Exploratory Data Analysis - Mini Project 1

Team Croatia

Prashanth Thirukkurungudi Sekar, Pravin Sundar,
Barathwaaj Parthasarathy, Bhavna Sinha
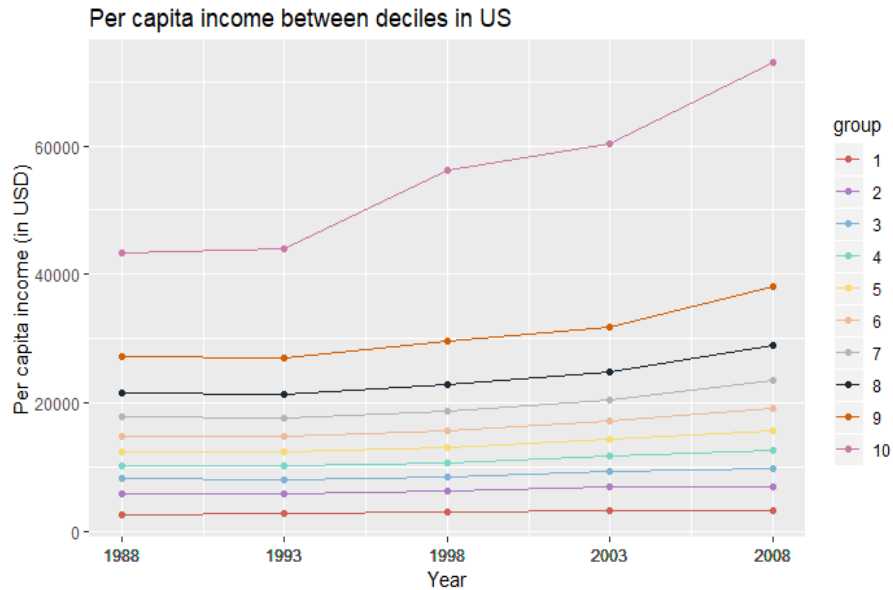
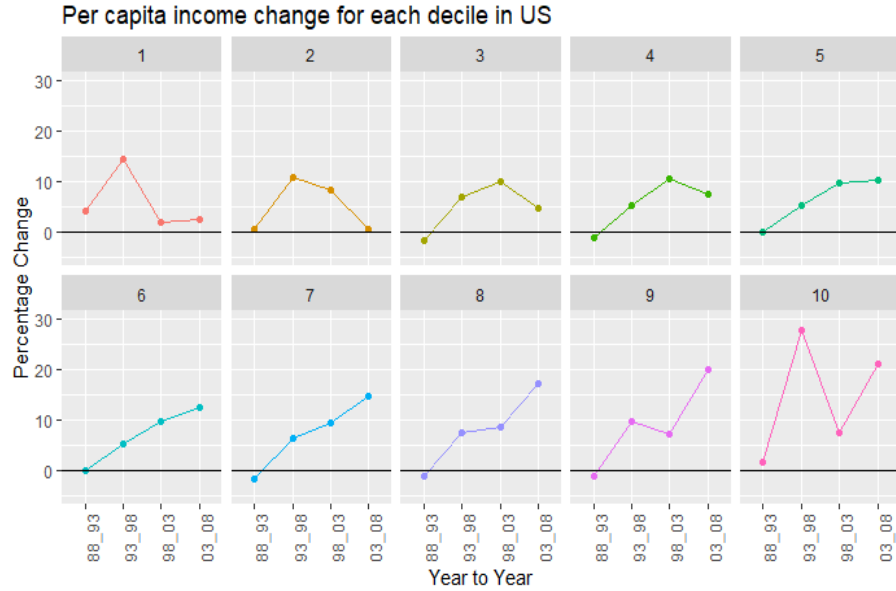February 13, 2019

1   Introduction

Income inequality is the extent to which the income is distributed in an uneven manner within the country and around the globe. Through exploring the distribution of per capita income using World Panel Income Distribution (WPID) and Ventile data sets, we find that income inequality is present throughout the world but substantially varies from country to country.

2   Per capita income distribution within U.S.A

We would like to first understand the distribution of per capita income within the United States across the income groups. We will look at the income of these groups across 5 year terms from 1988 to 2008.

From the graph above we can see a clear increasing trend for most of the deciles year on year; most notable being the 10th group(10% of individuals with highest income). What we can immediately conclude from the graph is that in the lower groups the increase is much smaller and these group incomes tend to be a lot closer to each other. The difference in income between ventiles in the lower ventiles is much smaller than the difference in income between ventiles 8,9 and 10.

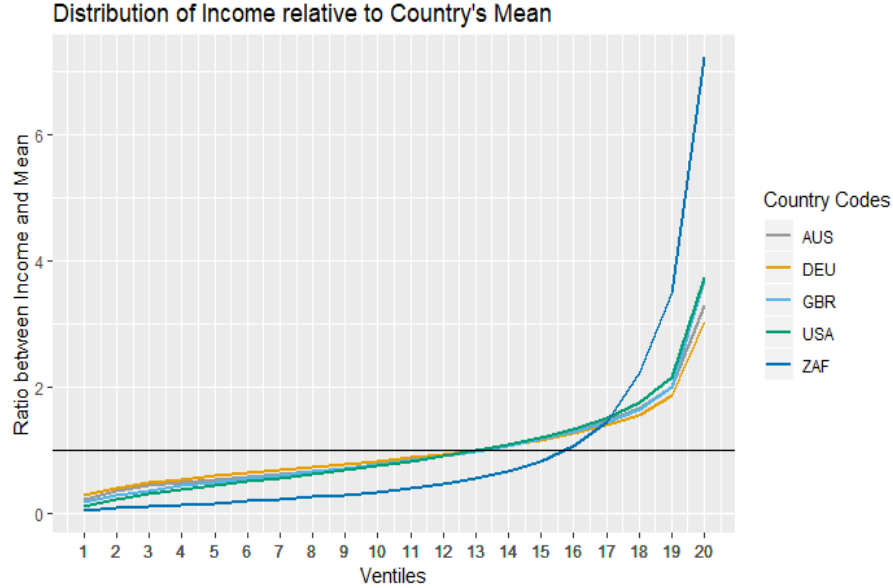## Per capita income change for each decile in US

Looking at the distribution of the difference in income for each decile. From the above facet graph we are trying to understand the "rate of increase" in each decile. It can be seen that groups 3 to 9 have a negative difference, i.e, the income has decreased from year 1988 to 1993. Every decile has seen a positive rate of increase in the year 1993 to 1998 with decile 10 seeing the highest increase (nearly 28%) and deciles 4 5 and 6 being the lowest (5% approx). From 1998 to 2003 deciles 1 2 9 and 10 has seen a decrease in rate of increase, in the sense that, although there is an increase in income in these deciles the rate of increase is smaller compared to the previous term. It is interesting to note that although 10th decile has the steepest decrease in the percentage, this decile still has a much higher income compared to other deciles. Finally in the last category (2003-2008) we observe a drop in rate of increase only in deciles 2 3 and 4 with marginal increase in decile 1 and 5. Overall we can observe a steady increase in deciles 5 6 7 and 8 and few fluctuations in the other groups.

Notably, we can see a diverging pattern within the deciles where decile 1 displayed a growth of 20% and 10th decile with 41%. The phenomenon of income inequality in the United States is clearly significant.

3 Distribution of income between selected countries

Now that we have seen the income inequality within US let us take a look at where US stands compared to other countries. The countries in consideration are : United States, United Kingdom, Germany, South Africa and Australia. To be able to compare the incomes across different countries we have decided to

take the ratio between the ventile income and the country's mean income. This will resolve the effect of the variation in income between countries across ventiles.
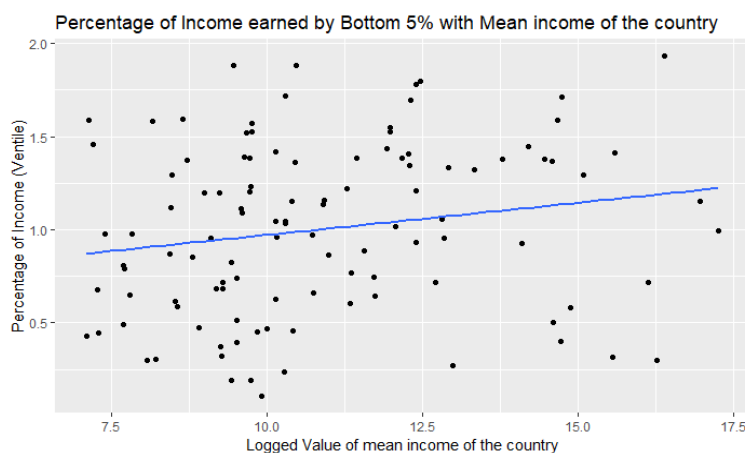

What can be instantly observed from the graph is that South Africa has the highest relative income among the countries in the last ventile and the lowest relative income at ventile 1, meaning, among all the income generated by 5% of individuals in a country with the lowest and highest income, South Africa places itself at the extreme (7 times it's mean in ventile 10 and lowest in ventile 1). Secondly, we can see that other 4 countries follow similar trends across the ventiles. We can see that in countries South Africa, USA, United Kingdom, Australia and Germany the degree of income inequality keeps increasing (in this order). USA has the second lowest ratio at ventile 1 and second highest in ventile 20 and so on. Which means the low income individuals in South Africa are earning even lesser than other countries and the high income individuals are earning much higher than other countries. The same goes for the other countries in the specified order above.
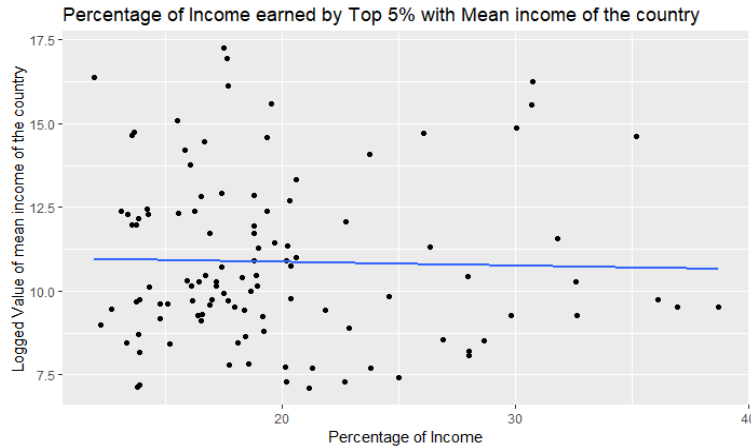
It is interesting to note that the mean income for South Africa occurs at ventile 15-16 but for all other countries it occurs at ventile 13. Leading to derive a conclusion that the income generated in South Africa in ventiles from 16-20 is much higher compared to all the other ventiles implying that the income inequality, compared to other countries is greater in South Africa.

4 Income contribution for top and bottom 5% around the globe

In this section, we will discuss how much the bottom 5% and top 5% income generating populations affect the mean income of the country. A country's economy is not just defined by the richest people. The mean income of the country will still depend on how much the poor people earn. We will see here which of the two has more influence on the mean income of a country.

In the data, China was divided into two - China Rural and China Urban, hence for this analysis they were grouped together as one country, China. In the two linear model plots below, we can see that the mean income has a positive relationship with percentage of income generated by bottom 5% of the population. There is a marginal decrease in mean income with increase in percentage of income generated by top 5%. There is reason to believe that the mean income does not depend on the percentage of income generated by top 5% of the population since this decrease is negligible.

Percentage of Income earned by Bottom 5% with Mean income of the country

Percentage of Income earned by Top 5% with Mean income of the country

Note: For top 5% Guinea-Bissau was excluded as it was an outlier with 60%

The regression line of the plots suggest that there is a positive relationship between percentage of income earned by bottom 5% and the mean income of the country. In the second plot, it is observed that the two variables are almost independent of each other. There is a slightly decreasing relationship which is negligible.

Looking at the two plots (World Map) in Appendix, we can see that there are three types of countries. The first set are Southeast Asian countries and some central African countries. These countries correspond to high share of income generated by the bottom 5% of the population(Pakistan,Armenia,Egypt,etc.,). We can see that these countries do not have a significant share of income generated by top 5% of the population. These countries have the least per capita income.
The second set of countries are those which have high share of income generated by top 5% of the population. These are generally South American and South African countries. The share of income generated by these countries in the bottom 5% of the population are very low. This suggests that there is imbalance in incomes of the people.
The third type of countries are those which are more economically stable countries. These countries have a balanced income distribution among the ventiles (USA,Argentina,Russia,etc.,) with contributing percentages occurring in the middle in both the plots.

5   Conclusion

Per Capita Income distribution in the US: There is a steady increase in the per capita income through the years in consideration with a much higher increase seen in decile 10.

Distribution of income between selected countries: Among selected countries South Africa shows distinctive income inequality with higher ventiles earning much more than the lower ventiles and other 4 countries show similar trends with different magnitudes of the same.

Relationship between top and bottom 5% in a country and its mean: The percentage income earned by top 5% in a country does not affect the mean but evidently the bottom 5% has a very small negative influence. This might be due to the regional differences as the variation can be grouped into different segments of countries.
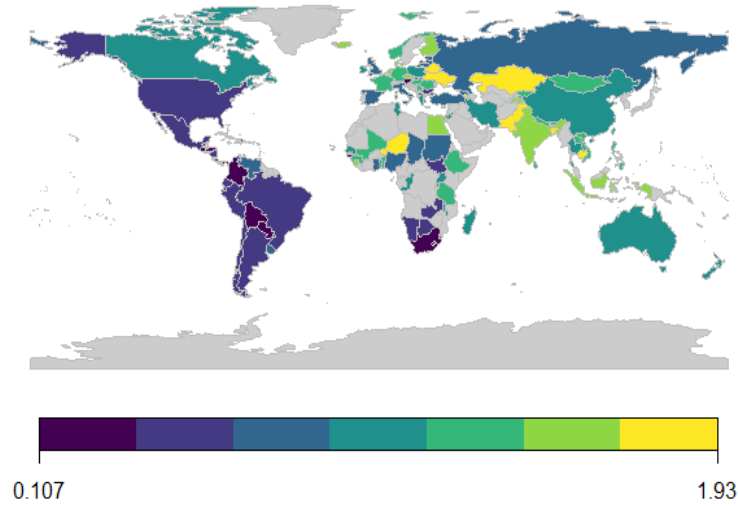
In conclusion, we can declare that the income inequality is a world wide phenomenon, prevalent in some and seldom in others.
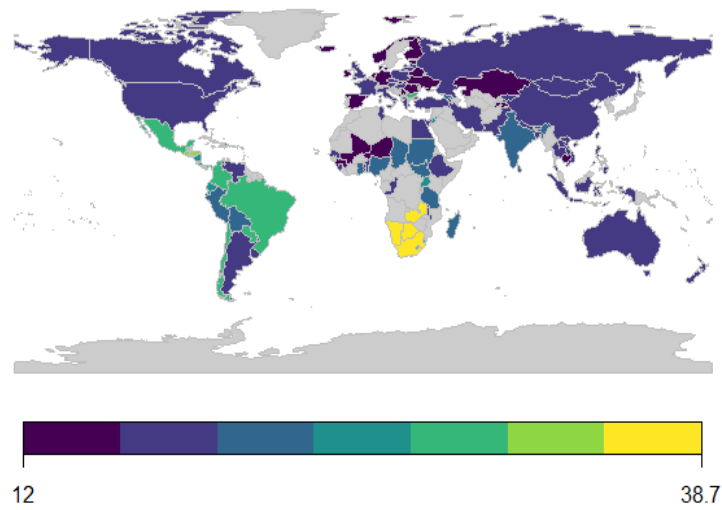
References
[1] https://inequality.org/facts/income-inequality/
[2] https://en.wikipedia.org/wiki/Income_inequality_in_the_United_States
[3] https://ourworldindata.org/global-economic-inequality

6   Appendix

Percentage of Income earned by Bottom 5%

0.107   1.93

Percentage of Income earned by Top 5%

12   38.7

Note: The grey parts indicate that the data is not available for those countries