Project Milestone 2: Group 11

Jinghao Liu, Marwan Lloyd, Skylar Shafer, Max Zou

As our group is working with age data in sports, we felt it would be most beneficial to begin with a range of prototypes across different sporting areas. As such, our prototypes make use of data from real life Barclays Premier League statistics, EA Sport’s FIFA series, and a wide range of Olympic games. As will be discussed in greater below, each visualization took a different approach on what aspects of their data were most important and that deserved the most highlighting relevant to our end goals.

Beginning with the Premier League visualization ([Here](https://mlloyd05.shinyapps.io/bpl_vis/)), the prototype encompasses tools for filtering the data by team, player, and up to two desired metrics of analysis, and a user can see this data displayed both on a graph, and in an accompanying table. Courtesy of this wide range of selections and possibilities, this tool readily offers versatility for any user interested in Premier League analysis. We believe this functionality will be helpful in making our product broadly useful to sports analysts, especially when augmented by other dynamic query features included other prototypes such as mouseover on individual players in each scatterplot to reveal names/club. Additionally, for age related analysis, it offers multiple statistics that can be compared for players across age ranges, to see differences in performance over those gaps.

A second soccer related visualization ([Here](https://mlloyd05.shinyapps.io/fifa/)) that we used was based on the “FIFA” series. This prototype includes the ability to select league, position, and game year, and outputs a pair of plots of player overall rating and potential score. This visualization is a box plot of age vs rating. There’s also a jitter layer on the boxplot shows every player’s age and rating. A researcher could easily find how many young players had joined a given professional league and how many older players are still active in that same league. This data is valuable for evaluation as it is already standardized by EA’s own scoring system during game development (e.g. a player rated 80 in the Premier League is the same as a player rated 80 in the Bundesliga). As opposed to our other prototypes, this visualization does not focus on a breadth of topics, but instead narrows down on one area, the player development, very deeply by allowing analysis of a large number of players.

Beyond soccer data, our team also built prototypes centered around the Olympic Games ([Here](https://mlloyd05.shinyapps.io/olympics_sports_ages/)). The Olympics offer a great format for comparison, as they have a wide range of sports, with different physical requirements, and different athlete make-ups, all while consistently being ranked Gold, Silver and Bronze. As such, we are able to use medal results as a proxy for athletic performance, and see how the age-medal distribution differs across sports. To do this, this prototype plots the number of total medals won by athletes of each age at the Olympics for a given sport. It allows selection of multiple sports for easy comparison, and differentiates their points at each age layered over each other with color selection. Additionally, this prototype offers a second tab, where a user can check a given number of sports and the gender/age distribution of medals are given in bar graph form. This serves as an additional athletic intensity proxy, as certain sports (e.g. Hockey) have different rules surrounding physicality for males and females. Both tabs provide the ability to analyze a set of sports and see how the age breakdown for medals in a given sports. This gives us insight into the overall effects of aging on performance in different kinds of physical activity, and is a functionality that goes well with our research focus and has an intuitive design for users to subset the large dataset as desired.

Finally, we also implement our last prototype ([Here](https://mlloyd05.shinyapps.io/general_olympics/)) with the same Olympics dataset. However, this visualization is a graph that compares of age vs. count of athletes, across genders. A slider to select years and a selection box to pick types of sports are also available to query subsets of the data. This view is chosen because the number of athletes can act as an additional proxy for sporting performance, as it requires high level abilities just to qualify for the Olympic Games in the first place. As such, a visualization of age vs. count of Olympic athletes helps answer the question of peak performance of age across sports, by seeing when a given sport has its peak number of Olympic athletes.

At this point, we now understand that there are a few major tradeoffs between the functionalities of our potential final product. Firstly, we understand that scope is a large question, that is still in progress of being narrowed. Many of these sport datasets are very large, and thus the potential for a large scope is very high. That being said, the final product must be intuitive and must be able to produce quick utility for the user. As shown consistently across our prototypes, our intended method to address this is user-available filtering, allowing the user to easily decide what they want to see on their screen. A downside to this is that we must use a graphical style, like a scatterplot, that is useful and makes sense for almost all data categories, without being able to necessarily get more specific. Another potential weakness that we must address is in our searching system, if we are going to give such a large potential set of axes. The UI must promote searching for categories, so that a user can find the category they desire instantly, rather than having to scroll through all 40 or more categories that could be axes for analysis. An upside is that our product then offers answers to more of the questions are user is likely to have.

The opposite of this tradeoff can be seen in the second prototype. Only two graphs and their associated axes are produced, but they are both uniquely designed to be informative on their topic and display their information succinctly. They thus provide less information but the information they do provide is better presented. We believe that a final prototype of the highest utility/ease of use could be somewhere between the two. It could allow opportunities to change between a smaller subset of axes easily and have a more in-depth and particularly suited graphing style that provides more utility to the user, though we have to be careful about data size limitations. We could also use a multi-tab system, one that allows changing between players and their position and club, and another focuses on clubs as a whole and their statistics rather than each player, as an example. Then, we are focusing on two different yet closely associated questions and producing results of significant value in either, and the UI for each will be similar and thus more intuitive to switch between.

Overall, we feel positive about where we are in our prototyping and that we have found a number of different ways to structure our problem to create value. Looking forward, although we are not finished yet, we feel that we are on the right path to making a remarkable final project that could be useful for sports analysts and fans alike.

**Prototype Links and notes**

[Public preprocessing and visualization code](https://github.com/Mlloyd05/Fall2022_Stat679_Group11_Project/tree/main/Milestone2_Public_Code)

Premier League: <https://mlloyd05.shinyapps.io/bpl_vis/>

FIFA: <https://mlloyd05.shinyapps.io/fifa/>

Olympics 1: <https://mlloyd05.shinyapps.io/olympics_sports_ages/>

Olympics 2: <https://mlloyd05.shinyapps.io/general_olympics/>

**Prototype Notes**

For any prototype issues please contact Marwan Lloyd ([Marwan.Lloyd@wisc.edu](mailto:Marwan.Lloyd@wisc.edu) / [melloyd2@wisc.edu](mailto:melloyd2@wisc.edu) )

FIFA: If the app disconnects from server frequently, please reach out and Marwan can do an app reset. Due to limitations on data limits shinyapp.io’s free plan if too many users use the app at once space maximums can cause server issues

Olympics 1: On the initial screen (when everything is blank) you might see a red error message. You can ignore that and proceed with using the app as normal, once a dropdown option is picked, everything works as intended.