

## DATA ANALYSIS AND VISUALIZATION

### About the dataset

- The data set is about the Summer Olympics medal from 1976 to 2008 which contains various fields like city, year, sport, discipline, event, gender, etc.
- Data observations and cleaning:
  - There was separate record for Russia, formerly known as Soviet Union. We have combined both of them together.
  - Also, East Germany, West Germany and Germany are clubbed together for better visualization, since they are all same.

### Storyline

We want to know which country is accumulating a greater number of medals and in which events and disciplines and who is the athlete who is contributing the most to his/her country.

### → Lollipop Chart

#### Why this chart?

- I used this chart because this chart is best to show the Ranking.
- Here, I used respective country's map in order to identify them easily.
- I placed count of medals variable on Y axis because, it was hard to see them on X axis.
- I used these encodings because I wanted to find out the country with the highest number of medals.

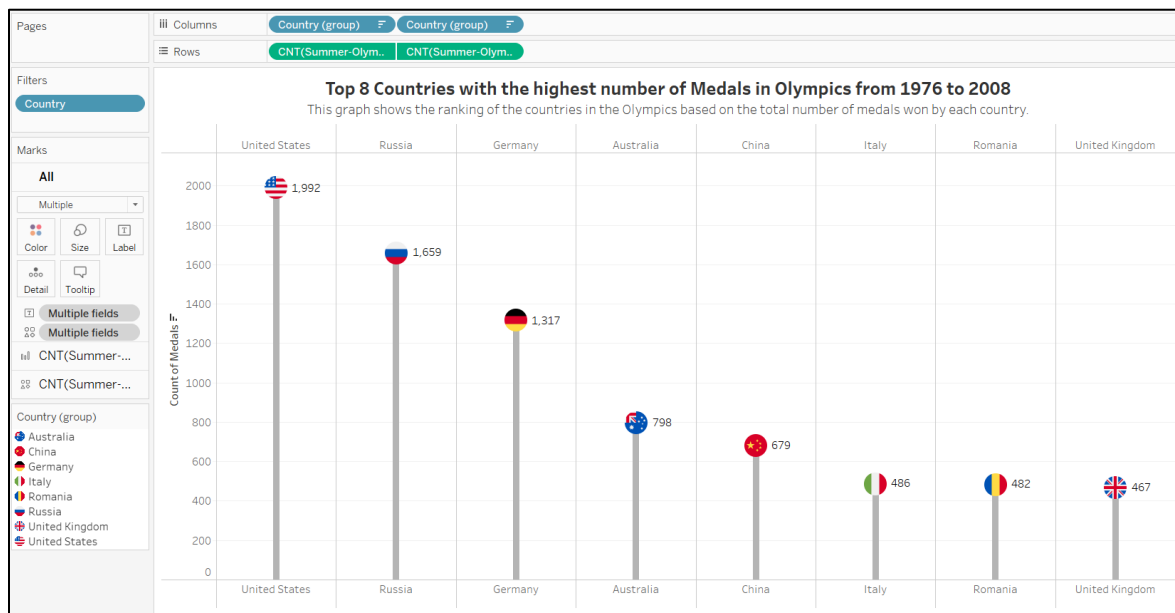


Figure 1: Top 8 countries with highest number of medals in Olympics from 1976 to 2008

- This chart is the Lollipop Chart.
- We can see that United States is leading the medal tally. Russia is followed by United States and is at the 2<sup>nd</sup> position, and at the 3<sup>rd</sup> position is Germany.
- All these three countries have more than 1000 medals won from all the sports from 1976 to 2008.
- Australia and China are the countries which has medals between 500 and 1000.
- The last three countries in the top 8 are Italy, Romania, and United Kingdom with total medals of 486, 482, and 467, respectively.

Now, as we know that United States is at the top in the medal tally, let us explore that who is the contributor of these medals based on gender. For that, we will use butterfly/distribution chart.

### → Butterfly Chart / Distribution Chart

#### Why this chart?

- I used this chart because this chart is good to show the distribution between two variables.
- Here, I used two different colors to distinguish between Men and Women.
- I have also added some annotations where required to provide the necessary information to the audience.
- I used Sport variable on y-axis because it was the best way to create a butterfly chart.
- I used this encoding to find out the proportion of men and women with their contribution of winning a medal for respective sport for United States.

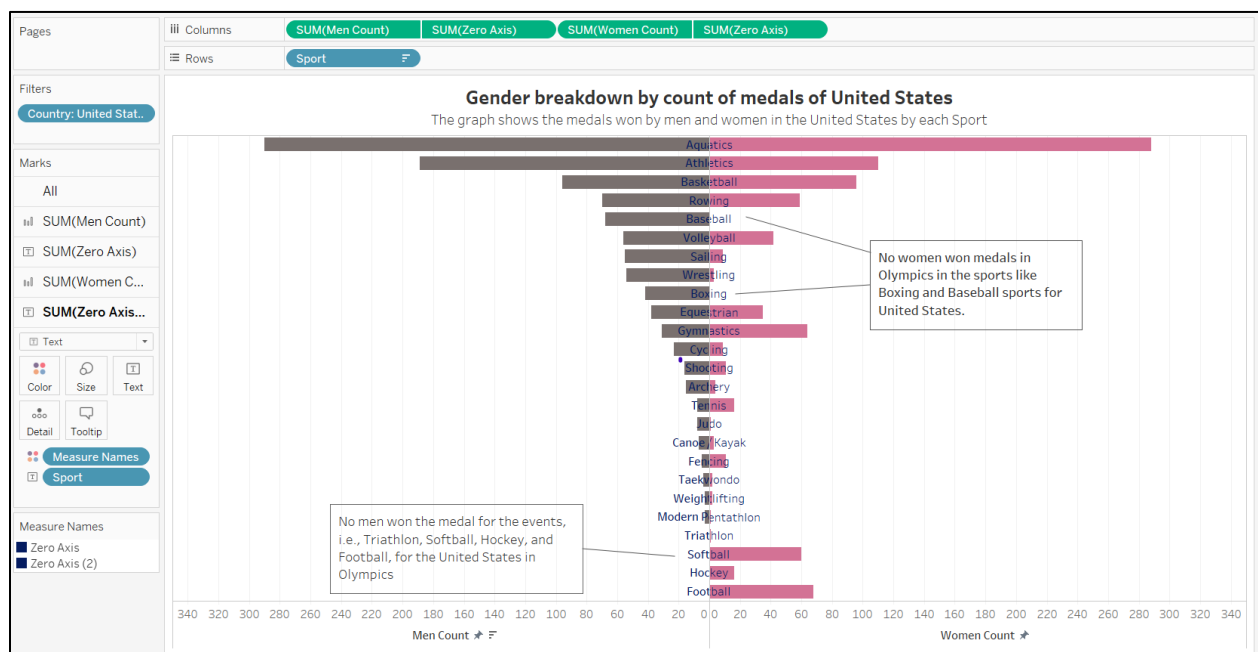


Figure 2: Gender breakdown by count of medals of United States in Olympics

- This chart is Distribution or Butterfly Chart.
- We can see from the above graph that, the greatest number of medals for the United States is won in the Aquatics, Athletics, Basketball, and Rowing.
- Even, the proportion of men and women for these sports are almost same, except Athletics, i.e., 189:110.
- There are some sports where men and women are not able to win any medals for United States. These sports include Baseball and Boxing for women, and Triathlon, Softball, Hockey, and Football for men.
- In total, we can compare that, for each sport, men and women are contributing equally to win the medals, except some for United States.

To dive further in the story, let's see that, were men and women contributing the same amount of medals to United States throughout the period or not? For that, let's compare it with line chart.

### → Line Chart

#### Why this chart?

- I used this chart because I wanted to see the top 2 sports for United States which brings highest number of medals by gender throughout the period.
- Here, I used two different colors to distinguish between Men and Women.
- I have also added some annotations where required to provide the necessary information to the audience.
- I used time as an independent variable, and it must be shown on horizontal axis.
- I used this encoding to find out the count of medals by male and female players in Aquatics and Athletics sport over the period from 1976 to 2008.

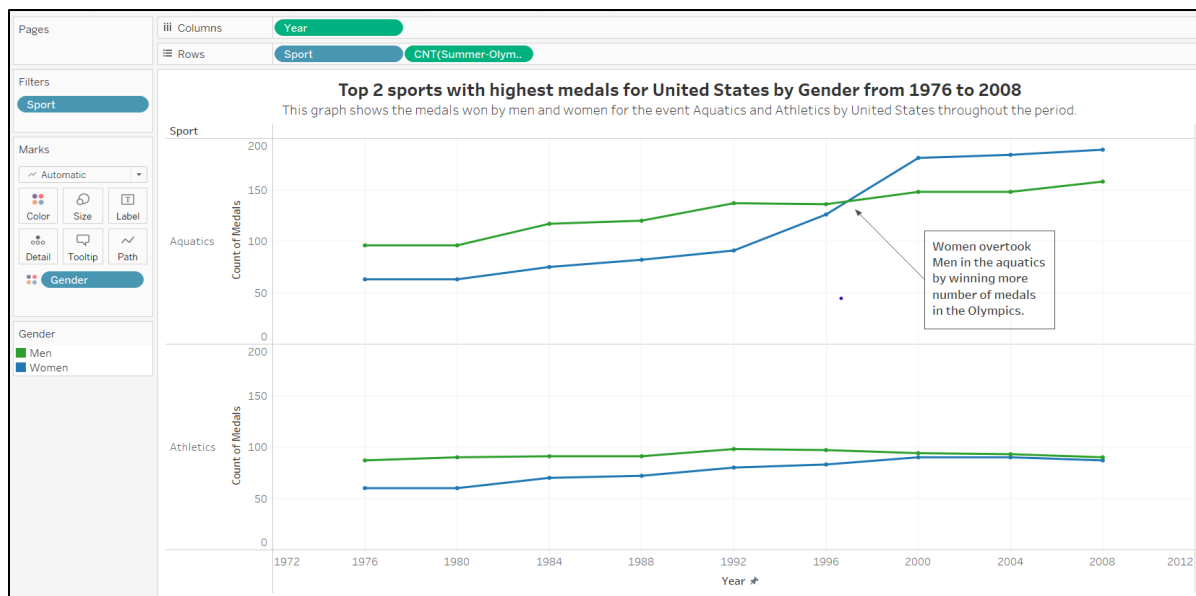


Figure 3: Top 2 sports with highest number of medals for U.S. by gender in Olympics from 1976 to 2008

- This chart is line chart.
- We are comparing the top 2 sports from the previous chart by gender for the United States from 1976 to 2008.
- We can see that, for Aquatics, men were dominating for U.S. from 1976 to 1996.
- In 1997, women crossed men by winning more number of medals in the Olympics of year 2000, and the trend continued till 2008.
- For Athletics, men's contribution of winning is same from the start till end of the period.
- But women's contribution of winning the medals in athletics is slightly increasing year by year, and the result can be seen in the year 2008, where men's and women's winning are almost same in the Athletics.

Furthermore in the story, we want to explore that, as aquatics being dominant sport for United States of winning medals, who are the top 10 Athlete with the highest number of medals i.e., Gold, Silver and Bronze?

### → **Stacked Bar Chart**

#### Why this chart?

- I used this chart because it was the best way here to compare the players with their count of medals in Aquatics sport.
- Here, I used three different colors to distinguish the counts and type of medals they won, Gold for 1<sup>st</sup> position, Silver for 2<sup>nd</sup> position, and Bronze for 3<sup>rd</sup> position.
- I have also added the number of medals won by each, that is Gold, Silver, and Bronze, along with the total number of counts of medals for each player.
- We can also invert this graph and change the axis for x and y respectively, but to view it properly, I used Athlete names on the X-axis.
- I used this encoding to find out the count of medals by the top 10 players in aquatics sport with their individual medals count as well for United States.

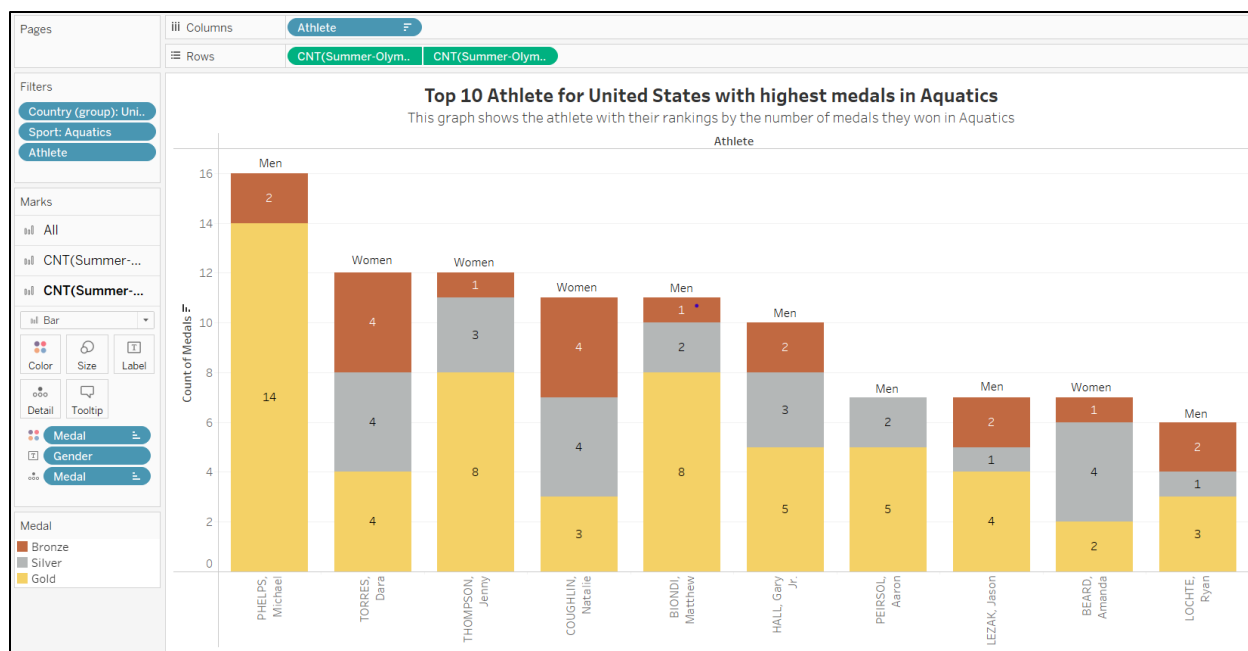


Figure 4: Top 10 Athlete for United States with highest number of medals in Aquatics

- This chart is bar chart.
- We can see in the plot that, in the top 10 list of Aquatics Sport, men and women are mostly same contributors for the medals for United States.
- But, Michael Phelps is dominating the list here, by achieving 14 Golds, and 2 Bronze in Aquatics.
- Second and third position is acquired by women and contributing the total of 12 medals each.
- So, in nutshell, we can derive the ranking of the athlete in the aquatics sport by the number of medals they won for United States.

Now, at last, we know that Michael Phelps is ranked 1<sup>st</sup> for bringing the most number of medals in Aquatics for United States. Then, let's explore that, for which events in the Aquatics, does Michael Phelps brings the medal from or which event he is dominating the most in aquatics?

### → Side-by-Side Chart

#### Why this chart?

- I used this chart because side by side chart is used to compare the values of different categories with each horizontal axis.
- Here, I used two different icons to distinguish the position/medals achieved by Michael Phelps.
- We can also invert this graph and change the axis for x and y respectively, but to view it properly, I used event names on the X-axis.

→ I used this encoding to find out the number of medals won by Michael Phelps in Aquatics in different event within swimming discipline.

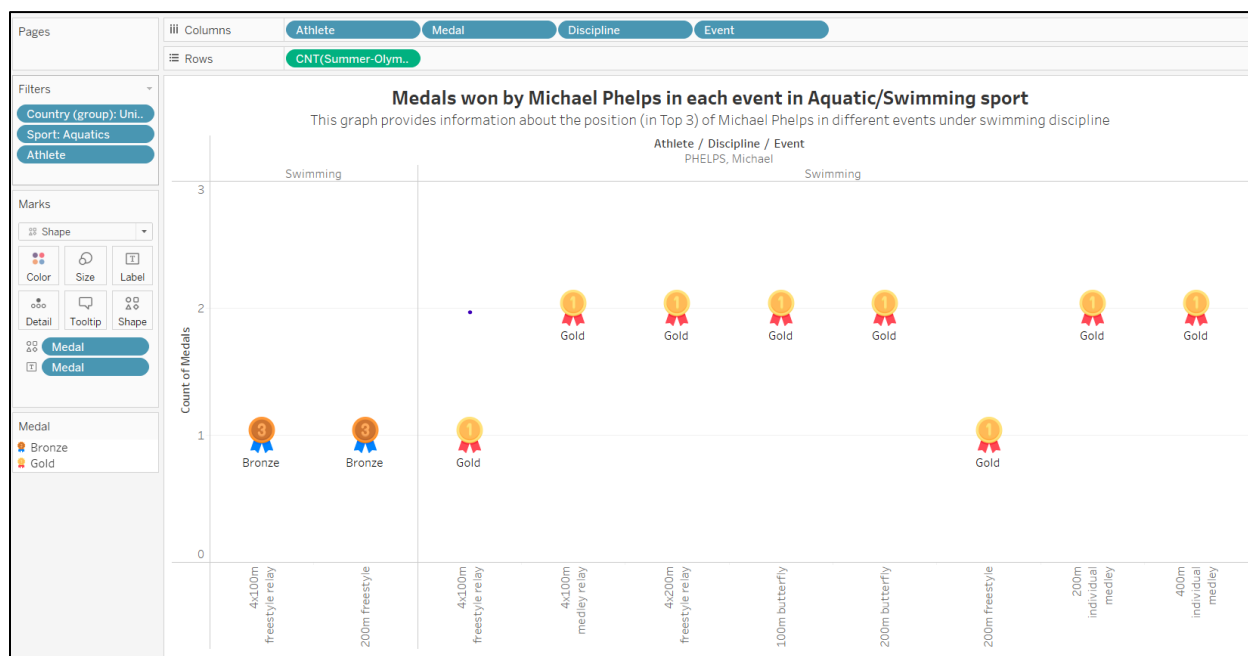


Figure 5: Number of medals won by Michael Phelps in each event in Aquatics within Swimming discipline

- This chart is side-by-side chart.
- From all our discussion so far, we reached till Michael Phelps and we found that he is bringing the most number of medals for United States in Aquatics.
- Now, this chart shows us that, the discipline he is dominating is Swimming.
- There are various events in swimming like 4x100 freestyle relay or 200m butterfly, etc.
- There are 18 events in the Swimming discipline in which Michael is dominating in 10 of them.
- He is bringing the gold medals in the events like 100m butterfly, 200m butterfly and many more, and bringing bronze medals in the events like 4x100 freestyle relay and 200m freestyle.
- One interesting thing is that, he has won only 3 silver medals throughout his career, which is not part of our dataset. (Source: <https://olympics.com/en/athletes/michael-phelps-ii>)