

Olympics2022

Proposal:

Introduction: For our final project, we have decided to investigate medal counts, specifically comparing First World vs Third World countries. This question will allow us to investigate how funding affects performance at the Olympics. For First World countries such as the US and Australia, there is more disposable income that can be attributed to athletics. This funding can help hire better coaches, build better facilities and equipment. Our data set is from Kaggle which is a website where data scientists upload data sets for public use. Our data set includes 2800 athletes with 15 disciplines, and 91 nations participating in the Winter Olympics 2022. The dataset was updated every 2-3 days with the new medals awarded and was finally updated when the Olympics ended on February 20th. There are 2892 rows and 3 columns.

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5      v purrr 0.3.4
## v tibble 3.1.4       v dplyr 1.0.7
## v tidyr 1.1.3        v stringr 1.4.0
## v readr 2.0.1        v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

glimpse(Olympics2022data) Data Analysis Plan:

Our research question that we decided on was looking at medal counts vs the economic status of the country the athlete is representing. We want to see if there is a correlation between First World countries and a higher number of medals vs a Third World country. Our outcome variable(Y) is the correlation value and our predictor variable(X) is the medal counts. The countries that we will include as First World will be USA, Australia, Canada, New Zealand and Japan. For our Third World countries, we will include Haiti, South Sudan, Afghanistan, Madagascar, and Sierra Leone. We will start with 5 countries in each category and will add more if we think that data analysis calls for it.