DATA
VISUALIZATION
PROJECT

PROCESS BOOK

Visualize the NBA

https://github.com/com-480-data-visualization/com-480-project-freespin

TEAM FREESPIN

Subject and goals

The NBA records and keeps track of all kind of statistics and information about players. You can't watch a single NBA game without seeing a statistic about a player, a record being approached, broken or set.

Our goal for this project is to visualize interesting statistics about the NBA and NBA's players, principally their performances and their geographical aspect covering all the data made available from the foundation of the NBA.

We're a team of three. Two of us are basketball players and obviously huge fans of the NBA, with a solid background in this field with a deep understanding of all the detailed aspects behind a basketball game. It's why we'd like to make available and share this hidden aspect of the game to everyone.
Moreover, our last member, discovering this new world in the context of this project, adds a different point of view, more outside of the basketball world, and links different level of the knowledge of this sport.

Our project targets from amateurs to the professionals of Basketball. We tried that our project is approachable by everyone not having a huge background in basketball. We aimed to find an engaging and interactive so that anyone could play with the data and understand it easily. And maybe with our visualization, some of them could try to watch basketball games and become fans.

Our main idea is to show the performances of a player, his contribution to the different statistics of this game. Is he a scorer ? Someone who makes his teammates shine ? A beast ready to protect his rim at all cost ?
All kind of characteristic showing its impact on the game and inside his team are obviously very interesting in such a sport and this was pretty straightforward to go for a way to visualize this.

But we also wanted something that is not only focused on the player itself but more on the ecosystem existing in the NBA.
Working with the dataset, we decided to show the geographical aspect of the league. States, and thus franchises, may be linked in some ways, around a player or a trade for example. Visualizing trades which may have impacted drastically the future of the league is very interesting.

NBA is always changing, each season rookies make their first step, players retire and rosters change. Focusing on a player, we want to show his journey in the NBA and how he did or is doing during his career.

Before the implementation

The Dataset

Our first step in this project was to decide from which dataset to begin. As the context of this project is axed more on the visualization rather than data processing, we searched for a preprocessed dataset and containing data which are interesting themselves to visualize. We found two datasets on Kaggle: nba-players-data and nba-players-stats.

- nba-players-data : containing all the information and statistics about the players who played between 1996 and 2019.

This dataset contains following statistics in addition to the personal information of each player: Points, Rebounds, Assists, Offensive/Defensive Rebounds Percentage, Usage Percentage, True Shooting Percentage and etc.

Despite that this dataset covers till the recent year 2019, it lacks former data, which is why we decided to select a second dataset.

- nba-players-stats : separated in three files: Players.csv, Player_data.csv and Seasons_Stats.csv.

First and second one contains players' personal information and basic information as a player such as in which team he was being part of: these two datasets are merged into one, to avoid that the data is dispersed. The last one contains following statistics in addition to all the statistics listed above: 2-Point Field Goals, 3-Point Field Goals, Turnover Percentage, Block Percentage, Steal Percentage, and etc.

In addition to these two main datasets, by necessity during the project we added the following data: teamState.csv. It's a data of the 30 teams of the NBA. It consists of the team name, team abbreviation, the home State of the team, and the abbreviation of that state.

Initial Idea

Our initial idea was to show the dominance of a player depending on his body, height and weight, and show the possible correlations of with the statistics of the game. We thought about a visualization of a human body that we can play with to make it taller, smaller, bigger and thinner to sort out the player around the same shape.

Then with the feedback of our TA and with more technical researches on this part, we decided to make this part as a possible extra idea of our project as it's a creative but challenging idea.

We also thought about an interactive map of the United States showing the performance of teams with a color code as a choropleth map. When clicking on a state, we could see which opponents were the easiest or toughest. But then we would need to make a model defining how much a game between two teams is tough.

For this second idea, beside the showing toughest opponent part, as it's not suited for the context of this course, we kept this idea of an interactive choropleth map. Moreover, as our first initial idea has now become an extra idea, we decided to elaborate more on this map and additionally add a new visualization.

Improved Idea

For the performance visualization, we decided to use a Sankey Diagram. It was presented as a flow diagram in which the width of the arrows is proportional to the flow rate. The advantage of this diagram is that we can have multiple attributes and can observe the dependency and the flow between these attributes.

For our interactive map, in short our idea is to have a single map where we can choose which visualization to use, a connection map for the franchises a player played for, a bubble map denoting the performances in the same franchises and a choropleth to show the number of players per states. It could be a way to visualize trades, but also decisions that had a huge impact in the history of the NBA.

Project’s root : The Visualization

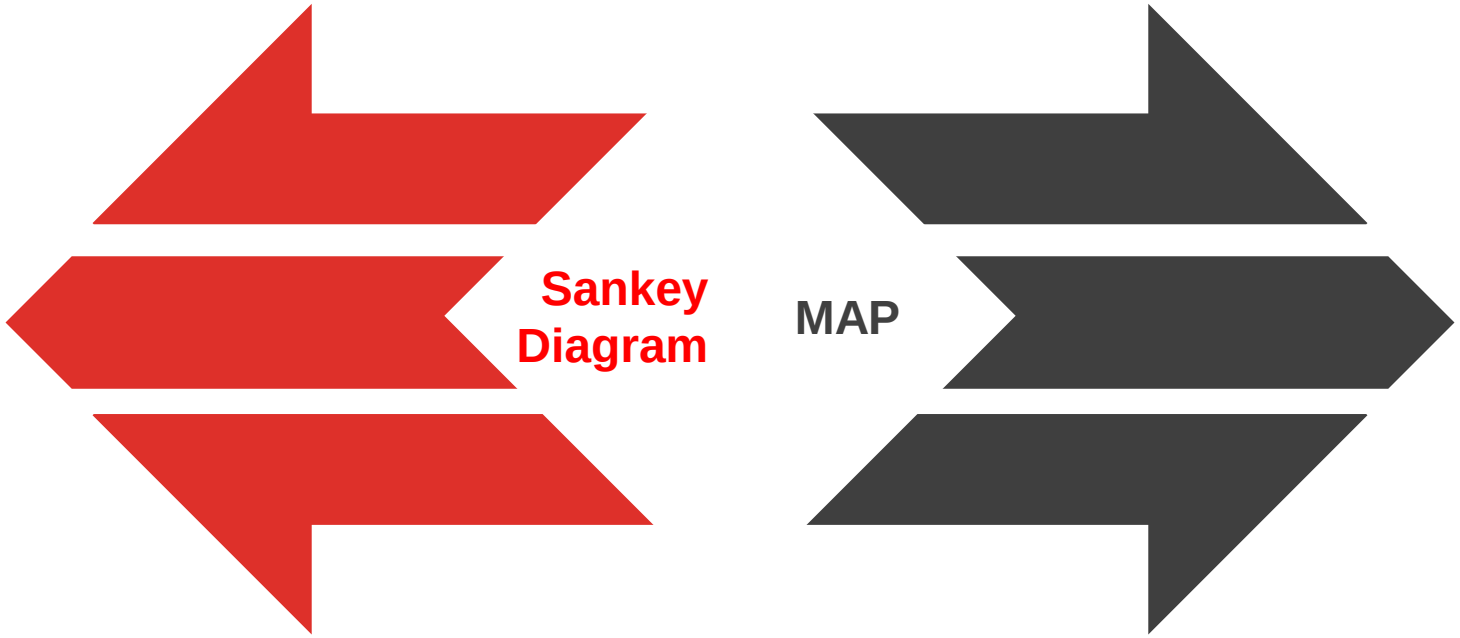
Final Plan

For the Sankey diagram, we let the users choose which statistics they’d like to see. To add to the diagram, we first choose them in the select box, a team in which the player belongs to and the year of the season then press the search button and all the players of the team of that precise season appear as a list of check boxes. Then we can select the player(s) that we’re interested in and click make Sankey !

For the Map, by default, we see the choropleth map showing the number of players per state. Then using the select box, situated just above the map, we can choose a player among the players checked in the check box on the left side of the map. For the chosen player, we visualize the state that the player has been played using a connection map and the performance that he had in each state using the bubble map.

The Sankey diagram was the best option for us in order to use our dataset efficiently and to show performances of players, alone or compared to team members. It can get a bit messy when there is a lot of players selected but in general it is a great visualization in order to show statistics and the weighted participation of players in different categories, like points, assists, rebounds,…

It shows the insides of a team, the role of each teammate, from the franchise player (arguably the best of the team) to a player on the bench who is sent on the floor for specific game plan or capacities useful to lock down in defense a player from the opposite team.



The MAP is a very versatile visualization tool. It declines in many different way and this is why we decided to use it.

A choropleth map, very basic, to show the number of players who played in a certain state/franchise. (based on the overall data).

A mix of bubble map and connection in order to show where a selected player played, connecting the franchises on the map with links, and how well he did there with a bubble sized with the ratio of his overall performance.

It’s a great tool to visualize an important side of the NBA, being a national league. It implies a lot more than just 2 teams playing a game of basketball.

Implementation process

01
Implementation

To implement Sankey Diagram, we used d3's Sankey diagram. The default layout of our diagram was with Player1, Player2 and with statistics free throws and points. We placed the diagram on the left side of the page letting the right side for the select boxes and check boxes, and mainly for the map.

02
Data Preprocessing

The Sankey diagram needs the statistic data based on the player. And the player can be selected among the team members of a specific season. So, we divided the data file into small numerous files based on the team and the year of the season. We wanted to do as much as possible the preprocessing so that the interaction with the visualization can be briefly done, as the size of the file to read and filter becomes significantly smaller.

03
Interactivity

We provide a way to choose players and statistics that they'd like to view. The former is done by adding the interface for the user to select team and the season through a select box. We added at the right of the Sankey diagram two select boxes: one for the team and the season. The latter is done by adding a checkbox so that the user can check all the statistics they need.

04
Design

We improved the design of our Sankey diagram with the select box's design. We wanted a more sophisticated design for our select box. We decided to exploit the advantage of the wide library of JavaScript. We used a jQuery plugin.

The implementation of the map has been done with d3's DataMaps library. The map is composed of three maps: choropleth map, connection map and bubble map. We first implemented the choropleth showing the number of players per state. Then we implemented the bubbles, using d3's bubbles, and the arcs, using d3's arcs.

For the choropleth, we calculated beforehand the number of players per state. Then for bubble and connection map, the data to show on the map limits to one player. So we split the original data file to small pieces of file corresponding to distinct player as we have done for the data processing part of the Sankey diagram.

The interact with our map is simple. First for the choropleth, the needed map is already present at the first sight on the page. For the connection and bubble map, it suffices to select one player among the players checked in the left side of the map. Once you have selected a player in the select box situated above the map, the map draws the arc between the states in which the player has been playing and finally draw the bubble proportionally to the performance of the player he showed in each state.

For the visual aspect of the choropleth and the bubble map, we decided to only use 8 different hues to limit too much diversity of colors as it can diverge the attention of the user. The choice of the color has been done using the course material. For the connection map, we needed as color of the connection, a flashy color: it's why we went for the hot red, which is harmonious with the purple of the bubbles and of the choropleth.

Challenges faced

Technical

Nothing will always go as expected. We had to readapt our ideas in order to make them feasible as we did from the body visualization to the Sankey diagram. We still have a way to show performances but now related to statistics instead of “body related variables”.
Having an image of the product in mind is great, but we had to not be too rigid about it. We had little knowledge prior the project in JS in general and it forced us to be flexible in how and what we wanted to work and show.
An important point, related to the NBA itself, is the fact that some teams changed name over the years (the actual New-Orleans Pelicans were the Hornets until 2002 for example), this implied for us to inform ourselves about all the changes that may have happened during the years.

Data

The data part actually turned in a real challenge. We had a lot of information to process and it happened that some values were missing for some teams. It forced us to make a choice between offering less stats options or less players options and we decided to the second one. This implied that we won’t be able to select players from the Phoenix Suns for example. We arguably decided this because it seems more interesting to us and because it’ll be easier to add data afterwards in that way than in the other. On the webpage, if you try to select a player not present in our data, a message appears and says that we currently don’t have the information for this player.

Implementation

The Sankey diagram was pretty straightforward to implement once we had the data processed and the rights basis to work with. The main challenge was the MAP, where we faced multiple problems. Our basis was a map of the USA but we totally missed the fact that the NBA has one Canadian team based in Toronto. We decided not to drop it but reworking the entire map was too much work at this point of the project. Making the visualization such that Toronto exists but is not “visible” (that we don’t see the area and borders of it on the map). It has been done before and it’s not harming what we want to express with the visualization thus seemed to be a great idea.
Merging the 3 maps was a mess, implying a lot of debugging. We couldn’t work all together on the maps directly due to the fact that 2 of us couldn’t pull from the repository, thus both of us worked on our side and then sent our code to the last member. This is definitely not optimal but we didn’t find any way to solve this and we still wanted to progress on the project.
Putting both visualizations on the same page implied for us that the selected players for the Sankey would be the same that can be shown on the map. As it wasn’t how we originally thought about the website arrangement, we had to restructure our code to avoid any serious bugs.

Design decisions

Sankey diagram

Boxes to select the stats we want to see/compare on the Sankey Diagram.
2 scrolling menus to select the year and the team, then boxes to select the players/players seems the most ergonomic way to go through the selection. Also, it allows to see way more potential players than by searching directly by name and eases to compare performances between players of the same team.

Interactivity

Map

Selecting the players by the same way as the Sankey is great to not have repetitive tasks to use our tool.
Once the players are selected, it updates the scrolling menu of the map and offers to select one the player previously chosen.
The initial state of the map is the choropleth one, since it's based on the whole dataset.
It is explained on the website and directly visualize something for the user before he interacts with the website !

Contemplative

Visualizations

If it happens that something requested by the user is missing, we handle this by showing a message that it may be available in the future.
Our data isn't totally complete but still allows a lot of possibilities so this is not something that would impact negatively the use of our visualizations.
We also make sure that everything is clear and explained to the user in order to not create frustration.

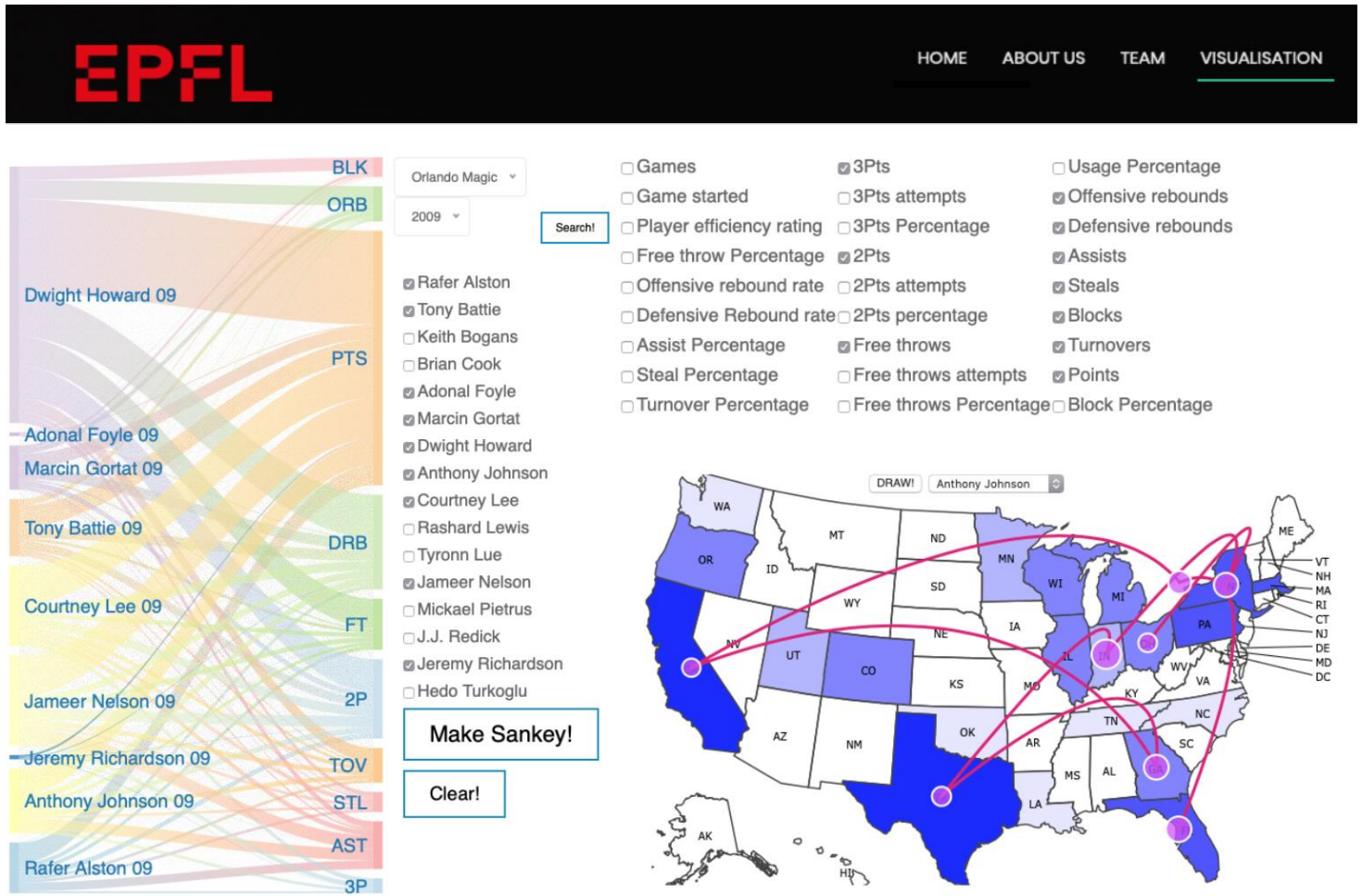
Ease for interpretation

Website

Putting both visualizations on the same page as suggested by the assistants.
Use a really basic template for the website that is minimalist, allowing the visualizations to be the main sections of the page.
Also, a small description of the tool and the data is defined at the top.

Simplicity

Final product



Two in One

Before, you had to move page between the Sankey diagram and the map. Now, you can see all our visualizations at one glance!

Sankey Diagram

You can analyze a player's different statistics and performances and compare them with other players!

Choropleth Map

You can now see the number of players per state with different color code. This is the basis state of the map!

Connection&Bubble Map

You can follow the trace of the player! A player can be selected and see how his performance changes through different teams!

Peer assessment

Maxime Fellrath

- Implemented alone the Sankey diagram visualization
- Implemented the final implementation of the map merging the 3 maps visualization in the website
- Managed the webpage and github repo
- Contributed to the process book with sketches and detailed explanations of problem faced
- Debugged the final product

Yura Tak

Worked on the data processing for the Sankey diagram -
Implemented the connection map visualization -
Worked on the choropleth map visualization -
Wrote half of the process book -
Searched all resources needed for our implementations -

Xavier Jeanmonod

- Worked on the data processing for the map
- Implemented the bubble and choropleth map visualization
- Recorded the screencast
- Designed the process book and wrote half of it
- Debugged the final product