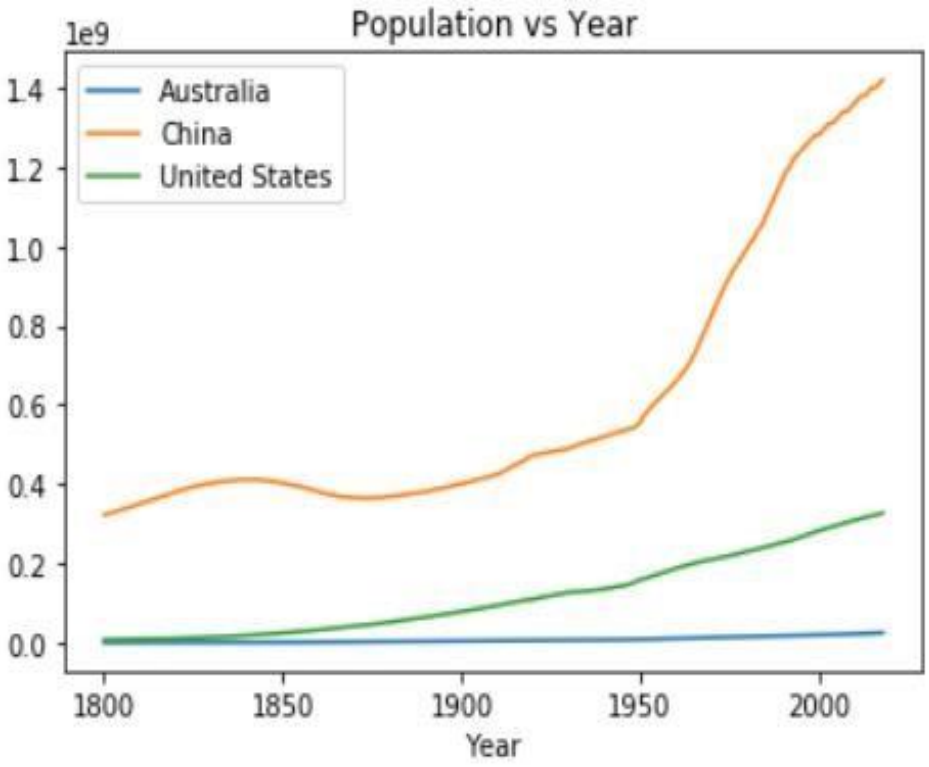
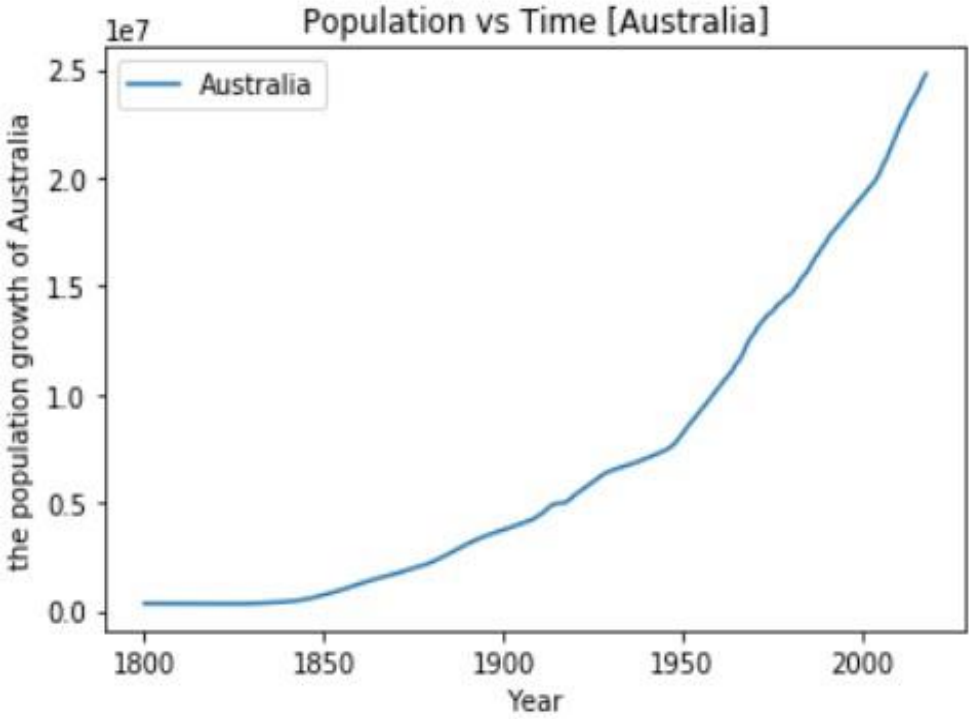
Education Analysis Project

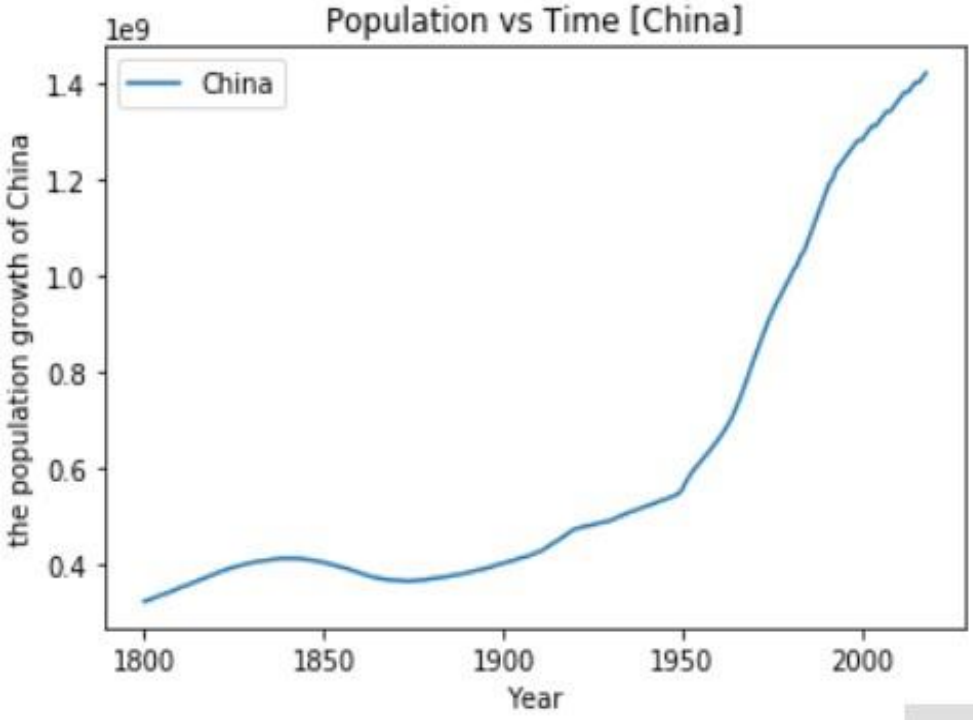
TASK-A:

A1:

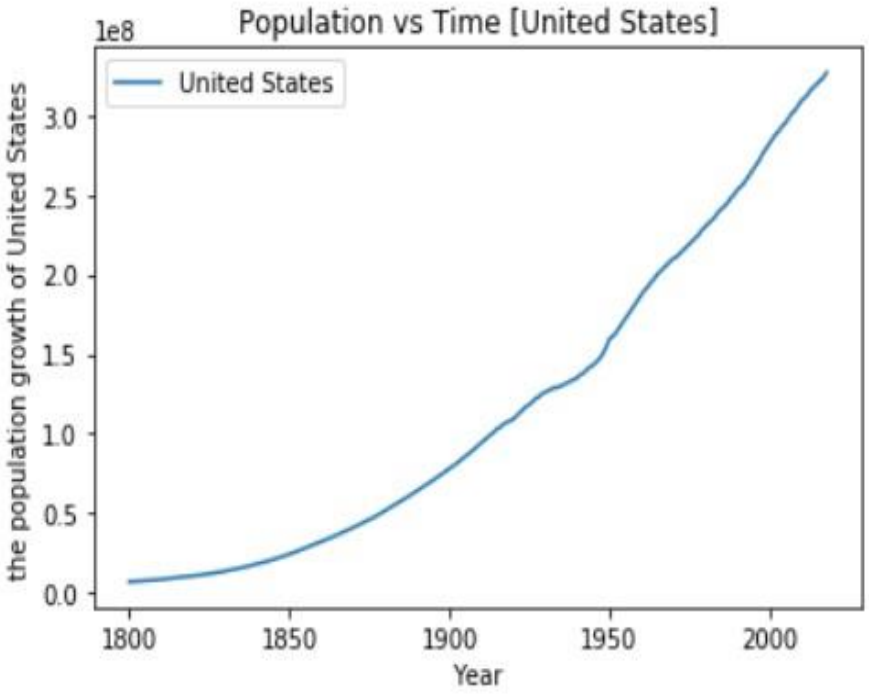
1)




Population vs Time [Australia]
1e7
Australia
the population growth of Australia
2.5
2.0
1.5
1.0
0.5
0.0
1800
1850
1900
1950
2000
Year

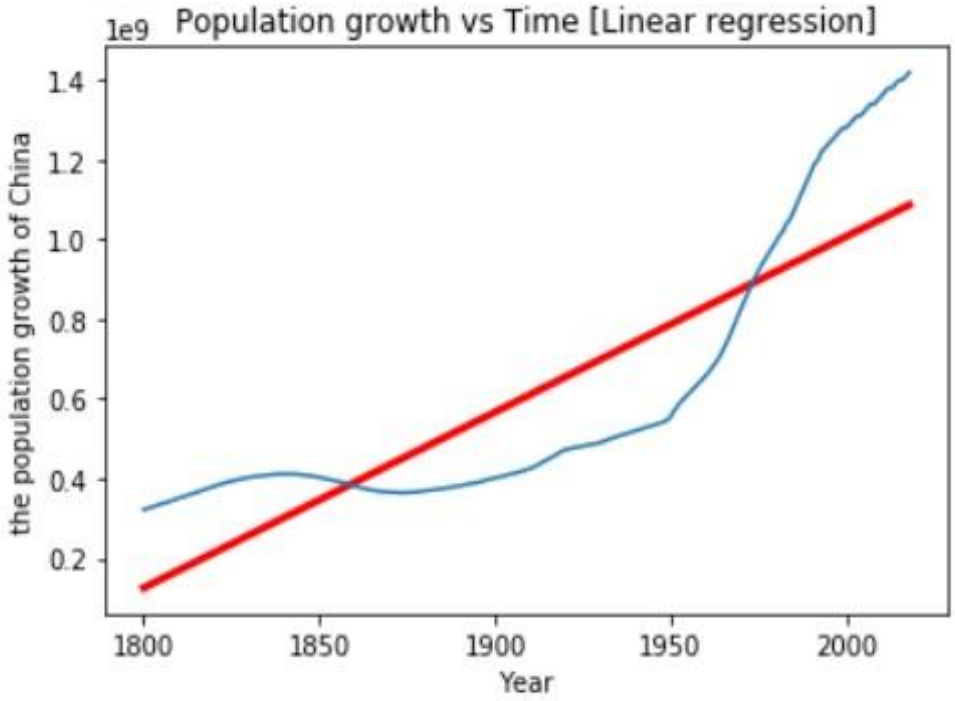

Population vs Time [China]
1e9
China
the population growth of China
1.4
1.2
1.0
0.8
0.6
0.4
1800
1850
1900
1950
2000
Year

Rename



As it can be seen in the graph above, the population values for all three countries (Australia, United States and China) are all increasing with respect to time. The overall graph of Population values of Australia is increasing over time where it is stable until almost 1850 and it starts increasing from then onwards. While the Population for China is overall increasing as well in which the graph starts increasing up till 1850 and got decreased linearly until 1875 and increases the same linear way for next one hundred years. After that, it got increased drastically up till now (even faster compared to how it increased during 1875 to 1975). In the end, the graph of United States population versus time is increased linearly almost everywhere. It has increased a bit slowly during first fifty years but then it has got stable linearization all the way (except a decade or two before 1950 where it went a bit slow as well).

2)



The Chinese population data linear regression graph is as it is indicated above.

The graph seems to be linear in most cases in the figure. At first, it starts increasing a little bit slowly for almost first fifty years compared to Linear regression line. And the graph is going down linearly from around 1850 till about halfway (1875) before 1900. From 1875, it is going up linearly until 1950. After that, it is increasing way too much quickly until the end which is linear as well. In conclusion, the graph is linear, but it is divided into a few parts in terms of the value of slope and the speed of incrementation of linearization. After all, the graph is definitely linear, and it is good fit in terms of linearization.

As the values of slope and interception are found from the graph above, we can use the linear fit line formula to predict the values of resident population in 2020 and 2010.

```
Intercept = -7814934677.299574 Slope
= 4411204.542598283
```

If we substitute these values in linear equation:

For year 2010,

Population value = slope*Time + Intercept

= 4411204.542598283* 2010 + (-7814934677.299574)

= 1095698498.748958

Population value = 1.051 * 10^-9 [1e9]

For year 2020,

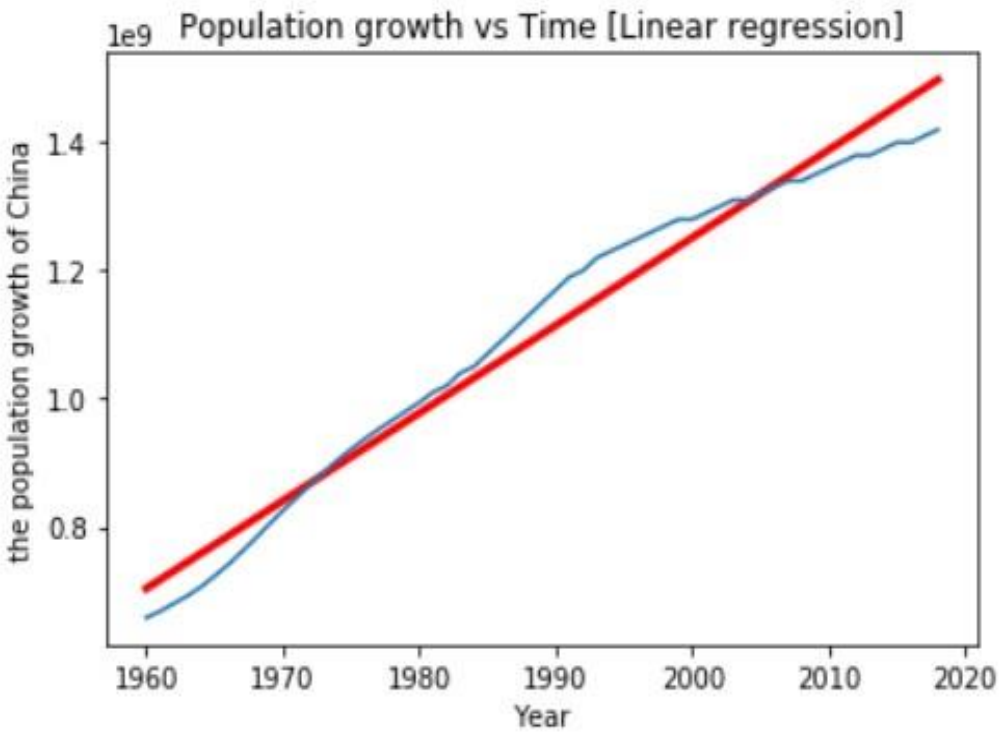
Population value = slope*Time + Intercept

= 4411204.542598283* 2020 + (-7814934677.299574)

= 1051586453.322975

Population value = 1.095 * 10^-9 [1e9]

Hence, the predicted population values in the year 2010 and 2020 are 1.095 * 10^-9 and 1.051 * 10^-9, respectively.



As it can be seen in the graph above, the plot is provided as the data is from 1960. This graph looks more linear as compared to the graph we got from the last question which was from data of every year. If we

try fitting the graph to more recent datapoints rather than everyone, it will give us more accurate results in this case. The graph seemed to be almost same as linear line which is great for linear fit. The better prediction of future population would be given more accurately by this graph as it is more like a linear. So, the future population values according to this graph would be as follow.

```
Intercept = -26144732612.507313
Slope     = 13698129.748684982
```

If we substitute these values in linear equation:

For year 2010,

Population value = slope*Time + Intercept

= 13698129.748684982*2010 + (-26144732612.507313)

= 1388508182.349649

Population value = 1.388 * 10^-9 [1e9]

For year 2020,

Population value = slope*Time + Intercept

= 13698129.748684982*2020 + (-26144732612.507313)

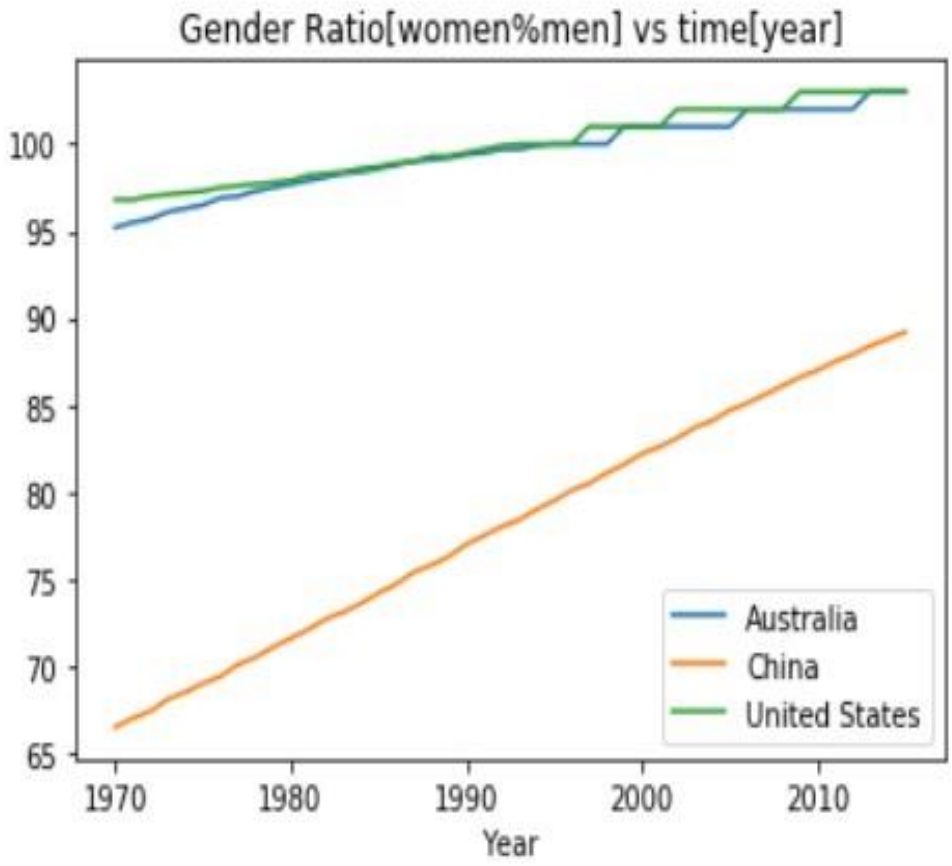
= 1525489479.83635

Population value = 1.525 * 10^-9 [1e9]

We can see that from 2010 to 2020, the predicted value of future population in china is increasing by greater value in recent data graph than in the graph we got from original data.

A2:

1)

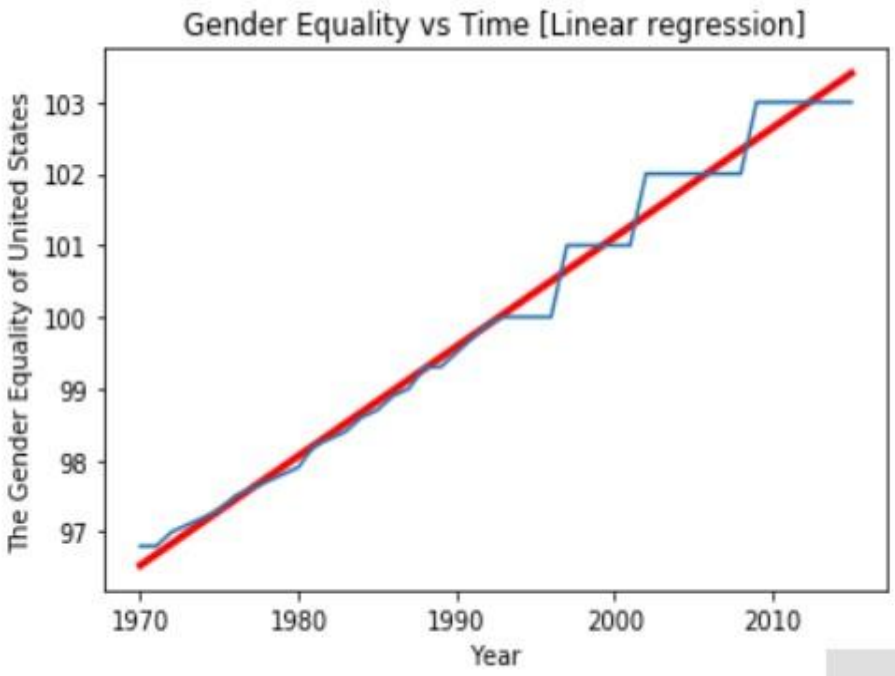


The plot for the gender ratio in schools of Australia, China and United States over time is shown in a graph above.

As it is shown in the python file, the maximum and minimum gender ratio (women% men) for schools in Australia over the time period is 103.0 and 95.2, respectively.

According to the graph shown above, two countries, Australia and United States are having similar growth trend in the graph. Those two countries are having high interception values while China is having proper linear trend with low intercept. Moreover, the gender ratio of US and Australia is having linear trends, but Australian Gender ratio's growth is increasing more faster than US. At some points, they both are intersecting with each other and after that the growth is going up and down three times for both countries until the end which is divided into segments and making a little difference between both trends.

2)



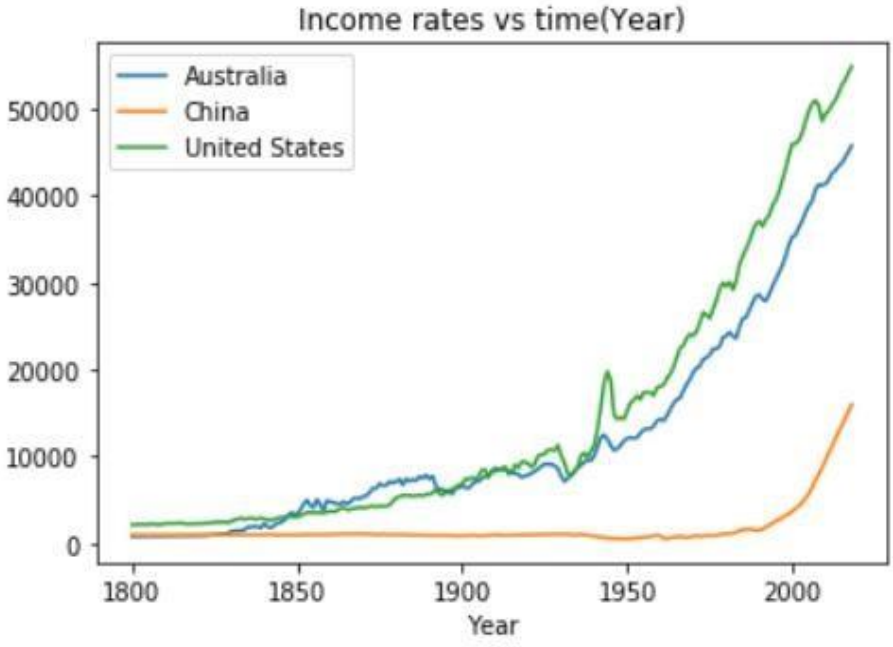
Rename

The gender ratio of United States over time is plotted as shown above in the picture. The graph is very close to linear until after year 1990. However, it has eventually started going down and up from then which is dividing graphs into little segments. But the next segment seems bigger than the last one and segments are increasing each time which is making the graph non-linear in the future. It seems from the graph that it will have even more bigger segments in further years, so it does not look like it is going to be linear graph as it was before 1990.

A3:

1)

```
<matplotlib.axes._subplots.AxesSubplot at 0x164a1ac0cc0>
```



The Income of Australia, China, and United States over time over time is plotted in a graph which is indicated in the image above. The trend is growing up either way for United States and Australia which is not proper linear, but it is increasing over time. For China, the graph is stable for most of the time period. It is not growing up until after 1950. After that, trend is drastically going up which is really surprising.

As shown in the python file, the minimum Income value of country China from the entire dataset is 530 which is happening in the year 1949. In that year, Australia's income rate was 11800.

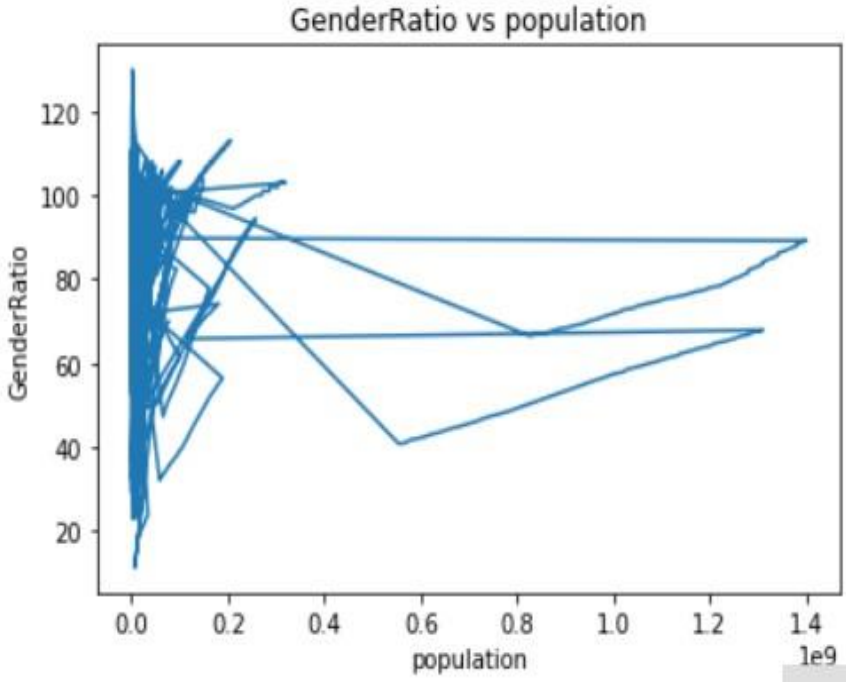
A4:

1)

As in the python file, all these datasets from population values, income, and gender ratio in schools for the different years and different countries are merged and it contains all intersecting values.

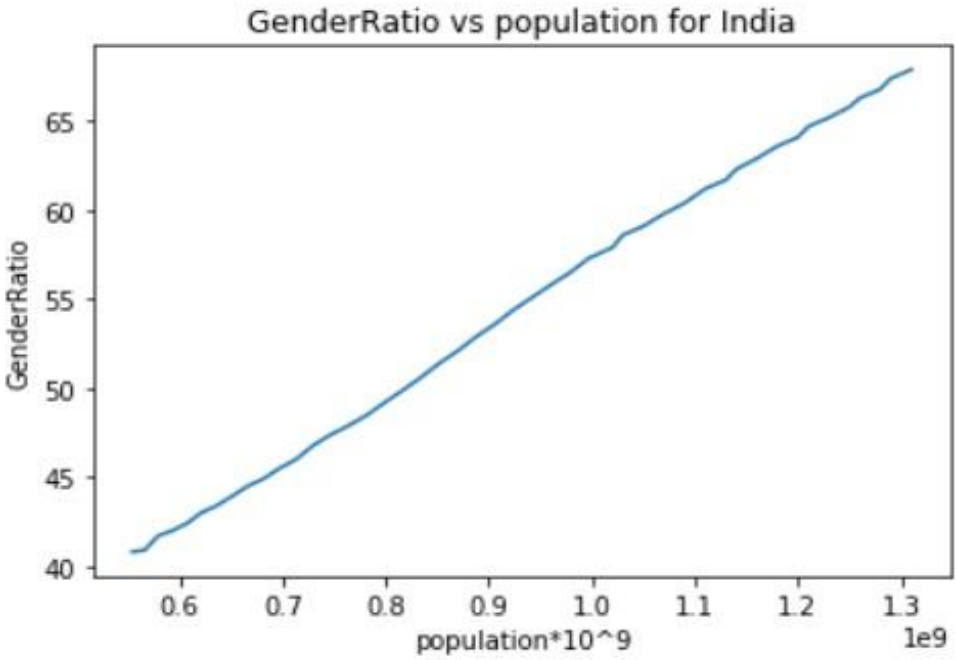
The first year and last year from the combined dataset are 1970 and 2015, respectively.

2)



The relationship between gender ratio and population is indicated above. It can be seen that the relationship is really stable and they both are on the same trend in the beginning. But it is so complicated when it goes further. We cannot see where the graph is going in that case because there are different values for different countries which are having their values on the graph. But it seems to be divided into several parts. Hence, the proper relationship is still unknown between these two variables.

3)



If I select data for any particular country, then it will be clearer than the one on last relationship. For the graph of the country India, it is proper linear graph. As you can see on the picture above, it is all linear at every point which is the relationship we can see for India. Hence, India is having a linear trend for their population values over their Gender ratio values.