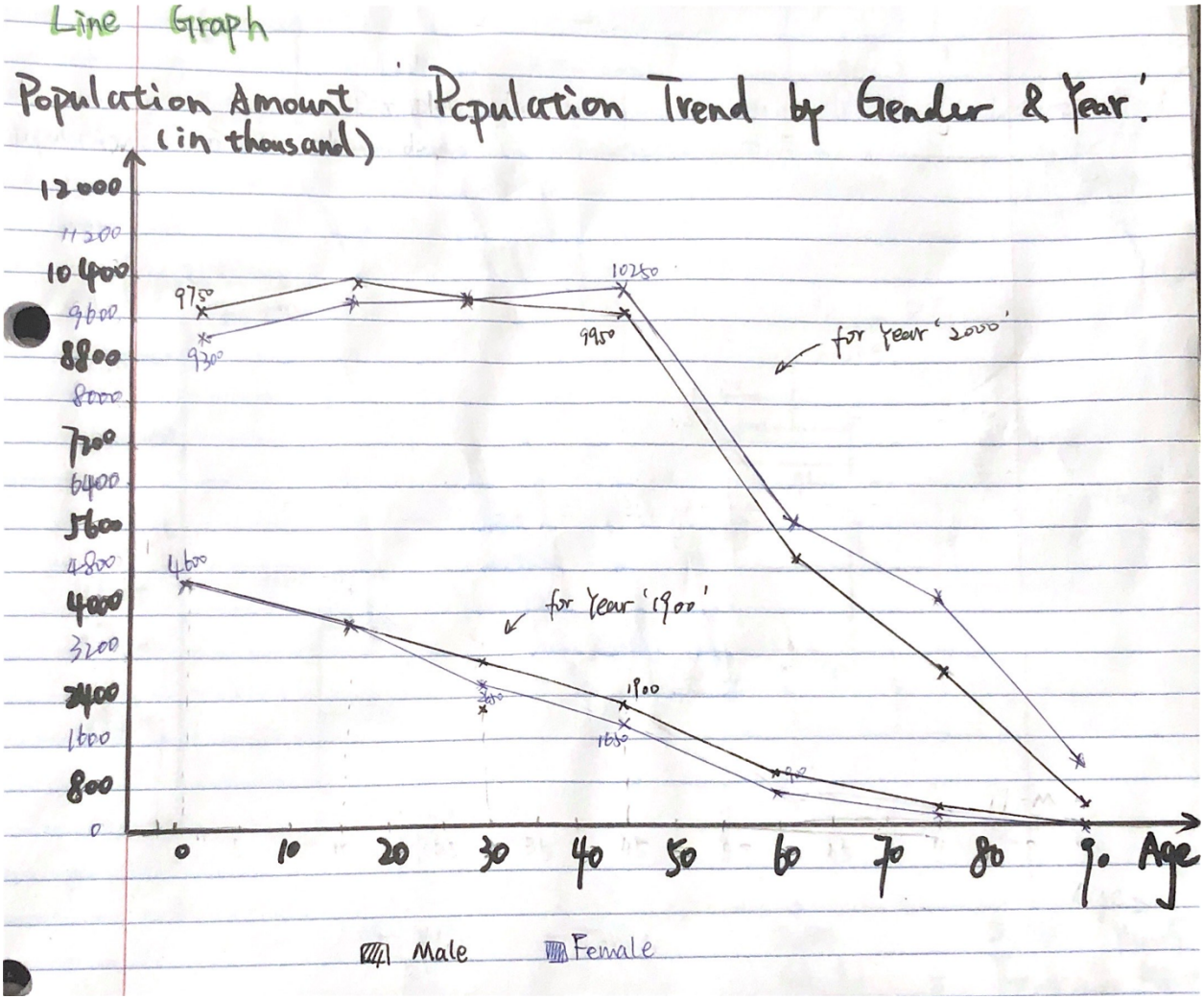
Question:
Which gender has the most changes in population based on different ages?

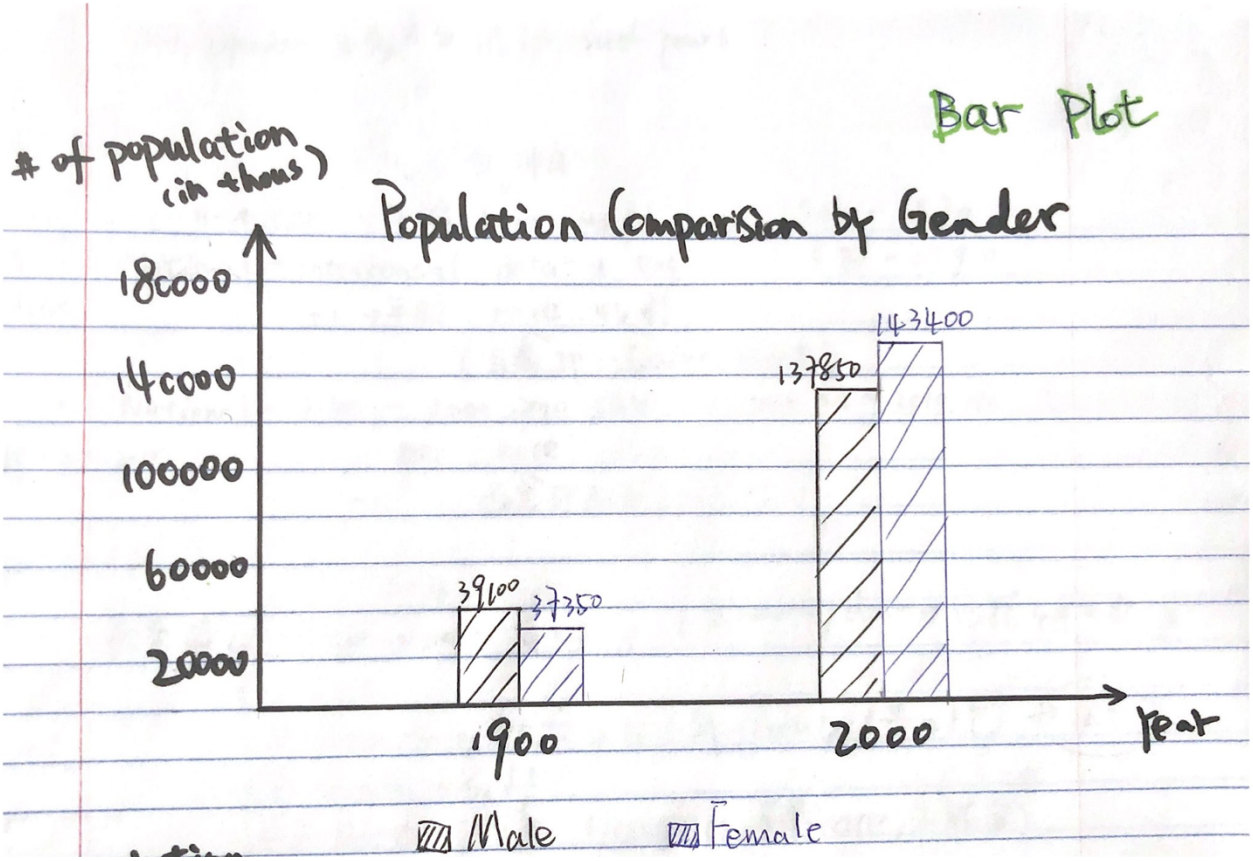
The Line graph:



This line plot is created for showing populations changes by age and gender in trend. I choose line graph because it can show data variables and trends very clearly. For example, being classified in different colors, the plot shows a very straight forward trend pattern of population change in both 1900 and 2000. This line chart communicates that although these four variable lines shared a decrease trend, the fluctuation changes of the process are different. For instance, there is a maximum dramatic decrease from age 45 to age 60 in 2000 for both male and female. However, for population change in 1900, the fluctuation is very gentle which might contribute to the population at age 0 in 1900 is already very low (4600,000). This line plot works very well for showing the overall view and capturing major changes on fluctuations. Moreover, it allows possible estimation of missing data. The disadvantage of line plot is that plotting too many lines

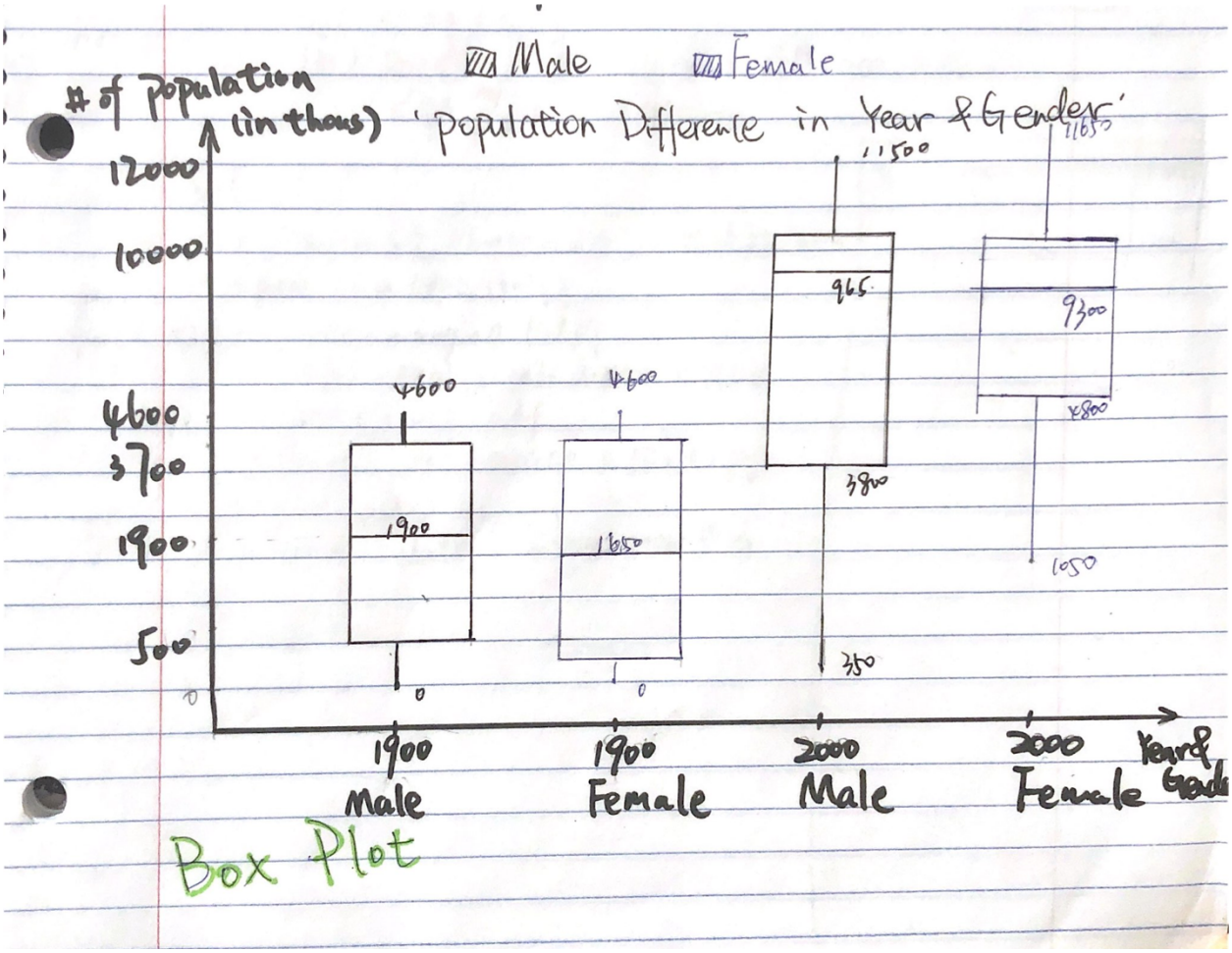
in one graph would make data difficult to compare. Also, it is time consuming to check out exact values for data and would be difficult to sketch if the dataset is too large.

The bar plot:



This sketched plot is generated for comparing the total change of population from 1900 to 2000 by gender. This is because bar plot can display multiple categories, summarize key values, and easily compare larger changes among groups. Benefited from the advantages of bar plot, it can communicate the total difference in 1900 and 2000 as well as the total increased number across genders. Moreover, I hope readers can easily understand the total population changes over time. From the plot, it can be clearly shown that there is a very high difference for population in 1900 and 2000. What's more interesting is that the population of 'female' in 1900 is the lowest, but in 2000 it's the highest one which shows the overall dramatic changes clearly. However, this plot still has disadvantages. Although we know the there is a huge difference between 1900 and 2000, it is hard to show the fluctuation of data vividly in bar plot. For example, compared with the line graph, it can not display population changes with a clear trend. In contrast with the scatter plot, bar plot lacks of precision in establishing the dataset since it does not contain sufficient details.

The box plot:



I sketch this plot to give insight on indicating the total difference in each year and gender by exploring the maximum and minimum number of populations in ages. I hope through this plot, readers can get a clear version of graphical display of the distribution of datasets. This box plot shows very specific median, lowest and highest quarters value, which can be used to compared one variable to the others. For instance, we can see that the population of male in 2000 have greatest difference in minimum and maximum by ages. Moreover, it shows that compared with the left skewed pattern of 1900 population, 2000 has a right skewed one since the median value is very close to the highest one. This shows that an overall increase population when year goes up. Also, it shows that a huge difference of 1050,000 in minimum population for 1900 and 2000, and a 7050,000 in highest population. Moreover, the box plot shows statistical info that the range of each gender's population also enlarged. However, it also has disadvantages of not keeping the exact values and details of the distribution results and only shows a summary of the dataset. For example, it's not convenient for users to check which specific age has the lowest population in 1900 and 2000.

Write **one paragraph** that reflects on all 3 of your sketches overall. Compare your designs with each other — what are their strengths and weaknesses?

Comparing three sketches, the most appropriate plot should be the line plot, but all three plots have strengths and weakness. For bar plot, the strength of it is showing the relationship between the population and categoric variable of gender. Also, it provides a summary for readers to easily understand the overall changes which provides insights of very high difference for population in 1900 and 2000.The weakness of bar plot is that, compared to line graph, readers would have no information about specific data point such as age. Also, compared to box plot which is appropriate for showing statistical info, bar plot can only show the population counts in total. For box plot, it shows very specific medians, minimum, and maximum by gender in 1900 and 2000. Also, using box plot gave an overview of the fluctuation range of each category. The advantage of box plot is that it can provide a clear dataset summary and very simple to sketch. However, compared to line plot, box plot does not show exact age of the lowest and highest population. Sometimes, readers might be confused since the information provided is not sufficient. For line graph, it is the most appropriate on among these three since line graph shows a trend of the whole dataset, and it also captures major fluctuation of the population changes. Moreover, line graph contains sufficient information of the age, and it is easy to match fluctuation with specific age. Besides, this line graph also has subcategories as gender and year which is useful to compare different variables trends. However, there are still weakness for line plot that it is not easy to sketch if the dataset is too big; and it's hard for readers to compare if there are too many subcategories in the same line plot. To take advantage of all three plots. It would be better if we can combine the details and trend of line plot, the clearness of bar plot, and the statistical info of box plot together.