Data624\_HW1

Violet Stoyanova

September 1, 2019

#loading the needed libraries   
library(tidyr)  
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.5.2

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(knitr)  
library(utils)  
library(ggplot2)  
library(forecast)

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

library(readxl)

## Warning: package 'readxl' was built under R version 3.5.2

library(fpp2)

## Warning: package 'fpp2' 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

#cleaning the environment  
rm(list=ls())

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

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

1. What is the frequency of each series? Hint: apply the frequency() function.

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

#help(gold)  
  
head(gold)

## Time Series:  
## Start = 1   
## End = 6   
## Frequency = 1   
## [1] 306.25 299.50 303.45 296.75 304.40 298.35

#Daily morning gold prices in US dollars. 1 January 1985 – 31 March 1989.

#help(woolyrnq)  
head(woolyrnq)

## Qtr1 Qtr2 Qtr3 Qtr4  
## 1965 6172 6709 6633 6660  
## 1966 6786 6800

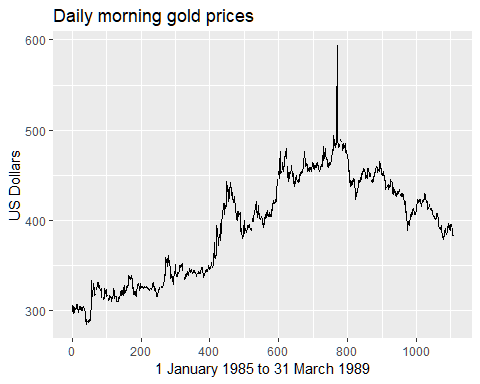
#Quarterly production of woollen yarn in Australia: tonnes. Mar 1965 – Sep 1994.

#help(gas)  
head(gas)

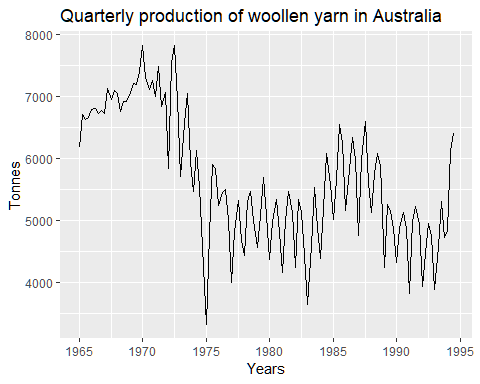
## Jan Feb Mar Apr May Jun  
## 1956 1709 1646 1794 1878 2173 2321

#Australian monthly gas production: 1956–1995.

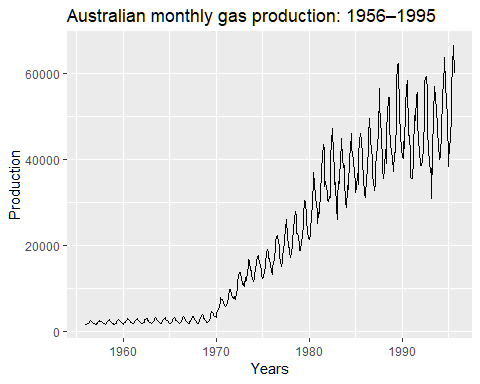
autoplot(gold) +  
 ggtitle("Daily morning gold prices") +  
 xlab("1 January 1985 to 31 March 1989") +  
 ylab("US Dollars")



autoplot(woolyrnq) +  
 ggtitle("Quarterly production of woollen yarn in Australia") +  
 xlab("Years") +  
 ylab("Tonnes")



autoplot(gas) +  
 ggtitle("Australian monthly gas production: 1956–1995") +  
 xlab("Years") +  
 ylab("Production")



frequency(gold)

## [1] 1

frequency(woolyrnq)

## [1] 4

frequency(gas)

## [1] 12

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.

Can you spot any seasonality, cyclicity and trend? What do you learn about the series?

#getting the working directory  
getwd()

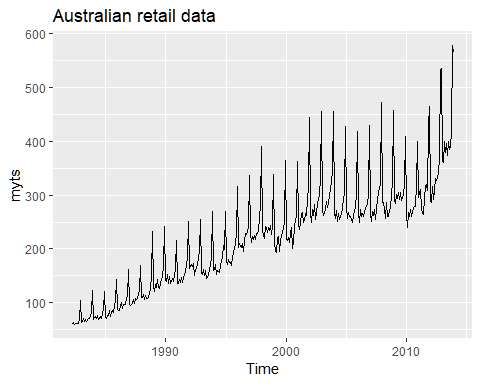
## [1] "C:/Users/Violet/Documents"

#install.packages(ts)  
  
library(openxlsx)

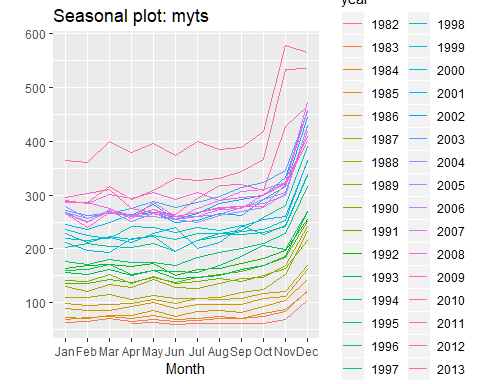
## Warning: package 'openxlsx' was built under R version 3.5.3

file2<-"retail.xlsx"  
retaildata <-read\_excel("retail.xlsx", skip=1)  
myts <- ts(retaildata[,"A3349873A"],  
 frequency=12, start=c(1982,4))

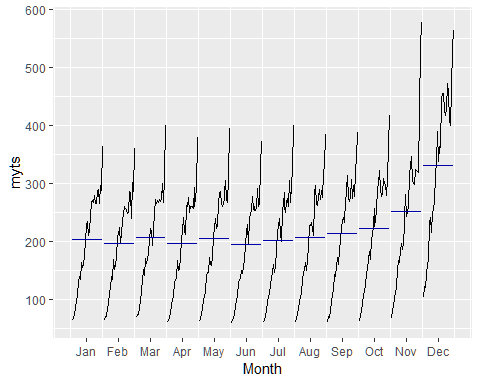
autoplot(myts) +  
 ggtitle("Australian retail data")

 The plot shows a steady inceras of retail throughout the years between early 1980’s until late 2010’s

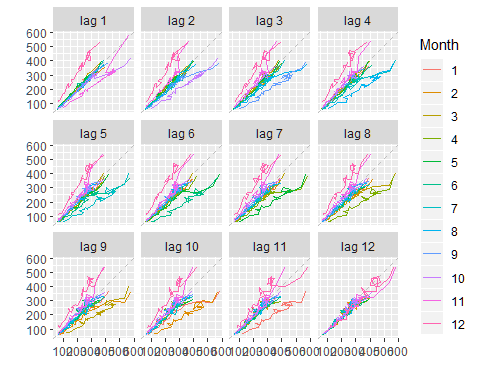
ggseasonplot(myts)

 In terms of seasonality there is a large jump in sales starting in October and lasting till end of December.

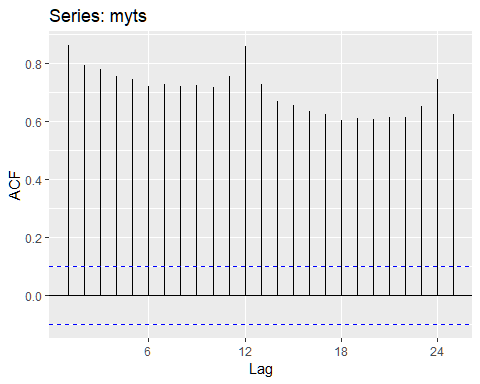
ggsubseriesplot(myts)

 The subseries plot shows the mean values of the time series. In this regard we see that the values are in an upward trend from October to December.

gglagplot(myts, lags=12)

 From the lagplot we see that there is a strong correlation for all lag values. It is strongest for lag=12.

ggAcf(myts)

 The ACF plot shows strong autocorrelation. The seasonality is reflected in the strongest ACF being observed at lag=12.