Time Series Analysis

Aymen Rumi

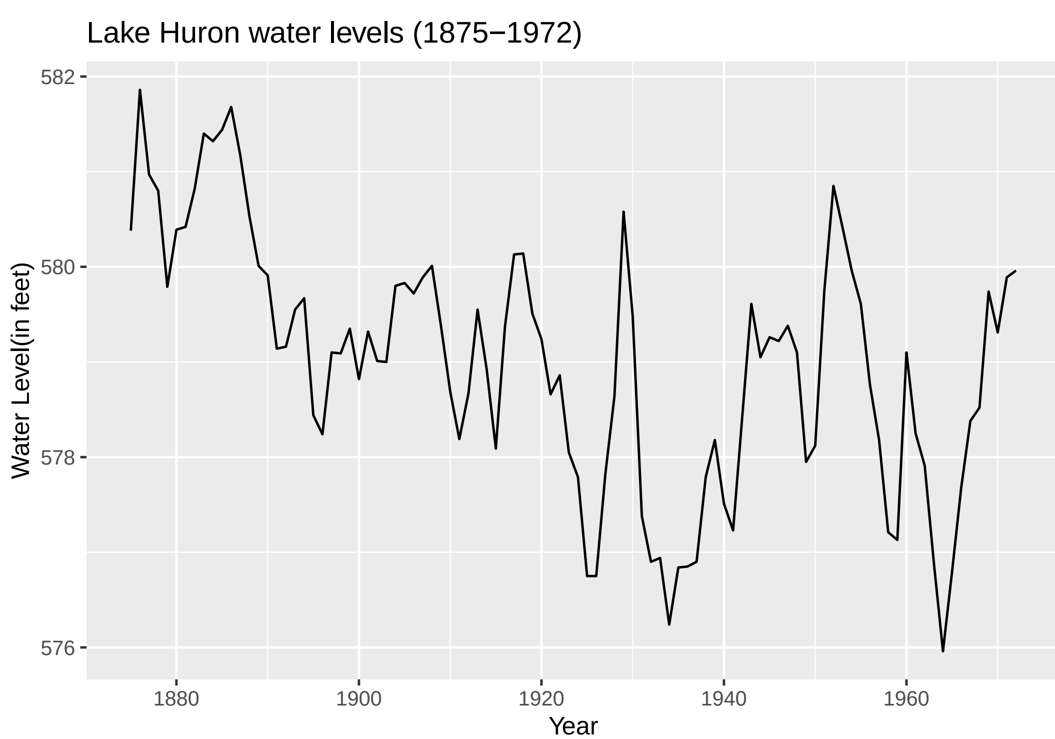
Overview

In this Markdown document we will work with different time series datasets and run various analyses. Areas of focus will include

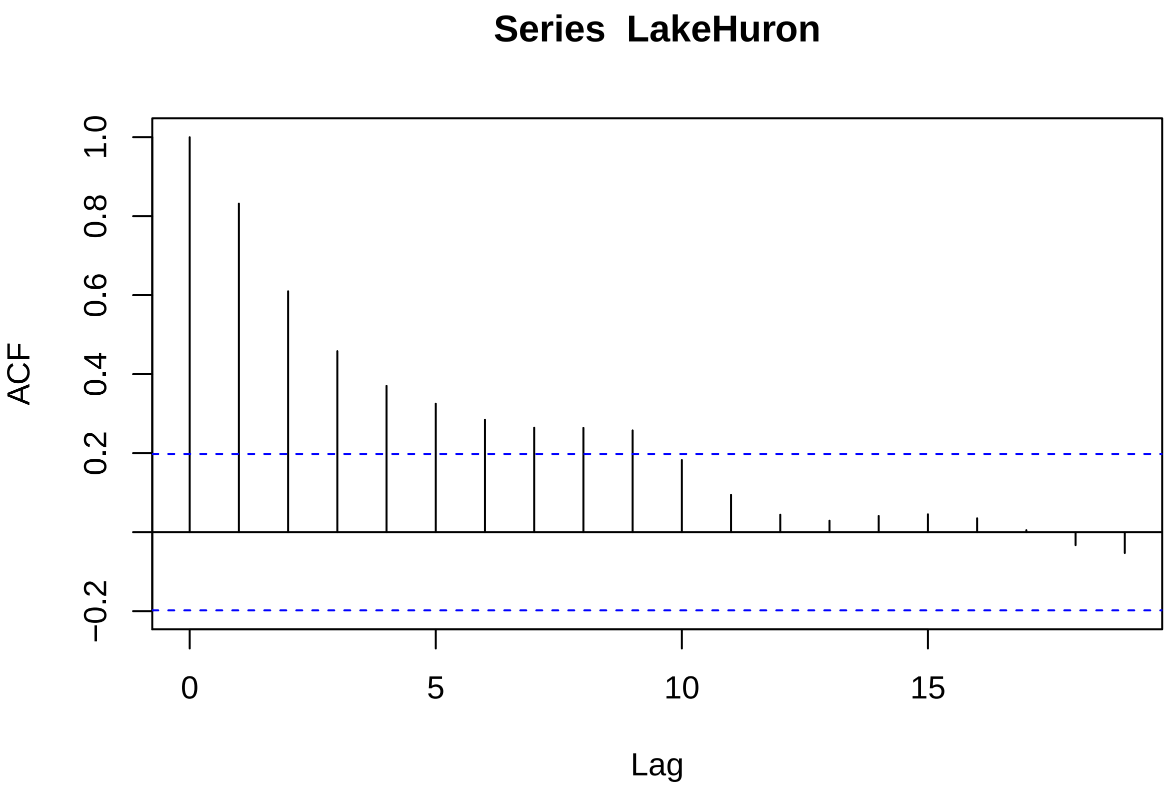
- Decomposition of Time Series Data: "Systematic Pattern" + "Noise"
- Redisual Analysis
- Modelling Time Series Data
- Model Selection Methods
- Time Series Forecasting

Time Series Decomposition

```
# Lake Huron Water Level Data
autoplot(LakeHuron)+ylab("Water Level(in feet)")+
  ggtitle("Lake Huron water levels (1875-1972)")+
  xlab("Year")
```

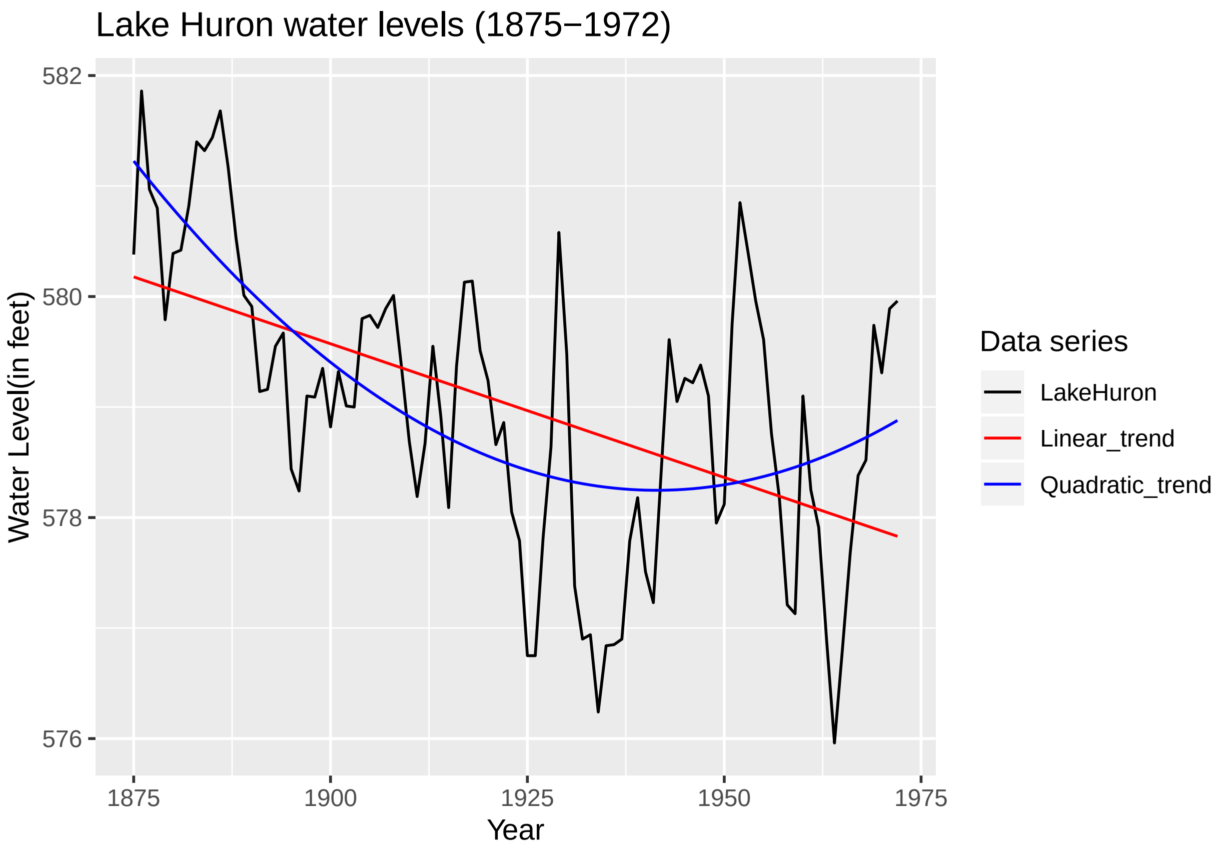


```r
# Autocorrelation Plot
acf(LakeHuron)
```



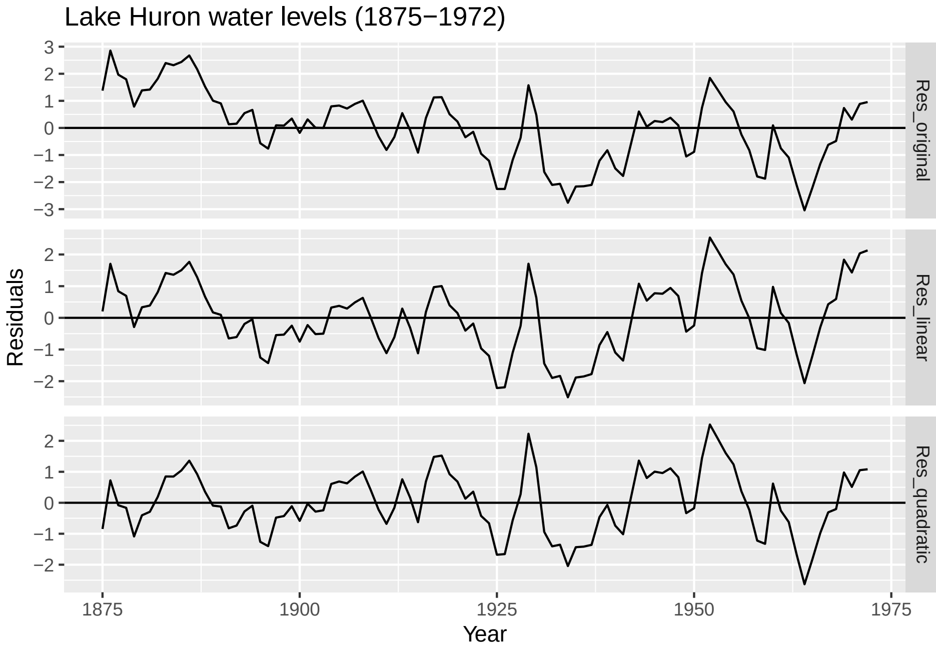
Computing Trend

```r
# Fitting linear trend
linear_trend<-tslm(LakeHuron~trend)
# Fitting quadratic trend
quadratic_trend<-tslm(LakeHuron~trend+I(trend^2))

LakeHuron_with_fits<-cbind(LakeHuron,Linear_trend=fitted(linear_trend),Quadratic_trend=fitted(quadratic_

autoplot(LakeHuron_with_fits)+ylab("Water Level(in feet)")+
  ggtitle("Lake Huron water levels (1875-1972)")+xlab("Year")+
  guides(colour=guide_legend(title="Data series"))+
  scale_colour_manual(values=c("black","red","blue"))
```

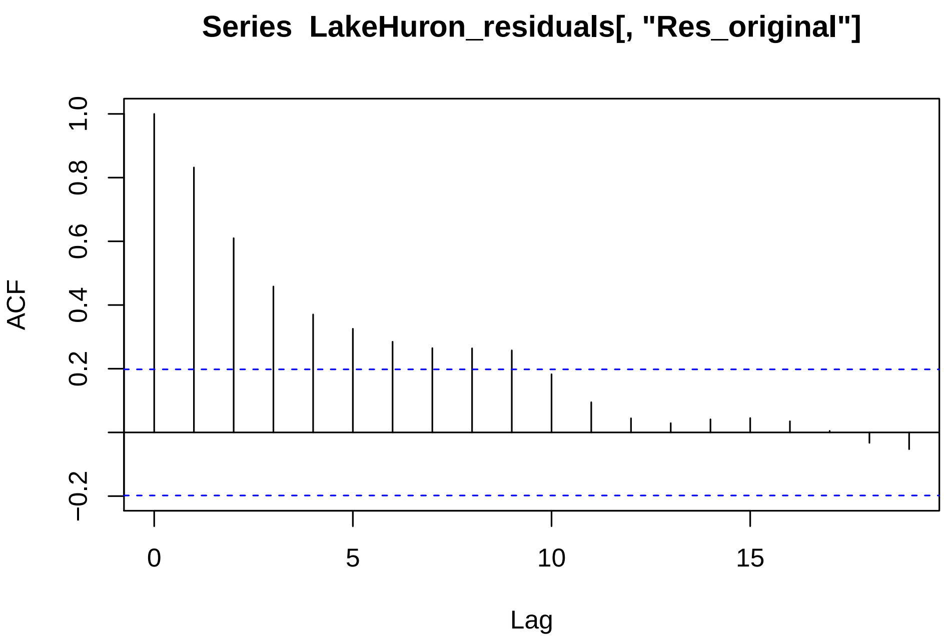


Residual Analysis

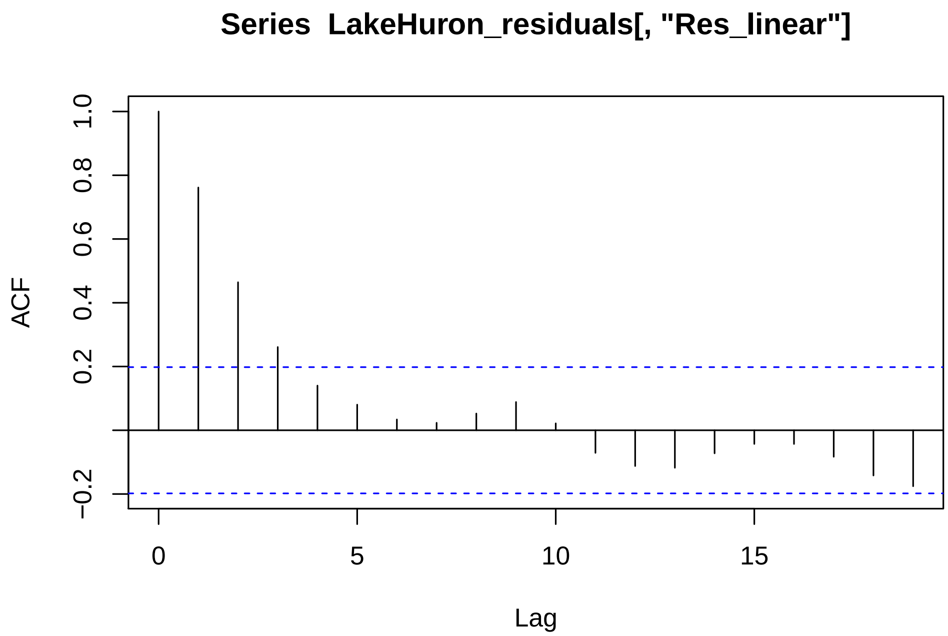
```
LakeHuron_residuals<-cbind(Res_original=LakeHuron-mean(LakeHuron),
                           Res_linear=LakeHuron-fitted(linear_trend),
                           Res_quadratic=LakeHuron-fitted(quadratic_trend))

autoplot(LakeHuron_residuals,facet=TRUE)+
  ylab("Residuals")+
  ggtitle("Lake Huron water levels (1875-1972)")+
  xlab("Year")+geom_hline(yintercept = 0)+guides(colour=guide_legend(title="Data series"))+scale_colour_
```



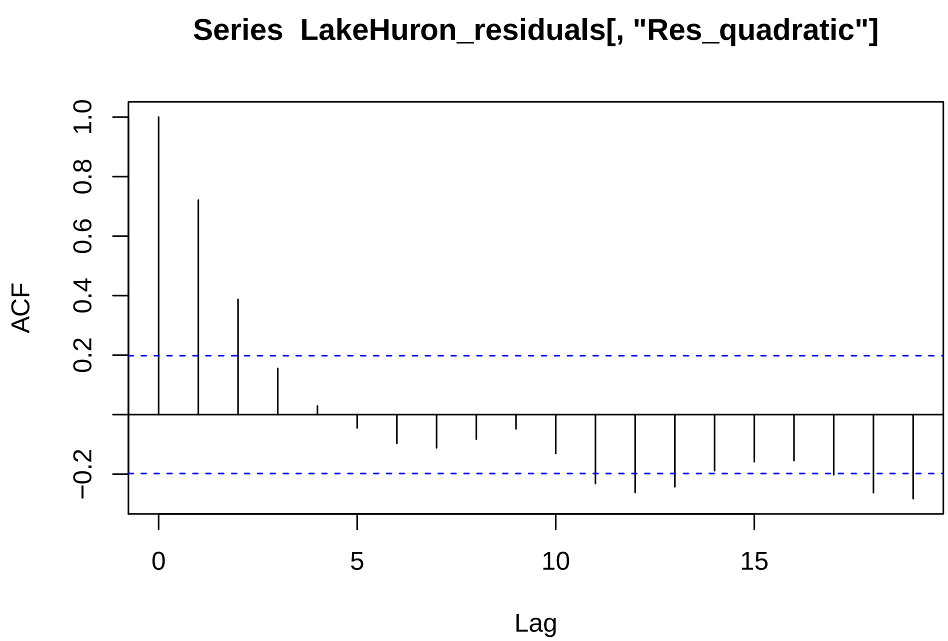

```r
acf(LakeHuron_residuals[,"Res_original"])
```




```r
acf(LakeHuron_residuals[,"Res_linear"])
```

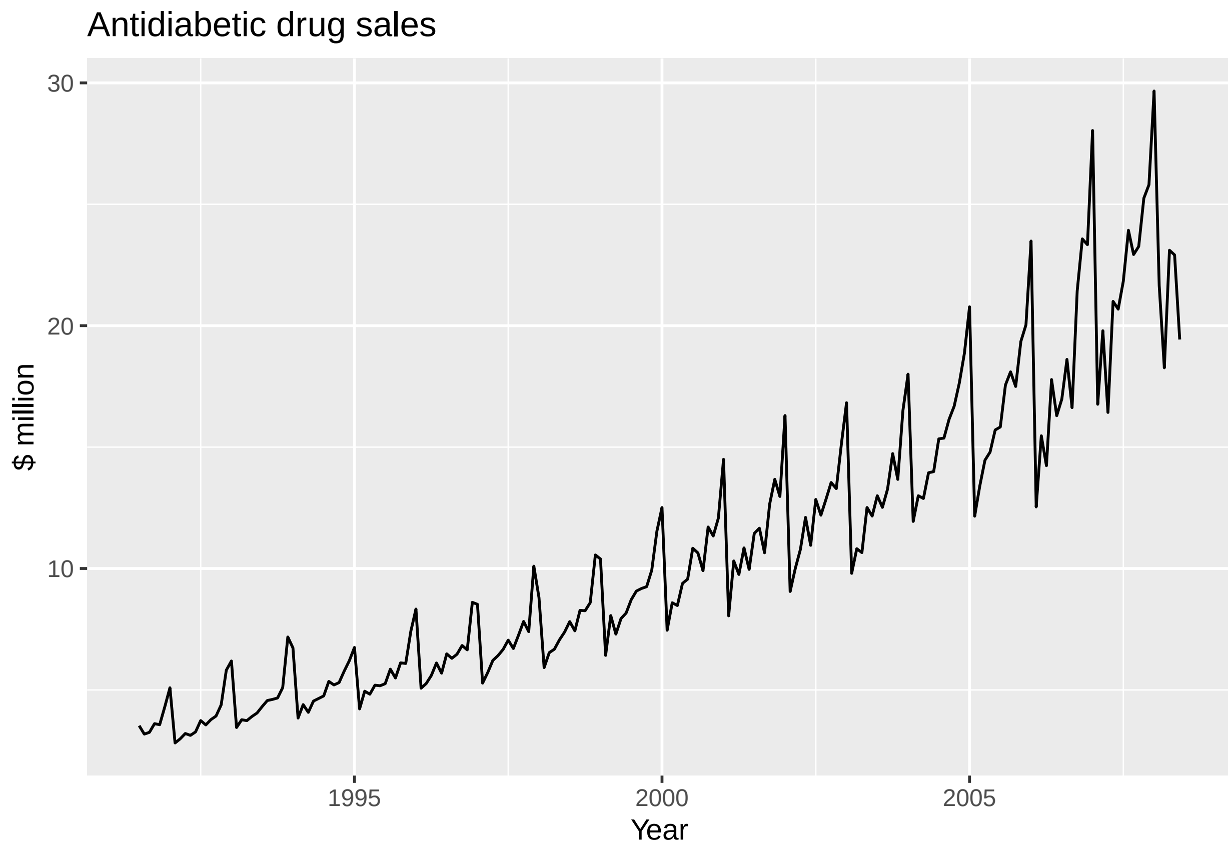



```
acf(LakeHuron_residuals[,"Res_quadratic"])
```

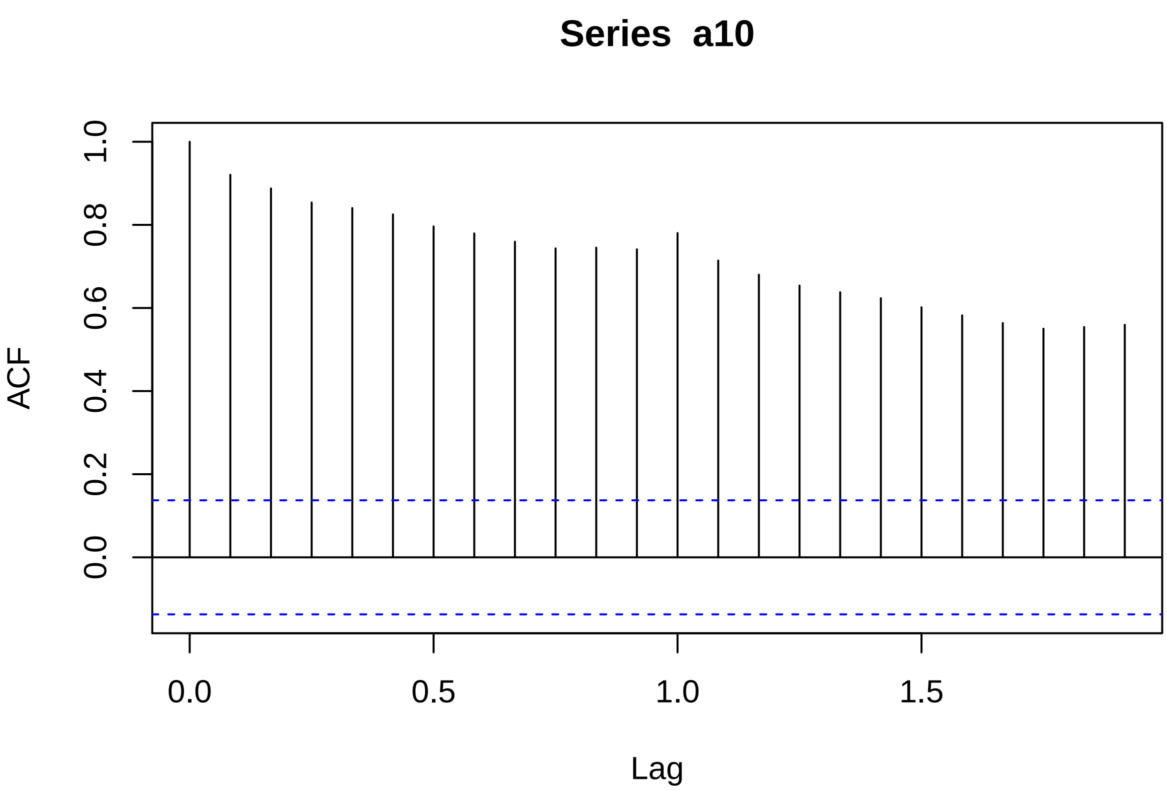


Computing Trend + Seasonality

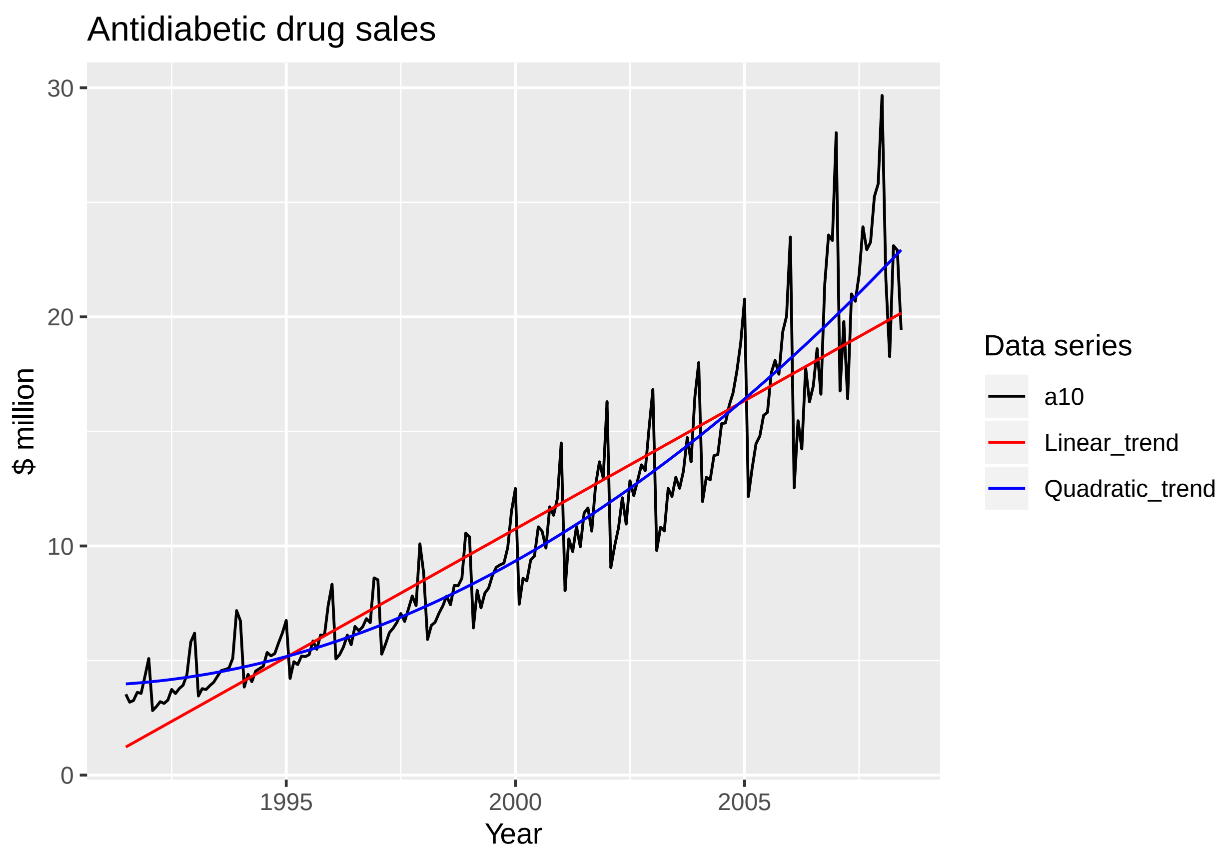
```
# Antidiabetic Drug Sale Data
autoplot(a10)+ggtitle("Antidiabetic drug sales")+ylab("$ million")+xlab("Year")
```


Antidiabetic drug sales
30
20
10
$ million
1995
2000
2005
Year

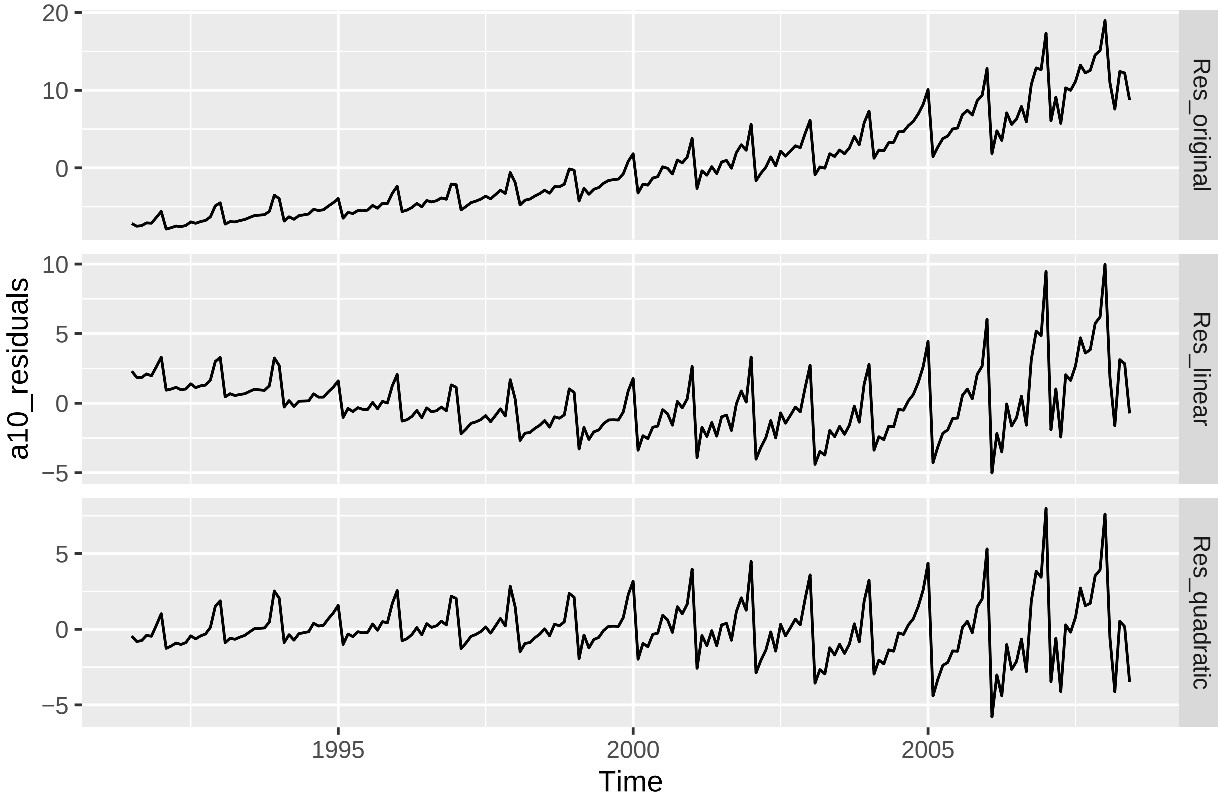
```r
acf(a10)
```


Series a10
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
Lag

```r
# Fitting & Removing Trend
linear_trend<-tslm(a10~trend)
quadratic_trend<-tslm(a10~trend+I(trend^2))

a10_with_fits<-cbind(a10,Linear_trend=fitted(linear_trend),Quadratic_trend=fitted(quadratic_trend))

autoplot(a10_with_fits)+ylab("$ million")+
  ggtitle("Antidiabetic drug sales")+xlab("Year")+
  guides(colour=guide_legend(title="Data series"))+
  scale_colour_manual(values=c("black","red","blue"))
```




```r
a10_residuals<-cbind(Res_original=a10-mean(a10),
                     Res_linear=a10-fitted(linear_trend),
                     Res_quadratic=a10-fitted(quadratic_trend))


autoplot(a10_residuals,facet=TRUE)
```

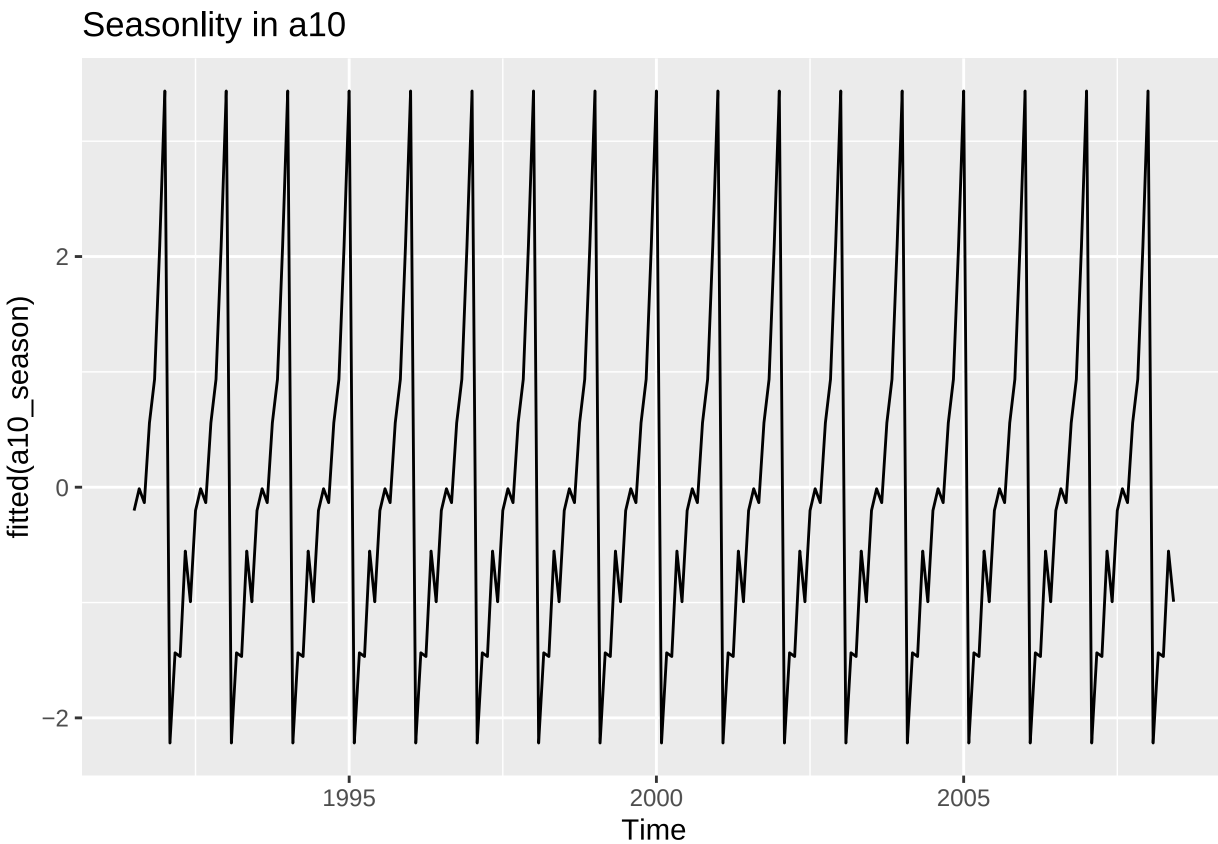


Estimating Seasonality

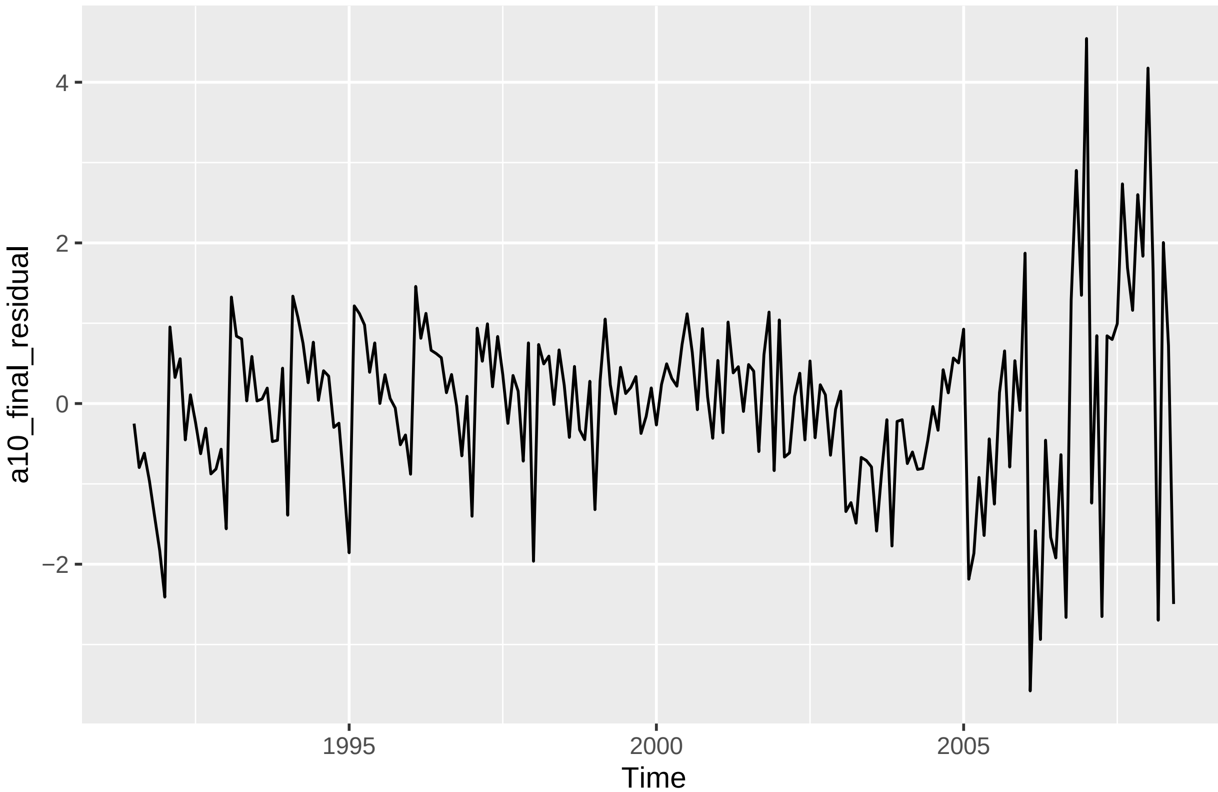
```
# We will estimate & remove seasonility from time series data withquadrtic  trend removed

frequency(a10_residuals[,"Res_quadratic"])
```

```
## [1] 12
```

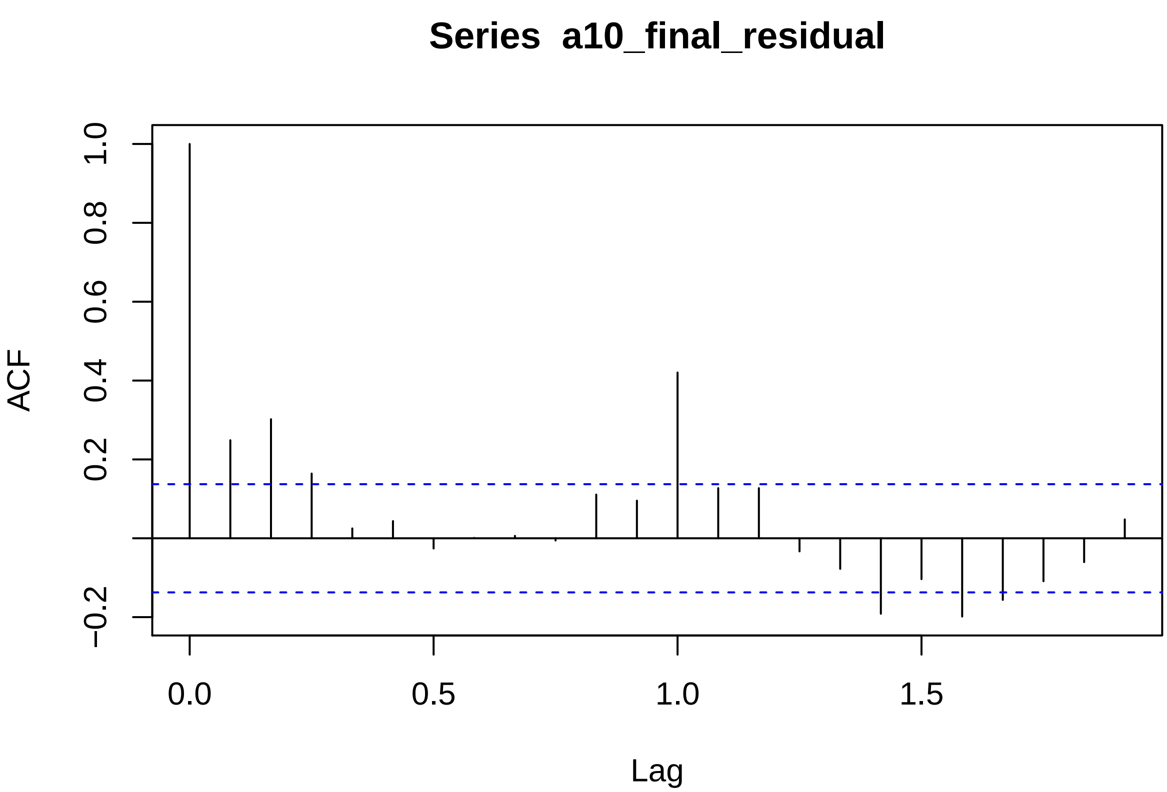
```
a10_season<-tslm(Res_quadratic~season,data=a10_residuals)
autoplot(fitted(a10_season))+ggtitle("Seasonlity in a10")
```




```r
# Removing Trend & Sesonality
a10_final_residual<-cbind("Residual"=a10-fitted(quadratic_trend)-fitted(a10_season))
autoplot(a10_final_residual)
```



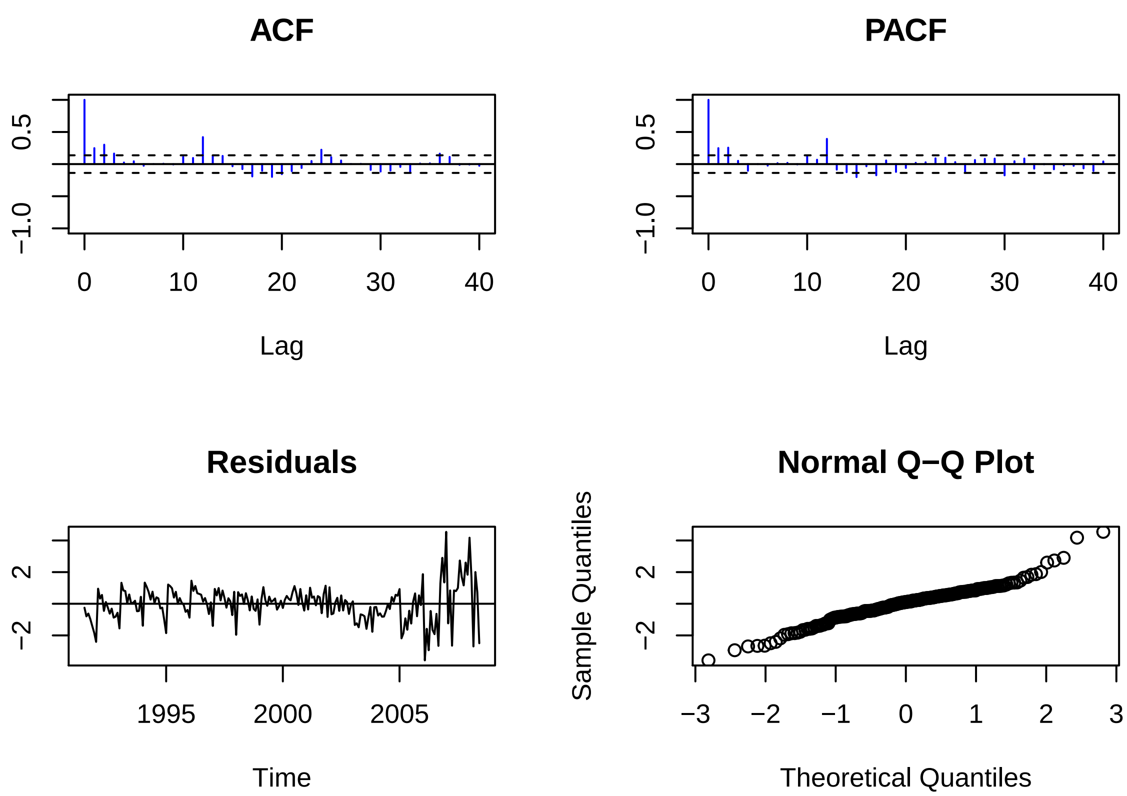

```r
acf(a10_final_residual)
```



Testing Residuals

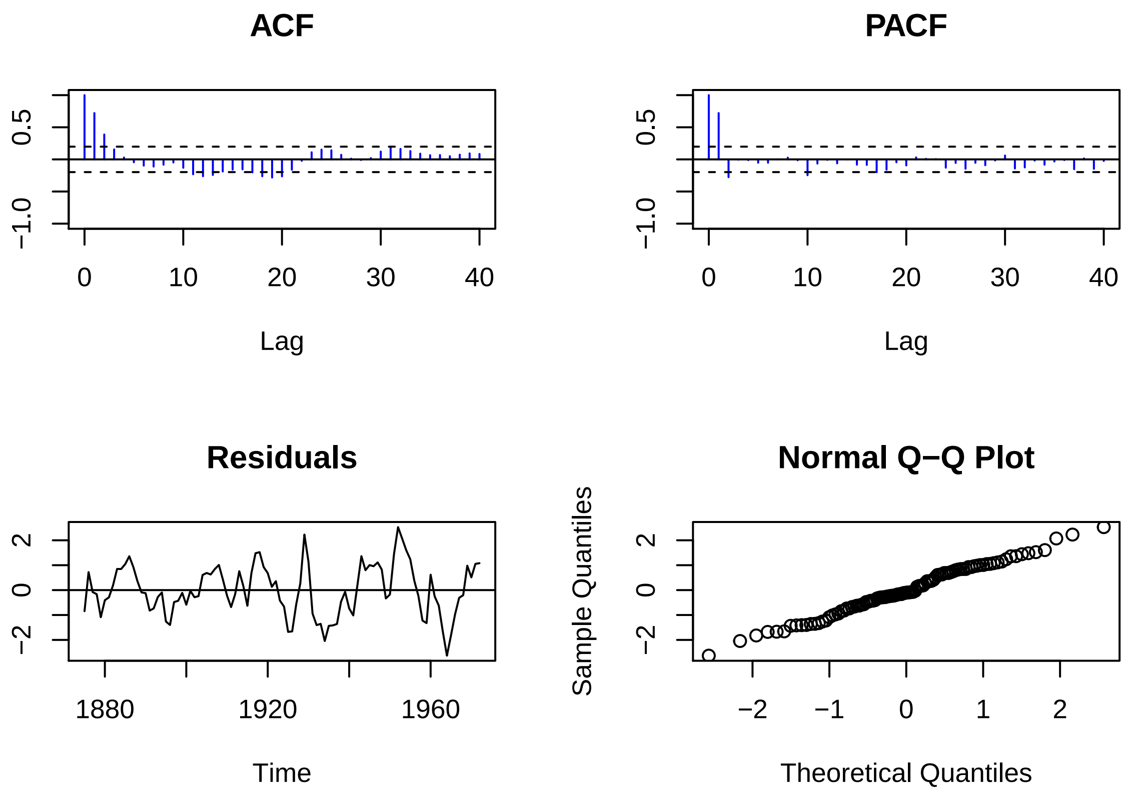
```r
test(a10_final_residual)
```

```
## Null hypothesis: Residuals are iid noise.
## Test                        Distribution Statistic   p-value
## Ljung-Box Q                 Q ~ chisq(20)    115.35         0 *
## McLeod-Li Q                 Q ~ chisq(20)    256.69         0 *
## Turning points T      (T-134.7)/6 ~ N(0,1)       148    0.0262 *
## Diff signs S        (S-101.5)/4.1 ~ N(0,1)        94    0.0696
## Rank P            (P-10353)/487.4 ~ N(0,1)     10210    0.7692
```



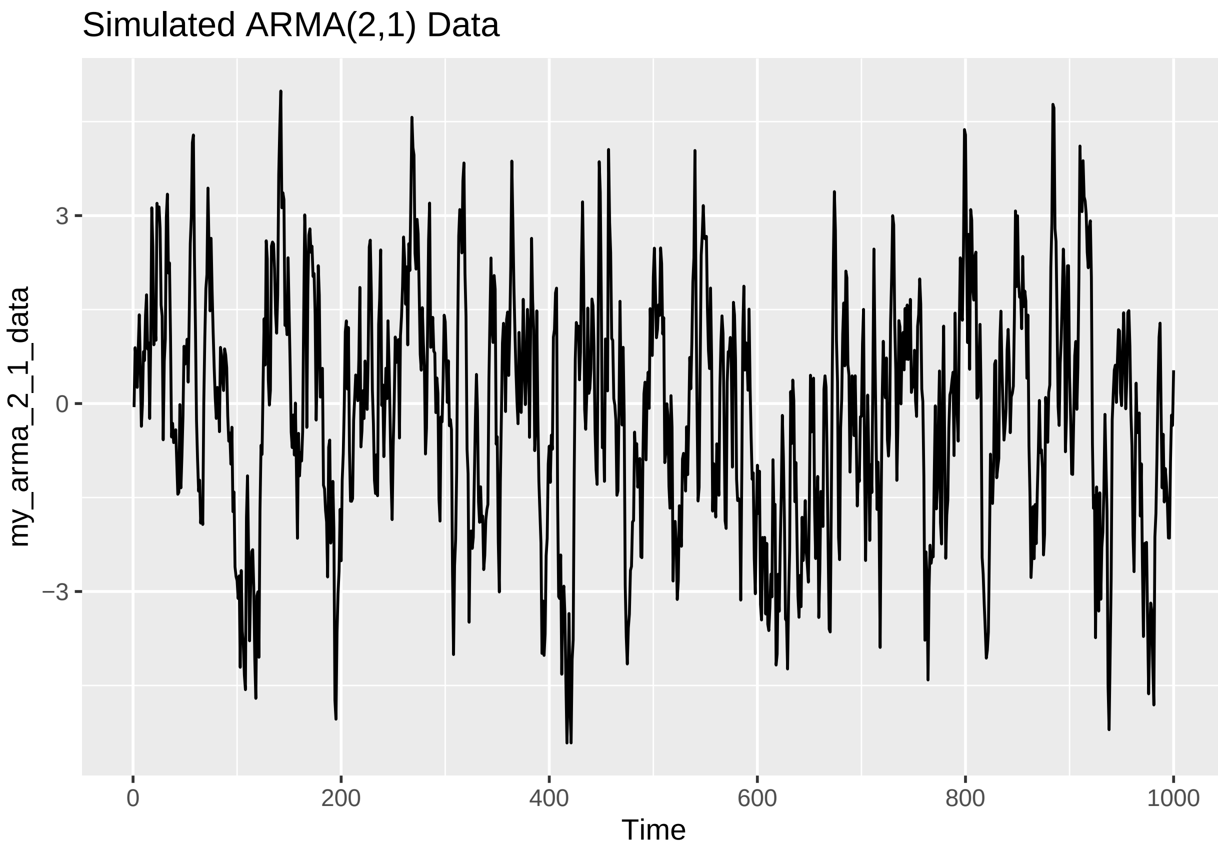

```r
test(LakeHuron_residuals[,"Res_quadratic"])
```

```
## Null hypothesis: Residuals are iid noise.
## Test                        Distribution Statistic   p-value
## Ljung-Box Q                 Q ~ chisq(20)    138.67         0 *
## McLeod-Li Q                 Q ~ chisq(20)     56.45         0 *
## Turning points T        (T-64)/4.1 ~ N(0,1)        40         0 *
## Diff signs S          (S-48.5)/2.9 ~ N(0,1)        50    0.6015
## Rank P           (P-2376.5)/162.9 ~ N(0,1)      2406    0.8563
```

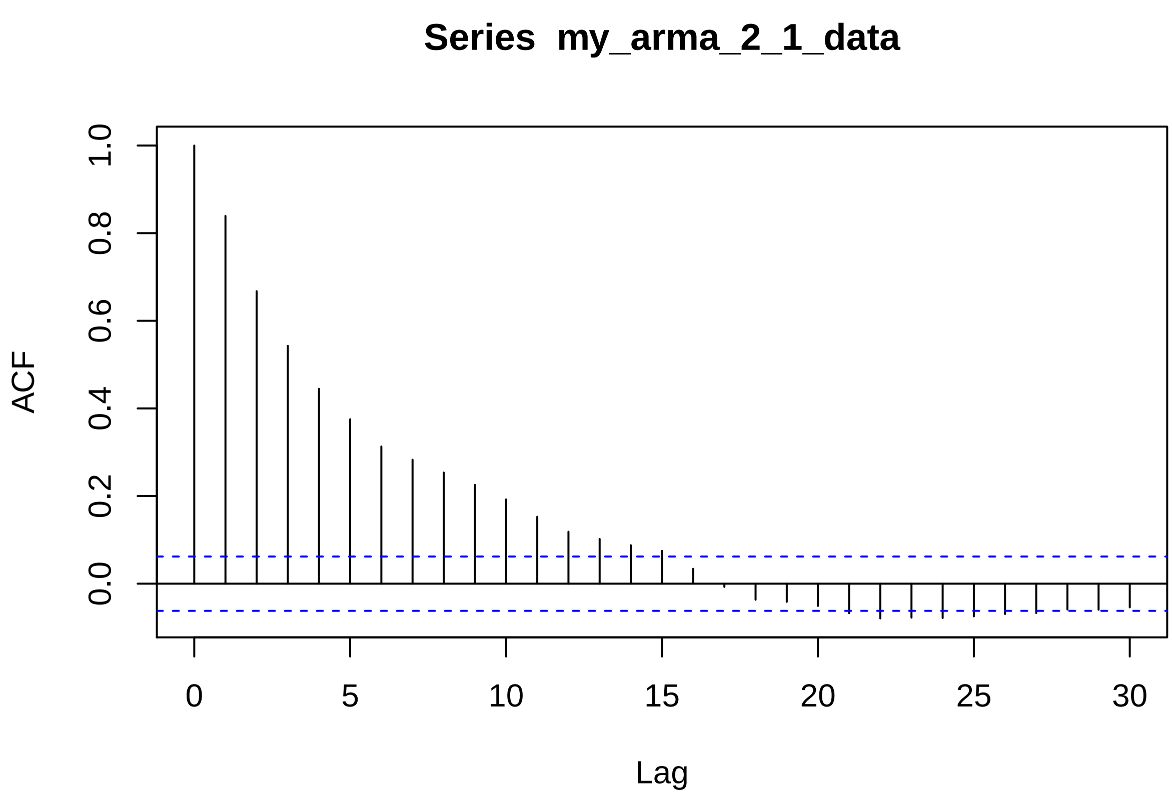


Modeling Time Series Data

```
set.seed(1)
#Simulated ARMA(2,1) Model

true_ar_coef=c(0.6,0.2)
true_ma_coef=c(0.4)
my_ARMA_2_1_model=list(ar=true_ar_coef,ma=true_ma_coef)

my_arma_2_1_data=arima.sim(my_ARMA_2_1_model,n=1000)

autoplot(my_arma_2_1_data)+ggtitle("Simulated ARMA(2,1) Data")
```




```
acf(my_arma_2_1_data)
```



Estimating ARMA coefficients

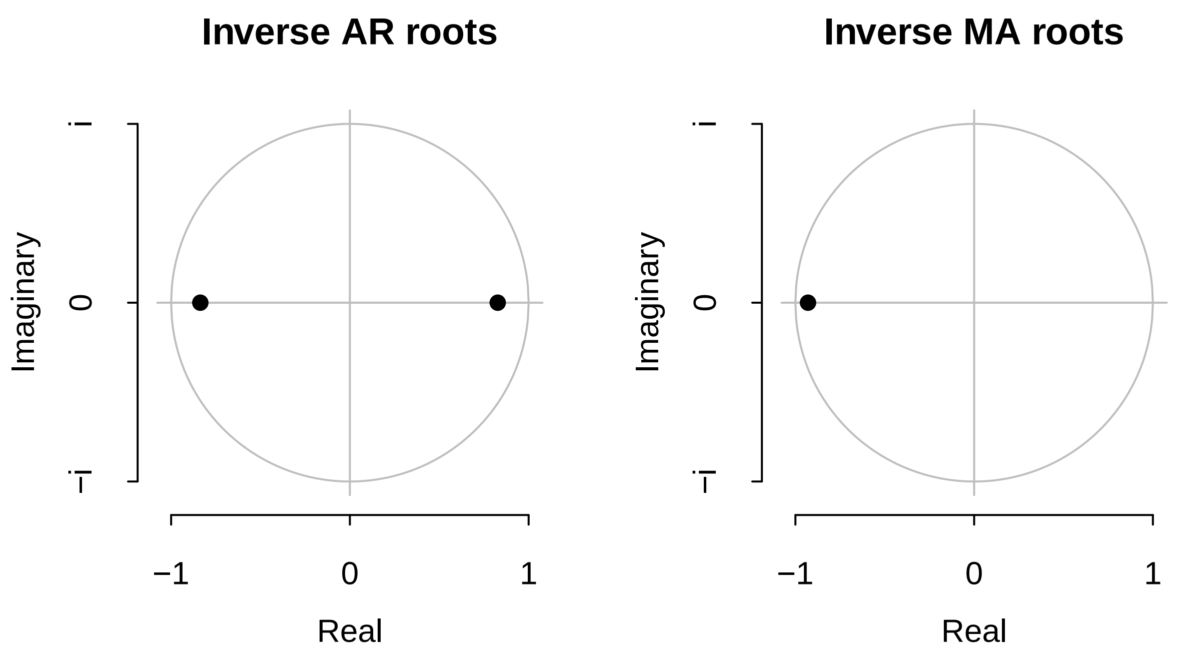
```r
hannan(my_arma_2_1_data-mean(my_arma_2_1_data),p=2,q=1)
```

```
## $phi
## [1] 0.6367820 0.1316353
##
## $theta
## [1] 0.3266316
##
## $sigma2
## [1] 1.071797
##
## $aicc
## [1] 2916.528
##
## $se.phi
## [1] 0.1380438 0.1171178
##
## $se.theta
## [1] 0.1420062
```

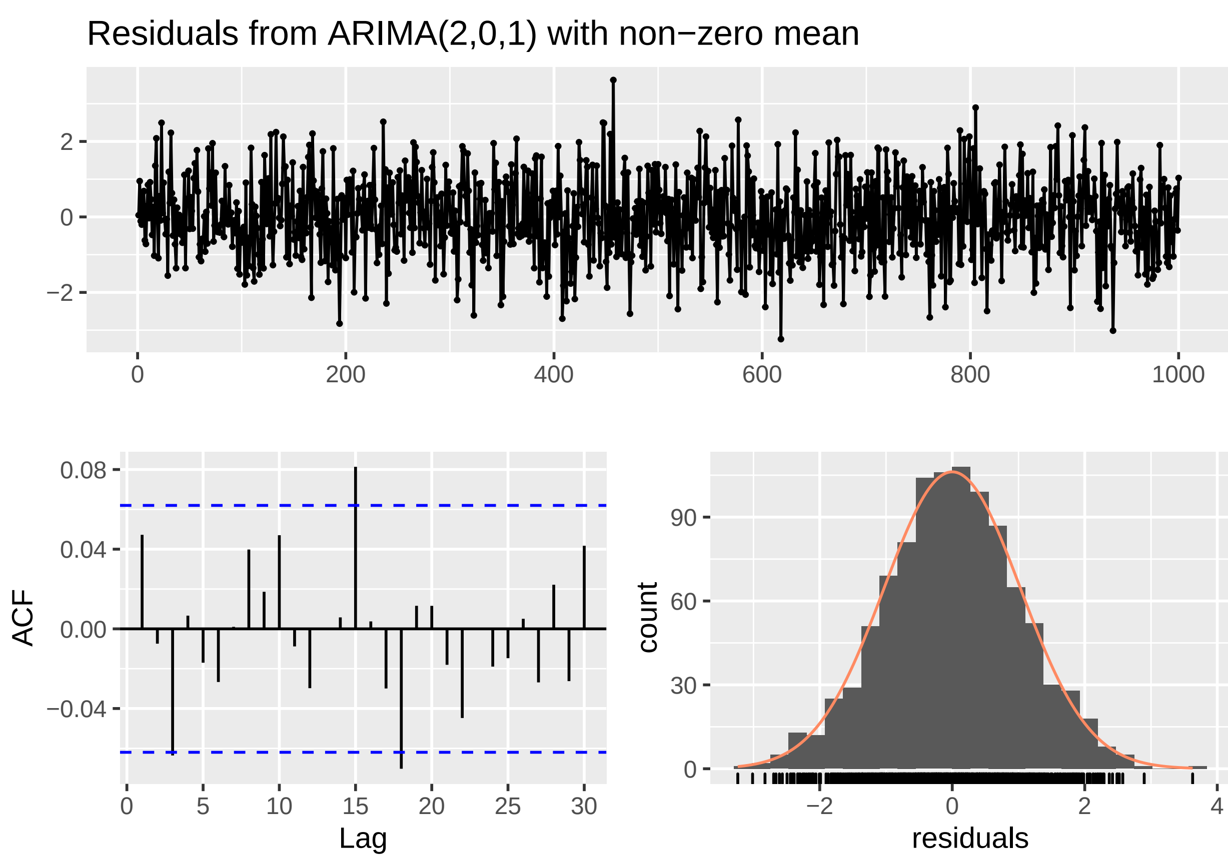
```r
summary(Arima(my_arma_2_1_data,order=c(2,0,1)))
```

```
## Series: my_arma_2_1_data
## ARIMA(2,0,1) with non-zero mean
##
## Coefficients:
##           ar1     ar2     ma1     mean
##       -0.0095  0.6921  0.9295  -0.1420
## s.e.   0.0491  0.0461  0.0340   0.1973
##
## sigma^2 estimated as 1.068:  log likelihood=-1450.37
## AIC=2910.75   AICc=2910.81   BIC=2935.29
##
## Training set error measures:
##                         ME     RMSE       MAE      MPE     MAPE     MASE
## Training set -0.0001669132 1.031195 0.8214051 44.66213 197.4916 0.941992
##                    ACF1
## Training set 0.04721335
```

```r
plot((Arima(my_arma_2_1_data,order=c(2,0,1))))
```




```r
checkresiduals((Arima(my_arma_2_1_data,order=c(2,0,1))))
```

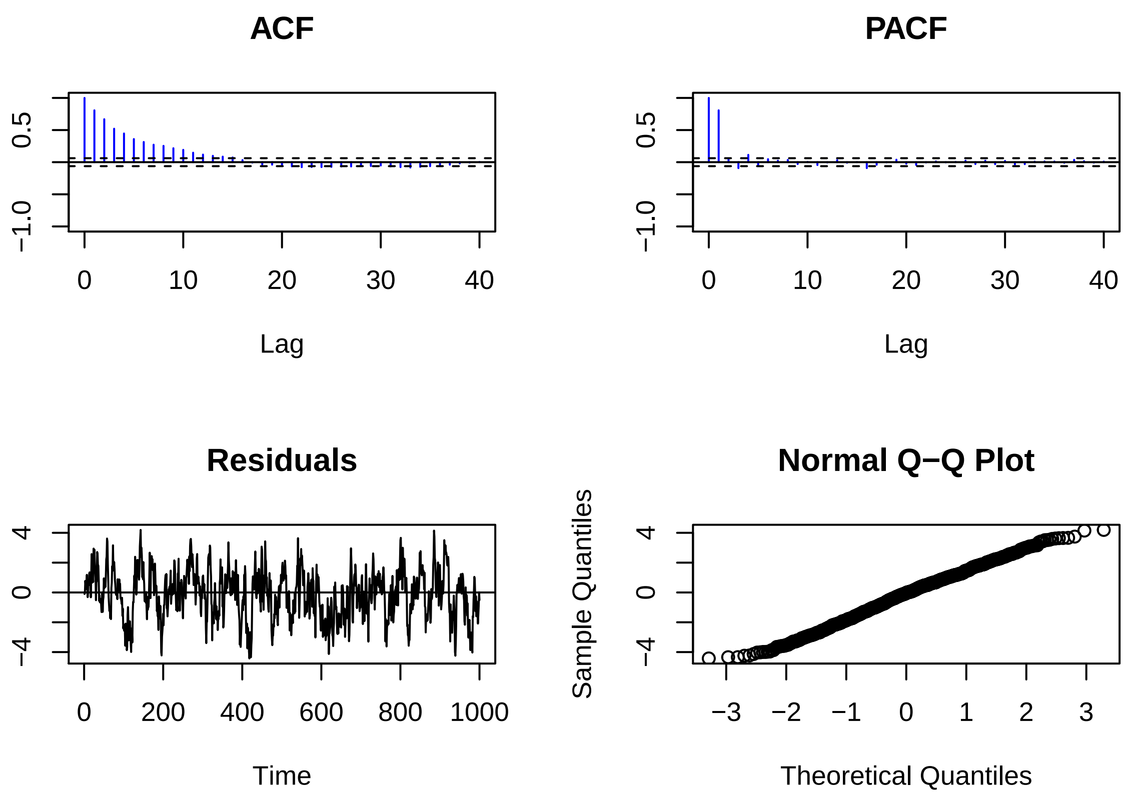



```
##
##  Ljung-Box test
##
```

```
## data:  Residuals from ARIMA(2,0,1) with non-zero mean
## Q* = 11.596, df = 6, p-value = 0.07161
##
## Model df: 4.   Total lags used: 10
```

```
test(fitted(Arima(my_arma_2_1_data,order=c(2,0,1))))
```

```
## Null hypothesis: Residuals are iid noise.
## Test                          Distribution Statistic   p-value
## Ljung-Box Q                   Q ~ chisq(20)   2096.65         0 *
## McLeod-Li Q                   Q ~ chisq(20)    667.85         0 *
## Turning points T (T-665.3)/13.3 ~ N(0,1)          543         0 *
## Diff signs S       (S-499.5)/9.1 ~ N(0,1)         500    0.9563
## Rank P         (P-249750)/5274.4 ~ N(0,1)      230054    2e-04 *
```

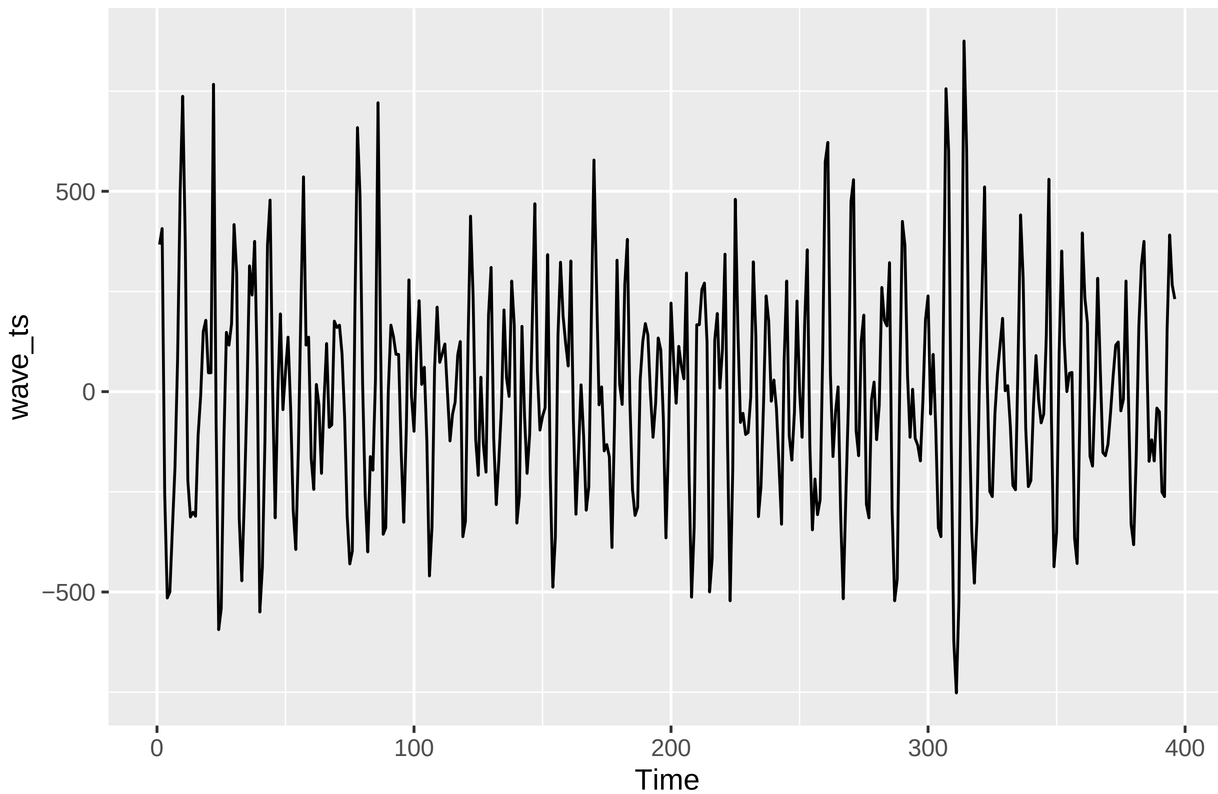


Model Selection Methods

```
wave_data=read_csv("wavedat_new.csv")
```

```
## Parsed with column specification:
## cols(
##   waveht = col_double()
## )
```

```
wave_ts=wave_data%>%pull(waveht)%>%ts(.,frequency = 1)
autoplot(wave_ts)
```




```
wave_auto_select_AIC=auto.arima(wave_ts,stepwise=FALSE,seasonal=FALSE,ic="aic",trace=TRUE)
```

```
##
##  Fitting models using approximations to speed things up...
##
##  ARIMA(0,0,0) with zero mean     : 5548.674
##  ARIMA(0,0,0) with non-zero mean : 5550.582
##  ARIMA(0,0,1) with zero mean     : 5334.636
##  ARIMA(0,0,1) with non-zero mean : 5336.596
##  ARIMA(0,0,2) with zero mean     : 5316.739
##  ARIMA(0,0,2) with non-zero mean : 5318.729
##  ARIMA(0,0,3) with zero mean     : 5218.578
##  ARIMA(0,0,3) with non-zero mean : Inf
##  ARIMA(0,0,4) with zero mean     : 5185.9
##  ARIMA(0,0,4) with non-zero mean : Inf
##  ARIMA(0,0,5) with zero mean     : 5187.523
##  ARIMA(0,0,5) with non-zero mean : Inf
##  ARIMA(1,0,0) with zero mean     : 5449.941
##  ARIMA(1,0,0) with non-zero mean : 5451.884
##  ARIMA(1,0,1) with zero mean     : 5333.608
##  ARIMA(1,0,1) with non-zero mean : 5335.598
##  ARIMA(1,0,2) with zero mean     : 5326.179
##  ARIMA(1,0,2) with non-zero mean : 5328.157
##  ARIMA(1,0,3) with zero mean     : 5212.332
```

```
##  ARIMA(1,0,3) with non-zero mean : 5202.621
##  ARIMA(1,0,4) with zero mean     : 5202.095
##  ARIMA(1,0,4) with non-zero mean : 5189.405
##  ARIMA(2,0,0) with zero mean     : 5257.361
##  ARIMA(2,0,0) with non-zero mean : 5259.053
##  ARIMA(2,0,1) with zero mean     : 5203.155
##  ARIMA(2,0,1) with non-zero mean : 5195.208
##  ARIMA(2,0,2) with zero mean     : 5171.268
##  ARIMA(2,0,2) with non-zero mean : 5163.619
##  ARIMA(2,0,3) with zero mean     : 5156.341
##  ARIMA(2,0,3) with non-zero mean : 5151.446
##  ARIMA(3,0,0) with zero mean     : 5256.836
##  ARIMA(3,0,0) with non-zero mean : 5258.614
##  ARIMA(3,0,1) with zero mean     : 5256.42
##  ARIMA(3,0,1) with non-zero mean : 5258.171
##  ARIMA(3,0,2) with zero mean     : 5141.022
##  ARIMA(3,0,2) with non-zero mean : Inf
##  ARIMA(4,0,0) with zero mean     : 5213.393
##  ARIMA(4,0,0) with non-zero mean : 5214.895
##  ARIMA(4,0,1) with zero mean     : 5111.274
##  ARIMA(4,0,1) with non-zero mean : 5090.584
##  ARIMA(5,0,0) with zero mean     : 5178.428
##  ARIMA(5,0,0) with non-zero mean : 5179.615
##
##  Now re-fitting the best model(s) without approximations...
##
##
##
##
##  Best model: ARIMA(4,0,1) with non-zero mean
```

```r
wave_auto_select_AIC=auto.arima(wave_ts,stepwise=FALSE,seasonal=FALSE,ic="bic",trace=TRUE)
```

```
##
##  Fitting models using approximations to speed things up...
##
##  ARIMA(0,0,0) with zero mean     : 5552.656
##  ARIMA(0,0,0) with non-zero mean : 5558.544
##  ARIMA(0,0,1) with zero mean     : 5342.599
##  ARIMA(0,0,1) with non-zero mean : 5348.54
##  ARIMA(0,0,2) with zero mean     : 5328.684
##  ARIMA(0,0,2) with non-zero mean : 5334.655
##  ARIMA(0,0,3) with zero mean     : 5234.503
##  ARIMA(0,0,3) with non-zero mean : Inf
##  ARIMA(0,0,4) with zero mean     : 5205.807
##  ARIMA(0,0,4) with non-zero mean : Inf
##  ARIMA(0,0,5) with zero mean     : 5211.412
##  ARIMA(0,0,5) with non-zero mean : Inf
##  ARIMA(1,0,0) with zero mean     : 5457.903
##  ARIMA(1,0,0) with non-zero mean : 5463.828
##  ARIMA(1,0,1) with zero mean     : 5345.552
##  ARIMA(1,0,1) with non-zero mean : 5351.524
##  ARIMA(1,0,2) with zero mean     : 5342.104
##  ARIMA(1,0,2) with non-zero mean : 5348.064
```

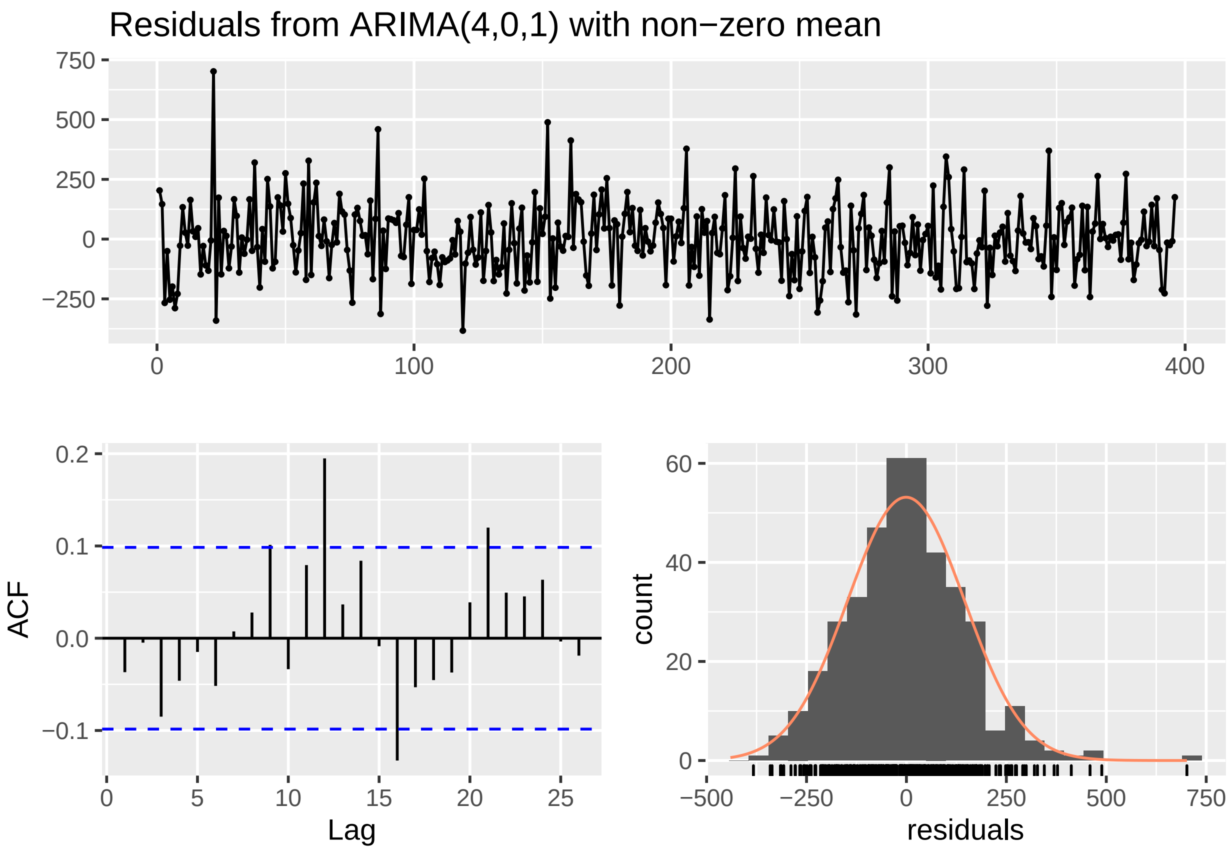
```
##  ARIMA(1,0,3) with zero mean     : 5232.239
##  ARIMA(1,0,3) with non-zero mean : 5226.509
##  ARIMA(1,0,4) with zero mean     : 5225.984
##  ARIMA(1,0,4) with non-zero mean : 5217.275
##  ARIMA(2,0,0) with zero mean     : 5269.305
##  ARIMA(2,0,0) with non-zero mean : 5274.978
##  ARIMA(2,0,1) with zero mean     : 5219.081
##  ARIMA(2,0,1) with non-zero mean : 5215.115
##  ARIMA(2,0,2) with zero mean     : 5191.175
##  ARIMA(2,0,2) with non-zero mean : 5187.507
##  ARIMA(2,0,3) with zero mean     : 5180.23
##  ARIMA(2,0,3) with non-zero mean : 5179.316
##  ARIMA(3,0,0) with zero mean     : 5272.762
##  ARIMA(3,0,0) with non-zero mean : 5278.521
##  ARIMA(3,0,1) with zero mean     : 5276.327
##  ARIMA(3,0,1) with non-zero mean : 5282.059
##  ARIMA(3,0,2) with zero mean     : 5164.91
##  ARIMA(3,0,2) with non-zero mean : Inf
##  ARIMA(4,0,0) with zero mean     : 5233.3
##  ARIMA(4,0,0) with non-zero mean : 5238.784
##  ARIMA(4,0,1) with zero mean     : 5135.163
##  ARIMA(4,0,1) with non-zero mean : 5118.454
##  ARIMA(5,0,0) with zero mean     : 5202.317
##  ARIMA(5,0,0) with non-zero mean : 5207.485
##
##  Now re-fitting the best model(s) without approximations...
##
##
##
##
##  Best model: ARIMA(4,0,1) with non-zero mean
```

```r
wave_auto_select_AIC=auto.arima(wave_ts,stepwise=FALSE,seasonal=FALSE,ic="aicc",trace=TRUE)
```

```
##
##  Fitting models using approximations to speed things up...
##
##  ARIMA(0,0,0) with zero mean     : 5548.684
##  ARIMA(0,0,0) with non-zero mean : 5550.612
##  ARIMA(0,0,1) with zero mean     : 5334.667
##  ARIMA(0,0,1) with non-zero mean : 5336.657
##  ARIMA(0,0,2) with zero mean     : 5316.801
##  ARIMA(0,0,2) with non-zero mean : 5318.831
##  ARIMA(0,0,3) with zero mean     : 5218.68
##  ARIMA(0,0,3) with non-zero mean : Inf
##  ARIMA(0,0,4) with zero mean     : 5186.054
##  ARIMA(0,0,4) with non-zero mean : Inf
##  ARIMA(0,0,5) with zero mean     : 5187.739
##  ARIMA(0,0,5) with non-zero mean : Inf
##  ARIMA(1,0,0) with zero mean     : 5449.971
##  ARIMA(1,0,0) with non-zero mean : 5451.945
##  ARIMA(1,0,1) with zero mean     : 5333.669
##  ARIMA(1,0,1) with non-zero mean : 5335.701
##  ARIMA(1,0,2) with zero mean     : 5326.281
```

```
##  ARIMA(1,0,2) with non-zero mean : 5328.311
##  ARIMA(1,0,3) with zero mean     : 5212.486
##  ARIMA(1,0,3) with non-zero mean : 5202.837
##  ARIMA(1,0,4) with zero mean     : 5202.311
##  ARIMA(1,0,4) with non-zero mean : 5189.694
##  ARIMA(2,0,0) with zero mean     : 5257.422
##  ARIMA(2,0,0) with non-zero mean : 5259.155
##  ARIMA(2,0,1) with zero mean     : 5203.257
##  ARIMA(2,0,1) with non-zero mean : 5195.362
##  ARIMA(2,0,2) with zero mean     : 5171.422
##  ARIMA(2,0,2) with non-zero mean : 5163.835
##  ARIMA(2,0,3) with zero mean     : 5156.557
##  ARIMA(2,0,3) with non-zero mean : 5151.735
##  ARIMA(3,0,0) with zero mean     : 5256.938
##  ARIMA(3,0,0) with non-zero mean : 5258.768
##  ARIMA(3,0,1) with zero mean     : 5256.574
##  ARIMA(3,0,1) with non-zero mean : 5258.387
##  ARIMA(3,0,2) with zero mean     : 5141.238
##  ARIMA(3,0,2) with non-zero mean : Inf
##  ARIMA(4,0,0) with zero mean     : 5213.547
##  ARIMA(4,0,0) with non-zero mean : 5215.111
##  ARIMA(4,0,1) with zero mean     : 5111.49
##  ARIMA(4,0,1) with non-zero mean : 5090.873
##  ARIMA(5,0,0) with zero mean     : 5178.644
##  ARIMA(5,0,0) with non-zero mean : 5179.903
##
##  Now re-fitting the best model(s) without approximations...
##
##
##
##
##  Best model: ARIMA(4,0,1) with non-zero mean
```

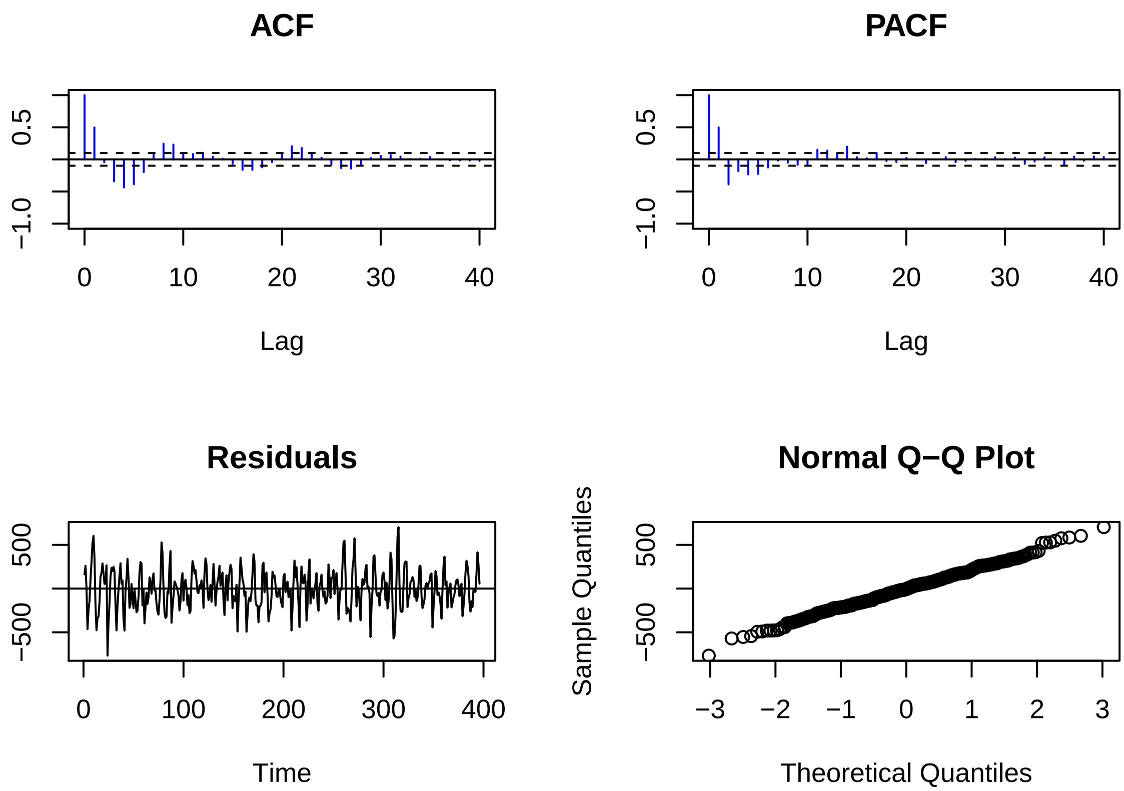
```
checkresiduals(wave_auto_select_AIC)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(4,0,1) with non-zero mean
## Q* = 10.436, df = 4, p-value = 0.03369
##
## Model df: 6.   Total lags used: 10
```

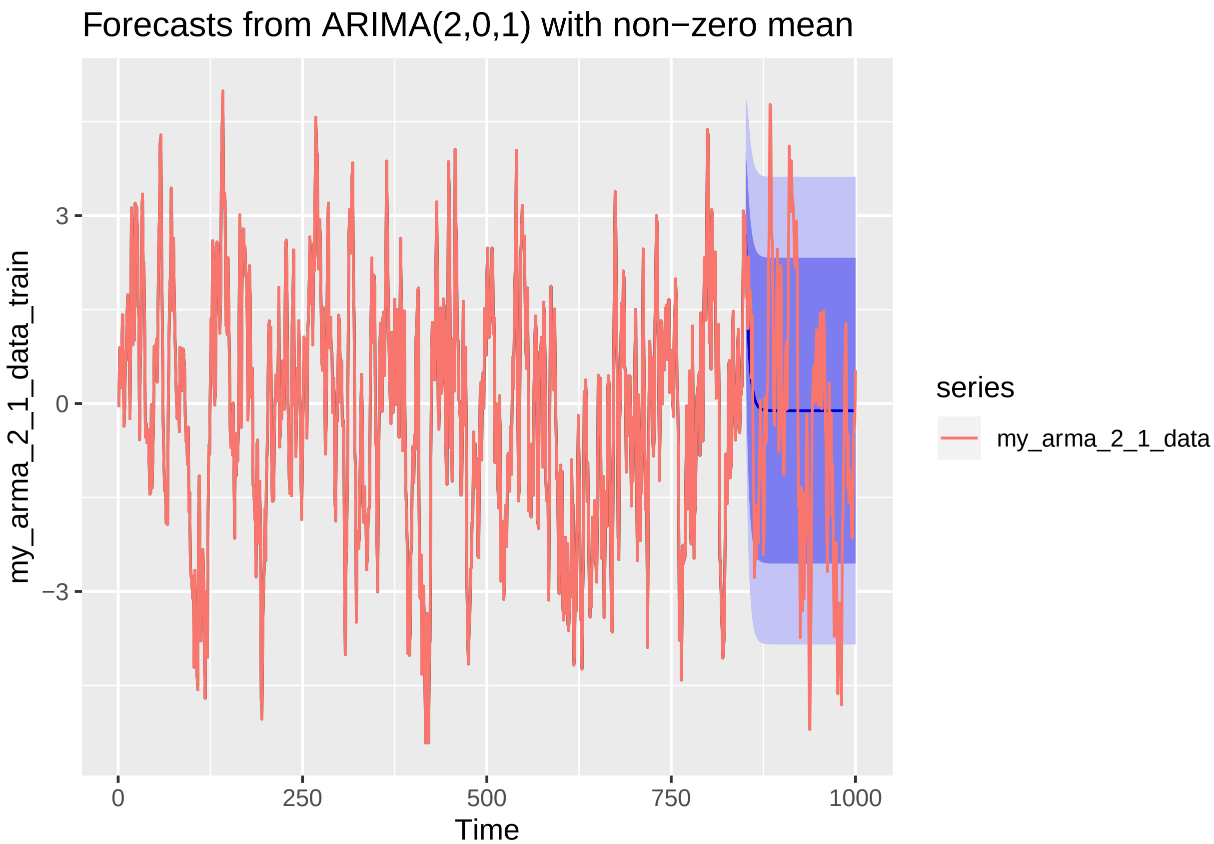
```
test(fitted(wave_auto_select_AIC))
```

```
## Null hypothesis: Residuals are iid noise.
## Test                        Distribution Statistic   p-value
## Ljung-Box Q                Q ~ chisq(20)    395.47         0 *
## McLeod-Li Q                Q ~ chisq(20)     67.05         0 *
## Turning points T  (T-262.7)/8.4 ~ N(0,1)       185         0 *
## Diff signs S      (S-197.5)/5.8 ~ N(0,1)       225         0 *
## Rank P         (P-39105)/1315.9 ~ N(0,1)     39542    0.7398
```



Forecasting

```
my_arma_2_1_data_train <- window(my_arma_2_1_data,end=850)
my_arma_2_1_data_test <- window(my_arma_2_1_data,start=851,end=1000)

my_arma_2_1_train_mod <- arima(my_arma_2_1_data_train,c(2,0,1))
myforecasts<-forecast::forecast(my_arma_2_1_train_mod,h=150)

autoplot(myforecasts) + autolayer(my_arma_2_1_data)
```




```r
forecast_table%>% dplyr::slice(1:15) %>%  select(c(1,6,7,2:5)) %>%
  round(2) %>% kable()
```

Point Forecast	observed	errors	Lo 80	Hi 80	Lo 95	Hi 95
2.70	1.96	0.74	1.37	4.03	0.67	4.73
2.04	1.70	0.34	0.20	3.87	-0.77	4.84
1.65	1.74	-0.09	-0.41	3.71	-1.50	4.80
1.29	1.20	0.10	-0.92	3.50	-2.08	4.66
1.02	2.35	-1.33	-1.27	3.31	-2.49	4.52
0.79	1.55	-0.75	-1.55	3.14	-2.79	4.38
0.62	1.79	-1.18	-1.76	2.99	-3.02	4.25
0.47	1.64	-1.17	-1.93	2.87	-3.20	4.14
0.36	0.40	-0.05	-2.06	2.77	-3.34	4.05
0.26	1.41	-1.15	-2.16	2.69	-3.44	3.97
0.19	-0.67	0.86	-2.24	2.62	-3.53	3.91
0.13	-1.48	1.61	-2.30	2.56	-3.59	3.85
0.08	-2.78	2.86	-2.35	2.52	-3.64	3.81
0.04	-2.39	2.43	-2.39	2.48	-3.68	3.77
0.01	-1.64	1.66	-2.43	2.45	-3.72	3.74

```r
forecast_table%>%  mutate(outside95=I(observed<`Lo 95` | observed>`Hi 95`)) %>%
  count(outside95)
```

```
## # A tibble: 2 x 2
##   outside95     n
##   <I<lgl>>  <int>
```

```
## 1 FALSE        140
## 2 TRUE          10
```