Problem Chosen	2025	Team Control Number
C	MCM/ICM Summary Sheet	2524922

Analysis and Predictions of the Summer Olympics: Will the Flame Keep Burning for Everyone?

Every four years, athletes from all over the world compete at a highly competitive level in the Olympic Games. This event is nothing short of magnificent, as countries all over the world unite to compete on the biggest stage to prove their athletic abilities.

The Games present a multitude of fascinating research topics that have important implications for future Games. Many challenges arise from attempting to model Olympic medal counts, making it a popular topic of investigation. Previous models have employed socioeconomic data which proves to be an effective tool for modeling medal counts.

In contrast, we developed a model to predict Olympic performance utilizing data surrounding medal counts per country from 1992 to present. Our model reflects purely the statistics from the past Games, ignoring other factors that influence the results.

Our method is split into two parts. We first determined which year to begin our model and decided that the sociopolitical landscape post-1992 would provide us the most accurate predictions due to the collapse of the Soviet Union. We then developed an ordinary least squares regression model to make medal count predictions for the 2028 games.

Our model provided us with linear equations for each country, and we highlighted the six countries with the most interesting results; three had the highest increasing trends and three had the highest decreasing trends.

Given the restriction on data usage for this problem, we believe that the precision of our model could be enhanced by including sociopolitical factors that contribute to Olympic performance. We hope to develop a more advanced model that adapts to population and GDP per capita in the future.

Analysis and Predictions of the Summer Olympics: Will the Flame Keep Burning for Everyone?

2524922

2025-01-27

Contents

1. Background Information **1**

2. Assumptions **2**

3. Model Description **2**

3.1. Parameter Selection 2

3.2. Linear Regression Model 2

4. Analysis **9**

4.1. Most Dramatic Changes in Olympic Performance 9

4.2. Great Coach Effect 10

5. References **11**

6. Appendix **12**

7. AI Use Report **14**

1. Background Information

We were asked to create a statistical model to predict the individual and team results for the Los Angeles, USA Summer Olympics in 2028 using data from the data provided to us. There are many existing models for this problem that account for a plethora of trends, such as population, GDP per capita, political status, climate, proximity of games, accessibility of resources, and even incentives from the country. Economist Dan Johnson constructed a model in 2000 that doesn't even consider athletes' performances before the games. His model executed a 95 percent success rate at the last Olympic Games [1]. Other models suggest that total GDP for each country is the most effective predictor for Olympic medal counts [2]. In this report we will perform statistical analysis from previous Olympic Games, investigating some of the most prevalent trends in medal earnings and extrapolating the predicted outcomes for the 2028 Los Angeles Games.

2. Assumptions

Due to the constraints of the project, we were only to use data provided to us. As such, factors such as socioeconomic status and geopolitical instability were not taken into consideration when creating our model. Other Olympic models take into account variables such as religion in how it affects medal outcome, since for example in countries that are predominantly Muslim, they send in less female athletes, therefore their performance is underrepresented in those events as a result. Our model did not assume that external factors could detract from overall team improvement by handicapping the female athletes projected growth [3]. Additionally, our model did not take into account changes in GDP that could impact the performance of Olympic athletes in the 2028 Olympics.

3. Model Description

3.1. Parameter Selection

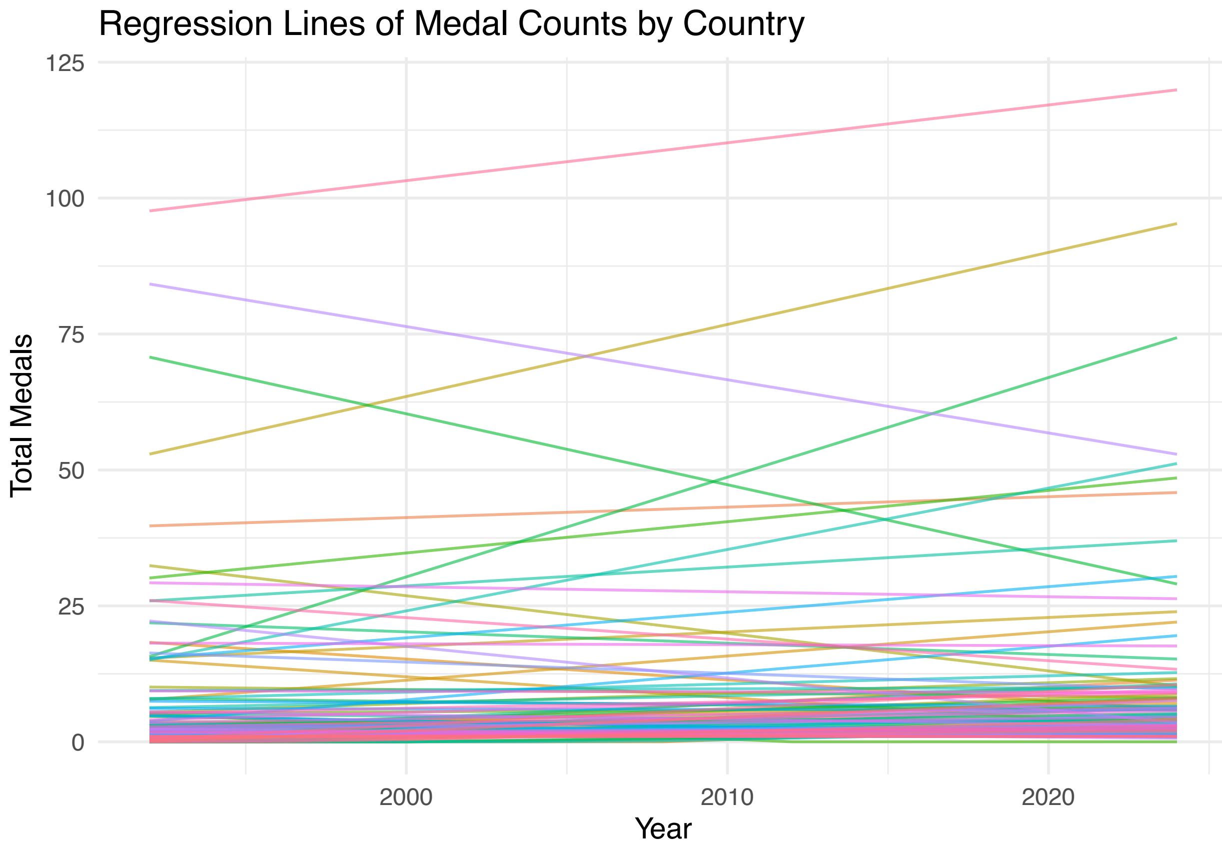
When creating our models, we decided to start by using the results for the individual and team events starting in the year 1992. The 1992 political landscape was pivotal in influencing the composition of countries competing which most accurately reflects our modern country composition. The collapse of the Soviet Union introduced 15 newly formed countries into the Olympics. This introduction of new countries alters the medal distribution from previous years and redistributes medals that would have previously been given to the Soviet Union. East and West Germany, as well as South and North Yemen had been reunified which would affect the model by consolidating medal wins that would have been previously divided. 1992 was the first time South Africa was able to compete in the Olympics since 1960 due to a ban from international sports because of the apartheid state. This introduction of a new country would affect the medal trends and diversity. With the end of the Cold War, countries were no longer competing at an ideological level and the battle between capitalism and communism drove a lot of the motivation for medal tallies [4]. Due to all these surrounding factors, for our model, we think the most accurate predictive data starts by modeling medals beginning in 1992.

Additionally, we only considered countries that have competed in the Olympic games for 2 or more competitions, since otherwise our model could not predict accurate trends in their general performance. Our model also bottoms out at zero since countries cannot earn a negative amount of medals. Hence, there are countries with decreasing medal trends that have a plateau at zero because otherwise their projected performance would be negative.

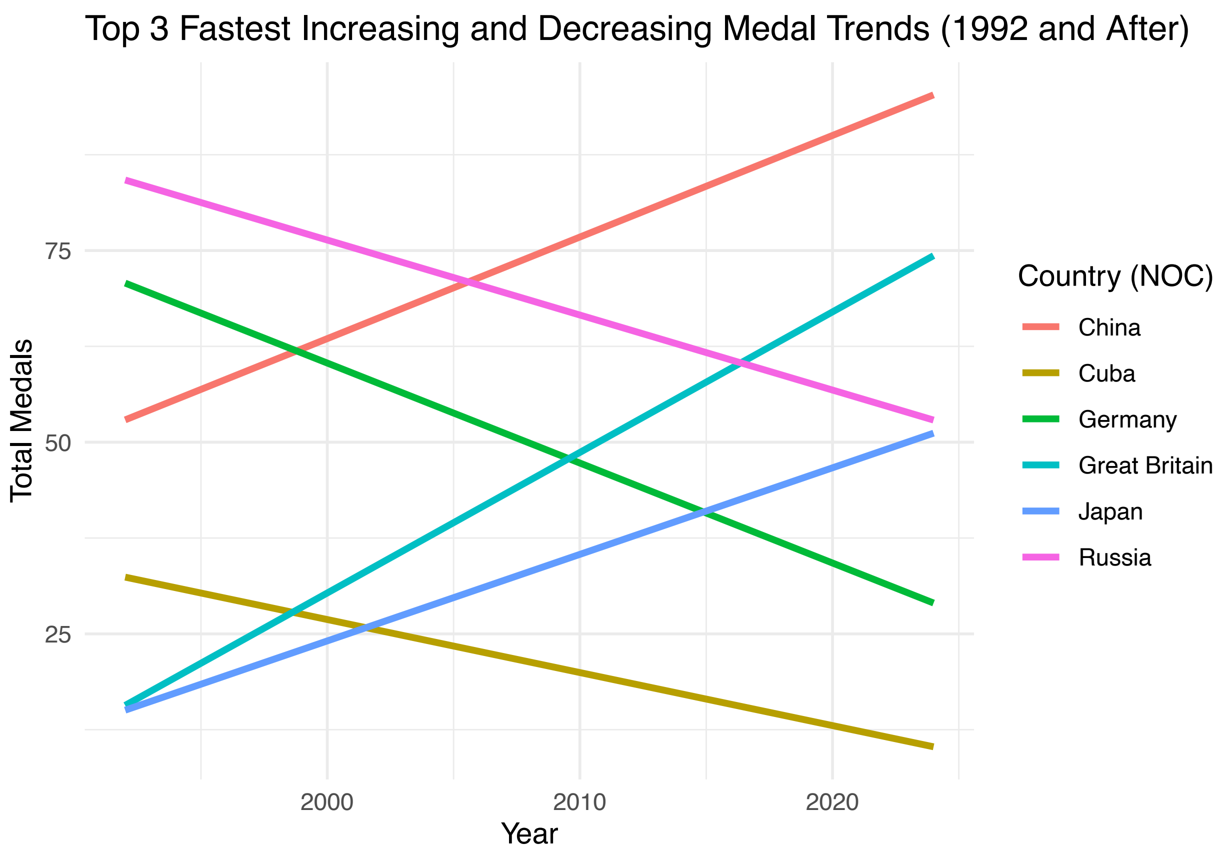
In order to measure improvements and declines in performance for specific sport events, we introduced a scale of zero to three for the type of medal that a country won, with zero being no medal, one being bronze, two being silver, and three being gold.

3.2. Linear Regression Model

We developed an ordinary least squares linear regression model to predict the relationship between the total number of medals won by each country in the Summer Olympics, taking the year as our independent variable. We applied a least squares analysis to create lines of best fit for each country, illustrated below.



Below, we highlight the three countries with the fastest increasing and fastest decreasing trends in total number of medals. The equations for their lines are given as well.



China: $Y = -2586 + 1.83X$.

Cuba: $Y = 1410 - 0.692X$.

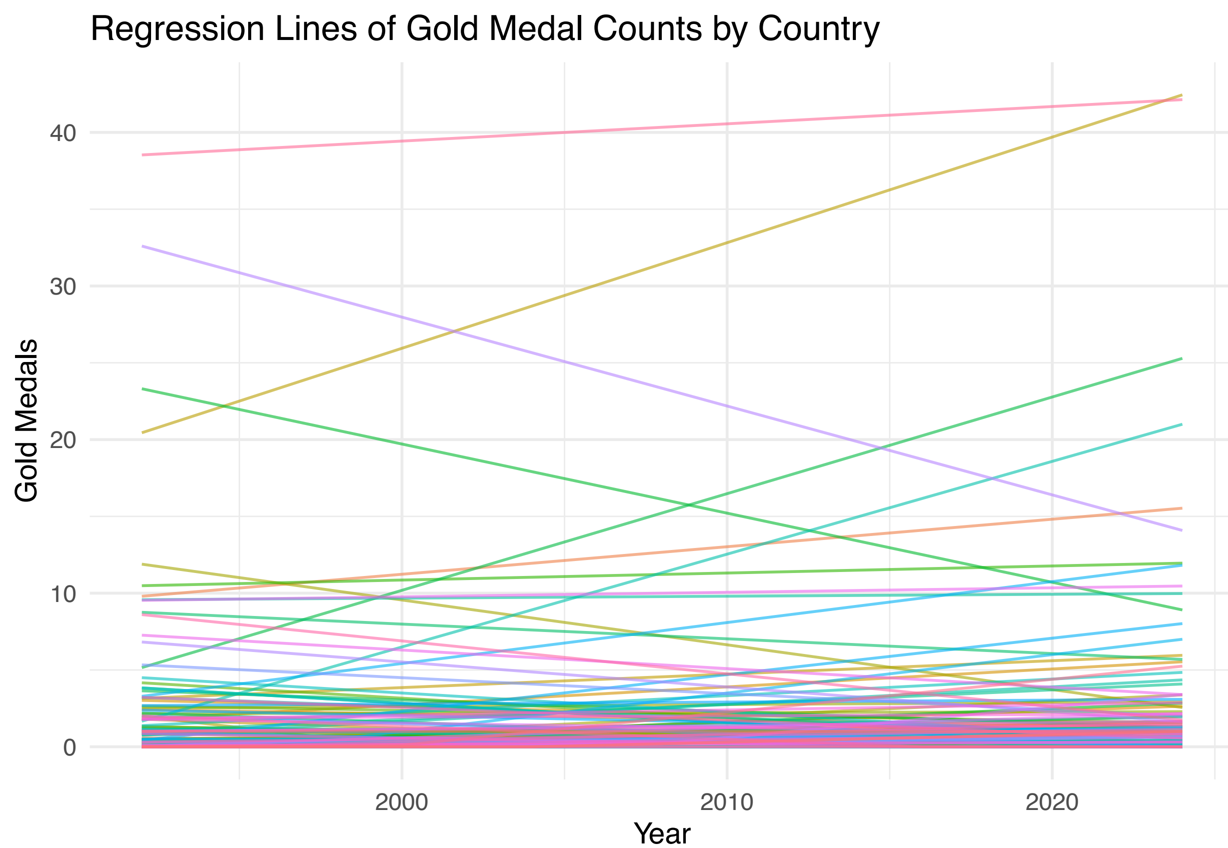
Germany: $Y = 2669 - 1.30X$.

Great Britain: $Y = -3686 + 1.83X$.

Japan: $Y = -2234 + 1.13X$.

Russia: $Y = 2034 - 0.979X$.

We then produced a similar orinary least squares regression model strictly for the number of gold medals.



From both models, we extrapolated the predictions for each country's medal count in the 2028 Olympics. The predictions for total medal count and gold medal count, respectively, for each country are given below. Additionally, a 95% confidence interval for each result is shown.

```
## # A tibble: 110 x 5
##    NOC                    model  Predicted Lower_CI Upper_CI
##    <chr>                  <list>     <dbl>    <dbl>    <dbl>
##  1 United States          <lm>      123.     110.     136.
##  2 China                  <lm>      101.      78.1    123.
##  3 Great Britain          <lm>       81.7     67.2     96.1
##  4 Japan                  <lm>       55.7     42.4     69.0
##  5 France                 <lm>       50.8     36.0     65.7
##  6 Russia                 <lm>       49.0      0      105.
##  7 Australia              <lm>       46.6     27.4     65.8
##  8 Italy                  <lm>       38.4     28.1     48.6
##  9 Netherlands            <lm>       32.3     22.3     42.3
## 10 South Korea            <lm>       25.9     18.0     33.9
## 11 Canada                 <lm>       25       18.1     31.9
## 12 Germany                <lm>       23.8     12.7     34.9
## 13 Brazil                 <lm>       23.8     18.3     29.3
## 14 New Zealand            <lm>       21.5     15.2     27.8
## 15 Spain                  <lm>       17.6     11.9     23.2
## 16 Hungary                <lm>       14.4      5.47    23.3
## 17 Kenya                  <lm>       13.3      8.37    18.3
## 18 Azerbaijan             <lm>       12.6      4.13    21.0
## 19 Denmark                <lm>       12.4      8.55    16.2
## 20 Uzbekistan             <lm>       11.8      5.37    18.1
```

```
## 21 Ukraine             <lm>     11.7     5.84    17.7
## 22 Iran                <lm>     11.7     6.66    16.7
## 23 Jamaica             <lm>     10.6     5.46    15.8
## 24 Switzerland         <lm>     10.2     5.10    15.3
## 25 Kazakhstan          <lm>     10.1     3.23    16.9
## 26 Croatia             <lm>      9.58    6.93    12.2
## 27 Turkey              <lm>      9.44    4.42    14.5
## 28 Sweden              <lm>      9.06    4.58    13.5
## 29 Costa Rica          <lm>      9.00  NaN      NaN
## 30 Zimbabwe            <lm>      9.00  NaN      NaN
## 31 Poland              <lm>      8.83    4.61    13.1
## 32 Georgia             <lm>      8.71    6.23    11.2
## 33 Chinese Taipei      <lm>      8.61    3.82    13.4
## 34 Serbia              <lm>      8.50    0.167   16.8
## 35 Czech Republic      <lm>      7.93    3.55    12.3
## 36 Belgium             <lm>      7.75    3.97    11.5
## 37 Cuba                <lm>      7.5     0       15.5
## 38 Colombia            <lm>      7.43    2.76    12.1
## 39 India               <lm>      7.18    4.08    10.3
## 40 South Africa        <lm>      6.56    1.95    11.2
## 41 Mexico              <lm>      6.47    3.07     9.87
## 42 Ethiopia            <lm>      6.44    2.42    10.5
## 43 Norway              <lm>      6.42    2.68    10.2
## 44 Thailand            <lm>      6.06    2.08    10.0
## 45 North Korea         <lm>      5.78    2.44     9.13
## 46 Kyrgyzstan          <lm>      5.68    0.621   10.7
## 47 Ireland             <lm>      5.64    2.41     8.87
## 48 Greece              <lm>      5.56    0       14.5
## 49 Austria             <lm>      5.34    1.21     9.48
## 50 Israel              <lm>      5.24    2.50     7.98
## 51 Hong Kong           <lm>      5.24    0.701    9.77
## 52 Slovenia            <lm>      4.86    3.09     6.63
## 53 Armenia             <lm>      4.70    2.21     7.18
## 54 Ecuador             <lm>      4.63    0.404    8.86
## 55 Venezuela           <lm>      4.60    0.190    9.01
## 56 Belarus             <lm>      4.46    0.616    8.31
## 57 Argentina           <lm>      4.28    1.45     7.10
## 58 Dominican Republic  <lm>      4.13    0.787    7.48
## 59 Bahrain             <lm>      4       0       10.2
## 60 Lithuania           <lm>      4.00    1.03     6.97
## 61 Egypt               <lm>      3.93    0        8.18
## 62 Philippines         <lm>      3.81    0.864    6.75
## 63 Indonesia           <lm>      3.33    1.32     5.35
## 64 Portugal            <lm>      3.29    1.06     5.51
## 65 Tunisia             <lm>      3.26    1.58     4.95
## 66 Mongolia            <lm>      3.18    0.195    6.17
## 67 Uganda              <lm>      3.04    0        8.12
## 68 Malaysia            <lm>      2.96    0.421    5.49
## 69 Tajikistan          <lm>      2.77    0        7.81
## 70 Slovakia            <lm>      2.68    0        5.65
## 71 Kosovo              <lm>      2.67    0       10.6
## 72 Moldova             <lm>      2.62    0        6.34
## 73 Guatemala           <lm>      2.33  NaN      NaN
## 74 Trinidad and Tobago <lm>      2.31    0        6.97
```

```
## 75 Latvia               <lm>       2.31     0          5.21
## 76 Vietnam              <lm>       2.20     0          5.44
## 77 Bahamas              <lm>       2.14     0.478      3.81
## 78 Algeria              <lm>       2.05     0          4.99
## 79 Chile                <lm>       2.04     0          7.79
## 80 Botswana             <lm>       2.00     0         10.3
## 81 Grenada              <lm>       2        0          4.04
## 82 Qatar                <lm>       1.96     0.161      3.75
## 83 Puerto Rico          <lm>       1.67     0.627      2.71
## 84 Estonia              <lm>       1.43     0          3.04
## 85 Fiji                 <lm>       1.33     0         17.2
## 86 Jordan               <lm>       1.33     0         17.2
## 87 Bulgaria             <lm>       1.22     0          6.65
## 88 Romania              <lm>       1.22     0          9.99
## 89 Mozambique           <lm>       1.00   NaN        NaN
## 90 Afghanistan          <lm>       1.00   NaN        NaN
## 91 Iceland              <lm>       1.00   NaN        NaN
## 92 United Arab Emirates <lm>       1      NaN        NaN
## 93 Cyprus               <lm>       1      NaN        NaN
## 94 Burundi              <lm>       1      NaN        NaN
## 95 Cameroon             <lm>       1        1          1
## 96 Ghana                <lm>       1      NaN        NaN
## 97 Kuwait               <lm>       1        1          1
## 98 Panama               <lm>       1      NaN        NaN
## 99 Syria                <lm>       1        1          1
## 100 Zambia              <lm>       1      NaN        NaN
## 101 Pakistan            <lm>       1      NaN        NaN
## 102 Peru                <lm>       1      NaN        NaN
## 103 Singapore           <lm>       0.971    0          4.01
## 104 Morocco             <lm>       0.806    0          2.72
## 105 Nigeria             <lm>       0.739    0          4.60
## 106 Namibia             <lm>       0.709    0          2.51
## 107 Finland             <lm>       0.702    0          2.70
## 108 Saudi Arabia        <lm>       0.421    0          5.98
## 109 Ivory Coast         <lm>       0.333    0          8.26
## 110 FR Yugoslavia       <lm>       0      NaN        NaN
```

Then we similarly predicted the number of gold medals each country will win in 2028.

```
## # A tibble: 110 x 5
##    NOC                  model  Predicted Lower_CI Upper_CI
##    <chr>                <list>     <dbl>    <dbl>    <dbl>
##  1 China                <lm>      45.2      30.5     59.9
##  2 United States        <lm>      42.6      34.5     50.6
##  3 Great Britain        <lm>      27.8      15.3     40.4
##  4 Japan                <lm>      23.4      14.3     32.5
##  5 Australia            <lm>      16.2       8.65    23.9
##  6 Netherlands          <lm>      12.9       7.38    18.4
##  7 France               <lm>      12.1       6.72    17.6
##  8 Russia               <lm>      11.8       0       25.3
##  9 South Korea          <lm>      10.6       5.50    15.7
## 10 Italy                <lm>      10.0       5.46    14.6
## 11 New Zealand          <lm>       8.97      6.42    11.5
## 12 Mozambique           <lm>       8.00    NaN      NaN
```

```
## 13 Germany             <lm>      7.11      0       16.4
## 14 Canada              <lm>      6.31      2.15    10.5
## 15 Uzbekistan          <lm>      6.11      2.40     9.81
## 16 Brazil              <lm>      6         2.51     9.49
## 17 Hungary             <lm>      5.31      1.82     8.79
## 18 Kenya               <lm>      5.28      2.40     8.15
## 19 Jamaica             <lm>      4.81      1.37     8.25
## 20 Iran                <lm>      4.47      1.33     7.61
## 21 Serbia              <lm>      4.20      2.98     5.42
## 22 Croatia             <lm>      3.78      1.53     6.02
## 23 Georgia             <lm>      3.18      1.46     4.90
## 24 Norway              <lm>      3.14      0.328    5.95
## 25 Sweden              <lm>      3         0.328    5.67
## 26 Czech Republic      <lm>      2.96      0.336    5.59
## 27 Belgium             <lm>      2.94      1.25     4.64
## 28 Spain               <lm>      2.92      0        8.19
## 29 Slovenia            <lm>      2.44      1.05     3.84
## 30 Denmark             <lm>      2.31      0.777    3.83
## 31 Chinese Taipei      <lm>      2.25      1.28     3.22
## 32 Switzerland         <lm>      2.17      0.120    4.21
## 33 Ireland             <lm>      2.09      0        4.87
## 34 Bahamas             <lm>      1.86      0.130    3.58
## 35 Uganda              <lm>      1.78      0        4.04
## 36 Argentina           <lm>      1.75      0        3.66
## 37 South Africa        <lm>      1.75      0        4.31
## 38 Poland              <lm>      1.58      0        4.31
## 39 Bahrain             <lm>      1.5       0        6.50
## 40 Latvia              <lm>      1.48      0.346    2.62
## 41 Thailand            <lm>      1.47      0        3.24
## 42 Philippines         <lm>      1.45      0        3.13
## 43 Ethiopia            <lm>      1.44      0        3.68
## 44 Hong Kong           <lm>      1.41      0        3.70
## 45 Ecuador             <lm>      1.40      0        4.91
## 46 Azerbaijan          <lm>      1.39      0        3.09
## 47 Cuba                <lm>      1.39      0        5.49
## 48 Austria             <lm>      1.39      0        3.28
## 49 Guatemala           <lm>      1.33    NaN      NaN
## 50 Israel              <lm>      1.33      0.141    2.51
## 51 Tunisia             <lm>      1.23      0        2.99
## 52 Bulgaria            <lm>      1.22      0        4.06
## 53 Venezuela           <lm>      1.20      0        3.40
## 54 Vietnam             <lm>      1.20      0        4.44
## 55 Belarus             <lm>      1.14      0        3.56
## 56 Pakistan            <lm>      1.13    NaN      NaN
## 57 Chile               <lm>      1.04      0        6.79
## 58 Indonesia           <lm>      1.03      0        2.13
## 59 Botswana            <lm>      1.00      0        9.32
## 60 FR Yugoslavia       <lm>      1.00    NaN      NaN
## 61 Zimbabwe            <lm>      1.00    NaN      NaN
## 62 Romania             <lm>      0.972     0        6.09
## 63 Trinidad and Tobago <lm>      0.962     0        2.97
## 64 Greece              <lm>      0.944     0        4.31
## 65 Colombia            <lm>      0.915     0        2.97
## 66 Algeria             <lm>      0.893     0        2.82
```

```
##   67 Ukraine              <lm>       0.893      0         5.20
##   68 North Korea          <lm>       0.807      0         4.36
##   69 Egypt                <lm>       0.800      0         2.31
##   70 Qatar                <lm>       0.761      0         2.53
##   71 Slovakia             <lm>       0.714      0         2.76
##   72 Portugal             <lm>       0.714      0         1.79
##   73 Puerto Rico          <lm>       0.670      0         1.71
##   74 Morocco              <lm>       0.639      0         1.97
##   75 Kazakhstan           <lm>       0.536      0         2.30
##   76 India                <lm>       0.464      0         1.39
##   77 Mexico               <lm>       0.361      0         1.69
##   78 Singapore            <lm>       0.343      0         3.51
##   79 Tajikistan           <lm>       0.343      0         3.51
##   80 Dominican Republic   <lm>       0.267      0         1.63
##   81 Turkey               <lm>       0.222      0         1.84
##   82 Estonia              <lm>       0.213      0         1.57
##   83 Mongolia             <lm>       0.207      0         1.57
##   84 Lithuania            <lm>       0.0833     0         1.56
##   85 Armenia              <lm>       0.0217     0         1.04
##   86 Afghanistan          <lm>       0        NaN       NaN
##   87 Burundi              <lm>       0        NaN       NaN
##   88 Cameroon             <lm>       0          0         3.28
##   89 Costa Rica           <lm>       0        NaN       NaN
##   90 Cyprus               <lm>       0        NaN       NaN
##   91 Fiji                 <lm>       0          0         7.59
##   92 Finland              <lm>       0          0         0.860
##   93 Ghana                <lm>       0        NaN       NaN
##   94 Grenada              <lm>       0          0         1.54
##   95 Iceland              <lm>       0        NaN       NaN
##   96 Ivory Coast          <lm>       0          0         7.26
##   97 Jordan               <lm>       0          0         7.26
##   98 Kosovo               <lm>       0          0        23.8
##   99 Kuwait               <lm>       0          0         0
##  100 Kyrgyzstan           <lm>       0          0         0
##  101 Malaysia             <lm>       0          0         0
##  102 Moldova              <lm>       0          0         0
##  103 Namibia              <lm>       0          0         0
##  104 Nigeria              <lm>       0          0         1.49
##  105 Panama               <lm>       0        NaN       NaN
##  106 Peru                 <lm>       0        NaN       NaN
##  107 Saudi Arabia         <lm>       0          0         0
##  108 Syria                <lm>       0          0         8.83
##  109 United Arab Emirates <lm>       0        NaN       NaN
##  110 Zambia               <lm>       0        NaN       NaN
```

4. Analysis

4.1. Most Dramatic Changes in Olympic Performance

As highlighted from the model, Russia has shown one of the greatest overall declines in total Olympic medals since 1992. Although a general decreasing trend did begin in Russia with the 2004 Games, the steep regression line can also be attributed to Russia's near absence from the 2024 Paris Olympics, a result of the war in Ukraine [5].

Due to Cuba's increasingly struggling economy, they have been unable to invest the necessary resources to fund elite athletics programs, hence why we see a similarly steep declining trend in their athletes' performances and their medal tallies. Their economy has not shown any recent improvements, so it can be assumed that this declining performance will continue for the 2028 Summer Olympics [6].

Finally, at the 2024 Paris Olympics, Germany had their worst Olympic performance since their reunification in 1990. While one cause of this substantial decline cannot be pinpointed, funding cuts and bureaucracy can be blamed for their worsening performance [7].

As far as the most positive trends in medals, we see from our model that China has had a steady increase in their Olympic performance throughout the years which is due to a concentrated decades-long effort by the country to invest more money into Chinese sports to show dominance and outperform other countries [8].

Japan has also had among the greatest improvements in medals. A major contributor to this may have been hosting the 2020 Summer Olympics, as well as the heavy funding Japan put into advanced training for their athletes leading up to these Games [9]. Public funding may have had a similar effect on Great Britain, who saw an even steeper rise in total medals [10].

4.2. Great Coach Effect

The "great coach" effect refers to the phenomenon that a coach has the ability to bring a team to victory based on the merit of their coaching. Since there are no stipulations around coaches being from different countries than the team they are representing, bringing in a new coach could greatly improve the success of the team they are coaching for. Based on our model, we have seen that in gymnastics, Russia has been improving steadily over the decades. Russia has had the most significant improvement in Gymanstics which could be attributed to them having great coaching. If we look at Germany's gymnastics medals, we can see that they have been decreasing in their medal gains over time. We did not want to look at a sport where they have historically always not gotten medals in, but rather a sport that they have had diminishing medal wins over time, which could be attributed to inadequate training and bad coaching. It could be suggested that Germany could investigate "stealing" a great coach from Russia's olympic team. Likewise, we see that in gymnastics Cuba has also been steadily declining in their medal results overtime, where the "great coach" effect could also help them improve as well. Finally, Russia has been declining in shooting over time, while we see that Germany has been improving in shooting; therefore, it could additionally be recommended that Russia could "steal" coaches from Germany to improve their shooting performance.

```
## `summarise()` has grouped output by 'Sport'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 5 x 4
##   Sport         medal_change final_medals initial_medals
##   <chr>                <int>        <int>          <int>
## 1 Gymnastics             -26            2             28
## 2 Handball                -6           16             22
## 3 Cycling Track           -4            5              9
## 4 Canoe Slalom            -2            2              4
## 5 Table Tennis            -2            0              2
```

```
## `summarise()` has grouped output by 'Team'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 1 x 4
##   Team   medal_change final_medals initial_medals
##   <chr>         <int>        <int>          <int>
## 1 Russia           16           16              0
```

```
## `summarise()` has grouped output by 'Sport'. You can override using the
## `.groups` argument.

## # A tibble: 5 x 4
##   Sport            medal_change final_medals initial_medals
##   <chr>                   <int>        <int>          <int>
## 1 Fencing                    -2            0              2
## 2 Canoe Sprint               -1            1              2
## 3 Taekwondo                  -1            1              2
## 4 Archery                     0            0              0
## 5 Art Competitions            0            0              0

## `summarise()` has grouped output by 'Sport'. You can override using the
## `.groups` argument.

## # A tibble: 5 x 4
##   Sport               medal_change final_medals initial_medals
##   <chr>                      <int>        <int>          <int>
## 1 Sailing                       -3            0              3
## 2 Weightlifting                 -3            0              3
## 3 Canoeing                      -2            2              4
## 4 Trampolining                  -2            0              2
## 5 Rhythmic Gymnastics           -1            6              7

## `summarise()` has grouped output by 'Team'. You can override using the
## `.groups` argument.

## # A tibble: 1 x 4
##   Team  medal_change final_medals initial_medals
##   <chr>        <int>        <int>          <int>
## 1 Japan            7            7              0
```

5. References

1. Interview: Dan Johnson discusses the correlation between economics and medal counts at this year's Winter Olympics. (2002, February 6). All Things Considered. https://link.gale.com/apps/doc/A162283474/AONE?u=gain40375&sid=bookmark-AONE&xid=9d5441ef
2. Bernard, A., & Busse, M. (2000). Who Wins the Olympic Games: Economic Development and Medal Totals. doi:10.3386/w7998
3. Bredtmann, J., Crede, C.J. and Otten, S. (2016), Olympic medals: Does the past predict the future?. Significance, 13: 22-25. https://doi.org/10.1111/j.1740-9713.2016.00915.x
4. Barcelona 1992 summer Olympics - athletes, medals & results. (n.d.). https://olympics.com/en/olympic-games/barcelona-1992
5. Maynes, C. (2024, July 28). Russia is mostly absent from the Paris Olympics, Fielding only a small team. NPR. https://www.npr.org/2024/07/28/nx-s1-5052855/russia-is-mostly-absent-from-the-paris-olympics-fielding-only-a-small-team
6. Robinson, C. (2024, August 11). Cuba disappoints at the Paris Olympics. Havana Times. https://havanatimes.org/cuba/cuba-disappoints-at-the-paris-olympics/
7. Guardian News and Media. (2024, August 13). Germany ponders its olympic future after disappointing Paris Medal Tally. The Guardian. https://www.theguardian.com/sport/article/2024/aug/13/germany-ponders-its-olympic-future-after-disappointing-medal-tally-at-paris-2024

8. Yuan, S. (2024, August 13). Gold medal rivalry: For China and the US, the Olympics are more than just sport. – The Diplomat. https://thediplomat.com/2024/08/gold-medal-rivalry-for-china-and-the-us-the-olympics-are-more-than-just-sport/#:~:text=China's%20rise%20in%20the%20Olympic,United%20States%2C%20i
9. Board of Audit of Japan. (December 4, 2019). Subsidies for supporting the performance improvement of athletes competing in the Tokyo 2020 Olympics in Japan from fiscal year 2015 to 2018 (in billion Japanese yen) [Graph]. In Statista. https://www.statista.com/statistics/1092043/japan-grants-for-performance-enhancement-of-tokyo-2020-olympics-athletes/
10. Goranova, D. (2014). The impact of public funding on olympic performance and mass participation in great britain (Order No. 10021317). Available from ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global. (1774181089). Retrieved from https://login.lp.hscl.ufl.edu/login?url=https://www.proquest.com/dissertations-theses/impact-public-funding-on-olympic-performance-mass/docview/1774181089/se-2

6. Appendix

```r
library(dplyr)
olympics <- read.csv("../2025_Problem_C_Data/summerOly_athletes.csv")
hosts <- read.csv("../2025_Problem_C_Data/summerOly_hosts.csv")
medal_counts <- read.csv("../2025_Problem_C_Data/summerOly_medal_counts.csv")
olympics <- olympics |>
  mutate(Medal = case_when(
    Medal == "No medal" ~ 0,
    Medal == "Bronze" ~ 1,
    Medal == "Silver" ~ 2,
    Medal == "Gold" ~ 3
  ))
```

```r
# Load necessary libraries
library(dplyr)
library(ggplot2)

# Filter data for years 1992 and later
medal_counts_filtered <- medal_counts %>%
  filter(Year >= 1992)

# Create regression models for each country
models <- medal_counts_filtered %>%
  group_by(NOC) %>%
  summarise(model = list(lm(Total ~ Year, data = cur_data())))

# Extract regression coefficients for each country
lines_data <- models %>%
  rowwise() %>%
  mutate(
    intercept = coef(model)[1],  # Intercept
    slope = coef(model)[2]       # Slope
  ) %>%
  ungroup() %>%
  select(NOC, intercept, slope)

# Create data for plotting regression lines
```

```r
plot_lines <- medal_counts_filtered %>%
  distinct(Year) %>%
  cross_join(lines_data) %>%
  mutate(Predicted = pmax(intercept + slope * Year, 0))  # Cap predictions at 0

# Plot all regression lines
ggplot() +
  geom_line(data = plot_lines, aes(x = Year, y = Predicted, group = NOC, color = NOC), alpha = 0.6) +
  labs(title = "Regression Lines of Medal Counts by Country",
       x = "Year",
       y = "Total Medals",
       color = "Country (NOC)") +
  theme_minimal() +
  theme(legend.position = "none")  # Remove legend if too many countries
```

```r
# Filter data for years 1992 and later
medal_counts_filtered <- medal_counts %>%
  filter(Year >= 1992)

# Create regression models for each country
models <- medal_counts_filtered %>%
  group_by(NOC) %>%
  summarise(model = list(lm(Gold ~ Year, data = cur_data())))

# Extract regression coefficients for each country
lines_data <- models %>%
  rowwise() %>%
  mutate(
    intercept = coef(model)[1],  # Intercept
    slope = coef(model)[2]       # Slope
  ) %>%
  ungroup() %>%
  select(NOC, intercept, slope)

# Create data for plotting regression lines
plot_lines <- medal_counts_filtered %>%
  distinct(Year) %>%
  cross_join(lines_data) %>%
  mutate(Predicted = pmax(intercept + slope * Year, 0))  # Cap predictions at 0

# Plot all regression lines
ggplot() +
  geom_line(data = plot_lines, aes(x = Year, y = Predicted, group = NOC, color = NOC), alpha = 0.6) +
  labs(title = "Regression Lines of Gold Medal Counts by Country",
       x = "Year",
       y = "Gold Medals",
       color = "Country (NOC)") +
  theme_minimal() +
  theme(legend.position = "none")  # Remove legend if too many countries
```

7. AI Use Report

OpenAI *ChatGPT* (ChatGPT-4)

Query 1: "My data set has columns NOC, Total, and Year. NOC is name of country. Total is the sum of gold, silver, and bronze medals the country won that year. Please write R code that will create linear regression models for each country, with Year as the predictor variable and Total as the response. Then, plot all of these lines on top of one another."

Output: # Load necessary libraries library(dplyr) library(ggplot2)

Assuming your dataset is named 'olympics' # Step 1: Split the data by country and create regression models models <- olympics %>% group_by(NOC) %>% summarise(model = list(lm(Total ~ Year, data = cur_data())))

Step 2: Extract regression line information for each country lines_data <- models %>% rowwise() %>% mutate(intercept = coef(model)[1], # Intercept of the model slope = coef(model)[2] # Slope of the model) %>% ungroup() %>% select(NOC, intercept, slope)

Step 3: Create data for plotting regression lines plot_lines <- olympics %>% distinct(Year) %>% cross_join(lines_data) %>% mutate(Predicted = intercept + slope * Year)

Step 4: Plot all regression lines ggplot() + geom_line(data = plot_lines, aes(x = Year, y = Predicted, group = NOC, color = NOC), alpha = 0.6) + labs(title = "Regression Lines of Medal Counts by Country", x = "Year", y = "Total Medals", color = "Country (NOC)") + theme_minimal() + theme(legend.position = "none") # Remove legend if too many countries"

Query 2: "Now how could I reduce the model to just the countries with the three fastest increasing and three fastest decreasing lines"

Output: # Step 1: Identify the three fastest increasing and decreasing countries top_countries <- lines_data %>% arrange(desc(slope)) %>% # Sort by slope in descending order slice_head(n = 3) %>% # Top 3 fastest increasing bind_rows(arrange(lines_data, slope) %>% # Sort by slope in ascending order slice_head(n = 3) # Top 3 fastest decreasing)

Step 2: Filter the plot_lines data to include only the selected countries filtered_plot_lines <- plot_lines %>% filter(NOC %in% top_countries$NOC)

Step 3: Plot the filtered regression lines ggplot() + geom_line(data = filtered_plot_lines, aes(x = Year, y = Predicted, group = NOC, color = NOC), size = 1.2) + labs(title = "Top 3 Fastest Increasing and Decreasing Medal Trends", x = "Year", y = "Total Medals", color = "Country (NOC)") + theme_minimal() + theme(legend.position = "right") # Keep legend for selected countries"