Dynamics of Passes in Football

**Process Book**

Data Visualization (CPSC 6030)

Fall 2021

Group 11

Anoop Kakkireni & Bashir Hamidi

Table of Contents

[**1.** **Overview and Motivation:** 3](#_Toc89609601)

[**2.** **Related Work:** 3](#_Toc89609602)

[**3.** **Questions:** 4](#_Toc89609603)

[**4.** **Data:** 4](#_Toc89609604)

[**5.** **Exploratory Data Analysis:** 4](#_Toc89609605)

[**6.** **Design Evolution:** 4](#_Toc89609606)

[**7.** **Implementation:** 4](#_Toc89609607)

[**8.** **Evaluation:** 5](#_Toc89609608)

[**9.** **References** 5](#_Toc89609609)

***Project report:****An important part of your project is your process book. Your process book details your steps in developing your solution, including the alternative designs you tried, and the insights you got. Develop your process book out of the project proposal. Equally important to your final results is how you got there! Your process book is the place you describe and document the space of possibilities you explored at each step of your project. It is not, however, a journal or lab notebook that describes every detail - you should think carefully about the important decisions you made and insights you gained and present your reasoning in a concise way. We strongly advise you to include many figures in your process book, including photos of your sketches of potential designs, screen shots from different visualization tools you explored, inspirations of visualizations you found online, etc. Several images illustrating changes in your design or focus over time will be far more informative than text describing those changes. Instead, use text to describe the rationale behind the evolution of your project. Your process book should include the following topics. Depending on your project type the amount of discussion you devote to each of them will vary:*

* *Overview and Motivation: Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.*
* *Related Work: Anything that inspired you, such as a paper, a web site, visualizations we discussed in class, etc.*
* *Questions: What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?*
* *Data: Source, scraping method, cleanup, etc.*
* *Exploratory Data Analysis: What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?*
* *Design Evolution: What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?*
* *Implementation: Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.*
* *Evaluation: What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?*

*As this will be your only chance to describe your project in detail make sure that your process book is a standalone document that fully describes your results and the final design.*

To do list:

1. video recording
2. Add differences between the dataset in publication and dataset you had. What were the differences, what were the limitations you had that prevented you to create a figure like the publication?
   * Machine learning algorithems to generate phases we analyzed
   * Format of the dataset was different
   * We did not player movement data without the ball
3. Figures introducing the dashboard. It does not have to be all the figures at once and can be in pieces of the dashboard.
4. ~~move the timeline axis to middle of the bar chart and nodes~~
5. ~~change color in the grey area for player based on positions~~
6. ~~match the color of grey area players to the nodes of timeline~~
7. add legend for start and end position color
8. move x-axis label to the end of the axis
9. debug color issues
10. debug 2 showing up instead of 22
11. debug minutes with no pattern
12. ~~add goalie to stats board~~
13. stats board, when you select a player, keep the highlight so we know which player is selected
14. when you select a bar graph, highlight the background bars of the node graph (b) to indicate that all those nodes are selected and are shown

## **Overview and Motivation**

*Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.*

Motivation:

Football is a passing game. Although the outcome of a game is determined by the number of goals scored, it wouldn’t be possible to score goals without passing the ball. It is also important to acknowledge the fact that goal scoring is a team effort. The contributions of players in different positions must be accounted for to develop a proper understanding of the impact of players in a particular game.

Moreover, there is a misconception that goal scorers are the most important players in a team. From an informed standpoint, this is not true in most cases and there are several players contributing to the success of a team. This dashboard will clear this misconception and enable the viewer to have a better understanding of the game and players’ importance represented through passing dynamics.

## **Related Work**

*Anything that inspired you, such as a paper, a web site, visualizations we discussed in class, etc.*

Publication:

We were inspired by the publication “PassVizor: Toward Better Understanding of the Dynamics of Soccer Passes”, published in IEEE Transactions on Visualization and Computer Graphics. The idea behind the paper is to derive underlying patterns in football passes to try to better understand the attacking styles. This helps in analyzing how players build an attack using different passing tactics.

ADD differences between the dataset here and dataset you had. What were the differences, what were the limitations you had that prevented you to create a figure like the publication?

## **Questions**

*What questions are you trying to answer?*

*How did these questions evolve over the course of the project?*

*What new questions did you consider in the course of your analysis?*

Initial Stage

In the early stages of the project implementation, we were very interested in exploring the relationship between the transfer fees of the players and the success of the teams in the season following the transfer. We were interested in exploring the following questions:

* 1. What are the passing patterns of various teams in the premier league?
  2. What positions (forward, center, defense) have the teams spend most of their funding on?
  3. Does the spending on transfer fees have a favorable impact on the team in the following season?
  4. What are the impact of transferred player age and position on the team’s outcome in the following season?

Intermediate Stage

As we explored the available data in our dataset as well as learned more about the requirements of the project, we learned that we needed to better align the project dataset and questions to mirror that of the selected research publication. At this stage, we moved to a different dataset and came up with the following research questions:

1. How does passing behavior affect the outcome of a game?
2. What is the impact of different players in the overall performance of the game?
3. What are the passing styles and patterns of a particular team?

From these questions it is still very evident that we were still interested in comparing across different teams. This proved challenging during the implementation of our project and availability of public datasets.

Final Stage

By now we had explored five different datasets, and each had their pros and cons. After selecting a final dataset and working on the project, it became evident that comparing performance across teams would not be feasible given the extent in which we had to manipulate the dataset to get it to a desirable state.

At this point, we shifted our focus from wanting to compare teams to comparing the passing of different players within a team. Our questions evolved to:

Have you ever wondered the extent to which passing impacts the outcome of football? To study this, we have built a dashboard with the primary goal of answering the following research questions:

1. How do the passing dynamics differ between players within a game?
2. What is the impact of different passing styles on the overall outcome of a game?
3. What are the passing styles and patterns of a particular team?

## **Data**

*Source, scraping method, cleanup, etc.*

Initial Stage

At this stage in the implementation of the project, we had selected two datasets. The first dataset (original data sourced from Football-Data [1], compiled and downloaded via a data hub[2]) was on the English Premier League teams such as home and away goals, yellow and red cards. This dataset is organized by specific seasons into individual comma separated value (CSV) files. In addition to this, we wanted to incorporate a second dataset that contained top 250 transfer fees per season from the year 2000 to 2018 [3].

Intermediate Stage

As discussed in other sections of this report (such as Section 3, Questions), this project evolved during the implementation cycle. At this stage, we considered several other datasets and decided to proceed with a FIFA World Cup dataset from 2010, 2014, and 2018 [4]. published under the Creative Commons Attribution 4.0 International license. We proceeded to use this dataset for some time but eventually moved away from it because it did not contain the individual pass-level data that we needed to mirror the original publication’s visualizations.

Final Stage

The data within the implemented project was compiled by StatsBomb and made freely available via the StatsBomb Open Data repository for public use and research projects [5]. Data was separated by individual events in JSON files. For our analysis, we selected a JSON file representing a match between Real Madrid and Liverpool with a goal outcome of 3-1. This single filed contained over 150 thousand lines of data with roughly 3500 observations of events.

Given that Real Madrid scored three goals within this game, we decided to look at their passing patterns. First, we subset the dataset to only include the events involving players from Real Madrid.

Because there are different SVG elements within our dashboard, depending on the need, we have a different subset of data. For example, to graph the bar chart and the nodes, we subset the dataset to only include the passes of Real Madrid players that were not intercepted. However, to calculate the statistics, we included all the data of Real Madrid players and used python to calculate the number of passes each individual player was involved with as well as those passes that were successful (not intercepted). Using this data, we were able to calculate the passing accuracy percentage for each player and create our own JSON dataset. Within this dataset, we also calculated and included the number of shots taken by each player, goal assists (passes to someone else who made a goal), and goals made.

## **Exploratory Data Analysis**

*What visualizations did you use to initially look at your data?*

*What insights did you gain?*

*How did these insights inform your design?*

Because our dataset and research questions evolved quite a bit during the implementation of this project, many of the exploratory steps performed were on datasets that we ended up not using for the final implementation of the dashboard. For example, in the “Initial Stage” when we were interested in delving deep into the trade-fee of players as well as assess trends within teams, we did a summary figure of each team and position within the team and saw several outliers. First, we noticed that some teams paid significant sums of money (figure shown is in Euros) to trade center-forward players. Of note were Arsenal, Chelsea, Manchester City, and Manchester United who often paid an overall order of magnitude for center-forward players as compared to all other positions. We were very interested in exploring these trends; however, due to the fact that this did not mirror the visualization in the publication of interest, we moved away from this.

Figure 1: An exploratory assessment of the "Initial Stage" dataset to assess the granularity and availability of trade-fee (in Euros) data for two consecutive seasons.

## **Design Evolution**

*What are the different visualizations you considered?*

*Justify the design decisions you made using the perceptual and design principles you learned in the course.*

*Did you deviate from your proposal?*

From the planning phase of the project all the way to the end of the implementation, our design ideas and visualizations changed significantly. There were three primary reasons for this: first, our datasets and research questions changed several times; and second, as we implemented the visualizations, we realized shortcomings in the data and as a result we had to make changes to our proposed designs; and third, during the process of implementation, incremental and design related changes had to be made to the visualizations to accommodate the viewer and the desired outcome of the dashboard.

In this section, we will only focus on the evolution of our dashboard and visualizations using the final dataset we selected.

Our proposed visualizations midway through the semester included several figures some of which are shown below. Some of these figures were sourced from analytics websites looking at football data. At the early stage we proposed to develop network graphs of passes between players of teams as shown in Figures 2 and 3 below [6, 7]. We believed that such network graphs would give us an idea of the passing style and the play style of a particular team. Additionally, we felt that network graphs will highlight the key players responsible for creating chances throughout the game. The network graph has the ability to convey the patterns in passing behavior because it actually depicts the location of players on the field. It is easy for the viewer to understand the visualization based on their understanding of the game. The thickness of the lines or the saturation of the lines used to connect different players can be used to portray the number of passes between them.

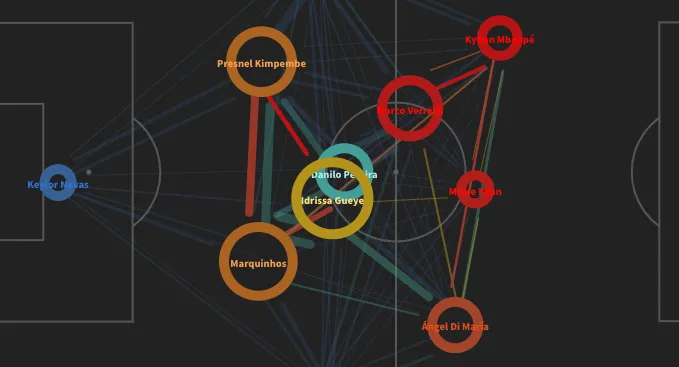


Figure 2: Network graph of passes [6]

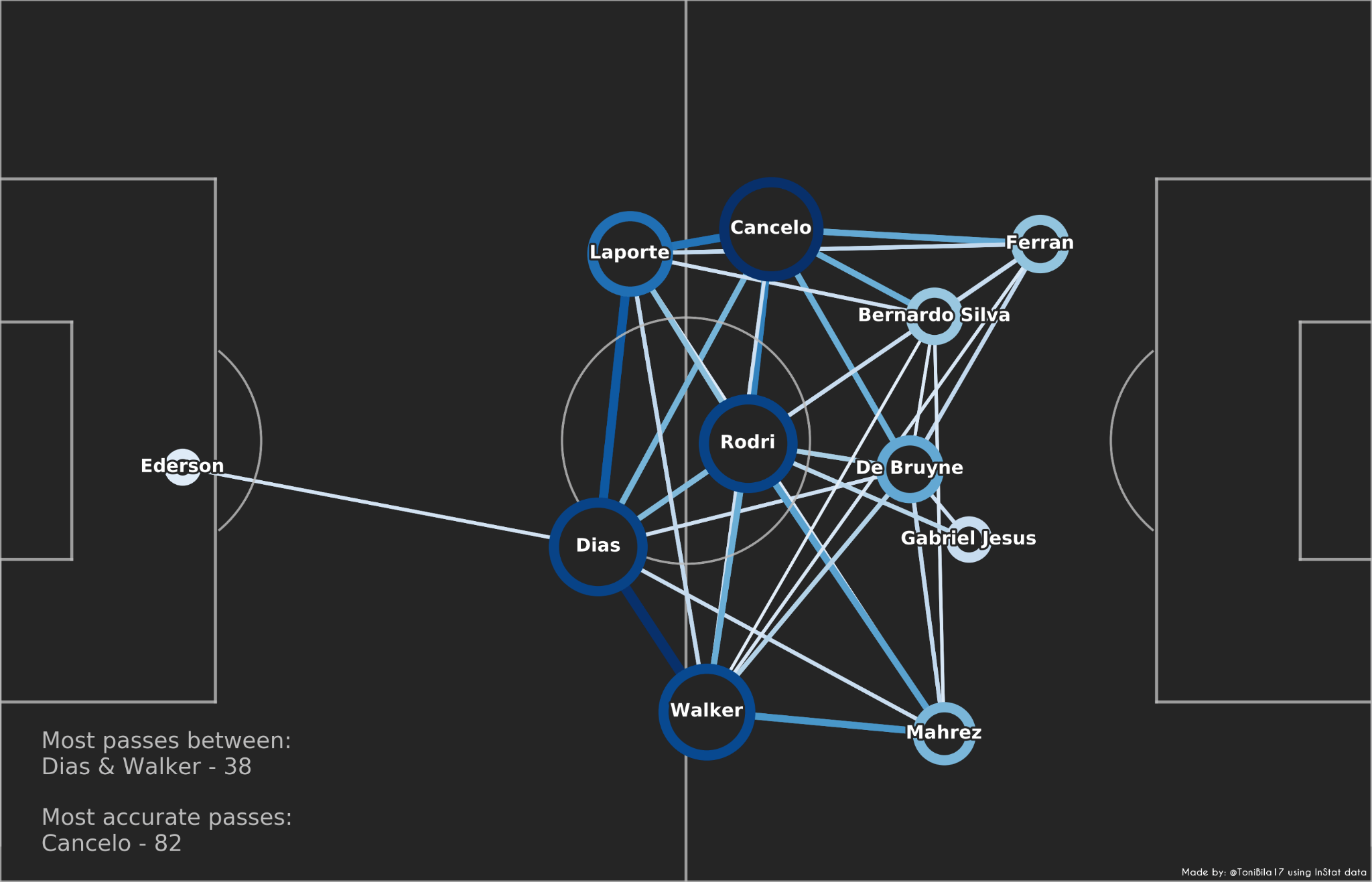


Figure 3: Network graph of passes [7]

Another one of our proposed visualizations at this early stage was the creation of a radial graph as shown in Figure 4 below to showcase the skill level of various players and teams[6]. The radial graph was appealing to us because of the fact that the metrics in such a figure will be normalized and hence will give a better picture of the contributions of different players. The intention behind using a radial graph is to estimate the competency of a player in each of the attributes revealing the weaknesses and strengths. We felt that the same idea can be extended by aggregating the data based on the position of the players namely, midfielders, attackers and defenders.

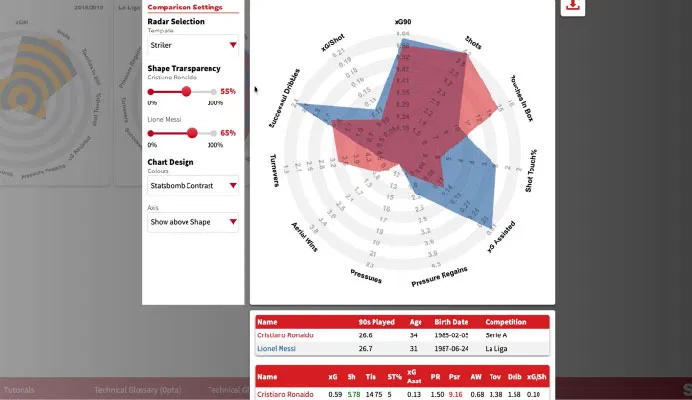


Figure 4: Radial graph depicting statistics of teams and players[6]

As we have briefly outlined in Sections 3 and 4 within this report, our datasets and research questions shifted throughout the project. This was primarily to mirror the implementation of the figures shown in the publication selected shown in Section 2, Related Work.

We would like to briefly acknowledge the importance of adhering to the requirements of projects as you move to research and implement said project. Whether the project is an assignment for a course, work project, or grant, the fact remains the same that adhering to the requirements are very important. The evolution of this project was not linear primarily because we failed to adhere to the requirements of the project in the early stages by properly selecting a dataset that was similar to the one in the publication as well as proposing figures that were identical to the dashboard of the publication. Acknowledging this experience is an important step in the evolution of this project as well as a learning experience for the both of us.

As we moved to a dataset [5] that related more closely to the data within the publication, we began working on replicating the dashboard within the publication. It should be made clear that although we made our dashboard to be similar to that of the publication, there are some major limitations. First, although our dataset contained similar information, it did not include near the extent of the data present within the publication’s dataset. And second, the publication included major methods in statistical analysis including deep learning and matrix computations and the code from these methods were not shared publicly to be reused. Therefore, we did not perform any of these analyses.

Within this dashboard, we began work on a bar chart that communicates the total number of passes of Real Madrid players within this particular match against Liverpool for every minute of the game. Shown below in Figure 5 we see one of the first implementations of our bar chart.

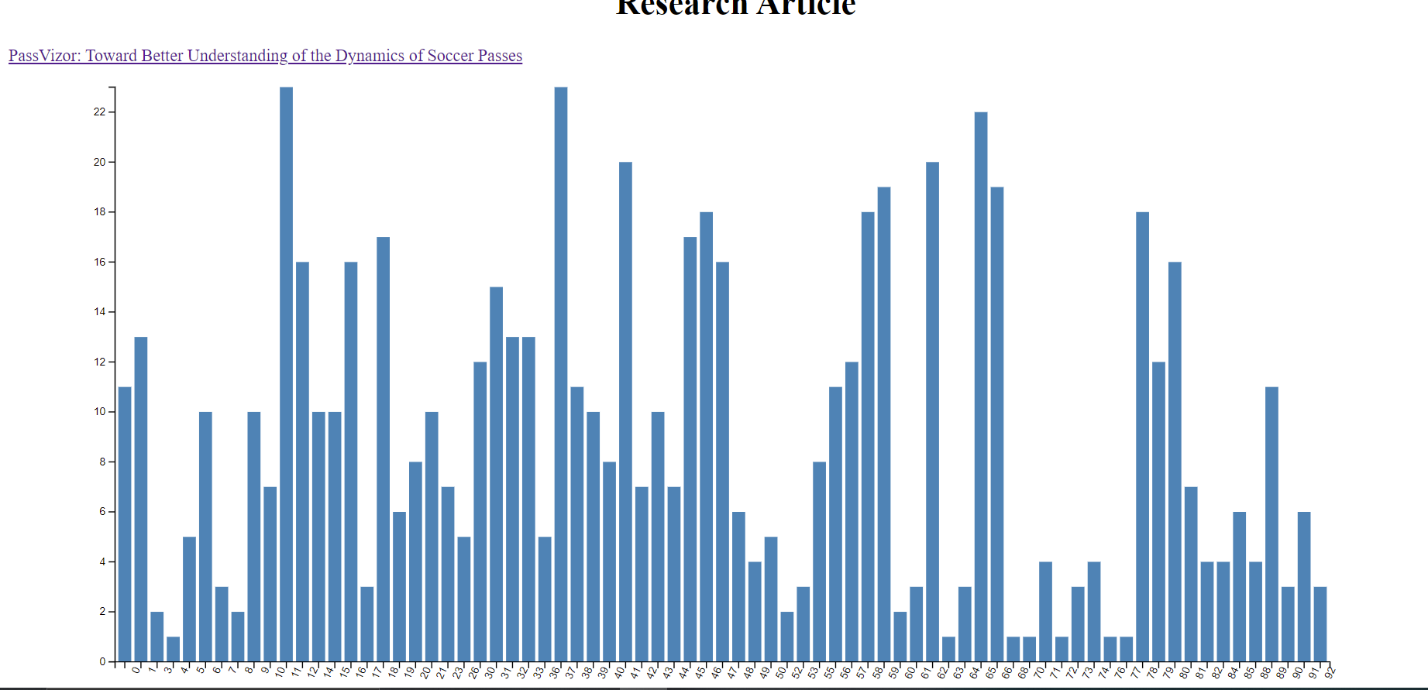


Figure : Bar chart within the dashboard showing the total number of passes within each minute.

Next, we began working on a node and line graph that separated the passes to each individual player shown across horizontal lines. We have shown an early implementation of this in Figure 6 below. As the Figure caption indicates, individual players are on the y-axis and jersey numbers are shown as labels. For every pass a player has made during a particular minute, a node should be drawn on this figure. As you see in this early implementation, all nodes are shown in the upper left-hand corner of the SVG element overlaying one another.

As you observe, the bar chart shown in Figure 5 has a roughly double the height of the one shown below in Figure 6. As we added SVG elements to our dashboard, we had to play with the height to overcome the issue of size.

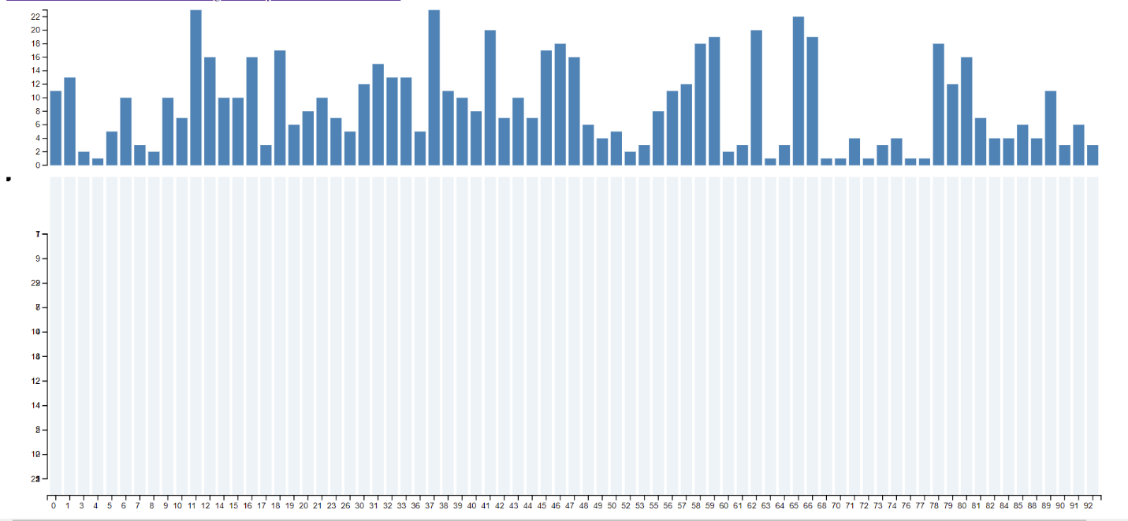


Figure : Node and line graph showing detailed passing information of individual players. On the left-hand side of the figure below the bar chart you see a node. This represents all the nodes for all players that are to be distributed across lines representing players. This is an early implementation showing our progress.

We were able to resolve the issues of the nodes and visualize the detailed player-level passing data on individual lines, as shown in Figure 7.

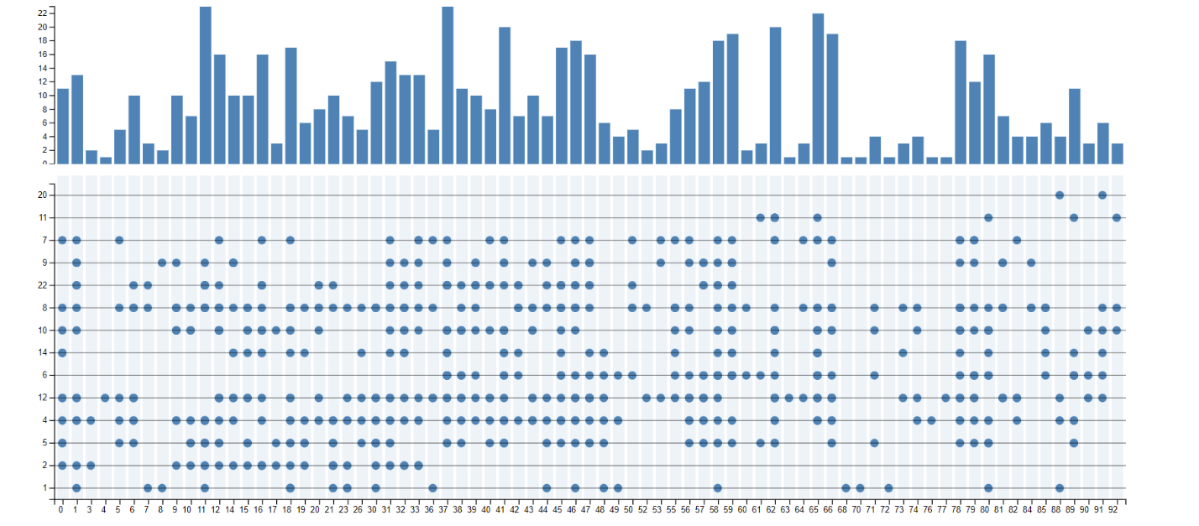


Figure : Node and line graph successfully showing player-level passing data.

Next, we began implementing our passing process figure. We were able to pull all the x- and y-coordinates for Real Madrid players involved in a pass and visualize them as nodes as seen in Figure 8A.

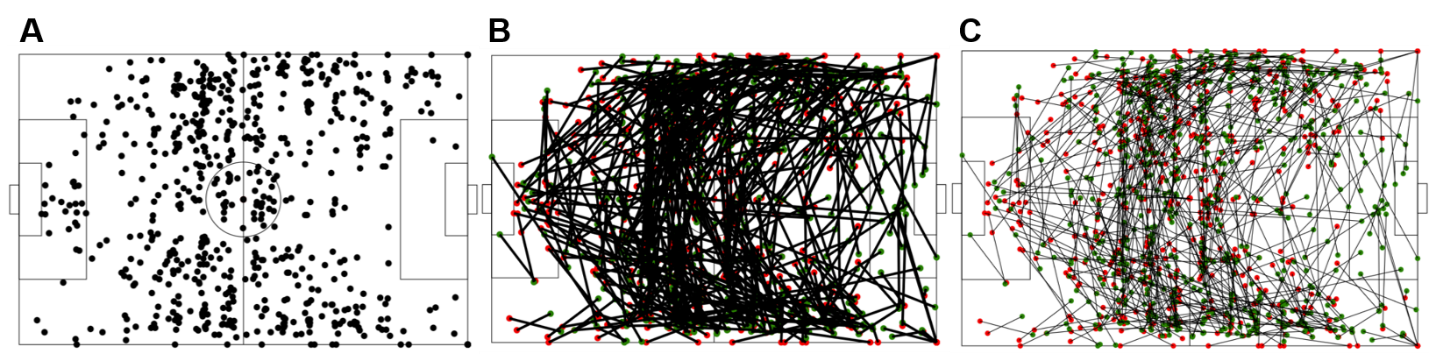
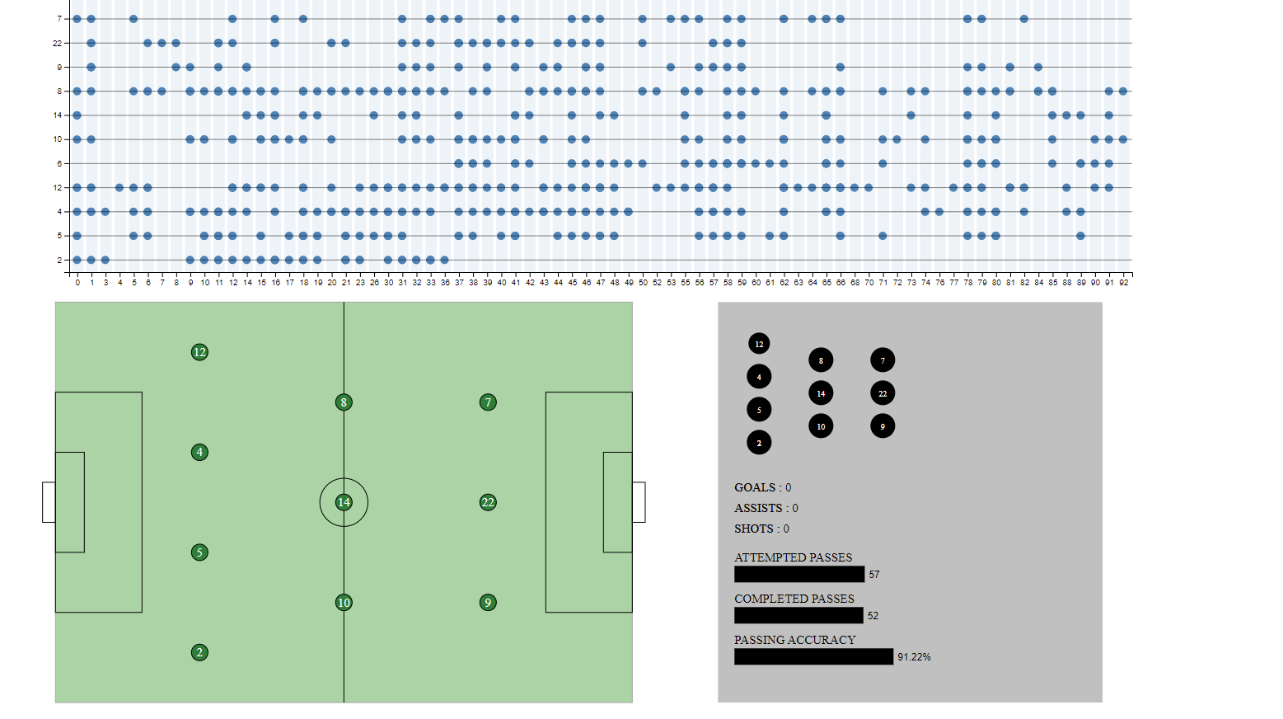


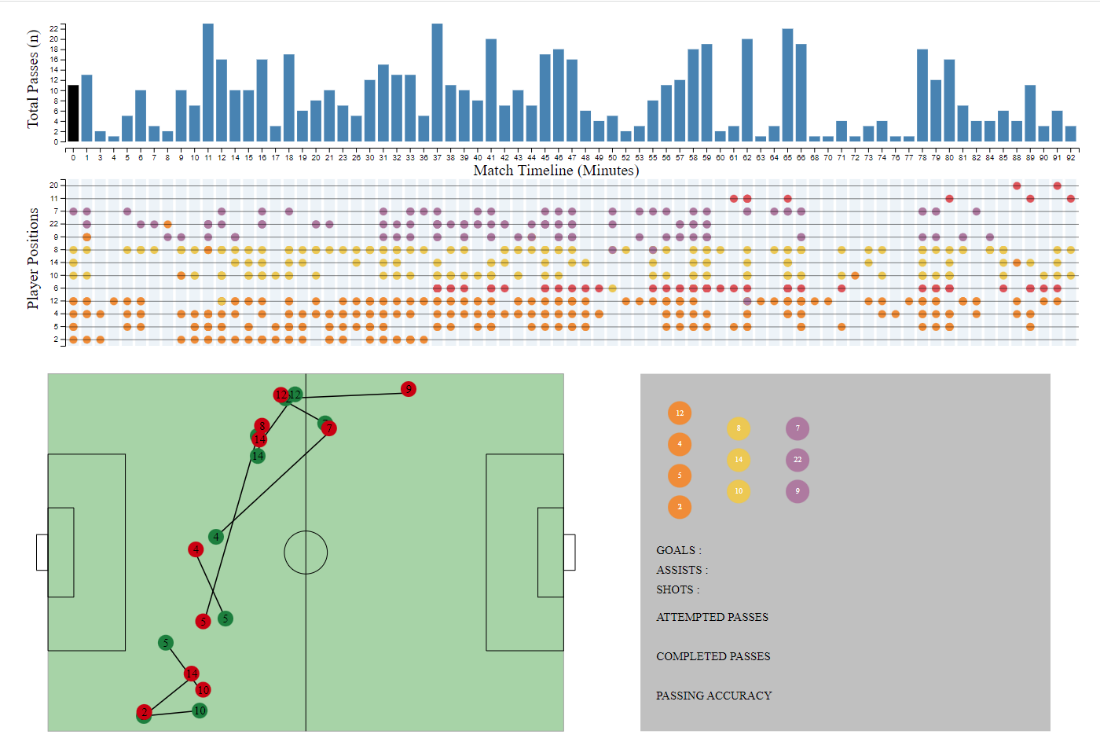
Figure : Position of Real Madrid players involved in passes is shown in A; B adds lines connecting the players involved in a pass event; and C reduces the weight of line connecting the nodes to allow better observations of the patterns.

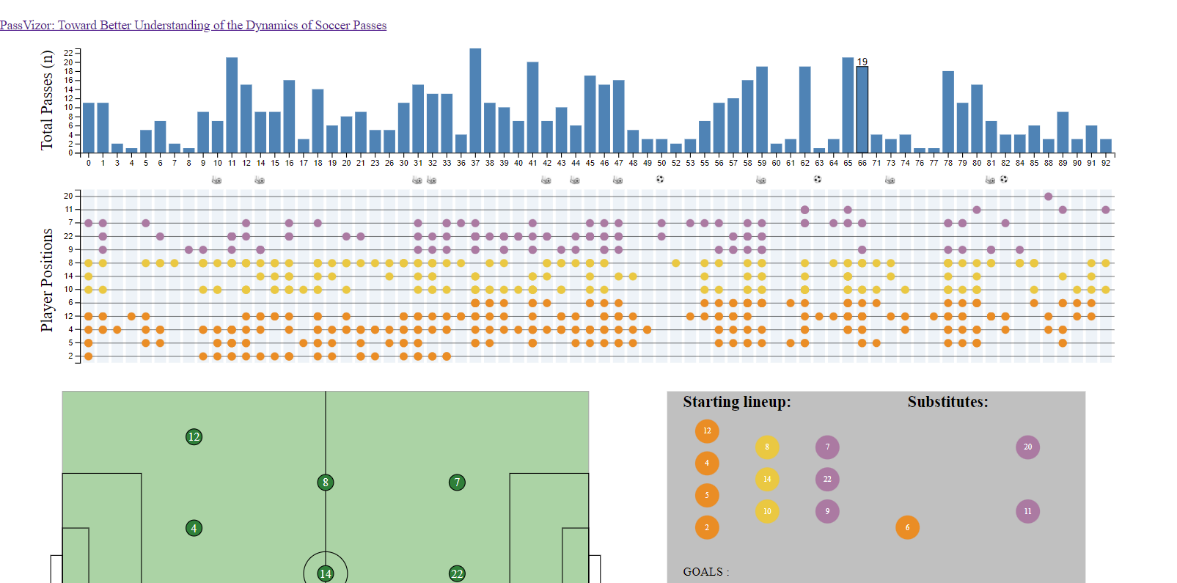
We were able to add to this figure by connecting the players that are involved in individual pass events involving two or more players (8B). We visualized the passing process by further reducing the weight of connecting lines as seen in 8C.

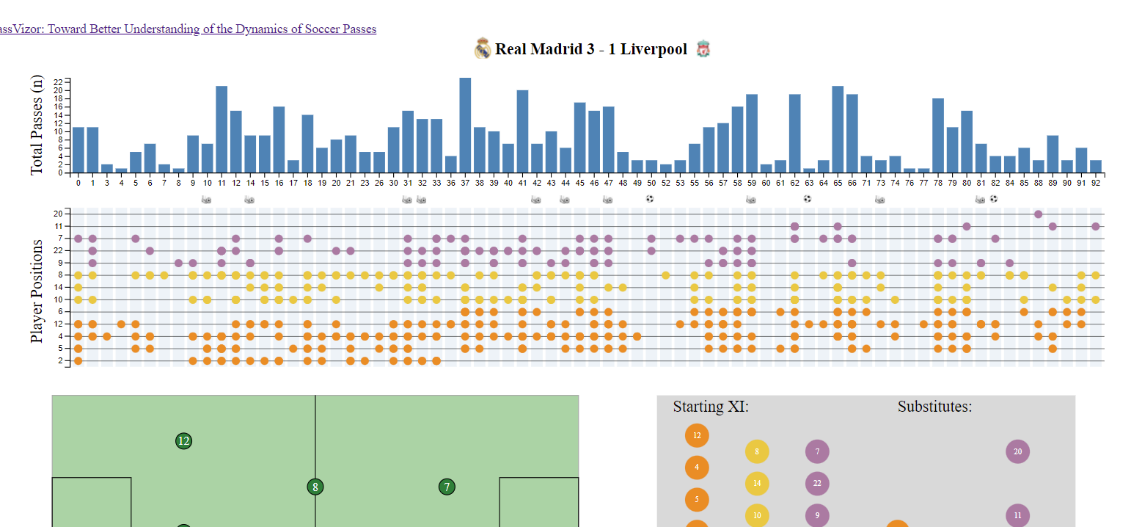
We made some significant We then added a statistics board











## **Implementation**

*Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.*

## **Evaluation**

*What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?*

## **References**

*Make reference format consistent*

1. *Football-Data*. [cited 2021 20 September 2021]; Available from: <http://www.football-data.co.uk/>.

2. *English Premier League (football)*. [cited 2021 20 September 2021]; Available from: <https://datahub.io/sports-data/english-premier-league>.

3. Lehkyi, S. *Football Transfers 2000-2018*. [cited 2021 20 September 2021]; Available from: <https://www.kaggle.com/slehkyi/football-transfers-2000-2018>.

4. Martinez Mejorado, D.A., Ramirez Marquez, Jose Emmanuel, *FIFA World Cups 2010, 2014 and 2018 matches' information datasets*. 2018: Zenodo.

5. *StatsBomb Open Data*. Available from: <https://github.com/statsbomb/open-data>.

6. *Match Analysis*. How to incorporate StatsBomb data and analytics 20 October 2021]; Available from: <https://statsbomb.com/>.

7. *Manchester City’s passing network vs. Tottenham*. Total Football Analysis; Available from: <https://totalfootballanalysis.com/article/manchester-citys-passing-network-vs-tottenham>.