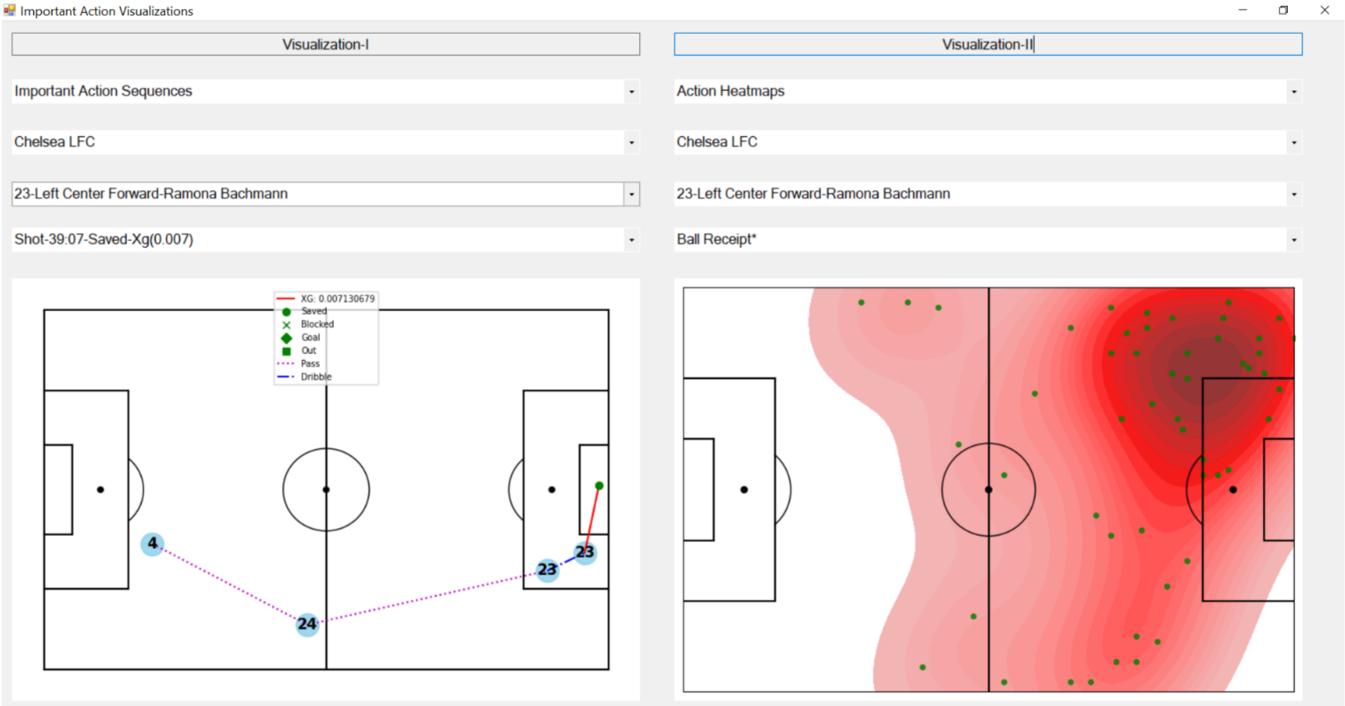
Sabancı Üniversitesi

Football Data Visualization CS 543 Term Project

Furkan Reha Tutaş
Master Student at Sabanci University
Pendik/Istanbul
Turkey
furkanreha@sabanciuniv.edu



Important Action Visualizations

Visualization-I	Visualization-II
Important Action Sequences	Action Heatmaps
Chelsea LFC	Chelsea LFC
23-Left Center Forward-Ramona Bachmann	23-Left Center Forward-Ramona Bachmann
Shot-39:07-Saved-Xg(0.007)	Ball Receipt*

Fig. 1. Using visualization interface: (left) An important action sequence filtered by player, (right) Ball receipt action heatmap for a filtered player.

1.ABSTRACT

This article provides a visualization interface for football data from Stats Bomb [1] using the idea in the selected paper Soccer Stories [2] which is player positions and phases of player actions are the most efficient way of analyzing football matches. The interface aims to give detailed insights about a football match by simultaneously providing average in-out positions of teams, heatmaps for different types of actions (ball receptions, passes, shots, duels, etc.) of each player, sequence of important actions (sequence of actions yield to shot) happened during a game, expected goals versus time for both oppositions, expected goals plot on the pitch for both oppositions and important actions involvement heatmap for a selected team.

2.INTRODUCTION

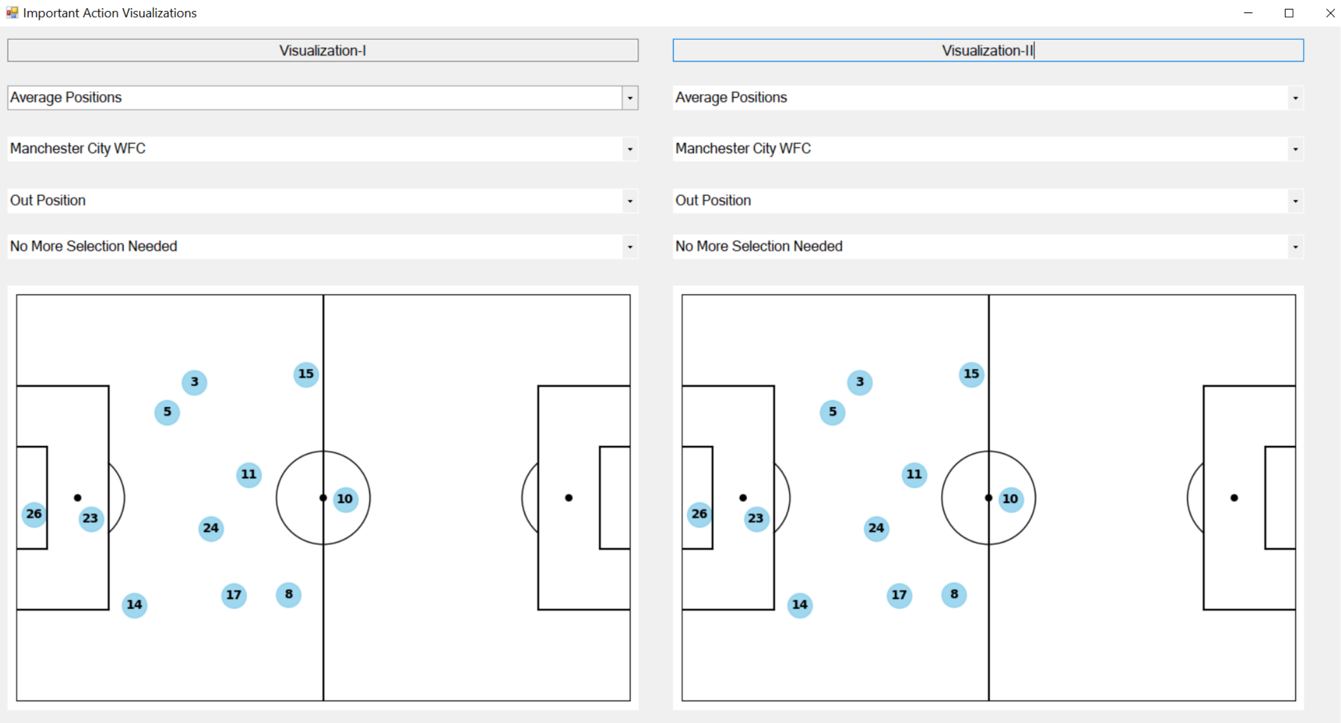
As stated in Soccer Stories [2], some companies provide data almost for all national leagues that consist of player positions, expected goals, action types, time of actions, etc.

This paper uses data from Stats Bomb [1] for Women Premier League provided as free. The developer of this visualization project is grateful to Stats Bomb football analytics experts [3].

Soccer Stories [2] indicates that a soccer game consists of “phases”. In this paper, phases are called important action sequences to be clearer in the definition. A phase is a sequence of actions by players of the same team which yields a shot at the of that sequence. For example, in [Figure 1-left], the visualization shows an important action sequence which consists of 2 passes, 1 dribble, and at the end 1 shot which is saved by the opposition goalkeeper.

In addition to important action sequences, heatmaps are used to see the distribution of an action type positions by a player on the pitch. For example, in [Figure 1-right], the visualization shows a heatmap for ball receipt action (positions where the filtered player receives the ball) for a player.

STATSBOMB
INSIDE EVERY PASS, EVERY SHOT, EVERY MOVE



Important Action Visualizations

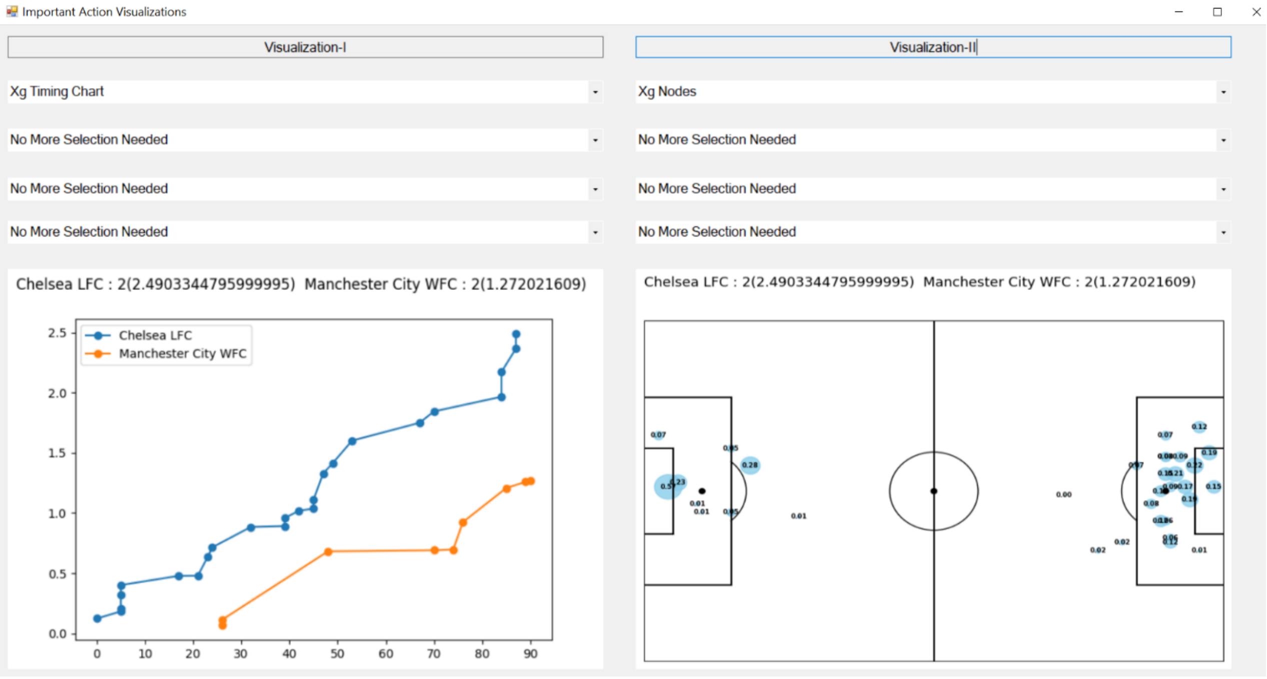
Visualization-I	Visualization-II
Average Positions	Average Positions
Manchester City WFC	Manchester City WFC
Out Position	Out Position
No More Selection Needed	No More Selection Needed

Fig. 2. Using visualization interface: (left) Average out-positions for a filtered team, (right) Average in-positions for a filtered team

Thirdly, scatter plots are used to see the average in/out positions (In-position means when the team has the possession/ball, out-position means when the opposition has the possession/ball) for a team. For example, [Figure 2-left] shows the average out-position scatter plot for a filtered team, on the one hand [Figure 2-right] shows the average in-position scatter plot for a filtered team.

In addition to these visualizations, connected scatter plot for expected goals versus game time is used. For example, in [Figure 3-left], in terms of expected goals, Chelsea overperformed Man City.

Another visualization is expected goal nodes that show the expected goal for each action weighted by XG for each team. For instance, in [Figure 3-left], although Chelsea is better in terms of



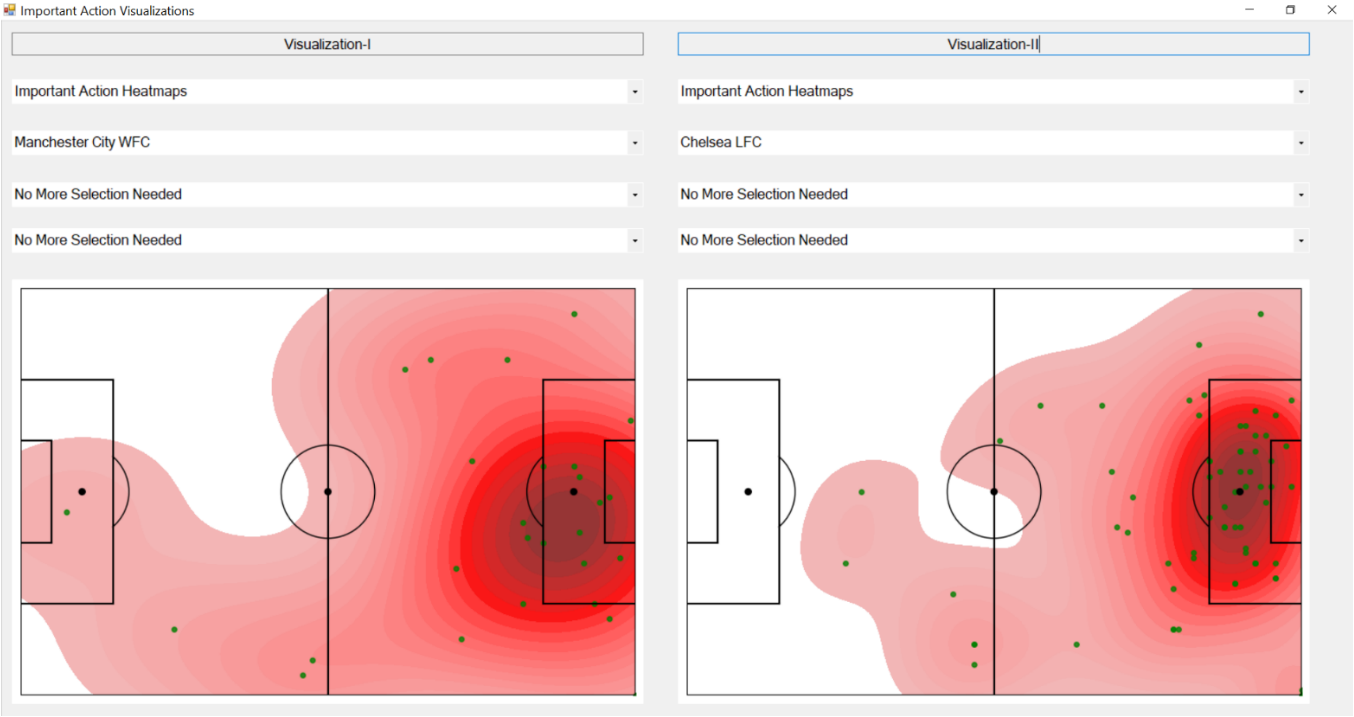
Important Action Visualizations

Visualization-I	Visualization-II
Xg Timing Chart	Xg Nodes
No More Selection Needed	No More Selection Needed
No More Selection Needed	No More Selection Needed
No More Selection Needed	No More Selection Needed

Chelsea LFC : 2(2.4903344795999995) Manchester City WFC : 2(1.272021609)

Chelsea LFC : 2(2.4903344795999995) Manchester City WFC : 2(1.272021609)

Fig. 3. Using visualization interface: (left) XG versus game time, (right) Weighted scatter plots w.r.t expected goals



Important Action Visualizations

Visualization-I	Visualization-II
Important Action Heatmaps	Important Action Heatmaps
Manchester City WFC	Chelsea LFC
No More Selection Needed	No More Selection Needed
No More Selection Needed	No More Selection Needed

Fig. 4. Using visualization interface: (right-left) Important action heatmaps for both oppositions.

accumulated expected goals, Man City creates the best chance during the game with the expected goal of 0.57.

The final visualization is an important action involvements heatmap for each action. This heatmap only considers the important action sequences and includes only shot-assist-final pass (assist of assist) actions in it. In [Figure 4], Man City's important actions are more on the right-hand side of the pitch while Chelsea's important actions are more equally distributed on the pitch.

3.DATA

In the below table, the necessary attributes from Stats Bomb data [1] for the implementation are explained.

Attribute	Functionality
Id	A unique id for each action.
Player	Name and Surname of the player.
Location	[X, Y] coordinates on the pitch where the action happened.
Team	Team name of the player who did the action.
Type	Action type (Ball Receipt, Shot, Duel, Block, Goalkeeper Save, etc.)
Possession Team	Name of the team that has the possession during the action.
Outcome	Outcome of the action. Successful/Unsuccessful for pass action, Goal/Blocked/Saved/Out for shot, etc.
End Location	This is only available for pass/shot actions. In this project only shot ending location is used.
XG	Expected goal for a goal action.
Second	Exact second when the action happened.
Minute	Exact minute when the action happened.

In the below table, newly processed attributes from Stats Bomb data [1] for the implementation are explained.

Attribute	How It Processed
Time	Define as minute:second
Previous Actions	Defined only for shot actions. It contains all actions until the shoot action by the same team [Pseudo Code-1].

Extract Previous Actions for Shots [Pseudo Code-1]:

```
For each shot in shot actions:
    For each action previously happened (in reverse order):
        If action.possession_team != shot.team:
            break
        else
            shot.previous_actions.add(action)
```

Stat Bombs [1] provides the data in JSON format. In the project, the Pandas library in python is used to read and pre-process the data.

4.VISUALIZATION AND SERVER

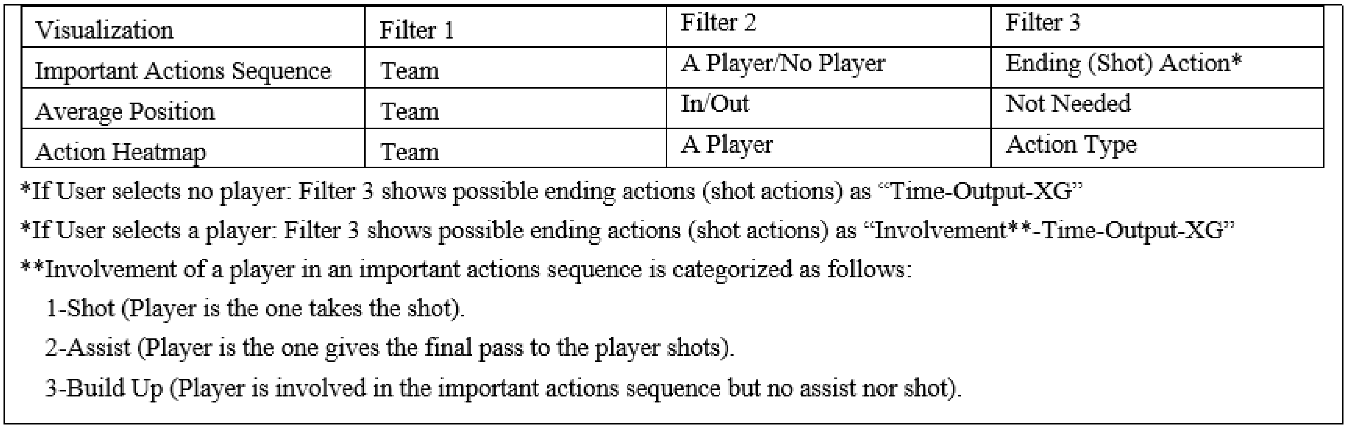
For visualization Matplotlib.pyplot library and seaborn library in python are used. Pyplot is used to visualize the pitch, whereas

Seaborn is used to visualizing scatter plots and heatmaps (Average Positions, Important Action Sequences, and Action Heatmaps, etc.).

4.1.Visualization of Important Actions Sequences

An important action sequence is a sequence of actions by the same team that results in a shot action. Please refer to [Pseudo Code-1] for how to extract the sequence of actions until the shot action.

As it can be seen in [Figure 1-left], each action represented as a point labelled with the player shirt number who did the action and connected to the next action depending on the action type (Purple dot line for pass action, blue dot line for dribble action, the red regular line for shot action) on the pitch. In addition to this, the outcome of the shot action represented with a shape (Circle for saved shot, square for shot goes out of the pitch, diamond for a goal and cross for blocked shot by an opposition team player) and corresponding expected goal (XG) is labeled in the legend.



Visualization	Filter 1	Filter 2	Filter 3
Important Actions Sequence	Team	A Player/No Player	Ending (Shot) Action*
Average Position	Team	In/Out	Not Needed
Action Heatmap	Team	A Player	Action Type

*If User selects no player: Filter 3 shows possible ending actions (shot actions) as "Time-Output-XG"

*If User selects a player: Filter 3 shows possible ending actions (shot actions) as "Involvement**-Time-Output-XG"

**Involvement of a player in an important actions sequence is categorized as follows:

1-Shot (Player is the one takes the shot).

2-Assist (Player is the one gives the final pass to the player shots).

3-Build Up (Player is involved in the important actions sequence but no assist nor shot).

Fig. 5. Shows types of filters for each visualization.

Labelled points are visualized by using seaborn.regplot and the rest (shapes (shot outcome), lines between points (actions), the pitch, and the legend) are visualized by using Matplotlib.pyplot.

4.2.Visualization of Average In/Out Positions

Average in/out positions shows the average formation by a team when they have/don't have the ball/possession. It is quite important to distinguish in-out possession formations of teams since most teams have different formations for defense (out-possession) and attack (in-possession).

As can be seen in [Figure 2], each player's average position is represented as a point labelled with the player's shirt number. The average position for a player simply calculated by taking the average location of all actions (filtered by whether team is in/out possession) by that player.

Labelled points (players) are visualized by using seaborn.regplot and the rest (the pitch) are visualized by using Matplotlib.pyplot.

4.3.Visualization of Action Heatmaps

Action heatmap shows the distribution of positions for a filtered action by a filtered player. It helps to understand the player contribution and positioning for an action type.

As it can be seen in [Figure 1-right], each action represented as a point colored as green and red colored areas represent the distribution of the action positions.

Green colored points are visualized by using seaborn.regplot, corresponding heatmap is visualized by using seaborn.kdeplot and the rest (the pitch) are visualized by using Matplotlib.pyplot.

4.4.Server and GUI

Server for GUI is developed using Flask-Restful in python.

GUI is developed using Windows form application in C#. The followings are the steps for the communication between GUI and server.

1. GUI sends get request to server GET ……./visualization_type?var="……"
   Where variable is the filters selected by the user.
2. Server saves the related visualization as a bitmap and sends location of the bitmap.
3. GUI opens the bitmap and shows to the user.

4.5.Filters for Visualizations

In [Figure 5], filter options for each visualization can be observed. Since filtering for important actions sequence visualization is the most complex and similar approach to the other two visualizations, it will be covered in this section.

1. GUI gets teams (home and away team) from the server and display teams in filter 1.

2. User selects a team from filter 1; GUI gets players of the filtered team (Player displayed as Shirt Number-Position-Name) from server. GUI adds players to filter 2. Since user may want to continue without selecting a player, GUI adds one more filtering option as “Continue without selecting a player”.
3. If user selects a player, GUI gets all important actions sequences that the filtered player involved and adds these actions sequences to filter 3.
4. If user selects no player, GUI gets all important actions sequences by the selected team and adds these actions sequences to filter 3.
5. User selects an important actions sequence; GUI gets bitmap for that important actions sequence from server and visualize it.

In order to determine players involvement in important actions sequences, the following idea is used:

```
For p in players:
  For s in important_action_sequences:
    If p = s.shot.player:
      p.important_actions_involvement.add(s.id, “Shot”)
    elif p = s.assist.player:
      p.important_actions_involvement.add(s.id,“Assist”)
    elif p is in s.other_actions.players:
      p.important_actions_involvement.add(s.id,“Build”)
```

4.6.Visualization of XG Timing Chart

XG timing chart shows the accumulated expected goal for two teams w.r.t game time.

As can be seen in [Figure 3-left], teams are represented with two different colors (red and orange), and each point on the scatter plot represents a shot action by that team. Each point shows the accumulated expected goal (y-axis) until the game minute (x-axis).

Connected multiple scatter plots are visualized by using Matplotlib.pyplot.

4.7.Visualization of XG Nodes

XG nodes show expected goals for each team’s shot actions and these nodes are weighted by expected goal values.

As can be seen in [Figure 3-right], Chelsea (Home Team)’s XG nodes are shown on the right-hand side of the pitch, whereas Man City (Away Team)’s XG nodes are shown on the left-hand side of the pitch.

XG nodes are visualized by using seaborn.regplot whereas, the rest (the pitch) are visualized by using Matplotlib.pyplot.

4.8.Visualization of Important Action Heatmaps

Important action heatmaps for a filtered team shows the distribution of coordinates where the filtered team creates chances.

As mentioned in [Section 2], the heatmap only considers the important action sequences (An action sequence by a team that yields a shot action) and includes only shot-assist-final pass (assist of assist) actions in it. For instance, in [Figure 4], both important action heatmaps for Chelsea and Man City can be observed.

Green colored points are visualized by using seaborn.regplot, the corresponding heatmap is visualized by using seaborn.kdeplot and the rest (the pitch) are visualized by using Matplotlib.pyplot.

5.DEMO FOR INTERFACE

You can use the following link to watch a short demo video for the developed visualization interface:

https://youtu.be/fi_c6__25eE

REFERENCES

[1]: Statsbomb. (n.d.). Statsbomb/open-data. Retrieved January 16, 2021, from https://github.com/statsbomb/open-data

[2]: Perin, C., Vuillemot, R., &amp; Fekete, J. (2013). SoccerStories: A Kick-off for Visual Soccer Analysis. IEEE Transactions on Visualization and Computer Graphics, 19(12), 2506-2515. doi:10.1109/tvcg.2013.192

[3]: Football Like Never Before. (2020, October 06). Retrieved January 16, 2021, from https://statsbomb.com/