INFO 2201 Write Up

Does Olympic success have any effect on the valuation of worth for a particular sport in that country?

And various other explorations of data.

Name:
Zaid Badawiyeh
zaid.badawiyeh@colorado.edu

Term:
Spring ‘23

Lecture:
010

Recitation:
013 (Jay Ghosh)

Coding Section Overview and Notes

Data types:
CSV Structured Plain-Text **(Type #1)**
JSON API **(Type #2)**

Sources:
Kaggle dataset '120 years of olympic history, athletes, and result' **(Source #1) (Type #1)**
Kaggle dataset 'Forbes Highest Paid Athletes 1990-2021' **(Source #2) (Type #1)**
The World Bank 'GDP (Current US$)' **(Source #3) (Type #2)**

Links:
(Source #1)
https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results

(Source #2)
https://www.kaggle.com/datasets/darinhawley/forbes-high-paid-athletes-19902021

(Source #3) (Homepage + api documentation)
https://data.worldbank.org/indicator/NY.GDP.MKTP.CD
https://datahelpdesk.worldbank.org/knowledgebase/articles/898599-indicator-api-queries

Notes for coding section:
When using functions in the coding section, enter years that correspond with olympic years (preferably summer). Additionally, some countries are listed as abbreviations ie. USA and UK while others are full names ie. Russia, Spain, Brazil, Argentina. This is how the data is set up unfortunately. Additionally, overfiltering the data can leave you with no results, omit optional inputs to see all listings of that input type (ie. all countries).

The goal of this project was to determine if success in the olympics correlates with an increased valuation of that sport in a country. Success would be measured as medals won in the olympics (independent variable) while sports valuation would use average earnings for different athletes by sport and country as its metric (Dependent). Medals won would be sourced by a kaggle csv dataset called, ‘120 years of olympic history, athletes, and results' while average earnings would be a product of a different kaggle csv dataset called, ‘Forbes Highest Paid Athletes 1990-2021'. In addition, the project will take a glance into how hosting the Olympics affects GDP of the host country. For this question, data will be provided by a World Bank JSON API called, ‘GDP (Current US$)’.

In order to answer the questions set forth by this project, the first step I took was to explore the data. In specific, I wanted to explore the Olympic medal dataset further. To do this I created my first function, ‘explore_data’, which was designed to explore and filter through the Olympic medal data by year, sport, and medal type. What this allowed me to do was view the data with filters to determine significant years, sports available, and to discard all values that didn’t win a medal. The function also served to help determine the quality of data and presence of nulls/missing values.

The second step in the process was to define a more complex version of the explore data function that would allow more filtration and serve a larger function that utilizes the results as a dependent variable data entry. Ultimately the goal of function 2, ‘filter_olympic_data’, was to filter by multiple sports or medal types as opposed to only one, filter by team (country), and aggregate the number of medals, grouped by sport, team, and medal type to get a count. This count will serve as our independent variable in a later function as the metric for olympic success. Additionally, the team condition was set as an optional parameter to explore all countries if desired. The result of this function gives a clean output consisting of ‘Sport’ ‘Country’ ‘Medal Type’ and ‘Count’ and will help in determining which countries are successful in which sports and during what year. From this function, I was able to determine that the United States performed exceptionally well in Basketball during the 2004 Olympics, winning 24 medals, across their mens, womens, and youth teams. Given this finding, availability of data, and that sports culture varies by country, I decided to use basketball in the United States as my main metric in determining if Olympic success resulted in increased sports valuation.

The next step was to calculate and isolate the dependent variable– average earnings. Function 3, ‘get_forbes’, is designed to return average earnings for athletes in a particular sport or nationality during a specified year. Average earnings will serve as our dependent variable in this study, representing increased or decreased 'sports valuation in a country'. Unfortunately, the forbes dataset this function pulls from is extremely limited in that it only holds the top 100 paid athletes per year. What this indicates is that our data will be limited to athletes in high paying sports and high paying countries, effectively decreasing the statistical significance of our study by lack of representative data. Despite this, I continued the analysis with caution to study the most popular sports and countries with the largest sports economies (Basketball in the United States). Overall, the average earnings still represents an increase or decreased sports valuation; however, the countries that metric may be relevant to is limited by the dataset.

In order to determine our second dependent variable, function 4, ‘get_country_gdp’ was created to query the world bank api for the gdp of a specified country in a specified year. The function would then run the same query for the following year and calculate the difference

between the averages. This difference would serve as the dependent variable to determine if hosting the olympics or succeeding in the olympics resulted in a significantly increased GDP. When utilizing, the console requests inputs for country and date and then displays the GDP for the date requested, GDP for the following year, and the difference between the two years

With all three functions created to isolate the relevant independent and dependent variables, it was time to put it together in a sequential manner that would allow you to explore how different successful Olympic ventures affected sports valuation and GDP for that country. Unfortunately, the forbes data limited the sports and countries available for study. As an alternative to creating a sequential program that allowed you to study the change in any country and sport I decided to limit the project by manually entering the sports and countries I wanted to explore. Through this method, the project puts more of a focus on the potential correlation between variables than creating a user interface that controls for the limitations of data but still allows you to explore. When I initially decided on this project, I was intending to create a user interface that allows you to explore the relationship while setting your own parameters, I realized very quickly that this is too large a scope for the project and decided to simplify by limiting the country and sport input to Basketball in the United States and manually changing the functions to determine different trends.

Instead of creating a comprehensive user interface, I made function 5, 'earnings_vis' to create a bar graph of average earnings by sport for a particular year to serve as a visualization aid in our data analysis. The function can filter by country and serves to visually showcase the changing average earnings year by year.

So, what did the data show? To start, using the fileter_olympic_data function we determined that during the 2004 Olympics in Athens, Greece The United States performed exceptionally well with 24 medals across all the basketball competitions. I will use this

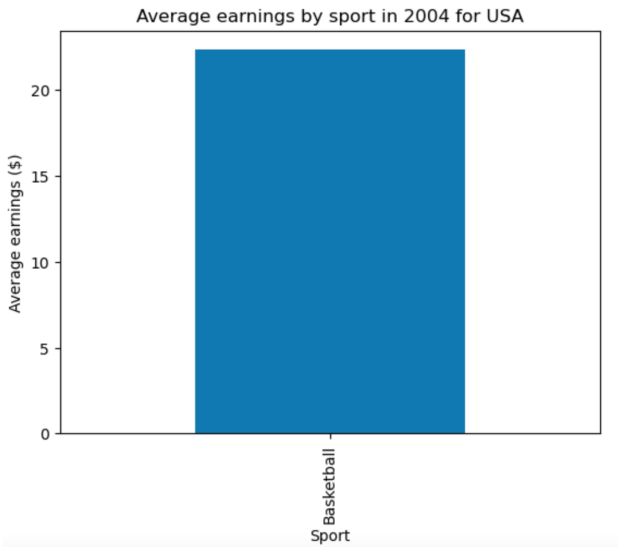
	Medal	Team	Sport	Count
0	Bronze	United States	Basketball	12
1	Gold	United States	Basketball	12
2	Silver	United States	Cycling	2
3	Silver	United States	Tennis	1

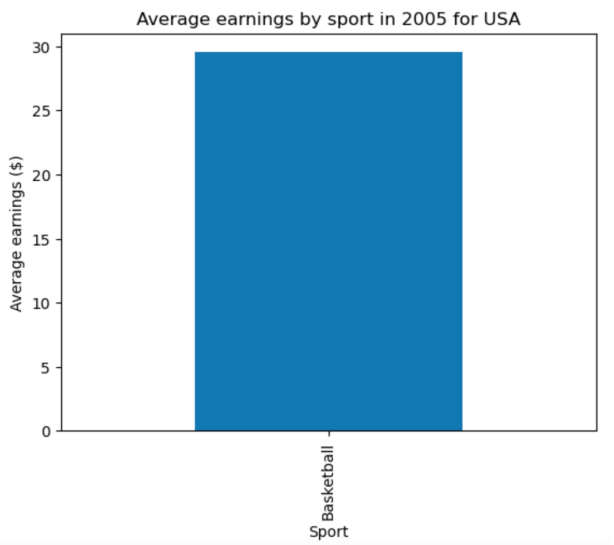
information to determine if average basketball earnings in the United States rose significantly, dropped, or did not change in the years after the Olympic Games in 2004.

Utilizing the 'earnings vis' and 'get_forbes' function I will explore the change in average earnings in the years following the 2004 USA Basketball Olympic success to determine if the success had any effect on the valuation of sports in that country.

```
Enter the year: 2004
Enter the sport: Basketball
Enter the nationality (optional): USA
                 Name Nationality  Earnings  Year       Sport
56       Allen Iverson         USA      19.7  2004  Basketball
90        Andre Miller         USA      15.0  2004  Basketball
103  Anfernee Hardaway         USA      15.1  2004  Basketball
601          Grant Hill         USA      25.9  2004  Basketball
686          Jason Kidd         USA      15.9  2004  Basketball
855       Kevin Garnett         USA      29.7  2004  Basketball
894        Kobe Bryant         USA      26.1  2004  Basketball
940       Lebron James         USA      21.1  2004  Basketball
1093    Michael Jordan         USA      35.0  2004  Basketball
1323   Rasheed Wallace         USA      15.3  2004  Basketball
1468   Shaquille ONeal         USA      31.9  2004  Basketball
1558     Tracy McGrady         USA      19.0  2004  Basketball
1598      Vince Carter         USA      20.2  2004  Basketball
Average earnings: $ 22.300000000000004
```

```
Enter the year: 2005
Enter the sport: Basketball
Enter the nationality (optional): USA
                 Name Nationality  Earnings  Year       Sport
893       Kobe Bryant         USA      28.8  2005  Basketball
939      Lebron James         USA      22.9  2005  Basketball
1092   Michael Jordan         USA      33.0  2005  Basketball
1467  Shaquille ONeal         USA      33.4  2005  Basketball
Average earnings: $ 29.525
```





Through our analysis, we can see that the average salary for a basketball player in the United States rose from $22 million to $29 million in the one year after the USA's successful basketball performance in the 2004 Athens, Greece Olympics. This would indicate that Olympic success does have some correlation to valuation of sports and ultimately answers our main question.

To explore our second question regarding if success or hosting has an impact on gdp, we will use the 'get_country_gdp' function to determine the change in gdp for the United States and Greece, before and after the 2004 olympics.

```
Enter a year: 2004
Enter a country name: Greece
GDP for Greece in 2004: 240963562236.127
GDP for Greece in 2005: 247875422204.414
Difference in GDP between 2004 and 2005: 6911859968
```

```
Enter a year: 2004
Enter a country name: United States
GDP for United States in 2004: 12217193198000
GDP for United States in 2005: 13039199193000
Difference in GDP between 2004 and 2005: 822005995000
```

The GDP function showcases that the GDP of both Greece (host) and United States (success at olympics) increased after their 2004 Olympic run. This would indicate that both success and hosting is somewhat correlated to GDP, although there are many extraneous variables that can also affect GDP.

Before you take the results of this study at face value, let's explore the limitations I encountered when conducting my analysis. First and foremost, the limit of 100 entries in the forbes dataset restricted the sports, countries and overall trends for observation. In addition, trends were determined based on the changing statistics for one singular case (basketball in the United States 2004-2005). A statistically significant study would likely incorporate multiple cases into a regression analysis to determine a genuine correlation pattern. Unfortunately, the dataset and timing factors hinder this project from exploring the data at a more in depth level; however, the project still succeeds in allowing you to explore trends assuming data availability. Beyond limitations of the study, I was unable to create a ui that allows a user to input a year, sport, and

country and receive a comprehensive study of how earnings changed based on their olympic success. Should I continue on this project in the future, creating the comprehensive exploration program with a UI would be a top priority. In addition I would aim to run a regression analysis as opposed to a single test case. Overall, the findings of this analysis present an interesting, but expected, trend that success in the Olympics correlates to increased valuation for that sport in a defined country. The results should be studied further to determine true statistical significance.