Project Milestone 3: Group 11

Jinghao Liu, Marwan Lloyd, Skylar Shafer, Max Zou

At its core, our project looked to provide useful visualizations on the relationship between aging and athletic performance. This final goal is a drilled down version of our loftier beginning goal of rating athletes across sports as they aged, as that initial problem involved too many parts to visualize well. Thus, the results are a product of quality over quantity.

Sports analytics is a topic that a lot of work has been done in already, and we feel as though our dashboard provides a collection of efficient visualizations for an analyst looking to learn more on the dynamics of gender, age, and more across sports. The Olympics are the pinnacle of sport, and so the perfect dataset for us to use. Additionally, the Olympics offer an accepted standardization across different sports and events. All of them take place over the same period, in the same location, and finish with a Gold, Silver, and Bronze medalist, thus the format sets up several layers of standardization.

Our final dashboard drew inspiration from several ideas we found across literature of previous sports analytics projects. We found much scholarly research had been done on many of these sub-topics, such as aging in sport, for example. In turn, we felt we could best make our project useful by making it a toolbox for analysts looking to learn more on a given sport in the large amount the Olympics provide, in terms of the traits of the associated athletes, particularly gender and age, among others. Rather than go for a more niche piece of the research we had uncovered, we decided to produce a dashboard that would be interesting and easy to use for sports enthusiasts of all backgrounds.

In terms of the design itself, many of the literature reviews we saw included a table to go along with and elaborate on a graph, an example that we followed in one of our tabs as well.

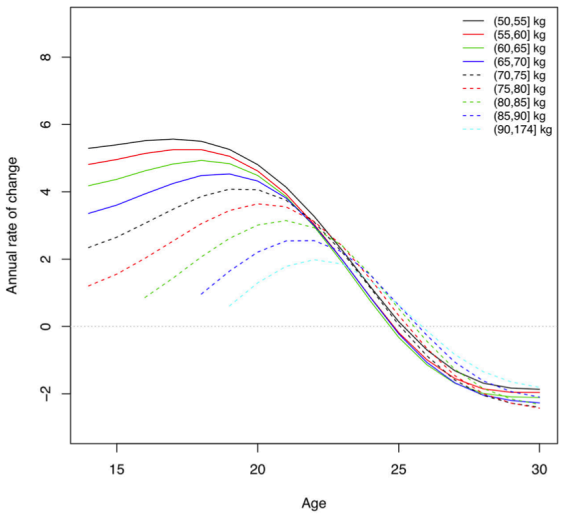


FIGURE 6 | Annual rate of change in performance levels, at 90th percentile, by age and body mass for women. Lines represent the rate of change in total weight lifted for athletes with different body mass. This corresponds to the derivative of the curves for performance levels, where zero indicates the age at which there was a maximum in the absolute performance level.

TABLE 2 | Performance development, at the 50th and 90th percentiles, for different body weights: age at maximal annual rate of increase and age at peak performance.

	Women				Men			
	Age at max increase		Age at peak performance		Age at max increase		Age at peak performance	
Body mass intervals (kg)*	**50th**	**90th**	**50th**	**90th**	**50th**	**90th**	**50th**	**90th**
(50, 55]	14.0	16.8	23.5	24.9				
(55, 60]	15.0	17.2	23.7	24.9	17.0	16.0	24.5	26.6
(60, 65]	16.0	17.6	24.1	24.7	17.4	16.0	24.1	26.6
(65, 70]	17.6	18.5	24.7	24.9	17.8	16.0	24.3	26.6
(70, 75]	19.9	19.5	25.5	25.1	18.2	16.0	24.7	26.6
(75, 80]	21.1	20.5	26.3	25.3	18.9	17.4	25.3	26.2
(80, 85]	21.7	21.1	27.0	25.3	19.9	17.6	26.2	25.7
(85, 90]	21.9	21.7	27.8	25.5	20.7	18.1	27.0	25.5
(90, 95]					21.1	18.7	28.0	25.5
(95, 100]					21.3	19.7	29.0	25.5
(100, 105]					21.3	20.5	29.0	25.5
W (90, 174] M (105, 174]	22.3	22.1	29.0	25.5	21.3	21.1	28.0	25.5

An example of a table elaborating on its relative visualization for weightlightfiting data (Huebner and Perperoglou)

Beyond that table, our choice of allowing multiple selected sports and genders to be displayed at once was also influenced by the literature we saw, as demonstrations with side by side comparisons proved to be the most effective overall.

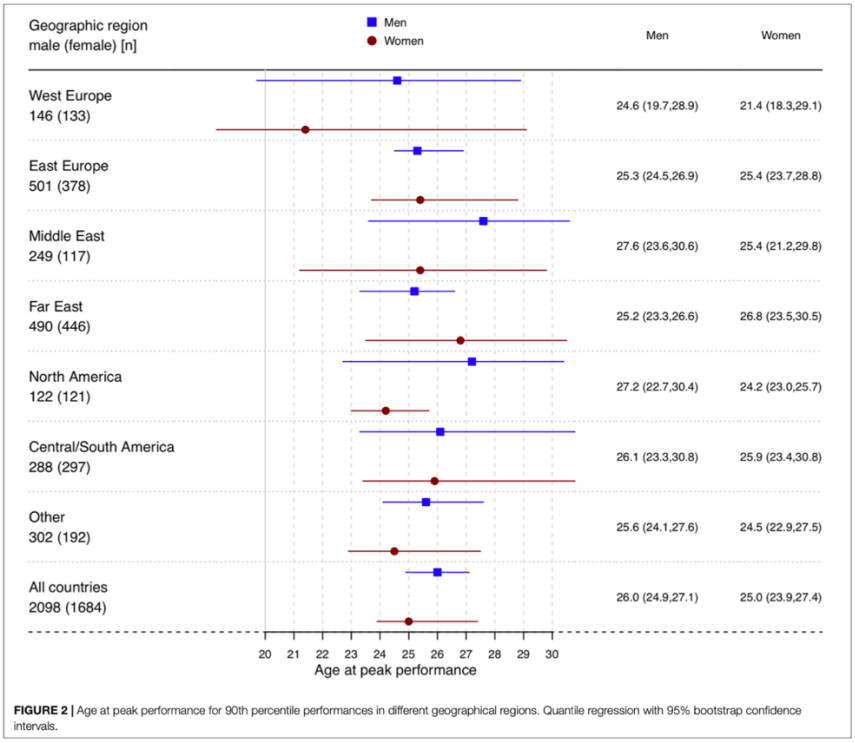


FIGURE 2 | Age at peak performance for 90th percentile performances in different geographical regions. Quantile regression with 95% bootstrap confidence intervals.

An example of a visualization comparing peak performance age in weightlifting across different geographic regions and genders (Huebner and Perperoglou)

Over the course of our design's lifecycle, we continually tried new ideas, discarding those that we feel did not bring optimal results. To begin, we worked with a few datasets to find multiple, in pursuit of an optimal solution. We experimented with datasets on FIFA player ratings, NBA players, as well as another on the traits of players in the English Premier League, before settling on a general Olympics dataset that allowed for filtering on the traits of historical athletes including the year of competition, where it was held, whether it was summer or winter games, as well as individual athlete characteristics like height, weight, age, and nationality.

After deciding on our dataset, we had a number of potential utilities that we wanted to pull out of the dataset. In particular, we initially decided we wanted to prioritize the effect of aging in sport on athletes of different genders. So, a few of our initial prototypes focused on this vision. We designed models that displayed medal count by age and gender filtered by sport, to

see what age range typically has the most success in each sport for each gender. We feel as though this is a valuable piece of insight, as you can then look deeper into sports with small age ranges and assess if this is due to the nature of the competition, for example, it being too physically strenuous for an athlete who participates in it to have a long career of doing so. From this, information can potentially be extracted on what kinds of exercise are best/worst for longevity, whether it be for an athlete training and hoping for a long career, or a regular person who wants to live a length and healthy life with as little of a reduction in athletic ability as possible.

Another topic that we later approached is what height and weight typically led to success for a given competition as the typical build of an optimal athlete in a sport changes over time, as the strategy of said sport evolves. We wanted to explore this topic as well as those of gender and age, and so another one of our prototypes created a scatterplot which provided information on the traits of competitors, as well as their gender, and whether they won medals, thus creating a scatterplot with extensive customizability. An interesting section of Shiny capabilities which we utilized, and did not know going into this project, was that you can use *aes_string* instead of *aes* for describing aesthetics, and this allows you to change these attributes based on reactive user input, as the user input comes in as strings. So, the user can then select axes to adjust the scatterplot by, out of height, weight, year of competition, and age of competitor, while learning of each competitor's medal success and gender via the color and shape attributes of a given point.

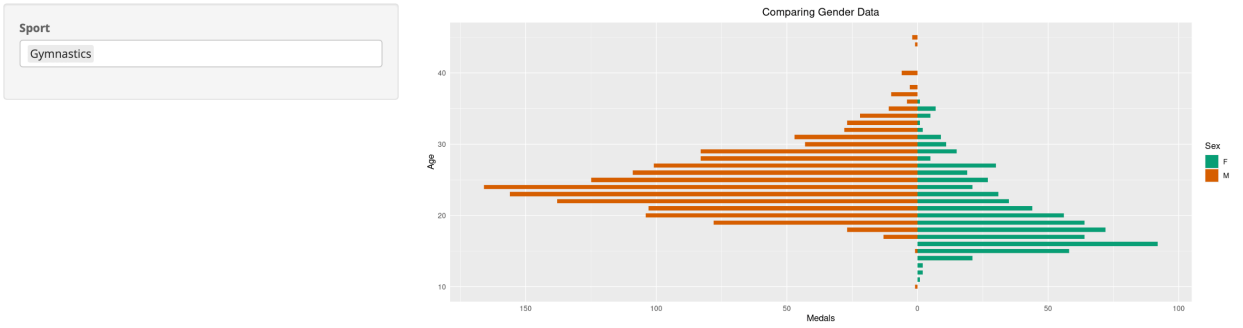
A final topic that we explored deeper was how the age distribution of competitors has changed over different intervals of competition. In doing so, we created another panel which gave the competitors at every age for a given sport, with an adjustable time interval in order to

see how the age distribution of athletes for sports by gender has changed in different time periods.

As our prototype period came to an end, we felt conflicted. We had initially decided on a reduced scope, focusing on age and gender in sport, but over time, we had come to produce many visualizations with their own utility for a user interested in sports analytics. In turn, we felt as though we best did our project justice by creating a dashboard including visualizations focused on multiple facets of these athletes and their traits, each focused on a different kind of insight that we felt was displayed efficiently and intuitively. Thus, we combined multiple prototypes, focusing them on the one dataset, and creating a tabbed dashboard that allowed the user to click between the different perspectives of the data offered.

When a user opens the dashboard, the first visualization they see is the trend between age in the Olympics and medals won by athletes at that age, broken down by selected spots. We chose this as the first visualization in order to get the point right away to a user, in case they have only minimal times. Olympic successes and failures are determined by medal counts, and so we felt that leading with this would be optimal for a user.

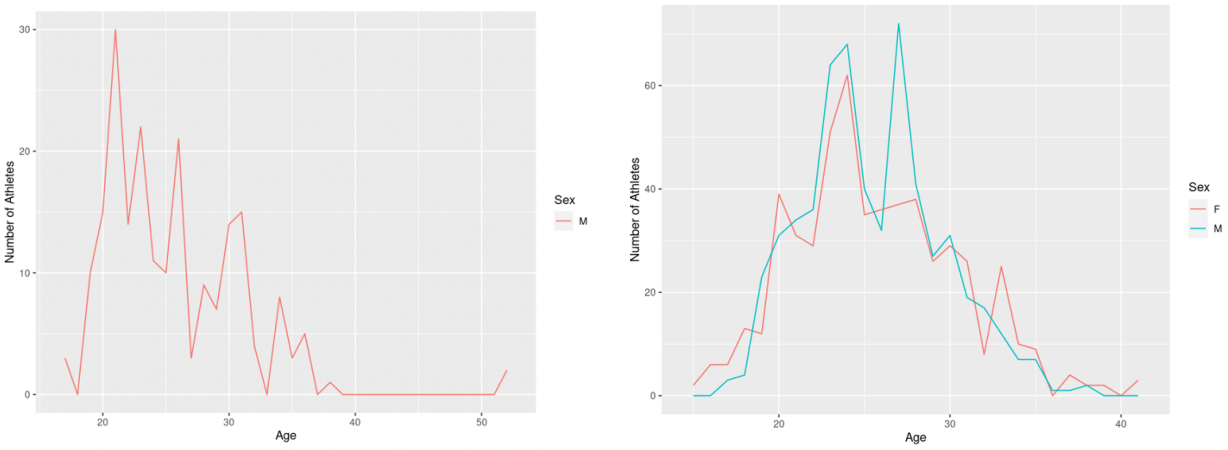
Following that headlining page, our next tab provides a similar medal breakdown by age and sport, but this time also including gender. We felt this was appropriate to include in case any users are researching differences in Male & Female sports at the Olympics or so that users can use the gender of athletes as a proxy for some things like intensity (e.g. Women's hockey doesn't allow body checking while Men's does). This graph is flipped 90 degrees compared to the first tab as the flip makes it easier to see between Male & Female numbers, as comparing the two in opposite vertical directions is not as intuitive for users.



Our next tab is a version of the Age vs. Medal visualization. It shows the age distribution of all athletes in the Olympics, for both genders, with option to choose among different sports and filter the years which these athletes participated. Because the Olympics is a already representation of the best athletes chosen from all over the world a simple count of number of participants is representative of the population of all athletes for the purpose of this research. The graph chosen is a simple line graph for a direct visualization of the distribution across all ages. The graph of genders is overlayed, it is included as the scope was increased and this addition feature shows the difference of distribution between genders.

It is interesting to compare the distribution of athletes between genders across different kind of sports. For some sports like gymnastics and swimming, the age distribution of female is clearly peaks earlier than the male graph, indicating that among the best athletes in the field, women gymnastics and swimmer reach their peak performance at a younger age compared to men. In fact, for most sports, female athletes have a lower age distribution. With the year slider input, the range of years of data to include can be adjusted. With a range set to 20 years and dragging the slider, the user can compare the change of athlete age distribution throughout the years. For example, in the case of water polo, female athletes started to participate only after the year 2000. Moreover, the change of age distribution can also be observed using this feature. When selecting “Speed Skating”, it is clear that the distribution of age changed from a peak

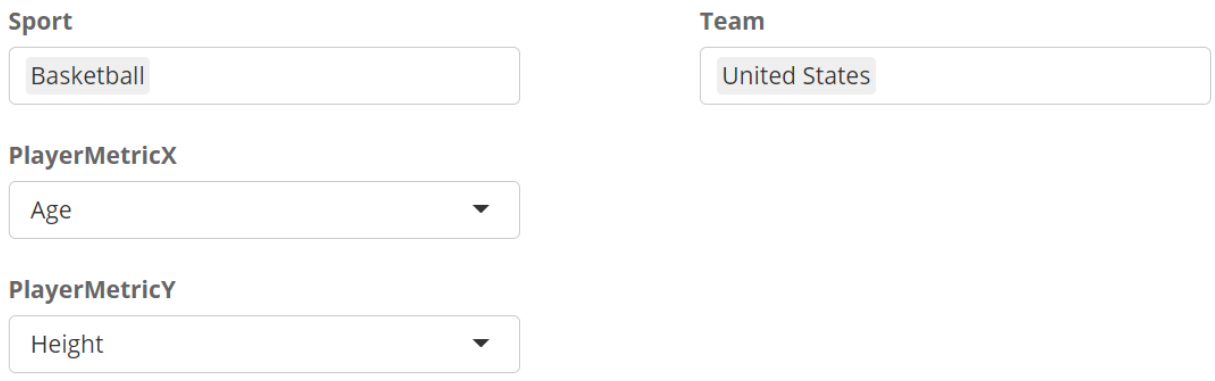
around 20 years old in the 1920s when they were first introduced to the Olympics to a peak of 25 years old nowadays.

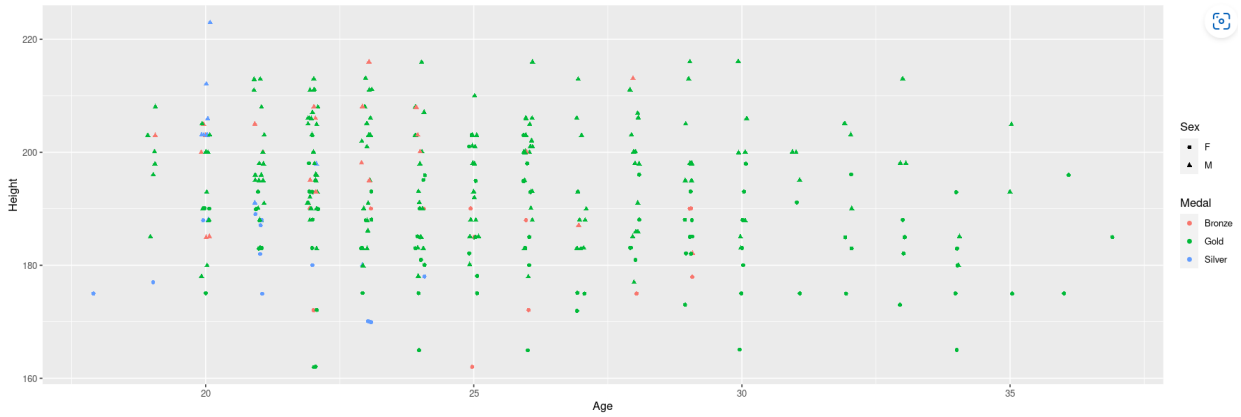


Speed Skating Athletes Age Distribution (Left: 1920-1930; Right: 2006-2016)

The groups consider this observation as a result on the medical and technology advances in the field, allowing athletes to perform the sport safer and with better support (coaching, dieting, health monitoring), which extended their careers, giving the more experienced athletes an advantage.

The fourth tab is the customizable scatterplot prototype aforementioned in the prototypes section, which looks like the following.

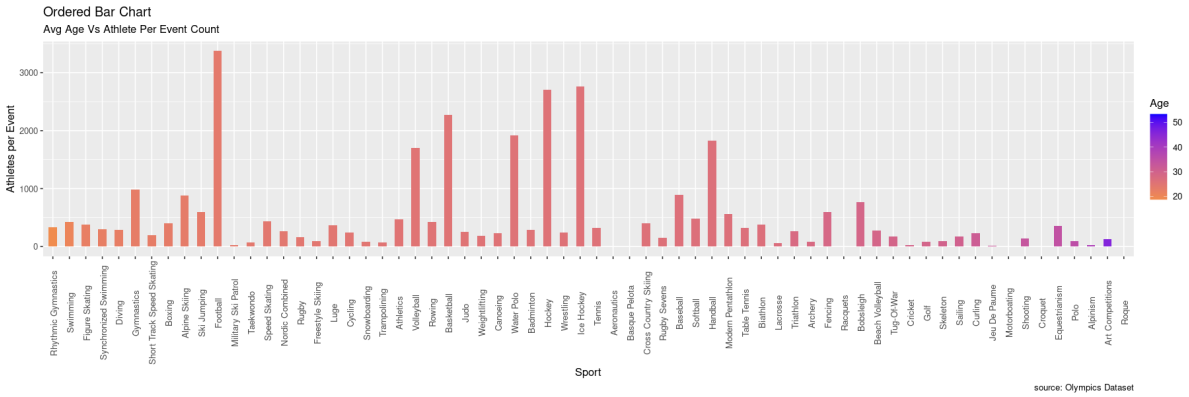




ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
351.00	Julius Shareef Abdur-Rahim	M	23.00	202.00	104.00	United States	USA	2000 Summer	2000.00	Summer	Sydney	Basketball	Basketball Men's Basketball	Gold
2636.00	Stephen Todd "Steve" Alford	M	19.00	185.00	74.00	United States	USA	1984 Summer	1984.00	Summer	Los Angeles	Basketball	Basketball Men's Basketball	Gold

The primary utility is its customizability, and the table at the bottom is a resource for further research on athletes within a given sport and competing for a given nation. You can thus locate particular athletes, getting more information on outliers or notable historical athletes of interest.

Finally, our last tab is new, not integrated from the prototype, and looks as follows.



We felt as though it would also be good to compare the sample size between events, and show the average age of these events. So, the vertical height of each bar is the number of competitors, and the color is the age, with the sports ordered along the x-axis by age, as well. You can thus see quite easily which sports are dominated by younger athletes, which are in the

middle, and so on, as well how large each event is, with football (soccer) as an example being the largest.

We acknowledge that we chose a topic that was broad and diverse, with many Olympic athletes competing across a range of unique sports. That being said, we feel as though our visualizations provide an efficient toolbox for analysts to gain insightful information on the traits of Olympic athletes with regard to year of competition, age, height, and weight, with a focus on differences in gender, as well.

Dashboard Link

https://mlloyd05.shinyapps.io/Combined_Prototypes/

Citations:

Huebner, Marianne and Perperoglou, Aris. "Performance Development From Youth to Senior and Age of Peak Performance in Olympic Weightlifting." *Frontiers*, Children's Exercise Physiology, 27 Aug. 2019, https://www.frontiersin.org/articles/10.3389/fphys.2019.01121/full