Final Exam: Data Visualization report

Name: Pei, Tian
Students ID: 002255408

Effective data visualization can show the disorganized data set to data analysts in a scientific and effective way in terms of discovering the patterns among them. In this dataset, athletes from different countries and their activities, to show the dataset effectively, it is necessary to post the goals in the first place. For instance, which country get the highest number of medals in all games; does these factors: Height, Weight, Sex, and ages affect athletes getting medals; is it true that the height and weight of those who won medals reached a balance.

The most effective way of showing which country acquired the highest number of medals in the past of events is using the world map. Figure 1 shows that the number of medals among different countries in the world. The color in the chart shows that counties get the various levels of medals, which means that the lighter of color, the higher of medal did countries get. It is necessary to mention that the number of medals in Russia could not be accuracy due to the whole dataset contains the data of Soviet Union which is no longer exist. Nevertheless, we can also know that America is the country which has owned the highest number of medals in throughout history.

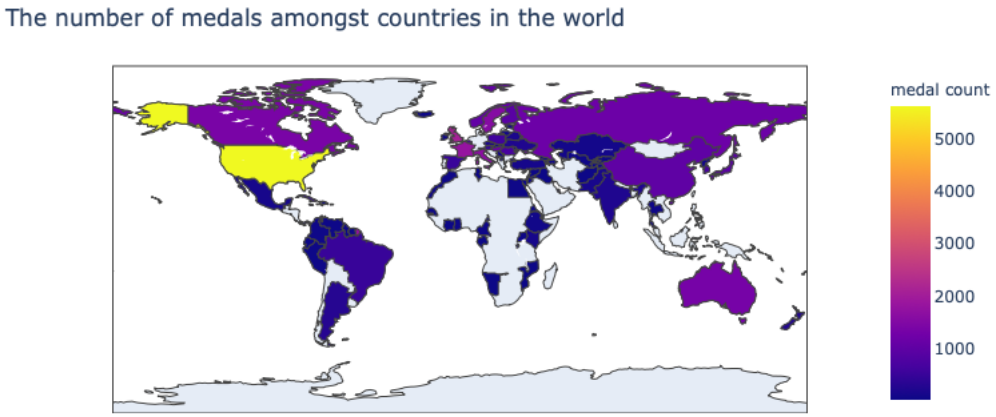

Figure 1: the number of medals among countries in the world

Throughout all players in the world, we expect to find the relationship between whether a player got a medal (gold, silver, bronze all included) or not and their physical features such as age, height, and weight. Considering there are missing values in the dataset, which could affect data visualization during coding, therefore, I decided to remove them because they just take a small part in the whole dataset. Another important processing is that I defined a new column representing if the player got a medal or not. As we can see in the table 2, it is clear to discover the distribution of data. The varied color of dots means if a player got a medal or not. The size of them

represents the age of competitors. In accordance with the distribution of data, there is no clear relationship as the assumption mentioned at first. However, we can find that the most dots of players who got medals are in the middle part of chart. In another word, most part of prize-winners have the relative balance physical quality.

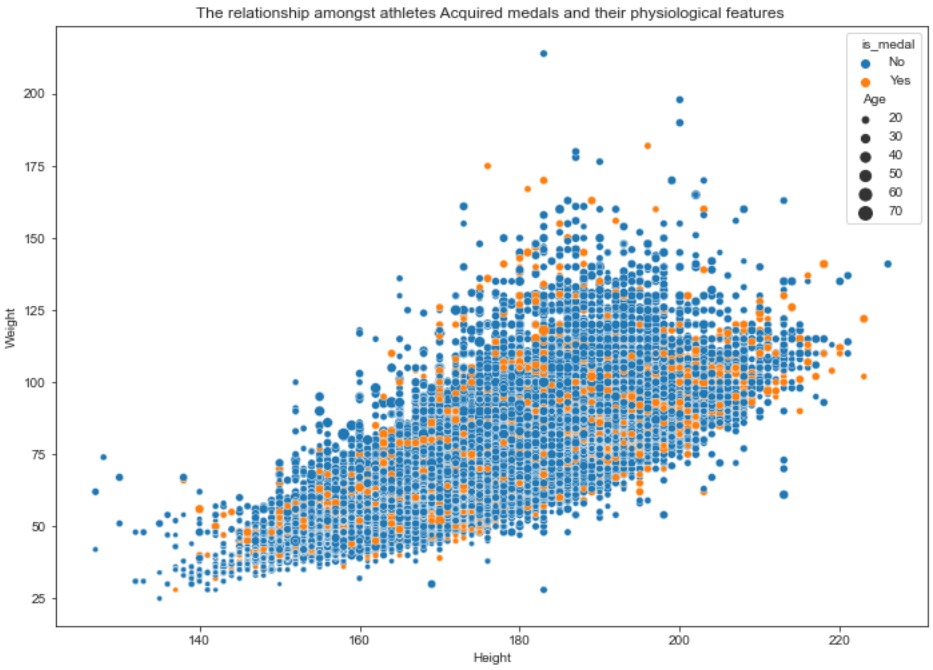

Figure 2: The relationship amongst athletes Acquired medals and their physiological features

To prove our assumption that the player owned the balanced height and weight takes the most part amongst prize-winners further, I used Body mass index (BMI) to group data as a new column. According to the BMI formula, I separated prize-winners into four diverse groups and visualized in the figure 3. It is obverse to know that athletes with a normal body index have won the most medals (76.4%), which proves that the assumption is correct.

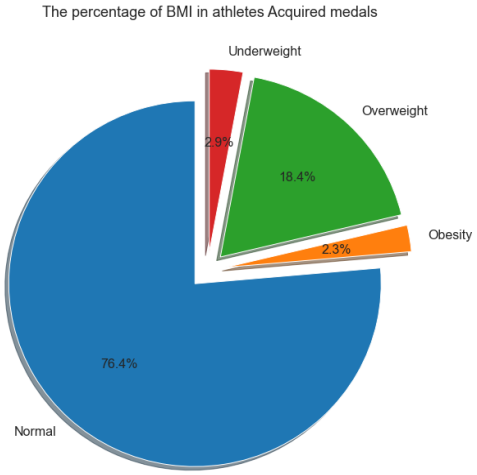

Figure 3: The percentage of BMI in athletes Acquired medals