DATA 624: PREDICTIVE ANALYTICS

Gabriel Campos

Last edited February 04, 2024

```
library('ggplot2')
library('forecast')
library('quantmod')
```

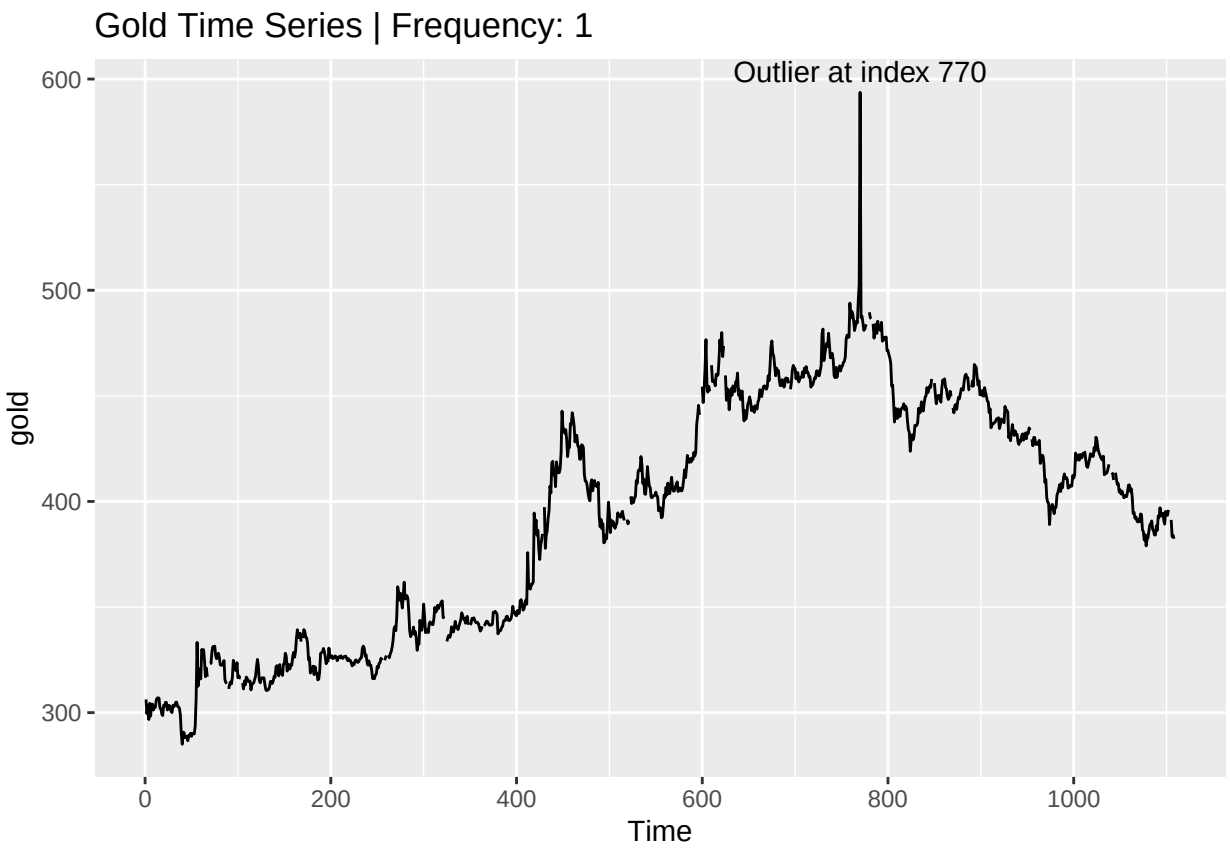
INSTRUCTIONS

Please submit exercises 2.1, 2.2, 2.3, 2.4, 2.5 and 2.8 from the Hyndman online Forecasting book. Please submit both your Rpubs link as well as attach the .pdf file with your code.

1. Use the help function to explore what the series gold, woolyrnq and gas represent. Use `autoplot()` to plot each of these in separate plots. What is the frequency of each series? Hint: apply the frequency() function. Use which.max() to spot the outlier in the gold series. Which observation was it?

```
gld_frq<-frequency(gold)
gld_max<-which.max(gold)

autoplot(gold)+
  ggtitle(paste("Gold Time Series | Frequency:",gld_frq))+
   annotate("text", x = gld_max, y = gold[gld_max],
            label = paste("Outlier at index", gld_max), vjust = -0.5)
```



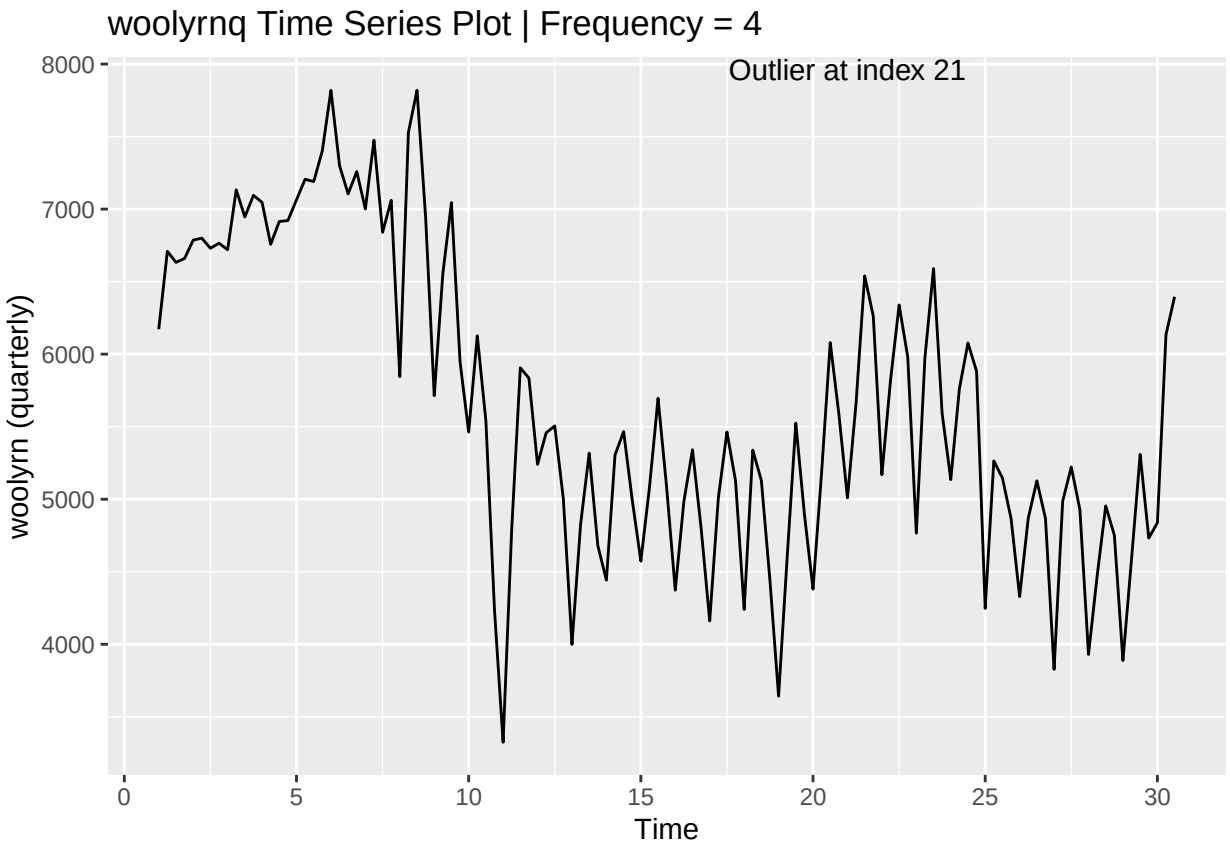

```
# help(woolyrnq)
head(woolyrnq,30)
```

```
##      Qtr1 Qtr2 Qtr3 Qtr4
## 1965 6172 6709 6633 6660
## 1966 6786 6800 6730 6765
## 1967 6720 7133 6946 7095
## 1968 7047 6757 6915 6921
## 1969 7064 7206 7190 7402
## 1970 7819 7300 7105 7259
## 1971 7001 7475 6840 7061
## 1972 5845 7529
```

```
wlyrnq_freq <- frequency(woolyrnq)
wlyrnq_max <- which.max(woolyrnq)


# use ts() b/c data as per Forecasting: Principles and Practice 2.1 ts object
# e.g. y <- ts(z, start=2003, frequency=12)

autoplot(ts(woolyrnq, frequency = wlyrnq_freq)) +
  ggtitle(paste("woolyrnq Time Series Plot | Frequency =", wlyrnq_freq," ")) +
   ylab("woolyrn (quarterly)") +
  annotate("text", x = wlyrnq_max, y = woolyrnq[wlyrnq_max],
           label = paste("Outlier at index", wlyrnq_max), vjust = -0.5)
```



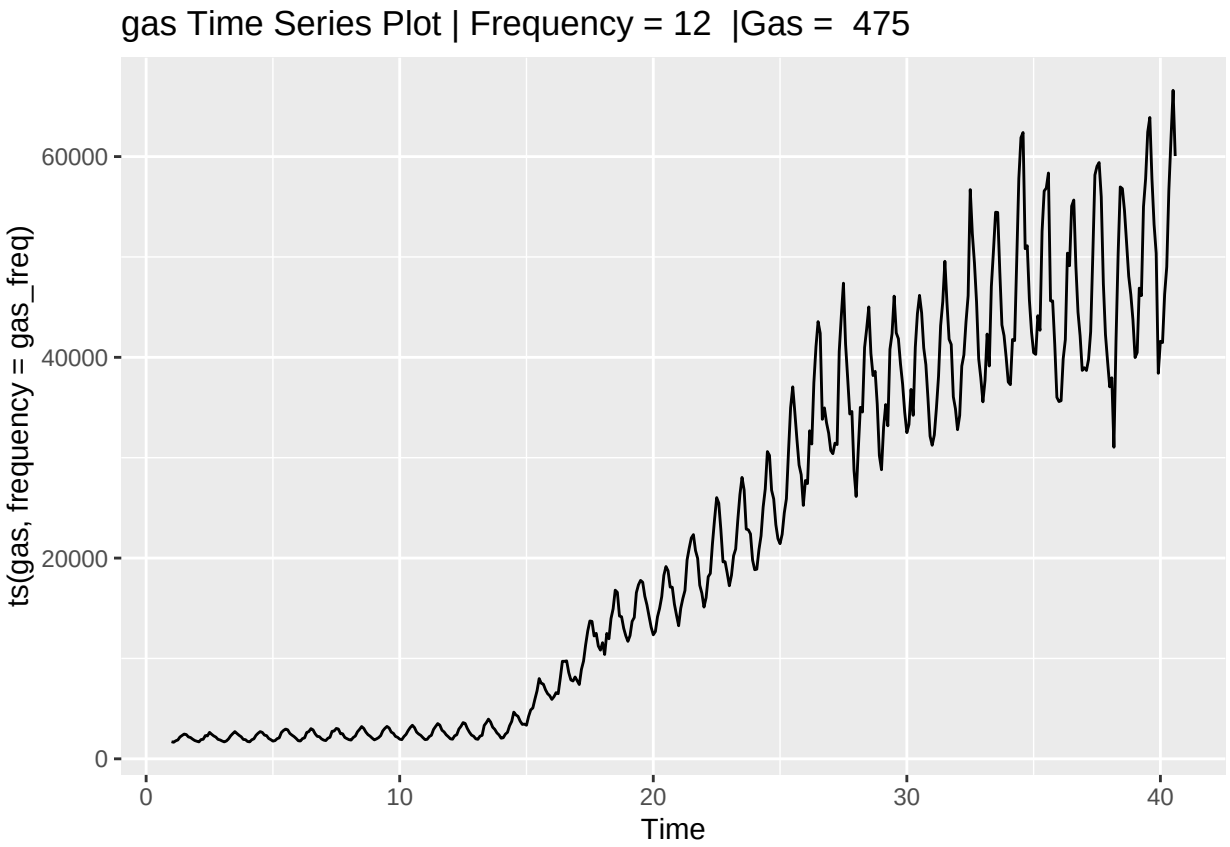

```
#help(gas)
head(gas,100)
```

```
##       Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
## 1956 1709 1646 1794 1878 2173 2321 2468 2416 2184 2121 1962 1825
## 1957 1751 1688 1920 1941 2311 2279 2638 2448 2279 2163 1941 1878
## 1958 1773 1688 1783 1984 2290 2511 2712 2522 2342 2195 1931 1910
## 1959 1730 1688 1899 1994 2342 2553 2712 2627 2363 2311 2026 1910
## 1960 1762 1815 2005 2089 2617 2828 2965 2891 2532 2363 2216 2026
## 1961 1804 1773 2015 2089 2627 2712 3007 2880 2490 2237 2205 1984
## 1962 1868 1815 2047 2142 2743 2775 3028 2965 2501 2501 2131 2015
## 1963 1910 1868 2121 2268 2690 2933 3218 3028 2659 2406 2258 2057
## 1964 1889 1984 2110 2311
```

```
#autoplot(gas)
gas_freq <- frequency(gas)
gas_max <- which.max(gas)

# use ts() b/c data as per Forecasting: Principles and Practice 2.1 ts object
# e.g. y <- ts(z, start=2003, frequency=12)

autoplot(ts(gas, frequency = gas_freq)) +
  ggtitle(paste("gas Time Series Plot | Frequency =", gas_freq," |Gas = ", gas_max))
```



2. Download the file `tute1.csv` from the book website, open it in Excel (or some other spreadsheet application), and review its contents. You should find four columns of information. Columns B through D each contain a quarterly series, labelled Sales, AdBudget and GDP. Sales contains the quarterly sales for a small company over the period 1981-2005. AdBudget is the advertising budget and GDP is the gross domestic product. All series have been adjusted for inflation.

a. You can read the data into R with the following script:

```
tute1 <- read.csv(tute1_csv, header=TRUE)
head(tute1,10)
```

```
##          X  Sales AdBudget   GDP
## 1  Mar-81 1020.2    659.2 251.8
## 2  Jun-81  889.2    589.0 290.9
## 3  Sep-81  795.0    512.5 290.8
## 4  Dec-81 1003.9    614.1 292.4
## 5  Mar-82 1057.7    647.2 279.1
## 6  Jun-82  944.4    602.0 254.0
## 7  Sep-82  778.5    530.7 295.6
## 8  Dec-82  932.5    608.4 271.7
## 9  Mar-83  996.5    637.9 259.6
## 10 Jun-83  907.7    582.4 280.5
```

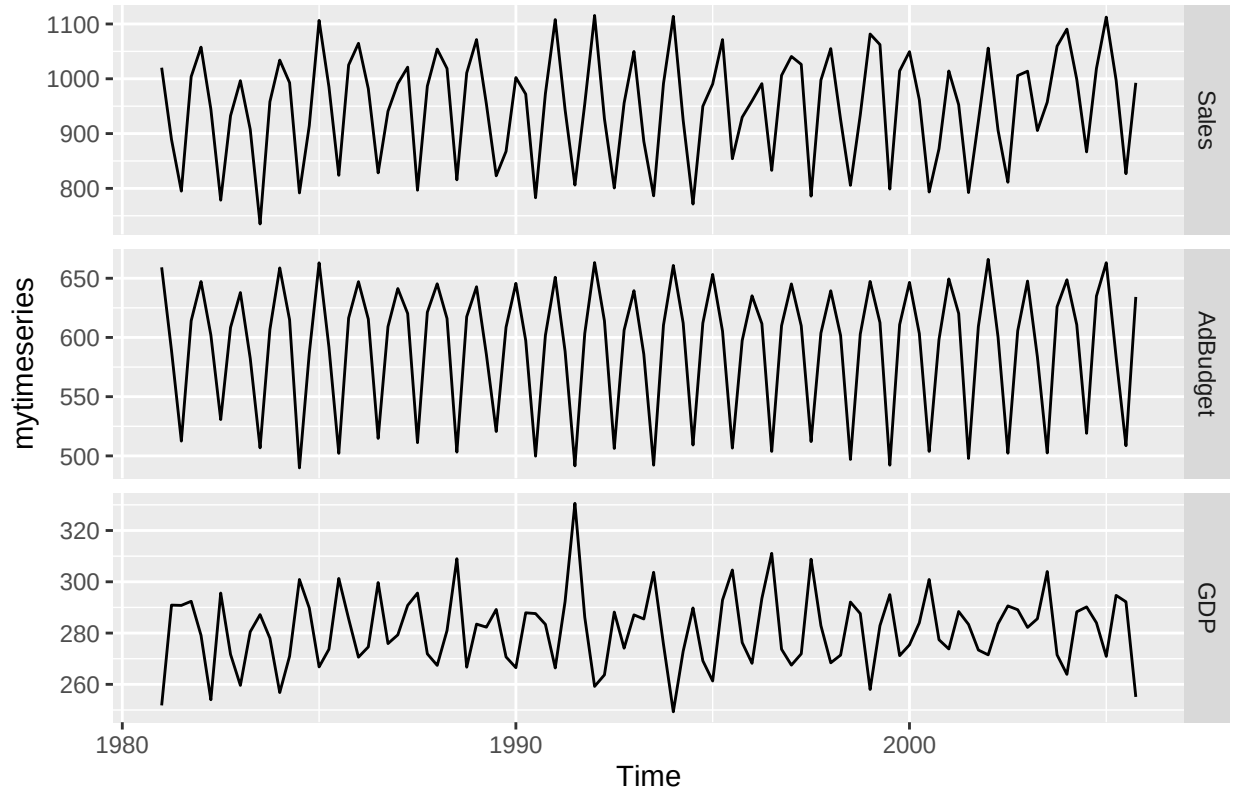
b. Convert the data to time series

```
mytimeseries <- ts(tute1[,-1], start=1981, frequency=4)
```

(The [,-1] removes the first column which contains the quarters as we don't need them now.)

c. Construct time series plots of each of the three series

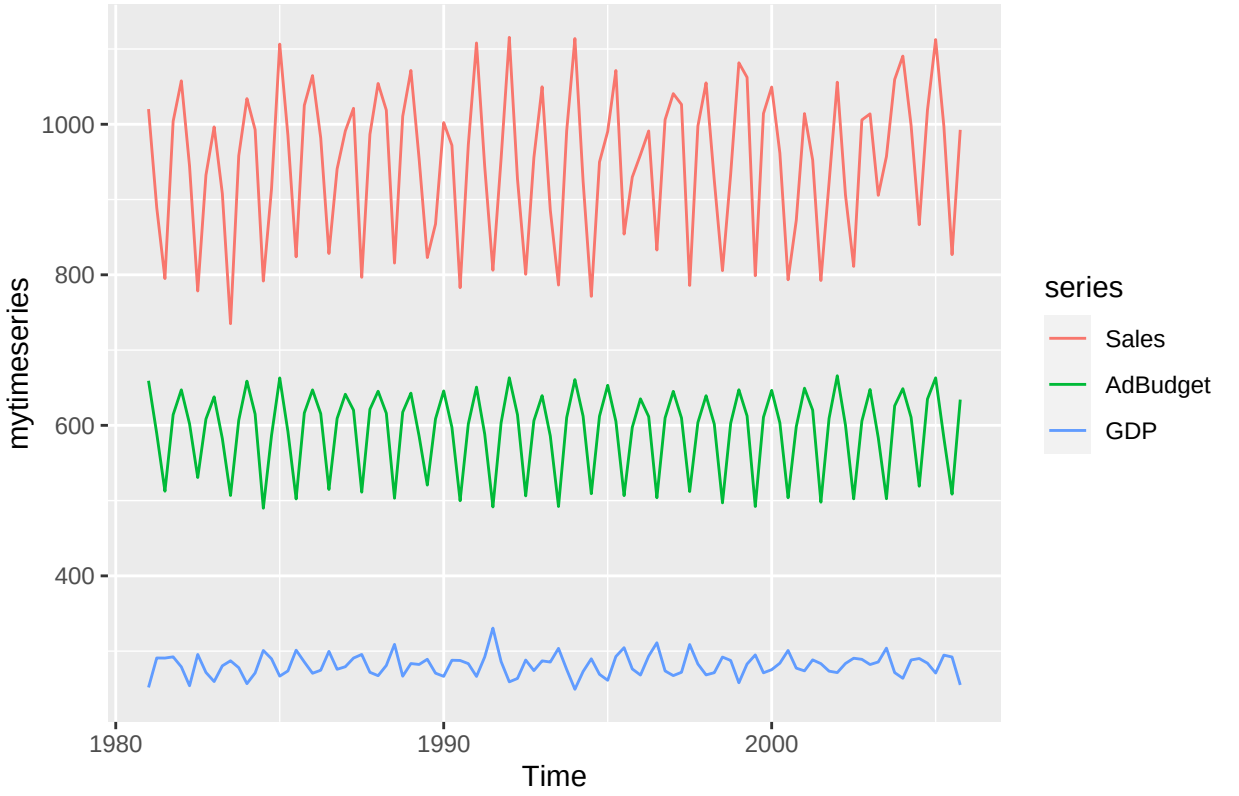
```
autoplot(mytimeseries, facets=TRUE)
```



Check what happens when you don't include facets=TRUE.

ANSWER

```
autoplot(mytimeseries, facets=FALSE)
```



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. You can read the data into R with the following script:

```
#retaildata <- readxl::read_excel("retail.xlsx", skip=1)
```

The second argument (skip=1) is required because the Excel sheet has two header rows.

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

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?

Create time plots of the following time series: `bicoal`, `chicken`, `dole`, `usdeaths`, `lynx`, `goog`, `writing`, `fancy`, `a10`, `h02`.