Tutorial 1 Answers

1. Reading in the csv file and printing out the first six rows.

```
# Read data file and create monthly ts object
dt <- read.csv("RetailSales.csv")
print(head(dt))
```

```
    Date   Food Hhold Clothing DeptStores Other Restaurants  Total
1 Apr-82 1162.6 592.3    359.9      460.1 479.1       342.4 3396.4
2 May-82 1150.9 629.6    386.6      502.6 486.1       342.1 3497.9
3 Jun-82 1160.0 607.4    350.5      443.8 467.5       328.7 3357.8
4 Jul-82 1206.4 632.4    359.3      459.1 491.1       338.5 3486.8
5 Aug-82 1152.5 622.6    325.2      438.4 485.7       331.5 3355.9
6 Sep-82 1189.1 622.0    346.3      465.1 489.9       341.9 3454.3
```

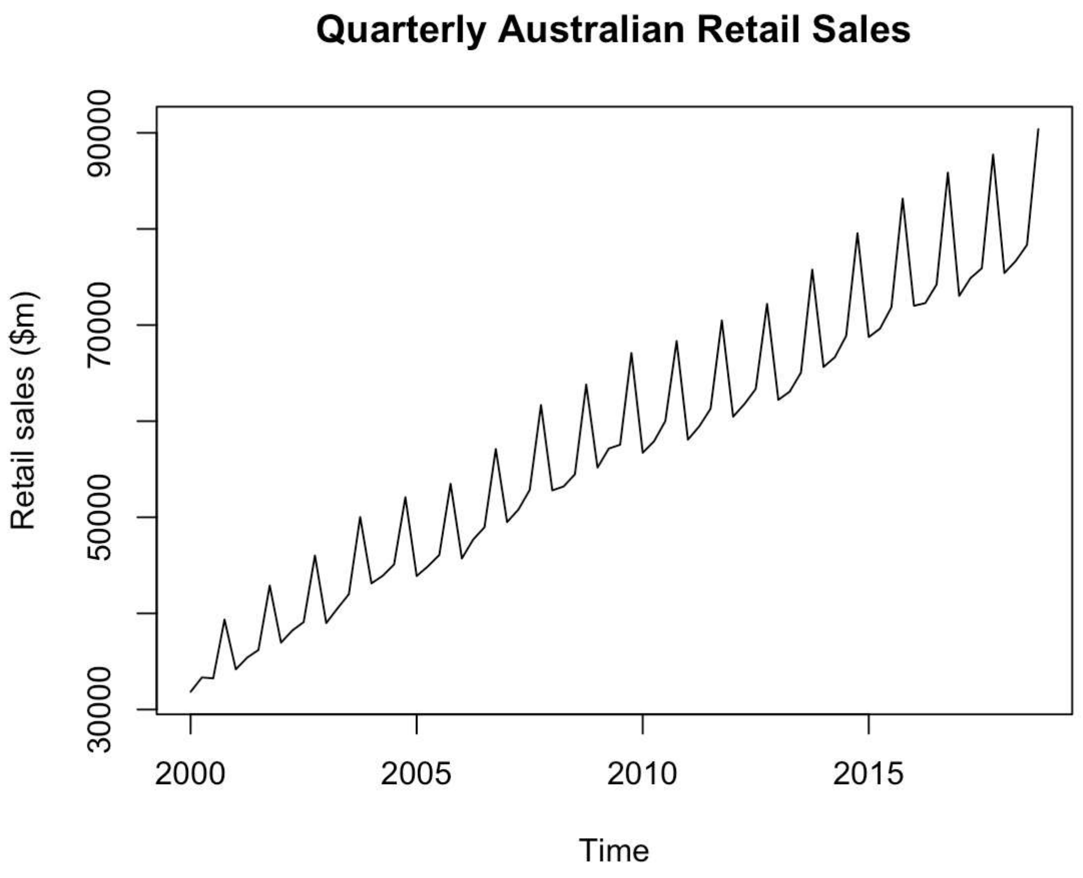
2. Creating the monthly time series object from April 1982 to September 2024, with a time series plot.

```
Retail_m <- ts(dt$Total, frequency=12, start=c(1982,4), end=c(2024,9))
plot(Retail_m, main="Monthly Australian Retail Sales",
  ylab="Retail sales ($m)")
```

3. Conversion to quarterly frequency, restrict sample (window) to 2000q1 to 2018q4.

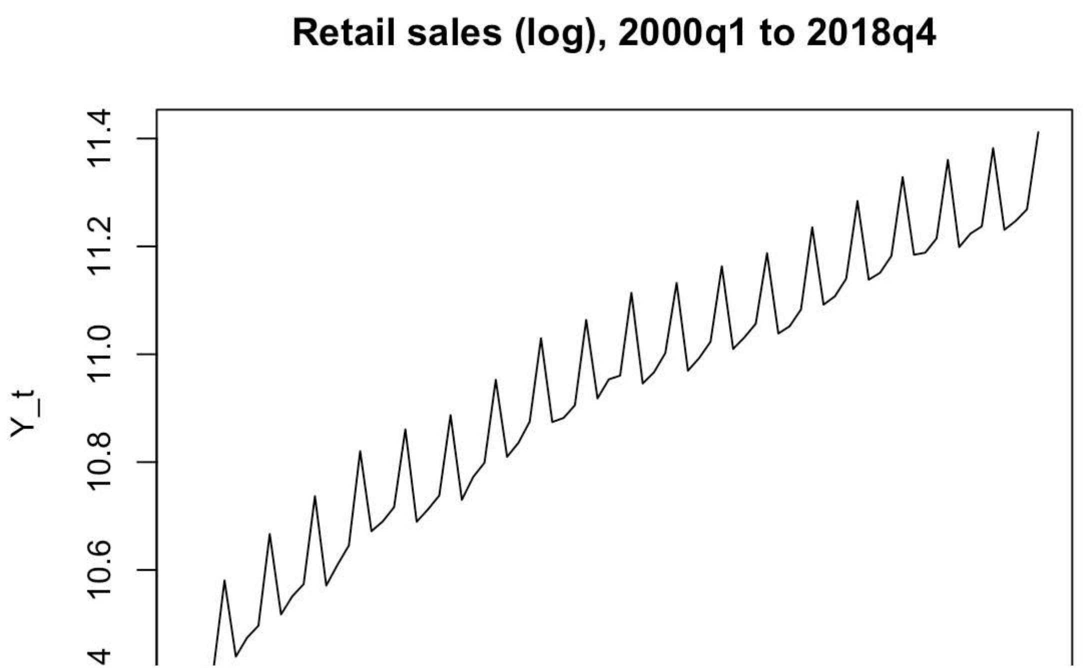
```
Retail_q <- aggregate(Retail_m, nfrequency=4)
S_t <- window(Retail_q, start=c(2000,1), end=c(2018,4))
plot(S_t, main="Quarterly Australian Retail Sales",
```

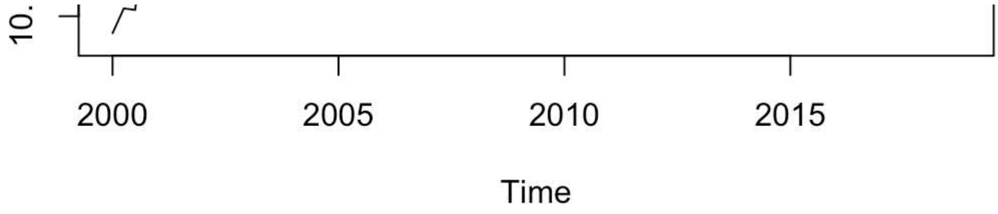
```
    ylab="Retail sales ($m)")
```



4. Taking the log of the quarterly time series gives the plot below. Note how the log stabilises the seasonal pattern, and makes evident the slowdown in the increase in the trend since around the middle of the sample.

```
Y_t <- log(S_t)
plot(Y_t, main="Retail sales (log), 2000q1 to 2018q4")
```





5. The first difference of log retail sales is found using the `diff' command.

```
# First difference of log retail sales, convert to pct.
D_Y_t <- diff(Y_t)*100
print(round(head(cbind(Y_t,D_Y_t)),4))
```

```
            Y_t     D_Y_t
2000 Q1 10.3685        NA
2000 Q2 10.4144    4.5997
2000 Q3 10.4116   -0.2842
2000 Q4 10.5805   16.8887
2001 Q1 10.4393  -14.1158
2001 Q2 10.4745    3.5186
```

Differencing is an important operation in time series analysis. In this case it is the *first* difference because the difference is taken between observations *one* time period apart. For example $4.5997 = (10.4144 - 10.3685) \times 100$. The first value in `D_Y_t` is missing (`NA`) because there is no value of `Y_t` available before 10.3685 to compute the difference. Taking a difference of a time series will necessarily result in a shorter time series because of this.

6. The percentage change in retail sales is computed directly as follows.

```
# Percentage change of retail sales
R_t <- diff(S_t)/lag(S_t,-1)*100
# Compare
print(round(head(cbind(D_Y_t, R_t)),4))
```

```
           D_Y_t      R_t
2000 Q2   4.5997   4.7071
2000 Q3  -0.2842  -0.2838
2000 Q4  16.8887  18.3986
2001 Q1 -14.1158 -13.1648
2001 Q2   3.5186   3.5813
2001 Q3   2.2103   2.2349
```

The values of $\log S_t - \log S_{t-1}$ and $(S_t - S_{t-1})/S_{t-1}$ are quite similar especially for the smaller changes. For the entire time series the two variables are very highly correlated:

```
print(round(cor(R_t, D_Y_t),3))
```

```
[1] 0.999
```

This is also evident on the plot shown below. For our purposes in forecasting, we will very frequently take the log of a time series because it tends to stablise / linearise features such as trend and seasonality. Changes in a logged time series can be interpreted as approximately percentage changes, which is often convenient.

```
plot(as.numeric(R_t), as.numeric(D_Y_t), pch=16, cex=0.7,
     xlab="Rt", ylab="DYt")
lines(as.numeric(R_t), as.numeric(R_t), col="blue")
```