In [58]:
```
#library(tidyverse)
library("readxl")
install.packages("forecast")
library(forecast)
library(fpp2)
library(ggplot2)
library(tseries)
#list.files(path = "../input")
library(readxl)
```

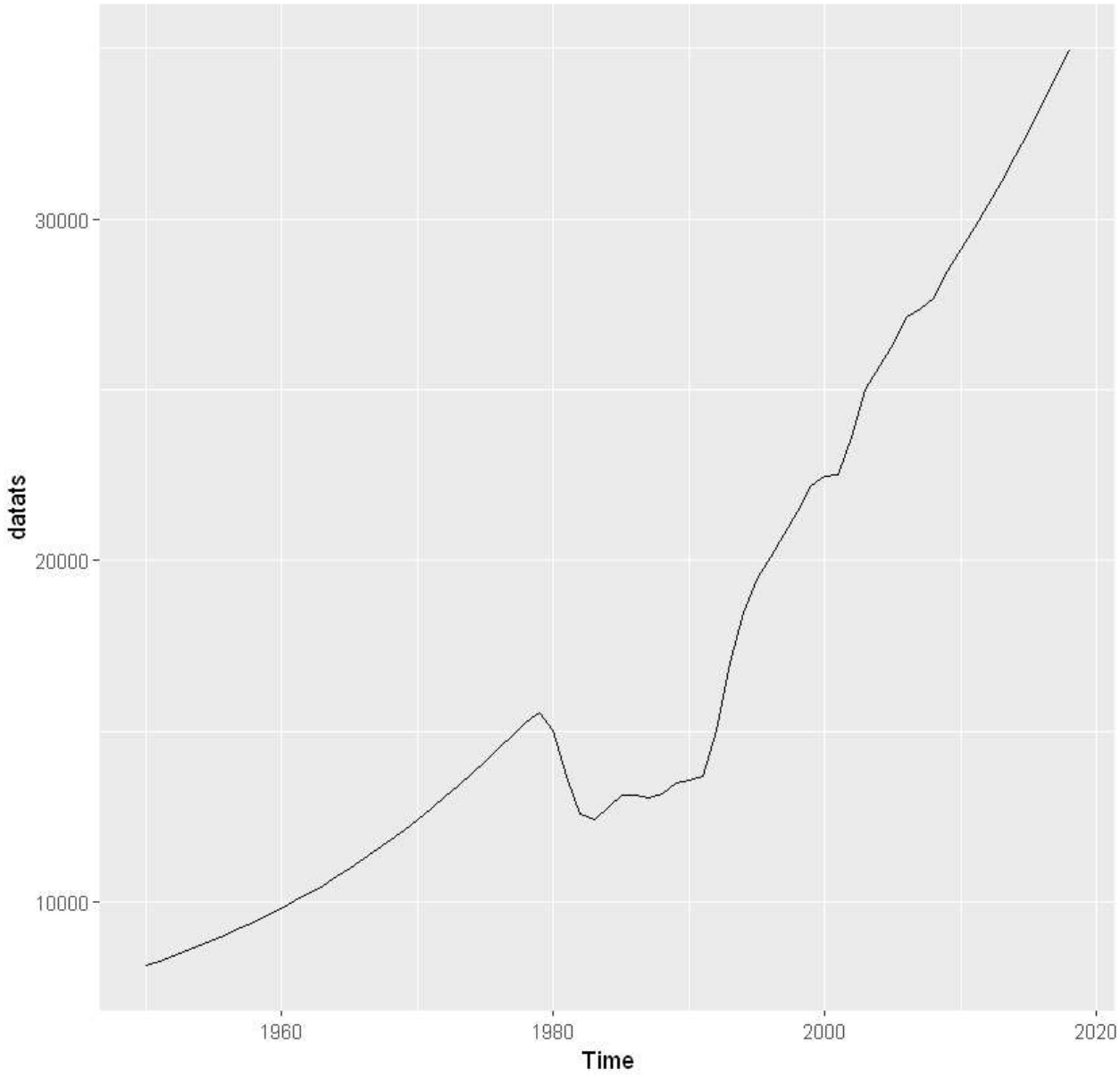
This dataset is of population of India from 1950 to 2018.It is a non seasonal yearly data.

In [98]:
```
#readig data from excel file
Data_popl <- read_excel("populationIndia.xlsx")
```

In [99]:
```
#changing the name of column
colnames(Data_popl) <- c('Year','Population')
head(Data_popl)
```

Year	Population
1950	8150
1951	8284
1952	8425
1953	8573
1954	8728
1955	8891

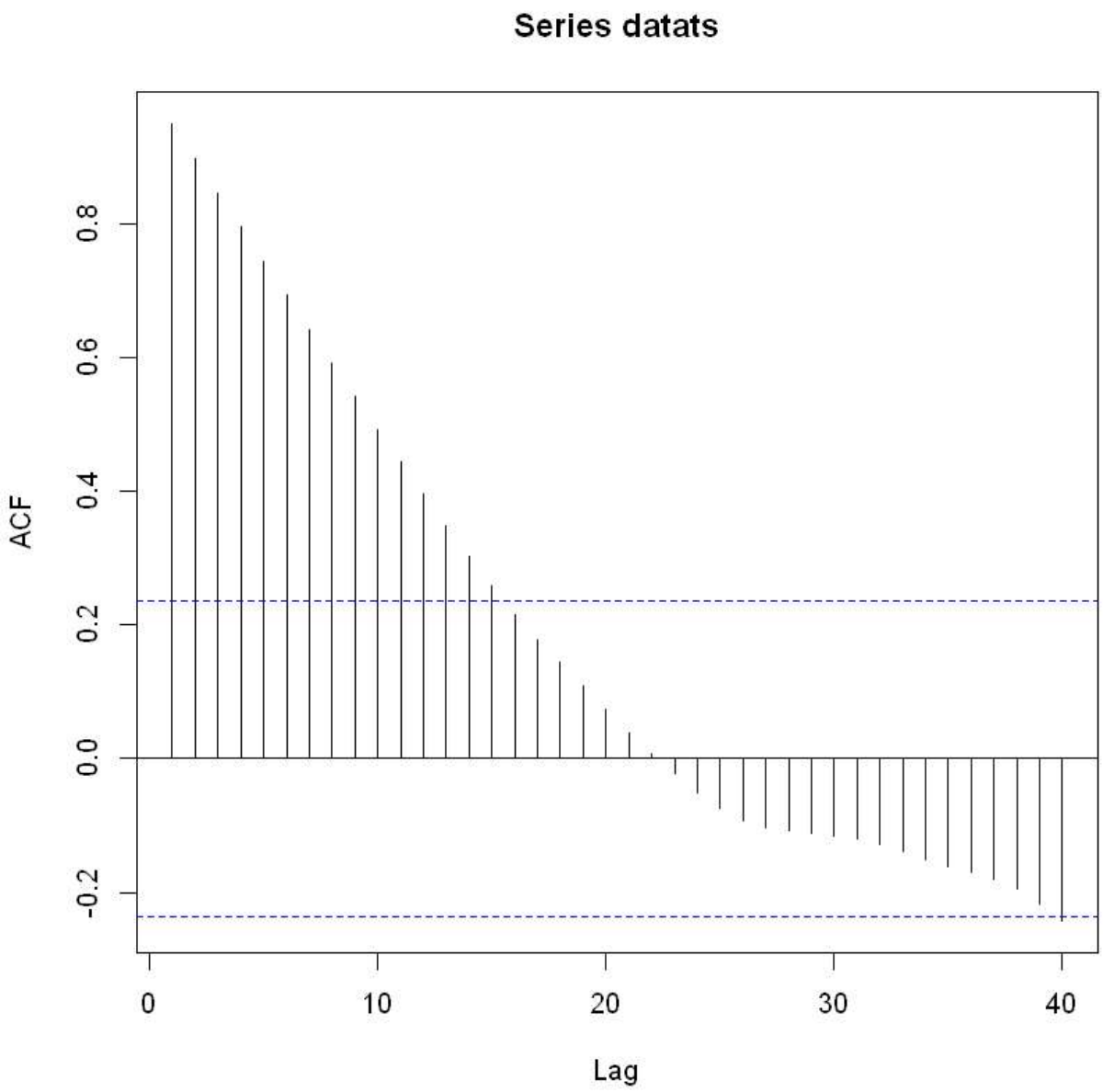
In [100…
```
datats <- ts(Data_popl$Population,start = c(1950,1),frequency=1)
#Data_popl$Population
autoplot(datats)
```



After analyzing the time series plot , it is clear that overall it is increasing, other than the period 1981 to 1991.

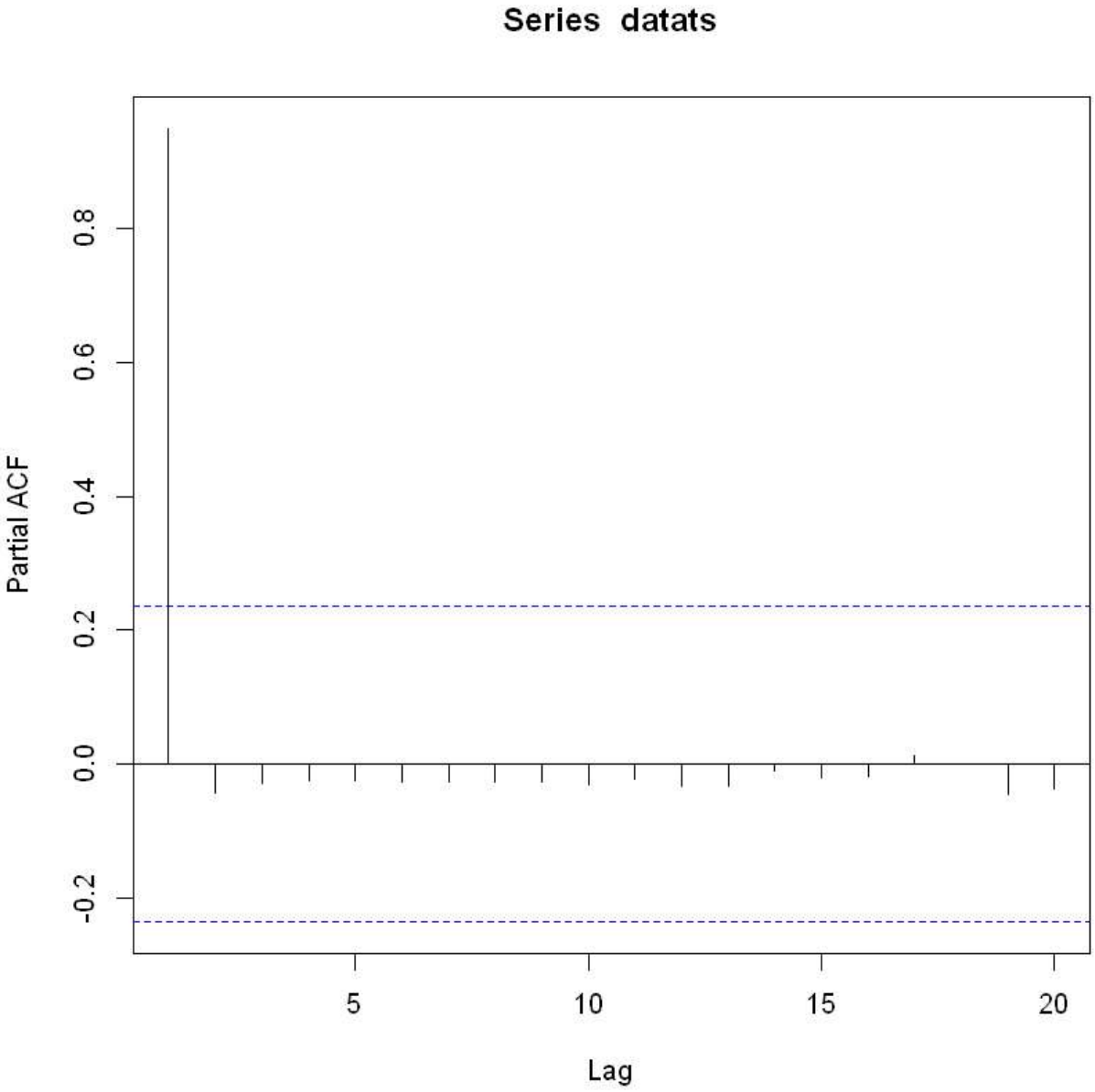
In [102…

```
acf(datats,lag.max=40)
```



In [103…

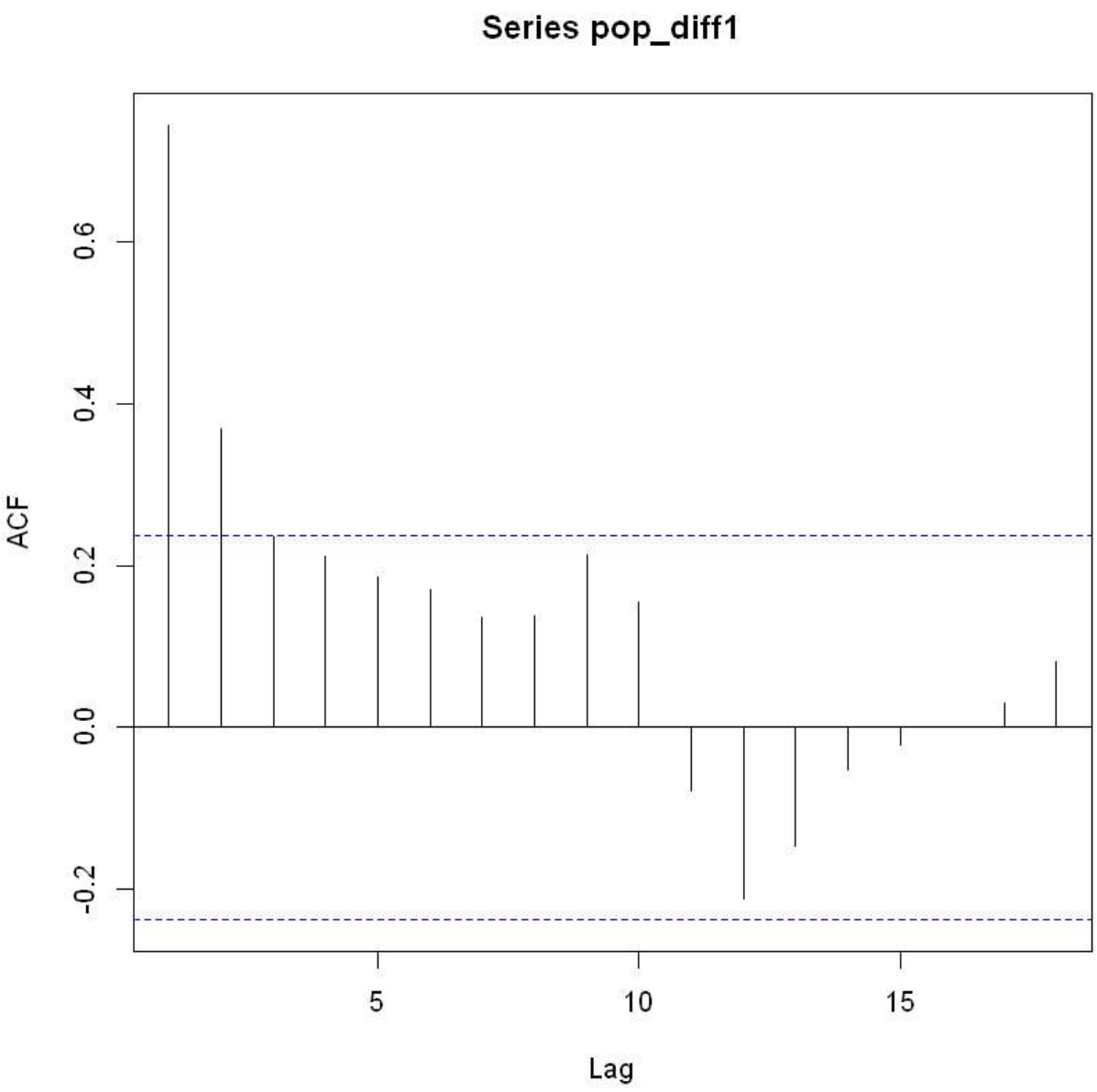
```
pacf <- pacf(datats,lag.max = 20,na.action = na.pass)
```



Here p value is greater than 0.05 hence this dataset is not stationary.We will take differencing in order to remove the trend and to make dataset stationary.

In [106…

```
acf(pop_diff1)
```



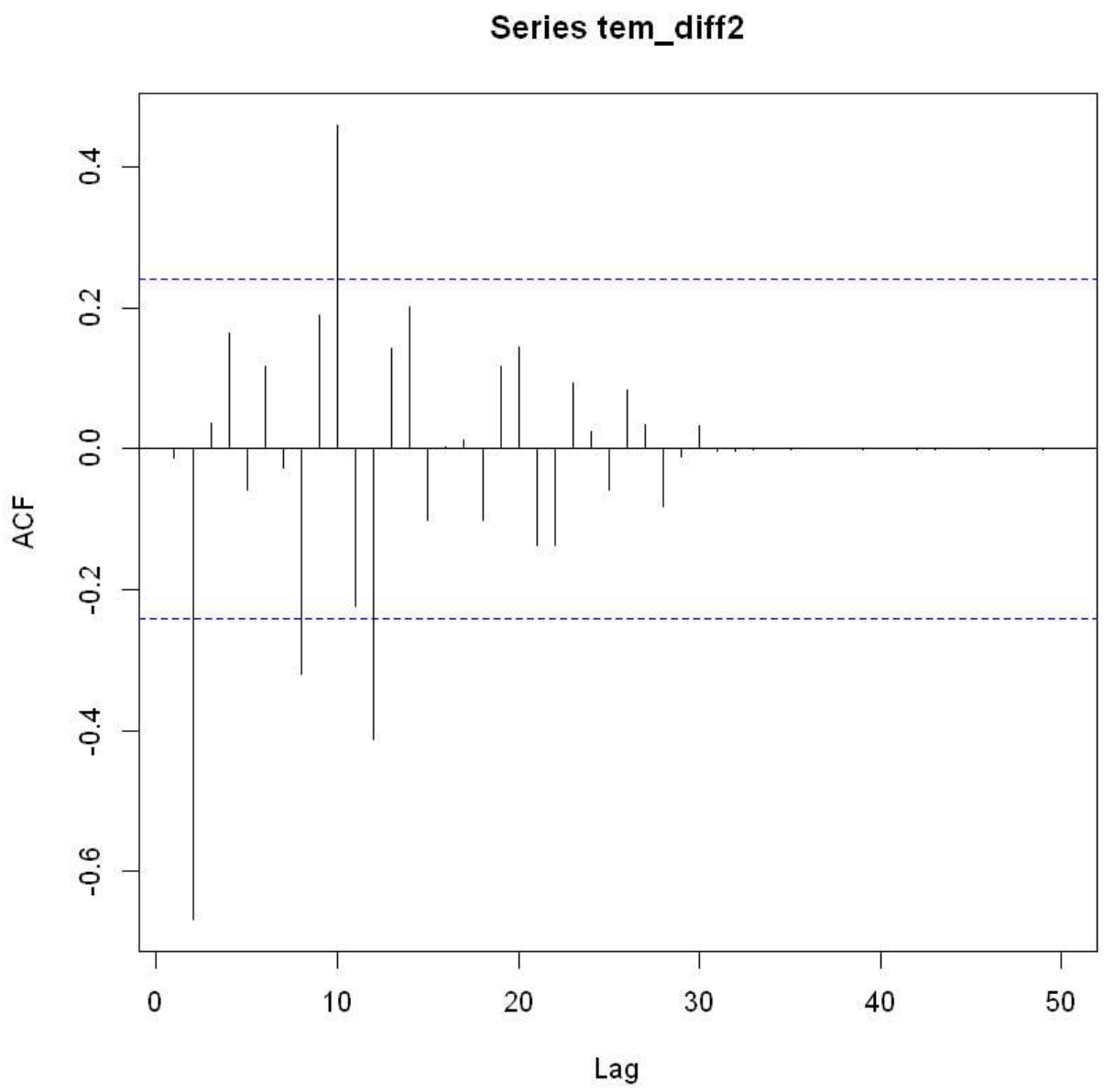
In [107…
```
adf.test(pop_diff1)
```

```
        Augmented Dickey-Fuller Test

data:  pop_diff1
Dickey-Fuller = -2.3786, Lag order = 4, p-value = 0.4214
alternative hypothesis: stationary
```

In [ ]:
```
Still p value is greater than 0.05 hence we need to take further differencing.
```

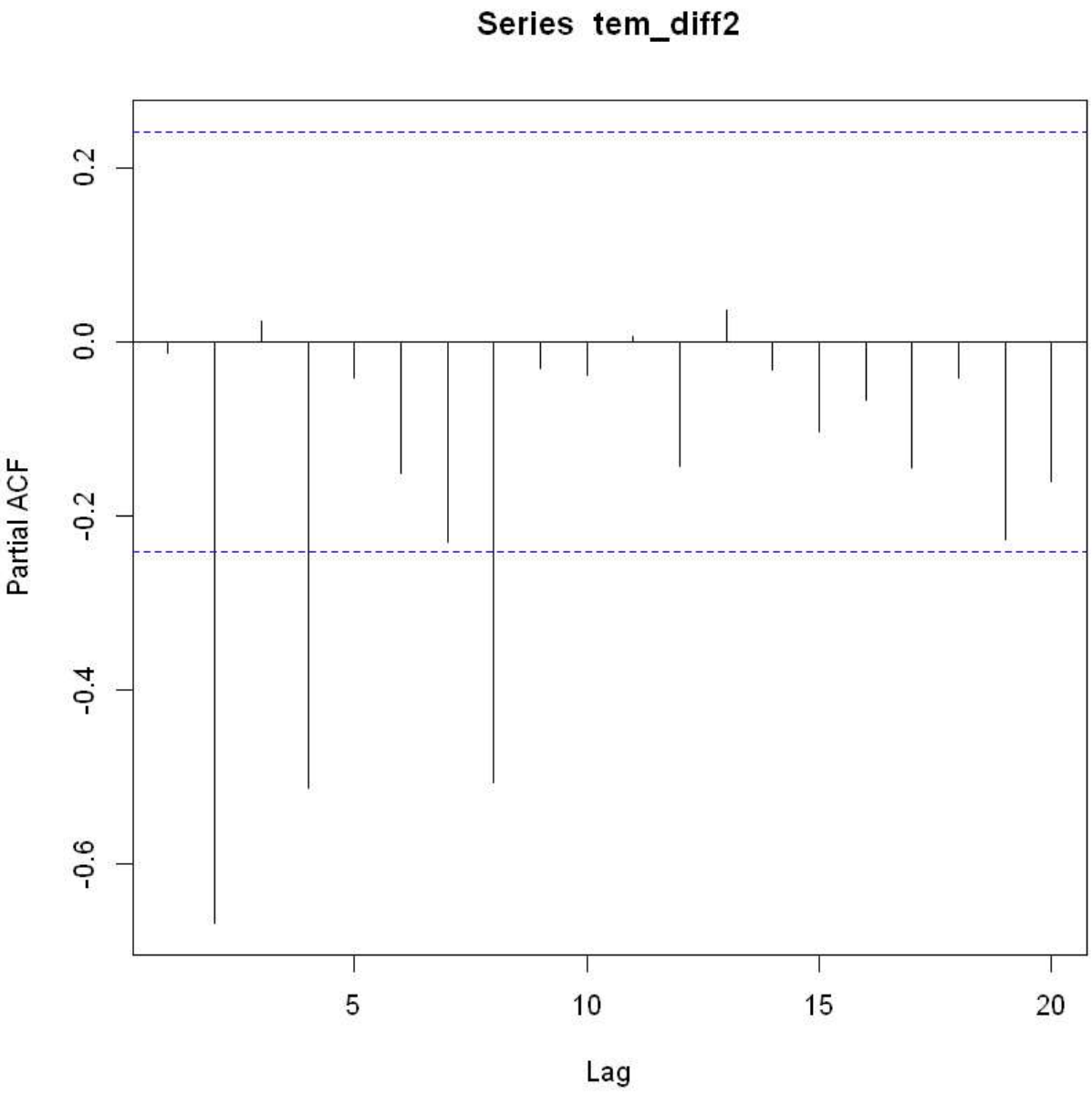
In [108…
```
tem_diff2=diff(pop_diff1,differences=2)
```

In [109…
```
acf(tem_diff2,lag.max=50)
```



After analyzing ACF plot (2/4/8,2,2/10.12)(non seasonal)

In [59]:
```
pacf <- pacf(tem_diff2,lag.max = 20,na.action = na.pass)
```



In [110…

```
adf.test(tem_diff2)
```

```
Warning message in adf.test(tem_diff2):
"p-value smaller than printed p-value"
        Augmented Dickey-Fuller Test

data:  tem_diff2
Dickey-Fuller = -5.7325, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

In [31]:

```
fit <- auto.arima(tem_diff2,trace = TRUE)
fit
```

```
 ARIMA(2,0,2) with non-zero mean : Inf
 ARIMA(0,0,0) with non-zero mean : 991.6024
 ARIMA(1,0,0) with non-zero mean : 993.7883
 ARIMA(0,0,1) with non-zero mean : Inf
 ARIMA(0,0,0) with zero mean     : 989.4745
```

```
 ARIMA(1,0,1) with non-zero mean : Inf

 Best model: ARIMA(0,0,0) with zero mean

Series: tem_diff2
ARIMA(0,0,0) with zero mean

sigma^2 estimated as 184040:  log likelihood=-493.71
AIC=989.41   AICc=989.47   BIC=991.6
```

In [111…

```
fit <- Arima(tem_diff2, order=c(8,2,2))
fit
```

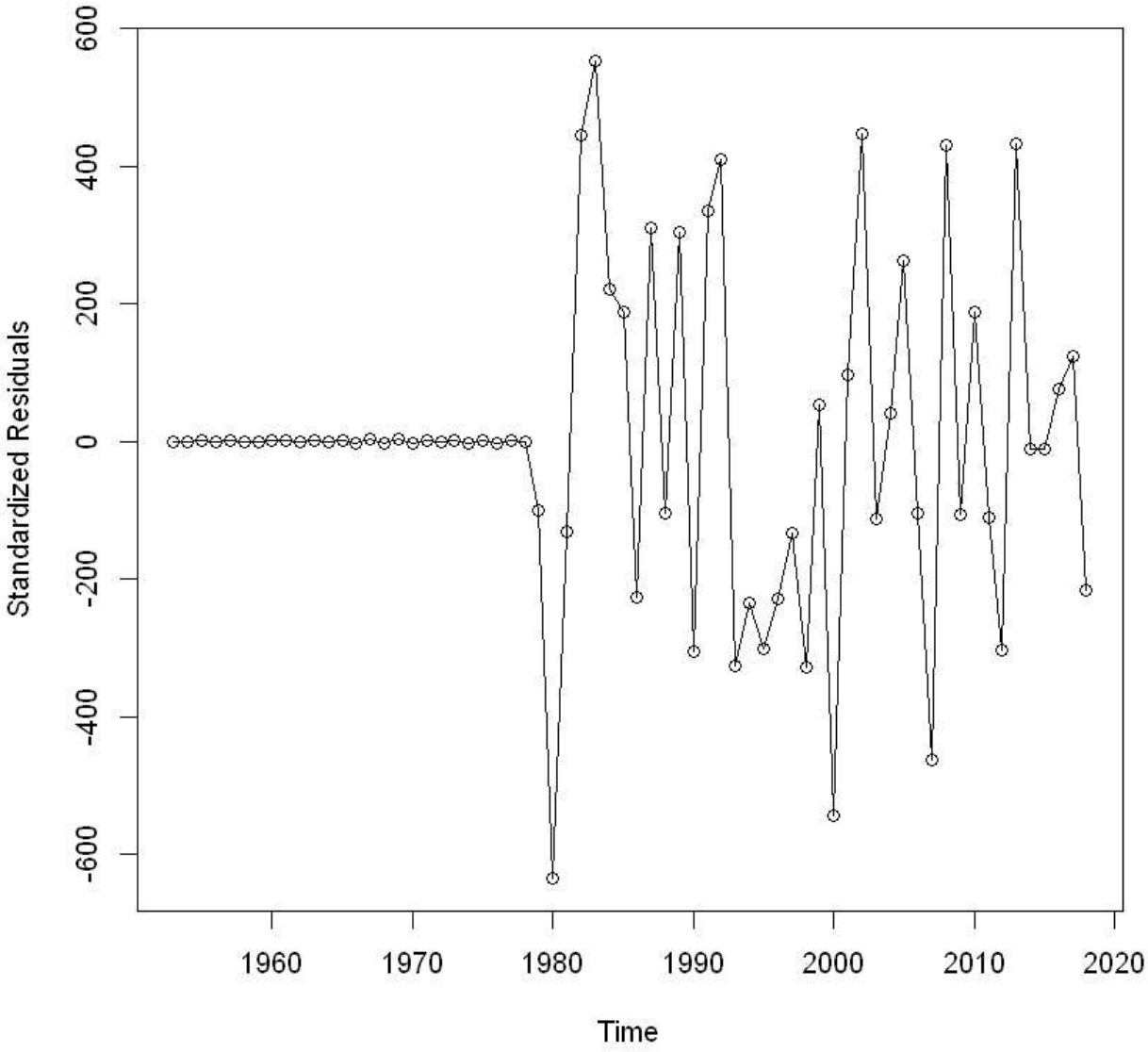
```
Series: tem_diff2
ARIMA(8,2,2)

Coefficients:
          ar1      ar2      ar3      ar4      ar5      ar6      ar7      ar8
      -0.1452  -1.1268  -0.2727  -0.9216  -0.3236  -0.6199  -0.2192  -0.4365
s.e.   0.1104   0.1056   0.1613   0.1561   0.1554   0.1540   0.1009   0.1024
          ma1      ma2
      -1.9961  0.9997
s.e.   0.0881  0.0881

sigma^2 estimated as 64415:  log likelihood=-451.28
AIC=924.56   AICc=929.64   BIC=948.31
```

Here Arima(8,2,2)has the lower AIC and BIC values , hence we will go with (8,2,2)

In [112…

```
plot(window(residuals(fit),start=c(1950,1)),ylab='Standardized Residuals',type='o')
```

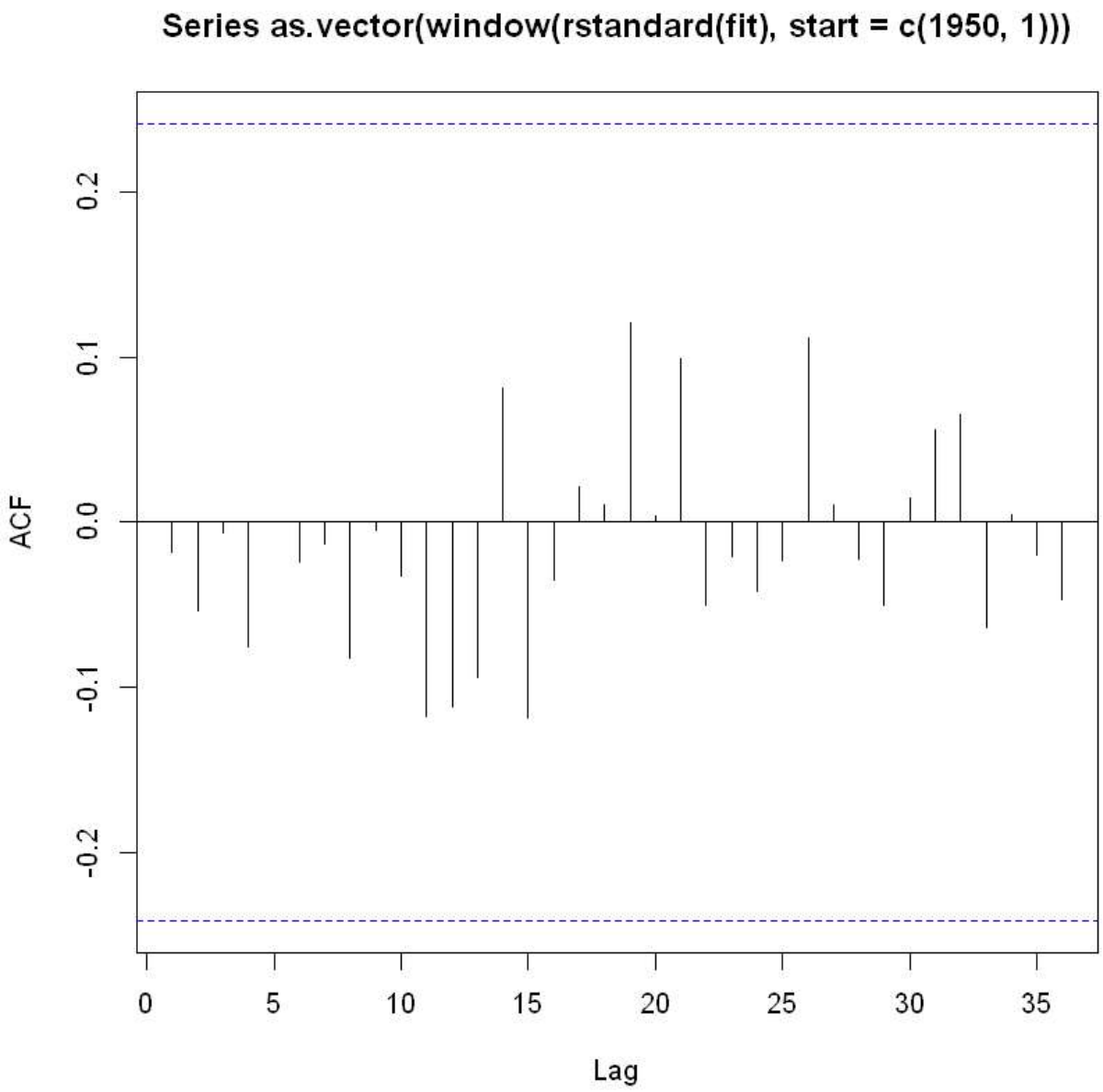
```
Warning message in window.default(x, ...):
"'start' value not changed"
```



In [113…

```
acf(as.vector(window(rstandard(fit),start=c(1950,1))),lag.max=36)
```

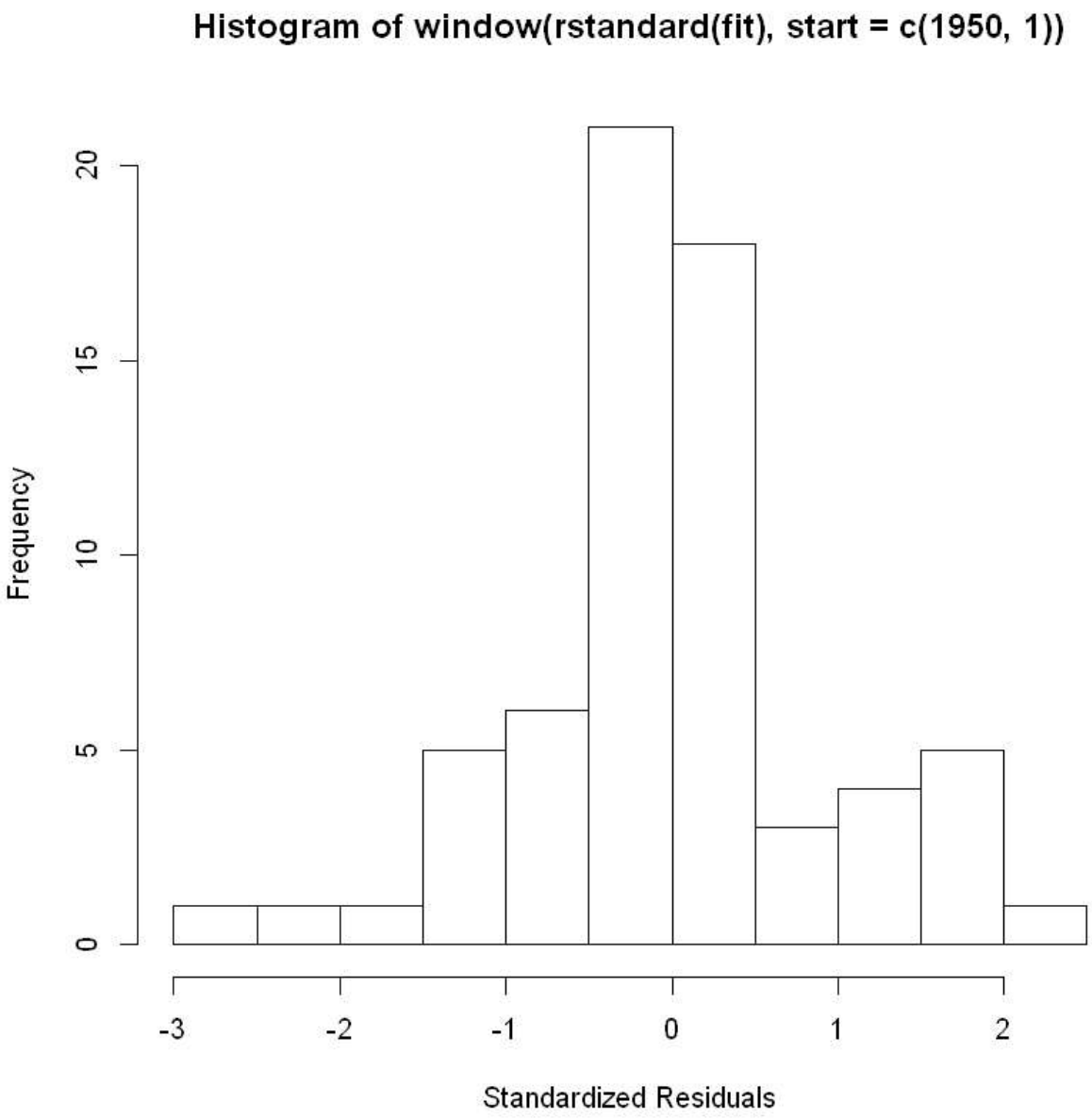
```
Warning message in window.default(x, ...):
"'start' value not changed"
```



In [114…

```
hist(window(rstandard(fit),start=c(1950,1)),xlab='Standardized Residuals')
```

```
Warning message in window.default(x, ...):
"'start' value not changed"
```

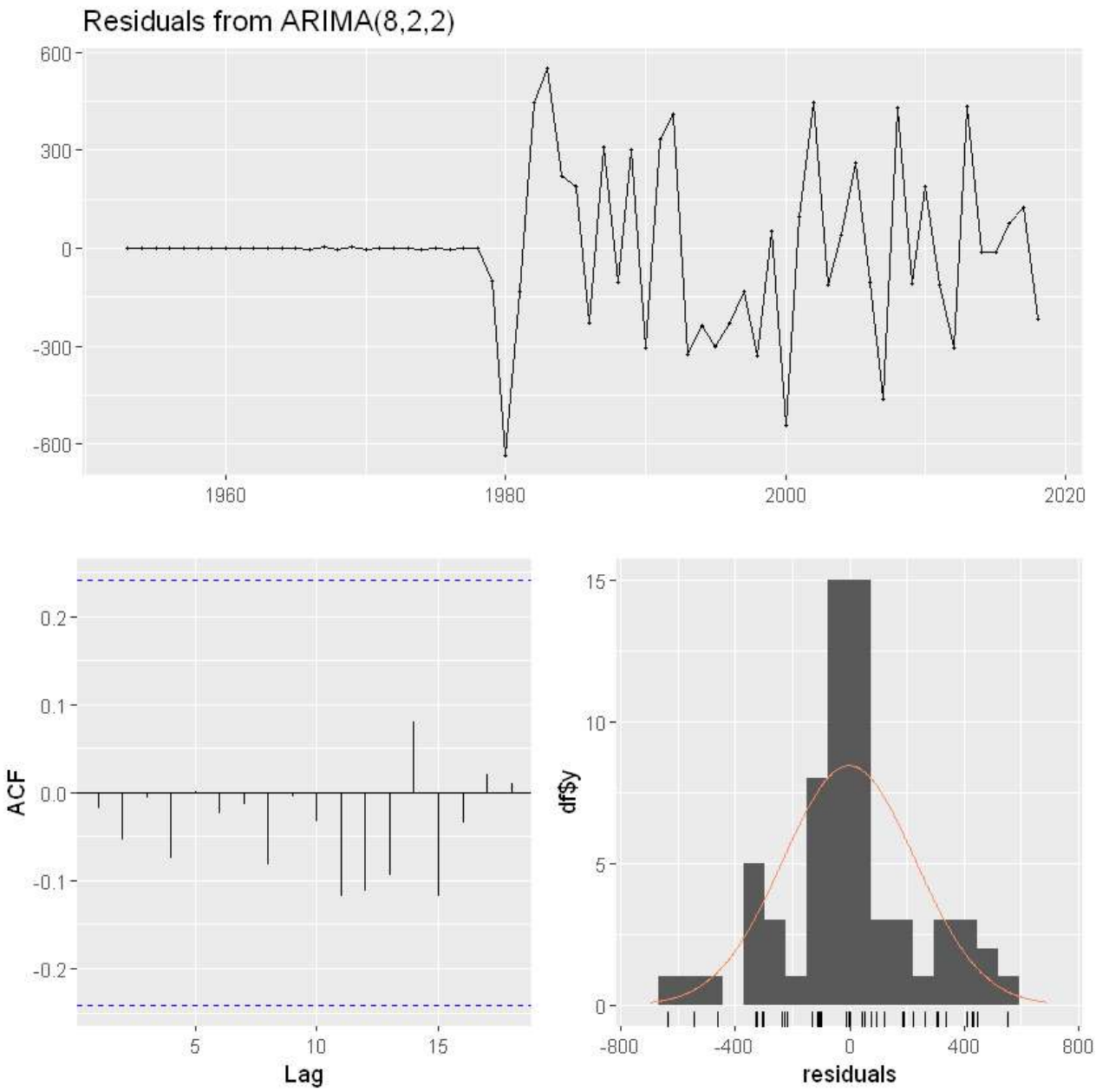


In [115…
```
shapiro.test(residuals(fit))
```

```
        Shapiro-Wilk normality test

data:  residuals(fit)
W = 0.93038, p-value = 0.001142
```

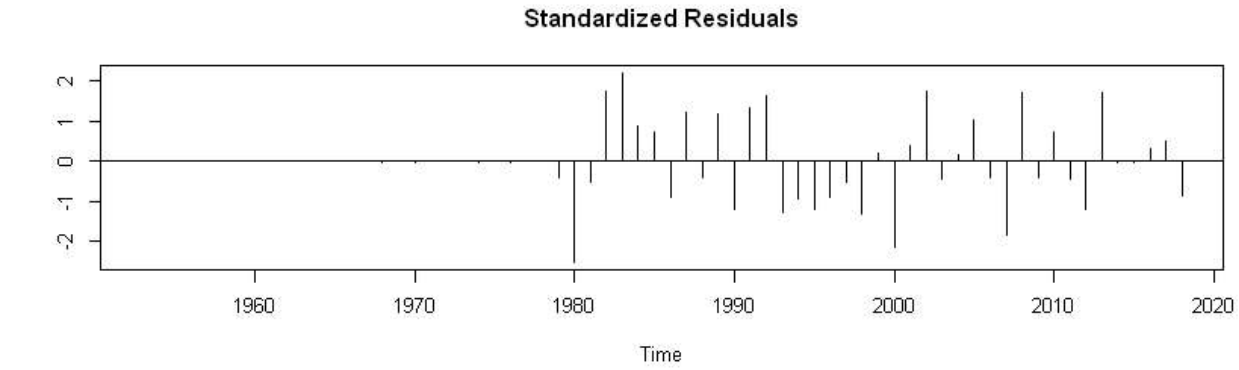
In [116…
```
checkresiduals(fit)
```

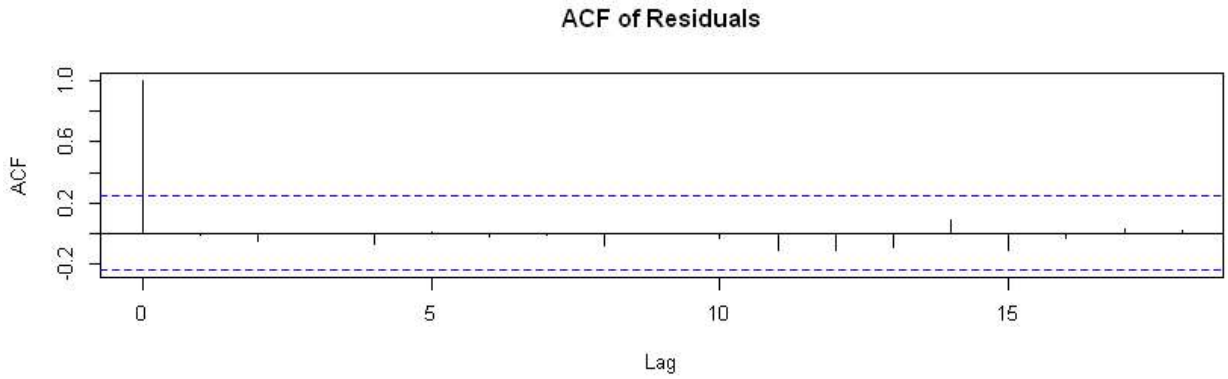
```
        Ljung-Box test

data:  Residuals from ARIMA(8,2,2)
Q* = 4.2014, df = 3, p-value = 0.2405

Model df: 10.   Total lags used: 13
```

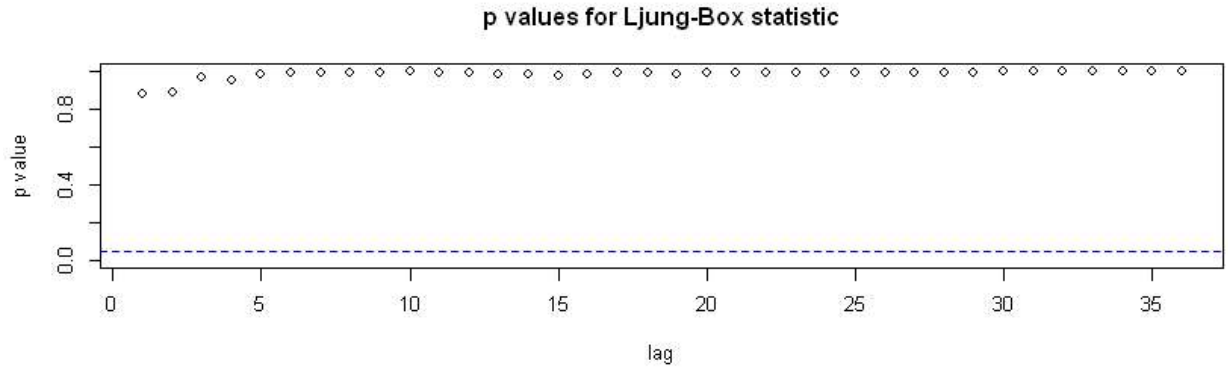


In [117…

```
tsdiag(fit,gof=36,omit.initial=F)
```

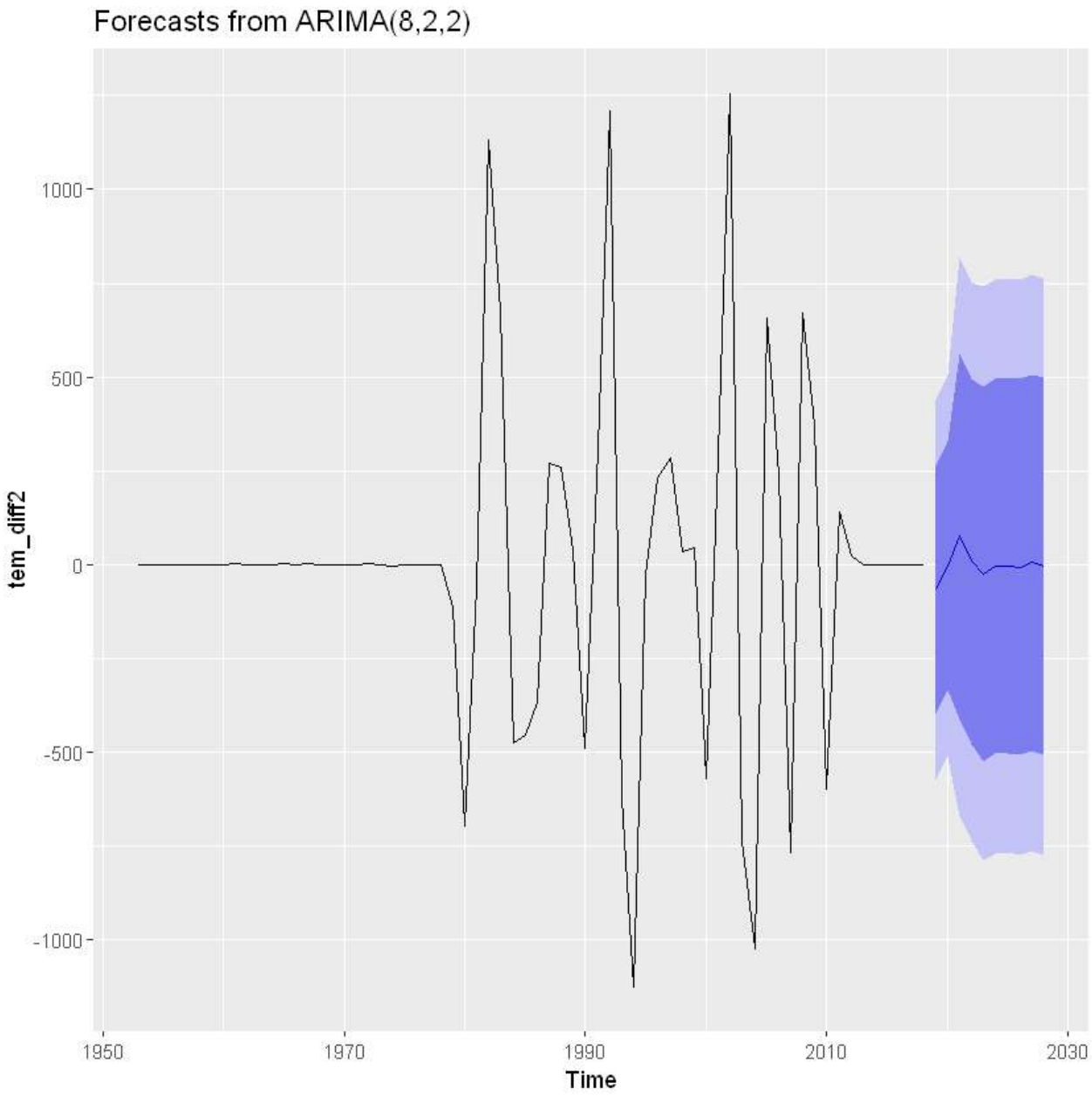






In [118…

```
autoplot(forecast(fit))
```


Forecasts from ARIMA(8,2,2)
1000
500
0
-500
-1000
tem_diff2
1950
1970
1990
2010
2030
Time

In [ ]: