Summer Olympic Medals

Danny Brown and Elizabeth Rodriguez
STAT2600
December 13, 2021

Introduction

The Summer Olympics is a very special event that happens every 4 years. In a time when the country is divided, the patriotism of the Olympics brings the whole United States together to root for a common goal: GOLD! We both enjoy watching the Olympics with family and friends when it's on, and we appreciate the bond that the Olympics creates. The first Olympics was held in Athens, Greece, and they began on April 6, 1896. Ever since they have been held every four years without fail besides a few times when they were canceled because of war or postponed by the most recent pandemic. People all over the world pay attention to the Olympics when it's on, and it is really hard to miss. Our datasets are from Kaggle, and they are called "Summer Olympic Medals (1976-2008)" and "Population of all Countries from 1960-Present". The Summer Olympics dataset is observation and categorical, but the Population dataset is observation and numerical. The Olympics dataset was collected from a database with all of the Olympic medal winners over the history of the Olympics. The population data was collected from a database with the populations, fertility rates, and life expectancy. We hope to answer the following questions using this data:

- How does gender affect how many medals are won over time by country?
- How does a country affect how many medals won over time?
- How does a country's population affect how many medals are won from 1976-2008?

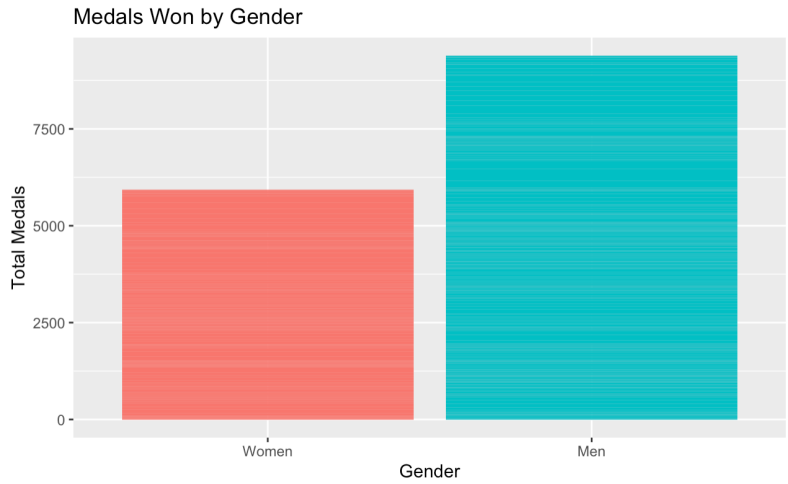
Methods and Results

Cleaning up the data

The data is loaded in as Summer_Olympics that had the following columns: City, Year, Sport, Discipline, Event, Athlete, Gender, Country_Code, Country, Event_gender, and Medal. We first removed the columns that were not needed, including City, Sport, Event, Country_Code, and Event_gender. Then we dropped all the rows that had missing data (116 rows were dropped).

ANOVA Tests

We first wanted to see if there was any variation of the medals won that could be explained by Gender. When looking at the total number of medals won just by gender, we see that there are more men that have won medals in the Summer Olympics (from 1976 to 2008) than women.



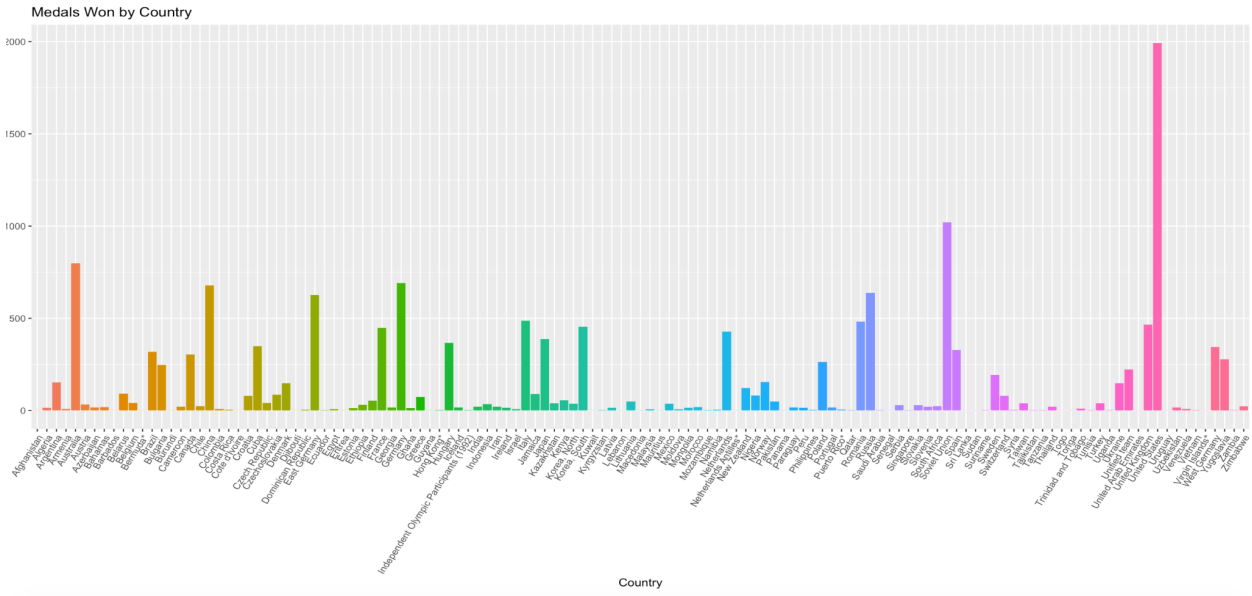
To see if Gender is a factor to explain medals won, we ran an ANOVA test. ANOVA stands for analysis of variance. It is testing if there is a difference in each of the group's means. We got an F-statistic of 0.059 and a p-value of 0.808. Because the F-statistic is small and the p-value is pretty large, we can assume gender does not affect the medals being won.

```
`summarise()` has grouped output by 'Country', 'Year'. You can override using the `.groups`
argument.
Call:
   aov(formula = n.sum ~ Gender, data = Gender)

Terms:
                  Gender Residuals
Sum of Squares      51.5  758352.1
Deg. of Freedom        1       869

Residual standard error: 29.54102
Estimated effects may be unbalanced
             Df Sum Sq Mean Sq F value Pr(>F)
Gender        1     51    51.5   0.059  0.808
Residuals   869 758352   872.7
```

We then tested the variable Country against the total amount of medals that have been won (n.sum). After running the ANOVA test we got an F-statistic of 25.87. This is quite large, so we can assume that there is a lot of variation in the mean for each country winning medals from 1976 to 2008. We also got a p-value of 2*10e^(-16) (which is very small). From the big F-statistic and small p-value, we can conclude that the country where the athletes are from (that are winning the medals) has an effect on the medals won.




```
Call:
   aov(formula = n.sum ~ Country, data = TotalMedals2)

Terms:
                  Country Residuals
Sum of Squares  1276018.6  169129.5
Deg. of Freedom       126       432

Residual standard error: 19.78644
Estimated effects may be unbalanced
             Df  Sum Sq Mean Sq F value Pr(>F)
Country     126 1276019   10127   25.87 <2e-16 ***
Residuals   432  169129     392
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

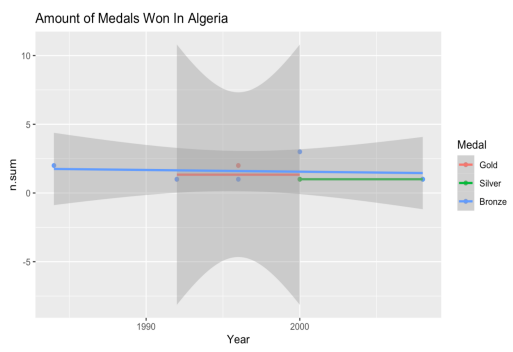
Its interesting to note from our bar graph, we see that the top five countries with the most medals won are The United States, Soviet Union, Australia, China, and Germany. The top ten countries with the most wins in the Summer Olympic Games overall are The United States, Soviet Union, Great Britain, France, Germany, China, Italy, Australia, Hungary, and Sweden.

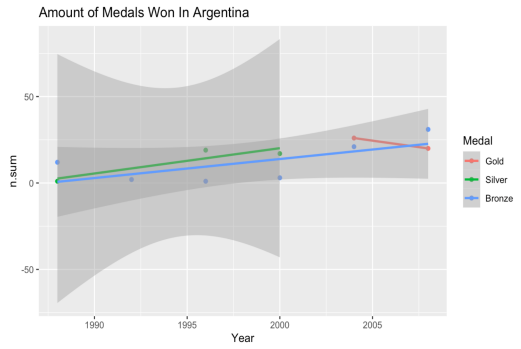
Linear Regression

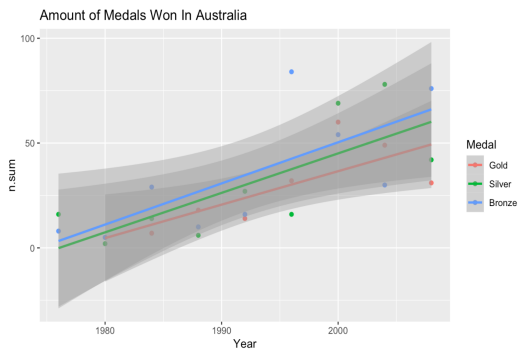
We then ran linear regression on each country from 1976 to 2008, but because some countries did not have enough data (they only won medals for one year), we only ran the linear regression on countries that had won medals at least 5 years and up.

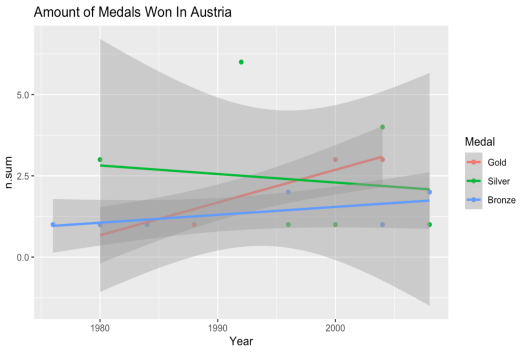
The countries we ran linear regression on are: Algeria, Argentina, Australia, Austria, Bahamas, Belgium, Brazil, Bulgaria, Canada, China, Columbia, Croatia, Cuba, Denmark, Ethiopia, Finland, France, Germany, Greece, Hungary, India, Indonesia, Iran, Ireland, Israel, Italy, Jamaica, Japan, Kenya, North Korea, South Korea, Latvia, Lithuania, Mexico, Mongolia, Morocco, Netherlands, New Zealand, Nigeria, Norway, Poland, Portugal, Romania, Slovenia, South Africa, Spain, Sweden, Switzerland, Taiwan, Thailand, Trinidad and Tobago, Turkey, United Kingdom, United States, Venezuela, and Yugoslavia.

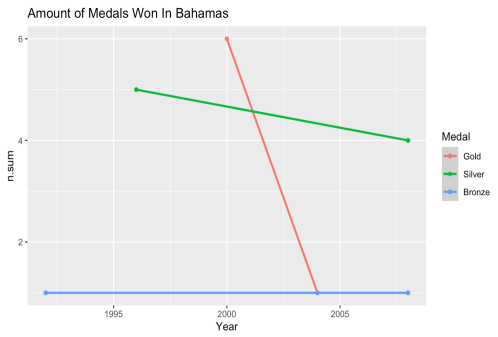
The countries that did not have enough data include: Afghanistan, Armenia, Azerbaijan, Barbados, Belarus, Bermuda, Burundi, Cameroon, Chile, Costa Rica, Cote D'lvoire, Czech Republic, Czechoslovakia, Djibouti, Dominican Republic, East Germany, Ecuador, Egypt, Eritrea, Estonia, Georgia, Ghana, Guyana, Hong Kong, Iceland, Kuwait, Kyrgyzstan, Lebanon, Macedonia, Malaysia, Mauritius, Moldova, Mozambique, Namibia, Pakistan, Panama, Paraguay, Peru, Phillippines, Puerto Rico, Qatar, Russia, Saudi Arabia, Senegal, Serbia, Singapore, Slovakia, Soviet Union, Sri Lanka, Sudan, Suriname, Syria, Tajikistan, Tanzania, Togo, Tonga, Tunisia, Uganda, Ukraine, Uruguay, Uzbekistan, Vietnam, Virgin Islands, West Germany, Zambia, and Zimbabwe.

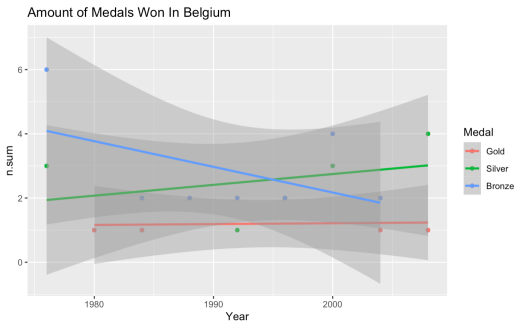












We ran it by the specific type of medals; gold, silver, and bronze, and added a confidence interval (in grey). From looking at all the graphs we can see how different and nonlinear the winnings are, which backs up the results we got from the ANOVA test.

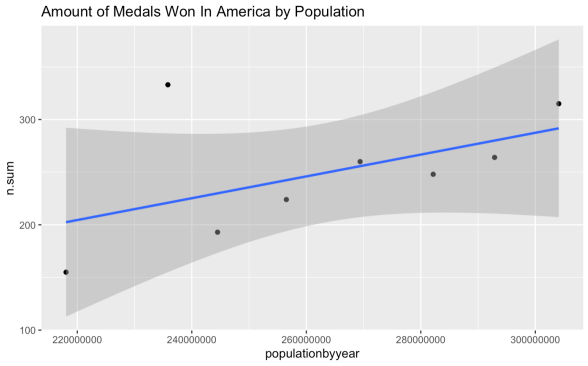
Next, we wanted to look at the population of several countries. We chose to run this test on the top ten most populated countries to see if the population affected how many medals are won. According to Population Education, the top ten most populated countries in the world are China, India, The United States, Indonesia, Brazil, Pakistan, Nigeria, Bangladesh, Russia, and Mexico. The first country that we looked into for population is The United States. From 1976-2008 The United States won medals in 1976, 1984, 1988, 1992, 1996, 2000, 2004, and 2008. We found the population numbers for each of these years and ran a linear regression looking at if the number of total medals won in each year was affected by the population of that year. When we ran the graph, the confidence interval is very large and there is one outlier that can be affecting that and the graph. It does look like it is linearly increasing. We then found the p-value to be quite large (0.903), meaning that for The United States it does not seem that rising population is a strong factor in determining the medals won in the Summer Olympics. This could also be because we are only looking at a total of 9 Olympic games (which is not a large amount of data). We also noticed that in 1980 The United States won zero medals. Our data did not include any years that medals were not won, so we imported a new dataset that includes that year to see how another outlier would affect our linear model.

```
Call:
lm(formula = n.sum ~ populationbyyear, data = USpop)

Residuals:
    Min      1Q  Median      3Q     Max
-47.504 -24.922 -17.194   9.066 112.070

Coefficients:
                    Estimate     Std. Error t value Pr(>|t|)
(Intercept)      -23.3251905069 183.7241050579  -0.127    0.903
populationbyyear   0.0000010357   0.0000006949   1.490    0.187

Residual standard error: 54.43 on 6 degrees of freedom
Multiple R-squared:  0.2702,    Adjusted R-squared:  0.1486
F-statistic: 2.221 on 1 and 6 DF,  p-value: 0.1867
```

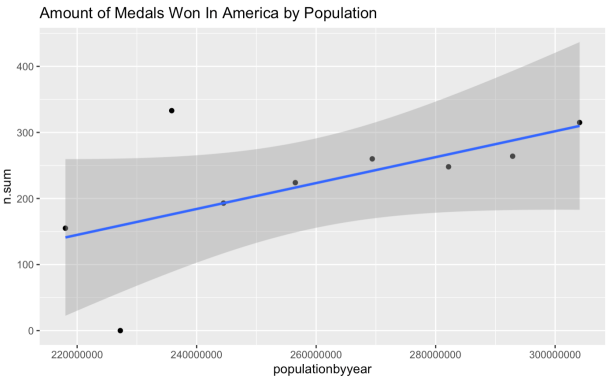



```
Call:
lm(formula = n.sum ~ populationbyyear, data = googlespreadus)

Residuals:
     Min       1Q   Median       3Q      Max
-159.084  -18.848    5.132   13.942  157.046

Coefficients:
                     Estimate     Std. Error t value Pr(>|t|)
(Intercept)      -286.632123239  262.126648798  -1.093   0.3104
populationbyyear    0.000001962    0.000001006   1.949   0.0922 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 85.78 on 7 degrees of freedom
Multiple R-squared:  0.3519,    Adjusted R-squared:  0.2593
F-statistic: 3.801 on 1 and 7 DF,  p-value: 0.09224
```



It turns out that adding in the zero medals won for the year 1989 affects our p-value and makes it slightly more significant, although it is still too high to be considered so. It should also be noted that each four-year interval had an increase of 10 million (or more) in population. When investigating the two outliers that are coincidentally next to each other, we found that the event was taking place in Moscow and The United States led a boycott to protest the late 1979 Soviet invasion into Afghanistan. There were 65 other nations that too refused to take part in the games, but there were 80 countries that still sent athletes to compete. When looking into the 1984 Summer Olympics we found that they took place in Los Angeles, so maybe there was home-field advantage. This game was also boycotted by fourteen Eastern Bloc countries (including the Soviet Union and East Germany) in response to the boycott from the last Summer Olympics. It's also interesting to note that Los Angeles will be hosting the Summer Olympics in 2028 for the third time.
The coefficient of determination tells us how much of the variation in the sum of the medals won is explained by the variable we test it against, so in our case; 25.9 percent of the variation in the total amount of medals won can be explained by population.

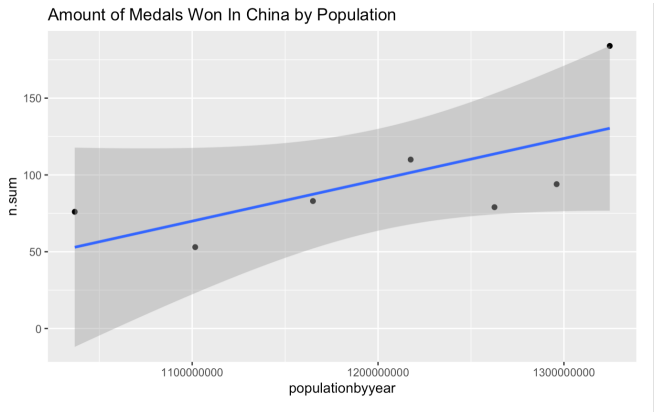
When looking at population vs the total amount of medals won in China, again we found no significance at first. Then we added the two years that they won zero medals (1976 and 1980) and got a statistically significant p-value of 0.00298. This tells us that the population in China likely affects how many medals are being won.

```
Call:
lm(formula = n.sum ~ populationbyyear, data = Chinapop)

Residuals:
      1       2       3       4       5       6       7
 23.077 -17.362  -4.406   8.445 -34.690 -28.686  53.623

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)      -2.261e+02  1.594e+02  -1.418   0.2154
populationbyyear  2.691e-07  1.323e-07   2.033   0.0977 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 34.14 on 5 degrees of freedom
Multiple R-squared:  0.4526,    Adjusted R-squared:  0.3431
F-statistic: 4.134 on 1 and 5 DF,  p-value: 0.0977
```

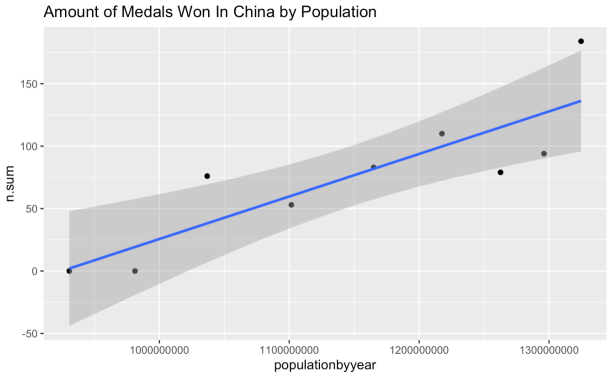


```
Call:
lm(formula = n.sum ~ populationbyyear, data = googlespreadchina)

Residuals:
    Min      1Q  Median      3Q     Max
-36.095 -19.230  -2.009  10.267  47.781

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -3.150e+02  8.837e+01  -3.565  0.00916 **
populationbyyear  3.407e-07  7.658e-08   4.449  0.00298 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.72 on 7 degrees of freedom
Multiple R-squared:  0.7387,    Adjusted R-squared:  0.7014
F-statistic: 19.79 on 1 and 7 DF,  p-value: 0.002977
```



After China started normalizing their relationship with The United States in the late 1970’s they were let onto the International Olympic Committee in 1979. They also were part of the boycott of the Olympics taking place in Moscow in 1980.
From the adjusted r-squared value on the second graph, we see that 70 percent of the variation in the total sum of the medals won in China is explained by the population.

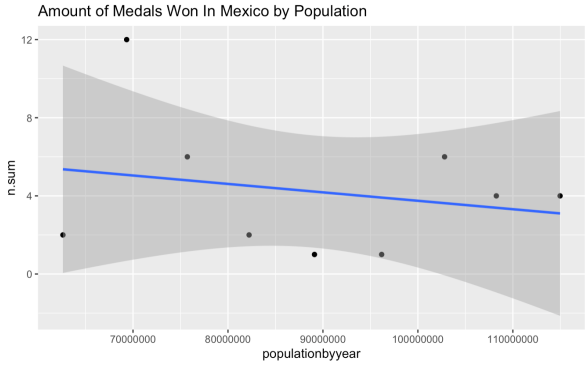
Unlike The United States and China, Mexico has won medals every year in the span of 1976 to 2008, so there was no need to add in any outliers that would be missing. It seems from the linear model that there is somewhat of a downward trend, but when running a regression test we find a p-value of 0.565, which is too high to say the population has any effect on medals won.

```
Call:
lm(formula = n.sum ~ populationbyyear, data = Mexicopop)

Residuals:
    Min      1Q  Median      3Q     Max
-3.3619 -2.9133  0.6079  1.2036  6.9278

Coefficients:
                     Estimate     Std. Error t value Pr(>|t|)
(Intercept)       8.06462564196  6.47795115122   1.245    0.253
populationbyyear -0.00000004316  0.00000007148  -0.604    0.565

Residual standard error: 3.64 on 7 degrees of freedom
Multiple R-squared:  0.04951,   Adjusted R-squared:  -0.08628
F-statistic: 0.3646 on 1 and 7 DF,  p-value: 0.565
```



We notice that the only outlier is in the year 1980. Looking further into this, we found that Mexico was one of the 80 countries that did not boycott those Olympic games.
The adjusted r-squared value can be negative because r-squared is very close to zero.

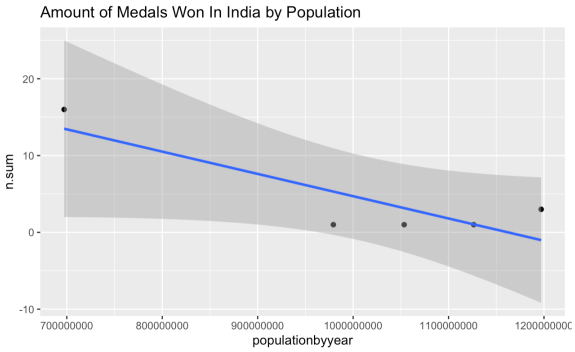
India only had 22 medals in the entire time span. The p-value is not statistically significant. The model also only has 5 points of data and therefore would not be a good predictor. They also got zero medals in 1976, 1984, 1988, and 1992. After adding those years into the data, we found that the p-value is still not statistically significant but that the linear model seems to have a smaller slope due to multiple years having the same amount of wins. Both models would not be good predictors.

```
Call:
lm(formula = n.sum ~ populationbyyear, data = Indiapop)

Residuals:
      1       2       3       4       5
 2.50452 -4.31131 -2.15863 -0.04228  4.00770

Coefficients:
                      Estimate     Std. Error t value Pr(>|t|)
(Intercept)      33.72604437642 10.34721923480   3.259   0.0472 *
populationbyyear -0.00000002902  0.00000001009  -2.875   0.0638 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.898 on 3 degrees of freedom
Multiple R-squared:  0.7337,    Adjusted R-squared:  0.645
F-statistic: 8.267 on 1 and 3 DF,  p-value: 0.06377
```

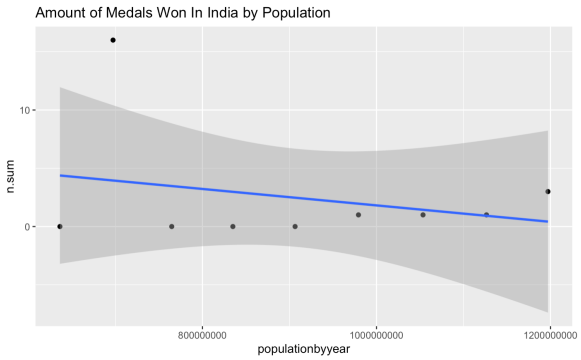



```
Call:
lm(formula = n.sum ~ populationbyyear, data = googlespreadindia)

Residuals:
    Min      1Q  Median      3Q     Max
-4.3806 -2.9784 -0.9601  0.0778 12.0501

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       8.869e+00  9.022e+00   0.983    0.358
populationbyyear -7.054e-09  9.713e-09  -0.726    0.491

Residual standard error: 5.335 on 7 degrees of freedom
Multiple R-squared:  0.07008,   Adjusted R-squared:  -0.06277
F-statistic: 0.5275 on 1 and 7 DF,  p-value: 0.4912
```



Again, we see there is an outlier in the year 1980. India is also one of the 80 countries that did compete in those Olympics and did not join the boycott.

The adjusted r-square being -0.062777 reiterated how population is not an explanatory variable for the total amount of medals won in India.

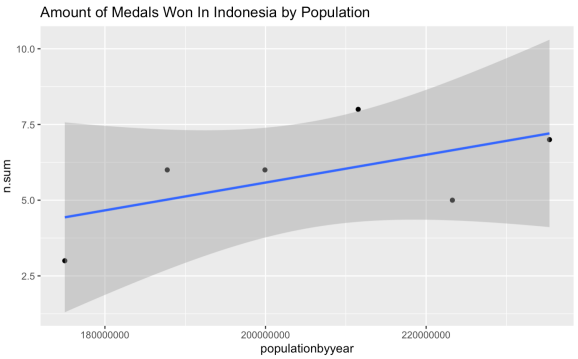
Indonesia did not win any medals in 1976, 1980, and 1984. When comparing adding those values and only modeling the years that they won, it seems the better model comes from the linear regression that used the years they did not win any medals. Not only is the p-value significant in that model and not in the other model, but the confidence interval for error is also smaller as well.

```
Call:
lm(formula = n.sum ~ populationbyyear, data = Indonesiapop)

Residuals:
      1       2       3       4       5       6
-1.4346  0.9797  0.4214  1.8884 -1.6499 -0.2049

Coefficients:
                      Estimate     Std. Error t value Pr(>|t|)
(Intercept)      -3.59720619499  6.34816637530  -0.567    0.601
populationbyyear  0.00000004590  0.00000003074   1.493    0.210

Residual standard error: 1.543 on 4 degrees of freedom
Multiple R-squared:  0.3578,    Adjusted R-squared:  0.1973
F-statistic: 2.229 on 1 and 4 DF,  p-value: 0.2097
```

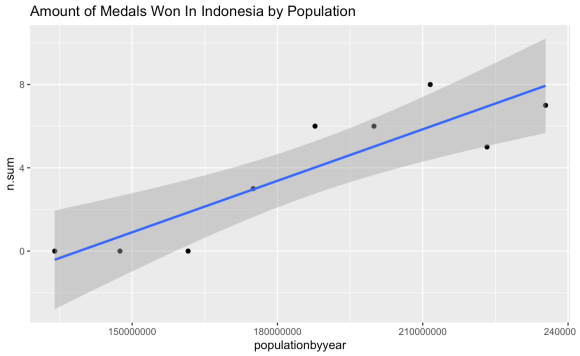


```
Call:
lm(formula = n.sum ~ populationbyyear, data = googlespreadindonesia)

Residuals:
     Min       1Q   Median       3Q      Max
-1.94454 -0.94167  0.03566  0.98027  2.02258

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.147e+01  3.076e+00  -3.727  0.00738 **
populationbyyear  8.246e-08  1.627e-08   5.067  0.00145 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.593 on 7 degrees of freedom
Multiple R-squared:  0.7858,    Adjusted R-squared:  0.7552
F-statistic: 25.68 on 1 and 7 DF,  p-value: 0.001451
```



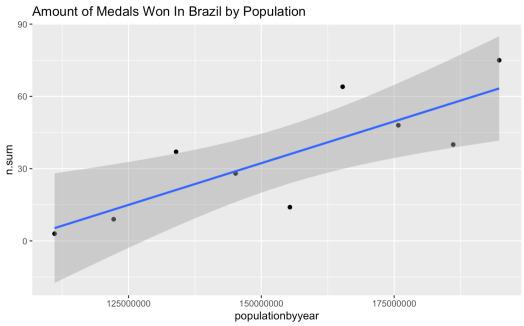
Indonesia did compete in all the games that they did not win medals in. It’s also interesting to note that Indonesia has never participated in the Winter Olympic Games and has never hosted any of the Games either.
The adjusted r-squared value is 0.7552; meaning 75 percent of the variation in the total amount of medals won in Indonesia can be explained by population.

Brazil’s linear model has a positive trend, meaning that as their population gets bigger they tend to win more medals. We also get a p-value of 0.007846, which is statistically significant and we can conclude that population is an explanatory variable in the number of medals won for Brazil.

```
Residuals:
    Min      1Q  Median      3Q     Max
-22.002  -4.011  -2.143  11.704  21.122

Coefficients:
                     Estimate     Std. Error t value Pr(>|t|)
(Intercept)      -71.6650077518  29.5108232861  -2.428  0.04553 *
populationbyyear   0.0000006929   0.0000001882   3.681  0.00785 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.33 on 7 degrees of freedom
Multiple R-squared:  0.6594,    Adjusted R-squared:  0.6107
F-statistic: 13.55 on 1 and 7 DF,  p-value: 0.007846
```



There are no outliers in this dataset.
61 percent of the variation in total medals won in Brazil is explained by population.

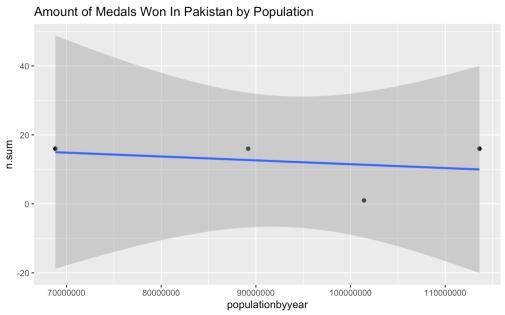
Pakistan was a country that had one of the highest populations but was one of our datasets that we chose not to run a linear regression on due to lack of data. Without adding the years they won zero medals, the p-value is fairly high; 0.715, and the confidence interval takes up almost the whole graph. This shows us that there isn’t a trend to how Pakistan is winning medals relating to population. After altering the model and adding in the years with zero wins, we can see that there is more of a trend to follow. The model shows a negative correlation between the medals won and population, but our p-value being 0.1203 is still not significant enough to claim that this trend is because of the population.

```
lm(formula = n.sum ~ populationbyyear, data = Pakistanpop)

Residuals:
      1       2       3       4
  1.018   3.294 -10.338   6.026

Coefficients:
                        Estimate    Std. Error t value Pr(>|t|)
(Intercept)      22.6762836001 25.2034597519    0.90    0.463
populationbyyear -0.0000001118  0.0000002661   -0.42    0.715

Residual standard error: 8.805 on 2 degrees of freedom
Multiple R-squared:  0.0811,    Adjusted R-squared:  -0.3783
F-statistic: 0.1765 on 1 and 2 DF,  p-value: 0.7152
```

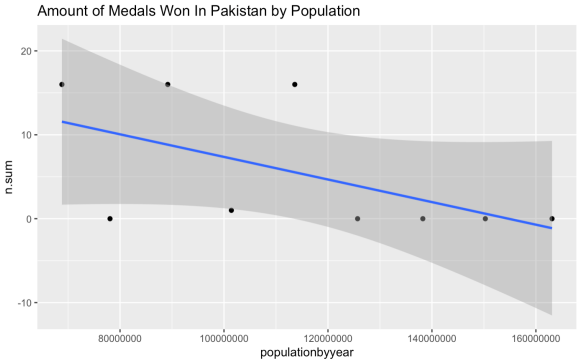



```
Call:
lm(formula = n.sum ~ populationbyyear, data = googlespreadpakistan)

Residuals:
     Min       1Q   Median       3Q      Max
-10.3173  -3.9060  -0.5983   4.4370  10.4676

Coefficients:
                        Estimate     Std. Error t value Pr(>|t|)
(Intercept)      20.82738219408  9.01058780739   2.311   0.0541 .
populationbyyear -0.00000013462  0.00000007613  -1.768   0.1203
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.042 on 7 degrees of freedom
Multiple R-squared:  0.3088,    Adjusted R-squared:   0.21
F-statistic: 3.127 on 1 and 7 DF,  p-value: 0.1203
```



Pakistan has participated in every Summer Olympic Game except for 1980 due to the boycott.

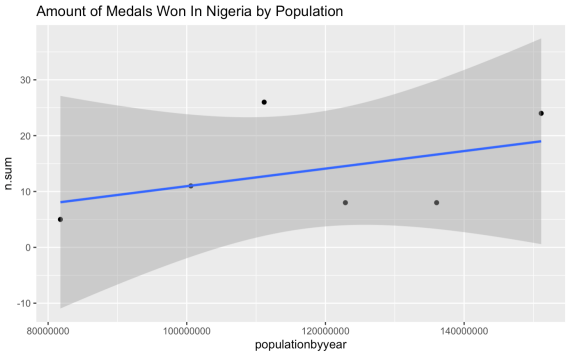
Nigeria is also one of the countries that we did not run linear regression on. From 1976 to 2008 Nigeria has only won medals in four years, 1976, 1984, 1988, and 1992. Both models are not statistically significant, suggesting that we cannot predict how many medals are being won from population in Nigeria.

```
Call:
lm(formula = n.sum ~ populationbyyear, data = Nigeriapop)

Residuals:
      1       2       3       4       5       6
-3.08350 -0.04423 13.29233 -6.55046 -8.62053  5.00639

Coefficients:
                       Estimate    Std. Error t value Pr(>|t|)
(Intercept)      -4.7830951418 19.4272831746  -0.246    0.818
populationbyyear  0.0000001573  0.0000001626   0.967    0.388

Residual standard error: 9.062 on 4 degrees of freedom
Multiple R-squared:  0.1896,    Adjusted R-squared:  -0.013
F-statistic: 0.9358 on 1 and 4 DF,  p-value: 0.3881
```

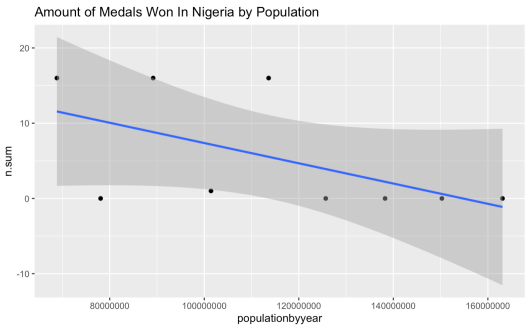


```
Call:
lm(formula = n.sum ~ populationbyyear, data = googlespreadnigeria)

Residuals:
     Min       1Q   Median       3Q      Max
-10.3173  -3.9060  -0.5983   4.4370  10.4676

Coefficients:
                      Estimate     Std. Error t value Pr(>|t|)
(Intercept)      20.82738219408  9.01058780739   2.311   0.0541 .
populationbyyear -0.00000013462  0.00000007613  -1.768   0.1203
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.042 on 7 degrees of freedom
Multiple R-squared:  0.3088,    Adjusted R-squared:   0.21
F-statistic: 3.127 on 1 and 7 DF,  p-value: 0.1203
```



There are no outliers in the data.
The adjusted r-squared value tells us that 21 percent of the variation in total medals won in Nigeria can be explained by their population.

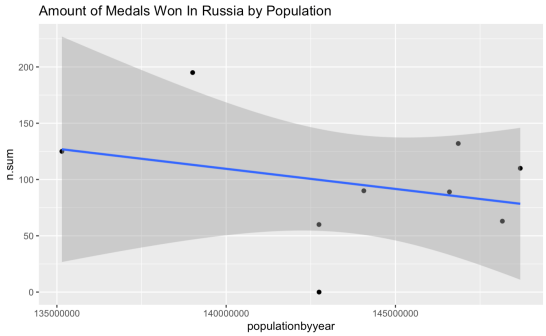
We had to create an entire new dataframe for Russia due to them participating in the Summer Olympic Games from 1976-2008 under three different names. In 1976 - 1988 they played as the Soviet Union (URS), in 1992 they played as part of the Unified Team, then as Russia. The graph looks different from the others because their population during that time span actually decreased. This model is not significant, we can tell from the p-value and just looking at the graph.

```
Call:
lm(formula = n.sum ~ populationbyyear, data = googlespreadrussia)

Residuals:
    Min      1Q  Median      3Q     Max
-99.696 -17.348  -1.842  31.541  81.960

Coefficients:
                     Estimate    Std. Error t value Pr(>|t|)
(Intercept)      609.699561836 633.852917332   0.962    0.368
populationbyyear  -0.000003573   0.000004407  -0.811    0.444

Residual standard error: 55.77 on 7 degrees of freedom
Multiple R-squared:  0.08585,   Adjusted R-squared:  -0.04474
F-statistic: 0.6574 on 1 and 7 DF,  p-value: 0.4442
```



The last country that was on the most populated list was Bangladesh, but we could not test anything due to having no data. Bangladesh has not won any medals ever - not just in that time period.

From the most populated list of countries, only 3 were significant using our data; China, Indonesia, and Brazil. India, The United States, Pakistan, Nigeria, Russia, and Mexico were not significant and there was no data on Bangladesh. We might have assumed the most populated countries would have more wins but is not what we are finding.

Conclusion

The Olympic Games are celebrated all over the world, and something that unites us all. It is interesting to test variables such as Country and Gender to see if we can predict who will win more medals. It might be obvious, but Country is a significant factor while Gender is not. Looking further into the variable Country, we can test other variables for each country. We tested the population with the idea that if a country has more people than they would win more medals. We found that is not the case for every country, while some countries can be predicted from population others cannot. We should reiterate that although this data spans over 32 years, because the Summer Olympic Games are every four years there is only a maximum of 9 years we can test over. This is a concern because of how small that dataset is, and this could potentially be fixing our models. If we were to do further research and take more years into account (before 1976 and after 2008) we might be able to find more consistent answers.

Sources

"1984 Summer Olympics." *Wikipedia*, Wikimedia Foundation, 12 Dec. 2021, https://en.wikipedia.org/wiki/1984_Summer_Olympics.

Agrawal, Divyansh. "Summer Olympics Medals (1976-2008)." *Kaggle*, 3 Feb. 2020, https://www.kaggle.com/divyansh22/summer-olympics-medals/metadata.

"Countries in the World by Population (2021)." *Worldometer*, https://www.worldometers.info/world-population/population-by-country/.

"Populations of All Countries from 1960-Present: Data Science and Machine Learning." *Kaggle*, https://www.kaggle.com/general/219676#1204240.

"The Olympic Boycott, 1980." *U.S. Department of State*, U.S. Department of State, https://2001-2009.state.gov/r/pa/ho/time/qfp/104481.htm.