My answers

Ricky Yang(GuangLiang Yang) 23205919

2024-03-14

```
# Write your answer here
```

Working with Data

(1) Import the dataset `Soccer_Table_2014-2015.csv` into `R` . (Thanks to Kirsten Spencer who supplied the dataset.)

```
# Write your answer here
df <- read.csv("~/Downloads/Soccer_Table_2014-2015.csv")
head(df)
```

```
##           League Pos              Team  P  W  D  L  F  A GD Pts   Ta
ble.Type
## 1 Premier League   1           Chelsea 38 26  9  3 73 32 41  87 Leag
ue Table
## 2 Premier League   2   Manchester City 38 24  7  7 83 38 45  79 Leag
ue Table
## 3 Premier League   3           Arsenal 38 22  9  7 71 36 35  75 Leag
ue Table
## 4 Premier League   4 Manchester United 38 20 10  8 62 37 25  70 Leag
ue Table
## 5 Premier League   5 Tottenham Hotspur 38 19  7 12 58 53  5  64 Leag
ue Table
## 6 Premier League   6         Liverpool 38 18  8 12 52 48  4  62 Leag
ue Table
##    Season                                      KEY  X
## 1 2014/15           Premier League|Chelsea|2014/15 NA
## 2 2014/15   Premier League|Manchester City|2014/15 NA
## 3 2014/15           Premier League|Arsenal|2014/15 NA
## 4 2014/15 Premier League|Manchester United|2014/15 NA
## 5 2014/15 Premier League|Tottenham Hotspur|2014/15 NA
## 6 2014/15         Premier League|Liverpool|2014/15 NA
```

```
head(df,2)
```

```
##           League Pos            Team  P  W D L  F  A GD Pts   Table.
Type
## 1 Premier League   1         Chelsea 38 26 9 3 73 32 41  87 League T
able
## 2 Premier League   2 Manchester City 38 24 7 7 83 38 45  79 League T
able
##    Season                                    KEY  X
## 1 2014/15         Premier League|Chelsea|2014/15 NA
## 2 2014/15 Premier League|Manchester City|2014/15 NA
```

```
data_shape <- dim(df)
data_shape[1]
```

```
## [1] 384
```

```
names(df)
```

```
##  [1] "League"     "Pos"        "Team"       "P"          "W"

##  [6] "D"          "L"          "F"          "A"          "GD"

## [11] "Pts"        "Table.Type" "Season"     "KEY"        "X"
```

```
df[3:5,]
```

```
##           League Pos              Team  P  W  D  L  F  A GD Pts   Ta
ble.Type
## 3 Premier League   3           Arsenal 38 22  9  7 71 36 35  75 Leag
ue Table
## 4 Premier League   4 Manchester United 38 20 10  8 62 37 25  70 Leag
ue Table
## 5 Premier League   5 Tottenham Hotspur 38 19  7 12 58 53  5  64 Leag
ue Table
##    Season                                    KEY  X
## 3 2014/15           Premier League|Arsenal|2014/15 NA
## 4 2014/15 Premier League|Manchester United|2014/15 NA
## 5 2014/15 Premier League|Tottenham Hotspur|2014/15 NA
```

```
#View(df)
```

(2) Compare the summary statistics of goal difference `GD` between the home and away league tables.

```
# Write your answer here
#summary(df$Table.Type)
#table(df$Table.Type)
#a <- df[which(df$League=='Premier League'), ]
#summary(a$GD)
#summary(df[which(df$Table.Type=='Home League Table'), ]$GD)
#summary(df[which(df$Table.Type=='Away League Table'), ]$GD)

home <-(df[df$Table.Type=='Home League Table' & df$League=="Premier Lea
gue", ]$GD)
away <- df[df$Table.Type=='Away League Table' & df$League=="Premier Lea
gue", ]$GD

summary(home)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -11.00   -2.75    6.00    7.25   13.50   30.00
```

```
summary(away)
```

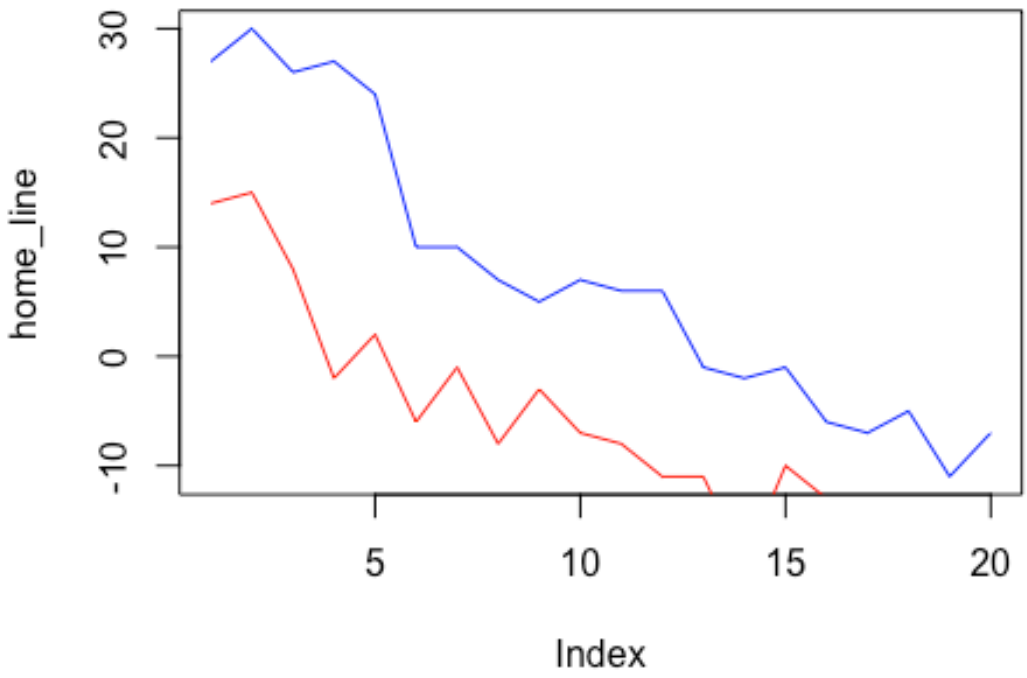
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -30.00  -13.50   -8.00   -7.25   -1.75   15.00
```

(3) Which one has more variation?

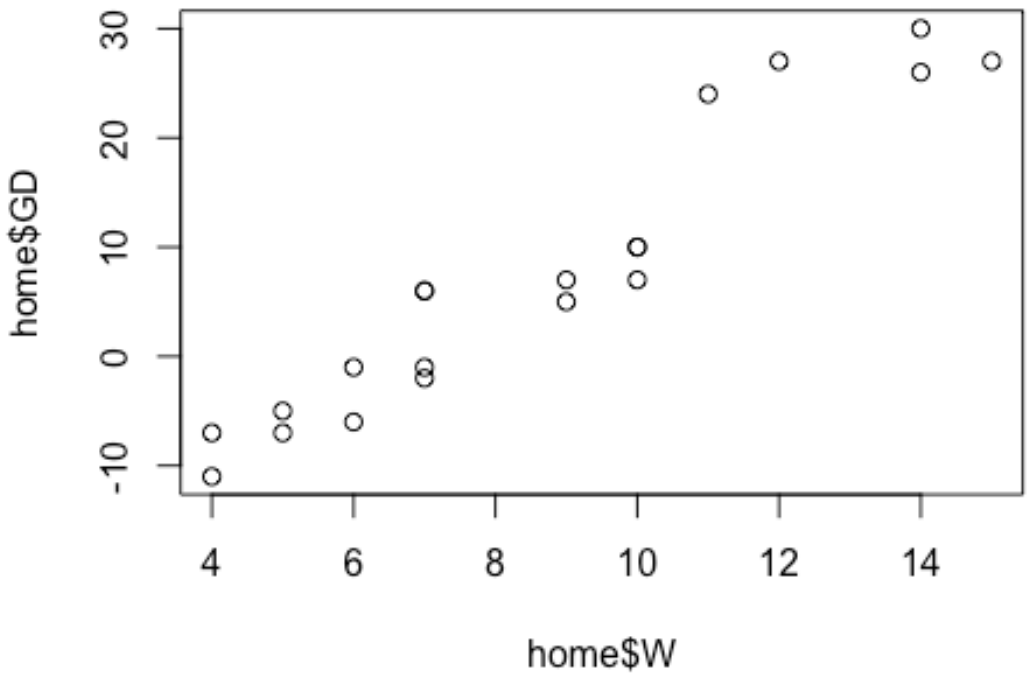
```
# Write your answer here

var(home)

## [1] 170.5132

var(away)

## [1] 129.9868
```

(4) Compare the relationship of number of wins and goal difference between the home and away games.

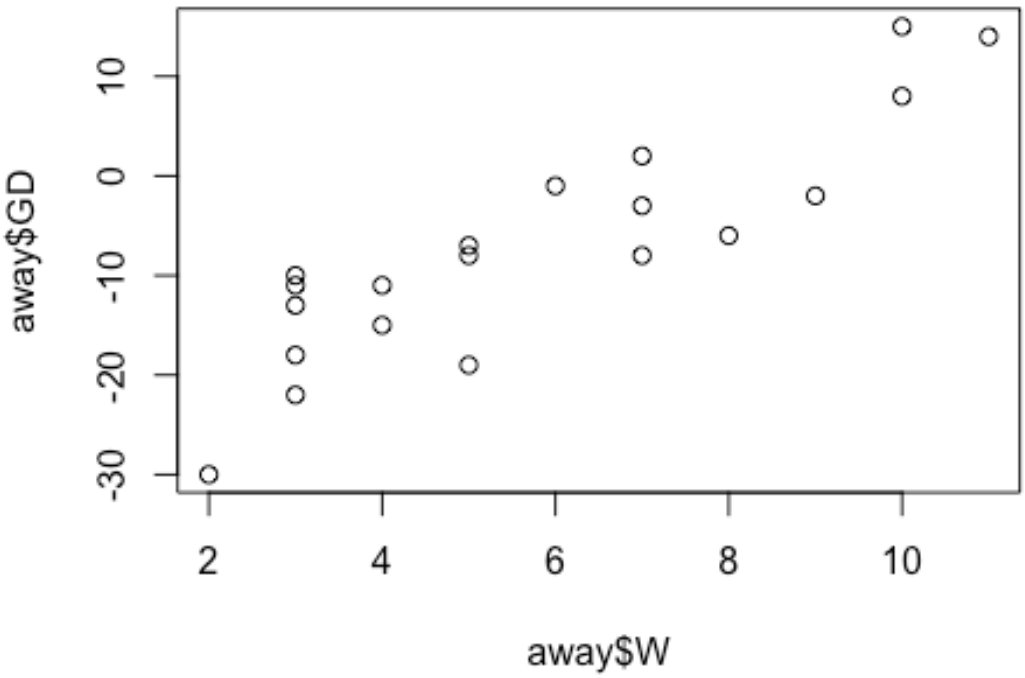
```
# Write your answer here

home_line <- df[df$Table.Type=='Home League Table' & df$League=="Premie
r League", ]$GD
away_line <- df[df$Table.Type=='Away League Table' & df$League=="Premie
r League", ]$GD



plot(home_line, type = "l", col = "blue")
lines(away_line, type="l", col = "red")
```




```
home <- df[df$Table.Type=='Home League Table' & df$League=="Premier Lea
gue", ]
away <- df[df$Table.Type=='Away League Table' & df$League=="Premier Lea
gue", ]
plot(home$W,home$GD)
```




```
plot(away$W,away$GD)
```



Further practice

You can practice more with other leagues in other countries from the file. You can also find more data from the `kaggle` open datasets (https://www.kaggle.com/datasets).