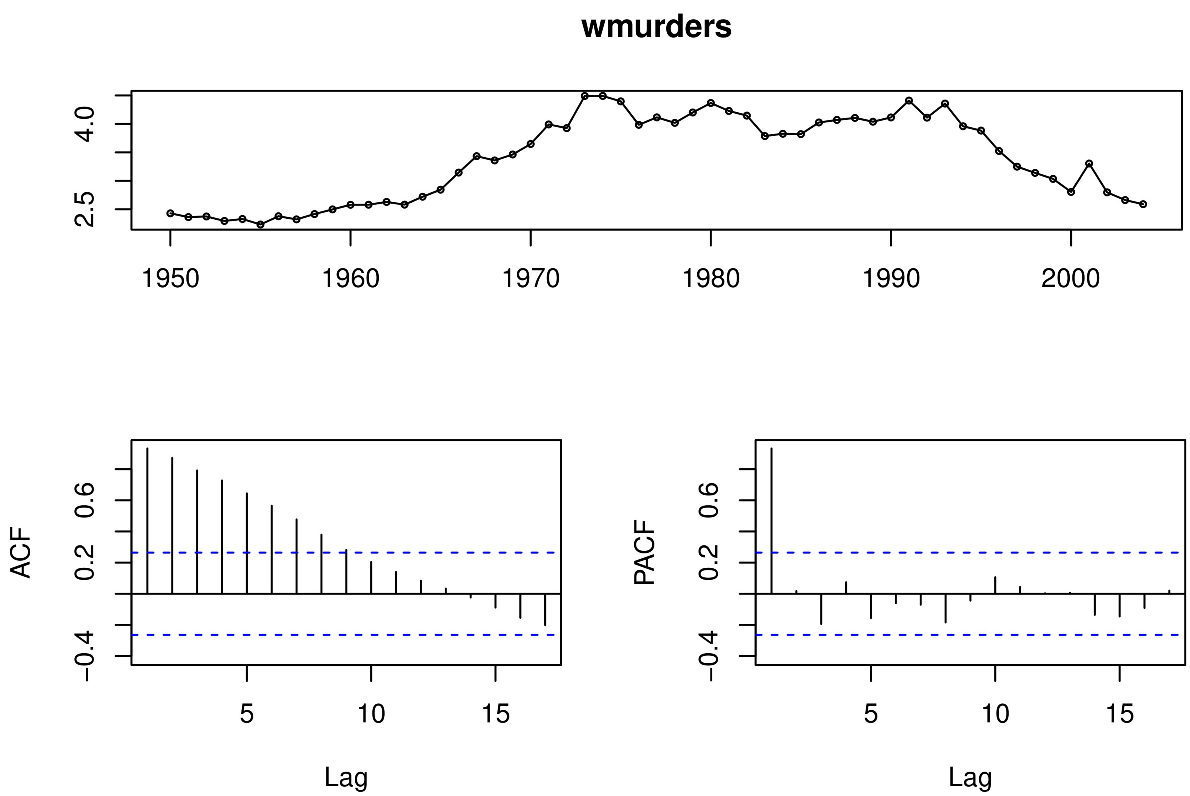
Female Murder Time Series Analysis

Dataset

Annual female murder rate (per 100,000 standard population) in the USA. 1950-2004.

```
data(wmurders)
tsdisplay(wmurders)
```



Looking at our initial ACF and PACF plots, we see that on ACF, there is an exponentially decaying of lags. On PACF, there is a spike at lag 1 but none beyond lag 1. It looks like we can choose AR model. However, on our time series plot, our data is seem non-stationary. We need to evaluate it.

Stationary Evaluation

```
kpss.test(wmurders)
```

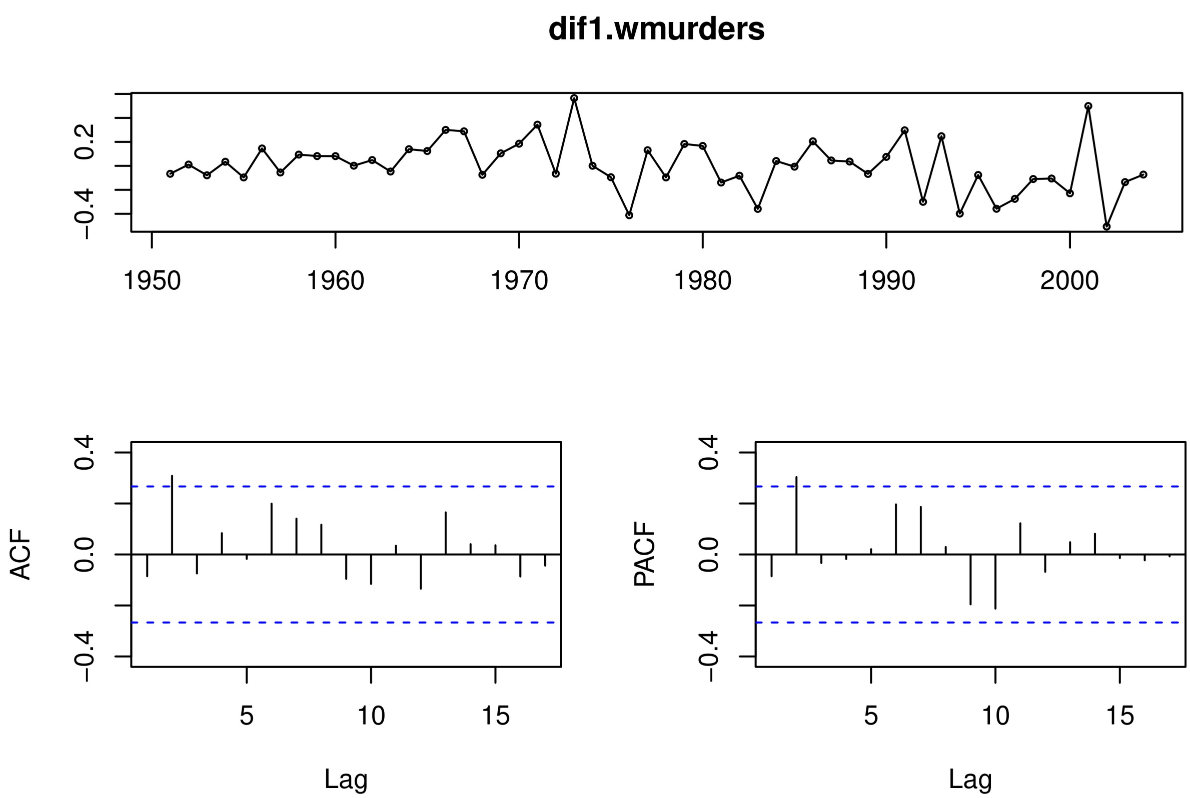
```
##
```

```
##  KPSS Test for Level Stationarity
##
## data:  wmurders
## KPSS Level = 0.63314, Truncation lag parameter = 3, p-value = 0.01962
```

From the test, our p-value = 0.01 is less than 0.05 and we reject the null hypothesis saying that the series is non-stationary. It is suggested that differencing is needed.

Differencing

```
dif1.wmurders <- diff(wmurders)
tsdisplay(dif1.wmurders)
```




```
kpss.test(dif1.wmurders)
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  dif1.wmurders
## KPSS Level = 0.46973, Truncation lag parameter = 3, p-value = 0.04848
```

Even though we tried first differencing, our test still reject null hypothesis saying that the series in non-stationary. We may need another differencing.

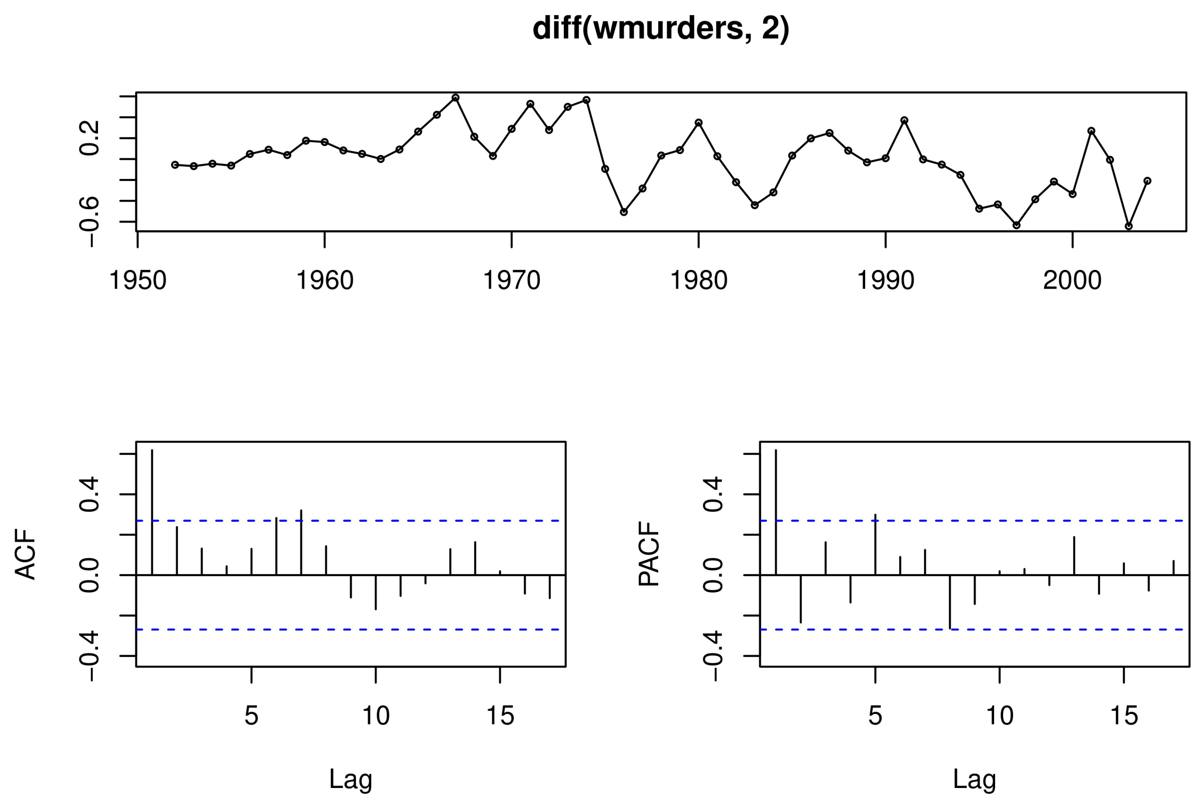
```r
dif2.wmurders <- diff(diff(wmurders))
kpss.test(dif2.wmurders)
```

```
## Warning in kpss.test(dif2.wmurders): p-value greater than printed p-value
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  dif2.wmurders
## KPSS Level = 0.045793, Truncation lag parameter = 3, p-value = 0.1
```

After another differencing, we can see our series is stationary with p-value > 0.05.

```r
tsdisplay(diff(wmurders,2))
```



Looking at ACF and PACF, we can see lag 1 has highest spike. We can assume p or q is 1 for our model.

ARIMA MA (Moving Average) Model

```r
fit <- Arima(wmurders, order=c(0,2,1))
summary(fit)
```

```
## Series: wmurders
## ARIMA(0,2,1)
##
## Coefficients:
##           ma1
##       -0.8995
## s.e.   0.0669
##
## sigma^2 estimated as 0.04747:  log likelihood=5.24
## AIC=-6.48   AICc=-6.24   BIC=-2.54
##
## Training set error measures:
##                       ME      RMSE       MAE        MPE     MAPE      MASE
## Training set -0.01306101 0.2118445 0.1559694 -0.3151353 4.447123 0.9591385
##                     ACF1
## Training set -0.2011523
```

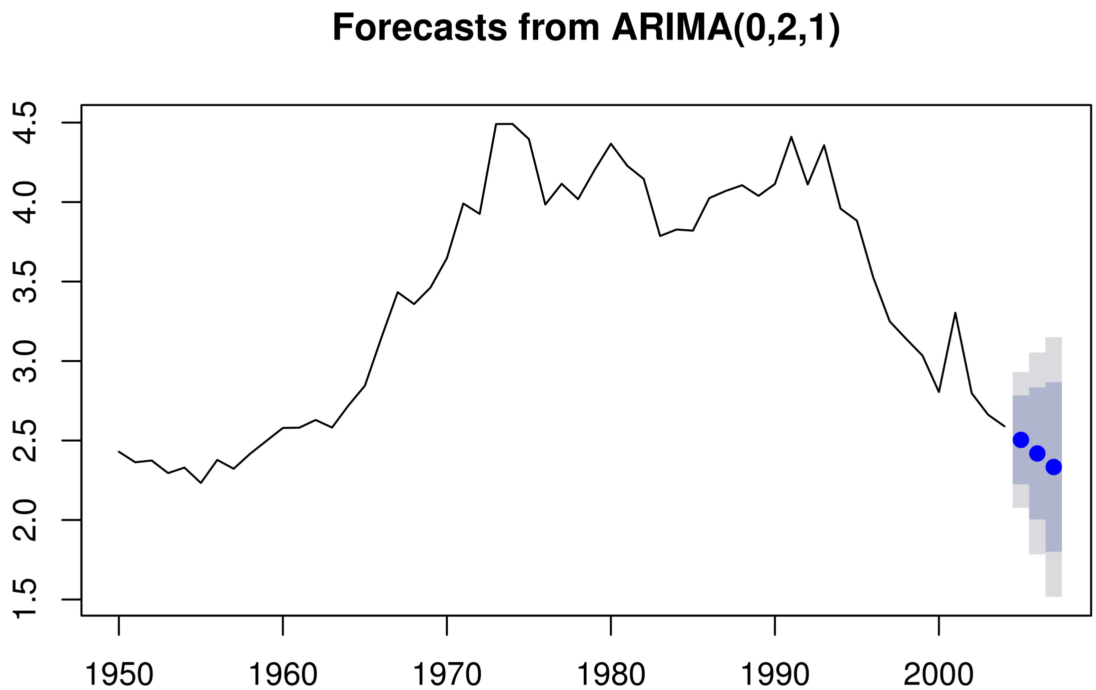
```
forecast(fit,h=3)
```

```
##      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 2005       2.504087 2.224876 2.783299 2.077070 2.931105
## 2006       2.418792 2.003603 2.833981 1.783815 3.053769
## 2007       2.333496 1.799788 2.867204 1.517260 3.149732
```

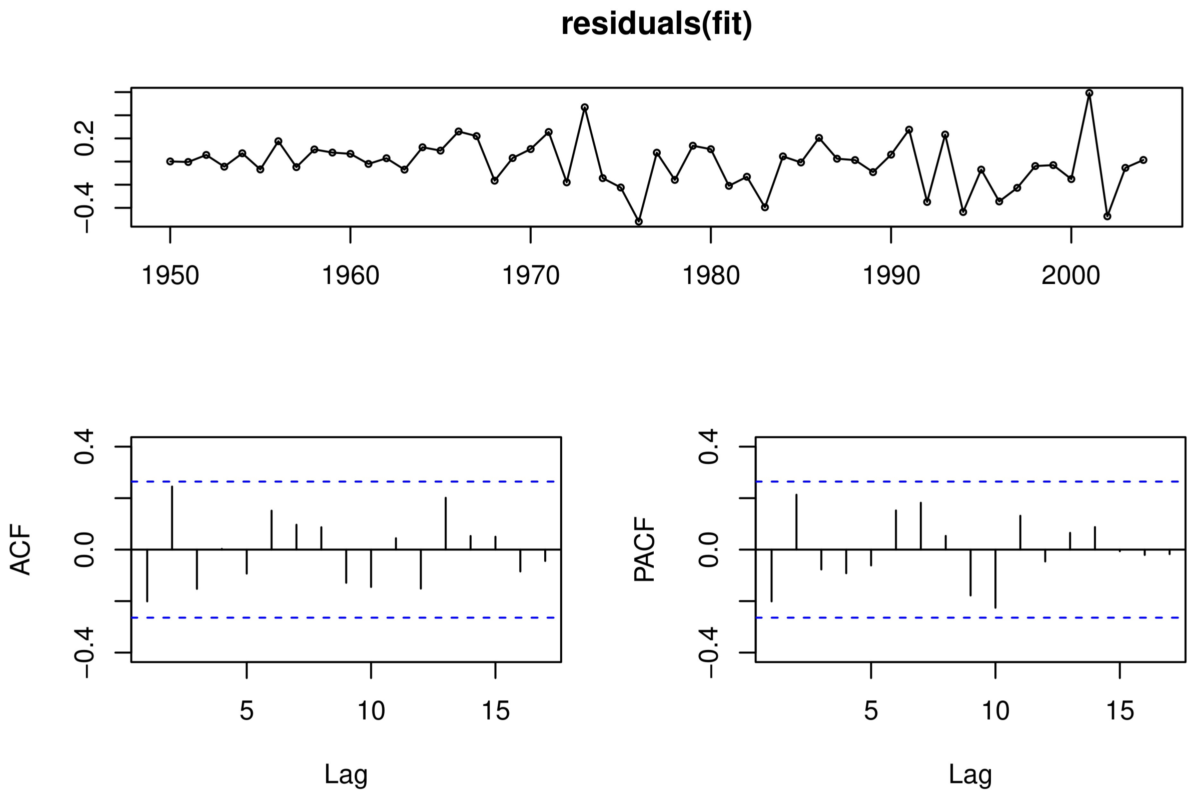
Should We Inlcude Constant?

By default in ARIMA, our constant equal mean and equal zero when degree of first difference is greater than 0.

```
plot(forecast(fit,h=3))
```




```
tsdisplay(residuals(fit))
```



As we can see that our model give us white noise for residual. This mean the model is optimal to gain information from our series.

Auto Fit ARIMA model

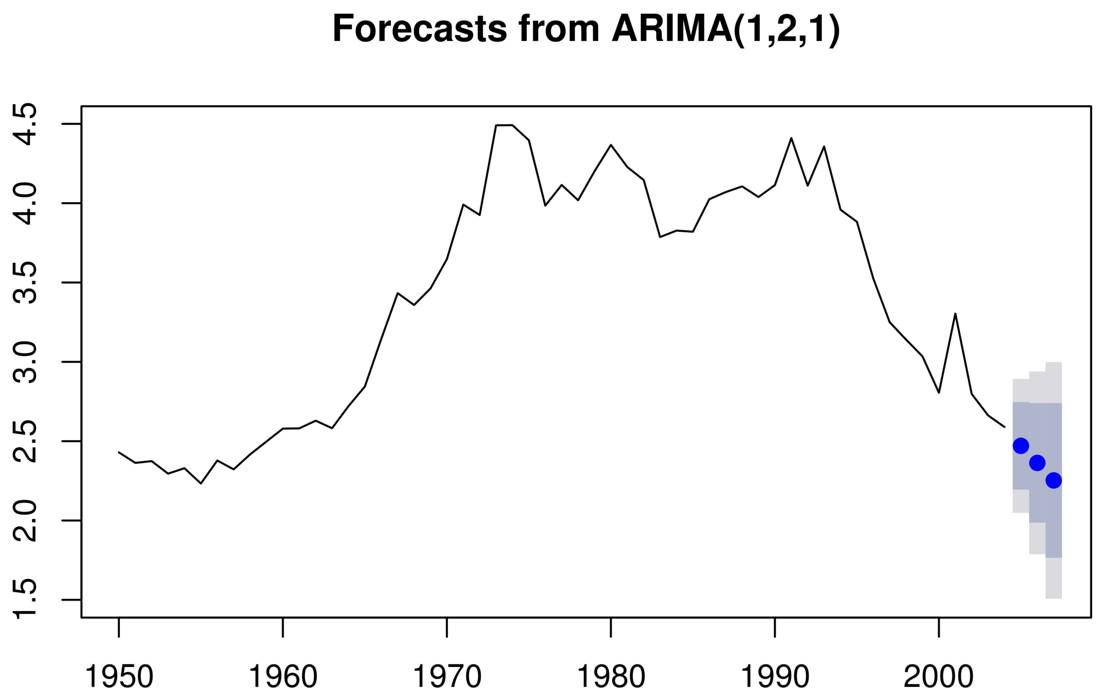
```
ar.fit<-auto.arima(wmurders)
summary(ar.fit)
```

```
## Series: wmurders
## ARIMA(1,2,1)
##
## Coefficients:
##           ar1      ma1
##       -0.2434  -0.8261
## s.e.   0.1553   0.1143
##
## sigma^2 estimated as 0.04632:  log likelihood=6.44
## AIC=-6.88   AICc=-6.39   BIC=-0.97
##
## Training set error measures:
##                       ME      RMSE       MAE        MPE     MAPE      MASE
## Training set -0.01065956 0.2072523 0.1528734 -0.2149476 4.335214 0.9400996
##                    ACF1
## Training set 0.02176343
```

```r
forecast(ar.fit,h=3)
```

```
##      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 2005       2.470660 2.194836 2.746484 2.048824 2.892496
## 2006       2.363106 1.986351 2.739862 1.786908 2.939304
## 2007       2.252833 1.765391 2.740276 1.507354 2.998313
```

```r
plot(forecast(ar.fit,h=3))
```



Conclusion:

The ARIMA model that auto fit has lower AICc than our picked MA model. We would use the ARIMA(1,2,1) for forecasting in our women murder time series.