Forecasting AAA Call Volume based on historical records

Basilica Gomes: Time series regression models

AAA Call Volume in Washington

It has been found that temperature has a significant effect on number of calls for road service assistance received by AAA. Temperature alone could explain about half of call volume variability.The goal is to find a way to predict emergency road service call volume received by AAA for future years. Their previous analysis addressed the effect of average daily temperature on emergency road service call volume. Additionally the Box-Jenkins models predicted the future AAA call volume based on the past records with average error rate about 6%.

Therefore we will combine the two approaches and explore the ADL models.

A conversation with the manager of the emergency road service center has led to two important observations: (1) Automakers seem to design cars to operate best at 60 degrees Fahrenheit and (2) call volume seems to increase more sharply when the average temperature drops a few degrees from an average temperature in the 30s than it does when similar drop occurs with an average in the 60s. This information suggests that the effect of temperature on emergency road service is nonlinear.

However the manager believes the number of road service calls received is related to the general economic cycle and that the Washington State unemployment rate is a good surrogate measurement for the general state of Washington's economy. Now he has observed that the cyclical trend of the time series seems to be lagging behind the general economic cycle.

The data on emergency road service call volume, average monthly temperature and the Washington State unemployment rate are given in **AAAdatacleaned.csv**.

1. Create time series objects for the calls, temperature, and unemployment rate variables and plot all variables to identify trends and seasonality.

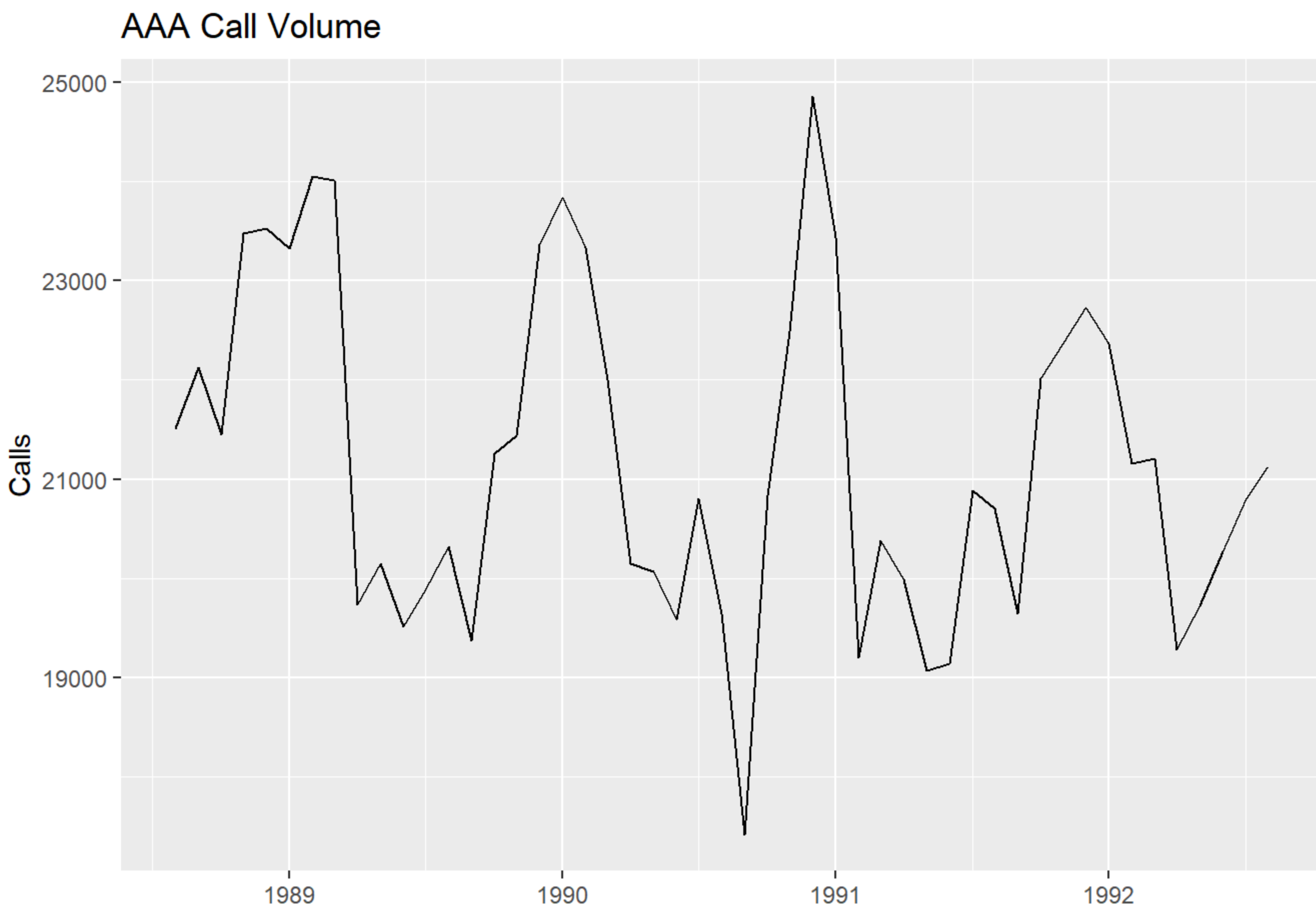
```
library(forecast)
```

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

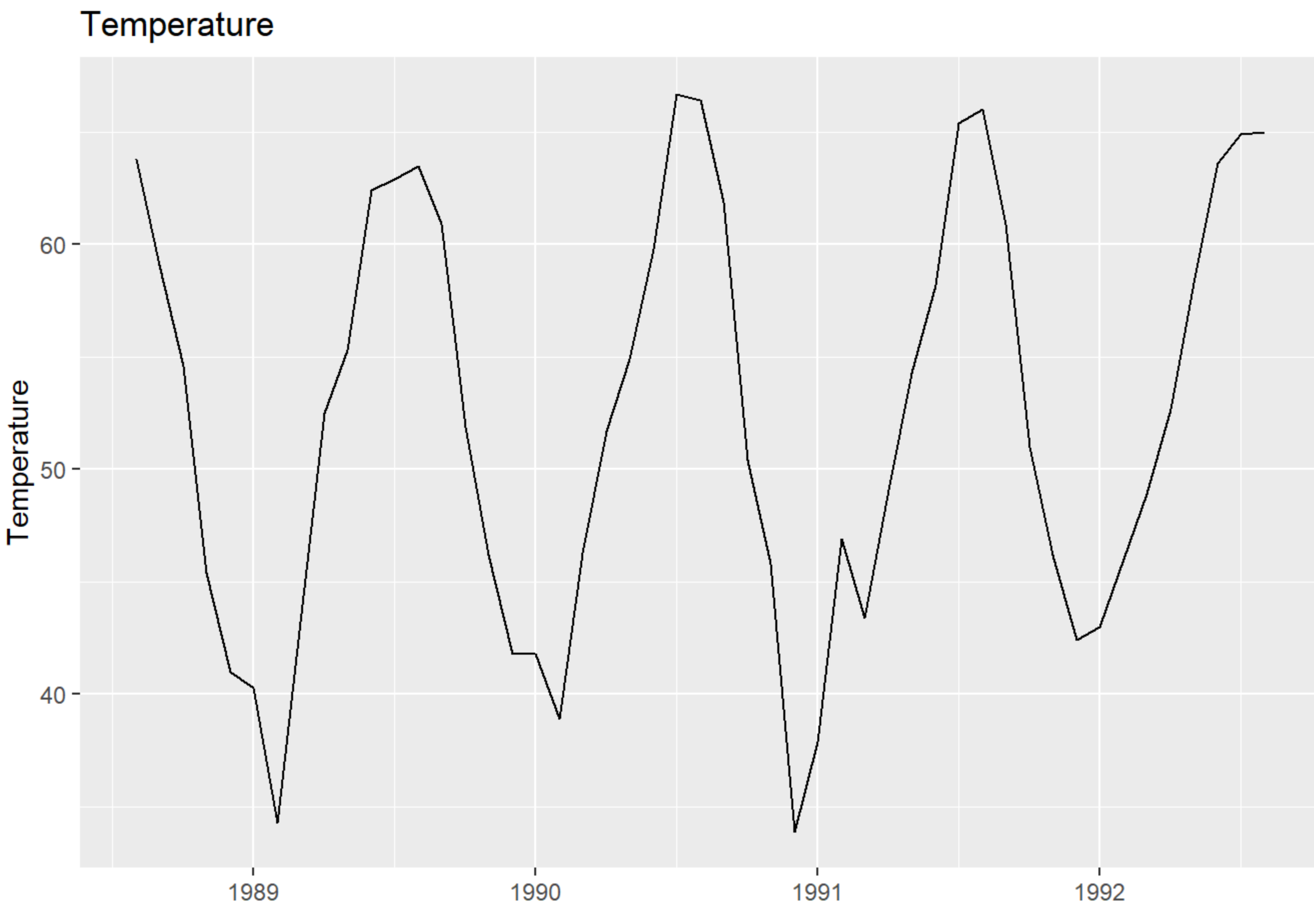
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

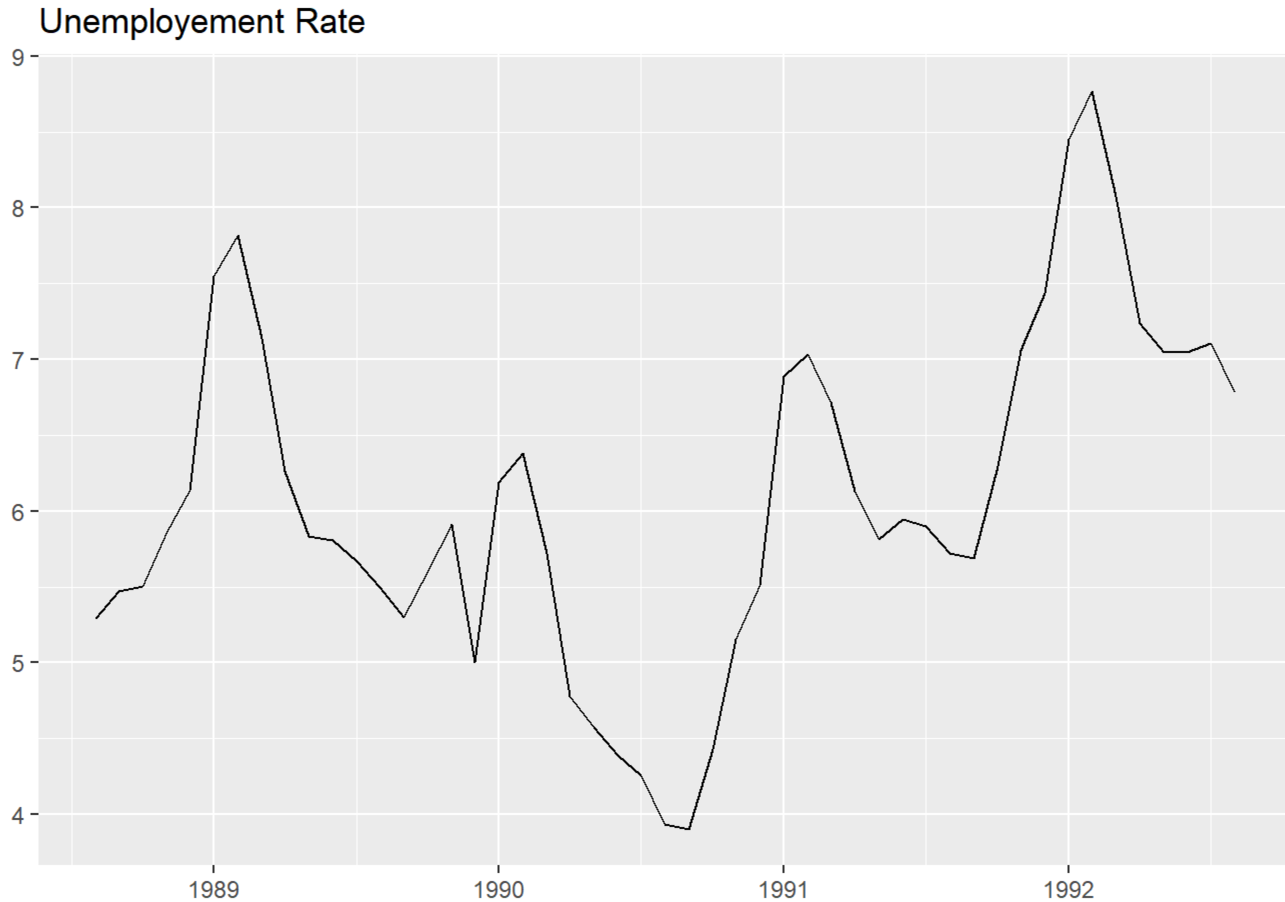
```
# read data
data = read.csv("AAAdatacleaned.csv")

# attach data to call variables without mentioning the dataset
attach(data)

# create ts object
calls.ts = ts(Calls, start=c(1988, 8), end = c(1992,8), 12)
temp.ts = ts(Temp, start=c(1988, 8), end = c(1992,8), 12)
unemp.ts = ts(Rate, start=c(1988, 8), end = c(1992,8), 12)


#plot variables
autoplot(calls.ts, main = "AAA Call Volume", xlab = "", ylab = "Calls")
```




```
autoplot(temp.ts, main = "Temperature", xlab = "", ylab = "Temperature")
```



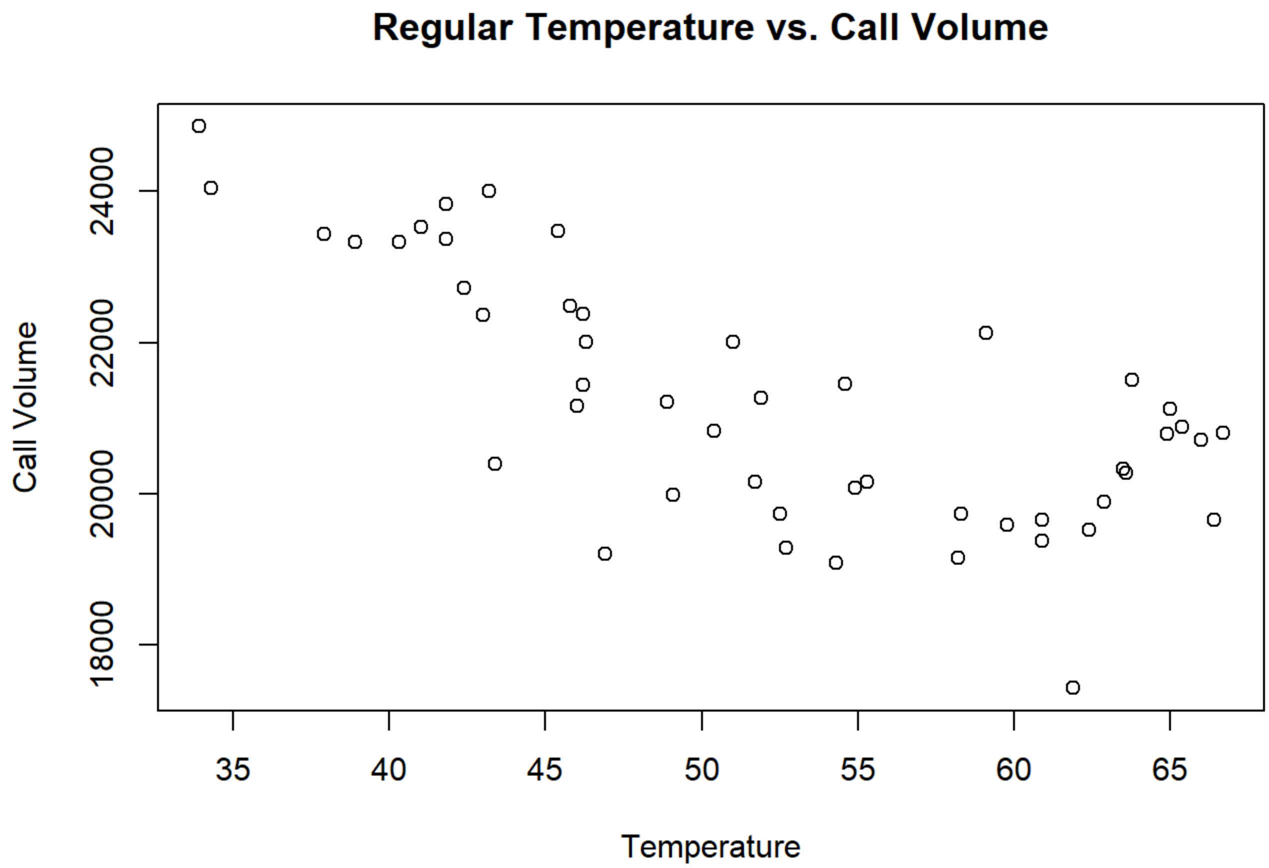

```
autoplot(unemp.ts, main = "Unemployement Rate", xlab = "", ylab = "")
```



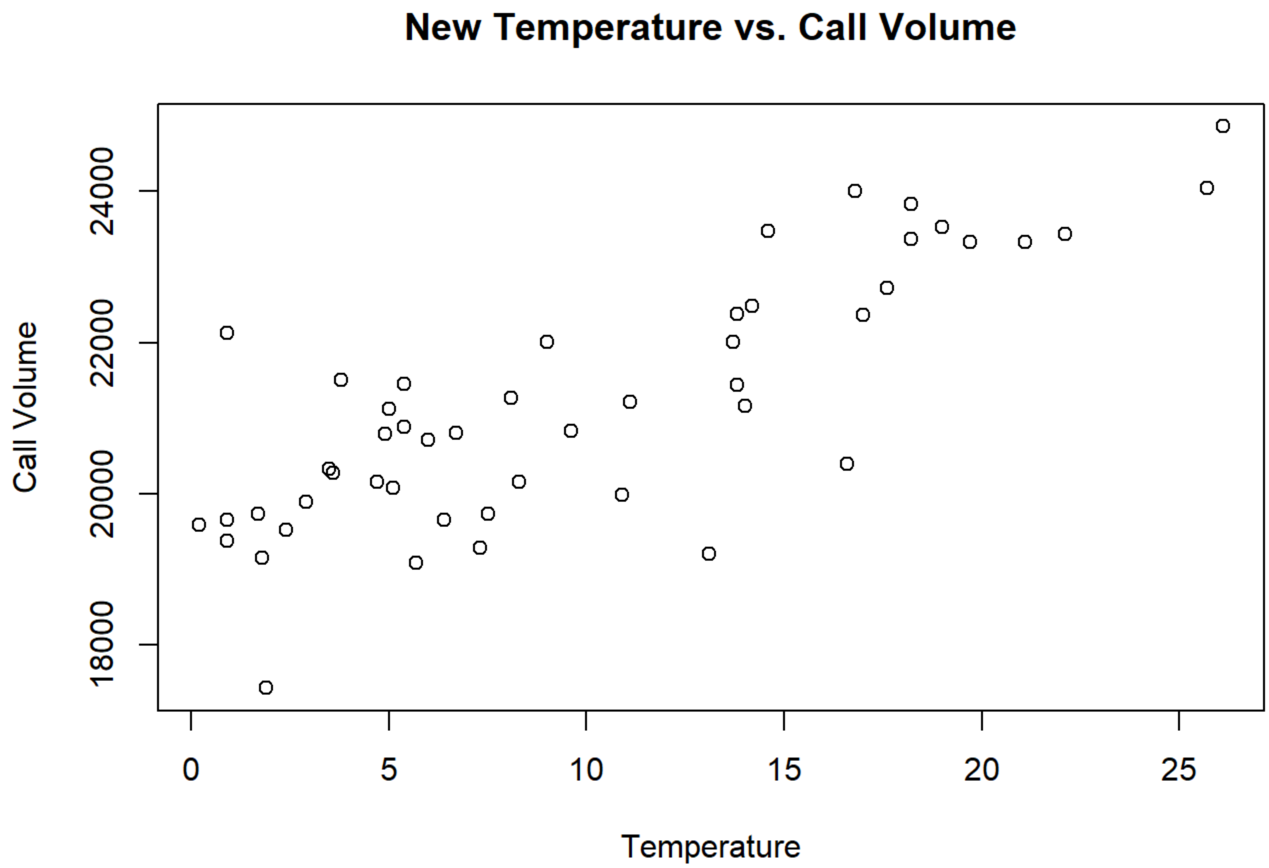
Reviewing calls we see no trend present, it has a seasonal component, level and noise is present. Reviewing Temperature we see no trend present, it has a seasonal component, level and noise present. Reviewing Unemployment rate we see a possible cubic trend, level and noise present, but no seasonal component present.

(@)As temperature has a relative scale and that the selection of the zero point is arbitrary. If vehicles are designed to operate best at 60 degrees Fahrenheit, then every degree above or below 60 degrees should make vehicles operate less reliably. To accomplish a transformation of the temperature data that simulates this effect, begin by subtracting 60 from average monthly temperature values. This re-positions “zero” to 60 degrees Fahrenheit. We will plot two scatter plots using the absolute value of this new temperature variable to identify pattrens and trends. 1. the regular temperature vs. calls volume; 2. the new temperature variable vs. calls volume.

```
# transform temp data
new.temp = abs(Temp-60)
new.temp.ts = ts(abs(Temp - 60), start=c(1988, 8), end = c(1992,8), 12)

# plot regular temperature vs call volume
plot(Temp, Calls, main = "Regular Temperature vs. Call Volume",
     xlab = "Temperature", ylab = "Call Volume")
```



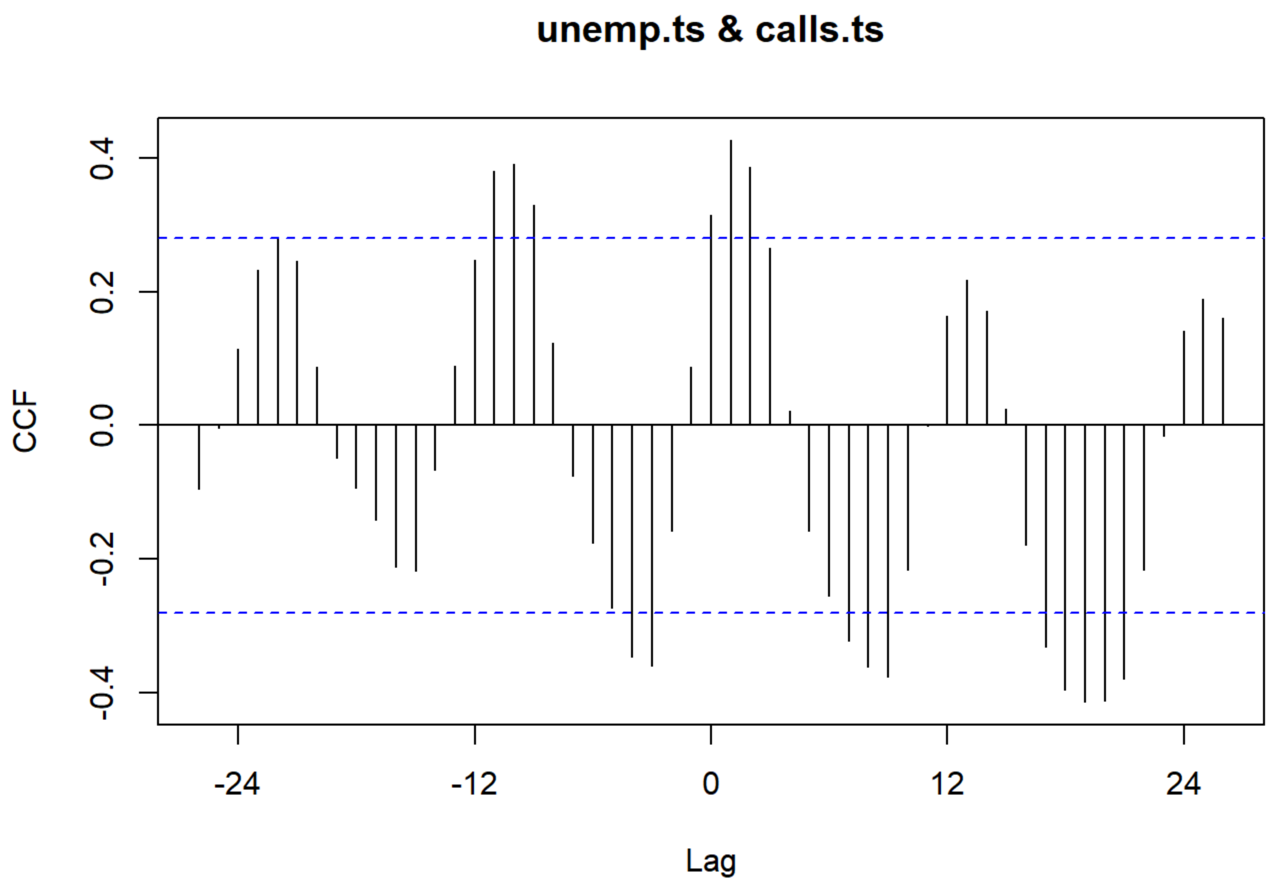

```
# plot regular temperature vs call volume
plot(new.temp, Calls, main = "New Temperature vs. Call Volume",
     xlab = "Temperature", ylab = "Call Volume")
```



Call volume is lowest at 60 degree Fahrenheit and increases significantly in temperatures below 60 degree Fahrenheit. After adjusting temperature call volume is lowest at 0 and increases as temperature increases.

2. Studying the cross-correlations between the unemployment rate and the calls volume. Is the unemployment rate leading the calls and if yes, by how much?

```
Ccf(unemp.ts, calls.ts, 26)
```



Max correlation occurs at lag -3 and -10 meaning that unemployment rate is significant on the call volume

3. Create a lagged unemployment rate variable and relate it to emergency road service. Give unemployment a lagged effect on emergency road service by using the unemployment rate: (1) *three* months prior to the current month and (2) *10* months prior to the current month. Fit two models using (1) and (2) lagged unemployment as predictors.

Are the coefficients of the independent variables significantly different from zero? Which model is better for prediction? Report adjusted coefficients of determination, RMSE, and MAPE for both models.

```
newdata = ts.intersect(calls.ts, unemp.ts, temp.ts,
                       ratelag3 = lag(unemp.ts, -3),
                       ratelag10 = lag(unemp.ts, -10))

m1 = tslm(calls.ts~ratelag3, newdata)
m2 = tslm(calls.ts~ratelag10, newdata)

summary(m1)
```

```
##
## Call:
## tslm(formula = calls.ts ~ ratelag3, data = newdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4256.3  -872.8    34.0   869.3  2948.7
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  23780.0     1272.2  18.692   <2e-16 ***
## ratelag3      -478.3      208.7  -2.292   0.0277 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1497 on 37 degrees of freedom
## Multiple R-squared:  0.1243, Adjusted R-squared:  0.1006
## F-statistic: 5.251 on 1 and 37 DF,  p-value: 0.02772
```

```
summary(m2)
```

```
##
## Call:
## tslm(formula = calls.ts ~ ratelag10, data = newdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3687.5  -789.5   -67.6   619.2  3289.4
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15235.7     1358.3  11.217 1.83e-13 ***
## ratelag10      993.5      234.7   4.234 0.000146 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1313 on 37 degrees of freedom
## Multiple R-squared:  0.3263, Adjusted R-squared:  0.3081
## F-statistic: 17.92 on 1 and 37 DF,  p-value: 0.0001458
```

```
accuracy(m1$fitted.values, calls.ts)
```

```
##                    ME    RMSE      MAE       MPE     MAPE      ACF1 Theil's U
## Test set 3.731773e-13 1457.95 1168.193 -0.482921 5.595074 0.5620345 0.9782629
```

```
accuracy(m2$fitted.values, calls.ts)
```

```
##                     ME     RMSE      MAE        MPE     MAPE      ACF1 Theil's U
## Test set 9.330286e-14 1278.754 982.8276 -0.3711263 4.716643 0.3594786 0.8603499
```

m1 Lag 3: RMSE: 1457.95 MAPE: 5.59 Adjusted Coefficient of Determination: 0.1006 p-value: 0.02772

m2 Lag 10: RMSE: 1278.75 MAPE: 4.71 Adjusted Coefficient of Determination: 0.3081 p-value 0.0001458

The independent variables are significant and different from 0 at both lags. Model 2 (lag 10) is better because it has a lower RMSE and MAPE compared to Model 1 and it has a higher Rsquared

4. Develop a multiple regression equation using the new transformed average temperature variable created in 1. and the lagged unemployment variable created in (3) that describes the calls volume the best.

Is this a good model? Report the adjusted coefficients of determination, RMSE, and MAPE.

Have any of the underlying assumptions been violated?

```
data_avg_temp = ts.intersect(calls.ts, new.temp.ts,
                       ratelag10 = lag(unemp.ts, -10))


m3 = tslm(calls.ts~ratelag10 + new.temp.ts, data_avg_temp)

summary(m3)
```

```
##
## Call:
## tslm(formula = calls.ts ~ ratelag10 + new.temp.ts, data = data_avg_temp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2274.7  -421.5   199.0   470.5  1387.1
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17008.46     881.99  19.284  < 2e-16 ***
## ratelag10     399.75     166.31   2.404   0.0215 *
## new.temp.ts   171.57      22.47   7.635 4.91e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 822.4 on 36 degrees of freedom
## Multiple R-squared:  0.7428, Adjusted R-squared:  0.7285
## F-statistic: 51.98 on 2 and 36 DF,  p-value: 2.428e-11
```

```
accuracy(m3$fitted.values, calls.ts)
```

```
##           ME     RMSE      MAE       MPE     MAPE      ACF1 Theil's U
## Test set  0 790.1433 600.8607 -0.148965 2.943097 0.3357316  0.548137
```

M3: RMSE: 790.14 MAPE: 2.94 Adjusted R-squared: 0.73 p-value 2.428e-11

Yes, this is a good model since it has a lower RMSE and MAPE compared to the previous models. Additionally it has a higher R Squared compared to previous models and the independent variables are significantly different from alpha at 0.05. No model assumptions have been violated.

5. Recommendation on the regression model that is more appropriate for predicting the emergency road service call volume. I recommend using average temperature and unemployment rates to predict the call volume for AAA. Since cars perform at their optimal level with temperatures at 60 degree Fahrenheit, it’s imperative to transform the temperature variable by setting 60 degree Fahrenheit to 0 degree Fahrenheit thus making this variable a robust predictor of call volume. Additionally call volume is affected by the economic state and unemployment rate is a good proxy and valuable in predicting call volume. Both temperature and unemployment have been statistically significant in predicting call volume. However here are a few suggestions to bear in mind while using this analysis to predict call volume. Since the analysis revealed maximum lags for unemployment rate at -3 and -10. upon performing a multiple linear regression using those variables we observed that the model with a lag at -10 performed better than the model with lag at -3. Therefore it is advisable to use a model to predict call volume by first adjusting temperature from 60 Degree F to 0 degree F and setting unemployment lag at -10