Final Project: Data Visiualization

Jian Wang

Part I: Introduction

I am a follower of the European Champions League and English Premier League, so it is in my interest to see how the modern soccer games have evolved over the years. And I am a regular player of FIFA soccer games, so it is also exciting to understand how the player's attributes changed over the years. Therefore it spurs my interests to find a related dataset on soccer games in Europe. And Kaggle.com provides a great dataset that combines actual match information with synthetical scores for team and players in the game of FIFA.

It is a dataset about European Soccer matches from 2008 to 2016, in the form of database tables. There are 7 tables, which are Player, Player Attributes, Team, Team Attributes, Matches, League, Country. And this dataset includes players' in-game position defined by coordinates, so it should be helpful to make a map graph.

Precisely, I want to investigate the trend of style changes in the modern game of soccer as well as the difference among teams from the different country/league. Here is a list of interesting questions to answer:
1. I want to investigate the team's FIFA attributes on their actual match performance.
   a. The idea is to investigate which build-up play can have the best offense or defense in terms of the metric defined in FIFA.
   b. This could be done by drawing a map graph for each year and see the difference, ie heatmap.
2. I want to know the change of players' attributes over the year.
   a. To see if the player getting faster, stronger over the years.[✓]
   b. To see which type of player is more popular, technical or physical.
   c. This could be shown by a bar graph, histogram or boxplot.
3. I want to see the difference between leagues in both fronts of offense and defense.
   a. To see the goal difference between the home team and away team.[✓]
   b. To see the playing style in different leagues, like the number of foul committed or the degree of possession.[Couldn't find possession as a number.]
   c. This could be shown by bubble-map and choropleth graph.[✓]
4. I want to know the frequency of game played by teams in different leagues.
   a. This could be shown by the connection map. [It turns out the match only happen within a league.]
   b. We can show the winning and losing ratio for teams from different leagues.
5. I want to know the composition of goals within a year in terms of leagues, teams.
   a. This can be shown by treemapping.[✓]

Part II: Summary of Data.

Histogram:
- On player's ability changes over the years.
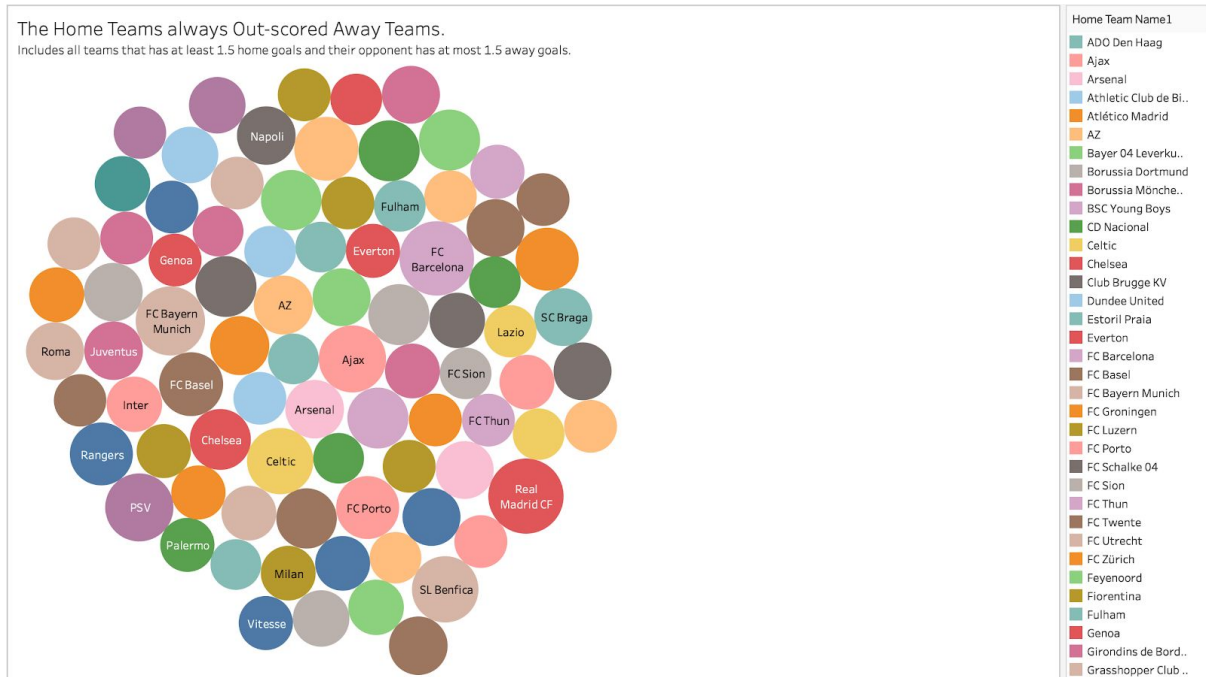

Build up speed over different Leagues.
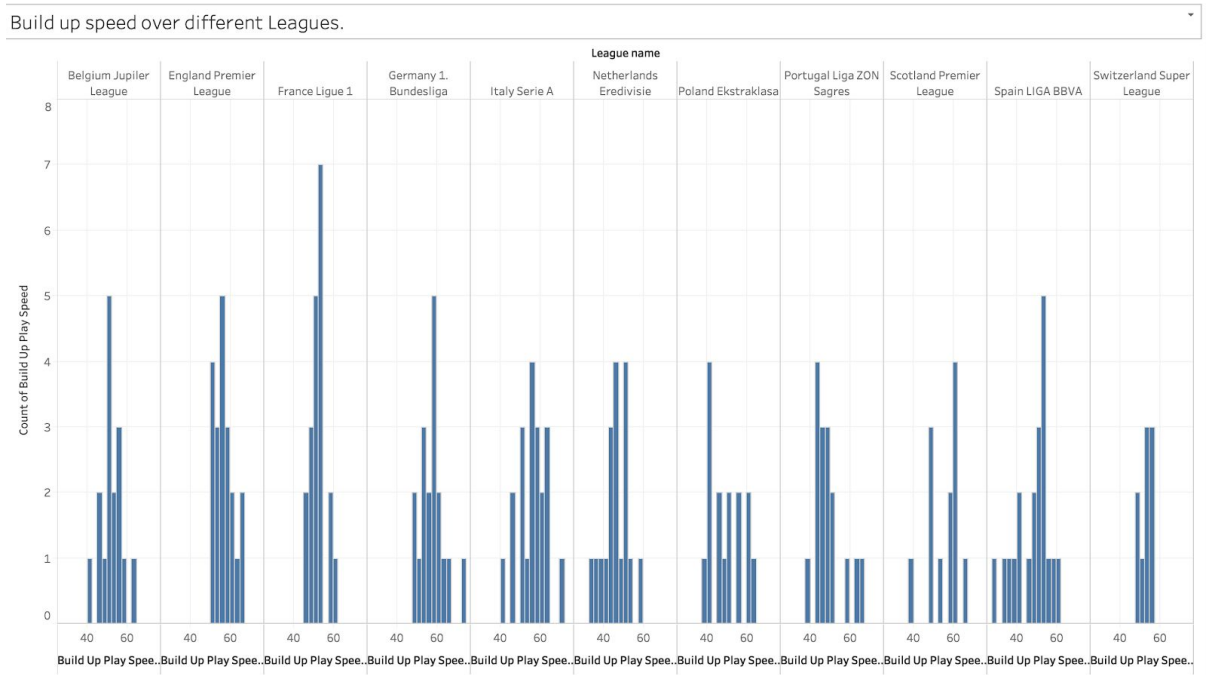
Description:

This plot show's different's league's game of speed. And we can see the distribution of teams in Spain LIGA BBVA is much slower than the counterparts in England Premier League. Looking at the lower's lend of the build-up speed, the minimum value is around 45 for teams in England but for teams in Spain, it can fall to 10.

This observation consolidates our common observation, for which Spanish teams hold the ball more while English teams play more directly. But surprisingly, we see that Italy has a very wide range of playing speed by counting the number of bins and so is the Spanish team.

Bubble map:
- On each league's away's team's performance

The Home Teams always Out-scored Away Teams.
Includes all teams that has at least 1.5 home goals and their opponent has at most 1.5 away goals.

Home Team Name1
- ADO Den Haag
- Ajax
- Arsenal
- Athletic Club de Bi..
- Atlético Madrid
- AZ
- Bayer 04 Leverku..
- Borussia Dortmund
- Borussia Mönche..
- BSC Young Boys
- CD Nacional
- Celtic
- Chelsea
- Club Brugge KV
- Dundee United
- Estoril Praia
- Everton
- FC Barcelona
- FC Basel
- FC Bayern Munich
- FC Groningen
- FC Luzern
- FC Porto
- FC Schalke 04
- FC Sion
- FC Thun
- FC Twente
- FC Utrecht
- FC Zürich
- Feyenoord
- Fiorentina
- Fulham
- Genoa
- Girondins de Bord..
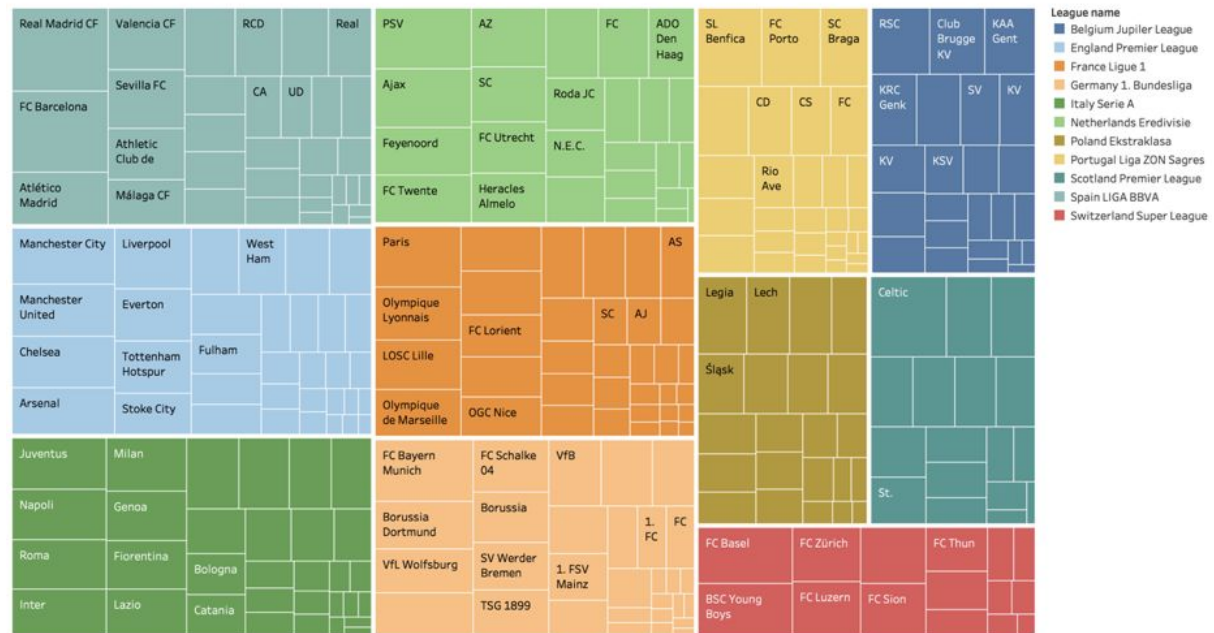- Grasshopper Club ..

This plot uses data from all the match and filtered in such a way that the home team's average scores are higher than their opponents. The color of the map represents different teams while the size of the bubble represents the average goals scored by home teams. It is interesting that FC Barcelona, Real Madrid CF, and FC Bayern Munich are so outstanding in the graph.

Treemaps:
- On goals scored by league and teams

Total Team Goals per League

Home Team Name1. Color shows details about League name. Size shows sum of Home Team Goal. The marks are labeled by Home Team Name1.
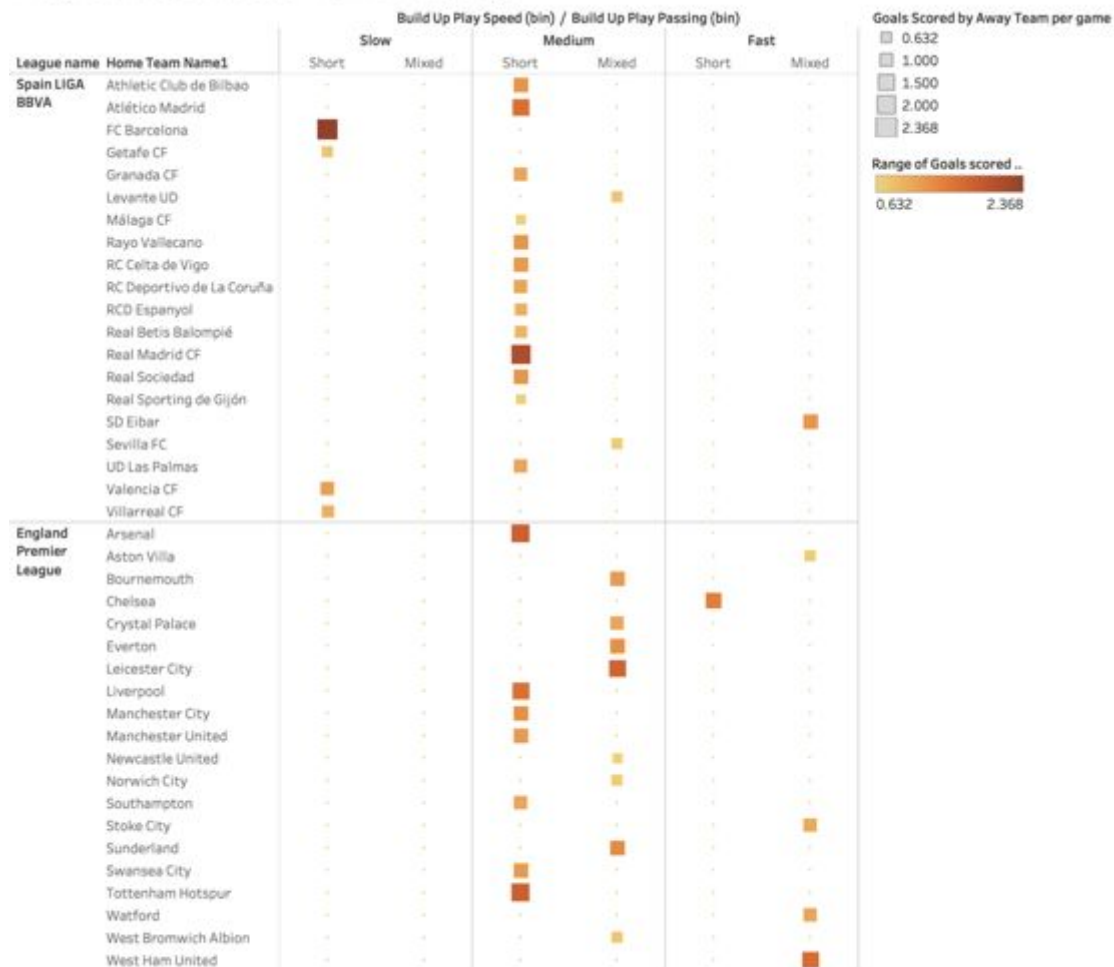
This graph gives us the top scorer team in each league that scores the most goals over the year from 2008 to 2016. The color indicates the leagues and the size of the block indicates the total number of goals.

This information summarizes the history of each league and witnesses the teams' attacking abilities.

Heatmap:
- On away teams' goals scored and its relationship with the team's build-up style.

Away Teams Goal Scored on Team Build Up style

League name	Home Team Name1	Slow – Short	Slow – Mixed	Medium – Short	Medium – Mixed	Fast – Short	Fast – Mixed
Spain LIGA BBVA	Athletic Club de Bilbao	·	·	■	·	·	·
	Atlético Madrid	·	·	■	·	·	·
	FC Barcelona	■	·	·	·	·	·
	Getafe CF	■	·	·	·	·	·
	Granada CF	·	·	■	·	·	·
	Levante UD	·	·	·	■	·	·
	Málaga CF	·	·	■	·	·	·
	Rayo Vallecano	·	·	■	·	·	·
	RC Celta de Vigo	·	·	■	·	·	·
	RC Deportivo de La Coruña	·	·	■	·	·	·
	RCD Espanyol	·	·	■	·	·	·
	Real Betis Balompié	·	·	■	·	·	·
	Real Madrid CF	·	·	■	·	·	·
	Real Sociedad	·	·	■	·	·	·
	Real Sporting de Gijón	·	·	■	·	·	·
	SD Eibar	·	·	·	·	·	■
	Sevilla FC	·	·	·	■	·	·
	UD Las Palmas	·	·	■	·	·	·
	Valencia CF	■	·	·	·	·	·
	Villarreal CF	■	·	·	·	·	·
England Premier League	Arsenal	·	·	■	·	·	·
	Aston Villa	·	·	·	·	·	■
	Bournemouth	·	·	·	■	·	·
	Chelsea	·	·	·	·	■	·
	Crystal Palace	·	·	·	■	·	·
	Everton	·	·	·	■	·	·
	Leicester City	·	·	·	■	·	·
	Liverpool	·	·	■	·	·	·
	Manchester City	·	·	■	·	·	·
	Manchester United	·	·	■	·	·	·
	Newcastle United	·	·	·	■	·	·
	Norwich City	·	·	·	■	·	·
	Southampton	·	·	■	·	·	·
	Stoke City	·	·	·	·	·	■
	Sunderland	·	·	·	■	·	·
	Swansea City	·	·	■	·	·	·
	Tottenham Hotspur	·	·	■	·	·	·
	Watford	·	·	·	·	·	■
	West Bromwich Albion	·	·	·	■	·	·
	West Ham United	·	·	·	·	·	■

Goals Scored by Away Team per game
- 0.632
- 1.000
- 1.500
- 2.000
- 2.368

Range of Goals scored ...
0.632 — 2.368

Sum of Away Team Goal (color) and sum of Away Team Goal (size) broken down by Build Up Play Speed (bin) and Build Up Play Passing (bin) vs. League name and Home Team Name1. The view is filtered on League name, which keeps England Premier League and Spain LIGA BBVA.

This heat map shows the individual teams' build-up styles in terms of directness. We pick teams in Span and England, which represents the two style of modern plays, the former represents the possession-focused while the latter likes to play more direct.

We can see from the plot that England teams like to pass longer and riskier. But for Spanish teams, they choose to pass safer and play more slowly.

Bar plot:

Player's technical abilities change over time.
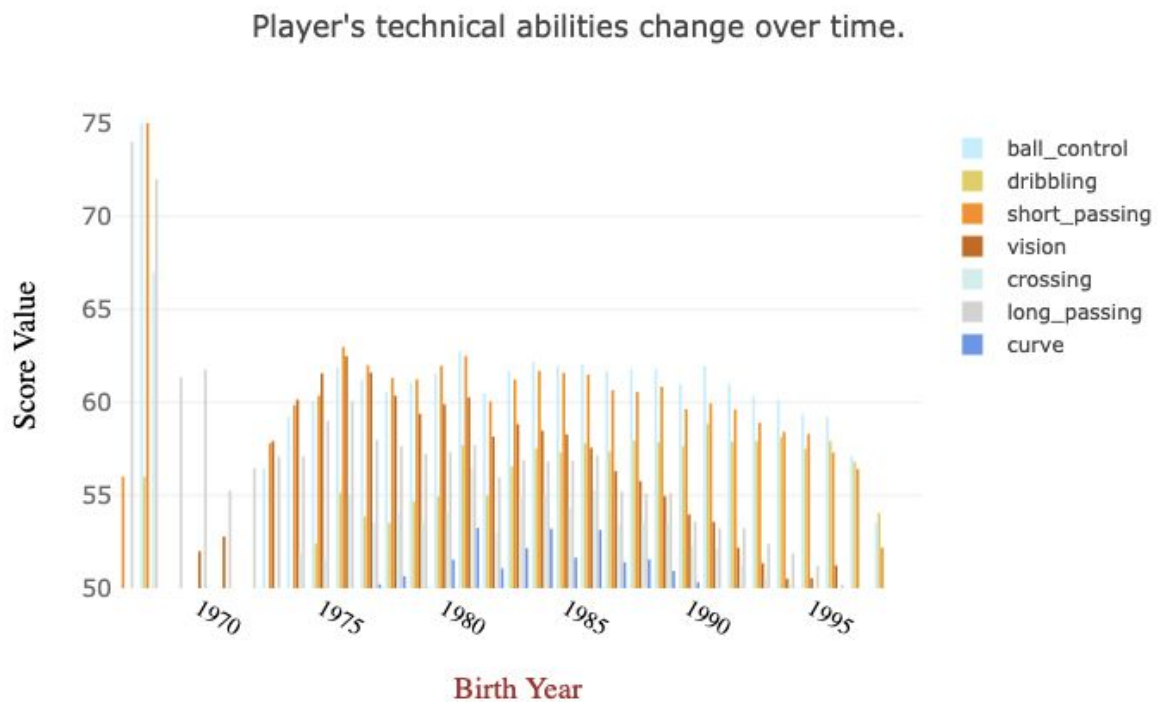
This bar plot shows the overall player's technical abilities changes over the years. We see that for vision, the older players are better at find opportunity in the pitch while the young players are better at dribbling. This is an interactive plot which you can find in the attached files.

Scatter Plot:

Relationship between passing style and crossing frequency.
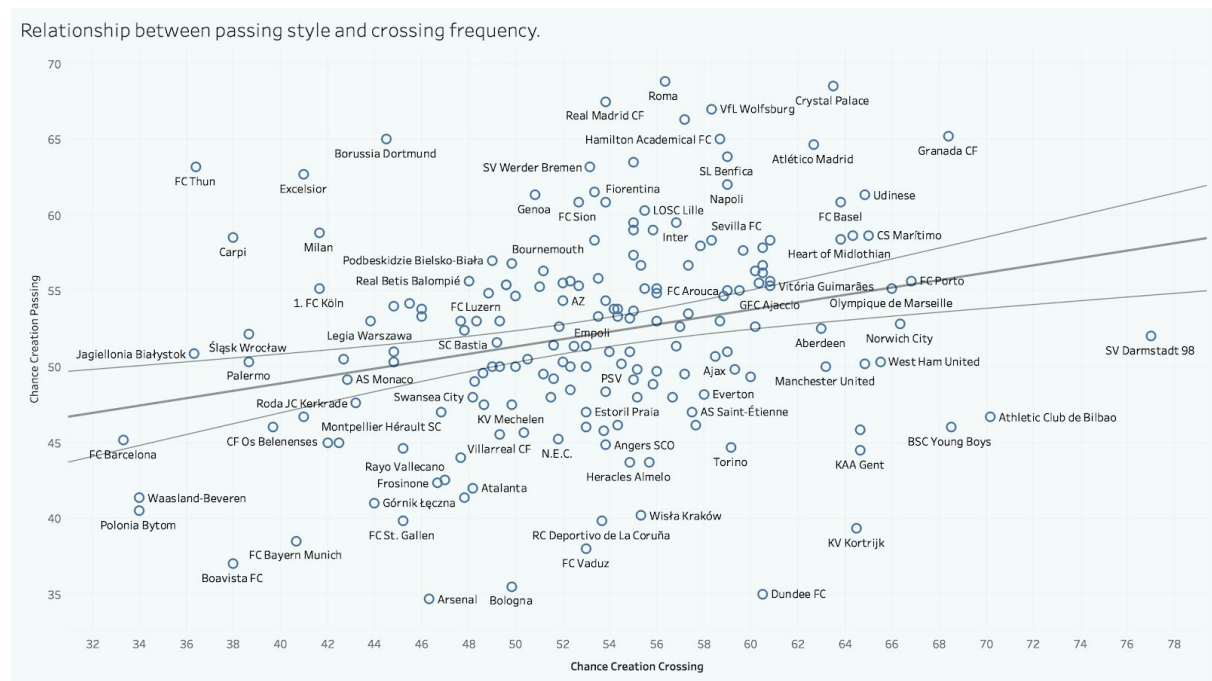
This plot shows the relationship between all teams' passing style and crossing frequency. It turns out that team will more cross tends to have more long-passing and through passes as well.

From this plot, we see that Barcelona has an extreme style of playing, which is almost no crossing. While Atletico Madrid, which in the same league as Barcelona, played a more direct game.

Box-Plot:

Frequency of shooting over different league.
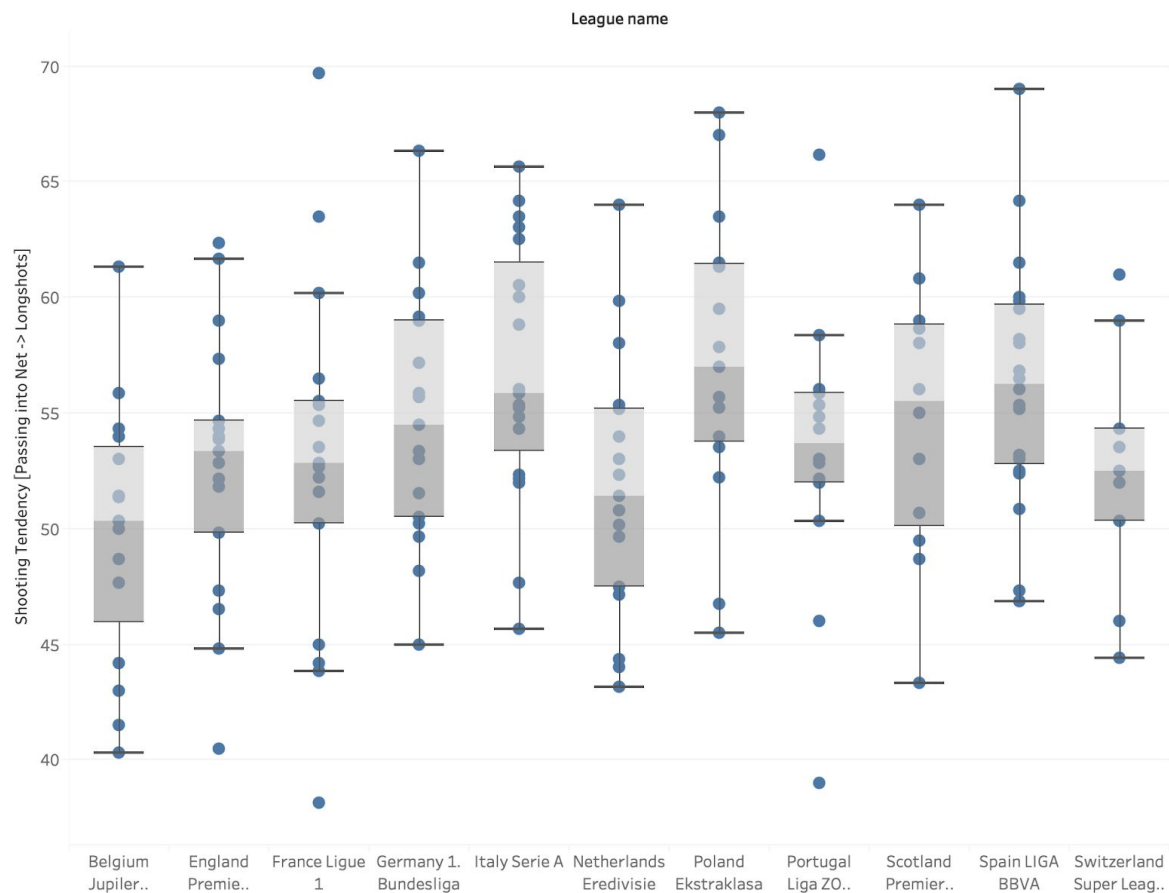
This box-plot shows the tendency of shooting for each league. It is a nice summary on how teams tend to score goals. Surprisingly, England is not as likely to make long shots as the major way to score and Netherland Eredivisie is the least likely to make long shots.
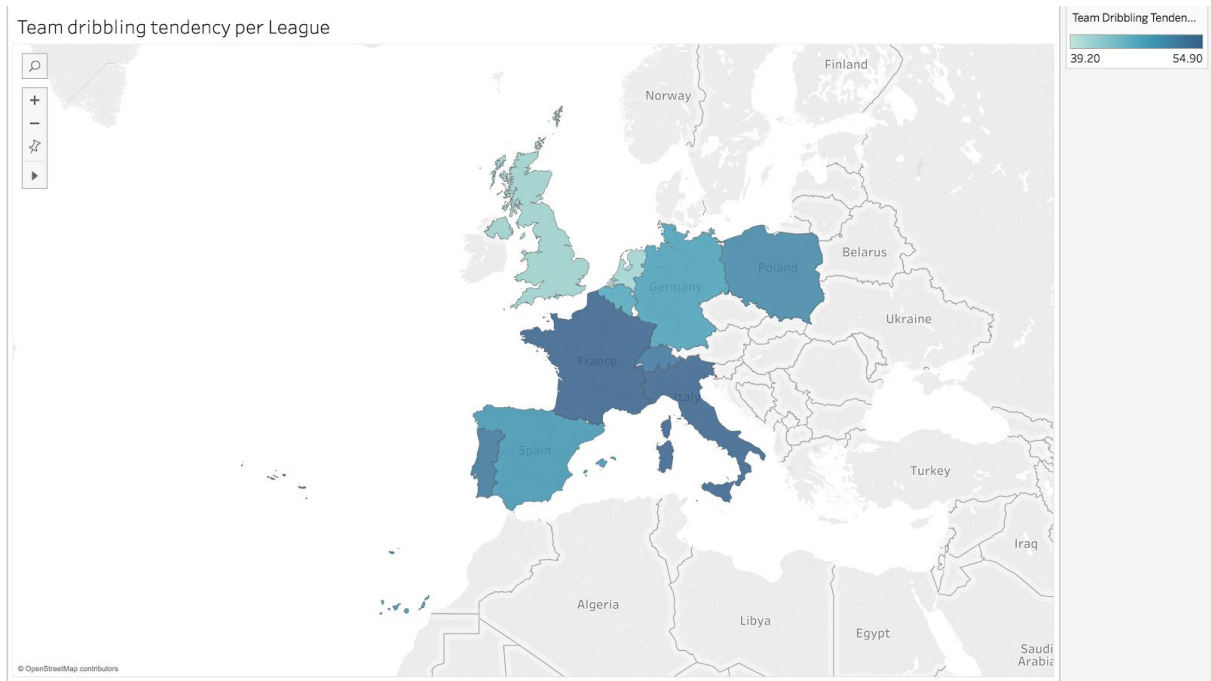
Stacked Bar plot:

Top 4 Attacking Teams On Each Season in England Premier League
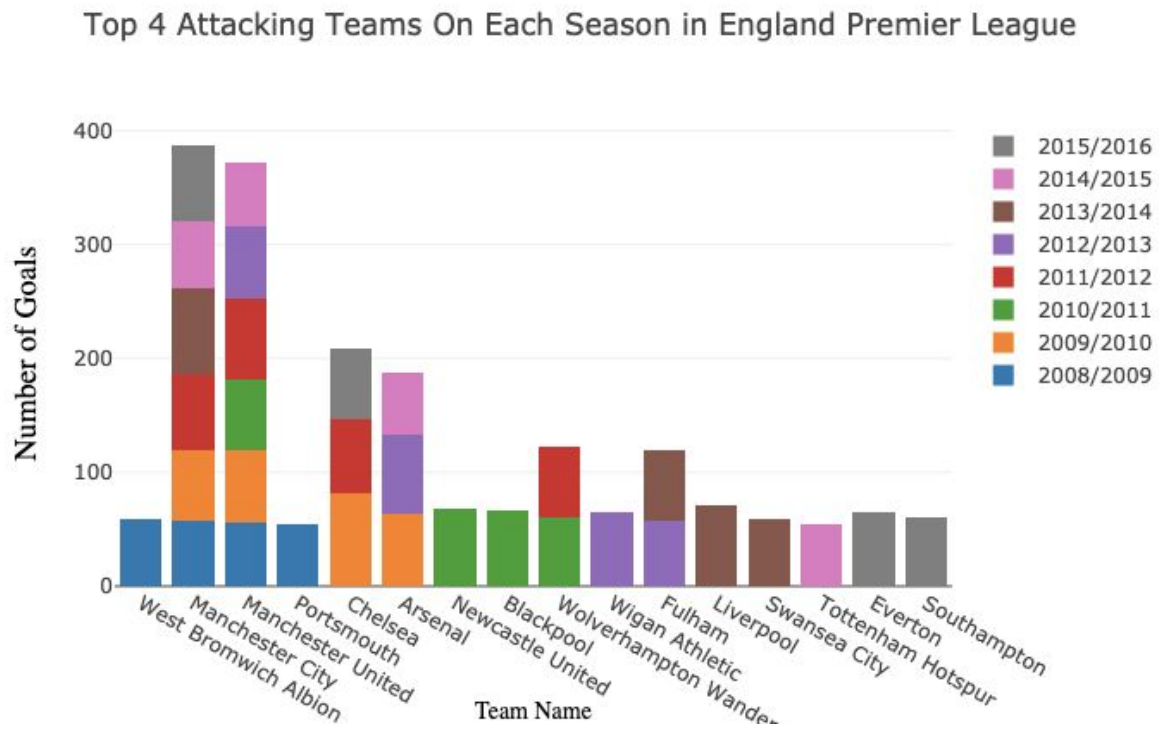
This stacked barplot shows all teams that have been the top four scored team in English Premier League. We can see Manchester City and Manchester United are the best attacking teams.

Choropleth graph:

Team dribbling tendency per League

This choropleth graph shows the tendency to dribble for the game across the continent. It turns out that teams in France and Italy tend to use dribbling to advance play.

Connection map:

- This connection map shows the most entertaining games between Manchester United and all other Premier League's team. We define the entertainment by the total goals scored over the year by the two teams. Surprisingly, it turns out that between Man Utd with Stoke City is one of the most exciting games to watch.

The Connection Map from Manchest united on To Other Team.
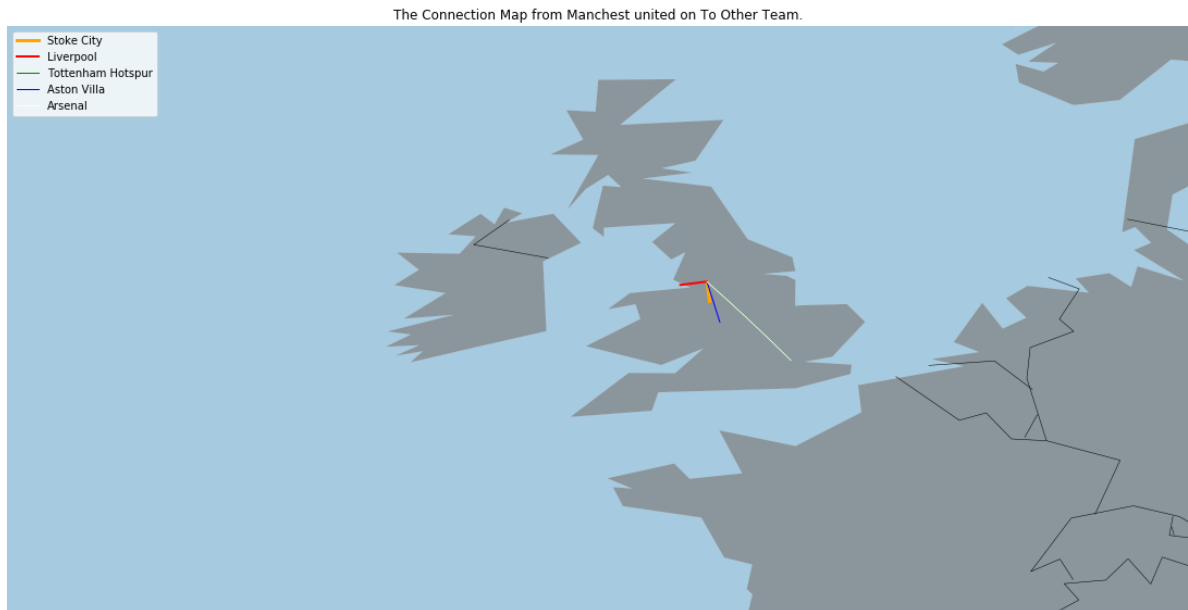
most_exct_games

executed in 8ms, finished 13:22:10 2019-05-15

	away_team_name1	home_team_goal	away_team_goal	Total_goals_btw_two_teams
0	Manchester City	14	17	31
1	Stoke City	25	6	31
2	Liverpool	16	13	29
3	Tottenham Hotspur	20	8	28
4	Aston Villa	21	6	27
5	Arsenal	18	7	25
6	Everton	14	6	20
7	Newcastle United	12	7	19
8	Fulham	14	4	18
9	Hull City	14	4	18

III A stroy to tell.

When I was investigating the trends of a player's attributes changes, I find one interesting phenomenon. It can best be summarized as the following plot:

Player's physical abilities change over time.
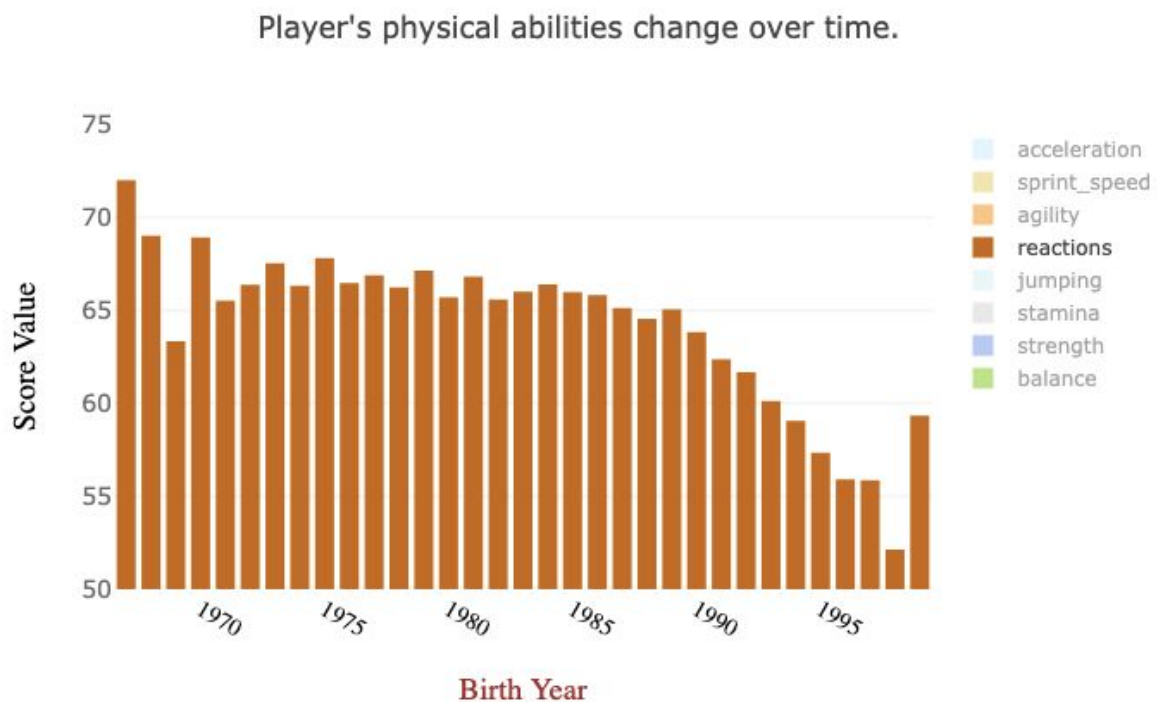
From this explanation of the mechanism behind this metric, we see that reaction is defined as the ability for a player to respond to his situation. And naturally, we think as a physical attribute, the young players should have a more active sense of their surrounding. But it turns out that such a trait is better equipped with older players. This is in some way contradicting what we have thought of, but after reviewing some intelligent player's report. We see that reaction is more about the player's analysis of the situation on the pitch and how he should participate and make a difference.

During the game, a lot of random things could happen, and for a young player, maybe due to inexperience, he may not see the opportunity but for a seasoned player, he knows under which situation the chance could appear on the pitch. In this sense, we can understand that the older athelete having higher average reaction scores do make sense.

For the team development, it is better to train the young player with older players and for any team, going for young promising players is not always a right move. You have to have some veteran players to educate new players and be aware of the change in one game.

Citations:

On treemapping: https://datavizcatalogue.com/methods/treemap.html

On colors: https://www.webucator.com/blog/2015/03/python-color-constants-module/

On python formating: https://pyformat.info/

On SQLite datetime manipulation: https://www.sqlite.org/lang_datefunc.html

On plotly axis: https://plot.ly/python/axes/