1. Introduction

The Olympic Games are a periodic sporting event held every 4 years by a designated country/city. During the Games, the world's best athletes compete in several events spanning a multitude of disciplines with the goal of winning gold, silver, and bronze medals. While winning a medal is the crowning achievement for most athletes, being an Olympian is also considered by many as one of the most important feats in their lifetime. It symbolizes to some athletes the fact that they have reached the top of their discipline and the results of their hard work will be on display for the world to see.

Beyond the individual or team aspect for the athletes, the Games also mean for those with a patriotic sense a chance to represent and compete for their country. Not only important for the athletes but also for the country/nation they represent; the countries keep a count of the medals won by the athletes in their delegation and the counts are compared at the end of the games with the country having the most medals at the end being named top nation.

The end of the second World War somewhat left the world in limbo, the major European countries were decimated and impoverished by the war while the United State and the Soviet Union found themselves at the top in terms of economic and military power. This bipolarization and the resulting tensions led to what we now know as the cold war, during that time those two major superpowers were involved in proxy wars, economical, geopolitical, ideological, and technological conflicts[1,2]. The rivalry and tensions between the two nations were extremely high and the threat of a nuclear war was a possibility. The rivalry also extended to the world of sports and the Olympics Games, the only 2 nations at the top of the medals count post second World War and before the dissolution of the Soviet Union were the United States and the Soviet Union. With 11 Olympic Games held, from 1948 to 1988, they split the top nation ranking with 5 for the United States and 6 for the Soviet Union.

The top nation accolade and the Olympics success being enjoyed and fought over by two of the richest and most powerful nations in the world does not appear illogical. However, sporting competitions are supposed to reward the gifted athletes at the top their disciplines and this aspect should not be overshadowed. Is it possible that just like during the cold war era, the Games can be turned into a competition between the world's wealthiest and most powerful nations? This project investigates and tries to answer this question, challenging the fairness of the competition. The project solely focuses on the post cold war era as the answer is already clear for that period, also the winter games are not included as countries with cold weather are already favored to host and compete in the

competition. For this investigation we examine the top nation for each Olympic Games and look at their Gross Domestic Product (GDP) per capita as compared to the rest of the world to determine if there are any clear advantages with dominating the games and rich country population.

Although not a controversy, if just presented in that light, it would make sense for the richest countries with more money than their counterparts to be able to send larger delegations with more athletes and provide them with the necessary tools and resources to increase their chances during the competition. Additionally, it is important to note that this investigation in no way attempts to discredit the talented athletes of the more successful and richest countries.

2. Methods

For this study we collected 2 sets of data and combined them for the analysis, Olympic game data and GDP per capita data. The following section details how we acquired the data and how we got them ready for the analyses.

a. Olympic Data

We gathered Olympic games data from the TidyTuesday[3] project repository, which contains sets of raw datasets added weekly with the objective of learning data analysis, exploration, and visualization. For the Olympics Games, the data included information about the competition starting from Athens 1896 to Rio 2016. The data was first obtained by sport statistics enthusiasts on [4,5] then using R code the owner of the Kaggle[6] dataset scrapped it to perform some analysis on the games and their history. the variables and description are given in Table 1. First, we filtered the dataset to remove the games prior to 1991, the year of the dissolution of the Soviet Union, and keep the games from the summer season.

Table 1: Olympic Data Dictionary

Variable	Description
Id	Athlete ID
Name	Athlete Name
Sex	Athlete Sex
Age	Athlete Age
Height	Athlete Height in cm
Weight	Athlete weight in kg
Team	Country/Team competing for

NOC (National Olympic Committees)	NOC region
Games	Olympic games name
Year	Year of Olympics
Season	Season either winter or summer
City	City of Olympic host
Sport	Sport
Event	Specific event
Medal	Medal (Gold, Silver, Bronze, or NA)

We worked with a compact dataset that did not include the individual athletes and their personal information, we focused on the countries/teams, the year and removed the non-medal winning rows. To ensure that the data was correct we made sure that the medals won were counted once when tallied, as the original dataset included multiple counts for team sports and events. The final check was done by comparing the medal count of the top countries in the year 1992 with the results from the official Olympic games[7]. A preview of the organized is given in Table 2.

Table 2: A top preview of the Olympic data used

	year	country	medals
1	1992	Unified Team	112
2	1992	United States	108
3	1992	Germany	82
4	1992	China	53
5	1992	Cuba	31
6	1992	Hungary	30

b. GDP per capita Data

We obtained the dataset for the GDP per capita from GAPMINDER[8], a reliable and independent resource for data. Their mission is to combat misconceptions within people thanks to the reliable data they have and put at the disposition of others for teaching and learning. The data contained the countries and their GDP per capita for each year, going back to 1959. For this study we were only interested in the data for the years in which Olympics Games were held, 1992, 1996, 2000, 2004, 2008, 2012 and 2016. After filtering the data for those years, we merged it with the Olympics data based on the names of the countries, we recognized that due to characters used and possible mistakes and

mismatches in the datasets we have the potential to lose some data in that process. One noticeable example is that the Unified Team from the year 1992 was excluded from the remaining data. The Unified Team consisted of several soviet athletes unable to compete under the umbrella of the dissolved union.

For the final dataset, we grouped the games by years, and we have the medal count for each country, their GDP per capita, their rank in the remaining Olympic nation and a scaled value of their GDP per capita as it offers a better way to view the wide data range. The scale used was a log 10 scale, a preview of the final dataset is given in Table 3.

Table 3: A top preview of the final dataset used for analysis

	country	year	medals	gdp_per_capita	o_rank	scaled_gdppc
1	United States	2016	121	58400	1	4.766413
2	China	2016	70	9050	2	3.956649
3	Russia	2016	56	9470	3	3.97635
4	France	2016	42	37700	4	4.576341
5	Germany	2016	42	42600	5	4.62941
6	Japan	2016	41	35900	6	4.555094

3. Results and Discussion

After acquiring a dataset, we believed was appropriate to answer our initial queries we proceeded with the analysis. As we intended, we looked at the plot of the medal count vs the scaled GDP per capita of each Olympic games, and medal counts for each country for all the games in the selected period. The remaining figures give a breakdown of our findings.

From the plots we can observe recurring nations in the top medals' winners, and this can lead to the conclusion that the games are in fact a competition among some of the most powerful nations in the world. There does not seem to be a linear relationship or correlation between the GDPs per capita and winning medals, but we see a clear separation between nations winning several medals every year and the rest. There's also a clear dominance of the United States throughout the years and a climb from China in both medals count and GDP per capita over the years. Additionally, the total medal count plot highlights the clear dominance and the recurring nations.

We can say that the results partially answer the questions we were asking ourselves at the beginning of the study. However, we still believe that although the GDP per capita is an important

metric it does not allow us to fully qualify the nations in terms of power and riches. To fully perform this study several additional metrics would have to be considered to properly rank the nations. The method used only offers a snapshot of the countries' economics, which we know are highly volatile and dynamics complex systems. Our current code removes the nation not winning any medals and adding them can have a significant impact on our current conclusions

Additionally, other factors specific to the games need to be considered, as an example we have purposefully excluded the winter games as some nations were already favored due to their climates, similarly some nations have advantages regarding certain disciplines or are prioritizing specific disciplines to be collect the maximum number of medals during the games.

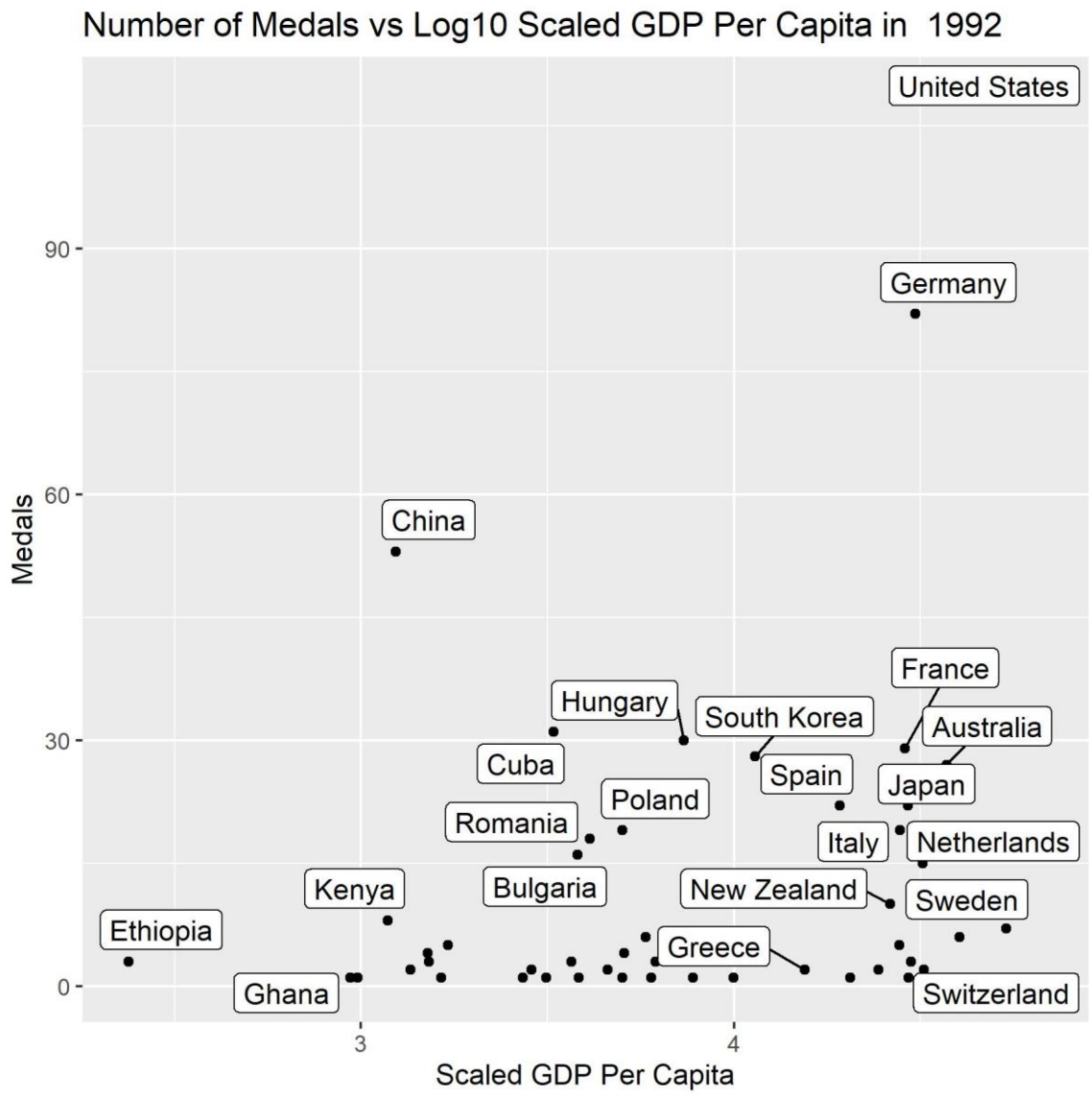

Figure 1: Number of Medals vs Scaled GDP Per Capita 1992

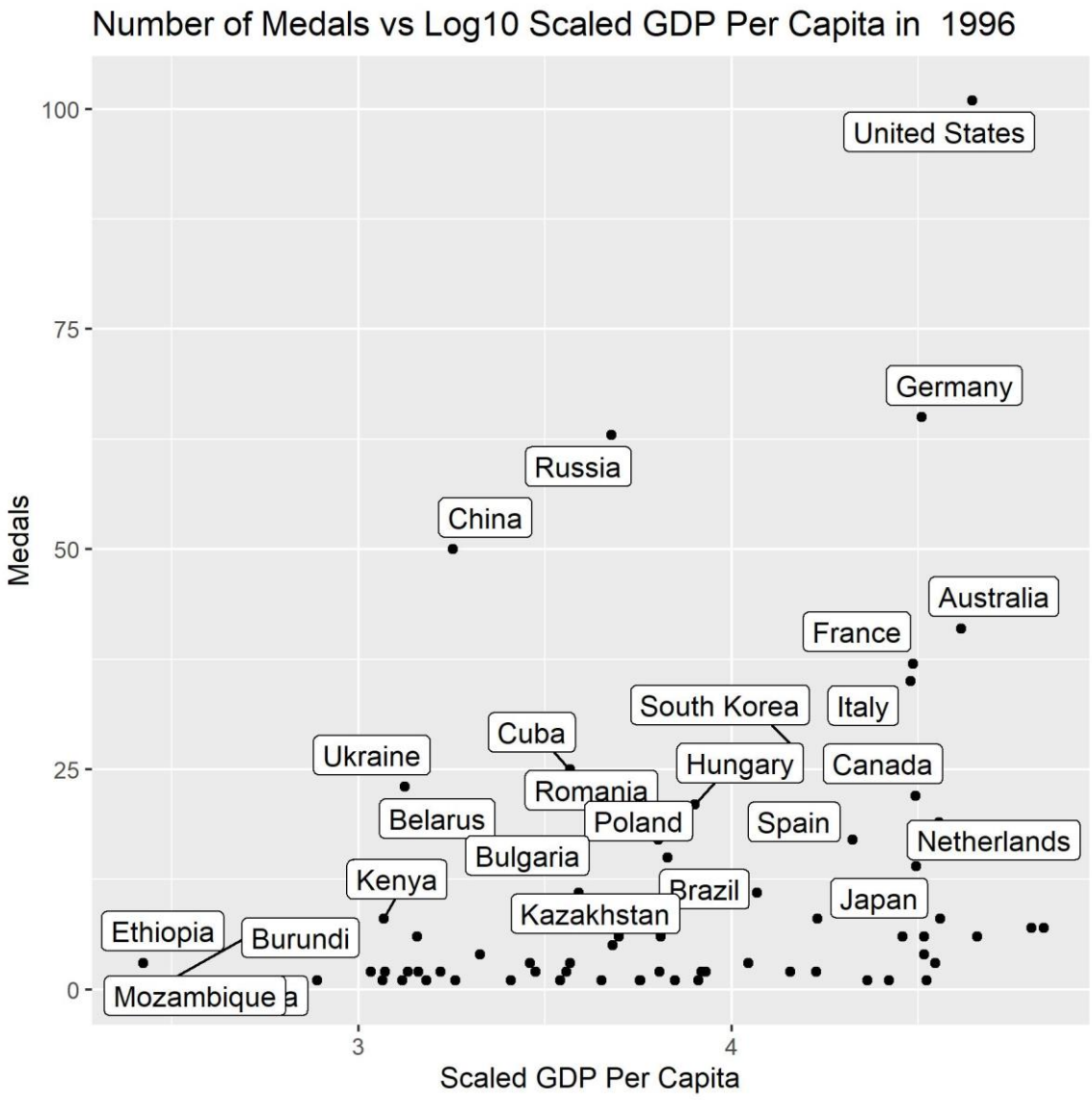

Figure 2: Number of Medals vs Scaled GDP Per Capita 1996

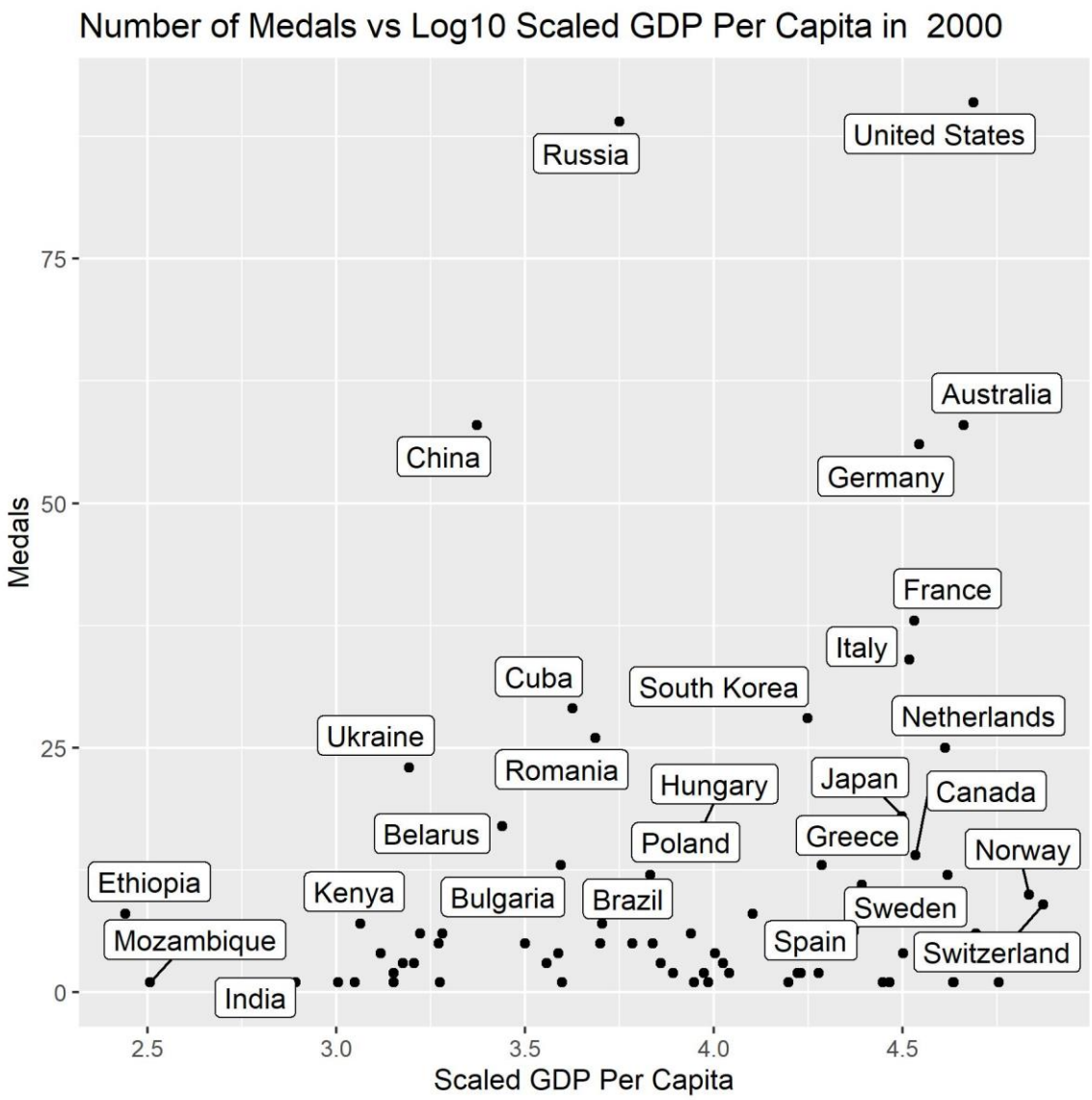

Figure 3: Number of Medals vs Scaled GDP Per Capita 2000

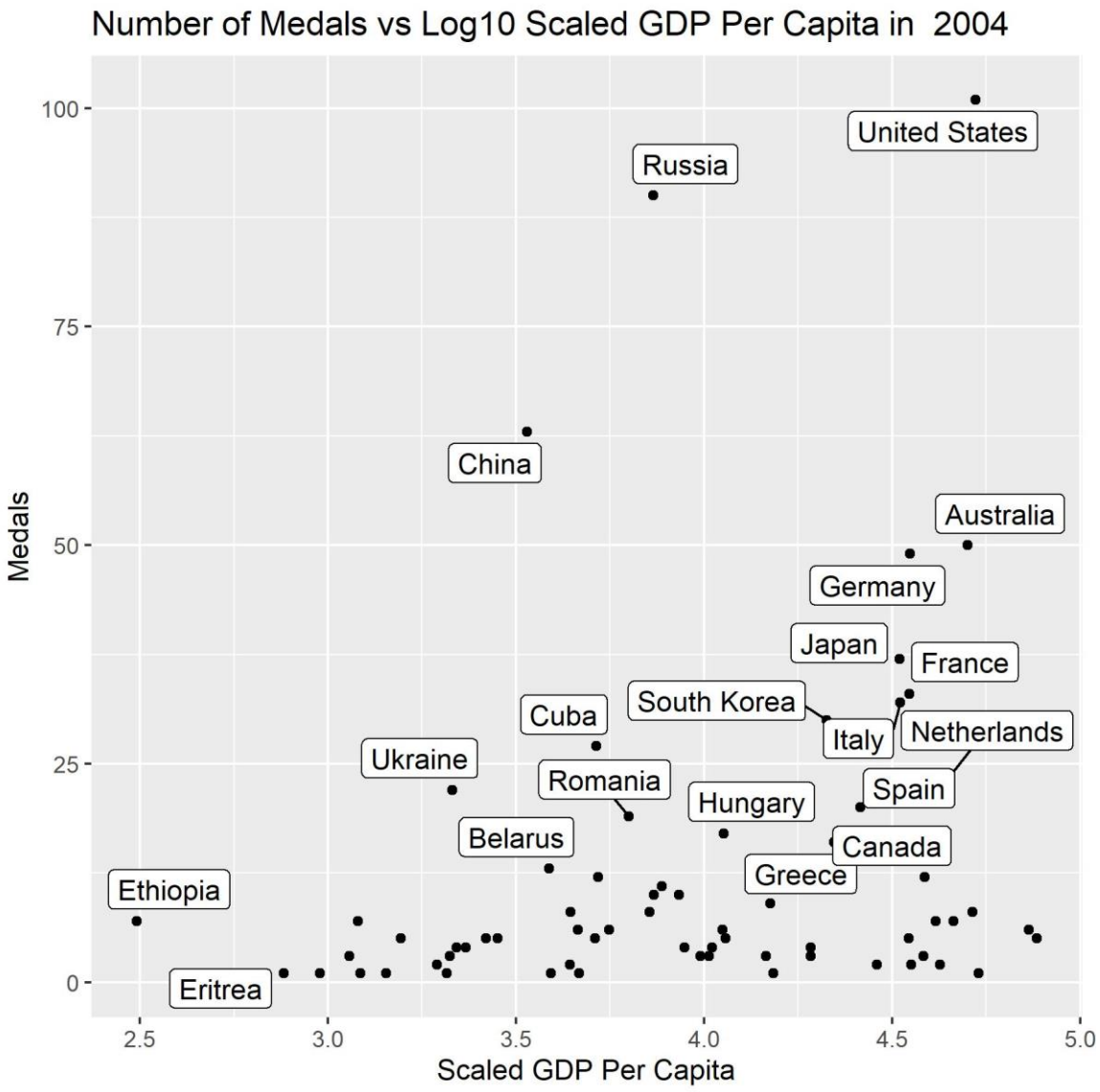

Figure 4: Number of Medals vs Scaled GDP Per Capita 2004

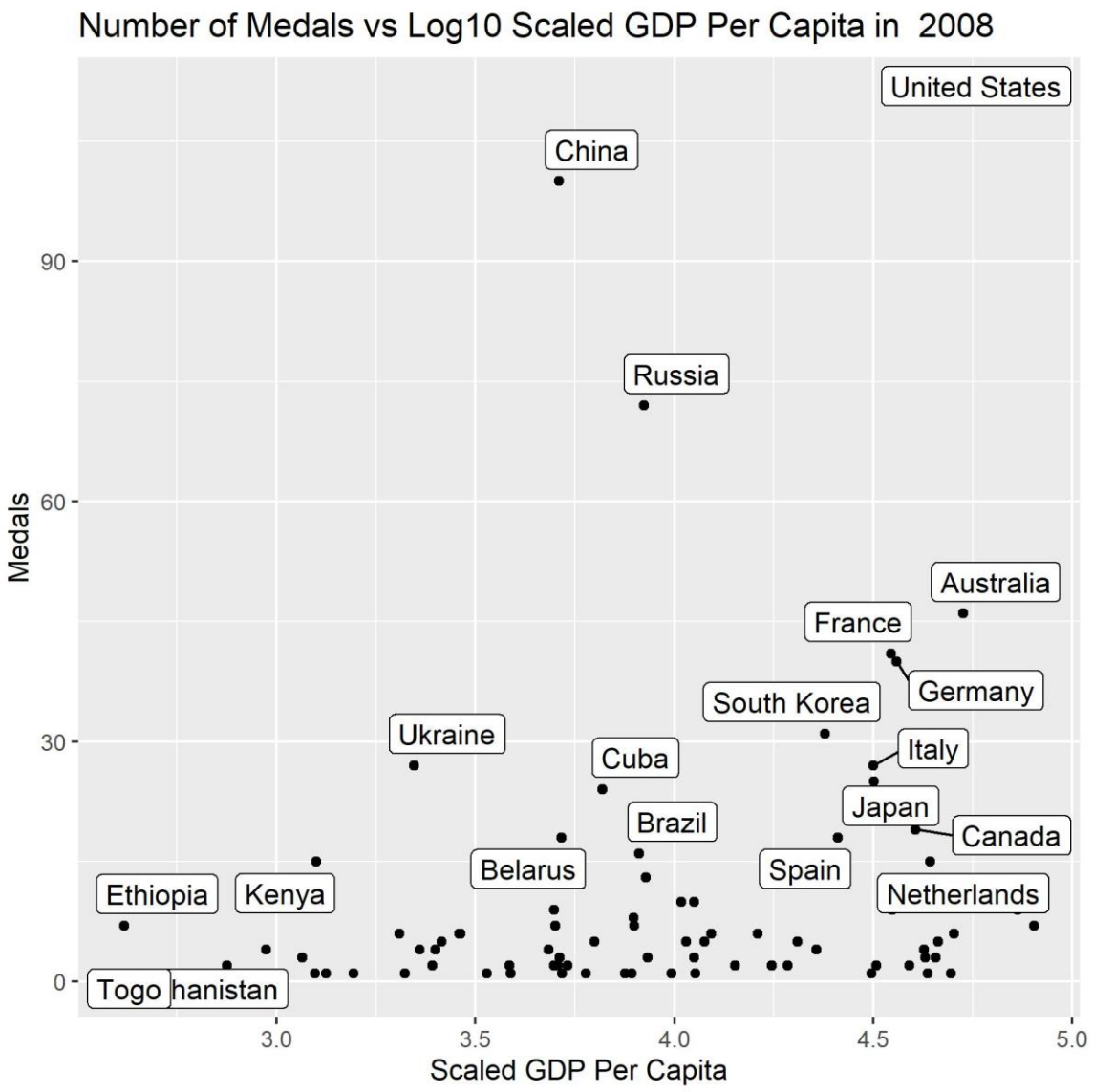

Figure 5: Number of Medals vs Scaled GDP Per Capita 2008

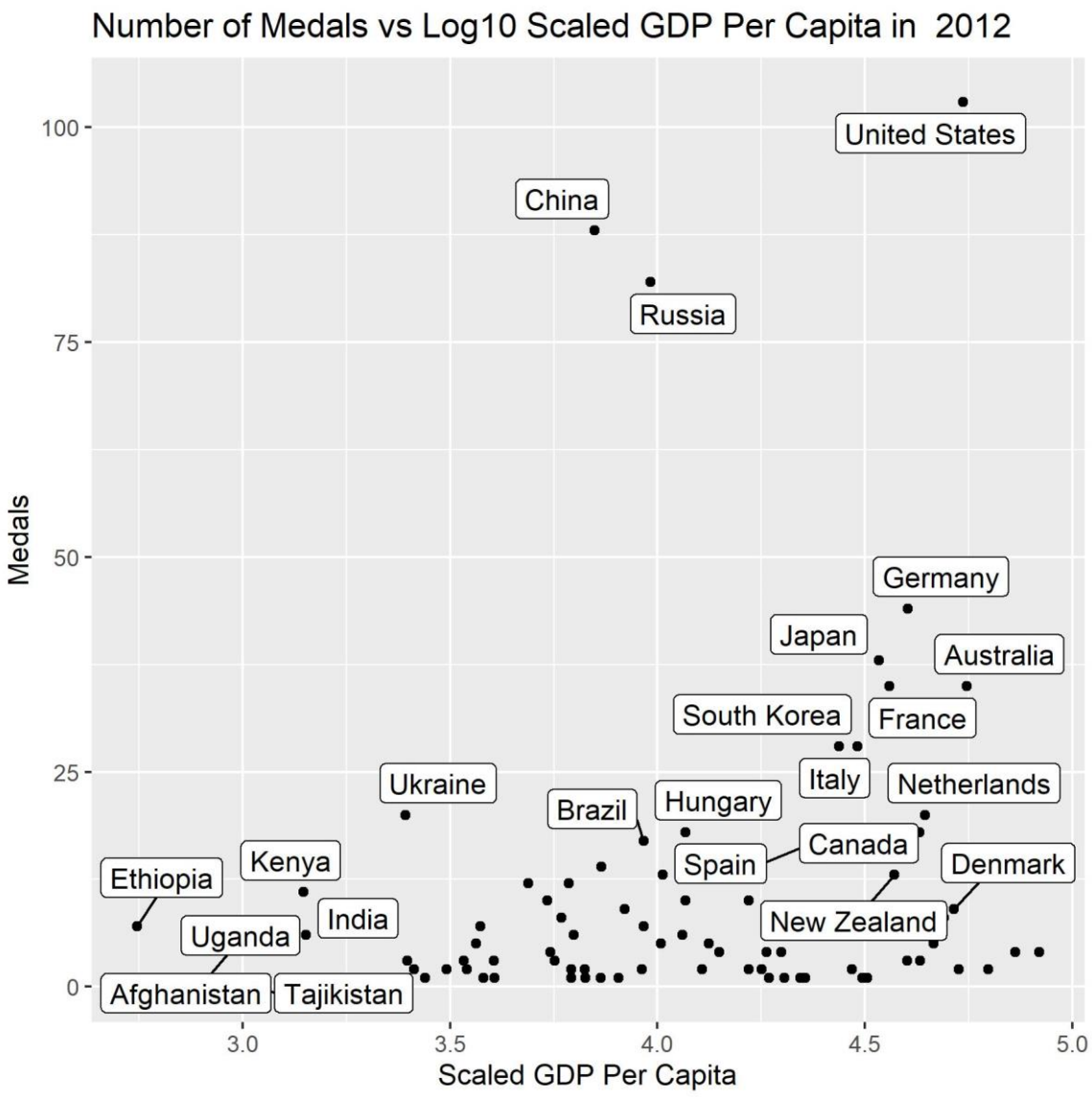

Figure 6: Number of Medals vs Scaled GDP Per Capita 2012

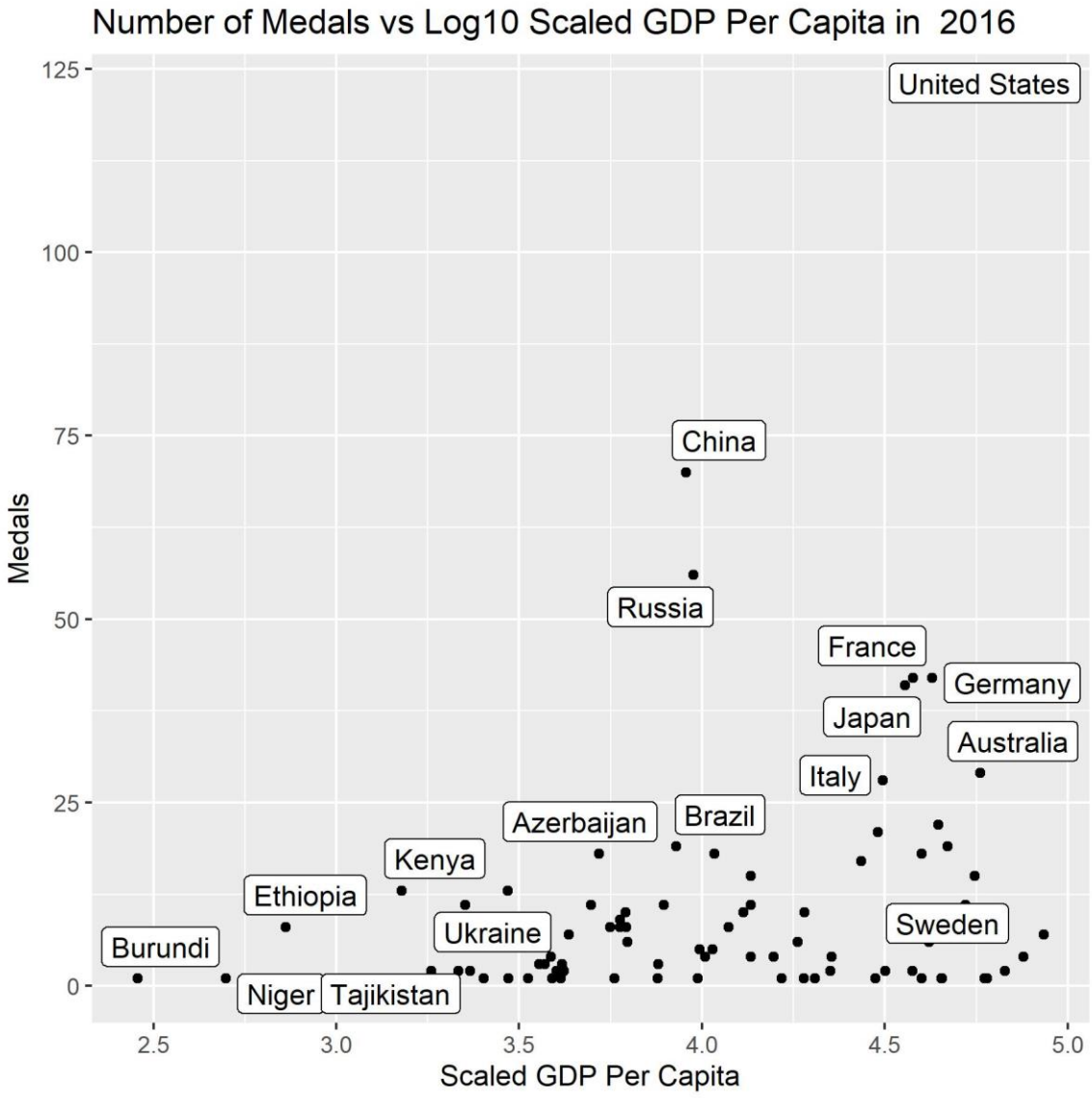

Figure 7: Number of Medals vs Scaled GDP Per Capita 201

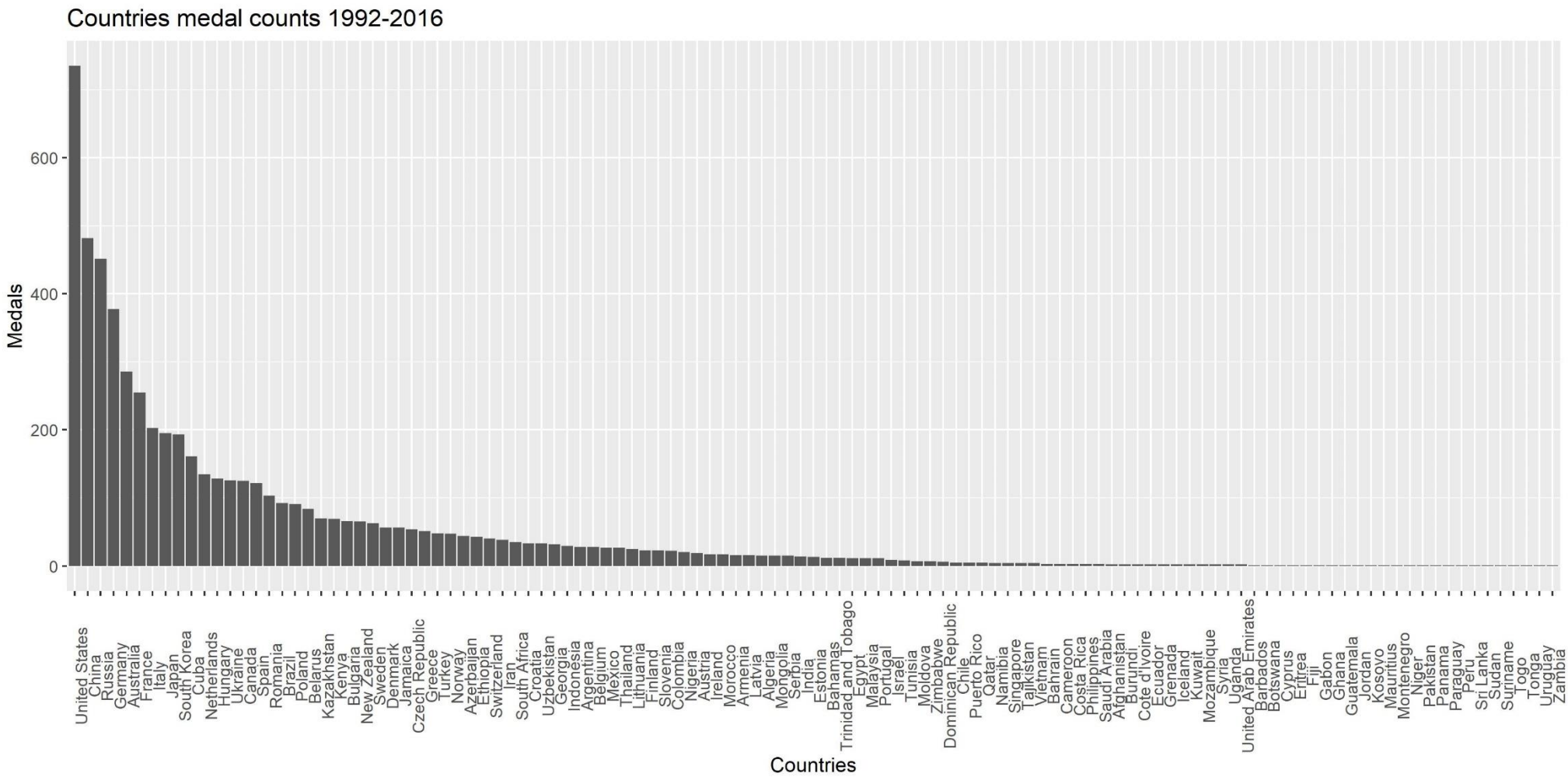

Figure 8: Medals count for countries 1992-2016

References

1. Rapkin, D. P., Thompson, W. R., & Christopherson, J. A. (1979). Bipolarity and Bipolarization in the Cold War Era: Conceptualization, Measurement, and Validation. The Journal of Conflict Resolution, 23(2), 261–295. http://www.jstor.org/stable/173832
2. Towards a bipolar world (1945–1953) - The Cold War (1945–1989). Towards a bipolar world (1945–1953) - The Cold War (1945–1989) - CVCE website. (n.d.). Retrieved August 4, 2022, from https://www.cvce.eu/en/education/unit-content/-/unit/55c09dcc-a9f2-45e9-b240-eaef64452cae/1dc7e103-8078-45e1-b8ac-2199a9be5783
3. Rfordatascience. (n.d.). Tidytuesday/data/2021/2021-07-27 at master · rfordatascience/tidytuesday. GitHub. Retrieved August 4, 2022, from https://github.com/rfordatascience/tidytuesday/tree/master/data/2021/2021-07-27
4. Sports reference: Sports stats, fast, easy, and up-to-date. Sports. (n.d.). Retrieved August 12, 2022, from https://www.sports-reference.com/
5. OlympStats. Retrieved August 12, 2022, from https://olympstats.com/
6. Rgriffin. (2018, June 15). 120 years of olympic history: Athletes and results. Kaggle. Retrieved August 12, 2022, from https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results?select=noc_regions.csv
7. Ioc. (n.d.). Barcelona 1992 olympic medal table - gold, silver & bronze. Olympics.com. Retrieved August 4, 2022, from https://olympics.com/en/olympic-games/barcelona-1992/medals
8. Gapmider Data. Gapminder. (n.d.). Retrieved August 4, 2022, from https://www.gapminder.org/data/