A Brief Analysis of the Olympics and Climate

Travis Mark and Ellie Sullivan

Ritchie School of Engineering, University of Denver

Code Repository:

https://github.com/speedytmark/Olympic_Climate_Project

Abstract:

Every two years, the world tunes in for the most elite athletic competitions, the Summer and Winter Olympic Games. Athletes from around the world train for years for a chance to compete against the best of the best. As the world becomes more connected, participation and support for the Olympic Games has increased. Does climate have an effect on the participation in a particular season's Games? How about success? Climate change, also referred to as global warming, may be a factor in whether a country performs better (more gold medals) in the summer or the winter. Using an Olympic athletes dataset and a temperature dataset, we can analyze the relationship between climate and the Olympic Games. Using KMeans clustering on the temperature data, each country can be assigned a climate which, with the season of the Games, can be used to determine if the climate of the country impacts the participation or success of a team in a specific season of the world's largest athletic competition. Overall

Introduction:

The Olympic Games are originally attributed to ancient Greece and Greek mythology. The first Olympics were held in Olympia from 776 BC to 393 AD. In ancient times, the Olympic Games were religious as well as an athletic competition. The modern day Olympics were founded by Baron Pierre de Coubertin. Coubertin is credited with founding the International Olympic Committee in 1894 and with the first modern Games in 1896. The modern era of the Olympics has set a world stage for the best athletes from each country to compete for the coveted gold medal. The gold medal has become an icon that transcends cultures and crosses language barriers and is synonymous with the highest attainment of success. The Olympic Games themselves have become an international symbol of cooperation and sportsmanship. The number of countries participating in the Summer Olympic Games has increased since the Games were revived in the late 19th century. The winter events were added in the 1920s and have seen an increase in competing teams and athletes since then. Women first competed in the Olympics in 1900 but there were not many women at each of the games until years later. The summer and winter Olympics were held in the same year until 1992 when the winter Olympics was moved to two years after the summer Olympics.

World events have been reflected in Olympic participation over the years. For example, during both World Wars, the games were canceled due to the focus on battles across Europe. Wars and conflicts during the Olympics can be seen in the countries participating, specifically the countries who boycott the Games. The Olympic Games have also been a focus of innovation,

both in sport and technology. Whether it is searching to shave precious hundredths of a second off an athlete's time with an aerodynamic suit, or precision timing mechanisms, or television broadcasts, the Olympics have been a boon for many innovations. In 1936, the Berlin Olympics were first televised to a small group nearby. In 1964, the Olympics were broadcast on television worldwide. Similarly, in recent years the Olympics have been live streamed on the internet, bringing even more coverage and sports to people all across the world. With the Olympics being on the cutting edge of technology they have provided many people all over the world the opportunity to observe and learn about new or different sports and, like the Olympic flame, ignite passion for sport for new generations of athletes. Over time, populations have changed, but so have climates. In recent years, a large focus has been placed on climate change, also known as global warming, and the environment. The interests of a population are often related to the surrounding culture and environment. It stands to reason that countries with a warmer climate may perform better or have more athletes in the summer games than in the winter games, and vice versa for cold countries.

The following analysis will delve into the relationship between a country's climate and the country's participation in and performance at the Olympics. First and foremost, is there a relationship between climate and participation or between climate and success at the Olympics? Then, if there is a relationship, what is it? How have cold weather climate countries participation in Summer games and warm weather countries participation in Winter games changed over time? Does a country's climate impact its success at each seasons' games? As our world continues to experience the effects of climate change, how may the changing climate impact countries' participation in and performance at the Olympic Games? In order to answer these questions, the datasets must be processed, joined, and analyzed. To fully understand the relationship, various comparisons between summer and winter as well as comparisons between climates will be analyzed.

Literature Review:

Olympic History Data: Thorough Analysis:[1]

In this analysis, Olympic History Data: Thorough Analysis, the analyst reviewed major patterns in Olympics history. The trends of the number of competing athletes and nations, as well as the number of events in the summer and winter games were reviewed. Over time the number of athletes, nations, and events have increased in both seasons of the Olympics. In analyzing these trends, rgriffin was able to find correlations to major historical events. For example, there were two periods in which the Olympics were not held that correspond to World War I and World War II. The analyst noted that an explanation for the decrease in participation in the 1956 Melbourne Olympics could be the boycotts of Iraq, Egypt and Lebanon due to the Suez Crisis. The Netherlands, Spain, Switzerland and Cambodia did not participate because of the Hungarian Revolution and China did not compete due to the addition of Taiwan.

Women were not allowed to compete in the Olympic Games until 1900. There was not a significant increase in female olympic participation until the 1980s. Around 2010, the difference in the number of male and female athletes decreased to less than 2,000. The percentage of

[1] See Griffin for more information on the analysis of Olympics.

female athletes on the United States Olympic Team increased from approximately 12% in 1936 to over 25% in 1976 and in 2016 about 50% of the team were women.

The geographic distribution of athletes has also changed overtime. In Amsterdam in 1928, there were a handful of countries with over 100 athletes competing. By Rio in 2016, most of North America, South America, Europe and Asia had greater than 100 athletes competing. African countries have historically had fewer competitors but that number has been increasing.

Olympic Games Medal Count Analysis Summer and Winter Olympic Games:[2]

The thesis, Olympic Games Medal Count Analysis Summer and Winter Olympic Games, was prepared by Jiaxin Si. Si's analysis focused on the statistics of events, medals, hosting cities, winners' characteristics, and a gender comparison. The main data analysis portion of the thesis included a statistical model using GDP (Gross Domestic Product) and average high temperature.

Si used a linear regression model to analyze the relationship between the number of medals and the average high temperature in winter. Jiaxin explains that, "studies have shown that different ethnic groups in different climate zones have their own strength sports." According to the referenced studies, different climates are more conducive to endurance activities. The conclusion from this model is that there is a linear dependence between the number of medals and average high winter temperature. Moderate winter high temperatures are correlated to a higher count of medals.

Data Exploration of Historical Olympics Dataset:[3]

Pandas, seaborn, matplotlib, numpy, and pandasql were used in this analysis to investigate the sex, height, weight and age of athletes and how that affects the medals won. For example, athletes with height less than 140 cm and age less than 20 years were found to have won more medals. Using a maximum height of 140 cm and maximum age of 20 years, the athletes remaining competed in diving and gymnastics.

Dataset Description:

Olympics Dataset:[4]

The Olympics dataset contains information for each athlete who has competed in the Olympic Games from 1896 to 2016. The dataset contains descriptive information for the athlete including name, age, sex, height, weight and team. Each row also includes data identifying the Olympics (year and season), sport, and event. The success for each athlete (specific to the year, season, sport, and event) is recorded as well. Over the years, teams (and their National Olympic Committees, aka NOC) have changed. As country borders and alliances change, so do the teams and NOCs at the Olympics. The data are not complete, which is to be expected. Before analyzing the data, missing values must be addressed.

[2] See Si for more information on the olympics and temperature thesis.
[3] See Hargurjeet for more information on the Olympics dataset data exploration.
[4] See Chadha for more information on the dataset and analysis.

National Olympic Committee (NOC) Dataset:[5]

In order to appropriately clean the team and NOC data in the Olympics dataset, a table of National Committee Codes (NOC) was synthesized from various online sources (including the Olympics website). Given changes in teams, historical and current teams were used to clean and group the Olympics data. For example, Germany has a complicated history in the Olympics. Due to the World Wars Germany was banned from competing in 1920, 1924, and 1948. In the years between 1956 and 1988, there were various teams that are now a part of Germany (GER): East Germany (GDR), West Germany (FRG), and Saarland (SAA). The United Team of Germany (formerly GER), where East and West Germany competed as one team, was retrospectively assigned the IOC code of EUA. In order to organize athletes, medals, and teams under their current designations, all German teams should be grouped under Germany with the NOC code of GER. The list of NOCs includes many historical codes that are no longer used or that reference countries that no longer exist. This list containing Code, National Olympic Committee, and Current Name will be used to update NOC and Team and to join Olympics with countries referenced in the next two datasets.

Temperature Dataset:[6]

The temperature dataset that was used in this analysis was downloaded from Kaggle, with the original data obtained from The Berkeley Earth Surface Temperature Study. This study combined temperature measurements from 16 other archives. Different methods were used to collect the temperature at various locations (i.e. mercury thermometers, electronic thermometers, etc.). The original purpose of this combined dataset was to prove that climate change theories are backed by science through data analysis. This dataset is composed of multiple files but GlobalLandTemperaturesByCountry.csv was the only file used for the Olympics analysis. This set of data contains dt (or date time as a string), AverageTemperature (in Celcius), AverageTemperatureUncertainty, and Country. The Average Temperature is reported for each month. The dates in this file range from 1743 to 2013. Given that the Olympics were revived in 1896, the joined data frame will not contain data from the 1700s and the majority of the 1800s. Formatting and units of measure will require some manipulation to improve the usability and clarity of the data. The cleaned data is then aggregated and used to determine the climate of each country for our analysis.

Country Polygons:[7]

[5] See National Olympic Committees (NOC) - Olympic Movement for more information on the committees and participating countries.
[6] See Climate Change: Earth Surface Temperature Data for more information on the temperature dataset.
[7] See Datasets/Geo-Countries for more information on the geojson/country polygon coordinate dataset.

For high-quality map visualizations, countries.geojson was used to join polygons detailing the borders of each country. The dataset includes country code, which is different from the NOC code, country name(which also has slight variations to the names in the temperature and olympics datasets), and the coordinates for the polygon on the map. The geojson file was downloaded from the Data Visualizations course at the University of Denver: Ritchie School of Engineering and Computer Science (COMP- 4433 taught by Ben Siebrase). The dataset can also be found on github.

Methods and Data Exploration:

Data Cleaning:

Each dataset was processed as its own data frame in order to convert data types, address null values and add necessary columns prior to joining together for analysis. There were 9474 null values in age and 231333 null values in medal. The null values in age were replaced with the mean age value for that sport. Sport was used to group rows before calculating the mean so the replaced values were more accurate to the sport. Gymnastics has an average age of 22.73 years while the average age for Polo is 35.33 years. By using the average age for a given sport rather than the overall average, the replacement values for each null are closer to the true distribution of the athletes competing in that sport. The null values in the medal column indicate that the athlete did not win a medal. Since participation is one of the key features in the analysis, the null values cannot be dropped. Instead, the null values can be replaced with NoMedal. The Games column is a string which contains the year and season of the Olympics the athlete competed in. This can be split into Year and Season columns to provide direct access to each value for the analysis.

In order to join the Olympics dataset to the Temperature dataset, the country names in both datasets must be cleaned and standardized. Another reason for cleaning the team names or adding the country column is that some NOC codes were listed with multiple teams. FRA had the most teams associated with 160 teams. In addition to multiple teams for one NOC code, the Olympics dataset has historical team names and NOC codes that do not align to a country today. The NOC list maps each of the NOC codes (historical and current) to the team name and the current country name, based on geography. For example, Malaya became Malaysia in 1963. Prior to 1963 Malaya competed once, every Olympic Games after that Malaya (and the other island nations that came together) competed as Malaysia. The current countries will be used to join the Olympics data to the temperature and the geojson polygon coordinates. Before merging the two data frames together, the data types were updated. Since there were no null values in this data frame, no replacements were required. A dictionary with Code as the key and National Olympic Committee as the value was generated from the NOC data frame to compare the two data frames. Any codes or teams / National Olympic Committees that did not match were recorded so they could be adjusted or added to the mapping list. For example LIB was used as the code for Lebanon but the official code is LBN.

The temperature dataset required data type conversion, formatting, filtering, unit conversion and null value handling. The date of the measurement had to be converted from a string to a date and formatted correctly. The temperature values were recorded in Celcius as that is the international unit of temperature. Since the United States uses Fahrenheit, converting the temperature values makes them more easily interpreted for those familiar with standard units of measure. The modern revival of the Olympic Games was in 1896 so the temperature data from 1743 to 1900 is not relevant to the Olympic analysis. The null values in the temperature column were mostly from prior to 1900, another reason for the cutoff of 1900 for dropping earlier data. The last date in the data is from 9/1/2013 and most of the data for this date was missing, so it was also dropped. Other missing values were from as far back as 1900. The missing values from dates other than 9/1/2013 were from Antarctica and a handful of other countries. Since Antarctica is not an Olympic country, those values can be removed without affecting the

Olympics analysis. After doing this, 65 null temperature observations remained out of over 570,000 initial observations. The majority of the remaining null values were from the early 1900's and tended to be in the first year or two of a country reporting and for small island nations. For example, the Federated State of Micronesia was missing the first 2 full years of observations (1900-01) and then a string of 7 months in the following year. Since the values of a monthly average temperature would likely experience seasonality, we did not feel that a forward or backward would be appropriate as we may be filling a value for January forward or backward for 2 years and skew our data. Another option of filling the null values using the countries average was discussed, but we determined that this was not a necessary solution, since our purpose was to, in the end, create an average temperature for each country over all months reported. Therefore, we decided to drop the remaining missing values (more detail can be found in the "Missing Data" of the temperature section of the notebook) Average Temperature was reported for each continent as well. Since these were aggregates of the country data and not necessary to conduct the analysis, they were also dropped. There were a handful of European countries which had multiple entries for each month (i.e. Denmark and Denmark (Europe)). After reviewing these values it was determined that the entry without "(Europe)" was likely representing the entire country and its colonies or holdings. Since we were interested in the average temperature for the countries as we traditionally consider them for the Olympics and how they are identified on the map, the entries without "(Europe)" were dropped and the other was renamed to just the country name. The mean temperature for each country was calculated using groupby('Country') and a data frame with Country and Avg Temp Fahrenheit was created for later use.

<u>Exploratory Data Analysis:</u>

The goals of this analysis are to determine the relationship between climate and participation and climate and success at the Olympic Games.

Specifically, do hot climate countries perform better or have more athletes in the summer Games? *And conversely, do cold climate countries perform better or have more athletes in the winter Games?*

In order to analyze the seasons separately, a filtered data frame was created for each season. To begin to understand the distribution of both the summer and winter Olympics data frames, some initial plots were generated.

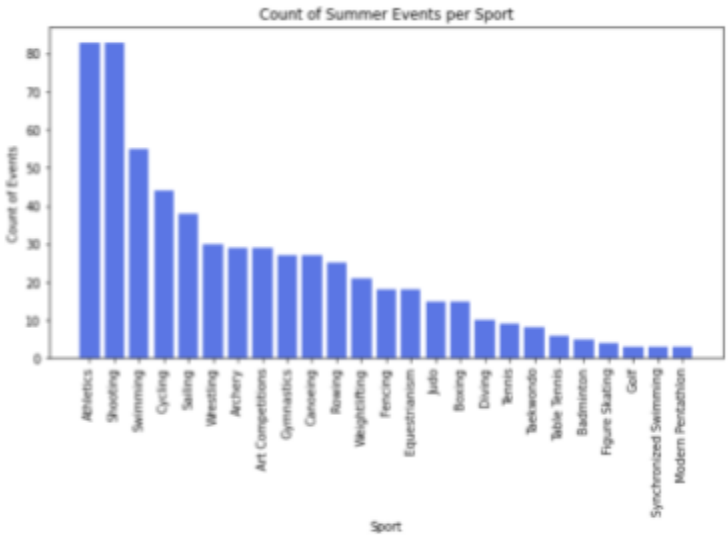

Figure 1: Count of Summer Events. Summer dataframe grouped by Sport, unique count of Event.

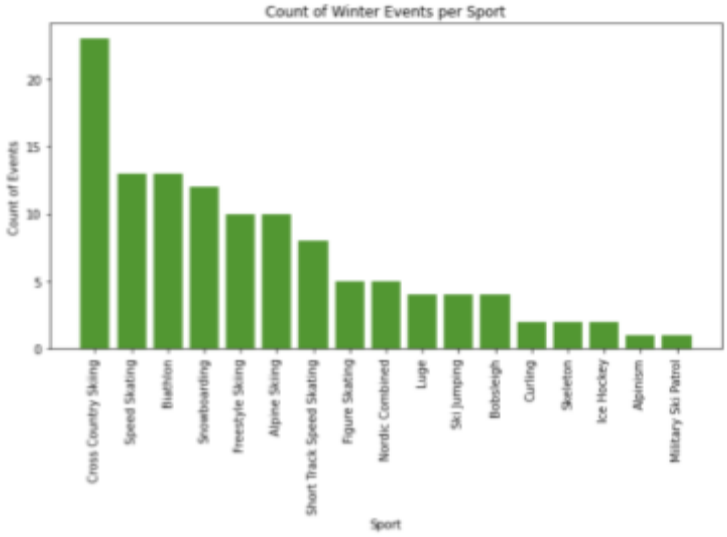

Figure 2: Count of Winter Events. Winter dataframe grouped by Sport, unique count of Event.

Figures 1 and 2 are the distinct count of events for each sport in summer (Figure 1) and winter (Figure 2). Each plot shows the top 25 sports, but since winter had only 17 sports all winter sports are shown. The summer Olympics are older than the winter Olympics and have significantly more events. Note that in addition to far more events at the summer games, there are more team sports (Basketball, Soccer, Handball, Water Polo) at the summer games while Hockey is the only large team sport with most other sports being individual. The ratio between number of events and number of team sports between the two seasons will likely skew the participation results.

In order to get an initial look at participation, a bar plot was created for both seasons using a distinct count of name (Figure 3 - summer and Figure 4 - winter). The top 25 countries are shown in each plot. The first female athletes to compete in the Olympics competed in 1900. The number of female athletes has increased substantially over the last 120 years. In 1908, 44 women competed in the summer olympics. By 2016, that count increased to 5026. In 1924, 13 women competed in the winter olympics. That number grew to 1102 women in 2014. Figures 5 and 6 highlight female participation in the summer (Figure 5) and winter (Figure 6) games. Again, showing the top 25 countries by count. The United States has the most athletes in both the summer and the winter, and also the most female athletes in both seasons. The summer has 2097 United States women and the winter has 522. Participation in the winter is about 20% of what it was in the summer for the overall athlete count. Female participation in the winter is approximately 25% of the summer totals.

Distinct Count of Athlete by NOC

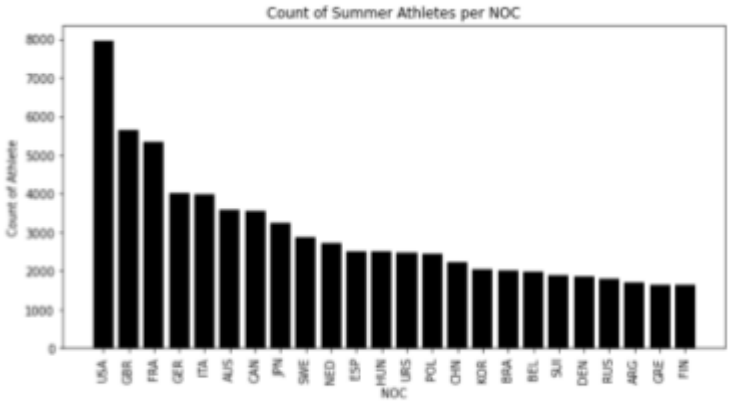

Figure 3: Count of Summer Athletes. Summer dataframe grouped by NOC, unique count of Name.

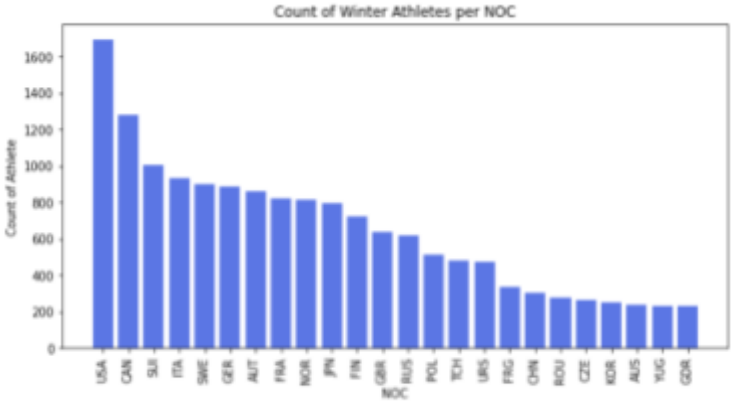

Figure 4: Count of Winter Athletes. Winter dataframe grouped by NOC, unique count of Name.

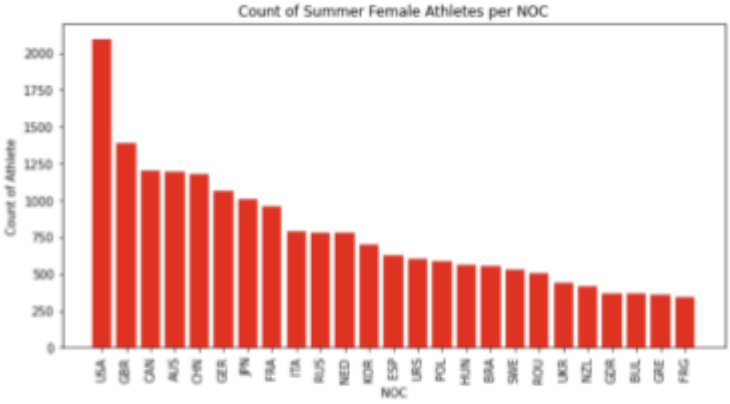

Figure 5: Distinct count of Female Summer Athletes. Summer dataframe grouped by NOC, unique count of Name.

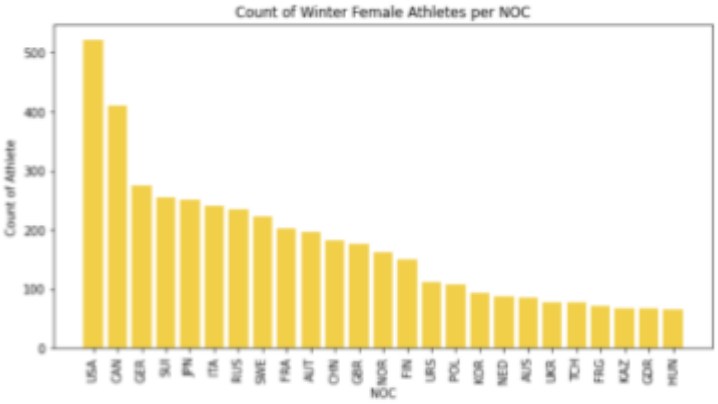

Figure 6: Distinct count of Female Winter Athletes. Winter dataframe grouped by NOC, unique count of Name.

The second key feature to be analyzed is the success, which translates to medal count. Using the categorical 'Medal' column, a stacked bar chart showing the "successes" was plotted. Figures 7 and 8 are summer and winter stacked bar charts showing the count of gold, silver, and bronze medals that a country won. The top performing country based on number of gold medals and number of total medals for the Summer Olympics (Figure 7) is the United States. The Winter Olympics (Figure 8) has two top performers, Finland for count of gold medals and the United States for total medal count.

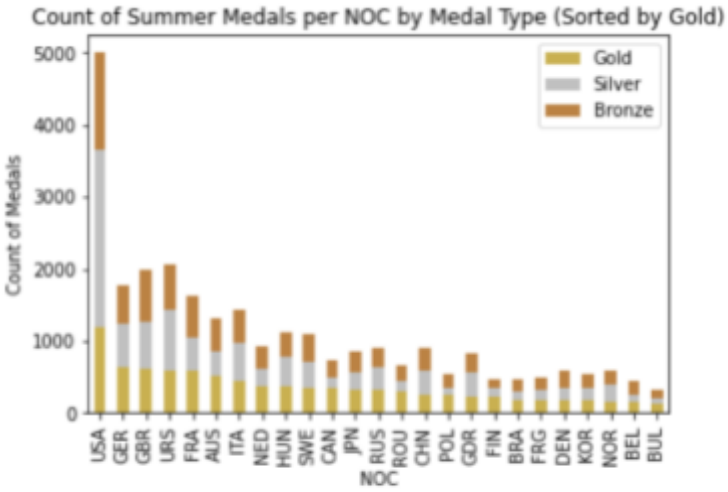

Figure 7: Total count of Summer medals per NOC, colored by type

Summer dataframe grouped by NOC.

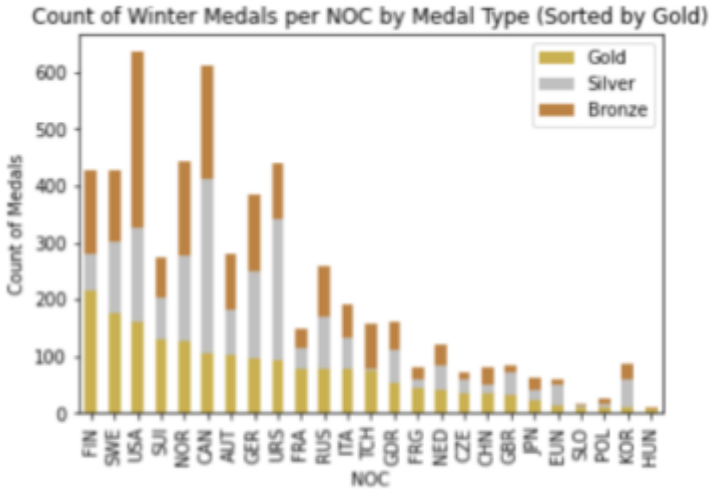

Figure 8: Total count of Winter medals per NOC, colored by type

Winter dataframe grouped by NOC.

In the interest of grouping countries into climate categories (hot, temperate, and cold), an initial estimate of the clusters was determined from the histogram of average temperature (Figure 9) and then the clusters were plotted on a scatter plot (Figure 10) and a swarm plot (Figure 11). In Figure 9, we see a small bimodal curve. The data are left skewed with the majority of observations occurring above 60 degrees. The median average temperature is 73.8 degrees and the mean average temperature is 66.6 degrees, being pulled downward by the outlying temperature of Greenland and Denmark.

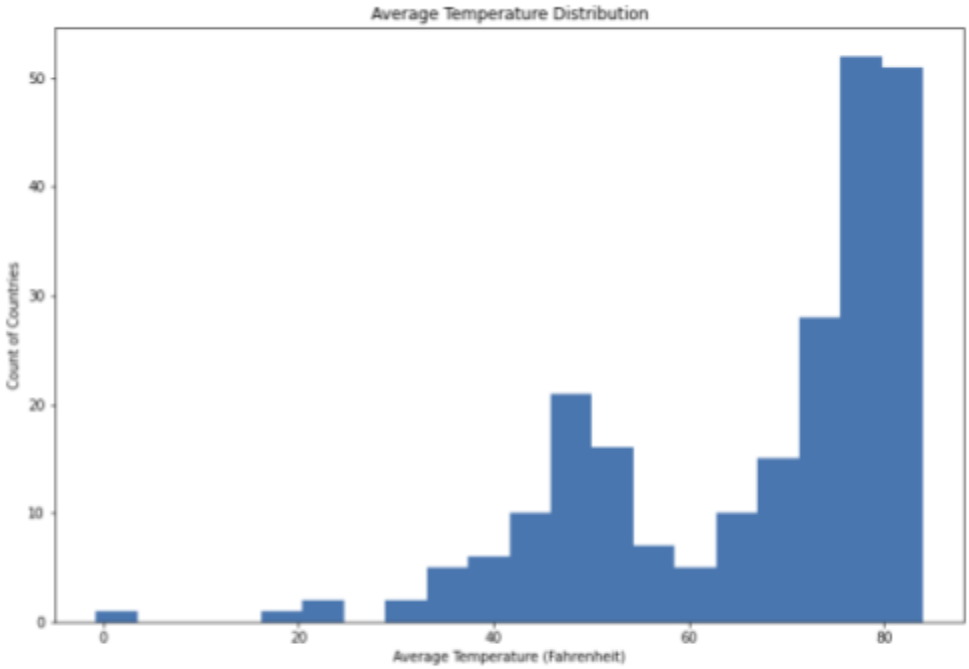

Figure 9: Histogram of Average Temperature highlighting the estimated three clusters.

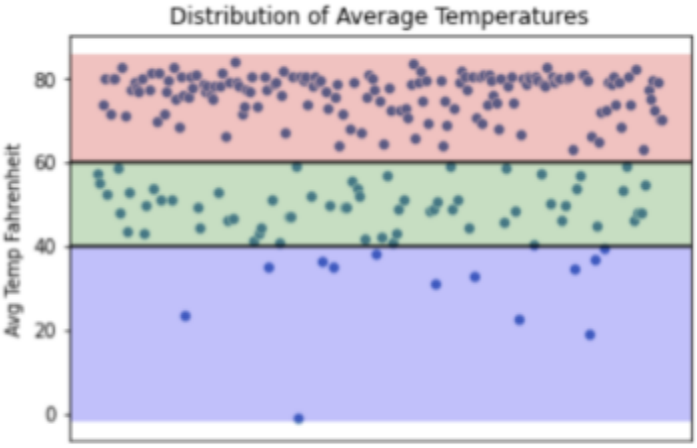

Figure 10: Distribution of Average Temperatures Scatter Plot, using estimated clusters

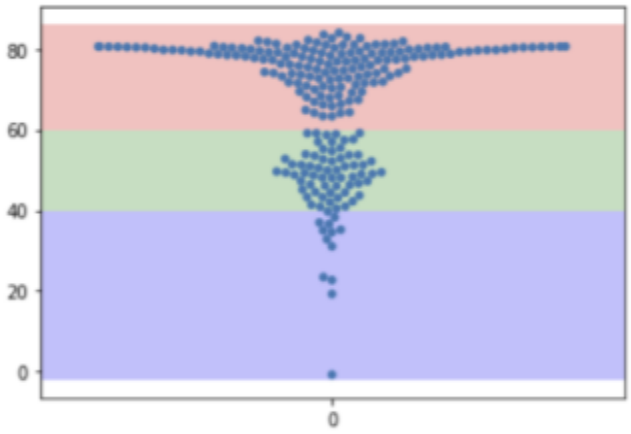

Figure 11: Distribution of Average Temperatures Swarm Plot, using estimated clusters.

To formally assign points (countries) to one of the climates, k-means, a form of unsupervised learning, was used to generate 3 distinct temperature clusters (code shown in Figure 12). The data had to be reshaped in order to use the sklearn.cluster function KMeans. The number of clusters (3), the random_state (0), and the number of iterations (10) were all defined. The random_state parameter is used to ensure reproducibility of results. Without the random state, each time the code is run may return a different result. The results of this unsupervised clustering were added as a new column of the data frame, replacing the numerical assignments with the more clear labels of hot, temperate, and cold. The centers of these clusters were 28.77, 49.60, and 76.50, which were very close to the estimates from the histogram.

```
#Convert the series to an array and reshape
temperatures=country_temps.array.reshape(-1, 1)

#Create the KMeans model with 3 clusters
kmeans=KMeans(n_clusters=3, random_state=0, n_init=10)
kmeans.fit(temperatures)

#Get the cluster labels for each data point
clusters=kmeans.labels_

#Review the output of the model by printing the country, its temperature, and the corresponding cluster
#for country, temp, cluster in zip(country_temps.index, temperatures, clusters):
#    print(f"Country: {country}, Temperature: {temp[0]}, Cluster: {cluster}")

#Create a new data frame with the country temps and add the associated cluster as a column
climate2=pd.DataFrame(country_temps)
climate2['climate']=clusters

#Map the cluster to a climate description
climate2['climate'].replace({0:'hot', 1:'temperate', 2:'cold'}, inplace=True)
climate2
```

Figure 12: SciKit Learn K-means clustering code for climate assignments.

The final cluster assignments (Figure 13) were plotted as a scatter plot, colored by the climate group. In the estimated clusters, there were 13 cold countries, 63 temperate countries, and 156 hot countries. In the k-means clusters, there were 12 cold, 64 temperate, and 156 hot countries. One of the estimated cold countries was assigned to the temperate k-means cluster. The rest of the countries were accurately assigned in the earlier estimate.

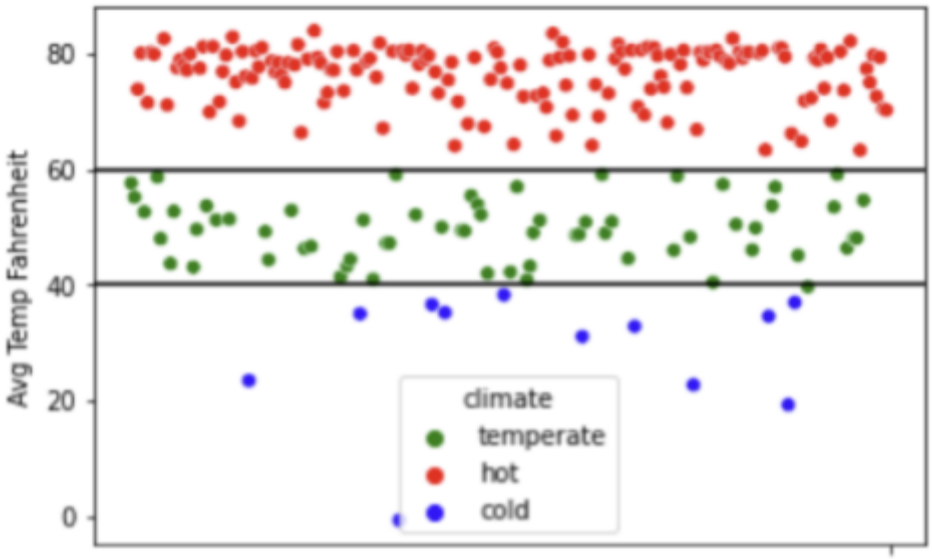

Figure 13: Color coded scatter plot of k-means clusters. Cluster assignments replaced climate category names.

Using the last data set, countries and their polygon coordinates for the map, countries were colored by the average temperature (Figure 14) and by the three k-means climate clusters (Figure 15). Some of the countries were not consistently named in the temperature data and the geojson data, so corrections were made to fill in all the countries in the temperature dataset. Many of the countries that had to be renamed were islands. For example, The Bahamas versus Bahamas. The 'United States' had to be renamed to the 'United States of America'. Countries (or regions) without temperature data, like Antarctica, are shaded with a cross pattern.

Two missing countries are readily apparent, which are South Sudan and Somaliland. South Sudan is not provided in the temperature dataset. South Sudan gained independence in 2011 and our temperature data goes through 2013. This short overlap and the establishment of a new nation may be reasons for South Sudan missing from the data. Somalia is listed in both datasets but there is no temperature data for Somaliland. Somaliland is an autonomous region in Northern Somalia and while it is self-governing no foreign countries recognize it. This may explain why it is missing from the dataset.

Most hot climate countries fall between the Tropic lines. All of Africa and Central America, most of South America and the Middle East, as well as most island nations, and the region of Australasia comprise the hot climate countries. Temperate climate countries generally fall between the Tropic lines and Arctic circle, but are closer to the Tropic lines. With a few exceptions, almost all temperate countries are in the Northern hemisphere and specifically comprise the United States and most of Europe. All cold climate countries are in the Northern Hemisphere and while they extend below the Arctic Circle, their position nearer the pole is expected. Canada, Russia, and the Nordic countries of Norway, Sweden, and Finland are the prominent cold countries.

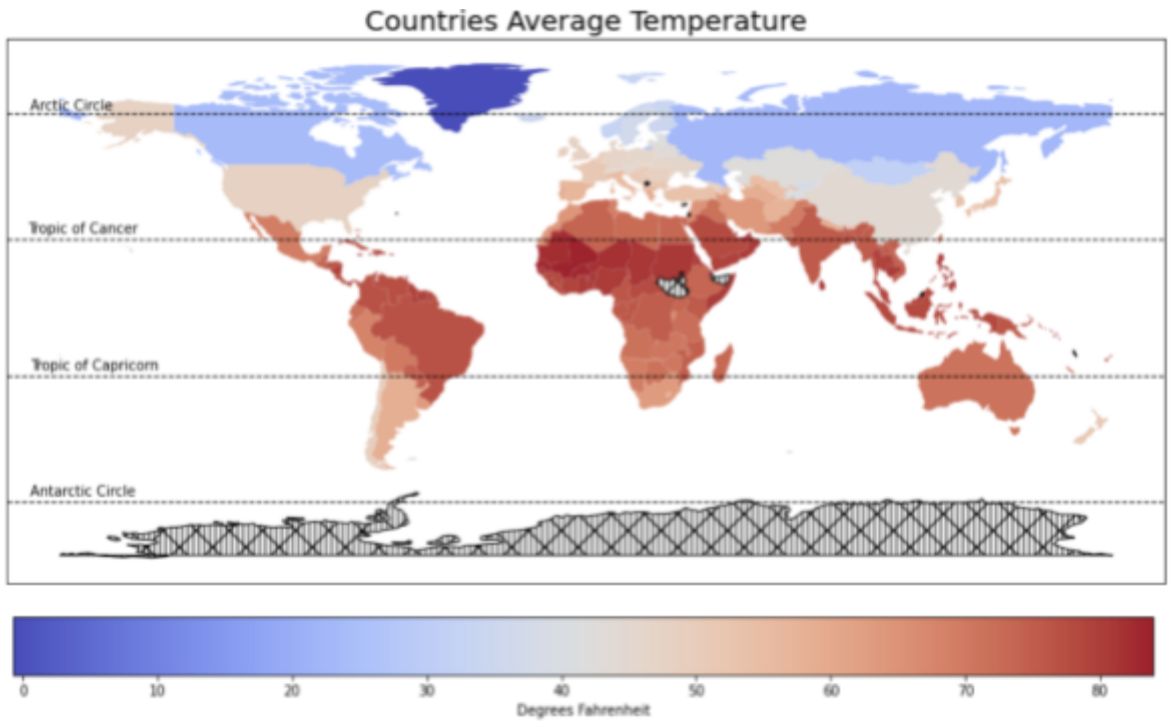

Figure 14: Map of Average Temperature Fahrenheit shaded by degrees Fahrenheit.

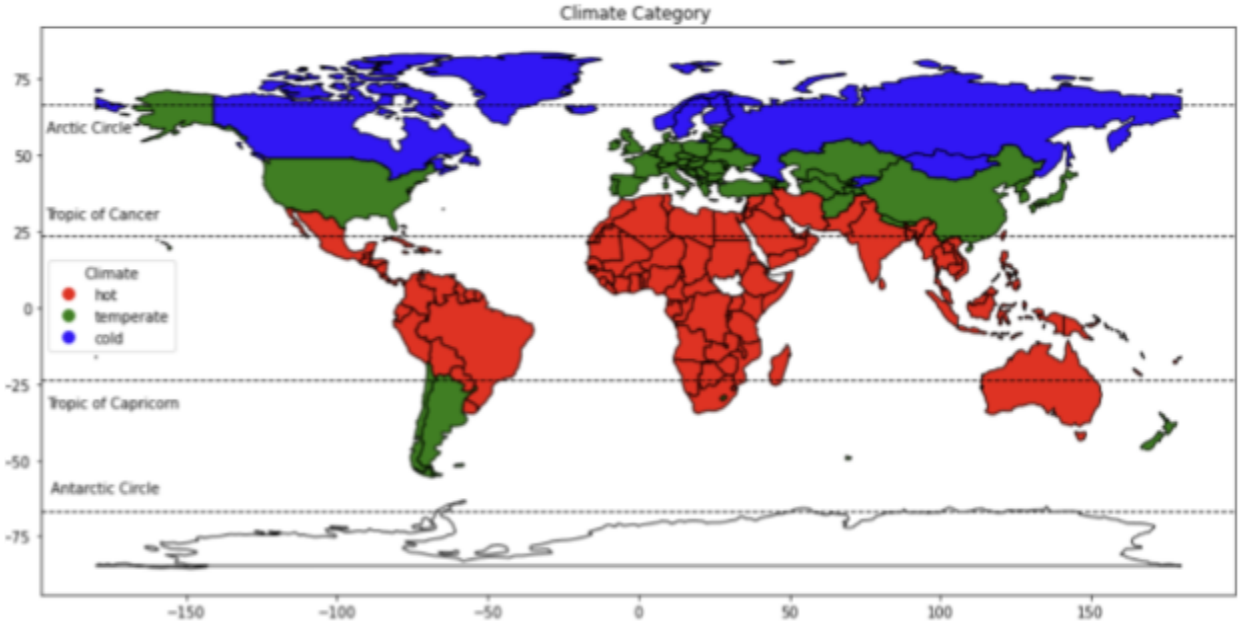

Figure 15: Map of Climate Assignment based on k-means clustering of average temperature.

Project Analysis:

The goal of this project was to analyze and define the relationship between a country's climate and participation in summer and winter Olympics as well as the relationship between a country's climate and success in summer and winter Olympics. Using the knowledge from data exploration, the Olympics-NOC merged data frame was merged with the Temperature-Polygon merged data frame. When merging the data frames, using the country column from both data frames provided the fewest nulls. Fewer than 800 of the approximate 270,000 athlete rows did not have matching data in the climate-polygon data. These were from 16 countries that did not have a match in the climate data frame and were predominantly small island nations. These countries had very few participants. The Olympics Art Competition was included in the Olympics from 1912 to 1948. At its peak, 1932, 1100 individuals competed. The vast majority of the Olympics data are sports. In keeping with the spirit of the research question, focused on climate impact on sporting events, for the rest of the analysis the Art Competition "sport" was dropped.

To analyze the participation, a new dataframe was created with a count of countries based on year, season, and climate. For example, in 1896 there was 1 cold country, 1 hot country, and 10 temperate countries competing in the summer games. In 2016, the most recent year in the dataset, 8 cold countries, 127 hot countries and 57 temperate countries competed in the summer games. Based on the total number of countries in each climate category, the percentage of countries that competed was calculated. In 2016, 66.7% of cold countries, 89.1% of temperate countries, and 81.4% of hot countries competed in the summer games (Figure 16 - summer and Figure 17 - winter).

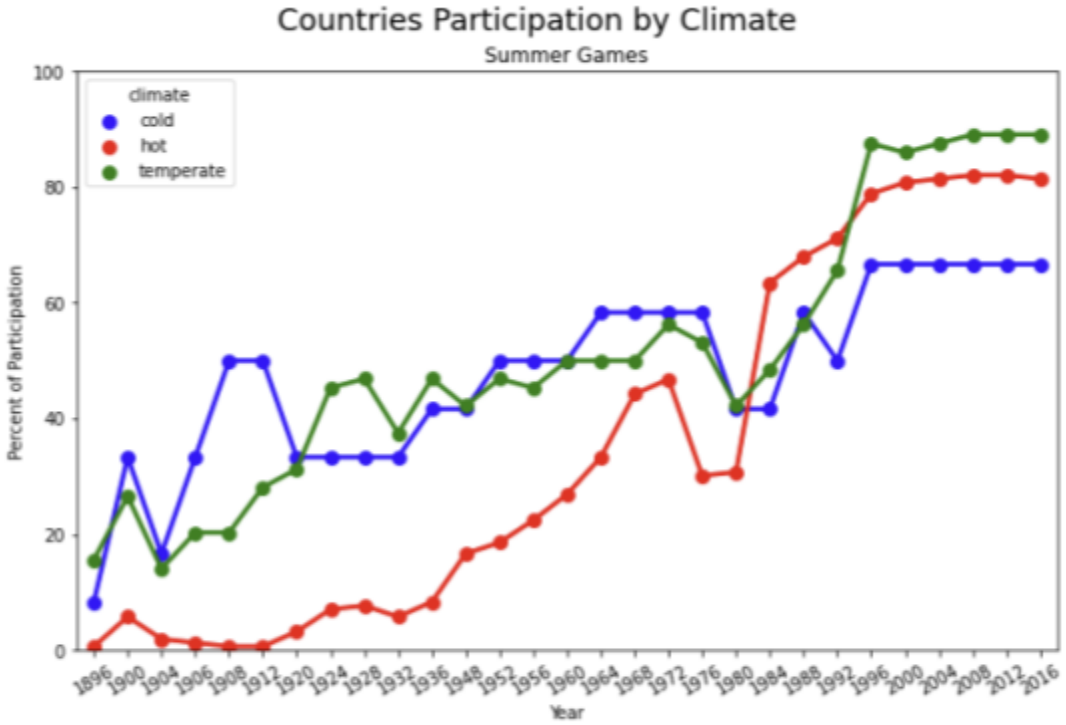

Figure 16: Percentage of countries in each climate category that participated in the summer games over time. Percentage determined using the total count in the entire dataset for that climate group.

Figure 17: Percentage of countries in each climate category that participated in the winter games over time. Percentage determined using the total count in the entire dataset for that climate group.

Separating the summer and winter games, the increase in countries competing in the games that are the opposite of the country's climate is apparent (cold countries competing in the summer games). Participation in the summer games as far back as 1908 was approximately 50% for both temperate and cold countries. The percentage of participation increased slightly for both to around 60% in the 1960's. Through the 1980's temperate countries saw a second period of growth and have leveled out to about 90% since the early 1990's. Interestingly, participation by hot climate countries started out quite low, 10% or less, until after WWII when the participation by hot countries began to grow steadily. Starting in 1948, participation by hot countries doubled, to 20%, and a period of steady increase in participation continued until 1996 when it reached 80%. Participation in the Summer Olympics has remained steady for each climate since the 1996 games. Since Figure 17 is showing the percentage of the total number of countries for that climate, higher percentages for cold and temperate countries are expected. For example, with 57 temperate countries participating in the 2016 summer olympic games, the temperate climate is at 89.1% participation, while hot countries, with 127 participating, is at 81.4%. The temperate climate has less than half the number of participating hot countries but a higher percentage. Due to the disparity in the number of countries in each climate, the percentage of countries participating was used to level the playing field.

In the Winter Olympics, which began in 1924, the participation by cold climate countries was highest until 1992. The percent of cold climate countries participating in the games largely followed a similar pattern to that of cold countries in the Summer Games. Temperate climate countries participated steadily around 40% until the 1980 games, where their participation began to increase until 2000, reaching about 80%. Starting in 1994, temperate climate countries have been the highest participating climate. Unsurprisingly, hot climate countries participate in the Winter Games at a significantly lower percentage than cold or temperate countries, below

10% until 1980. Although not as dramatic an increase as temperate countries, hot countries began to increase participation in 1980 and have slowly increased to just under 20% in the most recent games (2014).

Currently, about 60% of cold climate countries participate in the Summer and Winter Games. About 90% of temperate countries participate in the Summer Games and about 80% in the Winter Games. However, while about 80% of hot climate countries participate in the Summer Games, less than 20% participate in the Winter Games. Factors other than climate could impact the difference in participation by hot climate countries in the Summer and Winter Games. Factors ranging from qualified athletes and general interest and/or knowledge of the events to access to specialized equipment and facilities.

This summary, however, does not take the number of sports, events, or athletes into account. A single row for the games counts as participation in the percentage calculation, so a country with one athlete is weighted the same as a country with 1000 athletes. One way to get a better idea of participation is to review the participation specific to the sports. As long as a country has at least one athlete competing in at least one event of a given sport, that sport will be included in the percentage calculation for the country. This map will show us how specialized a country is. Are all of a country's athletes competing the same sport(s)?

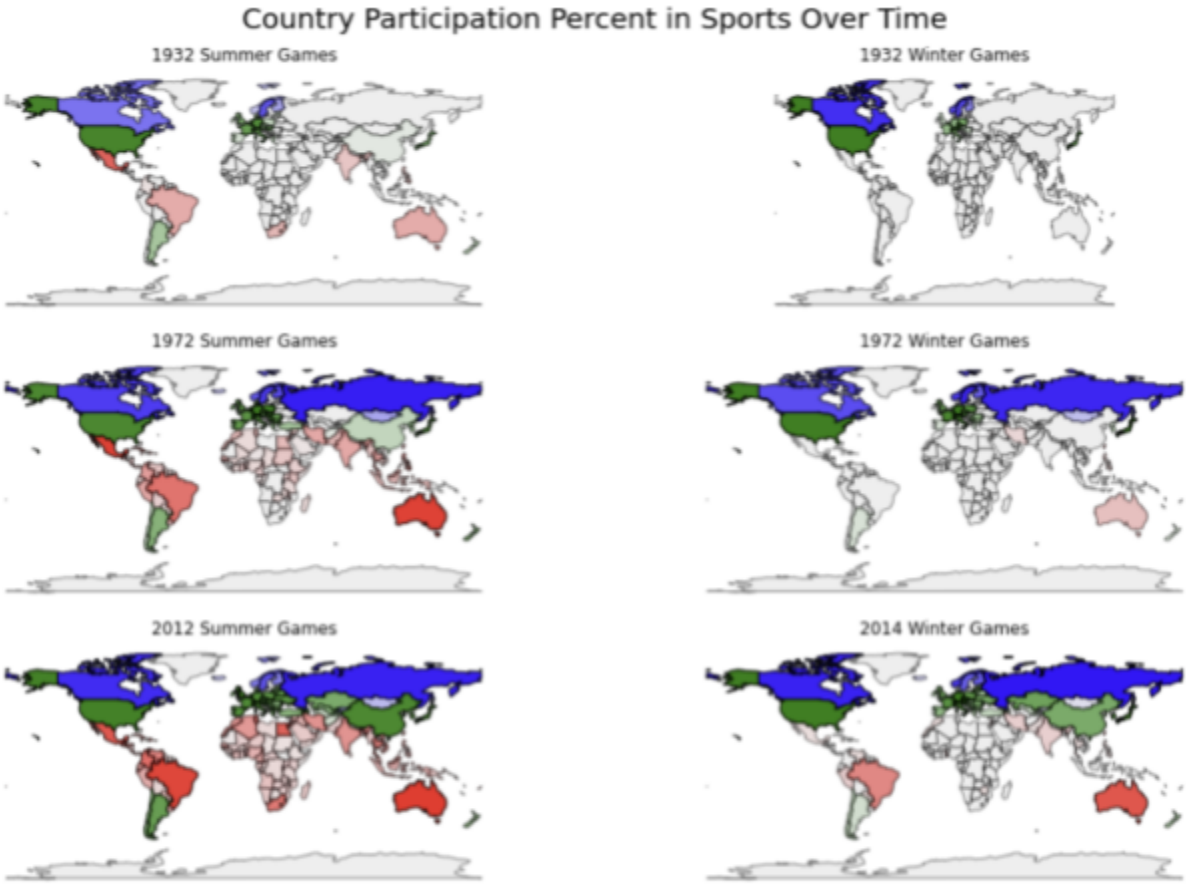

Figure 18: Shaded Map of Percentage of Events Participation. Color indicates climate and intensity indicates level of participation.

In order to clearly show the change over time, three years of the summer and winter games were chosen: one early (1932), one late (2012/2014) and one in between (1972). The maps are laid out to aid comparison from summer to winter and from year to year. From these maps, both the summer and winter games have seen an increase in the percentage of sports countries participate in. In 1932, few countries participated in the summer olympics and even fewer in the winter olympics. The countries that participated in the 1932 winter olympics participated in a high percentage of sports. Between the 1972 and 2012 summer games, the percentage of sports Brasil participated in increased, as did the percentage for many countries in Africa and Asia, majorly hot climate countries. A notable increase in percent of sports occurred between the 1972 and 2014 winter games for Australia and to a lesser extent, Brazil.

Another way to get more information about a country's participation is using a total count of the number of events for a specific year and season to calculate the percentage of events that a country had an athlete participate in. For example, the sport athletics (track and field) has 83 events in the dataset. The general trend for the maps of events is that summer has more countries participating compared to winter and as time passes, the percentage for participating countries increases.

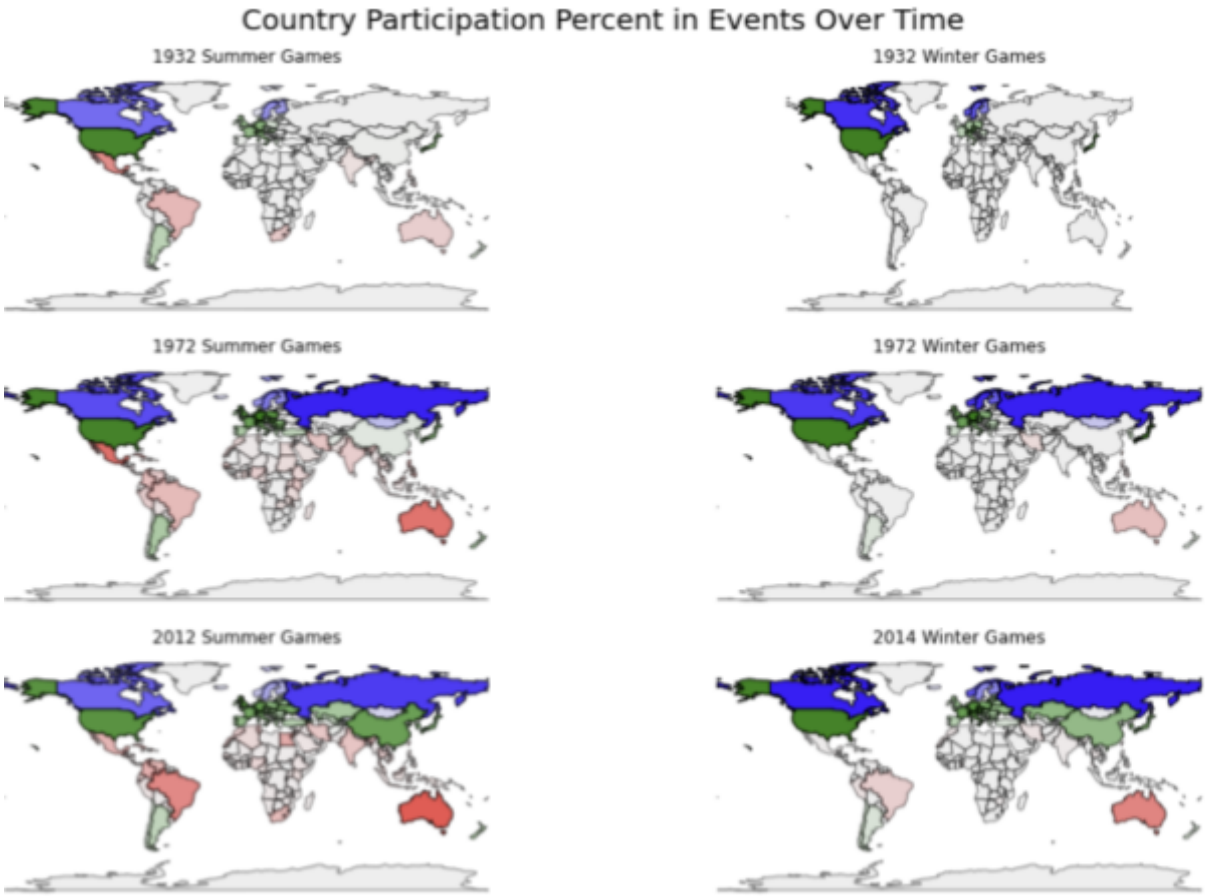

Figure 19: Shaded Map of Percentage of Events Participation. Color indicates climate and intensity indicates level of participation.

These maps show more countries participating in summer events than in winter events. In 1932, the percentage of events that Canada participated in was higher in the winter than in the summer. Over the years, more sports and events have been added which affects the percentage.

For the Summer Games, participation in sports was very low by hot climate countries early on and even into the 1970's. However, participation has increased markedly by 2012. In comparison to the previous maps for sport participation, here we see the very light shading red of Africa for the Summer Games of 1972 and 2012. This indicates that while the number of hot climate countries in Africa participating in the games has been increasing and they are participating in more sports than ever, they are still only participating in a limited number of events. Even in temperate climate countries, outside of the United States of America and Europe there was little participation in the early years. Now, most temperate climate countries have a strong participation in sports. The cold climate countries participation is driven strongly by Canada and Russia, while there is mixed participation in the Nordic countries. Today, the Olympic Summer Games have truly become the global games.

For the Winter Games, there was virtually no participation by hot countries outside of Australia through the 1970's. Early on, participation in the Winter Games was limited to the U.S., Canada, and Europe, with strong showing by the Nordic countries. Until at least 1972, essentially all participation was by the United States, Canada, Russia, and European countries. Australia was the only hot country of note but still had low participation. While participation has expanded as of 2014 in the Winter Games to include more countries from Asia, there is essentially no participation or very little by African countries, island nations of the Indopacific and most of South America.

Participation is not the only thing that matters when it comes to the Olympic Games; a country must also perform well to win. Top performers are rewarded with gold, silver, and bronze medals. Before analyzing the medals won, team sports need to be addressed. For team sports, the data frame contains a row for each athlete on the team. This causes an inflated medal count for countries that medaled in team sports. For example, the United States added 17 medals for the women's soccer (football) team medaling in 2008. To remedy this, medal counts can be grouped by team and event and then the duplicates can be dropped. This requires that id, name, sex, age, and city are dropped from the data frame, thus creating the "duplicates" to be removed. Figure 20 shows the average adjusted medal counts over time. The average was used to account for more countries participating, which increases a climate's chance of winning medals.

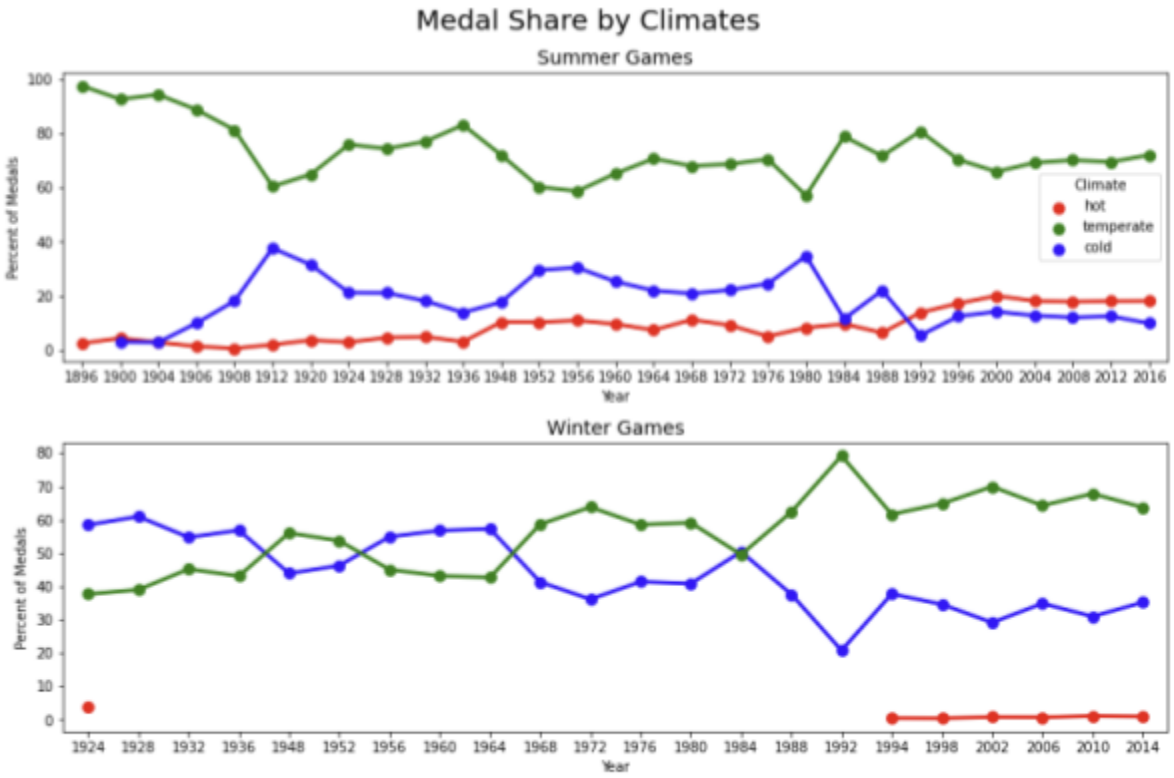

Figure 20: Percentage of Medals by Climate.

At the Summer Games, temperate climate countries have consistently won the majority of medals, between 60% and 80% at each Games. Similarly to what we observed in the total counts, cold climate countries won the second highest percentage, of between 20% and 40% until 1984. At which point, they dropped to closer to 10%-15% and hot climate countries have won about 20% since then.

At the Winter Games, temperate and cold climate countries each won between 40%-60% of medals, trading the highest share of medals back and forth through 1984. Starting in 1988, temperate climate countries began to dominate more of the medal share, peaking to almost 80% in 1992, before plateauing to 70% since. After 1988, cold climate countries saw their share of medals decrease to about 30%. In 1994, hot climate countries began to win a marginal percentage of the medals each year.

While this approach addresses the increasing number of medals awarded and allows for easier comparison of which climates had the most success, it does not address the issue of the small number of cold climate countries or larger number of hot countries in our model. We'll try another approach next, looking at the average number of medals won by countries in each climate.

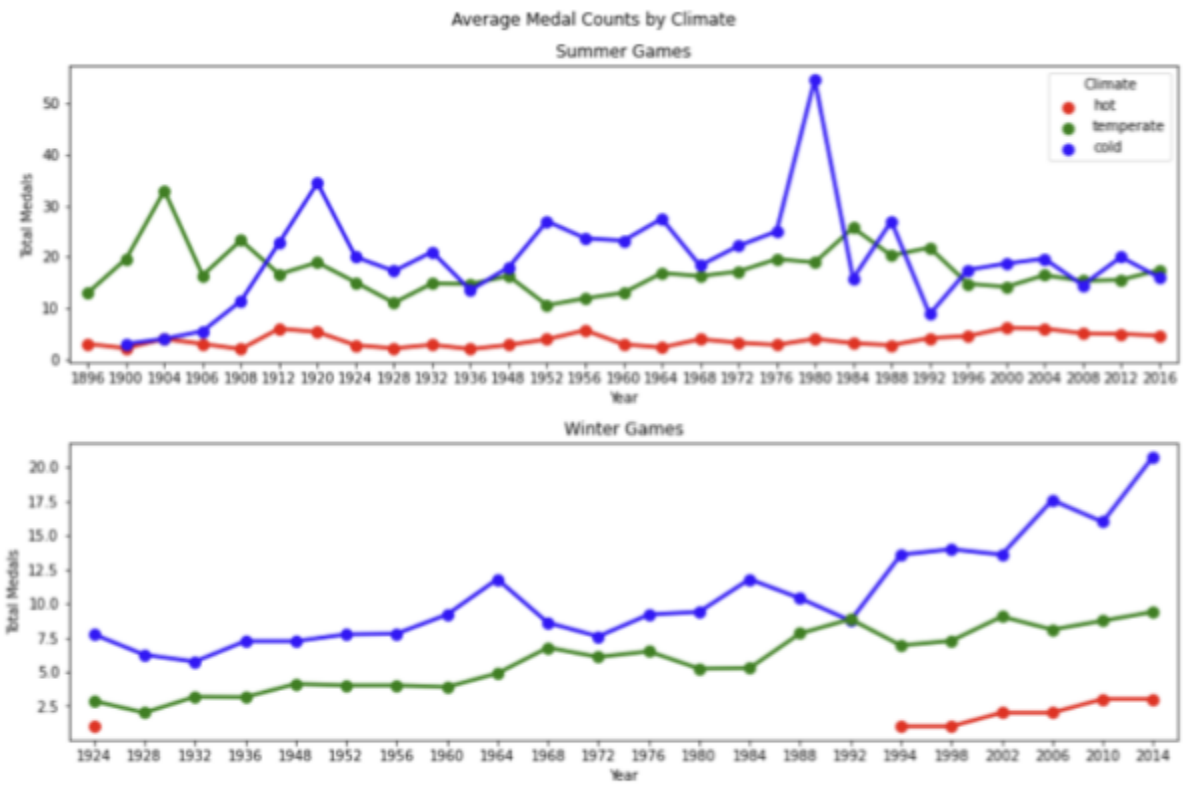

Figure 21: Average medal counts by climate.

Averaging the medal counts was necessary as the number of participating countries for the cold climate over the entire time period is significantly less than that of temperate and hot climates. With fewer countries, cold climate countries have performed very well at the Summer Games. In fact, on average cold climate countries outperformed temperate countries at all but 5 summer games since 1912. The only drastic domination by temperate countries was in 1992. With this normalization, it is evident that hot climate countries still have less success even at the Summer Games and the fact that they have more total medals in recent years is due to the increase in participation of the 153 hot climate countries. Less surprisingly, cold climate countries tend to dominate in the Winter Games, on average. While it was difficult to parse out the success of hot climate countries in recent years due to the overall low medal counts compared to other climates. On average they have been seeing a slow, but steady increase in success.

Again, this may not tell the entire story, as some of the most prolific countries may be compensating for other countries and skewing the average medal success. Some of those countries, identified earlier, are the United States, Russia, China, and Germany. Average medal counts by climate could be skewed by outlying countries with prolific medal counts. Let's look at the last 5 games and see the distribution of medal counts by countries in each climate. To identify outliers, a boxplot of medals, grouped by climate, from for each games since 2000 was plotted and is shown in Figure 21. In each year and each climate grouping, there are visible outliers with more medals than the majority of competing countries.

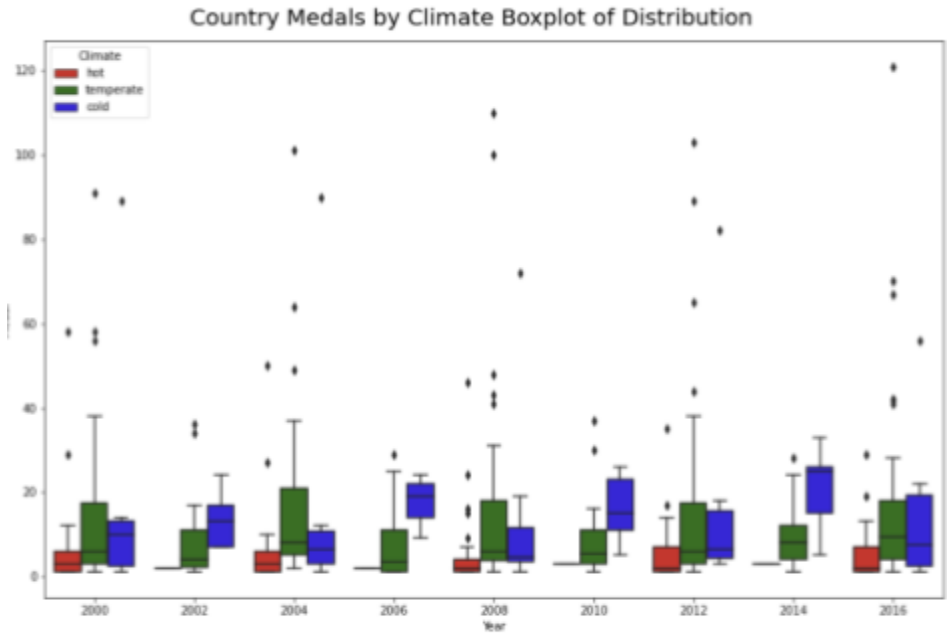

Figure 22: Boxplot of medal count by climate from 2000 to 2016.

Removing the outliers shown in Figure 21, the performance of the majority of each climate can be analyzed. Figure 22 shows the average medal counts for each climate from 2000 to 2016. This figure highlights the temperate climate as the best performing climate for all olympic games from 2000 forward. Countries that were significantly out performing the rest in their climate, were removed, which means that countries that were adding many medals to the climate's overall medal count are no longer considered.

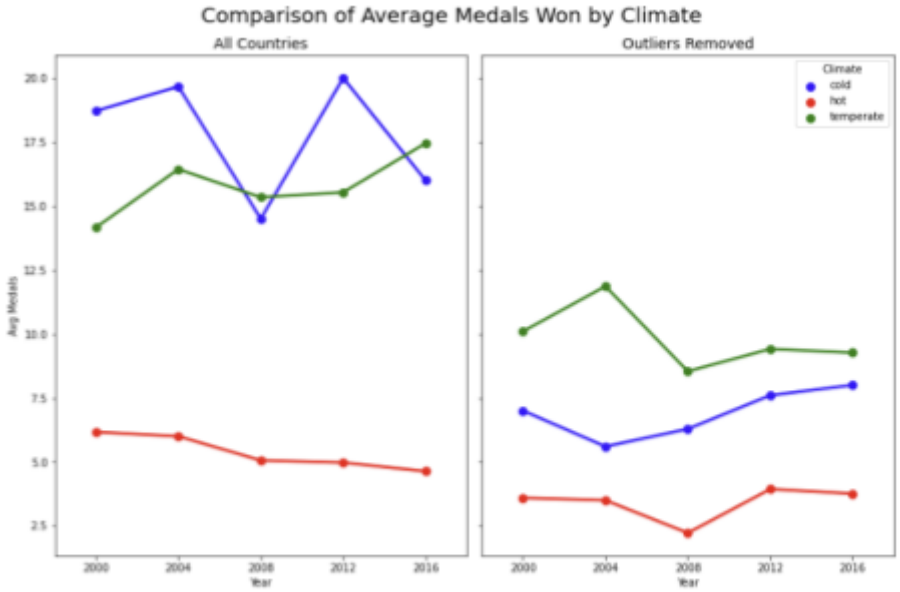

Figure 23: Line plot of average medal count per climate from 2000 to 2016.

In contrast, to finish the analysis of medal success as it's related to climate, the top 5 countries for each climate will be the only countries used. Since the last visual removed the top performing countries, this visual will be only the top performing countries in each climate at each of the games.

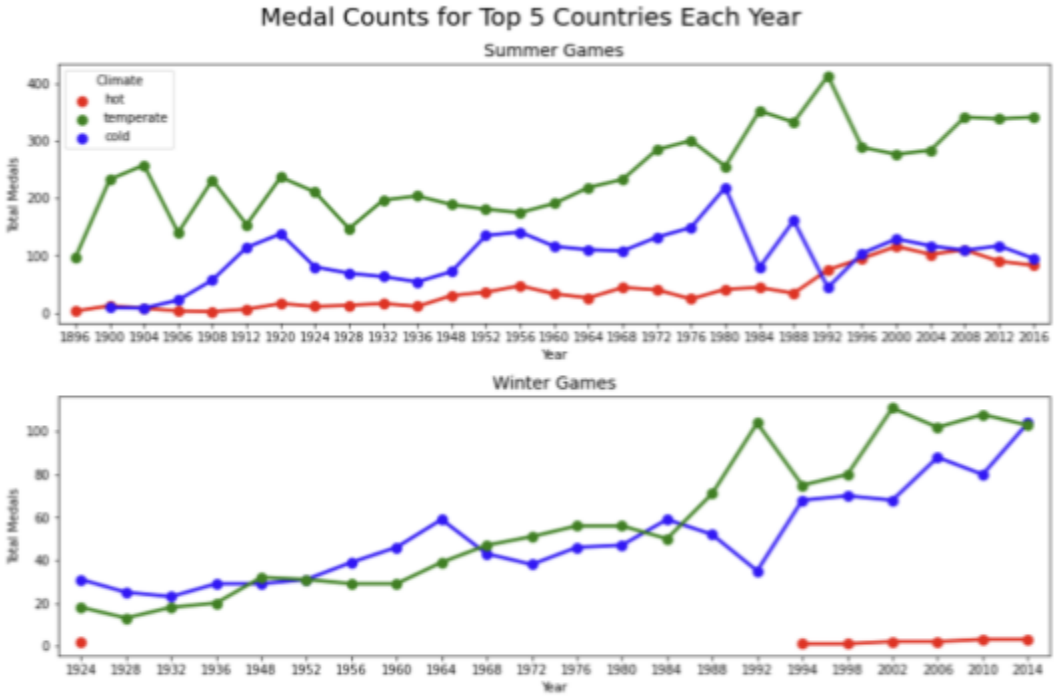

Figure 24: Medal Counts for the Top 5 Performing Countries for Each Climate.

In the Summer Games, we now see that the top temperate climate countries are consistently the most successful and by a good margin. Through most of the 1900's the top cold climate countries were more successful than hot climate countries. The top cold climate countries managed to produce nearly the amount of medals of temperate countries in 1980, although this could be explained in part by the USA's boycott of the Summer Games in Moscow. In 1992, the top hot climate countries began to improve and have since drawn consistently even with cold climate countries over the past 6 Summer Games.

In the Winter Games, the top cold climate countries performed slightly better than the top temperate countries until 1964. Then, beginning in 1968 the top temperate climate countries began to take the lead. Over the history of the Winter Olympics, both the top temperate and cold climate countries have seen steady increase in their total medals. Overall, the top cold and temperate climate countries have performed fairly closely to each other. The one exception to this is the 1992 Games, where the top temperate countries drastically outperformed cold climate. Although hot climate countries have begun to win medals in the past 6 Games, the success of the top hot climate countries is still far below that of the top cold climate countries. Both hot and cold countries remain well below the performance of the top temperate countries.

Some explanation for the new found success of hot climate countries can be found in 1992 at the Summer Games. Hot climate countries doubled their success from 1988 in Track and Field, Boxing, and Swimming events. The anomaly in the 1992 Winter Games where temperate

countries won 14 more medals in Cross Country Skiing, 9 more medals in Biathlon, and 7 more in Figure Skating than they won at 1988 Winter Games is matched by the sharp performance decrease of cold climate countries in the same year. Cold countries won 11 fewer medals in Cross Country Skiing and 7 fewer in Figure Skating.

At the Winter Games, Australia has won 13 (6 gold, 3 silver, 4 bronze) of the medals for hot climate countries and India has won 1 gold medal. India's medal was in Mixed Alpinism at the 1924 Games. At the same games, Australia won a gold in the same event. More recently, Australia has won medals in Alpine Skiing, Freestyle Skiing, Snowboarding, and Short Track Speed Skating, which accounts for all medal wins for hot climate countries.

Through 2016, the Summer Games have been held 29 times. At the Summer Games, the temperate countries: the United States (28), Germany (25), and France (18) have ranked in the top 5 of medal count the most often. Placing in the top 5 28 times, means the U.S. has been in the top 5 of temperate medal winners at every summer games except one, the 1980 Moscow Games, which was boycotted by the United States. Australia (27) and Brazil (15) have the most success for hot countries. Canada (27), Sweden (27), Finland (24), Norway (24), and Russia (17) have had the most consistent success out of the cold countries. Russia did not compete in several games in the first half of the 1900's (1904, 1920, 1928, 1932, 1936, 1948, and 1952 Summer Games). Despite not competing in 7 editions of the Summer Games, they have enjoyed marked success in the games they have attended.

Through 2016, the Winter Games had been held 22 times. The cold climate countries which appeared in the top 5 of cold climate medal count the most times were Canada, Finland, Norway, and Sweden, all appearing in the top 5 each year. Temperate climate countries with the most consistent success are Austria and the USA both with 21 and Germany with 19 times. For hot climate countries, Australia and India are the only countries which have won medals.

Conclusion:

Does a country's climate affect its likelihood to participate in the Olympics?

Our assumption was that hot climate countries would be more likely to participate in Summer Games than Winter Games, and that cold climate countries would be more active in the Winter Games than Summer.

This does seem to hold for the hot climate countries, where they have dramatically higher participation rates in the Summer Games than the Winter Games. However, cold climate countries have historically had strong participation in the Summer Games, rivaling that of temperate climate countries, and continue to have high participation rates to present day. Temperate climate countries have consistently been high participants in both games. So, our assumption does seem to hold for hot climate countries but not necessarily for cold climate countries.

Over the history of the modern Olympic Games, the participation of all countries has increased. As the games gained popularity and recognition, countries of all climates began to participate more and we see increased participation in games of the opposite climate. This is particularly true of the Winter Games, where hot climate countries had virtually no participation until 1980. While temperate climate had around 40% participation in the Winter Games until the 1980's, they have had strong participation in both games. A bit surprising is the participation rate of cold climate countries in the Summer Games. Their participation in the Summer Games has been historically near that of temperate climate countries, but more recently has plateaued while temperate countries continued to increase. The Summer Games are more popular and this is driven mostly by the higher rate of participation by the large number of hot climate countries.

Is a country's climate an indicator of success in the Olympic Games?

The assumption was similar to the previous research question. Cold climate countries would tend to be more successful, in terms of medals won, at the Winter Games and less successful at the Summer Games. While expecting the opposite to be true for hot climate countries.

Indeed, this appears to be true for hot climate countries, as they have more medal success at the Summer Games and until the 1994 Winter Games had no tradition of medal success. In the past 6 Winter Games hot climate countries have won a limited number of medals. However, hot climate countries lag far behind other climates in terms of average number of medals won at the Summer Games. It is not surprising to find that the cold climate countries have the most success at the Winter Games. However, it was surprising to find the amount of success that cold climate countries have had at the Summer Games. Historically cold climate countries actually won more medals on average than the temperate climate countries and even recently are winning the same number as the temperate climate countries.

Temperate climate countries have by far experienced the most success at the Summer Games. Recently, since 1988, they have also had the most success in total at the Winter Games. Cold climate countries historically had the most success at the Winter Games, but have also had good success at the Summer Games. When we look at the average number of medals won, because there are so few cold climate countries, we can see that cold climate countries have experienced more success than either hot or temperate climate countries. As expected, hot climate countries

have struggled at the Winter Games. Even though there are 10 times as many hot climate countries as cold climate countries by our model, they have only recently surpassed the cold climate countries in total success at the Summer Games. They continue to lag behind both cold and temperate climate countries in terms of average medals won at both games.

At the Summer Games, the top 5 temperate countries each year have about 3 times as many medals as either the top 5 cold or hot climate countries. The top 5 cold climate countries have historically more medals than hot countries and more recently they are performing equally at the Summer Games. At the Winter Games, the top 5 temperate and cold climate countries performed equally, with a few exceptions.

In conclusion, in terms of participation hot climate countries tend to participate more in the Summer Games and see more success at the Summer Games, but both their participation and success is much lower than either the temperate or cold climate countries. This does not appear to extend to cold climate countries. Temperate and cold climate countries have strong participation at both the Summer and Winter Games. In terms of success, while temperate climate countries have had more total medals at both the Summer and Winter Games, this can be attributed in part due to the higher number of temperate climate countries in our model. When observing the average number of medals, the cold climate countries have more success than temperate countries. The top performing temperate countries have more success at the Summer Games and equal success at the Winter Games. This indicates that temperate and cold climate countries are more likely to have success at the Olympic Games than hot climate countries.

However, it is important to note, as with many global phenomena, climate is likely not the only indicator of success at the Olympics. The lack of both participation and success of hot climate countries and the participation rates and higher success of cold and temperate climate countries may very likely have other factors. Some of those reasons, which are not explored in this analysis, could be population, GDP, global affairs, and political systems.

Future Analysis:

Analysis of Olympics data when paired with temperature is one perspective. Other datasets that may provide insights into which countries perform better or have a larger presence at the games. For example, including population data may explain why certain countries win more medals than others. Statistically speaking, a larger population increases the likelihood of extraordinary athletes. With the large number of events at each Olympic Games, the number of extraordinary athletes required to out-perform other countries is also high. Larger populations are likely to increase popularity and following for such athletes, which in turn may lead to training from a younger age, better coaches and trainers, more athletes working toward Olympic-level competition.

Another dataset that could be linked to a country and therefore its Olympic athletes and results is the country's gross domestic product (GDP). The financial situation for a country and those who live there may affect the country's Olympic prospects. Being an elite-level athlete can be expensive. Athletes may have to travel, purchase equipment and attire, pay trainers and coaches, as well as facilities costs or membership fees. When athletes eventually get to the national teams

and international competitions, the country is also responsible for providing at least some of the financial support for these activities and competitions. In addition a country's economic development status may afford its citizens greater access to sport and news media, increasing awareness and interest and also afford them more leisure and economic opportunities to pursue their passion in sport.

The Olympics data can be analyzed with data concerning international events. Bans and boycotts were identified in the participation section. Conflicts between nations can impact whether or not certain countries attend the Olympics at all. For example, when South Korea was chosen to host the games and North Korea's suggestion to co-host them was rejected, communist countries (North Korea, Cuba and Nicaragua) refused to participate. The games were not held during the world wars. The USA boycotted the 1980 Summer Games in Moscow at the height of the Cold War. In 1976 the nations of Africa protested racial segregation in South Africa and boycotted the Games. [8]

The political system of a country may also be related to the country's success at the Olympic Games. For example, countries under a communist system or those with a dictator, may perform better as the government has full control over the people. Any promising athletes can be sent to training centers from a very young age. An article from the New York Times explains that "sports clubs" in Russia, which is a semi-presidential authoritarian dictatorship, look for children who have talent in a specific sport. The children are sent to train and eventually when "they are no longer of competitive worth", they become coaches. This has been included as one of the reasons that Russia performs so well at the Olympics. Countries under other political systems, like the democratic presidential republic United States have also performed well, but the analysis of political systems and their correlation to Olympic medals would be interesting.

[8] See Council on Foreign Relations for more information on global affairs and the Olympic Games.

References:

Chadha, Sagar. "Olympics Data- Cleaning, Exploration, Prediction." *Kaggle.com*, 2018, www.kaggle.com/code/chadalee/olympics-data-cleaning-exploration-prediction. Accessed 20 Sept. 2023.

"Climate Change: Earth Surface Temperature Data." *Www.kaggle.com*, Berkeley Earth, 2016, www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data?select=GlobalLandTemperaturesByCountry.csv.

Council on Foreign Relations. "Timeline: Politics and Protest at the Olympics." *Council on Foreign Relations*, 2022, www.cfr.org/timeline/olympics-boycott-protest-politics-history. Accessed 5 Nov. 2023.

"Datasets/Geo-Countries." *GitHub*, 3 Nov. 2023, github.com/datasets/geo-countries. Accessed 10 Nov. 2023.

Griffin, R. "Olympic History Data: Thorough Analysis." *Kaggle.com*, 2018, www.kaggle.com/code/heesoo37/olympic-history-data-a-thorough-analysis. Accessed 30 Sept. 2023.

Hargurjeet. "Data Exploration of Historical Olympics Dataset." *Nerd for Tech*, 27 Aug. 2021, medium.com/nerd-for-tech/data-exploration-of-historical-olympics-dataset-2d50a7d0611d#62a6. Accessed 5 Nov. 2023.

Herman, Robin. "How Youngsters Are Recruited for the Future." *The New York Times*, 11 July 1976, www.nytimes.com/1976/07/11/archives/how-youngsters-are-recruited-for-the-future.html. Accessed 19 Nov. 2023.

International Olympic Committee. "What Is the Origin of the Olympic Games? - History and Origin of the Games - FAQ." *International Olympic Committee*, 28 Apr. 2021, olympics.com/ioc/faq/history-and-origin-of-the-games/what-is-the-origin-of-the-olympic-games. Accessed 15 Oct. 2023.

"National Olympic Committees (NOC) - Olympic Movement." *International Olympic Committee*, 28 Apr. 2021, olympics.com/ioc/national-olympic-committees. Accessed 29 Oct. 2023.

Si, Jianxi. *Olympic Games Medal Count Analysis Summer and Winter Olympic Games*. 2018, scholarworks.calstate.edu/downloads/n870zt00q. Accessed 3 Nov. 2023.

Wikipedia Contributors. "List of Countries by System of Government." *Wikipedia*, Wikimedia Foundation, 18 Feb. 2019, en.wikipedia.org/wiki/List_of_countries_by_system_of_government. Accessed 19 Nov. 2023.