ISOM 675: Data Visualization
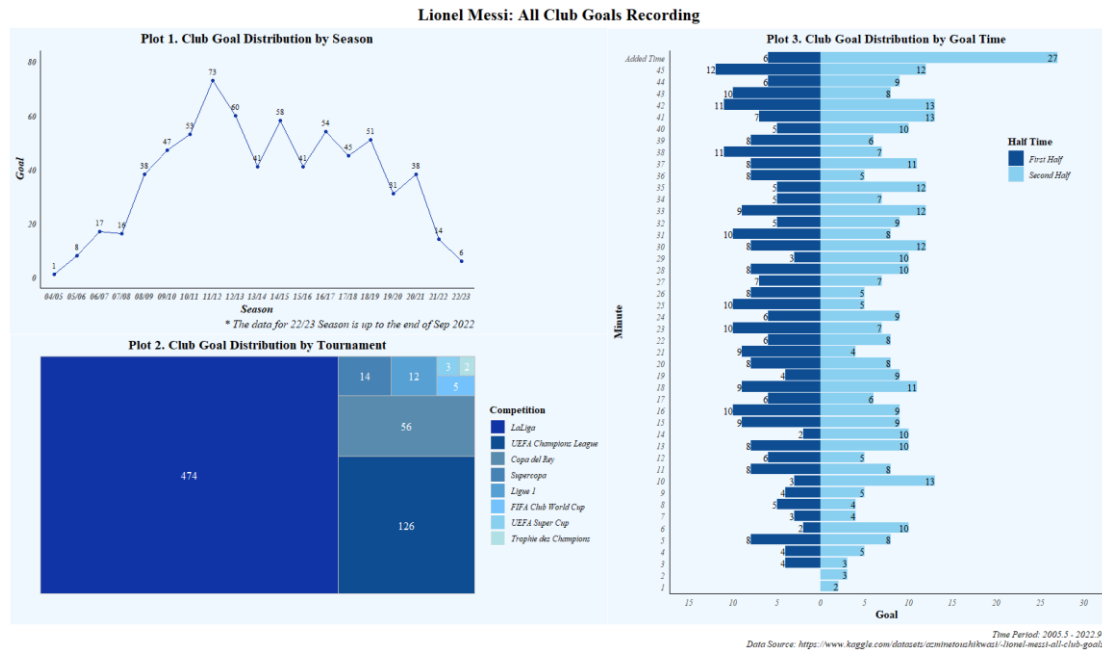**Homework Assignment 1**

Name: *Runqing (Roch) Jia*
NetID: *rjia24*
Submission Date: *2022/10/6*

1. Data Visualization Result

2. Visualization Notebook

1) Data Set Introduction

Lionel Messi, who is one of the greatest contemporary soccer players (sometimes even called "Greatest of All Times"), has experienced a superexcellent career in his field. As a huge fan of Messi, I would like to explore his brilliant achievements on the soccer pitch – and goal is the most straightforward index to measure his historical performance.

Therefore, I chose a goal recording dataset including the core information of his goals since he started his professional career: seasons, goal time, competitions, opponents, assists and so on. By applying this dataset, I can explore all his goals from different perspectives.

2) Data Selection & Manipulation

After understanding the data structure, I decided to display the goal distribution from three dimensions: season, competition (tournament) and goal minute:

- Season is the most generally used and comprehensive grouping criteria to measure the performance of a soccer player throughout a whole season
- I would like to apply tournament as a grouping measure to show the competition power of Messi in different types of competitions (for example, a player can perform very differently in domestic league and international cups)
- Goal by minutes is not very intuitive data, but I am still interested in whether Messi has "high incidence goal time" throughout the 90 minutes of a match

Thus, I tried to group the goal data by the above three criteria, made some manipulations (the data is not very clean though) and get the count of goals.

3) Visualization Approaches

Afterwards, I chose three kinds of plots to display the three tables:

- Season – Line: Since we only have 1 x and 1 y and we would like to track the changes in goals by season, line chart can be a good choice
- Tournament – Tree Map: There are 8 club tournaments that Messi has participated in. Tree Map is a method to intuitively reflect the distribution of all goals and make it comparable
- Goal minute – Pyramid (Bar): We transfer the data into two subgroups by half-time (first and second half), and try to compare the data by two dimensions: different time in the same half and same time in the different half. Pyramid can help us achieve this goal

4) Visualization Understanding

All the three plots are now shown in the first part. There are some interesting conclusions and discoveries in the visualization:

- Plot 1: We can easily find that the noontide of Messi was Season 11/12, when he scored amazingly 73 goals. He kept a great goal efficiency for many years.
- Plot 2: LaLiga (i.e. Spanish league) was the main battleground for Messi throughout his career. It's rational because he spent most of his career in Spain and domestic league takes the majority of match time. However, Messi also performed brilliantly in UEFA Champion League and Copa del Rey (i.e. Spanish cup)
- Plot 3: From the visualization result, we did not find some specific rules for in which minute Messi tends to score more goals. However, it is obvious that he tends to have more goals in the second half, especially during the added time. In regular time, Messi has most goals in the 10th, 41st and 42nd minute of the second half (13).

5) Go Further…

In this visualization assignment, I did not include any complicated calculation for the goals. It prevents me from figuring out more informative discoveries. If we would like to go further in the future, maybe we can measure some more complicated data. For example, the scoring efficiency, which can be represented by average goal per minute. Another example is to figure out for each goal, whether the goal helps Messi's team to equalize or win the match (that is, the goal's power of conclusiveness).