Question 1: Dplyr Review

Load the Champion's League dataset, Champions.csv, from the homework folder. The dataset records 100 Champion's League matches between different soccer clubs. Note that this dataset is generated from simulation (not the real match history).

1. Use filter to find out rows (games) that home team wins, i.e., HomeGoal > AwayGoal . These rows should be stored in a new tbl_df object. Also use filter to find out rows that the HomeTeam is either "Barcelona" or "Real Madrid".

2. Use select to create a new table which exactly includes all the variables about home team (and excludes variables about away team). Create another table which only includes 6 columns: HomeTeam, AwayTeam, HomeGoal, AwayGoal, HomeCorner, and AwayCorner. Hint : you may use the argument starts_with or contains in the function select .

3. Use arrange to reorder the dataset by the number home goals, and display the following 6 columns of the reordered data: HomeTeam, AwayTeam, HomeGoal, AwayGoal, HomeCorner, and AwayCorner.

4. For each HomeTeam, find out its average HomeGoal, average HomePossession (possession rate), and average HomeYellow (number of yellow cards). Summarise the results in a table.

5. (Optional) Find out the top 5 frequent score (i.e., the combination of HomeGoal:AwayGoal). Note that 1:0 should be treated the same as 0:1 .

Question 2: Scatterplot

The data frame cars in the datasets package records the speed (in mph) and stopping distance (in ft) for 50 cars. Load the dataset using data(cars)

1. Create a scatterplot of dist (y-axis) vs. speed (x-axis).

2. Refine the basic plot by labeling the x-axis with "Speed (mpg)" and the y-axis with "Stopping Distance (ft)". Also add a title to the plot.

3. Revise the plot by changing the every point from the default open circles to red filled triangles (col="red", pch=17).