Project 2 Report

Shrey Gupta (sg685), Luke Goodman (lng26), Wei Tang (wt265)

Data Sources

We obtained the 100 meter olympic data online through the www.olympic.org website. This website only included data like the names of the athlete, which country they represented, their speed, and their medal. We wanted to figure out more about the statistics for the medalists so we obtained information such as age, weight, height, and race through Google search, Wikipedia, and www.sports-reference.com. The information collection took longer than we had initially thought because the information was scattered and we had to enter the data manually.

Variables

The variables we ended up using in our infographic are the athlete name, country, competition year, speed, medal, age, and region. We decided that weight and height did not really provide any significant statistical patterns. We calculated the BMI of each athlete using height and weight, but the results showed that all the athletes except a handful were in the normal weight range for their height. Since this is not meaningful data, we decided not to show it.

Data Filtering

The data was simple to begin with, so we wanted to make an interesting visualization that has not been done before. We decided to use the athlete's average speeds and compared those to the fastest runner to calculate how far each athlete would be behind when the fastest runner crosses the finish line. This showed a pretty wide distribution and it is interesting to see how the speed has changed so much over the years. The formula we used to calculate the distance behind was: **(average speed of fastest runner - average speed of runner)*(time of fastest runner)**.

In order to create the visualizations as shown, more data filtering was necessary. We needed to go through the list of sprinters in order to figure out how many medals each country won (also how much of each medal). We did this by creating a map from country code to a list that contained the number of gold, silver, and bronze medals.

Also, in order to display which countries have participated and which have not, we needed to get all the country codes (which was a part of data collection). We also needed to get the map codes for each country because that is different from the country code.

Our data is contained in a file called **data/data_dash.csv**, but we converted the data to json and is contained in the directory **data/data_dash.js**. This file splits up the data into a complete list, male sprinters list, and female sprinters list. We split up the data this way because it made more sense when creating the bubble charts.

Visualizations: 100m Race Simulator

The race simulator was the element that cemented this dataset as our choice. It allows users to compare notable runners from throughout the years as if they had been recorded visually on the same track. It also allows them to compare themselves (or others) to these Olympic sprinters as one another. Each sprinter is given the silhouette of a running person on a horizontal plot line that represents the 100m track. The images are animated to “run” from the far left of the line to the scaled distance from the fastest runner on the far right on page load and when a user presses the “Enter” key or submits a time. Although this time is not actually representative of the run time of one of these 100m races (for example, it doesn’t actually take Usain Bolt 9.63 seconds to reach the end), we decided it was more user- and attention-span-friendly to speed up the process, so the duration of the animation is adjusted t0 approximately ⅕ of the time it would take for the top competitor to reach the finish line.

The interactive elements of this visualization include hover labels, a race slider, and the ability to add a custom sprinter. When a user mouses over a silhouette image or a name in the list to the left, a label appears above the representative silhouette with their name, the year of the race, and their time as well as a horizontal measuring line below them that displays their distance from the first place finisher at any given point. When a user slides the range input below the race line, the runners slide to their position at that time on the slider and all labels and distance measurements adjust as well. Users can also select different combinations of runners from the dropdown menus below the slider as well as add custom runners to the race. Each time one of those two inputs are used, the race resets and animates again. This will be discussed below, but when a custom runner is added, the bubble charts below the simulator are also affected.

We limited input times to 10-23 seconds because this is the average range for adults and so that times would be fairly accurately entered (as fun as it may be to say you ran the 100m faster than Usain Bolt). The race always initializes with Usain Bolt, the world record holder in this event, and 4 random runners from the list of notable medalists.

This is an interesting visualization because it can display the general 100m performance differences between notable runners from different time periods, backgrounds, genders, etc. It is also a representation of how much faster these athletes are than your average human being (the fastest of us average runners finish between 13-15 seconds) and how that 3-4 seconds might look on a track with these Olympians. This fun interactive tool is expanded upon by more traditional and informative visualizations that follow it.

Visualizations: Bubble Charts

We produced 2 bubble graphs for our project. The graphs are exactly the same in structure, but represent different data (male vs. female sprinters). Since it is not fair to compare female to the male sprinters, we decided to split up the graphs according to gender. On the x-axis is the meters behind the fastest runner for that gender (Usain Bolt for male, and on the y-axis is the Olympic competition year. We placed a circle on the plot for each athlete who obtained a medal and colored the circles according to the medal received. The scales for this graph were pretty straightforward because it is linear data.

The interactive elements for these graphs are through tooltips and hover events. If a user wishes to learn more about a particular point on the graph, a tooltip appears when you

hover over the circle to provide detailed information which includes: athlete name, country, year, time, and meters behind fastest. Another way to interact with the graph is to hover over the country list on the left of the graph. This highlights the circles (athletes) that represent the country. This is an important visualization because it can show the user when a country's athletes were doing well and how many medals a country has. They can even use this to compare how different countries did if comparing how many bronze, silver, or gold medals.

Also, if a person adds their own time in the inputs, then they have a separate circle appear on the graphs to show how they did in comparison to the Olympic runners. The color for the circle is selected randomly using a function. The new runner is defaulted to USA because we all currently study here. This circle has the same hover over functionality. The graphs scale accordingly with the input. Since all the olympic athletes would be faster than any average person, the points would not appear on the graph if times were greater than 14 seconds. That is why we re-scaled everything on the graph if the points exceeded the graph's dimensions.

We thought that these two graphs would be interesting because they show the distribution of how athlete's speeds have changed over the years. If we did a simple plot where we graph speed vs. year, the data would be more clustered and wouldn't help people visualize what that data means except that sprinters have gotten faster. With this graph, you can see how far behind each runner would be from the fastest time and this is better to visualize because people can compare to distance better than time.

In general, people got faster over the years, but there are some interesting things to notice in the graph. For male runners, there was a period of 30 years (1930 - 1960) where the fastest time was not broken. The fastest time was set at 10.30 seconds in 1932 and was not broken until 1964. This was also during the time of the second world war and the depression, so that may have caused a lack in talent or training. For male runners, the time has gotten better and better after 1964 and it was set to its current record at 9.63 seconds by Usain Bolt. For female runners, the time has gotten better as well over the years, but it was broken in 1988 by Florence Joyner with a time of 10.54 seconds and has not been broken since. No runner has even gotten close in the past 25 years.

Looking at countries who have won medals, it is interesting to see that the United States has dominated the games from the beginning. The European countries like Germany and the United Kingdom have a significant amount of medals as well, but they are more concentrated in the center of the graph which means that they had wins in older Olympic games. Countries like Jamaica and Trinidad have a lot more recent wins, and shows a trend of fast runners coming from the Caribbean.

Visualization: World Map

We produced the world map using the json file **data/world-50m.json** from lectures. We also did data filter of the original data_dash.csv. Attributes of name, time, country, gender, place were selected from data_dash.csv and we checked the ISO-3166 numeric ids of each countries (http://en.wikipedia.org/wiki/ISO_3166-1_numeric) because the data in the JSON file does not contain country names, only ISO-3166 numeric ids. After the data collection, we sorted the records by country, found the corresponding ISO-3166 code, grouped data by

gender and found the fastest man sprinter with his time and the fastest woman sprinter with her time. It is interesting to see the fastest man and woman in a country which has got a medal, but several countries don't have a record for both man and woman and we just showed no man medalist in the tooltip. For each country, we also calculated the number of gold, silver and bronze medals the country has got in Olympic Games. We converted the filtered data to a JSON file called **data/CountryOfMedals.json.**

There are a total of 23 countries that won at least one medal in 100 meter competition in the Olympic Games. We also considered the regions and countries which never participated in the Olympic Games, e.g. the Antarctica, Greenland, Guinea-Bissau.

We used a linear color scale (from light pink to dark purple) to represent the total number of medals received by a country. The very light pink represents the countries which never won a medal. The grey represents countries or regions that never participated in the Olympic Games. When the user hovers over the countries on the map, a tooltip appears with the medal counts, the fastest male runner, and fastest female runner from that country.

From the world map, we can see obviously that the USA owns the maximum number of medals in the past 116 Olympic Games. And some countries in Africa has never participated in Olympic Games. Although about 200 countries and regions have taken part in the Games, most of them haven't received a medal in the 100 meters competition. Contrary to many stereotypes that might imply that most medalists come from the Caribbean and Africa, the map shows that medalists actually come from mostly North America and Eastern Europe, followed closely by the Caribbean.

Conclusions

Overall, the project was very exciting. It was very interesting to see how older olympic runners have performed over the years. We really enjoyed adding dynamic elements to a static page. It really enhanced user experience and can show a lot more about the data. As with all projects, there were some challenges we ran into when we trying to code what we we envisioned, but we overcame those obstacles and we have a great end result. We consistently were able to help one another with bugs in the respective visualizations that we focused on, producing a fun and informative series of interactive final products.

Works Cited

1. www.olympic.org - Medalist information: name, country, competition year, medal, time, gender
2. www.sports-reference.com - Athlete information: age, weight, height
3. www.google.com - Athlete information: race
4. www.wikipedia.com - Athlete information: age, weight, height, race and ISO-3166-1 numeric code for each country.
5. http://www.nczonline.net/blog/2009/05/26/computer-science-in-javascript-bubble-sort/ - Swap and BubbleSort functions
6. http://stackoverflow.com/questions/4187146/display-two-decimal-places-no-rounding - Convert number to 1 decimal place
7. http://www.moqui.org/down.png - Image for down arrow
8. http://i874.photobucket.com/albums/ab308/shotklok13/runner.png - Image for runner
9. https://farm4.staticflickr.com/3114/2445679036_5d4ba2a592_o.jpg - bacground image of track
10. Lecture notes from 022315
11. http://colorbrewer2.org - choose sequential color scale