Predictions for the 2023 ICC Cricket World Cup

2023-11-30

```
library(ggplot2)
library(caTools)
library(randomForest)
```

```
## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:randomForest':
##
##     combine

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(fastDummies)
```

```
## Thank you for using fastDummies!

## To acknowledge our work, please cite the package:

## Kaplan, J. & Schlegel, B. (2023). fastDummies: Fast Creation of Dummy (Binary) Columns and Rows from
```

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.0
## v readr     2.1.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::combine()       masks randomForest::combine()
## x dplyr::filter()        masks stats::filter()
## x dplyr::lag()           masks stats::lag()
## x randomForest::margin() masks ggplot2::margin()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(caret)
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```r
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
##
## Attaching package: 'pROC'
##
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```r
library(gplots)
```

```
##
## Attaching package: 'gplots'
##
## The following object is masked from 'package:stats':
##
##     lowess
```

```r
library(corrplot)
```

```
## corrplot 0.92 loaded
```

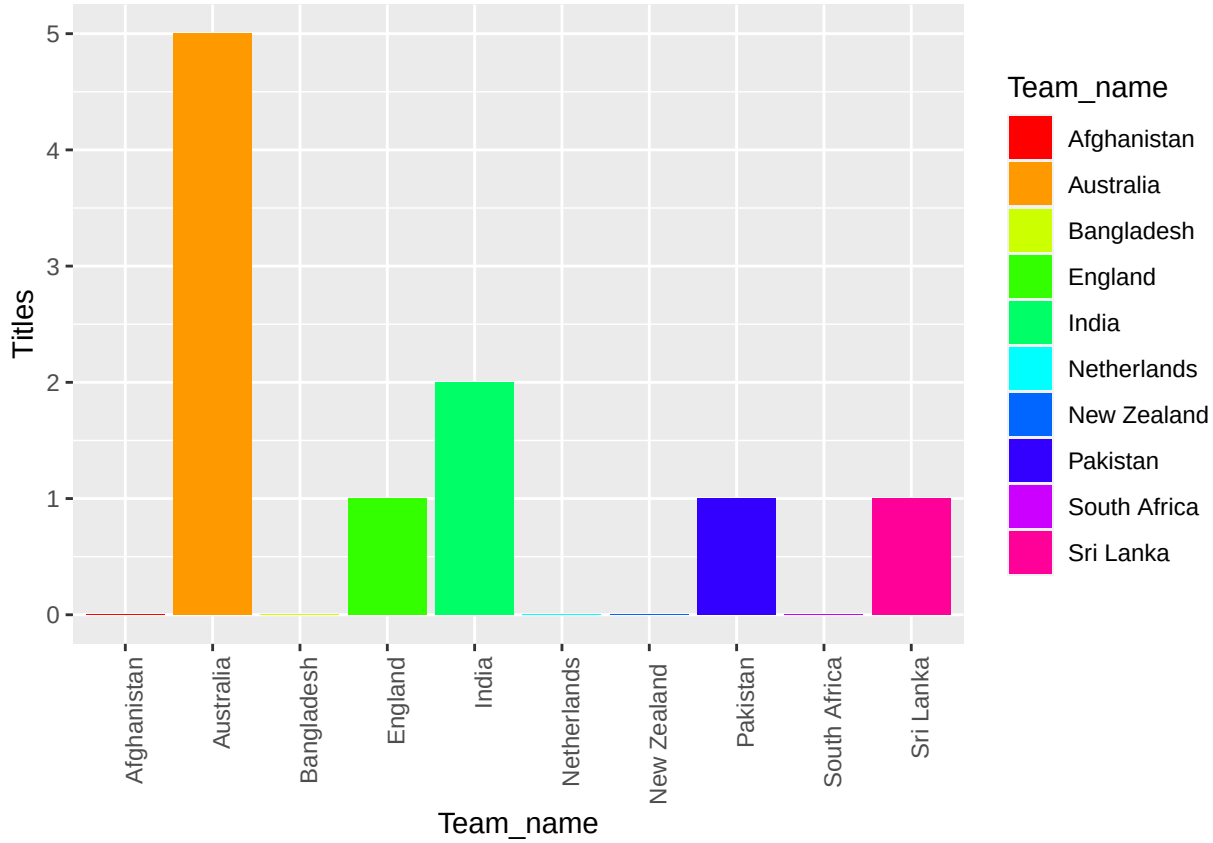
```r
library(ggcorrplot)
```

```r
#Reading the dataset
World_cup <- read.csv("World_cup_2023.csv")
results <- read.csv("results.csv")

World_cup
```

```
##       Team_name Team_ranking Titles Win_percentage_ODI WC_matches WC_match_won
## 1     Australia            1      5              60.73         94           69
## 2      Pakistan            2      1              52.78         79           45
## 3         India            3      2              52.38         84           53
## 4  New Zealand             4      0              45.89         89           54
## 5       England            5      1              50.32         83           48
## 6  South Africa            6      0              61.00         64           38
## 7    Bangladesh            7      0              36.65         40           14
## 8   Afghanistan            8      0              49.65         15            1
## 9     Sri Lanka            9      1              45.74         80           38
## 10  Netherlands           10      0              34.21         20            2
##    Win_percent_WC WC_match_loss Loss_percent_WC Tied No_result World_cup_winner
## 1           73.40            23           24.46    1         1              Yes
## 2           56.96            32           40.50    0         2              Yes
## 3           63.09            29           34.52    1         1              Yes
## 4           60.67            33           37.07    1         1               No
## 5           57.83            32           38.55    2         1              Yes
## 6           59.37            23           35.93    2         1               No
## 7           35.00            25           62.50    0         1               No
## 8            6.66            14           93.33    0         0               No
## 9           47.50            39           48.75    1         2              Yes
## 10          10.00            18           90.00    0         0               No
##    Recent_points Rating
## 1           2714    118
## 2           2316    116
## 3           3807    115
## 4           2806    104
## 5           2426    101
## 6           1910    101
## 7           2451     98
## 8           1361     91
## 9           2794     87
## 10          1044     37
```

No.of titles won by each teams

```r
ggplot(World_cup, aes(x = Team_name, y = Titles, fill = Team_name)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = rainbow(nrow(World_cup))) +  # Use rainbow colors
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

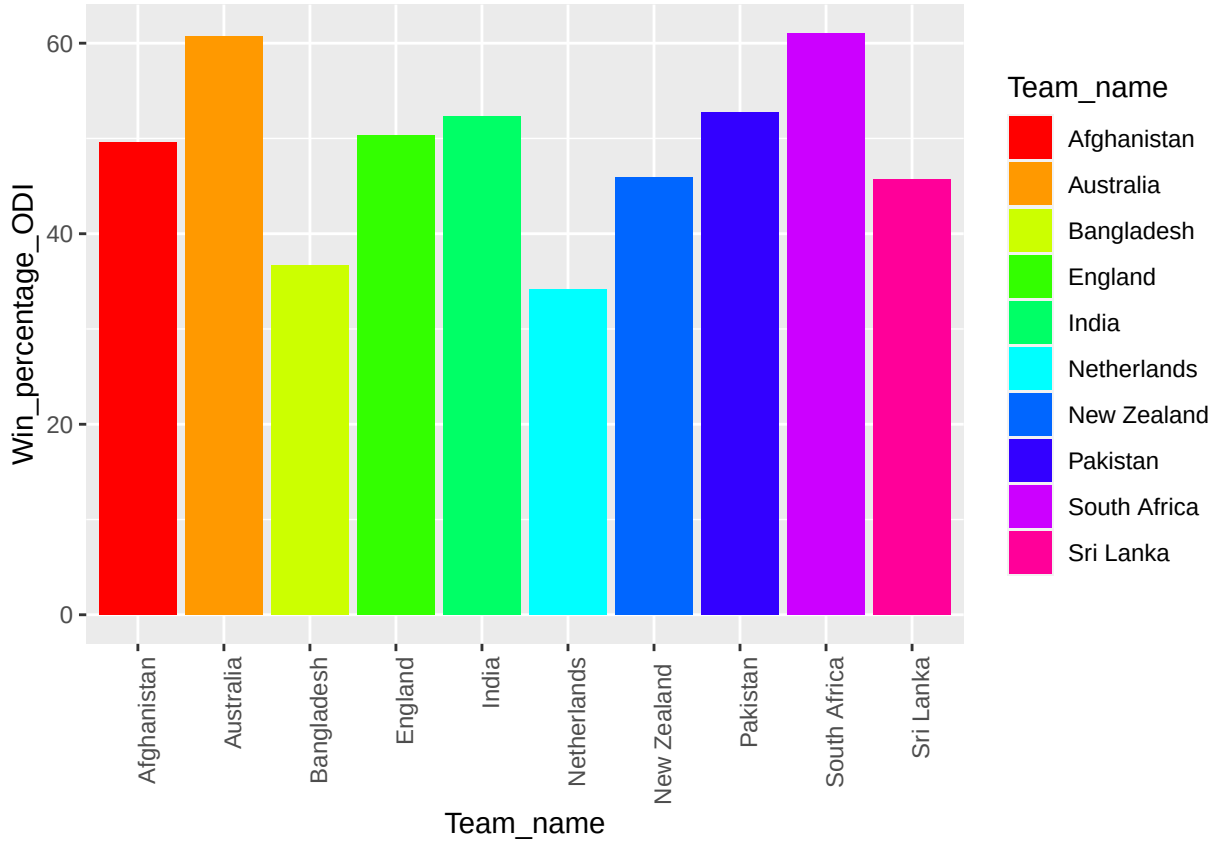



```
  labs(title = "Team Titles in World Cup",
       x = "Team Name",
       y = "Titles")
```

```
## $x
## [1] "Team Name"
##
## $y
## [1] "Titles"
##
## $title
## [1] "Team Titles in World Cup"
##
## attr(,"class")
## [1] "labels"
```

Win percentage in ODI by each team

```
ggplot(World_cup, aes(x = Team_name, y = Win_percentage_ODI, fill = Team_name)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = rainbow(nrow(World_cup))) +  # Use rainbow colors
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

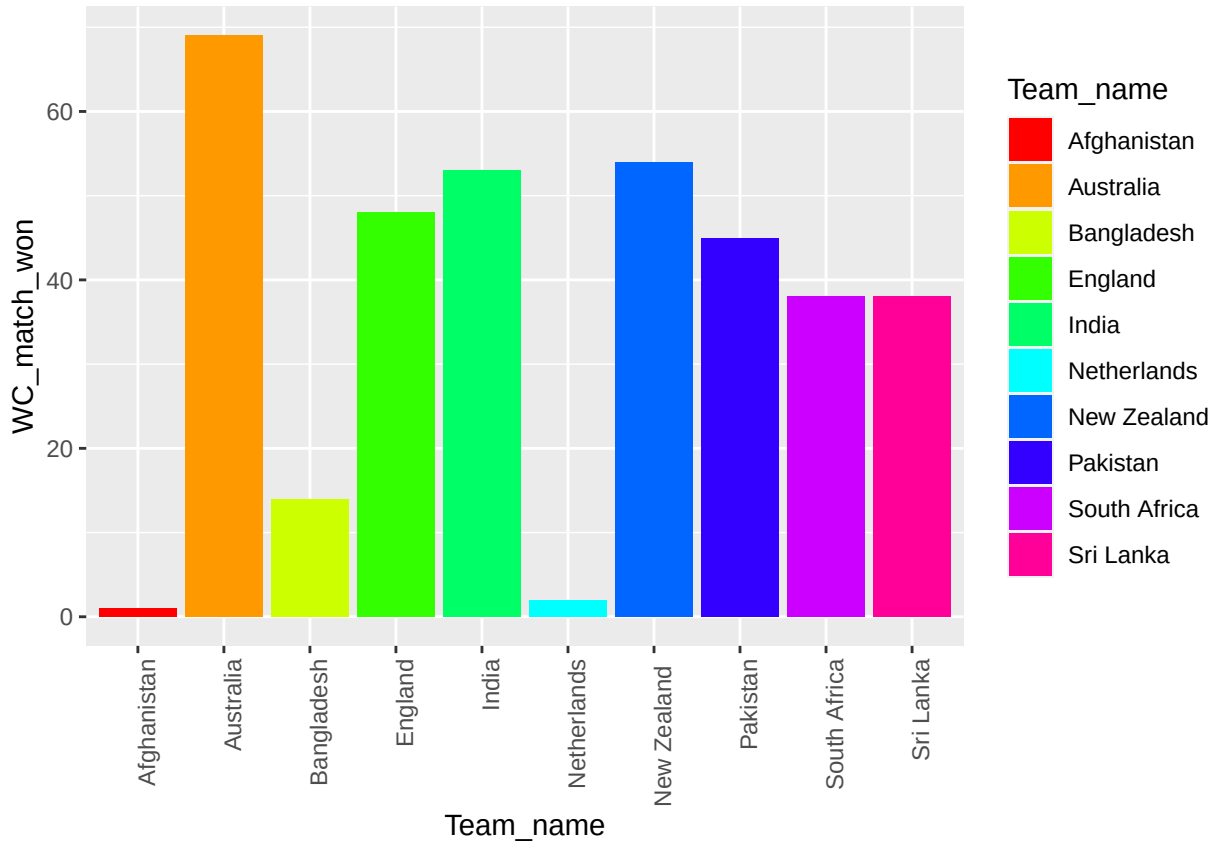



```
  labs(title = "Team Titles in World Cup",
       x = "Team Name",
       y = "Titles")
```

```
## $x
## [1] "Team Name"
##
## $y
## [1] "Titles"
##
## $title
## [1] "Team Titles in World Cup"
##
## attr(,"class")
## [1] "labels"
```

No.of matches won in world cup by each team

```
ggplot(World_cup, aes(x = Team_name, y = WC_match_won, fill = Team_name)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = rainbow(nrow(World_cup))) +  # Use rainbow colors
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

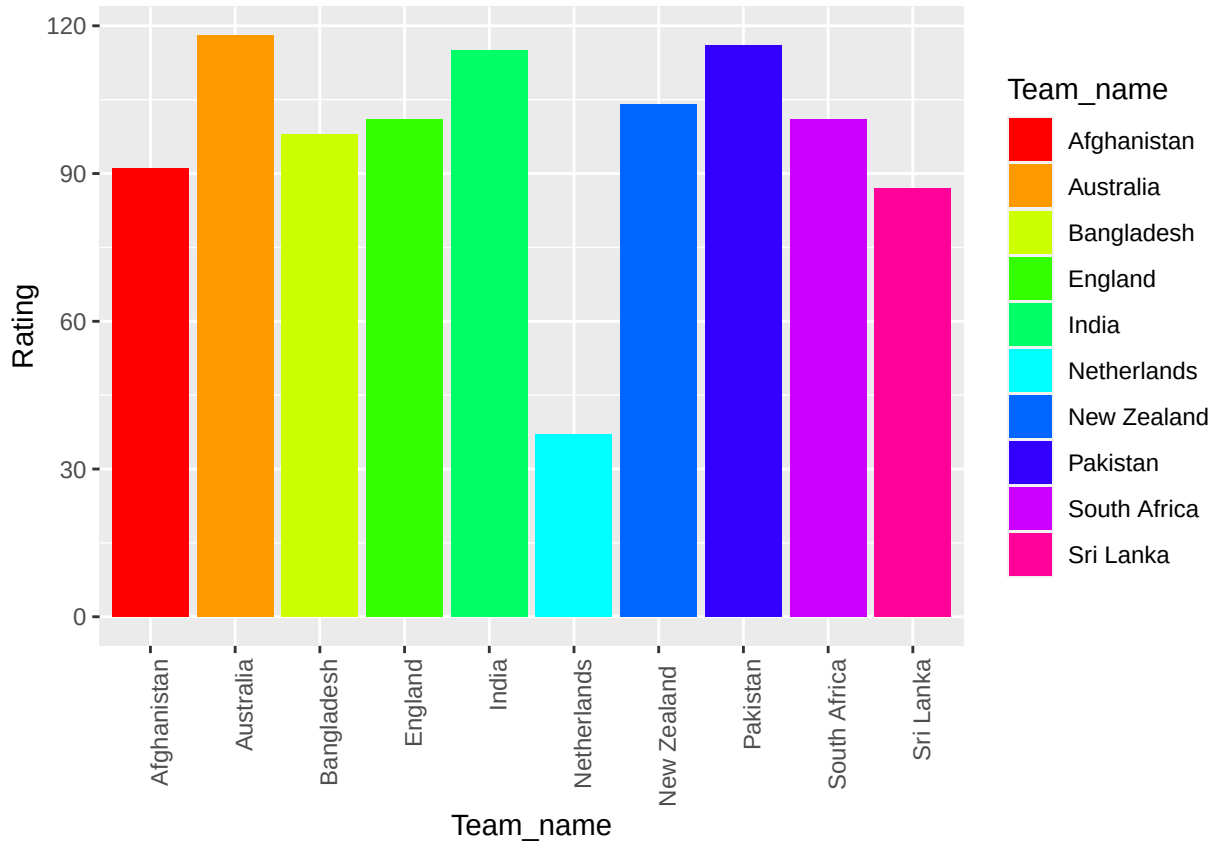



```
  labs(title = "Team Titles in World Cup",
       x = "Team Name",
       y = "Titles")
```

```
## $x
## [1] "Team Name"
##
## $y
## [1] "Titles"
##
## $title
## [1] "Team Titles in World Cup"
##
## attr(,"class")
## [1] "labels"
```

Recent ICC ODI rating

```
ggplot(World_cup, aes(x = Team_name, y = Rating, fill = Team_name)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = rainbow(nrow(World_cup))) +  # Use rainbow colors
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```




```
  labs(title = "Team Titles in World Cup",
       x = "Team Name",
       y = "Titles")
```

```
## $x
## [1] "Team Name"
##
## $y
## [1] "Titles"
##
## $title
## [1] "Team Titles in World Cup"
##
## attr(,"class")
## [1] "labels"
```

```
head(results)
```

```
##         Date      Team_1   Team_2      Winner           Margin
## 1 17/04/2015 Bangladesh  Pakistan Bangladesh    won by 79 runs
## 2 19/04/2015 Bangladesh  Pakistan Bangladesh  won by 7 wickets
## 3 22/04/2015 Bangladesh  Pakistan Bangladesh  won by 8 wickets
## 4 08/05/2015    Ireland   England   No result        No result
## 5 26/05/2015   Pakistan  Zimbabwe   Pakistan    won by 41 runs
## 6 29/05/2015   Pakistan  Zimbabwe   Pakistan  won by 6 wickets
```

```
##                          Ground
## 1 Shere Bangla National Stadium
## 2 Shere Bangla National Stadium
## 3 Shere Bangla National Stadium
## 4                   The Village
## 5               Gaddafi Stadium
## 6               Gaddafi Stadium
```

```
#Removing the Match abandoned and No result data from Winner column
results <- results[results$Winner != 'Match abandoned', ]
results <- results[results$Winner != 'No result', ]
```

Stats of top five teams in the ODI world cup 2023

Stats of Indian team

```
#Filtering India
df <- results[results$Team_1 == 'India' | results$Team_2 == 'India', ]
India <- df

head(India)
```

```
##          Date      Team_1 Team_2      Winner           Margin
## 12 18/06/2015  Bangladesh  India Bangladesh    won by 79 runs
## 14 21/06/2015 Bangladesh   India Bangladesh  won by 6 wickets
## 15 24/06/2015 Bangladesh   India       India   won by 77 runs
## 17 10/07/2015   Zimbabwe   India       India    won by 4 runs
## 20 12/07/2015   Zimbabwe   India       India   won by 62 runs
## 21 14/07/2015   Zimbabwe   India       India   won by 83 runs
##                          Ground
## 12 Shere Bangla National Stadium
## 14 Shere Bangla National Stadium
## 15 Shere Bangla National Stadium
## 17            Harare Sports Club
## 20            Harare Sports Club
## 21            Harare Sports Club
```

```
India_win <- India[India$Winner == 'India', ]
head(India_win)
```

```
##          Date      Team_1 Team_2 Winner           Margin
## 15 24/06/2015 Bangladesh   India  India   won by 77 runs
## 17 10/07/2015   Zimbabwe   India  India    won by 4 runs
## 20 12/07/2015   Zimbabwe   India  India   won by 62 runs
## 21 14/07/2015   Zimbabwe   India  India   won by 83 runs
## 74 23/01/2016  Australia   India  India won by 6 wickets
## 88 11/06/2016    Zimbabwe  India  India won by 9 wickets
##                          Ground
## 15 Shere Bangla National Stadium
## 17            Harare Sports Club
## 20            Harare Sports Club
## 21            Harare Sports Club
## 74         Sydney Cricket Ground
## 88            Harare Sports Club
```

No.of wins in ODIs against other teams

```
excluded_value <- 'India'
filtered_df <- India_win[India_win$Team_2 != excluded_value, ]
value_counts <- table(filtered_df$Team_2)
print(value_counts)
```

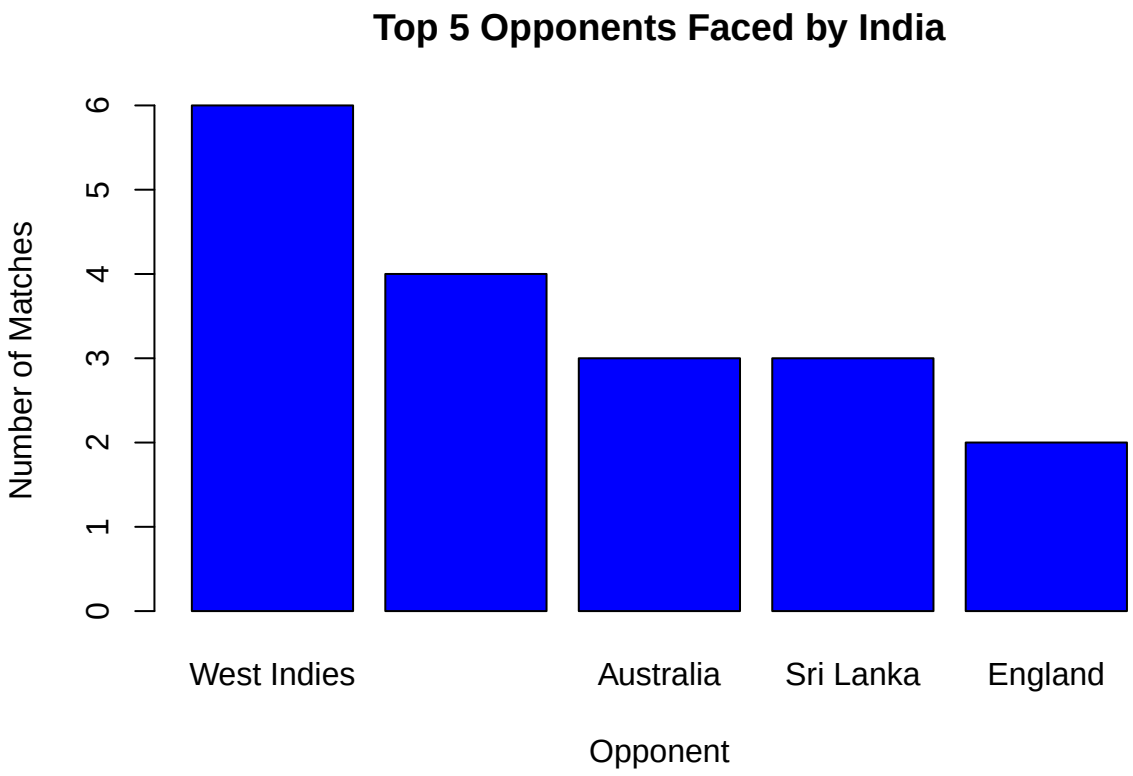
```
##
##    Australia      England  New Zealand South Africa    Sri Lanka  West Indies
##            3            2            4            2            3            6
```

```
excluded_value <- 'India'

filtered_df <- India_win[India_win$Team_1 != excluded_value, ]

value_counts <- table(filtered_df$Team_1)

print(value_counts)
```

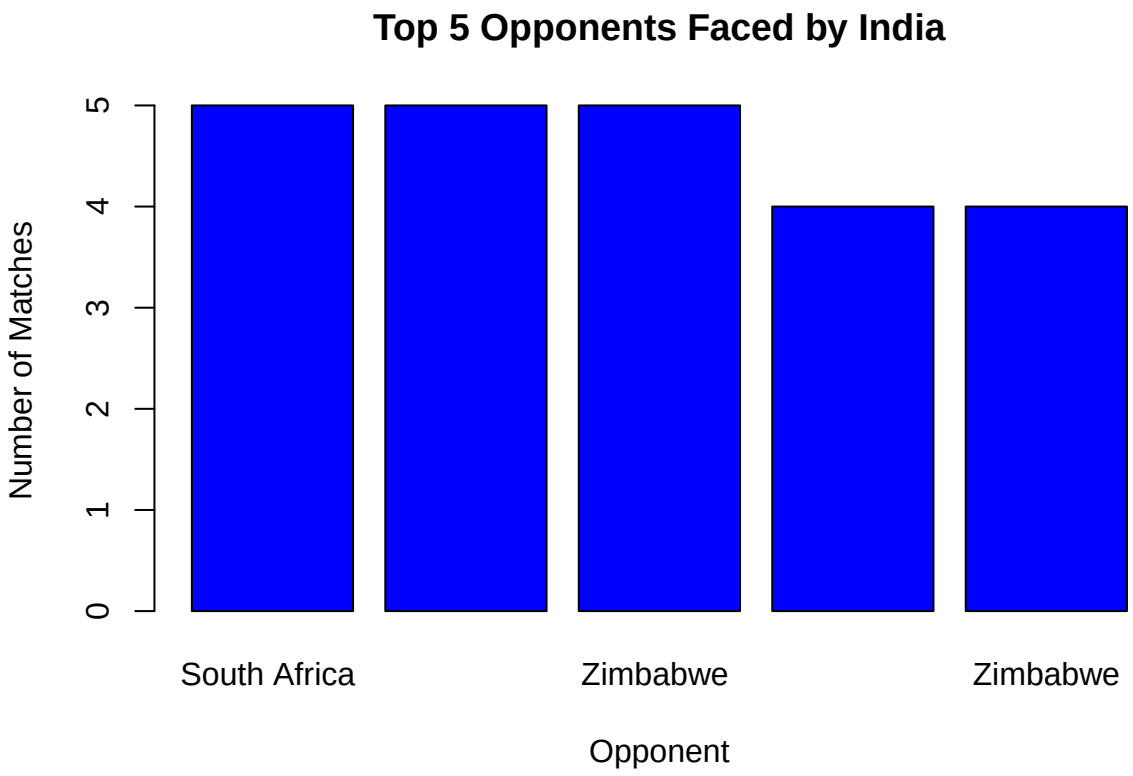
```
##
##   Afghanistan     Australia     Australia     Bangladesh   Bangladesh
##             1             1             2             1             4
##       England      England     Hong Kong   New Zealand  South Africa
##             2             1             1             1             5
##     Sri Lanka   West Indies      Zimbabwe     Zimbabwe
##             2             5             4             5
```

Displaying Team India's wins against other teams using bar graph

```
exclude <- 'India'

filtered_data <- India_win[India_win$Team_2 != exclude, ]

top_opponents <- head(names(sort(table(filtered_data$Team_2), decreasing = TRUE)), 5)
top_opponents_counts <- as.numeric(table(filtered_data$Team_2)[top_opponents])

barplot(top_opponents_counts, names.arg = top_opponents, col = "blue", main = "Top 5 Opponents Faced by
```

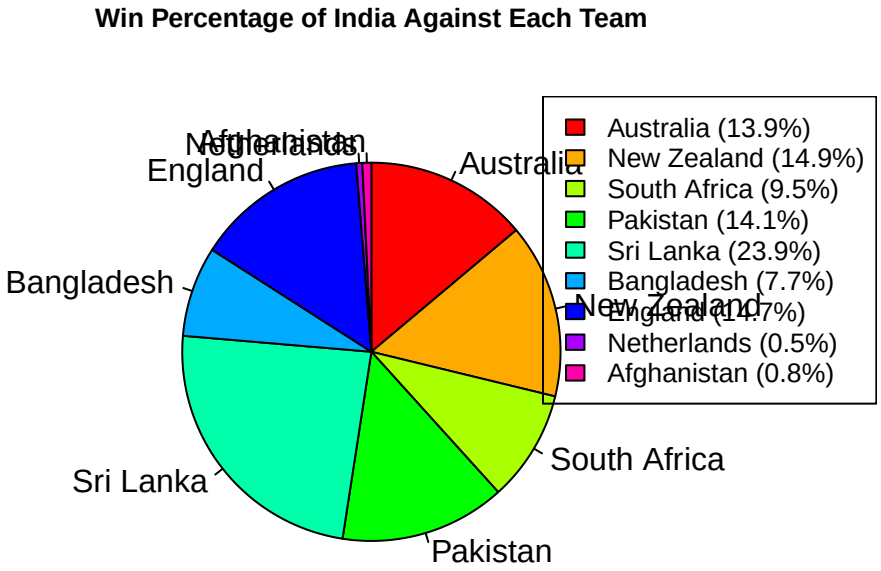



```r
exclude <- 'India'

filtered_data <- India_win[India_win$Team_1 != exclude, ]

top_opponents <- head(names(sort(table(filtered_data$Team_1), decreasing = TRUE)), 5)
top_opponents_counts <- as.numeric(table(filtered_data$Team_1)[top_opponents])

barplot(top_opponents_counts, names.arg = top_opponents, col = "blue", main = "Top 5 Opponents Faced by
```



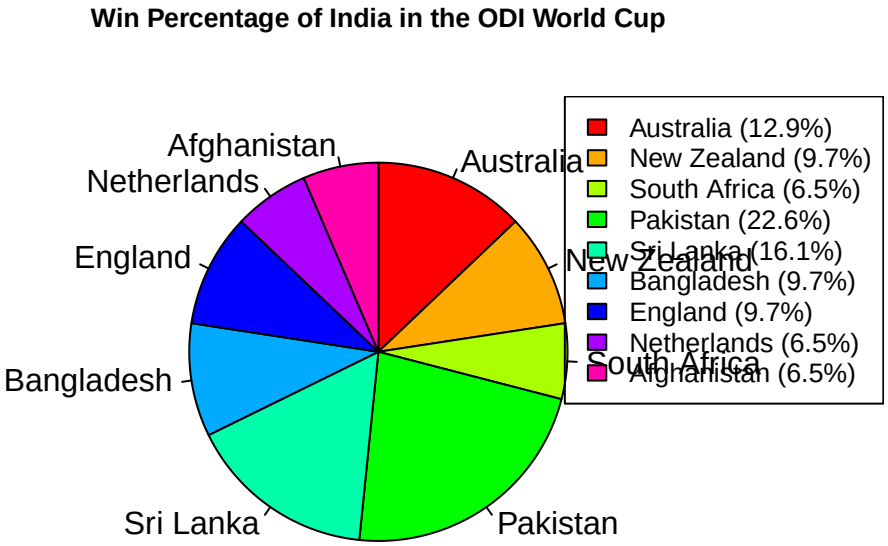
Win Percentage of India Against Each Team

```r
# Number of wins against each team
team_win_counts <- c(
  Australia = 54,
  `New Zealand` = 58,
  `South Africa` = 37,
  Pakistan = 55,
  `Sri Lanka` = 93,
  Bangladesh = 30,
  England = 57,
  Netherlands = 2,
  Afghanistan = 3
)

total_matches <- sum(team_win_counts)

win_percentages <- (team_win_counts / total_matches) * 100

# Pie chart
pie(win_percentages, labels = names(win_percentages), col = rainbow(length(win_percentages)),
    main = "Win Percentage of India Against Each Team", cex.main = 0.8, clockwise = TRUE)

legend("topright", legend = paste(names(win_percentages), sprintf("(%.1f%%)", win_percentages), sep = "
```



Win Percentage of India in the ODI world cup

```r
# Number of wins against each team in the ODI world cup
team_win_counts_wc_ind <- c(
  Australia = 4,
  `New Zealand` = 3,
  `South Africa` = 2,
  Pakistan = 7,
  `Sri Lanka` = 5,
  Bangladesh = 3,
  England = 3,
  Netherlands = 2,
  Afghanistan = 2
)

total_matches_wc_ind <- sum(team_win_counts_wc_ind)

win_percentages_wc_ind <- (team_win_counts_wc_ind / total_matches_wc_ind) * 100

pie(win_percentages_wc_ind, labels = names(win_percentages_wc_ind), col = rainbow(length(win_percentages
    main = "Win Percentage of India in the ODI World Cup", cex.main = 0.8, clockwise = TRUE)

legend("topright", legend = paste(names(win_percentages_wc_ind), sprintf("(%.1f%%)", win_percentages_wc
```



Stats of Australian team

```
# Filter the Australia Data
df1 <- results[results$Team_1 == 'Australia' | results$Team_2 == 'Australia', ]

Australia <- df1

head(Australia)
```

```
##          Date   Team_1    Team_2     Winner           Margin
## 33 27/08/2015 Ireland  Australia Australia    won by 23 runs
## 34 03/09/2015 England  Australia Australia    won by 59 runs
## 35 05/09/2015 England  Australia Australia    won by 64 runs
## 36 08/09/2015  England Australia    England   won by 93 runs
## 37 11/09/2015 England  Australia    England won by 3 wickets
## 38 12/09/2015 England  Australia  Australia won by 8 wickets
##                        Ground
## 33 Civil Service Cricket Club
## 34              The Rose Bowl
## 35                    Lord's
## 36               Old Trafford
## 37                 Headingley
## 38               Old Trafford
```

```r
Australia_win <- Australia[Australia$Winner == 'Australia', ]
head(Australia_win)
```

```
##           Date       Team_1       Team_2    Winner           Margin
## 38  12/09/2015     England     Australia Australia won by 8 wickets
## 80  06/02/2016 New Zealand     Australia Australia won by 4 wickets
## 86  05/06/2016 West Indies     Australia Australia won by 6 wickets
## 89  11/06/2016    Australia South Africa Australia   won by 36 runs
## 98  21/06/2016 West Indies     Australia Australia won by 6 wickets
## 102 26/06/2016 West Indies     Australia Australia   won by 58 runs
##                 Ground
## 38        Old Trafford
## 80         Sky Stadium
## 86  Providence Stadium
## 89         Warner Park
## 98     Kensington Oval
## 102    Kensington Oval
```

No.of wins for Australia against other teams

```r
excluded_value <- 'Australia'
filtered_df <- Australia_win[Australia_win$Team_2 != excluded_value, ]

value_counts <- table(filtered_df$Team_2)

print(value_counts)
```

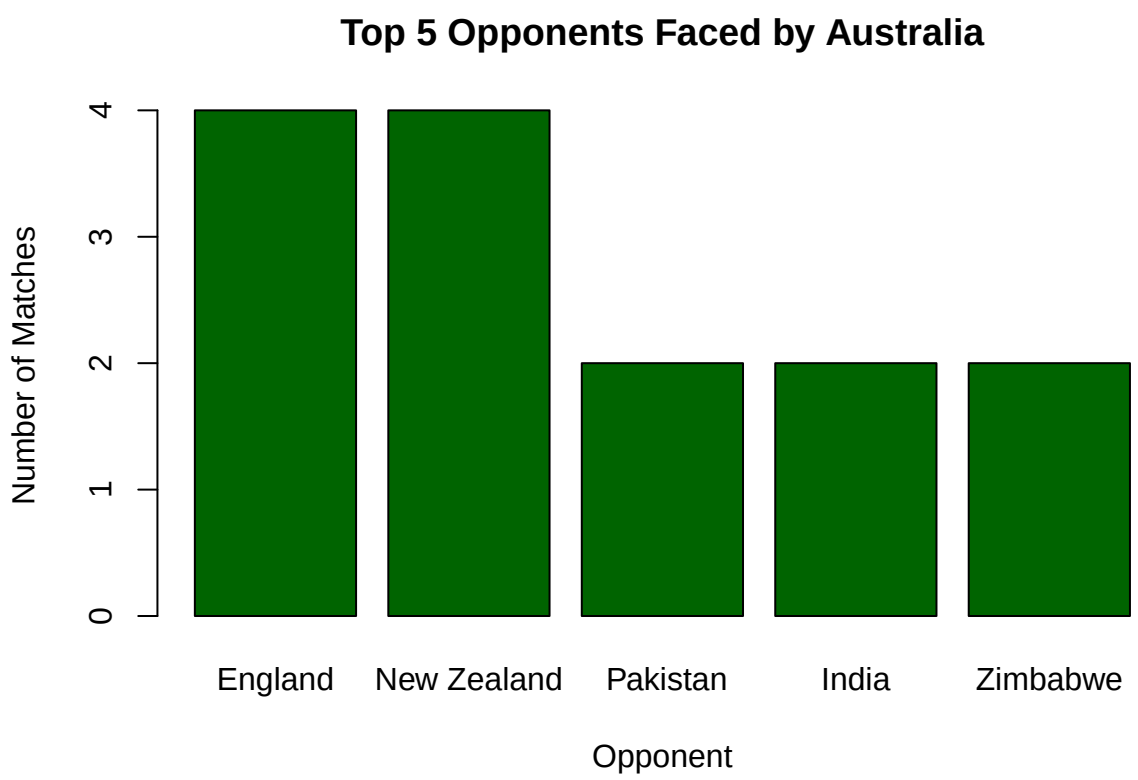
```
##
##     Pakistan      England        India  New Zealand     Pakistan South Africa
##            2            4            2            4            1            1
##     Zimbabwe
##            2
```

```r
excluded_value <- 'Australia'

filtered_df <- Australia_win[Australia_win$Team_1 != excluded_value, ]

value_counts <- table(filtered_df$Team_1)

print(value_counts)
```

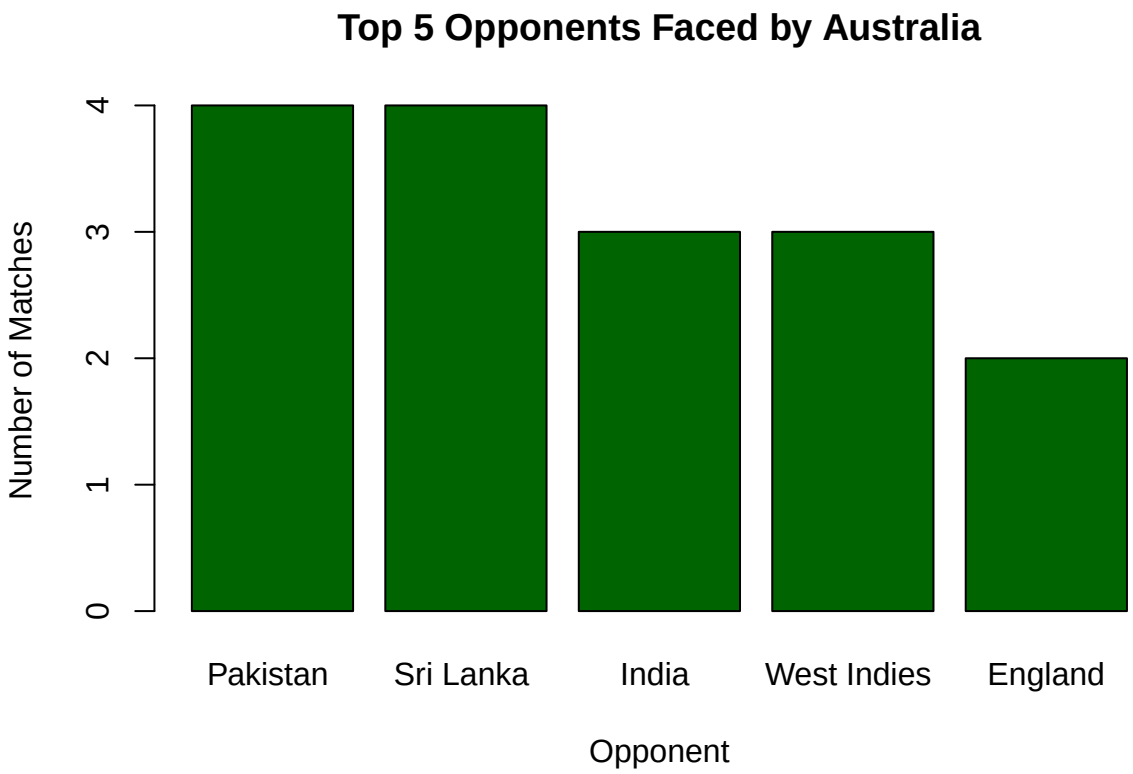
```
##
## Afghanistan       England     England         India       India  New Zealand
##           1            2           2            2           3            1
##     Pakistan    Pakistan     Sri Lanka   Sri Lanka   West Indies West Indies
##            2            4           2            4           2            3
```

Displaying Team Australia's wins against other teams using bar graph

```r
exclude <- 'Australia'

filtered_data <- Australia_win[Australia_win$Team_2 != exclude, ]

top_opponents <- head(names(sort(table(filtered_data$Team_2), decreasing = TRUE)), 5)
top_opponents_counts <- as.numeric(table(filtered_data$Team_2)[top_opponents])

barplot(top_opponents_counts, names.arg = top_opponents, col = "darkgreen", main = "Top 5 Opponents Face
```

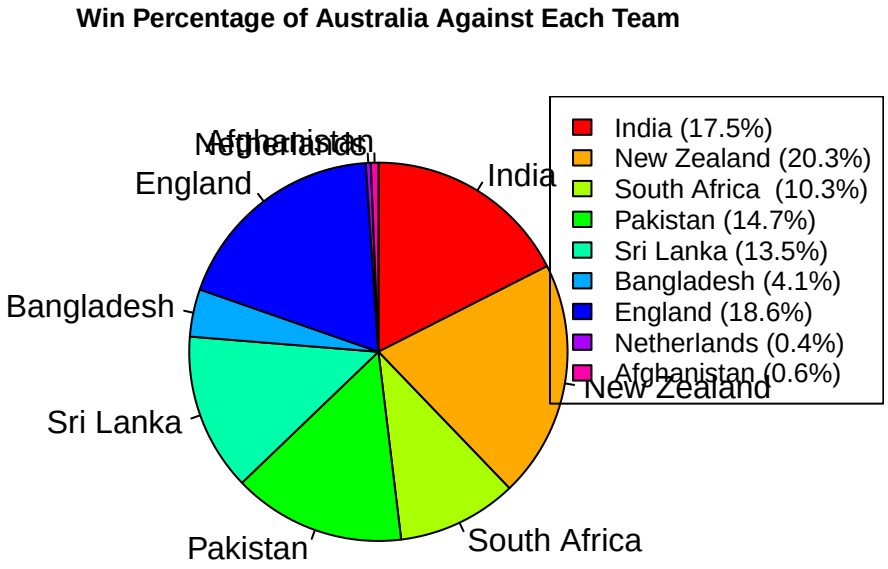



```r
exclude <- 'Australia'

filtered_data <- Australia_win[Australia_win$Team_1 != exclude, ]

top_opponents <- head(names(sort(table(filtered_data$Team_1), decreasing = TRUE)), 5)
top_opponents_counts <- as.numeric(table(filtered_data$Team_1)[top_opponents])

barplot(top_opponents_counts, names.arg = top_opponents, col = "darkgreen", main = "Top 5 Opponents Face
```



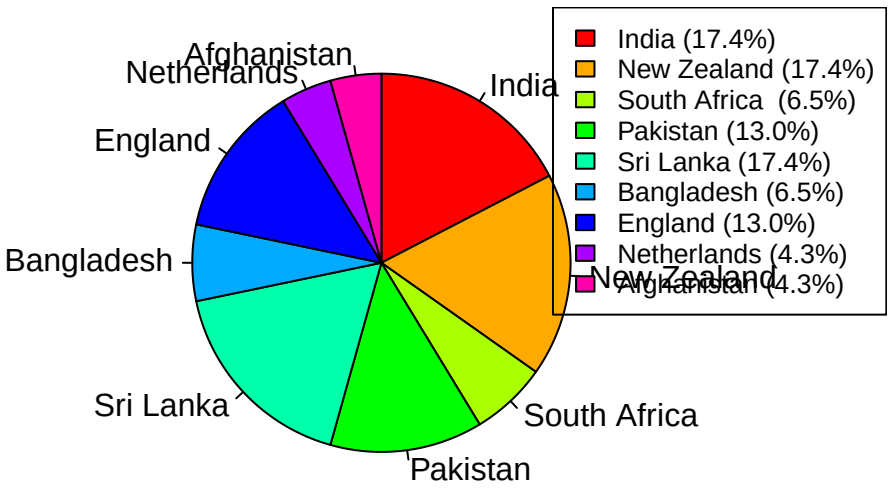
Win Percentage of Australia Against Each Team

```r
# Number of wins against each team
team_win_counts <- c(
  'India'= 82,
  'New Zealand'= 95,
  'South Africa '= 48,
  'Pakistan'= 69,
  'Sri Lanka'= 63,
  'Bangladesh'= 19,
  'England'= 87,
  'Netherlands'= 2,
  'Afghanistan'= 3
)

total_matches <- sum(team_win_counts)

win_percentages <- (team_win_counts / total_matches) * 100

pie(win_percentages, labels = names(win_percentages), col = rainbow(length(win_percentages)),
    main = "Win Percentage of Australia Against Each Team", cex.main = 0.8, clockwise = TRUE)


legend("topright", legend = paste(names(win_percentages), sprintf("(%.1f%%)", win_percentages), sep = "
```



Win Percentage of Australia in the ODI world cup

```r
# Number of wins against each team in the ODI world cup
team_win_counts_wc_ind <- c(
  'India'= 8,
  'New Zealand'= 8,
  'South Africa '= 3,
  'Pakistan'= 6,
  'Sri Lanka'= 8,
  'Bangladesh'= 3,
  'England'= 6,
  'Netherlands'= 2,
  'Afghanistan'= 2
)

total_matches_wc_ind <- sum(team_win_counts_wc_ind)

win_percentages_wc_ind <- (team_win_counts_wc_ind / total_matches_wc_ind) * 100

pie(win_percentages_wc_ind, labels = names(win_percentages_wc_ind), col = rainbow(length(win_percentages
    main = "Win Percentage of Australia in the ODI World Cup", cex.main = 0.8, clockwise = TRUE)

legend("topright", legend = paste(names(win_percentages_wc_ind), sprintf("(%.1f%%)", win_percentages_wc
```



Stats of Pakistan team

```
#Filter the Pakistan Data
df2 <- results[results$Team_1 == 'Pakistan' | results$Team_2 == 'Pakistan', ]
Pakistan <- df2

head(Pakistan)
```

```
##          Date      Team_1   Team_2      Winner           Margin
## 1  17/04/2015 Bangladesh  Pakistan Bangladesh    won by 79 runs
## 2  19/04/2015 Bangladesh  Pakistan Bangladesh  won by 7 wickets
## 3  22/04/2015 Bangladesh  Pakistan Bangladesh  won by 8 wickets
## 18 11/07/2015  Sri Lanka  Pakistan    Pakistan won by 6 wickets
## 22 15/07/2015  Sri Lanka  Pakistan  Sri Lanka  won by 2 wickets
## 24 19/07/2015  Sri Lanka  Pakistan    Pakistan  won by 135 runs
##                                    Ground
## 1            Shere Bangla National Stadium
## 2            Shere Bangla National Stadium
## 3            Shere Bangla National Stadium
## 18  Rangiri Dambulla International Stadium
## 22 Pallekele International Cricket Stadium
## 24                    R Premadasa Stadium
```

```
Pakistan_win <- Pakistan[Pakistan$Winner == 'Pakistan', ]
head(Pakistan_win)
```

```
##           Date     Team_1   Team_2   Winner           Margin
## 18  11/07/2015 Sri Lanka  Pakistan Pakistan won by 6 wickets
## 24  19/07/2015 Sri Lanka  Pakistan Pakistan  won by 135 runs
## 25  22/07/2015 Sri Lanka  Pakistan Pakistan won by 7 wickets
## 39  01/10/2015  Zimbabwe  Pakistan Pakistan  won by 131 runs
## 41  05/10/2015  Zimbabwe  Pakistan Pakistan won by 7 wickets
## 112 18/08/2016   Ireland  Pakistan Pakistan  won by 255 runs
##                                    Ground
## 18  Rangiri Dambulla International Stadium
## 24                      R Premadasa Stadium
## 25                      R Premadasa Stadium
## 39                       Harare Sports Club
## 41                       Harare Sports Club
## 112                             The Village
```

No.of wins for Pakistan against other teams

```
excluded_value <- 'Pakistan'

filtered_df <- Pakistan_win[Pakistan_win$Team_2 != excluded_value, ]

value_counts <- table(filtered_df$Team_2)

print(value_counts)
```

```
##
##   Australia New Zealand   Sri Lanka West Indies    Zimbabwe
##           2           5           2           3           2
```

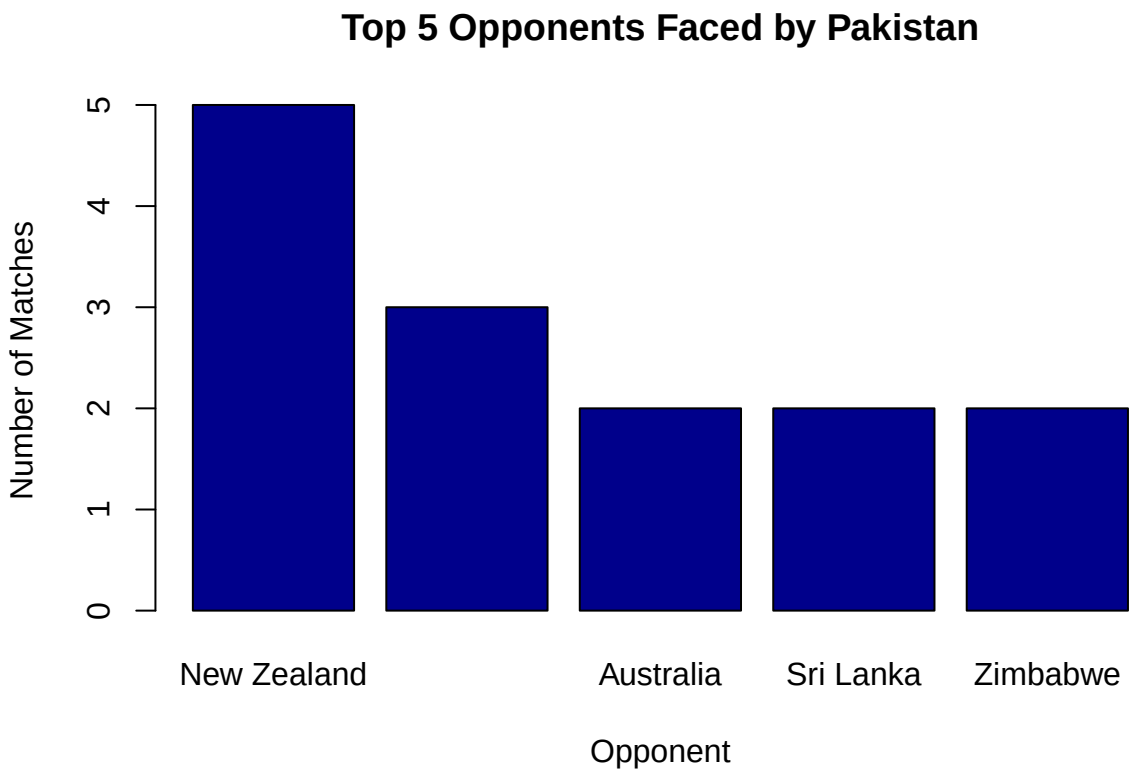
```
excluded_value <- 'Pakistan'

filtered_df <- Pakistan_win[Pakistan_win$Team_1 != excluded_value, ]

value_counts <- table(filtered_df$Team_1)

print(value_counts)
```

```
##
##  Afghanistan     Australia       England      Hong Kong       Ireland
##            2             1             2             1             1
##   Netherlands  New Zealand   South Africa South Africa     Sri Lanka
##            3             1             2             1             3
##     Zimbabwe
##            7
```

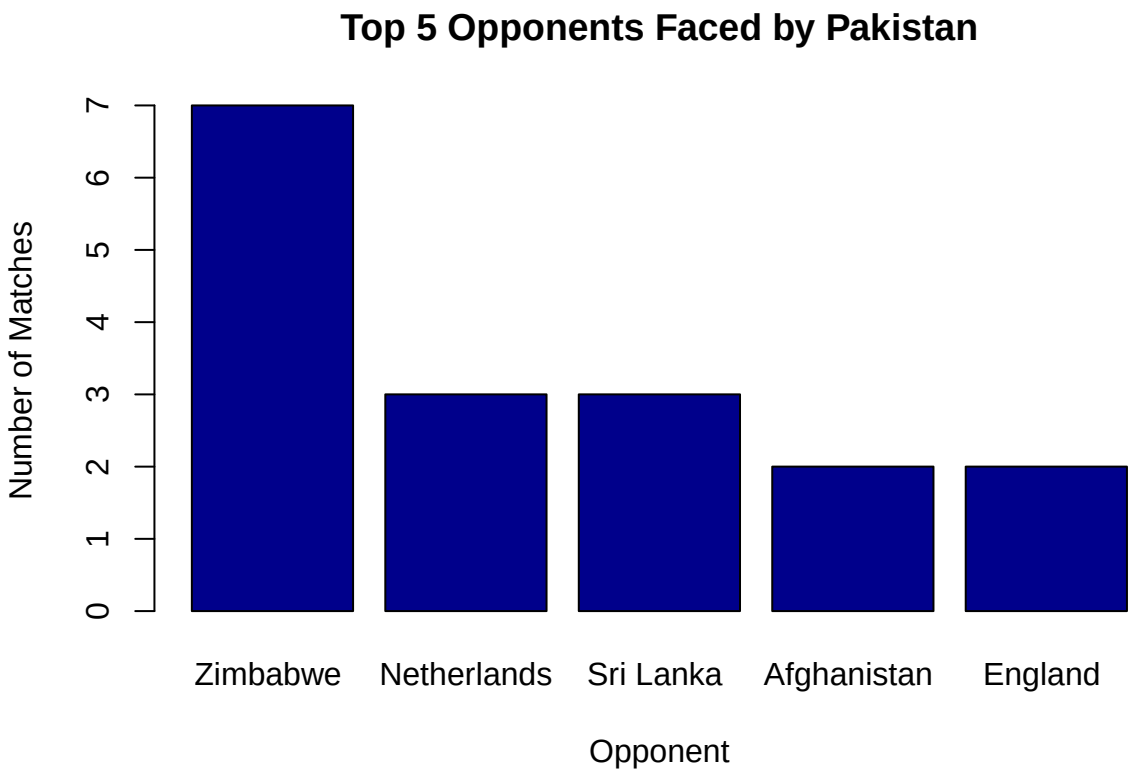
Displaying Team Pakistan's wins against other teams using bar graph

```
exclude <- 'Pakistan'

filtered_data <- Pakistan_win[Pakistan_win$Team_2 != exclude, ]

top_opponents <- head(names(sort(table(filtered_data$Team_2), decreasing = TRUE)), 5)
```

```r
top_opponents_counts <- as.numeric(table(filtered_data$Team_2)[top_opponents])

barplot(top_opponents_counts, names.arg = top_opponents, col = "darkblue", main = "Top 5 Opponents Faced
```

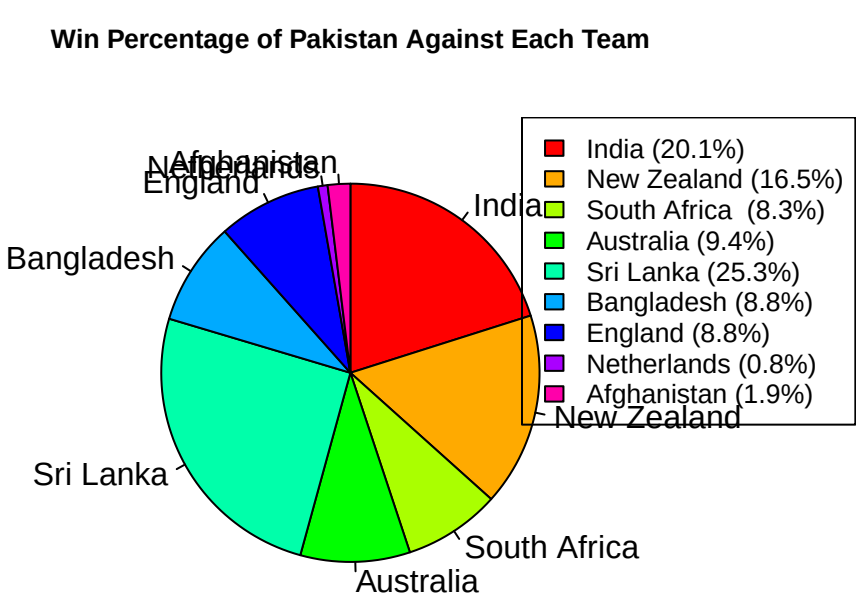



```r
exclude <- 'Pakistan'

filtered_data <- Pakistan_win[Pakistan_win$Team_1 != exclude, ]

top_opponents <- head(names(sort(table(filtered_data$Team_1), decreasing = TRUE)), 5)
top_opponents_counts <- as.numeric(table(filtered_data$Team_1)[top_opponents])

barplot(top_opponents_counts, names.arg = top_opponents, col = "darkblue", main = "Top 5 Opponents Faced
```



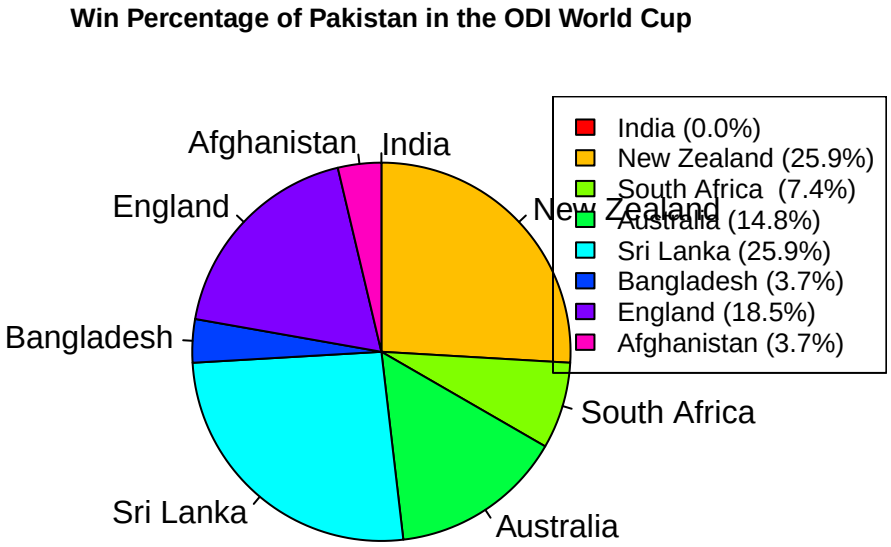
Win Percentage of Pakistan Against Each Team

```r
# Number of wins against each team
team_win_counts <- c(
  'India'= 73,
  'New Zealand'= 60,
  'South Africa '= 30,
  'Australia'= 34,
  'Sri Lanka'= 92,
  'Bangladesh'= 32,
  'England'= 32,
  'Netherlands'= 3,
  'Afghanistan'= 7
)

total_matches <- sum(team_win_counts)

win_percentages <- (team_win_counts / total_matches) * 100

pie(win_percentages, labels = names(win_percentages), col = rainbow(length(win_percentages)),
    main = "Win Percentage of Pakistan Against Each Team", cex.main = 0.8, clockwise = TRUE)


legend("topright", legend = paste(names(win_percentages), sprintf("(%.1f%%)", win_percentages), sep = "
```



Win Percentage of Pakistan in the ODI world cup

```
# Number of wins against each team in the ODI world cup
team_win_counts_wc_ind <- c(
  'India'= 0,
  'New Zealand'= 7,
  'South Africa '= 2,
  'Australia'= 4,
  'Sri Lanka'= 7,
  'Bangladesh'= 1,
  'England'= 5,
  'Afghanistan'= 1
)

total_matches_wc_ind <- sum(team_win_counts_wc_ind)

win_percentages_wc_ind <- (team_win_counts_wc_ind / total_matches_wc_ind) * 100

pie(win_percentages_wc_ind, labels = names(win_percentages_wc_ind), col = rainbow(length(win_percentages
    main = "Win Percentage of Pakistan in the ODI World Cup", cex.main = 0.8, clockwise = TRUE)

legend("topright", legend = paste(names(win_percentages_wc_ind), sprintf("(%.1f%%)", win_percentages_wc
```



Stats of New Zealand team

```
# Filter the New Zealand data
df3 <- results[results$Team_1 == 'New Zealand' | results$Team_2 == 'New Zealand', ]
New_Zealand <- df3

head(New_Zealand)
```

```
##          Date    Team_1      Team_2       Winner           Margin
## 8  09/06/2015  England  New Zealand     England   won by 210 runs
## 9  12/06/2015   England New Zealand New Zealand    won by 13 runs
## 10 14/06/2015  England  New Zealand New Zealand  won by 3 wickets
## 11 17/06/2015  England  New Zealand      England won by 7 wickets
## 13 20/06/2015  England  New Zealand      England won by 3 wickets
## 27 02/08/2015 Zimbabwe  New Zealand    Zimbabwe  won by 7 wickets
##                Ground
## 8           Edgbaston
## 9     Kennington Oval
## 10      The Rose Bowl
## 11       Trent Bridge
## 13   Riverside Ground
## 27 Harare Sports Club
```

```
New_Zealand_win <- New_Zealand[New_Zealand$Winner == 'New Zealand', ]
head(New_Zealand_win)
```

```
##           Date      Team_1      Team_2      Winner            Margin
## 28  04/08/2015   Zimbabwe  New Zealand New Zealand won by 10 wickets
## 29  07/08/2015   Zimbabwe  New Zealand New Zealand    won by 38 runs
## 77  31/01/2016 New Zealand    Pakistan New Zealand  won by 3 wickets
## 82  08/02/2016 New Zealand   Australia New Zealand    won by 55 runs
## 141 20/10/2016      India  New Zealand New Zealand     won by 6 runs
## 143 26/10/2016      India  New Zealand New Zealand    won by 19 runs
##                              Ground
## 28                  Harare Sports Club
## 29                  Harare Sports Club
## 77                          Eden Park
## 82                         Seddon Park
## 141               Arun Jaitley Stadium
## 143 JSCA International Stadium Complex
```

No.of wins for New Zealand in ODIs against other teams

```
excluded_value <- 'New Zealand'

filtered_df <- New_Zealand_win[New_Zealand_win$Team_2 != excluded_value, ]

value_counts <- table(filtered_df$Team_2)

print(value_counts)
```

```
##
##    Australia   Bangladesh        India  Netherlands     Pakistan South Africa
##            1            5            4            3            2            1
##    Sri Lanka
##            3
```

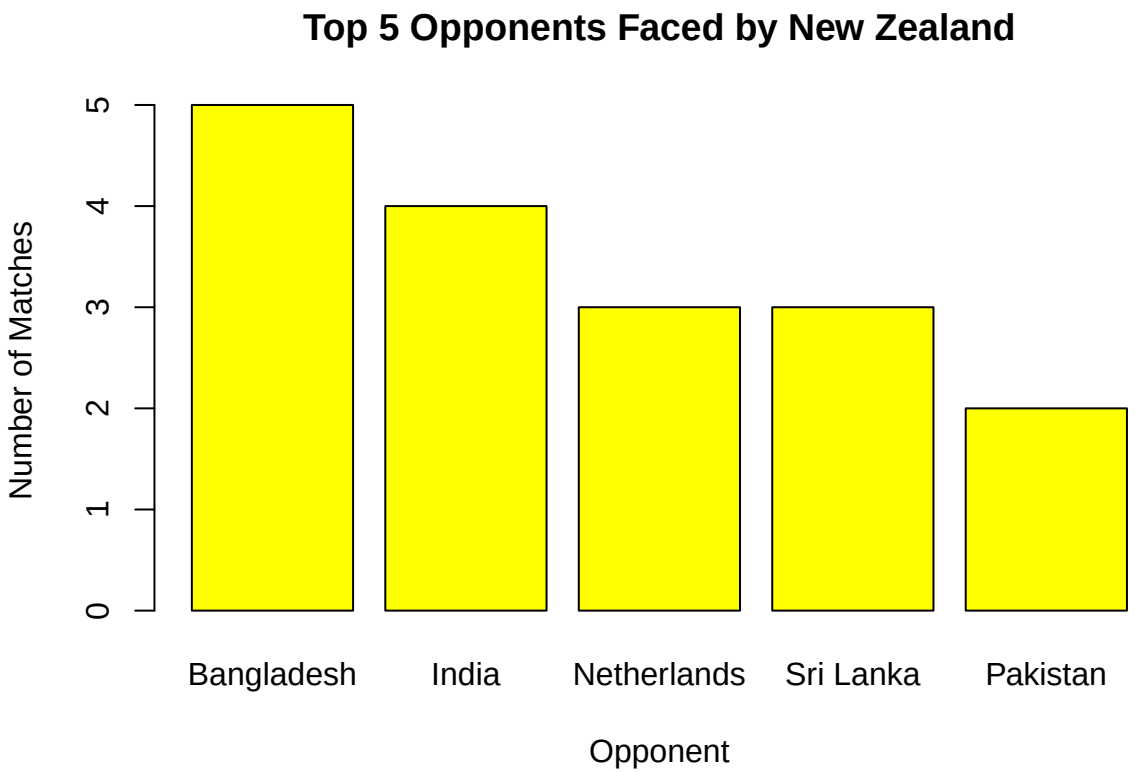
```
excluded_value <- 'New Zealand'

filtered_df <- New_Zealand_win[New_Zealand_win$Team_1 != excluded_value, ]

value_counts <- table(filtered_df$Team_1)

print(value_counts)
```

```
##
## Afghanistan   Bangladesh        India       Ireland     Ireland      Pakistan
##           1            1            3            3            1            3
##     Scotland  West Indies    Zimbabwe
##            1            2            2
```

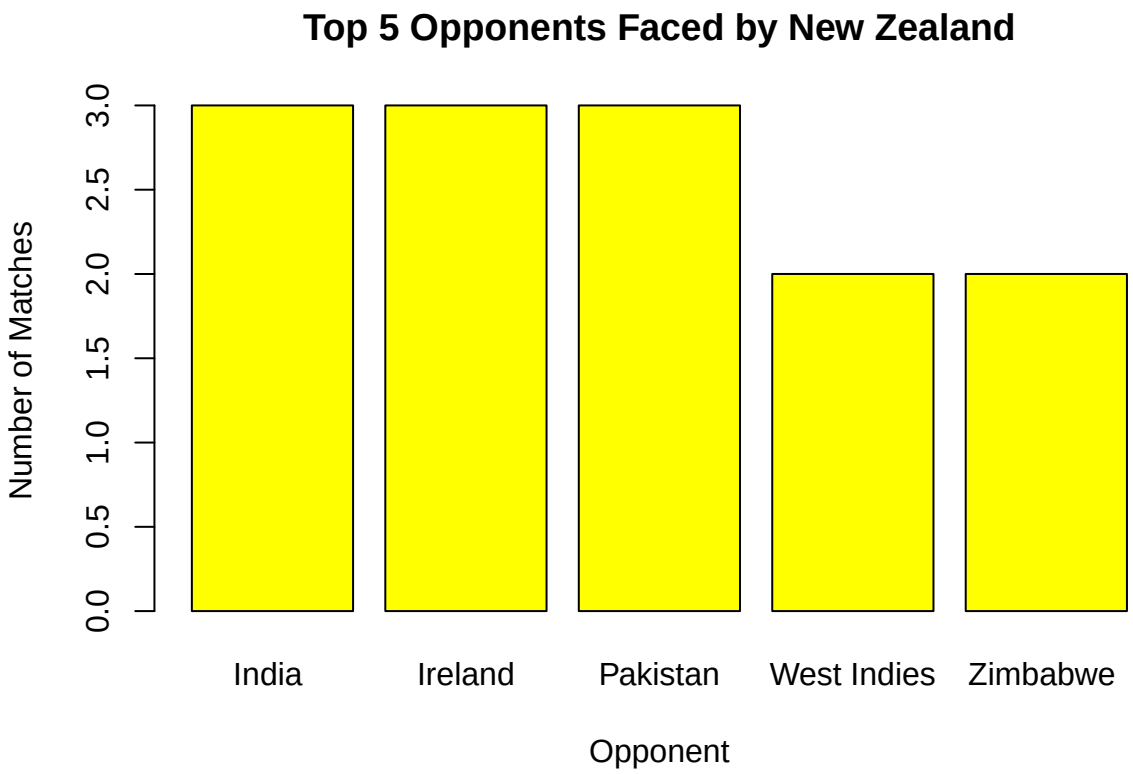
Displaying Team New Zealand's wins against other teams using bar graph

```
exclude <- 'New Zealand'

filtered_data <- New_Zealand_win[New_Zealand_win$Team_2 != exclude, ]

top_opponents <- head(names(sort(table(filtered_data$Team_2), decreasing = TRUE)), 5)
```

```r
top_opponents_counts <- as.numeric(table(filtered_data$Team_2)[top_opponents])

barplot(top_opponents_counts, names.arg = top_opponents, col = "yellow", main = "Top 5 Opponents Faced
```

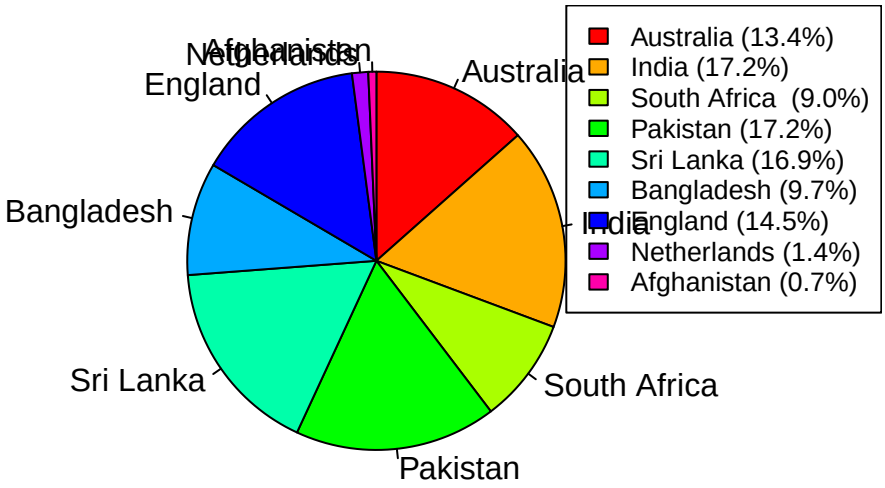



```r
exclude <- 'New Zealand'

filtered_data <- New_Zealand_win[New_Zealand_win$Team_1 != exclude, ]

top_opponents <- head(names(sort(table(filtered_data$Team_1), decreasing = TRUE)), 5)
top_opponents_counts <- as.numeric(table(filtered_data$Team_1)[top_opponents])

barplot(top_opponents_counts, names.arg = top_opponents, col = "yellow", main = "Top 5 Opponents Faced
```



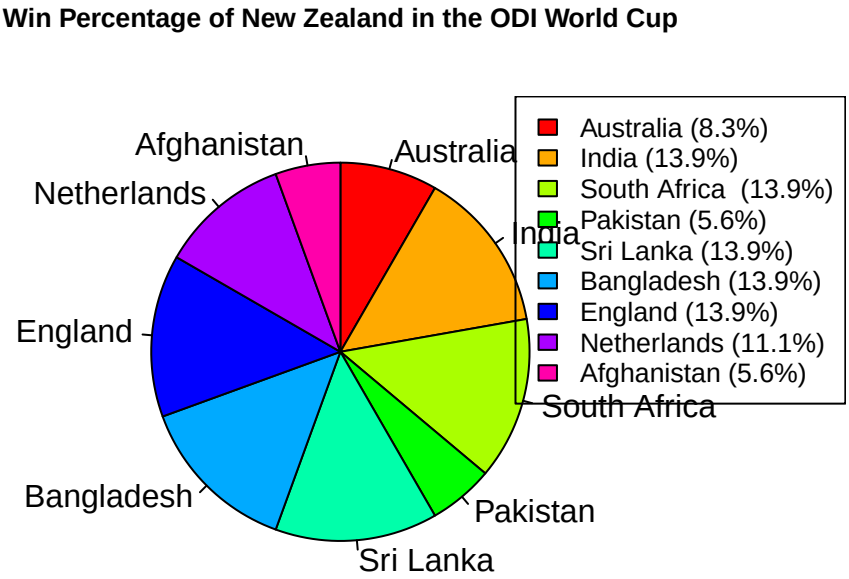
Win Percentage of New Zealand Against Each Team

```r
# Number of wins against each team
team_win_counts <- c(
  'Australia'= 39,
  'India'= 50,
  'South Africa '= 26,
  'Pakistan'= 50,
  'Sri Lanka'= 49,
  'Bangladesh'= 28,
  'England'= 42,
  'Netherlands'= 4,
  'Afghanistan'= 2
)

total_matches <- sum(team_win_counts)

win_percentages <- (team_win_counts / total_matches) * 100

pie(win_percentages, labels = names(win_percentages), col = rainbow(length(win_percentages)),
    main = "Win Percentage of New Zealand Against Each Team", cex.main = 0.8, clockwise = TRUE)


legend("topright", legend = paste(names(win_percentages), sprintf("(%.1f%%)", win_percentages), sep = "
```



Win Percentage of New Zealand in the ODI world cup

```
# Number of wins against each team in the ODI world cup
team_win_counts_wc_ind <- c(
  'Australia'= 3,
  'India'= 5,
  'South Africa '= 5,
  'Pakistan'= 2,
  'Sri Lanka'= 5,
  'Bangladesh'= 5,
  'England'= 5,
  'Netherlands'= 4,
  'Afghanistan'= 2
)

total_matches_wc_ind <- sum(team_win_counts_wc_ind)

win_percentages_wc_ind <- (team_win_counts_wc_ind / total_matches_wc_ind) * 100

pie(win_percentages_wc_ind, labels = names(win_percentages_wc_ind), col = rainbow(length(win_percentages
    main = "Win Percentage of New Zealand in the ODI World Cup", cex.main = 0.8, clockwise = TRUE)

legend("topright", legend = paste(names(win_percentages_wc_ind), sprintf("(%.1f%%)", win_percentages_wc
```



Stats of England team

```
# Filter the England data
df4 <- results[results$Team_1 == 'England' | results$Team_2 == 'England', ]

England <- df4

head(England)
```

```
##          Date    Team_1      Team_2       Winner                    Margin
## 9  12/06/2015   England New Zealand New Zealand             won by 13 runs
## 36 08/09/2015   England   Australia      England             won by 93 runs
## 56 11/11/2015 Pakistan      England     Pakistan Pakistan won by 6 wickets
## 58 13/11/2015 Pakistan      England     England     England won by 95 runs
## 59 17/11/2015 Pakistan      England     England   England won by 6 wickets
## 60 20/11/2015  Pakistan     England      England    England won by 84 runs
##                                Ground
## 9                     Kennington Oval
## 36                       Old Trafford
## 56                Sheikh Zayed Stadium
## 58                Sheikh Zayed Stadium
## 59             Sharjah Cricket Stadium
## 60 Dubai International Cricket Stadium
```

```
England_win <- England[England$Winner == 'England', ]
head(England_win)
```

```
##           Date        Team_1    Team_2  Winner                 Margin
## 36  08/09/2015       England Australia England         won by 93 runs
## 60  20/11/2015      Pakistan   England England England won by 84 runs
## 79  03/02/2016 South Africa    England England         won by 39 runs
## 81  06/02/2016 South Africa    England England       won by 5 wickets
## 135 07/10/2016   Bangladesh    England England         won by 21 runs
## 138 12/10/2016   Bangladesh    England England       won by 4 wickets
##                              Ground
## 36                         Old Trafford
## 60  Dubai International Cricket Stadium
## 79                        Mangaung Oval
## 81                      St George's Park
## 135       Shere Bangla National Stadium
## 138       Zahur Ahmed Chowdhury Stadium
```

No.of wins for England against other teams

```
excluded_value <- 'England'

filtered_df <- England_win[England_win$Team_2 != excluded_value, ]

value_counts <- table(filtered_df$Team_2)

print(value_counts)
```

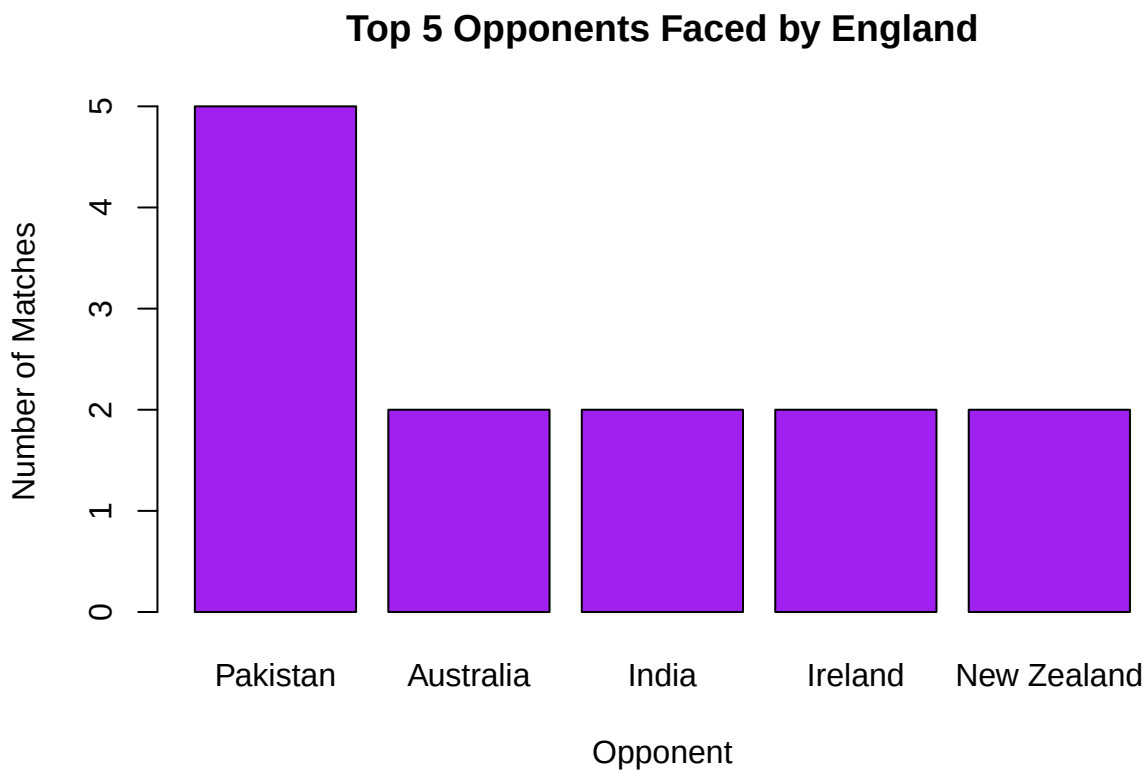
```
##
##    Australia        India      Ireland  New Zealand     Pakistan South Africa
##            2            2            2            2            5            1
##    Sri Lanka  West Indies
##            2            1
```

```
excluded_value <- 'England'

filtered_df <- England_win[England_win$Team_1 != excluded_value, ]

value_counts <- table(filtered_df$Team_1)

print(value_counts)
```

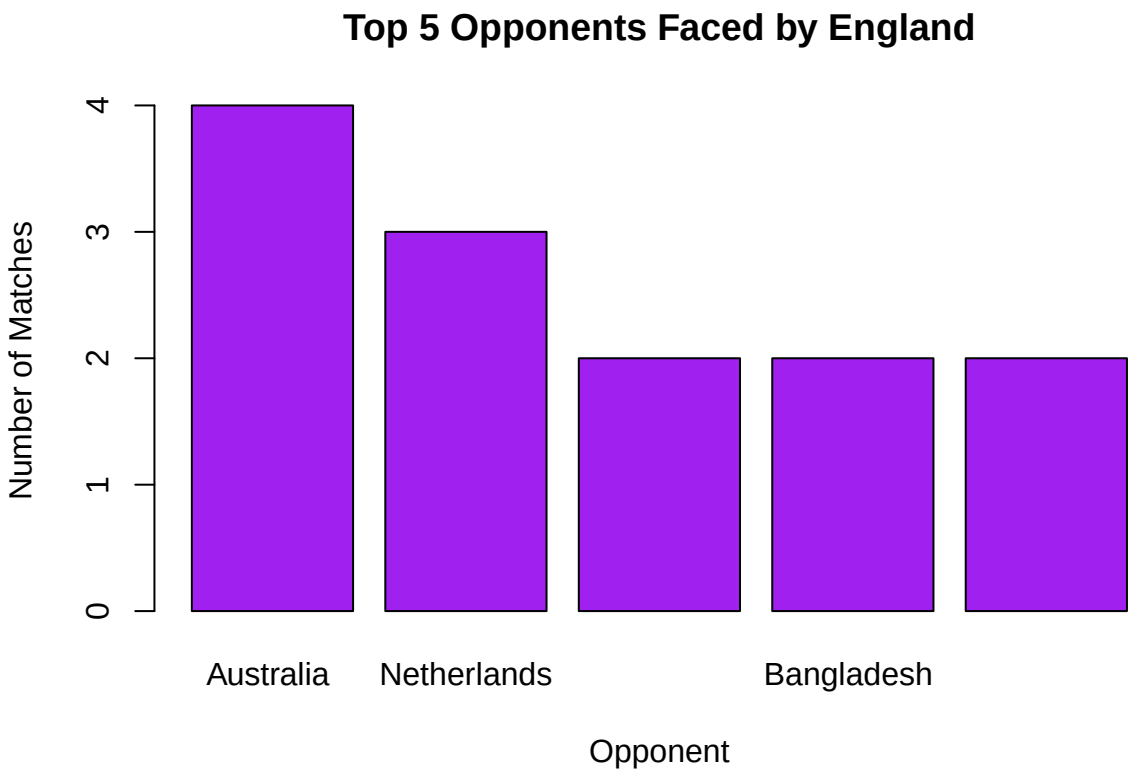
```
##
##    Australia     Bangladesh   Bangladesh          India      Ireland
##            4            2            2            1            1
##   Netherlands   New Zealand  New Zealand       Pakistan  South Africa
##            3            1            2            1            2
## South Africa      Sri Lanka    Sri Lanka   West Indies
##            2            1            2            1
```

Displaying Team England's wins against other teams using bar graph

```r
exclude <- 'England'

filtered_data <- England_win[England_win$Team_2 != exclude, ]

top_opponents <- head(names(sort(table(filtered_data$Team_2), decreasing = TRUE)), 5)
top_opponents_counts <- as.numeric(table(filtered_data$Team_2)[top_opponents])

barplot(top_opponents_counts, names.arg = top_opponents, col = "purple", main = "Top 5 Opponents Faced
```

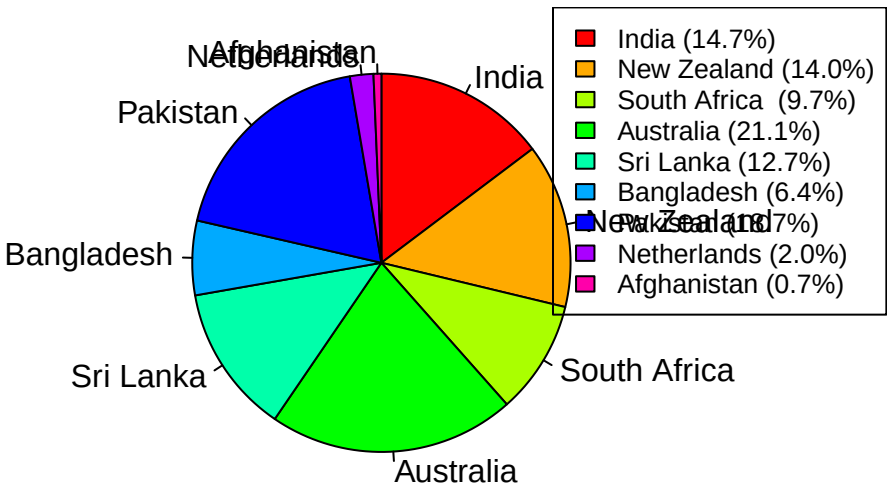



```r
exclude <- 'England'

filtered_data <- England_win[England_win$Team_1 != exclude, ]

top_opponents <- head(names(sort(table(filtered_data$Team_1), decreasing = TRUE)), 5)
top_opponents_counts <- as.numeric(table(filtered_data$Team_1)[top_opponents])

barplot(top_opponents_counts, names.arg = top_opponents, col = "purple", main = "Top 5 Opponents Faced
```

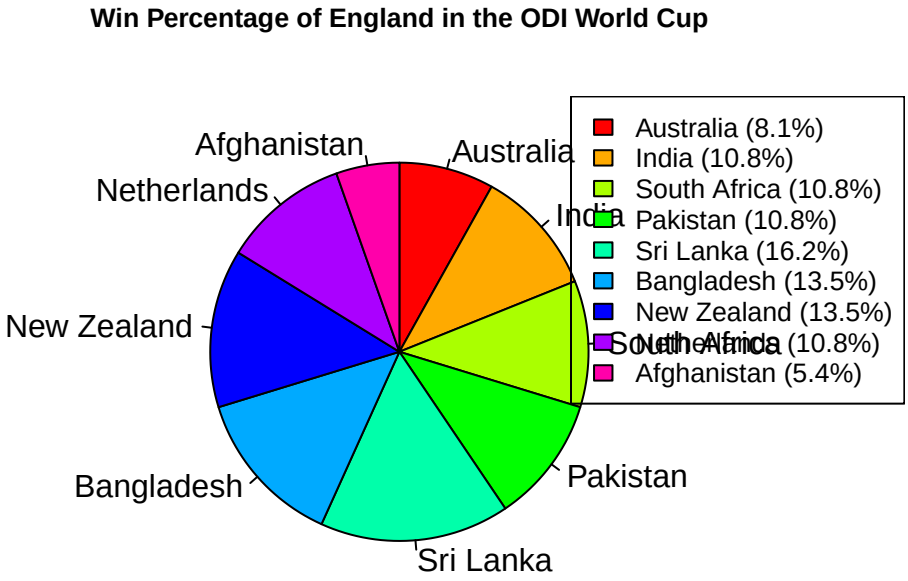


Win Percentage of England Against Each Team

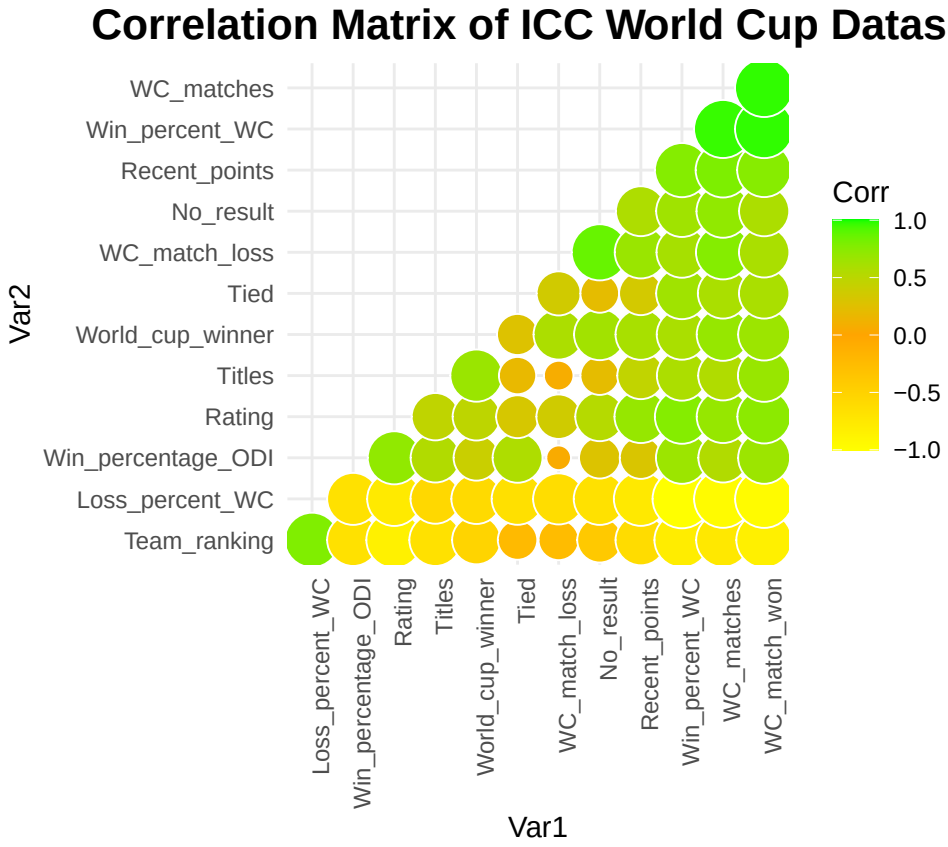
```r
# Number of wins against each team
team_win_counts <- c(
  'India'= 44,
  'New Zealand'= 42,
  'South Africa '= 29,
  'Australia'= 63,
  'Sri Lanka'= 38,
  'Bangladesh'= 19,
  'Pakistan'= 56,
  'Netherlands'= 6,
  'Afghanistan'= 2
)

total_matches <- sum(team_win_counts)

win_percentages <- (team_win_counts / total_matches) * 100

pie(win_percentages, labels = names(win_percentages), col = rainbow(length(win_percentages)),
    main = "Win Percentage of England Against Each Team", cex.main = 0.8, clockwise = TRUE)


legend("topright", legend = paste(names(win_percentages), sprintf("(%.1f%%)", win_percentages), sep = "
```




```r
# Number of wins against each team in the ODI world cup
team_win_counts_wc_ind <- c(
  'Australia'= 3,
  'India'= 4,
  'South Africa' = 4,
  'Pakistan'= 4,
  'Sri Lanka'= 6,
  'Bangladesh'= 5,
  'New Zealand'= 5,
  'Netherlands'= 4,
  'Afghanistan'= 2
)

total_matches_wc_ind <- sum(team_win_counts_wc_ind)

win_percentages_wc_ind <- (team_win_counts_wc_ind / total_matches_wc_ind) * 100

pie(win_percentages_wc_ind, labels = names(win_percentages_wc_ind), col = rainbow(length(win_percentages
    main = "Win Percentage of England in the ODI World Cup", cex.main = 0.8, clockwise = TRUE)

legend("topright", legend = paste(names(win_percentages_wc_ind), sprintf("(%.1f%%)", win_percentages_wc
```




```
modified_WorldCup_data = World_cup
modified_WorldCup_data$World_cup_winner <- ifelse(modified_WorldCup_data$World_cup_winner == "Yes", 1, 0
modified_WorldCup_data = subset(modified_WorldCup_data, select = -c(Team_name))

correlation_matrix <- cor(modified_WorldCup_data)

#corrplot(correlation_matrix, method = "color", type = "upper", addCoef.col = "white")
ggcorrplot(correlation_matrix,
           hc.order = TRUE,
           type = "lower",
           method = "circle",
           outline.color = "white",
           col = c("yellow", "orange", "green"),
           pch = c(19, 15, 17),
           tl.cex = 1.2,
           tl.col = "black") +
  labs(title = "Correlation Matrix of ICC World Cup Dataset") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold", size = 16),
        axis.text.x = element_text(angle = 90, hjust = 1))
```




```
# cricket teams participating in the world cup
worldcup_teams <- c('England', ' South Africa', '', 'West Indies',
                    'Pakistan', 'New Zealand', 'Sri Lanka', 'Afghanistan',
                    'Australia', 'Bangladesh', 'India')

# Filtering data involving only teams in the 'worldcup_teams' list
df_teams_1 <- results[results$Team_1 %in% worldcup_teams, ]
df_teams_2 <- results[results$Team_2 %in% worldcup_teams, ]

df_teams <- rbind(df_teams_1, df_teams_2)
nrow(df_teams)
```

```
## [1] 748
```

```
head(df_teams)
```

```
##          Date     Team_1      Team_2       Winner                  Margin
## 9  12/06/2015    England New Zealand New Zealand            won by 13 runs
## 12 18/06/2015 Bangladesh       India  Bangladesh            won by 79 runs
## 36 08/09/2015    England   Australia      England            won by 93 runs
## 54 07/11/2015  Sri Lanka West Indies    Sri Lanka  Sri Lanka won by 19 runs
## 57 11/11/2015 Bangladesh    Zimbabwe   Bangladesh Bangladesh won by 61 runs
## 60 20/11/2015   Pakistan     England      England    England won by 84 runs
##                                  Ground
## 9                        Kennington Oval
```

```
## 12           Shere Bangla National Stadium
## 36                             Old Trafford
## 54 Pallekele International Cricket Stadium
## 57           Shere Bangla National Stadium
## 60     Dubai International Cricket Stadium
```

```
df_teams_2019 <- subset(df_teams, select = -c(Date, Margin, Ground))

head(df_teams_2019)
```

```
##        Team_1      Team_2       Winner
## 9     England New Zealand New Zealand
## 12 Bangladesh       India  Bangladesh
## 36    England   Australia      England
## 54  Sri Lanka West Indies    Sri Lanka
## 57 Bangladesh    Zimbabwe   Bangladesh
## 60   Pakistan     England      England
```

Feature selection

```
df_teams_2019 <- df_teams_2019[ , !(names(df_teams_2019) %in% c("winning_team"))]
rownames(df_teams_2019) <- NULL

df_teams_2019$winning_team <- ifelse(df_teams_2019$Winner == df_teams_2019$Team_1, 1, NA)

df_teams_2019$winning_team[is.na(df_teams_2019$winning_team) & df_teams_2019$Winner == df_teams_2019$Tea

df_teams_2019 <- subset(df_teams_2019, select = -c(winning_team))

head(df_teams_2019)
```

```
##       Team_1      Team_2       Winner
## 1    England New Zealand New Zealand
## 2 Bangladesh       India  Bangladesh
## 3    England   Australia      England
## 4  Sri Lanka West Indies    Sri Lanka
## 5 Bangladesh    Zimbabwe   Bangladesh
## 6   Pakistan     England      England
```

Training and testing the data

```
final <- data.frame(model.matrix(~ Team_1 + Team_2 - 1, data = df_teams_2019), df_teams_2019)

X <- final[, !(names(final) %in% c("Winner"))]

y <- final$Winner

set.seed(42)

# training and testing partition
split <- sample.split(y, SplitRatio = 0.8)
```

```r
# splitting training and testing sets
X_train <- X[split, ]
X_test <- X[!split, ]
y_train <- y[split]
y_test <- y[!split]
```

```r
head(final)
```

```
##   Team_1Afghanistan Team_1Afghanistan. Team_1Australia Team_1Australia.
## 1                 0                  0               0                0
## 2                 0                  0               0                0
## 3                 0                  0               0                0
## 4                 0                  0               0                0
## 5                 0                  0               0                0
## 6                 0                  0               0                0
##   Team_1Bangladesh Team_1Bangladesh. Team_1England Team_1England.
## 1                0                 0             1              0
## 2                1                 0             0              0
## 3                0                 0             1              0
## 4                0                 0             0              0
## 5                1                 0             0              0
## 6                0                 0             0              0
##   Team_1Hong.Kong Team_1Hong.Kong. Team_1India Team_1India. Team_1India..
## 1               0                0           0            0             0
## 2               0                0           0            0             0
## 3               0                0           0            0             0
## 4               0                0           0            0             0
## 5               0                0           0            0             0
## 6               0                0           0            0             0
##   Team_1Ireland Team_1Ireland. Team_1Ireland.. Team_1Netherlands
## 1             0              0               0                 0
## 2             0              0               0                 0
## 3             0              0               0                 0
## 4             0              0               0                 0
## 5             0              0               0                 0
## 6             0              0               0                 0
##   Team_1New.Zealand Team_1New.Zealand. Team_1Oman Team_1Pakistan
## 1                 0                  0          0              0
## 2                 0                  0          0              0
## 3                 0                  0          0              0
## 4                 0                  0          0              0
## 5                 0                  0          0              0
## 6                 0                  0          0              1
##   Team_1Pakistan. Team_1Scotland Team_1Scotland. Team_1South.Africa
## 1               0              0               0                 0
## 2               0              0               0                 0
## 3               0              0               0                 0
## 4               0              0               0                 0
## 5               0              0               0                 0
## 6               0              0               0                 0
##   Team_1South.Africa. Team_1Sri.Lanka Team_1Sri.Lanka. Team_1West.Indies
## 1                   0               0                0                 0
## 2                   0               0                0                 0
```

```
## 3                   0               0                0                 0
## 4                   0               1                0                 0
## 5                   0               0                0                 0
## 6                   0               0                0                 0
##   Team_1West.Indies. Team_1Zimbabwe Team_1Zimbabwe. Team_2.England
## 1                  0              0               0              0
## 2                  0              0               0              0
## 3                  0              0               0              0
## 4                  0              0               0              0
## 5                  0              0               0              0
## 6                  0              0               0              0
##   Team_2.Pakistan Team_2.West.Indies Team_2Afghanistan Team_2Australia
## 1               0                  0                 0               0
## 2               0                  0                 0               0
## 3               0                  0                 0               1
## 4               0                  0                 0               0
## 5               0                  0                 0               0
## 6               0                  0                 0               0
##   Team_2Bangladesh Team_2England Team_2India Team_2Ireland Team_2Netherlands
## 1                0             0           0             0                 0
## 2                0             0           1             0                 0
## 3                0             0           0             0                 0
## 4                0             0           0             0                 0
## 5                0             0           0             0                 0
## 6                0             1           0             0                 0
##   Team_2New.Zealand Team_2Pakistan Team_2South.Africa Team_2Sri.Lanka
## 1                 1              0                  0               0
## 2                 0              0                  0               0
## 3                 0              0                  0               0
## 4                 0              0                  0               0
## 5                 0              0                  0               0
## 6                 0              0                  0               0
##   Team_2United.Arab.Emirates Team_2West.Indies Team_2Zimbabwe     Team_1
## 1                          0                 0              0    England
## 2                          0                 0              0 Bangladesh
## 3                          0                 0              0    England
## 4                          0                 1              0  Sri Lanka
## 5                          0                 0              1 Bangladesh
## 6                          0                 0              0   Pakistan
##        Team_2       Winner
## 1 New Zealand New Zealand
## 2       India  Bangladesh
## 3   Australia      England
## 4 West Indies    Sri Lanka
## 5    Zimbabwe   Bangladesh
## 6     England      England
```

Applying Random forest classifier algorithm

```
y_train <- as.factor(y_train)
y_test <- as.factor(y_test)

rf <- randomForest(y_train ~ ., data = cbind(X_train, y_train), ntree = 100, mtry = 20, nodesize = 1, in
```

```r
pred <- predict(rf, newdata = X_test)

levels_pred <- levels(pred)
levels_y_test <- levels(y_test)
common_levels <- union(levels_pred, levels_y_test)

pred <- factor(pred, levels = common_levels)
y_test <- factor(y_test, levels = common_levels)

score <- sum(predict(rf) == y_train) / length(y_train)

score2 <- sum(pred == y_test) / length(y_test)

cat("Training set accuracy:", sprintf("%.3f\n", score))
```

```
## Training set accuracy: 0.652
```

```r
cat("Test set accuracy:", sprintf("%.3f\n", score2))
```

```
## Test set accuracy: 0.547
```

Evaluating the best model on test data

```r
ranking <- read.csv('Icc_ranking.csv')

fixtures <- read.csv('Fixtures.csv')

pred_set <- list()
```

```r
fixtures <- fixtures %>%
  mutate(first_position = ranking$Team_ranking[match(Team_1, ranking$Team_name)])

fixtures <- fixtures %>%
  mutate(second_position = ranking$Team_ranking[match(Team_2, ranking$Team_name)])
fixtures <- head(fixtures, 45)
head(fixtures)
```

```
##   Round_number       Team_1      Team_2      Date
## 1            1      England New Zealand 5/10/2023
## 2            1     Pakistan  Netherlands 6/10/2023
## 3            1   Bangladesh Afghanistan 7/10/2023
## 4            1 South Africa   Sri Lanka 7/10/2023
## 5            1        India   Australia 8/10/2023
## 6            1  New Zealand Netherlands 9/10/2023
##                                                Location   Group Result
## 1                         Narendra Modi Stadium, Ahmedabad Group A     NA
## 2            Rajiv Gandhi International Stadium, Hyderabad Group A     NA
## 3 Himachal Pradesh Cricket Association Stadium, Dharamsala Group A     NA
## 4                              Arun Jaitley Stadium, Delhi Group A     NA
## 5                           MA Chidambaram Stadium, Chennai Group A     NA
## 6            Rajiv Gandhi International Stadium, Hyderabad Group A     NA
```

```
##   first_position second_position
## 1              5              NA
## 2             NA              10
## 3              7               8
## 4              6               9
## 5              3               1
## 6             NA              10
```

```r
fixtures$first_position <- ifelse(is.na(fixtures$first_position),
                                  mean(fixtures$first_position, na.rm = TRUE),
                                  fixtures$first_position)

fixtures$second_position <- ifelse(is.na(fixtures$second_position),
                                   mean(fixtures$second_position, na.rm = TRUE),
                                   fixtures$second_position)

head(fixtures)
```

```
##   Round_number       Team_1      Team_2      Date
## 1            1      England New Zealand 5/10/2023
## 2            1     Pakistan  Netherlands 6/10/2023
## 3            1   Bangladesh Afghanistan 7/10/2023
## 4            1 South Africa   Sri Lanka 7/10/2023
## 5            1        India   Australia 8/10/2023
## 6            1  New Zealand Netherlands 9/10/2023
##                                                 Location   Group Result
## 1                          Narendra Modi Stadium, Ahmedabad Group A     NA
## 2             Rajiv Gandhi International Stadium, Hyderabad Group A     NA
## 3 Himachal Pradesh Cricket Association Stadium, Dharamsala Group A     NA
## 4                              Arun Jaitley Stadium, Delhi Group A     NA
## 5                           MA Chidambaram Stadium, Chennai Group A     NA
## 6             Rajiv Gandhi International Stadium, Hyderabad Group A     NA
##   first_position second_position
## 1       5.000000        7.119048
## 2       4.411765       10.000000
## 3       7.000000        8.000000
## 4       6.000000        9.000000
## 5       3.000000        1.000000
## 6       4.411765       10.000000
```

```r
# Predicting winning teams
pred_set <- list()

for (i in 1:nrow(fixtures)) {
  if (fixtures$first_position[i] < fixtures$second_position[i]) {
    pred_set[[i]] <- data.frame(Team_1 = fixtures$Team_1[i],
                                Team_2 = fixtures$Team_2[i],
                                winning_team = NA)
  } else {
    pred_set[[i]] <- data.frame(Team_1 = fixtures$Team_2[i],
                                Team_2 = fixtures$Team_1[i],
                                winning_team = NA)
  }
```

```
}

pred_set <- do.call(rbind, pred_set)

backup_pred_set <- pred_set

head(pred_set)
```

```
##         Team_1      Team_2 winning_team
## 1      England New Zealand           NA
## 2    Pakistan  Netherlands           NA
## 3   Bangladesh Afghanistan           NA
## 4 South Africa   Sri Lanka           NA
## 5    Australia       India           NA
## 6  New Zealand Netherlands           NA
```

```
pred_set <- dummy_cols(pred_set, select_columns = c("Team_1", "Team_2"), remove_selected_columns = TRUE)

pred_set <- as.data.frame(pred_set)

missing_cols <- setdiff(names(final), names(pred_set))

for (c in missing_cols) {
  pred_set[[c]] <- 0
}

pred_set <- pred_set[, names(final)]

pred_set <- pred_set[, !names(pred_set) %in% 'Winner']

head(pred_set)
```

```
##   Team_1Afghanistan Team_1Afghanistan. Team_1Australia Team_1Australia.
## 1                 0                  0               0                0
## 2                 0                  0               0                0
## 3                 0                  0               0                0
## 4                 0                  0               0                0
## 5                 0                  0               0                0
## 6                 0                  0               0                0
##   Team_1Bangladesh Team_1Bangladesh. Team_1England Team_1England.
## 1                0                 0             0              0
## 2                0                 0             0              0
## 3                0                 0             0              0
## 4                0                 0             0              0
## 5                0                 0             0              0
## 6                0                 0             0              0
##   Team_1Hong.Kong Team_1Hong.Kong. Team_1India Team_1India. Team_1India..
## 1               0                0           0            0             0
## 2               0                0           0            0             0
## 3               0                0           0            0             0
## 4               0                0           0            0             0
## 5               0                0           0            0             0
```

```
## 6               0               0          0           0            0
##   Team_1Ireland Team_1Ireland. Team_1Ireland.. Team_1Netherlands
## 1             0              0              0                 0
## 2             0              0              0                 0
## 3             0              0              0                 0
## 4             0              0              0                 0
## 5             0              0              0                 0
## 6             0              0              0                 0
##   Team_1New.Zealand Team_1New.Zealand. Team_1Oman Team_1Pakistan
## 1                 0                  0          0              0
## 2                 0                  0          0              0
## 3                 0                  0          0              0
## 4                 0                  0          0              0
## 5                 0                  0          0              0
## 6                 0                  0          0              0
##   Team_1Pakistan. Team_1Scotland Team_1Scotland. Team_1South.Africa
## 1               0              0               0                  0
## 2               0              0               0                  0
## 3               0              0               0                  0
## 4               0              0               0                  0
## 5               0              0               0                  0
## 6               0              0               0                  0
##   Team_1South.Africa. Team_1Sri.Lanka Team_1Sri.Lanka. Team_1West.Indies
## 1                   0              0                0                 0
## 2                   0              0                0                 0
## 3                   0              0                0                 0
## 4                   0              0                0                 0
## 5                   0              0                0                 0
## 6                   0              0                0                 0
##   Team_1West.Indies. Team_1Zimbabwe Team_1Zimbabwe. Team_2.England
## 1                  0              0               0              0
## 2                  0              0               0              0
## 3                  0              0               0              0
## 4                  0              0               0              0
## 5                  0              0               0              0
## 6                  0              0               0              0
##   Team_2.Pakistan Team_2.West.Indies Team_2Afghanistan Team_2Australia
## 1               0                  0                 0               0
## 2               0                  0                 0               0
## 3               0                  0                 0               0
## 4               0                  0                 0               0
## 5               0                  0                 0               0
## 6               0                  0                 0               0
##   Team_2Bangladesh Team_2England Team_2India Team_2Ireland Team_2Netherlands
## 1                0             0          0             0                 0
## 2                0             0          0             0                 0
## 3                0             0          0             0                 0
## 4                0             0          0             0                 0
## 5                0             0          0             0                 0
## 6                0             0          0             0                 0
##   Team_2New.Zealand Team_2Pakistan Team_2South.Africa Team_2Sri.Lanka
## 1                 0              0                  0              0
## 2                 0              0                  0              0
## 3                 0              0                  0              0
```

```
## 4                  0             0                  0              0
## 5                  0             0                  0              0
## 6                  0             0                  0              0
##   Team_2United.Arab.Emirates Team_2West.Indies Team_2Zimbabwe Team_1 Team_2
## 1                          0                 0              0      0      0
## 2                          0                 0              0      0      0
## 3                          0                 0              0      0      0
## 4                          0                 0              0      0      0
## 5                          0                 0              0      0      0
## 6                          0                 0              0      0      0
```

Interpret the model results

```
# Predictions
predictions <- predict(rf, newdata = pred_set, type = "response")

for (i in 1:nrow(fixtures)) {
  cat(backup_pred_set[i, "Team_2"], " vs ", backup_pred_set[i, "Team_1"], "\n")

  if (predictions[i] == 1) {
    cat("Winner: ", backup_pred_set[i, "Team_2"], "\n")
  } else {
    cat("Winner: ", backup_pred_set[i, "Team_1"], "\n")
  }

  cat("\n")
}
```

```
## New Zealand  vs  England
## Winner:  England
##
## Netherlands  vs  Pakistan
## Winner:  Pakistan
##
## Afghanistan  vs  Bangladesh
## Winner:  Bangladesh
##
## Sri Lanka  vs  South Africa
## Winner:  South Africa
##
## India  vs  Australia
## Winner:  Australia
##
## Netherlands  vs  New Zealand
## Winner:  New Zealand
##
## Bangladesh  vs  England
## Winner:  England
##
## Afghanistan  vs  India
## Winner:  India
##
## Sri Lanka  vs  Pakistan
## Winner:  Pakistan
```

```
##
## South Africa  vs  Australia
## Winner:  Australia
##
## Bangladesh  vs  New Zealand
## Winner:  New Zealand
##
## Afghanistan  vs  England
## Winner:  England
##
## India  vs  Pakistan
## Winner:  Pakistan
##
## Sri Lanka  vs  Australia
## Winner:  Australia
##
## Netherlands  vs  South Africa
## Winner:  South Africa
##
## Afghanistan  vs  New Zealand
## Winner:  New Zealand
##
## Bangladesh  vs  India
## Winner:  India
##
## Pakistan  vs  Australia
## Winner:  Australia
##
## Netherlands  vs  Sri Lanka
## Winner:  Sri Lanka
##
## South Africa  vs  England
## Winner:  England
##
## New Zealand  vs  India
## Winner:  India
##
## Afghanistan  vs  Pakistan
## Winner:  Pakistan
##
## Bangladesh  vs  South Africa
## Winner:  South Africa
##
## Netherlands  vs  Australia
## Winner:  Australia
##
## Sri Lanka  vs  England
## Winner:  England
##
## South Africa  vs  Pakistan
## Winner:  Pakistan
##
## New Zealand  vs  Australia
## Winner:  Australia
```

```
##
## Netherlands  vs  Bangladesh
## Winner:  Bangladesh
##
## England  vs  India
## Winner:  India
##
## Sri Lanka  vs  Afghanistan
## Winner:  Afghanistan
##
## Bangladesh  vs  Pakistan
## Winner:  Pakistan
##
## South Africa  vs  New Zealand
## Winner:  New Zealand
##
## Sri Lanka  vs  India
## Winner:  India
##
## Netherlands  vs  Afghanistan
## Winner:  Afghanistan
##
## New Zealand  vs  Pakistan
## Winner:  Pakistan
##
## England  vs  Australia
## Winner:  Australia
##
## South Africa  vs  India
## Winner:  India
##
## Sri Lanka  vs  Bangladesh
## Winner:  Bangladesh
##
## Afghanistan  vs  Australia
## Winner:  Australia
##
## Netherlands  vs  England
## Winner:  England
##
## Sri Lanka  vs  New Zealand
## Winner:  New Zealand
##
## Afghanistan  vs  South Africa
## Winner:  South Africa
##
## Netherlands  vs  India
## Winner:  India
##
## Bangladesh  vs  Australia
## Winner:  Australia
##
## England  vs  Pakistan
## Winner:  Pakistan
```

```r
semi <- list(list('Pakistan', 'Australia'), list('England', 'India'))

print(semi)
```

```
## [[1]]
## [[1]][[1]]
## [1] "Pakistan"
##
## [[1]][[2]]
## [1] "Australia"
##
##
## [[2]]
## [[2]][[1]]
## [1] "England"
##
## [[2]][[2]]
## [1] "India"
```

```r
#function to predict Winner
clean_and_predict <- function(matches, ranking, final, rf) {
  positions <- c()

  for (match in matches) {
    positions <- c(positions, ranking$Team_ranking[ranking$Team_name == match[1]])
    positions <- c(positions, ranking$Team_ranking[ranking$Team_name == match[2]])
  }

  pred_set <- list()

  i <- 1
  j <- 1

  while (i <= length(positions)) {
    dict1 <- list(
      Team_1 = c(),
      Team_2 = c()
    )

    if (positions[i] < positions[i + 1]) {
      dict1$Team_1 <- c(dict1$Team_1, matches[[j]][[1]])
      dict1$Team_2 <- c(dict1$Team_2, matches[[j]][[2]])
    } else {
      dict1$Team_1 <- c(dict1$Team_1, matches[[j]][[2]])
      dict1$Team_2 <- c(dict1$Team_2, matches[[j]][[1]])
    }
    pred_set <- c(pred_set, list(dict1))
    i <- i + 2
    j <- j + 1
  }

  pred_set_df <- do.call(rbind, lapply(pred_set, data.frame, stringsAsFactors = FALSE))
```

```
  backup_pred_set <- pred_set_df

  pred_set_df <- dummy_cols(pred_set_df, select_columns = c("Team_1", "Team_2"), remove_selected_columns

  missing_cols2 <- setdiff(names(final), names(pred_set_df))

  for (c in missing_cols2) {
    pred_set_df[[c]] <- 0
  }

  pred_set_df <- pred_set_df %>%
    dplyr::select(names(final))

  pred_set_df <- pred_set_df[, !(names(pred_set_df) %in% c("Winner"))]

  predictions <- predict(rf, newdata = pred_set_df)

  for (i in 1:nrow(pred_set_df)) {
    cat(paste(backup_pred_set[i, "Team_2"], " vs ", backup_pred_set[i, "Team_1"], "\n"))

    if (predictions[i] == 1) {
      cat("Winner: ", backup_pred_set[i, "Team_2"], "\n")
    } else {
      cat("Winner: ", backup_pred_set[i, "Team_1"], "\n")
    }

    cat("\n")
  }

}
```

```
clean_and_predict(semi, ranking, final, rf)
```

```
## Pakistan  vs  Australia
## Winner:  Australia
##
## England  vs  India
## Winner:  India
```

```
finals = list(list('Australia', 'India'))
```

```
clean_and_predict(finals, ranking, final, rf)
```

```
## India  vs  Australia
## Winner:  Australia
```

```
# Confusion Matrix Creation
Y_predicted <- factor(pred)
y_test <- factor(y_test)

common_levels <- union(levels(Y_predicted), levels(y_test))
```

```r
Y_predicted <- factor(Y_predicted, levels = common_levels)
y_test <- factor(y_test, levels = common_levels)

conf_matrix <- confusionMatrix(Y_predicted, y_test)
```

```r
heatmap(conf_matrix$table,
        Colv = NA, Rowv = NA,
        margins=c(10,10),
        col = colorRampPalette(c("white", "steelblue"))(100),
        main = "Confusion Matrix for Random Forest",
        xlab = "Predicted Values",
        ylab = "Test Values")
```