ECS 272 Final Project Report

Jacob Lin, Xander Fang

March 2019

1 Goal and Approach

Our goal of the final project is to build a visualization tool for studying how different countries' achievements have changed over the last 100 years.

To achieve this goal, we found a data set that contains all records of medal winners in all Olympic games and select the 50 countries that win most medals in each game. We draw bubble set contours for these fifty countries to see what continent they are from and how many athletes from each country participate in an Olympic Game and how many of them won a medal for the country. We have a slide bar at the bottom of our graph for users to select an Olympic Game to analyze and a button to show the changes over years.

2 Experience

Our experience of this project has been smooth for the most part. However, there are some unexpected parts and complex parts during the implements.

The most unexpected part is about the parallel processing of setting time interval. At the beginning, we implemented one *setInterval* function for each continents. But we found the results of our system were unexpected, that some continents ran very fast, while some continents ran very slow. The result is the *setInterval* function is a parallel processing function. So, if the computing process is complex, the corresponding continents will run slower than others. At the end, we just use one *setInterval* function for all the continents and draw bubble sets over scatter plots of all the continents at the same time in each frame.

In addition, there are some complex parts. For example, to make our system runs smoothly, we decide draw 60 frames each second. As a result, we need to calculate the data and build virtual edge of continents for each frame. What's more, we need to implements the stable interfaces to insure there is no conflict between every operation in our system. What's more, we need to control the relationship of the data of each Olympic games and the data during the playing process between two Olympic games.

3 Implementation

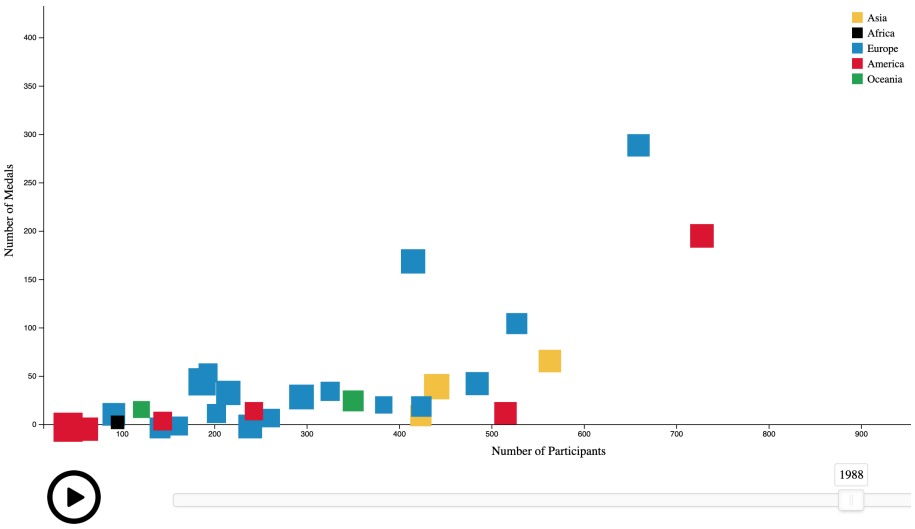

Figure 1: Normal Scatter Plot.

While the paper provides examples of bubble set contours over multiple graph types, we only implement the bubble set contour over a scatter plot. We believe that the bubble set contour makes more sense over a scatter plot than other graphs and we can implement more interactions with the scatter plot while other graphs has limitations of building interactions.

We have not only implemented the effects of bubble set over the scatter plots in the paper, but also finished some interactive visualization effects, which we'll talk in the *Differences from The Paper Version* parts.

We generate a virtual edge passing through member nodes, which to avoid overlapping within bubble boundaries as follows.

First, we generate a new control point at a corner of the obstacle's boundary. The side we choose at the beginning is the side with smaller area. Then make tests repeated. If the test fails, we use the diagonally opposite corner and generate additional control point at the corners. The algorithm loops until we get a route edge around the obstacle. The final set of virtual edges allow the set contour to avoid the obstacles and remain connected.

So, we believe that the contents in the paper are reproducible.

4 Differences from The Paper Version

We not only implemented the effects of bubble set over the scatter plots in the paper, but also finished some interactive visualization effects the paper didn't implement. We'll talk the difference as follows.

The most important difference is we generate a control button and a range-slider

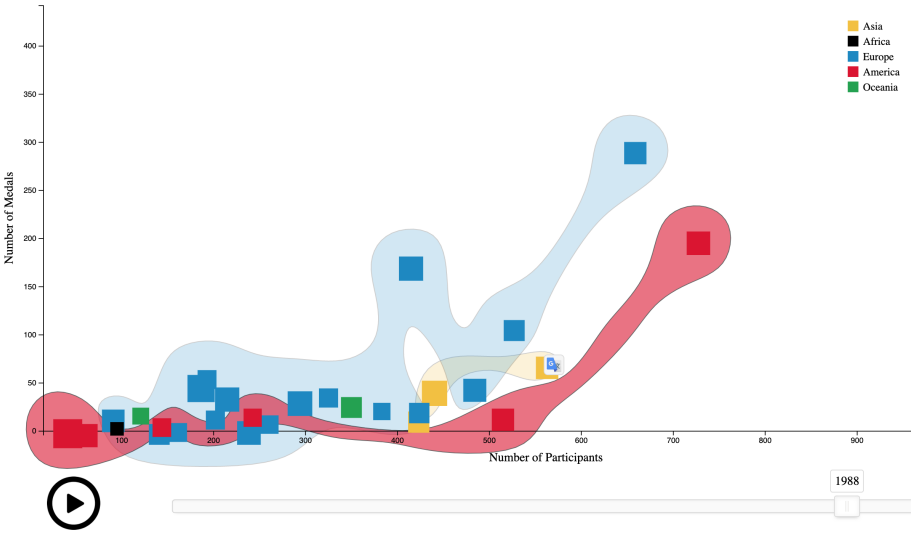

Figure 2: Bubble Sets over Scatter Plot(Hovering on the bubble set to change the opacity).

component. So, we not only could control the play and stop of our system as we want, but also have the ability to choose the beginning years when we re-play our system.

In addition, we implemented the hover function of each squares. When hovering on a square, there will be a label to display the corresponding information of the square, such as country, NOC, continent, number of participants(all, male and female) and number of medals.

Besides, The paper describes the algorithm for generating a contour over square data points, but the scatter plot with bubble contour in the paper is one with circular data points. The algorithm describes how the contour can go through each of the vertices of the square. We believe the circular version is based on squares too but the squares are invisible and only circles can be seen by users. For our project we simply draw squares on the scatter plot because it makes the overall logic much easier and squares and circles on a scatter plot does not make that much difference.

5 Dataset

Our project uses a dataset that contains records of participants for each Olympic Game.

The countries are represented by squares. The sizes of squares indicate proportional to the female-participants athlete ratio. The color of the squares means which continents the country belong to. The number of y-axis represent the number of athletes who won medals and the x-axis represent the total number of athletes that participate in an Olympic game.

To transform the dataset for our project, we aggregate all records by country and year. Because there are so many countries in Olympic Games in recent years, we only select the top fifty for our visualization tool. Another reason doing so is that countries with minimal achievements in Olympic Games always stack together in the scatter plot and creates difficulties for generating the bubble set contour. We also don't want to deal with cases that a country is split into two countries or a territory becomes independent and is counted as separate country in another game. The pool of the top fifty countries is fairly stable and consistent and therefore is the data we use for this visualization.

6 Case study and Analysis Results

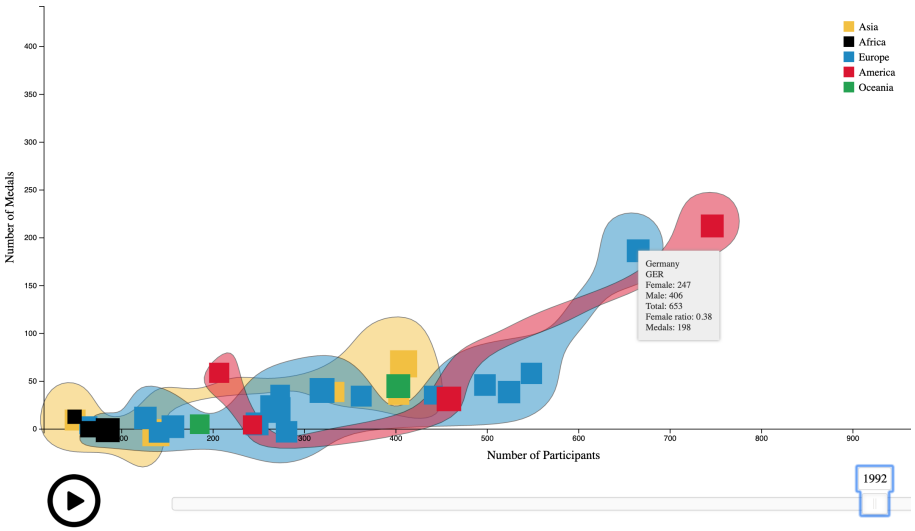

Figure 3: Hovering on the square to display information of the country.

By using our system, we could analyze the dataset more effectively and we have find some results over our dataset.

Firstly, as time goes on, there are more and more females participate in Olympic games.

Secondly, USA is the strongest country in the Olympic games. USA won the most number of medals at most time. Besides, China is the fastest growing country in the Olympic games.

Thirdly, Europe is the strongest continent in the Olympic games. Not only there are many European countries participate in Olympic games, but also the performance of European countries is the best. Meanwhile, Africa is the weakest continent in the Olympic games. There are only a few Africa countries have the ability to win medals.

7 Strength and Limitations

The bubble set contour is very powerful in showing spatial information about data points because it shows where a category of data points span in the global space and where they don't. Traditional ways of showing this is using either color or shapes but they are small channels that associated with each data point and users need to imagine such a contour by themselves by looking at these special disconnected points. The bubble set contour explicitly draws one big shape on the graph indicating the category of data which is more straightforward for users and skip the imagination step.

However, during the implementation, we also find some weakness of the bubble set contour.

First, We find that the algorithm is extremely slow if there are many data points close to each other. The contour tries to find a virtual edge to connect two data points but fails to find a relatively easy path. If blocked by multiple nodes from other categories, the contour tries to find a non-overlapping path and the calculation costs too much computing power of the browser.

Second, we only translate positions of bubble set every frame during playing our system. However, when it comes to squares, we use redrew method to implement the translation effect. As a result, we can only handle hover process during system playing and cannot handle click process until we stop our system.

8 Future Works

There are some improvements we can do in the future.

First, we can use a better dataset. In the paper version, the dataset is the fertility rate by life expectancy by country. If a country has more advanced technology and tools, or if a country puts more attention on fertility, the country will look better in the data set. So, the dispersal of dataset in the paper is better than ours. If we choose a better dataset, we could get a better result.

Second, we can use transitions to handle squares every frame when playing the animation. As a result, we can handle click and hover of squares when during the animation.

9 Acknowledge

Our bubble set implement code has referenced to krause's static bubble set code, which is on https://github.com/JosuaKrause/bubblesets-js.