TS-Project

Yirong

1/28/2017

```r
library(forecast)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: timeDate

## This is forecast 7.3
```

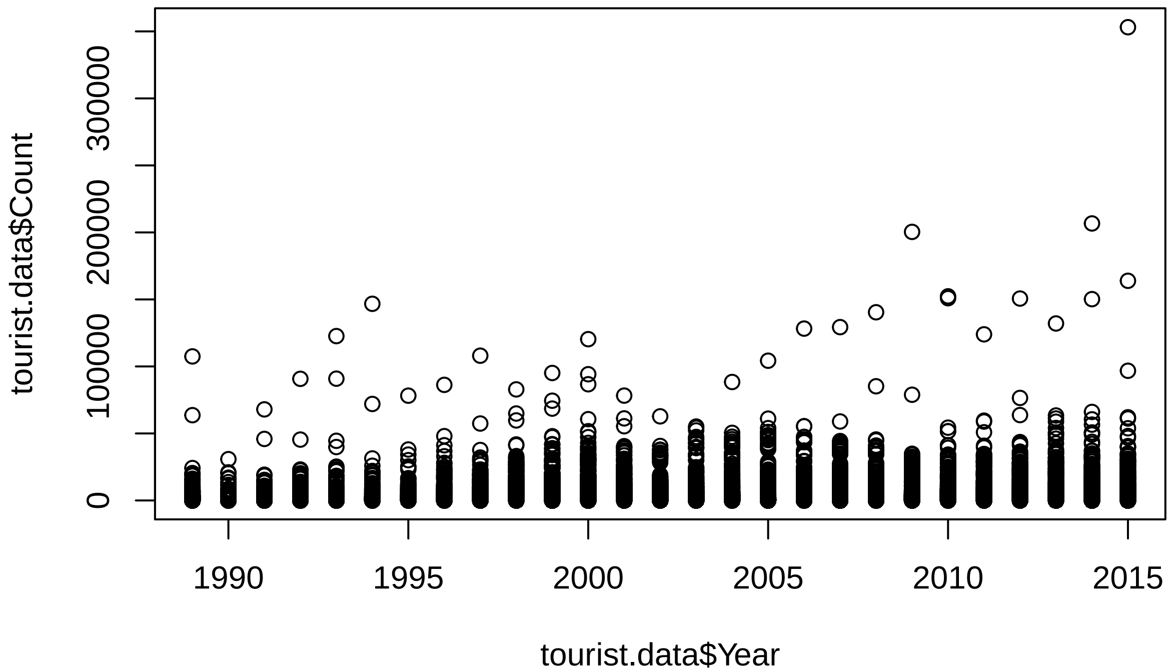
```r
library(fpp)
```

```
## Loading required package: fma

## Loading required package: tseries

## Warning: package 'tseries' was built under R version 3.3.2

## Loading required package: expsmooth

## Loading required package: lmtest
```

```r
setwd("~/Desktop/Yirong/Winter/TimeSeries")
```

Read Data into R; Aggregate and Clean Data

```r
tourist.data = read.csv("touristData.csv")
plot(tourist.data$Year,tourist.data$Count)
```



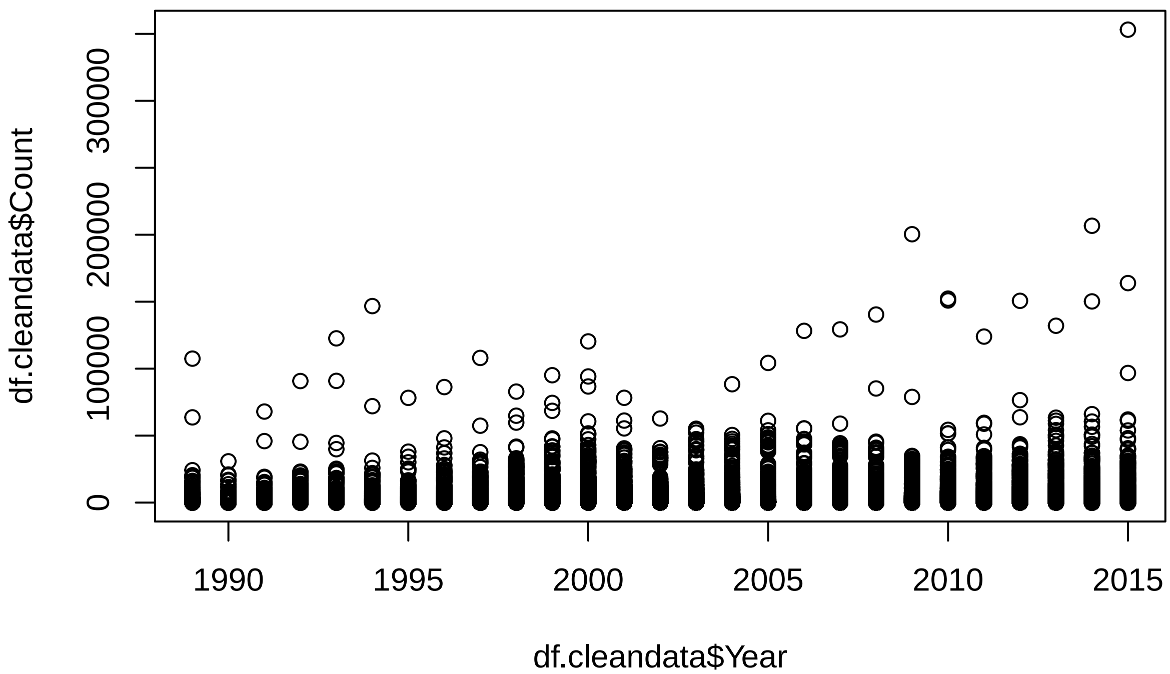

```r
summary(tourist.data)
```

```
##                              Continent                  Country
##  Europe                          :181044   Other countries: 51300
##  South America                   :123120   Angola         : 10260
##  Asia                            : 74568   Argentina      : 10260
##  Africa                          : 45948   Australia      : 10260
##  Central America and Caribbean: 44796   Austria        : 10260
##  North America                   : 30780   Belgium        : 10260
##  (Other)                         : 34536   (Other)        :432192
##                                    State           WayIn             Year
##  Outras Unidades da Federa<e7><e3>o: 66576   Air  :210060   Min.   :1989
##  Rio Grande do Sul                 : 65904   Land :100116   1st Qu.:1996
##  Paran<e1>                         : 62544   River: 80604   Median :2004
##  Amazonas                          : 41856   Sea  :144012   Mean   :2003
##  Santa Catarina                    : 38088                  3rd Qu.:2010
##  Par<e1>                           : 34800                  Max.   :2015
##  (Other)                           :225024
##        Month            Count
##  abril    : 44566   Min.   :     0
##  agosto   : 44566   1st Qu.:     0
##  dezembro : 44566   Median :     0
##  fevereiro: 44566   Mean   :   206
##  janeiro  : 44566   3rd Qu.:    11
##  julho    : 44566   Max.   :353122
##  (Other)  :267396   NA's   :6192
```

```r
df.cleandata = na.omit(tourist.data)
nrow(df.cleandata)
```

```
## [1] 528600
```

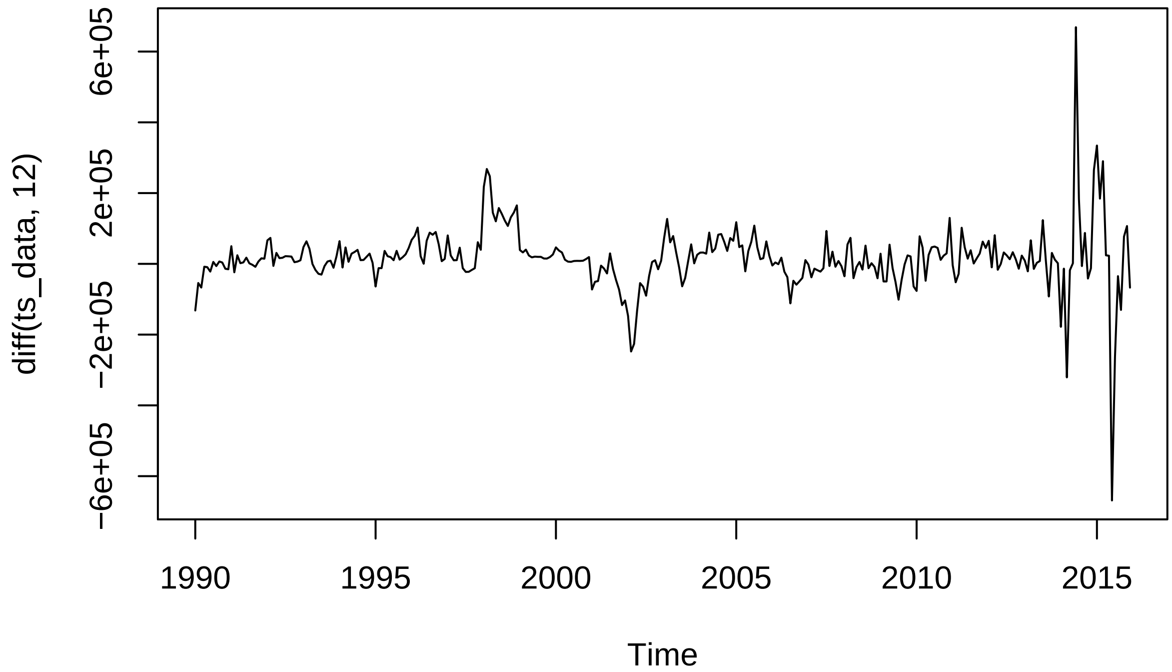
```r
df.cleandata = tourist.data[tourist.data$Count != 0,]
plot(df.cleandata$Year, df.cleandata$Count)
```




```r
df.cleandata$Month <- as.character(df.cleandata$Month)
df.cleandata$Month[df.cleandata$Month == "janeiro"] <- "01"
df.cleandata$Month[df.cleandata$Month == "fevereiro"] <- "02"
df.cleandata$Month[df.cleandata$Month == "mar\xe7o"] <- "03"
df.cleandata$Month[df.cleandata$Month == "abril"] <- "04"
df.cleandata$Month[df.cleandata$Month == "maio"] <- "05"
df.cleandata$Month[df.cleandata$Month == "junho"] <- "06"
df.cleandata$Month[df.cleandata$Month == "julho"] <- "07"
df.cleandata$Month[df.cleandata$Month == "agosto"] <- "08"
df.cleandata$Month[df.cleandata$Month == "setembro"] <- "09"
df.cleandata$Month[df.cleandata$Month == "outubro"] <- "10"
df.cleandata$Month[df.cleandata$Month == "novembro"] <- "11"
df.cleandata$Month[df.cleandata$Month == "dezembro"] <- "12"

df.cleandata$Month <- as.factor(df.cleandata$Month)


sortdata <- with(df.cleandata, aggregate(Count, list(Month = Month, Year = Year), sum)) #Sort by month,
```

Create Time Serie dataset

```r
# create time serise data. Plot data
count = sortdata$x
ts_data = ts(count,start=c(1989, 1), end=c(2015, 12), frequency=12)
ts.plot(diff(ts_data,12)) #Not stationary. Seasonality
```



Test Stationary and Take Difference

```r
adf.test(ts_data,alternative = "stationary")#p-value is smaller than 0.05
```

```
## Warning in adf.test(ts_data, alternative = "stationary"): p-value smaller
## than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  ts_data
## Dickey-Fuller = -7.7265, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

```r
kpss.test(ts_data) #p-value is smaller than 0.05
```

```
## Warning in kpss.test(ts_data): p-value smaller than printed p-value
```

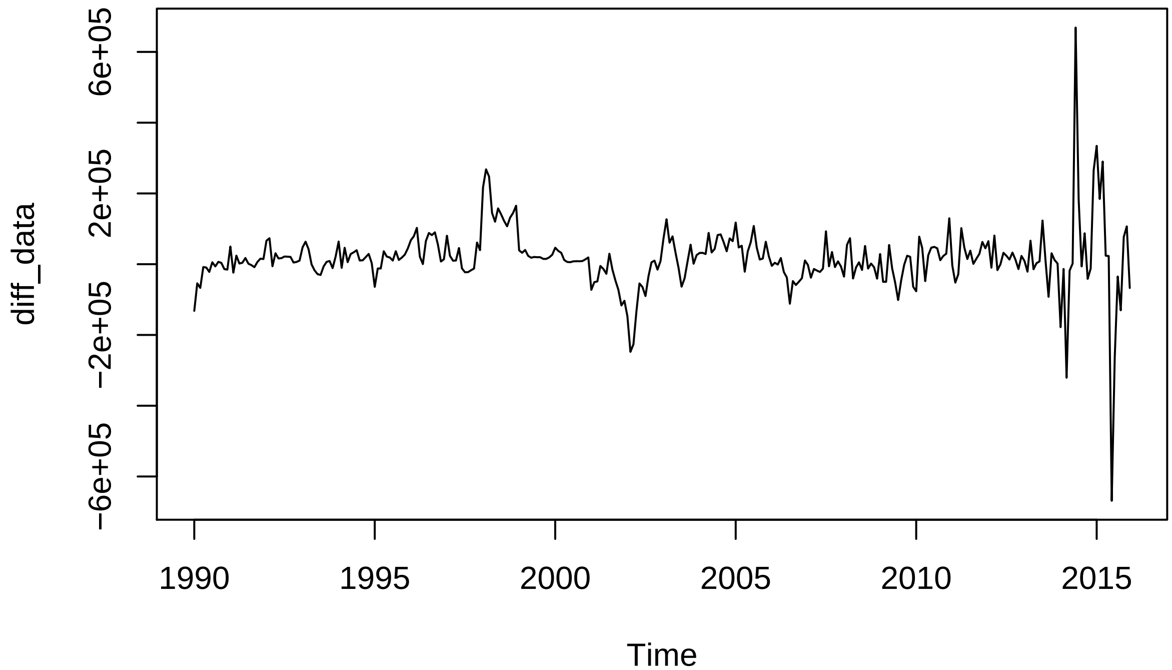
```
##
##  KPSS Test for Level Stationarity
##
## data:  ts_data
## KPSS Level = 4.4198, Truncation lag parameter = 4, p-value = 0.01
```

```r
kpss.test(diff(ts_data,12))
```

```
## Warning in kpss.test(diff(ts_data, 12)): p-value greater than printed p-
## value
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  diff(ts_data, 12)
## KPSS Level = 0.10137, Truncation lag parameter = 4, p-value = 0.1
```

```r
diff_data = diff(ts_data,12)

ts.plot(diff_data)
```

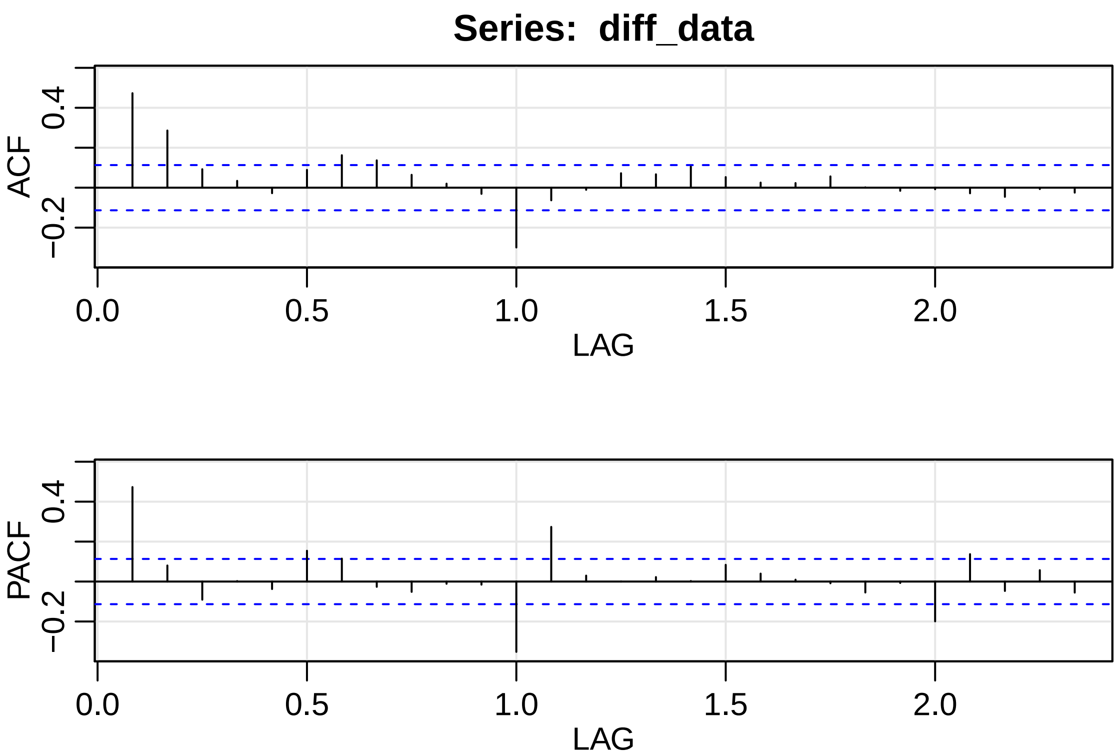


Build ARIMA model

```r
library(astsa)
```

```
## Warning: package 'astsa' was built under R version 3.3.2

##
## Attaching package: 'astsa'

## The following object is masked from 'package:fpp':
##
##     oil

## The following objects are masked from 'package:fma':
##
##     chicken, sales

## The following object is masked from 'package:forecast':
##
##     gas
```

```r
acf2(diff_data)
```




```
##        ACF  PACF
##  [1,]  0.47  0.47
##  [2,]  0.29  0.08
##  [3,]  0.09 -0.09
##  [4,]  0.03  0.00
##  [5,] -0.03 -0.04
##  [6,]  0.09  0.15
##  [7,]  0.16  0.11
##  [8,]  0.14 -0.03
##  [9,]  0.06 -0.05
## [10,]  0.02 -0.01
## [11,] -0.03 -0.02
## [12,] -0.30 -0.35
## [13,] -0.06  0.27
## [14,] -0.01  0.03
## [15,]  0.07  0.00
## [16,]  0.07  0.02
## [17,]  0.11  0.00
## [18,]  0.05  0.08
## [19,]  0.03  0.04
## [20,]  0.02  0.01
## [21,]  0.06 -0.01
## [22,]  0.00 -0.05
## [23,] -0.02 -0.01
## [24,] -0.01 -0.20
## [25,] -0.03  0.14
## [26,] -0.05 -0.05
```

```
## [27,] -0.01  0.06
## [28,] -0.02 -0.06
```

```r
# order = (1,0,2) or (1,0,4) or (2,0,4)
# seasonal = (1,0,1) or (2,0,1)
fit_1 <- Arima(diff_data, order = c(1,0,2), seasonal = c(1,0,1))
summary(fit_1) #AICc=7817.54
```

```
## Series: diff_data
## ARIMA(1,0,2)(1,0,1)[12] with non-zero mean
##
## Coefficients:
##          ar1      ma1      ma2     sar1     sma1  intercept
##       0.9418  -0.4400  -0.2476  -0.1917  -0.5546  16337.879
## s.e.  0.0359   0.0707   0.0577   0.1616   0.1532   7426.538
##
## sigma^2 estimated as 4.243e+09:  log likelihood=-3901.59
## AIC=7817.17   AICc=7817.54   BIC=7843.37
##
## Training set error measures:
##                    ME     RMSE      MAE      MPE     MAPE     MASE
## Training set -25.8819 64506.86 35288.75 22.46228 196.0868 0.470342
##                    ACF1
## Training set 0.05197197
```

```r
fit_2 <- Arima(diff_data, order = c(1,0,4), seasonal = c(1,0,1))
summary(fit_2) #AICc=7817.54
```

```
## Series: diff_data
## ARIMA(1,0,4)(1,0,1)[12] with non-zero mean
##
## Coefficients:
##          ar1      ma1      ma2      ma3     ma4     sar1     sma1
##       0.9549  -0.3823  -0.2005  -0.2093  0.0605  -0.2582  -0.5046
## s.e.  0.0252   0.0638   0.0652   0.0547  0.0618   0.1850   0.1783
##       intercept
##       16148.587
## s.e.   8332.304
##
## sigma^2 estimated as 4.084e+09:  log likelihood=-3894.65
## AIC=7807.31   AICc=7807.9   BIC=7840.99
##
## Training set error measures:
##                    ME     RMSE      MAE      MPE     MAPE      MASE
## Training set 76.37386 63083.48 35302.29 40.79722 198.9359 0.4705225
##                      ACF1
## Training set -0.000327513
```

```r
fit_3 <- Arima(diff_data, order = c(2,0,4), seasonal = c(1,0,1))
summary(fit_3) #AICc=7807.16
```

```
## Series: diff_data
```

```
## ARIMA(2,0,4)(1,0,1)[12] with non-zero mean
##
## Coefficients:
##          ar1     ar2     ma1      ma2      ma3      ma4     sar1     sma1
##       0.1788  0.7415  0.4058  -0.4925  -0.3703  -0.0633  -0.2936  -0.4837
## s.e.  0.1269  0.1210  0.1413   0.0787   0.0627   0.0703   0.1794   0.1736
##       intercept
##       16196.074
## s.e.   8477.456
##
## sigma^2 estimated as 4.058e+09:  log likelihood=-3893.22
## AIC=7806.43   AICc=7807.16   BIC=7843.86
##
## Training set error measures:
##                    ME     RMSE      MAE    MPE     MAPE      MASE
## Training set 100.0034 62775.01 35633.86 41.226 201.7773 0.4749418
##                     ACF1
## Training set 0.002159472
```

```r
fit_4 <- Arima(diff_data, order = c(1,0,2), seasonal = c(2,0,1))
summary(fit_4) #AICc=7813.98
```

```
## Series: diff_data
## ARIMA(1,0,2)(2,0,1)[12] with non-zero mean
##
## Coefficients:
##          ar1     ma1     ma2    sar1    sar2     sma1  intercept
##       0.5027  0.0899  0.1228  0.2148  0.3643  -0.9384   16751.01
## s.e.  0.1532  0.1649  0.0983  0.0856  0.0967   0.0811    2368.50
##
## sigma^2 estimated as 4.105e+09:  log likelihood=-3898.75
## AIC=7813.5   AICc=7813.98   BIC=7843.45
##
## Training set error measures:
##                    ME     RMSE      MAE     MPE     MAPE      MASE
## Training set 534.9231 63347.21 35958.41 48.9219 237.7623 0.4792675
##                     ACF1
## Training set -0.00577597
```

```r
fit_5 <- Arima(diff_data, order = c(1,0,4), seasonal = c(2,0,1))
summary(fit_5) #AICc=7803.81
```

```
## Series: diff_data
## ARIMA(1,0,4)(2,0,1)[12] with non-zero mean
##
## Coefficients:
##          ar1      ma1      ma2      ma3     ma4    sar1    sar2     sma1
##       0.9533  -0.3859  -0.1904  -0.2290  0.0572  0.1030  0.3391  -0.8807
## s.e.  0.0276   0.0634   0.0645   0.0551  0.0611  0.1026  0.1148   0.0876
##       intercept
##       16393.741
## s.e.   5009.177
##
```

```
## sigma^2 estimated as 3.974e+09:  log likelihood=-3891.54
## AIC=7803.08   AICc=7803.81   BIC=7840.51
##
## Training set error measures:
##                     ME     RMSE      MAE      MPE     MAPE      MASE
## Training set 101.6438 62126.55 35055.59 54.75469 222.5119 0.4672343
##                       ACF1
## Training set -0.0005188945
```

```
fit_6 <- Arima(diff_data, order = c(2,0,4), seasonal = c(2,0,1))
summary(fit_6) # AICc=7802.92
```

```
## Series: diff_data
## ARIMA(2,0,4)(2,0,1)[12] with non-zero mean
##
## Coefficients:
##          ar1     ar2     ma1      ma2      ma3      ma4    sar1    sar2
##       0.1674  0.7504  0.4098  -0.4920  -0.3858  -0.0818  0.0872  0.3507
## s.e.  0.1212  0.1160  0.1351   0.0791   0.0635   0.0709  0.1031  0.1149
##          sma1  intercept
##       -0.8803  16471.453
## s.e.   0.0873   5025.957
##
## sigma^2 estimated as 3.945e+09:  log likelihood=-3890.02
## AIC=7802.04   AICc=7802.92   BIC=7843.22
##
## Training set error measures:
##                     ME  RMSE      MAE      MPE    MAPE      MASE
## Training set 83.42126 61797 35363.08 54.92209 226.819 0.4713327
##                     ACF1
## Training set 0.003458095
```

```
fit_7 <- Arima(diff_data, order = c(2,0,4), seasonal = c(0,0,1))
summary(fit_7) # AICc=7809.98
```

```
## Series: diff_data
## ARIMA(2,0,4)(0,0,1)[12] with non-zero mean
##
## Coefficients:
##          ar1     ar2     ma1      ma2      ma3      ma4     sma1
##       0.2110  0.7227  0.3486  -0.5075  -0.3639  -0.0701  -0.7251
## s.e.  0.1462  0.1412  0.1565   0.0826   0.0642   0.0715   0.0508
##       intercept
##       16469.310
## s.e.   6262.053
##
## sigma^2 estimated as 4.066e+09:  log likelihood=-3894.69
## AIC=7807.38   AICc=7807.98   BIC=7841.07
##
## Training set error measures:
##                     ME     RMSE      MAE      MPE     MAPE      MASE
## Training set -172.997 62943.14 35306.26 38.28926 217.6053 0.4705753
##                     ACF1
## Training set 0.002815134
```

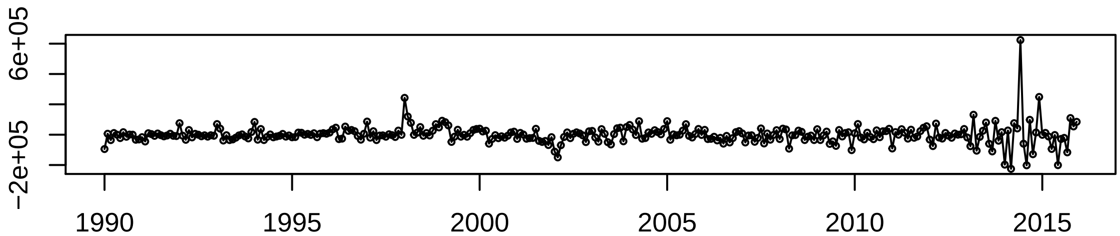
```r
fit_8 <- Arima(diff_data, order = c(1,0,0), seasonal = c(1,0,1))
summary(fit_8) # AICc=7817.35
```

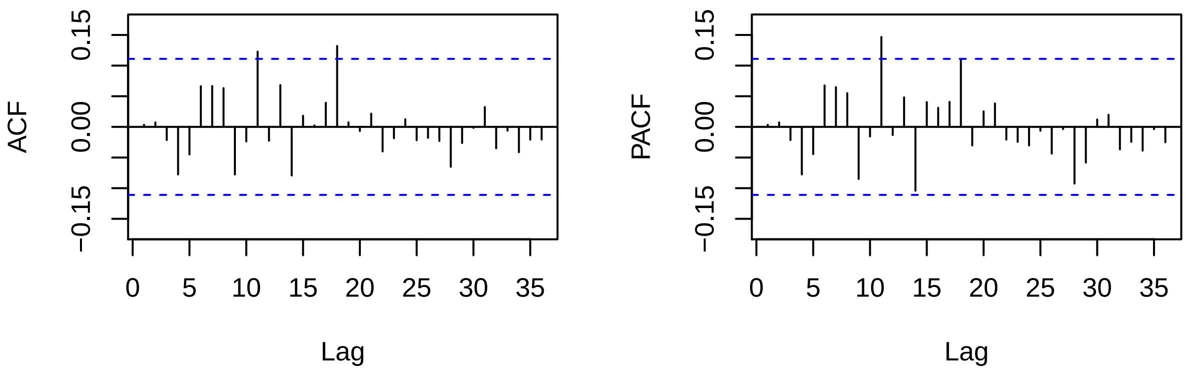
```
## Series: diff_data
## ARIMA(1,0,0)(1,0,1)[12] with non-zero mean
##
## Coefficients:
##          ar1     sar1     sma1  intercept
##       0.6325  -0.2410  -0.4627  16935.187
## s.e.  0.0460   0.1986   0.1913   4497.007
##
## sigma^2 estimated as 4.278e+09:  log likelihood=-3903.58
## AIC=7817.15   AICc=7817.35   BIC=7835.87
##
## Training set error measures:
##                     ME     RMSE      MAE      MPE     MAPE      MASE
## Training set -463.4918 64982.06 36745.84 27.82342 212.9508 0.4897626
##                    ACF1
## Training set -0.0395586
```

```r
auto.arima(diff_data, seasonal = TRUE) #ARIMA(1,0,0)(0,0,1)[12] AICc=7816.78
```

```
## Series: diff_data
## ARIMA(1,0,0)(0,0,1)[12] with non-zero mean
##
## Coefficients:
##          ar1     sma1  intercept
##       0.6180  -0.6700   16963.72
## s.e.  0.0458   0.0526    3442.21
##
## sigma^2 estimated as 4.271e+09:  log likelihood=-3904.33
## AIC=7816.65   AICc=7816.78   BIC=7831.62
```

```r
res <- residuals(fit_6)
tsdisplay(res, main = "")
```

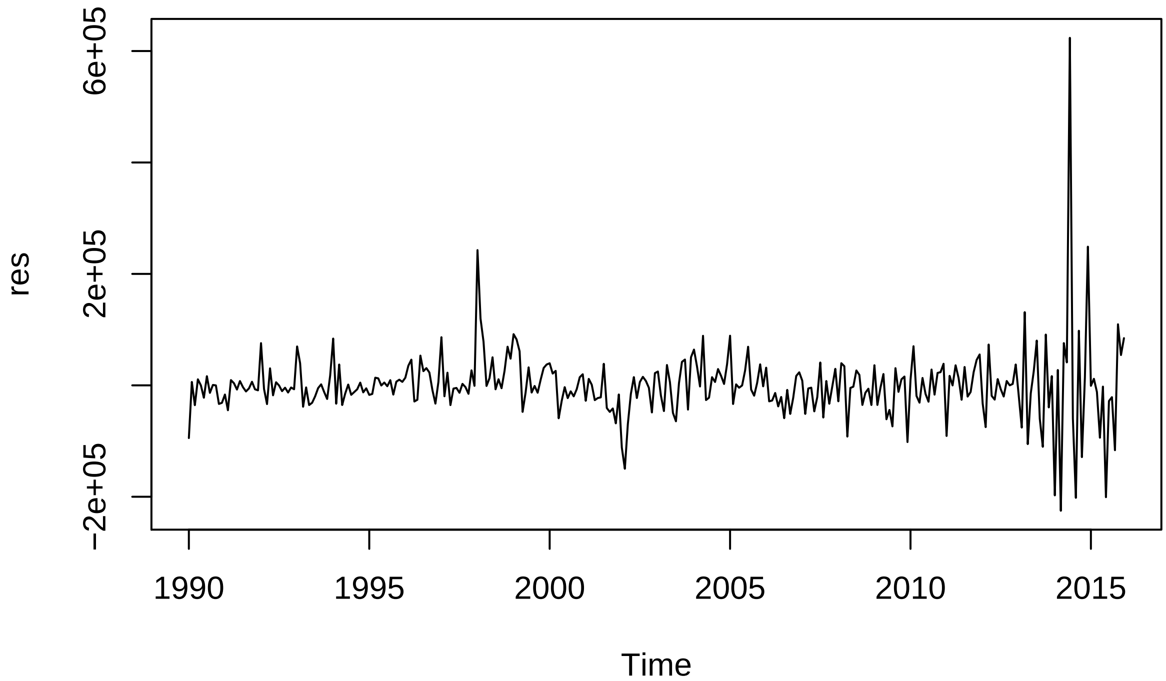





```r
Box.test(res, lag = 36, fitdf = 9, type = "Ljung")
```

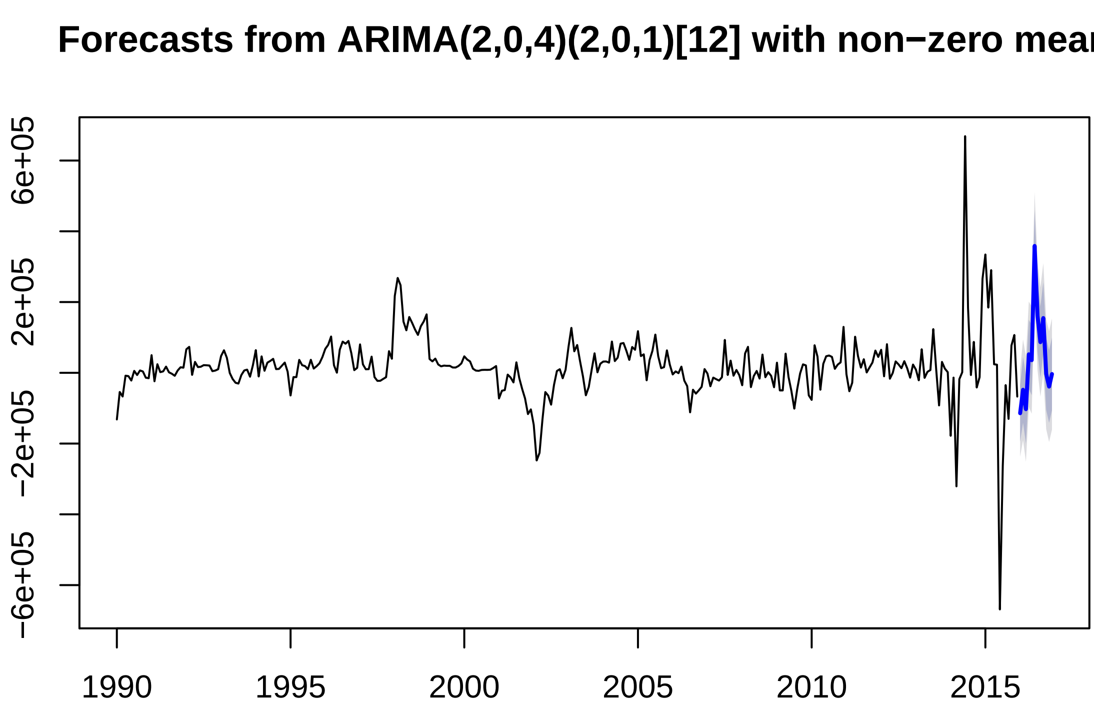
```
##
##  Box-Ljung test
##
## data:  res
## X-squared = 28.872, df = 27, p-value = 0.3671
```

```r
plot(res)
```



Forecast

```r
plot(forecast(fit_6, h = 12))
```




```r
fcast = forecast(fit_6, h = 12)
fcast
```

```
##          Point Forecast      Lo 80      Hi 80        Lo 95      Hi 95
```

```
## Jan 2016    -114133.579 -194634.79 -33632.371 -237249.5533   8982.394
## Feb 2016     -47990.836 -140940.70  44959.032 -190145.3924  94163.720
## Mar 2016    -102754.287 -200000.02  -5508.552 -251478.8050  45970.231
## Apr 2016      52258.925  -45366.29 149884.145  -97045.9641 201563.814
## May 2016      35788.915  -63188.13 134765.961 -115583.4137 187161.245
## Jun 2016     358546.711  259144.58 457948.841  206524.2729 510569.150
## Jul 2016     161089.047   60737.47 261440.619    7614.5603 314563.533
## Aug 2016      86732.028  -14040.11 187504.170  -67385.6645 240849.720
## Sep 2016     154796.853   53327.51 256266.193    -387.1102 309980.816
## Oct 2016      -3089.455 -104947.49  98768.576 -158867.8709 152688.960
## Nov 2016     -38835.442 -141222.91  63552.027 -195423.5623 117752.679
## Dec 2016      -3513.328 -106245.60  99218.940 -160628.7729 153602.117
```