and use the model with just the x.

The data is so scattered that we knew when we were looking for a model it was not going to be a perfect model. When interpreting our parameters we see that we predict that in January of 1961 we should have 32.1169 ppm of CO_2 Niamtougou and then our coefficients are giving us as amplitude and also give us a wave curve. Overall, there is a significant increase in the temperature over time due to the fact that we have good fit for our model and the x coefficient is positive and significant. The x coefficient is 0.000861335. This number is the slope and shows us how much we predict the temperature to be changing month to month. Below is the model placed over the maximum data, while examining the goodness of fit seems difficult from afar, zooming in on a small range of data points reveals the periodic nature of the dataset as well as how good our fit was.

Out[13]=

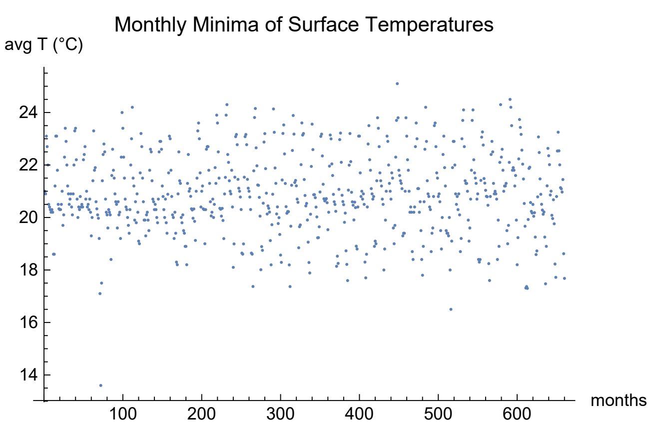


Out[14]=



Minimum Values

Out[17]=



In[18]:=
```
LM2 = LinearModelFit[min, {x, Sin[2 π x / 12], Cos[2 π x / 12]}, x, ConfidenceLevel → 0.95]
LM2["ParameterTable"]
LM2["ParameterConfidenceIntervals"]
```

Out[18]= FittedModel[$20.8637 + 0.0000822851\,x - 1.18964\,\mathrm{Cos}\left[\frac{\pi x}{6}\right] + 1.07753\,\mathrm{Sin}\left[\frac{\pi x}{6}\right]$]

Out[19]=

	Estimate	Standard Error	t-Statistic	P-Value
1	20.8637	0.0811293	257.166	$8.58284823559\times10^{-661}$
x	0.0000822851	0.000212673	0.386909	0.698949
$\mathrm{Sin}\left[\frac{\pi x}{6}\right]$	1.07753	0.057303	18.8041	2.01356×10^{-63}
$\mathrm{Cos}\left[\frac{\pi x}{6}\right]$	−1.18964	0.0572978	−20.7623	5.55514×10^{-74}

Out[20]= {{20.7044, 21.023}, {−0.000335317, 0.000499887}, {0.965014, 1.19005}, {−1.30215, −1.07713}}

For the minimum values, we tried starting out with a linear model to see what we would get. However, the t-statistic for the *x* value was not extreme, the p-value was too high for our liking, and the confidence interval included the probability of the parameter being 0. This meant the x term was most likely not significant and therefore we decided to make our model without x in it.

In[21]:=
```
LM3 = LinearModelFit[min, {Sin[2 π x/12], Cos[2 π x/12]}, x, ConfidenceLevel → 0.95]
LM3["ParameterTable"]
LM3["AdjustedRSquared"]
LM3["ParameterConfidenceIntervals"]
ListPlot[LM3["FitResiduals"], PlotLabel → "Residuals for Min", Filling → Axis,
 AxesLabel → {"Fitted Value", "Residual"}]
Histogram[LM3["FitResiduals"],
 PlotLabel → "Histogram of Max Residuals", AxesLabel → {"Residual", "Frequency"}]
```

Out[21]= FittedModel[$20.8909 - 1.18955 \cos\left[\frac{\pi x}{6}\right] + 1.07723 \sin\left[\frac{\pi x}{6}\right]$]

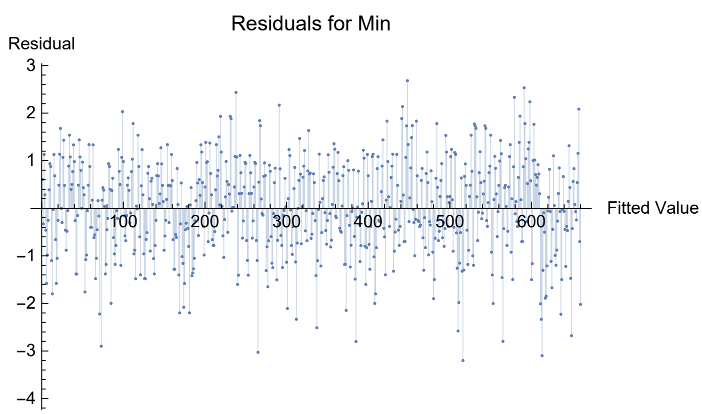
Out[22]=

	Estimate	Standard Error	t-Statistic	P-Value
1	20.8909	0.0404892	515.962	$3.327192082443 \times 10^{-859}$
$\sin\left[\frac{\pi x}{6}\right]$	1.07723	0.0572604	18.8128	1.73937×10^{-63}
$\cos\left[\frac{\pi x}{6}\right]$	−1.18955	0.0572604	−20.7745	4.51531×10^{-74}

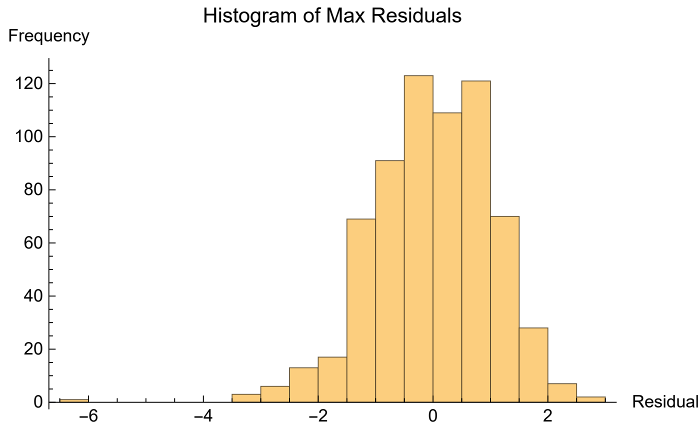
Out[23]= 0.543154

Out[24]= {{20.8114, 20.9704}, {0.964791, 1.18966}, {-1.30199, -1.07712}}

Out[25]=

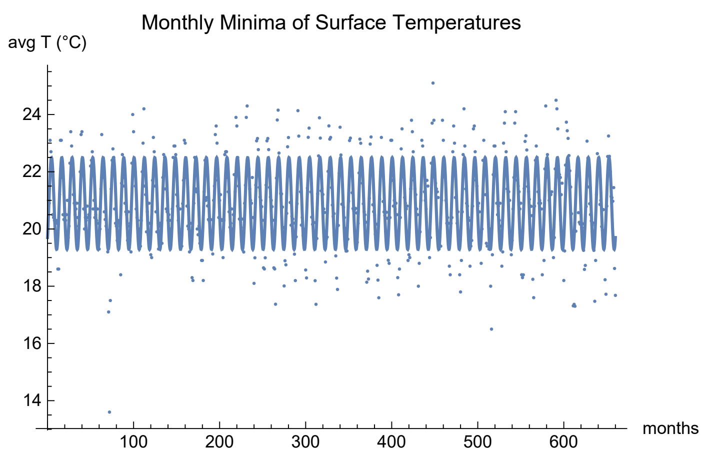


Out[26]=

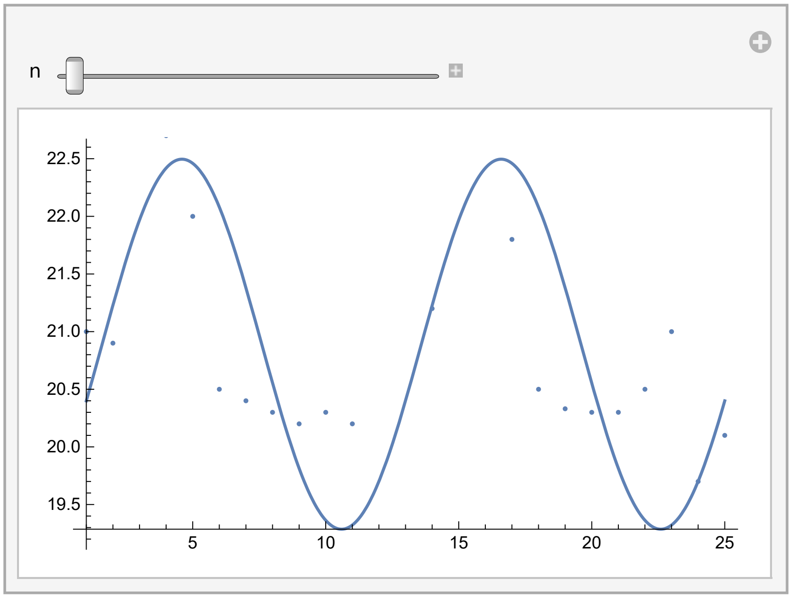


This then lead us to the conclusion that for Niamtougou there was not a significant increase in temperature over the data set for minimum temperatures, just fluctuation. The fluctuation between seasons was expected and accounted for using sines and cosines in our model. After graphing this model and looking at the parameters, residuals, and the histogram plot of the residuals, we were happy to see the parameters were significant and that the residuals seemed independent and randomly distributed. The histogram plot also had normal distribution centered about zero which made us confident in the model we choose. It also brought one data points (the one to the far left) under suspicion, however one “off” data point in a set of 660 was not enough to raise real alarm. Below is the model placed over the minimum data, while examining the goodness of fit seems difficult from afar, zooming in on a small range of data points reveals the periodic nature of the dataset as well as how good our fit was.

Out[27]=



Out[28]=



We were very pleased with the data we were able to get. However we had to do quite a bit of digging to get the data in a workable form since it was so scattered. Of the Togolese, we would like ask why the data was so scattered. We would also like to hear them explain their data collection methods, is it possible that the data is being collected at different points in the Kara region or just in Niamtougou? There is a weather station (DXNG) located in the Niamtougou International Airport, is this where data might be collected and if so, using what instrumentation? When they report the monthly average, are they reporting an average calculated using the average minimum and maximum of each day, or are they using a different time span, or perhaps a different method entirely?