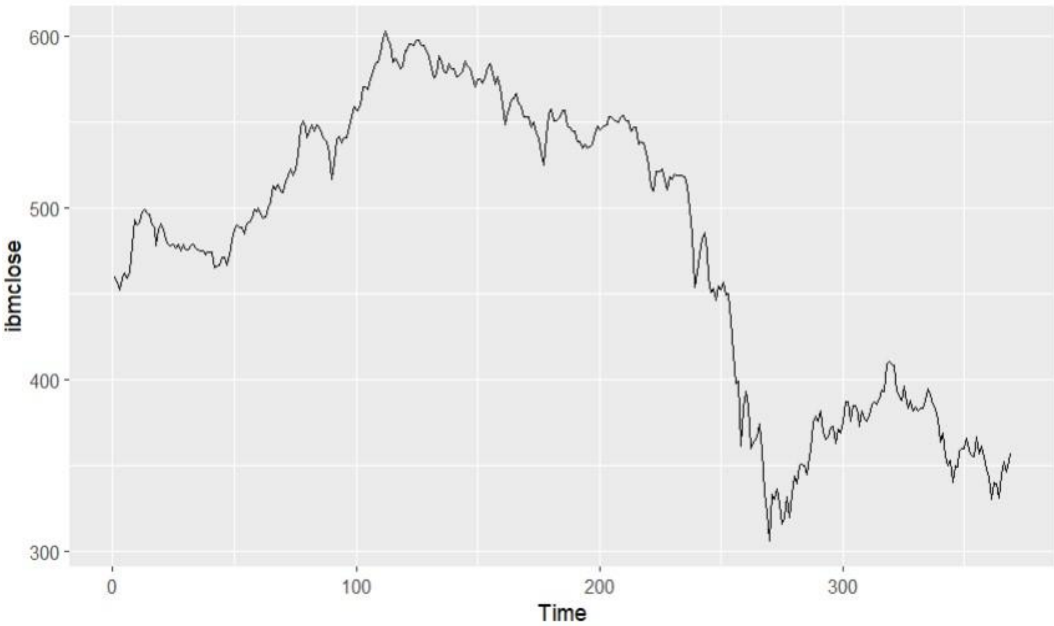
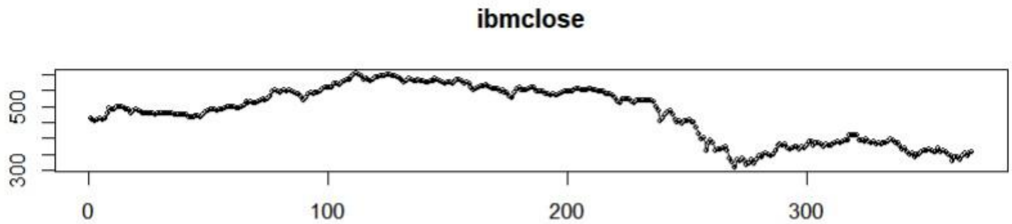
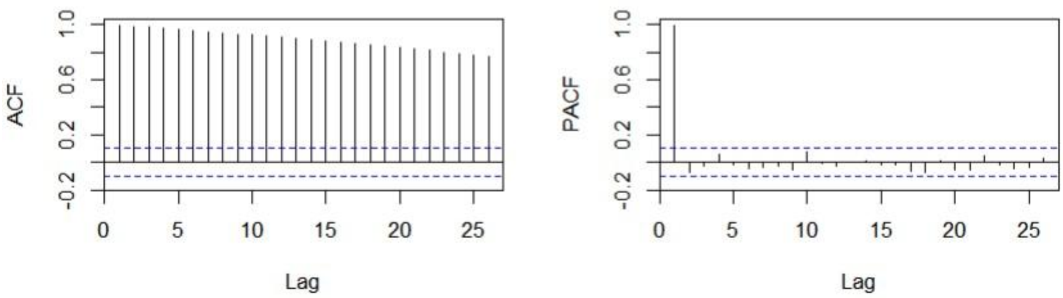
Q1.a Plot IBM CLOSE and its ACF and PACF and explain why it is non-stationary and need differenced

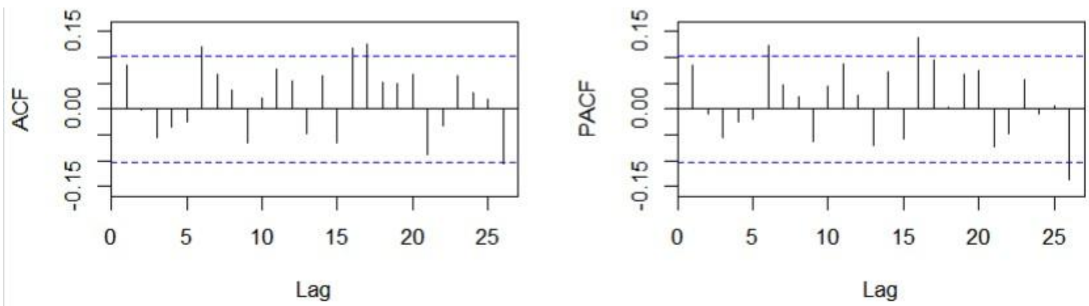






Explanation: If the auto-correlation function (ACF) shows a slow decay, it typically indicates that the time series data has persistence or a trend, and differencing is often needed.

Q1.b Form one or more hypothesis as to the nature of the underlying data, and fit an Arima model to the data in R.

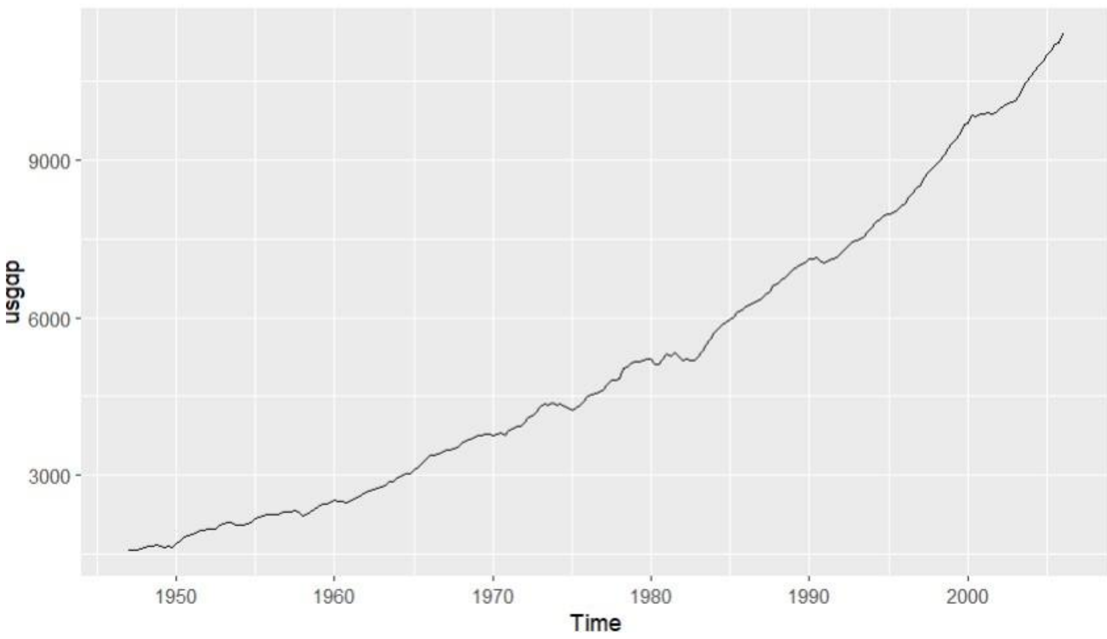


I first use the ndiff to test the times we need to difference the data,and do the difference,after that,i plot the ACF and PACF of the differenced data.As we can see above,I make a hypothesis of ARIMA(0,1,0).And I finally get the result as below.

```
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
```

Q2.a Find the suitable transformation for USGDP

I first plot the USGDP and it shows some convexity.



and the use the function to find the suitable lambda for USGDP

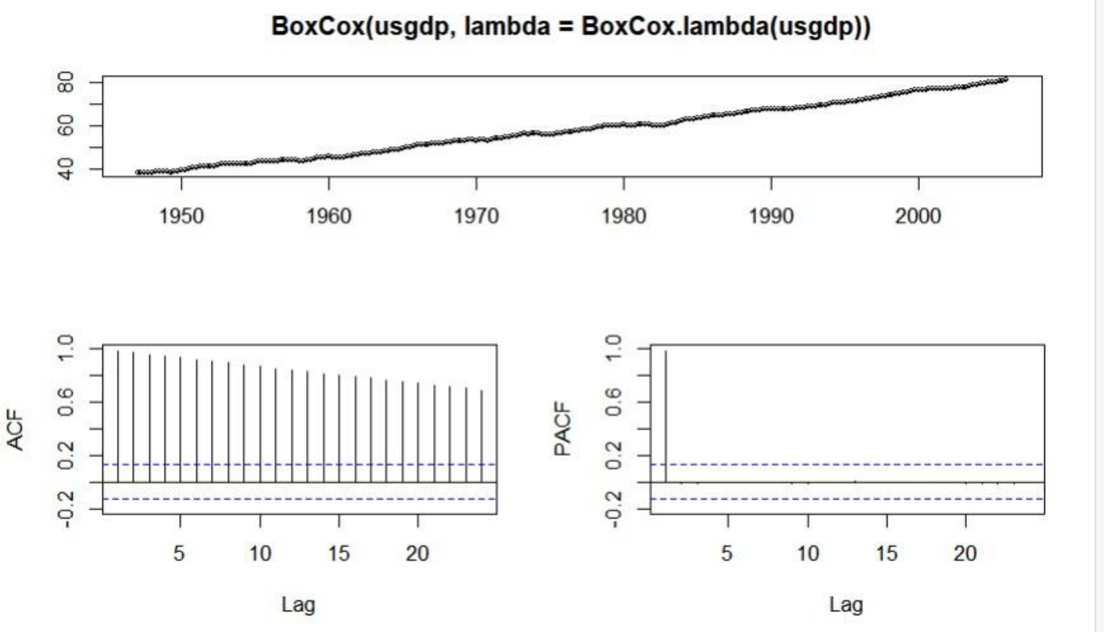
```
> BoxCox.lambda(usgdp)
[1] 0.366352
```

as shown above ,the best lambda for boxcox of USGDP is **0.366352**

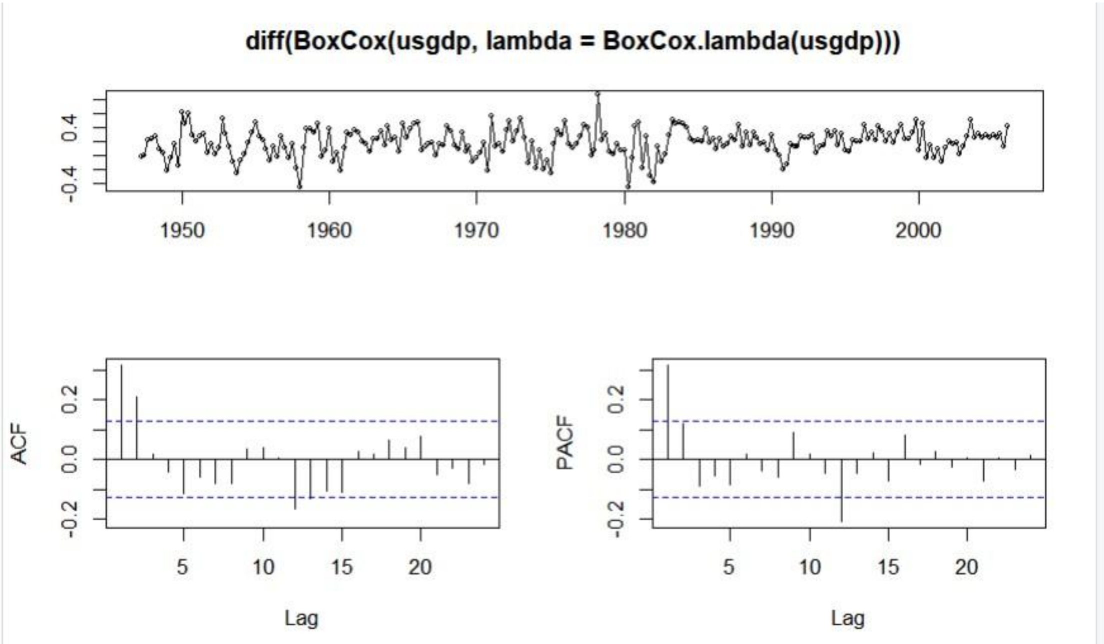
Q2.b Find suitable ARIMA for transformed data

```
> 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
```

Q2.c Try some other model by experimenting the orders chosen



Firstly I use tsdisplay to see the ACF and PACF for the transformed data,it’s clearly that the ACF is slowly decreasing,and the plot of transformed still shows a strongly time trend.So it is clearly that the transformed data still need to be differenced.After differencing,the tsdisplay of the differecnced data are as follows.



As we can see from the above ACF and PACF, it can be reasonable detected that the lag of AR is 1 or 2 lags,the lag of MA is 0 because there exist a sharply decreasing after lag0.
So i made the hypothesis of ARIMA(1,1,0) and ARIMA(2,1,0)

```
> 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
```

and do the residual test for ARIMA(1,1,0)

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

        Ljung-Box test

data:  Residuals from ARIMA(1,1,0) with drift
Q* = 10.274, df = 7, p-value = 0.1736

Model df: 1.   Total lags used: 8
```

The p-value is greater than 0.05, which means the residuals is close to normal distribution,which means our estimation is reasonable.

And we also run the order of (2,1,0),the result is as below.

```
> 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
```

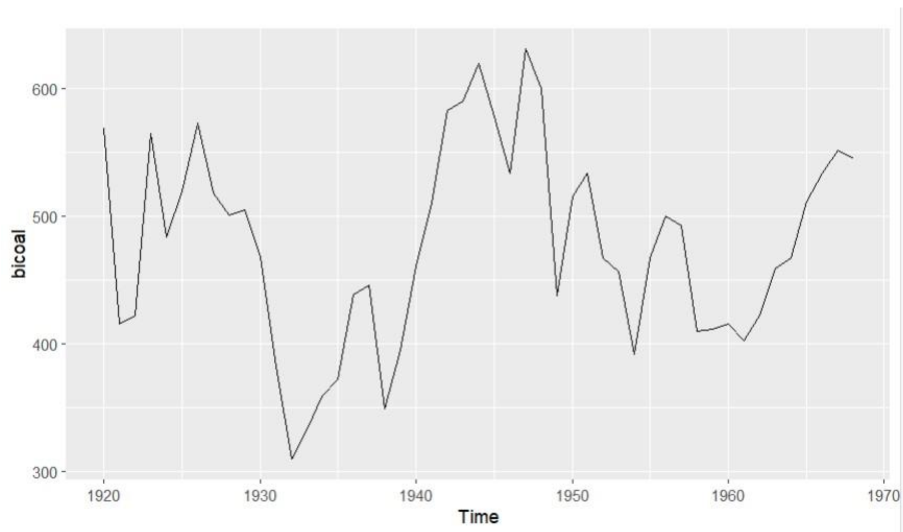
and we do the residual test.

```
> 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
```

the p-value is also greater than 0.05,which means the residuals is close to normal distribution,which means our estimation is reasonable.

Comparing the (1,1,0)and(2,1,0),the AICc is lower for(2,1,0) ,which means the (2,1,0) is a better ARIMA model for USGDP.

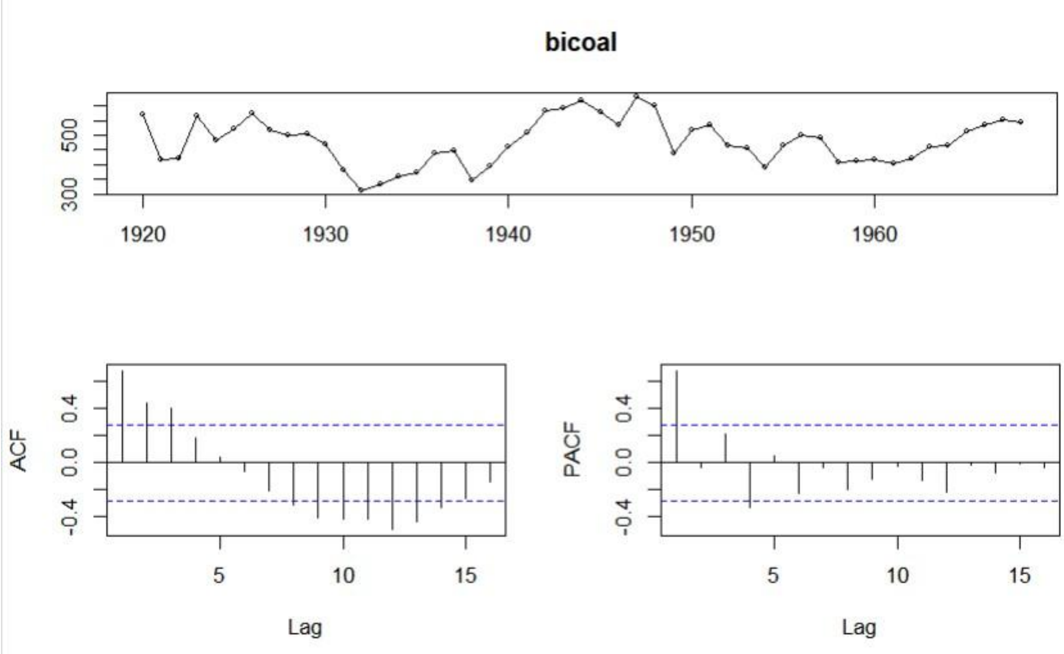
Q3.a Produce a timeplot of the data.



Q3.b

from the equation we can tell it is a (4,0,0) ARIMA

Q3.c Explain why this model is choose by ACF and PACF



As we can see from the above figure,the ACF is exponentially decay and it has a significant spike of PACF at lag 1 and lag 4,but it doesn't meet the requirement of MA(q),so q is 0,and it doesn't need differenced,so i assume the ARIMA is (1,0,0) and (4,0,0).

```
> Arima(bicoal,order=c(1,0,0))
Series: bicoal
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1      mean
      0.6954  484.9036
s.e.  0.1020   25.1873

sigma^2 = 3242:  log likelihood = -266.89
AIC=539.78   AICc=540.32   BIC=545.46
```

```
> checkresiduals(Arima(bicoal,order=c(1,0,0)))

      Ljung-Box test

data:  Residuals from ARIMA(1,0,0) with non-zero mean
Q* = 10.678, df = 9, p-value = 0.2985

Model df: 1.   Total lags used: 10
```

Above are the result of ARIMA(1,0,0)

```
> Arima(bicoal,order=c(4,0,0))
Series: bicoal
ARIMA(4,0,0) with non-zero mean

Coefficients:
         ar1      ar2     ar3      ar4      mean
      0.8334  -0.3443  0.5525  -0.3780  481.5221
s.e.  0.1366   0.1752  0.1733   0.1414   21.0591

sigma^2 = 2795:  log likelihood = -262.05
AIC=536.1   AICc=538.1   BIC=547.45
```

```
> checkresiduals(Arima(bicoal,order=c(4,0,0)))

        Ljung-Box test

data:  Residuals from ARIMA(4,0,0) with non-zero mean
Q* = 4.852, df = 6, p-value = 0.5629

Model df: 4.   Total lags used: 10
```

Above are the result of (4,0,0).

We can see that the p-value of both orders are greater than 0.05,which means the residuals is close to normal distribution.But the AICc for (4,0,0) is lower than (1,0,0),which means (4,0,0) is a better estimation of the ARIMA of bicoal.