Assignment 2: Time Series Decomposition

Peter Ye

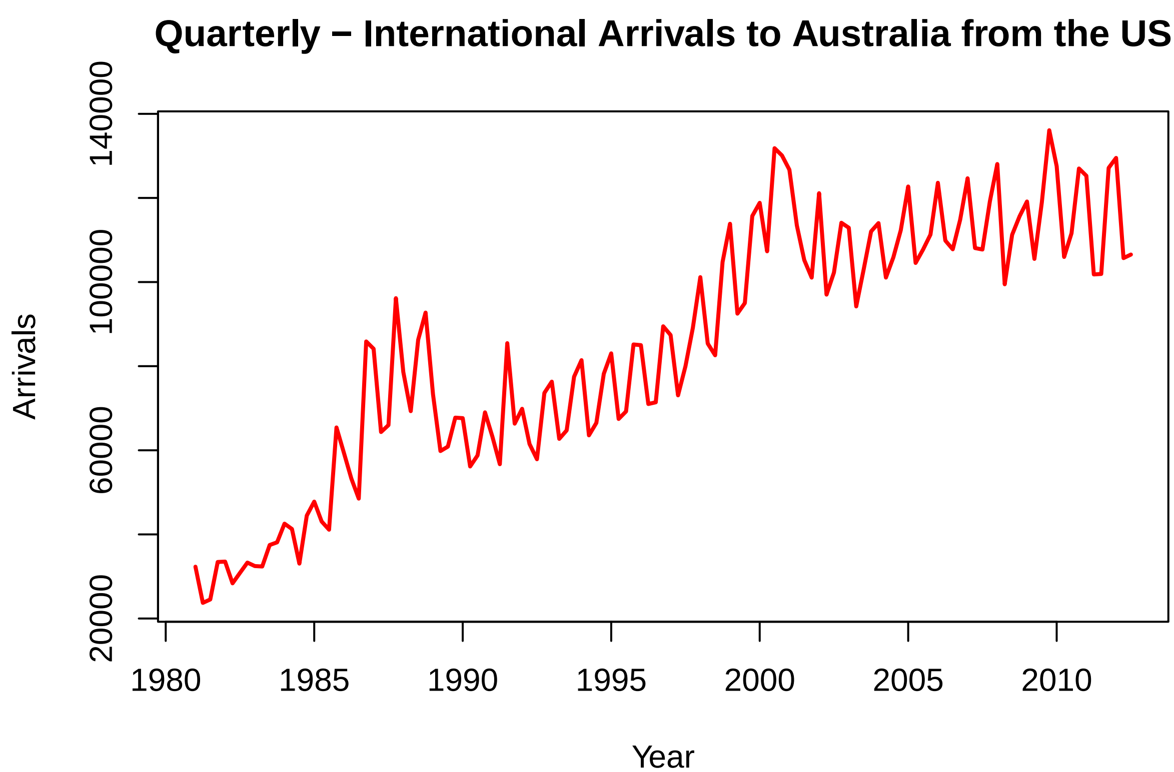
2024-03-31

Question 1

```r
# Load visitors.rda data
load("visitors.rda")

# Create a time series object
visitors_Quarterly <- ts(visitors$Arrivals, start=c(1981,1), frequency=4)
```

```r
# Plot the graph
plot(visitors_Quarterly, main = "Quarterly - International Arrivals to Australia from the US",
     ylab = "Arrivals", xlab = "Year", col = "red", lwd = 2, cex.main = 1.2, cex.axis = 0.9,
     cex.lab = 1, font.main = 2)
```

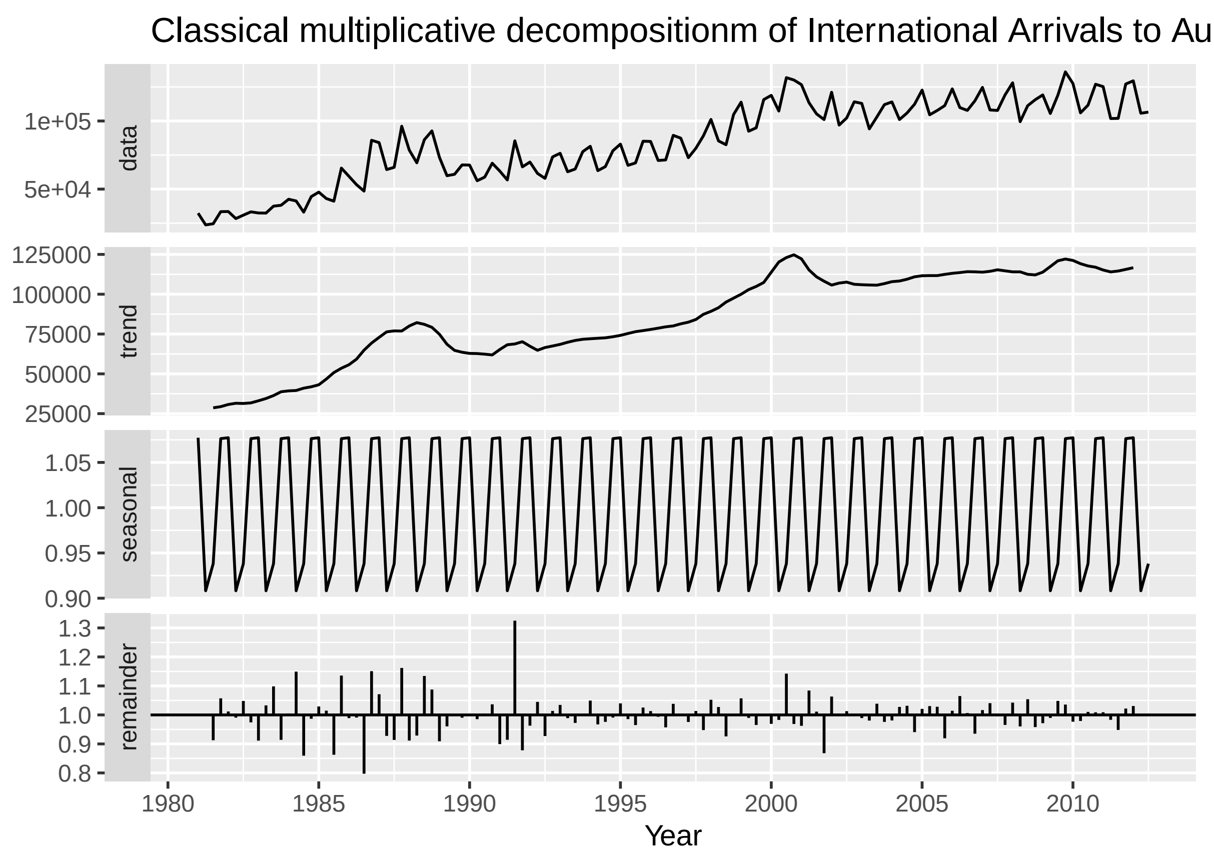


- Trend: : There is a clear upward trend over the period, indicating that the number of arrivals has generally increased.
- Seasonality: There appears to be a regular pattern within each year, which suggests seasonality. Peaks and troughs recur at similar intervals, suggesting certain times of the year have consistently higher or lower numbers of arrivals.
- Cyclic: There are broader fluctuations that occur over several years, which may suggest economic cycles or other long-term factors affecting travel habits.

Question 2

The previous graph demonstrates the amplitude of seasonal pattern is increasing as the average level of the seasonal data gets larger. Therefore, adopting multiplicative Holt-Winters method is more appropriate.

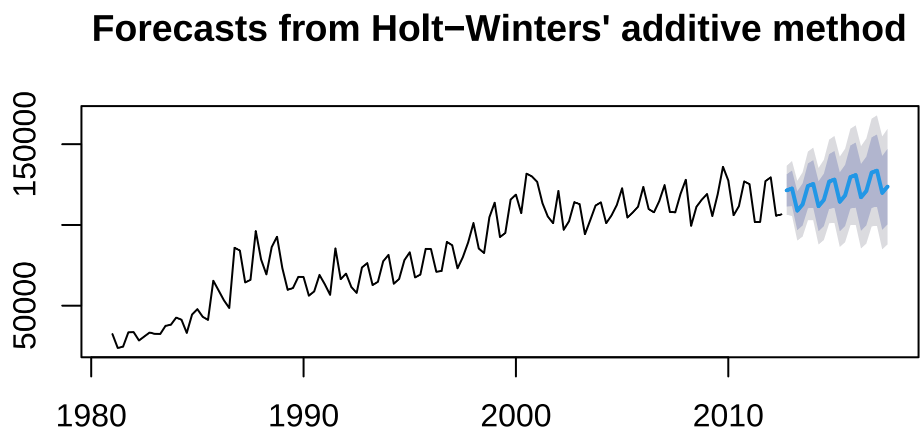
```
# Multiplicative decompositionm
visitors_Quarterly %>% decompose(type = "multiplicative") %>%
  autoplot() + xlab("Year") +
  ggtitle("Classical multiplicative decompositionm of International Arrivals to Australia from the US")
```



Question 3

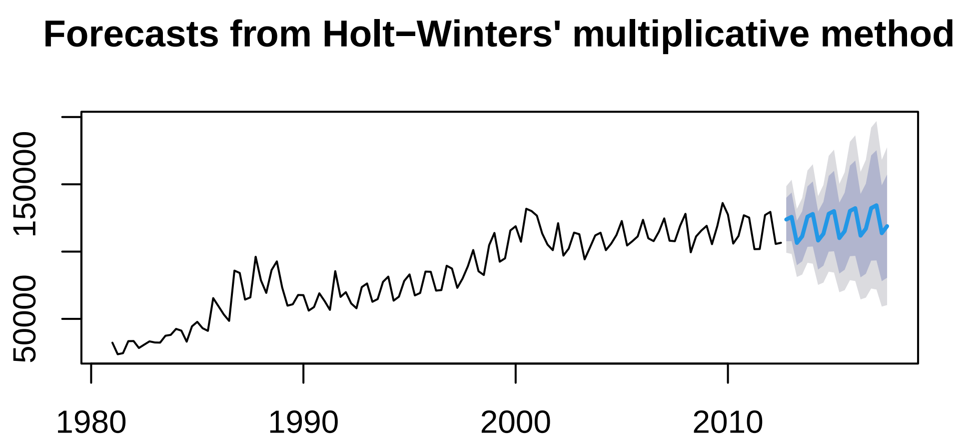
- Linear trend with additive seasonality

```
add_hw <- hw(visitors_Quarterly, h=20 , seasonal="additive",
           damped=FALSE)
plot(add_hw)
```



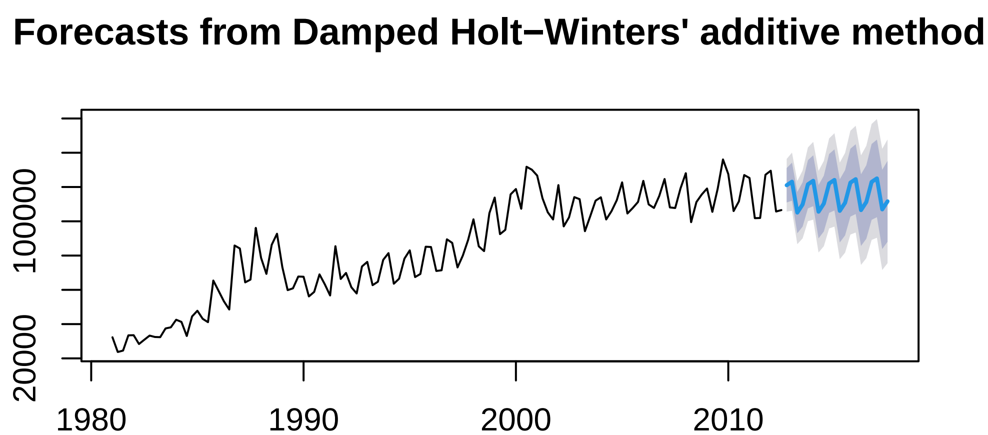
- Linear trend with multiplicative seasonality

```
multi_hw <- hw(visitors_Quarterly, h=20 , seasonal="multiplicative",
           damped=FALSE)
plot(multi_hw)
```



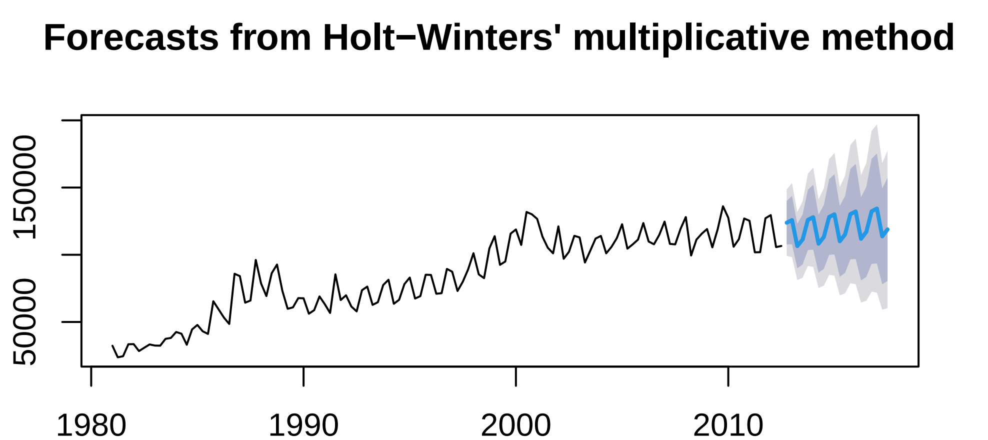
- Linear trend with additive seasonality and damping

```
add_damp_hw <- hw(visitors_Quarterly, h=20 , seasonal="additive",
           damped=TRUE)
plot(add_damp_hw)
```



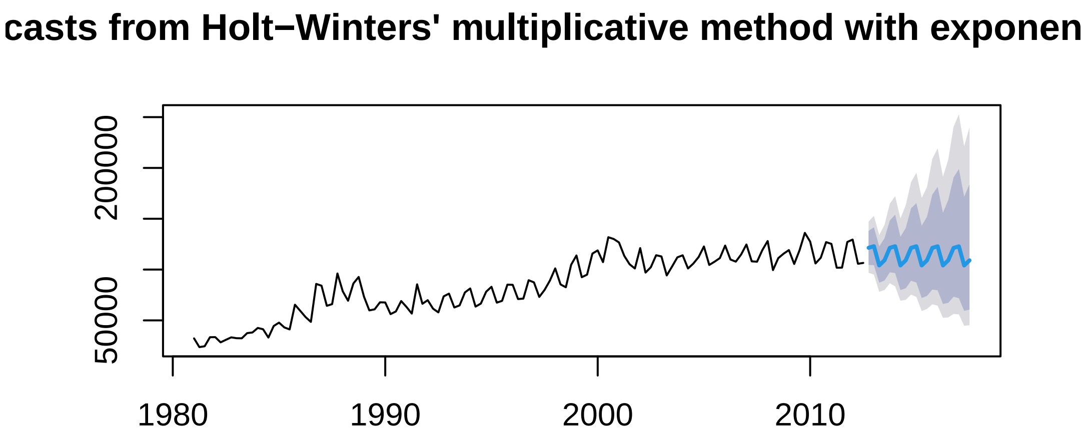
- Linear trend with multiplicative seasonality and damping

```
multi_damp_hw <- hw(visitors_Quarterly, h=20 , seasonal="multiplicative",
           damped=TRUE)
plot(multi_hw)
```



- Exponential trend with multiplicative seasonality

```
multi_expo_hw <- hw(visitors_Quarterly, h=20 , seasonal="multiplicative",
           damped=FALSE, exponential = TRUE)
plot(multi_expo_hw)
```



Question 4

```
# Fetch the RMSE value from each method
rmse_add_hw <- accuracy(add_hw)[1, "RMSE"]
rmse_multi_hw <- accuracy(multi_hw)[1, "RMSE"]
rmse_add_damp_hw <- accuracy(add_damp_hw)[1, "RMSE"]
rmse_multi_damp_hw <- accuracy(multi_damp_hw)[1, "RMSE"]
rmse_multi_expo_hw <- accuracy(multi_expo_hw)[1, "RMSE"]

# Create a data frame
methods_df <- data.frame(
  Method = c("add_hw", "multi_hw", "add_damp_hw", "multi_damp_hw", "multi_expo_hw"),
  RMSE = c(rmse_add_hw, rmse_multi_hw, rmse_add_damp_hw, rmse_multi_damp_hw, rmse_multi_expo_hw)
)

# Create a table
knitr::kable(methods_df, caption = "Forecasting Methods and Their RMSE") %>%
  kable_styling(bootstrap_options = c("striped", "hover"))
```

Table 1: Forecasting Methods and Their RMSE

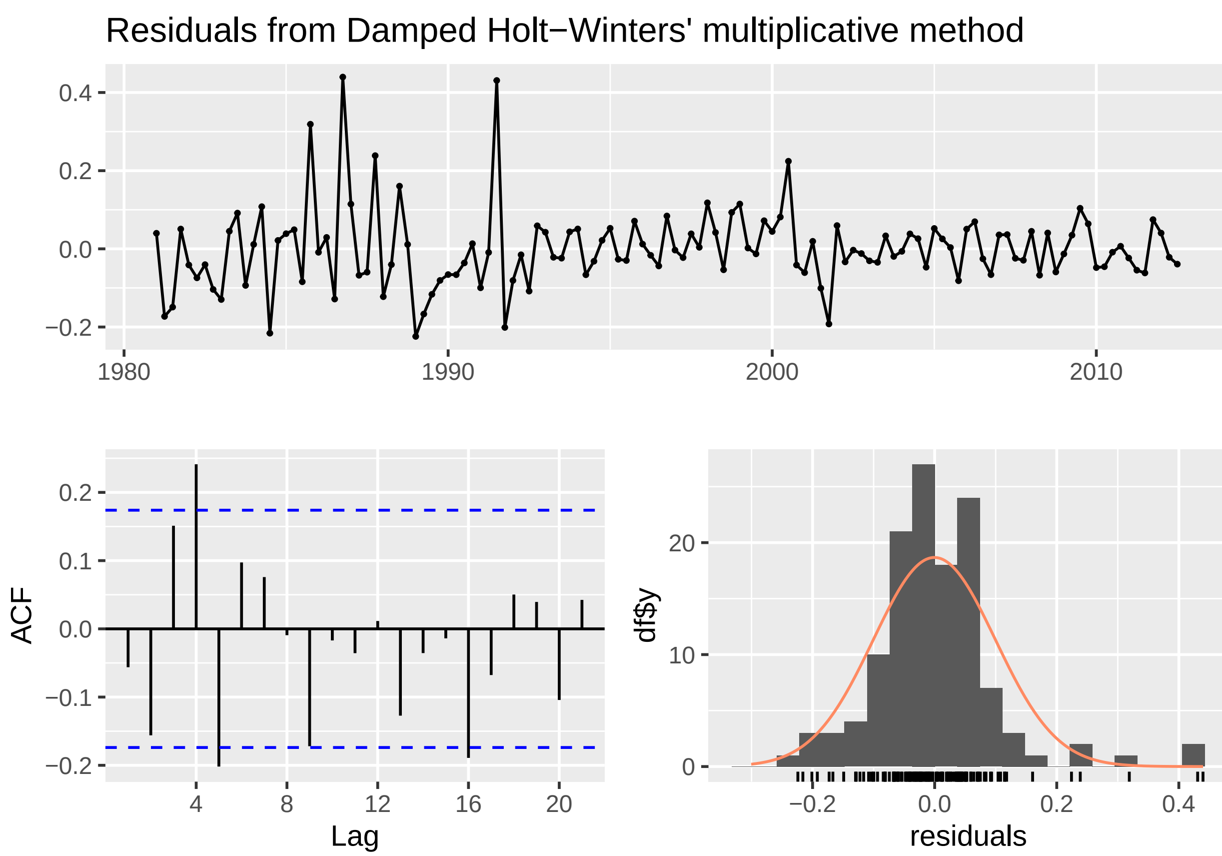
Method	RMSE
add_hw	7542.656
multi_hw	7550.956

add_damp_hw	7552.064
multi_damp_hw	7460.002
multi_expo_hw	7870.123

I prefer to use the method of Linear trend with multiplicative seasonality and damping. This method showcases the lowest Root Mean Square Error (RMSE) value of 7460.002, signifying its superior accuracy and reliability in forecasting with minimal errors.

Question 5

```
checkresiduals(multi_damp_hw)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from Damped Holt-Winters' multiplicative method
## Q* = 21.921, df = 8, p-value = 0.005064
##
## Model df: 0.   Total lags used: 8
```

- Residuals over time: The residuals do not show any apparent trends or patterns over time. However, the variance of residuals becomes smaller after 1950.

- ACF of residuals: The autocorrelation of the residuals is low (mostly within the blue confidence bounds), indicating the model has captured the data's time-related patterns well, leaving behind noise.
- Histogram of residuals: The residuals are roughly normally distributed as shown by the histogram and the smooth curve, meaning the model's errors are appropriate for the data.

However, the p-value is below the common significance level of 0.05, we reject the null hypothesis of no autocorrelation in the residuals at the 95% confidence level. This suggests that there is evidence of some autocorrelation in the residuals from the Damped Holt-Winters' multiplicative method, which contradicts the earlier assessment from the ACF plot.

```r
summary(multi_damp_hw)
```

```
##
## Forecast method: Damped Holt-Winters' multiplicative method
##
## Model Information:
## Damped Holt-Winters' multiplicative method
##
## Call:
##  hw(y = visitors_Quarterly, h = 20, seasonal = "multiplicative",
##
##  Call:
##      damped = TRUE)
##
##   Smoothing parameters:
##     alpha = 0.52
##     beta  = 0.0027
##     gamma = 1e-04
##     phi   = 0.98
##
##   Initial states:
##     l = 26914.2591
##     b = 2002.8599
##     s = 1.0638 0.9467 0.9133 1.0762
##
##   sigma:  0.1034
##
##      AIC     AICc      BIC
## 2913.642 2915.538 2942.084
##
## Error measures:
##                     ME     RMSE     MAE        MPE     MAPE      MASE
## Training set -1.834419 7460.002 5363.22 -0.9701216 6.852661 0.7266756
##                      ACF1
## Training set -0.001971966
##
## Forecasts:
##         Point Forecast     Lo 80    Hi 80    Lo 95    Hi 95
## 2012 Q4       122225.2 106033.32 138417.1 97461.86 146988.5
## 2013 Q1       123798.8 105276.90 142320.8 95471.97 152125.7
## 2013 Q2       105187.4  87816.96 122557.8 78621.62 131753.1
## 2013 Q3       109171.5  89578.75 128764.3 79206.97 139136.1
## 2013 Q4       122829.1  99138.55 146519.6 86597.54 159060.7
```

```
## 2014 Q1       124397.5  98830.37 149964.7 85295.93 163499.2
## 2014 Q2       105685.3  82691.88 128678.6 70519.92 140850.6
## 2014 Q3       109677.3  84552.31 134802.3 71251.94 148102.7
## 2014 Q4       123386.1  93754.23 153018.0 78068.05 168704.2
## 2015 Q1       124949.8  93607.44 156292.1 77015.81 172883.7
## 2015 Q2       106144.5  78421.35 133867.7 63745.60 148543.4
## 2015 Q3       110143.9  80270.02 140017.7 64455.77 155832.0
## 2015 Q4       123899.9  89083.75 158716.1 70653.18 177146.7
## 2016 Q1       125459.1  89009.06 161909.2 69713.57 181204.7
## 2016 Q2       106568.1  74614.21 138522.0 57698.83 155437.4
## 2016 Q3       110574.2  76411.57 144736.8 58326.96 162821.4
## 2016 Q4       124373.9  84836.81 163910.9 63907.17 184840.6
## 2017 Q1       125929.0  84794.32 167063.6 63018.96 188838.9
## 2017 Q2       106958.8  71100.32 142817.3 52117.97 161799.7
## 2017 Q3       110971.1  72828.41 149113.8 52636.88 169305.4
```

- alpha: 0.52 suggests moderate weighting to the recent past data
- beta: 0.0027 indicates a very small adjustment for the trend, implying the trend is fairly stable.
- gamma: 1e-04 is also quite small, suggesting the seasonal component changes very slowly.
- phi: 0.98 near 1 indicates a very slow damping of the trend.

Overall, the model seems to be a reasonable fit for the data. The forecast accuracy would be considered acceptable for many practical applications, especially considering the relatively low MAPE.

Question 6

```r
# Create a seasonal naive model
snaive_forecast <- snaive(visitors_Quarterly, h = 20)
accuracy(snaive_forecast)
```

```
##                    ME     RMSE      MAE      MPE     MAPE MASE      ACF1
## Training set 2885.382 10298.98 7380.488 3.728117 9.396353    1 0.4528293
```

Observing that the RMSE of the seasonal naive method exceeds that of our preferred model indicates that our selected model surpasses the seasonal naive method in forecasting accuracy.

Question 7

(a)

- Trend Component: The trend graph shows a general upward trend in the number of persons in the civilian labor force in Australia. This is evident from the consistent increase in the value over time.
- Seasonal Component: The seasonal graph reveals a consistent pattern that repeats annually, which is typical for labor force data influenced by seasonal hiring patterns, holidays, and similar periodic events. However, the seasonality component has values that fluctuate between approximately -50 and +100, indicating that the seasonal effect is significant but not dominant compared to the overall values of the time series, which are in the thousands.
- Remainder Component: The scale shows fluctuations that range up to about ±300, suggesting there are still some unexplained variations in the data after accounting for the trend and seasonal components. Some large spikes, especially downward ones from 1990 to 1995, may indicate outliers or periods of unusual activity that were not captured by the trend or seasonal components.

(b)

Based on the trend component, there's no obvious sharp decline around 1991/1992 that would be indicative of a recession. The trend continues its upward progression with only minor fluctuations. The seasonal and remainder components don't directly show economic conditions like a recession because they're designed to capture predictable patterns and random noise, respectively.