Micaila Marcelle & Jakob Garcia
DATA 363
Dr. Watkins & Dr. Mathias

Gender Disparities in Elite Cycling

Gender inequality has long been a pervasive issue within the realm of sports. Though major strides have been made towards improvement in the last few decades, when considering differences in monetary compensation, the size of competitive events, and the amount of external investment, it becomes clear that there is still a long way to go. The overall goal of our project is then to gain a deeper understanding of the extent of this problem within the particular context of elite cycling. By analyzing data associated with the men's and women's UCI World Tours, a set of races representing the highest level of competition for both groups in the realm of road cycling, we aim to further illuminate the gender disparities that currently exist, working to bring attention to the varying experiences of male and female athletes within this sport.

Relative to a number of other sports, road cycling has been somewhat slow in its movement towards equality between male and female riders. To put things into perspective, one of the biggest and most well-known events in road cycling, the Tour de France, first took place in 1903 for men (*Tour de France History | Continental Tires*, n.d), but the women's version of the race had its first official occurrence only as recently as 2022, ignoring previous failed attempts at organizing the event (Mallon, 2024). Such a reality points to the fact that female cyclists simply lack much of the infrastructure that has been built up around men's cycling. They frequently face fewer competitive opportunities, lower wages, fewer sponsors, and greater social barriers to entry than their male counterparts, all in addition to the common belief that female athletes just can't handle as much as male athletes (Colangelo et al., 2023). These factors then feed into a strong sense of inequality within the sport.

Our project then looks at some of the key data associated with men's and women's elite road cycling, specifically considering the races that overlap between the men's and women's UCI World Tours. We examine the differences in the men's and women's prize pools and individual winnings, along with the variation in the number of teams involved in each race and the total distance covered. The latter two pieces of information are also considered for different time points to see how these values have changed over the last few years. In examining this data, we aim to see how the actual structure of some of the most elite road races varies for men and

women, providing insight into the different competitive environments that male and female athletes face and considering the broader implications that these differences may have. We then seek to instigate further change within elite road cycling through this work.

Methods

As is mentioned above, for the purposes of this study, we restrict our scope to the elite road cycling races that overlap between the men's and women's UCI World Tours; this ensures that all of the data we examine within this project is sufficiently comparable and in line with the questions that we aim to answer. Furthermore, there are no known existing public databases containing all of the desired information on men's and women's elite road cycling. The researchers were thus required to manually collect data on the prize money, length, and number of participating teams (the latter two over time) for the races within our range of interest. This information was obtained from online sources, then compiled into a single data set, with the resulting spreadsheet being provided in the "Appendix" section of this paper.

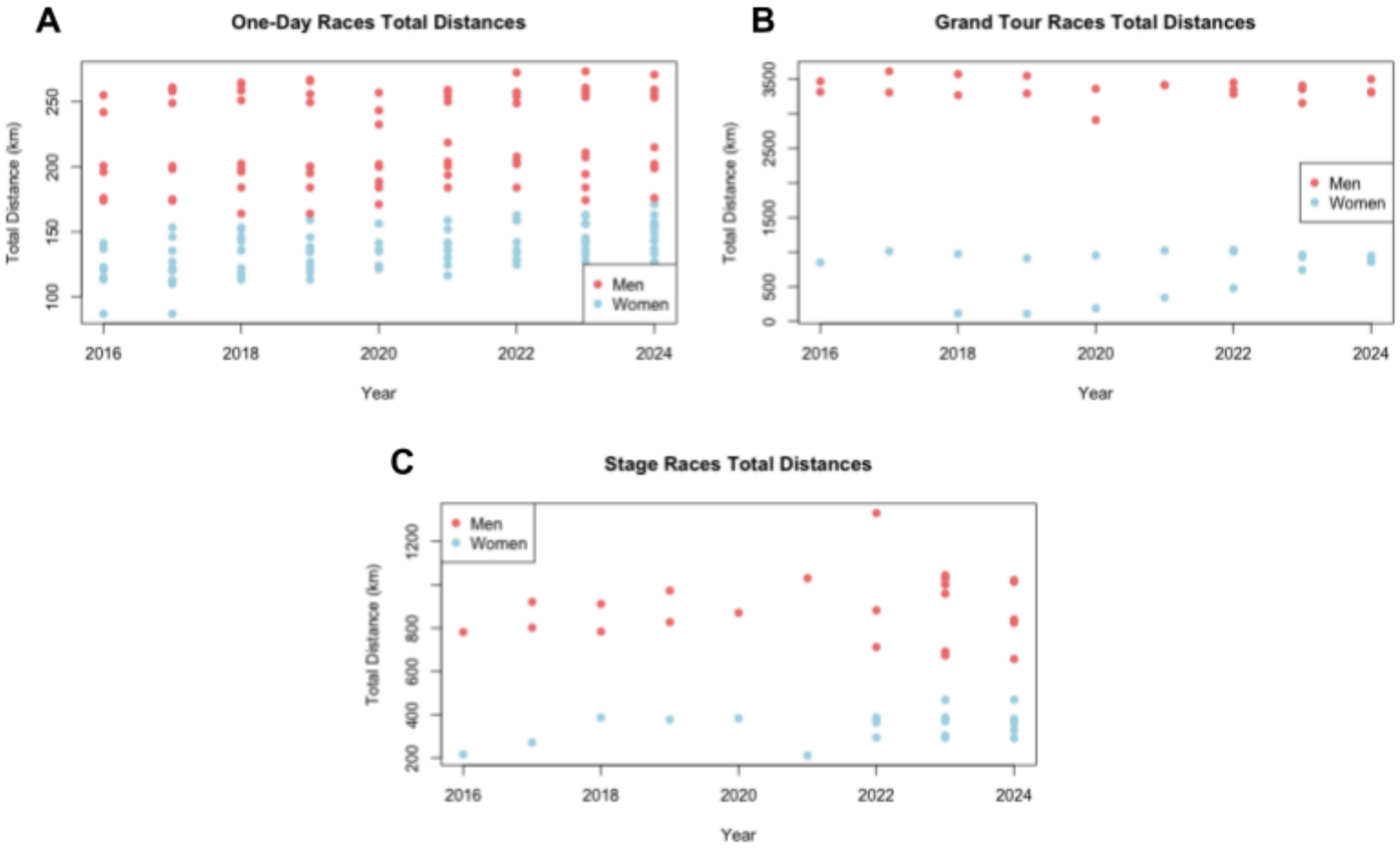
To actually collect all of the desired data, we first used the official sites for the men's and women's 2024 UCI World Tours to determine which races are included in both sets, ultimately finding 20 races to include in our study. Each race was additionally marked with its associated type (stage, grand tour, one-day), allowing the data to be further broken down into these categories during analysis. External websites were then used to collect information on the individual winner's prize money, the total prize pool, the total distance, and the total number of teams for both the men's and women's version of each race. The former two data points were only collected for 2024, due to the lack of availability of this information, while the latter two were collected for 2016 through 2024 (as available) in order to get an idea of how the total distance and number of teams have changed over time. The sources used for each data point are then included within the spreadsheet for greater clarity.

It's important to note, however, that our data set contains one major piece of bias. While information on prize money is readily available for all of the men's races in our data set, it is significantly more difficult to find for the women's races, and is frequently unavailable in the latter case. Furthermore, much of the data that was obtained on prize money for the women's races was a result of knowledge of equality in prize pools for the men's and women's events. As a result, the information that we do have on the prize money for women's races is heavily biased

towards equality with that for the men's races, which is highly unlikely to be an accurate reflection of reality. This key source of bias is important to consider when interpreting the results of our research project regarding prize money, and will be further addressed when discussing the validity of conclusions resulting from our data analysis. We also note that many of the women's races within our data set only came into existence quite recently, which means that we have slightly more time series data for the years towards the end of our 2016-2022 range, and less for those towards the beginning. This is then considered when analyzing how the total distance and the number of teams have changed over time, since it may impact the accuracy of our analysis for earlier years. With such biases kept in mind, the outcomes of our research can then be more effectively understood.

Data Summaries & Visualization

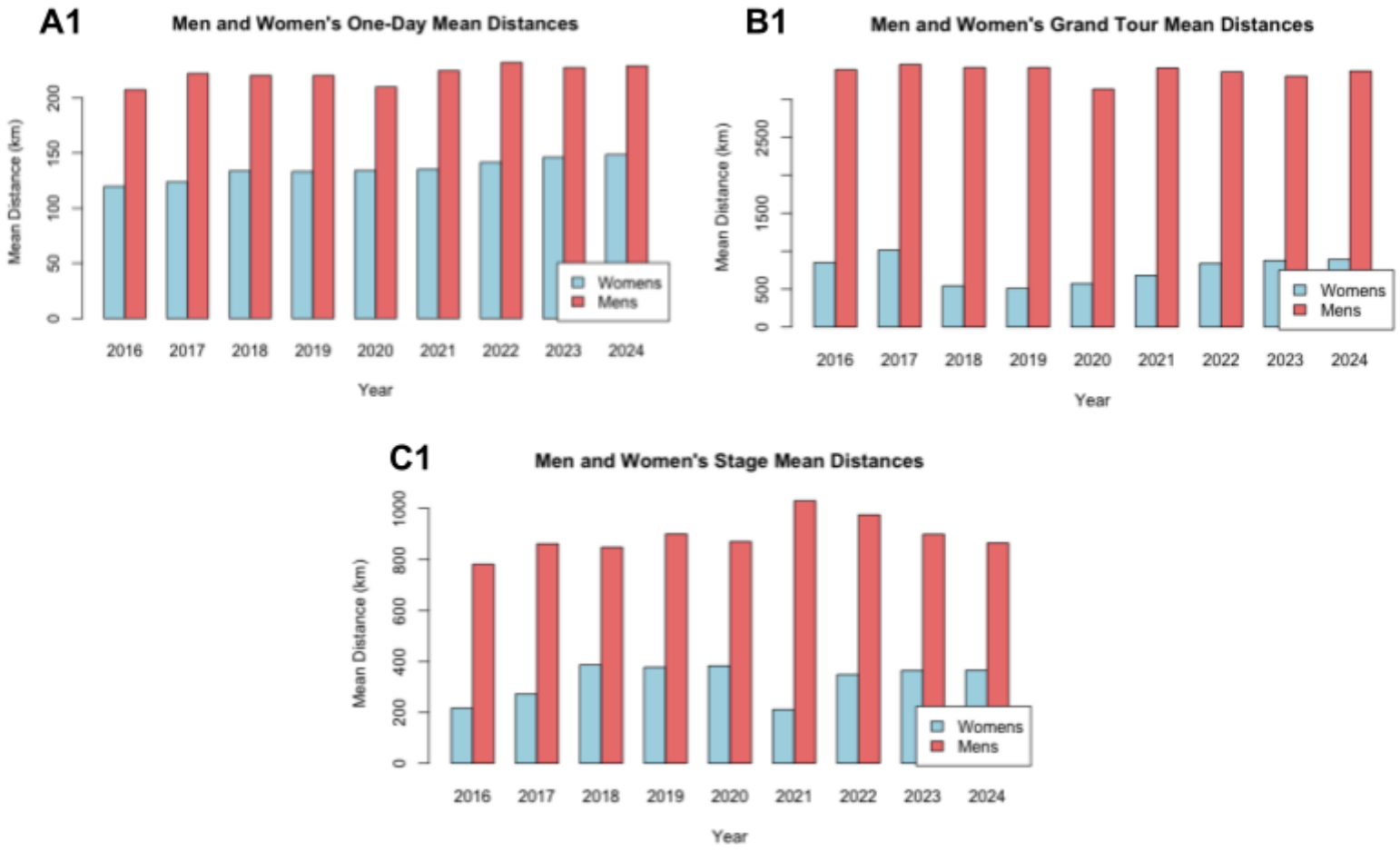
Using the data collected via the methods discussed above, which are provided within the spreadsheet found in the "Appendix" section, we constructed several figures in R to visually summarize this data on elite road cycling. This R code can be found within a GitHub repository linked in the "Appendix" section. We then first consider how the distance of races, in kilometers, has changed for both the men's and women's overlapping UCI World Tour races from 2016 to 2024, examining this for the three primary types of races considered within our research project. This leads to the figures below:



Figures A, B, C: For each year from 2016 to 2024, the data points for the total distance (in kilometers) of the men's (red) and women's (blue) overlapping UCI World Tour races are plotted within a quasi-scatterplot, with A corresponding to the one-day races, B corresponding to the Grand Tour races, and C corresponding to the stage races.

We can observe a sort of "split" within each of the plots above. The men's races in each category appear to consistently be longer than the women's races, with this divide being relatively small for the one-day races, somewhat larger for the stage races, and substantial for the Grand Tours. Notice as well that there does not appear to be a clear trend upwards or downwards in our data points with respect to time, indicating that in all three race categories, and for both men and women, the total distances of races have remained approximately constant from 2016 through 2024.

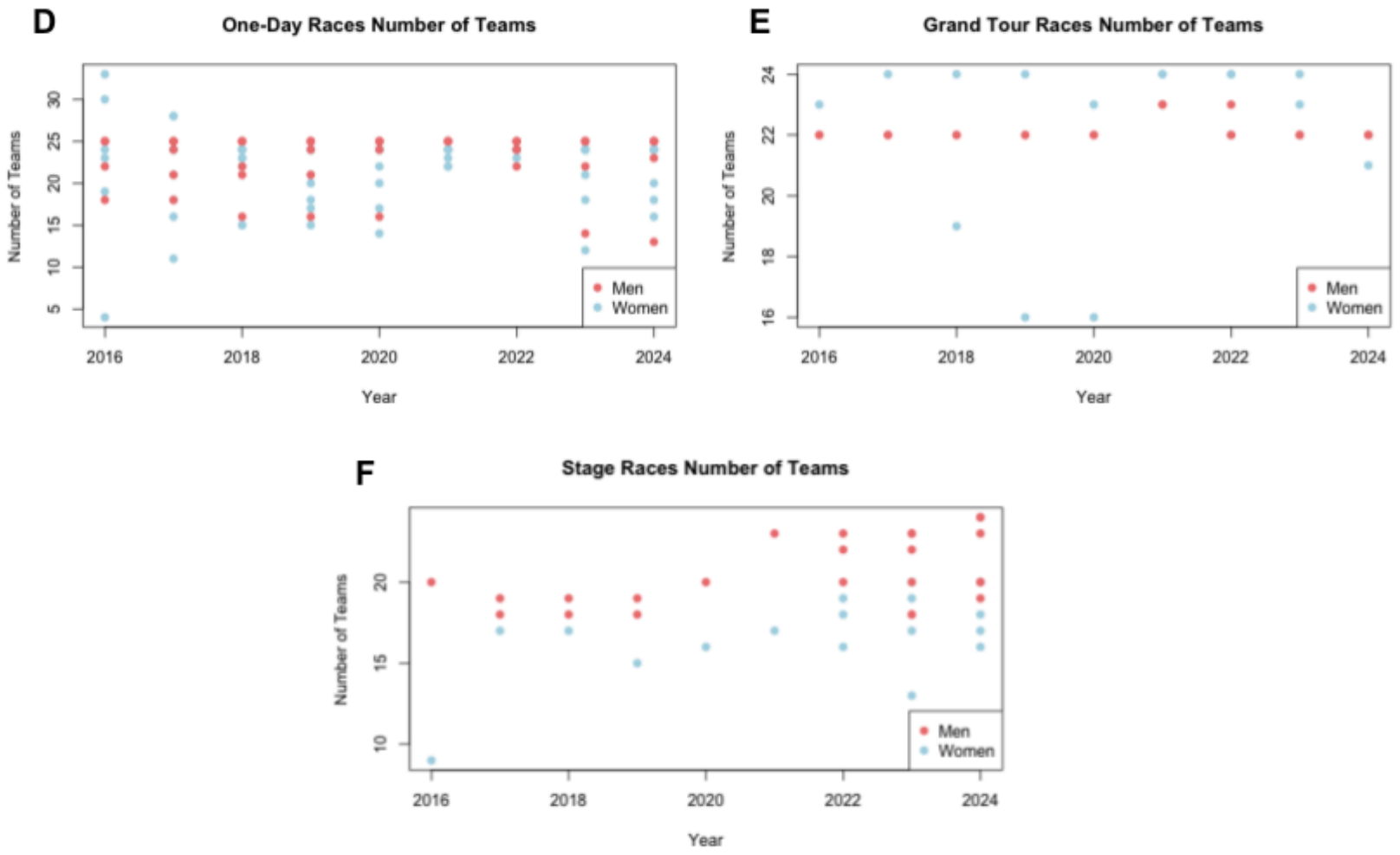
To further confirm these observations, an additional set of plots was constructed, which visualizes the mean distances (in kilometers) for the men's and women's races of each year in our dataset, with this information again being split up by the type of race. This leads to the following:



Figures A1, B1, C1: For each year from 2016 to 2024, the mean distance (in kilometers) is plotted within a bar chart for the men's (red) and women's (blue) overlapping UCI World Tour races. The information is divided by race type, with A1 representing one-day races, B1 representing Grand Tour races, and C1 representing stage races.

These figures then further confirm what was seen within our previous quasi-scatterplots of race length. By examining the plot for each category, we can clearly observe that for all three types of races, the mean distances of the men's races are consistently higher than those of the women's races over all of the years considered. This difference in means is then most stark for the Grand Tours, though it is still abundantly clear for the stage and one-day races, particularly when it comes to the former. There's also no obvious trend upwards or downwards in mean distance with respect to time in any of the categories considered, for either men or women.

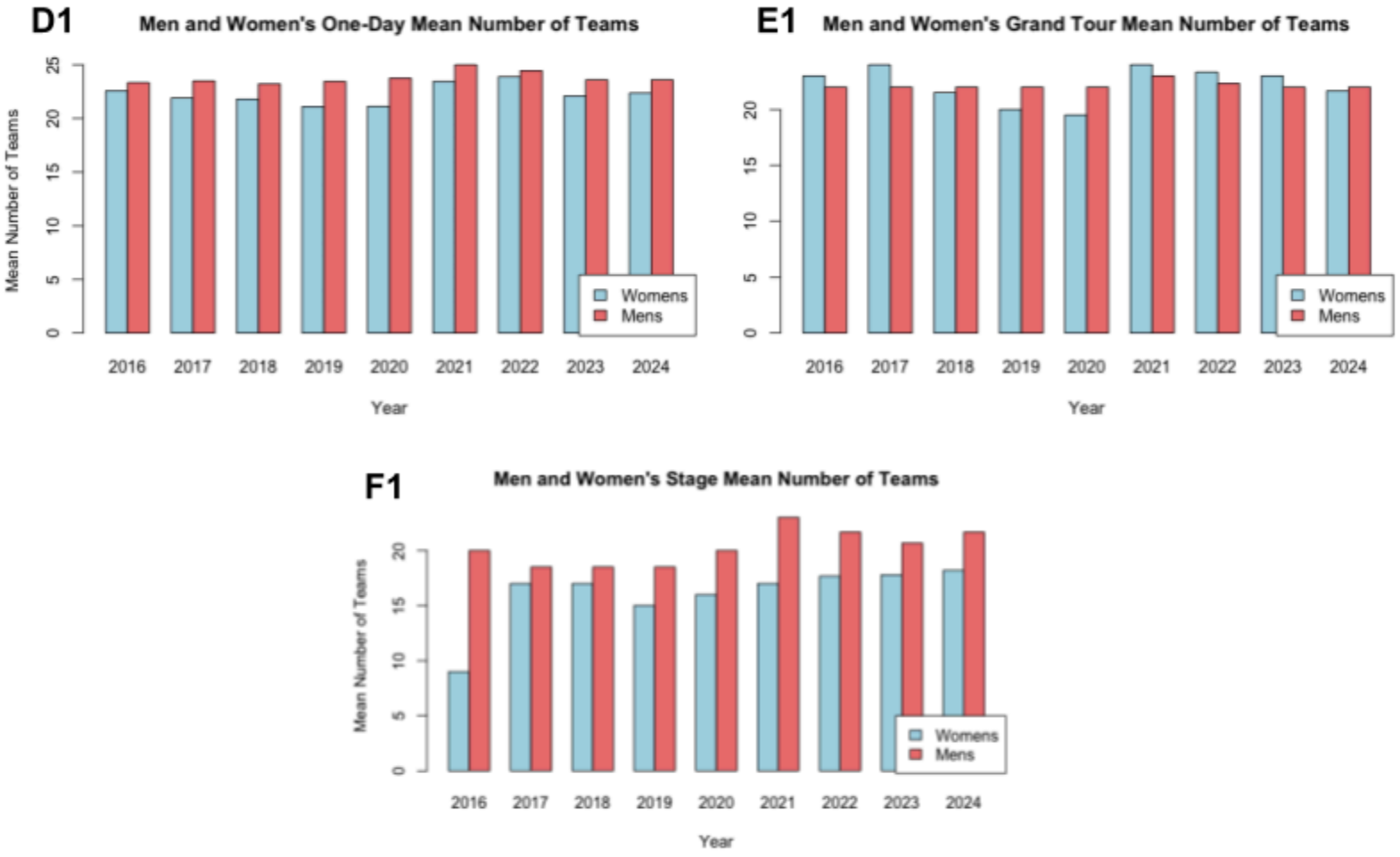
Next, we examine how the number of teams involved in the men's and women's races of our dataset has changed from 2016 to 2024, again breaking this down into the three main types of races that we chose to consider. This leads to the following figures:



Figures D,E,F: For each year from 2016 to 2024, the data points on the total number of participating teams in the men's (red) and women's (blue) overlapping UCI World Tour races are plotted within a quasi-scatterplot, with D corresponding to the one-day races, E to the Grand Tour races, and F to the stage races.

Within each of the plots above, we can observe that the data points corresponding to the men's and women's races seem to be relatively close to one another, with the points for these two groups tending to be the most tightly intermixed for the one-day races, and slightly more spread out for the stage races and the Grand Tours. Hence, we see that for each year, the numbers of men's and women's teams participating in the UCI World Tour races considered within our study are quite similar, with this similarity being strongest for one-day races, and slightly weaker for stage and Grand Tour races. As was the case with race distance, we also don't observe a clear trend upwards or downwards with respect to time in any of the plots above, for either men or women. It thus appears that for all three categories of race considered, the number of teams participating in the races of that category doesn't tend to greatly increase or decrease with time, with this being true for both the men's and women's versions of these races.

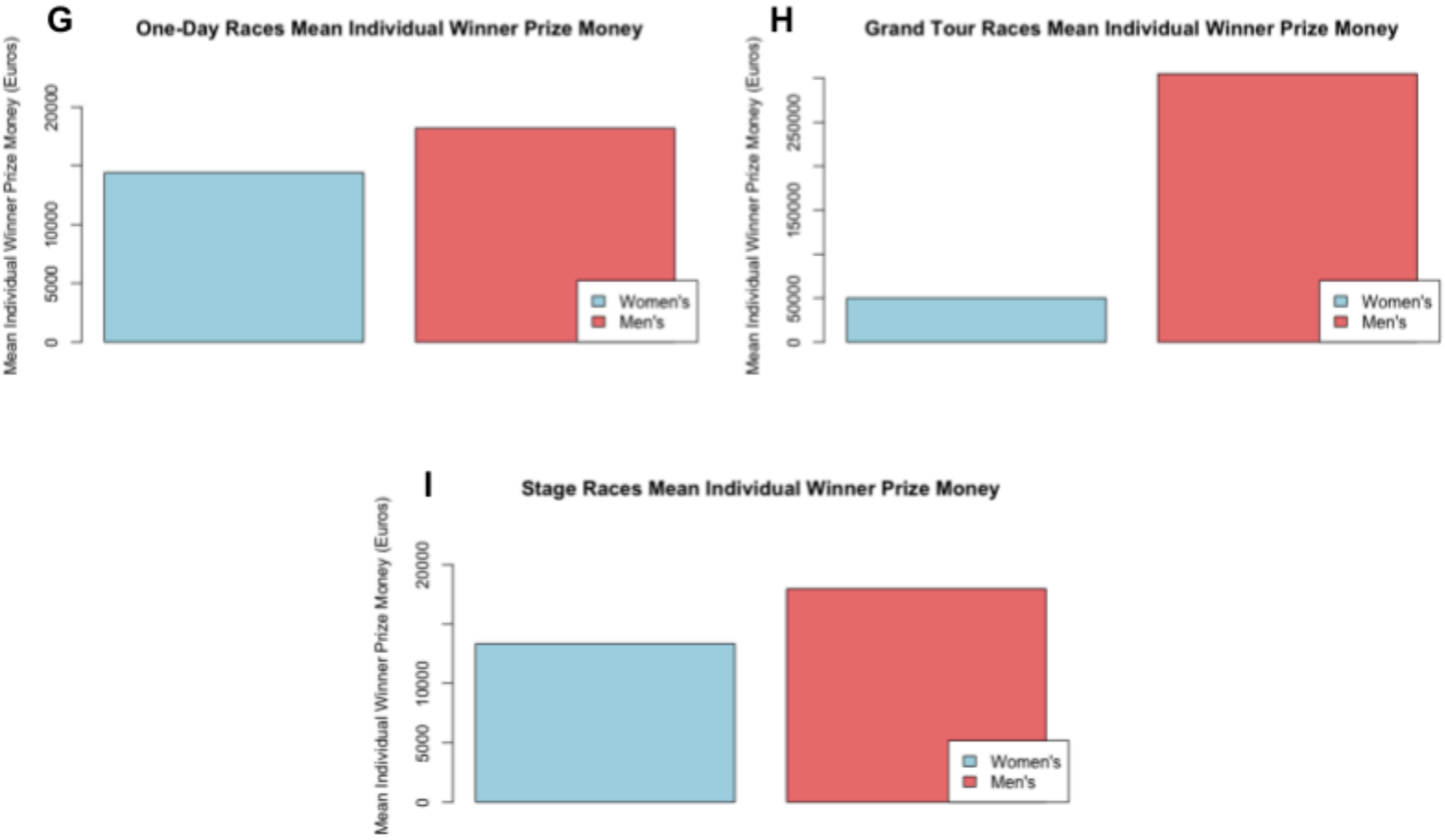
To further confirm these observations, we again constructed a secondary set of plots, which visualize the mean number of teams within the men's and women's races for each year from 2016 to 2024, with this similarly being divided by type of race. The plots are as follows:



Figures D1, E1, F1: For each year from 2016 to 2024, the mean number of teams is plotted within a bar chart for the men's (red) and women's (blue) overlapping UCI World Tour races. This information is divided by race type, with D1 representing one-day races, E1 representing Grand Tour races, and F1 representing stage races.

In support of what was previously seen, we find that the mean numbers of teams participating in the men's and women's races for each year tend to be relatively similar, with this being true across the race categories. There is also again a lack of any clear trend upwards or downwards with respect to time in any of the plots. We can now observe as well that the mean number of teams tends to be closest between the men's and women's races within the one-day and Grand Tour categories, with greater differences being seen within the stage category.

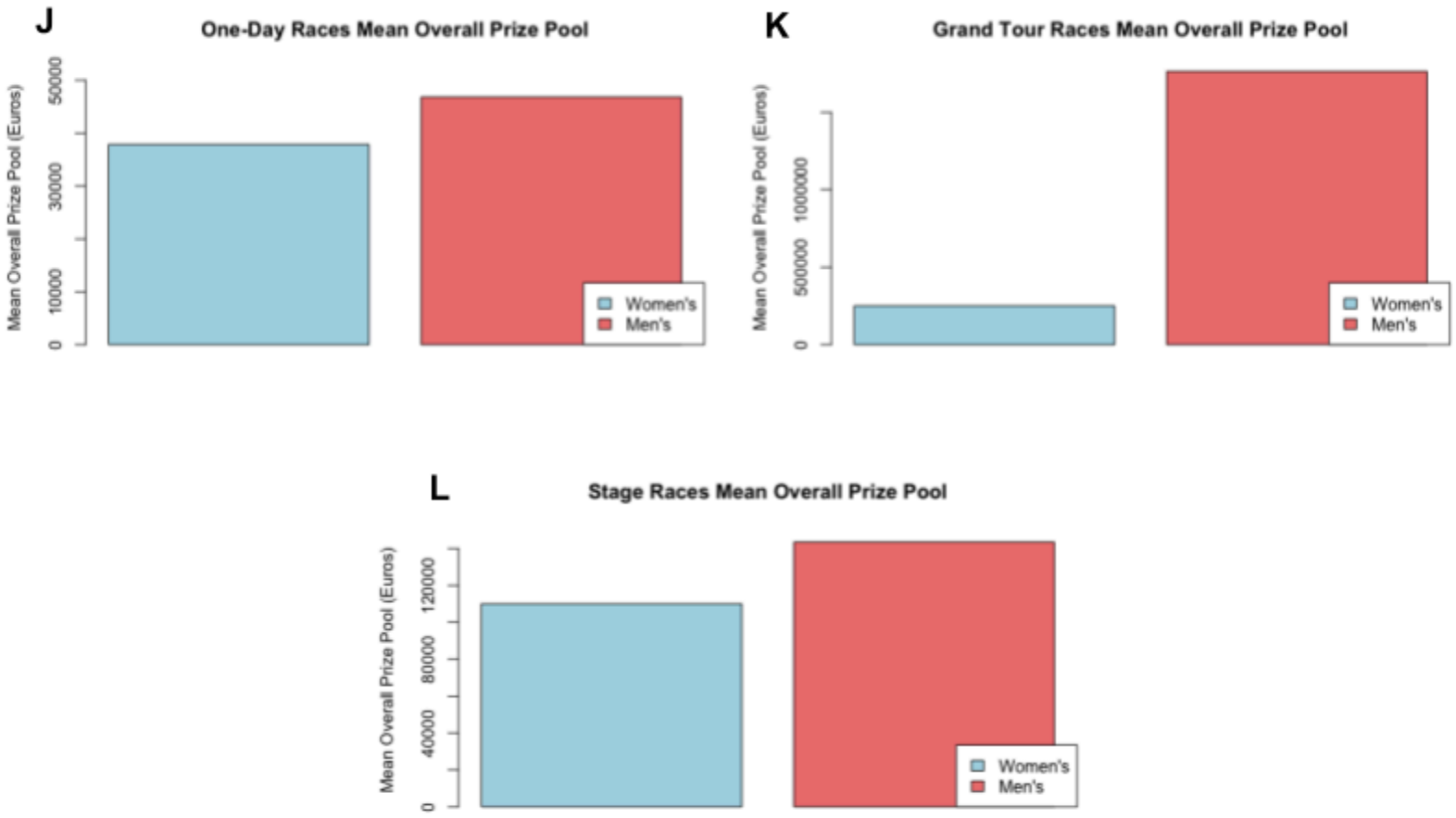
As is mentioned in the “Introduction” section above, the amount of prize money offered for races is also an important metric to consider when examining gender disparities within elite road cycling. Thus, we chose to examine two final statistics: mean individual winner prize money and mean overall prize pool, again dividing this by race category. A set of figures expressing how the former differs between the men’s and women’s UCI World Tour races that we considered is then provided below:



Figures G, H, I: For each of the three primary race categories, the mean individual winner prize money (in euros) is plotted as a bar chart for the men’s (red) and women’s (blue) UCI World Tour races that were considered. G then corresponds to the one-day races, H to the Grand Tour races, and I to the stage races.

When examining these plots, we can observe that the mean individual winner prize money for the men’s races is consistently higher than that of the women’s races within each category. This divide is then somewhat smaller for the one-day and stage races, while being

extremely large for the Grand Tour races. We then plotted the mean overall prize pool for the men's and women's races of each category in a similar manner, leading to the following set of figures:



Figures G, H, I: For each of the three primary race categories, the mean overall prize pool (in euros) is plotted as a bar chart for the men's (red) and women's (blue) UCI World Tour races that were considered. J then corresponds to the one-day races, K corresponds to the Grand Tour races, and L corresponds to the stage races.

A similar trend can be seen within these figures. The mean overall prize pool for the men's races is consistently higher than that of the women's races in each of our three categories, with this divide being relatively small for the one-day races, somewhat larger for the stage races, and extremely wide for the Grand Tour races.

Considering these observations made based on the visualizations of our collected data, we reach a key set of ideas that will be explored more rigorously within the "Analysis" section of this paper. For one, we want to see how significant the differences truly are between the mean distances of the men's and women's races in each category, and whether any substantial differences exist between the mean number of teams participating in these races. We also want to consider whether there has been a significant change in either mean race distance or the mean number of teams from 2016 to 2024, specifically seeing whether there has been any sort of increase with time. Finally, we then want to examine how significant the differences are between the men's and women's races in each category when it comes to mean individual winner prize money and mean overall prize pool. By more deeply investigating such questions, we can then delve further into the extent of gender disparities in elite road cycling.

Appendix

A link to the spreadsheet containing all of the data used within this study is given here, and a link to a GitHub repository containing the code used to generate our figures can be found here.

References

Colangelo, J., Smith, A., Buadze, A., & Liebrenz, M. (2023). "There just isn't any other option—so we just have to put up with it": mental health in women's cycling and the necessity of structural change. *Frontiers in Sports and Active Living*, *5*. https://doi.org/10.3389/fspor.2023.1270957

Mallon, R. (2024, August 10). *The Tour de France Femmes' Long and Winding Road: A brief history of the women's Tour de France*. road.cc. https://road.cc/content/feature/brief-history-womens-tour-de-france-302715

Tour de France History | Continental Tires. (n.d.). https://www.continental-tires.com/stories/tour-de-france-history/#:~:text=The%20first%20Tour%20de%20France,boost%20its%20declining%20sales%20figures