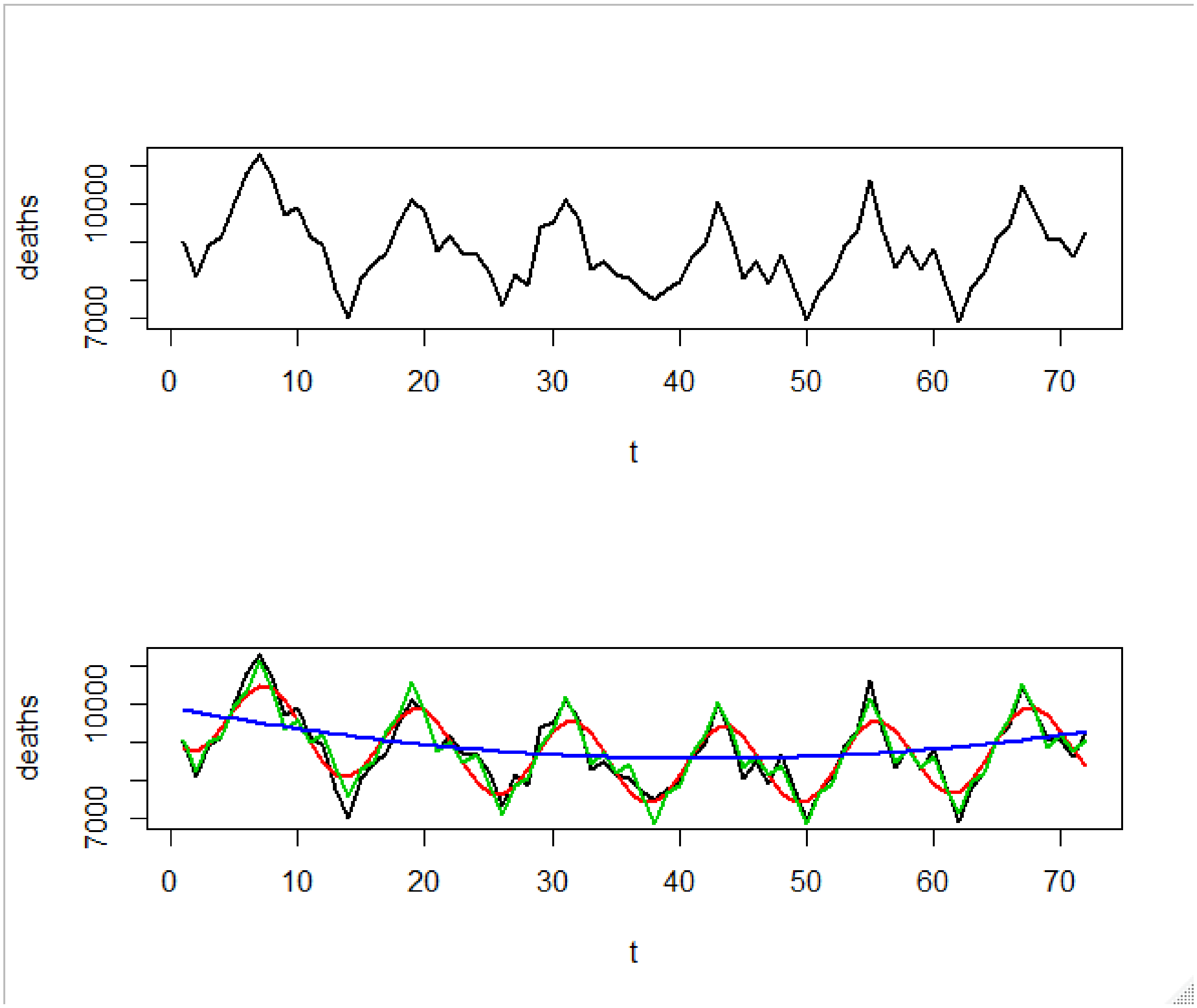
Time Series Analysis of US Accidental Deaths from 1973 to 1979

Ben Hairston

This report contains an analysis of time series data from the United States from 1973 to 1979. The main objective of the analysis is to find trends and try to use those trends to predict future accidental deaths. First there will be a cosine model compared to a monthly average model.

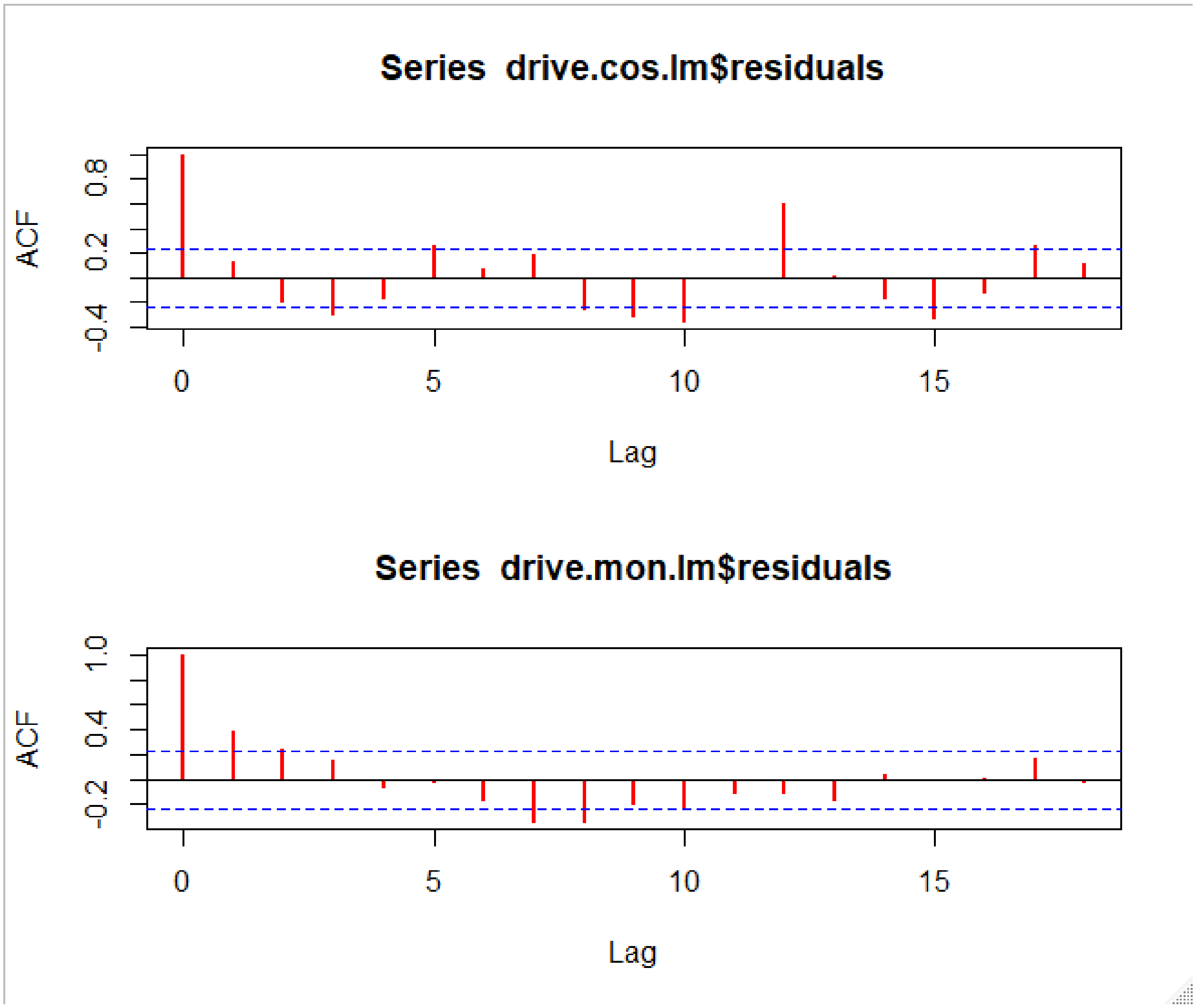


Based on the graph above the cosine model comes close to showing the trends between months and seasons, but the monthly average model tends to capture the trends better.

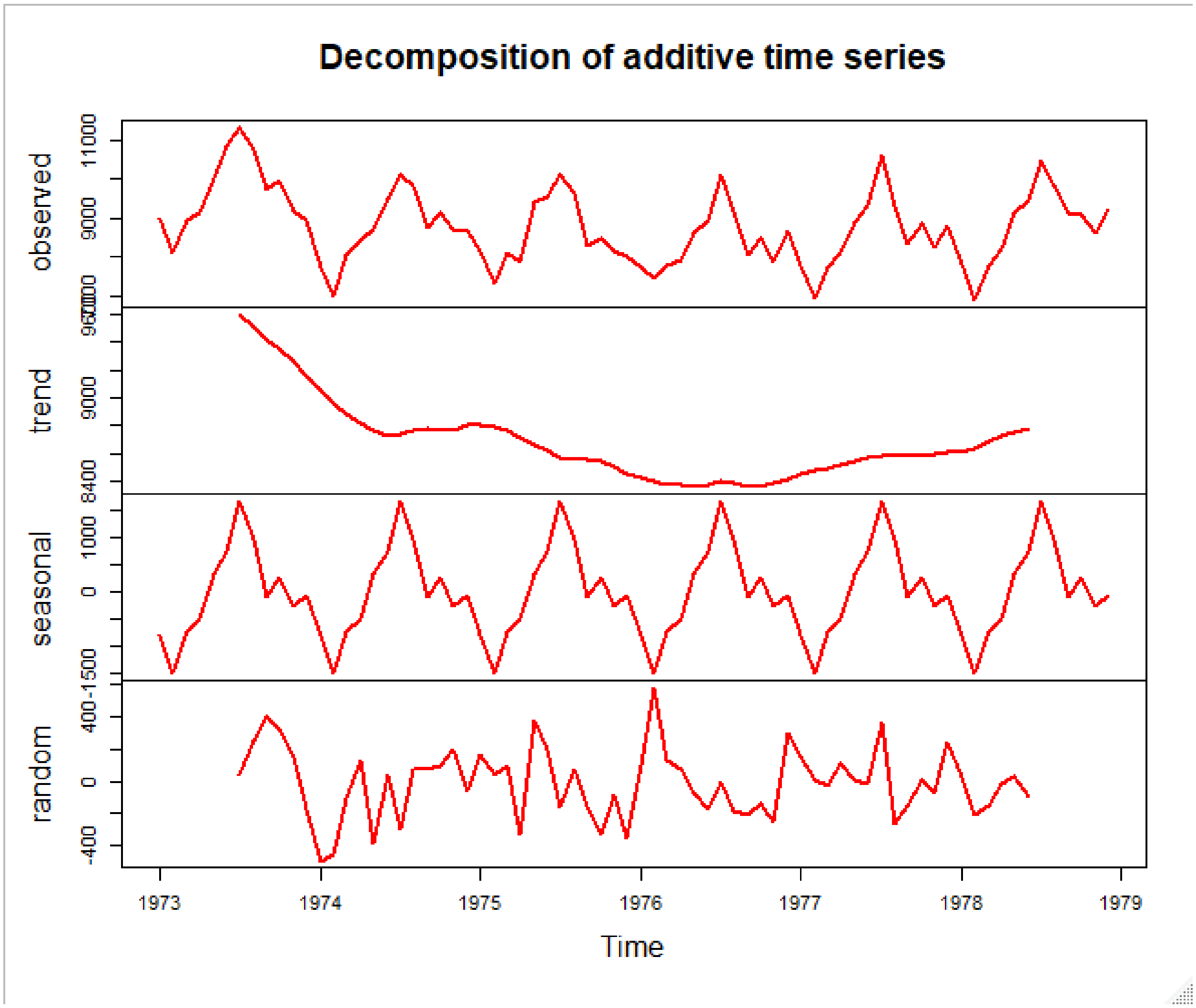
```
Call:
lm(formula = deaths ~ t + I(t^2) + Xcos + Xsin, data = drive)

Residuals:
     Min       1Q   Median       3Q      Max 
-1114.96  -259.16   -30.49   291.56  1097.49 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 9934.3500   180.8306  54.937  < 2e-16 ***
t            -71.6446    11.4069  -6.281 2.86e-08 ***
I(t^2)         0.8330     0.1514   5.503 6.33e-07 ***
Xcos        -734.9544    82.7114  -8.886 6.13e-13 ***
Xsin        -754.8832    83.3130  -9.061 2.98e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 495.8 on 67 degrees of freedom
Multiple R-squared:  0.7471,	Adjusted R-squared:  0.732 
F-statistic: 49.48 on 4 and 67 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = deaths ~ t + I(t^2) + month, data = drive)

Residuals:
    Min      1Q  Median      3Q     Max 
-587.05 -155.30   -8.24  177.54  609.34 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  9.419e+03  1.458e+02  64.626  < 2e-16 ***
t           -7.203e+01  6.324e+00 -11.390  < 2e-16 ***
I(t^2)       8.279e-01  8.388e-02   9.871 5.02e-14 ***
monthAug     1.523e+03  1.588e+02   9.596 1.40e-13 ***
monthDec     5.182e+02  1.592e+02   3.256  0.00189 ** 
monthFeb    -1.026e+03  1.587e+02  -6.468 2.28e-08 ***
monthJan    -2.860e+02  1.587e+02  -1.802  0.07679 .  
monthJul     2.217e+03  1.587e+02  13.970  < 2e-16 ***
monthJun     1.348e+03  1.587e+02   8.496 9.07e-12 ***
monthMar    -2.296e+02  1.586e+02  -1.447  0.15329    
monthMay     8.639e+02  1.586e+02   5.445 1.10e-06 ***
monthNov     2.614e+02  1.590e+02   1.643  0.10569    
monthOct     7.794e+02  1.589e+02   4.904 7.94e-06 ***
monthSep     4.829e+02  1.588e+02   3.041  0.00354 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 274.8 on 58 degrees of freedom
Multiple R-squared:  0.9328,	Adjusted R-squared:  0.9177 
F-statistic: 61.89 on 13 and 58 DF,  p-value: < 2.2e-16
```

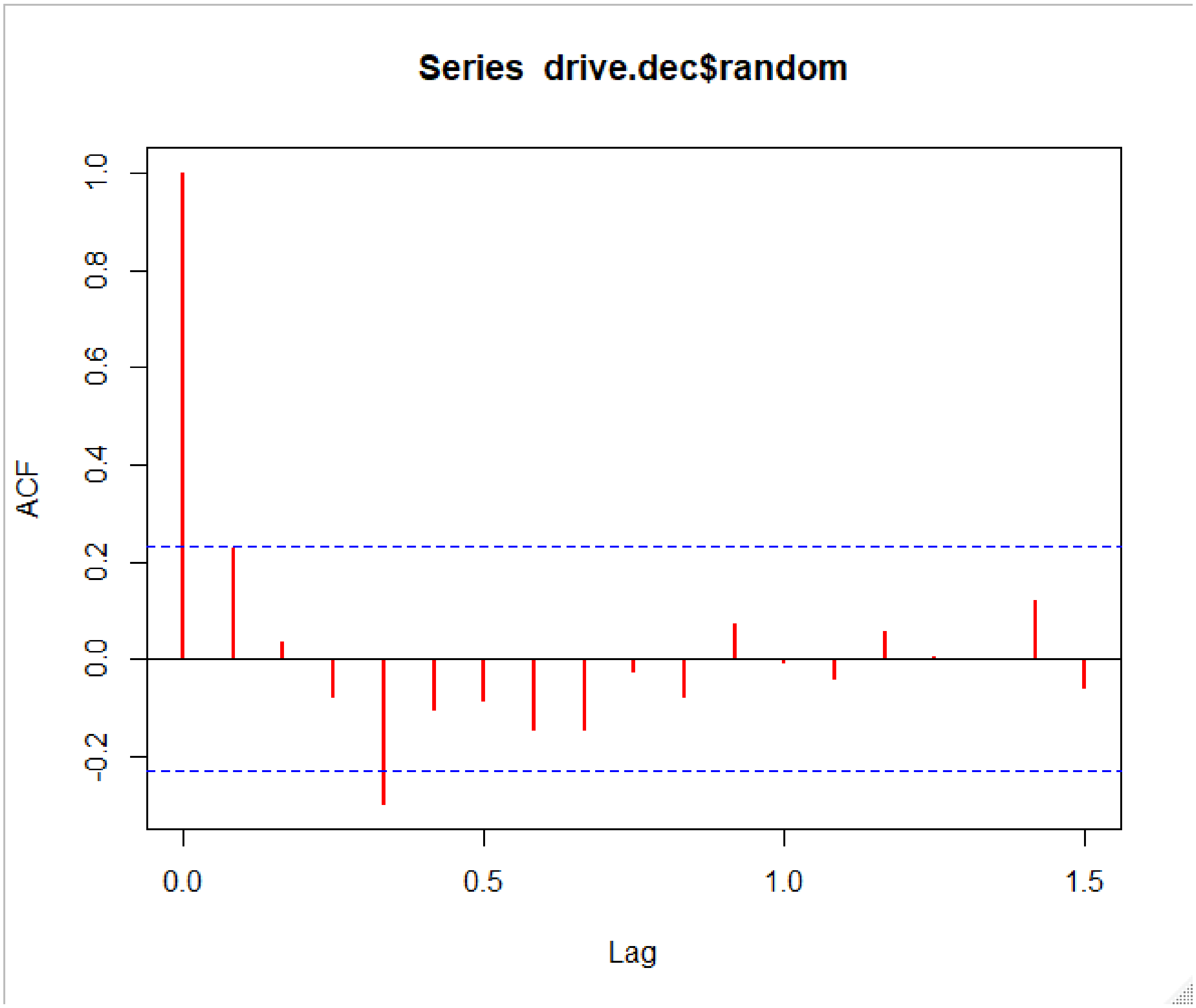
In both summaries the t-value is significant, and the cosine model explains about 73% of variability and the monthly averages model explains about 92% of the variability in the response.

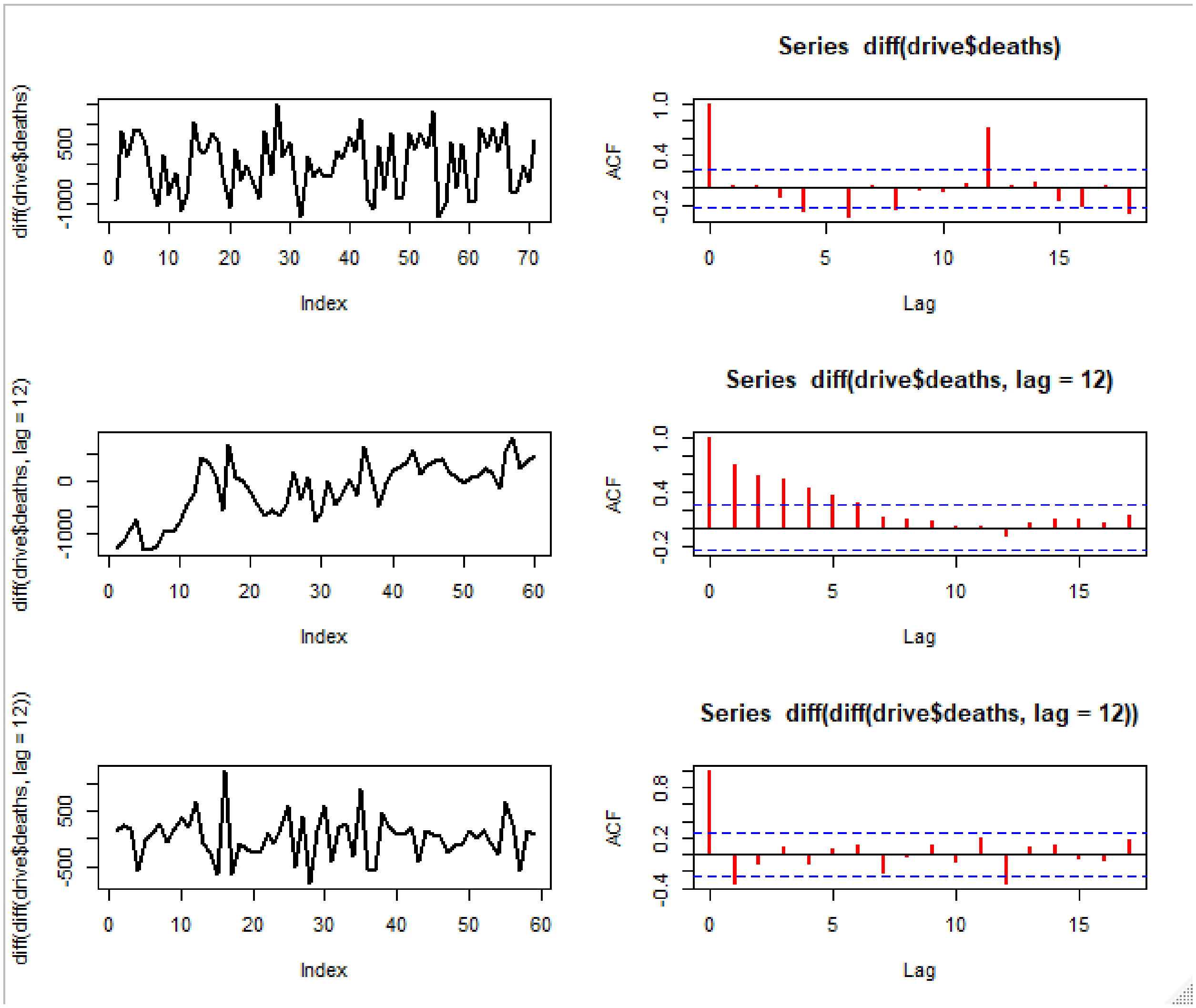


Based on the autocorrelation models, there is statistically significant autocorrelation in the residuals.



Decomposition of additive time series
observed
trend
seasonal
random
Time
1973
1974
1975
1976
1977
1978
1979


Series drive.dec$random
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0.0
0.5
1.0
1.5
Lag



Using variations of the original series and the autocorrelation functions, there is a plot that is different by lag1, the middle by lag 12, and the bottom one is the combination of both. The bottom graph shows significant autocorrelation near lag 1 and lag12 but lag 1 has a higher autocorrelation.

Arima Model 1

```
Call:
arima(x = USAccDeaths, order = c(1, 1, 0), seasonal = list(order = c(1, 1, 0)))

Coefficients:
          ar1      sar1
      -0.3325   -0.3515
s.e.   0.1214    0.1250

sigma^2 estimated as 116222:  log likelihood = -428.63,  aic = 863.27
```

Arima Model 2

```
Call:
arima(x = USAccDeaths, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1)))

Coefficients:
          ma1     sma1
      -0.4303  -0.5528
s.e.   0.1228   0.1784

sigma^2 estimated as 99347:  log likelihood = -425.44,  aic = 856.88
```

Arima Model 3

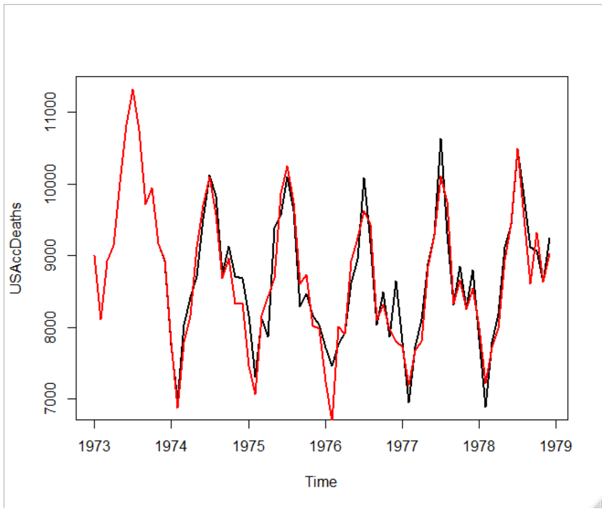
```
Call:
arima(x = USAccDeaths, order = c(1, 1, 1), seasonal = list(order = c(1, 1, 1)))

Coefficients:
         ar1      ma1    sar1     sma1
      0.0762  -0.4867  0.3004  -0.9940
s.e.  0.2967   0.2601  0.1776   1.2269

sigma^2 estimated as 81497:  log likelihood = -424.99,  aic = 859.98
>
```

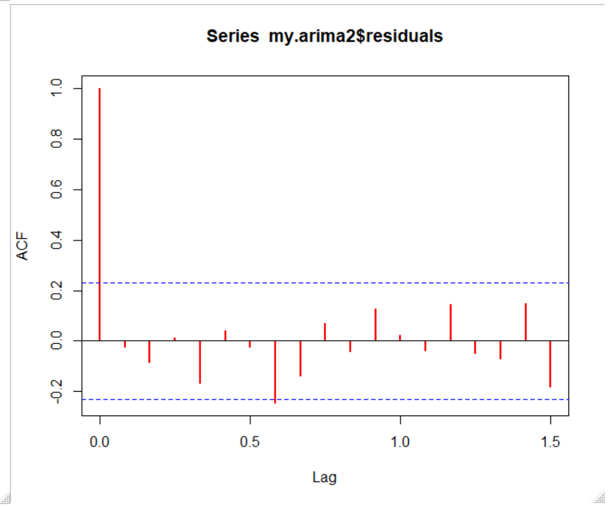
Similar to the UKDriverDeaths analysis, the first 2 models have their coefficients with 2 or more standard errors from zero, but the third model is less than one standard error from 0. Again, the sigma^2 for the third model is the lowest out of all all three models. For model 2 the BIC is the lowest, beating model 1 by about 7 BIC.

```
> AIC(my.arima1,my.arima2,my.arima3)
          df      AIC
my.arima1  3 863.2667
my.arima2  3 856.8800
my.arima3  5 859.9815
> BIC(my.arima1,my.arima2,my.arima3)
          df      BIC
my.arima1  3 869.4993
my.arima2  3 863.1126
my.arima3  5 870.3692
```

```
> predict(my.arima1,3)$pred
          Jan      Feb      Mar
1979 8319.918 7413.229 8266.869
> predict(my.arima2,3)$pred
          Jan      Feb      Mar
1979 8336.061 7531.829 8314.644
> predict(my.arima3,3)$pred
          Jan      Feb      Mar
1979 8256.991 7417.119 8238.220
```

Predictions are all different between the arima models. The autocorrelation has been reduced and comparing the original values with the fitted ones shows that the model follows the original series very closely.





Conclusion

After analyzing the time series of US accidental deaths between 1973 and 1979. The cosine model did a good job at predicting values if we didn't take into account the highs and lows between seasons. The monthly average model was better at predicting extremes but it required many more coefficients, yet both models had a lot of autocorrelation at several lags. Using the ACF function, the autocorrelation was reduced and good predictions were made as evidenced by the last 2 graphs.