Summer Olympics Performance Analysis

Furong Bai

2017年4月27日

PROBLEM STATEMENT

Many people like me enjoy watching the Olympic Games, having the valuable and fun time to watch the world top athletes to compete in various sports. People are proud of the success of their country's athletes, thus number of medals won of a nation is important to national prestige. A glance at the by-country medal tables can clearly tell us that not all the nations have the same ability to win the game. While some countries have dominated the medals in specific sports, others are hardly come into the picture. Hence, there is one question we may want to ask while viewing the games: what makes these countries so successful in the Olympic games?

In this exercise, I explored Summer Olympics data from 1896-2012, in order to understand the performance of leading countries over year. The objective is to find answers to the following questions:

- What are the top 10 countries in the history of Summer Olympics?
- How're their performances change over time?
- Which gender won more medals in summer Olympic? Did the pattern changed over time?
- What is the impact of socioeconomic variables on national performance in Summer Olympic games? (GDP, Population & per capita GDP)

Data Preparation

Explore the data sets

First Check the structure of the data.Original data sets were from Kaggle.

"summer" data-set listed information for each medal winner and each medal won. It contained 31165 observations and 9 variable as following:

1. Year: year of the Olympics
2. City: host city
3. Sport: type of sport
4. Discipline: field of sport
5. Athlete: athlete's name
6. Country: which country the athlete belonged to
7. Gender: athlete's gender
8. Event: specific Olympic event
9. Medal: type of medal won: gold, silver or bronze

"dictionary" data-set included the general information of participated countries. It contained 201 observations and 4 variables listed as following:

1. Country: country name
2. Code: country code
3. Population: country's population
4. GDP.per.Capita: per capita GDP in the country

Missing values & outliers

Then check the missing values and outliers within the data.

```
sapply(summer, function(x) sum(is.na(x)))
```

```
##      Year      City     Sport Discipline    Athlete    Country
##         0         0         0          0          0          0
##    Gender     Event     Medal
##         0         0         0
```
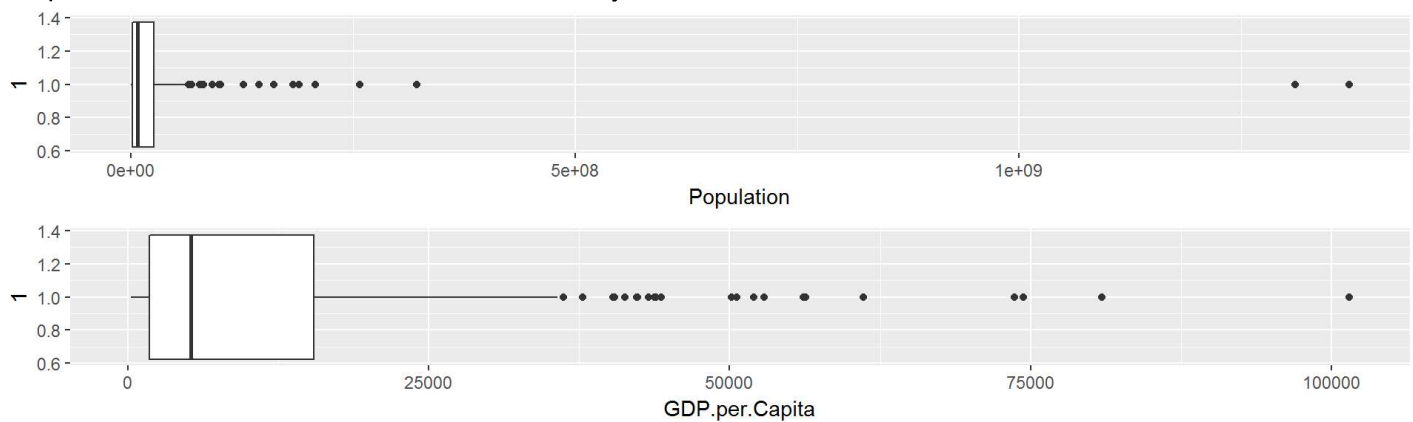
```
sapply(dictionary, function(x) sum(is.na(x)))
```

```
##      Country        Code   Population GDP.per.Capita
##            0           0            5             25
```

- no missing values in "summer" data set
- "dictionary" data set contained missing values: 5 missing in Population and 25 missing in GDP.per.Capita

Boxplots show that outliers or extreme values may exist in the data-set.

- Following showed the extreme values in Population:

```
##    Country Code Population GDP.per.Capita
## 1    China  CHN 1371220000       8027.684
## 2    India  IND 1311050527       1598.259
```

- Following showed the extreme values in GDP.per.Capita:

```
##        Country Code Population GDP.per.Capita
## 1  Luxembourg  LUX     569676      101449.97
## 2      Norway  NOR    5195921       74400.37
## 3       Qatar  QAT    2235355       73653.39
## 4 Switzerland  SUI    8286976       80945.08
```

- Values for these outliers were reasonable.
  - Two extreme values in Population were India and China.
  - Four extreme values in per capita GDP were Luxembourg, Norway, Qatar and Switzerland.
- Their values were actual and make sense

- Thus I didn't exclude these entries in analysis.

Data Analysis

What are the top 10 countries in history of Olympics?

```
## Source: local data frame [10 x 5]
## Groups: Country [10]
##
##     Country  Gold Silver Bronze Total
##      <fctr> <dbl>  <dbl>  <dbl> <dbl>
## 1       USA  2235   1252   1098  4585
## 2       URS   838    627    584  2049
## 3       GBR   546    621    553  1720
## 4       FRA   408    491    497  1396
## 5       GER   452    378    475  1305
## 6       ITA   476    416    404  1296
## 7       AUS   312    405    472  1189
## 8       HUN   412    316    351  1079
## 9       SWE   349    367    328  1044
## 10      NED   233    279    339   851
```

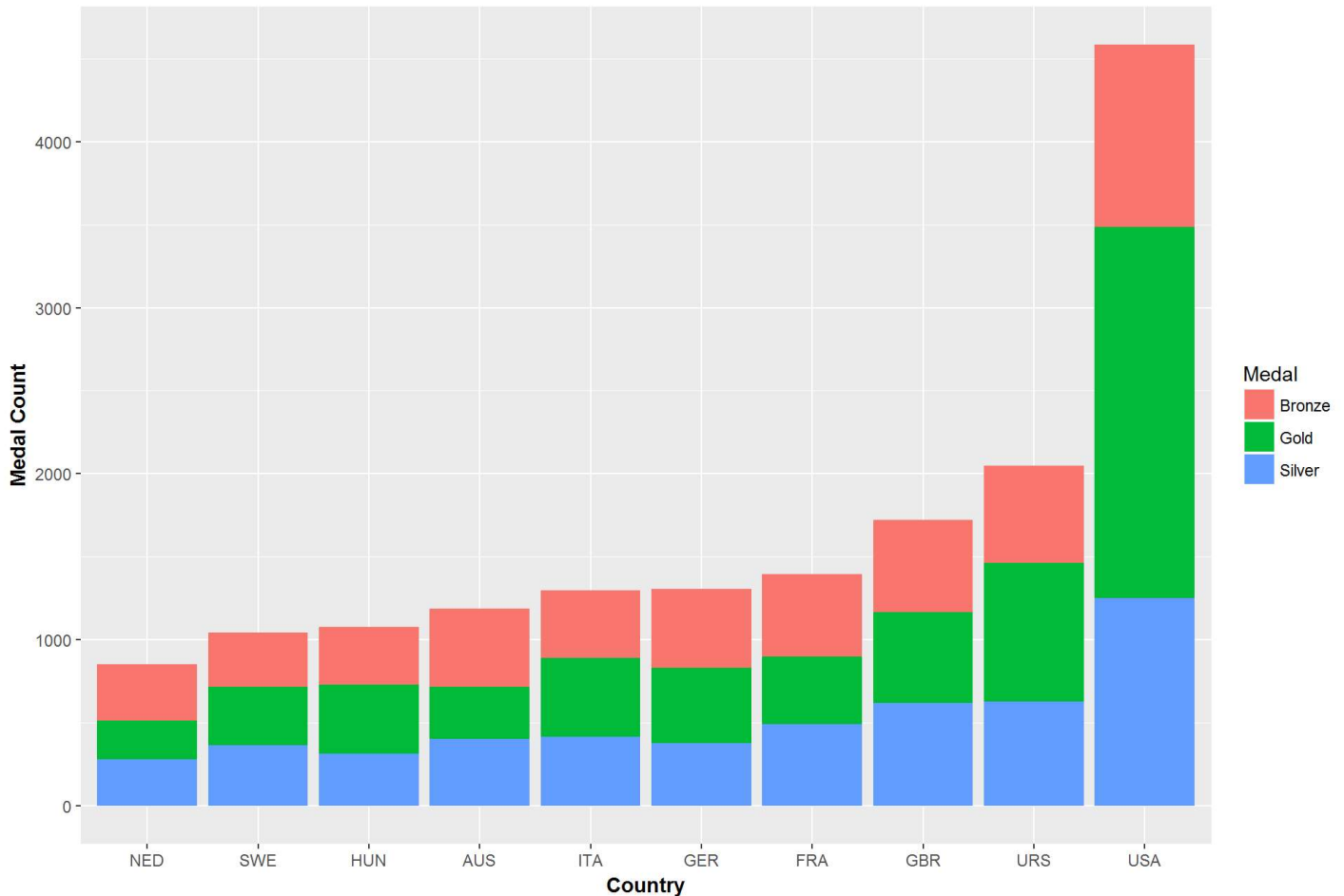## What are the top 10 countries in 2012 Olympics?

- Some countries were leading in 2012 Olympics Medal Tally but didn't show up in overall Medal Tally, such as Russia, China, Japan.
- USA won the most medals in 2012 and overall history.

```
## Source: local data frame [10 x 3]
## Groups: Country [10]
##
##     Country  Year TotalMedal
##      <fctr> <int>      <dbl>
## 1       USA  2012        250
## 2       RUS  2012        130
## 3       CHN  2012        128
## 4       GBR  2012        126
## 5       AUS  2012        114
## 6       GER  2012         94
## 7       JPN  2012         84
## 8       FRA  2012         82
## 9       NED  2012         69
## 10      ITA  2012         68
```

Visualization

Which countries won the most medals in history of Olympics?

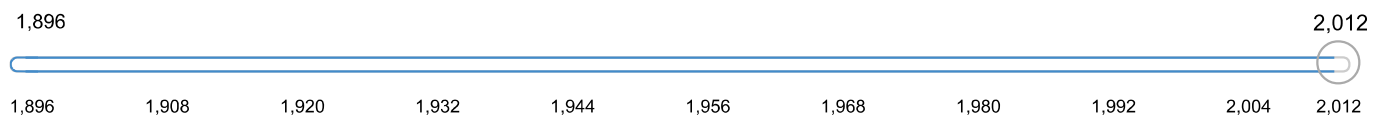**Bar Chart of Medal Counts For Top 10 Countries**

The above graph shows the leading countries in medal tally. Followings are some findings in the graph:

- U.S. won the most amount of medals in history of summer Olympics. Its total medals was above 4500 and was about double of 2nd country in medal tally. Gold medals occupied the highest share compared to silver and bronze.
- Although Soviet Union broke down in 1980s, it still won 2nd highest amount of medals in the history of Olympics.
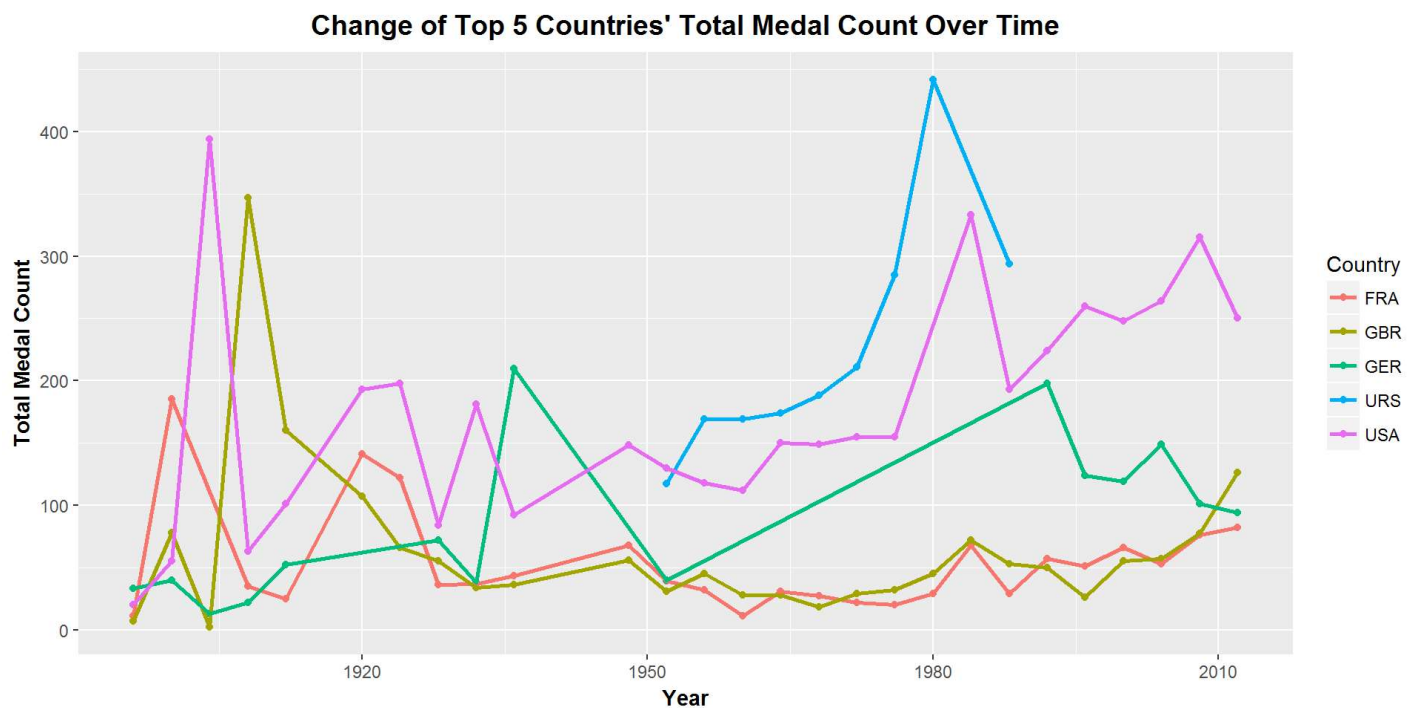
What was the top 10 countries in each year?

- In different year, top 10 countries in the medal tally changed.
- Some countries dropped out of the top list over time and some countries emerged into top list in recent years.
- **Year 1916 1940, & 1944 didn't hold Olympics due to the world war.**
- Select year you want to check which are the top 10 countries in that year.
  Select Year:

1,896   2,012

| 1,896 | 1,908 | 1,920 | 1,932 | 1,944 | 1,956 | 1,968 | 1,980 | 1,992 | 2,004 | 2,012 |

How did performance of top 5 countries in Olympic history change over time?

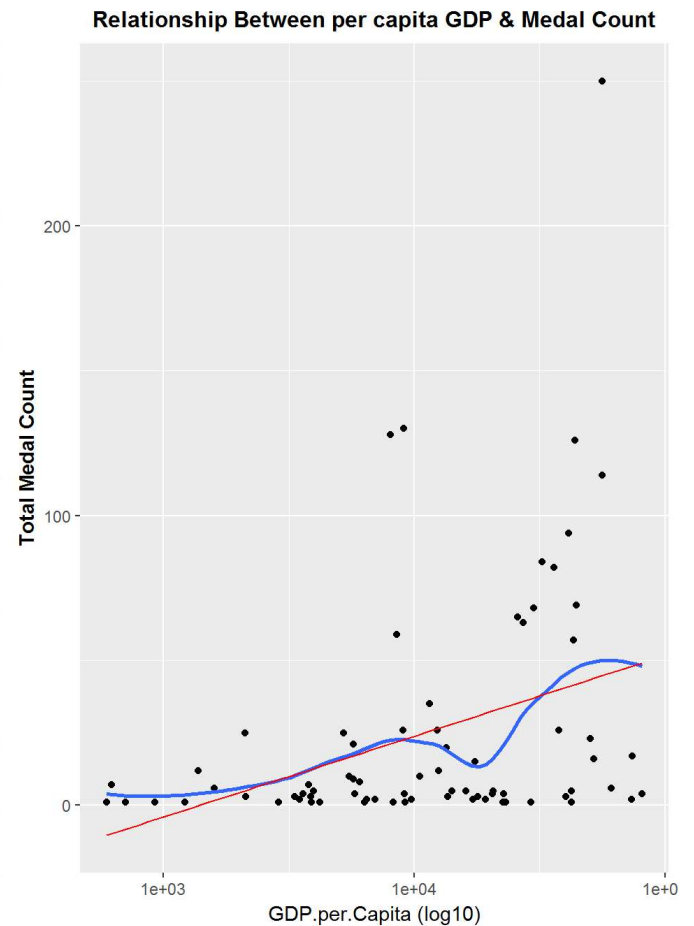**Change of Top 5 Countries' Total Medal Count Over Time**
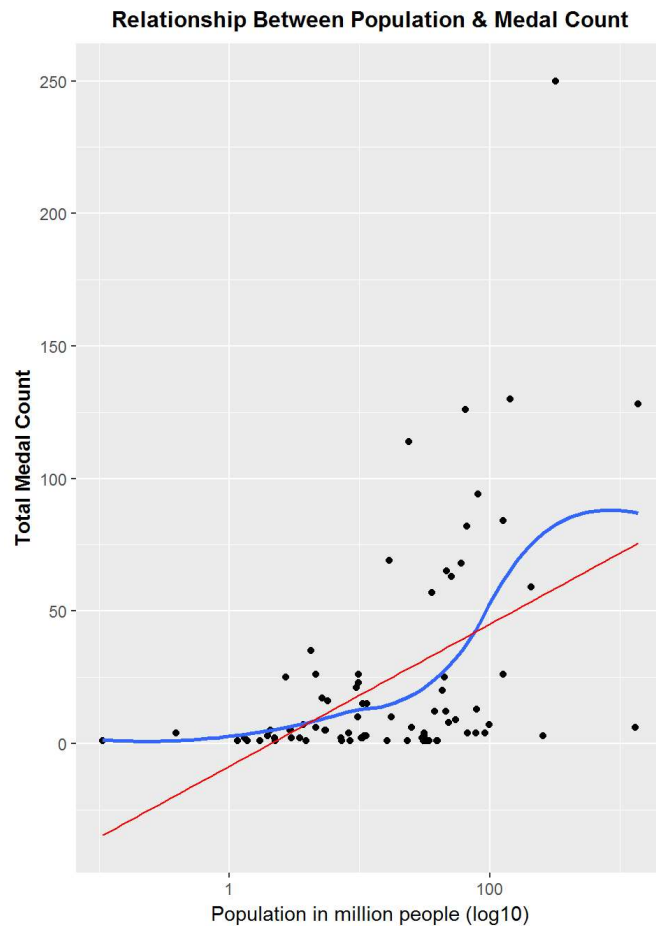

As we can see from the graph above:

- Most of these countries attended Olympic games since invention of the game.
- USA was keeping consistently high performer over time.
- Soviet Union won about 450 medals in 1980, highest ever among all summer Olympic games.
- China didn't come into the picture maybe because it joined Olympics late.

What makes some countries so succeful in Olymics?

- Is there an relationship between some social-economic variables and total medal won?
- Focus on Year 2012.
- Assumption:
  - With larger population, the country may have larger pool of talents.
  - With greater GDP per capita, the country may have greater ability to invest on sports.

- Graphs show the scatter plot and relationship between two social-economic variables and total medals.

- Finally, tit a linear regression model, the output of the model is listed below.
    - Intercepts for log(Pop.million) - 14.563;
    - Intercepts for log(Pop.million) - 16.551;
    - Detailed results were listed below:

```
##
## Call:
## lm(formula = TotalMedal ~ log(Pop.million) + log(GDP.per.Capita),
##     data = country_table)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -51.312 -23.256   0.123  13.383 154.266
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -169.318     31.017  -5.459 6.66e-07 ***
## log(Pop.million)      14.563      2.223   6.552 7.70e-09 ***
## log(GDP.per.Capita)   16.551      3.102   5.335 1.08e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32.03 on 71 degrees of freedom
##    (6 observations deleted due to missingness)
## Multiple R-squared:  0.4534, Adjusted R-squared:  0.438
## F-statistic: 29.45 on 2 and 71 DF,  p-value: 4.859e-10
```