Week 1 DATA624 Time Series Graphics Exercises

Jonathan Hernandez

January 30, 2019

```
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. You can read the data into R with the following script:
retaildata <- readxl::read_excel("retail.xlsx", skip=1)
```

- Reading in the data:

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

```
b. Select one of the time series as follows (but replace the column name with
your own chosen column):
myts <- ts(retaildata[,"A3349873A"],
  frequency=12, start=c(1982,4))
```

- Let's first read the data and see the column names and replace it with a column I choose. I'll choose say the 10th column.

```r
column_name <- names(retaildata)[10] # 10th column
myts <- ts(retaildata[,column_name],
      frequency=12, start=c(1982,4))
```

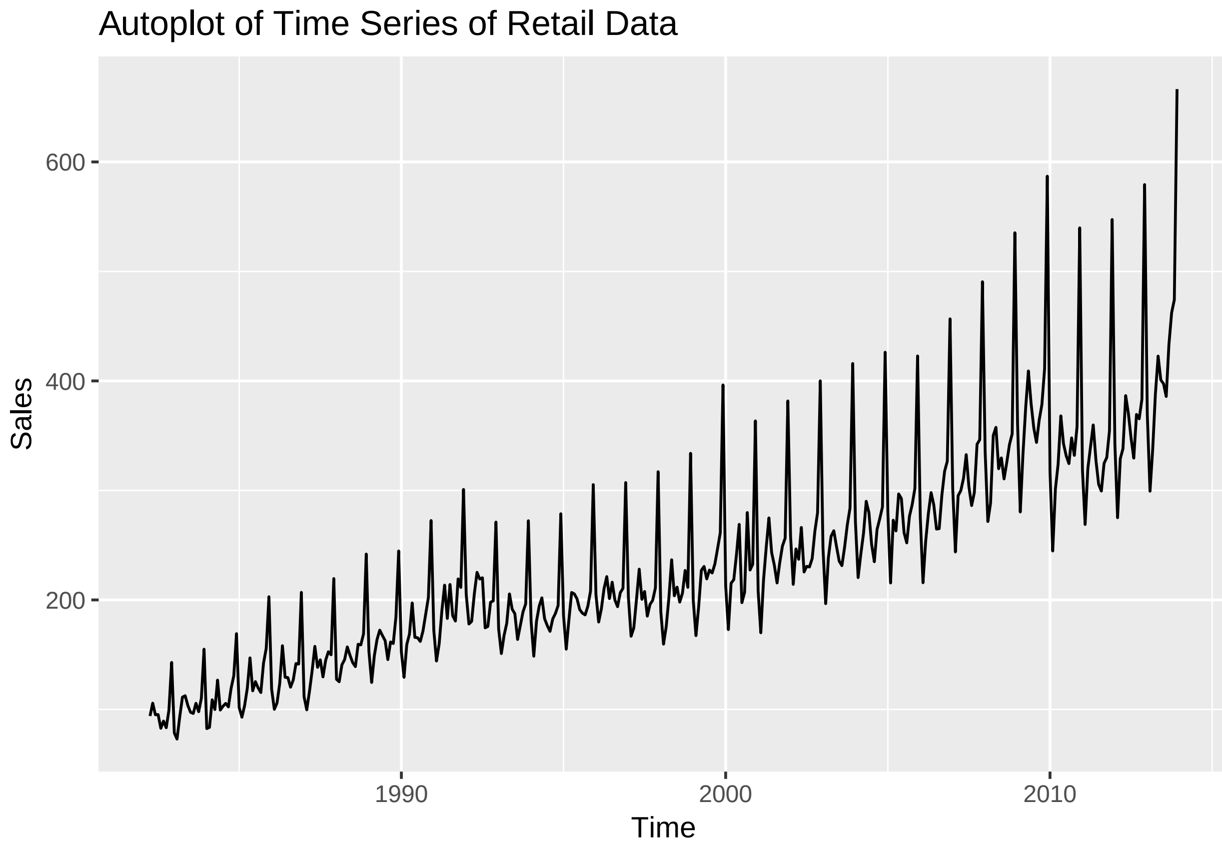
```
c. Explore your chosen retail time series using the following functions:
autoplot(), ggseasonplot(), ggsubseriesplot(), gglagplot(), ggAcf()
Can you spot any seasonality, cyclicity and trend? What do you learn about the series?
```

- Load the ggplot2 library and plot as requested and spot any trends, seasonality and cyclicity.

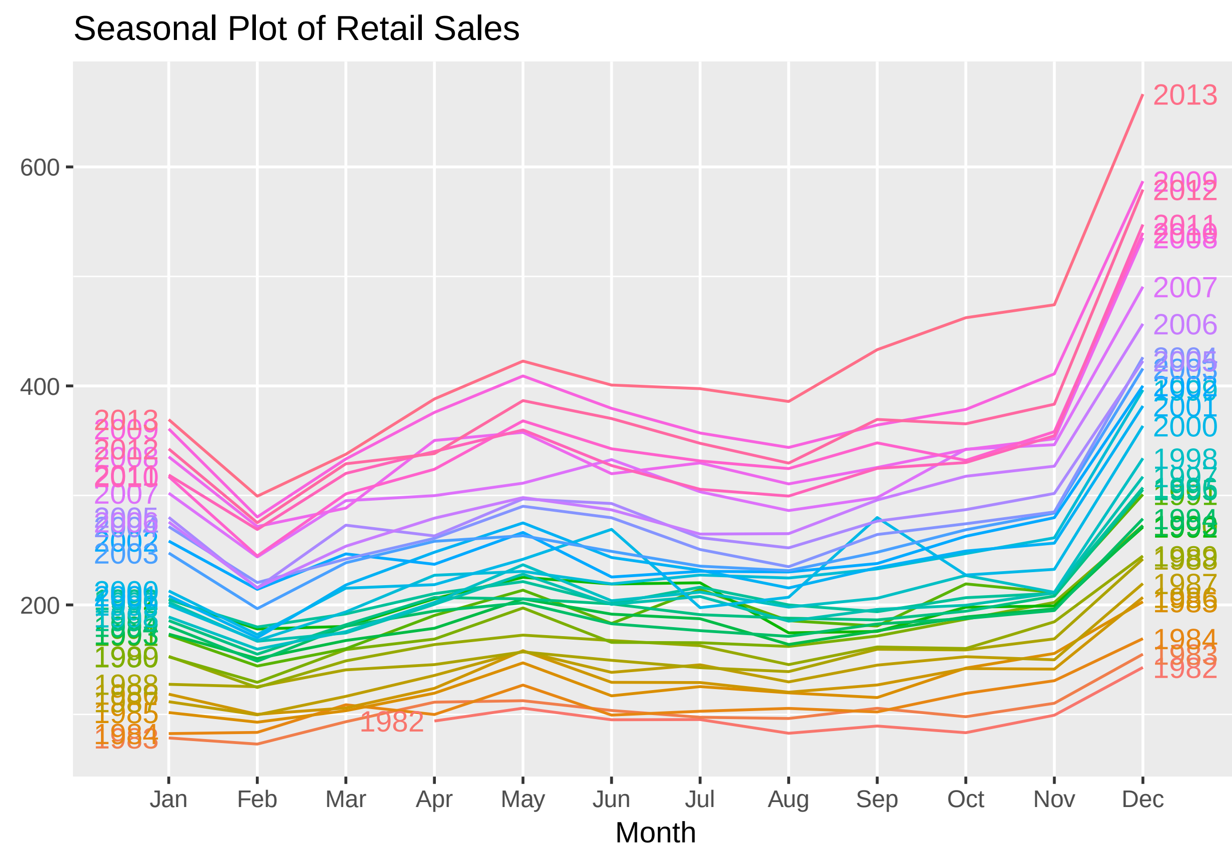
```r
library(ggplot2)
library(ggfortify)
library(fpp2)
```

```
## Loading required package: forecast

## Loading required package: fma

## Loading required package: expsmooth
```

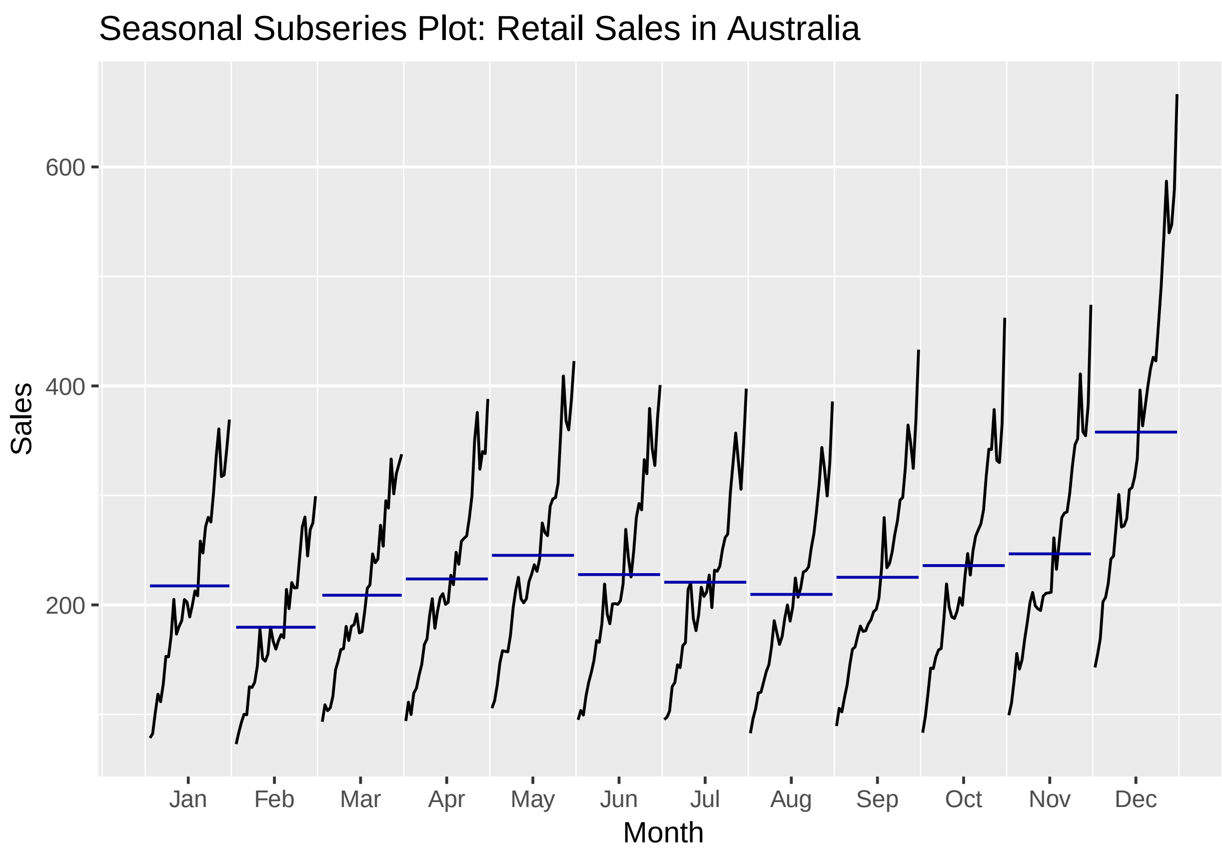
```r
#autoplot
autoplot(myts) +
  ggtitle("Autoplot of Time Series of Retail Data") +
  xlab("Time") +
  ylab("Sales")
```




```
#ggseasonplot
ggseasonplot(myts, year.labels=TRUE, year.labels.left=TRUE) +
  ggtitle("Seasonal Plot of Retail Sales")
```



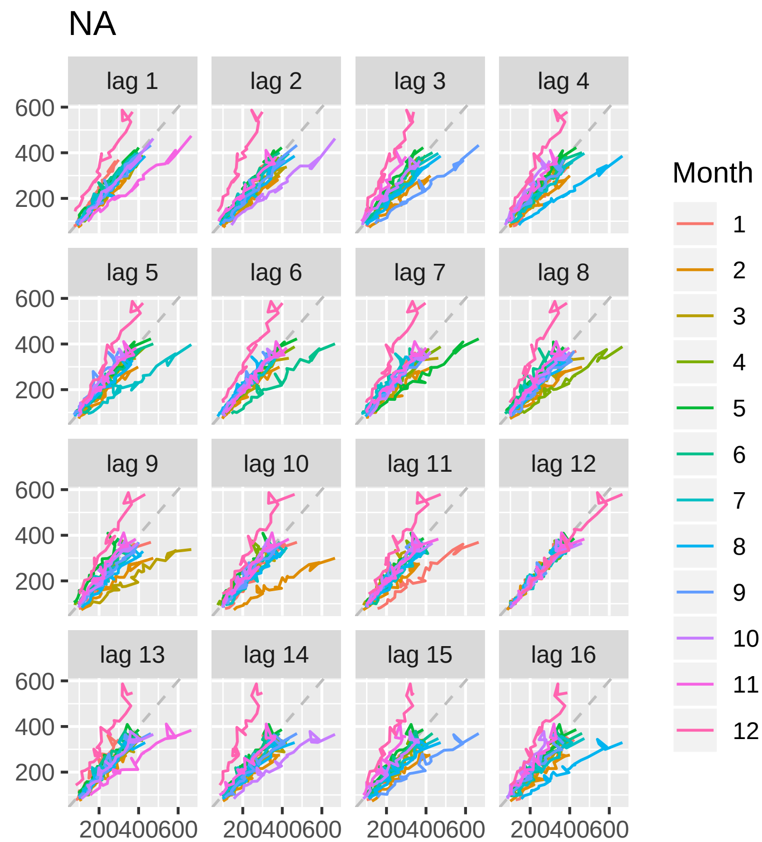

```r
# ggsubseriesplot
ggsubseriesplot(myts) +
  ylab("Sales") +
  xlab("Month") +
  ggtitle("Seasonal Subseries Plot: Retail Sales in Australia")
```



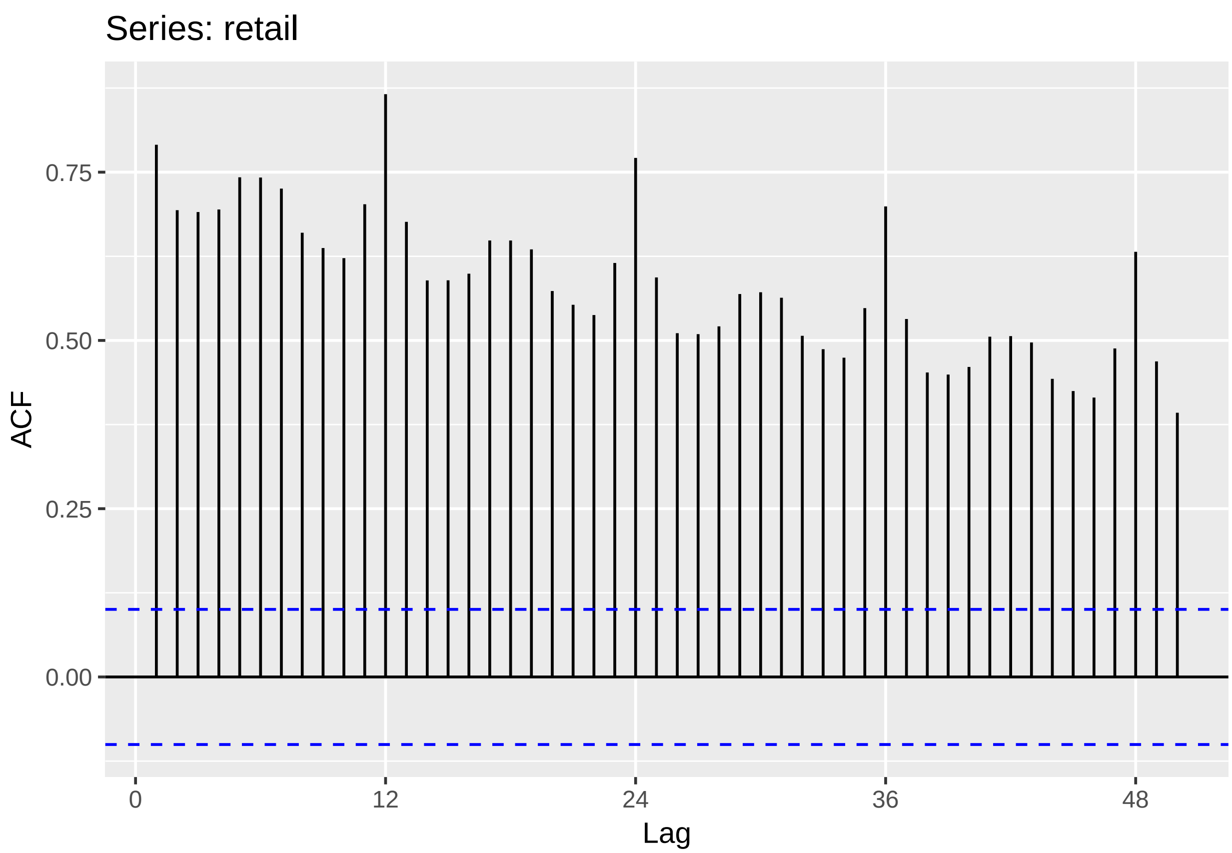

```
## gglagplot
retail <- window(myts, start=1982)
```

```
## Warning in window.default(x, ...): 'start' value not changed
```

```
gglagplot(retail)
```




```
# ggAcf
ggAcf(retail, lag=50)
```



- From the the autoplot(), there appears to be a increasing trend as well as seasonality. This is due to the increase in data and retail sales over the years. There is also seasonality as about every year or so, there is a spike in sales perhaps in a certain month.
- From the ggseasonplot(), we see that there is seasonality which now showing months, reveals that major increases appear going into December as well as sale spikes from February to around May.
- Overall, there is an increasing trend and sales spike up entering in December and after February, for most years increases until May.

  2.7 The arrivals data set comprises quartely international arrivals (in thousands) to Australia from Japan, New Zealand, UK and the US. Use autoplot(), ggseasonplot() and ggsubseriesplot() to compare the differences between the arrivals from these four countries. Can you identify any unusual observations?
- We'll take a look at the arrivals datasets from the fpp2 package as that has already been loaded.

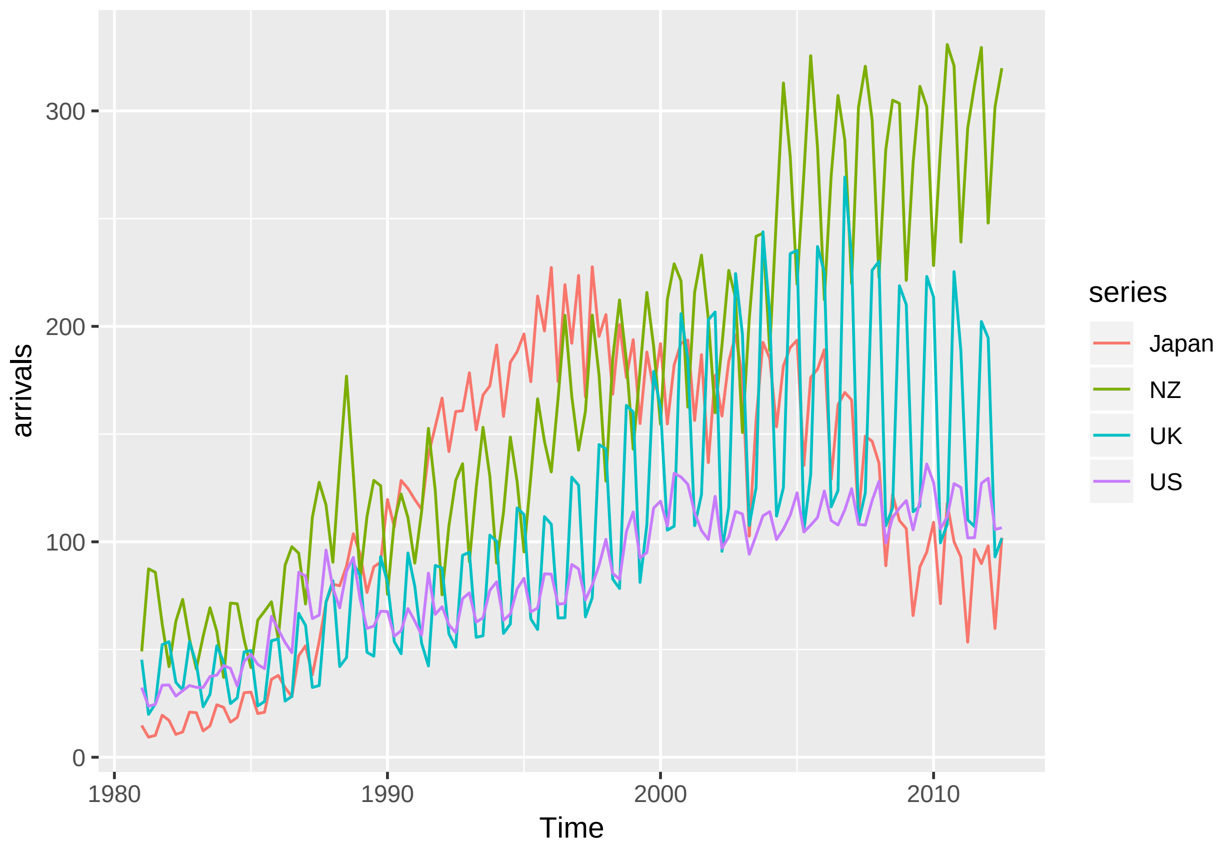
```
summary(arrivals)
```

```
##      Japan              NZ                UK                US
##  Min.   :  9.321   Min.   : 37.04   Min.   : 19.89   Min.   : 23.72
##  1st Qu.: 74.135   1st Qu.: 96.52   1st Qu.: 53.89   1st Qu.: 63.95
##  Median :135.461   Median :154.54   Median : 95.56   Median : 85.88
##  Mean   :122.080   Mean   :170.59   Mean   :106.86   Mean   : 84.85
##  3rd Qu.:176.752   3rd Qu.:228.60   3rd Qu.:128.13   3rd Qu.:108.98
##  Max.   :227.641   Max.   :330.81   Max.   :269.29   Max.   :136.09
```

```r
str(arrivals)
```

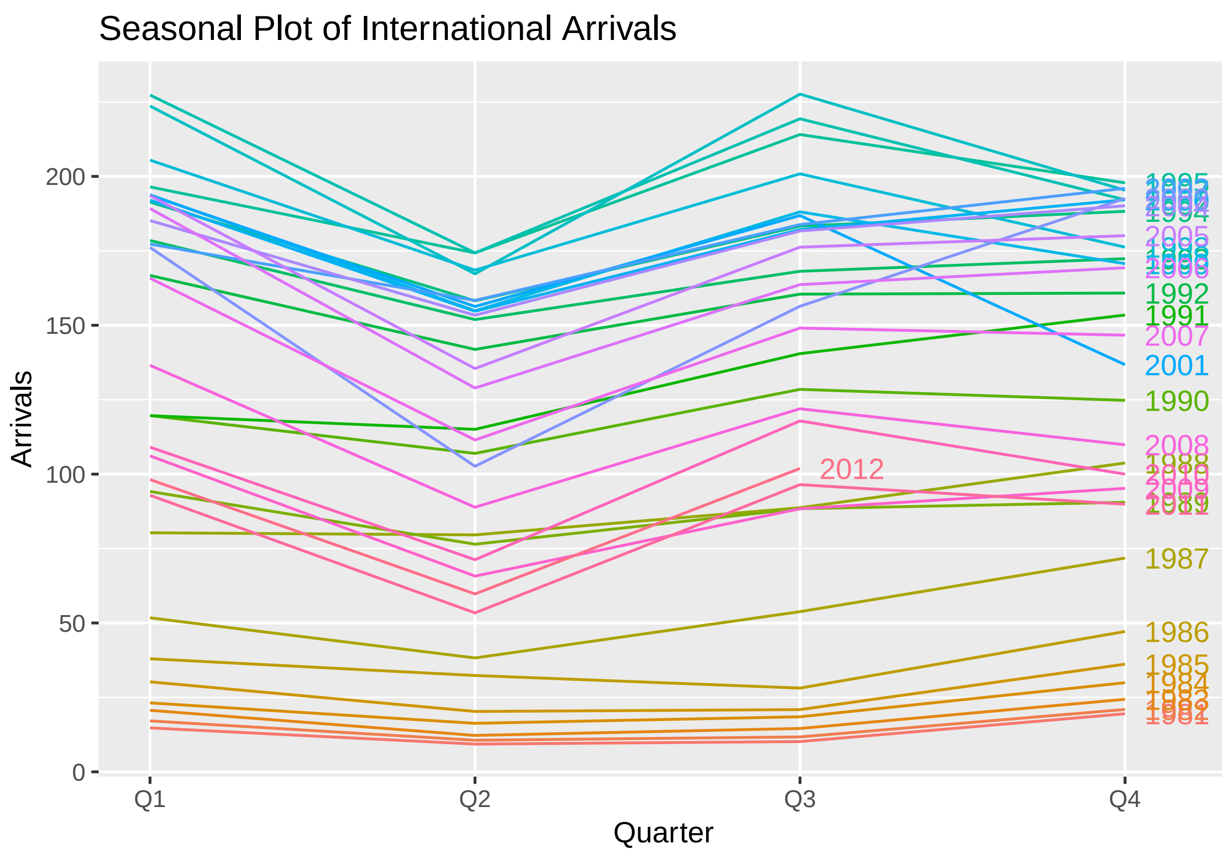
```
##  Time-Series [1:127, 1:4] from 1981 to 2012: 14.76 9.32 10.17 19.51 17.12 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : NULL
##   ..$ : chr [1:4] "Japan" "NZ" "UK" "US"
```

```r
autoplot(arrivals)
```



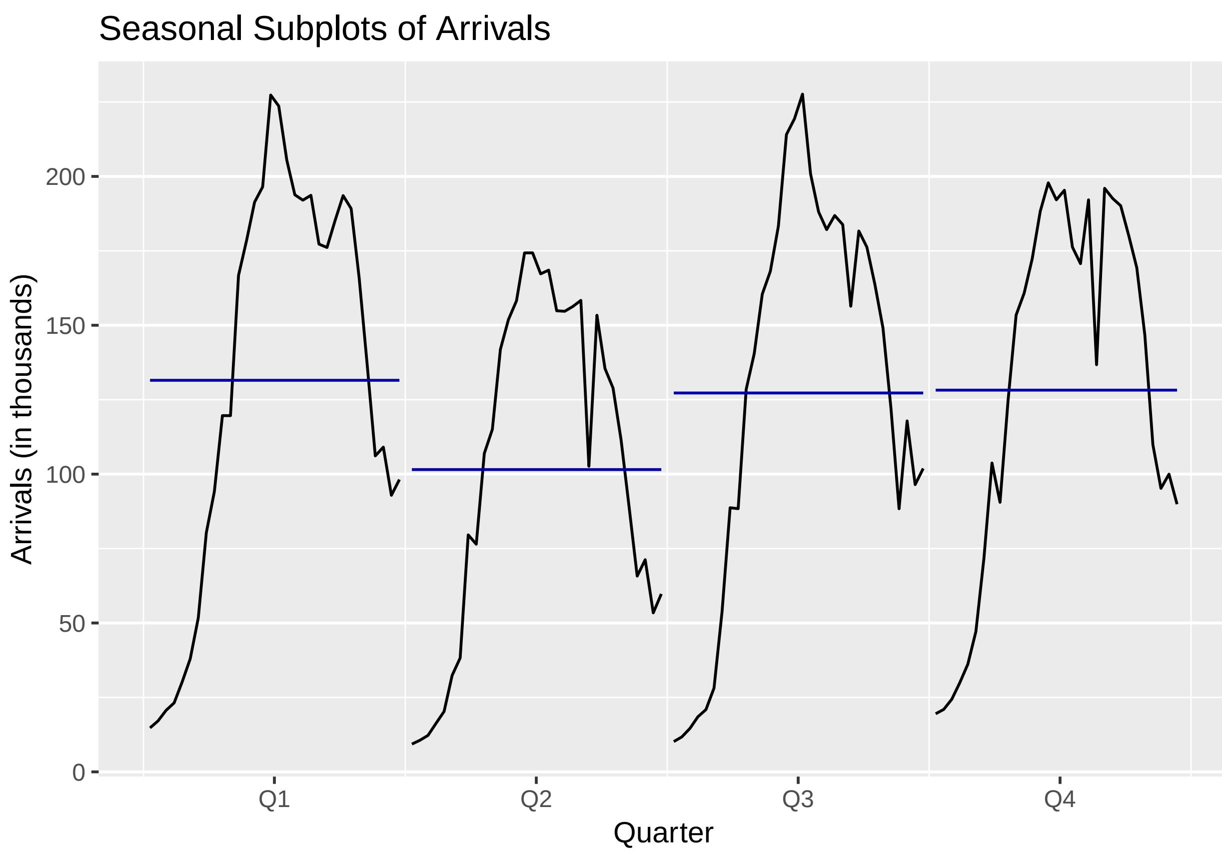
- For these arrivals, there seems to be strong seasonalilty. For United States flights, there seems to be cyclic behavior. Arrivals to New Zealand show a strong increasing trend. Flights to the UK also have a cyclic behavior as well.
- Plotting using ggseasonplot()

```r
#ts_arrivals <- ts(arrivals, frequency = 4, start=c(1981,1))
ggseasonplot(arrivals[,1], year.labels = TRUE) +
  ggtitle("Seasonal Plot of International Arrivals") +
  ylab("Arrivals")
```



- From this seasonal plot, for most years, total arrivals starts with a downward trend followed by a increasing trend from Q2 to Q3. Also several years there are both increasing and decreasing trends.

```
ggsubseriesplot(arrivals[,1]) +
  ylab("Arrivals (in thousands)") +
  ggtitle("Seasonal Subplots of Arrivals")
```



- Here we see that for each quarter, the amount of arrivals is roughly the same and follow a similar pattern. The plot also shows that the average arrivals for Q1, Q3 and Q4 are approximately the same. Q2 average arrivals are lower than the rest. So throughout the years, the average arrivals were always the same in the first quarter and the last half of the years.
- It is quite strange that Q2 arrivals would be slighly lower than other times throughout the years. One would think that international flights would increase during the spring season.

  2.10 dj contains 292 consecutive trading days of the Dow Jones Index. use `ddj <- diff(dj)` to compute the daily changes in the index. Plot ddj and its ACF. Do the changes in the Dow Jones Index look like white noise?
- Let's load the dj dataset and plot the data and examine its correlogram and check for white noise.

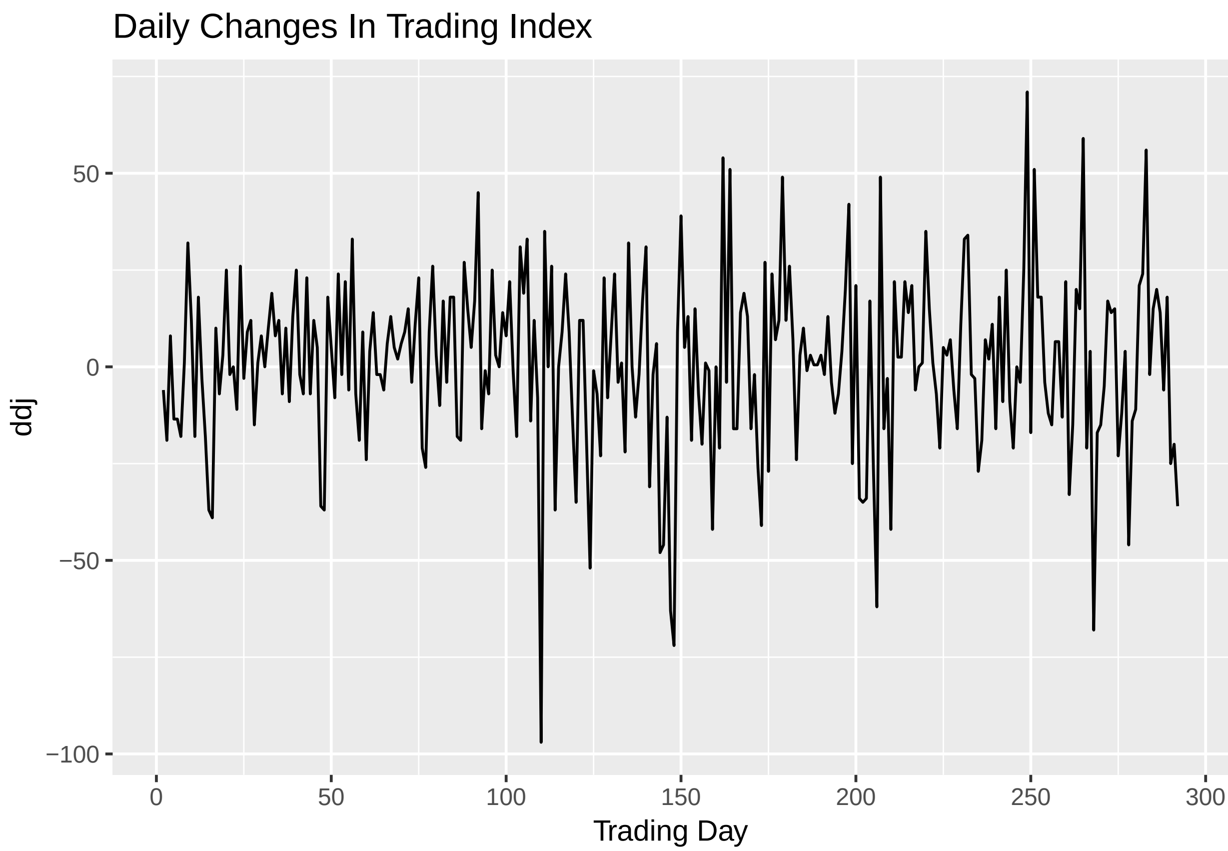
```
str(dj)
```

```
##  Time-Series [1:292] from 1 to 292: 3651 3645 3626 3634 3620 ...
```

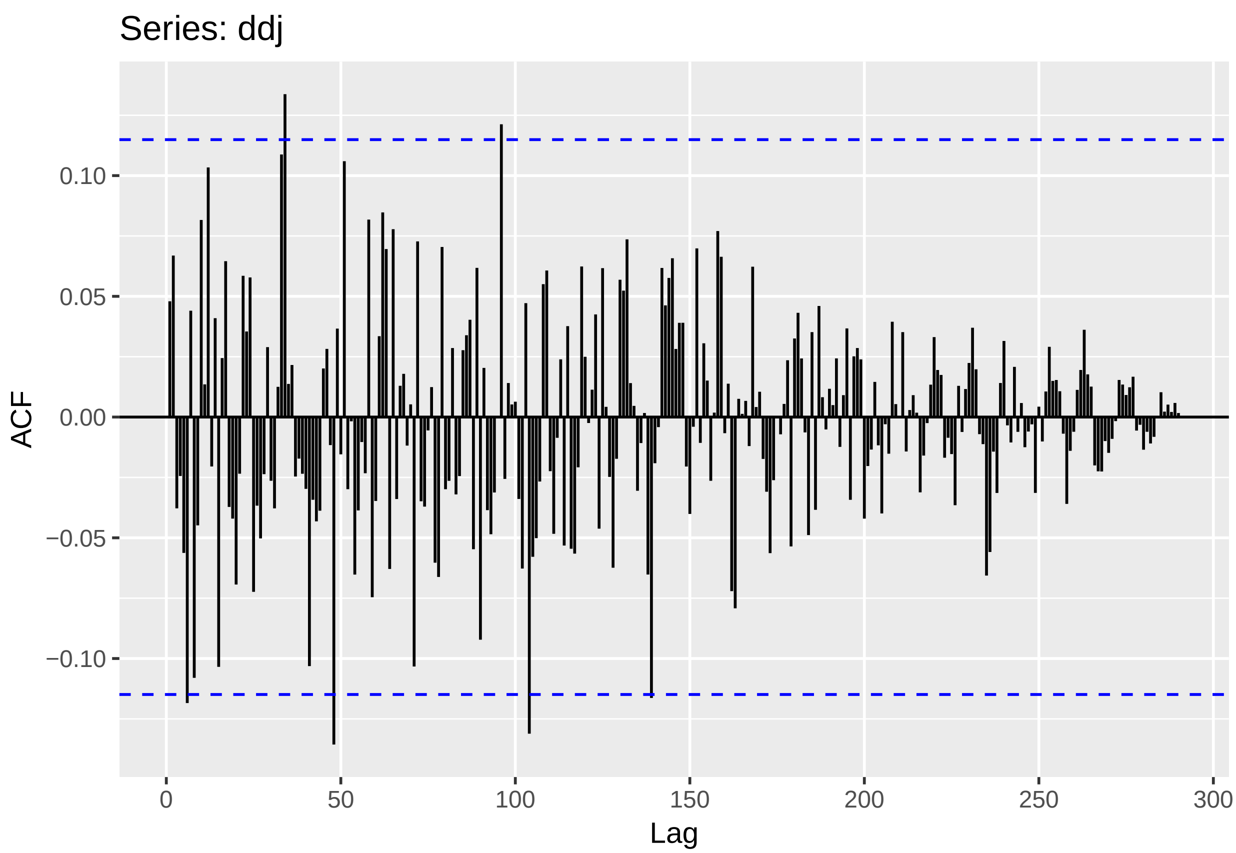
```
summary(dj)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3537    3672    3752    3755    3849    3978
```

```r
ddj <- diff(dj)

autoplot(ddj) +
  xlab("Trading Day") +
  ggtitle("Daily Changes In Trading Index")
```




```r
# ACF
ggAcf(ddj, lag = 290)
```



- By Looking at the daily changes in the autoplot, the daily changes look similar to white noise meaning there is no trend or seasonality. The ACF plot can show otherwise as more than one lag value is above the bounds $\pm 2/\sqrt{291} = 0.11$ which may not be white noise.

  3.1 For the following series, find an appropriate Box-Cox transformation in order to stabilise the variance

  - usnetelec
  - usgdp
  - mcopper
  - enplanements

- Let's look at each dataset starting with its structure, then apply the Box-Cox transformation to stabilise the variance.

```r
str(usnetelec)
```

```
##  Time-Series [1:55] from 1949 to 2003: 296 334 375 404 447 ...
```

```r
str(usgdp)
```

```
##  Time-Series [1:237] from 1947 to 2006: 1570 1569 1568 1591 1616 ...
```

```r
str(mcopper)
```

```
##  Time-Series [1:564] from 1960 to 2007: 255 260 249 258 244 ...
```

```r
str(enplanements)
```

```
##  Time-Series [1:282] from 1979 to 2002: 21.1 22.9 25.9 24.4 23.4 ...
```

“‘