Business Forecasting

Final Exam: Crime Data

Ronak Parikh

Business Forecasting Final Exam

Introduction

Crime in US has been steadily decreasing over the years. In the US, FBI tracks crime data. Data is tracked by type of crime. For this exercise, we will focus on robberies, theft, and larceny. See https://www.fbi.gov/services/cjis/ucr for more detail.

Import Data

Note: For convenience and ease of analysis, I have merged the month and the year columns as date column in the CSV.

	A	B	C	D
1	Month	Year	Date	Data
2	January	2008	Jan-08	1392
3	February	2008	Feb-08	1031
4	March	2008	Mar-08	1301
5	April	2008	Apr-08	1420
6	May	2008	May-08	1574
7	June	2008	Jun-08	1587
8	July	2008	Jul-08	1731
9	August	2008	Aug-08	1743
10	Septembe	2008	Sep-08	1473
11	October	2008	Oct-08	1422

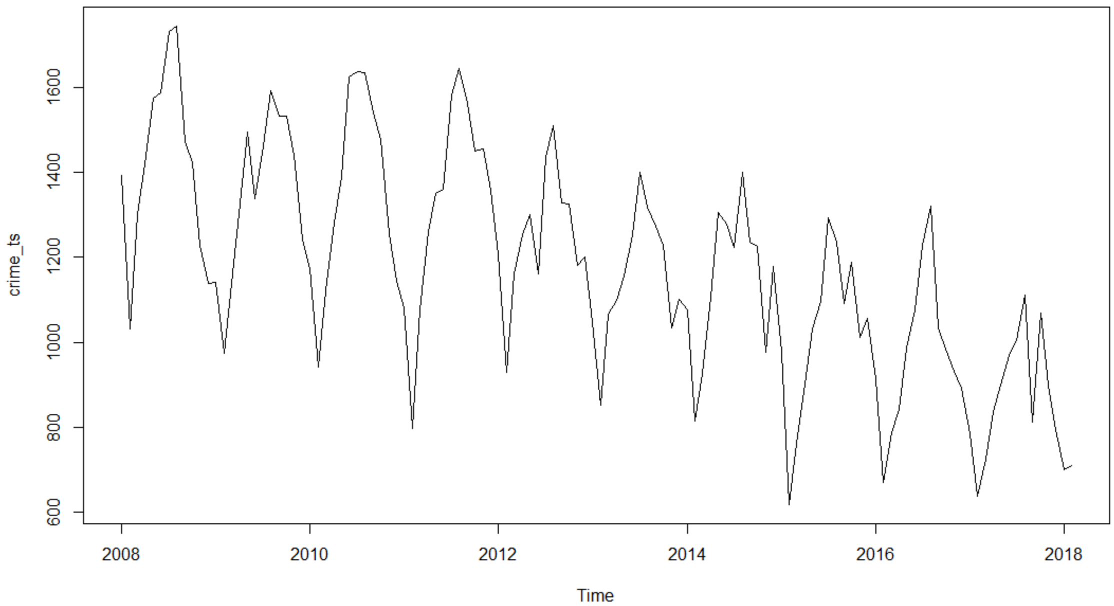
Please do the following steps once the csv file is on your desktop.

```
> library(readr)
Warning message:
package ‘readr’ was built under R version 3.4.4
> crimes <- read_csv("C:/Users/asher/Desktop/Asher_MS/BF/Final/Crimes.csv")
Parsed with column specification:
cols(
  Month = col_character(),
  Year = col_integer(),
  Date = col_character(),
  Data = col_integer()
)
> View(Crimes)
Error in View : object 'Crimes' not found
> View(crimes)
```

```
crime_ts <- ts(crimes$Data, start=c(2008,1),frequency = 12)
  plot(crime_ts)
```

Plot and Inference

- Show a time series plot.
- `plot(crime_ts)`



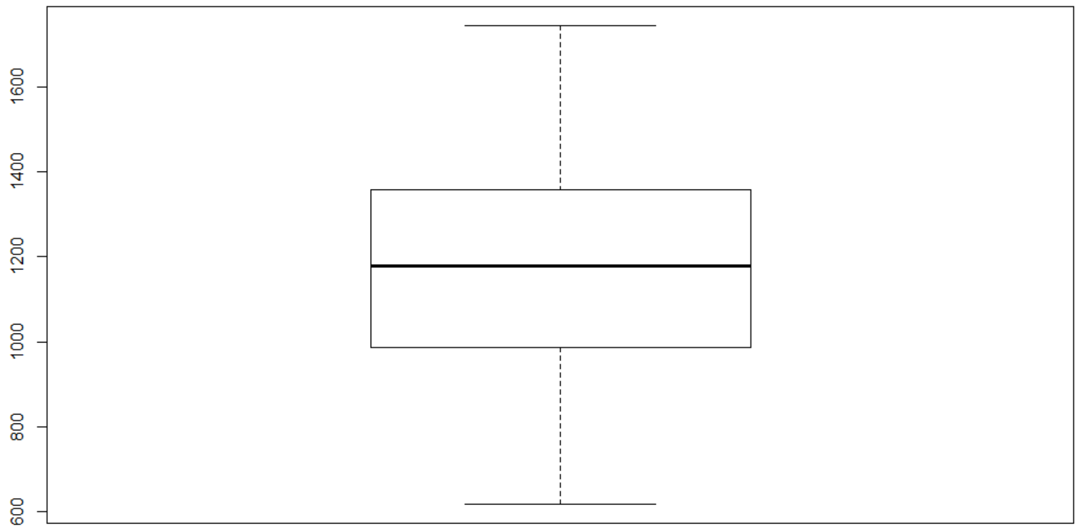
- Please summaries your observations of the times series plot
  - Overall the trend is downward, i.e. the crime rate is decreasing year by year.
  - We have a cyclical trend with regular upward and downward peaks.
  - Seasonality is observed as well. Crimes are high in the month of June-July and lowest in the month of Jan-Feb. From Jan to June there is rising trend and July to Dec is falling trend.

Central Tendency

- What are the min, max, mean, median, 1st and 3rd Quartile values of the times series?

```
> summary(crime_ts)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  619.0   991.5  1179.5  1179.1  1355.2  1743.0
```

- Show the box plot.
  boxplot(crime_ts)

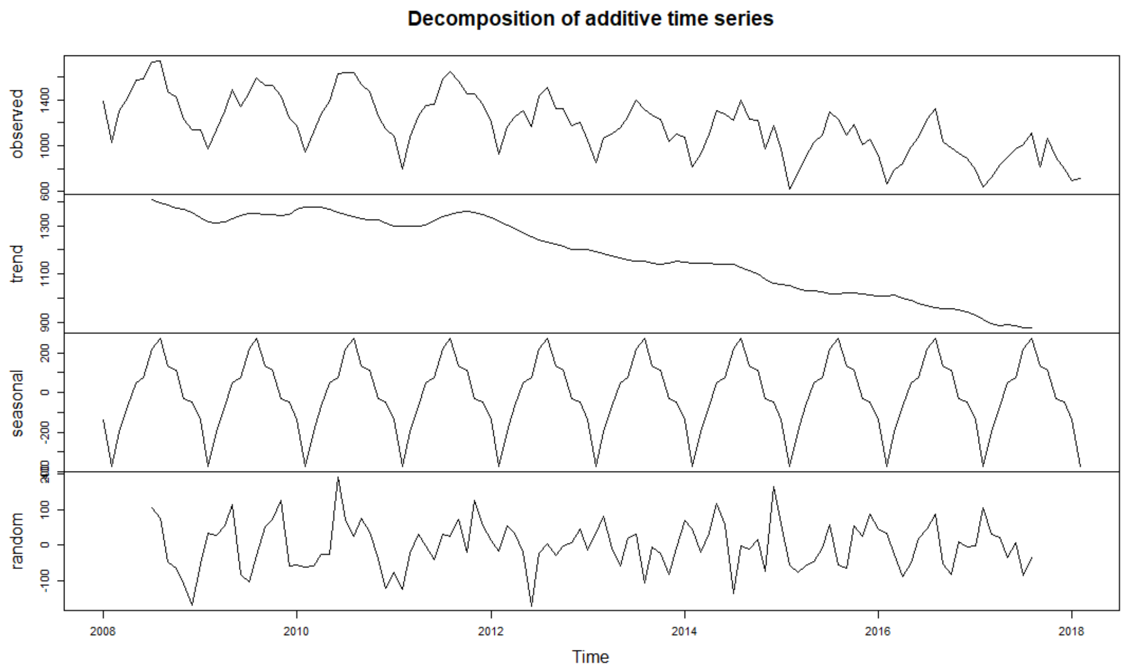


- Can you summarize your observation about the time series from the summary stats and box plot?
    - The mean and median are almost same, 1179.1 and 1179.5 respectively.
    - Min is 619 and max is 1743
    - First and third quartile value are 991.5 and 1355.2 respectively.

Decomposition

- Plot the decomposition of the time series.

```
decomp<-decompose(crime_ts)
plot(decomp)
```



- Is the times series seasonal?
    - Yes, it is seasonal

- Is the decomposition additive or multiplicative?
    - It is additive

```
> decomp$type
[1] "additive"
```

- If seasonal, what are the values of the seasonal monthly indices?

```
> decomp$seasonal
           Jan        Feb        Mar        Apr        May        Jun        Jul
2008 -140.35579 -374.62894 -198.86505  -69.94375   48.32477   77.65810  216.51968
2009 -140.35579 -374.62894 -198.86505  -69.94375   48.32477   77.65810  216.51968
2010 -140.35579 -374.62894 -198.86505  -69.94375   48.32477   77.65810  216.51968
2011 -140.35579 -374.62894 -198.86505  -69.94375   48.32477   77.65810  216.51968
2012 -140.35579 -374.62894 -198.86505  -69.94375   48.32477   77.65810  216.51968
2013 -140.35579 -374.62894 -198.86505  -69.94375   48.32477   77.65810  216.51968
2014 -140.35579 -374.62894 -198.86505  -69.94375   48.32477   77.65810  216.51968
2015 -140.35579 -374.62894 -198.86505  -69.94375   48.32477   77.65810  216.51968
2016 -140.35579 -374.62894 -198.86505  -69.94375   48.32477   77.65810  216.51968
2017 -140.35579 -374.62894 -198.86505  -69.94375   48.32477   77.65810  216.51968
2018 -140.35579 -374.62894
           Aug        Sep        Oct        Nov        Dec
2008  272.14468  132.35255  111.50069  -29.23542  -45.47153
2009  272.14468  132.35255  111.50069  -29.23542  -45.47153
2010  272.14468  132.35255  111.50069  -29.23542  -45.47153
2011  272.14468  132.35255  111.50069  -29.23542  -45.47153
2012  272.14468  132.35255  111.50069  -29.23542  -45.47153
2013  272.14468  132.35255  111.50069  -29.23542  -45.47153
2014  272.14468  132.35255  111.50069  -29.23542  -45.47153
2015  272.14468  132.35255  111.50069  -29.23542  -45.47153
2016  272.14468  132.35255  111.50069  -29.23542  -45.47153
2017  272.14468  132.35255  111.50069  -29.23542  -45.47153
2018
> 
```

- For which month is the value of time series high and for which month is it low?

- Lowest in February and highest in August

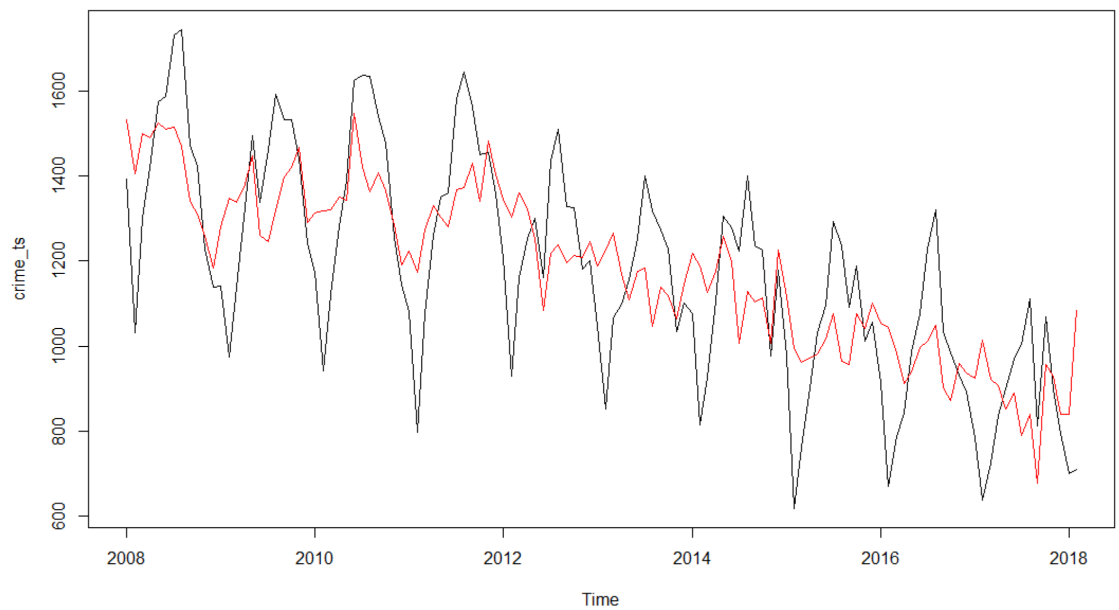
- Can you think of the reason behind the value being high in those months and low in those months?

  The weather is cold in February and hence people do not venture outdoors much. Also, the homeless are provided shelters by the government during this time.

  August is summer time, people socialize more and the homeless take shelter on the curbside, hence crimes are possibly high.

- Show the plot for time series adjusted for seasonality. Overlay this with the line for actual time series? Does seasonality have big fluctuations to the value of time series?
  - No Major fluctuations are observed

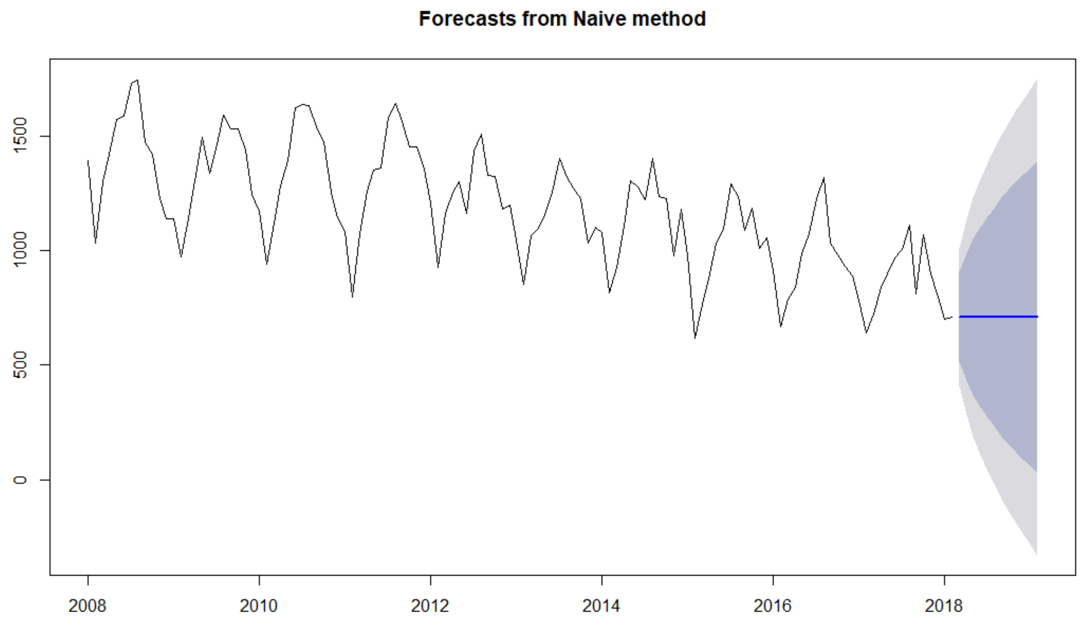
```
library(fpp)
temp_sesadjust<-seasadj(decomp)
plot(crime_ts)
lines(temp_sesadjust,col='red')
```



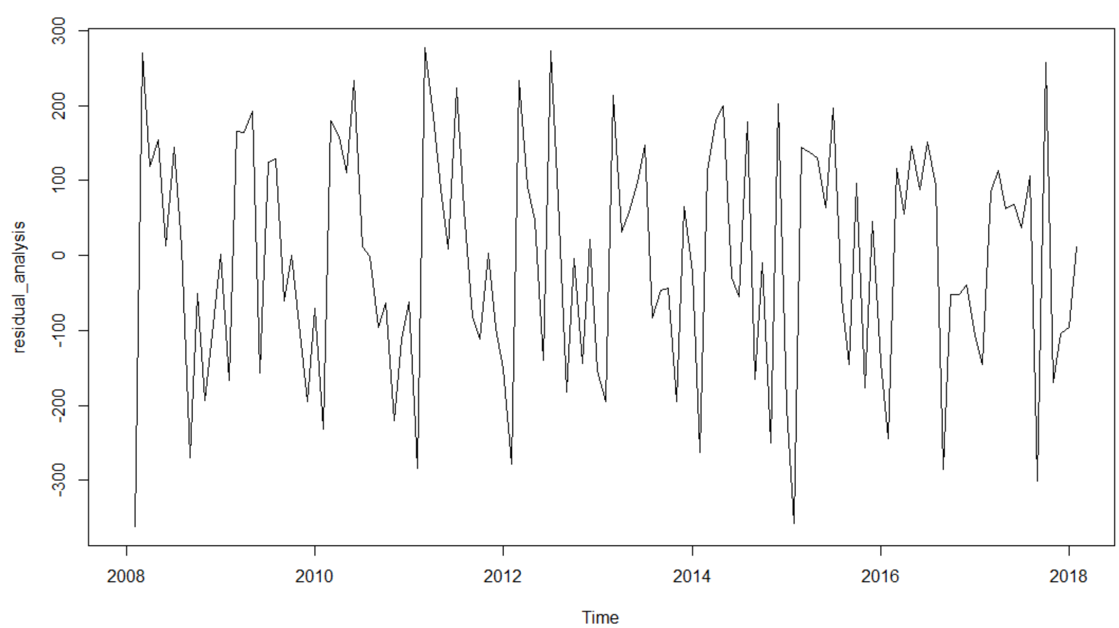
Naïve Method

- Output

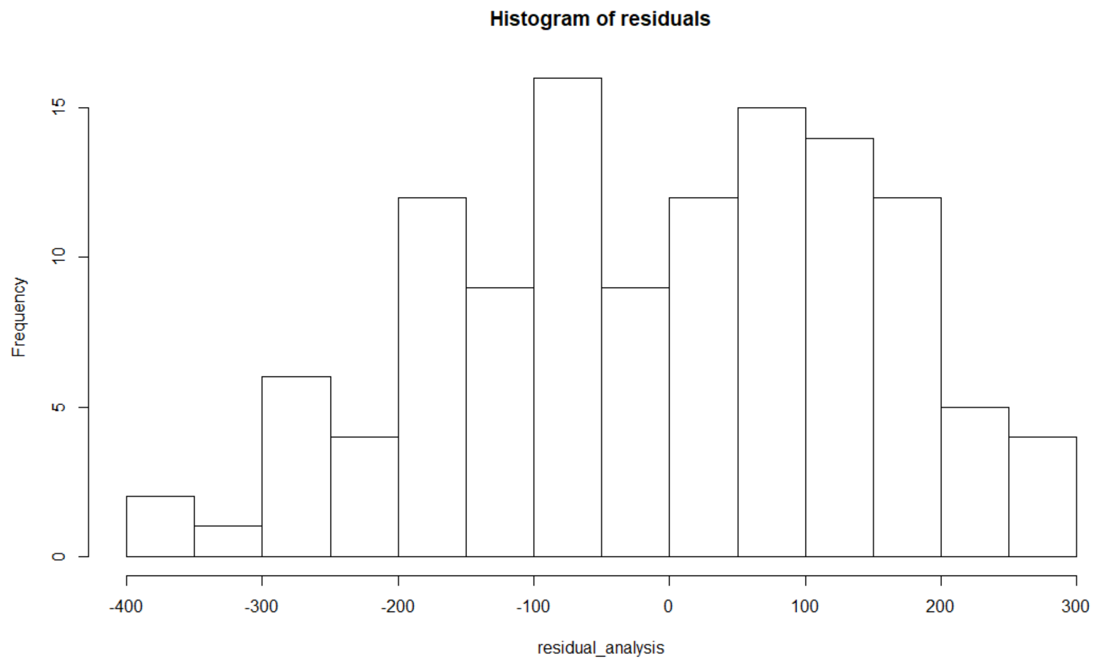
```
naive_forecast<-naive(crime_ts,12)
plot(naive_forecast)
```



- Perform Residual Analysis for this technique.
    - Do a plot of residuals. What does the plot indicate?
      The values fluctuate between -300 to 250. Values are significant when residuals are zero, which is not the case. The values are not very significant.



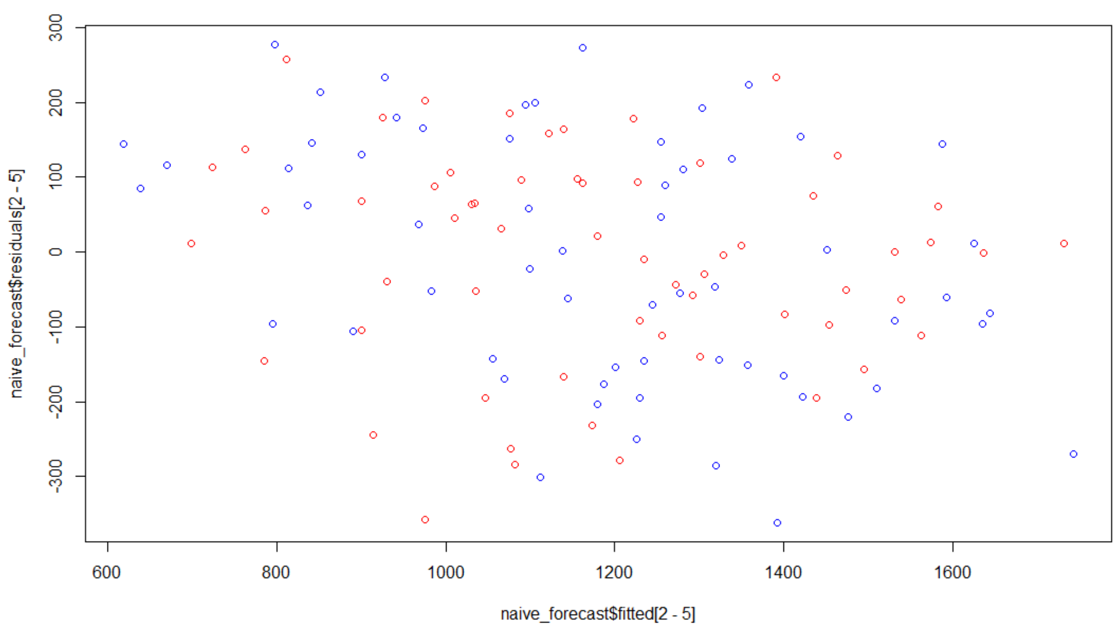
    - Do a Histogram plot of residuals. What does the plot indicate?
        - Residual values are left skewed.



- Do a plot of fitted values vs. residuals. What does the plot indicate?

  plot(naive_forecast$fitted[2-5],naive_forecast$residuals[2-5],col=c("red","blue"))

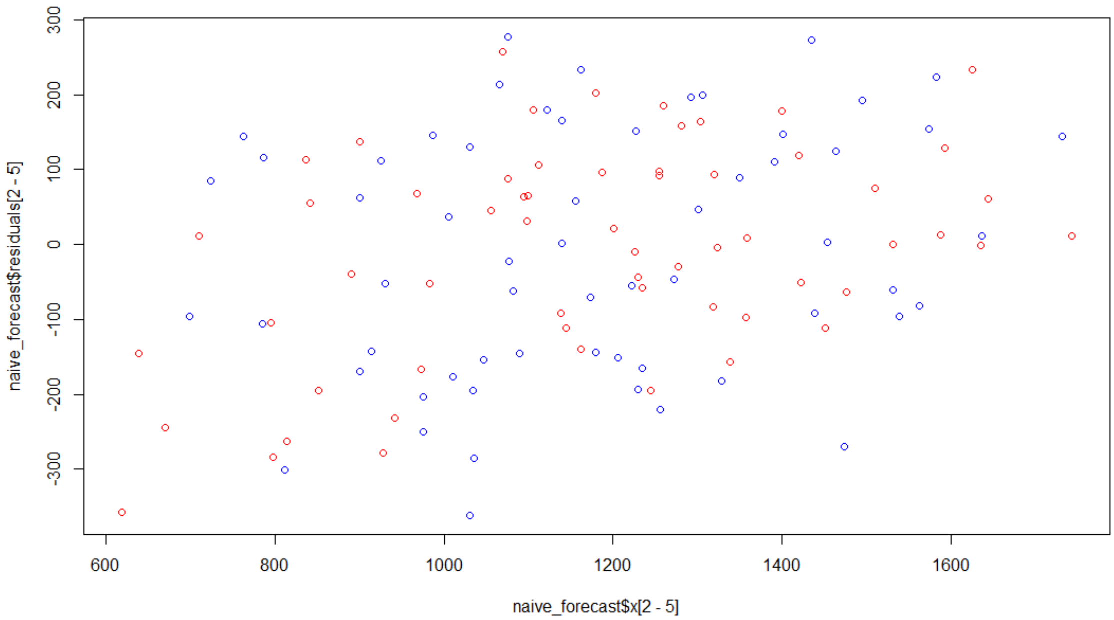
  The variation in values is high, hence interpretation is difficult.



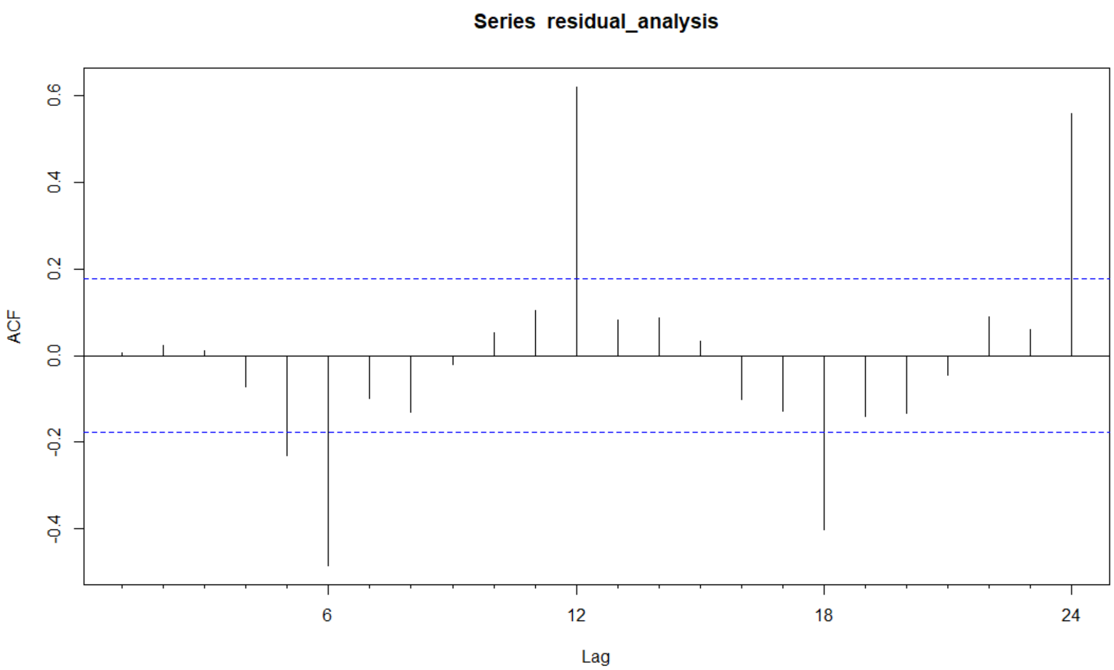
- Do a plot of actual values vs. residuals. What does the plot indicate?

plot(naive_forecast$x[2-5],naive_forecast$residuals[2-5],col=c("red","blue"))

The variation in values is high, hence interpretation is difficult.



- Do an ACF plot of the residuals? What does this plot indicate?
  Acf(residual_analysis)



We can say that very few values are significant.

- Print the 5 measures of accuracy for this forecasting technique

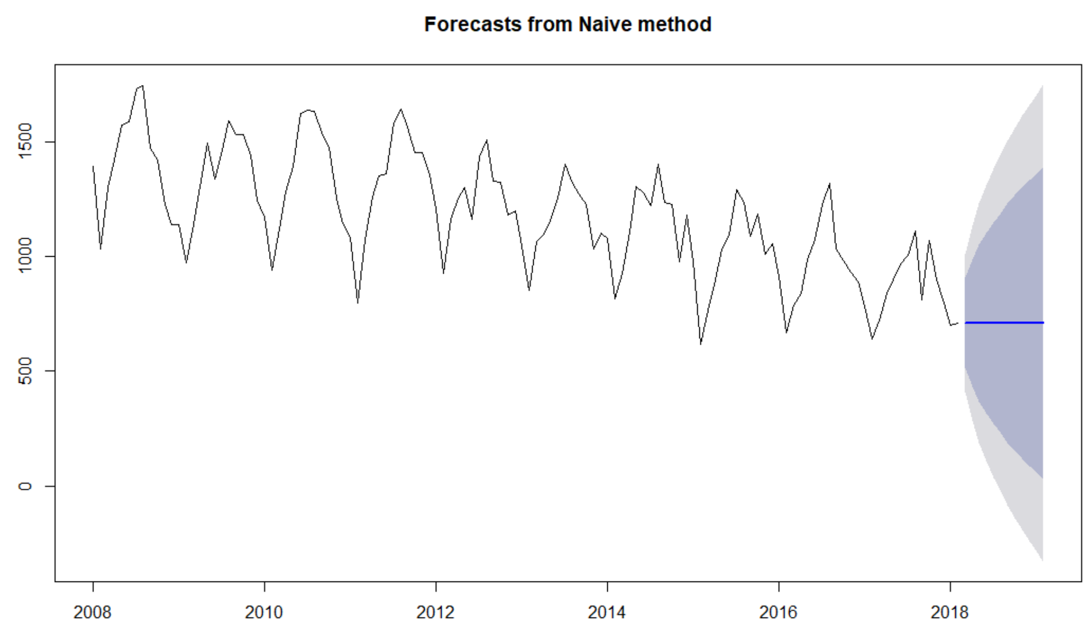
```
> accuracy(naive_forecast)
                    ME     RMSE      MAE       MPE     MAPE     MASE        ACF1
Training set -5.636364 153.4663 129.1074 -1.626318 12.00476 1.193129 0.006097097
```

- Forecast
  - Time series value for next year. Show table and plot

    naive_fore<- forecast(naive_forecast, h=12)

```
> naive_fore
         Point Forecast     Lo 80     Hi 80      Lo 95    Hi 95
Mar 2018            710 513.32491  906.6751  409.21140 1010.789
Apr 2018            710 431.85941  988.1406  284.62068 1135.379
May 2018            710 369.34874 1050.6513  189.01886 1230.981
Jun 2018            710 316.64981 1103.3502  108.42280 1311.577
Jul 2018            710 270.22112 1149.7789   37.41624 1382.584
Aug 2018            710 228.24637 1191.7536  -26.77859 1446.779
Sep 2018            710 189.64661 1230.3534  -85.81183 1505.812
Oct 2018            710 153.71883 1266.2812 -140.75864 1560.759
Nov 2018            710 119.97472 1300.0253 -192.36580 1612.366
Dec 2018            710  88.05874 1331.9413 -241.17707 1661.177
Jan 2019            710  57.70251 1362.2975 -287.60293 1707.603
Feb 2019            710  28.69749 1391.3025 -331.96228 1751.962
```

plot(naive_fore)



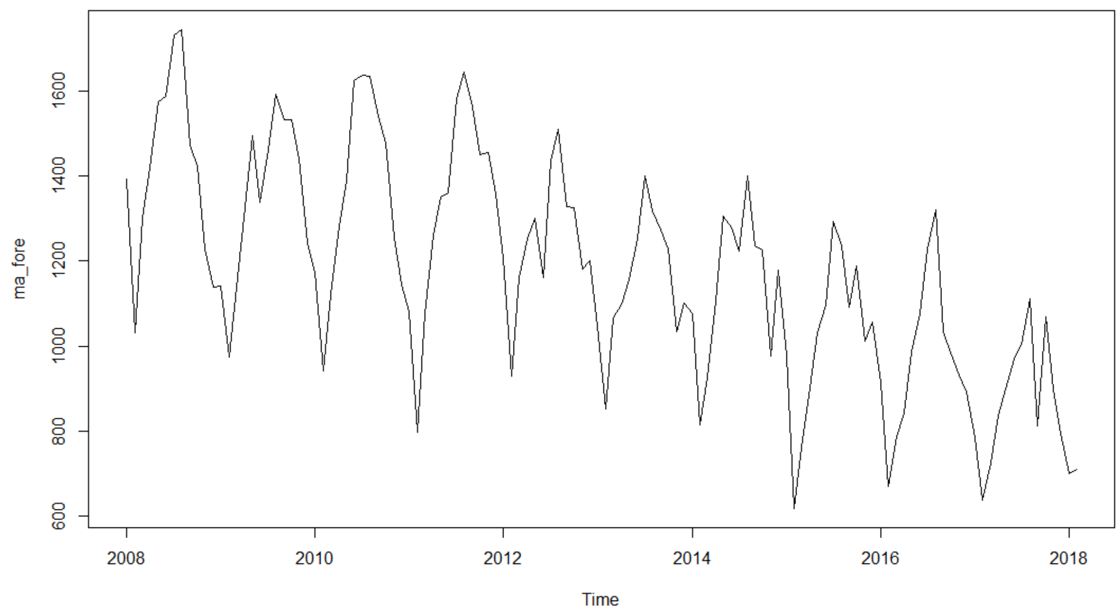
- Summarize this forecasting technique
  - How good is the accuracy?
    - The MAPE is 12.004, which is good. Hence there is 12% difference in actual and predicted values. Accuracy is good.

- What does it predict the value of time series will be in one year?
  The predicted value is 710

- Other observation
  - No such important observation

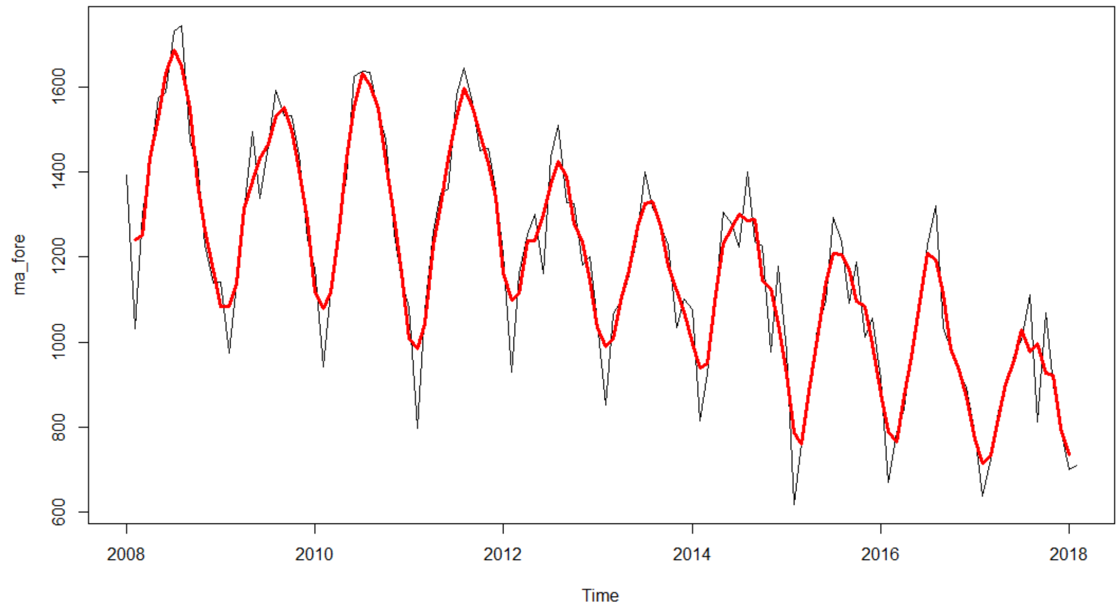
Simple Moving Averages

- Plot the graph for time series.
  ma_fore<-ma(crime_ts, order=1)
  plot(ma_fore)

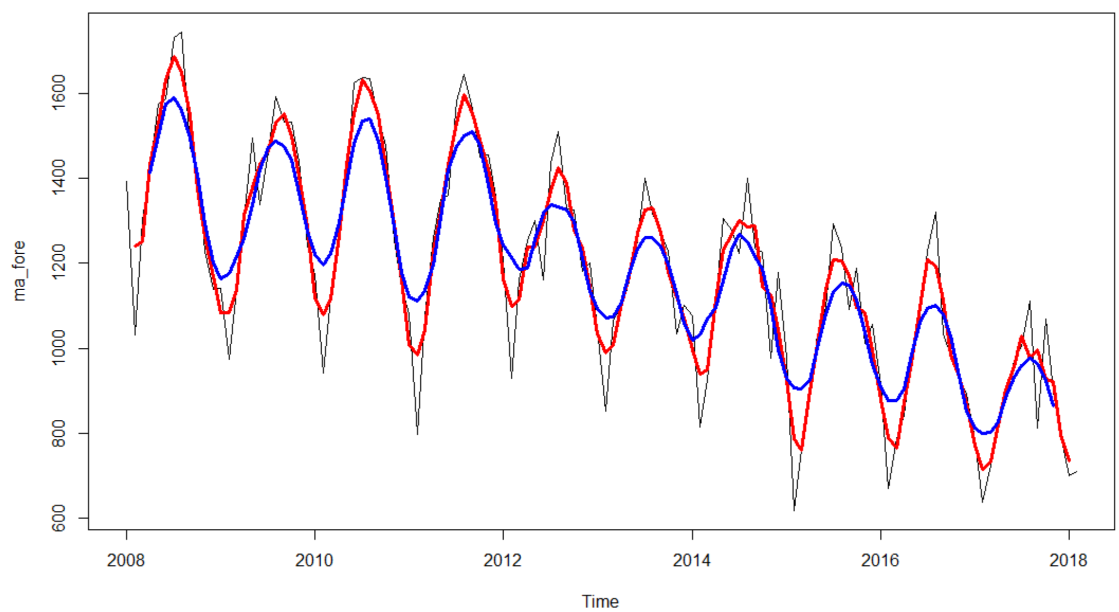


- Show the Simple Moving average of order 3 on the plot above in Red

  ma3_fore<-ma(crime_ts, order=3)
  lines(ma3_fore,col="red",lwd=3)

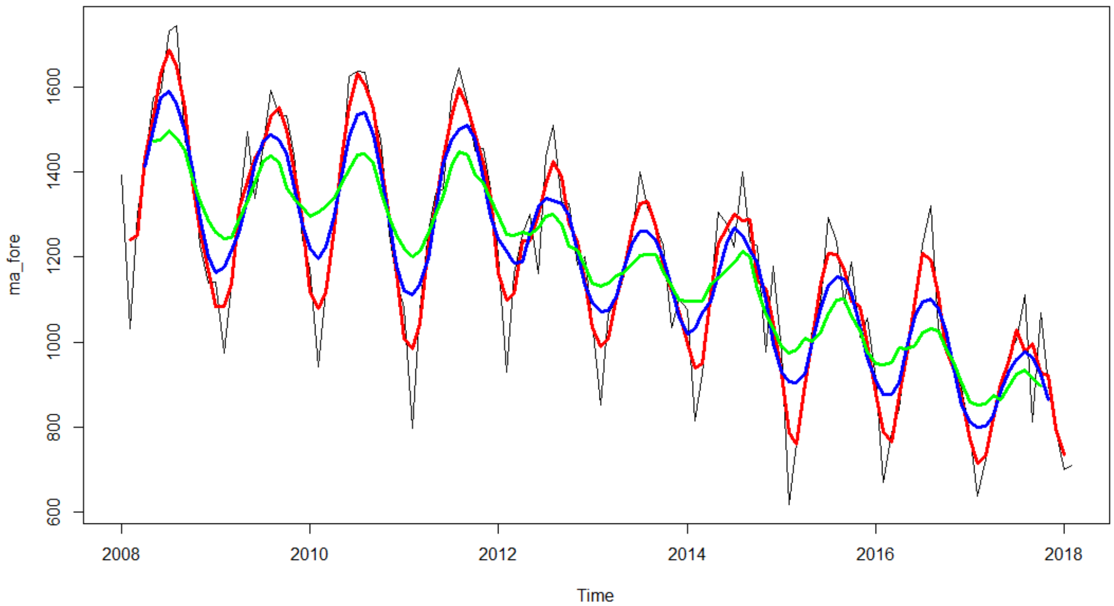


- Show the Simple Moving average of order 6 on the plot above in Blue
  ma6_fore<-ma(crime_ts, order=6)
  lines(ma6_fore,col="blue",lwd=3)



- Show the Simple Moving average of order 9 on the plot above in Green

  ma9_fore<-ma(crime_ts, order=9)
  lines(ma9_fore,col="green",lwd=3)



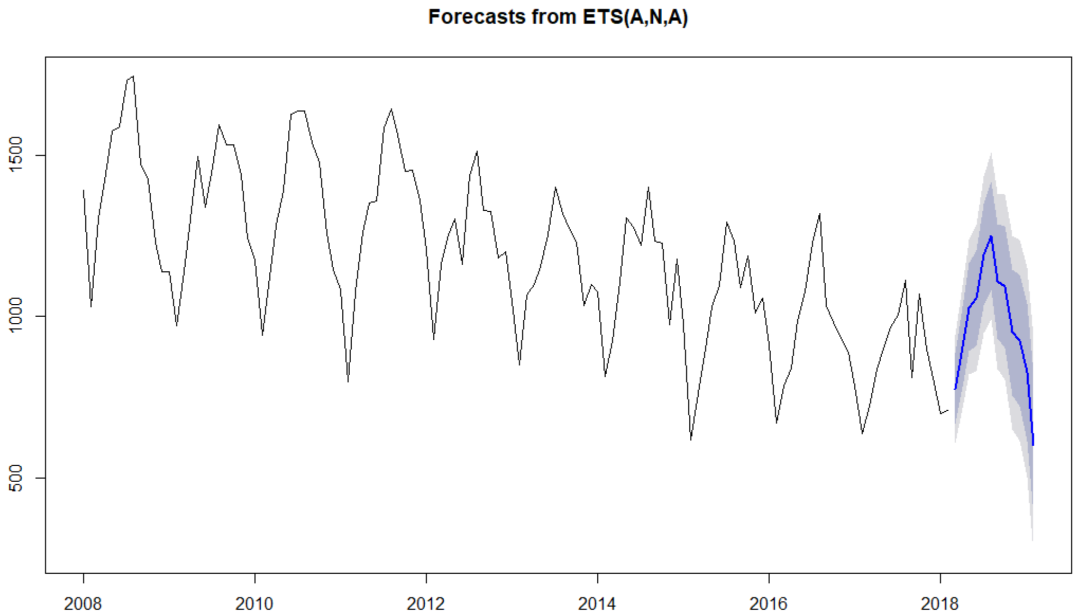
- (Bonus) show the forecast of next 12 months using one of the simple average order that you feel works best for time series

  USING ETS:

  ets_forecast <- ets(crime_ts)
  fore_ets<-forecast.ets(ets_forecast, h=12)
  fore_ets

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Mar 2018	776.4469	667.0042	885.8895	609.0687	943.8250
Apr 2018	906.9292	783.4354	1030.4230	718.0618	1095.7967
May 2018	1028.3387	892.2367	1164.4406	820.1887	1236.4886
Jun 2018	1057.1163	909.4791	1204.7536	831.3247	1282.9080
Jul 2018	1190.9695	1032.6351	1349.3039	948.8180	1433.1210
Aug 2018	1249.7781	1081.4249	1418.1312	992.3041	1507.2520
Sep 2018	1108.7453	930.9370	1286.5537	836.8109	1380.6798
Oct 2018	1091.2937	904.5082	1278.0792	805.6299	1376.9575
Nov 2018	950.9327	755.5821	1146.2833	652.1698	1249.6956
Dec 2018	926.0483	722.4928	1129.6039	614.7370	1237.3597
Jan 2019	827.6569	616.2145	1039.0992	504.2836	1151.0301
Feb 2019	601.5649	382.5165	820.6133	266.5593	936.5705

plot(fore_ets)
-



USING: Simple moving averages (forecast for accuracy table (extra))
ma_forecast<-forecast(ma_fore, h=12)
ma_forecast
accuracy(ma_forecast)

```
         Point Forecast       Lo 80     Hi 80    Lo 95     Hi 95
Mar 2018       776.4469  667.0042  885.8895 609.0687  943.8250
Apr 2018       906.9292  783.4354 1030.4230 718.0618 1095.7967
May 2018      1028.3387  892.2367 1164.4406 820.1887 1236.4886
Jun 2018      1057.1163  909.4791 1204.7536 831.3247 1282.9080
Jul 2018      1190.9695 1032.6351 1349.3039 948.8180 1433.1210
Aug 2018      1249.7781 1081.4249 1418.1312 992.3041 1507.2520
Sep 2018      1108.7453  930.9370 1286.5537 836.8109 1380.6798
Oct 2018      1091.2937  904.5082 1278.0792 805.6299 1376.9575
Nov 2018       950.9327  755.5821 1146.2833 652.1698 1249.6956
Dec 2018       926.0483  722.4928 1129.6039 614.7370 1237.3597
Jan 2019       827.6569  616.2145 1039.0992 504.2836 1151.0301
Feb 2019       601.5649  382.5165  820.6133 266.5593  936.5705
> accuracy(ma_forecast)
                    ME     RMSE     MAE        MPE     MAPE      MASE       ACF1
Training set -6.398486 80.34939 62.6294 -0.8653392 5.664747 0.5787813 0.05803802
>
```

- What are your observations of the plot as the moving average order goes up?
    - Error of prediction increases with it.

Smoothing

- Perform a smoothing forecast for next 12 months for the time series.
  ses(crime_ts, h=12)

```
> ses(crime_ts, h=12)
         Point Forecast     Lo 80     Hi 80      Lo 95    Hi 95
Mar 2018       709.9989 512.50595  907.4918  407.95950 1012.038
Apr 2018       709.9989 430.71569  989.2821  282.87209 1137.126
May 2018       709.9989 367.95393 1052.0439  186.88627 1233.112
Jun 2018       709.9989 315.04269 1104.9551  105.96551 1314.032
Jul 2018       709.9989 268.42665 1151.5711   34.67242 1385.325
Aug 2018       709.9989 226.28235 1193.7154  -29.78172 1449.780
Sep 2018       709.9989 187.52656 1232.4712  -89.05359 1509.051
Oct 2018       709.9989 151.45347 1268.5443 -144.22261 1564.220
Nov 2018       709.9989 117.57284 1302.4250 -196.03857 1616.036
Dec 2018       709.9989  85.52770 1334.4701 -245.04739 1665.045
Jan 2019       709.9989  55.04857 1364.9492 -291.66119 1711.659
Feb 2019       709.9989  25.92611 1394.0717 -336.20014 1756.198
> |
```

- What is the value of alpha? What does that value signify?
  summary(ses(crime_ts, h=12))

  The value of alpha is 0.99. It signifies the optimal smoothing parameter for the model to get minimum error

```
> summary(ses(crime_ts, h=12))

Forecast method: Simple exponential smoothing

Model Information:
Simple exponential smoothing

Call:
 ses(y = crime_ts, h = 12)

  Smoothing parameters:
    alpha = 0.9999

  Initial states:
    l = 1391.4554

  sigma:  154.1046
```

```
     AIC     AICc      BIC
1819.256 1819.459 1827.668

Error measures:
                    ME     RMSE      MAE       MPE     MAPE    MASE        ACF1
Training set -5.586269 152.8362 128.0557 -1.612822 11.90683 1.18341 0.005428101

Forecasts:
```


```
         Point Forecast     Lo 80     Hi 80      Lo 95    Hi 95
Mar 2018       709.9989 512.50595  907.4918  407.95950 1012.038
Apr 2018       709.9989 430.71569  989.2821  282.87209 1137.126
May 2018       709.9989 367.95393 1052.0439  186.88627 1233.112
Jun 2018       709.9989 315.04269 1104.9551  105.96551 1314.032
Jul 2018       709.9989 268.42665 1151.5711   34.67242 1385.325
Aug 2018       709.9989 226.28235 1193.7154  -29.78172 1449.780
Sep 2018       709.9989 187.52656 1232.4712  -89.05359 1509.051
Oct 2018       709.9989 151.45347 1268.5443 -144.22261 1564.220
Nov 2018       709.9989 117.57284 1302.4250 -196.03857 1616.036
Dec 2018       709.9989  85.52770 1334.4701 -245.04739 1665.045
Jan 2019       709.9989  55.04857 1364.9492 -291.66119 1711.659
Feb 2019       709.9989  25.92611 1394.0717 -336.20014 1756.198
```

- What is the value of initial state?
  The value of initial state is 1391.4554

- What is the value of sigma? What does the sigma signify?
  The value of sigma is 154.1056. It signifies that the residuals have more variation around the residual mean.

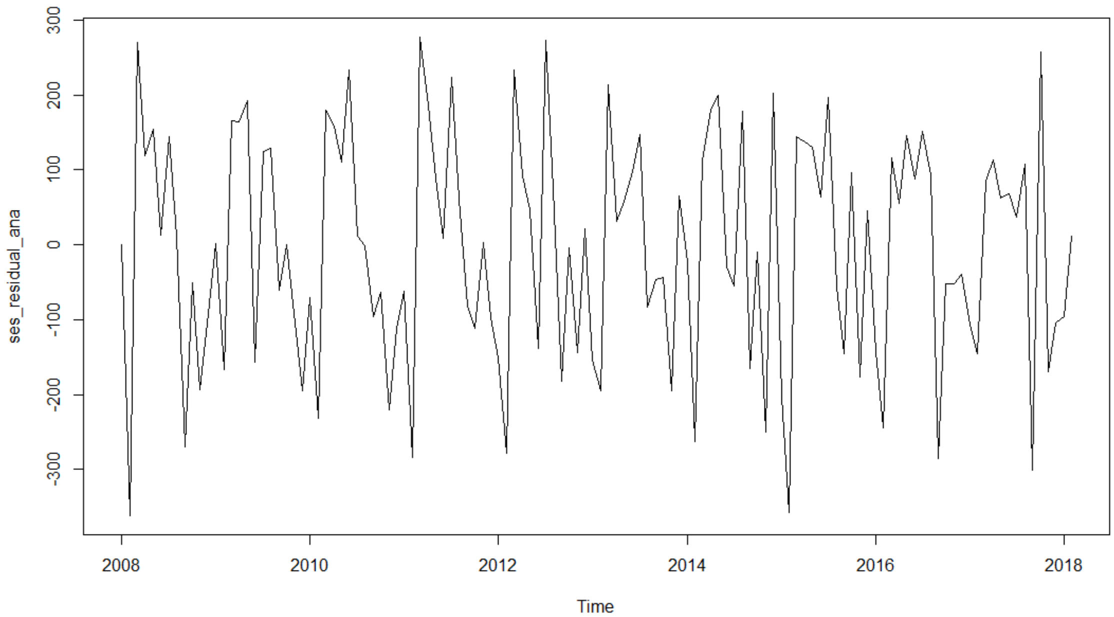
- Perform Residual Analysis for this technique.

  ses_fore <- ses(crime_ts, h=12)
  ses_residual_ana <- residuals(ses_fore)
  ses_residual_ana

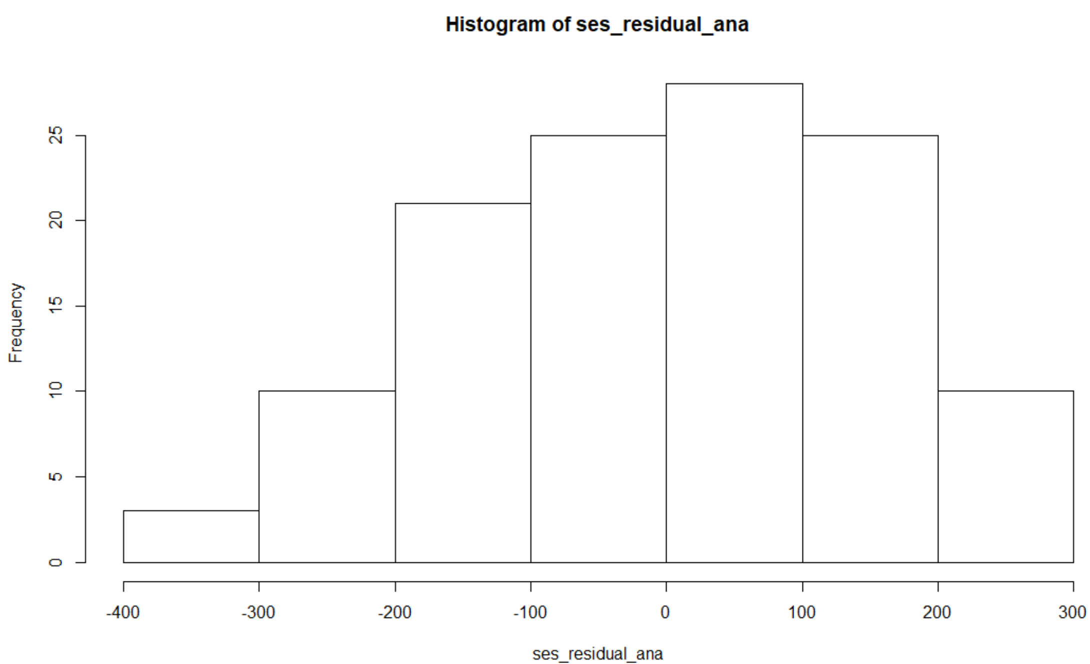
```
> ses_fore <- ses(crime_ts, h=12)
> ses_residual_ana <- residuals(ses_fore)
> ses_residual_ana
               Jan           Feb           Mar           Apr           May
2008  5.446450e-01 -3.609999e+02  2.699638e+02  1.190271e+02  1.540119e+02
2009  1.990876e+00 -1.669998e+02  1.659833e+02  1.640166e+02  1.920164e+02
2010 -7.101955e+01 -2.310071e+02  1.799768e+02  1.590180e+02  1.100159e+02
2011 -6.201123e+01 -2.840062e+02  2.769715e+02  1.850278e+02  9.001855e+01
2012 -1.510097e+02 -2.780151e+02  2.339721e+02  9.202345e+01  4.700922e+01
2013 -1.539979e+02 -1.950154e+02  2.139805e+02  3.202145e+01  5.800321e+01
2014 -2.299339e+01 -2.630023e+02  1.119736e+02  1.800112e+02  2.000180e+02
2015 -2.029797e+02 -3.570203e+02  1.439642e+02  1.380144e+02  1.300138e+02
2016 -1.419955e+02 -2.440142e+02  1.159755e+02  5.501163e+01  1.460055e+02
2017 -1.060040e+02 -1.460106e+02  8.498536e+01  1.130085e+02  6.301133e+01
2018 -9.601053e+01  1.099038e+01
               Jun           Jul           Aug           Sep           Oct
2008  1.301544e+01  1.440013e+02  1.201443e+01 -2.699988e+02 -5.102707e+01
2009 -1.569808e+02  1.249843e+02  1.290125e+02 -6.098707e+01 -6.113418e-03
2010  2.340110e+02  1.102346e+01 -9.988950e-01 -9.600010e+01 -6.300962e+01
2011  9.009024e+00  2.240009e+02  6.102245e+01 -8.199388e+01 -1.110082e+02
2012 -1.389953e+02  2.729861e+02  7.502736e+01 -1.819925e+02 -4.018243e+00
2013  9.800581e+01  1.470098e+02 -8.298526e+01 -4.600832e+01 -4.300461e+01
2014 -2.897995e+01 -5.500290e+01  1.779945e+02 -1.649822e+02 -9.016538e+00
2015  6.401303e+01  1.970064e+02 -5.698025e+01 -1.450057e+02  9.698546e+01
2016  8.801464e+01  1.520088e+02  9.301524e+01 -2.849907e+02 -5.202857e+01
2017  6.800632e+01  3.700682e+01  1.070037e+02 -3.009893e+02  2.579698e+02
2018
               Nov           Dec
2008 -1.930051e+02 -9.101935e+01
2009 -9.200000e+01 -1.950092e+02
2010 -2.200063e+02 -1.120221e+02
2011  2.988872e+00 -9.699970e+01
2012 -1.440004e+02  2.098557e+01
2013 -1.950043e+02  6.598045e+01
2014 -2.500009e+02  2.029749e+02
2015 -1.759903e+02  4.498236e+01
2016 -5.200522e+01 -4.000521e+01
2017 -1.689741e+02 -1.050169e+02
2018
> plot(ses_residual_ana)
```

  - Do a plot of residuals. What does the plot indicate?

    plot(ses_residual_ana)



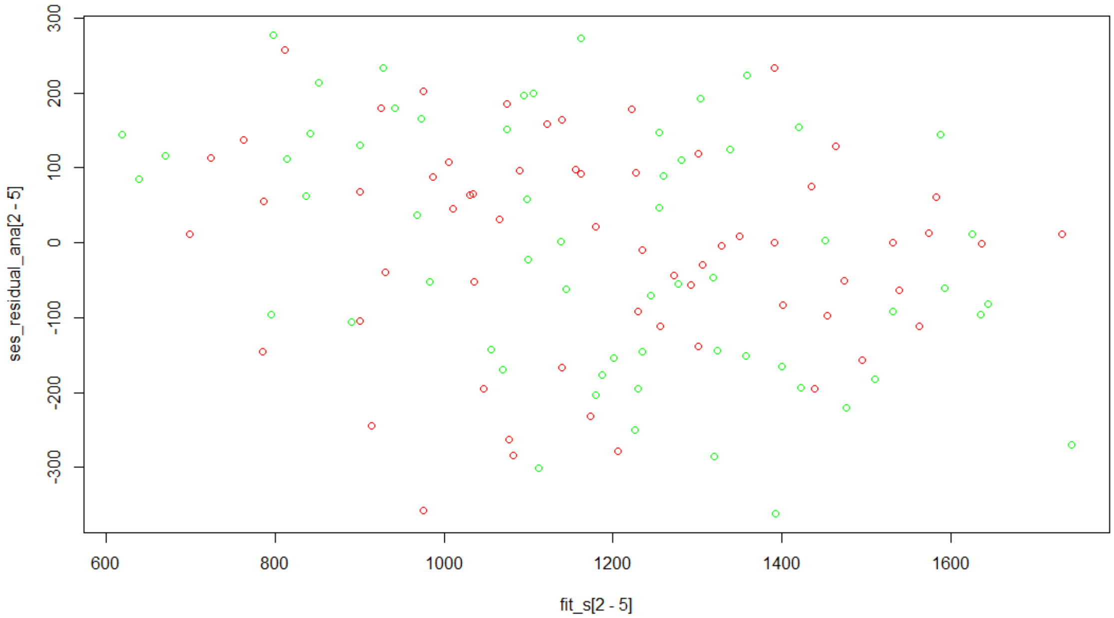
- Do a Histogram plot of residuals. What does the plot indicate?
  - From the plot we can see that data is left skewed.



- Do a plot of fitted values vs. residuals. What does the plot indicate?

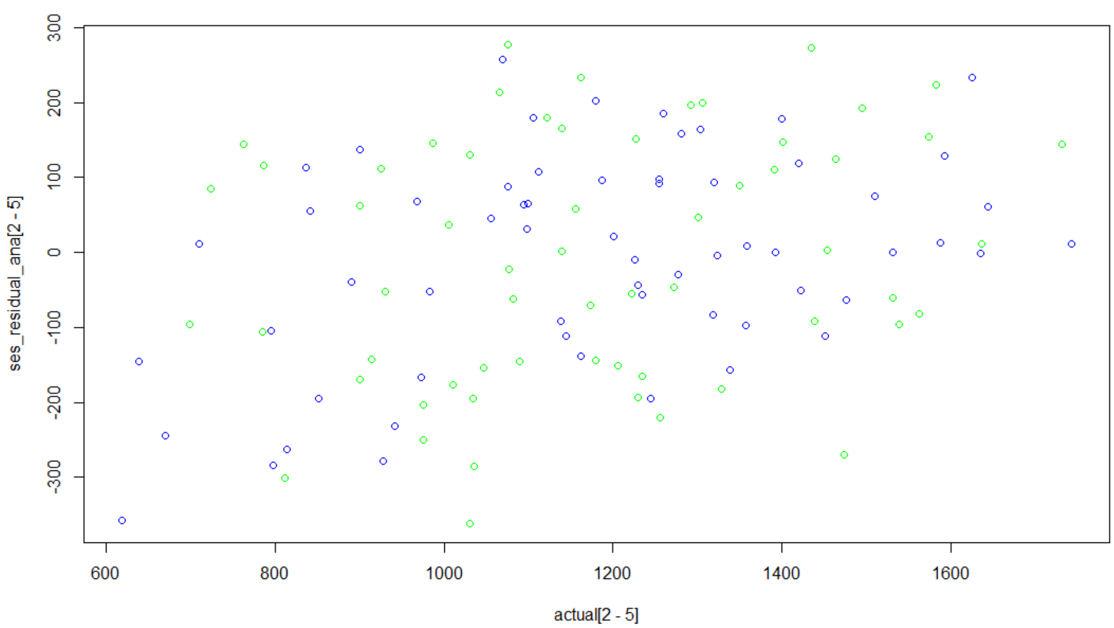
  We observe cyclic and seasonal pattern in both, fitted and residual values.

  fit_s<-ses_fore$fitted
  plot(fit_s[2-5],ses_residual_ana[2-5], col=c("Red","Green"))

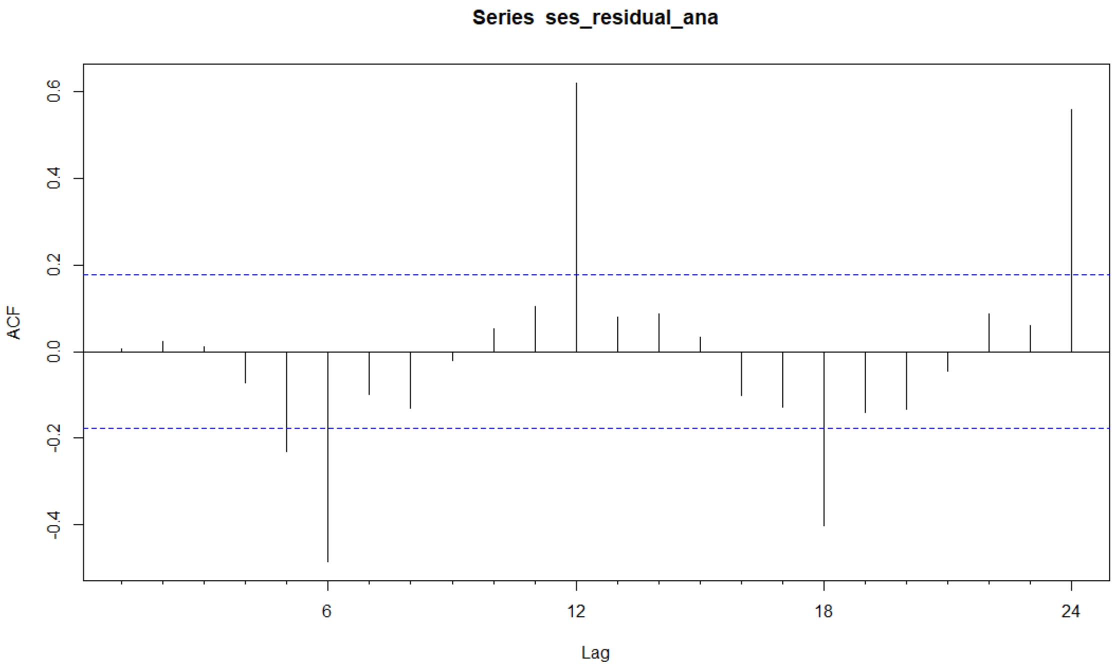


- Do a plot of actual values vs. residuals. What does the plot indicate?

  We observe cyclic and seasonal pattern in both, fitted and residual values.
  actual<-ses_fore$x
  plot(actual[2-5],ses_residual_ana[2-5],col=c("Blue","Green"))



- Do an ACF plot of the residuals? What does this plot indicate?
  Acf(ses_residual_ana)
  We can say that only few values are significant, at 6, 12, 18, 24

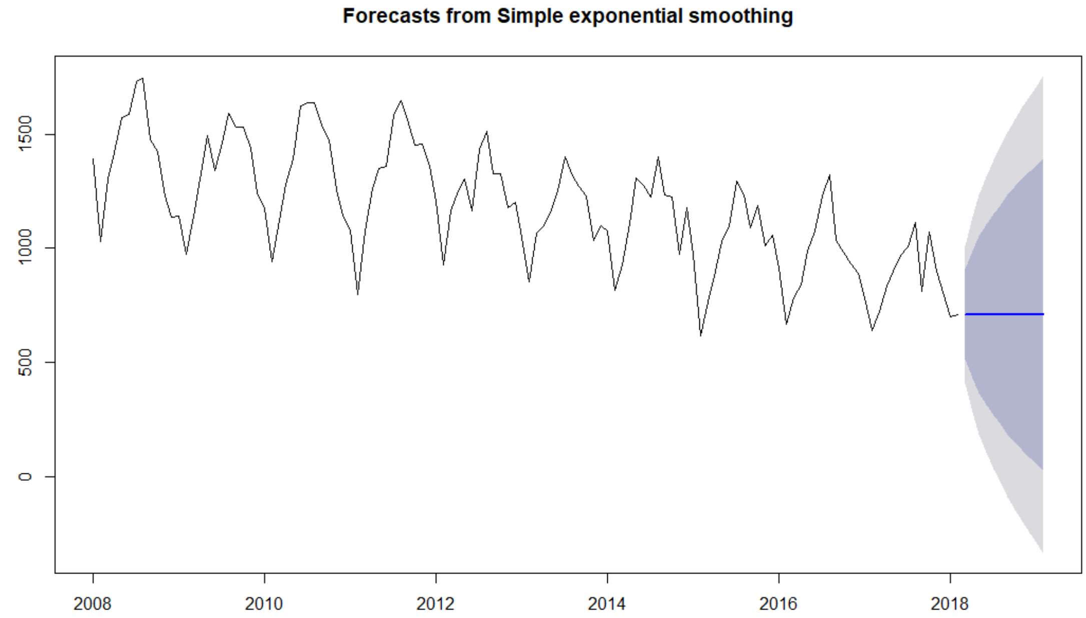


- Print the 5 measures of accuracy for this forecasting technique

```
> Acf(ses_residual_ana)
> accuracy(ses_fore)
                    ME     RMSE      MAE       MPE     MAPE    MASE        ACF1
Training set -5.586269 152.8362 128.0557 -1.612822 11.90683 1.18341 0.005428101
>
```

- Forecast
  - Time series value for next year. Show table and plot

```
> forecast(ses_fore, h=12)
         Point Forecast       Lo 80     Hi 80      Lo 95    Hi 95
Mar 2018       709.9989 512.50595  907.4918  407.95950 1012.038
Apr 2018       709.9989 430.71569  989.2821  282.87209 1137.126
May 2018       709.9989 367.95393 1052.0439  186.88627 1233.112
Jun 2018       709.9989 315.04269 1104.9551  105.96551 1314.032
Jul 2018       709.9989 268.42665 1151.5711   34.67242 1385.325
Aug 2018       709.9989 226.28235 1193.7154  -29.78172 1449.780
Sep 2018       709.9989 187.52656 1232.4712  -89.05359 1509.051
Oct 2018       709.9989 151.45347 1268.5443 -144.22261 1564.220
Nov 2018       709.9989 117.57284 1302.4250 -196.03857 1616.036
Dec 2018       709.9989  85.52770 1334.4701 -245.04739 1665.045
Jan 2019       709.9989  55.04857 1364.9492 -291.66119 1711.659
Feb 2019       709.9989  25.92611 1394.0717 -336.20014 1756.198
```



- Summarize this forecasting technique

```
> summary(ses_fore)

Forecast method: Simple exponential smoothing

Model Information:
Simple exponential smoothing

Call:
 ses(y = crime_ts, h = 12)

  Smoothing parameters:
    alpha = 0.9999

  Initial states:
    l = 1391.4554

  sigma:  154.1046

     AIC     AICc      BIC
1819.256 1819.459 1827.668

Error measures:
                    ME     RMSE      MAE       MPE     MAPE    MASE        ACF1
Training set -5.586269 152.8362 128.0557 -1.612822 11.90683 1.18341 0.005428101
```

  - How good is the accuracy?
    MAPE is 11.9 so accuracy is good
  - What does it predict the value of time series will be in one year?
    The predicted value in 1 year is 709.99

- Other observation
  MPE, MAPE, MASE values show good accuracy, but RMSE, MAE show poor accuracy measures.

Holt-Winters

- Perform Holt-Winters forecast for next 12 months for the time series.

```
holt<-HoltWinters(crime_ts)
holt_forecast<-forecast(holt, h=12)
holt_forecast
plot(holt_forecast)
```

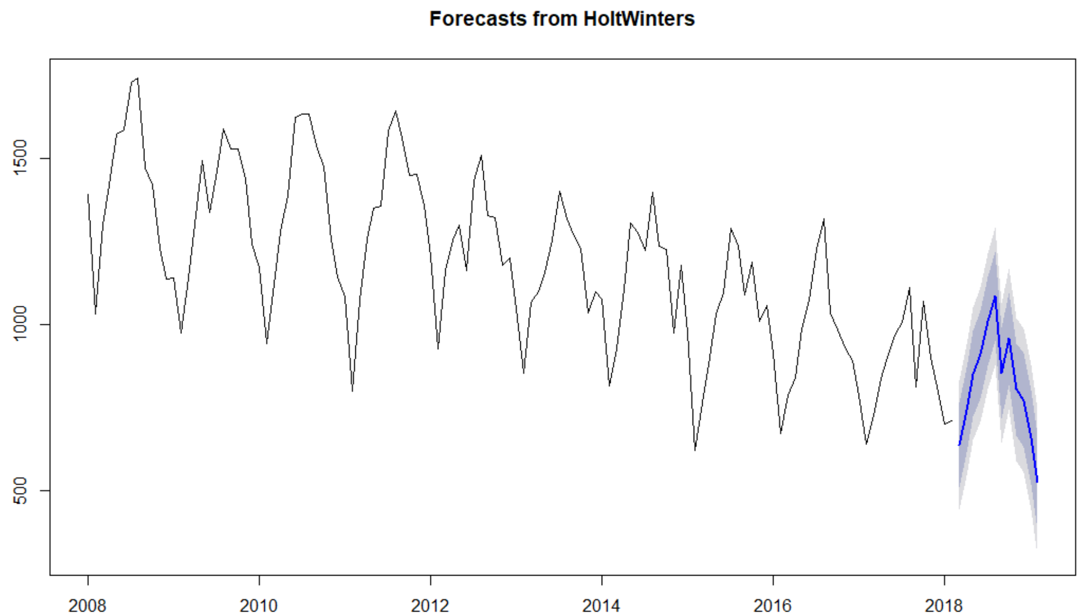
```
Holt-Winters exponential smoothing with trend and additive seasonal component.

Call:
HoltWinters(x = crime_ts)

Smoothing parameters:
 alpha: 0.1583224
 beta : 0.005786356
 gamma: 0.4467106

Coefficients:
           [,1]
a    818.848314
b     -6.780014
s1  -177.049490
s2   -63.916884
s3    52.668209
s4   117.920159
s5   225.125281
s6   307.869360
s7    81.796516
s8   192.420461
s9    45.961763
s10   20.161964
s11  -84.438994
s12 -213.571013
```

```
> holt_forecast
         Point Forecast    Lo 80     Hi 80    Lo 95     Hi 95
Mar 2018       635.0188 507.7725  762.2651 440.4124  829.6252
Apr 2018       741.3714 612.5219  870.2209 544.3132  938.4296
May 2018       851.1765 720.7254  981.6276 651.6687 1050.6842
Jun 2018       909.6484 777.5971 1041.6998 707.6933 1111.6035
Jul 2018      1010.0735 876.4231 1143.7239 805.6729 1214.4741
Aug 2018      1086.0376 950.7893 1221.2859 879.1931 1292.8820
Sep 2018       853.1847 716.3394  990.0300 643.8979 1062.4715
Oct 2018       957.0287 818.5872 1095.4701 745.3007 1168.7566
Nov 2018       803.7899 663.7530  943.8269 589.6220 1017.9579
Dec 2018       771.2101 629.5783  912.8420 554.6030  987.8173
Jan 2019       659.8292 516.6029  803.0554 440.7835  878.8748
Feb 2019       523.9171 379.0968  668.7375 302.4336  745.4007
```



- What is the value of alpha? What does that value signify?
  The value of alpha is 0.1583. It signifies that the predictions are stable and random variations are smoothed.

- What is the value of beta? What does that value signify?
  The value of beta is 0.0057. it signifies that trend completely depends on the previous period value.

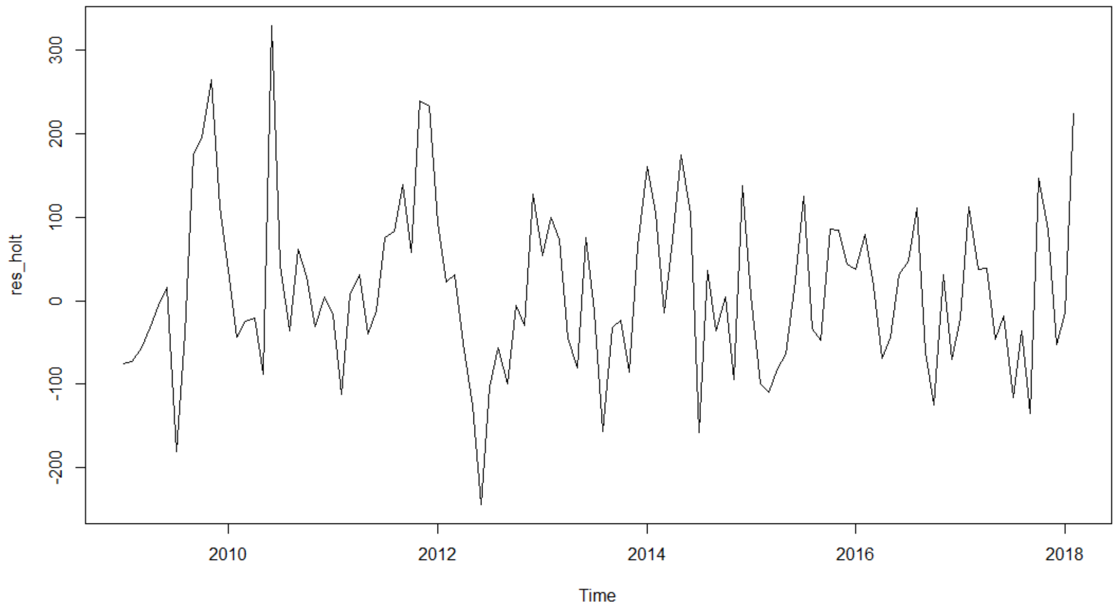
- What is the value of gamma? What does that value signify?
  The value of gamma is 0.4467. It signifies that the seasonality repeats according to cycles at regular time intervals

- What is the value of initial states for the level, trend and seasonality? What do these values signify?
  From the above:
  a is Level,
  b is Trend ,
  s1 to s12 are the initial seasonality values for each month

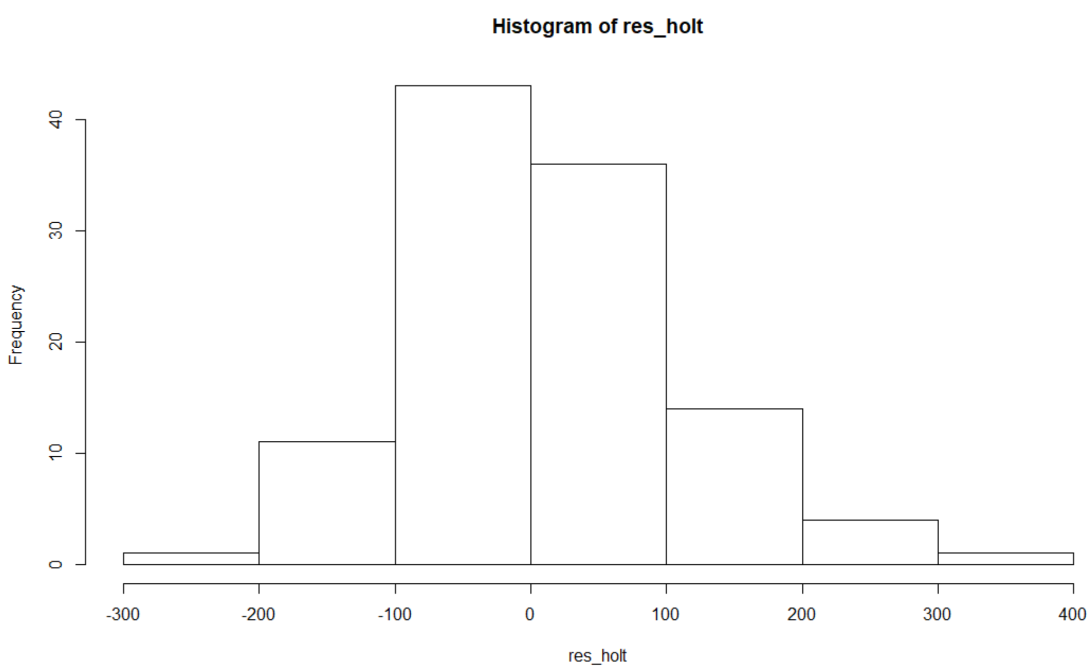
- What is the value of sigma? What does the sigma signify?
  sd(complete.cases(holt_forecast$residuals))
  0.2990297
  The value of sigma signifies the value of standard deviation.

- Perform Residual Analysis for this technique.
  - Do a plot of residuals. What does the plot indicate?
    res_holt<-residuals(holt)
    plot(res_holt)

    We can say that the variation in residuals is high.

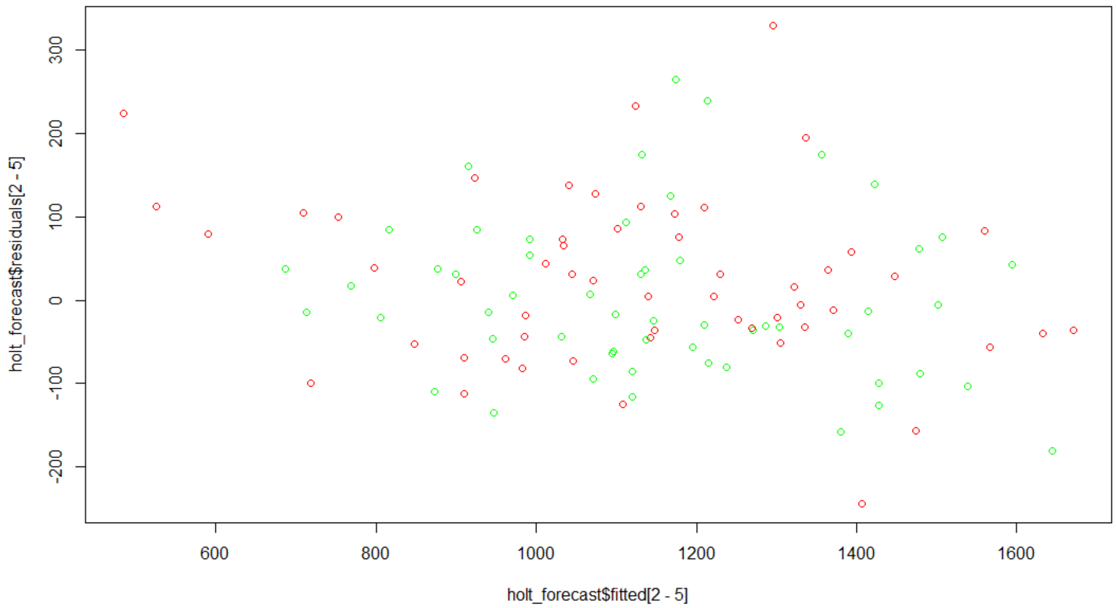


- Do a Histogram plot of residuals. What does the plot indicate?
  We can see that data is right skewed.

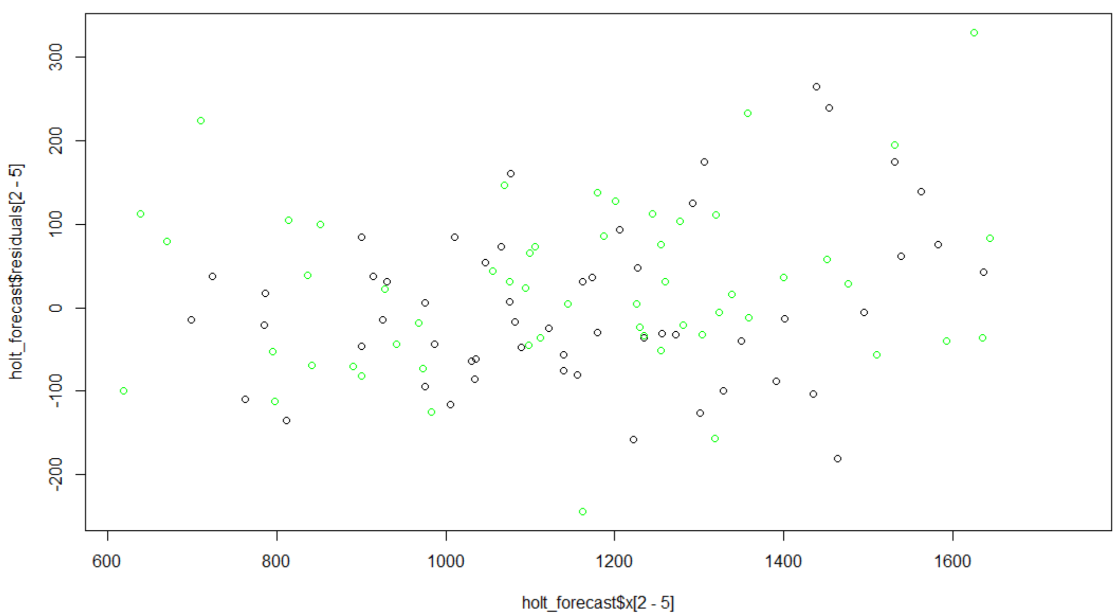


- Do a plot of fitted values vs. residuals. What does the plot indicate?

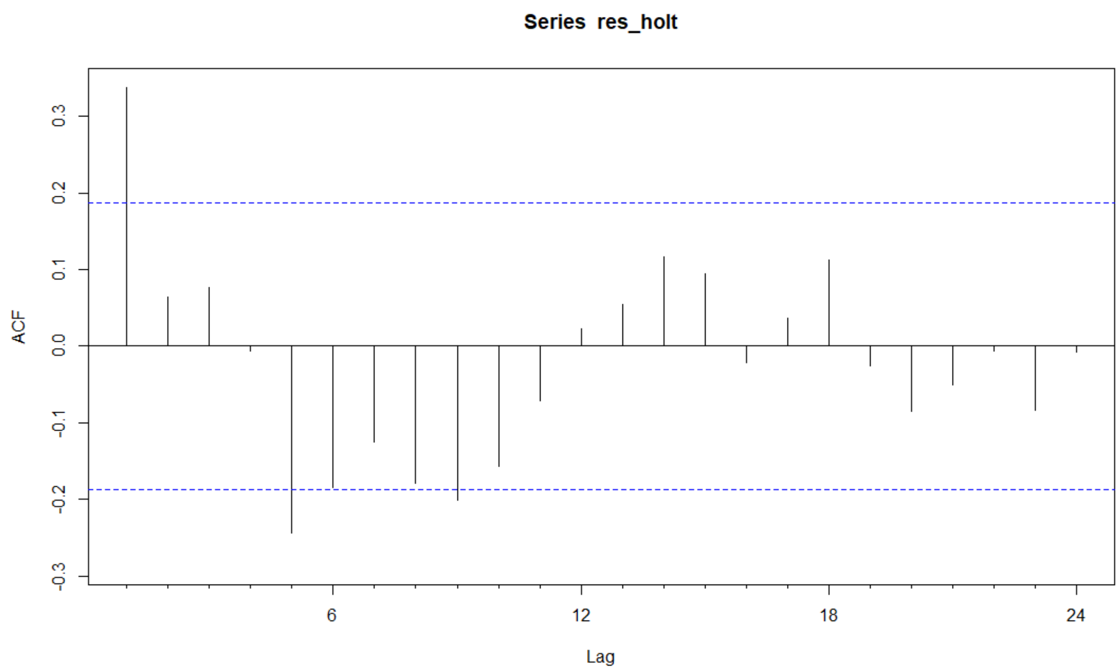
  The fitted and residual variation is initially high, which eventually reduces
  plot(holt_forecast$fitted[2-5],holt_forecast$residuals[2-5],col=c("red","green"))



- Do a plot of actual values vs. residuals. What does the plot indicate?
  plot(holt_forecast$x[2-5],holt_forecast$residuals[2-5],col=c("Green","Black"))
  The forecasted and residual variation is initially high, which eventually reduces



- Do an ACF plot of the residuals? What does this plot indicate?
  We can see that values are not very significant
  Acf(res_holt)



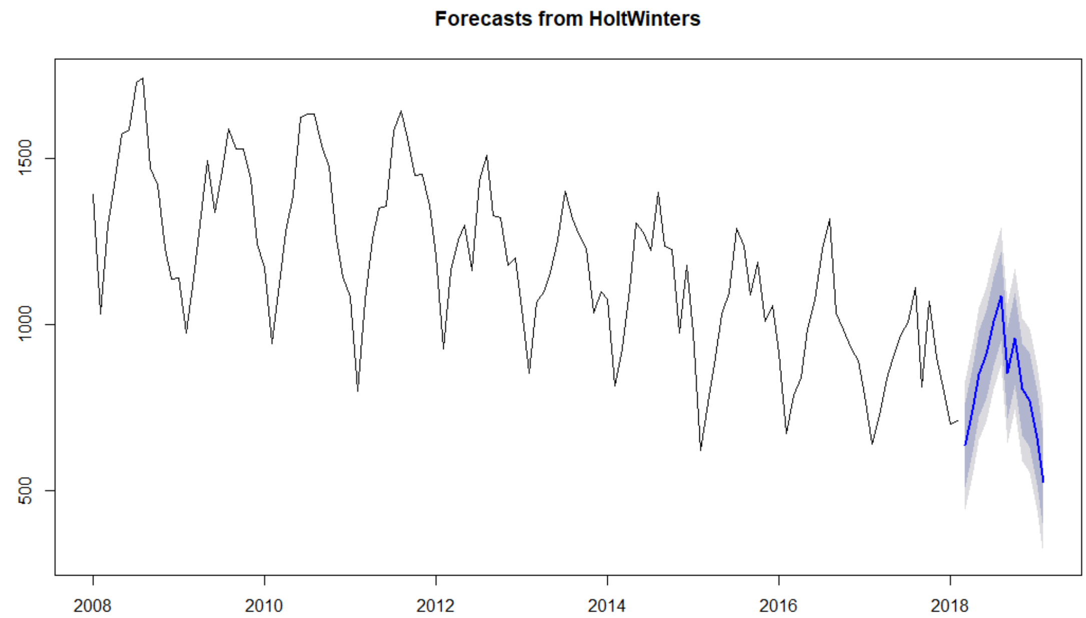
- Print the 5 measures of accuracy for this forecasting technique
  accuracy(holt_forecast)

```
> accuracy(holt_forecast)
                   ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
Training set 12.20904 99.58964 77.59005 0.7944291 6.939642 0.7170382 0.3374136
```

- Forecast
  - Time series value for next year. Show table and plot

forecast(holt, h=12)
plot(forecast(holt, h=12))

```
> forecast(holt, h=12)
         Point Forecast    Lo 80     Hi 80    Lo 95     Hi 95
Mar 2018       635.0188 507.7725  762.2651 440.4124  829.6252
Apr 2018       741.3714 612.5219  870.2209 544.3132  938.4296
May 2018       851.1765 720.7254  981.6276 651.6687 1050.6842
Jun 2018       909.6484 777.5971 1041.6998 707.6933 1111.6035
Jul 2018      1010.0735 876.4231 1143.7239 805.6729 1214.4741
Aug 2018      1086.0376 950.7893 1221.2859 879.1931 1292.8820
Sep 2018       853.1847 716.3394  990.0300 643.8979 1062.4715
Oct 2018       957.0287 818.5872 1095.4701 745.3007 1168.7566
Nov 2018       803.7899 663.7530  943.8269 589.6220 1017.9579
Dec 2018       771.2101 629.5783  912.8420 554.6030  987.8173
Jan 2019       659.8292 516.6029  803.0554 440.7835  878.8748
Feb 2019       523.9171 379.0968  668.7375 302.4336  745.4007
> plot(forecast(holt, h=12))
> 
```



- Summarize this forecasting technique
  - How good is the accuracy?
    The MSE and MAPE is low, hence accuracy is high
  - What does it predict the value of time series will be in one year?
    The predicted value for Feb 2019 is 523.9171
  - Other observation
    No such notable observation.

ARIMA or Box-Jenkins

- Is Time Series data stationary? How did you verify? Please post the output from one of the test

We perform the adf and kpss test to see if model is stationery. Since p value is <0.
05 for ADF and not >0.05 for kpss, we can say that model is NOT stationery.

adf.test(crime_ts)
kpss.test(crime_ts)

```
> adf.test(crime_ts)

        Augmented Dickey-Fuller Test

data:  crime_ts
Dickey-Fuller = -9.1282, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(crime_ts) : p-value smaller than printed p-value
> kpss.test(crime_ts)

        KPSS Test for Level Stationarity

data:  crime_ts
KPSS Level = 2.1832, Truncation lag parameter = 2, p-value = 0.01

Warning message:
In kpss.test(crime_ts) : p-value smaller than printed p-value
>
```

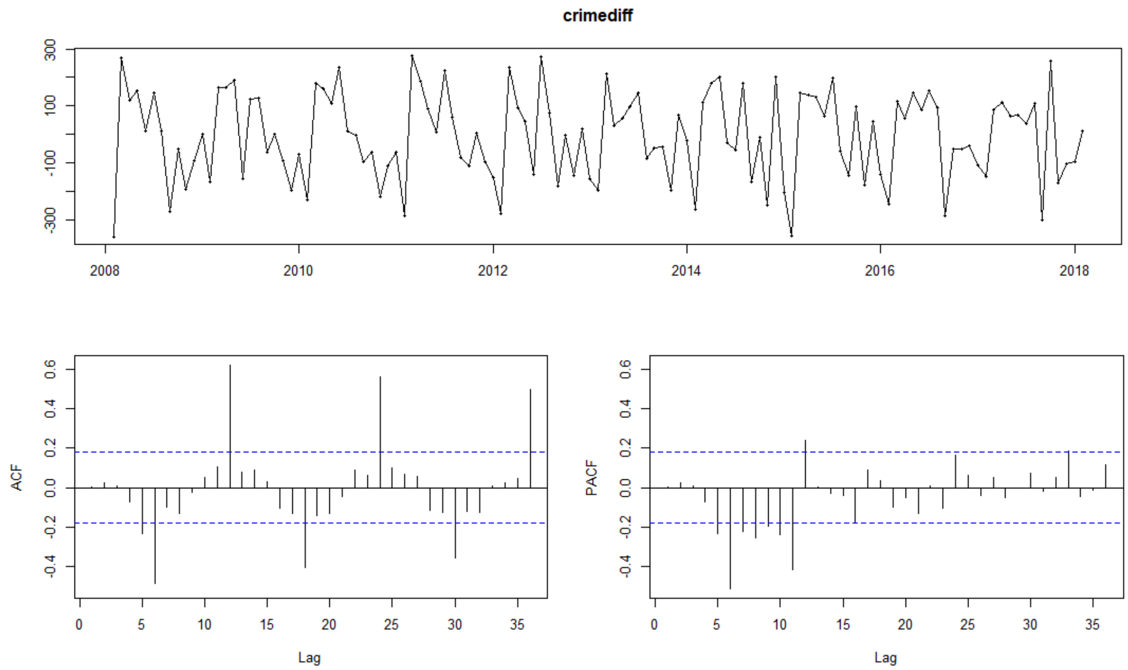
- How many differences are needed to make it stationary?

```
> ndiffs(crime_ts)
[1] 1
```

- Is Seasonality component needed?
  No, seasonal component is not needed.

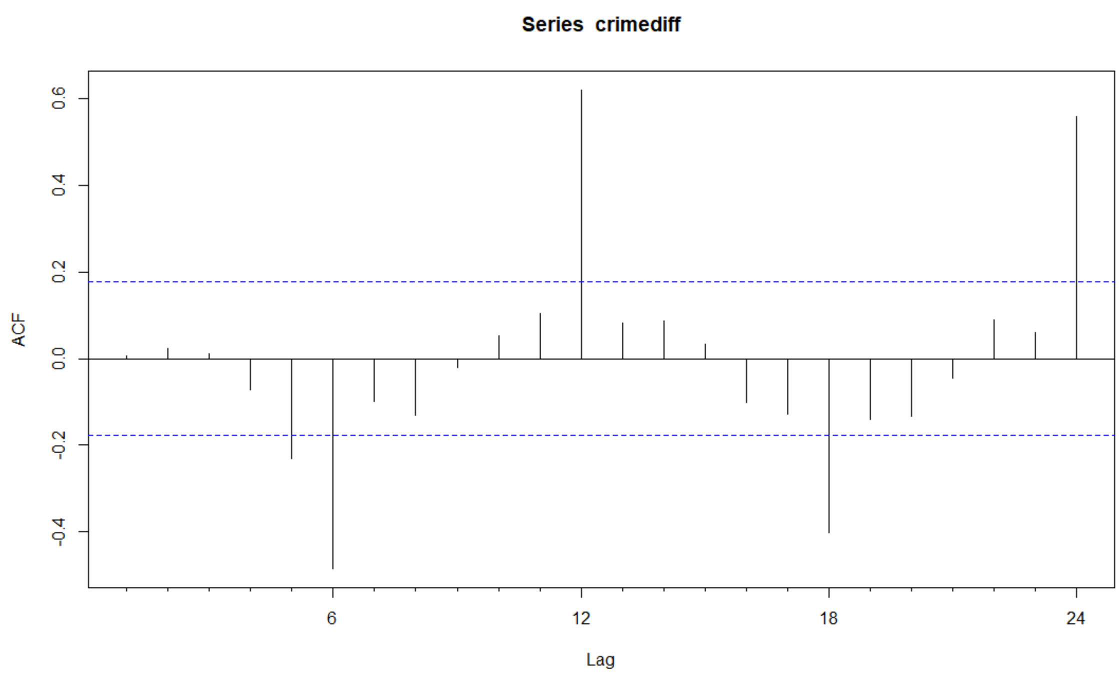
- Plot the Time Series chart of the differenced series.

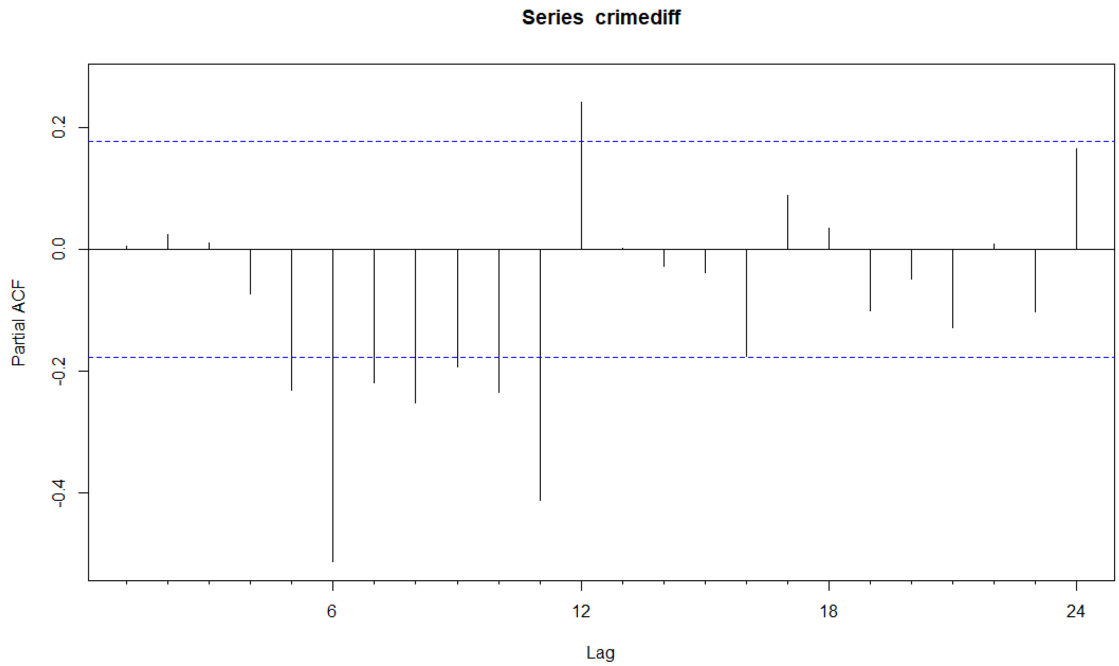
crimediff<- diff(crime_ts, differences=1)
tsdisplay(crimediff)



- Plot the ACF and PACF plot of the differenced series.

Acf(crimediff)
Pacf(crimediff)





- Based on the ACF and PACF, which are the possible ARIMA model possible?
  ARIMA(1,0,1)(0,1,1)
  auto.arima(crimediff)

```
> auto.arima(crimediff)
Series: crimediff
ARIMA(1,0,1)(0,1,1)[12]

Coefficients:
         ar1      ma1     sma1
      0.4834  -0.9652  -0.7218
s.e.  0.1014   0.0512   0.1113

sigma^2 estimated as 7415:  log likelihood=-644.68
AIC=1297.37   AICc=1297.75   BIC=1308.13
>
```

- Show the AIC, BIC and Sigma^2 for the possible models?

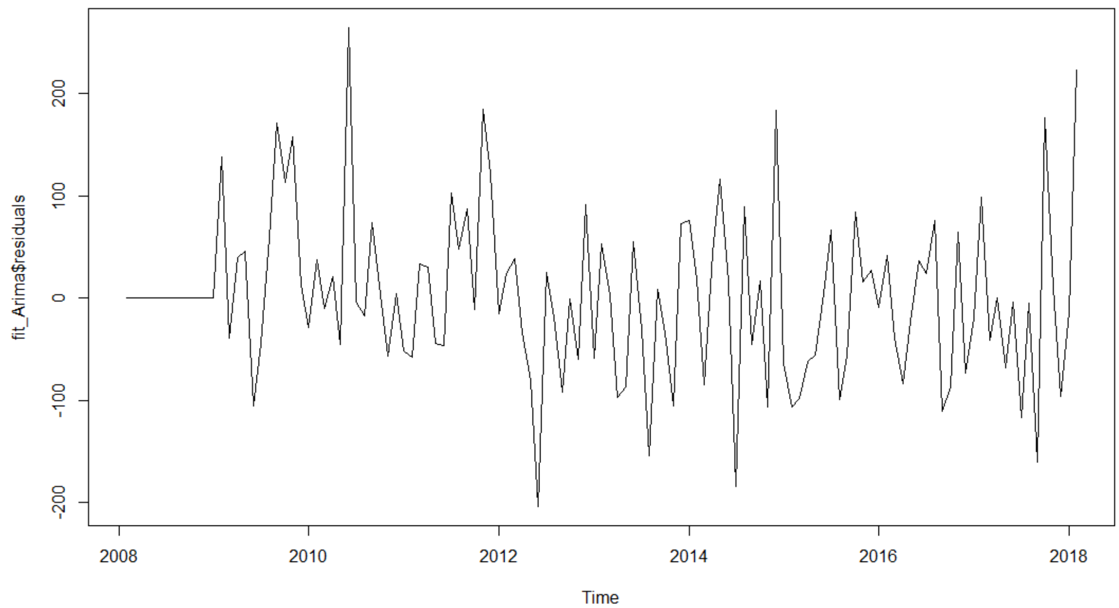
```
sigma^2 estimated as 7415:  log likelihood=-644.68
AIC=1297.37   AICc=1297.75   BIC=1308.13
```

- Based on the above AIC, BIC and Sigma^2 values, which model will you select?
  ARIMA(1,0,1)(0,1,1)

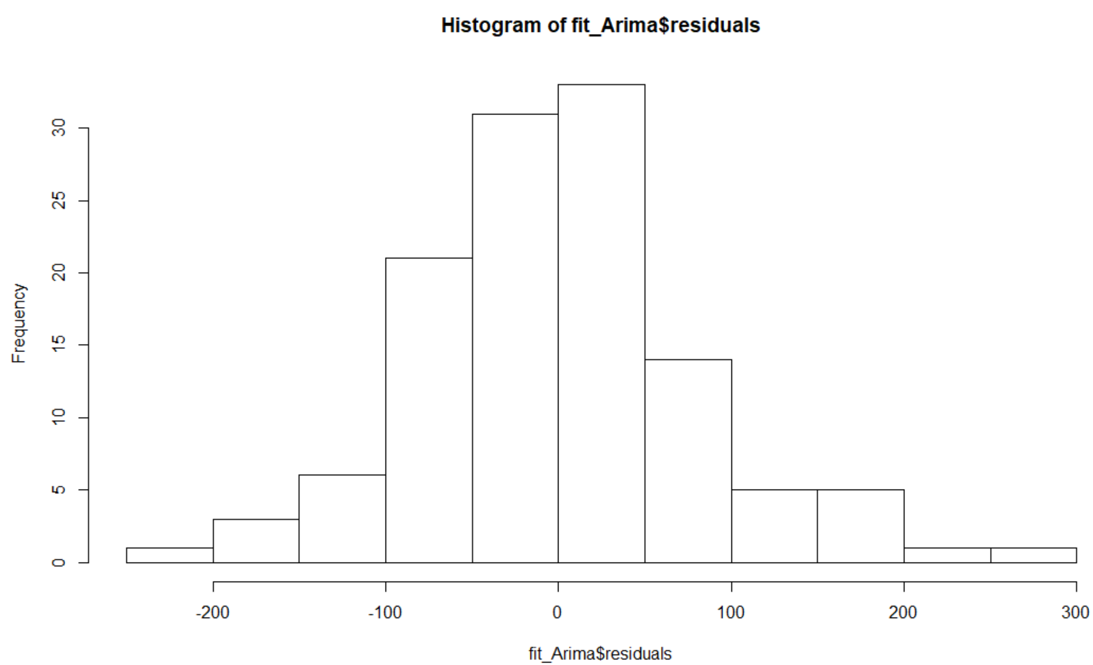
- What is the final formula for ARIMA with the coefficients?
  p=1, d=0, q=1
  P=0, D=1, Q=1

- Perform Residual Analysis for this technique.
  - Do a plot of residuals. What does the plot indicate?
    Initially, residuals are zero, which fluctuate after 2010.

fit_Arima<-auto.arima(crimediff)
fit_Arima
plot(fit_Arima$residuals)



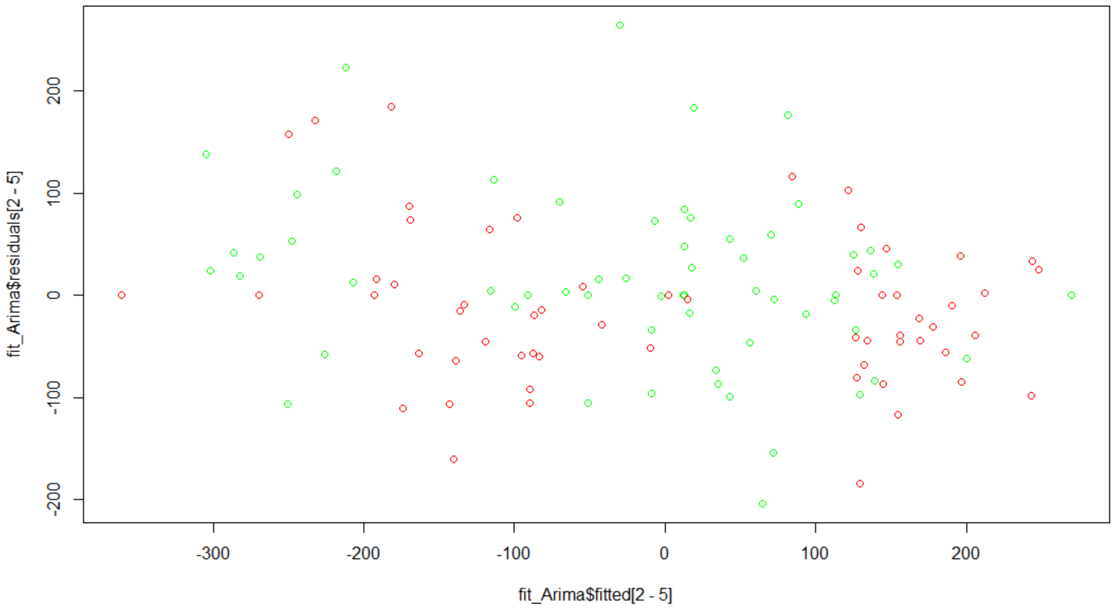
- Do a Histogram plot of residuals. What does the plot indicate?
  Data is almost symmetrical. Slightly right skewed
  hist(fit_Arima$residuals)



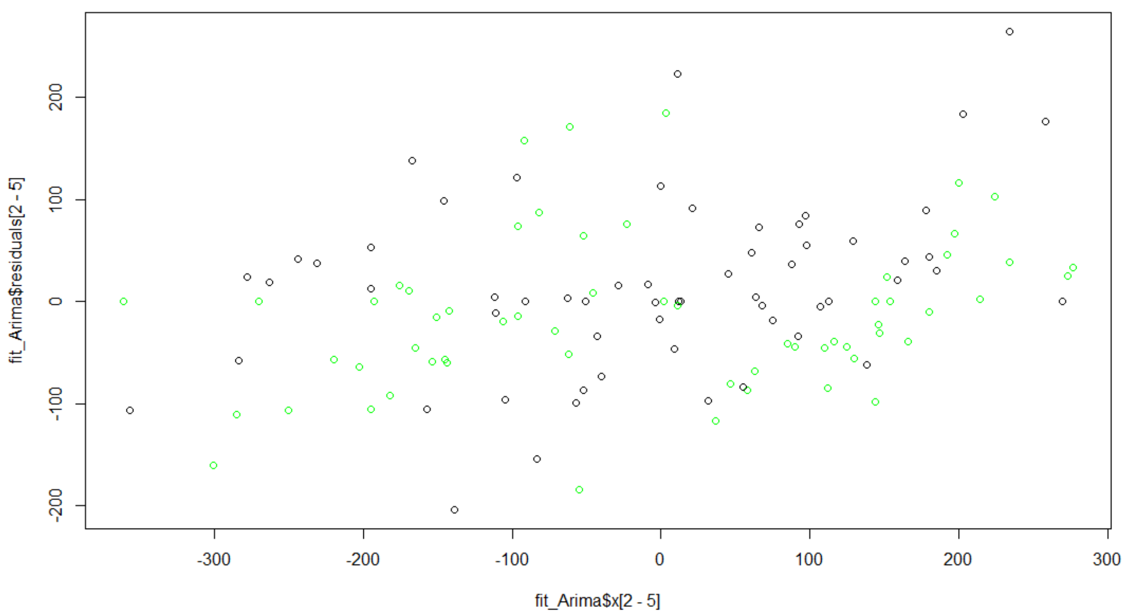
- Do a plot of fitted values vs. residuals. What does the plot indicate?

  Initially, residuals are zero, which fluctuate later.

plot(fit_Arima$fitted[2-5],fit_Arima$residuals[2-5],col=c("red","green"))

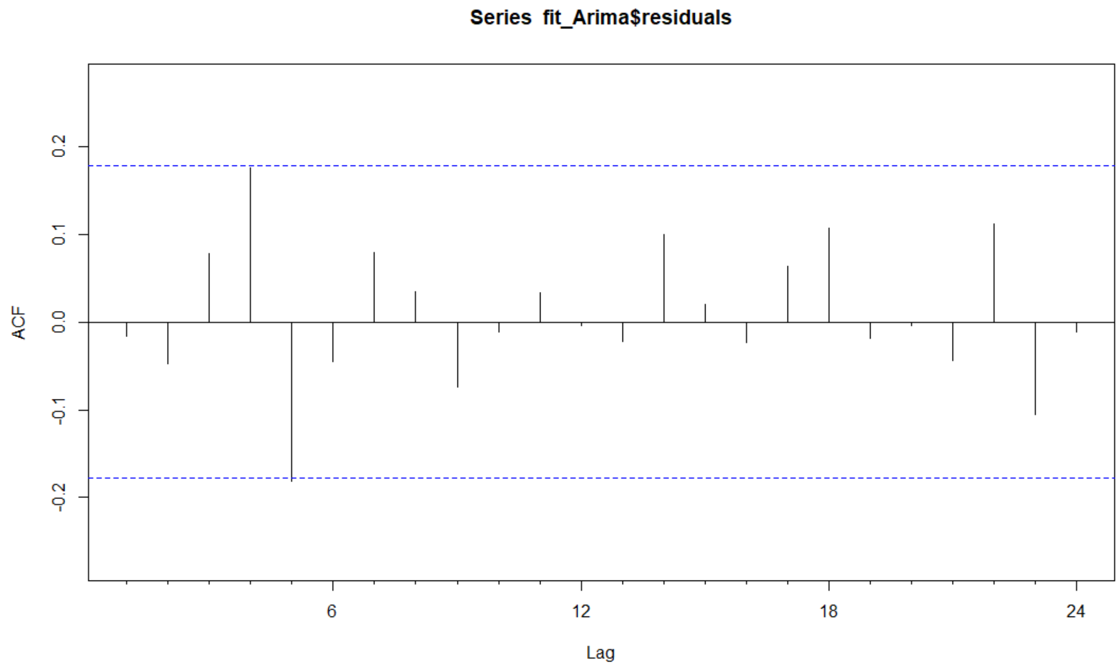


- Do a plot of actual values vs. residuals. What does the plot indicate?
  Initially, residuals are zero, which fluctuate later.
  (fit_Arima$x[2-5],fit_Arima$residuals[2-5],col=c("Green","Black"))



- Do an ACF plot of the residuals? What does this plot indicate?

  We can see that only couple of values are significant
  Acf(fit_Arima$residuals)

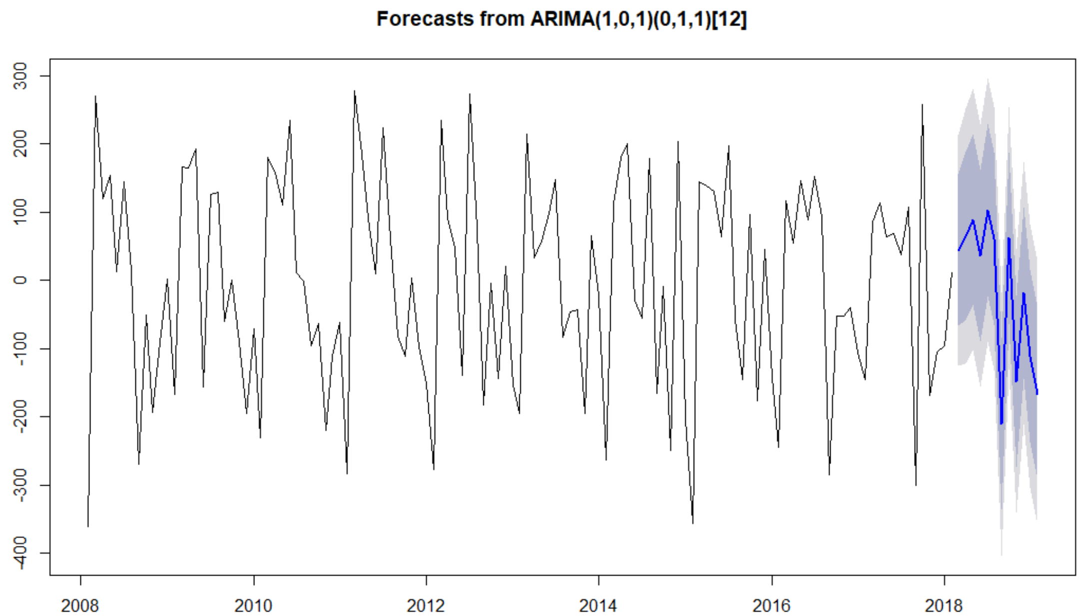


- Print the 5 measures of accuracy for this forecasting technique.

```
> accuracy(fit_Arima)
                    ME     RMSE      MAE MPE MAPE      MASE        ACF
Training set 0.5585606 80.59614 59.39053 Inf  Inf 0.6288069 -0.0152400
>
```

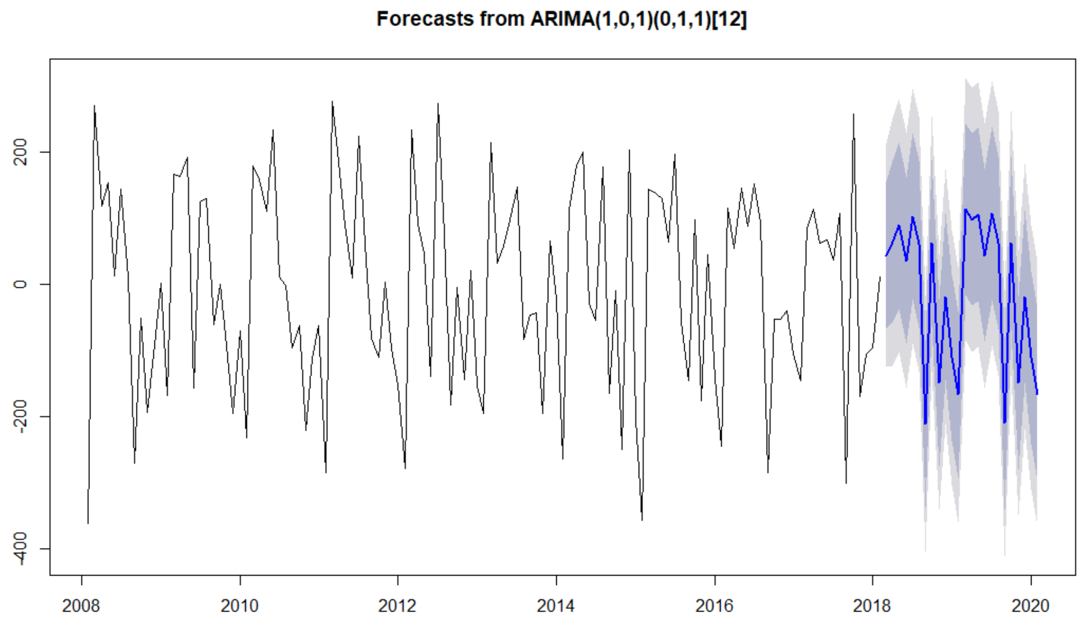
- Forecast
  - Next one year. Show table and plot

```
> forecast(fit_Arima, h=12)
         Point Forecast      Lo 80     Hi 80      Lo 95     Hi 95
Mar 2018       44.04921  -66.41948 154.51791 -124.89808 212.99651
Apr 2018       64.50171  -58.04954 187.05296 -122.92425 251.92768
May 2018       89.14110  -36.06586 214.34807 -102.34643 280.62864
Jun 2018       36.07212  -89.74741 161.89164 -156.35225 228.49648
Jul 2018      103.18465  -22.77760 229.14690  -89.45799 295.82729
Aug 2018       61.15065  -64.84493 187.14624 -131.54296 253.84427
Sep 2018     -210.95138 -336.95473 -84.94802 -403.65688 -18.24587
Oct 2018       62.44329  -63.56184 188.44842 -130.26493 255.15151
Nov 2018     -148.34246 -274.34781 -22.33711 -341.05101  44.36610
Dec 2018      -18.38280 -144.38736 107.62177 -211.09015 174.32455
Jan 2019     -110.83451 -236.83532  15.16629 -303.53611  81.86708
Feb 2019     -166.79065 -292.77505 -40.80626 -359.46716  25.88585
> plot(forecast(fit_Arima, h=12))
>
```



- Next two years. Show table and plot

```
> forecast(fit_Arima, h=24)
         Point Forecast       Lo 80     Hi 80      Lo 95     Hi 95
Mar 2018       44.04921  -66.41948 154.51791 -124.89808 212.99651
Apr 2018       64.50171  -58.04954 187.05296 -122.92425 251.92768
May 2018       89.14110  -36.06586 214.34807 -102.34643 280.62864
Jun 2018       36.07212  -89.74741 161.89164 -156.35225 228.49648
Jul 2018      103.18465  -22.77760 229.14690  -89.45799 295.82729
Aug 2018       61.15065  -64.84493 187.14624 -131.54296 253.84427
Sep 2018     -210.95138 -336.95473 -84.94802 -403.65688 -18.24587
Oct 2018       62.44329  -63.56184 188.44842 -130.26493 255.15151
Nov 2018     -148.34246 -274.34781 -22.33711 -341.05101  44.36610
Dec 2018      -18.38280 -144.38736 107.62177 -211.09015 174.32455
Jan 2019     -110.83451 -236.83532  15.16629 -303.53611  81.86708
Feb 2019     -166.79065 -292.77505 -40.80626 -359.46716  25.88585
Mar 2019      113.19226  -16.49498 242.87949  -85.14725 311.53176
Apr 2019       97.92719  -32.60138 228.45576 -101.69903 297.55341
May 2019      105.29982  -25.42459 236.02422  -94.62591 305.22554
Jun 2019       43.88364  -86.88649 174.65377 -156.11202 243.87930
Jul 2019      106.96094  -23.81988 237.74175  -93.05106 306.97293
Aug 2019       62.97620  -67.80710 193.75951 -137.03961 262.99202
Sep 2019     -210.06886 -340.85274 -79.28498 -410.08554 -10.05217
Oct 2019       62.86992  -67.91405 193.65390 -137.14690 262.88675
Nov 2019     -148.13621 -278.92002 -17.35241 -348.15278  51.88035
Dec 2019      -18.28310 -149.06605 112.49986 -218.29837 181.73218
Jan 2020     -110.78631 -241.56563  19.99300 -310.79601  89.22339
Feb 2020     -166.76735 -297.53086 -36.00384 -366.75288  33.21818
> plot(forecast(fit_Arima, h=24))
> plot(forecast(fit_Arima, h=12))
```



- Summarize this forecasting technique
  - How good is the accuracy?
    MASE is 0.62, which is low, hence accuracy is good.

  - What does it predict time series will be in one year and next two years?
    Prediction for Feb 2019 is -166.790
    Prediction for Feb 2020 is -166.767

  - Other observation
    Point forecast is highly fluctuating every month.

Accuracy Summary

- Show a table of all the forecast method above with their accuracy measures.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Naïve	-5.63636	153.4663	129.1074	-1.6263	12.004	1.1931	0.006097
Simple Moving Averages	-6.3984	80.3493	62.6294	-0.86532	5.6647	0.5787	0.05803
Simple Smoothing	-5.5862	152.83	128.0557	-1.612	11.906	1.18341	0.00542
Holt winters	12.20904	99.589	77.59	0.7944	6.939642	0.717038	0.3374
ARIMA	0.558	80.596	59.390	inf	inf	0.6288	-0.0152

- Separately define each forecast method and why it is useful. Show the best and worst forecast method for each of the accuracy measures.

For naïve forecasts, we simply set all forecasts to be the value of the last observation. This method works remarkably well for many economic and financial time series.
Simple Moving Average is a method of time series smoothing and is a very basic forecasting technique. It does not need estimation of parameters, but rather is based on order selection.
The simple exponential smoothing method does not account for any trend or seasonal components, rather, it only uses the decreasing weights to forecast future results. This makes the method suitable only for time series without trend and seasonality
Holt-Winters method, is one of the many methods or algorithms that can be used to forecast data points in a series, provided that the series is "seasonal", i.e. repetitive over some period.
An autoregressive integrated moving average (ARIMA) model is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables. The model's goal is to predict future securities or financial market moves by examining the differences between values in the series instead of through actual values.

ME is best for ARIMA
RMSE is best for SMA
MAE is best for ARIMA
MPE is best for SMA
MAPE is best for SMA
MASE is best for SMA

Conclusion

- Summarize your analysis of time series value over the time-period.
    - We observe a seasonality with crime rate being low in February and high in August.
    - Since there is overall downward trend, crime rate is predicted to drop in the coming years.
    - From Feb to June there is rising trend and Aug to Dec is falling trend.

- Based on your analysis and forecast above, do you think the value of the time series will increase, decrease or stay flat over the next year? How about next 2 years?

  Based on our analysis and forecast, the crime rate is predicted to reduce over the next year and the following year as well

- Rank forecasting methods that best forecast for this time series based on historical values.
    - ARIMA
    - Simple Moving Averages
    - Holts Winters
    - Simple Smoothing
    - Naive

Final Question

- If you were me, what final grade would you give yourself for this class?

  I would give myself an A.

- Indicate the reasons why you gave yourself this grade?

  I feel I have had an active participation in the class throughout, completed the assignments and exams satisfactorily in a timely manner.

R code:

```
#Final_exam_Asher

View(crimes)
crime_ts <- ts(crimes$Data, start=c(2008,1),frequency = 12)
plot(crime_ts)
summary(crime_ts)
boxplot(crime_ts)
decomp<-decompose(crime_ts)
plot(decomp)
decomp$type
decomp$seasonal

library(fpp)
temp_sesadjust<-seasadj(decomp)
plot(crime_ts)
lines(temp_sesadjust,col='red')

#Naive
naive_forecast<-naive(crime_ts,12)
plot(naive_forecast)

residual_analysis<-residuals(naive_forecast)
plot(residual_analysis)

histo<-hist(residual_analysis,breaks =10,main = "Histogram of residuals")
plot(naive_forecast$fitted[2-5],naive_forecast$residuals[2-5],col=c("red","blue"))
plot(naive_forecast$x[2-5],naive_forecast$residuals[2-5],col=c("red","blue"))
Acf(residual_analysis)

accuracy(naive_forecast)

naive_fore<- forecast(naive_forecast, h=12)
naive_fore
plot(naive_fore)

#SMA
ma_fore<-ma(crime_ts, order=1)
plot(ma_fore)
```

```
#forecast for accuracy table

ma_forecast<-forecast(ma_fore, h=12)
ma_forecast
accuracy(ma_forecast)


ma3_fore<-ma(crime_ts, order=3)
lines(ma3_fore,col="red",lwd=3)

ma6_fore<-ma(crime_ts, order=6)
lines(ma6_fore,col="blue",lwd=3)

ma9_fore<-ma(crime_ts, order=9)
lines(ma9_fore,col="green",lwd=3)

ets_forecast <- ets(crime_ts)
fore_ets<-forecast.ets(ets_forecast, h=12)
fore_ets
plot(fore_ets)

#smoothing
library (fpp)
ses(crime_ts, h=12)
summary(ses(crime_ts, h=12))

ses_fore <- ses(crime_ts, h=12)
ses_residual_ana <- residuals(ses_fore)
ses_residual_ana
plot(ses_residual_ana)

hist(ses_residual_ana)

fit_s<-ses_fore$fitted
fit_s
plot(fit_s[2-5],ses_residual_ana[2-5], col=c("Red","Green"))

actual<-ses_fore$x
plot(actual[2-5],ses_residual_ana[2-5],col=c("Blue","Green"))

Acf(ses_residual_ana)
accuracy(ses_fore)
forecast(ses_fore, h=12)
plot(forecast(ses_fore, h=12))
summary(ses_fore)

#Holt winters
holt<-HoltWinters(crime_ts)
holt_forecast<-forecast(holt, h=12)
holt_forecast
plot(holt_forecast)
```

```
sd(complete.cases(holt_forecast$residuals))

res_holt<-residuals(holt)
plot(res_holt)
hist(res_holt)

plot(holt_forecast$fitted[2-5],holt_forecast$residuals[2-5],col=c("red","green"))
plot(holt_forecast$x[2-5],holt_forecast$residuals[2-5],col=c("Green","Black"))
Acf(res_holt)
accuracy(holt_forecast)

forecast(holt, h=12)
plot(forecast(holt, h=12))

# ARIMA
crime_ts
adf.test(crime_ts)
kpss.test(crime_ts)
ndiffs(crime_ts)

crimediff<- diff(crime_ts, differences=1)
tsdisplay(crimediff)

Acf(crimediff)
Pacf(crimediff)
tsdisplay(crime_ts)

acf(crimediff, lag.max=20)

Acf(crimediff, lag.max=20, plot=FALSE)

Pacf(crimediff, lag.max=20)
Pacf(crimediff, lag.max=20, plot=FALSE)

fit_Arima<-auto.arima(crimediff)
fit_Arima
plot(fit_Arima$residuals)
hist(fit_Arima$residuals)
Acf(fit_Arima$residuals)

plot(fit_Arima$fitted[2-5],fit_Arima$residuals[2-5],col=c("red","green"))
plot(fit_Arima$x[2-5],fit_Arima$residuals[2-5],col=c("Green","Black"))
accuracy(fit_Arima)
forecast(fit_Arima, h=12)
plot(forecast(fit_Arima, h=12))

forecast(fit_Arima, h=24)
    plot(forecast(fit_Arima, h=24))
```