Assignment1

Athena Rodrigues

2025-03-19

Set Up

The tidyverse, forecast, readr, and ggplot2 packages were used in this assignment.

Question 1

Loading the data set into R.

```r
rd <- read_csv("AFG_IRL_refugeeData.csv", show_col_types = FALSE)
head(rd)
```

```
## # A tibble: 6 x 6
##    Year 'Country of origin' 'Country of asylum' 'Country of origin (ISO)'
##   <dbl> <chr>               <chr>               <chr>
## 1  1999 Afghanistan         Ireland             AFG
## 2  2000 Afghanistan         Ireland             AFG
## 3  2001 Afghanistan         Ireland             AFG
## 4  2002 Afghanistan         Ireland             AFG
## 5  2003 Afghanistan         Ireland             AFG
## 6  2004 Afghanistan         Ireland             AFG
## # i 2 more variables: 'Country of asylum (ISO)' <chr>,
## #   'Refugees under UNHCR's mandate' <dbl>
```

```r
summary(rd)
```

```
##       Year      Country of origin  Country of asylum  Country of origin (ISO)
##   Min.   :1999   Length:23          Length:23          Length:23
##   1st Qu.:2004   Class :character   Class :character   Class :character
##   Median :2010   Mode  :character   Mode  :character   Mode  :character
##   Mean   :2010
##   3rd Qu.:2016
##   Max.   :2021
##   Country of asylum (ISO) Refugees under UNHCR's mandate
##   Length:23               Min.   :  5.0
##   Class :character        1st Qu.: 67.5
##   Mode  :character        Median :149.0
##                           Mean   :161.1
##                           3rd Qu.:202.0
##                           Max.   :431.0
```

Question 2

Creating Time Series Object

```
rdts <- ts(rd$`Refugees under UNHCR's mandate`, start = c(1999), end = c(2021), frequency = 1)
print(rdts)
```

```
## Time Series:
## Start = 1999
## End = 2021
## Frequency = 1
##  [1]    5   30   41   41   38   58   77   99  106  127  149  157  170  181  129  158  182  222  230
## [20]  289  366  420  431
```

Question 3

Time Series Plot

```
autoplot(rdts) +
  labs(title = "Afghan Refugees into Ireland from 1999 - 2021") +
  xlab("Year") +
  ylab("Number of Refugees")
```
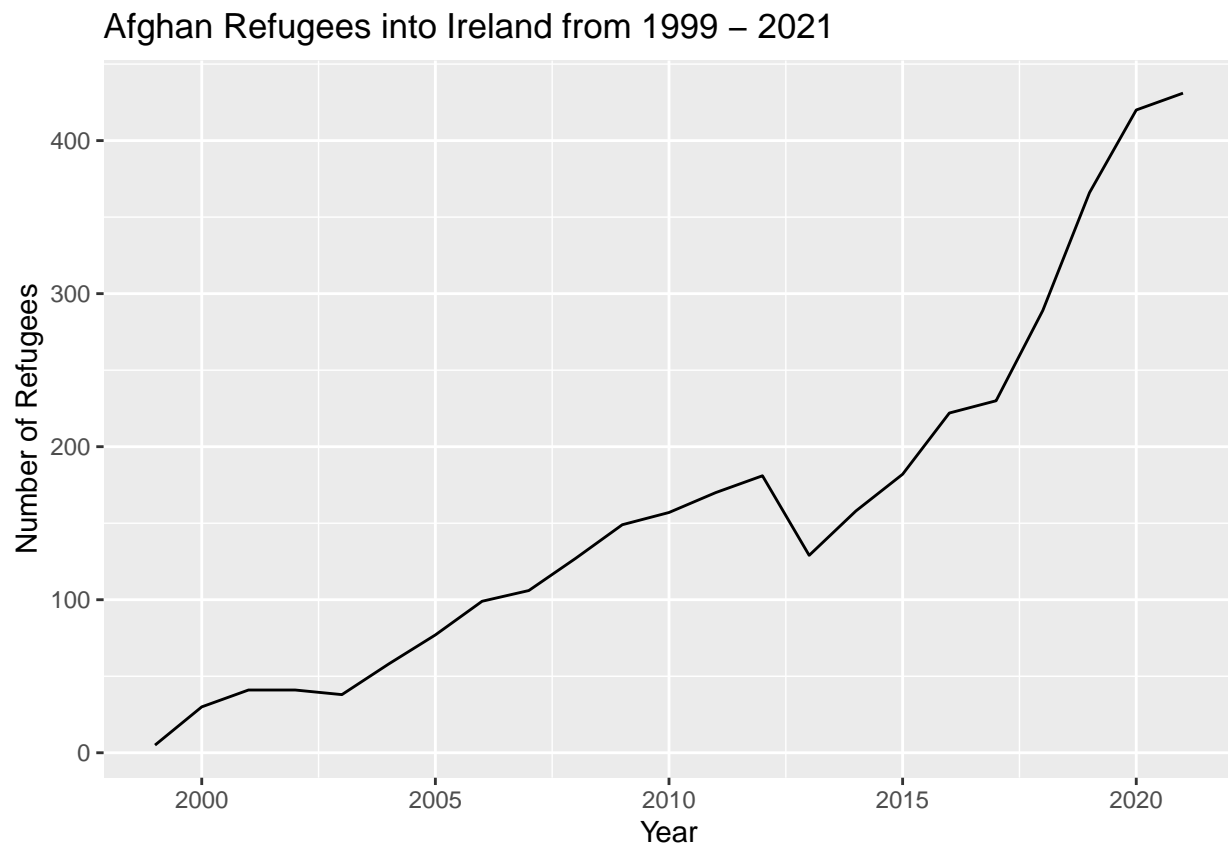
Question 4

Forecasting Report for 2016 to 2021 using 1999 to 2015

```
train <- window(rdts, start = 1999, end = 2015)
```

Average Method

```
average <- meanf(train, h = 6)
summary(average)
```

```
##
## Forecast method: Mean
##
## Model Information:
## $mu
## [1] 102.8235
##
## $mu.se
## [1] 14.30941
##
## $sd
## [1] 58.99919
##
## $bootstrap
## [1] FALSE
##
## $call
## meanf(y = train, h = 6)
##
## attr(,"class")
## [1] "meanf"
##
## Error measures:
##                       ME     RMSE      MAE       MPE    MAPE     MASE
## Training set -4.177214e-15 57.23762 51.01038 -148.0652 179.707 2.843784
##                    ACF1
## Training set 0.7909365
##
## Forecasts:
##      Point Forecast    Lo 80    Hi 80     Lo 95    Hi 95
## 2016       102.8235 21.66945 183.9776 -25.87521 231.5223
## 2017       102.8235 21.66945 183.9776 -25.87521 231.5223
## 2018       102.8235 21.66945 183.9776 -25.87521 231.5223
## 2019       102.8235 21.66945 183.9776 -25.87521 231.5223
## 2020       102.8235 21.66945 183.9776 -25.87521 231.5223
## 2021       102.8235 21.66945 183.9776 -25.87521 231.5223
```

```
autoplot(train) +
  autolayer(average, series="Average", PI=TRUE) +
  labs(title = "Average Method Forecasting") +
  xlab("Year") +
  ylab("Number of Refugees")
```

Average Method Forecasting
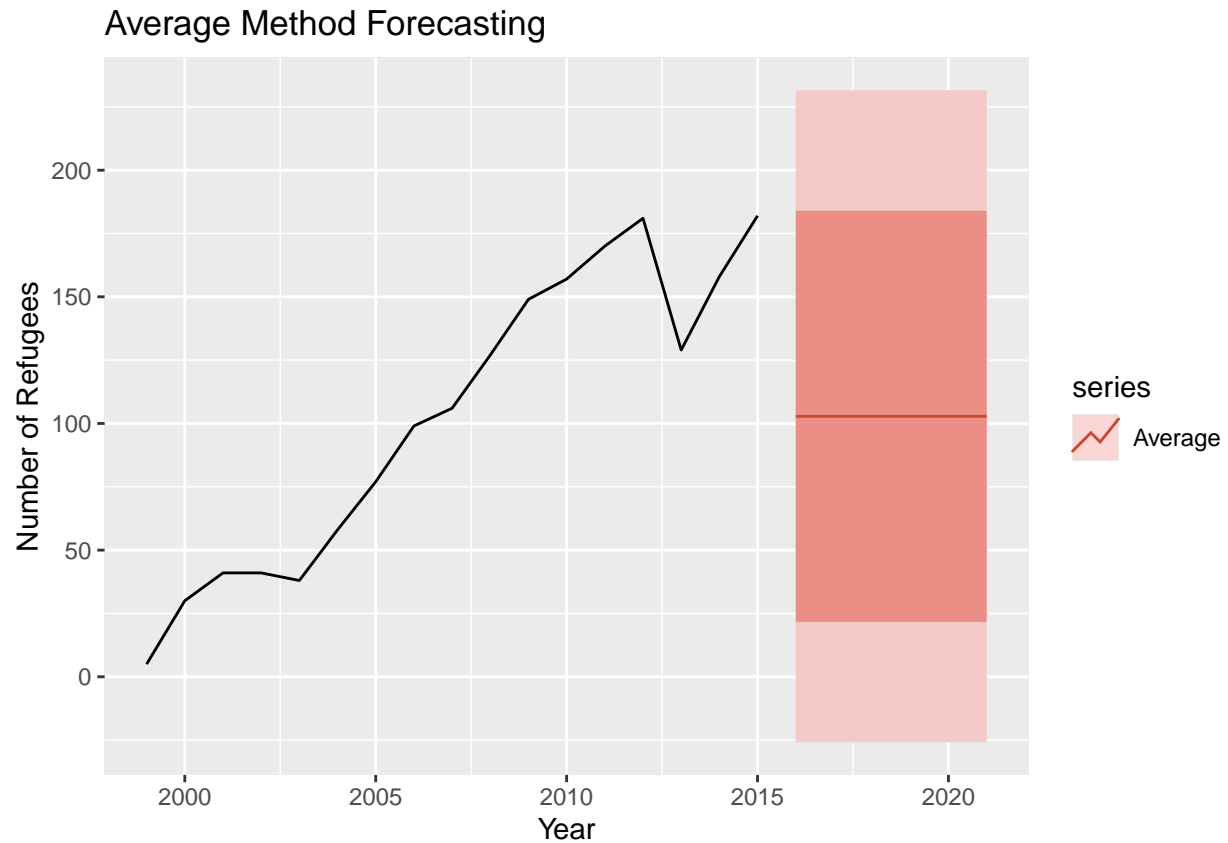
Naive Method

```
naive <- naive(train, h=6)
summary(naive)
```

```
##
## Forecast method: Naive method
##
## Model Information:
## Call: naive(y = train, h = 6)
##
## Residual sd: 21.5769
##
## Error measures:
##                   ME    RMSE     MAE      MPE     MAPE MASE       ACF1
## Training set 11.0625 21.5769 17.9375 14.47522 20.50083    1 -0.1317405
##
## Forecasts:
##      Point Forecast    Lo 80    Hi 80     Lo 95    Hi 95
## 2016            182 154.3481 209.6519 139.71006 224.2899
## 2017            182 142.8943 221.1057 122.19299 241.8070
## 2018            182 134.1055 229.8945 108.75167 255.2483
## 2019            182 126.6962 237.3038  97.42012 266.5799
## 2020            182 120.1685 243.8315  87.43682 276.5632
## 2021            182 114.2669 249.7331  78.41122 285.5888
```

```
autoplot(train) +
  autolayer(naive, series="Naive", PI=TRUE)+
  labs(title = "Naive Method Forecasting") +
  xlab("Year") +
  ylab("Number of Refugees")
```
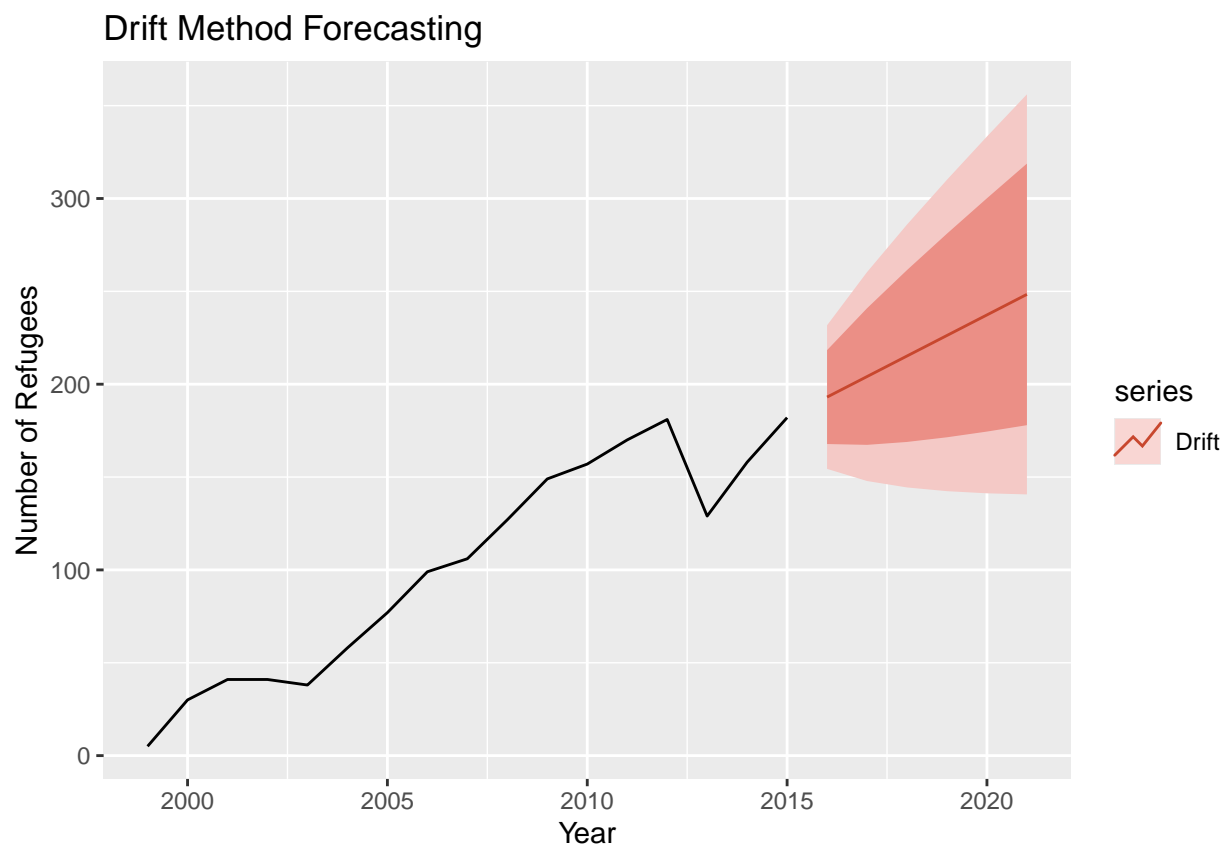
Naive Method Forecasting
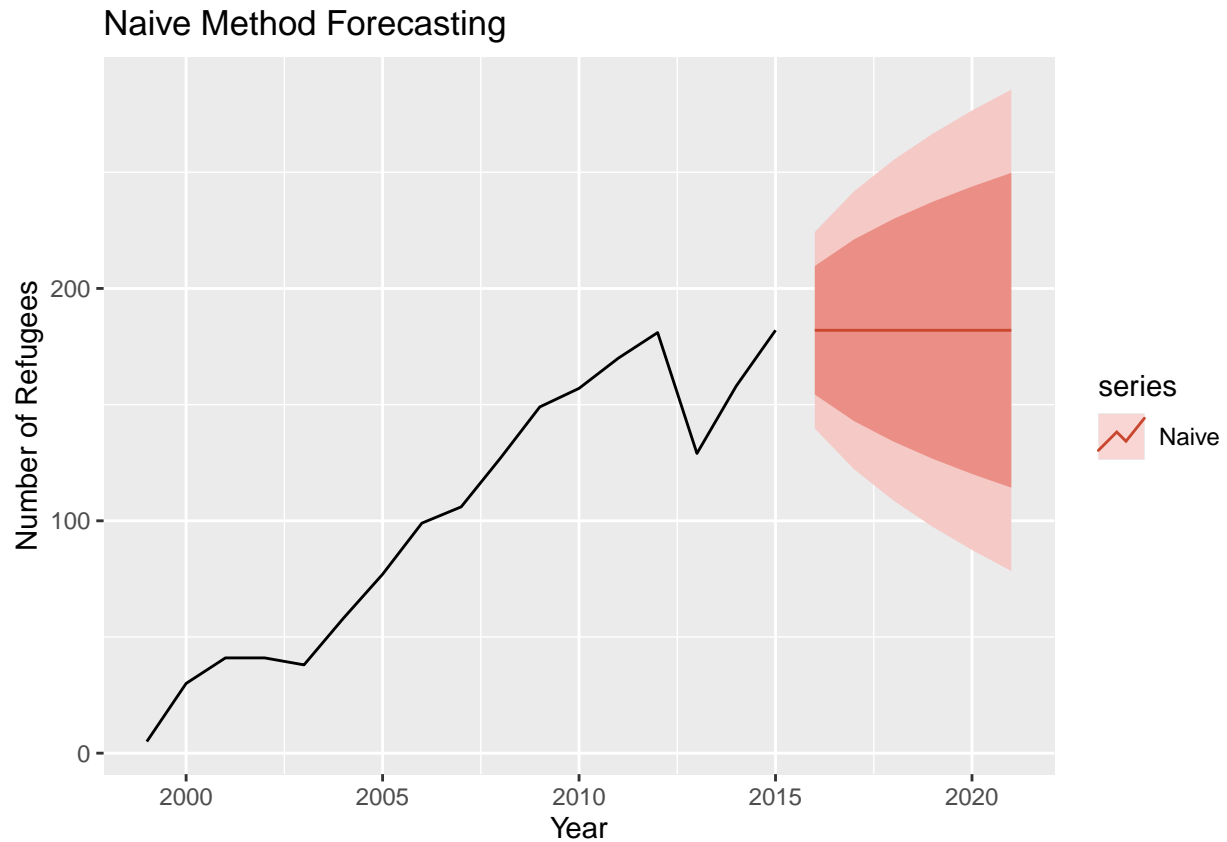
Drift Method

```
drift <- rwf(train, h=6, drift=TRUE)
summary(drift)
```

```
##
## Forecast method: Random walk with drift
##
## Model Information:
## Call: rwf(y = train, h = 6, drift = TRUE)
##
## Drift: 11.0625  (se 4.7832)
## Residual sd: 19.1328
##
## Error measures:
##                ME     RMSE      MAE        MPE     MAPE      MASE       ACF1
## Training set    0 18.52522 11.92969 -0.05333648 14.80218 0.6650697 -0.1317405
##
## Forecasts:
```

```
##      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
## 2016         193.0625  167.7883  218.3367  154.4089  231.7161
## 2017         204.1250  167.3456  240.9044  147.8757  260.3743
## 2018         215.1875  168.9077  261.4673  144.4087  285.9663
## 2019         226.2500  171.4225  281.0775  142.3985  310.1015
## 2020         237.3125  174.4997  300.1253  141.2486  333.3764
## 2021         248.3750  177.9478  318.8022  140.6658  356.0842
```

```
autoplot(train) +
  autolayer(drift, series="Drift", PI=TRUE) +
  labs(title = "Drift Method Forecasting") +
  xlab("Year") +
  ylab("Number of Refugees")
```


Drift Method Forecasting

Overall Forecast of Methods

```
forecast_table <- data.frame(
  Year = 2016:2021,
  Average = average$mean,
  Naive = naive$mean,
  Drift = drift$mean
)
forecast_table
```

```
##   Year  Average Naive    Drift
```

```
## 1 2016 102.8235    182 193.0625
## 2 2017 102.8235    182 204.1250
## 3 2018 102.8235    182 215.1875
## 4 2019 102.8235    182 226.2500
## 5 2020 102.8235    182 237.3125
## 6 2021 102.8235    182 248.3750
```

Question 5

```
autoplot(rdts) +
  autolayer(average, series="Average", PI=FALSE) +
  autolayer(naive, series="Naive", PI=FALSE) +
  autolayer(drift, series="Drift", PI=FALSE) +
  labs(title = "Forecasts for Afghan Refugees into Ireland (2016-2021)",
       x = "Year", y = "Number of Refugees") +
  guides(colour=guide_legend(title="Method"))
```
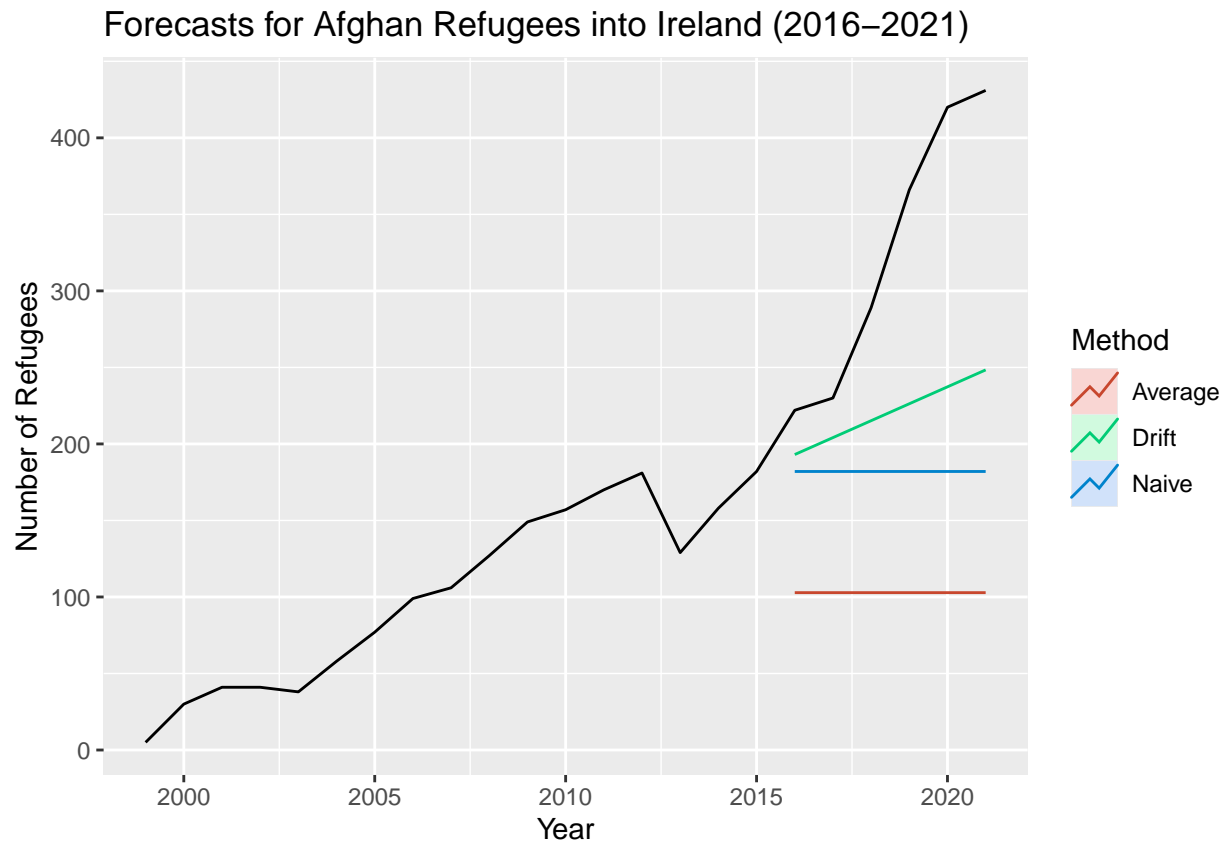

Question 6

The Drift Method is the best method out of the three for forecasting the number of Afghan refugees entering Ireland. The other two methods show no change in the number of refugees over time, which is known to be false when compared to the actual data. The Average and Naive Methods' focus on past observations fails to correctly forecast the increase in refugee numbers, however, the Drift Method's approximation attempts

using average and last observations incorporate this increase over time. To improve the drift method, rolling or time series cross-validation could be added to assess the model's predictive accuracy over time.