**Implementation:**

Tableau Link: <https://public.tableau.com/views/OlympicAnalysisReport/OlympicReportStory?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link>

**Executive Summary:**

Teaser image:

A graph of different colored lines

Description automatically generated

* This is our teaser image as it presents the problem we highlighted in our analysis in which we will touch on more throughout the report.

Storyline:

* The USA has been dominating the summer Olympics the past 20 years comparatively.
* So what is leading to this success?
  + They have been excelling particularly in swimming, track & field, basketball, and gymnastics.
  + They have seen top athlete medalists in these sports but mostly swimming at large with people like Michael Phelps leading the way.
* However, the US has been experiencing little growth in their performance.
  + We are not improving our performance but rather staying constant.
  + While countries like the United Kingdom are growing their medal count rapidly.
* The US should identify areas in which we have been less successful to focus on to improve our performance.
  + Sports like golf, the pentathlon, rugby, rhythmic gymnastics, etc. are all sports that we have earned no more than 1 medal in the last 20 years.
  + These are the sports that we should investigate to improve.
* Funding lower earning medal sports will lead to better overall performance.
  + Currently the top funded sports include swimming, track and field and gymnastics, all of which we already perform extremely well in!
  + Switching some of our funding from these sports to less successful sports will allow us to improve our overall performance as a country and continue to dominate the Olympics.

Big Idea:

Highlighting past Olympic trends for the US compared to other countries helps Americans see US strengths, weaknesses, and ways that the US can continue to grow and succeed going forward.

Summary of each slide:

* Slide 1: This acts as our title slide to give a super brief overview of what the rest of the visualizations are going to be about.
* Slide 2: This slide is our interactive dashboard. All the graphs are connected in the sense that if one country is selected on either of the top two graphs, the rest of them line up. This slide also mentions the scope of our data by having the titles say that they are from 1996-present and the summer games specifically. The top left is a choropleth graph that shows the higher-ranking countries highlighted in darker colors. The top right graph is a line graph that shows a timeline of the past 20 years and each country’s medal count. We have the top 5 countries highlighted as well as the UK, the rest as background lines shaded grey. The bottom graph shows the breakdown of medals (gold, silver, bronze). It Is easy to see that the US dominates the other countries by being so far above in the line graph and by the US being shaded the darked on the map graph. Along with being able to filter by country in the top right, some other interactivity on this dashboard includes hovering to see medal count and zooming on the map.
* Slide 3: On this slide, we took the dashboard from the previous slide and narrowed it down to just show the United States performance in the games. This visual shows the total medals that the US has earned (5002) along with the breakdown of each medal type for the US. From this slide on, we focus all the visualizations on the United States.
* Slide 4: The graph on this slide is a bar graph that shows the distribution of medals earned by sports. The top two sports (swimming, track & field) so we have highlighted them. While there are many different events in the summer Olympics, this graph shows only the top 10. The viewer can sort the Y axis by the field values, data source order, or alphabetically in either ascending or descending order. The top two sports have the medal count at the end of the bar, but for the other bars, the viewer can hover with their mouse to see the medal count.
* Slide 5: After seeing the sports we excel in the most, we went into the top 30 athletes that make up these sports. The key in the right corner shows that blue represents swimming, and it is clear that the majority of our top athletes are swimmers. This could be explained by the fact that swimming has multiple events meaning multiple times to earn a medal. Another sport with multiple athletes being in the top is track & field which is in the same situation as swimming with multiple events. We also have a box of extra information that dives deeper into the statistics on Michael Phelps because he performs so far above everyone else.
* Slide 6: This graph is the same line graph that is part of the dashboard. We highlight the fact that the United States is far above the rest of the countries in medal counts consistently over the past 20 years. We point out that while we are staying at the top, our performance is plateauing. Other countries like the United Kingdom (shown in purple) are increasing in medal counts after each Olympic games.
* Slide 7: This slide is similar to the slide that shows the United States top events except we are highlighting our bottom sports in which we can improve our performance. We show the sports that we have scored the least number of medals in over the past 20 years. The bar chart has golf and the pentathlon with only 1 medal but that is because we excluded the events with no medals from the graph (badminton, handball, rugby, rhythmic gymnastics, hockey, table tennis, and trampolining). Like with the top sports, we only focused on the bottom 10 events. We wanted to acknowledge these sports and the fact that we have very few medals in them because we will go into our suggestion for how to improve their performance.
* Slide 8: This slide is where we make our suggestions to solve the issue moving forward. After some research, we discovered that the top 3 funded summer Olympic sports in the United States are swimming, track & field, and gymnastics. As previous graphs have shown, there is a direct correlation between the top funded sports and the sports that we earn the most medals in. After coming to this conclusion, we want to suggest that we should start funding some of the lower performing sports like rugby, badminton, pentathlon, and golf as shown with the icons along the bottom. This change in funding will improve our overall performance as a country, allowing us to remain the top competitor going forward.
* Slide 9: On this slide we reiterated our takeaways. We included this slide because we wanted to make sure that the audience saw the information we wanted them to internalize again after reviewing this visual that might not have stood out at first glance when going through the visualization.

**Basic Info:**

**Project title**:

Olympic Data Analysis

**A list of all team members**:

Chloe Adams, Brianna Mulcahy

**Data:**

**Data Background**

We got our main data from Kaggle about the [Olympics](https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results). This dataset contains historical data on the Olympic Games, from Athens 1896 to Rio 2016. Each row represents an individual athlete competing in an individual Olympic event, with various variables describing the athlete and the event. The Kaggle dataset contains 27,117 rows and 15 columns of data.

We also used articles from [sports.yahoo.com](https://sports.yahoo.com/us-olympic-funding-leads-medals-160008387.html#:~:text=Since%20the%20turn%20of%20the,U.S.%20athletes%20on%20the%20podium.) and [swimswam.com](https://swimswam.com/usa-swimming-athletes-received-4-5-million-in-grant-money-from-usopc-in-2021/) to get more information on the funding habits of the US for the Olympics which helped us investigate where funding is currently being invested and where we may improve more from it being invested.

**Description of Data**

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **Type** | **Description** | **Domain Specific Language** |
| ID | Text | Unique ID of each athlete in the Olympics | 27,117 Unique categories |
| Name | Text | Name of the athlete competing in the Olympics | 134721 Unique Categories  Mode: Robert Tait McKenzie |
| Sex | Text | Sex of athlete competing | 2 Unique Categories (Male, Female)  Mode: Male |
| Age | Numeric | Age of athlete competing | Range: (10,97) |
| Height | Numeric | Height of athlete competing in cm | Range: (127,226) |
| Weight | Numeric | Weight of athlete competing in kg | Range: (25,214) |
| Team | Text | Name of the team the athlete is competing for | 237 Unique Categories  Mode: United States |
| NOC | Text | 3 Letter code for National Olympic Committee | 230 Unique Categories  Mode: USA |
| Games | Text | Year and season of the Olympic Games | 51 Unique Categories  Mode: 2000 Summer |
| Year | Numeric | Integer of the year the Olympic games took place in | Range: (1896, 2016) |
| Season | Text | Season in which the Olympic games took place in (summer/winter) | 2 Unique Categories (Summer, Winter)  Mode: Summer |
| City | Text | Host city that the Olympics took place in | 41 Unique Categories  London |
| Sport | Text | Sport/activity that the athlete is competing in at the Olympics | 66 Unique Categories  Mode: Track & Field |
| Event | Text | Olympic event the athlete is competing in | 765 Unique Categories  Mode: Men’s Football |
| Medal | Text | Medal earned in the event (Gold, Silver, Bronze, NA) | 4 Unique Categories  Mode: NA |

**Data preprocessing steps**

We used Excel to remove NA values from data file, so tableau sees the value as missing instead of a category of NA which makes the data easier to use.

We then further utilized Excel to separate the medal column into a gold column, silver column, and bronze column having 0/1 binary values to allow us to take the count of each medal category.

Lastly, we resourced Excel again to clean the team column in the dataset so there are distinct values for each country and resolve inconsistencies like Soviet Union and Russia or United States and United States-1. This took the longest time as we had to go through 27,117 rows of the excel file and manually update the team columns to match the NOC code for the country. In our final analysis, we ended up using only the data after 1996 making the Soviet Union and Russia issue irrelevant to the analysis because we wanted our data to be recent, so our analysis was up to date.

**Visualizations:**

A logo with a flag and rings

Description automatically generated

* Title slide
  + The main purpose of this slide is to give a glance into the topic of the United States Summer Olympics analysis
  + This slide also sets the color scheme of the slides to follow

A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

* Interactive Overview Dashboard
  + Views
    - 1st graphic represents the main view of the visual
    - 2nd graphic represents the view of the dashboard when specifying a specific country
  + Visual encodings:
    - Map
      * X location encodes longitude
      * Y location encodes latitude
      * Color saturation/luminance encodes medal count
    - Line chart
      * Y axis encodes medal count
      * X axis encodes year
      * Color encodes country
    - Bar chart
      * Y axis encodes medal type
      * X axis encodes medal count
      * Color encodes medal type
  + Interactivity:
    - Hovering over lines, bars, or countries shows more information on the variable and the count associated with it. The map has additional interactivity for zooming in on the countries. Furthermore, all the graphs are linked to one another so if you select a country it switches to that country on all other graphs. There is also a filtering option by country linked to all the graphs.
  + Rational:
    - The color of the lines or bars show either what country or what medal type earned which are both categorical meaning it needs hue to encode it.
    - We used saturation/luminance to encode medal count in the map as it is a quantitative value and has numerical meaning. We also chose to use a continuous color palette so that are audience sees the big ideas without focusing too much on exact values.
    - We redundantly encoded the medal type but thought it added another layer of remembering the medal by the color it usually was.
    - We greyed out the many other country lines to reduce clutter and highlight the key countries of importance in the line chart.
    - One thing that we also noticed was that the medal count did not add up to equal values. After seeing this we did some research on the data values and found that they were uneven because of ties for medals that occurred in different events.

A screenshot of a graph

Description automatically generated

* Top Sports graph
  + Visual encodings:
    - X axis encodes medal count
    - Y axis encodes sport
  + Interactivity:
    - Hovering over each bar shows the name of the sport and how many medals the US has accumulated in that sport between 1996-2016. The Y axis can be changed by the viewer so that you can sort however you want. The options are alphabetically, data source order, or by the field in ascending or descending order.
  + Rational:
    - This is helpful to show which sports the United States excels in throughout the summer games.
    - We used the blue bars and the sports highlighted in blue in the title to show the top two sports and also used the data labels to highlight them even more and greyed out the other bars to highlight the top two medal earning sports even more

A screenshot of a graph

Description automatically generated

* Top Athletes graph
  + Visual encodings:
    - X axis encodes medal count
    - Y axis encodes athlete
    - Color hue encodes sport
  + Interactivity:
    - Hovering over each bar shows the name of the athlete and how many medals they have. The Y axis can be changed by the viewer so that you can sort however you want. The options are alphabetically, data source order, or by the field in ascending or descending order.
  + Rational:
    - The color of the bars shows what sport that athlete medaled in and is a categorical variable which needs hue to encode it as.
    - This graph shows the top athletes by sport highlighting the most prevalent medaling sport by athletes.

A graph of different colored lines

Description automatically generated

* No Growth Graph
  + Visual encodings:
    - X axis encodes year of Olympic Summer Games
    - Y axis encodes medal count
    - Color hue encodes team
    - Color containment encodes area of interest
  + Interactivity:
    - Hovering over each line shows the country’s team and how many medals that the country has in each particular year.
  + Rational:
    - This is helpful because it shows that while the US is performing well, they are not improving their performance at all while some other countries are suggesting they should find ways to improve their performance.
    - We used the blue containment shading to draw eyes to the United States and showing that they are experiencing constant performance.
    - We used color to encode the teams because it is a categorical variable which means it needs to be encoded by color hue.

A graph of a golfing competition

Description automatically generated with medium confidence

* Bottom Sports graph
  + Visual encodings:
    - X axis encodes medal count
    - Y axis encodes sport
  + Interactivity:
    - Hovering over each bar shows the name of the sport and how many medals the US has accumulated in that sport between 1996-2016. The Y axis can be changed by the viewer so that you can sort however you want. The options are alphabetically, data source order, or by the field in ascending or descending order.
  + Rational:
    - This is helpful to show which sports the United States perform poorly in throughout the summer games and shows an area that the US can improve their performance in.
    - We used the red bars and the sports in the title to highlight the bottom two sports and used the data labels to highlight them even more and greyed out the other bars to highlight the bottom two sports even more.

A screenshot of a sports information

Description automatically generated

* Suggestion Slide
  + The main purpose of this slide is to highlight our suggestion for improving USA growth in the Olympics.
  + The top left graphic shows the top 3 summer Olympic sports receiving funding and the amounts. This visual is important in showing that the top medaling sports are the one that are receiving the most money.
  + The bottom visual shows 4 sports in which we suggest the US increase funding to as they are lower performing sports that would benefit heavily from more funding. These sports include rugby, badminton, pentathlon, and golf.

A blue and white text on a blue background

Description automatically generated

* Appendix slide
  + The purpose of this slide is to synthesize all that was covered in our project so that the main ideas and takeaways are clear.

**Usage Scenario:**

Include a usage scenario walking through how your current visualization can be used during an interactive session, illustrated with screenshots of your system in action.

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

* Hovering
  + By hovering the mouse over the countries on the map, the name of the country and their medal count from 1996-2016 pops up. The same goes for points on the line graphs and each bar graph. The line graph will also say what year the data point you are hovered over is from.

A screenshot of a graph

Description automatically generated

* Zooming
  + The map on the dashboard has a zoom feature on the left side. There is a plus and minus button to zoom in/out. It’s also possible to zoom with two fingers on the mouse pad. The viewer can also drag the map along after zooming in. This allows the audience to better view countries that are smaller in size and compare the color accordingly.

A screenshot of a computer screen

Description automatically generatedA screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generatedA screenshot of a computer screen

Description automatically generated

* Linking Graphs
  + On the interactive dashboard all the graphs are linked in the sense that if a country is clicked on the map, the line graph will show only that country, and the bar chart will show only that country’s medal breakdown (the top pictures). The same is true if the viewer clicks on a certain point on the line graph (the bottom pictures). If the viewer wants to see a country’s performance in a specific year, they can click on that data point on the line graph and the medal count will adjust to just that specific Olympic games.

A screenshot of a graph

Description automatically generated

A screenshot of a computer

Description automatically generated

* Filtering Graphs
  + If the viewer wants to view data for a specific country but might not know where it is located on the map to be able to click on it, they can click the drop-down menu in the top right corner and search for the country. This will make all of the visuals adapt to show that specific country’s data the same way it would if the country was clicked on the map.

A screenshot of a graph

Description automatically generatedA graph of athletes with numbers and text

Description automatically generated with medium confidence

A graph of athletes making up

Description automatically generated

* Sorting In Bar Graphs
  + Each of the bar graphs in our visualization analysis have a button at the top of the Y axis that allows the viewer to decide the order of the items. The choices to sorting by the field values, data source order, and alphabetically in either ascending or descending order.

**Reflection:**

Most enjoyable part of the project?

* The most enjoyable part of the project for us was creating the visualizations. We liked the process of creating the visuals, dashboards, and graphs once we had our storyline in place. It was super rewarding to use all the skills that we learned throughout the semester like using the correct types of charts, color, maps, and interactivity skills to the test to create effective visualizations.

Least enjoyable part of project?

* The least enjoyable part of the project was cleaning the Kaggle dataset. There were many inconsistencies in the team column of the data. This was super important for us to fix as we knew we wanted to use the map functionality and would need to fix all the team names, so they aligned. One of the problems with the team names were that country names changed over time like Russia and the Soviet Union, and we had to use our best judgement with cases like these to either keep them the same or change them. We ended up not changing the team's name for the Soviet Union and just used data from 1996-2016 to eliminate this issue. We also had issues with team names with extra values added to the end of them like United States-1 and Russia-3. For this issue we had to go through the entire excel file and manually change the values back to just the country name. This process ended up taking us a long time because of the many rows in the file.

How has the project developed from beginning to the final product?

* Our project has really developed from the beginning into a better story. Initially we were planning to just look at an overview of our past performance in the Olympics and what the US should expect for the upcoming summer games. However, we shifted our story to look at our performance to see our strengths and our weaknesses to highlight areas we should improve so that we can continue to remain successful in the games going forward. This really helped progress our visualization as we provided a suggestion that was guided by our data analysis in which we suggested that we should start funding lower medal earning sports.

How have your visualization goals changed?

* Our visualizations goals changed when we changed our storyline to looking at US Olympic weaknesses and how we could improve overall performance. This changed the visuals we needed to support this new storyline. We implemented a bottoms sport visual and a table of the top 3 funded summer Olympic sports to show that the bottom performing sports is where we should fund as it will lead to higher overall performance for the US.

How have your technical goals changed?

* Our technical goals did not change, we were able to implement all our visualizations the way we wanted using Tableau.

How realistic was your original proposal in terms of what is possible in tableau?

* Our original proposal was realistic in terms of what is possible in tableau. We created the initial drawn draft with tableau and the data in mind so that our visuals would be realistic to create in Tableau. The one issue we ran into with our initial draft was the map graph. In our initial graph we were going to use a symbol map with dots representing the total medal count for each country. However, when we created this graph in Tableau the dots overlapped and were hard to discern on smaller countries. This created an issue of being able to roughly estimate the medal count values comparatively among different countries.

Was there anything you wanted to implement that you couldn’t?

* We wanted to implement a graph showing the US funding for differing Olympic sports overtime. However, we could not find a source or data that contained this information. Instead we found a data source that showed the top 5 most funded Olympic sports and used this to create an informational slide rather than creating a graph. In the end it worked out fine, but we wish we could have had a graph showing funding trends.
* One issue we ran into when working on this project was when our medal counts did not add up to equal values as we discussed in an earlier part of the report. We found that these differences were due to ties in the Olympic events leading to unequal counts.

If you were to do the project again, what would you do differently?

* If we were to do this project again, we would spend more time exploring the data and really figuring out what we wanted our story to be. This is because we spent a lot of time cleaning data and creating graphs that did not really support our final storyline. If we would have known that we were looking at the Olympics with a focus on our successes and where we were to improve based on recent data, we would not have spent so much time cleaning the data that we did not even end up using in our analysis. In our final story and analysis, we only ended up using data from 1996 and onward. However, before we knew this, we cleaned the entire excel file fixing all the country names which took us several hours when if we knew what our final analysis was we would have a smaller subset cutting the time cleaning by a lot.

**Project Management and Team Assessment**

Throughout the entirety of the project, we used Snapchat as our main form of communication. We were both able to respond in a timely manner on this platform. We also used free time in class and after class to discuss changes we want to make to the project and set guidelines for what we want to complete and when we want it done by.

**Work Breakdown**

* Chloe
  + Searched for datasets (10/30) (1 hour)
  + Preprocessed Kaggle file (5 hours) (11/2)
  + Created Tableau graphs and dashboards (7 hours) (11/27)
  + Created Tableau Story (3 hours) (11/27)
  + Data section of initial project proposal (40 minutes) (10/30)
  + Team communication plan section of initial project proposal (10 minutes) (10/30)
  + Data processing section of initial project proposal and project updates (15 minutes) (10/30)
  + Edit and look over Project update 1 (45 min) (11/10)
  + Edit and look over Project update 2 (45 min) (11/27)
  + Used past project updates to write the report and added additional information (2 hours)
  + Worked on project presentation (1 hour)
  + Recorded project presentation and finalized the report (1.5 hours) (12/5)
* Brianna
  + Searched for possible data sets (15 minutes) (10/30)
  + Created sample graphs on Tableau (45 minutes) (11/4)
  + Project proposal write up (1.5 hours) (11/2)
  + Project Update 1 report writeup (1.5 hours) (11/9)
  + Project Update 2 report writeup (1.5 hours) (11/27)
  + Worked on project presentation (1 hour)
  + Used past project updates to write the report and added additional information (1 hour)
  + Recorded project presentation and finalized the report (1.5 hours) (12/5)

**Credits:**

Any Sources of Inspiration:

* We got inspiration for our tableau story by a video that was posted for our reference for our mini project 2 by visual intelligence consulting. They provided a sample [template](https://vizualintelligenceconsulting.com/leadlanding/tableau-executive-story-template/) for use of the public. In this story, we learned a lot about using a color theme and adding text explainers to our story to make it seem connected and seamless. We used this source as our inspiration for our mini project and now this final project.

Explain what changes you made and their magnitude (e.g., unchanged vs. minor tweaks vs. major functionality additions) for any code that you built upon.

* The main changes we have made were mainly from changing our storyline, this led us to create more graphs looking at bottom and top sports. This ended up being a major update to our project.
* We also added a title slide, suggestion slide, and appendix slide which enhances our story and ensures understanding for the audience. This was a minor tweak we made so that our story flowed better.
* We also changed our dashboards a little bit to increase readability by placing keys, updating fonts and font sizes, and simplified text explainers to solidify story line and understanding. These were minor tweaks we made to make the story clearer and connected.