USERGUIDE

This program uses data from the 2018 Soccer World Cup group stage and predicts the winner between teams within a group.

The Soccer World Cup is an event that takes place every 4 years and that hosts teams representing 32 countries around the world. The World Cup begins with a group stage, in which there are eight groups with 4 teams each. In this stage, every team plays the other three teams in the same group, and a ranking is formed depending on how many matches each team wins (3 points), losses (0 points) or ties (1 point). Only the two teams with the best ranking of each group go to the next round. After the group stage, the World Cup takes a tournament setting, where teams play one match against their opponent once, and who wins goes to the next round and who losses is eliminated. Since 16 teams make it pass the group stage, the next round after the group stage is the round of 16, then the quarterfinals, the semifinals, and by last, the final.

In world soccer, there are six regional confederations: the Asian, the African, the Central American, the South American, the Oceanian, and the European. In the lead up to the World Cup, each confederation holds "World Cup Qualifier" matches. In 2018, out of the 32 teams that qualified, 14 were from the European region, 5 were from the South American region, 3 were from the Central American region, 5 were from the Asian region, and 5 were from the African region. No team for the Oceanian region qualified.

Prior to each world cup, there is a meeting where the groups are selected at random. Teams are placed in pots based on their FIFA (International Federation of Association Football) world ranking, the highest ranking teams go to the first pot, the second highest to the second pot, and so on. There are 4 pots in total, and one team is selected from each pot via random draw to form each group. This process is repeated 8 times to create the 8 groups.

In the 2018 FIFA World Cup, the eight groups were the following:

- Group A:
  - Uruguay
  - Russia
  - Saudi Arabia
  - Egypt
- Group B:
  - Spain
  - Portugal
  - Iran
  - Morocco
- Group C:
  - France
  - Denmark
  - Peru
  - Australia
- Group D:
  - Croatia
  - Argentina
  - Nigeria
  - Iceland
- Group E:
  - Brazil
  - Switzerland
  - Serbia
  - Costa Rica
- Group F:
  - Sweden
  - Mexico
  - South Korea
  - Germany
- Group G:
  - Belgium
  - England
  - Tunisia
  - Panama
- Group H:
  - Colombia
  - Japan
  - Senegal
  - Poland

What this program does is to calculate the predicted probability of a team winning against another one from the same group based on the 2014 FIFA ranking of both teams, the history of wins and losses against each other, and the history of goals against each other.

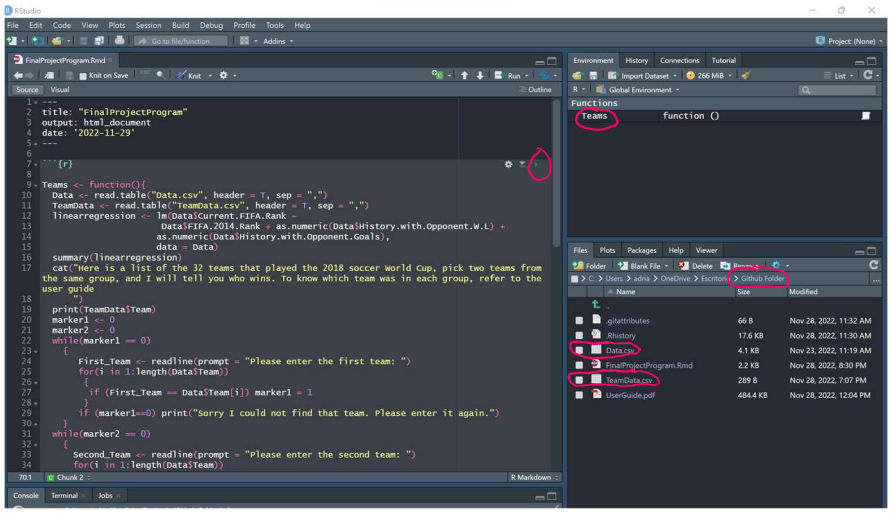
To be able to use the program, the first thing you should do is to Download R and RStudio to your computer, I am attaching here a YouTube video with the instructions: https://www.youtube.com/watch?v=cX532N_XLIs&t=64s

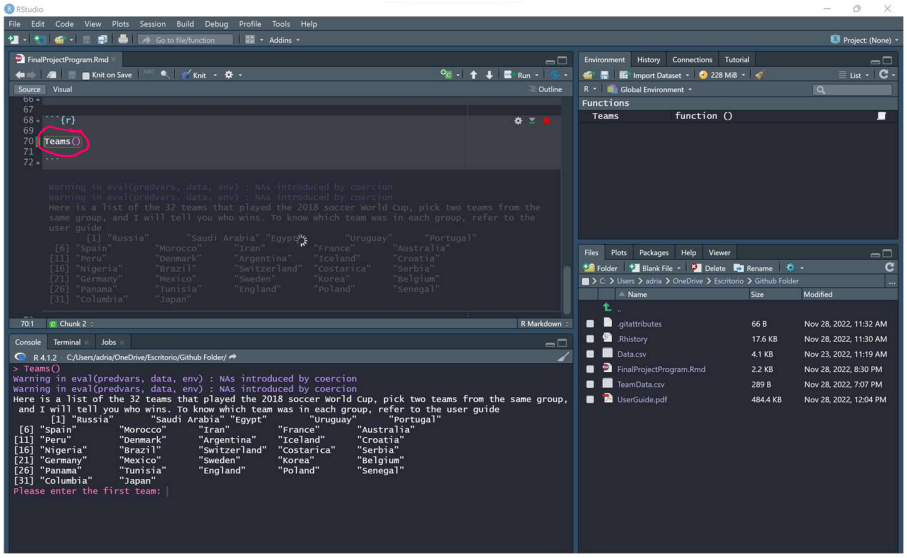
Next thing you should do is to download the data to your personal computer. It is very important that your R directory is the same as your folder containing the datasets. To do so:

1. Open RStudio
2. On the top left of your screen, click Session
3. A taskbar will be displayed under it, and you should be able to click on "Set Working Directory"
4. Then another taskbar will be displayed, click on "Choose Directory"
5. Your computer desktop will then appear, and you should be able to click on your folder containing the dataset as your working directory.

If you have successfully set the folder containing the datasets as your working directory, you should be able to see them in your right-hand side of your screen.

To run the program, click the green play symbol at the end of the 7th line of code, this will create the function that runs the program. Then, if you scroll down, there is a new chunk of code starting on line 68. To start the program, select "Teams()" on line 70, and then click Ctrl + Enter.

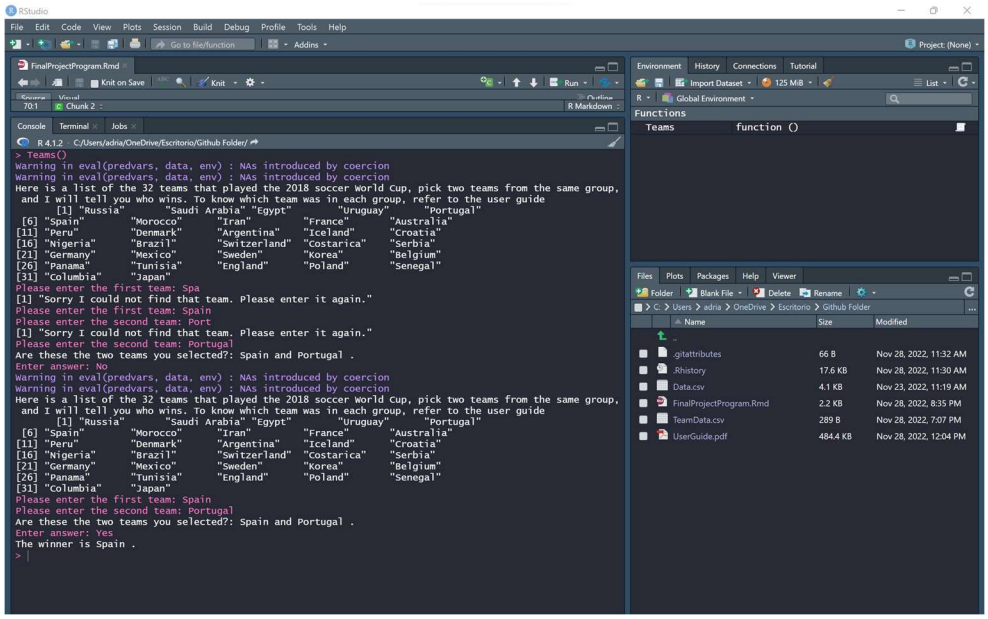




When you start the program, it is going to display the 32 teams that played the 2018 FIFA World Cup and how they are spelled. Then it is going to ask you to enter the first team you want to look at, input it and press “Enter” on your keyboard. Make sure that any word you enter has the first letter capitalized. If the team you have entered does not match one of the 32 teams on the list previously shown by the program, it will ask you to re-enter the first team again until you input one of the 32 teams listed at the beginning. Once you have entered a team on the list as the first team, the program will ask you to enter the second team, the team the first team you selected plays against. Input it and press “Enter”. Again, if the team you have entered does not match one of the 32 teams on the list previously shown, the program will ask you to re-enter the second team until you input one of the 32 teams listed at the beginning. In addition, if the second team you entered is the same as the first one, the program will ask you to re-input it again. Once you have entered a team on the list as the second team different from the first team, it will ask you to confirm that those are the two teams you selected. If they are, you should enter “Yes” with the first letter as a capital letter and press “Enter”. If they are not, the program will re-start. If you confirm that those are the two teams you selected, the program will calculate the winning probabilities of both teams and output the winner team, the one with the highest winning probability. Do not worry about these two warnings, the program will run anyways:

Warning in eval(predvars, data, env) : NAs introduced by coercion
Warning in eval(predvars, data, env) : NAs introduced by coercion

Here is an example where I have entered the incorrect teams the first try and where I have not confirmed that those were the two teams I selected. See how the program re-started. I ended up inputting the correct teams and the program output the winning team.



In the previous example, I picked two teams from Group B: Spain and Portugal. In the bottom left-hand side of the screen, you can see how I typed them with the first letter as a capital letter. I did the same when I typed “Yes”, and the program, then, output the winner.

If you try and input every single match that happens in every group, you will know the teams that the program predicted to advance to the round of 16. You should consider that the model is not the best predictor because of the time elapsed between observations. The ranking used to predict current ranking is from 2014, and the history of wins, losses and goals can be from 1990. In the 2018 FIFA World Cup, Spain and Portugal actually tied.