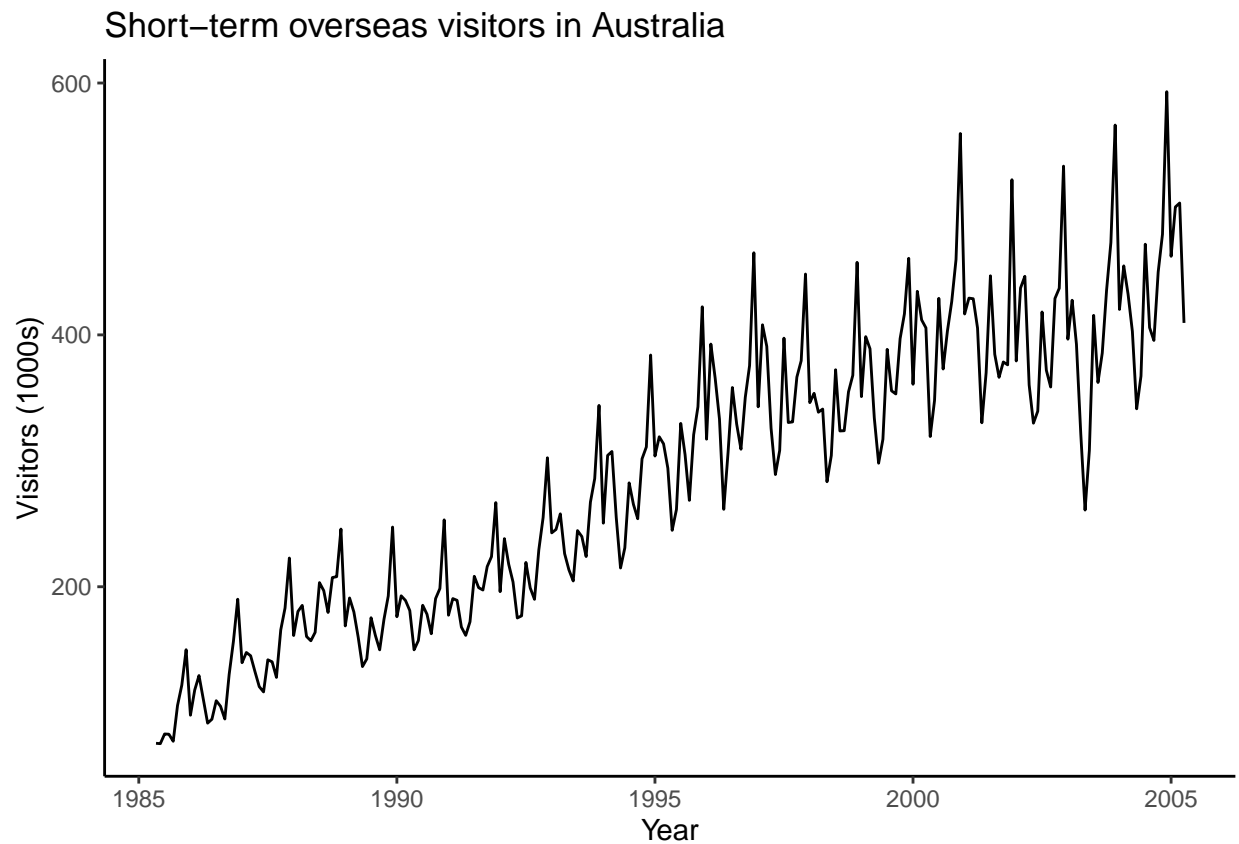


BUAN6357_Homework2_Noorani

Part A: This time plot shows both a trend & seasonality. We see a general positive trend as the years progress. We also see seasonality, as we see regular/periodic troughs and peaks.

```
#Part A - make time plot & describe main features  
autoplot(visitors) +  
  ggtitle("Short-term overseas visitors in Australia") +  
  xlab("Year") + ylab("Visitors (1000s)")
```



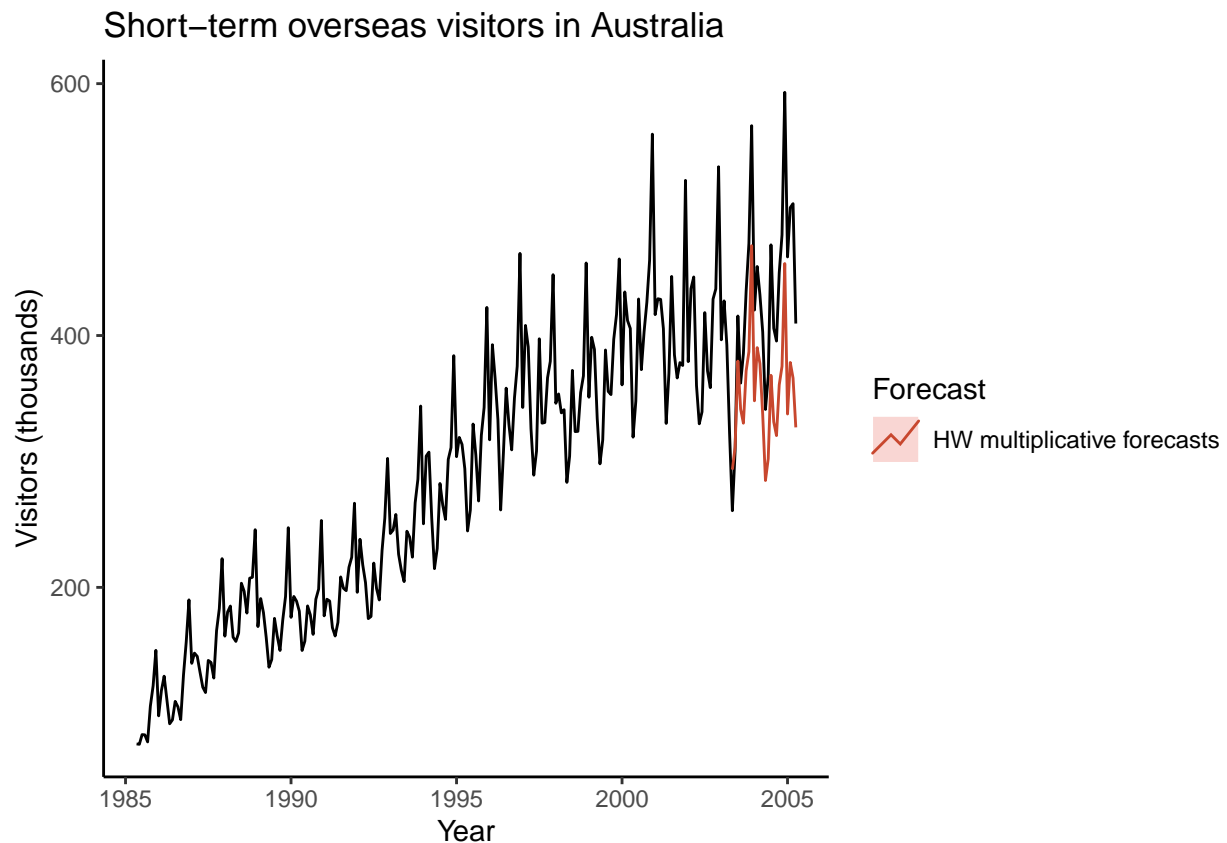
Part B - Split data into train & test set comprising the last two years of available data. Forecast the test set using Holt-Winters' multiplicative method.

```
#split so that test set includes last 24 data points (last 2 years)  
train <- window(visitors, end = c(2003,4))  
test <- window(visitors, start = c(2003,5))  
  
#Forecasting for next 24 periods, based on train data set  
hwfc <- hw(train, seasonal="multiplicative", h = 24)  
  
### Check the models  
hwfc[["model"]]
```

```
## Holt-Winters' multiplicative method
```

```
##
## Call:
## hw(y = train, h = 24, seasonal = "multiplicative")
##
## Smoothing parameters:
##   alpha = 0.4379
##   beta  = 0.0164
##   gamma = 1e-04
##
## Initial states:
##   l = 90.9387
##   b = 2.584
##   s = 0.945 1.0553 1.0882 0.9681 1.3072 1.0711
##       1.0264 0.9095 0.938 1.0401 0.8509 0.8001
##
## sigma: 0.0571
##
##      AIC      AICc      BIC
## 2326.608 2329.699 2383.988
```

```
### Plotting forecast as layer on original data, so we can visually see how close our forecast is to ac
autoplot(visitors) +
  autolayer(hwfc, series="HW multiplicative forecasts", PI=FALSE) +
  xlab("Year") + ylab("Visitors (thousands)") +
  ggtitle("Short-term overseas visitors in Australia") +
  guides(colour=guide_legend(title="Forecast"))
```



Part C: is multiplicative seasonality necessary here? Why or why not?

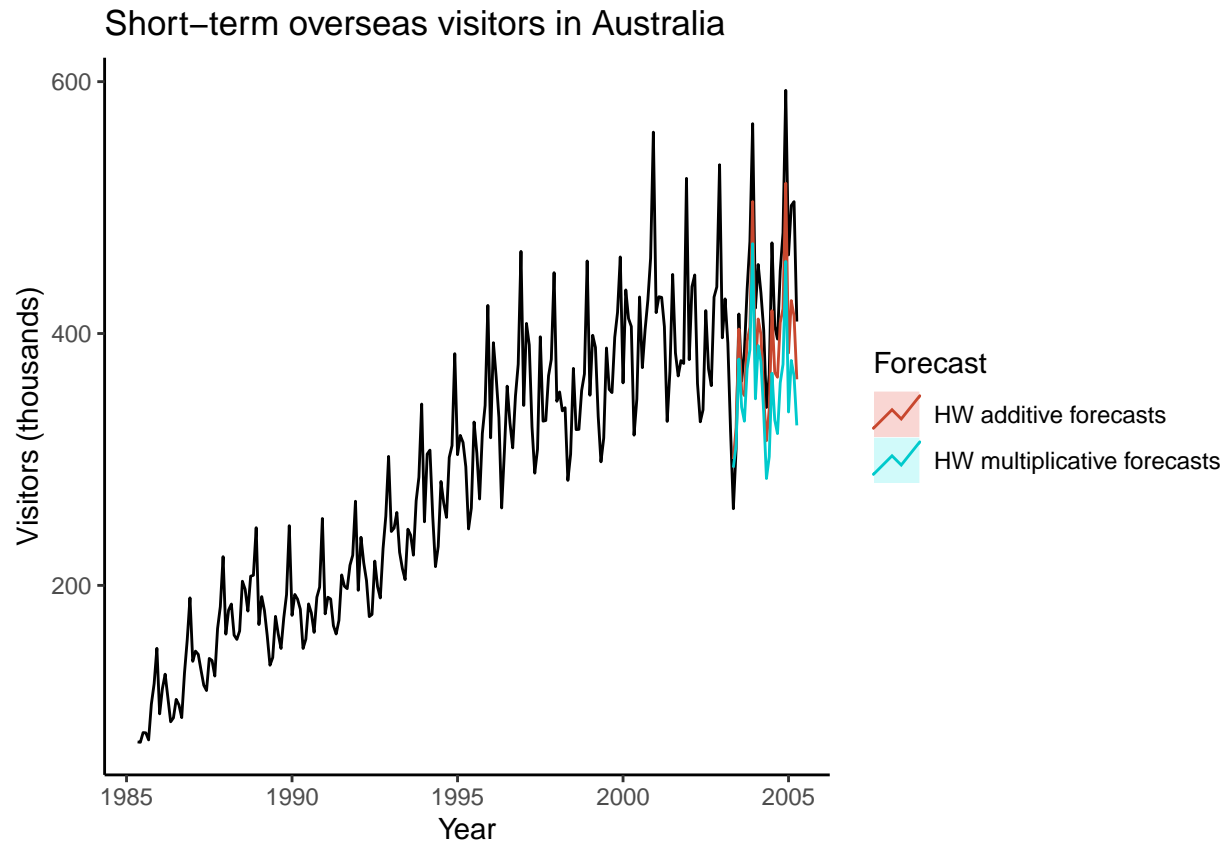
Multiplicative seasonality does seem to be necessary because the magnitude of the seasonal variation does seem to be proportional to the year. In other words, the magnitude of the troughs & peaks varies by year. In addition, if we use AIC as our comparison criteria, then the multiplicative seasonality model performs slightly better than the additive. AIC for multiplicative model: 2327, AIC for additive model: 2423

```
#comparing part B w/ HW's additive model
hwfc2 <- hw(train, seasonal="additive", h = 24)

### Check the models
hwfc2[["model"]]
```

```
## Holt-Winters' additive method
##
## Call:
## hw(y = train, h = 24, seasonal = "additive")
##
## Smoothing parameters:
##   alpha = 0.4095
##   beta  = 1e-04
##   gamma = 0.3117
##
## Initial states:
##   l = 102.8697
##   b = 1.2155
##   s = -13.6952 13.9731 23.5123 -9.7538 81.8144 18.6175
##         5.9491 -30.6165 -15.7017 15.2181 -38.6217 -50.6956
##
## sigma: 17.8275
##
##      AIC      AICc      BIC
## 2422.918 2426.008 2480.297
```

```
### Plotting forecast as layer on original data, so we can visually see how close our forecast is to actual
autoplot(visitors) +
  autolayer(hwfc2, series="HW additive forecasts", PI=FALSE) +
  autolayer(hwfc, series="HW multiplicative forecasts", PI=FALSE) +
  xlab("Year") + ylab("Visitors (thousands)") +
  ggtitle("Short-term overseas visitors in Australia") +
  guides(colour=guide_legend(title="Forecast"))
```



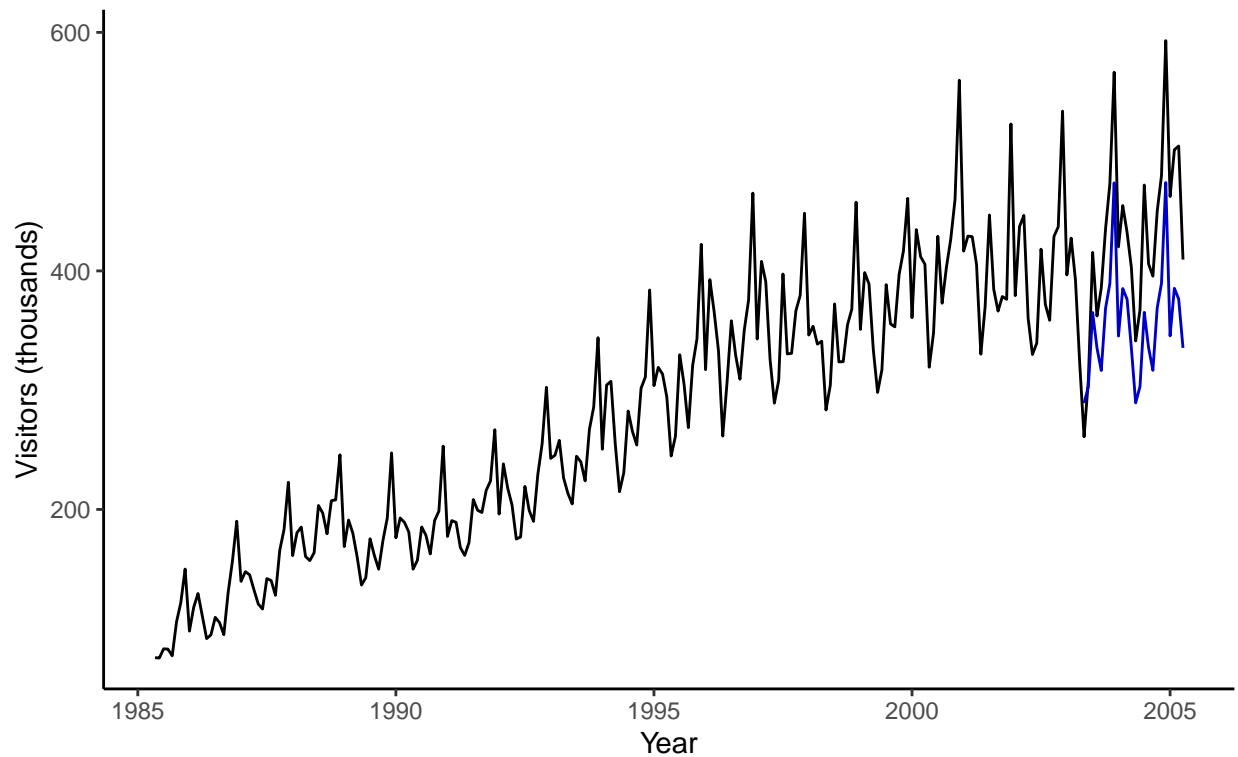
Part D - Forecast the two-year test set using each of the following methods: i. ETS ii. additive ETS model applied to a Box-Cox transformed series iii. seasonal naïve method

```
#i. an ETS model
ets_fit <- ets(train)
ets_forecast <- forecast(ets_fit, h=24)

autoplot(visitors) +
  autolayer(ets_forecast, PI=FALSE) +
  ggtitle("ETS Model\nShort-term overseas visitors in Australia") +
  xlab("Year") + ylab("Visitors (thousands)")
```

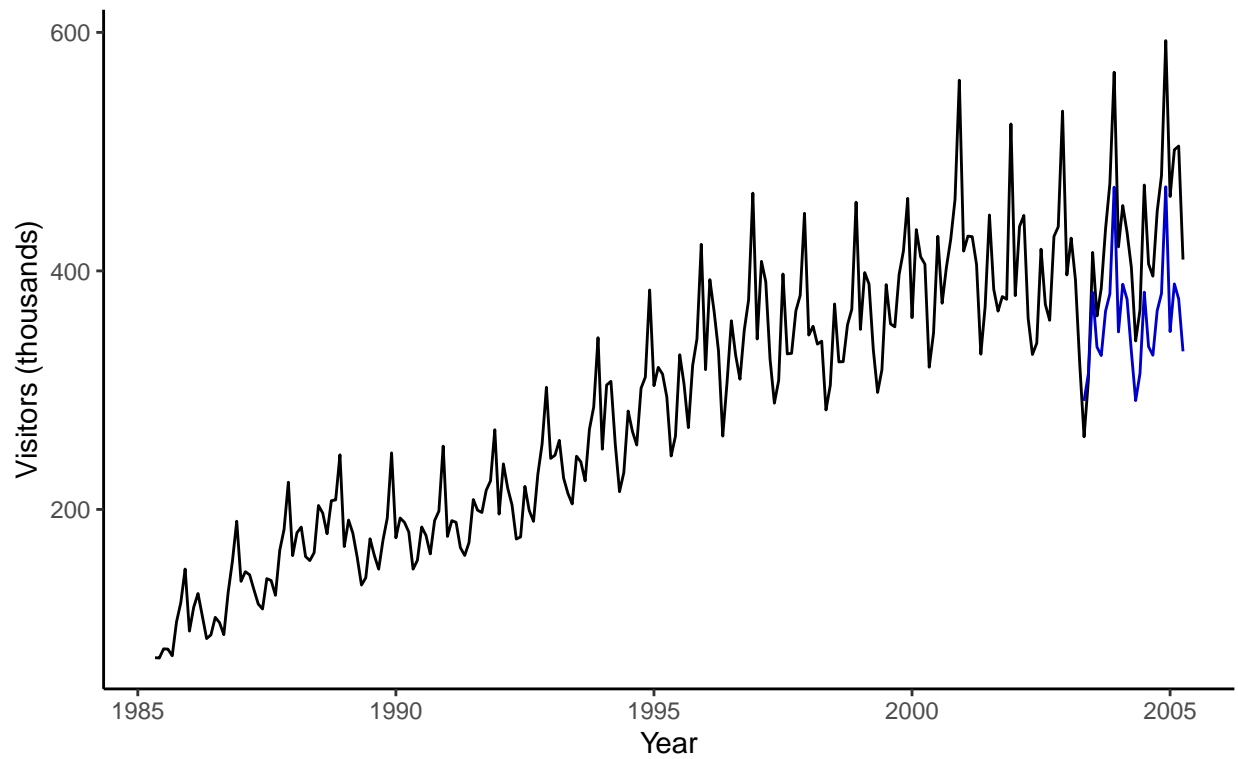
ETS Model

Short-term overseas visitors in Australia



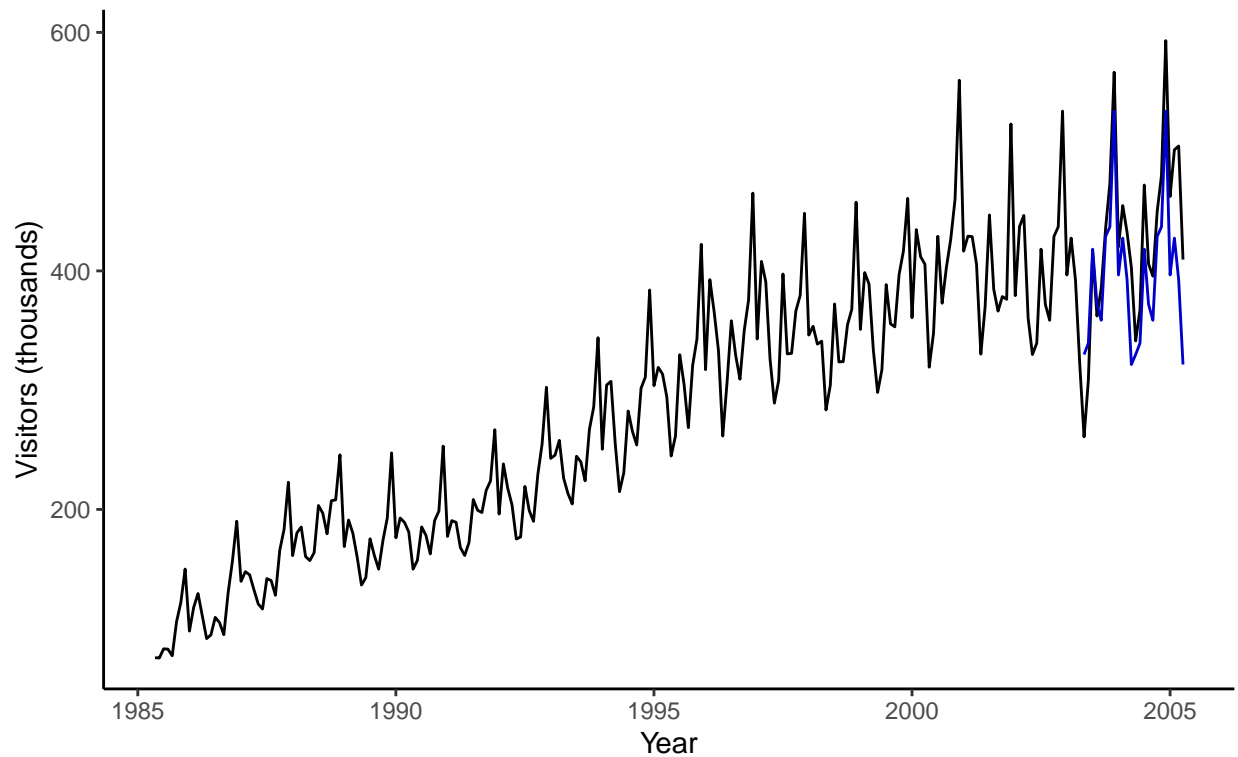
```
# -----#  
  
#ii. an additive ETS model applied to a Box-Cox transformed series  
(lambda <- BoxCox.lambda(train))  
  
## [1] 0.3624899  
  
ETS_bc_add_forecast <- ets(train, lambda=lambda, additive.only = TRUE) %>% forecast(h=24)  
  
autoplot(visitors) +  
  autolayer(ETS_bc_add_forecast, PI = FALSE) +  
  theme_classic() +  
  ggtitle("Additive Box-Cox transformed\nShort-term overseas visitors in Australia (thousands)") +  
  xlab("Year") + ylab("Visitors (thousands)")
```

Additive Box–Cox transformed
Short-term overseas visitors in Australia (thousands)



```
# -----#  
  
    ###. a seasonal naïve method  
snaive_forecast <- snaive(train, h=24)  
  
autoplot(visitors) +  
  autolayer(snaive_forecast, PI = FALSE) +  
  ggtitle("Seasonal Naive\nShort-term overseas visitors in Australia (thousands)") +  
  xlab("Year") + ylab("Visitors (thousands)")
```

Seasonal Naive Short-term overseas visitors in Australia (thousands)

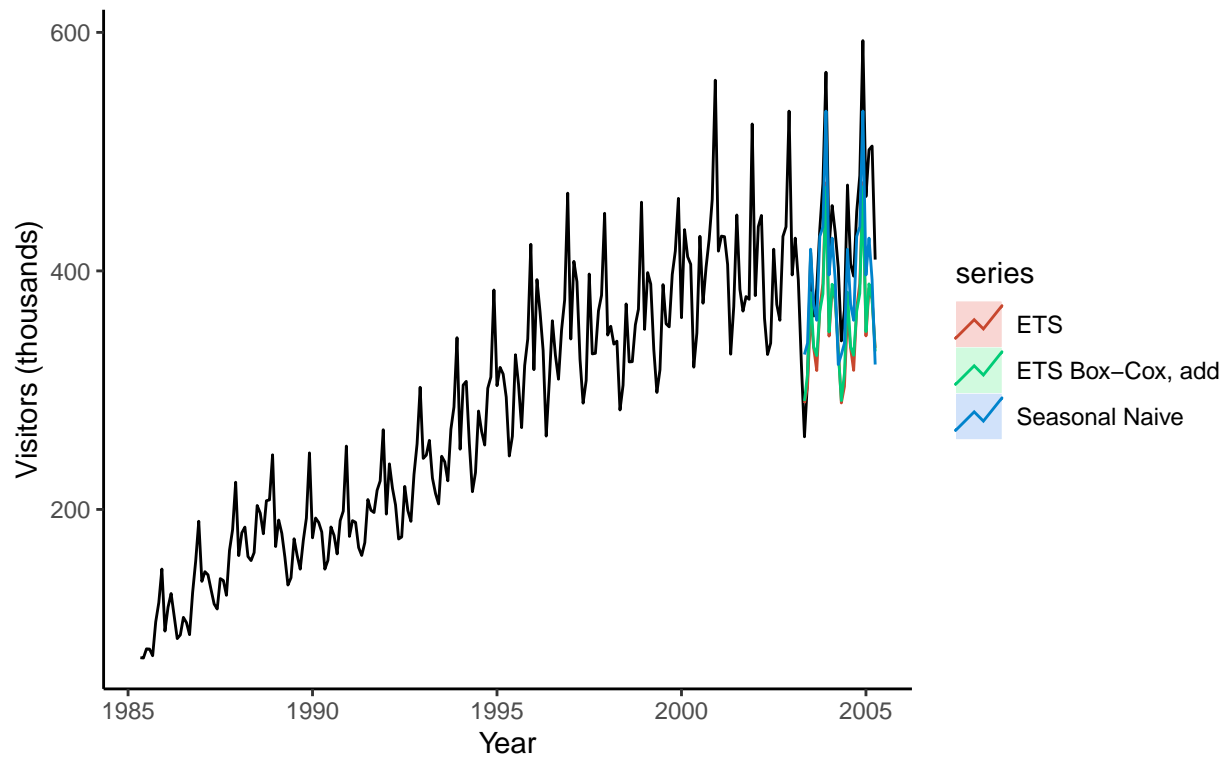


Part E - Which method gives the best forecasts? Does this method pass the residuals test? According to the accuracy function, the ETS(M,Ad,M) model is the best, since it gives the lowest RMSE. This method does not quite pass the residuals test, since the p-value is less than 0.05, which means we reject the null hypothesis.

```
#first, let's do a visual comparison of all 4 methods
autoplot(visitors) +
  autolayer(ets_forecast, PI=FALSE, series="ETS") +
  autolayer(ETS_bc_add_forecast, PI=FALSE, series="ETS Box-Cox, add") +
  autolayer(snaive_forecast, PI=FALSE, series="Seasonal Naive") +
  ggtitle("All Forecasts\nShort-term overseas visitors in Australia (thousands)") +
  xlab("Year") + ylab("Visitors (thousands)")
```

All Forecasts

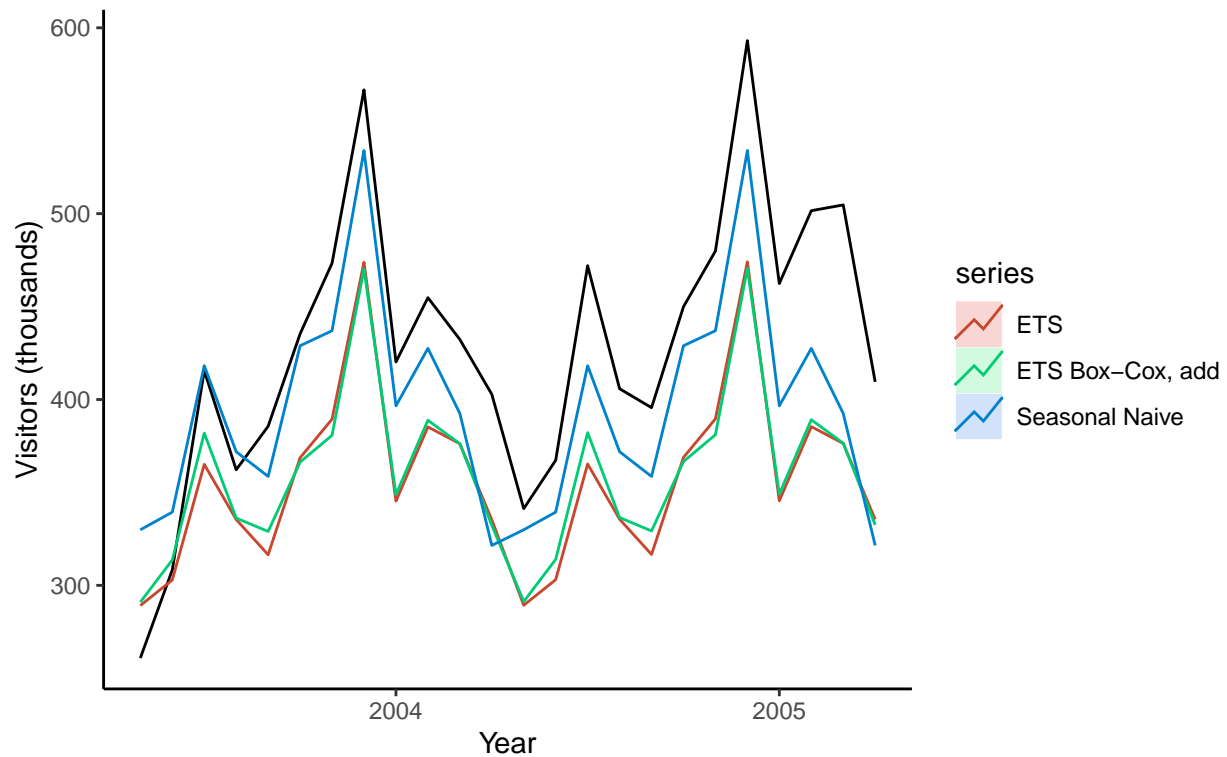
Short-term overseas visitors in Australia (thousands)



```
#now let's do a visual comparison, but focusing only on the test data, so that we zoom in  
#on the last 24 points  
autoplot(test) +  
  autolayer(ets_forecast, PI=FALSE, series="ETS") +  
  autolayer(ETS_bc_add_forecast, PI=FALSE, series="ETS Box-Cox, add") +  
  autolayer(snaive_forecast, PI=FALSE, series="Seasonal Naive") +  
  ggtitle("All Forecasts\nShort-term overseas visitors in Australia (thousands)") +  
  xlab("Year") + ylab("Visitors (thousands)")
```


All Forecasts

Short-term overseas visitors in Australia (thousands)

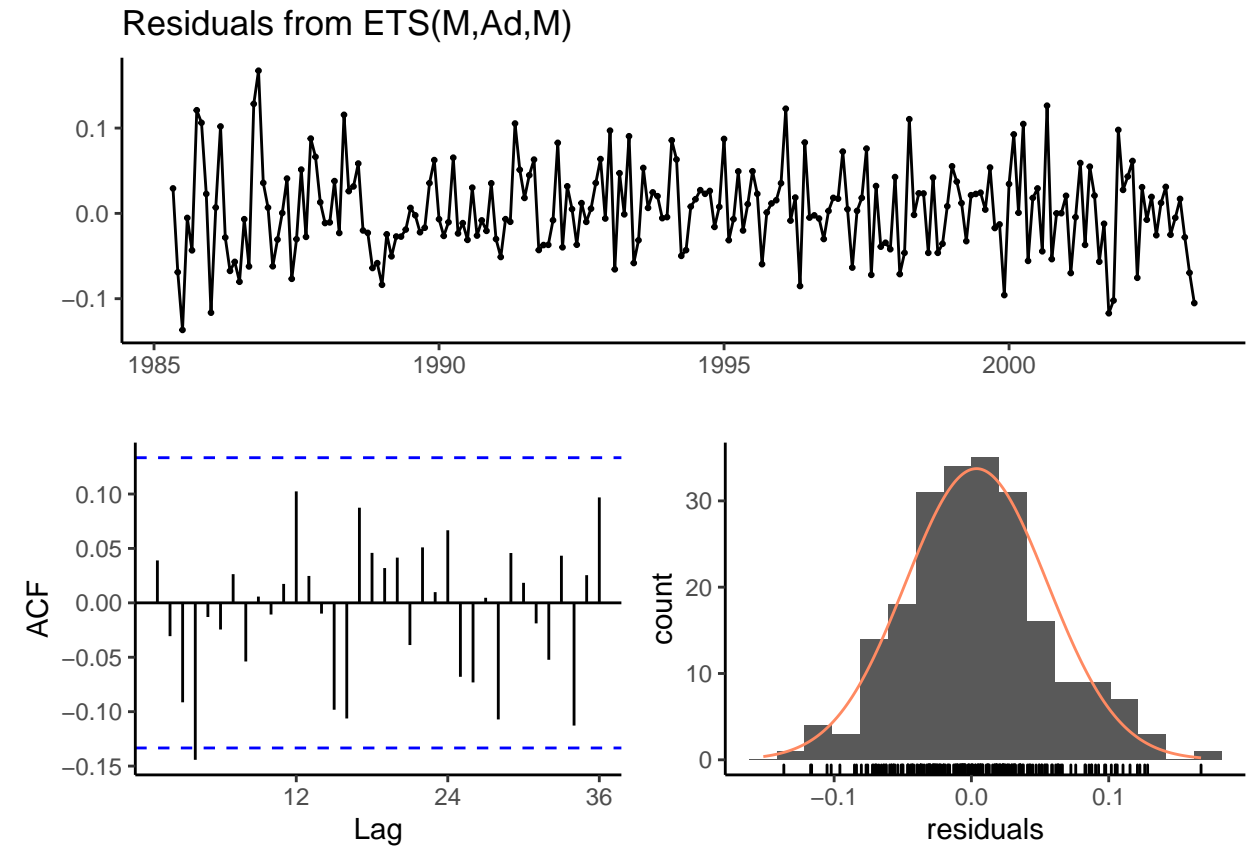


```
#comparing forecasts using accuracy fcn
ETS_acc <- accuracy(ets_forecast, visitors)
ETS_bc_add_fc_acc <- accuracy(ETS_bc_add_forecast, visitors)
snaive_acc <- accuracy(snaive_forecast, visitors)

#print results from accuracy fcn. Let's use just the RMSE to compare
RMSE <- c(ETS_acc['Training set', 'RMSE'], ETS_bc_add_fc_acc['Training set', 'RMSE'], snaive_acc['Training
names(RMSE) <- c('ETS', 'additive_ETS_Box_Cox', 'Seasonal_Naive')
print(RMSE)
```

```
##          ETS additive_ETS_Box_Cox          Seasonal_Naive
##          14.53480          14.97096          31.15613
```

```
#using checkresiduals on ETS, best model
checkresiduals(ets_forecast)
```



```
##
##  Ljung-Box test
##
## data:  Residuals from ETS(M,Ad,M)
## Q* = 20.677, df = 7, p-value = 0.004278
##
## Model df: 17.   Total lags used: 24
```