CS590V: Data Visualization and Exploration– Spring 2017
Homework 1: Plotting of Graphs using D3 and Python
Submitted by: Ankita Mehta (30450796)

Preprocessing:

1. Commas have been removed from CSV files using python "re" tool.
2. Some of the columns have been grouped manually.

Analysis: Tables 2A-2J and table 1 have been analyzed and various graphs have been shown below:

Interactions: Tooltips, mouse hover interactions have been added in all graphs.

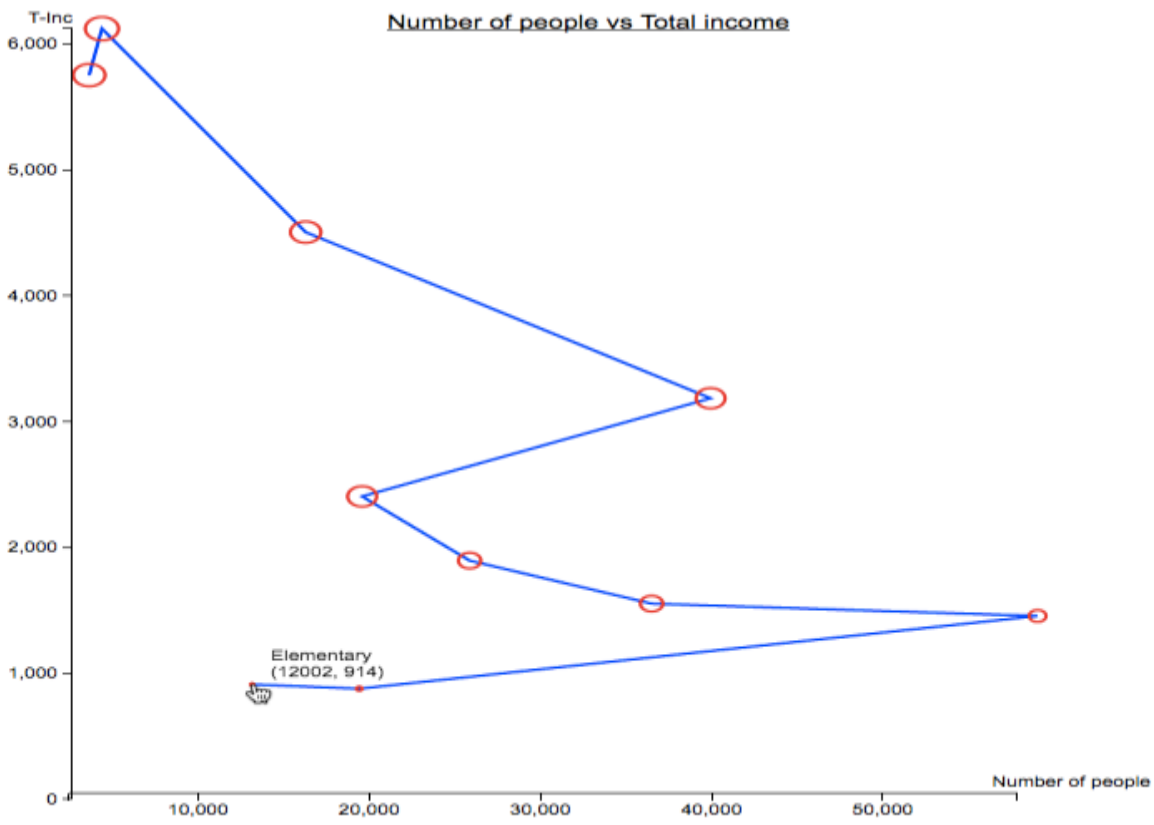
Tables 2A(Mean) and 2J(Median):

General trend which has been analyzed with the two tables having Average and Median values of Monthly income, earnings and average months worked is:

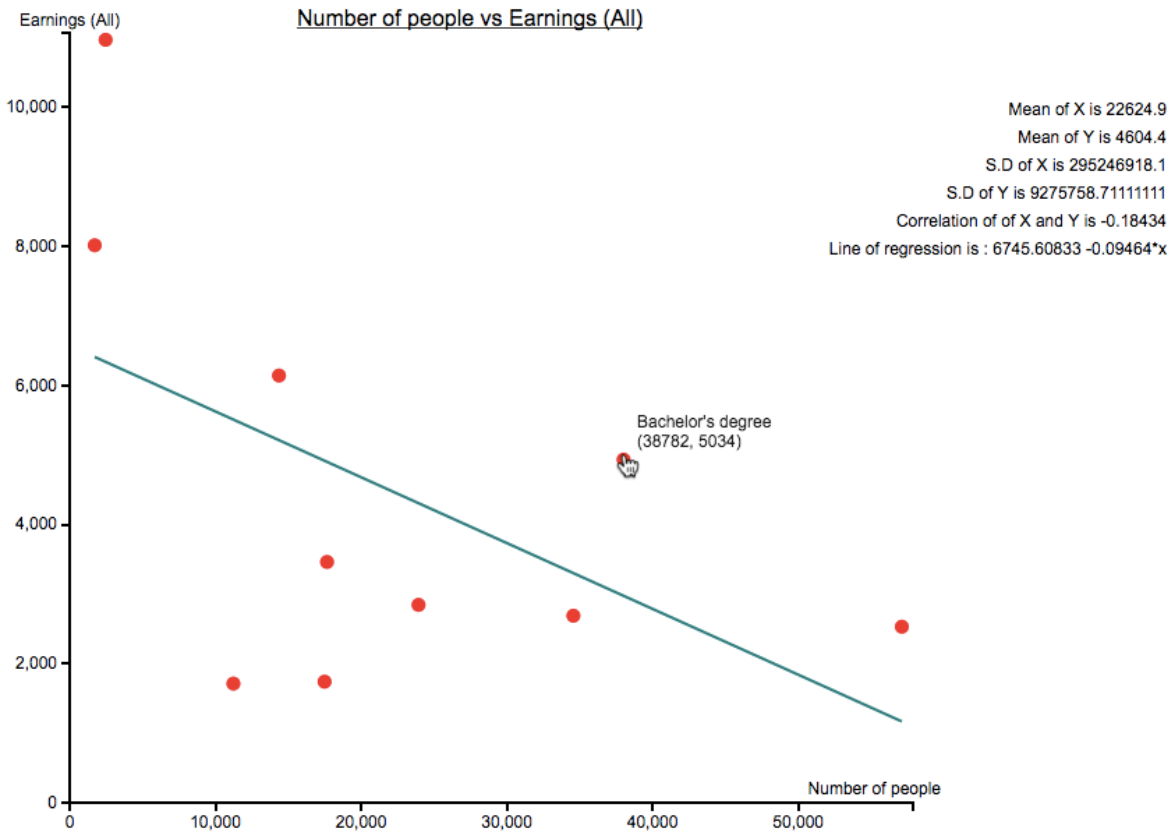
1. With the increase in the level of education from Elementary to Doctoral, number of people has been **decreased** but their average, mean income and earnings have been increased.
2. Also, the average number of months worked have also been increased with the increase in the level of education.

The analysis can be shown in the following graphs:

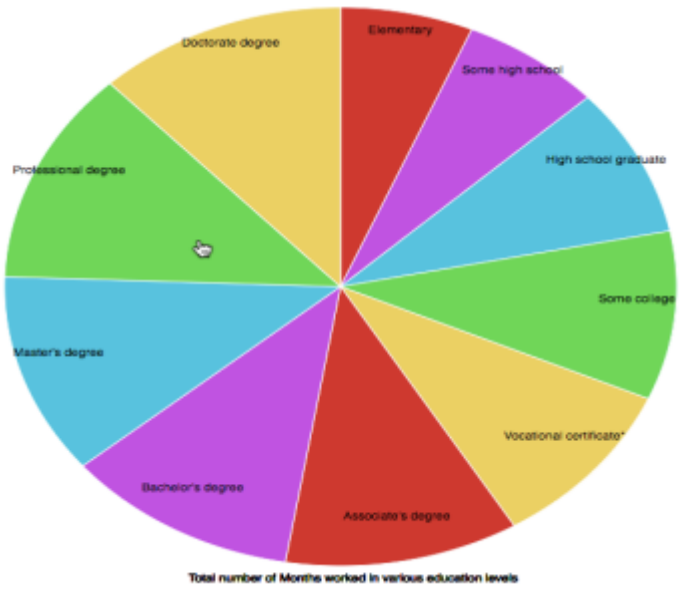
1. The below scatterplot shows the Number of people and their Average income for various levels of education. Here, the tooltip has been included to show the education level.



2. The below scatterplot shows the various statistics and the line of regression between number of people and their Average Earnings(All).
It can be inferred that as the level of education has been increased, the number of people have been decreased with the increase in their Average earnings(All). Hence the negative correlation coefficient justifies.



3. The below graph is the pie chart showing Average number of months worked in various degree levels.

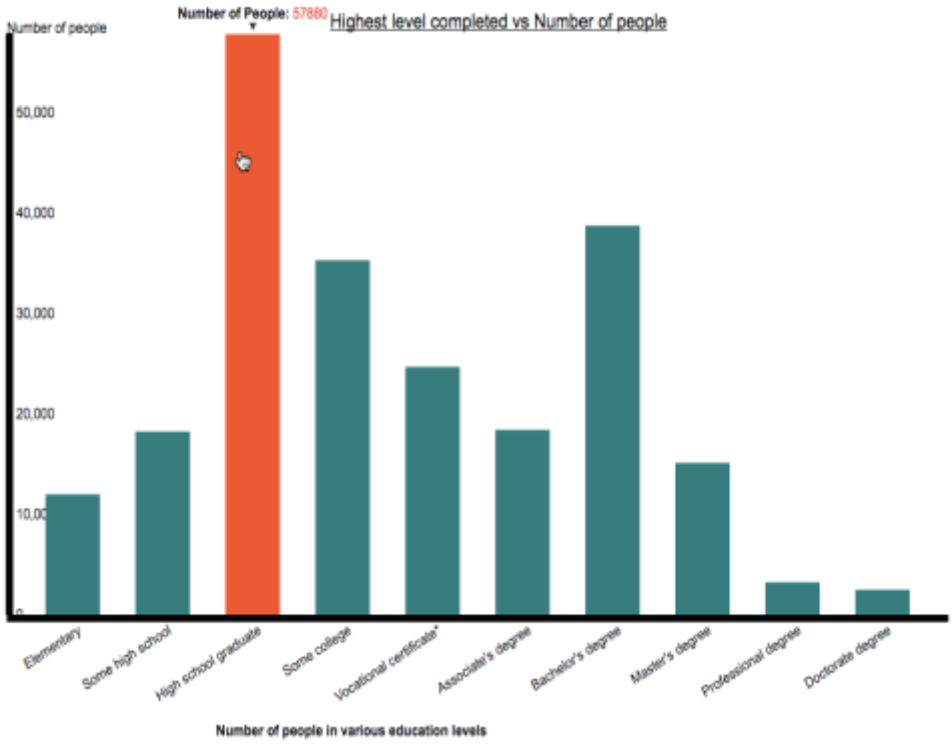


4. Next graph is the merged scatterplot of the Number of people with their Mean and Median Earnings (All) for the various education levels.
It can be seen that "Mean" values are always higher than the "Median" values which means the data present is right skewed data i.e, people with higher education have more earnings.

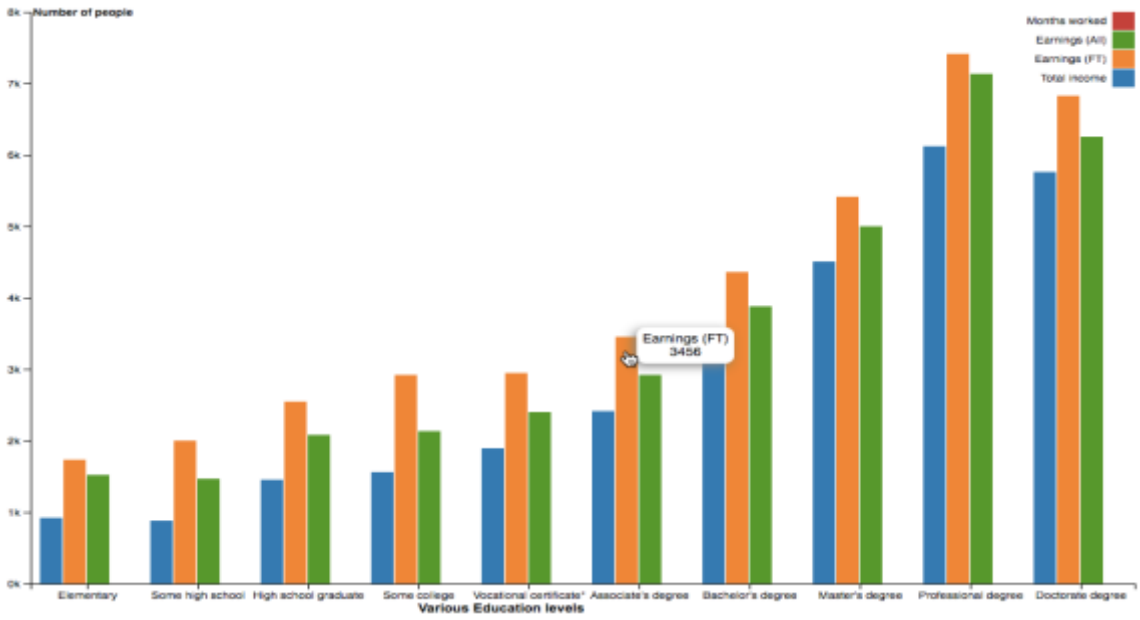
Here radius of circle represents the Average number of months worked.

This figure illustrates that the mean values are always larger than the median values since there are a number of large Earnings (FT) that make it larger.

5. Below is the bar chart showing the number of people in different education levels.
A tooltip has been included in this to make it interactive.



6. Following is the grouped bar chart which is showing the increasing trend of Average Income, Earnings and Months worked.
Months worked are very small as compared to other fields, therefore not visible.
It can also be seen that Full-time Earnings are always larger than the All earnings (FT + PT)



7. Next is the stacked bar chart showing the Mean and Median data of Professional degree.
It can be seen that its mean values are very larger than the median values.

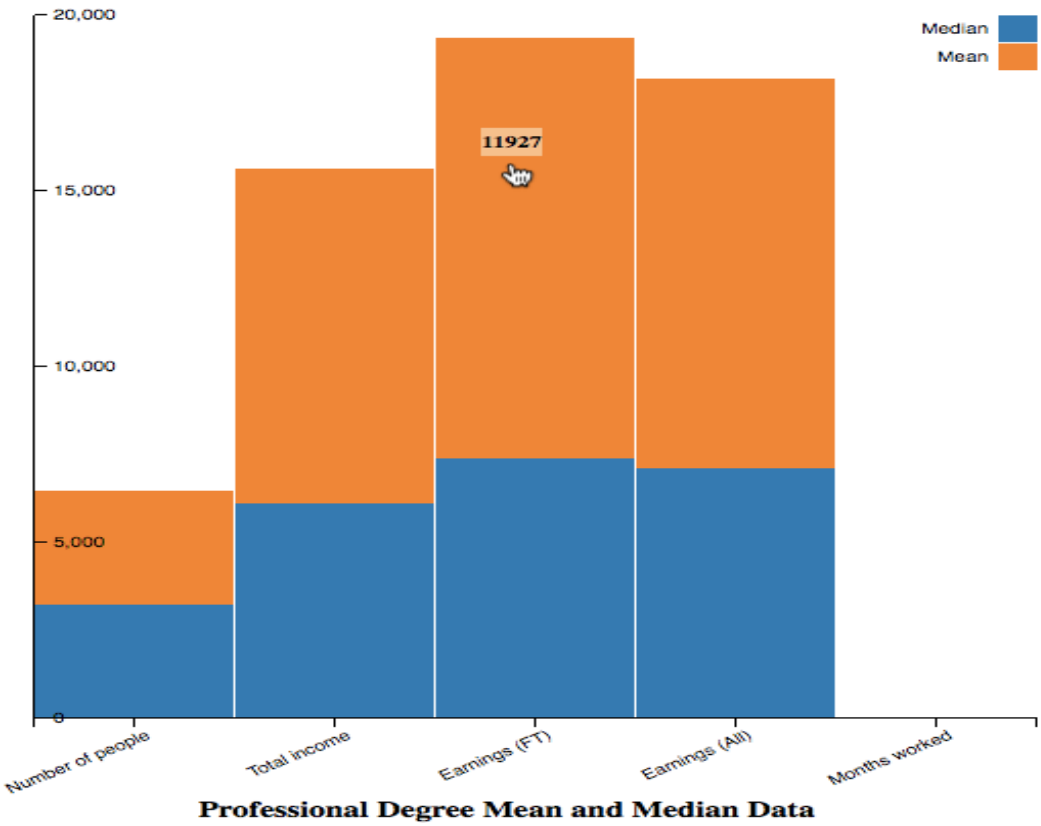
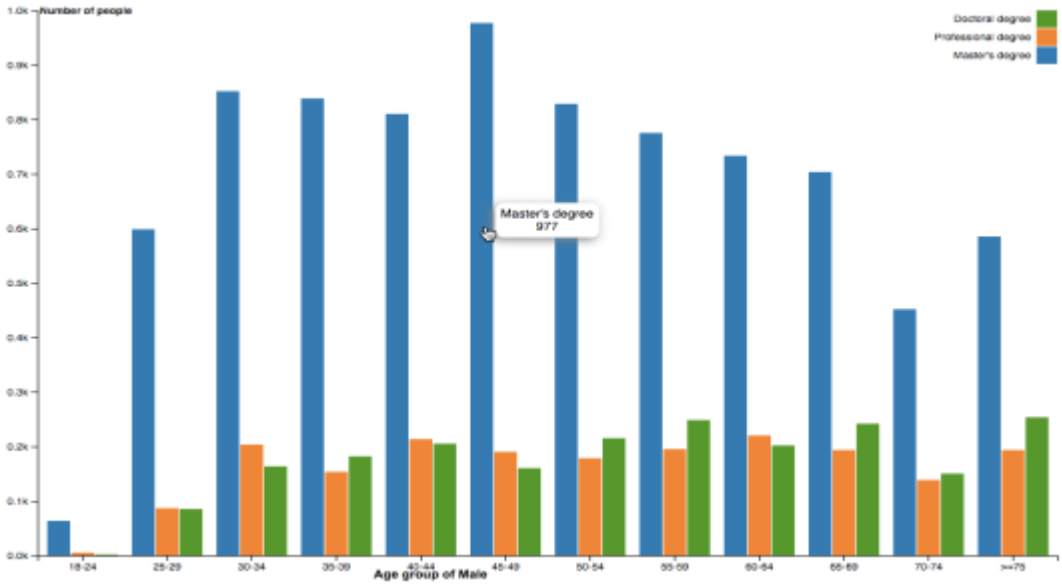


Table – 1 (All Races):

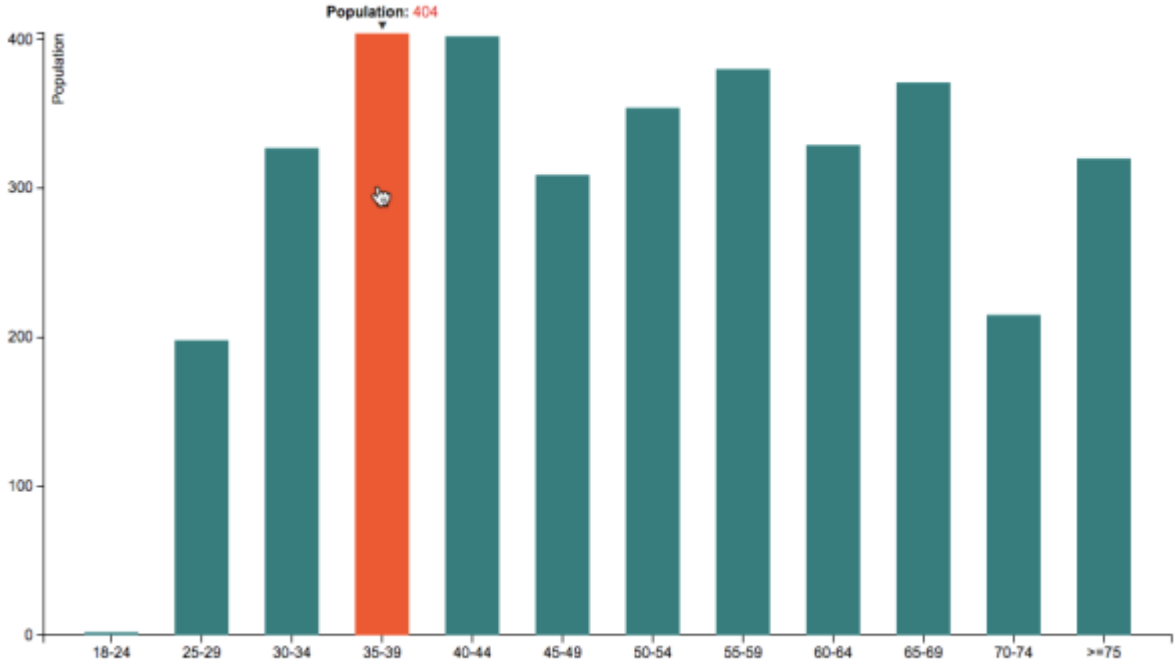
Following are the graphs showing the number of Male and Female of various age groups in different education levels.

1. First is the donout chart showing the total number of Female (of any degree) in various age groups. It can be seen that maximum number of females are in the 18-24 age group who are taking any sort of education.

2. Next is the group chart showing the number of males of various age groups who are taking Professional, master's and Doctoral degree.
   It can be inferred that large number of males take master's degree in 45-49 age group.

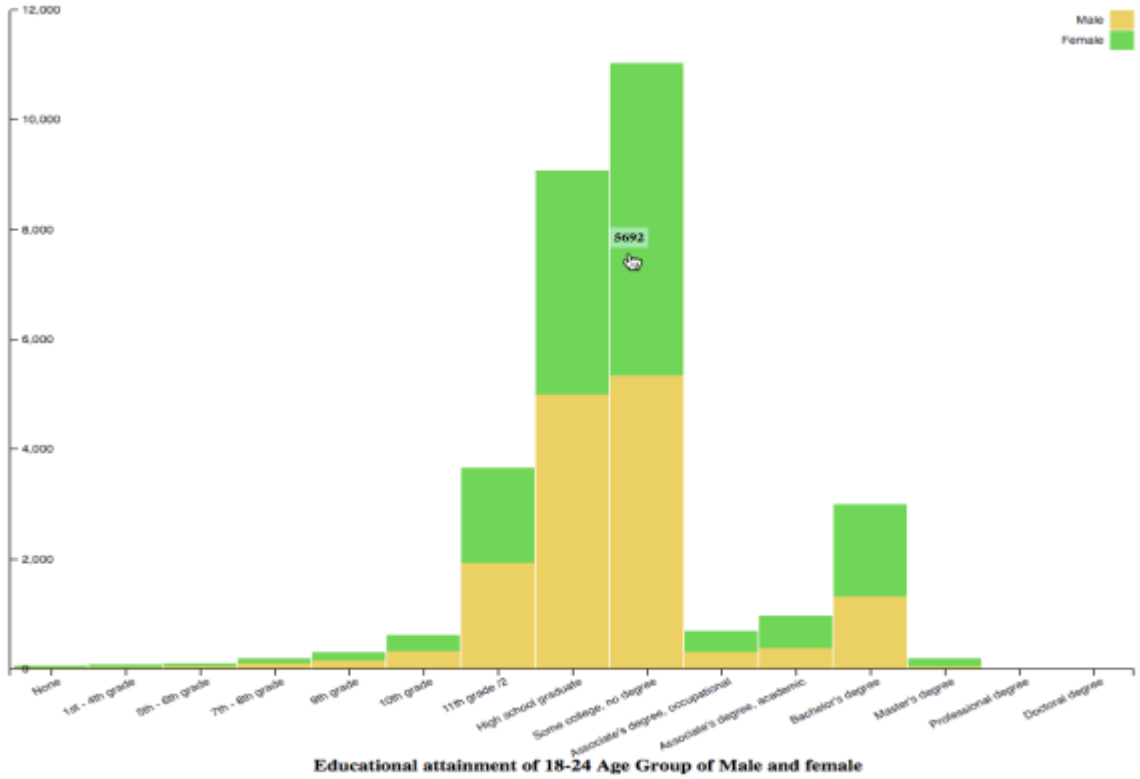


3. Next is bar chart showing the total number of people in Doctoral Degree of various age groups.
   It can be seen that maximum people who are taking Doctoral Degree are of 35-39 age groups.



Population in various age groups of Doctoral degree of Both sexes

4. Next is the stacked bar chart showing Educational attainment of 18-24 age group of male and female. It can be seen that till high school males were larger in number, but afterwards females grew in number as the education level increases.



Educational attainment of 18-24 Age Group of Male and female