Project Milestone 2: Group 11

Jinghao Liu, Marwan Lloyd, Skylar Shafer, Max Zou

**Choice of Design Prototypes:** We approached this problem in a number of interesting ways that we felt offered different utility to the user. For starters, one of them focused particularly on the topic of changing the x and y axis of a comparatively simple scatterplot, using a dataset on the Barclays Premier League. Teams can be selected, as well as positions, and then you can select x and y variables out of a set of more than 40 variables and compare them against each other. In turn, the user gets an incredibly large number of comparisons they can make, and in turn can derive insights from many relationships incredibly easily, and switch between topics entirely in a moment’s notice. We believe this functionality will be good for making our product broadly useful to sports analysts, especially when included with a number of other dynamic query features that are showcased in our other prototypes such as mouseover on individual players in each scatterplot to reveal names/club as well as updating color-coding with a legend of players by position or club. We chose this functionality in that it offers a lot of options for the user to analyze different subsets of the data, and thus significantly more utility and adaptability than a model that does not accept this kind of dynamic query.

Another design prototype we utilized operated on an athlete event dataset at the Olympics. It would plot the number of total medals won by athletes of each age at the Olympics for a given sport, and multiple sports could be selected and their points at each age layered over each other with color coordination. In another tab, one could check a given number of sports and the gender/age distribution of medals are given in a horizontal bar graph, with the coloration of each gender’s side of the graph noted in the legend. Both tabs provide the ability to analyze a set of sports and see how age contributes to the probability of winning a medal in that set in either gender. This gives us insight into the overall effects of aging on performance in different kinds of physical activity. We chose this functionality because it goes well with our research focus and has an intuitive design that allows the easy creation of subsets of an otherwise large dataset.

Another visualization approach was also implemented for the Olympics dataset. This visualization is a line graph 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 is another indication of sport performance. By the nature of Olympics, among all population, only the athletes with top sport performance in their fields are selected to compete. Therefore, a visualization of age vs. count of Olympic athletes helps answer the question of peak performance of age across sports. Gender is also included for comparison of distribution of ages among different genders.

Another visualization was based on the game “FIFA” series. It would plot the overall rating and potential score at same time for each player played different positions in different league for every generation of game. 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 professional league and how many older players were still active in professional league and how they performed. It also shows when young players reach their potential. Researchers could also compare the age architecture in different league, different positions and different league. Since the rating is calculated by the most famous Sports games maker “EA sprots”, it’s largely convincing but still doubtful. The plot delivers as much as possible information in not bad looking.

**Demonstration and Implementation:**

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

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

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

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

**Critical Evaluation:**  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. 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 in one of our prototypes, the first mentioned, we can create a final product that incorporates changing axes depending on the user’s input. 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. 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. 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 great about where we are in our prototyping and have found a number of unique ways of structuring our problem that create value. We have looked at the problem in a number of ways and found a middle-ground between models that we believe will provide the most value to the user in an easy-to-understand user interface that is sleek and efficient. 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.