Time Series Week 1 Assignment

Vishal Arora & Samriti Malhotra

June 6, 2020

library(fpp2)

## Warning: package 'fpp2' was built under R version 3.5.3

## Loading required package: ggplot2

## Loading required package: forecast

## Warning: package 'forecast' was built under R version 3.5.3

## Loading required package: fma

## Warning: package 'fma' was built under R version 3.5.3

## Loading required package: expsmooth

## Warning: package 'expsmooth' was built under R version 3.5.3

library(kableExtra)

## Warning: package 'kableExtra' was built under R version 3.5.3

## Question 2.1

Use the help function to explore what the series gold, woolyrnq and gas represent.

Using help function to know more about the various datasets like a short description about the dataset,format of the dataset which in this case is Time Series data for all three datasets, what is Source of the data and example of how to dsiplay to learn about data using tsdisplay function.

help(gold)

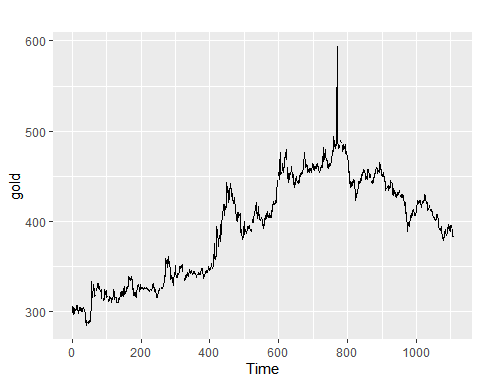
## starting httpd help server ... done

help("woolyrnq")  
help("gas")

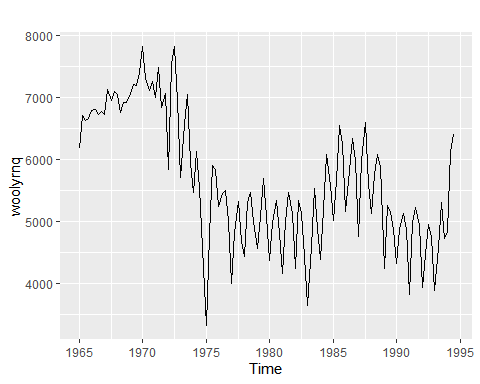
Use autoplot() to plot each of these in separate plots.

Using autoplot function can give us a graphic diaply of the ts datasets ,looking at graphs we can see if there is any trend , seanality or cyclic patterns in the datasets which can help us in selecting our model for prediction.

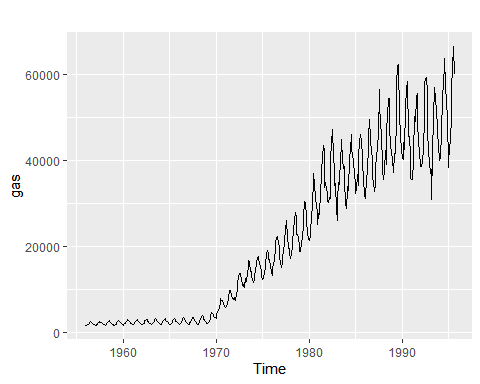
autoplot(gold)



autoplot(woolyrnq)



autoplot(gas)



What is the frequency of each series? Hint: apply the frequency() function.  
Using frequency function we find that Gold has Yearly frequency , Woolrnq has Quarterly frequency and Gas has monthly freqnency.

frequency(gold)

## [1] 1

frequency(woolyrnq)

## [1] 4

frequency(gas)

## [1] 12

Use which.max() to spot the outlier in the gold series. Which observation was it?

Using the which.max() function gives us the outlier with maximum value as also seen from the graph above, which is 770.

which.max(gold)

## [1] 770

## Question 2.3 :-

Download some monthly Australian retail data from the book website. These represent retail sales in various categories for different Australian states, and are stored in a MS-Excel file.

### a) Read the data from excel file

retaildata <- readxl::read\_excel("retail.xlsx", skip=1)  
  
head(retaildata)

## # A tibble: 6 x 190  
## `Series ID` A3349335T A3349627V A3349338X A3349398A A3349468W  
## <dttm> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1982-04-01 00:00:00 303. 41.7 63.9 409. 65.8  
## 2 1982-05-01 00:00:00 298. 43.1 64 405. 65.8  
## 3 1982-06-01 00:00:00 298 40.3 62.7 401 62.3  
## 4 1982-07-01 00:00:00 308. 40.9 65.6 414. 68.2  
## 5 1982-08-01 00:00:00 299. 42.1 62.6 404. 66   
## 6 1982-09-01 00:00:00 305. 42 64.4 412. 62.3  
## # ... with 184 more variables: A3349336V <dbl>, A3349337W <dbl>,  
## # A3349397X <dbl>, A3349399C <dbl>, A3349874C <dbl>, A3349871W <dbl>,  
## # A3349790V <dbl>, A3349556W <dbl>, A3349791W <dbl>, A3349401C <dbl>,  
## # A3349873A <dbl>, A3349872X <dbl>, A3349709X <dbl>, A3349792X <dbl>,  
## # A3349789K <dbl>, A3349555V <dbl>, A3349565X <dbl>, A3349414R <dbl>,  
## # A3349799R <dbl>, A3349642T <dbl>, A3349413L <dbl>, A3349564W <dbl>,  
## # A3349416V <dbl>, A3349643V <dbl>, A3349483V <dbl>, A3349722T <dbl>,  
## # A3349727C <dbl>, A3349641R <dbl>, A3349639C <dbl>, A3349415T <dbl>,  
## # A3349349F <dbl>, A3349563V <dbl>, A3349350R <dbl>, A3349640L <dbl>,  
## # A3349566A <dbl>, A3349417W <dbl>, A3349352V <dbl>, A3349882C <dbl>,  
## # A3349561R <dbl>, A3349883F <dbl>, A3349721R <dbl>, A3349478A <dbl>,  
## # A3349637X <dbl>, A3349479C <dbl>, A3349797K <dbl>, A3349477X <dbl>,  
## # A3349719C <dbl>, A3349884J <dbl>, A3349562T <dbl>, A3349348C <dbl>,  
## # A3349480L <dbl>, A3349476W <dbl>, A3349881A <dbl>, A3349410F <dbl>,  
## # A3349481R <dbl>, A3349718A <dbl>, A3349411J <dbl>, A3349638A <dbl>,  
## # A3349654A <dbl>, A3349499L <dbl>, A3349902A <dbl>, A3349432V <dbl>,  
## # A3349656F <dbl>, A3349361W <dbl>, A3349501L <dbl>, A3349503T <dbl>,  
## # A3349360V <dbl>, A3349903C <dbl>, A3349905J <dbl>, A3349658K <dbl>,  
## # A3349575C <dbl>, A3349428C <dbl>, A3349500K <dbl>, A3349577J <dbl>,  
## # A3349433W <dbl>, A3349576F <dbl>, A3349574A <dbl>, A3349816F <dbl>,  
## # A3349815C <dbl>, A3349744F <dbl>, A3349823C <dbl>, A3349508C <dbl>,  
## # A3349742A <dbl>, A3349661X <dbl>, A3349660W <dbl>, A3349909T <dbl>,  
## # A3349824F <dbl>, A3349507A <dbl>, A3349580W <dbl>, A3349825J <dbl>,  
## # A3349434X <dbl>, A3349822A <dbl>, A3349821X <dbl>, A3349581X <dbl>,  
## # A3349908R <dbl>, A3349743C <dbl>, A3349910A <dbl>, A3349435A <dbl>,  
## # A3349365F <dbl>, A3349746K <dbl>, ...

#kable(head(retaildata)) %>%   
# kable\_styling(bootstrap\_options = c("striped","hover","condensed","responsive"),full\_width = F,position = "left",font\_size = 12) %>%row\_spec(0, background ="gray")

### b) Select one of the time series as follows (but replace the column name with your own chosen column):

# choosing column 'A3349721R'  
myts <- ts(retaildata[,"A3349721R"],frequency=12, start=c(1982,1), end=c(2005,6))  
head(myts)

## Jan Feb Mar Apr May Jun  
## 1982 161.8 158.7 166.6 172.9 165.9 169.5

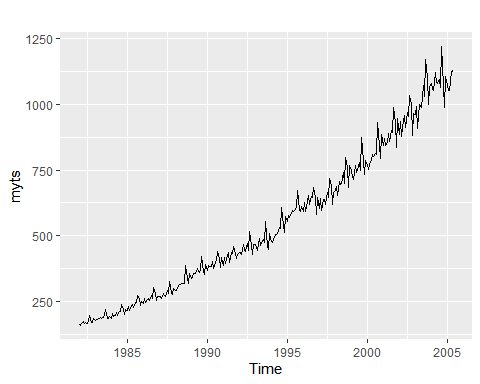
tail(myts,6)

## Jan Feb Mar Apr May Jun  
## 2005 1076.2 1053.5 1053.7 1126.6 1129.4 1122.8

### c) Explore your chosen retail time series using the following functions:autoplot(), ggseasonplot(), ggsubseriesplot(), gglagplot(), ggAcf().Can you spot any seasonality, cyclicity and trend? What do you learn about the series?

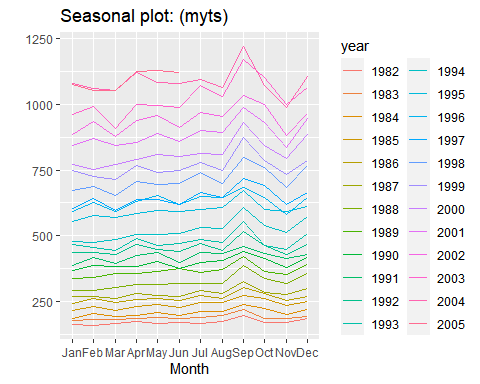
Using autoplot function on ts dataset , we can see that there is indeed a trend and seasonality in the graph.

autoplot(myts)



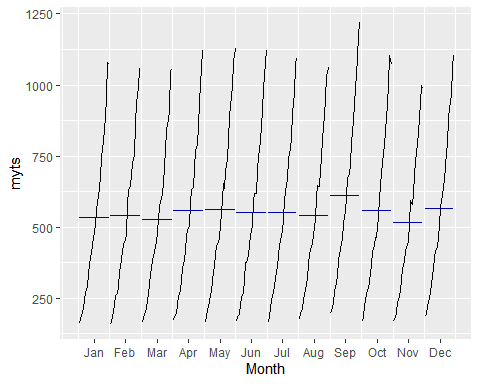
Using the seasonplot , one can confirm what we deduced from the above plot that there is slight seasonlity pattern also in the data. Thus in sept is the peak and march is the trough season.

ggseasonplot((myts))



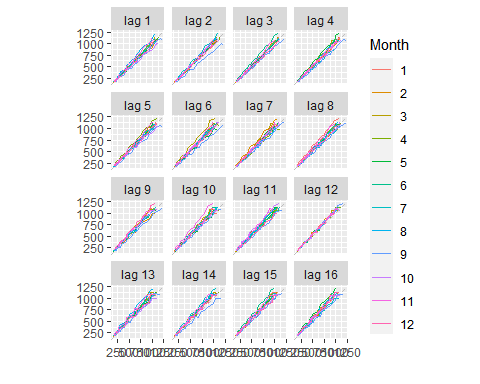
Using subseriesplot we can see the seasonality for each month for time span of time series.We can clearly see that september month is the peak and march and November are the troughs.

ggsubseriesplot(myts)

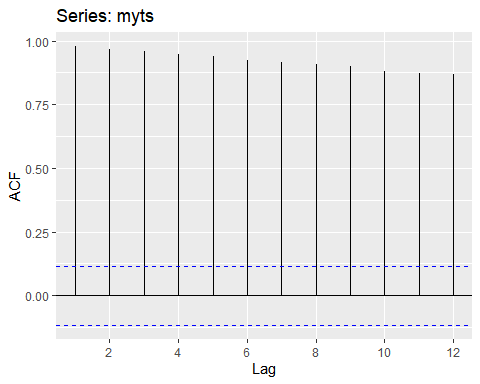


Lag plot show the scatterplots for each month and all the graphs show linearity thus it suggests a strong autocorrelations exits.

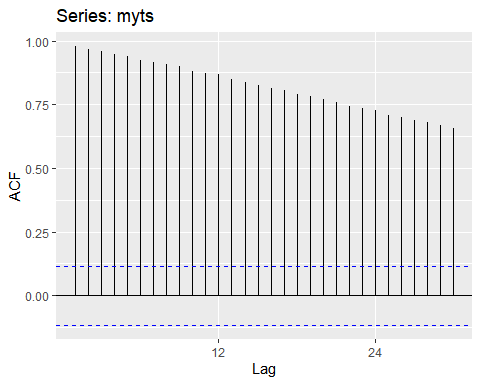
gglagplot(myts)

 Using ggACF function to plot ACF graphs with differrent lag periods, It shows there exists a strong autocorrelations and also we can see with the increase in lag period the graphs show decrease positive values. With increase lag period in second graph we can there is a slight season pattern also.

ggAcf(myts, lag=12)



ggAcf(myts, lag=30)

 ## Summary

As we can clearly see that out data in second questions has a positive trend with a slight seasonality with no cyclic patterns. This can be corroborated by plotting using the decompose function and then autoplotting .

decmyts <-decompose(myts)  
autoplot(decmyts, type="multiplicative")

