

# main

*Harrison Burke*

*12/3/2019*

## **The research question(s)**

What is the relationship between rowing success and economic success?

## **Background/significance of the research**

My goal in this project was to take a deeper dive into factors that influence rowing success at the Olympics. The stereotype of the sport is that it's elitist and only for the wealthy, so I made a point to look at the relationship between wealth metrics and rowing results. Furthermore, there is a surprising lack of interesting rowing data on the web, so even exploring within the dataset will uncover some interesting stories.

## **The methods used to obtain and analyze the data**

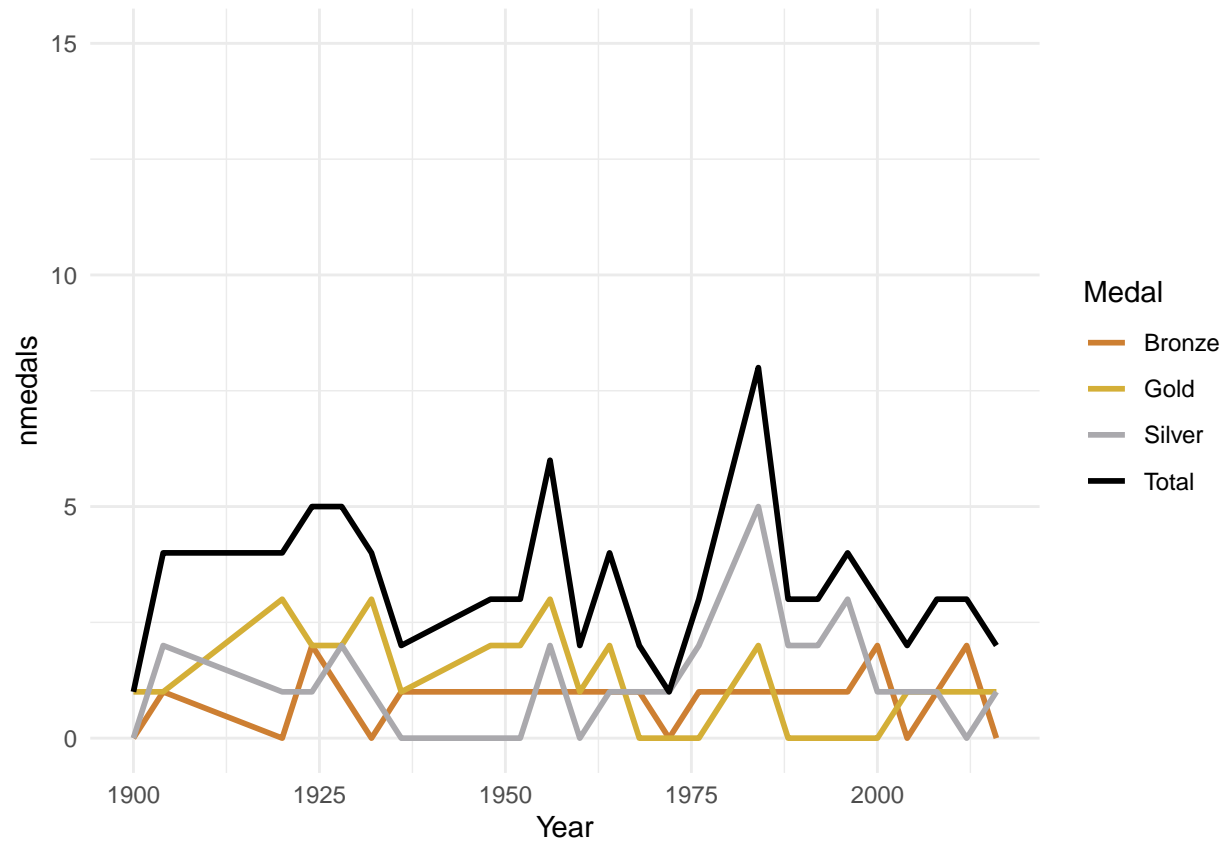
I completed the project entirely in R. I downloaded datasets from Kaggle and the World Banks, and mainly used the tidyverse library to clean, transform, and merge the data. Once aggregated, I used many different libraries to help me manipulate the data to get the information I wanted. A list of libraries that I used will be attached here:

shiny plotly shinyWidgets fs shinythemes moderndive gt tidyverse countrycode

I performed standard linear regression with GDP Per Capita as the explanatory variable, and the medal count at the olympics for the dependent variable. I visualized this relationship with a scatter plot and a regression line, and I printed the output of the model to a table.

## **A discussion of the research, the limitations of the current research, reasonableness of any assumptions made, possibilities of future work/studies that should be conducted, etc.**

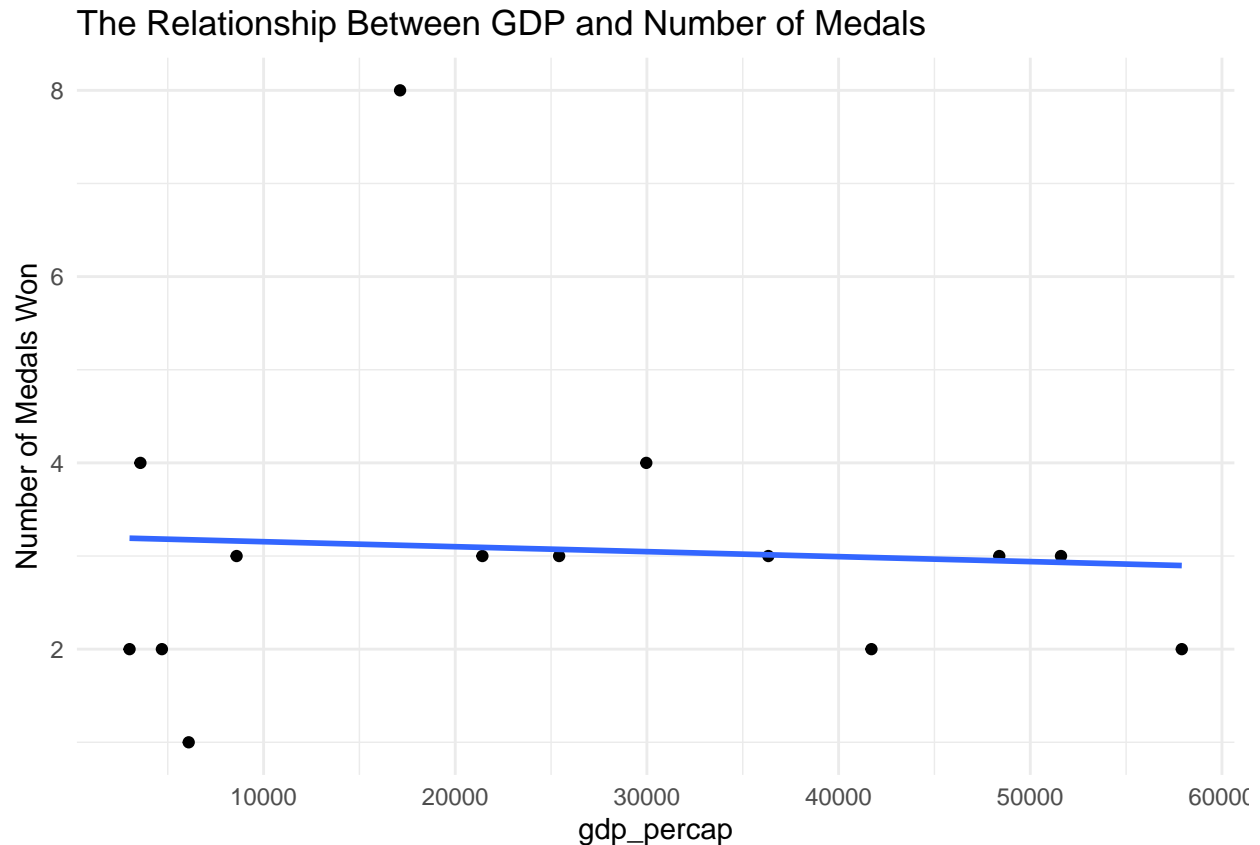
The research was fascinating, as I had not explored data of this type before. I was somewhat surprised in the variability of the model between countries, but considering how many factors there are to explore, I also knew that there was no way it was going to be a perfect fit. I was focused on the data that I had, but there are many more factors that I would like to explore. For instance, if the percentage of a country that is water has any effect, the number of rowing clubs, height, weight, skin color. The more factors that can be explored, the better the understanding will be from a meta-perspective.



The results of the analysis (tables, charts, graphs, significance, confidence intervals, descriptive text)

Here, I am only looking at the data for the United States. As is clear in the graph, there is not a strong relationship between GDP per capita and the number of medals won at the olympics.

```
## Warning: NAs introduced by coercion
```



The table more clearly describes the model. The model predicts that for the United States, when GDP is 0, the number of medals that will be won will be around 3. The table is helped by the graph. The project allows users to manipulate the country that they are looking at, so the data ostensibly varies from country to country. Some countries have a strong positive correlation, others have a strong negative correlation, and many, like the US, have little to no correlation. I expected more consistency across countries, but there are obviously many more factors at play than what is accounted for by GDP Per Capita.

```
## Warning: NAs introduced by coercion
```

```
## Adding missing grouping variables: `total_pop`
```

Country GDP Per Capita and Number of Medals Won  
The relationship between GDP per capita and medal count varies by country

	Coefficient	5th percentile	95th percentile
Intercept	3.207	1.526	4.888
Height	−0.000	−0.000	0.000

Most Medals per Million people at the 2016 Olympics

```
## Warning: NAs introduced by coercion
```

Country	# of Medals per Million People
Estonia	0.759600784
Lithuania	0.696306581
New Zealand	0.639290813
Croatia	0.479547307
Norway	0.382195134
Denmark	0.348972050
Ireland	0.209507667
Netherlands	0.176279708
Australia	0.124341204
Switzerland	0.119444373
Czechia	0.094682328
United Kingdom	0.076176269
Poland	0.052703678
Romania	0.050747766
Germany	0.036289876
Italy	0.033002979
France	0.029897105
Canada	0.027558515
South Africa	0.017886251
United States	0.006189507
China	0.001450679
Algeria	0.000000000
Angola	0.000000000
Argentina	0.000000000
Austria	0.000000000
Azerbaijan	0.000000000
Bahamas	0.000000000
Belarus	0.000000000
Belgium	0.000000000
Bermuda	0.000000000
Brazil	0.000000000
Bulgaria	0.000000000
Chile	0.000000000
Cuba	0.000000000
Ecuador	0.000000000
Egypt	0.000000000
Greece	0.000000000
Hong Kong SAR China	0.000000000
Hungary	0.000000000
India	0.000000000
Indonesia	0.000000000
Iran	0.000000000
Iraq	0.000000000
Japan	0.000000000
Kazakhstan	0.000000000
Libya	0.000000000
Mexico	0.000000000
Nigeria	0.000000000
Paraguay	0.000000000
Peru	0.000000000
Russia	0.000000000
Serbia	0.000000000

Singapore	0.000000000
South Korea	0.000000000
Spain	0.000000000
Sweden	0.000000000
Thailand	0.000000000
Togo	0.000000000
Trinidad & Tobago	0.000000000
Tunisia	0.000000000
Turkey	0.000000000
Ukraine	0.000000000
Uruguay	0.000000000
Uzbekistan	0.000000000
Vanuatu	0.000000000
Venezuela	0.000000000
Vietnam	0.000000000
Zimbabwe	0.000000000

---

## References

<https://raw.githubusercontent.com/datasets/population/master/data/population.csv>  
<https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results>