

Data Visualization Project Process Book

- **Basic Info.**
- Project Title : Soccer Stats
- Names : Sravan Kumar Neerati, Sreekanth Reddy Konda
- Emails : u0931101@utah.edu , u1140489@utah.edu
- UID : 0931101 , u1140489
- Project Repository Link : <https://github.com/Meerkat3/SoccerStats>

Overview: Soccer stats is a data visualization designed to see various details of top 50 soccer players and teams in various clubs. This includes two visualizations.

1. Showing various attributes of the player, how is his performance over the years, comparison of players performance based on different attributes.
2. Showing a team's performance in a club over the years and the points scored by each club in each season.

Background and Motivation: Though there are many visualizations online which shows these data, we couldn't find a place where we can compare multiple players based on their score in various areas like dribbling, free kick accuracy etc.

We are trying to answer the above questions. Our visualization is different in showing as many as 20 attributes in various areas to see a player's performance. Also we are trying to provide a comparison tool to compare two or more player's performance based on different attributes.

Project Objectives: We aim to show the highly sought after mainstream Soccer stats in overview and details of the Player attributes and their performance over time, Insights of a club's performance and how the betting fared on the matches(tentative).

Main questions we are trying to answer:

1. What are the basic details and latest scores of a player in different areas like penalties, marking, dribbling etc.
2. How well is a player improving in various areas over the years?
3. Comparing two or more player's performance over the years?
4. Points scored by a club in a league per season?
5. Comparison of club's performances in specific leagues over the years?

Related Work and Inspiration:

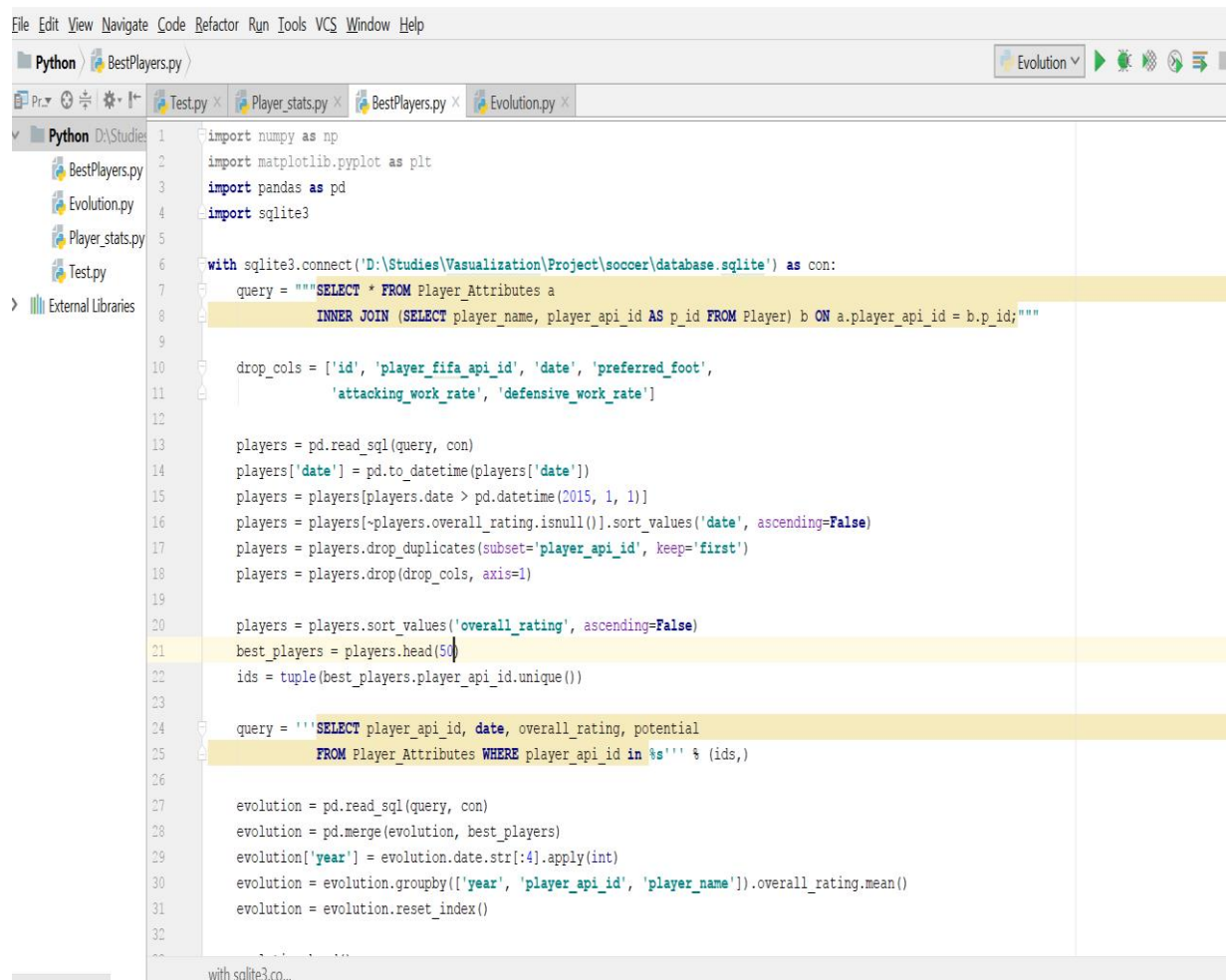
Data: The data is presented in SQLite format and we intend to do some pre-processing before storing them as csv/json for our final usage. If time permits we intend to have a backend server, serving the data.

- Link: <https://www.kaggle.com/hugomathien/soccer/data>

Data Processing: We initially thought there won't be much processing needed as we have all the data in various tables in SQLite. But as we started writing python code to deduce the data we realized that we needed many joins among tables. Also there were multiple records corresponding to single year. We had to clean all this data and finally saved them in different csv's which makes it easy to use them in our visualization.

We have implemented python and R code to retrieve relevant data from sqlite format.

Sample codes below:



```
File Edit View Navigate Code Refactor Run Tools VCS Window Help
Python BestPlayers.py
Test.py Player_stats.py BestPlayers.py Evolution.py
Python D:\Studies
BestPlayers.py
Evolution.py
Player_stats.py
Test.py
> External Libraries
1 import numpy as np
2 import matplotlib.pyplot as plt
3 import pandas as pd
4 import sqlite3
5
6 with sqlite3.connect('D:\Studies\Visualization\Project\soccer\database.sqlite') as con:
7     query = """SELECT * FROM Player_Attributes a
8             INNER JOIN (SELECT player_name, player_api_id AS p_id FROM Player) b ON a.player_api_id = b.p_id;"""
9
10
11     drop_cols = ['id', 'player_fifa_api_id', 'date', 'preferred_foot',
12                 'attacking_work_rate', 'defensive_work_rate']
13
14     players = pd.read_sql(query, con)
15     players['date'] = pd.to_datetime(players['date'])
16     players = players[players.date > pd.datetime(2015, 1, 1)]
17     players = players[~players.overall_rating.isnull()].sort_values('date', ascending=False)
18     players = players.drop_duplicates(subset='player_api_id', keep='first')
19     players = players.drop(drop_cols, axis=1)
20
21     players = players.sort_values('overall_rating', ascending=False)
22     best_players = players.head(50)
23     ids = tuple(best_players.player_api_id.unique())
24
25     query = '''SELECT player_api_id, date, overall_rating, potential
26             FROM Player_Attributes WHERE player_api_id in %s''' % (ids,)
27
28     evolution = pd.read_sql(query, con)
29     evolution = pd.merge(evolution, best_players)
30     evolution['year'] = evolution.date.str[:4].apply(int)
31     evolution = evolution.groupby(['year', 'player_api_id', 'player_name']).overall_rating.mean()
32     evolution = evolution.reset_index()
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
with sqlite3.co...
```

```

Evolution.py
Test.py Player_stats.py BestPlayers.py Evolution.py
4 import sqlite3
5
6 with sqlite3.connect('D:\Studies\Visualization\Project\soccer\database.sqlite') as con:
7     query = """SELECT * FROM Player_Attributes a
8             INNER JOIN (SELECT player_name, birthday, height, weight, player_api_id AS p_id FROM Player) b ON a.player_api_id = b.p_id;"""
9     drop_cols = ['id', 'player_fifa_api_id', 'date', 'preferred_foot',
10                'attacking_work_rate', 'defensive_work_rate']
11
12 players = pd.read_sql(query, con)
13 players['date'] = pd.to_datetime(players['date'])
14 players = players[players.date > pd.datetime(2015, 1, 1)]
15 players = players[~players.overall_rating.isnull()].sort_values('date', ascending=False)
16 players = players.drop_duplicates(subset='player_api_id', keep='first')
17 players = players.drop(drop_cols, axis=1)
18
19 players = players.sort_values('overall_rating', ascending=False)
20 best_players = players[['player_api_id', 'player_name', 'birthday', 'height', 'weight']].head(50)
21 ids = tuple(best_players.player_api_id.unique())
22
23 query = '''SELECT player_api_id, date, overall_rating, crossing, finishing,
24            heading_accuracy,
25            short_passing,
26            dribbling,
27            volleys,
28            free_kick_accuracy,
29            long_passing,
30            ball_control,
31            acceleration,
32            sprint_speed,
33            balance,
34            jumping,
35            stamina,

```

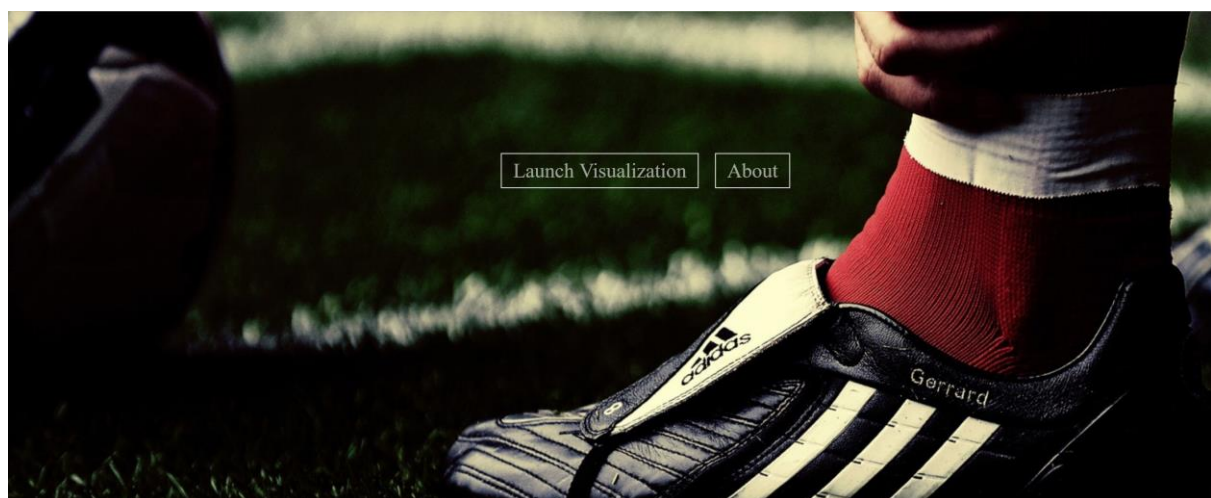
Design Evaluation and Implementation:

We originally planned to show all the data belonging to top 50 players at once. But in feed back session we got a very good suggestion that helps in comparing players one by one. i.e we initially show that data of one player, then we can have the ability to include more players into that visualization. This will keep on adding multiple lines to the line chart to show the trend of all the selected players over the years. The following picture shows the trends of two players selected (current stage which doesn't involve selection of players yet.)

We will have 2 tabs in the site. One for player specific visualizations and another for clubs related data visualizations.

Landing Page :

We have improvised on the initial design by adding an extra launch page which provides the motivation for picking the dataset and insights for the same. It summarizes the whole visualizations beautifully and sets the mood for the end user.



Clicking on "About" show the mentioned summary and provides link to the demo video

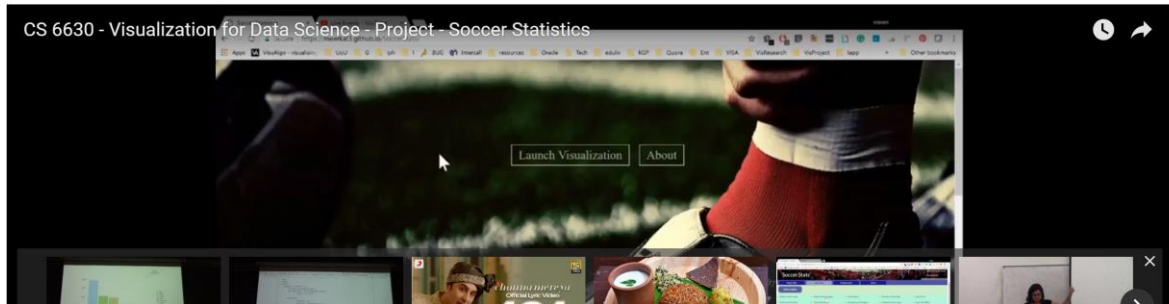
Soccer Statistics

As passionate Soccer Enthusiasts and avid Football fans, the two of us, Sravan Kumar Neerati and Sreekanth Reddy Konda, set out to make a visualization to celebrate the success of the top Soccer players and Clubs around the world. Soccer fandom has increased tremendously over the years, and the two of us have been lucky to witness the growth of players and clubs from different parts of the world. We hope our project can be both enjoyed by all other soccer enthusiasts and provide helpful insights to the avid followers and fans of the players and clubs.

To launch our visualization, simply scroll up to the top of the page and press the 'Launch Visualization' button. Select a player, to see his summary and ratings in different attributes, scroll through the timeline, and to compare the performance over the years with another player, search and select the player(s) from the list. The color codings help differentiate the players from one another, the color of the player name added and the line help relate the same player. When we wish to compare a lot of players and are interested in specific years rather than the while timeline, one can brush along the x-axis timeline and see the bar charts with more emphasis on the ratings for that particular attribute. Feel Free to add as many players as possible to the line chart and brush on the x-axis timeline to zoom into the specific years. Although a player has been performing really well in a specific attribute, he might be lagging other attributes of the game and that can be seen clearly in the bar charts. Hovering over the bars or the lines will show the details that are specific to the visualization and get more insights

Our Visualization also dwells into how the clubs from different leagues have performed over the seasons. The total points scored in a season are color coded which puts up a heatmap that clearly shows the trend in a league.

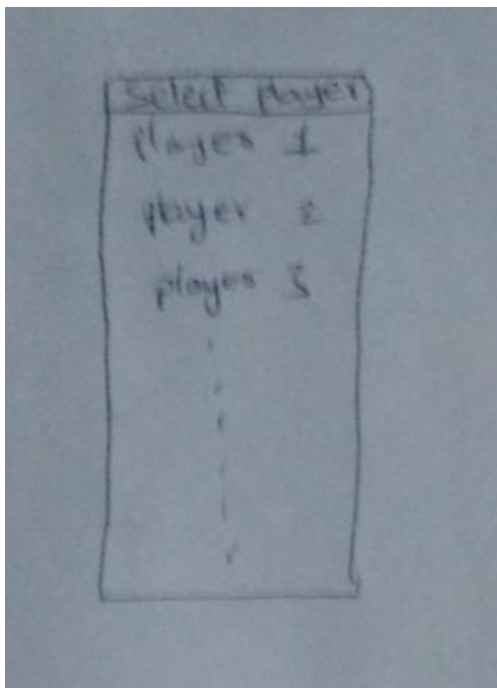
While not all the clubs are not present in all seasons, the reasons for that can be clearly seen from its performance in the previous year. At the same time, we can see the top clubs being highlighted as a result of their consistent performances over the seasons.



Clicking on “Launch visualization” would take the end user to the main visualization.

View 1: (Selection of player)

We initially think about having a drop down to select the player. The following picture describes this



Then we came to know that it is difficult to select a player from drop down by scrolling in a small section. Then we decided to have a button which on click shows the names of all 50 players at one place, which makes it easy for the user to select.

Player

Select Player

- Alexis Sanchez
- Carlos Tevez
- Diego Costa
- Gerard Pique
- Ivan Rakitic
- Marco Reus
- Paul Pogba
- Sergio Aguero
- Wayne Rooney
- Luka Modric
- Andres Iniesta
- Cesc Fabregas
- Diego Godin
- Giorgio Chiellini
- James Rodriguez
- Mats Hummels
- Petr Cech
- Sergio Busquets
- Yaya Toure
- Thiago Silva
- Angel Di Maria
- Cristiano Ronaldo
- Eden Hazard
- Gonzalo Higuain
- Jerome Boateng
- Mesut Oezil
- Philipp Lahm
- Sergio Ramos
- Karim Benzema
- Thibaut Courtois
- Arjen Robben
- David De Gea
- Franck Ribery
- Henrik Mkhitaryan
- Kevin de Bruyne
- Miranda
- Robert Lewandowski
- Thomas Mueller
- Lionel Messi
- Xavi Hernandez
- Bastian Schweinsteiger
- David Silva
- Gareth Bale
- Hugo Lloris
- Manuel Neuer
- Neymar
- Santi Cazorla
- Toni Kroos
- Luis Suarez
- Zlatan Ibrahimovic

We have improvised on this as below to enhance user experience and make the transition seamless

Choose another player

- Alexis Sanchez
- Carlos Tevez
- Diego Costa
- Gerard Pique
- Ivan Rakitic
- Marco Reus
- Paul Pogba
- Sergio Aguero
- Wayne Rooney
- Luka Modric
- Andres Iniesta
- Cesc Fabregas
- Diego Godin
- Giorgio Chiellini
- James Rodriguez
- Mats Hummels
- Petr Cech
- Sergio Busquets
- Yaya Toure
- Thiago Silva
- Angel Di Maria
- Cristiano Ronaldo
- Eden Hazard
- Gonzalo Higuain
- Jerome Boateng
- Mesut Oezil
- Philipp Lahm
- Sergio Ramos
- Karim Benzema
- Thibaut Courtois
- Arjen Robben
- David De Gea
- Franck Ribery
- Henrik Mkhitaryan
- Kevin de Bruyne
- Miranda
- Robert Lewandowski
- Thomas Mueller
- Lionel Messi
- Xavi Hernandez
- Bastian Schweinsteiger
- David Silva
- Gareth Bale
- Hugo Lloris
- Manuel Neuer
- Neymar
- Santi Cazorla
- Toni Kroos
- Luis Suarez
- Zlatan Ibrahimovic

View 2 (basic details of the player):

We are decided to show the picture of the player upon selection, basic details and the details of the points (latest) he scored in various areas. We decided to go with the radar chart to show these attributes. This helps us in deciding in what areas the player is strong/weak. For example, we can clearly see that Lionel Messi is good in attacking attributes.



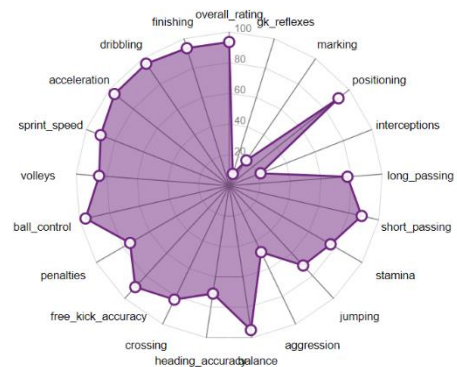
name : Lionel Messi

birthday : 11-May-92

height : 198.12

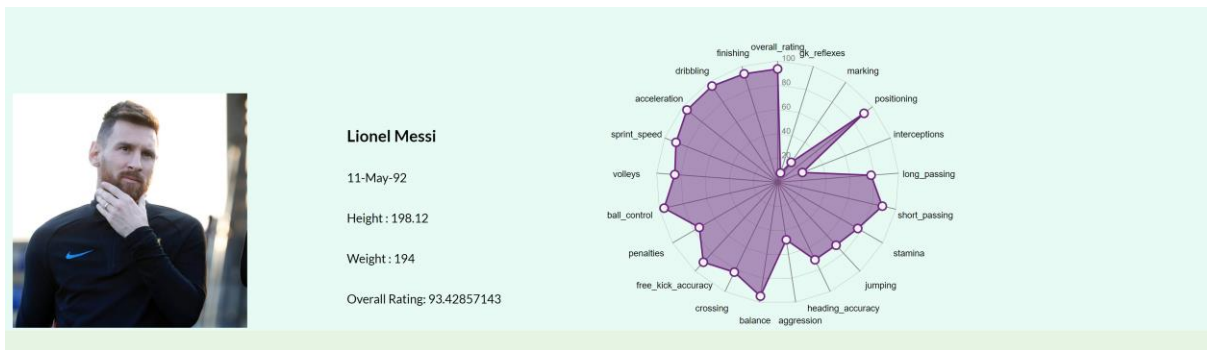
weight : 194

overall_rating: 93.42857143



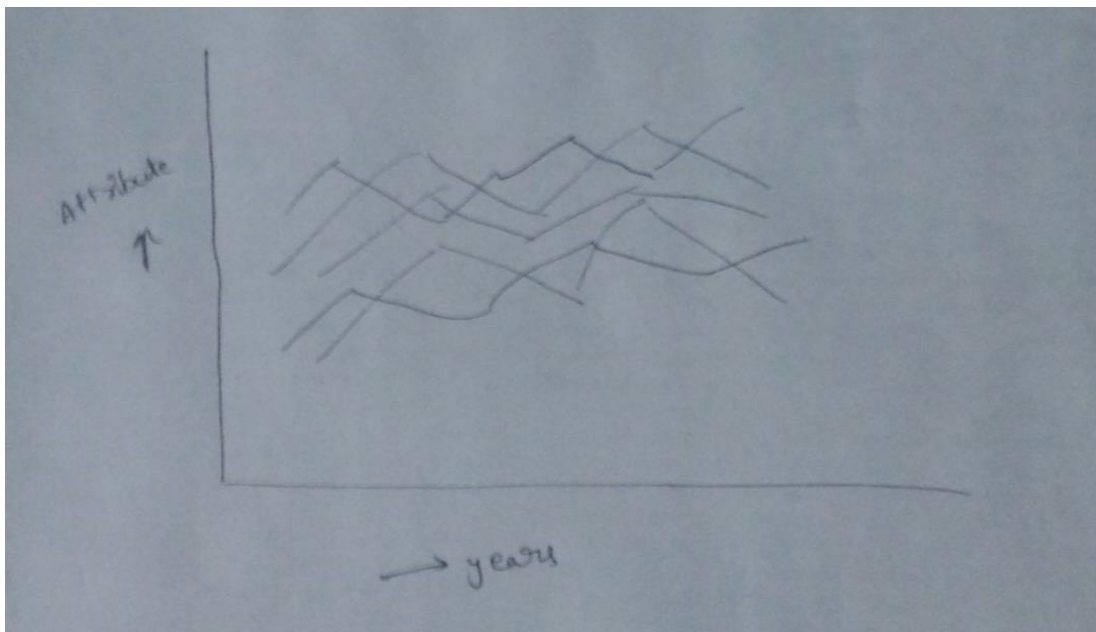
Like wise, we can see in what attributes a player is good at like attacking, defending, goal keeping etc.

We have improvised on the player details making it more visually appealing.

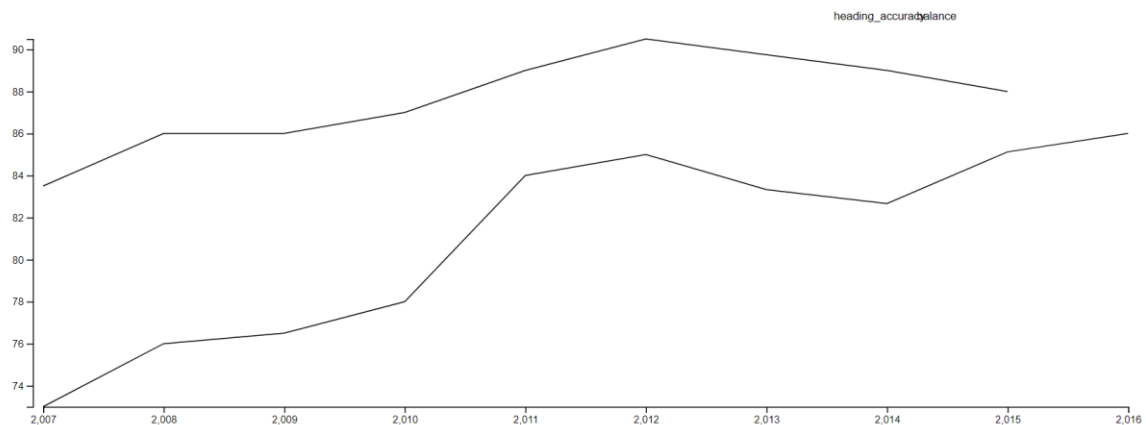


View 3:

Initially we thought to have all the data corresponding to 50 players in a line chart to show how their performance was changing over the years. Then we thought to have an interaction, when the user hovers over the line, we will show the information of that particular player. Following picture helps to describe this.



Then we changed it to initially have the data corresponding to a single selected player from view 1. Also the user will have the facility to add one more player to compare. For this, we are thinking to have a search box, when user types part of the player's name, we show the suggestions. Once user clicks the player's name, we will add one more line to the line chart. Colours are used to distinguish between players. Also we will have a tooltip to show player's information on mouseover.

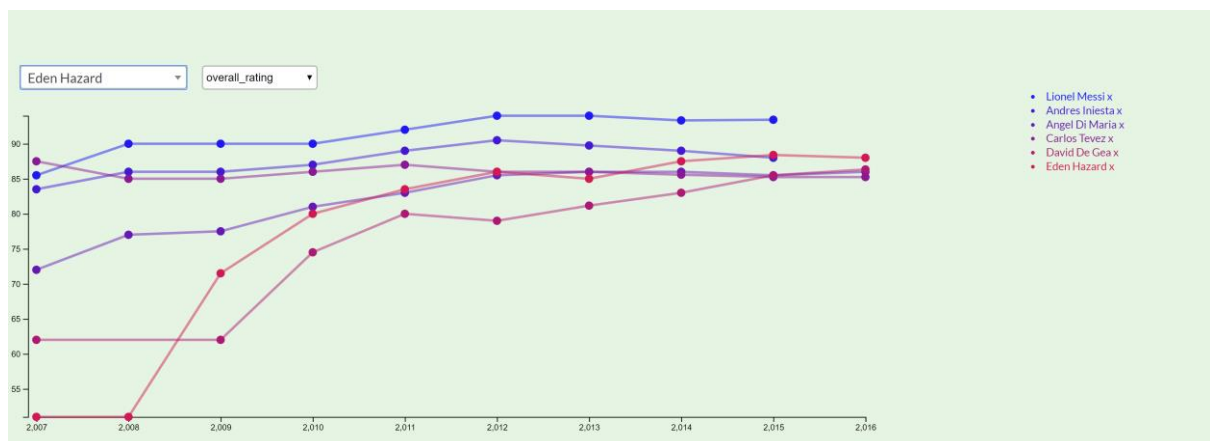


The above picture shows the trend of two players over the years. Along with this we will have a search bar on right side to include more players.

Initial interaction planned was to click on the year or place circles to click on the year to show details of the set of players selected in a bar chart that aligns with the y axis to highlight comparison of the players selected.

Addition of circles can lead to a bit of cluttering of the visualization. Hence we moved on to use a brush selection over the linecharts based on which we add new charts below the linecharts to highlight the comparison of players in those years.

We have improvised on the visualization to add color coding between the players being added and the corresponding line charts



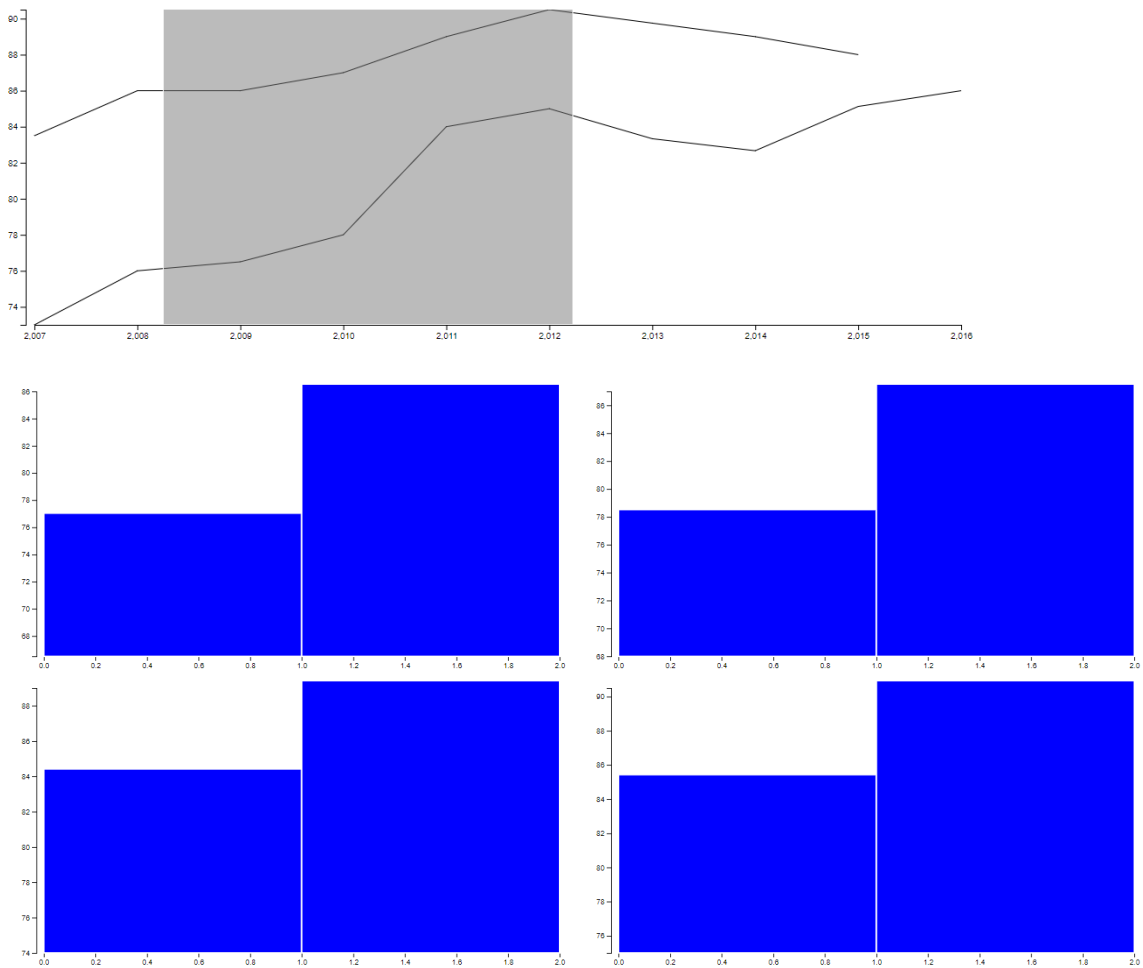
View 4:

Based on the selection from the brush, We add respective barcharts for year the number of years.

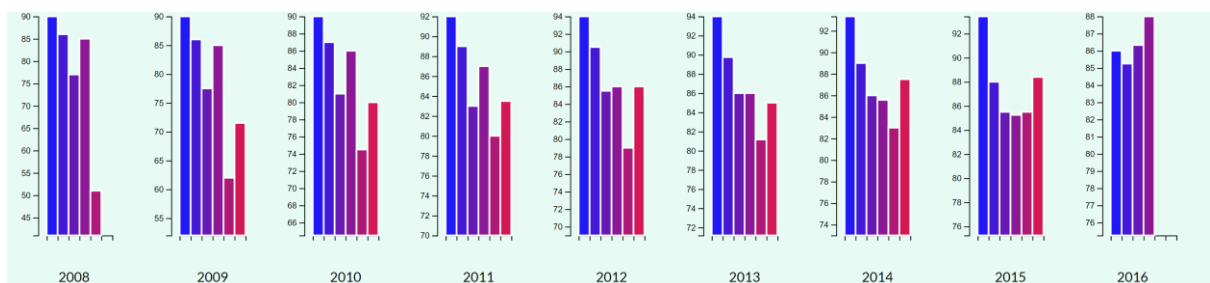
The bars would correspond to the number of players selected for the line charts' coparison.

The number of barcharts is dependent on the number of years brushed.

The following pictures show the implementation when the brush includes four years and two players.



We have improvised on the planned visualization to make the bar charts reflect the color from the line chart for the player and change the width of the rectangles depending on the number of players selected for comparison.



View 5:

We plan to show the heatmap of the clubs in a league in terms of the points scored over a set of seasons. The data spans over six seasons and there are a number of leagues at hand with good number of clubs in each of them.

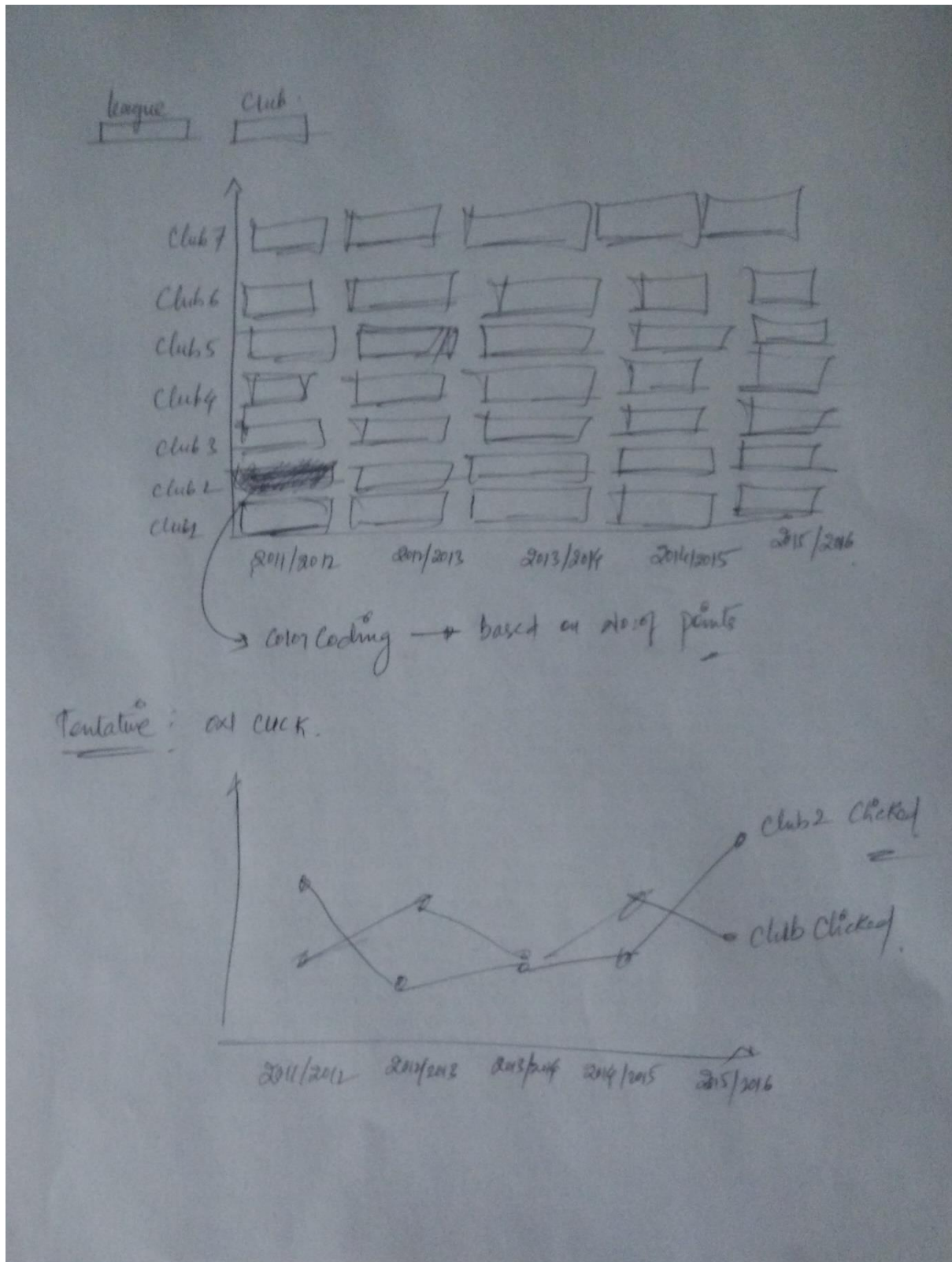
The heatmap includes color coding for the number of points that the club scored in that season. This provides a good overview of the clubs performances and comparisons in a league at the same time.

View 6(tentative):

On click of the rectangle corresponding to a club over any of the seasons, we display a line chart of the clubs performance over the seasons. This is an enhancement over the color coding in the heatmap and provides a clear picture of the progress over years.

The user can click on the multiple rectangles and the linecharts corresponding to those clubs would be added to the linechart. The user can click again on the rectangle to remove that line from the line

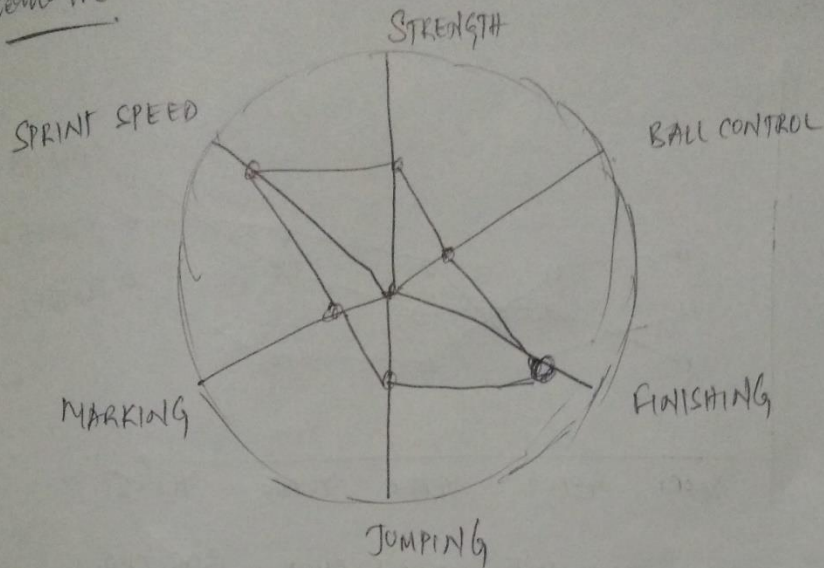
chart.



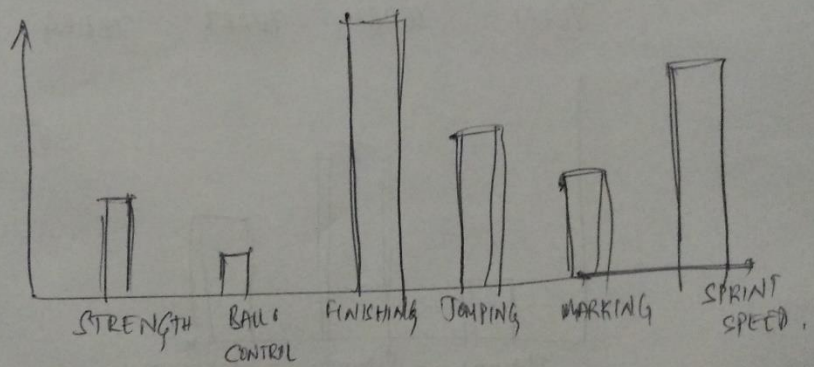
Initial Design pics to start with :

Individual Player Attributes:

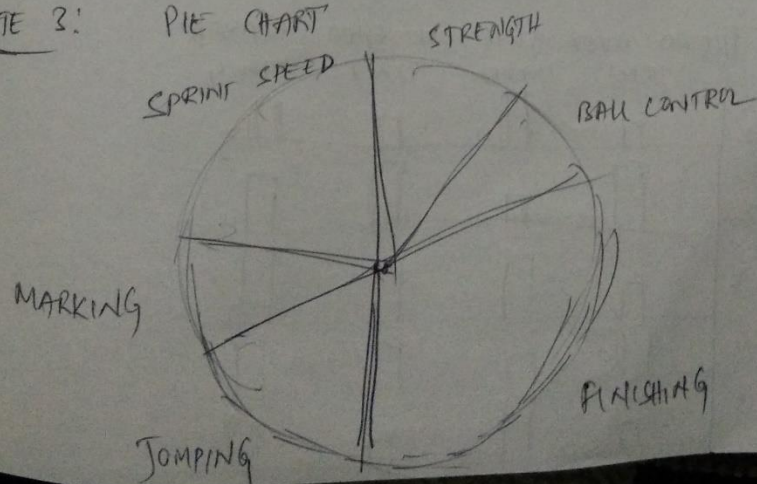
Alternative 1: RADAR CHART



ALTERNATE 2: BAR CHART



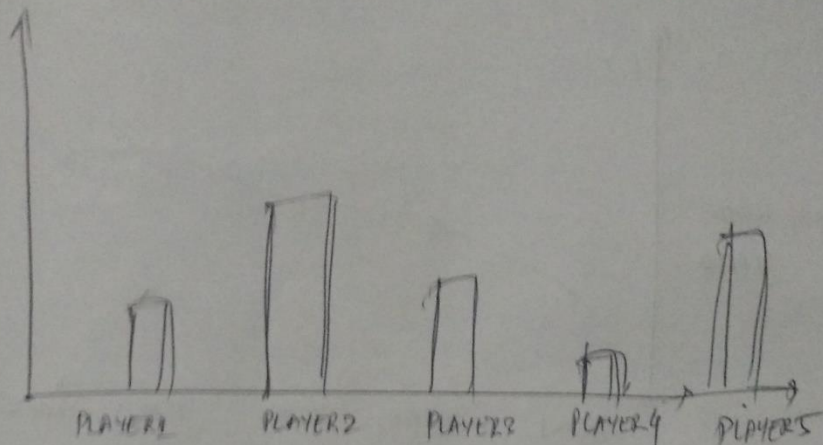
ALTERNATE 3: PIE CHART



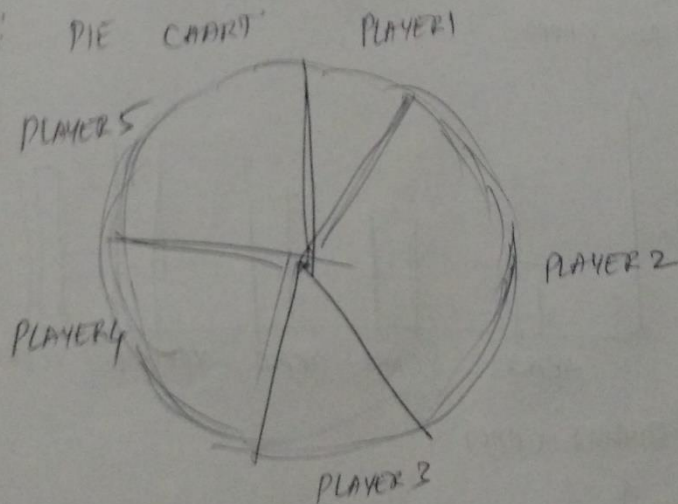
COMPARISON OF PLAYERS PER YEAR (INTERACTIVE)

ON SELECTION OF YEAR.

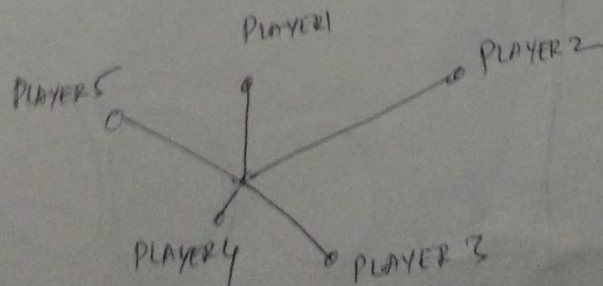
ALTERNATE 1: BAR CHART



ALTERNATE 2: PIE CHART

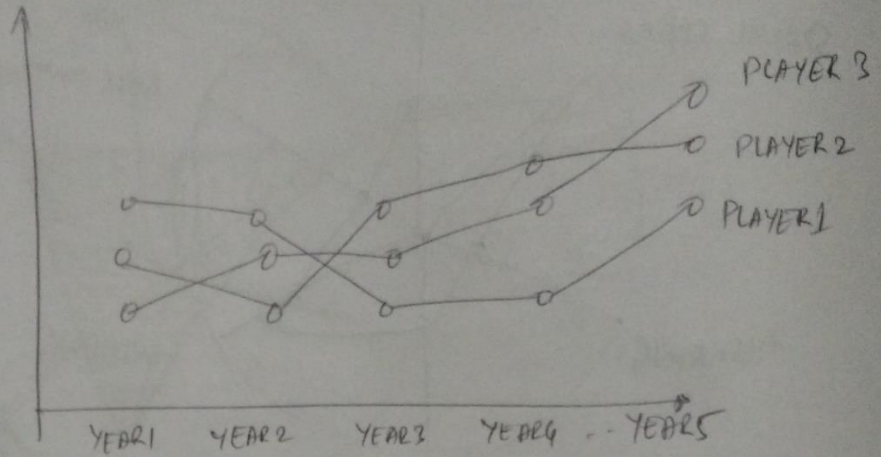


ALTERNATE 3: RADIAL LINES

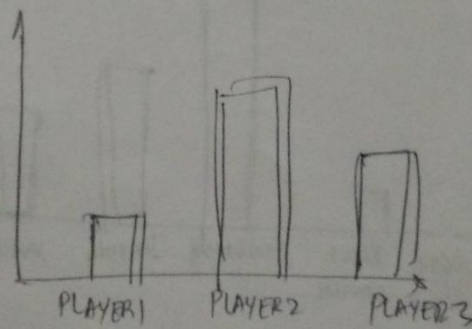
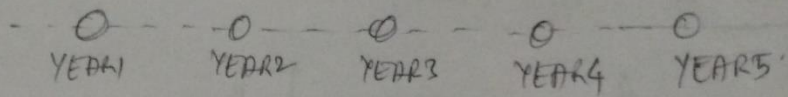


EVOLUTION OF BEST PLAYERS!

ALTERNATIVE 1: LINE CHART

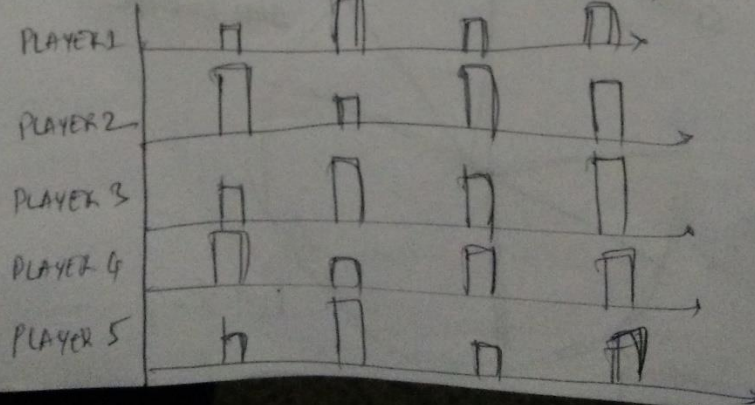


ALTERNATIVE 2: YEAR TIMELINE + ON CLICK BAR CHART



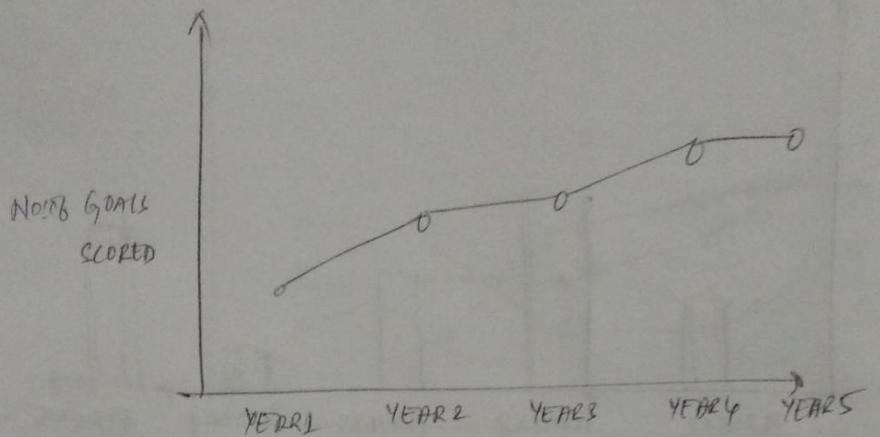
ALTERNATIVE 3:

TREND OVER TIME FOR EACH PLAYER!

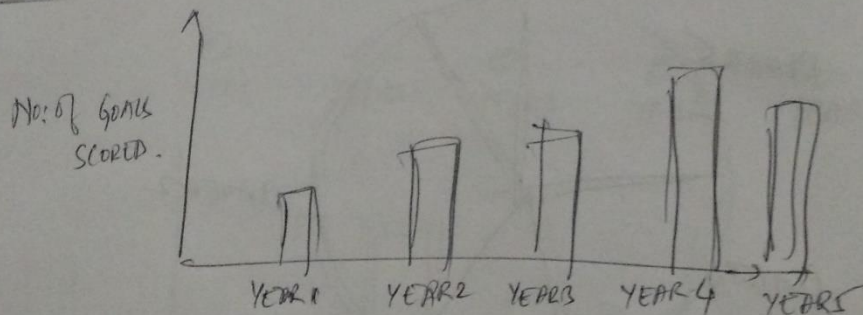


CLUBS' PERFORMANCE OVER YEARS PER LEAGUE!

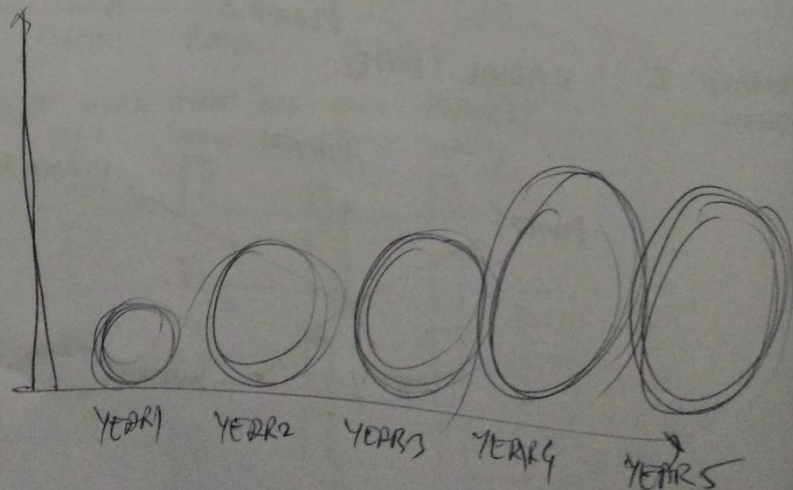
ALTERNATE 1: LINE CHART!



ALTERNATE 2: BAR CHART

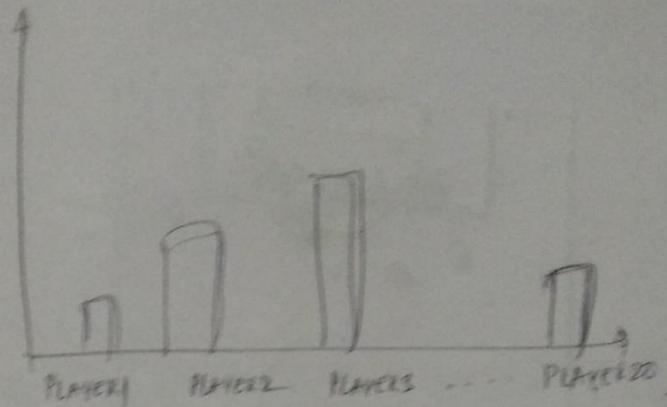


ALTERNATE 3: BUBBLE CHART

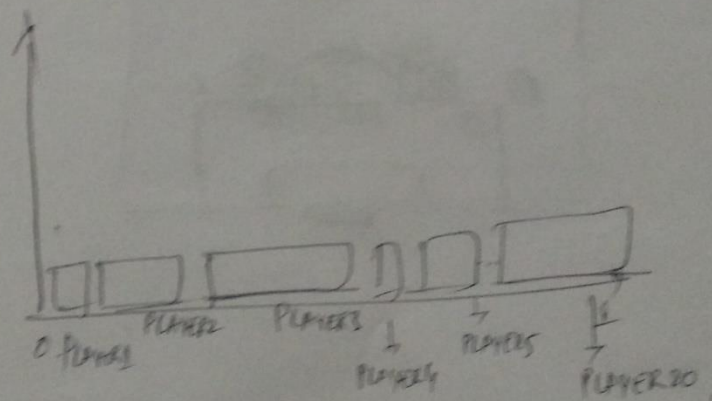


SPARE OF INFLUENCE BY PLAYER ON GETTING

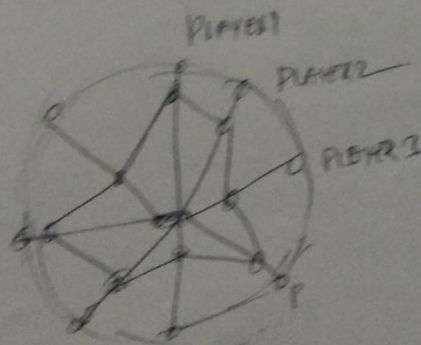
ALTERNATE 1: BAR CHART



ALTERNATE 2: STACKED BAR CHART



ALTERNATE 3: RADAR CHART



MATCH STATE : OPTIONAL :

INCIDENT DATA: EX: SHOTS TAKEN.

