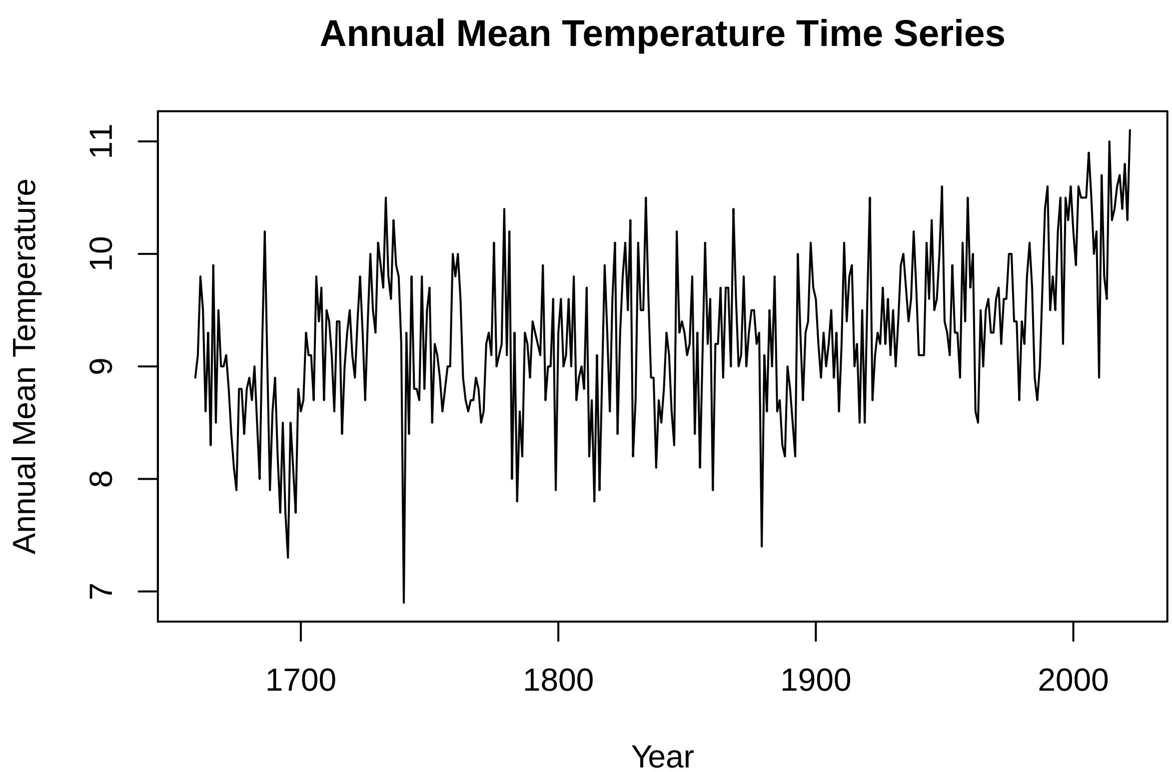
A Time Series Exploration of Central England's Climate

I intend to do an analysis on the (monthly) central England temperature series from year 1659 to 2023. I intend on assessing whether there truly is a presence of a warming trend, find prominent periods of oscillation, and whether there truly is a slowdown or hiatus of global warming. I decided to filter out the data from 2023, since the data is not complete. The following code loads and filters the data.

```
url <- "https://www.metoffice.gov.uk/hadobs/hadcet/data/meantemp_monthly_totals.txt"
local_file <- "/Users/nicholasliagridonis/Desktop/meantemp_monthly_totals.txt"
download.file(url, local_file)
temp <- read.table(local_file, header = TRUE, skip = 4)
temp <- temp[!apply(temp[, 2:13] == -99.9, 1, any), ]
```

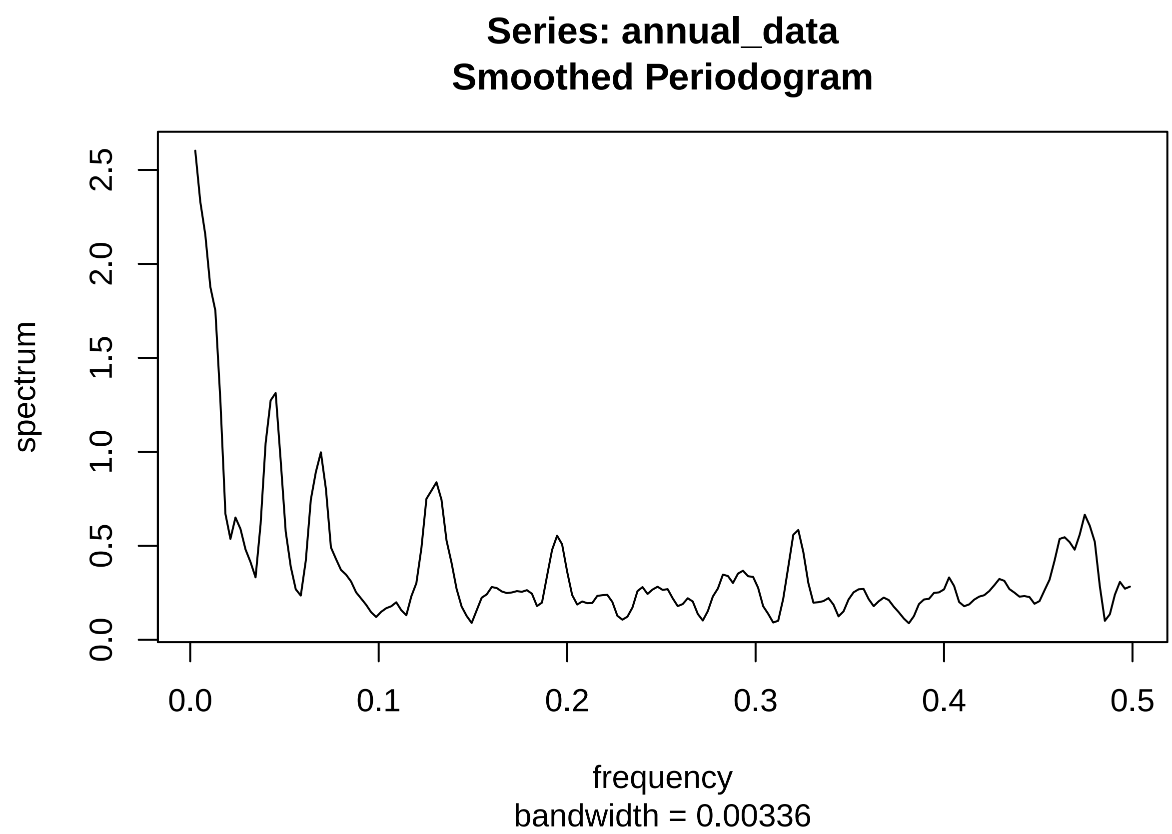
I first wanted to just plot the time series data to see if there was anything particularly notable.

```
plot(temp$Year, temp$Annual, type = "l", xlab = "Year", ylab = "Annual Mean Temperature", main = "Annual
```



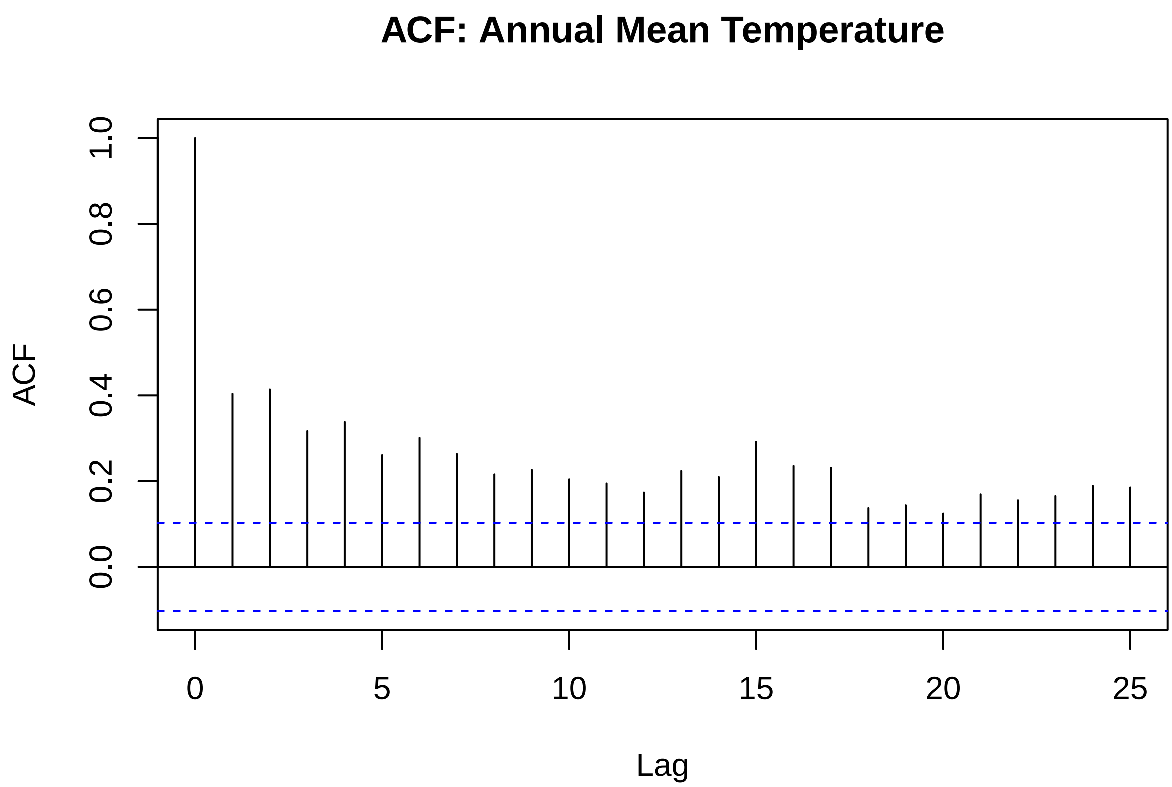
There seem to be a couple big dips from 1700 to 1900, and quite a bit of oscillations in between. From 1900 on, there seems like there is potentially a gradual increase. I'll investigate this further. First, I'll create a periodogram for Spectral Analysis.

```r
annual_data <- temp$Annual
spec_result <- spec.pgram(annual_data, spans = 5, taper = 0, log = "no")
```

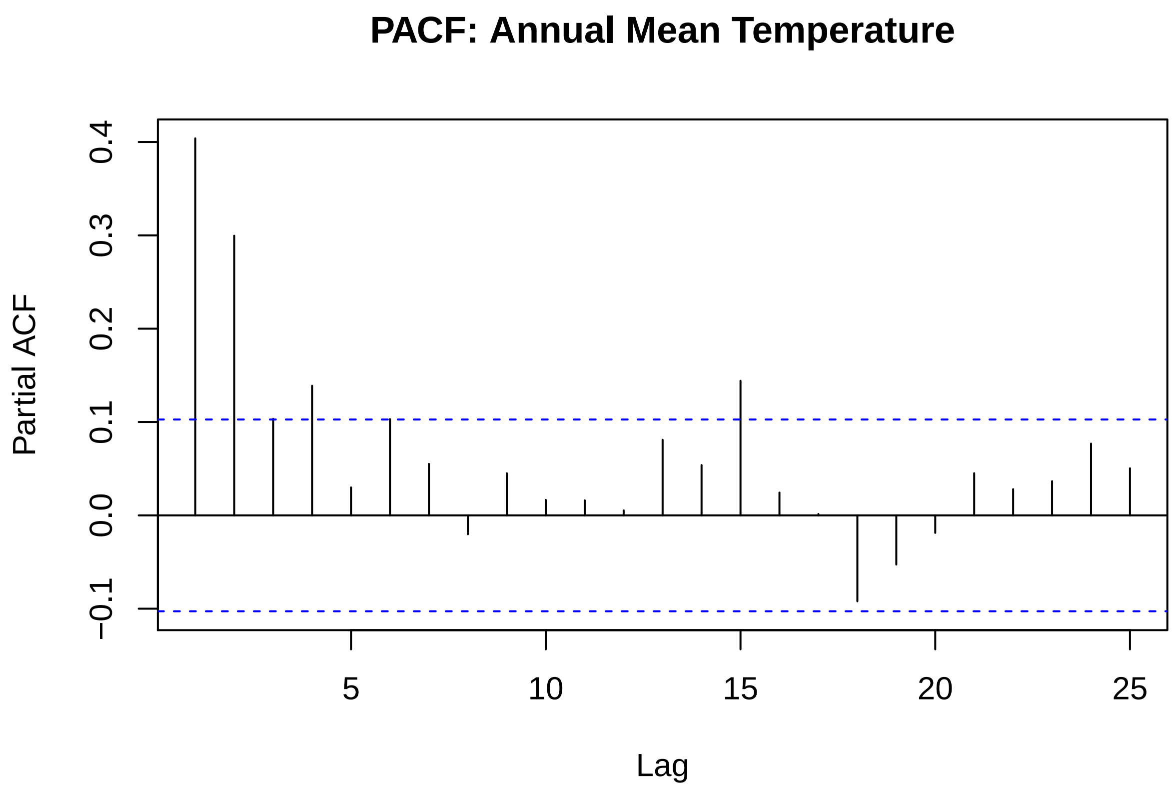


Here, we can make certain comments about periodicity. There are prominent oscillations in the 0.0 to 0.1 range, and then after that, the gaps between oscillations seem to slowly increase. The prominent oscillations after 0.1 center at 0.13, then 0.19, then 0.32, then 0.47. I want to do a further analysis on this data, so I'm going to go through the steps to fit an ARIMA model. First, I will plot the ACF AND PACF.

```r
ts_data <- ts(temp$Annual, start = temp$Year[1])
acf_result <- acf(ts_data, main = "ACF: Annual Mean Temperature")
```

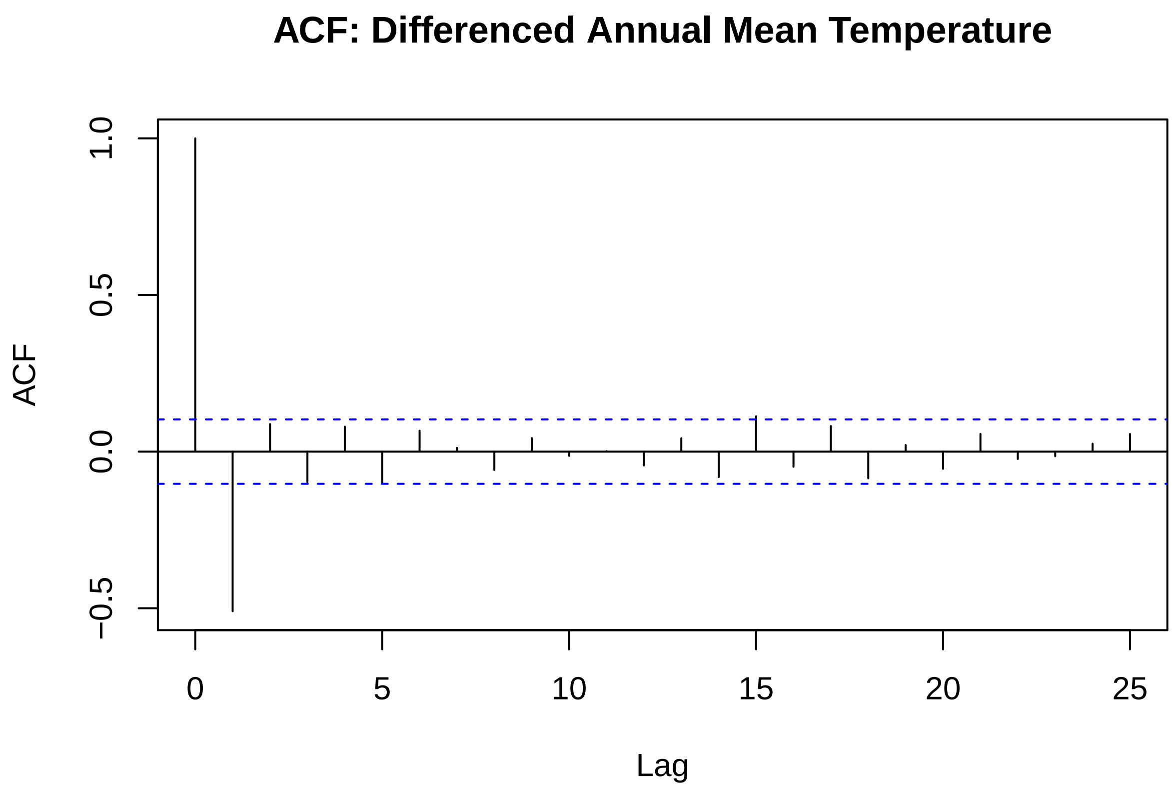



```
pacf_result <- pacf(ts_data, main = "PACF: Annual Mean Temperature")
```

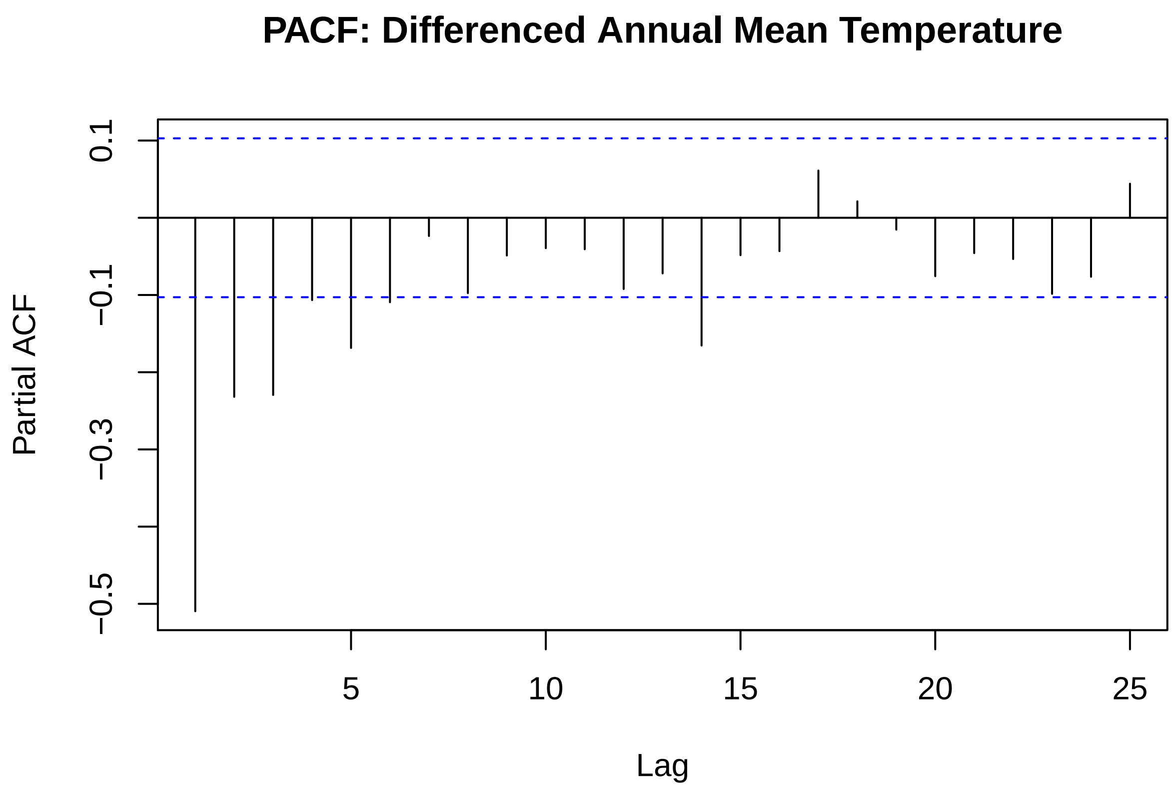


As you can see in the ACF plot, there are significant spikes at virtually every lag. Additionally, the PACF plot has several early lags that are quite significant. These facts may suggest an autoregressive structure with a level of non-stationarity that could be fixed by differencing the data. As a result, I will calculate the differenced ACF and PACF and plot that.

```
diff_ts_data <- diff(ts_data)
diff_acf <- acf(diff_ts_data, main = "ACF: Differenced Annual Mean Temperature")
```




```
diff_pacf <- pacf(diff_ts_data, main = "PACF: Differenced Annual Mean Temperature")
```



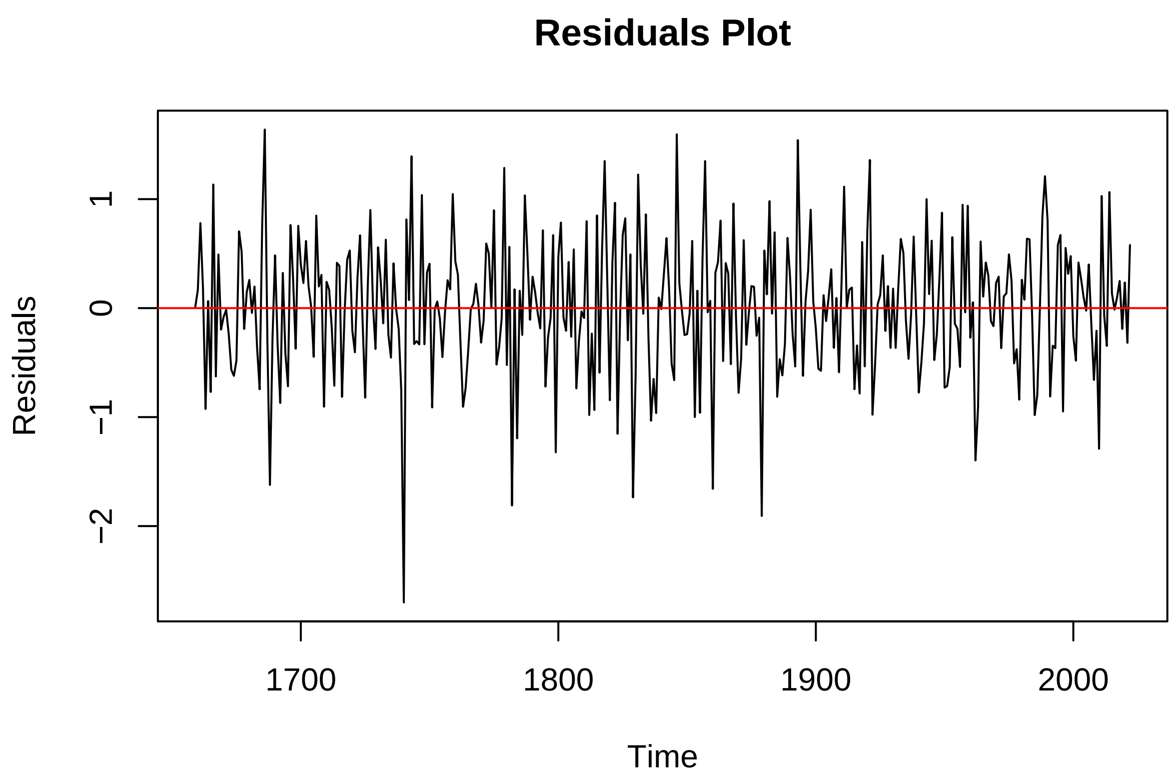
We see some interesting things with these plots. Based on these results, we can fit an ARIMA model.Because of the significant negative values in the PACF at lag 1 and, I'll fit an AR(2). Because of the differencing, I have a first order differenced series. Because the ACF values beyond lag 1 are not significant, we do not need a moving average, so I'll use 0. Therefore, it'll be (2,1,0).

```r
arima_model <- arima(ts_data, order = c(2, 1, 0))
summary(arima_model)
```

```
##           Length Class  Mode
## coef        2    -none- numeric
## sigma2      1    -none- numeric
## var.coef    4    -none- numeric
## mask        2    -none- logical
## loglik      1    -none- numeric
## aic         1    -none- numeric
## arma        7    -none- numeric
## residuals 364    ts     numeric
## call        3    -none- call
## series      1    -none- character
## code        1    -none- numeric
## n.cond      1    -none- numeric
## nobs        1    -none- numeric
## model      10    -none- list
```

Looking at the summary here, the ACF1 value near zero indicates that the model residuals do not exhibit strong autocorrelation at lag 1, suggesting that this model is good choice for representing the temporal dependencies in the data. The negative values of ar1 and ar2 suggest a tendency for the series to revert to the mean. This makes sense with the oscillations we've observed. Just to double check that this model is good, I'll additionally plot the models residuals.

```r
residuals <- residuals(arima_model)
plot(residuals, type = "l", xlab = "Time", ylab = "Residuals", main = "Residuals Plot")
abline(h = 0, col = "red")
```



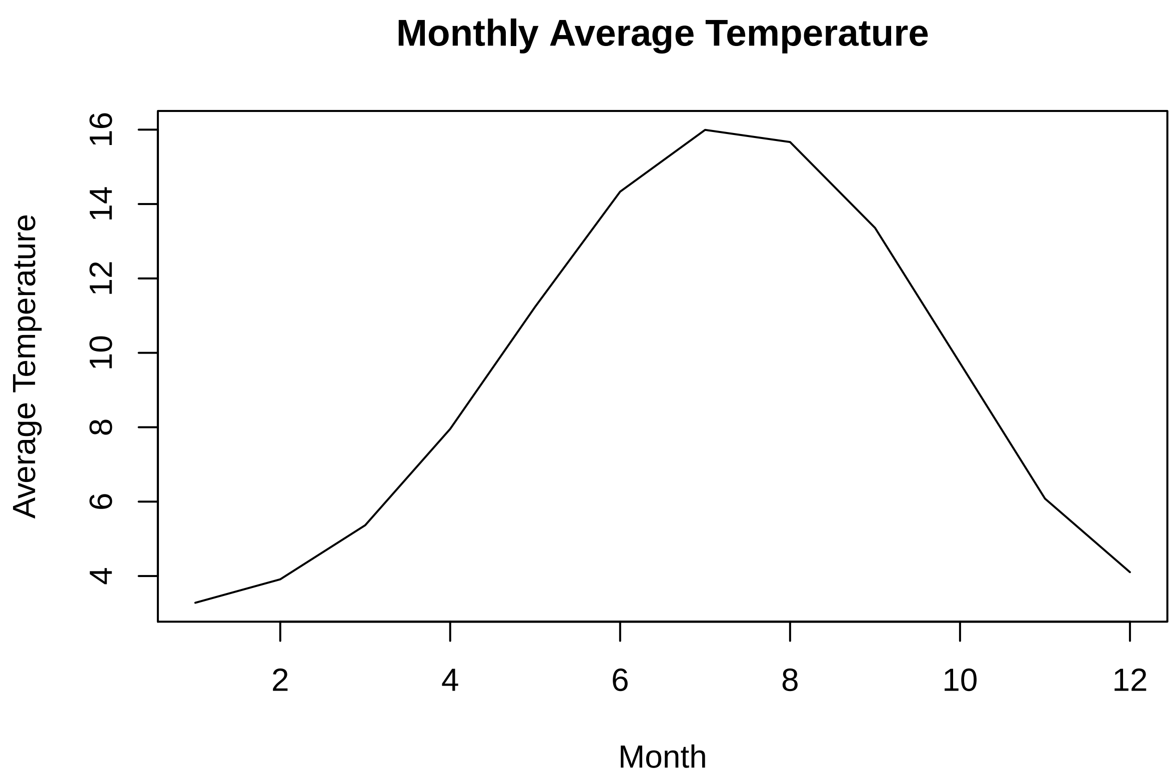
Thankfully, these residuals don't seem to show any patterns, mostly random oscillating around 0. I will also perform a Ljung-Box test to check the residuals independence.

```r
ljung_box_test <- Box.test(residuals, lag = 20, type = "Ljung-Box")
print(ljung_box_test)
```

```
##
##  Box-Ljung test
##
## data:  residuals
## X-squared = 48.647, df = 20, p-value = 0.0003449
```

Now, this is interesting, this low p-value indicates that there is evidence to reject the null hypothesis of independence in the residuals. This suggests there may be some autocorrelation present in the residuals. I realized now that I hadn't accounted for seasonality in the data, and should try to implement that in my model. Because of this, I will now try to fit a SARIMA model. To help inform the creation of this model, I first want to plot the monthly average temperatures to see what exactly the pattern typically is.

```r
monthly_averages <- colMeans(temp[, 2:13], na.rm = TRUE)
# Plot monthly averages
plot(monthly_averages, type = "l", xlab = "Month", ylab = "Average Temperature", main = "Monthly Averag
```



As you can see, average monthly temperatures appear to be normally distributed.

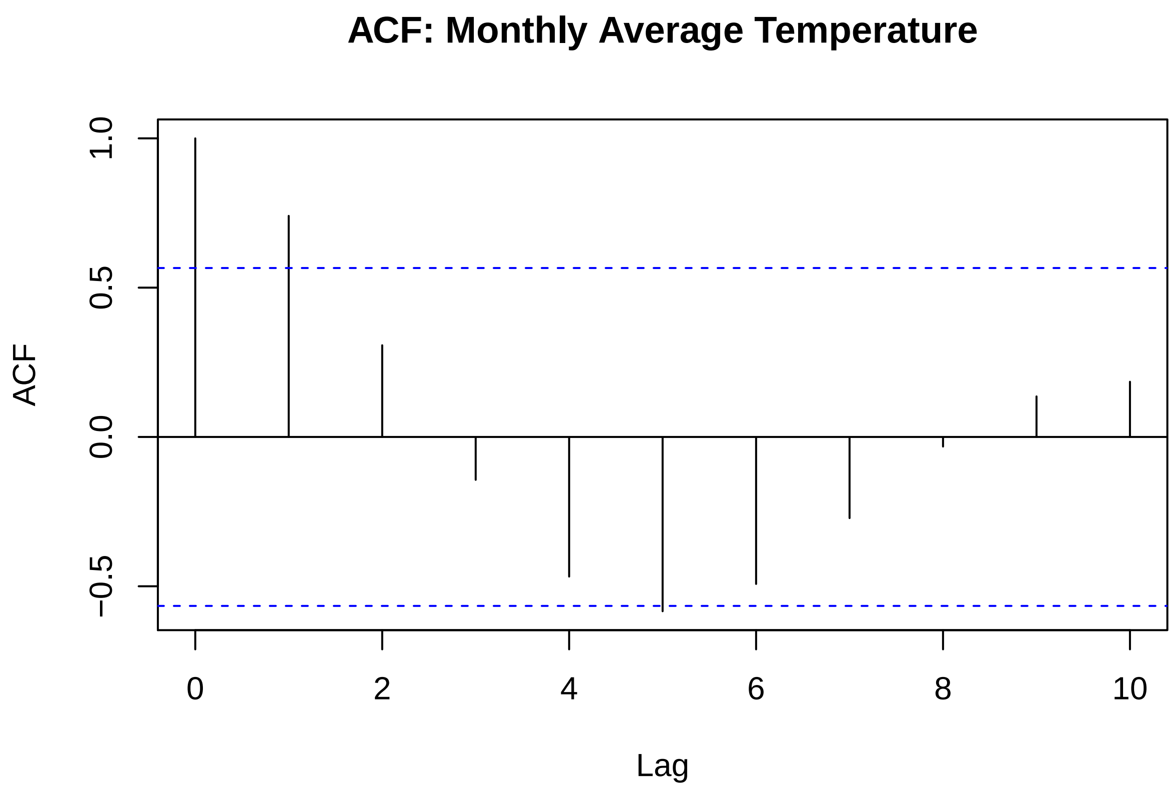
```r
shapiro.test(monthly_averages)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  monthly_averages
## W = 0.90468, p-value = 0.1823
```

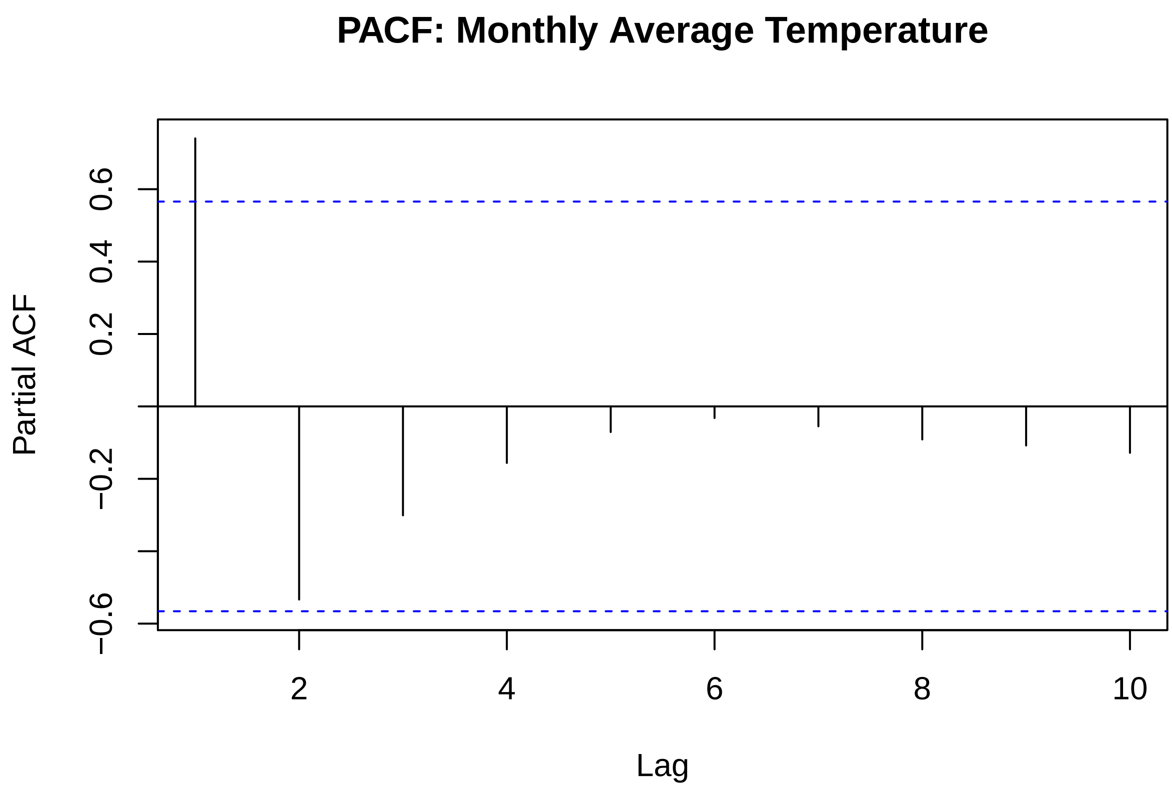
A Shapiro-Wilk test confirms this perceived normality.

Now, I'm going to start the process of fitting a SARIMA model. First, I'll create ACF and PACF plots for the monthly averages.

```r
acf_monthly <- acf(monthly_averages, main = "ACF: Monthly Average Temperature")
```




```
pacf_monthly <- pacf(monthly_averages, main = "PACF: Monthly Average Temperature")
```



In the ACF plot, at lag 0 it is over the dashed threshold at around 1 and at lag 1 it is over the dashed threshold at about 0.7, and the rest of the lags are within the dashed lines. In the PACF plot, lag 0 is the only lag with a significant value, at around 0.7. This data appears to be stationary, so I can proceed with fitting the model. I decided to fit a SARIMA model and see if the function chose a particular fit. I tried auto.arima to see if it would include seasonality:

```r
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.0.5
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

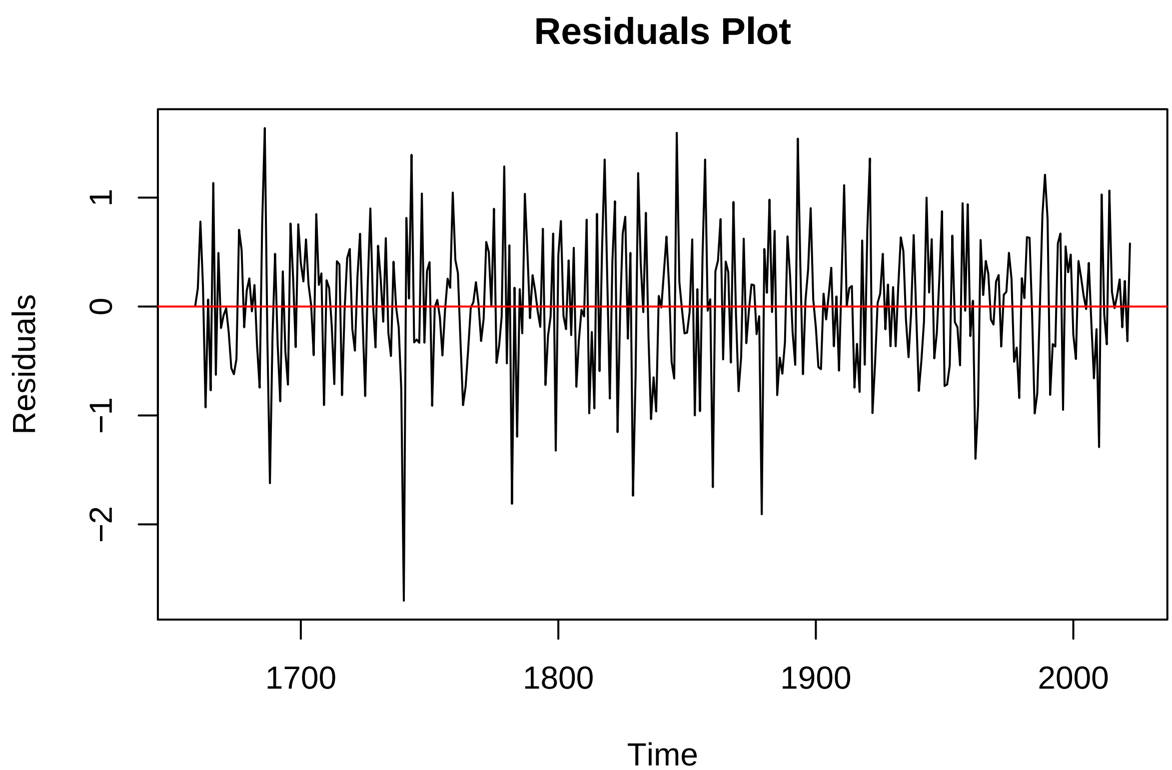
```r
autoarima_model <- auto.arima(ts_data)
summary(autoarima_model)
```

```
## Series: ts_data
## ARIMA(0,1,1)
##
## Coefficients:
##           ma1
##       -0.8353
## s.e.   0.0407
##
```

```
## sigma^2 = 0.3398:  log likelihood = -319.25
## AIC=642.49   AICc=642.52   BIC=650.28
##
## Training set error measures:
##                      ME      RMSE       MAE        MPE     MAPE      MASE
## Training set 0.02400012 0.5812944 0.4516679 -0.1323316 4.947438 0.7763043
##                    ACF1
## Training set 0.04675082
```

It turns out it didn't. The fit is similar to the one I fitted earlier, but I want to see how the residuals look with this model.

```
autoarima_residuals <- residuals(autoarima_model)
plot(residuals, type = "l", xlab = "Time", ylab = "Residuals", main = "Residuals Plot")
abline(h = 0, col = "red")
```



This residuals plot looks the way it should, similar to the past ARIMA residuals plot. Now, I'll also perform a Ljung-Box test to check the residuals independence.

```
ljung_box_test_autoarima <- Box.test(autoarima_residuals, lag = 20, type = "Ljung-Box")
print(ljung_box_test_autoarima)
```

```
##
##  Box-Ljung test
##
## data:  autoarima_residuals
## X-squared = 29.693, df = 20, p-value = 0.07499
```

This model looks better, we do not have evidence to reject the null hypothesis.

I want to be certain that this model works better, so I'll compare the BIC values of the two models.

```
bic_arima <- BIC(arima_model)
bic_autoarima <- BIC(autoarima_model)

cat("BIC for ARIMA model:", bic_arima, "\n")
```

```
## BIC for ARIMA model: 699.6523
```

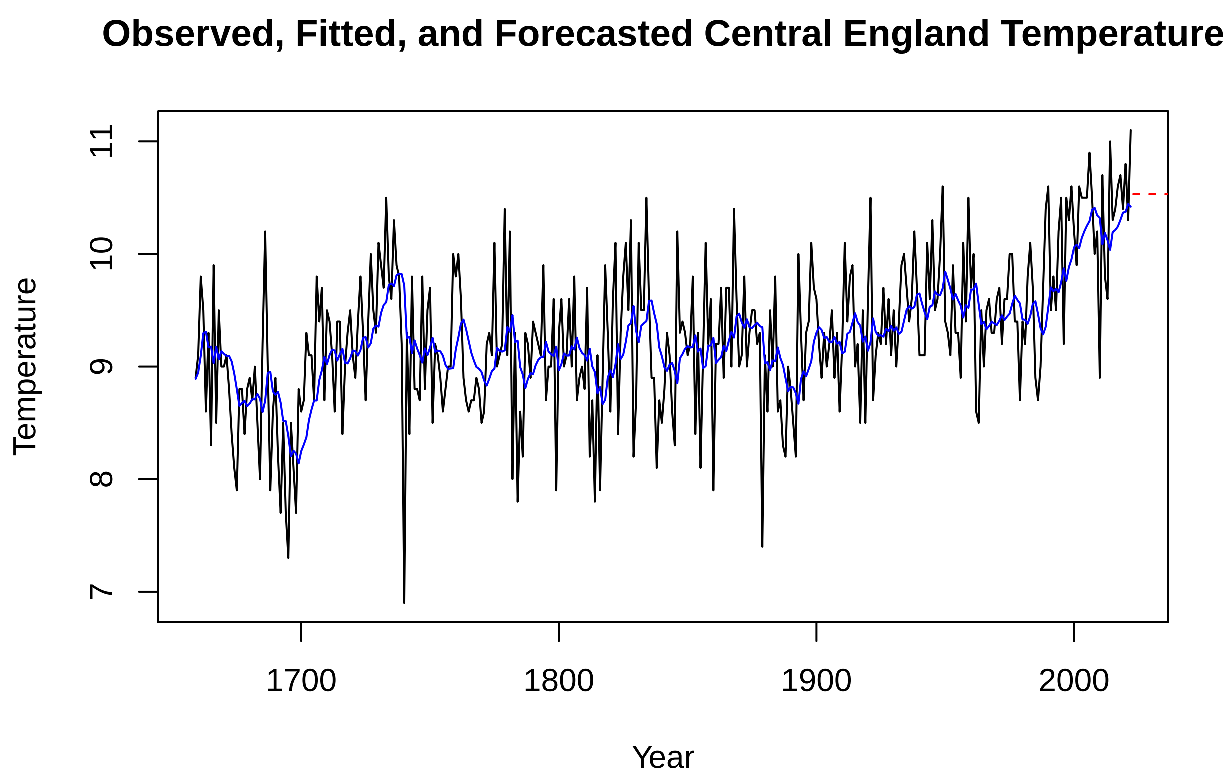
```
cat("BIC for AutoARIMA model:", bic_autoarima, "\n")
```

```
## BIC for AutoARIMA model: 650.2793
```

As expected, the autoarima model ended up also having a lower BIC. I'm now going to use this model to analyze the overall trend in the central England temperature record. I'll plot the observed, fitted, and forecasted values to see if there are warming trends, particularly in recent years, to see if my work justifies Benner's claim.

```
observed_values <- ts_data

plot(observed_values, col = "black", type = "l", lwd = 1, xlab = "Year", ylab = "Temperature", main = "(
lines(fitted(autoarima_model), col = "blue", lty = 1, lwd = 1)
lines(forecast(autoarima_model, h = 24)$mean, col = "red", lty = 2, lwd = 1)
```



```r
lm_model <- lm(observed_values ~ time(observed_values))
summary(lm_model)
```

```
##
## Call:
## lm(formula = observed_values ~ time(observed_values))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.08724 -0.36284 -0.00802  0.43129  1.53303
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           3.9481251  0.5633669   7.008 1.19e-11 ***
## time(observed_values) 0.0028960  0.0003056   9.477  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6126 on 362 degrees of freedom
## Multiple R-squared:  0.1988, Adjusted R-squared:  0.1966
## F-statistic: 89.81 on 1 and 362 DF,  p-value: < 2.2e-16
```

We see some really interesting things in this regression plot. First is that we have a positive coefficient for the year variable, and it is statistically significant. This suggests a warming trend over time in the central England temperature data. When viewing the fitted values, this is also apparent in recent years.

My analysis concurs with Benner's. It agrees with Benner by identifying a warming trend, acknowledging cyclic patterns in temperature variability, and recognizing the non-stationary nature of temperature fluctuations. With respect to Vaidyanathan, my analysis does not coincide with the notion of a hiatus. The warming trend is shown in both the overall trend analysis and the observed, fitted, and forecasted values. The inclusion of cyclic patterns and seasonality in my models further assists this thought. However, it is also important to take into consideration that this analysis is specific to the central England temperature series and may not necessarily represent global temperature trends.