

FIFA 20 EDA

Naga Santhosh Kartheek Karnati

3/27/2020

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
#SM Project
#FIFA player ratings datasets

library(ggplot2)
library(dplyr)

## 
## Attaching package: 'dplyr'

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

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

library(tidyr)
library(stringr)
library(data.table)

## 
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
## 
##     between, first, last

library(rio)

## Warning: package 'rio' was built under R version 3.6.2

library(modelr)
library(purrr)

## 
## Attaching package: 'purrr'
```

```

## The following object is masked from 'package:data.table':
##
##      transpose

#loading the datasets:
#FIFA20 players dataset:

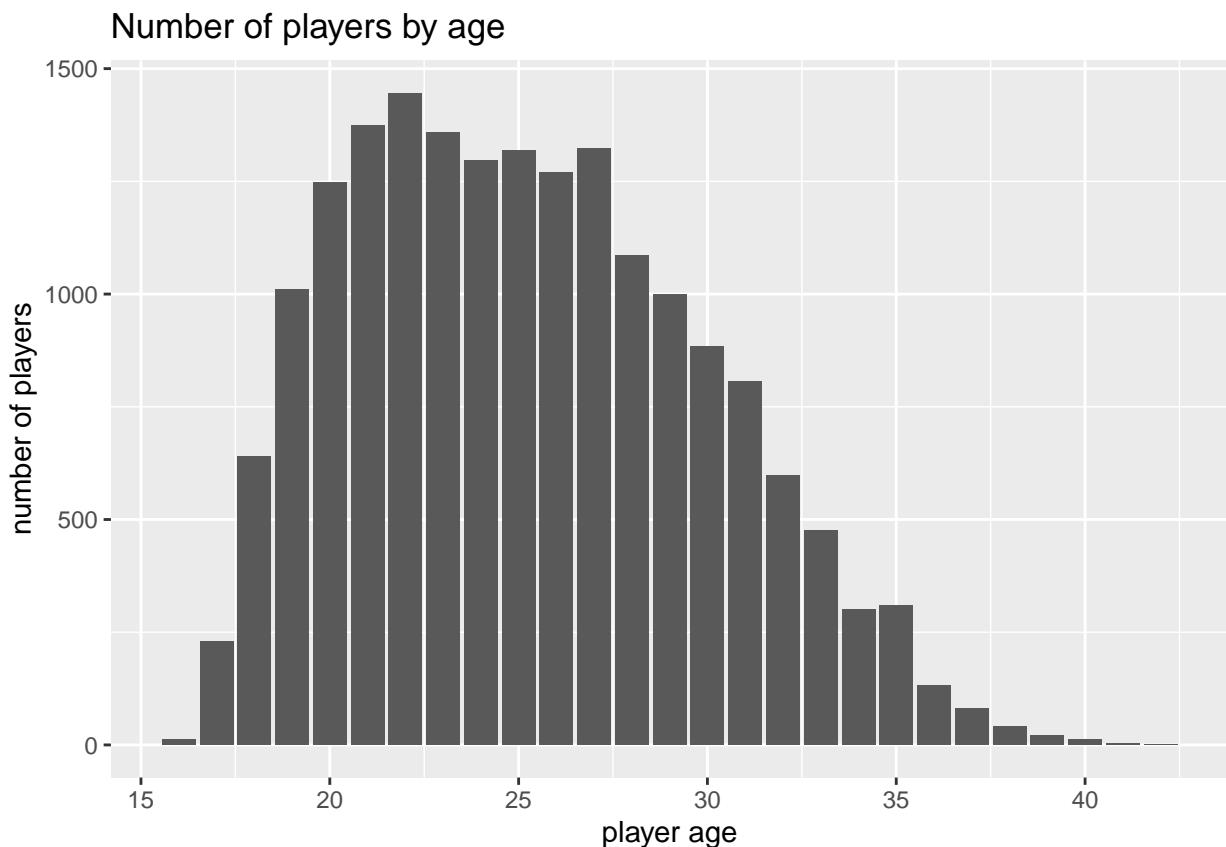
fifa20 <- fread('D:/NEU/Spring 2020/SML/Project/Datasets/players_20.csv')
class(fifa20)

## [1] "data.table" "data.frame"

#View(fifa20)
#fifa20 as a tibble
fifa20 <- as_tibble(fifa20)

#EDA and visualization:
#Basic visualization:
#Histogram of player age
fifa20 %>% ggplot(aes(age))+geom_bar()+labs(x="player age", y="number of players",
                                               title="Number of players by age" )

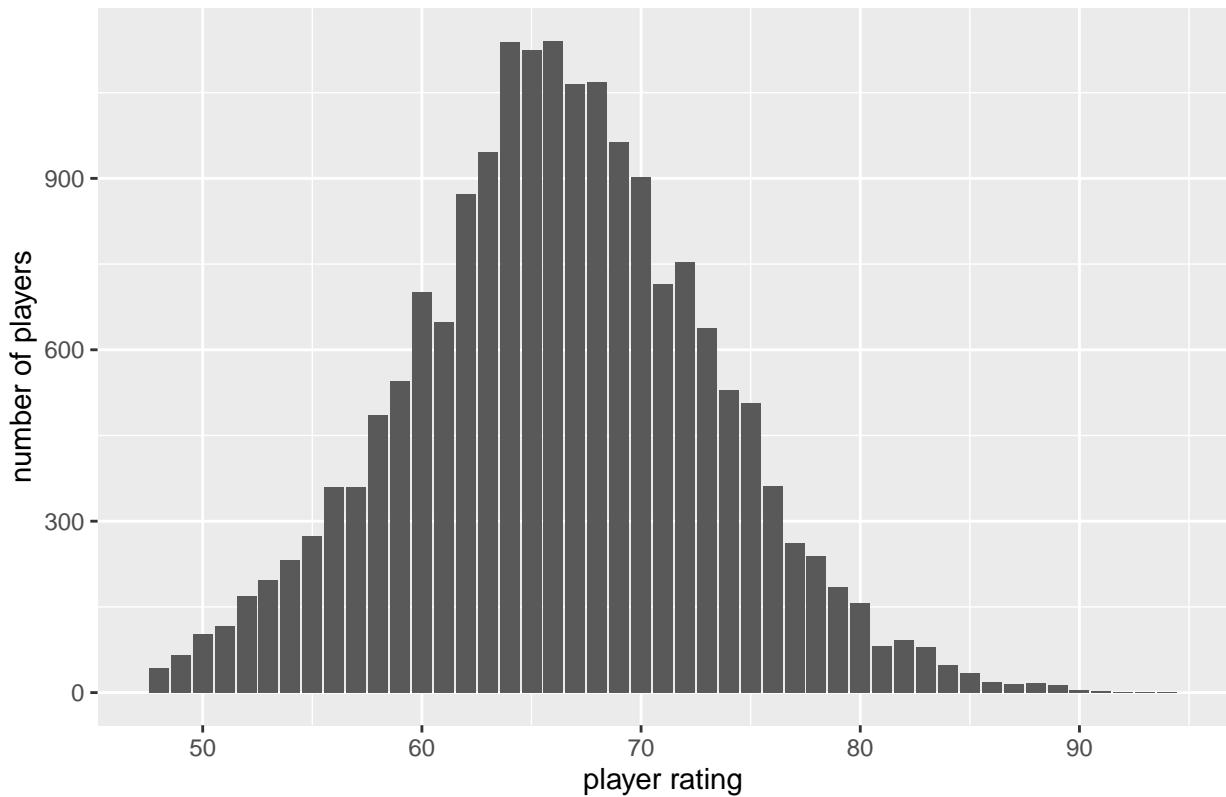
```



```
#Most common age for players is 22.
```

```
#Histogram of player rating
fifa20 %>% ggplot(aes(overall)) + geom_bar() + labs(x = "player rating", y = "number of players",
                                               title = "Number of players by rating" )
```

Number of players by rating



#most common rating range for players is 64-66.

```
#Which country has the most players on FIFA20:
fifa20 %>% group_by(nationality) %>% count() %>%
  arrange(desc(n))
```

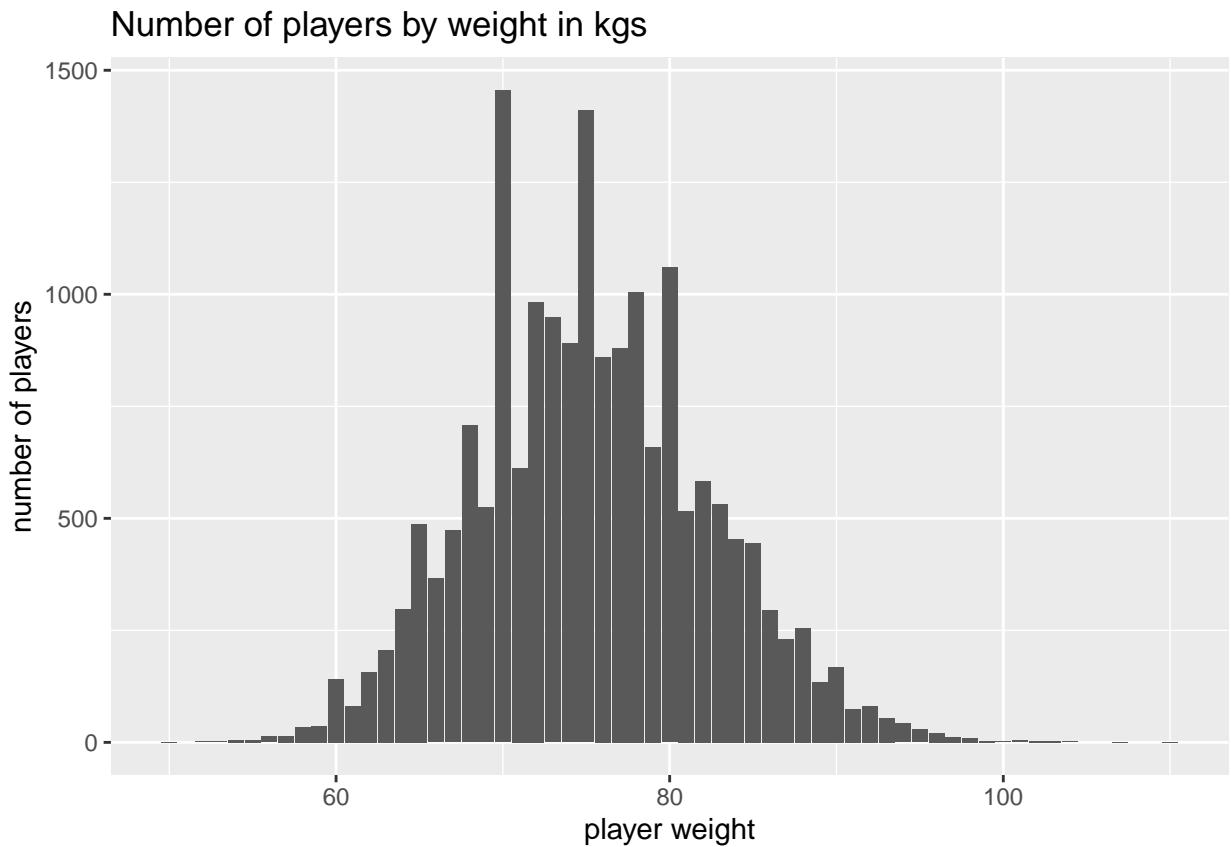
```
## # A tibble: 162 x 2
## # Groups:   nationality [162]
##   nationality     n
##   <chr>        <int>
## 1 England      1667
## 2 Germany      1216
## 3 Spain         1035
## 4 France        984
## 5 Argentina     886
## 6 Brazil         824
## 7 Italy          732
## 8 Colombia       591
## 9 Japan          453
## 10 Netherlands    416
```

```
## # ... with 152 more rows
```

```
#The country with most players on FIFA20 is England.
```

```
#Histogram of weight in kgs
```

```
fifa20 %>% ggplot(aes(weight_kg)) + geom_bar() + labs(x = "player weight", y = "number of players",  
title = "Number of players by weight in kgs")
```

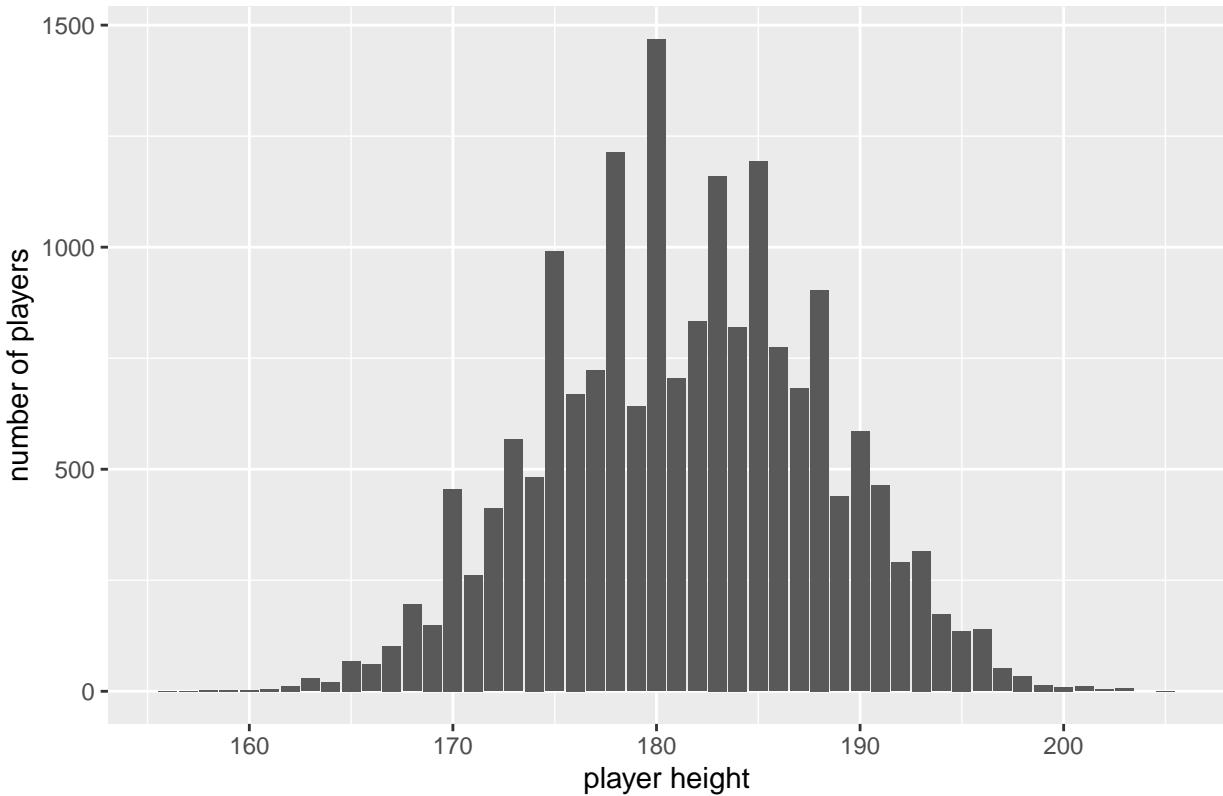


```
#70 and 75 kgs seem to be the most and 2nd most common player weights on FIFA20.
```

```
#Histogram of height in cms
```

```
fifa20 %>% ggplot(aes(height_cm)) + geom_bar() + labs(x = "player height", y = "number of players",  
title = "Number of players by height in cms")
```

Number of players by height in cms



```
#180 cms seems to be the most common player height on fifa20.
```

```
#Top 10 players with highest value in euros
fifa20 %>% arrange(desc(value_eur)) %>% select(short_name, club, nationality,
overall, value_eur)
```

```
## # A tibble: 18,278 x 5
##   short_name   club           nationality overall value_eur
##   <chr>        <chr>          <chr>       <int>    <int>
## 1 Neymar Jr   Paris Saint-Germain Brazil      92 105500000
## 2 L. Messi     FC Barcelona     Argentina    94 95500000
## 3 K. Mbapp    Paris Saint-Germain France     89 93500000
## 4 E. Hazard    Real Madrid     Belgium      91 90000000
## 5 K. De Bruyne Manchester City  Belgium      91 90000000
## 6 H. Kane      Tottenham Hotspur England     89 83000000
## 7 M. Salah     Liverpool       Egypt       90 80500000
## 8 V. van Dijk  Liverpool       Netherlands  90 78000000
## 9 J. Oblak     Atl  tico Madrid Slovenia    91 77500000
## 10 P. Dybala   Juventus      Argentina    88 76500000
## # ... with 18,268 more rows
```

```
#Tibble of 10 most valuable players on fifa20.
```

```
#Top 10 players with highest weekly wage in euros
fifa20 %>% arrange(desc(wage_eur)) %>% select(short_name, club, nationality,
                                              overall, wage_eur)
```

```
## # A tibble: 18,278 x 5
##   short_name       club      nationality overall wage_eur
##   <chr>         <chr>      <chr>        <int>    <int>
## 1 L. Messi     FC Barcelona Argentina     94    565000
## 2 E. Hazard    Real Madrid  Belgium      91    470000
## 3 Cristiano Ronaldo Juventus Portugal    93    405000
## 4 K. De Bruyne Manchester City Belgium    91    370000
## 5 A. Griezmann FC Barcelona France      89    370000
## 6 L. Suárez    FC Barcelona Uruguay     89    355000
## 7 L. Modrić    Real Madrid  Croatia     90    340000
## 8 T. Kroos     Real Madrid  Germany     88    330000
## 9 S. Agüero    Manchester City Argentina 89    300000
## 10 Sergio Ramos Real Madrid  Spain       89    300000
## # ... with 18,268 more rows
```

#Tibble of players with highest weekly wage in euros.

```
#Top 10 players with best rating
fifa20 %>% arrange(desc(overall)) %>% select(short_name, club, nationality, overall)
```

```
## # A tibble: 18,278 x 4
##   short_name       club      nationality overall
##   <chr>         <chr>      <chr>        <int>
## 1 L. Messi     FC Barcelona Argentina     94
## 2 Cristiano Ronaldo Juventus Portugal    93
## 3 Neymar Jr    Paris Saint-Germain Brazil     92
## 4 J. Oblak     Atlético Madrid Slovenia     91
## 5 E. Hazard    Real Madrid  Belgium      91
## 6 K. De Bruyne Manchester City Belgium     91
## 7 M. ter Stegen FC Barcelona Germany     90
## 8 V. van Dijk  Liverpool   Netherlands   90
## 9 L. Modrić    Real Madrid  Croatia     90
## 10 M. Salah    Liverpool   Egypt       90
## # ... with 18,268 more rows
```

#Tibble of top 10 rated players on fifa20.

```
#####ABSTRACT HYPOTHESIS: there exists a positive
#correlation between player overall and value in euros
#Is there correlation b/w pairs of features in {overall, value_eur, wage_eur}
ovr_val_wag <- fifa20 %>% select(overall, value_eur, wage_eur)
cor_ovr_val_wag <- cor(ovr_val_wag)
round(cor_ovr_val_wag, 2)
```

```
##          overall value_eur wage_eur
## overall     1.00     0.64     0.57
```

```

## value_eur      0.64      1.00      0.86
## wage_eur      0.57      0.86      1.00

#one would expect value and wage to be highly and positively correlated with overall but
#the correlation isn't that strong. That being said, overall and wage are +vely correlated.
#High +ve correlation exists b/w value and wage though.
#Therefore, our hypothesis is true.

```

```

#Top 10 players with the best potential overall
fifa20 %>% arrange(desc(potential)) %>% select(short_name, club,
                                              nationality, overall, potential)

```

```

## # A tibble: 18,278 x 5
##   short_name       club     nationality overall potential
##   <chr>          <chr>     <chr>        <int>     <int>
## 1 K. MbappÃ©    Paris Saint-Germain France        89      95
## 2 L. Messi       FC Barcelona Argentina      94      94
## 3 Cristiano Ronaldo Juventus Portugal        93      93
## 4 J. Oblak       AtlÃ©tico Madrid Slovenia      91      93
## 5 M. ter Stegen  FC Barcelona Germany        90      93
## 6 M. de Ligt     Juventus Netherlands      85      93
## 7 JoÃ£o FÃ©lix   AtlÃ©tico Madrid Portugal       80      93
## 8 Neymar Jr     Paris Saint-Germain Brazil       92      92
## 9 P. Dybala     Juventus Argentina      88      92
## 10 L. SanÃ©      Manchester City Germany       86      92
## # ... with 18,268 more rows

```

```
#top 10 players with best potential on fifa20.
```

```

#Players aged 25 or less with best potential:
fifa20 %>%
  filter(age <= 25) %>%
  arrange(desc(potential)) %>%
  select(short_name, club, nationality, overall, potential)

```

```

## # A tibble: 9,933 x 5
##   short_name       club     nationality overall potential
##   <chr>          <chr>     <chr>        <int>     <int>
## 1 K. MbappÃ©    Paris Saint-Germain France        89      95
## 2 M. de Ligt     Juventus Netherlands      85      93
## 3 JoÃ£o FÃ©lix   AtlÃ©tico Madrid Portugal       80      93
## 4 P. Dybala     Juventus Argentina      88      92
## 5 L. SanÃ©      Manchester City Germany       86      92
## 6 G. Donnarumma Milan Italy             85      92
## 7 J. Sancho     Borussia Dortmund England       84      92
## 8 K. Havertz    Bayer 04 Leverkusen Germany      84      92
## 9 VinÃ©cius Jr. Real Madrid Brazil           79      92
## 10 H. Kane      Tottenham Hotspur England       89      91
## # ... with 9,923 more rows

```

```
#Players aged 21 or less with best potential:
fifa20 %>%
  filter(age <= 21) %>%
  arrange(desc(potential)) %>%
  select(short_name, club, nationality, overall, potential)
```

```
## # A tibble: 4,514 x 5
##   short_name       club      nationality overall potential
##   <chr>        <chr>      <chr>     <int>     <int>
## 1 K. Mbappé          Paris Saint-Germain France      89      95
## 2 M. de Ligt         Juventus Netherlands      85      93
## 3 João Félix        Atlético Madrid Portugal      80      93
## 4 G. Donnarumma      Milan      Italy          85      92
## 5 J. Sancho          Borussia Dortmund England      84      92
## 6 K. Havertz         Bayer 04 Leverkusen Germany      84      92
## 7 Vinícius Jr.       Real Madrid Brazil          79      92
## 8 P. Foden           Manchester City England      76      90
## 9 T. Alexander-Arnold Liverpool England          83      89
## 10 L. Jović          Real Madrid Serbia          83      89
## # ... with 4,504 more rows
```

```
#Player with highest scope for increase in overall
# ie Potential-Overall
pot_inc <- fifa20 %>%
  mutate(pot_increase = potential-overall) %>%
  arrange(desc(pot_increase)) %>%
  select(short_name, club, nationality, age, overall, potential, pot_increase)
pot_inc
```

```
## # A tibble: 18,278 x 7
##   short_name   club      nationality     age overall potential pot_increase
##   <chr>        <chr>      <chr>     <int>     <int>     <int>     <int>
## 1 G. Bazunu    Manchester ~ Republic of I~     17      59      84      25
## 2 S. Ramos Mi~ Boca Juniors Argentina      17      56      81      25
## 3 B. Mumba     Sunderland England          17      55      80      25
## 4 S. Spasov    Oxford Unit~ Bulgaria      17      49      74      25
## 5 Tao Qianglo~ Hebei China~ China PR      17      48      73      25
## 6 B. McPherson Grimsby Town England      17      48      73      25
## 7 K. Bafounta  Borussia Do~ France          17      59      83      24
## 8 L. Chevalier LOSC Lille France          17      58      82      24
## 9 J. García    Cruz Azul Mexico          17      55      79      24
## 10 H. Mnoga    Portsmouth England         17      53      77      24
## # ... with 18,268 more rows
```

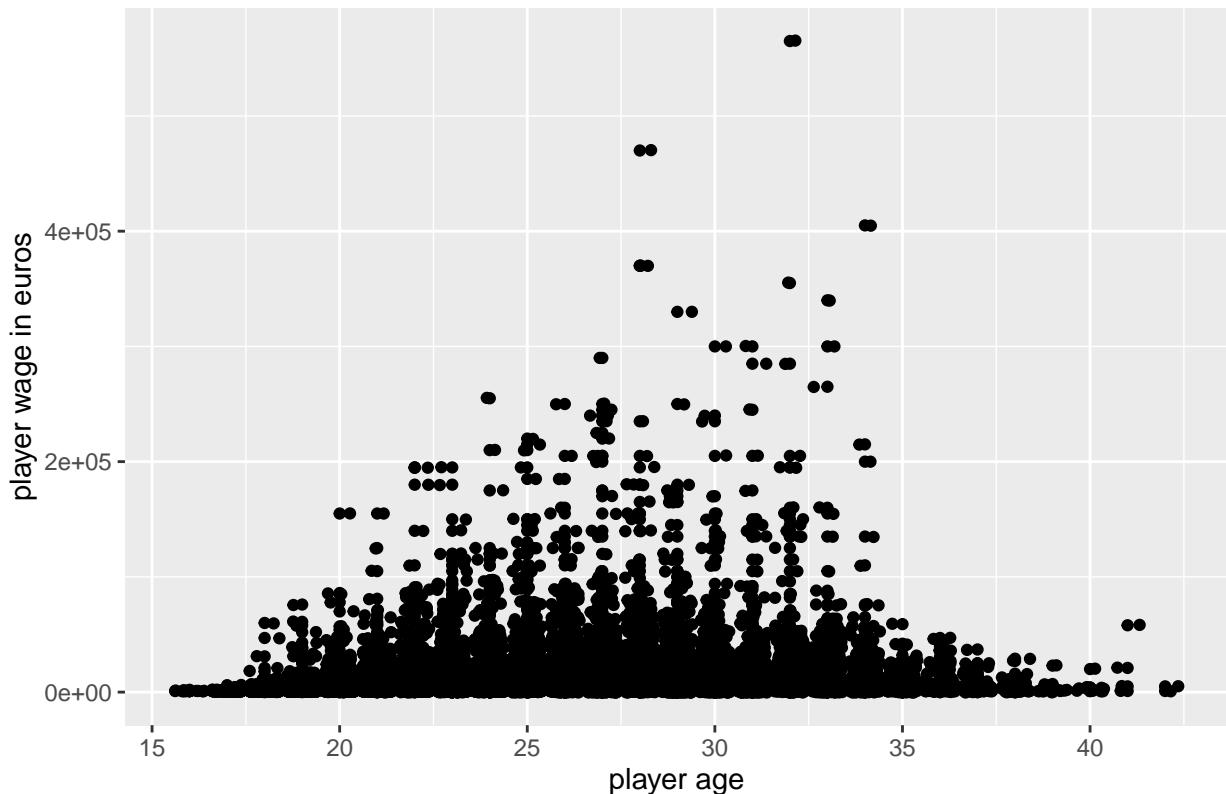
```
#we see that young players have scope for highest increase in rating.
#so are pot_increase and age positively correlated?
pot_inc_age_pot_inc <- pot_inc %>% select(age, pot_increase)
cor_pot_inc_age_pot_inc <- cor(pot_inc_age_pot_inc)
round(cor_pot_inc_age_pot_inc, 2)
```

```
##               age pot_increase
## age            1.00      -0.87
## pot_increase -0.87       1.00
```

```
#age and pot_increase have high -ve correlation.
```

```
#####ABSTRACT HYPOTHESIS:  
##### WAGE AND AGE CORRELATION:  
#scatter plot of wage and age  
fifa20 %>% ggplot(aes(age, wage_eur))+geom_point()+geom_jitter() +  
  labs(x="player age", y="player wage in euros",  
    title = "wage vs age scatter plot")
```

wage vs age scatter plot



```
## our hypothesis:  
#player wage and age are positively correlated upto the age of 31 and  
#negatively correlated after that  
players31_andless <- fifa20 %>% filter(age <= 31)%>% select(age, wage_eur, value_eur)  
players_over31 <- fifa20 %>% filter(age > 31)%>% select(age, wage_eur, value_eur)  
  
cor_31andless <- cor(players31_andless)  
round(cor_31andless, 2)
```

```
##          age wage_eur value_eur  
## age      1.00    0.19     0.15  
## wage_eur  0.19    1.00     0.87  
## value_eur 0.15    0.87     1.00
```

```
#age and wage are +vely correlated but the correlation is not very high.
```

```
cor_over31 <- cor(players_over31)
round(cor_over31, 2)
```

```
##           age wage_eur value_eur
## age       1.00   -0.12   -0.18
## wage_eur  -0.12    1.00    0.92
## value_eur -0.18    0.92    1.00
```

```
#age and wage are -vely correlated but the correlation is not very high.
#Therefore we dont reject the hypothesis.
```

#####ABSTRACT HYPOTHESIS

```
#left footed players have a higher overall rating compared to right footed players
fifa20 %>% filter(preferred_foot=="Left")%>%
  summarise(avg_player_overall = mean(overall))
```

```
## # A tibble: 1 x 1
##   avg_player_overall
##             <dbl>
## 1             66.7
```

```
#average overall rating of left footed player on fifa20 is 66.7
```

```
fifa20 %>% filter(preferred_foot=="Right")%>%
  summarise(avg_player_overall = mean(overall))
```

```
## # A tibble: 1 x 1
##   avg_player_overall
##             <dbl>
## 1             66.1
```

```
#average overall rating of left footed player on fifa20 is 66.1
#Our hypothesis is true but not by a huge margin.
```

#####6 PHYSICAL ATTRIBUTES EDA

```
#EDA on pace, shooting, passing, dribbling, defending and physique
#Top 10 fastest players on fifa20
```

```
fifa20 %>% arrange(desc(pace)) %>% select(short_name, club, nationality, overall,
  team_position, pace)
```

```
## # A tibble: 18,278 x 6
##   short_name      club          nationality overall team_position  pace
##   <chr>        <chr>          <chr>      <int> <chr>      <int>
## 1 K. Mbapp    Paris Saint-Germain  France        89  RW         96
## 2 Adama Traor   Wolverhampton Wander~ Spain        74  SUB         96
## 3 L. San       Manchester City   Germany       86  SUB         95
```

```

## 4 Gelson Martins AS Monaco Portugal 82 RM 95
## 5 K. Nagai FC Tokyo Japan 69 LS 95
## 6 A. ChalÃ¡ Deportivo Toluca Ecuador 66 LWB 95
## 7 P. Aubameyang Arsenal Gabon 88 LM 94
## 8 S. ManÃ© Liverpool Senegal 88 LW 94
## 9 Douglas Costa Juventus Brazil 84 SUB 94
## 10 IÃ±aki Williams Athletic Club de Bil~ Spain 82 RW 94
## # ... with 18,268 more rows

```

```

#Top 10 best shooters of the ball on fifa20
fifa20 %>% arrange(desc(shooting)) %>% select(short_name, club, nationality, overall,
team_position, shooting)

```

```

## # A tibble: 18,278 x 6
##   short_name     club      nationality overall team_position shooting
##   <chr>       <chr>      <chr>     <int> <chr>        <int>
## 1 Cristiano Ronaldo Juventus Portugal    93 LW        93
## 2 L. Messi      FC Barcelona Argentina  94 RW        92
## 3 H. Kane       Tottenham Hotspur England   89 ST        91
## 4 S. AgÃ¼ero   Manchester City Argentina  89 ST        90
## 5 L. SuÃ¡rez    FC Barcelona Uruguay    89 ST        89
## 6 F. Quagliarella Sampdoria Italy      83 ST        89
## 7 M. Reus       Borussia Dortmund Germany   88 CAM       88
## 8 Z. IbrahimoviÄ‡ LA Galaxy Sweden     85 ST        88
## 9 R. Lewandowski FC Bayern MÃ¼nchen Poland    89 ST        87
## 10 G. Bale      Real Madrid Wales      85 SUB       87
## # ... with 18,268 more rows

```

#Strikers are the best shooters

```

#Top 10 best passers on fifa20
fifa20 %>% arrange(desc(passing)) %>% select(short_name, club, nationality, overall,
team_position, passing)

```

```

## # A tibble: 18,278 x 6
##   short_name     club      nationality overall team_position passing
##   <chr>       <chr>      <chr>     <int> <chr>        <int>
## 1 L. Messi      FC Barcelona Argentina  94 RW        92
## 2 K. De Bruyne  Manchester City Belgium   91 RCM       92
## 3 C. Eriksen    Tottenham Hotspur Denmark   88 RM        90
## 4 T. Kroos      Real Madrid Germany    88 LCM       90
## 5 L. ModriÄ‡    Real Madrid Croatia    90 RCM       89
## 6 David Silva  Manchester City Spain     88 LCM       88
## 7 Parejo        Valencia CF Spain      86 RCM       88
## 8 Neymar Jr    Paris Saint-Germain Brazil    92 CAM       87
## 9 M. PjanÄ‡      Juventus Bosnia Herzegovna 86 CDM       87
## 10 Bruno Fernandes Sporting CP Portugal   85 RCM       87
## # ... with 18,268 more rows

```

#midfielders are the best passers with the exception of messi who is a winger.

#Top 10 best dribblers on fifa20

```
fifa20 %>% arrange(desc(dribbling)) %>% select(short_name, club, nationality, overall,
                                                team_position, dribbling)
```

```
## # A tibble: 18,278 x 6
##   short_name     club      nationality overall team_position dribbling
##   <chr>       <chr>      <chr>      <int> <chr>        <int>
## 1 L. Messi    FC Barcelona Argentina     94 RW         96
## 2 Neymar Jr   Paris Saint-Germa~ Brazil      92 CAM        95
## 3 E. Hazard    Real Madrid    Belgium     91 LW         94
## 4 Bernardo Silva Manchester City Portugal     87 RW         92
## 5 D. Mertens   Napoli       Belgium     87 SUB        91
## 6 L. Insigne   Napoli       Italy       87 LS         91
## 7 Isco        Real Madrid    Spain       86 SUB        91
## 8 Douglas Costa Juventus     Brazil      84 SUB        91
## 9 K. Mbapp     Paris Saint-Germa~ France     89 RW         90
## 10 P. Dybala   Juventus     Argentina    88 RW         90
## # ... with 18,268 more rows
```

#Wingers are the best dribblers.

#Top 10 best defenders on fifa20

```
fifa20 %>% arrange(desc(defending)) %>% select(short_name, club, nationality, overall,
                                                team_position, defending)
```

```
## # A tibble: 18,278 x 6
##   short_name     club      nationality overall team_position defending
##   <chr>       <chr>      <chr>      <int> <chr>        <int>
## 1 V. van Dijk  Liverpool  Netherlands    90 LCB        90
## 2 G. Chiellini Juventus  Italy       89 LCB        90
## 3 K. Koulibaly Napoli     Senegal     89 LCB        89
## 4 D. God n    Inter      Uruguay     88 RCB        89
## 5 M. Hummels   Borussia Dortmund Germany    87 LCB        89
## 6 Piqu     FC Barcelona Spain      88 RCB        88
## 7 A. Laporte   Manchester City France     87 LCB        88
## 8 T. Alderweireld Tottenham Hotspur Belgium    87 RCB        88
## 9 N. Kant     Chelsea     France      89 RCM        87
## 10 Sergio Ramos Real Madrid  Spain      89 LCB        87
## # ... with 18,268 more rows
```

#Its surprising to see a midfielder who is very good at defending.

#Top 10 most physical players

```
fifa20 %>% arrange(desc(physic)) %>% select(short_name, club, nationality, overall,
                                                team_position, physic)
```

```
## # A tibble: 18,278 x 6
##   short_name     club      nationality overall team_position physic
##   <chr>       <chr>      <chr>      <int> <chr>        <int>
## 1 M. Marega    FC Porto    Mali        80 RS         90
## 2 Casemiro    Real Madrid  Brazil      87 CDM        89
## 3 M. Sissoko   Tottenham Hotspur France     81 SUB        89
## 4 Garrido     C diz CF     Spain      72 RDM        89
```

```

## 5 G. Kondogbia Valencia CF Central African~ 81 SUB 88
## 6 F. KessiÃ© Milan Ivory Coast 79 RCM 88
## 7 Wesley Aston Villa Brazil 79 ST 88
## 8 J. Kucka Parma Slovakia 77 RCM 88
## 9 M. Fellaini Shandong Luneng T~ Belgium 76 LDM 88
## 10 M. Bostwick Lincoln City England 68 RCB 88
## # ... with 18,268 more rows

```

#midfielders dominate the top10 list.

```

#####Goalkeeping stats
#Top 10 best goalkeepers on fifa20
fifa20 %>% filter(team_position=="GK") %>%
  arrange(desc(overall)) %>%
  select(short_name, club, nationality, overall)

```

```

## # A tibble: 662 x 4
##   short_name       club      nationality overall
##   <chr>          <chr>      <chr>        <int>
## 1 J. Oblak       AtlÃ©tico Madrid Slovenia     91
## 2 M. ter Stegen FC Barcelona Germany      90
## 3 Alisson        Liverpool    Brazil       89
## 4 De Gea         Manchester United Spain       89
## 5 Ederson        Manchester City Brazil       88
## 6 T. Courtois   Real Madrid    Belgium     88
## 7 "S. HandanoviÄ\u008d" Inter      Slovenia     88
## 8 M. Neuer        FC Bayern MÃ¼nchen Germany     88
## 9 H. Lloris       Tottenham Hotspur France      88
## 10 W. SzczÄsny Juventus    Poland      86
## # ... with 652 more rows

```

#EDA on gk stats like gk_diving, gk_handling, gk_kicking, gk_reflexes, gk_speed, gk_positioning

```

#Top 10 best divers on fifa20
fifa20 %>% arrange(desc(gk_diving)) %>% select(short_name, club, nationality, overall,
gk_diving)

```

```

## # A tibble: 18,278 x 5
##   short_name       club      nationality overall gk_diving
##   <chr>          <chr>      <chr>        <int>    <int>
## 1 De Gea         Manchester United Spain       89      90
## 2 K. Navas       Real Madrid    Costa Rica   87      90
## 3 G. Donnarumma Milan        Italy        85      90
## 4 H. Lloris      Tottenham Hotspur France     88      89
## 5 M. ter Stegen FC Barcelona  Germany     90      88
## 6 "S. HandanoviÄ\u008d" Inter      Slovenia     88      88
## 7 J. Oblak       AtlÃ©tico Madrid Slovenia     91      87
## 8 M. Neuer        FC Bayern MÃ¼nchen Germany     88      87
## 9 A. Lopes        Olympique Lyonnais Portugal   85      87
## 10 Ederson       Manchester City Brazil       88      86
## # ... with 18,268 more rows

```

```
#Top 10 best handlers on fifa20
fifa20 %>% arrange(desc(gk_handling)) %>% select(short_name, club, nationality, overall,
gk_handling)
```

```
## # A tibble: 18,278 x 5
##   short_name       club      nationality  overall gk_handling
##   <chr>        <chr>      <chr>        <int>     <int>
## 1 J. Oblak      Atl tico Madrid Slovenia      91      92
## 2 T. Courtois  Real Madrid    Belgium      88      89
## 3 M. Neuer      FC Bayern M nchen Germany      88      87
## 4 P. G l csyi RB Leipzig      Hungary      85      86
## 5 M. ter Stegen FC Barcelona  Germany      90      85
## 6 "S. Handanovi \u008d" Inter Slovenia      88      85
## 7 Y. Sommer     Borussia M nchengladbach Switzerland 84      85
## 8 Alisson       Liverpool      Brazil       89      84
## 9 De Gea        Manchester United Spain       89      84
## 10 S. Ruffier   AS Saint- tienne France       84      84
## # ... with 18,268 more rows
```

```
#Top 10 best kickers on fifa20
fifa20 %>% arrange(desc(gk_kicking)) %>% select(short_name, club, nationality, overall,
gk_kicking)
```

```
## # A tibble: 18,278 x 5
##   short_name       club      nationality  overall gk_kicking
##   <chr>        <chr>      <chr>        <int>     <int>
## 1 Ederson        Manchester City Brazil       88      93
## 2 M. Neuer       FC Bayern M nchen Germany      88      91
## 3 I. Khune       Kaizer Chiefs South Africa 78      90
## 4 M. ter Stegen FC Barcelona  Germany      90      88
## 5 J. Pickford   Everton       England      83      87
## 6 Kepa           Chelsea       Spain       84      86
## 7 A. Onana       Ajax          Cameroon    82      86
## 8 Alisson        Liverpool      Brazil      89      85
## 9 M. Pigliacelli Universitatea Craiova Italy       73      85
## 10 Herrer n    Athletic Club de Bilbao Spain      79      84
## # ... with 18,268 more rows
```

```
#Top 10 best reflexes on fifa20
fifa20 %>% arrange(desc(gk_reflexes)) %>% select(short_name, club, nationality, overall,
gk_reflexes)
```

```
## # A tibble: 18,278 x 5
##   short_name       club      nationality  overall gk_reflexes
##   <chr>        <chr>      <chr>        <int>     <int>
## 1 De Gea         Manchester United Spain       89      92
## 2 H. Lloris     Tottenham Hotspur France      88      91
## 3 M. ter Stegen FC Barcelona  Germany      90      90
## 4 K. Navas       Real Madrid      Costa Rica 87      90
## 5 G. Donnarumma Milan          Italy       85      90
## 6 J. Oblak       Atl tico Madrid Slovenia      91      89
## 7 Alisson        Liverpool      Brazil      89      89
```

```

## 8 "S. Handanovi\u00e1" Inter Slovenia 88 89
## 9 A. Lopes Olympique Lyonnais Portugal 85 89
## 10 R. B\u00e1rki Borussia Dortmund Switzerland 85 89
## # ... with 18,268 more rows

```

```
#Top 10 best speed on fifa20
fifa20 %>% arrange(desc(gk_speed)) %>% select(short_name, club, nationality, overall,
gk_speed)
```

```

## # A tibble: 18,278 x 5
##   short_name   club      nationality overall gk_speed
##   <chr>        <chr>      <chr>     <int>   <int>
## 1 Jordi Masip Real Valladolid CF Spain      81    65
## 2 R. Zentner 1. FSV Mainz 05 Germany    73    65
## 3 A. Lopes Olympique Lyonnais Portugal 85    64
## 4 P. Gazzaniga Tottenham Hotspur Argentina 75    64
## 5 Ederson Manchester City Brazil      88    63
## 6 H. Lloris Tottenham Hotspur France      88    63
## 7 A. Onana Ajax Cameroon    82    63
## 8 Alberto Rayo Vallecano Spain      72    63
## 9 A. Maisonnial Paris FC France      64    63
## 10 J. Rodr\u00edguez Puebla FC Mexico      60    63
## # ... with 18,268 more rows

```

```
#Top 10 best positioning on fifa20
fifa20 %>% arrange(desc(gk_positioning)) %>% select(short_name, club, nationality, overall,
gk_positioning)
```

```

## # A tibble: 18,278 x 5
##   short_name   club      nationality overall gk_positioning
##   <chr>        <chr>      <chr>     <int>   <int>
## 1 G. Buffon Juventus Italy      83    91
## 2 J. Oblak Atl\u00e1tico Madrid Slovenia 91    90
## 3 Alisson Liverpool Brazil    89    90
## 4 "S. Handanovi\u00e1" Inter Slovenia 88    89
## 5 M. ter Stegen FC Barcelona Germany 90    88
## 6 Ederson Manchester City Brazil    88    86
## 7 W. Szcz\u0144sny Juventus Poland    86    86
## 8 Casillas FC Porto Spain      83    86
## 9 R. Jarstein Hertha BSC Norway    83    86
## 10 De Gea Manchester United Spain    89    85
## # ... with 18,268 more rows

```

```
####ATTACKING STATS
#stats like attacking_crossing, attacking_finishing and attacking_heading_accuracy
#Top 10 best crossers on fifa20
fifa20 %>% arrange(desc(attacking_crossing)) %>%
  select(short_name, club, nationality, overall,attacking_crossing)
```

```

## # A tibble: 18,278 x 5
##   short_name   club      nationality overall attacking_crossing
##   <chr>        <chr>      <chr>     <int>   <int>

```

```

## 1 K. De Bruyne Manchester City Belgium 91 93
## 2 J. Kimmich FC Bayern MÃ¼nchen Germany 86 91
## 3 Quaresma BeÃ½iktaÃ½ JK Portugal 81 91
## 4 Pedro LeÃ³n SD Eibar Spain 80 91
## 5 J. RodrÃºguez Real Madrid Colombia 85 90
## 6 A. Kolarov Roma Serbia 82 90
## 7 A. Robertson Liverpool Scotland 85 89
## 8 Marcelo Real Madrid Brazil 85 89
## 9 Alex Telles FC Porto Brazil 84 89
## 10 J. Ward-Prowse Southampton England 79 89
## # ... with 18,268 more rows

```

```

#Top 10 best finishers on fifa20
fifa20 %>% arrange(desc(attacking_finishing)) %>%
  select(short_name, club, nationality, overall, attacking_finishing)

```

```

## # A tibble: 18,278 x 5
##   short_name     club      nationality overall attacking_finishing
##   <chr>       <chr>      <chr>      <int>            <int>
## 1 L. Messi    FC Barcelona Argentina 94 95
## 2 Cristiano Ronaldo Juventus Portugal 93 94
## 3 H. Kane    Tottenham Hotspur England 89 94
## 4 S. AgÃ¼ero Manchester City Argentina 89 93
## 5 L. SuÃ¡rez FC Barcelona Uruguay 89 91
## 6 M. Salah    Liverpool Egypt 90 90
## 7 K. MbappÃ© Paris Saint-Germain France 89 89
## 8 A. Griezmann FC Barcelona France 89 89
## 9 P. Aubameyang Arsenal Gabon 88 89
## 10 M. Icardi Inter Argentina 85 89
## # ... with 18,268 more rows

```

#ARE FINISHING AND SHOOTING CORRELATED?

```

#Top 10 players most likely to score a goal via a header on fifa20
fifa20 %>% arrange(desc(attacking_heading_accuracy)) %>%
  select(short_name, club, nationality, overall, attacking_heading_accuracy)

```

```

## # A tibble: 18,278 x 5
##   short_name     club      nationality overall attacking_heading_accuracy
##   <chr>       <chr>      <chr>      <int>            <int>
## 1 B. Dost    Sporting CP Netherlands 82 93
## 2 L. de Jong Sevilla FC Netherlands 82 93
## 3 L. Pavoletti Cagliari Italy 78 93
## 4 Sergio Ramos Real Madrid Spain 89 92
## 5 Aduriz    Athletic Club de Bi~ Spain 82 92
## 6 M. Fellaini Shandong Luneng Tai~ Belgium 76 92
## 7 O. Giroud Chelsea France 82 91
## 8 M. Smith    Millwall England 68 91
## 9 M. MandÅ¾uki~ Juventus Croatia 84 90
## 10 RaÃºl GarcÃ¡a Athletic Club de Bi~ Spain 80 90
## # ... with 18,268 more rows

```

```
#### MOVEMENT STATS
#stats like movement_acceleration and movement_balance
#Top 10 players with best acceleration
fifa20 %>% arrange(desc(movement_acceleration)) %>%
  select(short_name, club, nationality, overall,movement_acceleration)
```

```
## # A tibble: 18,278 x 5
##   short_name     club           nationality overall movement_accelerati~
##   <chr>       <chr>           <chr>      <int>          <int>
## 1 Adama Traoré  Wolverhampton Wanderers Spain        74          97
## 2 K. Mbappé    Paris Saint-Germain France       89          96
## 3 R. Sterling  Manchester City   England       88          96
## 4 S. Mane      Liverpool        Senegal       88          95
## 5 Douglas Costa Juventus        Brazil        84          95
## 6 Lucas Moura  Tottenham Hotspur  Brazil        83          95
## 7 Gelson Martins AS Monaco     Portugal       82          95
## 8 I. Sarr      Watford         Senegal       78          95
## 9 A. Musa      Al Nassr        Nigeria       73          95
## 10 K. Nagai     FC Tokyo        Japan        69          95
## # ... with 18,268 more rows
```

```
#Top 10 players with best balance
fifa20 %>% arrange(desc(movement_balance)) %>%
  select(short_name, club, nationality, overall,movement_balance)
```

```
## # A tibble: 18,278 x 5
##   short_name     club           nationality overall movement_balance
##   <chr>       <chr>           <chr>      <int>          <int>
## 1 E. Oztumer   Charlton Athletic England       69          97
## 2 R. Fraser    Bournemouth     Scotland      81          96
## 3 "T. Italo"
## 4 L. Messi     FC Barcelona    Argentina     94          95
## 5 Bernard      Everton         Brazil        80          95
## 6 S. Kaneko    Shimizu S-Pulse Japan        70          95
## 7 Aridai       RCD Mallorca    Spain        70          95
## 8 S. Horvath   SG Dynamo Dresden Austria      67          95
## 9 Isi Ros     AD Alcorcón     Spain        65          95
## 10 K. Holzweiler Viktoria Köln Germany      64          95
## # ... with 18,268 more rows
```

#ARE BALANCE AND DRIBBLING CORRELATED?

```
#### POWER STATS
#stats like power_shot_power, power_jumping, power_strength
#Top 10 players with best shot power on fifa20
fifa20 %>% arrange(desc(power_shot_power)) %>%
  select(short_name, club, nationality, overall,power_shot_power)
```

```
## # A tibble: 18,278 x 5
##   short_name     club           nationality overall power_shot_power
##   <chr>       <chr>           <chr>      <int>          <int>
```

```

##      <chr>          <chr>          <chr>      <int>      <int>
## 1 Cristiano Ronaldo Juventus Portugal     93       95
## 2 A. Kolarov        Roma    Serbia      82       95
## 3 Hulk             Shanghai SIPG FC Brazil      80       94
## 4 G. Bale           Real Madrid   Wales      85       92
## 5 K. De Bruyne     Manchester City Belgium     91       91
## 6 H. Kane           Tottenham Hotspur England     89       90
## 7 P. Pogba          Manchester United France    88       90
## 8 Z. Ibrahimović LA Galaxy   Sweden      85       90
## 9 F. Quagliarella Sampdoria   Italy      83       90
## 10 G. Xhaka          Arsenal    Switzerland 81       90
## # ... with 18,268 more rows

```

#ARE SHOOTING AND SHOT POWER CORRELATED?

```

#Top 10 jumpers on fifa20
fifa20 %>% arrange(desc(power_jumping)) %>%
  select(short_name, club, nationality, overall,power_jumping)

```

```

## # A tibble: 18,278 x 5
##   short_name      club      nationality  overall power_jumping
##   <chr>          <chr>          <chr>      <int>      <int>
## 1 Cristiano Ronaldo Juventus Portugal     93       95
## 2 E. Sabbi         Hobro IK      United States 68       95
## 3 T. Hasegawa     Kawasaki Frontale Japan      66       95
## 4 M. Icardi        Inter        Argentina    85       94
## 5 O. Lewicki       Malmø FF    Sweden      72       94
## 6 M. Barbieri     Rosario Central Argentina    71       94
## 7 T. Sugimoto      Matsumoto Yamaga Japan      61       94
## 8 Sergio Ramos   Real Madrid   Spain      89       93
## 9 E. Boateng       Dalian YiFang FC Ghana      77       93
## 10 E. Dennis       Club Brugge KV Nigeria    75       93
## # ... with 18,268 more rows

```

#ARE JUMPING AND ATTACKING_HEADING_ACCURACY CORRELATED?

```

#Top 10 players with best power strength on fifa20
fifa20 %>% arrange(desc(power_strength)) %>%
  select(short_name, club, nationality, overall,power_strength)

```

```

## # A tibble: 18,278 x 5
##   short_name      club      nationality  overall power_strength
##   <chr>          <chr>          <chr>      <int>      <int>
## 1 A. Akinfenwa  Wycombe Wanderers England     65       97
## 2 K. Koulibaly Napoli      Senegal      89       95
## 3 R. Lukaku      Inter        Belgium     85       95
## 4 Wesley         Aston Villa   Brazil      79       95
## 5 N. Sále        FC Bayern München Germany    85       94
## 6 S. Coates      Sporting CP    Uruguay    82       94
## 7 D. Zapata      Atalanta     Colombia   82       94
## 8 K. Waston      FC Cincinnati Costa Rica 73       94
## 9 A. Cerri        Cagliari     Italy      72       94
## 10 O. Dularăș Standard de Liège Belgium    71       94
## # ... with 18,268 more rows

```

```
#### MENTALITY STATS
#stats like mentality_positioning, mentality_penalties and mentality_vision
#Top 10 players with best position sense on fifa20
fifa20 %>% arrange(desc(mentality_positioning)) %>%
  select(short_name, club, nationality, team_position, overall, mentality_positioning)
```

```
## # A tibble: 18,278 x 6
##   short_name     club      nationality team_position overall mentality_positi~
##   <chr>       <chr>      <chr>        <chr>      <int>      <int>
## 1 Cristiano Ro~ Juventus    Portugal     LW          93        95
## 2 L. Messi       FC Barcelo~ Argentina   RW          94        94
## 3 H. Kane        Tottenham ~ England    ST          89        93
## 4 S. AgÃ¼ero    Manchester~ Argentina   ST          89        93
## 5 E. Cavani     Paris Sain~ Uruguay    ST          88        93
## 6 M. Salah       Liverpool    Egypt       RW          90        92
## 7 L. SuÃ¡rez     FC Barcelo~ Uruguay    ST          89        92
## 8 T. MÃ¶ller     FC Bayern ~ Germany   SUB         86        92
## 9 M. Icardi      Inter       Argentina  RES         85        92
## 10 R. Lewandows~ FC Bayern ~ Poland    ST          89        91
## # ... with 18,268 more rows
```

```
#Top 10 best penalty takes on fifa20
fifa20 %>% arrange(desc(mentality_penalties)) %>%
  select(short_name, club, nationality, overall, mentality_penalties)
```

```
## # A tibble: 18,278 x 5
##   short_name     club      nationality overall mentality_penalties
##   <chr>       <chr>      <chr>      <int>      <int>
## 1 M. Kruse      FenerbahÃ§e SK    Germany    83        92
## 2 Fabinho       Liverpool    Brazil      85        91
## 3 S. Haller     West Ham United France    83        91
## 4 M. Balotelli  Brescia     Italy       82        91
## 5 L. MilivojeviÄ‡ Crystal Palace Serbia    81        91
## 6 Neymar Jr    Paris Saint-Germain Brazil    92        90
## 7 H. Kane       Tottenham Hotspur England   89        90
## 8 M. Reus        Borussia Dortmund Germany   88        90
## 9 A. KramariÄ‡   TSG 1899 Hoffenheim Croatia   83        90
## 10 B. Dost       Sporting CP    Netherlands 82        90
## # ... with 18,268 more rows
```

#ARE ATTACKING FINISHING AND PENALTIES CORRELATED?

```
#Top 10 players with best vision on fifa20
fifa20 %>% arrange(desc(mentality_vision)) %>%
  select(short_name, club, nationality, overall, mentality_vision)
```

```
## # A tibble: 18,278 x 5
##   short_name     club      nationality overall mentality_vision
##   <chr>       <chr>      <chr>      <int>      <int>
## 1 L. Messi      FC Barcelona Argentina  94        94
## 2 K. De Bruyne   Manchester City Belgium   91        94
## 3 C. Eriksen    Tottenham Hotspur Denmark  88        92
```

```

## 4 L. Modrić Real Madrid Croatia 90 91
## 5 David Silva Manchester City Spain 88 91
## 6 Neymar Jr Paris Saint-Germain Brazil 92 90
## 7 M. Čilic Arsenal Germany 84 90
## 8 E. Hazard Real Madrid Belgium 91 89
## 9 T. Kroos Real Madrid Germany 88 89
## 10 L. Insigne Napoli Italy 87 89
## # ... with 18,268 more rows

```

#ARE VISION AND PASSING CORRELATED?

```

#####DEFENDING STATS
#stats like defending_marking, defending_standing_tackle and
#defending_sliding_tackle
#Top 10 best markers on fifa20
fifa20 %>% arrange(desc(defending_marking)) %>%
  select(short_name, club, nationality, team_position, overall, defending_marking)

```

```

## # A tibble: 18,278 x 6
##   short_name   club     nationality team_position overall defending_marking
##   <chr>        <chr>    <chr>      <chr>       <int>           <int>
## 1 G. Chiellini Juventus Italy       LCB          89            94
## 2 M. Čokrić Inter    Slovakia   LCB          86            92
## 3 V. van Dijk Liverpool Netherlands LCB          90            91
## 4 K. Koulibaly Napoli    Senegal   LCB          89            91
## 5 N. Kanté    Chelsea   France    RCM          89            90
## 6 Sergio Busquets FC Barcelona Spain      CDM          89            90
## 7 D. Godín    Inter    Uruguay   RCB          88            90
## 8 M. Hummels Borussia Dortmund Germany   LCB          87            90
## 9 T. Alderweirel Tottenham Hotspur Belgium   RCB          87            90
## 10 L. Bonucci Juventus Italy       RCB          86            90
## # ... with 18,268 more rows

```

#Top 10 best tacklers on fifa20

```

fifa20 %>% arrange(desc(defending_standing_tackle)) %>%
  select(short_name, club, nationality, team_position, overall, defending_standing_tackle)

```

```

## # A tibble: 18,278 x 6
##   short_name   club     nationality team_position overall defending_standing_tackle
##   <chr>        <chr>    <chr>      <chr>       <int>           <int>
## 1 V. van Dijk Liverpool Netherlands LCB          90            92
## 2 N. Kanté    Chelsea   France    RCM          89            91
## 3 G. Chiellini Juventus Italy       LCB          89            91
## 4 K. Koulibaly Napoli    Senegal   LCB          89            90
## 5 T. Alderweirel Tottenham Hotspur Belgium   RCB          87            90
## 6 I. Gueye    Paris Saint-Germain Senegal   SUB          83            90
## 7 D. Godín    Inter    Uruguay   RCB          88            89
## 8 A. Laporte Manchester United France   LCB          87            89
## 9 M. Hummels Borussia Dortmund Germany   LCB          87            89
## 10 N. Sául    FC Bayern Munich Germany   RCB          85            89
## # ... with 18,268 more rows

```

```

#Top 10 best sliding tacklers on fifa20
fifa20 %>% arrange(desc(defending_sliding_tackle)) %>%
  select(short_name, club, nationality, team_position, overall, defending_sliding_tackle)

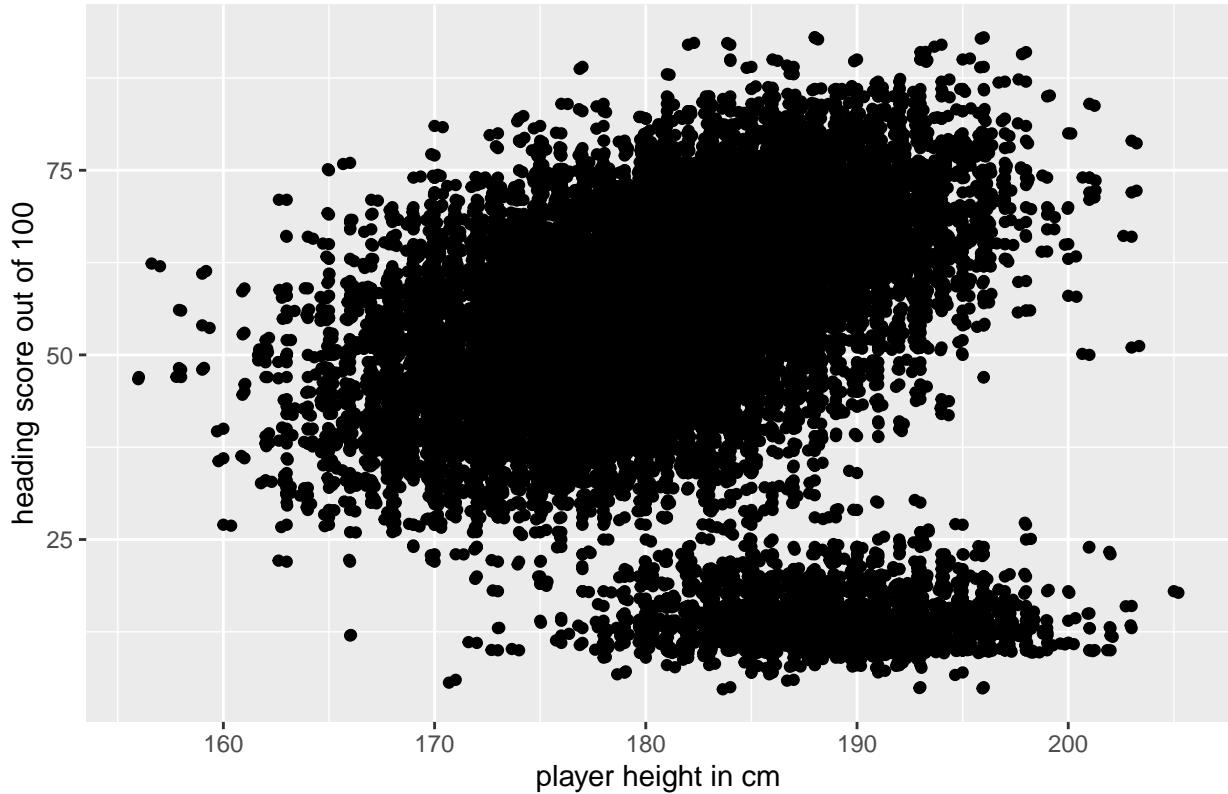
## # A tibble: 18,278 x 6
##   short_name    club      nationality team_position overall defending_sliding_~
##   <chr>        <chr>      <chr>       <chr>      <int>            <int>
## 1 Sergio Ramos Real Madr~ Spain        LCB         89             90
## 2 R. Nainggol~ Cagliari   Belgium     CAM         83             90
## 3 G. Chiellini Juventus  Italy       LCB         89             89
## 4 K. Manolas Napoli    Greece     RCB         85             89
## 5 A. Laporte Manchester~ France     LCB         87             88
## 6 J. Vertongh~ Tottenham~ Belgium    LCB         87             88
## 7 A. Vidal FC Barcel~ Chile      SUB         84             88
## 8 T. Delaney Borussia ~ Denmark   SUB         82             88
## 9 K. Koulibaly Napoli    Senegal    LCB         89             87
## 10 PiquÃ© FC Barcel~ Spain      RCB         88             87
## # ... with 18,268 more rows

#####ABSTRACT HYPOTHESIS
#tall, short and strong players are
#statistically good at heading, dribbling and tackling respectively

#Tall players and heading
#scatter plot of height vs heading
fifa20 %>% ggplot(aes(height_cm, attacking_heading_accuracy))+geom_point()+
  geom_jitter()+labs(x="player height in cm", y="heading score out of 100",
  title = "Plot of height vs heading score")

```

Plot of height vs heading score



```
#THIS IS A VERY INTERESTING FIND. WHY?  
cor(fifa20$height_cm, fifa20$attacking_heading_accuracy,  
    method = "spearman", use = "complete.obs")
```

```
## [1] 0.2054838
```

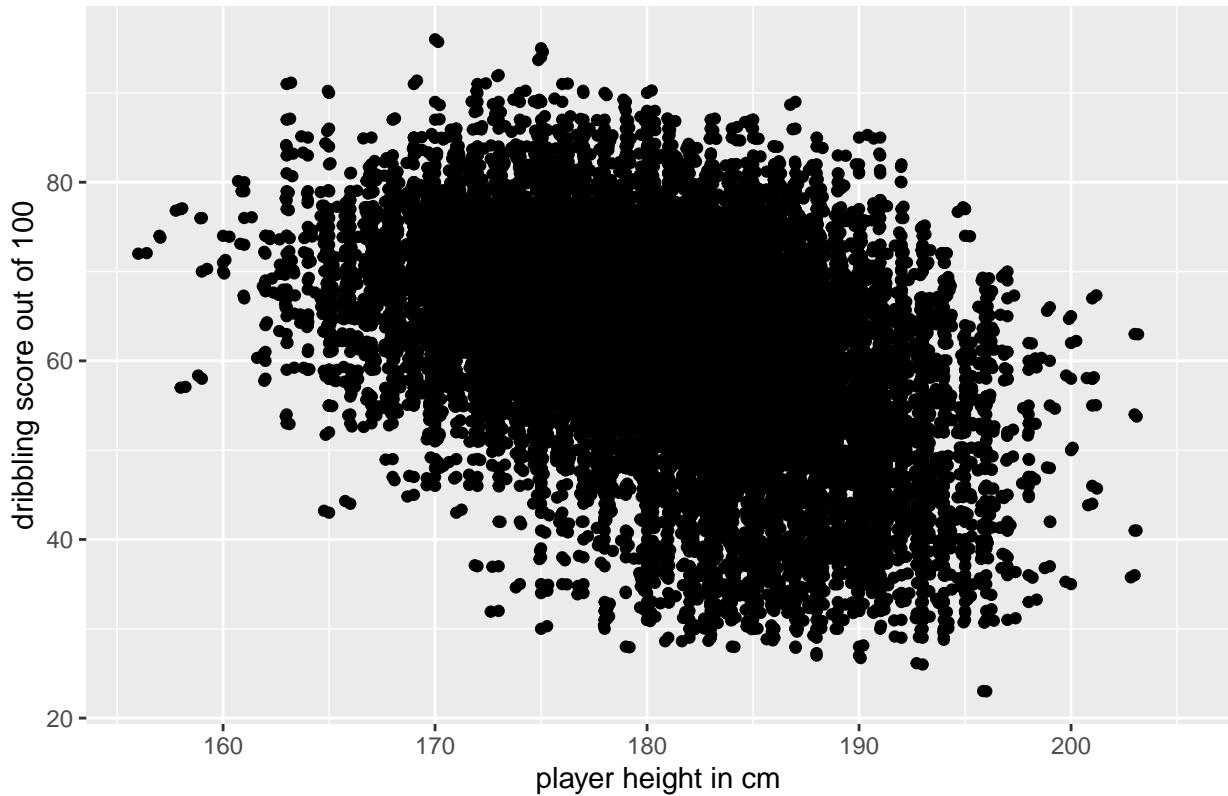
```
#There exists a weak positive correlation between height and heading.
```

```
#Short players and dribbling  
#scatter plot of height vs dribbling  
fifa20 %>% ggplot(aes(height_cm, dribbling)) + geom_point() +  
  geom_jitter() + labs(x = "player height in cm", y = "dribbling score out of 100",  
    title = "Plot of height vs dribbling score")
```

```
## Warning: Removed 2036 rows containing missing values (geom_point).
```

```
## Warning: Removed 2036 rows containing missing values (geom_point).
```

Plot of height vs dribbling score



```
cor(fifa20$height_cm, fifa20$dribbling,
    method = "spearman", use = "complete.obs")
```

```
## [1] -0.3946403
```

#There is a weak negative correlation b/w height and dribbling ie short players are good at dribbling.

#Dribbling and movement attributes:

```
cor(fifa20$dribbling, fifa20$movement_agility,
    method = "pearson", use = "complete.obs")
```

```
## [1] 0.7241454
```

#Dribbling and movement_agility are highly +vely correlated

```
cor(fifa20$dribbling, fifa20$movement_balance,
    method = "pearson", use = "complete.obs")
```

```
## [1] 0.5581589
```

#Dribbling and movement_balance are moderately +vely correlated

#correlation matrix for dribbling and movement attributes:

```
library(corrplot)
```

```

## Warning: package 'corrplot' was built under R version 3.6.3

## corrplot 0.84 loaded

library(rquery)

## Warning: package 'rquery' was built under R version 3.6.3

## Loading required package: wprapr

## Warning: package 'wprapr' was built under R version 3.6.3

## 
## Attaching package: 'wprapr'

## The following object is masked from 'package:modelr':
## 
##     qae

## The following object is masked from 'package:data.table':
## 
##     :=

## The following object is masked from 'package:tidyর':
## 
##     unpack

## The following object is masked from 'package:dplyr':
## 
##     coalesce

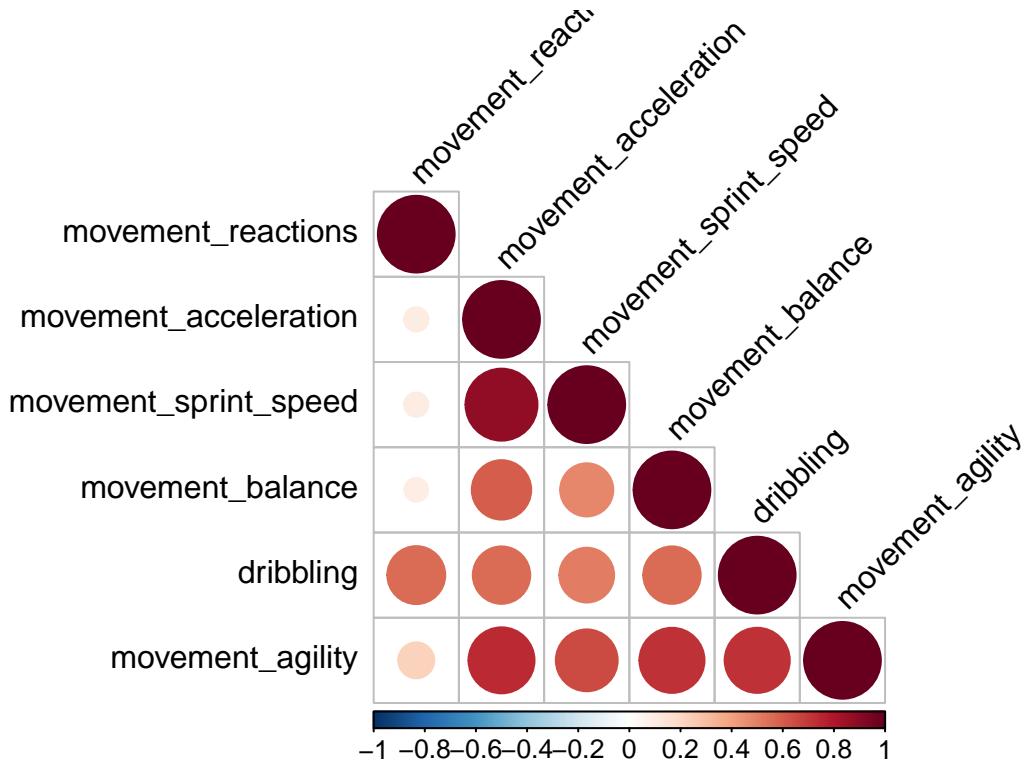
## 
## Attaching package: 'rquery'

## The following object is masked from 'package:tidyর':
## 
##     expand_grid

## The following object is masked from 'package:ggplot2':
## 
##     arrow

drib_move <- fifa20 %>% select(dribbling,movement_acceleration,
                                    movement_sprint_speed, movement_agility,
                                    movement_reactions, movement_balance)
source("http://www.sthda.com/upload/rquery_cormat.r")
require("corrplot")
rquery.cormat(drib_move)

```



```

## $r
##           movement_reactions movement_acceleration
## movement_reactions          1
## movement_acceleration      0.1          1
## movement_sprint_speed      0.1          0.88
## movement_balance           0.096         0.6
## dribbling                   0.57          0.56
## movement_agility           0.22          0.74
##           movement_sprint_speed movement_balance dribbling
## movement_reactions
## movement_acceleration
## movement_sprint_speed          1
## movement_balance              0.48          1
## dribbling                     0.51          0.56          1
## movement_agility              0.65          0.72          0.72
##           movement_agility
## movement_reactions
## movement_acceleration
## movement_sprint_speed
## movement_balance
## dribbling
## movement_agility             1
## 
## $p
##           movement_reactions movement_acceleration
## movement_reactions          0

```

```

## movement_acceleration      2.8e-151      0
## movement_sprint_speed      1.7e-155      0
## movement_balance           3.1e-97       0
## dribbling                  0             0
## movement_agility          5.89999999999934e-312 0
##                               movement_sprint_speed movement_balance dribbling
## movement_reactions
## movement_acceleration
## movement_sprint_speed      0
## movement_balance           0             0
## dribbling                  0             0             0
## movement_agility          0             0             0
##                               movement_agility
## movement_reactions
## movement_acceleration
## movement_sprint_speed
## movement_balance
## dribbling
## movement_agility          0
##
## $sym
##                               movement_reactions movement_acceleration
## movement_reactions      1
## movement_acceleration
## movement_sprint_speed      +
## movement_balance           .
## dribbling                  .
## movement_agility          ,
##                               movement_sprint_speed movement_balance dribbling
## movement_reactions
## movement_acceleration
## movement_sprint_speed 1
## movement_balance          .
## dribbling                  .
## movement_agility          ,
##                               movement_agility
## movement_reactions
## movement_acceleration
## movement_sprint_speed
## movement_balance
## dribbling
## movement_agility          1
## attr(,"legend")
## [1] 0 ' ' 0.3 '.' 0.6 ',' 0.8 '+' 0.9 '*' 0.95 'B' 1

```

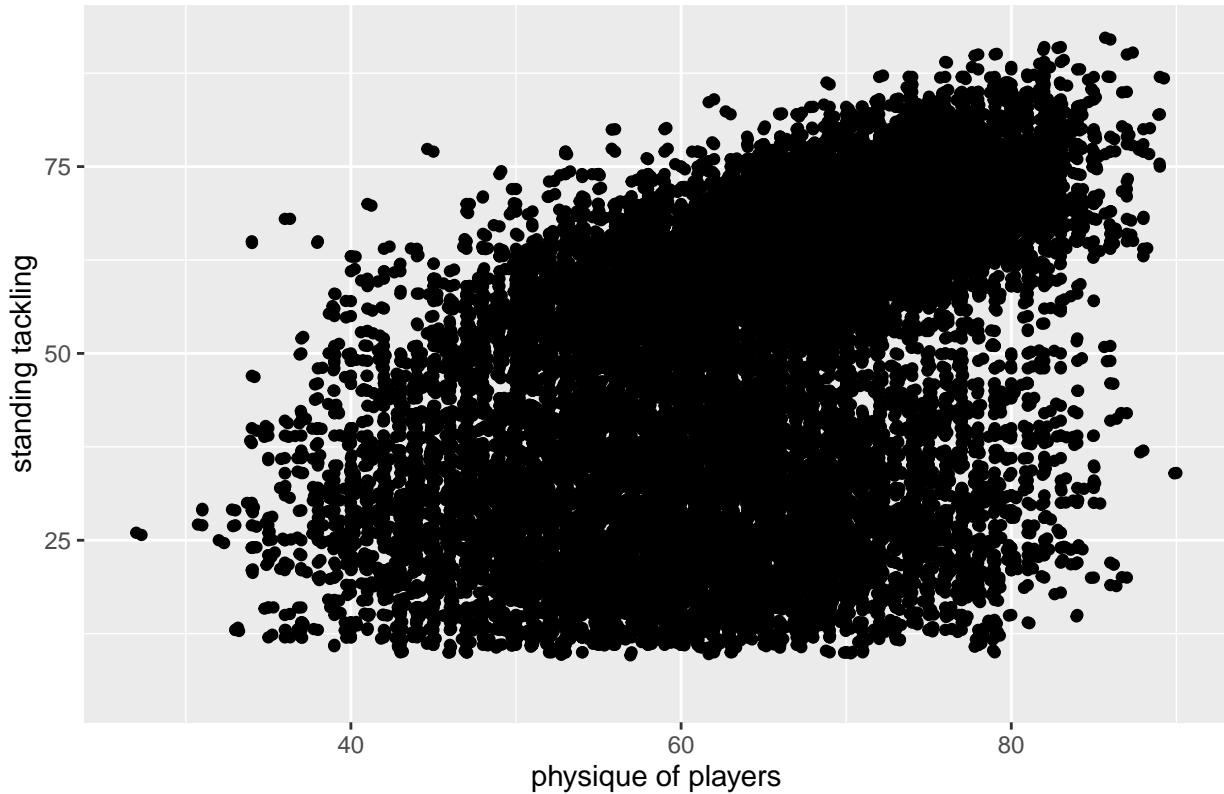
*#we can see from the corrplot that dribbling is highly +vely
#correlated with the movement attributes.*

*#Strong players and tackling(both standing and sliding)
#scatter plot of physique vs standing tackling*
fifa20 %>% ggplot(aes(physic, defending_standing_tackle))+geom_point()+
 geom_jitter()+labs(x="physique of players", y="standing tackling",
 title = "physique vs standing tackling")

```
## Warning: Removed 2036 rows containing missing values (geom_point).
```

```
## Warning: Removed 2036 rows containing missing values (geom_point).
```

physique vs standing tackling



```
cor(fifa20$physic, fifa20$defending_standing_tackle,  
    method = "pearson", use = "complete.obs")
```

```
## [1] 0.4896467
```

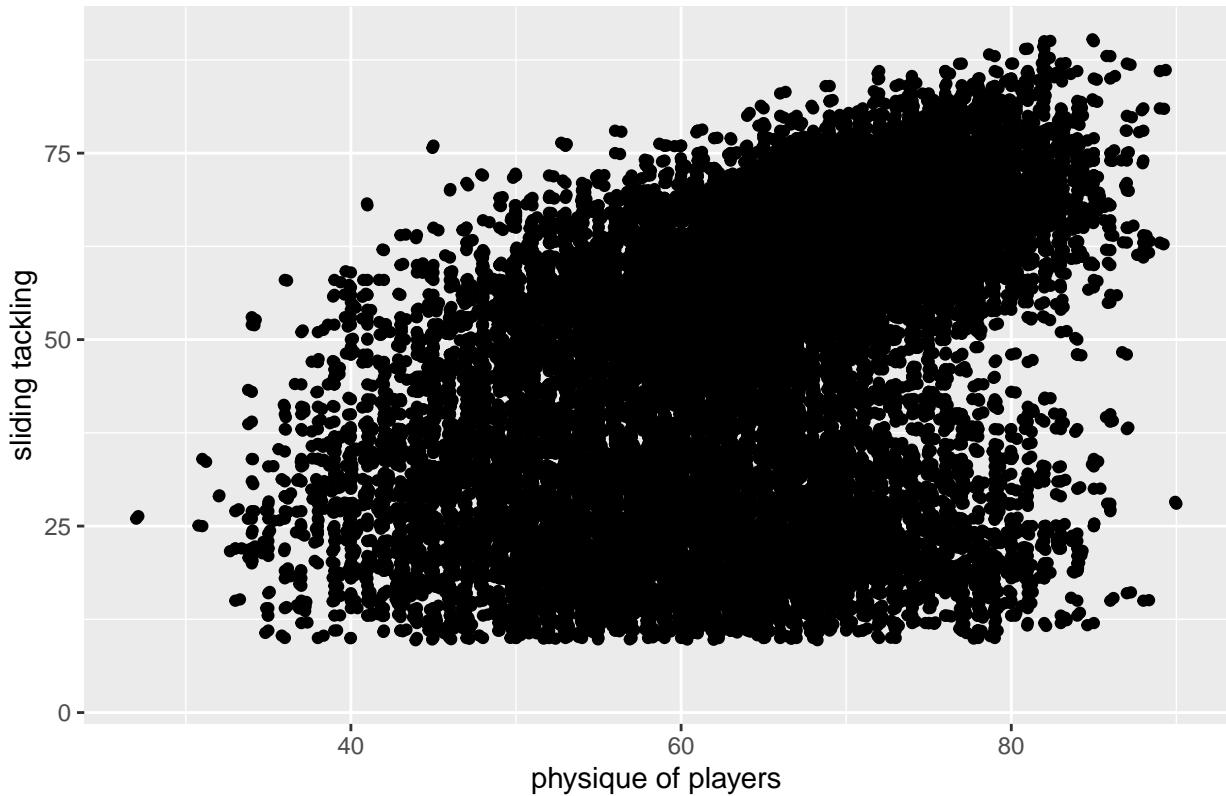
```
#moderate +ve correlation
```

```
#scatter plot of physique vs sliding tackling  
fifa20 %>% ggplot(aes(physic, defending_sliding_tackle))+geom_point() +  
  geom_jitter() + labs(x="physique of players", y="sliding tackling",  
    title = "physique vs sliding tackling")
```

```
## Warning: Removed 2036 rows containing missing values (geom_point).
```

```
## Warning: Removed 2036 rows containing missing values (geom_point).
```

physique vs sliding tackling



```
cor(fifa20$physic, fifa20$defending_sliding_tackle,
  method = "pearson", use = "complete.obs")  
  
## [1] 0.4515019  
  
#moderate +ve correlation  
  
####TEAM STATS:  
#Top 10 teams with highest average player overall  
fifa20 %>% filter(!club %in% c("Uruguay", "Colombia", "Mexico", "Netherlands")) %>%
  group_by(club) %>%
  summarise(avg_player_ovr = mean(overall)) %>%
  arrange(desc(avg_player_ovr))  
  
## # A tibble: 694 x 2  
##   club           avg_player_ovr  
##   <chr>             <dbl>  
## 1 1 FC Bayern MÃ¼nchen      81.3  
## 2 Real Madrid            80.1  
## 3 Juventus                80.1  
## 4 FC Barcelona             78.4  
## 5 Bayer 04 Leverkusen     77.3  
## 6 Chelsea                  77.1
```

```

## 7 Manchester City          77
## 8 Napoli                   76.9
## 9 Manchester United        76.8
## 10 Tottenham Hotspur       76.5
## # ... with 684 more rows

#These are 10 best teams or the top 10 teams with the highest average player
#overall out of 100.

#Top 10 teams with highest average player overall in the starting 11
#exclude subs and reserves
best_teams <- fifa20 %>% filter(!club %in% c("Uruguay", "Colombia", "Mexico", "Netherlands")) %>%
  filter(!team_position %in% c("SUB", "RES")) %>%
  group_by(club) %>%
  summarise(avg_player_ovr = mean(overall)) %>%
  arrange(desc(avg_player_ovr))
best_teams

## # A tibble: 690 x 2
##   club           avg_player_ovr
##   <chr>          <dbl>
## 1 FC Barcelona    87.6
## 2 Real Madrid     86.7
## 3 Manchester City 86.4
## 4 Liverpool       85.7
## 5 FC Bayern MÃ¼nchen 85.5
## 6 Juventus        85.3
## 7 Paris Saint-Germain 85.2
## 8 Tottenham Hotspur 84.1
## 9 Borussia Dortmund 83.6
## 10 AtlÃ©tico Madrid 83.2
## # ... with 680 more rows

#view(best_teams)
#These are the top 10 teams with the highest avg. starting 11 player overall.

#Worst teams:
worst_teams <- fifa20 %>% filter(!club %in% c("Uruguay", "Colombia", "Mexico", "Netherlands", "India")) %>%
  filter(!team_position %in% c("SUB", "RES")) %>%
  group_by(club) %>%
  summarise(avg_player_ovr = mean(overall)) %>%
  arrange(avg_player_ovr)
worst_teams

## # A tibble: 689 x 2
##   club           avg_player_ovr
##   <chr>          <dbl>
## 1 Bolton Wanderers 54.5
## 2 UCD AFC          54.7
## 3 Waterford FC     56.4
## 4 Finn Harps        56.9
## 5 Bury              59.1
## 6 Leyton Orient     59.4

```

```

## 7 Sligo Rovers           59.5
## 8 Derry City            59.7
## 9 Macclesfield Town     59.9
## 10 Bohemian FC          60.1
## # ... with 679 more rows

#View(worst_teams)

#Teams with most valuable squads:
fifa20 %>% group_by(club)%>%
  summarise(squad_value_eur = sum(value_eur))%>%
  arrange(desc(squad_value_eur))

## # A tibble: 698 x 2
##   club           squad_value_eur
##   <chr>             <int>
## 1 Real Madrid      897850000
## 2 FC Barcelona     869300000
## 3 Manchester City 845745000
## 4 Juventus        735475000
## 5 Liverpool        693265000
## 6 FC Bayern MÃ¼nchen 688775000
## 7 Paris Saint-Germain 687550000
## 8 Tottenham Hotspur 649850000
## 9 AtlÃ©tico Madrid 590375000
## 10 Borussia Dortmund 532325000
## # ... with 688 more rows

#Teams with highest weekly wage:
fifa20 %>% group_by(club)%>%
  summarise(wage_value_eur = sum(wage_eur))%>%
  arrange(desc(wage_value_eur))

## # A tibble: 698 x 2
##   club           wage_value_eur
##   <chr>             <int>
## 1 Real Madrid      5354000
## 2 FC Barcelona     4950000
## 3 Manchester City 3984000
## 4 Juventus        3750000
## 5 Manchester United 2874000
## 6 Chelsea          2806000
## 7 Liverpool         2667000
## 8 Tottenham Hotspur 2603000
## 9 FC Bayern MÃ¼nchen 2516000
## 10 Paris Saint-Germain 2396000
## # ... with 688 more rows

#####BEST 11 ON FIFA20
unique(fifa20$team_position)

## [1] "RW"   "LW"   "CAM"  "GK"   "RCM"  "LCB"  "ST"   "CDM"  "LDM"  "RM"   "RCB"  "LCM"

```

```
## [13] "LM"   "CF"   "SUB"  "LB"   "LS"   "RB"   "RDM"  "RES"  "RAM"  "RS"   "RF"   "CM"
## [25] "CB"   "LF"   "LAM"  ""     "RWB"  "LWB"
```

```
#Assuming a 4231 formation, we come up with the best team on fifa20 or the
#fifa world 11
#striker:
#positions to be considered: ST, CF, LS, RS, LF, RF
fifa20 %>% filter(team_position %in% c("ST", "CF", "LS", "RS", "LF", "RF")) %>%
  arrange(desc(overall)) %>%
  select(short_name, team_position, overall)
```

```
## # A tibble: 900 x 3
##   short_name      team_position overall
##   <chr>            <chr>          <int>
## 1 H. Kane           ST              89
## 2 S. Agüero         ST              89
## 3 L. Suárez         ST              89
## 4 R. Lewandowski   ST              89
## 5 E. Cavani         ST              88
## 6 K. Benzema        CF              87
## 7 L. Insigne        LS              87
## 8 Roberto Firmino   CF              86
## 9 C. Immobile       ST              86
## 10 A. Lacazette     ST              86
## # ... with 890 more rows
```

```
#can pick either of Kane, Suarez, Aguero or Lewandowski.
```

```
#Wingers: 2 wingers to be picked, one on the left and one on the right.
#positions to be considered: RW, LW
fifa20 %>% filter(team_position %in% c("LW", "RW")) %>%
  arrange(desc(overall)) %>%
  select(short_name, team_position, overall)
```

```
## # A tibble: 323 x 3
##   short_name      team_position overall
##   <chr>            <chr>          <int>
## 1 L. Messi          RW              94
## 2 Cristiano Ronaldo LW              93
## 3 E. Hazard          LW              91
## 4 M. Salah           RW              90
## 5 K. Mbappé          RW              89
## 6 A. Griezmann       LW              89
## 7 P. Dybala          RW              88
## 8 R. Sterling         LW              88
## 9 S. Mané             LW              88
## 10 Bernardo Silva    RW              87
## # ... with 313 more rows
```

```
#we pick Messi on the right and Ronaldo on the left
```

```
#one AM, positions: CAM, LAM, RAM
```

```
fifa20 %>% filter(team_position %in% c("CAM", "LAM", "RAM")) %>%
  arrange(desc(overall)) %>%
  select(short_name, team_position, overall)
```

```
## # A tibble: 357 x 3
##   short_name     team_position overall
##   <chr>           <chr>          <int>
## 1 Neymar Jr      CAM             92
## 2 M. Reus         CAM             88
## 3 H. Ziyech       RAM             85
## 4 A. Gómez        CAM             85
## 5 D. Alli         CAM             84
## 6 M. Hamed        CAM             84
## 7 M. Ázil         CAM             84
## 8 R. Nainggolan   CAM             83
## 9 M. Kruse        CAM             83
## 10 G. Sigurðsson  CAM             83
## # ... with 347 more rows
```

#we pick Neymar

#2 midfielders, one on the right and one on the left
#right midfielder: RCM, CDM, RDM, CM

```
fifa20 %>% filter(team_position %in% c("RCM", "CDM", "RDM", "CM")) %>%
  arrange(desc(overall)) %>%
  select(short_name, team_position, overall)
```

```
## # A tibble: 912 x 3
##   short_name     team_position overall
##   <chr>           <chr>          <int>
## 1 K. De Bruyne   RCM             91
## 2 L. Modrić      RCM             90
## 3 N. Kanté       RCM             89
## 4 Sergio Busquets CDM            89
## 5 Casemiro       CDM             87
## 6 Fernandinho    CDM             87
## 7 Thiago         CDM             87
## 8 M. Verratti    RCM             86
## 9 I. Rakitić     RCM             86
## 10 M. Pjanić      CDM             86
## # ... with 902 more rows
```

#we pick De Bruyne

#left midfielder: positions to be considered: LCM, CDM, LDM, CM

```
fifa20 %>% filter(team_position %in% c("LCM", "CDM", "LDM", "CM")) %>%
  arrange(desc(overall)) %>%
  select(short_name, team_position, overall)
```

```
## # A tibble: 910 x 3
##   short_name     team_position overall
##   <chr>           <chr>          <int>
## 1 Sergio Busquets CDM             89
```

```

## 2 P. Pogba      LDM      88
## 3 David Silva   LCM      88
## 4 T. Kroos       LCM      88
## 5 Casemiro      CDM      87
## 6 Fernandinho    CDM      87
## 7 Thiago         CDM      87
## 8 M. Pjanic     CDM      86
## 9 Coutinho        LCM      86
## 10 F. de Jong    LCM      85
## # ... with 900 more rows

```

#we pick Busquets

```

#2wing backs, one on the right and one on the left:
#left back, positions to be considered: LB, LWB
fifa20 %>% filter(team_position %in% c("LWB","LB")) %>%
  arrange(desc(overall))%>%
  select(short_name, team_position, overall)

```

```

## # A tibble: 618 x 3
##   short_name  team_position overall
##   <chr>        <chr>        <int>
## 1 Jordi Alba    LB            87
## 2 A. Robertson  LB            85
## 3 D. Alaba      LB            85
## 4 Marcelo        LB            85
## 5 Alex Sandro    LB            85
## 6 Alex Telles    LB            84
## 7 Grimaldo       LB            83
## 8 L. Digne        LB            83
## 9 N. Tagliafico  LB            82
## 10 N. Schulz     LB            82
## # ... with 608 more rows

```

```

# we pick Jordi Alba
#right back, positions to be considered: RB, RWB
fifa20 %>% filter(team_position %in% c("RWB","RB")) %>%
  arrange(desc(overall))%>%
  select(short_name, team_position, overall)

```

```

## # A tibble: 618 x 3
##   short_name  team_position overall
##   <chr>        <chr>        <int>
## 1 J. Kimmich     RB            86
## 2 Carvajal       RB            85
## 3 Azpilicueta    RB            84
## 4 K. Walker       RB            84
## 5 T. Alexander-Arnold RB            83
## 6 NÃ©lson Semedo RB            82
## 7 Ricardo Pereira RB            82
## 8 JesÃ³s Navas     RB            82
## 9 L. Piszczek     RB            81

```

```

## 10 AndrÃ© Almeida      RB          81
## # ... with 608 more rows

#we pick Joshua Kimmich

#2 centre backs
#positions to be considered: LCB, RCB
fifa20 %>% filter(team_position %in% c("LCB", "RCB")) %>%
  arrange(desc(overall))%>%
  select(short_name, team_position, overall)

## # A tibble: 1,320 x 3
##   short_name  team_position overall
##   <chr>        <chr>       <int>
## 1 V. van Dijk  LCB           90
## 2 K. Koulibaly LCB           89
## 3 G. Chiellini LCB           89
## 4 Sergio Ramos LCB           89
## 5 PiquÃ©      RCB           88
## 6 D. GodÃ¡n    RCB           88
## 7 A. Laporte   LCB           87
## 8 Thiago Silva LCB           87
## 9 J. Vertonghen LCB           87
## 10 M. Hummels  LCB           87
## # ... with 1,310 more rows

# we pick VVD and one of Koulibaly, Chiellini and Ramos

#Goalkeeper, positions: GK
fifa20 %>% filter(team_position %in% c("GK")) %>%
  arrange(desc(overall))%>%
  select(short_name, team_position, overall)

## # A tibble: 662 x 3
##   short_name  team_position overall
##   <chr>        <chr>       <int>
## 1 J. Oblak     GK            91
## 2 M. ter Stegen GK            90
## 3 Alisson      GK            89
## 4 De Gea       GK            89
## 5 Ederson      GK            88
## 6 T. Courtois  GK            88
## 7 "S. HandanoviÄ\u008d" GK            88
## 8 M. Neuer      GK            88
## 9 H. Lloris     GK            88
## 10 W. SzczÄsny   GK            86
## # ... with 652 more rows

#we pick Jan Oblak

#Therefore the team comprises of Lewandowski, Messi, Ronaldo, De Bruyne, Busquets,
#Neymar, Alba, Van Dijk, Ramos, Kimmich and Oblak.

```

```
#####HIGHEST RATED PLAYERS FOR FAMOUS JERSEY NUMBERS: 1,2,3,4,5,6,7,8,9,10,11
fifa20 %>% filter(team_jersey_number==1) %>%
  arrange(desc(overall))%>%
  select(short_name, team_jersey_number, overall)
```

```
## # A tibble: 575 x 3
##   short_name      team_jersey_number overall
##   <chr>                <int>     <int>
## 1 M. ter Stegen            1        90
## 2 Alisson                  1        89
## 3 De Gea                     1        89
## 4 "S. Handanovi\u00e1"
## 5 M. Neuer                  1        88
## 6 H. Lloris                  1        88
## 7 K. Navas                   1        87
## 8 W. Szcz\u0144sny             1        86
## 9 A. Lopes                   1        85
## 10 P. G\u00f3l\u00e1cs
## # ... with 565 more rows
```

#Ter Stegen

```
fifa20 %>% filter(team_jersey_number==2) %>%
  arrange(desc(overall))%>%
  select(short_name, team_jersey_number, overall)
```

```
## # A tibble: 499 x 3
##   short_name      team_jersey_number overall
##   <chr>                <int>     <int>
## 1 D. God\u00e3o                 2        88
## 2 Thiago Silva               2        87
## 3 J. Gim\u00e3nez                2        85
## 4 Carvajal                   2        85
## 5 K. Walker                   2        84
## 6 D. Djene\u00e7o                 2        83
## 7 N\u00e3\u00e3o Semedo              2        82
## 8 A. R\u00e3\u00e3diger                2        82
## 9 M\u00e3\u00e3rio Fernandes           2        82
## 10 V. Lindel\u00e4uf                2        81
## # ... with 489 more rows
```

#Godin

```
fifa20 %>% filter(team_jersey_number==3) %>%
  arrange(desc(overall))%>%
  select(short_name, team_jersey_number, overall)
```

```
## # A tibble: 526 x 3
##   short_name      team_jersey_number overall
```

```

##      <chr>          <int>  <int>
## 1 G. Chiellini           3     89
## 2 PiquÃ©             3     88
## 3 Fabinho            3     85
## 4 Pepe               3     84
## 5 Grimaldo           3     83
## 6 RaÃ°ol Albiol       3     83
## 7 JosuÃ© Chiamulera   3     83
## 8 P. Kimpembe         3     82
## 9 Juiano Mestres      3     82
## 10 Ã¶der MilitÃ±o       3     81
## # ... with 516 more rows

```

#Chiellini

```

fifa20 %>% filter(team_jersey_number==4) %>%
  arrange(desc(overall))%>%
  select(short_name, team_jersey_number, overall)

```

```

## # A tibble: 553 x 3
##   short_name    team_jersey_number overall
##   <chr>          <int>     <int>
## 1 V. van Dijk           4     90
## 2 Sergio Ramos         4     89
## 3 T. Alderweireld      4     87
## 4 I. RakitiÃ‡           4     86
## 5 M. de Ligt            4     85
## 6 N. SÃ¼le              4     85
## 7 J. Tah                4     83
## 8 V. Kompany            4     83
## 9 R. Nainggolan          4     83
## 10 Illarramendi         4     83
## # ... with 543 more rows

```

#Van Dijk

```

fifa20 %>% filter(team_jersey_number==5) %>%
  arrange(desc(overall))%>%
  select(short_name, team_jersey_number, overall)

```

```

## # A tibble: 580 x 3
##   short_name    team_jersey_number overall
##   <chr>          <int>     <int>
## 1 Sergio Busquets        5     89
## 2 J. Vertonghen          5     87
## 3 Marquinhos             5     86
## 4 M. Pjanic              5     86
## 5 R. Varane               5     85
## 6 Allan                  5     85
## 7 Sokratis                5     84
## 8 G. Wijnaldum            5     84
## 9 J. Stones                 5     83
## 10 Bartra                 5     83
## # ... with 570 more rows

```

```
#Busquets
```

```
fifa20 %>% filter(team_jersey_number==6) %>%
  arrange(desc(overall))%>%
  select(short_name, team_jersey_number, overall)
```

```
## # A tibble: 565 x 3
##   short_name      team_jersey_number overall
##   <chr>                <int>     <int>
## 1 P. Pogba                  6        88
## 2 Thiago                   6        87
## 3 M. Verratti               6        86
## 4 Koke                      6        85
## 5 S. de Vrij                 6        84
## 6 Lucas Leiva               6        84
## 7 D. Sánchez                6        83
## 8 S. Khedira                 6        83
## 9 T. Delaney                 6        82
## 10 Nacho Fernández             6        82
## # ... with 555 more rows
```

```
#Pogba
```

```
fifa20 %>% filter(team_jersey_number==7) %>%
  arrange(desc(overall))%>%
  select(short_name, team_jersey_number, overall)
```

```
## # A tibble: 599 x 3
##   short_name      team_jersey_number overall
##   <chr>                <int>     <int>
## 1 Cristiano Ronaldo            7        93
## 2 E. Hazard                   7        91
## 3 K. Mbappé                     7        89
## 4 N. Kanté                      7        89
## 5 R. Sterling                  7        88
## 6 H. Son                        7        87
## 7 J. Sancho                     7        84
## 8 José Callejón                  7        84
## 9 E. Villa                       7        84
## 10 Gonçalo Guedes                7        83
## # ... with 589 more rows
```

```
#Ronaldo
```

```
fifa20 %>% filter(team_jersey_number==8) %>%
  arrange(desc(overall))%>%
  select(short_name, team_jersey_number, overall)
```

```
## # A tibble: 596 x 3
##   short_name      team_jersey_number overall
##   <chr>                <int>     <int>
## 1 T. Kroos                      8        88
```

```

## 2 Saõl                      8    85
## 3 Bruno Fernandes            8    85
## 4 Arthur                     8    84
## 5 N. Fekir                   8    84
## 6 Felipe Anderson             8    84
## 7 Javi Martínez              8    84
## 8 I. GóndolaYan              8    84
## 9 Paulinho                   8    84
## 10 Fabián                    8    83
## # ... with 586 more rows

```

#Toni Kroos

```

fifa20 %>% filter(team_jersey_number==9) %>%
  arrange(desc(overall))%>%
  select(short_name, team_jersey_number, overall)

```

```

## # A tibble: 577 x 3
##   short_name      team_jersey_number overall
##   <chr>                <int>     <int>
## 1 L. Suárez                  9     89
## 2 R. Lewandowski              9     89
## 3 E. Cavani                  9     88
## 4 K. Benzema                 9     87
## 5 Roberto Firmino             9     86
## 6 A. Lacazette                9     86
## 7 R. Lukaku                  9     85
## 8 Z. Ibrahimović              9     85
## 9 E. Dáviko                  9     84
## 10 A. Martial                 9     83
## # ... with 567 more rows

```

#Lewandowski

```

fifa20 %>% filter(team_jersey_number==10) %>%
  arrange(desc(overall))%>%
  select(short_name, team_jersey_number, overall)

```

```

## # A tibble: 581 x 3
##   short_name team_jersey_number overall
##   <chr>                <int>     <int>
## 1 L. Messi                  10     94
## 2 Neymar Jr                  10     92
## 3 L. Modrić                  10     90
## 4 H. Kane                   10     89
## 5 S. Agüero                  10     89
## 6 P. Dybala                  10     88
## 7 S. Mané                      10     88
## 8 Coutinho                   10     86
## 9 Parejo                      10     86
## 10 A. Gómez                   10     85
## # ... with 571 more rows

```

```
#Messi

fifa20 %>% filter(team_jersey_number==11) %>%
  arrange(desc(overall))%>%
  select(short_name, team_jersey_number, overall)
```

```
## # A tibble: 582 x 3
##   short_name  team_jersey_number overall
##   <chr>          <int>     <int>
## 1 M. Salah            11      90
## 2 M. Reus             11      88
## 3 A. Di MarÃa         11      86
## 4 G. Bale              11      85
## 5 O. DembÃ©lÃ©           11      84
## 6 M. Depay             11      84
## 7 Douglas Costa        11      84
## 8 T. Werner             11      83
## 9 T. Lemar              11      83
## 10 W. Zaha              11      83
## # ... with 572 more rows
```

#Salah

```
#BEST TEAMS AGED 23 OR LESS AND MORE THAN 23:
#<=23
fifa20leq23 <- fifa20 %>% filter(age <= 23)
#Assuming a 4231 formation, we come up with the best team on fifa20 or the
#fifa world 11
#striker:
#positions to be considered: ST, CF, LS, RS, LF, RF
fifa20leq23 %>% filter(team_position %in% c("ST", "CF", "LS", "RS", "LF", "RF")) %>%
  arrange(desc(overall))%>%
  select(short_name, team_position, overall)
```

```
## # A tibble: 189 x 3
##   short_name  team_position overall
##   <chr>          <chr>     <int>
## 1 K. Havertz    RF          84
## 2 A. Martial    ST          83
## 3 T. Werner     LS          83
## 4 L. MartÃnez   RS          81
## 5 JoÃ£o FÃ©lix  RS          80
## 6 Diogo Jota    LS          80
## 7 M. DembÃ©lÃ©   ST          80
## 8 Santi Mina    LS          80
## 9 Wesley        ST          79
## 10 Joelinton   RS          78
## # ... with 179 more rows
```

#Havertz

```
#Wingers: 2 wingers to be picked, one on the left and one on the right.
#positions to be considered: RW, LW
fifa20leq23 %>% filter(team_position %in% c("LW", "RW")) %>%
  arrange(desc(overall)) %>%
  select(short_name, team_position, overall)
```

```
## # A tibble: 116 x 3
##   short_name  team_position overall
##   <chr>        <chr>          <int>
## 1 K. Mbappé    RW              89
## 2 K. Coman     LW              84
## 3 S. Gnabry    RW              84
## 4 Oyarzabal    LW              82
## 5 H. Lozano    RW              82
## 6 Rony Lopes   LW              80
## 7 Vinícius Jr. RW              79
## 8 F. Chiesa    RW              79
## 9 C. Pulisic   LW              79
## 10 M. Rashica  LW              79
## # ... with 106 more rows
```

#Mbappe, Coman

```
#one AM, positions: CAM, LAM, RAM
fifa20leq23 %>% filter(team_position %in% c("CAM", "LAM", "RAM")) %>%
  arrange(desc(overall)) %>%
  select(short_name, team_position, overall)
```

```
## # A tibble: 90 x 3
##   short_name  team_position overall
##   <chr>        <chr>          <int>
## 1 D. Alli      CAM             84
## 2 S. Bergwijn  CAM             82
## 3 D. van de Beek CAM            81
## 4 David Neres  LAM             81
## 5 Lucas Paquetá CAM            79
## 6 F. Neuhaus   CAM             78
## 7 M. Zaracho   CAM             78
## 8 M. Á-degaard  CAM            78
## 9 A. Miranchuk CAM             78
## 10 Y. Yazıcı    CAM            78
## # ... with 80 more rows
```

#Alli

```
#2 midfielders, one on the right and one on the left
#right midfielder: RCM, CDM, RDM, CM
fifa20leq23 %>% filter(team_position %in% c("RCM", "CDM", "RDM", "CM")) %>%
  arrange(desc(overall)) %>%
  select(short_name, team_position, overall)
```

```
## # A tibble: 230 x 3
```

```

##   short_name  team_position overall
##   <chr>        <chr>          <int>
## 1 L. Torreira    RDM           82
## 2 RÃ©ben Neves   CM            82
## 3 W. Ndidi       CDM           81
## 4 N. Barella     RCM           80
## 5 Y. Tielemans   RCM           80
## 6 H. Winks       RDM           80
## 7 M. Eggestein   RCM           80
## 8 J. Weigl       RDM           80
## 9 L. Pellegrini   RDM           79
## 10 F. KessiÃ©      RCM           79
## # ... with 220 more rows

```

```

#Torreira
#left midfielder: positions to be considered: LCM, CDM, LDM, CM
fifa201eq23 %>% filter(team_position %in% c("LCM", "CDM", "LDM", "CM")) %>%
  arrange(desc(overall)) %>%
  select(short_name, team_position, overall)

```

```

## # A tibble: 229 x 3
##   short_name  team_position overall
##   <chr>        <chr>          <int>
## 1 F. de Jong    LCM           85
## 2 FabiÃ¡n       LCM           83
## 3 RÃ©ben Neves   CM            82
## 4 H. Aouar       LCM           81
## 5 T. Ndombele   LDM           81
## 6 W. Ndidi       CDM           81
## 7 Merino         LCM           79
## 8 S. Berge       LCM           79
## 9 Dani Olmo     LCM           79
## 10 F. Grillitsch LCM           79
## # ... with 219 more rows

```

```

#Frenkie De Jong
#2wing backs, one on the right and one on the left:
#left back, positions to be considered: LB, LWB
fifa201eq23 %>% filter(team_position %in% c("LWB", "LB")) %>%
  arrange(desc(overall)) %>%
  select(short_name, team_position, overall)

```

```

## # A tibble: 175 x 3
##   short_name  team_position overall
##   <chr>        <chr>          <int>
## 1 Grimaldo      LB            83
## 2 L. Shaw        LB            81
## 3 B. Chilwell    LB            79
## 4 O. Zinchenko   LB            78
## 5 AarÃ³n MartÃ¡n LB            78
## 6 ReguilÃ³n      LB            77
## 7 Renan Lodi     LB            77

```

```

## 8 M. Sarr      LB      76
## 9 Rubén Duarte LB      76
## 10 Pedraza     LB      76
## # ... with 165 more rows

```

```

#Grimaldo
#right back, positions to be considered: RB, RWB
fifa20leq23 %>% filter(team_position %in% c("RWB", "RB")) %>%
  arrange(desc(overall)) %>%
  select(short_name, team_position, overall)

```

```

## # A tibble: 162 x 3
##   short_name      team_position overall
##   <chr>            <chr>        <int>
## 1 T. Alexander-Arnold RB          83
## 2 A. Wan-Bissaka   RB          79
## 3 T. Kehrer       RB          79
## 4 L. Klostermann  RWB         79
## 5 D. Dumfries    RB          78
## 6 D. Calabria    RB          77
## 7 Y. Atal         RB          77
## 8 J. Mähle       RB          77
## 9 N. Mazraoui    RB          77
## 10 F. Bustos     RB          77
## # ... with 152 more rows

```

```

#Alexander Arnold
#2 centre backs
#positions to be considered: LCB, RCB
fifa20leq23 %>% filter(team_position %in% c("LCB", "RCB")) %>%
  arrange(desc(overall)) %>%
  select(short_name, team_position, overall)

```

```

## # A tibble: 258 x 3
##   short_name      team_position overall
##   <chr>            <chr>        <int>
## 1 N. Sále         RCB          85
## 2 L. Hernández   LCB          84
## 3 J. Tah          RCB          83
## 4 M. Akanji       RCB          83
## 5 J. Gomez        RCB          80
## 6 A. Christensen RCB          80
## 7 Ráben Dias     RCB          80
## 8 Ferro           LCB          78
## 9 J. Andersen    RCB          78
## 10 I. Diop         LCB          78
## # ... with 248 more rows

```

```
#Sule and Lucas Hernandez
```

```
#Goalkeeper, positions: GK
```

```
fifa20[team_position %in% c("GK")] %>%
  arrange(desc(overall)) %>%
  select(short_name, team_position, overall)
```

```
## # A tibble: 106 x 3
##   short_name  team_position overall
##   <chr>        <chr>       <int>
## 1 G. Donnarumma GK            85
## 2 A. Onana      GK            82
## 3 M. Maignan    GK            80
## 4 D. Livaković GK            80
## 5 A. Lafont     GK            79
## 6 A. Meret      GK            78
## 7 E. Audero     GK            77
## 8 P. Bernardoni GK            77
## 9 Rubén Blanco  GK            76
## 10 A. Gunn      GK            76
## # ... with 96 more rows
```

#Donnarumma

#Therefore the best 11 under-23 or 23 is: Havertz, Mbappe, Coman, Torreira,
#Frenkie De Jong, Alli, Grimaldo, TAA, Sule, Lucas Hernandez and Donnarumma

#ONE INTERESTING OBSERVATION MADE IS THAT NO PLAYER AGED 23 OR LESS MADE IT
#INTO THE FIFA WORLD 11 SO THE BEST TEAM AGED MORE THAN 23 IS THE SAME AS THE
#FIFA WORLD 11.

####BEST PLAYER FOR TOP 10 RANKED COUNTRIES IN THE FIFA RANKINGS

#Current FIFA rankings:

1.Belgium, 2.France, 3.Brazil, 4.England, 5.Uruguay, 6.Croatia, 7.Portugal,
#8.Spain, 9.Argentina, 10.Colombia

#Belgium:

```
fifa20 %>% filter(nationality=="Belgium") %>% arrange(desc(overall)) %>%
  select(short_name, club, overall) %>% top_n(1)
```

Selecting by overall

```
## # A tibble: 2 x 3
##   short_name  club          overall
##   <chr>        <chr>       <int>
## 1 E. Hazard    Real Madrid    91
## 2 K. De Bruyne Manchester City 91
```

#Hazard and De bruyne

#France

```
fifa20 %>% filter(nationality=="France") %>% arrange(desc(overall)) %>%
  select(short_name, club, overall) %>% top_n(1)
```

Selecting by overall

```

## # A tibble: 3 x 3
##   short_name   club      overall
##   <chr>        <chr>     <int>
## 1 K. Mbappé    Paris Saint-Germain     89
## 2 N. Kanté     Chelsea            89
## 3 A. Griezmann FC Barcelona        89

#Mbappe, Kante and Griezmann
#Brazil
fifa20 %>% filter(nationality=="Brazil") %>% arrange(desc(overall)) %>%
  select(short_name, club, overall) %>% top_n(1)

```

Selecting by overall

```

## # A tibble: 1 x 3
##   short_name   club      overall
##   <chr>        <chr>     <int>
## 1 Neymar Jr   Paris Saint-Germain     92

```

```

#Neymar
#England
fifa20 %>% filter(nationality=="England") %>% arrange(desc(overall)) %>%
  select(short_name, club, overall) %>% top_n(1)

```

Selecting by overall

```

## # A tibble: 1 x 3
##   short_name   club      overall
##   <chr>        <chr>     <int>
## 1 H. Kane     Tottenham Hotspur       89

```

```

#Harry Kane
#Uruguay
fifa20 %>% filter(nationality=="Uruguay") %>% arrange(desc(overall)) %>%
  select(short_name, club, overall) %>% top_n(1)

```

Selecting by overall

```

## # A tibble: 1 x 3
##   short_name   club      overall
##   <chr>        <chr>     <int>
## 1 L. Suárez   FC Barcelona        89

```

```

#Luis Suarez
#Croatia
fifa20 %>% filter(nationality=="Croatia") %>% arrange(desc(overall)) %>%
  select(short_name, club, overall) %>% top_n(1)

```

Selecting by overall

```

## # A tibble: 1 x 3
##   short_name club      overall
##   <chr>     <chr>      <int>
## 1 L. Modrić Real Madrid      90

#Luka Modric
#Portugal
fifa20 %>% filter(nationality=="Portugal")%>%arrange(desc(overall))%>%
  select(short_name, club, overall) %>% top_n(1)

```

Selecting by overall

```

## # A tibble: 1 x 3
##   short_name club      overall
##   <chr>     <chr>      <int>
## 1 Cristiano Ronaldo Juventus      93

```

```

#Cristiano Ronaldo
#Spain
fifa20 %>% filter(nationality=="Spain")%>%arrange(desc(overall))%>%
  select(short_name, club, overall) %>% top_n(1)

```

Selecting by overall

```

## # A tibble: 3 x 3
##   short_name club      overall
##   <chr>     <chr>      <int>
## 1 De Gea      Manchester United    89
## 2 Sergio Ramos Real Madrid      89
## 3 Sergio Busquets FC Barcelona    89

```

```

#David De Gea, Sergio Ramos and Sergio Busquets
#Argentina
fifa20 %>% filter(nationality=="Argentina")%>%arrange(desc(overall))%>%
  select(short_name, club, overall) %>% top_n(1)

```

Selecting by overall

```

## # A tibble: 1 x 3
##   short_name club      overall
##   <chr>     <chr>      <int>
## 1 L. Messi    FC Barcelona      94

```

```

#Lionel Messi
#Colombia
fifa20 %>% filter(nationality=="Colombia")%>%arrange(desc(overall))%>%
  select(short_name, club, overall) %>% top_n(1)

```

Selecting by overall

```

## # A tibble: 1 x 3
##   short_name    club      overall
##   <chr>        <chr>     <int>
## 1 J. Rodríguez Real Madrid     85

#James Rodriguez

####PLAYERS WITH THE BEST SKILL MOVES WITH OVERALL >=85
fifa20 %>% filter(overall>=85,skill_moves==5)%>%arrange(desc(skill_moves))%>%
  select(short_name, club, nationality, team_position,overall,skill_moves)%>%
  group_by(skill_moves) %>% top_n(1)%>%ungroup()

## Selecting by skill_moves

## # A tibble: 9 x 6
##   short_name    club      nationality team_position overall skill_moves
##   <chr>        <chr>     <chr>       <chr>      <int>     <int>
## 1 Cristiano Ronaldo Juventus Portugal    LW          93        5
## 2 Neymar Jr      Paris Saint-Germain Brazil     CAM         92        5
## 3 K. Mbappé      Paris Saint-Germain France    RW          89        5
## 4 P. Pogba       Manchester United France    LDM         88        5
## 5 Thiago         FC Bayern München Spain     CDM         87        5
## 6 A. Di María    Paris Saint-Germain Argentina LW          86        5
## 7 Coutinho        FC Bayern München Brazil    LCM         86        5
## 8 Z. Ibrahimović LA Galaxy Sweden       ST          85        5
## 9 Marcelo        Real Madrid Brazil       LB          85        5

####PLAYERS WITH BEST WEAK FOOT WITH OVERALL >=85
fifa20 %>% filter(overall>=85,weak_foot==5)%>%arrange(desc(weak_foot))%>%
  select(short_name, club, nationality, team_position,overall,weak_foot)%>%
  group_by(weak_foot) %>% top_n(1)%>%ungroup()

## Selecting by weak_foot

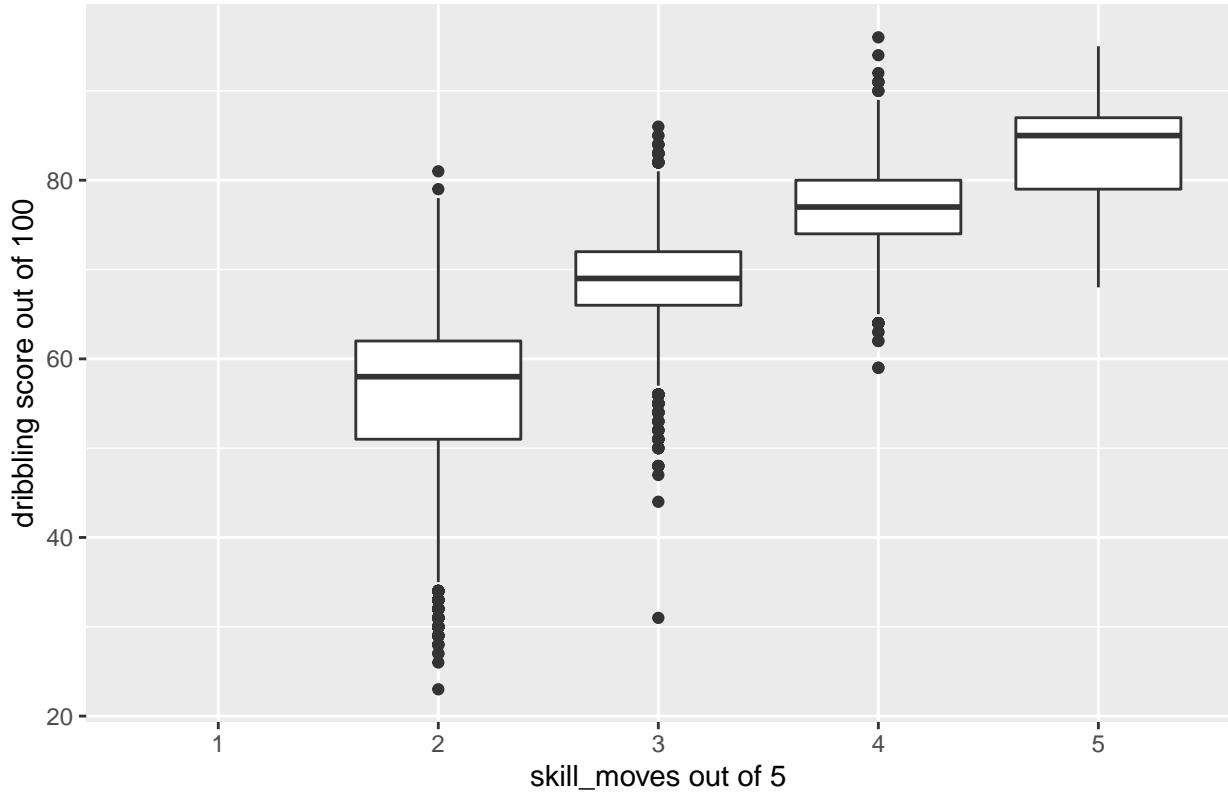
## # A tibble: 5 x 6
##   short_name    club      nationality team_position overall weak_foot
##   <chr>        <chr>     <chr>       <chr>      <int>     <int>
## 1 Neymar Jr      Paris Saint-Germain Brazil     CAM         92        5
## 2 K. De Bruyne    Manchester City   Belgium    RCM         91        5
## 3 C. Eriksen     Tottenham Hotspur Denmark   RM          88        5
## 4 T. Kroos        Real Madrid     Germany   LCM         88        5
## 5 H. Son          Tottenham Hotspur Korea Republ LM          87        5

##Are skill_moves and dribbling correlated?
#Boxplot of skill_moves vs dribbling
fifa20 %>% ggplot(aes(as.factor(skill_moves),dribbling))+geom_boxplot()+
  labs(x="skill_moves out of 5",y="dribbling score out of 100",
       title = "Boxplot of skill_moves vs dribbling")

## Warning: Removed 2036 rows containing non-finite values (stat_boxplot).

```

Boxplot of skill_moves vs dribbling



```
#They are positively correlated
#Strength of correlation:
cor(fifa20$skill_moves, fifa20$dribbling, method = "spearman",
     use = "complete.obs")
```

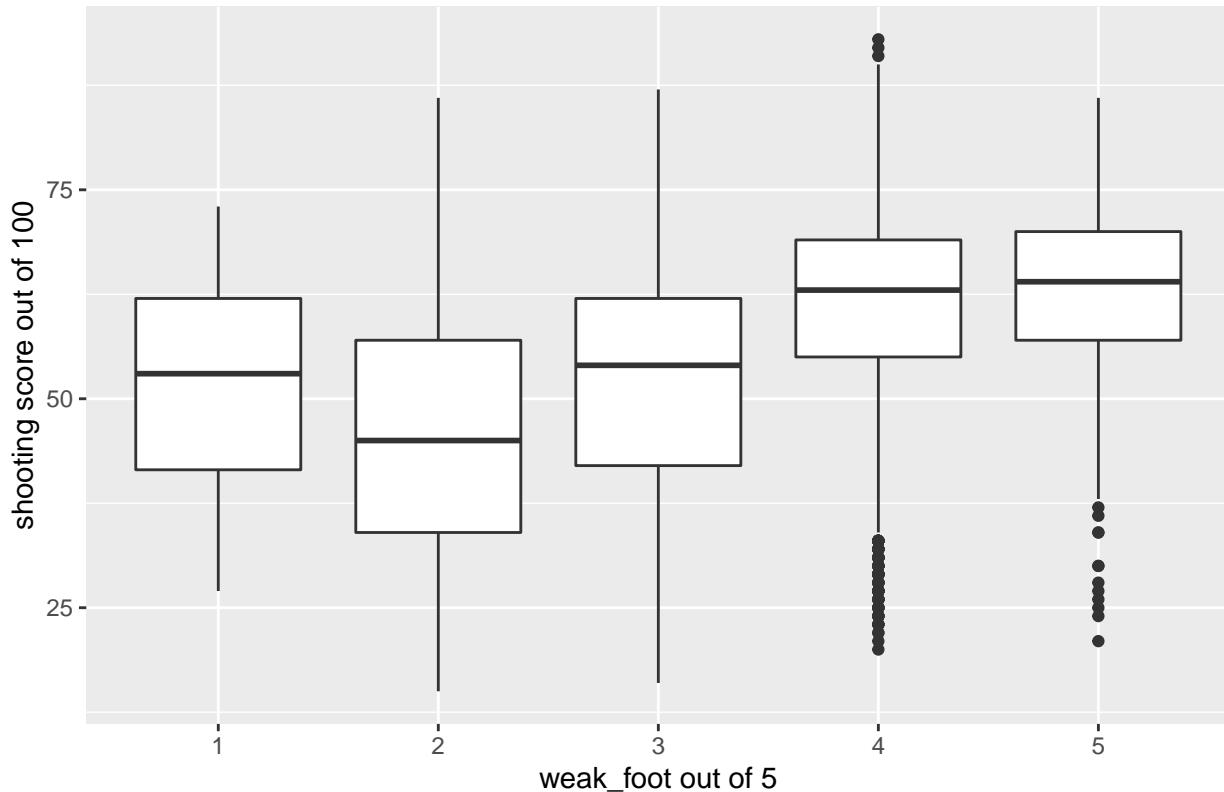
```
## [1] 0.7873962
```

```
#high positive correlation.
```

```
##Are players with a strong weak foot good at shooting?
#Boxplot of weak foot vs shooting
fifa20 %>% ggplot(aes(as.factor(weak_foot), shooting)) + geom_boxplot() +
  labs(x = "weak_foot out of 5", y = "shooting score out of 100",
       title = "Boxplot of weak foot vs shooting")
```

```
## Warning: Removed 2036 rows containing non-finite values (stat_boxplot).
```

Boxplot of weak foot vs shooting



```
#Strength of correlation:
cor(fifa20$weak_foot, fifa20$shooting, method = "spearman",
    use = "complete.obs")

## [1] 0.3237961

#weak positive correlation

####split work rate into 2 columns: attack and defence work rate:
fifa20 <- separate(fifa20, work_rate, into = c("attack_workrate","defence_workrate"),
                     sep = "/")

#good players with high attack and defence workrates:
fifa20 %>% filter(overall>=85, attack_workrate=="High", defence_workrate=="High") %>%
  arrange(desc(overall)) %>%
  select(short_name, club, nationality, team_position,
         attack_workrate, defence_workrate, overall) %>% group_by(overall) %>%
  top_n(1) %>% ungroup()

## Selecting by overall

## # A tibble: 18 x 7
##   short_name club  nationality team_position attack_workrate defence_workrate
##   <chr>      <chr> <chr>       <chr>           <chr>           <chr>
## 1 ...
```

```

## 1 K. De Bruyne Manc~ Belgium RCM High High
## 2 L. Modrić Real~ Croatia RCM High High
## 3 H. Kane Tott~ England ST High High
## 4 A. Griezmann FC Barça France LW High High
## 5 C. Eriksen Tott~ Denmark RM High High
## 6 E. Cavani Pari~ Uruguay ST High High
## 7 Bernardo Silva Manc~ Portugal RW High High
## 8 H. Son Tott~ Korea Repu~ LM High High
## 9 J. Vertonghen Tott~ Belgium LCB High High
## 10 Roberto Soriano Live~ Brazil CF High High
## 11 T. Müller FC Bayern Germany SUB High High
## 12 A. Lacazette Arse~ France ST High High
## 13 Saúl Ñíguez Atlético Spain LCM High High
## 14 A. Robert Live~ Scotland LB High High
## 15 Bruno Fernandes Spor~ Portugal RCM High High
## 16 Carvajal Real~ Spain RB High High
## 17 Iago Aspas RC Celta~ Spain RS High High
## 18 Koke Atlético Madrid Spain RM High High
## # ... with 1 more variable: overall <int>

```

```

#Laziest players on fifa20
fifa20 %>% filter(attack_workrate == "Low", defence_workrate == "Low") %>%
  arrange(desc(overall)) %>%
  select(short_name, club, nationality, team_position,
         attack_workrate, defence_workrate, overall) %>% group_by(overall) %>%
  top_n(1) %>% ungroup()

```

```
## Selecting by overall
```

```

## # A tibble: 35 x 7
##   short_name club  nationality team_position attack_workrate defence_workrate
##   <chr>      <chr> <chr>        <chr>          <chr>          <chr>
## 1 C. Strandberg Århus~ Sweden    LS     Low    Low
## 2 M. Kramer   ADO ~ Netherlands SUB    Low    Low
## 3 A. Fink     Karl~ Germany   SUB    Low    Low
## 4 S. Ameobi   Nott~ England   SUB    Low    Low
## 5 P. Forsell  HJK ~ Finland  RES    Low    Low
## 6 C. Kazimoglu Tibi~ Turkey   RS     Low    Low
## 7 N. Schmidt  VfL ~ Germany  SUB    Low    Low
## 8 A. Ba       AJ Al~ Mauritania SUB    Low    Low
## 9 Abraham Gómez Tibi~ Spain   LDM    Low    Low
## 10 E. Kujović Djur~ Sweden   RES    Low    Low
## # ... with 25 more rows, and 1 more variable: overall <int>

```

```

#####SUMMARIES OF FEW COLUMNS
sum_stats <- fifa20 %>% select(overall, potential, value_eur, wage_eur,
                                    release_clause_eur,
                                    pace, shooting, passing,
                                    dribbling, physic, defending)
summary(sum_stats)

```

```
##   overall      potential      value_eur      wage_eur
```

```

## Min. :48.00  Min. :49.00  Min. :      0  Min. :      0
## 1st Qu.:62.00 1st Qu.:67.00 1st Qu.: 325000 1st Qu.: 1000
## Median :66.00 Median :71.00 Median : 700000 Median : 3000
## Mean   :66.24 Mean   :71.55 Mean   : 2484038 Mean   : 9457
## 3rd Qu.:71.00 3rd Qu.:75.00 3rd Qu.: 2100000 3rd Qu.: 8000
## Max.   :94.00  Max.   :95.00  Max.   :105500000 Max.   :565000
##
## release_clause_eur      pace      shooting      passing
## Min.   : 13000  Min.   :24.0  Min.   :15.0  Min.   :24.00
## 1st Qu.: 563000 1st Qu.:61.0 1st Qu.:42.0 1st Qu.:50.00
## Median : 1200000 Median :69.0  Median :54.0  Median :58.00
## Mean   : 4740717 Mean   :67.7  Mean   :52.3  Mean   :57.23
## 3rd Qu.: 3700000 3rd Qu.:75.0 3rd Qu.:63.0 3rd Qu.:64.00
## Max.   :195800000 Max.   :96.0  Max.   :93.0  Max.   :92.00
## NA's   :1298    NA's   :2036  NA's   :2036  NA's   :2036
## dribbling      physic      defending
## Min.   :23.00  Min.   :27.00  Min.   :15.00
## 1st Qu.:57.00 1st Qu.:59.00 1st Qu.:36.00
## Median :64.00  Median :66.00  Median :56.00
## Mean   :62.53  Mean   :64.88  Mean   :51.55
## 3rd Qu.:69.00 3rd Qu.:72.00 3rd Qu.:65.00
## Max.   :96.00  Max.   :90.00  Max.   :90.00
## NA's   :2036  NA's   :2036  NA's   :2036

```

#release_clause_eur is NA for 1298 players, this is because these players do not have a release clause included in their current contract.

#pace, shooting, passing, dribbling, defending and physic is NA for 2036 players, lets find out why

```

pace_na <- fifa20 %>% filter(is.na(pace),is.na(dribbling),is.na(shooting),
                                is.na(passing),is.na(defending),is.na(phovic))%>%
  select(short_name, club, nationality, team_position, overall)
dim(pace_na) #2036 players with NAs.

```

```
## [1] 2036     5
```

#team positions of these players:

```
unique(pace_na$team_position)
```

```
## [1] "GK"   "SUB"  "RES"  "
```

#We can see that goalkeepers, subs and reserve team players do not have values for these 6 attributes.

```
pace_na %>% filter(team_position=="")
```

```

## # A tibble: 40 x 5
##   short_name   club   nationality team_position overall
##   <chr>        <chr>  <chr>       <chr>          <int>
## 1 J. Serendero Uruguay Uruguay     ""              80
## 2 A. Lunev      Russia Russia     ""              79
## 3 L. SÁ;reda    Uruguay Uruguay     ""              79
## 4 P. DÁ;renas   Uruguay Uruguay     ""              75

```

```

## 5 H. Lindner Austria Austria    ""
## 6 M. Borjan Canada Canada    ""
## 7 J. Santigaro Ecuador Ecuador    ""
## 8 A. El-Shenawy Egypt Egypt    ""
## 9 A. Shunin Russia Russia    ""
## 10 V. Belec Slovenia Slovenia    ""
## # ... with 30 more rows

```

#There are 40 rows in this dataframe where the position is "". These players are fictional and do not belong to any club.

```

#RELEASE CLAUSE EDA
#players with the highest release clauses in euros:
fifa20 %>% arrange(desc(release_clause_eur)) %>%
  select(short_name, club, nationality, overall, wage_eur,
         value_eur, release_clause_eur)

```

```

## # A tibble: 18,278 x 7
##   short_name   club      nationality overall wage_eur value_eur release_clause_~
##   <chr>        <chr>      <chr>       <int>    <int>    <int>          <int>
## 1 L. Messi    FC Barce~ Argentina     94    565000  95500000  195800000
## 2 Neymar Jr   Paris Sa~ Brazil       92    290000  105500000  195200000
## 3 K. MbappÃ© Paris Sa~ France      89    155000  93500000  191700000
## 4 E. Hazard    Real Mad~ Belgium     91    470000  90000000  184500000
## 5 K. De Bruy~ Manchester~ Belgium    91    370000  90000000  166500000
## 6 J. Oblak     AtlÃ©tic~ Slovenia    91    125000  77500000  164700000
## 7 H. Kane      Tottenham~ England     89    220000  83000000  159800000
## 8 V. van Dijk  Liverpool~ Netherlands 90    200000  78000000  150200000
## 9 M. Salah     Liverpool~ Egypt      90    240000  80500000  148900000
## 10 M. ter Ste~ FC Barce~ Germany     90    250000  67500000  143400000
## # ... with 18,268 more rows

```

#Messi and Neymar have the highest release clauses in euros

#Are value and release clause correlated?

```

cor(fifa20$value_eur, fifa20$release_clause_eur, method = "pearson",
  use = "complete.obs")

```

```

## [1] 0.9937346

```

#We see that value and release clause have very high positive correlation, correlation coefficient almost equal to 1.

##Are wage and release clause correlated?

```

cor(fifa20$wage_eur, fifa20$release_clause_eur, method = "pearson",
  use = "complete.obs")

```

```

## [1] 0.8538087

```

```

#We also see that wage and release clause have high positive correlation.

#Do talented young players have high release clauses?
elite_young <- fifa20 %>% filter(age <= 23)%>%
  mutate(elite_pot_inc = potential-overall)%>%
  select(short_name, club, nationality, overall, potential,
         elite_pot_inc, release_clause_eur) %>%
  arrange(desc(elite_pot_inc))
elite_young

## # A tibble: 7,318 x 7
##   short_name club  nationality overall potential elite_pot_inc release_clause_eur
##   <chr>      <chr> <chr>       <int>     <int>        <int>            <int>
## 1 G. Bazunu Manc~ Republic o~      59        84          25        827000
## 2 S. Ramos ~ Boca~ Argentina      56        81          25        432000
## 3 B. Mumba Sund~ England         55        80          25        468000
## 4 S. Spasov Oxfo~ Bulgaria       49        74          25        193000
## 5 Tao Qiang~ Hebe~ China PR     48        73          25        153000
## 6 B. McPher~ Grim~ England       48        73          25        165000
## 7 K. Bafoun~ Boru~ France        59        83          24        861000
## 8 L. Cheval~ LOSC~ France        58        82          24        667000
## 9 J. GarcÃa Cruz~ Mexico       55        79          24        432000
## 10 H. Mnoga Port~ England      53        77          24        330000
## # ... with 7,308 more rows

#Check if 'elite_pot_inc' and 'release_clause_eur' are correlated
cor(elite_young$elite_pot_inc, elite_young$release_clause_eur,
    method = "pearson", use = "complete.obs")

## [1] -0.2187238

#They have a low negative correlation. One would expect, higher possibility of
#increase in overall would result in the player having a bigger release clause
#but that is not the case here.
elite_young %>% ggplot(aes(as.factor(elite_pot_inc), release_clause_eur))+
  # geom_boxplot()

#Do older players have smaller release clauses?
cor(fifa20$age, fifa20$release_clause_eur, method = "pearson",
    use = "complete.obs")

## [1] 0.06574811

#There is negligible correlation between these 2 variables which is surprising.

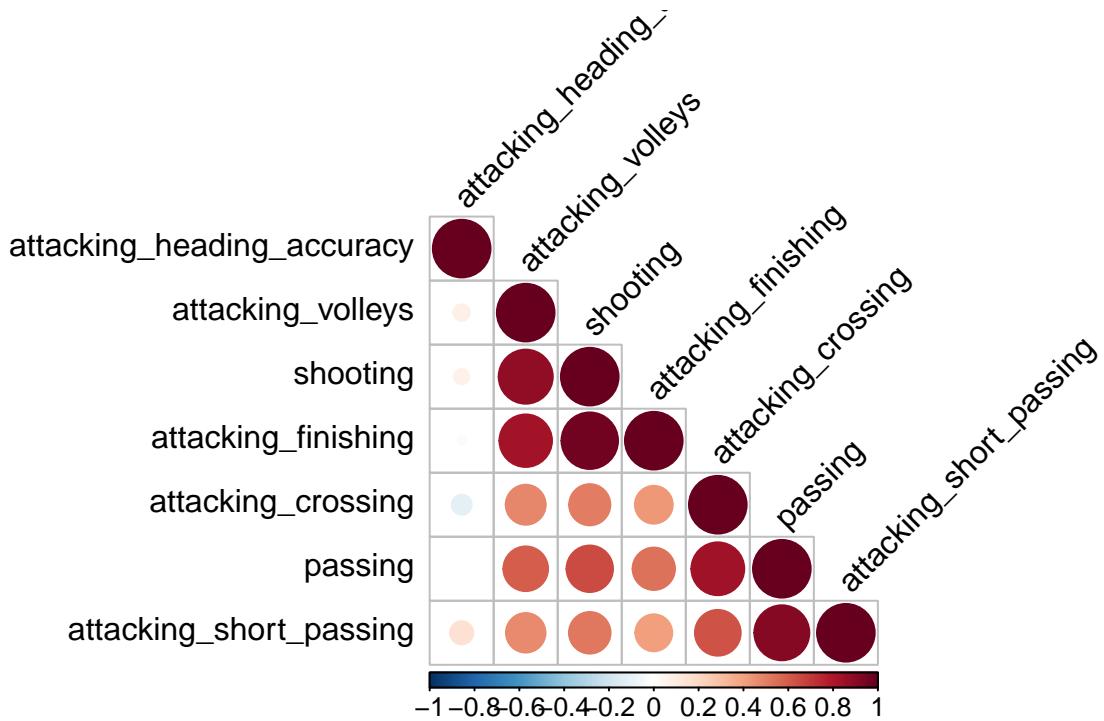
#####ATTACK STATS:
attack_stats <- fifa20%>% select(shooting, passing, attacking_crossing,
                                      attacking_finishing, attacking_heading_accuracy,
                                      attacking_short_passing, attacking_volleys)

```

```

source("http://www.sthda.com/upload/rquery_cormat.r")
require("corrplot")
rquery.cormat(attack_stats)

```



```

## $r
##                                     attacking_heading_accuracy attacking_volleys
## attacking_heading_accuracy                               1
## attacking_volleys                                0.081          1
## shooting                                         0.071          0.88
## attacking_finishing                            0.021          0.83
## attacking_crossing                           -0.12           0.48
## passing                                         0.0092          0.6
## attacking_short_passing                         0.16           0.47
##                                     shooting attacking_finishing attacking_crossing
## attacking_heading_accuracy
## attacking_volleys
## shooting
## attacking_finishing
## attacking_crossing
## passing
## attacking_short_passing
##                                     passing attacking_short_passing
## attacking_heading_accuracy
## attacking_volleys
## shooting

```

```

## attacking_finishing
## attacking_crossing
## passing 1
## attacking_short_passing 0.91 1
##
## $p
##          attacking_heading_accuracy attacking_volleys
## attacking_heading_accuracy 0
## attacking_volleys 0 0
## shooting 2e-19 0
## attacking_finishing 0 0
## attacking_crossing 0 0
## passing 0.24 0
## attacking_short_passing 0 0
##          shooting attacking_finishing attacking_crossing
## attacking_heading_accuracy
## attacking_volleys
## shooting 0
## attacking_finishing 0 0
## attacking_crossing 0 0 0
## passing 0 0 0
## attacking_short_passing 0 0 0
##          passing attacking_short_passing
## attacking_heading_accuracy
## attacking_volleys
## shooting
## attacking_finishing
## attacking_crossing
## passing 0
## attacking_short_passing 0 0
##
## $sym
##          attacking_heading_accuracy attacking_volleys
## attacking_heading_accuracy 1
## attacking_volleys 1
## shooting +
## attacking_finishing +
## attacking_crossing .
## passing .
## attacking_short_passing .
##          shooting attacking_finishing attacking_crossing
## attacking_heading_accuracy
## attacking_volleys
## shooting 1
## attacking_finishing B 1
## attacking_crossing . . 1
## passing , . +
## attacking_short_passing . . ,
##          passing attacking_short_passing
## attacking_heading_accuracy
## attacking_volleys
## shooting
## attacking_finishing
## attacking_crossing

```

```

## passing           1
## attacking_short_passing    *      1
## attr(,"legend")
## [1] 0 ' ' 0.3 '.' 0.6 ',' 0.8 '+' 0.9 '*' 0.95 'B' 1

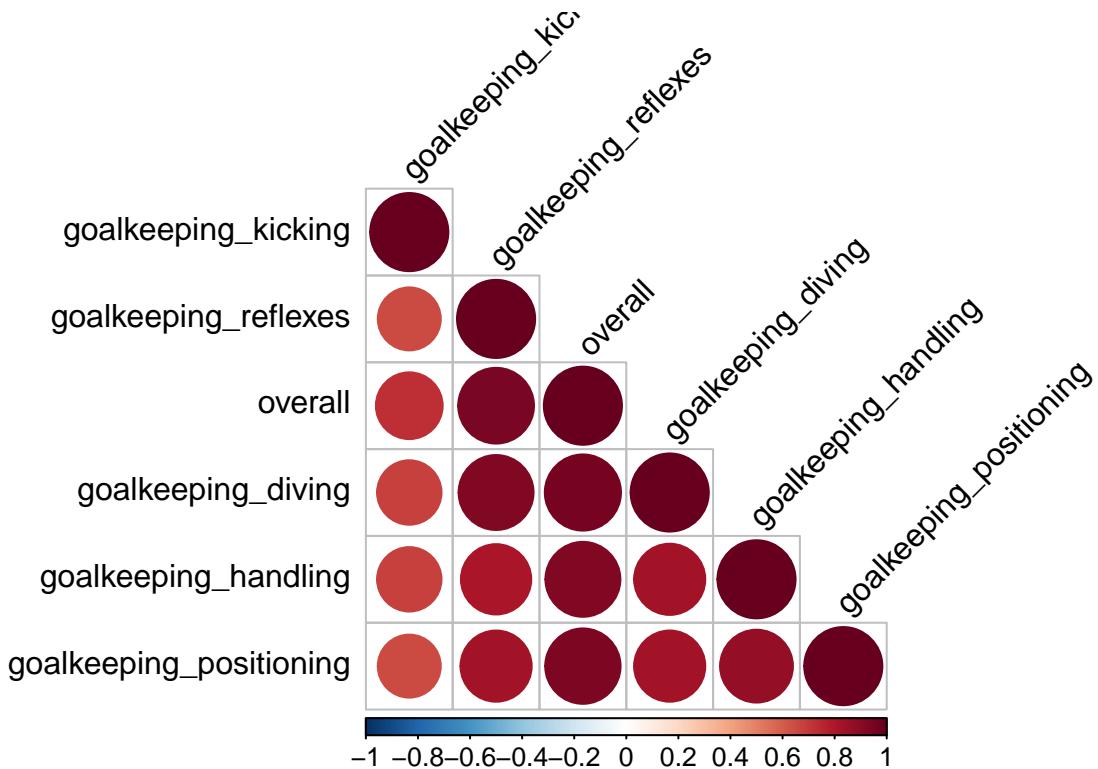
#Almost all pairs are positively correlated to each other.
#One exception is the correlation b/w heading and crossing. There is a -ve
#correlation b/w these 2 variables and it makes sense since a good crosser of the
#ball is the one delivering the crosses to the player whos good at heading.

```

```

#GOALKEEPING STATS:
gk_stats <- fifa20%>% filter(team_position=="GK")%>%
  select(overall, goalkeeping_diving, goalkeeping_handling,goalkeeping_kicking,
         goalkeeping_positioning, goalkeeping_reflexes)
source("http://www.sthda.com/upload/rquery_cormat.r")
require("corrplot")
rquery.cormat(gk_stats)

```



```

## $r
##               goalkeeping_kicking goalkeeping_reflexes overall
## goalkeeping_kicking           1
## goalkeeping_reflexes          0.65          1
## overall                         0.73          0.94        1

```

```

## goalkeeping_diving          0.68          0.92          0.95
## goalkeeping_handling        0.68          0.81          0.92
## goalkeeping_positioning     0.65          0.83          0.93
##                                     goalkeeping_diving  goalkeeping_handling
## goalkeeping_kicking
## goalkeeping_reflexes
## overall
## goalkeeping_diving          1
## goalkeeping_handling        0.83          1
## goalkeeping_positioning     0.83          0.87
##                                     goalkeeping_positioning
## goalkeeping_kicking
## goalkeeping_reflexes
## overall
## goalkeeping_diving
## goalkeeping_handling
## goalkeeping_positioning     1
##
## $p
##                                     goalkeeping_kicking  goalkeeping_reflexes  overall
## goalkeeping_kicking           0
## goalkeeping_reflexes         2.8e-82          0
## overall                      2.3e-111 1.50000000019285e-315  0
## goalkeeping_diving           1.1e-89          4.9e-272          0
## goalkeeping_handling         2.9e-90          1.1e-154 1.5e-272
## goalkeeping_positioning      1e-81           1.5e-169 2.7e-290
##                                     goalkeeping_diving  goalkeeping_handling
## goalkeeping_kicking
## goalkeeping_reflexes
## overall
## goalkeeping_diving           0
## goalkeeping_handling         2.9e-168          0
## goalkeeping_positioning      1.9e-168          1e-204
##                                     goalkeeping_positioning
## goalkeeping_kicking
## goalkeeping_reflexes
## overall
## goalkeeping_diving
## goalkeeping_handling
## goalkeeping_positioning      0
##
## $sym
##                                     goalkeeping_kicking  goalkeeping_reflexes  overall
## goalkeeping_kicking           1
## goalkeeping_reflexes          ,
## overall                      ,          *
## goalkeeping_diving            ,
## goalkeeping_handling          ,
## goalkeeping_positioning       ,
##                                     goalkeeping_diving  goalkeeping_handling
## goalkeeping_kicking
## goalkeeping_reflexes
## overall
## goalkeeping_diving           1

```

```

## goalkeeping_handling      +
## goalkeeping_positioning +           1
##                                     +
##                                     goalkeeping_positioning
## goalkeeping_kicking
## goalkeeping_reflexes
## overall
## goalkeeping_diving
## goalkeeping_handling
## goalkeeping_positioning 1
## attr(),"legend")
## [1] 0 ' ' 0.3 '.' 0.6 ',' 0.8 '+' 0.9 '*' 0.95 'B' 1

```

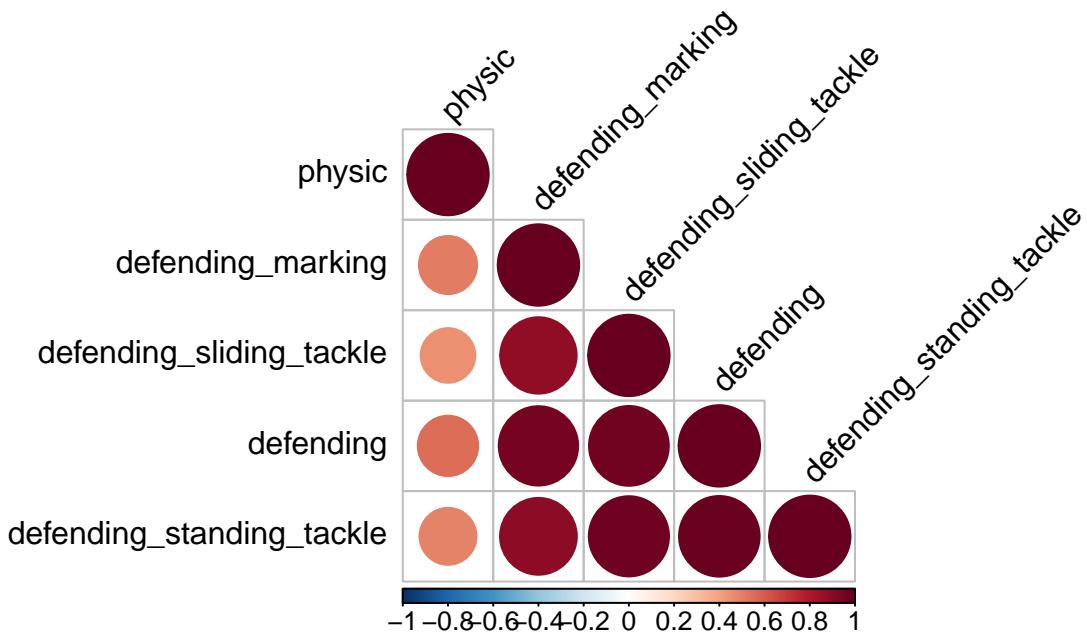
#kicking is not as highly correlated to GK overall when compared to other goalkeeping attributes.

#DEFENDING STATS:

```

defence_stats <- fifa20 %>% select(defending, defending_marking,
                                         defending_standing_tackle, defending_sliding_tackle,
                                         physic)
source("http://www.sthda.com/upload/rquery_cormat.r")
require("corrplot")
rquery.cormat(defence_stats)

```



```
## $r
```

```

##                                     physic defending_marking defending_sliding_tackle
## physic                               1
## defending_marking                  0.51                         1
## defending_sliding_tackle          0.45                         0.88
## defending                           0.55                         0.95
## defending_standing_tackle        0.49                         0.89
##                                         defending defending_standing_tackle
## physic
## defending_marking
## defending_sliding_tackle
## defending                           1
## defending_standing_tackle        0.98                         1
##
## $p
##                                     physic defending_marking defending_sliding_tackle
## physic                            0
## defending_marking                 0                         0
## defending_sliding_tackle          0                         0
## defending                           0                         0
## defending_standing_tackle        0                         0
##                                         defending defending_standing_tackle
## physic
## defending_marking
## defending_sliding_tackle
## defending                           0
## defending_standing_tackle        0                         0
##
## $sym
##                                     physic defending_marking defending_sliding_tackle
## physic                            1
## defending_marking                .
## defending_sliding_tackle         .      +
## defending                           .
## defending_standing_tackle        .      +
##                                         defending defending_standing_tackle
## physic
## defending_marking
## defending_sliding_tackle
## defending                           1
## defending_standing_tackle        B      1
## attr(,"legend")
## [1] 0 ' ' 0.3 '.' 0.6 ',' 0.8 '+' 0.9 '*' 0.95 'B' 1

```

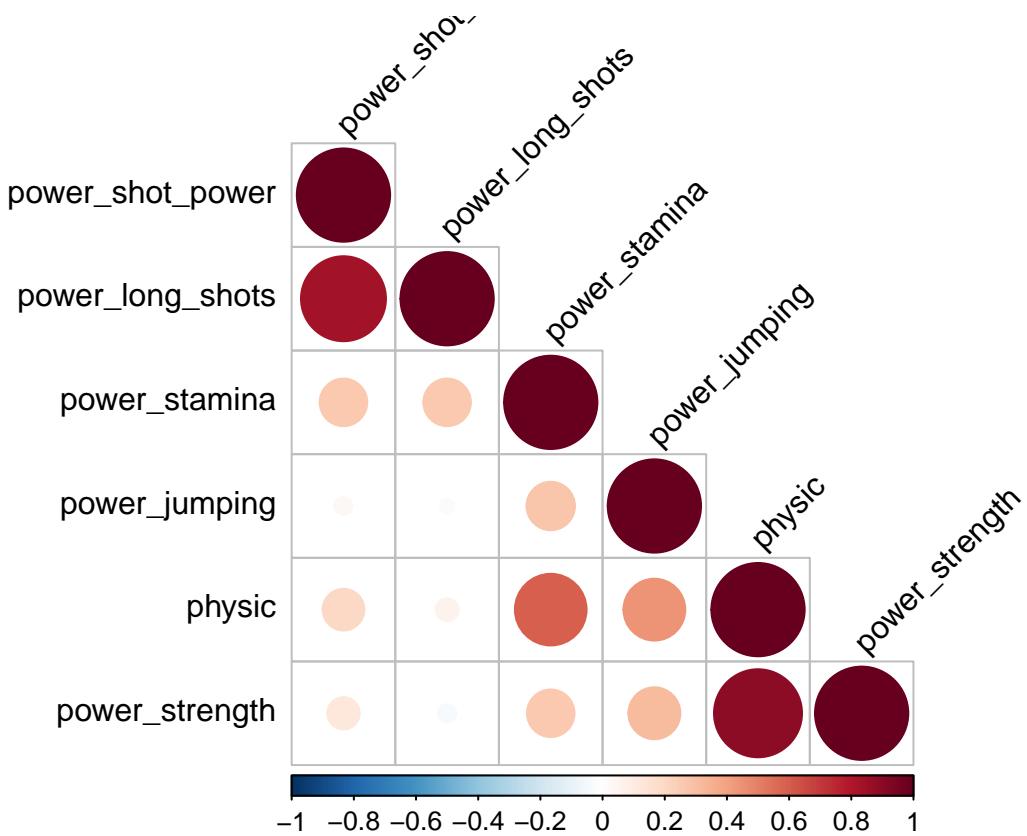
#Though physique is not highly correlated to defence stats, having a high physique rating out of 100 implies that the player is good at defending.

#POWER STATS:

```

power_stats <- fifa20%>%select(physic, power_shot_power, power_jumping,
                                 power_stamina, power_strength, power_long_shots)
source("http://www.sthda.com/upload/rquery_cormat.r")
require("corrplot")
rquery.cormat(power_stats)

```



```

## $r
##          power_shot_power power_long_shots power_stamina power_jumping
## power_shot_power           1
## power_long_shots        0.83           1
## power_stamina         0.26           0.26           1
## power_jumping        0.037          -0.023          0.27           1
## physic                 0.2            0.06          0.59          0.44
## power_strength       0.12          -0.038          0.26          0.31
##                  physic power_strength
## power_shot_power           1
## power_long_shots
## power_stamina
## power_jumping
## physic                1
## power_strength      0.89           1
## 
## $p
##          power_shot_power power_long_shots power_stamina
## power_shot_power           0
## power_long_shots           0           0
## power_stamina              0           0           0
## power_jumping        2.3e-56        9.8e-72           0
## physic                  4.3e-152      1.5e-14           0
## power_strength       6e-110       5.1e-17 1.6999999988919e-314
##          power_jumping physic power_strength
## power_shot_power

```

```

## power_long_shots
## power_stamina
## power_jumping          0
## physic                 0      0
## power_strength          0      0          0
##
## $sym
##                  power_shot_power power_long_shots power_stamina power_jumping
## power_shot_power 1
## power_long_shots +           1
## power_stamina          1
## power_jumping           .
## physic                 .
## power_strength          .
##
##                  physic power_strength
## power_shot_power
## power_long_shots
## power_stamina
## power_jumping
## physic          1
## power_strength +     1
## attr(,"legend")
## [1] 0 ' ' 0.3 '.' 0.6 ',' 0.8 '+' 0.9 '*' 0.95 'B' 1

```

*#long shots and shot power have high +ve correlation which makes sense
#power strength and physique have high +ve correlation too which makes sense.*

#####INTERCEPTIONS AND DEFENDING:

```
cor(fifa20$defending, fifa20$mentality_interceptions, method = "pearson",
    use = "complete.obs")
```

```
## [1] 0.961099
```

#Very high +ve correlation

#INTERCEPTIONS FOR DEFENSIVE MIDS, WING BACKS AND CENTRE BACKS:

```
unique(fifa20$team_position)
```

```
## [1] "RW"   "LW"   "CAM"  "GK"   "RCM"  "LCB"  "ST"   "CDM"  "LDM"  "RM"   "RCB"  "LCM"
## [13] "LM"   "CF"   "SUB"  "LB"   "LS"   "RB"   "RDM"  "RES"  "RAM"  "RS"   "RF"   "CM"
## [25] "CB"   "LF"   "LAM"  ""     "RWB"  "LWB"
```

#DEFENSIVE MIDS:

```
mid_int_def <- fifa20 %>% filter(team_position %in% c("CDM", "LDM", "RCM", "LCM", "CM"))
cor(mid_int_def$defending, mid_int_def$mentality_interceptions)
```

```
## [1] 0.9254998
```

#Very high positive correlation(.93)

#WING BACKS:

```

wb_int_def <- fifa20 %>% filter(team_position %in% c("LB", "RB", "LWB", "RWB"))
cor(wb_int_def$defending, wb_int_def$mentality_interceptions)

## [1] 0.9085694

#CENTRE BACKS:
cb_int_def <- fifa20 %>% filter(team_position %in% c("LCB", "RCB"))
cor(cb_int_def$defending, cb_int_def$mentality_interceptions)

## [1] 0.9202419

#mid way between defensive mids and wingbacks(.92)

#####POSITIONING AND FINISHING:
cor(fifa20$mentality_positioning, fifa20$attacking_finishing, method = "pearson",
  use = "complete.obs")

## [1] 0.895442

#very high positive correlation.

#####STANDING TACKLING AND DEFENDING
#STANDING TACKLING FOR DEFENSIVE MIDS, WING BACKS AND CENTRE BACKS:
unique(fifa20$team_position)

## [1] "RW"   "LW"   "CAM"  "GK"   "RCM"  "LCB"  "ST"   "CDM"  "LDM"  "RM"   "RCB"  "LCM"
## [13] "LM"   "CF"   "SUB"  "LB"   "LS"   "RB"   "RDM"  "RES"  "RAM"  "RS"   "RF"   "CM"
## [25] "CB"   "LF"   "LAM"  ""     "RWB"  "LWB"

#DEFENSIVE MIDS:
mid_sttack_def <- fifa20 %>% filter(team_position %in% c("CDM", "LDM", "RCM", "LCM", "CM"))
cor(mid_sttack_def$defending, mid_sttack_def$defending_standing_tackle)

## [1] 0.9506021

#Very high positive correlation(.95)

#WING BACKS:
wb_sttack_def <- fifa20 %>% filter(team_position %in% c("LB", "RB", "LWB", "RWB"))
cor(wb_sttack_def$defending, wb_sttack_def$defending_standing_tackle)

## [1] 0.9444266

#CENTRE BACKS:
cb_sttack_def <- fifa20 %>% filter(team_position %in% c("LCB", "RCB"))
cor(cb_sttack_def$defending, cb_sttack_def$defending_standing_tackle)

## [1] 0.9554827

```

```

#best(.955)

#####SLIDING TACKLING AND DEFENDING
#SLIDING TACKLING FOR DEFENSIVE MIDS, WING BACKS AND CENTRE BACKS:
unique(fifa20$team_position)

## [1] "RW"   "LW"   "CAM"  "GK"   "RCM"  "LCB"  "ST"   "CDM"  "LDM"  "RM"   "RCB"  "LCM"
## [13] "LM"   "CF"   "SUB"  "LB"   "LS"   "RB"   "RDM"  "RES"  "RAM"  "RS"   "RF"   "CM"
## [25] "CB"   "LF"   "LAM"  ""     "RWB"  "LWB"

#DEFENSIVE MIDS:
mid_sltack_def <- fifa20%>%filter(team_position %in% c("CDM", "LDM", "RCM", "LCM", "CM"))
cor(mid_sltack_def$defending, mid_sltack_def$defending_sliding_tackle)

## [1] 0.9170229

#Very high positive correlation (.917)

#WING BACKS:
wb_sltack_def <- fifa20%>%filter(team_position %in% c("LB", "RB", "LWB", "RWB"))
cor(wb_sltack_def$defending, wb_sltack_def$defending_sliding_tackle)

## [1] 0.912707

#CENTRE BACKS:
cb_sltack_def <- fifa20%>%filter(team_position %in% c("LCB", "RCB"))
cor(cb_sltack_def$defending, cb_sltack_def$defending_sliding_tackle)

## [1] 0.9137909

#similar to wingbacks (.91)

#####VISION AND PASSING:
cor(fifa20$passing, fifa20$mentality_vision, method = "pearson",
  use = "complete.obs")

## [1] 0.8792501

#very high +ve correlation.

#####FINISHING AND PENALTIES:
cor(fifa20$attacking_finishing, fifa20$mentality_penalties, method = "pearson",
  use = "complete.obs")

## [1] 0.8471015

```

```

#very high +ve correlation.

#####DIFFERENCES BETWEEN FORWARD LINE, MIDFIELD AND DEFENCE:
unique(fifa20$team_position)

## [1] "RW"   "LW"   "CAM"  "GK"   "RCM"  "LCB"  "ST"   "CDM"  "LDM"  "RM"   "RCB"  "LCM"
## [13] "LM"   "CF"   "SUB"  "LB"   "LS"   "RB"   "RDM"  "RES"  "RAM"  "RS"   "RF"   "CM"
## [25] "CB"   "LF"   "LAM"  ""     "RWB"  "LWB"

####PACE
#wingers:
fifa20 %>% filter(team_position %in% c("RW", "LW")) %>%
  summarise(avg_pace = mean(pace)) #78.6

## # A tibble: 1 x 1
##   avg_pace
##       <dbl>
## 1      78.6

#Strikers and forwards:
fifa20 %>% filter(team_position %in% c("ST", "CF", "LS", "RS", "RF", "LF")) %>%
  summarise(avg_pace = mean(pace)) #70.4

## # A tibble: 1 x 1
##   avg_pace
##       <dbl>
## 1      70.4

#Attacking mids:
fifa20 %>% filter(team_position %in% c("CAM", "RM", "LM", "RAM", "LAM")) %>%
  summarise(avg_pace = mean(pace)) #75.9

## # A tibble: 1 x 1
##   avg_pace
##       <dbl>
## 1      75.9

#Defensive mids:
fifa20 %>% filter(team_position %in% c("RCM", "CDM", "LDM", "LCM", "RDM", "CM")) %>%
  summarise(avg_pace = mean(pace)) #64.7

## # A tibble: 1 x 1
##   avg_pace
##       <dbl>
## 1      64.7

```

```
#Wing backs:  
fifa20 %>% filter(team_position %in% c("RB", "LB", "LWB", "RWB")) %>%  
  summarise(avg_pace = mean(pace)) #73.7
```

```
## # A tibble: 1 x 1  
##   avg_pace  
##     <dbl>  
## 1      73.7
```

```
#Centre backs:  
fifa20 %>% filter(team_position %in% c("RCB", "LCB", "CB")) %>%  
  summarise(avg_pace = mean(pace)) #57.5
```

```
## # A tibble: 1 x 1  
##   avg_pace  
##     <dbl>  
## 1      57.5
```

#Wingers are the fastest and cbs are the slowest.

```
####SHOOTING:  
#wingers:  
fifa20 %>% filter(team_position %in% c("RW", "LW")) %>%  
  summarise(avg_shooting = mean(shooting)) #65.7
```

```
## # A tibble: 1 x 1  
##   avg_shooting  
##     <dbl>  
## 1      65.7
```

```
#Strikers and forwards:  
fifa20 %>% filter(team_position %in% c("ST", "CF", "LS", "RS", "RF", "LF")) %>%  
  summarise(avg_shooting = mean(shooting)) #68.6
```

```
## # A tibble: 1 x 1  
##   avg_shooting  
##     <dbl>  
## 1      68.6
```

```
#Attacking mids:  
fifa20 %>% filter(team_position %in% c("CAM", "RM", "LM", "RAM", "LAM")) %>%  
  summarise(avg_shooting = mean(shooting)) #63.1
```

```
## # A tibble: 1 x 1  
##   avg_shooting  
##     <dbl>  
## 1      63.1
```

```
#Defensive mids:  
fifa20 %>% filter(team_position %in% c("RCM", "CDM", "LDM", "LCM", "RDM", "CM")) %>%  
  summarise(avg_shooting = mean(shooting)) #57.8
```

```
## # A tibble: 1 x 1  
##   avg_shooting  
##       <dbl>  
## 1      57.8
```

```
#Wing backs:  
fifa20 %>% filter(team_position %in% c("RB", "LB", "LWB", "RWB")) %>%  
  summarise(avg_shooting = mean(shooting)) #47.8
```

```
## # A tibble: 1 x 1  
##   avg_shooting  
##       <dbl>  
## 1      47.8
```

```
#Centre backs:  
fifa20 %>% filter(team_position %in% c("RCB", "LCB", "CB")) %>%  
  summarise(avg_shooting = mean(shooting)) #38.3
```

```
## # A tibble: 1 x 1  
##   avg_shooting  
##       <dbl>  
## 1      38.3
```

Strikers have the best shooting and centre backs have the worst shooting.

```
###PASSING  
#wingers:  
fifa20 %>% filter(team_position %in% c("RW", "LW")) %>%  
  summarise(avg_passing = mean(passing)) #63.8
```

```
## # A tibble: 1 x 1  
##   avg_passing  
##       <dbl>  
## 1      63.8
```

```
#Strikers and forwards:  
fifa20 %>% filter(team_position %in% c("ST", "CF", "LS", "RS", "RF", "LF")) %>%  
  summarise(avg_passing = mean(passing)) #57.6
```

```
## # A tibble: 1 x 1  
##   avg_passing  
##       <dbl>  
## 1      57.6
```

```
#Attacking mids:  
fifa20 %>% filter(team_position %in% c("CAM", "RM", "LM", "RAM", "LAM")) %>%  
  summarise(avg_passing = mean(passing)) #64.9
```

```
## # A tibble: 1 x 1  
##   avg_passing  
##       <dbl>  
## 1      64.9
```

```
#Defensive mids:  
fifa20 %>% filter(team_position %in% c("RCM", "CDM", "LDM", "LCM", "RDM", "CM")) %>%  
  summarise(avg_passing = mean(passing)) #65.5
```

```
## # A tibble: 1 x 1  
##   avg_passing  
##       <dbl>  
## 1      65.5
```

```
#Wing backs:  
fifa20 %>% filter(team_position %in% c("RB", "LB", "LWB", "RWB")) %>%  
  summarise(avg_passing = mean(passing)) #60.6
```

```
## # A tibble: 1 x 1  
##   avg_passing  
##       <dbl>  
## 1      60.6
```

```
#Centre backs:  
fifa20 %>% filter(team_position %in% c("RCB", "LCB", "CB")) %>%  
  summarise(avg_passing = mean(passing)) #50.8
```

```
## # A tibble: 1 x 1  
##   avg_passing  
##       <dbl>  
## 1      50.8
```

#defensive mids have the best passing and centre backs have the worst passing.

```
####Dribbling  
fifa20 %>% filter(team_position %in% c("RW", "LW")) %>%  
  summarise(avg_dribbling = mean(dribbling)) #72.3
```

```
## # A tibble: 1 x 1  
##   avg_dribbling  
##       <dbl>  
## 1      72.3
```

```

#Strikers and forwards:
fifa20 %>% filter(team_position %in% c("ST", "CF", "LS", "RS", "RF", "LF")) %>%
  summarise(avg_dribbling = mean(dribbling)) #67.7

## # A tibble: 1 x 1
##   avg_dribbling
##       <dbl>
## 1           67.7

#Attacking mids:
fifa20 %>% filter(team_position %in% c("CAM", "RM", "LM", "RAM", "LAM")) %>%
  summarise(avg_dribbling = mean(dribbling)) #71.1

## # A tibble: 1 x 1
##   avg_dribbling
##       <dbl>
## 1           71.1

#Defensive mids:
fifa20 %>% filter(team_position %in% c("RCM", "CDM", "LDM", "LCM", "RDM", "CM")) %>%
  summarise(avg_dribbling = mean(dribbling)) #67.3

## # A tibble: 1 x 1
##   avg_dribbling
##       <dbl>
## 1           67.3

#Wing backs:
fifa20 %>% filter(team_position %in% c("RB", "LB", "LWB", "RWB")) %>%
  summarise(avg_dribbling = mean(dribbling)) #65.3

## # A tibble: 1 x 1
##   avg_dribbling
##       <dbl>
## 1           65.3

#Centre backs:
fifa20 %>% filter(team_position %in% c("RCB", "LCB", "CB")) %>%
  summarise(avg_dribbling = mean(dribbling)) #51.9

## # A tibble: 1 x 1
##   avg_dribbling
##       <dbl>
## 1           51.9

#wingers have the best dribbling and centre backs have the worst dribbling.

####DEFENDING
fifa20 %>% filter(team_position %in% c("RW", "LW")) %>%
  summarise(avg_defending = mean(defending)) #37.0

```

```

## # A tibble: 1 x 1
##   avg_defending
##       <dbl>
## 1      37.0

#Strikers and forwards:
fifa20 %>% filter(team_position %in% c("ST", "CF", "LS", "RS", "RF", "LF")) %>%
  summarise(avg_defending = mean(defending)) #33.4


## # A tibble: 1 x 1
##   avg_defending
##       <dbl>
## 1      33.4

#Attacking mids:
fifa20 %>% filter(team_position %in% c("CAM", "RM", "LM", "RAM", "LAM")) %>%
  summarise(avg_defending = mean(defending)) #42.6


## # A tibble: 1 x 1
##   avg_defending
##       <dbl>
## 1      42.6

#Defensive mids:
fifa20 %>% filter(team_position %in% c("RCM", "CDM", "LDM", "LCM", "RDM", "CM")) %>%
  summarise(avg_defending = mean(defending)) #61.8


## # A tibble: 1 x 1
##   avg_defending
##       <dbl>
## 1      61.8

#Wing backs:
fifa20 %>% filter(team_position %in% c("RB", "LB", "LWB", "RWB")) %>%
  summarise(avg_defending = mean(defending)) #64.3


## # A tibble: 1 x 1
##   avg_defending
##       <dbl>
## 1      64.3

#Centre backs:
fifa20 %>% filter(team_position %in% c("RCB", "LCB", "CB")) %>%
  summarise(avg_defending = mean(defending)) #68.4


## # A tibble: 1 x 1
##   avg_defending
##       <dbl>
## 1      68.4

```

```
#Centre backs have the best defense and forwards/strikers have the worst defense.
```

```
####PHYSIC
#wingers
fifa20 %>% filter(team_position %in% c("RW", "LW")) %>%
  summarise(avg_physic = mean(physic)) #61.5
```

```
## # A tibble: 1 x 1
##   avg_physic
##       <dbl>
## 1      61.5
```

```
#Strikers and forwards:
fifa20 %>% filter(team_position %in% c("ST", "CF", "LS", "RS", "RF", "LF")) %>%
  summarise(avg_physic = mean(physic)) #68.9
```

```
## # A tibble: 1 x 1
##   avg_physic
##       <dbl>
## 1      68.9
```

```
#Attacking mids:
fifa20 %>% filter(team_position %in% c("CAM", "RM", "LM", "RAM", "LAM")) %>%
  summarise(avg_physic = mean(physic)) #61.6
```

```
## # A tibble: 1 x 1
##   avg_physic
##       <dbl>
## 1      61.6
```

```
#Defensive mids:
fifa20 %>% filter(team_position %in% c("RCM", "CDM", "LDM", "LCM", "RDM", "CM")) %>%
  summarise(avg_physic = mean(physic)) #69.4
```

```
## # A tibble: 1 x 1
##   avg_physic
##       <dbl>
## 1      69.4
```

```
#Wing backs:
fifa20 %>% filter(team_position %in% c("RB", "LB", "LWB", "RWB")) %>%
  summarise(avg_physic = mean(physic)) #68.5
```

```
## # A tibble: 1 x 1
##   avg_physic
##       <dbl>
## 1      68.5
```

```
#Centre backs:
fifa20 %>% filter(team_position %in% c("RCB", "LCB", "CB")) %>%
  summarise(avg_physic = mean(physic)) #73.6
```

```
## # A tibble: 1 x 1
##   avg_physic
##   <dbl>
## 1 73.6
```

#Centre backs have the best physic and wingers/attacking mids have the worst physic.

```
####Very good players who are slow:
fifa20 %>% filter(overall >= 85) %>%
  arrange(pace) %>%
  select(short_name, club, team_position, overall, pace)
```

```
## # A tibble: 108 x 5
##   short_name     club      team_position overall  pace
##   <chr>       <chr>      <chr>        <int> <int>
## 1 Parejo       Valencia CF    RCM          86    41
## 2 Sergio Busquets FC Barcelona CDM          89    42
## 3 T. Kroos     Real Madrid   LCM          88    45
## 4 M. Hummels   Borussia Dortmund LCB          87    51
## 5 J. Rodríguez Real Madrid   SUB          85    55
## 6 Piqué        FC Barcelona  RCB          88    56
## 7 Z. Ibrahimović LA Galaxy   ST           85    56
## 8 L. Bonucci   Juventus     RCB          86    58
## 9 D. Godán     Inter        RCB          88    60
## 10 Casemiro    Real Madrid   CDM          87    62
## # ... with 98 more rows
```

#Dani Parejo, Busquets, Kroos
####very fast players who are bad overall

```
fifa20 %>% filter(overall <= 70) %>%
  arrange(desc(pace)) %>%
  select(short_name, club, team_position, overall, pace)
```

```
## # A tibble: 13,522 x 5
##   short_name   club      team_position overall  pace
##   <chr>       <chr>      <chr>        <int> <int>
## 1 K. Nagai     FC Tokyo    LS            69    95
## 2 A. Chalá     Deportivo Toluca LWB           66    95
## 3 K. Manneh    FC Cincinnati SUB           68    94
## 4 C. Bărbulescu Universitatea Craiova RW           68    94
## 5 J. Aguirre   Rosario Central SUB           68    94
## 6 M. Bolly     Molde FK    SUB           65    94
## 7 E. List      Gillingham  SUB           62    94
## 8 A. Kiwomya   Doncaster Rovers SUB           61    94
## 9 I. Openda    Club Brugge KV    SUB           70    93
```

```
## 10 Y. Niakat  Al Wehda ST 70 93
## # ... with 13,512 more rows
```

```
####good players who are physically weak:
fifa20 %>% filter(overall >= 85) %>%
  arrange(physic)%>%
  select(short_name, club, team_position, overall, physic)
```

```
## # A tibble: 108 x 5
##   short_name   club      team_position overall physic
##   <chr>        <chr>      <chr>          <int>  <int>
## 1 L. Insigne Napoli       LS            87    47
## 2 D. Mertens Napoli       SUB           87    53
## 3 A. G mez Atalanta     CAM           85    55
## 4 R. Sterling Manchester City LW            88    57
## 5 David Silva Manchester City LCM           88    57
## 6 Neymar Jr  Paris Saint-Germain CAM           92    58
## 7 Isco        Real Madrid    SUB           86    60
## 8 C. Eriksen Tottenham Hotspur RM            88    63
## 9 Thiago      FC Bayern M nchen CDM           87    63
## 10 J. Rodr guez Real Madrid    SUB           85    63
## # ... with 98 more rows
```

```
#Insigne, Mertens
####strong players who are bad overall:
fifa20 %>% filter(overall <= 70) %>%
  arrange(desc(physic))%>%
  select(short_name, club, team_position, overall, physic)
```

```
## # A tibble: 13,522 x 5
##   short_name   club      team_position overall physic
##   <chr>        <chr>      <chr>          <int>  <int>
## 1 M. Bostwick Lincoln City RCB            68    88
## 2 B. Fofana   Gaz Metan Media   SUB           68    88
## 3 A. Coly     Kristiansund BK RB            65    88
## 4 Fali        C diz CF      SUB           70    87
## 5 J. Marquis  Portsmouth   ST             69    87
## 6 D. Hediger   FC Thun     SUB           68    87
## 7 A. Nsiala   Ipswich Town SUB           69    86
## 8 F. Di Tacchio US Salernitana 1919 RCM           67    86
## 9 S. Sarr     N mes Olympique SUB           66    86
## 10 A. Ogogo   Bristol Rovers SUB           65    86
## # ... with 13,512 more rows
```

```
#####centre backs attack the goal during set pieces to chip in with headers
#which centre backs are good at attacking the goal with their heads?
fifa20 %>% filter(team_position %in% c("LCB", "RCB", "CB"))%>%
  arrange(desc(attacking_heading_accuracy))%>%
  select(short_name, club, overall, attacking_heading_accuracy)
```

```
## # A tibble: 1,420 x 4
##   short_name   club      overall attacking_heading_accuracy
```

```

##      <chr>      <chr>      <int>      <int>
## 1 Sergio Ramos Real Madrid      89       92
## 2 F. Fazio     Roma          80       90
## 3 S. SanÃ©     FC Schalke 04    80       89
## 4 D. GodÃ¡n    Inter         88       88
## 5 M. Hummels   Borussia Dortmund 87       87
## 6 H. Maguire   Manchester United 82       87
## 7 V. van Dijk  Liverpool      90       86
## 8 A. Laporte   Manchester City   87       86
## 9 N. SÃ¼le     FC Bayern MÃ¼nchen 85       86
## 10 W. Orban    RB Leipzig      81       86
## # ... with 1,410 more rows

```

#Sergio Ramos expected.

```

####wing backs overlap with the wingers while attacking to deliver
#key crosses into the box
#which wing/full backs are the best crossers:
fifa20 %>% filter(team_position %in% c("RB", "RWB", "LWB", "LB"))%>%
  arrange(desc(attacking_crossing))%>%
  select(short_name, club, overall, attacking_crossing)

```

```

## # A tibble: 1,236 x 4
##      short_name      club      overall  attacking_crossing
##      <chr>        <chr>      <int>      <int>
## 1 J. Kimmich  FC Bayern MÃ¼nchen      86       91
## 2 A. Kolarov   Roma          82       90
## 3 A. Robertson Liverpool      85       89
## 4 Marcelo     Real Madrid      85       89
## 5 Alex Telles  FC Porto       84       89
## 6 T. Alexander-Arnold Liverpool      83       88
## 7 Jordi Alba   FC Barcelona     87       87
## 8 M. Plattenhardt Hertha BSC      77       87
## 9 R. RodrÃguez Milan          80       86
## 10 P. KadeÅ¡bek TSG 1899 Hoffenheim 80       86
## # ... with 1,226 more rows

```

#Joshua Kimmich expected.

```

####wingers who are very good crossers:
fifa20 %>% filter(team_position %in% c("RW", "LW"))%>%
  arrange(desc(attacking_crossing))%>%
  select(short_name, club, overall, attacking_crossing)

```

```

## # A tibble: 323 x 4
##      short_name      club      overall  attacking_crossing
##      <chr>        <chr>      <int>      <int>
## 1 L. Messi     FC Barcelona     94       88
## 2 M. AcuÃ±a     Sporting CP      81       87
## 3 A. Di MarÃa  Paris Saint-Germain 86       86
## 4 Bernardo Silva Manchester City   87       85

```

```

## 5 Cristiano Ronaldo Juventus 93 84
## 6 D. Payet Olympique de Marseille 81 84
## 7 P. Griezmann Brighton & Hove Albion 78 84
## 8 A. Griezmann FC Barcelona 89 83
## 9 M. Depay Olympique Lyonnais 84 83
## 10 P. Dybala Juventus 88 82
## # ... with 313 more rows

```

####Target men are strikers who are tall, strong and good at heading.

#Best target men:

```

fifa20 %>% filter(height_cm >= 180, physic >= 75) %>%
  filter(team_position %in% c("ST", "CF", "LS", "RS", "RF", "LF", "SUB")) %>%
  arrange(desc(attacking_heading_accuracy)) %>%
  select(short_name, club, overall, attacking_heading_accuracy)

```

```

## # A tibble: 945 x 4
##   short_name     club      overall attacking_heading_accuracy
##   <chr>        <chr>      <int>                <int>
## 1 B. Dost      Sporting CP    82                  93
## 2 L. de Jong   Sevilla FC    82                  93
## 3 L. Pavoletti Cagliari     78                  93
## 4 O. Giroud    Chelsea       82                  91
## 5 M. Smith     Millwall      68                  91
## 6 M. MandÅžukÅ‡ Juventus    84                  90
## 7 G. PellÅ“    Shandong Luneng TaiShan FC  77                  90
## 8 E. Cavani    Paris Saint-Germain 88                  89
## 9 Javi MartÃ±ez FC Bayern MÃ¼nchen 84                  87
## 10 W. Weghorst VfL Wolfsburg  81                  87
## # ... with 935 more rows

```

####Very good players who are poor at penalties: excluding fbs, wbs, cbs, gks and subs
since they dont generally take penalties

```

fifa20 %>% filter(overall >= 85) %>%
  filter(!team_position %in% c("RB", "RWB", "LWB", "LB", "GK", "LCB", "RCB", "CB", "SUB")) %>%
  arrange(mentality_penalties) %>%
  select(short_name, club, overall, mentality_penalties)

```

```

## # A tibble: 55 x 4
##   short_name     club      overall mentality_penalties
##   <chr>        <chr>      <int>                <int>
## 1 F. de Jong   FC Barcelona 85                  45
## 2 Allan       Napoli      85                  53
## 3 N. KantÅ©   Chelsea     89                  54
## 4 SaÃ±ol      AtlÃ©tico Madrid 85                  55
## 5 S. MilinkoviÄ‡-SaviÄ‡ Lazio     85                  56
## 6 Koke        AtlÃ©tico Madrid 85                  59
## 7 Sergio Busquets FC Barcelona 89                  60
## 8 Fernandinho Manchester City 87                  61
## 9 M. Verratti  Paris Saint-Germain 86                  64
## 10 Casemiro   Real Madrid  87                  66
## # ... with 45 more rows

```

```
#Frenkie De Jong
```

```
####Wingers who cut in and shoot to score goals: They are false wingers  
#ie right wingers with preferred foot as left and have good shooting and vice versa  
fifa20 %>% filter(team_position=="RW", preferred_foot=="Left")%>%  
  arrange(desc(shooting))%>%  
  select(short_name, club, shooting, overall)
```

```
## # A tibble: 65 x 4  
##   short_name     club      shooting  overall  
##   <chr>        <chr>      <int>    <int>  
## 1 L. Messi     FC Barcelona       92      94  
## 2 M. Salah     Liverpool        86      90  
## 3 Hulk         Shanghai SIPG FC    85      80  
## 4 P. Dybala    Juventus        82      88  
## 5 Anderson Talisca Guangzhou Evergrande Taobao FC 82      83  
## 6 C. Vela       Los Angeles FC    82      83  
## 7 F. Thauvin   Olympique de Marseille 80      83  
## 8 S. Berghuis Feyenoord          77      80  
## 9 Munir        Sevilla FC        77      79  
## 10 D. Berardi  Sassuolo          77      78  
## # ... with 55 more rows
```

```
#Leo Messi, Salah
```

```
fifa20 %>% filter(team_position=="LW", preferred_foot=="Right")%>%  
  arrange(desc(shooting))%>%  
  select(short_name, club, shooting, overall)
```

```
## # A tibble: 130 x 4  
##   short_name     club      shooting  overall  
##   <chr>        <chr>      <int>    <int>  
## 1 Cristiano Ronaldo Juventus       93      93  
## 2 E. Hazard     Real Madrid     83      91  
## 3 S. Man  o    Liverpool        83      88  
## 4 Morales       Levante UD      81      82  
## 5 E. Zahavi    Guangzhou R&F FC  81      81  
## 6 M. Depay      Olympique Lyonnais 80      84  
## 7 L. Trossard   Brighton & Hove Albion 80      79  
## 8 R. Sterling   Manchester City   79      88  
## 9 S. El Shaarawy Shanghai Greenland Shenhua FC 78      82  
## 10 M. Rashica   SV Werder Bremen  78      79  
## # ... with 120 more rows
```

```
#Ronaldo, Hazard
```

```
#####COMPARISON BETWEEN MESSI AND RONALDO: MAJOR STATS OUT OF 100
```

```
#BASIC STATS:
```

```
messi_ronaldo_basic <- fifa20 %>%
```

```

filter(short_name == "L. Messi" | short_name == "Cristiano Ronaldo") %>%
  select(age, height_cm, weight_kg, overall, value_eur, wage_eur, release_clause_eur,
         international_reputation,
         weak_foot, skill_moves)
messi_ronaldo_basic

## # A tibble: 2 x 10
##   age height_cm weight_kg overall value_eur wage_eur release_clause_eur
##   <int>      <int>     <int>    <int>    <int>      <int>
## 1    32        170       72      94  95500000  565000    195800000
## 2    34        187       83      93  58500000  405000    96500000
## # ... with 3 more variables: international_reputation <int>, weak_foot <int>,
## #   skill_moves <int>

#6 MAIN ATTRIBUTES:
messi_ronaldo_6main <- fifa20 %>%
  filter(short_name == "L. Messi" | short_name == "Cristiano Ronaldo") %>%
  select(pace, shooting, passing, dribbling, defending, physic)

#ATTACK ATTRIBUTES:
messi_ronaldo_attack <- fifa20 %>%
  filter(short_name == "L. Messi" | short_name == "Cristiano Ronaldo") %>%
  select(attacking_crossing, attacking_finishing, attacking_heading_accuracy,
         attacking_short_passing, attacking_volleys)

#MOVEMENT ATTRIBUTES:
messi_ronaldo_move <- fifa20 %>%
  filter(short_name == "L. Messi" | short_name == "Cristiano Ronaldo") %>%
  select(movement_acceleration, movement_sprint_speed, movement_agility,
         movement_reactions, movement_balance)

#POWER ATTRIBUTES:
messi_ronaldo_power <- fifa20 %>%
  filter(short_name == "L. Messi" | short_name == "Cristiano Ronaldo") %>%
  select(power_shot_power, power_jumping, power_stamina,
         power_strength, power_long_shots)

#DEFENSE ATTRIBUTES:
messi_ronaldo_def <- fifa20 %>%
  filter(short_name == "L. Messi" | short_name == "Cristiano Ronaldo") %>%
  select(defending_marking, defending_standing_tackle, defending_sliding_tackle)

#MENTALITY ATTRIBUTES:
messi_ronaldo_ment <- fifa20 %>%
  filter(short_name == "L. Messi" | short_name == "Cristiano Ronaldo") %>%
  select(mentality_aggression, mentality_interceptions, mentality_positioning,
         mentality_penalties, mentality_vision, mentality_composure)

#Difference in basic stats:
diff_basic <- data.frame(diff(as.matrix(messi_ronaldo_basic)))
diff_basic

##   age height_cm weight_kg overall value_eur wage_eur release_clause_eur

```

```

## 1 2 17 11 -1 -37000000 -160000 -99300000
##   international_reputation weak_foot skill_moves
## 1 0 0 1

#Difference in 6 main attributes:
diff_6main <- data.frame(diff(as.matrix(messi_ronaldo_6main)))
diff_6main

##   pace shooting passing dribbling defending physic
## 1 3 1 -10 -7 -4 12

#Difference in attack attributes:
diff_attack <- data.frame(diff(as.matrix(messi_ronaldo_attack)))
diff_attack

##   attacking_crossing attacking_finishing attacking_heading_accuracy
## 1 -4 -1 19
##   attacking_short_passing attacking_volleys
## 1 -9 -1

#Difference in movement attributes:
diff_move <- data.frame(diff(as.matrix(messi_ronaldo_move)))
diff_move

##   movement_acceleration movement_sprint_speed movement_agility
## 1 -2 7 -6
##   movement_reactions movement_balance
## 1 1 -24

#Difference in defence attributes:
diff_def <- data.frame(diff(as.matrix(messi_ronaldo_def)))
diff_def

##   defending_marking defending_standing_tackle defending_sliding_tackle
## 1 -5 -5 -2

#difference in power attributes:
diff_power <- data.frame(diff(as.matrix(messi_ronaldo_power)))
diff_power

##   power_shot_power power_jumping power_stamina power_strength power_long_shots
## 1 9 27 10 10 -1

#difference in mentality attributes:
diff_ment <- data.frame(diff(as.matrix(messi_ronaldo_ment)))
diff_ment

##   mentality_aggression mentality_interceptions mentality_positioning
## 1 15 -11 1
##   mentality_penalties mentality_vision mentality_composure
## 1 10 -12 -1

```

```

#These are the differences between messi and ronaldo.
#Who has got the best overall stats?
total_diff <- bind_cols(diff_6main, diff_attack, diff_def, diff_ment,
                        diff_move, diff_power)
total_diff

##   pace shooting passing dribbling defending physic attacking_crossing
## 1      3          1     -10       -7      -4     12           -4
##   attacking_finishing attacking_heading_accuracy attacking_short_passing
## 1                  -1                      19           -9
##   attacking_volleys defending_marking defending_standing_tackle
## 1                  -1                  -5           -5
##   defending_sliding_tackle mentality_aggression mentality_interceptions
## 1                  -2                   15          -11
##   mentality_positioning mentality_penalties mentality_vision
## 1                  1                   10          -12
##   mentality_composure movement_acceleration movement_sprint_speed
## 1                  -1                  -2            7
##   movement_agility movement_reactions movement_balance power_shot_power
## 1                  -6                   1          -24            9
##   power_jumping power_stamina power_strength power_long_shots
## 1                 27                  10           10          -1

#adding values across columns:
sum <- 0
for (i in 1:length(total_diff))
{
  sum <- sum + total_diff[[i]]
}
print(sum)

## [1] 20

#####OVERALL RONALDO LEADS MESSI BY 20 POINTS IN NUMERIC STATS.

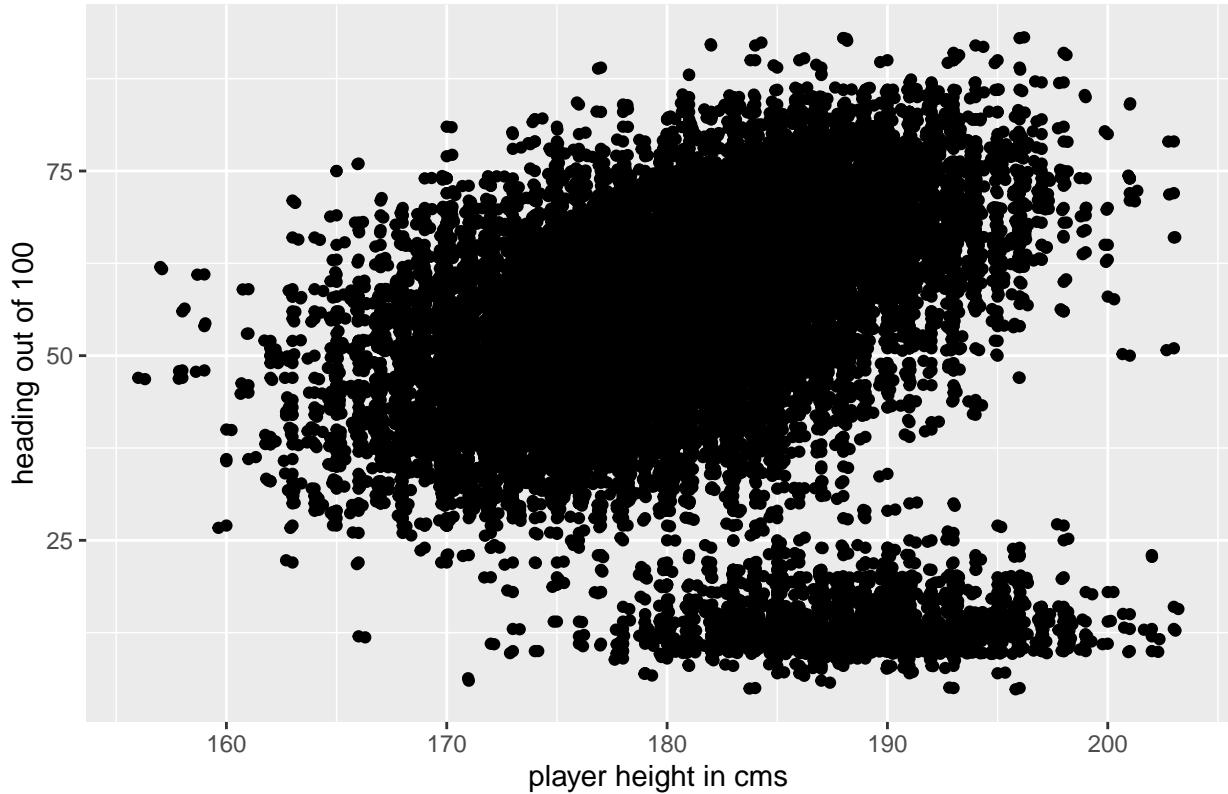
####RANGE OF PLAYER OVERALL:
summary(fifa20$overall)

##      Min. 1st Qu. Median    Mean 3rd Qu.    Max.
## 48.00  62.00  66.00  66.24  71.00  94.00

#####ABSTRACT HYPOTHESIS EXTENSION:
#Tall players are good at heading
#Exclude goalkeepers in the comparison
fifa20_sans_gks <- fifa20 %>% filter(team_position != "GK")
fifa20_sans_gks %>% ggplot(aes(height_cm, attacking_heading_accuracy)) + geom_point() +
  geom_jitter() + labs(x = "player height in cms", y = "heading out of 100",
    title = "height vs heading")

```

height vs heading



```
cor(fifa20_sans_gks$height_cm, fifa20_sans_gks$attacking_heading_accuracy,  
method = "spearman", use = "complete.obs")
```

```
## [1] 0.2990911
```

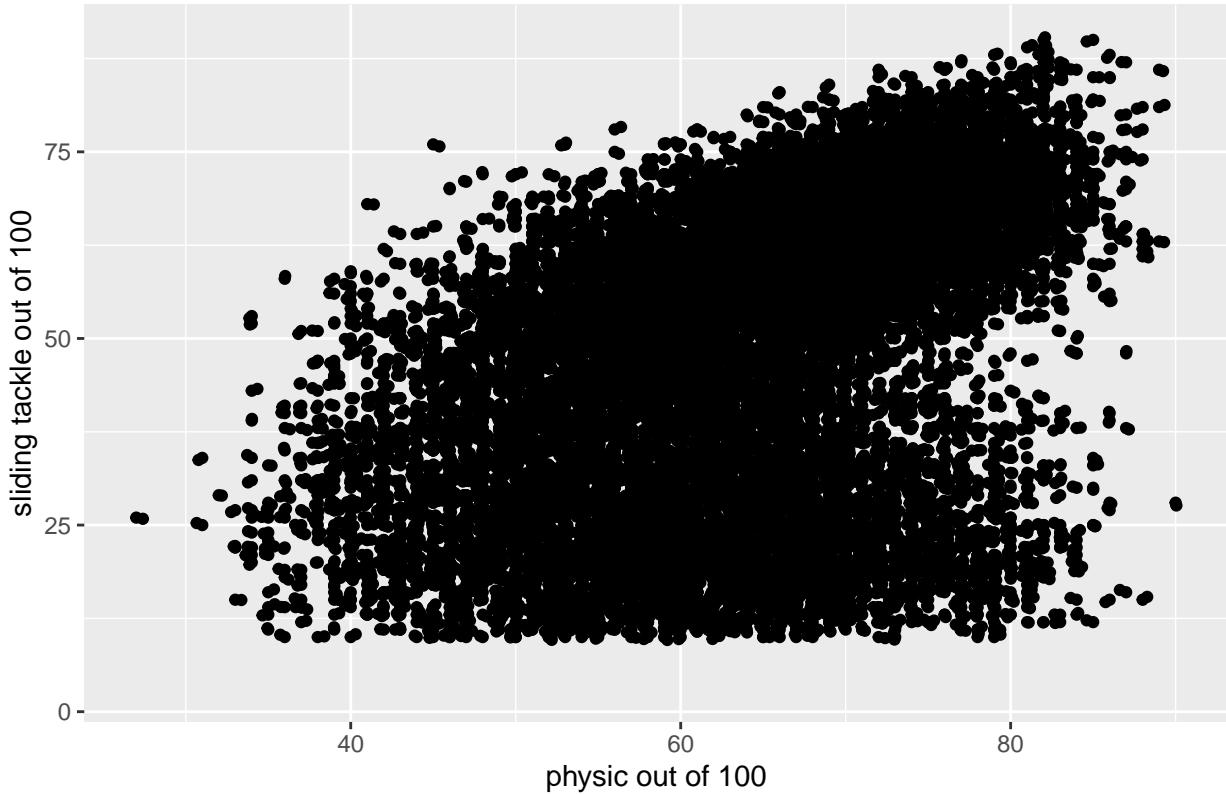
```
#correlation is higher without GKS since they dont really head the ball  
#in a match.
```

```
#Strong players are good at tackling:  
#excluding GKS  
fifa20_sans_gks %>% ggplot(aes(physic, defending_sliding_tackle)) + geom_point() +  
  geom_jitter() + labs(x = "physic out of 100", y = "sliding tackle out of 100",  
    title = "physic vs sliding tackling")
```

```
## Warning: Removed 1374 rows containing missing values (geom_point).
```

```
## Warning: Removed 1374 rows containing missing values (geom_point).
```

physic vs sliding tackling



```
cor(fifa20_sans_gks$physic, fifa20_sans_gks$defending_sliding_tackle,
    method = "spearman", use = "complete.obs")
```

```
## [1] 0.4983606
```

#correlation is higher without GKS since they dont really tackle in matches either.

```
fifa20 <- fread('D:/NEU/Spring 2020/SML/Project/Datasets/players_20.csv')
class(fifa20)
```

```
## [1] "data.table" "data.frame"
```

```
#View(fifa20)
#fifa20 as a tibble
fifa20 <- as_tibble(fifa20)
```

####ANALYSIS OF PLAYER POSITION VS VALUE:
`unique(fifa20$team_position)`

```
## [1] "RW"   "LW"   "CAM"  "GK"   "RCM"  "LCB"  "ST"   "CDM"  "LDM"  "RM"   "RCB"  "LCM"
## [13] "LM"   "CF"   "SUB"  "LB"   "LS"   "RB"   "RDM"  "RES"  "RAM"  "RS"   "RF"   "CM"
## [25] "CB"   "LF"   "LAM"  ""     "RWB"  "LWB"
```

```

fifa20_dup <- fifa20
unique(fifa20_dup$team_position)

## [1] "RW"   "LW"   "CAM"  "GK"   "RCM"  "LCB"  "ST"   "CDM"  "LDM"  "RM"   "RCB"  "LCM"
## [13] "LM"   "CF"   "SUB"  "LB"   "LS"   "RB"   "RDM"  "RES"  "RAM"  "RS"   "RF"   "CM"
## [25] "CB"   "LF"   "LAM"  ""     "RWB"  "LWB"

fifa20_dup$team_position[fifa20_dup$team_position %in%
                           c("RW", "LW", "ST", "CF", "LS", "RS", "RF", "LF")] <- "Attack"

fifa20_dup$team_position[fifa20_dup$team_position %in%
                           c("LCB", "RCB", "LB", "RB", "CB", "LWB", "RWB")] <- "Defense"

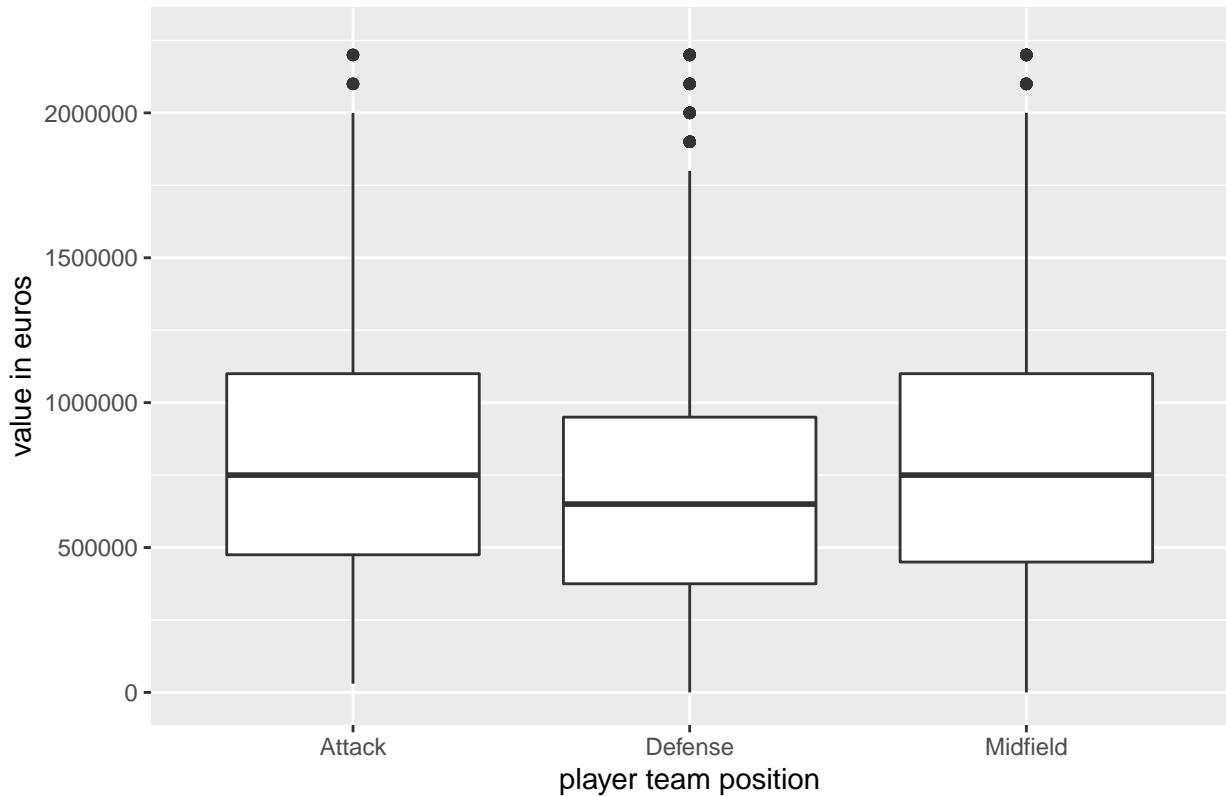
fifa20_dup$team_position[fifa20_dup$team_position %in%
                           c("CAM", "RCM", "CDM", "LDM", "RM", "LCM", "LM",
                             "RDM", "RAM", "CM", "LAM")] <- "Midfield"

#POSITION VS VALUE
fifa20_dup %>% filter(team_position %in% c("Attack", "Midfield", "Defense")) %>%
  ggplot(aes(team_position, value_eur)) + geom_boxplot() + ylim(0, 2250000) +
  labs(x = "player team position", y = "value in euros", title = "position vs value")

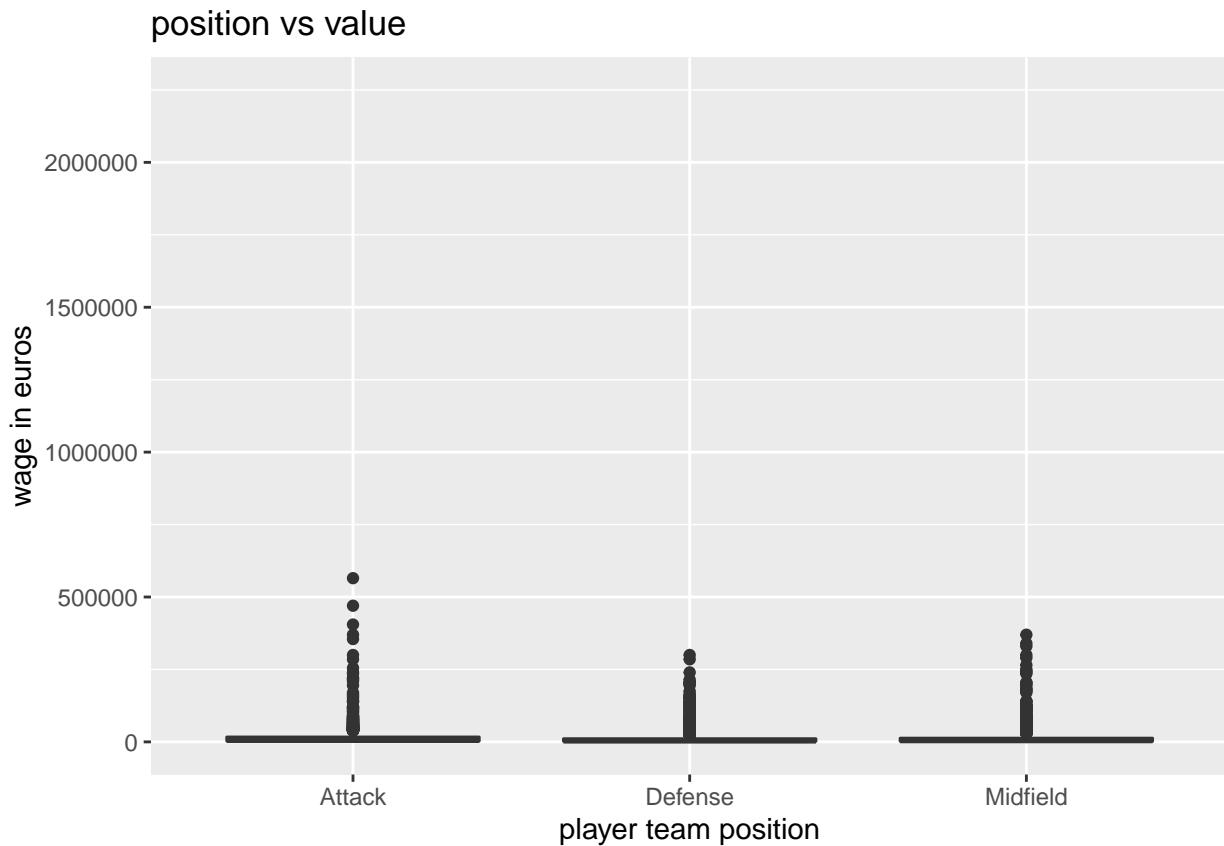
```

Warning: Removed 2448 rows containing non-finite values (stat_boxplot).

position vs value



```
#POSITION VS WAGE
fifa20_dup %>% filter(team_position %in% c("Attack", "Midfield", "Defense")) %>%
  ggplot(aes(team_position, wage_eur)) + geom_boxplot() + ylim(0, 2250000) +
  labs(x = "player team position", y = "wage in euros", title = "position vs value")
```



```
#ATTACK:
fifa20 %>% filter(team_position %in% c("RW", "LW", "ST", "CF", "LS", "RS", "RF", "LF")) %>%
  summarise(avval = mean(value_eur))
```

```
## # A tibble: 1 x 1
##       avval
##   <dbl>
## 1 5420626.
```

```
#MIDFIELD:
fifa20 %>% filter(team_position %in% c("CAM", "RCM", "CDM", "LDM", "RM", "LCM", "LM",
                                         "RDM", "RAM", "CM", "LAM")) %>%
  summarise(avval = mean(value_eur))
```

```
## # A tibble: 1 x 1
##       avval
##   <dbl>
## 1 4112979.
```

```

#DEFENSE :
fifa20 %>% filter(team_position %in% c("LCB", "RCB", "LB", "RB", "CB", "LWB", "RWB")) %>%
  summarise(avval = mean(value_eur))

## # A tibble: 1 x 1
##       avval
##   <dbl>
## 1 2953213.

#####ANALYSIS OF PLAYER POSITION VS WAGE:
unique(fifa20$team_position)

## [1] "RW"    "LW"    "CAM"   "GK"    "RCM"   "LCB"   "ST"    "CDM"   "LDM"   "RM"    "RCB"   "LCM"
## [13] "LM"    "CF"    "SUB"   "LB"    "LS"    "RB"    "RDM"   "RES"   "RAM"   "RS"    "RF"    "CM"
## [25] "CB"    "LF"    "LAM"   ""      "RWB"   "LWB"

#ATTACK:
fifa20 %>% filter(team_position %in% c("RW", "LW", "ST", "CF", "LS", "RS", "RF", "LF")) %>%
  summarise(avwag = mean(wage_eur))

## # A tibble: 1 x 1
##       avwag
##   <dbl>
## 1 17597.

#MIDFIELD:
fifa20 %>% filter(team_position %in% c("CAM", "RCM", "CDM", "LDM", "RM", "LCM", "LM",
                                         "RDM", "RAM", "CM", "LAM")) %>%
  summarise(avwag = mean(wage_eur))

## # A tibble: 1 x 1
##       avwag
##   <dbl>
## 1 12802.

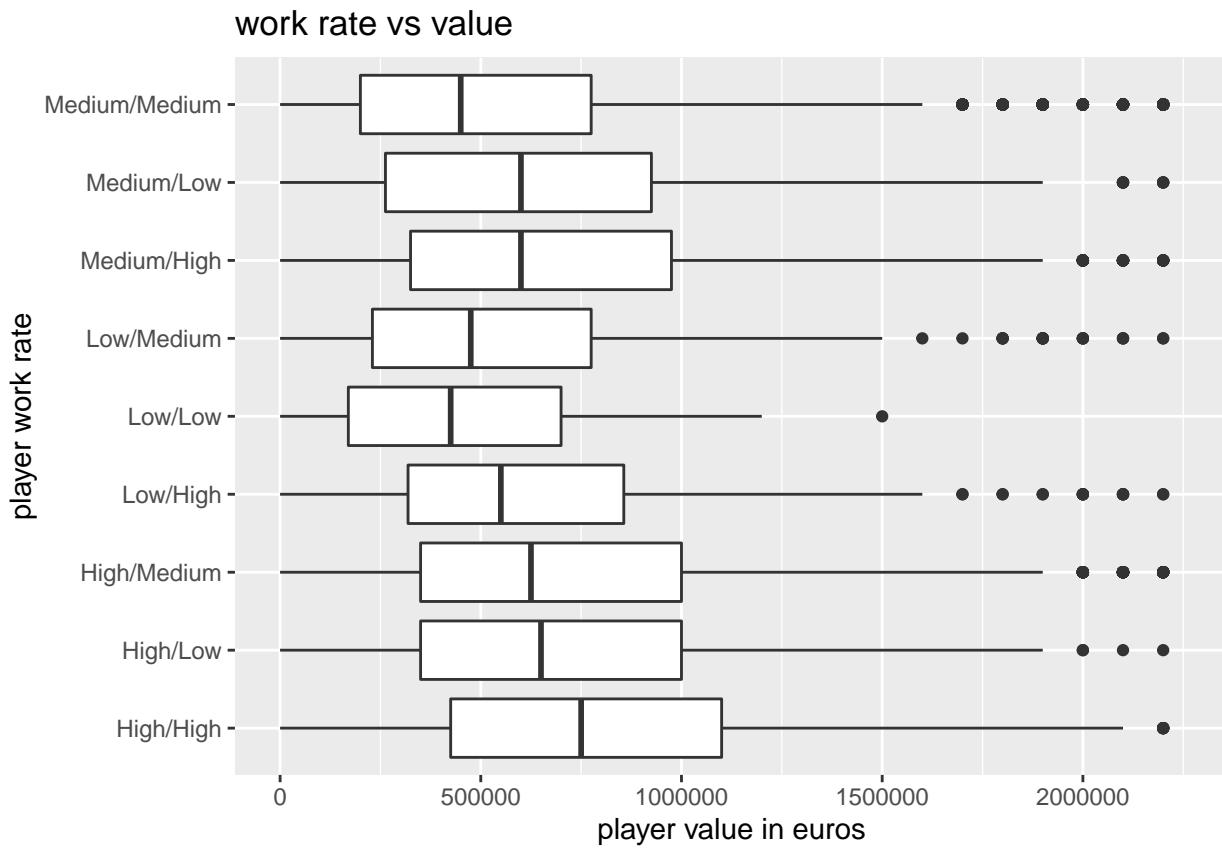
#DEFENSE :
fifa20 %>% filter(team_position %in% c("LCB", "RCB", "LB", "RB", "CB", "LWB", "RWB")) %>%
  summarise(avwag = mean(wage_eur))

## # A tibble: 1 x 1
##       avwag
##   <dbl>
## 1 11210.

#####WORK RATE VS VALUE:
fifa20 %>% ggplot(aes(work_rate, value_eur)) + geom_boxplot() + ylim(0, 2250000) + coord_flip() +
  labs(x = "player work rate", y = "player value in euros", title = "work rate vs value")

## Warning: Removed 4388 rows containing non-finite values (stat_boxplot).

```



```
##### REGEX INJURY PRONE:
```

```
fifa20 <- fifa20 %>% mutate(injury_prone = str_detect(fifa20$player_traits, "Injury Prone"))

fifa20$injury_prone[fifa20$injury_prone == "TRUE"] <- 1
fifa20 %>% group_by(work_rate, injury_prone) %>% count()
```

```
## # A tibble: 18 x 3
## # Groups:   work_rate, injury_prone [18]
##   work_rate      injury_prone     n
##   <chr>          <dbl> <int>
## 1 High/High        0     909
## 2 High/High        1     77
## 3 High/Low         0    680
## 4 High/Low         1     4
## 5 High/Medium      0   3132
## 6 High/Medium      1     84
## 7 Low/High          0   373
## 8 Low/High          1     44
## 9 Low/Low           0    34
## 10 Low/Low          1     1
## 11 Low/Medium       0   435
## 12 Low/Medium       1     26
## 13 Medium/High      0  1486
## 14 Medium/High      1    232
## 15 Medium/Low       0   883
```

## 16 Medium/Low	1	3
## 17 Medium/Medium	0	9664
## 18 Medium/Medium	1	211