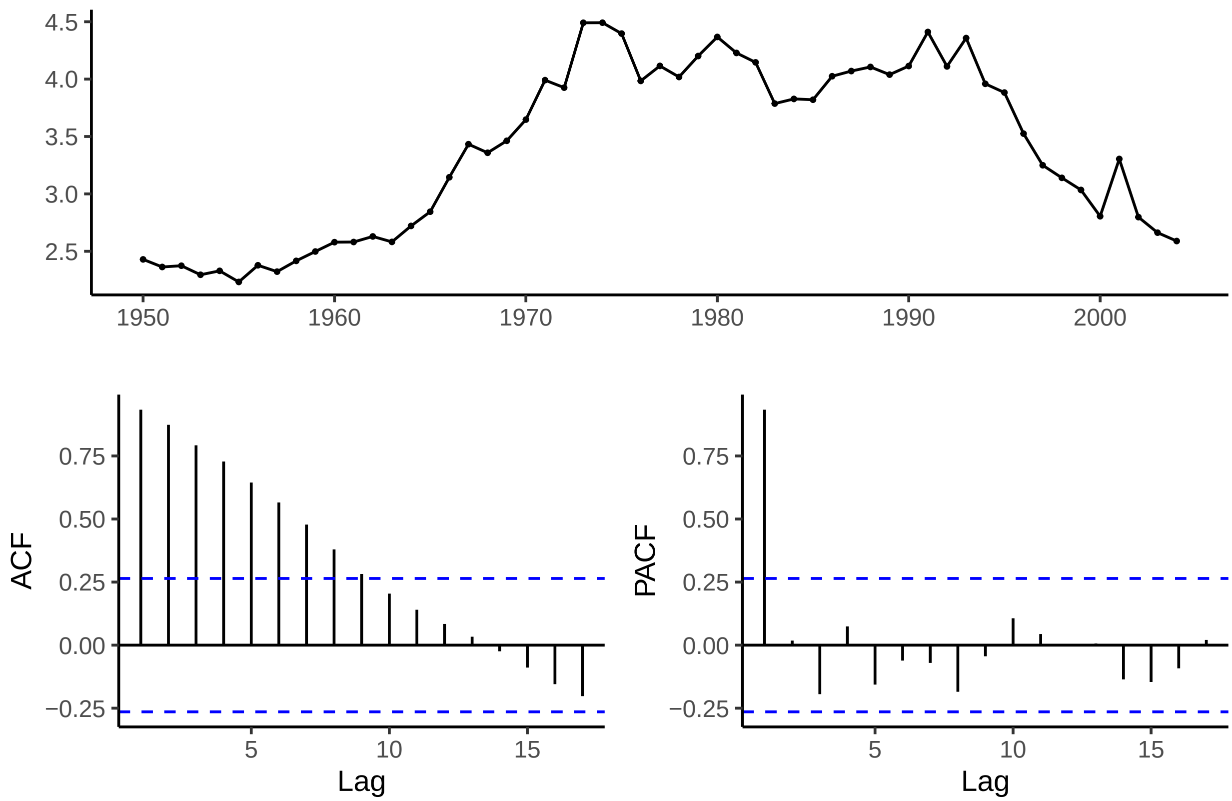
HW2

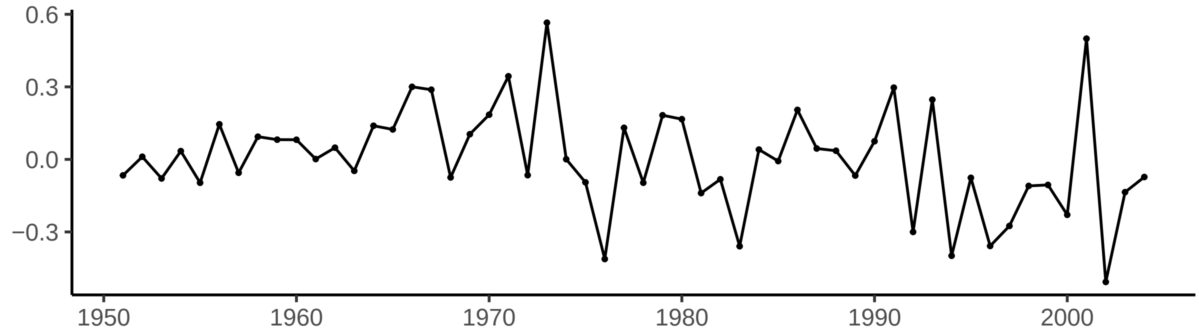
```r
# Originial time series
autoplot(wmurders) + ylab('murders')
```

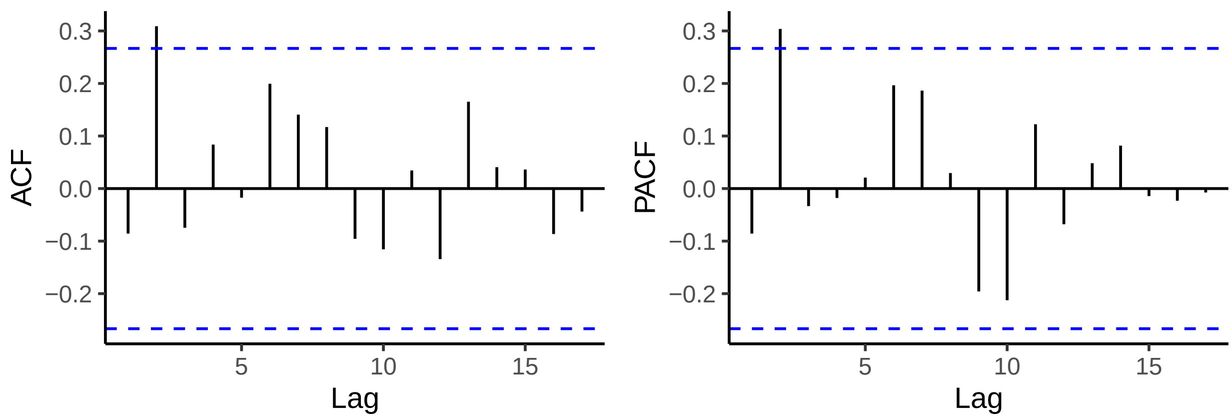
The murders time seriers does not seem to be stationary. We can notice a trend in this time series hence we must difference the data.

```r
# ACF and PACF plots of original time series
ggtsdisplay(murders$x, plot.type = 'partial')
```




```
# ACF and PACF plots for first differenced time series
ggtsdisplay(diff(murders$x), plot.type = 'partial')
```





Since the lag 1 in ACF plot is negative and it displays a sharp cutoff, we consider adding MA term to the model. Since the autocorrelation becomes zero at fifth lag, the range of q is 1-5.

```
# Choosing our model
fit1 <- Arima(murders$x, order = c(0,1,1))
summary(fit1)
```

```
## Series: murders$x
## ARIMA(0,1,1)
##
## Coefficients:
##           ma1
##       -0.0527
## s.e.   0.1070
##
## sigma^2 estimated as 0.04628:  log likelihood=6.85
## AIC=-9.7   AICc=-9.47   BIC=-5.72
##
## Training set error measures:
##                       ME      RMSE       MAE        MPE     MAPE      MASE
## Training set 0.003198768 0.2111909 0.1600467 -0.0627303 4.650887 0.9842122
##                     ACF1
## Training set -0.01729024
```

```
fit2 <- Arima(murders$x, order = c(0,1,2))
summary(fit2)
```

```
## Series: murders$x
## ARIMA(0,1,2)
##
## Coefficients:
##           ma1     ma2
##       -0.0660  0.3712
## s.e.   0.1263  0.1640
##
## sigma^2 estimated as 0.0422:  log likelihood=9.71
## AIC=-13.43   AICc=-12.95   BIC=-7.46
##
## Training set error measures:
##                        ME      RMSE       MAE         MPE     MAPE      MASE
## Training set 0.0007242355 0.1997392 0.1543531 -0.08224024 4.434684 0.9491994
##                     ACF1
## Training set 0.005880608
```

```
fit3 <- Arima(murders$x, order = c(0,1,3))
summary(fit3)
```

```
## Series: murders$x
## ARIMA(0,1,3)
##
## Coefficients:
##           ma1     ma2     ma3
##       -0.0557  0.3881  0.0274
## s.e.   0.1401  0.2000  0.1720
##
## sigma^2 estimated as 0.04298:  log likelihood=9.73
## AIC=-11.45   AICc=-10.64   BIC=-3.5
##
## Training set error measures:
##                        ME      RMSE       MAE         MPE     MAPE      MASE
## Training set 0.0003896156 0.1996238 0.1543043 -0.08814268 4.430501 0.9488991
##                     ACF1
## Training set -0.00324186
```

```
fit4 <- Arima(murders$x, order = c(0,1,4))
summary(fit4)
```

```
## Series: murders$x
## ARIMA(0,1,4)
##
## Coefficients:
##           ma1     ma2     ma3      ma4
##       -0.0951  0.4282  0.0893  -0.1151
## s.e.   0.1407  0.1559  0.1559   0.1407
##
## sigma^2 estimated as 0.04313:  log likelihood=10.01
## AIC=-10.03   AICc=-8.78   BIC=-0.08
##
## Training set error measures:
##                          ME      RMSE       MAE         MPE     MAPE      MASE
```

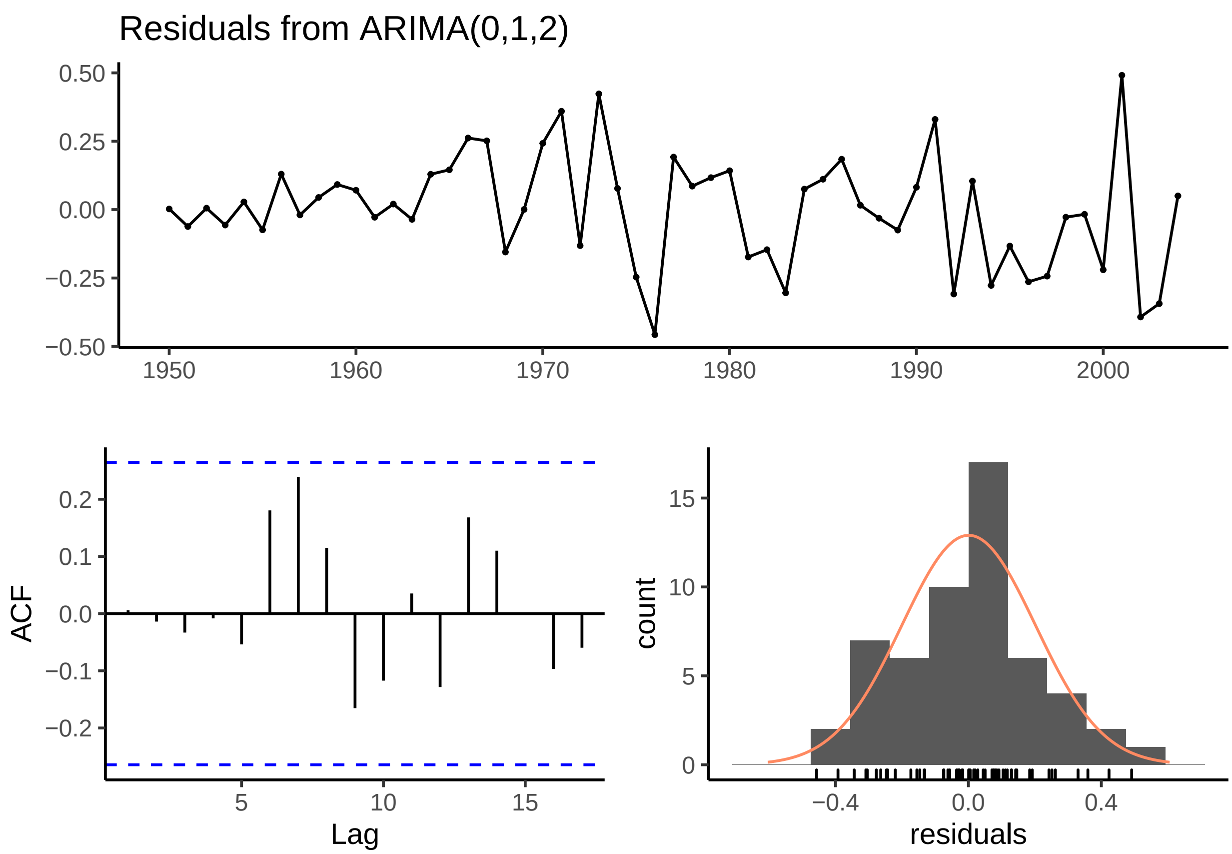
```
## Training set -0.0002754692 0.1980042 0.1514636 -0.1166496 4.357698 0.9314303
##                      ACF1
## Training set 0.01820526
```

```r
fit5 <- Arima(murders$x, order = c(0,1,5))
summary(fit5)
```

```
## Series: murders$x
## ARIMA(0,1,5)
##
## Coefficients:
##           ma1     ma2     ma3      ma4      ma5
##       -0.0490  0.5802  0.1867  -0.1214  -0.3067
## s.e.   0.1346  0.1821  0.1484   0.1238   0.1296
##
## sigma^2 estimated as 0.03864:  log likelihood=11.12
## AIC=-10.25   AICc=-8.46   BIC=1.69
##
## Training set error measures:
##                      ME      RMSE       MAE       MPE     MAPE      MASE
## Training set 0.00135435 0.1855318 0.1360048 -0.059442 3.978605 0.8363659
##                      ACF1
## Training set -0.02008706
```

Since Fit 2 has the lowest AICc and BIC, we choose p,d,q as (0,1,2) A model with one order of differencing assumes that the original series has a constant average trend, hence a constant term should be included. Since our model is of order one, we must include a constant term.

```r
# Checking for residuals
fit <- Arima(murders$x, order = c(0,1,2))
checkresiduals(fit)
```



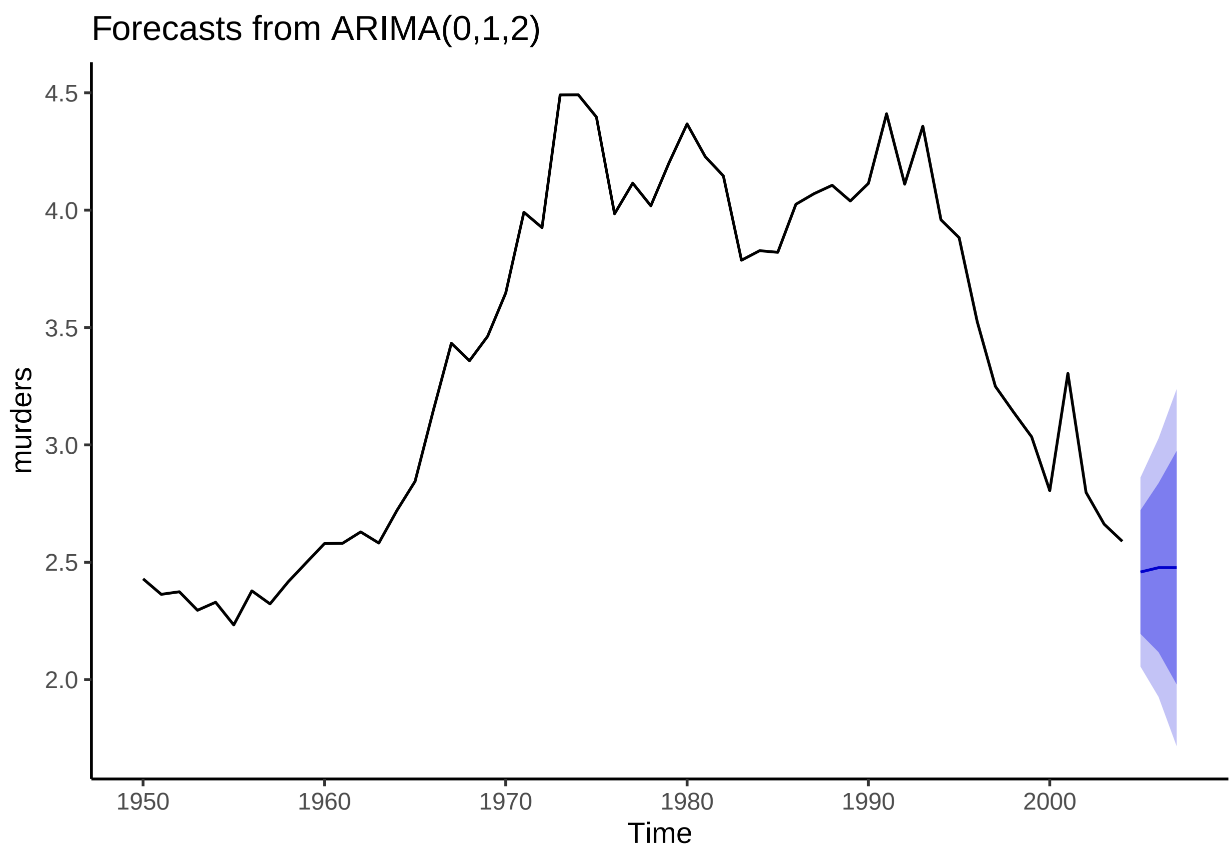

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,2)
## Q* = 9.7748, df = 8, p-value = 0.2812
##
## Model df: 2.   Total lags used: 10
```

The residual plot looks like white noise, which means there is no information left in them, The ACF plot confirms this as there is no Auto-correlation between the residuals.This means that our model is satisfactory.Since the p value is high, we cannot reject null hypothesis which is that the series is white noise.

```
#Generating forecasts
(forecast <- fit %>%
             forecast(h = 3))
```

```
##      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 2005       2.458450 2.195194 2.721707 2.055834 2.861066
## 2006       2.477101 2.116875 2.837327 1.926183 3.028018
## 2007       2.477101 1.979272 2.974929 1.715738 3.238464
```

```
forecast %>% autoplot() + ylab('murders')
```




```r
# let ARIMA() choose the model
auto <- auto.arima(murders$x, d = 1)
summary(auto)
```

```
## Series: murders$x
## ARIMA(0,1,0)
##
## sigma^2 estimated as 0.04563:  log likelihood=6.73
## AIC=-11.46   AICc=-11.38   BIC=-9.47
##
## Training set error measures:
##                       ME      RMSE       MAE         MPE     MAPE      MASE
## Training set 0.00295268 0.2116718 0.1597016 -0.0635393 4.638393 0.9820898
##                     ACF1
## Training set -0.08564966
```

The ARIMA has chosen (0,1,0). Since the model with pre existing p,d,q have lower AICc and BIC, that is a better model than the model that ARIMA has chosen

```r
# Force run all combinations
auto_fit <- auto.arima(murders$x, d = 1, stepwise = FALSE, approximation = FALSE)
summary(auto_fit)
```

```
## Series: murders$x
```

```
## ARIMA(0,1,2)
##
## Coefficients:
##           ma1     ma2
##       -0.0660  0.3712
## s.e.   0.1263  0.1640
##
## sigma^2 estimated as 0.0422:  log likelihood=9.71
## AIC=-13.43   AICc=-12.95   BIC=-7.46
##
## Training set error measures:
##                        ME      RMSE       MAE         MPE     MAPE      MASE
## Training set 0.0007242355 0.1997392 0.1543531 -0.08224024 4.434684 0.9491994
##                     ACF1
## Training set 0.005880608
```

When we force run all the combinations,ARIMA selects (0,1,2), which is the model that we have chosen to be the best model as well.