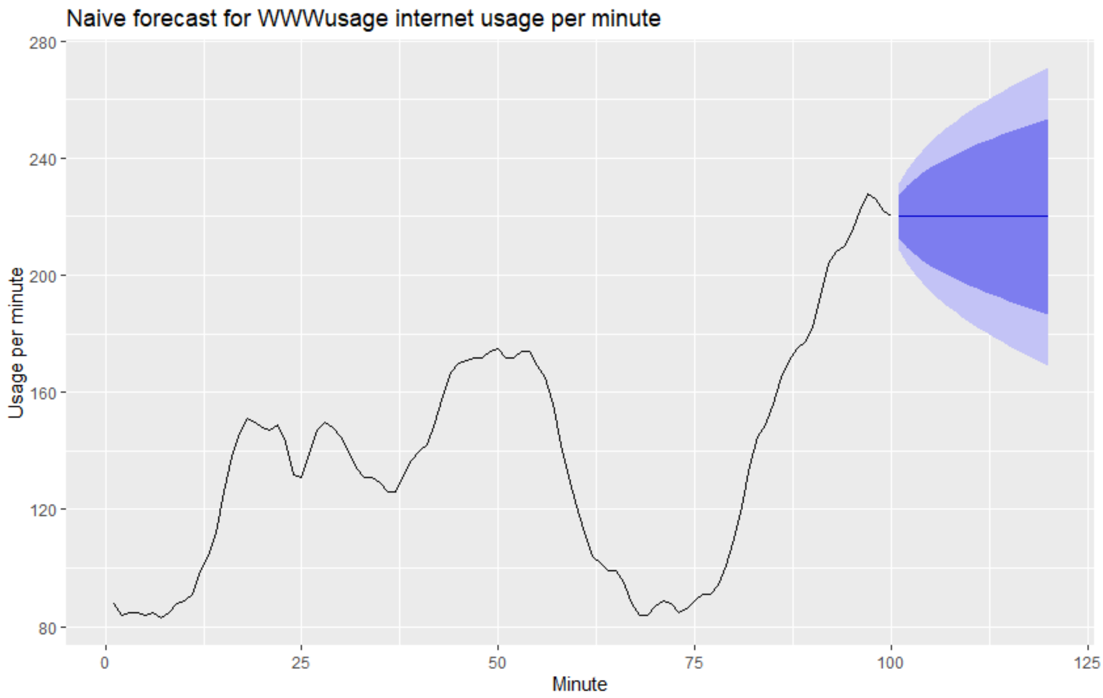
Exercise 1.

Consider the time series WWWusage, which shows internet usage per minute.

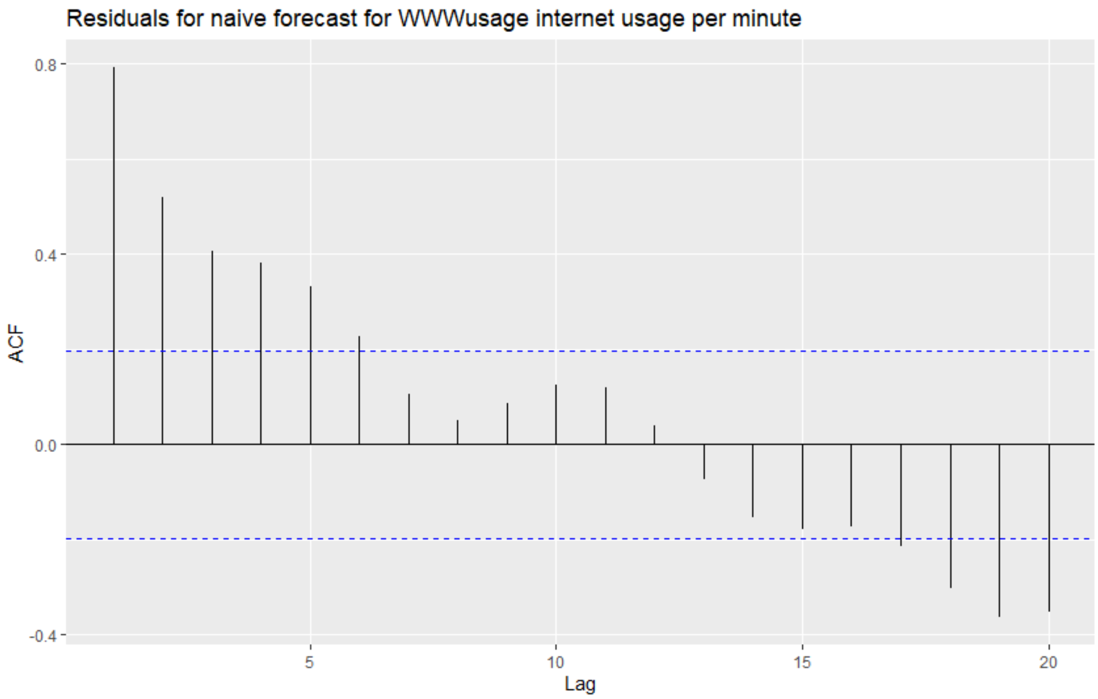
(i) Plot the time series along with a forecast for the next 20 minutes using the na ̈ive method.

```
help(WWWusage)

WWWusage

usagefc <- naive(WWWusage, h=20)

usagefc

autoplot(usagefc)+ggtitle('Naive forecast for WWWusage internet usage per minute')+xlab('Minute')+ylab('Usage per minute')
```



(ii) Plot a correlogram of the residuals.

```
ggAcf(residuals(usagefc))+ggtitle('Residuals for naive forecast for WWWusage internet usage per minute')
```



(iii) Perform a Ljung-Box test with lag 8 on the residuals.

```
usageres <- residuals(usagefc)

Box.test(usageres, type='L', lag=8)

        Box-Ljung test

data:  usageres
X-squared = 142.99, df = 8, p-value < 2.2e-16
```

(iv) Use parts (ii) and (iii) to comment on whether there is autocorrelation or seasonality.

There is significant evidence to reject null hypothesis

H0: there is no autocorrelation with lag 8

No seasonality (not seasonal data)

Exercise 2.

Consider the time series woolyrnq, which shows quarterly production of woollen yarn in Australia.

(i) Plot the time series along with a forecast for the next 3 years using the seasonal na ̈ive method.

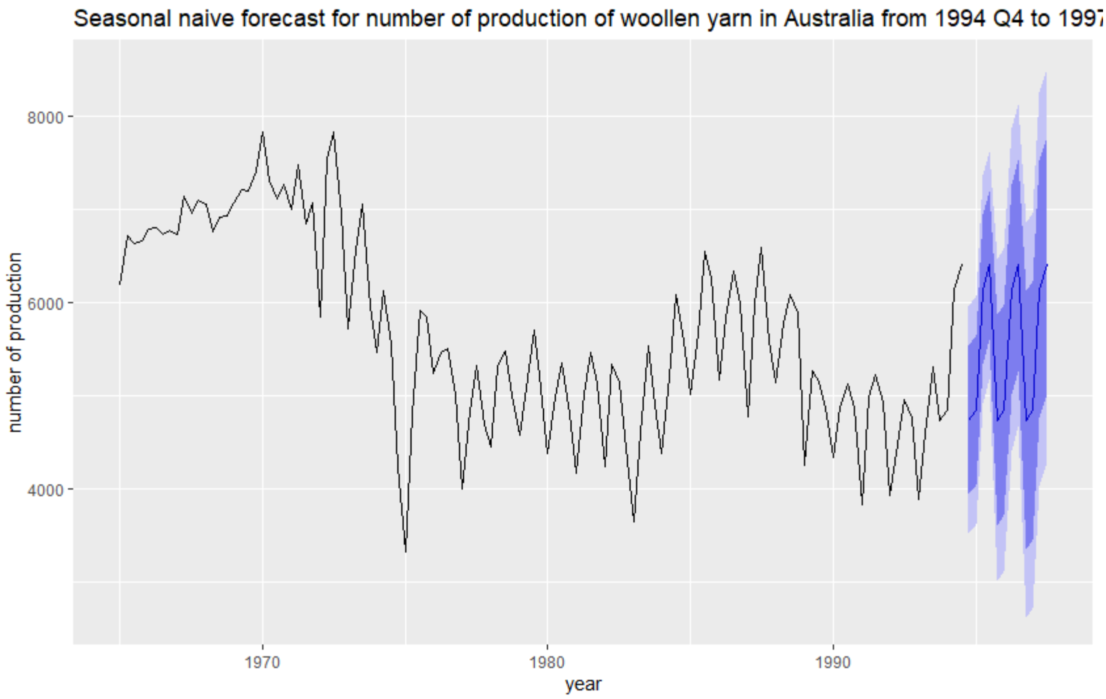
help(woolyrnq)

woolyrnq

woofc <- snaive(woolyrnq, h=12)

woofc

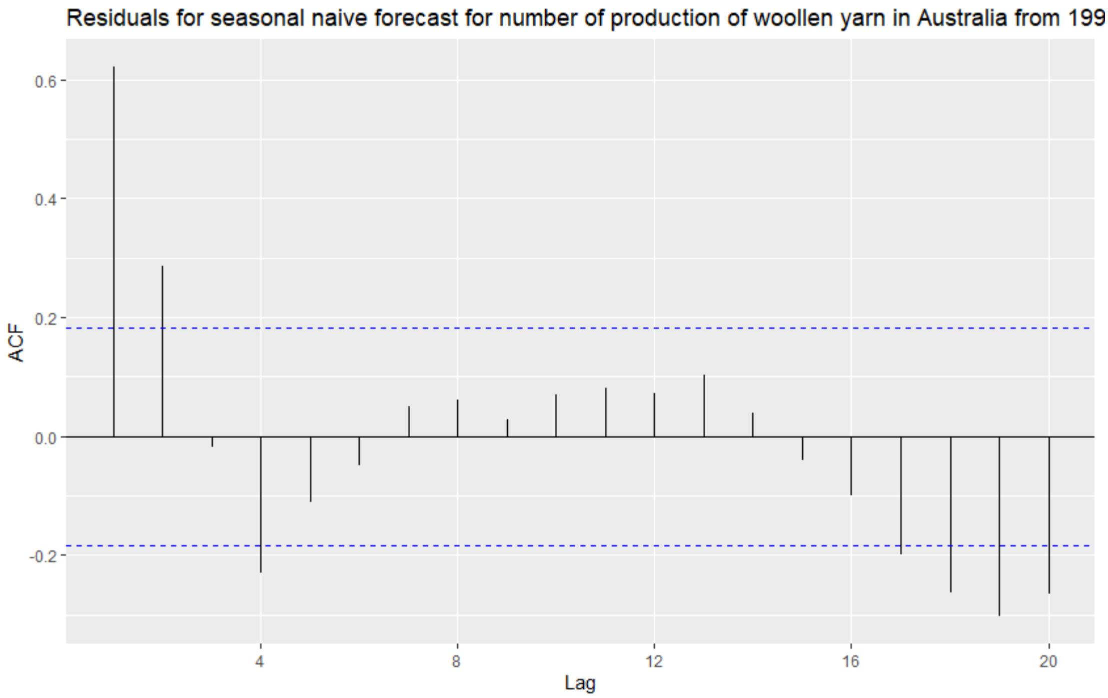
autoplot(woofc)+ggtitle('Seasonal naive forecast for number of production of woollen yarn in Australia from 1994 Q4 to 1997 Q3')+xlab('year')+ylab('number of production')



(ii) Plot a correlogram of the residuals.

```
woores <- residuals(woofc)
```

```
ggAcf(woores)+ggtitle('Residuals for seasonal naive forecast for number of production of woollen yarn in Australia from 1994 Q4 to 1997 Q3')
```



(iii) Perform a Ljung-Box test with lag 8 on the residuals.

```
Box.test(woores, type='L', lag=8)
```

        Box-Ljung test

data: woores

X-squared = 64.299, df = 8, p-value = 6.641e-11

(iv) Use parts (ii) and (iii) to comment on whether there is autocorrelation or seasonality.

There is significant evidence to reject the null hypothesis

H0: there is no autocorrelation up to lag 8

i.e. we have reason to believe there is autocorrelation. There is also strong seasonality.

Exercise 3.

Consider the time series ibmclose, which shows the closing IBM stock price on a number of successive days.

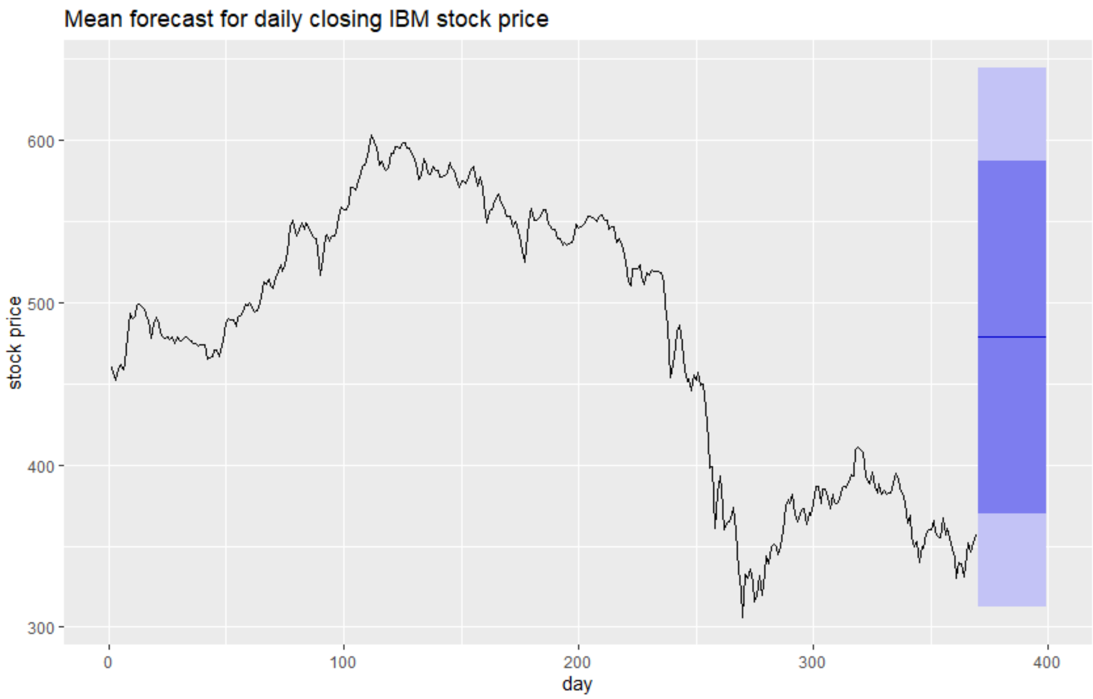
(i) Plot the time series along with a forecast for the next 30 days using the mean method.

```
help(ibmclose)
```

```
ibmfc <- meanf(ibmclose, h=30)
```

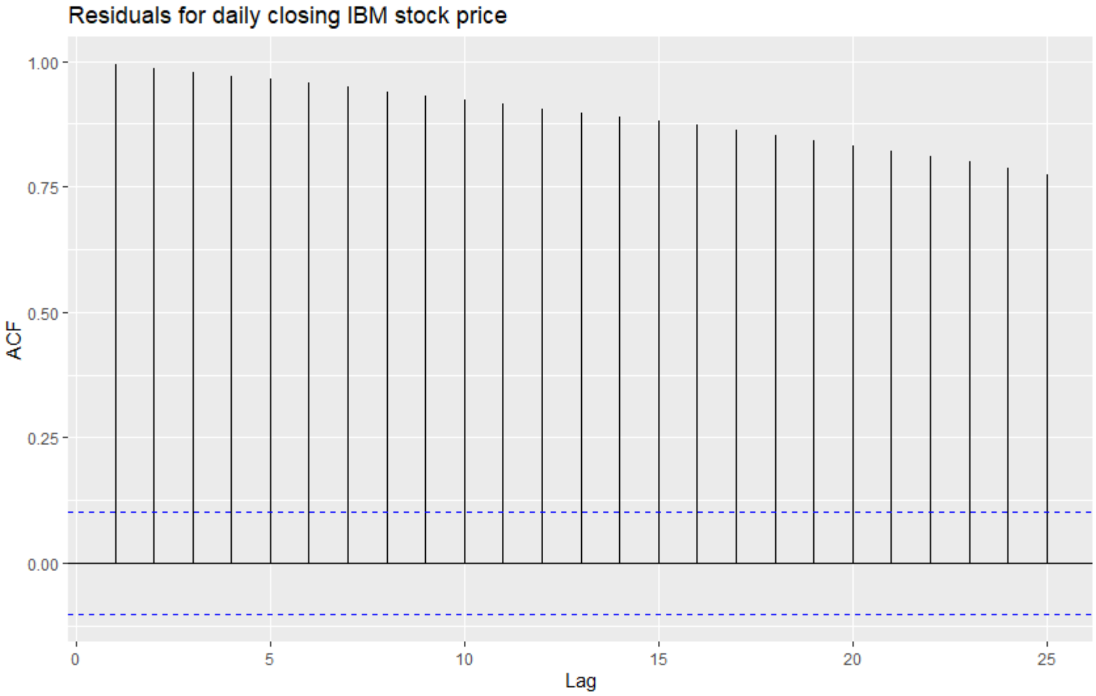
```
ibmfc
```

```
autoplot(ibmfc)+ggtitle('Mean forecast for daily closing IBM stock price')+xlab('day')+ylab('stock price')
```



(ii) Plot a correlogram of the residuals.

```
ibmres <- residuals(ibmfc)
ggAcf(ibmres)+ggtitle('Residuals for daily closing IBM stock price')
```



(iii) Perform a Ljung-Box test with lag 8 on the residuals.

```
Box.test(ibmres, type='L', lag=8)
```

```
	Box-Ljung test

data:  ibmres
X-squared = 2809.9, df = 8, p-value < 2.2e-16
```

(iv) Use parts (ii) and (iii) to comment on whether there is autocorrelation or seasonality.

There is significant evidence to reject the null hypothesis

Ho: there is autocorrelation up to lag 8

No seasonality

(v) What might be a more appropriate method of forecast for this time series?

ARIMA Model

The model selection for stock price data is typically more complex because they often exhibit intricate trends and periodicity, and the ARIMA model could be a good choice.

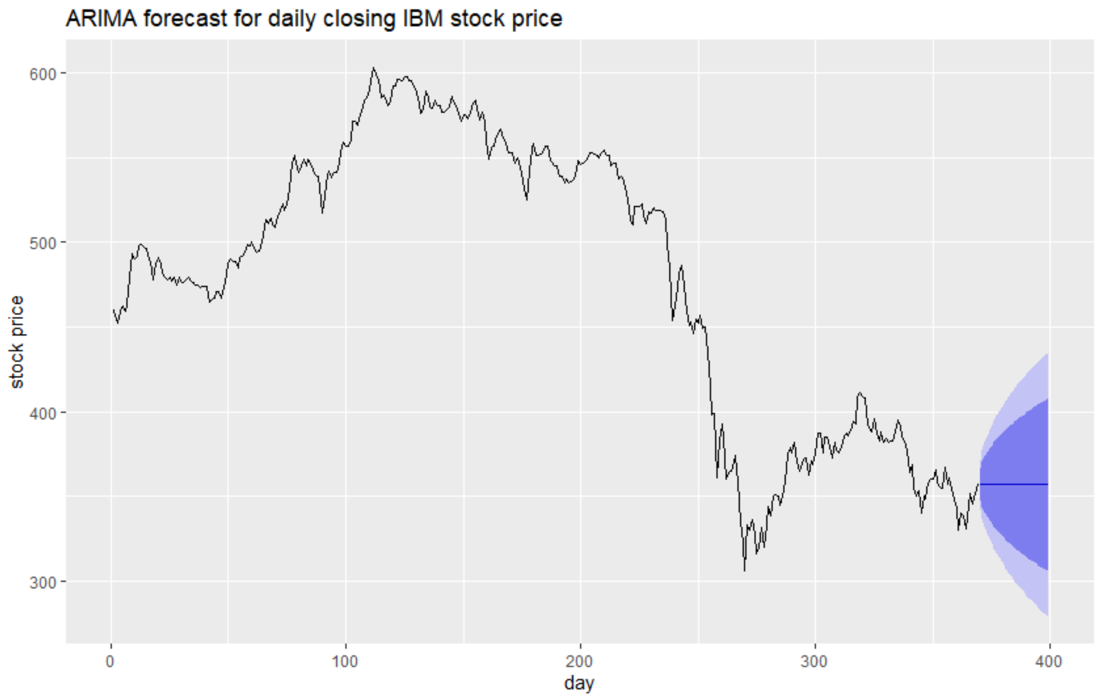
(vi) Implement your chosen method from part (v) and investigate whether it is indeed a better choice than the mean method.

```
arima_model <- auto.arima(ibmclose)
```

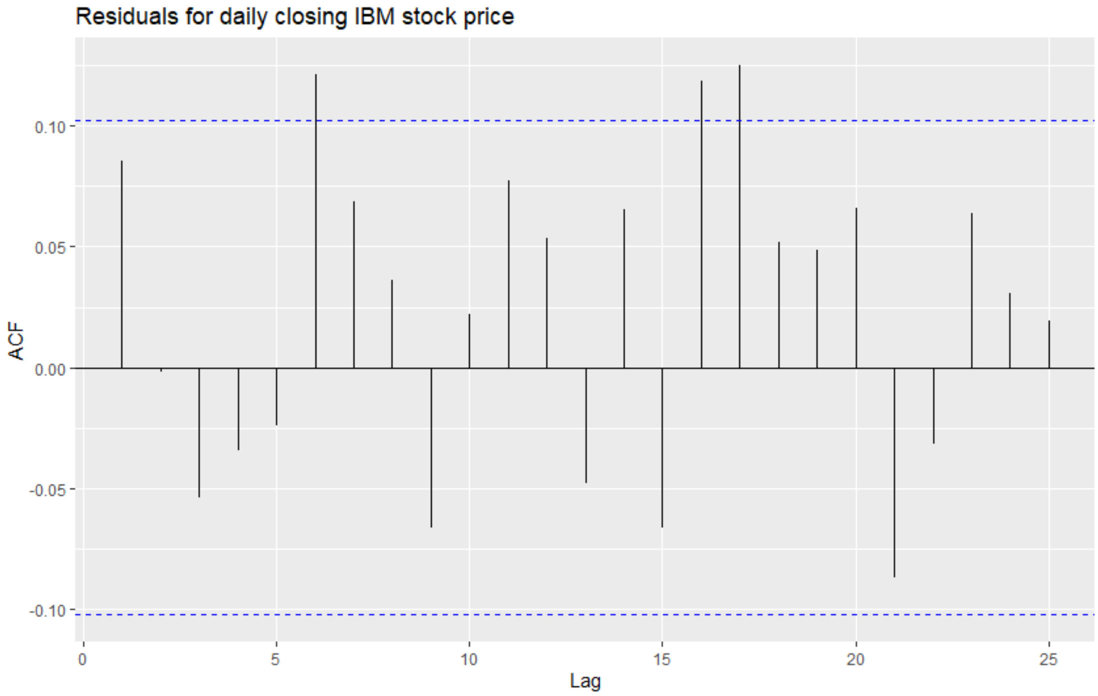
```
ibmfc2 <- forecast(arima_model, h=30)
```

```
ibmfc2
```

```
autoplot(ibmfc2)+ggtitle('ARIMA forecast for daily closing IBM stock
price')+xlab('day')+ylab('stock price')
```



```
ibmres2 <- residuals(ibmfc2)
ggAcf(ibmres2)+ggtitle('Residuals for daily closing IBM stock price')
```



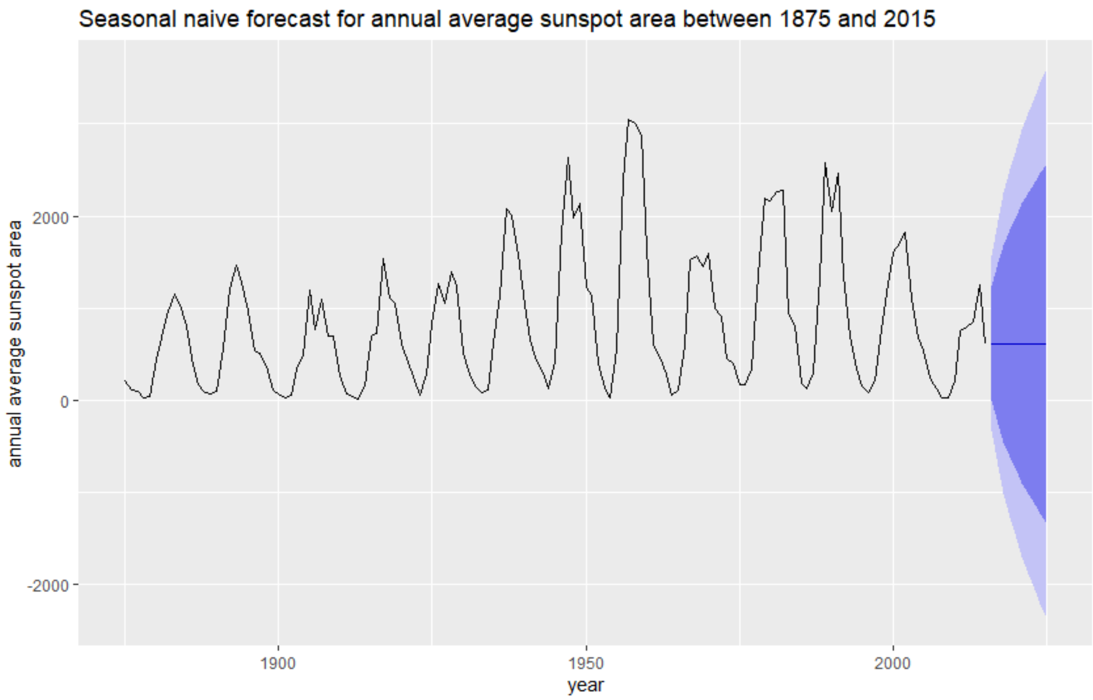

```
Box.test(ibmres2, type='L', lag=8)
```

```
        Box-Ljung test

data:  ibmres2
X-squared = 12.291, df = 8, p-value = 0.1387
```

Exercise 4.

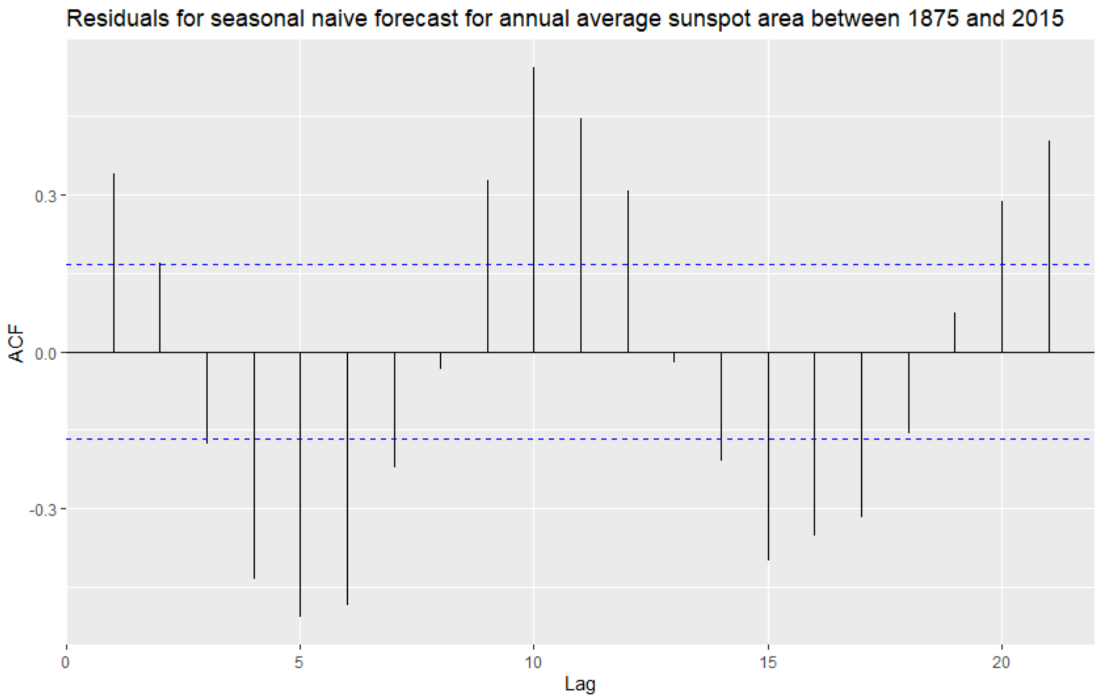
Consider the time series sunspotarea, which shows the annual average sunspot area between 1875 and 2015. Investigate whether the seasonal na ̈ive method is a good way to forecast this series.

```
help(sunspotarea)

sunspotarea

sunfc <- snaive(sunspotarea, h=10)

sunfc

autoplot(sunfc)+ggtitle('Seasonal naive forecast for annual average sunspot area between 1875 and 2015')+xlab('year')+ylab('annual average sunspot area')
```



sunres <- residuals(sunfc)

ggAcf(sunres)+ggtitle('Residuals for seasonal naive forecast for annual average sunspot area between 1875 and 2015')



Box.test(sunres, type='L', lag=8)

Box-Ljung test

data: sunres

X-squared = 133.13, df = 8, p-value < 2.2e-16

seasonal naive method is not a good way to forecast sunspotarea series

There is sufficient evidence to reject the null hypothesis

H0: there is no autocorrelation up to lag 8

Exercise 5.

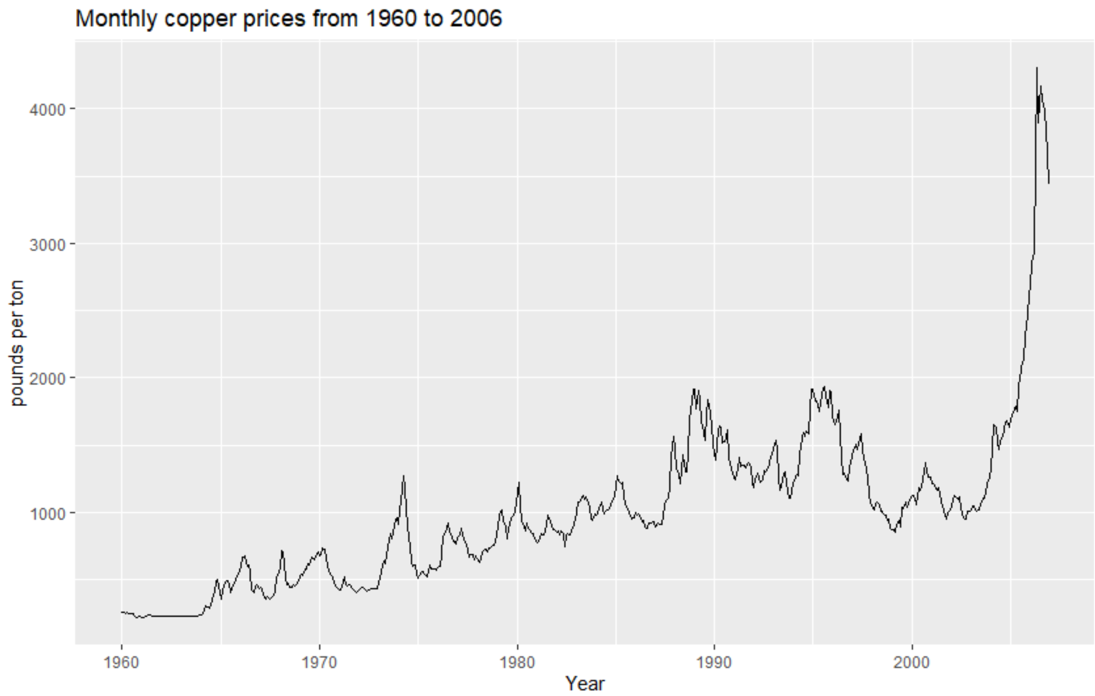
Consider the time series mcopper, which shows the price of copper.

(i) Plot the time series.

```
help(mcopper)
```

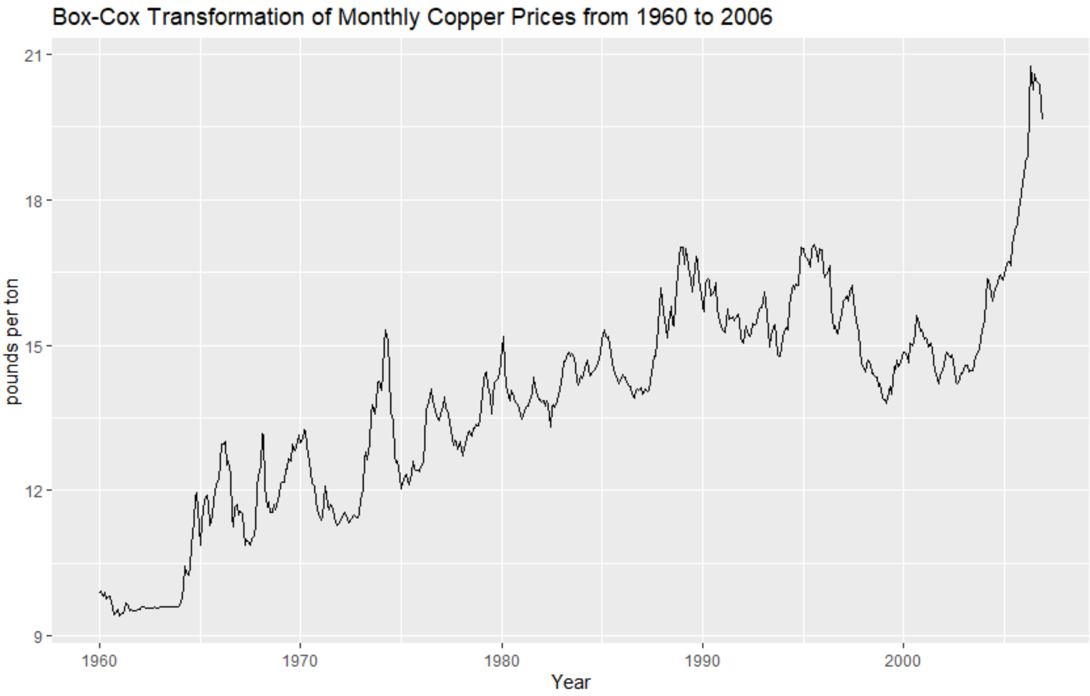
```
mcopper
```

```
autoplot(mcopper)+ggtitle('Monthly copper prices from 1960 to 2006')+xlab('Year')+ylab('pounds per ton')
```



(ii) Perform an appropriate Box Cox transformation.

```
lambda <- BoxCox.lambda(mcopper)

lambda

autoplot(BoxCox(mcopper,lambda))+ggtitle('Box-Cox Transformation of Monthly Copper Prices from 1960 to 2006')+xlab('Year')+ylab('pounds per ton')
```

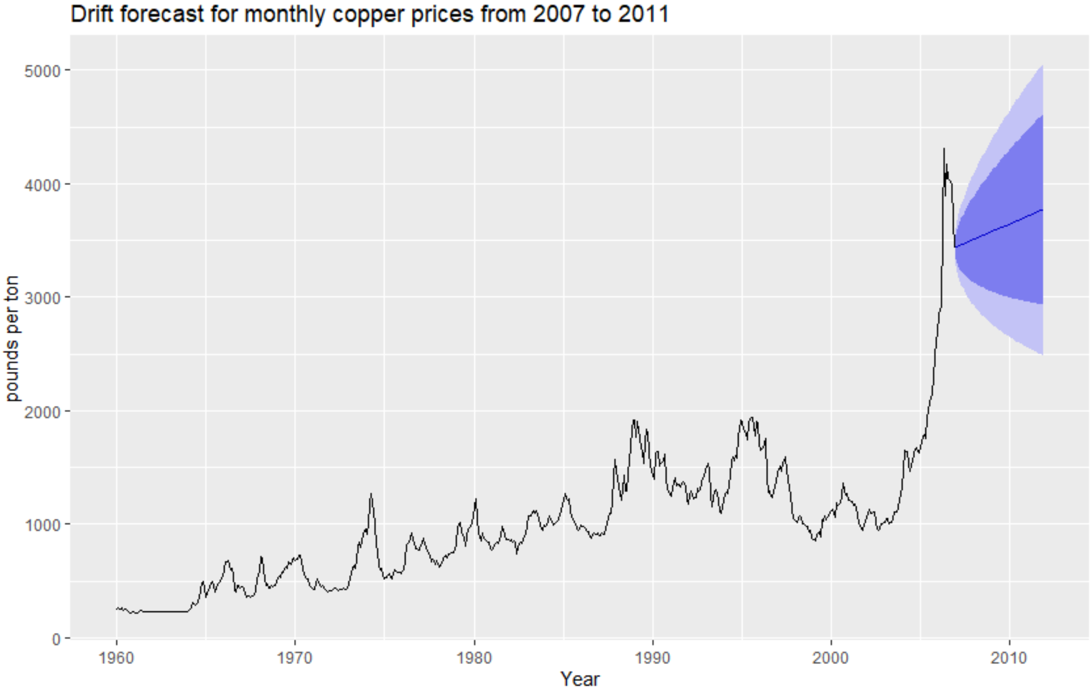


(iii) Plot the transformed data along with a five year forecast using the drift method.

```
mcopperfc <- rwf(mcopper, h=60, drift=TRUE)
```

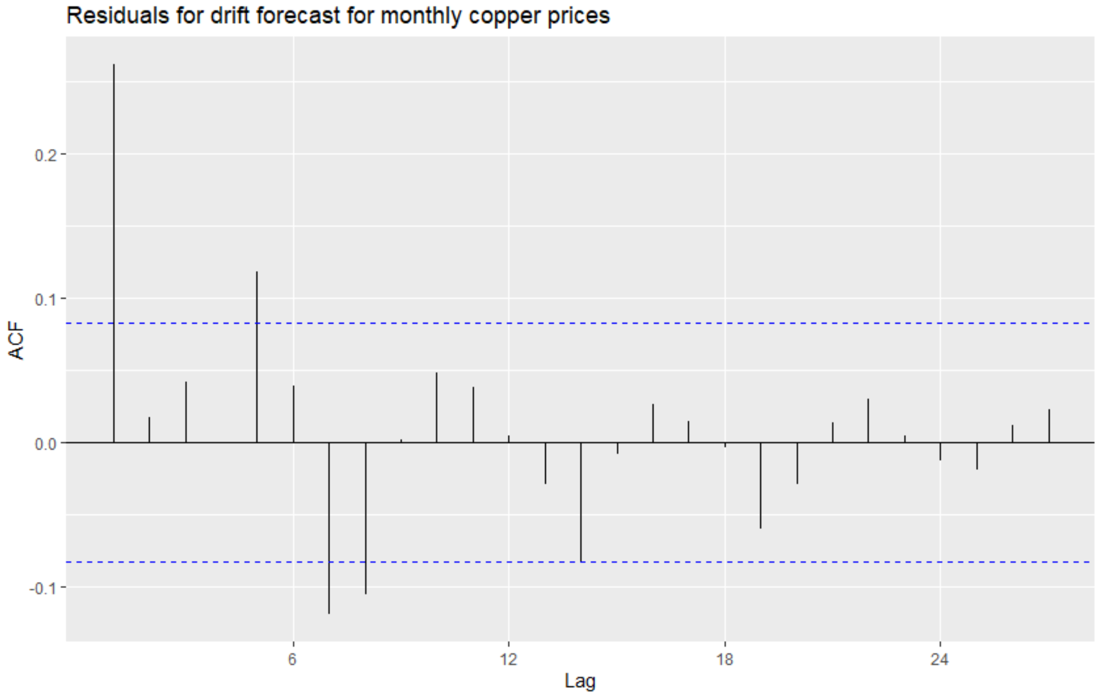
```
mcopperfc
```

```
autoplot(mcopperfc)+ggtitle("Drift forecast for monthly copper prices from 2007 to 2011")+xlab("Year")+ylab("pounds per ton")
```



(iv) Plot a correlogram of the residuals.

```
mcopperres <- residuals(mcopperfc)
ggAcf(mcopperres)+ggtitle('Residuals for drift forecast for monthly copper prices')
```



(v) Perform a Ljung-Box test with lag 8 on the residuals.

```
Box.test(mcopperres, type="L", lag=8)
```

Box-Ljung test

data: mcopperres

X-squared = 63.469, df = 8, p-value = 9.684e-11

(vi) Use parts (iv) and (v) to comment on whether there is autocorrelation or seasonality.

There is significant evidence to reject the null hypothesis

H0: there is no autocorrelation up to lag 8

However, there is no evidence of seasonality

Exercise 6.

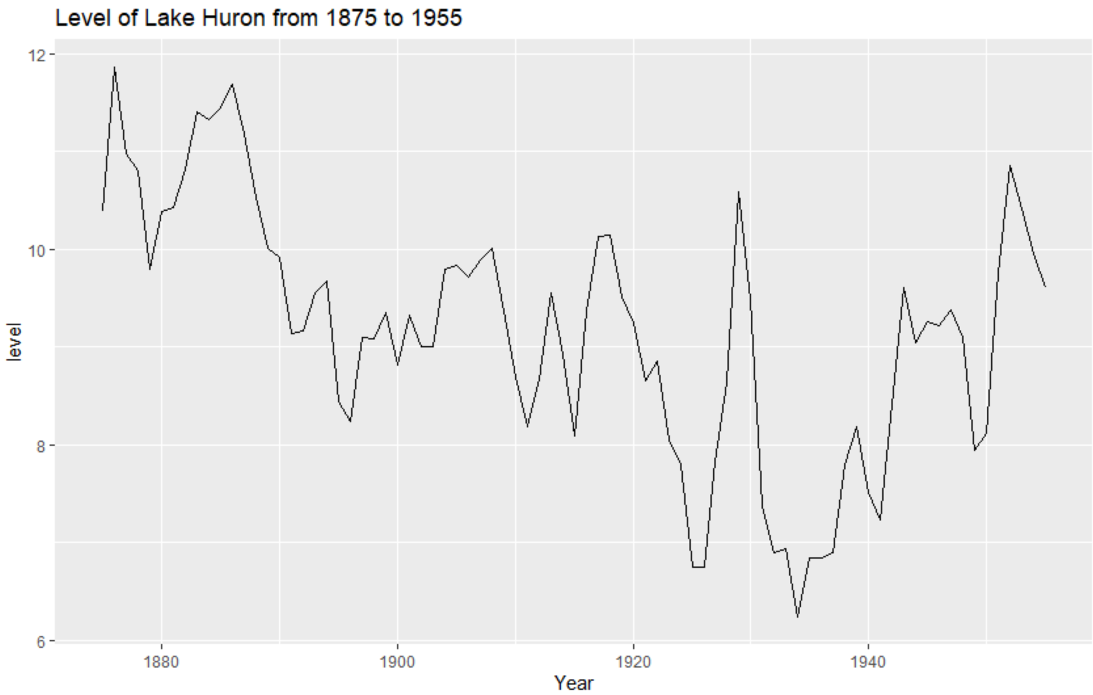
Consider the time series huron, which shows the level of Lake Huron between 1875 and 1972.

(i) Use the window function to split the time series into a training set consisting of the data from 1875 to 1955 and the test set consisting of the data from 1956 to 1972.

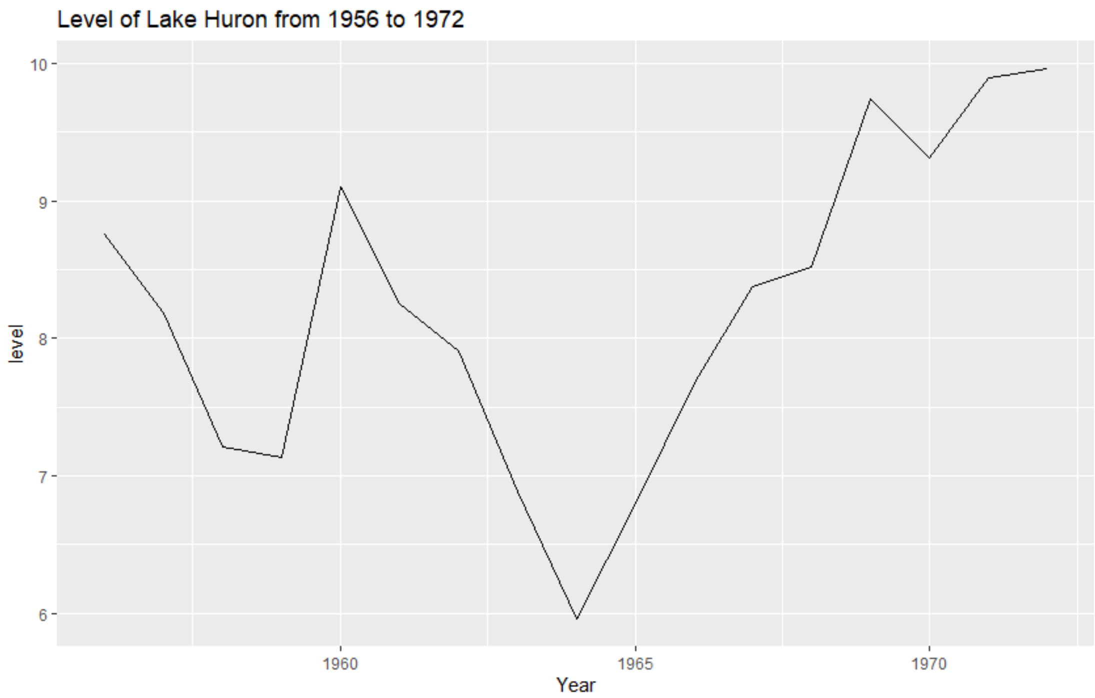
```
help(huron)

huron

huron1 <- window(huron, end=c(1955, 1))

huron2 <- window(huron, start=c(1956, 1))

huron1

huron2
```

(ii) Check you have correctly split up the data by producing two separate plots.

```
autoplot(huron1)+ggtitle("Level of Lake Huron from 1875 to
1955")+xlab("Year")+ylab("level")
```



```
autoplot(huron2)+ggtitle("Level of Lake Huron from 1956 to 1972")+xlab("Year")+ylab("level")
```



(iii) Explain why the seasonal na¨ive forecast is not appropriate here.

Because the time series of lake level heights has a trend, and seasonal naive forecasting may produce inaccurate predictions for time series with trends or other non-seasonal patterns.

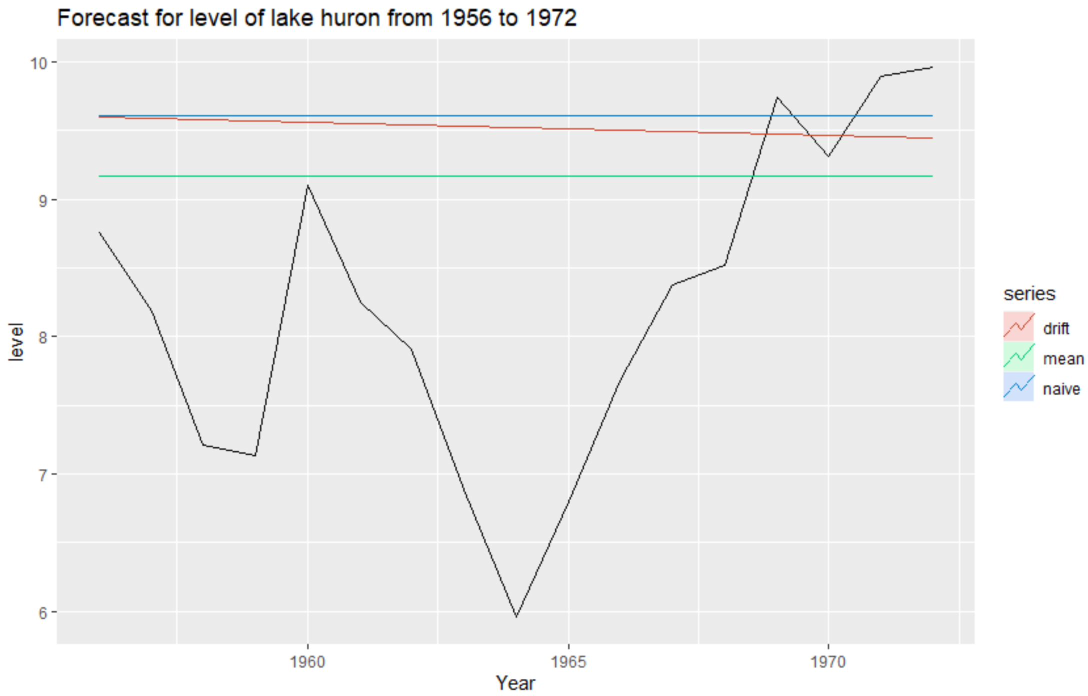
(iv) Use the training set to produce three forecasts for the period 1956 to 1972, using the na¨ive method, the drift method and the mean method. Plot each forecast alongside the test data.

```
huronfc1 <- naive(huron1, 17)
```

```
huronfc2 <- rwf(huron1, h=17, drift=TRUE)
```

```
huronfc3 <- meanf(huron1, 17)
```

```
autoplot(huron2)+
 autolayer(huronfc1, series = 'naive', PI=FALSE)+
 autolayer(huronfc2, series = 'drift', PI=FALSE)+
 autolayer(huronfc3, series = 'mean', PI=FALSE)+
 ggtitle('Forecast for level of lake huron from 1956 to 1972')+
 xlab('Year')+ylab('level')
```




```
accuracy(huronfc1, huron2)
accuracy(huronfc2, huron2)
accuracy(huronfc3, huron2)
```

```
> accuracy(huronfc1, huron2)
                    ME      RMSE      MAE         MPE      MAPE    MASE      ACF1 Theil's U
Training set -0.009625 0.7178988 0.554375  -0.4238952  6.188552 1.00000 0.1407863        NA
Test set     -1.394118 1.7991534 1.483529 -19.3542423 20.257760 2.67604 0.6532090  2.225255
> accuracy(huronfc2, huron2)
                        ME      RMSE      MAE         MPE      MAPE     MASE      ACF1 Theil's U
Training set -2.442430e-16 0.7178343 0.554375  -0.3165107  6.186713 1.000000 0.1407863        NA
Test set     -1.307493e+00 1.7480878 1.450125 -18.3123389 19.755083 2.615784 0.6619984  2.159213
> accuracy(huronfc3, huron2)
                        ME     RMSE      MAE        MPE     MAPE     MASE     ACF1 Theil's U
Training set -1.867803e-16 1.285300 1.016327  -2.130276 11.69820 1.833284 0.839739        NA
Test set     -9.536238e-01 1.484179 1.215033 -13.883399 16.54082 2.191716 0.653209  1.848391
```

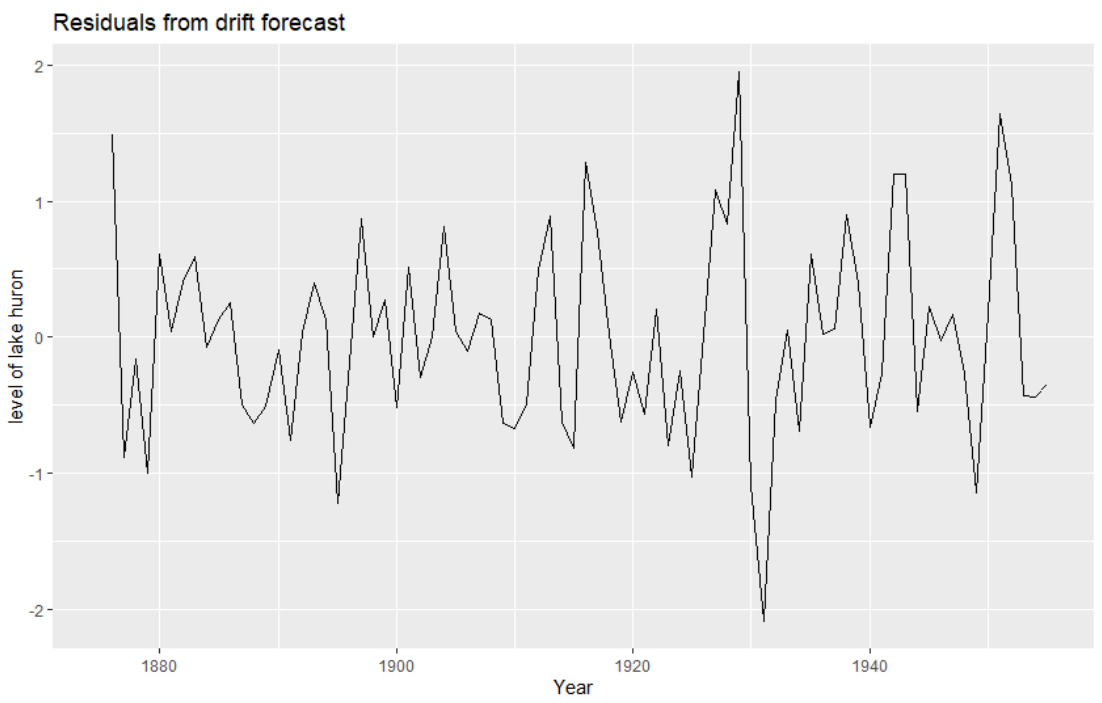
(v) Compare the accuracy of each of the forecasts you calculated in part (iv) by comparing their errors. Which forecast performs the best?

drift method

(vi) For the best performing method, compute the residuals and plot them.

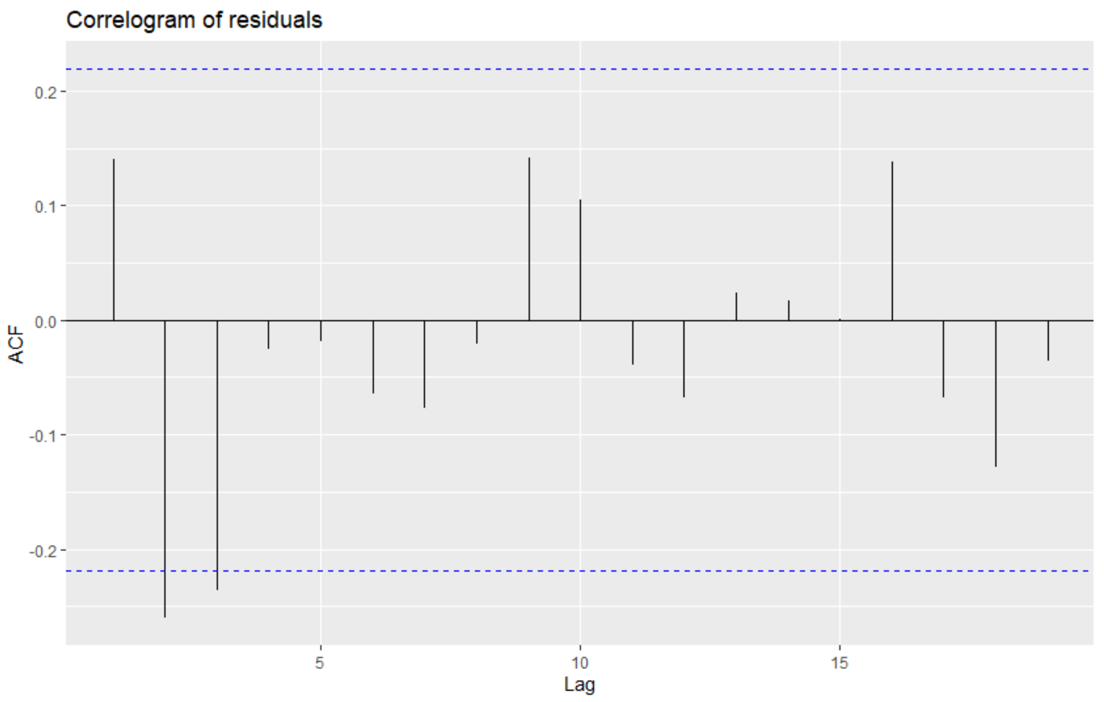
```
huronres <- residuals(huronfc2)
```

```
autoplot(huronres)+ggtitle('Residuals from drift forecast')+xlab('Year')+ylab('level of lake huron')
```



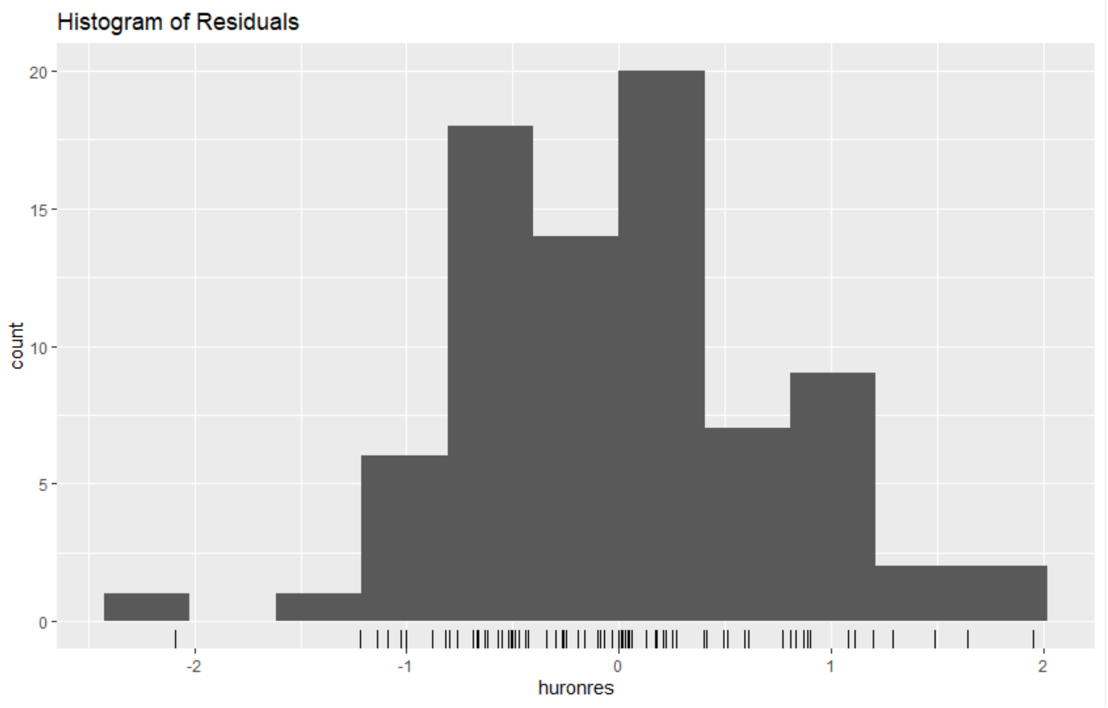
(vii) Do the residuals appear to be uncorrelated and normally distributed?

```
# yes: uncorrelated
# yes: normally distributed
ggAcf(huronres)+ggtitle('Correlogram of residuals')
```

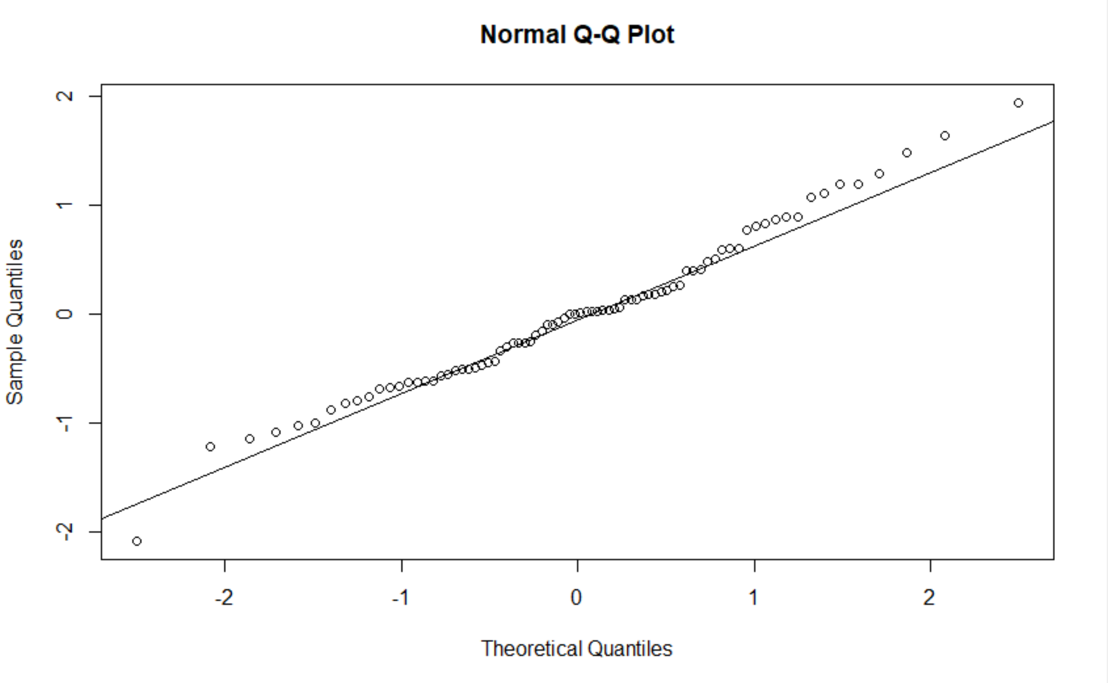



```
Box.test(huronres, type="L", lag=8)

	Box-Ljung test

data:  huronres
X-squared = 13.073, df = 8, p-value = 0.1094
```

```
gghistogram(huronres)+ggtitle('Histogram of Residuals')
```




```
qqnorm(huronres)
qqline(huronres)
```



(viii) In light of your answer to part (vii), would you like to change your choice of forecast? If so, check the residuals of your second choice.

I wouldn't change drift forecast

checkresiduals(huronres)