Abstract

From the first modern Olympics in 1896 to Rio 2016, the Summer Olympic Games has been through 120 years. This study analyzed the historical data about the Olympics and tried to answer four questions, i.e., how female participation changed over the years, whether there is a home-field advantage at the Olympics, what is the medal efficiency for each participating country and region, and which sports have the highest number of participants. Results showed that both total number of athletes and the rate of female participation have been increasing in the past 120 years. Also, there seems to exist a home-field advantage. Third, medal efficiency is highly correlated with participating country or region's economic development. Finally, athletics, gymnastics, and swimming have the highest number of athletes.

Contents

Visualizing 120 years of Summer Olympic Games

Chris Fang, & Hongtao Hao

(Equal contribution)

December 2019

1   Introduction

The modern Olympic Games are the greatest sports mega-event in the world (Grix, 2013). It generates a massive number of audiences worldwide. The latest Rio 2016 Olympics, for example, attracted 3.5 billion viewers (Roxbourough, 2016), half of the world population. Apart from this mobilizing power, the Olympic Games have significant political (Giulianotii, 2015; Cottrell & Nelson, 2010; MacAloon, 1995), economic (Madden, 2002; Blake, 2005; Osada, Ojima, Kurachi, Miura, & Kawamoto, 2016), and socio-cultural (Malfas, Theodoraki, & Houlihan, 2004) impacts. It also contributes to global friendship and cooperation (Beutler, 2008).

Gender equality in sports has been a hot topic (Mervosh & Caron, 2019). Debates on this issue have revolved around equal pay and media coverage (Baker, Seymour, & Zimbalist, 2019). There has long been an obvious gender pay gap in sports (Abrams, 2019; Farmer, 2017). For example, of the 100 athletes on the 2019 Forbes list of the world's highest-paid athletes, only one of them is a woman. i.e., Serena Williams, who is ranked 63th on the list. There has also been a lack of media coverage on women in sports. Although 40% of all sports participants are female, only 4% of the sports media coverage were about female athletes (MacKenzie, 2019). A study (Cooky, Messner, & Musto, 2015) examining sports

coverage from 1989 to 2014 found that despite a dramatic increase in the number of women playing sports (Good, 2015), there had been a decrease in the amount of coverage of female athletes. Under this backdrop, we aim to look at female participation across the globe in the Summer Olympic Games of the past 120 years.

Recently, there has been a trend that fewer cities want to host the Olympics, challenging the future of a century-old tradition (Goldblatt, 2016). For example, 12 cities bid for 2004 Summer Olympics but only two for the 2020 Winter games (Ludacer, 2018). Considering the importance of Olympic success in national pride (Mower, 2012), a home-field advantage in the Olympics might encourage countries to bid for hosting the mega-event. Therefore, it is worth investigating whether this advantage exists in the Olympics, and how significant it is.

Although earning medals can boost national pride, countries are not "born equal" in terms of their ability to achieve Olympic success. Studies have demonstrated that population sizes and economic development for a large part determined a country's Olympic performance (Soos, Martinez, & Szabo, 2017; Xun, 2005; Bernard & Busse, 2004). However, we think it's unfair to compare Singapore with the United States in terms of total number of medals earned, because the two countries have totally different population size. What is more important is "medal efficiency", i.e., medal counts per athlete participating in the Olympics. In other words, we are more concerned about how effective a country is to earn medals, rather than the total number of medals earned. We aim to look at how this "medal efficiency" differs between countries.

Apart from its significance for societies at large, Olympics is associated with individuals' health as well (Sandercocok, Beedie, & Mann, 2016). People might be inspired by the Olympics and more actively engage in physical activities. Besides its benefits for physical health (Warburton, Nicol, & Bredin, 2006), it has psychological benefits as well (Ghildiyal, 2015). For example, playing sports can help build one's characters. With the importance of sports in mind, we want to examine which sports in the Olympics history have the highest number of participants.

2 Literature Review

2.1 Female Participation

Many attempts have already been made to visualize female participation in the Olympics. The most basic one (Bmallion, 2015) is using tables to display the information about countries with the highest and lowest percentages of female athletes, and to show female percentages across the 120 years of the Olympics. See Table 1.

This method is simple and informative. This is useful for researchers looking for information on this topic, but it is not effective in terms of visualization. For example, looking at the plain number won't help viewers have a clear idea of the trend in female participation in the past 120 years.

NOC	TOTAL	MEN	WOMEN	FEM%%%
East Timor	5	2	3	60.0%
Bhutan	19	8	11	57.9%
Saint Kitts and Nevis	17	8	9	52.9%
China	2076	1015	1061	51.1%
Palau	18	9	9	50.0%
DPR Korea (North)	325	172	153	47.1%
Saint Lucia	17	9	8	47.1%
São Tomé and Principe	11	6	5	45.5%
Belarus	537	295	242	45.1%
Angola	148	82	66	44.6%

(a) Countries with the highest rate of female participation

EVENTS	MEN	WOMEN	MIXED	TOTAL	FEM%%%	FEMELIGIBLE
1896	43	0	0	43	0.0%	0.0%
1900	71	2	22	95	2.1%	25.3%
1904	92	3	0	95	3.2%	3.2%
1906	72	1	1	74	1.4%	2.7%
1908	96	3	7	106	2.8%	9.4%
1912	91	5	6	102	4.9%	10.8%
1920	130	7	15	152	4.6%	14.5%
1924	112	10	4	126	7.9%	11.1%
1928	92	14	3	109	12.8%	15.6%
1932	99	14	4	117	12.0%	15.4%
1936	110	15	4	129	11.6%	14.7%
1948	112	19	5	136	14.0%	17.6%
1952	117	25	7	149	16.8%	21.5%
1956	116	26	9	151	17.2%	23.2%
1960	113	29	8	150	19.3%	24.7%
1964	119	33	11	163	20.2%	27.0%
1968	115	39	18	172	22.7%	33.1%
1972	132	43	20	195	22.1%	32.3%
1976	130	49	19	198	24.7%	34.3%
1980	134	50	19	203	24.6%	34.0%
1984	144	62	15	221	28.1%	34.8%
1988	151	72	14	237	30.4%	36.3%

(b) Female percentage in Olympic events

Table 1: Tables provided by Bmalliion (2015)

Another simple but effective method is using a line graph displaying the number of female and male athletes in each Olympic Games from 1896 to 2016 (Nunes, 2019). See Figure 1.
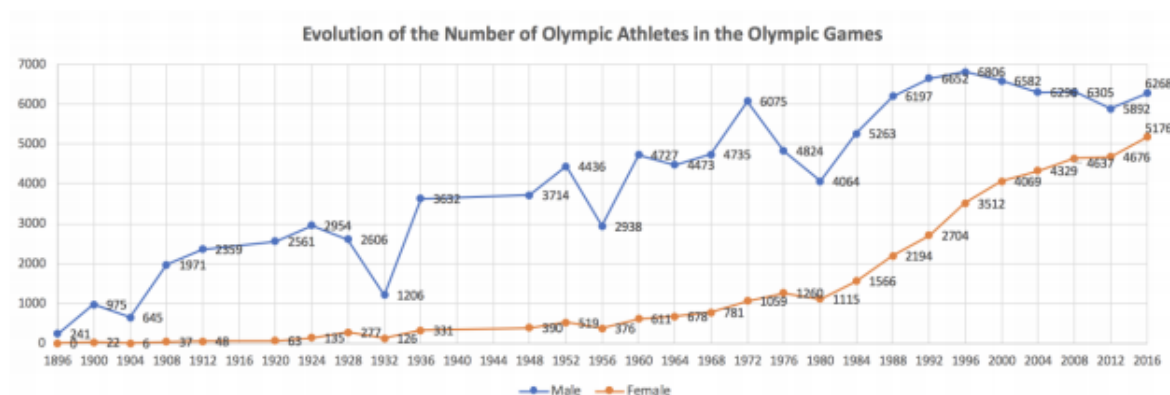
Figure 1: Evolution of the number of Olympic athletes, male and female (1986-2016), Nunes (2019)

This graph is successful in showing the growth in the number of female athletes. However, it fails to show the percentage change. Although viewers can see directly the increase in the total number, they might find it difficult to notice the changes in the percentage.

Washington Street Journal did an interactive visualization comparing the inclusion of female and male events in the Olympics history (Serkez, 2018). It clearly shows that most men's events were already established before the 1960s, whereas most women's events were only introduced after the 1980s. See Figure 2.
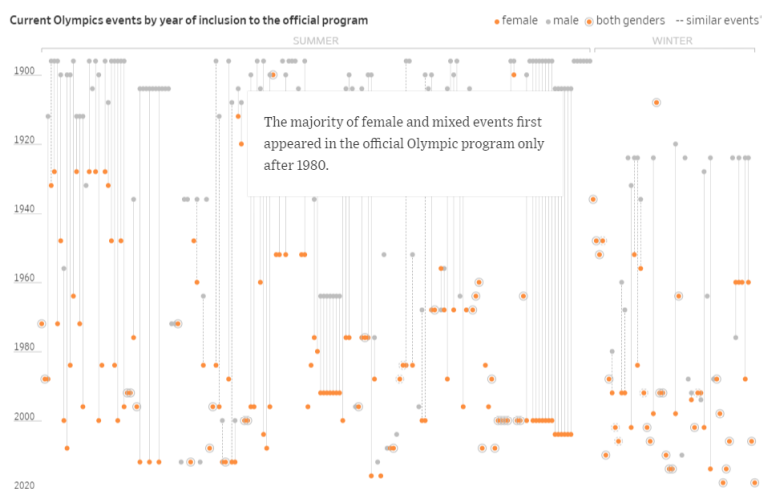

Figure 2: Interactive plot by *Washington Post* showing inclusion of male and female Olympic events

Each individual dot represents a sport of a gender. When clicked, there will be a line connecting to the same sport of a different gender, and information on the inclusion of this sport for both genders will appear. Overall, this visualization clearly shows how late the introduction of female events was in the Olympics history, but it did not show how the number and percentage of female athletes changed over the years.

The official website of the International Olympic Committee (IOC) also made an attempt to show the growth in the percentage of sportswomen in the Olympics. See Figure 3.

As can be seen, it is effective in the way that it shows the dramatic changes in female participation by showing four Olympics in each of the charts. However, its drawback is obvious: the number of Olympic Games that are able to be displayed is extremely limited. It is impossible to show the information of all the Olympics in the past 120 years using this method.

Although there were many attempts visualizing the evolution of female participation in general, few visualizations exist showing the percentage change by continent. The only piece we found is several tables showing men and women participation in the Olympics from 1996 to 2016. No visualizations on this information were made. We decided that there is a gap here. Although both the total number and relative percentage of female athletes have been growing (Nunes, 2019), the increase might be different between continents. We decided to take a closer look at this in our study.
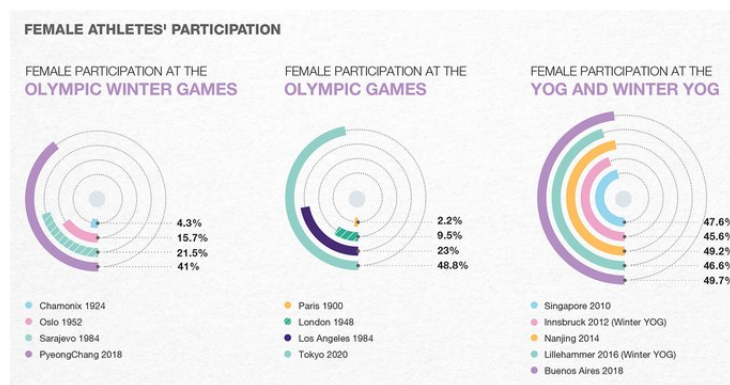

Figure 3: IOC graph visualizing female athletes' participation

2.2   Home-field Advantage

Similar to the attempt at showing female participation, simple and basic tables were utilized to examine the existence of home-field advantage at the Olympics (Pettigrew & Reiche, 2016). See Figure 4.

This visualization shows a table comparing the total number of medals earned at the host year and the previous Olympics. Changes between the two were shown in the last column. Among the 16 host countries from 1952 to 2012, only two countries showed negative changes, meaning that hosting the Olympics helped the country earn more medals. However, the problem with this method is that it is a little bit arbitrary to compare the host year and the adjacently previous Olympics. Changes might have been positive simply because these countries did not perform well only in the previous Games, even though they had performed well eight or more years before the host year. Therefore, it is a more robust choice to show these countries' Olympic performances in all years.

Clarke (2000) did this by calculating the "Home: Away Ratio". See Figure 5.

He listed the percentage of all available medals won by countries that have ever hosted an Olympic both at the Home years and the Away years. The ratio of "Home: Away" was calculated. Obviously, a ratio larger than 1 indicates the existence of home-field advantage

		MEDALS WON BY HOST COUNTRY		
HOST COUNTRY	YEAR	PREVIOUS OLYMPICS	HOST YEAR	CHANGE
Finland	1952	24	22	-2
Australia	1956	11	35	+24
Italy	1960	25	36	+11
Japan	1964	18	29	+11
Mexico	1968	1	9	+8
West Germany	1972	26	40	+14
Canada	1976	5	11	+6
Soviet Union	1980	125	195	+70
United States	1984	94	174	+80
South Korea	1988	19	33	+14
Spain	1992	4	22	+18
United States	1996	108	101	-7
Australia	2000	41	58	+17
Greece	2004	13	16	+3
China	2008	63	100	+37
Great Britain	2012	47	65	+18

Country	Percentage of available medals won		Ratio
	Away	Home	Home:Away
Australia	2.3	7.4	3.2
Belgium	1.3	8.1	6.0
Canada	2.1	1.8	0.9
Finland	3.3	4.8	1.4
France	5.0	23.7	4.8
Great Britain	5.3	25.3	4.8
Germany	6.6	14.8	2.2
Greece	0.4	26.8	62.3
The Netherlands	1.7	5.8	3.4
Italy	4.3	7.8	1.8
Japan	3.1	5.8	1.9
Korea	1.2	4.5	3.6
Mexico	0.4	1.7	3.9
Soviet Union	16.9	30.9	1.8
Spain	0.5	2.7	5.1
Sweden	4.5	21.0	4.7
USA	17.0	37.9	2.2

Figure 4: Medals won by host countries at host year and the previous Olympics, Pettigrew & Reiche (2016)

Figure 5: Percentage of available medals won by host countries at home and away, Clarke (2016)

at the Olympics. The drawback of this method is that it only shows the summary of all the years, lacking information about these countries' performance in each year.

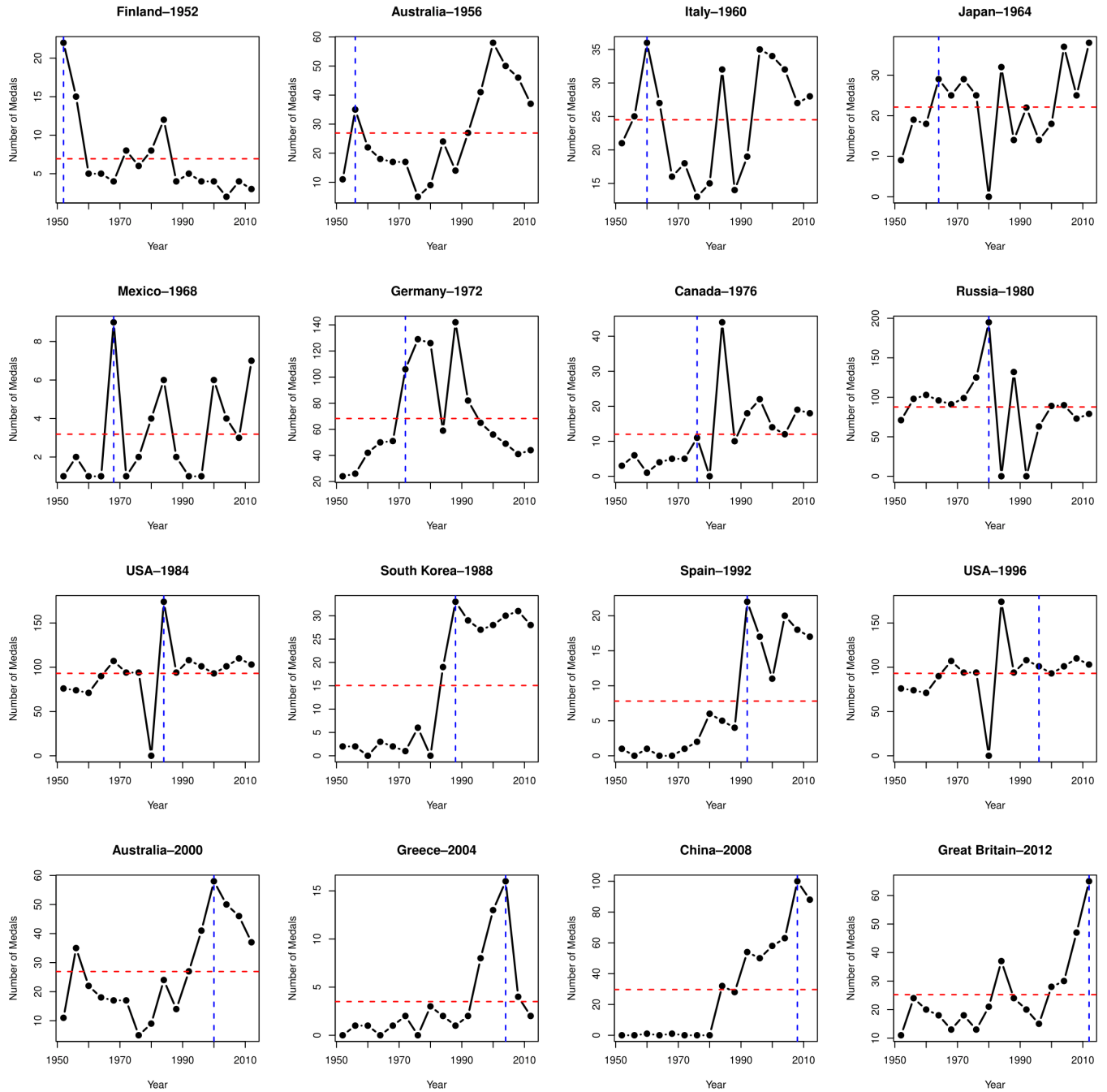Visualizations showing all the years do exist (Grange, 2016). See Figure 6.

Figure 6: Visualizing home-field advantage by Grange (2006)

For example, in this small multiple made by R, each graph has a blue vertical line emphasizing the number of medals earned during the host year, and a red horizontal line indicating the total average. It is very effective in the sense that it clearly compares the performance during the host year and all other years. The problem with this method is that lines are not very good at showing density distribution of medal numbers over the years. To improve this drawback, we decided to use kernel density estimation (KDE) coupled with small multiples.

2.3 Medal Efficiency

The most straightforward way of measuring a country's efficiency of earning medals is to count the number of medals per athlete participating. This method was used by Pettigrew and Reiche (2016) in their analysis of home-field advantage.

Some scholars argued that when assessing the efficiency of medal production, we need to take into account the resources that countries possess, such as GDP and population. From an economics perspective, Rathke and Woitek (2007) came up with a sophisticated formula calculating this efficiency, and visualized the results using multiple box plots that were stacked horizontally. See Figure 7.
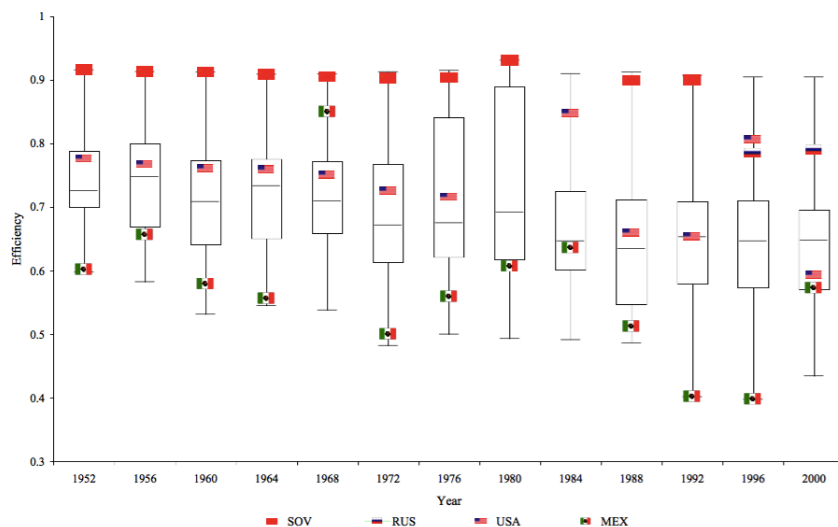

Figure 7: The production of Olympic Medals, 1952-2000, Rathke & Wotiek (2007)
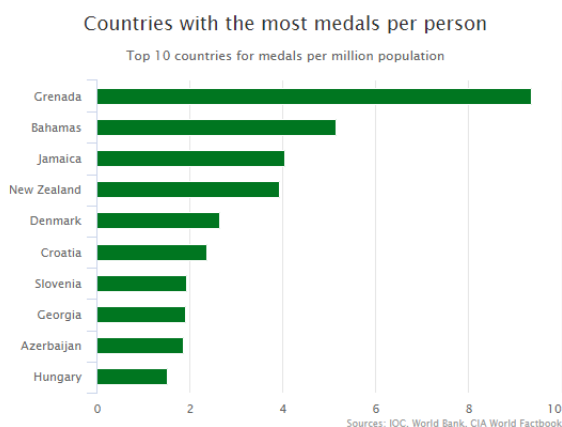
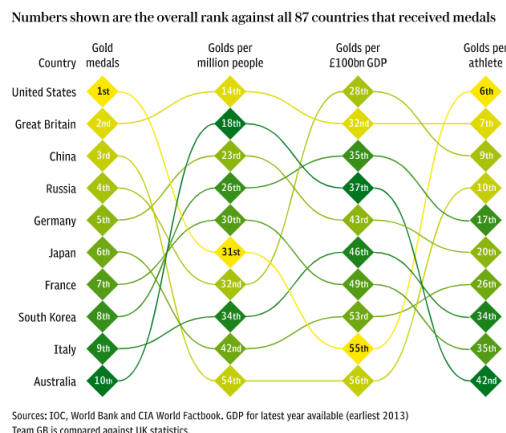Figure 8: Top 10 countries for medals per million population, Eirk (2006)

Figure 9: Ranking medal production considering population and GDP, Eirk (2006)

Visualizations by Eirk (2016) for the *Telegraph*, and those by medalspercapita.com also highlighted the importance of GDP and the size of population. For example, Eirk (2016) ranked Olympic nations based on "Gold per million people" and "Gold per £100 bn GDP". See Figure 8 and 9.

Similarly, using the concept of "medals per capita", medalspercapita.com ranked countries based on "population per medal", i.e., how many people are needed to generate a medal. They created a choropleth map based on the results where darker colors indicate lower "medals per capita". See Figure 10 and 11.
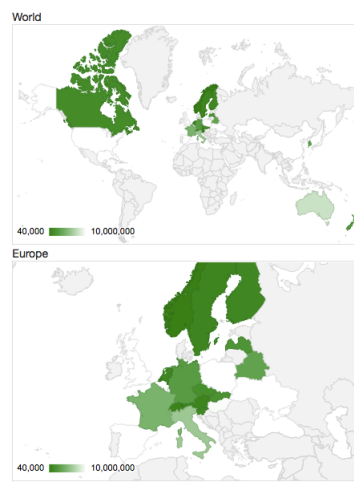

Figure 10: Population per medal

Figure 11: World map of population per medal

However, we think it is problematic to divide medal counts by population or GDP in measuring medal efficiency. First of all, this method is unfair for countries who joined the Olympics late. For example, China (PRC) attended the Olympics for the first time in 1952 and due to political issues, it was not a participant till 1984 (COC, 2004). If the total number of medals is the basis for medal efficiency, it would be unfair for countries like China and many newly founded countries in Africa.

Second, it might be unfair to divide the number of medals by population. This formula assumes that a country can send as many athletes as it wants, but this is not the case. For example, for Tokyo 2020, each country can send at most three male and three female athletes (International Table Tennis Federation, 2018). As a result, China will send six table tennis athletes, as will Australia, Japan, Brazil, Egypt, Germany, and the United States (Wikipedia, n.d.), even though China's population size is larger than that of all the six countries combined (World Bank, 2018).

Third, it might be simplistic to only count the total number of medals. There has been a debate on whether the ranking of Olympic countries should be based on the number of total medals or only golds (Johnson, 2008). We decided to adopt the method by New York Times, i.e., "medal points" (Klein, 2008). In this system, a gold medal is given 4 points, silver 2 and bronze 1. Center for Strategic International Studies (n.d.) used this method and ranked countries based on what they called as "weighted medals".

In this interactive, viewers can see a country's ranking on each category of sport in a specific Olympic Games by clicking the country's name. This visualization is clear and direct, but it has some drawbacks. First, only the top countries can be shown interactively but there are over 200 participating countries and regions. Second, and because of the first drawback, this visualization did not provide viewers a general idea of the distribution of media efficiency across countries. See Figure 12.

To solve these problems, we decided to rank countries by "medal points" as well, but on an interactive choropleth world map.
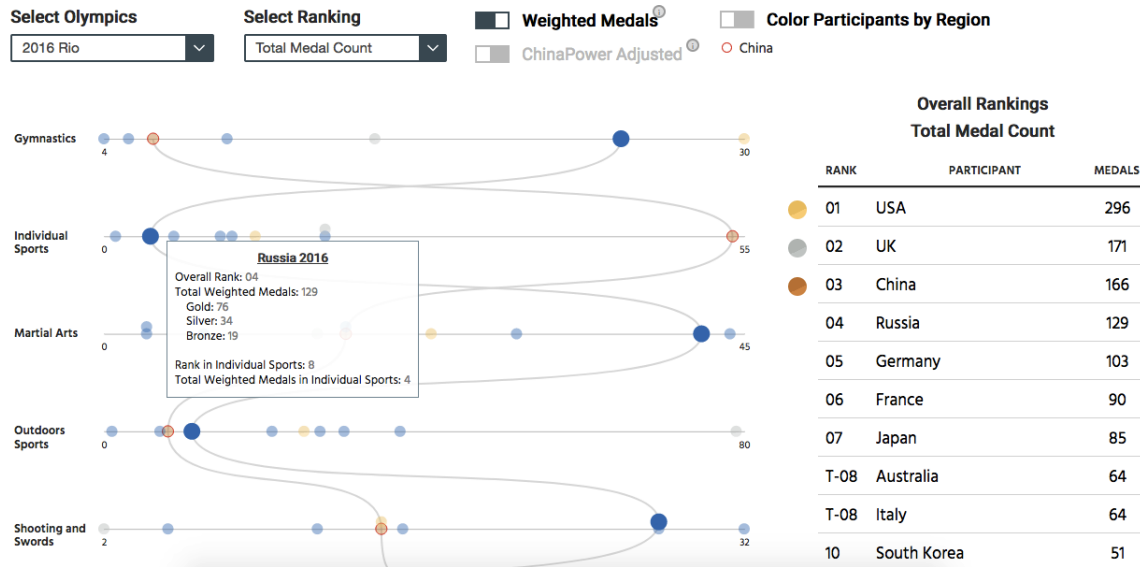
Figure 12: Ranking by weighted medals

2.4 Ranking Sports by Number of Athletes

Dutta (2018) in his Kaggle post visualized the top 10 sports that USA excel by putting 10 boxes vertically with sports winning the most gold medals on top. See Figure 13.

This visualization is transparent and clear but the problem is that it is limited to only about 10 items, which is not enough for us to visualize all the sports in the Summer Olympics. Instead, we decided to try a bar chart and word cloud for our visualizations.
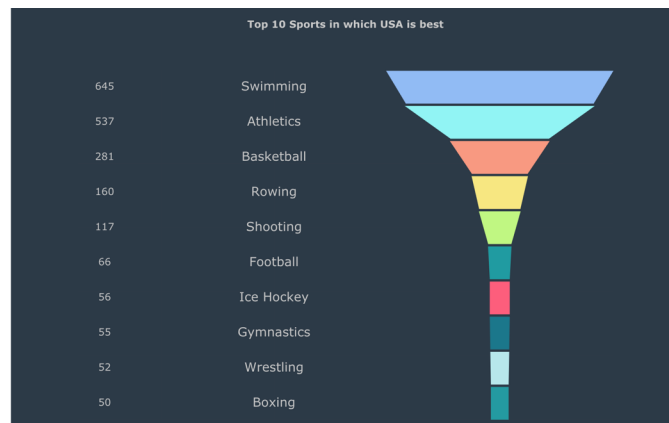

Figure 13: Ranking by weighted medals

3   Unique Contribution

Based on the above discussions of the previous visualizations on the topics we are addressing, our contributions are:

- One of first few attempts to explore female participation in the Summer Olympics over 120 years by continent. This new insight will be important because although data shows that overall female participation in the Olympics has been increasing dramatically, this might have been the case for each continent. Further investigation into the percentage of female athletes in each of the six continents (Asia, Europe, Africa, South America, North America, and Oceania) will reveal deeper information about the trend of female athletes participation.

- Introducing KDE in examining home-field advantage in the Olympics. This method makes it possible to show the density of each host country's performance distribution, which makes the comparison between "home" and "away" performance clearer.

- Introduction of the concept of "medal efficiency". Earlier attempts either lacked taking into account of "medal points" or unfairly included GDP and population in their formula. Our way of calculating medal efficiency, i.e., weighted medal points per athlete participating, can more accurately estimate each country's efficiency of producing medals in the Summer Olympics.

4   Visualization Plans

4.1   Plans: Female Participation

As discussed above, we will use a line graph, stacked bar chart, and area chart to show the overall trend of female participation in the Summer Olympics. For female participation in each individual continent, we will use KDE in a small multiple.

In a line graph, the x-axis will be time and the y-axis the percentage against the total number of participants. We will show both female and male athletes. See Figure 14.
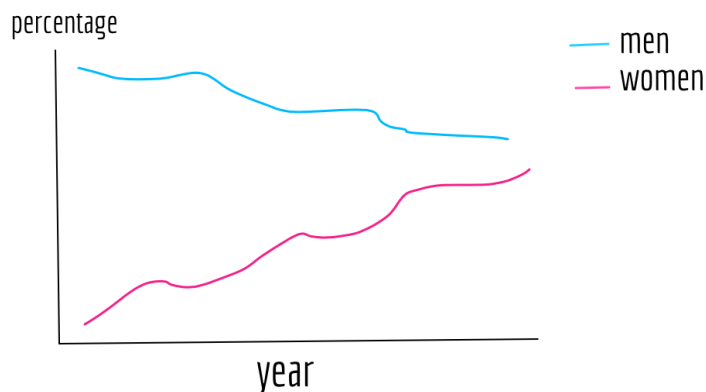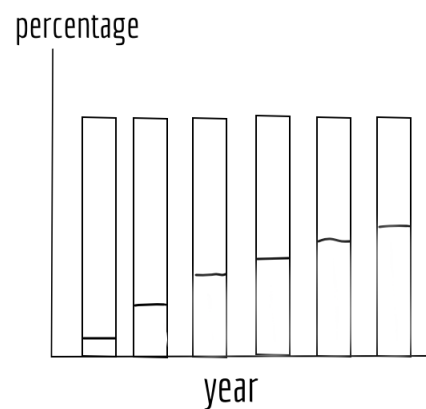
Figure 14: Line graph for female participation

Figure 15: Stacked bar chart for female participation

For stacked bar charts, the x-axis will be time and y-axis percentage against the total number of athletes. Two segments in each bar will represent male and female respectively, and the two will sum up to 100% for every year presented. This will make the comparison between male and female participation very clear. See Figure 15.

In order to also show the changes in the total number of athletes, beside the changes in percentages, we will use a stacked area chart. The x-axis will be again time, and the y-axis will be the total number of athletes for each Olympic Games. See Figure 16.

For female participation in each individual continent over the years, we will line graphs in a
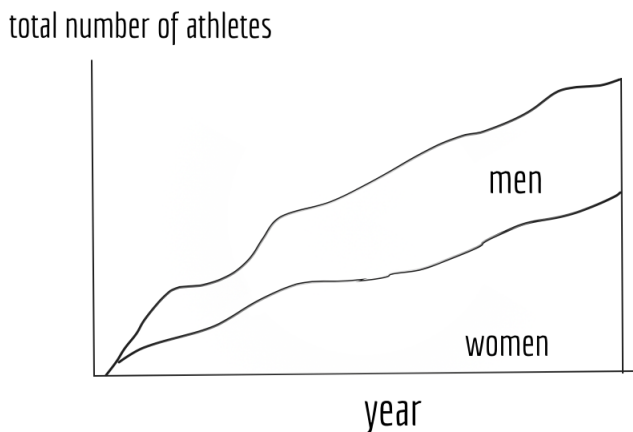
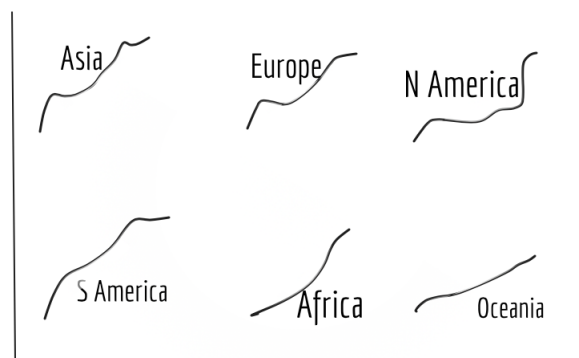Figure 16: staked area chart for female participation

Figure 17: Line graph in small multiple for female participation

small multiple. The x-axis will be time and the y-axis will be the percentage of female
athletes against the total number of athletes. To make comparisons between continents
and with the global average clearer, for each plot in the small multiple, we will put the line
graph of other continents in the background, and we will also plot the global average as the
benchmark. See Figure 17.

4.2   Plans: Home-field Advantage

To examine whether there exists a home-field advantage at the Olympics, we will first use a
scatter plot with jitter. The x-axis will be countries that have ever been a host, and the
y-axis will be the percentage against the total number of medals a country earned. Each
dot represents an Olympic Games the country has participated in. Black dots denote data
for a "non-host" year and orange dots denote data for host year. See Figure 18.

To better show the density distribution of medal gains, we will employ kernel density
estimation in a small multiple. The x-axis will be the percentage against the total number
of medals at a Games and the y-axis will be density. We will use an arrow to denote the
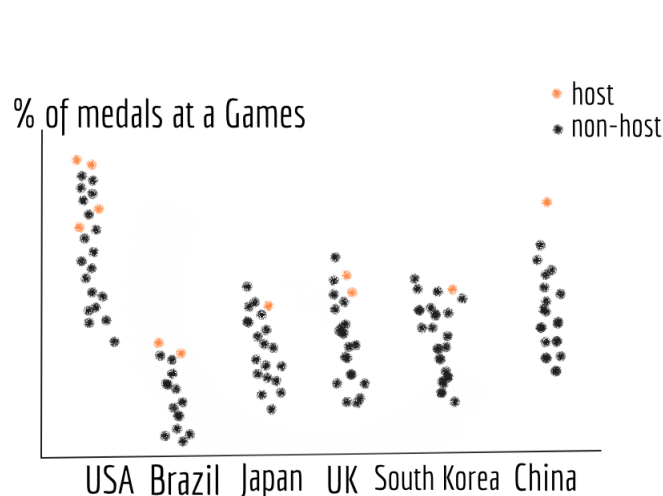density of a year when the country was a host. See Figure 19.

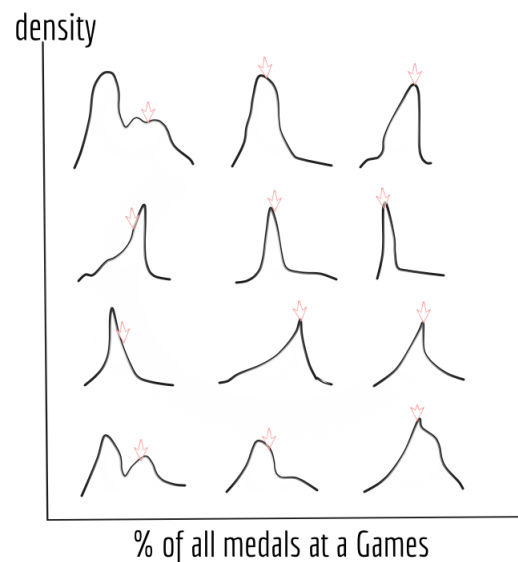Figure 18: Scatter plot with jittering for
home-field advantage

Figure 19: KDE for home-field advantage

Arrows located at points with high densities, would be signs of the existence of home-field advantage at the Olympics.

4.3   Plans: Medal Efficiency

The index of "medal efficiency" will be calculated as medal points per athlete participating. We will use an interactive choropleth world map to show each participating country and region's score. See Figure 20.

4.4   Plans: Ranking Sports by Number of Athletes

First, we will use a bar chart where the x-axis will be the different sports and the y-axis the corresponding number of athletes. See Figure 21.

Second, we will show this ranking using a word cloud in which larger font size denotes a higher number of participants. This graph will be clearer and more direct because viewers can intuitively get a sense of the relative size of participation of a sport. See Figure 22.
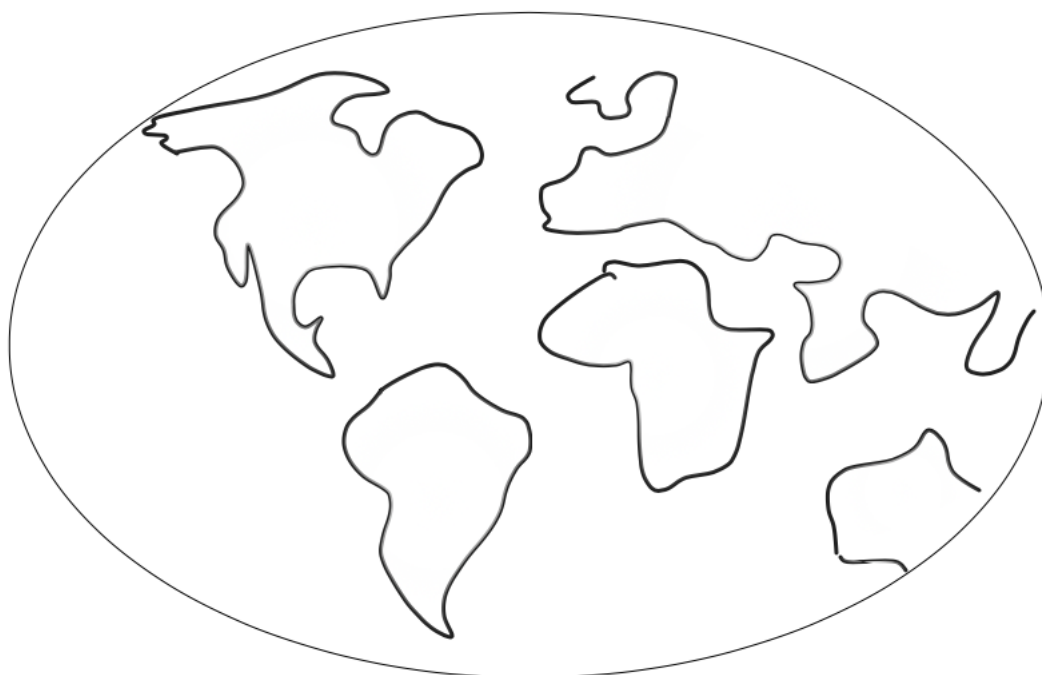

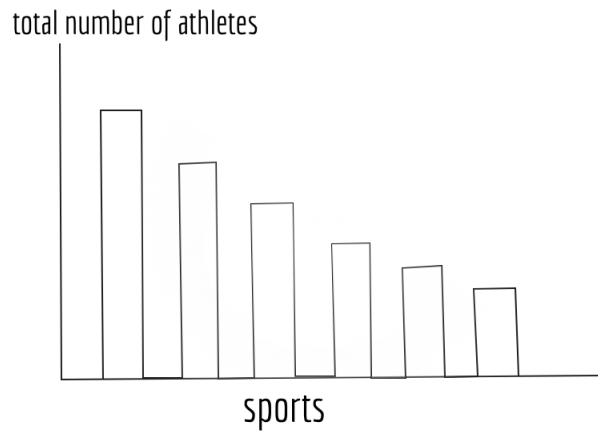Figure 20: Choropleth for medal efficiency

Figure 21: Bar chart for sports ranking

Figure 22: Word cloud for sports ranking

5 Plotting

5.1 Plotting: Female Participation

As planned above, we started with a simple line graph displaying both male and female athletes' percentage against all athletes. The strength of this method is that it is simple and clear. It shows that the female percentage has been steadily growing in the past century at the Olympics. The drawback is that it does not generate a stark contrast between male and female as a stacked bar chart can do. See Figure 23.
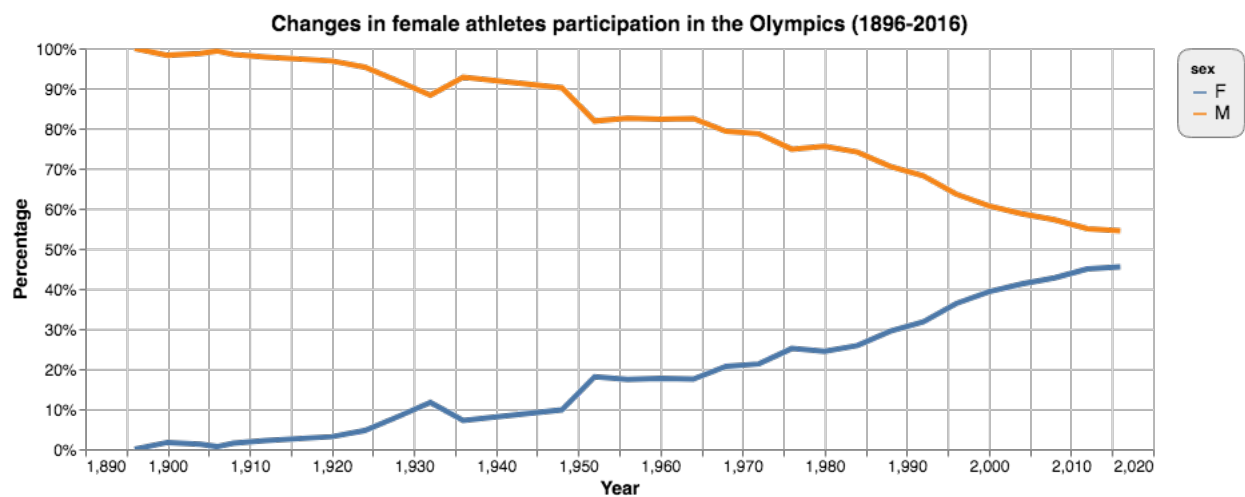

Figure 23: Line graph for female participation
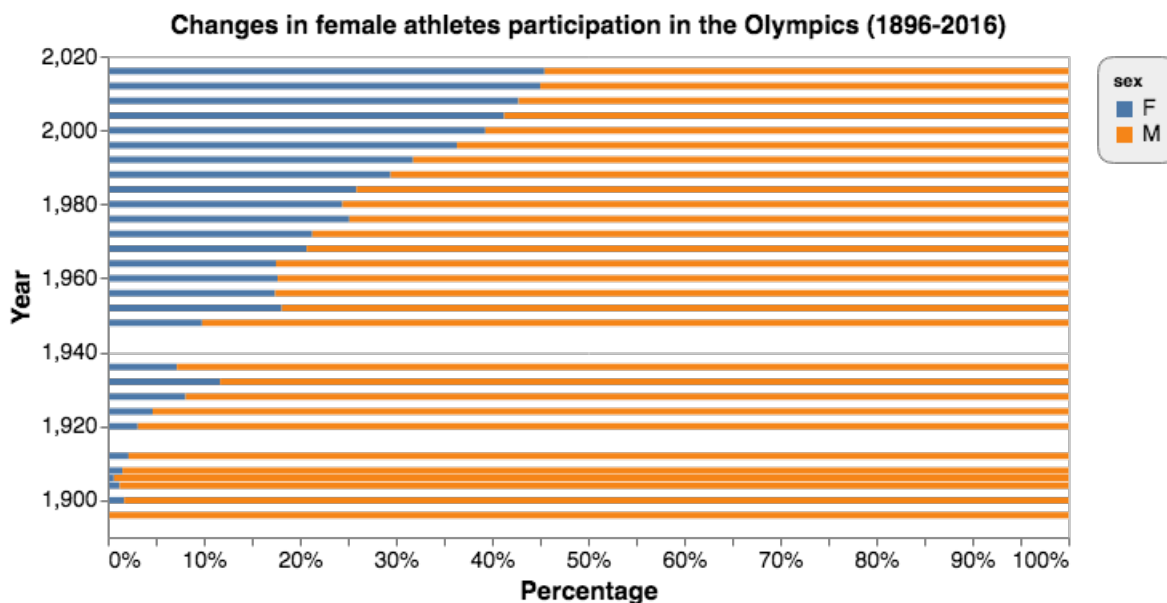
Figure 24: Horizontal stacked bar chart for female participation

We then used stacked bar graphs. We first tried stacking the bar horizontally with percentage on the x-axis and time on y-axis. See Figure 24.

It was not as beautiful as we wanted, so we tried to put it upright and replace the axis labels. To highlight the trend, we set the male section grey and the female part bright blue. We also drew a red line at the intersection of the two parts. This time, female participation was obviously highlighted and the trend shown very clearly. See Figure 25.

Figure 25: Vertical stacked bar chart for female participation

However, as discussed above, stacked bar charts can only show the changes in percentages, not those in the actual number of participan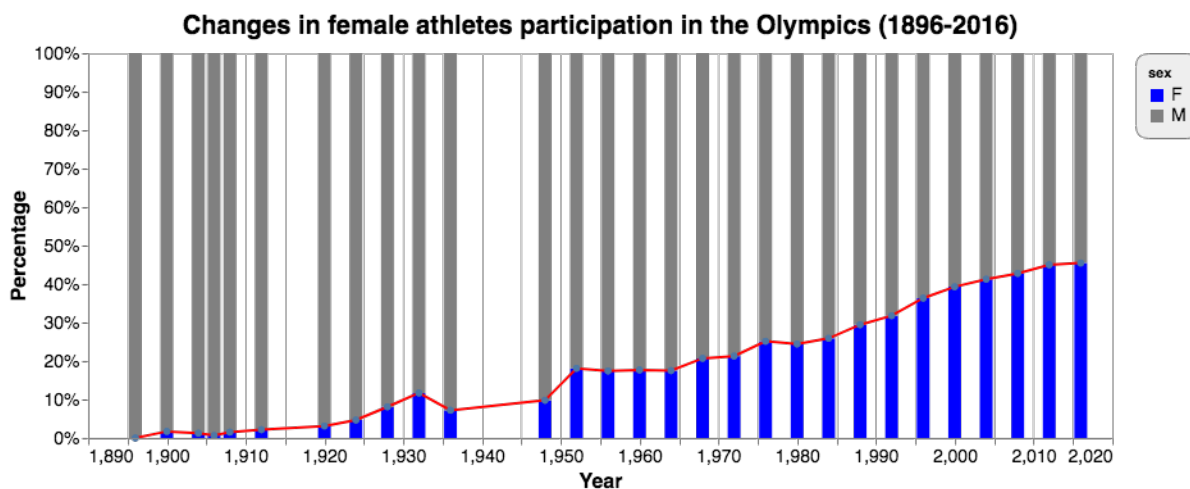ts. To also visualize how the total number of athletes participating in the Olympics changed over the century, we opted for a stacked area chart. See Figure 26.

The x-axis denotes time and the y-axis number of athletes. Areas are made up of male and female participants. The two added up to the total number of athletes. This stacked area chart not only shows the changes in female participation, it also displays how the total number of athletes has been increasing over the years. The drawback is that area is not a much preferred visual encoding. As a result, the comparison of percentages of male and female is not very accurate. That said, we think its strengths outweigh its disadvantages.

To visualize changes in female participation in each continent, we first used a simple line graph where we put all the continent together. The x-axis is time, and the y-axis is female participation. The legend shows the colors associated with the continents. See Figure 27.
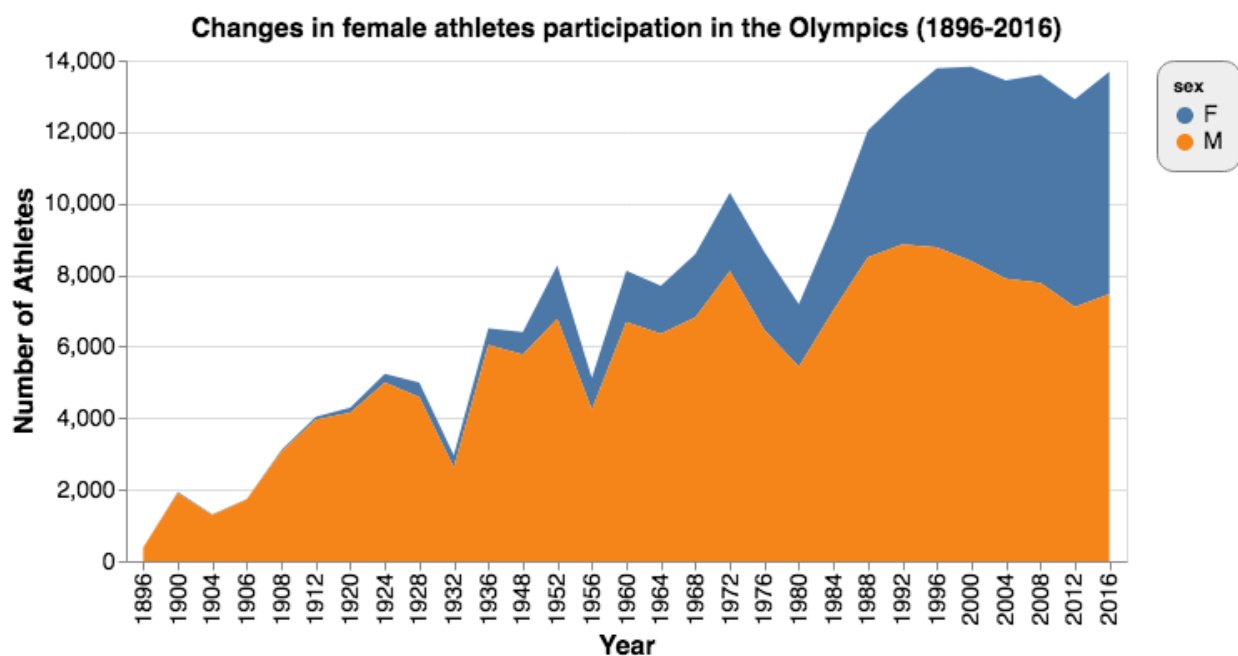

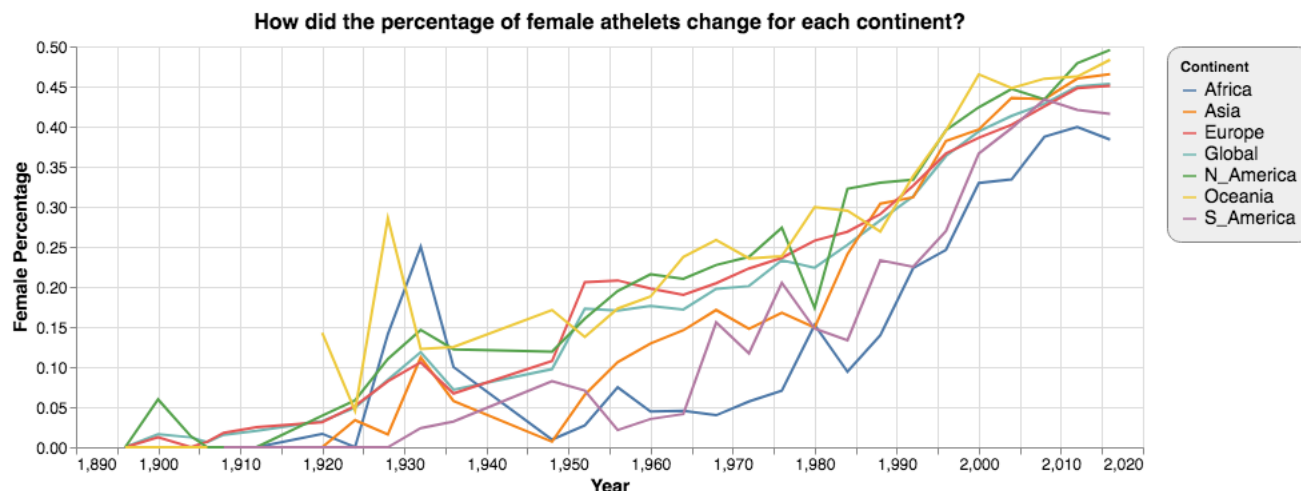Figure 26: Stacked area chart for female participation

Figure 27: Line graph for female participation by continent

The graph was very messy. As there were six continents and one global average, we could not distinguish between the seven categories very easily. So we gave up this option.

Then we tried a small multiple in which each continent and the global average were shown in simple line graphs. This chart showed every continent clearly but it was not very easy to make a comparison. See Figure 28.

To make comparisons clearer, we decided to use an area chart. We stacked each area chart each representing a continent along a vertical line. However, comparisons were still difficult. Most importantly, in terms of using color as the visual encoding, this graph has too many categories. Besides, it has red and green at the same time, and therefore is not accessible to colorblind populations. See Figure 29.

We then went back to the small multiple. Since we cared about the comparison between each continent, we decided to plot every continent and the global average in each of the six graphs. We highlighted only one continent in each graph, and set all other continents grey in the background. It worked much better than out earlier attempts, but one drawback is that it did not allow comparison with the global average.

Finally, we decided to highlight the global average with a black line. This time, comparisons both between continents and with the global average were clear. See Figure 30
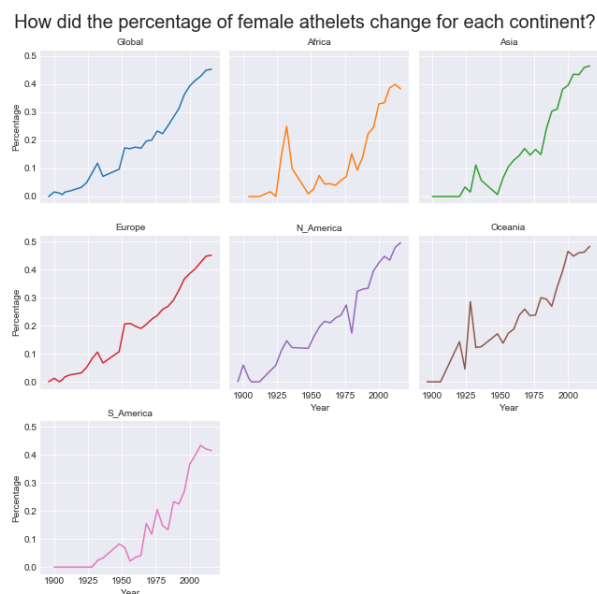
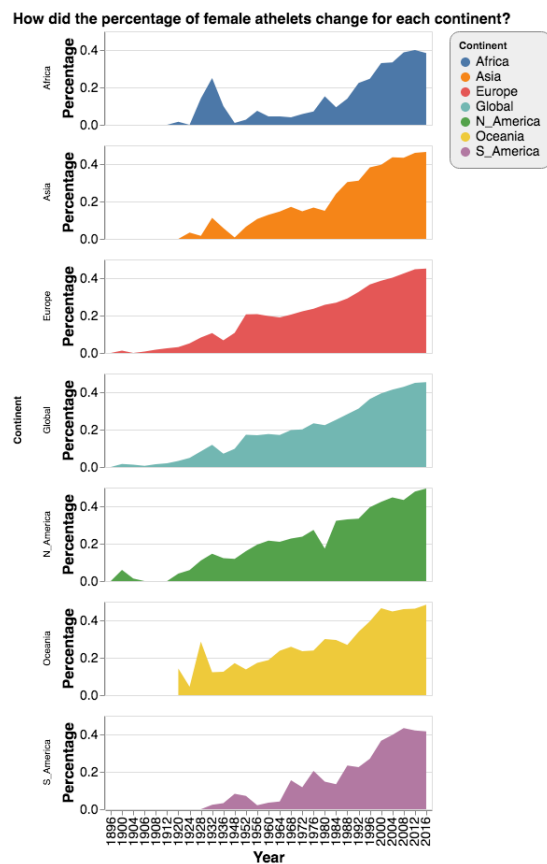Figure 28: Line graph in small multiple for female participation by continent

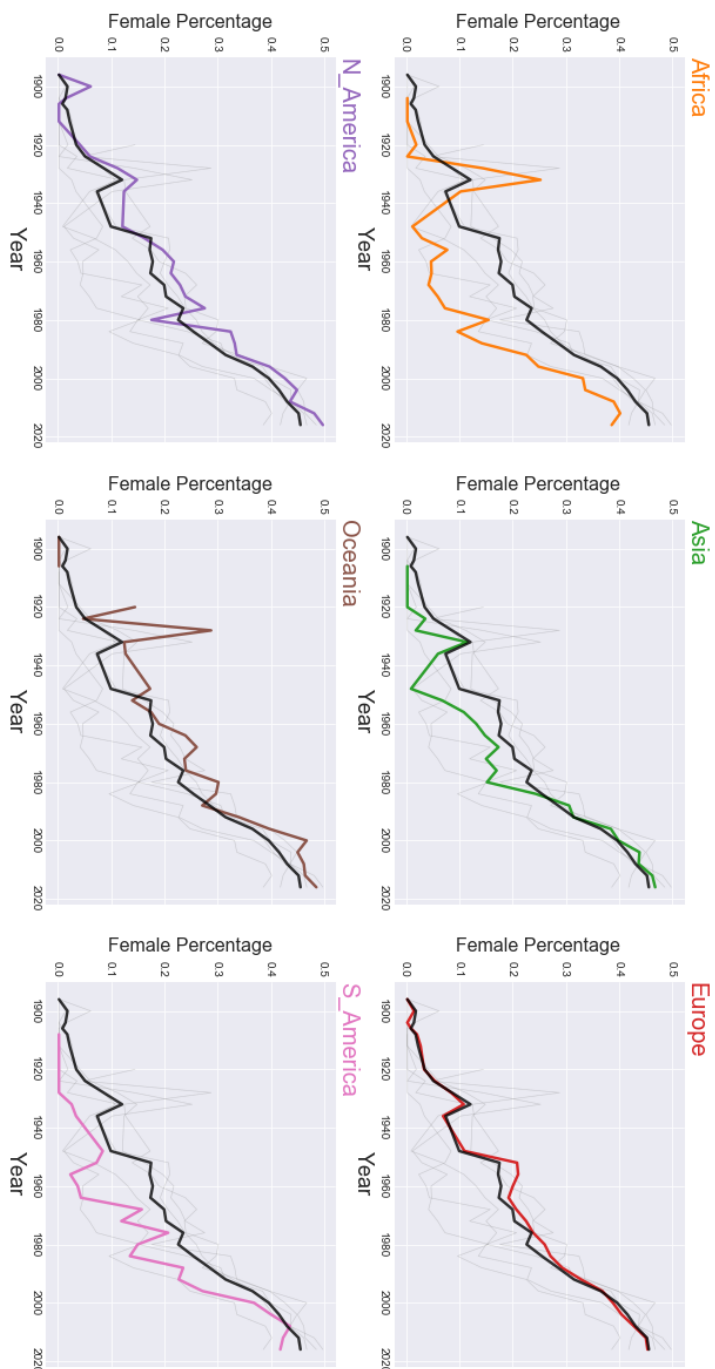Figure 29: Stacked bar chart for female participation by continent

Figure 30: Line graph with reference for female participation by continent

5.2   Plotting: Home-field Advantage

As discussed above, we first tried scatter plot with jitter. The x-axis is the countries that have ever hosted the Olympics and the y-axis denotes the percentage of medals earned by a country against the total number of medals in that year. Blue dots represent data when the country was not the host and the orange ones for when it was the host. See Figure 31.

The blue dots were too packed. We later tried beeswarm plot coupled with box plot. Dots were shown much more clearly but one drawback is that we could not see the density distribution of all the dots very well. Density distribution was important in this case because it would allow an easier comparison between the medal percentage when a country was a host and that when it was not. See Figure 32.

To show density distribution of medal percentages, we used kernel density estimation in a small multiple. The x-axis is the medal percentage and the y-axis is the probability density. We used a red arrow to denote when the country was an Olympics host. An arrow located at the tail would indicate the existence of home-field advantage. See Figure 33.
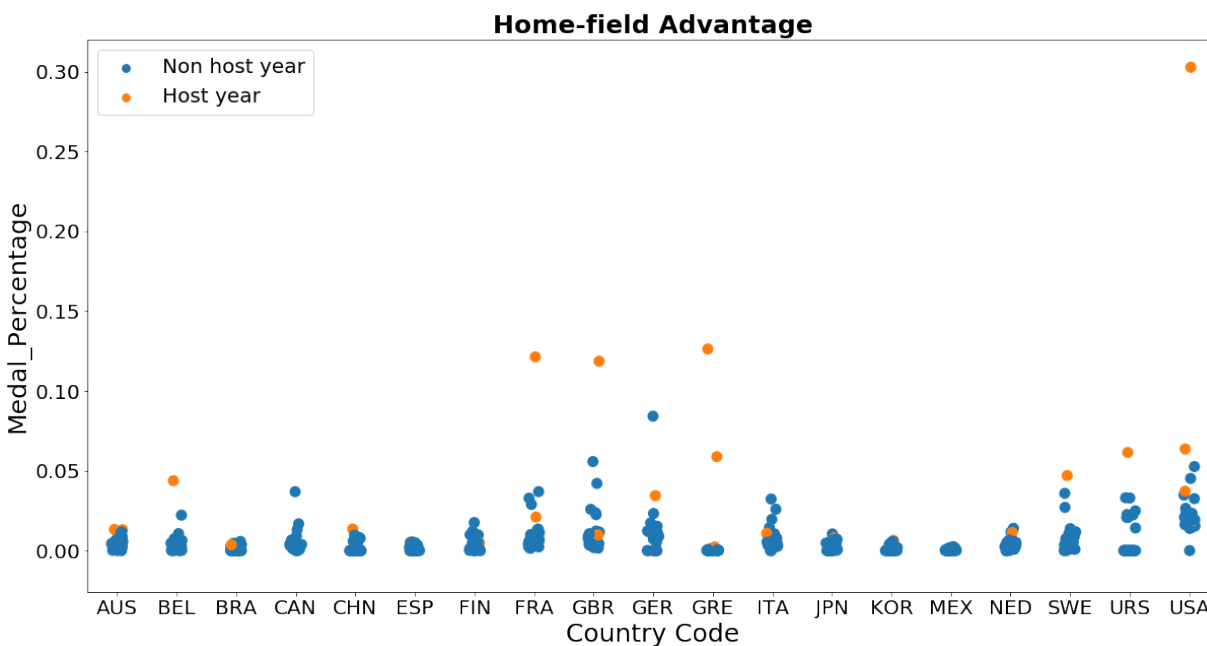

Figure 31: Scatter plot with jittering for home-field advantage
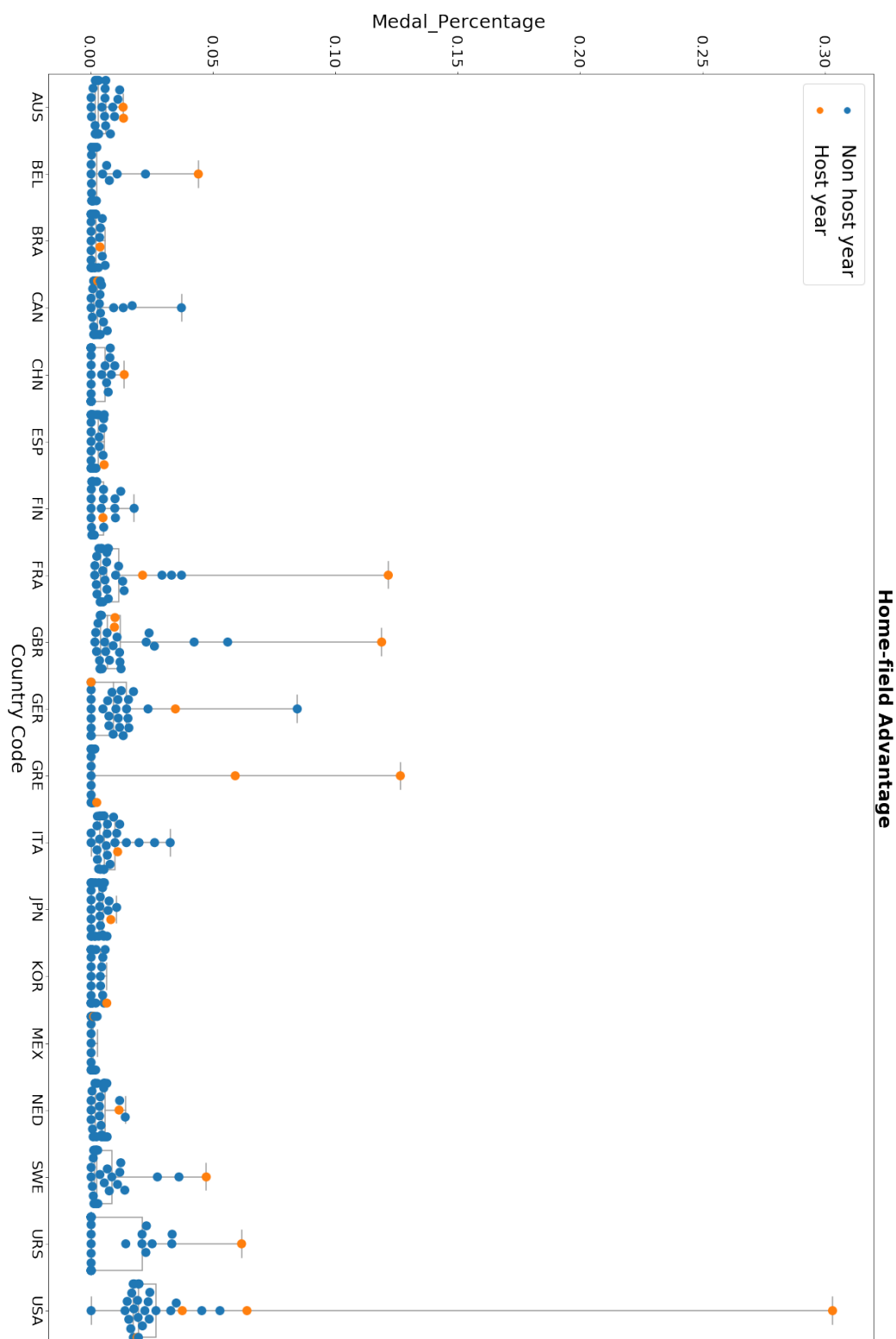
Figure 32: Beeswarm plot for home-field advantage
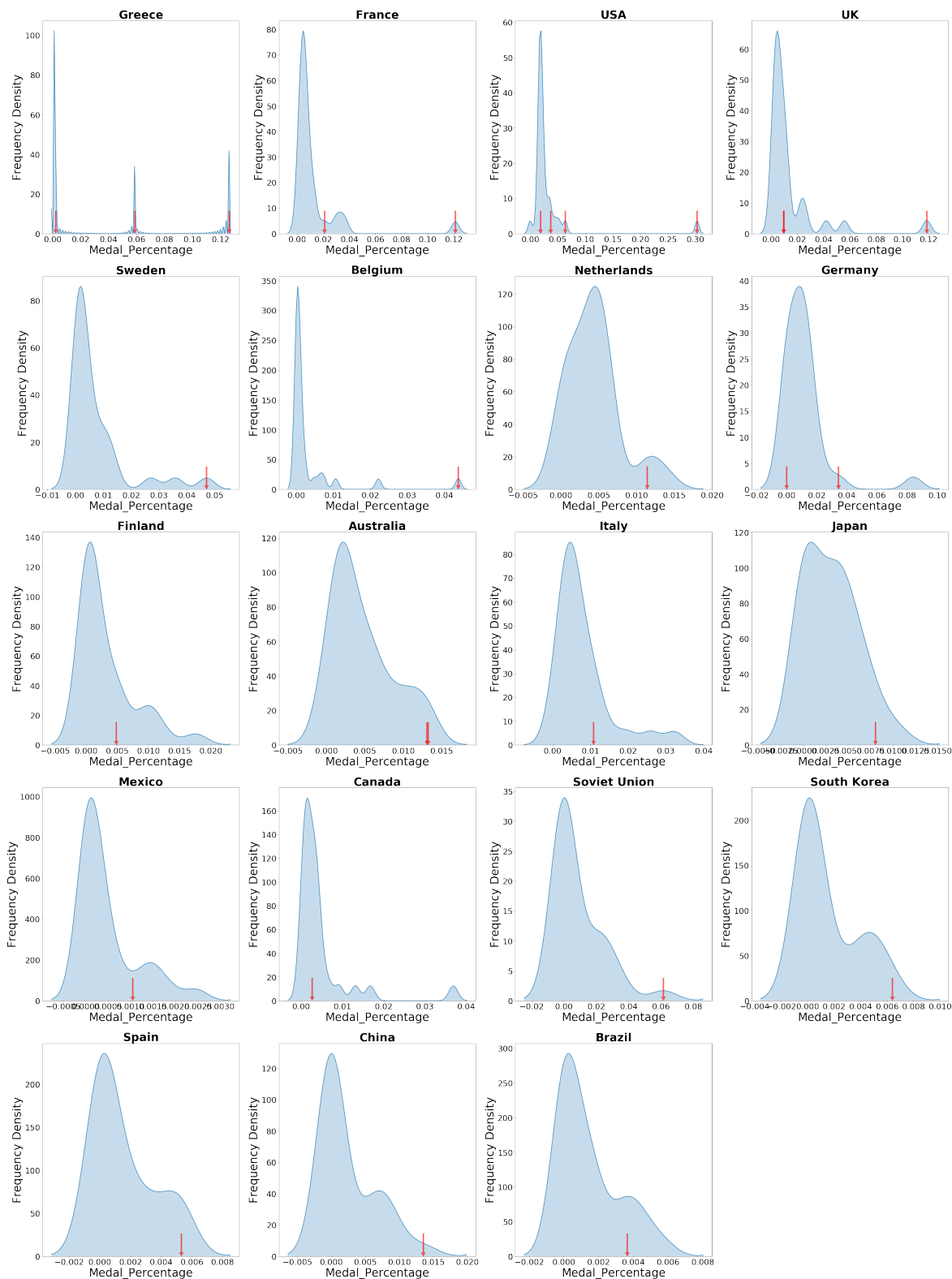
Figure 33: KDE in small multiple for home-field advantage

5.3 Plotting: Medal Efficiency

As planned, first of all, we did data manipulation before plotting. Golds were given four points, Silver 2, and Bronze 1. Attendance without any medals was assigned to 0 point. A country's "medal efficiency" was calculated as the quotient of total medal points and the total number of athletes participating over 120 years.

We then plotted a choropleth map using orthographic projection. We applied Viridis colormap where darker shades denote higher medal efficiency. To ensure interactivity, we made this map with the Plotly package. See Figure 34.

We found this index of medal efficiency to be both more accurate and intuitive. The reasons why it is more accurate have been outlined above. We believe that it is also more intuitive because, for example, for countries that still exist today, the highest medal efficiency is 0.91 for the United States of America. The fact that the highest is close to one makes it easier to compare between countries.
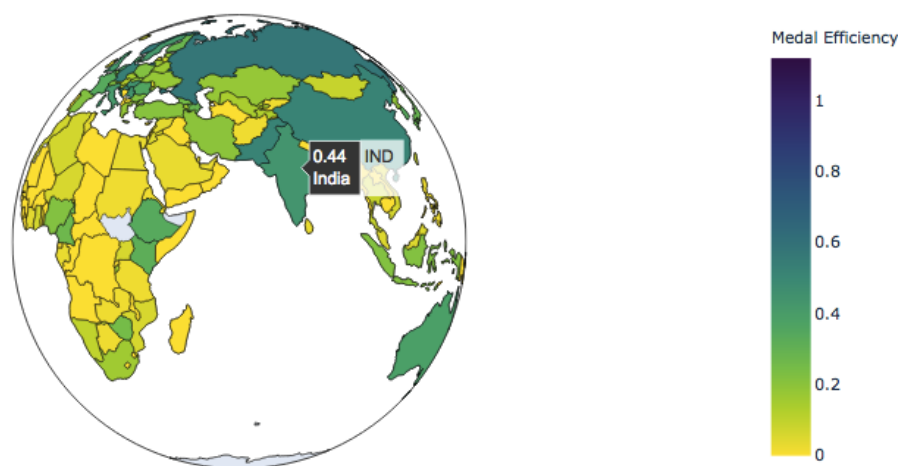
Figure 34: Interactive choropleth for medal efficiency

5.4 Plotting: Ranking Sports by Number of Athletes

To rank sports according to the number of athletes, we first plotted a bar chart, each bar representing a category. As can be seen below, since there were too many categories, the names of sports overlapped to a degree that most of them were indistinguishable. Since there are around thirty sports, using colors is not an ideal option. See Figure 35.
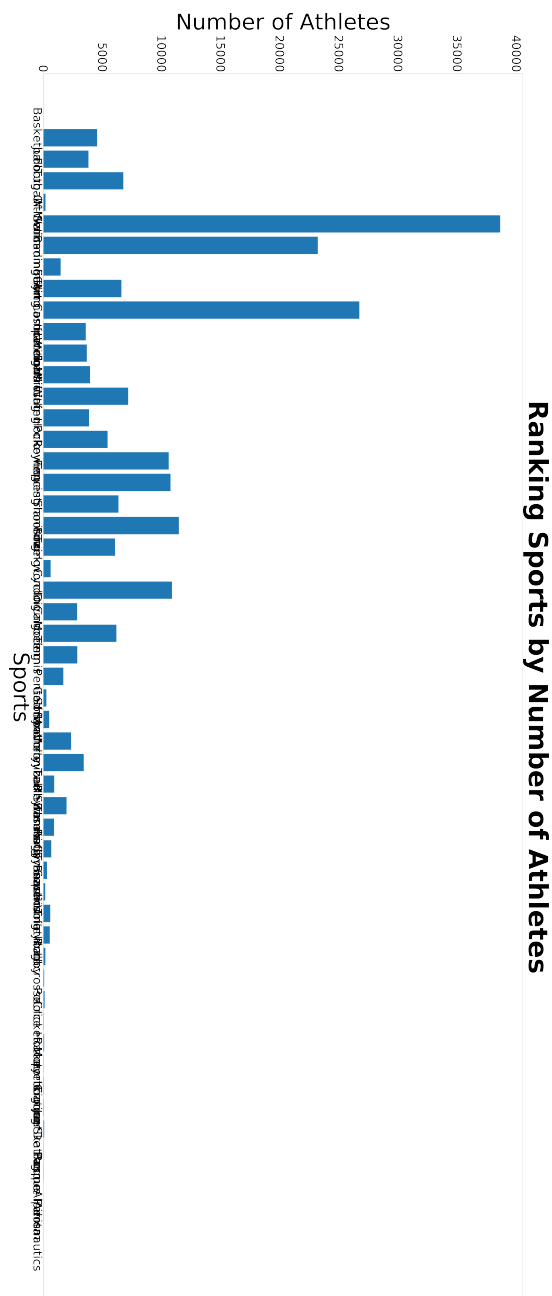
Figure 35: Bar chart for sports ranking
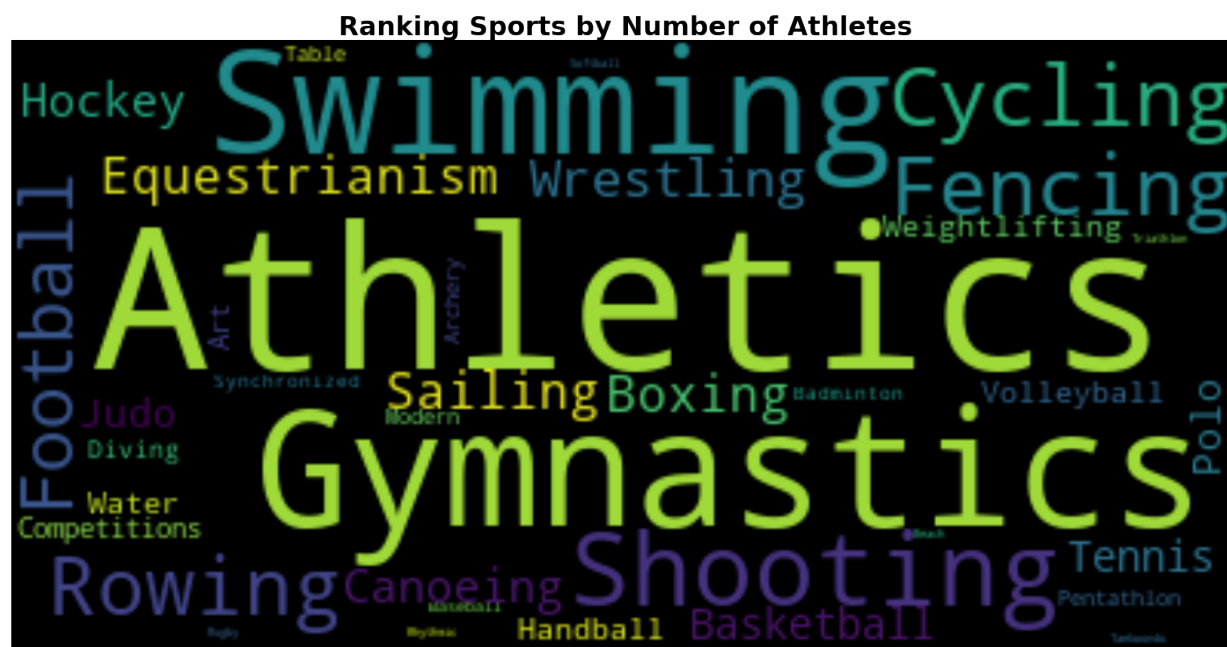
Figure 36: Word cloud for sports ranking

We then thought about the word cloud. In a word cloud, the size of words is associated with the frequencies. This would make our ranking clearer. We did this visualization with the wordcloud package. See Figure 36.

6 Discussion

6.1 Research Question Recap

This paper answered four major questions:

1. How does female participation at the Olympics, both overall and in each continent, change over time?

2. Is there a home-field advantage at the Olympics?

3. What is the medal efficiency for each country and region?

4. Which sports have the highest number of participating athletes at the Olympics?

6.2 Findings

Our visualizations demonstrated that:

1. Both the total number of participating athletes and the percentage of female athletes in the Olympics has been increasing. Female participation started from 0% in Athens 1896 to 45% in Rio 2016.

2. However, this growth in the rate of participation of women varies across continents. The rate in Africa has always been lower than the global average except for during the 1930s. The rate in Asia and South America had also been always lower than the global average but caught up during the 1980s and 2000s, respectively. The rate in Europe, North America and Oceania has always been the same as or higher than the global mean.

3. A country is more likely to have a higher degree of medal share when it hosts the Olympics. In other words, home-field advantage does seem to exist in the Olympics.

4. Among today's countries and regions, the United States of America has the highest medal efficiency, 0.91. This means that on average, every American athlete almost earned a Bronze medal in the past Olympic Games. Medal efficiency for Russia, China, Pakistan, India, Germany, and Australia is also very high. We also found that medal efficiency had a high correlation with countries' economic development.

5. Athletics, gymnastics, swimming, shooting and football are among the sports with the highest number of participating athletes. This might be because these sports have greater shares of total medals. It might also because there are many sub-categories within them.

6.3 Effectiveness and future work

First, when visualizing changes in female participation, we found that stacked bar charts are effective in showing the comparisons between male and females. Stacked area charts do

not provide very accurate percentage change, but it is good if we also want to show how the total number of athletes has been increasing. The seaborn FacetGrid function is useful for visualizing female participation by continent. Reference information set in grey in the background and highlighting the global average helps us make comparisons between continents and with the global average much clearer. Our next step is to do this graph with Altair so as to enable interactivity when we post it online.

In our study into the home-field advantage, we realized that even with jittering, scatter plot is not effective at showing a large number of data points. Beeswarm plot with box plot is good at displaying all the points without overlapping, but its representation of density distribution is not very intuitive and direct. KDE in a small multiple is a much better method for showing the distribution clearly. KDE in this case is especially useful for examining home-field advantage because it shows the level of probability that a country gets a certain share of medals.

Our introduction of "medal efficiency" contributes to the assessment of a country's efficiency at obtaining Olympic medals. The choropleth we used allows us to show all the countries and regions interactively. However, we acknowledged that it would have been better if it were a cartogram. This might be our next step.

Using word cloud is a very unique way to show the relative size of participation in various sports. The drawback is that some sports have two or more words in their names, and our processing of word clouds might not have been rendering their frequencies very accurately. That is why we see "water", "art" and "competition" in the word cloud. We will fix this problem in our next step.

We will publish it online as our portfolio later.

7   References

Abrams, O. (2019, June 23). Why Female Athletes Earn Less Than Men Across Most
    Sports. Retrieved from https://www.forbes.com/sites/oliviaabrams/2019/06/23/why-

female-athletes-earn-less-than-men-across-most-sports/#646f8eeb40fb.

Baker, C. N., Seymour, E., &Zimbalist, A. (2019, April 18). Less Money, Less Media: How
Female Athletes are Undervalued. Retrieved from
https://msmagazine.com/2019/04/18/less-money-less-media-how-female-athletes-are-
undervalued/.

Bian, X. (2005). Predicting Olympic medal counts: The effects of economic development
on Olympic performance. *The park place economist, 13* (1), 37-44.

Beutler, I. (2008). Sport serving development and peace: Achieving the goals of the United
Nations through sport. *Sport in society, 11* (4), 359-369.

Bernard, A. B., & Busse, M. R. (2004). Who wins the Olympic Games: Economic
resources and medal totals. *Review of economics and statistics, 86* (1), 413-417.

COC. (2014, March 27). 10th-15th Olympic Games: 1936-1952. Retrieved from
http://en.olympic.cn/games/summer/2004-03-27/121663.html.

Cooky, C., Messner, M. A., & Musto, M. (2015). "It's dude time!" A quarter century of
excluding women's sports in televised news and highlight shows. *Communication &
Sport, 3* (3), 261-287.

Cottrell, M. P., Nelson, T. (2011). Not just the Games? Power, protest and politics at the
Olympics. *European Journal of International Relations, 17* (4), 729-753.

Darnell, S. C. (2012). Olympism in action, Olympic hosting and the politics of 'sport for
development and peace': Investigating the development discourses of Rio 2016. *Sport in
Society, 15* (6), 869-887.

Dutta, D. (2018). Analysing the Olympics (for last 120 yrs.). Retrieved from
https://www.kaggle.com/duttadebadri/analysing-the-olympics-for-last-120-yrs.

Farmer, I. (2017, December 18). Why are female athletes paid less than male athletes?

Retrieved from https://news.jrn.msu.edu/2017/12/why-are-female-athletes-paid-less-than-male-athletes/.

Forbes (2019). World's Highest Paid Athletes. Retrieved from https://www.forbes.com/athletes/list/#tab:overall.

Ghildiyal, R. (2015). Role of sports in the development of an individual and role of psychology in sports. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4381313/.

Giulianotti, R. (2015). The Beijing 2008 Olympics: Examining the interrelations of China, globalization, and soft power. *European Review, 23* (2), 286-296.

Goldblatt, D. (2016, August 4). Nobody wants to host the Olympic Games anymore. Can you blame them? Retrieved from https://qz.com/748894/nobody-wants-to-host-the-olympic-games-anymore-can-you-blame-them/.

Good, A. (2015, June 5). When it comes to women in sports, TV news tunes out. Retrieved from https://news.usc.edu/82382/when-it-comes-to-women-in-sports-tv-news-tunes-out/.

Grix, J. (2013). Sport politics and the Olympics. *Political studies review, 11* (1), 15-25.

Center for Strategic International Studies (n.d.). How dominant is China at the Olympic Games? Retrieved from https://chinapower.csis.org/dominant-china-olympic-games/.

International Table Tennis Federation. (2018). Tyko 2020 Qualification System Table Tennis (Eng). Retrieved from https://www.ittf.com/wp-content/uploads/2018/05/2018-05-14-Tokyo-2020-Qualification-system-Table-Tennis-eng.pdf

Johnson, I. (2008, August 13). Who's on First in Medals Race? Retrieved from https://www.wsj.com/articles/SB121856271893833843?ns=prod/accounts-wsj.

Klein, J. Z. (2008, August 23). The Medal Rankings: Which Country Leads the Olympics? [FINAL UPDATE]. Retrieved from https://beijing2008.blogs.nytimes.com/2008/08/23/the-medal-rankings-which-country-leads-the-olympics/.

Ludacer, R. (2018, February 6). No one wants to host the Olympics anymore - will they go away? Retrieved from https://www.businessinsider.com/future-olympics-no-country-wants-to-host-games-2018-2.

MacAloon, J. (1995). Politics and the Olympics: some new dimensions. Received from https://www.recercat.cat/bitstream/handle/2072/10632/WP053$_e$ng.pdf?sequence = 1

MacKenzie, M. (2019, July 16). Female Athletes Receive Only 4% of Sports Media Coverage-Adidas Wants to Change That. Retrieved from https://www.glamour.com/story/female-athletes-receive-only-4-of-sports-media-coverage-adidas-wants-to-change-that.

Malfas, M., Theodoraki, E., & Houlihan, B. (2004, September). Impacts of the Olympic Games as mega-events. In *Proceedings of the Institution of Civil Engineers-Municipal Engineer* (Vol. 157, No. 3, pp. 209-220). Thomas Telford Ltd.

Mervosh, S., & Caron, C. (2019, March 8). 8 Times Women in Sports Fought for Equality. Retrieved from https://www.nytimes.com/2019/03/08/sports/women-sports-equality.html.

Mower, J. (2012, January 1). London 2012: Olympic success is key to national pride. Retrieved from https://www.bbc.com/news/world-16245075.

Nunes, R.A. (2019). Women athletes in the Olympic Games. Journal of Human Sport and Exercise. 2019, 14(3): 674-683. doi:10.14198/jhse.2019.143.17

Osada, M., Ojima, M., Kurachi, Y., Miura, K., & Kawamoto, T. (2016). Economic impact of the Tokyo 2020 Olympic games. Bank of Japan Reports and Research Paper, Tokyo, January. Retrieved from

http://www.trustfu.com/Uploads/image/20180827/$20180827212024_95674.pdf$

Rathke, A., & Woitek, U. (2007). Economics and Olympics: An efficiency analysis. Available at SSRN 967629.

Sandercock, G., Beedie, C., & Mann, S. (2016). Is Olympic inspiration associated with fitness and physical activity in English schoolchildren? A repeated cross-sectional comparison before and 18 months after London 2012. *British Medical Journal Open*, 6

Soós, I., Martínez, J. C. F., & Szabo, A. (2017). Before the Rio Games: A retrospective evaluation of the effects of the population size, GDP and national temperature on winning medals at the 2012 London Olympic Games. *Journal of Human Sport and Exercise*, 12(1), 246-250.

Wikipedia (n.d.). Retrieved from https://en.wikipedia.org/wiki/Table$_tennis_at_the_2020_summer_Olympics_{-Qualification.}$

Warburton, D. E., Nicol, C. W., & Bredin, S. S. (2006). Health benefits of physical activity: the evidence. *Cmaj, 174* (6), 801-809.

World Bank (2018). Population 2018. Retrieved from https://databank.worldbank.org/data/download/POP.pdf