Arrivals Data Set

The arrivals data set comprises quarterly international arrivals (in thousands) to Australia from Japan, New Zealand, UK and the US.

(a)

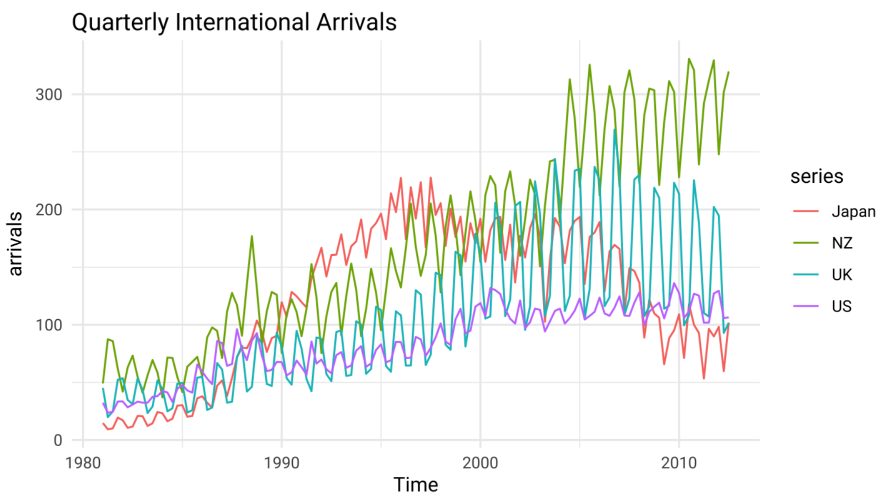

Figure 1: Quarterly International Arrivals (in thousands) to Australia from Japan, New Zealand, UK and the US

As seen in figure 1, the number of arrivals from all four countries increased overall, even though some fluctuated over the years. The number of arrivals from Japan to Australia steadily increased until around 1998, peaking between 1995 and 2000 at approximately 220 000 before declining afterward. New Zealand shows a steady increasing trend despite some fluctuations. Overall, it has an upward trend and shows the highest number of people arriving in Australia. This could be because New Zealand is close to Australia, but despite that, the high results for New Zealand are notable, considering its population size of around 5 million people. Regarding the UK, it also has an upward trend; however, after around 2007, the trend gradually decreases. The US has an overall upward trend as well, which is consistent but increases very gradually.

In this analysis, the terms Q1, Q2, Q3, and Q4 are used to represent the four quarters of the year in Australia, each corresponding to specific weather seasons. To avoid confusion, these terms are defined as follows:

Q1 (January to March) corresponds to summer.
Q2 (April to June) corresponds to fall.
Q3 (July to September) corresponds to winter.
Q4 (October to December) corresponds to spring.

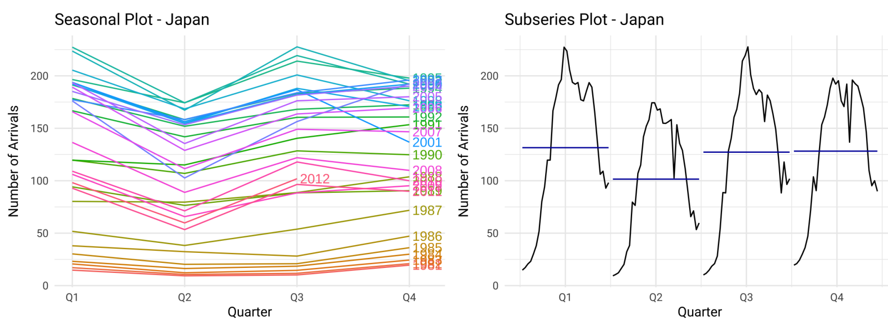

Figure 2: Seasonal and Subseries Plots - Japan

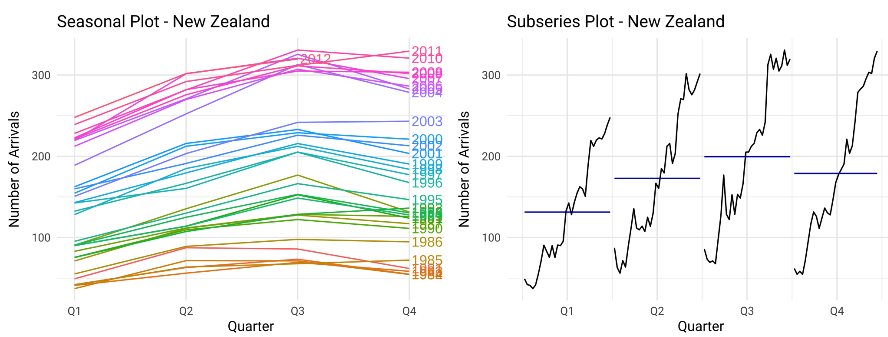

Figure 3: Seasonal and Subseries Plots - New Zealand

Japan clearly show a strong seasonal pattern in arrivals to Australia, as seen in figure 2. People from Japan tend to visit Australia mostly during the 1st and 3rd quarters, which correspond to summer and winter. They visit less during spring and fall. The seasonal plot clearly shows that arrivals peak in the 3rd quarter and drop in the 2nd quarter. In the subseries plot, quarters 1, 3, and 4 show a similar number of arrivals, with the 2rd quarter having the lowest average. The blue horizontal lines in the subseries plot highlight that the 2rd quarter consistently has the lowest number of arrivals.

Looking at figure 3, we can see that New Zealand also shows a seasonal pattern. The seasonal plot shows that most people arrive from New Zealand in the 3rd and 4th quarters, which correspond to winter and spring. In the subseries plot, the most visited time is the 3rd quarter (winter). Fall and spring have similar average numbers of arrivals, while the 1st quarter has the lowest.

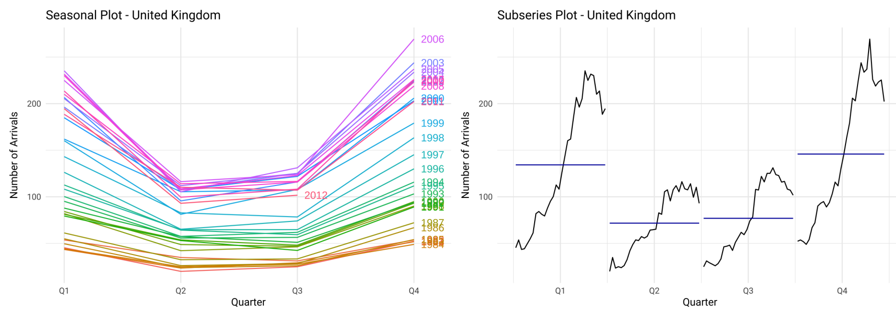

Figure 4: Seasonal and Subseries Plots - United Kingdom

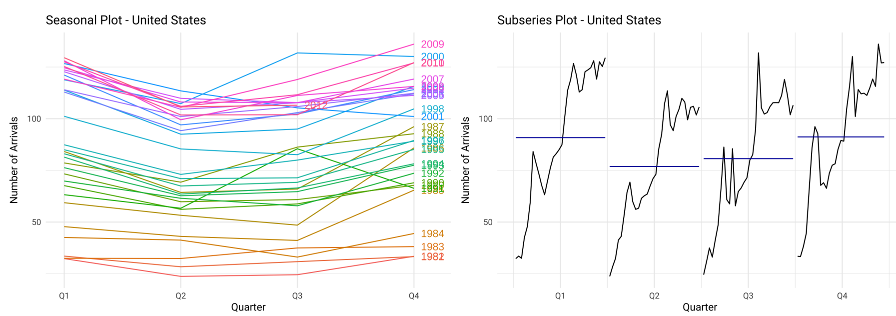

Figure 5: Seasonal and Subseries Plots - United States

Looking at figure 4, we can notice that the United Kingdom has a very strong seasonal pattern. In the seasonal plot, it is clear that people from the UK consistently travel to Australia during the first and fourth quarters, which correspond to summer and spring. The subseries plot also shows this trend. The difference in arrivals between the first and fourth quarters and the second and third quarters is substantial, with a gap of around 70k.

Looking at figure 5, the United States does not show as strong a seasonal pattern as other countries, with some outliers. In the seasonal plot, we can observe that in the early 1980s, there was not a very noticeable pattern regarding when people from the U.S. traveled to Australia. However, starting from the 1990s, people tend to travel mostly during the first and fourth quarters, corresponding to summer and spring. Additionally, we can notice the outlier in 2000, where a significant number of people traveled to Australia during winter and spring. It could have been related to the 2000 Summer Olympics, which took place from September 15 to October 1 in Australia. In the subseries plot, people from the U.S. typically travel during summer and spring, as I mentioned earlier, with the lowest numbers during the fall. I believe that the outlier in 2000 made a significant difference in the average number of arrivals.

(b)

As for me, I expected people to travel to Australia during the summer and spring, but as I reported before, the number of people arriving in Australia during the winter is also quite large. People from the United Kingdom and the United States mostly travel to Australia during the summer and spring. People from Japan travel during both summer and winter, while people from New Zealand typically travel during the winter, which is interesting. Another thing to note is that the northern part of Australia is quite warm during the winter, so people might go there since it is not cold at all. This could explain why there are so many people arriving from Japan and New Zealand during the winter. In most countries, there was not a clear seasonal pattern in the 1980s, but patterns started to form in the 1990s. An unusual pattern was the significant drop in the number of people traveling from Japan to Australia starting around 1998, while other countries showed an upward trend. It was also interesting to see the unusually high number of people arriving from the United States to Australia in 2000 during the winter, which did not follow the typical seasonal pattern.

Consecutive Trading Days of the Dow Jones Index

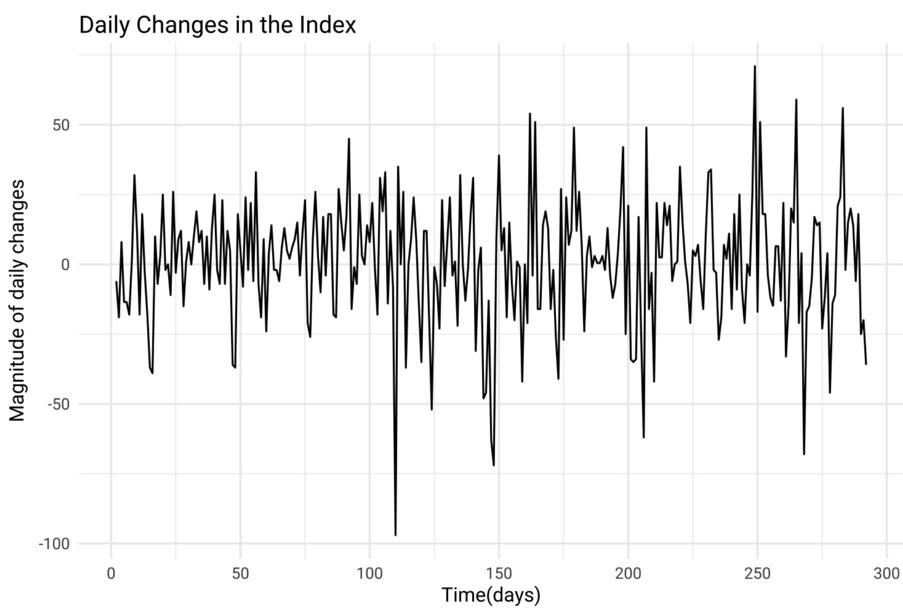

Figure 6: Daily Changes in the Dow Jones Index

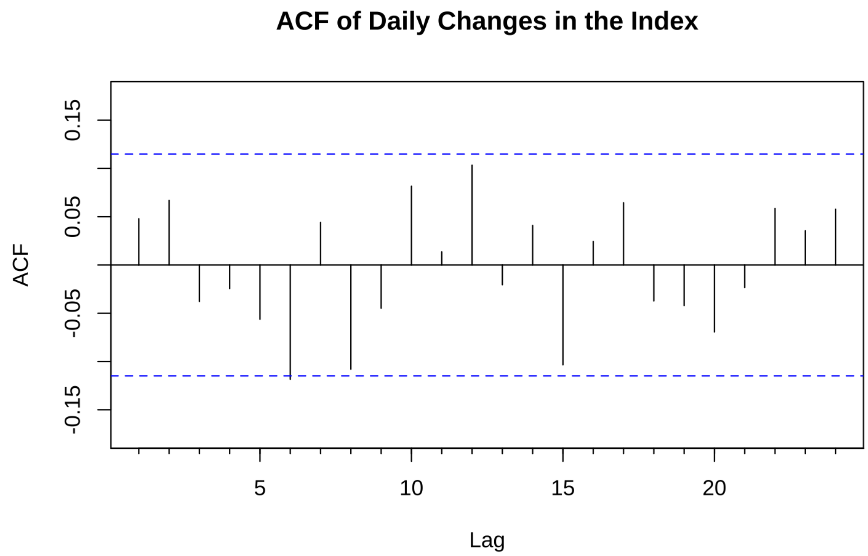

Figure 7: ACF of Daily Changes in the Dow Jones Index

The plot contains 292 consecutive trading days of the Dow Jones Index.

Looking at figure 6, we can see that there is no noticeable clear pattern, just a lot of variance from negative to positive around zero. Over the 292 consecutive days, we do not observe much pattern. Additionally, looking at figure 7, our assumption about the lack of a particular pattern is confirmed, as the majority of autocorrelations fall within the blue lines, confirming that this is white noise. There is only one line touching the blue line, meaning the majority fall inside the blue lines. Therefore, we can conclude that the changes during these days are random and unpredictable, and our analysis suggests that the changes in the Dow Jones Index are white noise.

Data dole, usdeaths and bricksq

This report looks at three time series datasets (dole, usdeaths, and bricksq) and explores whether transforming the data is good choice. To do this, I used the BoxCox.lambda() function to find the best transformation for each series. After applying the recommended transformations, I compared the ACF plots for the original and transformed data. This helped me see if the transformations actually improved the series.

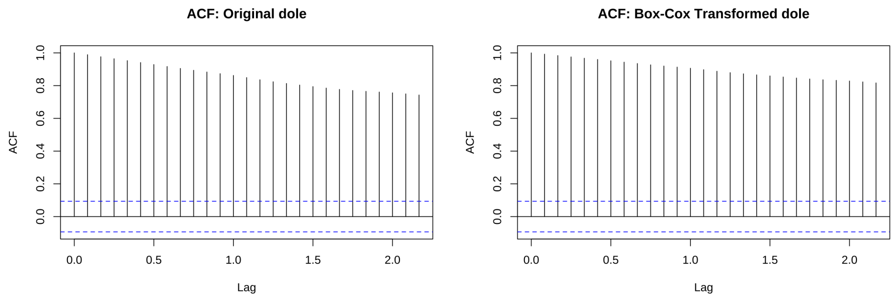

Figure 8: ACF of dole

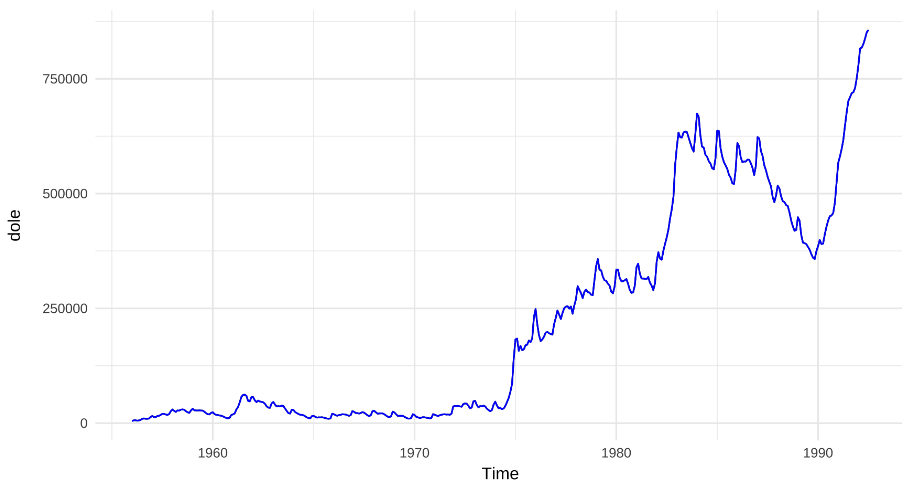

Figure 9: Autoplot of Dole Dataset

Looking at figure 8, we can see that the Box-Cox transformation did not significantly improve the ACF, as it shows a similar pattern to the original data. This suggests that using the original data without transformation may be more appropriate. In figure 9, we observe an overall upward trend, with a sharp spike around 1982 followed by a decline, and then another rise starting around 1990.

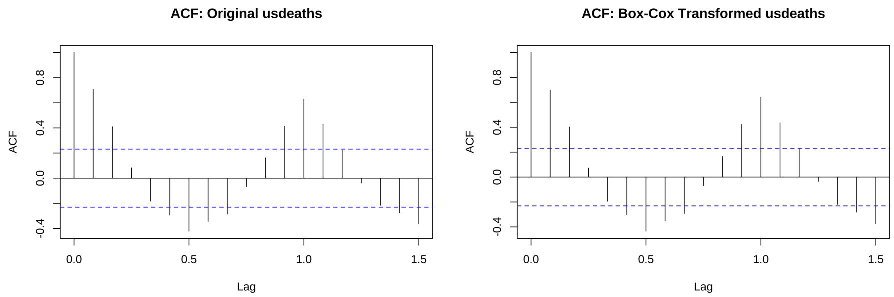

Figure 10: ACF of usdeaths

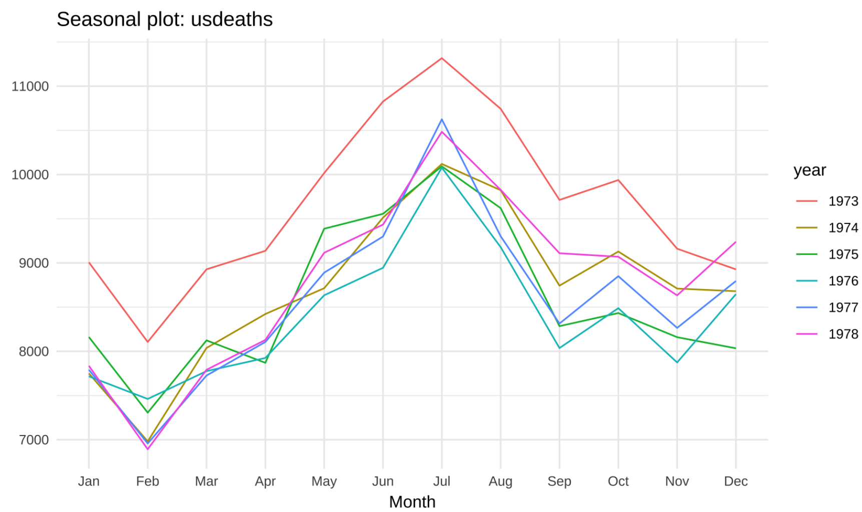

Figure 11: Seasonal Plot of USdeaths Dataset

Looking at figure 10, we observe a clear seasonal pattern in the usdeaths dataset. Additionally, the transformation applied did not significantly change the ACF plot, indicating that it had minimal impact. Therefore, it is better to use the original data without transformation. From the seasonal plot of usdeaths (figure 11), we can see that the number of deaths peaks in July and reaches its lowest point in February.

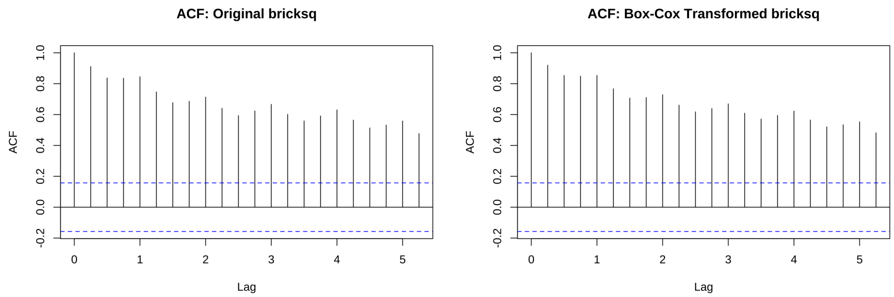

Figure 12: ACF of bricksq

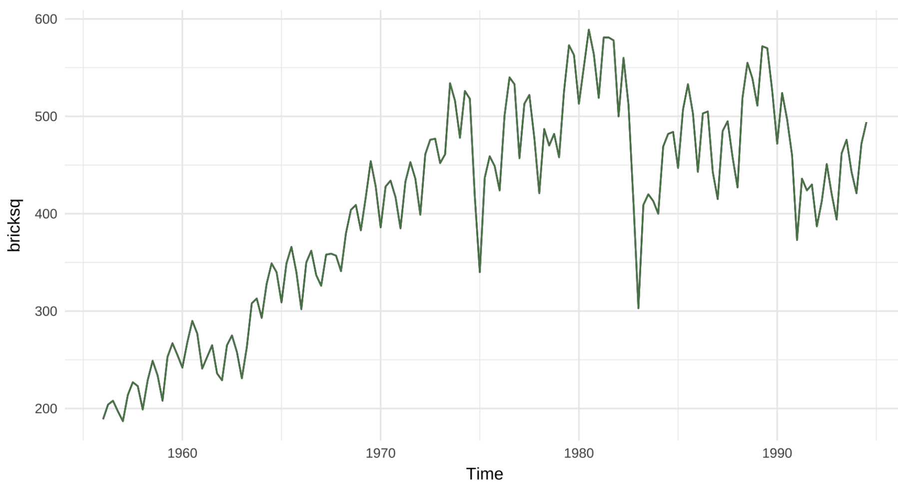

Figure 13: Autoplot of Bricksq Dataset

Looking at figure 12, we can see that the Box-Cox transformation did not significantly change the autocorrelation of the data. There is no clear evidence of a seasonal pattern in either the original or transformed series. For this reason, we proceed with the original data and use it to create the plot shown in figure 13. Looking at the plot, there is an upward trend in the earlier decades, but it is not consistent increase. After a certain point, the values fluctuate and even decrease at times.

Australian Beer Production

In this report, I will calculate the residuals from a seasonal naive forecast applied to the quarterly Australian beer production data from 1992.

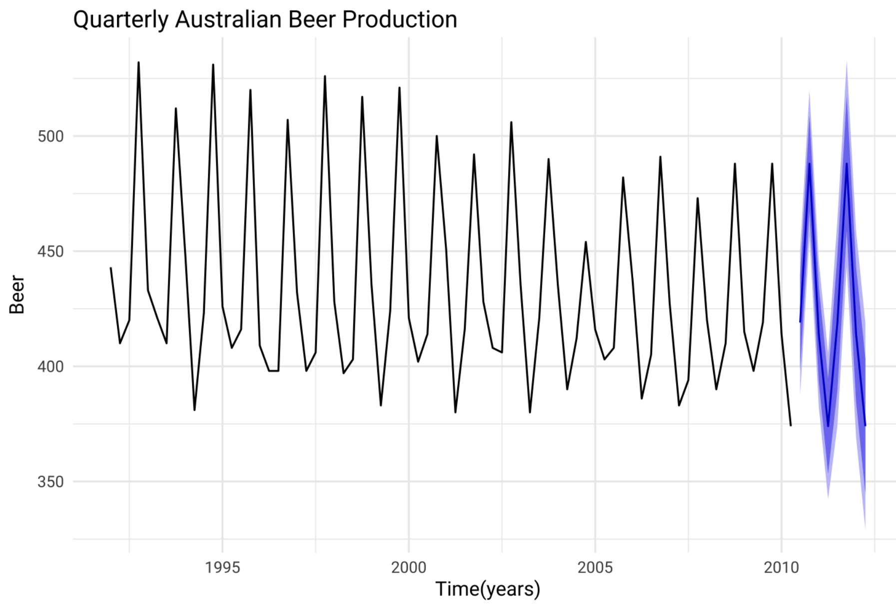

Figure 14: Quarterly Australian Beer Production

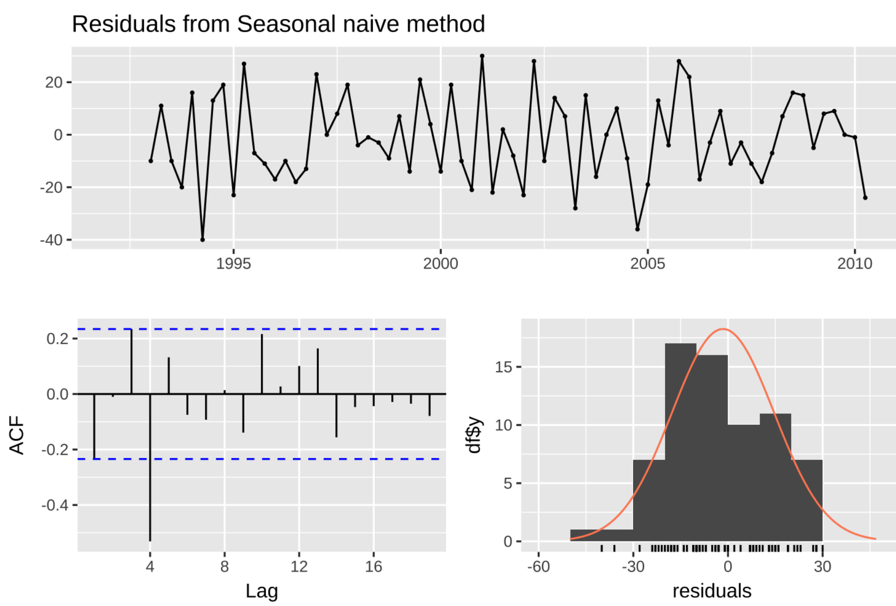

Figure 15: Residuals for Seasonal Naive Method

Looking at figure 14, we can see a seasonal pattern with periodic increases and drops. Overall, there seems to be a slight downward trend. The blue shaded area represents the forecast, capturing uncertainty, and it follows the seasonal pattern. This means we are more confident about the seasonal pattern in the forecast, but predicting the overall increase or decrease is harder.

Looking at the residual plots in figure 15, we see that the residuals are around zero, staying fairly consistent over time. The ACF plot shows that most lags are within the confidence bounds, except for lag 4, which shows some autocorrelation (not white noise). Finally, the histogram of residuals looks close to normally distributed.

Daily Closing IBM Stock Prices

(a)

Figure 16: Autoplot of Daily Closing IBM Stock Prices

Looking at figure 16, we can see that there is no clear trend or seasonality. The series starts with an increase, then around 120, it begins to decrease, followed by another increase around 270.

(b)

As shown in figure 17, the black line represents the training set, which consists of the first 300 observations and is used to fit the model. The red line represents the test set, which contains the 69 observations and is used to evaluate the model's forecasting.

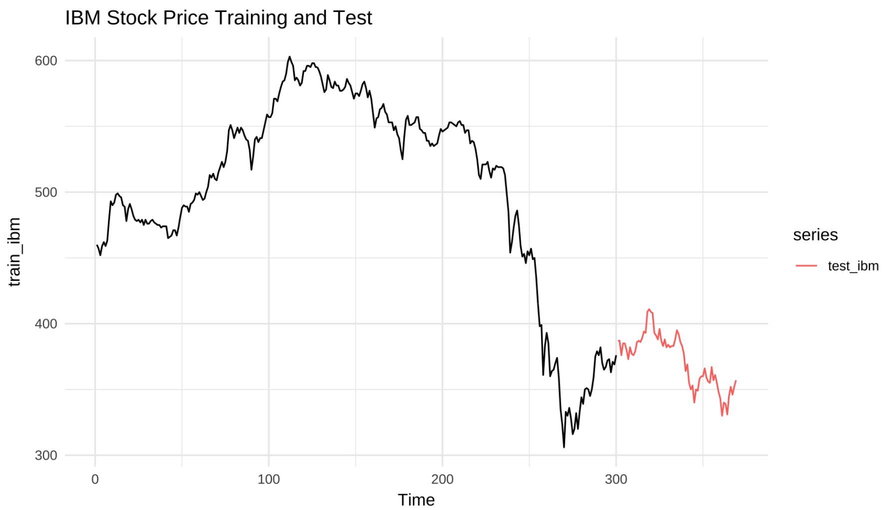

Figure 17: Training and Test (Daily Closing IBM Stock Prices)

Figure 18: Forecast Methods (Daily Closing IBM Stock Prices)

(c)

As shown in figure 18, the mean forecast overestimated the stock price, predicting values that are too high. This outcome is understandable since IBM's stock price experiences both upward and downward movements, making it difficult for the mean method to provide accurate predictions. Both the naive and seasonal naive forecasts performed similarly, producing predictions that were closer to the test set but still somewhat off in capturing the precise direction of the price changes. The drift forecast performed the best, as it was able to account for the slight downward trend present in the data.

(d)

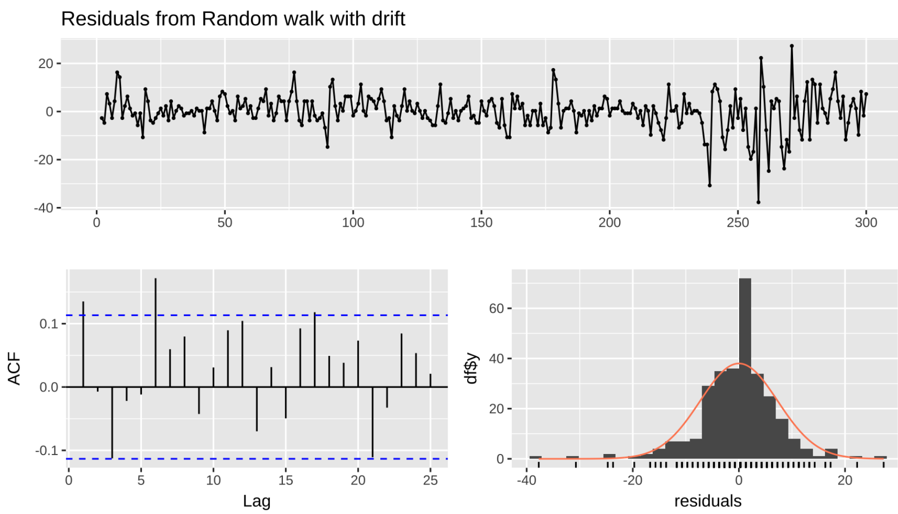

Figure 19: Residuals of Drift

Looking at the figure 19, the residuals from the drift model are mostly around zero and have roughly constant variance, which is a good sign. But we can see some significant autocorrelation at certain lags and a few large outliers, suggesting the model does not fully capture the data's structure. The residuals are not perfectly white noise, meaning there might still be some patterns that the model cannot fully predict.

Men's 400 Meters Final in Olympic Games

The data set mens400 contains the winning times (in seconds) for the men's 400 meters final in each Olympic Games from 1896 to 2016.

(a)

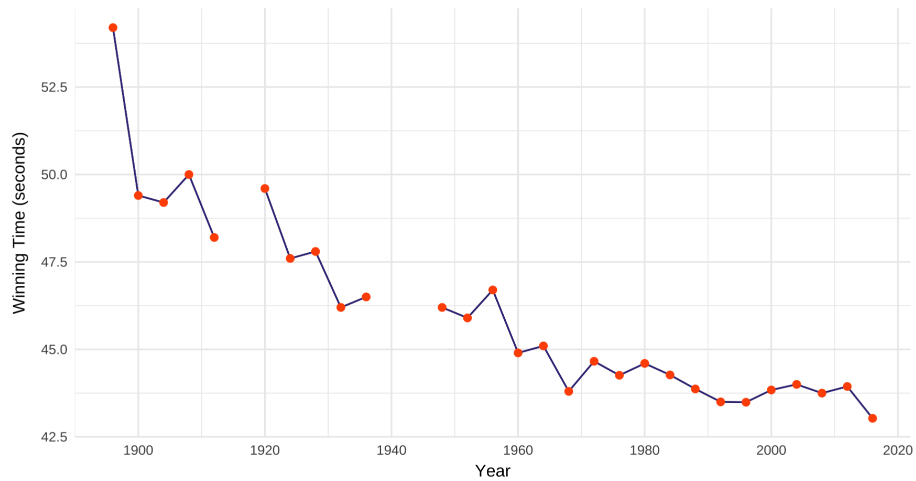

Figure 20: Autoplot of Men's 400 Meters Final

As shown in figure 20, the overall winning times (in seconds) for the men's 400 meters final in each Olympic Games are decreasing. Another thing to note is that there are three missing values.

(b)

Statistic	Value
R-squared	0.823
Adjusted R-squared	0.817
F-statistic	121.2
Prob (F-statistic)	2.75e-11
Residual Std. Error	1.136
Degrees of Freedom	26
Intercept Estimate	50.31
Trend Estimate	-0.258

Table 1: Model Summary Statistics for Regression Line on Men's 400m Data

According to table 1, the R-squared value for this model is 0.823, meaning about 82.3% of the variation in men's 400m winning times can be explained by the linear trend over time. The adjusted R-squared of 0.817 shows the model still has strong explanatory power, even after accounting for the number of predictors. The highly significant p-value confirms that the trend variable is a strong predictor of winning times. The estimated intercept is 50.31 seconds, and the trend coefficient of -0.258 means that winning times have decreased by around 0.258 seconds per year. The residual standard error is 1.136, which means the typical difference between observed times and the fitted line is just over one second. The model is based on a solid number of observations with 26 degrees of freedom, although three years were skipped due to missing data.

(c)

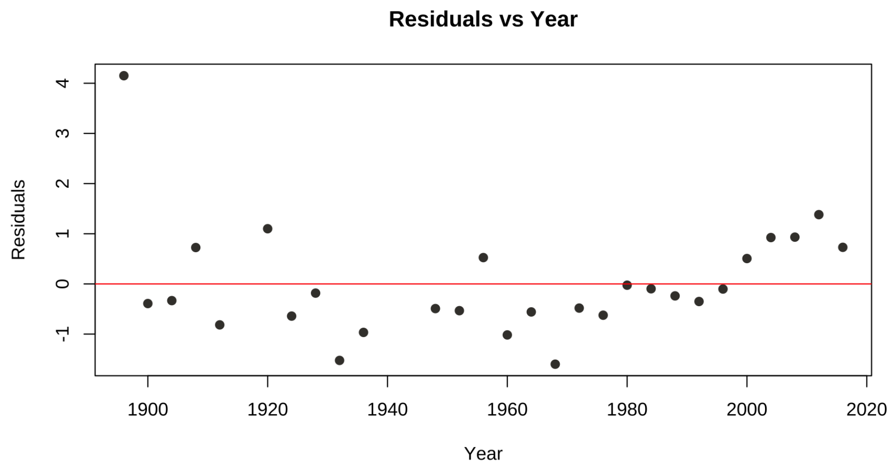

Figure 21: Residuals of mens400

Overall, looking at the residual plot in figure 21, the linear trend model seems to work pretty well for most of the data. However, sometimes model tends to underestimate, especially in recent years, along with a few outliers. But there are no serious violations of linear model assumptions.

(d)

The predicted winning time for the men's 400 meters final in the 2020 Olympics is 43.85038 seconds. The actual winning time in the Tokyo 2020 Olympics was 43.85 seconds, achieved by Steven Gardiner. Our prediction was extremely close, making it highly accurate.