**Team**: MCMC Hammer

**Slide 1**: Outside Data

To extend our analysis, we thought it would be helpful to include player statistics in order to better understand how the variables we were provided interacted with personal data like age, height, weight, and average points per game. We found a dataset that provided this information, but we did not know how PlayerIDs mapped to actual players. To solve this problem, we used the gps data to identify the substitutes in each world cup game. This matched many of the players with their ID, but some of the players never started out as substitutes, so we identified these players by their tries after approximating the orientation of the field.

**Slide 2**: Recovery Variable

Our focus for the analysis, was how different variables interact with the amount of time it takes to 'recover' from fatigue after a game. In order to be able to do any kind of inference or analysis we had to be able to quantify recovery. The best way we found to be able to do this was by measuring the number of days it took each player to recover from each tournament. We chose to look at tournaments because all of the games occur either on the same day or in quick succession, and would therefore have no individual recovery times. We used the player's average fatigue as a threshold for recovery. We then looked at individual player's fatigue over time after each tournament and compiled it. We calculated each players average recovery time to be 3.38. Something to note is we discarded player 9, because she overall did not have reliable data, and tended to not have fatigue measurements around tournaments anyway. We also had to drop the Kitakyushu tournament, since it took place so soon after the Commonwealth tournament, and had significantly increased recovery times, as the players were recovering from two tournaments.

**Slide 3**: Inference

We used multiple linear regression to explore what variables best explained Average Days to Recover from fatigue. We were concerned with recovery from fatigue because the Canadian National Women's Rugby Team wanted to understand and manage player fatigue. By doing so, they could better protect their players from dehydration, injuries, and other harms.

We found several variables that were significant in explaining a player's average recovery time. Although some of the variables cannot be directly changed--Age and Menstruation, for example--the variables that can are worth looking into. Since we found duration to be significant, team managers can take that into account when planning training sessions; however, we found that the session type itself was not significant. It is possible that more detailed data about the session may be produce more insight. Additionally, Player 3 had a high average points scored, and when we removed her, it changed one of the model coefficients from a small, negative number to a larger positive number.

Last Spring I participated in Duke's Datafest. For those of you who don't know, a datafest is when Duke gives us a big dataset and asks us essentially to play and give our insights on it. It was structured similarly to a hackathon where there are teams and everyone is working in the same room over the weekend to work with the data. Last year, the data came from a Canadian women's rugby, the Sevens, and they posed the focus to be fatigue.

So we were given self reported data from each game and practice, about the players wellness, how they were feeling that day, and they workload, how hard they thought they worked that day. We were also given GPS data, that told us where exactly each player was during the games at all times. This ended up being very useful for our analysis and visualization. We decided that we wanted to know more about the players, and since the players were de identified, we didn't know their age or weight or records, which were all data point that could be useful in determining their responses to fatigue. So we decided we needed a way to identify the players beyond player 1, player 2, etc.

So what we did was graph the players locations during key points in the game, and saw whether they were playing or on the bench, seeing who was a substitute, to cross reference this with the official game website as to identify who each player corresponded with in the data. We were able to recover data about their age, their weight, and their average points scored.

Here is a more zoomed in version of the graph, we were able to make this with ggplot, go tidyverse!

After this excitement, we still needed to make some sort of conclusion and give our insight. We decided to focus on fatigue recovery, as to see how different factors about a player affected their ability to recover from fatigue and perform at the games. We were given a lot of data from different tournaments, so seeing how they recovered from a tournament was one of our focuses. We decided to make a threshold for fatigue for each player, since every player is different and will obviously report fatigue differently. We found that the average days to recover from each players "fatigue threshold" was about 3.38 days. This graph shows how we determined this, the green line is player 3's fatigue threshold, and it seems as though during this tournament it took player 3 5 days to recover.

We also made some basic multivariate regression inferences, and found that the variables that seemed to affect fatigue recovery the most were SleepHours, SleepQuality, Pain, Illness, Menstruation, Nutrition, USG, TrainingReadiness, Duration, Age, Average Points Scored. One of my team members made this graph to show fatigue during the Dubai tournament.