Final Project

Derek Elvis Rui Wang, Jose Ventura

2024-04-29

```r
require(forecast)
```

```
## Loading required package: forecast

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
require(tseries)
```

```
## Loading required package: tseries
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
require(tinytex)
```

```
## Loading required package: tinytex
```

```r
library(fpp2)
```

```
## -- Attaching packages ---------------------------------------------- fpp2 2.5 --

## v ggplot2   3.4.4     v expsmooth 2.3
## v fma       2.5

##
```

```r
require(ggplot2)
library(readr)
require(lubridate)
```

```
## Loading required package: lubridate

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

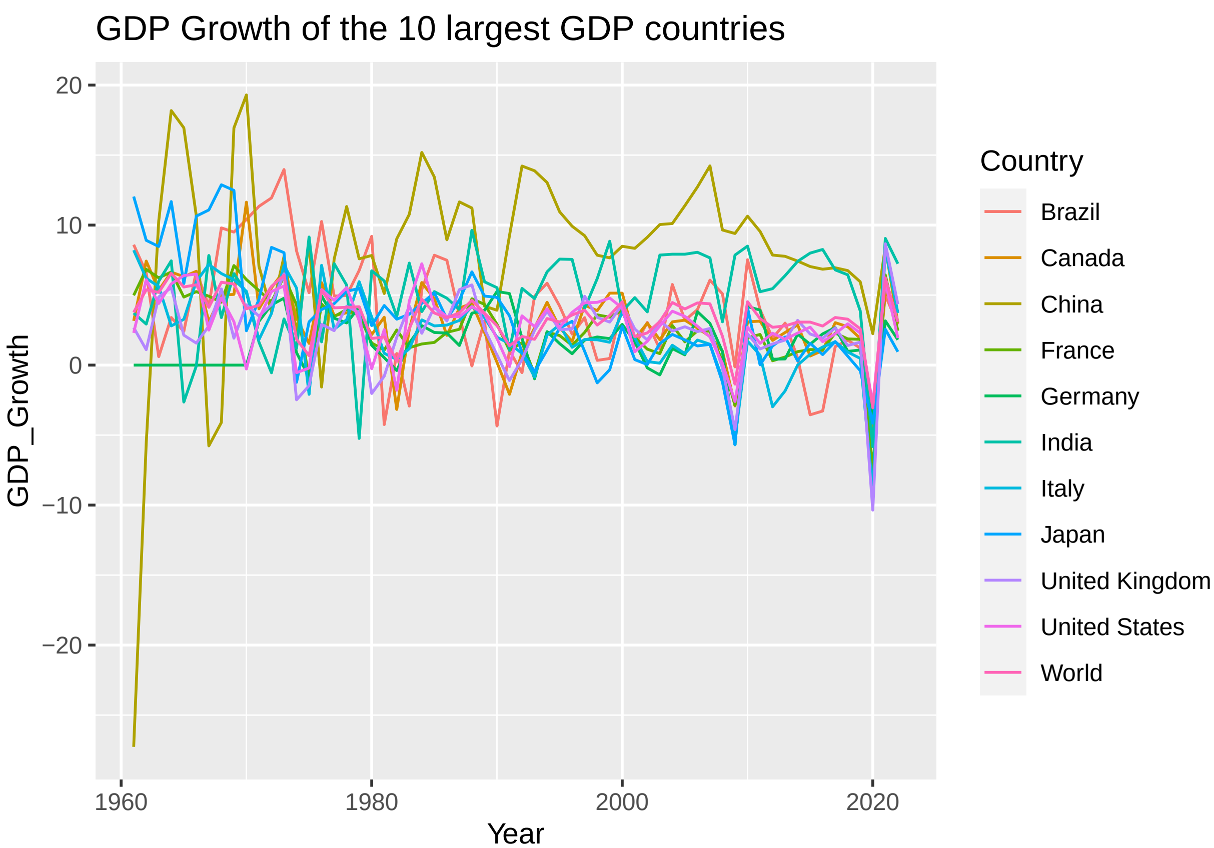
```r
library(urca)
```

For our project we are finding the GDP and GDP growth of the ten countries with the highest GDP and the average of the world

```r
gdp <- gdp <- read_csv("~/Documents/Derek/R/TS_GDP_Dataset.csv",
                col_types = cols(Year = col_date(format = "%Y"),
                                 `GDP Growth` = col_number()))

names(gdp) <- c("Year", "Country", "Country_Code", "GDP", "GDP_Growth")
country_data <- unique(gdp$Country)
```

After uploading the csv file, we changed the name to have the names of the columns to be read easier as well as to have two names. Then created the 'country_data' to show the unique countries we will be working with.

```r
ggplot(gdp, aes(x = Year, y = GDP_Growth, color = Country)) +
  geom_line() +
  labs(title = "GDP Growth of the 10 largest GDP countries")
```

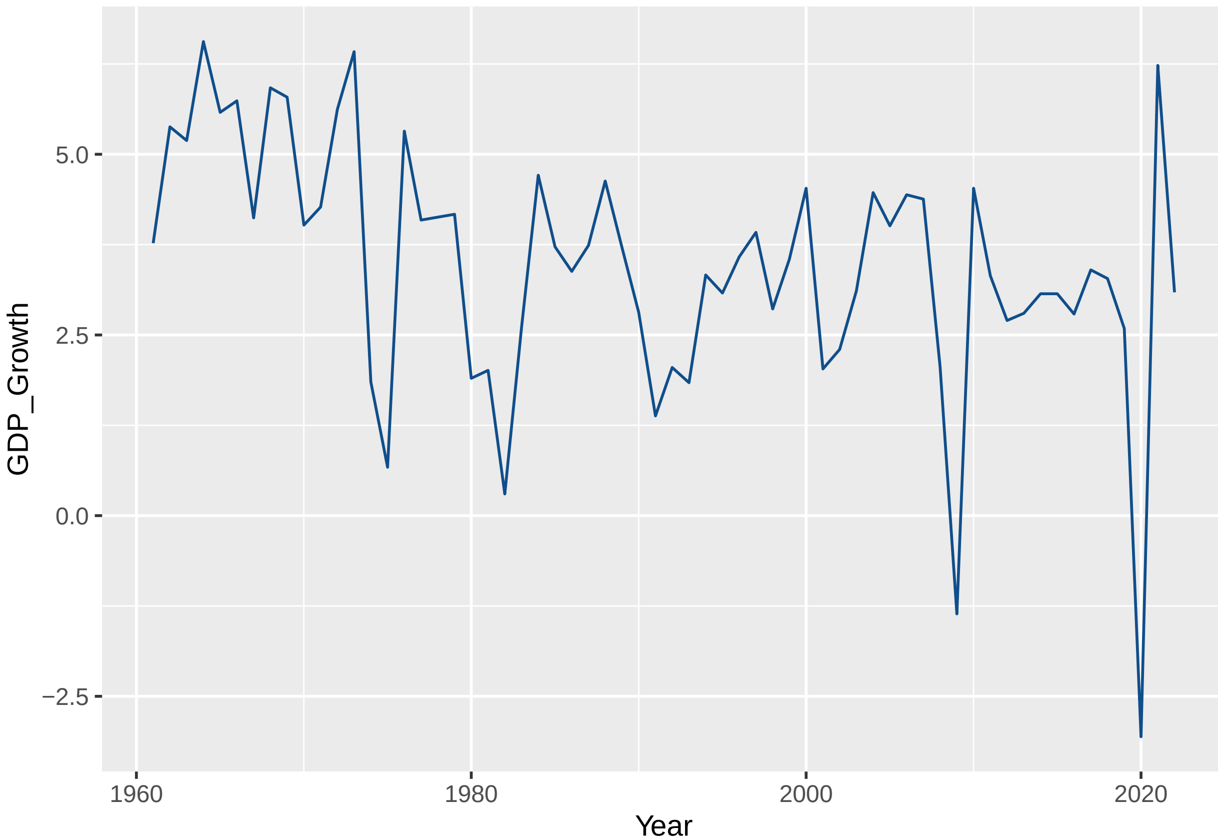


Here is a graph of the 10 countries and the world avg in one graph

```r
world_gdp <- gdp %>%
  filter(Country == "World") %>%
  select(Year, GDP, GDP_Growth)
```

We are first going to look at the world gdp, so first we filter the data so it only has world data

```r
ggplot(world_gdp) +
  geom_line(aes(Year, GDP_Growth, group = 1), color = "dodgerblue4", linewidth = .5)
```

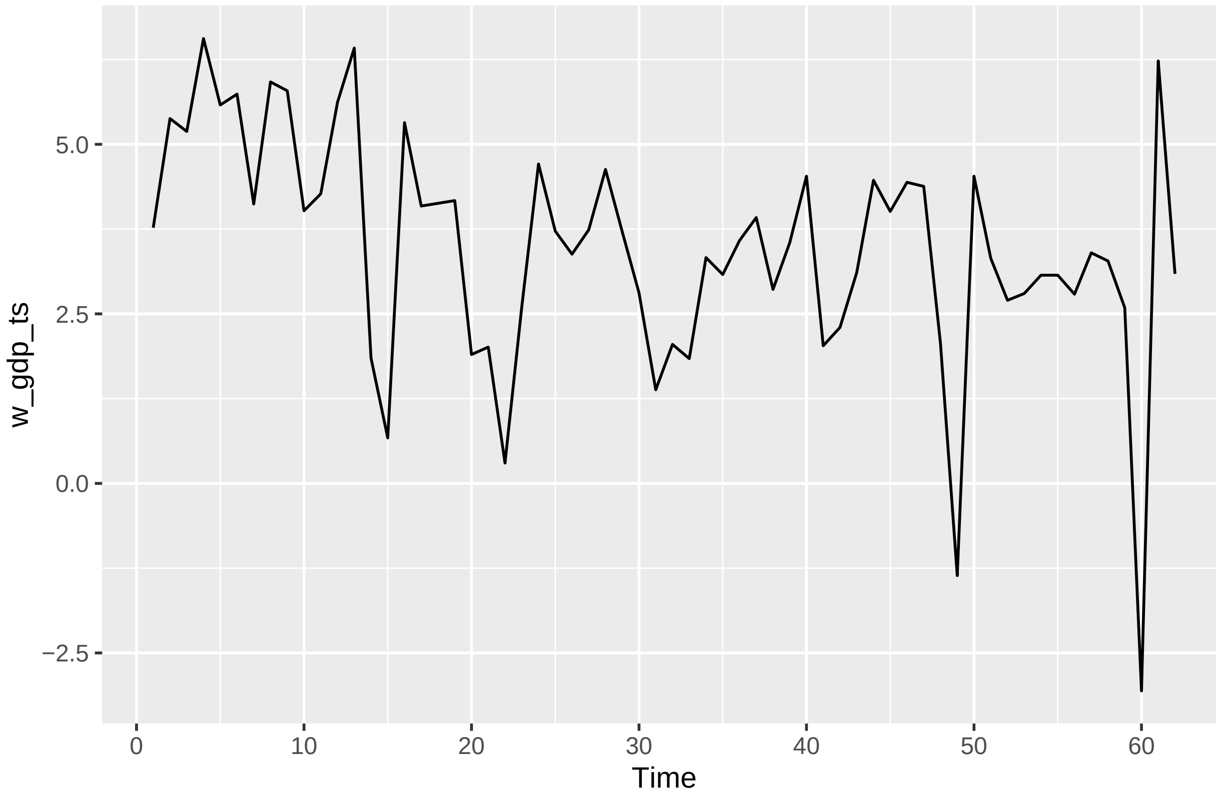


Here is the graph of the gdp growth of the world gdp

```
wtrain <- world_gdp[1:50,]
wtest <- world_gdp[51:72,]
wntest <- nrow(wtest)
```

Then we make a train and testing set

```
w_gdp_ts <- ts(world_gdp$GDP_Growth, frequency = 1)

#Stationarity
autoplot(w_gdp_ts)
```

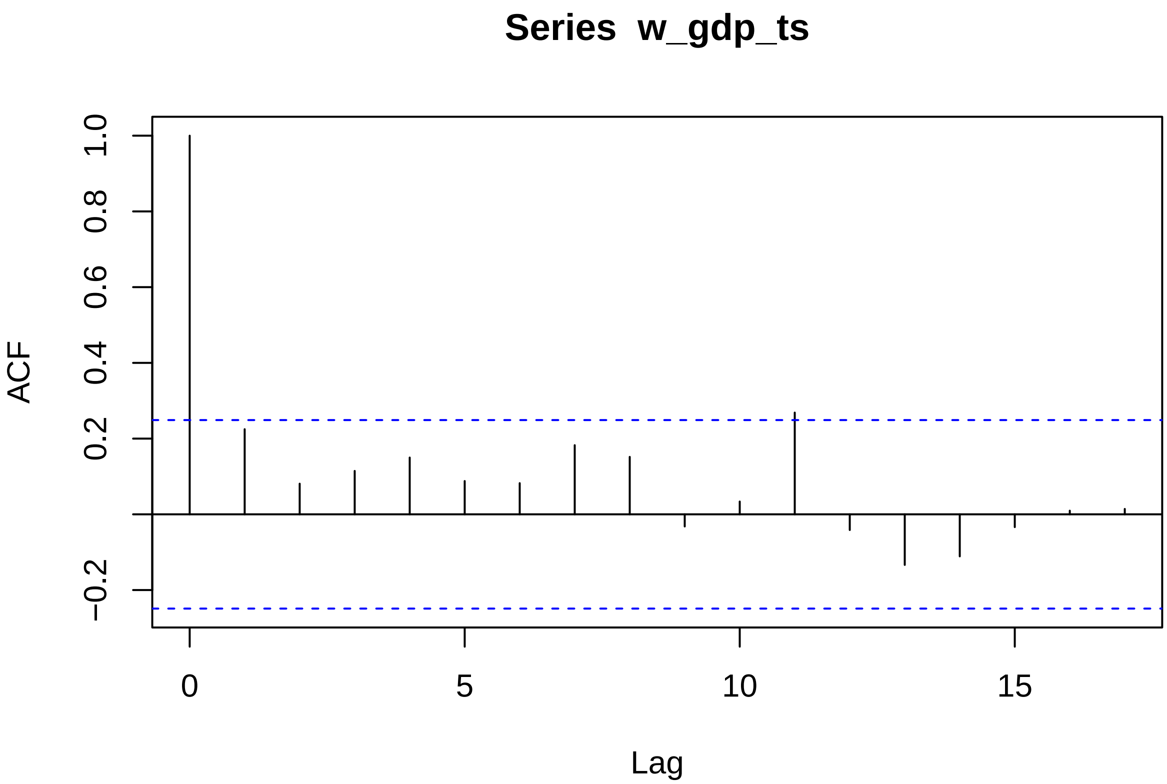


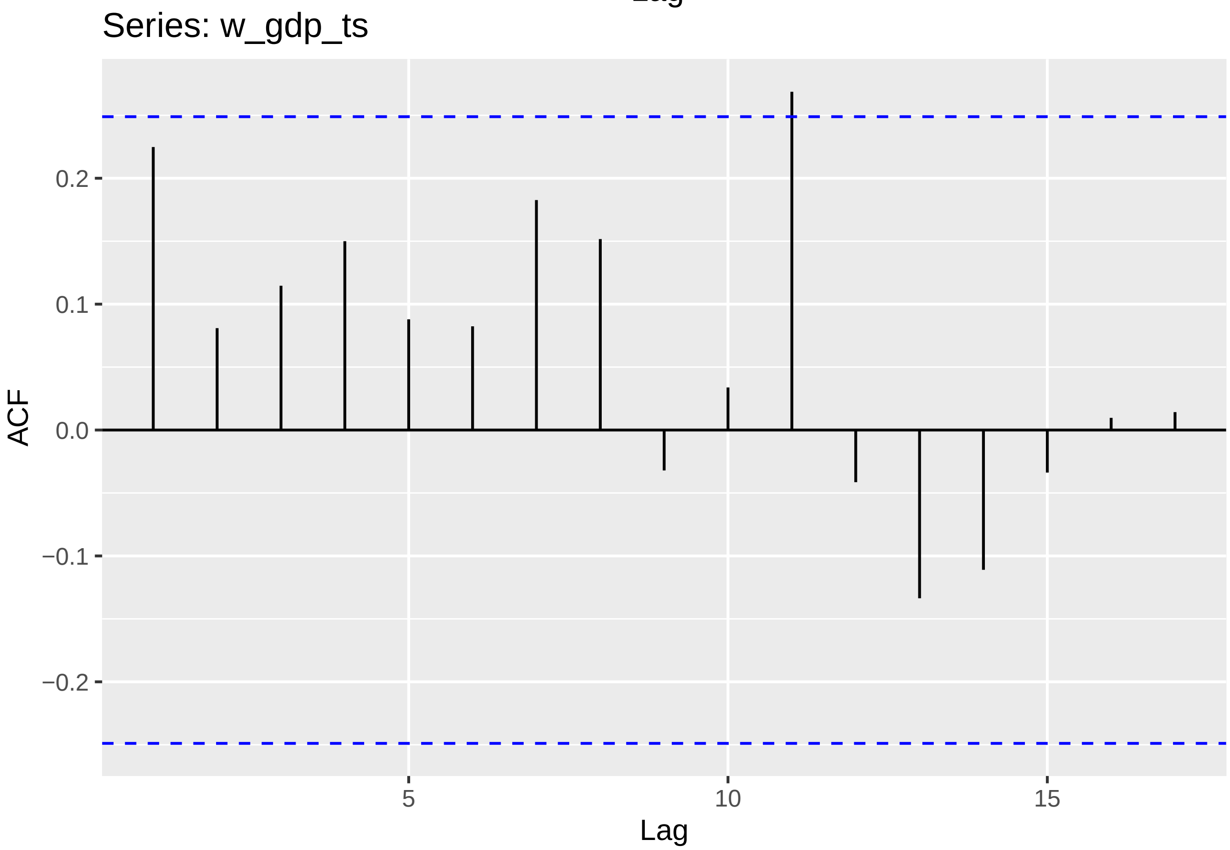

```r
ur.kpss(w_gdp_ts) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.7126
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Then we create a time series model and then check for stationary. We use the kpss test and see if it needs to be diffrenced. Seeing that the test-statisic is near the test data, we can diffrence the data once.

```r
autoplot(acf(w_gdp_ts))
```

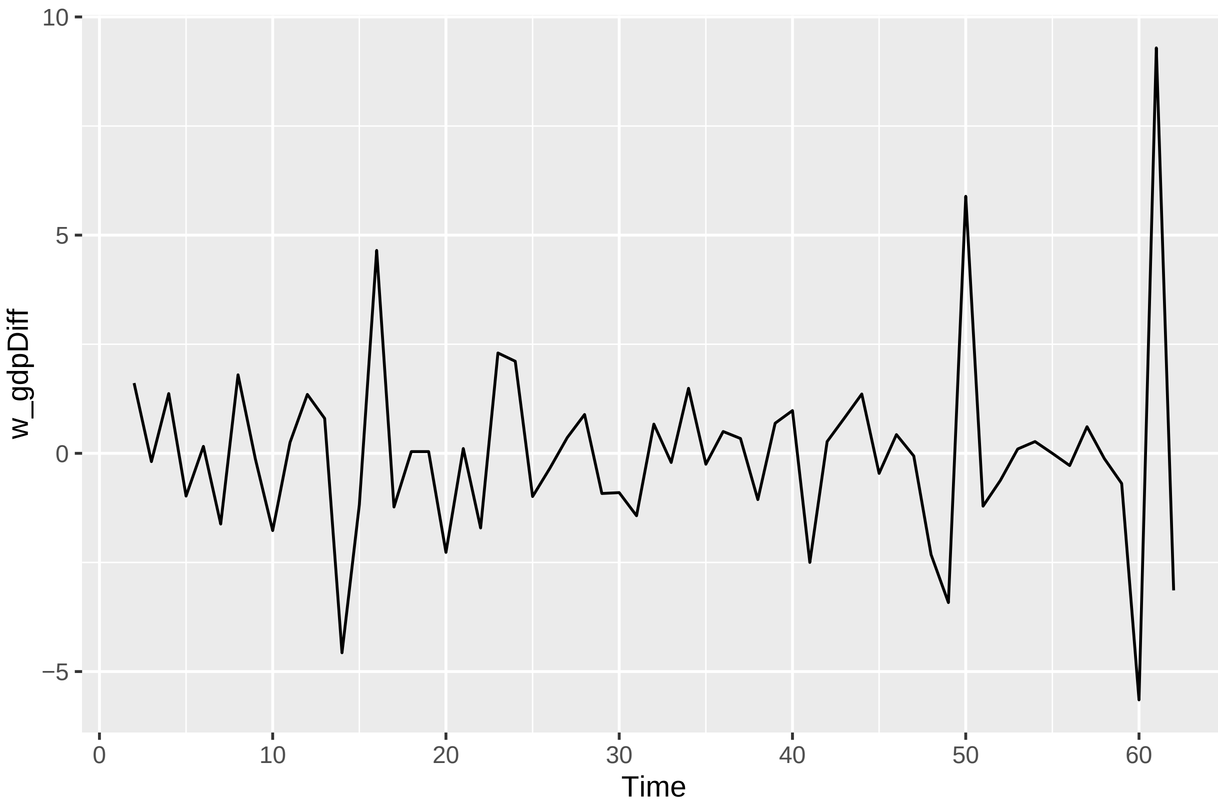





```
w_gdpDiff = diff(w_gdp_ts, lag = 1)

ur.kpss(w_gdpDiff) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0364
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

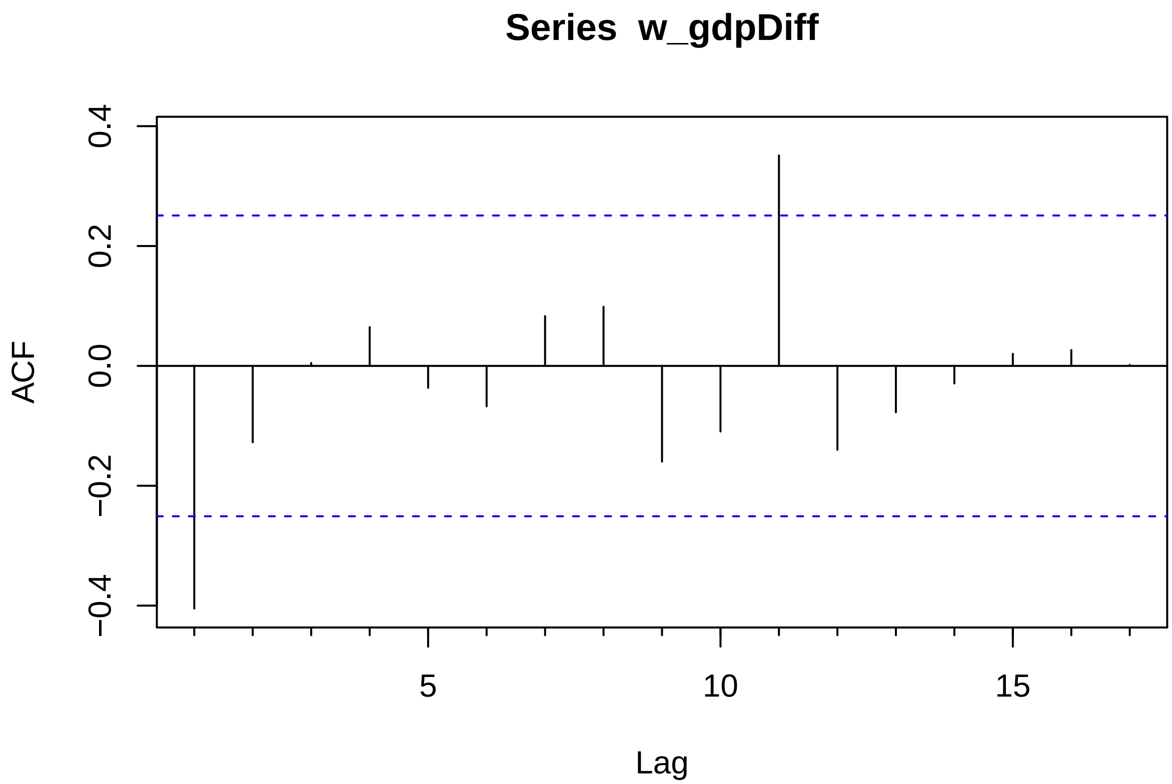
```
autoplot(w_gdpDiff)
```



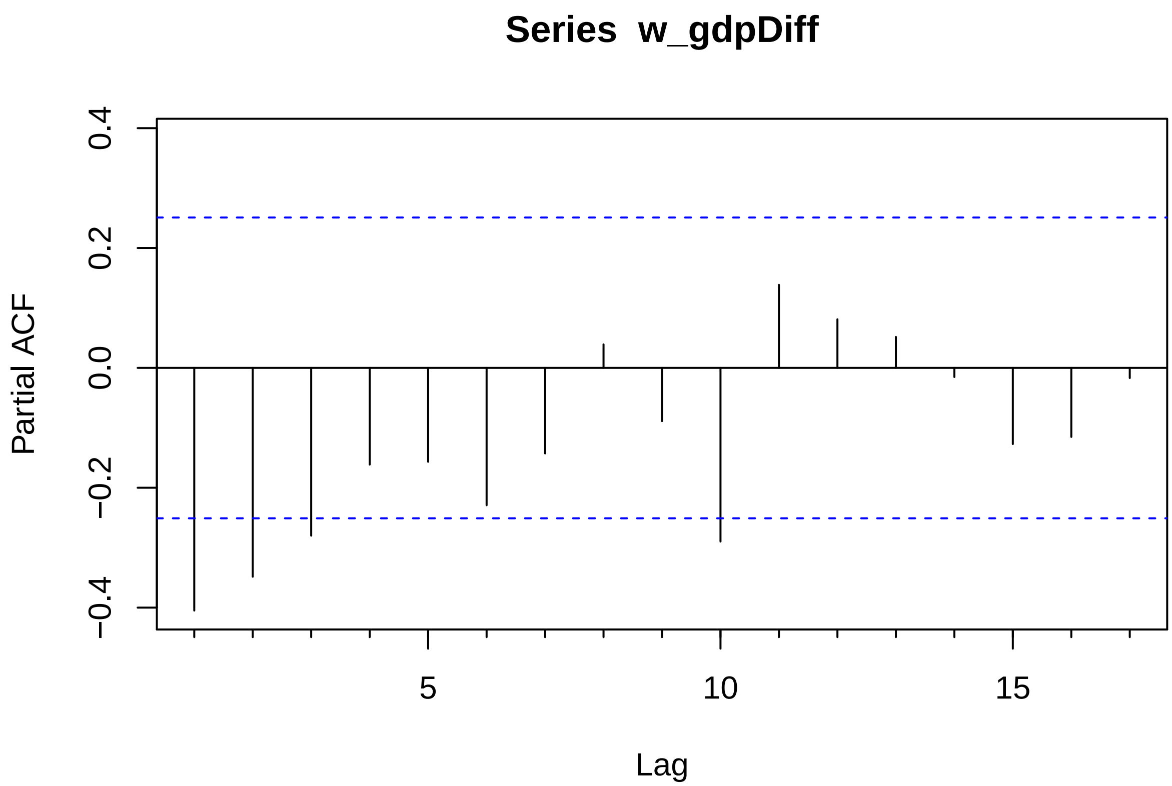
We go ahead and difference the data once and then check the test statistic again Seeing that the test statistic is way lower, this data is stationary.

We then look at the Acf and pacf graphs to see our q and p value for the ARIMA model

```
Acf(w_gdpDiff) #1,11
```




```
Pacf(w_gdpDiff) #1,2,3,10
```

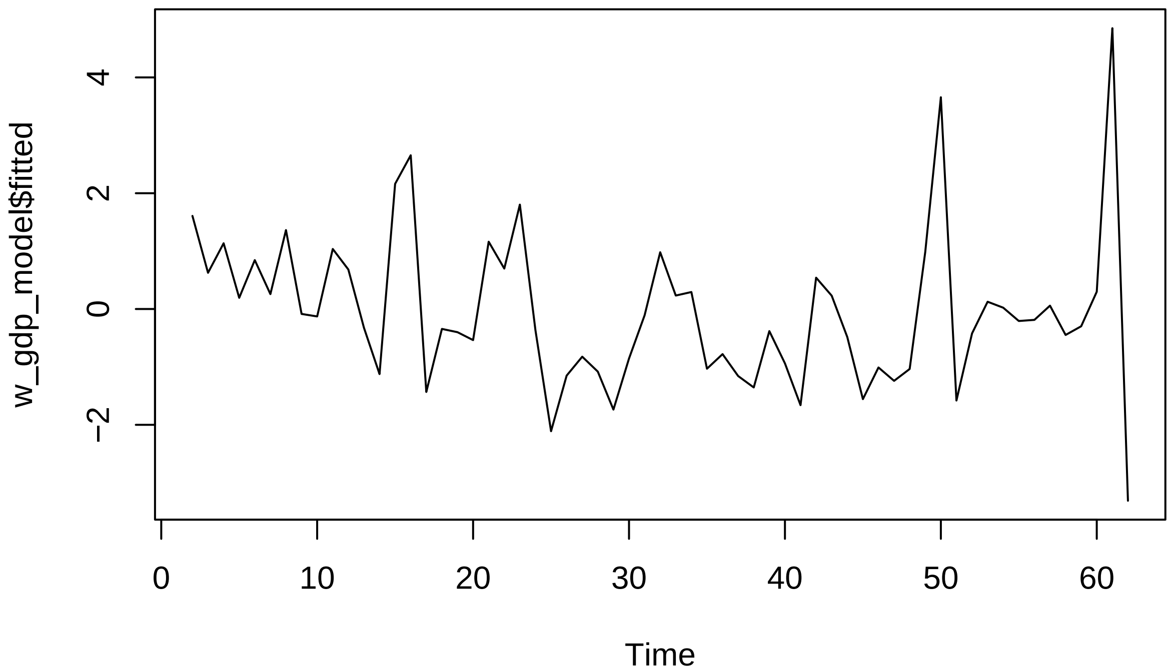



```
w_gdp_model <- Arima(w_gdpDiff, order = c(1, 1, 2), method = "ML")
summary(w_gdp_model)
```

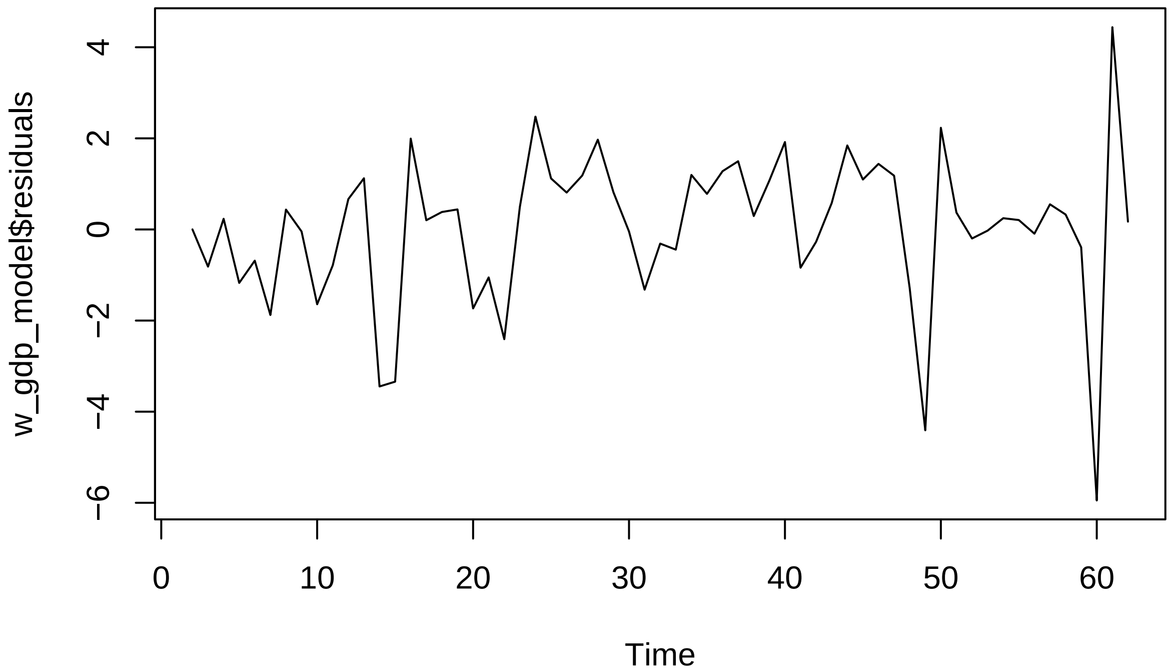
```
## Series: w_gdpDiff
## ARIMA(1,1,2)
##
## Coefficients:
##          ar1      ma1     ma2
##       0.0706  -1.8825  0.8909
## s.e.  0.1463   0.0763  0.0749
##
## sigma^2 = 2.93:  log likelihood = -119.65
## AIC=247.3   AICc=248.03   BIC=255.68
##
## Training set error measures:
##                      ME     RMSE      MAE MPE MAPE      MASE          ACF1
## Training set 0.0410129 1.654553 1.175059 Inf  Inf 0.5157162 -0.003446656
```

Once we find the p and q values, we try them for the ARIMA models to find the best AICc value Using the ARIMA(1,1,2) model gives the best AICc value.

```r
plot(w_gdp_model$fitted)
```




```r
plot(w_gdp_model$residuals)
```

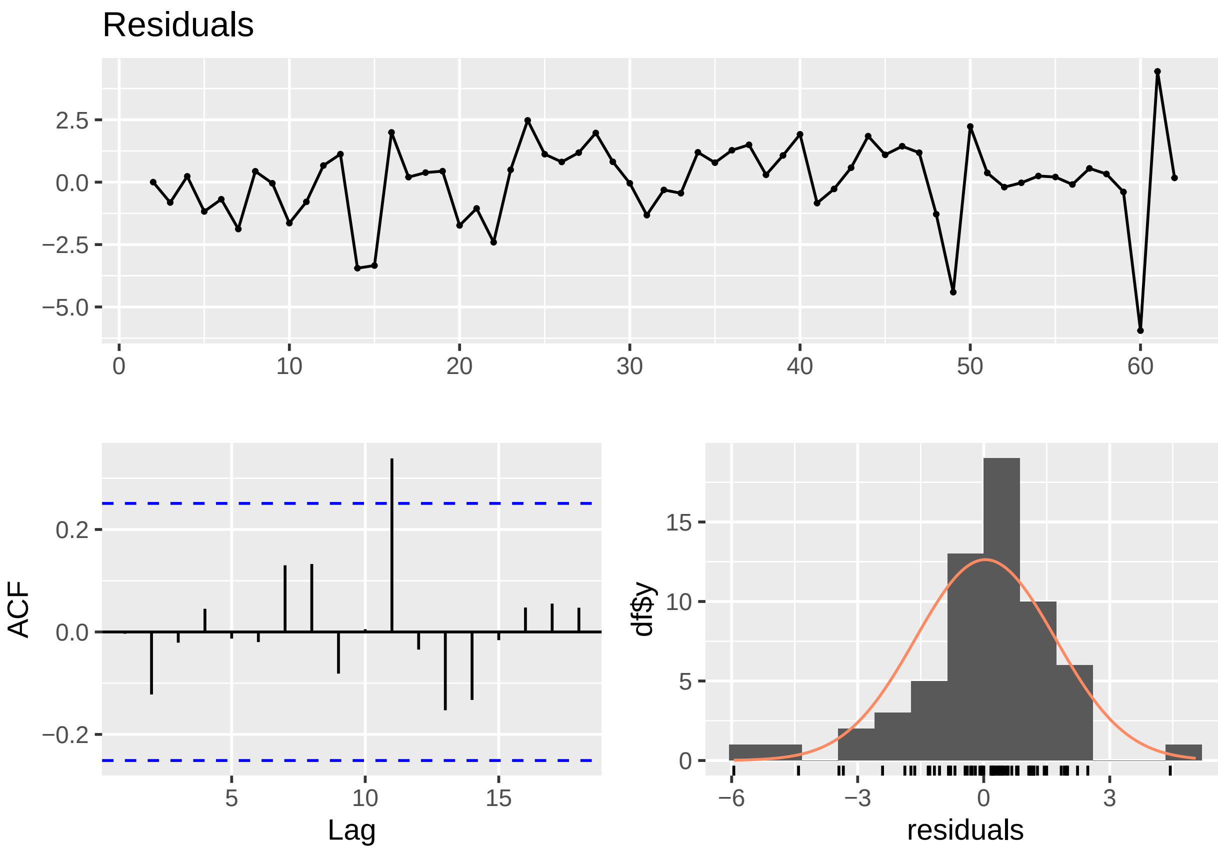



```r
#Check stationary of the residuals
ur.kpss(w_gdp_model$residuals) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.2704
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

We then look at the fitted and residual models and then conduct another kpss test for our model with the residuals which looks good

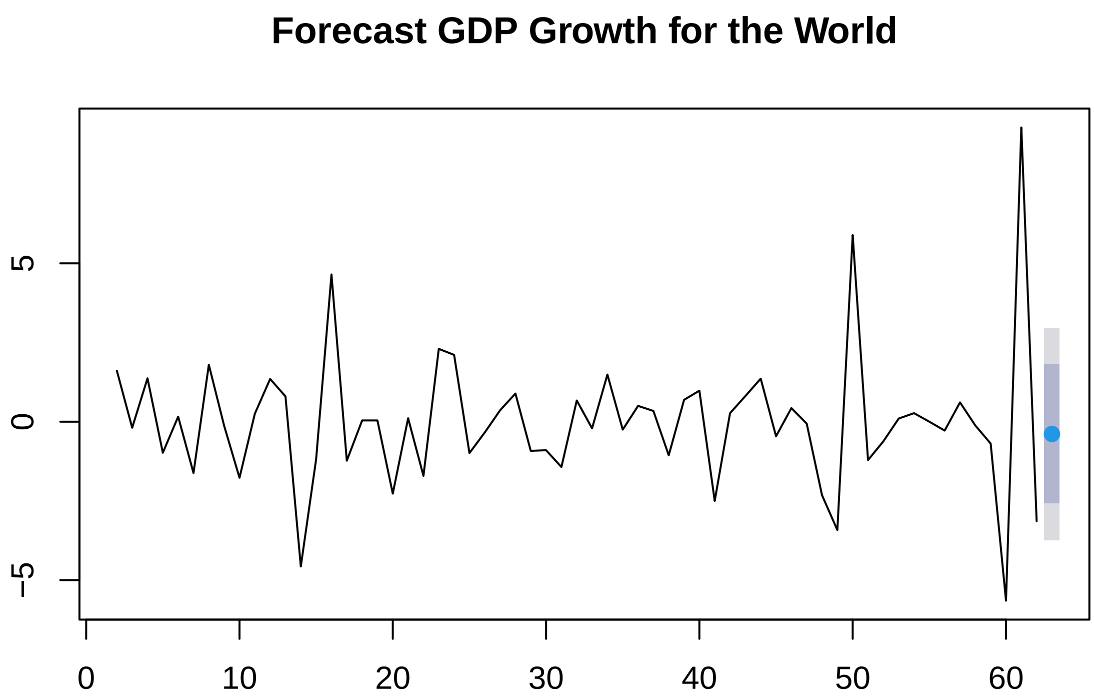
```r
checkresiduals(w_gdp_model$residuals) #This may have to be diffrenced again
```




```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 4.148, df = 10, p-value = 0.9404
##
## Model df: 0.   Total lags used: 10
```

We check if there are any white noise in our data however, it passes the test

```
w.forecast_values <- forecast(w_gdp_model, h=1)
plot(w.forecast_values, main = "Forecast GDP Growth for the World")
```

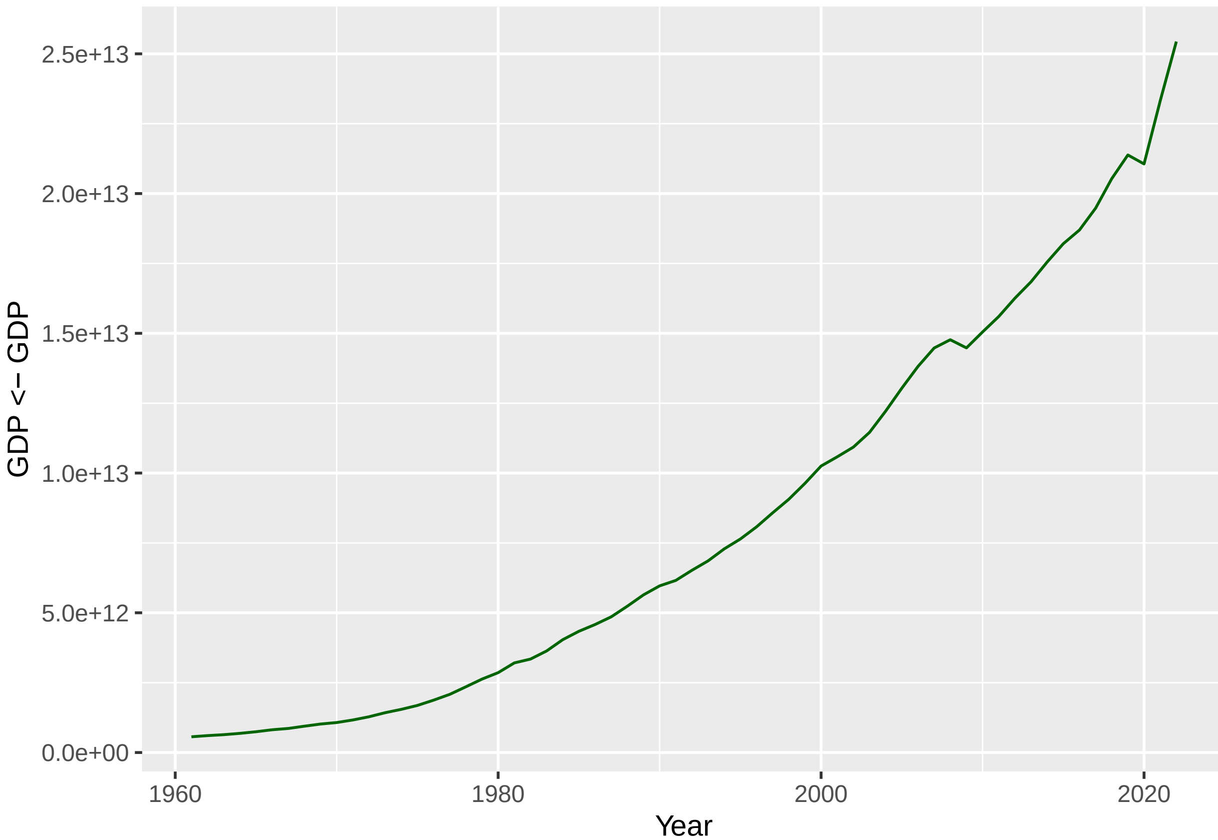


Then we can forecast the data for the future

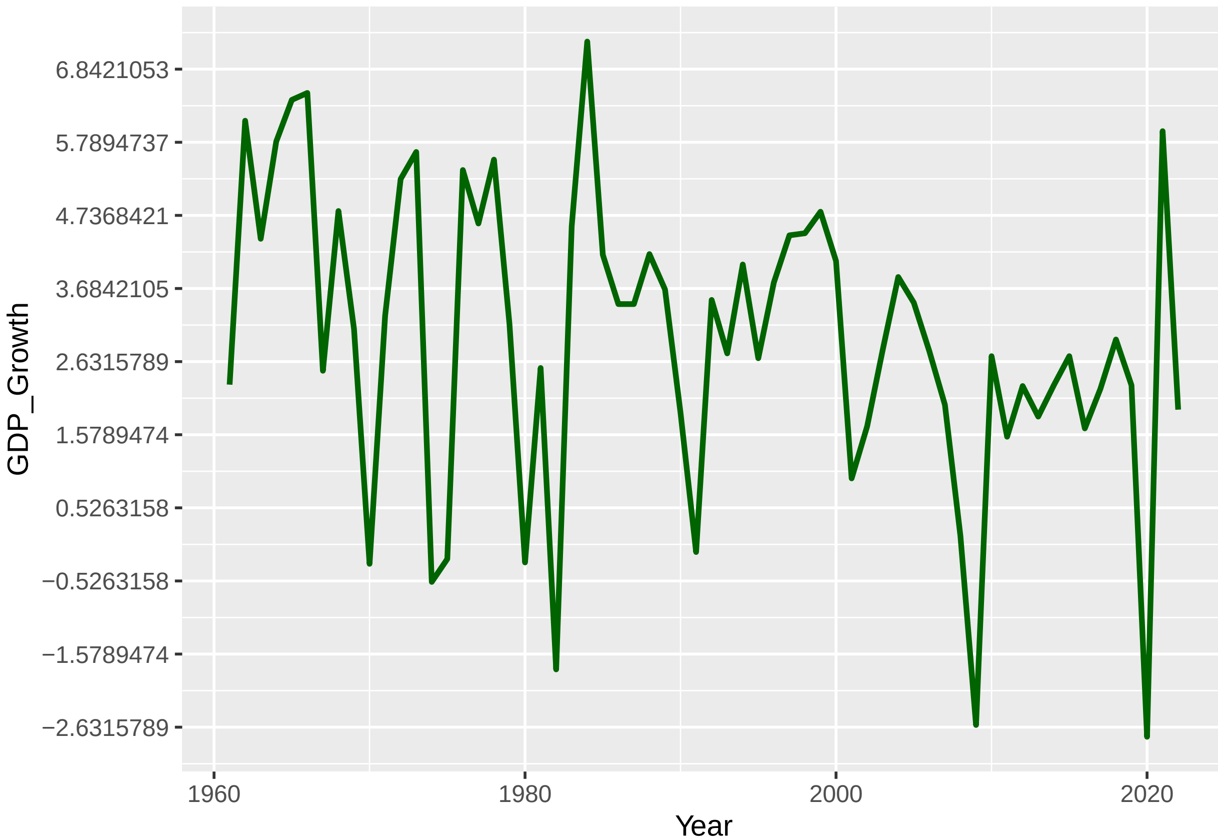
```
us_gdp <- gdp %>%
  filter(Country == "United States") %>%
  select(Year, GDP, GDP_Growth)
```

Next we are looking at the gdp and the gdp growth of the United States

```
ggplot(us_gdp) +
  geom_line(aes(Year, GDP <- GDP, group = 1), color = "darkgreen", linewidth = .5)
```




```r
ggplot(us_gdp, aes(Year, GDP_Growth, group = 1)) +
  geom_line(color = "darkgreen", linewidth = 1) +
  scale_y_continuous(
    breaks = seq(-10, 10, by = 20/19)
  )
```

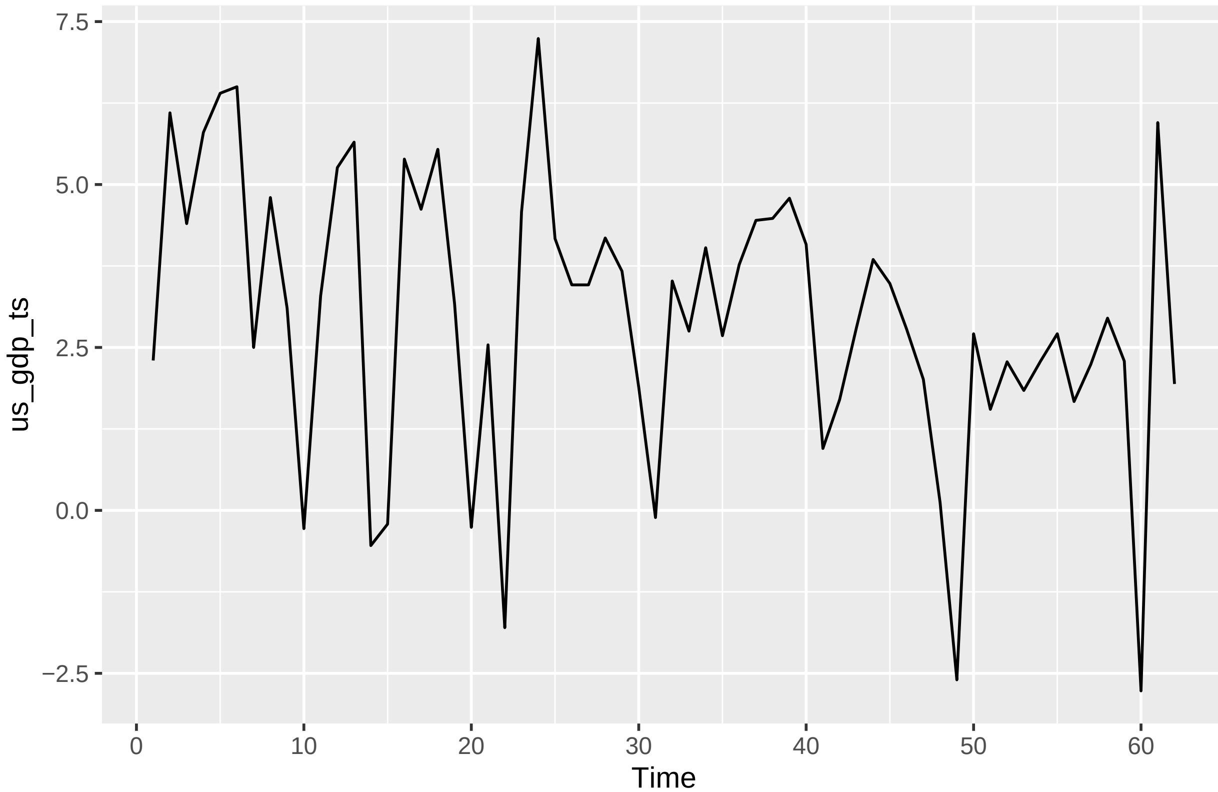


Here is the graph of the gdp growth of the us gdp. The graphs show dips in the data during the 1980s recession, the 2008 recession, and the 2020 COVID19 pandemic.

```
ustrain <- us_gdp[1:50,]
ustest <- us_gdp[51:62,]
usntest <- nrow(ustest)
```

Then we make a train and testing set

```
us_gdp_ts <- ts(us_gdp$GDP_Growth)
autoplot(us_gdp_ts)
```

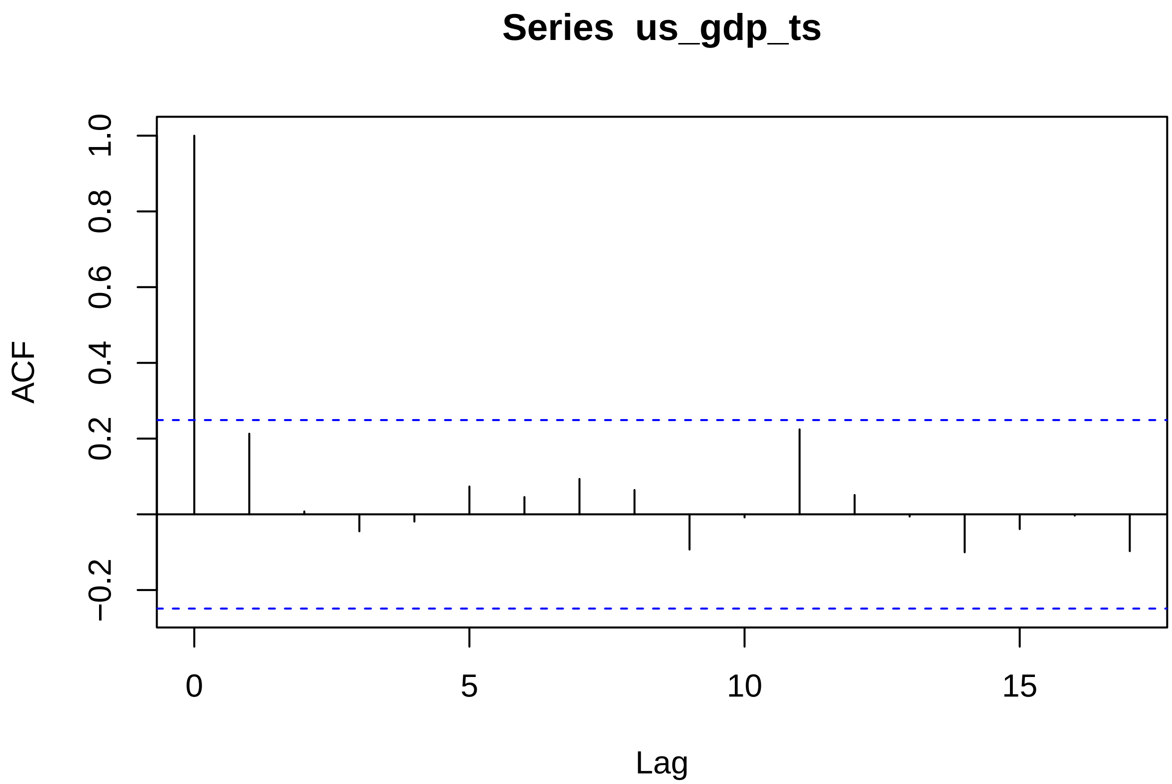


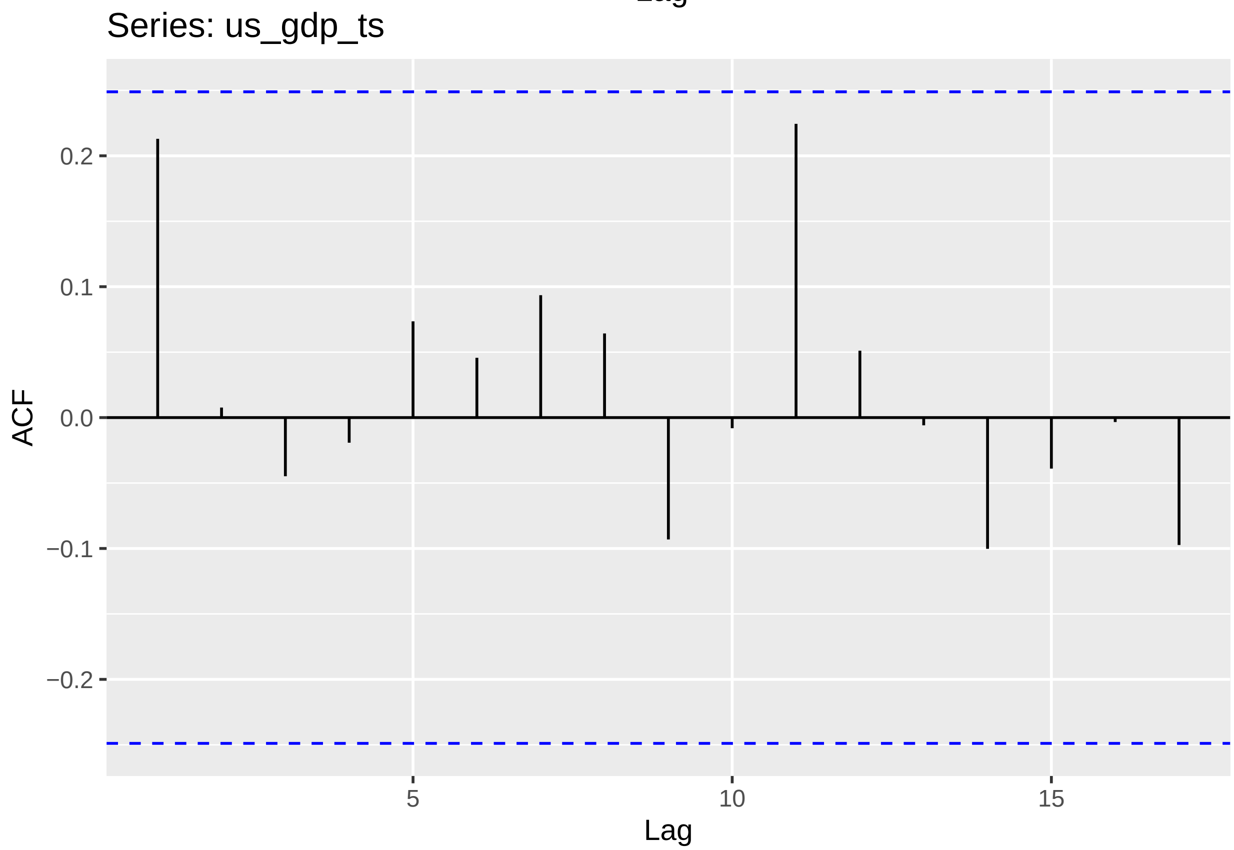

```r
ur.kpss(us_gdp_ts) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.5418
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Then we create a time series model and then check for stationary. We use the kpss test and see if it needs to be diffrenced. Seeing that the test-statisic is near the test data, we can diffrence the data once.

```r
autoplot(acf(us_gdp_ts))
```

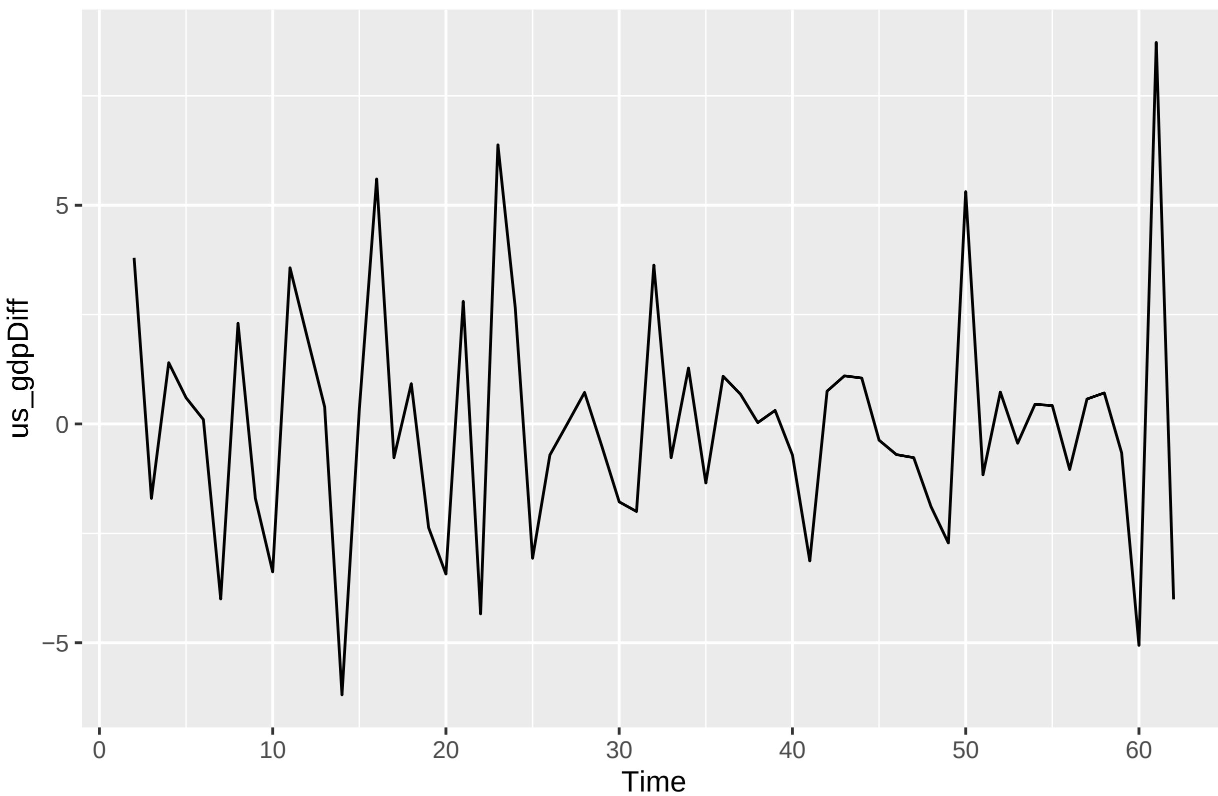





```
us_gdpDiff = diff(us_gdp_ts, lag = 1)

ur.kpss(us_gdpDiff) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0359
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

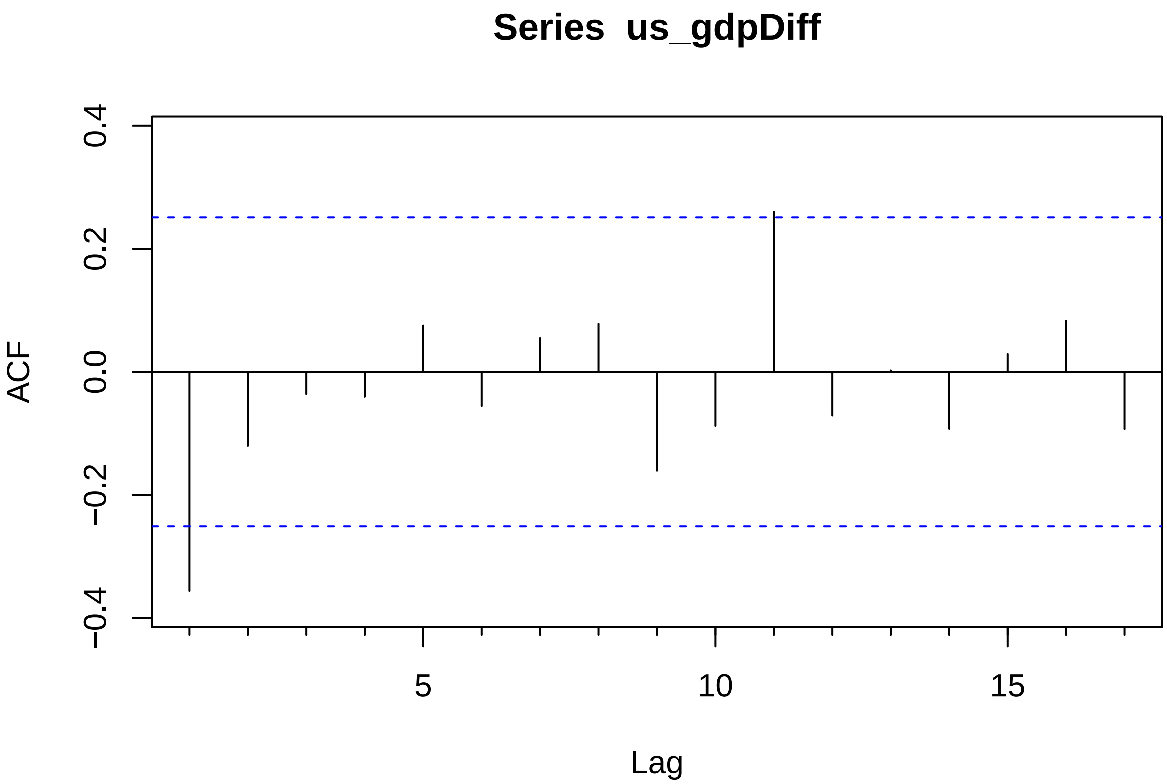
```
autoplot(us_gdpDiff)
```



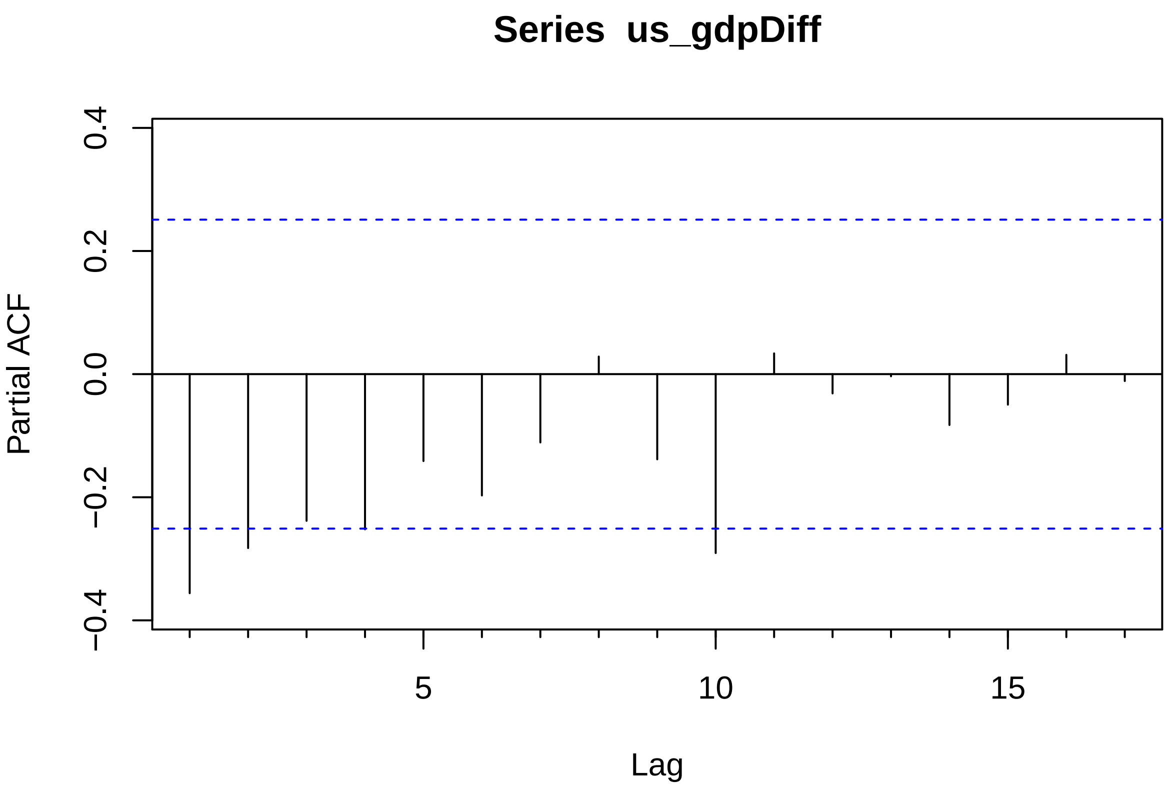
We go ahead and difference the data once and then check the test statistic again Seeing that the test statistic is way lower, this data is stationary.

We then look at the Acf and pacf graphs to see our q and p value for the ARIMA model

```
Acf(us_gdpDiff) #1,11
```




```r
Pacf(us_gdpDiff) #1,2,3,4,10
```

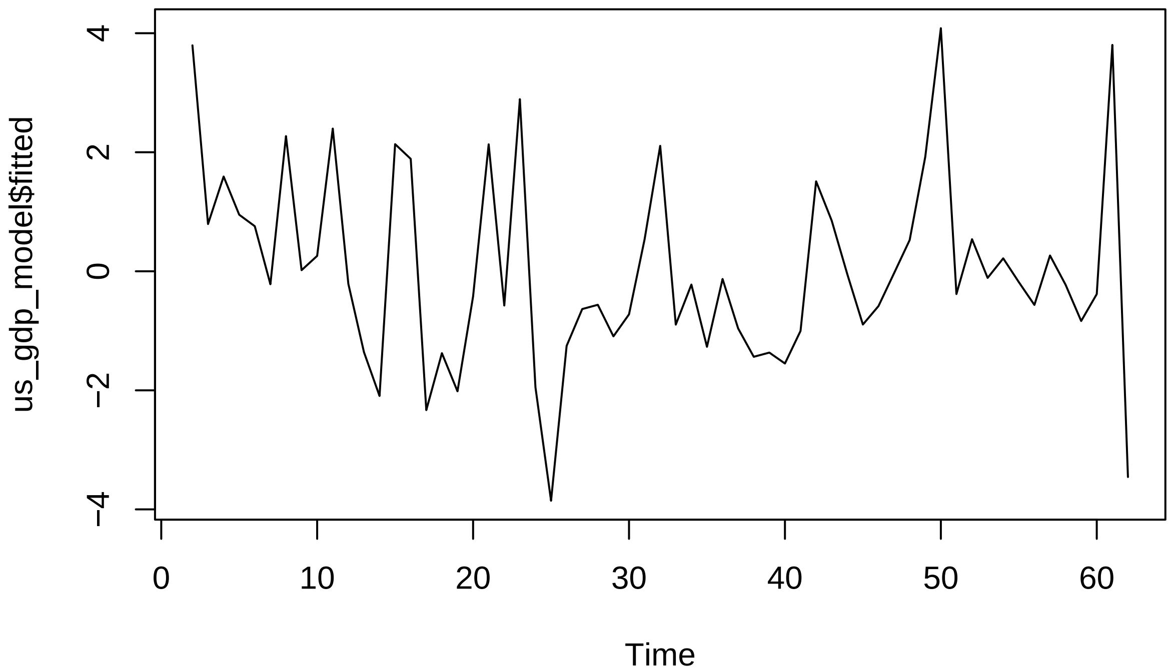



```r
us_gdp_model <- Arima(us_gdpDiff, order = c(1, 1, 2), method = "ML")
summary(us_gdp_model)
```

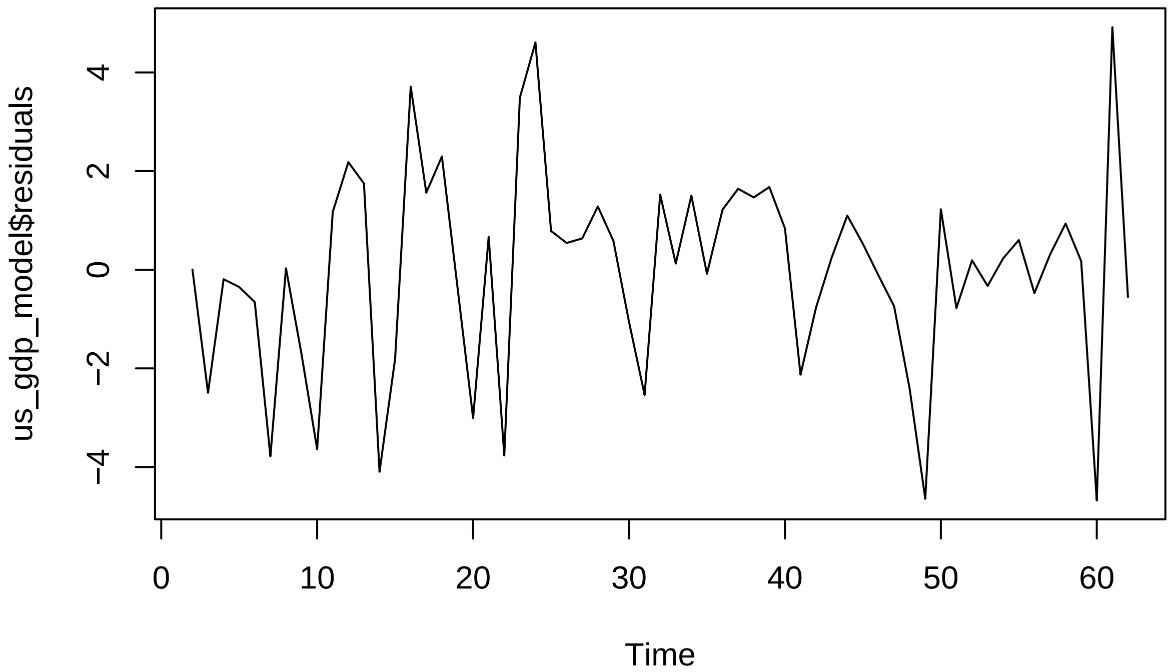
```
## Series: us_gdpDiff
## ARIMA(1,1,2)
##
## Coefficients:
##          ar1      ma1     ma2
##       0.1635  -1.9963  0.9998
## s.e.  0.1332   0.0819  0.0819
##
## sigma^2 = 4.48:  log likelihood = -134.5
## AIC=276.99   AICc=277.72   BIC=285.37
##
## Training set error measures:
##                       ME     RMSE      MAE MPE MAPE      MASE       ACF1
## Training set -0.02305184 2.046136 1.522929 Inf  Inf 0.4786827 0.02275276
```

Once we find the p and q values, we try them for the ARIMA models to find the best AICc value Using the ARIMA(1,1,2) model gives the best AICc value.

```
plot(us_gdp_model$fitted)
```




```
plot(us_gdp_model$residuals)
```

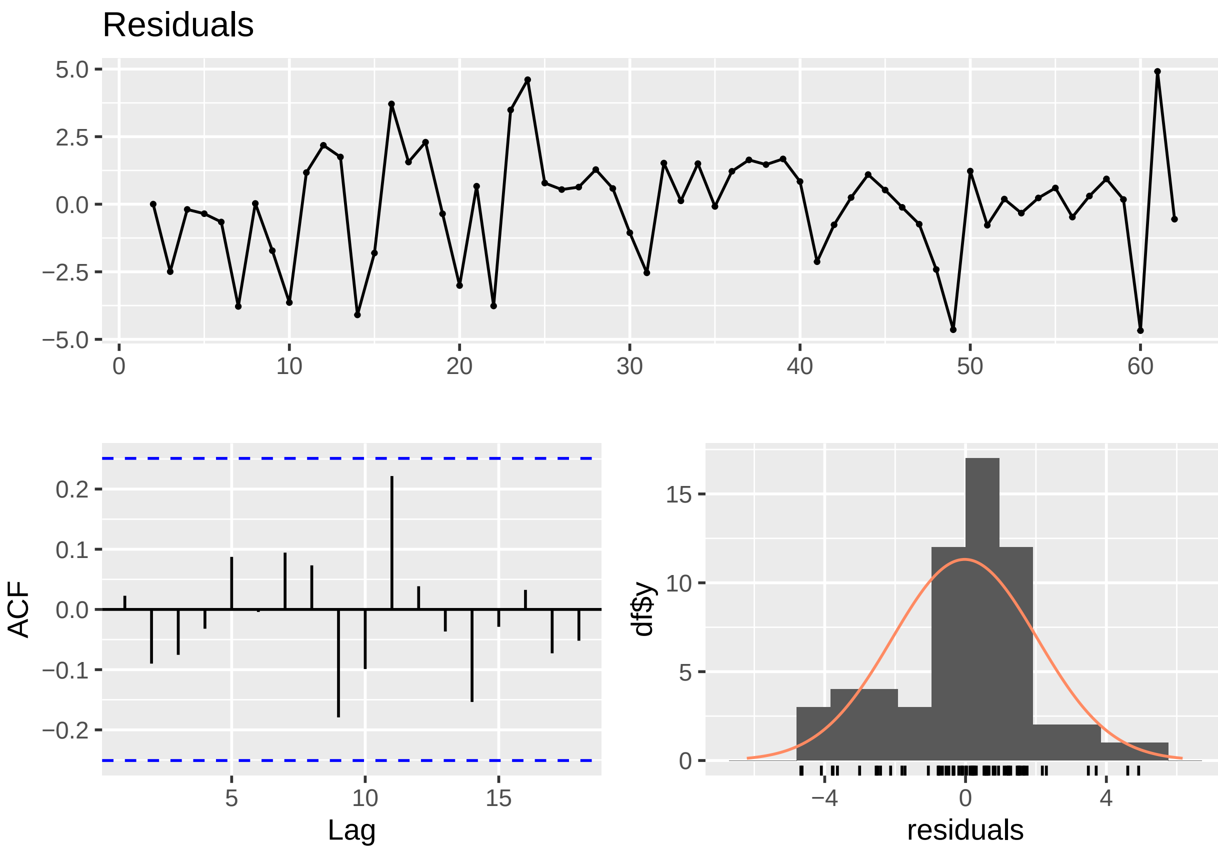



```r
#Check stationary of the residuals
ur.kpss(us_gdp_model$residuals) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.1355
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

We then look at the fitted and residual models and then conduct another kpss test for our model with the residuals which looks good

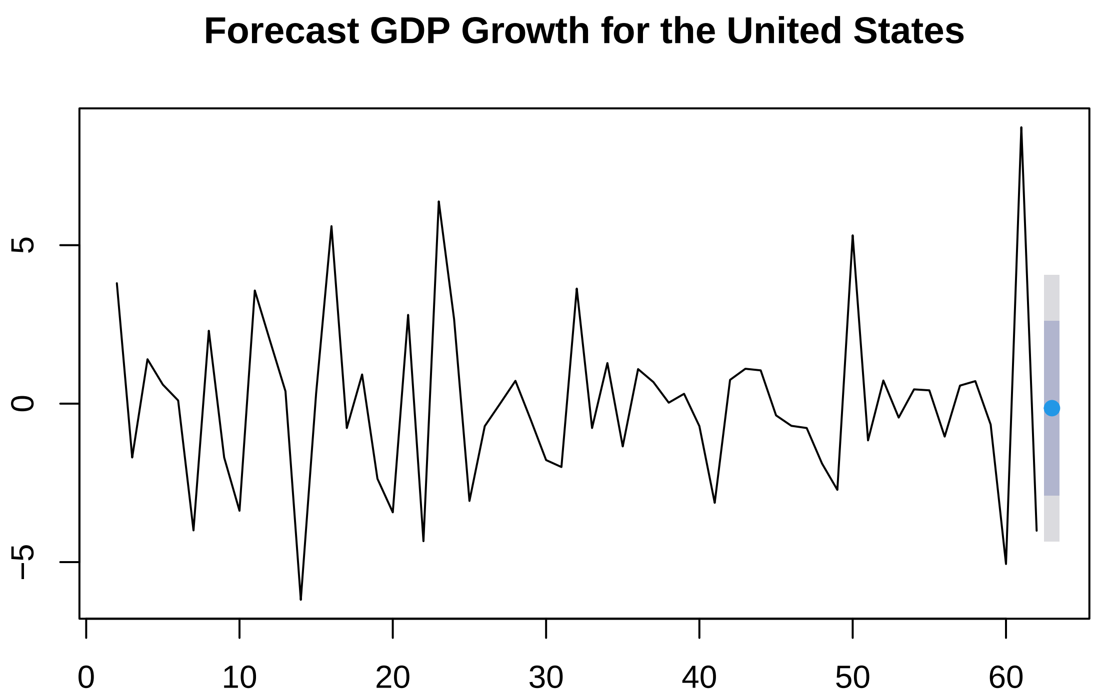
```r
checkresiduals(us_gdp_model$residuals)
```




```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 5.6686, df = 10, p-value = 0.8423
##
## Model df: 0.   Total lags used: 10
```

We check if there are any white noise in our data however, it passes the test

```
us.forecast_values <- forecast(us_gdp_model, h=1)
plot(us.forecast_values, main = "Forecast GDP Growth for the United States")
```

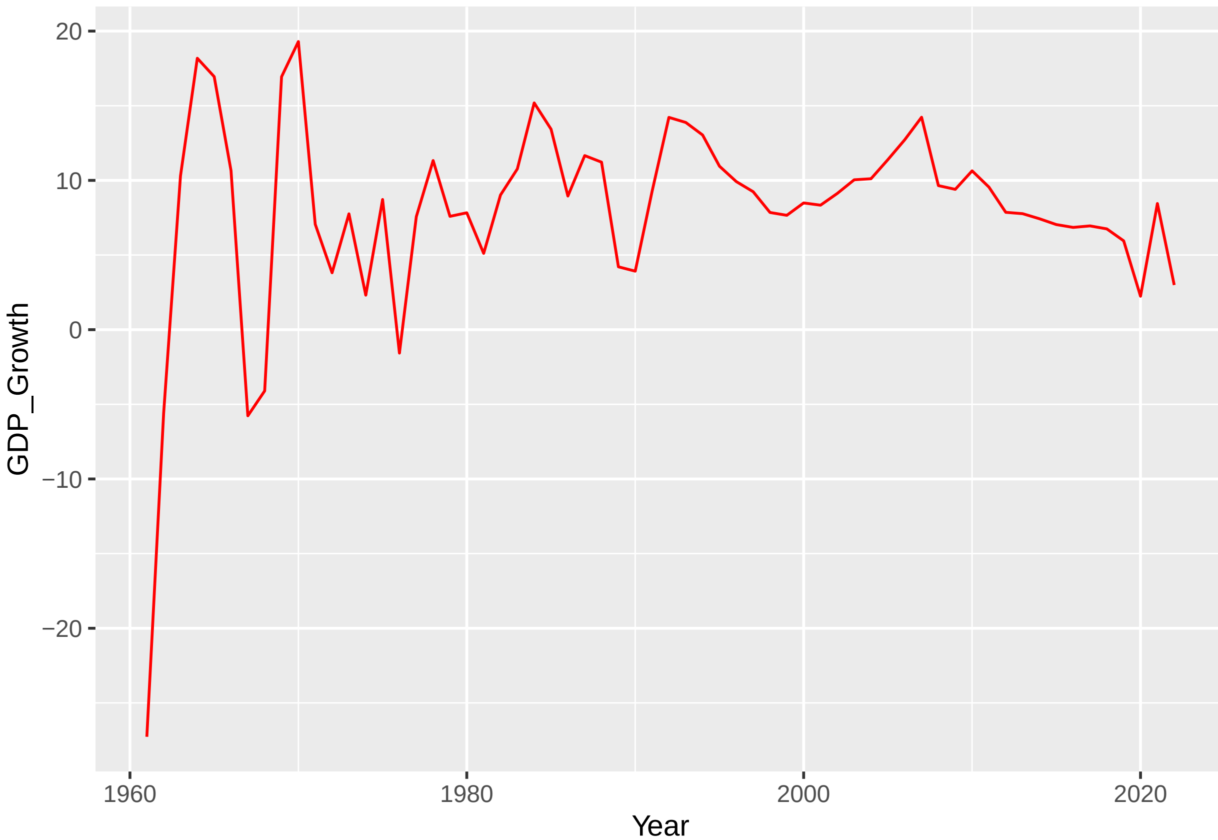


Then we can forecast the data for the future

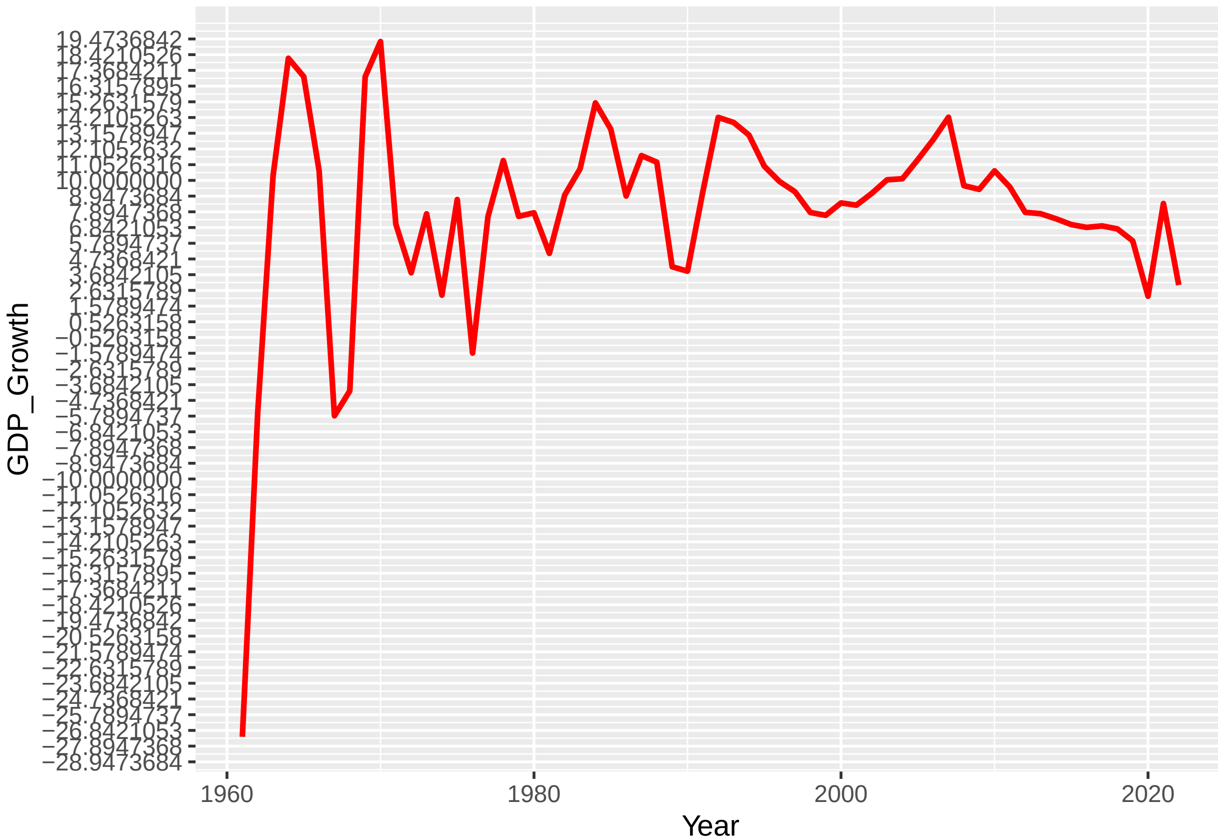
```
china_gdp <- gdp %>%
  filter(Country == "China") %>%
  select(Year, GDP, GDP_Growth)
```

Next is China

```
ggplot(china_gdp) +
  geom_line(aes(Year, GDP_Growth, group = 1), color = "red", linewidth = .5)
```




```r
ggplot(china_gdp, aes(Year, GDP_Growth, group = 1)) +
  geom_line(color = "red", linewidth = 1) +
  scale_y_continuous(
    breaks = seq(-30, 20, by = 20/19)
  )
```

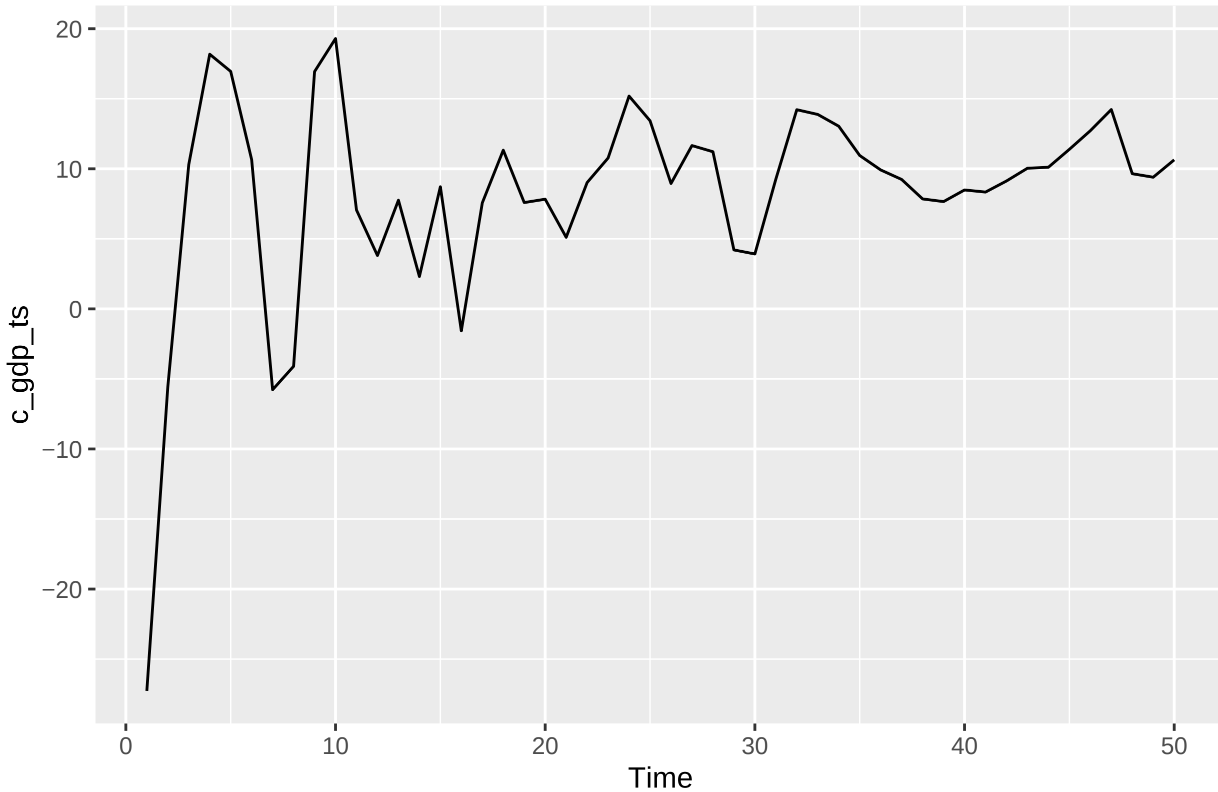


Here are the graphs for China. You can see a large rise in the early 1960s due to China's transition from agricultural to industrial

```r
ctrain <- china_gdp[1:50,]
ctest <- china_gdp[51:62,]
nctest <- nrow(ctest)
```

Then we make a train and testing set

```r
c_gdp_ts <- ts(ctrain$GDP_Growth)
autoplot(c_gdp_ts)
```




```r
ur.kpss(c_gdp_ts) %>% summary()
```

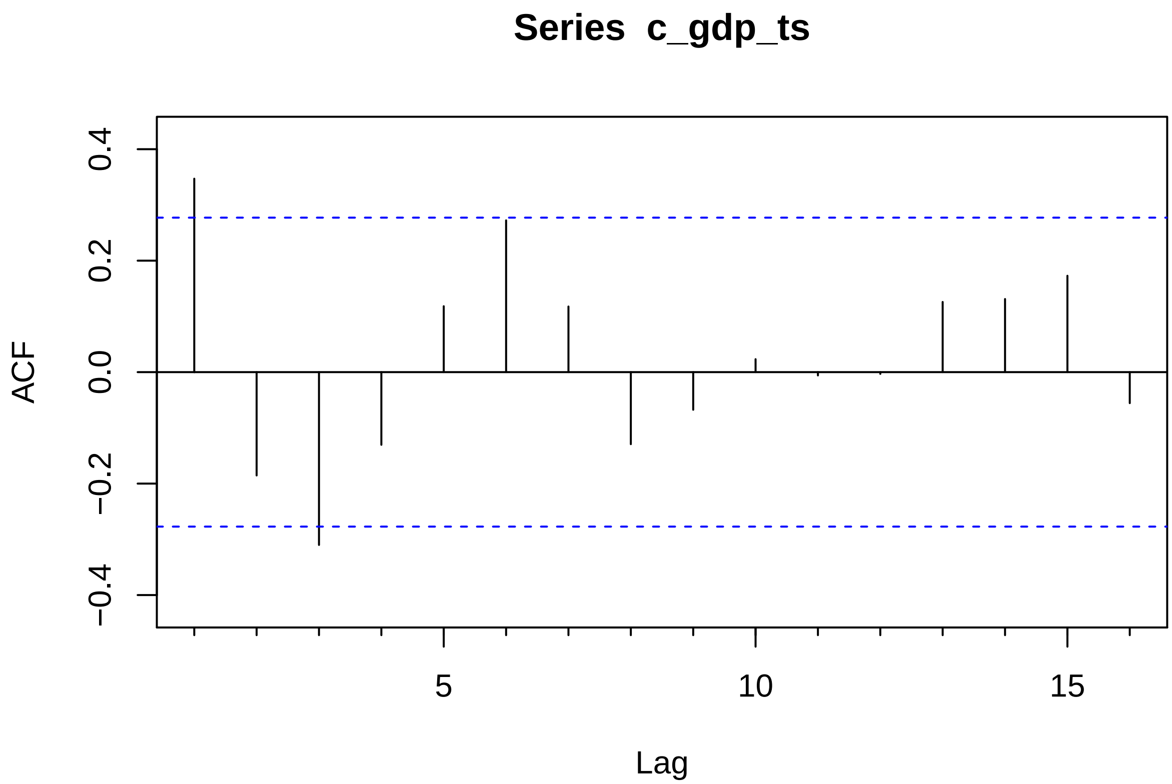
```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.4514
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Then we create a time series model and then check for stationary. We use the kpss test and see if it needs to be diffrenced. Seeing that the test-statisic is near the test data, we can diffrence the data once.

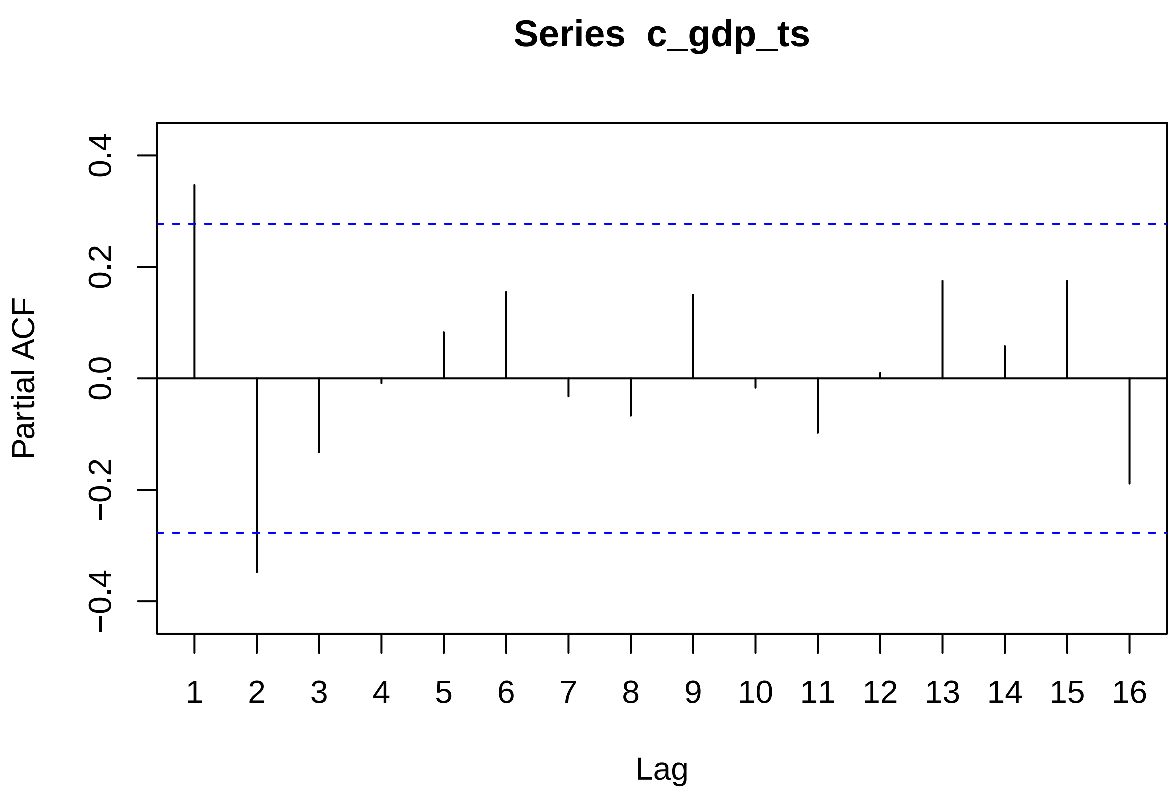
Looking at the data, it seems that the data does not need to be diffrenced

We then look at the Acf and pacf graphs to see our q and p value for the ARIMA model

```r
Acf(c_gdp_ts) #1,3,6
```




```r
Pacf(c_gdp_ts) #1,2
```



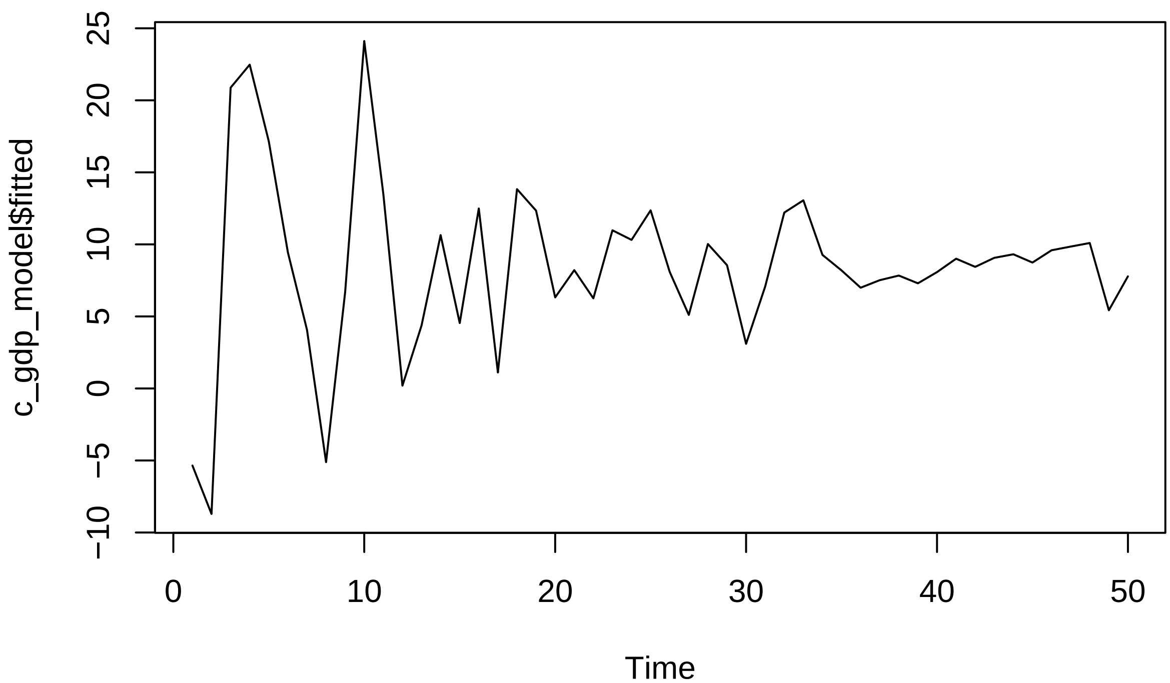

```r
#d=0

#(3,0,1) has been the best so far (AICc = 327.05)
```

```r
c_gdp_model <- Arima(c_gdp_ts, order = c(3, 0, 1), method = "ML")
summary(c_gdp_model)
```

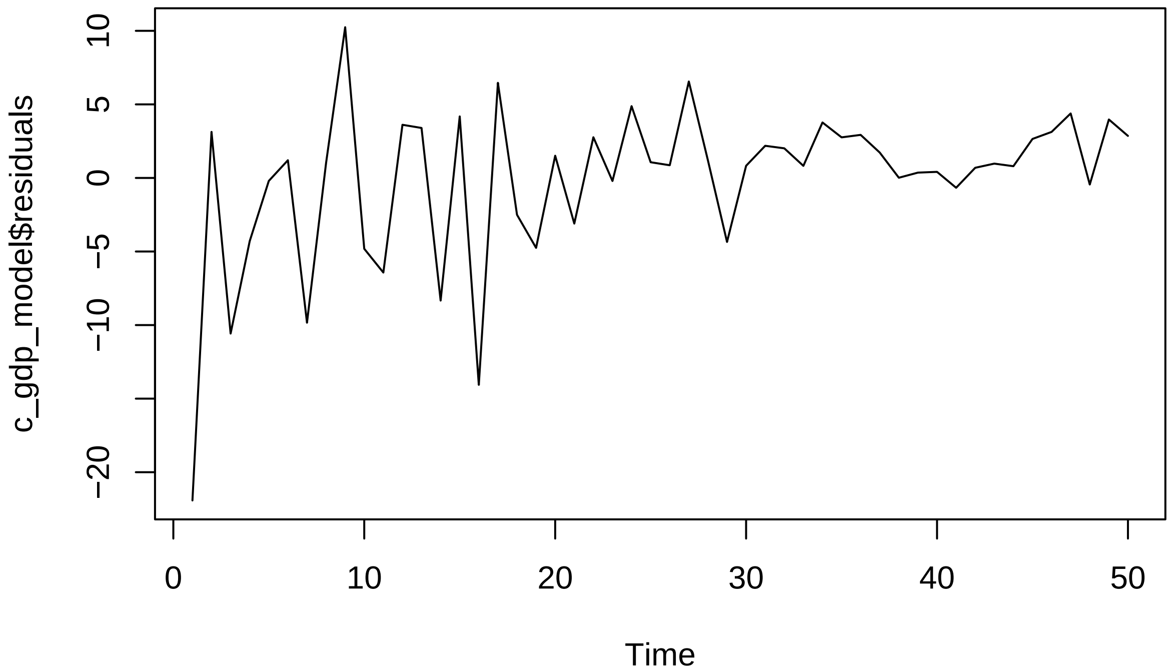
```
## Series: c_gdp_ts
## ARIMA(3,0,1) with non-zero mean
##
## Coefficients:
##          ar1      ar2      ar3      ma1    mean
##       0.9385  -0.7180  -0.0170  -0.1525  8.4569
## s.e.  0.7121   0.6214   0.4989   0.6915  0.8549
##
## sigma^2 = 32.99:  log likelihood = -156.55
## AIC=325.09   AICc=327.05   BIC=336.57
##
## Training set error measures:
##                      ME     RMSE      MAE      MPE     MAPE      MASE
## Training set -0.1429325 5.448842 3.717147 21.16002 59.09941 0.8582207
##                    ACF1
## Training set -0.1725466
```

Once we find the p and q values, we try them for the ARIMA models to find the best AICc value Using the ARIMA(3,0,1) model gives the best AICc value.

```r
plot(c_gdp_model$fitted)
```




```r
plot(c_gdp_model$residuals)
```

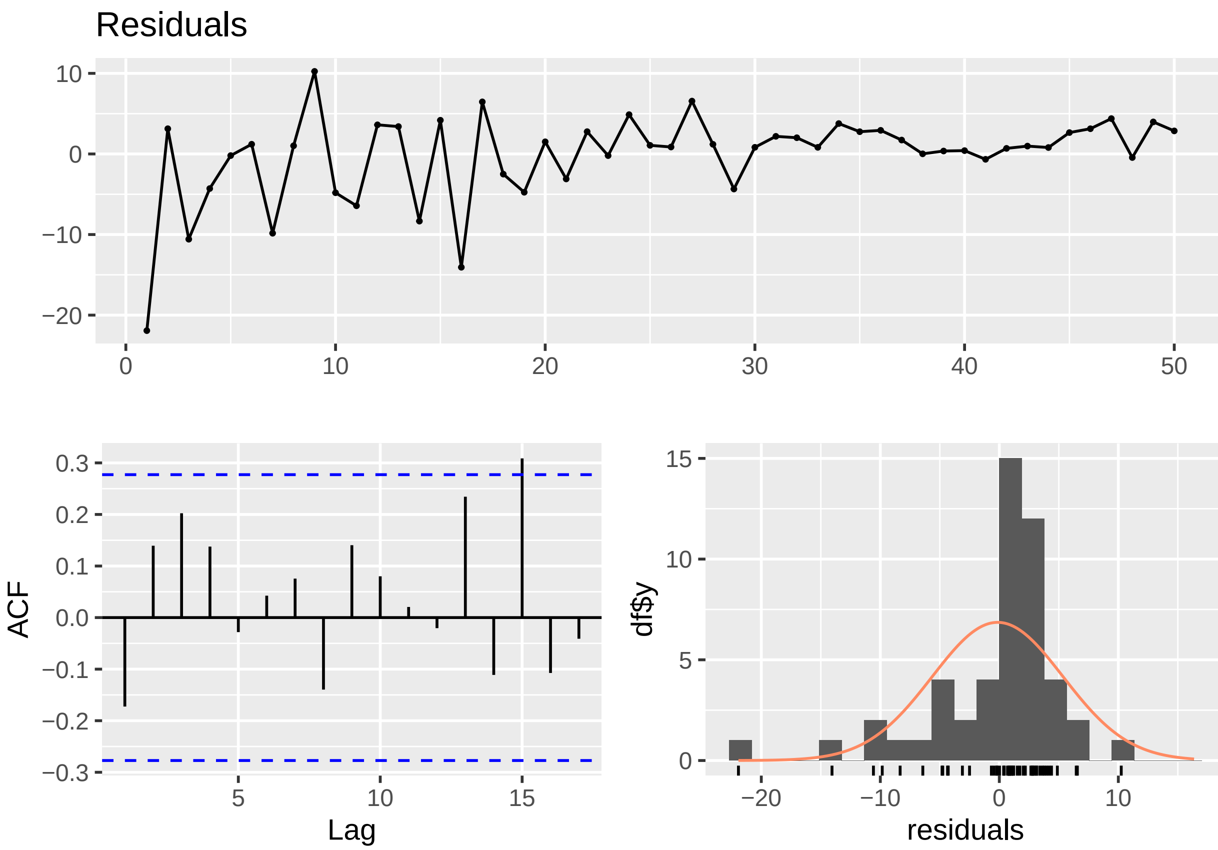



```r
#Check stationary of the residuals
ur.kpss(c_gdp_model$residuals) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.7965
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

We then look at the fitted and residual models and then conduct another kpss test for our model with the residuals which looks good

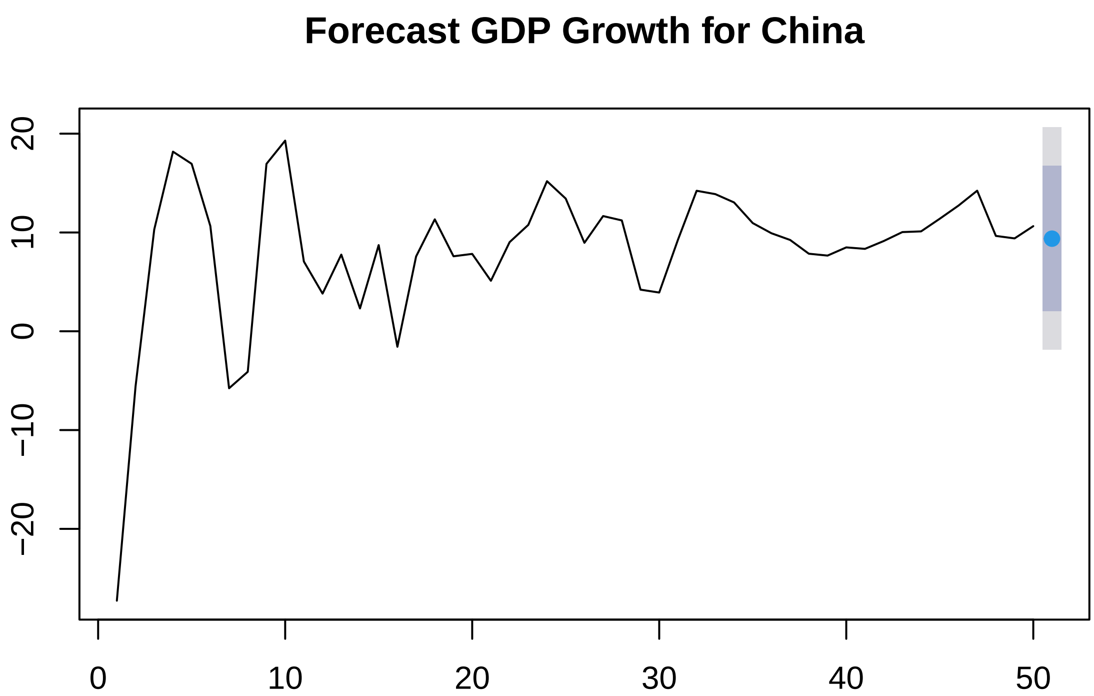
```r
checkresiduals(c_gdp_model$residuals)
```




```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 9.3385, df = 10, p-value = 0.5003
##
## Model df: 0.   Total lags used: 10
```

Look for white noise

```
c.forecast_values <- forecast(c_gdp_model, h=1)
plot(c.forecast_values, main = "Forecast GDP Growth for China")
```

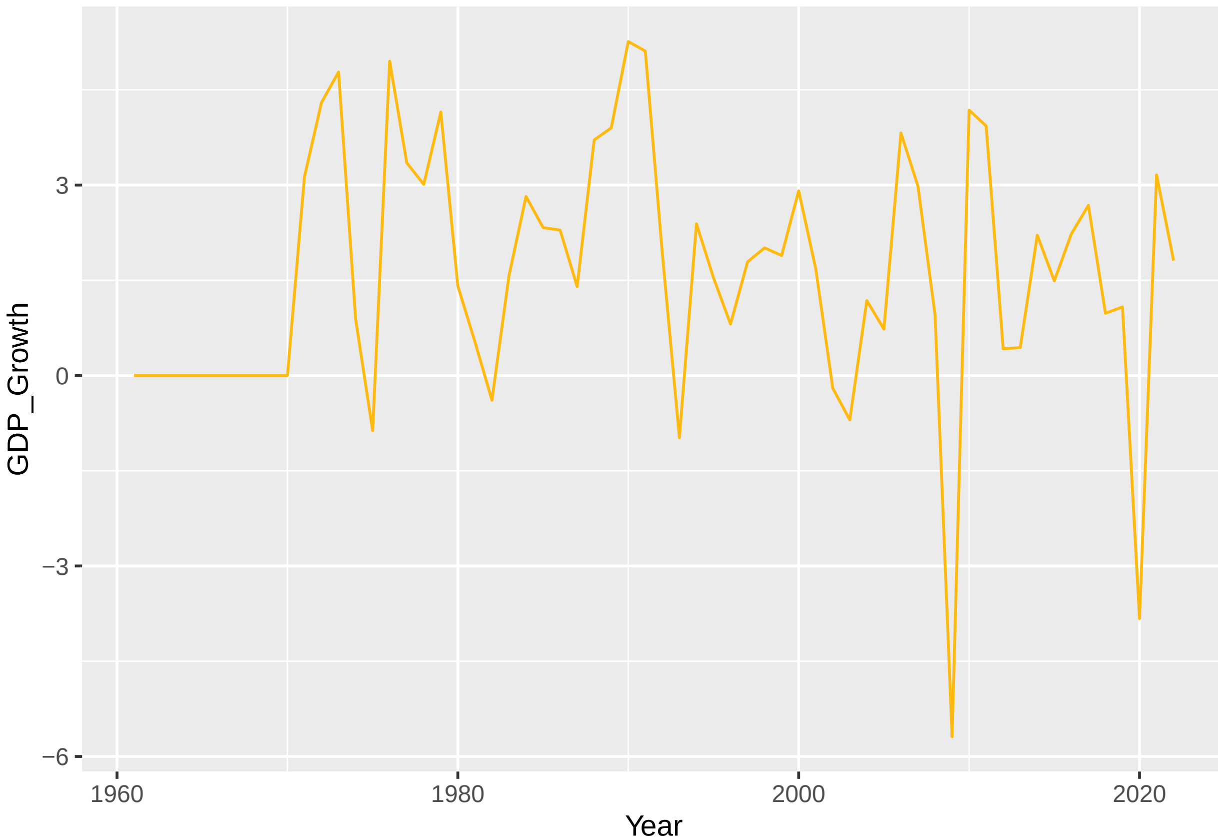


forecast the data for the future

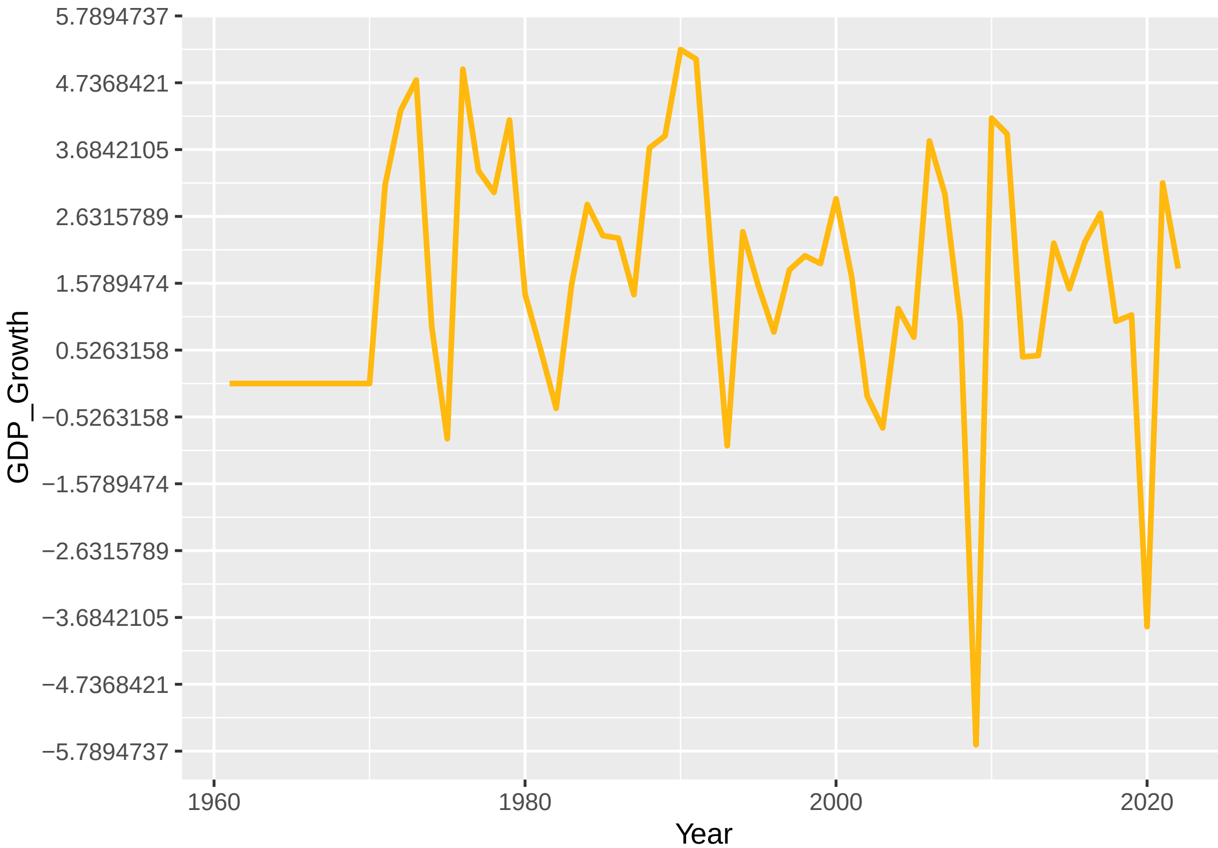
```
germany_gdp <- gdp %>%
  filter(Country == "Germany") %>%
  select(Year, GDP, GDP_Growth)
```

Here is the data for Germany

```
ggplot(germany_gdp) +
  geom_line(aes(Year, GDP_Growth, group = 1), color = "darkgoldenrod1", linewidth = .5)
```




```r
ggplot(germany_gdp, aes(Year, GDP_Growth, group = 1)) +
  geom_line(color = "darkgoldenrod1", linewidth = 1) +
  scale_y_continuous(
    breaks = seq(-10, 10, by = 20/19)
  )
```

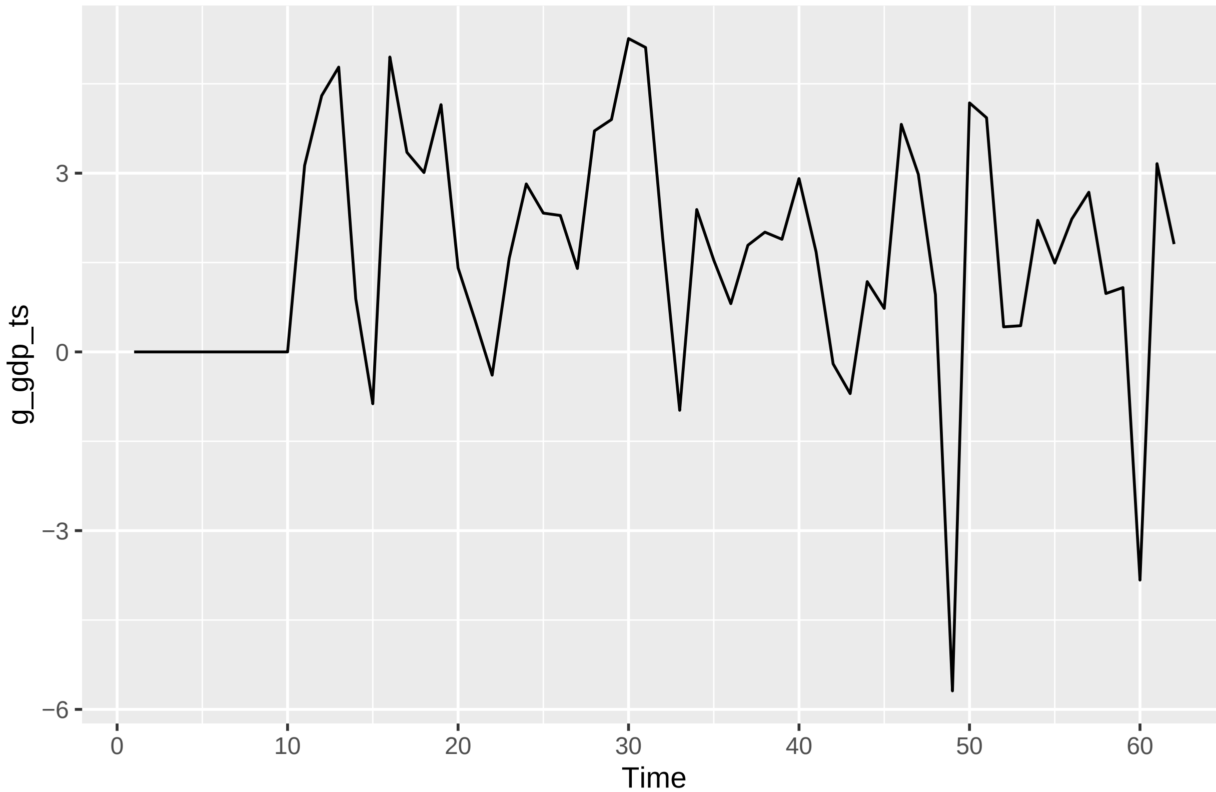


Here are the graphs

```
gtrain <- germany_gdp[1:50,]
gtest <- germany_gdp[51:62,]
gntest <- nrow(gtest)
```

Then we make a train and testing set

```
g_gdp_ts <- ts(germany_gdp$GDP_Growth)
autoplot(g_gdp_ts)
```

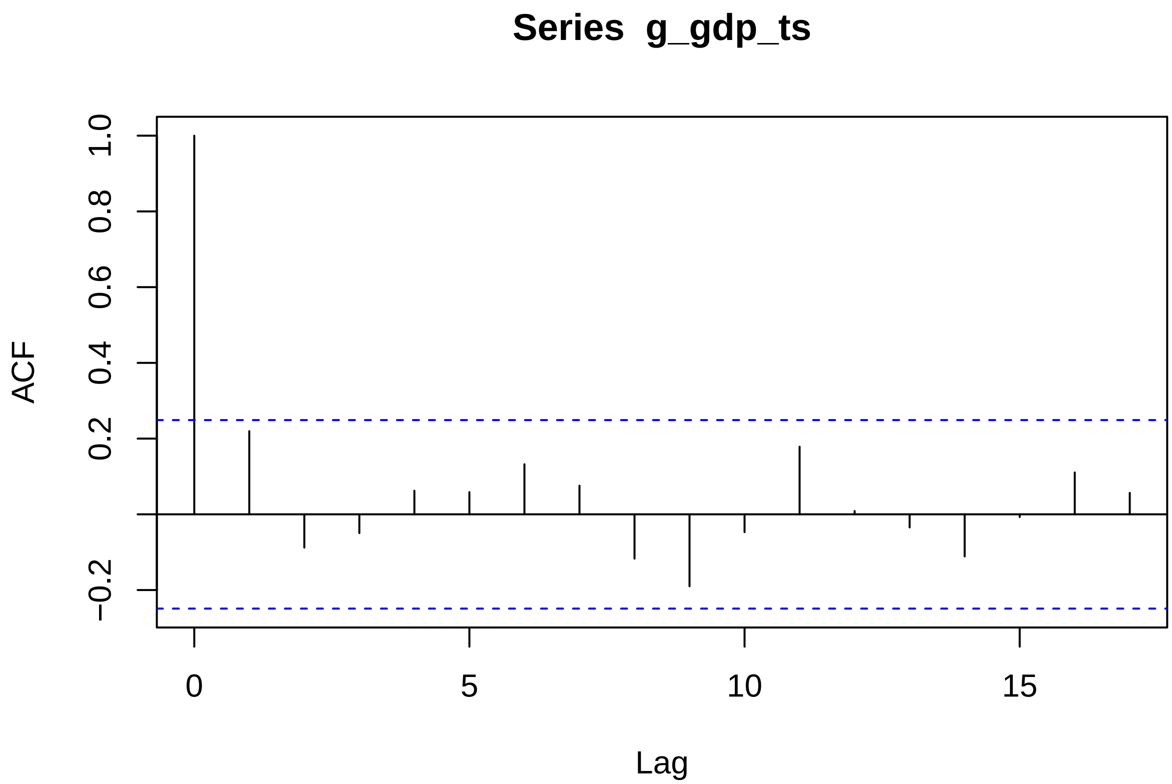


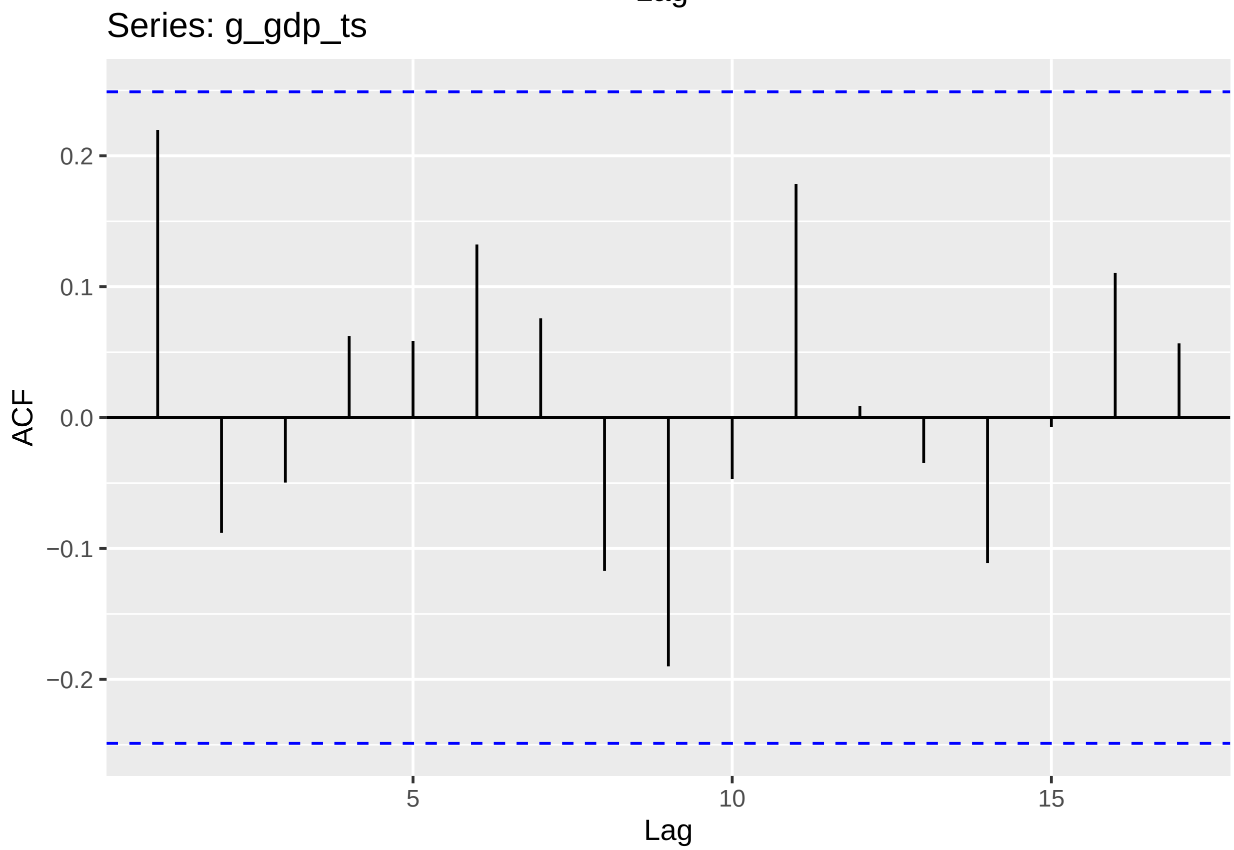

```r
ur.kpss(g_gdp_ts) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.1566
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Then we create a time series model and then check for stationary. We use the kpss test and see if it needs to be diffrenced. Seeing that the test-statisic is near the test data, we can diffrence the data once.

```r
autoplot(acf(g_gdp_ts))
```

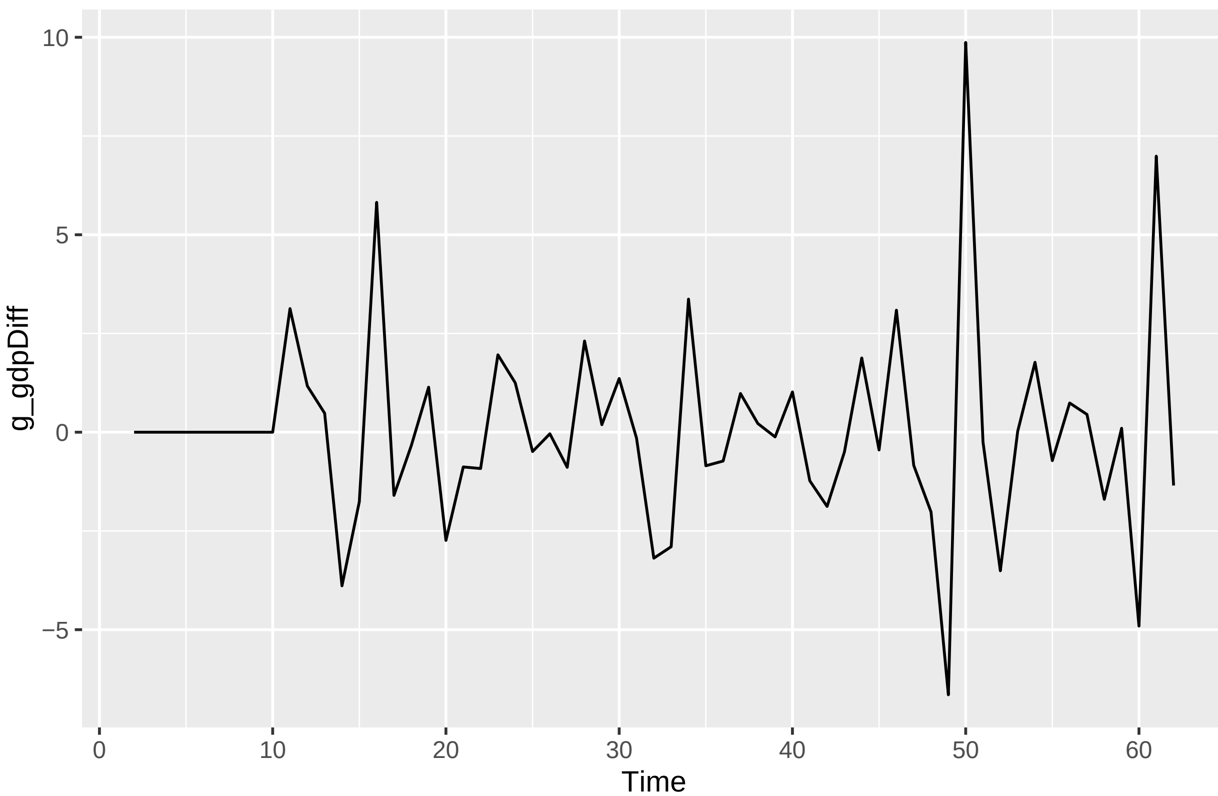





```r
g_gdpDiff = diff(g_gdp_ts, lag = 1)

ur.kpss(g_gdpDiff) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0419
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

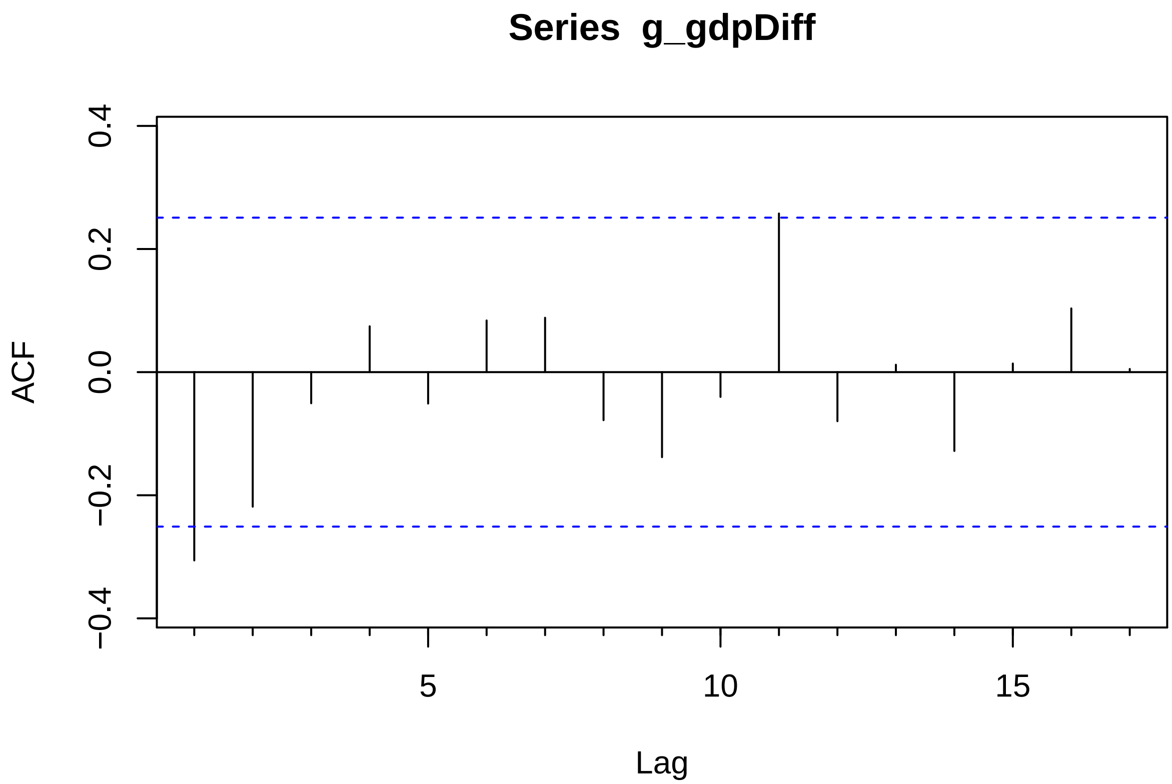
```
autoplot(g_gdpDiff)
```



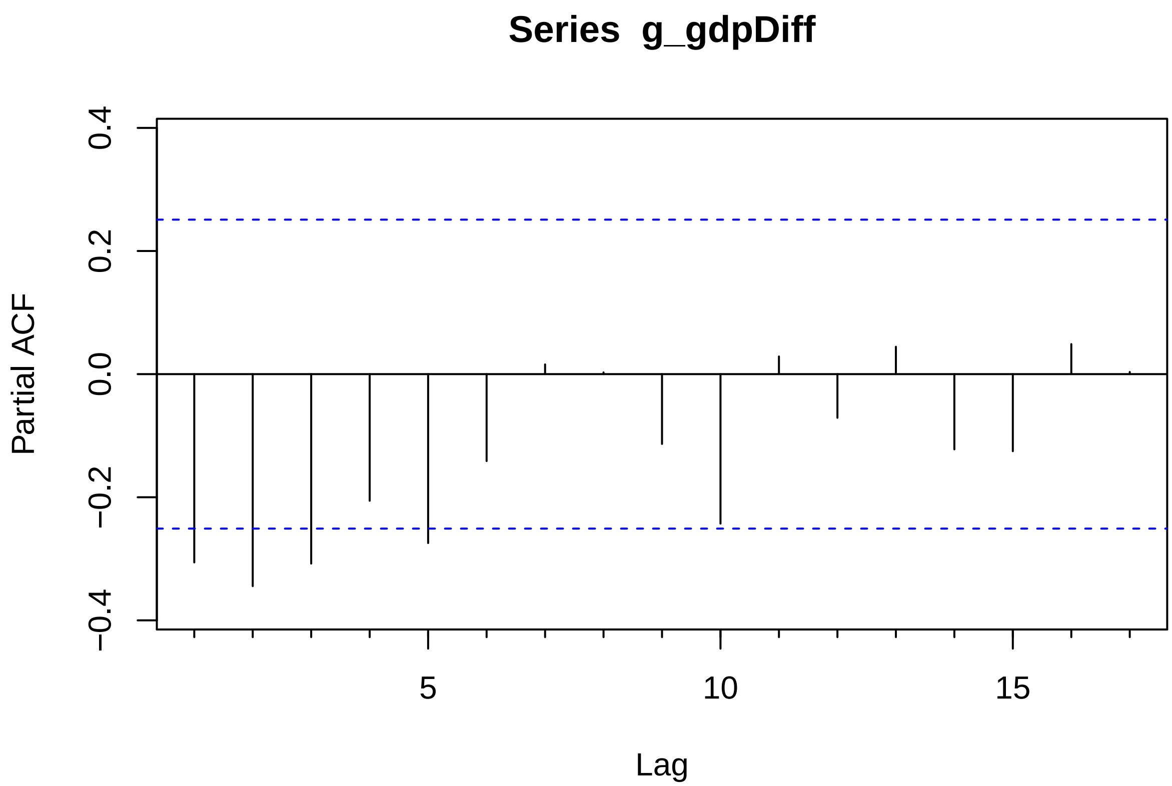
We go ahead and difference the data once and then check the test statistic again Seeing that the test statistic is way lower, this data is stationary.

We then look at the Acf and pacf graphs to see our q and p value for the ARIMA model

```
Acf(g_gdpDiff) #1,11
```




```r
Pacf(g_gdpDiff) #1,2,3,5
```



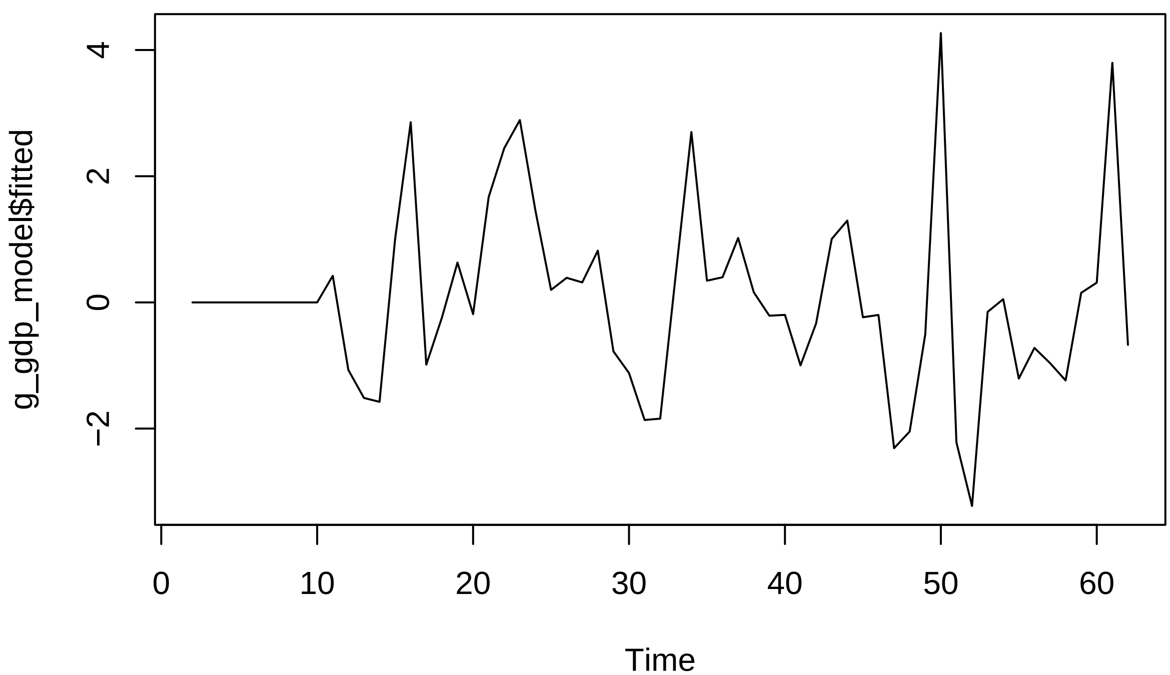

```r
#d=1

#(1,1,3) has been the best so far (AICc = 272.33)
```

```r
g_gdp_model <- Arima(g_gdpDiff, order = c(1, 1, 3), method = "ML")
summary(g_gdp_model)
```

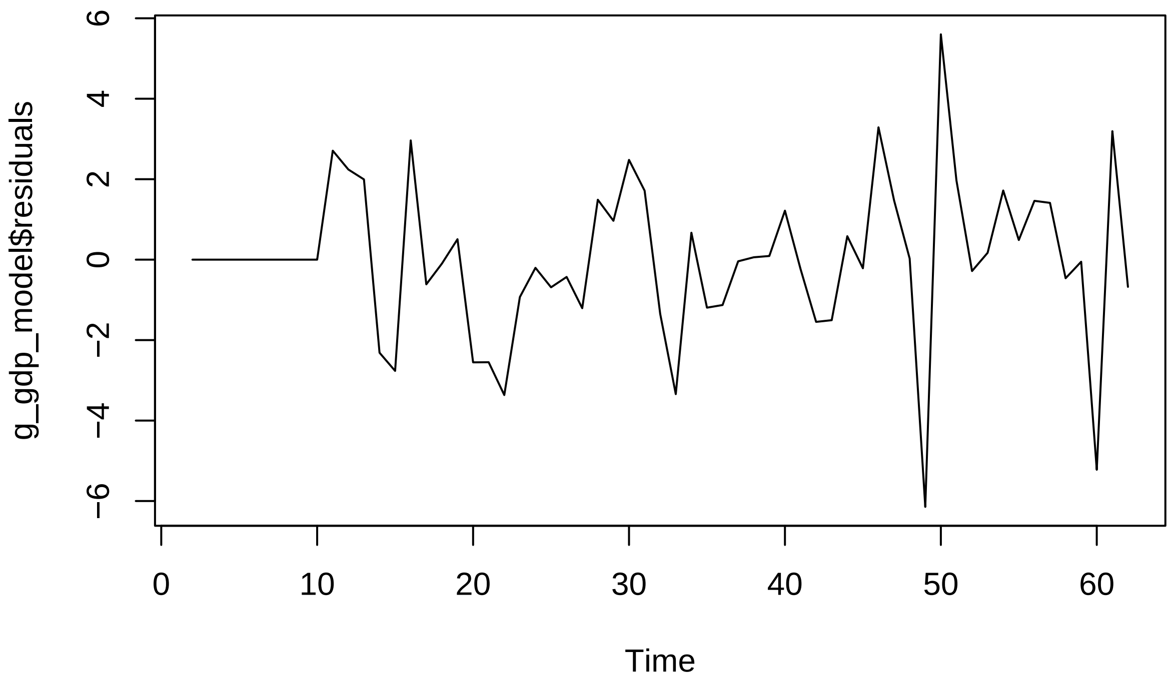
```
## Series: g_gdpDiff
## ARIMA(1,1,3)
##
## Coefficients:
##           ar1      ma1     ma2     ma3
##       -0.1133  -1.5902  0.2507  0.3803
## s.e.   0.3659   0.3447  0.6551  0.3330
##
## sigma^2 = 4.155:  log likelihood = -130.61
## AIC=271.22   AICc=272.33   BIC=281.69
##
## Training set error measures:
##                       ME     RMSE      MAE    MPE     MAPE      MASE
## Training set -0.01027594 1.953087 1.337679 88.652 179.8038 0.5041186
##                     ACF1
## Training set -0.02810031
```

Once we find the p and q values, we try them for the ARIMA models to find the best AICc value Using the ARIMA(1,1,3) model gives the best AICc value.

```r
plot(g_gdp_model$fitted)
```




```r
plot(g_gdp_model$residuals)
```

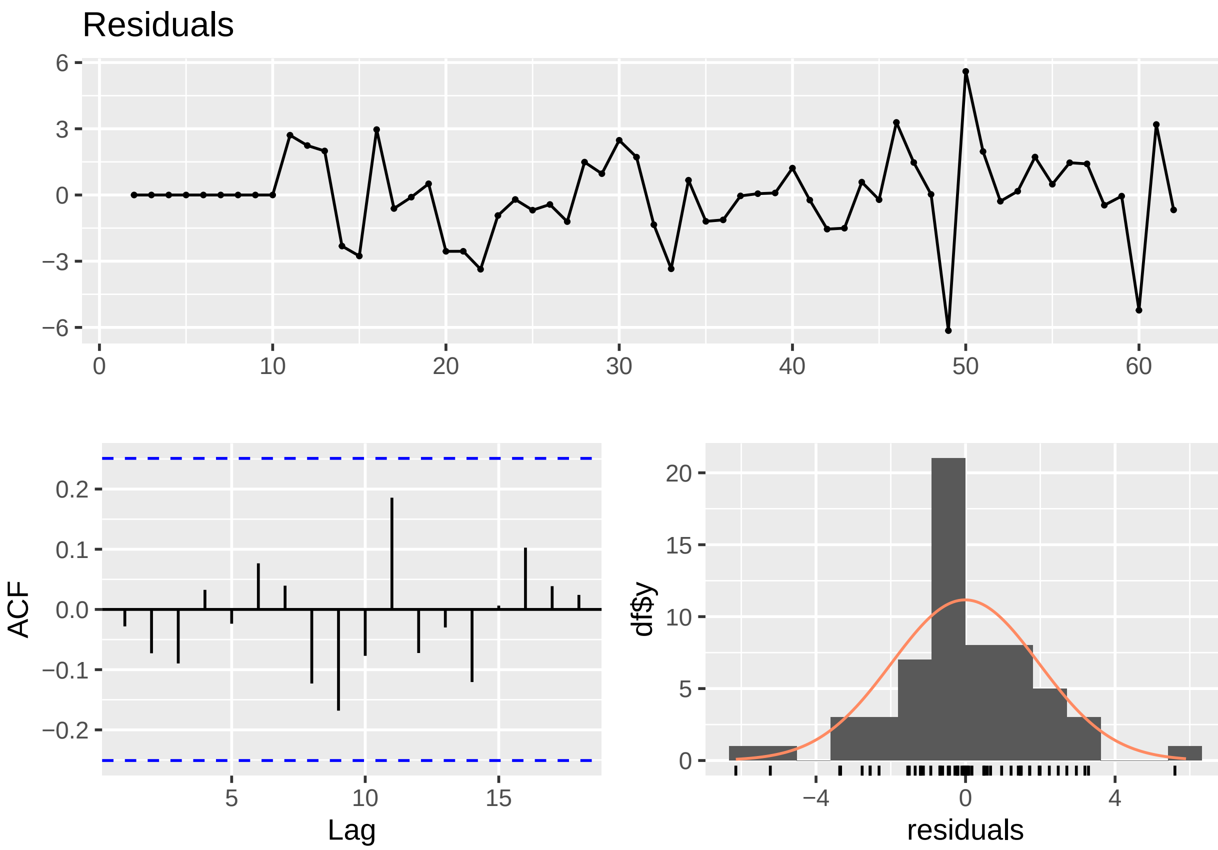



```
#Check stationary of the residuals
ur.kpss(g_gdp_model$residuals) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0992
##
## Critical value for a significance level of:
##                  10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

We then look at the fitted and residual models and then conduct another kpss test for our model with the residuals which looks good

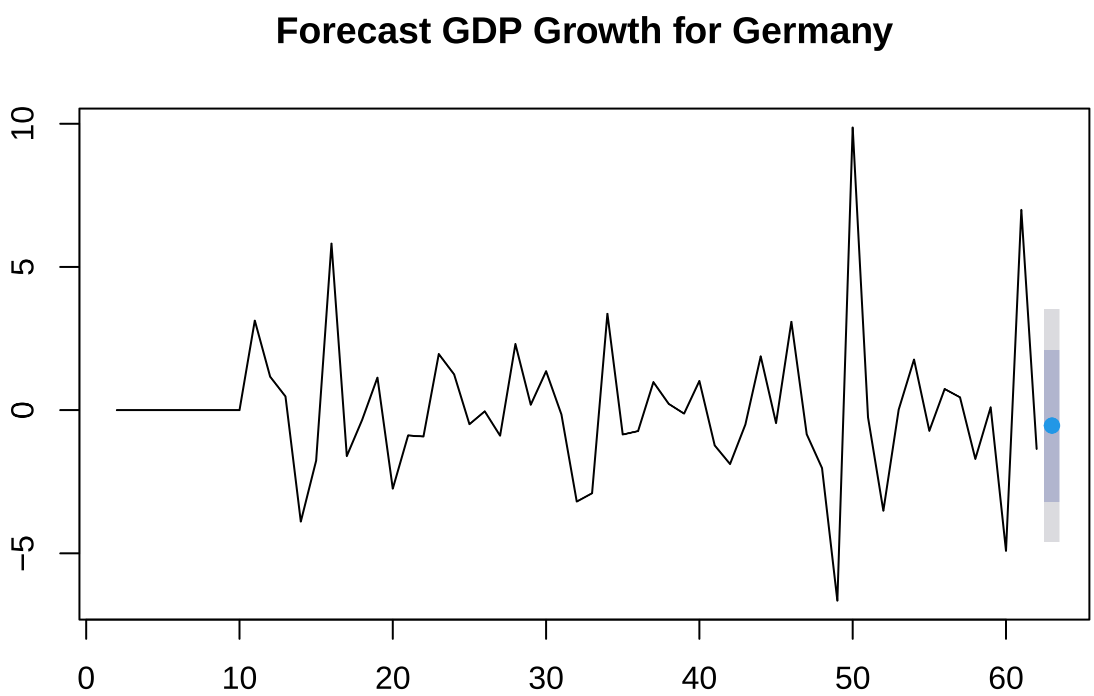
```
checkresiduals(g_gdp_model$residuals)
```




```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 5.1922, df = 10, p-value = 0.878
##
## Model df: 0.   Total lags used: 10
```

Look for white noise

```
g.forecast_values <- forecast(g_gdp_model, h=1)
plot(g.forecast_values, main = "Forecast GDP Growth for Germany")
```

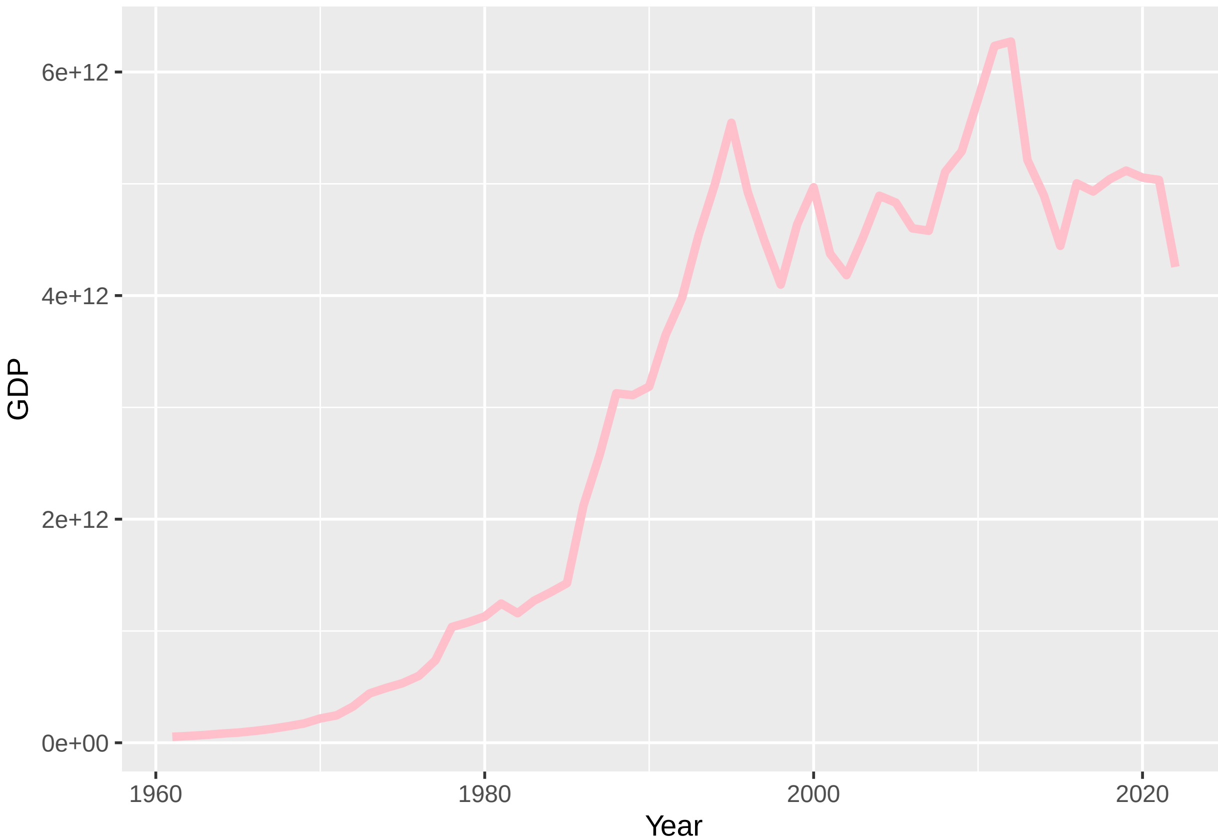


Forecast the data for the future

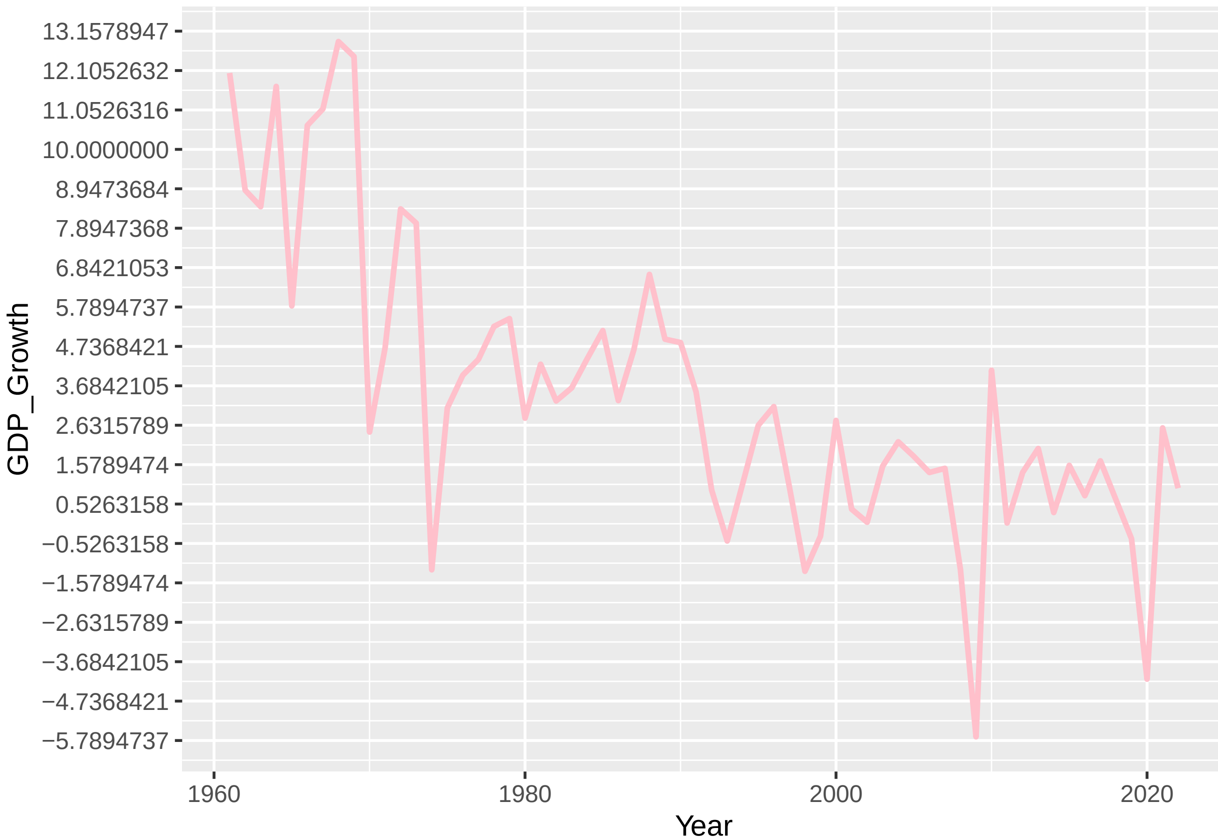
```
japan_gdp <- gdp %>%
  filter(Country == "Japan") %>%
  select(Year, GDP, GDP_Growth)
```

Japan is the country next up

```
ggplot(japan_gdp) +
  geom_line(aes(Year, GDP, group = 1), color = "pink", linewidth = 1.5)
```




```
ggplot(japan_gdp, aes(Year, GDP_Growth, group = 1)) +
  geom_line(color = "pink", linewidth = 1) +
  scale_y_continuous(
    breaks = seq(-10, 15, by = 20/19)
  )
```

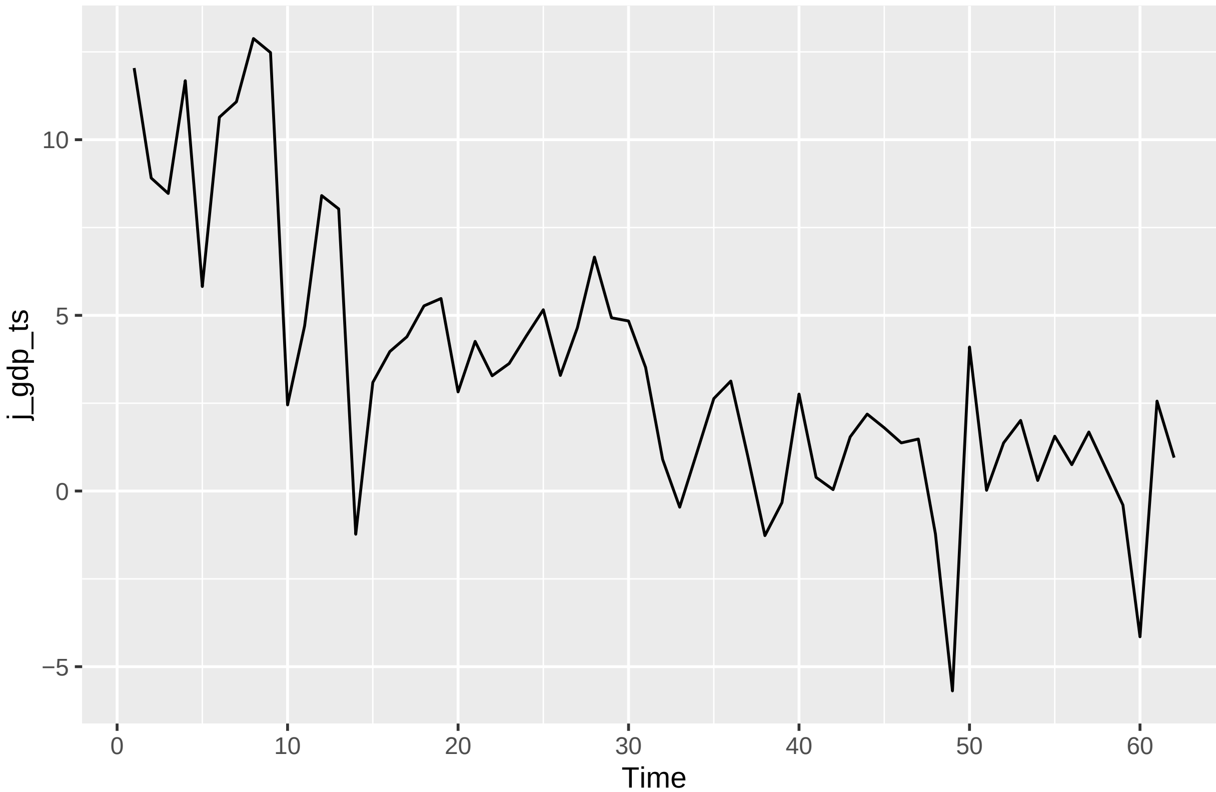


Here is the gdp and gdp growth for Japan over the last 62 years

```
jtrain <- japan_gdp[1:50,]
jtest <- japan_gdp[51:62,]
jntest <- nrow(jtest)
```

Then we make a train and testing set

```
j_gdp_ts <- ts(japan_gdp$GDP_Growth)
autoplot(j_gdp_ts)
```

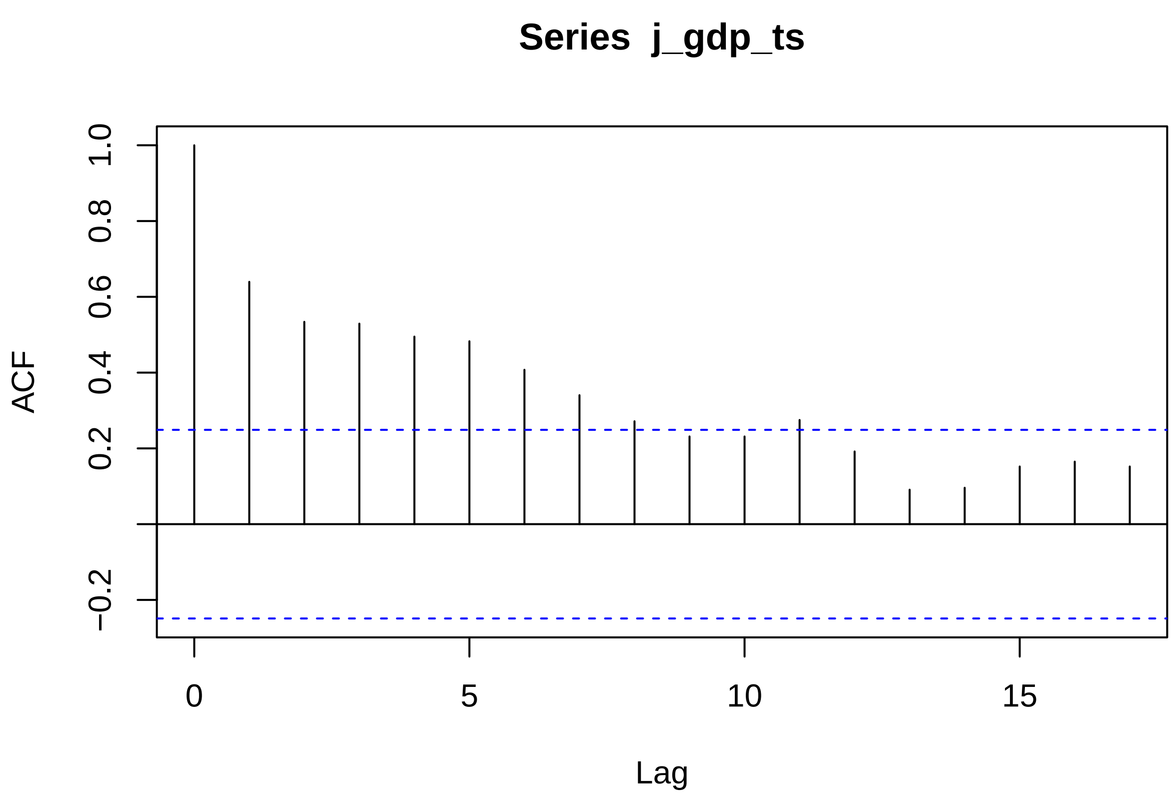


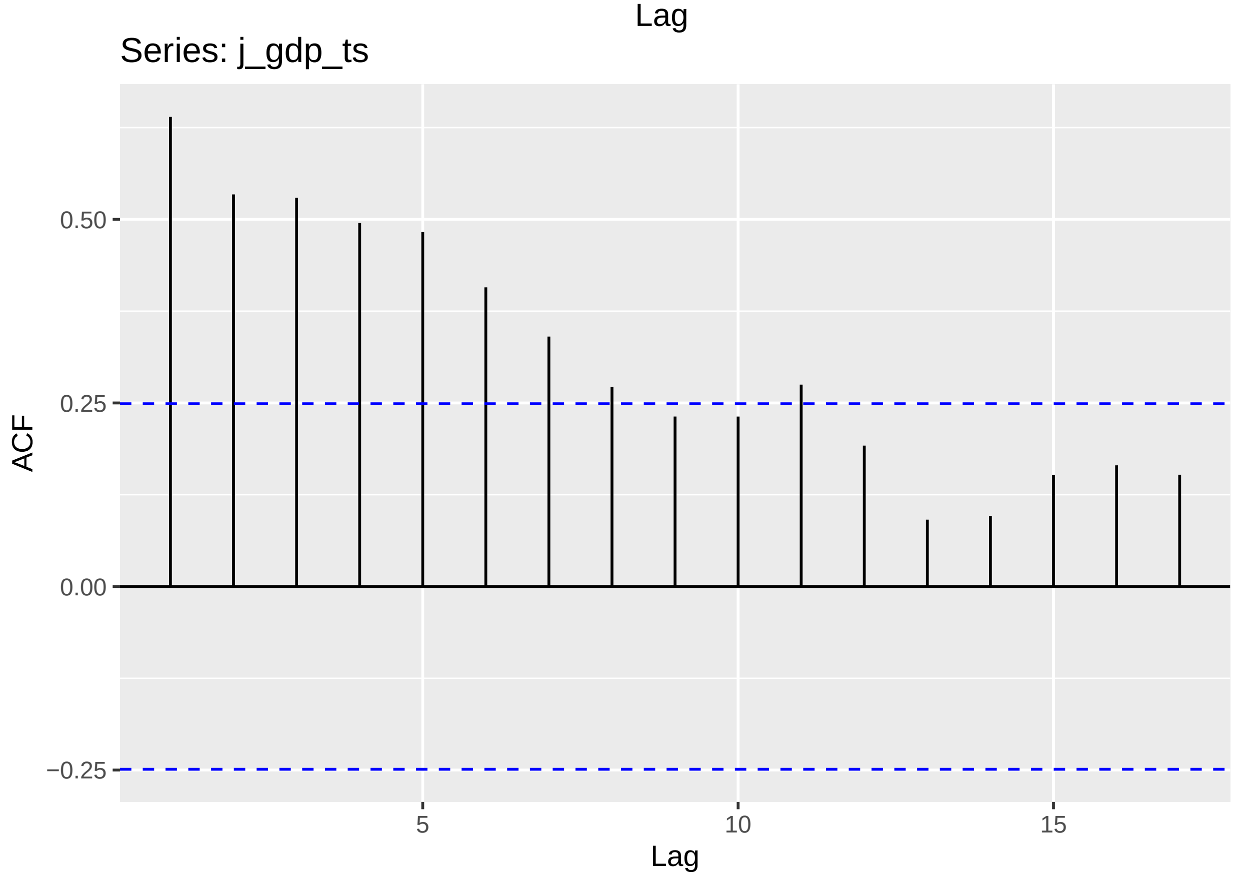

```
ur.kpss(j_gdp_ts) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 1.3021
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Then we create a time series model and then check for stationary. We use the kpss test and see if it needs to be diffrenced. Seeing that the test-statisic is near the test data, we can diffrence the data once.

```
autoplot(acf(j_gdp_ts))
```

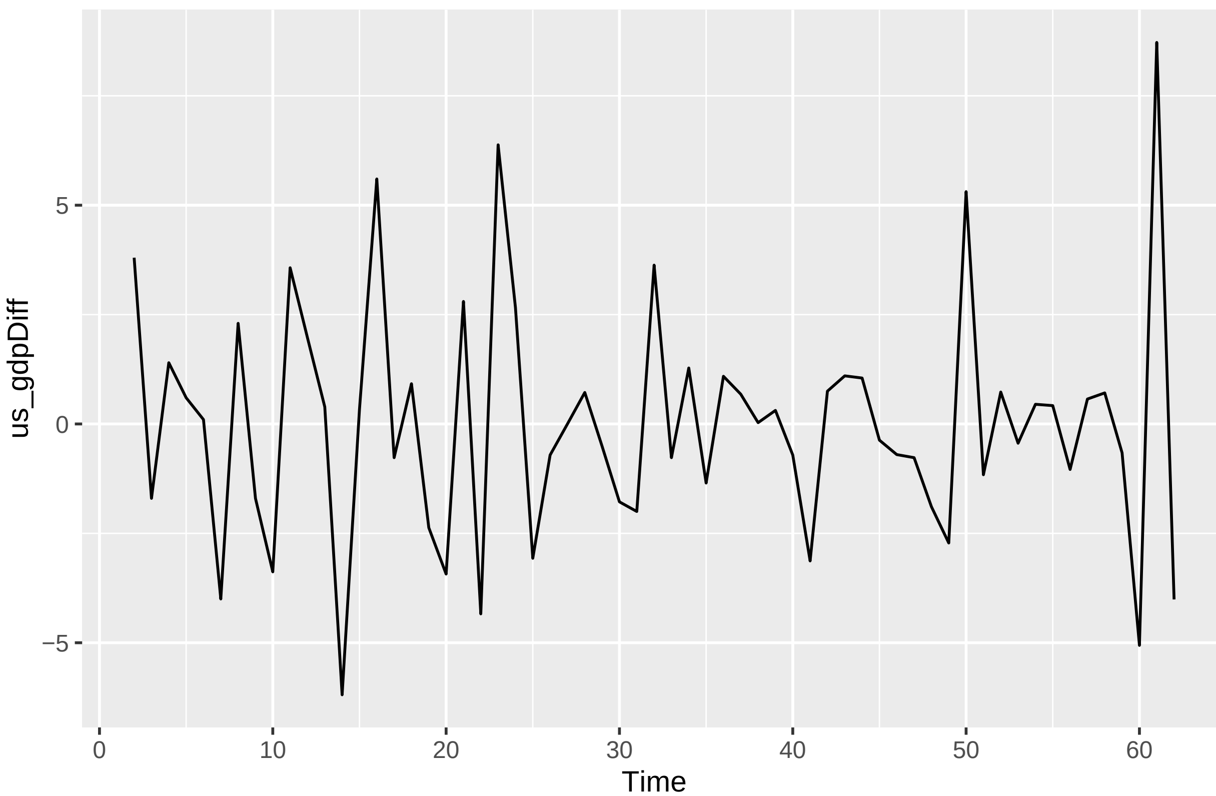





```r
j_gdpDiff = diff(j_gdp_ts, lag = 1)

ur.kpss(j_gdpDiff) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.095
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

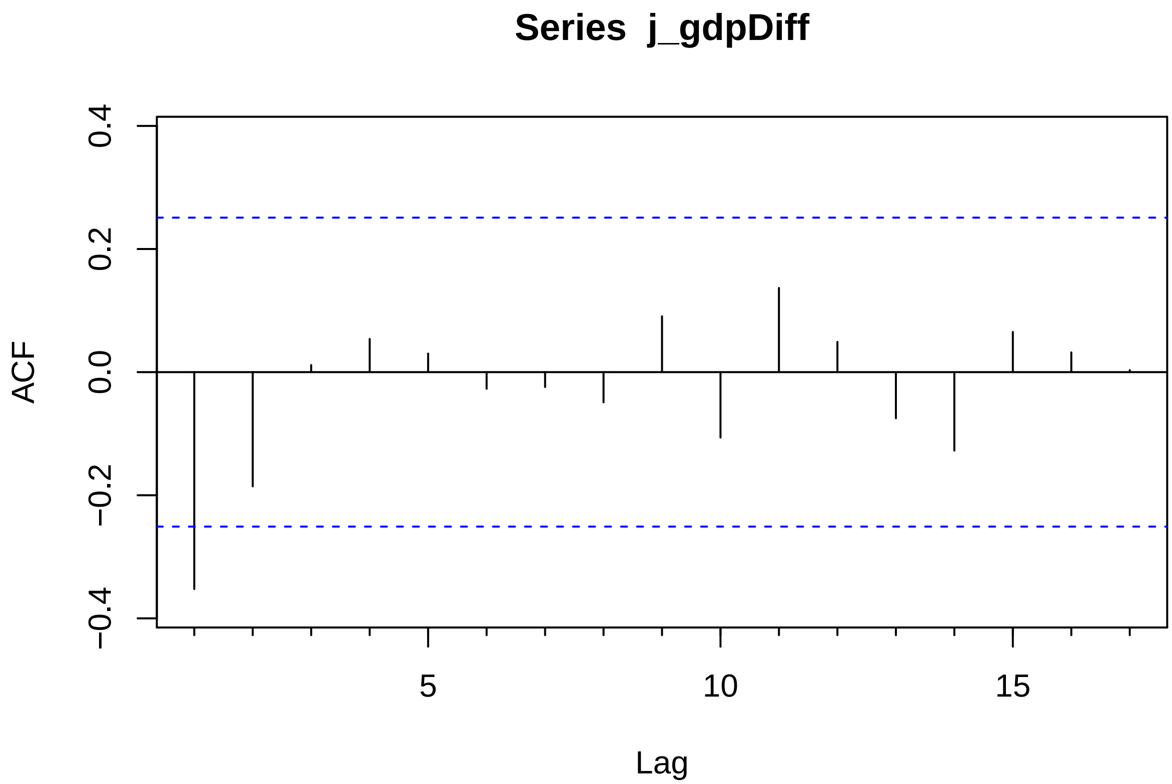
```r
autoplot(us_gdpDiff)
```



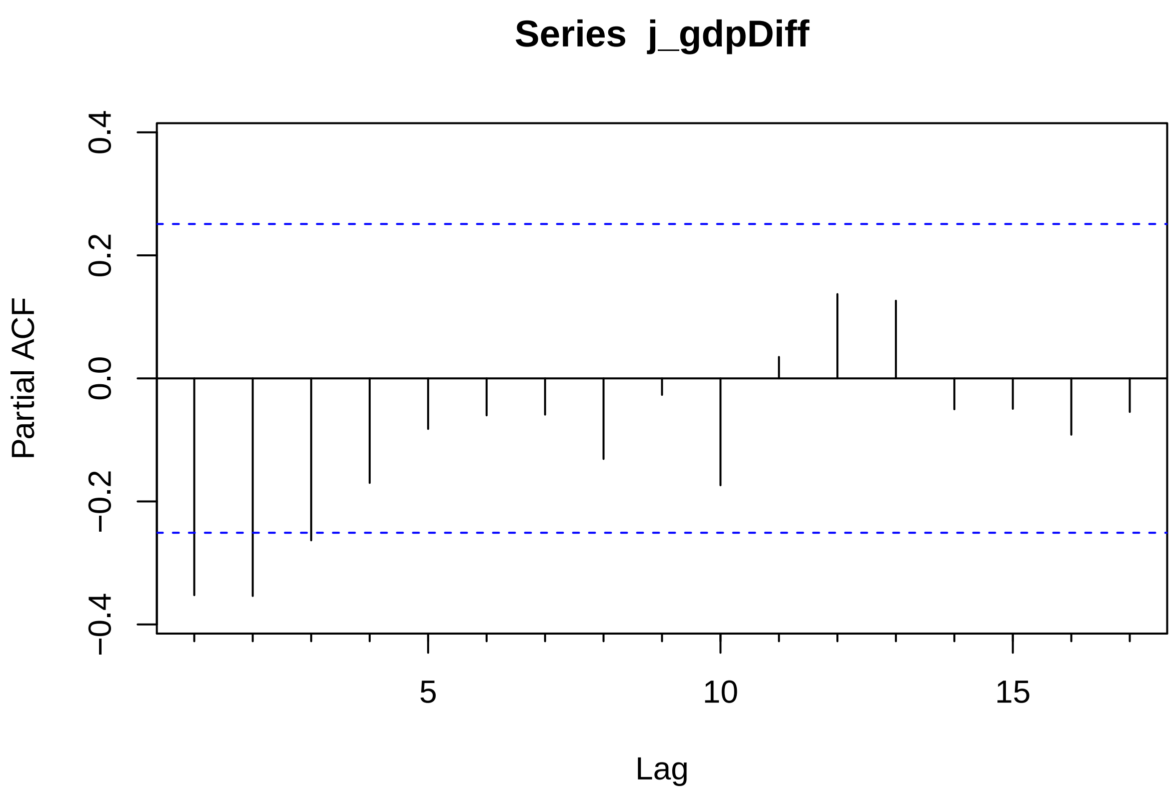
We go ahead and difference the data once and then check the test statistic again Seeing that the test statistic is way lower, this data is stationary.

We then look at the Acf and pacf graphs to see our q and p value for the ARIMA model

```r
Acf(j_gdpDiff) #1
```




```
Pacf(j_gdpDiff) #1,2,3
```



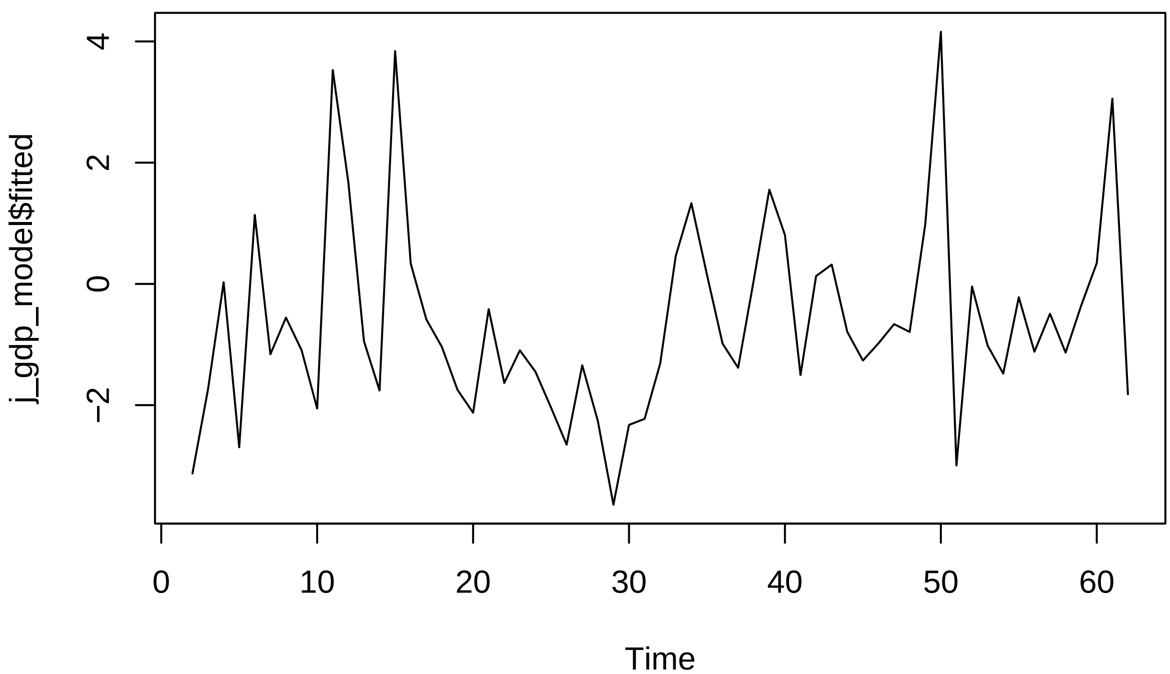

```
#d=1

#(11,1,6) has been the best so far (AICc = 297.67)
```

```r
j_gdp_model <- Arima(j_gdpDiff, order = c(1, 1, 2), method = "ML")
summary(j_gdp_model)
```

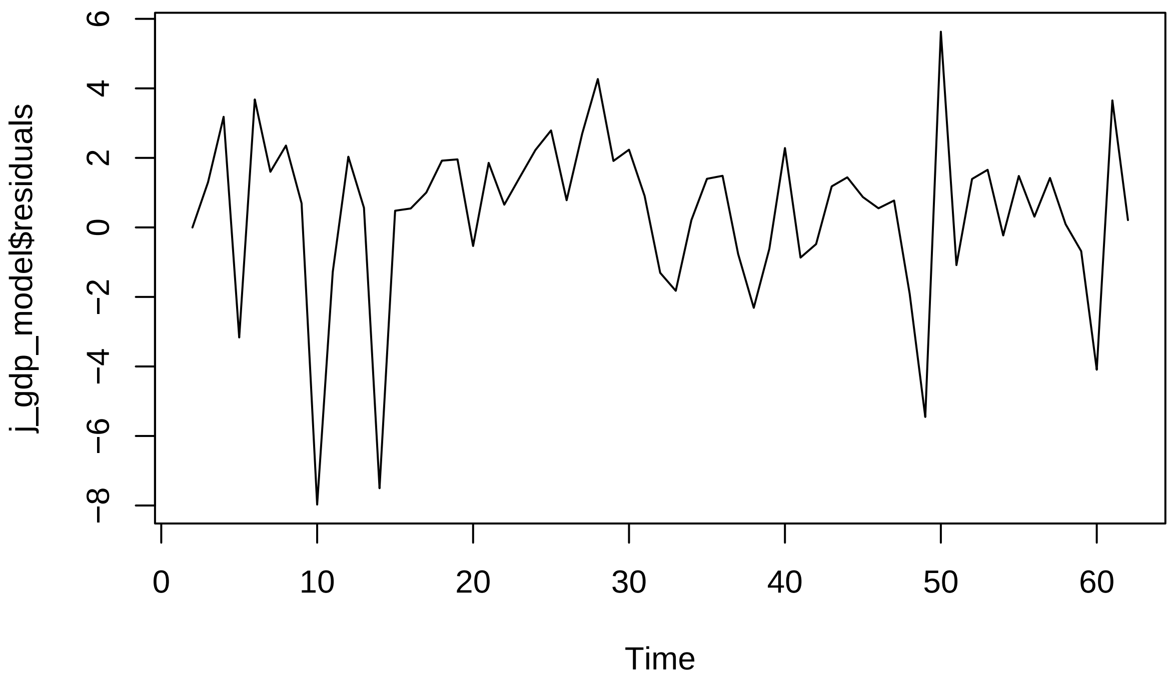
```
## Series: j_gdpDiff
## ARIMA(1,1,2)
##
## Coefficients:
##          ar1      ma1     ma2
##       0.2637  -1.9872  0.9999
## s.e.  0.1295   0.1217  0.1224
##
## sigma^2 = 6.428:  log likelihood = -144.47
## AIC=296.95   AICc=297.67   BIC=305.32
##
## Training set error measures:
##                     ME     RMSE     MAE      MPE    MAPE      MASE         ACF1
## Training set 0.4436566 2.450901 1.82339 53.04981 181.679 0.5182049 -0.02642518
```

Once we find the p and q values, we try them for the ARIMA models to find the best AICc value Using the ARIMA(11,1,6) model gives the best AICc value.

```r
plot(j_gdp_model$fitted)
```




```r
plot(j_gdp_model$residuals)
```

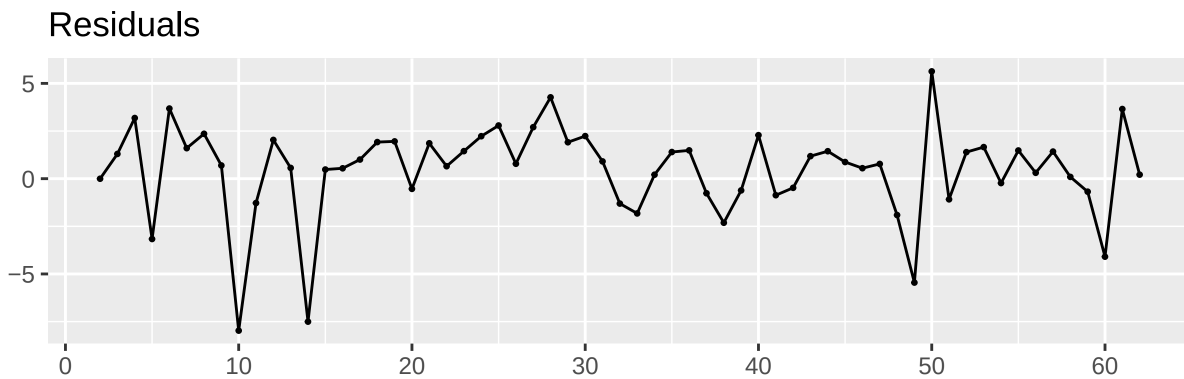


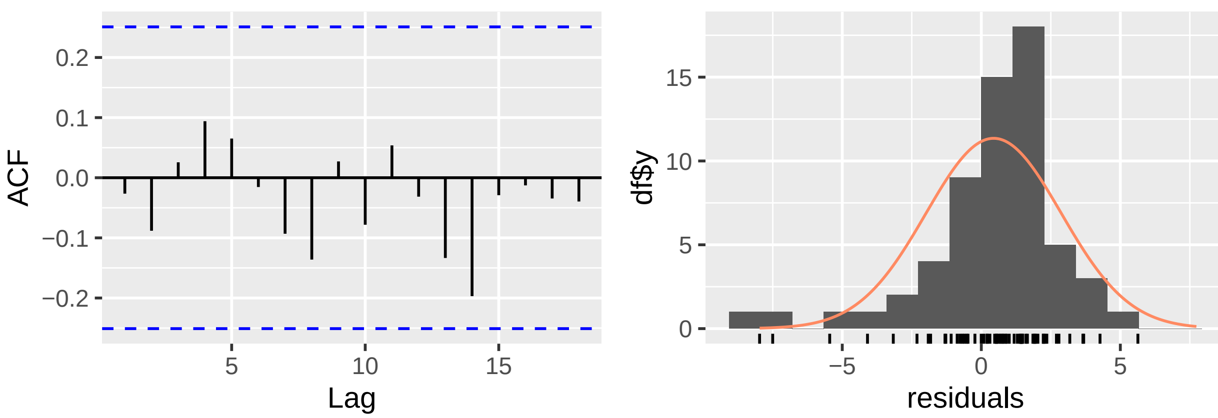

```r
#Check stationary of the residuals
ur.kpss(j_gdp_model$residuals) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0767
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

We then look at the fitted and residual models and then conduct another kpss test for our model with the residuals which looks good

```r
checkresiduals(j_gdp_model$residuals)
```

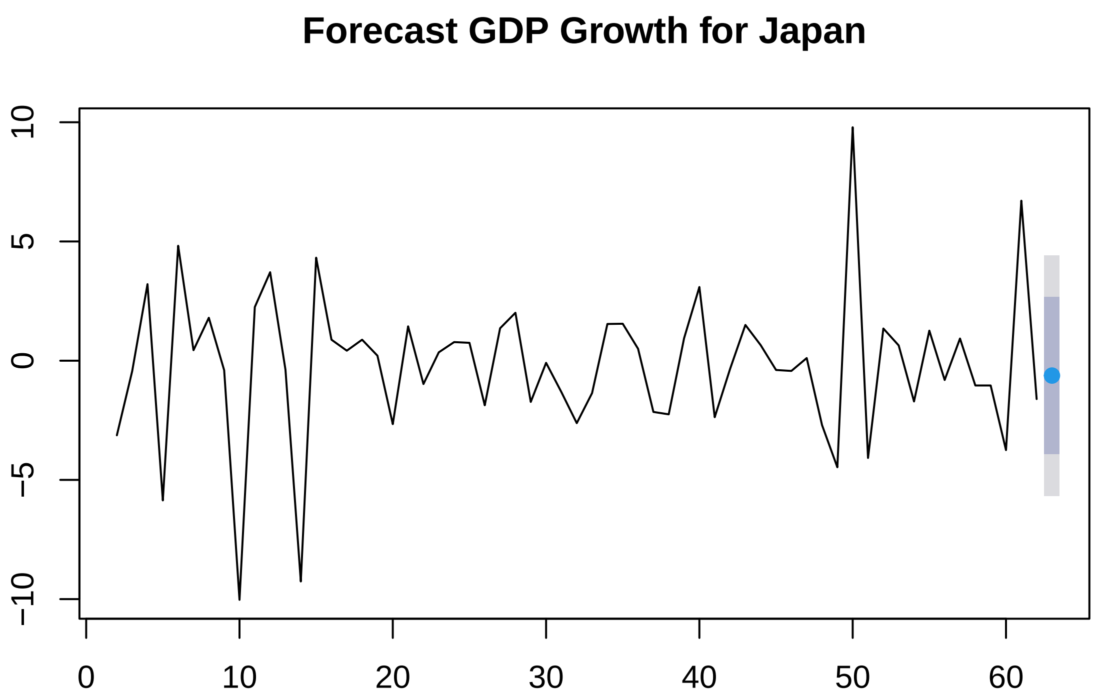





```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 3.9736, df = 10, p-value = 0.9485
##
## Model df: 0.   Total lags used: 10
```

Look for white noise

```
j.forecast_values <- forecast(j_gdp_model, h=1)
plot(j.forecast_values, main = "Forecast GDP Growth for Japan")
```

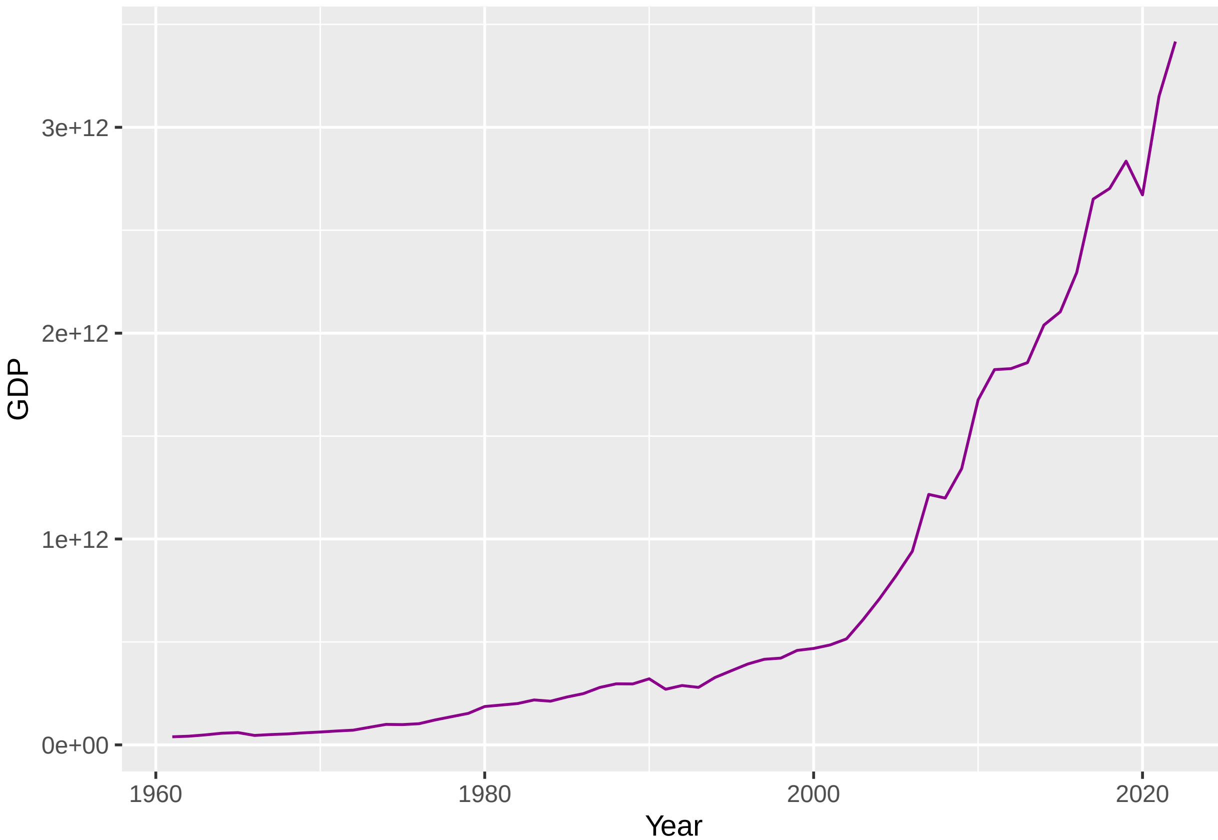


Plot for the future

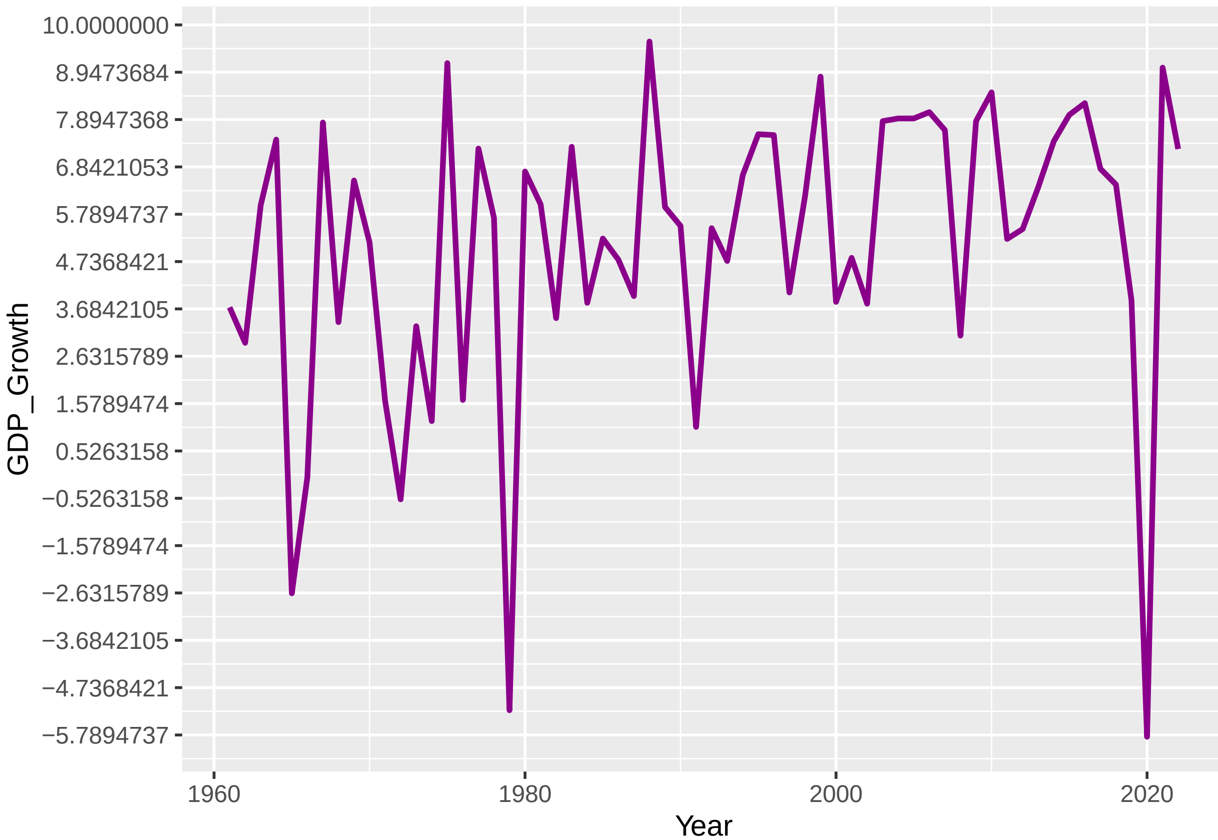
```r
india_gdp <- gdp %>%
  filter(Country == "India") %>%
  select(Year, GDP, GDP_Growth)
```

The next country is India

```r
ggplot(india_gdp) +
  geom_line(aes(Year, GDP, group = 1), color = "darkmagenta", linewidth = .5)
```




```r
ggplot(india_gdp, aes(Year, GDP_Growth, group = 1)) +
  geom_line(color = "darkmagenta", linewidth = 1) +
  scale_y_continuous(
    breaks = seq(-10, 10, by = 20/19) )
```

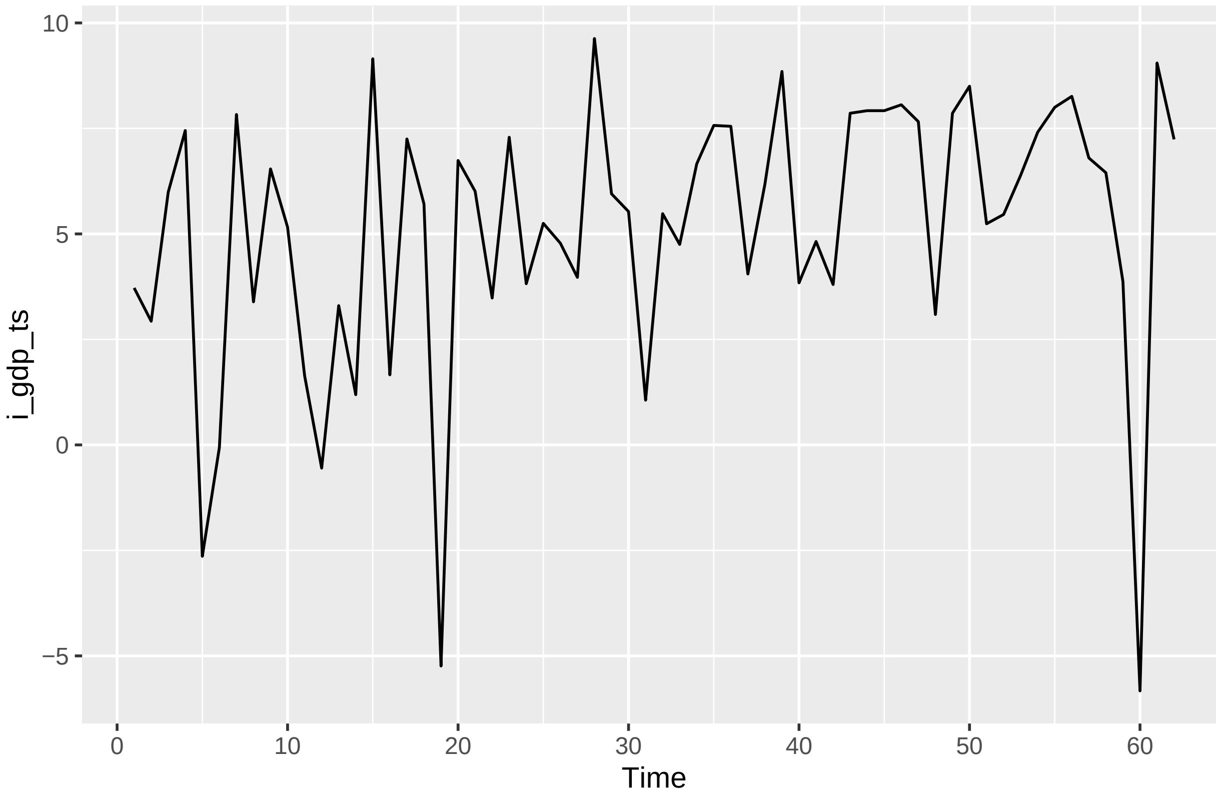


Here are the grpahs for gdp and gdp growth

```
itrain <- india_gdp[1:50,]
itest <- india_gdp[51:62,]
intest <- nrow(itest)
```

Train and testing sets

```
i_gdp_ts <- ts(india_gdp$GDP_Growth)
autoplot(i_gdp_ts)
```

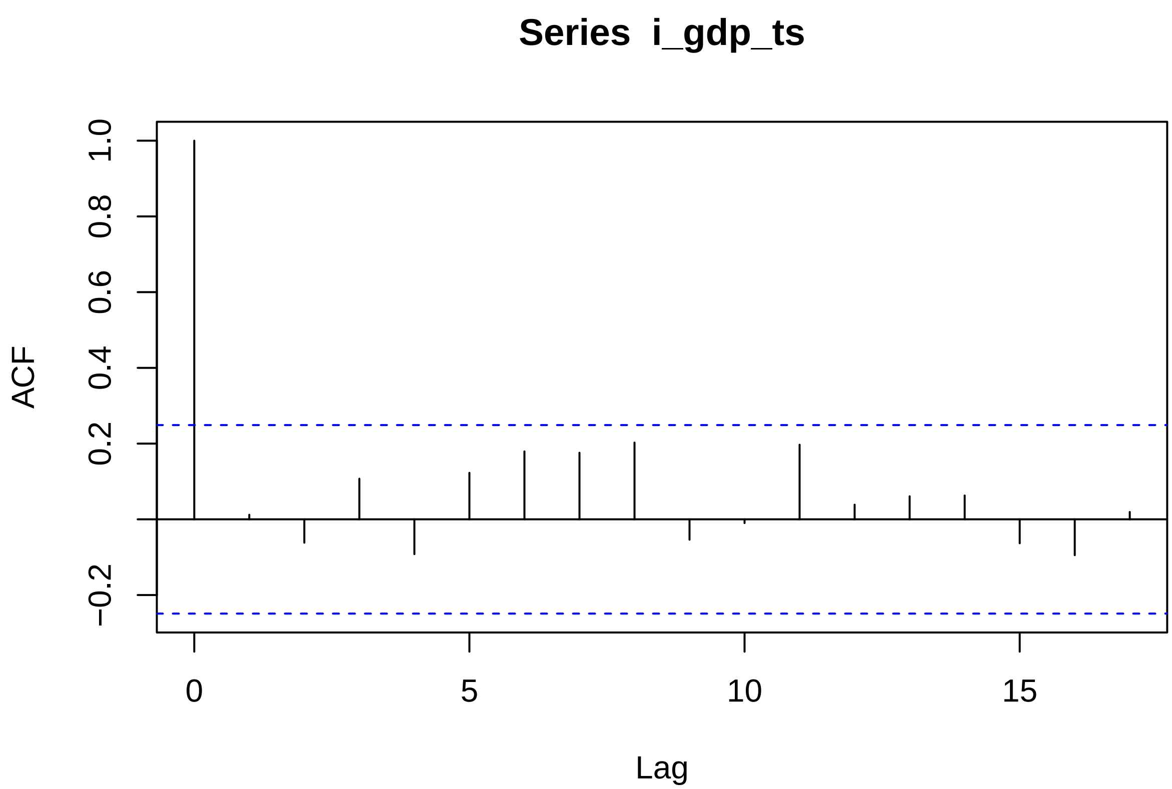


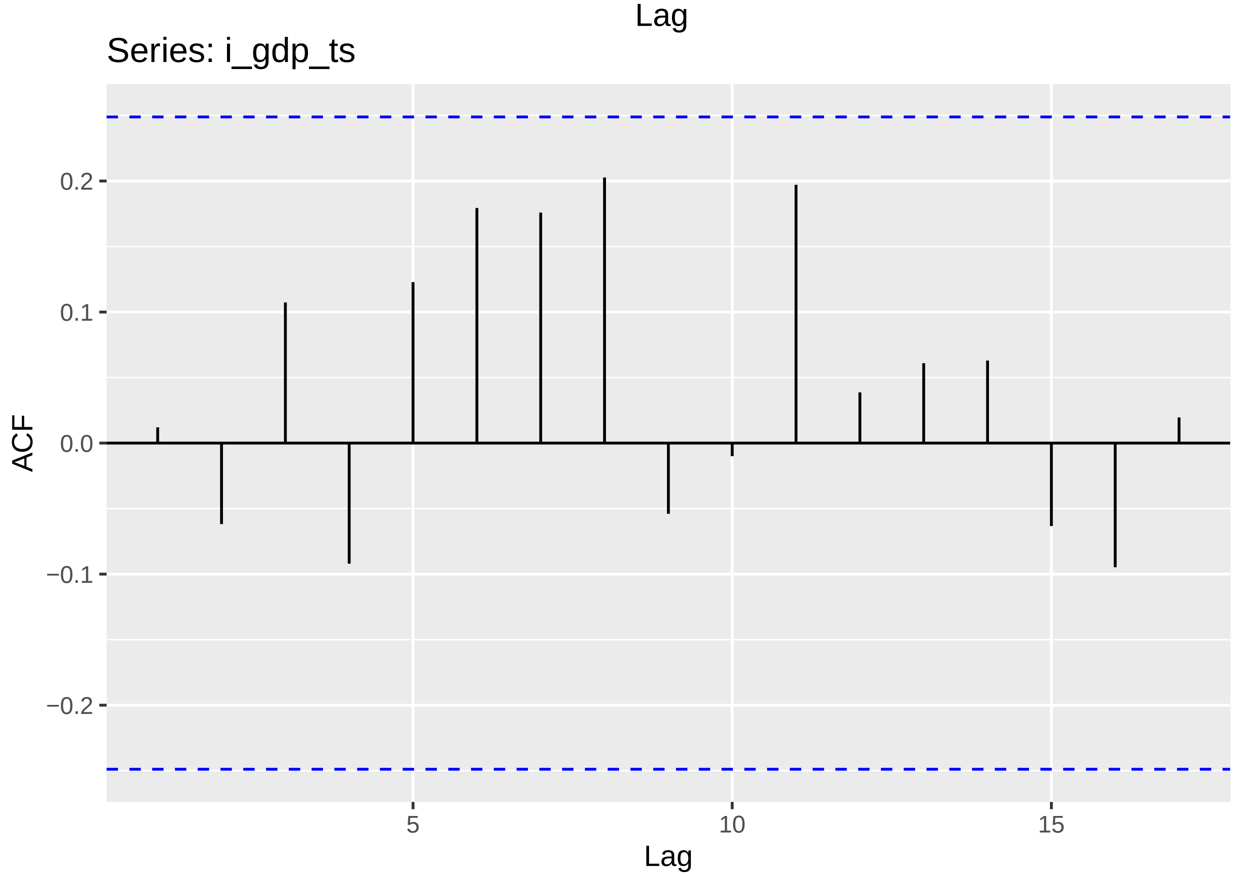

```
ur.kpss(i_gdp_ts) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.6758
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Then we create a time series model and then check for stationary. We use the kpss test and see if it needs to be diffrenced. Seeing that the test-statisic is near the test data, we can diffrence the data once.

```
autoplot(acf(i_gdp_ts))
```

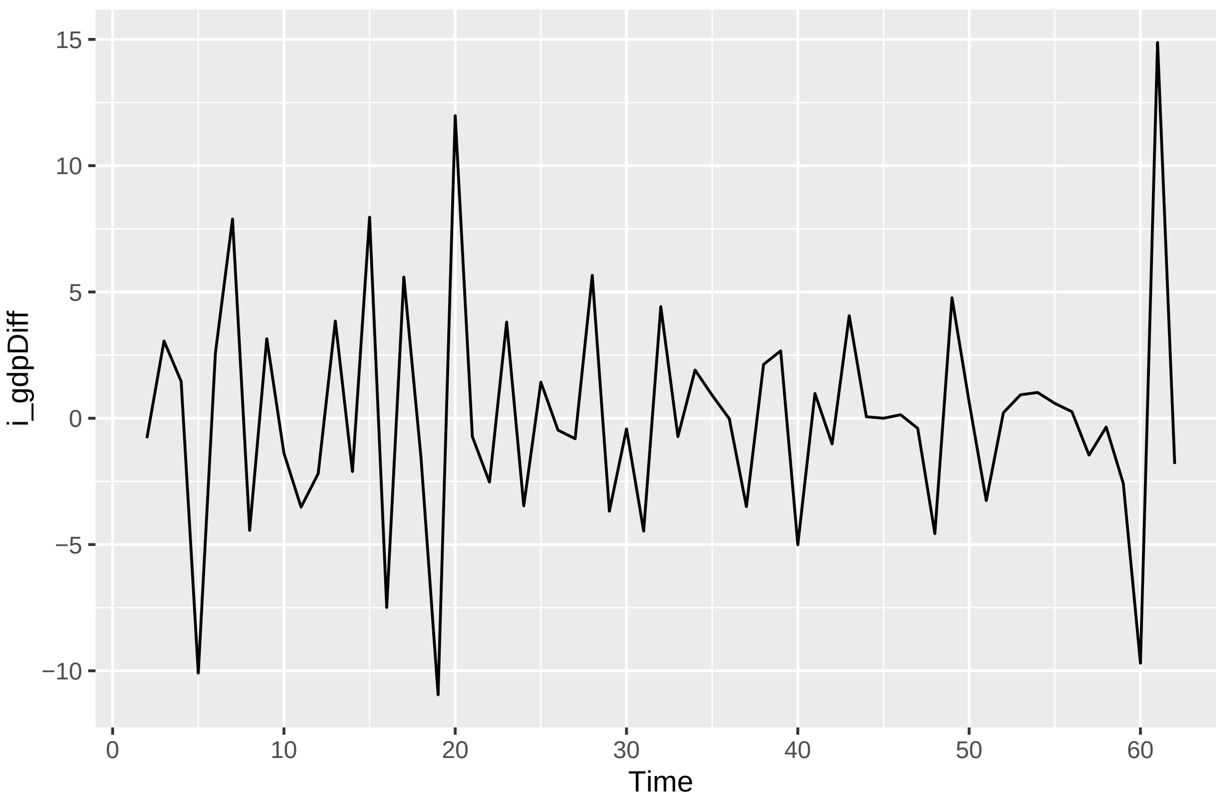





```r
i_gdpDiff = diff(i_gdp_ts, lag = 1)

ur.kpss(i_gdpDiff) %>% summary()
```

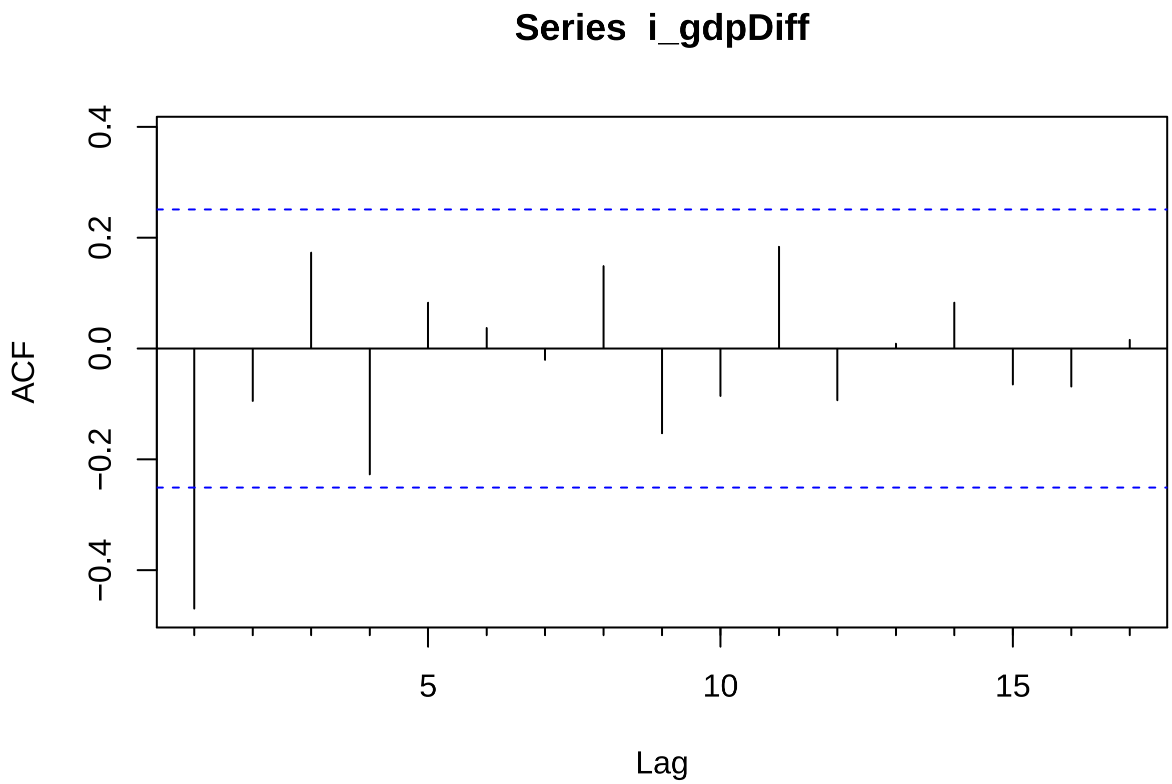
```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0267
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

```
autoplot(i_gdpDiff)
```

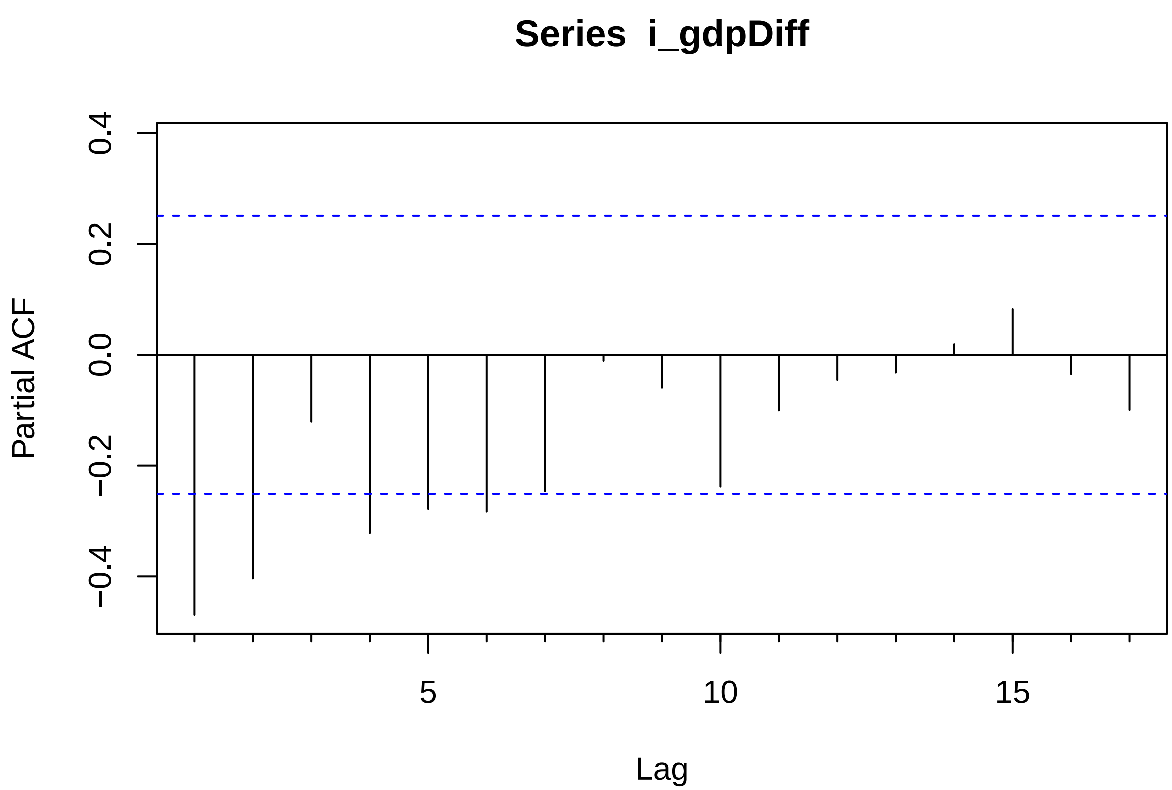


We go ahead and difference the data once and then check the test statistic again Seeing that the test statistic is way lower, this data is stationary. We then look at the Acf and pacf graphs to see our q and p value for the ARIMA model

```
Acf(i_gdpDiff) #1
```




```r
Pacf(i_gdpDiff) #1,2,4,5,6
```



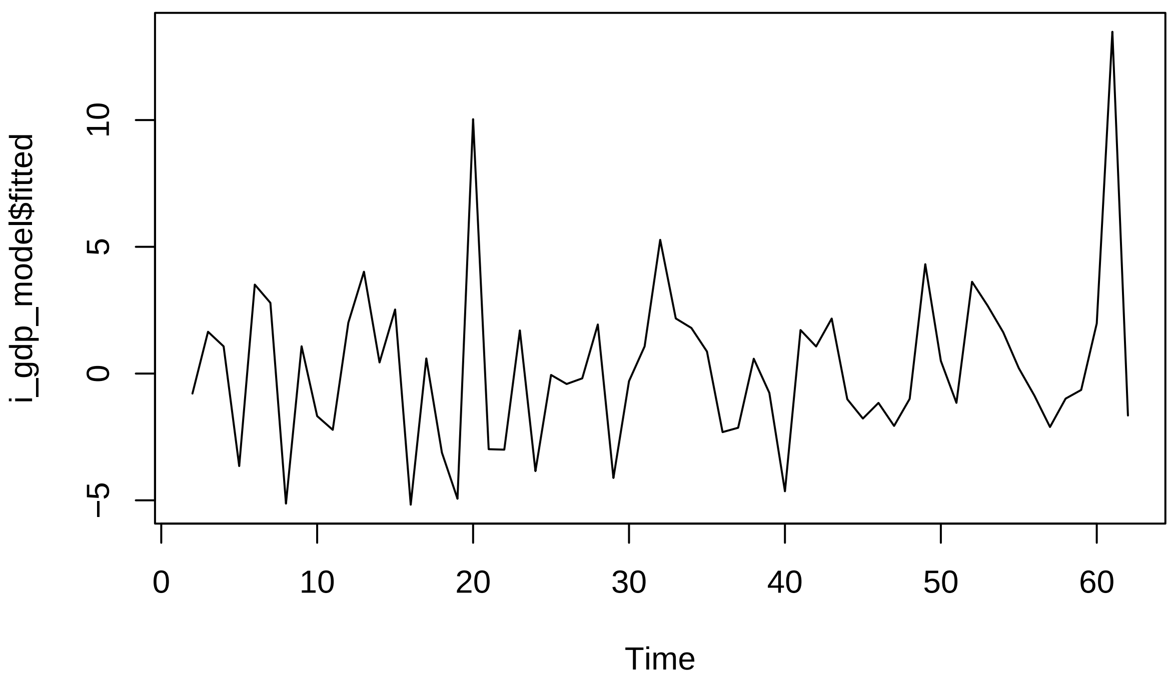

```r
#d=1

#(1,1,6) has been the best so far (AICc = 327.73)
```

```r
i_gdp_model <- Arima(i_gdpDiff, order = c(1, 1, 6), method = "ML")
summary(i_gdp_model)
```

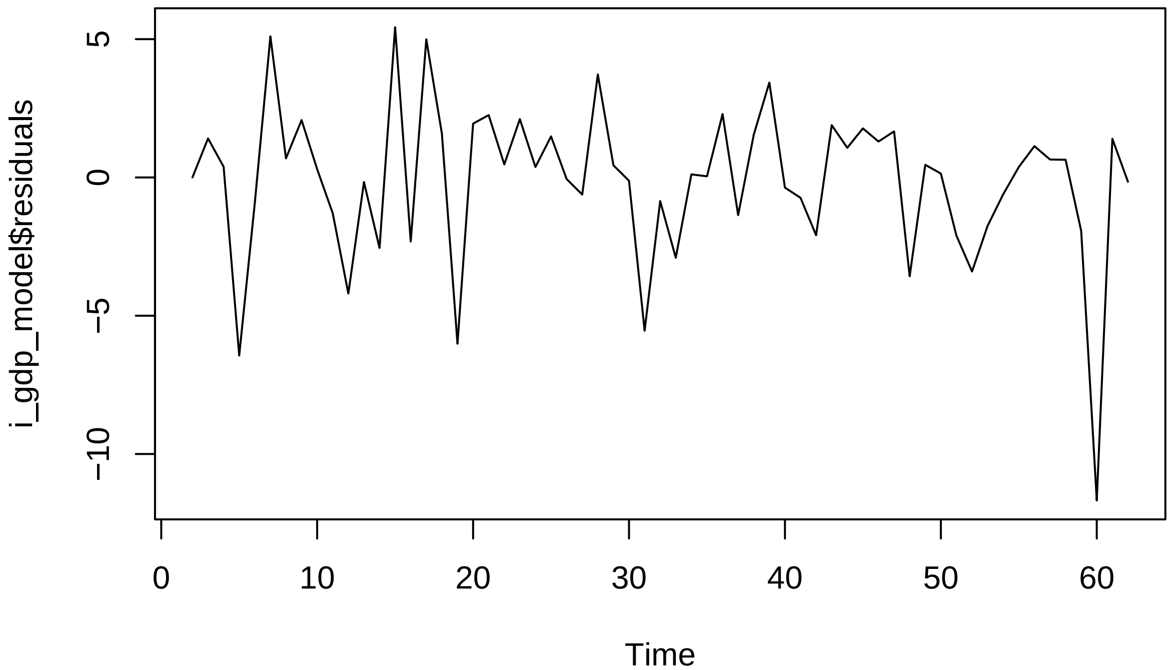
```
## Series: i_gdpDiff
## ARIMA(1,1,6)
##
## Coefficients:
##           ar1      ma1     ma2     ma3      ma4     ma5      ma6
##       -0.0376  -2.0889  1.0259  0.3719  -0.4762  0.4818  -0.3144
## s.e.   0.4141   0.3993  0.8966  0.5577   0.4266  0.4442   0.1667
##
## sigma^2 = 9.157:  log likelihood = -154.45
## AIC=324.9   AICc=327.73   BIC=341.66
##
## Training set error measures:
##                      ME     RMSE     MAE MPE MAPE      MASE        ACF1
## Training set -0.1494359 2.820582 1.94168 Inf  Inf 0.3375466 -0.01933381
```

Once we find the p and q values, we try them for the ARIMA models to find the best AICc value Using the ARIMA(1,1,6) model gives the best AICc value.

```r
plot(i_gdp_model$fitted)
```




```r
plot(i_gdp_model$residuals)
```

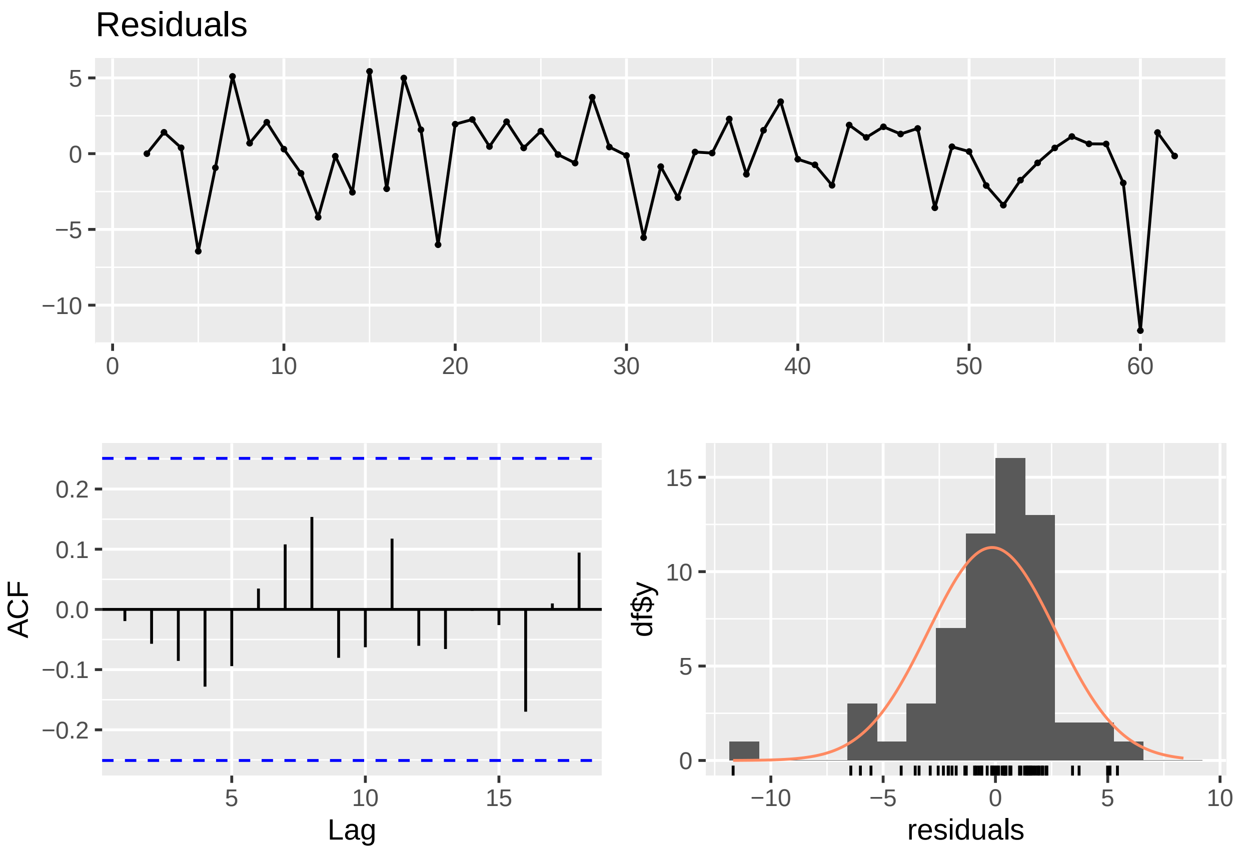



```r
#Check stationary of the residuals
ur.kpss(i_gdp_model$residuals) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.1983
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

We then look at the fitted and residual models and then conduct another kpss test for our model with the residuals which looks good

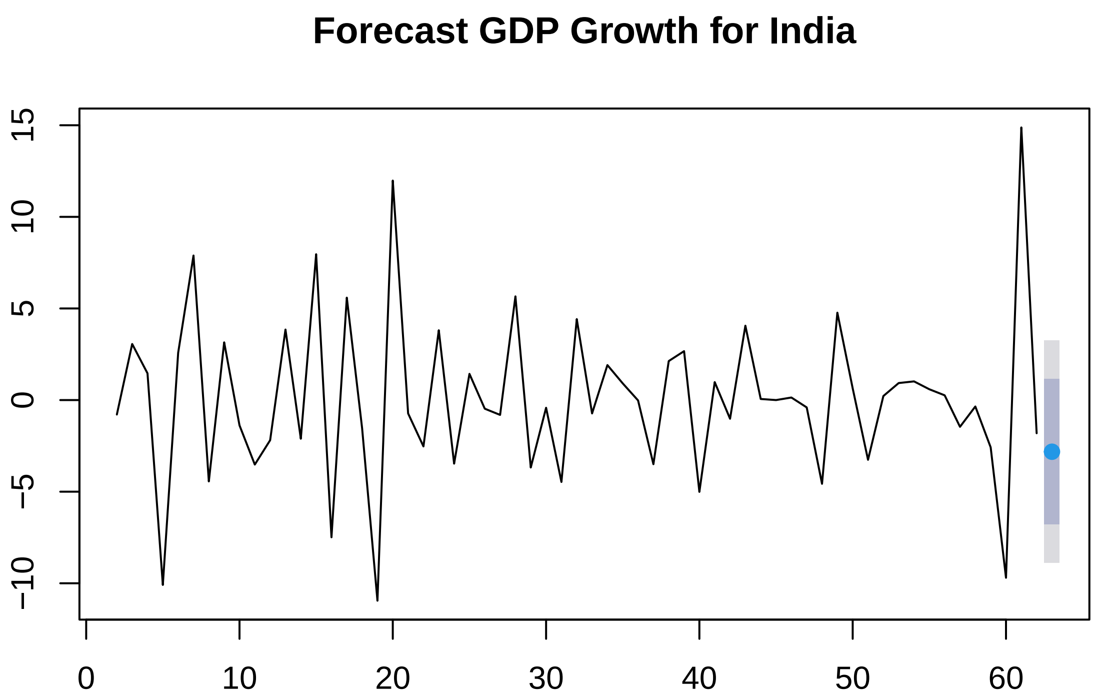
```r
checkresiduals(i_gdp_model$residuals)
```




```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 5.8415, df = 10, p-value = 0.8284
##
## Model df: 0.   Total lags used: 10
```

Check for white noise

```
i.forecast_values <- forecast(i_gdp_model, h=1)
plot(i.forecast_values, main = "Forecast GDP Growth for India")
```

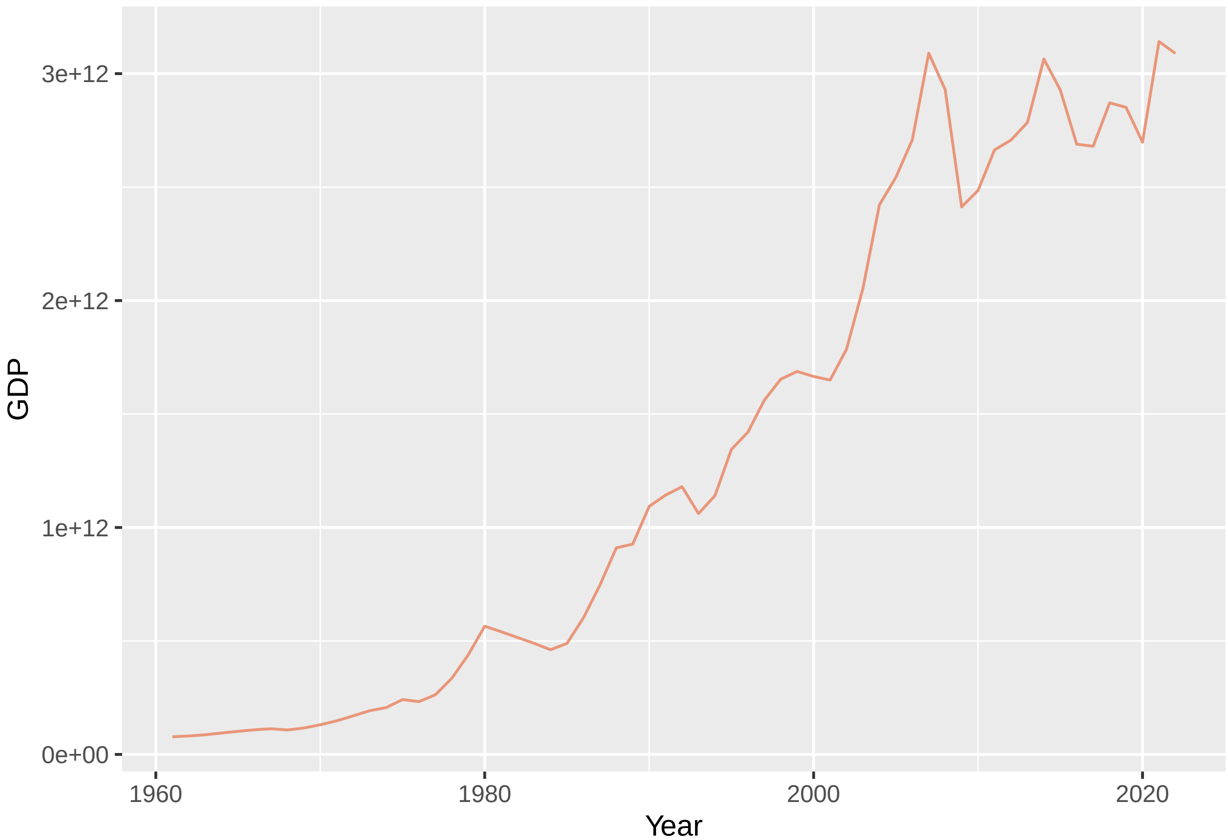


Then forecast for the future

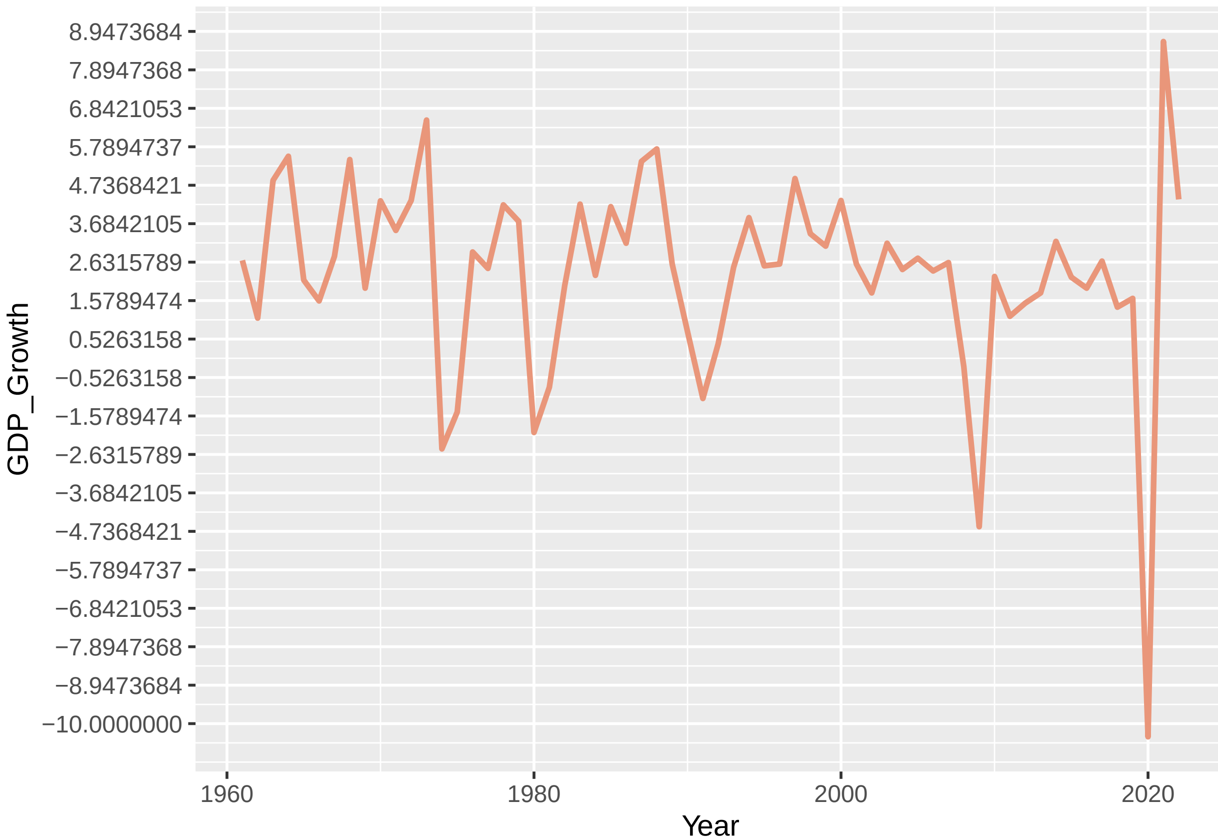
```
uk_gdp <- gdp %>%
  filter(Country == "United Kingdom") %>%
  select(Year, GDP, GDP_Growth)
```

The United Kingdom is up next

```
ggplot(uk_gdp) +
  geom_line(aes(Year, GDP, group = 1), color = "darksalmon", linewidth = .5)
```




```
ggplot(uk_gdp, aes(Year, GDP_Growth, group = 1)) +
  geom_line(color = "darksalmon", linewidth = 1) +
  scale_y_continuous(
    breaks = seq(-10, 10, by = 20/19)   )
```

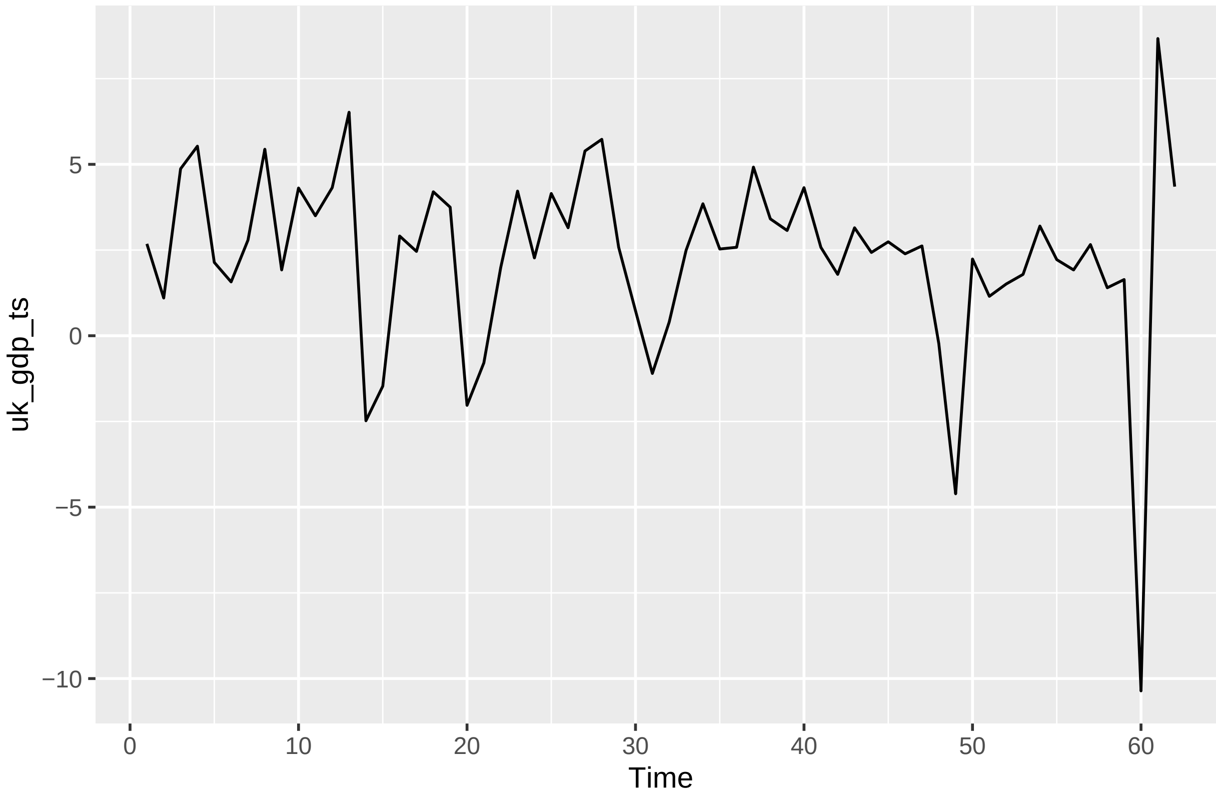


Here are the graphs which shows a very constant graowth until 2008 and 2020

```
uktrain <- uk_gdp[1:50,]
uktest <- uk_gdp[51:62,]
ukntest <- nrow(uktest)
```

Training and testing sets

```
uk_gdp_ts <- ts(uk_gdp$GDP_Growth)
autoplot(uk_gdp_ts)
```

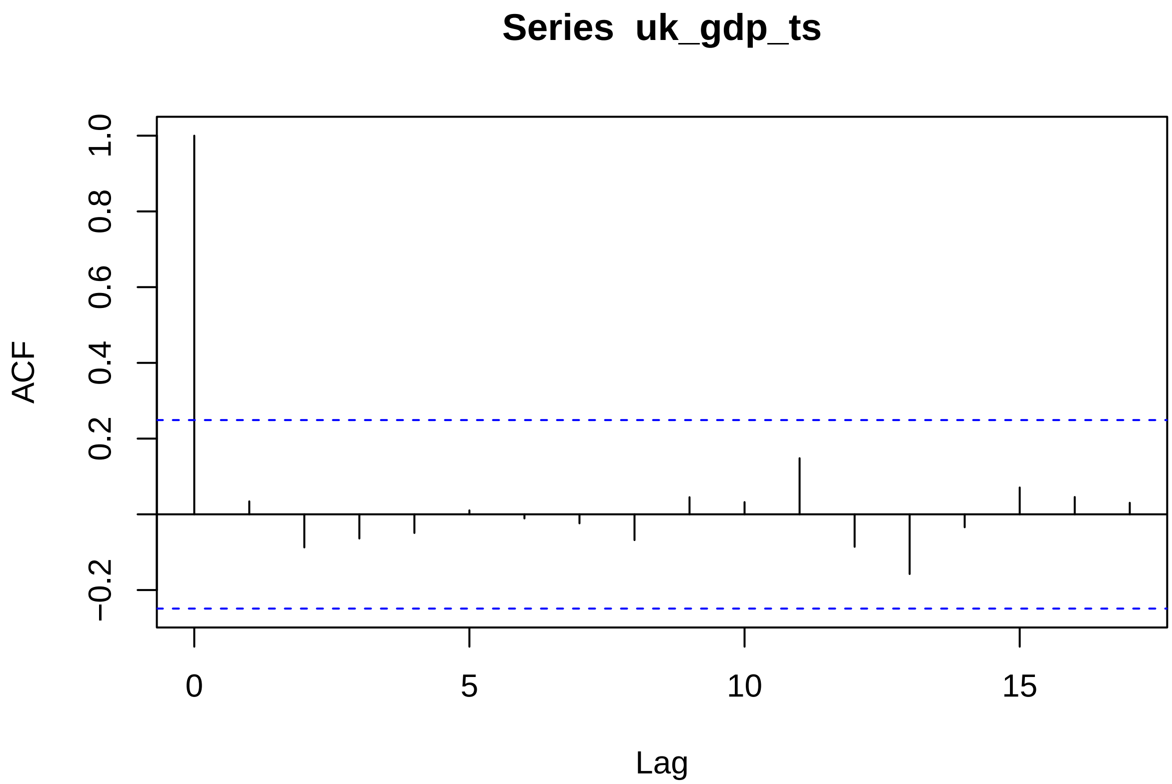


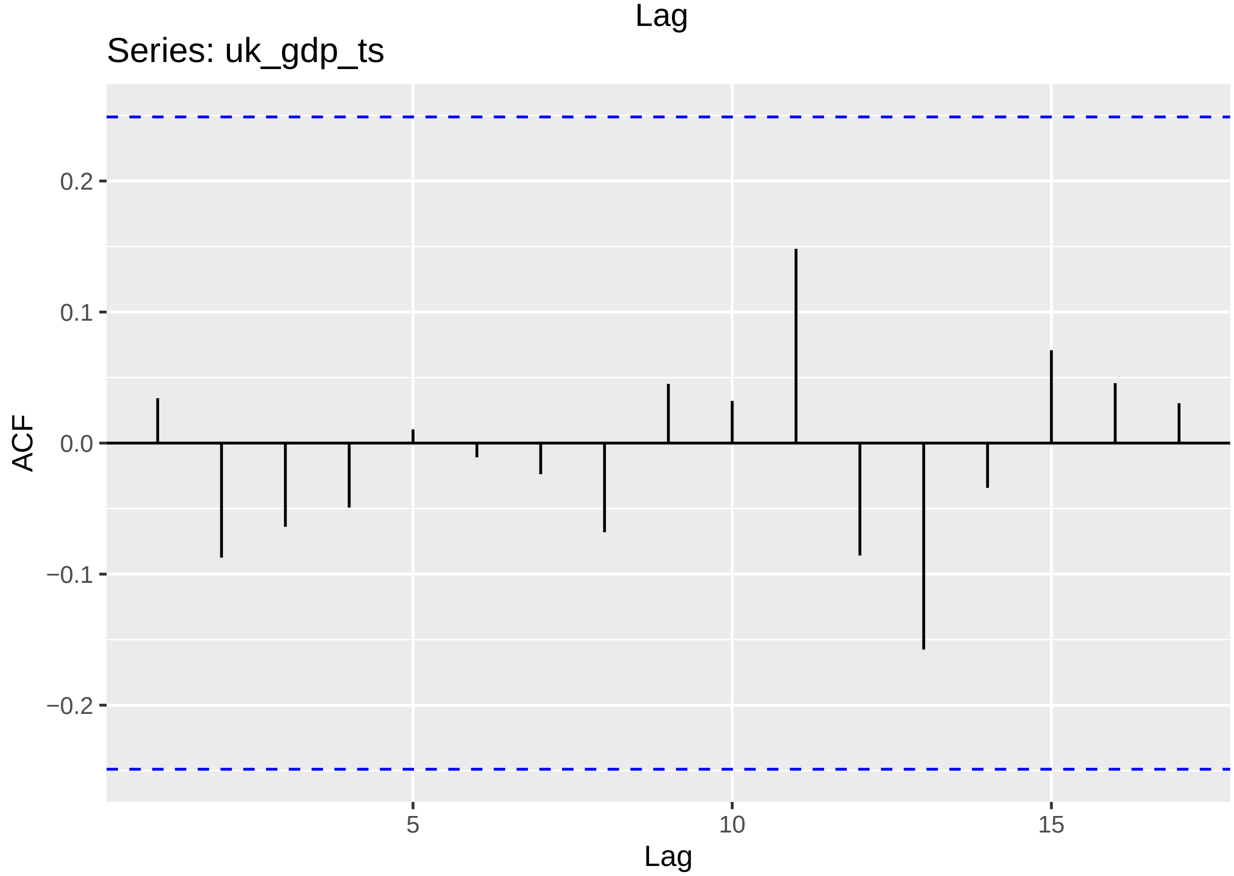

```r
ur.kpss(uk_gdp_ts) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.2418
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Create a time series model and then check for stationarity

```r
autoplot(acf(uk_gdp_ts))
```

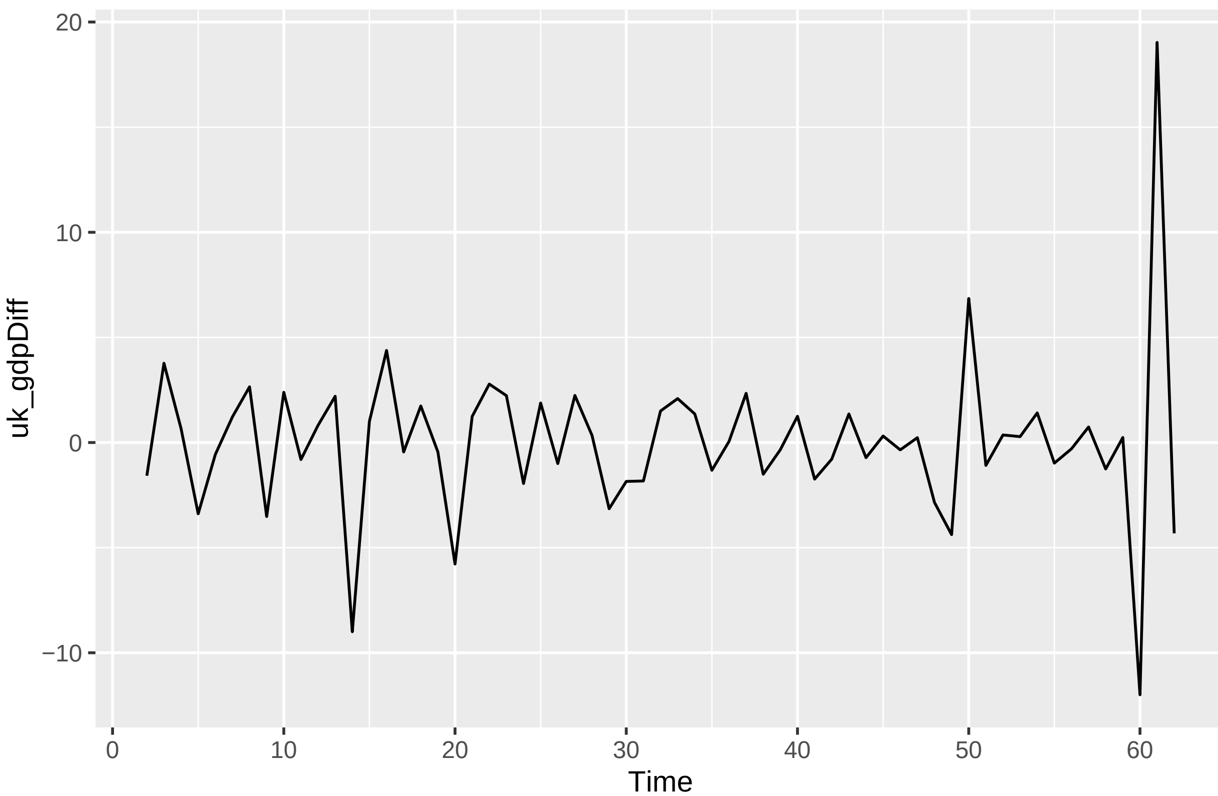





```
uk_gdpDiff = diff(uk_gdp_ts, lag = 1)

ur.kpss(uk_gdpDiff) %>% summary()
```

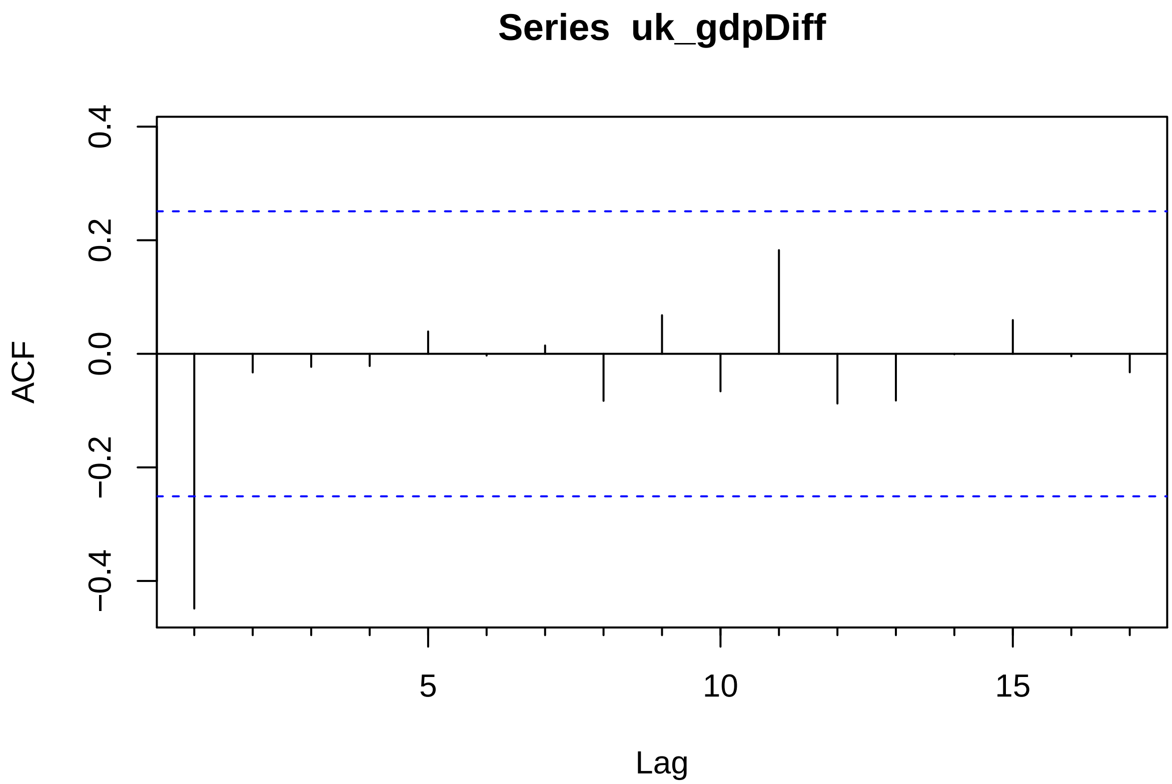
```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0388
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

```
autoplot(uk_gdpDiff)
```

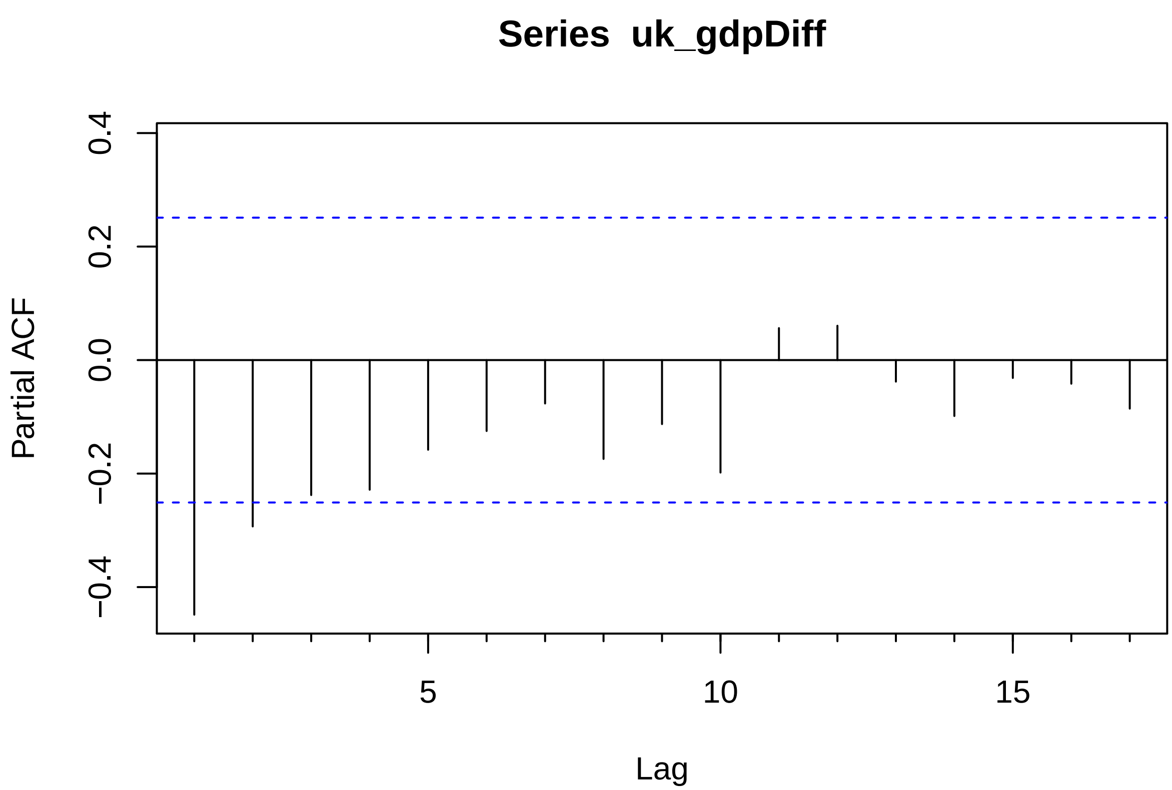


Difference once to get a better test statistic value. Then plot the afc and the pafc graphs to find the q and p values for ARIMA model

```
Acf(uk_gdpDiff) #1
```




```
Pacf(uk_gdpDiff) #1,2
```



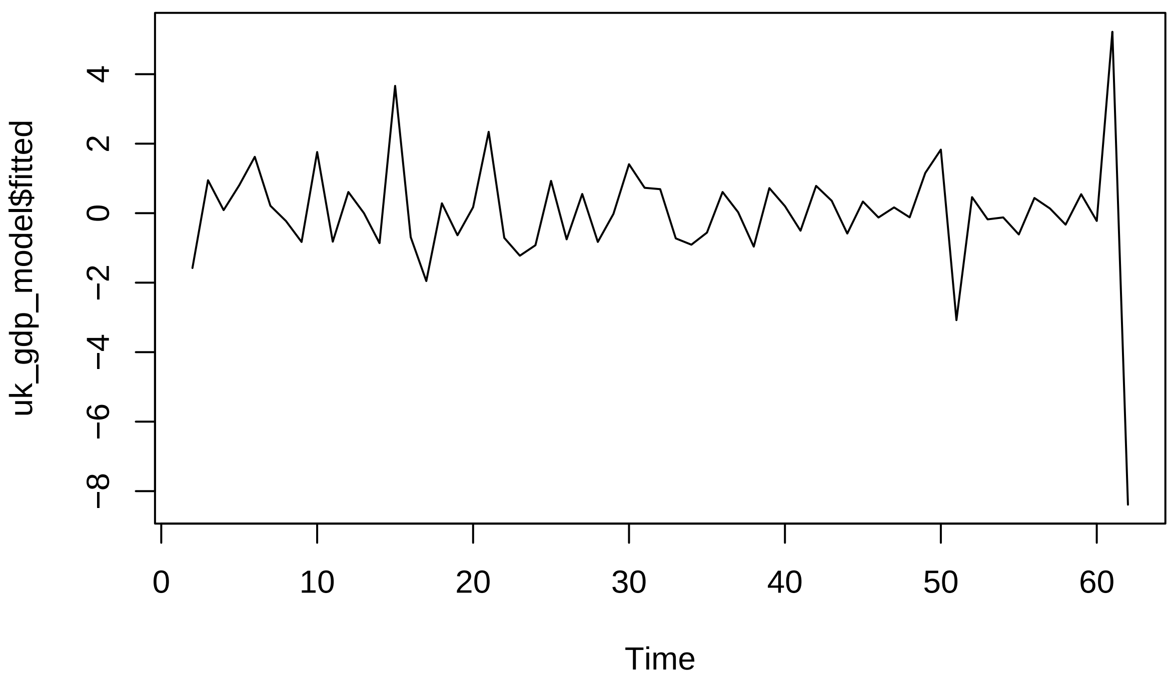

```
#d=1

#(1,1,1) has been the best so far (AICc = 296.07)
```

```r
uk_gdp_model <- Arima(uk_gdpDiff, order = c(1, 1, 1), method = "ML")
summary(uk_gdp_model)
```

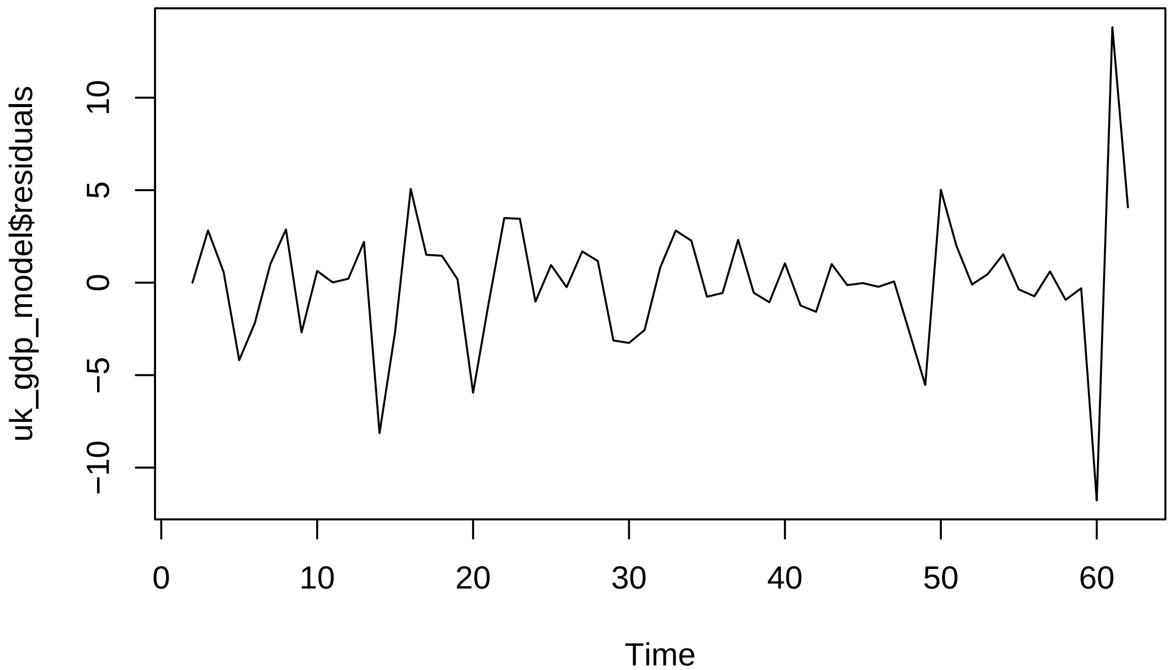
```
## Series: uk_gdpDiff
## ARIMA(1,1,1)
##
## Coefficients:
##           ar1      ma1
##       -0.4435  -1.0000
## s.e.   0.1155   0.0444
##
## sigma^2 = 11.99:  log likelihood = -161.16
## AIC=328.31   AICc=328.74   BIC=334.6
##
## Training set error measures:
##                     ME     RMSE      MAE      MPE     MAPE      MASE       ACF1
## Training set 0.0223211 3.375813 2.177859 49.41083 124.3914 0.5569497 -0.1140573
```

Once we find the p and q values, we try them for the ARIMA models to find the best AICc value Using the ARIMA(1,1,1) model gives the best AICc value.

```r
plot(uk_gdp_model$fitted)
```




```r
plot(uk_gdp_model$residuals)
```

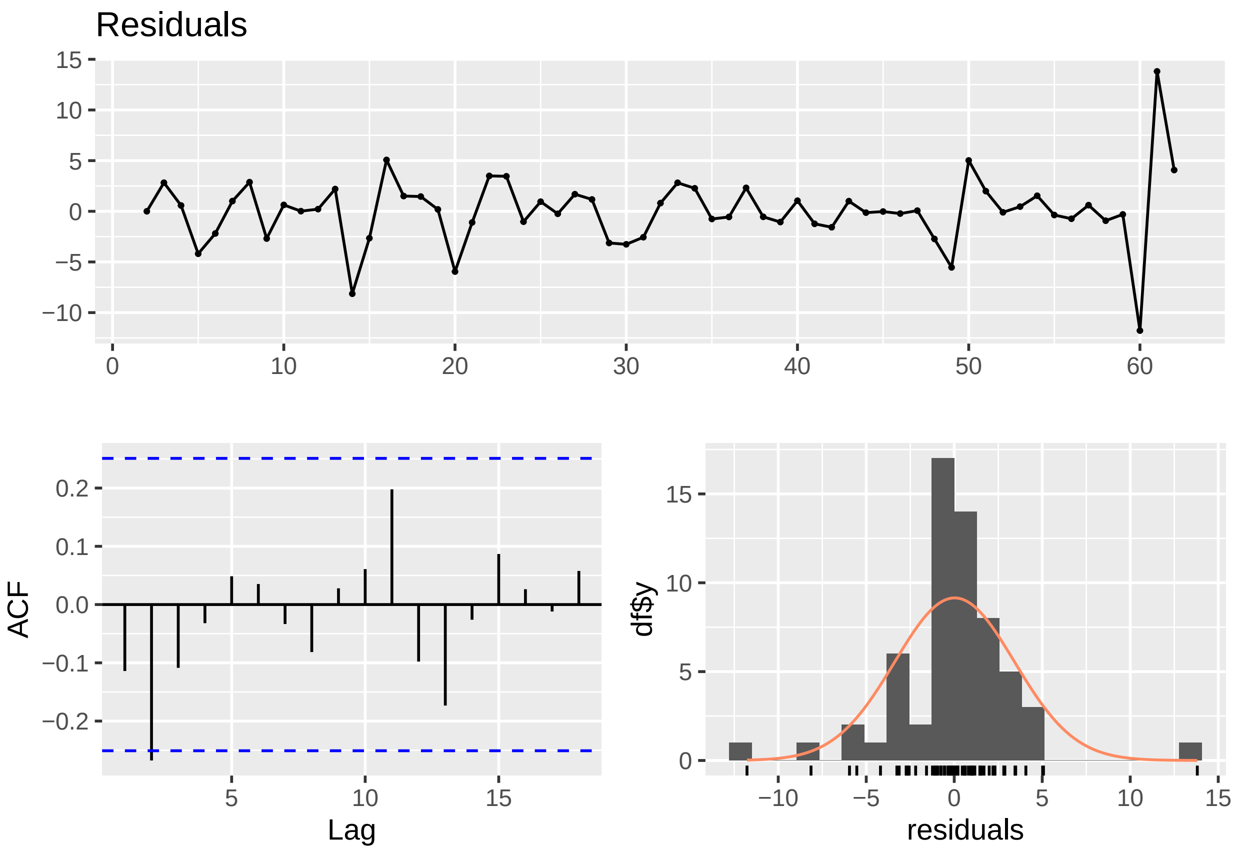



```r
#Check stationary of the residuals
ur.kpss(uk_gdp_model$residuals) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0796
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

We then look at the fitted and residual models and then conduct another kpss test for our model with the residuals which looks good

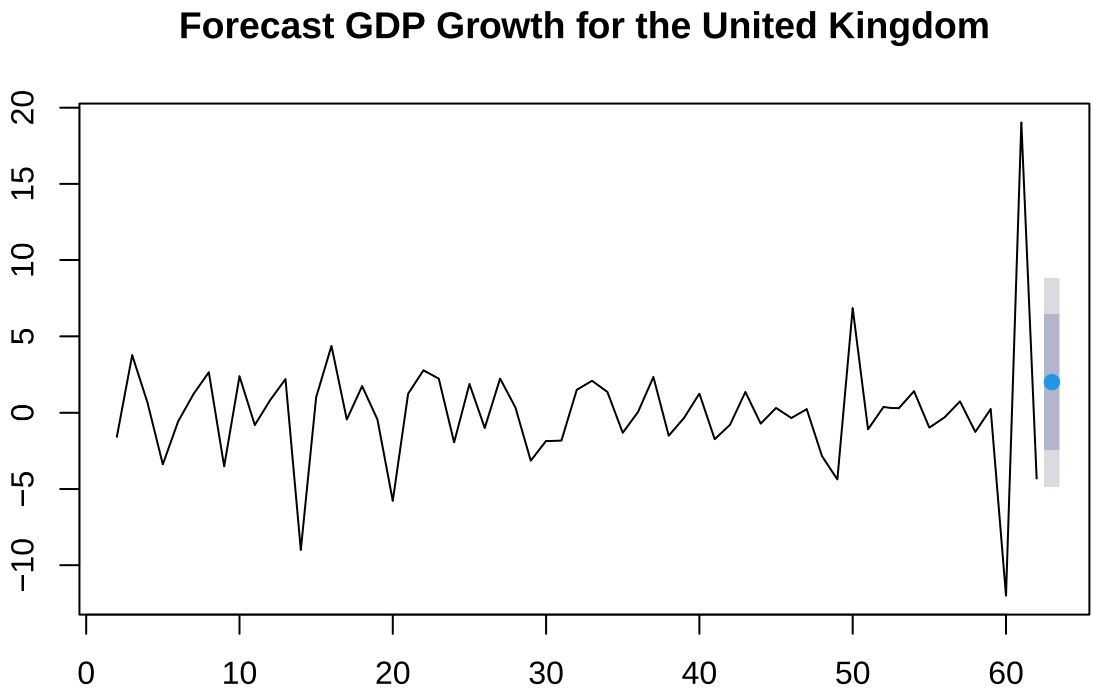
```r
checkresiduals(uk_gdp_model$residuals)
```




```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 7.4969, df = 10, p-value = 0.6778
##
## Model df: 0.   Total lags used: 10
```

Look for white noise which looks good

```
uk.forecast_values <- forecast(uk_gdp_model, h=1)
plot(uk.forecast_values, main = "Forecast GDP Growth for the United Kingdom")
```

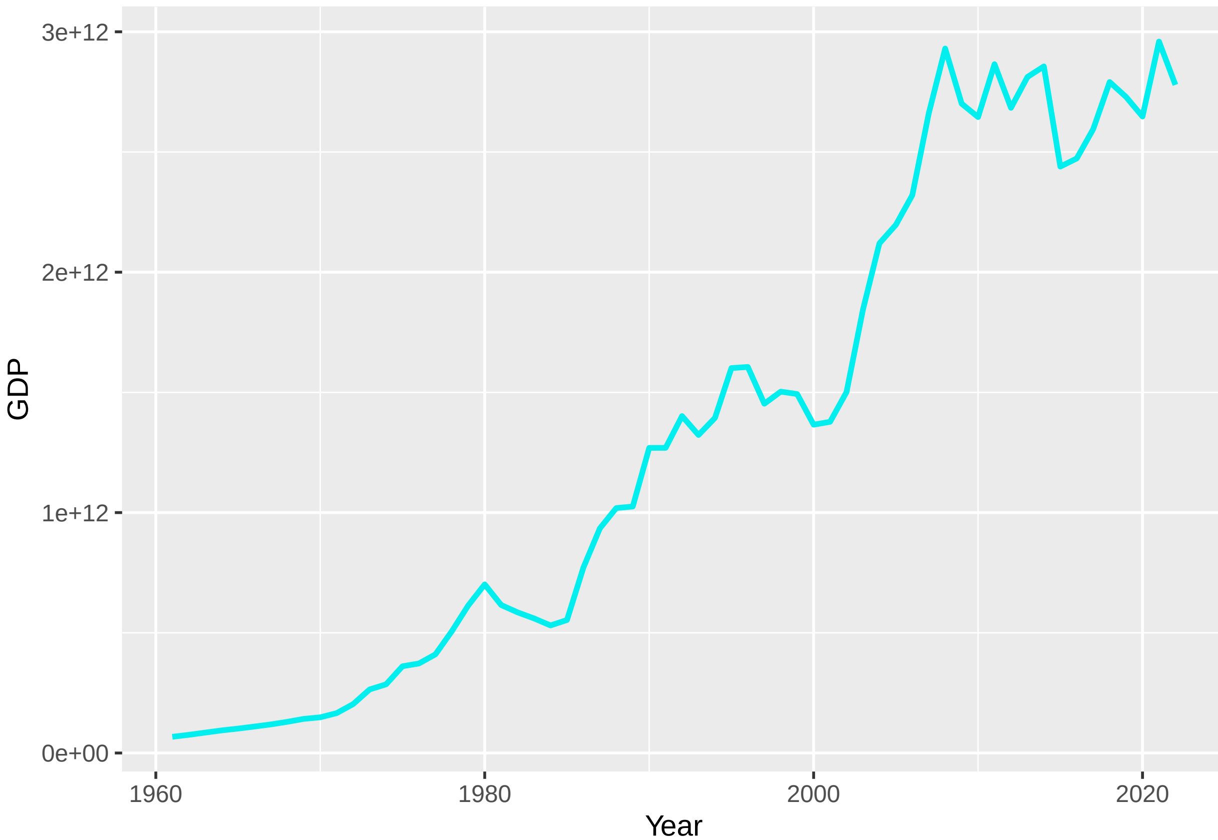


We then forecast the data for the future. Looking at the graph we can see that the gdp growth will decrease then go back to its mean value.

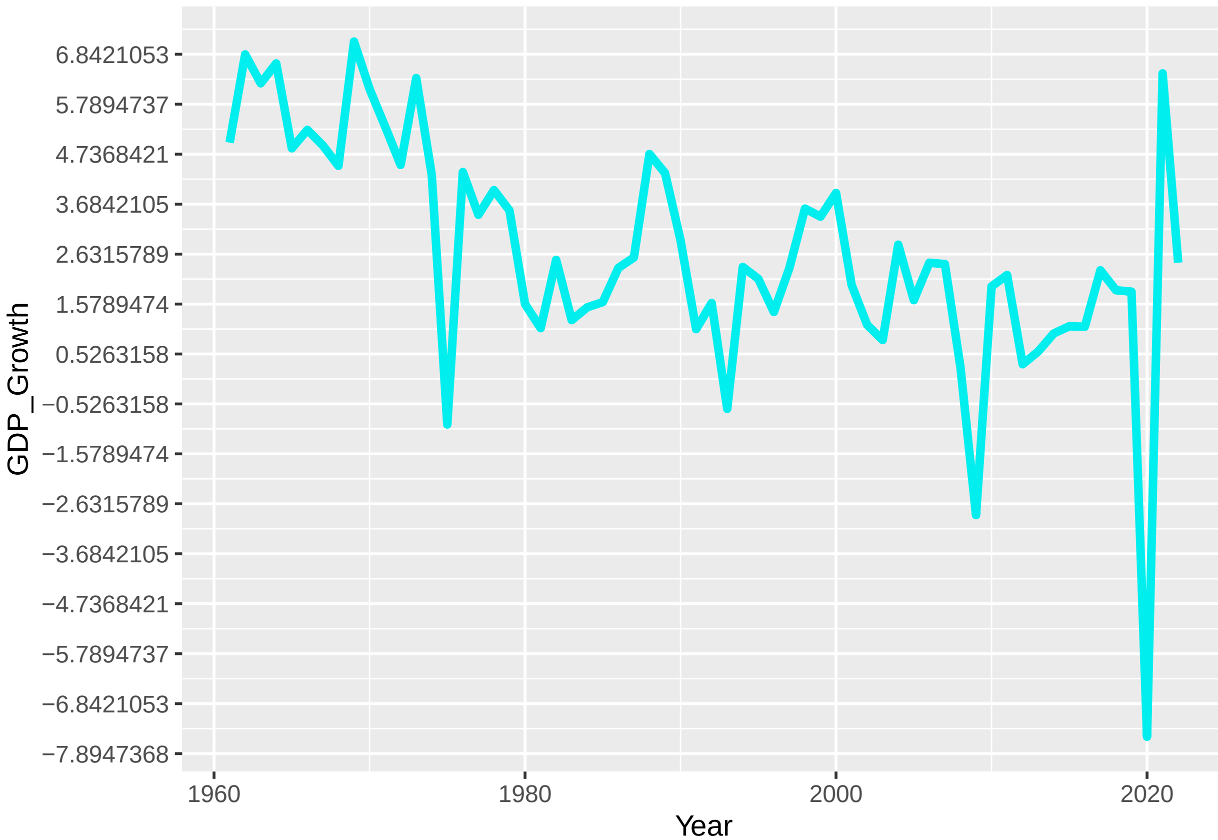
```
france_gdp <- gdp %>%
  filter(Country == "France") %>%
  select(Year, GDP, GDP_Growth)
```

France is the next country

```
ggplot(france_gdp) +
  geom_line(aes(Year, GDP, group = 1), color = "cyan2", linewidth = 1)
```




```r
ggplot(france_gdp, aes(Year, GDP_Growth, group = 1)) +
  geom_line(color = "cyan2", linewidth = 1.5) +
  scale_y_continuous(
    breaks = seq(-10, 10, by = 20/19))
```

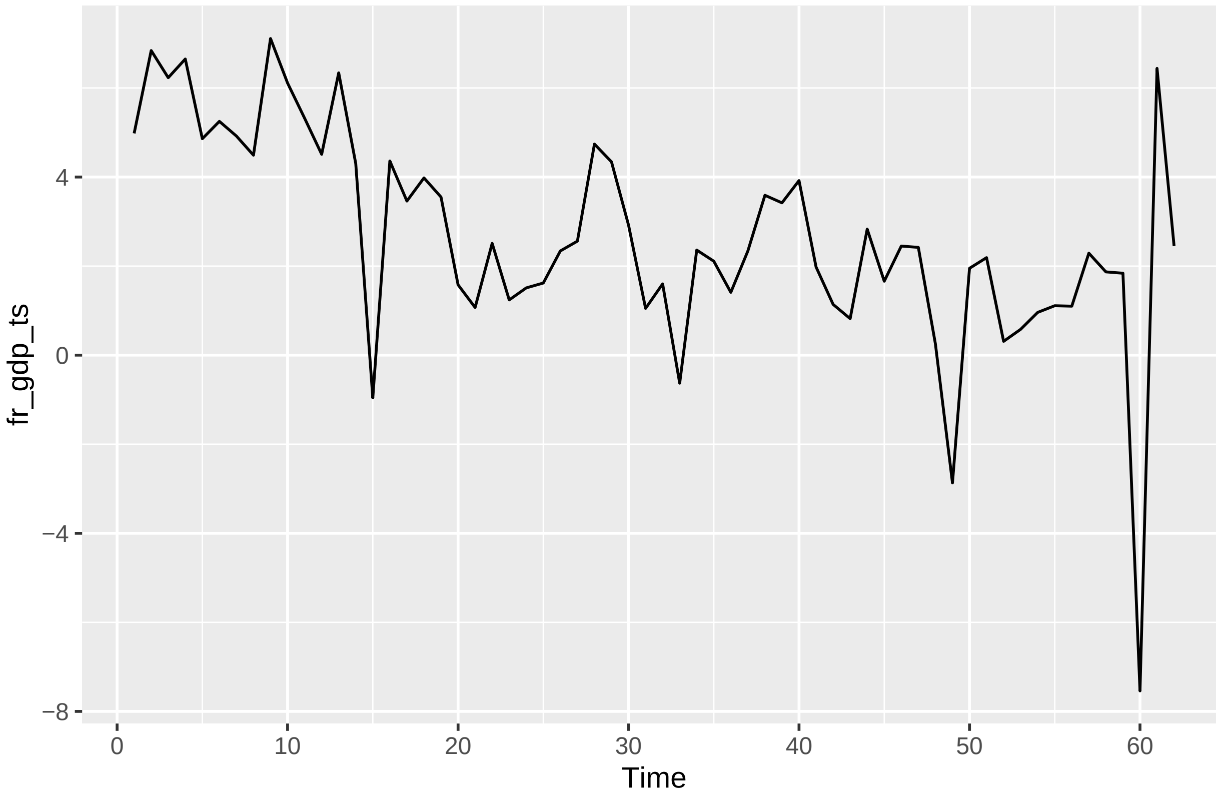


Here are the graphs showing the gdp and gdp growth and the gradual decrease in total gdp

```
ftrain <- france_gdp[1:50,]
ftest <- france_gdp[51:62,]
fntest <- nrow(ftest)
```

Then make the train and test statistic

```
fr_gdp_ts <- ts(france_gdp$GDP_Growth)
autoplot(fr_gdp_ts)
```

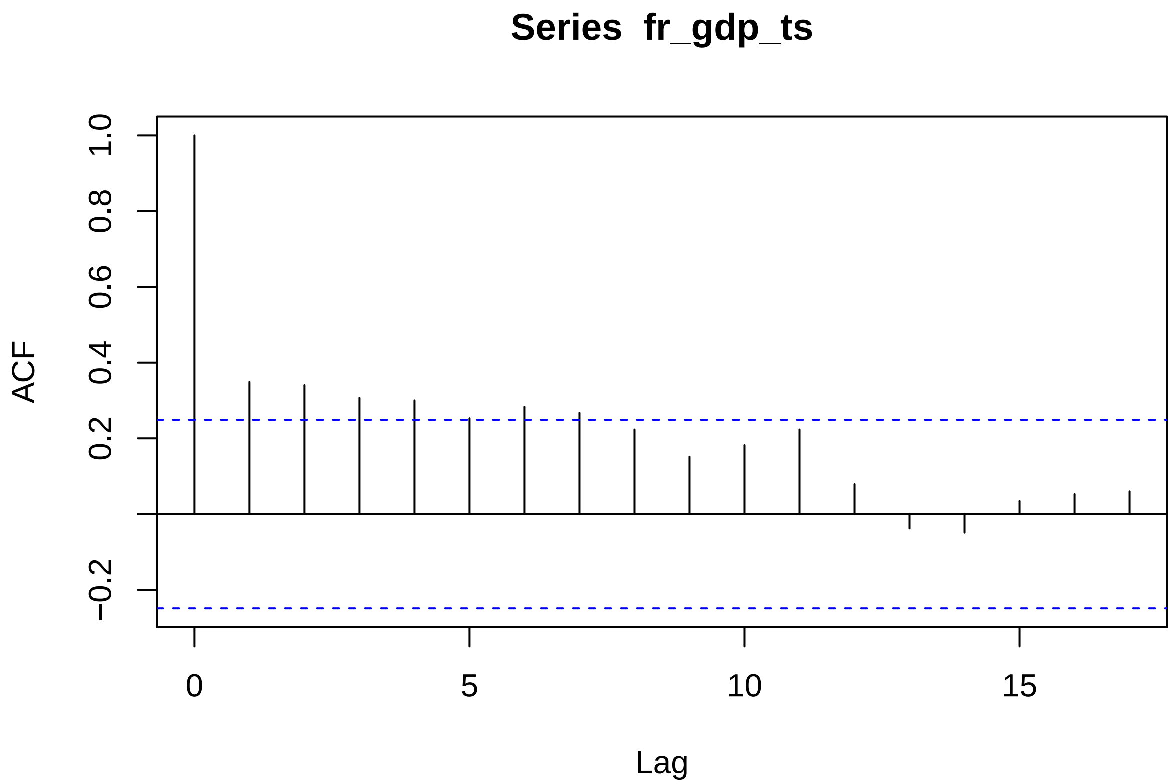


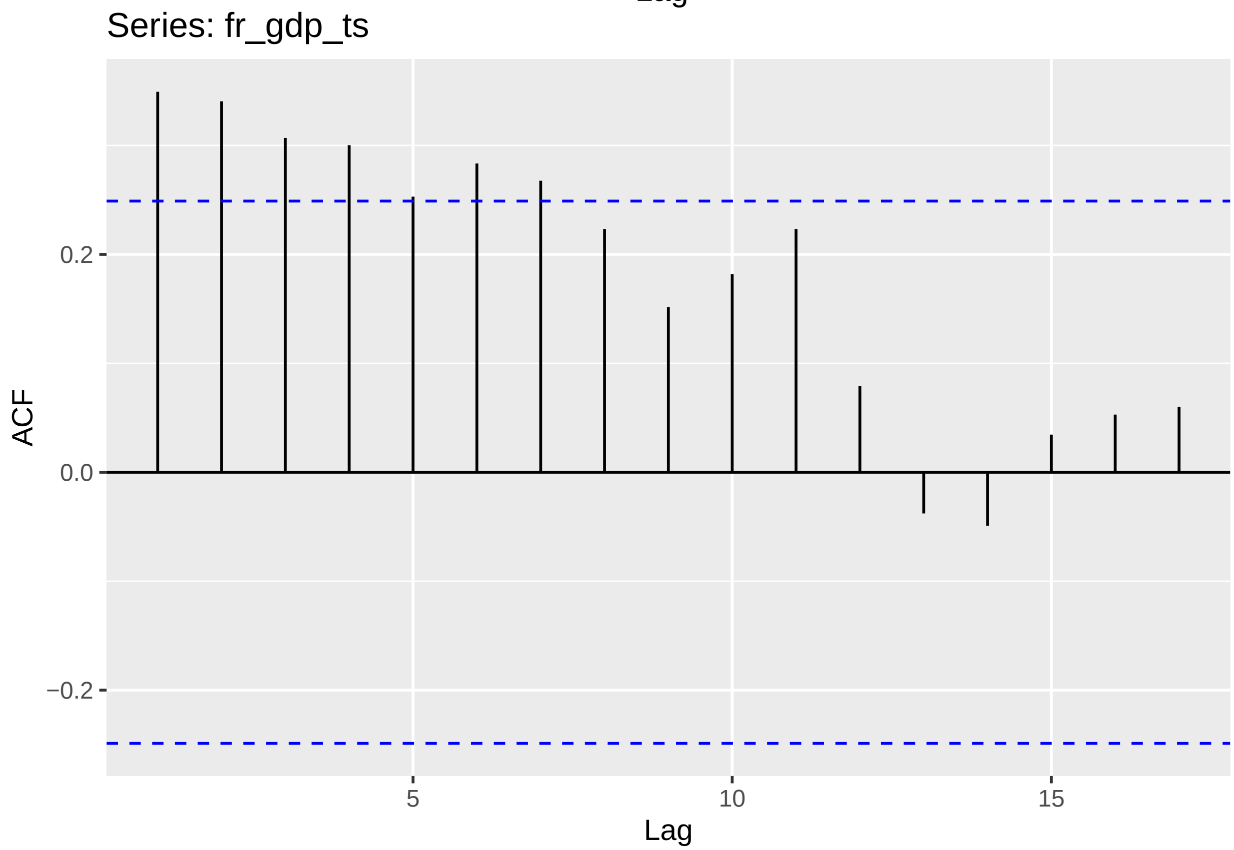

```
ur.kpss(fr_gdp_ts) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 1.1105
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Create time series and the plot the data. Preform kpss test to see if the null hypothesis is accepted or not. If not, then differentiate

```
autoplot(acf(fr_gdp_ts))
```

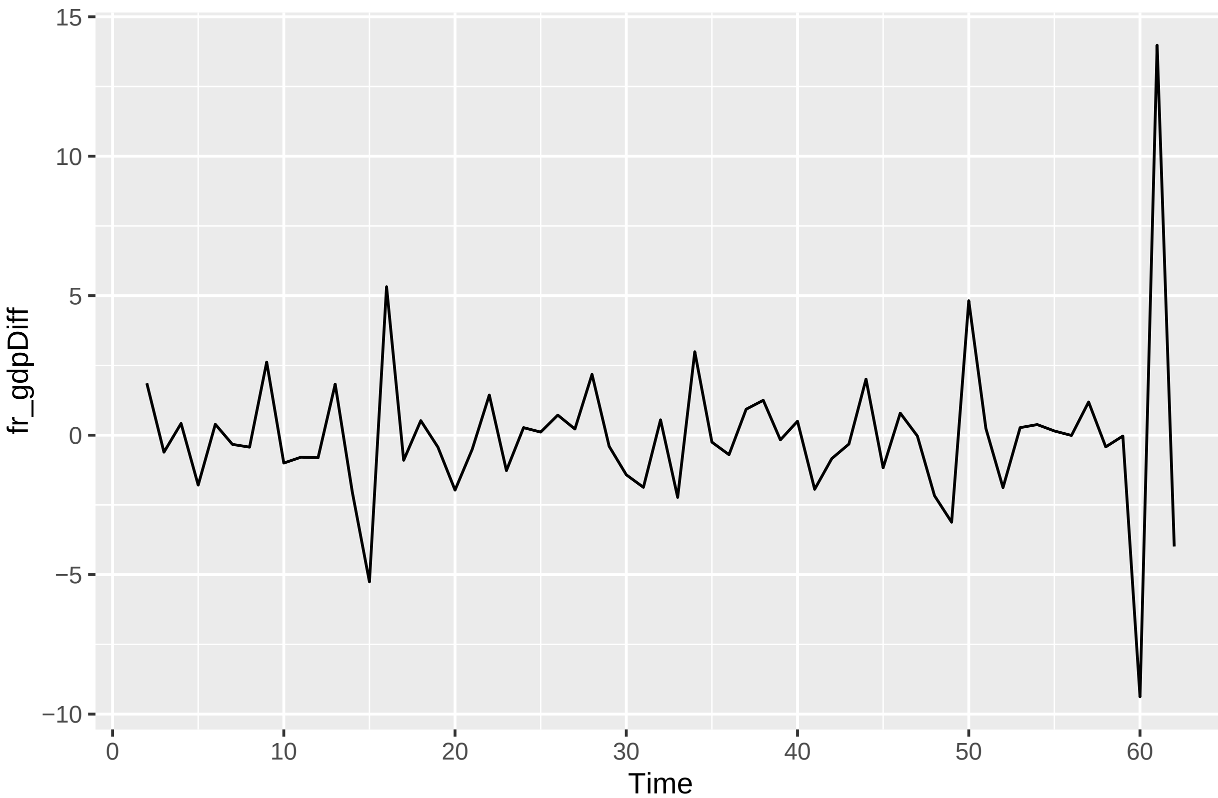





```
fr_gdpDiff = diff(fr_gdp_ts, lag = 1)

ur.kpss(fr_gdpDiff) %>% summary()
```

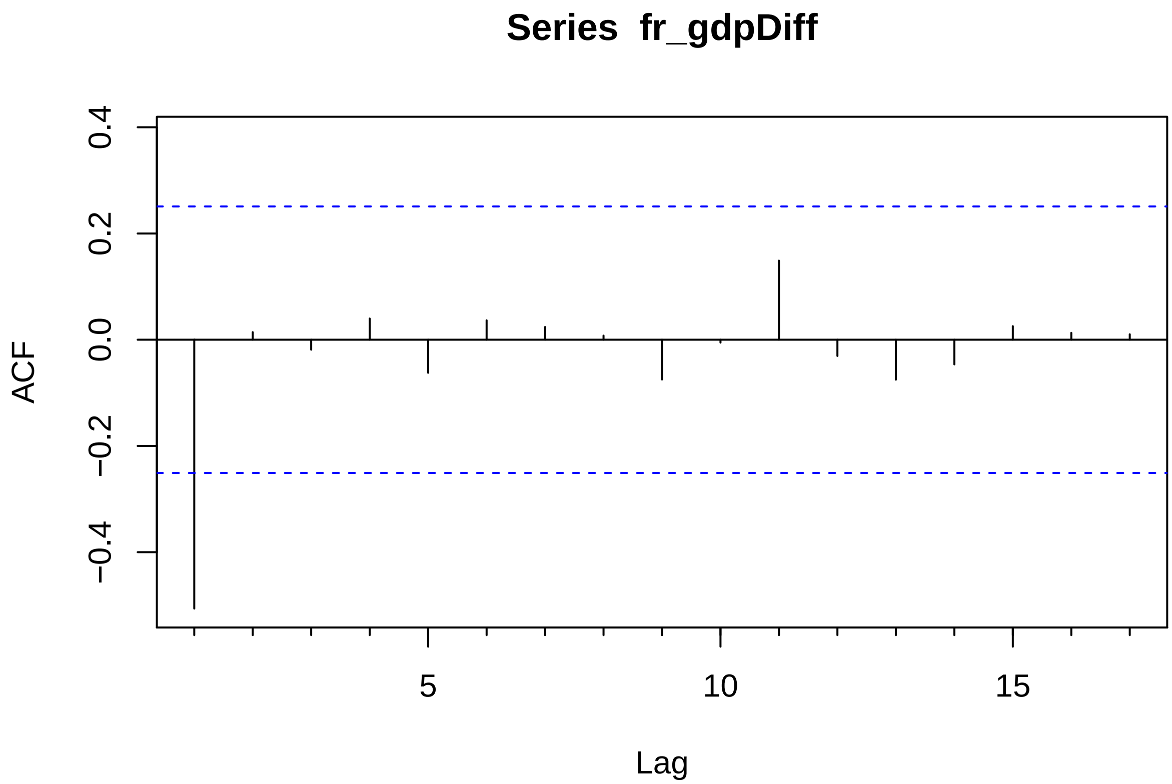
```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0473
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

```
autoplot(fr_gdpDiff)
```

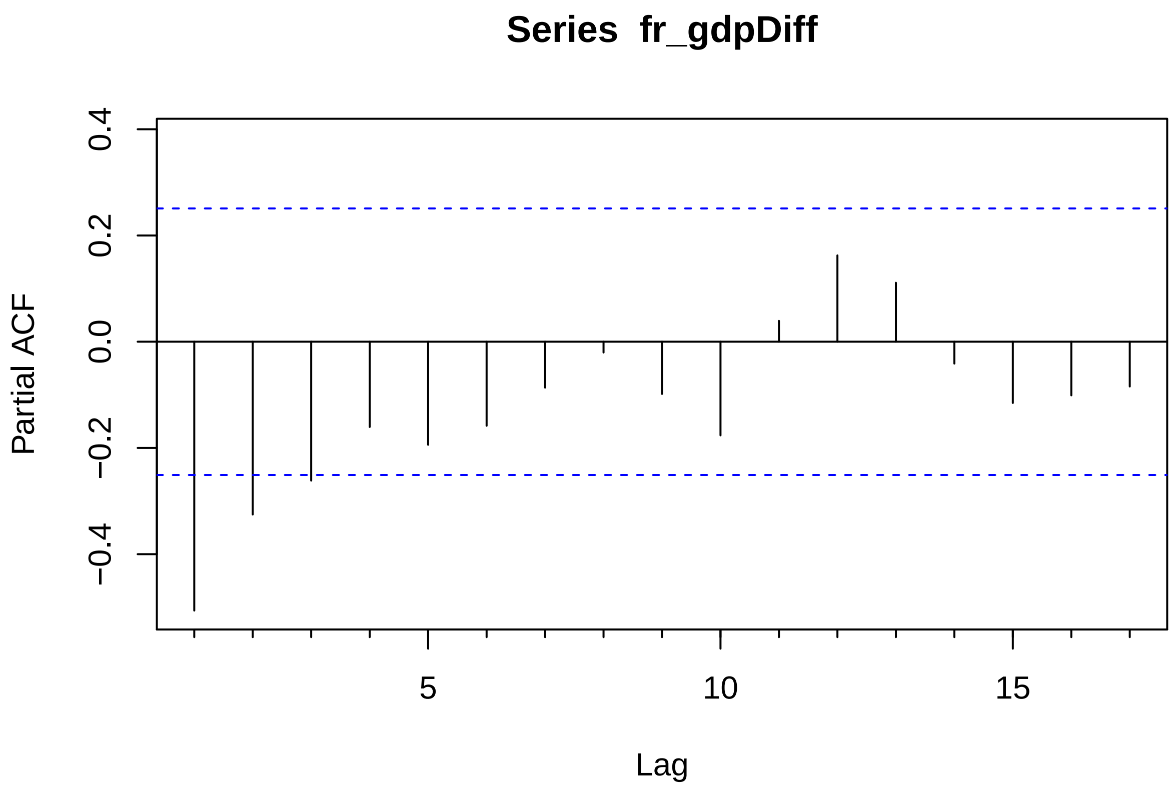


After preforming one level of diffrencing, we can then see that we can accept the null hypothesis and then use the afc and pacf plots to use for the ARIMA model of the data

```
Acf(fr_gdpDiff) #1
```




```
Pacf(fr_gdpDiff) #1,2,3
```



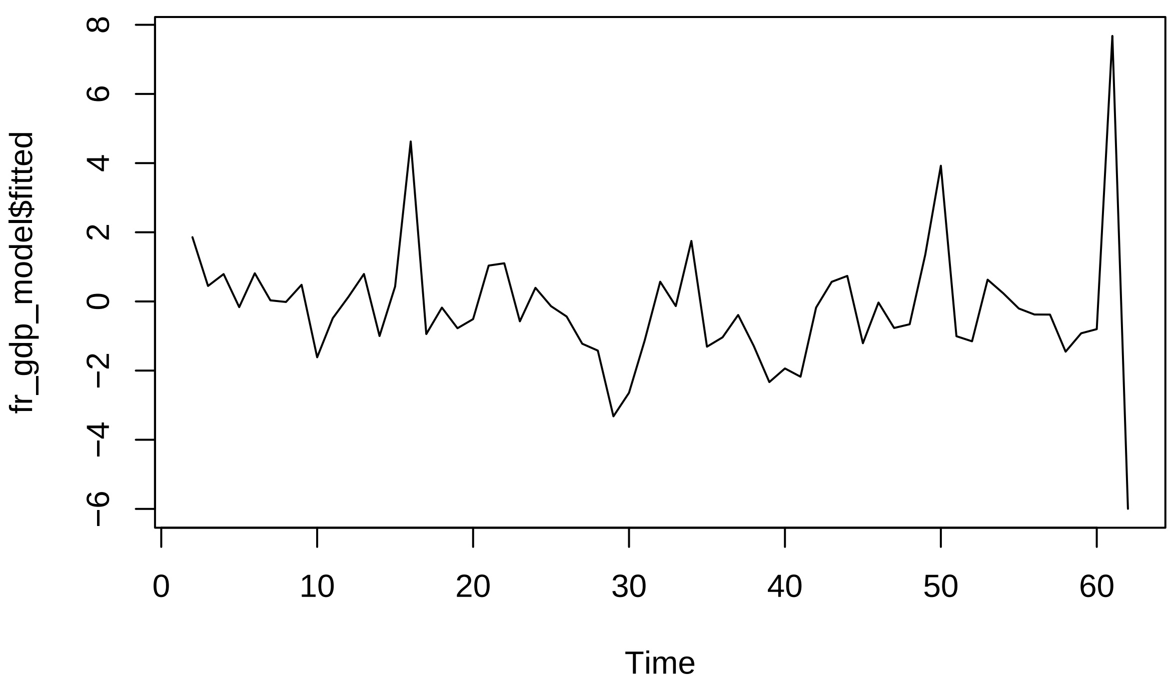

```
#d=1

#(1,1,2) has been the best so far (AICc = 273.42)
```

```r
fr_gdp_model <- Arima(fr_gdpDiff, order = c(1, 1, 2), method = "ML")
summary(fr_gdp_model)
```

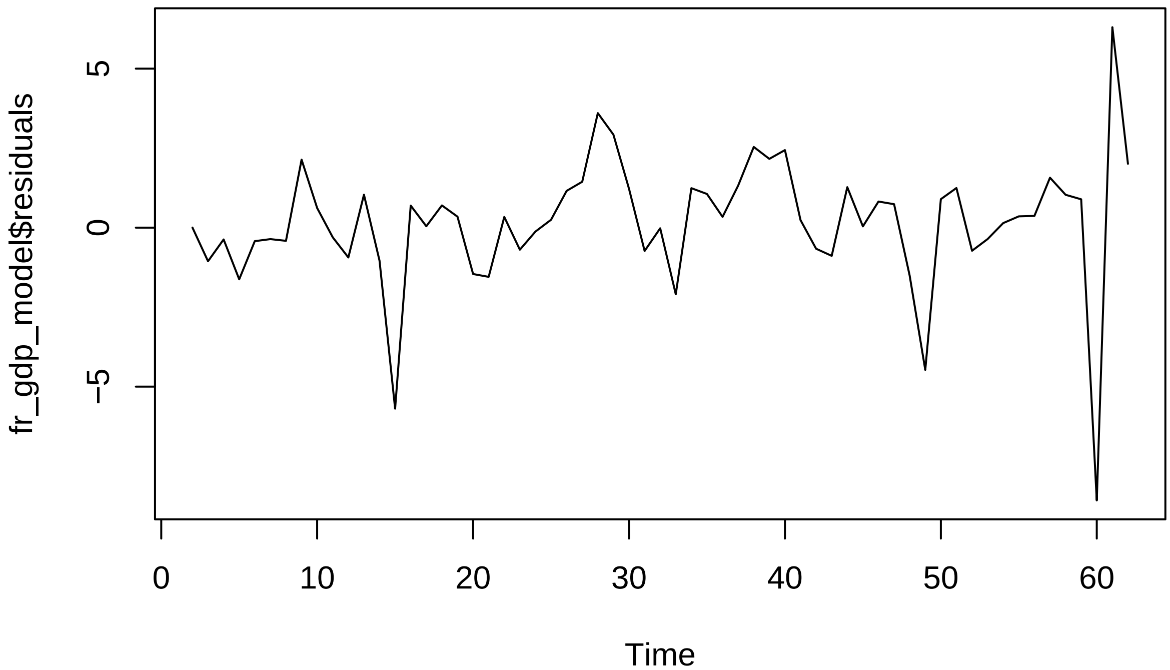
```
## Series: fr_gdpDiff
## ARIMA(1,1,2)
##
## Coefficients:
##           ar1      ma1     ma2
##       -0.0321  -1.8627  0.8730
## s.e.   0.1485   0.0895  0.0884
##
## sigma^2 = 4.481:  log likelihood = -132.35
## AIC=272.69   AICc=273.42   BIC=281.07
##
## Training set error measures:
##                     ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 0.1543819 2.046353 1.337992 -113.3343 296.7573 0.4843408
##                     ACF1
## Training set -0.02545161
```

Once we find the p and q values, we try them for the ARIMA models to find the best AICc value Using the ARIMA(1,1,2) model gives the best AICc value. France has the best AICc value of all the countries in our data.

```r
plot(fr_gdp_model$fitted)
```




```r
plot(fr_gdp_model$residuals)
```

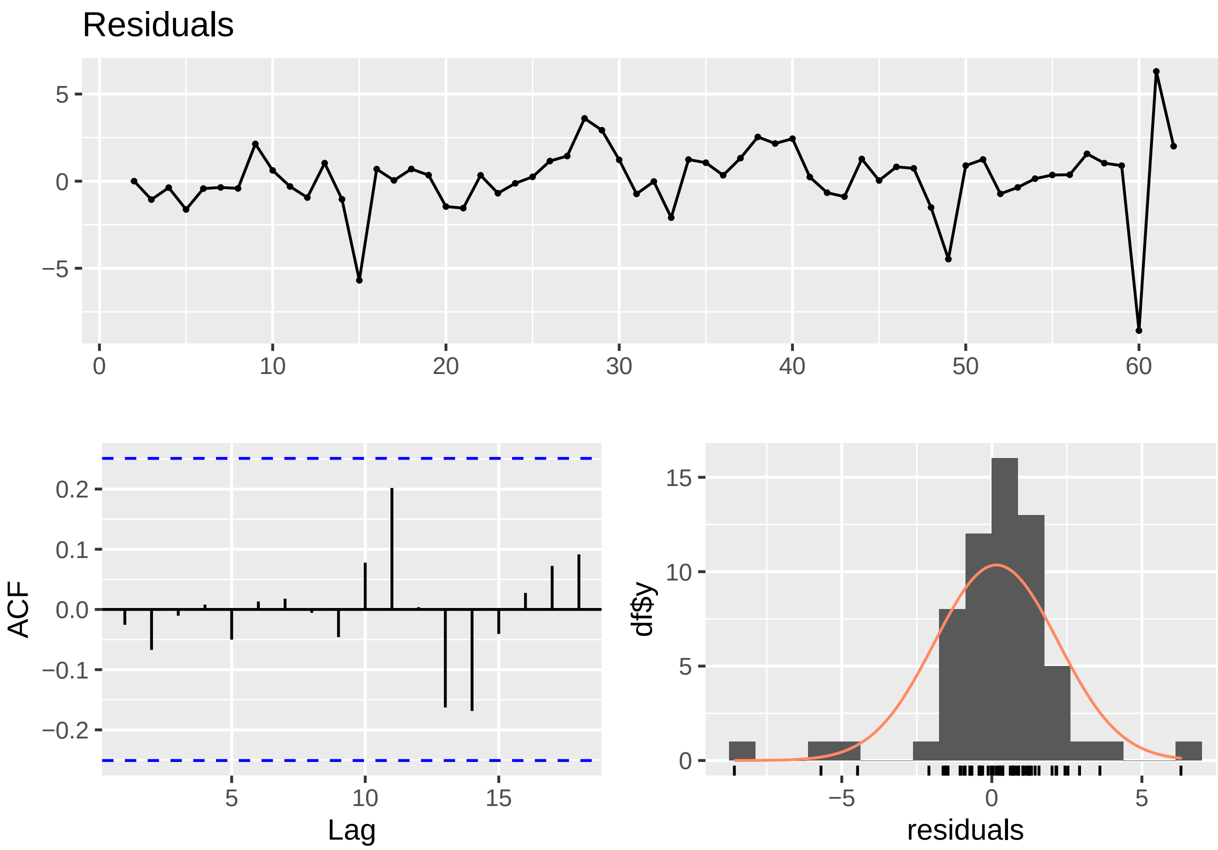



```r
#Check stationary of the residuals
ur.kpss(fr_gdp_model$residuals) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.1769
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

THen plot the fitted and residual models and check the stationary of the residuals

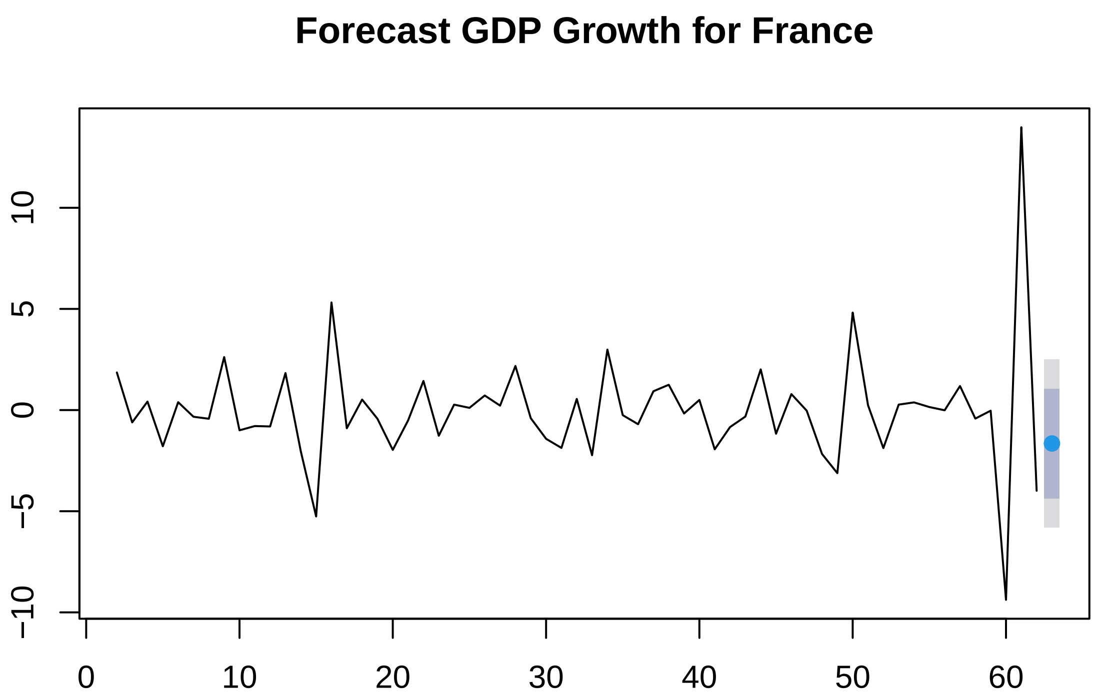
```r
checkresiduals(fr_gdp_model$residuals)
```




```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 1.1671, df = 10, p-value = 0.9997
##
## Model df: 0.   Total lags used: 10
```

Then we can look for white noise which our data shoes none of

```
f.forecast_values <- forecast(fr_gdp_model, h=1)
plot(f.forecast_values, main = "Forecast GDP Growth for France")
```

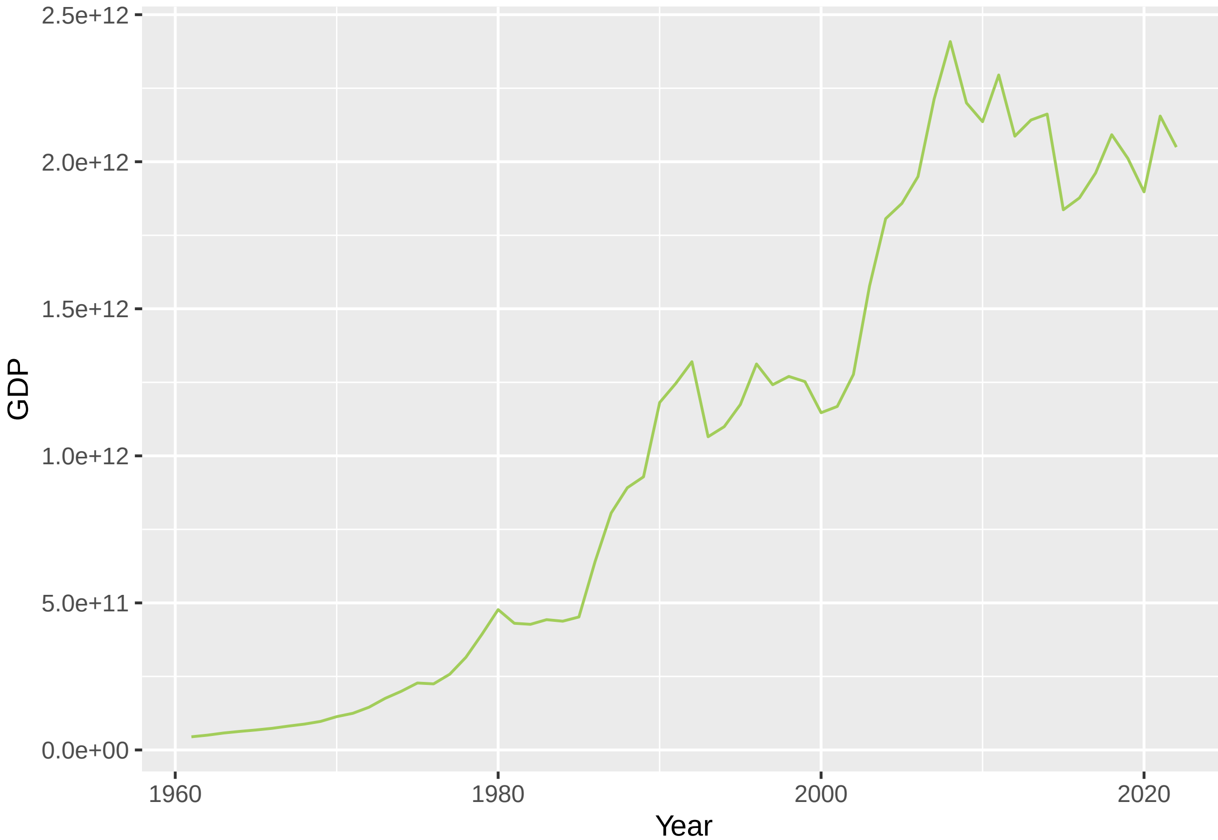


Then we can forecast the data. Like most of the graphs, the gdp decreased during covid and jumped up afterwards and is now starting to mellow out following the large spike.

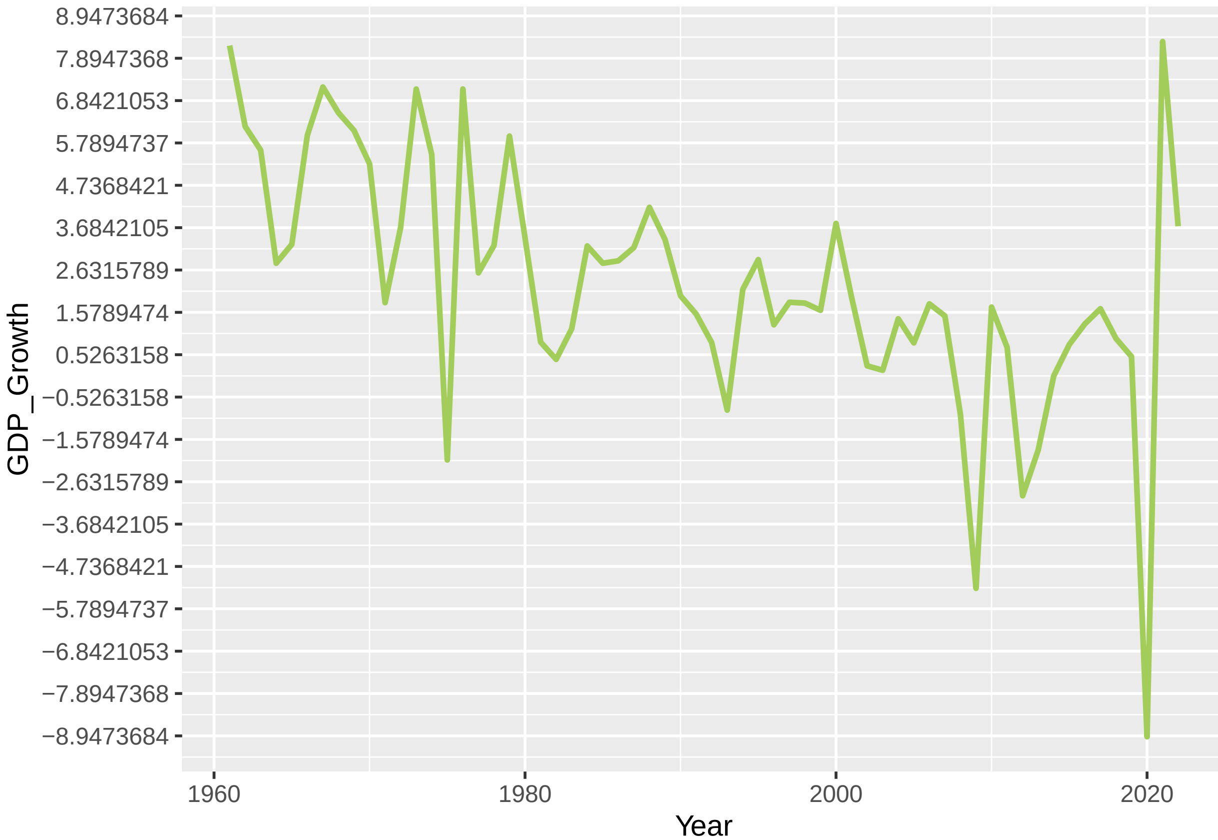
```
italy_gdp <- gdp %>%
  filter(Country == "Italy") %>%
  select(Year, GDP, GDP_Growth)
```

Italy is up next

```
ggplot(italy_gdp) +
  geom_line(aes(Year, GDP, group = 1), color = "darkolivegreen3", linewidth = .5)
```




```r
ggplot(italy_gdp, aes(Year, GDP_Growth, group = 1)) +
  geom_line(color = "darkolivegreen3", linewidth = 1) +
  scale_y_continuous(
    breaks = seq(-10, 10, by = 20/19))
```

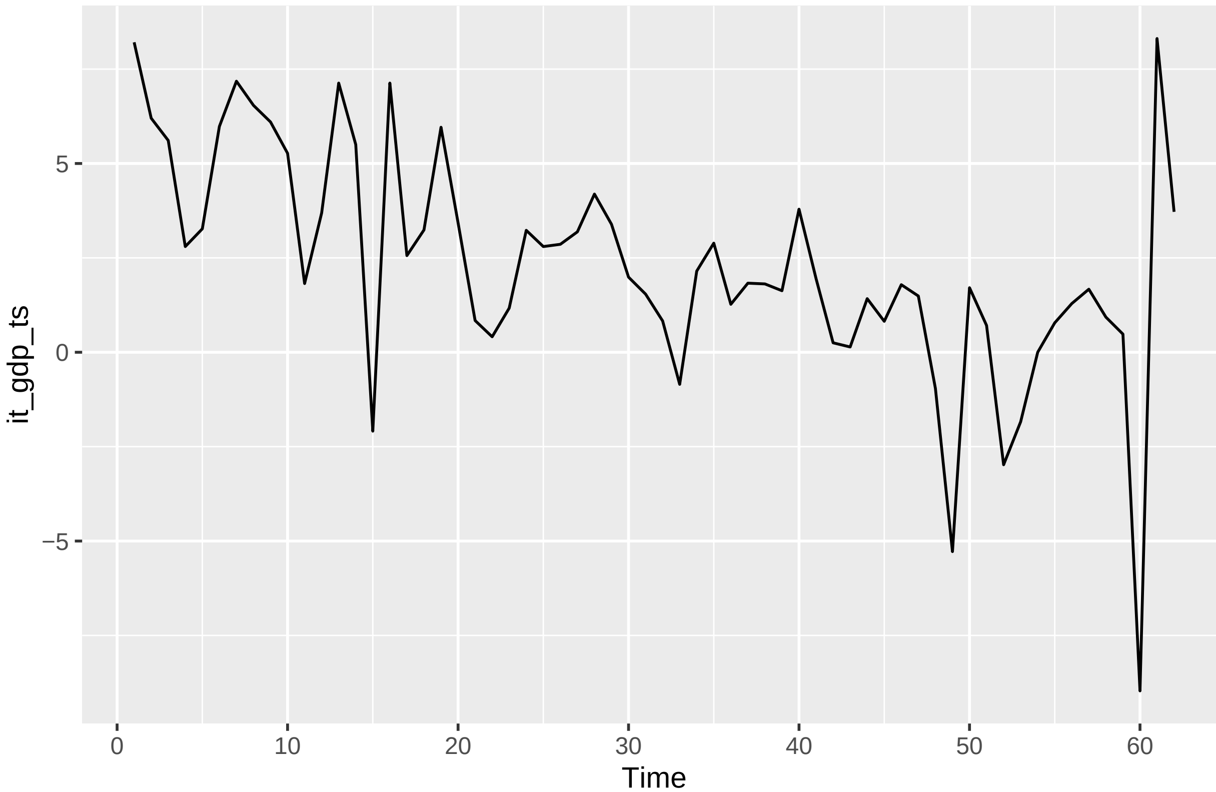


The data shows us that Italy has a gradual decerase of their gdp over time and then plummits during covid.

```
ittrain <- italy_gdp[1:50,]
ittest <- italy_gdp[51:62,]
itntest <- nrow(ittest)
```

Training and testing sets

```
it_gdp_ts <- ts(italy_gdp$GDP_Growth)
autoplot(it_gdp_ts)
```

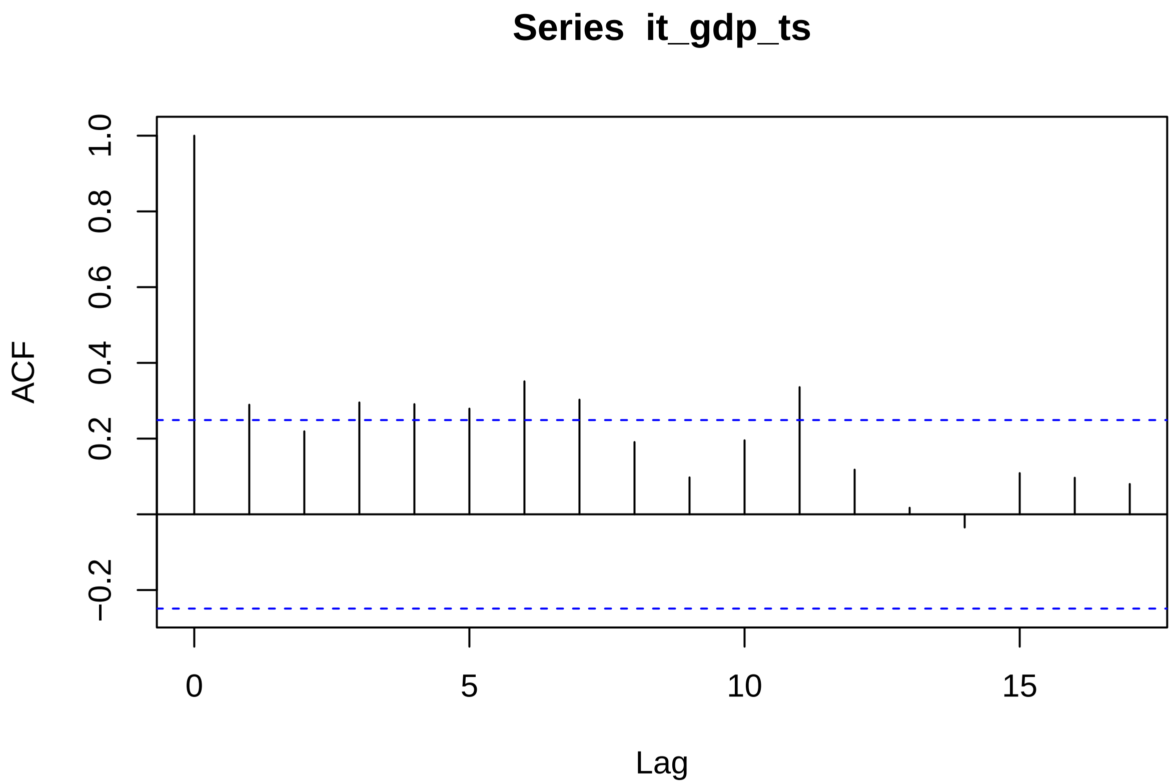


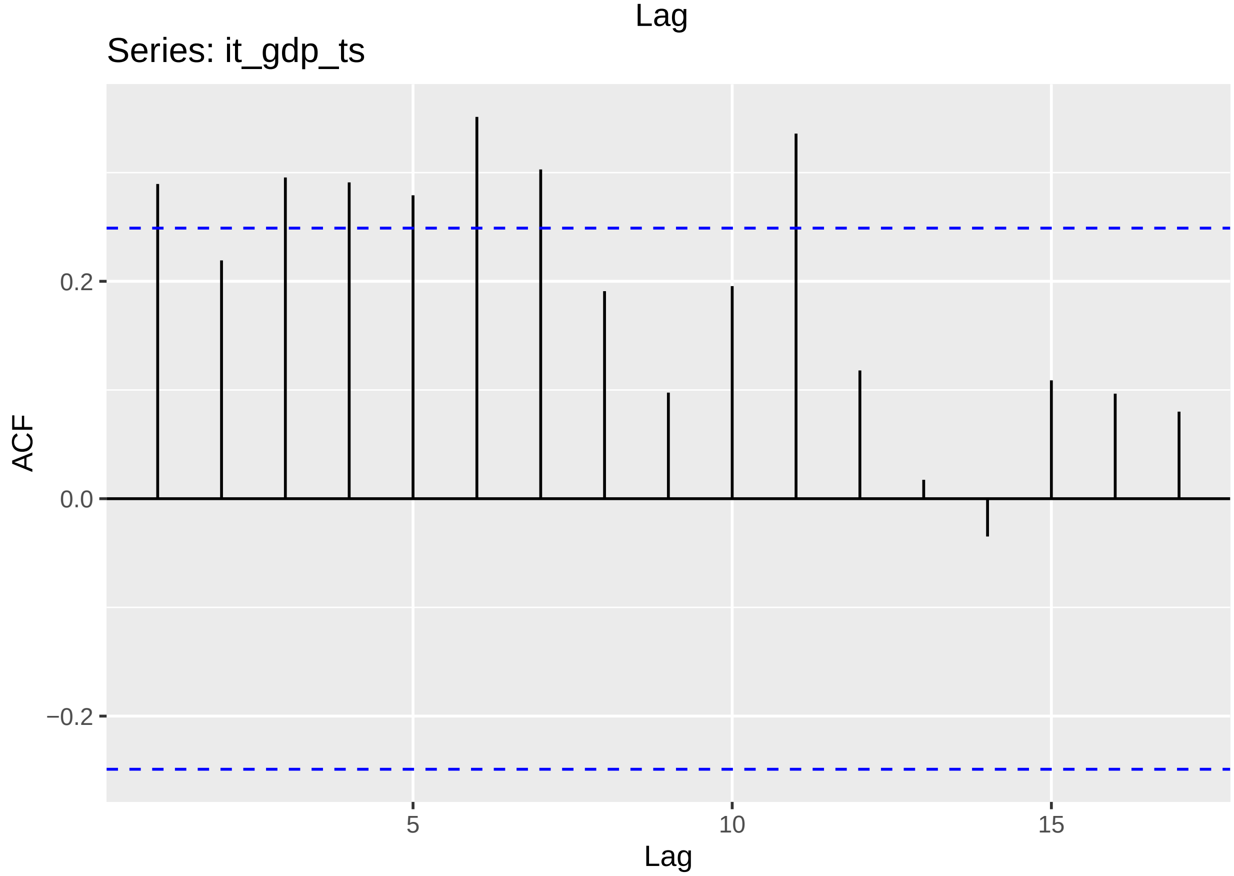

```r
ur.kpss(it_gdp_ts) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 1.2217
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Then we made a time series for the italy data and then preform a kpss test to check the stationary of it

```r
autoplot(acf(it_gdp_ts))
```

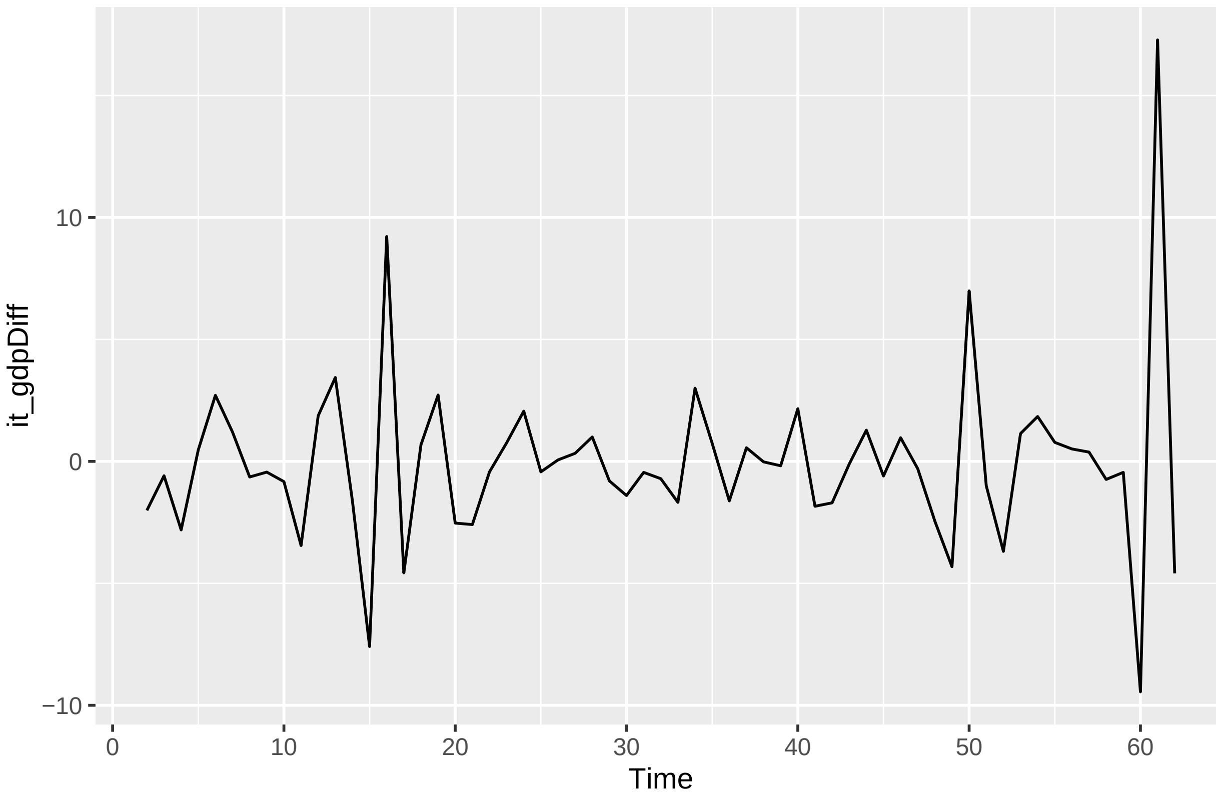





```
it_gdpDiff = diff(it_gdp_ts, lag = 1)

ur.kpss(it_gdpDiff) %>% summary()
```

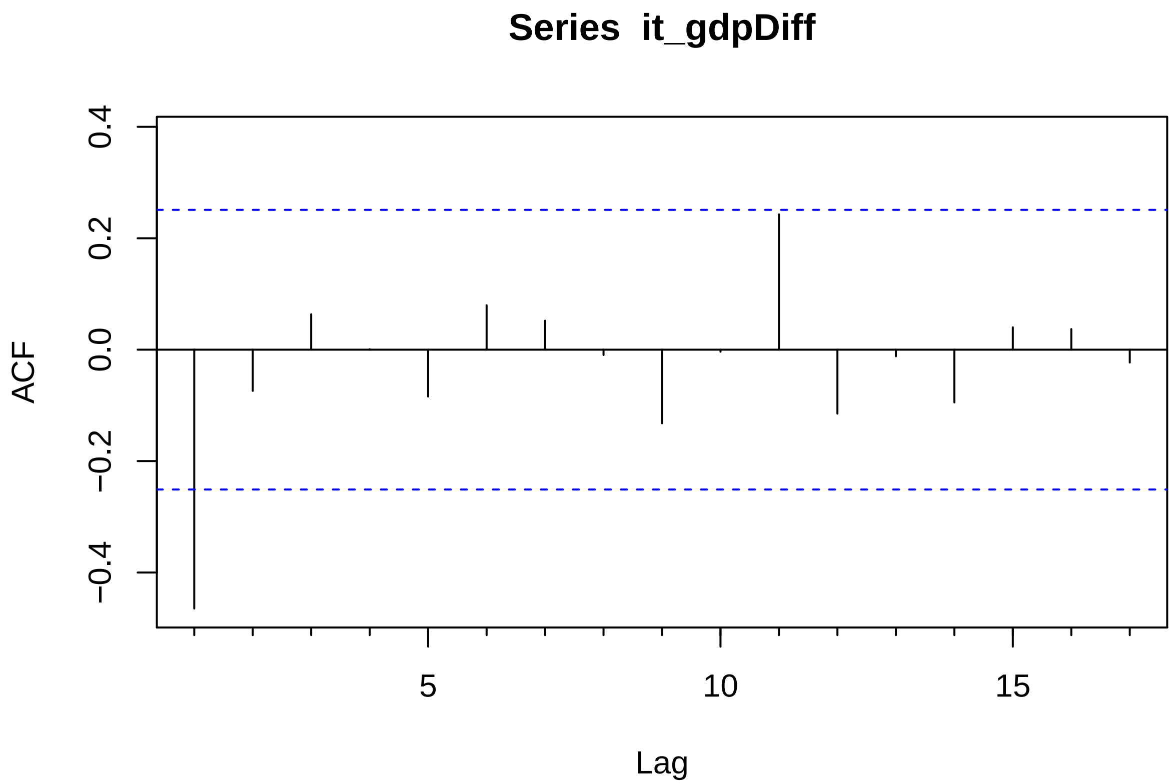
```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0982
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

```
autoplot(it_gdpDiff)
```

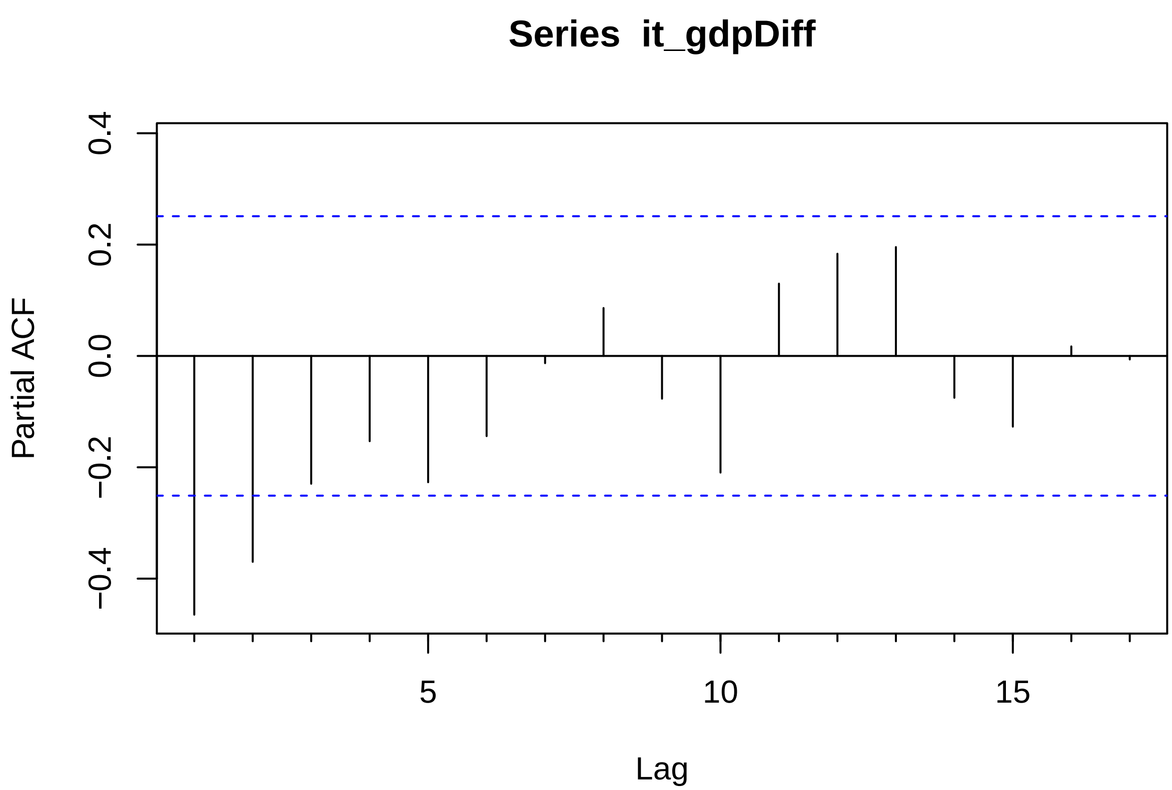


We then diffrence the data to find a better value for our test statistic. Once found, we can then preform the ARIMA test

```
Acf(it_gdpDiff) #1
```




```r
Pacf(it_gdpDiff) #1,2
```



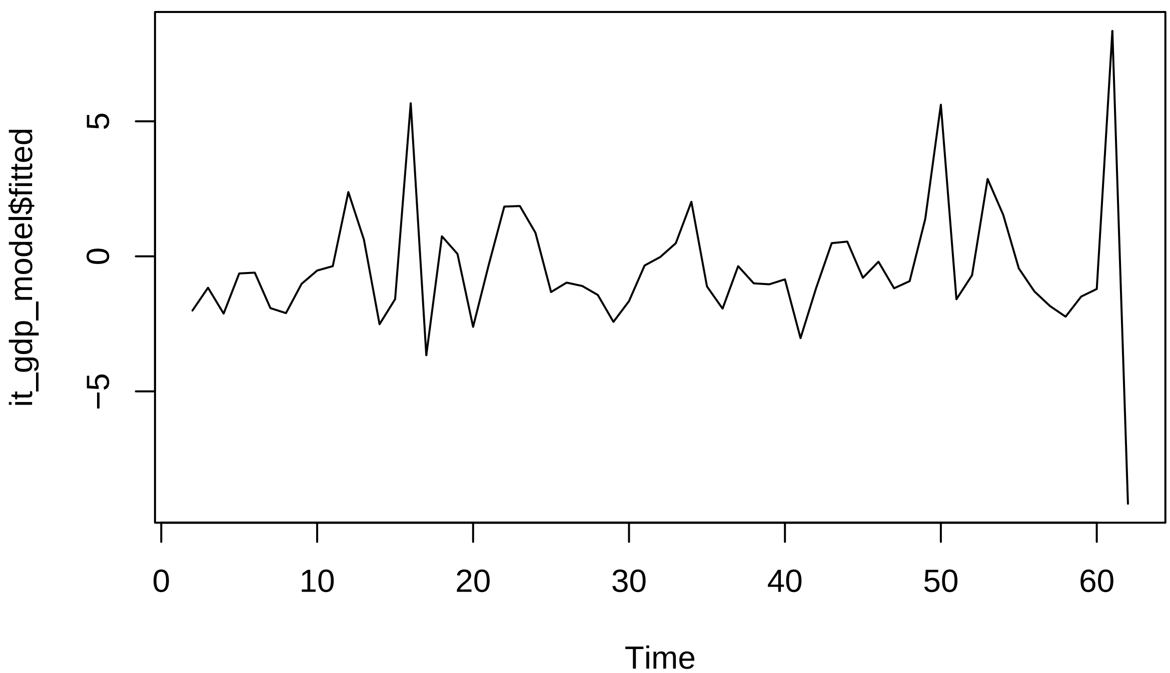

```r
#d=1

#(1,1,2) has been the best so far (AICc = 301.75)
```

```r
it_gdp_model <- Arima(it_gdpDiff, order = c(1, 1, 2), method = "ML")
summary(it_gdp_model)
```

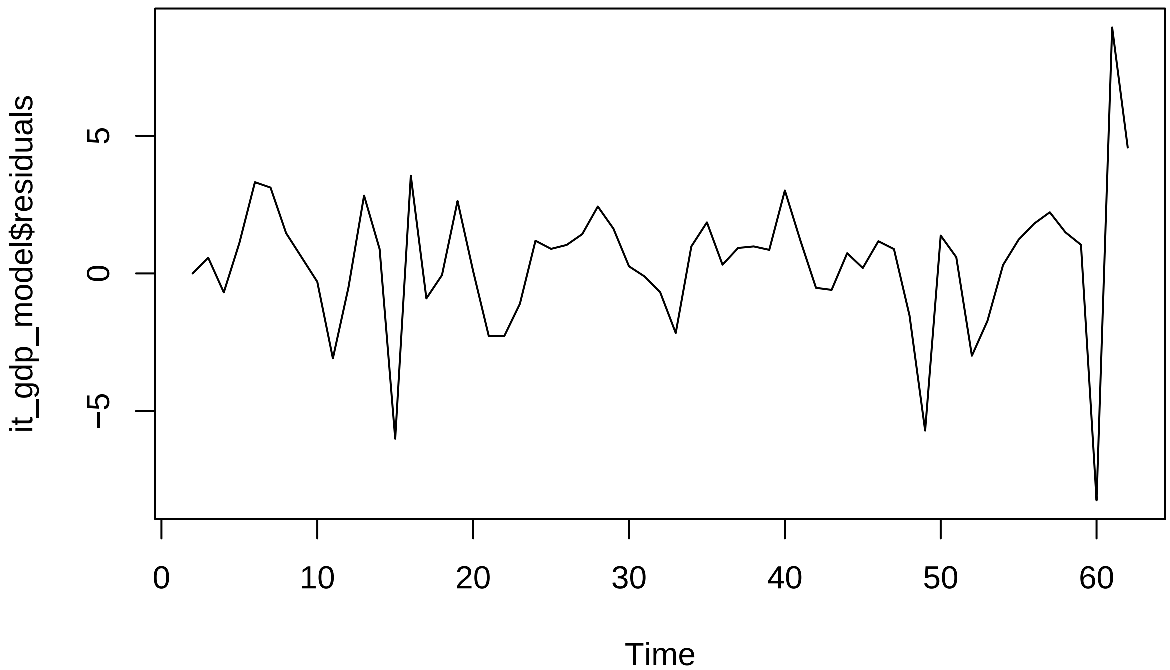
```
## Series: it_gdpDiff
## ARIMA(1,1,2)
##
## Coefficients:
##           ar1      ma1     ma2
##       -0.0241  -1.9902  0.9999
## s.e.   0.1346   0.0859  0.0861
##
## sigma^2 = 6.712:  log likelihood = -146.51
## AIC=301.02   AICc=301.75   BIC=309.4
##
## Training set error measures:
##                     ME     RMSE      MAE       MPE     MAPE     MASE
## Training set 0.3963182 2.504411 1.756155 -6.003682 245.1163 0.505611
##                     ACF1
## Training set -0.03897398
```

Once we find the p and q values, we try them for the ARIMA models to find the best AICc value Using the ARIMA(1,1,2) model gives the best AICc value.

```r
#Check the fitted and residuals
plot(it_gdp_model$fitted)
```




```r
plot(it_gdp_model$residuals)
```

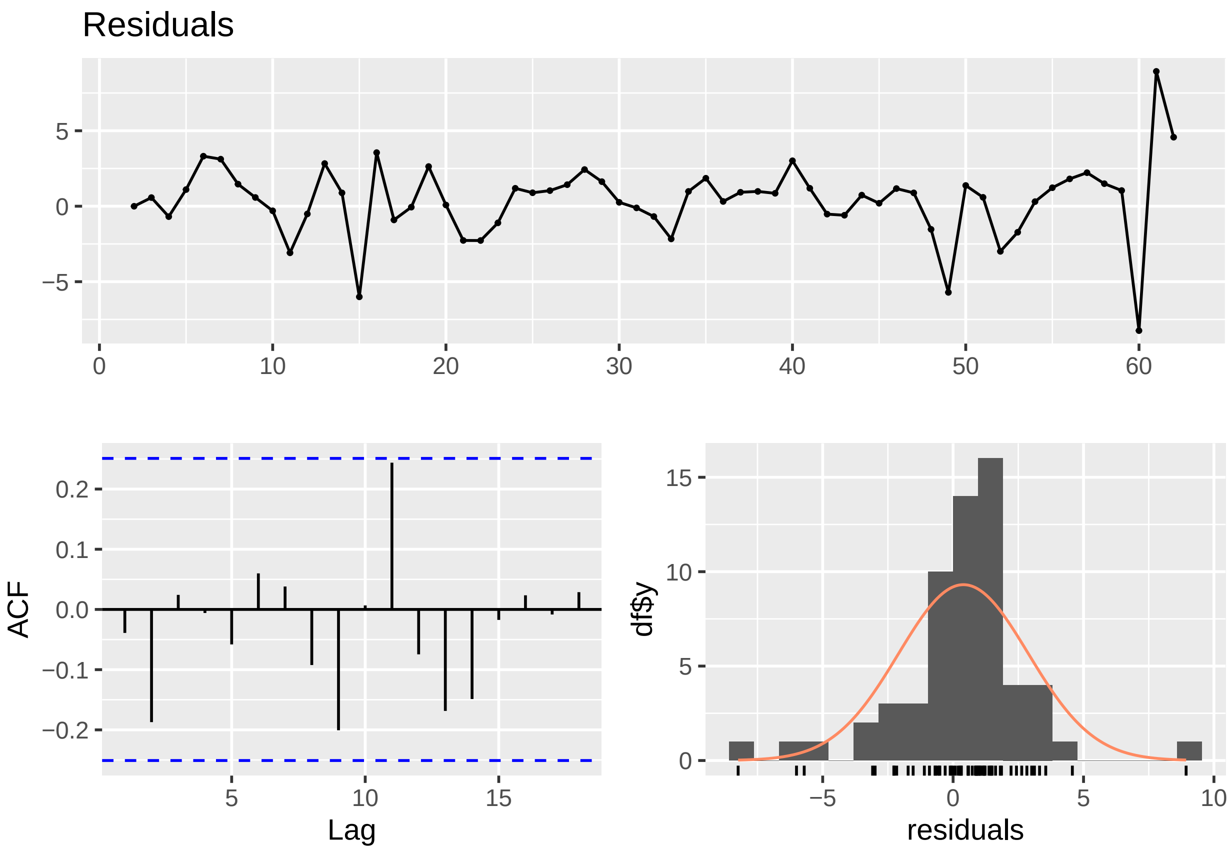



```r
#Check stationary of the residuals
ur.kpss(it_gdp_model$residuals) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0629
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

We then look at the fitted and residual models and then conduct another kpss test for our model with the residuals which looks good

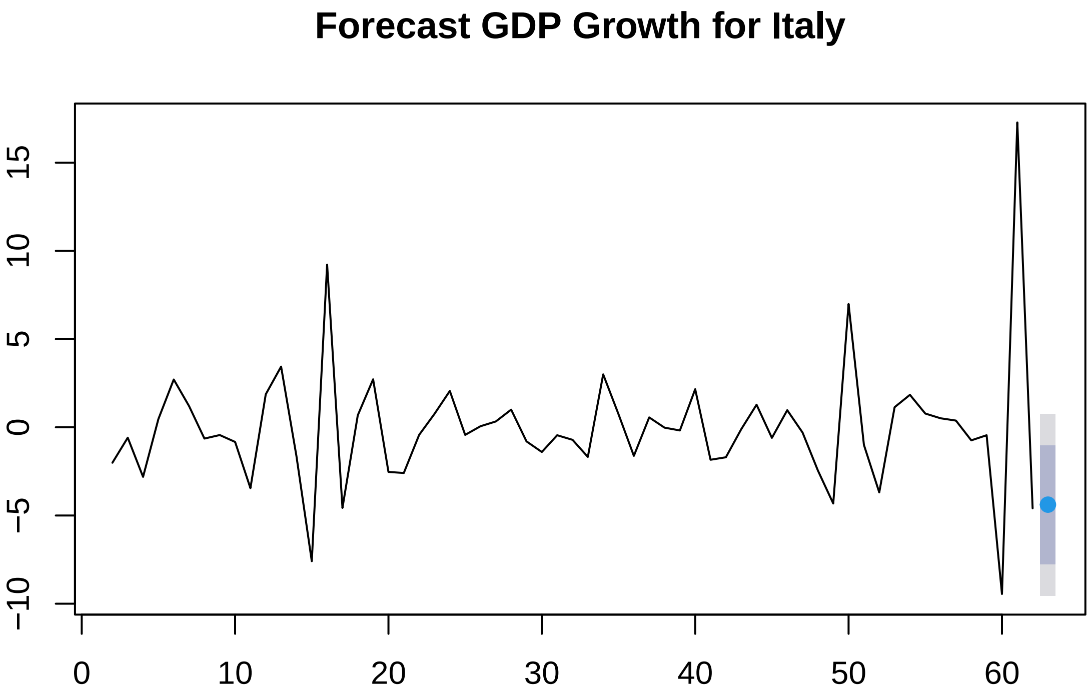
```r
checkresiduals(it_gdp_model$residuals) #Might have white noise
```




```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 6.6026, df = 10, p-value = 0.7624
##
## Model df: 0.   Total lags used: 10
```

We then check the residuals to see if we have white noise. Looking at the data it looks like me might have a little bit of white noise but it shouldn't affect the data too much

```
it.forecast_values <- forecast(it_gdp_model, h=1)
plot(it.forecast_values, main = "Forecast GDP Growth for Italy")
```



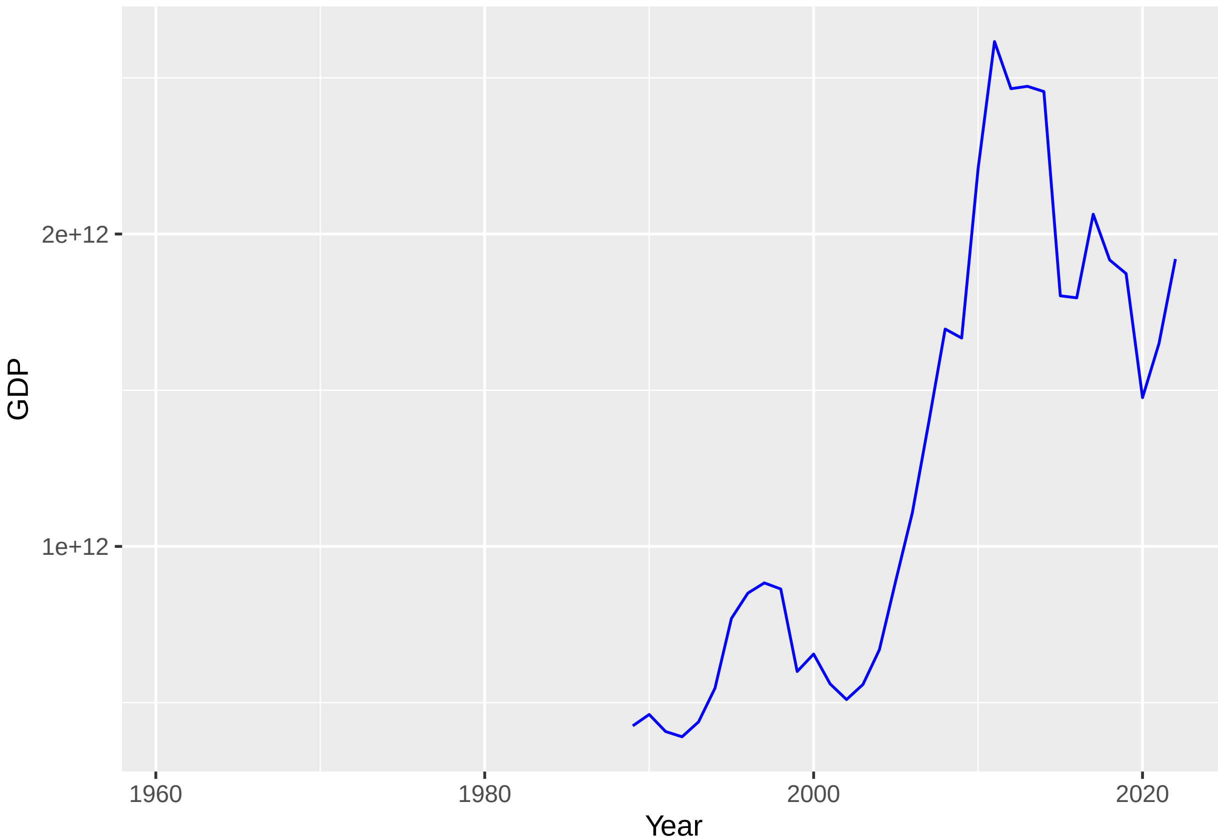
Then we go ahead and forecast the data which shows us that the gdp for Italy might me lower than the average from before covid

```
brazil_gdp <- gdp %>%
  filter(Country == "Brazil") %>%
  select(Year, GDP, GDP_Growth)
```

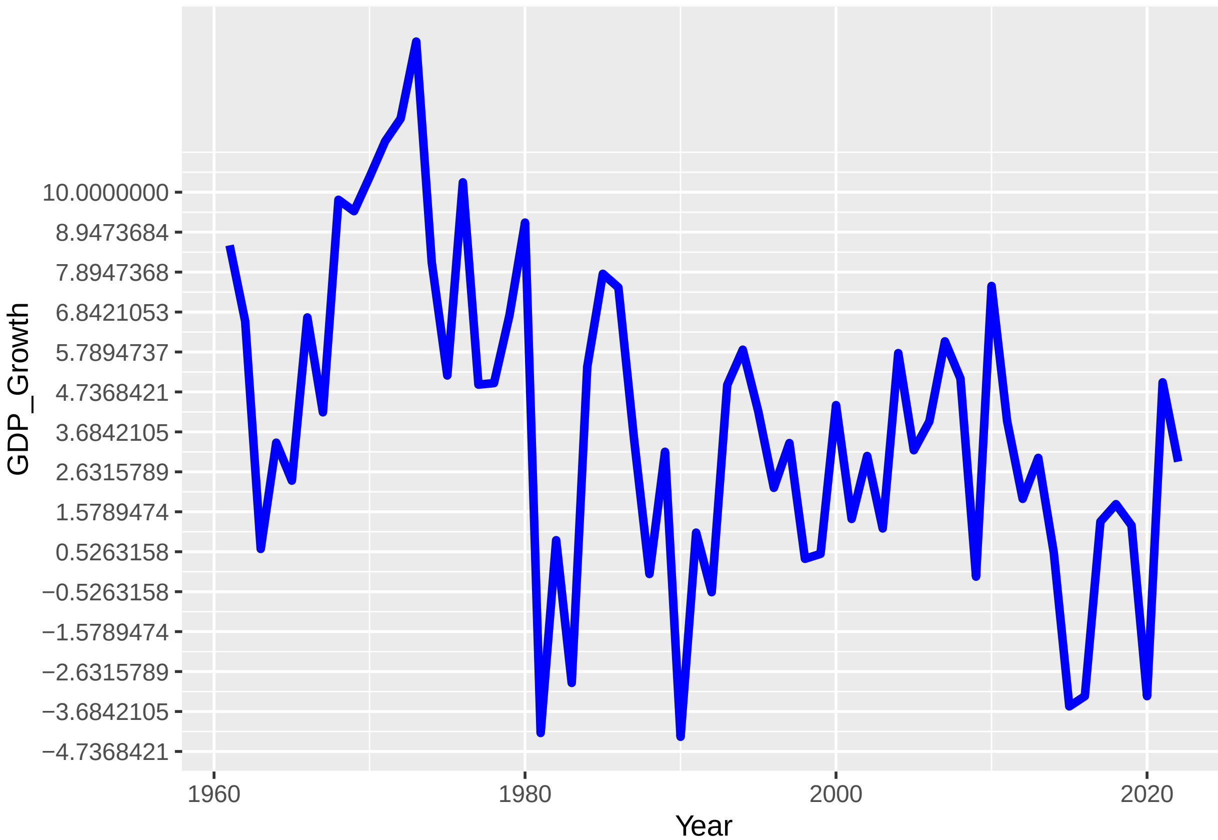
Brazil up next

```
ggplot(brazil_gdp) +
  geom_line(aes(Year, GDP, group = 1), color = "blue", linewidth = .5)
```

```
## Warning: Removed 28 rows containing missing values (`geom_line()`).
```




```r
ggplot(brazil_gdp, aes(Year, GDP_Growth, group = 1)) +
  geom_line(color = "blue", linewidth = 1.5) +
  scale_y_continuous(
    breaks = seq(-10, 10, by = 20/19) )
```

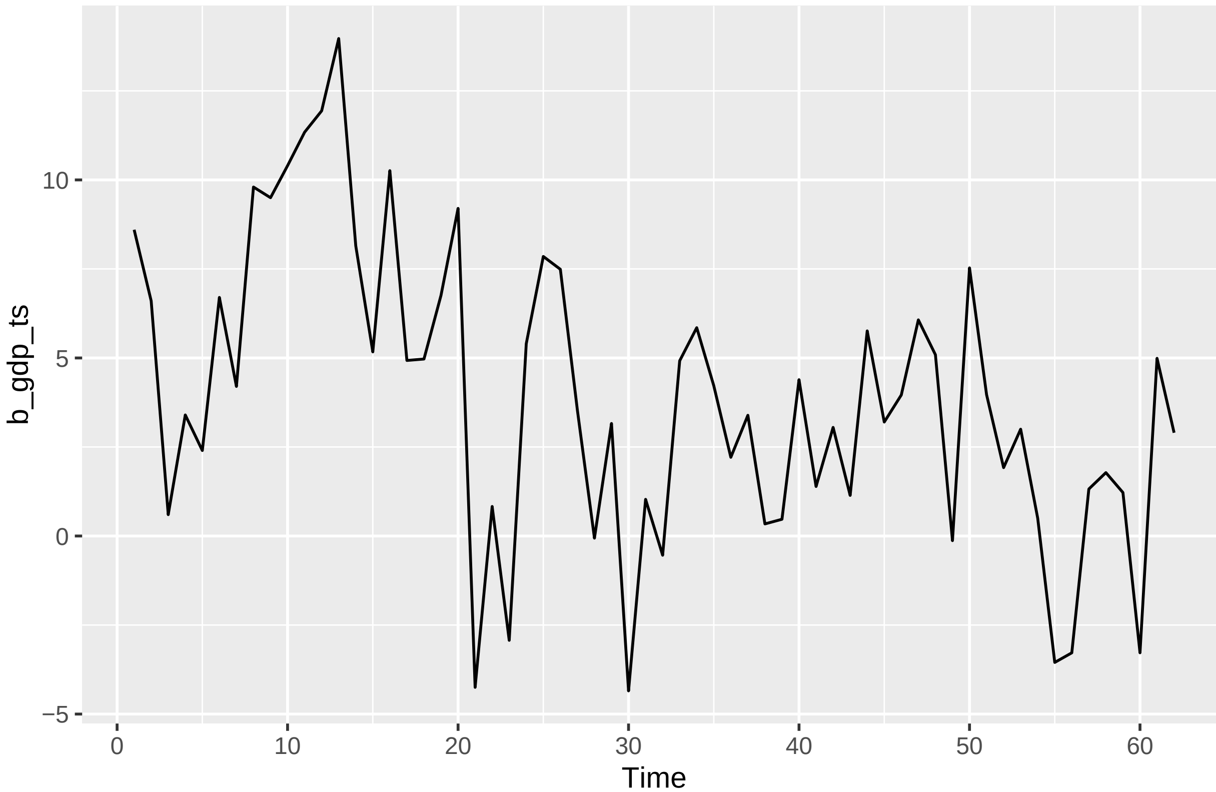


These plots show us that there is a lot of negative spikes thorughout the 1980s which can becasue of political unrest and corruption

```
btrain <- brazil_gdp[1:50,]
btest <- brazil_gdp[51:62,]
bntest <- nrow(btest)
```

Train and test sets

```
b_gdp_ts <- ts(brazil_gdp$GDP_Growth)
autoplot(b_gdp_ts)
```

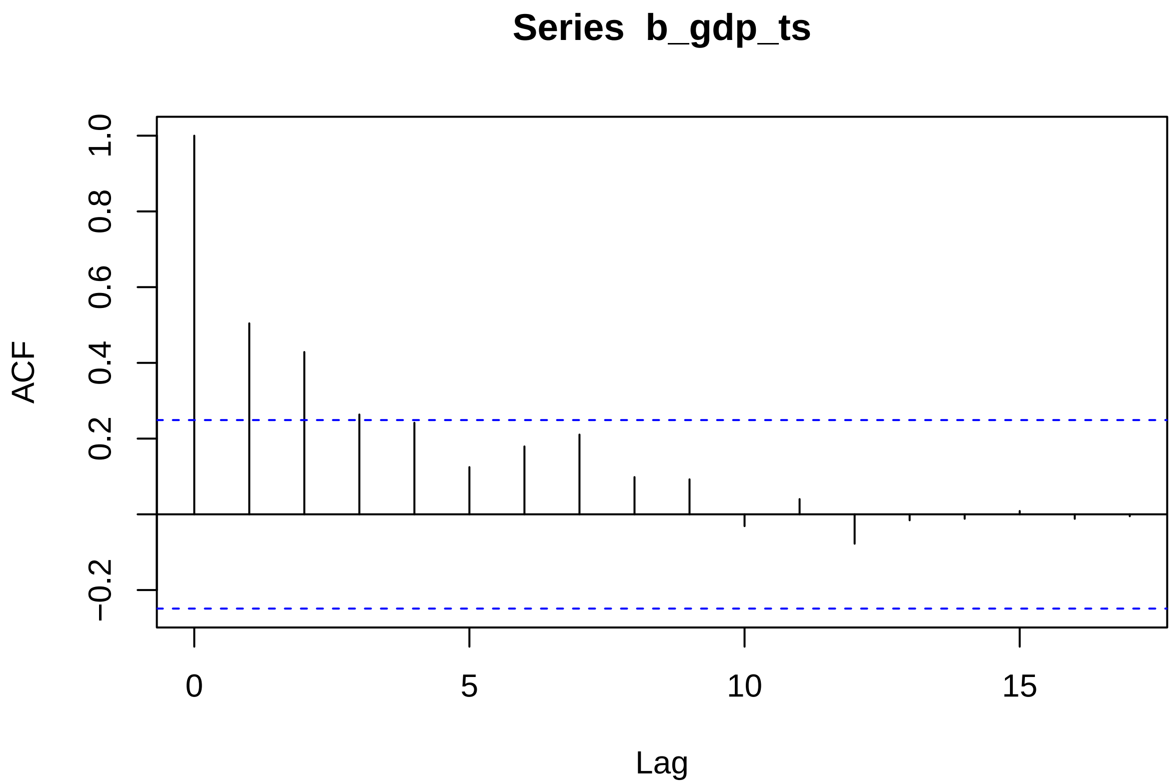


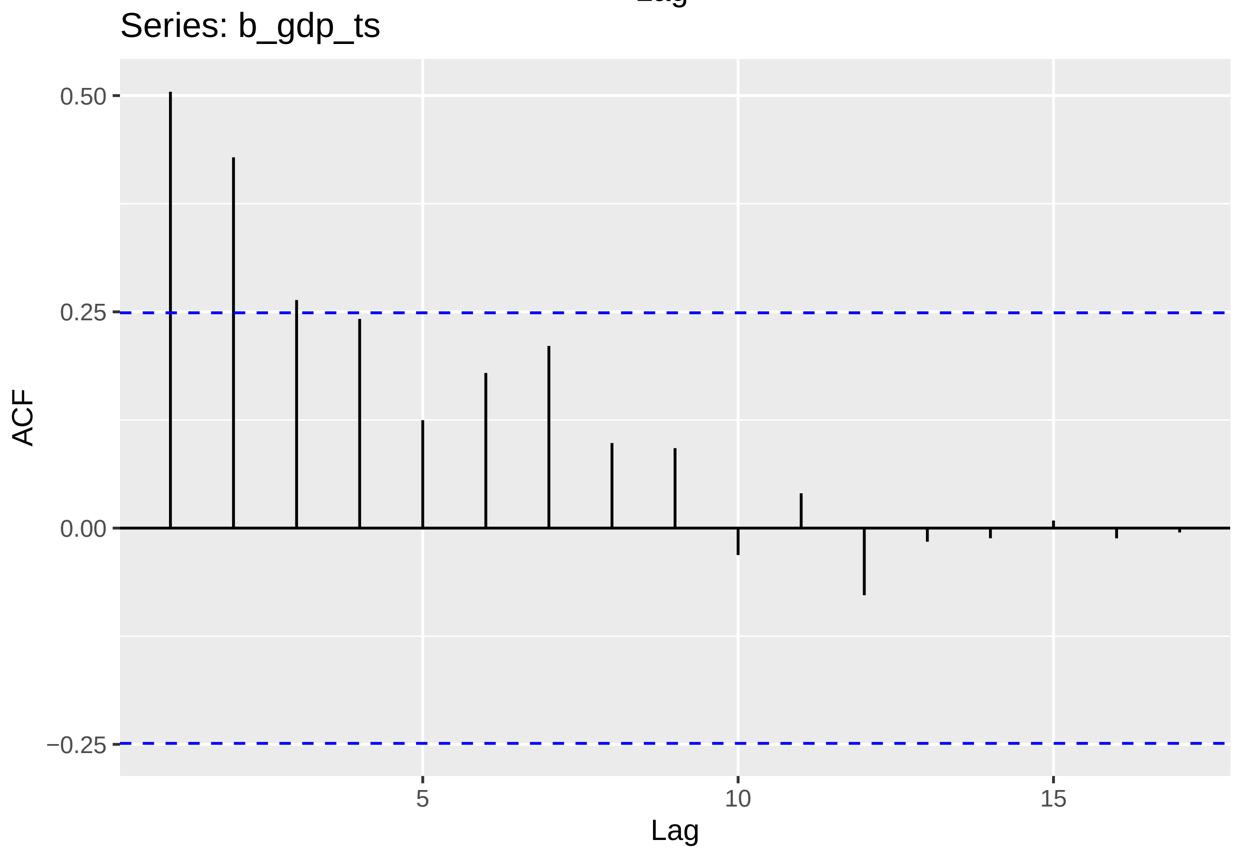

```r
ur.kpss(b_gdp_ts) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.6821
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Create a time series set and then check the stationary of the data

```r
autoplot(acf(b_gdp_ts))
```

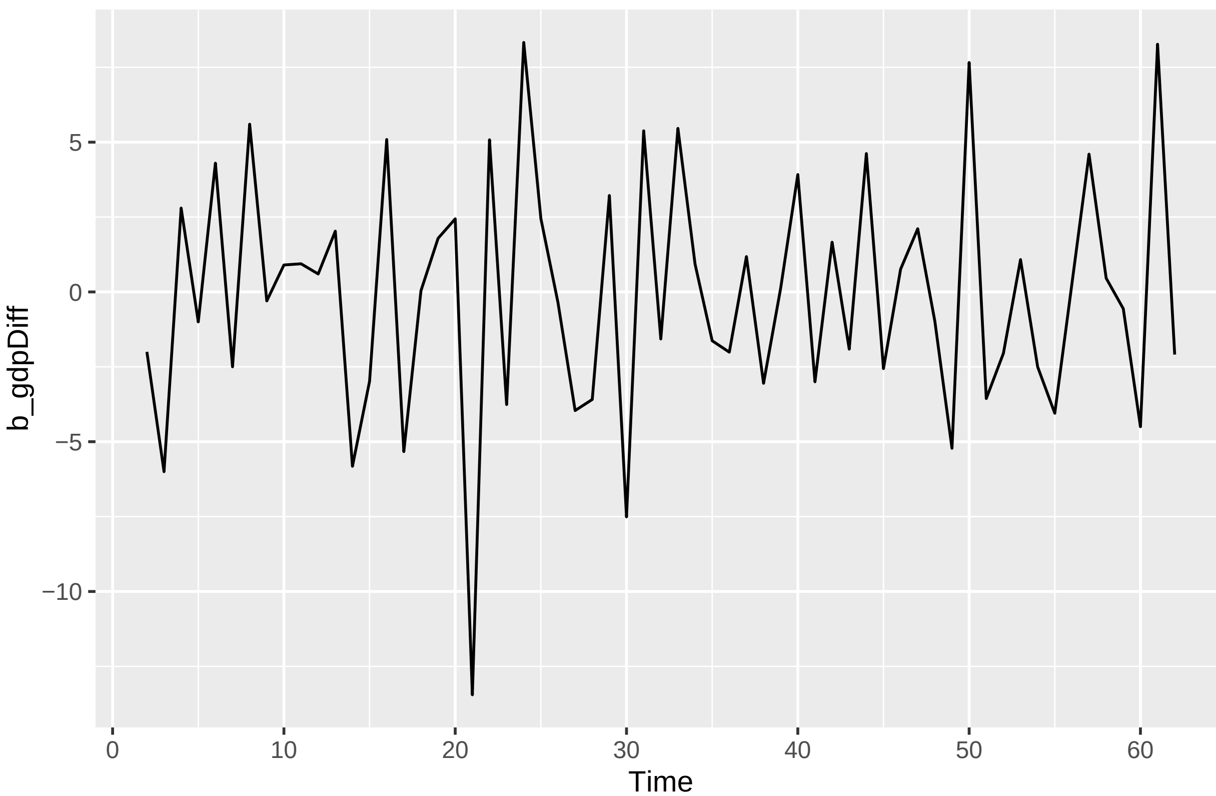





```r
b_gdpDiff = diff(b_gdp_ts, lag = 1)

ur.kpss(b_gdpDiff) %>% summary()
```

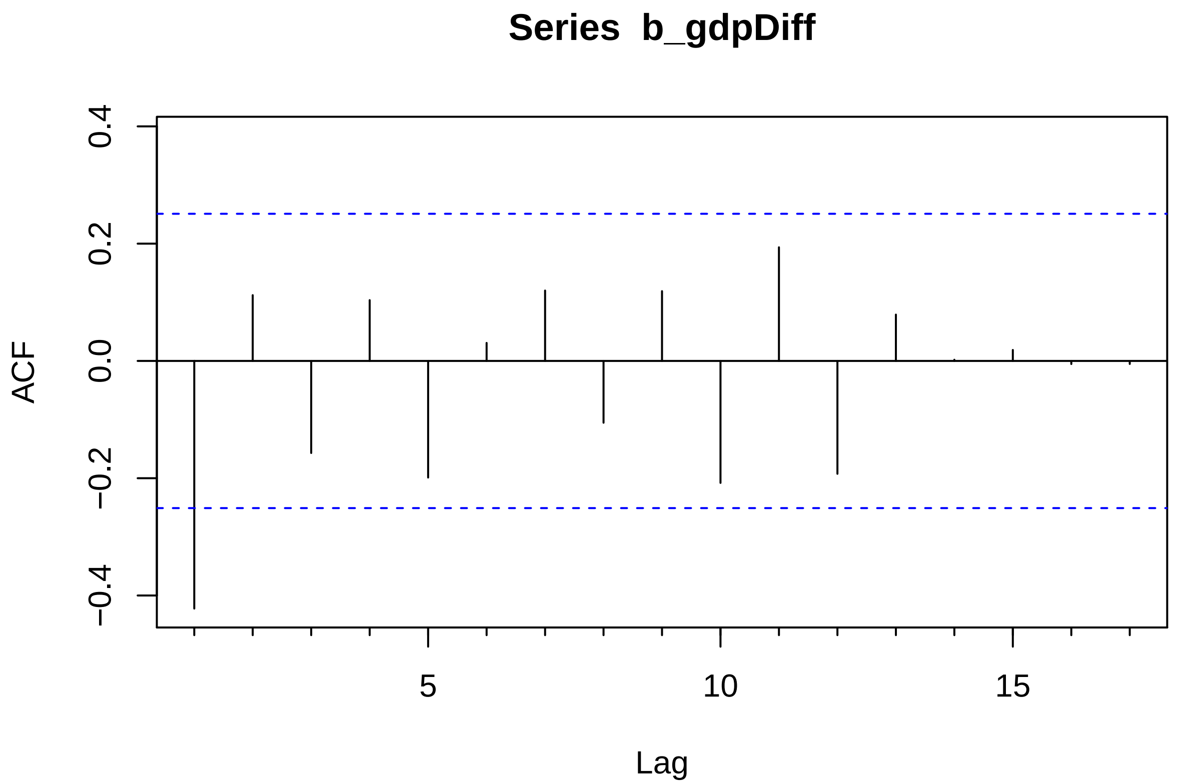
```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0418
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

```r
autoplot(b_gdpDiff)
```

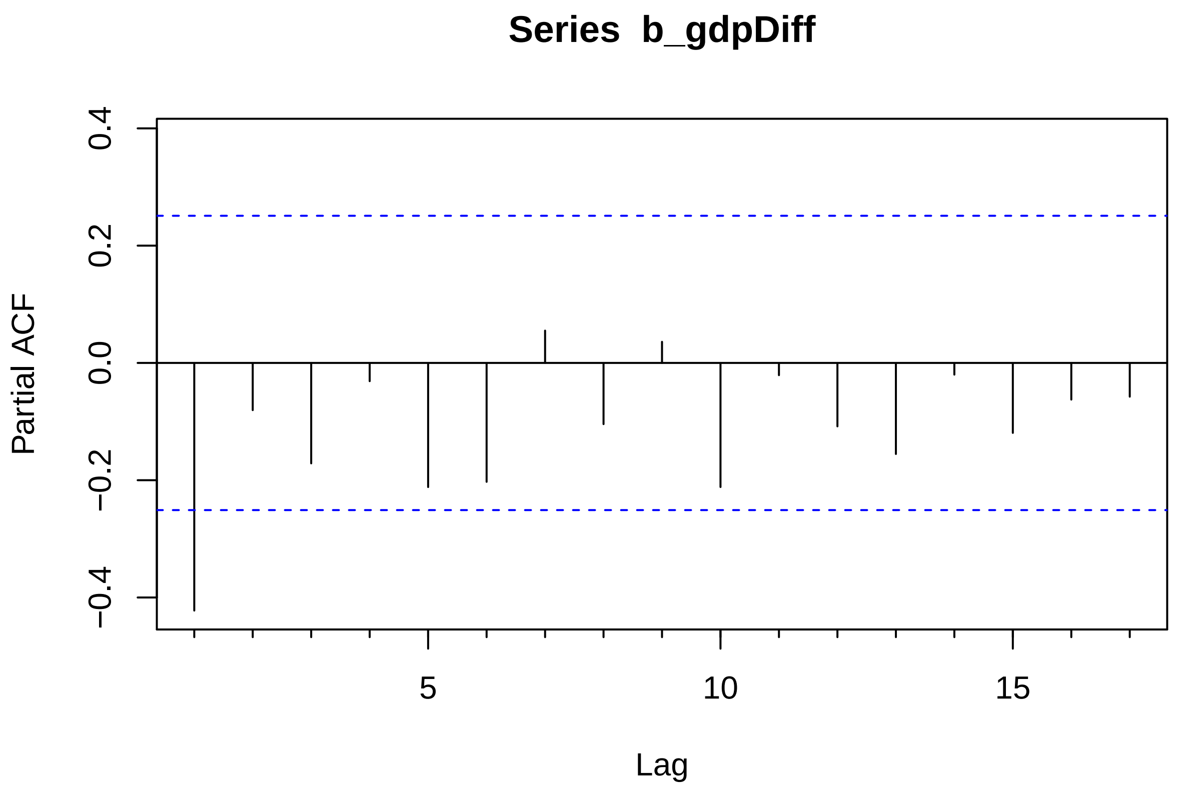


We go ahead and difference the data once and then check the test statistic again Seeing that the test statistic is way lower, this data is stationary.

```r
Acf(b_gdpDiff) #1
```




```r
Pacf(b_gdpDiff) #1
```



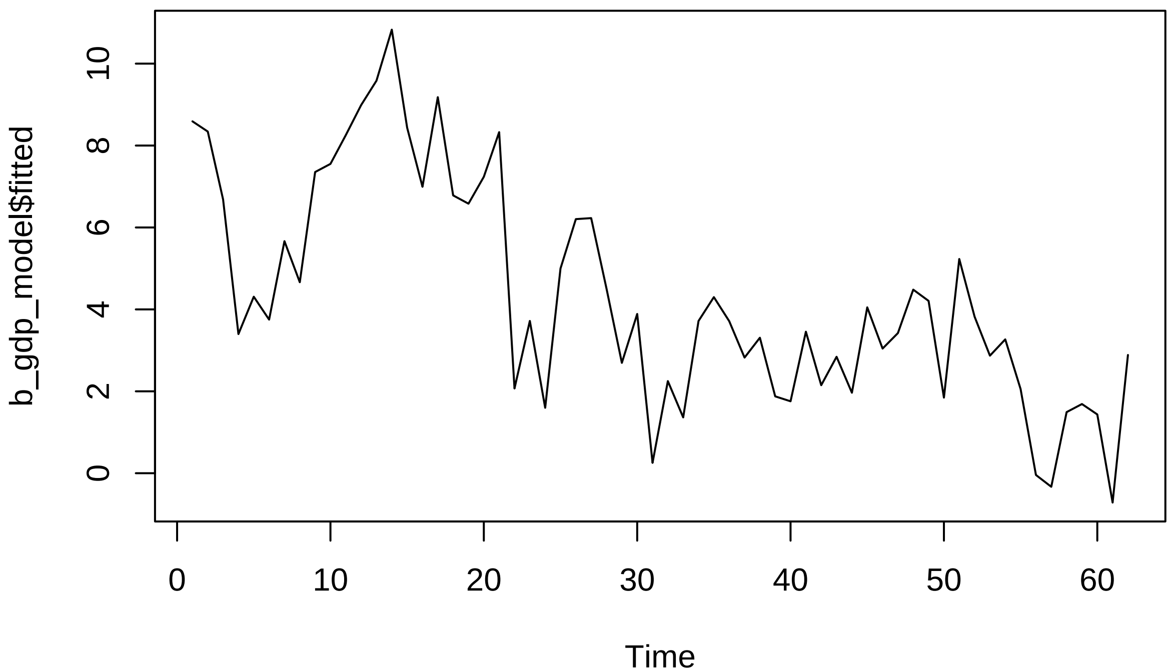

```r
#d=1

#(11,1,6) has been the best so far (AICc = 337.27)
```

```r
b_gdp_model <- Arima(b_gdp_ts, order = c(1, 1, 1), method = "ML")
summary(b_gdp_model)
```

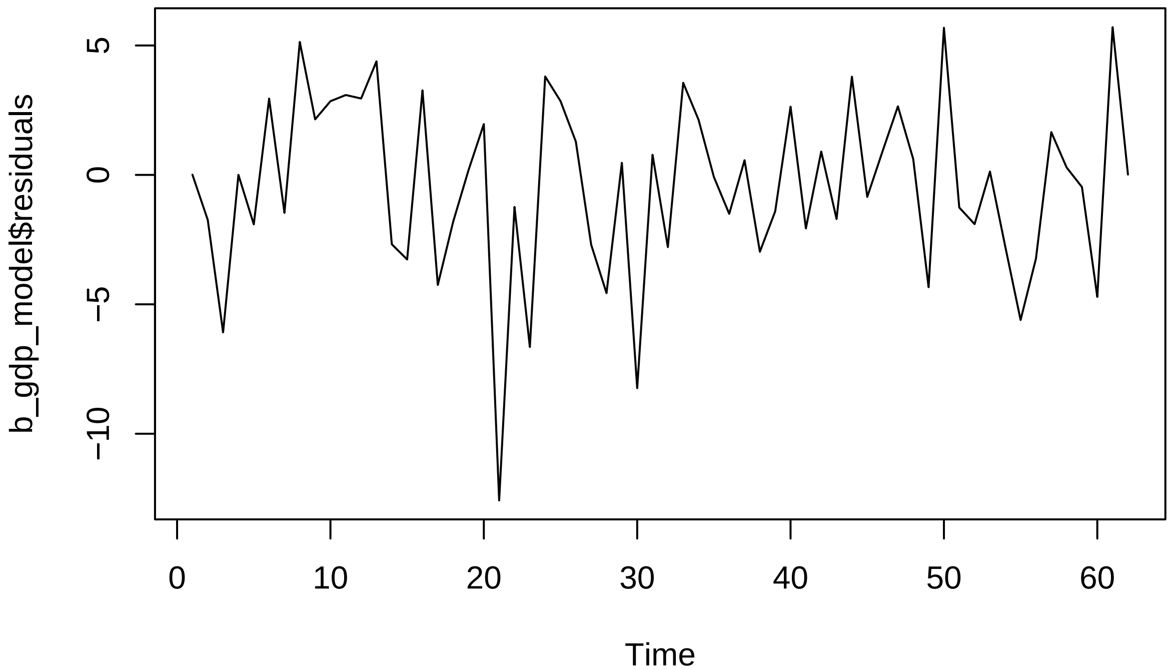
```
## Series: b_gdp_ts
## ARIMA(1,1,1)
##
## Coefficients:
##          ar1      ma1
##       0.3539  -0.8814
## s.e.  0.1684   0.0983
##
## sigma^2 = 12.72:  log likelihood = -163.54
## AIC=333.08   AICc=333.5   BIC=339.41
##
## Training set error measures:
##                      ME     RMSE      MAE     MPE     MAPE      MASE
## Training set -0.4434965 3.478597 2.680306 159.402 284.8521 0.8432112
##                     ACF1
## Training set -0.07695265
```

Then we can preform the ARIMA model to find the best AICc value which came from ARIMA(11,1,6)

```r
plot(b_gdp_model$fitted)
```




```r
plot(b_gdp_model$residuals)
```

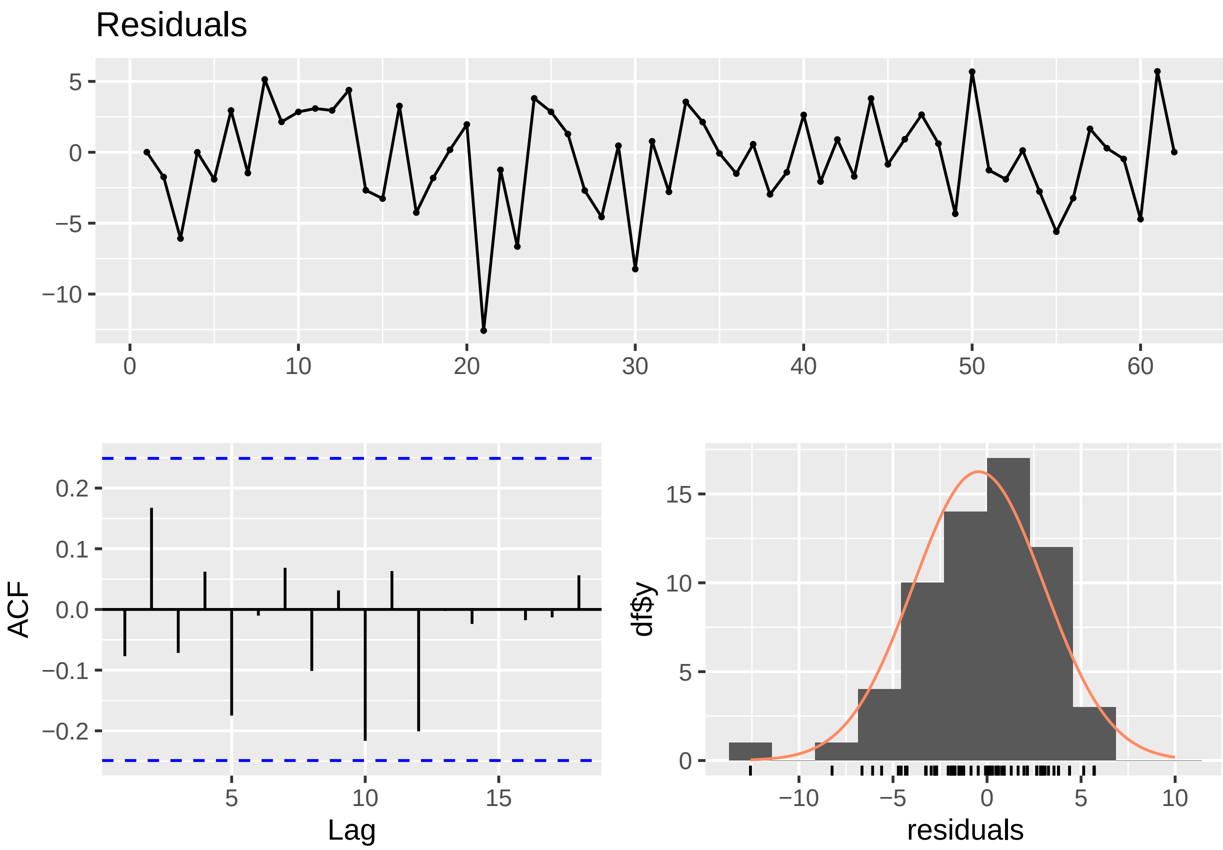



```
#Check stationary of the residuals
ur.kpss(b_gdp_model$residuals) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0636
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

We then look at the fitted and residual models and then conduct another kpss test for our model with the residuals which looks good

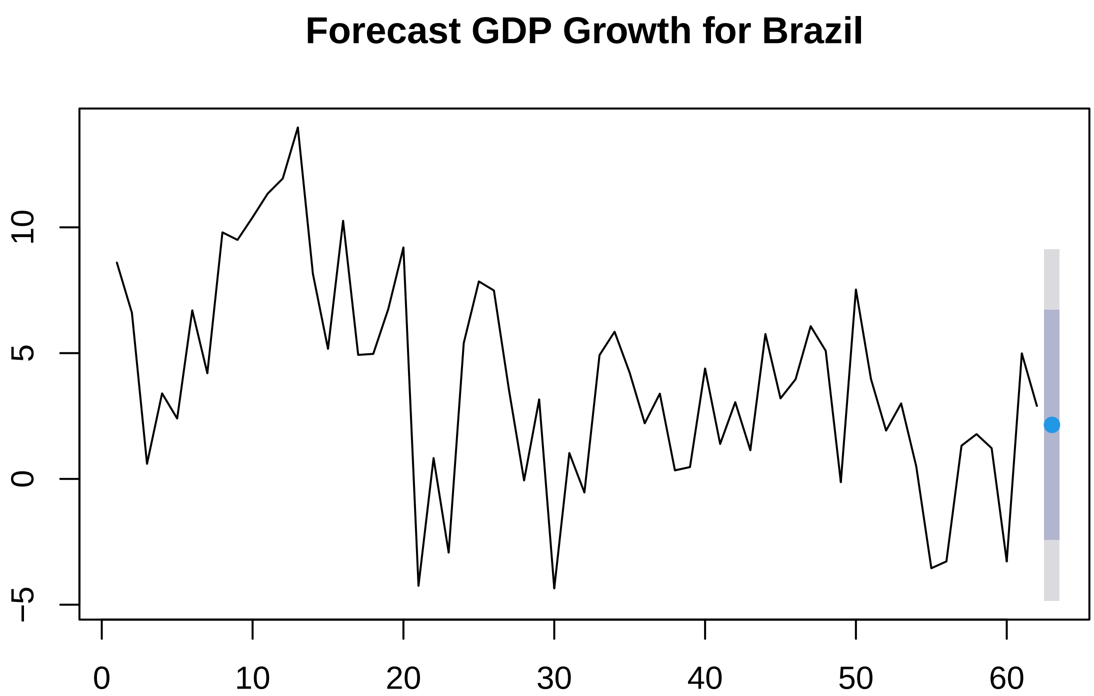
```
checkresiduals(b_gdp_model$residuals)
```




```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 9.7301, df = 10, p-value = 0.4645
##
## Model df: 0.   Total lags used: 10
```

Looking at the residuals, it does not look like there is any white noise in the data that can fluctuate the data and predictions

```
b.forecast_values <- forecast(b_gdp_model, h=1)
plot(b.forecast_values, main = "Forecast GDP Growth for Brazil")
```

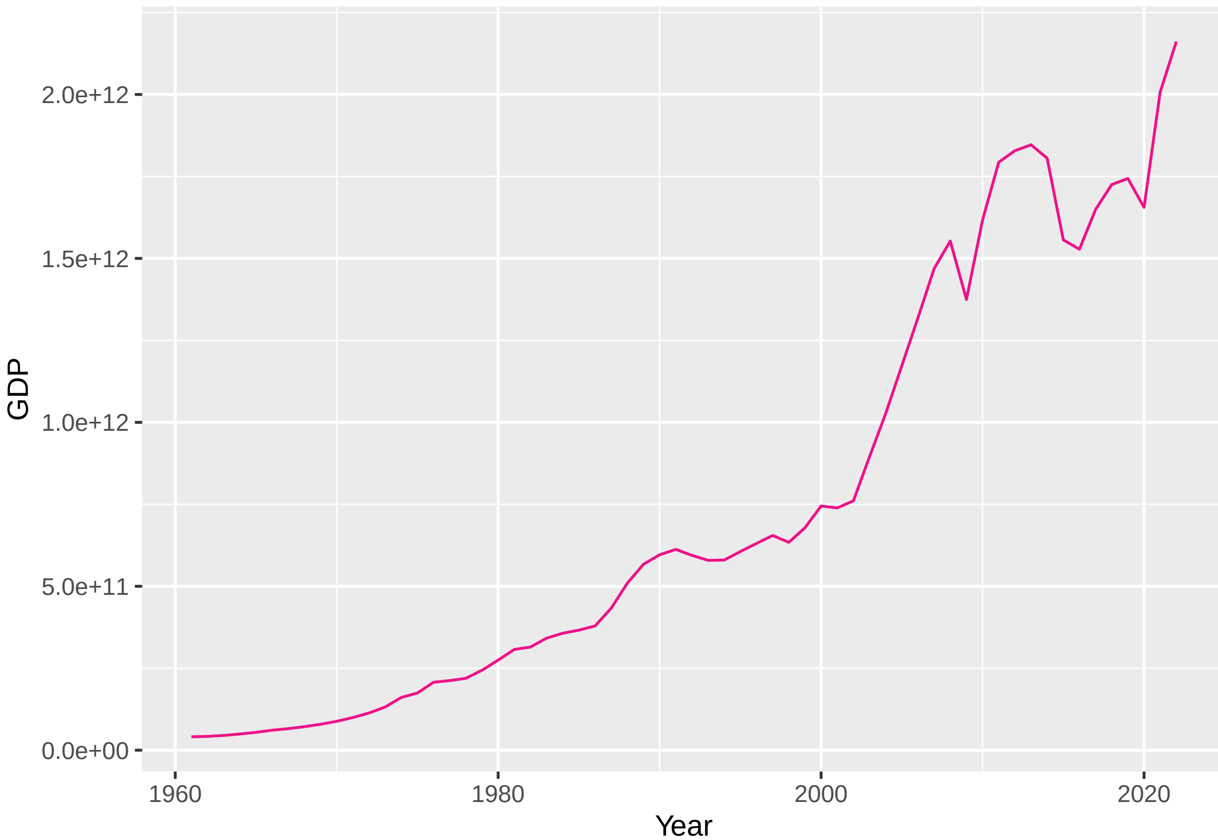


Then we can predict the forecast. Looking at the data shows that the gdp is on the rise but looks like it will fall a bit in the future

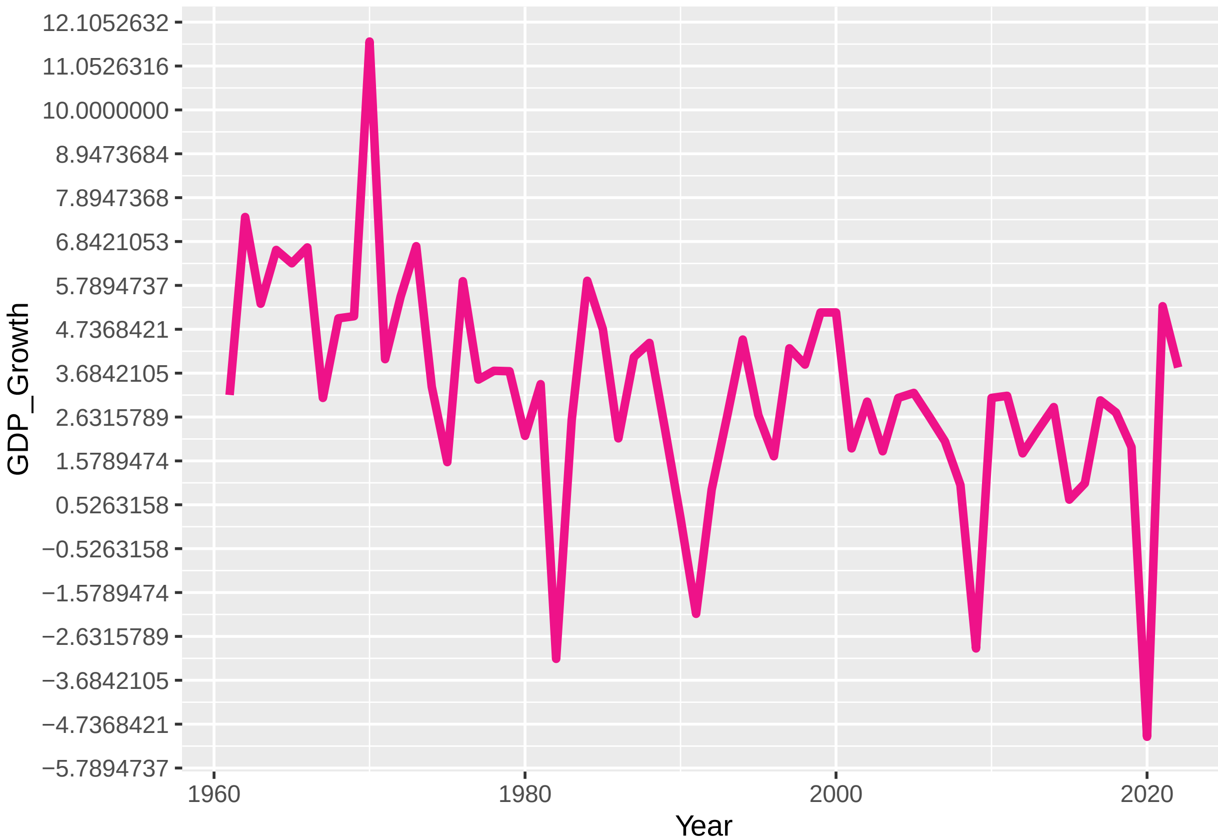
```
canada_gdp <- gdp %>%
  filter(Country == "Canada") %>%
  select(Year, GDP, GDP_Growth)
```

Our last country is Canada

```
ggplot(canada_gdp) +
  geom_line(aes(Year, GDP, group = 1), color = "deeppink2", linewidth = .5)
```




```r
ggplot(canada_gdp, aes(Year, GDP_Growth, group = 1)) +
  geom_line(color = "deeppink2", linewidth = 1.5) +
  scale_y_continuous(
    breaks = seq(-10, 15, by = 20/19))
```

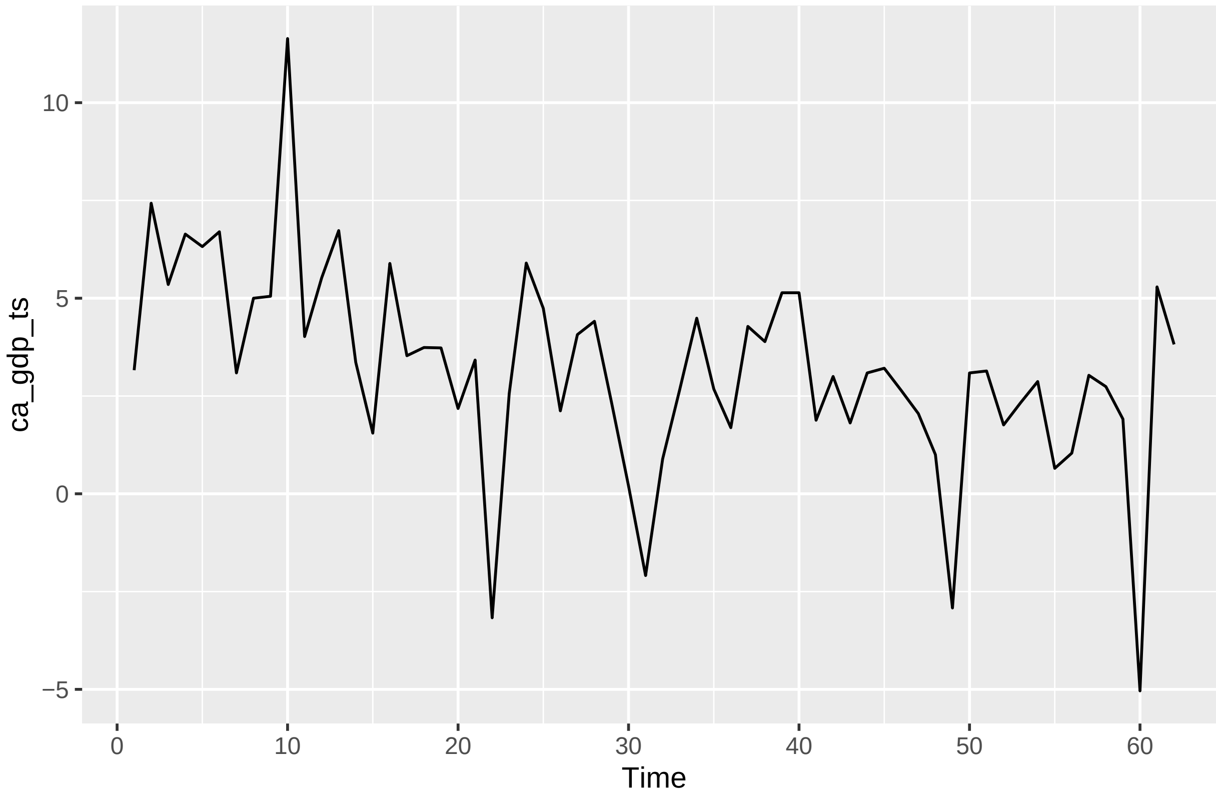


Canada looks like they have a very similar gdp growth with the United States, even showing falls in the early 1980s, 2008, and 2020

```
catrain <- canada_gdp[1:50,]
catest <- canada_gdp[51:62,]
cantest <- nrow(catest)
```

Train and testing sets

```
ca_gdp_ts <- ts(canada_gdp$GDP_Growth)
autoplot(ca_gdp_ts)
```

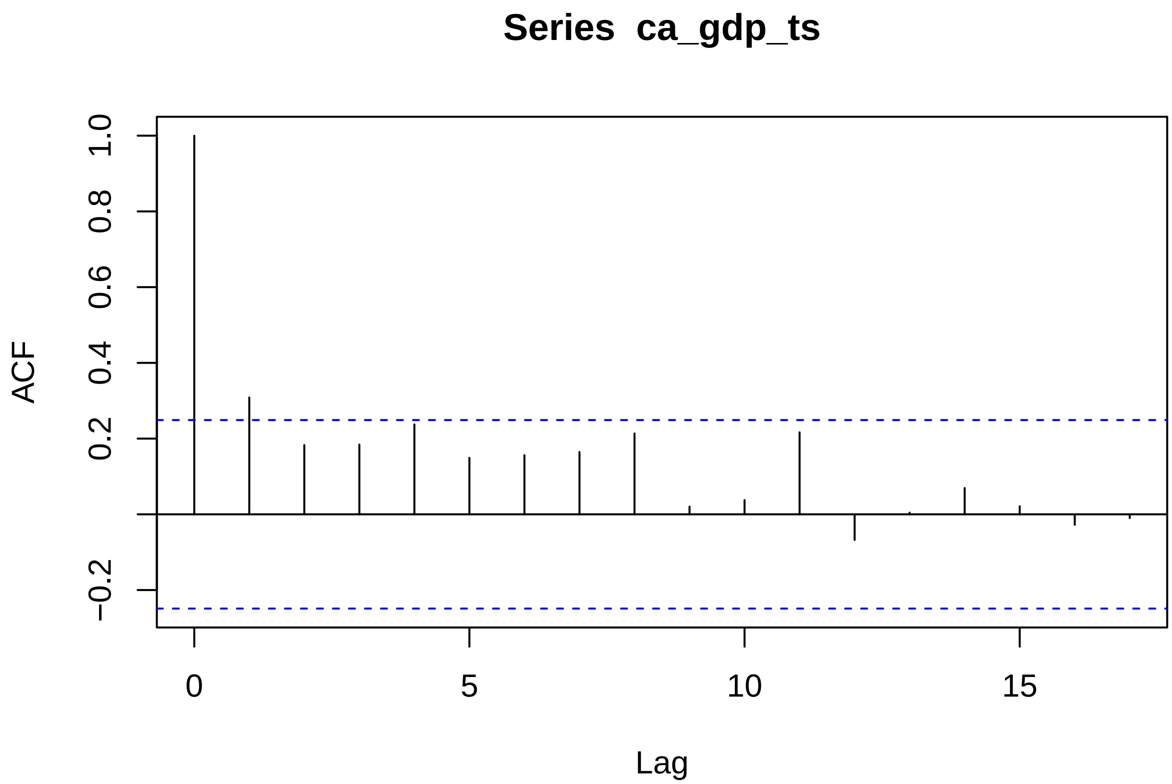


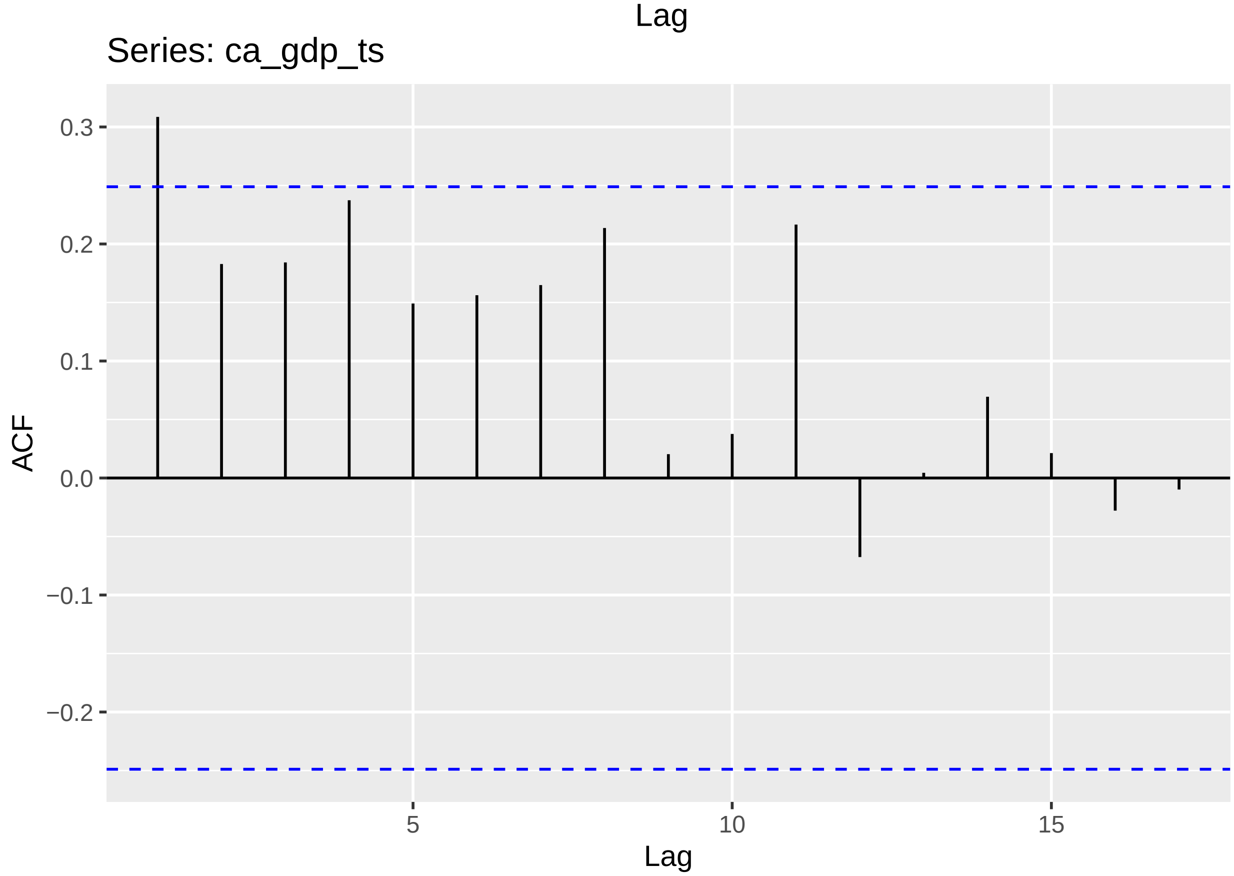

```r
ur.kpss(ca_gdp_ts) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.856
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Then we make a time series model and then used the kpss test to see if the null hypothesis is accepted or not

```r
autoplot(acf(ca_gdp_ts))
```

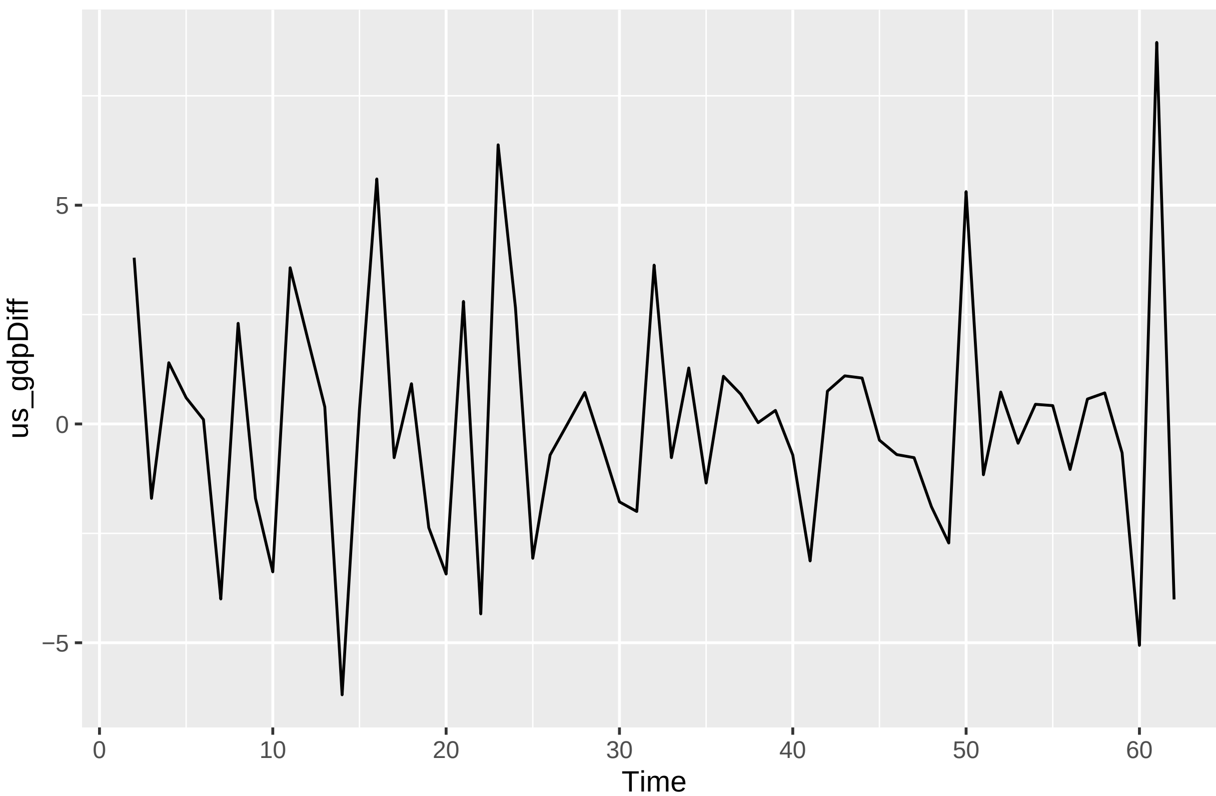





```r
ca_gdpDiff = diff(ca_gdp_ts, lag = 1)

ur.kpss(ca_gdpDiff) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.046
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

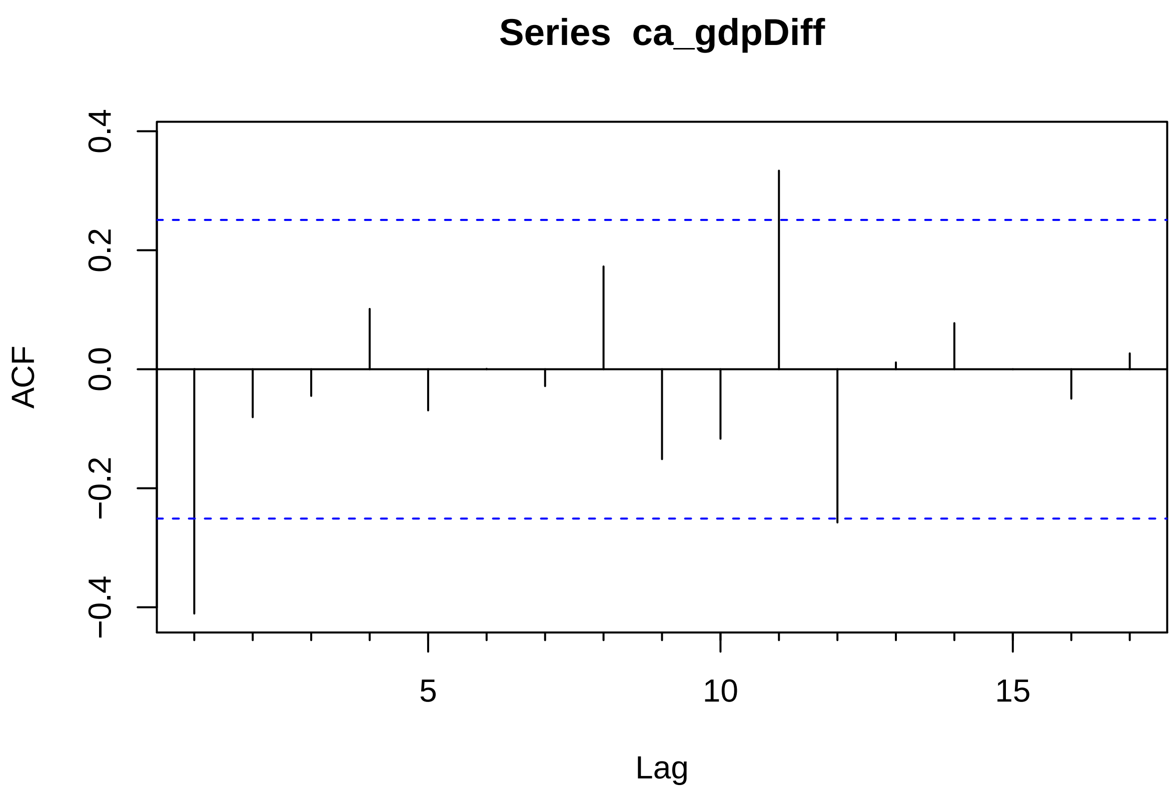
```
autoplot(us_gdpDiff)
```



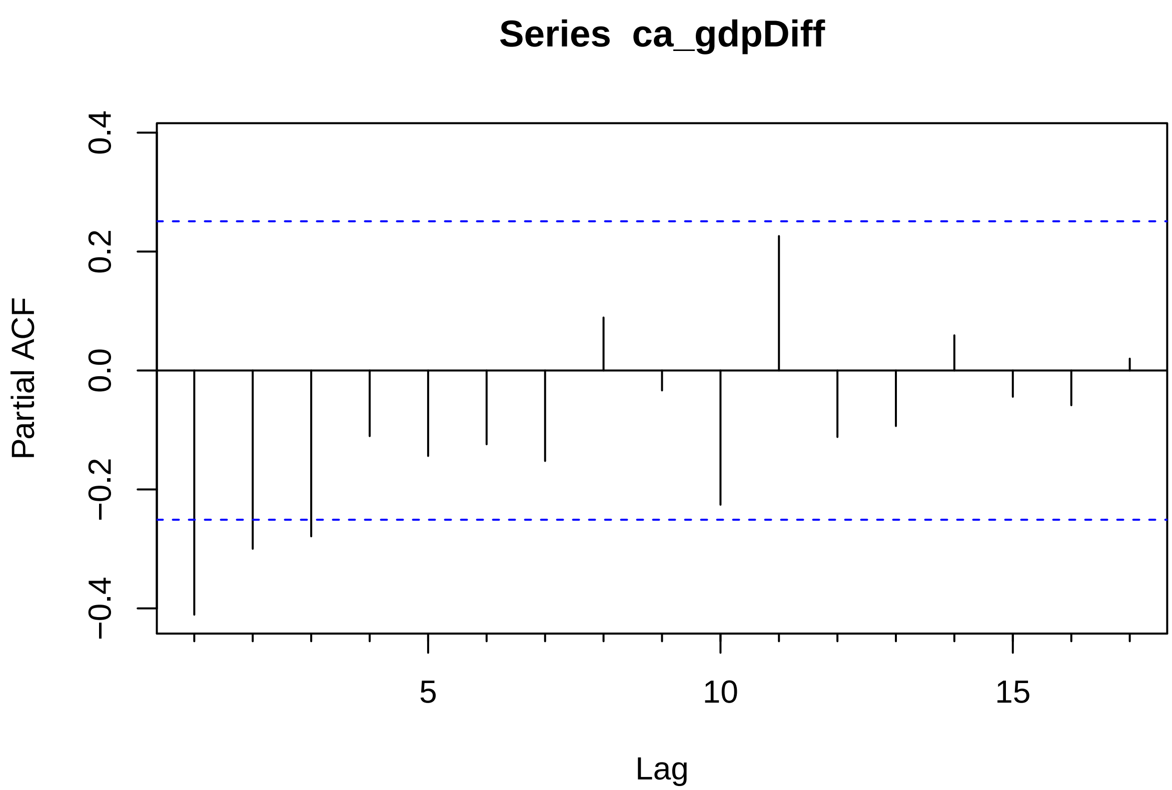
We go ahead and difference the data once and then check the test statistic again Seeing that the test statistic is way lower, this data is stationary.

We then look at the Acf and pacf graphs to see our q and p value for the ARIMA model

```
Acf(ca_gdpDiff) #1,11,12
```




```
Pacf(ca_gdpDiff) #1,2,3
```



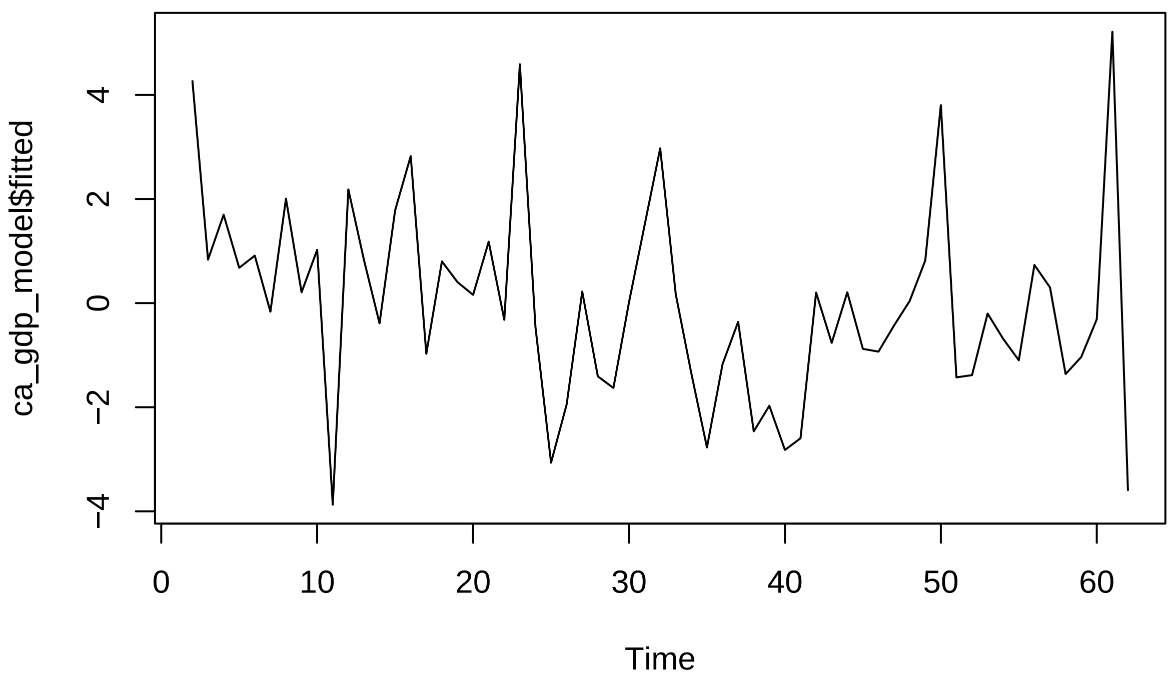

```
#d=1

#(1,1,2) has been the best so far (AICc = 291.81)
```

```r
ca_gdp_model <- Arima(ca_gdpDiff, order = c(1, 1, 2), method = "ML")
summary(ca_gdp_model)
```

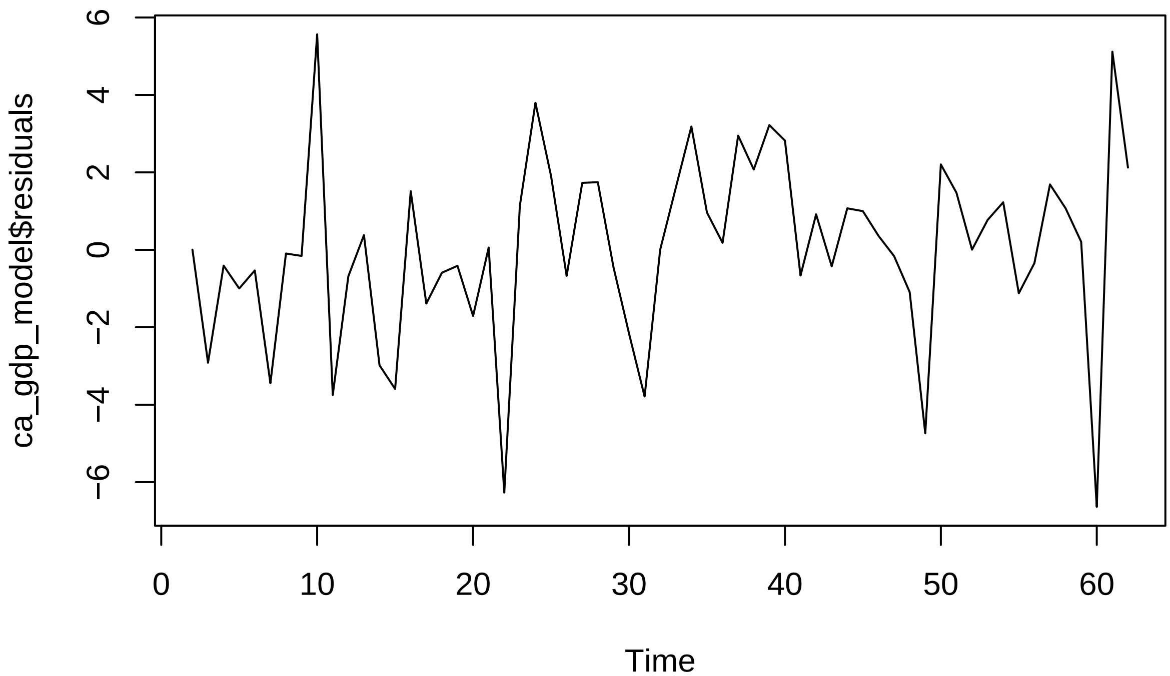
```
## Series: ca_gdpDiff
## ARIMA(1,1,2)
##
## Coefficients:
##          ar1      ma1     ma2
##       0.1047  -1.9002  0.9055
## s.e.  0.1579   0.0912  0.0918
##
## sigma^2 = 6.019:  log likelihood = -141.54
## AIC=291.09   AICc=291.81   BIC=299.46
##
## Training set error measures:
##                      ME    RMSE      MAE MPE MAPE      MASE       ACF1
## Training set 0.03147252 2.37148 1.741973 Inf  Inf 0.4765131 0.01578978
```

Once we find the p and q values, we try them for the ARIMA models to find the best AICc value Using the ARIMA(1,1,2) model gives the best AICc value.

```r
plot(ca_gdp_model$fitted)
```




```r
plot(ca_gdp_model$residuals)
```

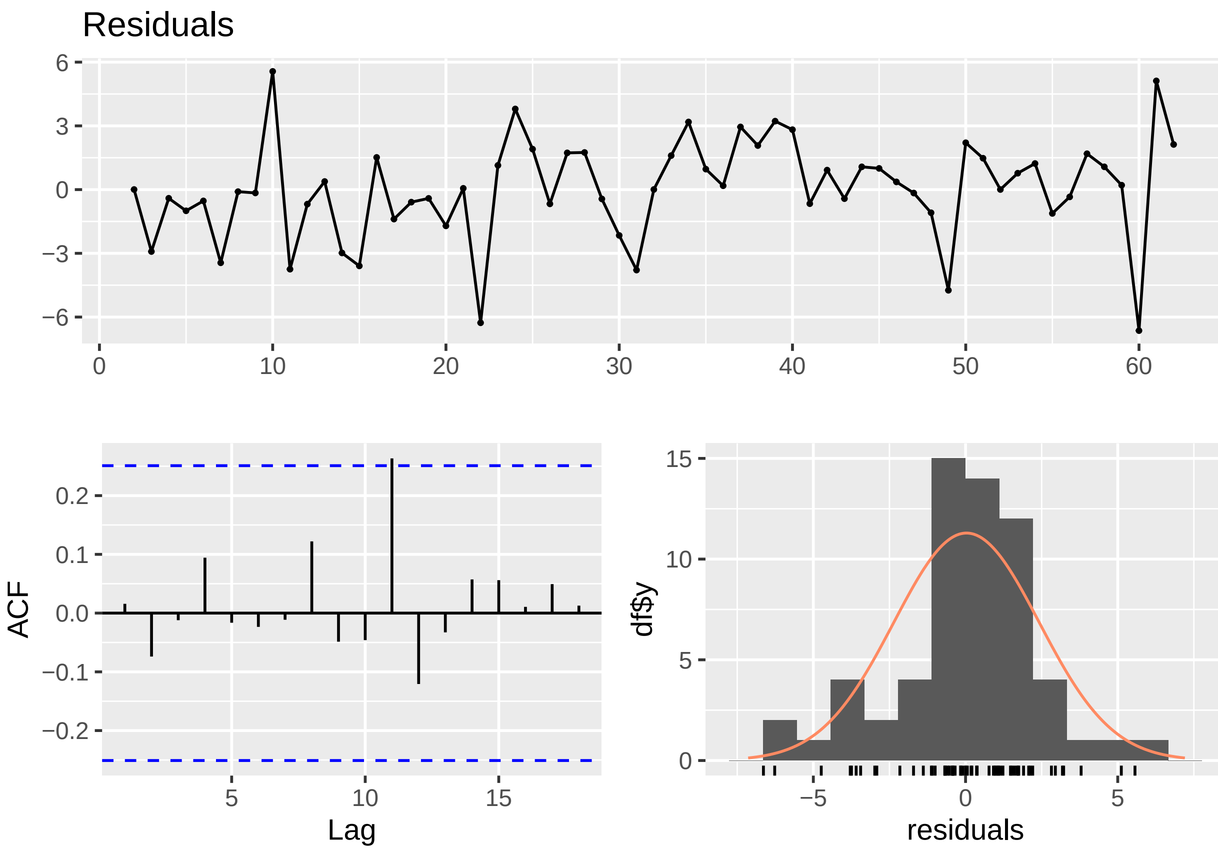



```r
#Check stationary of the residuals
ur.kpss(ca_gdp_model$residuals) %>% summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.3667
##
## Critical value for a significance level of:
##                  10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Then we can look at the fitted and residual models and check the stationary of them

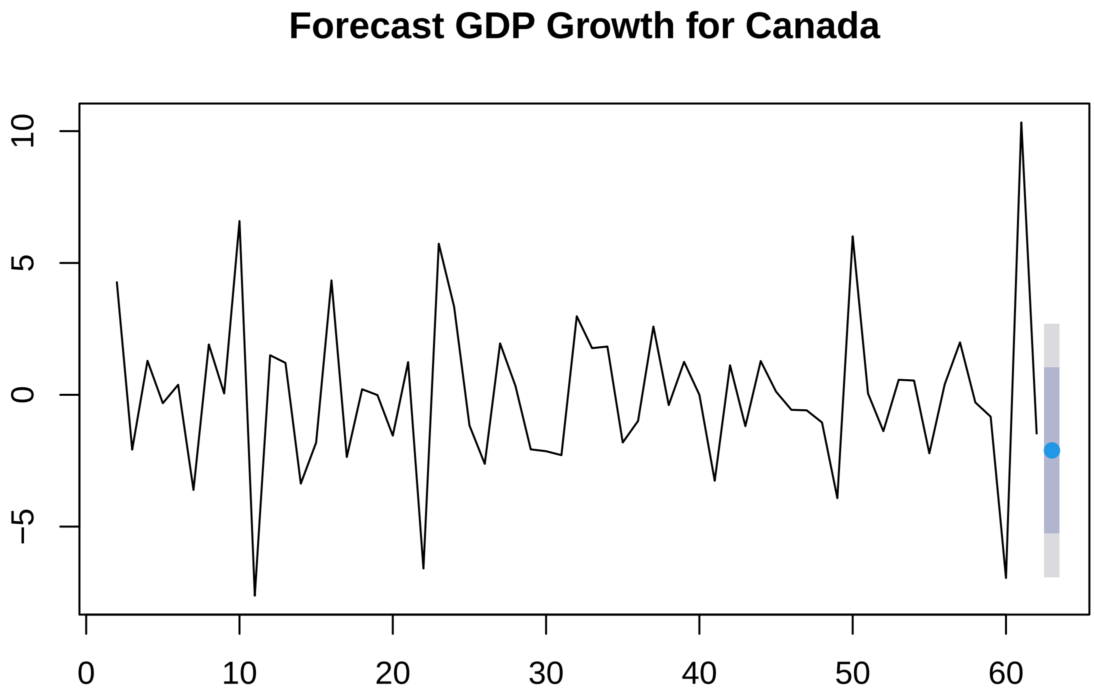
```r
#Check for white noise
checkresiduals(ca_gdp_model$residuals)
```




```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 2.4622, df = 10, p-value = 0.9914
##
## Model df: 0.   Total lags used: 10
```

We can then check the white noise and it looks like there one significant spike on the 11th lag but thats it

```
ca.forecast_values <- forecast(ca_gdp_model, h=1)
plot(ca.forecast_values, main = "Forecast GDP Growth for Canada")
```



We can then forecast the values for the future and looks like Canada might have a higher gdp than they had before the covid pandemic