Top Clubs and Their Claim to be Comeback Kings of Europe

Group Team Number: 14
Joshua Matt Abraham, jma672
Ayan Agarwal, aka398
Gauri Sarode, gss383 (Online Class)

Project goal

From a fan's perspective, comebacks are one of the most exhilarating things in football. Who wouldn't love watching the team they support go from 0-2 down to win 3-2 with the last kick of the game? (The opponent maybe?). Even the players themselves get a rush from such a dramatic finish. The goal of our project is to determine which club in world football deserves the title of “Comeback Kings”. Every club wants it, but surely only a few deserve the title. We go on to examine the reasons for these comebacks and analyze why certain clubs achieve comebacks more frequently than others.

Data Set

We are using 2-3 datasets that can be merged based on a particular match.

- ❏ Football Europe Statistics dataset – https://www.kaggle.com/jangot/ligue1-match-statistics
  The dataset contains statistical information for matches across Europe along with their results. Attributes such as possession, shots, formations, tackles, dribbles and goals by the home and away teams will be used to determine the results of these matches, whether teams made comebacks and a few statistics that could have fueled the comeback. These attributes are mainly quantitative.

- ❏ Football Match dataset - http://www.football-data.co.uk/
  This dataset contains attributes corresponding to important events in matches such as when comebacks occurred in matches, with their half time and full time scores.

Analytical Questions and Proxy Tasks

- ❏ ***Question***: Which are the biggest clubs in world football? (Introductory question that will further lead to clubs which can be considered comeback kings)
  Proxy Tasks:
  Which are the clubs with the most wins since their inception?
  Which are the clubs that have gained the most points in each season of a league?
  Which are the clubs with the largest fan attendances in their matches?

- ***Question***: Which clubs are the best at comebacks?
  Proxy Tasks:
  Which clubs have secured the most draws or wins after trailing first?
  What are the largest margins by which a club has comeback to draw or win a match and which are these clubs?
  How frequently have these clubs made comebacks and often have they failed to do so?

- ***Question***: What are the reasons that clubs attain comebacks?
  Proxy Tasks:
  Does playing in your home ground increase the chances of achieving a comeback?
  Are comebacks a result of drastic substitutions or do they occur without many substitutions as well?
  Does increased possession and fouls against a club inflate the chances of a comeback by that club?

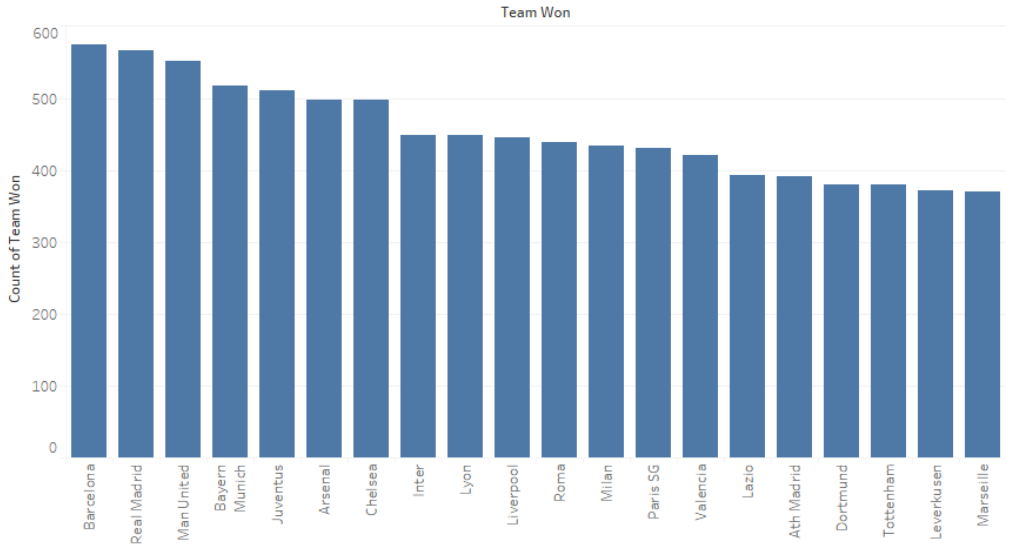
- ***Question***: How does attaining a comeback affect a club's mentality?
  Proxy Tasks:
  What were the results in the next 5 matches after a comeback?
  Does the losing team also start losing their next matches?

Story Design

Data Analysis

- ***Question***: Which are the biggest clubs in world football? (Introductory question that will further lead to clubs which can be considered comeback kings)
  Proxy Tasks:
  Which are the clubs with the most wins since their inception?
  Which are the clubs that have gained the most points in each season of a league?
  Which are the clubs with the largest fan attendances in their matches?

Sheet 4

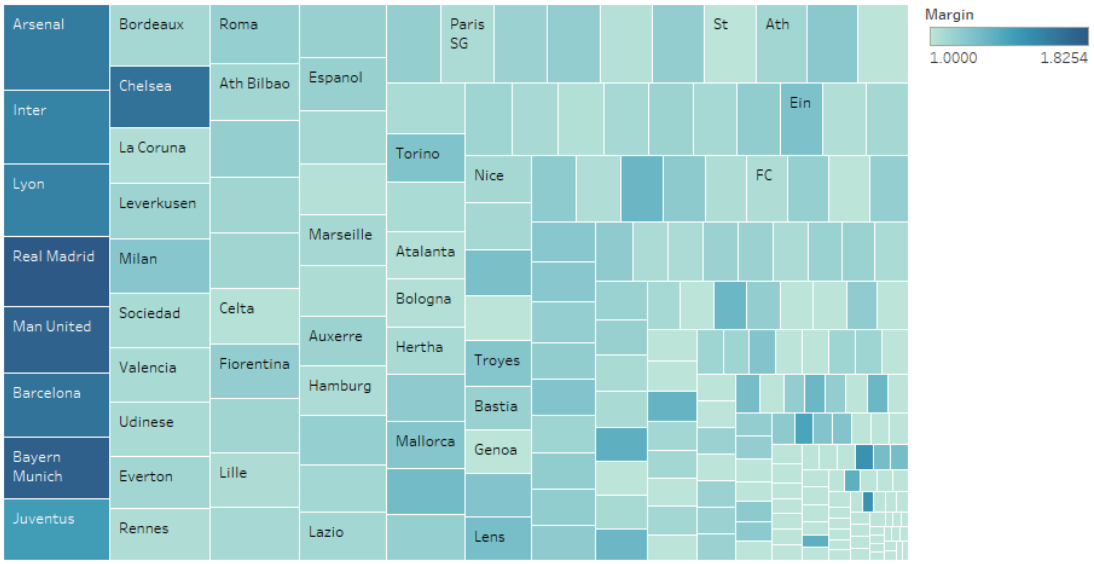


Count of Team Won for each Team Won. The view is filtered on Team Won, which keeps 20 of 229 members.

The top 20 teams based on points and most number of wins are shown above.
The top 10 teams are Barcelona, Real Madrid, Man United, Bayern Munich, Juventus, Arsenal, Chelsea, Inter, Lyon and Liverpool.

- ***Question***: Which clubs are the best at comebacks?
  Proxy Tasks:
  Which clubs have secured the most draws or wins after trailing first?
  What are the largest margins by which a club has comeback to draw or win a match and which are these clubs?
  How frequently have these clubs made comebacks and often have they failed to do so?

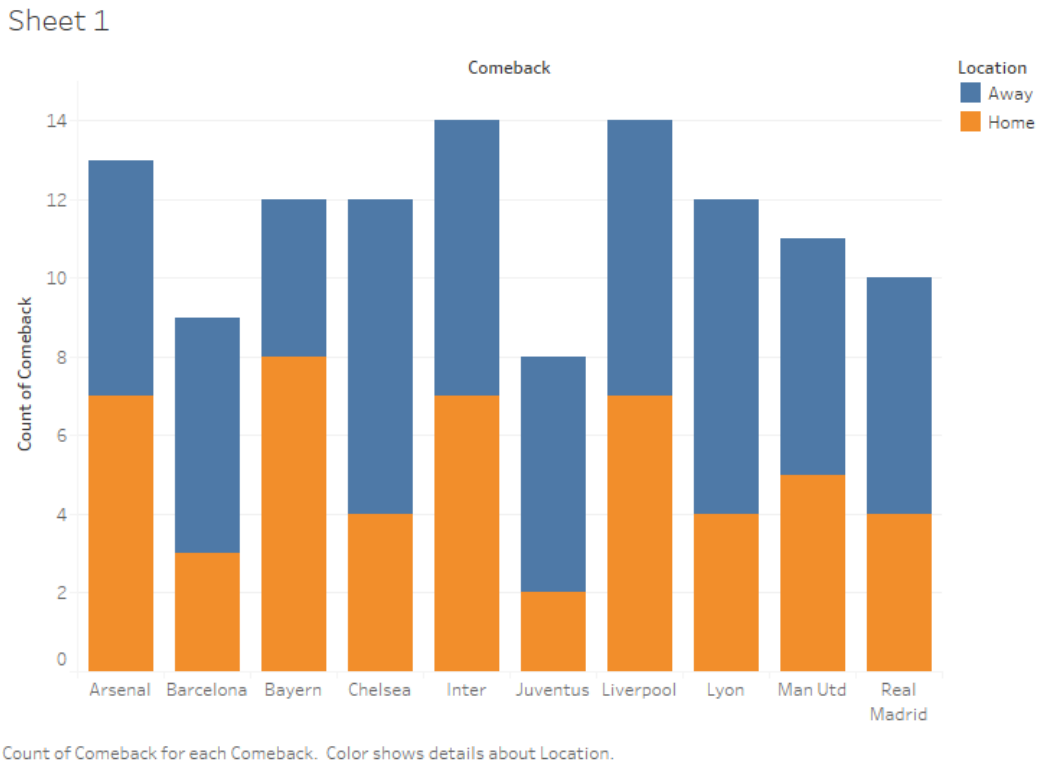
Sheet 2 (2)



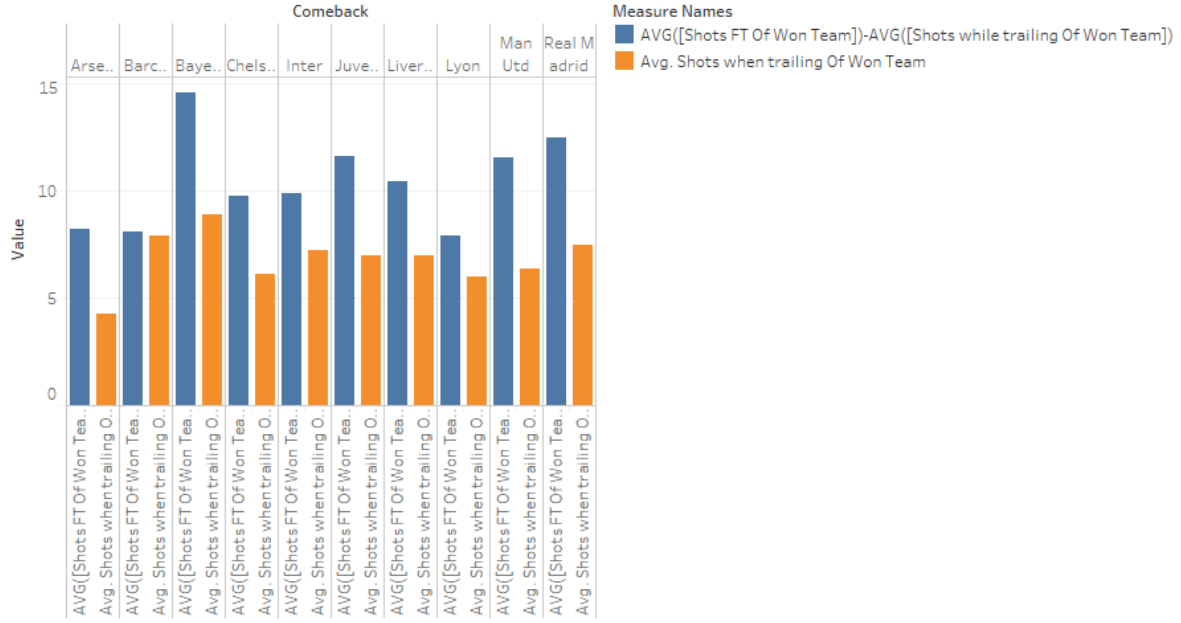
Comeback. Color shows average of Margin. Size shows count of Comeback. The marks are labeled by Comeback. The data is filtered on Team Won, which has multiple members selected. The view is filtered on Comeback, which excludes Aachen, Evian Thonon Gaillard, Extremadura and False.

This treemap shows the teams that have the most comebacks (with size), and highest margins of comebacks (with intensity). It has been filtered based on Teams that won the most matches. As we can see, the top teams generally have the most comebacks and even the highest margins of comebacks.
Chelsea is one of the standout clubs outside the top 10 that have a high comeback margin.

- ***Question***: What are the reasons that clubs attain comebacks?
  Proxy Tasks:
  Does playing in your home ground increase the chances of achieving a comeback?
  Does increased possession and fouls against a club inflate the chances of a comeback by that club?
  Do teams making comebacks attempt more shots than when they were trailing?

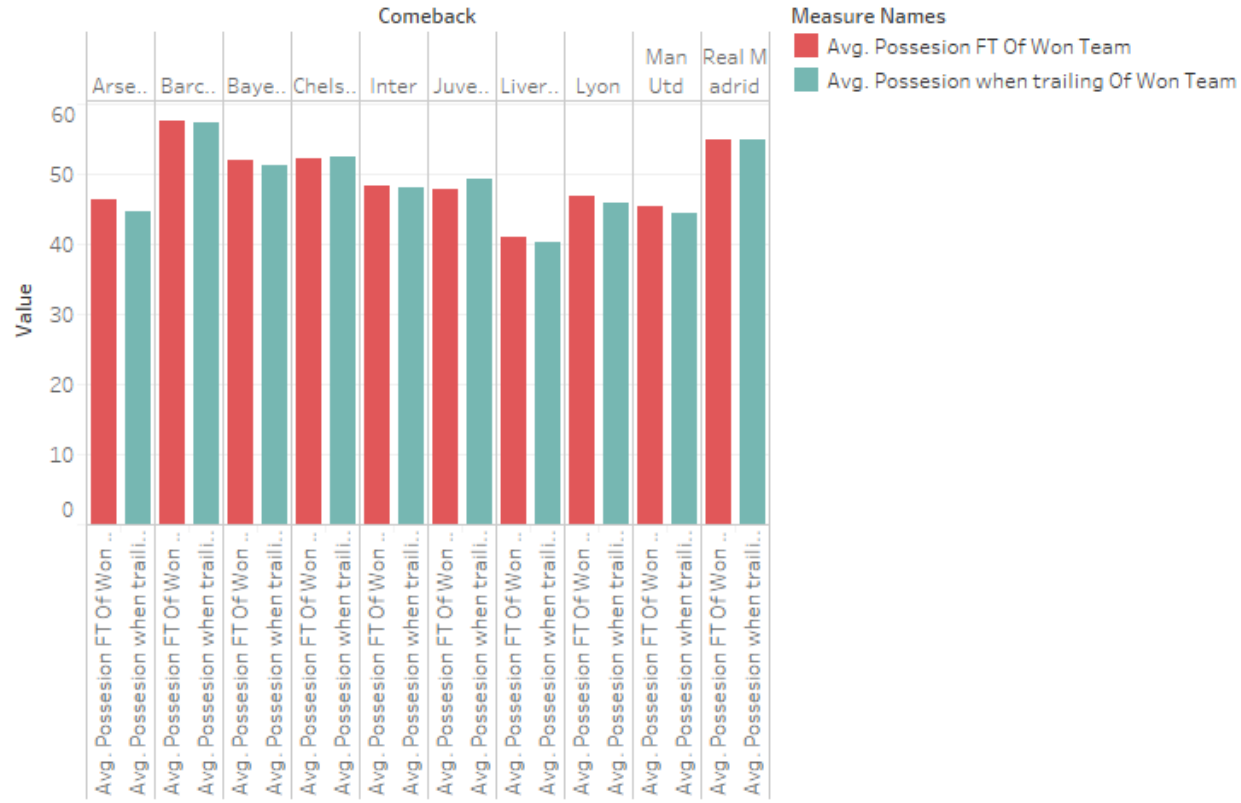


Sheet 2



AVG([Shots FT Of Won Team])-AVG([Shots while trailing Of Won Team]) and Avg. Shots when trailing Of Won Team for each Comeback. Color shows details about AVG([Shots FT Of Won Team])-AVG([Shots while trailing Of Won Team]) and Avg. Shots when trailing Of Won Team.

Sheet 3



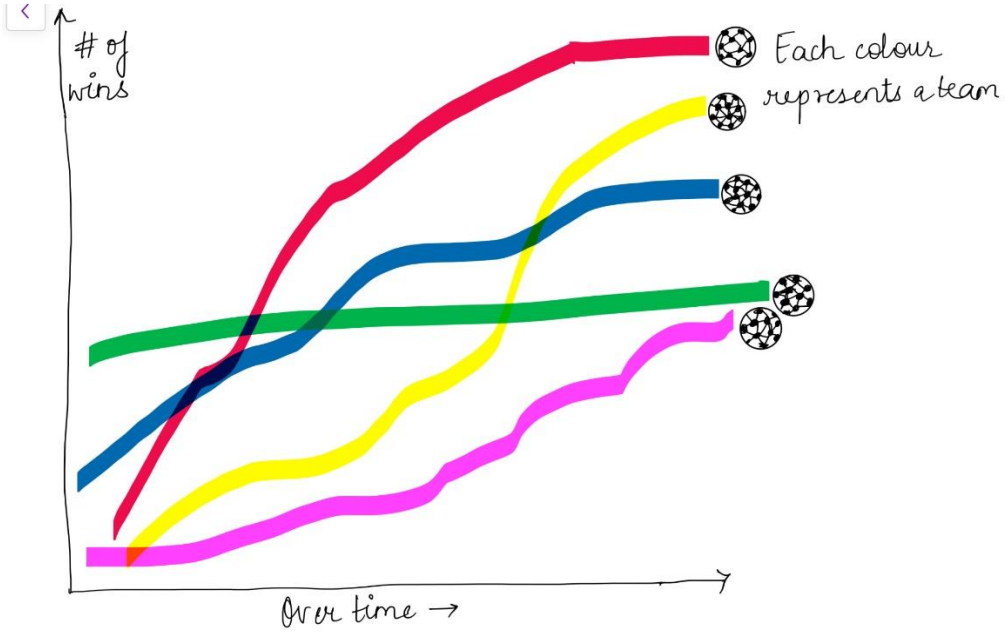
Avg. Possesion FT Of Won Team and Avg. Possesion when trailing Of Won Team for each Comeback. Color shows details about Avg. Possesion FT Of Won Team and Avg. Possesion when trailing Of Won Team.

From the 4 graphs above, we can say that the top 10 teams achieve comebacks more often playing away from home. This could indicate a tendency for the bigger teams to grow in the game slower, while the losing team starts brighter because they are at home.
Also, in most cases the, the teams trying to comeback attempt much more shots to do so, and also the opponents tend to concede more fouls in this period.
Most surprisingly, the possession before and after a comeback is almost identical in all the cases. So, possession does not have a bearing on the chances of a comeback.

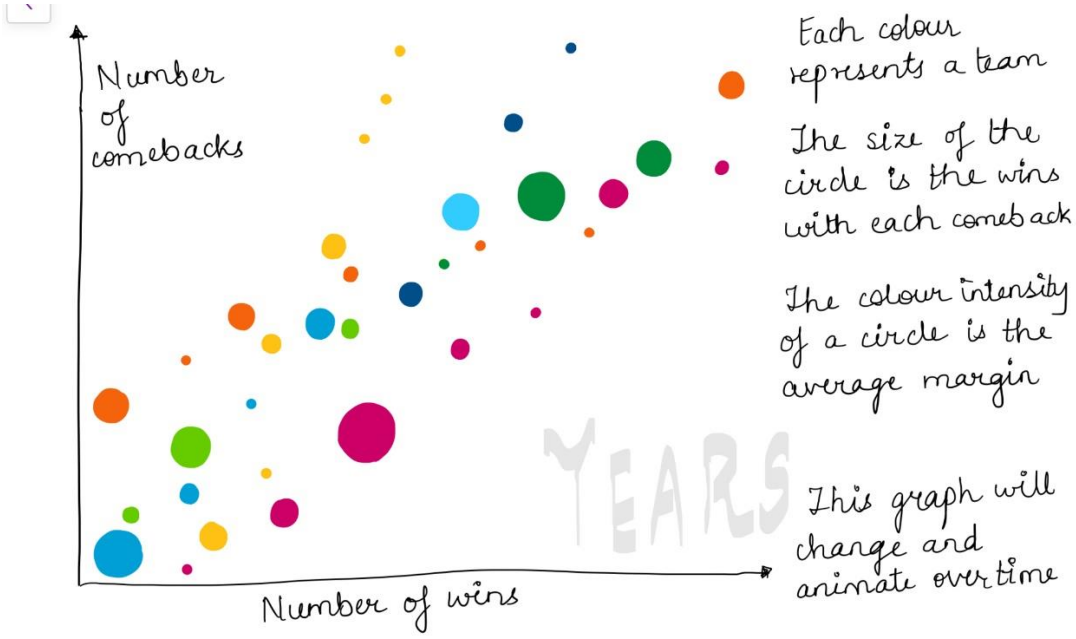
Storyboard

Section 1:
In this section we propose to determine which clubs should be considered the best in football. We determine this by simply observing the number of wins for each club since their inception. We would like to show this with the help of a multi-line graph. The path of the lines would show how these clubs won over time and dominated different eras. X-axis would represent time, and Y-axis represents number of wins.

Imagine the below graph, with each line growing. In the end the top 10 clubs would cluster high up to the right of the graph.

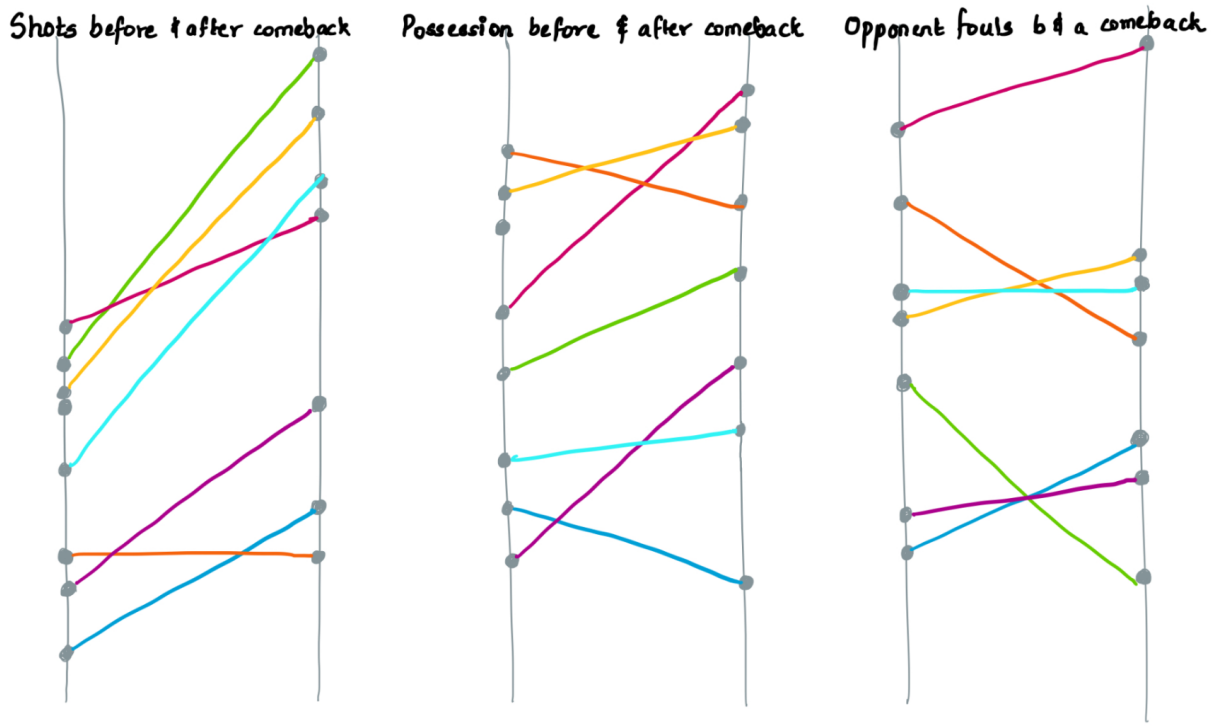


Section 2:
We then filter out the top 10 or 5 clubs and visualize how these clubs perform comebacks. It will be in the form of scatter plot with each plot representing 4 different attributes.
X-axis represents total number of wins.
Y-axis represents total number of comebacks (including draws and wins)
The size of the plot shows the average of margin of a comeback.
One plot represents one club.



Section 3:
We try to show how these comebacks correlate with certain statistics like possession, shot attempted, fouls conceded by opposing team, home or away win, etc. With this we visualize

how these attributes affect or help comebacks. We can use a slope plot for the same, where the left axis would represent the statistic before a comeback, and the right would represent the statistic after a comeback.



Implementation:

Github Link – https://github.com/NYU-VIS-FALL2018/storytelling-group-14
Demo Link – https://nyu-vis-fall2018.github.io/storytelling-group-14/

Changelog:

1. Preformed various data processing on the original dataset to fit our visualization needs.
2. Removed the last analytics question as from feedback.
3. Improved every basic tableau visualisation of analytical questions with effective solutions in D3.
4. As per the feedback, we changed certain aspects in our visualisation like, not including intensity and removing the time relation as the 2nd graph had too many attributes to understand.
5. Changed the section 3 graph from a sunburst to parallel slope graphs as it was a simple an effective way to represent our problem.
6. Made selection of a team uniform across all graphs at once, so that a user can visualise the performance of 1 club at a time.
7. Added a story highlight feature in the first graph to highlight some important stories in the growth of certain clubs.