Assignment 1: Introduction to Time Series

Peter Ye

2024-03-23

```r
# Set seed
set.seed(66)

# Import packages
suppressPackageStartupMessages(library(quantmod))
library(tseries)
library(ggplot2)
library(forecast)
```

Question 1

(a)

$$\begin{aligned} Var(X+Y) &= Var(X) + Var(Y) + 2 * Cov(X,Y) \\ &= Var(X) + Var(Y) + 2 * Corr(X,Y) * \sqrt{Var(X)} * \sqrt{Var(Y)} \\ &= 9 + 4 + 2 * 0.25 * \sqrt{9} * \sqrt{4} \\ &= 16 \end{aligned}$$

(b)

$$\begin{aligned} Cov(X, X+Y) &= Cov(X,X) + Cov(X,Y) \\ &= Var(X) + Corr(X,Y) * \sqrt{Var(X)} * \sqrt{Var(Y)} \\ &= 9 + 0.25 * \sqrt{9} * \sqrt{4} \\ &= 10.5 \end{aligned}$$

(c)

$$Corr(X+Y, X-Y) = \frac{Cov(X+Y, X-Y)}{\sqrt{Var(X+Y) * Var(X-Y)}}$$
$$= \frac{Cov(X,X) - Cov(X,Y) + Cov(Y,X) - Cov(Y,Y)}{\sqrt{(Var(X) + Var(Y) + 2 * Corr(X,Y) * \sqrt{Var(X)} * \sqrt{Var(Y)}) * (Var(X) + Var(Y) - 2 * Corr(X,Y) * \sqrt{Var(X)} * \sqrt{Var(Y)})}}$$
$$= \frac{Var(X) - Var(Y)}{\sqrt{(Var(X) + Var(Y) + 2 * Corr(X,Y) * \sqrt{Var(X)} * \sqrt{Var(Y)}) * (Var(X) + Var(Y) - 2 * Corr(X,Y) * \sqrt{Var(X)} * \sqrt{Var(Y)})}}$$
$$= \frac{9-4}{\sqrt{(9+4+2*0.25*\sqrt{9}*\sqrt{4})*(9+4-2*0.25*\sqrt{9}*\sqrt{4})}}$$
$$= \frac{5}{\sqrt{16*10}}$$
$$= \frac{5}{\sqrt{160}} \approx 0.39528$$

Question 2

$$\begin{aligned} Cov(X+Y, X-Y) &= Cov(X,X) - Cov(X,Y) + Cov(Y,X) - Cov(Y,Y) \\ &= Cov(X,X) - Cov(Y,Y) \\ &= Var(X) - Var(Y) \\ &= 0 \end{aligned}$$

Question 3

(a)

$$\begin{aligned} E(Y_t) &= E(5 + 2t + X_t) \\ &= E(5) + E(2t) + E(X_t) \\ &= 5 + 2t + 0 \\ &= 5 + 2t \end{aligned}$$

(b)

$$\begin{aligned} \gamma_Y(k) = Cov(Y_t, Y_{t-k}) &= Cov(5 + 2t + X_t, 5 + 2(t+k) + X_{t+k}) \\ &= Cov(X_t, X_{t+k}) \\ &= \gamma_k \end{aligned}$$

The terms 5, $2t$, and $2(t+k)$ involve constants and deterministic time trends that do not contribute to the covariance. Therefore, eliminating the constant and deterministic time trend terms is reasonable.

(c)

For Y_t, the mean function is a linear function of time which depends on t. Therefore, it is not stationary because it violates the condition for stationarity, which its mean function is not constant over time.

Question 4

- (1) — (A): The plot (1) shows a series that exhibits fluctuations at a consistent and known interval, indicating a seasonal pattern. Its ACF plot should display a peak at a lag of 12, which is monthly.
- (2) — (C): There is an upward trend observed, leading to a conclusion that its Autocorrelation Function (ACF) plot would exhibit a positive trend, particularly since the autocovariance decreases as the lag increases.
- (3) — (B): The plot (3) shows a series that exhibits fluctuations at a consistent and known interval, indicating a seasonal pattern. Its ACF plot should display a peak at a lag of 10, which is 10 years spanning from plot (3).

Question 5

```r
# Define parameters
mu <- 2.3
sigma <- 1.2

# Generate a normal white noise sample of size 10
sample_10 <- rnorm(n = 10, mean = mu, sd = sigma)

# Calculate and print the sample mean and sample standard deviation for size 10
sample_mean_10 <- mean(sample_10)
sample_sd_10 <- sd(sample_10)
cat("Sample of size 10 - Mean:", sample_mean_10, "Standard Deviation:", sample_sd_10, "\n")
```

```
## Sample of size 10 - Mean: 2.555579 Standard Deviation: 1.119797
```

The sample mean and standard deviation are not very close to the expected values of 2.3 and 1.2, respectively. This is expected due to the small sample size, which can lead to greater variability and less accuracy in estimating population parameters.

```r
# Generate a normal white noise sample of size 10000
sample_10000 <- rnorm(n = 10000, mean = mu, sd = sigma)

# Calculate and print the sample mean and sample standard deviation for size 10000
sample_mean_10000 <- mean(sample_10000)
sample_sd_10000 <- sd(sample_10000)
cat("Sample of size 10000 - Mean:", sample_mean_10000, "Standard Deviation:", sample_sd_10000)
```

```
## Sample of size 10000 - Mean: 2.275309 Standard Deviation: 1.175565
```

The results are much closer to the expected values, demonstrating the law of large numbers, as the sample size increases, the sample mean tends to get closer

Question 6

```r
data("AirPassengers")
data("JohnsonJohnson")
# help("AirPassengers")
# help("JohnsonJohnson")
```

(a)

With the help() function, the descriptions of two data are following:

- AirPassengers: The classic Box & Jenkins airline data. Monthly totals of international airline passengers, 1949 to 1960.
- JohnsonJohnson: Quarterly earnings (dollars) per Johnson & Johnson share 1960–80.

(b)

AirPassengers:

- Start: Jan 1949
- End: Dec 1960
- Frequency: Monthly

```r
# Start date
start_date_ap <- start(AirPassengers)
# End date
end_date_ap <- end(AirPassengers)
# Frequency
frequency_ap <- frequency(AirPassengers)

# Print the results
cat("AirPassengers Start Date:", start_date_ap, "\n")
```

```
## AirPassengers Start Date: 1949 1
```

```r
cat("AirPassengers End Date:", end_date_ap, "\n")
```

```
## AirPassengers End Date: 1960 12
```

```r
cat("AirPassengers Frequency:", frequency_ap, "data points per year\n")
```

```
## AirPassengers Frequency: 12 data points per year
```

JohnsonJohnson:

- Start: 1st quarter 1960
- End: 4th quarter 1980
- Frequency: Quarterly

```r
# Start date
start_date_jj <- start(JohnsonJohnson)
# End date
end_date_jj <- end(JohnsonJohnson)
# Frequency
frequency_jj <- frequency(JohnsonJohnson)

# Print the results
cat("JohnsonJohnson Start Date:", start_date_jj, "\n")
```

```
## JohnsonJohnson Start Date: 1960 1
```

```r
cat("JohnsonJohnson End Date:", end_date_jj, "\n")
```

```
## JohnsonJohnson End Date: 1980 4
```

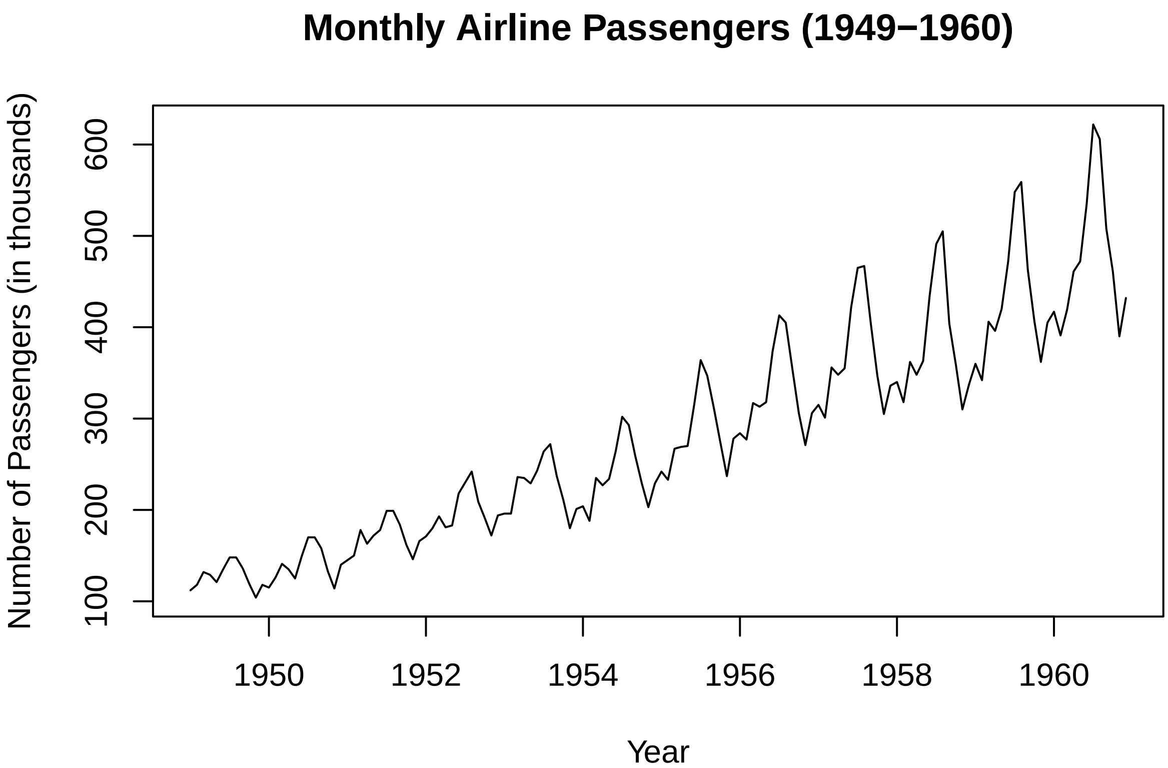
```r
cat("JohnsonJohnson Frequency:", frequency_jj, "data points per year\n")
```

```
## JohnsonJohnson Frequency: 4 data points per year
```

(c)

AirPassengers:

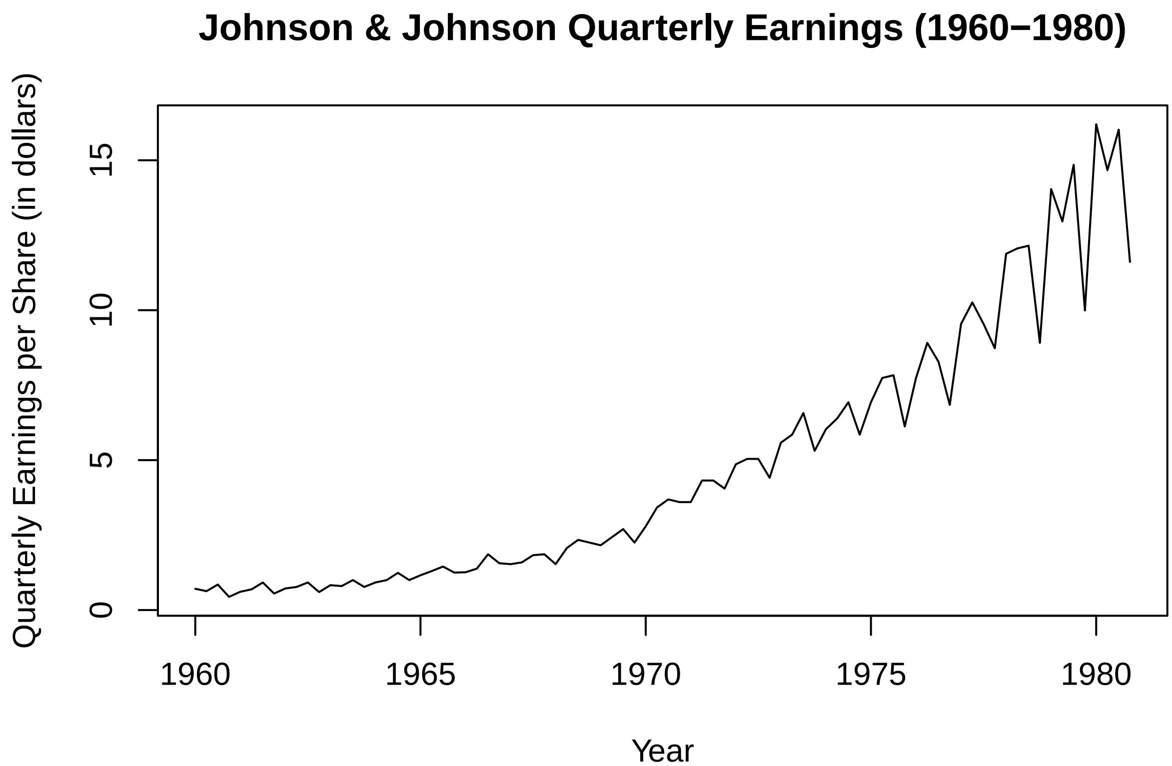
```r
plot(AirPassengers, xlab = "Year", ylab = "Number of Passengers (in thousands)",
     main = "Monthly Airline Passengers (1949-1960)")
```



- Trend: It shows a clear upward trend, indicating a consistent increase in the number of international airline passengers over the 12-year period. This demonstrates growth in the airline industry or increased popularity of air travel during this time.
- Seasonality: There is a noticeable seasonal pattern within each year, with peaks typically in the middle of the year (June/July) and troughs at the beginning and end of the year. This seasonality reflects higher travel demand during summer months, possibly due to vacation travel.
- Cyclic: There is no clear cyclic pattern

JohnsonJohnson:

```r
plot(JohnsonJohnson, xlab = "Year", ylab = "Quarterly Earnings per Share (in dollars)",
     main = "Johnson & Johnson Quarterly Earnings (1960-1980)")
```



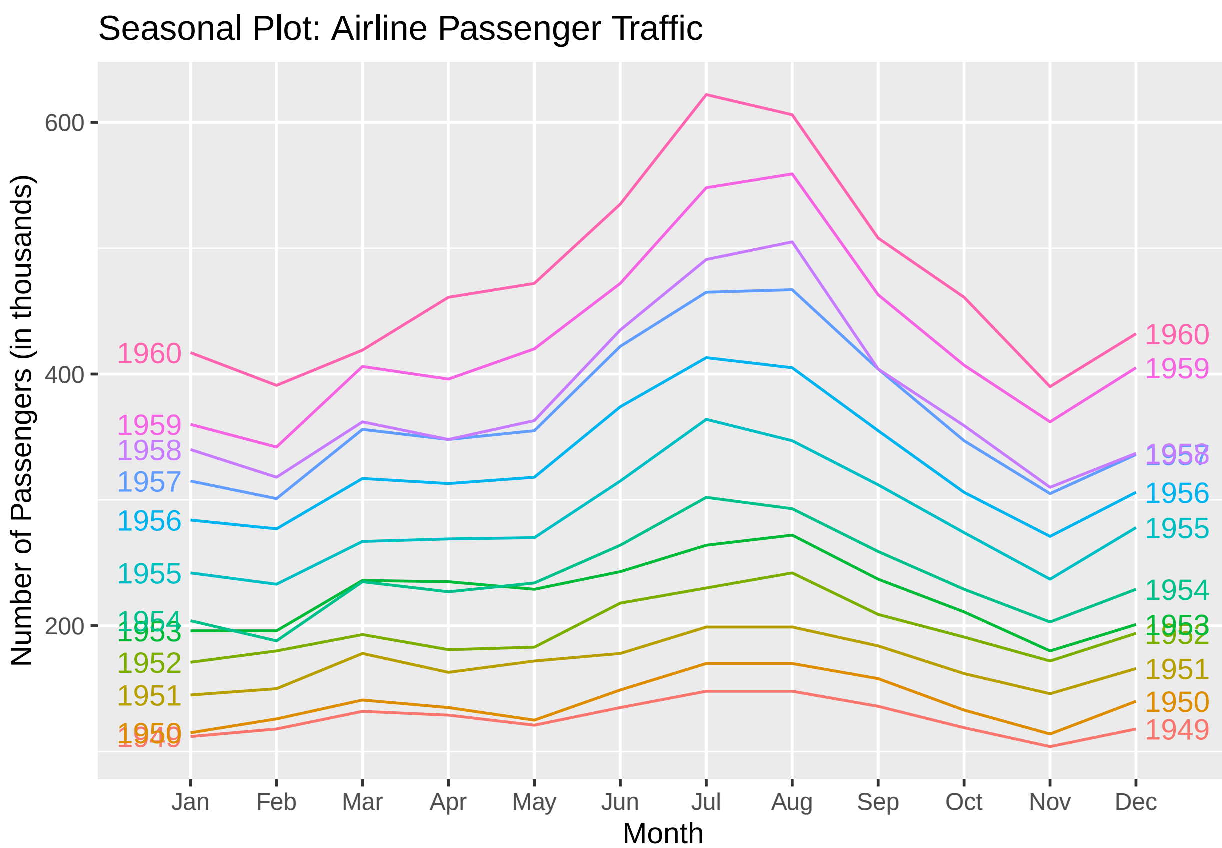
- Trend: Similar to AirPassengers data, it shows an upward trend over time, indicating growth in Johnson & Johnson's quarterly earnings per share from 1960 to 1980. This reflects the company's financial growth and possibly overall economic growth during this period.
- Seasonality: This time series also exhibits some seasonality, though the pattern may not be as pronounced or as regular as in the AirPassengers data.The seasonality in earnings could be related to fluctuations in sales due to various factors, including market demand and fiscal policies.
- Cyclic: There is a sign of cyclic pattern that earning appearing to become more volatile over time, with larger fluctuations in more recent quarters. This could indicate increasing uncertainty in the market or changes in the business environment affecting the company.

(d)

AirPassengers:

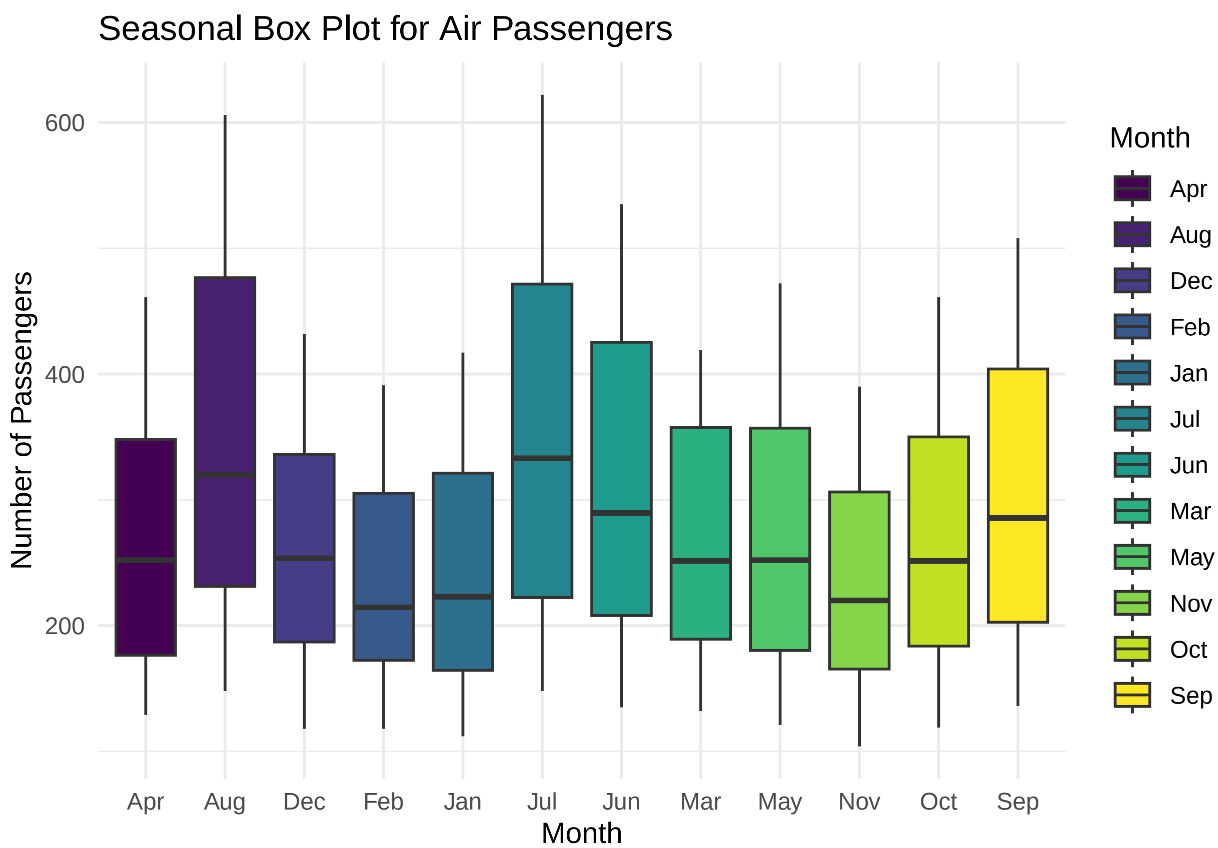
```
# Create the ggseasonplot
p <- ggseasonplot(AirPassengers, year.labels = TRUE, year.labels.left = TRUE) +
  ylab("Number of Passengers (in thousands)") +
  ggtitle("Seasonal Plot: Airline Passenger Traffic")

# Print the plot
print(p)
```




```r
# Create a dataframe with a 'Month' column for the x-axis and 'Passengers' for the y-axis
air_passengers_df <- data.frame(
  Month = factor(month.abb[cycle(AirPassengers)]), # 'month.abb' gives month abbreviations
  Passengers = as.numeric(AirPassengers)
)

# Plot the boxplot
ggplot(air_passengers_df, aes(x = Month, y = Passengers, fill = Month)) +
  geom_boxplot() +
  scale_fill_viridis_d() +
  labs(title = "Seasonal Box Plot for Air Passengers",
       x = "Month",
       y = "Number of Passengers") +
  theme_minimal()
```

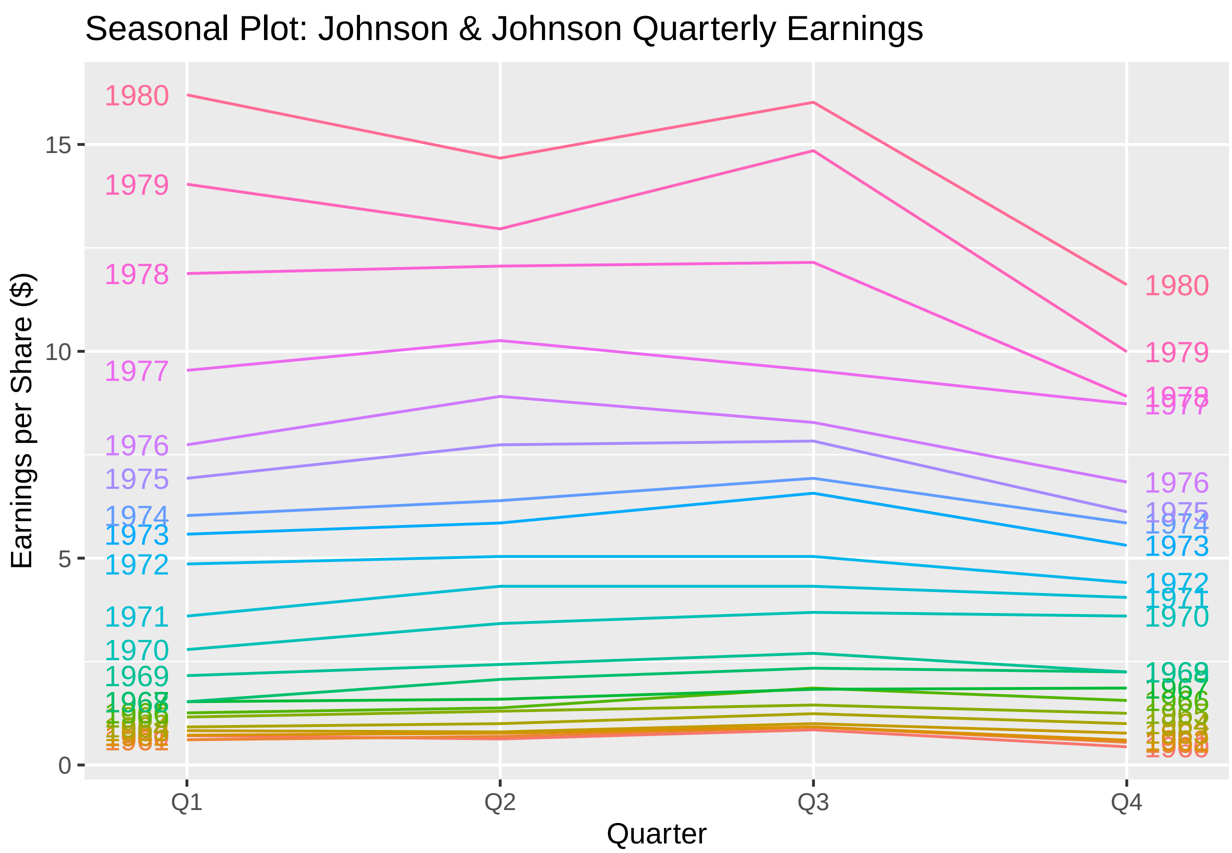


- Seasonality: There's a clear seasonal pattern, with the middle months (especially June, July, August) showing higher median values and a larger interquartile range (IQR), which corresponds to the peak travel season in summer.
- Trend: The range of the boxes and whiskers seems to increase as the months progress from January to August, suggesting a growing number of passengers as the year progresses towards the peak season.
- Outlier: There are a few outliers, especially in the later months of the year, which may represent years with exceptionally high passenger numbers.

JohnsonJohnson:

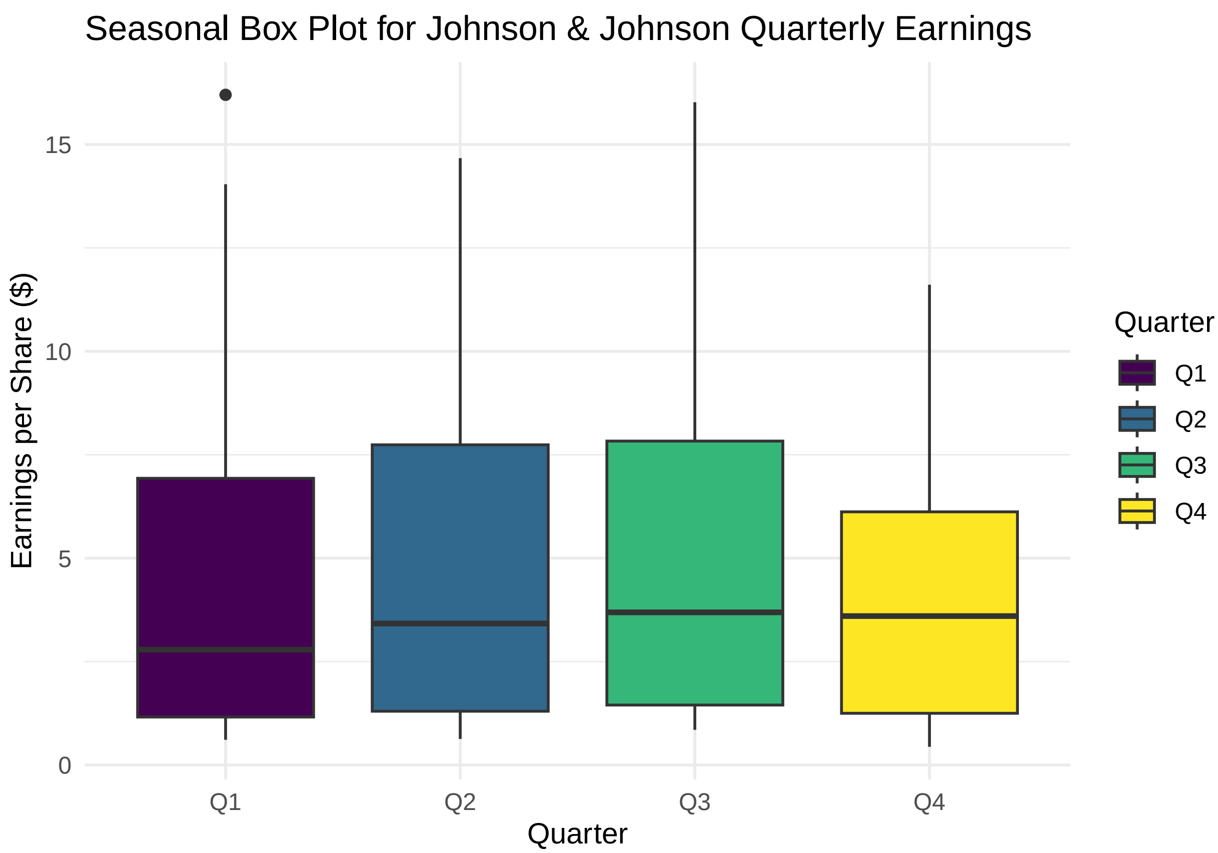
```r
# Create the ggseasonplot
p <- ggseasonplot(JohnsonJohnson, year.labels = TRUE, year.labels.left = TRUE) +
  ylab("Earnings per Share ($)") +
  ggtitle("Seasonal Plot: Johnson & Johnson Quarterly Earnings")

# Print the plot
print(p)
```




```r
# Create a dataframe with a 'Quarter' column for the x-axis and 'Earnings' for the y-axis
jj_df <- data.frame(
  Quarter = factor(rep(c("Q1", "Q2", "Q3", "Q4"), length.out = length(JohnsonJohnson))),
  Earnings = as.numeric(JohnsonJohnson)
)

# Plot the boxplot
ggplot(jj_df, aes(x = Quarter, y = Earnings, fill = Quarter)) +
  geom_boxplot() +
  scale_fill_viridis_d() +
  labs(title = "Seasonal Box Plot for Johnson & Johnson Quarterly Earnings",
       x = "Quarter",
       y = "Earnings per Share ($)") +
  theme_minimal()
```



- Seasonality: There's less pronounced seasonality in the earnings data compared to the airline passengers. However, there might be a slight increase in median values in Q2 and Q3 compared to Q1 and Q4.
- Trend: The spread of the data, as indicated by the IQR, is quite stable across the quarters, with Q3 showing a slightly higher variance.
- Outlier: There's a noticeable outlier in Q1, which might represent an unusual spike in earnings for that particular quarter in a specific year.