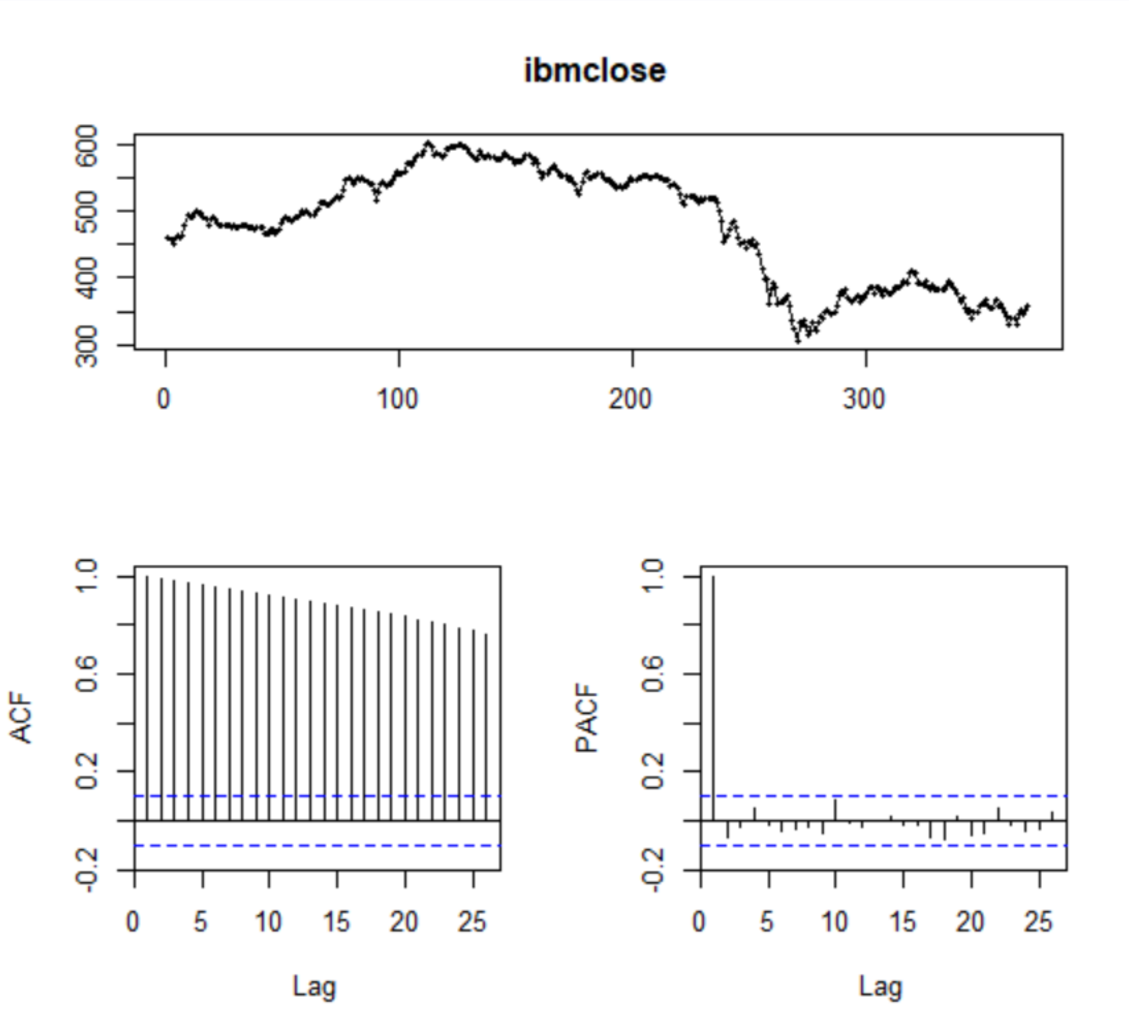
Q1

A classic example of a non-stationary series is the daily closing IBM stock price series(data set ibmclose).

Use R to plot the daily closing prices for IBM stock and the ACF and PACF.

```
library(fpp2)
```

```
tsdisplay(ibmclose)
```



Explain how each plot shows that the series is non-stationary and should be differenced

For the time series plot, it may show that the mean and variance of ibmclose is not constant. So the series may be non-stationary.

Very slowly decaying ACF may show that the series maybe non-stationary.

PACF cannot tell it is non-stationary.

```
> ndiffs(ibmclose)
[1] 1
> nsdiffs(ibmclose)
Error in nsdiffs(ibmclose) : Non seasonal data
```

Q1b

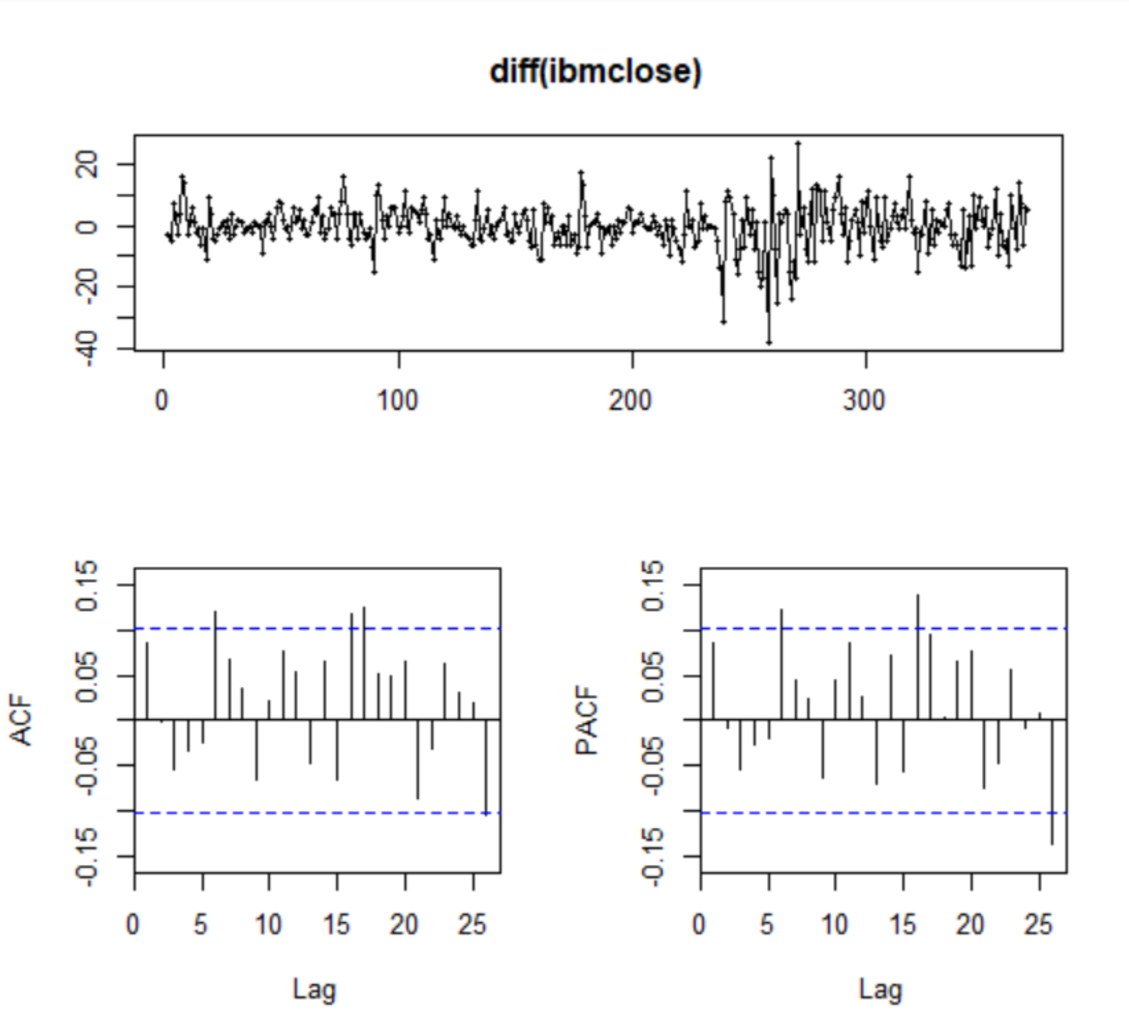
Form one or more hypothesis as to the nature of the underlying data

1.Staying with ARIMA(p,d,q) model and using ndiffs(ibmclose) and nsdiffs(ibmclose)

2.d=1

3.p will be order of AR component. This can be determined using PACF

```
tsdisplay(diff(ibmclose))
```



4.Hence p = 0

5.q is order of MA component. This can be determined using the ACF

6.Hence q = 0

Fit an Arima model to the data in R. See class slides for syntax.

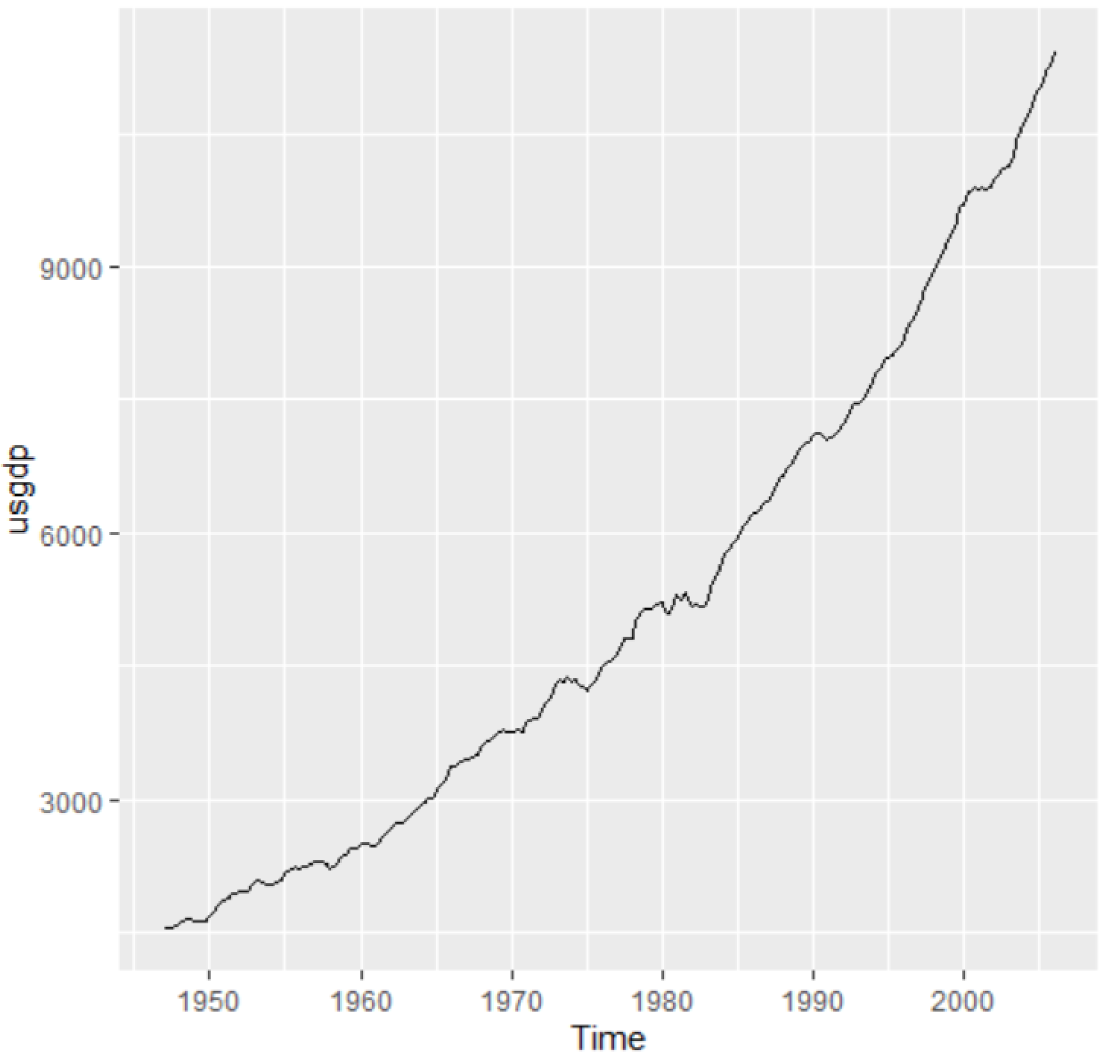
```
> Arima(ibmclose, order=c(0,1,0),include.drift=TRUE)
Series: ibmclose
ARIMA(0,1,0) with drift

Coefficients:
        drift
      -0.2799
s.e.   0.3778

sigma^2 = 52.68:  log likelihood = -1251.09
AIC=2506.19   AICc=2506.22   BIC=2514
> Arima(ibmclose, order=c(0,1,0))
Series: ibmclose
ARIMA(0,1,0)

sigma^2 = 52.62:  log likelihood = -1251.37
AIC=2504.74   AICc=2504.75   BIC=2508.64
```

Q2

For the usgdp series:

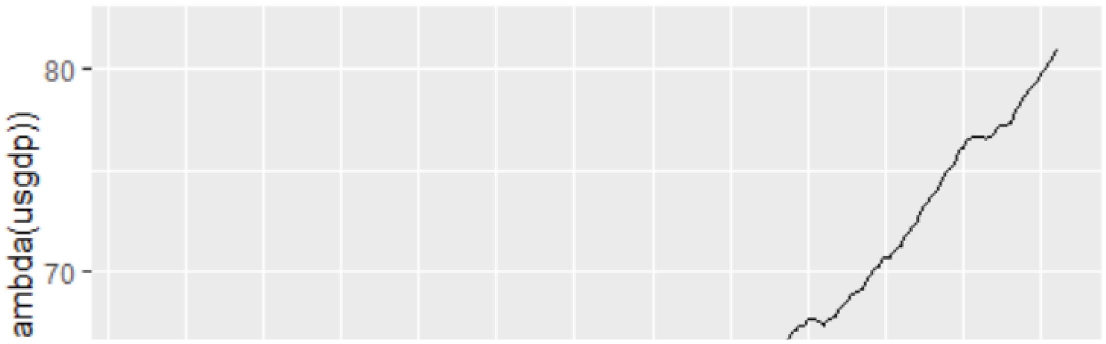
a.if necessary, find a suitable Box-Cox transformation for the data;

```
1.Raw time series (visually) shows some convexity
```



2.BoxCox transformed time series appears more linear

```
autoplot(BoxCox(usgdp, lambda=BoxCox.lambda(usgdp)))
```



3.lambda value

```
> BoxCox.lambda(usgdp)
[1] 0.366352
```

b.fit a suitable ARIMA model to the transformed data using auto.arima();

1.uses auto.arima on BoxCox transformed version of USGDP

```
> auto.arima(BoxCox(usgdp, lambda=BoxCox.lambda(usgdp)))
Series: BoxCox(usgdp, lambda = BoxCox.lambda(usgdp))
ARIMA(2,1,0) with drift

Coefficients:
         ar1     ar2   drift
      0.2795  0.1208  0.1829
s.e.  0.0647  0.0648  0.0202

sigma^2 = 0.03518:  log likelihood = 61.56
AIC=-115.11   AICc=-114.94   BIC=-101.26
```

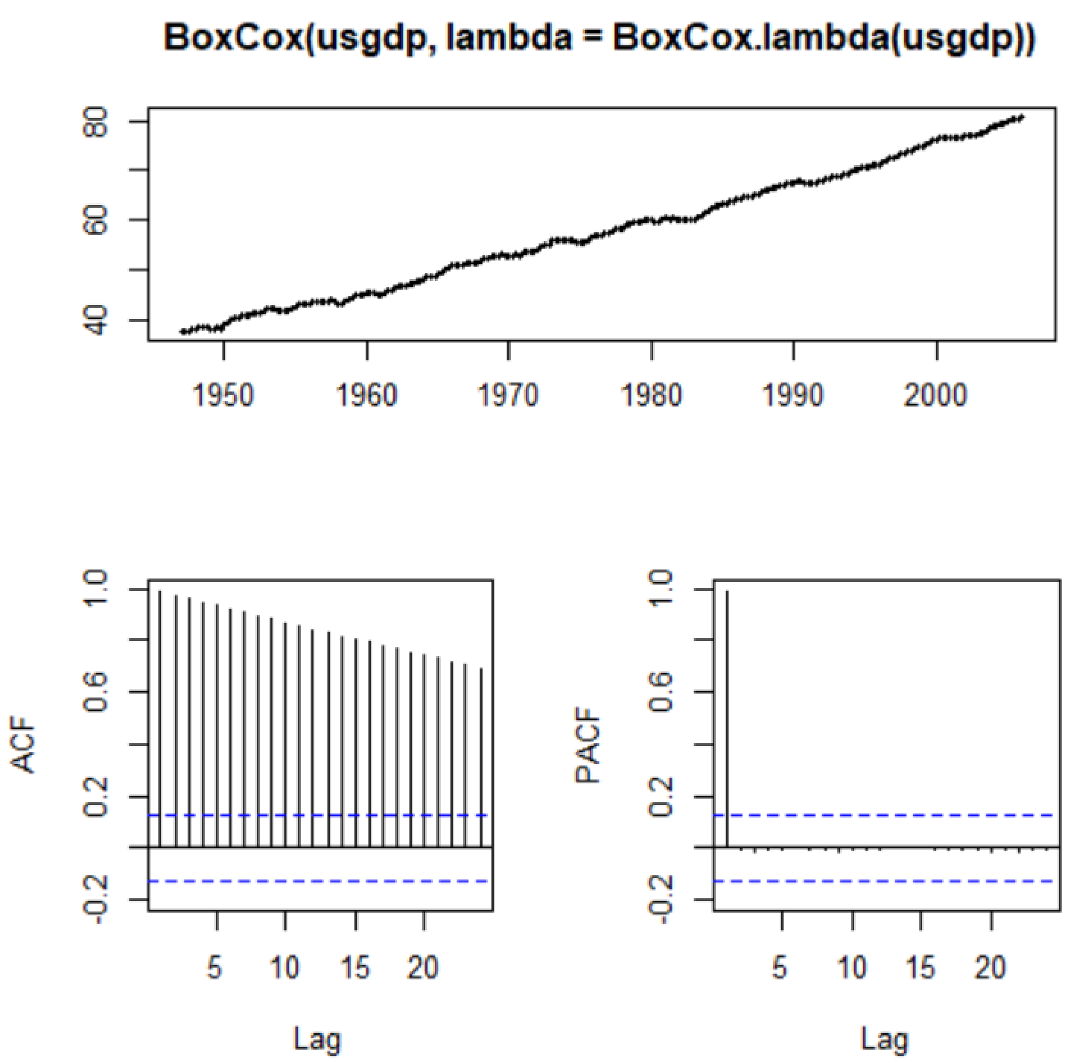
2.Replicates the same model using Arima on raw variable

```
> Arima(usgdp, order=c(2,1,0),include.drift=TRUE, lambda='auto')
Series: usgdp
ARIMA(2,1,0) with drift
Box Cox transformation: lambda= 0.3663571

Coefficients:
         ar1     ar2   drift
      0.2795  0.1208  0.1829
s.e.  0.0647  0.0648  0.0202

sigma^2 = 0.03519:  log likelihood = 61.55
AIC=-115.09   AICc=-114.92   BIC=-101.24
```

c.try some other plausible models by experimenting with the orders chosen;

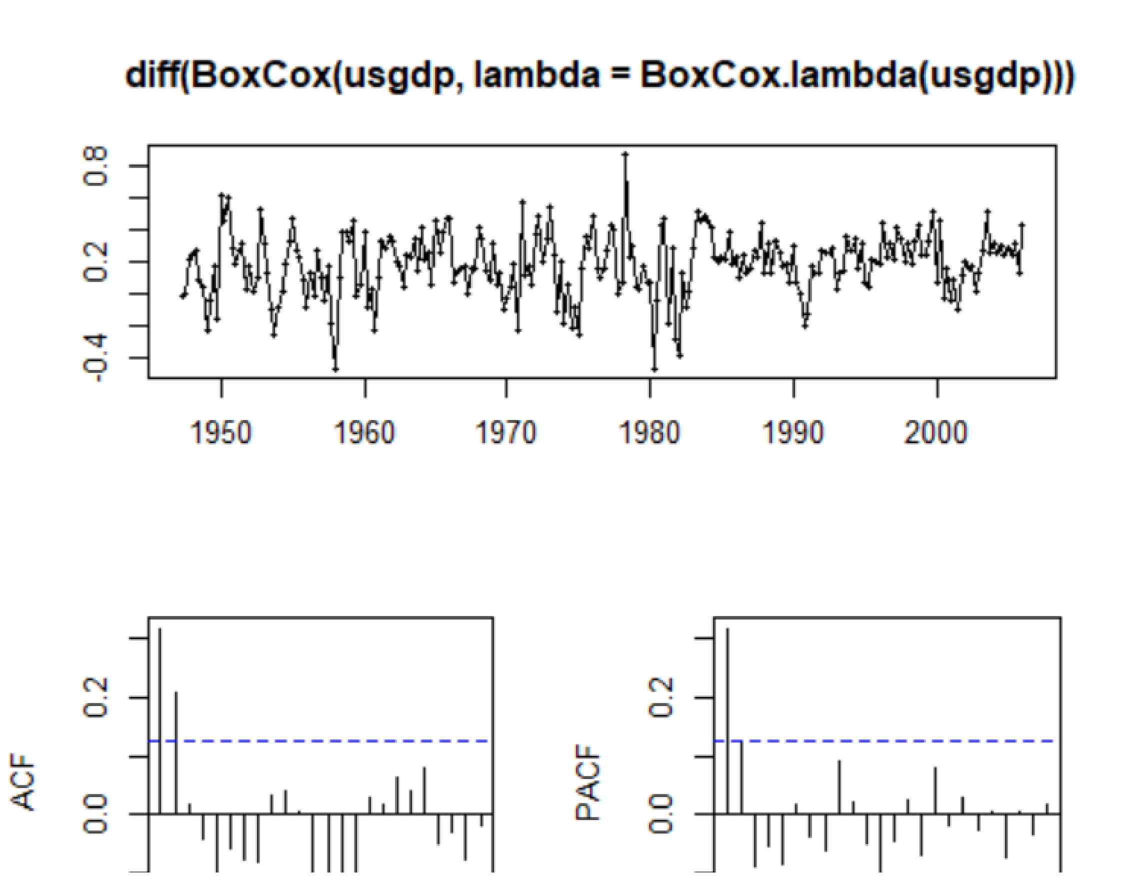
Note that for 2c, you should follow the process overviewed in class where we inspect ACF/PACF, form hypotheses, fit the models, check AICc AND Ljung-Box test

```
tsdisplay(BoxCox(usgdp, lambda=BoxCox.lambda(usgdp)))
```




```
> nsdiffs(BoxCox(usgdp, lambda=BoxCox.lambda(usgdp)))
[1] 0
> ndiffs(BoxCox(usgdp, lambda=BoxCox.lambda(usgdp)))
[1] 1
```

```
tsdisplay(diff(BoxCox(usgdp, lambda=BoxCox.lambda(usgdp))))
```




```
tsdisplay(diff(BoxCox(usgdp, lambda=BoxCox.lambda(usgdp))))
```

Fast decaying pattern in PACF probably eliminates possibility of MA model

Possible hypotheses:

ARIMA(1,1,0)

ARIMA(2,1,0)

ARIMA(1,1,0)[Want to be certain]

ARIMA(1,1,0)

```
> Arima(usgdp, order=c(1,1,0),include.drift=TRUE, lambda='auto')
Series: usgdp
ARIMA(1,1,0) with drift
Box Cox transformation: lambda= 0.3663571

Coefficients:
         ar1   drift
      0.3180  0.1831
s.e.  0.0619  0.0179

sigma^2 = 0.03556:  log likelihood = 59.82
AIC=-113.64   AICc=-113.54   BIC=-103.25
```

ARIMA(2,1,0)

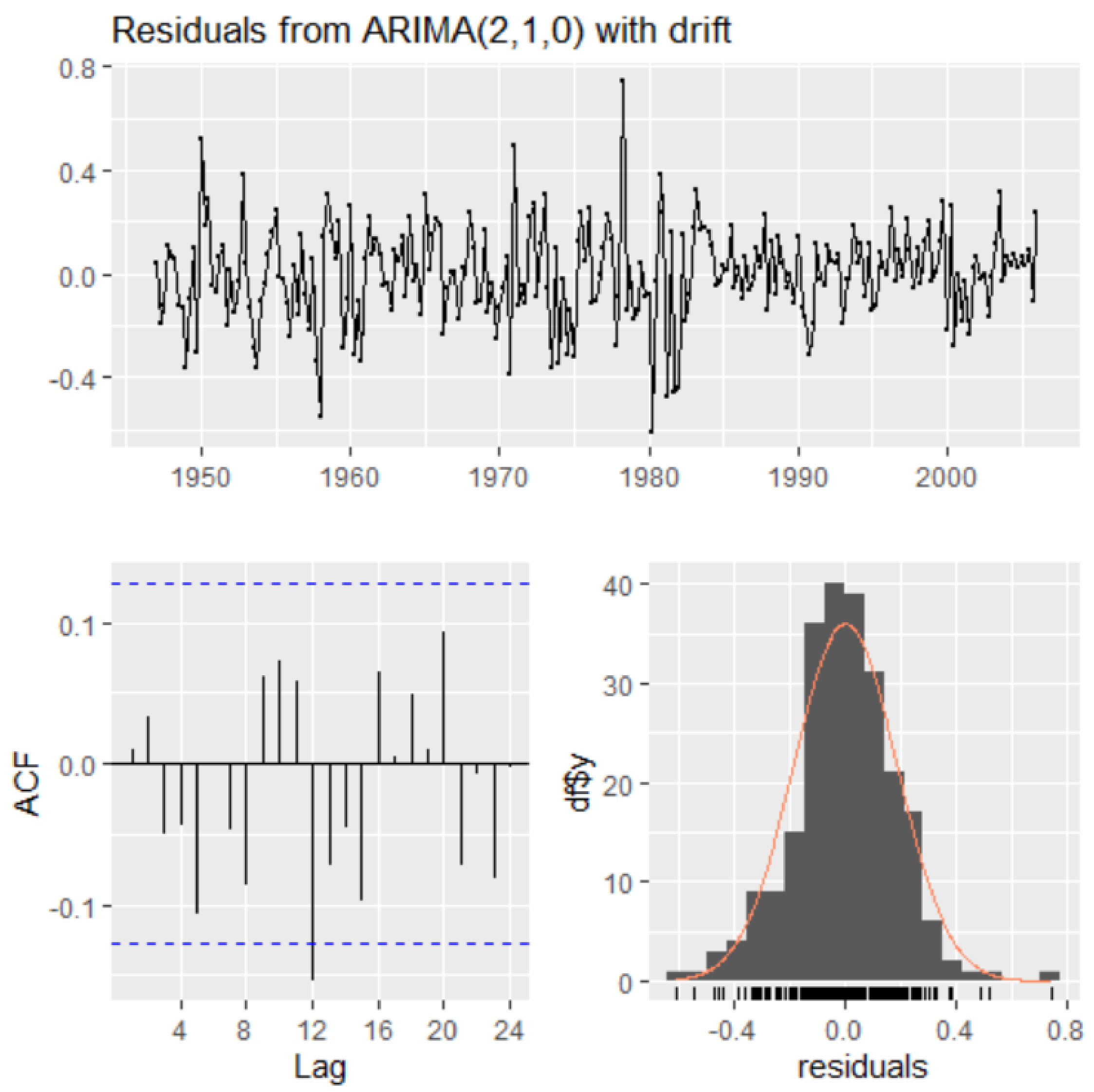
```
> Arima(usgdp, order=c(2,1,0),include.drift=TRUE, lambda='auto')
Series: usgdp
ARIMA(2,1,0) with drift
Box Cox transformation: lambda= 0.3663571

Coefficients:
         ar1     ar2   drift
      0.2795  0.1208  0.1829
s.e.  0.0647  0.0648  0.0202

sigma^2 = 0.03519:  log likelihood = 61.55
AIC=-115.09   AICc=-114.92   BIC=-101.24
```

d.choose what you think is the best model and check the residual diagnostics;

You can access the data similar to Q1.

Choose AICC are smallest one, so choose ARIMA(2,1,0)

```
> checkresiduals(Arima(usgdp, order=c(2,1,0),include.drift=TRUE, lambda='auto'))

        Ljung-Box test

data:  Residuals from ARIMA(2,1,0) with drift
Q* = 6.5772, df = 6, p-value = 0.3617

Model df: 2.   Total lags used: 8
```

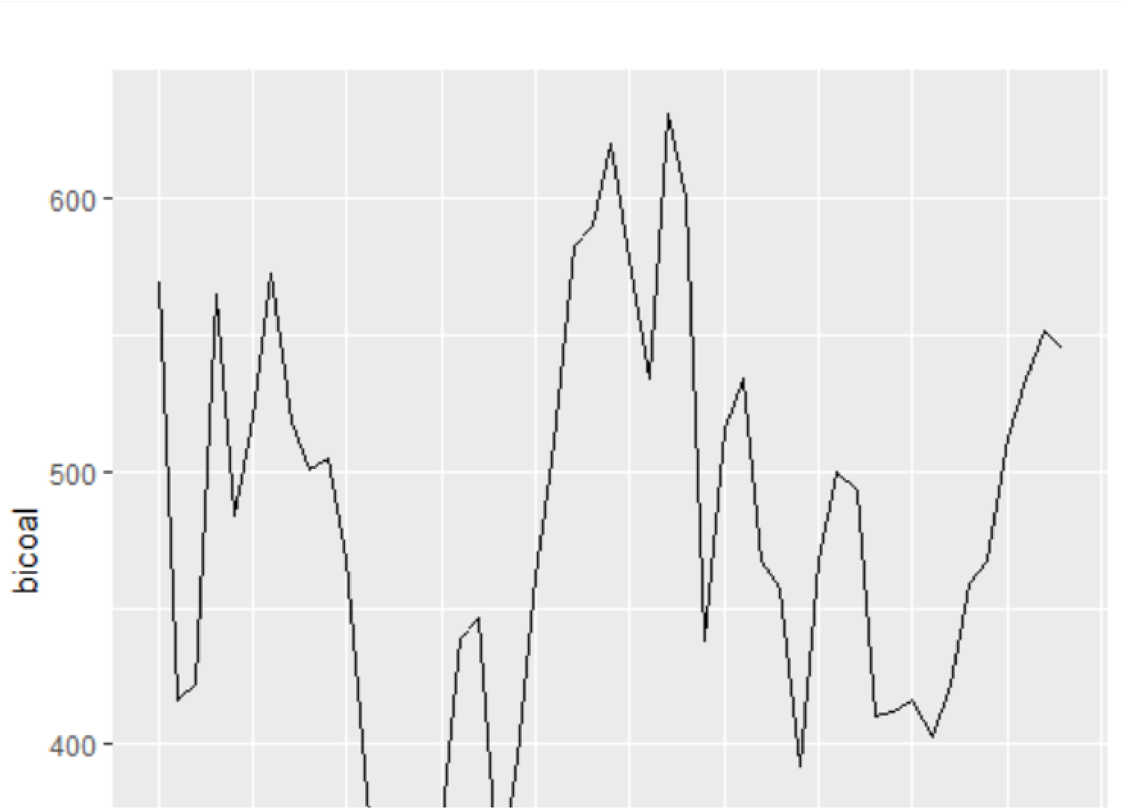


Q3

The annual bituminous coal production in the United States from 1920 to 1968 is in data set bicoal.

a.Produce a time plot of the data

```
autoplot(bicoal)
```



b.You decide to fit the following model to the series:

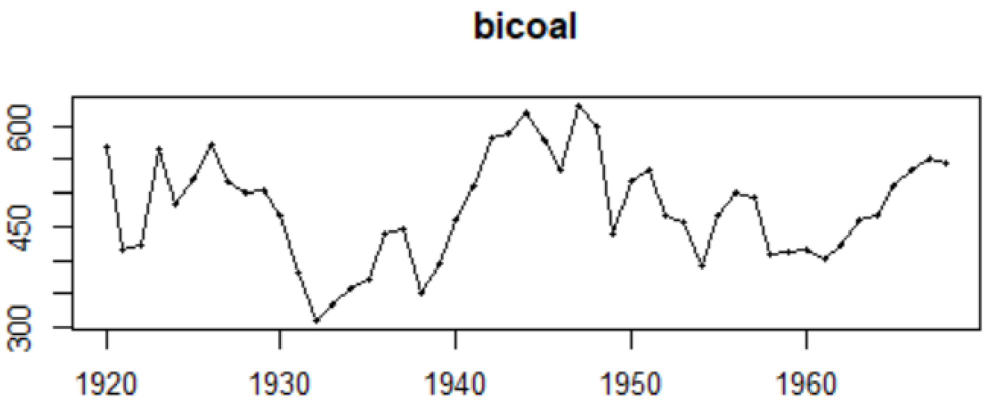
$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \varepsilon_t$$

where y_t is the coal production in year t and epsilon_t is a white noise series. What sort of ARIMA model is this?

```
ARIMA(4,0,0)
```

c.Explain why this model was chosen using the ACF and PACF

```
tsdisplay(bicoal)
```



PACF has some ambiguity. 1st column is clearly significant. So is column 4. On the other hand, column 2 and 3 are marginal at best. So the p should be 4

ACF shows a gradually decaying pattern, consistent with AR