Project Milestone 1: Group 11

Jinghao Liu, Marwan Lloyd, Skylar Shafer, Max Zou

**Modality Decision:** For our project, our group has decided to work with aging data in sports and athletics, both at the professional level and recreational level. More specifically, we are looking to analyze the impact of aging on athletic performance, seeing why age impacts athletes differently, both within the same sport and across different sports, and see what factors might contribute to these differences. **(NOTE: CHANGE/UPDATE THIS IF WE DECIDE ON SOMETHING ELSE)** We believe that the design studio modality will be the better option for this analysis, as it can allow users to tailor the data they are looking at to sports and athlete groups that they are familiar with and better able to relate to, thus providing them with greater understanding of the data they are viewing.

When thinking about the dashboard that this data could produce, several key factors come to mind. These include, but are not limited to, options to switch between sport displayed, handling increased specificity if a user wants to narrow in on an age range, and may also include a table sub-section as well for easier diving into a specific professional player, if so desired. (**NOTE**: **Include screenshot of a mockup below here, probably will look something similar to the in class example that included a dropdown and a table (I believe right before the wildfire data)**)

**Motivation and Goals:** The possible uses for this data are widespread, with clients ranging from sports teams, athletic training groups, and individuals who are looking to stay athletic at an above average level into their 30’s, 40’s and beyond. Given the disparities between these groups, it will be a challenge to satisfy them all. However, doing this will require successful implementation of the customization tools described above. Most importantly will be having relatable data across a range of sports, as there is not large amount of insight that an aging golfer can gain from an aging soccer player. It may be a challenge to find the exact same data for every sport and so a normalization process will be required, most likely via the creation of a type of rubric, in order to provide meaningful comparisons on the factors that do stay similar across sports, like injuries. This normalization will allow for our display to change between sports without constant jarring visual changes impacting the user’s experience.

**Literature Review (NEED TO DEFINITELY FLESH OUT HERE MORE):** When reviewing previous statistical research into this subject, as well as tangential ones, a few themes of visualization became apparent. Firstly, the prevalence of numerical tables, it is very common to find a table with a subset of raw numbers accompanying the visualization. Typically they serve to provide further elaborations of the data (e.g. by including information like sample size, n, or % differences between points) and support the images by ensuring that this information is easily available without reducing the image’s legibility. (**NOTE: PROVIDE SCREENSHOTS HERE OF A FEW TABLE/GRAPH COMBINATIONS IN THE LITERATURE)**

**Packages:** One major benefit of working with sports data is its widespread prevalence which significant helps make the overall data collection process easier. There are many publicly available datasets on Kaggle which provide pre-packaged information, however, as with any pre-made dataset, the risk with using these is that we do not entirely know what the data collection process was. As such, in addition to these pre-packaged datasets, our group will be able to utilize several packages in R that have been designed to easily scrape and aggregate sports data. These include nflscrapR and cfbfastR, which are packages for aggregating and scraping both pro & college level football information; nbastatR, hoopR, and wehoop, for NBA and NCAA levels of basketball; baseballR for the MLB; engsoccerdata for top level European soccer; and hockeyR for the NHL. With these available to our team, we will be able to verify and supplement pre-packaged datasets as necessary through data collection work that we can ensure matches our desired standards.

**Conclusion: (PUT STUFF HERE AT SOME POINT)**

**Citations:** (Add in Citations as we add in charts/images/screenshots, then once done put it on its own page at the end)