DATA 624 - Homework 3

OMER OZEREN

Table of Contents

Exercises 6.2 and 6.3 from the Hyndman online Forecasting book.

```
#clear the workspace
rm(list = ls())
#load req's packages
library(forecast)
library(readxl)
library(RCurl)
library(fpp2)
library(seasonal)
```
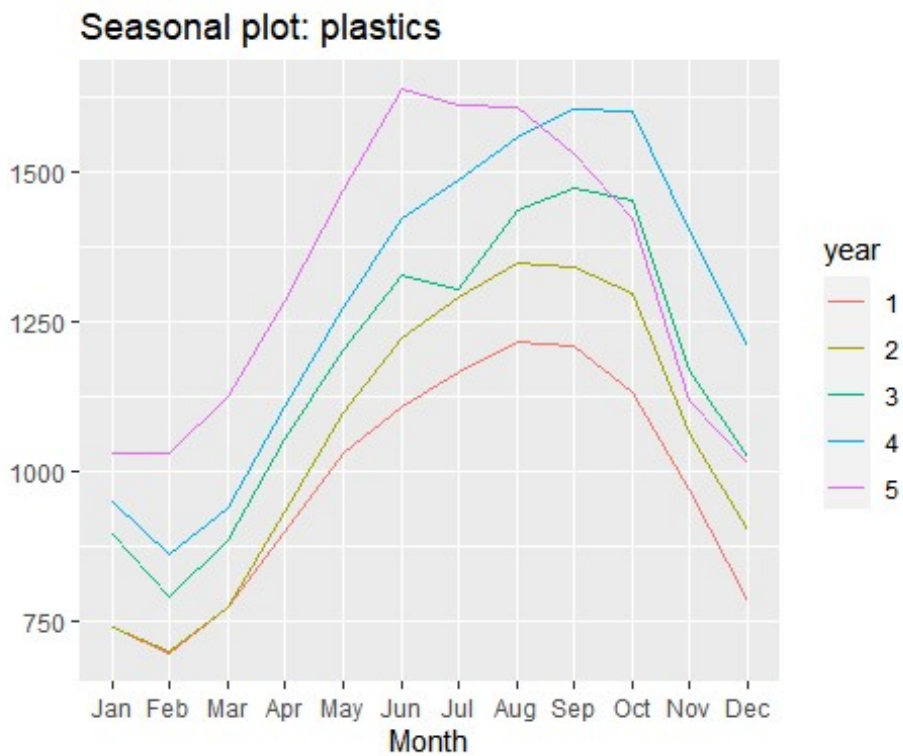
Question 6.2

The plastics data set consists of the monthly sales (in thousands) of product A for a plastics manufacturer for five years.
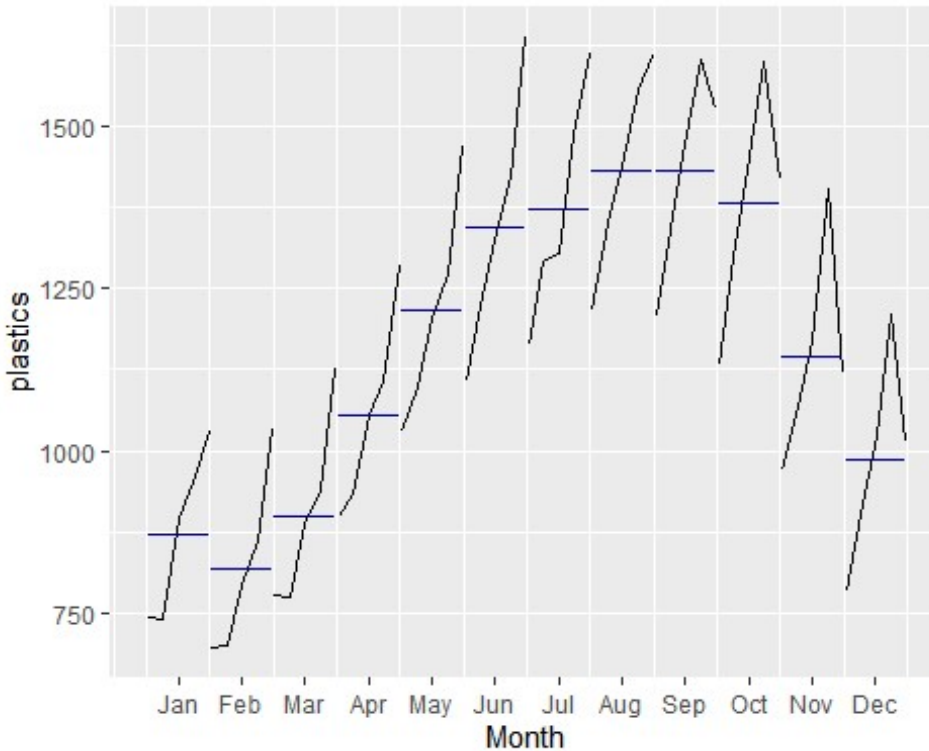
A - Plot the timeseries

```
autoplot(plastics)
```
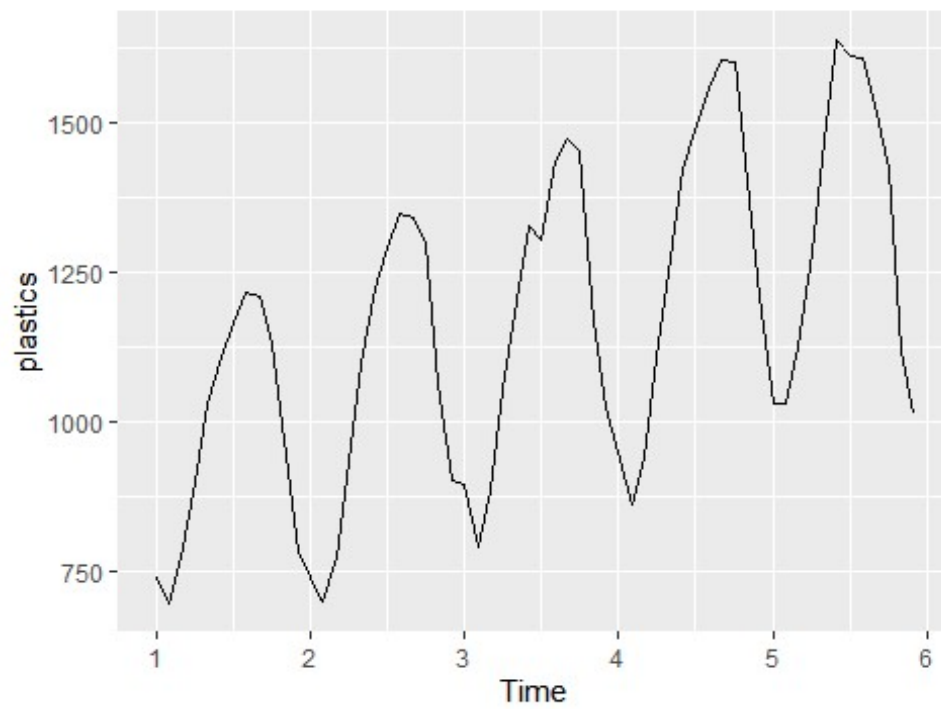
```
ggseasonplot(plastics)
```


Seasonal plot: plastics

```
ggsubseriesplot(plastics)
```

We see an apparent seasonality in the timeseries which is confirmed by the seasonal plot. Sales are at their seasonal low in Feb and at their seasonal high in Aug / Sept.

The sub-series plot shows an interesting pattern. For Sept to Dec, we see a shart reversal in the final data-points of the subseries plot. It could be indicitive of a change in the overall trend, or possibly a change in the season. There is not enough data here to draw a reasonable conclusion either way.

B - Classical Multiplicative Decomposition

```
mult.decomp <- decompose(plastics, type = "multiplicative")
autoplot(mult.decomp )
```

Decomposition of multiplicative time series
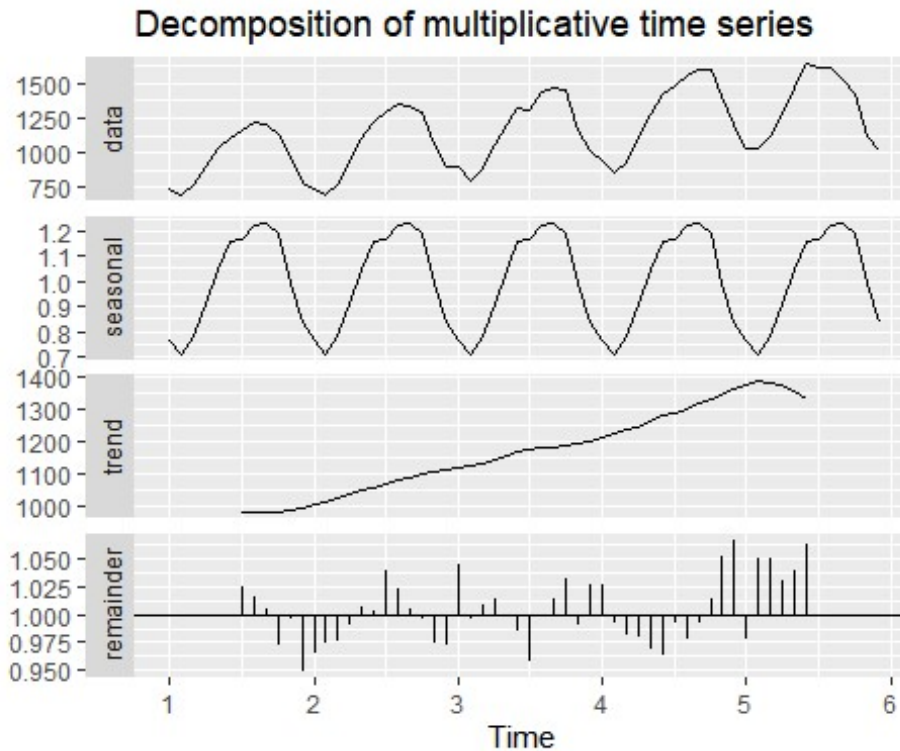
C - Do your results from part B support interpretation from part A?

Yes. We can see a clear seasonal pattern and in addition, we see a downturn in the trend in year 5, as suggested as a possibility based on the sub-series plots.

D - Compute and plot the seasonally adjusted data

```
autoplot(plastics, series="Raw") +
  autolayer(trendcycle(mult.decomp), series="Trend") +
  autolayer(seasadj(mult.decomp), series="Seasonally Adj") +
  xlab("Year Num") + ylab("Sales in 000s") +
  ggtitle("Monthly Sales (Plastics) - 000s") +
  scale_colour_manual(values=c("gray","blue","red"),
              breaks=c("Raw","Seasonally Adj","Trend"))
```

Monthly Sales (Plastics) - 000s
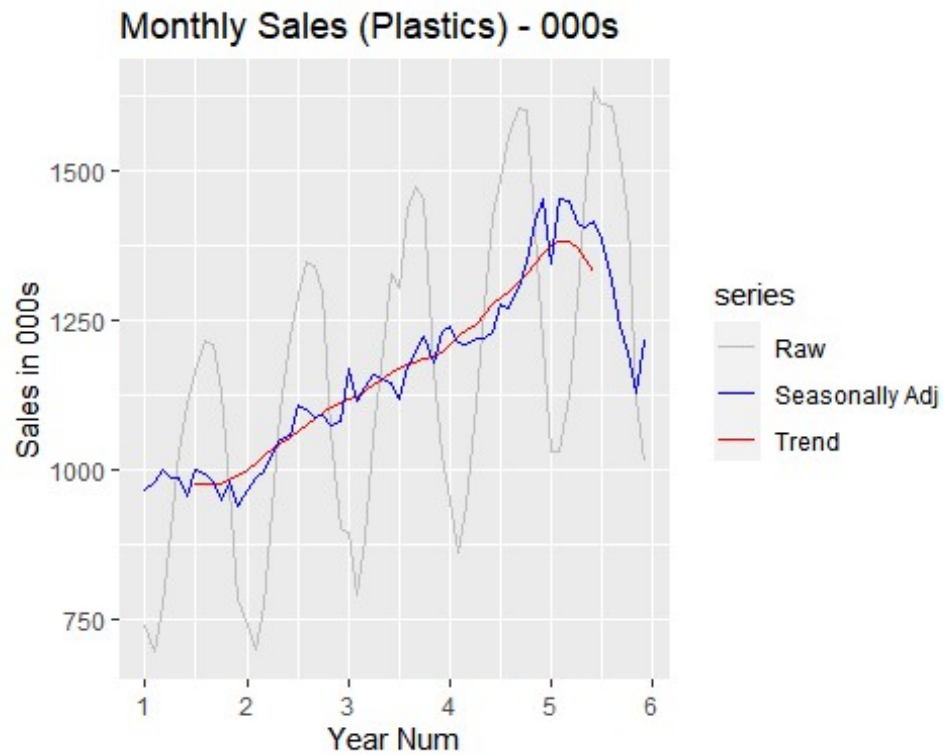
E - Inject an outlier

```
#randomly inject an outlier between 500-1000
p.mod <- plastics
rnd.index <- sample(1:length(plastics),1)
p.mod[rnd.index] <- p.mod[rnd.index]+ (sample(500:1000,1) * sample(c(-
1,1),1))
#re-run all the above
mult.decomp <- decompose(p.mod, type = "multiplicative")
autoplot(mult.decomp )
```
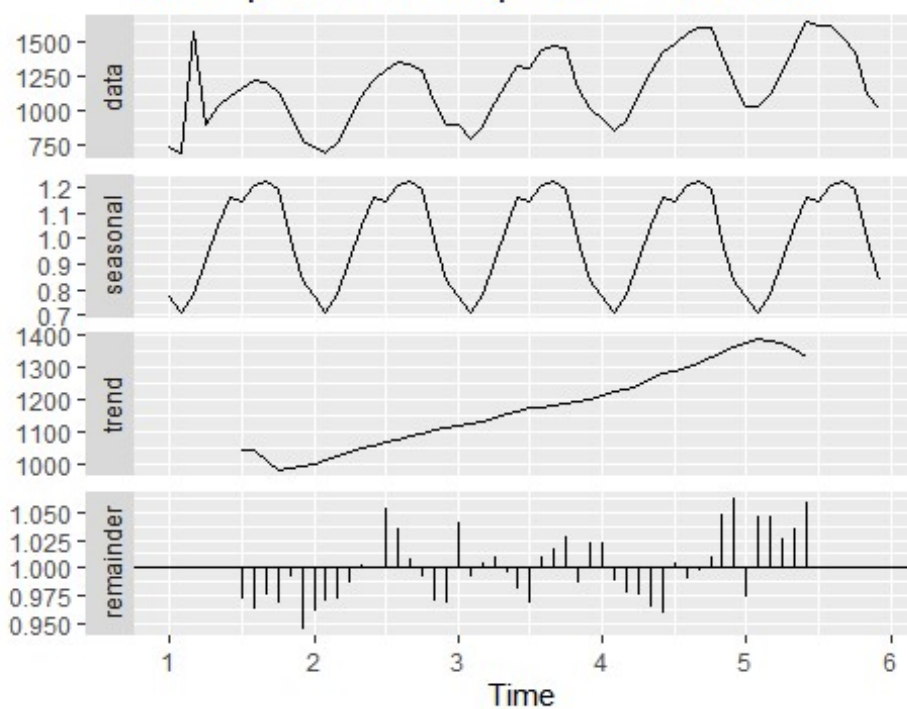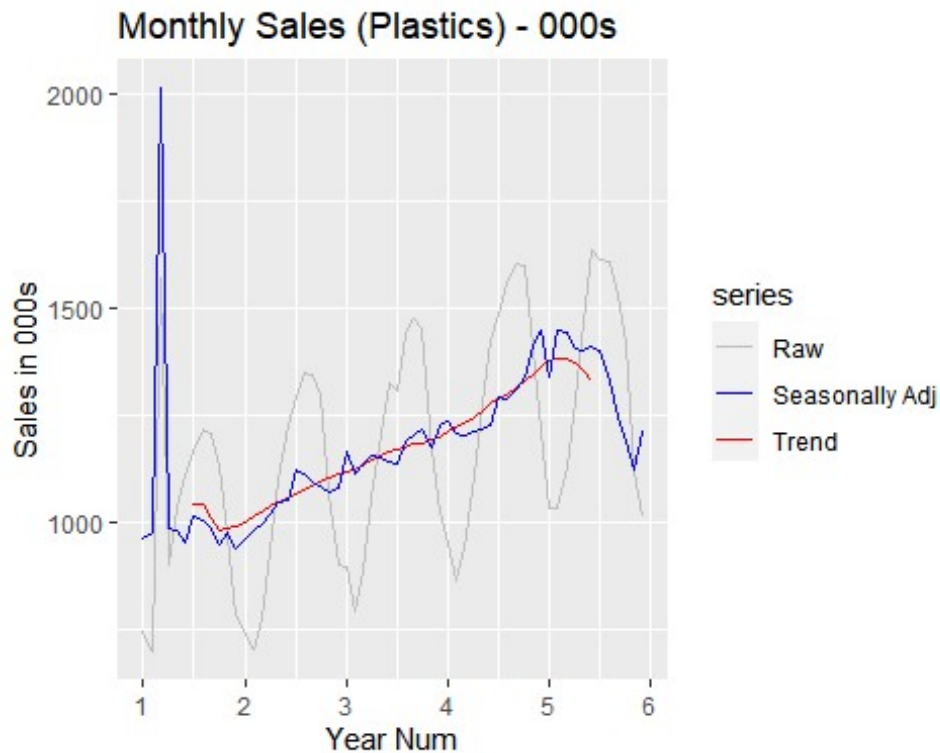
Decomposition of multiplicative time series


```
autoplot(p.mod , series="Raw") +
  autolayer(trendcycle(mult.decomp), series="Trend") +
  autolayer(seasadj(mult.decomp), series="Seasonally Adj") +
  xlab("Year Num") + ylab("Sales in 000s") +
  ggtitle("Monthly Sales (Plastics) - 000s") +
  scale_colour_manual(values=c("gray","blue","red"),
            breaks=c("Raw","Seasonally Adj","Trend"))
```

Monthly Sales (Plastics) - 000s

The trend component is generally insensitive to outliers of the magnitude selected and that the seasonal component is highly sensitive - in this case, we appear to get a major cyclical blip in the seasonally adjusted data. The sign of the outlier doesn't appear to be important as it pertains to either of these observations.

F - Does outlier location matter?

```
#randomly inject an outlier between 500-1000
p.mod <- plastics
rnd.index <- 47
p.mod[rnd.index] <- p.mod[rnd.index]+ 500
autoplot(p.mod , series="Raw") +
  autolayer(trendcycle(mult.decomp), series="Trend") +
  autolayer(seasadj(mult.decomp), series="Seasonally Adj") +
  xlab("Year Num") + ylab("Sales in 000s") +
  ggtitle("Monthly Sales (Plastics) - 000s") +
  scale_colour_manual(values=c("gray","blue","red"),
            breaks=c("Raw","Seasonally Adj","Trend"))
```
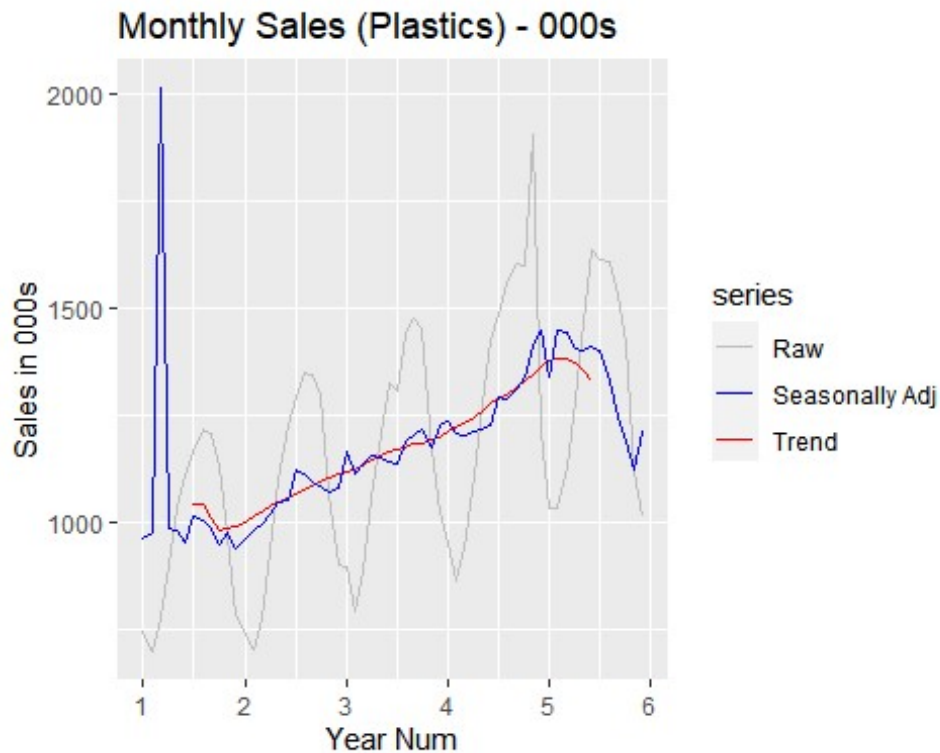
Monthly Sales (Plastics) - 000s

The position of the outlier does not seem to matter greatly as it related to the seasonally adjusted data. My observation is that it has a disruptive effect regardless of position.

In terms of the trend component, there is a minor effect related to oulier position. In the above chart, we can see the blip in the trend component around x=4.25

Question 6.3

Recall your retail time series data (from Exercise 3 in Section 2.10). Decompose the series using X11. Does it reveal any outliers, or unusual features that you had not noticed previously?

```
#borrowed code from week1 hw to load the aussie retail data
temp_file <- tempfile(fileext = ".xlsx")
download.file(url =
"https://github.com/omerozeren/DATA624/raw/master/HMW1/retail.xlsx",
            destfile = temp_file,
            mode = "wb",
            quiet = TRUE)
retaildata <- readxl::read_excel(temp_file,skip=1)
aussie.retail <- ts(retaildata[,"A3349388W"],
  frequency=12, start=c(1982,4))
#run decomp as per the text
x11.decomp <- seas(aussie.retail, x11="")
autoplot(x11.decomp, main = "Aussie Retail - X11 Decomposition" )
```

Aussie Retail - X11 Decomposition
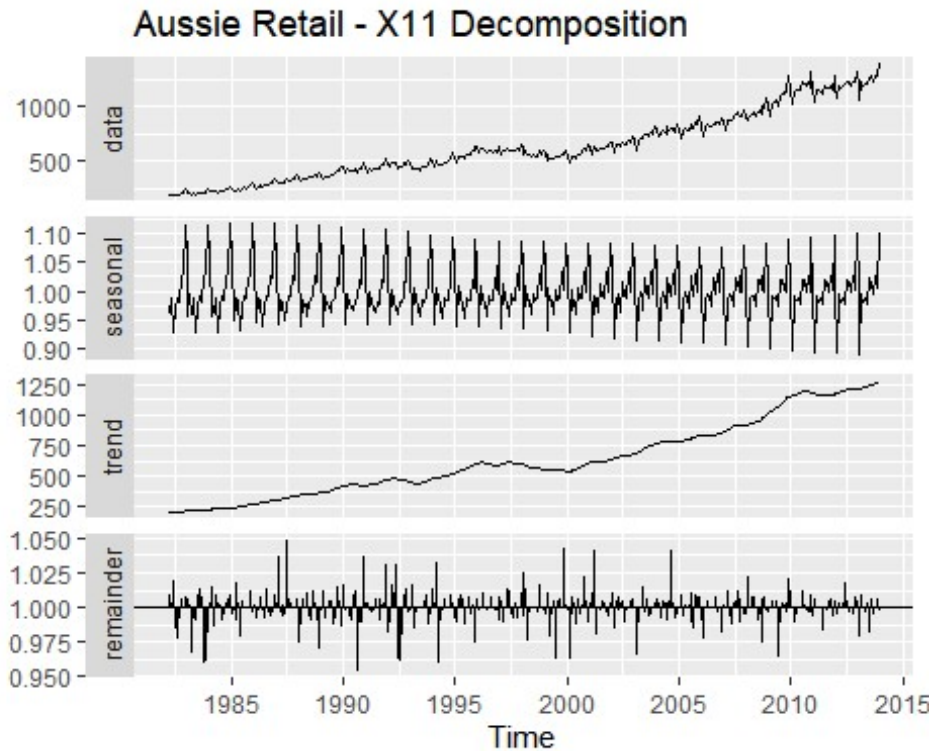
At a high level, the aussie retail data looks similar to previous representations and there are no real surprises here. There is a strong seasonal effect and there is a reasonably conssitent positive trend. The remainder values are reasonably consistent through time with no major standouts. Retail data above shows a strong seasonal and trend. Trend is increasing over the period of time. Seasonal part is also varying over the period of time. Seasonal component has low amplitudes to start with but as we go further in time, seasonal variations increases in aplitude. This is a clear example of multiplicative time series. Remainder component of the decomposed graph also shows two outliers near year 1990. Hence,In most years, the sales drop off from May to June, and then increase in July.However, in 2000, the raw data series shows a sharp increase in June 2000, which is not present in other years.