# BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

## MBA - Business Analytics

## Group 8

## REGRESSION ANALYSIS OF BOWLERS IN IPL AND WPL

## Contents

**Course Name :** Statistics and Basic Econometrics

**Course Code :** MPBA G505

**Group 8 -**

2023H1540813P - G Ranjana Jeysree

2023H1540837P - Eldho John Menachery

2023H1540841P - Riya Gupta

2023H1540862P - Karthik Thilakan E

2023H1540869P - Sanjana V

2023H1540870P - Shantanu Kumar Panda

2023H1540821P - Yash Goyal

## Introduction :

Cricket is one of the most popular games in our country. India is blessed to have some of the best cricket players. The Indian Premier League (IPL) is a professional Twenty20 cricket league in India. The BCCI founded the league in 2007. Women's Premier League is the newest addition in the Premier league.

In cricket, a bowler's performance holds great influence over the result of a match. The quantity of wickets a bowler manages to capture acts as a measure of their impact. Nevertheless, the bowler's contribution goes beyond merely taking wickets. Through this research, we aim to determine the factors that define a bowler's performance in a game, enhancing our understanding.

## Problem Statement

- Data on at least 50 cricket players is to be collected and a regression analysis is to be done of their performance in the recent Indian Premier League (IPL 2023) for men and Women's Premier League (WPL 2023) for women.
- There must be at least five independent variable and the regression of the dependent variable is to be done on all the independent variables.
- The data for at least fifty cricket players must be collected for analysis.
- All the analysis must be done in the R environment and submitted as .qmd (Quarto) + docx files, the data set (.csv format)

## Need for study (Motivation) :

Multiple Linear regression is applied in various areas of business and academic study. It is a reliable statistical procedure to predict the future. In this project, we examine the correlation between a bowler's wicket count and various factors like age, economy rate, bowling style, number of overs bowled, and other relevant metrics. This study seeks to offer a valuable understanding of the importance of a bowler's function and impact within the sporting framework. The insights gained from this analysis may assist teams in making educated choices regarding player recruitment, strategy development, and enhancements in overall team performance.

## Objectives :

- To conduct multiple linear regression analysis on bowlers' performance in IPL 2023.
- To ascertain the impact of variables like age, economy rate, bowling style, and number of overs bowled (independent variables) on the bowlers' performance, measured by the total number of wickets taken (dependent variable).

- To look into different transformations and interactions between different independent variables and come up with the best possible model.
- To gain valuable insights that can be extracted from the identified correlations, offering a deeper understanding of the relationships and their implications.

## Proposed methodology :

- Determine the dependent and independent variable(s) to be used in this analysis.
- Dependent variable:
  - Total number of wickets taken by each bowler
- Independent variables:
  - Number of matches played
  - Number of Innings
  - Player economy rate
  - Age
  - Instagram followers
  - Twitter followers
  - IPL Auction Price
  - Number of dot balls
  - Gender
  - Right/Left Arm
  - Spin/Pace
- Regress the dependent variable on all the independent variables
- Derive necessary insights from the established correlation that can be further used for predicting future results

## Project Timeline :

The anticipated time frame for the project tasks is as follows:

| Task | Due Date |
|---|---|
| Project Proposal Submission | 31st August 2023 |
| Finalize Data Collection | 2nd October 2023 |
| Mid Term Review | 4th October 2023 |
| Final Submission | 27th November 2023 |
| Final Presentation | 29th November 2023 |

## Bowler Data Set :

Our data set consists of data relating to 108 bowlers who participated in the Domestic Cricket League (IPL/WPL) season 2023.

### Dependent Variable :

#### Wickets

- This variable refers to the total number of wickets taken by each bowler in IPL/WPL season 2023.
- A wicket is when the ball hits one of the two sets of three stumps and two bails at either end of the pitch to get the batter out.

### Independent Variables :

#### Gender

- This variable includes two values – M and F.
- M stands for Male, and F stands for Female.
- The data set consists of 57 male bowlers and 51 female bowlers.

#### Matches

- Matches refers to the total number of matches the bowler has participated in IPL/WPL season 2023.
- A cricket match typically involves two teams, each with batting and bowling innings. Therefore, a player may participate in a match as a batsman, a bowler, a fielder, or any combination of these roles. The number of matches gives an overall count of how many times a player has taken the field, whether they batted, bowled, fielded, or did a combination of these activities.

### Innings

- Innings refers to the total number of times each bowler has bowled in the IPL/WPL season 2023.
- It only considers the number of matches a player has come to bowl during the season. In this statistic, batters and fielders do not have innings counted against their names. For example, if a player bowls in every inning of a match, their number of innings will be equal to the number of matches they have played. However, if they do not get a chance to bowl in some innings, there will be a difference between the number of innings and matches.

### Economy_Rate

- Economy_Rate refers to the average number of runs a bowler conceded in every over he/she bowls, that is, how economical a bowler has been in the season.
- The lesser the Economy rate, the more economical the bowler has been in giving runs.

### Age

- Age refers to the age of the bowlers in years.
- The range for the age of the bowlers in this data set is 15 to 40 years.
- We have considered the age of the bowlers as of 29th May 2023 (Final match of IPL 2023).

### Insta_Followers

- Insta_Followers refers to each bowler's total number of followers on Instagram.
- The followers are critical to the success of an instagram account since they indicate its popularity and influence within the platform.
- We have considered the followers data as of 29th September 2023.
- The number of followers is a dynamic value; thus, we have considered the approximate value for each player. Since the number of followers is a huge number and the daily changes are minor, we assume the change would be marginal and thus would not affect the final outcome.

### Twitter_Followers

- Twitter_Followers refers to each bowler's total number of followers on Twitter (X).
- The followers are critical to the success of a twitter account since they indicate its popularity and influence within the platform.
- We have considered the followers data as of 29th September 2023.
- The number of followers is a dynamic value; thus, we have considered the approximate value for each player. Since the number of followers is a huge number

and the daily changes are minor, we assume the change would be marginal and thus would not affect the final outcome.

### Auction_Price

- Auction_Price refers to the highest bid the bowler gets by teams in rupees.
- If the bowler has a good track record, more teams would want to include them, and thus, the bidding value increases.
- The data collected is from the 2023 auction that took place for all the sold players, and for the retained players we have mentioned the last auctioned price.
- We have considered the latest auction prices for each bowler. Some of the bowlers were retained by the teams and thus we considered their auction prices from previous years.

### Dot_Balls

- Dot_Balls refers to the total number of dot balls bowled by a bowler throughout the IPL/WPL season 2023.
- A delivery is considered a dot ball when no runs are conceded off it. While facing a ball, when a batting side fails to score runs off the bat or by any other means, that particular delivery is called a dot ball.

### Right_Left_Arm

- This variable indicates whether a player is a right-arm or left-arm bowler.
- The data set consists of 79 right-arm bowlers and 29 left-arm bowlers.

### Spin_Pace

- This variable indicates whether a player is a spinner or a pacer.
- There are two broad categories of bowlers in cricket - pace and spin. Pace bowlers rely primarily on the speed of the ball to dismiss batsmen, whereas spin bowlers rely on the rotation and turn of the ball.
- The data set consists of 48 spinners and 60 pacers.

## Steps Followed :

- For the purpose of this project, the entire regression analysis was performed in the R environment using multiple libraries and the data-set collected for the purpose of this project.
- The first step was to import the data into RStudio.
- Then we performed analysis to check whether a linear relationship exists between all the independent variables with the dependent variable.

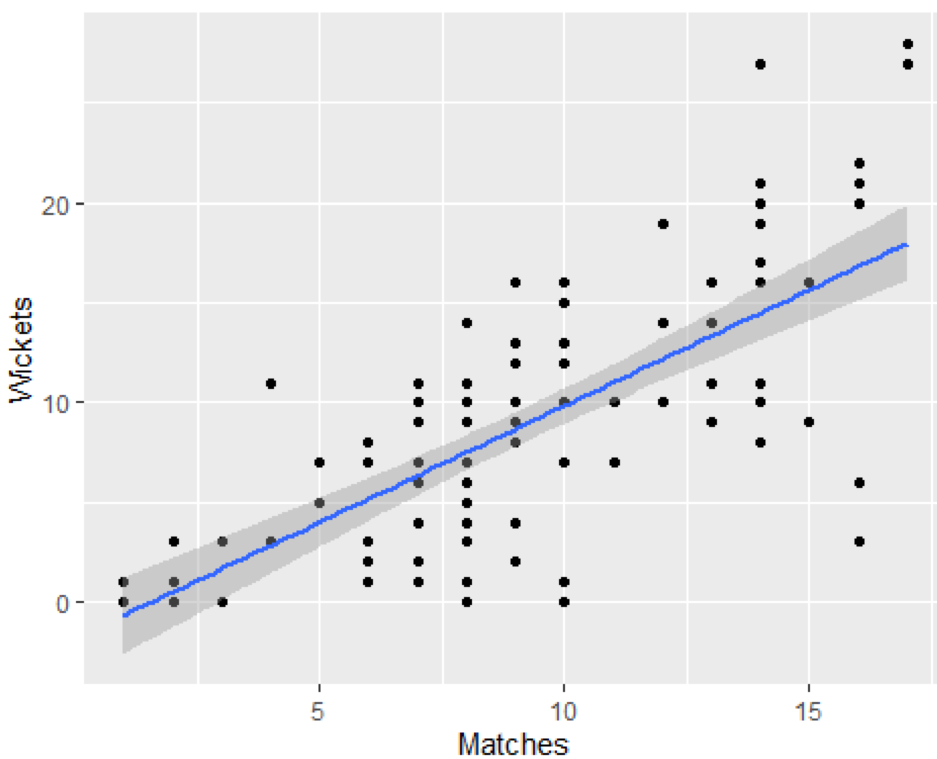
- After checking for linear relationships, we will check for the other assumptions/conditions of linear regression namely homoscedasticity, normality, independence of residuals, no outliers, no multi-collinearity, etc.
- Then we form the correlation matrix between multiple variables in a data-set. It provides a comprehensive summary of the pairwise correlations between variables, revealing the direction and strength of these relationships. We thus figure out which variables are to be removed.
- From that, we start our model-making process. After making models involving removal, addition, and interaction of various independent variables, we came up with the best model which describes the linear relationship in the best possible manner through the multiple R squared value and adjusted R squared value.
- We will conduct residual analysis and ANOVA test to find the RSS values for all the models made. Whichever has the lowest RSS value, will be the model we go forward with.
- Lastly, we will derive insights from the chosen model and use it to make predictions as per the need of the user.

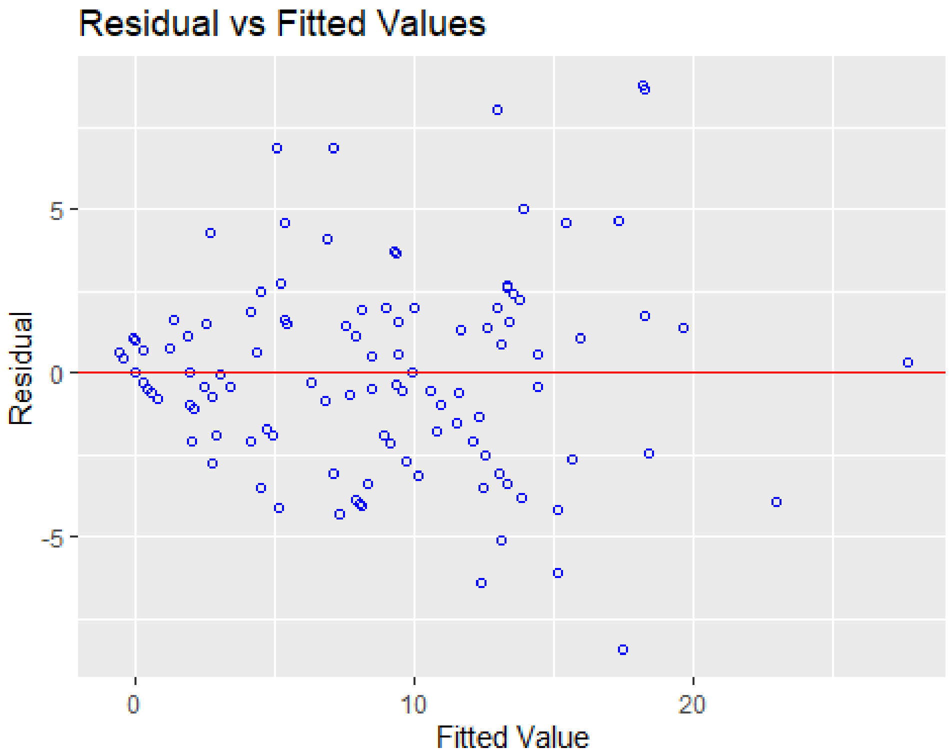
## R Code :

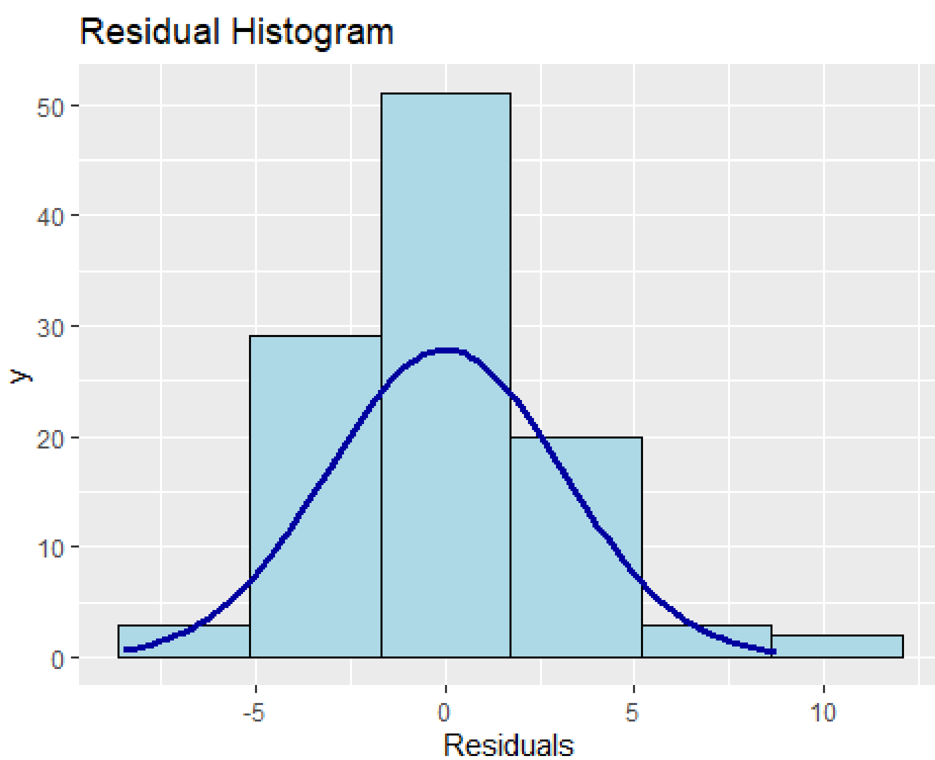
```
library(tidyverse)

Warning: package 'tidyverse' was built under R version 4.3.2

Warning: package 'readr' was built under R version 4.3.2

Warning: package 'dplyr' was built under R version 4.3.2

library(UsingR)

Warning: package 'UsingR' was built under R version 4.3.2

Warning: package 'HistData' was built under R version 4.3.2

Warning: package 'Hmisc' was built under R version 4.3.2

library(readr)
library(jtools)

Warning: package 'jtools' was built under R version 4.3.2

library(moments)
library(lmtest)

Warning: package 'lmtest' was built under R version 4.3.2

Warning: package 'zoo' was built under R version 4.3.2

library(olsrr)

Warning: package 'olsrr' was built under R version 4.3.2

library(plyr)

Warning: package 'plyr' was built under R version 4.3.2

library(dplyr)
library(caret)

Warning: package 'caret' was built under R version 4.3.2

library(reshape2)

Warning: package 'reshape2' was built under R version 4.3.2

library(corrplot)

Warning: package 'corrplot' was built under R version 4.3.2

library(coefplot)

Warning: package 'coefplot' was built under R version 4.3.2

library(ggplot2)
```

```
Data<- read_csv("D:/R_dataset.csv")

Rows: 108 Columns: 14
── Column specification
────────────────────────────────────────────────────────
Delimiter: ","
chr (4): Name, Gender, Right_Left_Arm, Spin_Pace
dbl (7): SNo, Wickets, Matches, Innings, Economy_Rate, Age, Dot_Balls
num (3): Insta_Followers, Twitter_Followers, Auction_Price

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.

view(Data)

predictor_vars <-
c("Matches","Innings","Age","Insta_Followers","Twitter_Followers","Auction_Pr
ice","Dot_Balls","Economy_Rate")
response_var <- "Wickets"

#Checking linearity between all DVs and IV and validating the linear
regression assumptions
for (var in predictor_vars) {

  # simple linear regression
  lm_model <- lm(reformulate(predictor_vars, response_var), data = Data)

  # display summary
  print(summary(lm_model))

  #plotting regression line
  print(
    ggplot(Data, aes_string(x = var, y = response_var)) + geom_point() +
      geom_smooth(method = "lm", se = TRUE) + labs(y = response_var, x = var)
  )

  # checking for Homoscedasticity
  ols_plot_resid_fit(lm_model)

  # checking for normality of residuals
  ols_plot_resid_hist(lm_model)
}


Call:
lm(formula = reformulate(predictor_vars, response_var), data = Data)

Residuals:
    Min      1Q  Median      3Q     Max
```

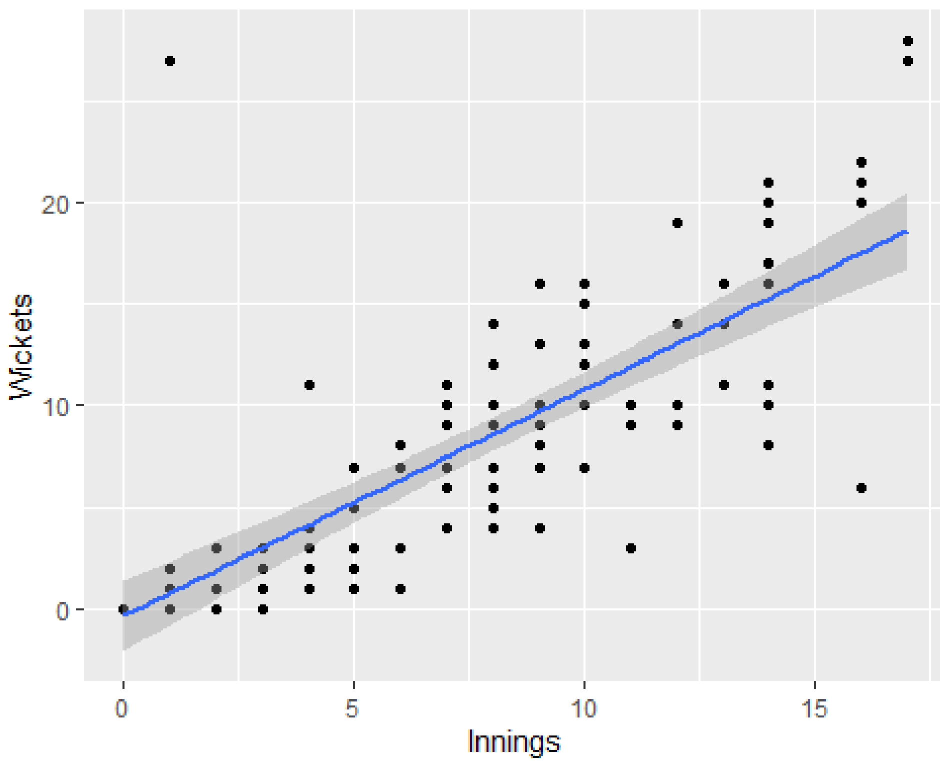
```
-8.4567 -2.0740 -0.1778  1.5762  8.7860

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)       -3.245e+00  2.715e+00  -1.195   0.2349    
Matches            4.284e-01  1.753e-01   2.444   0.0163 *  
Innings           -4.826e-01  2.100e-01  -2.298   0.0236 *  
Age                5.464e-02  7.023e-02   0.778   0.4384    
Insta_Followers   -1.930e-07  1.697e-07  -1.137   0.2582    
Twitter_Followers  2.004e-07  3.285e-07   0.610   0.5431    
Auction_Price      4.214e-09  9.148e-09   0.461   0.6460    
Dot_Balls          1.550e-01  1.779e-02   8.715 6.91e-14 ***
Economy_Rate       1.011e-01  1.863e-01   0.543   0.5886    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.215 on 99 degrees of freedom
Multiple R-squared:  0.773, Adjusted R-squared:  0.7546 
F-statistic: 42.14 on 8 and 99 DF,  p-value: < 2.2e-16

Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
ℹ Please use tidy evaluation idioms with `aes()`.
ℹ See also `vignette("ggplot2-in-packages")` for more information.

`geom_smooth()` using formula = 'y ~ x'
```



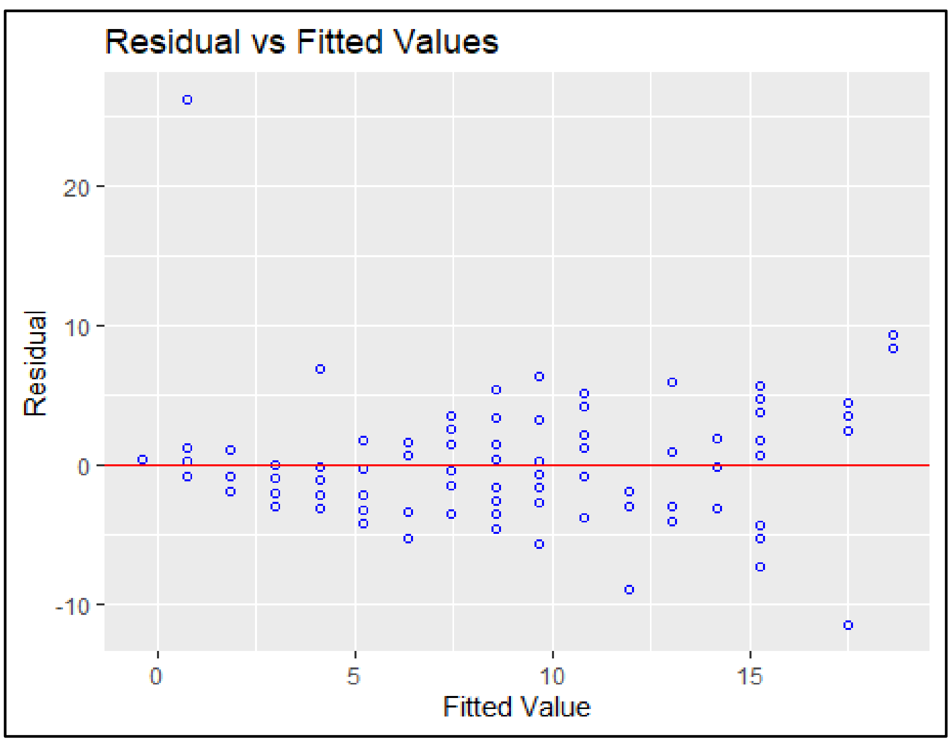

Residual vs Fitted Values
Residual
Fitted Value
5
0
-5
0
10
20


Residual Histogram
y
Residuals
50
40
30
20
10
0
-5
0
5
10

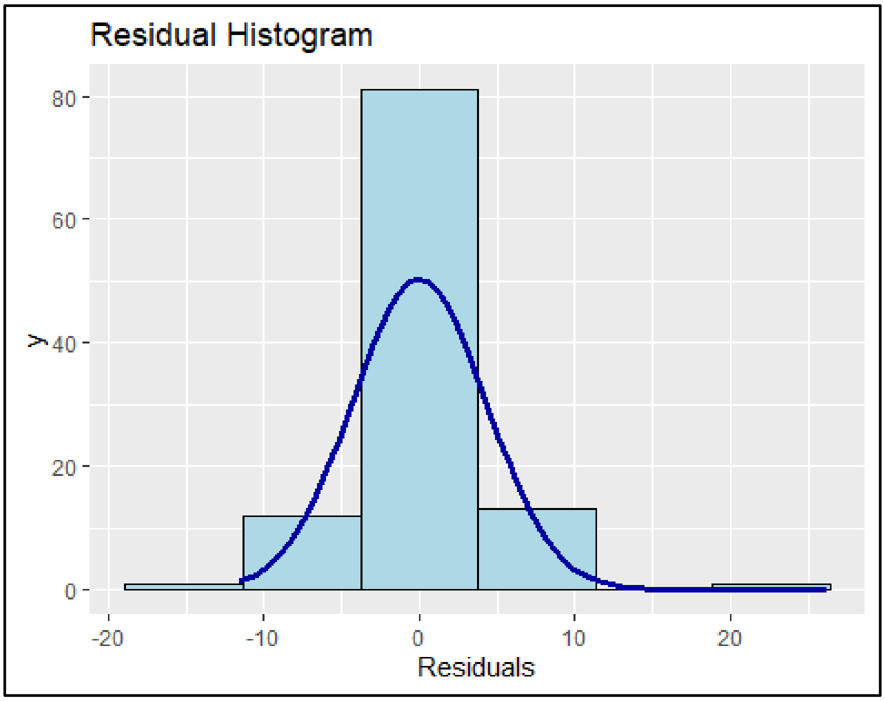
```
Call:
lm(formula = reformulate(predictor_vars, response_var), data = Data)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.4567 -2.0740 -0.1778  1.5762  8.7860 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)       -3.245e+00  2.715e+00  -1.195   0.2349    
Matches            4.284e-01  1.753e-01   2.444   0.0163 *  
Innings           -4.826e-01  2.100e-01  -2.298   0.0236 *  
Age                5.464e-02  7.023e-02   0.778   0.4384    
Insta_Followers   -1.930e-07  1.697e-07  -1.137   0.2582    
Twitter_Followers  2.004e-07  3.285e-07   0.610   0.5431    
Auction_Price      4.214e-09  9.148e-09   0.461   0.6460    
Dot_Balls          1.550e-01  1.779e-02   8.715 6.91e-14 ***
Economy_Rate       1.011e-01  1.863e-01   0.543   0.5886    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.215 on 99 degrees of freedom
Multiple R-squared:  0.773,	Adjusted R-squared:  0.7546 
F-statistic: 42.14 on 8 and 99 DF,  p-value: < 2.2e-16

`geom_smooth()` using formula = 'y ~ x'
```



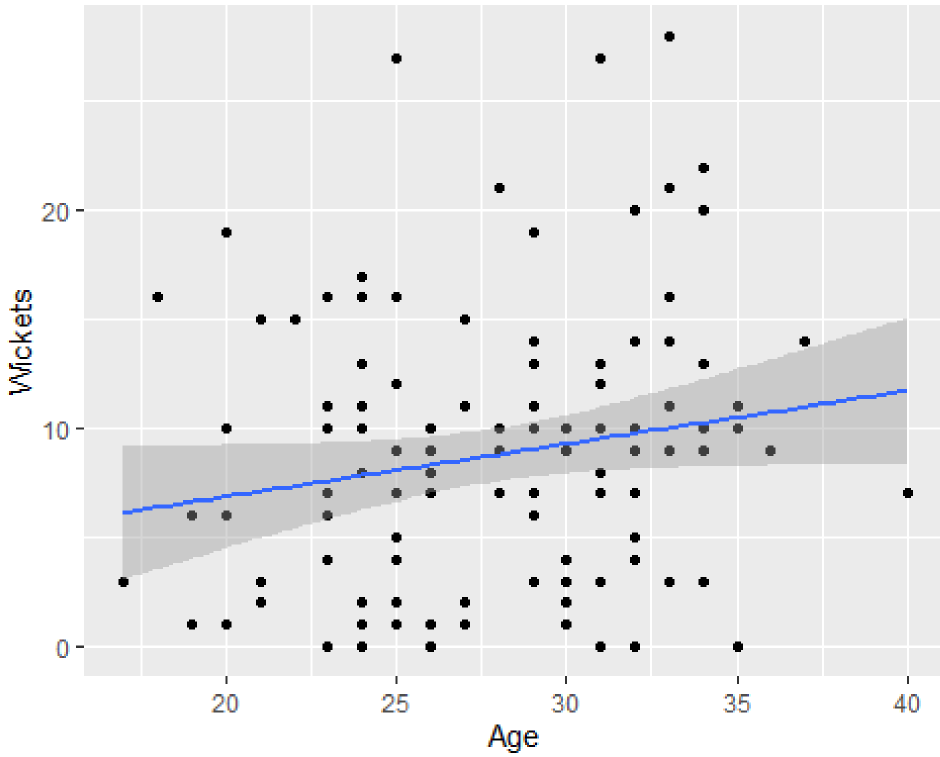
Wickets
20
10
0
0
5
10
15
Innings

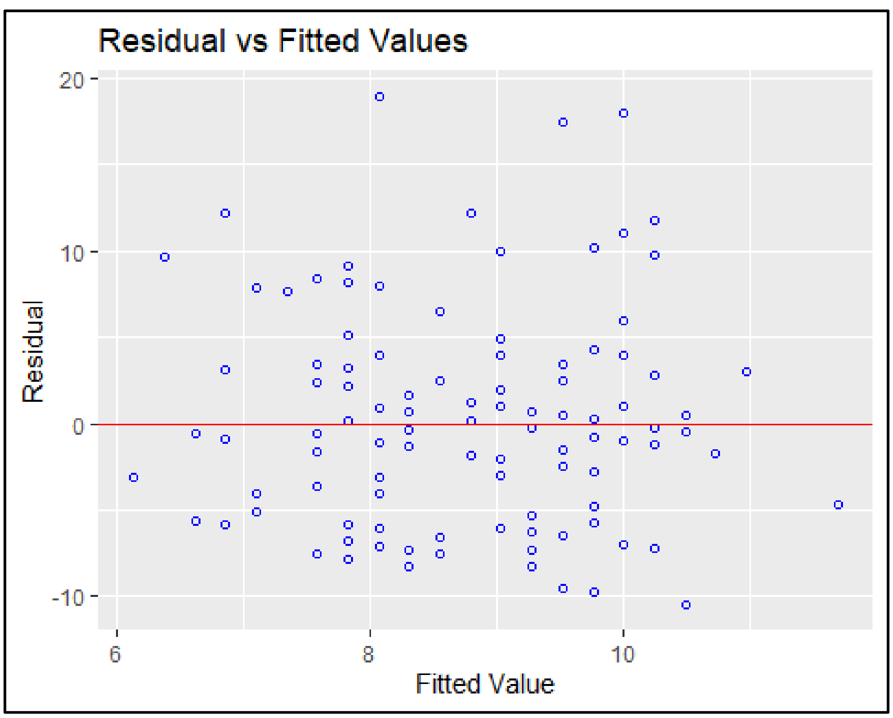



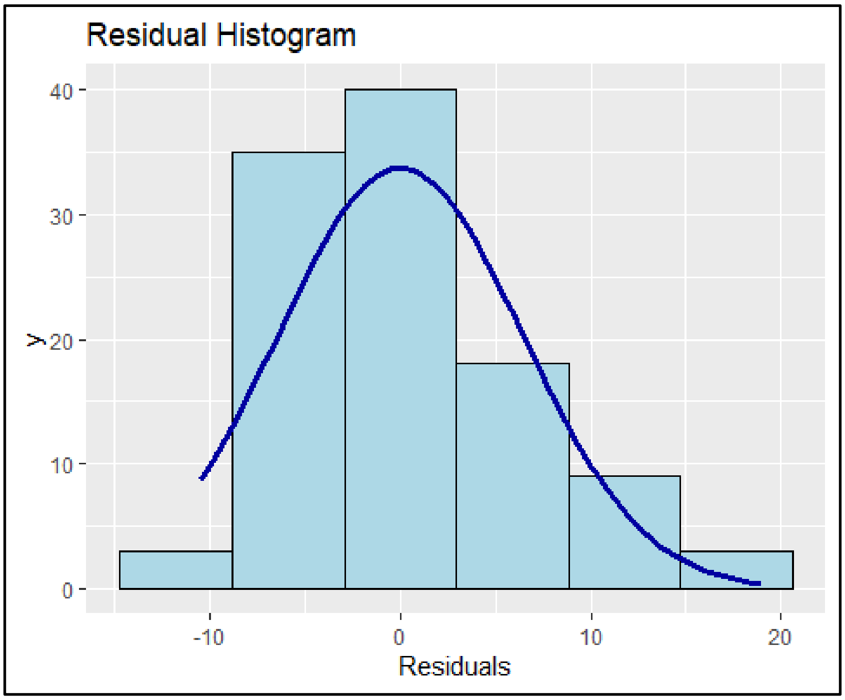
Residual vs Fitted Values
Residual
20
10
0
-10
0
5
10
15
Fitted Value



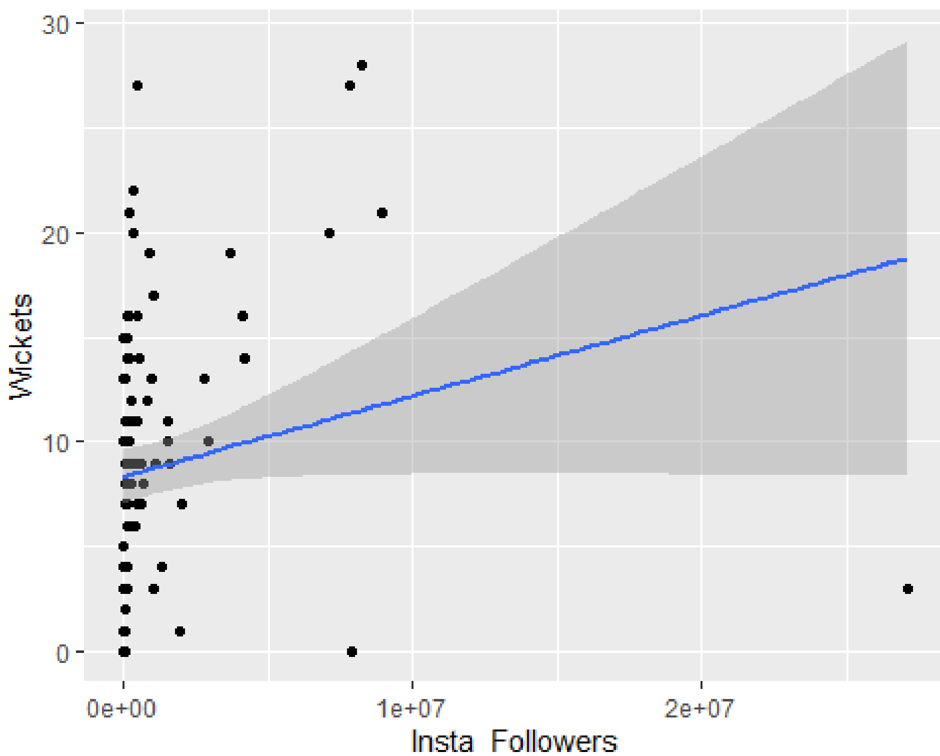

```
Call:
lm(formula = reformulate(predictor_vars, response_var), data = Data)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.4567 -2.0740 -0.1778  1.5762  8.7860 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)      -3.245e+00  2.715e+00  -1.195   0.2349    
Matches           4.284e-01  1.753e-01   2.444   0.0163 *  
Innings          -4.826e-01  2.100e-01  -2.298   0.0236 *  
Age               5.464e-02  7.023e-02   0.778   0.4384    
Insta_Followers  -1.930e-07  1.697e-07  -1.137   0.2582    
Twitter_Followers  2.004e-07  3.285e-07   0.610   0.5431    
Auction_Price     4.214e-09  9.148e-09   0.461   0.6460    
Dot_Balls         1.550e-01  1.779e-02   8.715 6.91e-14 ***
Economy_Rate      1.011e-01  1.863e-01   0.543   0.5886    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.215 on 99 degrees of freedom
Multiple R-squared:  0.773, Adjusted R-squared:  0.7546 
F-statistic: 42.14 on 8 and 99 DF,  p-value: < 2.2e-16

`geom_smooth()` using formula = 'y ~ x'
```

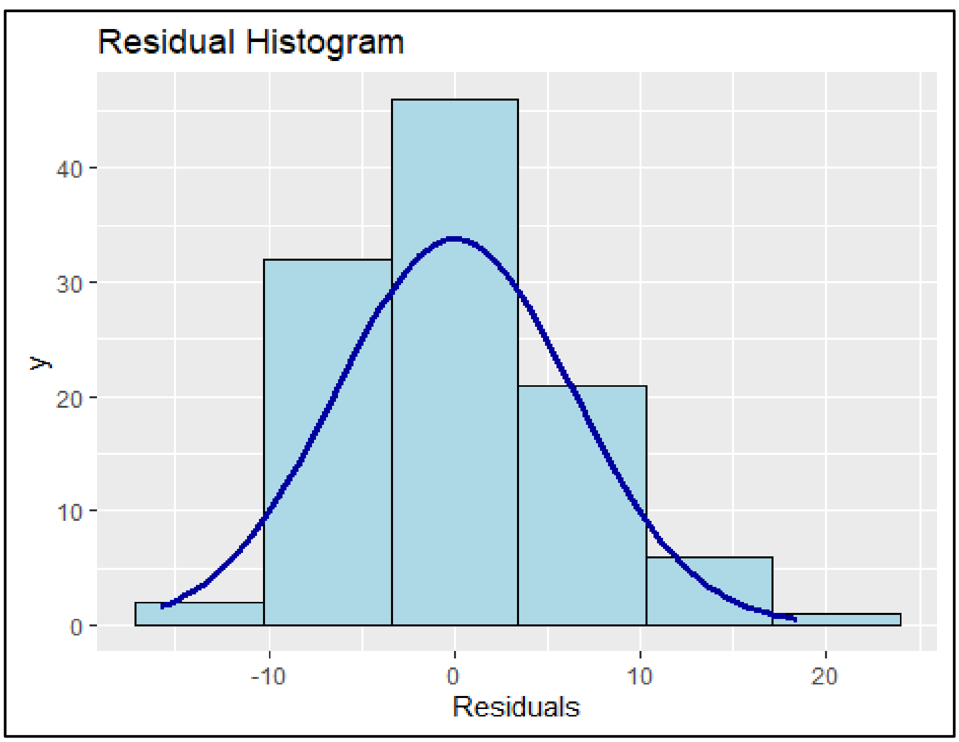

Wickets
Age
0
10
20
20
25
30
35
40


Residual vs Fitted Values
Residual
Fitted Value
-10
0
10
20
6
8
10

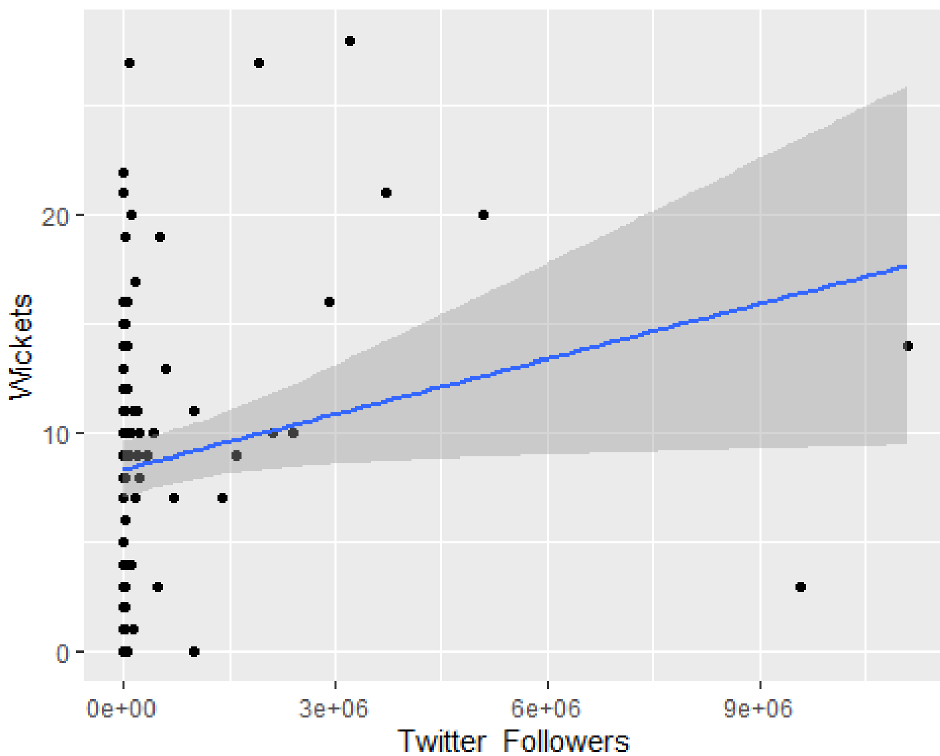



```
Call:
lm(formula = reformulate(predictor_vars, response_var), data = Data)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.4567 -2.0740 -0.1778  1.5762  8.7860 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)       -3.245e+00  2.715e+00  -1.195   0.2349    
Matches            4.284e-01  1.753e-01   2.444   0.0163 *  
Innings           -4.826e-01  2.100e-01  -2.298   0.0236 *  
Age                5.464e-02  7.023e-02   0.778   0.4384    
Insta_Followers   -1.930e-07  1.697e-07  -1.137   0.2582    
Twitter_Followers  2.004e-07  3.285e-07   0.610   0.5431    
Auction_Price      4.214e-09  9.148e-09   0.461   0.6460    
Dot_Balls          1.550e-01  1.779e-02   8.715 6.91e-14 ***
Economy_Rate       1.011e-01  1.863e-01   0.543   0.5886    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.215 on 99 degrees of freedom
Multiple R-squared:  0.773,	Adjusted R-squared:  0.7546 
F-statistic: 42.14 on 8 and 99 DF,  p-value: < 2.2e-16

`geom_smooth()` using formula = 'y ~ x'
```

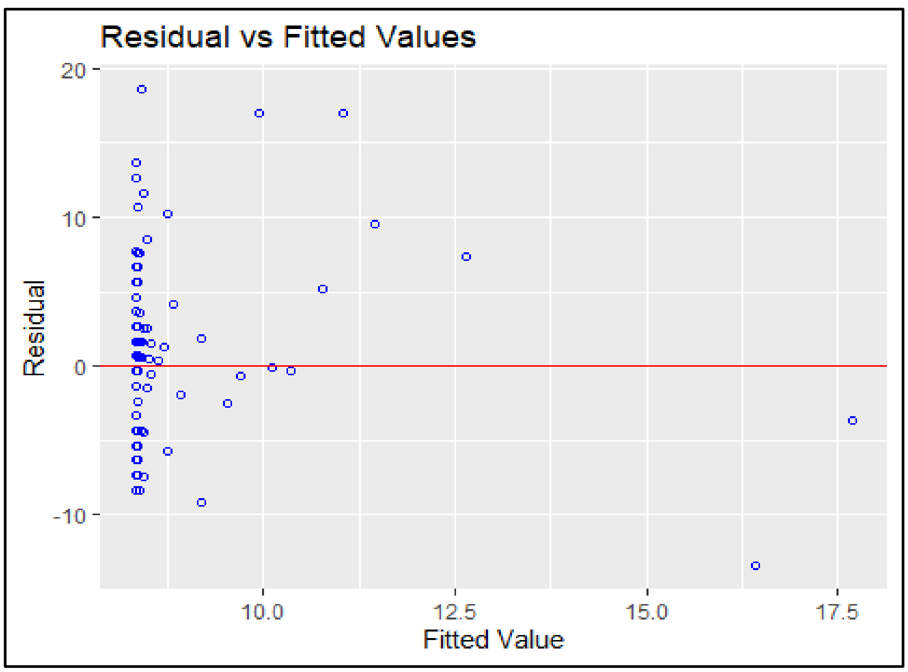

30
20
10
0
Wickets
0e+00
1e+07
2e+07
Insta_Followers


Residual Histogram
40
30
20
10
0
y
-10
0
10
20
Residuals

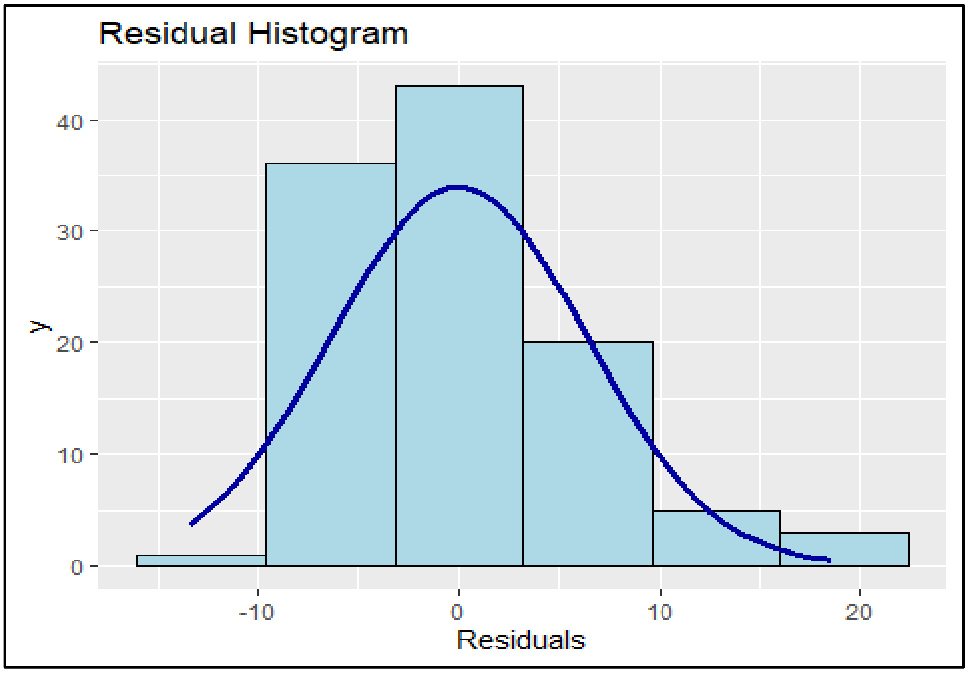
```
Call:
lm(formula = reformulate(predictor_vars, response_var), data = Data)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.4567 -2.0740 -0.1778  1.5762  8.7860 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)       -3.245e+00  2.715e+00  -1.195   0.2349    
Matches            4.284e-01  1.753e-01   2.444   0.0163 *  
Innings           -4.826e-01  2.100e-01  -2.298   0.0236 *  
Age                5.464e-02  7.023e-02   0.778   0.4384    
Insta_Followers   -1.930e-07  1.697e-07  -1.137   0.2582    
Twitter_Followers  2.004e-07  3.285e-07   0.610   0.5431    
Auction_Price      4.214e-09  9.148e-09   0.461   0.6460    
Dot_Balls          1.550e-01  1.779e-02   8.715 6.91e-14 ***
Economy_Rate       1.011e-01  1.863e-01   0.543   0.5886    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.215 on 99 degrees of freedom
Multiple R-squared:  0.773, Adjusted R-squared:  0.7546 
F-statistic: 42.14 on 8 and 99 DF,  p-value: < 2.2e-16

`geom_smooth()` using formula = 'y ~ x'
```



20
10
0
Wickets
0e+00
3e+06
6e+06
9e+06
Twitter_Followers

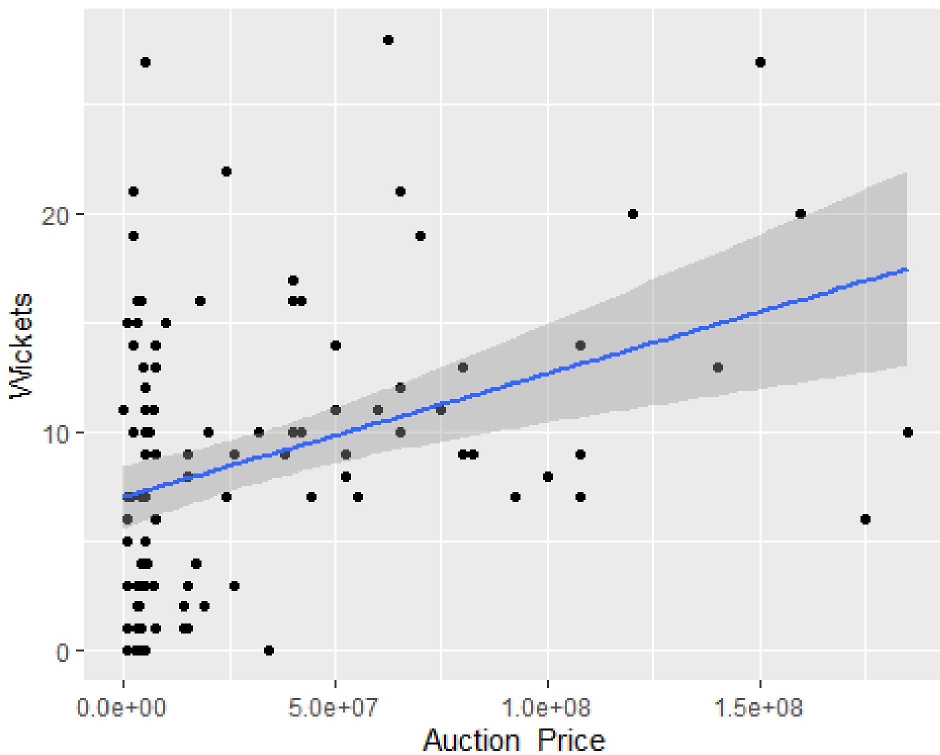



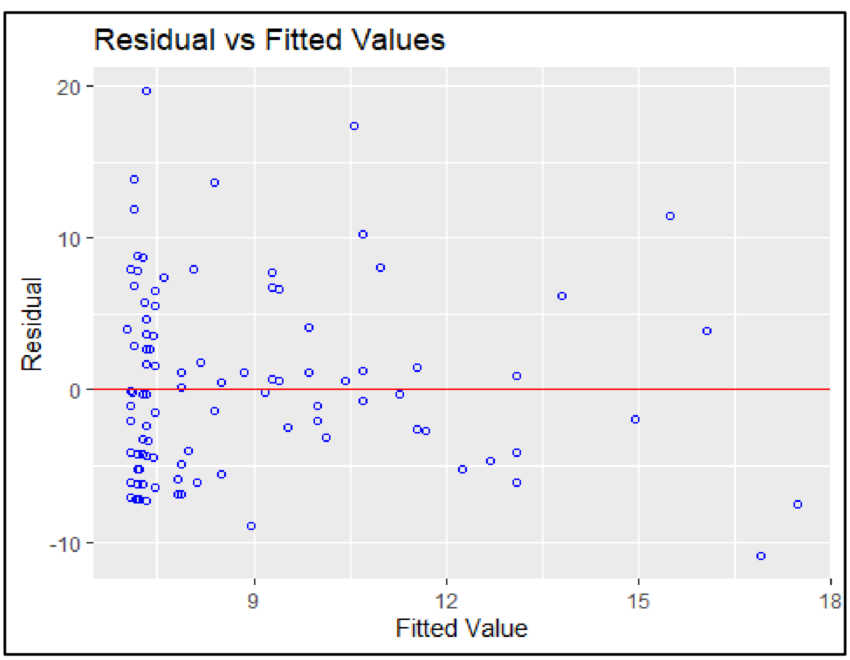
Residual vs Fitted Values
20
10
0
-10
Residual
10.0
12.5
15.0
17.5
Fitted Value

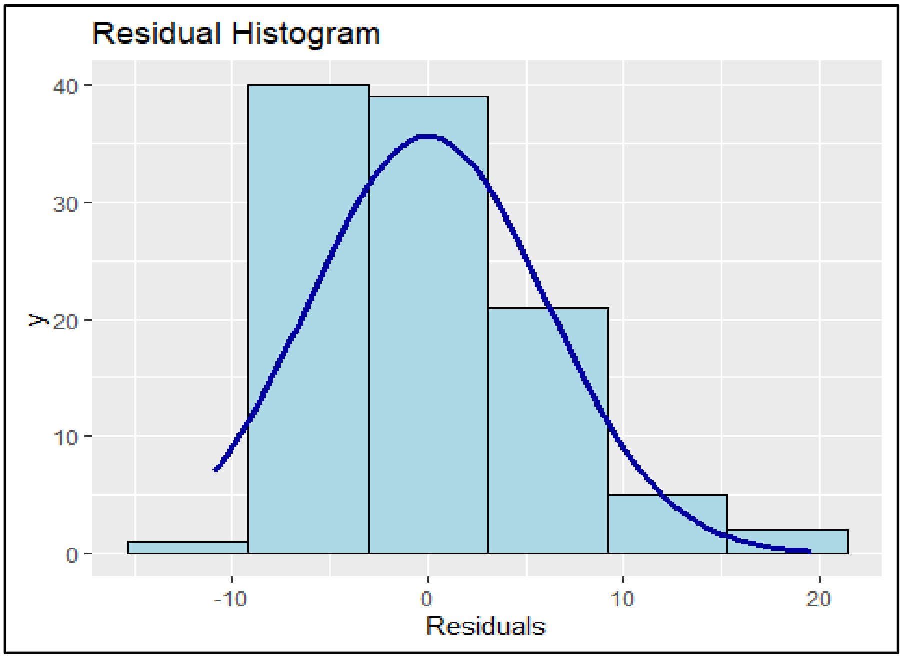



```
Call:
lm(formula = reformulate(predictor_vars, response_var), data = Data)

Residuals:
    Min      1Q  Median      3Q     Max
-8.4567 -2.0740 -0.1778  1.5762  8.7860

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       -3.245e+00  2.715e+00  -1.195   0.2349
Matches            4.284e-01  1.753e-01   2.444   0.0163 *
Innings           -4.826e-01  2.100e-01  -2.298   0.0236 *
Age                5.464e-02  7.023e-02   0.778   0.4384
Insta_Followers   -1.930e-07  1.697e-07  -1.137   0.2582
Twitter_Followers  2.004e-07  3.285e-07   0.610   0.5431
Auction_Price      4.214e-09  9.148e-09   0.461   0.6460
Dot_Balls          1.550e-01  1.779e-02   8.715 6.91e-14 ***
Economy_Rate       1.011e-01  1.863e-01   0.543   0.5886
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

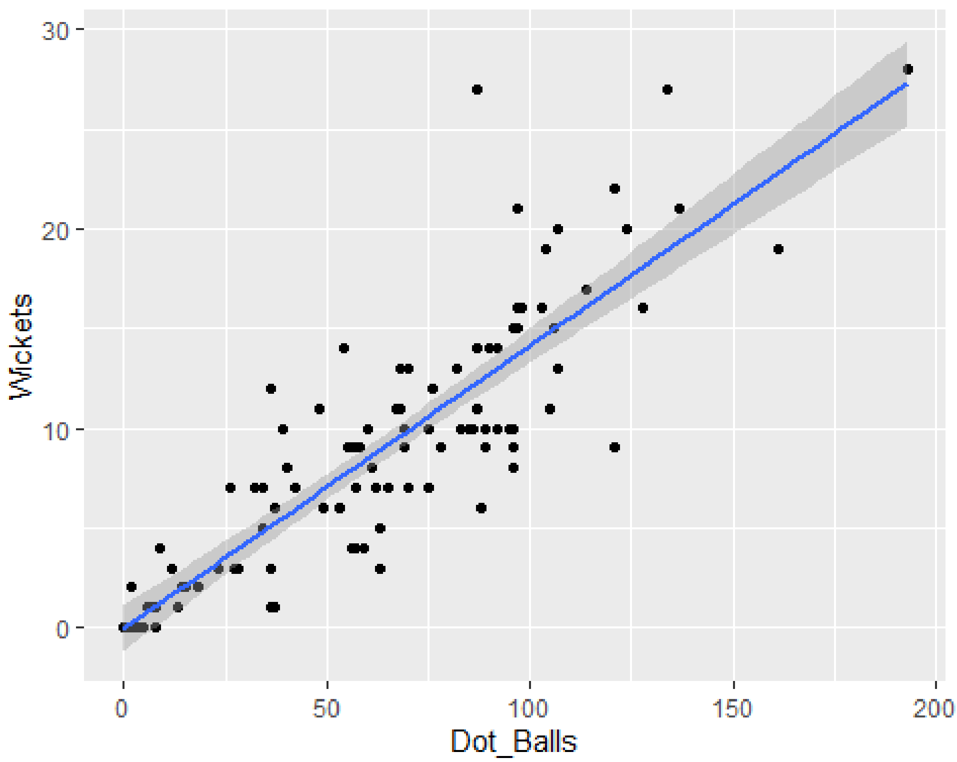
```
Residual standard error: 3.215 on 99 degrees of freedom
Multiple R-squared:  0.773, Adjusted R-squared:  0.7546
F-statistic: 42.14 on 8 and 99 DF,  p-value: < 2.2e-16

`geom_smooth()` using formula = 'y ~ x'
```



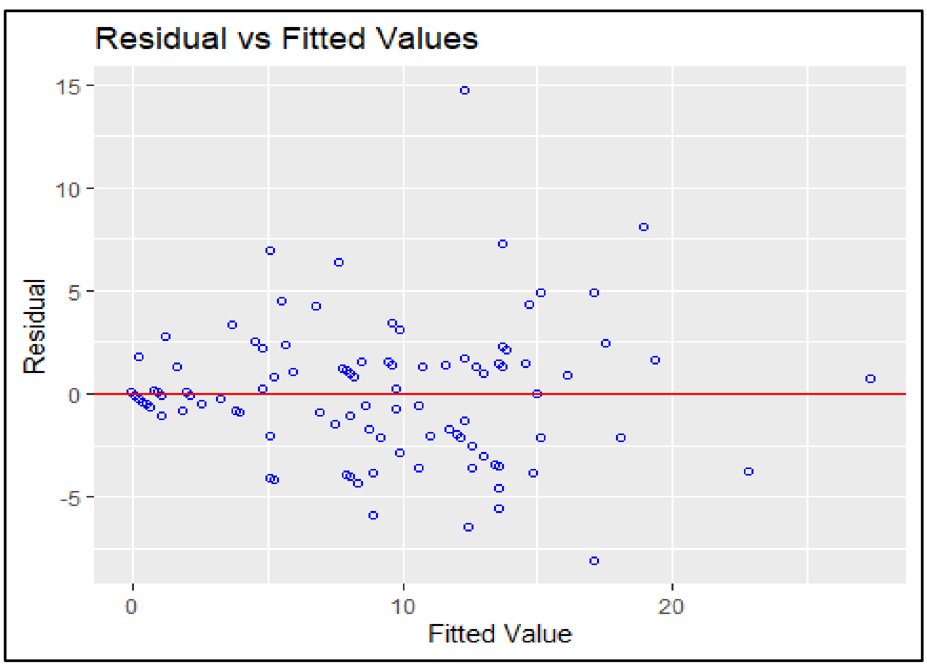





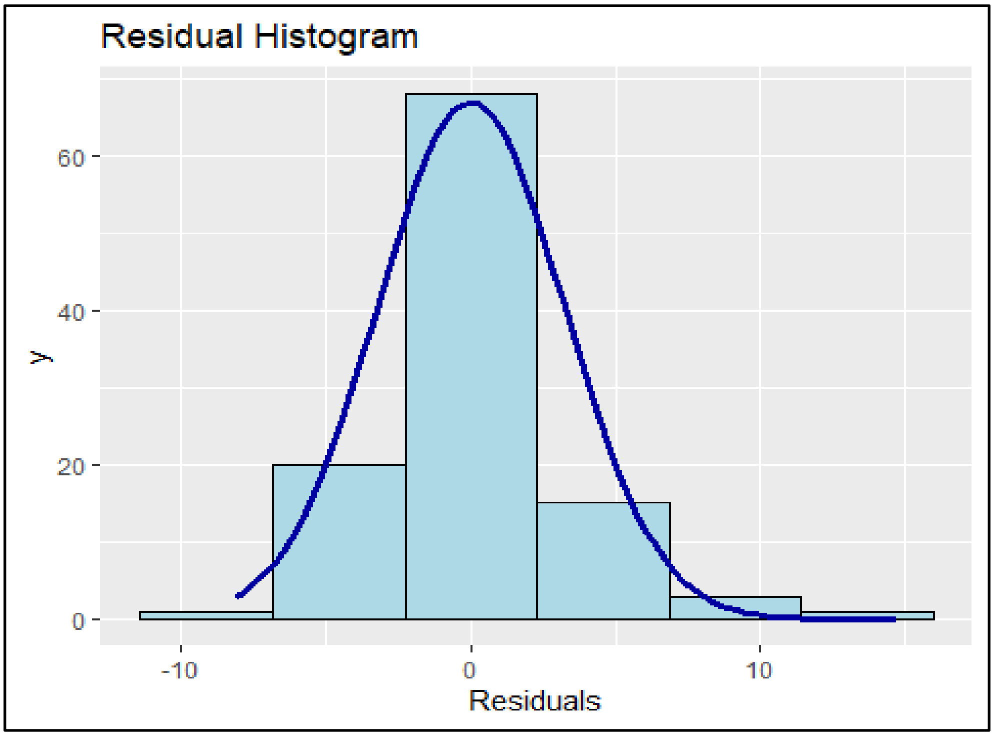

```
Call:
lm(formula = reformulate(predictor_vars, response_var), data = Data)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.4567 -2.0740 -0.1778  1.5762  8.7860 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)       -3.245e+00  2.715e+00  -1.195   0.2349    
Matches            4.284e-01  1.753e-01   2.444   0.0163 *  
Innings           -4.826e-01  2.100e-01  -2.298   0.0236 *  
Age                5.464e-02  7.023e-02   0.778   0.4384    
Insta_Followers   -1.930e-07  1.697e-07  -1.137   0.2582    
Twitter_Followers  2.004e-07  3.285e-07   0.610   0.5431    
Auction_Price      4.214e-09  9.148e-09   0.461   0.6460    
Dot_Balls          1.550e-01  1.779e-02   8.715 6.91e-14 ***
Economy_Rate       1.011e-01  1.863e-01   0.543   0.5886    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.215 on 99 degrees of freedom
Multiple R-squared:  0.773,	Adjusted R-squared:  0.7546 
F-statistic: 42.14 on 8 and 99 DF,  p-value: < 2.2e-16

`geom_smooth()` using formula = 'y ~ x'
```



Wickets
Dot_Balls
30
20
10
0
0
50
100
150
200

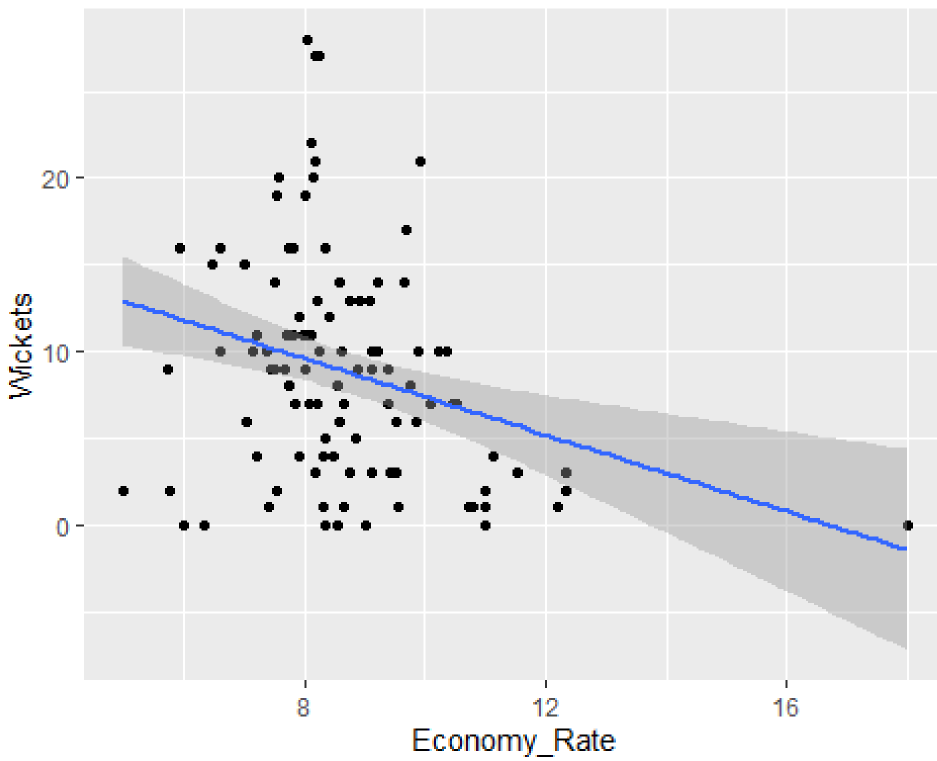



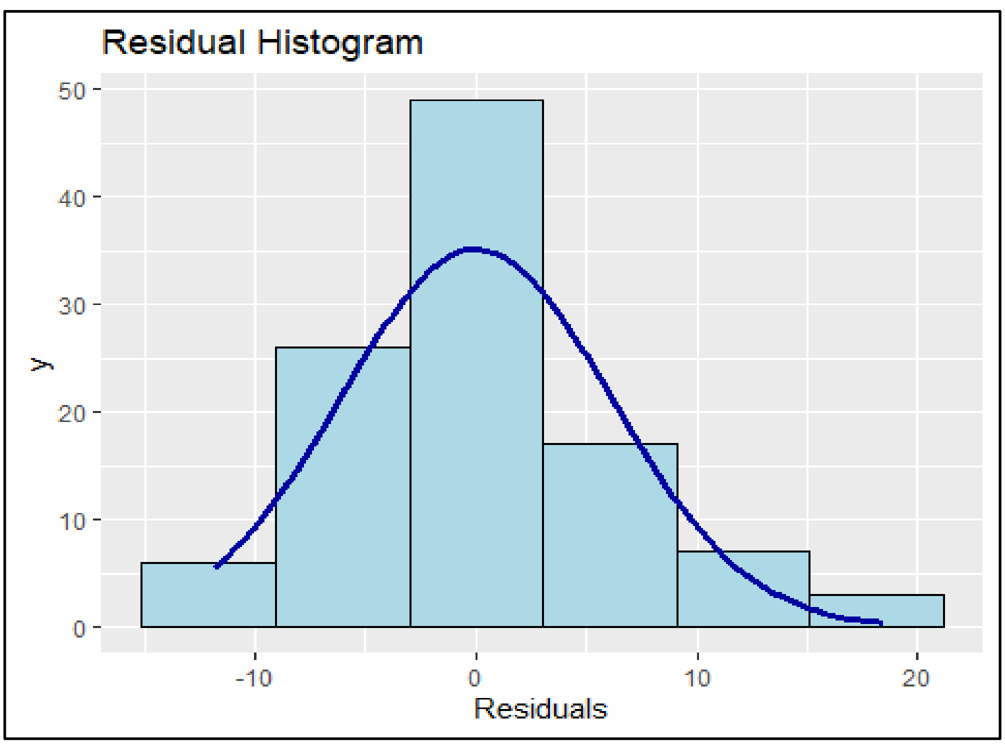
Residual vs Fitted Values
Residual
Fitted Value
15
10
5
0
-5
0
10
20




```
Call:
lm(formula = reformulate(predictor_vars, response_var), data = Data)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.4567 -2.0740 -0.1778  1.5762  8.7860 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)       -3.245e+00  2.715e+00  -1.195   0.2349    
Matches            4.284e-01  1.753e-01   2.444   0.0163 *  
Innings           -4.826e-01  2.100e-01  -2.298   0.0236 *  
Age                5.464e-02  7.023e-02   0.778   0.4384    
Insta_Followers   -1.930e-07  1.697e-07  -1.137   0.2582    
Twitter_Followers  2.004e-07  3.285e-07   0.610   0.5431    
Auction_Price      4.214e-09  9.148e-09   0.461   0.6460    
Dot_Balls          1.550e-01  1.779e-02   8.715 6.91e-14 ***
Economy_Rate       1.011e-01  1.863e-01   0.543   0.5886    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.215 on 99 degrees of freedom
Multiple R-squared:  0.773, Adjusted R-squared:  0.7546 
F-statistic: 42.14 on 8 and 99 DF,  p-value: < 2.2e-16
```

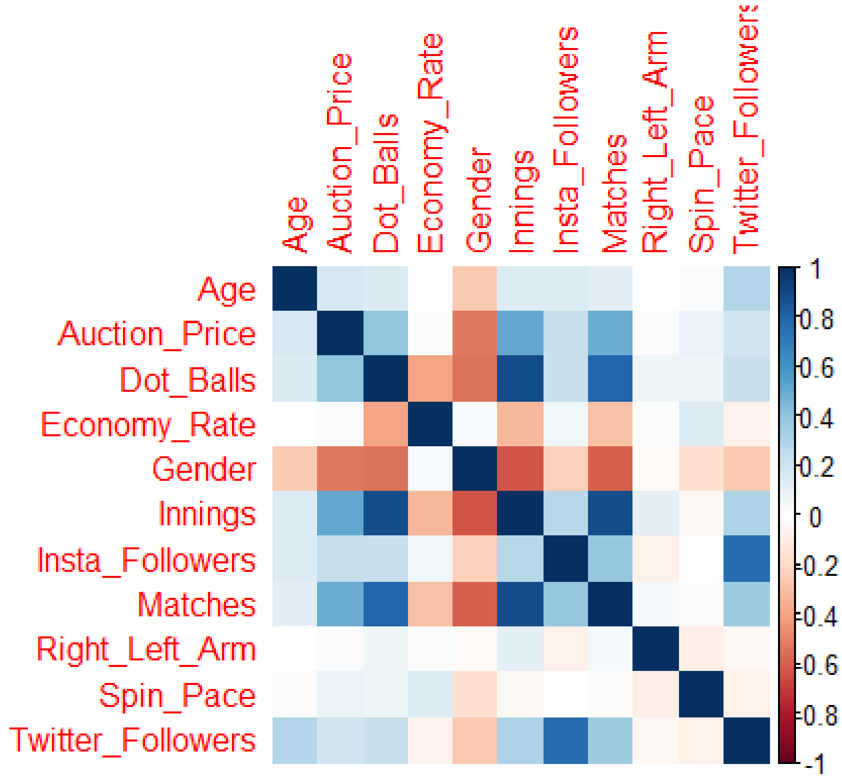
```
`geom_smooth()` using formula = 'y ~ x'
```



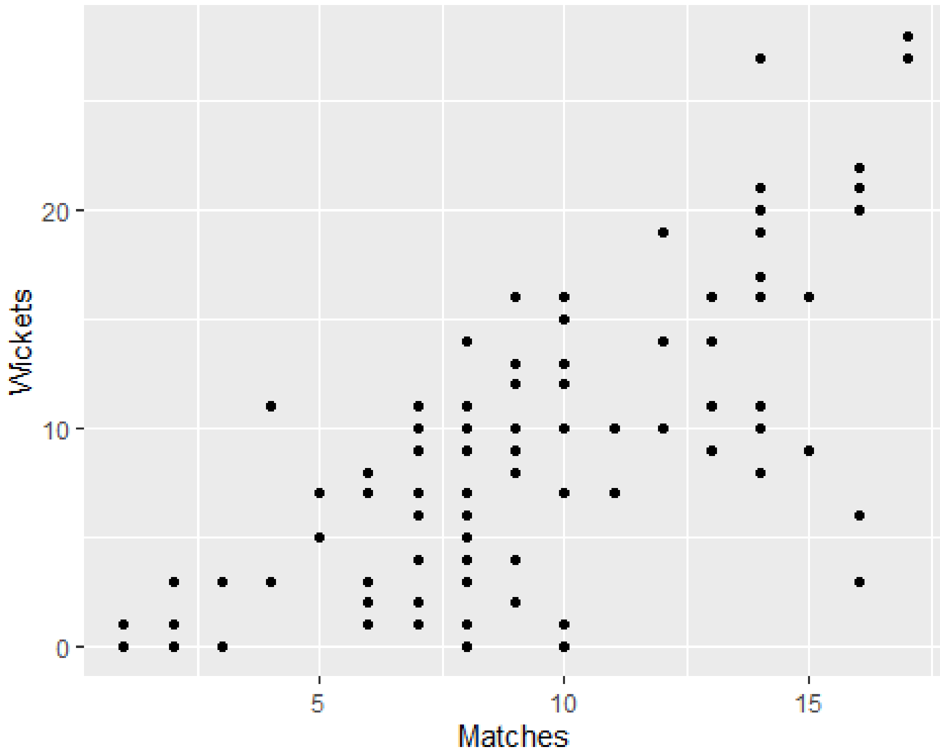



```
#Correlation Matrix

Data_Correlation <- read_csv("D:/R_dataset.csv")

Rows: 108 Columns: 14
── Column specification
────────────────────────────────────────────────────────
Delimiter: ","
chr (4): Name, Gender, Right_Left_Arm, Spin_Pace
dbl (7): SNo, Wickets, Matches, Innings, Economy_Rate, Age, Dot_Balls
num (3): Insta_Followers, Twitter_Followers, Auction_Price

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.

head(Data_Correlation)

# A tibble: 6 × 14
    SNo Name   Wickets Gender Matches Innings Economy_Rate   Age
Insta_Followers
  <dbl> <chr>    <dbl> <chr>    <dbl>   <dbl>        <dbl> <dbl>
<dbl>
1     1 Adam …       8 M            6       6         8.54    31
28200
2     2 Akash…      14 M            8       8         8.58    29
161000
3     3 Alice…       6 F            8       8         7.02    19
124000
4     4 Alzar…       7 M            7       7         9.37    26
59100
5     5 Amanj…       0 F           10       3         6.33    23
8974
6     6 Ameli…      15 F           10      10         6.45    22
0
# ℹ 5 more variables: Twitter_Followers <dbl>, Auction_Price <dbl>,
#   Dot_Balls <dbl>, Right_Left_Arm <chr>, Spin_Pace <chr>

# Encode categorical variables

# Male = 0 and Female = 1
Data_Correlation$Gender <- ifelse(Data_Correlation$Gender == "M", 0, 1)

# Right Arm = 0 and Left Arm = 1
Data_Correlation$Right_Left_Arm <- ifelse(Data_Correlation$Right_Left_Arm ==
"Right", 0, 1)
```

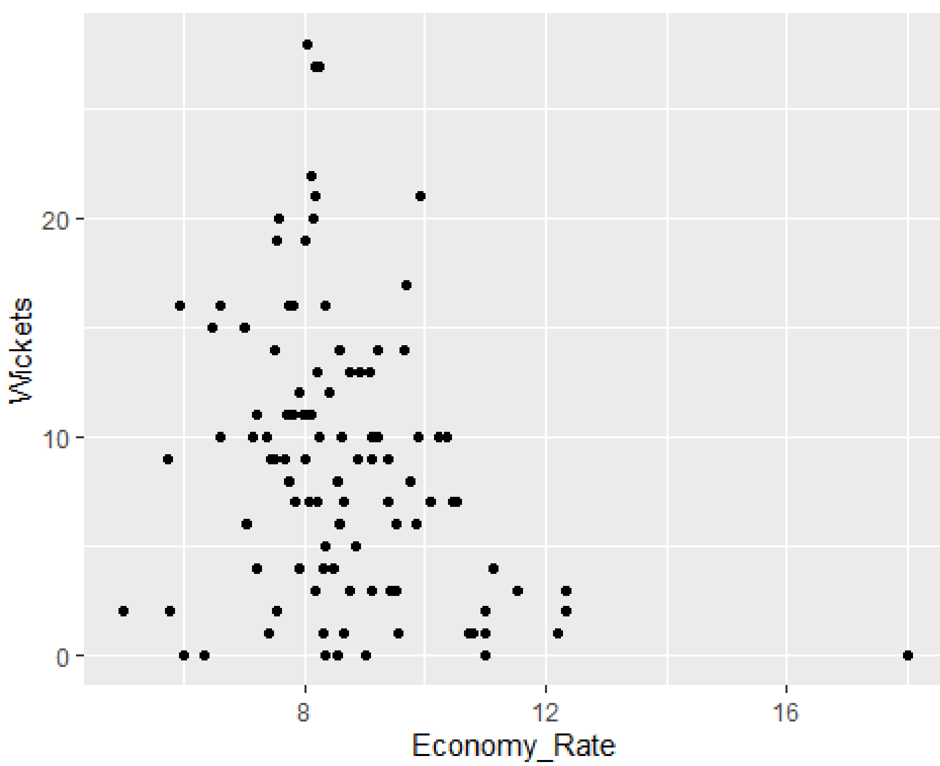
```r
# Spin = 0 and Pace = 1
Data_Correlation$Spin_Pace <- ifelse(Data_Correlation$Spin_Pace == "Spin", 0,
1)

columns_of_interest <- c("Gender", "Matches", "Innings", "Economy_Rate",
"Age", "Insta_Followers", "Twitter_Followers", "Auction_Price", "Dot_Balls",
"Right_Left_Arm", "Spin_Pace")
subset_data <- Data_Correlation[, columns_of_interest]
head(subset_data)

# A tibble: 6 × 11
  Gender Matches Innings Economy_Rate   Age Insta_Followers Twitter_Followers
   <dbl>   <dbl>   <dbl>        <dbl> <dbl>           <dbl>             <dbl>
1      0       6       6         8.54    31           28200                 0
2      0       8       8         8.58    29          161000                 0
3      1       8       8         7.02    19          124000             12900
4      0       7       7         9.37    26           59100                 0
5      1      10       3         6.33    23            8974                 0
6      1      10      10         6.45    22               0              2876
# ℹ 4 more variables: Auction_Price <dbl>, Dot_Balls <dbl>,
#   Right_Left_Arm <dbl>, Spin_Pace <dbl>

# Create a correlation matrix
M <- cor(subset_data)

# Plot the correlation matrix
corrplot(M, method = 'color', order = 'alphabet')
```



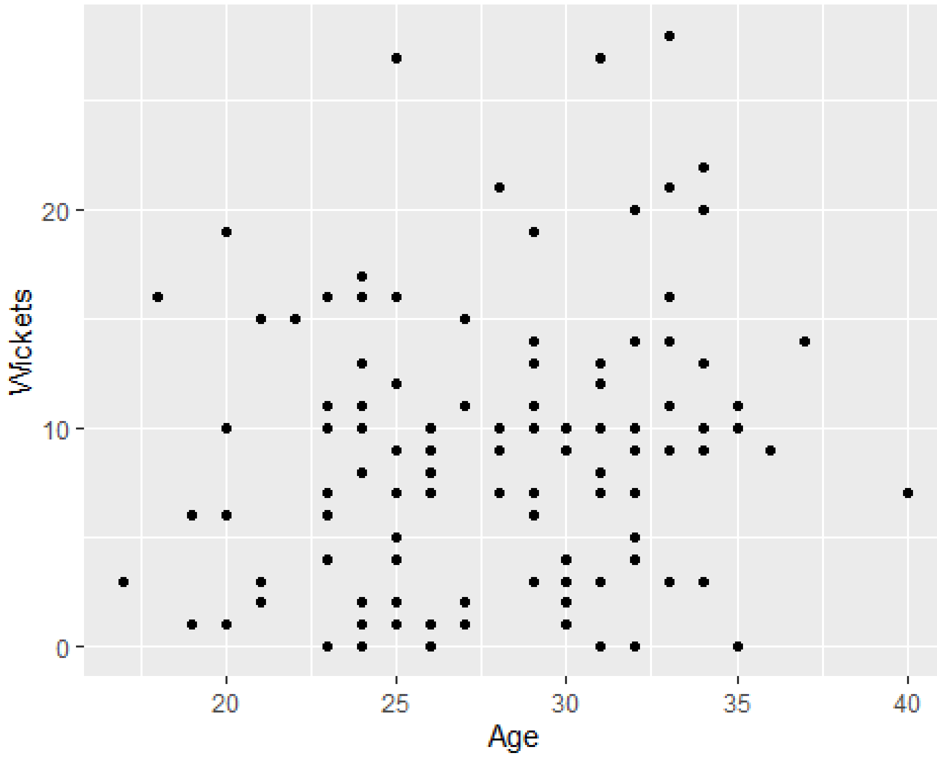

```
#Checking and removing outliers from quantitative variables

ggplot(Data, aes(x=Matches, y=Wickets)) + geom_point()
```



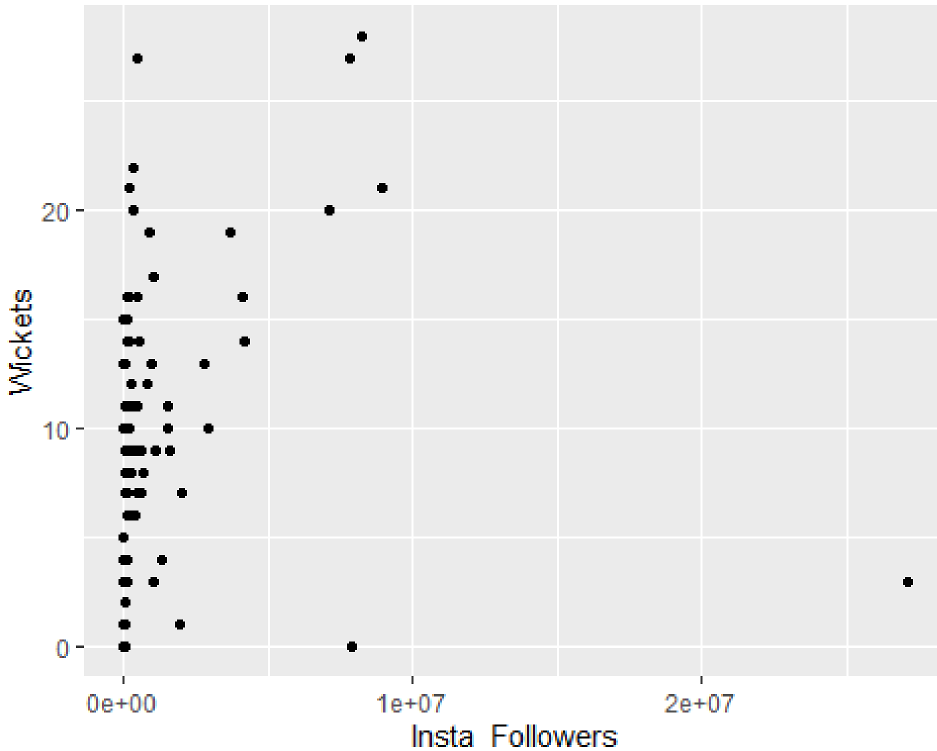

```
ggplot(Data, aes(x=Economy_Rate, y=Wickets)) + geom_point() # Outlier present
```



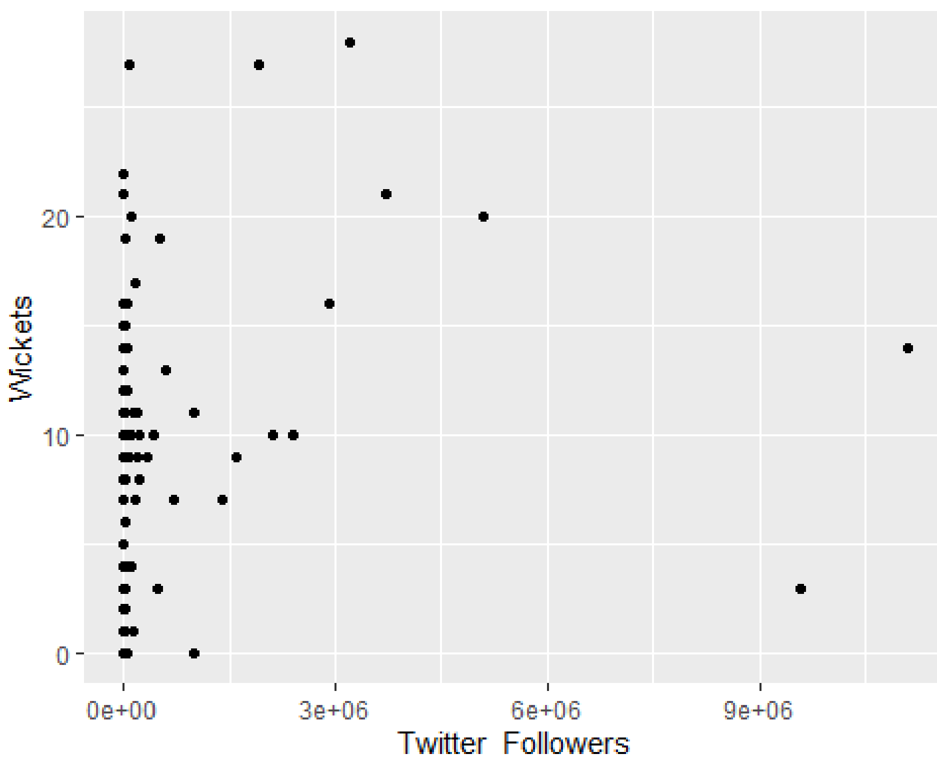
```
ggplot(Data, aes(x=Age, y=Wickets)) + geom_point()
```



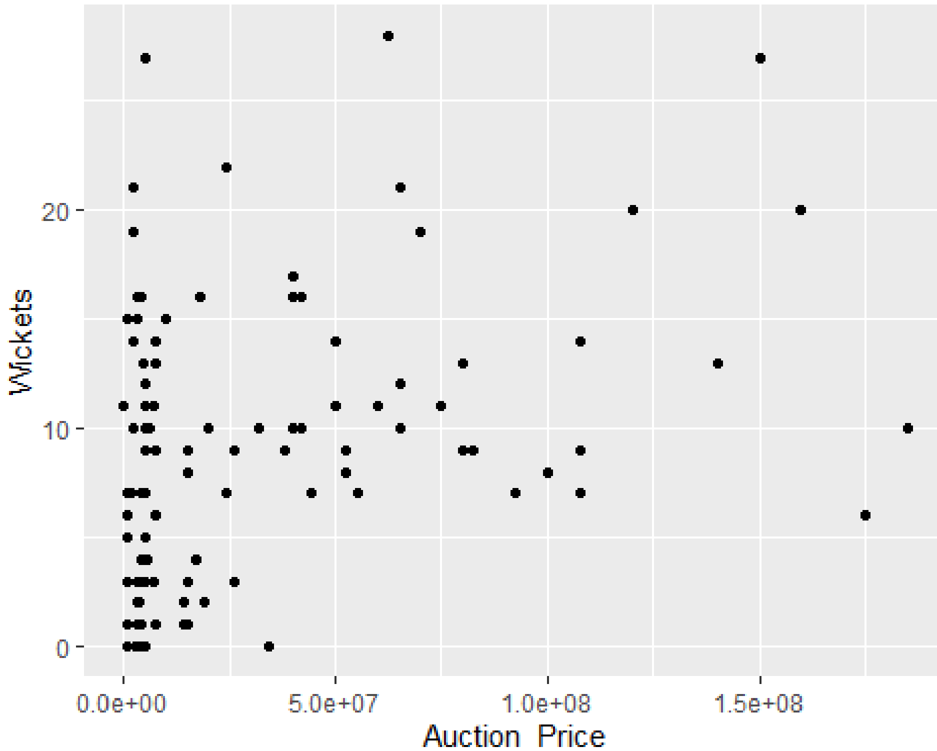

```
ggplot(Data, aes(x=Insta_Followers, y=Wickets)) + geom_point() #Hardik Pandya
outlier
```



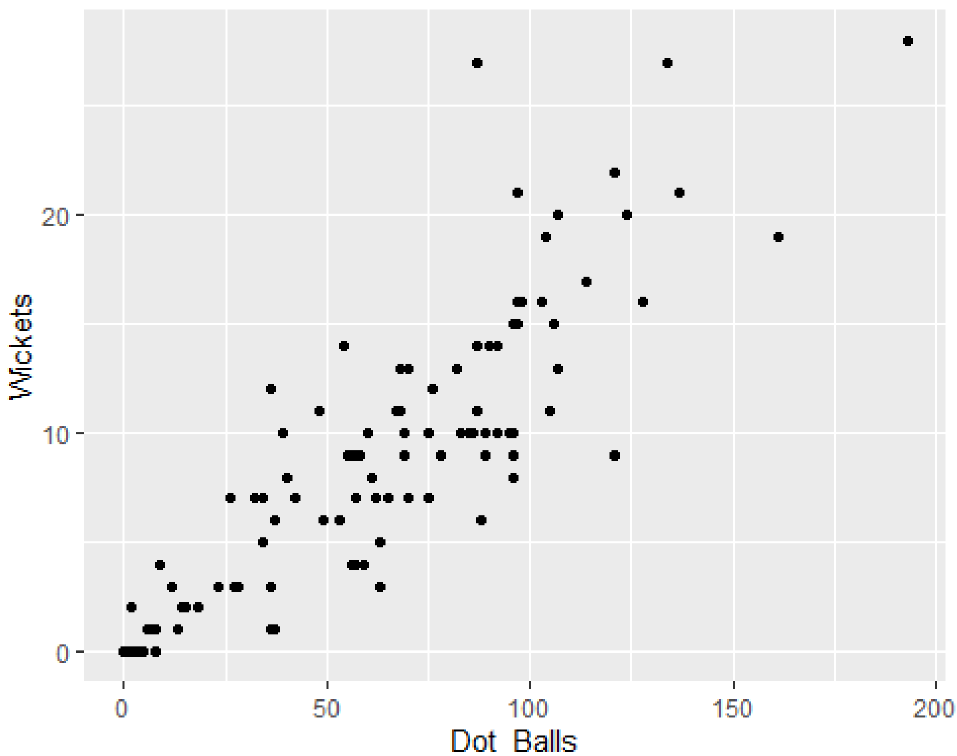

```r
ggplot(Data, aes(x=Twitter_Followers, y=Wickets)) + geom_point()
```



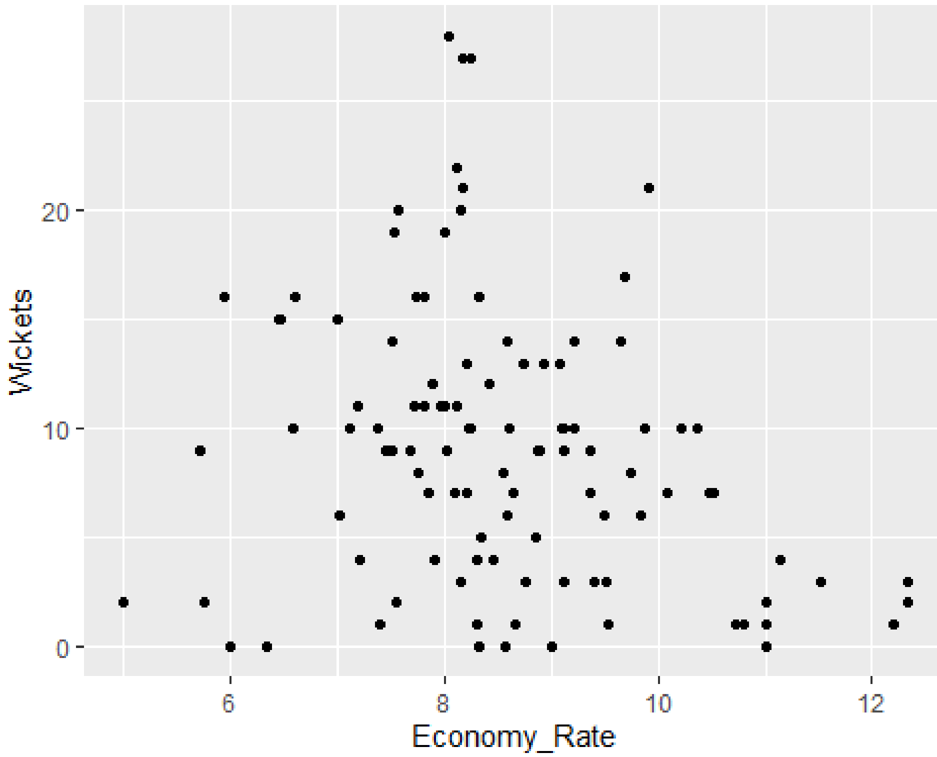
```
ggplot(Data, aes(x=Auction_Price, y=Wickets)) + geom_point()
```



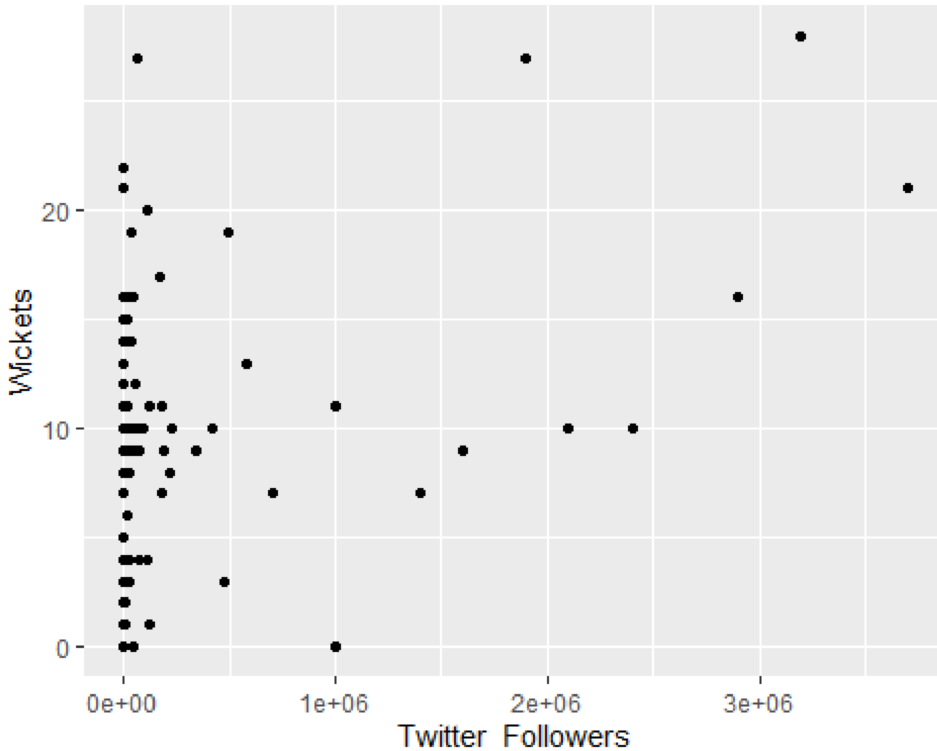

```
ggplot(Data, aes(x=Dot_Balls, y=Wickets)) + geom_point()
```



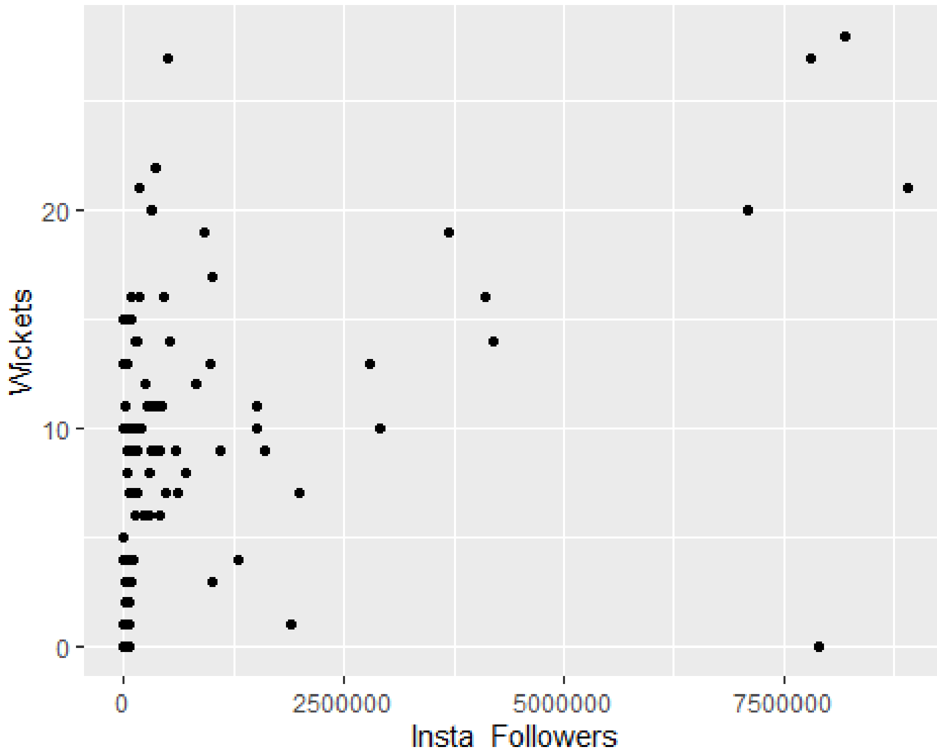

```
ggplot(Data[Data$Economy_Rate < 18, ], aes(x=Economy_Rate, y=Wickets)) +
geom_point()
```




```
sum(Data$Economy_Rate >= 18)

[1] 2

ggplot(Data[Data$Twitter_Followers < 5100000, ], aes(x=Twitter_Followers,
y=Wickets)) + geom_point()
```




```
sum(Data$Twitter_Followers >= 5100000)

[1] 3

ggplot(Data[Data$Insta_Followers < 9000000, ], aes(x=Insta_Followers,
y=Wickets)) + geom_point()
```



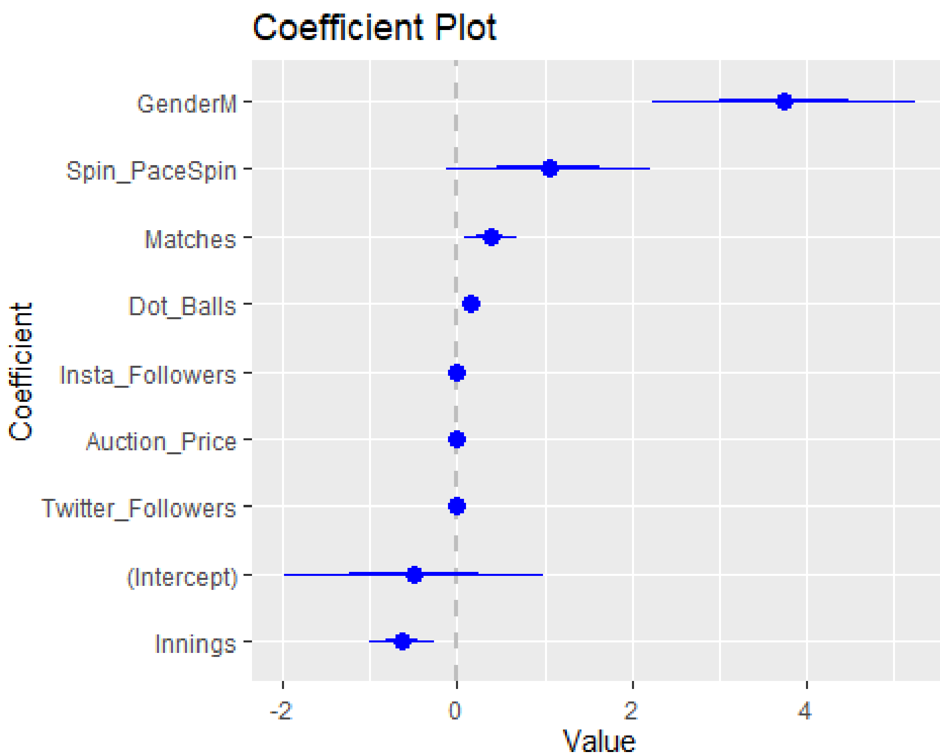

```
sum(Data$Insta_Followers >= 9000000)

[1] 1

Data <- Data[Data$Insta_Followers < 9000000,]
Data <- Data[Data$Twitter_Followers < 5100000,]
Data <- Data[Data$Economy_Rate < 18,]

view(Data)

#Multiple Linear Regression
Model_1 <- lm(Wickets ~
Matches+Innings+Insta_Followers+Twitter_Followers+Auction_Price+Dot_Balls+Gen
der+Spin_Pace, data = Data)
summary(Model_1)


Call:
lm(formula = Wickets ~ Matches + Innings + Insta_Followers +
    Twitter_Followers + Auction_Price + Dot_Balls + Gender +
    Spin_Pace, data = Data)

Residuals:
    Min      1Q  Median      3Q     Max
-6.3166 -1.7977 -0.1541  1.6479  6.3838
```

```
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)       -5.043e-01  7.398e-01  -0.682 0.497094    
Matches            3.760e-01  1.501e-01   2.505 0.013963 *  
Innings           -6.369e-01  1.889e-01  -3.371 0.001089 ** 
Insta_Followers    1.406e-06  3.969e-07   3.542 0.000621 ***
Twitter_Followers -2.101e-06  8.236e-07  -2.551 0.012347 *  
Auction_Price     -2.192e-08  9.096e-09  -2.410 0.017878 *  
Dot_Balls          1.435e-01  1.567e-02   9.158 1.13e-14 ***
GenderM            3.739e+00  7.481e-01   4.998 2.68e-06 ***
Spin_PaceSpin      1.045e+00  5.810e-01   1.798 0.075381 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.718 on 94 degrees of freedom
Multiple R-squared:  0.8332,    Adjusted R-squared:  0.819 
F-statistic: 58.71 on 8 and 94 DF,  p-value: < 2.2e-16
```

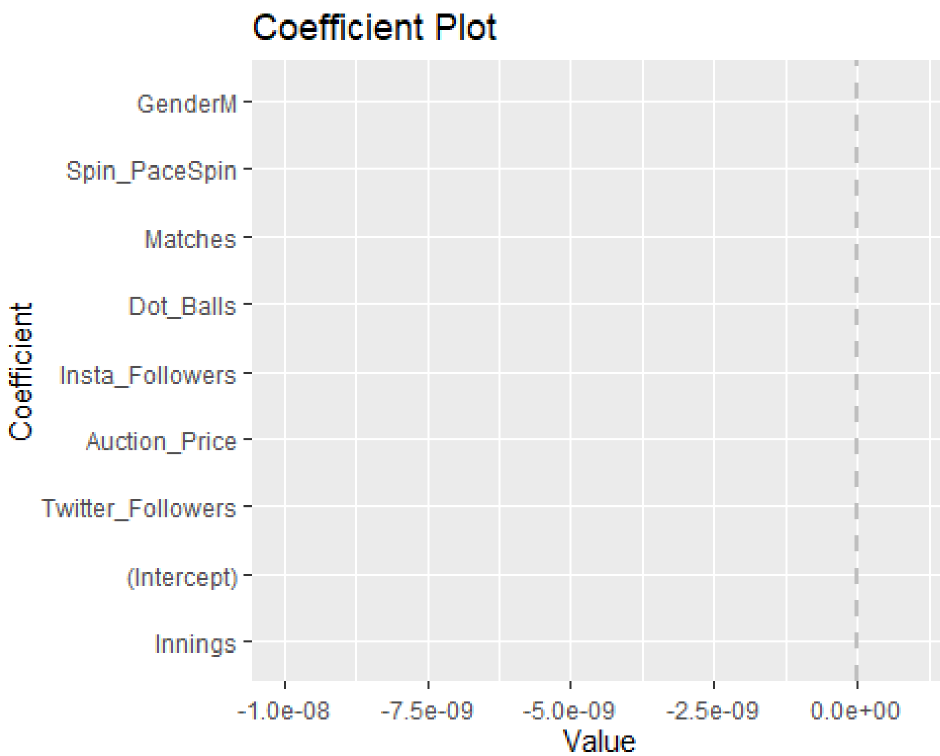
```
coef(Model_1)
```

```
      (Intercept)           Matches           Innings   Insta_Followers 
    -5.043430e-01      3.759895e-01     -6.368640e-01      1.405542e-06 
Twitter_Followers     Auction_Price         Dot_Balls           GenderM 
    -2.101354e-06     -2.192469e-08      1.435159e-01      3.738828e+00 
    Spin_PaceSpin 
     1.044583e+00 
```

```
coefplot(Model_1, sort='mag')
```

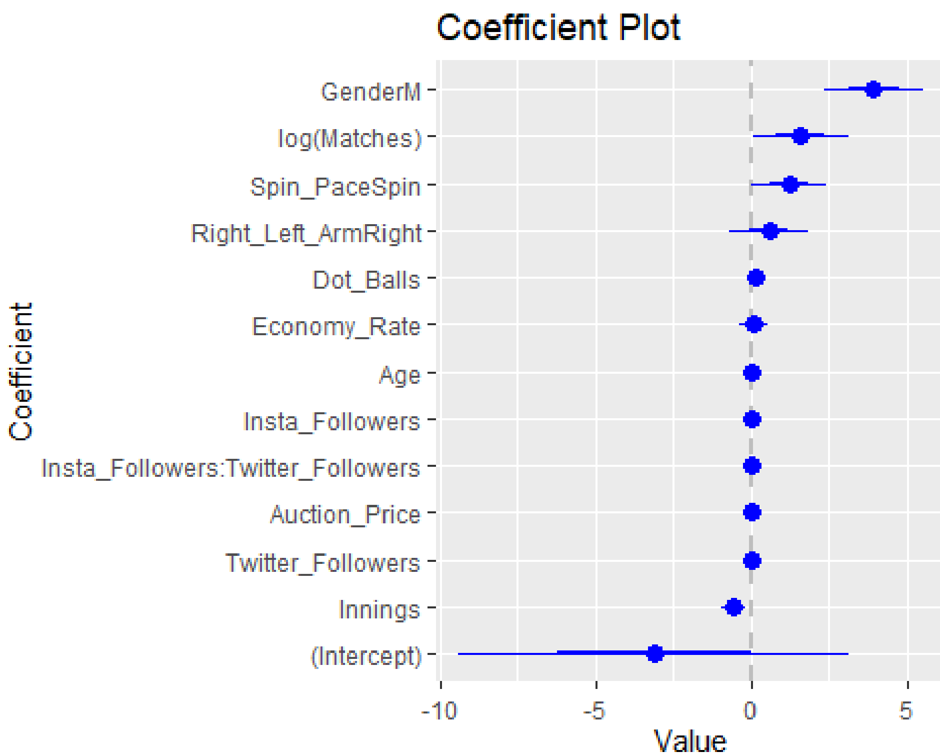



```
coefplot(Model_1, sort='mag') + scale_x_continuous(limits=c(-.00000001,
.000000001))

Warning: Removed 9 rows containing missing values (`geom_point()`).
```

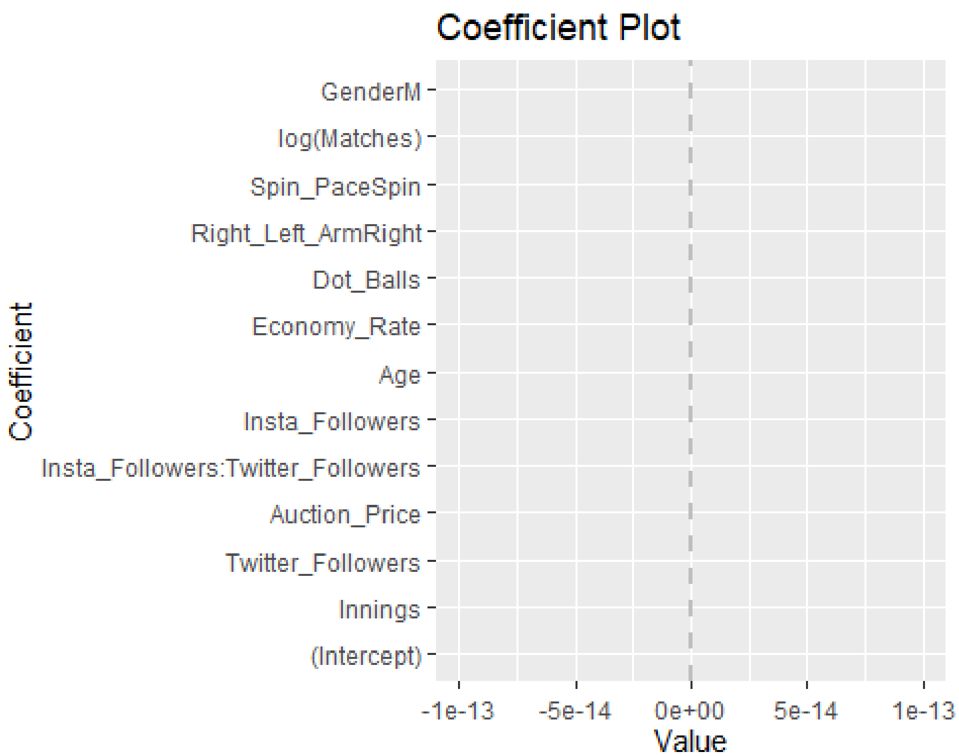



```
Model_2 <- lm(Wickets ~
log(Matches)+Innings+Economy_Rate+Age+Insta_Followers*Twitter_Followers+Aucti
on_Price+Dot_Balls+Gender+Right_Left_Arm+Spin_Pace, data = Data)
summary(Model_2)


Call:
lm(formula = Wickets ~ log(Matches) + Innings + Economy_Rate +
    Age + Insta_Followers * Twitter_Followers + Auction_Price +
    Dot_Balls + Gender + Right_Left_Arm + Spin_Pace, data = Data)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7387 -1.7782 -0.2874  1.7693  7.5861

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                      -3.125e+00  3.120e+00  -1.002  0.31914
log(Matches)                      1.586e+00  7.640e-01   2.076  0.04073 *
Innings                          -5.628e-01  1.870e-01  -3.010  0.00339 **
Economy_Rate                      7.775e-02  2.348e-01   0.331  0.74127
Age                               2.387e-02  6.253e-02   0.382  0.70362
Insta_Followers                   7.532e-07  5.818e-07   1.295  0.19878
Twitter_Followers                -2.737e-06  9.450e-07  -2.897  0.00474 **
Auction_Price                    -1.687e-08  9.578e-09  -1.762  0.08150 .
```

```
Dot_Balls                          1.484e-01  1.713e-02   8.662 1.74e-13 ***
GenderM                            3.940e+00  7.904e-01   4.986 2.98e-06 ***
Right_Left_ArmRight                5.806e-01  6.359e-01   0.913  0.36364
Spin_PaceSpin                      1.234e+00  6.091e-01   2.027  0.04566 *
Insta_Followers:Twitter_Followers  3.175e-13  1.825e-13   1.740  0.08525 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.745 on 90 degrees of freedom
Multiple R-squared:  0.8373,    Adjusted R-squared:  0.8156
F-statistic: 38.58 on 12 and 90 DF,  p-value: < 2.2e-16

coef(Model_2)

                      (Intercept)                       log(Matches)
                    -3.125124e+00                       1.586175e+00
                          Innings                       Economy_Rate
                    -5.627696e-01                       7.774951e-02
                              Age                    Insta_Followers
                     2.386645e-02                       7.532078e-07
                Twitter_Followers                      Auction_Price
                    -2.737269e-06                      -1.687497e-08
                        Dot_Balls                            GenderM
                     1.483830e-01                       3.940315e+00
              Right_Left_ArmRight                      Spin_PaceSpin
                     5.805996e-01                       1.234408e+00
Insta_Followers:Twitter_Followers
                     3.175435e-13

coefplot(Model_2, sort='mag')
```




```
coefplot(Model_2, sort='mag') + scale_x_continuous(limits=c(-.0000000000001,
.0000000000001))

Warning: Removed 13 rows containing missing values (`geom_point()`).
```



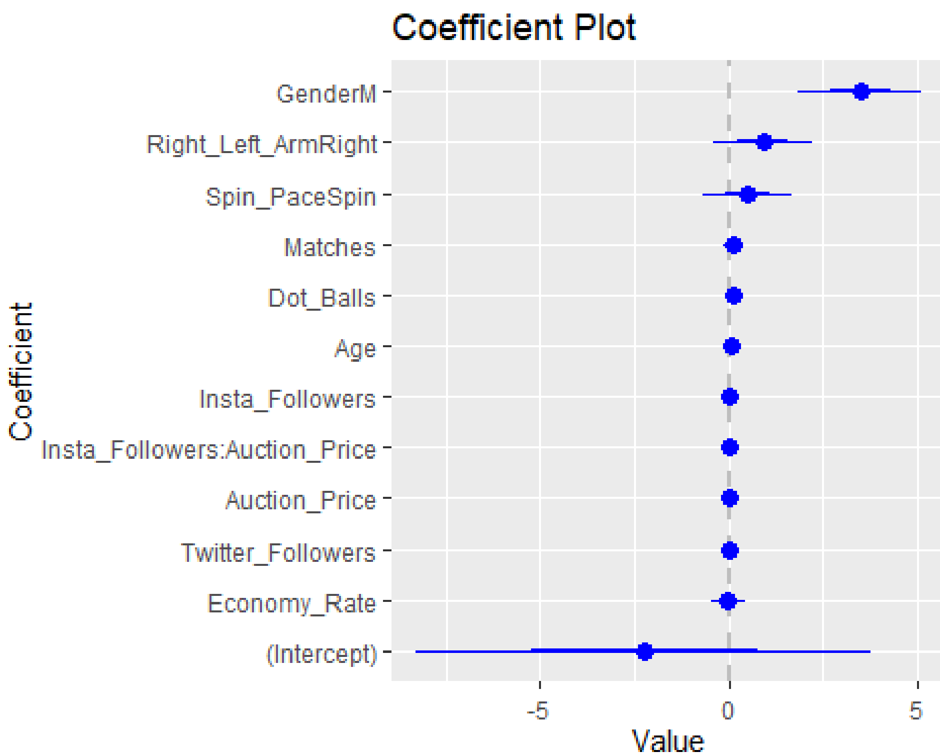

```
Model_3 <- lm(Wickets ~
Matches+Economy_Rate+Age+Insta_Followers*Auction_Price+Dot_Balls+Gender+
Right_Left_Arm+Spin_Pace+Twitter_Followers, data = Data)
summary(Model_3)


Call:
lm(formula = Wickets ~ Matches + Economy_Rate + Age + Insta_Followers *
    Auction_Price + Dot_Balls + Gender + Right_Left_Arm + Spin_Pace +
    Twitter_Followers, data = Data)

Residuals:
    Min      1Q  Median      3Q     Max
-6.1709 -1.3719 -0.0949  1.0862 12.2962

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    -2.245e+00  3.003e+00  -0.748 0.456653
Matches                         1.200e-01  1.320e-01   0.910 0.365420
Economy_Rate                   -2.156e-02  2.332e-01  -0.092 0.926529
Age                             5.109e-02  6.384e-02   0.800 0.425694
Insta_Followers                 2.877e-07  6.991e-07   0.411 0.681695
Auction_Price                  -3.992e-08  1.029e-08  -3.878 0.000199 ***
Dot_Balls                       1.106e-01  1.301e-02   8.500 3.49e-13 ***
GenderM                         3.468e+00  8.065e-01   4.301 4.27e-05 ***
```

```
Right_Left_ArmRight            9.030e-01  6.504e-01   1.388 0.168442
Spin_PaceSpin                  4.924e-01  5.974e-01   0.824 0.411963
Twitter_Followers             -1.547e-06  9.772e-07  -1.583 0.116802
Insta_Followers:Auction_Price  1.165e-14  5.099e-15   2.284 0.024674 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.812 on 91 degrees of freedom
Multiple R-squared:  0.8272,	Adjusted R-squared:  0.8063
F-statistic: 39.61 on 11 and 91 DF,  p-value: < 2.2e-16
```

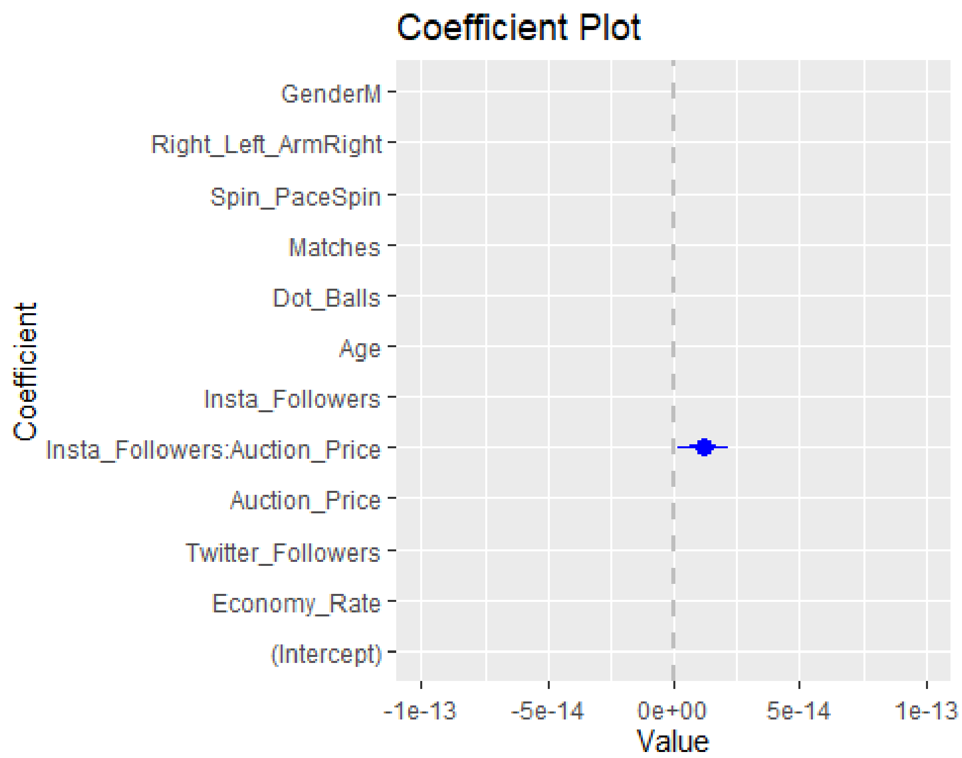
```r
coef(Model_3)
```

```
                  (Intercept)                       Matches
                -2.245164e+00                  1.200488e-01
                 Economy_Rate                           Age
                -2.156050e-02                  5.108681e-02
              Insta_Followers                 Auction_Price
                 2.876749e-07                 -3.992422e-08
                    Dot_Balls                       GenderM
                 1.106137e-01                  3.468459e+00
          Right_Left_ArmRight                 Spin_PaceSpin
                 9.029625e-01                  4.923698e-01
            Twitter_Followers Insta_Followers:Auction_Price
                -1.547240e-06                  1.164715e-14
```

```r
coefplot(Model_3, sort='mag')
```




```
coefplot(Model_3, sort='mag') + scale_x_continuous(limits=c(-.0000000000001,
.0000000000001))

Warning: Removed 11 rows containing missing values (`geom_point()`).
```



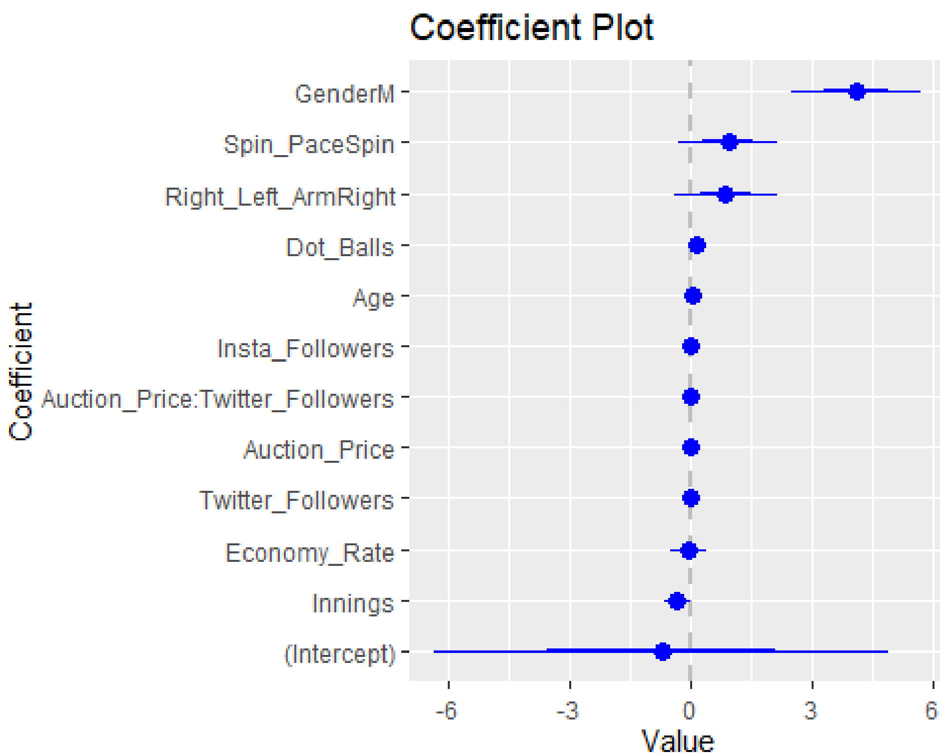

```
Model_4 <- lm(Wickets ~
Economy_Rate+Innings+Age+Insta_Followers+Auction_Price*Twitter_Followers+Dot_
Balls+Right_Left_Arm+Spin_Pace+Gender, data = Data)
summary(Model_4)


Call:
lm(formula = Wickets ~ Economy_Rate + Innings + Age + Insta_Followers +
    Auction_Price * Twitter_Followers + Dot_Balls + Right_Left_Arm +
    Spin_Pace + Gender, data = Data)

Residuals:
    Min      1Q  Median      3Q     Max
-6.0851 -1.6869 -0.0061  1.6734  9.6462

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                    -7.231e-01  2.822e+00  -0.256  0.79835
Economy_Rate                   -3.870e-02  2.280e-01  -0.170  0.86560
Innings                        -3.425e-01  1.633e-01  -2.097  0.03880 *
Age                             3.460e-02  6.373e-02   0.543  0.58856
Insta_Followers                 7.144e-07  5.870e-07   1.217  0.22673
Auction_Price                  -2.899e-08  1.013e-08  -2.863  0.00521 **
Twitter_Followers              -2.671e-06  9.197e-07  -2.904  0.00463 **
Dot_Balls                       1.453e-01  1.729e-02   8.405 5.53e-13 ***
```

```
Right_Left_ArmRight              8.710e-01  6.480e-01   1.344  0.18222
Spin_PaceSpin                    9.241e-01  6.167e-01   1.499  0.13744
GenderM                          4.116e+00  8.001e-01   5.145 1.53e-06 ***
Auction_Price:Twitter_Followers  3.389e-14  1.815e-14   1.867  0.06516 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.779 on 91 degrees of freedom
Multiple R-squared:  0.8313,    Adjusted R-squared:  0.8109
F-statistic: 40.77 on 11 and 91 DF,  p-value: < 2.2e-16
```

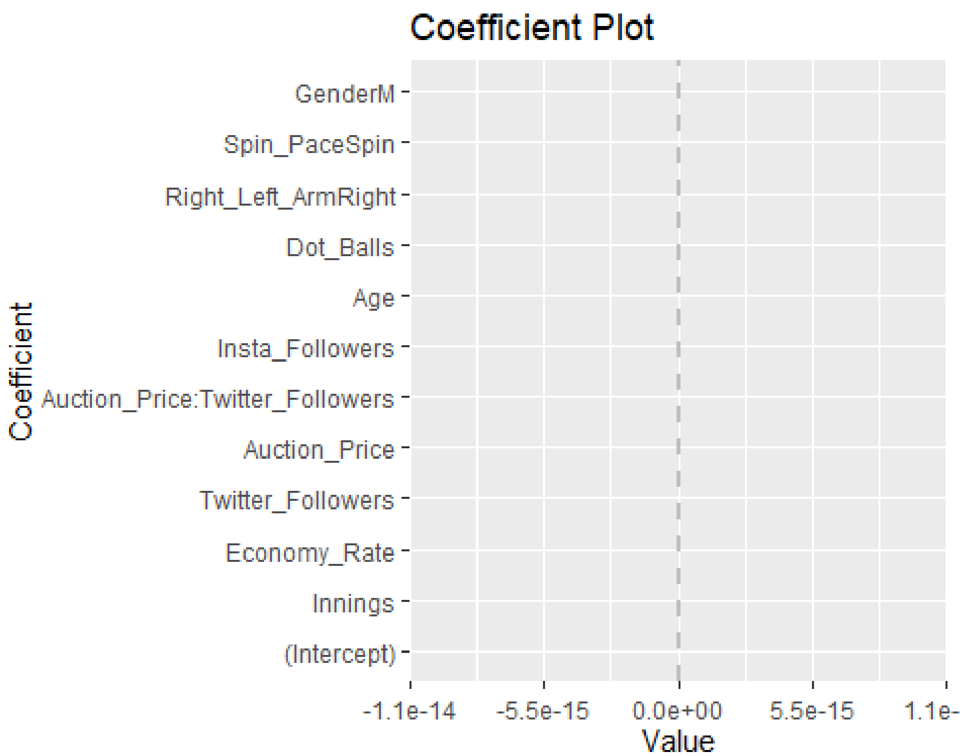
```r
coef(Model_4)
```

```
                    (Intercept)                    Economy_Rate
                  -7.231432e-01                   -3.870388e-02
                        Innings                             Age
                  -3.424693e-01                    3.459573e-02
                Insta_Followers                   Auction_Price
                   7.143977e-07                   -2.898760e-08
              Twitter_Followers                       Dot_Balls
                  -2.670615e-06                    1.452811e-01
            Right_Left_ArmRight                   Spin_PaceSpin
                   8.710031e-01                    9.241304e-01
                        GenderM Auction_Price:Twitter_Followers
                   4.116307e+00                    3.388986e-14
```

```r
coefplot(Model_4, sort='mag')
```




```
coefplot(Model_4, sort='mag') + scale_x_continuous(limits=c(-.00000000000001,
.00000000000001))

Warning: Removed 12 rows containing missing values (`geom_point()`).
```



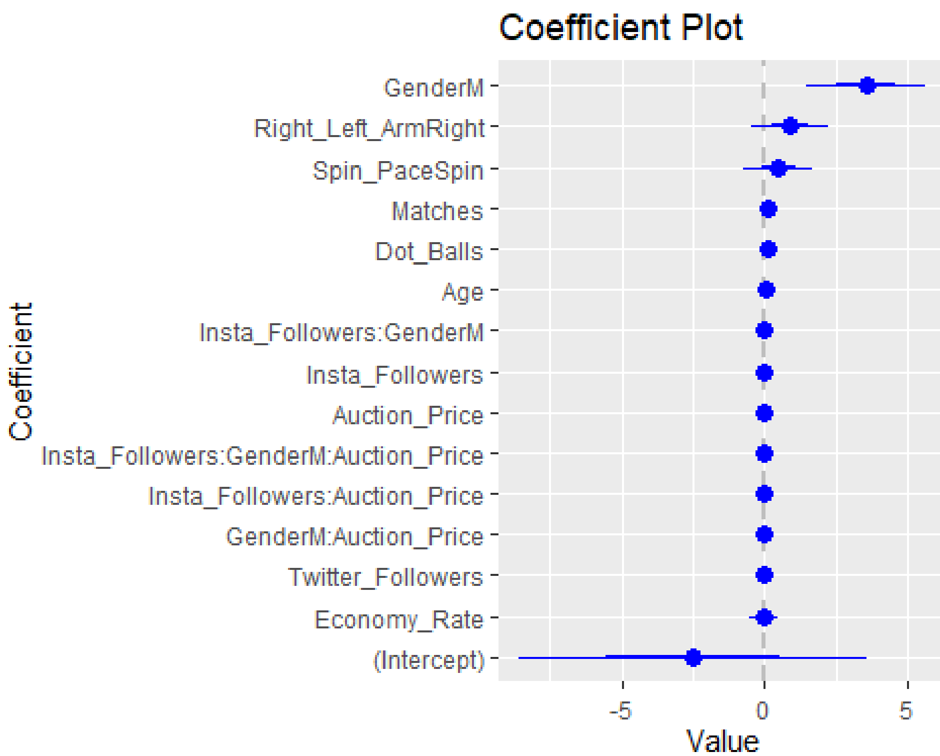

```
Model_5 <- lm(Wickets ~
Matches+Economy_Rate+Age+Insta_Followers*Gender*Auction_Price+Twitter_Followe
rs+Dot_Balls+Right_Left_Arm+Spin_Pace, data = Data)
summary(Model_5)


Call:
lm(formula = Wickets ~ Matches + Economy_Rate + Age + Insta_Followers *
    Gender * Auction_Price + Twitter_Followers + Dot_Balls +
    Right_Left_Arm + Spin_Pace, data = Data)

Residuals:
    Min      1Q  Median      3Q     Max
-6.1632 -1.4119 -0.3449  1.2319 12.1847

Coefficients:
                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                           -2.489e+00  3.046e+00  -0.817 0.416038
Matches                                1.282e-01  1.360e-01   0.943 0.348429
Economy_Rate                          -2.110e-02  2.394e-01  -0.088 0.929953
Age                                    5.772e-02  6.501e-02   0.888 0.377013
Insta_Followers                        6.722e-08  2.187e-06   0.031 0.975555
GenderM                                3.569e+00  1.038e+00   3.440 0.000892
Auction_Price                          2.389e-09  6.657e-08   0.036 0.971452
Twitter_Followers                     -1.963e-06  1.059e-06  -1.854 0.067094
```

```
Dot_Balls                               1.083e-01  1.351e-02   8.014 4.35e-12
Right_Left_ArmRight                     8.871e-01  6.642e-01   1.336 0.185138
Spin_PaceSpin                           4.940e-01  6.063e-01   0.815 0.417468
Insta_Followers:GenderM                 5.930e-07  2.303e-06   0.257 0.797411
Insta_Followers:Auction_Price          -1.116e-13  2.190e-13  -0.510 0.611489
GenderM:Auction_Price                  -4.231e-08  6.704e-08  -0.631 0.529632
Insta_Followers:GenderM:Auction_Price   1.212e-13  2.191e-13   0.553 0.581510

(Intercept)
Matches
Economy_Rate
Age
Insta_Followers
GenderM                                ***
Auction_Price
Twitter_Followers                      .
Dot_Balls                              ***
Right_Left_ArmRight
Spin_PaceSpin
Insta_Followers:GenderM
Insta_Followers:Auction_Price
GenderM:Auction_Price
Insta_Followers:GenderM:Auction_Price
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.836 on 88 degrees of freedom
Multiple R-squared:  0.8301,    Adjusted R-squared:  0.8031
F-statistic: 30.71 on 14 and 88 DF,  p-value: < 2.2e-16
```

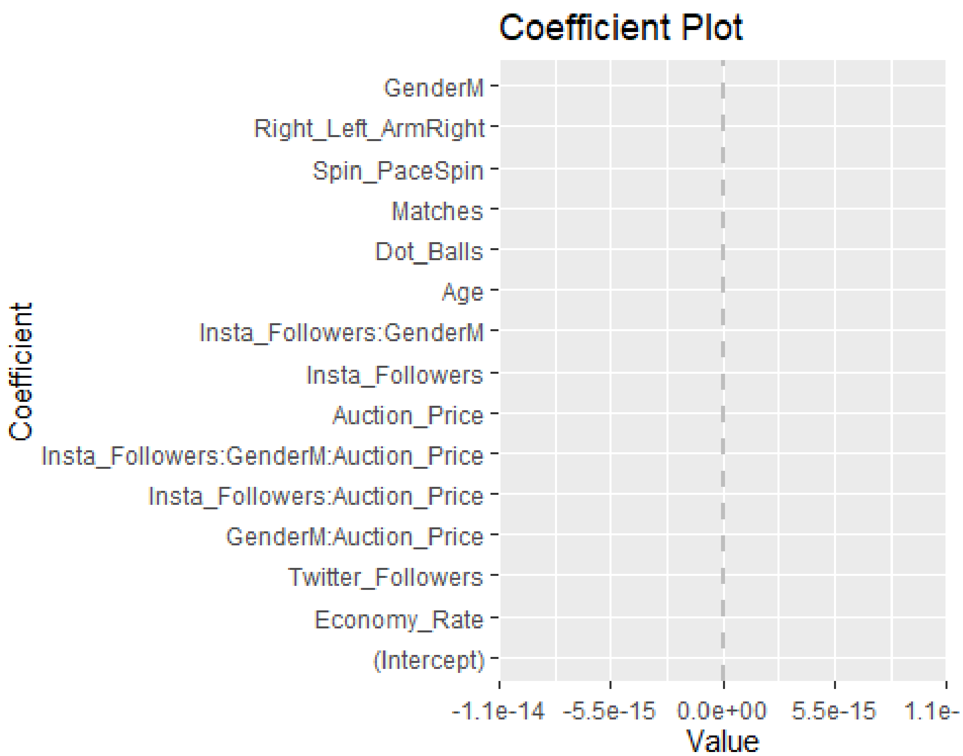
```
coef(Model_5)
```

```
                          (Intercept)                               Matches
                        -2.489154e+00                          1.282012e-01
                         Economy_Rate                                   Age
                        -2.110088e-02                          5.771818e-02
                      Insta_Followers                               GenderM
                         6.721658e-08                          3.568978e+00
                        Auction_Price                     Twitter_Followers
                         2.389246e-09                         -1.963285e-06
                            Dot_Balls                   Right_Left_ArmRight
                         1.083016e-01                          8.871248e-01
                        Spin_PaceSpin               Insta_Followers:GenderM
                         4.939640e-01                          5.929837e-07
        Insta_Followers:Auction_Price                 GenderM:Auction_Price
                        -1.116151e-13                         -4.230811e-08
Insta_Followers:GenderM:Auction_Price
                         1.212331e-13
```

```
coefplot(Model_5, sort='mag')
```

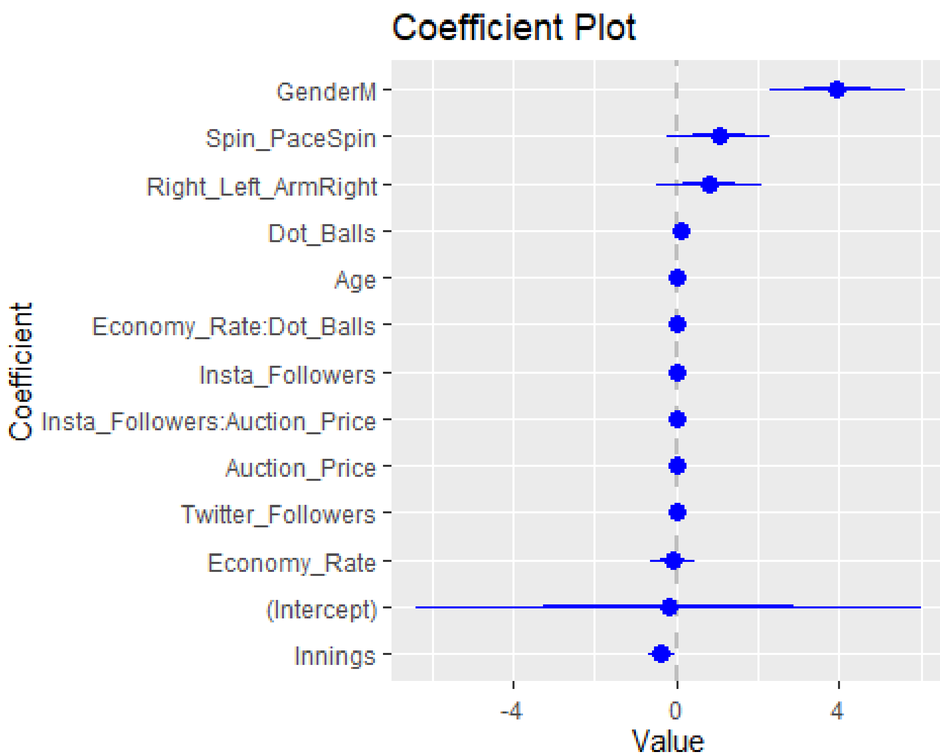



```
coefplot(Model_5, sort='mag') + scale_x_continuous(limits=c(-.00000000000001,
.00000000000001))

Warning: Removed 15 rows containing missing values (`geom_point()`).
```

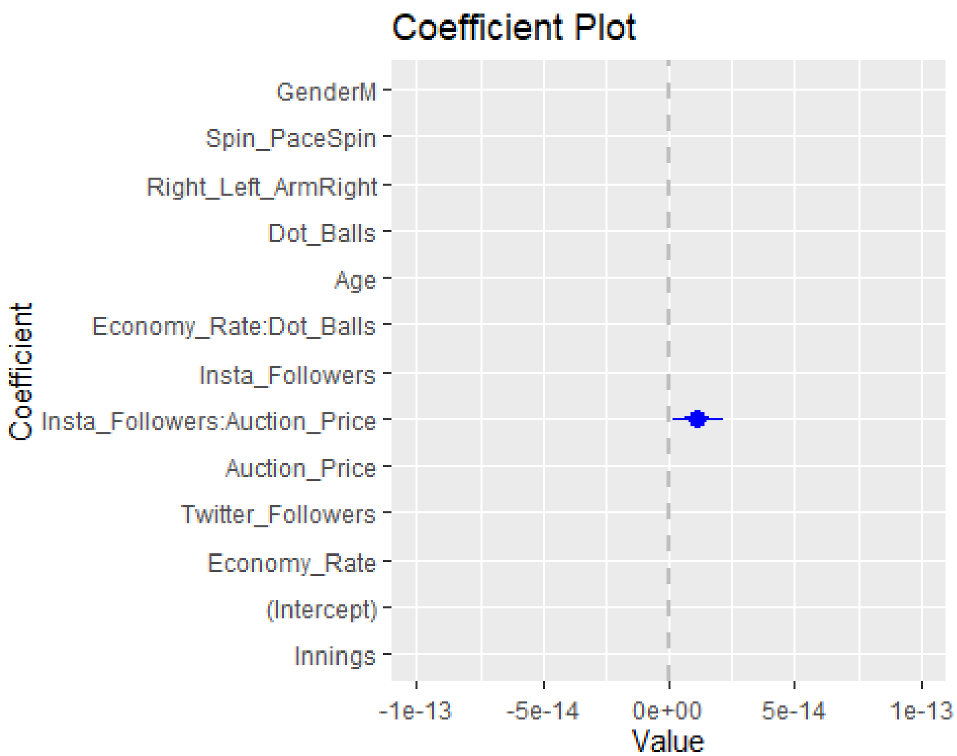



```
Model_6 <- lm(Wickets ~
Economy_Rate*Dot_Balls+Innings+Age+Insta_Followers*Auction_Price+Twitter_Foll
owers+Gender+Right_Left_Arm+Spin_Pace, data = Data)
summary(Model_6)


Call:
lm(formula = Wickets ~ Economy_Rate * Dot_Balls + Innings + Age +
    Insta_Followers * Auction_Price + Twitter_Followers + Gender +
    Right_Left_Arm + Spin_Pace, data = Data)

Residuals:
    Min      1Q  Median      3Q     Max
-5.6851 -1.6571 -0.1629  1.2408  9.5949

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                   -1.787e-01  3.095e+00  -0.058  0.95409
Economy_Rate                  -8.495e-02  2.846e-01  -0.298  0.76603
Dot_Balls                      1.291e-01  4.529e-02   2.850  0.00542 **
Innings                       -3.702e-01  1.665e-01  -2.223  0.02870 *
Age                            3.010e-02  6.335e-02   0.475  0.63577
Insta_Followers                2.543e-07  6.873e-07   0.370  0.71232
Auction_Price                 -3.186e-08  1.028e-08  -3.098  0.00260 **
Twitter_Followers             -1.381e-06  9.645e-07  -1.432  0.15564
```

```
GenderM                         3.957e+00  8.332e-01   4.749 7.68e-06 ***
Right_Left_ArmRight             8.053e-01  6.407e-01   1.257  0.21207
Spin_PaceSpin                   1.049e+00  6.315e-01   1.661  0.10022
Economy_Rate:Dot_Balls          2.564e-03  5.759e-03   0.445  0.65720
Insta_Followers:Auction_Price   1.126e-14  5.009e-15   2.248  0.02704 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.766 on 90 degrees of freedom
Multiple R-squared:  0.8347,    Adjusted R-squared:  0.8127
F-statistic: 37.88 on 12 and 90 DF,  p-value: < 2.2e-16

coef(Model_6)

                  (Intercept)                  Economy_Rate
                -1.786594e-01                 -8.495493e-02
                    Dot_Balls                       Innings
                 1.290792e-01                 -3.702482e-01
                          Age               Insta_Followers
                 3.010442e-02                  2.542533e-07
                Auction_Price             Twitter_Followers
                -3.186464e-08                 -1.381044e-06
                      GenderM            Right_Left_ArmRight
                 3.957129e+00                  8.053065e-01
                Spin_PaceSpin        Economy_Rate:Dot_Balls
                 1.048756e+00                  2.564056e-03
Insta_Followers:Auction_Price
                 1.125921e-14

coefplot(Model_6, sort='mag')
```

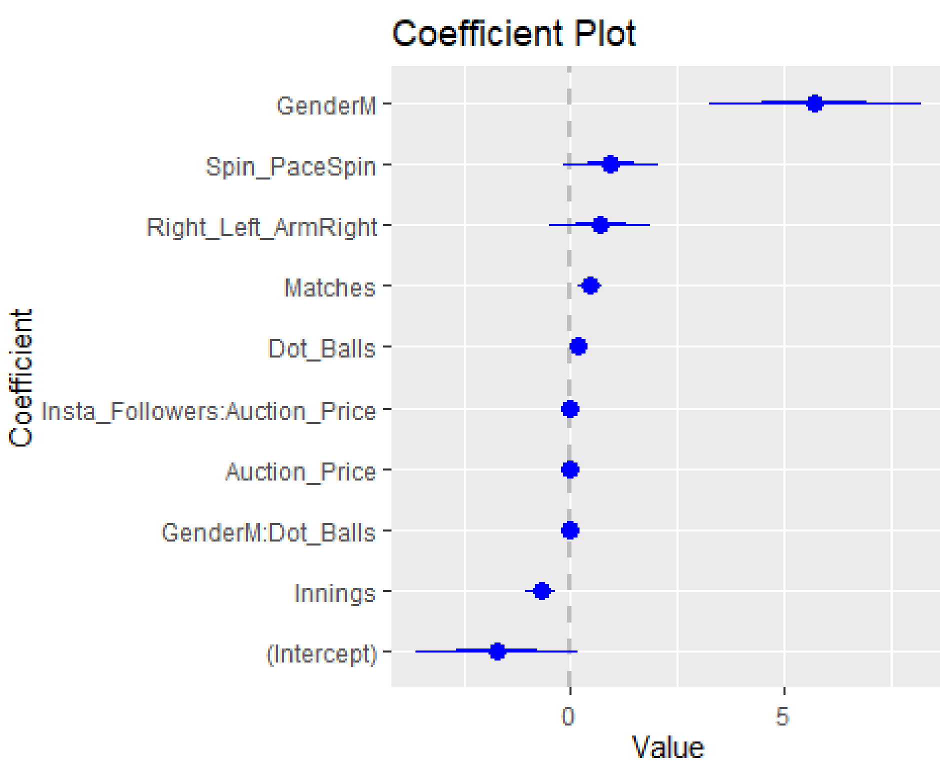



```
coefplot(Model_6, sort='mag') + scale_x_continuous(limits=c(-.0000000000001,
.0000000000001))

Warning: Removed 12 rows containing missing values (`geom_point()`).
```



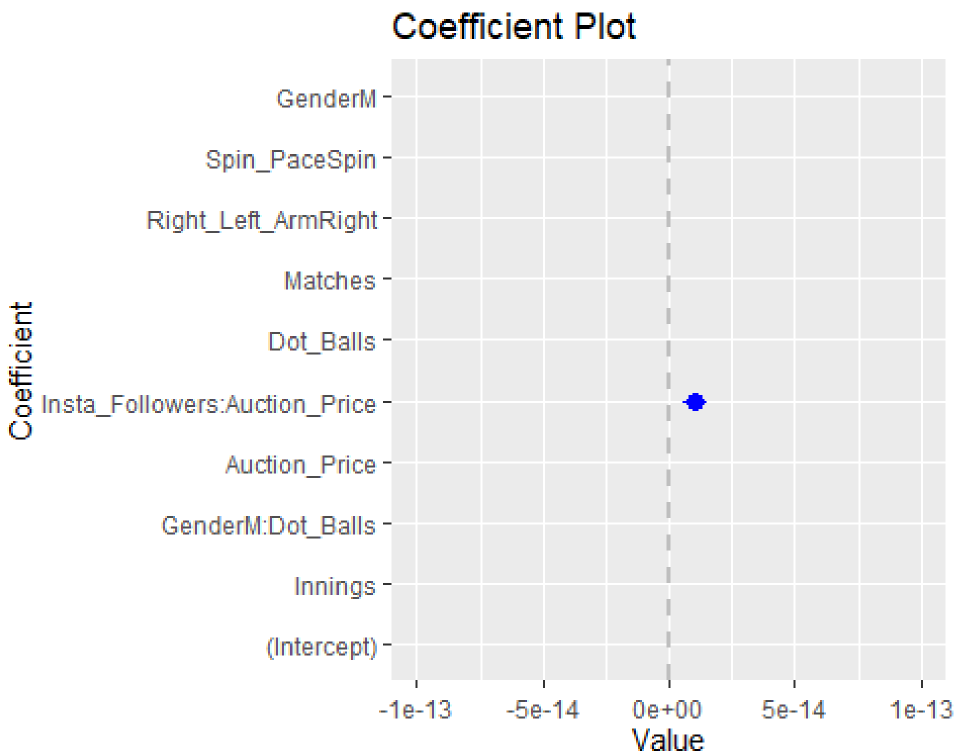

```
Model_7 <- lm(Wickets ~
Matches+Innings+Gender*Dot_Balls+Insta_Followers:Auction_Price+Auction_Price+
Spin_Pace+Right_Left_Arm, data = Data)
summary(Model_7)


Call:
lm(formula = Wickets ~ Matches + Innings + Gender * Dot_Balls +
    Insta_Followers:Auction_Price + Auction_Price + Spin_Pace +
    Right_Left_Arm, data = Data)

Residuals:
    Min      1Q  Median      3Q     Max
-6.8430 -1.8130 -0.2342  1.5058  5.9047

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                   -1.715e+00  9.425e-01  -1.819 0.072093 .
Matches                        4.605e-01  1.484e-01   3.104 0.002532 **
Innings                       -6.984e-01  1.842e-01  -3.792 0.000266 ***
GenderM                        5.730e+00  1.246e+00   4.599 1.34e-05 ***
Dot_Balls                      1.605e-01  1.674e-02   9.586 1.54e-15 ***
Auction_Price                 -2.887e-08  9.248e-09  -3.122 0.002392 **
Spin_PaceSpin                  9.519e-01  5.630e-01   1.691 0.094248 .
Right_Left_ArmRight            6.992e-01  5.919e-01   1.181 0.240514
```

```
GenderM:Dot_Balls             -3.459e-02  1.680e-02  -2.059 0.042301 *
Insta_Followers:Auction_Price  1.039e-14  2.380e-15   4.367 3.27e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.622 on 93 degrees of freedom
Multiple R-squared:  0.8465,    Adjusted R-squared:  0.8317
F-statistic: 56.99 on 9 and 93 DF,  p-value: < 2.2e-16

coef(Model_7)

                  (Intercept)                       Matches
                -1.714575e+00                  4.605067e-01
                     Innings                       GenderM
                -6.984056e-01                  5.729561e+00
                    Dot_Balls                 Auction_Price
                 1.604696e-01                 -2.887414e-08
               Spin_PaceSpin            Right_Left_ArmRight
                 9.518665e-01                  6.991749e-01
            GenderM:Dot_Balls Insta_Followers:Auction_Price
                -3.458697e-02                  1.039170e-14

coefplot(Model_7, sort='mag')
```




```
coefplot(Model_7, sort='mag') + scale_x_continuous(limits=c(-.0000000000001,
.0000000000001))
```

```
Warning: Removed 9 rows containing missing values (`geom_point()`).
```




```
# Step Function to find best model
nullModel <- lm(Wickets ~ 1, data=Data)
fullModel <- lm(Wickets ~
Spin_Pace*Twitter_Followers*Right_Left_Arm*Auction_Price*
                  Dot_Balls*Gender*Innings*Economy_Rate*Age*Insta_Followers,
data = Data)

WicketsStep <- step(nullModel,scope=list(lower=nullModel, upper=fullModel),
direction="both")

Start:  AIC=383.09
Wickets ~ 1

                    Df Sum of Sq    RSS    AIC
+ Dot_Balls          1   3117.00 1048.5 243.00
+ Innings            1   2286.89 1878.6 303.07
+ Gender             1   1434.41 2731.1 341.61
+ Insta_Followers    1   1110.15 3055.3 353.16
+ Twitter_Followers  1    644.87 3520.6 367.76
+ Auction_Price      1    445.91 3719.6 373.42
+ Economy_Rate       1    283.70 3881.8 377.82
+ Age                1     86.66 4078.8 382.92
<none>                           4165.5 383.09
```

```
+ Spin_Pace          1      7.41 4158.1 384.90
+ Right_Left_Arm     1      5.76 4159.7 384.94

Step:  AIC=243
Wickets ~ Dot_Balls

                    Df Sum of Sq    RSS    AIC
+ Gender             1     99.59  948.9 234.72
+ Insta_Followers    1     63.51  985.0 238.56
+ Right_Left_Arm     1     26.64 1021.9 242.35
<none>                           1048.5 243.00
+ Twitter_Followers  1     19.19 1029.3 243.10
+ Innings            1      9.08 1039.4 244.10
+ Spin_Pace          1      5.76 1042.7 244.43
+ Age                1      5.65 1042.8 244.44
+ Economy_Rate       1      4.22 1044.3 244.58
+ Auction_Price      1      0.09 1048.4 244.99
- Dot_Balls          1   3117.00 4165.5 383.09

Step:  AIC=234.72
Wickets ~ Dot_Balls + Gender

                    Df Sum of Sq     RSS    AIC
+ Innings            1     49.70  899.21 231.18
+ Insta_Followers    1     47.29  901.61 231.45
+ Right_Left_Arm     1     25.21  923.69 233.95
+ Auction_Price      1     23.67  925.24 234.12
<none>                            948.90 234.72
+ Spin_Pace          1     16.01  932.89 234.97
+ Twitter_Followers  1      7.86  941.04 235.86
+ Economy_Rate       1      2.03  946.88 236.50
+ Age                1      0.27  948.64 236.69
+ Dot_Balls:Gender   1      0.00  948.90 236.72
- Gender             1     99.59 1048.50 243.00
- Dot_Balls          1   1782.18 2731.08 341.61

Step:  AIC=231.18
Wickets ~ Dot_Balls + Gender + Innings

                    Df Sum of Sq     RSS    AIC
+ Insta_Followers    1     52.79  846.41 226.95
+ Spin_Pace          1     37.95  861.25 228.74
+ Right_Left_Arm     1     19.21  880.00 230.96
<none>                            899.21 231.18
+ Auction_Price      1      9.46  889.75 232.09
+ Twitter_Followers  1      9.36  889.85 232.10
+ Dot_Balls:Innings  1      3.95  895.25 232.72
+ Gender:Innings     1      3.01  896.19 232.83
+ Economy_Rate       1      1.14  898.07 233.05
```

```
+ Age                1      0.04  899.17 233.17
+ Dot_Balls:Gender   1      0.02  899.18 233.18
- Innings            1     49.70  948.90 234.72
- Gender             1    140.22 1039.42 244.10
- Dot_Balls          1    856.19 1755.39 298.08

Step:  AIC=226.95
Wickets ~ Dot_Balls + Gender + Innings + Insta_Followers

                            Df Sum of Sq     RSS    AIC
+ Spin_Pace                  1     31.41  815.00 225.05
+ Twitter_Followers          1     28.44  817.97 225.43
+ Auction_Price              1     27.32  819.10 225.57
+ Gender:Innings             1     20.25  826.16 226.45
<none>                                    846.41 226.95
+ Right_Left_Arm             1     11.53  834.89 227.53
+ Innings:Insta_Followers    1     10.17  836.24 227.70
+ Dot_Balls:Gender           1      8.34  838.07 227.93
+ Gender:Insta_Followers     1      3.73  842.69 228.49
+ Dot_Balls:Insta_Followers  1      3.48  842.93 228.52
+ Dot_Balls:Innings          1      3.38  843.03 228.53
+ Economy_Rate               1      2.78  843.64 228.61
+ Age                        1      1.49  844.92 228.76
- Insta_Followers            1     52.79  899.21 231.18
- Innings                    1     55.20  901.61 231.45
- Gender                     1    124.52  970.93 239.08
- Dot_Balls                  1    759.98 1606.39 290.94

Step:  AIC=225.05
Wickets ~ Dot_Balls + Gender + Innings + Insta_Followers + Spin_Pace

                            Df Sum of Sq     RSS    AIC
+ Spin_Pace:Dot_Balls        1     44.96  770.04 221.21
+ Spin_Pace:Insta_Followers  1     36.89  778.11 222.28
+ Twitter_Followers          1     32.91  782.09 222.81
+ Spin_Pace:Innings          1     30.57  784.43 223.11
+ Gender:Innings             1     22.48  792.52 224.17
+ Auction_Price              1     18.84  796.16 224.64
<none>                                    815.00 225.05
+ Right_Left_Arm             1     14.44  800.56 225.21
+ Dot_Balls:Gender           1      9.42  805.58 225.85
+ Innings:Insta_Followers    1      9.05  805.95 225.90
+ Dot_Balls:Innings          1      5.54  809.46 226.35
+ Gender:Insta_Followers     1      3.60  811.40 226.59
+ Dot_Balls:Insta_Followers  1      2.53  812.47 226.73
+ Age                        1      1.49  813.51 226.86
- Spin_Pace                  1     31.41  846.41 226.95
+ Spin_Pace:Gender           1      0.27  814.73 227.02
+ Economy_Rate               1      0.14  814.86 227.03
```

```
- Insta_Followers            1     46.25  861.25 228.74
- Innings                    1     75.10  890.10 232.13
- Gender                     1    149.16  964.16 240.36
- Dot_Balls                  1    790.15 1605.15 292.86

Step:  AIC=221.21
Wickets ~ Dot_Balls + Gender + Innings + Insta_Followers + Spin_Pace + 
    Dot_Balls:Spin_Pace

                            Df Sum of Sq    RSS    AIC
+ Twitter_Followers          1    33.596 736.45 218.61
+ Spin_Pace:Gender           1    21.839 748.21 220.24
+ Auction_Price              1    19.509 750.54 220.56
+ Right_Left_Arm             1    16.710 753.33 220.95
<none>                                   770.04 221.21
+ Gender:Innings             1    11.584 758.46 221.65
+ Spin_Pace:Insta_Followers  1    11.203 758.84 221.70
+ Innings:Insta_Followers    1     9.081 760.96 221.99
+ Spin_Pace:Innings          1     3.720 766.32 222.71
+ Dot_Balls:Gender           1     2.205 767.84 222.91
+ Gender:Insta_Followers     1     1.519 768.52 223.00
+ Dot_Balls:Innings          1     0.763 769.28 223.10
+ Economy_Rate               1     0.294 769.75 223.17
+ Age                        1     0.104 769.94 223.19
+ Dot_Balls:Insta_Followers  1     0.003 770.04 223.21
- Dot_Balls:Spin_Pace        1    44.955 815.00 225.05
- Insta_Followers            1    51.916 821.96 225.93
- Innings                    1   101.375 871.42 231.95
- Gender                     1   154.402 924.45 238.03

Step:  AIC=218.61
Wickets ~ Dot_Balls + Gender + Innings + Insta_Followers + Spin_Pace + 
    Twitter_Followers + Dot_Balls:Spin_Pace

                                    Df Sum of Sq    RSS    AIC
+ Auction_Price                      1    42.340 694.11 214.51
+ Twitter_Followers:Insta_Followers  1    34.470 701.98 215.68
+ Spin_Pace:Gender                   1    16.899 719.55 218.22
+ Gender:Innings                     1    16.480 719.97 218.28
<none>                                           736.45 218.61
+ Right_Left_Arm                     1    12.254 724.19 218.88
+ Spin_Pace:Insta_Followers          1     7.272 729.18 219.59
+ Gender:Insta_Followers             1     6.079 730.37 219.76
+ Innings:Insta_Followers            1     5.579 730.87 219.83
+ Dot_Balls:Gender                   1     5.369 731.08 219.86
+ Twitter_Followers:Gender           1     4.033 732.41 220.05
+ Dot_Balls:Innings                  1     2.536 733.91 220.26
+ Spin_Pace:Innings                  1     2.348 734.10 220.28
+ Age                                1     1.501 734.95 220.40
```

```
+ Twitter_Followers:Dot_Balls        1     0.865 735.58 220.49
+ Dot_Balls:Insta_Followers          1     0.164 736.28 220.59
+ Economy_Rate                       1     0.092 736.36 220.60
+ Spin_Pace:Twitter_Followers        1     0.060 736.39 220.60
+ Twitter_Followers:Innings          1     0.005 736.44 220.61
- Twitter_Followers                  1    33.596 770.04 221.21
- Dot_Balls:Spin_Pace                1    45.647 782.09 222.81
- Insta_Followers                    1    78.082 814.53 226.99
- Innings                            1   105.376 841.82 230.39
- Gender                             1   171.266 907.71 238.15

Step:  AIC=214.51
Wickets ~ Dot_Balls + Gender + Innings + Insta_Followers + Spin_Pace + 
    Twitter_Followers + Auction_Price + Dot_Balls:Spin_Pace

                                    Df Sum of Sq    RSS    AIC
+ Spin_Pace:Gender                   1    27.367 666.74 212.37
+ Auction_Price:Insta_Followers      1    19.718 674.39 213.54
+ Twitter_Followers:Insta_Followers  1    18.320 675.79 213.76
+ Gender:Innings                     1    16.593 677.51 214.02
<none>                                           694.11 214.51
+ Right_Left_Arm                     1    12.427 681.68 214.65
+ Spin_Pace:Innings                  1    12.156 681.95 214.69
+ Gender:Insta_Followers             1    10.310 683.80 214.97
+ Auction_Price:Dot_Balls            1     9.573 684.53 215.08
+ Dot_Balls:Gender                   1     7.927 686.18 215.33
+ Twitter_Followers:Auction_Price    1     6.814 687.29 215.50
+ Twitter_Followers:Gender           1     5.107 689.00 215.75
+ Dot_Balls:Innings                  1     4.049 690.06 215.91
+ Spin_Pace:Insta_Followers          1     3.861 690.25 215.94
+ Age                                1     3.624 690.48 215.97
+ Innings:Insta_Followers            1     3.381 690.73 216.01
+ Auction_Price:Innings              1     2.713 691.39 216.11
+ Twitter_Followers:Dot_Balls        1     2.347 691.76 216.16
+ Spin_Pace:Auction_Price            1     1.687 692.42 216.26
+ Dot_Balls:Insta_Followers          1     1.152 692.95 216.34
+ Twitter_Followers:Innings          1     0.586 693.52 216.43
+ Auction_Price:Gender               1     0.540 693.57 216.43
+ Spin_Pace:Twitter_Followers        1     0.018 694.09 216.51
+ Economy_Rate                       1     0.008 694.10 216.51
- Auction_Price                      1    42.340 736.45 218.61
- Dot_Balls:Spin_Pace                1    46.944 741.05 219.25
- Twitter_Followers                  1    56.428 750.54 220.56
- Innings                            1    62.498 756.61 221.39
- Insta_Followers                    1   114.734 808.84 228.27
- Gender                             1   209.446 903.55 239.68

Step:  AIC=212.37
Wickets ~ Dot_Balls + Gender + Innings + Insta_Followers + Spin_Pace + 
    Twitter_Followers + Auction_Price + Dot_Balls:Spin_Pace + 
```

```
    Gender:Spin_Pace

                                    Df Sum of Sq    RSS    AIC
+ Auction_Price:Insta_Followers      1    19.108 647.63 211.38
+ Twitter_Followers:Insta_Followers  1    17.784 648.96 211.59
+ Right_Left_Arm                     1    15.342 651.40 211.97
<none>                                           666.74 212.37
+ Spin_Pace:Auction_Price            1    12.426 654.31 212.43
+ Gender:Insta_Followers             1    10.028 656.71 212.81
+ Gender:Innings                     1     8.513 658.23 213.05
+ Twitter_Followers:Auction_Price    1     7.180 659.56 213.25
+ Spin_Pace:Insta_Followers          1     6.634 660.11 213.34
+ Innings:Insta_Followers            1     5.637 661.10 213.50
+ Age                                1     5.171 661.57 213.57
+ Auction_Price:Innings              1     4.693 662.05 213.64
+ Auction_Price:Dot_Balls            1     4.666 662.07 213.65
+ Dot_Balls:Gender                   1     4.521 662.22 213.67
+ Twitter_Followers:Gender           1     3.521 663.22 213.82
+ Twitter_Followers:Dot_Balls        1     2.007 664.73 214.06
+ Dot_Balls:Innings                  1     1.768 664.97 214.10
+ Spin_Pace:Twitter_Followers        1     0.841 665.90 214.24
+ Spin_Pace:Innings                  1     0.758 665.98 214.25
+ Economy_Rate                       1     0.565 666.17 214.28
+ Auction_Price:Gender               1     0.452 666.29 214.30
+ Dot_Balls:Insta_Followers          1     0.430 666.31 214.30
+ Twitter_Followers:Innings          1     0.113 666.63 214.35
- Gender:Spin_Pace                   1    27.367 694.11 214.51
- Innings                            1    50.648 717.39 217.91
- Twitter_Followers                  1    52.651 719.39 218.20
- Auction_Price                      1    52.809 719.55 218.22
- Dot_Balls:Spin_Pace                1    73.603 740.34 221.16
- Insta_Followers                    1   119.235 785.97 227.32

Step:  AIC=211.38
Wickets ~ Dot_Balls + Gender + Innings + Insta_Followers + Spin_Pace +
    Twitter_Followers + Auction_Price + Dot_Balls:Spin_Pace +
    Gender:Spin_Pace + Insta_Followers:Auction_Price

                                    Df Sum of Sq    RSS    AIC
+ Twitter_Followers:Insta_Followers  1    26.164 621.47 209.13
+ Right_Left_Arm                     1    17.215 630.42 210.60
+ Auction_Price:Dot_Balls            1    14.576 633.06 211.03
<none>                                           647.63 211.38
+ Gender:Innings                     1     9.003 638.63 211.93
- Insta_Followers:Auction_Price      1    19.108 666.74 212.37
+ Age                                1     5.621 642.01 212.48
+ Spin_Pace:Auction_Price            1     5.476 642.16 212.50
+ Dot_Balls:Gender                   1     3.930 643.70 212.75
- Twitter_Followers                  1    22.413 670.04 212.88
+ Gender:Insta_Followers             1     2.412 645.22 212.99
```

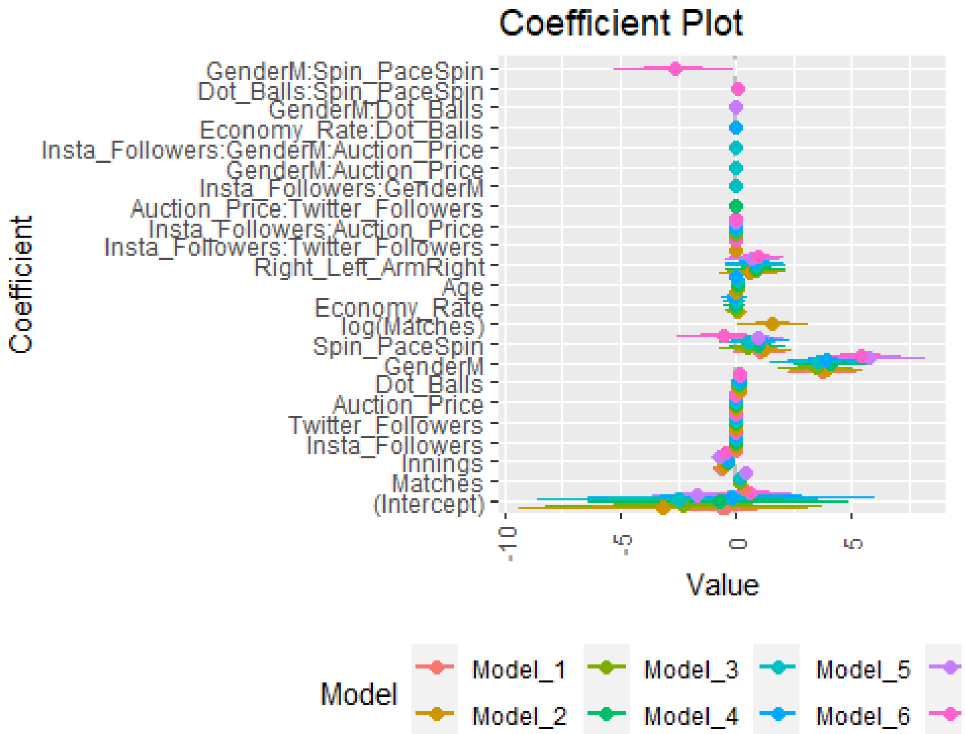
```
+ Dot_Balls:Innings                 1     2.216 645.42 213.02
+ Dot_Balls:Insta_Followers         1     2.034 645.60 213.05
+ Spin_Pace:Insta_Followers         1     1.951 645.68 213.06
+ Auction_Price:Gender              1     1.371 646.26 213.16
+ Twitter_Followers:Dot_Balls       1     1.224 646.41 213.18
+ Twitter_Followers:Gender          1     0.782 646.85 213.25
+ Auction_Price:Innings             1     0.535 647.10 213.29
+ Twitter_Followers:Auction_Price   1     0.415 647.22 213.31
+ Spin_Pace:Innings                 1     0.405 647.23 213.31
+ Economy_Rate                      1     0.395 647.24 213.31
+ Spin_Pace:Twitter_Followers       1     0.337 647.29 213.32
+ Innings:Insta_Followers           1     0.083 647.55 213.36
+ Twitter_Followers:Innings         1     0.057 647.57 213.37
- Gender:Spin_Pace                  1    26.757 674.39 213.54
- Innings                           1    47.744 695.38 216.70
- Dot_Balls:Spin_Pace               1    54.200 701.83 217.65

Step:  AIC=209.13
Wickets ~ Dot_Balls + Gender + Innings + Insta_Followers + Spin_Pace +
    Twitter_Followers + Auction_Price + Dot_Balls:Spin_Pace +
    Gender:Spin_Pace + Insta_Followers:Auction_Price +
Insta_Followers:Twitter_Followers

                                   Df Sum of Sq    RSS    AIC
+ Right_Left_Arm                    1    16.167 605.30 208.41
+ Auction_Price:Dot_Balls           1    12.241 609.23 209.08
+ Twitter_Followers:Innings         1    12.044 609.42 209.11
<none>                                          621.47 209.13
+ Twitter_Followers:Dot_Balls       1    10.148 611.32 209.43
+ Innings:Insta_Followers           1    10.010 611.46 209.46
+ Gender:Innings                    1     9.610 611.86 209.52
+ Dot_Balls:Insta_Followers         1     9.088 612.38 209.61
+ Twitter_Followers:Auction_Price   1     8.280 613.19 209.75
+ Age                               1     6.554 614.91 210.03
+ Spin_Pace:Auction_Price           1     3.912 617.55 210.48
+ Dot_Balls:Gender                  1     3.184 618.28 210.60
+ Dot_Balls:Innings                 1     2.434 619.03 210.72
+ Auction_Price:Gender              1     1.695 619.77 210.85
+ Gender:Insta_Followers            1     0.881 620.59 210.98
+ Economy_Rate                      1     0.593 620.87 211.03
+ Twitter_Followers:Gender          1     0.442 621.03 211.05
+ Auction_Price:Innings             1     0.311 621.16 211.08
+ Spin_Pace:Innings                 1     0.097 621.37 211.11
+ Spin_Pace:Insta_Followers         1     0.034 621.43 211.12
+ Spin_Pace:Twitter_Followers       1     0.031 621.44 211.12
- Gender:Spin_Pace                  1    25.973 647.44 211.34
- Insta_Followers:Twitter_Followers 1    26.164 647.63 211.38
- Insta_Followers:Auction_Price     1    27.488 648.96 211.59
- Dot_Balls:Spin_Pace               1    49.017 670.48 214.95
- Innings                           1    51.717 673.18 215.36
```

```
Step:  AIC=208.41
Wickets ~ Dot_Balls + Gender + Innings + Insta_Followers + Spin_Pace + 
    Twitter_Followers + Auction_Price + Right_Left_Arm + Dot_Balls:Spin_Pace 
+ 
    Gender:Spin_Pace + Insta_Followers:Auction_Price + 
Insta_Followers:Twitter_Followers

                                    Df Sum of Sq    RSS    AIC
<none>                                           605.30 208.41
+ Auction_Price:Dot_Balls            1    10.241 595.06 208.66
+ Twitter_Followers:Dot_Balls        1     8.954 596.35 208.88
+ Gender:Innings                     1     8.530 596.77 208.95
+ Dot_Balls:Insta_Followers          1     8.410 596.89 208.97
+ Innings:Insta_Followers            1     8.379 596.92 208.98
+ Twitter_Followers:Innings          1     7.939 597.36 209.05
- Right_Left_Arm                     1    16.167 621.47 209.13
+ Age                                1     6.347 598.95 209.33
+ Right_Left_Arm:Dot_Balls           1     6.297 599.00 209.34
+ Twitter_Followers:Right_Left_Arm   1     5.623 599.68 209.45
+ Twitter_Followers:Auction_Price    1     5.124 600.18 209.54
+ Right_Left_Arm:Gender              1     2.815 602.48 209.93
+ Dot_Balls:Innings                  1     2.782 602.52 209.94
+ Spin_Pace:Auction_Price            1     2.625 602.67 209.97
+ Dot_Balls:Gender                   1     2.368 602.93 210.01
+ Right_Left_Arm:Insta_Followers     1     1.452 603.85 210.16
+ Right_Left_Arm:Auction_Price       1     1.345 603.95 210.18
+ Right_Left_Arm:Innings             1     1.052 604.25 210.23
+ Auction_Price:Gender               1     0.568 604.73 210.32
+ Gender:Insta_Followers             1     0.431 604.87 210.34
+ Auction_Price:Innings              1     0.397 604.90 210.34
+ Spin_Pace:Twitter_Followers        1     0.173 605.13 210.38
+ Spin_Pace:Insta_Followers          1     0.124 605.18 210.39
+ Spin_Pace:Innings                  1     0.111 605.19 210.39
+ Spin_Pace:Right_Left_Arm           1     0.097 605.20 210.40
+ Twitter_Followers:Gender           1     0.002 605.30 210.41
+ Economy_Rate                       1     0.000 605.30 210.41
- Insta_Followers:Twitter_Followers  1    25.117 630.42 210.60
- Gender:Spin_Pace                   1    28.891 634.19 211.22
- Insta_Followers:Auction_Price      1    29.414 634.71 211.30
- Innings                            1    46.784 652.08 214.08
- Dot_Balls:Spin_Pace                1    52.372 657.67 214.96
```

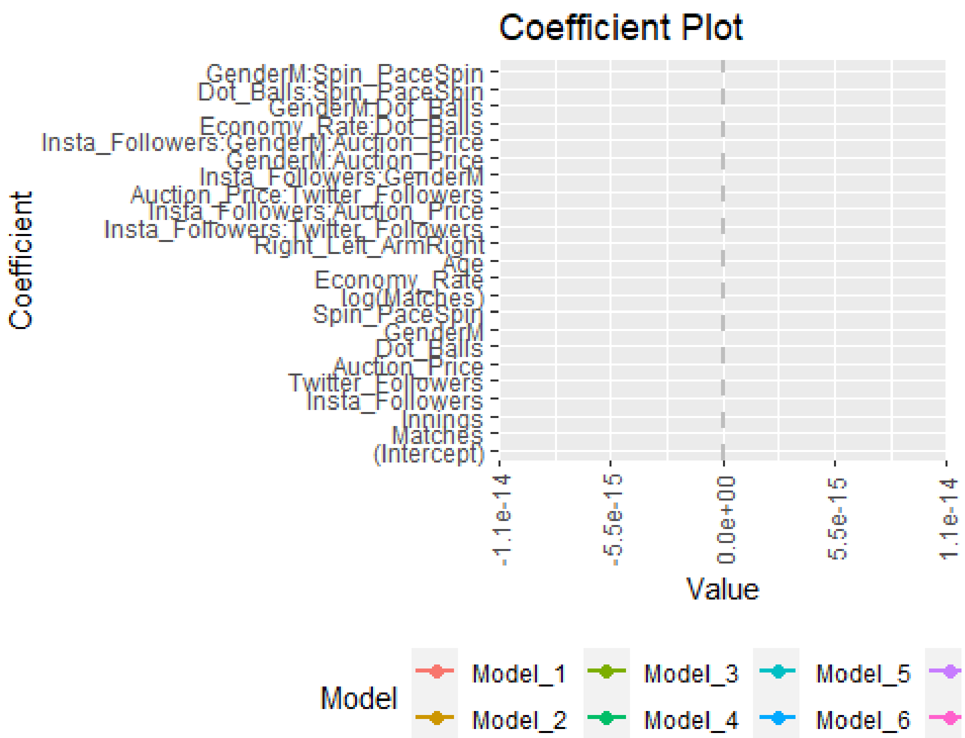
```
summary(WicketsStep)
```

```
Call:
lm(formula = Wickets ~ Dot_Balls + Gender + Innings + Insta_Followers + 
    Spin_Pace + Twitter_Followers + Auction_Price + Right_Left_Arm + 
    Dot_Balls:Spin_Pace + Gender:Spin_Pace + Insta_Followers:Auction_Price + 
```

```
    Insta_Followers:Twitter_Followers, data = Data)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.4996 -1.3658 -0.0243  1.4448  9.2545 

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)                        5.868e-01  9.257e-01   0.634  0.52772    
Dot_Balls                          1.343e-01  1.616e-02   8.308 9.43e-13 ***
GenderM                            5.393e+00  8.869e-01   6.081 2.84e-08 ***
Innings                           -4.136e-01  1.568e-01  -2.637  0.00984 ** 
Insta_Followers                   -3.801e-07  8.159e-07  -0.466  0.64244    
Spin_PaceSpin                     -5.104e-01  1.002e+00  -0.509  0.61177    
Twitter_Followers                 -1.895e-06  9.052e-07  -2.094  0.03912 *  
Auction_Price                     -2.878e-08  9.857e-09  -2.920  0.00442 ** 
Right_Left_ArmRight                9.166e-01  5.912e-01   1.550  0.12454    
Dot_Balls:Spin_PaceSpin            4.621e-02  1.656e-02   2.791  0.00642 ** 
GenderM:Spin_PaceSpin             -2.661e+00  1.284e+00  -2.073  0.04107 *  
Insta_Followers:Auction_Price      1.028e-14  4.914e-15   2.091  0.03932 *  
Insta_Followers:Twitter_Followers  3.378e-13  1.748e-13   1.932  0.05644 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.593 on 90 degrees of freedom
Multiple R-squared:  0.8547,    Adjusted R-squared:  0.8353 
F-statistic: 44.11 on 12 and 90 DF,  p-value: < 2.2e-16

multiplot(Model_1, Model_2, Model_3, Model_4,
Model_5,Model_6,Model_7,WicketsStep,pointSize=2)
```

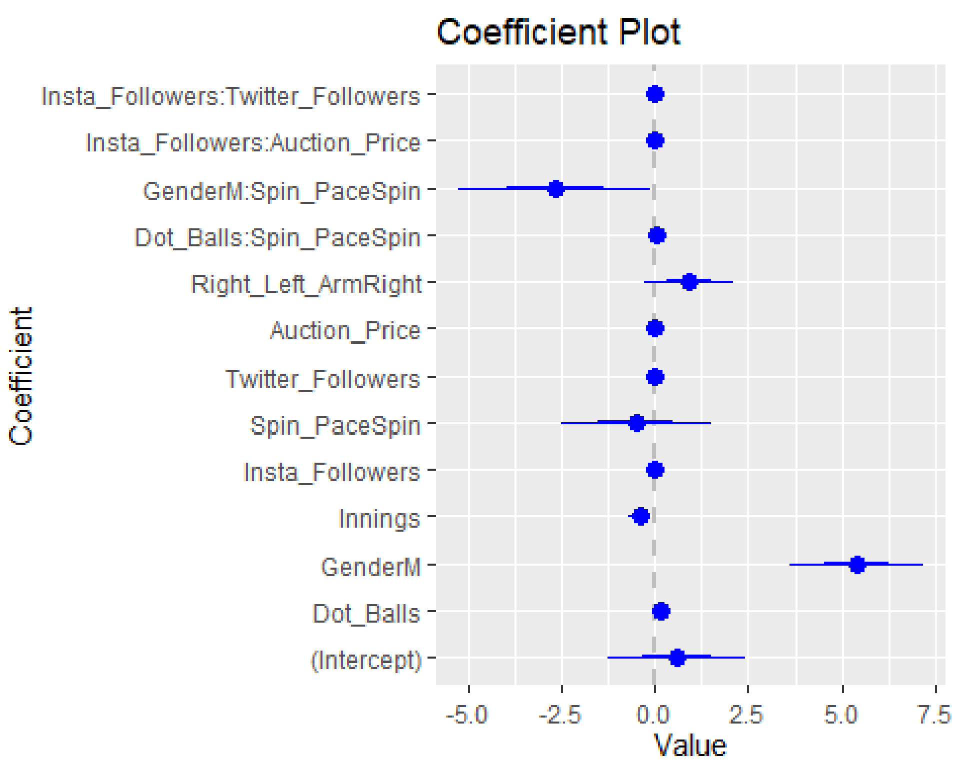



```
multiplot(Model_1, Model_2, Model_3, Model_4,
Model_5,Model_6,Model_7,WicketsStep, pointSize=2) +
scale_x_continuous(limits=c(-.00000000000001, .00000000000001))

Warning: Removed 97 rows containing missing values (`geom_point()`).
```

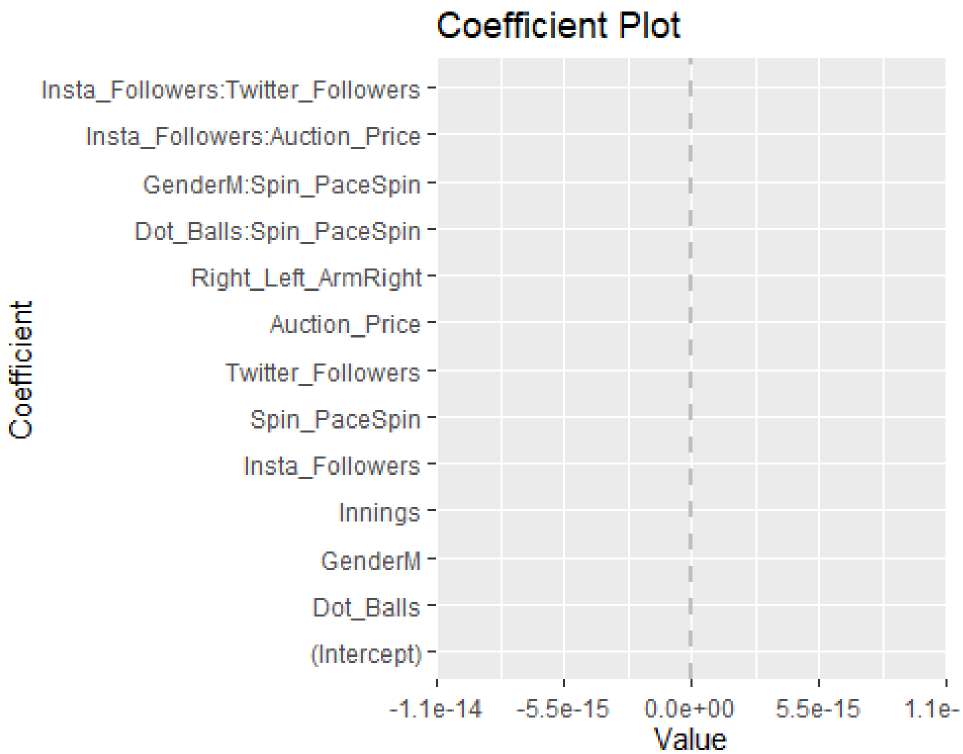



```
anova(Model_1, Model_2, Model_3, Model_4,
Model_5,Model_6,Model_7,WicketsStep)

Analysis of Variance Table

Model 1: Wickets ~ Matches + Innings + Insta_Followers + Twitter_Followers +
    Auction_Price + Dot_Balls + Gender + Spin_Pace
Model 2: Wickets ~ log(Matches) + Innings + Economy_Rate + Age +
Insta_Followers *
    Twitter_Followers + Auction_Price + Dot_Balls + Gender +
    Right_Left_Arm + Spin_Pace
Model 3: Wickets ~ Matches + Economy_Rate + Age + Insta_Followers *
Auction_Price +
    Dot_Balls + Gender + Right_Left_Arm + Spin_Pace + Twitter_Followers
Model 4: Wickets ~ Economy_Rate + Innings + Age + Insta_Followers +
Auction_Price *
    Twitter_Followers + Dot_Balls + Right_Left_Arm + Spin_Pace +
    Gender
Model 5: Wickets ~ Matches + Economy_Rate + Age + Insta_Followers * Gender *
    Auction_Price + Twitter_Followers + Dot_Balls + Right_Left_Arm +
    Spin_Pace
Model 6: Wickets ~ Economy_Rate * Dot_Balls + Innings + Age + Insta_Followers
*
    Auction_Price + Twitter_Followers + Gender + Right_Left_Arm +
    Spin_Pace
```

```
Model 7: Wickets ~ Matches + Innings + Gender * Dot_Balls +
Insta_Followers:Auction_Price +
    Auction_Price + Spin_Pace + Right_Left_Arm
Model 8: Wickets ~ Dot_Balls + Gender + Innings + Insta_Followers + Spin_Pace
+
    Twitter_Followers + Auction_Price + Right_Left_Arm + Dot_Balls:Spin_Pace
+
    Gender:Spin_Pace + Insta_Followers:Auction_Price +
Insta_Followers:Twitter_Followers
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     94 694.67
2     90 677.92  4    16.756 0.5208 0.72062
3     91 719.66 -1   -41.743 5.1902 0.02514 *
4     91 702.70  0    16.963
5     88 707.75  3    -5.047
6     90 688.39 -2    19.360
7     93 639.36 -3    49.022
8     90 605.30  3    34.065 1.4118 0.24469
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#WicketsStep has one of the least residual sum of squares and all variables
are significant, so we proceed with model 7
coefplot(WicketsStep)
```

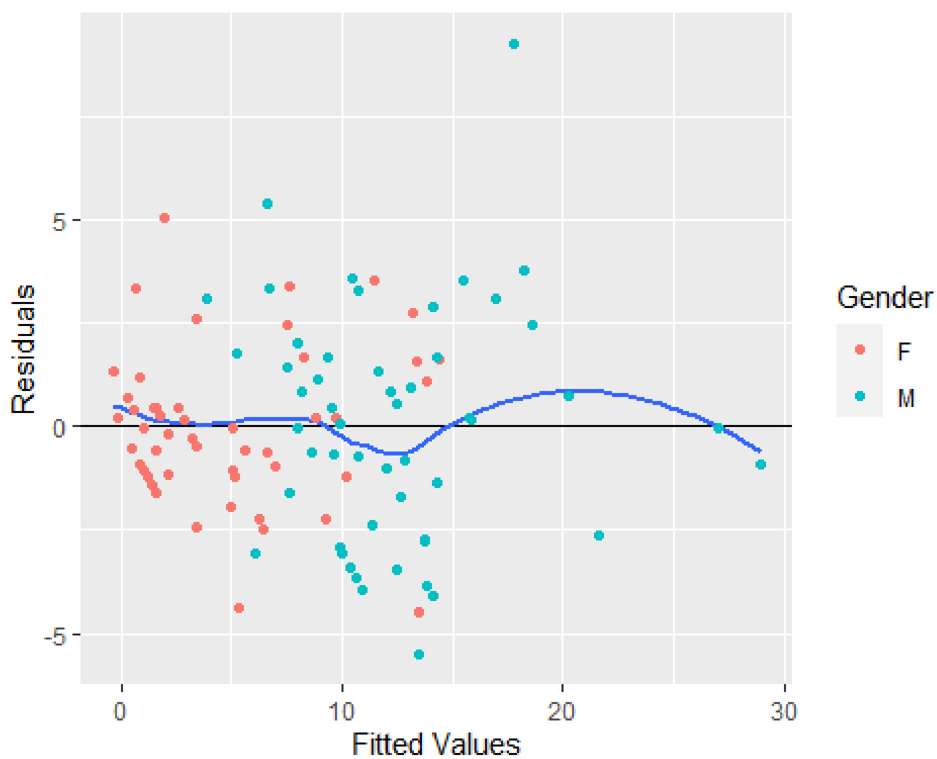


```
coefplot(WicketsStep) + scale_x_continuous(limits=c(-.00000000000001,
.00000000000001))

Warning: Removed 13 rows containing missing values (`geom_point()`).
```



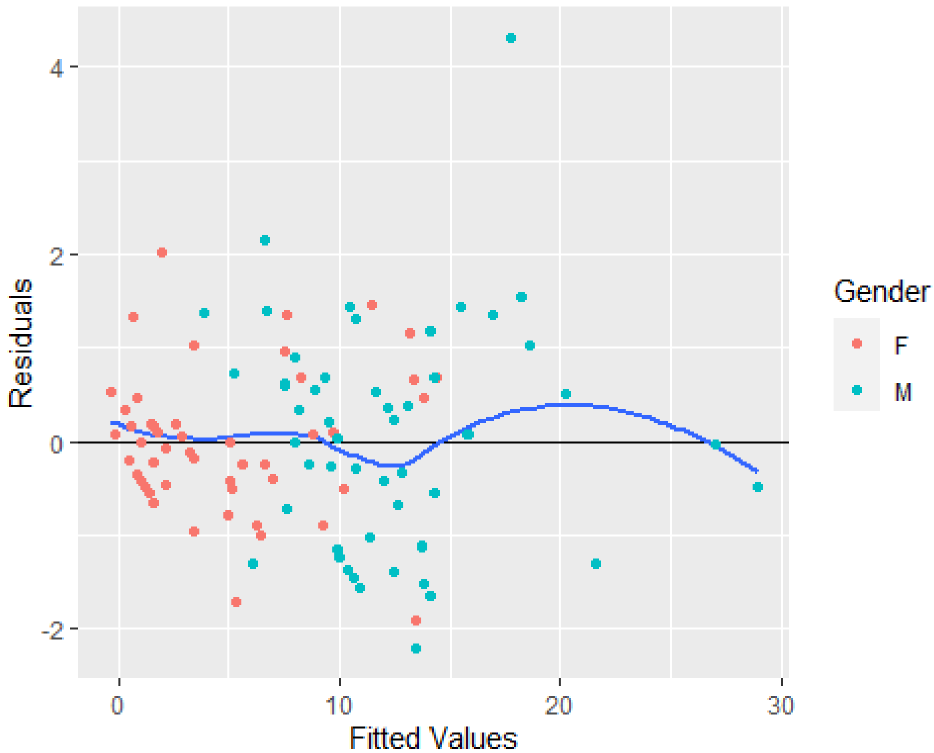

```
head(fortify(WicketsStep))

  Wickets Dot_Balls Gender Innings Insta_Followers Spin_Pace
Twitter_Followers
1       8        40      M       6           28200      Spin
0
2      14        54      M       8          161000      Pace
0
3       6        53      F       8          124000      Spin
12900
4       7        57      M       7           59100      Pace
0
5       0         8      F       3            8974      Pace
0
6      15        96      F      10               0      Spin
2876
  Auction_Price Right_Left_Arm        .hat    .sigma      .cooksd   .fitted
1       1.5e+07          Right 0.12516645 2.607894 1.108406e-06  8.024343
2       2.0e+06          Right 0.07494186 2.582759 1.075982e-02 10.722483
3       7.5e+06          Right 0.05651532 2.605737 6.860931e-04  6.972028
4       2.4e+07          Right 0.05443415 2.572000 1.089571e-02 10.955665
```

```
5        5.0e+06          Right 0.06561477 2.604629 1.216987e-03  1.189882
6        1.0e+07          Right 0.12808906 2.604846 2.377132e-03 13.889347
       .resid   .stdresid
1 -0.02434258 -0.01003552
2  3.27751679  1.31400332
3 -0.97202830 -0.38587579
4 -3.95566505 -1.56859083
5 -1.18988178 -0.47465371
6  1.11065309  0.45864685

#fitted (y hat) values against residuals
W1 <- ggplot(aes(x=.fitted, y=.resid), data = WicketsStep) +  geom_point() +
  geom_hline(yintercept = 0) +    geom_smooth(se = FALSE) +
  labs(x="Fitted Values", y="Residuals")

W1 + geom_point(aes(color=Gender))

`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

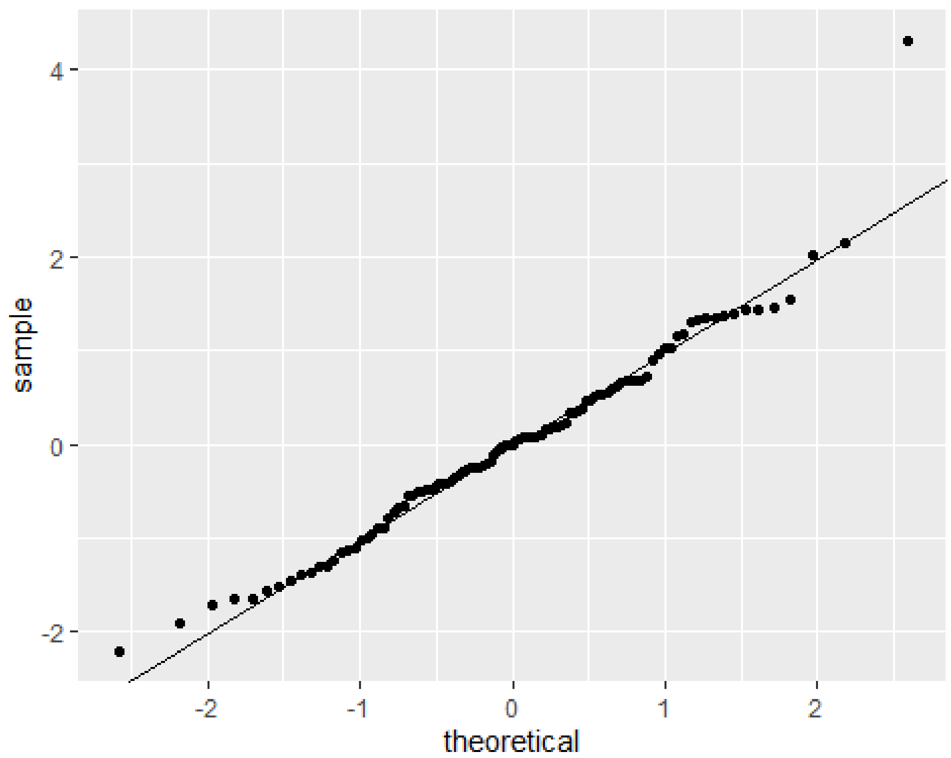



```
W2 <- ggplot(aes(x=.fitted, y=.stdresid), data = WicketsStep) +
geom_point() +
  geom_hline(yintercept = 0) +    geom_smooth(se = FALSE) +
  labs(x="Fitted Values", y="Residuals")

W2 + geom_point(aes(color=Gender))
```

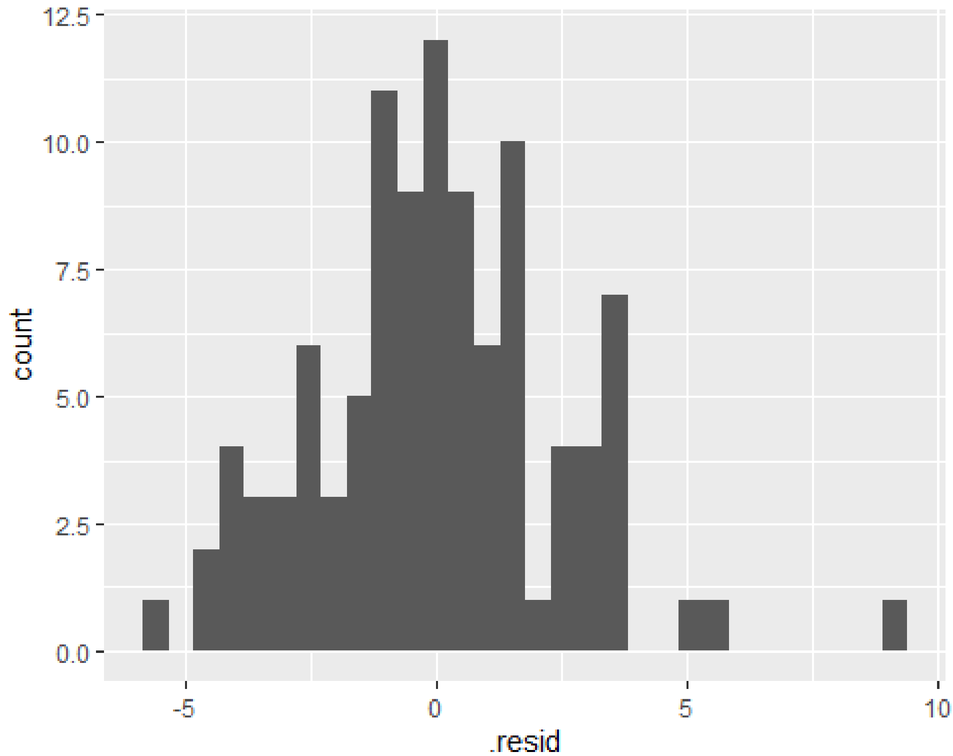
```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



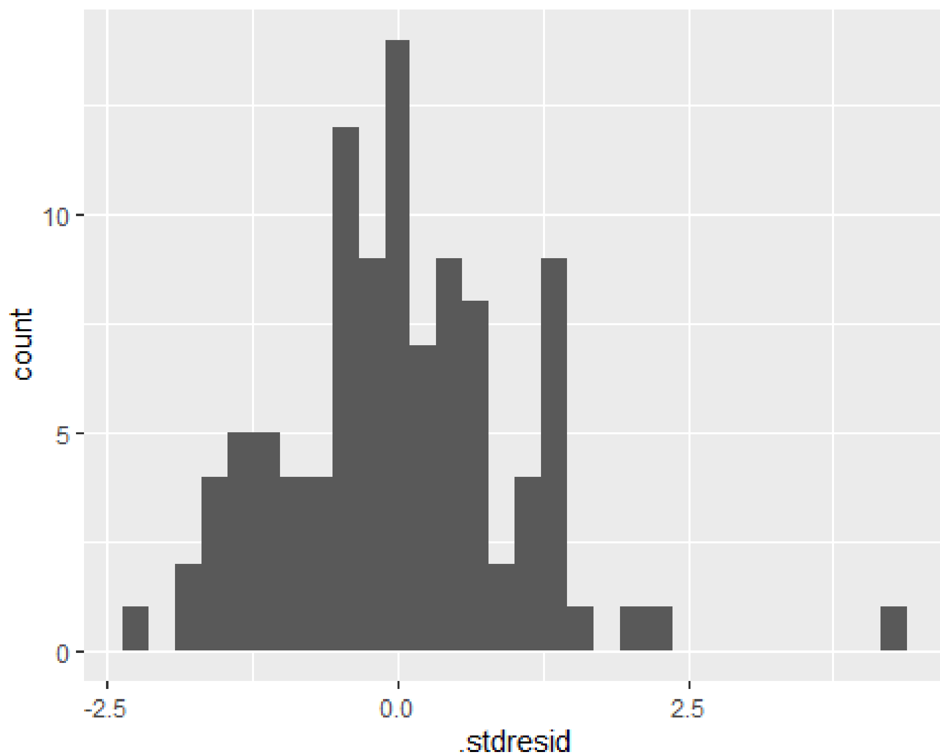

```
#Q-Q plot
ggplot(WicketsStep, aes(sample=.stdresid)) + stat_qq() + geom_abline()
```




```
# a histogram of the residuals.
ggplot(WicketsStep, aes(x=.resid)) + geom_histogram()

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
ggplot(WicketsStep, aes(x=.stdresid)) + geom_histogram()

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




## Insights :

- While performing simple linear regression between each independent variable with the dependent variable we reached the following conclusions -
  - Matches, Auction_Price and Dot_Balls have a strong linear relationship with wickets and obeys homoscedasticity and normality of residuals.
  - Innings and Wickets have a strong linear relationship, however the residuals follow weak normal distribution.
  - For Insta_Followers and Wickets, there is a weak linear relationship, but they obey the conditions of homoscedasticity and normality of residuals.
  - Age, Twitter_Followers and Economy_Rate have a weak linear relationship with Wickets, the residuals show a weak normal distribution and they do not satisfy homoscedasticity, therefore, they might be removed in the final models.
- We have selected WicketsStep as the final model since it shows the least RSS value.

- WicketsStep has an R-squared value of 0.8547 and adjusted R-squared of 0.8353. This means that the predictor variables included in the model explain around 85% of the variation in the number of wickets taken by bowlers.
- The adjusted R-squared value is also very close to the R-squared value, indicating that model has not been overfitted by including extraneous variables.
- The residual standard error is 2.593, which gives an estimate of the standard deviation of the residuals or prediction errors. This value is small relative to the scale of the response variable (Wickets), indicating decent predictive performance.
- Most of the predictor variables included in the final WicketsStep model seem statistically significant based on the p-values.
- Interaction effects between variables like Gender & Dot balls, Spin pace & Dot balls improve the predictive capability and are retained in the final model.

## Conclusion :

In summary, the high R-squared, low prediction error, significant variables and interactions make this a good fit model with good predictive performance as visible from the diagnostic plots as well. This model explains a large portion of variability in the wicket taking capability of bowlers based on important explanatory variables related to skill, experience, bowling style, no. of matches/innings played and the auction price/relevance of the individual players.

## References :


IPL 2023 Stats

IPL Auction Data

Women IPL Stats

Women IPL Auction Data

X (Twitter) Followers

Instagram Followers