MATH2270 Assignment 1

Code

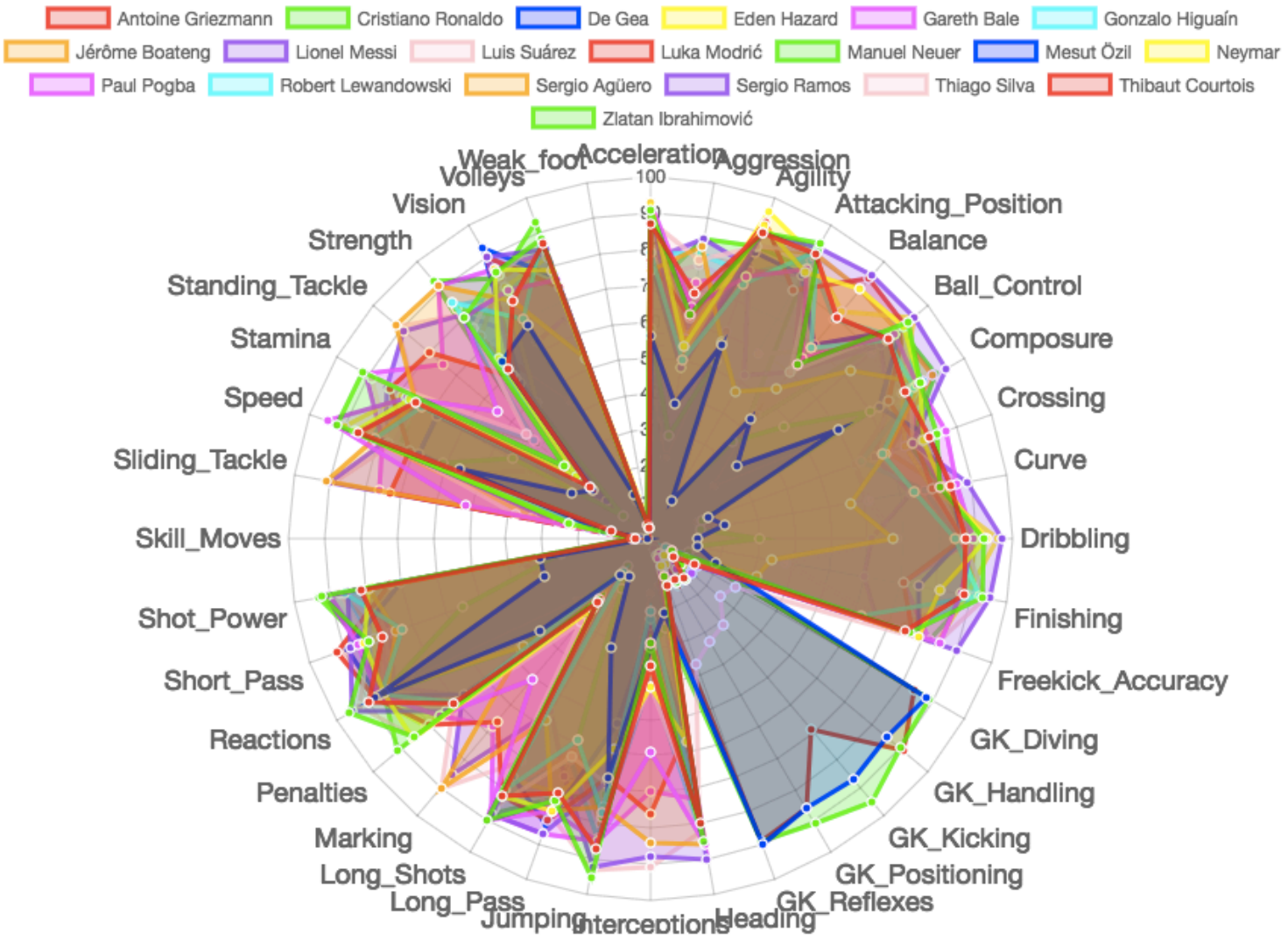
Deconstruct and Reconstruct

Student Details

- Carlos Eduardo Castillo Garzon (s3643370)

Part I - Data Visualisation

The video game Fifa 2017 has released its data to Kaggle as an open source for users to analyze it, where plenty of data analytics and scientists have access to. For the Assignment 1, I decided to evaluate a data visualisation (see bellow) made by the Kaggle User Hitesh Palamada about Fifa 2017 players.



Taken from: Kaggle (https://www.kaggle.com/hiteshp/exploring-fifa-2017-dataset/notebook)

Part II - Deconstruct

The chosen data visualization had a large number of flaws that under my perspective as follow: - Excessive amount of variables - Non-logical comparisons - Overlapping lines - The axis is not shown In general, it is an overcharged plot, where the author wanted to express the top 20 best players and all their skills in the polar plane. The author has tryed to elucidate the 20 best players in the game and which their individual skills were, which lead us to identify that he has used the wrong visualisation, as I could simply have used a bar chart plotting the overall player rating and then plot their skills by their position in the field. The data source has sufficient information to make a great data analysis with sufficient knowledge of the

subject. Taking into consideration that the question was “which are the top 20 best players in Fifa 2017 and which their individual skills are?”, We can conclude that the visualization was not properly telling the story to answer the question, leading us to supose that he failed two parts in the Kaiser Fung’s Trifecta.

Part III - Reconstruct

TOP 20 Best Players

Hide

Hide

```
##Merge the data to get the positions
Dataset=merge(FullData,Positions,by.x = "Preffered_Position",by.y ="Pos",all.x = T
RUE)
##Sorting by the best Overall Ranking
sorted<-arrange(Dataset,desc(Rating))
##Top 20 of the best Midfielders
top20<- head(sorted, 20)
##Retrieving the right data to plot
Pl <- top20 %>% select(Name, 11,54) %>% as.data.frame()
##Creating the data frame to plot(to fix the names)
PT <- data.frame(Name = c('Cristiano Ronaldo','Lionel Messi','Manuel Neuer','Neyma
r',
                          'Luis   Suares','De Gea','Gareth Bale', 'Robert Lewandow
ski',
                          'Zlatan Ibrahimovich','Mesut Ozil','Sergio Ramos','Jerom
e Boateng',
                          'Thiago Silva','Luka Modrich','Thibaut Courtois','Eden H
azard',                                 'Gonzalo Higuain','Sergio Aguero',
                          'Kevin De Bruyne','Mats Hummels'),
                 Rating = c(94,93,92,92,92,90,90,90,90,89,89,89,89,89,89,89,89,89
,88,88),
                 Position = c('Attacker','Attacker','GoalKeeper','Attacker',
'Attacker','GoalKeeper','Attacker','Attacker',
                              'Attacker','Midfielder','Defender','Defender',
'Defender','Midfielder','GoalKeeper','Attacker',
                              'Attacker','Attacker','Midfielder','Defender'))
# Basic barplot
p<-ggplot(data=PT, aes(reorder(Name,Rating),Rating , fill=Position)) +
  geom_bar(stat="identity",width=0.8) +theme(axis.text.x = element_text(angle = 90
, hjust = 1)) + ggtitle("Top 20 Best Players - FIFA 2017 (s3643370)")
p + coord_flip()
```

TOP 5 Best GoalKeepers

Hide

```
##Filter by GoalKepper
GoalKeeper <- Dataset %>% filter(Preffered_Position=='GK')
##Sorting by the best Overall Ranking
GoalKeepersorted<-arrange(GoalKeeper,desc(Rating))
##Top 5 of the best Goalkeepers
top5GK<- head(GoalKeepersorted, 5)
##Retrieving the right data to plot (Goalkeeper important skills)
(PlGoalKeeper <- top5GK %>% select(Name, 49:54) %>% as.data.frame())
```

	Name	GK_Positioning	GK_Diving	GK_Kicking	GK_Handling	GK_Reflexes	F
	<chr>	<int>	<int>	<int>	<int>	<int>	<
1	Manuel Neuer	91	89	95	90	89	0
2	De Gea	86	88	87	85	90	0
3	Thibaut Courtois	86	84	69	91	89	0
4	Gianluigi Buffon	90	87	68	88	84	0
5	Hugo Lloris	82	87	68	87	90	0

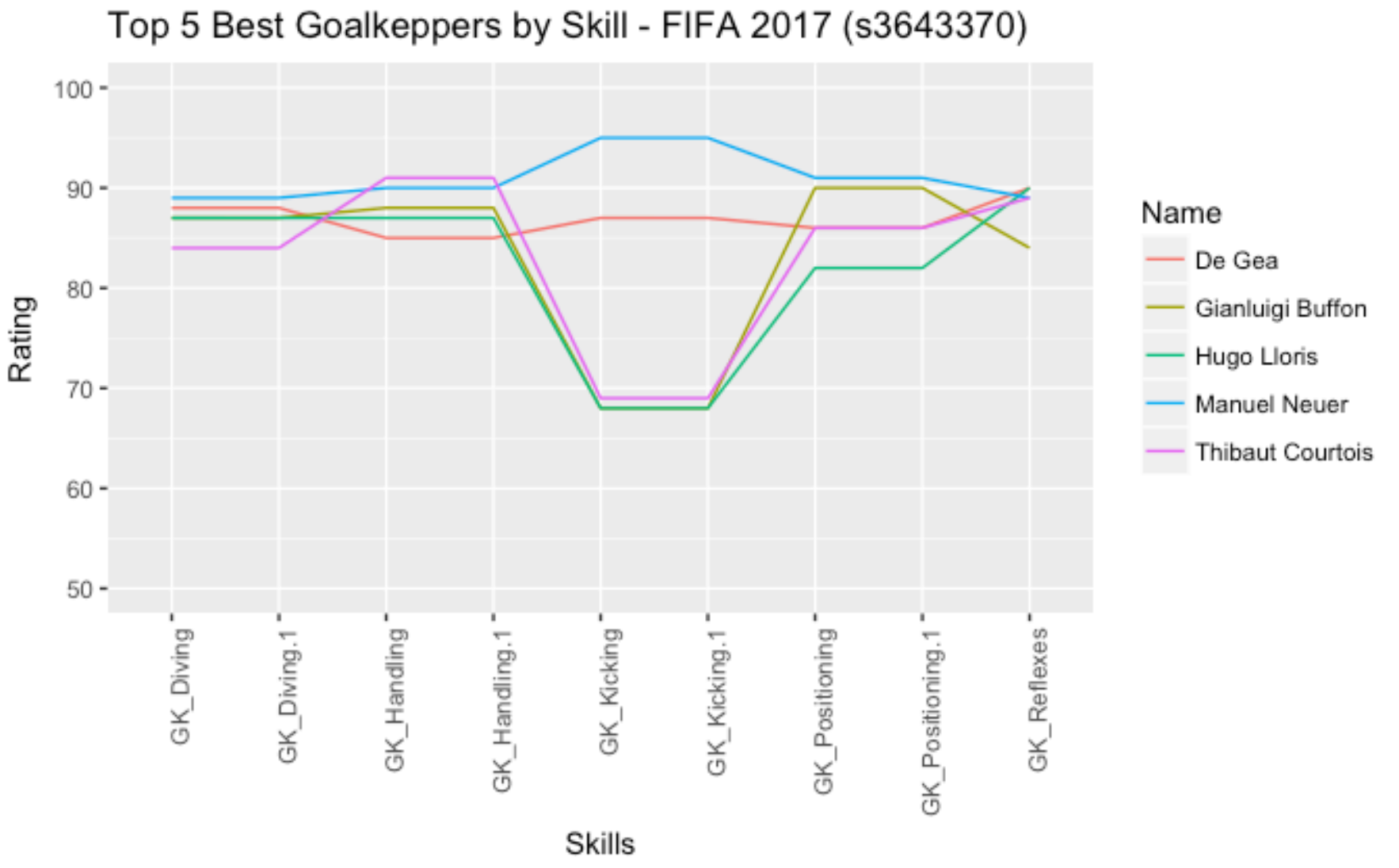
5 rows

Hide

Hide

Hide

```
##Creating the data frame to plot
GKPT <- data.frame(Name = c('Manuel Neuer','De Gea','Thibaut Courtois','Gianluigi
Buffon','Hugo Lloris'),
                   GK_Positioning = c(91,86,86,90,82),
                   GK_Diving = c(89,88,84,87,87),
                   GK_Kicking = c(95,87,69,68,68),
                   GK_Handling = c(90,85,91,88,87),
                   GK_Reflexes = c(89,90,89,84,90),
                   GK_Positioning= c(91,86,86,90,82),
                   GK_Diving= c(89,88,84,87,87),
                   GK_Kicking= c(95,87,69,68,68),
                   GK_Handling= c(90,85,91,88,87),
                   GK_Reflexes= c(89,90,89,84,90))
df.gathered <- gather(GKPT,key = 'Skills',value = 'Rating',2:10)
##Plot
ggplot(df.gathered, aes(x = Skills, y = Rating , group = Name, col = Name))+geom_l
ine()+
  scale_y_continuous(limits = c(50,100), breaks = seq(0,100,10)) + theme(axis.text
.x = element_text(angle = 90, hjust = 1)) + ggtitle("Top 5 Best Goalkeppers by Ski
ll - FIFA 2017 (s3643370)")
```



TOP 5 Best Defenders

Hide

Hide

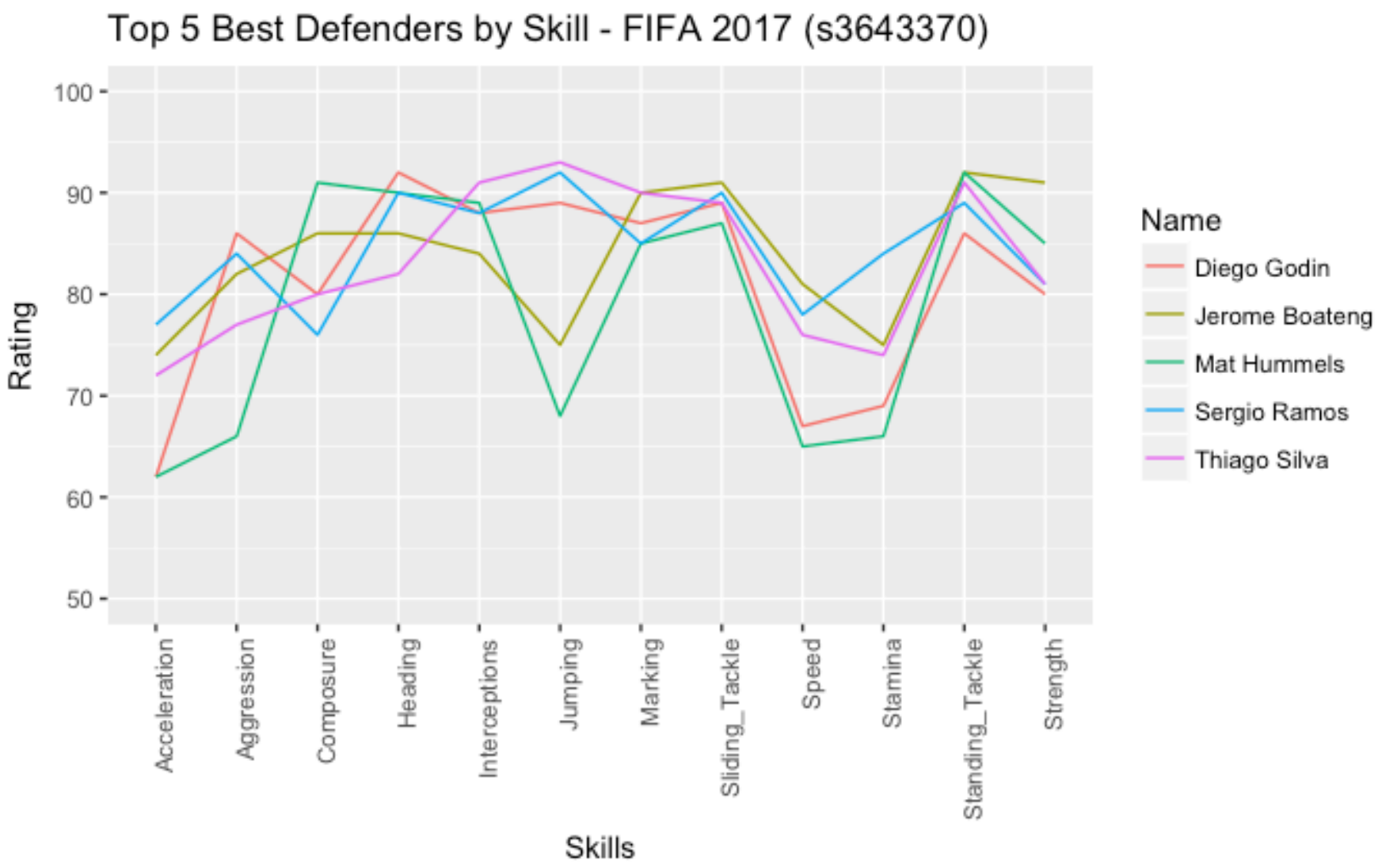
```
##Filter by Defender
Defender <- Dataset %>% filter(Position=='Defender')
##Sorting by the best Overall Ranking
Defendersorted<-arrange(Defender,desc(Rating))
##Top 5 of the best Defenders
top5DF<- head(Defendersorted, 5)
##Retrieving the right data to plot (Defender important skills)
(PlDefender <- top5DF %>% select(Name, 22:25,28,30,34:37,40:41) %>% as.data.frame(
))
```

	Name <chr>	Marki... <int>	Sliding_Tackle <int>	Standing_Tackle <int>	Aggressio <in
1	Sergio Ramos	85	90	89	8
2	J\x92\xa9r\x92\x82me Boateng	90	91	92	8
3	Thiago Silva	90	89	91	7
4	Mats Hummels	85	87	92	6
5	Diego God\x92_n	87	89	86	8

5 rows | 1-7 of 13 columns

Hide

```
##Creating the data frame to plot
DFPT <- data.frame(Name = c('Sergio Ramos','Jerome Boateng','Thiago Silva','Mat Hu
mmels','Diego Godin'),
                   Marking = c(85,90,90,85,87),
                   Sliding_Tackle = c(90,91,89,87,89),
                   Standing_Tackle = c(89,92,91,92,86),
                   Aggression = c(84,82,77,66,86),
                   Interceptions = c(88,84,91,89,88),
                   Composure= c(76,86,80,91,80),
                   Acceleration= c(77,74,72,62,62),
                   Speed= c(78,81,76,65,67),
                   Stamina= c(84,75,74,66,69),
                   Strength= c(81,91,81,85,80),
                   Jumping= c(92,75,93,68,89),
                   Heading= c(90,86,82,90,92))
df.gathered1 <- gather(DFPT,key = 'Skills',value = 'Rating',2:13)
##Plot
ggplot(df.gathered1, aes(x = Skills, y = Rating , group = Name, col = Name))+geom_
line()+
  scale_y_continuous(limits = c(50,100), breaks = seq(0,100,10)) + theme(axis.text
.x = element_text(angle = 90, hjust = 1)) + ggtitle("Top 5 Best Defenders by Skill
- FIFA 2017 (s3643370)")
```



TOP 5 Best Midfielders

Hide

```
##Filter by Midfielder
Midfielder <- Dataset %>% filter(Position=='Midfielder')
##Sorting by the best Overall Ranking
Midfieldersorted<-arrange(Midfielder,desc(Rating))
##Top 5 of the best Midfielders
top5MF<- head(Midfieldersorted, 5)
##Retrieving the right data to plot (Midfielder important skills)
(PlMidfielder <- top5MF %>% select(Name, 20:22,25,29,31:37) %>% as.data.frame())
```

	Name	Ball_Control	Dribbling	Marki...	Aggression	Vision	Crossing	S
	<chr>	<int>	<int>	<int>	<int>	<int>	<int>	
1	Mesut \x92_zil	90	86	22	48	93	82	
2	Luka Modri\xf3\xf6	92	86	66	62	90	78	
3	Kevin De Bruyne	86	85	30	68	88	90	
4	Iniesta	93	90	57	58	94	79	
5	Paul Pogba	90	89	68	72	86	78	

5 rows | 1-9 of 13 columns

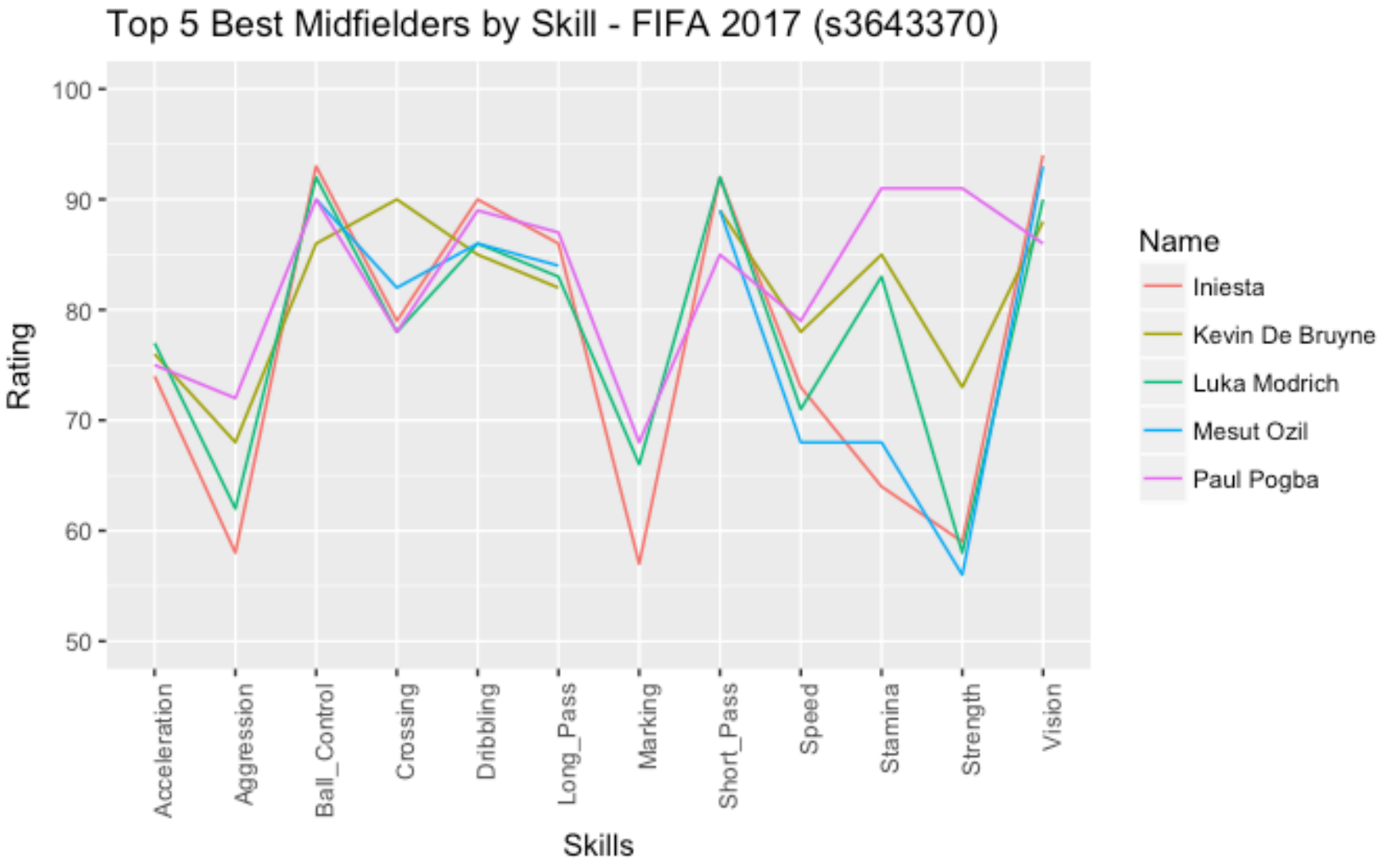
Hide

Hide

Hide

```
##Creating the data frame to plot
MFPT <- data.frame(Name = c('Mesut Ozil','Luka Modrich','Kevin De Bruyne','Iniesta
','Paul Pogba'),

                   Ball_Control = c(90,92,86,93,90),
                   Dribbling = c(86,86,85,90,89),
                   Marking = c(22,66,30,57,68),
                   Aggression = c(48,62,68,58,72),
                   Vision = c(93,90,88,94,86),
                   Crossing = c(82,78,90,79,78),
                   Short_Pass= c(89,92,89,92,85),
                   Long_Pass= c(84,83,82,86,87),
                   Acceleration= c(77,77,76,74,75),
                   Speed= c(68,71,78,73,79),
                   Stamina= c(68,83,85,64,91),
                   Strength= c(56,58,73,59,91))
df.gathered2 <- gather(MFPT,key = 'Skills',value = 'Rating',2:13)
##Plot
ggplot(df.gathered2, aes(x = Skills, y = Rating , group = Name, col = Name))+geom_
line()+
  scale_y_continuous(limits = c(50,100), breaks = seq(0,100,10)) + theme(axis.text
.x = element_text(angle = 90, hjust = 1)) + ggtitle("Top 5 Best Midfielders by Ski
ll - FIFA 2017 (s3643370)")
```



TOP 5 Best Attackers

Hide

Hide

```
##Filter by Attacker
Attacker <- Dataset %>% filter(Position=='Attacker')
##Sorting by the best Overall Ranking
Attackersorted<-arrange(Attacker,desc(Rating))
##Top 5 of the best Midfielders
top5AT<- head(Attackersorted, 5)
##Retrieving the right data to plot (Attacker important skills)
(lAttacker <- top5AT %>% select(Name, 20:21,26:27,32,34:36,39:48) %>% as.data.fram
e())
```

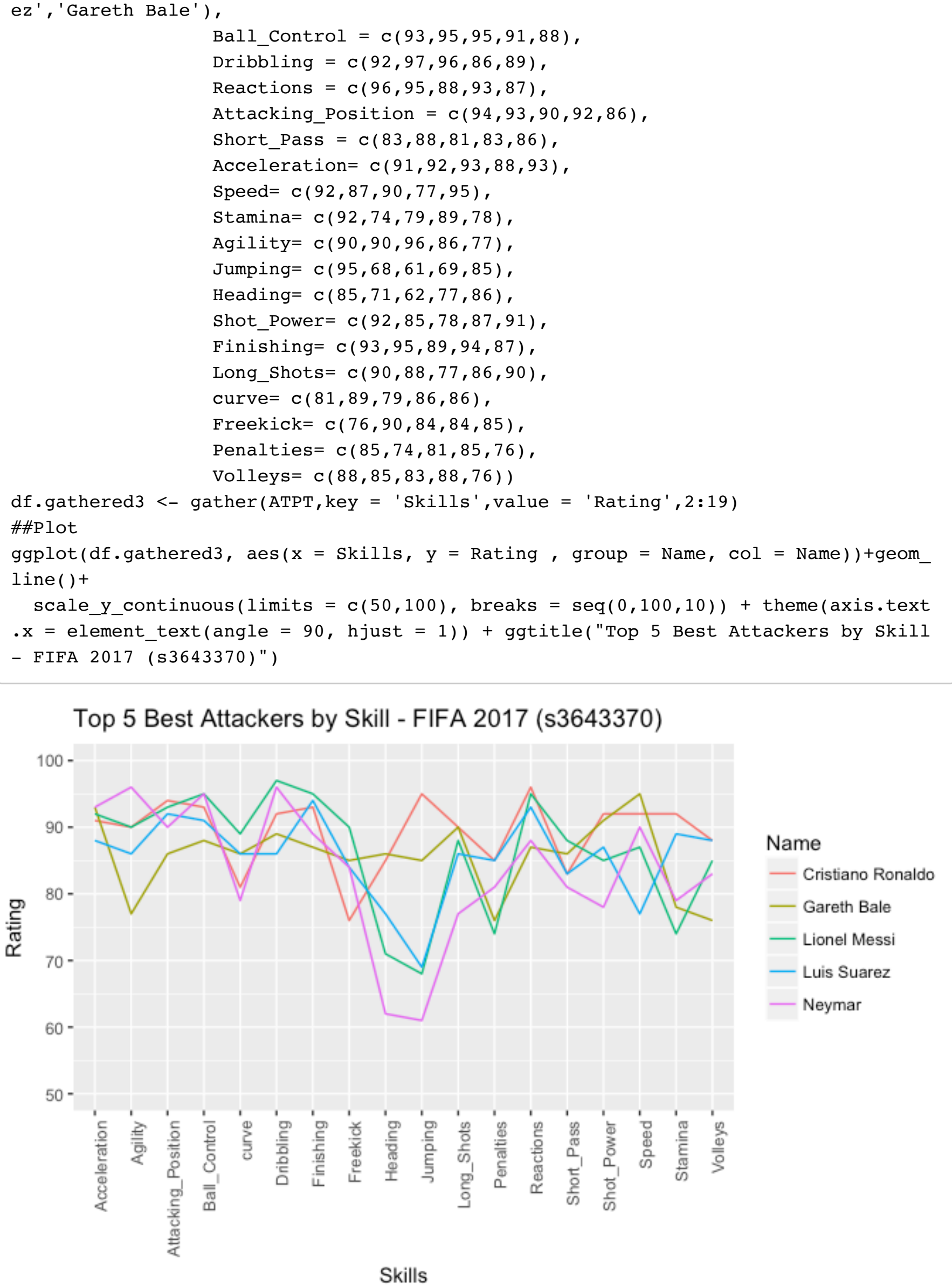
	Name	Ball_Control	Dribbling	Reactions	Attacking_Position	Short_
	<chr>	<int>	<int>	<int>	<int>	
1	Cristiano Ronaldo	93	92	96	94	
2	Lionel Messi	95	97	95	93	
3	Neymar	95	96	88	90	
4	Luis Su\x92\x8drez	91	86	93	92	
5	Gareth Bale	88	89	87	86	

5 rows | 1-8 of 19 columns

Hide

Hide

```
##Creating the data frame to plot
ATPT <- data.frame(Name = c('Cristiano Ronaldo','Lionel Messi','Neymar','Luis Suar
ez','Gareth Bale'),
                   Ball_Control = c(93,95,95,91,88),
                   Dribbling = c(92,97,96,86,89),
                   Reactions = c(96,95,88,93,87),
                   Attacking_Position = c(94,93,90,92,86),
                   Short_Pass = c(83,88,81,83,86),
                   Acceleration= c(91,92,93,88,93),
                   Speed= c(92,87,90,77,95),
                   Stamina= c(92,74,79,89,78),
                   Agility= c(90,90,96,86,77),
                   Jumping= c(95,68,61,69,85),
                   Heading= c(85,71,62,77,86),
                   Shot_Power= c(92,85,78,87,91),
                   Finishing= c(93,95,89,94,87),
                   Long_Shots= c(90,88,77,86,90),
                   curve= c(81,89,79,86,86),
                   Freekick= c(76,90,84,84,85),
                   Penalties= c(85,74,81,85,76),
                   Volleys= c(88,85,83,88,76))
df.gathered3 <- gather(ATPT,key = 'Skills',value = 'Rating',2:19)
##Plot
ggplot(df.gathered3, aes(x = Skills, y = Rating , group = Name, col = Name))+geom_
line()+
  scale_y_continuous(limits = c(50,100), breaks = seq(0,100,10)) + theme(axis.text
.x = element_text(angle = 90, hjust = 1)) + ggtitle("Top 5 Best Attackers by Skill
- FIFA 2017 (s3643370)")
```



Conclusions

The reconstruction process focused in visualise what I conclude the author was trying to visualise. Firstly, the creation of a bar plot was required to express the Top 20 best players in overall ranking, splitting them by their position in the field, which gives the user a better message in which players are the best and what role they perform in their National or club teams. Secondly, and probably the most important one was the creation of four plots that shows the top 5 best players in overall, but in this case filtered by their position in the field, such as Goalkeeper, Defender, Midfielder and Attacker, and using only the most important skills for that roles, which makes the plots way less overcharged and more logical for soccer fans. The use of a Cartesian plane with line plot, makes the information understanding a lot better than the original plot, with the presence of axes and nice colour codification. Note: apart from the original data, I had to go to Fifa 2017 website (http://www.fifauteam.com/fifa-17-position-change-cards-guide/#Bonus (http://www.fifauteam.com/fifa-17-position-change-cards-guide/#Bonus)) to collect the common name of the roles, as in the data set those were using coded names, to finally create my own file to be merged with the original data file.

References

-Original visualisation: https://www.kaggle.com/hiteshp/exploring-fifa-2017-dataset/notebook (https://www.kaggle.com/hiteshp/exploring-fifa-2017-dataset/notebook)

-Original data sources https://www.kaggle.com/hiteshp/exploring-fifa-2017-dataset/data (https://www.kaggle.com/hiteshp/exploring-fifa-2017-dataset/data)

-Additional data sources http://www.fifauteam.com/fifa-17-position-change-cards-guide/#Bonus (http://www.fifauteam.com/fifa-17-position-change-cards-guide/#Bonus)