

Fédération Internationale de Football Association

Dataset Link: https://www.kaggle.com/karangadiya/fifa19

Abstract

The dataset has been taken from the above website and has a good collection of variables that includes the player strengths, team strengths.

The column data is the dataset are as follows:

- Age, Nationality, Overall, Potential, Club, Value, Wage, Preferred Foot
- Position, Jersey Number, Height, Weight
- Crossing, Finishing, Heading, Accuracy, Short Passing, Volleys, Dribbling, Curve, FK
 Accuracy, Long Passing, Ball Control, Acceleration, Sprint Speed, Agility, Reactions,
 Balance, Shot Power, Jumping, Stamina, Strength, Long Shots, Aggression, Interceptions,
 Positioning, Vision, Composure, and Release Clause.

About Data

This dataset is a 9-megabyte .csv file according to survey discovered in year early 2019, which has more than 89 columns, which has been shaped and modified in the Rstdio (.R File ATTACHMENT). The csv file has more than 18000 rows of data for each player including attacking, defensing and goalkeeping. There are 'NA' Values in some of the columns. We plan to clean the data and gather valuables intelligence from the dataset.

Specifications of the data

The data has few categorical variables like Nationality (country), whether player is right footed or left footed, we have a number of numerical variable like Age of the player, overall player value (out of 100). The height and weight ratio of the player, which says more about the fitness level of the player in the year 2019. The distribution of the data is projected by the age numerical quantity.

The central limit theorem is defined for understanding how exactly the shape of the distribution of mean, when we pluck a sample size from the whole data. We plot a histogram to show the values inside

Sampling of data is done for understanding the random size of plucking. Which makes the graph lean more towards the average or the mean of the data.

```
#football mentioned here is also called
 #american soccer
 library(tidyverse)
 library(UsingR)
 library(prob)
 library(sampling)
 football_data<-data.frame(read.csv("data.csv" ,header = TRUE))</pre>
 #view(football_data)
head(football_data, n=3)
                           Name Age
    0 158023
                       L. Messi
                                  31
      20801 Cristiano Ronaldo
                                  33
    2 190871
                      Neymar Jr
                                              Photo Nationality
1 https://cdn.sofifa.org/players/4/19/158023.png
                                                      Argentina
https://cdn.sofifa.org/players/4/19/20801.png
https://cdn.sofifa.org/players/4/19/190871.png
                                                       Portugal
                                                         Brazil
                                   Flag Overall Potential
                                                                            c1ub
1 https://cdn.sofifa.org/flags/52.png
                                             94
                                                        94
                                                                   FC Barcelona
2 https://cdn.sofifa.org/flags/38.png
                                             94
                                                        94
                                                                       Juventus
3 https://cdn.sofifa.org/flags/54.png
                                              92
                                                        93 Paris Saint-Germain
                                       Club.Logo
                                                      Value
                                                                Wage Special
1 https://cdn.sofifa.org/teams/2/light/241.png a,-110.5M a,-565K
                                                                        2202
  https://cdn.sofifa.org/teams/2/light/45.png
                                                     â,¬77M â,¬405K
                                                                        2228
  https://cdn.sofifa.org/teams/2/light/73.png â,-118.5M â,-290K
                                                                        2143
  Preferred.Foot International.Reputation Weak.Foot Skill.Moves
                                                                         work.Rat
1
            Left
                                                                  4 Medium/ Mediu
<mark>m</mark>
2
                                                                         High/ Lo
           Right
                                           5
                                                                  5
W
3
           Right
                                                     5
                                                                      High/ Mediu
m
   Body.Type Real.Face Position Jersey.Number
                                                       Joined Loaned.From
       Messi
1
                    Yes
                               RF
                                                  Jul 1, 2004
                                             10
                                                 Jul 10,
2
     Ronaldo
                    Yes
                               ST
                                                         2018
                                             10
                                                  Aug 3, 2017
      Neymar
                    Yes
                               LW
  Contract.Valid.Until Height Weight
                                         LS
                                               ST
                                                    RS
                                                         LW
                                                                    CF
                             '7 159 lbs 88+2 88+2
                                                  88+2 92+2 93+2
                   2021
                           6'2 1831bs 91+3 91+3 91+3 89+3 90+3 90+3 90+3 89+3
2
                   2022
3
                   2022
                           5'9 150lbs 84+3 84+3 84+3 89+3 89+3 89+3 89+3 89+3
   LAM
        CAM
             RAM
                    LM
                        LCM
                               \mathsf{CM}
                                 RCM
                                         RM
                                             LWB
                                                  LDM
                                                        CDM
                                                             RDM
                                                                   RWB
                            84+2
                                       91+2
            93+2
                  91+2
       93+2
                       84+2
                                 84+2
                                            64+2
                                                  61+2
                                                       61+2
                                                            61+2
                                                                  64+2
 88+3 88+3 88+3 88+3 81+3 81+3 81+3 88+3 65+3 61+3 61+3 61+3 65+3 61+3 53+3
 89+3 89+3 89+3 88+3 81+3 81+3 81+3 88+3 65+3 60+3 60+3 60+3 65+3 60+3 47+3
        RCB
              RB Crossing Finishing HeadingAccuracy ShortPassing Volleys
 47+2 47+2
            59 + 2
                        84
                                   95
                                                                  90
                                                                          86
                                   94
 53+3 53+3 61+3
                        84
                                                    89
                                                                  81
                                                                          87
3 47+3 47+3 60+3
                        79
                                   87
                                                    62
  Dribbling Curve FKAccuracy LongPassing BallControl Acceleration SprintSpeed
               93
                           94
1
         97
                                        87
                                                     96
                                                                   91
                                                                                86
         88
                81
                           76
                                                     94
                                                                   89
                                        77
                                                                                91
         96
               88
                           87
                                        78
                                                     95
                                                                   94
                                                                                90
  Agility Reactions Balance ShotPower Jumping Stamina Strength LongShots
```

1	91	95	95	85	68	72	59	94
2	87	96	70	95	95	88	79	93
3	96	94	84	80	61	81	49	82
	Aggression	Intercep	tions Po:	sitioning	Vision	Penalties	Composure	Marking
1	48		22	94	94	75	96	33
2	63		29	95	82	85	95	28
3	56		36	89	87	81	94	27
	StandingTag	kle Slid	lingTackl	e GKDiving	g GKHand	lling GKKi	cking GKPo:	sitioning
1		28	2	δ	5	11	15	14
2		31	2	3 7	7	11	15	14
3		24	3	3 9)	9	15	15
	GKReflexes	Release.	Clause					
1	8	â,-	226.5M					
2	11	â,-	127.1M					
3	11	â,-	228.1M					

After Cleaning the data:

```
> head(football_data,n=3)
                             Name Age Nationality Overall Potential
           ID
    0 158023
                                                           94
                                                                      94
                         L. Messi
                                          Argentina
                                    31
       20801 Cristiano Ronaldo
                                    33
                                           Portugal
                                                           94
                                                                      94
    2 190871
                                             Brazil
                                    26
                                                           92
                                                                      93
                       Neymar Jr
                   club
                          Value Wage Preferred.Foot International.Reputation
          FC Barcelona 110.5M 565K
1
                                                  Left
2
              Juventus
                            77M 405K
                                                Right
 Paris Saint-Germain 118.5M 290K
                                                Right
  Weak.Foot Skill.Moves
                            Body.Type Real.Face Position Jersey.Number
                                Messi
                                              Yes
                                                         RF
2
                              Ronaldo
                                                         ST
                                              Yes
                                                                         10
                         5
                               Neymar
                                              Yes
                                                         LW
  Contract.Valid.Until Height Crossing Finishing HeadingAccuracy
                                                                         ShortPassing
                             5'7
6'2
                    2021
                                        84
                                                    95
1
2
3
                    2022
                                                                                     81
                                        84
                                                    94
                                                                      89
                             5'9
                    2022
                                        79
                                                                      62
  Volleys Dribbling Curve FKAccuracy LongPassing BallControl Acceleration 86 97 93 94 87 96 91
                   88
       87
                          81
                                      76
                                                    77
                                                                  94
                                                                                89
                   96
                          88
                                                    78
                                                                  95
                                                                                94
  SprintSpeed Agility Reactions Balance ShotPower
                                                        Jumping Stamina Strength
                     91
                                95
                                         95
            86
                                                     85
                                                              68
                                                                       72
                                                                                 59
                     87
                                          70
            91
                                96
                                                     95
                                                              95
                                                                       88
                                                                                 79
            90
                     96
                                94
                                          84
                                                     80
                                                              61
                                                                       81
                                                                                 49
  LongShots Aggression Interceptions
                                                       Vision Penalties
                                                                          Composure
                                         Positioning
                                      22
29
                                                                                  96
                      48
                                                    94
                                                            94
                                                                       75
1
2
3
                                                            82
                                                    95
                                                                       85
          93
                      63
                                                                                  95
                                      36
                                                    89
                                                            87
                                                                                  94
                      56
                                                                       81
           StandingTackle SlidingTackle GKDiving GKHandling GKKicking
        33
                         28
                                        26
                                                    6
                                                               11
        28
                         31
                                        23
                                                               11
                                                                          15
                         24
                                                    9
                                                                9
                                                                          15
  GKPositioning
                 GKReflexes Release.Clause
              14
                                       226.5M
2
              14
                           11
                                       127.1M
              15
                           11
                                       228.1M
```

R-code:

```
x<-table(football_data$Nationality)
x<-head(sort(x ,decreasing = TRUE))</pre>
```

Explanation:

Tabulating the player's country and sorting it in decreasing would fetch the most number of players from a particular country .

Output:

X

```
England Germany Spain Argentina France Brazil 1662 1198 1072 937 914 827
```

Numeric variable

Now we try to understand the how many players are right-footed and how many player are left-footed

For that to implement, we need to understand the game has a goalkeeper, for whom the right-footed and left-footed does not imply correctly . We need to subset and reduce the vision of the equation to players on the field except the goalkeepers.

We introduce this as factor as the values are either right or left footed.

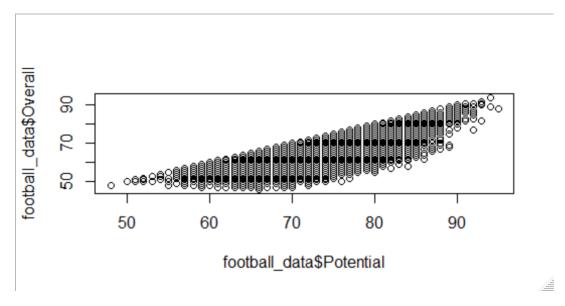
```
p<-as.factor(football_data$Preferred.Foot)
subdf<- subset(p,football_data$Preferred.Foot!="")
subdf
subdf<- table(subdf)
prop.table(subdf)</pre>
```

Now we look at the scatterplot between the overall ranking and the potential to which the player is capable off.

R-code:

plot(football_data\$Potential,football_data\$Overall)

Graph:



Explanation:

The above graph deals with exemplifying the coordination of potential, which is expected to be higher from a player when compared to the overall values.

Now we understand the sentiment of the player's Jersey Number as the player is free to choose the number what are the general selections made by the player's to which the adapt the jersey.

R-code:

fivenum(football_data\$Jersey.Number,na.rm = TRUE)

median(football_data\$Jersey.Number,na.rm = TRUE)

Explanation:

Fivenum summary is to know the q1 and q3 of the Jercy number, under which interval most of the players choose and some of the number are not present which has 'NA' values, which we make true to remove them during operations like mean, median, etc.

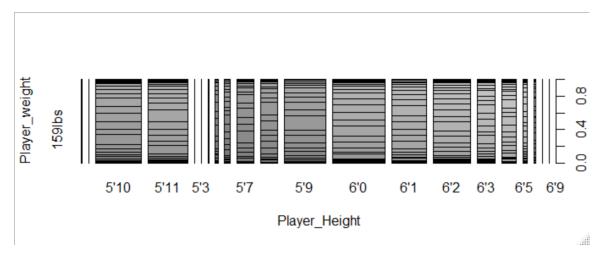
Output:

```
> fivenum(football_data$Jersey.Number,na.rm = TRUE)
[1]  1  8  17  26  99

median(football_data$Jersey.Number,na.rm = TRUE)
[1]  17
```

Health of the player's is one of the most important part as the players are provided with the best facilities and wages ,which indicates that the health and the height to body ratio of players should be really good.

Graph:



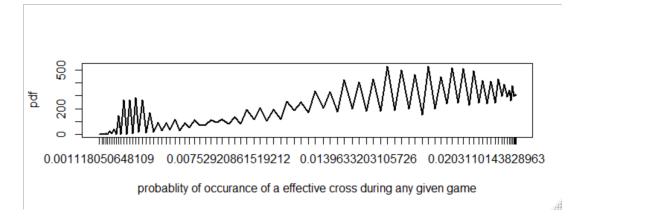
Crossing is one of major perks of a breakthrough in a game points .we calculate the capability of players who can give more number of connecting crosses to another player during a given running game .

R-code:

```
cross_mean<-mean(football_data$Crossing,na.rm =TRUE)
cross_sd<-sd(football_data$Crossing,na.rm =TRUE)
pdf1<-dnorm(football_data$Crossing,mean = cross_mean,sd = cross_sd)
pdf<-table(pdf1)
sum(pdf)

plot( x = pdf,type = "l",abline(h=0),xlab ="probablity of occurance")</pre>
```

Graph:



Top Ranked player in the having overall greater than 85 and the vision for observing the ball movement plays vital role for predicting the player is a world class player or not .

R-code:

```
ranked<-football data[(football data$Overall >= 85|football data$Vision >=85),]
```

head(ranked,n =18)

Output:

head(ranked, n =18)							
	ï	ID	Name	Age	Nationality	Overall	Potential
1	0	158023	L. Messi	31	Argentina	94	94
2	1	20801	Cristiano Ronaldo	33	Portugal	94	94
3	2	190871	Neymar Jr	26	Brazil	92	93
4	3	193080	De Gea	27	Spain	91	93
5	4	192985	K. De Bruyne	27	Belgium	91	92
6	5	183277	E. Hazard	27	Belgium	91	91
7	6	177003	L. Modrić	32	Croatia	91	91
8	7	176580	L. Suárez	31	Uruguay	91	91
9	8	155862	Sergio Ramos	32	Spain	91	91
10	9	200389	J. Oblak	25	Slovenia	90	93
11	10	188545	R. Lewandowski	29	Poland	90	90
12	11	182521	T. Kroos	28	Germany	90	90
13	12	182493	D. GodÃn	32	Uruguay	90	90
14	13	168542	David Silva	32	Spain	90	90
15	14	215914	N. Kanté	27	France	89	90
16	15	211110	P. Dybala	24	Argentina	89	94
17	16	202126	H. Kane	24	England	89	91
18	17	194765	A. Griezmann	27	France	89	90

Few Rows and columns are omitted

As by observing the games we understand that the player who has the opportunity to score the goal (striker) is payed to a great extent when compared to defense player. We try find the strikers of the football data set

R-code:

```
all_st<-football_data[(football_data$Position %in% "ST"),]
```

all_st

Output:

```
all_st
                               Name Age
naldo 33
             ID
                                                 Nationality Overall Potential
         20801 Cristiano Ronaldo
                                                    Portugal
      1
                                                                    94
                                                                                94
     10 188545
11
                    R. Lewandowski
                                      29
                                                       Poland
                                                                    90
                                                                                90
17
                        H. Kane
S. Agüero
     16 202126
                                                                    89
                                                                                91
                                                      England
24
     23
        153079
                                      30
                                                                                89
                                                   Argentina
                                                                    89
37
     36 173731
                            G. Bale
                                      28
                                                        wales
                                                                    88
                                                                                88
44
     43 201399
                                      25
                                                                                90
                         M. Icardi
                                                   Argentina
                                                                    87
                                                                                89
48
                                      25
     47 192505
                         R. Lukaku
                                                      Belgium
                                                                    87
                                                                                87
49
     48 192387
                       c. Immobile
                                      28
                                                        Italy
                                                                    87
                                      27
     87 193301
88
                      A. Lacazette
                                                       France
                                                                    85
                                                                                86
     98 180930
                         E. Då¾eko
                                      32 Bosnia Herzegovina
```

106 105 165153	K. Benzema	30	France	85	85
125 124 202556	M. Depay	24	Netherlands	84	89
131 130 192629	Iago Aspas	30	Spain	84	84
140 139 181783	M. Mandå¾ukiä‡	32	Croatia	84	84
146 145 176769	Jonas	34	Brazil	84	84
157 156 230666	Gabriel Jesus	21	Brazil	83	92
160 159 230294	Louri Beretta	26	Brazil	83	83
180 179 206113	S. Gnabry	22	Germany	83	88

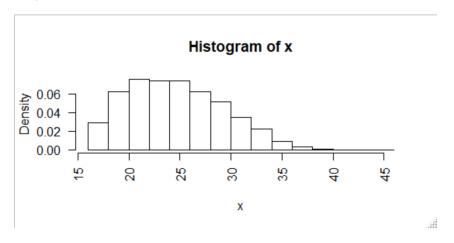
Centralized limit theorem

R-code:

```
#centrailzed limit theroem for the football data w.r.t Age element

x<-football_data$Age|
hist(x, prob = TRUE, breaks = 15, las =2)
#Distribution of the data
#normal
par
size<-20
xbar <- numeric(length(football_data$Age))
for(i in 1:length(football_data$Age)) {
   xbar[i] <- mean(rnorm(size, mean = mean(football_data$Age), sd = sd(football_data$Age))
}
hist(xbar, prob = TRUE, main = "Sample Size = 20")</pre>
```

Output:



Normal Distribution of the sample size varying from 20,30,40

R-code:

```
size<-20
xbar <- numeric(length(football_data$Age))
for(i in 1:length(football_data$Age)) {
   xbar[i] <- mean(rnorm(size, mean = mean(football_data$Age), sd = sd(football_data$Age)))
}</pre>
```

```
hist(xbar, prob = TRUE, main = "Sample Size = 20")

size<-30

xbar <- numeric(length(football_data$Age))

for(i in 1:length(football_data$Age)) {

    xbar[i] <- mean(rnorm(size, mean = mean(football_data$Age), sd = sd(football_data$Age)))

}

hist(xbar, prob = TRUE, main = "Sample Size = 30")

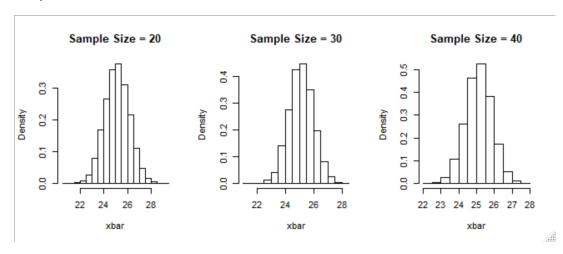
size<-40

xbar <- numeric(length(football_data$Age)) {

    xbar[i] <- mean(rnorm(size, mean = mean(football_data$Age), sd = sd(football_data$Age)))
}

hist(xbar, prob = TRUE, main = "Sample Size = 40")
```

Output:



Brazilians footballers are called the flair of soccer. Their touch to football is majestic, flexible .As a fifth largest data for Brazil I plan to sample some of the player details.

R-code:

```
> nrow(ST)
[1] 827
> size <-20
> sample<-srswor(size ,nrow(ST))
> table(sample)
sample
    0    1
807    20
```

```
> y<-ST[sample!=0,]
> f<-data.frame(table(y$Age))
> colnames(f)<-c("Age","samples")</pre>
 size <-30
 > sample<-srswor(size ,nrow(ST))</pre>
 > table(sample)
 sample
   0
          1
 797
        30
 y
Output:
     Age samples
                    1
1
 123456789
       20
       21
                    1
       23
       25
                    2
       26
                    4
                    2
1
       27
       28
       29
                    \bar{1}
                    1
       30
 10
                    1
       31
                    3
2
 11
       33
 12
       34
 У
                                           Name Age Nationality Overall Potential
          431 221540
                                          Raúĺ
 432
                                                                                                80
                                        Danilo
 646
          645 199304
                                                    26
                                                                Brazil
                                                                                 79
                                                                                                80
                                                                                 77
                                                                                                77
 1013 1012 140384
                                                    31
                                                                Brazil
                                             JÃ
 1370 1369 230250
                             Emilio Piodã£o
                                                    26
                                                                                 76
                                                                                                76
                                                                Brazil
 1436 1435 168607
                                       Rafinha
                                                    32
                                                                                 76
                                                                                                76
                                                                Brazil
 1818 1817 207734
2144 2143 230437
2827 2826 210016
                                     Paulinho
                                                    23
                                                                                                82
                                                                Brazil
                                                                                 75
                             Elias Fontoira
                                                    34
                                                                                 74
                                                                                                74
                                                               Brazil
                                 Diego Lopes
                                                    24
                                                                Brazil
                                                                                 73
                                                                                                77
   f
     Age samples 17 1
 1
2
3
4
5
6
7
8
9
10
                    1
       21
       23
                    3
       24
                    4
       25
                    1
       26
27
                    4
2
1
2
2
1
       28
       29
       30
 11
       31
                    1
1
 12
       32
 13
       33
       34
 14
   f<-data.frame(table(y$Age))
colnames(f)<-c("Age", "samples")</pre>
     Age samples
19 1
 1
2
       21
                    1
```

```
22
24
                        2
1
2
6
1
1
2
13
3456789
        25
        26
        27
        28
        29
10
        30
11
        31
                          1
                          8
1
12
        34
13
        36
```

France is the birth place of FIFA, one of the pride of French is PSG [Paris Saint-Germain] CLUB

R-code:

```
psg<-football_data[(football_data$Club =="Paris Saint-Germain"),]</pre>
psg
head(psg)
Output:
> head(psg)
                         Name Age Nationality Overall
                                                           Potential
                    Neymar Jr
       190871
                                26
                                          Brazil
                                                        92
                                                                    93
22
                                                        89
                                                                    89
    21 179813
                    E. Cavani
                                 31
                                         Uruguay
                                19
                                                        88
                                                                    95
26
    25 231747
                  K. Mbappé
                                          France
40
    39 164240
                                                                    88
                Thiago Silva
                                 33
                                          Brazil
                                                        88
42
                                                                    88
    41
          1179
                                40
                                           Italy
                                                        88
                    G. Buffon
    63 199556
                     Verratti
                                25
                                           Italy
                                                                    89
                                                        86
                     c1ub
                            Value Wage Preferred.Foot
                                   290K
   Paris Saint-Germain
                           118.5M
                                                   Right
```

Spain is famous for their playing style which is called tiki-taka (passing continuously)

81M 100K

60M

200K

Right

Right

```
R-code:
```

22 Paris Saint-Germain

26 Paris Saint-Germain

```
spain<-football_data[(football_data$Nationality == "Spain"),]</pre>
spain
head(spain)
Output:
head(spain)
                             Name Age Nationality Overall Potential
        193080
                                    27
                                                           91
                                                                       93
                          De Gea
                                              Spain
     8 155862
                    Sergio Ramos
                                    32
                                              Spain
                                                           91
                                                                       91
                                                                       90
    13 168542
                    David Silva
                                    32
                                                           90
                                              Spain
21
                                                           89
                                                                       89
    20 189511 Sergio Busquets
                                    29
                                              Spain
                                                                       91
31
    30 197781
                                    26
                                                           88
                             Isco
                                              Spain
                                                                       87
    49 189332
                                    29
                                                           87
                      Jordi Alba
                                              Spain
                  Club Value Wage Preferred. Foot International. Reputation
                                               Right
   Manchester United
                               260K
                           72M
ġ
                           51M 380K
                                                                                 4
          Real Madrid
                                               Right
                          60M 285K
                                                Left
                                                                                 4
14
     Manchester City
21
         FC Barcelona 51.5M 315K
                                               Right
```

Sampling:

Random sampling without replacement

For this we take the variables age, which is merely constant among most of the footballers . As the number of samples in the dataset are 18K I'd take a sample of the data of the Brazilians to apply sampling

```
Rcode: [size =21]
> age<-football_data[( football_data$Nationality == "Brazil"),]</pre>
> nrow(age)
[1] 827
> size<-21
> age_sample <-srswor(size,nrow(age))</pre>
> x<-table(age_sample)</pre>
> sample_age<-age[x!=0,]</pre>
Output:
> sample_age
    ï..
2
                             Name Age Nationality Overall Potential
         190871
                        Neymar Jr
                                    26
                                             Brazil
                                                           92
                                                                       93
28
                                                           88
                                                                       90
     27 200145
                         Casemiro
                                    26
                                             Brazil
     32 189242
33
                                                           88
                                                                       89
                         Coutinho
                                    26
                                             Brazil
36
     35 176676
                          Marcelo
                                    30
                                             Brazil
                                                           88
                                                                       88
40
     39 164240
                    Thiago Silva
                                    33
                                                           88
                                                                       88
                                             Brazil
                                                                       90
58
     57 210257
                          Ederson
                                    24
                                             Brazil
                                                           86
                                                                       87
62
     61 201942 Roberto Firmino
                                    26
                                                           86
                                             Brazil
65
     64 191043
                     Alex Sandro
                                    27
                                                           86
                                                                       86
                                             Brazil
     65 190483
                                             Brazil
66
                   Douglas Costa
                                    27
                                                           86
                                                                       86
         135507
76
     75
                     Fernandinho
                                    33
                                             Brazil
                                                           86
                                                                       86
     81 212831
                                    25
82
                                                           85
                                                                       90
                          Alisson
                                             Brazil
92
     91 187961
                         Paulinho |
                                    29
                                                                       85
                                             Brazil
                                                           85
103 102 171919
                            Naldo
                                    35
                                                                       85
                                             Brazil
                                                           85
105 104 168609
                          Miranda
                                    33
                                             Brazil
                                                           85
                                                                       85
107 106 164169
                    Filipe LuÃs
                                   32
                                            Brazil
                                                          85
                                                                      85
                                    25
                     Alex Telles
                                                                       87
113 112 212462
                                             Brazil
                                                           84
119 118 209499
                          Fabinho
                                    24
                                             Brazil
                                                           84
                                                                       88
Rcode:[size =31]
> size<-31
> age_sample <-srswor(size,nrow(age))</pre>
 x<-table(age_sample)</pre>
  sample_age < -age[x!=0,]
  sample_age
    i..
2 190871
                             Name Age Nationality Overall Potential
3
                                    26
                        Neymar Jr
                                             Brazil
                                                           92
                                                                       93
28
                                                           88
                                                                       90
     27 200145
                         Casemiro
                                    26
                                              Brazil
     32
                                                                       89
33
        189242
                         Coutinho
                                    26
                                             Brazil
                                                           88
36
     35 176676
                                                                       88
                          Marcelo
                                    30
                                             Brazil
                                                           88
                    Thiago Silva
40
     39 164240
                                    33
                                             Brazil
                                                           88
                                                                       88
58
                                                           86
                                                                       90
     57 210257
                          Ederson
                                    24
                                             Brazil
                                                                       87
62
     61 201942 Roberto Firmino
                                    26
                                             Brazil
                                                           86
65
                                                                       86
     64 191043
                     Alex Sandro
                                    27
                                             Brazil
                                                           86
66
     65 190483
                                    27
                                                                       86
                   Douglas Costa
                                                           86
                                             Brazil
76
     75
        135507
                     Fernandinho
                                    33
                                                                       86
                                             Brazil
                                                           86
                                             Brazil
     81 212831
                                    25
                                                                       90
82
                          Alisson
                                                           85
92
     91 187961
                         Paulinho Paulinho
                                    29
                                             Brazil
                                                           85
                                                                       85
                                                                       85
103 102 171919
                                    35
                            Naldo
                                             Brazil
                                                           85
105 104 168609
                                                           85
                                                                       85
                          Miranda
                                    33
                                             Brazil
107 106 164169
                    Filipe LuÃs
                                   32
                                                                      85
                                            Brazil
                                                          85
                                    25
113 112 212462
                     Alex Telles
                                             Brazil
                                                           84
                                                                       87
Rcode:
> size<-41
> age_sample <-srswor(size,nrow(age))</pre>
> x<-table(age_sample)</pre>
> sample_age<-age[x!=0,]</pre>
```

```
> sample_age
                              Name Age Nationality Overall Potential
              ID
3
       2
         190871
                                     26
                        Neymar Jr
                                              Brazil
                                                            92
                                                                        93
     27
         200145
28
                         Casemiro
                                                            88
                                                                        90
                                              Brazil
                                     26
33
     32
         189242
                         Coutinho
                                              Brazil
                                                            88
                                                                        89
                                     26
                                                                        88
36
     35
                                                            88
         176676
                          Marcelo
                                     30
                                              Brazil
         164240
40
     39
                    Thiago Silva
                                     33
                                                            88
                                                                        88
                                              Brazil
58
     57
         210257
                          Ederson
                                     24
                                              Brazil
                                                            86
                                                                        90
62
     61 201942 Roberto Firmino
                                     26
                                              Brazil
                                                                        87
                                                            86
65
     64 191043
                      Alex Sandro
                                     27
                                                            86
                                                                        86
                                              Brazil
66
     65 190483
                                     27
                                                            86
                                                                        86
                   Douglas Costa
                                              Brazil
                                                                        86
76
     75 135507
                                     33
                                                            86
                      Fernandinho
                                              Brazil
82
92
     81 212831
                          Alisson
                                     25
                                              Brazil
                                                            85
                                                                        90
     91 187961
                         Paulinho Paulinho
                                     29
                                                            85
                                                                        85
                                              Brazil
```

Systematic sampling of Brazilian footballers

Rcode:

```
a<-sample(ceiling(nrow(age)/size),1)</pre>
> b<-seq(a, by=ceiling(nrow(age)/size), length=size)</pre>
  ab <-age[b,]
 head(ab)
      ï.
                                 Name Age Nationality Overall Potential
28
       27 200145
                            Casemiro
                                        26
                                                Brazil
                                                              88
                                                                         90
159
      158 230375 Josuã© Chiamulera
                                       26
                                                Brazil
                                                              83
                                                                         83
                                                              82
276
      275 195093
                       Willian Josã©
                                        26
                                                                         83
                                                Brazil
      564 230312
                                                              79
                                                                         79
565
                      Gabri Prestã£o
                                        26
                                                Brazil
      758 224116
759
                         Dyego Sousa
                                        28
                                                Brazil
                                                              78
                                                                         78
1007 1006 207707
                                                              77
                                                                         77
                                        29
                              Marã§al
                                                Brazil
                     Club Value Wage Preferred.Foot
28
             Real Madrid 59.5M 285K
                                                Right
159
                 Grãamio 22.5M
                                  43K
                                                Right
276
           Real Sociedad 26.5M
                                  44K
                                                Right
```

Extra

Tabulate the length of players name in dataset
> table(str_length(football_data\$Name))

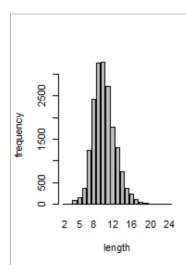
```
10
   1
         6
              89
                         365 1252 2409
                                         3248 3259
                   166
  11
              13
                          15
                                16
                                      17
                                            18
                                                  19
        12
                    14
2718 1774 1317
                   757
                         371
                               244
                                     106
                                            59
                                                  34
                    23
  20
        21
              22
                          24
        12
  16
                           1
```

Plotting a barplot for all the frequency of player name length occurance

Rcode:

barplot(table(str_length(football_data\$Name)), xlab = "length", ylab = "frequency")

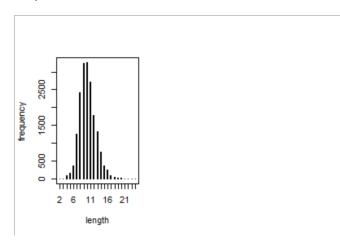
Graph:



Rcode:

plot(table(str_length(football_data\$Name)), xlab = "length", ylab = "frequency")

Graph:



Smallest player's name in the dataset

Rcode:

football_data\$Name[str_length(football_data\$Name) == min(str_length(football_data\$Name))]

Output:

> football_data\$Name[str_length(football_data\$Name) == min(str_length(footbal
l_data\$Name))]
[1] PC

All the player names which start with the letter P

football_data\$Name[str_detect(football_data\$Name, "^P")]

> football_data\$Name[str_detect(football_data\$Name, "^P")] [1] P. Dybala P. Aubameyang P. Pogba [4] Piqué [7] Pepe Parejo Paulinho P. Kimpembe Pizzi [10] Pepe Reina Pau López Pablo Sarabia [13] Pedro [16] P. Zieliński [19] Pacheco [22] P. Gulácsi [25] P. Lees-Melou P. ÄŒech Pablo Fornals Paco Alcã;cer Pedro León P. KadeÅ™Ã;bek P. Wernbloom Portu P. Jones Petros Pablo Maffeo [28] P. Sisto P. Max Pere Pons P. Hernã¦ndez P. Diop P. Jagielka P. Hetemaj P. Bargfrede [31] Pozuelo Pedro Mendes [34] P. Pérez P. Groß [37] Pedro Obiang P. Cutrone [40] P. Skjelbred [43] Paulinho P. Herrmann P. Schick P. Bargfrede

There are many rows and columns which are not present here as the space won't be sufficient.