**CS 5830 Project 3: Doping in the Olympics**

Joshua Hales and Stockton Smith

## Introduction

For this project, we analyzed relationships among Olympic events, doping, and medal tallies to provide information to sports enthusiasts on the integrity of the Olympic Games. The International Olympic Committee (IOC) banned doping in 1967 and started testing in 1968 (IOC, n.d.). A doping instance is when the IOC caught a team member violating anti-doping rules and provided a penalty, typically resulting in stripping any illegitimate medals and records. Team members were primarily athletes, but sometimes included others (e.g. Otto Trefny as medical staff). Penalties primarily took the form of disqualification after performing in events or being banned from even competing in them. Team members caught without sanctions were not counted. Further information on our findings is given in our slide presentation and in our GitHub repository.

## Dataset

We scraped data from Wikipedia pages on the Olympic Games. Doping instances came from the page “Doping at the Olympic Games” and include data from 1968-2022. Total medal counts per country came from the page “All-time Olympic Games medal table” and include data from 1896-2022. We iterated over rows of relevant tables to collect data, and we split doping data between summer and winter events. Country names came from the displayed text of hyperlinks. Data split across names did not result in double counting.[1]

## Analysis Techniques

Bar charts showed rates and counts for country categories. A paired t-test for average doping instances by country in the Summer Olympics and Winter Olympics was appropriate because each country had data in both categories. Scatterplots, linear regression, and Pearson correlations were appropriate in showing relationships and trends of a population, although it is important to note that some of the data is likely influenced by size differences among countries.

## Results

We first considered potential differences in doping rates between the Summer and Winter Olympics. Among all countries caught doping in each season of games at least once, most had more violations per Olympics in the summer (Figure 1). Some, like the United States, had no instances in the winter. A paired t-test comparing average doping instances by country for summer and winter gave $t\_stat=2.332$ and $p\_value=0.033$. The $p\_value<0.05$ suggests that athletes dope more often in the Summer Games than in the Winter Games.

[1] For example, Russian athletes competed under names including the Soviet Union, Russia, and Olympic Athletes from Russia.

**Figure 1**

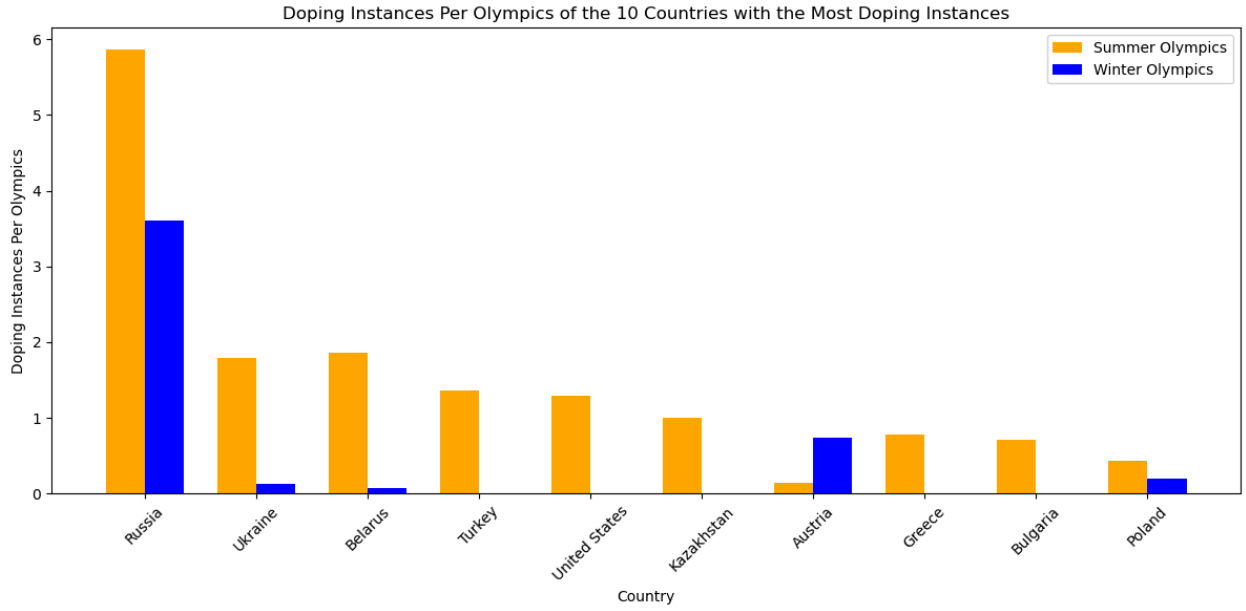
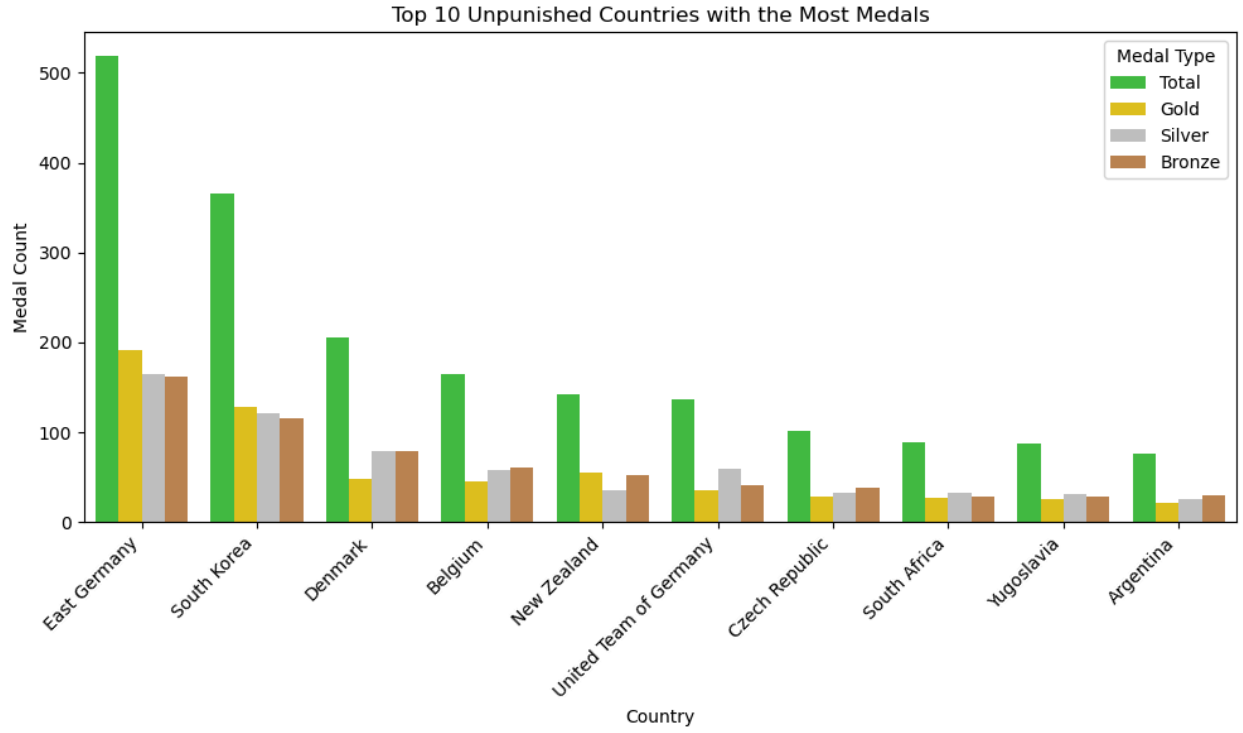



**Figure 2**




We next considered the countries with no doping instances. From those, we show the 10 with the highest medal totals (Figure 2). East Germany is notable for the number of medals earned over 20 years (1968-1988). However, later investigations showed that East Germany had a massive state-sponsored doping program, the details of which were kept hidden even from the athletes themselves. Despite this, the IOC refused to revisit past records (Franke & Berendonk, 1997).

**Figure 3**

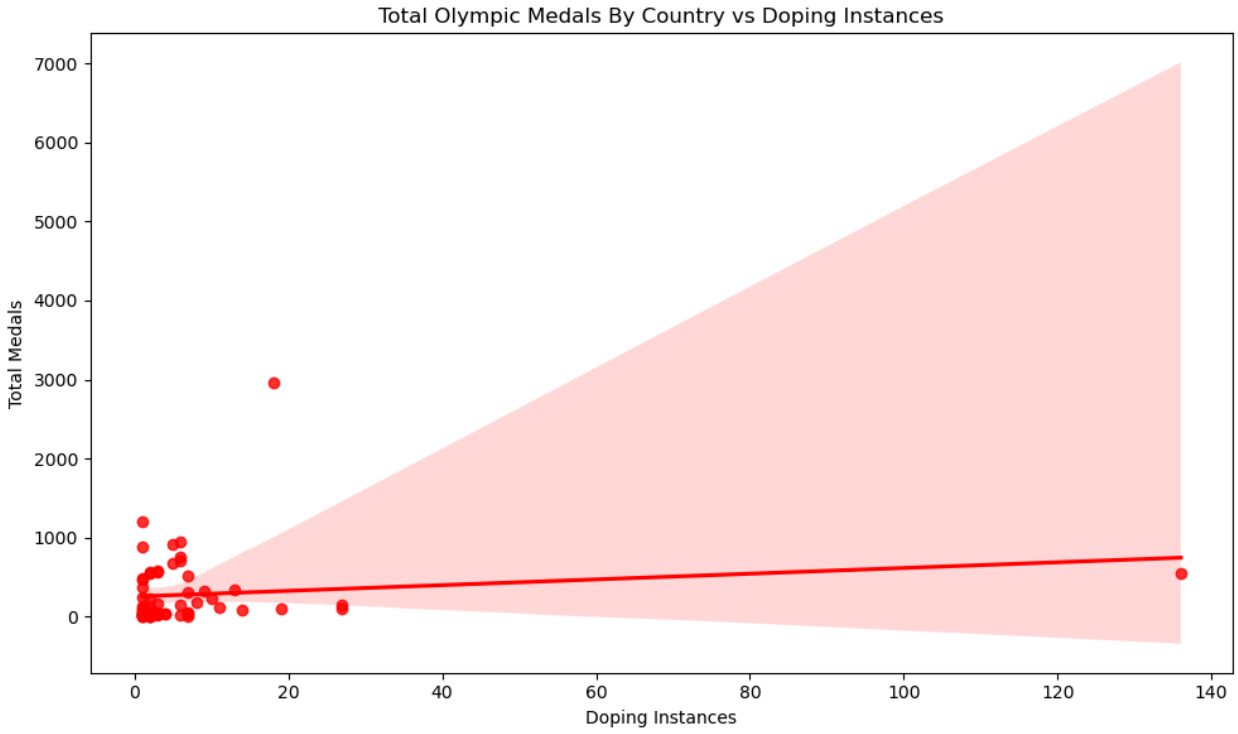
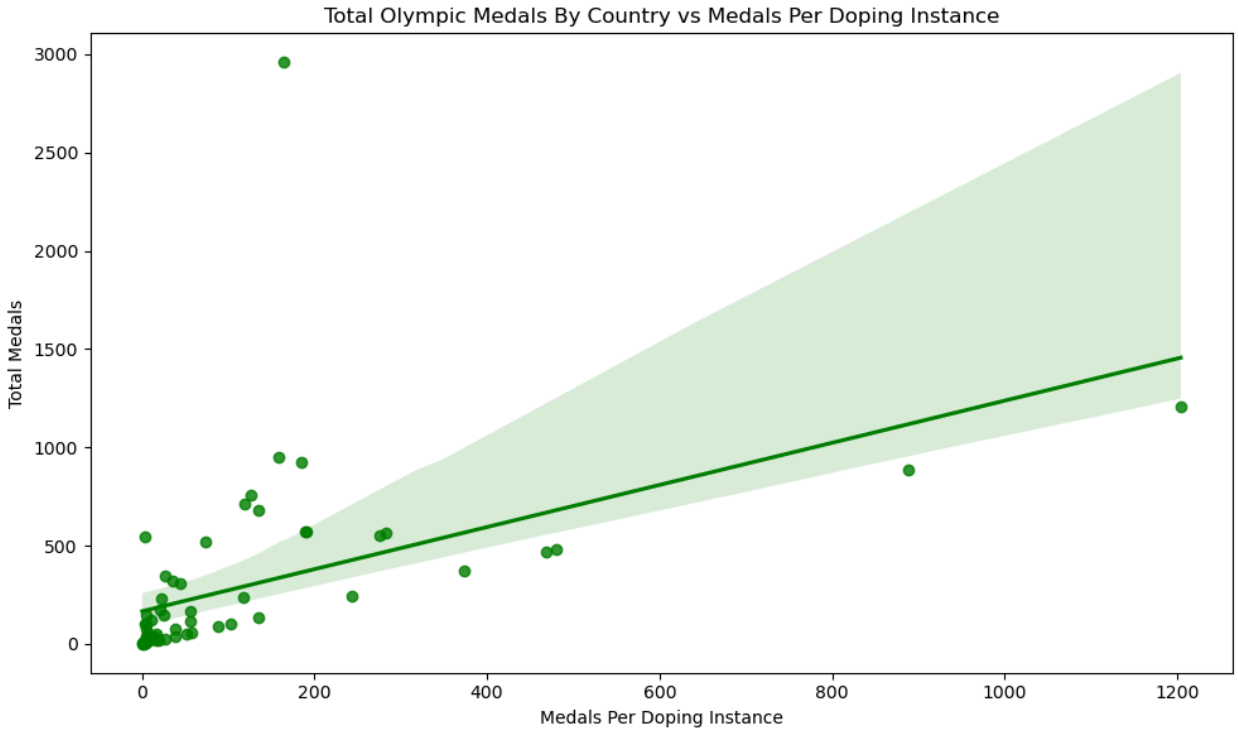



**Figure 4**



We last considered correlations for doping instances and medals among all countries. We found no correlation for the total number of doping instances and the total number of medals earned by that country (Figure 3). However, this data does not and cannot account for the number of dopers who were never caught or punished. We found a statistically significant positive correlation for the number of medals earned by a country per doper of that country and the total number of medals earned by that country (Figure 4), which more closely ties medal counts to doping. It seems to be the case that each of the two figures offers a different perspective of the same findings.

Overall, our analysis shows the prevalence of doping in the Olympics, as well as challenges in addressing it. We found clusters in the number of instances by country, and we also found that widespread doping, including instances resulting from state-sponsored programs, did not always result in sanctions or punishments; one reason for this may be that scandals that were discovered decades after occurring were likely to be treated much less harshly. Drug testing also had varying success, which could have influenced our findings. Ultimately, our analyses showed that drawing general conclusions on the integrity of the Olympic Games is actually more difficult than expected; this is because data collected on the topic may or may not often be reliable.

## Technical Section

We used tables from Wikipedia because we could not use requests on the official Olympics site. Tables required manual indexing to categorize since they did not have identifying labels. Bar charts helped compare numeric variables with categorical variables, such as country and season. In our figures, we arbitrarily showed the top 10 countries in order to reduce clutter and make it easier to visually parse our findings. The t-test was limited because not all countries participated in the Winter Olympics, and some had no doping instances. The scatterplot, linear regression, and Pearson correlation were appropriate for Figure 3 since they showed relationships and trends over 2 variables and multiple data points. Similar data in Figure 4 shows stronger correlation partly because medals are included in both variables.

If we categorized data by country, Olympic game, and number of participating athletes, we could have performed a more robust analysis. This would have allowed filtering medal totals for those earned during the doping ban, but we could not easily obtain or organize data in this way. Additionally, our analyses are sensitive to the changes and accuracy of the Wikipedia pages; pinning our analyses to page revision identifiers would allow consistent reproduction of the data used.

## References

*All-time Olympic Games medal table*. (n.d.). Wikipedia. Retrieved February 13, 2024, from https://en.wikipedia.org/wiki/All-time_Olympic_Games_medal_table

*Doping at the Olympic Games*. (n.d.). Wikipedia. Retrieved February 13, 2024, from https://en.wikipedia.org/wiki/Doping_at_the_Olympic_Games

Franke, W. W., & Berendonk, B. (1997, July 1). Hormonal doping and androgenization of athletes: a secret program of the German Democratic Republic government. *Clinical Chemistry*, *43*(7), 1262-1279. https://doi.org/10.1093/clinchem/43.7.1262

IOC. (n.d.). *1967: Creation of the IOC Medical Commission*. Olympics.com. Retrieved February 13, 2024, from https://olympics.com/ioc/1967-creation-of-the-ioc-medical-commission