[Gabriel A. Santiago Plaza](mailto:gabriel.santiago21@upr.edu)   
Alejandro M. Álvarez Bagnarol

[Christopher L Ayala Griffin](mailto:christopher.ayala@upr.edu)

[Carlos J Diaz Vargas](mailto:carlos.diaz44@upr.edu)

**Notes on the Importing and Cleaning Data**

The first thing we did as a team was give each other a role:  
 1. Recorder - Gabriel

2. Spokesperson- Alejandro

3. Manager- Carlos

4. Reflector - Christopher

After that, we went straight to the project and as we went through some problems that we could not solve, we went to the prerequisites to seek more help, and youtube. Since we did not have 3 hours for each course and the project, we decided to do everything as fast as we could. Each problem was fairly easy and each of us gave some good and important ideas to solve each of the test cases. In the end, we finished the project.

It all went as follows;   
1. In the first exercise we used library functions to add readr and dplyr so that we can read the csv and then check the structure of what returned the read\_csv function.

2. In the second exercise we all decided that creating the dataset wwr\_raw using that format is like the best way to pass all the tests

3. In this exercise we started hesitating a bit on how we were going to remove all NA inside the dataset, so what we did was filter out all NA using is.na, to remove all the rows that had NA

4. What we did was find the column where date was NA and replace it with another date and then change the venue one from NA to Groupama Stadium

5. For this part we created wwc\_3 from wwc\_2 and used functions separate the score, to away\_score and home\_score, so we can separate the 2 - 0 into\_home score = 2, away\_score = 0. The same thing with pks, away\_pks and home\_pks. At the end, replace every NA column from away\_score or home\_score with a 0

6. We loaded ggplot2 and then used ggplot with wwc\_3 and created a boxplot of attendance by venue using geom\_boxplot, geom\_jitter to put some color and to represent the data more beautifully.

7. We grouped everything by venue to summarize the number of games, minimum attendance and minimum attendance. We used mutate inside a replace and which to remove an extra zero from attendance. After updating the outlier we did the same thing as the beginning but with the updated values.

8. We added the correct geometries to create a boxplot of the attendance by venue. We used geom\_jitter to add the color red and reduce the size to 0.5. And inside of labs we added the title and subtitle.

9. We added the x and y parameters to give titles to the X axis and the Y axis, and color to add colors to each line, plus adding geom\_line to graph it in line.

10. We added arrange(desc(attendance)) to then know which were the highest attendance and store them in variables  
  
 Even if we divided our team with specific roles, we worked using teamwork to help each other with our strengths and weaknesses. In the end, the roles did not matter too much since we all worked in our own unique way, which is exactly what teamwork is about.