4.1 Exploratory Data Analysis

Based on this whole data set of Olympics, we did some exploratory data analysis, there are total nine attributes in the data set of Olympics. There are Year, City, Sport, Discipline, Athlete, Country, Gender, Event and Medal. Based on these attributes, we found a lot of interest things.

表格

描述已自动生成

4.1.1 Relationship between Medal and Year

We have generated a bar chart based on the sum of the total number of medals produced by all countries each year. The y-axis is the number of medals, the x-axis is Year (1896 – 2012). The blue color represents the Gold medal, the orange color represents the Silver medal and the green color represents the Bronze medal.

图表

描述已自动生成

Based on the bar chart, we could clearly know the relationship between year and number of medal. As time going on, the number of medals produced in the whole world has gradually increased.

In 1896, the number of three medals was less than 100, and perhaps in 1896, very few countries participated in the Olympics, and only a few sports were carried out at the Olympics, so the number of medals produced was very, very small. But in 1900, the number of medal has a significant improvement, because of the successful Olympic in 1896, more and more country join the Olympics in 1900. And then, in 1908, it is also a big improvement of Olympics. In the following time, the Olympic Games have been gradually and steadily developing.

Based on this bar chart, we found a interest thing, In our general understanding, the number of gold, silver and bronze medals should be the same. But the results show that the number of gold, silver and bronze medals in each Olympic Games is different. Why is that? After we look up the information and the rules of the Olympic Games in the internet. We know the reason is that in the Olympic Games sports, it is common to have a side-by-side first or second-place finish. If there is a parallel first place, the sport produces two gold medals and one bronze. If there is a parallel second place, then the sport produces one gold medal and two silver medals, which is why the number of gold, silver and bronze medals produced at each Olympic Games is not the same.

4.1.2 Medal for each country

After the analysis of the total number of medal for each year. We plan to see how many medals won by each country in all of the summer Olympics. First, we generated a data frame. We could clearly see the first country is USA, in all of the summer Olympics, it got 2235 gold medal, 1252 silver medal, 1098 bronze medal. It is amazing. The United States is the best medal-winning country in the Olympics.

表格

描述已自动生成

Then, we generated a chart about the relationship between number of Medal and Country. It is clear to us that there is not much difference in the number of medals in other countries except the top five countries.

图表, 折线图

描述已自动生成

4.1.3 Gold, Silver, Bronze Rate of Medal

Then, we think using the number of medal to see the relationship between medal and country is not so good method. So we plan to use another way to see the relation. It is the Gold rate, silver rate and bronze rate for each country. First, we generated a data frame like this:

表格

描述已自动生成

We can clearly see the medal rate for each country, the Gold rate of USA is 4.83834, we can also say that almost half of USA medals are gold.

Then we generated a chart which is easy to see this relation:

图表, 条形图

描述已自动生成

Also the blue is gold medal, orange is silver medal and green is bronze medal. Using the chart with different color, it is easy to found the real relation. It is easy to found which country is really good at Olympics (Big Gold Rata).

4.1.4 How to win more medals?

For us stakeholder, such as our country, how to get more medals is an important problem. So we did some analysis for the solving problem.

First, we generated a bar chart for the Number of Medals for each Sport.

图表, 直方图

描述已自动生成

Based on this bar chart, we could see that Aquatics is the sport with most awarded medals. It is about 1300+ gold medal, silver medal and bronze medal in Aquatics. And the second is Athletics and the third is Rowing, the forth is Gymnastics. Therefore, some suggestion for stakeholders, if the country want to win more medals in Olympics, the country could pay more attention on these four sports and pay more money on these four sports’ athletes. It is a useful way to help the country win more medals.

After we generated this bar chart, we found an interest thing in the bar chart, the sports named Boxing and Judo, the number of bronze medal is twice of gold and silver medal.

Two bronze medals were won because of the single-defeat elimination system in boxing and judo. All players defeated by the winners of group A1, A2, B1 and B2 will compete in the resurrection of their respective groups. For example, the first player defeated by the winner of Group A1 will compete with the second player defeated by the winner of Group A1, the winner of the match will compete with the third player defeated by the winner of Group A1, and so on.

4.1.5 Relationship between gender and medal

Based on the gender data in the dataset, we generated a bar chart which could see the relationship between medal and gender.

图表, 条形图

描述已自动生成

The light blue is male and the orange is female. We could clearly see in the first five Olympics, there is a little number of female sport in the Olympics. As the time goes on, more and more female athletes attend the Olympics and get a great result. Even in 2012, the number of male and female almost same.

For stakeholders, countries could not only pay attention to male athletes, they will also pay more attention to female athletes, in this time, female even could win more medals than male athletes.