<u>R Programming Lab Assignment List</u>

BE Sem VIII Jan-May 2020

1. Installation of R and R Studio, Basics of R [LO1].
2. Datatypes and Data structures in R [LO1, LO2,].
3. Loops and Conditions in R [LO1].
4. EDA on inbuilt R dataset [LO3, LO4].
5. Graphs Plotting using ggplot2 library [LO2, LO5].
6. Regression and Correlation [LO4].
7. Mini-Project Session 1 [LO6].:

   Choose Dataset from Kaggle, extracting data from large dataset
8. Mini-Project Session 2:

   Cleaning of the Dataset
9. Mini-Project Session 3:

   EDA on the dataset
10. Mini-Project Session 4:

    Regression analysis
11. Mini-Project Session 5:

    Data Visualization using ggplot2

<u>Subject In charges:</u> Dr. Shachi Natu Dr. Darshan Ingle.

Anirudh Poroorkara
Roll: 44
IT-1 B12

Lab Assignment: 01

Aim:

Installation of R and RStudio

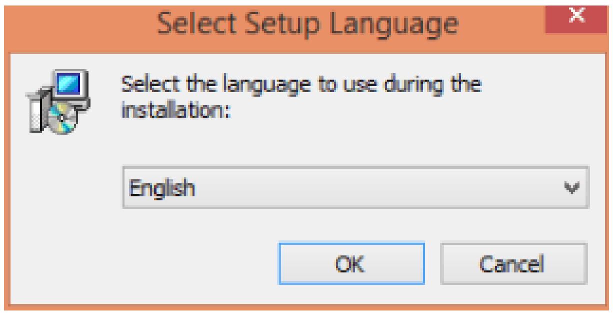
Theory:

R is an open source programming language and software environment for statistical computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis.

A major advantage that R has over many other statistical packages and is that it's free in the sense of free software. R is both flexible and powerful. It has an amazing ecosystem for developers and It has wide range of packages for data access, data cleaning or munging, performing Analysis, creating Reports etc.

Steps to install R on Windows:

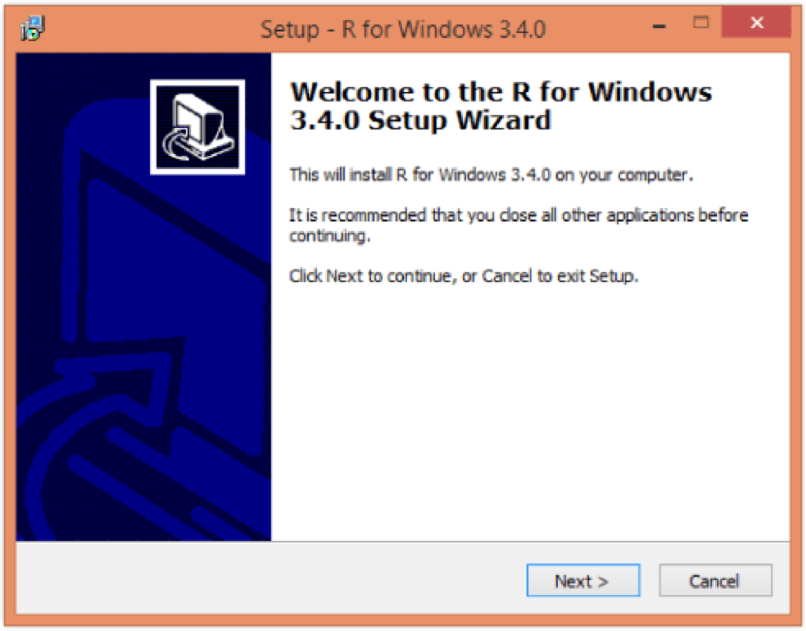
1. Go to http://ftp.heanet.ie/mirrors/cran.r-project.org.
   or https://cran.rstudio.com/bin/windows/base/
2. Under "Download and Install R", click on the "Windows" link.
3. Download the required the .exe file. You should see a link saying something like "Download R 3.4.0 for Windows" (or R X.X.X, where X.X.X gives the version of R, eg. R 3.4.0). Click on that link.
4. After downloading double-click on the R-3.4.0-win.exe to run it.
5. You will be asked what language to install it in – choose English.


Select Setup Language

Select the language to use during the installation:

English

6. The R Setup Wizard will appear in a window. Click "Next" at the bottom of the R setup wizard window.



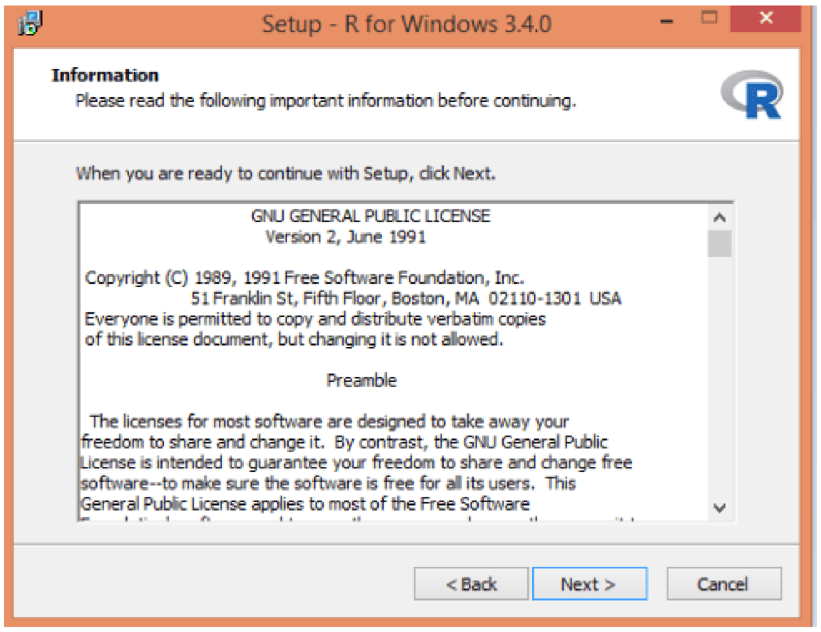
Welcome to the R for Windows 3.4.0 Setup Wizard

This will install R for Windows 3.4.0 on your computer.

It is recommended that you close all other applications before continuing.

Click Next to continue, or Cancel to exit Setup.

7. Click "Next" at the bottom of the R Setup wizard window.



Information

Please read the following important information before continuing.

When you are ready to continue with Setup, click Next.

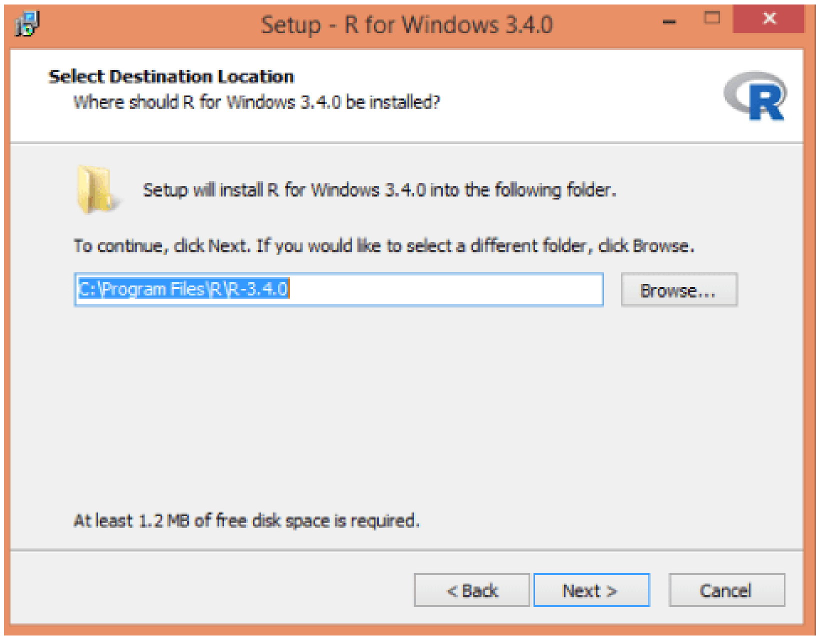
```
                    GNU GENERAL PUBLIC LICENSE
                       Version 2, June 1991

 Copyright (C) 1989, 1991 Free Software Foundation, Inc.
                 51 Franklin St, Fifth Floor, Boston, MA  02110-1301  USA
 Everyone is permitted to copy and distribute verbatim copies
 of this license document, but changing it is not allowed.

                            Preamble

  The licenses for most software are designed to take away your
freedom to share and change it.  By contrast, the GNU General Public
License is intended to guarantee your freedom to share and change free
software--to make sure the software is free for all its users.  This
General Public License applies to most of the Free Software
```

8. The next page says "Select Destination Location" at the top. By default, it will suggest to install R in "C:\Program Files" on your computer.


Setup - R for Windows 3.4.0

Select Destination Location
Where should R for Windows 3.4.0 be installed?

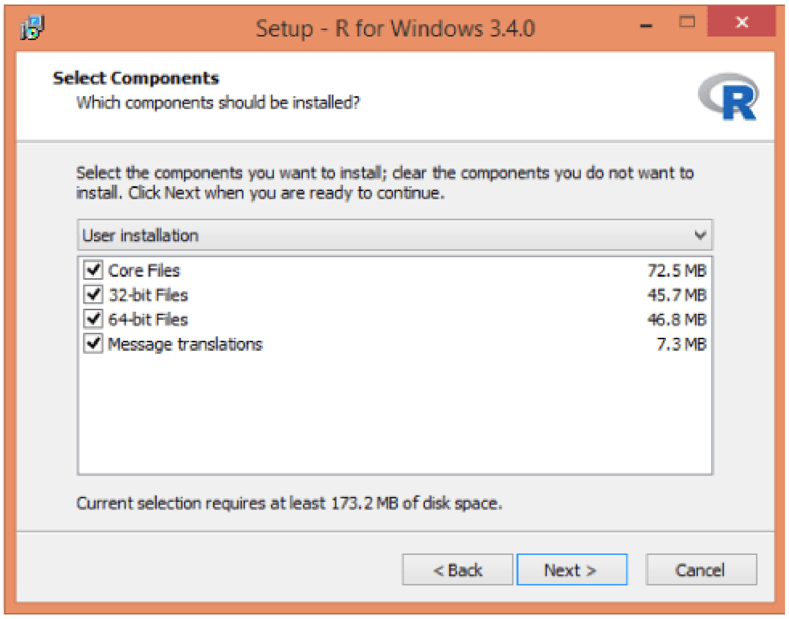
Setup will install R for Windows 3.4.0 into the following folder.

To continue, click Next. If you would like to select a different folder, click Browse.

C:\Program Files\R\R-3.4.0

At least 1.2 MB of free disk space is required.

9. The next page says “Select components” at the top. Click “Next” again.


Setup - R for Windows 3.4.0

Select Components
Which components should be installed?

Select the components you want to install; clear the components you do not want to install. Click Next when you are ready to continue.

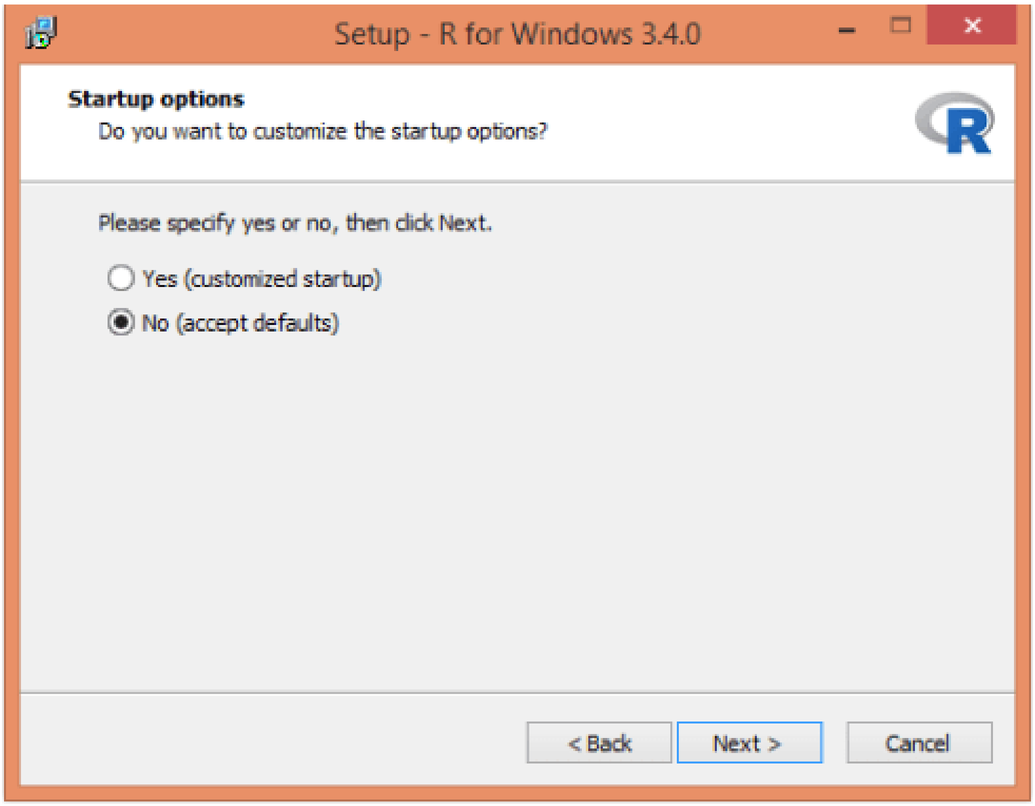
User installation

✓ Core Files	72.5 MB
✓ 32-bit Files	45.7 MB
✓ 64-bit Files	46.8 MB
✓ Message translations	7.3 MB

Current selection requires at least 173.2 MB of disk space.

10. The next page says “Startup options” at the top. Click “Next” again.

Anirudh Poroorkara
Roll: 44
IT-1 B12


Setup - R for Windows 3.4.0

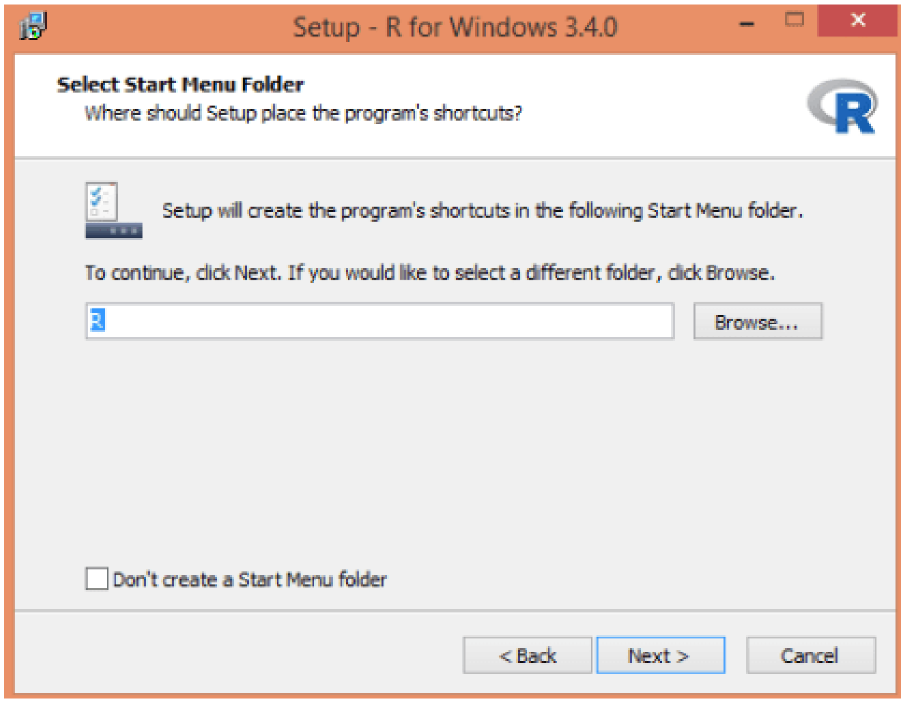
Startup options

Do you want to customize the startup options?

Please specify yes or no, then click Next.

Yes (customized startup) [✓ No (accept defaults)]

11. The next page says “Select start menu folder” at the top. Click “Next” again.


Setup - R for Windows 3.4.0

Select Start Menu Folder

Where should Setup place the program's shortcuts?

Setup will create the program's shortcuts in the following Start Menu folder.

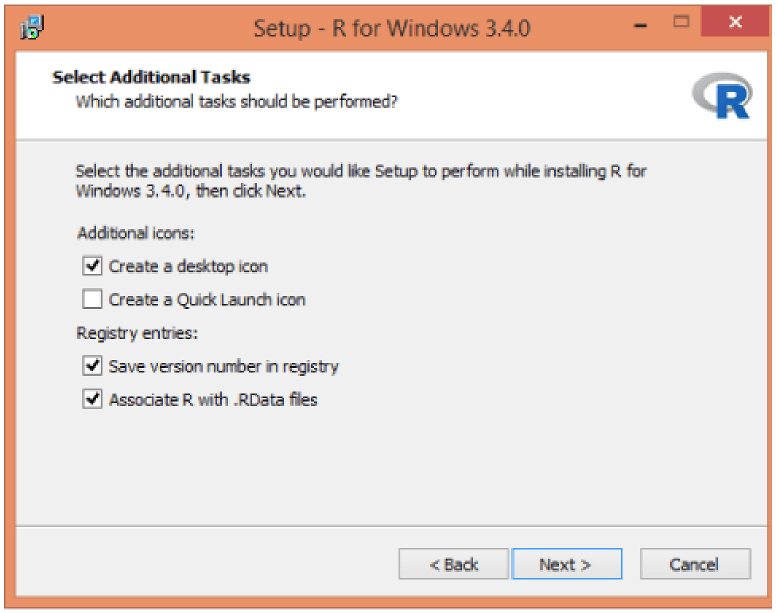
To continue, click Next. If you would like to select a different folder, click Browse.

R

☐ Don't create a Start Menu folder

Anirudh Poroorkara
Roll: 44
IT-1 B12

12. The next page says “Select additional tasks” at the top. Click “Next” again.



Select Additional Tasks
Which additional tasks should be performed?

Select the additional tasks you would like Setup to perform while installing R for Windows 3.4.0, then click Next.

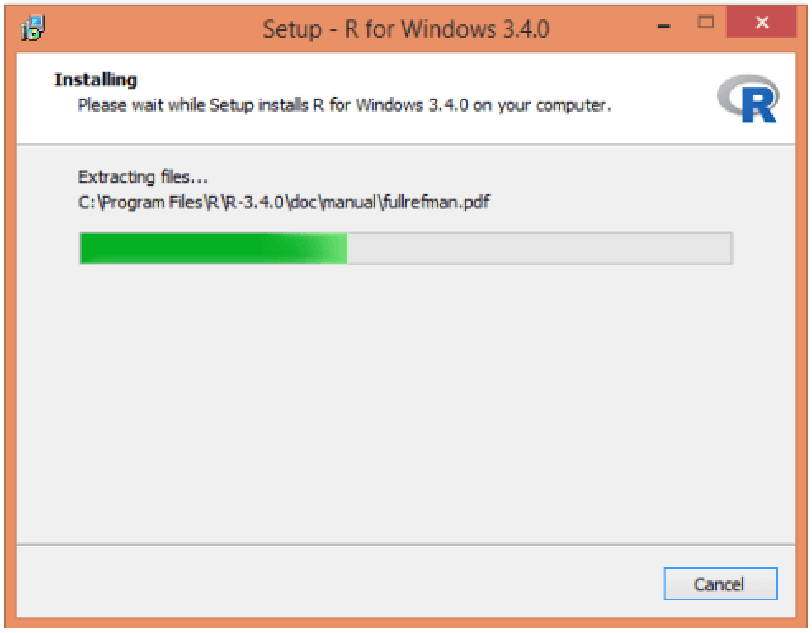
Additional icons:

- [✓] Create a desktop icon
- [ ] Create a Quick Launch icon

Registry entries:

- [✓] Save version number in registry
- [✓] Associate R with .RData files

13. Now it will be installing in your machine. You will see something like this.

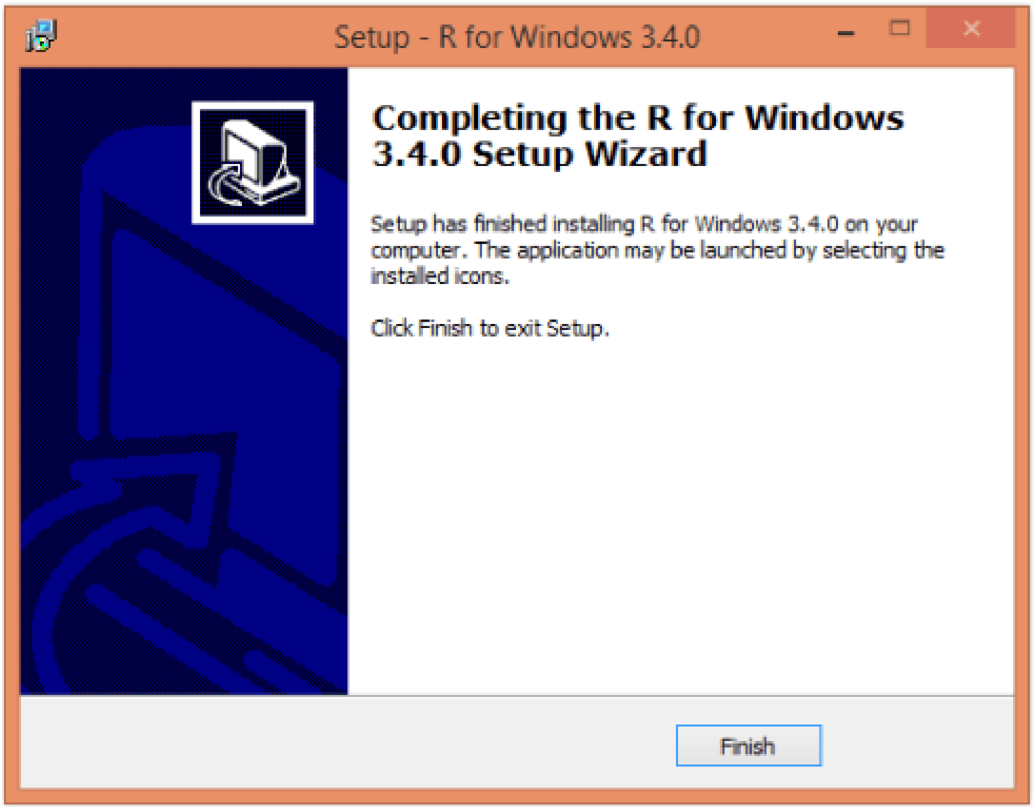


Installing
Please wait while Setup installs R for Windows 3.4.0 on your computer.

Extracting files...
C:\Program Files\R\R-3.4.0\doc\manual\fullrefman.pdf

14. Now you will see “Completing the R for Windows Setup Wizard” appear. Click “Finish”.

Anirudh Poroorkara
Roll: 44
IT-1 B12


Setup - R for Windows 3.4.0

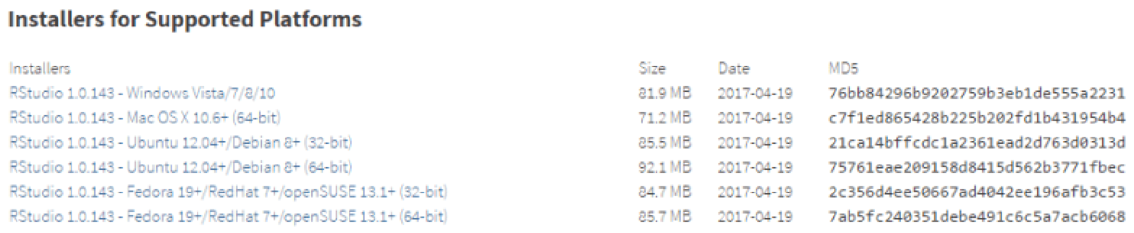
Completing the R for Windows 3.4.0 Setup Wizard

Setup has finished installing R for Windows 3.4.0 on your computer. The application may be launched by selecting the installed icons.

Click Finish to exit Setup.

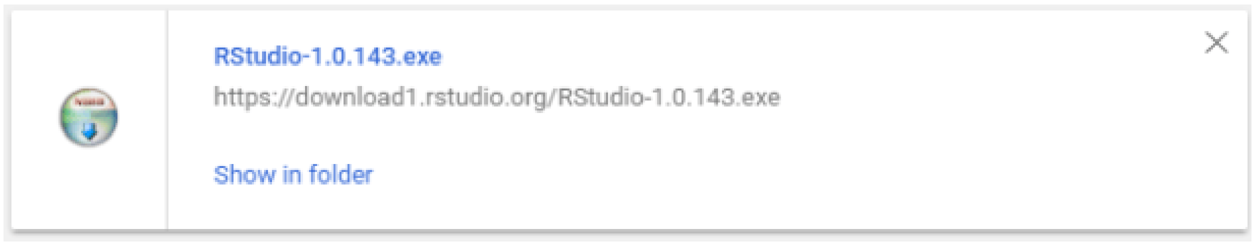
Steps to install R Studio:

1. First, go to https://www.rstudio.com/products/rstudio/ and click on DOWNLOAD RSTUDIO DESKTOP.
2. Then find out the installers for supported for Windows platforms. You will be getting something like this RStudio 1.0.143 – Windows Vista/7/8/10 . Click on that to download RStudio.

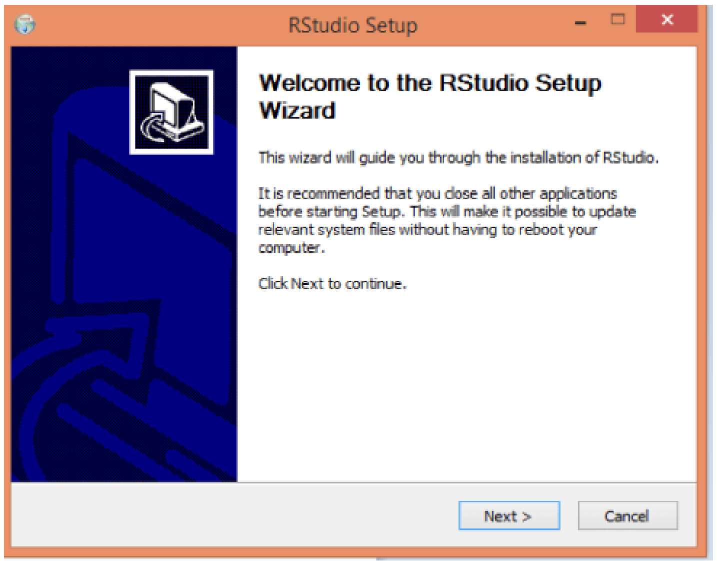

Installers for Supported Platforms

Installers	Size	Date	MD5
RStudio 1.0.143 - Windows Vista/7/8/10	81.9 MB	2017-04-19	76bb84296b9202759b3eb1de555a2231
RStudio 1.0.143 - Mac OS X 10.6+ (64-bit)	71.2 MB	2017-04-19	c7f1ed865428b225b202fd1b431954b4
RStudio 1.0.143 - Ubuntu 12.04+/Debian 8+ (32-bit)	85.5 MB	2017-04-19	21ca14bffcdc1a2361ead2d763d0313d
RStudio 1.0.143 - Ubuntu 12.04+/Debian 8+ (64-bit)	92.1 MB	2017-04-19	75761eae209158d8415d562b3771fbec
RStudio 1.0.143 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	84.7 MB	2017-04-19	2c356d4ee50667ad4042ee196afb3c53
RStudio 1.0.143 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	85.7 MB	2017-04-19	7ab5fc240351debe491c6c5a7acb6068

3. Next, find the file that was downloaded in your system and double click it. It will be named something like RStudio-1.0.143.exe. This will start the install process.


RStudio-1.0.143.exe
https://download1.rstudio.org/RStudio-1.0.143.exe
Show in folder

4. Click next to continue when the install wizard opens.


RStudio Setup

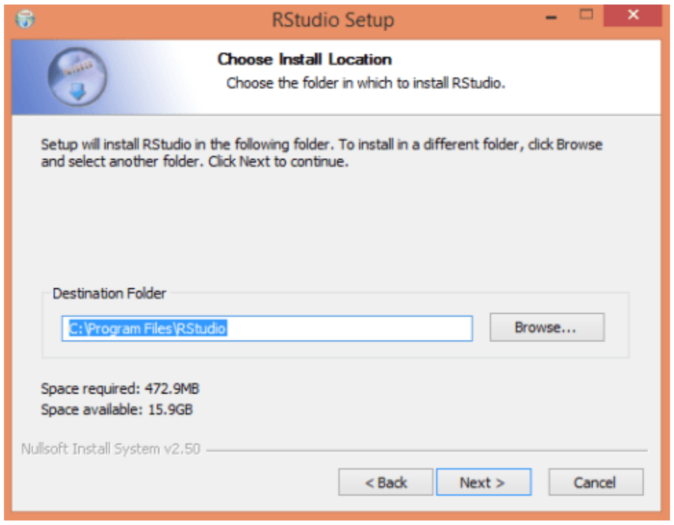
Welcome to the RStudio Setup Wizard

This wizard will guide you through the installation of RStudio.

It is recommended that you close all other applications before starting Setup. This will make it possible to update relevant system files without having to reboot your computer.

Click Next to continue.

5. Click next to accept the default install location.


RStudio Setup

Choose Install Location
Choose the folder in which to install RStudio.

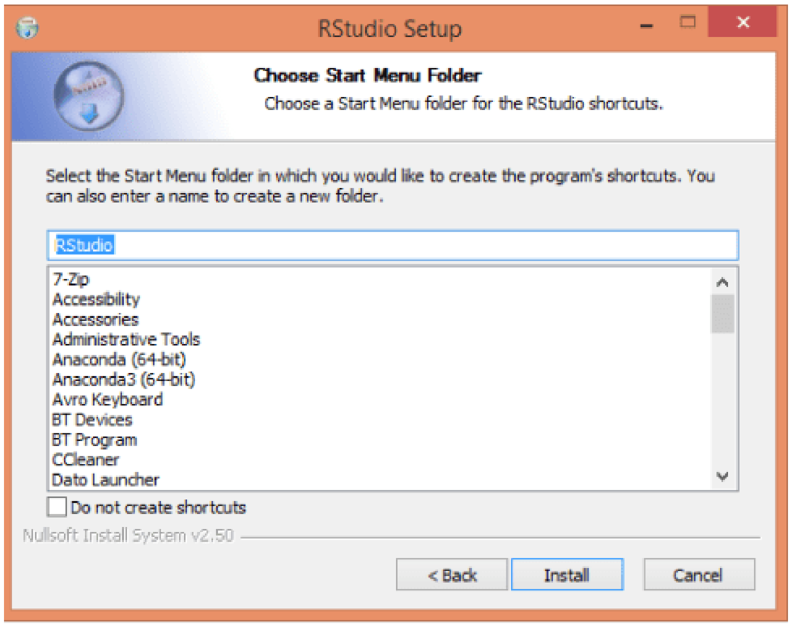
Setup will install RStudio in the following folder. To install in a different folder, click Browse and select another folder. Click Next to continue.

Destination Folder
C:\Program Files\RStudio

Space required: 472.9MB
Space available: 15.9GB

Nullsoft Install System v2.50

6. Click Install to accept the default start menu folder and install RStudio.



RStudio Setup

Choose Start Menu Folder

Choose a Start Menu folder for the RStudio shortcuts.

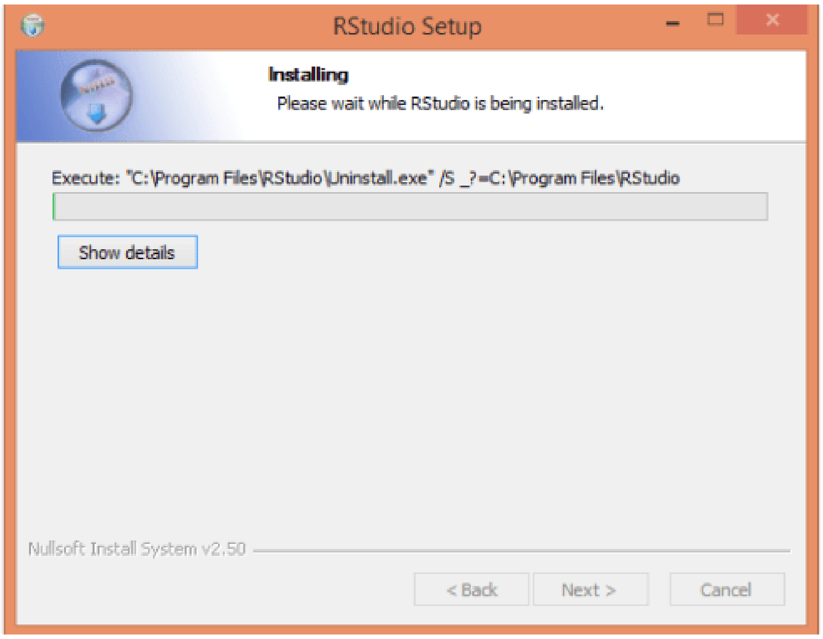
Select the Start Menu folder in which you would like to create the program's shortcuts. You can also enter a name to create a new folder.

RStudio

7-Zip
Accessibility
Accessories
Administrative Tools
Anaconda (64-bit)
Anaconda3 (64-bit)
Avro Keyboard
BT Devices
BT Program
CCleaner
Dato Launcher

Do not create shortcuts

7. Now it will be installing in your system.



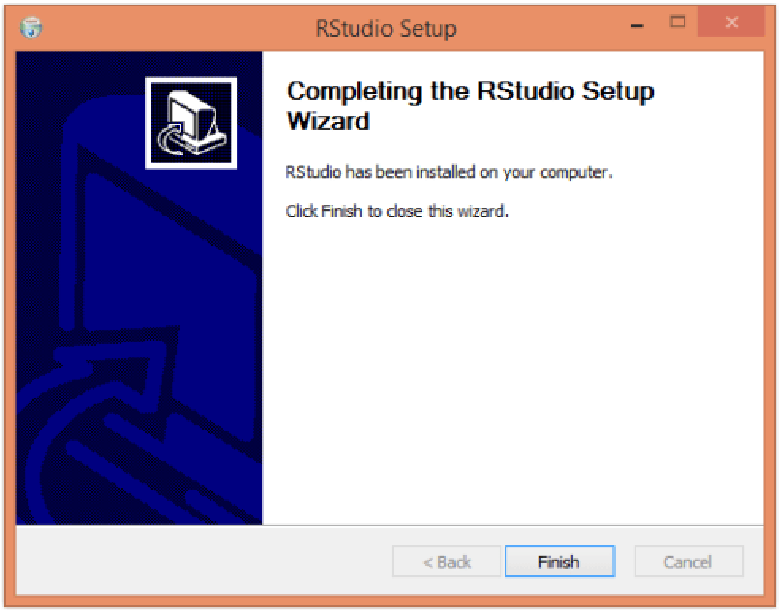
RStudio Setup

Installing

Please wait while RStudio is being installed.

Execute: "C:\Program Files\RStudio\Uninstall.exe" /S _?=C:\Program Files\RStudio

8. Next, click Finish to close the wizard.



Completing the RStudio Setup Wizard

RStudio has been installed on your computer.

Click Finish to close this wizard.

9. Check if there is an “RStudio” icon on the desktop of the computer. If so, double-click on the “RStudio” icon to start RStudio.

Conclusion:

Thus, R and R Studio has been installed in windows 10.

Lab Outcome: LO1

Anirudh Poroorkara
Roll: 44
IT-1 B12

Lab Assignment: 02

Aim:

Datatypes and Data structures in R.

Code:

```
> #Experiment 2: Datatypes in R
> #Anirudh Poroorkara
> #Roll: 44
>
> #To check your current working directory, you can run the command
getwd() in the RStudio console.
> #Present working directory
> getwd()
[1] "/home/lab1003"
>
> #To change your working directory, use setwd and specify the path to
the desired folder.
> setwd("~/Desktop")
> getwd()
[1] "/home/lab1003/Desktop"
>
> #The dir R function returns a character vector of file and/or folder
names within a directory.
> dir()
[1] "abc"                          "Assignment No. 7TCPDUMP.docx"
"c.c"
[4] "dsa softfin.odt"              "ds.py"
"EDU1664_Launcher"
[7] "KitcheNette-master"           "RSA.class"
"RSA.java"
> #ls and objects return a vector of character strings giving the
names of the objects in the specified environment.
> ls()
 [1] "A"               "B"               "C"               "c_emp.data"
"colnames"        "column.names"
 [7] "factor_traffic" "list2"           "matrix.names"    "merged.list"
"R"               "RandomList"
[13] "result"         "resultAdd"       "resultDiv"       "resultMul"
"resultSub"       "rownames"
```

```
[19] "row.names"      "traffic"         "v"               "v1"
"v2"              "vector1"
[25] "vector2"        "w"
> objects()
 [1] "A"              "B"               "C"               "c_emp.data"
"colnames"        "column.names"
 [7] "factor_traffic" "list2"           "matrix.names"    "merged.list"
"R"               "RandomList"
[13] "result"         "resultAdd"       "resultDiv"       "resultMul"
"resultSub"       "rownames"
[19] "row.names"      "traffic"         "v"               "v1"
"v2"              "vector1"
[25] "vector2"        "w"
>
> #The help() function and ? help operator in R provide access to the
documentation pages for R functions, data sets, and other objects,
both for packages in the standard R distribution and for contributed
packages.
> help.start()
> help("sum")
> ?summary
> ?mean.Date
>
> #declared variable x with value 1
> x<-44
>
> #The function class prints the vector of names of classes an object
inherits from.
> class(x)
[1] "numeric"
> #The R specific function typeof returns the type of an R object
> typeof(x)
[1] "double"
> #print x
> x
[1] 44
>
> #declared variable y with value Anirudh
> y<-'Anirudh'
> #Typeof y
> typeof(y)
[1] "character"
> #class of y
> class(y)
```

```
[1] "character"
> #Check if character
> is.character(y)
[1] TRUE
> #as. integer attempts to coerce its argument to be of integer type.
> as.integer(y)
[1] NA
Warning message:
NAs introduced by coercion
>
> #Creating Sequence from 4 to 14
> z <- 4:14
> z
 [1]  4  5  6  7  8  9 10 11 12 13 14
> length(z)
[1] 11
>
> #Vector
> #A vector is a collection of elements that are most commonly of mode
character, logical, integer or numeric.
>
> v <- c(1, 2, 3)
> v
[1] 1 2 3
> w <- c("Ani", "Nivi", "Dhi")
> w
[1] "Ani"  "Nivi" "Dhi"
> length(v)
[1] 3
> class(w)
[1] "character"
>
> #Adding Elements in vectors
> w <- c(w, "Muni")
>
> # Accessing vector elements using position.
> w[c(2,3)]
[1] "Nivi" "Dhi"
> # Accessing vector elements using 0/1 indexing.
> w[c(0,0,0,1)]
[1] "Ani"
>
> #Lists
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> # Create a list containing strings, numbers, vectors and a logical
values.
> RandomList <- list("Black", c(21,32,11), TRUE, 44)
> RandomList
[[1]]
[1] "Black"

[[2]]
[1] 21 32 11

[[3]]
[1] TRUE

[[4]]
[1] 44

>
> list2 <- list(45,65.5, "Red")
>
> # Merge the two lists.
> merged.list <- c(RandomList,list2)
>
> # Print the merged list.
> print(merged.list)
[[1]]
[1] "Black"

[[2]]
[1] 21 32 11

[[3]]
[1] TRUE

[[4]]
[1] 44

[[5]]
[1] 45

[[6]]
[1] 65.5

[[7]]
[1] "Red"
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
>
> # Convert the lists to vectors.
> v1 <- unlist(RandomList)
> v2 <- unlist(list2)
> v1
[1] "Black" "21"    "32"    "11"    "TRUE"  "44"
> v2
[1] "45"   "65.5" "Red"
>
> #Matrix
> #Matrices are the R objects in which the elements are arranged in a
two-dimensional rectangular layout. They contain elements of the same
atomic types.
>
> # Elements are arranged sequentially by row.
> A <- matrix(c(4:15), nrow = 4, byrow = TRUE)
> A
     [,1] [,2] [,3]
[1,]    4    5    6
[2,]    7    8    9
[3,]   10   11   12
[4,]   13   14   15
>
> # Elements are arranged sequentially by column.
> B <- matrix(c(4:15), nrow = 4, byrow = FALSE)
> B
     [,1] [,2] [,3]
[1,]    4    8   12
[2,]    5    9   13
[3,]    6   10   14
[4,]    7   11   15
>
> # Define the column and row names.
> rownames = c("R1", "R2", "R3", "R4")
> colnames = c("C1", "C2", "C3")
> C <- matrix(c(10:21), nrow = 4, byrow = TRUE, dimnames =
list(rownames, colnames))
> C
   C1 C2 C3
R1 10 11 12
R2 13 14 15
R3 16 17 18
R4 19 20 21
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
>
> #Add matrix A and B
> resultAdd <- A + B
> resultAdd
     [,1] [,2] [,3]
[1,]    8   13   18
[2,]   12   17   22
[3,]   16   21   26
[4,]   20   25   30
>
> #Subtract matrix A and B
> resultSub <- A - B
> resultSub
     [,1] [,2] [,3]
[1,]    0   -3   -6
[2,]    2   -1   -4
[3,]    4    1   -2
[4,]    6    3    0
>
> #Multiply matrix A and B
> resultMul <- A * B
> resultMul
     [,1] [,2] [,3]
[1,]   16   40   72
[2,]   35   72  117
[3,]   60  110  168
[4,]   91  154  225
>
> #Divide matrix A and B
> resultDiv <- A / B
> resultDiv
         [,1]      [,2]      [,3]
[1,] 1.000000 0.6250000 0.5000000
[2,] 1.400000 0.8888889 0.6923077
[3,] 1.666667 1.1000000 0.8571429
[4,] 1.857143 1.2727273 1.0000000
>
> #Arrays
> #Arrays are the R data objects which can store data in more than two
dimensions
> # Create two vectors of different lengths.
> v1 <- c(4,90,23)
> v2 <- c(20,45,27,44,87,39)
> column.names <- c("C1","C2","C3")
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> row.names <- c("R1","R2","R3")
> matrix.names <- c("Matrix_1","Matrix_2")
>
> # Take these vectors as input to the array.
> R <- array(c(vector1,vector2),dim = c(3,3,2),dimnames =
list(row.names,column.names,matrix.names))
> R
, , Matrix_1

   C1 C2 C3
R1  4 14 17
R2 90 15 18
R3 23 16 19

, , Matrix_2

   C1 C2 C3
R1  4 14 17
R2 90 15 18
R3 23 16 19

>
>
> # Print the third row of the second matrix of the array.
> R[3,,2]
C1 C2 C3
23 16 19
>
> # Print the element in the 1st row and 3rd column of the 1st matrix.
> R[1,3,1]
[1] 17
>
> #Calculate sum of all each row in Array
> rowSums(R)
 R1  R2  R3
 70 246 116
> #Calculate sum of columns in Array
> colSums(R)
   Matrix_1 Matrix_2
C1      117      117
C2       45       45
C3       54       54
>
>
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> #Factor
> #Factors are the data objects which are used to categorize the data
and store it as levels. They can store both strings and integers. They
are useful in the columns which have a limited number of unique
values.
>
> # Create a vector as input.
> traffic <-
c("Yello","Green","Yello","Red","Red","Yello","Green","Green","Green",
"Yello","Red")
> traffic
 [1] "Yello" "Green" "Yello" "Red"   "Red"   "Yello" "Green" "Green"
"Green" "Yello" "Red"
> is.factor(traffic)
[1] FALSE
>
> # Apply the factor function.
> factor_traffic <- factor(traffic)
> factor_traffic
 [1] Yello Green Yello Red   Red   Yello Green Green Green Yello Red
Levels: Green Red Yello
> is.factor(factor_traffic)
[1] TRUE
>
> #DataFrame
> #A data frame is a table or a two-dimensional array-like structure
in which each column contains values of one variable and each row
contains one set of values from each column.
>
> # Create the data frame.
>
> c_emp.data <- data.frame(
+   c_id = c (1:5),
+   c_name = c("Vedant","Amol","Muneeb","Maitreyee","Shah"),
+   c_roll = c(43,132,32,28,39) ,
+   stringsAsFactors = FALSE
+ )
>
> # Print the data frame.
> c_emp.data
  c_id    c_name c_roll
1    1    Vedant     43
2    2      Amol    132
3    3    Muneeb     32
```

```
4    4 Maitreyee     28
5    5      Shah     39
>
> # Get the structure of the data frame.
> str(c_emp.data)
'data.frame':	5 obs. of  3 variables:
 $ c_id  : int  1 2 3 4 5
 $ c_name: chr  "Vedant" "Amol" "Muneeb" "Maitreyee" ...
 $ c_roll: num  43 132 32 28 39
>
> # Print the summary.
> print(summary(c_emp.data))
      c_id      c_name              c_roll     
 Min.   :1   Length:5           Min.   : 28.0  
 1st Qu.:2   Class :character   1st Qu.: 32.0  
 Median :3   Mode  :character   Median : 39.0  
 Mean   :3                      Mean   : 54.8  
 3rd Qu.:4                      3rd Qu.: 43.0  
 Max.   :5                      Max.   :132.0  
>
> # Extract first two rows.
> result <- c_emp.data[1:2,]
> result
  c_id c_name c_roll
1    1 Vedant     43
2    2   Amol    132
>
> # Extract Specific columns.
> result <- data.frame(c_emp.data$c_name,c_emp.data$c_roll)
> result
  c_emp.data.c_name c_emp.data.c_roll
1            Vedant                43
2              Amol               132
3            Muneeb                32
4         Maitreyee                28
5              Shah                39
>
> # Add the batch coulmn.
> c_emp.data$batch <- c("B1","B3","B6","B5","B4")
> v <- c_emp.data
> v
  c_id    c_name c_roll batch
1    1    Vedant     43    B1
2    2      Amol    132    B3
```

```
3    3    Muneeb     32    B6
4    4 Maitreyee     28    B5
5    5      Shah     39    B4
```

Conclusion:

Thus, datatypes and structures were executed in R.

Lab Outcome: LO1, LO2

Anirudh Poroorkara
Roll: 44
IT-1 B12

Lab Assignment: 03

Aim:

Loops and Conditions in R

Code:

```
> #Lab Assignment 3
> #Loops and conditions
>
> #get the working directory
> getwd()
[1] "/cloud/project"
>
> #Control Structures
> #Control structures in R allow you to control the flow of execution
of a series of R expressions. Basically, control structures allow you
to put some “logic” into your R code, rather than just always executin
g the same R code every time. Control structures allow you to respond
to inputs or to features of the data and execute different R expressio
ns accordingly.

> #1.  IF-ELSE
> #The if-else combination is probably the most commonly used control
structure not only in R but also for other programming languages. This
structure allows you to test a condition and act on it depending on wh
ether it’s true or false.
> x<-1000
> if(x > 99){
+   print("X is greater than 100")
+ }
[1] "X is greater than 100"
>
> x <- 9
> if(x > 10){
+   print("Value is greater than 10")
+ } else {
+   print("Value is less than or equal to 10")
+ }
[1] "Value is less than or equal to 10"
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> #2. FOR loop
> #For loops are most commonly used for iterating over the elements of
an object (list, vector, etc.)
>
> for(i in 4:12) {
+   print(i)
+ }
[1] 4
[1] 5
[1] 6
[1] 7
[1] 8
[1] 9
[1] 10
[1] 11
[1] 12
>
> x <- c("ani", "nivi", "dhi", "parth")
> for(letter in x) {
+   print(letter)
+ }
[1] "ani"
[1] "nivi"
[1] "dhi"
[1] "parth"

> #3. WHILE loop
> #While loops begin by testing a condition. If it is true, then they
execute the loop body. Once the loop body is executed, the condition i
s tested again, and so forth, until the condition is false, after whic
h the loop exits.
>
> c <- 0
> while(c < 5) {
+   print(c)
+   c <- c + 1
+ }
[1] 0
[1] 1
[1] 2
[1] 3
[1] 4
>
> x <- c("ani", "nivi", "dhi", "parth")
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> i <- 0
> while(i < 5) {
+   print(x[i])
+   i <- i+1
+ }
character(0)
[1] "ani"
[1] "nivi"
[1] "dhi"
[1] "parth"

> #4. REPEAT loop
> #A repeat loop is used to iterate over a block of code multiple numb
er of times. There is no condition check in repeat loop to exit the lo
op. The only way to exit a repeat loop is to call break. These are not
commonly used in statistical or data analysis applications but they do
have their uses.
> #When the break statement is encountered inside a loop, the loop is
immediately terminated and program control resumes at the next stateme
nt following the loop.
>
> x <- 5
>
> repeat {
+   print(x)
+   x = x+1
+   if (x == 10){
+     break
+   }
+ }
[1] 5
[1] 6
[1] 7
[1] 8
[1] 9
> #The next statement in R programming language is useful when we want
to skip the current iteration of a loop without terminating it. On enc
ountering next, the R parser skips further evaluation and starts next
iteration of the loop.

> for(i in 1:10) {
+
+   if(i >= 5) {
+     # Skip the first 5 iterations
+     next
+   }
+   print(i)
+ }
[1] 1
[1] 2
[1] 3
[1] 4

> #Functions
> #Functions are defined using the function() directive and are stored
as R objects
>
> #Basic Simple function
> #define the function
> f <- function() {
+
+   print("Hi Anirudh!!!")
+ }
> #Call the function
> f()
[1] "Hi Anirudh!!!"
>
> #Build-in Functions
> # Create a sequence of numbers 4 to 14
> print(seq(4,14))
 [1]  4  5  6  7  8  9 10 11 12 13 14
>
> # Find mean of numbers 4 to 14
> print(mean(4:14))
[1] 9
>
> # Find sum of numbers 4 to 14
> print(sum(4:14))
[1] 99
>
> #User defined function
> pow <- function(x, y = 2) {
+   result <- x^y
+   print(paste(x,"raised to the power", y, "is", result))
+ }
>
> #Give both arguments
> pow(2,3)
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
[1] "2 raised to the power 3 is 8"
>
> #Give one arguement, the default value of other argument is taken
> pow(2)
[1] "2 raised to the power 2 is 4"
>
> #Return statement
> #The return statement terminates the execution of a function and ret
urns control to the calling function.
>
> calculate <- function(x) {
+   if (x > 0) {
+     result <- "Positive"
+   } else if (x < 0) {
+     result <- "Negative"
+   } else {
+     result <- "Zero"
+   }
+   return(result)
+ }
>
> calculate(9)
[1] "Positive"
> calculate(-5)
[1] "Negative"
> calculate(0)
[1] "Zero"

> #Different math functions
> x = 3.456
> n = 2
> #absolute value
> abs(x)
[1] 3.456
> #ceiling(3.475) is 4
> ceiling(x)
[1] 4
> #square root
> sqrt(x)
[1] 1.859032
> #floor(3.475) is 3
> floor(x)
[1] 3
> #natural logarithm
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> log(x)
[1] 1.240112
> #trunc(5.99) is 5
> trunc(x)
[1] 3
> #round(3.475, digit=2) is 3.48
> round(x, digits=n)
[1] 3.46
> #common logarithm
> log10(x)
[1] 0.5385737
> #signif(3.475, digit=2) is 3.5
> signif(x, digits=n)
[1] 3.5
> #e^x
> exp(x)
[1] 31.68996
> #Trigonometric functions
> cos(x)
[1] -0.9509798
> sin(x)
[1] -0.3092529
> tan(x)
[1] 0.325194
```

Conclusion:

Thus functions, loops and conditions was studied and implemented in R.

Lab Outcome: LO1

Anirudh Poroorkara
Roll: 44
IT-1 B12

Lab Assignment: 04

Aim:

EDA on inbuilt R dataset

Code:

```
> #Lab Assignment 4
> #Exploratory Data Analysis
>
> #Dataset name - UFC Fight Data
> #Kaggle link
>
>
> #get the working directory
> getwd()
[1] "/cloud/project"
>
> #Choosing dataset into R
> #data() function is used to list all the internal datasets present
> data()
>
> #we load the chicken weights dataset
> data("Orange")
>
> # ? is used to get infomation about any entity in R
> # here we use ?(Orange) to get information about the Orange dataaset
> ?(Orange)
>
> #load the dataset in a variable
> df <- Orange
>
> #Check the class before doing any cleaning
> class(df)
[1] "nfnGroupedData" "nfGroupedData"
[3] "groupedData"    "data.frame"
>
> #Check the number of rows and columns the data frame has
> dim(df)
[1] 35  3
> #We can see that the data frame has 35 rows and 3 columns
>
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

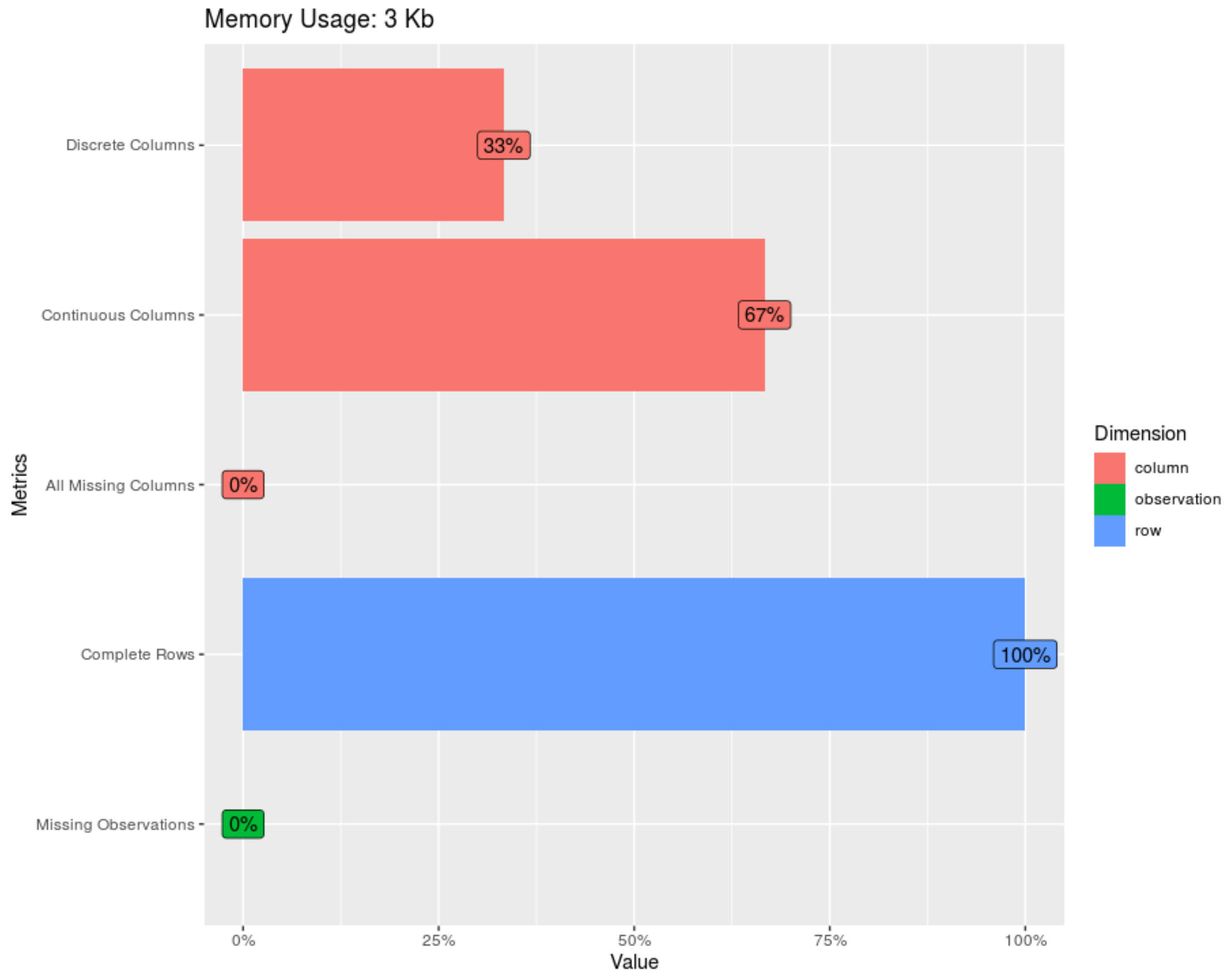
```
> #Finding summary of entire dataset
> summary(df)
 Tree        age         circumference  
 3:7   Min.   : 118.0   Min.   : 30.0  
 1:7   1st Qu.: 484.0   1st Qu.: 65.5  
 5:7   Median :1004.0   Median :115.0  
 2:7   Mean   : 922.1   Mean   :115.9  
 4:7   3rd Qu.:1372.0   3rd Qu.:161.5  
       Max.   :1582.0   Max.   :214.0  
> 
> #nrow() function is used to print the number of rows in dataset
> nrow(df)
[1] 35
> 
> #ncol() function is used to print the number of columns in the datas
et
> ncol(df)
[1] 3
> 
> #check for null values in dataset
> is.null(df)
[1] FALSE
> 
> #duplicated function is used to display duplicate values
> #it prints 'false' for non duplicate values and 'true' for duplicate
values
> duplicated(df)
 [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [9] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[17] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[33] FALSE FALSE FALSE
> 
> #to remove the duplicate value we use Unique function
> #it returns only the unique values in the dataset
> dfd<-unique(df)
> 
> #check again if all duplicate values are removed
> duplicated(df)
 [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [9] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[17] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[33] FALSE FALSE FALSE
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

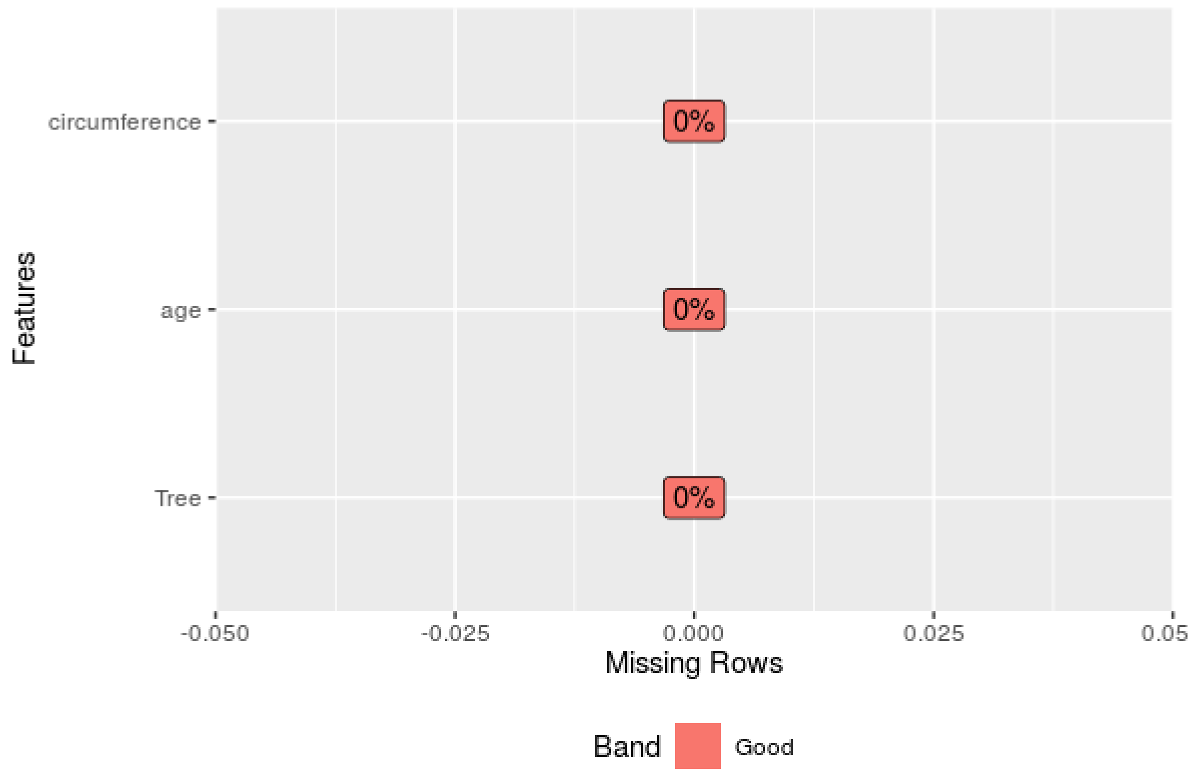
```
> #Exploratory data analysis is the process to get to know your data,
so that you can generate and test your hypothesis. Visualization techn
iques are usually applied.
> #EDA consists of univariate (1-variable) and bivariate (2-variables)
analysis.
> #DataExplorer can help you with different tasks throughout your data
exploration process.
> #Install the package if not installed
> #install.packages('DataExplorer')
> #install.packages('GGally')
>
> #import the library
> library(DataExplorer)
> library(GGally)
>
> #To get introduced to your newly created dataset:
> introduce(df)
  rows columns discrete_columns continuous_columns all_missing_columns
total_missing_values complete_rows total_observations
1   35       3                1                  2                   0
0            35                105
  memory_usage
1         3040
>
> #plot_intro gives a brief Introduction to the dataset.
> #It covers basic information and gives us an idea about the content.
> plot_intro(df)
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

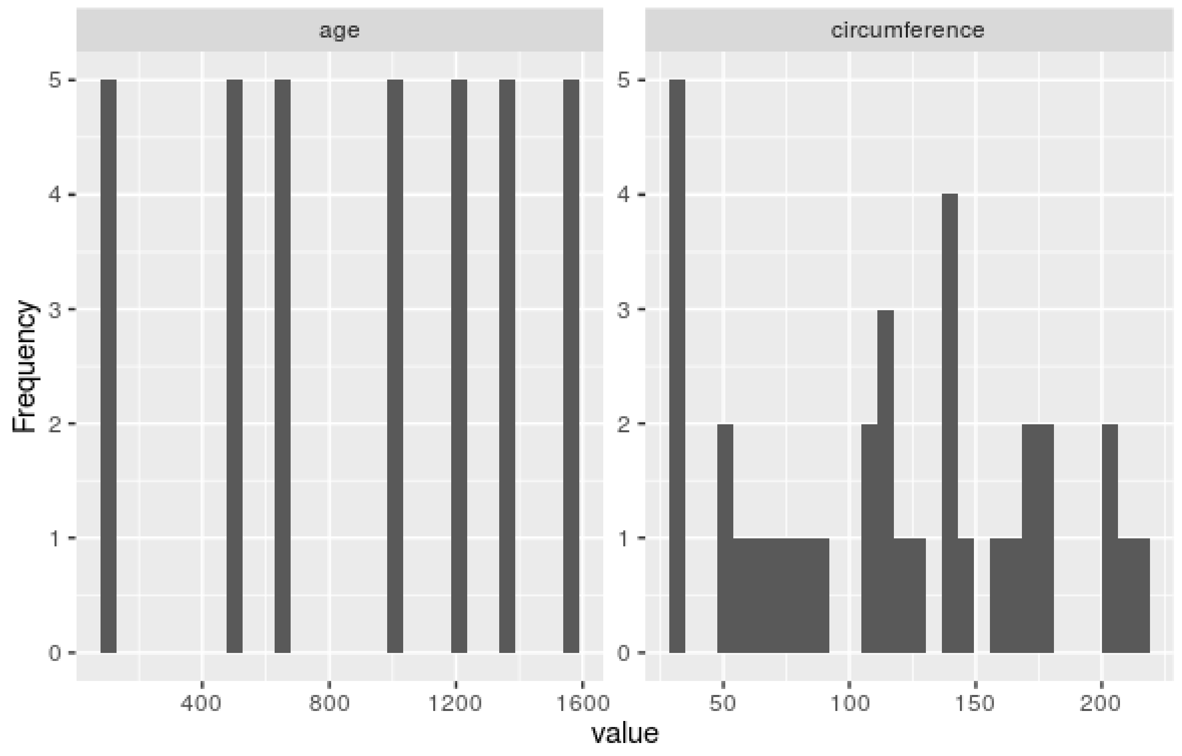
```
>
```



```
> #str gives the entire list of features in a network graph. We have t
he choice of viewing it radially or diagonally
> plot_str(df)
>
> #missing function returns and plots frequency of missing values for
each feature.
> #It also advices us whether to remove certain columns before carryin
g out our analysis
> #All our features are in acceptable form
> plot_missing(df)
```

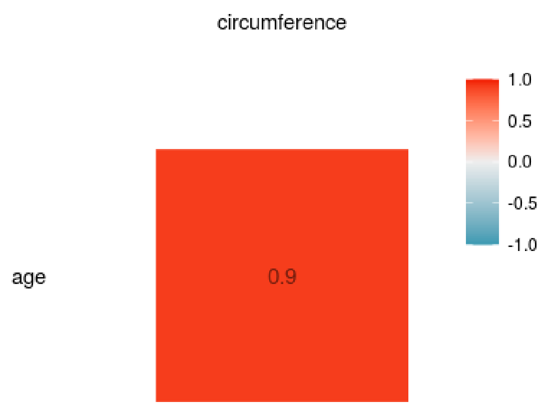


```
>
> #Let us first analyse and represent the continuous variables
> #Histograms can be used to analyse continuous variables
> # you can use a Histogram to organize and display the data in a more
userfriendly format.
> #A Histogram will make it easy
> #to see where the majority of values falls in a measurement scale, a
nd how much variation there is.
> plot_histogram(df)
```

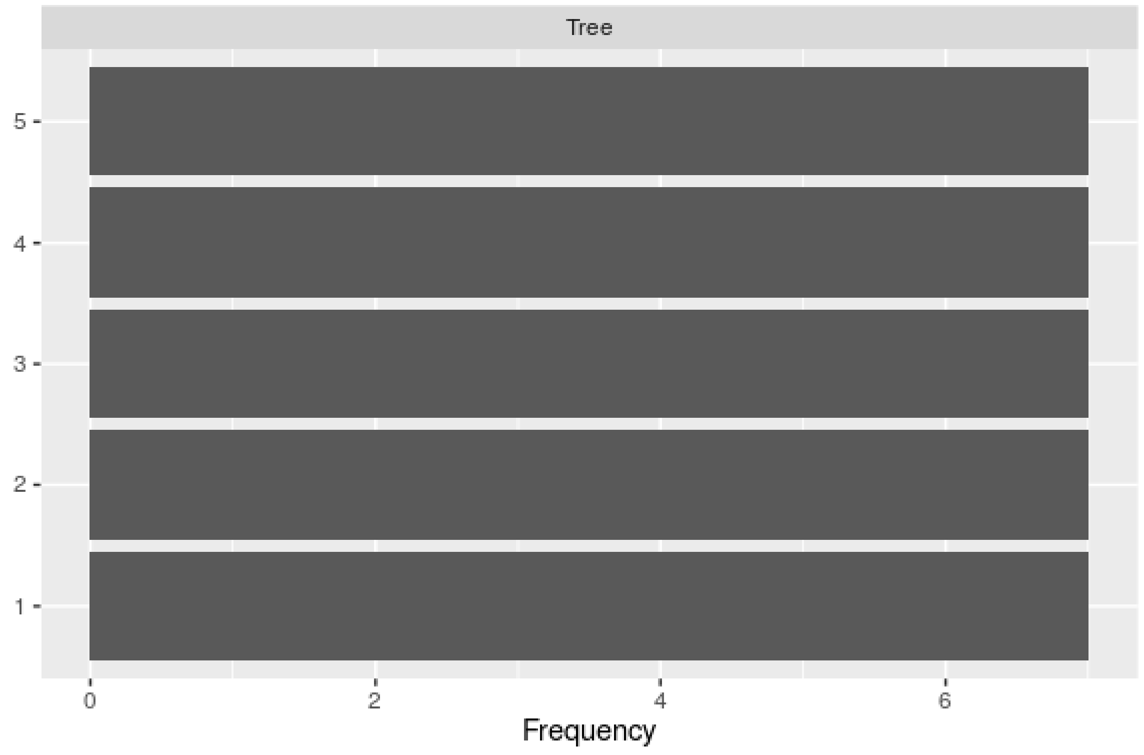


```
>
> #For multivariate analysis, let us do correlation analysis
> #Correlation is used to test relationships between quantitative vari
ables or categorical variables.
> #Tt's a measure of how things are related and how well they are rela
ted.
> ggcorr(df, label =TRUE, label_alpha =TRUE)
Warning message:
In ggcorr(df, label = TRUE, label_alpha = TRUE) :
  data in column(s) 'Tree' are not numeric and were ignored
```

Anirudh Poroorkara
Roll: 44
IT-1 B12



```
>
> #For categorical analysis, we use a bar graph
> #Each bar represents one value. When the bars are stacked next to on
e another,
> #the viewer can compare the different bars, or values, at a glance.
> plot_bar(df)
```



Anirudh Poroorkara
Roll: 44
IT-1 B12

Conclusion:

EDA was conducted on the inbuilt dataset.

Lab Outcome: LO3, LO4

Anirudh Poroorkara
Roll: 44
IT-1 B12

Lab Assignment: 05

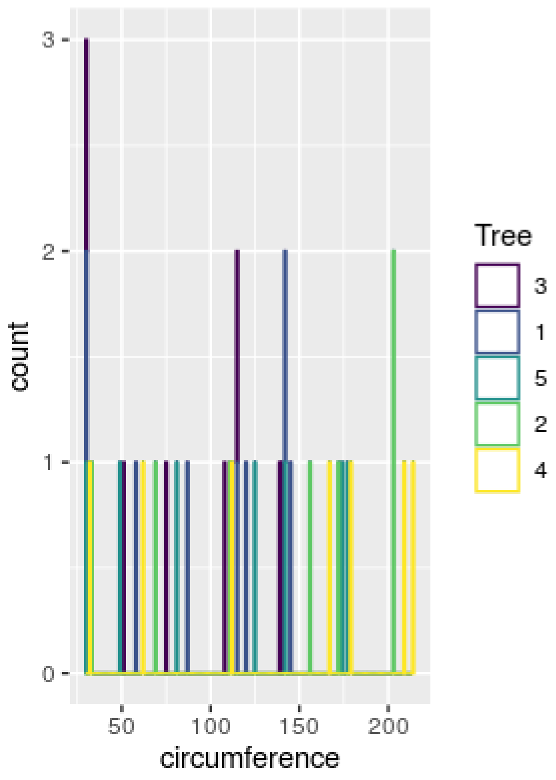
Aim:

Plot graphs using GGPLOT2 on the internal dataset.

Code:

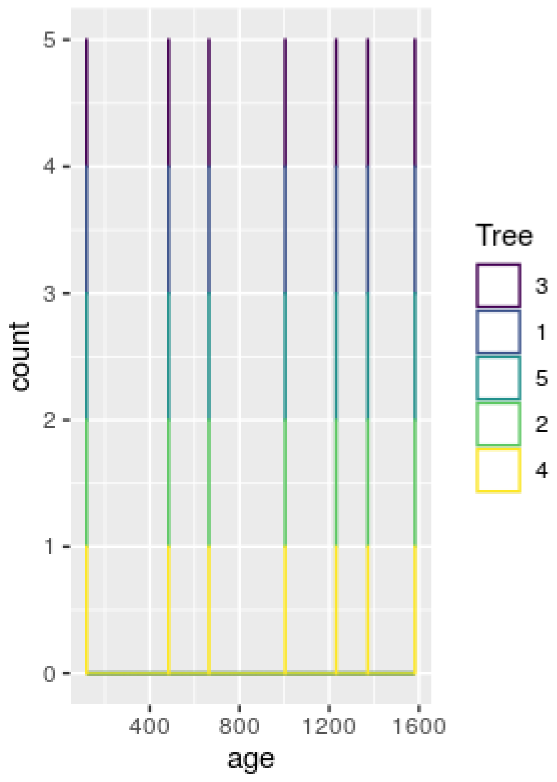
```
> #Lab Assignment 5
> #Data Visualization using ggplot2
>
> #get the working directory
> getwd()
[1] "/cloud/project"
>
> #load the dataset in a variable
> df <- Orange
>
> #ggplot2 is a robust and a versatile R package for generating aesthetic plots a
nd charts.
> #Plot = data + Aesthetics + Geometry
> #1. Data refers to a data frame
> #2. Aesthetics indicates x and y variables. It is also used to tell R how data
are displayed in a plot, e.g. color, size and shape of points etc.
> #3. Geometry refers to the type of graphics
>
> #install the package if not installed
> #install.packages("ggplot2")
> library(ggplot2)
> df
   Tree  age circumference
1     1  118            30
2     1  484            58
3     1  664            87
4     1 1004           115
5     1 1231           120
6     1 1372           142
7     1 1582           145
8     2  118            33
9     2  484            69
10    2  664           111
11    2 1004           156
12    2 1231           172
13    2 1372           203
14    2 1582           203
15    3  118            30
16    3  484            51
17    3  664            75
18    3 1004           108
```

```
19     3 1231           115
20     3 1372           139
21     3 1582           140
22     4  118            32
23     4  484            62
24     4  664           112
25     4 1004           167
26     4 1231           179
27     4 1372           209
28     4 1582           214
29     5  118            30
30     5  484            49
31     5  664            81
32     5 1004           125
33     5 1231           142
34     5 1372           174
35     5 1582           177
> #Plotting with R_Reach and winner using a group histogram
> ggplot(data  = df, aes( x = circumference, color=Tree)) + geom_histogram(fill="
white", binwidth = 1)
```

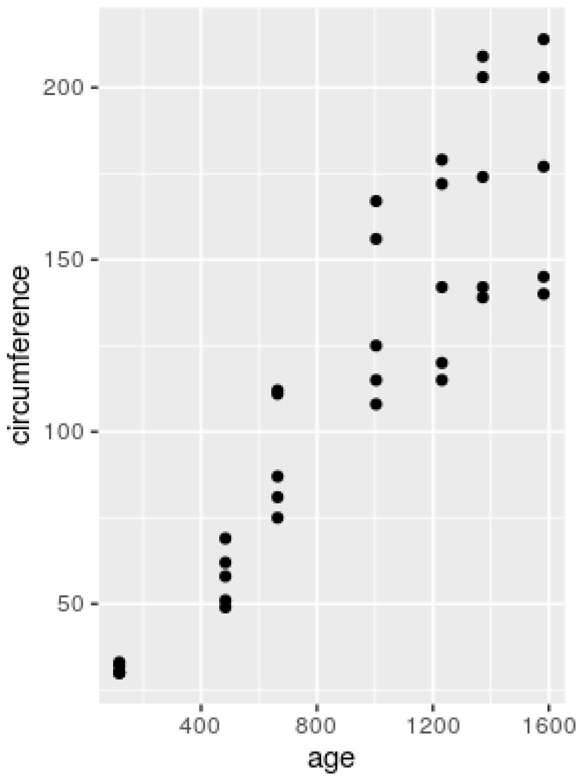


```
> ggplot(data  = df, aes( x = age, color=Tree)) + geom_histogram(fill="white", bi
nwidth = 1)
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

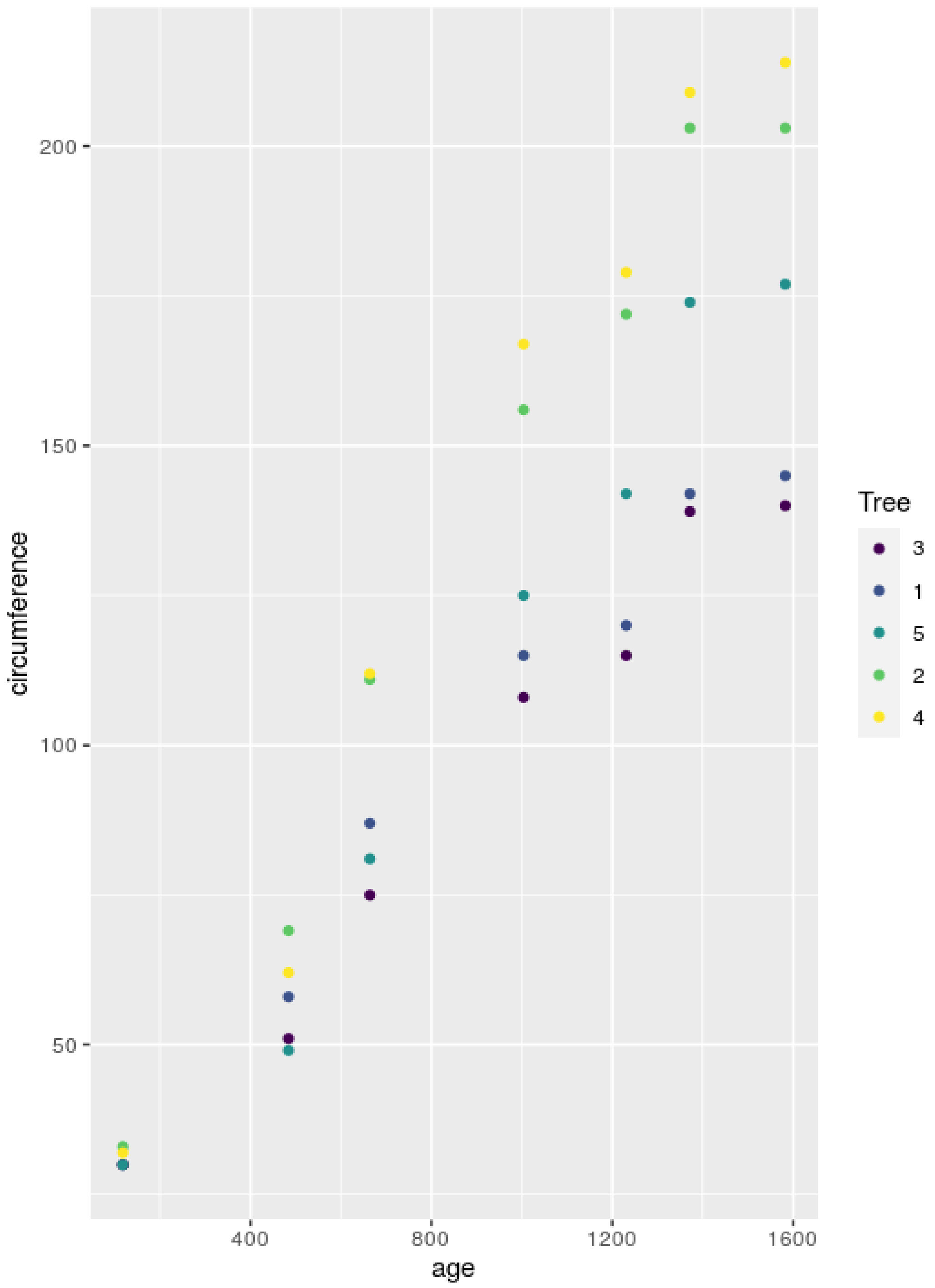


```
> ggplot(data = df, mapping = aes(x = age, y = circumference)) +geom_point()
```



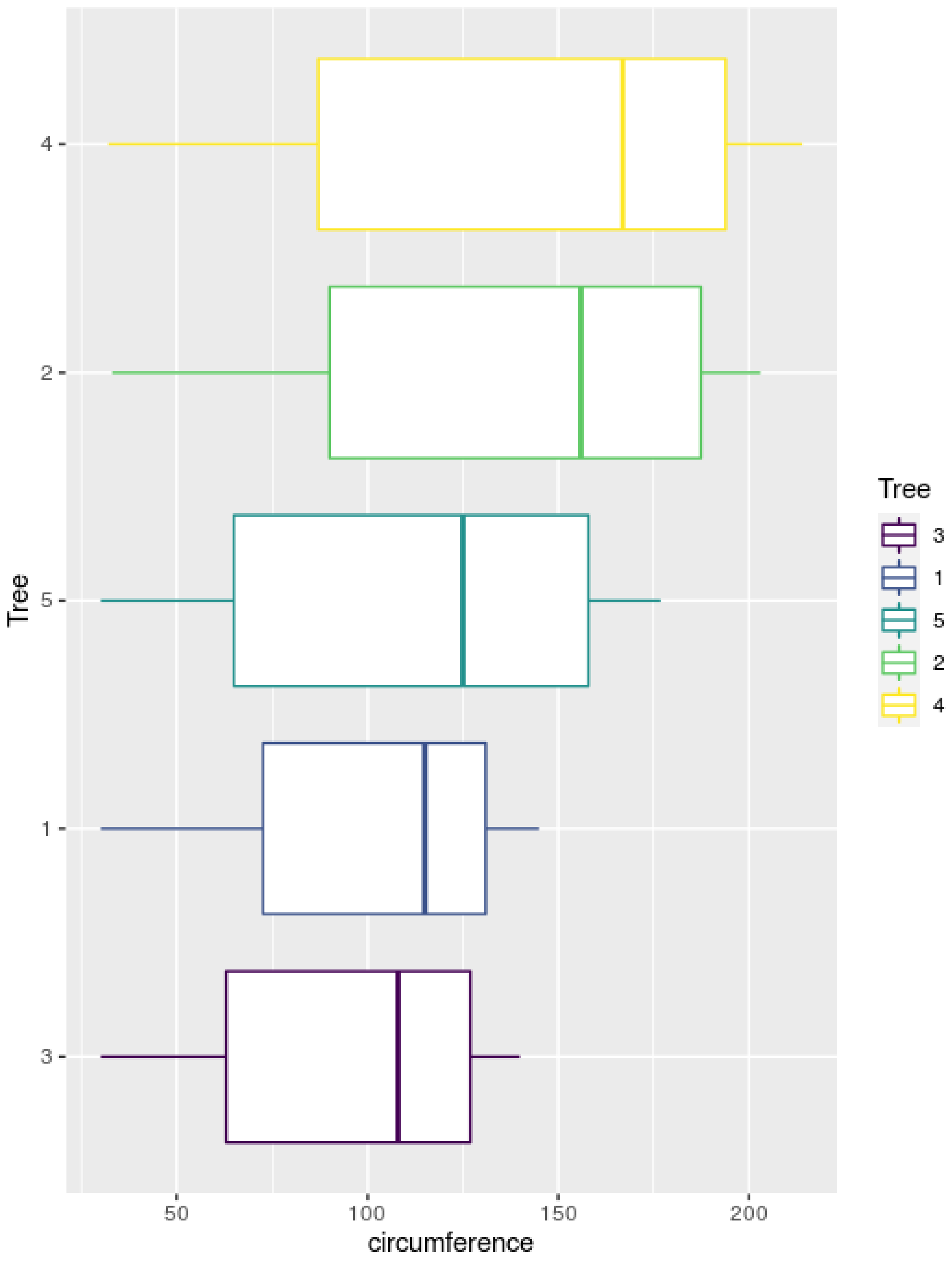
Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> ggplot(data = df, mapping = aes(x = age, y = circumference)) +geom_point(aes(co
lor = Tree))
```



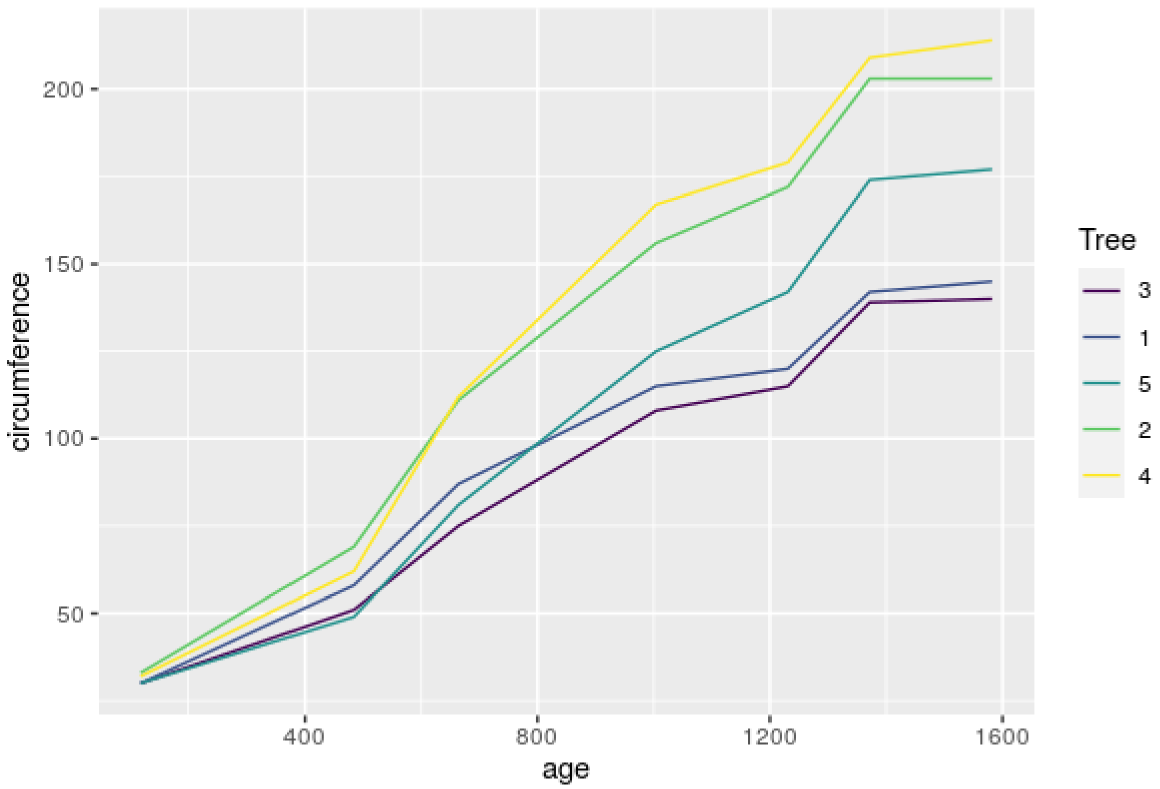
```
> ggplot(data = df, mapping = aes(x = circumference, y = Tree)) +geom_boxplot(aes
(color = Tree))
```

Anirudh Poroorkara
Roll: 44
IT-1 B12



```
> ggplot(data = df, mapping = aes(x = age, y = circumference, color = Tree)) + ge
om_line()
```

Anirudh Poroorkara
Roll: 44
IT-1 B12



Conclusion:

Thus, graphs were plotted using ggplot2.

Lab Outcome: LO2, LO5

Anirudh Poroorkara
Roll: 44
IT-1 B12

Lab Assignment: 06

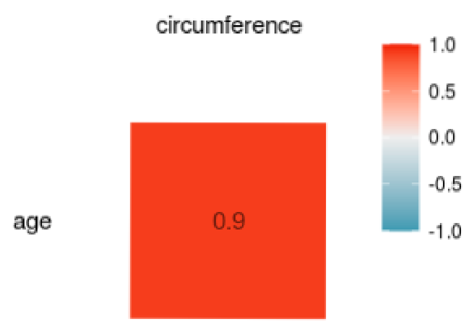
Aim:

Regression and Correlation

Code:

```
> #Lab Assignment 6
> #Regression and Correlation
>
>
> #get the working directory
> getwd()
[1] "/cloud/project"
>
> #Choosing dataset into R
> #data() function is used to list all the internal datasets present
> data()
>
> library(ggplot2)
> #install.packages("GGally")
> library(GGally)
> #we load the chicken weights dataset
> data("Orange")
>
> #For multivariate analysis, let us do correlation analysis
> #Correlation is used to test relationships between quantitative vari
ables or categorical variables.
> #Tt's a measure of how things are related and how well they are rela
ted.
> ggcorr(df, label =TRUE, label_alpha =TRUE)
Warning message:
In ggcorr(df, label = TRUE, label_alpha = TRUE) :
  data in column(s) 'Tree' are not numeric and were ignored
```

Anirudh Poroorkara
Roll: 44
IT-1 B12



```
                    >
> #Regression analysis is used in stats to find trends in data
> #It will provide you with an equation for a graph so that you can ma
ke predictions about your data
> #Since we have a lot of values, we use multiple regression analysis
> #Multiple linear regression is an extension of simple linear regress
ion used to
> #predict an outcome variable (y) on the basis of multiple distinct p
redictor variables (x).
>
> #Here we will predict the circumference based on age of the tree.
>
> model <- lm(circumference ~ age ,data = df)
> summary(model)

Call:
lm(formula = circumference ~ age, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-46.310 -14.946  -0.076  19.697  45.111 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 17.399650   8.622660   2.018   0.0518 .  
age          0.106770   0.008277  12.900 1.93e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 23.74 on 33 degrees of freedom
Multiple R-squared:  0.8345,	Adjusted R-squared:  0.8295 
F-statistic: 166.4 on 1 and 33 DF,  p-value: 1.931e-14
```

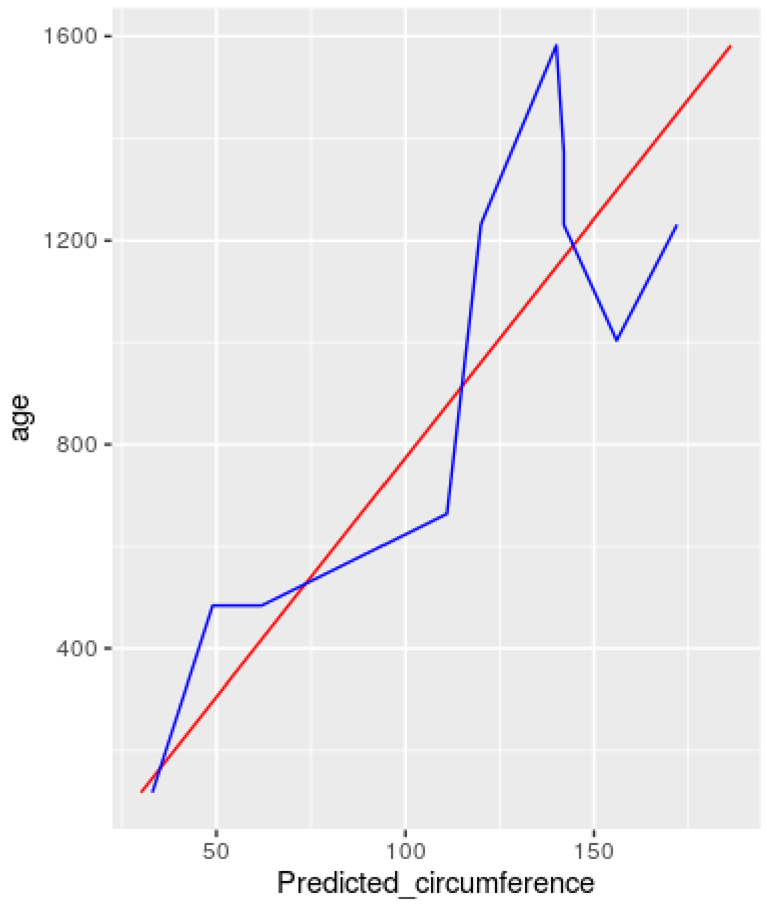
Anirudh Poroorkara
Roll: 44
IT-1 B12

```
>
>
> #The first step in interpreting the regression analysis is
> #to examine the F-statistic and the associated p-value, at the botto
m of model summary.
> summary(model)$coefficient
              Estimate  Std. Error   t value     Pr(>|t|)
(Intercept) 17.3996502 8.622659801  2.017898 5.179267e-02
age          0.1067703 0.008276623 12.900228 1.930596e-14
>
> #For a given the predictor, the t-statistic evaluates whether or not
there is significant association between the predictor and the outcome
variable,
> #that is whether the beta coefficient of the predictor is significan
tly different from zero.
>
> #The RSE estimate gives a measure of error of prediction. The lower
the RSE, the more accurate the model (on the data in hand).
> #The error rate can be estimated by dividing the RSE by the mean out
come variable:
> sigma(model)/mean(df$circumference)
[1] 0.2048874
>
> #We get a 0.204 which means it has a 20% error rate
>
> #Let us now test the created prediction model
> #install.packages("caTools")
> library(caTools)
> #Splitting the data set into training and testing data
> split = sample.split(df$circumference,SplitRatio = 2/3)
> #Training is 2 parts
> training_set = subset(df, split == TRUE)
> #testing is 1 part
> test_set = subset(df, split == FALSE)
>
> test_set
   Tree  age circumference
2     1  484            58
5     1 1231           120
6     1 1372           142
8     2  118            33
10    2  664           111
11    2 1004           156
```

```
12     2 1231           172
16     3  484            51
21     3 1582           140
23     4  484            62
30     5  484            49
33     5 1231           142
> 
> #Do a prediction on the model
> pred = predict(model, newdata = test_set)
> 
> #Store and compare the result in the test cases
> result <- data.frame(age = test_set$age, circumference = test_set$ci
rcumference, Predicted_circumference = pred)
> result
    age circumference Predicted_circumference
2   484            58                69.07649
5  1231           120               148.83392
6  1372           142               163.88854
8   118            33                29.99855
10  664           111                88.29515
11 1004           156               124.59706
12 1231           172               148.83392
16  484            51                69.07649
21 1582           140               186.31030
23  484            62                69.07649
30  484            49                69.07649
33 1231           142               148.83392

> ggplot(data = result) + geom_line(aes(x=Predicted_circumference, y=a
ge), color='red') + geom_line(aes(x=circumference, y=age), color='blue
')
```

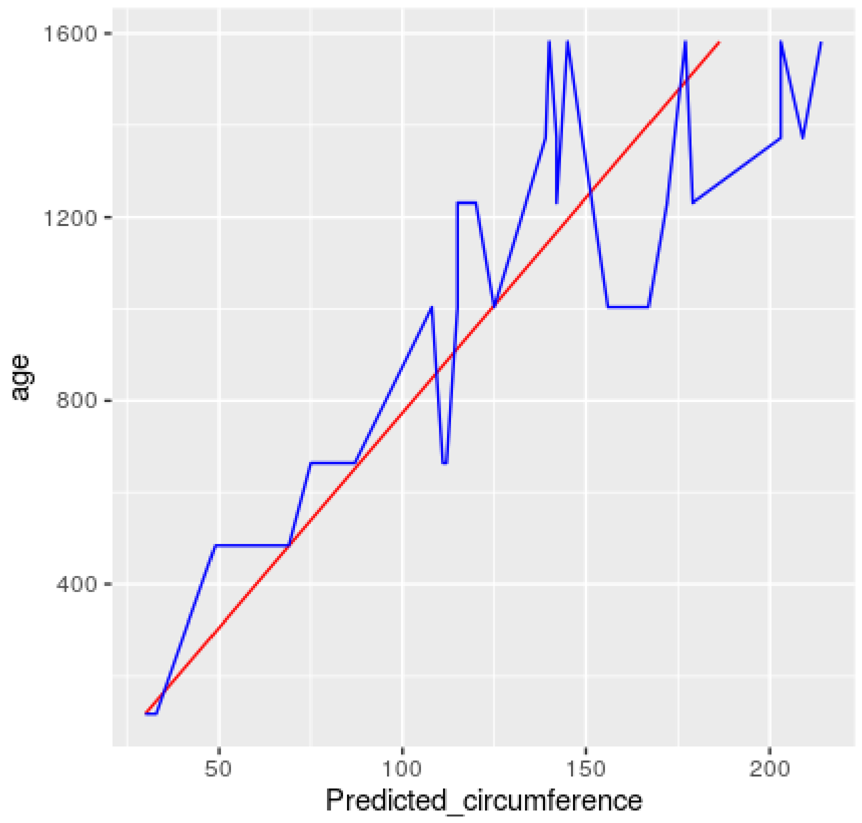
Anirudh Poroorkara
Roll: 44
IT-1 B12



```
> #Do entire df prediction on the model
> pred = predict(model, newdata = df)
> result <- data.frame(age = df$age, circumference = df$circumference,
Predicted_circumference = pred)
> result
    age circumference Predicted_circumference
1   118            30                29.99855
2   484            58                69.07649
3   664            87                88.29515
4  1004           115               124.59706
5  1231           120               148.83392
6  1372           142               163.88854
7  1582           145               186.31030
8   118            33                29.99855
9   484            69                69.07649
10  664           111                88.29515
11 1004           156               124.59706
12 1231           172               148.83392
13 1372           203               163.88854
```

```
14 1582             203                186.31030
15  118              30                 29.99855
16  484              51                 69.07649
17  664              75                 88.29515
18 1004             108                124.59706
19 1231             115                148.83392
20 1372             139                163.88854
21 1582             140                186.31030
22  118              32                 29.99855
23  484              62                 69.07649
24  664             112                 88.29515
25 1004             167                124.59706
26 1231             179                148.83392
27 1372             209                163.88854
28 1582             214                186.31030
29  118              30                 29.99855
30  484              49                 69.07649
31  664              81                 88.29515
32 1004             125                124.59706
33 1231             142                148.83392
34 1372             174                163.88854
35 1582             177                186.31030
> #Plotting model with results
> ggplot(data = result) + geom_line(aes(x=Predicted_circumference, y=a
ge), color='red') + geom_line(aes(x=circumference, y=age), color='blue
')
```

Anirudh Poroorkara
Roll: 44
IT-1 B12



Conclusion:

Thus, regression and correlation analysis were done on the internal dataset and graphs were plotted.

Lab Outcome: LO4

Anirudh Poroorkara
Roll: 44
IT-1 B12

Lab Assignment: 07

Aim:

Choose Dataset from Kaggle, extracting data from large dataset

Code:

```
> #Lab Assignment 7
> #Mini-Project Session 1
>
> #Dataset name - UFC Fight Data
> #Kaggle link
>
>
> #get the working directory
> getwd()
[1] "/cloud/project"
>
> #Loading dataset into R
> #Header is set to to true if file contains Header information
> #Sep stands for seperator which is , in our csv file
>
> df <- read.csv("data.csv", header = TRUE, sep =",")
>
> #Display data file
> head(df)
             R_fighter      B_fighter         Referee       date
location Winner
1          Henry Cejudo  Marlon Moraes    Marc Goddard 2019-06-08 Chica
go, Illinois, USA    Red
2 Valentina Shevchenko    Jessica Eye Robert Madrigal 2019-06-08 Chica
go, Illinois, USA    Red
3         Tony Ferguson Donald Cerrone Dan Miragliotta 2019-06-08 Chica
go, Illinois, USA    Red
4         Jimmie Rivera       Petr Yan Kevin MacDonald 2019-06-08 Chica
go, Illinois, USA   Blue
5           Tai Tuivasa  Blagoy Ivanov Dan Miragliotta 2019-06-08 Chica
go, Illinois, USA   Blue
6        Tatiana Suarez  Nina Ansaroff Robert Madrigal 2019-06-08 Chica
go, Illinois, USA    Red
  title_bout        weight_class no_of_rounds B_current_lose_streak B_
current_win_streak B_draw
1       True        Bantamweight            5                     0
4      0
```

```
2       True   Women's Flyweight            5                     0
3      0
3      False         Lightweight            3                     0
3      0
4      False        Bantamweight            3                     0
4      0
5      False         Heavyweight            3                     0
1      0
6      False Women's Strawweight            3                     0
4      0
  B_avg_BODY_att B_avg_BODY_landed B_avg_CLINCH_att B_avg_CLINCH_lande
d B_avg_DISTANCE_att
1        9.20000           6.00000         0.200000            0.00000
0           62.60000
2       14.60000           9.10000        11.800000            7.30000
0          124.70000
3       15.35484          11.32258         6.741935            4.38709
7           84.74194
4       17.00000          14.00000        13.750000           11.00000
0          109.50000
5       17.00000          14.50000         2.500000            2.00000
0          201.00000
6       19.50000          12.33333        11.833333            7.16666
7          142.33333
  B_avg_DISTANCE_landed B_avg_GROUND_att B_avg_GROUND_landed B_avg_HEA
D_att B_avg_HEAD_landed
1              20.60000         2.600000            2.000000       48.
60000          11.20000
2              42.10000         2.400000            1.900000      112.
00000          32.00000
3              38.58065         5.516129            3.806452       67.
64516          23.25806
4              48.75000        13.000000           10.500000      116.
25000          53.75000
5              59.50000         0.000000            0.000000      184.
50000          45.00000
6              63.83333         6.000000            4.166667      117.
83333          42.66667
   B_avg_KD B_avg_LEG_att B_avg_LEG_landed B_avg_PASS  B_avg_REV B_avg
_SIG_STR_att
1 0.8000000       7.60000          5.40000  0.4000000 0.00000000
65.4000
2 0.0000000      12.30000         10.20000  0.8000000 0.00000000
138.9000
```

```
3 0.6451613      14.00000         12.19355  0.9354839 0.09677419
97.0000
4 0.5000000       3.00000          2.50000  0.5000000 0.25000000
136.2500
5 0.0000000       2.00000          2.00000  0.0000000 0.00000000
203.5000
6 0.0000000      22.83333         20.16667  1.3333333 0.16666667
160.1667
  B_avg_SIG_STR_landed B_avg_SIG_STR_pct B_avg_SUB_ATT B_avg_TD_att B_
avg_TD_landed B_avg_TD_pct
1             22.60000          0.466000     0.4000000    0.8000000
0.2000000    0.1000000
2             51.30000          0.399000     0.7000000    1.0000000
0.5000000    0.2250000
3             46.77419          0.496129     0.3548387    2.1612903
0.6774194    0.2954839
4             70.25000          0.550000     0.2500000    2.5000000
1.2500000    0.2875000
5             61.50000          0.310000     0.0000000    0.0000000
0.0000000    0.0000000
6             75.16667          0.470000     0.6666667    0.8333333
0.3333333    0.2500000
  B_avg_TOTAL_STR_att B_avg_TOTAL_STR_landed B_longest_win_streak B_lo
sses B_avg_opp_BODY_att
1            66.4000               23.60000                   4
1           6.40000
2            158.7000              69.60000                   3
6          13.00000
3            103.7097              52.54839                   8
8          17.90323
4            154.7500              86.75000                   4
0          12.25000
5            204.0000              62.00000                   1
1          42.50000
6            183.5000              95.66667                   4
2          12.00000
  B_avg_opp_BODY_landed B_avg_opp_CLINCH_att B_avg_opp_CLINCH_landed B
_avg_opp_DISTANCE_att
1              4.000000             1.000000                 0.60000
51.20000
2              9.300000            12.800000                 9.60000
101.70000
3             11.870968             8.419355                 5.83871
84.54839
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
4              6.000000             6.000000                3.75000
94.25000
5             23.500000             0.500000                0.50000
205.00000
6              7.333333             9.666667                7.00000
95.16667
  B_avg_opp_DISTANCE_landed B_avg_opp_GROUND_att B_avg_opp_GROUND_land
ed B_avg_opp_HEAD_att
1                  17.40000             0.600000                0.20000
00            39.60000
2                  32.00000             8.100000                6.90000
00            97.70000
3                  38.06452             1.741935                0.93548
39            67.64516
4                  26.75000             1.750000                1.25000
00            82.50000
5                  89.50000             0.000000                0.00000
00           152.50000
6                  38.33333             5.166667                3.50000
00            86.66667
  B_avg_opp_HEAD_landed B_avg_opp_KD B_avg_opp_LEG_att B_avg_opp_LEG_l
anded B_avg_opp_PASS
1                9.40000    0.2000000           6.80000             4.8
00000     0.00000000
2               30.80000    0.1000000          11.90000             8.4
00000     1.40000000
3               25.48387    0.2258065           9.16129             7.4
83871     0.03225806
4               21.50000    0.2500000           7.25000             4.2
50000     0.00000000
5               56.50000    0.0000000          10.50000            10.0
00000     0.00000000
6               33.16667    0.0000000          11.33333             8.3
33333     1.50000000
  B_avg_opp_REV B_avg_opp_SIG_STR_att B_avg_opp_SIG_STR_landed B_avg_o
pp_SIG_STR_pct
1    0.00000000              52.80000                 18.20000
0.2360000
2    0.00000000             122.60000                 48.50000
0.4080000
3    0.03225806              94.70968                 44.83871
0.4532258
4    0.00000000             102.00000                 31.75000
0.3375000
```

```
5    0.00000000             205.50000                 90.00000
0.4300000
6    0.16666667             110.00000                 48.83333
0.4266667
  B_avg_opp_SUB_ATT B_avg_opp_TD_att B_avg_opp_TD_landed B_avg_opp_TD_
pct B_avg_opp_TOTAL_STR_att
1        0.00000000         1.000000           0.4000000       0.10000
000                53.8000
2        0.70000000         2.300000           0.9000000       0.23100
000               151.5000
3        0.09677419         2.096774           0.2258065       0.06354
839               100.3871
4        0.00000000         4.500000           0.7500000       0.09750
000               104.7500
5        0.00000000         0.500000           0.0000000       0.00000
000               205.5000
6        0.00000000         6.000000           1.1666667       0.14000
000               131.5000
  B_avg_opp_TOTAL_STR_landed B_total_rounds_fought B_total_time_fought
.seconds. B_total_title_bouts
1                  19.20000                     9
419.400                  0
2                  75.40000                    29
849.000                  0
3                  49.77419                    68
581.871                  1
4                  34.25000                     9
652.000                  0
5                  90.00000                     8
1200.000                 0
6                  68.66667                    18
886.500                  0
  B_win_by_Decision_Majority B_win_by_Decision_Split B_win_by_Decision
_Unanimous B_win_by_KO.TKO
1                          0                       1
0               2
2                          0                       2
1               0
3                          0                       0
7              10
4                          0                       0
2               2
5                          0                       0
1               0
```

```
6                          0                        0
3               0
  B_win_by_Submission B_win_by_TKO_Doctor_Stoppage B_wins B_Stance B_H
eight_cms B_Reach_cms
1                   1                            0      4 Orthodox
167.64      170.18
2                   0                            1      4 Orthodox
167.64      167.64
3                   6                            0     23 Orthodox
185.42      185.42
4                   0                            0      4   Switch
170.18      170.18
5                   0                            0      1 Southpaw
180.34      185.42
6                   1                            0      4 Orthodox
165.10      162.56
  B_Weight_lbs R_current_lose_streak R_current_win_streak R_draw R_avg
_BODY_att R_avg_BODY_landed
1          135                     0                    4      0
21.90000         16.400000
2          125                     0                    2      0
12.00000          7.714286
3          155                     0                   11      0
13.86667          8.666667
4          135                     1                    0      0
18.25000         10.250000
5          250                     1                    0      0
7.75000          6.750000
6          115                     0                    4      0
8.75000          7.500000
  R_avg_CLINCH_att R_avg_CLINCH_landed R_avg_DISTANCE_att R_avg_DISTAN
CE_landed R_avg_GROUND_att
1        17.000000           11.000000           75.00000
26.50000         9.400000
2         9.285714            6.857143           88.14286
36.14286        18.428571
3         2.866667            1.733333          116.13333
49.46667         5.333333
4         5.875000            4.125000          104.87500
41.00000         1.000000
5        11.000000            7.250000           50.75000
24.75000         0.500000
6         3.000000            2.250000           12.75000
4.75000        42.250000
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
  R_avg_GROUND_landed R_avg_HEAD_att R_avg_HEAD_landed R_avg_KD R_avg_
LEG_att R_avg_LEG_landed
1            6.500000       74.20000             23.90    0.400      
5.30000          3.70000
2           16.428571       84.57143             37.00    0.000      1
9.28571         14.71429
3            4.266667       96.73333             35.60    0.200      1
3.73333         11.20000
4            0.625000       80.50000             24.00    0.375      1
3.00000         11.50000
5            0.500000       50.75000             22.75    0.500      
3.75000          3.00000
6           35.750000       44.75000             31.25    0.000      
4.50000          4.00000
  R_avg_PASS R_avg_REV R_avg_SIG_STR_att R_avg_SIG_STR_landed R_avg_SI
G_STR_pct R_avg_SUB_ATT
1  1.2000000 0.0000000          101.4000             44.00000          
0.4660000     0.1000000
2  1.7142857 0.1428571          115.8571             59.42857          
0.5757143     0.4285714
3  0.3333333 0.1333333          124.3333             55.46667          
0.4300000     1.0000000
4  0.1250000 0.0000000          111.7500             45.75000          
0.3662500     0.0000000
5  0.2500000 0.0000000           62.2500             32.50000          
0.5450000     0.0000000
6  7.7500000 0.0000000           58.0000             42.75000          
0.6375000     0.5000000
  R_avg_TD_att R_avg_TD_landed R_avg_TD_pct R_avg_TOTAL_STR_att R_avg_
TOTAL_STR_landed
1    5.3000000        1.900000    0.4580000            129.9000       
69.1000
2    5.1428571        2.428571    0.6014286            161.5714       
102.8571
3    0.9333333        0.400000    0.2773333            133.0000       
63.4000
4    2.2500000        0.625000    0.1037500            117.3750       
50.7500
5    0.5000000        0.000000    0.0000000             63.5000       
32.7500
6    5.5000000        4.500000    0.8175000            101.5000       
80.5000
  R_longest_win_streak R_losses R_avg_opp_BODY_att R_avg_opp_BODY_land
ed R_avg_opp_CLINCH_att
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
1                    4        2           13.30000                8.8000
00               7.50000
2                    2        2           24.57143               14.1428
57              10.57143
3                   11        1           14.46667                8.1333
33               2.80000
4                    5        2           20.25000               13.3750
00               6.87500
5                    3        1            6.25000                4.7500
00               4.50000
6                    4        0            3.00000                2.2500
00               3.50000
  R_avg_opp_CLINCH_landed R_avg_opp_DISTANCE_att R_avg_opp_DISTANCE_la
nded R_avg_opp_GROUND_att
1               5.1000000               90.50000                  26.8
0000             0.800000
2               7.8571429               98.57143                  32.5
7143             6.428571
3               0.7333333               91.06667                  32.2
0000             4.866667
4               5.6250000              103.12500                  38.5
0000             0.875000
5               3.5000000               42.75000                  16.2
5000             7.750000
6               3.0000000                5.75000                   2.0
0000             2.000000
  R_avg_opp_GROUND_landed R_avg_opp_HEAD_att R_avg_opp_HEAD_landed R_a
vg_opp_KD R_avg_opp_LEG_att
1                0.300000           76.10000              17.30000
0.1000000           9.40000
2                4.285714           61.85714              12.42857
0.0000000          29.14286
3                2.800000           78.26667              23.20000
0.2666667           6.00000
4                0.750000           77.37500              20.37500
0.1250000          13.25000
5                2.750000           43.25000              14.00000
0.2500000           5.50000
6                1.500000            8.00000               4.00000
0.0000000           0.25000
  R_avg_opp_LEG_landed R_avg_opp_PASS R_avg_opp_REV R_avg_opp_SIG_STR_
att R_avg_opp_SIG_STR_landed
1              6.10000      0.0000000     0.0000000              98.80
000                 32.20000
```

```
2              18.14286      1.1428571     0.0000000              115.57
143                 44.71429
3               4.40000      0.3333333     0.1333333               98.73
333                 35.73333
4              11.12500      0.0000000     0.0000000              110.87
500                 44.87500
5               3.75000      0.7500000     0.0000000               55.00
000                 22.50000
6               0.25000      0.0000000     0.5000000               11.25
000                  6.50000
  R_avg_opp_SIG_STR_pct R_avg_opp_SUB_ATT R_avg_opp_TD_att R_avg_opp_T
D_landed R_avg_opp_TD_pct
1              0.3360000        0.00000000         0.900000           0
.1000000        0.0500000
2              0.4371429        0.28571429         3.285714           0
.8571429        0.1471429
3              0.3400000        0.06666667         2.866667           0
.6666667        0.1313333
4              0.4462500        0.00000000         2.375000           0
.0000000        0.0000000
5              0.3975000        0.00000000         1.000000           0
.0000000        0.0000000
6              0.5400000        0.75000000         0.500000           0
.0000000        0.0000000
  R_avg_opp_TOTAL_STR_att R_avg_opp_TOTAL_STR_landed R_total_rounds_fo
ught
1                 110.5000                   43.30000
27
2                 158.1429                   82.28571
25
3                 102.1333                   38.60000
33
4                 115.1250                   48.87500
20
5                  60.5000                   27.75000
7
6                  38.0000                   26.50000
8
  R_total_time_fought.seconds. R_total_title_bouts R_win_by_Decision_M
ajority
1                        742.60                   3
0
2                       1062.00                   2
0
```

```
3                       604.40                  2
0
4                       690.25                  0
0
5                       440.75                  0
0
6                       540.00                  1
0
  R_win_by_Decision_Split R_win_by_Decision_Unanimous R_win_by_KO.TKO
R_win_by_Submission
1                       2                           4               2
0
2                       1                           2               0
2
3                       1                           3               3
6
4                       1                           4               1
0
5                       0                           1               2
0
6                       0                           1               1
2
  R_win_by_TKO_Doctor_Stoppage R_wins R_Stance R_Height_cms R_Reach_cm
s R_Weight_lbs B_age R_age
1                            0      8 Orthodox       162.56      162.5
6         135    31    32
2                            0      5 Southpaw       165.10      167.6
4         125    32    31
3                            1     14 Orthodox       180.34      193.0
4         155    36    35
4                            0      6 Orthodox       162.56      172.7
2         135    26    29
5                            0      3 Southpaw       187.96      190.5
0         264    32    26
6                            0      4                165.10      167.6
4         115    33    28
>
> #Extracting relevant information of data needed for analysis
> #For this datset we will use the following columns extracted.
> #To extract certain columns, we will use the subset function.
> #Within the subset function, we need to specify the name of our data
matrix (i.e. df)
> #and the columns we want to select (i.e. 6 to 100)
> #We can also make use of column names to extract needed columns
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> #Notice that the data is extracted in our order of preference
> final_df <- subset(df, select = c(4, 6:12, 38, 63, 64, 65, 72, 73:76
, 144, 77:79 , 105, 130:132, 139:143, 145 ))
>
> #Display dataframe
> final_df
          date Winner title_bout          weight_class no_of_rounds B_c
urrent_lose_streak
1  2019-06-08    Red       True          Bantamweight            5
0
2  2019-06-08    Red       True     Women's Flyweight            5
0
3  2019-06-08    Red      False           Lightweight            3
0
4  2019-06-08   Blue      False          Bantamweight            3
0
5  2019-06-08   Blue      False           Heavyweight            3
0
6  2019-06-08    Red      False   Women's Strawweight            3
0
7  2019-06-08    Red      False          Bantamweight            3
0
8  2019-06-08   Blue      False   Women's Strawweight            3
1
9  2019-06-08   Blue      False         Featherweight            3
0
10 2019-06-08    Red      False   Women's Strawweight            3
0
11 2019-06-08   Blue      False          Middleweight            3
1
12 2019-06-08    Red      False          Bantamweight            3
0
13 2019-06-08    Red      False     Women's Flyweight            3
0
14 2019-06-01   Blue      False     Light Heavyweight            5
1
15 2019-06-01   Blue      False     Light Heavyweight            3
0
16 2019-06-01    Red      False         Featherweight            3
0
17 2019-06-01   Blue      False           Lightweight            3
0
18 2019-06-01    Red      False         Featherweight            3
0
```

```
19 2019-06-01   Blue      False          Welterweight            3
0
20 2019-06-01   Blue      False  Women's Bantamweight            3
1
21 2019-06-01   Blue      False           Lightweight            3
0
22 2019-06-01   Blue      False           Lightweight            3
2
23 2019-06-01    Red      False  Women's Bantamweight            3
0
24 2019-06-01   Blue      False     Light Heavyweight            3
1
25 2019-06-01    Red      False           Lightweight            3
0
26 2019-05-18    Red      False          Welterweight            5
1
27 2019-05-18   Blue      False          Middleweight            3
0
28 2019-05-18   Blue      False Women's Featherweight            3
0
29 2019-05-18    Red      False          Welterweight            3
0
30 2019-05-18    Red      False           Lightweight            3
0
31 2019-05-18    Red      False           Lightweight            3
0
32 2019-05-18    Red      False  Women's Bantamweight            3
0
   B_current_win_streak B_draw B_losses B_total_rounds_fought B_total_
time_fought.seconds.
1                     4      0        1                     9
419.4000
2                     3      0        6                    29
849.0000
3                     3      0        8                    68
581.8710
4                     4      0        0                     9
652.0000
5                     1      0        1                     8
1200.0000
6                     4      0        2                    18
886.5000
7                     3      0        4                    23
495.2500
```

```
8                    0      0        2                   10
716.0000
9                    1      0        1                   10
670.7500
10                   1      0        6                   26
763.1000
11                   0      0        5                   14
488.7143
12                   0      0        0                    0
NA
13                   2      0        3                   18
628.6250
14                   0      0        4                   27
625.0000
15                   3      0        0                    7
681.6667
16                   1      0        1                    3
310.5000
17                   1      0        3                   10
558.2000
18                   0      0        0                    0
NA
19                   0      0        0                    0
NA
20                   0      0        3                   13
720.4000
21                   4      0        1                   14
656.3333
22                   0      0        2                    8
727.6667
23                   0      0        0                    0
NA
24                   0      0        3                   13
620.6667
25                   0      0        0                    0
NA
26                   0      0        4                   36
724.8571
27                   1      0        0                    3
900.0000
28                   0      0        0                    0
NA
29                   0      0        0                    0
NA
```

```
30                   2      0        8                    59
754.9091
31                   0      0        0                     0
NA
32                   2      0        0                     6
900.0000
   B_total_title_bouts B_wins B_Stance B_Height_cms B_Reach_cms B_Weig
ht_lbs B_age
1                     0      4 Orthodox       167.64      170.18
135    31
2                     0      4 Orthodox       167.64      167.64
125    32
3                     1     23 Orthodox       185.42      185.42
155    36
4                     0      4   Switch       170.18      170.18
135    26
5                     0      1 Southpaw       180.34      185.42
250    32
6                     0      4 Orthodox       165.10      162.56
115    33
7                     0      8 Orthodox       167.64      165.10
135    32
8                     0      2 Orthodox       165.10      167.64
115    25
9                     0      3 Orthodox       180.34      182.88
145    31
10                    0      4 Orthodox       160.02      162.56
115    34
11                    0      2 Orthodox       182.88      187.96
185    28
12                    0      0   Switch       170.18      172.72
135    35
13                    0      5 Orthodox       167.64      165.10
125    33
14                    1      7 Orthodox       193.04      193.04
205    30
15                    0      3 Orthodox       193.04      198.12
205    27
16                    0      1 Orthodox       172.72      172.72
145    26
17                    0      2 Orthodox       177.80      180.34
155    29
18                    0      0 Orthodox       182.88      182.88
145    26
```

```
19                    0      0 Orthodox       185.42      187.96
170    27
20                    0      2 Orthodox       170.18      165.10
135    37
21                    1      5 Orthodox       182.88      190.50
155    39
22                    0      1 Orthodox       177.80      185.42
170    30
23                    0      0 Orthodox       170.18      185.42
145    23
24                    0      3 Orthodox       182.88      190.50
205    29
25                    0      0 Southpaw       182.88      187.96
155    24
26                    1     10 Orthodox       175.26      195.58
170    26
27                    0      1 Orthodox       180.34      182.88
185    30
28                    0      0 Orthodox       167.64      172.72
145    28
29                    0      0 Orthodox       177.80      182.88
170    31
30                    0     14 Orthodox       172.72      172.72
155    34
31                    0      0 Orthodox       177.80      180.34
155    27
32                    0      2                162.56      170.18
135    34
   R_current_lose_streak R_current_win_streak R_draw R_losses R_total_
rounds_fought
1                     0                    4      0        2
27
2                     0                    2      0        2
25
3                     0                   11      0        1
33
4                     1                    0      0        2
20
5                     1                    0      0        1
7
6                     0                    4      0        0
8
7                     0                    3      0        3
32
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
8                      2                    0      0        4
25
9                      0                    1      0        5
34
10                     0                    3      0        0
9
11                     1                    0      0        1
3
12                     2                    0      0        7
29
13                     1                    0      0        2
18
14                     1                    0      0        5
38
15                     3                    0      0        5
21
16                     0                    1      0        1
11
17                     0                    2      0        2
12
18                     3                    0      0        3
8
19                     0                    0      0        0
0
20                     2                    0      0        2
4
21                     0                    1      0        3
22
22                     2                    0      0        3
19
23                     0                    0      0        0
0
24                     0                    1      0        0
1
25                     1                    0      0        1
3
26                     2                    0      0        9
74
27                     0                    4      0        3
21
28                     0                    1      0        1
4
29                     0                    4      0        2
18
```

```
30                      0                    4      0        9
42
31                      0                    3      0        1
8
32                      0                    2      0        0
3
   R_total_time_fought.seconds. R_total_title_bouts R_wins R_Stance R_
Height_cms R_Reach_cms
1                       742.6000                   3       8 Orthodox
162.56      162.56
2                      1062.0000                   2       5 Southpaw
165.10      167.64
3                       604.4000                   2      14 Orthodox
180.34      193.04
4                       690.2500                   0       6 Orthodox
162.56      172.72
5                       440.7500                   0       3 Southpaw
187.96      190.50
6                       540.0000                   1       4
165.10      167.64
7                       750.6667                   0       9 Orthodox
170.18      180.34
8                       800.1111                   1       5 Orthodox
160.02      162.56
9                       624.0667                   1      10 Orthodox
172.72      180.34
10                      900.0000                   0       3 Orthodox
165.10      160.02
11                      692.0000                   0       0 Orthodox
190.50      200.66
12                      668.0000                   1       5 Orthodox
170.18      175.26
13                      900.0000                   0       4 Orthodox
175.26      172.72
14                      660.8667                   3      10 Orthodox
195.58      200.66
15                      455.7273                   0       6 Orthodox
185.42      200.66
16                      561.8000                   0       4 Southpaw
177.80      182.88
17                      610.8000                   0       3 Orthodox
175.26      177.80
18                      681.6667                   0       0 Orthodox
165.10      175.26
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
19                             NA                   0      0   Switch
177.80      182.88
20                       461.0000                   1      0 Orthodox
170.18      177.80
21                       686.8889                   0      6 Southpaw
177.80      177.80
22                       807.8571                   0      4 Southpaw
167.64      167.64
23                             NA                   0      0 Orthodox
175.26      187.96
24                       193.0000                   0      1 Orthodox
182.88      193.04
25                       900.0000                   0      0 Orthodox
190.50      195.58
26                       794.7692                   4     17 Southpaw
172.72      177.80
27                       634.2222                   1      6 Orthodox
187.96      200.66
28                       480.5000                   0      1 Orthodox
182.88      182.88
29                       453.4000                   0      8 Orthodox
180.34      190.50
30                       400.5652                   0     14 Orthodox
177.80      187.96
31                       480.5000                   0      3 Orthodox
167.64      177.80
32                       329.5000                   0      2 Orthodox
167.64      167.64
   R_Weight_lbs R_age
1           135    32
2           125    31
3           155    35
4           135    29
5           264    26
6           115    28
7           135    29
8           115    33
9           145    37
10          115    29
11          185    28
12          135    34
13          125    30
14          205    32
15          205    39
```

```
16          145    30
17          155    32
18          145    31
19          170    27
20          135    37
21          155    29
22          155    35
23          135    27
24          205    27
25          155    25
26          170    34
27          185    29
28          145    29
29          170    27
30          155    29
31          155    32
32          135    24
 [ reached 'max' / getOption("max.print") -- omitted 5112 rows ]
>
> #Save this data frame for future lab assignments using write.csv com
mands
> #To save a dataframe as CSV is easy.
> #Simply need to use the write.csv function with the name of the data
frame and the name of the file you want to write to
> write.csv(final_df, file = "df.csv")
>
> #Structure of dataframe
> str(final_df)
'data.frame':	5144 obs. of  31 variables:
 $ date                        : chr  "2019-06-08" "2019-06-08" "2019-
06-08" "2019-06-08" ...
 $ Winner                      : chr  "Red" "Red" "Red" "Blue" ...
 $ title_bout                  : chr  "True" "True" "False" "False" ..
.
 $ weight_class                : chr  "Bantamweight" "Women's Flyweigh
t" "Lightweight" "Bantamweight" ...
 $ no_of_rounds                : int  5 5 3 3 3 3 3 3 3 3 ...
 $ B_current_lose_streak       : num  0 0 0 0 0 0 0 1 0 0 ...
 $ B_current_win_streak        : num  4 3 3 4 1 4 3 0 1 1 ...
 $ B_draw                      : num  0 0 0 0 0 0 0 0 0 0 ...
 $ B_losses                    : num  1 6 8 0 1 2 4 2 1 6 ...
 $ B_total_rounds_fought       : num  9 29 68 9 8 18 23 10 10 26 ...
 $ B_total_time_fought.seconds.: num  419 849 582 652 1200 ...
 $ B_total_title_bouts         : num  0 0 1 0 0 0 0 0 0 0 ...
 $ B_wins                      : num  4 4 23 4 1 4 8 2 3 4 ...
 $ B_Stance                    : chr  "Orthodox" "Orthodox" "Orthodox"
"Switch" ...
 $ B_Height_cms                : num  168 168 185 170 180 ...
 $ B_Reach_cms                 : num  170 168 185 170 185 ...
 $ B_Weight_lbs                : num  135 125 155 135 250 115 135 115
145 115 ...
 $ B_age                       : num  31 32 36 26 32 33 32 25 31 34 ..
.
 $ R_current_lose_streak       : num  0 0 0 1 1 0 0 2 0 0 ...
 $ R_current_win_streak        : num  4 2 11 0 0 4 3 0 1 3 ...
 $ R_draw                      : num  0 0 0 0 0 0 0 0 0 0 ...
 $ R_losses                    : num  2 2 1 2 1 0 3 4 5 0 ...
 $ R_total_rounds_fought       : num  27 25 33 20 7 8 32 25 34 9 ...
 $ R_total_time_fought.seconds.: num  743 1062 604 690 441 ...
 $ R_total_title_bouts         : num  3 2 2 0 0 1 0 1 1 0 ...
 $ R_wins                      : num  8 5 14 6 3 4 9 5 10 3 ...
 $ R_Stance                    : chr  "Orthodox" "Southpaw" "Orthodox"
"Orthodox" ...
 $ R_Height_cms                : num  163 165 180 163 188 ...
 $ R_Reach_cms                 : num  163 168 193 173 190 ...
 $ R_Weight_lbs                : num  135 125 155 135 264 115 135 115
145 115 ...
 $ R_age                       : num  32 31 35 29 26 28 29 33 37 29 ..
.
>
> #Summary of entire dataframe
> print(summary(final_df))
     date               Winner            title_bout          weight_class
no_of_rounds
 Length:5144        Length:5144        Length:5144        Length:5144
Min.   :1.000
 Class :character   Class :character   Class :character   Class :chara
cter   1st Qu.:3.000
 Mode  :character   Mode  :character   Mode  :character   Mode  :chara
cter   Median :3.000

Mean   :3.119

3rd Qu.:3.000

Max.   :5.000
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
 B_current_lose_streak B_current_win_streak     B_draw     B_losses    
B_total_rounds_fought
 Min.   :0.0000        Min.   : 0.0000      Min.   :0   Min.   : 0.000   
Min.   : 0.000       
 1st Qu.:0.0000        1st Qu.: 0.0000      1st Qu.:0   1st Qu.: 0.000   
1st Qu.: 1.000       
 Median :0.0000        Median : 0.0000      Median :0   Median : 1.000   
Median : 5.000       
 Mean   :0.4298        Mean   : 0.8373      Mean   :0   Mean   : 1.464   
Mean   : 8.921       
 3rd Qu.:1.0000        3rd Qu.: 1.0000      3rd Qu.:0   3rd Qu.: 2.000   
3rd Qu.:13.000       
 Max.   :6.0000        Max.   :13.0000      Max.   :0   Max.   :13.000   
Max.   :75.000       
                                                                         
 B_total_time_fought.seconds. B_total_title_bouts     B_wins          B
_Stance         
 Min.   :   7.0               Min.   : 0.0000     Min.   : 0.000   Len
gth:5144        
 1st Qu.: 445.9               1st Qu.: 0.0000     1st Qu.: 0.000   Cla
ss :character  
 Median : 610.4               Median : 0.0000     Median : 1.000   Mod
e  :character  
 Mean   : 592.4               Mean   : 0.2799     Mean   : 2.484                      
 3rd Qu.: 767.2               3rd Qu.: 0.0000     3rd Qu.: 4.000                      
 Max.   :1500.0               Max.   :16.0000     Max.   :23.000                      
 NA's   :1265                                                                         
  B_Height_cms    B_Reach_cms     B_Weight_lbs        B_age       R_cur
rent_lose_streak
 Min.   :152.4   Min.   :152.4   Min.   :115.0   Min.   :18.00   Min.   
:0.0000       
 1st Qu.:172.7   1st Qu.:177.8   1st Qu.:145.0   1st Qu.:26.00   1st Q
u.:0.0000       
 Median :180.3   Median :182.9   Median :170.0   Median :29.00   Media
n :0.0000       
 Mean   :179.2   Mean   :183.3   Mean   :172.1   Mean   :29.17   Mean   
:0.5509       
 3rd Qu.:185.4   3rd Qu.:190.5   3rd Qu.:185.0   3rd Qu.:32.00   3rd Q
u.:1.0000       
 Max.   :210.8   Max.   :213.4   Max.   :770.0   Max.   :51.00   Max.   
:7.0000       
 NA's   :8       NA's   :666     NA's   :6       NA's   :172                      
 R_current_win_streak     R_draw     R_losses      R_total_rounds_foug
ht
```

```
 Min.   : 0          Min.   :0   Min.   : 0.000   Min.   : 0.00  
 1st Qu.: 0          1st Qu.:0   1st Qu.: 0.000   1st Qu.: 3.00  
 Median : 0          Median :0   Median : 1.000   Median : 9.00  
 Mean   : 1          Mean   :0   Mean   : 1.951   Mean   :12.85  
 3rd Qu.: 1          3rd Qu.:0   3rd Qu.: 3.000   3rd Qu.:19.00  
 Max.   :16          Max.   :0   Max.   :14.000   Max.   :80.00  
                                                                 
 R_total_time_fought.seconds. R_total_title_bouts     R_wins          R
_Stance
 Min.   :   7.0               Min.   : 0.000      Min.   : 0.000   Len
gth:5144
 1st Qu.: 470.6               1st Qu.: 0.000      1st Qu.: 1.000   Cla
ss :character
 Median : 620.3               Median : 0.000      Median : 2.000   Mod
e  :character
 Mean   : 603.8               Mean   : 0.597      Mean   : 3.598      
 3rd Qu.: 762.3               3rd Qu.: 1.000      3rd Qu.: 5.000      
 Max.   :1500.0               Max.   :16.000      Max.   :20.000      
 NA's   :650                                                          
  R_Height_cms    R_Reach_cms     R_Weight_lbs       R_age      
 Min.   :152.4   Min.   :152.4   Min.   :115.0   Min.   :19.00  
 1st Qu.:172.7   1st Qu.:177.8   1st Qu.:145.0   1st Qu.:26.00  
 Median :180.3   Median :182.9   Median :170.0   Median :29.00  
 Mean   :179.3   Mean   :183.7   Mean   :172.1   Mean   :29.44  
 3rd Qu.:185.4   3rd Qu.:190.5   3rd Qu.:185.0   3rd Qu.:32.00  
 Max.   :210.8   Max.   :213.4   Max.   :345.0   Max.   :47.00  
 NA's   :4       NA's   :316     NA's   :3       NA's   :64     
```

Conclusion:

Thus dataset has been selected from kaggle loaded into Rstudio.

Lab Outcome: LO6

Anirudh Poroorkara
Roll: 44
IT-1 B12

Lab Assignment: 08

Aim:

Cleaning of the Dataset

Code:

```
> #Lab Assignment 8
> #Mini-Project Session 2
> 
> #Cleaning of Dataset
> 
> #get the working directory
> getwd()
[1] "/cloud/project"
> 
> #Loading dataset into R
> #Header is set to to true if file contains Header information
> #Sep stands for seperator which is , in our csv file
> df <- read.csv("data.csv", header = TRUE, sep =",")
> 
> #Check the class before doing any cleaning
> class(df)
[1] "data.frame"
> 
> #Check the number of rows and columns the data frame has
> dim(df)
[1] 5144  145
> #We can see that the data frame has 5144 rows and 32 columns
> 
> #Finding summary of entire dataset
> summary(df)
  R_fighter          B_fighter           Referee              date            
location          
 Length:5144        Length:5144        Length:5144        Length:5144        
Length:5144       
 Class :character   Class :character   Class :character   Class :chara
cter   Class :character  
 Mode  :character   Mode  :character   Mode  :character   Mode  :chara
cter   Mode  :character  
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
    Winner           title_bout        weight_class        no_of_round
s   B_current_lose_streak
 Length:5144        Length:5144        Length:5144        Min.   :1.00
0   Min.   :0.0000
 Class :character   Class :character   Class :character   1st Qu.:3.00
0   1st Qu.:0.0000
 Mode  :character   Mode  :character   Mode  :character   Median :3.00
0   Median :0.0000
                                                          Mean   :3.11
9   Mean   :0.4298
                                                          3rd Qu.:3.00
0   3rd Qu.:1.0000
                                                          Max.   :5.00
0   Max.   :6.0000
 B_current_win_streak     B_draw  B_avg_BODY_att   B_avg_BODY_landed B
_avg_CLINCH_att
 Min.   : 0.0000      Min.   :0   Min.   : 0.000   Min.   : 0.000    M
in.   : 0.000
 1st Qu.: 0.0000      1st Qu.:0   1st Qu.: 3.500   1st Qu.: 2.333    1
st Qu.: 3.000
 Median : 0.0000      Median :0   Median : 7.000   Median : 5.000    M
edian : 6.333
 Mean   : 0.8373      Mean   :0   Mean   : 8.689   Mean   : 6.083    M
ean   : 8.241
 3rd Qu.: 1.0000      3rd Qu.:0   3rd Qu.:12.225   3rd Qu.: 8.500    3
rd Qu.:11.422
 Max.   :13.0000      Max.   :0   Max.   :49.000   Max.   :39.000    M
ax.   :87.000
 B_avg_CLINCH_landed B_avg_DISTANCE_att B_avg_DISTANCE_landed B_avg_GR
OUND_att B_avg_GROUND_landed
 Min.   : 0.000      Min.   :  0.00     Min.   :  0.000       Min.   :
0.000   Min.   : 0.000
 1st Qu.: 2.000      1st Qu.: 22.00     1st Qu.:  7.667       1st Qu.:
2.500   1st Qu.: 1.667
 Median : 4.200      Median : 44.67     Median : 15.200       Median :
6.500   Median : 4.333
 Mean   : 5.556      Mean   : 53.16     Mean   : 19.329       Mean   :
8.754   Mean   : 5.773
 3rd Qu.: 7.739      3rd Qu.: 74.33     3rd Qu.: 27.143       3rd Qu.:
12.167   3rd Qu.: 8.000
 Max.   :68.000      Max.   :271.00     Max.   :130.000       Max.   :
88.000   Max.   :47.000
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
 B_avg_HEAD_att   B_avg_HEAD_landed    B_avg_KD      B_avg_LEG_att   
B_avg_LEG_landed
 Min.   :  0.00   Min.   :  0.00    Min.   :0.0000   Min.   : 0.000   
Min.   : 0.000  
 1st Qu.: 29.41   1st Qu.: 11.00    1st Qu.:0.0000   1st Qu.: 2.000   
1st Qu.: 1.500  
 Median : 49.00   Median : 17.75    Median :0.1111   Median : 4.500   
Median : 3.600  
 Mean   : 55.45   Mean   : 19.82    Mean   :0.2464   Mean   : 6.009   
Mean   : 4.759  
 3rd Qu.: 74.69   3rd Qu.: 26.00    3rd Qu.:0.4000   3rd Qu.: 8.500   
3rd Qu.: 6.750  
 Max.   :277.00   Max.   :137.00    Max.   :5.0000   Max.   :61.000   
Max.   :47.000  
   B_avg_PASS        B_avg_REV      B_avg_SIG_STR_att B_avg_SIG_STR_lan
ded B_avg_SIG_STR_pct
 Min.   : 0.000   Min.   :0.0000   Min.   :  0.00    Min.   :  0.00   
Min.   :0.0000   
 1st Qu.: 0.200   1st Qu.:0.0000   1st Qu.: 38.50    1st Qu.: 17.25   
1st Qu.:0.3875   
 Median : 1.000   Median :0.0000   Median : 63.00    Median : 28.26   
Median :0.4550   
 Mean   : 1.278   Mean   :0.1692   Mean   : 70.15    Mean   : 30.66   
Mean   :0.4563   
 3rd Qu.: 1.879   3rd Qu.:0.2500   3rd Qu.: 94.53    3rd Qu.: 40.83   
3rd Qu.:0.5240   
 Max.   :15.000   Max.   :3.0000   Max.   :299.00    Max.   :154.00   
Max.   :1.0000   
 B_avg_SUB_ATT     B_avg_TD_att     B_avg_TD_landed  B_avg_TD_pct     B
_avg_TOTAL_STR_att
 Min.   :0.0000   Min.   : 0.0000   Min.   : 0.00   Min.   :0.0000   M
in.   :  0.00   
 1st Qu.:0.0000   1st Qu.: 0.8819   1st Qu.: 0.25   1st Qu.:0.1000   1
st Qu.: 57.78   
 Median :0.3333   Median : 2.0000   Median : 1.00   Median :0.3000   M
edian : 87.50   
 Mean   :0.5481   Mean   : 2.8049   Mean   : 1.19   Mean   :0.3155   M
ean   : 92.64   
 3rd Qu.:0.9000   3rd Qu.: 4.0000   3rd Qu.: 1.75   3rd Qu.:0.4861   3
rd Qu.:122.65   
 Max.   :8.0000   Max.   :19.0000   Max.   :10.00   Max.   :1.0000   M
ax.   :360.00   
 B_avg_TOTAL_STR_landed B_longest_win_streak    B_losses      B_avg_op
p_BODY_att
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
 Min.   :  0.00         Min.   : 0.000       Min.   : 0.000   Min.   : 
0.000  
 1st Qu.: 31.00         1st Qu.: 0.000       1st Qu.: 0.000   1st Qu.: 
3.358  
 Median : 47.67         Median : 1.000       Median : 1.000   Median : 
7.000  
 Mean   : 50.71         Mean   : 1.588       Mean   : 1.464   Mean   : 
8.297  
 3rd Qu.: 66.50         3rd Qu.: 2.000       3rd Qu.: 2.000   3rd Qu.:
11.639  
 Max.   :230.00         Max.   :16.000       Max.   :13.000   Max.   :
61.000  
 B_avg_opp_BODY_landed B_avg_opp_CLINCH_att B_avg_opp_CLINCH_landed B_
avg_opp_DISTANCE_att
 Min.   : 0.000        Min.   :  0.000      Min.   : 0.000          Mi
n.   :  0.00       
 1st Qu.: 2.333        1st Qu.:  2.857      1st Qu.: 1.667          1s
t Qu.: 21.41       
 Median : 4.600        Median :  5.909      Median : 3.667          Me
dian : 43.67       
 Mean   : 5.639        Mean   :  7.455      Mean   : 4.922          Me
an   : 51.88       
 3rd Qu.: 7.714        3rd Qu.: 10.000      3rd Qu.: 6.667          3r
d Qu.: 71.78       
 Max.   :48.000        Max.   :105.000      Max.   :84.000          Ma
x.   :361.00       
 B_avg_opp_DISTANCE_landed B_avg_opp_GROUND_att B_avg_opp_GROUND_lande
d B_avg_opp_HEAD_att
 Min.   :  0.00            Min.   : 0.000       Min.   : 0.000         
Min.   :  0.00    
 1st Qu.:  7.00            1st Qu.: 1.667       1st Qu.: 1.000         
1st Qu.: 27.00    
 Median : 14.67            Median : 4.714       Median : 3.000         
Median : 46.57    
 Mean   : 18.23            Mean   : 7.100       Mean   : 4.633         
Mean   : 52.09    
 3rd Qu.: 25.32            3rd Qu.: 9.454       3rd Qu.: 6.000         
3rd Qu.: 69.20    
 Max.   :150.00            Max.   :94.000       Max.   :84.000         
Max.   :335.00    
 B_avg_opp_HEAD_landed  B_avg_opp_KD    B_avg_opp_LEG_att B_avg_opp_LE
G_landed B_avg_opp_PASS  
 Min.   :  0.000       Min.   :0.0000   Min.   : 0.000    Min.   : 0.0
00       Min.   : 0.0000 
```

```
 1st Qu.:  8.857       1st Qu.:0.0000   1st Qu.: 2.143    1st Qu.: 1.8
00       1st Qu.: 0.0000  
 Median : 15.000       Median :0.0000   Median : 4.800    Median : 3.7
50       Median : 0.6667  
 Mean   : 17.368       Mean   :0.1609   Mean   : 6.048    Mean   : 4.7
74       Mean   : 1.0892  
 3rd Qu.: 23.000       3rd Qu.:0.2500   3rd Qu.: 8.114    3rd Qu.: 6.5
64       3rd Qu.: 1.5000  
 Max.   :126.000       Max.   :3.0000   Max.   :57.000    Max.   :50.0
00       Max.   :19.0000  
 B_avg_opp_REV    B_avg_opp_SIG_STR_att B_avg_opp_SIG_STR_landed B_avg
_opp_SIG_STR_pct
 Min.   :0.0000   Min.   :  0.00        Min.   :  0.00           Min.   
:0.0000       
 1st Qu.:0.0000   1st Qu.: 35.35        1st Qu.: 15.00           1st Q
u.:0.3533       
 Median :0.0000   Median : 60.25        Median : 25.07           Media
n :0.4157       
 Mean   :0.1536   Mean   : 66.43        Mean   : 27.78           Mean   
:0.4272       
 3rd Qu.:0.2000   3rd Qu.: 87.45        3rd Qu.: 36.55           3rd Q
u.:0.4946       
 Max.   :3.0000   Max.   :401.00        Max.   :202.00           Max.   
:1.0000       
 B_avg_opp_SUB_ATT B_avg_opp_TD_att B_avg_opp_TD_landed B_avg_opp_TD_p
ct B_avg_opp_TOTAL_STR_att
 Min.   :0.0000    Min.   : 0.000   Min.   : 0.0000     Min.   :0.0000   
Min.   :  0.00         
 1st Qu.:0.0000    1st Qu.: 1.000   1st Qu.: 0.2817     1st Qu.:0.0827   
1st Qu.: 54.00         
 Median :0.2500    Median : 2.364   Median : 0.8571     Median :0.2394   
Median : 83.00         
 Mean   :0.4625    Mean   : 2.901   Mean   : 1.0941     Mean   :0.2765   
Mean   : 86.98         
 3rd Qu.:0.6667    3rd Qu.: 4.000   3rd Qu.: 1.5000     3rd Qu.:0.3928   
3rd Qu.:113.63         
 Max.   :7.0000    Max.   :20.000   Max.   :11.5000     Max.   :1.0000   
Max.   :404.00         
 B_avg_opp_TOTAL_STR_landed B_total_rounds_fought B_total_time_fought.
seconds. B_total_title_bouts
 Min.   :  0.00             Min.   : 0.000        Min.   :   7.0              
Min.   : 0.0000     
 1st Qu.: 28.68             1st Qu.: 1.000        1st Qu.: 445.9              
1st Qu.: 0.0000     
```

```
 Median : 43.33            Median : 5.000        Median : 610.4   
Median : 0.0000  
 Mean   : 46.16            Mean   : 8.921        Mean   : 592.4   
Mean   : 0.2799  
 3rd Qu.: 59.00            3rd Qu.:13.000        3rd Qu.: 767.2   
3rd Qu.: 0.0000  
 Max.   :232.00            Max.   :75.000        Max.   :1500.0   
Max.   :16.0000  
 B_win_by_Decision_Majority B_win_by_Decision_Split B_win_by_Decision_
Unanimous B_win_by_KO.TKO 
 Min.   :0.00000            Min.   :0.0000          Min.   : 0.0000   
Min.   : 0.0000  
 1st Qu.:0.00000            1st Qu.:0.0000          1st Qu.: 0.0000   
1st Qu.: 0.0000  
 Median :0.00000            Median :0.0000          Median : 0.0000   
Median : 0.0000  
 Mean   :0.01691            Mean   :0.2123          Mean   : 0.7801   
Mean   : 0.8727  
 3rd Qu.:0.00000            3rd Qu.:0.0000          3rd Qu.: 1.0000   
3rd Qu.: 1.0000  
 Max.   :2.00000            Max.   :5.0000          Max.   :10.0000   
Max.   :11.0000  
 B_win_by_Submission B_win_by_TKO_Doctor_Stoppage     B_wins          B
_Stance         
 Min.   : 0.0000     Min.   :0.0000               Min.   : 0.000   Len
gth:5144        
 1st Qu.: 0.0000     1st Qu.:0.0000               1st Qu.: 0.000   Cla
ss :character   
 Median : 0.0000     Median :0.0000               Median : 1.000   Mod
e  :character   
 Mean   : 0.5515     Mean   :0.0453               Mean   : 2.484                     
 3rd Qu.: 1.0000     3rd Qu.:0.0000               3rd Qu.: 4.000                     
 Max.   :11.0000     Max.   :2.0000               Max.   :23.000                     
  B_Height_cms    B_Reach_cms     B_Weight_lbs   R_current_lose_streak 
R_current_win_streak
 Min.   :152.4   Min.   :152.4   Min.   :115.0   Min.   :0.0000        
Min.   : 0          
 1st Qu.:172.7   1st Qu.:177.8   1st Qu.:145.0   1st Qu.:0.0000        
1st Qu.: 0          
 Median :180.3   Median :182.9   Median :170.0   Median :0.0000        
Median : 0          
 Mean   :179.2   Mean   :183.3   Mean   :172.1   Mean   :0.5509        
Mean   : 1          
```

```
 3rd Qu.:185.4   3rd Qu.:190.5   3rd Qu.:185.0   3rd Qu.:1.0000   
3rd Qu.: 1  
 Max.   :210.8   Max.   :213.4   Max.   :770.0   Max.   :7.0000   
Max.   :16  
     R_draw  R_avg_BODY_att   R_avg_BODY_landed R_avg_CLINCH_att R_avg
_CLINCH_landed
 Min.   :0   Min.   : 0.000   Min.   : 0.000    Min.   : 0.000   Min.   
: 0.000   
 1st Qu.:0   1st Qu.: 4.000   1st Qu.: 2.800    1st Qu.: 3.400   1st Q
u.: 2.000   
 Median :0   Median : 7.297   Median : 5.000    Median : 6.545   Media
n : 4.333   
 Mean   :0   Mean   : 8.811   Mean   : 6.138    Mean   : 8.174   Mean   
: 5.525   
 3rd Qu.:0   3rd Qu.:12.000   3rd Qu.: 8.500    3rd Qu.:11.000   3rd Q
u.: 7.444   
 Max.   :0   Max.   :51.000   Max.   :39.000    Max.   :82.000   Max.   
:52.000   
 R_avg_DISTANCE_att R_avg_DISTANCE_landed R_avg_GROUND_att R_avg_GROUN
D_landed R_avg_HEAD_att  
 Min.   :  0.00     Min.   :  0.000       Min.   : 0.000   Min.   : 0.
000      Min.   :  0.00  
 1st Qu.: 24.14     1st Qu.:  8.437       1st Qu.: 3.333   1st Qu.: 2.
146      1st Qu.: 32.67  
 Median : 46.00     Median : 16.000       Median : 7.439   Median : 5.
000      Median : 50.03  
 Mean   : 53.58     Mean   : 19.551       Mean   : 9.557   Mean   : 6.
325      Mean   : 56.17  
 3rd Qu.: 74.47     3rd Qu.: 27.851       3rd Qu.:13.000   3rd Qu.: 8.
587      3rd Qu.: 75.00  
 Max.   :287.50     Max.   :131.000       Max.   :96.000   Max.   :62.
000      Max.   :264.00  
 R_avg_HEAD_landed    R_avg_KD      R_avg_LEG_att    R_avg_LEG_landed 
R_avg_PASS    
 Min.   :  0.00    Min.   :0.0000   Min.   : 0.000   Min.   : 0.000   
Min.   : 0.0000  
 1st Qu.: 12.17    1st Qu.:0.0000   1st Qu.: 2.000   1st Qu.: 1.667   
1st Qu.: 0.3333  
 Median : 18.45    Median :0.1667   Median : 4.786   Median : 3.905   
Median : 1.0000  
 Mean   : 20.27    Mean   :0.2577   Mean   : 6.329   Mean   : 4.997   
Mean   : 1.4136  
 3rd Qu.: 26.33    3rd Qu.:0.4000   3rd Qu.: 8.990   3rd Qu.: 7.000   
3rd Qu.: 2.0000  
```

```
 Max.   :119.00    Max.   :4.0000   Max.   :63.000   Max.   :44.000   
Max.   :14.0000  
   R_avg_REV      R_avg_SIG_STR_att R_avg_SIG_STR_landed R_avg_SIG_STR
_pct R_avg_SUB_ATT   
 Min.   :0.0000   Min.   :  0.00    Min.   :  0.00       Min.   :0.000
0    Min.   :0.0000  
 1st Qu.:0.0000   1st Qu.: 42.52    1st Qu.: 19.50       1st Qu.:0.393
3    1st Qu.:0.0000  
 Median :0.0000   Median : 64.18    Median : 28.71       Median :0.457
0    Median :0.3750  
 Mean   :0.1533   Mean   : 71.31    Mean   : 31.40       Mean   :0.463
8    Mean   :0.5431  
 3rd Qu.:0.2133   3rd Qu.: 95.25    3rd Qu.: 41.50       3rd Qu.:0.530
0    3rd Qu.:0.8451  
 Max.   :3.0000   Max.   :298.50    Max.   :141.00       Max.   :1.000
0    Max.   :9.0000  
  R_avg_TD_att    R_avg_TD_landed   R_avg_TD_pct    R_avg_TOTAL_STR_at
t R_avg_TOTAL_STR_landed
 Min.   : 0.000   Min.   : 0.000   Min.   :0.0000   Min.   :  0.00    
Min.   :  0.00        
 1st Qu.: 1.000   1st Qu.: 0.375   1st Qu.:0.1650   1st Qu.: 61.16    
1st Qu.: 33.61        
 Median : 2.333   Median : 1.000   Median :0.3221   Median : 89.79    
Median : 48.52        
 Mean   : 3.027   Mean   : 1.305   Mean   :0.3336   Mean   : 93.84    
Mean   : 51.52        
 3rd Qu.: 4.333   3rd Qu.: 2.000   3rd Qu.:0.4900   3rd Qu.:121.00    
3rd Qu.: 65.40        
 Max.   :30.000   Max.   :11.000   Max.   :1.0000   Max.   :325.50    
Max.   :202.50        
 R_longest_win_streak    R_losses      R_avg_opp_BODY_att R_avg_opp_BO
DY_landed R_avg_opp_CLINCH_att
 Min.   : 0.000       Min.   : 0.000   Min.   : 0.000     Min.   : 0.0
00        Min.   : 0.000      
 1st Qu.: 1.000       1st Qu.: 0.000   1st Qu.: 3.600     1st Qu.: 2.5
00        1st Qu.: 3.000      
 Median : 2.000       Median : 1.000   Median : 7.000     Median : 4.8
00        Median : 6.000      
 Mean   : 2.254       Mean   : 1.951   Mean   : 8.157     Mean   : 5.4
99        Mean   : 7.345      
 3rd Qu.: 3.000       3rd Qu.: 3.000   3rd Qu.:11.495     3rd Qu.: 7.5
00        3rd Qu.: 9.744      
 Max.   :16.000       Max.   :14.000   Max.   :75.000     Max.   :41.0
00        Max.   :82.000      
```

```
 R_avg_opp_CLINCH_landed R_avg_opp_DISTANCE_att R_avg_opp_DISTANCE_lan
ded R_avg_opp_GROUND_att
 Min.   : 0.000          Min.   :  0.00         Min.   :  0.00
Min.   :  0.000
 1st Qu.: 2.000          1st Qu.: 23.50         1st Qu.:  7.50
1st Qu.:  1.750
 Median : 3.833          Median : 45.00         Median : 14.81
Median :  4.750
 Mean   : 4.813          Mean   : 52.16         Mean   : 17.89
Mean   :  6.611
 3rd Qu.: 6.352          3rd Qu.: 73.00         3rd Qu.: 25.00
3rd Qu.:  9.000
 Max.   :51.000          Max.   :440.00         Max.   :144.00
Max.   :104.000
 R_avg_opp_GROUND_landed R_avg_opp_HEAD_att R_avg_opp_HEAD_landed  R_a
vg_opp_KD   R_avg_opp_LEG_att
 Min.   : 0.000          Min.   :  0.00     Min.   :  0.00        Min.
:0.000   Min.   : 0.000
 1st Qu.: 1.000          1st Qu.: 28.25     1st Qu.:  9.00        1st
Qu.:0.000   1st Qu.: 2.431
 Median : 3.000          Median : 46.67     Median : 14.83        Medi
an :0.000   Median : 4.931
 Mean   : 4.273          Mean   : 52.00     Mean   : 16.86        Mean
:0.154   Mean   : 5.958
 3rd Qu.: 5.893          3rd Qu.: 69.33     3rd Qu.: 22.48        3rd
Qu.:0.250   3rd Qu.: 8.000
 Max.   :53.000          Max.   :400.00     Max.   :132.00        Max.
:3.000   Max.   :63.000
 R_avg_opp_LEG_landed R_avg_opp_PASS    R_avg_opp_REV    R_avg_opp_SIG
_STR_att
 Min.   : 0.000       Min.   : 0.0000   Min.   :0.0000   Min.   :  0.0
0
 1st Qu.: 2.000       1st Qu.: 0.1667   1st Qu.:0.0000   1st Qu.: 36.7
9
 Median : 3.800       Median : 0.6667   Median :0.0000   Median : 60.5
3
 Mean   : 4.615       Mean   : 1.0138   Mean   :0.1603   Mean   : 66.1
2
 3rd Qu.: 6.333       3rd Qu.: 1.4000   3rd Qu.:0.2222   3rd Qu.: 88.2
6
 Max.   :41.000       Max.   :17.0000   Max.   :3.0000   Max.   :454.0
0
 R_avg_opp_SIG_STR_landed R_avg_opp_SIG_STR_pct R_avg_opp_SUB_ATT R_av
g_opp_TD_att
```

```
 Min.   :  0.00          Min.   :0.0000        Min.   :0.0000    Min.   
: 0.000  
 1st Qu.: 15.62          1st Qu.:0.3440        1st Qu.:0.0000    1st 
Qu.: 1.000  
 Median : 24.67          Median :0.4091        Median :0.2727    Medi
an : 2.400  
 Mean   : 26.97          Mean   :0.4133        Mean   :0.4491    Mean   
: 2.831  
 3rd Qu.: 35.58          3rd Qu.:0.4743        3rd Qu.:0.6667    3rd 
Qu.: 4.000  
 Max.   :151.00          Max.   :1.0000        Max.   :8.0000    Max.   
:22.000  
 R_avg_opp_TD_landed R_avg_opp_TD_pct R_avg_opp_TOTAL_STR_att R_avg_op
p_TOTAL_STR_landed
 Min.   : 0.0000     Min.   :0.0000   Min.   :  0.00         Min.   :  
0.00  
 1st Qu.: 0.3333     1st Qu.:0.0900   1st Qu.: 54.50         1st Qu.: 
29.33  
 Median : 0.8000     Median :0.2246   Median : 82.79         Median : 
42.75  
 Mean   : 1.0230     Mean   :0.2579   Mean   : 86.28         Mean   : 
45.03  
 3rd Qu.: 1.5000     3rd Qu.:0.3697   3rd Qu.:111.00         3rd Qu.: 
57.31  
 Max.   :11.0000     Max.   :1.0000   Max.   :461.00         Max.   :
202.00  
 R_total_rounds_fought R_total_time_fought.seconds. R_total_title_bout
s R_win_by_Decision_Majority
 Min.   : 0.00         Min.   :   7.0               Min.   : 0.000    
Min.   :0.00000  
 1st Qu.: 3.00         1st Qu.: 470.6               1st Qu.: 0.000    
1st Qu.:0.00000  
 Median : 9.00         Median : 620.3               Median : 0.000    
Median :0.00000  
 Mean   :12.85         Mean   : 603.8               Mean   : 0.597    
Mean   :0.02761  
 3rd Qu.:19.00         3rd Qu.: 762.3               3rd Qu.: 1.000    
3rd Qu.:0.00000  
 Max.   :80.00         Max.   :1500.0               Max.   :16.000    
Max.   :2.00000  
 R_win_by_Decision_Split R_win_by_Decision_Unanimous R_win_by_KO.TKO 
R_win_by_Submission
 Min.   :0.0000          Min.   : 0.000             Min.   : 0.000  
Min.   : 0.0000  
```

```
 1st Qu.:0.0000          1st Qu.: 0.000           1st Qu.: 0.000  
1st Qu.: 0.0000  
 Median :0.0000          Median : 1.000           Median : 1.000  
Median : 0.0000  
 Mean   :0.2809          Mean   : 1.177           Mean   : 1.255  
Mean   : 0.7776  
 3rd Qu.:0.0000          3rd Qu.: 2.000           3rd Qu.: 2.000  
3rd Qu.: 1.0000  
 Max.   :5.0000          Max.   :10.000           Max.   :11.000  
Max.   :13.0000  
 R_win_by_TKO_Doctor_Stoppage     R_wins          R_Stance           R_H
eight_cms    R_Reach_cms   
 Min.   :0.00000              Min.   : 0.000   Length:5144        Min.   
:152.4   Min.   :152.4  
 1st Qu.:0.00000              1st Qu.: 1.000   Class :character   1st 
Qu.:172.7   1st Qu.:177.8  
 Median :0.00000              Median : 2.000   Mode  :character   Medi
an :180.3   Median :182.9  
 Mean   :0.07135              Mean   : 3.598                      Mean   
:179.3   Mean   :183.7  
 3rd Qu.:0.00000              3rd Qu.: 5.000                      3rd 
Qu.:185.4   3rd Qu.:190.5  
 Max.   :2.00000              Max.   :20.000                      Max.   
:210.8   Max.   :213.4  
  R_Weight_lbs       B_age           R_age      
 Min.   :115.0   Min.   :18.00   Min.   :19.00  
 1st Qu.:145.0   1st Qu.:26.00   1st Qu.:26.00  
 Median :170.0   Median :29.00   Median :29.00  
 Mean   :172.1   Mean   :29.17   Mean   :29.44  
 3rd Qu.:185.0   3rd Qu.:32.00   3rd Qu.:32.00  
 Max.   :345.0   Max.   :51.00   Max.   :47.00  
 [ reached getOption("max.print") -- omitted 1 row ]
> 
> #Let us now check for missing values in the entire dataset column by 
column
> #This function returns the number of all missing values in each of t
he columns
> colSums(is.na(df))
                   R_fighter                     B_fighter                       
Referee 
                           0                             0                             
0 
                        date                      location                        
Winner 
```

```
                              0                             0
0
                     title_bout                  weight_class
no_of_rounds
                              0                             0
0
          B_current_lose_streak          B_current_win_streak
B_draw
                              0                             0
0
                   B_avg_BODY_att              B_avg_BODY_landed
B_avg_CLINCH_att
                           1265                          1265
1265
            B_avg_CLINCH_landed           B_avg_DISTANCE_att          B_avg
_DISTANCE_landed
                           1265                          1265
1265
               B_avg_GROUND_att           B_avg_GROUND_landed
B_avg_HEAD_att
                           1265                          1265
1265
              B_avg_HEAD_landed                      B_avg_KD
B_avg_LEG_att
                           1265                          1265
1265
               B_avg_LEG_landed                    B_avg_PASS
B_avg_REV
                           1265                          1265
1265
              B_avg_SIG_STR_att          B_avg_SIG_STR_landed             B
_avg_SIG_STR_pct
                           1265                          1265
1265
                  B_avg_SUB_ATT                   B_avg_TD_att
B_avg_TD_landed
                           1265                          1265
1265
                   B_avg_TD_pct           B_avg_TOTAL_STR_att         B_avg_
TOTAL_STR_landed
                           1265                          1265
1265
           B_longest_win_streak                      B_losses            B_
avg_opp_BODY_att
```

```
                           0                            0 
1265 
       B_avg_opp_BODY_landed         B_avg_opp_CLINCH_att       B_avg_o
pp_CLINCH_landed 
                        1265                         1265 
1265 
      B_avg_opp_DISTANCE_att    B_avg_opp_DISTANCE_landed          B_av
g_opp_GROUND_att 
                        1265                         1265 
1265 
     B_avg_opp_GROUND_landed           B_avg_opp_HEAD_att         B_avg
_opp_HEAD_landed 
                        1265                         1265 
1265 
                B_avg_opp_KD            B_avg_opp_LEG_att          B_av
g_opp_LEG_landed 
                        1265                         1265 
1265 
              B_avg_opp_PASS                B_avg_opp_REV         B_avg
_opp_SIG_STR_att 
                        1265                         1265 
1265 
    B_avg_opp_SIG_STR_landed        B_avg_opp_SIG_STR_pct             B
_avg_opp_SUB_ATT 
                        1265                         1265 
1265 
            B_avg_opp_TD_att          B_avg_opp_TD_landed              
B_avg_opp_TD_pct 
                        1265                         1265 
1265 
     B_avg_opp_TOTAL_STR_att   B_avg_opp_TOTAL_STR_landed         B_tot
al_rounds_fought 
                        1265                         1265 
0 
B_total_time_fought.seconds.          B_total_title_bouts    B_win_by_D
ecision_Majority 
                        1265                            0 
0 
     B_win_by_Decision_Split  B_win_by_Decision_Unanimous              
B_win_by_KO.TKO 
                           0                            0 
0 
        B_win_by_Submission B_win_by_TKO_Doctor_Stoppage              
B_wins 
```

```
                           0                            0
0
                    B_Stance                 B_Height_cms
B_Reach_cms
                           0                            8
666
                B_Weight_lbs        R_current_lose_streak          R_cu
rrent_win_streak
                           6                            0
0
                      R_draw               R_avg_BODY_att             R
_avg_BODY_landed
                           0                          650
650
             R_avg_CLINCH_att          R_avg_CLINCH_landed            R_
avg_DISTANCE_att
                         650                          650
650
        R_avg_DISTANCE_landed             R_avg_GROUND_att           R_a
vg_GROUND_landed
                         650                          650
650
               R_avg_HEAD_att            R_avg_HEAD_landed
R_avg_KD
                         650                          650
650
                R_avg_LEG_att             R_avg_LEG_landed
R_avg_PASS
                         650                          650
650
                   R_avg_REV           R_avg_SIG_STR_att          R_av
g_SIG_STR_landed
                         650                          650
650
            R_avg_SIG_STR_pct               R_avg_SUB_ATT
R_avg_TD_att
                         650                          650
650
              R_avg_TD_landed                 R_avg_TD_pct           R_a
vg_TOTAL_STR_att
                         650                          650
650
       R_avg_TOTAL_STR_landed         R_longest_win_streak
R_losses
```

```
                          650                           0
0
         R_avg_opp_BODY_att        R_avg_opp_BODY_landed          R_av
g_opp_CLINCH_att
                          650                         650
650
     R_avg_opp_CLINCH_landed       R_avg_opp_DISTANCE_att     R_avg_opp
_DISTANCE_landed
                          650                         650
650
        R_avg_opp_GROUND_att      R_avg_opp_GROUND_landed             R_
avg_opp_HEAD_att
                          650                         650
650
       R_avg_opp_HEAD_landed                 R_avg_opp_KD             R
_avg_opp_LEG_att
                          650                         650
650
        R_avg_opp_LEG_landed               R_avg_opp_PASS
R_avg_opp_REV
                          650                         650
650
       R_avg_opp_SIG_STR_att     R_avg_opp_SIG_STR_landed         R_avg
_opp_SIG_STR_pct
                          650                         650
650
            R_avg_opp_SUB_ATT              R_avg_opp_TD_att           R_a
vg_opp_TD_landed
                          650                         650
650
             R_avg_opp_TD_pct      R_avg_opp_TOTAL_STR_att   R_avg_opp_
TOTAL_STR_landed
                          650                         650
650
       R_total_rounds_fought R_total_time_fought.seconds.           R_t
otal_title_bouts
                            0                         650
0
  R_win_by_Decision_Majority      R_win_by_Decision_Split  R_win_by_De
cision_Unanimous
                            0                           0
0
              R_win_by_KO.TKO          R_win_by_Submission R_win_by_TKO
_Doctor_Stoppage
```

```
                              0                             0
0
                         R_wins                      R_Stance
R_Height_cms
                              0                             0
4
                    R_Reach_cms                  R_Weight_lbs
B_age
                            316                             3
172
                          R_age
                             64
>
> #Store and identify all columns that contain missing values.
> list_na <- colnames(df)[ apply(df, 2, anyNA) ]
> #Display list
> list_na
  [1] "B_avg_BODY_att"              "B_avg_BODY_landed"           "B
_avg_CLINCH_att"
  [4] "B_avg_CLINCH_landed"         "B_avg_DISTANCE_att"          "B
_avg_DISTANCE_landed"
  [7] "B_avg_GROUND_att"            "B_avg_GROUND_landed"         "B
_avg_HEAD_att"
 [10] "B_avg_HEAD_landed"           "B_avg_KD"                    "B
_avg_LEG_att"
 [13] "B_avg_LEG_landed"            "B_avg_PASS"                  "B
_avg_REV"
 [16] "B_avg_SIG_STR_att"           "B_avg_SIG_STR_landed"        "B
_avg_SIG_STR_pct"
 [19] "B_avg_SUB_ATT"               "B_avg_TD_att"                "B
_avg_TD_landed"
 [22] "B_avg_TD_pct"                "B_avg_TOTAL_STR_att"         "B
_avg_TOTAL_STR_landed"
 [25] "B_avg_opp_BODY_att"          "B_avg_opp_BODY_landed"       "B
_avg_opp_CLINCH_att"
 [28] "B_avg_opp_CLINCH_landed"     "B_avg_opp_DISTANCE_att"      "B
_avg_opp_DISTANCE_landed"
 [31] "B_avg_opp_GROUND_att"        "B_avg_opp_GROUND_landed"     "B
_avg_opp_HEAD_att"
 [34] "B_avg_opp_HEAD_landed"       "B_avg_opp_KD"                "B
_avg_opp_LEG_att"
 [37] "B_avg_opp_LEG_landed"        "B_avg_opp_PASS"              "B
_avg_opp_REV"
```

```
 [40] "B_avg_opp_SIG_STR_att"        "B_avg_opp_SIG_STR_landed"     "B
_avg_opp_SIG_STR_pct"
 [43] "B_avg_opp_SUB_ATT"            "B_avg_opp_TD_att"             "B
_avg_opp_TD_landed"
 [46] "B_avg_opp_TD_pct"             "B_avg_opp_TOTAL_STR_att"      "B
_avg_opp_TOTAL_STR_landed"
 [49] "B_total_time_fought.seconds." "B_Height_cms"                 "B
_Reach_cms"
 [52] "B_Weight_lbs"                 "R_avg_BODY_att"               "R
_avg_BODY_landed"
 [55] "R_avg_CLINCH_att"             "R_avg_CLINCH_landed"          "R
_avg_DISTANCE_att"
 [58] "R_avg_DISTANCE_landed"        "R_avg_GROUND_att"             "R
_avg_GROUND_landed"
 [61] "R_avg_HEAD_att"               "R_avg_HEAD_landed"            "R
_avg_KD"
 [64] "R_avg_LEG_att"                "R_avg_LEG_landed"             "R
_avg_PASS"
 [67] "R_avg_REV"                    "R_avg_SIG_STR_att"            "R
_avg_SIG_STR_landed"
 [70] "R_avg_SIG_STR_pct"            "R_avg_SUB_ATT"                "R
_avg_TD_att"
 [73] "R_avg_TD_landed"              "R_avg_TD_pct"                 "R
_avg_TOTAL_STR_att"
 [76] "R_avg_TOTAL_STR_landed"       "R_avg_opp_BODY_att"           "R
_avg_opp_BODY_landed"
 [79] "R_avg_opp_CLINCH_att"         "R_avg_opp_CLINCH_landed"      "R
_avg_opp_DISTANCE_att"
 [82] "R_avg_opp_DISTANCE_landed"    "R_avg_opp_GROUND_att"         "R
_avg_opp_GROUND_landed"
 [85] "R_avg_opp_HEAD_att"           "R_avg_opp_HEAD_landed"        "R
_avg_opp_KD"
 [88] "R_avg_opp_LEG_att"            "R_avg_opp_LEG_landed"         "R
_avg_opp_PASS"
 [91] "R_avg_opp_REV"                "R_avg_opp_SIG_STR_att"        "R
_avg_opp_SIG_STR_landed"
 [94] "R_avg_opp_SIG_STR_pct"        "R_avg_opp_SUB_ATT"            "R
_avg_opp_TD_att"
 [97] "R_avg_opp_TD_landed"          "R_avg_opp_TD_pct"             "R
_avg_opp_TOTAL_STR_att"
[100] "R_avg_opp_TOTAL_STR_landed"   "R_total_time_fought.seconds." "R
_Height_cms"
[103] "R_Reach_cms"                  "R_Weight_lbs"                 "B
_age"
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
[106] "R_age"
> #Import dplyr library
> #The dplyr package consists of many functions specifically used for
data manipulation.
> #These functions process data faster than Base R functions and are k
nown the best for data exploration and transformation, as well.
> library(dplyr)
> #Dropping all rows that contain missing values using na.omit()
> #This function removes all incomplete cases of a data object typical
ly of a data frame, matrix or vector
> df_drop <-df %>% na.omit()
> #Storing in another file
> dim(df_drop)
[1] 3355  145
> #Finding missing values in the new dataframe
> colSums(is.na(df_drop))
                   R_fighter                    B_fighter
Referee
                           0                            0
0
                        date                     location
Winner
                           0                            0
0
                  title_bout                 weight_class
no_of_rounds
                           0                            0
0
       B_current_lose_streak         B_current_win_streak
B_draw
                           0                            0
0
              B_avg_BODY_att            B_avg_BODY_landed
B_avg_CLINCH_att
                           0                            0
0
         B_avg_CLINCH_landed           B_avg_DISTANCE_att        B_avg
_DISTANCE_landed
                           0                            0
0
            B_avg_GROUND_att          B_avg_GROUND_landed
B_avg_HEAD_att
                           0                            0
0
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
           B_avg_HEAD_landed                      B_avg_KD
B_avg_LEG_att
                           0                             0
0
            B_avg_LEG_landed                    B_avg_PASS
B_avg_REV
                           0                             0
0
           B_avg_SIG_STR_att         B_avg_SIG_STR_landed            B
_avg_SIG_STR_pct
                           0                             0
0
               B_avg_SUB_ATT                  B_avg_TD_att
B_avg_TD_landed
                           0                             0
0
                B_avg_TD_pct         B_avg_TOTAL_STR_att        B_avg_
TOTAL_STR_landed
                           0                             0
0
        B_longest_win_streak                      B_losses            B_
avg_opp_BODY_att
                           0                             0
0
         B_avg_opp_BODY_landed         B_avg_opp_CLINCH_att       B_avg_o
pp_CLINCH_landed
                           0                             0
0
      B_avg_opp_DISTANCE_att    B_avg_opp_DISTANCE_landed          B_av
g_opp_GROUND_att
                           0                             0
0
      B_avg_opp_GROUND_landed           B_avg_opp_HEAD_att         B_avg
_opp_HEAD_landed
                           0                             0
0
                B_avg_opp_KD            B_avg_opp_LEG_att          B_av
g_opp_LEG_landed
                           0                             0
0
              B_avg_opp_PASS                B_avg_opp_REV         B_avg
_opp_SIG_STR_att
                           0                             0
0
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
    B_avg_opp_SIG_STR_landed        B_avg_opp_SIG_STR_pct             B
_avg_opp_SUB_ATT 
                           0                            0 
0 
            B_avg_opp_TD_att          B_avg_opp_TD_landed              
B_avg_opp_TD_pct 
                           0                            0 
0 
     B_avg_opp_TOTAL_STR_att   B_avg_opp_TOTAL_STR_landed         B_tot
al_rounds_fought 
                           0                            0 
0 
B_total_time_fought.seconds.          B_total_title_bouts   B_win_by_D
ecision_Majority 
                           0                            0 
0 
     B_win_by_Decision_Split  B_win_by_Decision_Unanimous              
B_win_by_KO.TKO 
                           0                            0 
0 
         B_win_by_Submission B_win_by_TKO_Doctor_Stoppage              
B_wins 
                           0                            0 
0 
                    B_Stance                 B_Height_cms              
B_Reach_cms 
                           0                            0 
0 
                B_Weight_lbs        R_current_lose_streak          R_cu
rrent_win_streak 
                           0                            0 
0 
                      R_draw               R_avg_BODY_att             R
_avg_BODY_landed 
                           0                            0 
0 
            R_avg_CLINCH_att          R_avg_CLINCH_landed            R_
avg_DISTANCE_att 
                           0                            0 
0 
        R_avg_DISTANCE_landed             R_avg_GROUND_att           R_a
vg_GROUND_landed 
                           0                            0 
0 
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
              R_avg_HEAD_att            R_avg_HEAD_landed            
R_avg_KD
                           0                            0              
0
               R_avg_LEG_att             R_avg_LEG_landed            
R_avg_PASS
                           0                            0              
0
                   R_avg_REV            R_avg_SIG_STR_att           R_av
g_SIG_STR_landed
                           0                            0              
0
           R_avg_SIG_STR_pct                R_avg_SUB_ATT            
R_avg_TD_att
                           0                            0              
0
             R_avg_TD_landed                 R_avg_TD_pct           R_a
vg_TOTAL_STR_att
                           0                            0              
0
      R_avg_TOTAL_STR_landed         R_longest_win_streak            
R_losses
                           0                            0              
0
            R_avg_opp_BODY_att        R_avg_opp_BODY_landed           R_av
g_opp_CLINCH_att
                           0                            0              
0
     R_avg_opp_CLINCH_landed       R_avg_opp_DISTANCE_att     R_avg_opp
_DISTANCE_landed
                           0                            0              
0
          R_avg_opp_GROUND_att      R_avg_opp_GROUND_landed             R_
avg_opp_HEAD_att
                           0                            0              
0
        R_avg_opp_HEAD_landed                R_avg_opp_KD             R
_avg_opp_LEG_att
                           0                            0              
0
          R_avg_opp_LEG_landed               R_avg_opp_PASS            
R_avg_opp_REV
                           0                            0              
0
```

```
       R_avg_opp_SIG_STR_att     R_avg_opp_SIG_STR_landed        R_avg
_opp_SIG_STR_pct
                           0                            0
0
           R_avg_opp_SUB_ATT              R_avg_opp_TD_att          R_a
vg_opp_TD_landed
                           0                            0
0
            R_avg_opp_TD_pct      R_avg_opp_TOTAL_STR_att   R_avg_opp_
TOTAL_STR_landed
                           0                            0
0
       R_total_rounds_fought R_total_time_fought.seconds.          R_t
otal_title_bouts
                           0                            0
0
  R_win_by_Decision_Majority      R_win_by_Decision_Split  R_win_by_De
cision_Unanimous
                           0                            0
0
             R_win_by_KO.TKO         R_win_by_Submission R_win_by_TKO
_Doctor_Stoppage
                           0                            0
0
                      R_wins                     R_Stance
R_Height_cms
                           0                            0
0
                 R_Reach_cms                 R_Weight_lbs
B_age
                           0                            0
0
                       R_age
                           0
>
> #Another method of cleaning data.
> #Let us now use complete.cases to find the number of complete rows
> #We can also create a complete subset of our data by using the compl
ete.cases function.
> comp_df<-df[complete.cases(df), ]
> #Finding the columns and rows in this dataset
> dim(comp_df)
[1] 3355  145
> #Finding missing values in the new dataframe
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> colSums(is.na(comp_df))
                   R_fighter                    B_fighter                   
Referee 
                           0                            0                     
0 
                        date                     location                   
Winner 
                           0                            0                     
0 
                  title_bout                 weight_class                 
no_of_rounds 
                           0                            0                     
0 
       B_current_lose_streak         B_current_win_streak                   
B_draw 
                           0                            0                     
0 
              B_avg_BODY_att            B_avg_BODY_landed              
B_avg_CLINCH_att 
                           0                            0                     
0 
         B_avg_CLINCH_landed          B_avg_DISTANCE_att          B_avg
_DISTANCE_landed 
                           0                            0                     
0 
            B_avg_GROUND_att          B_avg_GROUND_landed                
B_avg_HEAD_att 
                           0                            0                     
0 
           B_avg_HEAD_landed                     B_avg_KD                 
B_avg_LEG_att 
                           0                            0                     
0 
            B_avg_LEG_landed                   B_avg_PASS                    
B_avg_REV 
                           0                            0                     
0 
           B_avg_SIG_STR_att         B_avg_SIG_STR_landed                B
_avg_SIG_STR_pct 
                           0                            0                     
0 
               B_avg_SUB_ATT                 B_avg_TD_att              
B_avg_TD_landed 
```

```
                           0                             0                
0
                B_avg_TD_pct          B_avg_TOTAL_STR_att         B_avg_
TOTAL_STR_landed
                           0                             0                
0
        B_longest_win_streak                      B_losses              B_
avg_opp_BODY_att
                           0                             0                
0
       B_avg_opp_BODY_landed         B_avg_opp_CLINCH_att       B_avg_o
pp_CLINCH_landed
                           0                             0                
0
      B_avg_opp_DISTANCE_att    B_avg_opp_DISTANCE_landed           B_av
g_opp_GROUND_att
                           0                             0                
0
     B_avg_opp_GROUND_landed           B_avg_opp_HEAD_att          B_avg
_opp_HEAD_landed
                           0                             0                
0
                B_avg_opp_KD            B_avg_opp_LEG_att           B_av
g_opp_LEG_landed
                           0                             0                
0
              B_avg_opp_PASS                B_avg_opp_REV           B_avg
_opp_SIG_STR_att
                           0                             0                
0
    B_avg_opp_SIG_STR_landed        B_avg_opp_SIG_STR_pct               B
_avg_opp_SUB_ATT
                           0                             0                
0
            B_avg_opp_TD_att          B_avg_opp_TD_landed                 
B_avg_opp_TD_pct
                           0                             0                
0
     B_avg_opp_TOTAL_STR_att   B_avg_opp_TOTAL_STR_landed          B_tot
al_rounds_fought
                           0                             0                
0
B_total_time_fought.seconds.          B_total_title_bouts    B_win_by_D
ecision_Majority
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
                           0                            0 
0 
     B_win_by_Decision_Split  B_win_by_Decision_Unanimous 
B_win_by_KO.TKO 
                           0                            0 
0 
         B_win_by_Submission B_win_by_TKO_Doctor_Stoppage 
B_wins 
                           0                            0 
0 
                    B_Stance                 B_Height_cms 
B_Reach_cms 
                           0                            0 
0 
                B_Weight_lbs        R_current_lose_streak         R_cu
rrent_win_streak 
                           0                            0 
0 
                      R_draw               R_avg_BODY_att            R
_avg_BODY_landed 
                           0                            0 
0 
            R_avg_CLINCH_att          R_avg_CLINCH_landed           R_
avg_DISTANCE_att 
                           0                            0 
0 
       R_avg_DISTANCE_landed             R_avg_GROUND_att          R_a
vg_GROUND_landed 
                           0                            0 
0 
              R_avg_HEAD_att            R_avg_HEAD_landed 
R_avg_KD 
                           0                            0 
0 
               R_avg_LEG_att             R_avg_LEG_landed 
R_avg_PASS 
                           0                            0 
0 
                   R_avg_REV            R_avg_SIG_STR_att         R_av
g_SIG_STR_landed 
                           0                            0 
0 
           R_avg_SIG_STR_pct                R_avg_SUB_ATT 
R_avg_TD_att 
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
                           0                             0
0
             R_avg_TD_landed                 R_avg_TD_pct             R_a
vg_TOTAL_STR_att
                           0                             0
0
      R_avg_TOTAL_STR_landed         R_longest_win_streak
R_losses
                           0                             0
0
          R_avg_opp_BODY_att        R_avg_opp_BODY_landed            R_av
g_opp_CLINCH_att
                           0                             0
0
     R_avg_opp_CLINCH_landed       R_avg_opp_DISTANCE_att     R_avg_opp
_DISTANCE_landed
                           0                             0
0
        R_avg_opp_GROUND_att      R_avg_opp_GROUND_landed              R_
avg_opp_HEAD_att
                           0                             0
0
       R_avg_opp_HEAD_landed                 R_avg_opp_KD               R
_avg_opp_LEG_att
                           0                             0
0
        R_avg_opp_LEG_landed               R_avg_opp_PASS
R_avg_opp_REV
                           0                             0
0
       R_avg_opp_SIG_STR_att     R_avg_opp_SIG_STR_landed           R_avg
_opp_SIG_STR_pct
                           0                             0
0
           R_avg_opp_SUB_ATT                R_avg_opp_TD_att             R_a
vg_opp_TD_landed
                           0                             0
0
            R_avg_opp_TD_pct      R_avg_opp_TOTAL_STR_att    R_avg_opp_
TOTAL_STR_landed
                           0                             0
0
       R_total_rounds_fought R_total_time_fought.seconds.             R_t
otal_title_bouts
```

```
                           0                             0
0
  R_win_by_Decision_Majority      R_win_by_Decision_Split  R_win_by_De
cision_Unanimous
                           0                             0
0
             R_win_by_KO.TKO           R_win_by_Submission R_win_by_TKO
_Doctor_Stoppage
                           0                             0
0
                      R_wins                      R_Stance
R_Height_cms
                           0                             0
0
                 R_Reach_cms                  R_Weight_lbs
B_age
                           0                             0
0
                       R_age
                           0
>
>
> #We can also replace missing values in our datset using the mean or
median
> #Find and identify the columns like done previously
> list_na<- colnames(df)[ apply(df, 2, anyNA) ]
> #Display list
> list_na
  [1] "B_avg_BODY_att"             "B_avg_BODY_landed"          "B
_avg_CLINCH_att"
  [4] "B_avg_CLINCH_landed"        "B_avg_DISTANCE_att"         "B
_avg_DISTANCE_landed"
  [7] "B_avg_GROUND_att"           "B_avg_GROUND_landed"        "B
_avg_HEAD_att"
 [10] "B_avg_HEAD_landed"          "B_avg_KD"                   "B
_avg_LEG_att"
 [13] "B_avg_LEG_landed"           "B_avg_PASS"                 "B
_avg_REV"
 [16] "B_avg_SIG_STR_att"          "B_avg_SIG_STR_landed"       "B
_avg_SIG_STR_pct"
 [19] "B_avg_SUB_ATT"              "B_avg_TD_att"               "B
_avg_TD_landed"
 [22] "B_avg_TD_pct"               "B_avg_TOTAL_STR_att"        "B
_avg_TOTAL_STR_landed"
```

```
 [25] "B_avg_opp_BODY_att"           "B_avg_opp_BODY_landed"         "B
_avg_opp_CLINCH_att"
 [28] "B_avg_opp_CLINCH_landed"      "B_avg_opp_DISTANCE_att"        "B
_avg_opp_DISTANCE_landed"
 [31] "B_avg_opp_GROUND_att"         "B_avg_opp_GROUND_landed"       "B
_avg_opp_HEAD_att"
 [34] "B_avg_opp_HEAD_landed"        "B_avg_opp_KD"                  "B
_avg_opp_LEG_att"
 [37] "B_avg_opp_LEG_landed"         "B_avg_opp_PASS"                "B
_avg_opp_REV"
 [40] "B_avg_opp_SIG_STR_att"        "B_avg_opp_SIG_STR_landed"      "B
_avg_opp_SIG_STR_pct"
 [43] "B_avg_opp_SUB_ATT"            "B_avg_opp_TD_att"              "B
_avg_opp_TD_landed"
 [46] "B_avg_opp_TD_pct"             "B_avg_opp_TOTAL_STR_att"       "B
_avg_opp_TOTAL_STR_landed"
 [49] "B_total_time_fought.seconds." "B_Height_cms"                  "B
_Reach_cms"
 [52] "B_Weight_lbs"                 "R_avg_BODY_att"                "R
_avg_BODY_landed"
 [55] "R_avg_CLINCH_att"             "R_avg_CLINCH_landed"           "R
_avg_DISTANCE_att"
 [58] "R_avg_DISTANCE_landed"        "R_avg_GROUND_att"              "R
_avg_GROUND_landed"
 [61] "R_avg_HEAD_att"               "R_avg_HEAD_landed"             "R
_avg_KD"
 [64] "R_avg_LEG_att"                "R_avg_LEG_landed"              "R
_avg_PASS"
 [67] "R_avg_REV"                    "R_avg_SIG_STR_att"             "R
_avg_SIG_STR_landed"
 [70] "R_avg_SIG_STR_pct"            "R_avg_SUB_ATT"                 "R
_avg_TD_att"
 [73] "R_avg_TD_landed"              "R_avg_TD_pct"                  "R
_avg_TOTAL_STR_att"
 [76] "R_avg_TOTAL_STR_landed"       "R_avg_opp_BODY_att"            "R
_avg_opp_BODY_landed"
 [79] "R_avg_opp_CLINCH_att"         "R_avg_opp_CLINCH_landed"       "R
_avg_opp_DISTANCE_att"
 [82] "R_avg_opp_DISTANCE_landed"    "R_avg_opp_GROUND_att"          "R
_avg_opp_GROUND_landed"
 [85] "R_avg_opp_HEAD_att"           "R_avg_opp_HEAD_landed"         "R
_avg_opp_KD"
 [88] "R_avg_opp_LEG_att"            "R_avg_opp_LEG_landed"          "R
_avg_opp_PASS"
```

```
 [91] "R_avg_opp_REV"                "R_avg_opp_SIG_STR_att"        "R
_avg_opp_SIG_STR_landed"
 [94] "R_avg_opp_SIG_STR_pct"        "R_avg_opp_SUB_ATT"            "R
_avg_opp_TD_att"
 [97] "R_avg_opp_TD_landed"          "R_avg_opp_TD_pct"             "R
_avg_opp_TOTAL_STR_att"
[100] "R_avg_opp_TOTAL_STR_landed"   "R_total_time_fought.seconds." "R
_Height_cms"
[103] "R_Reach_cms"                  "R_Weight_lbs"                 "B
_age"
[106] "R_age"
>
> #Compute the mean with the argument na.rm = TRUE.
> #This argument is compulsory because the columns have missing data,
and this tells R to ignore them.
> # sapply() function takes list, vector or data frame as input and gi
ves output in vector or matrix.
> #It is useful for operations on list objects and returns a list obje
ct of same length of original set.
> df_new <-data.frame(
+   sapply(
+     df, #Your dataset
+     function(x) ifelse(is.na(x), #Function to check for missing valu
es and replace it with mean
+                        mean(x, na.rm = TRUE),
+                        x)))
>
>
> #Check the missing values now
> colSums(is.na(df_new))
                   R_fighter                    B_fighter
Referee
                           0                            0
0
                        date                     location
Winner
                           0                            0
0
                  title_bout                 weight_class
no_of_rounds
                           0                            0
0
       B_current_lose_streak         B_current_win_streak
B_draw
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
                           0                            0 
0
              B_avg_BODY_att            B_avg_BODY_landed 
B_avg_CLINCH_att
                           0                            0 
0
         B_avg_CLINCH_landed           B_avg_DISTANCE_att          B_avg
_DISTANCE_landed
                           0                            0 
0
            B_avg_GROUND_att          B_avg_GROUND_landed 
B_avg_HEAD_att
                           0                            0 
0
           B_avg_HEAD_landed                     B_avg_KD 
B_avg_LEG_att
                           0                            0 
0
            B_avg_LEG_landed                   B_avg_PASS 
B_avg_REV
                           0                            0 
0
           B_avg_SIG_STR_att        B_avg_SIG_STR_landed              B
_avg_SIG_STR_pct
                           0                            0 
0
               B_avg_SUB_ATT                 B_avg_TD_att 
B_avg_TD_landed
                           0                            0 
0
                B_avg_TD_pct          B_avg_TOTAL_STR_att        B_avg_
TOTAL_STR_landed
                           0                            0 
0
        B_longest_win_streak                     B_losses             B_
avg_opp_BODY_att
                           0                            0 
0
       B_avg_opp_BODY_landed        B_avg_opp_CLINCH_att       B_avg_o
pp_CLINCH_landed
                           0                            0 
0
      B_avg_opp_DISTANCE_att    B_avg_opp_DISTANCE_landed            B_av
g_opp_GROUND_att
```

```
                           0                             0 
0 
     B_avg_opp_GROUND_landed            B_avg_opp_HEAD_att          B_avg
_opp_HEAD_landed 
                           0                             0 
0 
                B_avg_opp_KD             B_avg_opp_LEG_att           B_av
g_opp_LEG_landed 
                           0                             0 
0 
              B_avg_opp_PASS                 B_avg_opp_REV          B_avg
_opp_SIG_STR_att 
                           0                             0 
0 
    B_avg_opp_SIG_STR_landed        B_avg_opp_SIG_STR_pct              B
_avg_opp_SUB_ATT 
                           0                             0 
0 
            B_avg_opp_TD_att          B_avg_opp_TD_landed              
B_avg_opp_TD_pct 
                           0                             0 
0 
     B_avg_opp_TOTAL_STR_att   B_avg_opp_TOTAL_STR_landed          B_tot
al_rounds_fought 
                           0                             0 
0 
B_total_time_fought.seconds.          B_total_title_bouts   B_win_by_D
ecision_Majority 
                           0                             0 
0 
     B_win_by_Decision_Split  B_win_by_Decision_Unanimous              
B_win_by_KO.TKO 
                           0                             0 
0 
         B_win_by_Submission B_win_by_TKO_Doctor_Stoppage              
B_wins 
                           0                             0 
0 
                    B_Stance                  B_Height_cms              
B_Reach_cms 
                           0                             0 
0 
                B_Weight_lbs        R_current_lose_streak           R_cu
rrent_win_streak 
```

```
                           0                             0
0
                      R_draw                R_avg_BODY_att             R
_avg_BODY_landed
                           0                             0
0
            R_avg_CLINCH_att           R_avg_CLINCH_landed            R_
avg_DISTANCE_att
                           0                             0
0
       R_avg_DISTANCE_landed              R_avg_GROUND_att           R_a
vg_GROUND_landed
                           0                             0
0
              R_avg_HEAD_att               R_avg_HEAD_landed
R_avg_KD
                           0                             0
0
               R_avg_LEG_att                R_avg_LEG_landed
R_avg_PASS
                           0                             0
0
                   R_avg_REV             R_avg_SIG_STR_att          R_av
g_SIG_STR_landed
                           0                             0
0
           R_avg_SIG_STR_pct                 R_avg_SUB_ATT
R_avg_TD_att
                           0                             0
0
             R_avg_TD_landed                  R_avg_TD_pct           R_a
vg_TOTAL_STR_att
                           0                             0
0
      R_avg_TOTAL_STR_landed         R_longest_win_streak
R_losses
                           0                             0
0
          R_avg_opp_BODY_att         R_avg_opp_BODY_landed          R_av
g_opp_CLINCH_att
                           0                             0
0
     R_avg_opp_CLINCH_landed       R_avg_opp_DISTANCE_att     R_avg_opp
_DISTANCE_landed
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
                           0                            0                
0
        R_avg_opp_GROUND_att       R_avg_opp_GROUND_landed             R_
avg_opp_HEAD_att
                           0                            0                
0
       R_avg_opp_HEAD_landed                  R_avg_opp_KD              R
_avg_opp_LEG_att
                           0                            0                
0
        R_avg_opp_LEG_landed                R_avg_opp_PASS                
R_avg_opp_REV
                           0                            0                
0
       R_avg_opp_SIG_STR_att     R_avg_opp_SIG_STR_landed            R_avg
_opp_SIG_STR_pct
                           0                            0                
0
           R_avg_opp_SUB_ATT              R_avg_opp_TD_att            R_a
vg_opp_TD_landed
                           0                            0                
0
            R_avg_opp_TD_pct      R_avg_opp_TOTAL_STR_att   R_avg_opp_
TOTAL_STR_landed
                           0                            0                
0
       R_total_rounds_fought R_total_time_fought.seconds.            R_t
otal_title_bouts
                           0                            0                
0
  R_win_by_Decision_Majority      R_win_by_Decision_Split  R_win_by_De
cision_Unanimous
                           0                            0                
0
             R_win_by_KO.TKO          R_win_by_Submission R_win_by_TKO
_Doctor_Stoppage
                           0                            0                
0
                      R_wins                     R_Stance                
R_Height_cms
                           0                            0                
0
                 R_Reach_cms                 R_Weight_lbs                
B_age
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
                           0                            0
0
                       R_age
                           0
>
> #Save cleaned file for future lab session
> write.csv(comp_df, file = "df_clean.csv")
```

Conclusion:

Thus, dataset has been cleaned and made ready for analysis.

Lab Outcome:LO6

Anirudh Poroorkara
Roll: 44
IT-1 B12

Lab Assignment: 09

Aim:

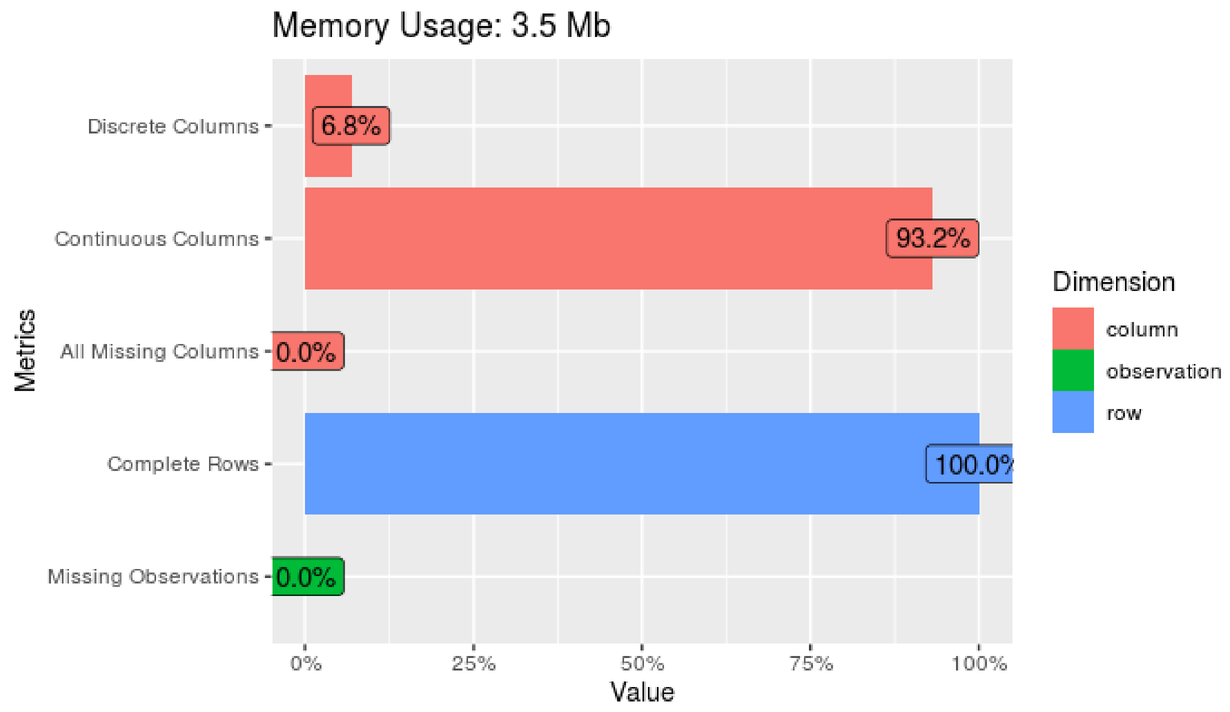
EDA on the dataset

Code:

```
> #Lab Assignment 9
> #Mini-Project Session 3
> 
> #Exploratory Data Analysis
> 
> #get the working directory
> getwd()
[1] "/cloud/project"
> 
> #Loading dataset into R
> #Header is set to to true if file contains Header information
> #Sep stands for seperator which is , in our csv file
> df <- read.csv("df_clean.csv", header = TRUE, sep =",")
> 
> 
> #Exploratory data analysis is the process to get to know your data,
so that you can generate and test your hypothesis. Visualization techn
iques are usually applied.
> #EDA consists of univariate (1-variable) and bivariate (2-variables)
analysis.
> #DataExplorer can help you with different tasks throughout your data
exploration process.
> #Install the package if not installed
> #install.packages('DataExplorer')
> 
> #import the library
> library(DataExplorer)
> library(GGally)
Loading required package: ggplot2
Registered S3 method overwritten by 'GGally':
  method from   
  +.gg   ggplot2
> 
> #To get introduced to your newly created dataset:
> introduce(df)
```

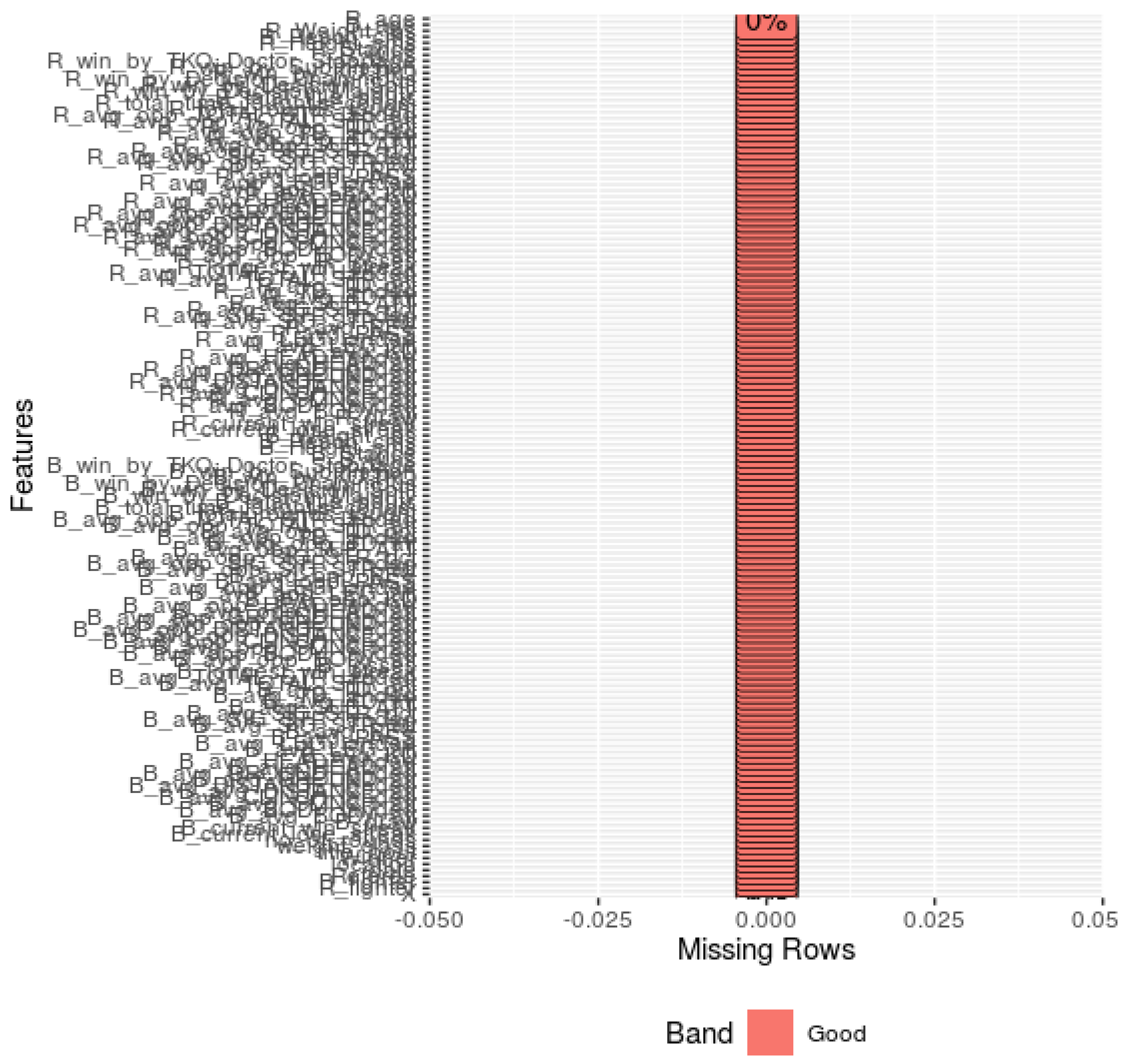
Anirudh Poroorkara
Roll: 44
IT-1 B12

```
  rows columns discrete_columns
1 3355     146               10
  continuous_columns all_missing_columns
1                136                   0
  total_missing_values complete_rows
1                    0          3355
  total_observations memory_usage
1             489830      3683888
>
> #plot_intro gives a brief Introduction to the dataset.
> #It covers basic information and gives us an idea about the content.
> plot_intro(df)
                                  >
```



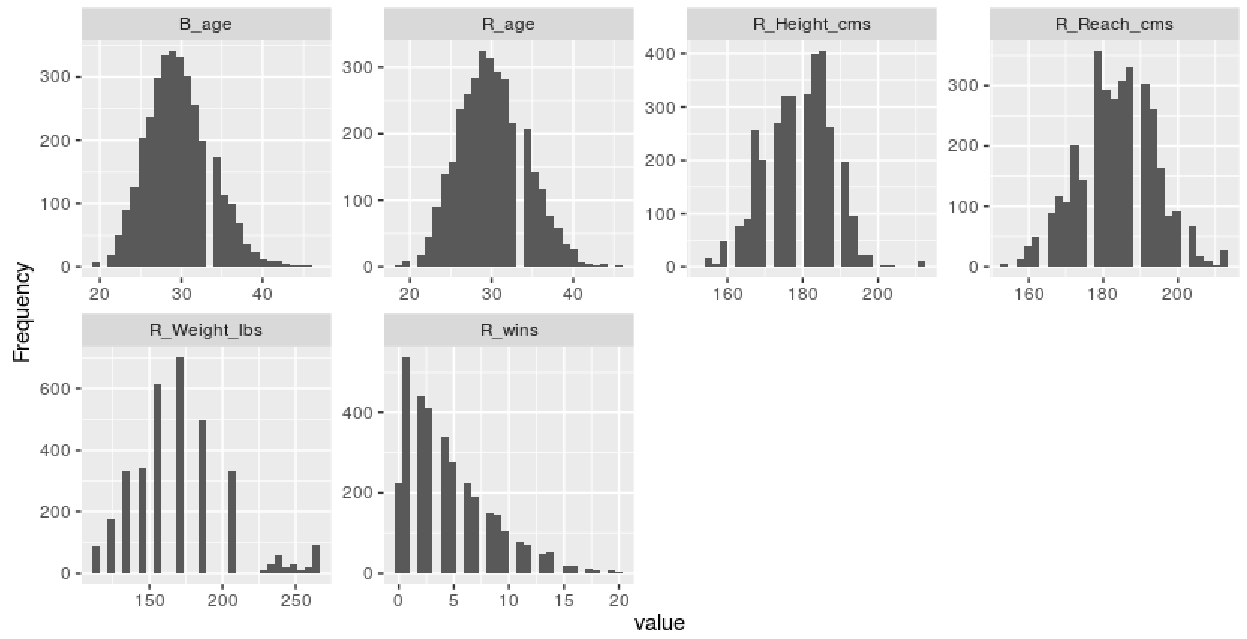
```
> #str gives the entire list of features in a network graph. We have t
he choice of viewing it radially or diagonally
> plot_str(df)
>
> #missing function returns and plots frequency of missing values for
each feature.
> #It also advices us whether to remove certain columns before carryin
g out our analysis
> #All our features are in acceptable form
> plot_missing(df)
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

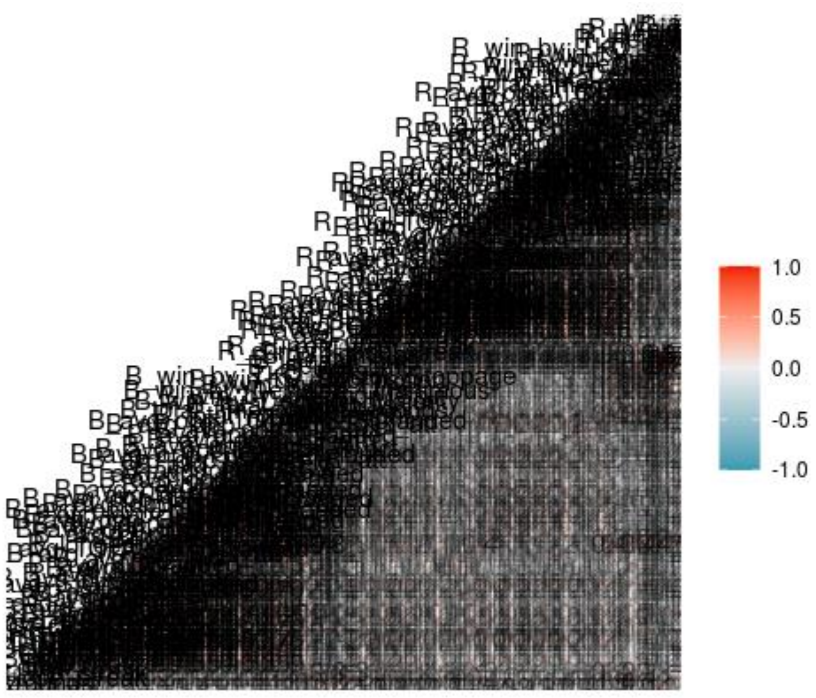


```
>
> #Let us first analyse and represent the continuous variables
> #Histograms can be used to analyse continuous variables
> # you can use a Histogram to organize and display the data in a more
userfriendly format.
> #A Histogram will make it easy
> #to see where the majority of values falls in a measurement scale, a
nd how much variation there is.
> plot_histogram(df)
```

```
>
```

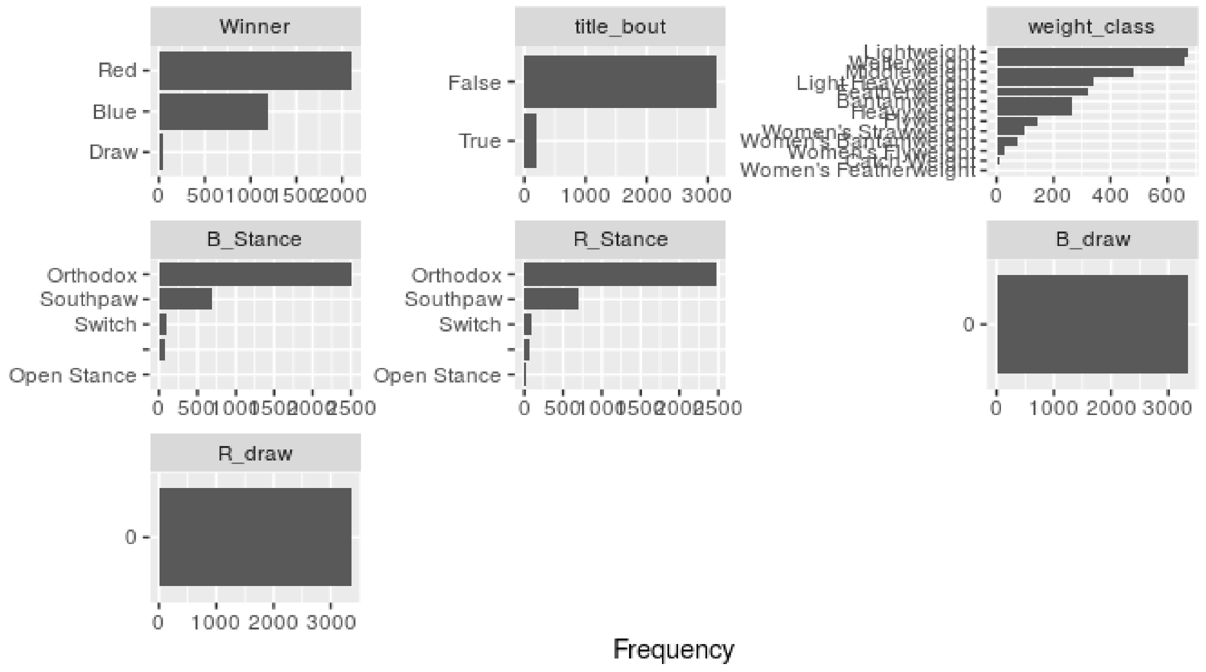


```
> #For multivariate analysis, let us do correlation analysis
> #Correlation is used to test relationships between quantitative vari
ables or categorical variables.
> #Tt's a measure of how things are related and how well they are rela
ted.
> ggcorr(df, label =TRUE, label_alpha =TRUE)
                                                  >
```



Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> #For categorical analysis, we use a bar graph
> #Each bar represents one value. When the bars are stacked next to on
e another,
> #the viewer can compare the different bars, or values, at a glance.
> plot_bar(df)
5 columns ignored with more than 50 categories.
R_fighter: 923 categories
B_fighter: 1114 categories
Referee: 171 categories
date: 445 categories
location: 146 categories
```



```
>
> #For additional complete reports, we can use the create_report funct
ion from dataexplorer
> #It creates an automatic EDA report and generates the report in an h
tml file.
> #report_EDA.pdf is attched in the folder
> #create_report(df)
```

Conclusion:

Thus, EDA is performed on the dataset.

Anirudh Poroorkara
Roll: 44
IT-1 B12

Lab Outcome:LO6

Anirudh Poroorkara
Roll: 44
IT-1 B12

Lab Assignment: 10

Aim:

Regression analysis

Code:

```
> #Lab Assignment 10
> #Mini-Project Session 4
>
> #Regression analysis
>
> #get the working directory
> getwd()
[1] "/cloud/project"
>
> #Loading dataset into R
> #Header is set to to true if file contains Header information
> #Sep stands for seperator which is , in our csv file
> df <- read.csv("df_clean.csv", header = TRUE, sep =",")
>
>
> #Regression analysis is used in stats to find trends in data
> #It will provide you with an equation for a graph so that you can ma
ke predictions about your data
> #Since we have a lot of values, we use multiple regression analysis
> #Multiple linear regression is an extension of simple linear regress
ion used to
> #predict an outcome variable (y) on the basis of multiple distinct p
redictor variables (x).
> #Here we will predict the Wins of R and B based on variables
> #1. Height
> #2. Weight
> #3. Age
> #4. Reach
> #5. Total rounds fought
> #6. Total time fought
>
>
> modelR <- lm(R_wins ~ R_Height_cms + R_Reach_cms + R_Weight_lbs +  R
_age + R_total_rounds_fought + R_total_time_fought.seconds.  ,data = d
f)
> summary(modelR)
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
Call:
lm(formula = R_wins ~ R_Height_cms + R_Reach_cms + R_Weight_lbs + 
    R_age + R_total_rounds_fought + R_total_time_fought.seconds., 
    data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.1217 -0.7885 -0.0042  0.7619  6.0647 

Coefficients:
                               Estimate
(Intercept)                  -2.6123404
R_Height_cms                  0.0020585
R_Reach_cms                   0.0267762
R_Weight_lbs                  0.0002106
R_age                        -0.0270183
R_total_rounds_fought         0.2679589
R_total_time_fought.seconds. -0.0027533
                             Std. Error
(Intercept)                   0.7106651
R_Height_cms                  0.0065516
R_Reach_cms                   0.0051411
R_Weight_lbs                  0.0012112
R_age                         0.0064898
R_total_rounds_fought         0.0019300
R_total_time_fought.seconds.  0.0001199
                             t value Pr(>|t|)
(Intercept)                   -3.676 0.000241
R_Height_cms                   0.314 0.753386
R_Reach_cms                    5.208 2.02e-07
R_Weight_lbs                   0.174 0.861988
R_age                         -4.163 3.22e-05
R_total_rounds_fought        138.836  < 2e-16
R_total_time_fought.seconds. -22.966  < 2e-16
                                     
(Intercept)                  ***
R_Height_cms                    
R_Reach_cms                  ***
R_Weight_lbs                    
R_age                        ***
R_total_rounds_fought        ***
R_total_time_fought.seconds. ***
---
Signif. codes:
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.338 on 3348 degrees of freedom
Multiple R-squared:  0.8795,	Adjusted R-squared:  0.8793 
F-statistic:  4074 on 6 and 3348 DF,  p-value: < 2.2e-16

> 
> #The first step in interpreting the multiple regression analysis is
> #to examine the F-statistic and the associated p-value, at the botto
m of model summary.
> summary(modelR)$coefficient
                                Estimate
(Intercept)                -2.6123403679
R_Height_cms                0.0020585422
R_Reach_cms                 0.0267761544
R_Weight_lbs                0.0002105744
R_age                      -0.0270182895
R_total_rounds_fought       0.2679588625
R_total_time_fought.seconds. -0.0027532817
                              Std. Error
(Intercept)                0.7106650709
R_Height_cms               0.0065516249
R_Reach_cms                0.0051410955
R_Weight_lbs               0.0012111899
R_age                      0.0064897934
R_total_rounds_fought      0.0019300441
R_total_time_fought.seconds. 0.0001198842
                               t value
(Intercept)                 -3.6759093
R_Height_cms                 0.3142033
R_Reach_cms                  5.2082585
R_Weight_lbs                 0.1738575
R_age                       -4.1631972
R_total_rounds_fought      138.8356197
R_total_time_fought.seconds. -22.9661718
                                Pr(>|t|)
(Intercept)                 2.407215e-04
R_Height_cms                7.533862e-01
R_Reach_cms                 2.021309e-07
R_Weight_lbs                8.619880e-01
R_age                       3.217298e-05
R_total_rounds_fought       0.000000e+00
R_total_time_fought.seconds. 1.630348e-108
> 
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> #For a given the predictor, the t-statistic evaluates whether or not
there is significant association between the predictor and the outcome
variable,
> #that is whether the beta coefficient of the predictor is significan
tly different from zero.
> # so for our given set of predictors we find that height and weight
is not significant in analysing the wins because its value is close to
0
> #Therefore we remove height and weight
> modelR <- lm(R_wins ~ R_Reach_cms +  R_age + R_total_rounds_fought +
R_total_time_fought.seconds.  ,data = df)
> summary(modelR)

Call:
lm(formula = R_wins ~ R_Reach_cms + R_age + R_total_rounds_fought +
    R_total_time_fought.seconds., data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-7.1208 -0.7902 -0.0062  0.7637  6.0593

Coefficients:
                               Estimate
(Intercept)                  -2.5845949
R_Reach_cms                   0.0287740
R_age                        -0.0265275
R_total_rounds_fought         0.2678970
R_total_time_fought.seconds. -0.0027602
                             Std. Error
(Intercept)                   0.4772775
R_Reach_cms                   0.0023339
R_age                         0.0062450
R_total_rounds_fought         0.0019240
R_total_time_fought.seconds.  0.0001187
                             t value Pr(>|t|)
(Intercept)                   -5.415 6.55e-08
R_Reach_cms                   12.329  < 2e-16
R_age                         -4.248 2.22e-05
R_total_rounds_fought        139.242  < 2e-16
R_total_time_fought.seconds. -23.247  < 2e-16

(Intercept)                  ***
R_Reach_cms                  ***
R_age                        ***
```

```
R_total_rounds_fought        ***
R_total_time_fought.seconds. ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.338 on 3350 degrees of freedom
Multiple R-squared:  0.8795,	Adjusted R-squared:  0.8794
F-statistic:  6114 on 4 and 3350 DF,  p-value: < 2.2e-16

>
> #The RSE estimate gives a measure of error of prediction. The lower
the RSE, the more accurate the model (on the data in hand).
> #The error rate can be estimated by dividing the RSE by the mean out
come variable:
> #Our outcome variable is R_wins
> sigma(modelR)/mean(df$R_wins)
[1] 0.2850224
>
> #We get a 0.28 which means it has a 28% error rate
> #install.packages('caTools')
> library(caTools)
> #Let us now test the created prediction model
> #Splitting the data set into training and testing data
> split = sample.split(df$R_wins,SplitRatio = 2/3)
> #Training is 2 parts
> training_set = subset(df, split == TRUE)
> #testing is 1 part
> test_set = subset(df, split == FALSE)
>
> #Do a prediction on the model
> R_pred = predict(modelR, newdata = test_set)
>
> #Store and compare the result in the test cases
> resultR <- data.frame(Reach = test_set$R_Reach_cms, Age = test_set$R
_age, Total_rounds= test_set$R_total_rounds_fought, Total_seconds=test
_set$R_total_time_fought.seconds., Actual_Value = test_set$R_wins, Pre
dicted_Value = R_pred)
> resultR
     Reach Age Total_rounds Total_seconds
3   193.04  35           33      604.4000
5   190.50  26            7      440.7500
6   167.64  28            8      540.0000
8   162.56  33           25      800.1111
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
10  160.02  29              9       900.0000
15  182.88  30             11       561.8000
17  177.80  37              4       461.0000
19  167.64  35             19       807.8571
20  193.04  27              1       193.0000
26  193.04  38             26       565.8333
28  177.80  29              9       858.3333
29  165.10  26             23       768.5000
33  172.72  31             12       900.0000
34  190.50  42             24       576.1818
36  182.88  28             22       656.6667
37  177.80  40             74       787.5769
39  195.58  35             22       355.6429
40  177.80  35             46       807.9375
41  162.56  27              6       900.0000
44  180.34  28              2       383.0000
46  182.88  39             23       509.1667
49  193.04  39             36       566.3750
56  213.36  26              1       261.0000
57  175.26  36             13       953.0000
58  187.96  34              6       554.0000
63  190.50  31             18       876.0000
65  203.20  36             41       565.7222
69  190.50  26              4       500.0000
72  198.12  37             19       611.6250
73  177.80  34             15       711.3333
78  177.80  29             18       900.0000
83  170.18  33             29       685.1667
93  187.96  31             14       510.8571
94  185.42  29             11       606.2000
99  200.66  34             36       568.5000
101 190.50  35             41       608.5556
111 187.96  36             32       723.8333
116 182.88  37             73       738.7500
119 198.12  35             26       736.1000
124 193.04  34              3       900.0000
135 185.42  24              3       900.0000
137 175.26  35             21       868.8571
140 203.20  29             11       802.0000
143 187.96  22              3       891.0000
144 193.04  25              1        46.0000
149 177.80  32             43      1027.9167
154 182.88  23              9       839.3333
156 180.34  32              9       436.8000
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
158 165.10  24               20       728.2857
159 193.04  39               35       590.7333
161 167.64  32               39       681.6875
162 182.88  30               19       776.8571
163 213.36  31               55       844.0556
165 190.50  34               43       801.4000
168 195.58  39               56       507.1481
177 180.34  31               15       497.7500
180 172.72  30                6       527.3333
190 177.80  29               26       697.5000
194 180.34  32               19       561.3333
195 190.50  28               12       695.8000
197 193.04  37               45       654.7778
198 182.88  44               43       656.1176
200 190.50  37               52       711.5000
204 177.80  34               69       766.5600
205 165.10  32               21       549.2000
206 165.10  34               39       737.6000
212 182.88  23                2       453.0000
213 160.02  29                6       900.0000
223 170.18  30               26       850.3333
226 177.80  32                7       601.6667
230 198.12  34               32       621.4615
233 195.58  34               21       360.6154
234 180.34  26               21       746.7500
235 170.18  33                3       900.0000
236 177.80  29                6       837.5000
240 205.74  26                5       659.5000
249 190.50  32                5       659.5000
254 200.66  33               31       555.2143
256 175.26  25               34       906.6364
266 177.80  40               42       693.4706
268 198.12  37               54       746.6500
271 195.58  39               53       492.0385
273 185.42  30               13       413.8571
274 190.50  26                3       899.0000
277 157.48  26               32       705.8462
278 185.42  27                8       777.3333
286 180.34  35               69       653.5517
290 185.42  32               44       660.5000
291 170.18  31               20       728.7500
293 167.64  34               35       739.3077
296 177.80  26                2       528.0000
297 187.96  30               12       822.5000
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
300 170.18  33              35      685.0000
302 167.64  31              64     1091.2941
305 165.10  31              18      510.2222
308 177.80  25              21      737.8750
313 185.42  28              18      581.6250
316 177.80  31               2      104.0000
317 172.72  33              33      654.5000
318 172.72  33              20      684.3750
319 165.10  29              33      778.1667
323 193.04  38              32      568.6429
324 198.12  32              19      874.5000
326 167.64  34              16      792.5000
328 203.20  29               7      645.3333
329 175.26  32               8      702.0000
330 180.34  22              16      569.1429
332 198.12  28              16      692.3333
337 160.02  36               6      900.0000
345 180.34  26               9      585.5000
350 185.42  30              48      744.8333
352 187.96  28               9      592.0000
356 185.42  35              55      527.0000
360 185.42  30              28      733.8182
363 160.02  24               9      900.0000
364 177.80  28               5      463.6667
365 172.72  29               9      900.0000
370 203.20  38              25      472.2308
374 190.50  38              61      784.6818
381 180.34  33              39      519.7368
385 185.42  28               9      900.0000
388 180.34  25              18      724.8571
392 165.10  36               7      600.6667
394 182.88  40              73      745.7037
397 175.26  26               8      707.6667
399 190.50  33              16      549.1250
402 177.80  29              15      900.0000
403 175.26  37               9      544.5000
404 182.88  22               2      415.0000
407 160.02  25               3      900.0000
411 203.20  40              10      389.8333
413 185.42  31              15      876.6000
416 195.58  36               6      511.3333
420 198.12  30               4      526.5000
422 180.34  28              24      742.8889
426 200.66  28              25      674.5000
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
432 187.96  37           47      538.0000
436 190.50  31            3      900.0000
441 182.88  28           11      772.2500
442 185.42  27            5      716.0000
444 162.56  32           18      880.5000
449 172.72  24            5      614.0000
450 187.96  26           22      773.6250
458 193.04  34           36      499.4444
462 177.80  32           13      774.6000
463 157.48  26           29      689.6667
468 170.18  26           10      697.5000
470 170.18  33           32      679.8462
477 167.64  31            1      236.0000
481 182.88  43           39      665.8667
482 190.50  24            1      275.0000
484 200.66  26            5      465.6667
485 172.72  33           11      612.8000
487 170.18  32           23      661.9000
489 175.26  26           15      681.0000
490 182.88  29            6      900.0000
492 170.18  36           25      787.5556
498 180.34  37           37      626.6875
522 180.34  33           43      702.2353
524 172.72  30            3      894.0000
525 185.42  30           12      883.2500
526 152.40  28            9      900.0000
527 175.26  23            3      686.0000
532 190.50  33           38      817.2308
534 187.96  24            7      690.0000
542 187.96  36           19      777.8571
    Actual_Value Predicted_Value
3             14      9.21381664
5              3      2.86585879
6              4      2.14898046
8              5      5.70646758
10             3      1.17742806
15             4      3.27795971
17             0      1.34904229
19             4      4.17082164
20             1      1.98886925
26             6      7.36540580
28             2      1.80403643
29             6      5.51670423
33             2      2.29349337
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
34              6     6.62185220
36              6     6.01603310
37             12    19.12084126
39              9     7.02664934
40             10    11.69616308
41              2     0.49987782
44              0     1.34037557
46              9     6.39925395
49             11    10.01635351
56              0     2.41239136
57              1     2.35554924
58              1     2.00006562
63              3     4.47872077
65             11    11.72956635
69              1     1.89862710
72              4     5.53643259
73              4     3.68452738
78              5     4.10010226
83              8     7.31458047
93              5     4.34190639
94              2     3.25502112
99             12    10.36238320
101            10    11.27243697
111             9     8.44356305
116            17    19.21346948
119             6     7.12119356
124             0     0.38752433
135             1     0.43354185
137             5     4.75750233
140             4     3.22617902
143             1     0.58452432
144             1     2.44767056
149             9    10.36487265
154             3     2.16181655
156             5     2.96104715
158             4     4.87706676
159            10     9.68122313
161             9     9.95659557
162             5     4.82753979
163            16    15.13685592
165             7    11.30247338
168            16    15.61085669
177             5     4.42672394
180             2     1.74126534
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
190             7      6.80221473
194             6      5.29628319
195             3      3.44830532
197             9     12.23647433
198             8     11.21894607
200            14     13.88210423
204            17     17.99853161
205             6      5.42705235
206            12      9.67612621
212             0      1.35288643
213             2      0.37373696
223             6      6.13458235
226             2      1.89710575
230            10      9.07152591
233             9      6.77155475
234             4      5.47945913
235             1     -0.24372077
236             2      1.05784906
240             2      2.16479068
249             1      1.56711057
254            11      9.08609649
256             8      8.40116139
266            12     10.80788657
268            12     14.54013554
271            16     14.84887104
273             6      4.29518837
274             0      0.52941871
277             9      7.88145580
278             3      2.03202643
286            17     18.35701300
290             9     11.86616293
291             3      4.83626440
293             8      8.67291040
296             1      0.92011880
297             2      2.97244981
300            10      8.92242269
302            15     15.54996134
305             5      4.75747440
308             4      5.45739737
313             4      5.22465924
316             2      1.95779711
317             8      8.54389993
318             6      4.97877817
319             7      8.08941030
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
323            10      8.96503320
324             3      4.94348810
326             4      3.43604651
328             3      2.58701891
329             1      1.81497899
330             5      4.73631147
332             4      4.74871961
337             0      0.18804442
345             1      2.70977359
350             8     12.75803099
352             2      2.85803495
356            20     15.10193164
360             5      7.43049415
363             2      1.31006559
364             1      1.84832639
365             1      1.54285729
370             8      7.64821197
374            14     16.06465510
381            10     10.74251007
385             1      1.93481402
388             4      4.76272374
392             1      1.42832668
394            19     19.11469350
397             2      1.95850301
399             6      4.79210516
402             4      3.29641117
403             1      2.38496667
404             0      1.48430073
407             1     -0.32384411
411             4      3.80413314
413             4      3.52720188
416             3      2.28403577
420             2      1.93862988
422             6      6.24075257
426             6      7.28210193
432            12     12.94842426
436             1      0.39402100
441             3      2.75013508
442             2      1.39762631
444             4      3.63582816
449             1      1.39331784
450             6      5.89243389
458            10     10.33373135
462             4      3.02716101
```

```
463             8      7.12242298
468             3      2.37618721
470             9      8.13295713
477             0      1.03321309
481             8     10.14697641
482             1      1.76903126
484             2      2.55363366
485             2      2.76526469
487             6      5.79794595
489             3      3.90738700
490             1      1.03150957
492             7      5.88079818
498             9      9.80540304
522            13     11.31036983
524             0     -0.07449134
525             2      2.73168310
526             2      0.98469803
527             0      0.75840425
532             7      9.94581998
534             2      2.15785338
542             5      4.81178627
 [ reached 'max' / getOption("max.print") -- omitted 952 rows ]
>
>
> #We now repeat the same with B
> modelB <- lm(B_wins ~ B_Height_cms + B_Reach_cms + B_Weight_lbs +  B
_age + B_total_rounds_fought + B_total_time_fought.seconds.  ,data = d
f)
> summary(modelB)

Call:
lm(formula = B_wins ~ B_Height_cms + B_Reach_cms + B_Weight_lbs + 
    B_age + B_total_rounds_fought + B_total_time_fought.seconds., 
    data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.7887 -0.7017 -0.0318  0.7241  5.7063 

Coefficients:
                               Estimate
(Intercept)                   0.0427642
B_Height_cms                  0.0006284
B_Reach_cms                   0.0085128
```

```
B_Weight_lbs                  0.0041488
B_age                        -0.0317970
B_total_rounds_fought         0.2743816
B_total_time_fought.seconds. -0.0023744
                             Std. Error
(Intercept)                   0.6512501
B_Height_cms                  0.0057748
B_Reach_cms                   0.0045366
B_Weight_lbs                  0.0010759
B_age                         0.0057644
B_total_rounds_fought         0.0019104
B_total_time_fought.seconds.  0.0001005
                             t value Pr(>|t|)
(Intercept)                    0.066 0.947649
B_Height_cms                   0.109 0.913358
B_Reach_cms                    1.876 0.060680
B_Weight_lbs                   3.856 0.000117
B_age                         -5.516 3.73e-08
B_total_rounds_fought        143.623  < 2e-16
B_total_time_fought.seconds. -23.616  < 2e-16

(Intercept)
B_Height_cms
B_Reach_cms                  .
B_Weight_lbs                 ***
B_age                        ***
B_total_rounds_fought        ***
B_total_time_fought.seconds. ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.189 on 3348 degrees of freedom
Multiple R-squared:  0.8785,	Adjusted R-squared:  0.8783
F-statistic:  4035 on 6 and 3348 DF,  p-value: < 2.2e-16

> summary(modelB)$coefficient
                                 Estimate
(Intercept)                   0.0427642265
B_Height_cms                  0.0006283718
B_Reach_cms                   0.0085127503
B_Weight_lbs                  0.0041488348
B_age                        -0.0317970175
B_total_rounds_fought         0.2743815706
```

```
B_total_time_fought.seconds. -0.0023744479
                               Std. Error
(Intercept)                  0.6512501322
B_Height_cms                 0.0057747983
B_Reach_cms                  0.0045366119
B_Weight_lbs                 0.0010759183
B_age                        0.0057644288
B_total_rounds_fought        0.0019104324
B_total_time_fought.seconds. 0.0001005447
                                  t value
(Intercept)                    0.06566483
B_Height_cms                   0.10881276
B_Reach_cms                    1.87645548
B_Weight_lbs                   3.85608725
B_age                         -5.51607433
B_total_rounds_fought        143.62275475
B_total_time_fought.seconds. -23.61583478
                                 Pr(>|t|)
(Intercept)                   9.476486e-01
B_Height_cms                  9.133575e-01
B_Reach_cms                   6.067968e-02
B_Weight_lbs                  1.173829e-04
B_age                         3.729752e-08
B_total_rounds_fought         0.000000e+00
B_total_time_fought.seconds. 3.522706e-114
>
> #In B_wins, Height has a T value close to zero, hence we can remove
the height
> modelB <- lm(B_wins ~ B_Reach_cms + B_Weight_lbs +  B_age + B_total_
rounds_fought + B_total_time_fought.seconds.  ,data = df)
> summary(modelB)

Call:
lm(formula = B_wins ~ B_Reach_cms + B_Weight_lbs + B_age + B_total_rou
nds_fought +
    B_total_time_fought.seconds., data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7880 -0.7020 -0.0320  0.7239  5.7072

Coefficients:
                               Estimate
(Intercept)                   0.0841889
```

```
B_Reach_cms                   0.0088642
B_Weight_lbs                  0.0041952
B_age                        -0.0318252
B_total_rounds_fought         0.2743800
B_total_time_fought.seconds. -0.0023751
                             Std. Error
(Intercept)                   0.5283127
B_Reach_cms                   0.0031853
B_Weight_lbs                  0.0009879
B_age                         0.0057578
B_total_rounds_fought         0.0019101
B_total_time_fought.seconds.  0.0001004
                             t value Pr(>|t|)
(Intercept)                    0.159  0.87340
B_Reach_cms                    2.783  0.00542
B_Weight_lbs                   4.246 2.23e-05
B_age                         -5.527 3.50e-08
B_total_rounds_fought        143.647  < 2e-16
B_total_time_fought.seconds. -23.664  < 2e-16

(Intercept)
B_Reach_cms                  **
B_Weight_lbs                 ***
B_age                        ***
B_total_rounds_fought        ***
B_total_time_fought.seconds. ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.188 on 3349 degrees of freedom
Multiple R-squared:  0.8785,	Adjusted R-squared:  0.8783
F-statistic:  4844 on 5 and 3349 DF,  p-value: < 2.2e-16

>
> sigma(modelB)/mean(df$B_wins)
[1] 0.3349714
>
> #We get a 0.33 which means it has a 33% error rate which is not very
efficient to solve models.
>
> #Let us now test the created prediction model
> #Splitting the data set into training and testing data
> split = sample.split(df$B_wins,SplitRatio = 2/3)
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> #Training is 2 parts
> training_set = subset(df, split == TRUE)
> #testing is 1 part
> test_set = subset(df, split == FALSE)
>
> #Do a prediction on the model
> B_pred = predict(modelB, newdata = test_set)
> #Store and compare the result in the test cases
> resultB <- data.frame(Reach = test_set$B_Reach_cms, Weight = test_se
t$B_Weight_lbs, Age = test_set$B_age, Total_rounds= test_set$B_total_r
ounds_fought, Total_seconds=test_set$B_total_time_fought.seconds., Act
ual_Value = test_set$B_wins, Predicted_Value = B_pred)
> resultB
     Reach Weight Age Total_rounds
1   170.18    135  31            9
2   167.64    125  32           29
4   170.18    135  26            9
5   185.42    250  32            8
7   165.10    135  32           23
8   167.64    115  25           10
9   182.88    145  31           10
14  198.12    205  27            7
21  195.58    170  26           36
22  182.88    185  30            3
27  187.96    185  26            5
32  180.34    170  26            3
33  162.56    135  35           21
37  177.80    155  37           67
41  172.72    135  28            9
42  170.18    135  31           13
45  203.20    265  27            3
46  195.58    185  30           13
48  180.34    170  27           20
53  162.56    115  32            6
59  193.04    235  37           15
68  182.88    170  30           22
69  177.80    135  30           14
71  177.80    155  30           10
73  185.42    145  32           50
74  157.48    115  33           15
75  182.88    145  28            9
77  185.42    155  29           15
79  195.58    185  31            9
84  177.80    145  28            6
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
87  162.56    125  30             3
88  162.56    115  34            22
89  187.96    170  34            41
90  182.88    170  30            23
94  190.50    185  33             4
95  182.88    145  31             8
98  165.10    125  30             2
102 198.12    265  37            23
111 193.04    170  31            28
112 165.10    135  32            22
123 170.18    135  26             6
125 172.72    145  26             1
126 160.02    125  23             4
127 180.34    155  34             9
130 177.80    155  34            27
132 182.88    170  29            19
136 175.26    125  30             3
140 195.58    185  43            55
142 167.64    125  23             3
143 190.50    205  32            34
145 182.88    145  27             8
147 177.80    135  32             3
148 170.18    135  30             8
149 182.88    145  29            15
155 170.18    125  32            42
156 190.50    155  31            21
157 165.10    125  30            34
161 182.88    155  28             2
173 185.42    170  34            14
175 177.80    155  31            28
177 175.26    135  25            37
178 187.96    155  29            39
181 195.58    185  30            11
182 190.50    185  34            19
189 162.56    115  33            15
190 180.34    155  32            19
191 167.64    125  32            26
192 190.50    185  31            15
193 175.26    135  31            12
194 187.96    170  29            11
197 200.66    205  27             7
198 198.12    264  31             7
201 165.10    125  30            11
206 165.10    135  26             6
```

```
214 190.50    185  33            1
215 185.42    170  32           21
220 180.34    125  29           15
221 180.34    145  26           20
226 182.88    155  32            3
227 172.72    125  27           12
228 172.72    135  27            9
229 200.66    260  33           34
231 195.58    185  34           15
232 185.42    185  28            8
238 180.34    155  26           12
241 193.04    205  30           19
243 193.04    205  37           25
249 190.50    249  26            4
250 185.42    155  26            6
251 187.96    155  30           20
258 187.96    155  35           15
259 172.72    135  28            1
262 190.50    185  31           12
267 195.58    170  35           30
268 180.34    185  40           26
269 203.20    240  41           11
272 190.50    155  34           16
274 187.96    205  30            1
282 162.56    135  28            9
287 170.18    125  24            9
293 172.72    125  30            5
295 182.88    170  27           20
296 180.34    135  38           38
300 170.18    135  32            8
303 182.88    145  29           14
307 165.10    125  26            2
309 182.88    155  29           43
311 152.40    115  28           23
313 182.88    170  27           12
318 172.72    125  29           12
320 162.56    115  32           12
324 213.36    265  30           38
326 177.80    155  31            7
327 190.50    170  32            6
335 170.18    135  28           18
340 180.34    170  26           16
343 182.88    185  27            5
346 193.04    170  32           10
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
349 203.20    185  28            5
356 187.96    170  26           25
358 167.64    125  31           23
362 185.42    170  23           21
364 180.34    125  21            3
366 167.64    125  30            3
367 185.42    185  41           30
374 193.04    205  29           17
375 175.26    135  24           31
377 180.34    145  26            3
379 182.88    155  35           11
381 195.58    170  35           29
389 187.96    170  36           20
393 190.50    185  37           23
395 167.64    115  27            4
399 170.18    170  36           28
402 182.88    135  28            6
403 167.64    135  37           21
404 195.58    155  34            2
405 170.18    135  29           21
407 162.56    115  26            7
409 187.96    185  39           61
412 167.64    155  34           15
414 187.96    170  33            4
    Total_seconds Actual_Value Predicted_Value
1        419.4000            4      2.64578048
2        849.0000            4      7.01675887
4        652.0000            4      2.25246512
5       1200.0000            1      1.10313008
7        495.2500            8      6.23009638
8        716.0000            2      2.30024746
9        670.7500            3      2.47771361
14       681.6667            3      2.14274681
21       724.8571           10      9.85966630
22       900.0000            1      0.21220050
27       610.5000            2      1.62087420
32       900.0000            1      0.25405870
33       702.3750            4      5.07140933
37       702.1538           14     17.84875885
41       900.0000            1      1.62231243
42       754.2000            2      2.94812710
45       658.0000            1      1.39817682
46       378.0000            6      4.30836258
48       603.1111            5      5.59182576
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
53        900.0000          0     0.49790804
59        700.5000          4     4.05562946
68        792.2500          5     5.61840719
69        793.2000          2     3.22924977
71        611.7500          2     2.64658947
73        684.4500         11    13.41106584
74        687.6667          4     3.39477946
75        890.0000          2     1.77807505
77        900.0000          2     3.43324601
79        344.6667          4     3.25818631
84        900.0000          1     0.88615416
87        900.0000          0    -0.21962998
88        808.3750          3     5.04195379
89        797.7333          9    10.73633365
90        548.3636          8     6.47203441
94        541.5000          0     1.31011275
95        551.7500          2     2.21158691
98        565.0000          0     0.32415353
102       558.4000          6     6.75905219
111       894.1111          9     7.08099510
112       513.7273          7     5.91183155
123       569.3333          3     1.62566419
125       251.0000          0     1.07429478
126       284.0000          2     1.71805446
127       410.6000          4     2.74516910
130       688.8182          7     7.00070673
132       655.7500          5     5.15128938
136       900.0000          1    -0.10705467
140       639.9545         17    14.79643559
142       900.0000          1     0.04817647
143       570.0000         10     9.58955238
145       776.0000          1     1.80627820
147       900.0000          0    -0.10623832
148       475.0000          3     2.27117175
149       707.3333          5     3.82637611
155       745.8000         12    10.85132143
156       413.0000          6     6.21756499
157       735.9231          8     8.69835990
161       314.0000          0     1.26740692
173       607.1667          4     3.75816722
175       671.0000          8     7.41288175
177       906.0833          8     9.40849529
178       418.2381         12    11.18509983
181       451.6667          4     3.58463905
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
182        588.1111          6      5.28328322
189        883.8000          3      2.97397917
190        583.3333          6      5.14236591
191        843.3333          3      6.20707754
192        871.2000          3      3.60888276
193        633.6000          2      3.00521066
194        406.0000          1      3.59445309
197        361.0000          3      2.92686767
198        651.0000          3      2.33579635
201        447.8333          4      3.07185276
206        513.3333          3      1.71363799
214        238.0000          0      1.20780646
215        566.9000          8      5.83811389
220        466.0000          4      4.29314130
221        794.1429          6      5.06505802
226        900.0000          1      0.02269513
227        900.0000          1      2.43532602
228        619.5000          1      2.32034477
229        577.4667         12      9.86078773
231        563.4286          3      4.28941602
232        532.7500          2      2.54251052
238        648.4000          1      3.25811901
241        501.0000          6      5.72389735
243        602.7273          6      6.90579175
249        514.0000          1      1.86669414
250        900.0000          1      1.05930139
251        487.8000          9      5.77483986
258        704.0000          5      3.73032367
259        205.0000          0      1.07794596
262        864.0000          3      2.80284320
267        492.0000          9      8.48001151
268        628.3636          3      6.82732941
269        349.1429          5      3.77634254
272        725.5000          4      4.00797993
274        263.0000          0      1.30529353
282        900.0000          3      1.53225219
287        900.0000          2      1.68514646
293        737.0000          2      0.80632673
295        682.7500          6      5.42519289
296        593.7059         10     10.05609631
300        494.2500          2      2.16180127
303        797.8000          4      3.33713142
307        475.0000          1      0.66521059
309        540.6500         15     11.94685294
```

```
311       831.6250           6      5.36200431
313       686.4000           3      3.22148372
318       900.0000           3      2.37167564
320       879.7500           2      2.19228334
324       450.7500          12     11.48829548
326       577.0000           1      1.87415791
327       900.0000           2      0.97630788
335       619.3750           6      4.73572159
340       606.0000           4      4.51926956
343       275.3333           3      2.34006320
346       648.0000           2      2.69486067
349       708.5000           2      1.45955733
356       787.5556           7      6.62502776
358       836.2500           2      5.43258608
362       659.1111           6      5.90553274
364       900.0000           1      0.22440216
366       900.0000           0     -0.17459985
367       758.2000           9      7.62968397
374       623.8571           4      4.91516818
375       907.3000           7      7.79115069
377       900.0000           0      0.14917954
379       610.4000           4      2.81008002
381       518.8571           8      8.14184389
389       688.0000           6      5.17132720
393       708.4444           5      5.99952764
395       561.5000           2      0.95726528
399       798.8000           8      6.94560416
402       900.0000           1      0.88923262
403       873.4286           4      4.64652486
404        78.5000           1      1.74836010
405       787.1250           6      5.12861846
407       474.0000           2      1.97501904
409       729.7826          15     16.28912334
412       900.0000           4      3.11651463
414       531.0000           1      1.24960843
 [ reached 'max' / getOption("max.print") -- omitted 974 rows ]
```

Conclusion:

Thus, Regression analysis is performed on the dataset.

Anirudh Poroorkara
Roll: 44
IT-1 B12

Lab Outcome: LO6

Anirudh Poroorkara
Roll: 44
IT-1 B12

Lab Assignment: 11

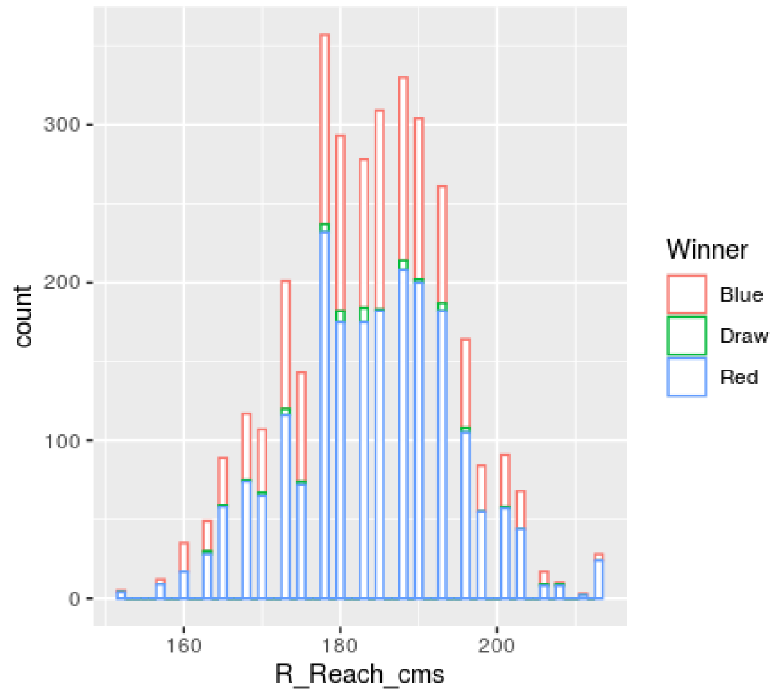
Aim:

Data Visualization using ggplot2

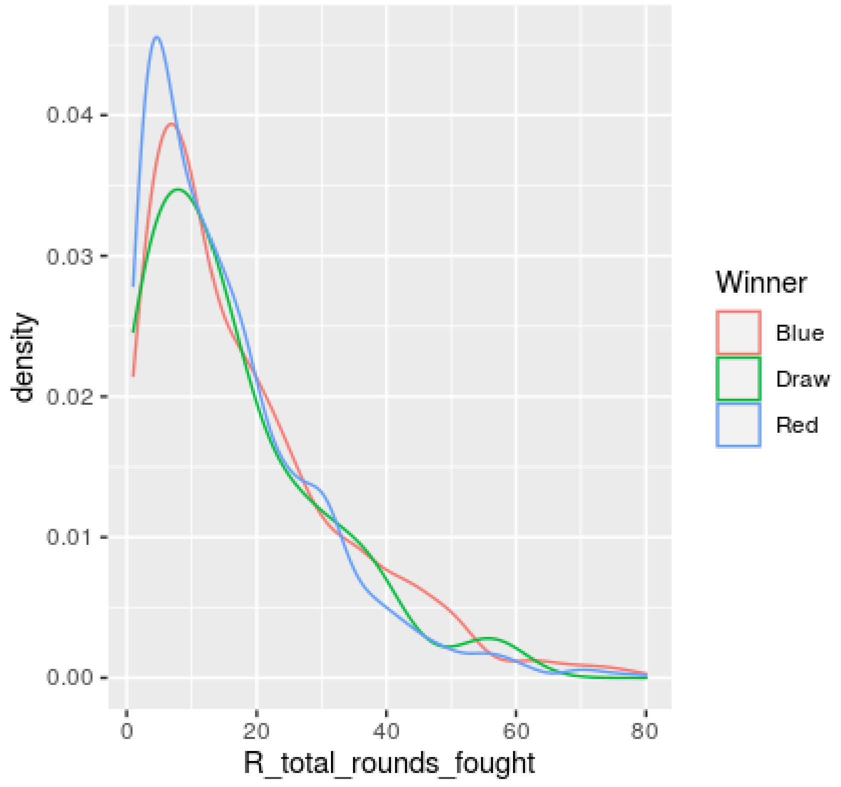
Code:

```
> #Lab Assignment 11
> #Mini-Project Session 5
> 
> #Data Visualization using ggplot2
> 
> #get the working directory
> getwd()
[1] "/cloud/project"
> 
> #Loading dataset into R
> #Header is set to to true if file contains Header information
> #Sep stands for seperator which is , in our csv file
> df <- read.csv("df_clean.csv", header = TRUE, sep =",")
> 
> #ggplot2 is a robust and a versatile R package for generating aesthetic plots a
nd charts.
> #Plot = data + Aesthetics + Geometry
> #1. Data refers to a data frame
> #2. Aesthetics indicates x and y variables. It is also used to tell R how data
are displayed in a plot, e.g. color, size and shape of points etc.
> #3. Geometry refers to the type of graphics
> 
> #install the package if not installed
> #install.packages("ggplot2")
> library(ggplot2)
> 
> #Plotting with R_Reach and winner using a group histogram
> ggplot(data  = df, aes( x = R_Reach_cms, color=Winner)) + geom_histogram(fill="
white", binwidth = 1)
```

Anirudh Poroorkara
Roll: 44
IT-1 B12



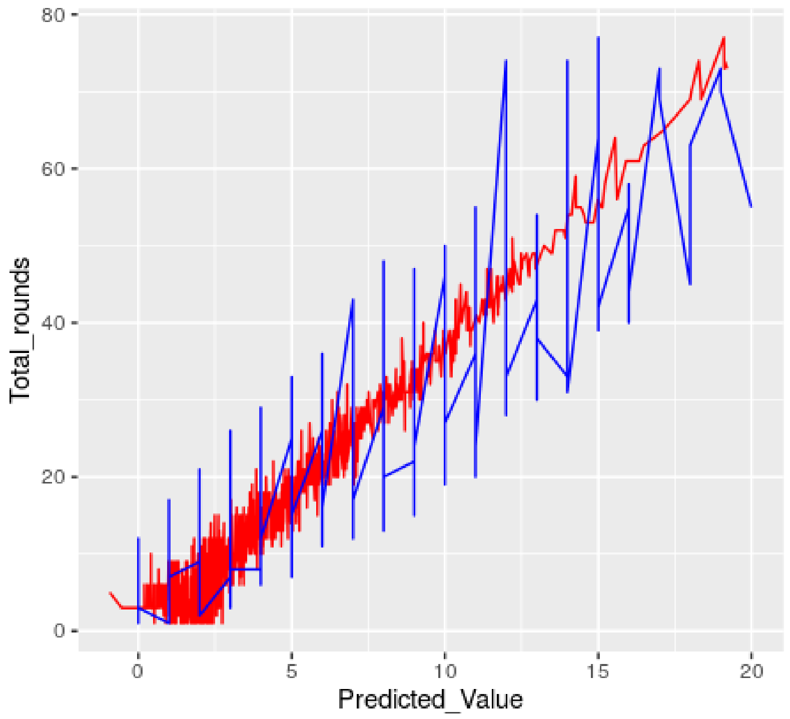
```
> #Plotting R_total_rounds_fought with winner using density graph
> ggplot(df, aes( x = R_total_rounds_fought, color=Winner)) + geom_density( )
```



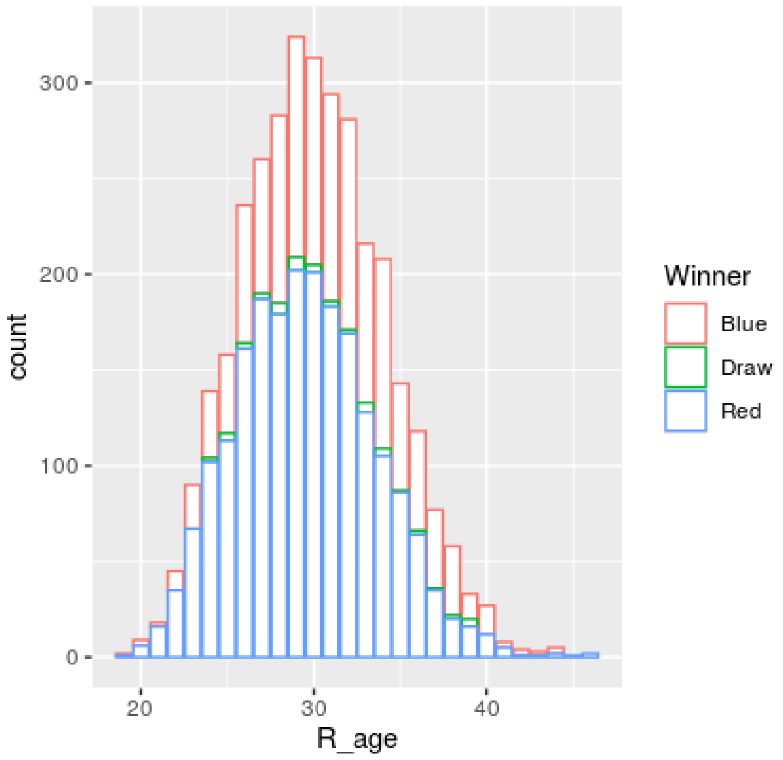
```
> #These two features were choosen as they had a significant hand in determining
the chances of R winning
>
```

Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> #Plotting regression model results of model R
> ggplot(data = resultR) + geom_line(aes(x=Predicted_Value, y=Total_rounds), colo
r='red') + geom_line(aes(x=Actual_Value, y=Total_rounds), color='blue')
```

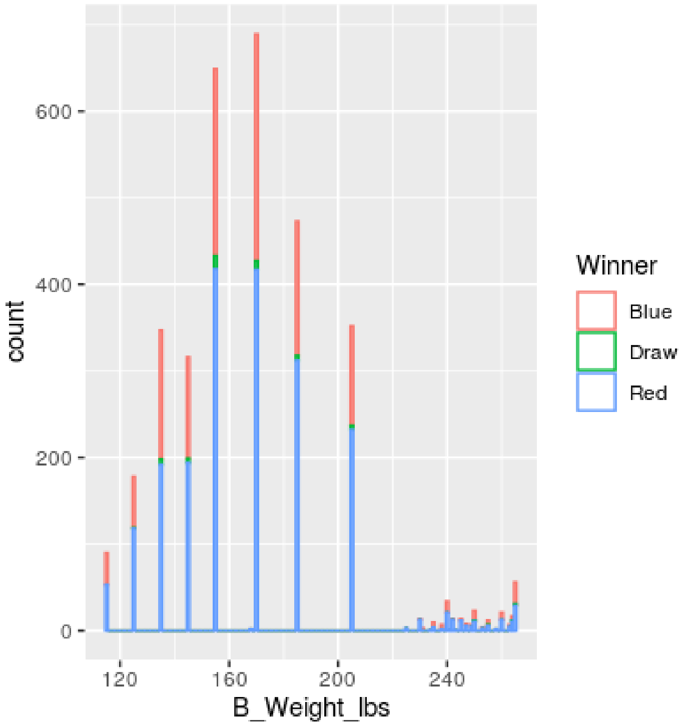


```
> #Plotting age of fighter to get an idea of the general age range of fighters
> ggplot(df, aes(x= R_age, color = Winner)) + geom_bar(fill='white')
```

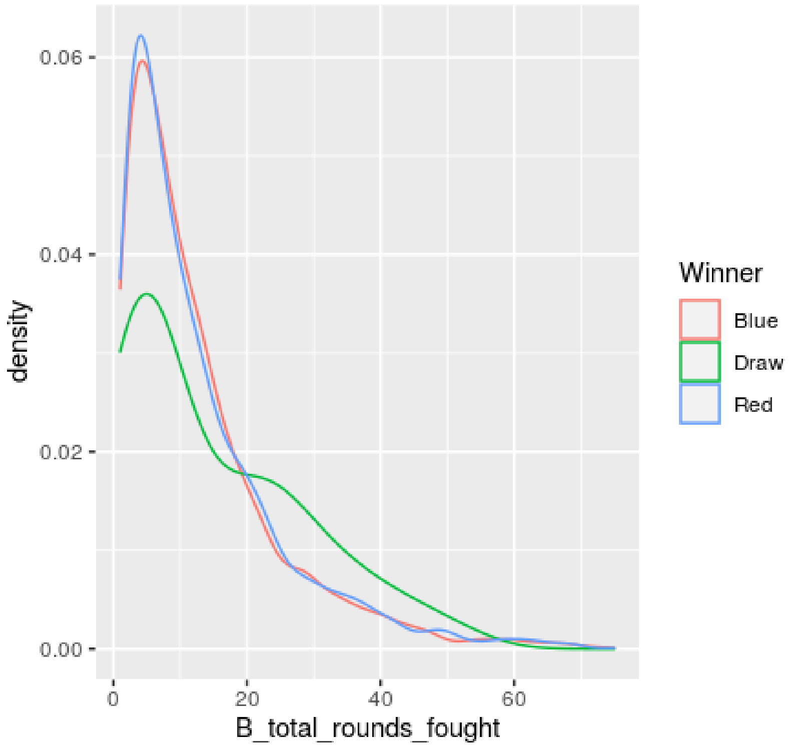


Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> #Plotting with B_weight and winner using a group histogram
> ggplot(data  = df, aes( x = B_Weight_lbs, color=Winner)) + geom_histogram(fill=
"white", binwidth = 1)
```

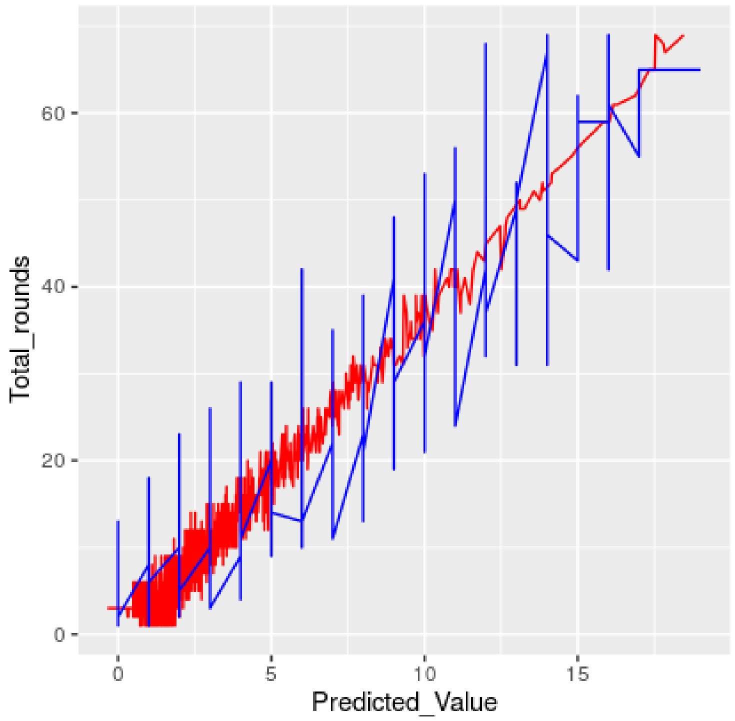


```
> #Plotting B_total_rounds_fought with winner using density graph
> ggplot(df, aes( x = B_total_rounds_fought, color=Winner)) + geom_density( )
```

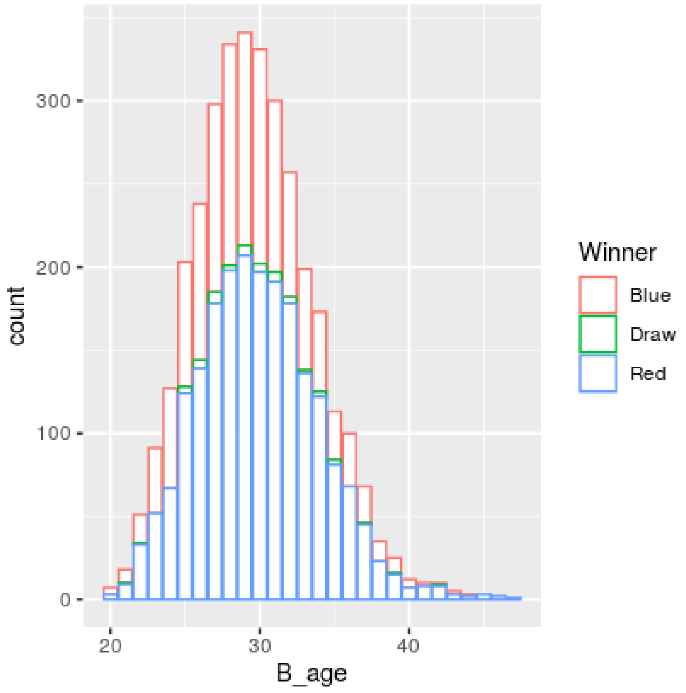


Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> #These two features were choosen as they had a significant hand in determining
the chances of B winning
> #Plotting regression model results of model B
> ggplot(data = resultB) + geom_line(aes(x=Predicted_Value, y=Total_rounds), colo
r='red') + geom_line(aes(x=Actual_Value, y=Total_rounds), color='blue')
```

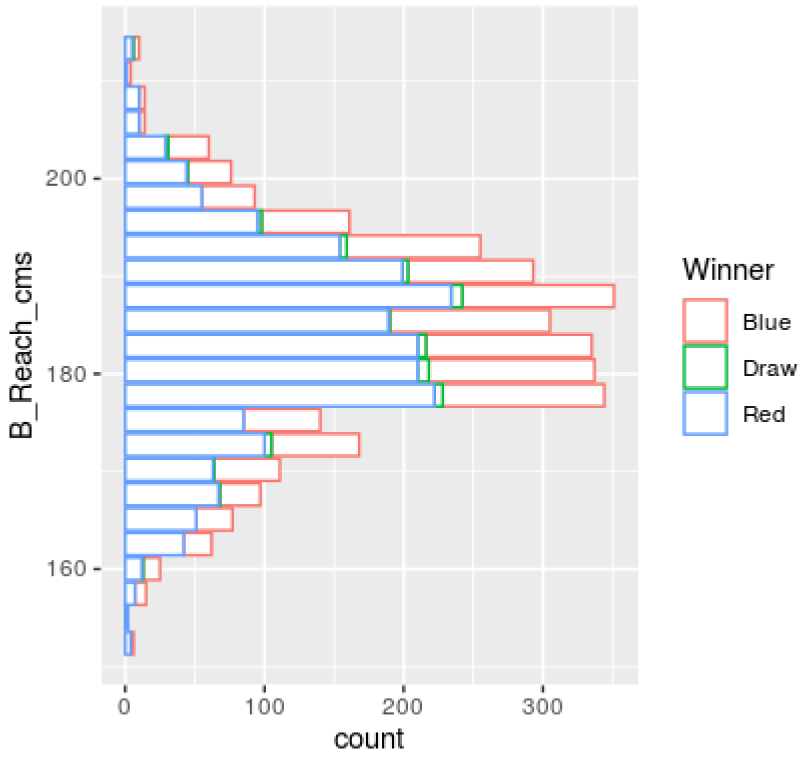


```
> #Plotting age of fighter to get an idea of the general age range of fighters
> ggplot(df, aes(x= B_age, color = Winner)) + geom_bar(fill='white')
```



Anirudh Poroorkara
Roll: 44
IT-1 B12

```
> #Plotting reach of B fighter
> ggplot(df, aes(x= B_Reach_cms , color = Winner)) + geom_bar(fill='white') + coo
rd_flip()
```



Conclusion:

Thus, data is visualized using ggplot2.

Lab Outcome: LO6