Generally speaking, when we play FIFA, in addition to the simple and direct way to win a strong player, it is extremely important to train players and trade players. The player's potential can't just look at its potential value, because the potential is the upper limit in FIFA. So how do you combine the other indicators to see the true potential of a player? What is the variable space for a player's development? What kind of relationship do they have? We are about to announce this answer through the following procedure.

According to feedback from many of my friends, we found that in the FIFA game, age, overall, sprintspeed, dribbling, passing, shoot, these are several indicators that are more important to a player. Then we will compare the potential values and their relationship against the above indicators.

I found relevant player data on the Kaggle website. The whole data is very complete and the amount of data is very large. The data is in CSV format and has a size of 8.72MB. There are a total of 87 columns. However, some of the players will have some missing data. For the sake of analysis, we set this part of the data to 0.

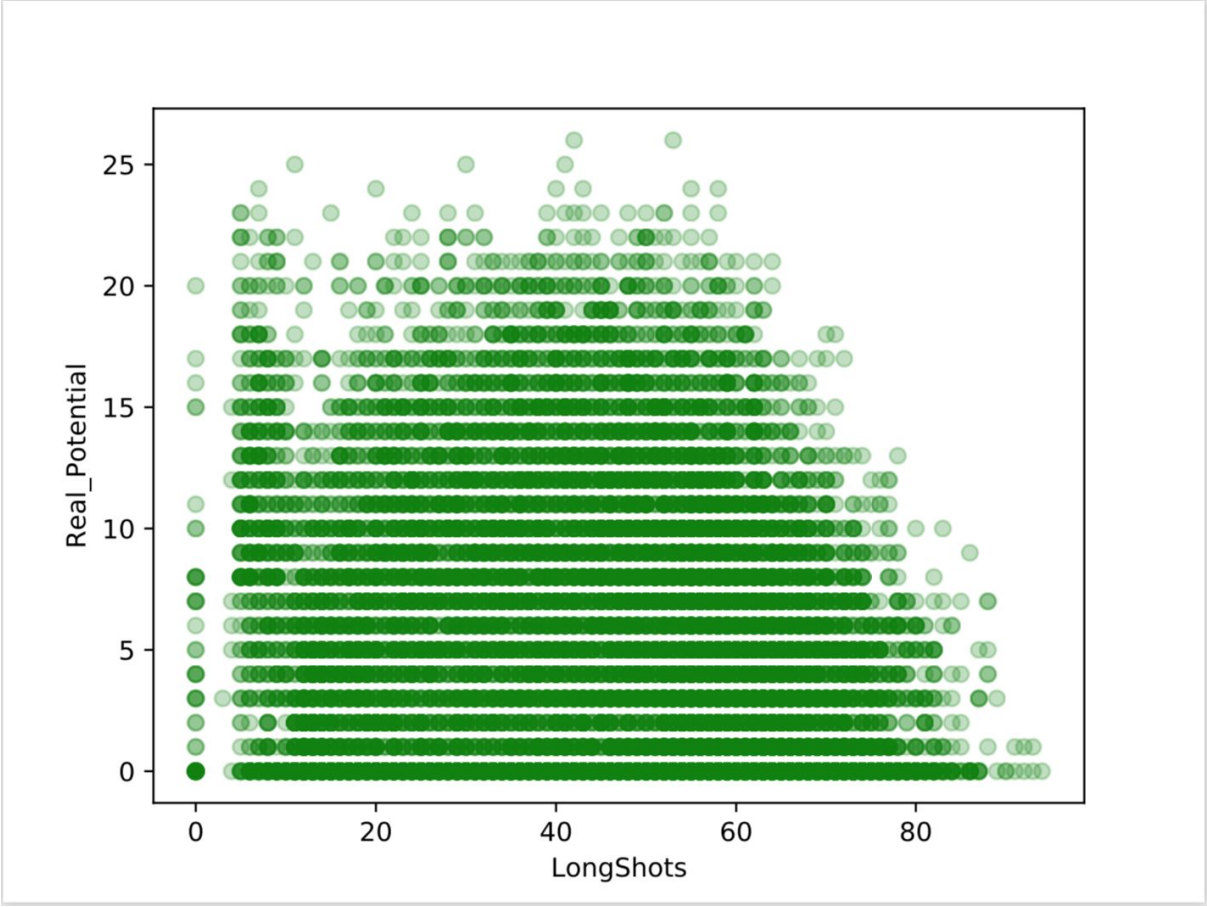
There is only one table in this data, so reading the csv format directly will be faster. However, I also converted it to data in sqlite format. But in any case, in the program, the extracted data exists in the form of a dictionary of list. So it would be easier to read the csv format directly.

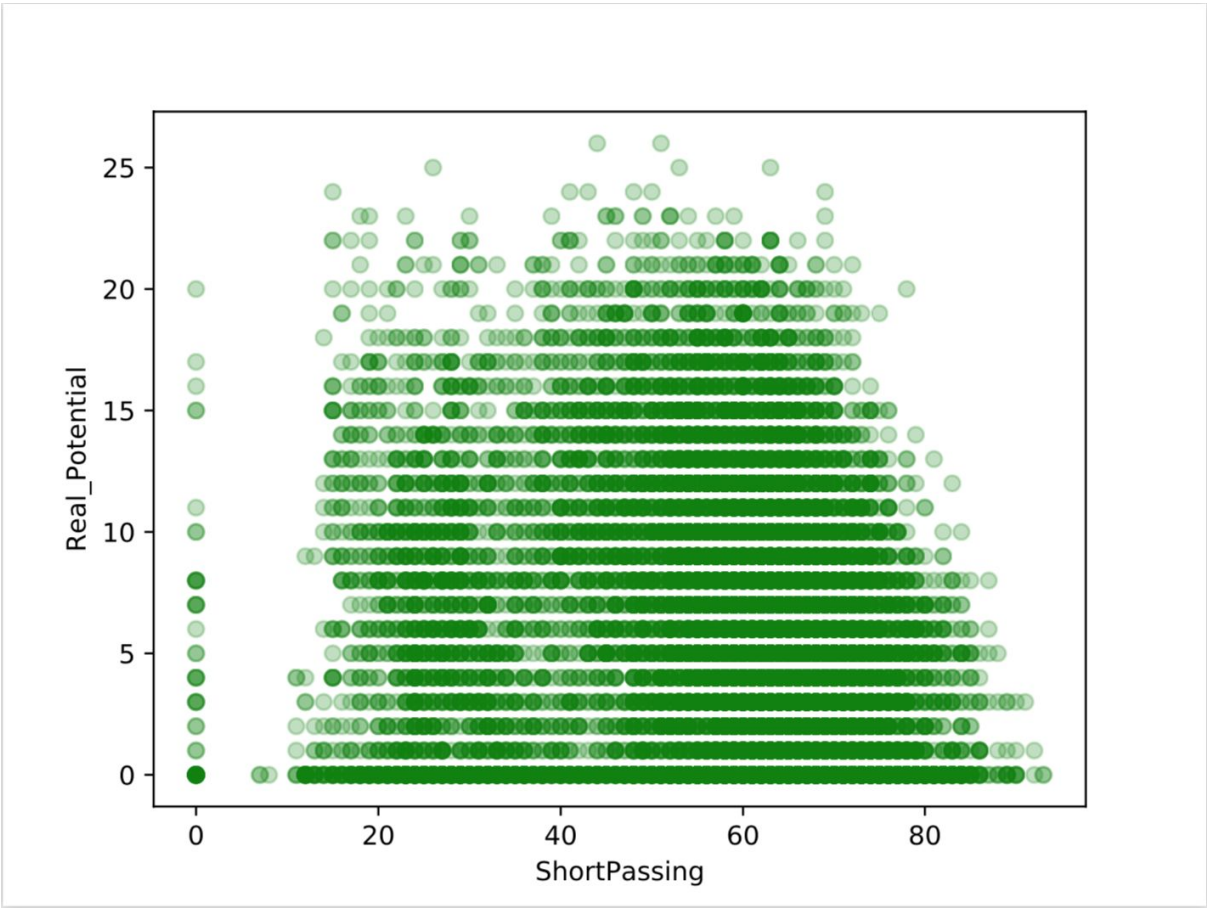
```python
#Programmer: Daiwei Li
#Date: 2019-04-28

import matplotlib.pyplot as plt
import numpy as np
#import pandas as pd
import csv

def read_csv(file_name):
    with open('data.csv') as f:
        reader = csv.reader(f)

    # eliminate blank rows if they exist
        rows = [row for row in reader if row]

        headings = rows[0]
        data = []
        for r in rows[1:]:
            r = [x for x in r]
            d = dict(zip(headings, r))
            data.append(d)

    return data



def find_data(data, bullet):
    raw_data = data
    x = []#data of bullet
    y = []#data of target which is potential
    for index in range(len(raw_data)):
        if raw_data[index][bullet] == "":
            raw_data[index][bullet] = "0"
        x.append(float(raw_data[index][bullet]))
        y.append(float(raw_data[index]["Potential"])-float(raw_data[index]["Overall"]))

    return x,y
```

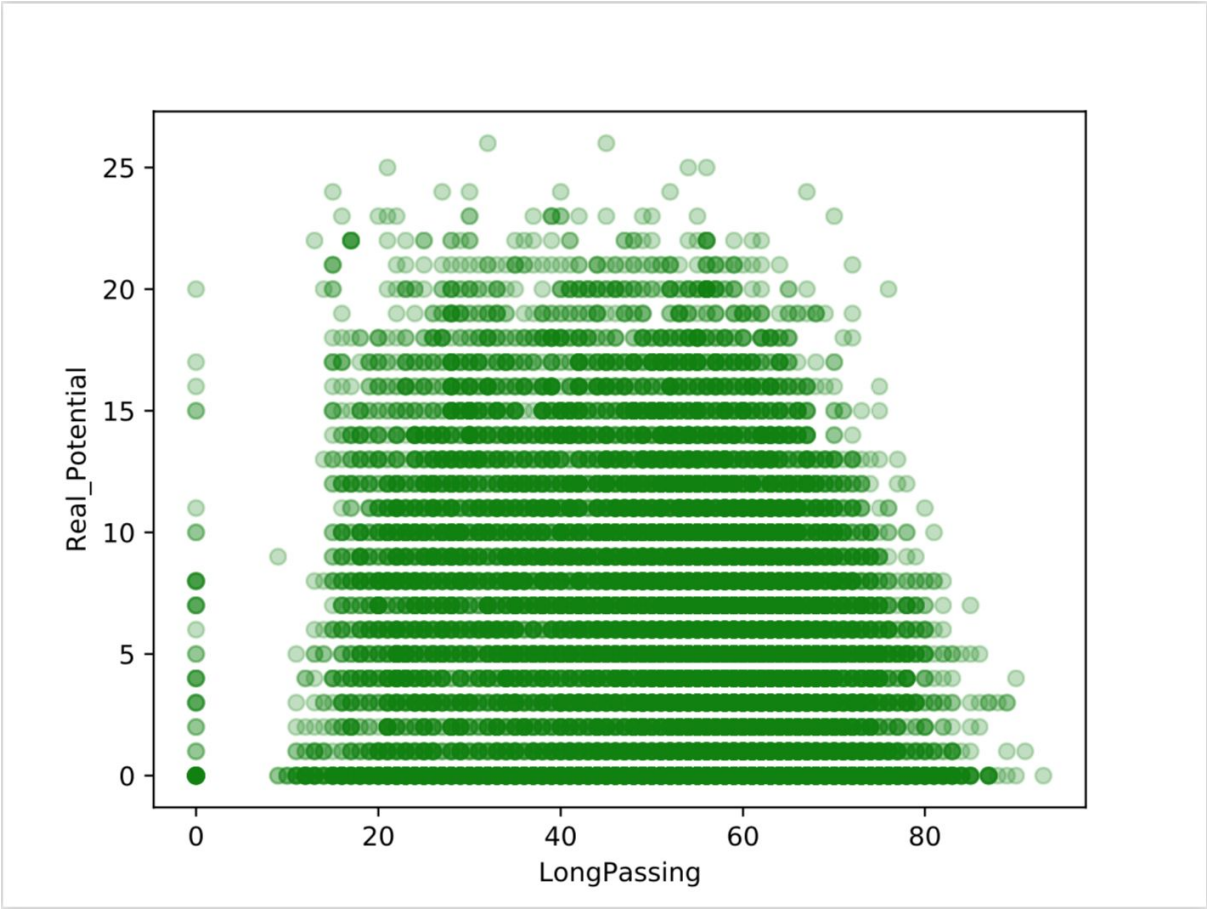
```
def scattering_visualization(x, y, bullet):
    plt.xlabel(bullet)
    plt.ylabel("Real_Potential")
    plt.scatter(np.array(x), np.array(y), c='g', alpha=0.25)
    fname = bullet + ".pdf"
    plt.savefig(fname)
    plt.show()


def main():
    x = []
    y = []
    print("Let analyze soccer player informaiton in FIFA-2019")
    print("We are trying to find all cases could effect a player to improve himself")
    print("In this way, we will see some charts about potential with age, overall rate, sprintspeed, drippling, pass, and shoot")
    print("Please input file name:(/ --> data.csv) ")
    data = read_csv("data.csv")
    #age vs potential
    print("Analyze with potential:(/ --> Age) ")
    x,y = find_data(data, "Age")
    scattering_visualization(x, y, "Age")
    #overall vs potential
    print("Analyze with potential:(/ --> Overall) ")
    x,y = find_data(data, "Overall")
    scattering_visualization(x, y, "Overall")
    #sprintspeed vs potential
    print("Analyze with potential:(/ --> SprintSpeed) ")
    x,y = find_data(data, "SprintSpeed")
    scattering_visualization(x, y, "SprintSpeed")
    #drippling vs potential
    print("Analyze with potential:(/ --> Dribbling) ")
    x,y = find_data(data, "Dribbling")
    scattering_visualization(x, y, "Dribbling")
    #longpass vs potential
    print("Analyze with potential:(/ --> LongPassing) ")
    x,y = find_data(data, "LongPassing")
    scattering_visualization(x, y, "LongPassing")
    #shortpass vs potential
    print("Analyze with potential:(/ --> ShortPassing) ")
```

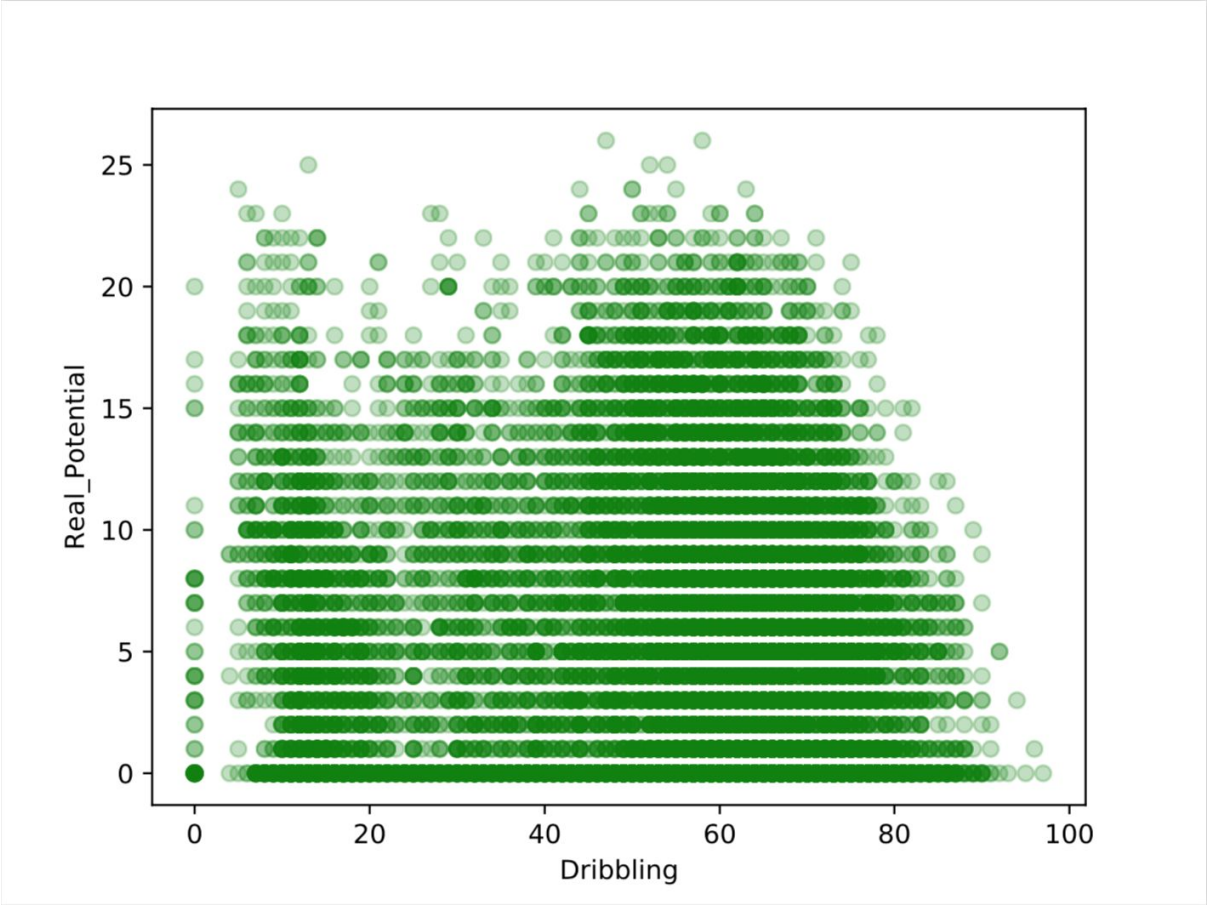
```
    print("Analyze with potential:(/ --> SprintSpeed) ")
    x,y = find_data(data, "SprintSpeed")
    scattering_visualization(x, y, "SprintSpeed")
    #drippling vs potential
    print("Analyze with potential:(/ --> Dribbling) ")
    x,y = find_data(data, "Dribbling")
    scattering_visualization(x, y, "Dribbling")
    #longpass vs potential
    print("Analyze with potential:(/ --> LongPassing) ")
    x,y = find_data(data, "LongPassing")
    scattering_visualization(x, y, "LongPassing")
    #shortpass vs potential
    print("Analyze with potential:(/ --> ShortPassing) ")
    x,y = find_data(data, "ShortPassing")
    scattering_visualization(x, y, "ShortPassing")
    #longshots vs potential
    print("Analyze with potential:(/ --> LongShots) ")
    x,y = find_data(data, "LongShots")
    scattering_visualization(x, y, "LongShots")
    print("All graphs are saved as pdf. Thank you~")
main()
```

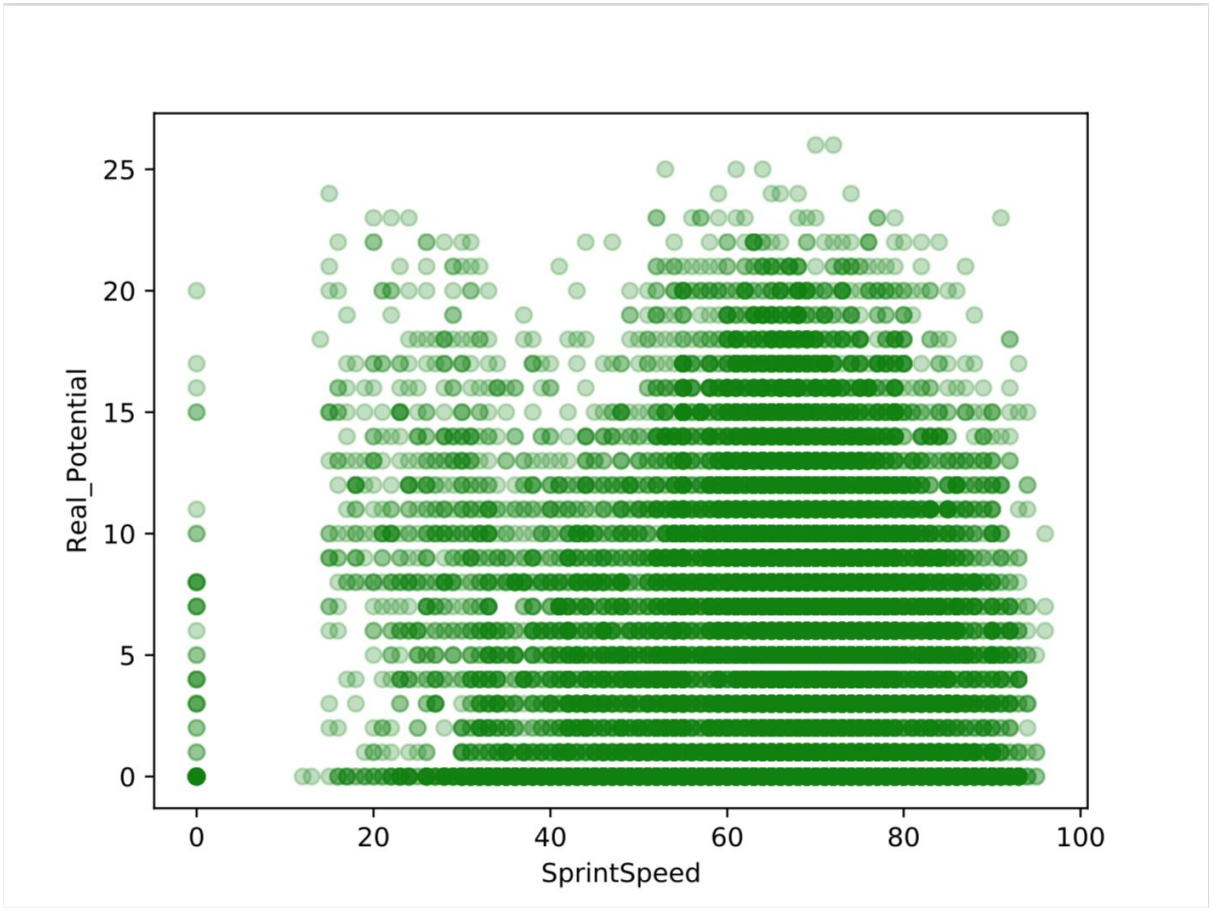
From the above data we have obtained the following diagram. I prefer the scatter plot because we don't know the exact mathematical relationship between the two values. Instead, we formally try to find the relationship between the two. Therefore, the scatter plot will show the overall trend between potential and individual data more clearly.

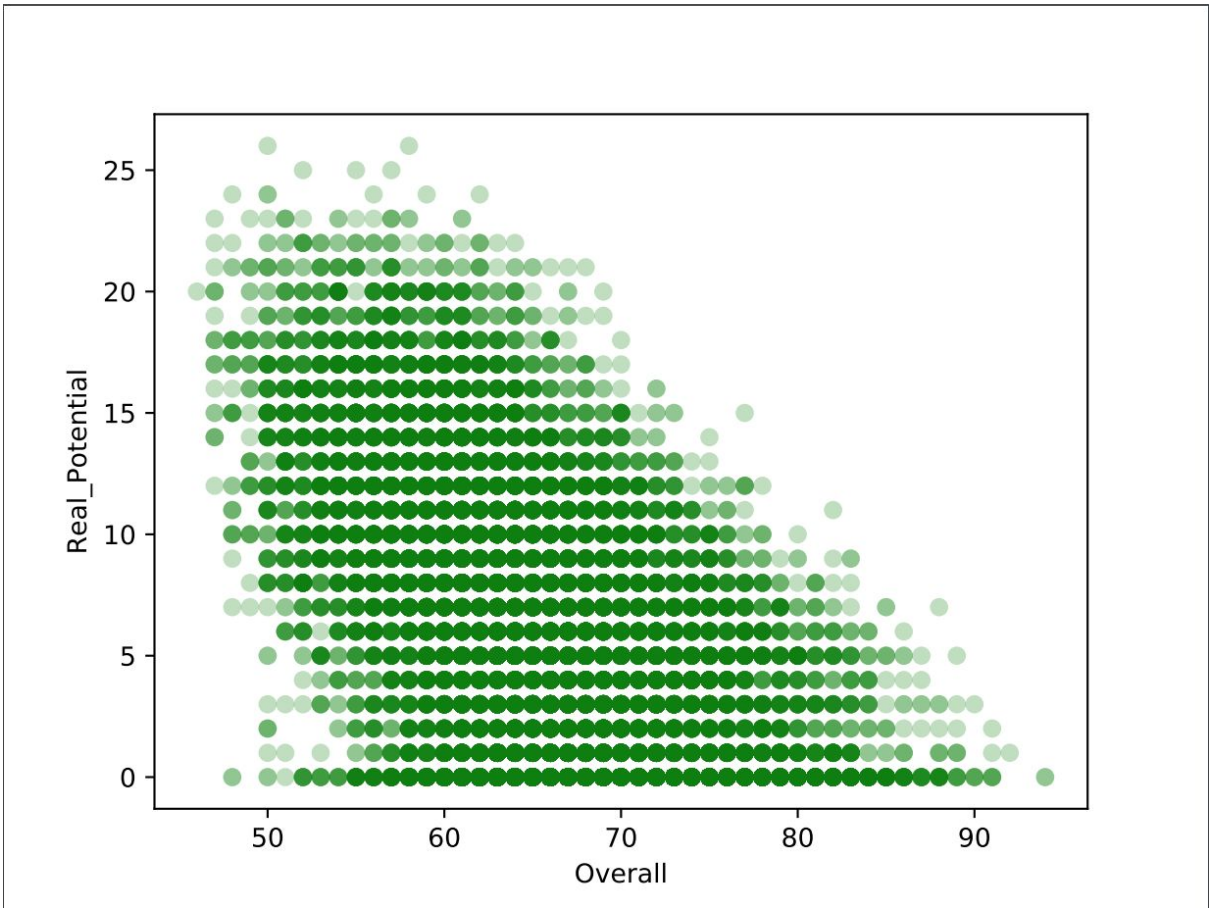


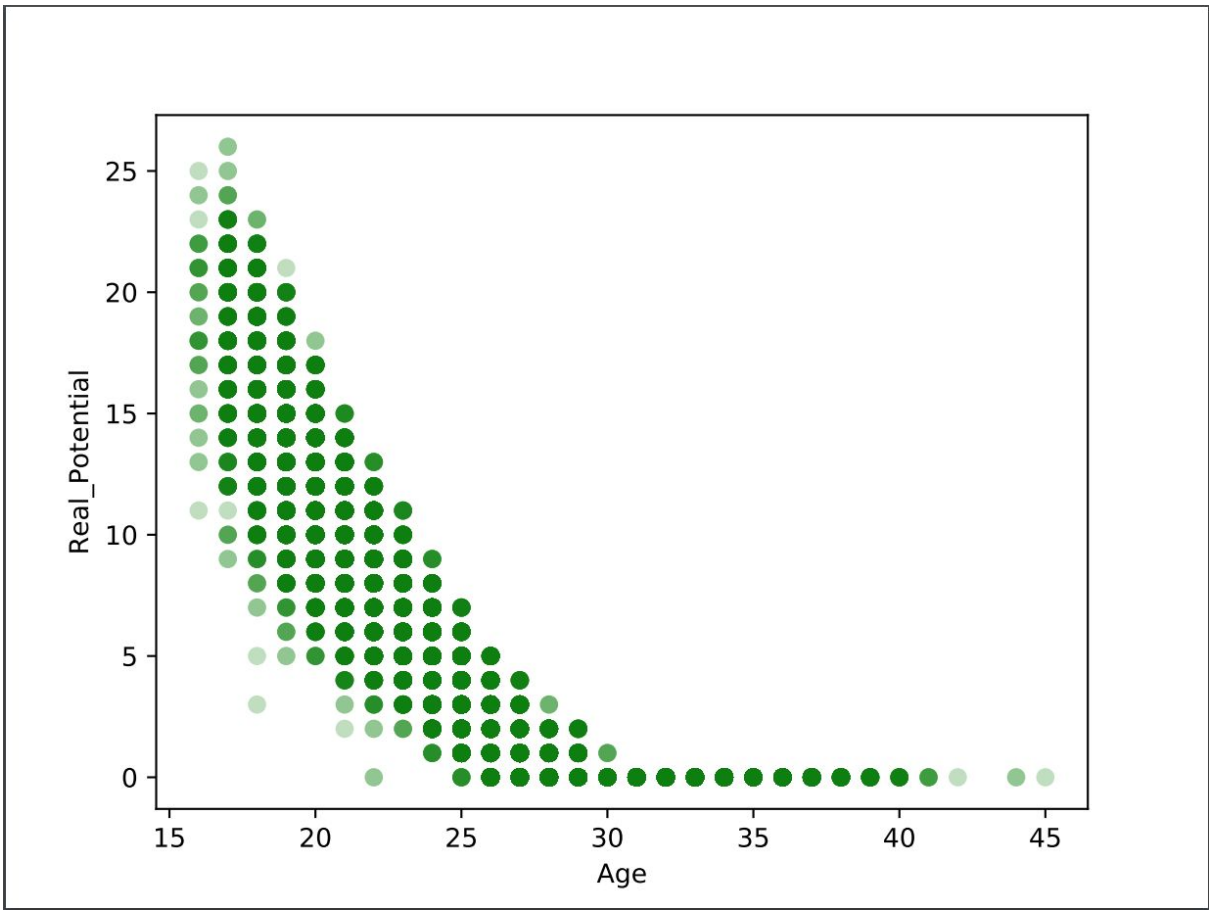

Real_Potential
ShortPassing
0
5
10
15
20
25
0
20
40
60
80


Real_Potential
LongPassing
0
5
10
15
20
25
0
20
40
60
80


Real_Potential
Dribbling
0
5
10
15
20
25
0
20
40
60
80
100


Real_Potential
SprintSpeed
0
5
10
15
20
25
0
20
40
60
80
100


Real_Potential
Overall
0
5
10
15
20
25
50
60
70
80
90



First, we found that age is most closely related to its potential capabilities. According to our analysis of the picture, the potential value may be inversely proportional to the square of the age or even the cubic value. So, you definitely don't want to train a player over the age of 30. Next, the data comparison is characterized by the value of overall. The scatter image of the slope shape represents that it is inversely proportional to the potential capability and the overall itself, that is, a linear function. At the same time, the player who has the stronger ability value is not the space that can be excavated. This is in line with our perception of the usual abilities. Haha! The two data, Dribbling and short passing, are also linear, even if not so obvious. This lateral proof proves that these two values are in a linear relationship with the current ability of a player. The rest of the data does not clearly show some specific trends. Perhaps the conditions we analyzed are not enough, or the rest of the item itself has many other factors, which is very complicated. But in simple terms, we can still conclude that most of the players with high scores in a single item don't have much room for development, because the scatter of the data is generally in the lower right corner. Of course, it is not so concentrated on an extreme point location.