Group: Bradley Paliska (bp355) , Euna Lee (el565) , Karim Pareja (kgp33) , Sahiba Dhindsa (skd77)

FIFA World Cup 2022 Data Analysis

Description of Data:

We obtained our data from Kaggle (titled: FIFA World Cup 2022: Complete Dataset) and we did not combine it with any other dataset. This dataset contains the statistics for all of the matches in the FIFA 2022 World cup, with each row being a particular match. The relevant variables include the teams that played in each specific match (labeled *team 1* and *team 2*). Number of goals scored by each team, possession percentage (how often teams possessed the ball in a match), and assists are all variables included in the data. Attempt variables detail how many attempts on target were made for each team in the match. In addition yellow card, red card, and foul variables detail the how many yellow cards, red cards, and fouls each team received respectively. Our dataset comprised 88 variables, so it was crucial for us to pick a relevant and meaningful subset to represent in our visualizations. For our visualizations, we decided on a choropleth map and and a bar graph, using nominal data which contrasts the countries' total goals and attempts on target during the FIFA 2022 World Cup.

Design Overview and Rationale

Design-wise, we decided to create two visualizations: a choropleth map and bar graph.

Choropleth Map:
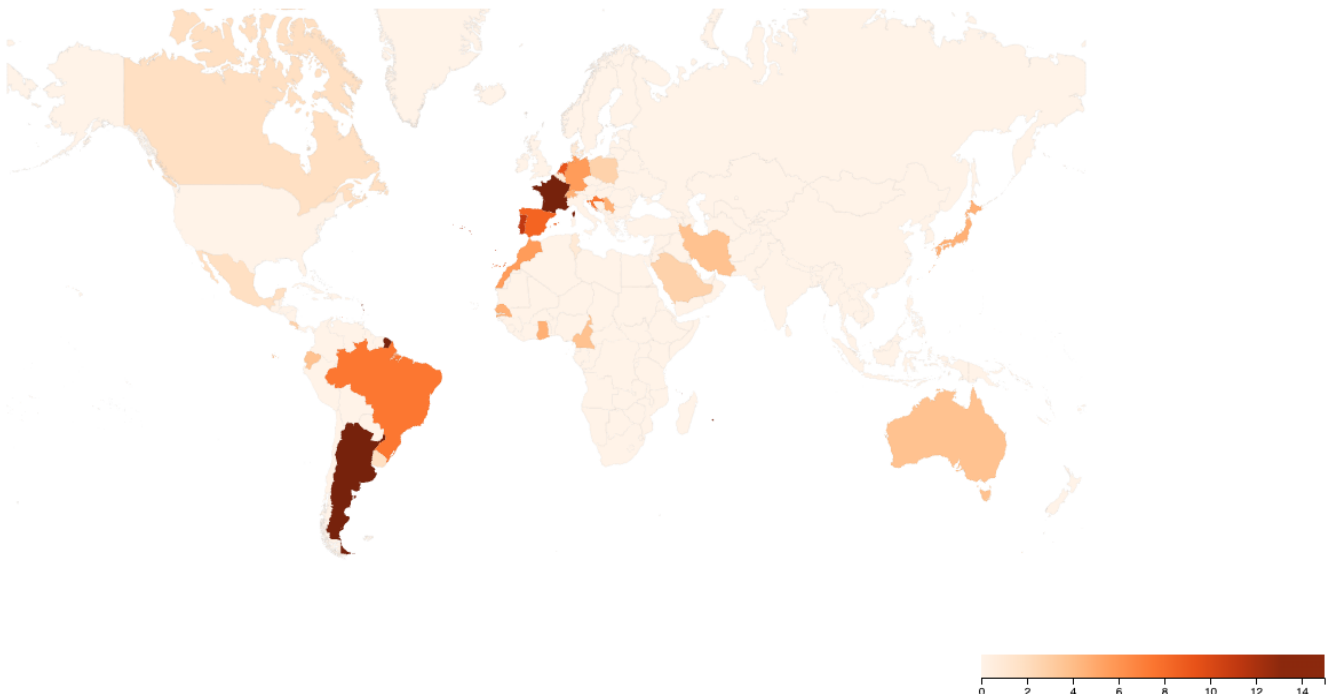
Marks: The marks in this visualization are represented by the countries since they constitute the data points.

Channels: The channels depict a change in saturation corresponding to the number of goals scored by each country. Additionally, the marks are aligned horizontally and vertically along a common axis defined by longitude and latitude.
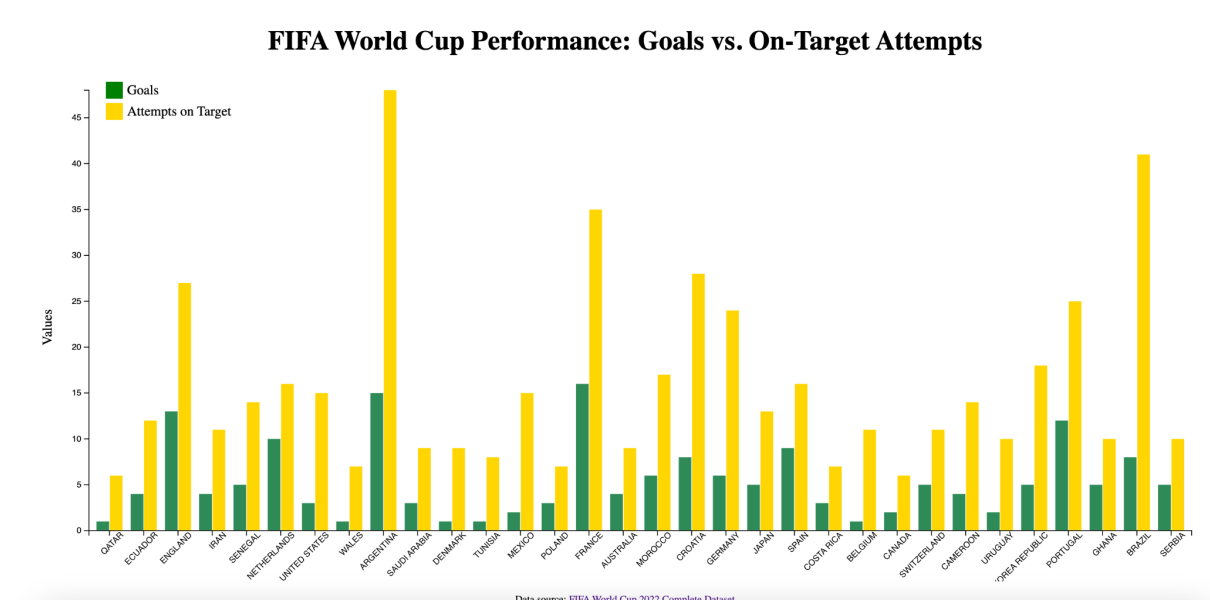
Visualization Summary and Analysis:
Choropleth graphs effectively display patterns in data spanning geographical regions. Due to the international scope of our data, we decided to utilize a choropleth map to illustrate the distribution of goals scored by country. The sequential color scale uses varying the saturation for every country to represent the number of goals they scored in the World Cup with dark orange/red colors indicating a higher number of goals. The most color-contrasting countries are located in South America and Europe.

From this visualization, we can discern that both European countries and certain countries in South America exhibit a notably higher number of goals scored. For the European nations, this observation may reflect the level of facilities they have at their disposal. Robust soccer programs from a young age coupled with top tier training programs can support the higher number of goals. Similarly, countries in South America have strong soccer traditions reflecting the elevated goal counts.

One trade-off in this design is the readability of the map. With our current visualization, it is unclear whether certain countries didn't participate in the World Cup or simply didn't score a goal. However, this tradeoff allows us to track patterns across the countries that have scored and present our findings in a clear and aesthetically pleasing manner. This choice provides the ability to obtain a summary of top-scoring countries at a glance.

Bar Graph:

FIFA World Cup Performance: Goals vs. On-Target Attempts

Data source: FIFA World Cup 2022 Complete Dataset

<u>Marks</u>: In this visualization the marks are bars representing the goals that each country has scored and their attempts on the target.

<u>Channels</u>: The channels display a shift in hues based on whether the bar represents the number of goals or attempts on target. Additionally, the positioning along a common x-scale corresponds to the bars associated with each country. A third channel is the length of the bars which depends on the data values for goals and attempts on target.

<u>Visualization Summary and Analysis:</u>
Bar graphs are incredibly useful for comparing different categories. Since we aimed to compare the categories of the number of goals and the number of attempts on target for each country that participated in the World Cup, we utilized a comparative bar graph. For the two categories, we used contrasting colors (yellow and green) to display a clear distinction between them, making it easier for viewers to differentiate and interpret the data. In addition, this color contrast allows the viewer to track patterns within each category.

From the visualization above, it can be ascertained that Argentina made the most attempts on targets but France made the highest number of goals. Brazil made the second most attempts on targets but made comparatively less goals than other countries. The top three countries who made the most attempts at scoring a goal were Argentina, Brazil, and France, but the top three countries with the most goals made were France, Argentina, and England. This does not seem to be a coincidence. Teams that advanced further into the World Cup would understandably have much higher attempts on target and goals scored. The high goal and attempt numbers for Argentina and France are evidenced by the fact that they played the World Cup final match. Likewise, Croatia's impressive figures can be attributed to their advancement to play in the semifinal match against Argentina.

Morocco is an intriguing case for analysis. Despite advancing to the semifinal match against France, they recorded lower numbers in both goals scored and attempts on target. For example, their goal and attempt figures were lower than those of Portugal and England, even though they progressed further in the tournament than these teams. One plausible explanation for this anomaly could be their higher goal-scoring accuracy compared to other countries. Alternatively, it is possible that they lucked out with the teams they played to get to the semifinals matches and their path to the semifinals involved matchups with teams that allowed for goal-scoring opportunities.

Overall, there seems to be a relationship between number of goals and number of attempts made as the more attempts you make, the more likely you are to score a goal. However, there are a few discernible countries that stray from this statement such as Brazil and Croatia. From this visualization, we can see that an increase in the number of attempts will most likely increase the number of goals scored for different countries.

The Story

The choropleth graph and the bar graph paired together interweave two distinct but complementary perspectives that display a comprehensive narrative of the FIFA 2022 World Cup. The sequential color scale of the choropleth graph helps paint a clear geographical picture of how countries fared in terms of goal scoring. The varying saturation highlights standout goal-scorers sparking questions of factors. This visualization ties in neatly with the bar graph which goes one step further to relate the number of goals to attempts made on target. The bar graph goes into detailed statistics for individual countries, showing the stark differences between countries and the impact number of attempts on goal has on total goals scored. It dives deeper than the holistic view the choropleth graph provides, making each country's statistics more distinctive.

When considering the audience for these visualizations, sports fanatics may prefer detailed statistics about individual countries while the general audience may prefer the choropleth map for a better summary. While the bar graph may not provide immediate data on which regions performed well or poorly, it allows users to consider specific countries. The choropleth graph has a strong visual impact with the varying saturations allowing the audience to identify regions with a high or low number of goals. Together, these two visualizations come together to fill in information one may lack, providing an efficient summary for the 2022 FIFA World Cup.

Contributions:

Bradley:
- Participated in a round table discussion to determine the project idea.
- Set up a GitHub repository for the group project.
- Gave ideas for the project's visualization.
- Started coding the project.

Euna:
- Attended all group meetings to plan and brainstorm projects.
- Attended office hours to inquire about our visualizations to the TA.
- Worked on the write-up for documentation.
- Presented project to class

Karim:
- Participated in a round table discussion to determine the project idea.
- Set up a When2Meet to find a time for all team members to meet.
- Set up GDrive and Milestone 1 document.
- Started coding the project.
- Presented Project

Sahiba:
- Attended all brainstorming meetings for project ideas and follow-up progress check-ins
- Wrote up a description of data, overview of design rationale for both visualizations and the story
- Set up and formatted Final Written Report document
- Attended office hours to clarify and discuss aspects of the visualization and write up with the TA

Time Breakdown:
- Week 1:
    - Organization (meet team, scheduling, drive, chats, …)
    - Brainstorming ideas for the dataset and potential visualizations
    - Initial visualization + Milestone 1
- Week 2:
    - Improve Visualization
    - Analysis and writing up findings from the visualizations
- Average of 7 hours a week