

Research Presentation

Analysis of the 2016 Rio de Janeiro Summer Olympic Games



Jeff Thomson
IST 652

DATA QUESTIONS:

1) Is there a relationship between the most medals won and:

Country

Age group

2) Does a country's population or GDP per capita influence total medals they will win?

The initial hypothesis is that countries with higher GDP per capita and population won the most medals. Correlation of 0.5 or greater

Data Importation

Two files:

1) Athletes.csv: Twelve columns and 11,539 rows

	id	name	nationality	sex	dob	height	weight	sport	gold	silver	bronze	age
0	736041664	A Jesus Garcia	ESP	male	10/17/1969	1.72	64.0	athletics	0	0	0	47
1	532037425	A Lam Shin	KOR	female	9/23/1986	1.68	56.0	fencing	0	0	0	30
2	435962603	Aaron Brown	CAN	male	5/27/1992	1.98	79.0	athletics	0	0	1	24
3	521041435	Aaron Cook	MDA	male	1/2/1991	1.83	80.0	taekwondo	0	0	0	26
4	33922579	Aaron Gate	NZL	male	11/26/1990	1.81	71.0	cycling	0	0	0	26

2) Countries.csv: List of every country, its three-letter abbreviation, population and GDP per Capita

country	code	population	gdp_per_capita
Afghanistan	AFG	32526562.0	594.323081
Albania	ALB	2889167.0	3945.217582
Algeria	ALG	39666519.0	4206.031232

EXPLORATORY ANALYSIS

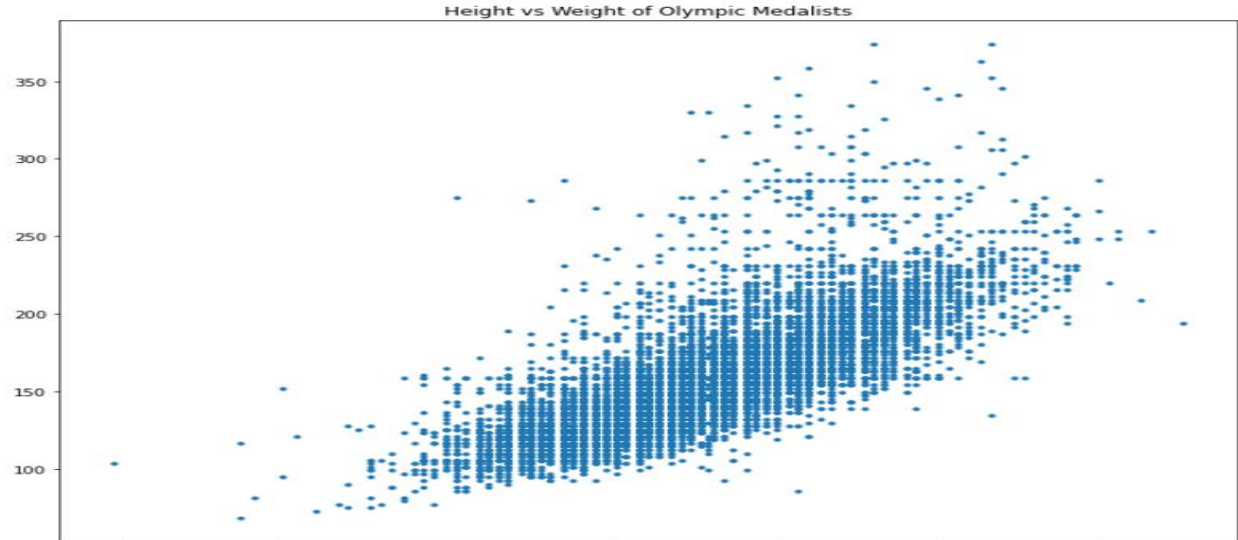
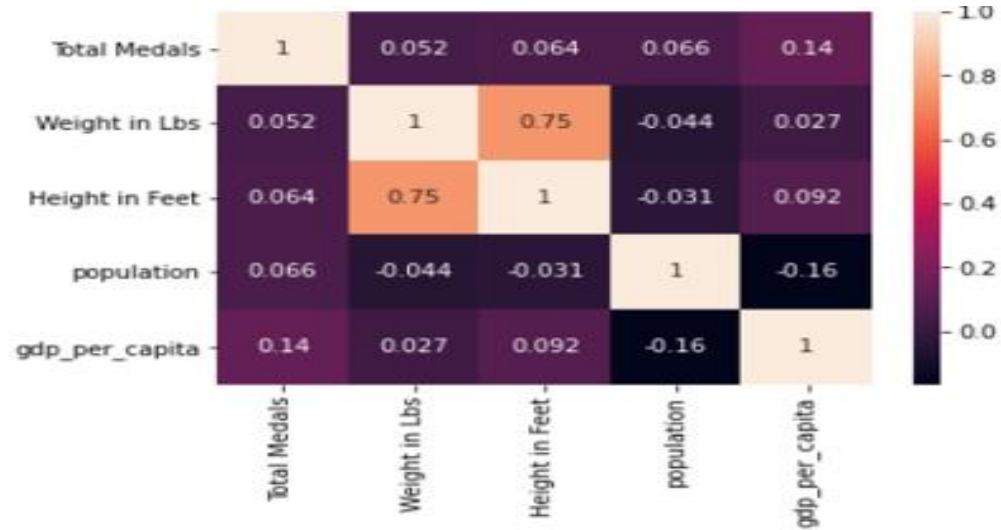


Data Importation and Clean-Up

- Removing/update NA's:
 - Deleted one NA from DOB column.
 - 989 NAs between weight and height columns. Replaced with that columns mean value.
- Checking data types: Luckily, these were already in the proper format
- Converting weight to lbs, height to feet
- Add a total medals column
- Joining additional data sets
 - Renamed athletes "nationality" column to "code"
 - Used Merge on Left Function to join "code" columns
 - Country data including GDP and population
- Remove redundant columns and round GDP to two decimals

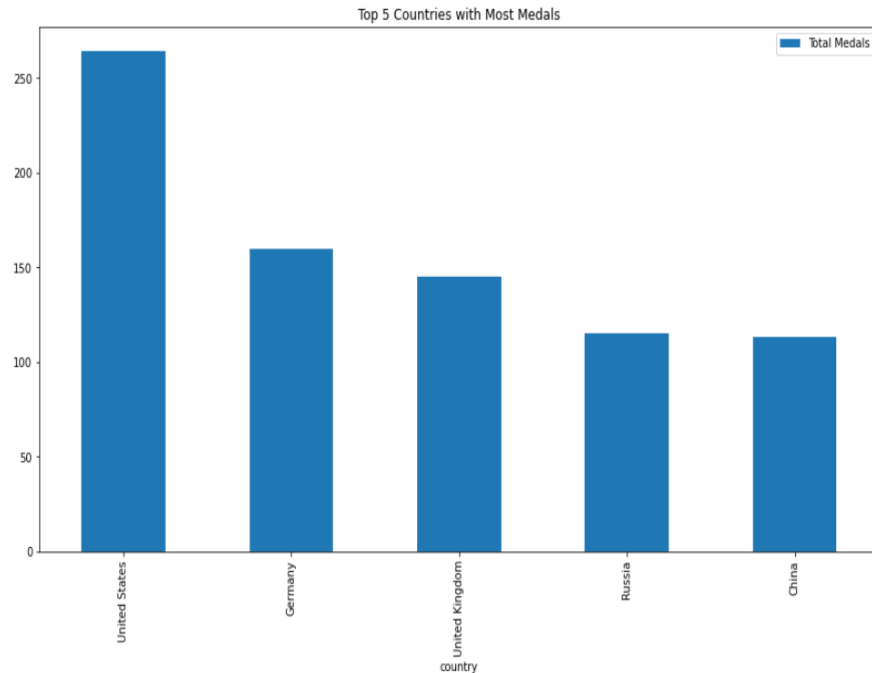
Correlation ANALYSIS

The correlation matrix shows the highest correlation between height and weight. Interestingly, total medals and population or GDP per capita are not highly correlated

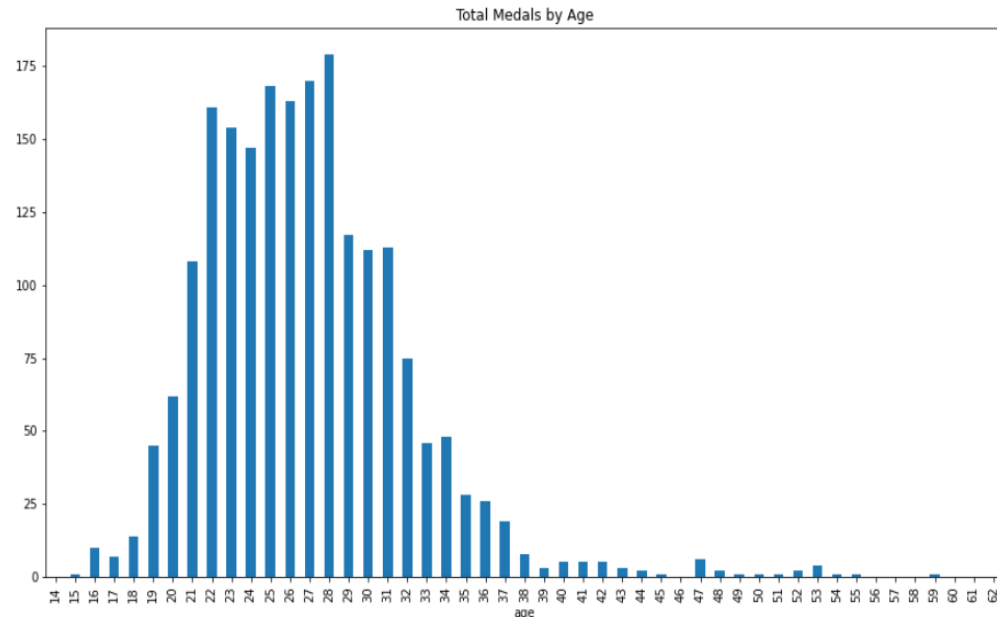


Which country won the most medals, which age group?

Utilized the groupby and sort functions to organize the data to help answer the research questions

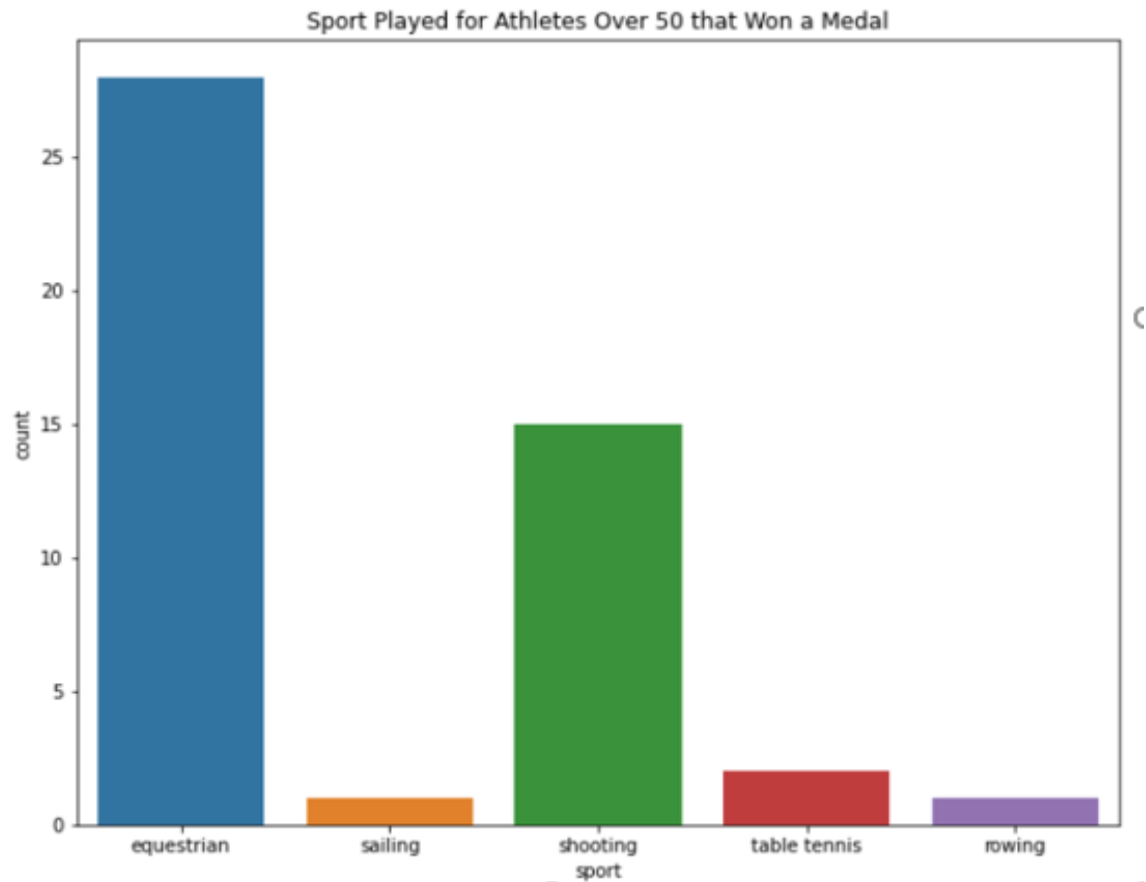


USA, Germany and UK are probably higher on the GDP per Capita list while China has a higher population.



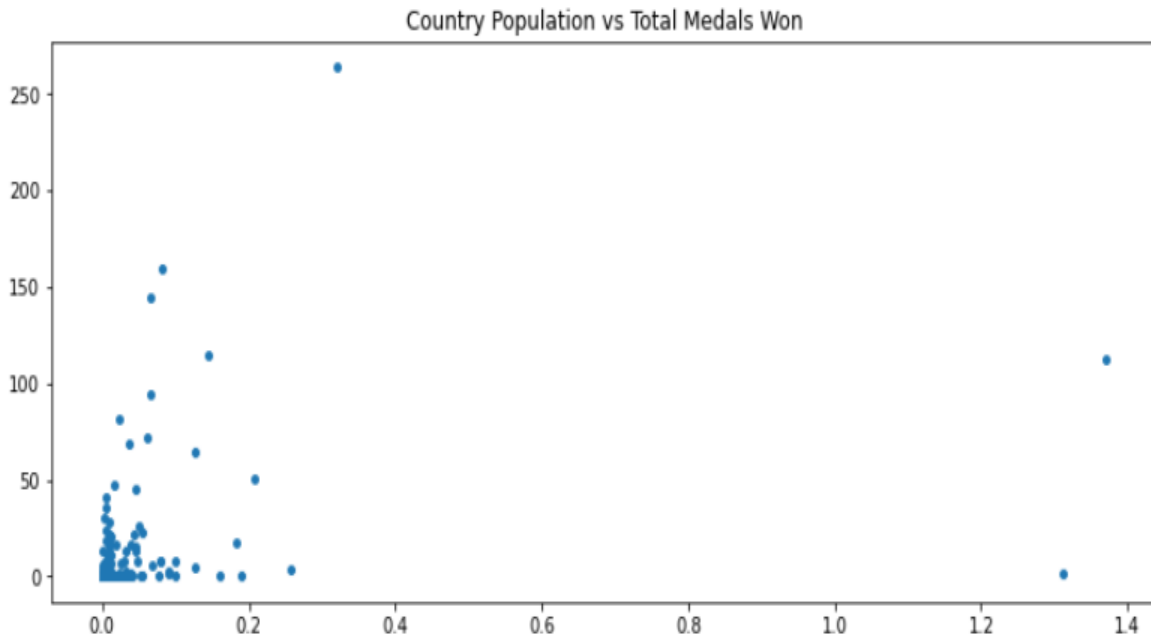
Age looks mostly as you'd expect, with younger ages winning more medals. What's interesting are the medals won for the 50+ group

BAR CHART: Medals by Sport for Athletes over 50 Years of Age



Not that these sports are not demanding, but perhaps require less physical endurance than other sports which allows for higher age groups to compete and win.

Does a country's population influence how many medals are won?

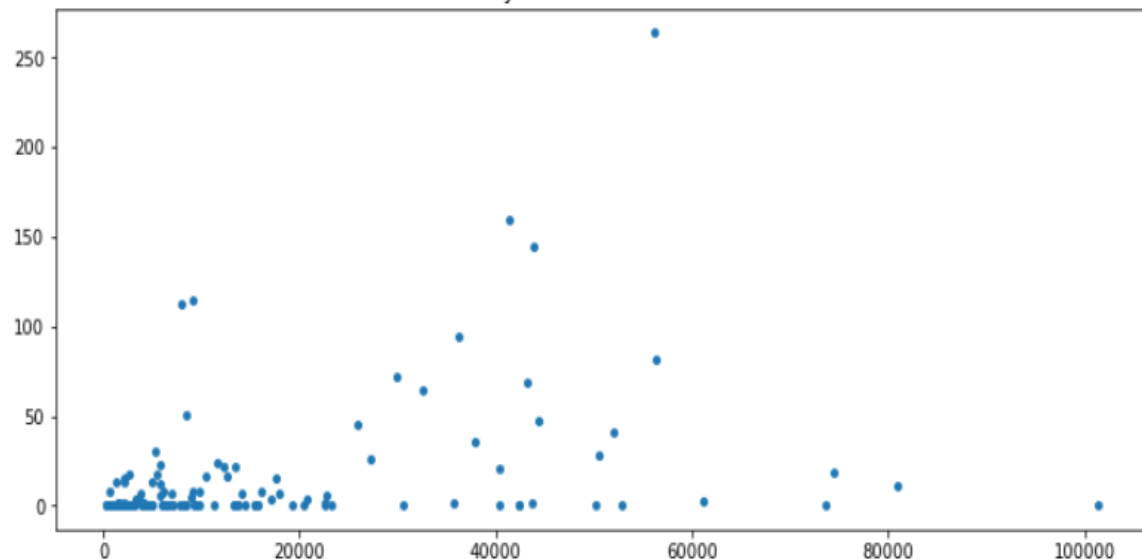


country	population	Total Medals
China	1.371220e+09	113
India	1.311051e+09	2
United States	3.214188e+08	264
Indonesia	2.575638e+08	4
Brazil	2.078475e+08	51
Pakistan	1.889249e+08	0
Nigeria	1.822020e+08	18
Bangladesh	1.609956e+08	0
Russia	1.440968e+08	115
Mexico	1.270172e+08	5
Japan	1.269585e+08	65
Philippines	1.006994e+08	1
Ethiopia	9.939075e+07	8
Vietnam	9.170380e+07	2
Egypt	9.150808e+07	3
Germany	8.141314e+07	160
Iran	7.910927e+07	8
Turkey	7.866583e+07	8
Czech Republic	1.055122e+07	15
Dominican Republic	1.052839e+07	1
Portugal	1.034865e+07	1
Hungary	9.844686e+06	22
Sweden	9.798871e+06	28
Azerbaijan	9.651349e+06	18
Belarus	9.513000e+06	12
United Arab Emirates	9.156963e+06	1
Austria	8.611088e+06	2
Tajikistan	8.481855e+06	1
Israel	8.380400e+06	2
Switzerland	8.286976e+06	11

- The results do look positive but are not linear. Four of the top five countries with most medals won are in the top 15 most populated countries.
- We see that highly populated countries like India, Indonesia and Pakistan won very little medals like Hungary won 22.

Does a country's GDP per Capita influence how many medals are won?

Country GDP vs Total Medals Won



country	gdp_per_capita	Total Medals
Luxembourg	101449.97	0
Switzerland	80945.08	11
Norway	74400.37	19
Qatar	73653.39	1
Ireland	61133.67	3
Australia	56310.96	82
United States	56115.72	264
Singapore	52888.74	1
Denmark	51989.29	41
Sweden	50579.67	28
Iceland	50173.34	0
Netherlands	44299.77	47
United Kingdom	43875.97	145
Austria	43774.99	2
Canada	43248.53	69
Hong Kong*	42327.84	0
Finland	42311.04	1
Germany	41313.31	160
United Arab Emirates	40438.76	1

- Like population, there seems to be a positive relationship between GDP per Capita and Total Medals won, though not very strong. Only three of our top five countries with most medals won are in the top 20 of GDP per Capita.
- We see that the top 5 countries only won a total of 34 medals.

Conclusions:

While there appears to be a positive correlation between total medals won and a countries with higher population and GDP per capita, though it's not as strong as originally believed.

Expected a correlation of ≥ 0.5 but got 0.066 for population and 0.14 for GDP per Capita

Perhaps raw genetic talent plays a much more significant role

Next Steps:

Expand the time frame to include Olympics from 1900 forward

Include the GDP (in addition to GDP per Capita)

Run a regression analysis and test/train model to see if we can predict medals won by certain attributes.

THANK YOU!
Questions?