

Data Analysis

A Look at FIFA Players and Attributes

Sau Kha

Dataset

- European Soccer Database (from Kaggle.com)
- <https://www.kaggle.com/hugomathien/soccer/data>
- +25,000 matches
- +10,000 players
- 11 European Countries with their lead championship
- Seasons 2008 to 2016
- Players' and Teams' attributes sourced from EA Sports' FIFA video game series

Python – SQLite3 & Pandas

- Create connection object
- `conn = sqlite3.connect ('database.sqlite')`
- Define SQL
- `sql = """SELECT * FROM sqlite_master WHERE type='table';"""`
- Pandas command; get data into a DataFrame
- `df = pd.read_sql(sql, conn)`

Data Exploration: All Tables

```
df_tables = pd.read_sql("""SELECT * FROM sqlite_master
WHERE type='table';""", conn)
```

Tables in the database

df_tables shape: (8, 5)

	type	name	tbl_name	rootpage	\
0	table	sqlite_sequence	sqlite_sequence	4	
1	table	Player_Attributes	Player_Attributes	11	
2	table	Player	Player	14	
3	table	Match	Match	18	
4	table	League	League	24	
5	table	Country	Country	26	
6	table	Team	Team	29	
7	table	Team_Attributes	Team_Attributes	2	

Data Exploration: All Tables

```
                                sql
0      CREATE TABLE sqlite_sequence(name,seq)
1  CREATE TABLE "Player_Attributes" (\n\t`id`\tIN...
2  CREATE TABLE `Player` (\n\t`id`\tINTEGER PRIMA...
3  CREATE TABLE `Match` (\n\t`id`\tINTEGER PRIMAR...
4  CREATE TABLE `League` (\n\t`id`\tINTEGER PRIMA...
5  CREATE TABLE `Country` (\n\t`id`\tINTEGER PRIM...
6  CREATE TABLE "Team" (\n\t`id`\tINTEGER PRIMARY...
7  CREATE TABLE `Team_Attributes` (\n\t`id`\tINTE...
```

Data Exploration: Player Table

Player table:

```
CREATE TABLE `Player` (  
  `id`      INTEGER PRIMARY KEY AUTOINCREMENT,  
  `player_api_id` INTEGER UNIQUE,  
  `player_name` TEXT,  
  `player_fifa_api_id` INTEGER UNIQUE,  
  `birthday` TEXT,  
  `height`  INTEGER,  
  `weight`  INTEGER  
)
```

Data Exploration: Player Attributes Table

```
Player_Attributes table:  
CREATE TABLE "Player_Attributes" (  
  `id`      INTEGER PRIMARY KEY AUTOINCREMENT,  
  `player_fifa_api_id`  INTEGER,  
  `player_api_id`  INTEGER,  
  `date`  TEXT,  
  `overall_rating`      INTEGER,  
  `potential`  INTEGER,  
  `preferred_foot`      TEXT,  
  `attacking_work_rate`  TEXT,  
  `defensive_work_rate`  TEXT,  
  `crossing`      INTEGER,  
  `finishing`      INTEGER,  
  `heading_accuracy`      INTEGER,  
  `short_passing`  INTEGER,  
  `volleys`      INTEGER,  
  `dribbling`      INTEGER,  
  `curve`  INTEGER,  
  `free_kick_accuracy`  INTEGER,  
  `long_passing`  INTEGER,  
  `ball_control`  INTEGER,  
  `acceleration`  INTEGER,  
  `sprint_speed`  INTEGER,
```

Data Exploration: Player Attributes Table

```
`agility`      INTEGER,  
`reactions`    INTEGER,  
`balance`      INTEGER,  
`shot_power`   INTEGER,  
`jumping`      INTEGER,  
`stamina`      INTEGER,  
`strength`     INTEGER,  
`long_shots`   INTEGER,  
`aggression`   INTEGER,  
`interceptions` INTEGER,  
`positioning`  INTEGER,  
`vision`       INTEGER,  
`penalties`    INTEGER,  
`marking`      INTEGER,  
`standing_tackle`  INTEGER,  
`sliding_tackle`  INTEGER,  
`gk_diving`     INTEGER,  
`gk_handling`   INTEGER,  
`gk_kicking`    INTEGER,  
`gk_positioning` INTEGER,  
`gk_reflexes`   INTEGER,  
FOREIGN KEY(`player_fifa_api_id`) REFERENCES `Player`(`player_fifa_api_id`),  
FOREIGN KEY(`player_api_id`) REFERENCES `Player`(`player_api_id`)
```


Data Cleaning

- `df_all_col.replace(r'\s+', np.nan, regex=True, inplace = True)`
- #drop row with any NA entry
`df_all_col.dropna(axis=0, how='any', inplace=True)`
- `df_all_col.sort_values('player_name',axis=0, inplace=True)`
- `df_all_col.drop_duplicates(inplace=True)`
- `df_all_col.to_csv('df_all_col.csv')`
- Before: 183929 rows, 50 columns
- After: 10898 rows, 50 columns
- 38 columns with numeric attributes for principal component analysis

Research Questions

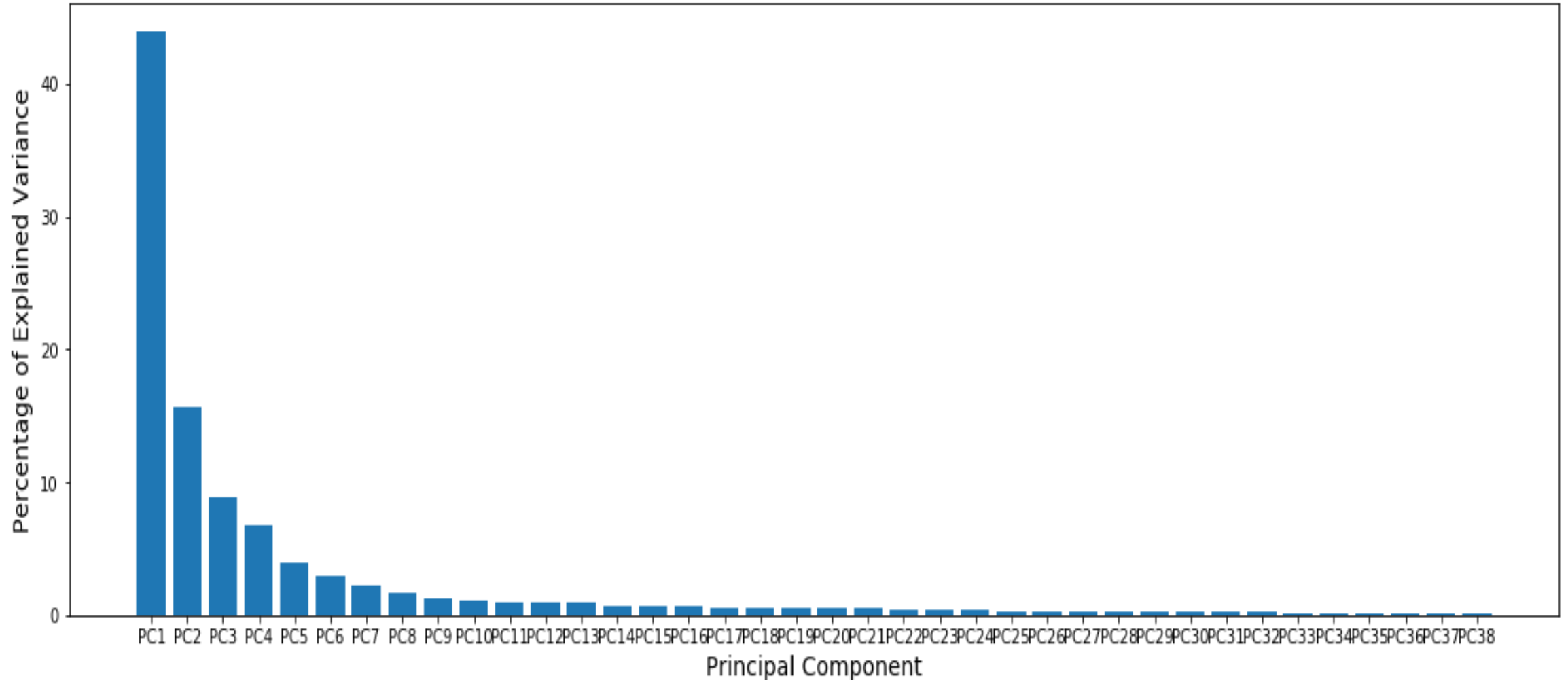
1. Which player's attribute contributes most to player's overall rating?
2. What attributes set players apart?

Intended Audience

1. Soccer fans
2. FIFA Video Game Players

Principal Component Analysis (PCA)

Scree Plot



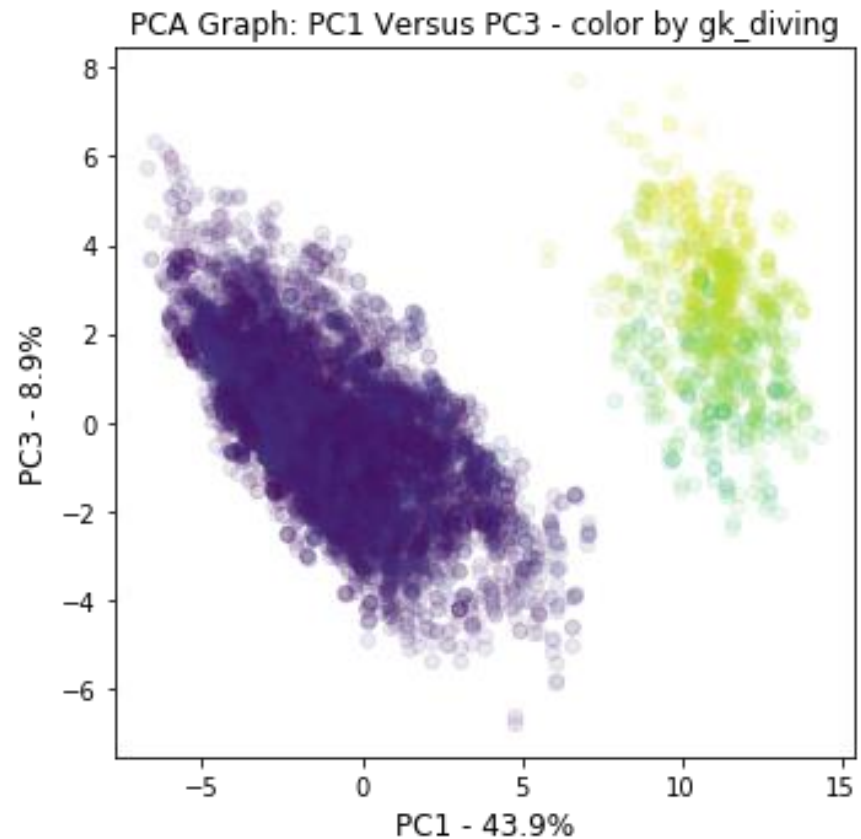
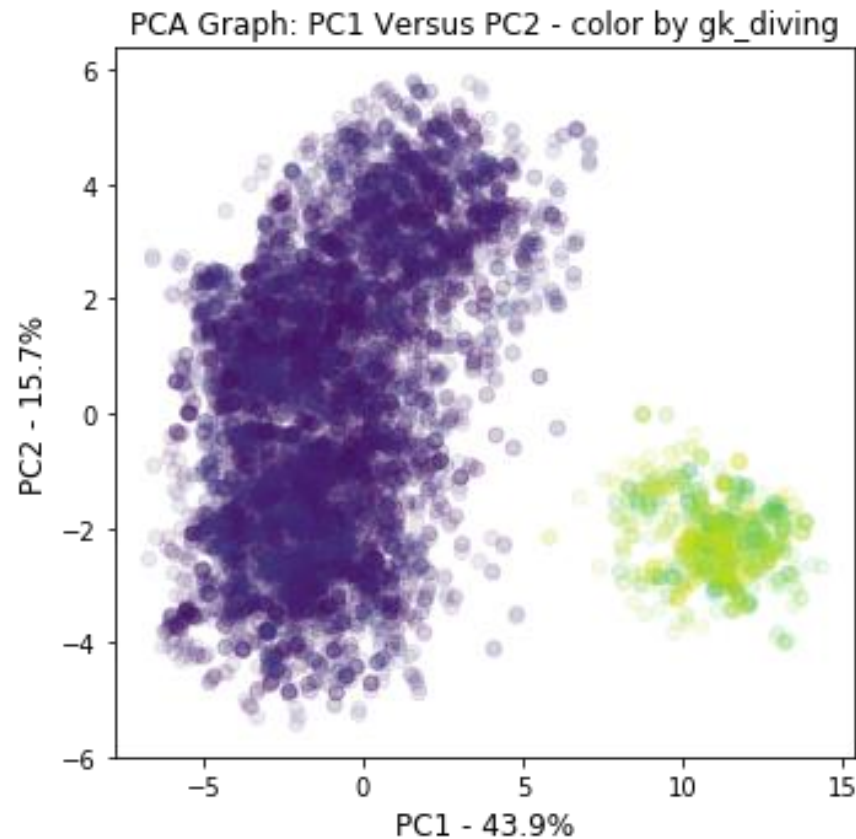
PCA 1 Loading Scores

Player Attributes	Loading Scores	Player Attributes	Loading Scores	Player Attributes	Loading Scores
ball_control	0.233895	gk_positioning	0.193662	gk_kicking	0.121982
dribbling	0.226609	volleys	0.193077	reactions	0.105162
short_passing	0.220168	finishing	0.191773	aggression	0.087644
crossing	0.213529	acceleration	0.185480	overall_rating	0.083033
curve	0.211442	penalties	0.183370	potential	0.080174
long_shots	0.211257	sprint_speed	0.182038	interceptions	0.071469
positioning	0.204851	long_passing	0.177636	standing_tackle	0.065230
shot_power	0.199655	agility	0.168734	sliding_tackle	0.059237
vision	0.197614	stamina	0.158548	marking	0.052163
gk_diving	0.196600	balance	0.153811	strength	0.028138
free_kick_accuracy	0.196366	height	0.134852	age	0.005573
gk_reflexes	0.195234	heading_accuracy	0.132158	jumping	0.001561
gk_handling	0.195011	weight	0.125473		

PC1 Vs PC2

PC1 Vs PC3

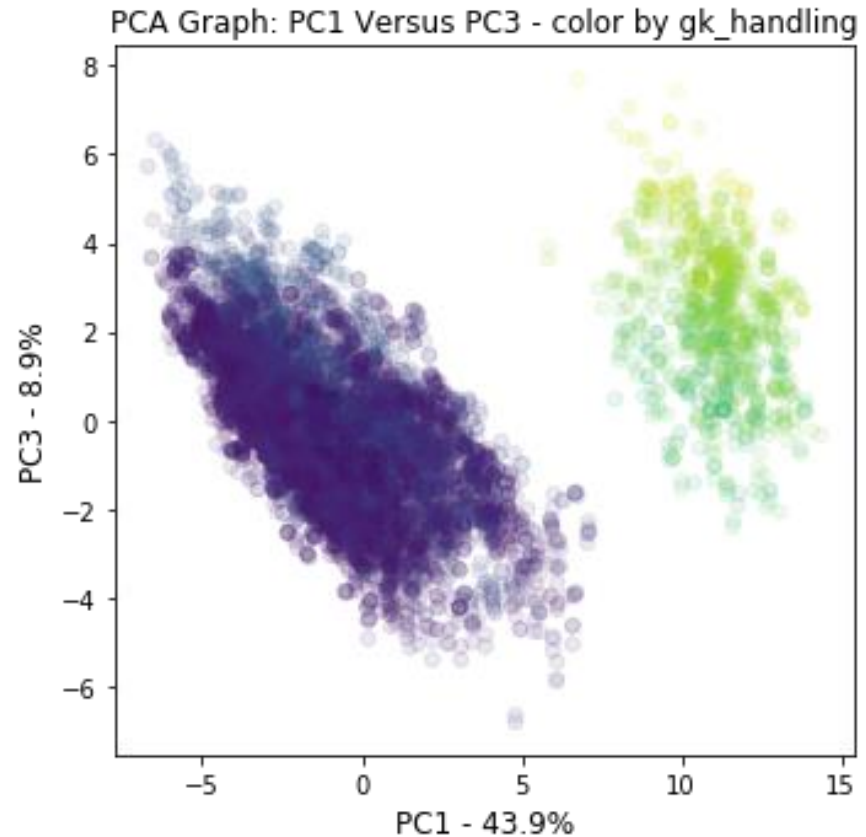
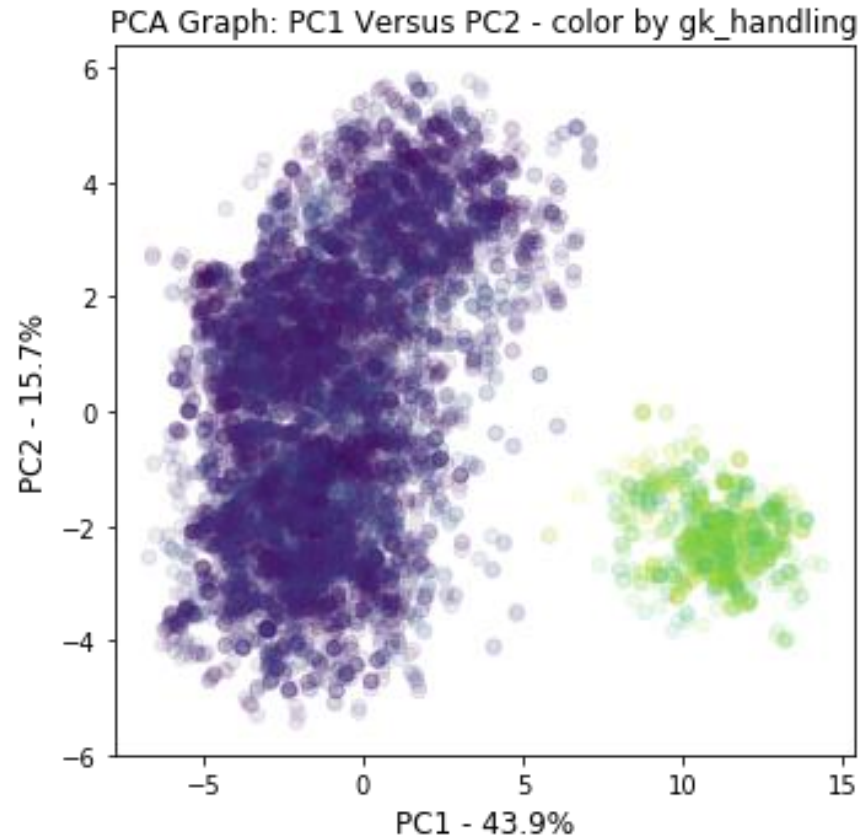
color by gk_diving



PC1 Vs PC2

PC1 Vs PC3

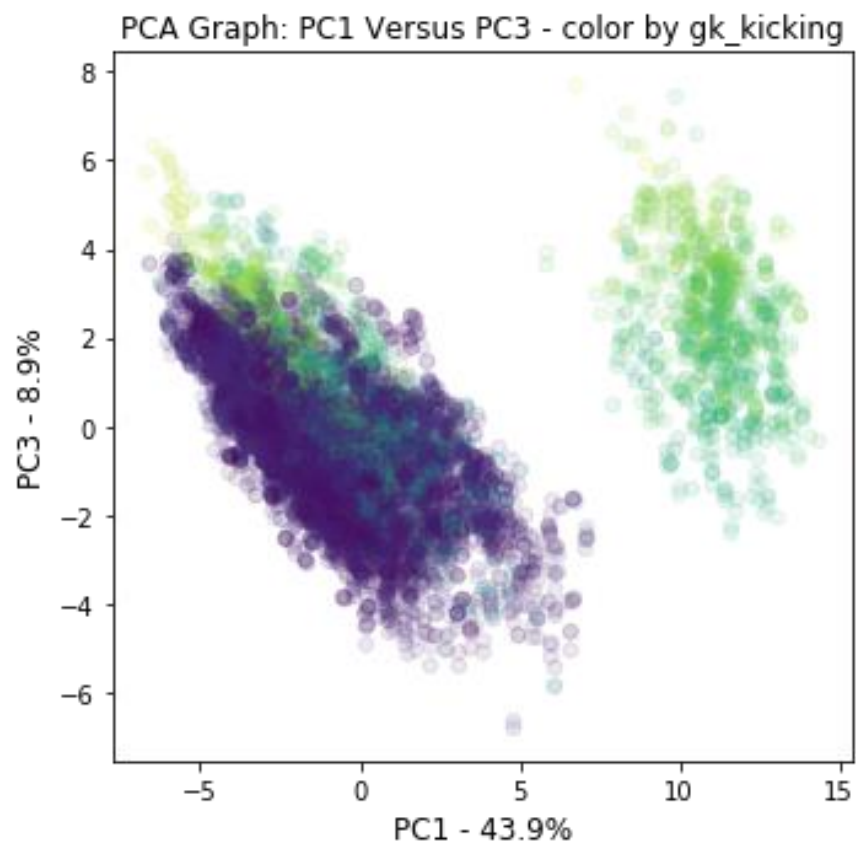
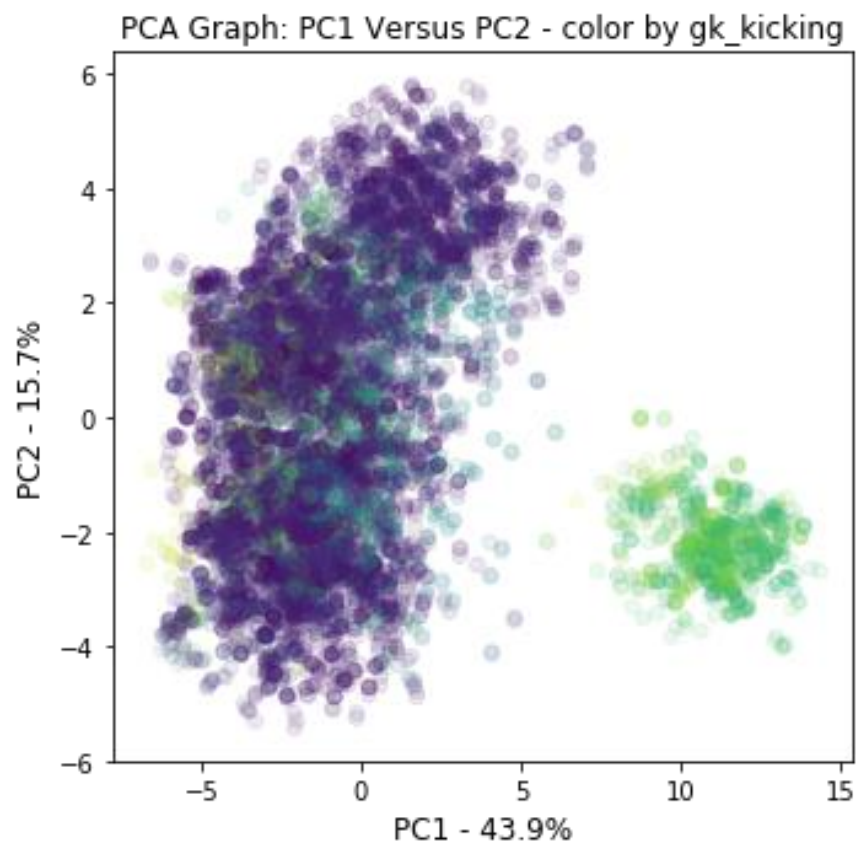
color by gk_handling



PC1 Vs PC2

PC1 Vs PC3

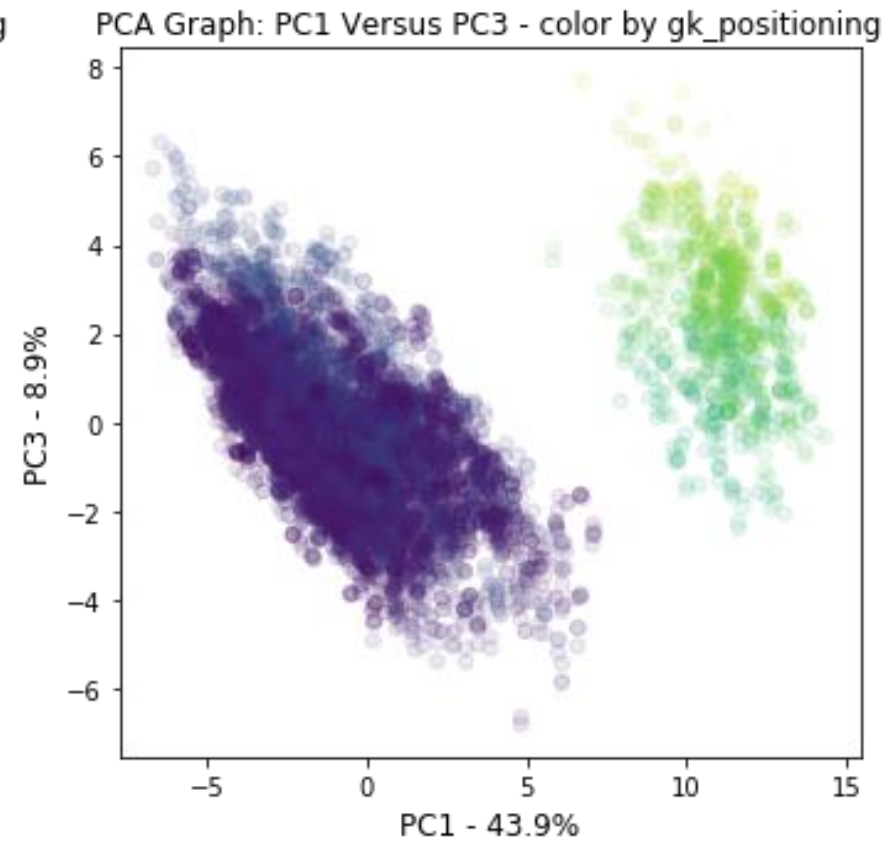
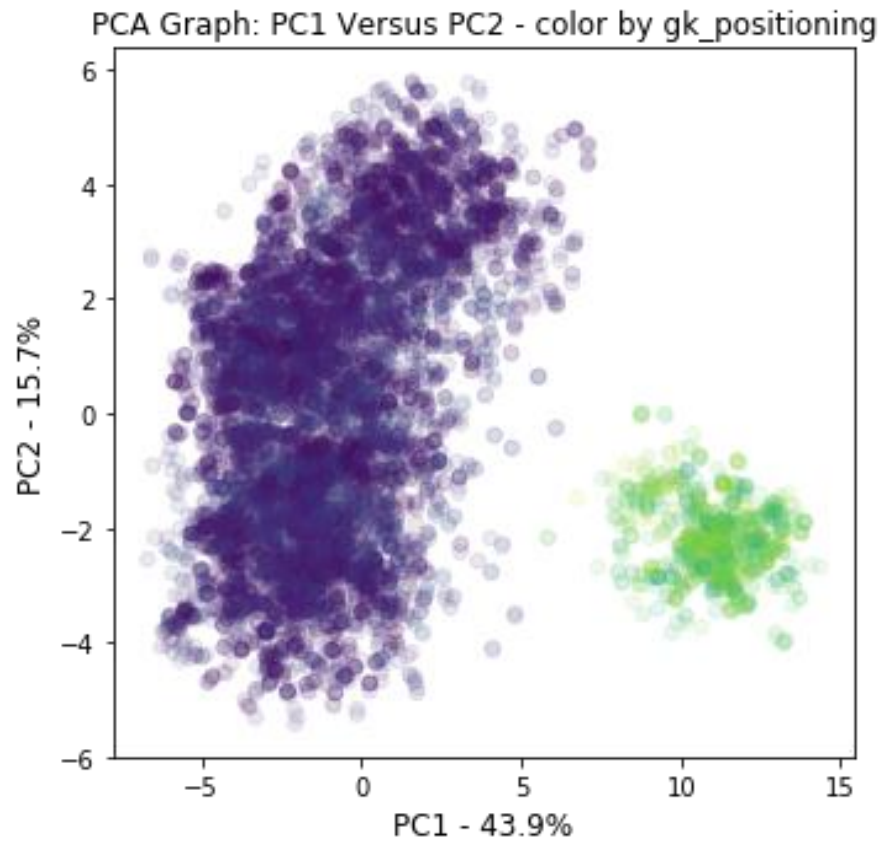
color by gk_kicking



PC1 Vs PC2

PC1 Vs PC3

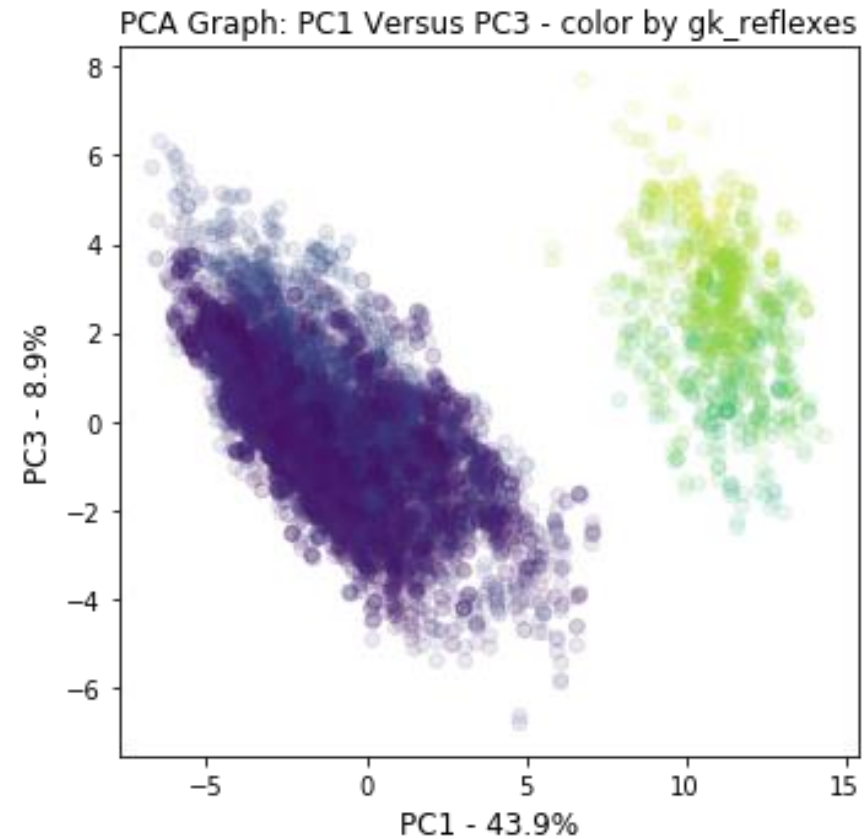
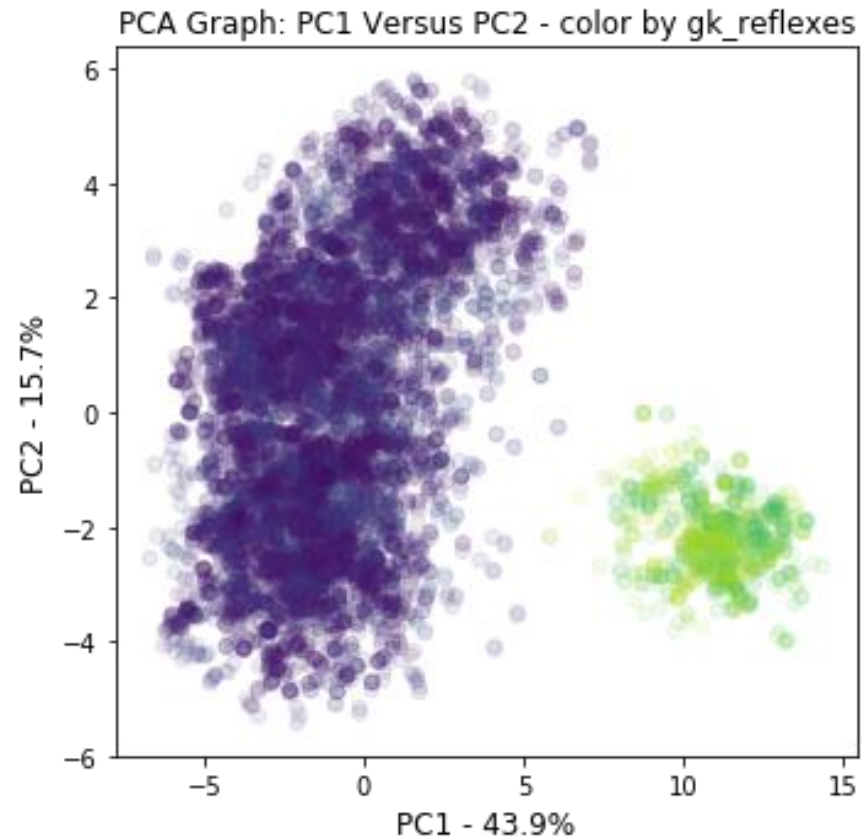
color by gk_positioning

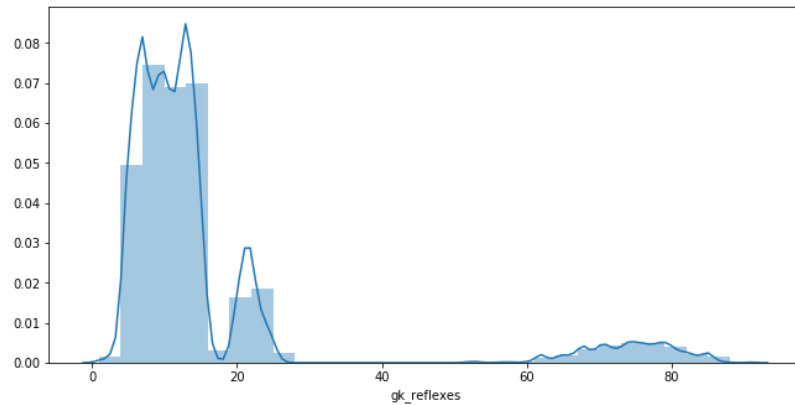
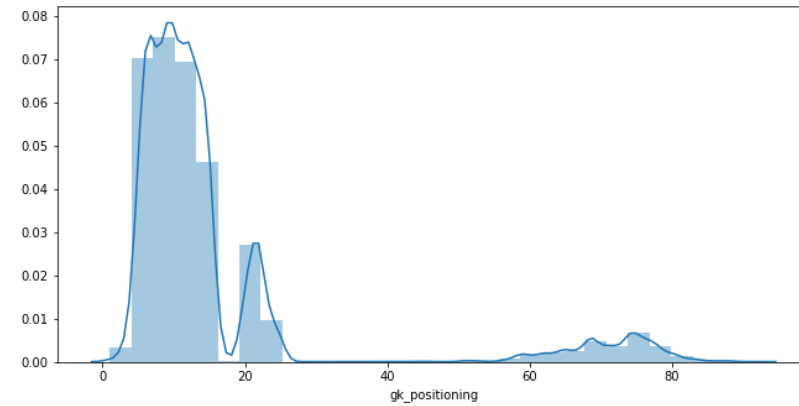
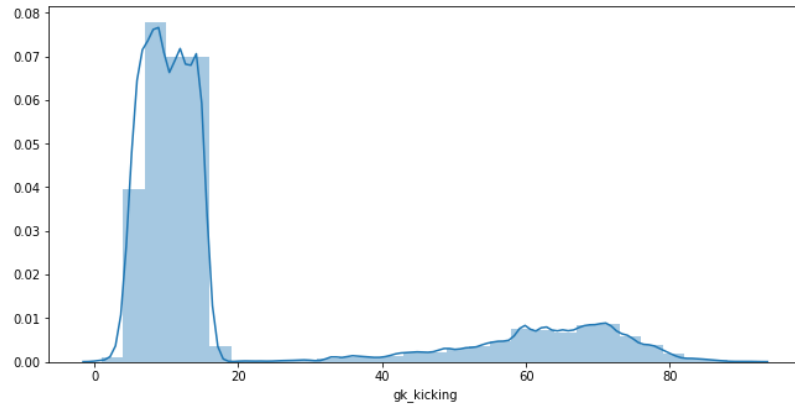
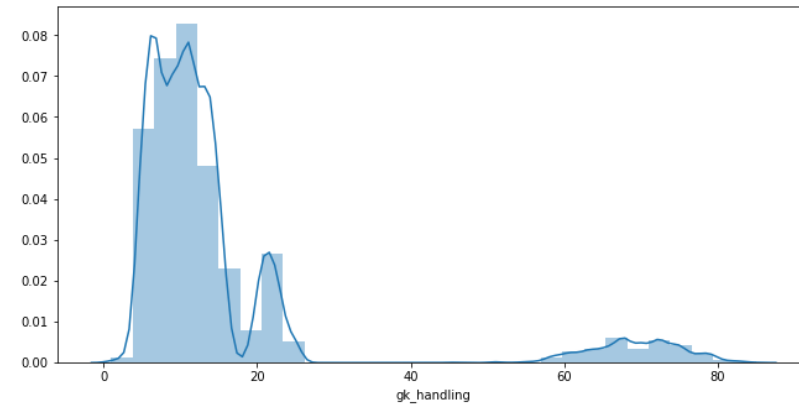
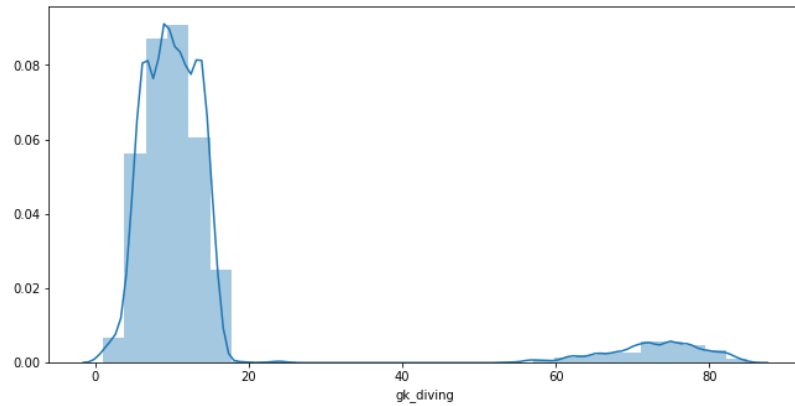


PC1 Vs PC2

PC1 Vs PC3

color by gk_reflexes

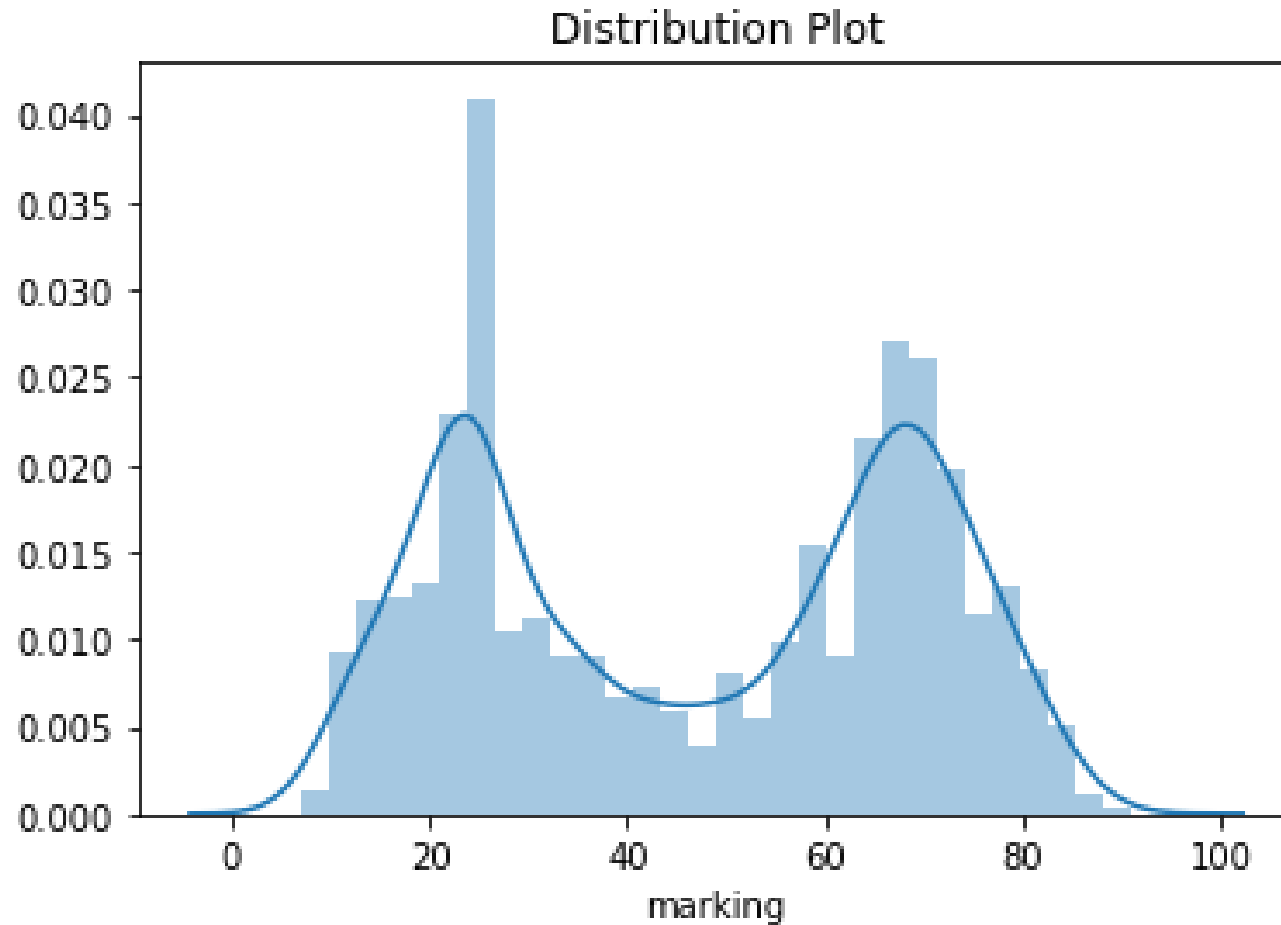




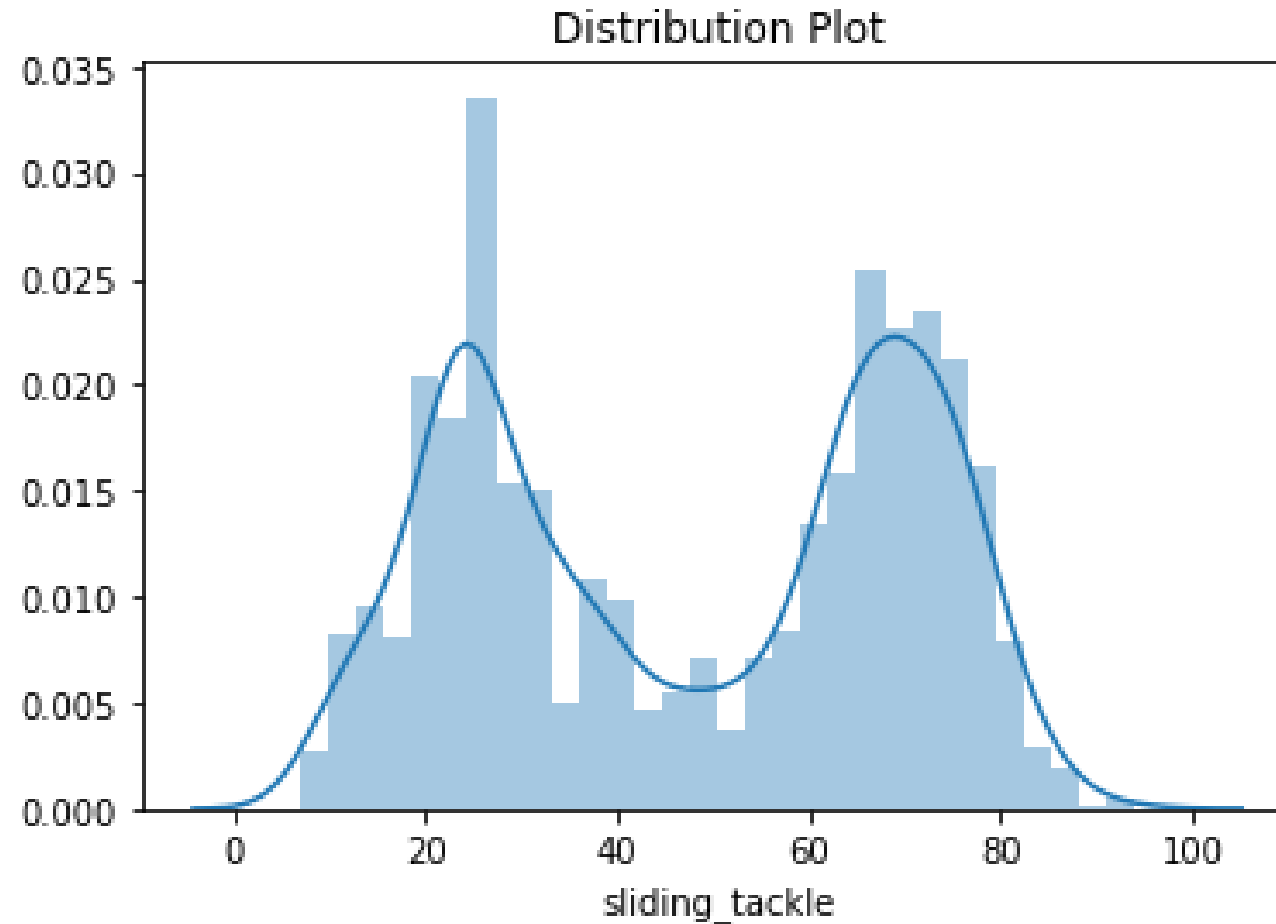
Distribution Plots – goalkeeping attributes

1. goalkeepers – one small subgroup on far right
2. additional small subgroup
 - a. gk_reflexes, gk_handling and gk_positioning
 - b. scores ranging from 18 – 28

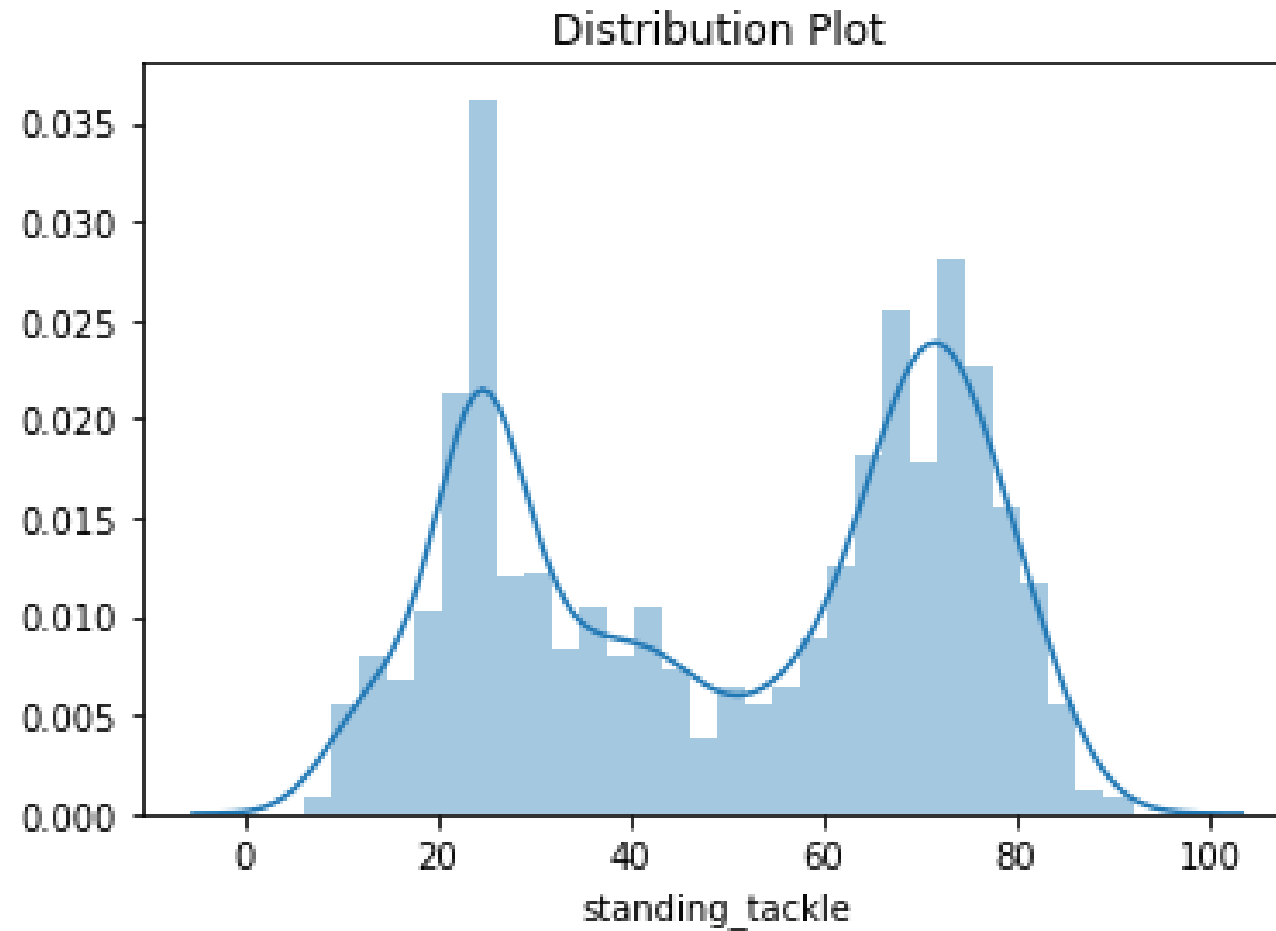
Bimodal Distribution Marking (Defending)



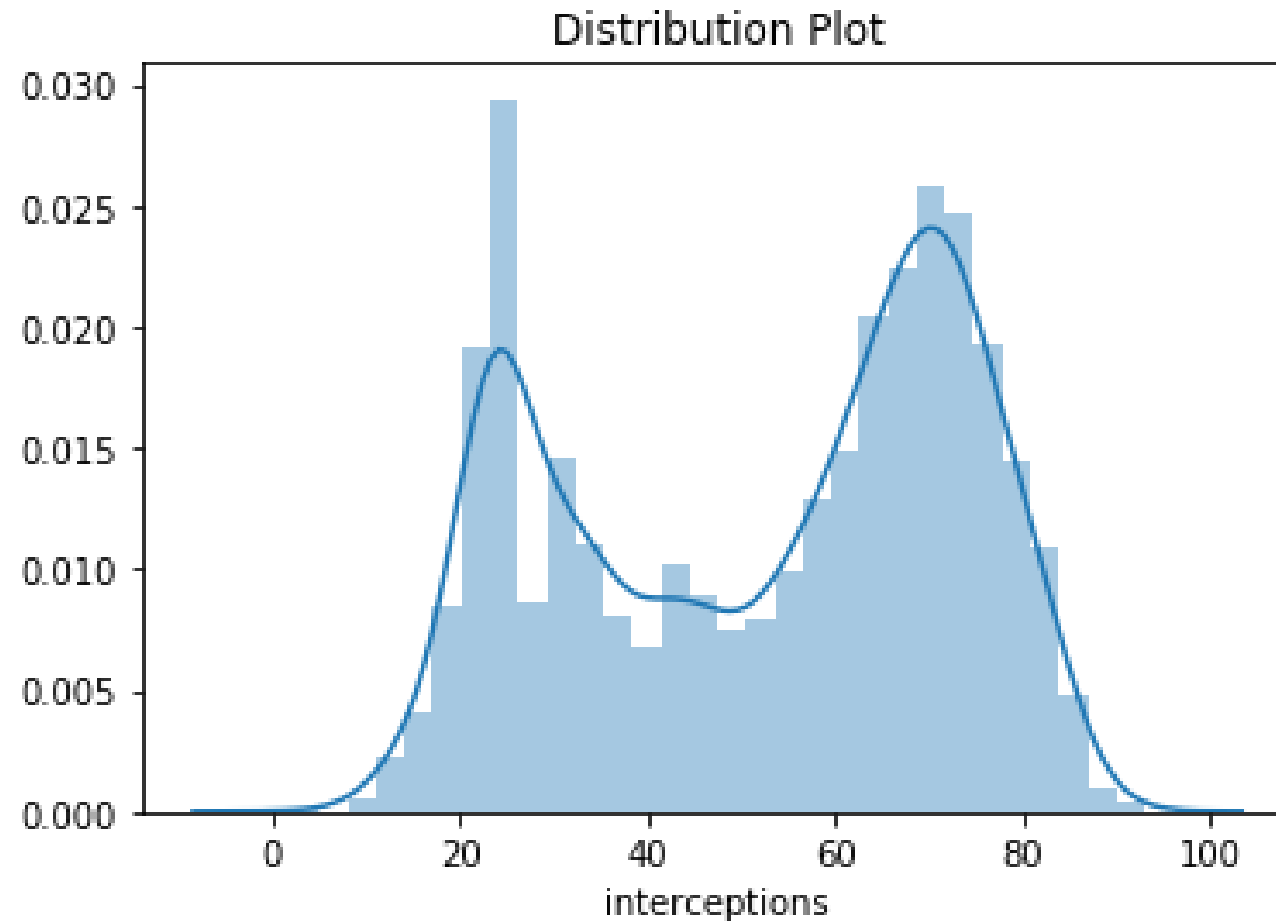
Bimodal Distribution Sliding Tackle (Defending)



Bimodal Distribution Standing Tackle (Defending)

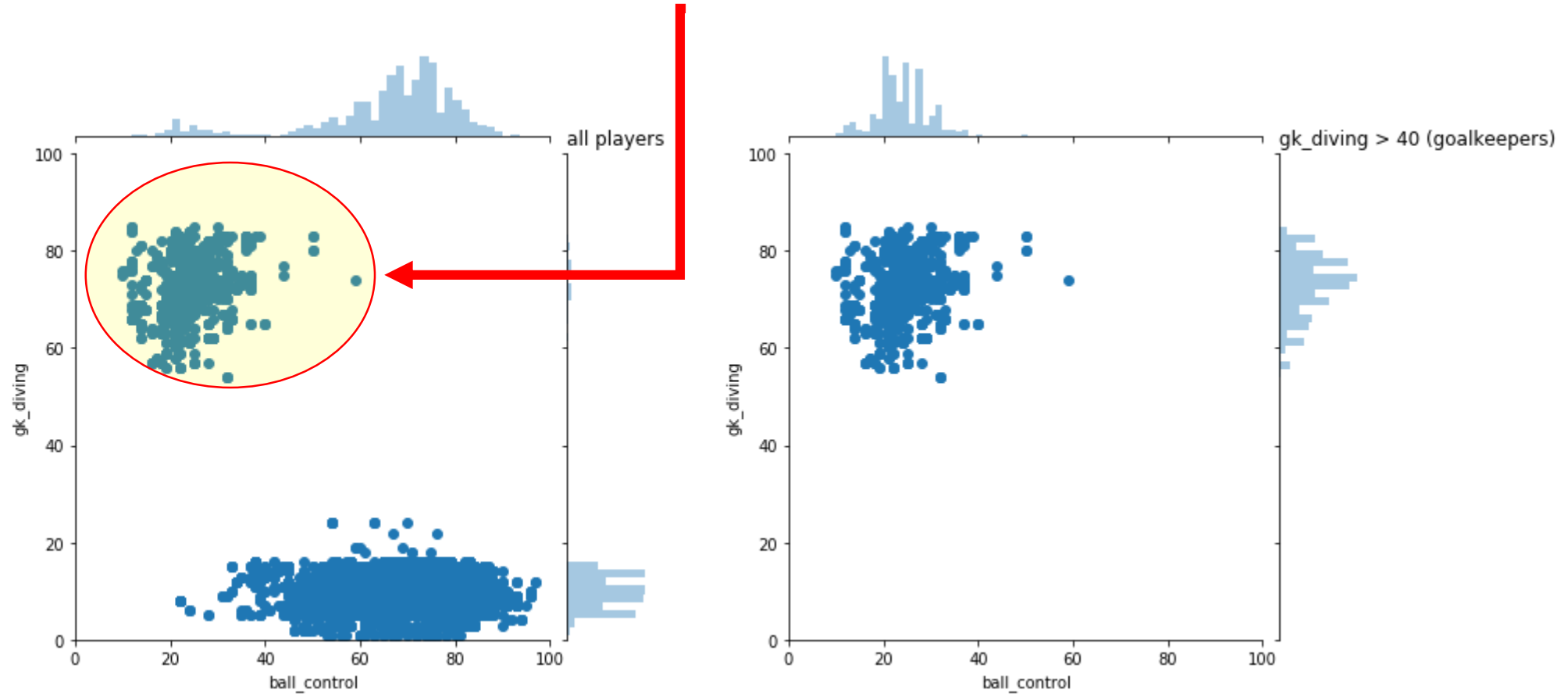


Bimodal Distribution Interceptions (Mentality)



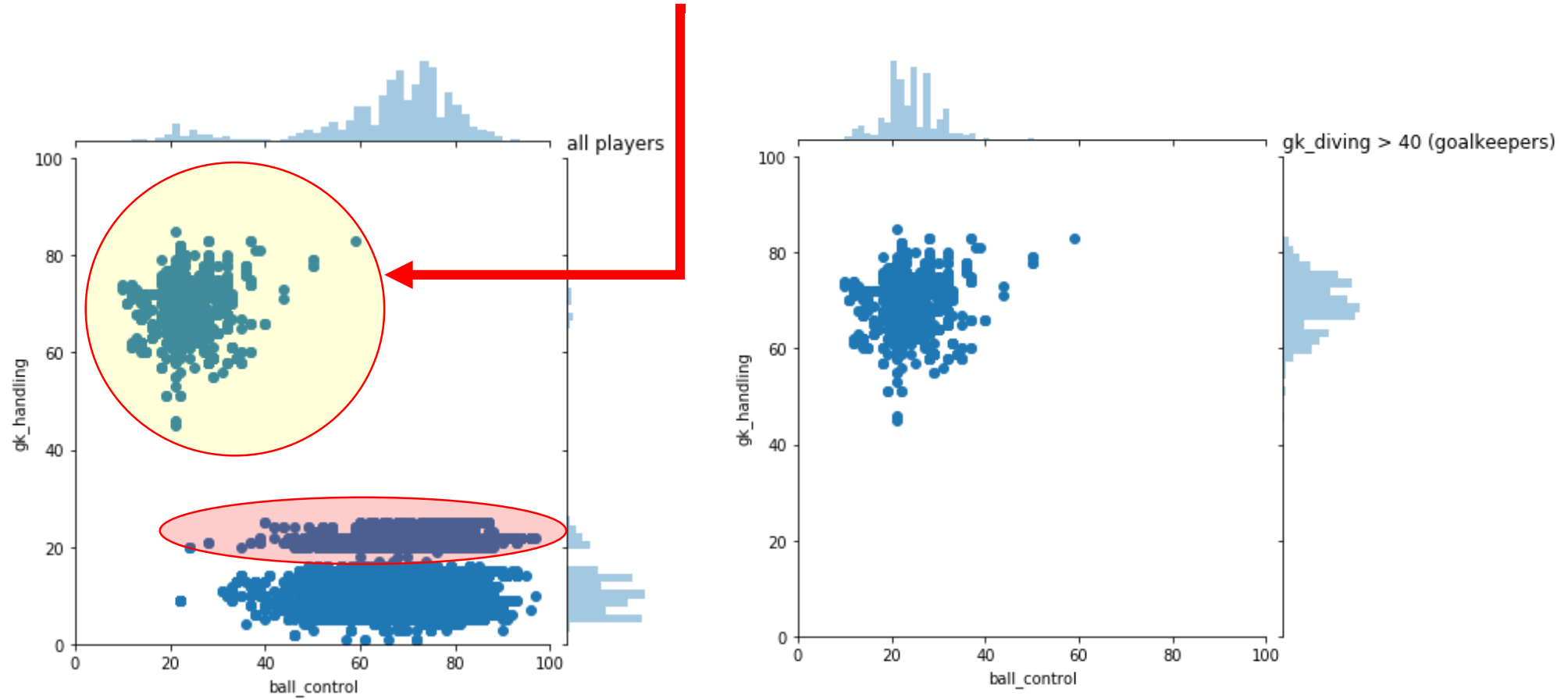
Ball Control Vs gk_Diving

Goalkeepers: gk_diving attributes > 40; relatively lower ball control scores



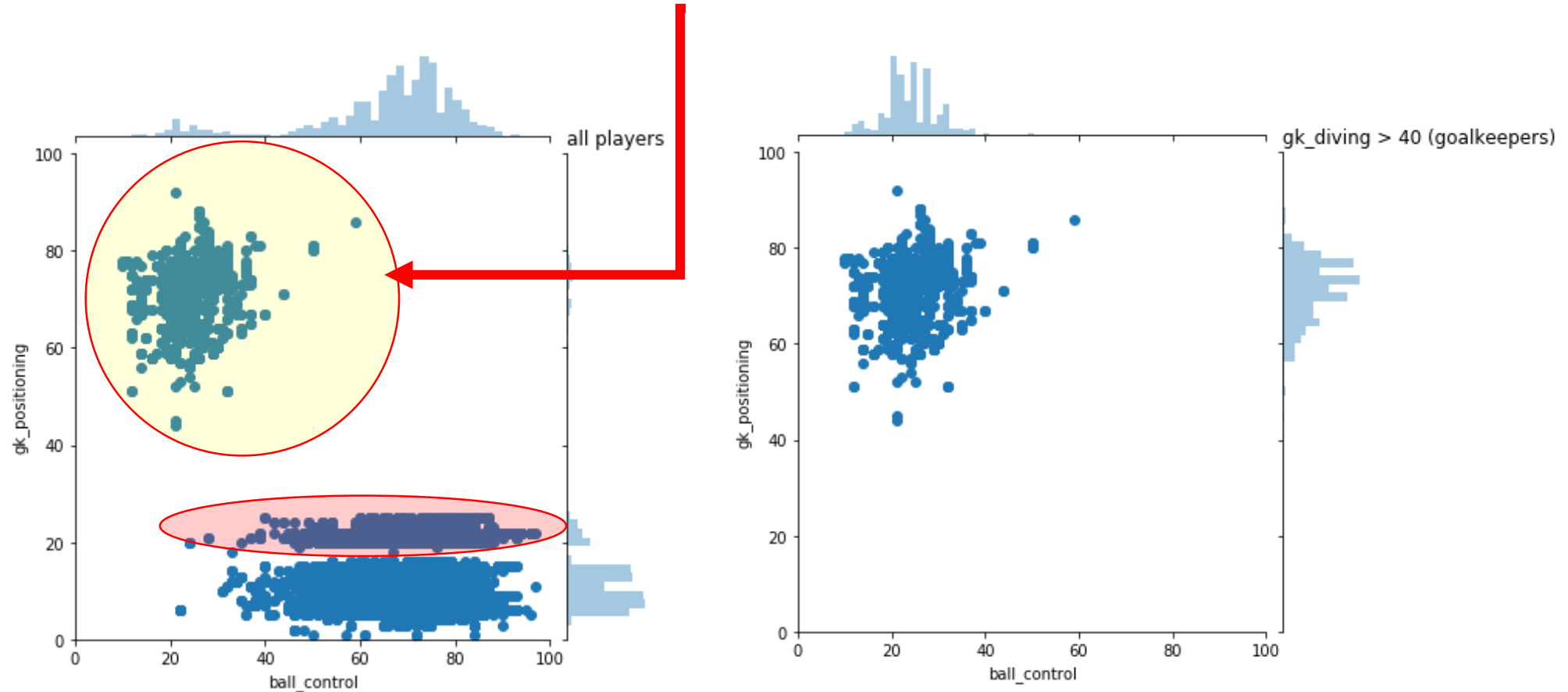
Ball Control Vs gk_Handling

Goalkeepers: gk_handling attributes > 40; relatively lower ball control scores



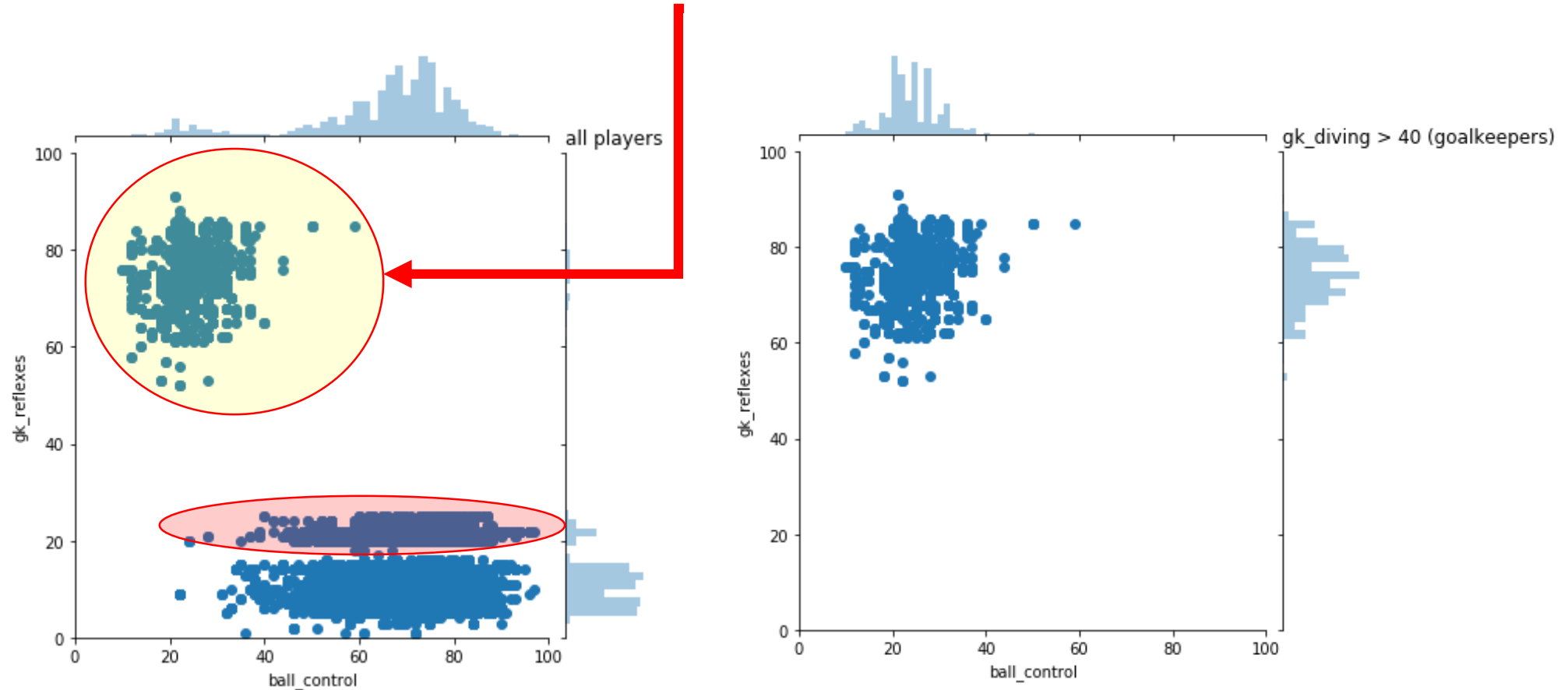
Ball Control Vs gk_Positioning

Goalkeepers: gk_positioning attributes > 40; relatively lower ball control scores



Ball Control Vs gk_Reflexes

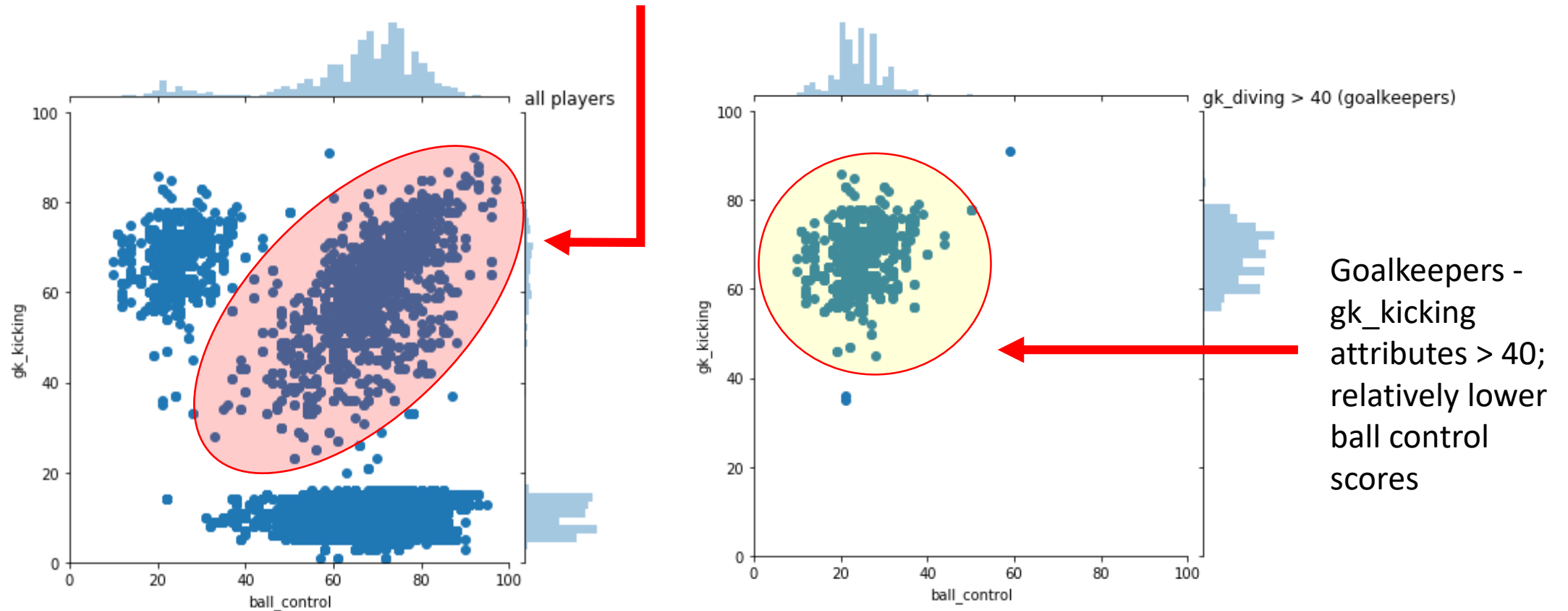
Goalkeepers: gk_reflexes attributes > 40; relatively lower ball control scores



Ball Control Vs gk_Kicking

non-goalkeepers - $\text{gk_diving} < 40$

gk_kicking attribute has a moderate +ve linear correlation with ball control attribute



gk_kicking attribute further separates the bigger subgroup into two smaller subgroups

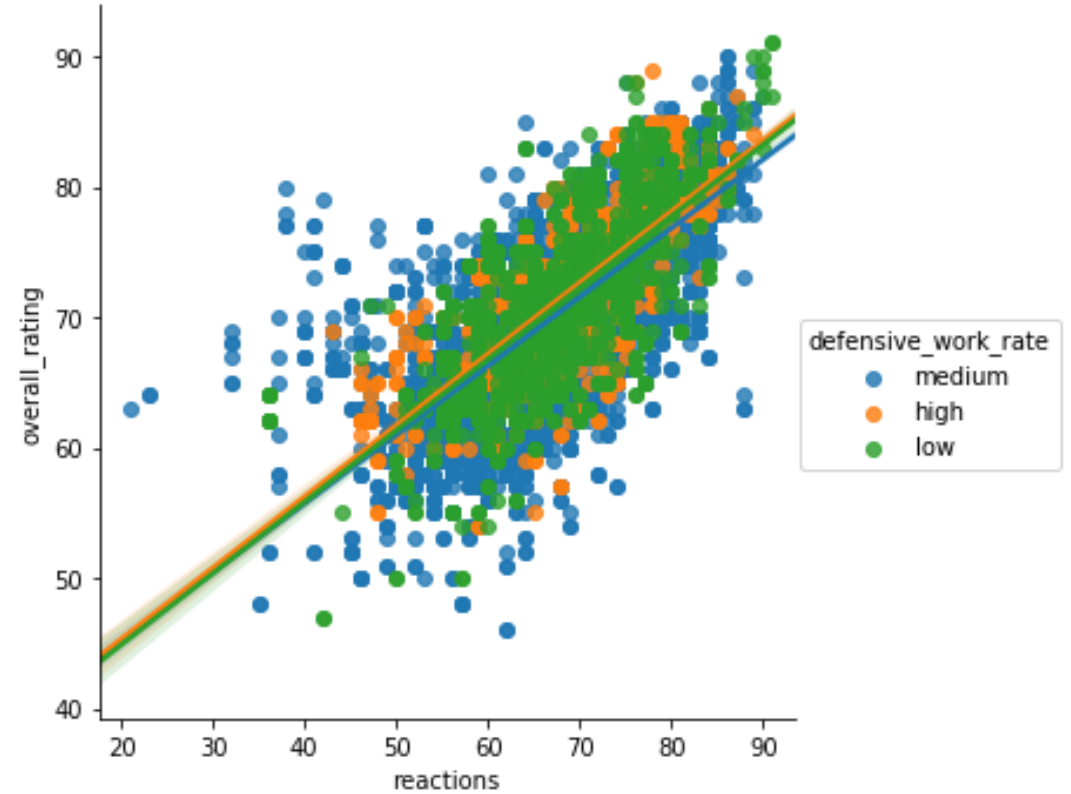
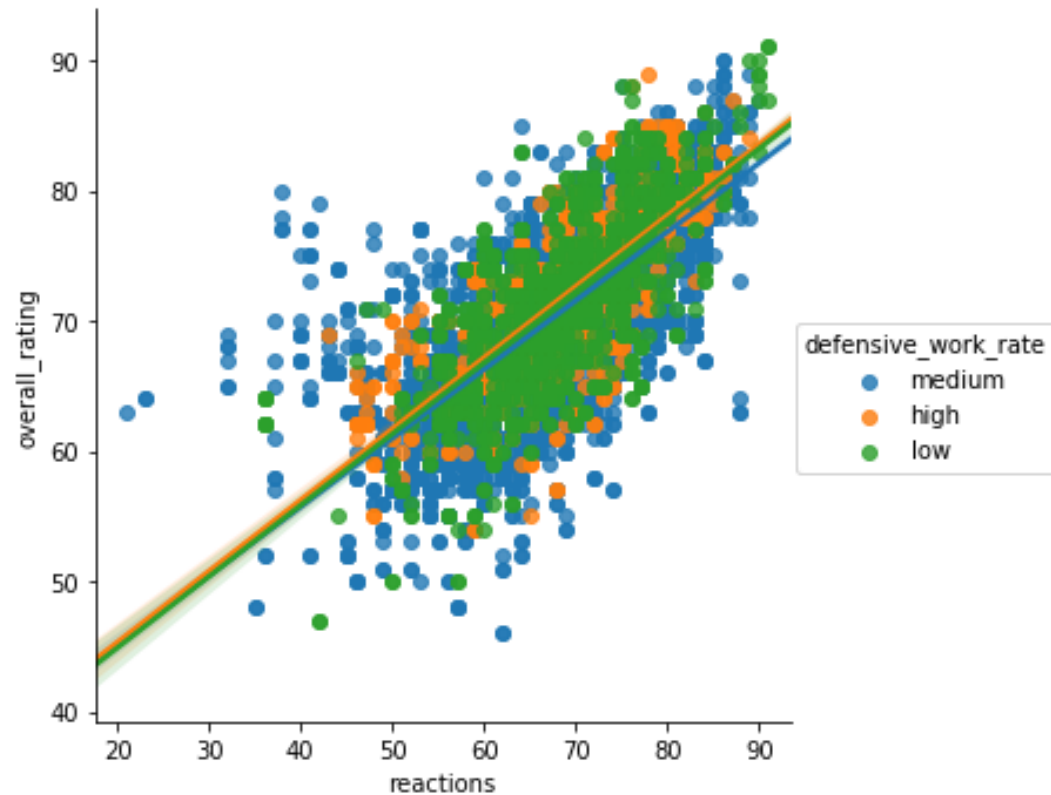
Correlation Coefficients

	overall_rating		overall_rating		overall_rating
age	0.3826	long_passing	0.4300	interceptions	0.2373
height	0.0259	ball_control	0.3726	positioning	0.2782
weight	0.0351	acceleration	0.2137	vision	0.3992
overall_rating	1.0000	sprint_speed	0.2184	penalties	0.3373
potential	<u>0.7840</u>	agility	0.2211	marking	0.1106
crossing	0.2911	reactions	0.7248	standing_tackle	0.1456
finishing	0.2694	balance	0.1094	sliding_tackle	0.1159
heading_accuracy	0.2403	shot_power	0.3703	gk_diving	0.0501
short_passing	0.4161	jumping	0.2112	gk_handling	0.0373
volleys	0.2908	stamina	0.2487	gk_kicking	0.0795
dribbling	0.2991	strength	0.2404	gk_positioning	0.0361
curve	0.2987	long_shots	0.3267	gk_reflexes	0.0437
free_kick_accuracy	0.3048	aggression	0.2615		

- Strong Positive Linear Correlation Between
- *Overall Rating & Reactions: Coeff = 0.7248*
- Overall Rating & Potential: Coeff = 0.7840
- Potential: not attribute, rather as another rating of the players
- Strong correlation explained by rating on similar set of attributes

Overall Rating Versus Reactions Attribute

Strong Positive Linear Correlation: 0.7248

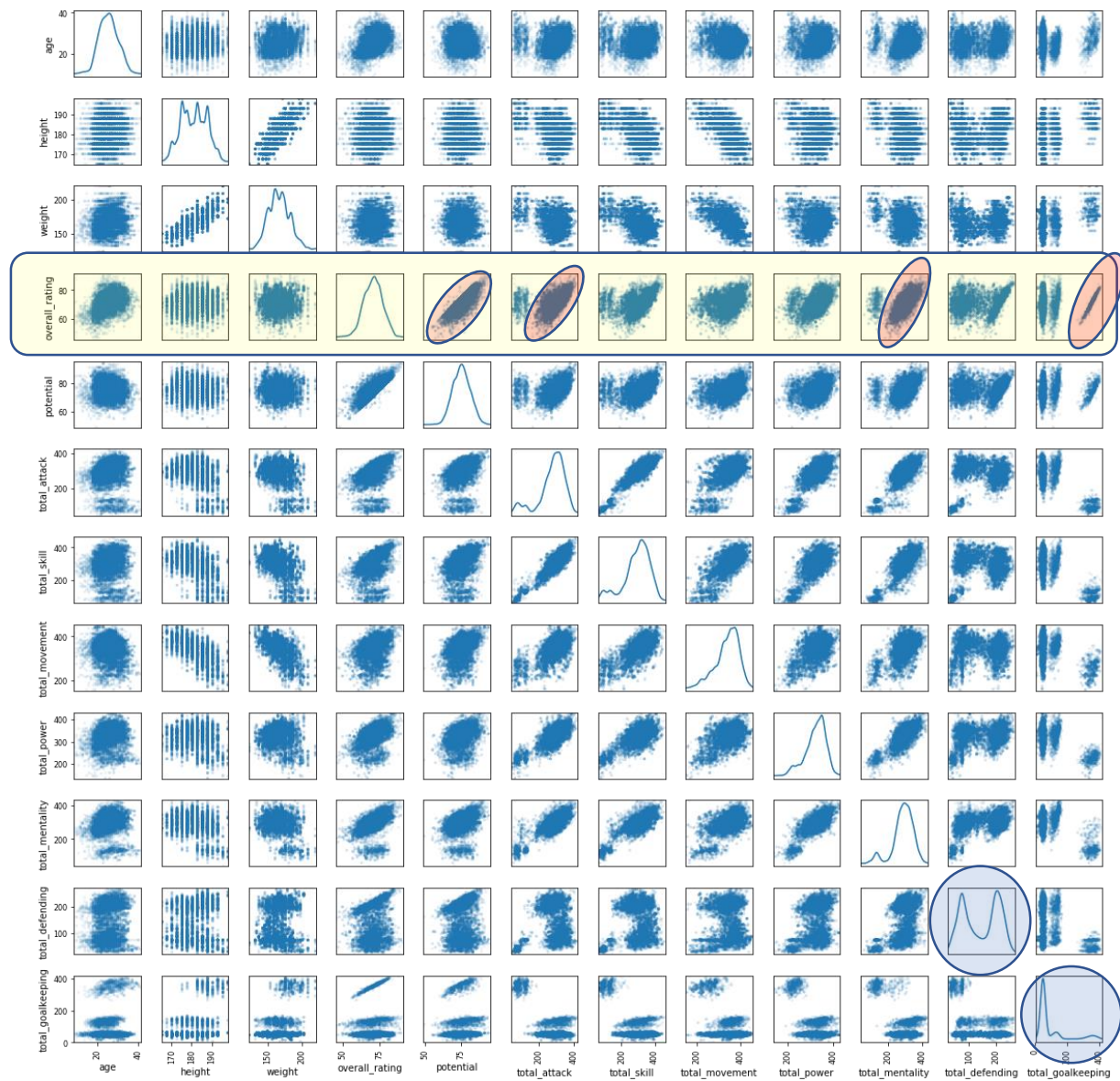


Attribute Categories

Total Attack	Total Skill	Total Movement	Total Power	Total Mentality	Total Defending	Total Goalkeeping
Crossing	Dribbling	Acceleration	Shot Power	Aggression	Marking	gk_Diving
Finishing	Curve	Sprint Speed	Jumping	Interceptions	Standing Tackle	gk_Handling
Heading Accuracy	Free Kicking Accuracy	Agility	Stamina	Positioning	Sliding Tackle	gk_Kicking
Short Passing	Long Passing	Reactions	Strength	Vision		gk_Positioning
Volleys	Ball Control	Balance	Long Shots	Penalties		gk_Refelexes

Scatter Plot Matrix

Overall Rating



Strong + ve
Linear
Correlation
Between
Overall Rating
and
Total
Goalkeeping
Attribute
Only In the
Goalkeeper
Subgroup

Total Defending

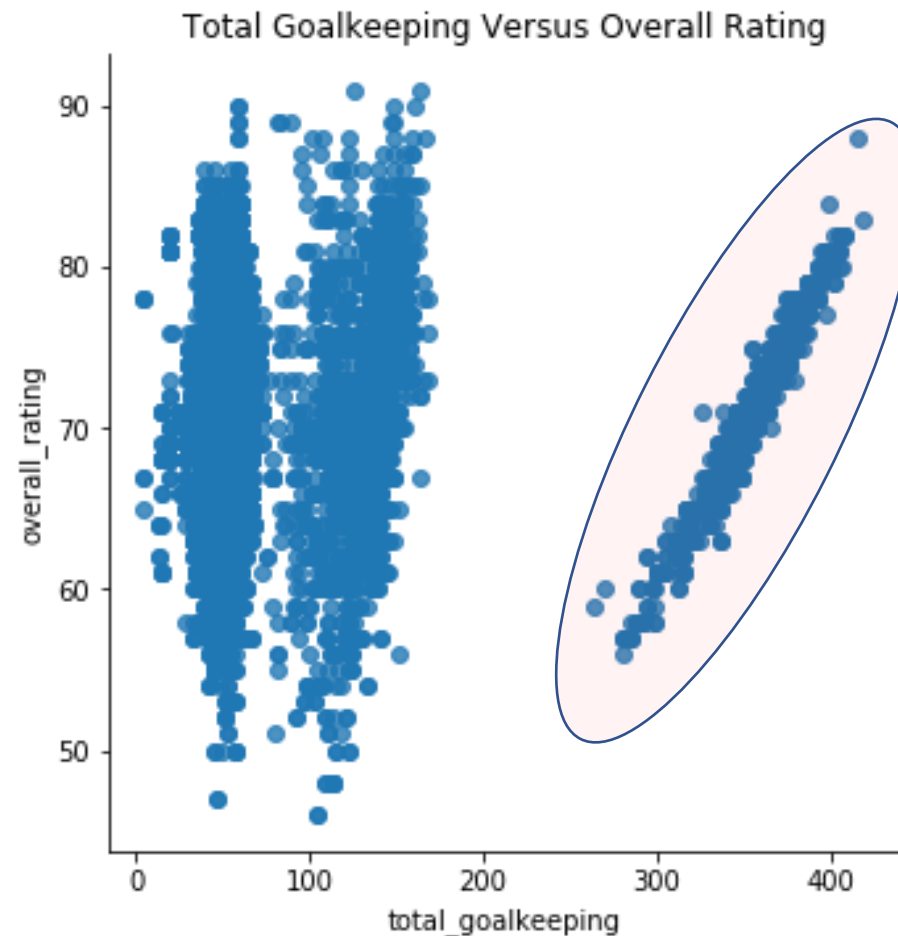
Total Goalkeeping

2 subgroups per Total Defending

1 large and 1 small subgroup per Total Goalkeeping

Goalkeeper Subgroup

Coeff. Overall Rating & Total Goalkeeping



Conclusion 1

1. Reactions attribute has a strong positive linear correlation (coeff. = 0.7248) and contribute most to overall rating
2. Total goalkeeping category has a very strong positive linear correlation (coeff. = 0.9782) with overall rating.

Conclusion 2

1. Defending & Goalkeeping Attributes
 - a) Bimodal distributions
 - b) Set players into subgroups
2. Total Defending Attributes:
Marking, Standing tackle, Sliding tackle
3. Total Goalkeeping Attributes:
gk_Diving, gk_Handling, gk_Kicking,
gk_Positioning, gk_Reflexes

Limitation

Team formation information: <https://sofifa.com/>

Tactics

Build Up Play

Speed 49 (Balanced)

Dribbling 76 (Lots)

Passing 53 (Mixed)

Positioning Organised

Chance Creation

Passing 50 (Normal)

Crossing 56 (Normal)

Shooting 45 (Normal)

Positioning Free Form


Defence

Pressure 67 (High)

Aggression 66 (Press)


Team Width 66 (Normal)

Defender Line Cover



Liverpool (ID: 9)
English Premier League (1)

Overall 81 Attack 85 Midfield 81 Defence 80



Home Stadium Anfield
Rival Team Manchester United
International Prestige 7
Domestic Prestige 9
Transfer Budget €76M
Starting 11 Average Age 25.09
Whole Team Average Age 23.15
Captain J. Henderson
Short Free Kick J. Henderson
Long Free Kick J. Henderson
Penalties M. Salah
Left Corner J. Henderson
Right Corner M. Salah

Like(350) Dislike(63)


Jump to Version ▾



History

Attack Midfield Defence Overall Rating

Limitation

Player Position: <https://sofifa.com/>



L. Karius (ID: 203775)
Loris   GK Age 24 (Jun 22, 1993) 6'2" 192lbs

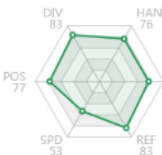
Overall Rating 80


Potential 84

Value €14M

Wage €71K

Preferred Foot **Right**
International Reputation **2** ★
Weak Foot **3** ★
Skill Moves **1** ★
Work Rate **Medium / Medium**
Body Type **Normal**
Real Face **Yes**
Release Clause **€27.7M**













**Liverpool**
81 ★★★★★
Position **GK**
Jersey Number **1**
Joined **Jul 1, 2016**
Contract Valid Until **2021**

Follow (488)

Like (56)

Dislike (5)

Jump to Version ▾

NAME	OVA	POT
 A. Lopes GK	84	86
 Alisson GK	83	89
 W. Szczesny GK	83	85
 Rui Patricio GK	83	83
 G. Donnarumma GK	82	93
 G. Rulli GK	82	87
 Sergio Rico GK	82	87
 Kepa GK	82	87
 Pau Lopez GK	81	86
 J. Butland GK	81	87

Attacking

14

Crossing

18

Finishing

14

Heading Accuracy

36

Short Passing

14

Volleys

Mentality

21

Aggression

13

Interceptions

8

Positioning

48

Vision

19

Penalties

34

Composure

Skill

16

Dribbling

12

Curve

14

FK Accuracy

38

Long Passing

23

Ball Control

Defending

13

Marking

11

Standing Tackle

13

Sliding Tackle

Movement

55

Acceleration

52

Sprint Speed

51

Agility

77

Reactions

43

Balance

Goalkeeping

83

GK Diving

76

GK Handling

76

GK Kicking

77

GK Positioning

83

GK Reflexes

Power

20

Shot Power

74

Jumping

36

Stamina

57

Strength

16

Long Shots

Traits

Puncher

GK Long Throw

GK Rushes Out of Goal