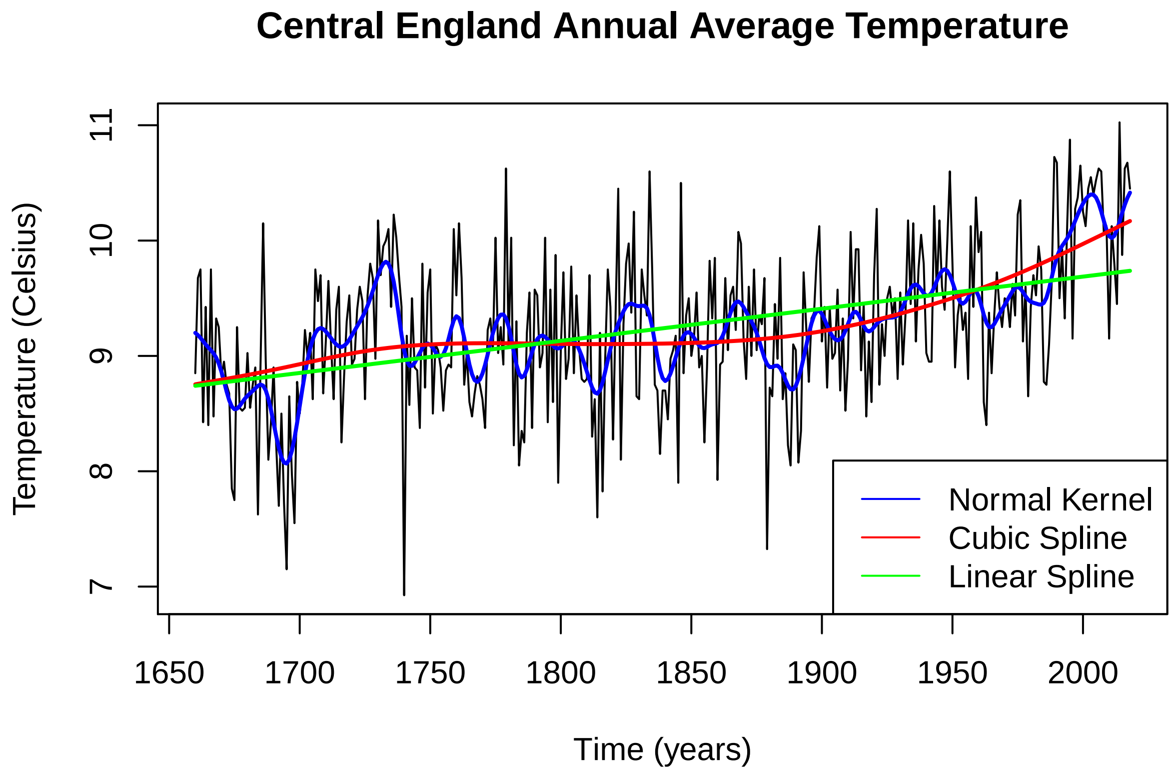
Trend Analysis of Central England Temperature Data

Tyson Miller

12/5/2019

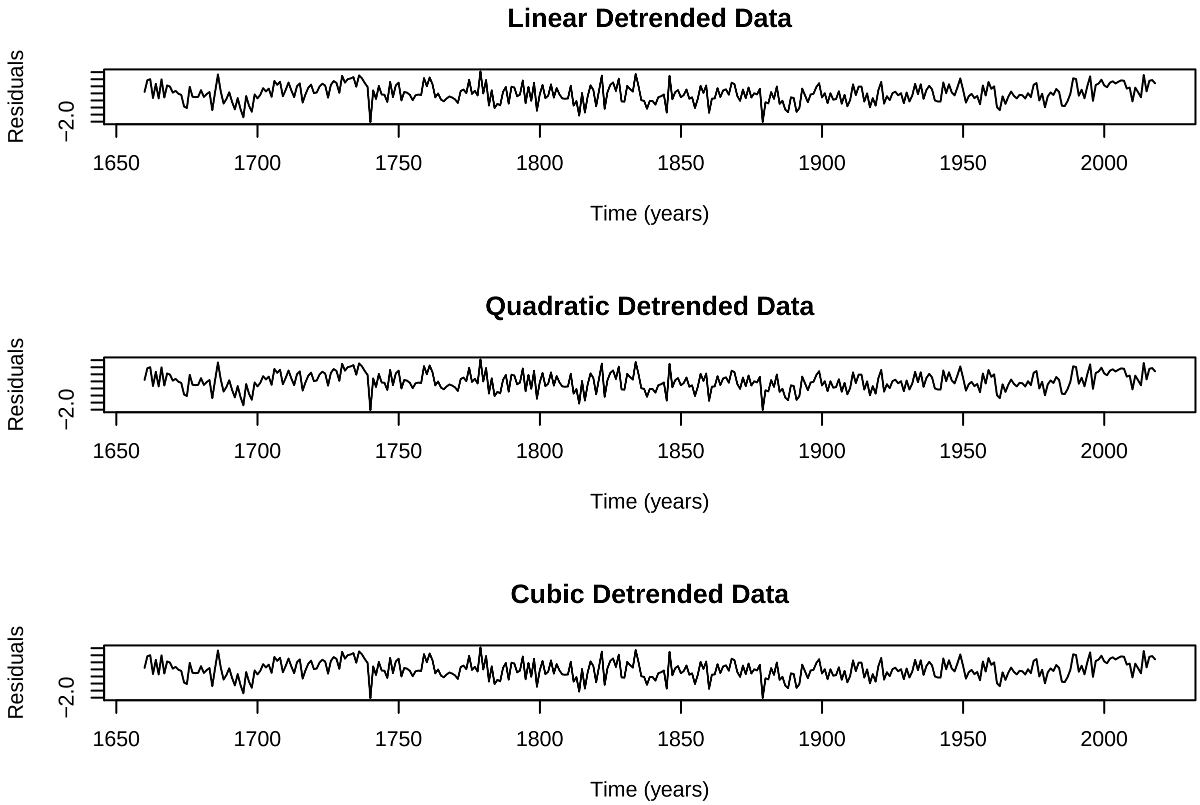
1. Introduction

Over the past few decades, global warming has become an increasingly critical issue as we are already experiencing the early effects and beginning to speculate the later, much more dire effects. The Central England temperature series is the longest instrumental temperature time series, making it a valuable tool to discuss long-term climate change. The CET data is a measurement of average seasonal temperature in a region representative of Central England every year dating back to 1659. The seasons are Winter, Spring, Summer and Fall and the measurements are recorded as the average of the 3 monthly averages corresponding to each season. December-January-February represents Winter, March-April-May represents Spring, June-July-August represents Summer, and September-November-December represents Fall. By taking the average of the four seasonal columns, we can get a time series of average temperature per year and by using time series analysis techniques we can get a sense of the long-term trend, if there is one, in Central England temperature variation. Given the importance of global warming, our focus will be to explore the issue of estimation of the trend component. Our goal is to investigate the trend in this time series data and, if the trend exists, try and find the best model to understand this trend. Trend models for this data have been discussed thoroughly with many different possibilities suggested however they mostly all use regression methods. Some of these, which we will make explicit references to are, the linear trend model suggested by Jones and Bradley (1992) and Jones and Hulme (1997), the quadratic trend model suggested by Benner (1999) and Harvey and Mills (2003), and the cubic trend model suggested by Harvey and Mills (2003) and Zhang and WU (2011).

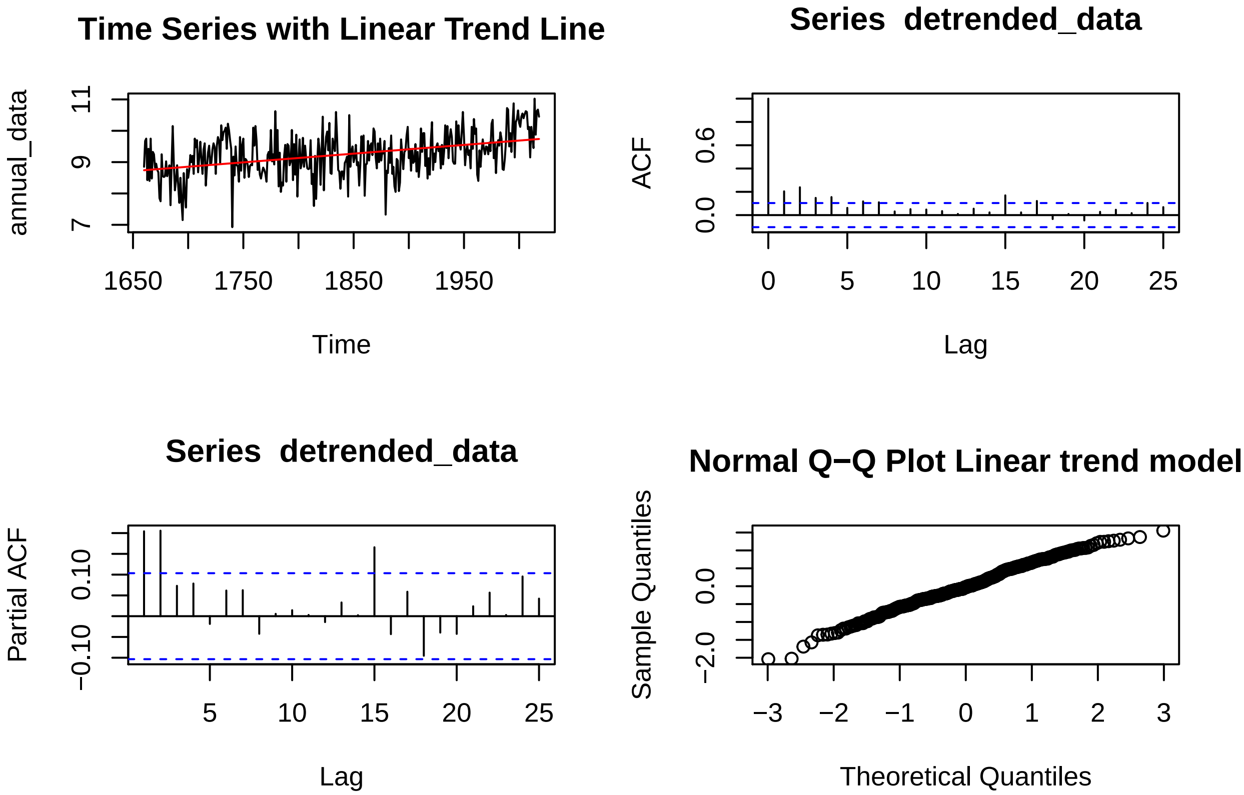


2. Analysis of Annual Trend

The first thing we notice about this time series is that it does not appear to be stationary. Looking at the moving average smoother (blue) using a normal kernel function, there seems to be a clear upward trend over time and especially in the last 20 years but it is somewhat hard to decipher visually and will require more in-depth analysis. Furthermore, looking at the spline method smooth curve (red) we can see that the long-term trend might be best fit by a cubic or that there are three periods of temperature change between 1660 and 2019. If we increase the smoothing paramater and smooth again using spline method (green) we can see that even a long-term linear trend looks possible like Jones and Hulme (1997) and Jones and Bradley (1992) postulated. We will begin by detrending the data in wach of the three forms suggested by the previous authors and compare their models using

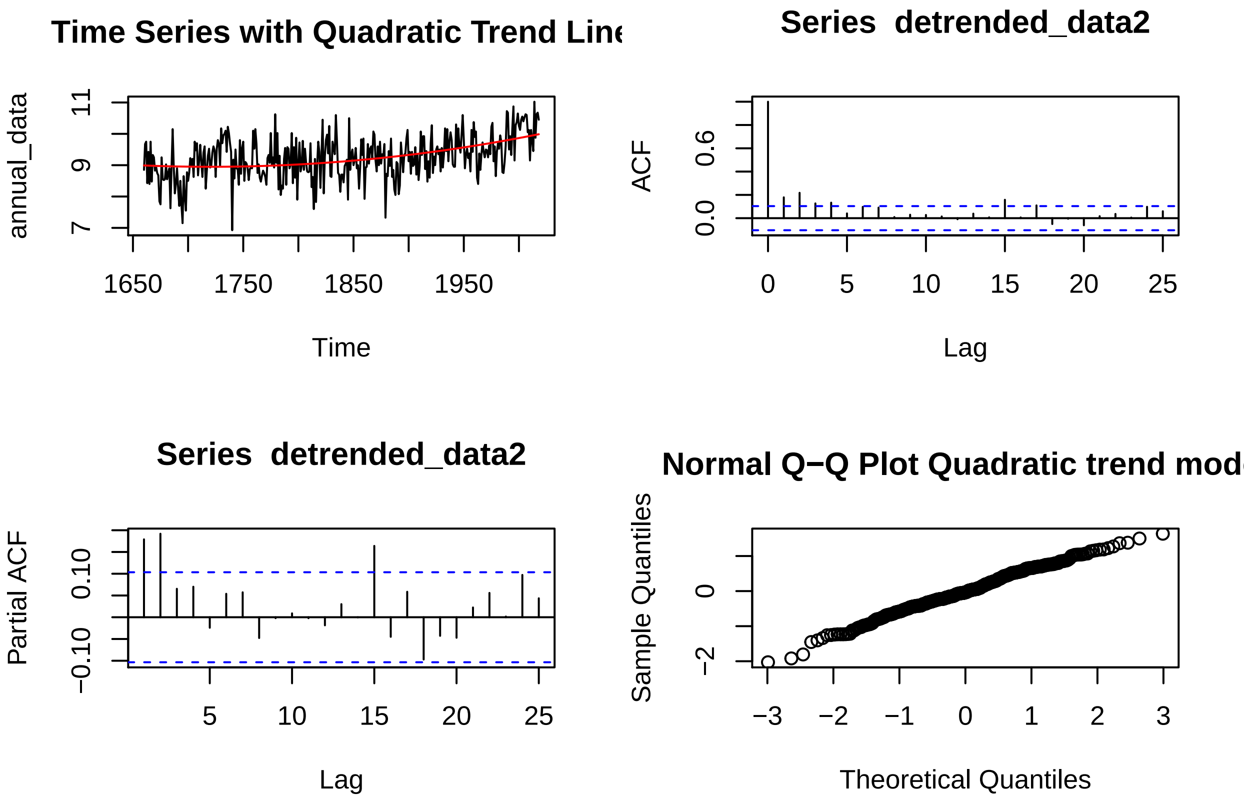


All 3 forms of detrended data look pretty similar and all look feasible. Our goal is ultimately to choose one that we feel best models the trend. To do so we will compare the ACF and PACF plots as well as compare some model diagnostic statistics such as the adjusted R-squared values, the BIC values and the histograms of the data. We will also be making references to previous pieces on the matter such as: Jones and Hulme (1997), Jones and Bradley (1992), Benner (1999), Harvey and Mills (2003), and Zhang and WU (2011).



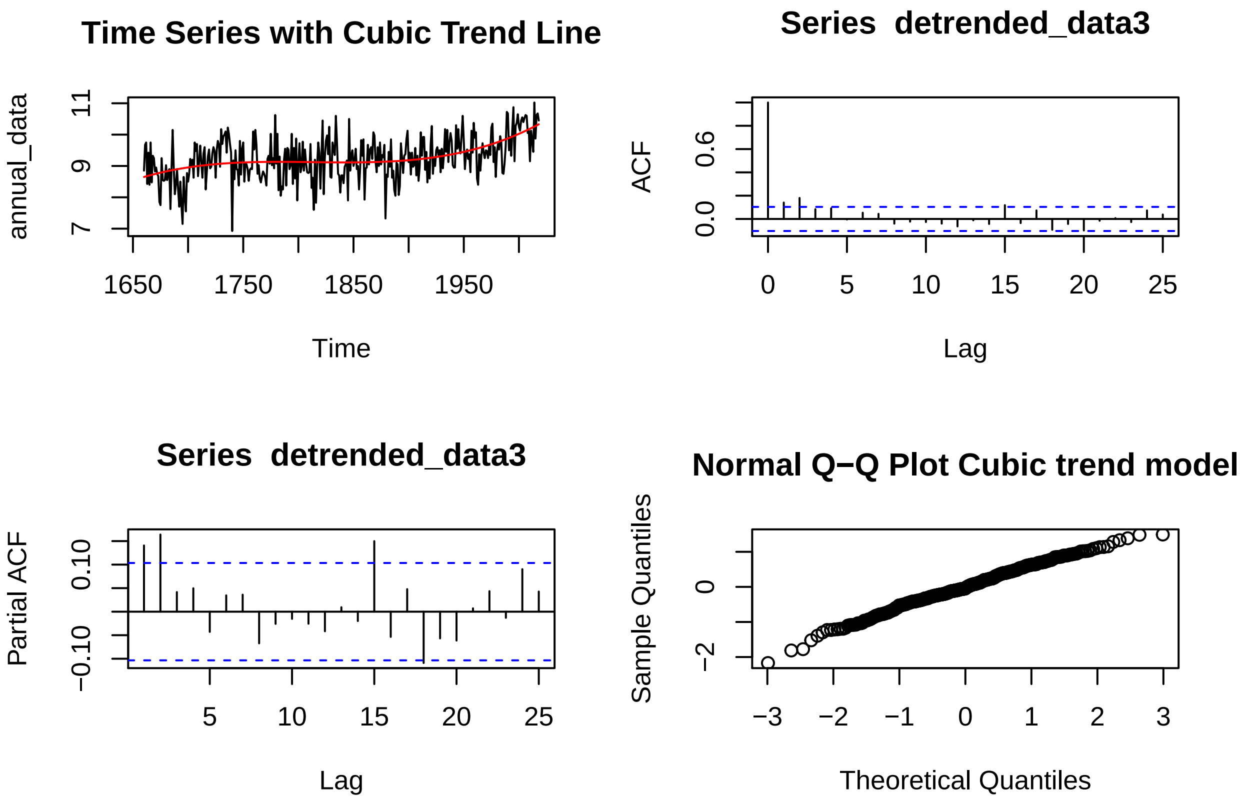

```
Call:
lm(formula = annual_data ~ time_data, na.action = NULL)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0393 -0.3801 -0.0267  0.4813  1.5521

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 8.738888   0.066613  131.19   <2e-16 ***
time_data   0.002783   0.000321    8.68   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.63 on 357 degrees of freedom
Multiple R-squared:  0.174, Adjusted R-squared:  0.172
```

```
F-statistic: 75.3 on 1 and 357 DF,  p-value: <2e-16
```




```
Call:
lm(formula = annual_data ~ time_data + I(time_data^2))

Residuals:
    Min      1Q  Median      3Q     Max
-2.0283 -0.4114 -0.0369  0.4805  1.6357

Coefficients:
                  Estimate  Std. Error t value Pr(>|t|)
(Intercept)     8.99235145  0.09878762   91.03  < 2e-16 ***
time_data      -0.00142966  0.00126723   -1.13  0.26000
I(time_data^2)  0.00001170  0.00000341    3.43  0.00067 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.62 on 356 degrees of freedom
Multiple R-squared:  0.201, Adjusted R-squared:  0.196
F-statistic: 44.7 on 2 and 356 DF,  p-value: <2e-16
```




```
Call:
lm(formula = annual_data ~ time_data + I(time_data^2) + I(time_data^3))

Residuals:
    Min      1Q  Median      3Q     Max
-2.1703 -0.3757 -0.0445  0.4331  1.4936

Coefficients:
                    Estimate    Std. Error t value Pr(>|t|)
(Intercept)     8.6383027649  0.1295085696   66.70  < 2e-16 ***
time_data       0.0102905364  0.0031111617    3.31  0.00104 **
I(time_data^2) -0.0000695752  0.0000200670   -3.47  0.00059 ***
I(time_data^3)  0.0000001505  0.0000000366    4.11  0.00005 ***
```

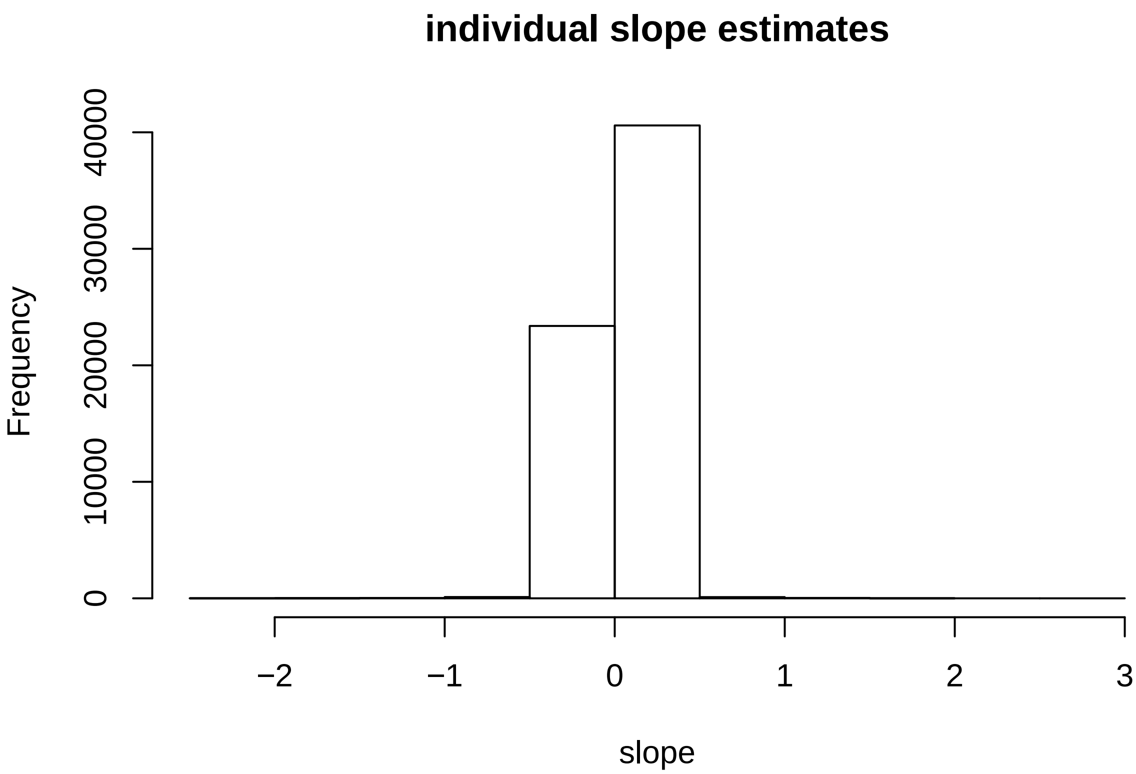
```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.607 on 355 degrees of freedom
Multiple R-squared:  0.237, Adjusted R-squared:  0.23
F-statistic: 36.7 on 3 and 355 DF,  p-value: <2e-16
```

Comparing the diagnostics for each of the proposed models, we see that all of the ACF's and PACF's look reasonable and point to some characteristics of stationarity which means that each of the trend models could have successfully described the trend we are interested in. The model diagnostics point to the cubic trend being the preferred model for the trend as the adjusted R-squared value is highest and the BIC lowest for this model. The normal quantile plots suggests that the linear or quadratic model is the best for the trend. The coefficient for time in the quadratic model is not significant while every coefficient in either of the other two models is. Since most of the model diagnostics, specifically the adjusted R-squared value point towards the cubic trend, there is a compelling case to choose this trend model over the others and we will investigate further if this trend is acceptable using spectral analysis. The cubic trend suggests 3 periods in Central England temperature change between 1660 and 2019, a cooling period in early part of the era, followed by a period of flunctuation around the mean, and ending with a warming trend in the latter part of the era into the present. This conclusion is in line with both Benner (1999) and Zhang and WU (2011). To investigate this further, a spectral analysis would give more insight on different periods in the trend throughout the years. However, at very low frequencies, spectral analysis is not always reliable and since we are considering such few oscillations in global temperature over such a long period of time, a spectral analysis might not yield reliable results.

MannKendall Test

Another non-parametric approach to test for trend is to use the MannKendall test put forth by Hirsch and Slack (1984). This test is used to determine if there is a linear trend in a time series by calculating a test statistic, tau, which is a measurement of the proportion of up-movements against time vs the proportion of down-movements, looking at all possible pairwise time-differences. This test statistic is then compared to a theoretical test statistic to conclude how likely this outcome would have been given there was no trend. One thing to be clear of for this test is that it is a test for a monotonic trend in a time series, which is only 1 of 3 of our postulated models. It can still be useful to see if the time series changes over time monotonically but would not necessarily mean that the linear model is the best fit for this trend as both the quadratic and the cubic fit also indicate some sort of overarching monotonic trend.

```
tau = 0.279, 2-sided pvalue =<2e-16
```

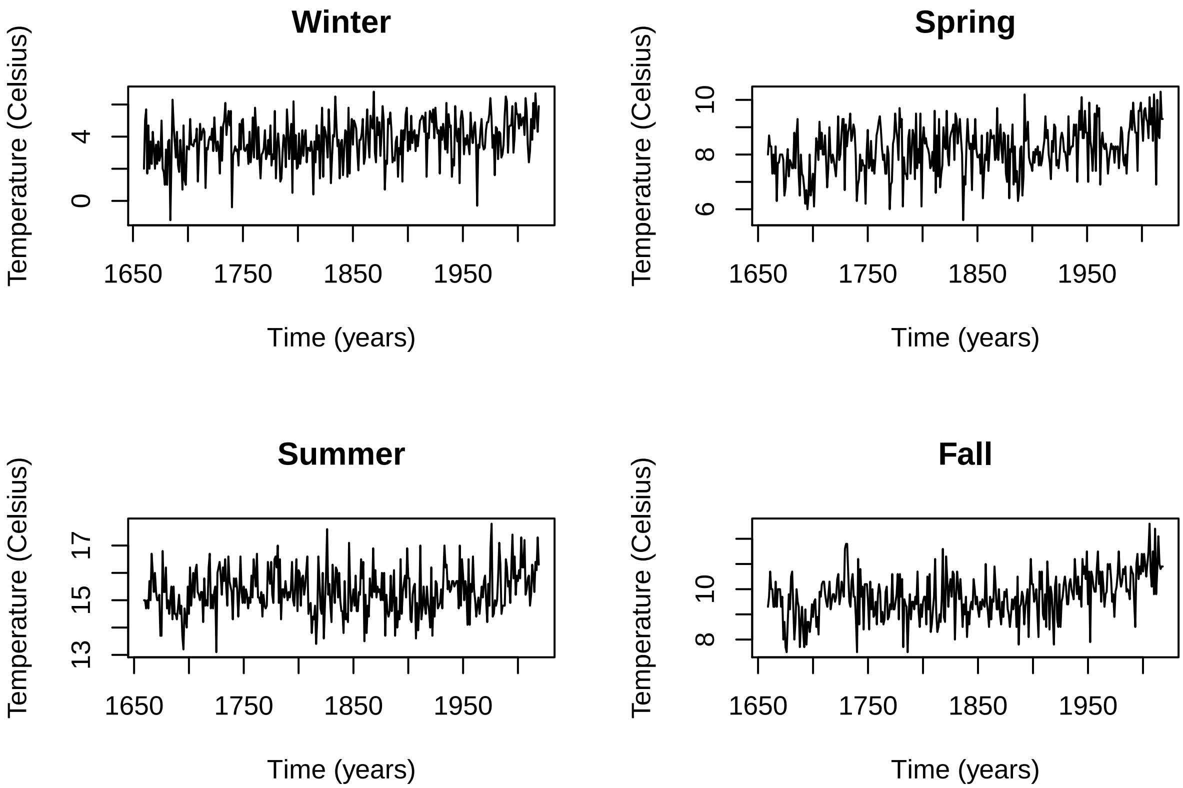



```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-2.5000 -0.0030  0.0027  0.0030  0.0087  2.6000
```

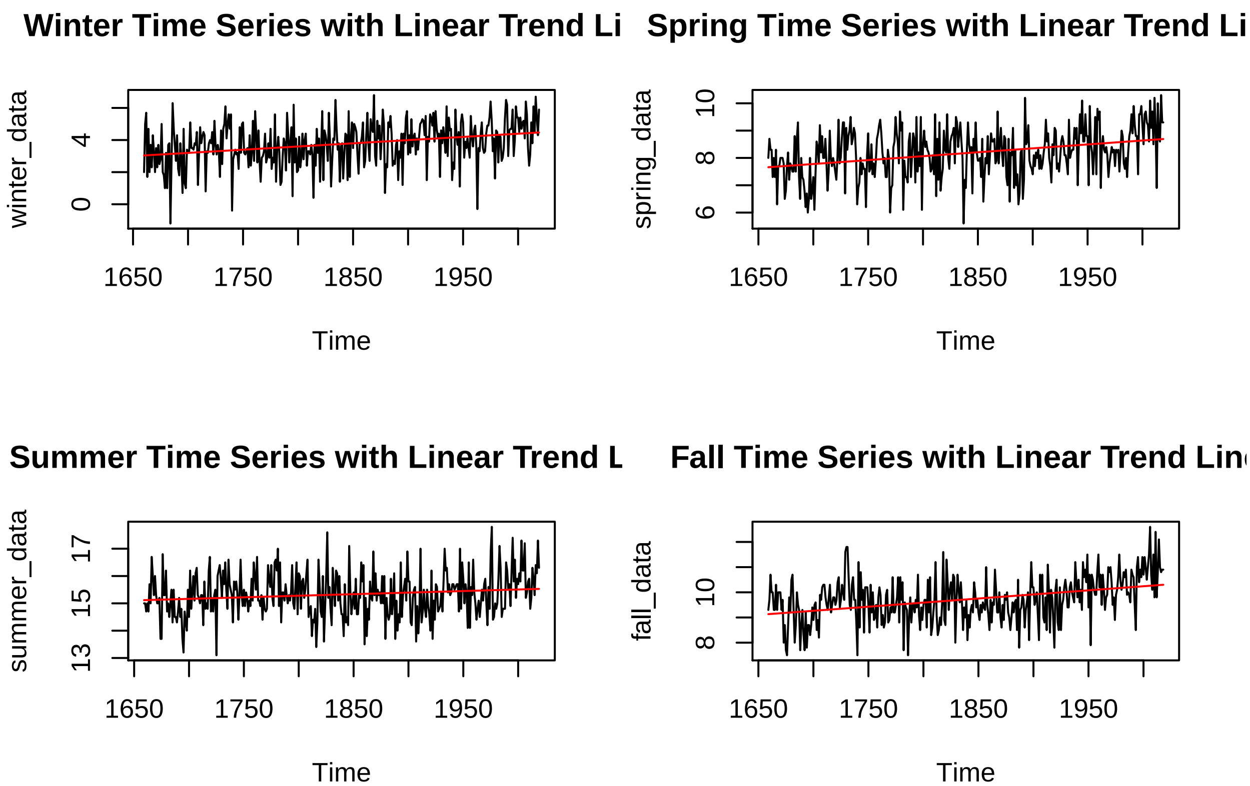
```
[1] 0.002749
```

This test tells us that there is strong evidence of a monotonic trend in this time series. Furthermore, computing the slope estimate for this trend proposed also by Hirsch and Slack (1984), we get a slope of .002749. Comparing this to our linear regression slope estimate (.002783) we see that these values are very very similar which gives even more evidence in favor of the positive linear model for long term global climate trend.

Analysis of Seasonal Trends



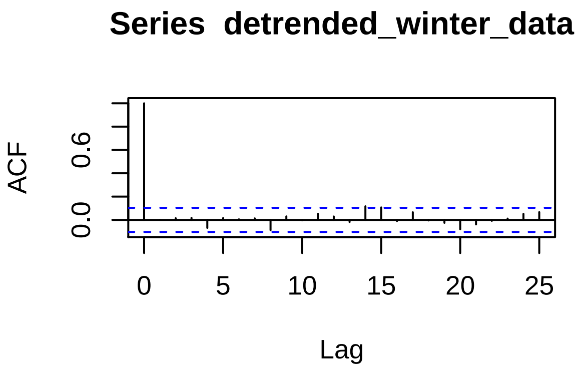
Another point of interest when trying to determine the trends in this time series data is to determine trends within specific seasons. It is interesting to note which season is effected most by the overall annual trend and whether or not certain seasons contribute to the warming pattern significantly more than others. To do this, we will proceed with each seasonal time series much like we did with the annual data, except only proceeding with a linear trend fit. We will check the model diagnostics for each seasonal linear trend and apply the non parametric MannKendall test for monotonic trends to see if there is an overall linear trend in any of the data and to compare the slopes with our linear regression methods. Our goal will be to see if any of the seasons have more significance when determining the overall annual linear trend than the others.

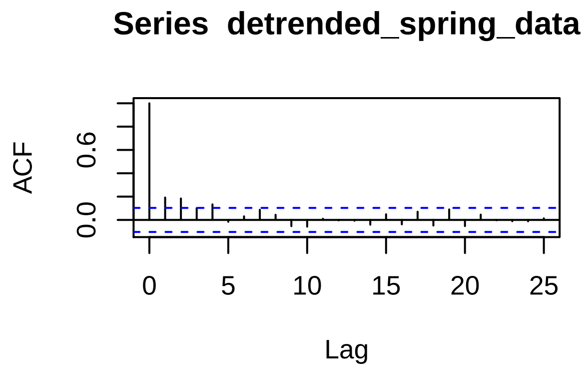


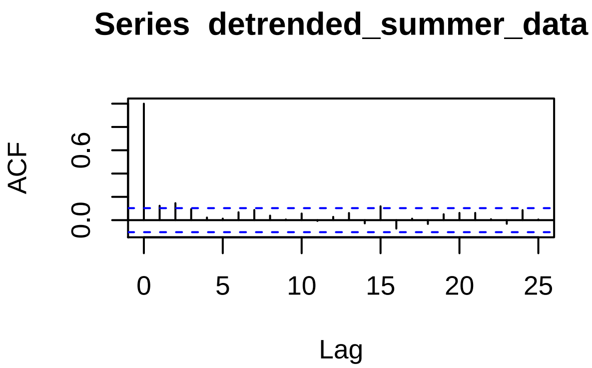

```
Call:
lm(formula = winter_data ~ time_winter_data, na.action = NULL)

Residuals:
   Min     1Q Median     3Q    Max
-4.546 -0.844  0.098  0.916  3.157

Coefficients:
                 Estimate Std. Error t value     Pr(>|t|)
(Intercept)      3.034968   0.137693   22.04      < 2e-16 ***
time_winter_data 0.003984   0.000661    6.03 0.0000000042 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.3 on 358 degrees of freedom
Multiple R-squared:  0.0921,    Adjusted R-squared:  0.0896
```

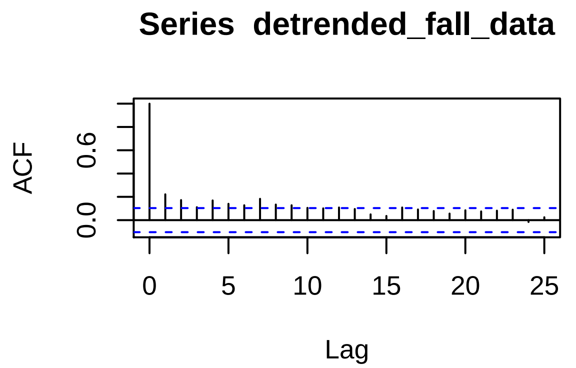
```
F-statistic: 36.3 on 1 and 358 DF,  p-value: 0.00000000415


Call:
lm(formula = spring_data ~ time_spring_data, na.action = NULL)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.5699 -0.5154  0.0086  0.6021  1.8695 

Coefficients:
                 Estimate Std. Error t value       Pr(>|t|)    
(Intercept)      7.659529   0.088583   86.47        < 2e-16 ***
time_spring_data 0.002867   0.000426    6.73 0.000000000066 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.843 on 359 degrees of freedom
Multiple R-squared:  0.112, Adjusted R-squared:  0.11 
F-statistic: 45.3 on 1 and 359 DF,  p-value: 0.0000000000664


Call:
lm(formula = summer_data ~ time_summer_data, na.action = NULL)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.0926 -0.5258 -0.0413  0.5182  2.3203 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      15.117086   0.085160  177.51   <2e-16 ***
time_summer_data  0.001144   0.000409    2.79   0.0055 ** 
---
```

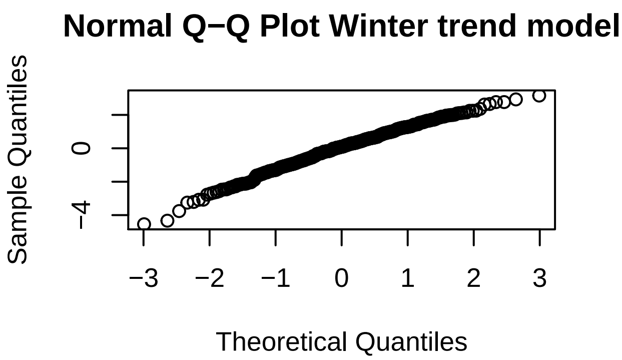
```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.811 on 359 degrees of freedom
Multiple R-squared:  0.0213,    Adjusted R-squared:  0.0186
F-statistic: 7.81 on 1 and 359 DF,  p-value: 0.00549


Call:
lm(formula = fall_data ~ time_fall_data, na.action = NULL)

Residuals:
    Min      1Q  Median      3Q     Max
-2.1846 -0.5504  0.0143  0.5812  2.4367

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.132558   0.088227  103.51   <2e-16 ***
time_fall_data 0.003249   0.000425    7.64    2e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.839 on 358 degrees of freedom
Multiple R-squared:  0.14,  Adjusted R-squared:  0.138
F-statistic: 58.4 on 1 and 358 DF,  p-value: 2.02e-13
```

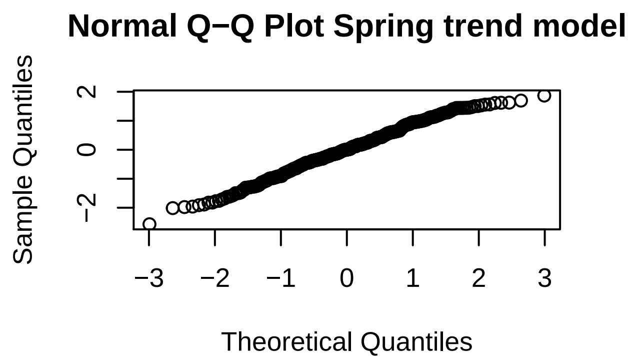

Series detrended_winter_data
ACF
0.0
0.6
0
5
10
15
20
25
Lag

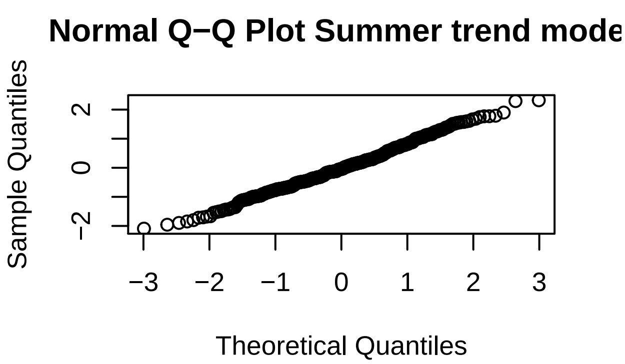

Series detrended_spring_data
ACF
0.0
0.6
0
5
10
15
20
25
Lag

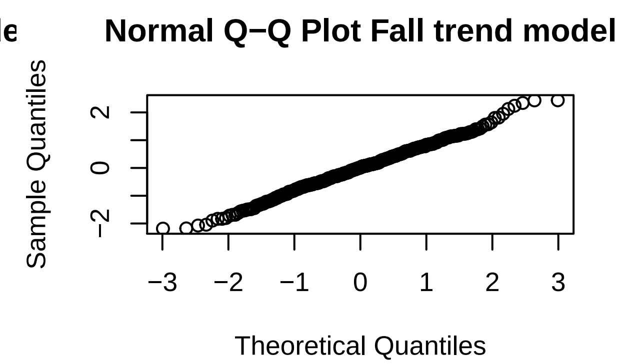

Series detrended_summer_data
ACF
0.0
0.6
0
5
10
15
20
25
Lag


Series detrended_fall_data
ACF
0.0
0.6
0
5
10
15
20
25
Lag

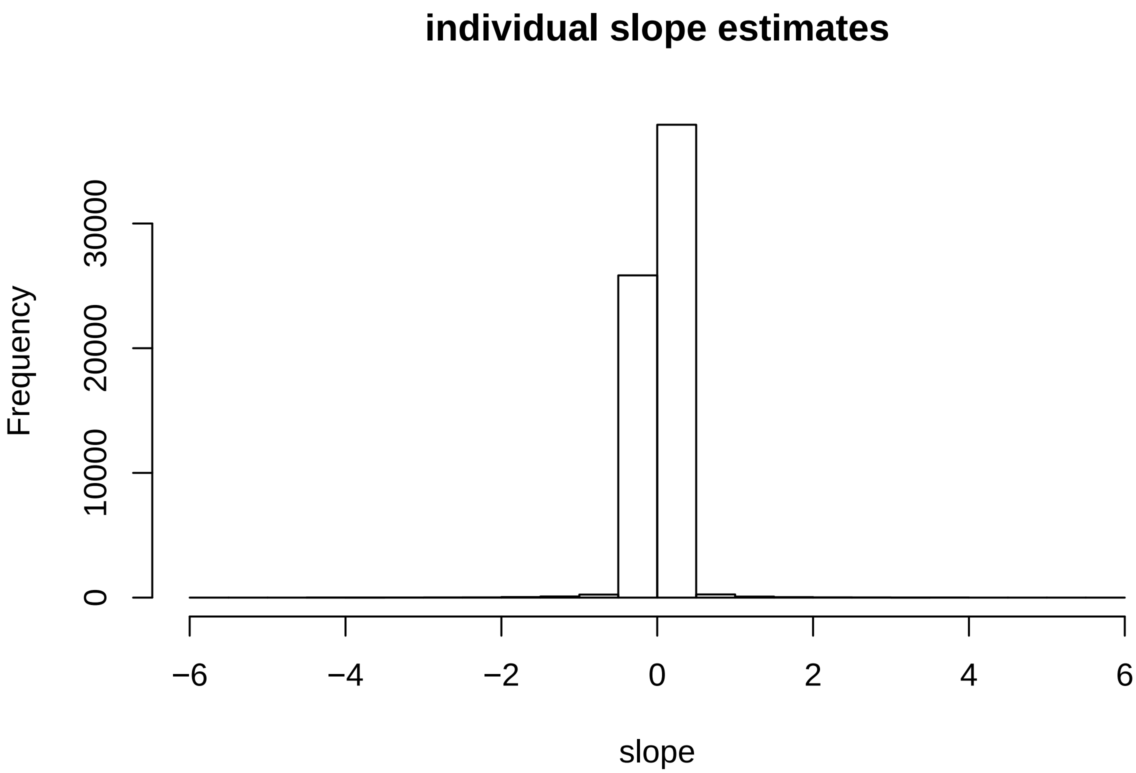

Normal Q−Q Plot Winter trend model
Sample Quantiles
−4
0
−3
−2
−1
0
1
2
3
Theoretical Quantiles


Normal Q−Q Plot Spring trend model
Sample Quantiles
−2
0
2
−3
−2
−1
0
1
2
3
Theoretical Quantiles


Normal Q−Q Plot Summer trend mode
Sample Quantiles
−2
0
2
−3
−2
−1
0
1
2
3
Theoretical Quantiles

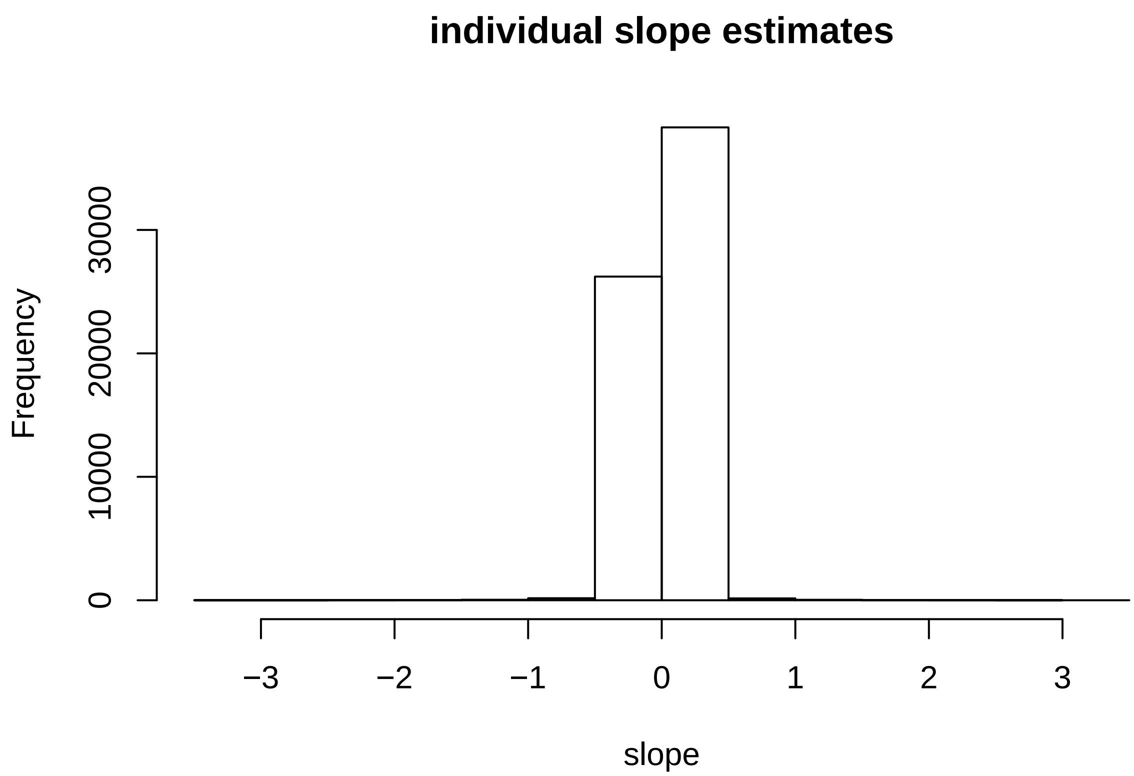

Normal Q−Q Plot Fall trend model
Sample Quantiles
−2
0
2
−3
−2
−1
0
1
2
3
Theoretical Quantiles

```r
MKslope(winter_data)
```




```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 -6.000  -0.008   0.004   0.004   0.016   5.700 

[1] 0.004075
```

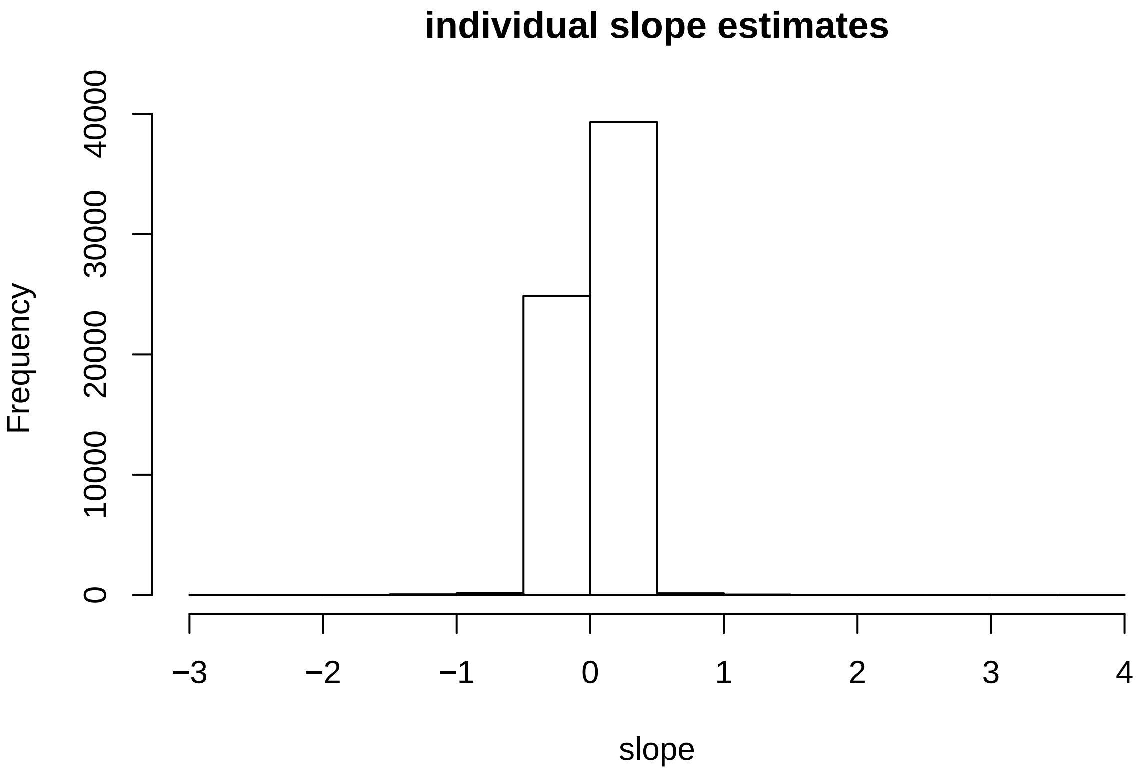
```r
MKslope(spring_data)
```




```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -3.400  -0.005   0.003   0.003   0.011   3.100
```

```
[1] 0.002804
```

```
MKslope(fall_data)
```




```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -2.800  -0.004   0.003   0.004   0.011   3.700
```

```
[1] 0.003306
```

Based on the results of our analysis, a linear trend is a sufficient model to fit for each of the seasons although it is less significant during the summer months. All of the models diagnostics indicate that the model fits as the Q-Q normal plots are normal and the ACF's all indicate stationarity. Looking at the MannKendall tests, they all indicate a significant monotonic positive slope although much less significant during the summer months. The greatest slope is seen during the winter months followed by fall and then spring and then summer. This tells us that the colder the season, the more affected by the overall warming trend. # Conclusions

Using linear and polynomial regressions to detrend the data, we can get a sense of the overall trend in Central England temperatures. As global warming is a very topical and important issue, many different models have been proposed for the trend in temperatures using the Central England Surface Temperature data. Some of these hypotheses are: a linear trend, (Jones and Hulme (1997) and Jones and Bradley (1992)) a quadratic trend, (Benner (1999) and Harvey and Mills (2003)) or higher order or cubic trends (Harvey and Mills (2003) and Zhang and WU (2011)). In our analysis, we considered all three trends as possible and looked at model diagnostics for each trend model to decide which one accurately described the data. All three were plausible based off of their diagnostics, but the cubic trend accounted for more of the variance in the data while keeping the significance in the coefficients. Because of this, I would say that the cubic fit does describe the data the best which is in line with the theory that there have been three distinct periods of cooling, fluctuation around the mean, and warming argued by Benner (1999) and Zhang and WU (2011). Analyzing further, we found that by using the MannKendall test for monotonic trends in time series, there is strong evidence of an overall monotonic or linear trend in annual Central England temperatures since 1660. The cubic fit in trend is still consistent with this conclusion as with it, there is still an overarching, small positive trend over time. Using the results from the MannKendall tests which indicated strong evidence for a monotonic trend over time, we analyzed each of the seasonal time series individually to determine which if any experienced this warming trend more significantly than the others. Our results were that the linear warming trend is felt more by the colder months than the warmer months, with winter having the greatest and most significant slope and summer having the lowest and least significant.

References

Benner, T.C. 1999. "Central England Temperatures: Long-Term Variability and Teleconnections." *International Journal of Climatology* 19: 391–403. doi:.

Harvey, D.I., and T.C. Mills. 2003. "Modelling Trends in Central Englandtemperatures." *Journal of Forecasting* 22: 35–47. doi:.

Hirsch, P.D., and James R. Slack. 1984. "A Nonparametric Trend Test for Seasonal Data with Serial Dependence." *Climates of the British Isles: Present, Past and Future* 20 (6): 727–32. doi:.

Jones, P.D., and R.S. Bradley. 1992. "Climatic Variations in the Longest Instrumental Records. in Climate Since a.D. 1500." Routledge London. doi:.

Jones, P.D., and M Hulme. 1997. "The Changing Temperature of Central England." *Climates of the British Isles: Present, Past and Future.* Routledge London, 173–96. doi:.

Zhang, T., and W.B. WU. 2011. "Testing Parametric Assumptions of Trends of Nonstationary Time Series." *Biometrika* 98 (3): 599–614. doi:.