**The World’s Game Process Book**

Michael Linnebach and Brady Smith

Initial Project Proposal

Basic Info

Project Title: Soccer Trends

Brady Smith [u0903295@utah.edu](mailto:u0903295@utah.edu) u0903295

Shawn Linnebach [michael\_linnebach08@yahoo.com](mailto:michael_linnebach08@yahoo.com) u1053546

https://github.com/BradySmith1019/dataviscourse-pr-soccertrends

Background and Motivation

When we started brainstorming ideas both of us decided that we wanted to do some type of sports visualization since we are familiar with how sports visualizations look from watching games. We then made the decision to design a visualization around something soccer related since it was a sport that we both interested in and familiar with. Next, we had to decide what information about the sport would make for a good visualization and because soccer is considered “the world’s game” we thought it would be best to make a visualization about its biggest event, the world cup.

Project Objectives

Are there any trends in soccer matches/scores/player stats/world cup stats over the last 90 years? How has the sport grown? We would like to create an all-in-one visualization for all the world cups since they started, containing match stats from all the games in each world cup, individual player stats, attendance, and any other relevant information. Overall, we want to make a fun and interactive visualization that will allow users to immerse themselves in the data and learn everything they wanted to know about the world cup. Overall, one of the main things we want our visualization to do is to highlight the worldwide nature of the game and how people all over the world play it.

Data

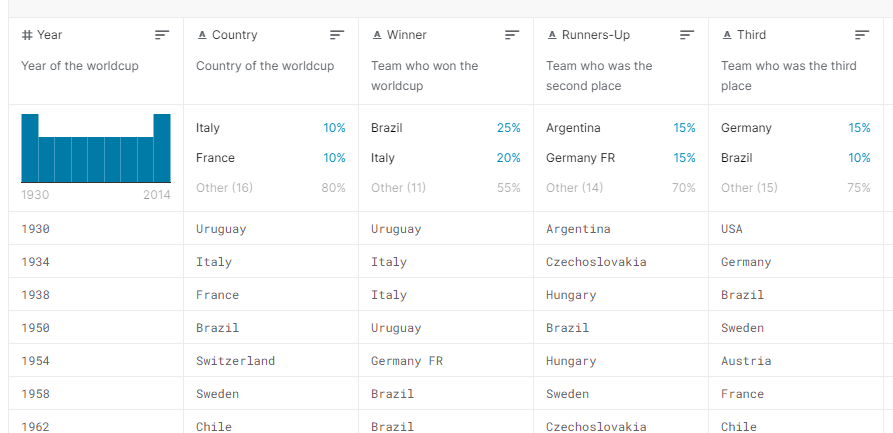
We searched the internet for historic soccer data and found a comprehensive dataset containing various data points for each world cup going back to 1930. Source: <https://www.kaggle.com/abecklas/fifa-world-cup>. The format the data is in is shown in figure 1 and 2 below. We feel like this dataset will provide all necessary information needed to create our visualization. It contains the match information of every game played at a world cup, details host cities, winners, and also provides some individual player statistics that we will be able to use for our visualization. The only other information we may need is world json data to draw the map of the world if we decide to incorporate a world map into our visualization and this information will be easy to obtain because we already have it stored in homework 4 for this class.

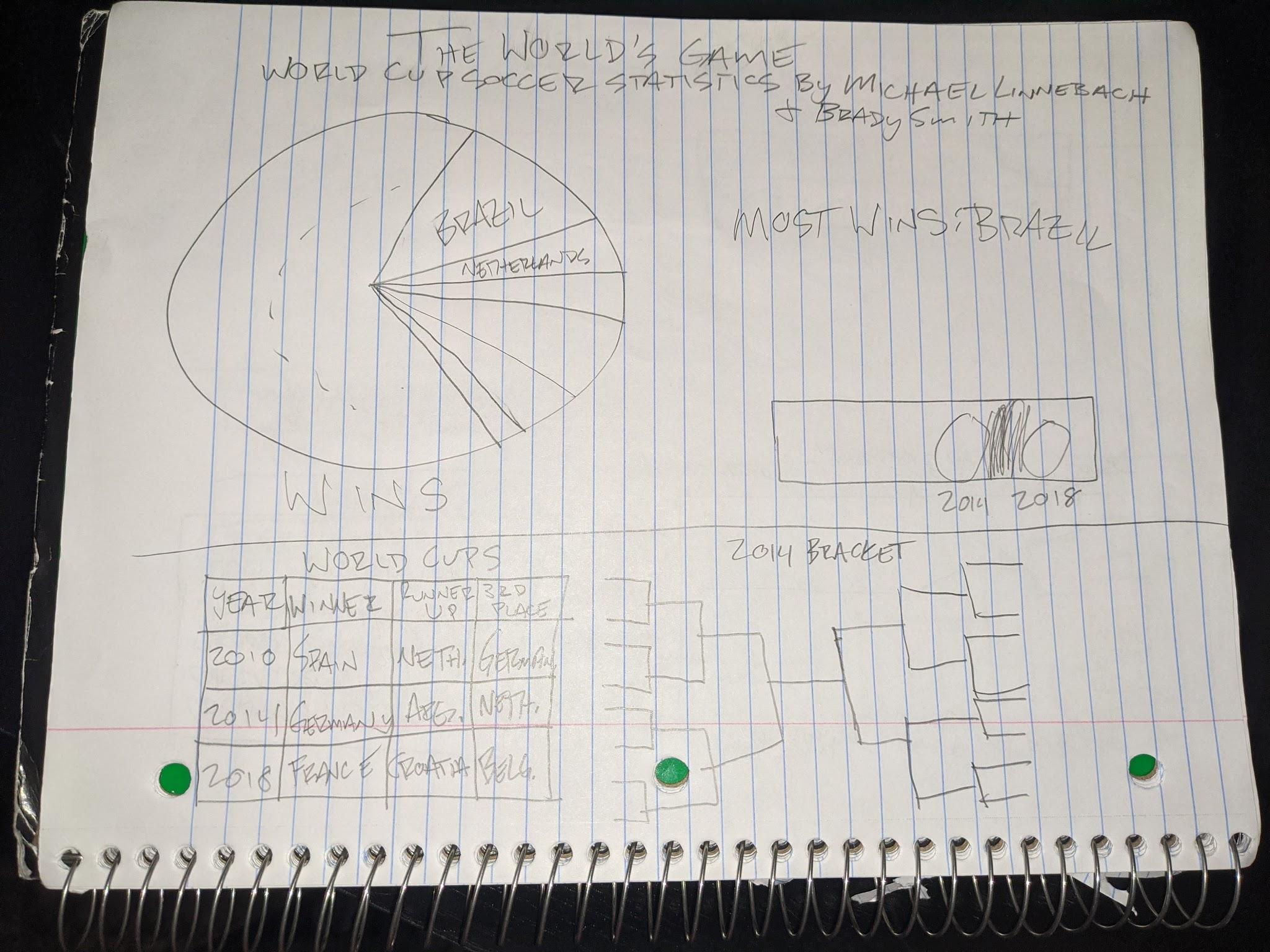
Fig. 1 - World Cup Data Entries

Fig. 2 - World cup match data entries

Data Processing

We do not anticipate having to do substantial data cleanup. The data is split up among three different csv files and each match is connected throughout the files by a shared id, so there will be a little bit of joining and sorting, but not a large amount. We plan to use as much of the data as possible to create as comprehensive a visualization as we can. Data processing will be implemented using javascript as a part of the data wrangling process and we may decide to loop through the game data at the start of our visualization to create even more data points that we could use for our visualization. For example, we might loop through every game of a specific year to generate the number of goals scored per game or we may loop through all data to determine a country's overall record for every world cup.

Visualization Design Prototypes

Fig. 3, 4, and 5 are all prototype visualization designs which helped us arrive at our final visualization design presented in the following section.

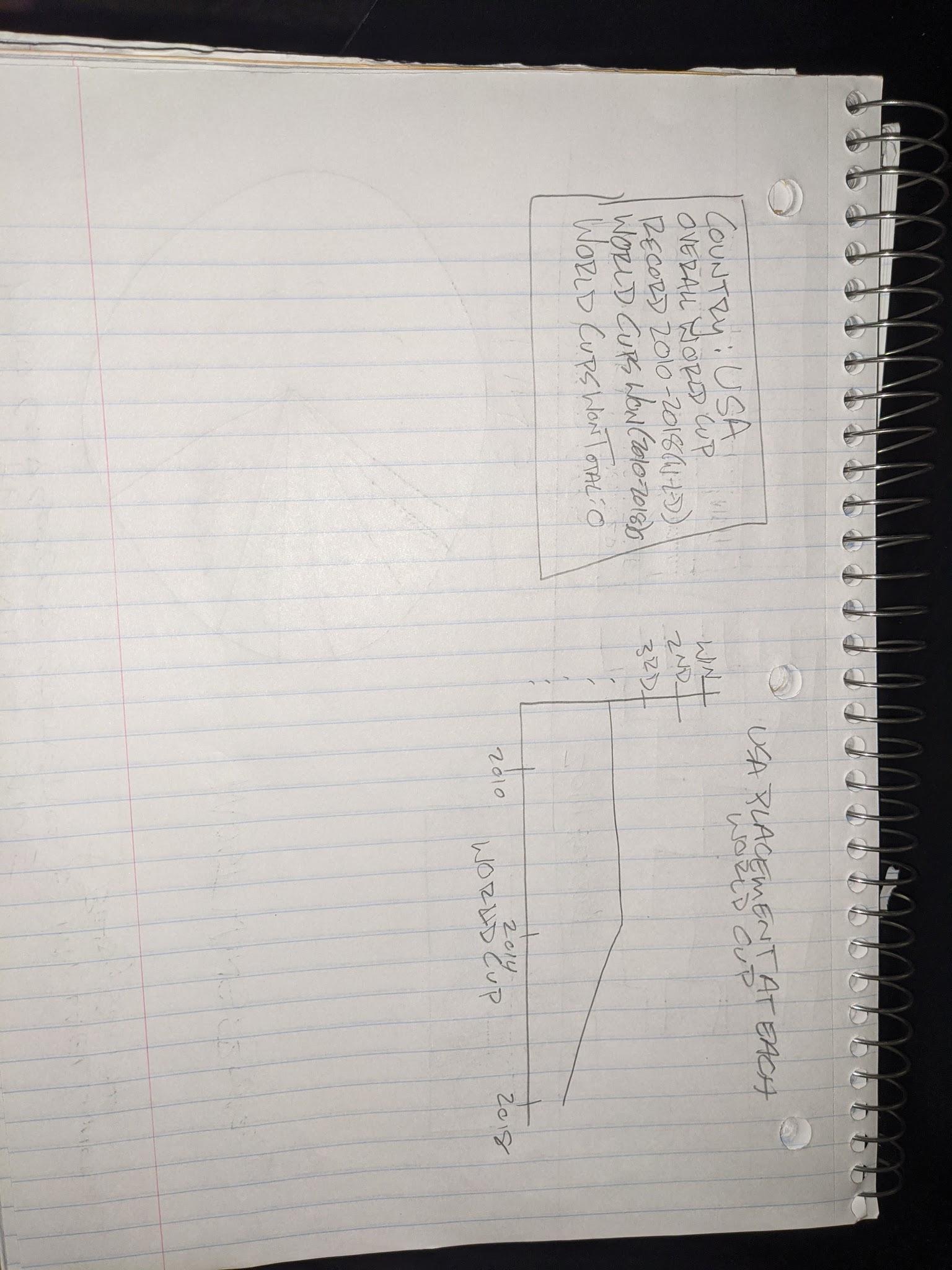


Fig. 3 - Prototype Design #1

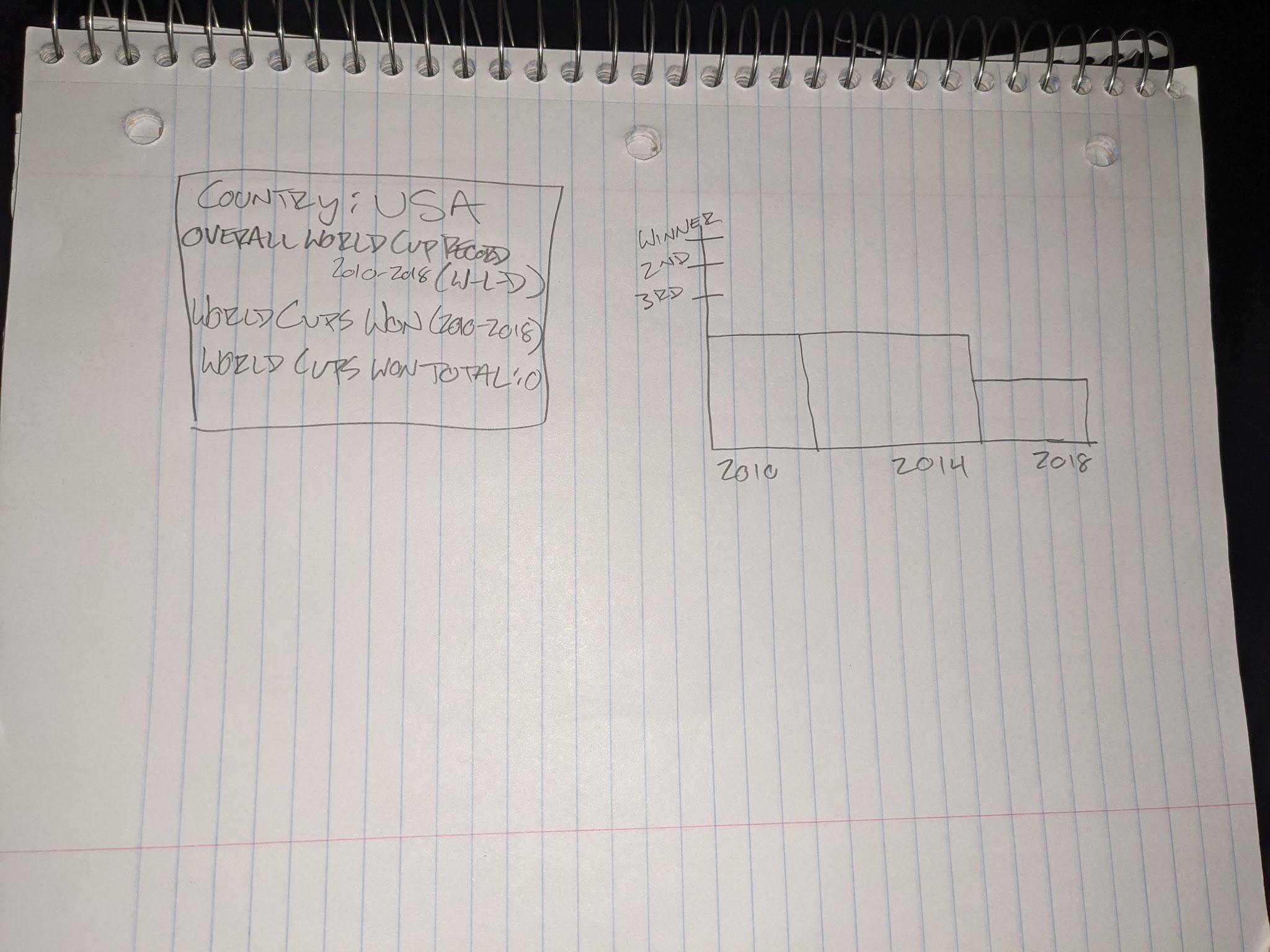
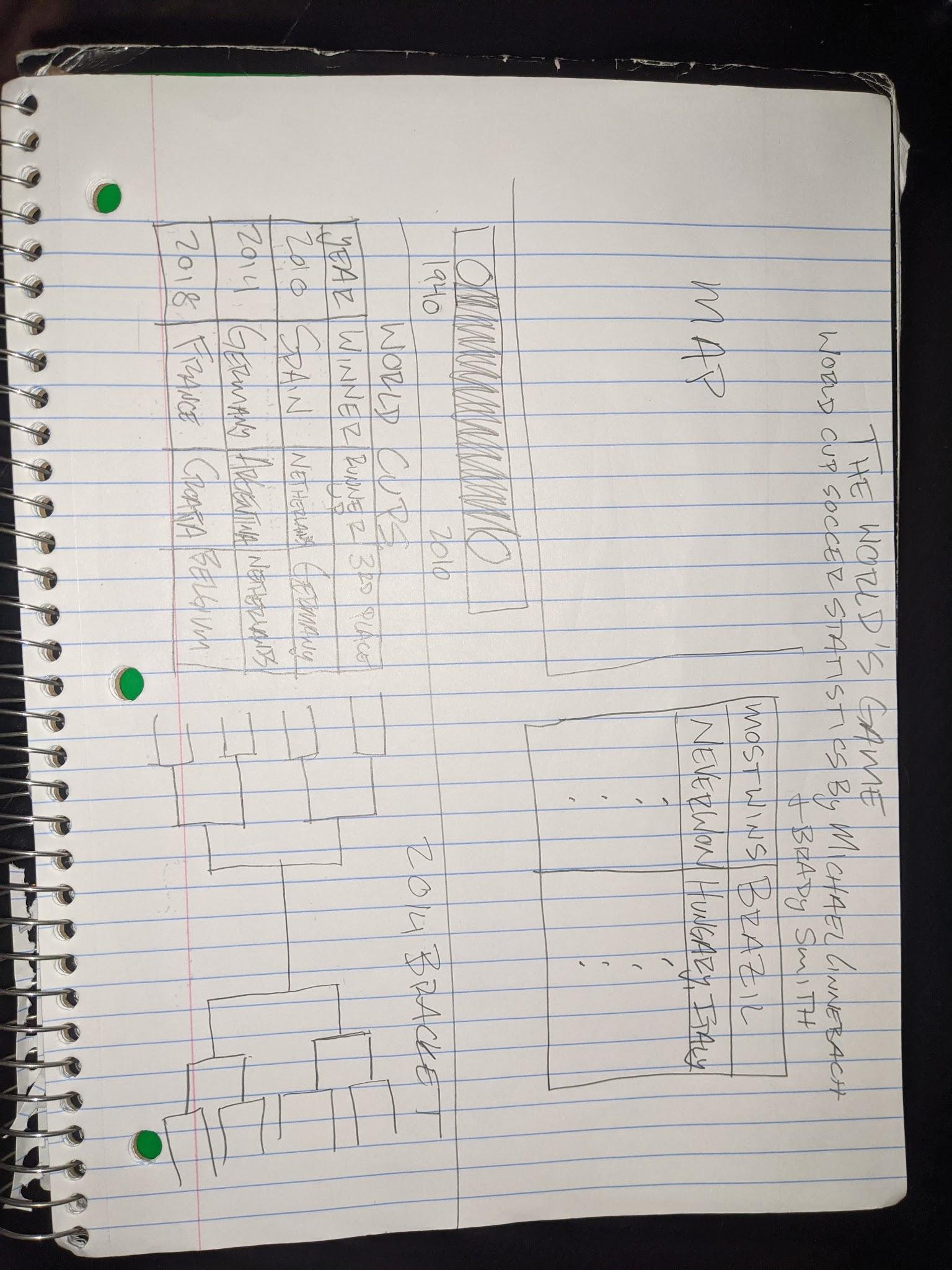
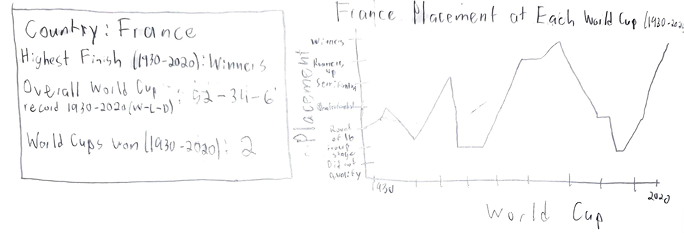
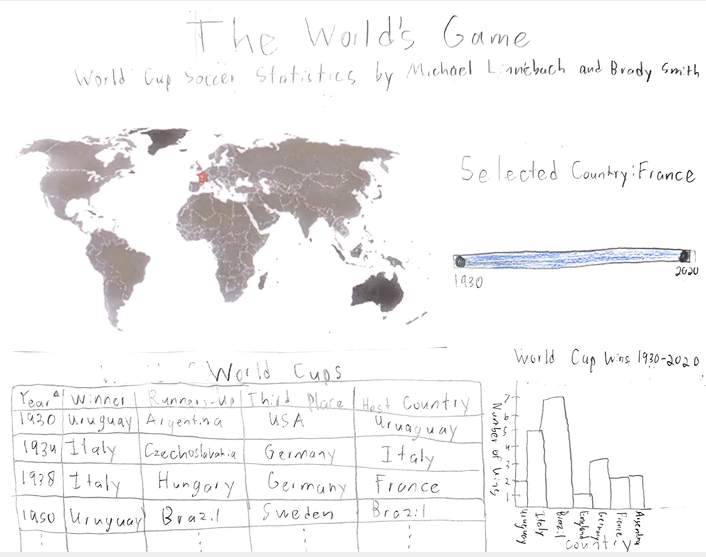
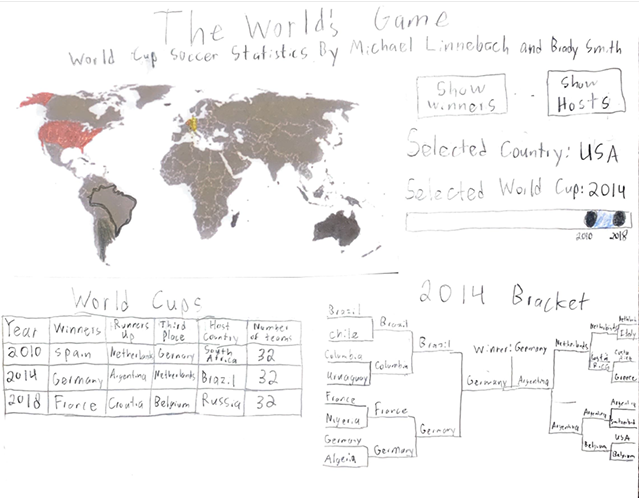
Fig. 4 - Prototype Design #2

Fig. 5 - Prototype Design #3

Visualization Design

The data will be displayed across multiple linked views. For the opening view we are planning on displaying a world map with every country in the world displayed on it and a year slider to select a range of years to select world cups in that year range. A table would show all the general data points for the world cups that occurred in the year range selected such as the year, the host country, the winner, runner up, the 3rd place team, and the number of teams that participated. Clinking a table header will sort the table based on that column. Clicking an entry in the table would cause that world cup year to be selected and the views would be updated to show specific data for that world cup. On the map, the winning country, the runners up, and the country that took third would all be filed in with different colors to make them stand out on the map. Additionally, we may also make it so that the host country and every country that participated in that world cup may be filled in with a different color as well. However, we are not sure how to resolve the duplicates since the same country that hosted would have participated as well so we might use another encoding such as making the border thicker to designate the host country to avoid this. Furthermore, another idea we have for when a specific year is selected is that we may be able to use an svg to build the entire knockout round bracket for that world cup. It would consist of lines and text and would show exactly how the tournament progressed at each round in the same way you would view any type of knockout tournament bracket. Then, by hovering over a team in the table it will bring up a small toolbox showing all the games that team played at the world cup. Furthermore, by clicking on a country in the map that country will be selected and individual data for that country will be shown, a selected country may be filled in with a different color. For the country and infobox will be brought up showing the country’s name, what continent it is located in, their overall record in the world cup that, their highest finish in that time frame, and how many world cups they have won in the time frame, and how many world cups they have won total. Similarly a line chart will be shown which shows the last stage that the country got to at each world cup before being eliminated in the currently selected time frame. Then, if we have time we may also incorporate adding another table for the country showing their top goalscorer at each world cup in the selected time frame. Then, just as an added note, upon startup the most recent world cup and most recent world cup winners will be selected as the initial world cup selected and the initial country selected and the slider for the year selector will be at 1930-2018 (the first world cup to the last world cup) and there were be two boxes one to display the world cup selected (its year) and the other to display the country selected (its name). Additionally we may add two storytelling buttons one when clicked would use the map to fill in all countries that have won the world cup and in an infobox would display the information about how many countries and what countries have won the world cup before. Additionally it would mention the country that has won the most world cups. Would use the map to fill in all countries that have hosted the world cup and in an infobox would display the information about how many countries have hosted and which countries they are and may highlight the fact that in 2010 South Africa became the first African nation to host and in 2002 South Korea became the first Asian nation to host. This whole design is shown on the following page in Fig. 6. We have decided to use a map to encode the main data this way because a map is easy to look at and gather information and it is visually pleasing. The year slider was a fairly straight-forward decision in that it is the simplest and easiest way to choose a range of years. The choice of a bracket to show all the games for a particular world cup was due to the fact that we want the visualization to be familiar and engaging for the user, and a bracket was the best and most informative way to do that.



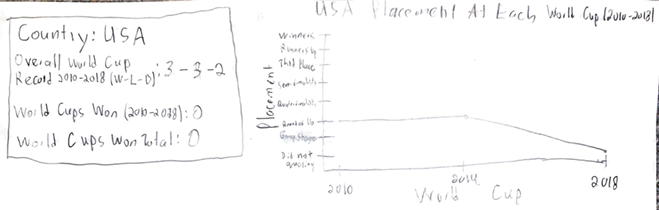


Fig. 6 - Final Visualization Design

Must-Have Features

* Main view - the world map
* Year slider - To select a range of years
* Table showing general data points for currently selected World Cups
* Bracket - if we can make it look good
* The ability to select a world cup by clicking on a table row
* The ability to select a country by clicking it on the map
* Filling in countries on the map for the selected world cup
* The infobox showing general data for the selected country
* The line chart to show the selected countries finishes for world cups in the selected time frame

Optional Features

* Specific player information for the selected country
* The storytelling button for the countries that have won the world cup
* The storytelling button for the countries that have hosted the world cup.
* A bar chart showing the goals per game for each world cup in the selected time frame.
* Hover feature on the world cup bracket to show in infobox about the games that team played.
* Display a country’s flag next to their name on the bracket or when the country is selected.

Project Schedule

* Week 1: Data acquisition, data joining and wrangling, draw map, year slider, table.
* Week 2: Add in the ability to select years with the slider and cause the table to change, add in the ability to select a world cup by clicking a table row, make the map update accordingly for the selected world cup.
* Week 3: Add in the ability to select a country and have the infobox and line chart display properly.
* Week 4: Draw the bracket for the selected world cup, add in the hover feature on the text of the bracket for each team if there’s time and we don’t run into too many issues.

Peer Review Session Feedback

Feedback:

* The World Map is very effective, another thing that could be added to it is that all countries that have ever qualified could be shaded in a certain.
* The bracket is really good the way it is drawn and probably the coolest and most unique part of the visualization. Although it seems like it would be hard to implement.
* The infographic for the selected country is cool and another thing that would be cool to display is the number of goals that country has scored at world cups.
* The line chart seems like the best option for displaying the countries finish at all previous world cups.
* The two way slider to select a time frame seems like it would add a lot of complexity to the visualization code while not adding much to the visualization.
* Selecting a world cup year by clicking on a table row seems kind of clunky and not very user friendly.
* All encodings seem effective and color appears to be used well.
* The optional storytelling buttons feature is very interesting.

Modifications made based on Feedback:

* Based on the feedback we decided to get rid of the two-way slider to select a time frame for the visualization and now instead it is just a normal slider with one point to drag the same way it was done in homework 4.
* The one way slider will now be used to select the specific world cup year to focus on and draw the bracket for instead of selecting a table row.
* Other than this the visualization remained the same because it seemed like the other group really liked the way our design prototype looked.
* The storytelling buttons became the optional features we would implement first if we have time.