

36104- Data Visualization and Narratives

VISUAL ANALYSIS OF OLYMPICS DATA

Abhisek Gautam

13679042

Introduction

Olympic Games is a well-known sporting platform which is recognized all over the world, has been distinguished from the late 19th century. Its origin however can be traced back to the Greek empire, at around 3,000 years ago, which consisted only of sprint race, and was held in Greece's city Olympia only accessible to freeborn Greek people (Young & Abraham, 2020). It has grown since then and now has become a hub for all the athletes worldwide to demonstrate their abilities in more than 28 individual sporting contests. Currently, it is held every 2 years in different countries, with names Summer Olympics and Winter Olympics, both having their own set of games (Young & Abraham, 2020). It has become a place which reflects the power of the individual contestants and has become a source of pride for the countries they represent.

Motivation and Research Questions

Olympics has a rich history, spanning from 1896 till 2018, and has been a part of history. So, it is an interesting topic to see how the historical events have affected the specifics of Olympics and how it has been changing till date. Hence, this report attempts to build around the following questions, with some connection to historical happenings where suitable:

1. What effect does the host country have in the medals won at the Olympics?
2. Is the performance of countries in Olympic games affected by the economic factors of the country?
3. Is the age of winning Olympics changing?

Literature Review

The advantage of host country in any sporting activity is well known, as the participants will have familiarity of the field, and also there is a great support from the home crowd. Host countries are expected to win 3 times the medals that they were winning while playing as away (Clarke, 2000). Being a host nation and also having a communist background is also going to have a positive effect in the number of medals won (Bian, 2005).

It is found that a country's socio-economic variables, such as GDP affects the country's performance in the Olympics by a great factor. Country's population and its GDP is seen to have a correlation with the number of medals won in Olympics (Bian, 2005).

Age factor is also one of the important ones when it comes to sports and even among the athletes of the same age, relative age effect (RAE) comes into factor which determines who triumphs (Fletcher & Sarkar, 2012). RAE states that an athlete can have more advantage as compared to another who is younger by almost a year with respect to maturity, experience and early specialization (Neill, Cotton, Cuadros & Connor, 2016).

Olympics has been a part of the history, and has affected, and also has been affected by the history. Olympic Games has made major social and political impacts throughout the history like including women in sports, has taken a stand against racial matters, promoted civil rights, has unified countries and even has been a tool to demonstrate power by different countries (O'Connell). Similarly, the politics of the countries such as racial separation, terrorism, World Wars and the Cold War have also affected Olympics at different times in the history (Dwyer & McMaster, 2018).

Approach

To answer the questions, three datasets were used namely, “120 years of Olympic history: athletes and results¹”, “Gapminder GDP per capita, constant PPP dollars- v25²” and “Gapminder Total Population v6³”. The Olympics dataset had the names of participants, their demographics, which sport they participated in and on which Olympic games. Gapminder’s GDP per capita dataset consisted of GDP of countries from 1960 forecasted till 2040, and their Population dataset had populations of all the countries of the world from 1800 till 2019 and forecasted from there onwards till 2100. A custom dataset was also created that maps the cities mentioned in the Olympics dataset to country names.

Visualizations were created in Tableau and Python. Initial data cleaning was done on Excel, and visualization specific data manipulation were carried out as needed on Tableau and Python.

Findings and Discussion

Host Country Effect

Figure 1 shows the number of times a country has hosted Olympics. The USA has hosted it the most, i.e. 8 times followed by France, which had hosted it 5 times.

Which countries have hosted Olympics how many times?

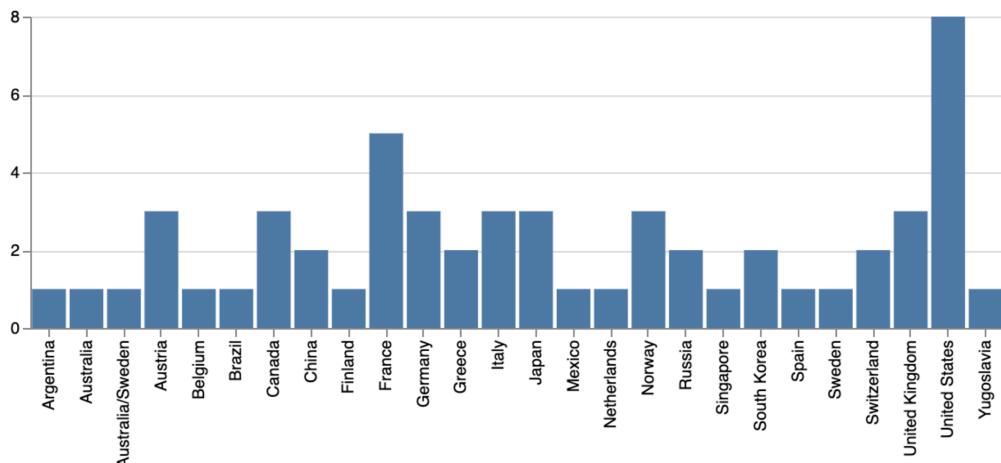


Figure 1- Countries that have hosted Olympics

Figure 2 shows the total medal winnings of countries that hosted Olympics, and shows that USA has won the most medals, and then it is Germany, followed by France. This hints that the countries which have hosted more Olympics have won more medals. The reasons why the host countries win more medals are due to home field advantage, crowd support and easier qualification criteria, hence more participation (Clarke, 2000).

¹ <https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results>

² <https://www.gapminder.org/data/documentation/gd001/>

³ https://gapm.io/d_popv6

Medals by Host Country

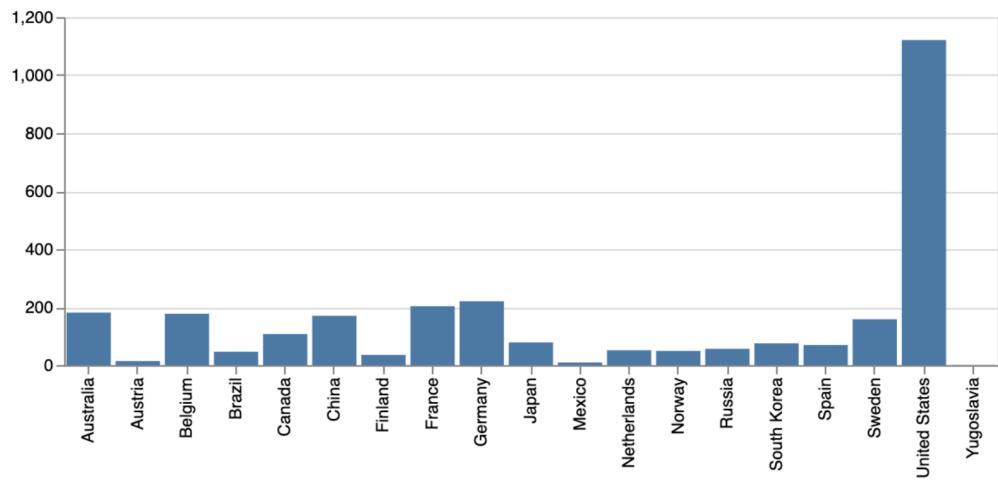


Figure 2- Medals won by Host Countries

Now let us see how the number of won medals vary with regards to the host country. Firstly, looking at the USA's medals won (Figure 3-4), we can see that the number of medals won by it has always increased when it hosted the Olympics. Note the sharp increase in the graph from 1976 till 1984. USA had protested against Russia's invasion of Afghanistan by boycotting the 1980's Olympics held in Russia, so the data of that year is not available. And immediately in 1984, Russia boycotted 1984's Olympics which was held in the USA, thus the USA won more medals.

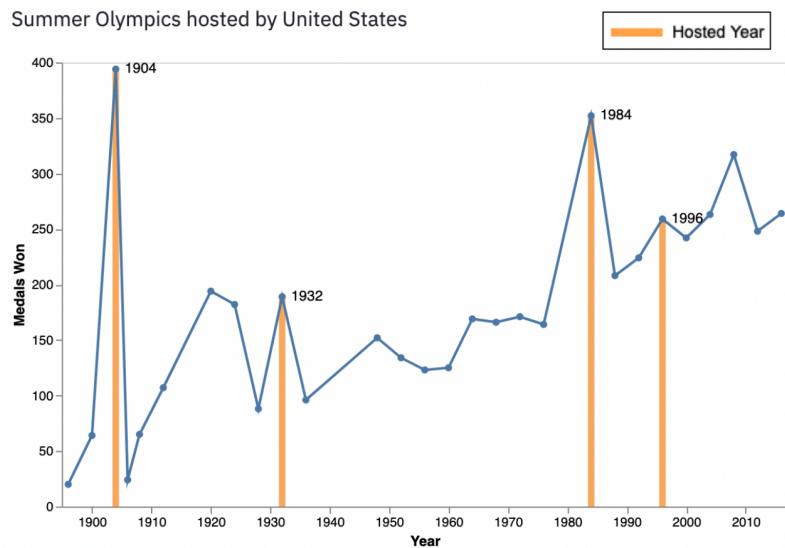


Figure 3- Summer Olympics medal counts of the USA

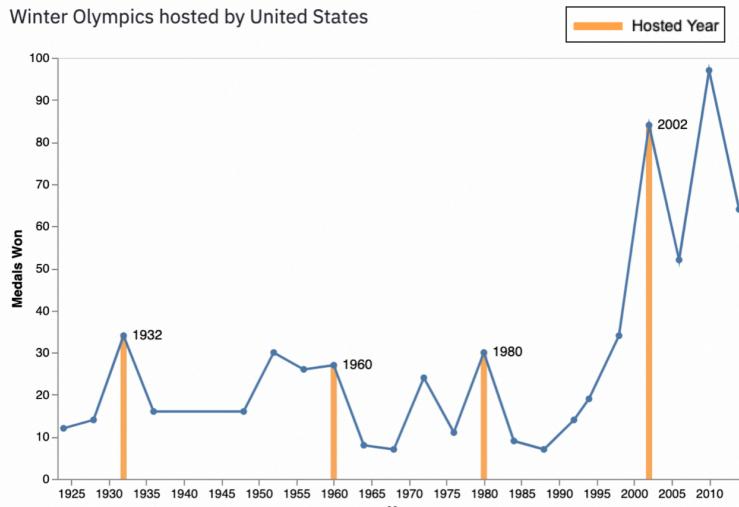


Figure 4- Winter Olympics medal counts of the USA

In some cases, the host country prevails in successive Olympics after they host, such as Spain (Figure 5), which had a tremendous increase in medal count when they hosted the Summer Olympics in 1992 (from 5 to 69) and have won more medals in the following years.

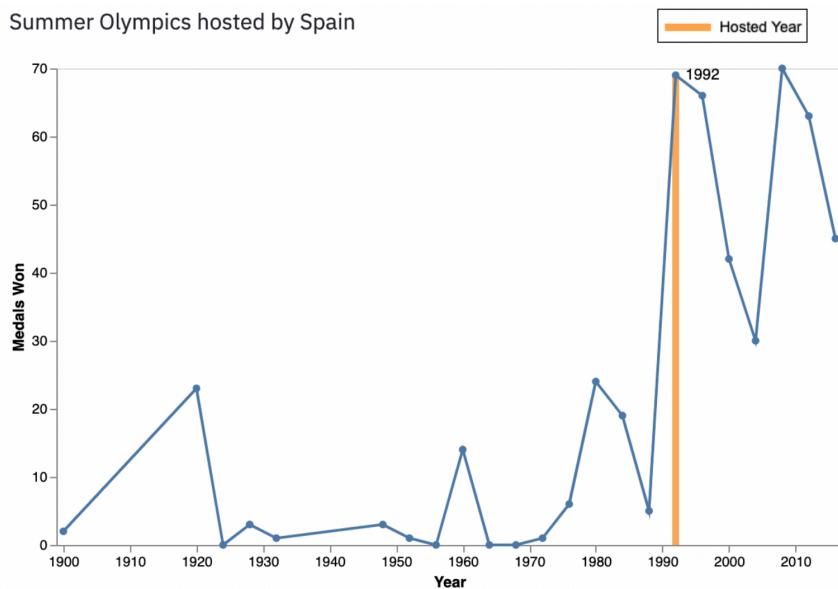


Figure 5- Summer Olympics medal counts of Spain

A similar trend is shown by other countries such as Japan (1964), South Korea (1998), Mexico (1968) and Norway (1994) where they have increased their medal counts in average after holding the Olympic games. However, countries such as Finland (1952), Canada (1976) and Russia (1980), though they won more medals when hosting, were not able to increase their medals in the successive years (Refer to Appendix I).

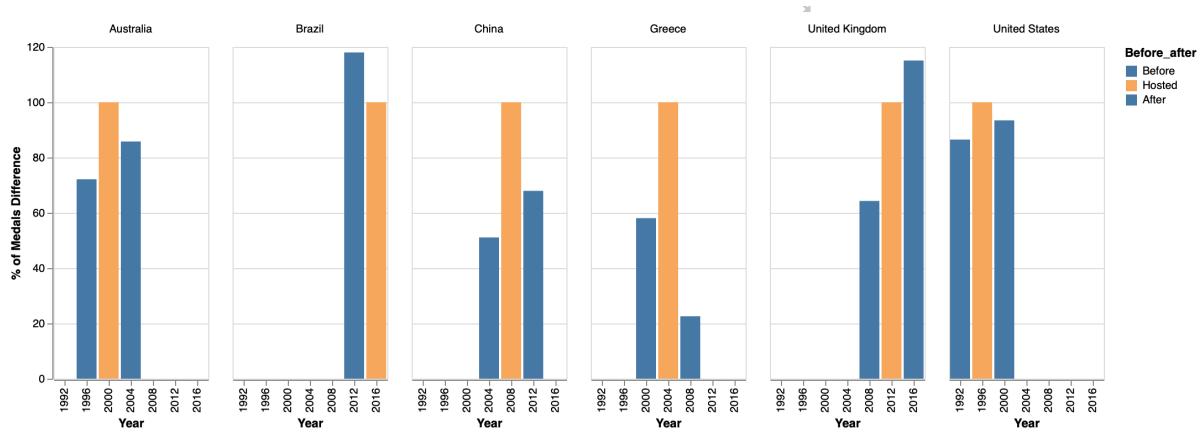


Figure 6- Percentage of winnings before, when hosting and after in Olympics, (100% when hosted)

Countries such as Australia (2000), China (2008) and Greece (2004) United Kingdom (2012) and United States (1996), have won between 10-20 % more medals when they were the hosts (Figure 6). Its only Brazil (2016) which failed to score more when they had hosted it (20 per cent less than when UK Olympics in 2012), whereas Greece managed to bag only 20 percent medals in 2008 after the Greece Olympics.

Economic Effect

The most basic and important economic predictors, the Gross Domestic Product (GDP) can be used to evaluate the performance of a country in the Olympics (Bernard & Busse, 2000). One of the derivatives of the GDP is the GDP per capita, which gives the earning of an individual person in the country can also be looked into. The GDP and GDP per capita are plotted in Figures 7-8 to see how they relate to the winnings. The graph is observed to be more scattered in the plot with GDP per capita (Figure 7), with correlation coefficient only 0.1, whereas that for total GDP was 0.17 (Figure 8), meaning that for the 2004 Olympics (and other Olympics- Appendix II), the number of medals won is a factor of GDP.

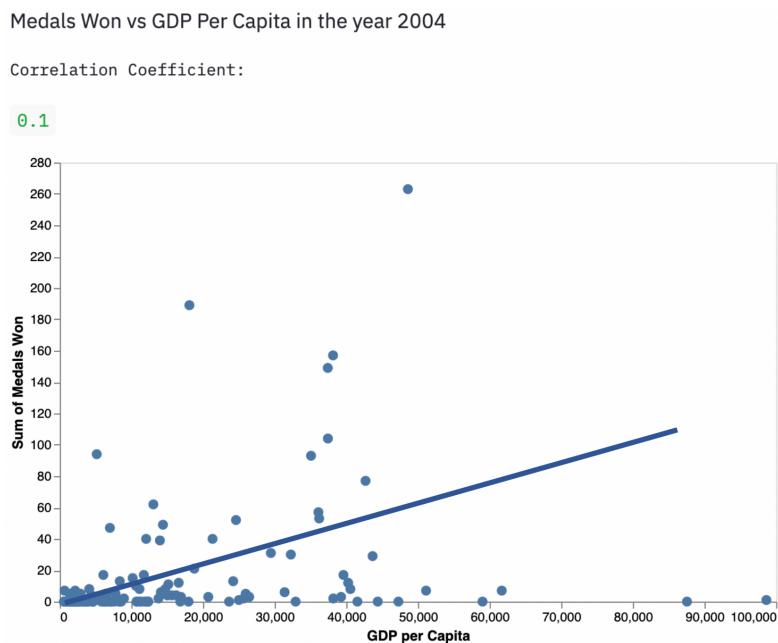


Figure 7- Medals Won vs GDP Per Capita in 2004 Olympics

Medals Won vs Total GDP in the year 2004

Correlation Coefficient:

0.17

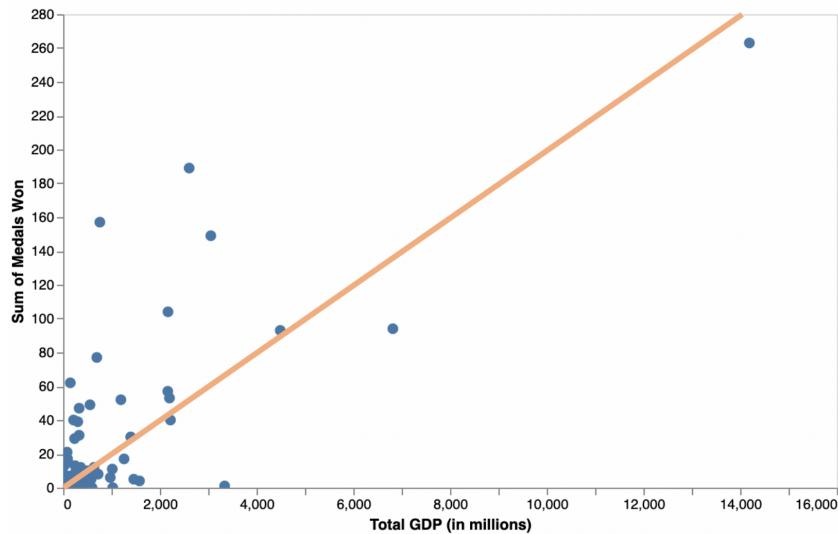


Figure 8- Medals Won vs Total GDP in 2004 Olympics

Analysing the correlations of total population along with GDP and total GDP to see how they relate to the medal counts (Figure 9), the GDP has always had a better correlation with medals won. This is not surprising because just having a better economy does not guarantee an Olympic medal. It also requires a large population pool from which best participants can be selected, much like selecting from a large sample space.

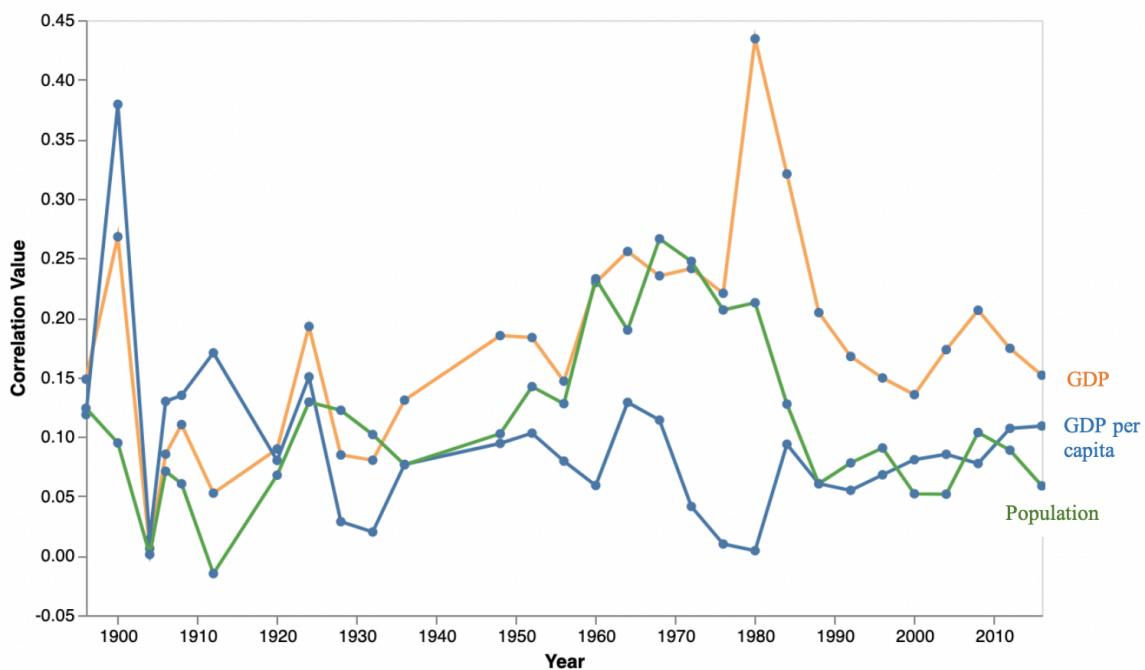


Figure 9- Correlation of GDP, GDP Per Capita and Population plotted with Year

Figure 9 also shows the correlation values have gone up and down for Population and GDP per capita along the years. The early periods, from 1896 till 1936 show that the country with the most GDP per capita had better chance of winning medals. This may have been due to

rich countries only participating in the Olympics. This changed after the World War II until 1988 (and the Population was more correlated to medal count in comparison). In this time, there was Cold War between The Soviet Union and USA, rise of communism (esp. in China), and prevalence of drug use by athletes. Olympics was a good stage for the countries to show their power. Countries which were able to select from a greater population, keep their athletes paid and trained won more medals. The concept of paying Olympic athletes also started at that time. Soon after the downfall of Soviet Union in 1991 and the unification of Germany, the competition to demonstrate power had subsided. So, in the recent years, the role of population and GDP per capita has very less correlation (less than 0.1).

The number of medals won, and total GDP of the China and Tuvalu are shown in Figure 10-11.

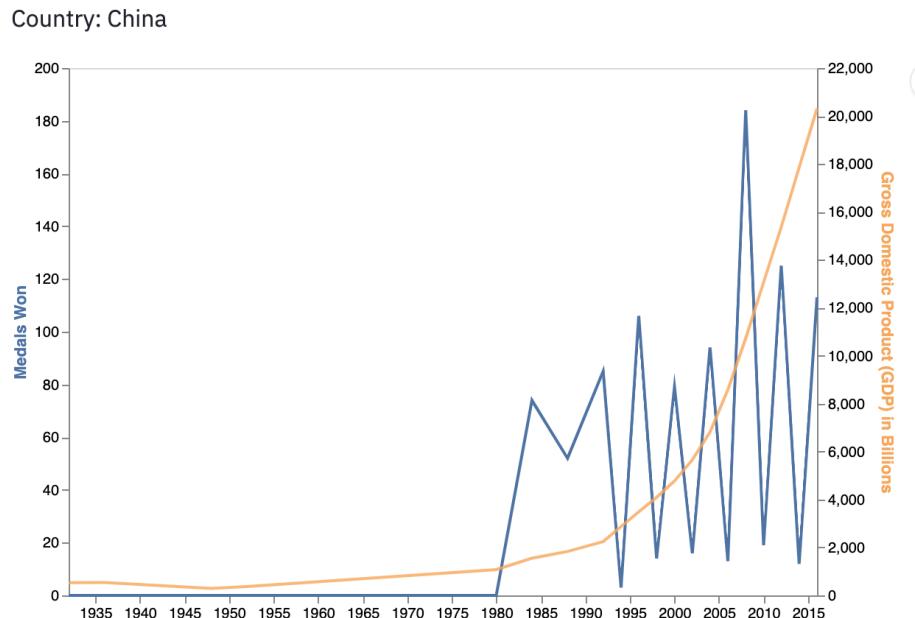


Figure 10- Medals won by China (highest GDP)

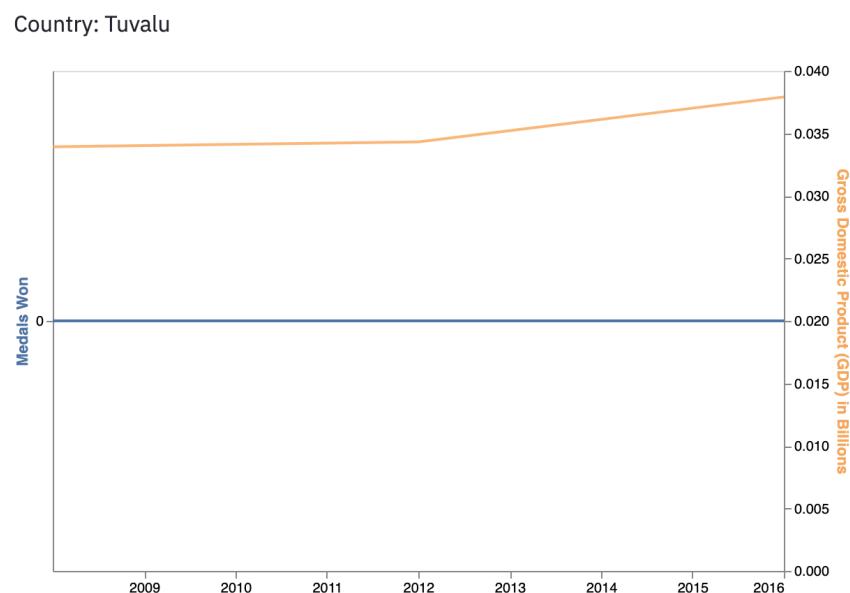


Figure 11- Medals won by Tuvalu (lowest GDP)

If we look into the top and bottom GDP countries (Appendix II), we see that the countries that had the least GDP have zero medals, whereas the top have all won more than 50 medals, with the highest of 264 by the United States. All these hint that the total medal winnings are related to the total GDP of a country in the Olympics.

Is the age of winning Olympics changing?

Figure 12 shows the mean age of the medal winners from the start of Olympic Games. The mean age was high during the early 1900s (around 27-28), had some fluctuations till 1948. It decreases to 24-25 there onwards till the 90s, from where it goes up to 26-27 till 2016 Olympics.

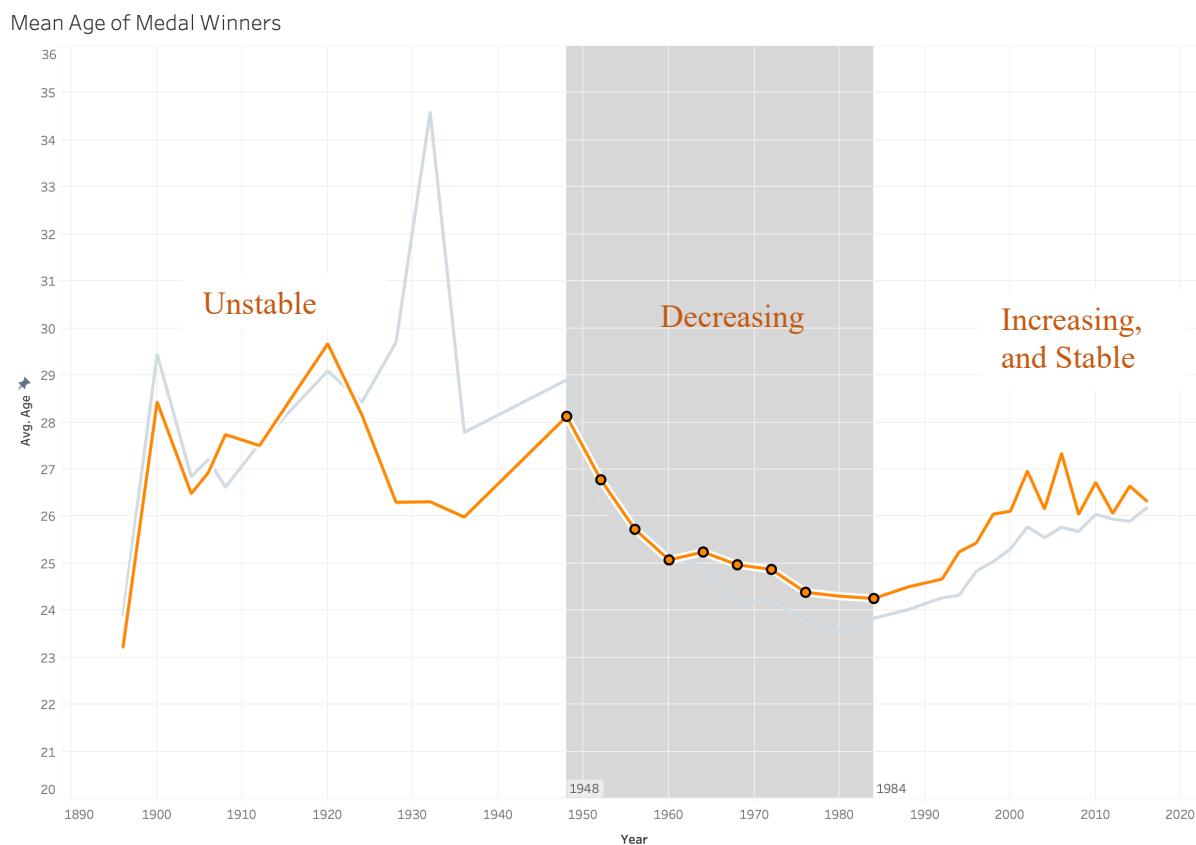


Figure 12- Mean Age of Medal Winners in Olympics

Now, let us look at how the ages change for the sports that have been continuing since the start of the modern Olympics.

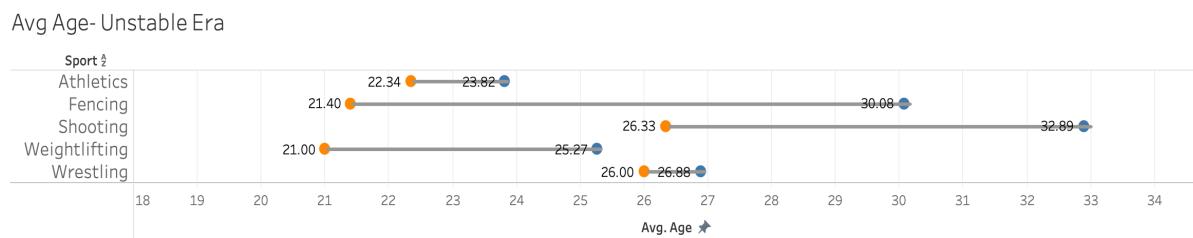


Figure 13- Age of winners from 1896 till 1948

Avg Age- Decreasing Era

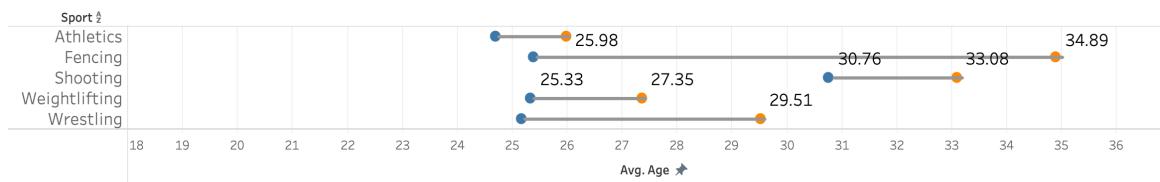


Figure 14- Age of winners from 1948 till 1984

Avg Age- Stable Era

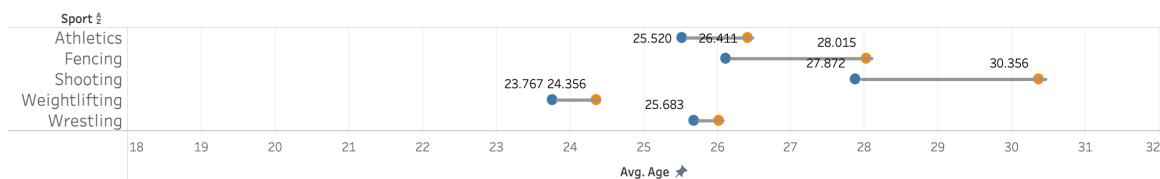


Figure 15- Age of winners from 1988 till 2016

Figure 13 shows that the age of winners had a big variation during the unstable time period, which decreased a little on the decreasing period (Figure 14), and finally during the modern time (Figure 15), it has become the least. The optimum age to win certain sport is seen to be narrowing down. Appendix III shows the variation for all sports from 1988 till 2016.

Conclusion

It is clear that the host countries have always a better chance of winning medals in the Olympics; they can win at least 10-20 percent more medals. Looking at the economic effect, even though country's population and per capita GDP affected the number of medals won in the past, the total GDP of the country is more significant to determine the winnings in the recent years. With the age factor, the age range of players winning medals has decreased over the years, and an optimal age for each sport can be identified in the recent years. Thus, there is a high chance for an athlete from a host country with high GDP, whose age range falls in the optimum age range for the sport to win a medal in the Olympics.

Recommendations

The study is a high-level analysis of the factors that come into play in the performance in the Olympics. There are many areas where deeper analysis into each of the factor can be made. Some of the questions that may be looked upon can be as follows:

- Host Country Effect-** What is the probability that a participant of the host country can win a medal? Are there any type of sports that the host country has more chance of winning?
- Economic Effect-** Are there any sports that are won mostly by countries with low GDP? Does the rise/fall in economy affect the Olympic winnings of a country, and if so, in how many years is the effect seen?
- Age Effect-** Does the Relative Age Effect influence Olympic winnings? At what age is optimum age for each sport? How does the performance of each athlete improve/degrade over time?

Moreover, looking at the medals count in Olympics official site⁴ and comparing with the Olympics history dataset used in this report, there are discrepancies, hence the dataset needs to be revised.

⁴ <https://www.olympic.org/>

References

- Mahtani, K.R., Protheroe, J., Slight, S.P., Demarzo, M.M.P., Blakeman, T., Barton, C. A., Brijnath, B., Roberts, N. (2012). *Can the London 2012 Olympics 'inspire a generation' to do more physical or sporting activities? An overview of systematic reviews*. Retrieved from <https://bmjopen.bmjjournals.org/content/3/1/e002058.full>
- Clarke, S. R. (2000, June). Home advantage in the Olympic Games. In *Proceedings of the 5th Australian Conference on Mathematics and Computers in Sport. University of Technology Sydney, Australia* (pp. 76-85). Retrieved from <https://researchbank.swinburne.edu.au/file/3bbcf005-ec1e-4def-9602-f7dd18d0a711/1/PDF%20%28Published%20version%29.pdf>
- Fletcher, D., & Sarkar, M. (2012). A grounded theory of psychological resilience in Olympic champions. *Psychology of sport and exercise*, 13(5), 669-678. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S1469029212000544>
- Bian, X. (2005). Predicting Olympic medal counts: The effects of economic development on Olympic performance. *The park place economist*, 13(1), 37-44. Retrieved from https://www.researchgate.net/profile/Xun_Bian/publication/28328103_Predicting_Olympic_Medal_Counts_the_Effects_of_Economic_Development_on_Olympic_Performance/links/5525b9850cf24b822b4058f9/Predicting-Olympic-Medal-Counts-the-Effects-of-Economic-Development-on-Olympic-Performance.pdf
- O'Connel, C. *13 Olympic Moments that Changed History*. Retrieved from <https://www.rd.com/culture/13-olympic-moments-that-changed-history/>
- Dwyer, B. B., & McMaster, A. (2018). *18 Times Politics Trumped Sport in Olympic Games' History*. Retrieved from <https://www.globalcitizen.org/en/content/history-political-activism-olympics-rio/>
- Bernard, A. B., & Busse, M. R. (2000). *Who wins the Olympic Games: Economic development and medal totals* (No. w7998). National Bureau of Economic Research. Retrieved from <https://www.nber.org/papers/w7998.pdf>
- O'Neill, K. S., Cotton, W. G., Cuadros, J. P., & O'Connor, D. (2016). An investigation of the relative age effect amongst Olympic athletes. *Talent Development & Excellence*, 8(1), 27-39. Retrieved from https://www.researchgate.net/profile/Juan_Cuadros2/publication/317582562_An_Investigation_of_the_Relative_Age_Effect_amongst_Olympic_Athletes/links/5cb4a254299bf120976828_56/An-Investigation-of-the-Relative-Age-Effect-amongst-Olympic-Athletes.pdf
- Young, D. C., Abrahams, H. M. (2020). *Olympic Games*. Retrieved from <https://www.britannica.com/sports/Olympic-Games>
- Olympic Sports - How Does Peak Age Vary?*. Retrieved from <https://www.thestatszone.com/archive/olympic-sports-how-does-peak-age-vary-13812>
- James, L. Does money buy medals? Analyzing the affect of GDP on Olympic success.
Retrieved from <https://www.yellowfinbi.com/blog/2012/08/yfcommunitynews-does-money-buy-medals-analyzing-the-affect-of-gdp-on-olympic-success-117473>

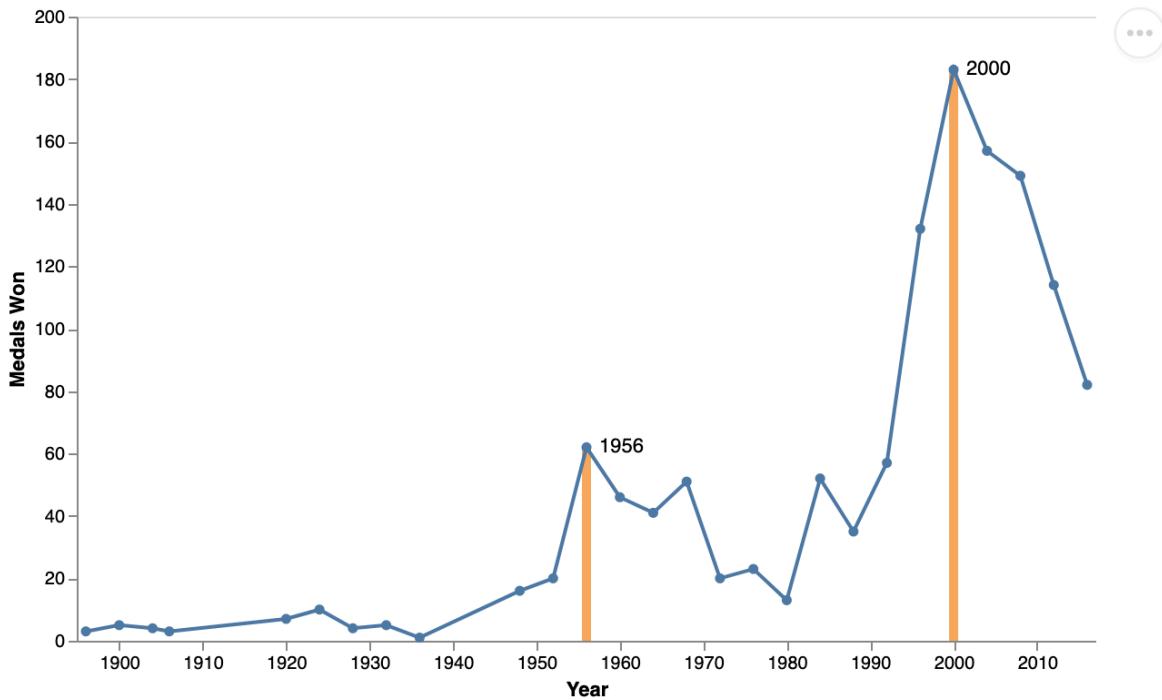
Toshkov, D. (2016). Olympic medals, economic power and population size. Retrieved from
<http://re-design.dimiter.eu/?p=868>

Appendix I- Olympic Performance of different hosting countries

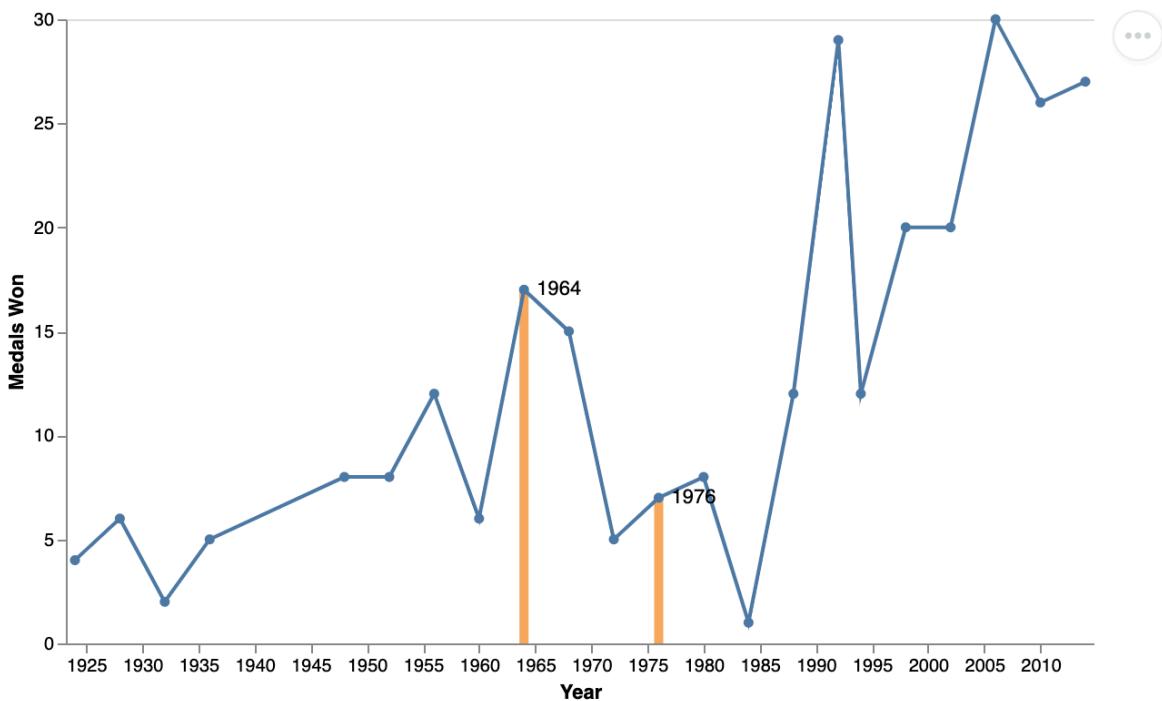
Note- the Orange line indicates the Hosted year for all graphs.



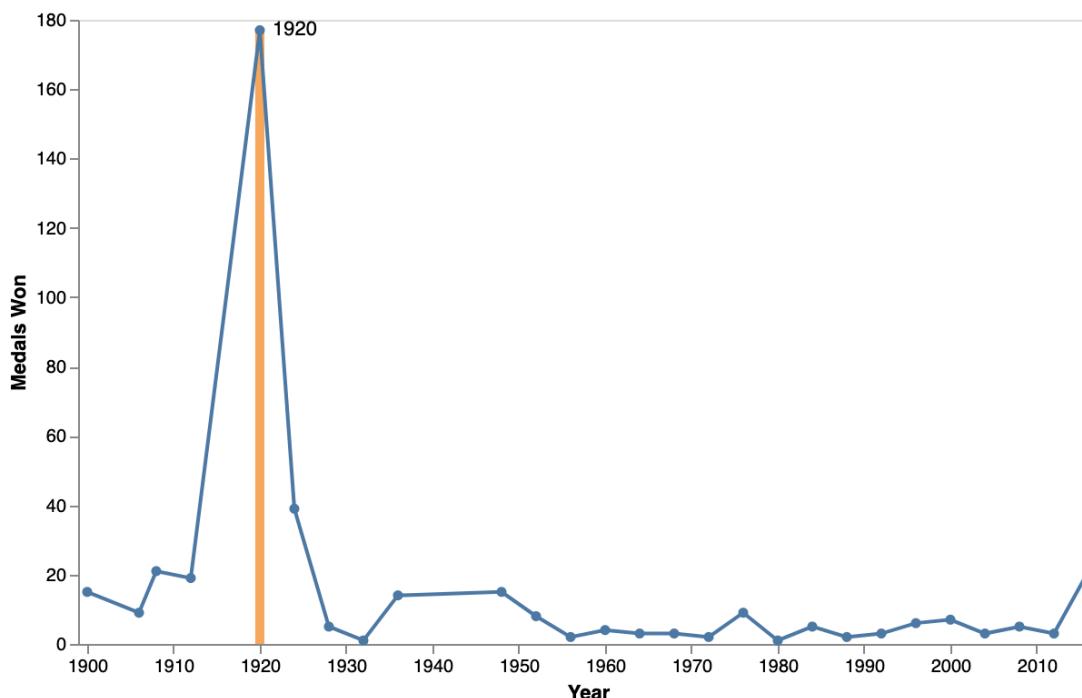
Summer Olympics hosted by Australia



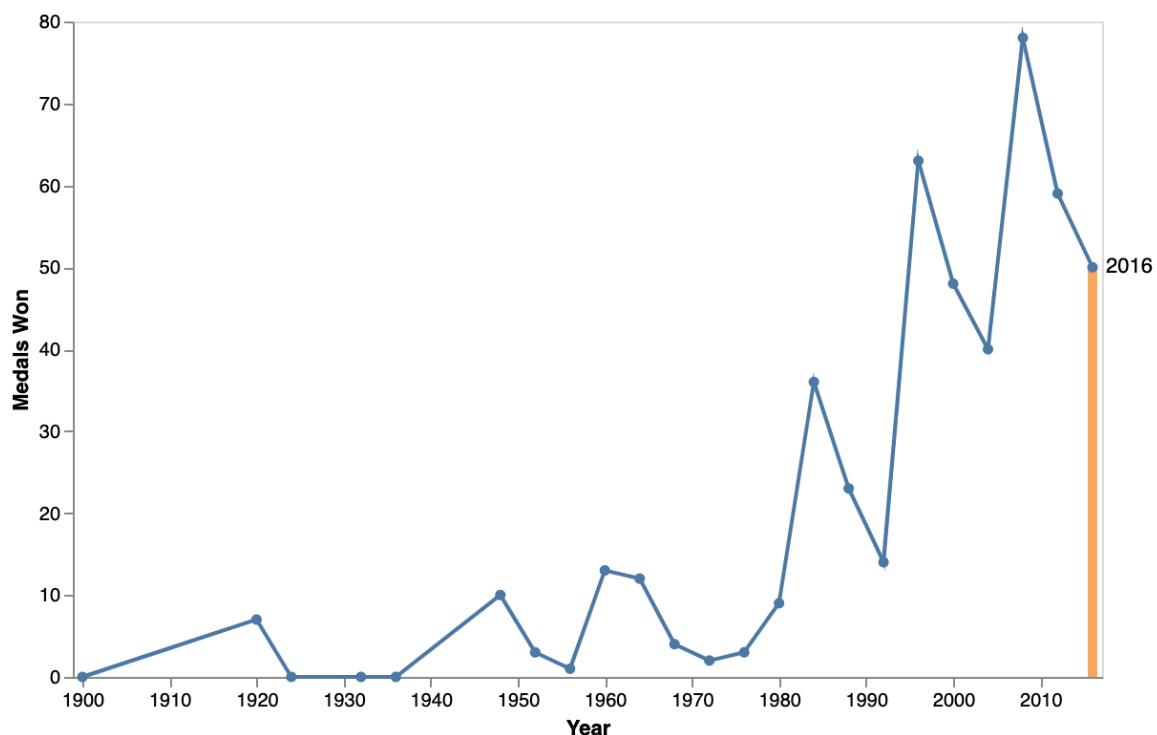
Winter Olympics hosted by Austria



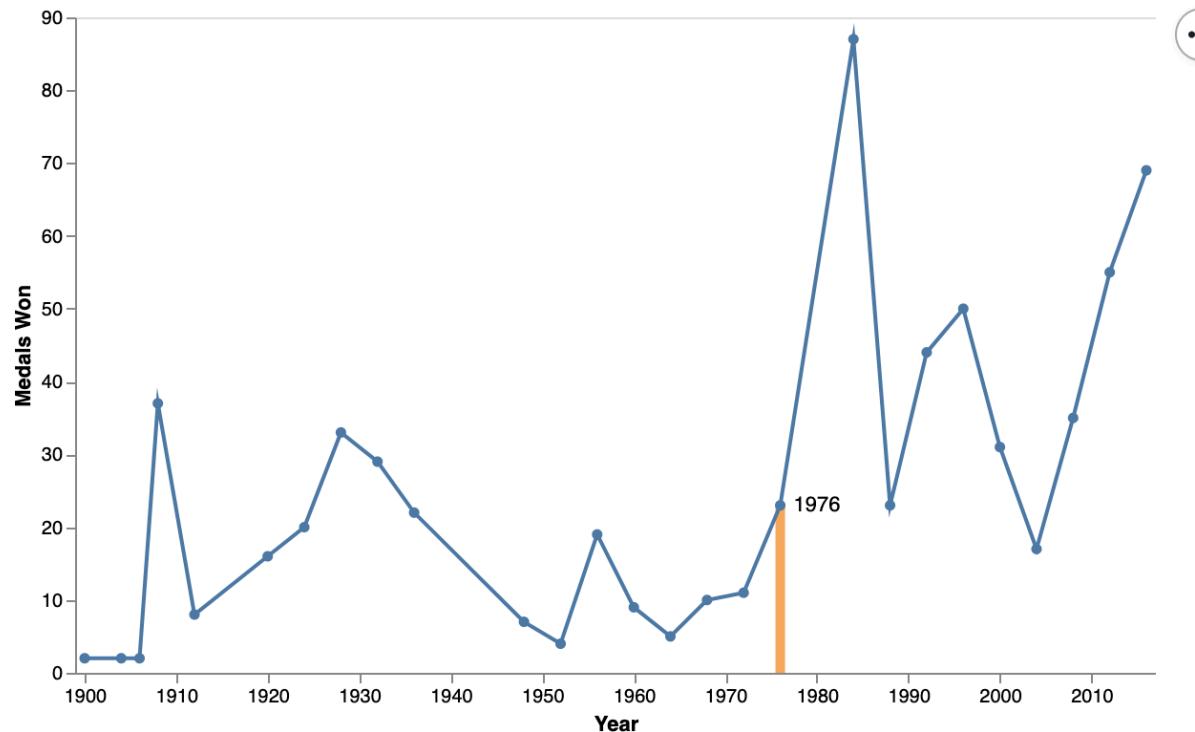
Summer Olympics hosted by Belgium



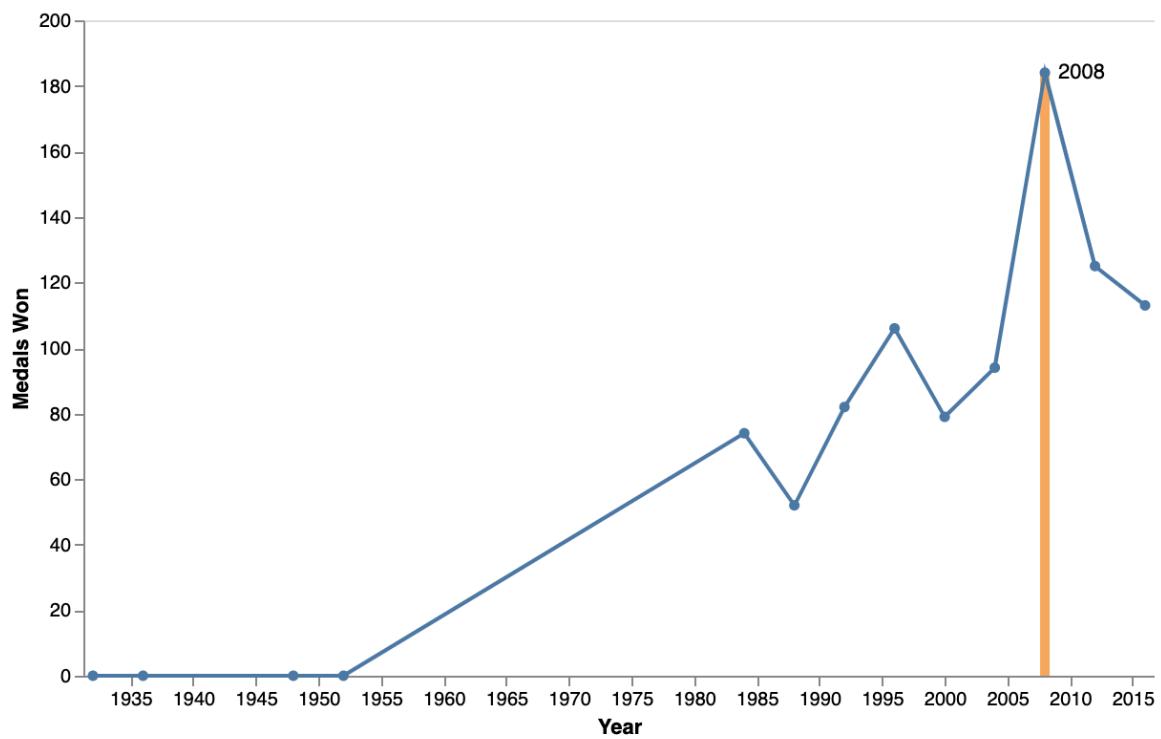
Summer Olympics hosted by Brazil



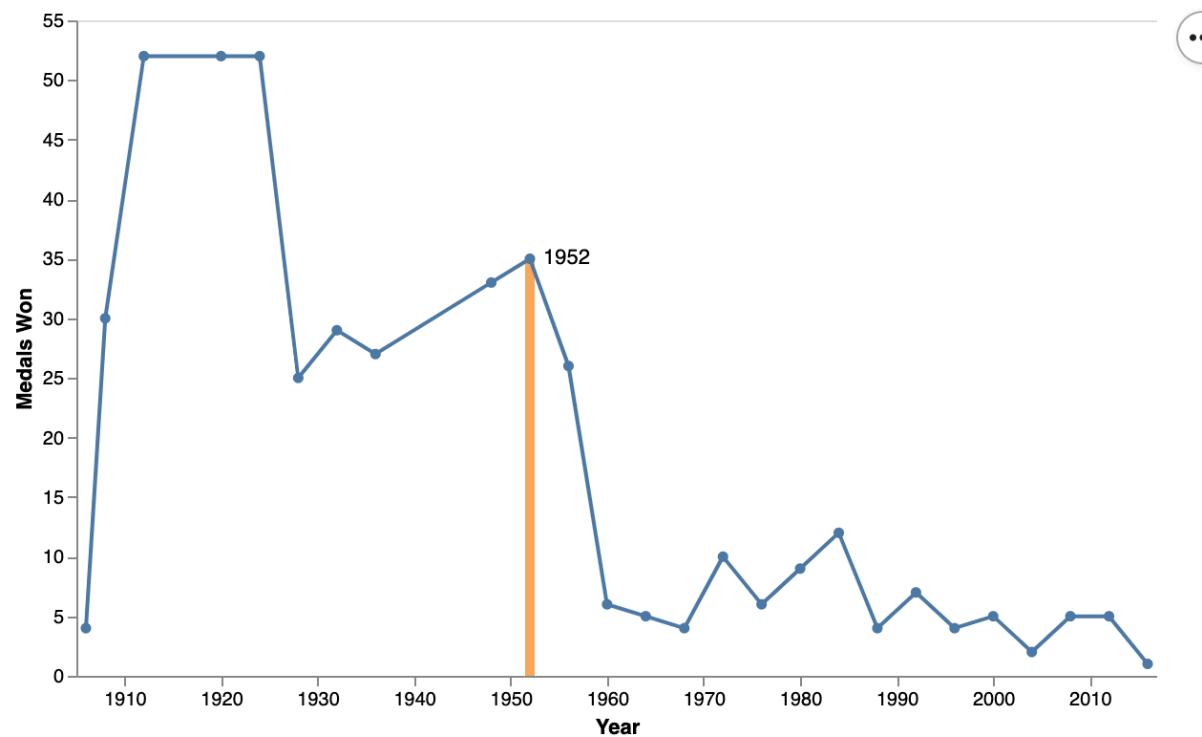
Summer Olympics hosted by Canada



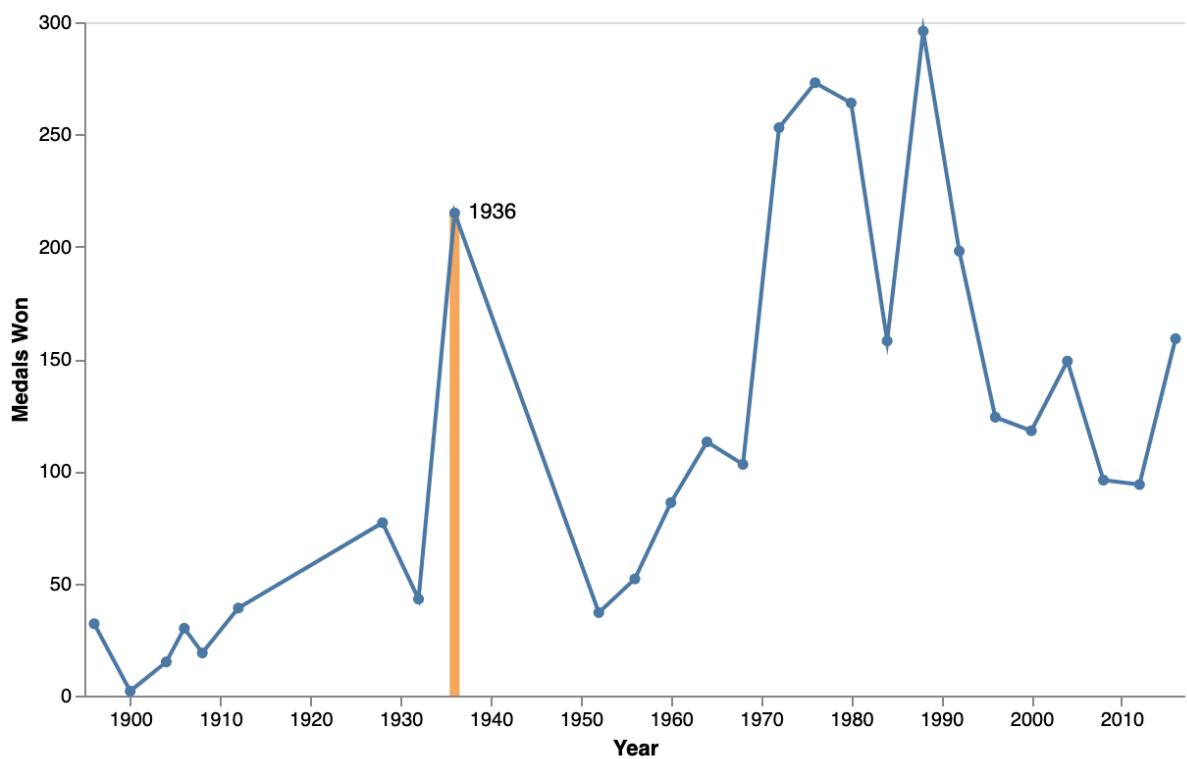
Summer Olympics hosted by China



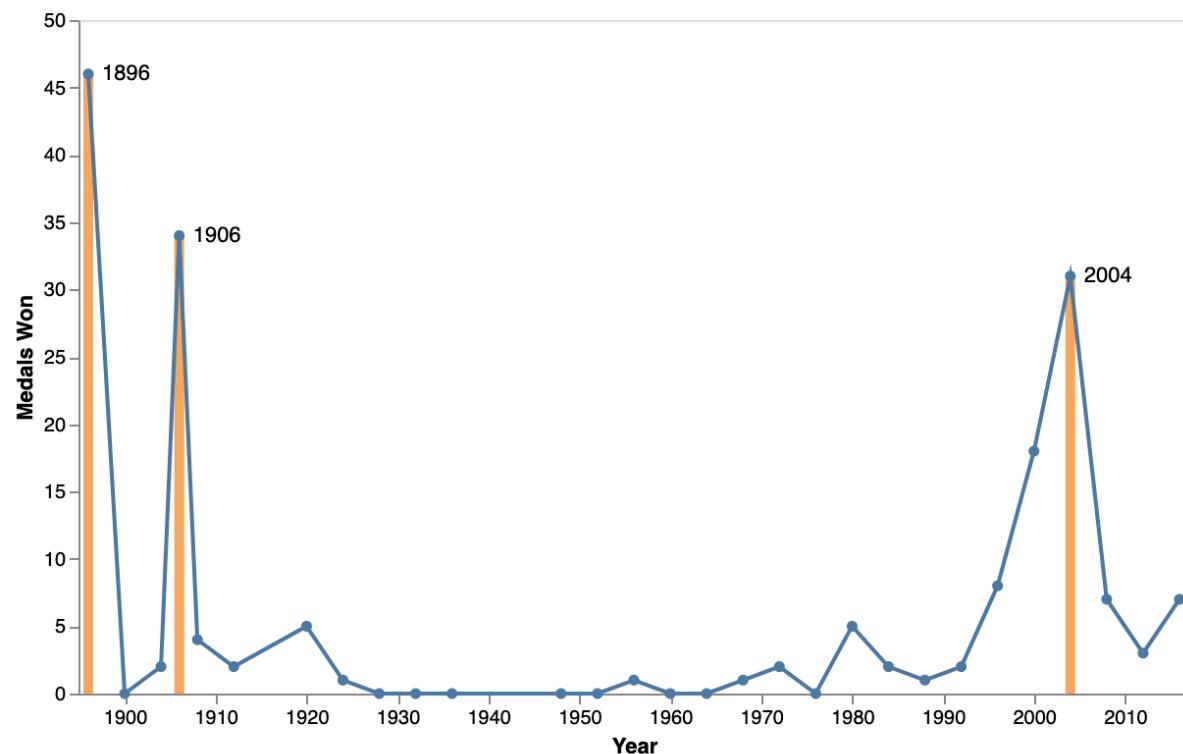
Summer Olympics hosted by Finland



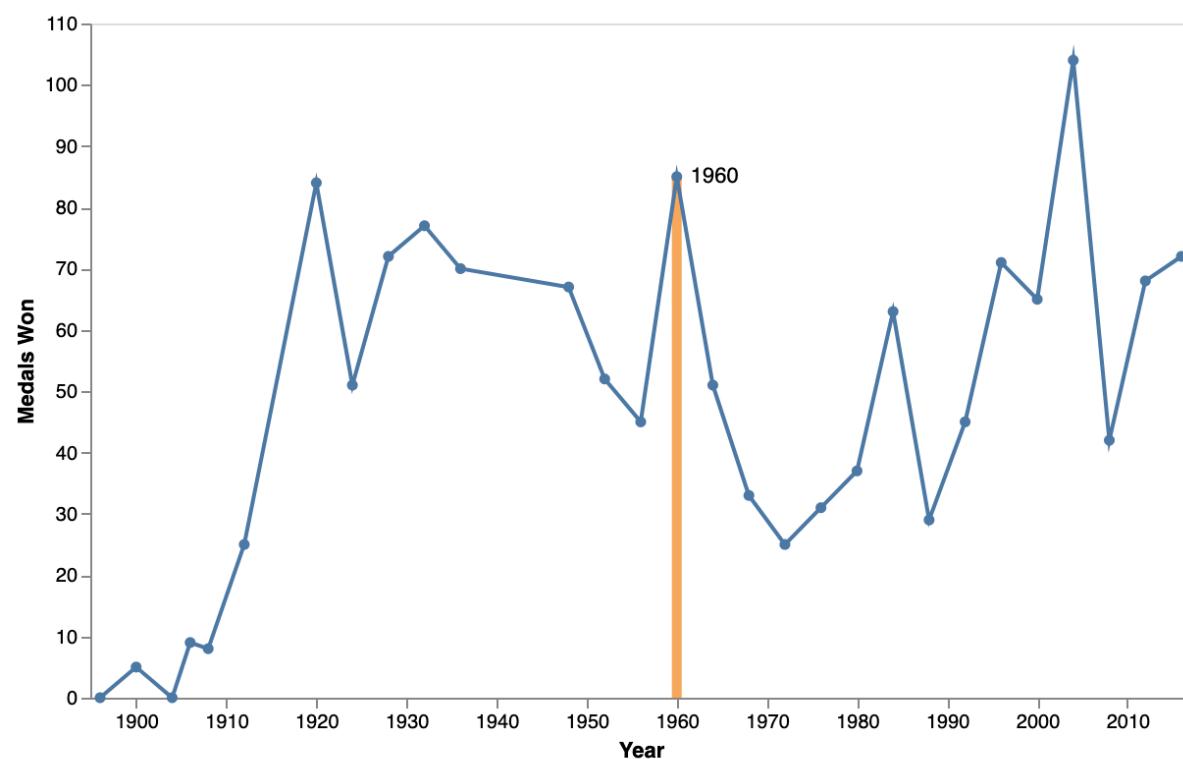
Summer Olympics hosted by Germany



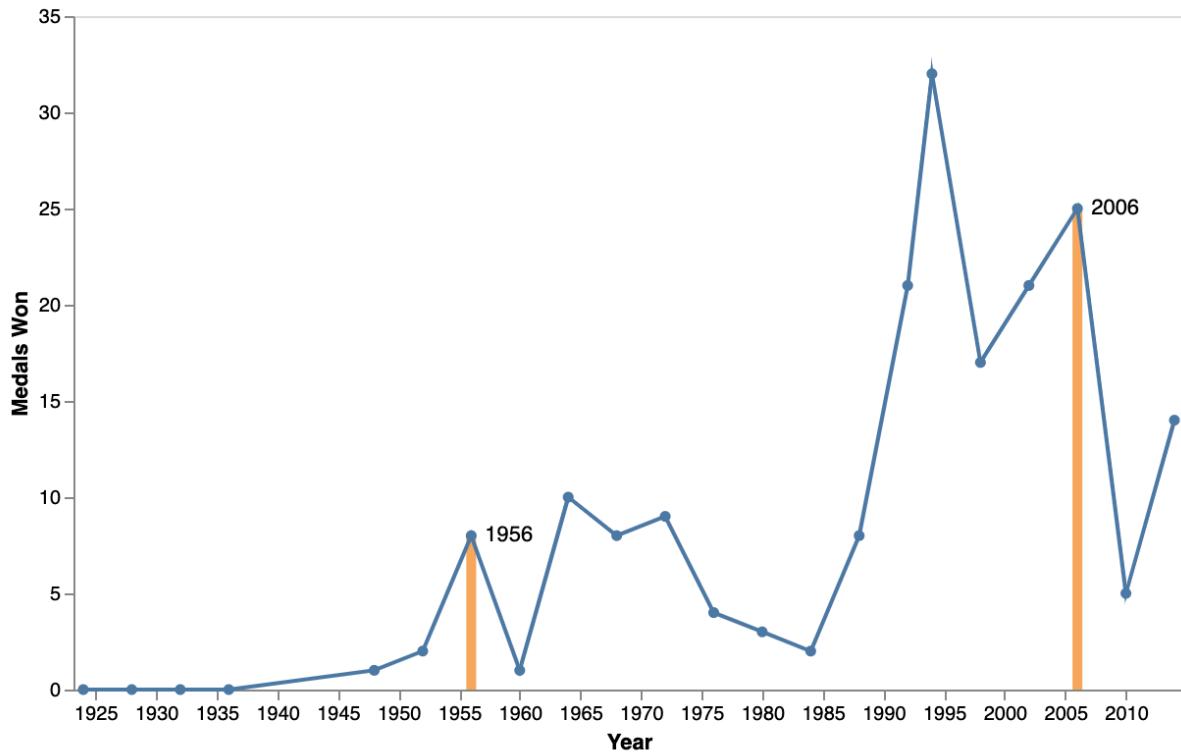
Summer Olympics hosted by Greece



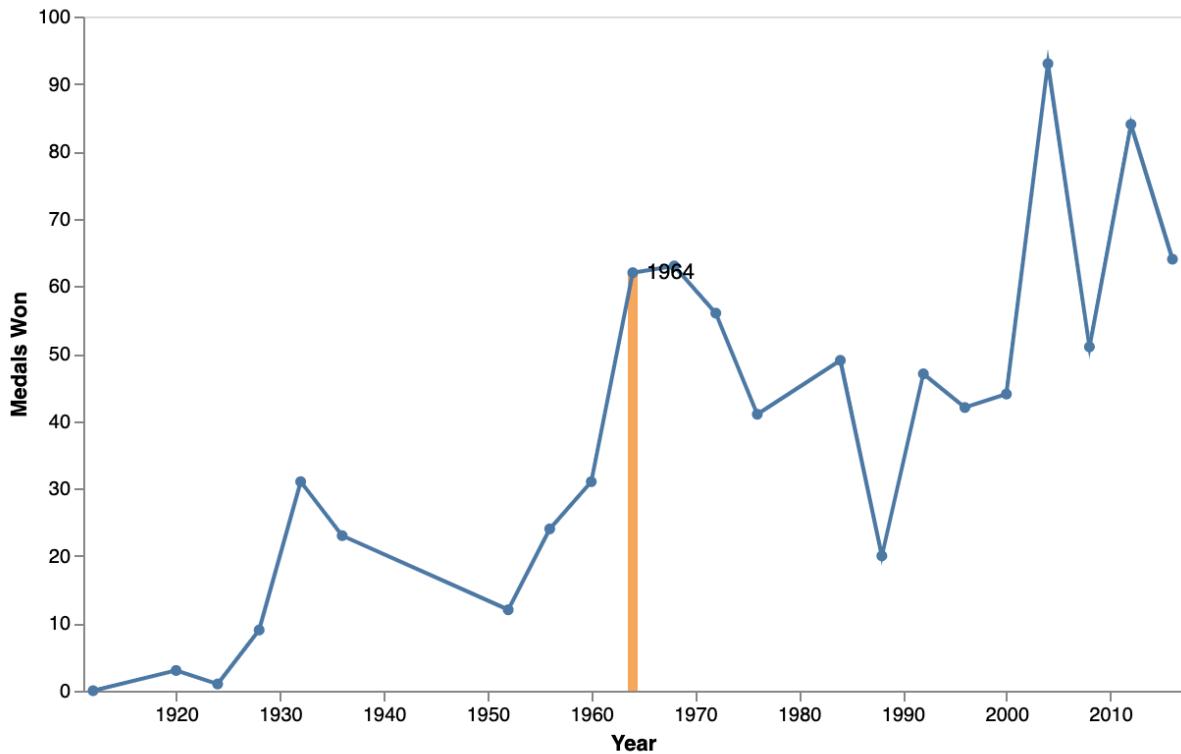
Summer Olympics hosted by Italy



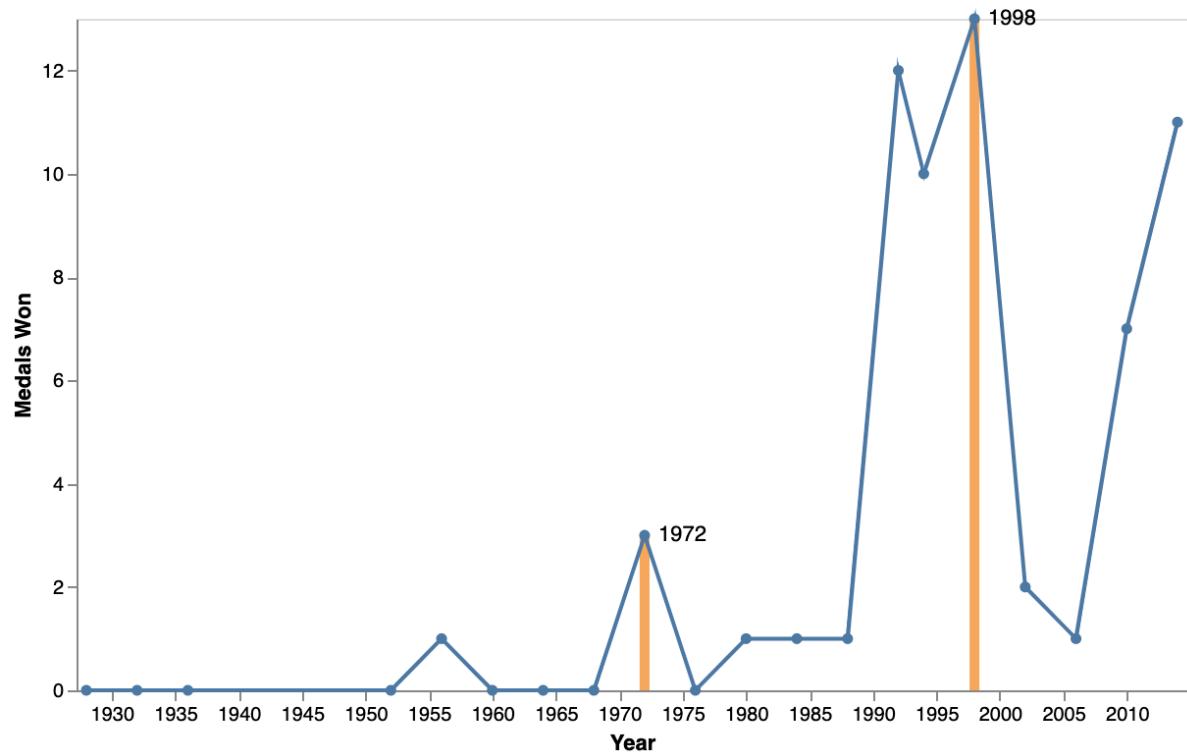
Winter Olympics hosted by Italy



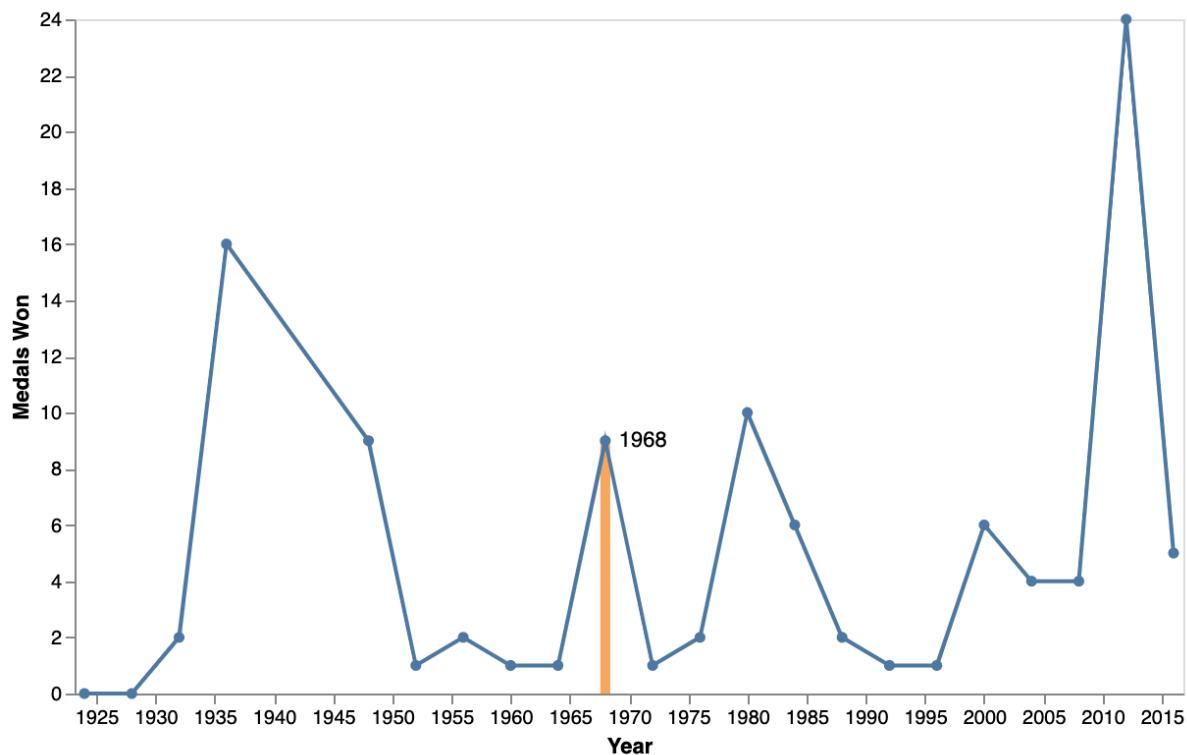
Summer Olympics hosted by Japan



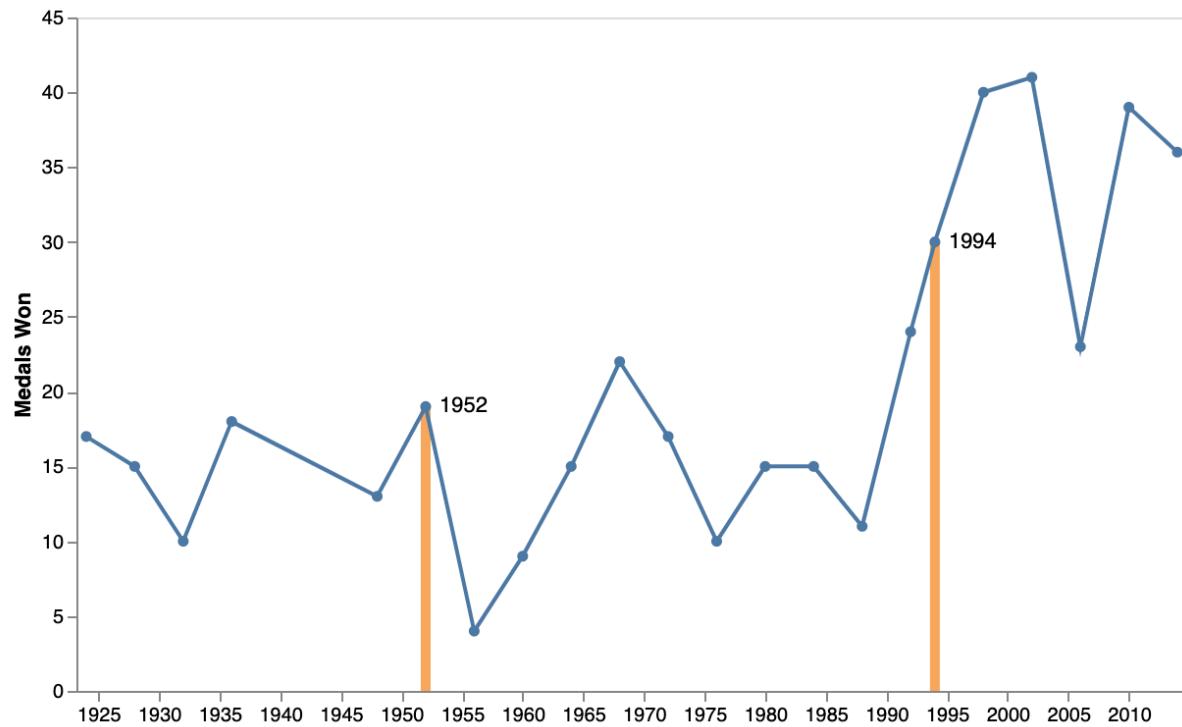
Winter Olympics hosted by Japan



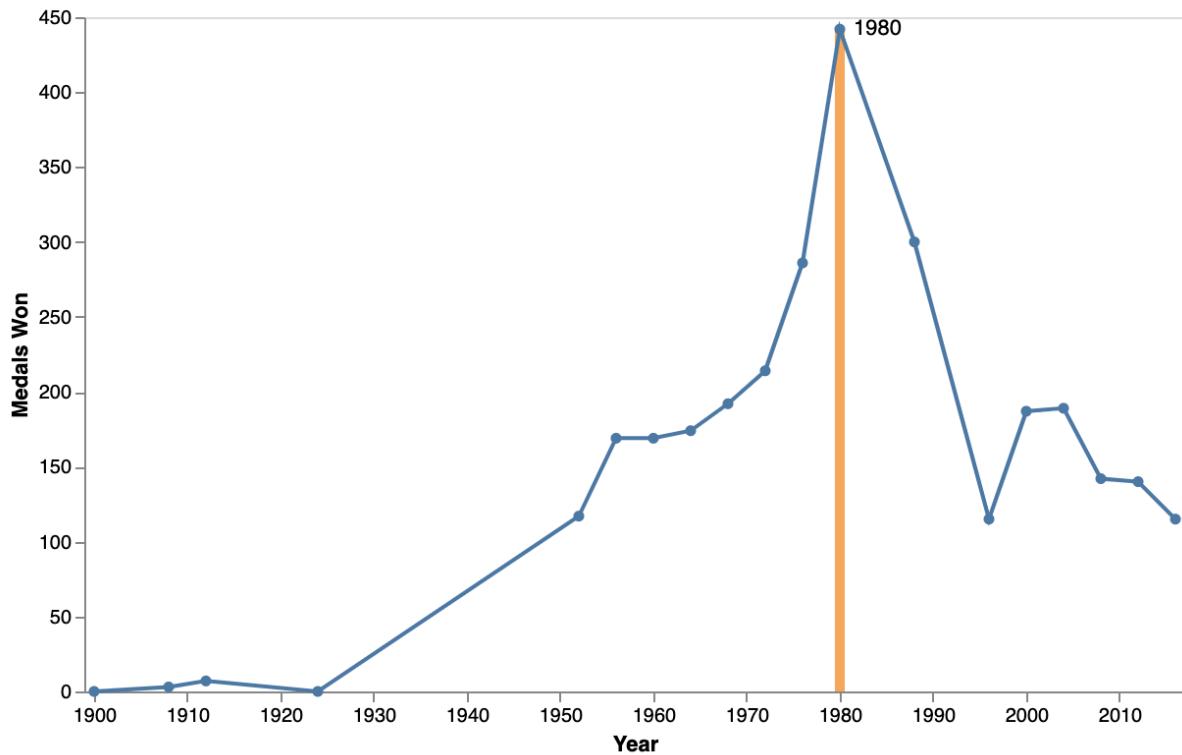
Summer Olympics hosted by Mexico



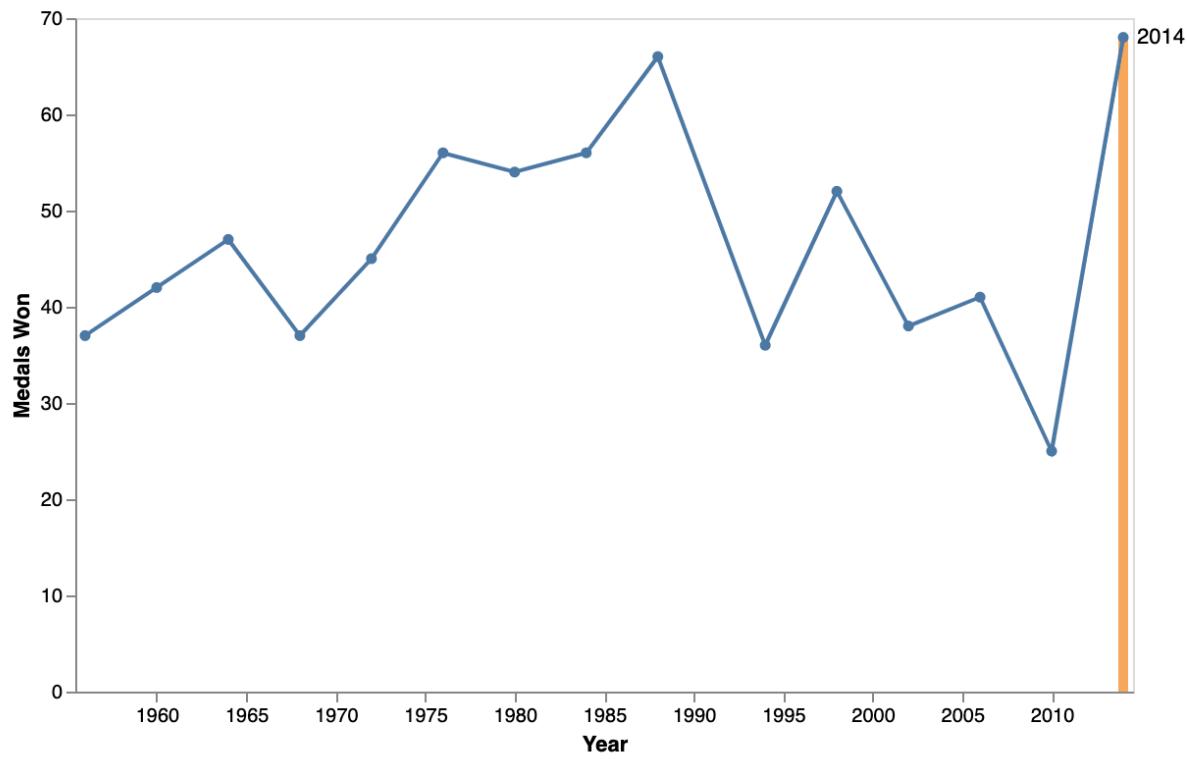
Winter Olympics hosted by Norway



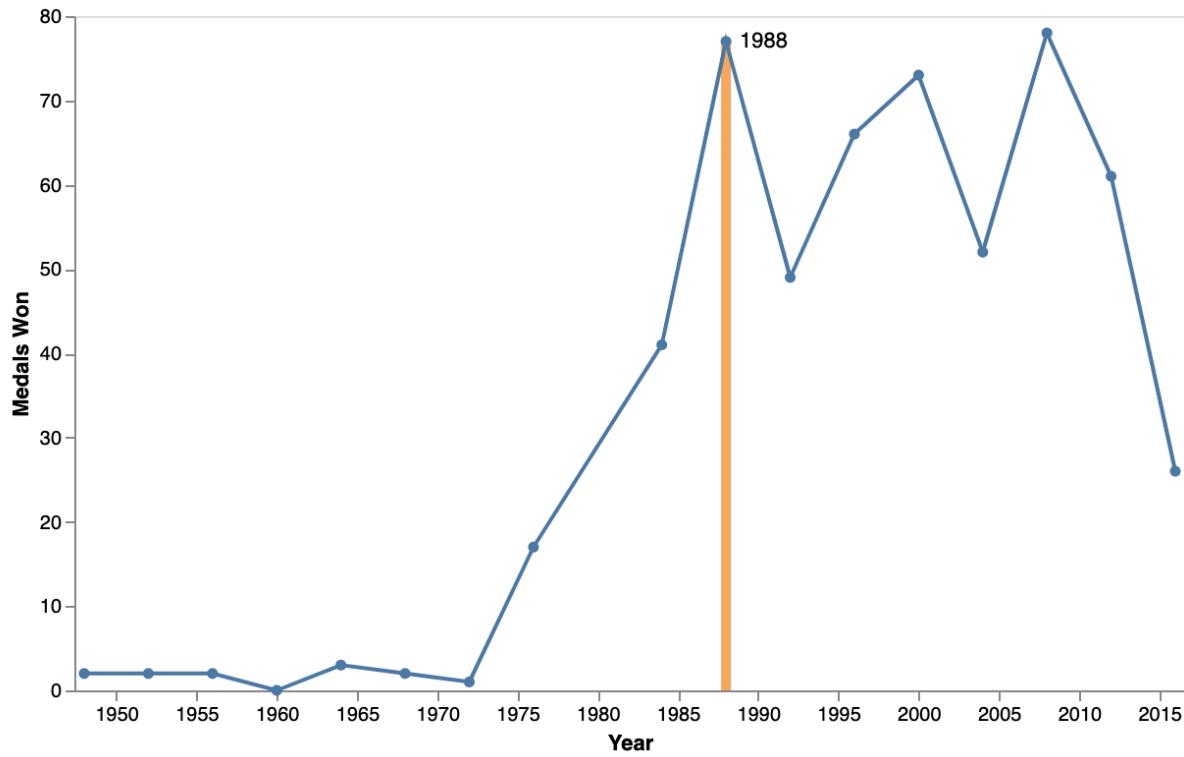
Summer Olympics hosted by Russia



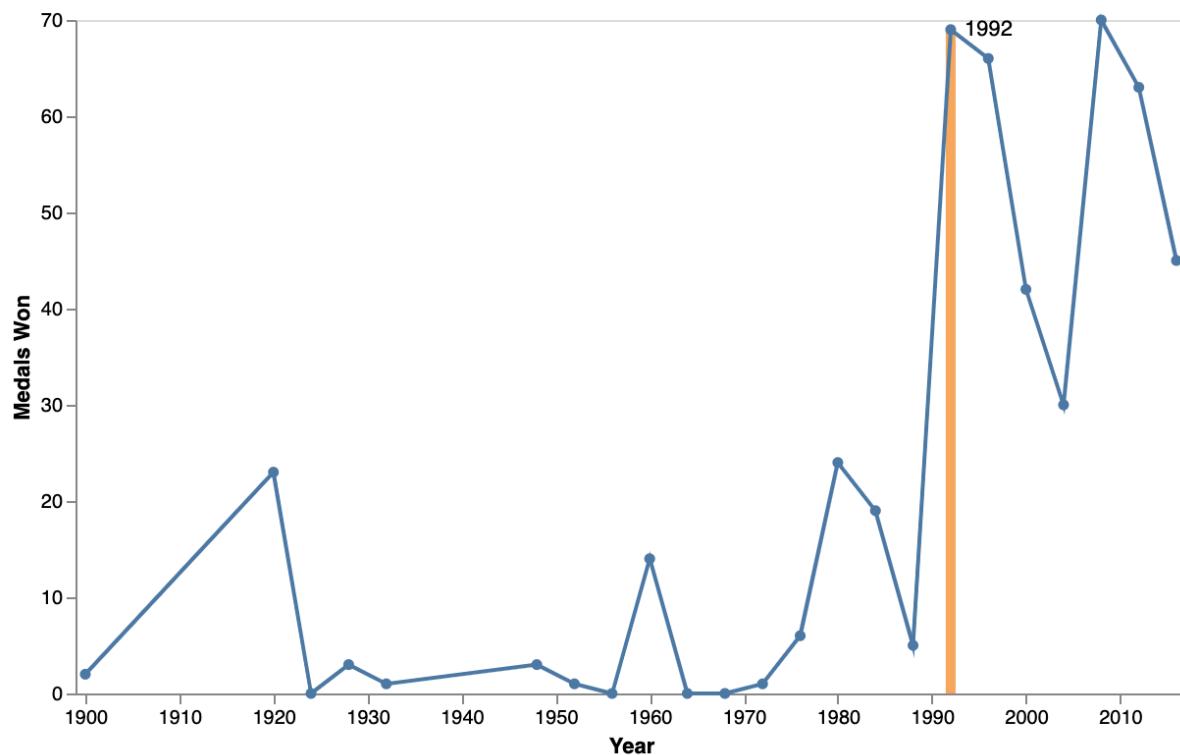
Winter Olympics hosted by Russia



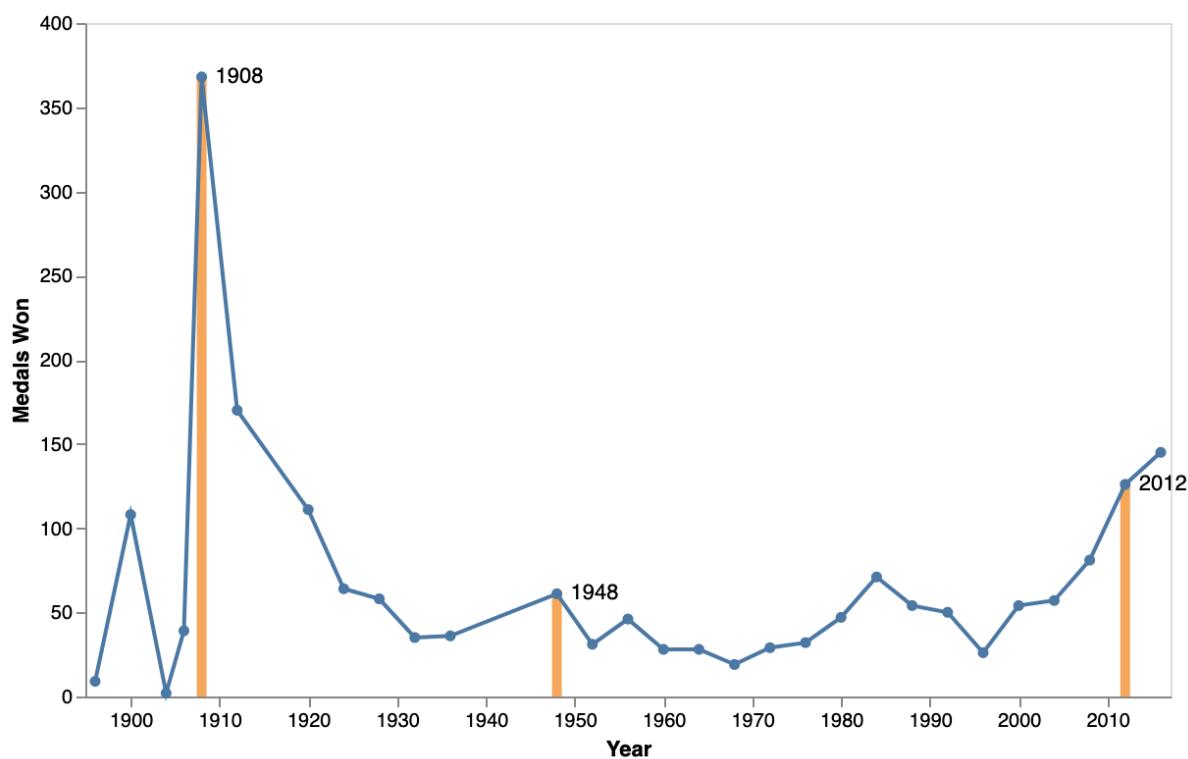
Summer Olympics hosted by South Korea



Summer Olympics hosted by Spain



Summer Olympics hosted by United Kingdom

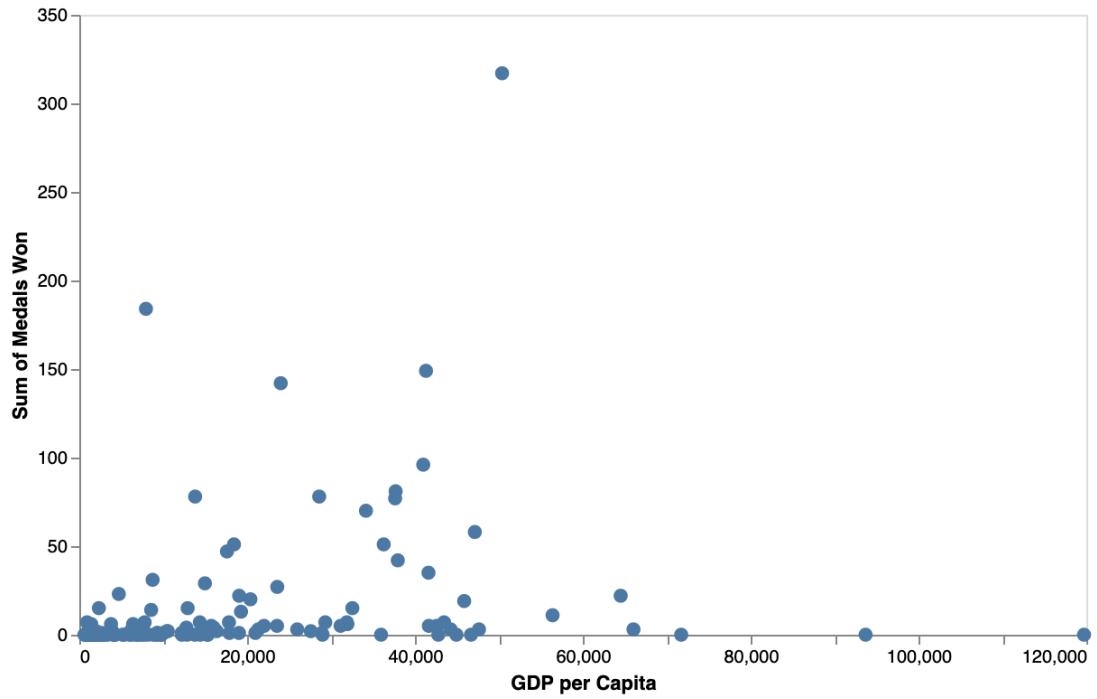


Appendix II- Correlation Values for GDP and GDP Per Capita

Correlation Scatter Plots

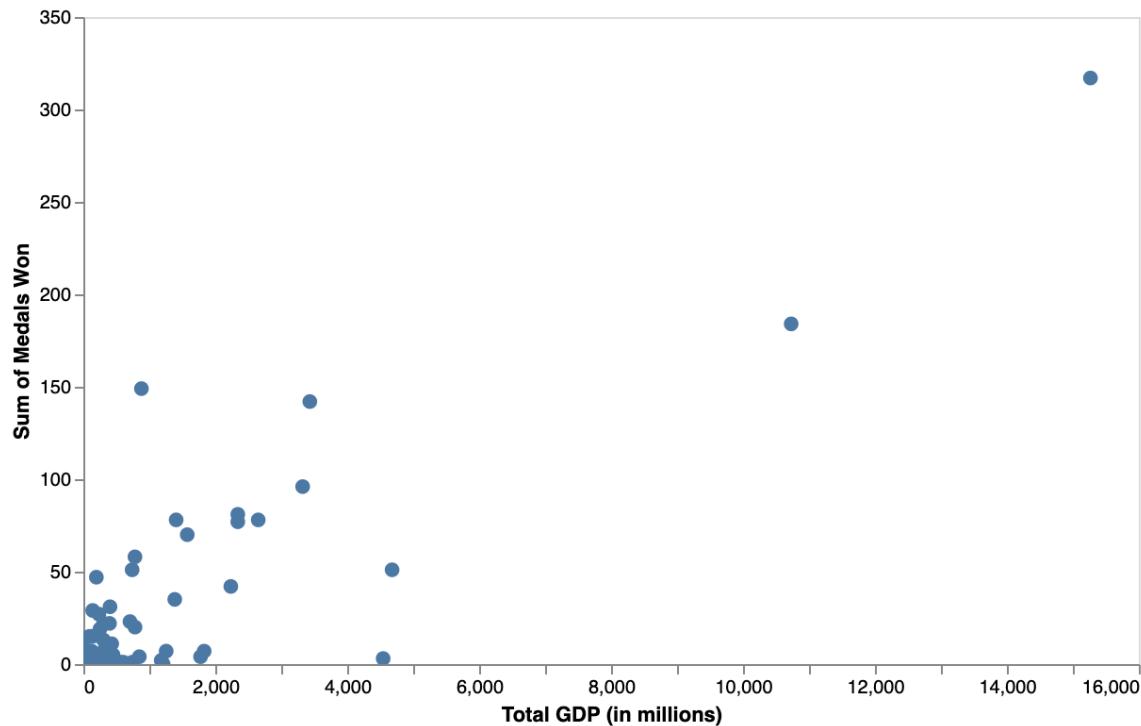
Medals Won vs GDP Per Capita in the year 2008

Correlation Coefficient: 0.08



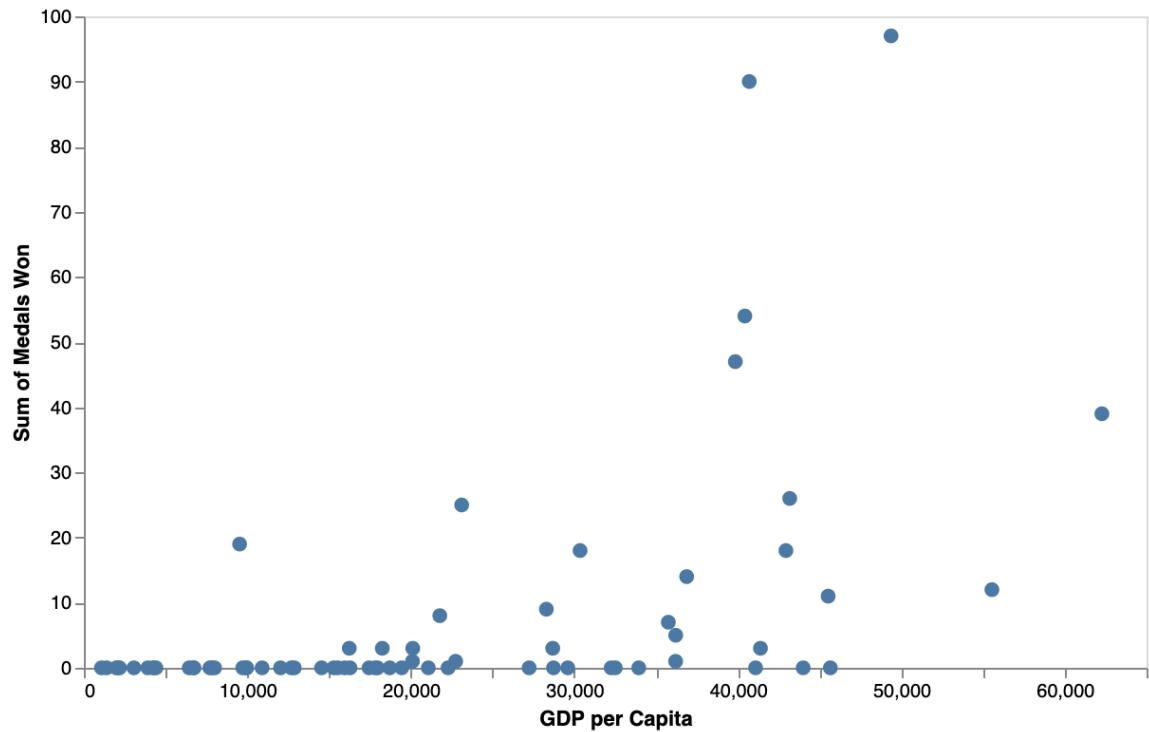
Medals Won vs Total GDP in the year 2008

Correlation Coefficient: 0.21



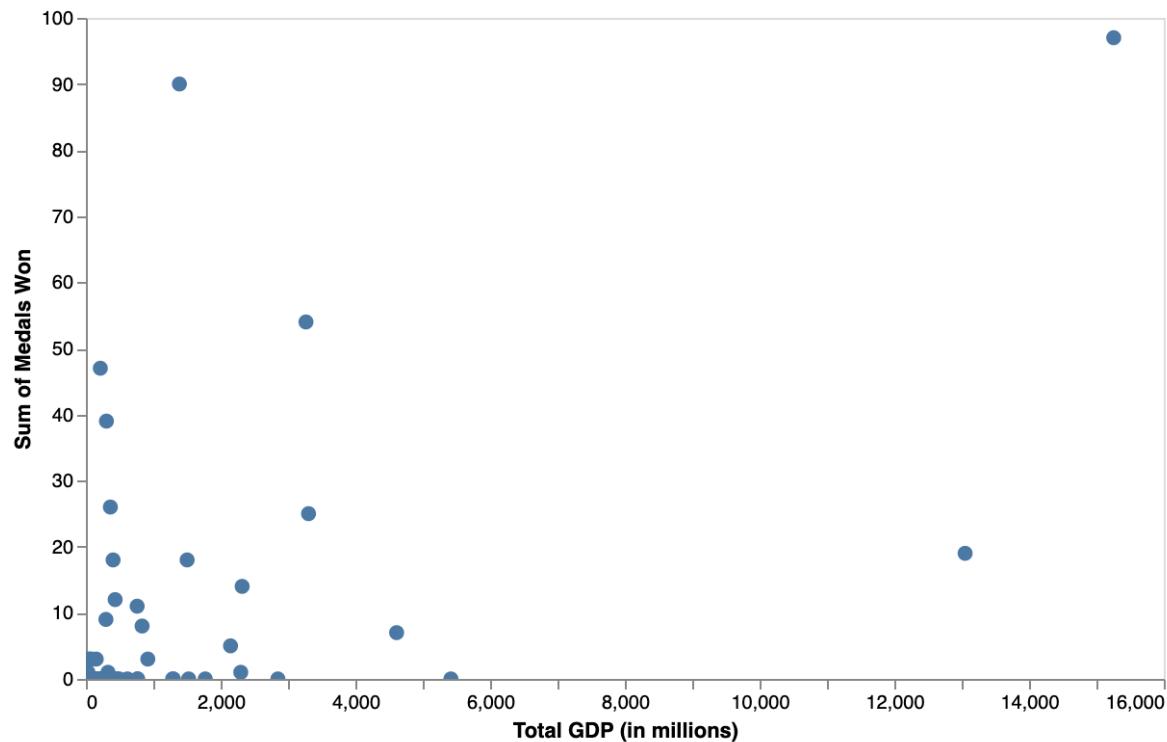
Medals Won vs GDP Per Capita in the year 2010

Correlation Coefficient: 0.18



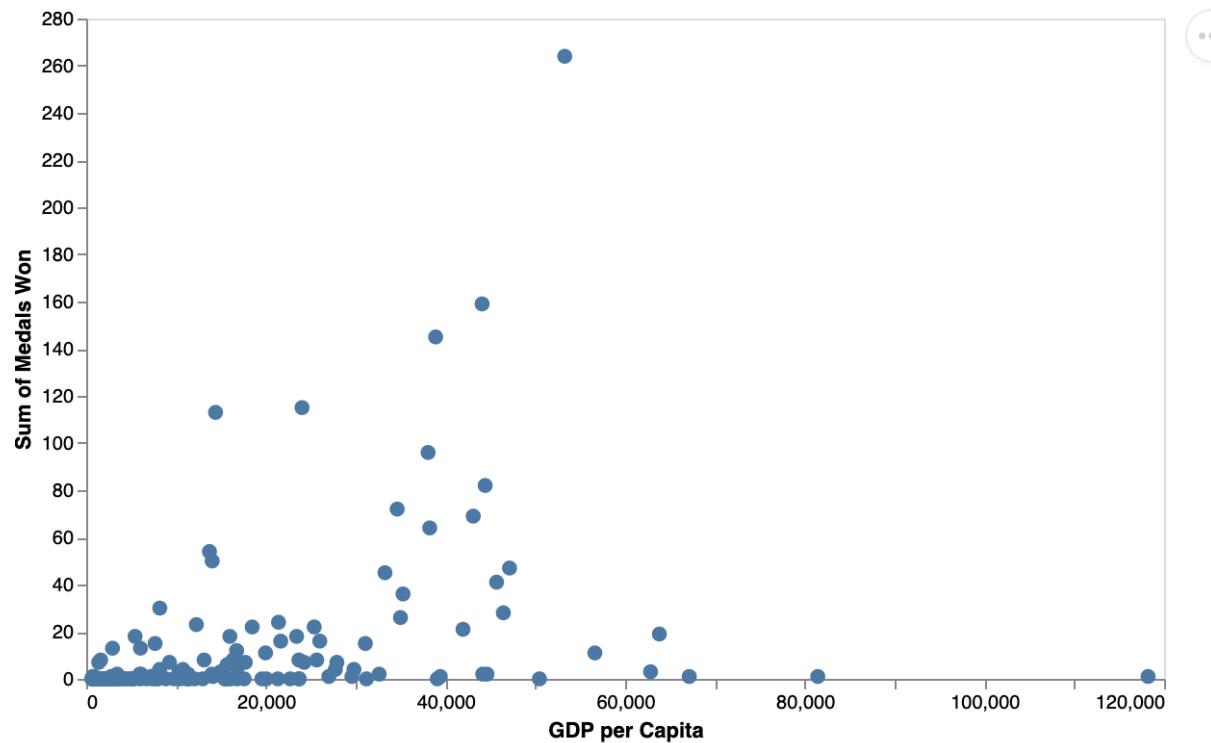
Medals Won vs Total GDP in the year 2010

Correlation Coefficient: 0.14



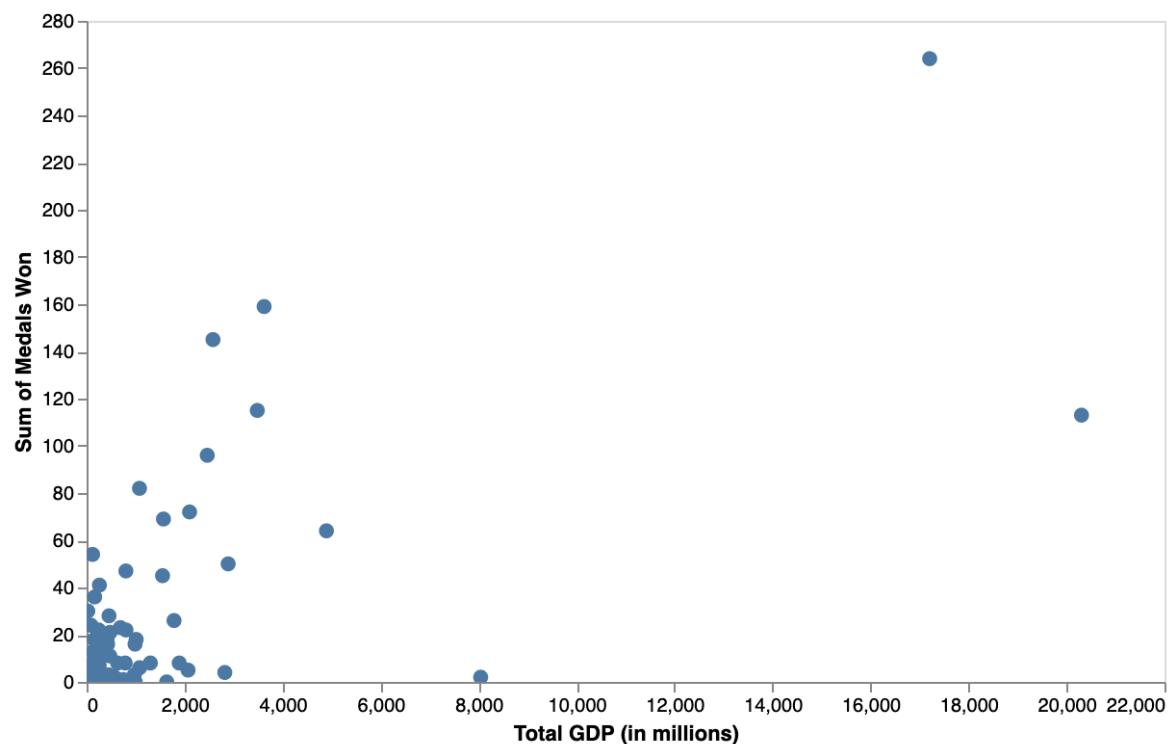
Medals Won vs GDP Per Capita in the year 2016

Correlation Coefficient: 0.11



Medals Won vs Total GDP in the year 2016

Correlation Coefficient: 0.15



Correlation Data- GDP, Population, GDP Per Capita

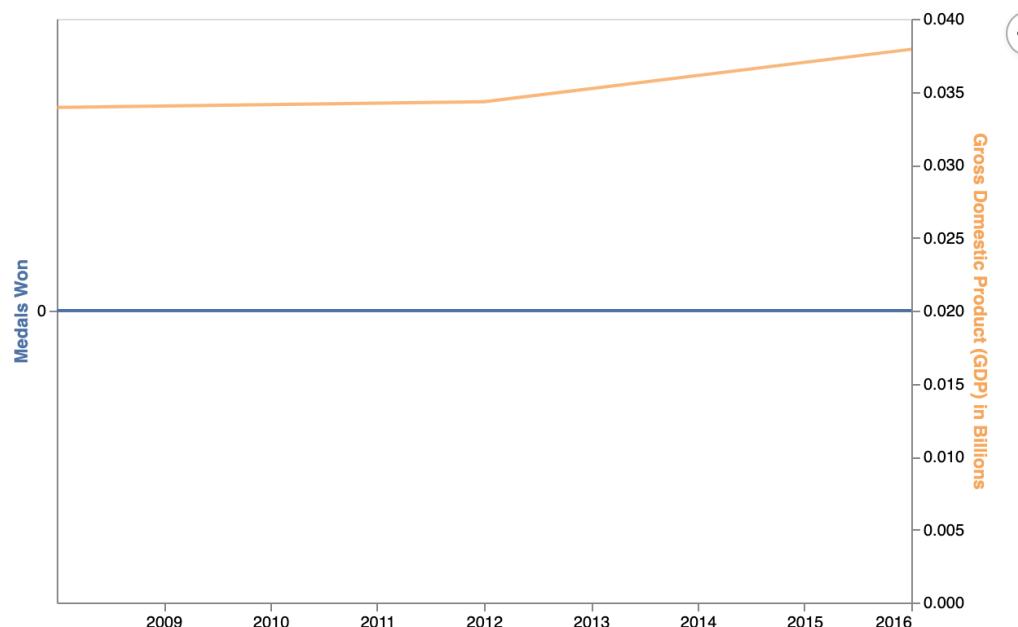
	Year	Corr_GDP	Corr_Popn	Corr_GDP_pc
0	1896-01-01	0.1482556396167299	0.12385713197190039	0.11846139970776402
1	1900-01-01	0.26792362626670096	0.09470385667315558	0.3791930555197616
2	1904-01-01	0.005948945339004744	0.0009640730958988231	0.013833059230800882
3	1906-01-01	0.08528332375323115	0.07091293373660627	0.12963199969332212
4	1908-01-01	0.11014020740686996	0.06039741946107625	0.13469856026136964
5	1912-01-01	0.05252786064952849	-0.015065802797246284	0.17036173286164302
6	1920-01-01	0.08960793601237702	0.06760417347524574	0.07999295162784635
7	1924-01-01	0.19255664820589127	0.12906081414306	0.15016322756776943
8	1928-01-01	0.08460473455503692	0.12216495024341524	0.0285489250704508
9	1932-01-01	0.08020133898928913	0.10175890661576996	0.01989051603573178
10	1936-01-01	0.1306380763450608	0.07631182549721102	0.07669018458596673
11	1948-01-01	0.1848550784010514	0.10238947913501707	0.09429437491998845
12	1952-01-01	0.18324530218095667	0.14204614940841073	0.10291062679733474
13	1956-01-01	0.14656239670854	0.1277088343704316	0.07941785099346724
14	1960-01-01	0.22955101838398856	0.23269373589822728	0.058973595105263975
15	1964-01-01	0.25559800559685436	0.18953781105475576	0.1287639714573382
16	1968-01-01	0.2350936155005985	0.2661856057018205	0.11390383046890479
17	1972-01-01	0.24119973119381136	0.24728046773295392	0.04133802451901201
18	1976-01-01	0.22041639197787502	0.20645220460086833	0.009781487293699475
19	1980-01-01	0.4343201834795484	0.2124990729183301	0.004178597680168503
20	1984-01-01	0.32066815742634874	0.12731596845044488	0.09362372090949984
21	1988-01-01	0.20421549498506217	0.060476065774674036	0.06040559996927415
22	1992-01-01	0.16734526900634708	0.07789011726449438	0.054891175178935894
23	1996-01-01	0.14927735670188508	0.09048416261399049	0.06793423852021203
24	2000-01-01	0.1352328261569431	0.05201608006961378	0.080627604504598
25	2004-01-01	0.1730841890431268	0.051686559918290076	0.085146470103365
26	2008-01-01	0.20618625179687775	0.10345919613592047	0.0773985706617476
27	2012-01-01	0.17415596464888447	0.08858794147294878	0.1069147579840293
28	2016-01-01	0.15159696580085594	0.058637377863085476	0.10893468661545762

Highest and Lowest GDPs

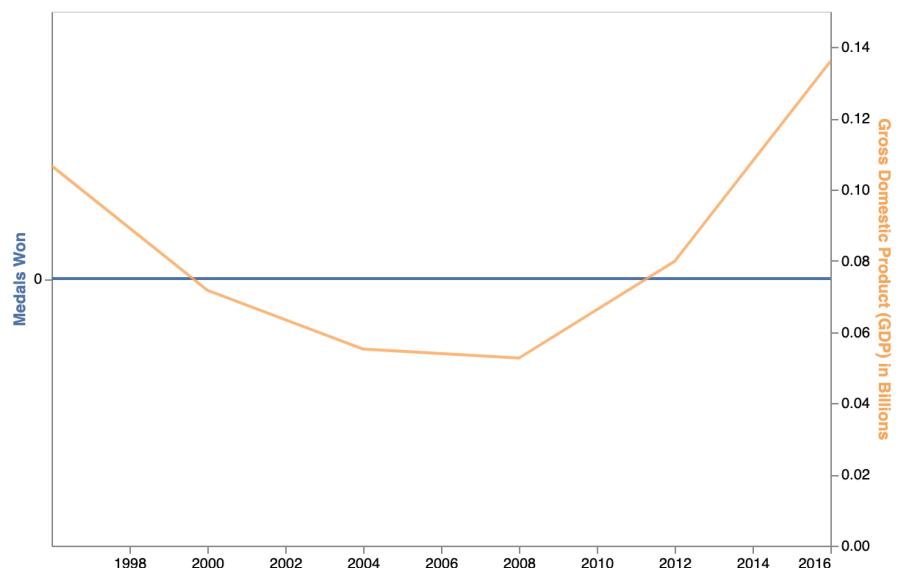
Bottom GDP Countries

	Country Name	Year	GDP_pc	country	Population	GDP
24426	Tuvalu	2016	3385	Tuvalu	11200	0.0379
24366	Nauru	2016	12951	Nauru	10500	0.1360
24332	Kiribati	2016	1898	Kiribati	113000	0.2145
24353	Marshall Islands	2016	3775	Marshall Islands	57700	0.2178
24376	Palau	2016	14251	Palau	17700	0.2522

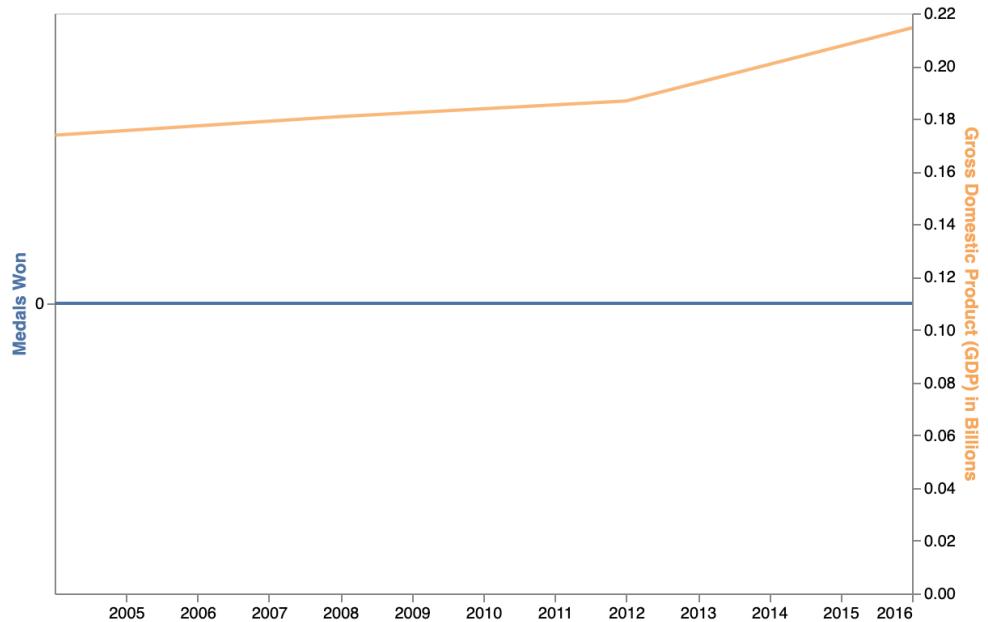
Country: Tuvalu



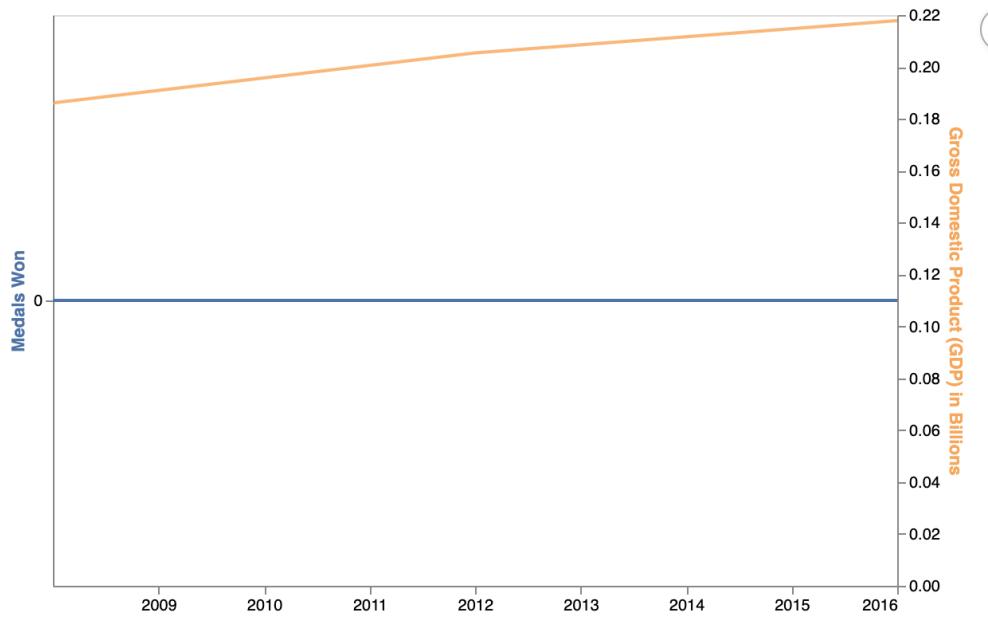
Country: Nauru



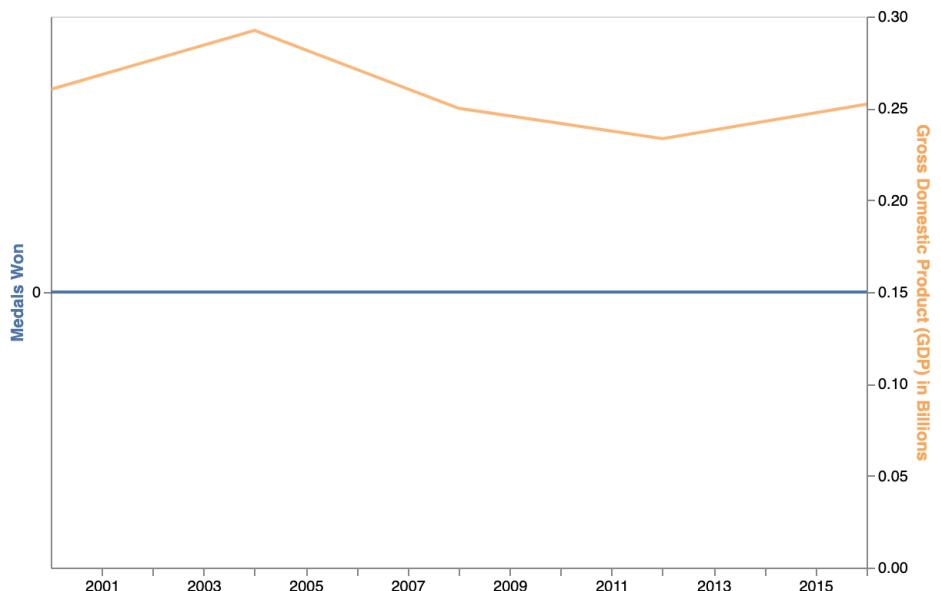
Country: Kiribati



Country: Marshall Islands



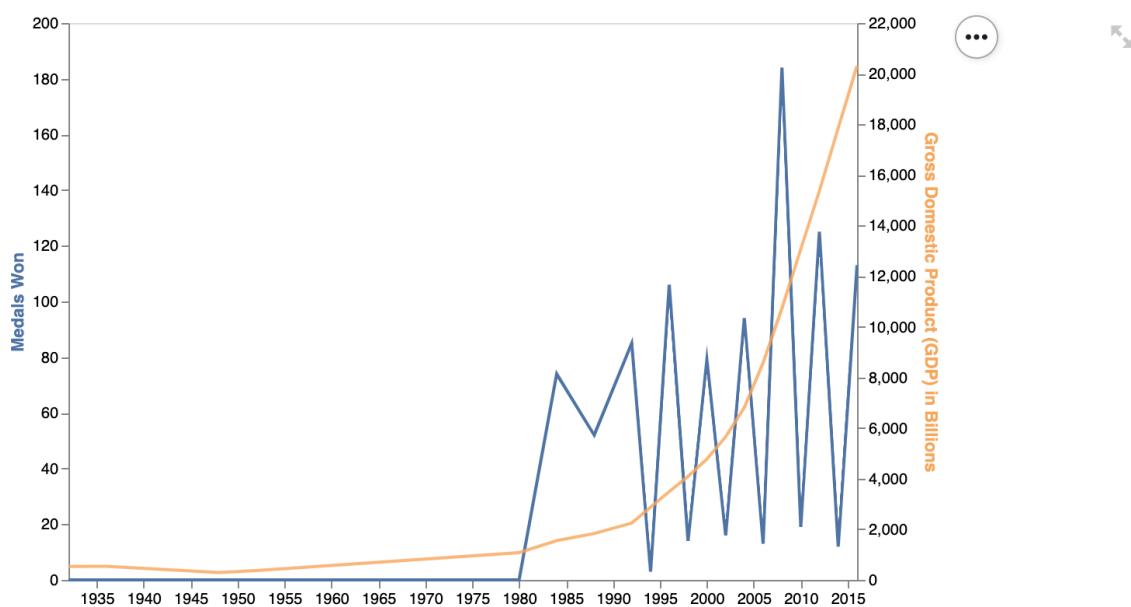
Country: Palau



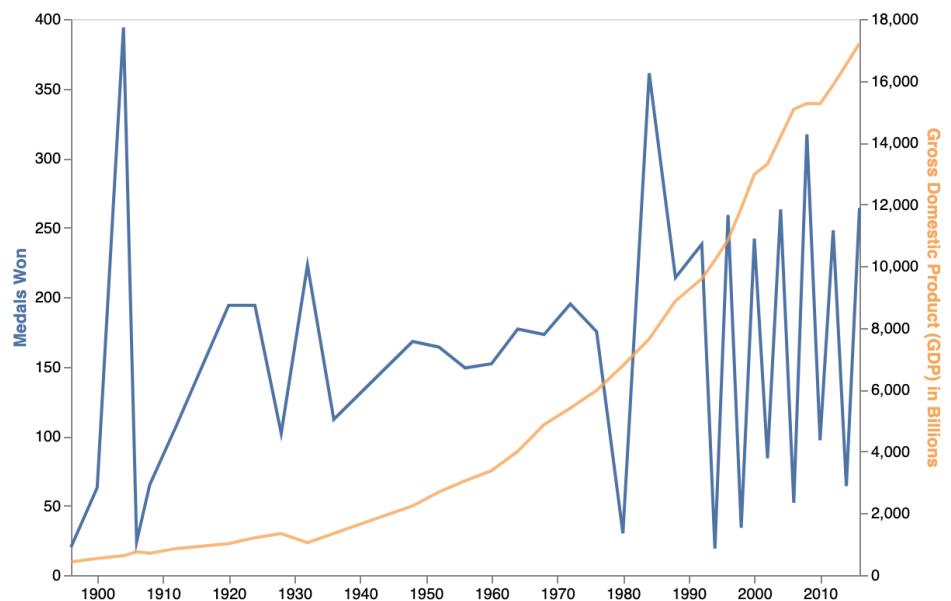
Top GDP Countries

	Country	Name	Year	GDP_pc	country	Population	GDP
24278		China	2016	14401	China	1410000000	20,305.4100
24431	United States	United States	2016	53273	United States	3230000000	17,207.1790
24320		India	2016	6093	India	13200000000	8,042.7600
24328		Japan	2016	38240	Japan	1280000000	4,894.7200
24306		Germany	2016	44072	Germany	82200000	3,622.7184

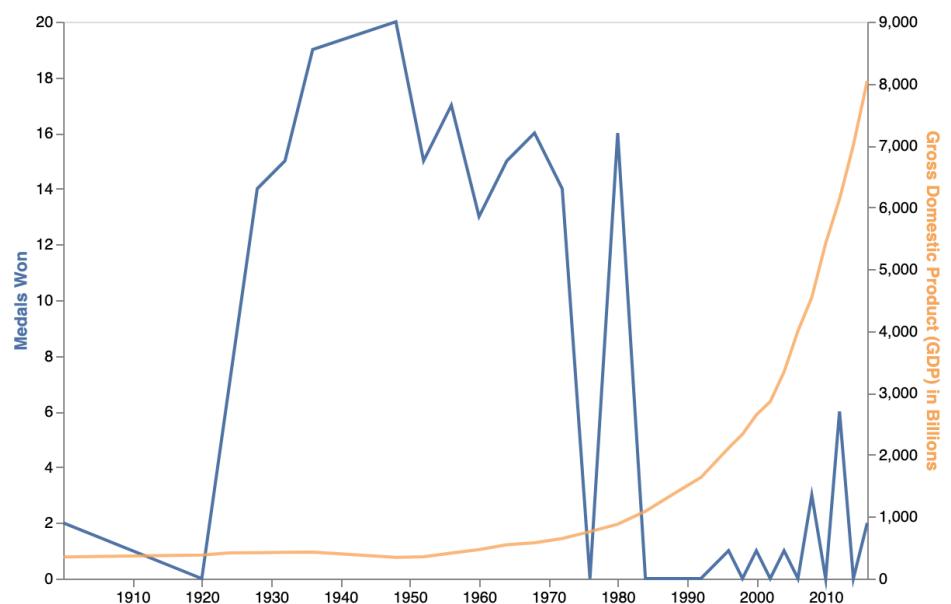
Country: China



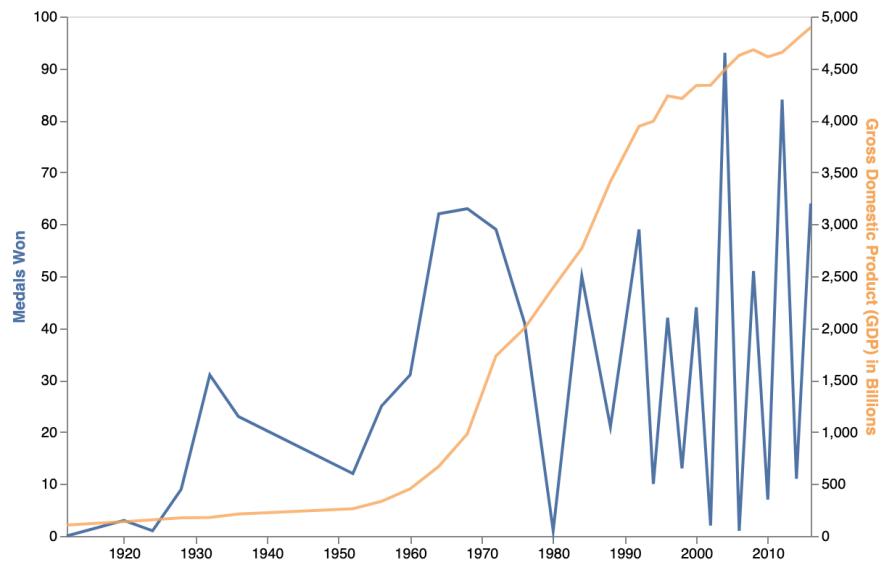
Country: United States



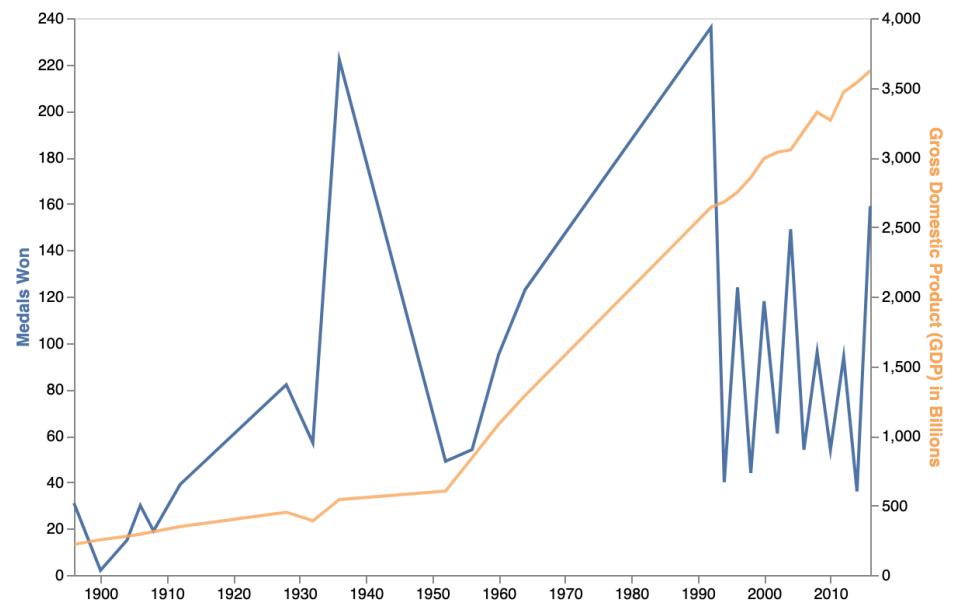
Country: India



Country: Japan



Country: Germany



Appendix III- Age range for all sports

Avg Age- Stable Era- All Sports

