

AAA Case

Yuxin Yao

2024-12-15

AAA Washington Case

Based on excellent performance and great recommendations from Mr. Tux Michael DeCoria had offered you a contract as a consultant to AAA. Your goal is to assist Mr DeCoria in finding a way to predict emergency road service call volume for future years. Their previous analysis addressed the effect of average daily temperature on emergency road service call volume. It has been found that the temperature's effect is significant and could explain about half of calls volume variability. You have also used the Box-Jenkins models and found that you can predict the future AAA calls volume based on the past records with average error rate about 6%.

After discussing your findings with Michael you decide to combine the two approached 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.

Michael DeCoria stated that he 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**.

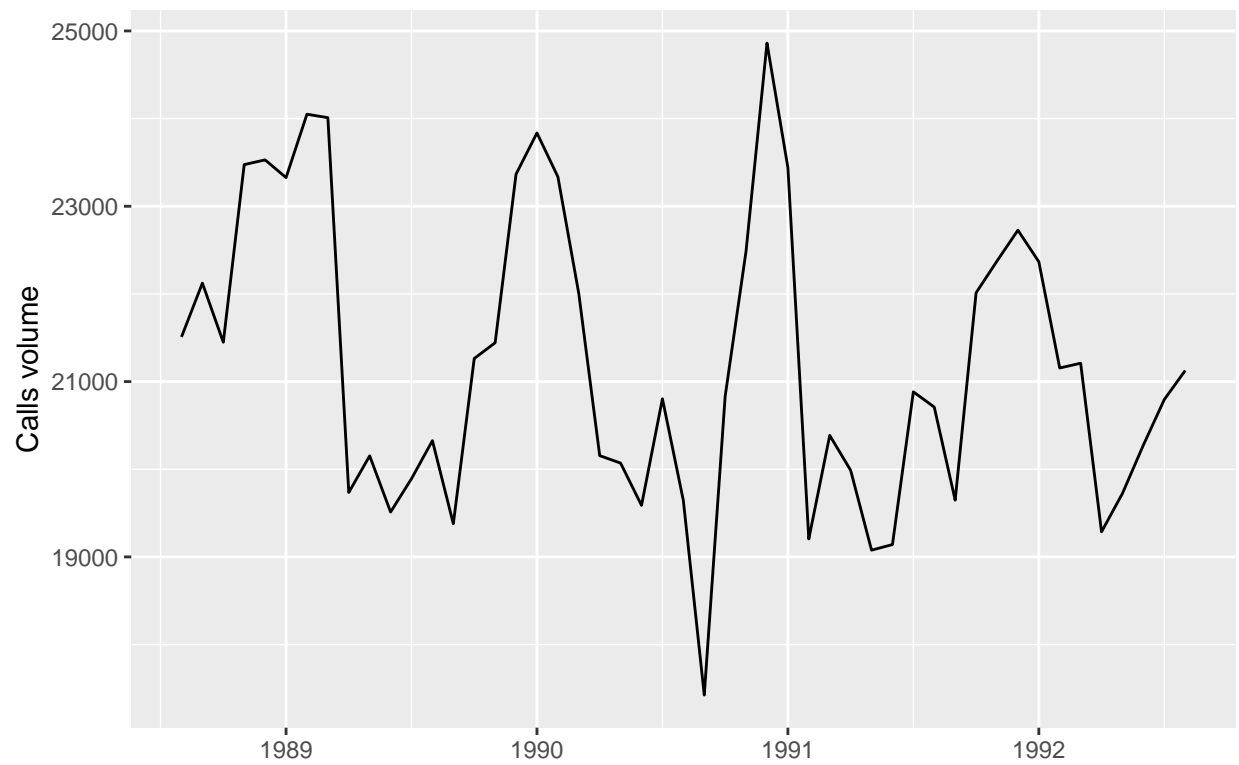
- (1) Upload the data into R. Create time series objects for the calls, temperature, and unemployment rate variables. Plot all three variables, label the axis. Comment on the time series.

```
data = read.csv("AAAdatacleaned.csv")
Call.ts = ts(data$Calls, start = c(1988,8),
             frequency = 12)
Temp.ts = ts(data$Temp, start = c(1988,8), frequency = 12)
Rate.ts = ts(data$Rate, start = c(1988,8), frequency = 12)

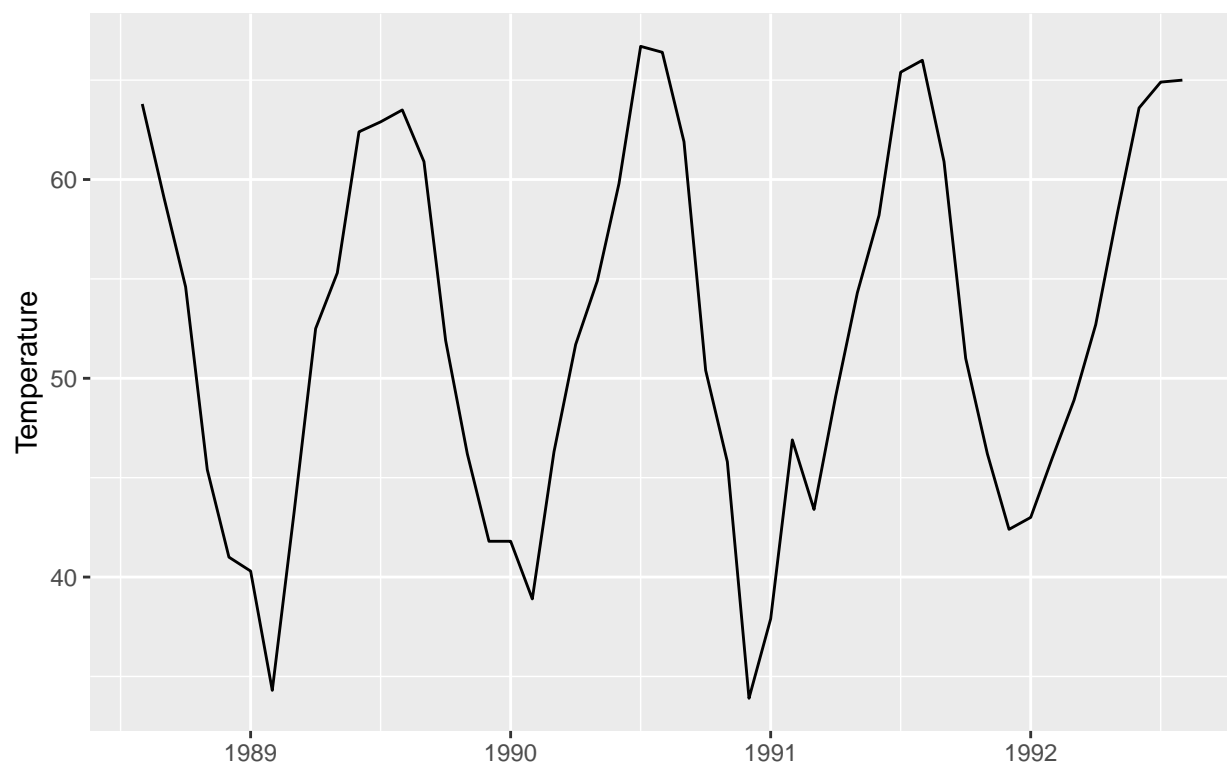
library(forecast)

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

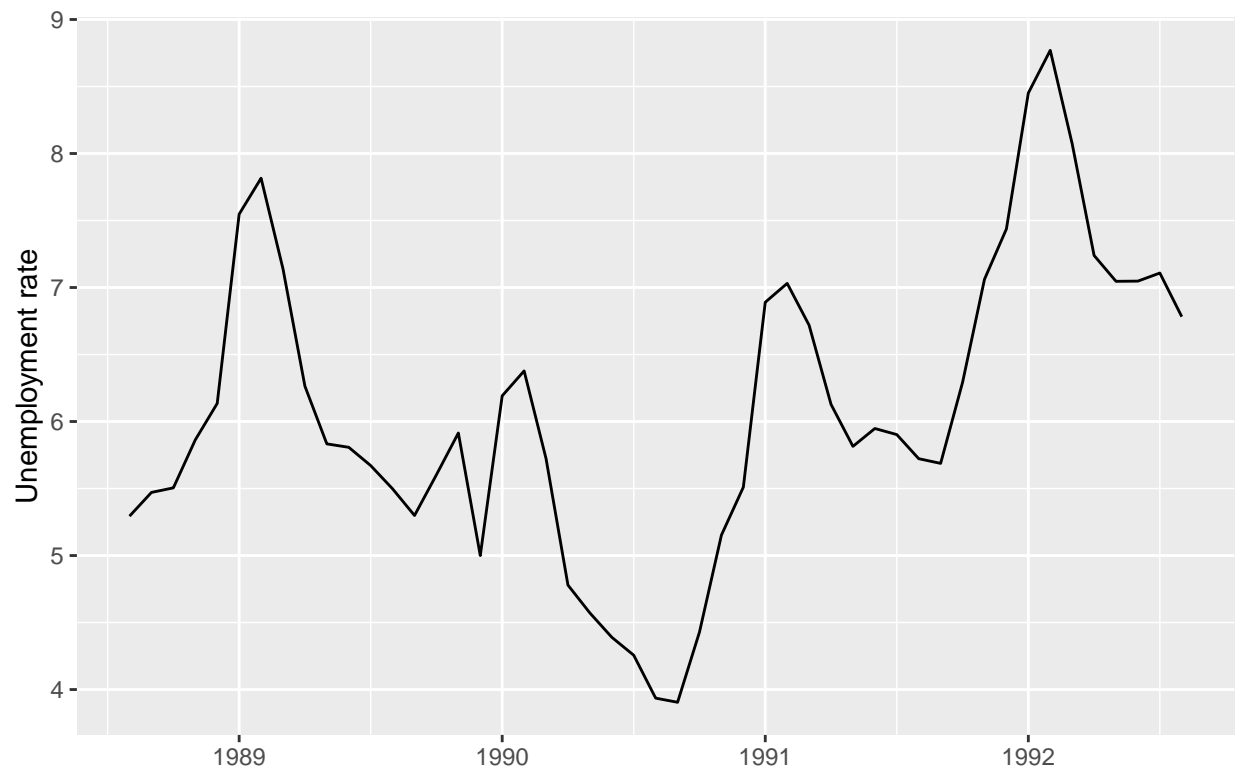
```
autoplot(Call.ts, xlab='', ylab='Calls volume')
```



```
autoplot(Temp.ts, xlab='', ylab='Temperature')
```



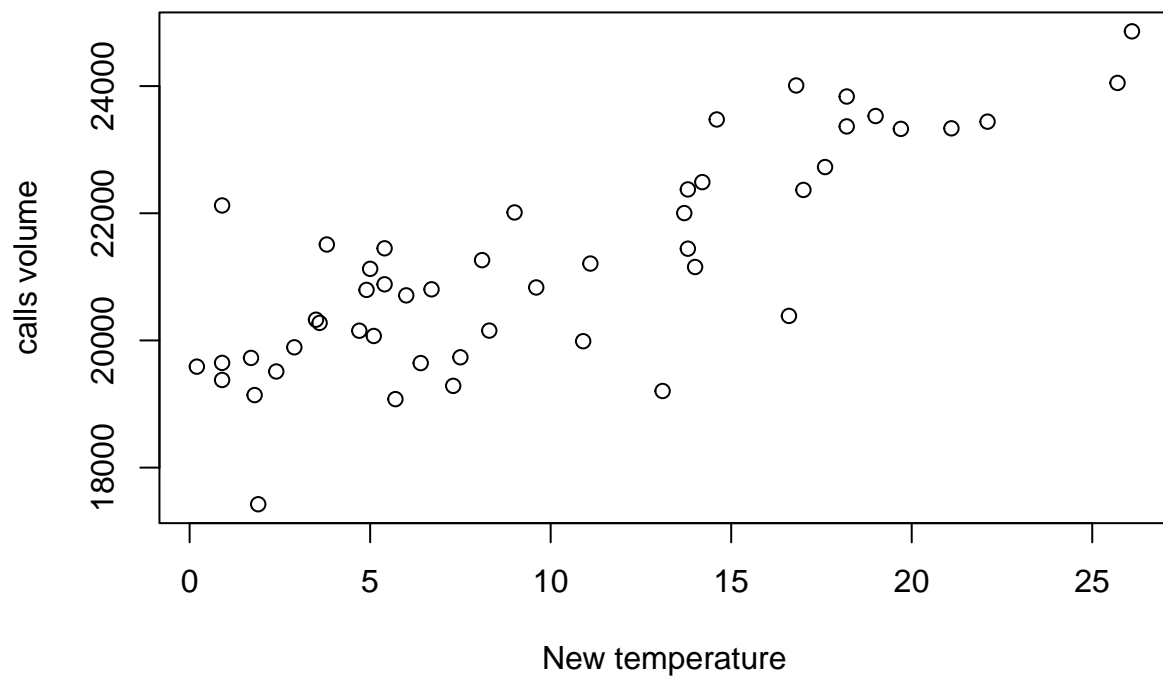
```
autoplot(Rate.ts, xlab='', ylab='Unemployment rate')
```



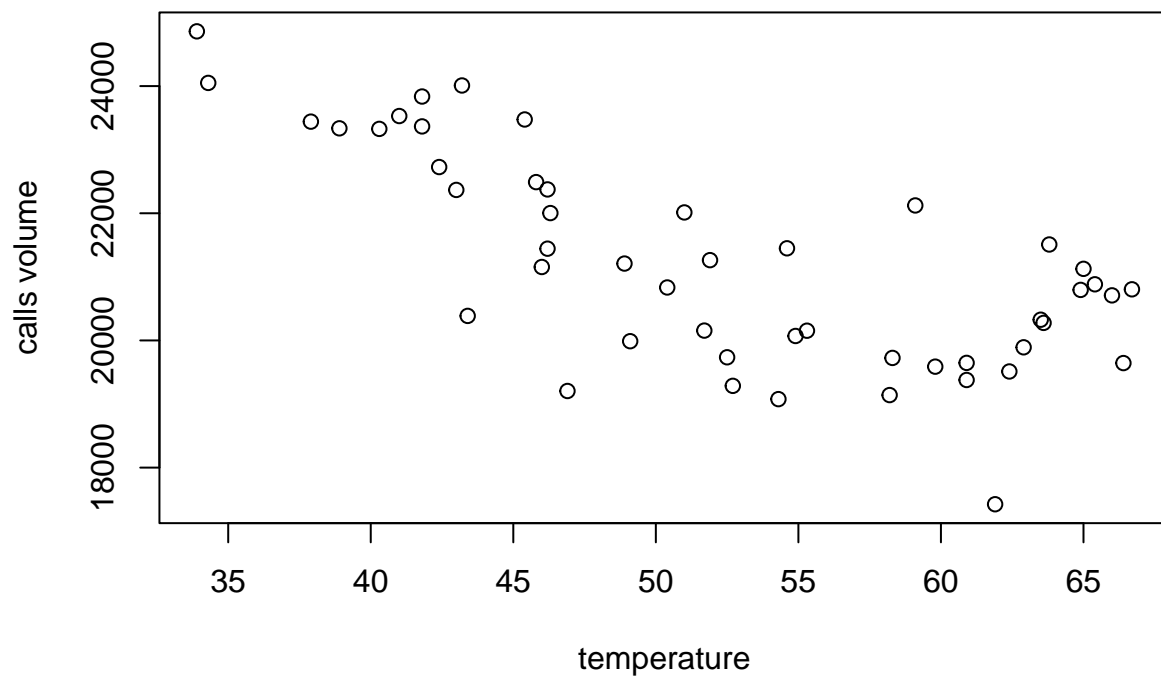
there is level, noise, seasonality

- (2) Create a new temperature variable. Remember that temperature is 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 repositions “zero” to 60 degrees Fahrenheit. Use the absolute value of this new temperature variable. Plot two scatterplots: 1. the regular temperature vs. calls volume; 2. the new temperature variable vs. calls volume. Comment on the patterns. Include your script and the plot.

```
newtemp.ts = abs(Temp.ts - 60)
plot(as.vector(newtemp.ts), as.vector(Call.ts), xlab = "New temperature", ylab = "calls volume")
```



```
plot(as.vector(Temp.ts), as.vector(Call1.ts), xlab = "temperature", ylab = "calls volume")
```

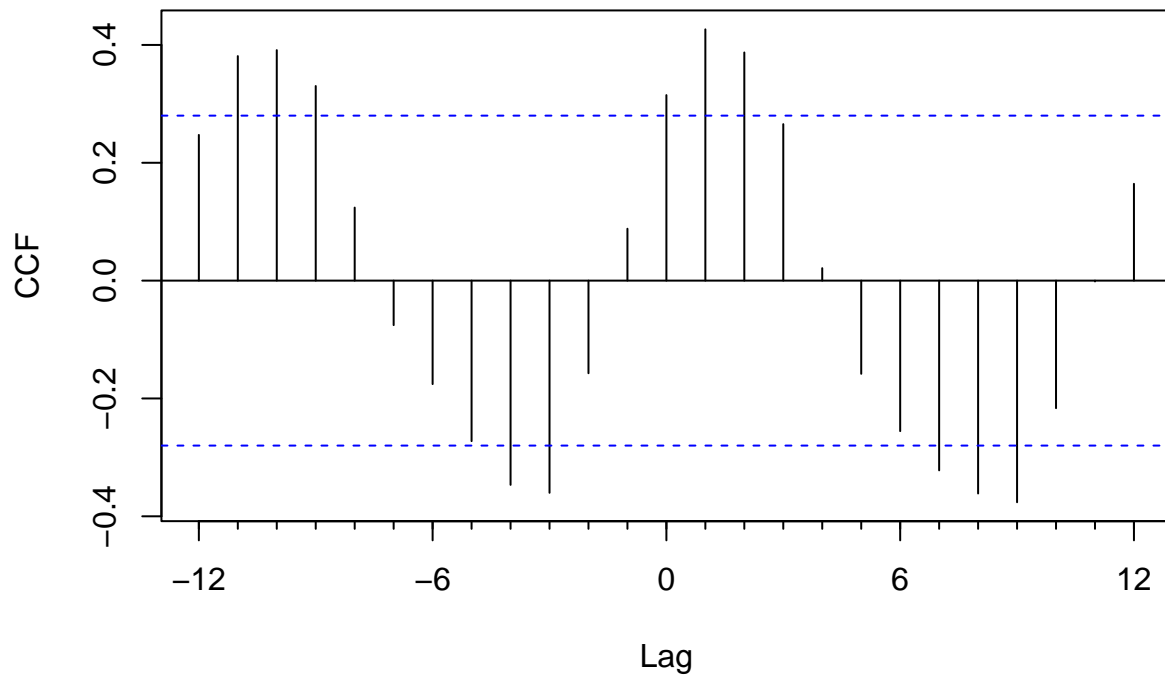


It is not pretty linear. the temperature below 60 seems to be downward trend, the temperature above 60 seems to be upward trend.

- (3) Study the cross-correlations between the unemployment rate and the calls volume. Is the unemployment rate leading the calls and if yes, by how much? Include your script, comments, and the ccf plot.

```
tmp=Ccf(Rate.ts, Call.ts, 12)
```

Rate.ts & Call.ts



tmp

```
##
## Autocorrelations of series 'X', by lag
##
##      -12      -11      -10      -9      -8      -7      -6      -5      -4      -3      -2
##  0.247  0.381  0.391  0.330  0.124 -0.076 -0.176 -0.272 -0.347 -0.360 -0.157
##      -1       0       1       2       3       4       5       6       7       8       9
##  0.088  0.315  0.426  0.387  0.266  0.021 -0.158 -0.255 -0.322 -0.361 -0.376
##      10      11      12
## -0.216 -0.001  0.164
```

There is a significant correlation between calls volume and unemployment rates at lags -3 and -10. Let's build two models.

- (4) Create a lagged unemployment rate variable and relate it to emergency road service based on your ccf analysis. Fit the model using lagged unemployment as a predictor. Are the coefficients of the independent variables significantly different from zero? Report adjusted coefficients of determination, RMSE, and MAPE for both models.

```
data$LaggedRate = lag(data$Rate, -4)

lagged_model = lm(Calls ~ LaggedRate, data = data)

summary(lagged_model)
```

```
##
## Call:
## lm(formula = Calls ~ LaggedRate, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2767.6 -1239.7  -329.8   963.8  3921.2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  18371.7     1267.1   14.500  <2e-16 ***
## LaggedRate    466.1       204.9    2.274   0.0275 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1598 on 47 degrees of freedom
## Multiple R-squared:  0.09915,    Adjusted R-squared:  0.07998
## F-statistic: 5.173 on 1 and 47 DF,  p-value: 0.02755
```

```
# Predicted values
predicted_lagged <- predict(lagged_model)
accuracy(predicted_lagged, data$Calls)
```

```
##              ME      RMSE      MAE      MPE      MAPE
## Test set 4.603014e-12 1564.789 1319.937 -0.5377443 6.208598
```

Both the intercept and the coefficient for LaggedRate are significantly different from zero. and the r-squared is low, RMSE is a little high.

- (5) Develop a multiple regression equation using the new transformed average temperature variable created in and the lagged unemployment variable created earlier that describe 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?

```
multi_model <- lm(Calls ~ LaggedRate + abs(Temp - 60), data = data)
summary(multi_model)
```

```
##
## Call:
## lm(formula = Calls ~ LaggedRate + abs(Temp - 60), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2573.8  -380.5    92.7   492.6  2643.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  19544.54     802.16   24.365  < 2e-16 ***
## LaggedRate   -43.83     140.81   -0.311    0.757
## abs(Temp - 60)  194.08     22.45    8.644 3.39e-11 ***
```

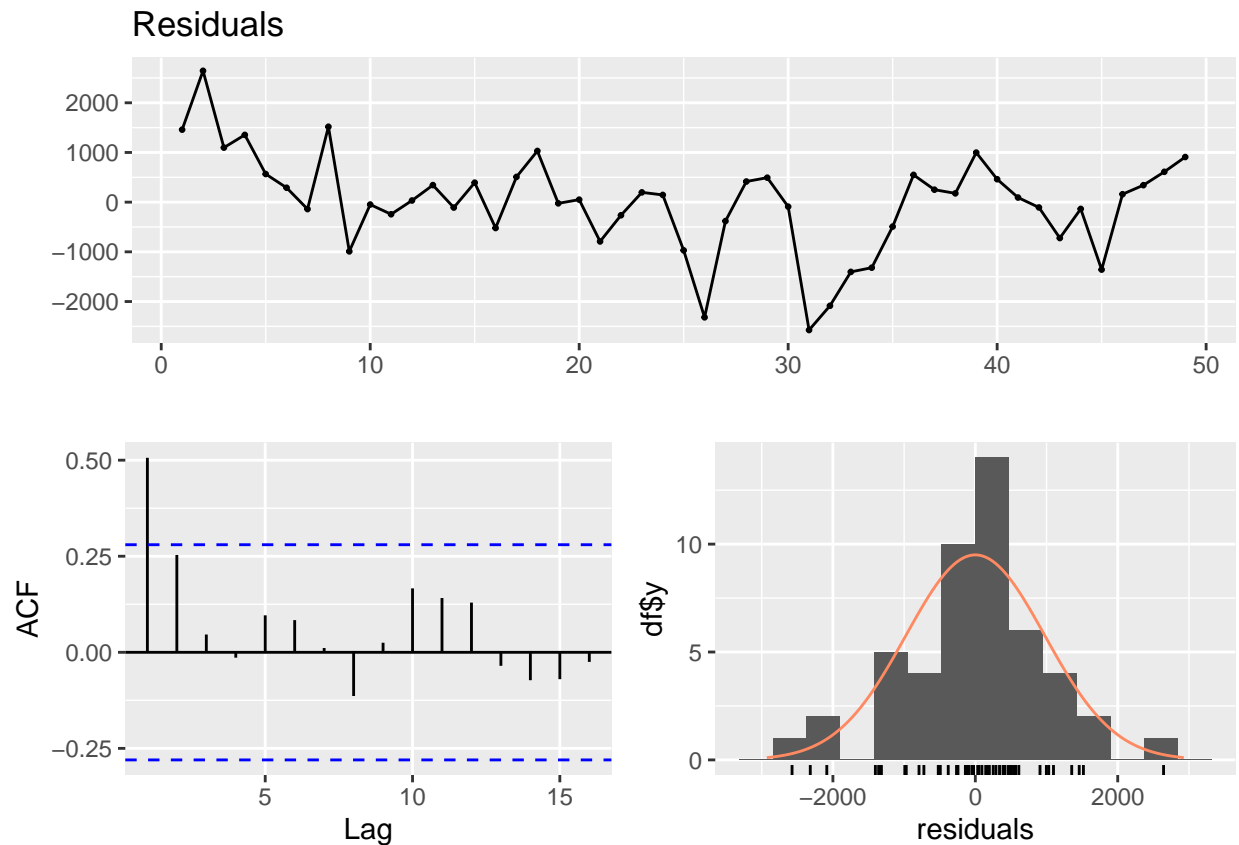


```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 996.9 on 46 degrees of freedom
## Multiple R-squared:  0.6567, Adjusted R-squared:  0.6418
## F-statistic:    44 on 2 and 46 DF,  p-value: 2.088e-11
```

```
predicted_multi <- predict(multi_model)
accuracy(predicted_multi, data$Calls)
```

```
##              ME      RMSE      MAE      MPE      MAPE
## Test set 7.424521e-12 965.9355 697.4009 -0.2154717 3.359482
```

```
checkresiduals(multi_model)
```



```
##
## Breusch-Godfrey test for serial correlation of order up to 10
##
## data: Residuals
## LM test = 18.421, df = 10, p-value = 0.04827
```

R-squared is not bad, but RMSE is a little high. It is fairly normal, However, they are not independent, see ACF at lag 1. So the model needs to improve.

- (6) Prepare a memo to Michael recommending the regression model you believe is more appropriate for predicting the emergency road service call volume. Write it in a paragraph form and keep in mind that Mr. DeCoria is not an expert in time series analysis or R. Therefore, don't be too technical.

Based on the analysis performed, I recommend using a multiple regression model that combines lagged unemployment rates and a transformed temperature variable (absolute deviation from 60°F) to predict emergency road service call volume. The model is effective because it combines the nonlinear effects of economic factors and temperature on vehicle performance. The results show that the transformed temperature variable is a significant predictor with a strong relationship with call volume. The model has a high adjusted R-squared value of 64.18%, which means that it explains a large portion of the variation in call volume. what's more, it provides a low RMSE and MAPE, confirming the accuracy of its predictions.

While the unemployment rate not show a strong independent effect, combining it with the temperature variable helps capture other trends in the data. The model's residuals were tested and while some slight serial correlation was detected, the model was generally robust and provided reliable forecasts. I believe this approach balances complexity and practicality to ensure accurate predictions while maintaining comprehensibility for operational use.