

Income Inequality?

An Analysis of U.S. Census Bureau Income Data From 1967 to 2019

Abdirizak Yussuf
Mentor: Dr. Xiaoyin Li

St. Cloud State University

eUSR
November 4th, 2022

1.1 Introduction

In the last few decades, the U.S. economy transformed from preindustrial colonies mainly dealing in subsistence farming, to mercantilist merchants competing in the trade of basic commodities, to corporations with massive production capacity able to produce goods for a global market.

Over the years corporations have become more effective in generating great revenue from smaller costs, as evidenced by the ratio between their revenue and costs incurred in producing products and services; still, this increase in scale of production is not reflected in the income of the average person, such as those that are producing these goods and services.

1.2 Problem Statement

I analyze income data from U.S. Census Bureau to see if there is income inequality between income groups; that is, to see if there is a difference in the relative change of income for income groups over the years.

1.4 Why should I care?

Income inequality is associated societal problems such as child mortality, and political instability. Having stated these implications which are detrimental to society, I offer two remedies.

1.3 Packages used in R

- ▶ tidyverse : for organizing and manipulating data
 - ▶ tibble: for turning columns to index
 - ▶ ggplot2: for data visualization
 - ▶ dplyr: for data transformation
 - ▶ readr: for data importing
 - ▶ tidyr: for data tidying
- ▶ CGPfunctions: for making slope graphs
- ▶ reshape2: for restructuring data
- ▶ Ckmeans.1d.dp: for univariate clustering

2.1 About Data and Variables (`income_mean`)

Mean household income received by each fifth and top 5 percent.

▶ Variables

- ▶ `year` - Year
- ▶ `dollar_type` - Dollar relative to that year or normalized to 2019
- ▶ `income_quintile` - Income quintile and top 5%
- ▶ `income_dollars` - Income dollar average

<code>year</code>	<code>income_quintile</code>	<code>income_dollars</code>
2019	Lowest	15286
2019	Second	40652
2019	Middle	68938
2019	Fourth	111112
2019	Highest	254449
2019	Top 5%	451122

2.2 About Data and Variables (`income_distribution`)

Households by total money income, grouped by year and income bracket.

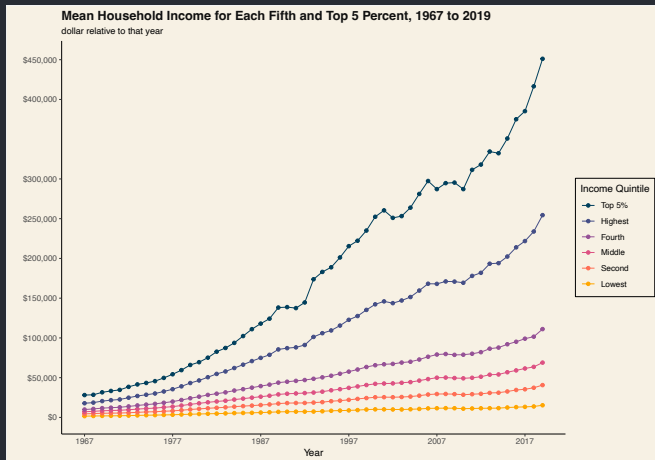
▶ Variables

- ▶ `year` - Year
- ▶ `income_bracket` - Income bracket
- ▶ `income_distribution` - Percentage distribution of household income

<code>year</code>	<code>income_bracket</code>	<code>income_distribution</code>
2019	Under \$15,000	9.1
2019	\$15,000 to \$24,999	8.0
2019	\$25,000 to \$34,999	8.3
2019	\$35,000 to \$49,999	11.7
2019	\$50,000 to \$74,999	16.5
2019	\$75,000 to \$99,999	12.3

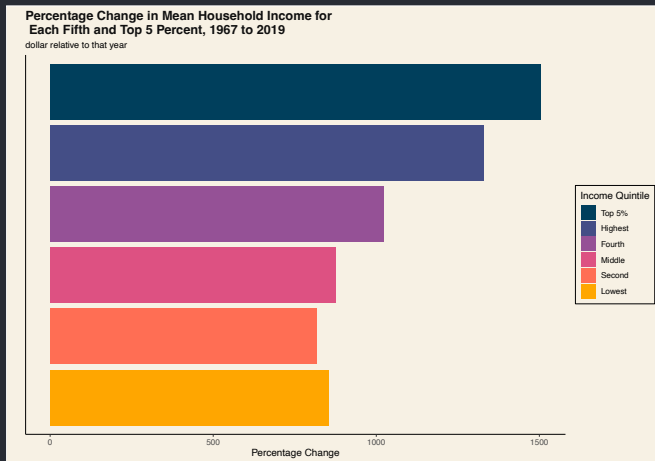
3.1 Mean Household Income Scatter Plot

The mean household income increased for all groups over the whole period. But most noticeable is the drastic increase for the top 5 percent.



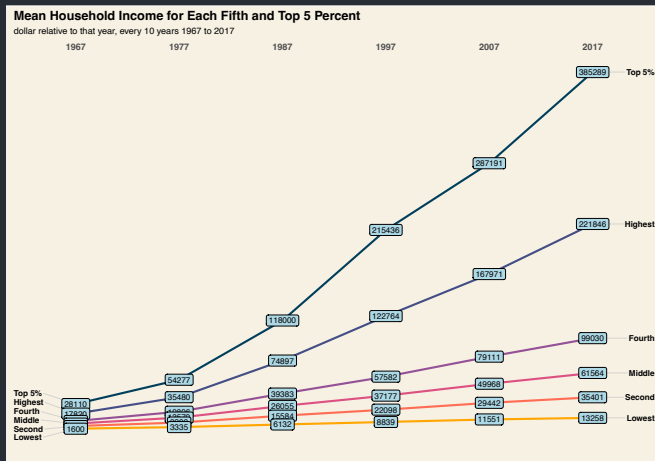
3.2 Percentage Change in Mean Household

The rate of change of income greatly differs for the income groups. Top 5 percent has twice the percentage increase than lowest quintile.



3.3 Slope Graph of Mean Household Income

Income groups are in one cluster in 1967. Over time, the difference in rate of change becomes apparent.



3.4 Analysis of Variance Model

- ▶ Null hypothesis: The mean `income_dollars` is equal for the 5 segments.
- ▶ Alternate hypothesis: At least one of the 5 segments is significantly different from the others.

With ($F = 71.93$, $P < 0.001$), we reject the null and conclude at least one of the 5 segments is significantly different from the others.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
<code>income_quintile</code>	4	326222713733	81555678433	71.9319	0
Residuals	260	294785415471	1133790060	NA	NA

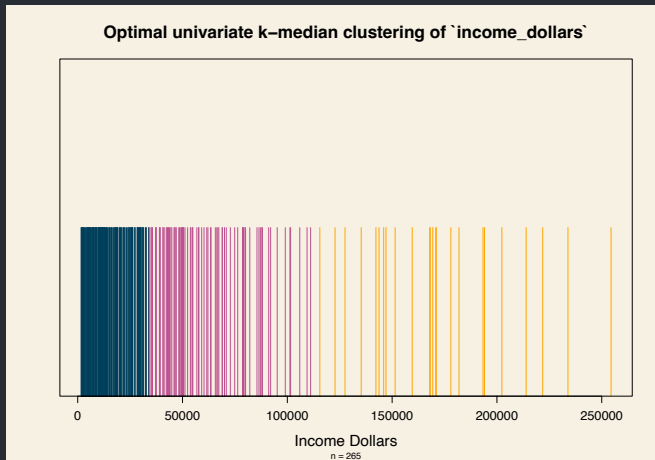
3.5 Tukey Pairwise Comparison

We see that only 3 pairs are not statistically different ($p < 0.05$); Second-Lowest, Second-Middle, and Middle-Fourth being on the cusp. Notice the difference between Lowest-Highest, as was obvious from the plot, Highest income quintile differs greatly from the others.

	income_quintile.diff	income_quintile.lwr	income_quintile.upr	income_quintile.p adj
Highest-Fourth	56410.08	38441.866	74378.2845	0.0000000
Lowest-Fourth	-43639.49	-61607.700	-25671.2815	0.0000000
Middle-Fourth	-18396.09	-36364.303	-427.8853	0.0418534
Second-Fourth	-31781.92	-49750.134	-13813.7155	0.0000201
Lowest-Highest	-100049.57	-118017.775	-82081.3570	0.0000000
Middle-Highest	-74806.17	-92774.379	-56837.9608	0.0000000
Second-Highest	-88192.00	-106160.209	-70223.7909	0.0000000
Middle-Lowest	25243.40	7275.187	43211.6053	0.0013430
Second-Lowest	11857.57	-6110.643	29825.7751	0.3683690
Second-Middle	-13385.83	-31354.039	4582.3789	0.2469747

5.1 Univariate Clustering

I use k-median clustering to partition observations into 3 clusters based on median `income_dollars`.



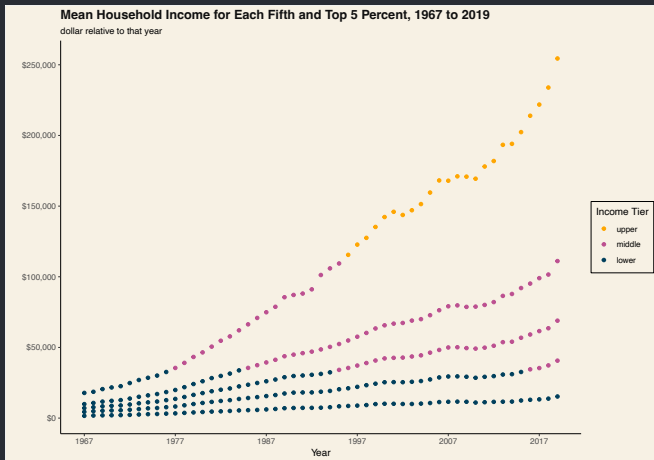
5.2 Income Tiers

$$\begin{array}{l} \text{Lower} \left\{ \begin{array}{l} \text{Under } \$15,000 \\ \$15,000 \text{ to } \$24,999 \\ \$25,000 \text{ to } \$34,999 \end{array} \right. \quad \text{Middle} \left\{ \begin{array}{l} \$35,000 \text{ to } \$49,999 \\ \$50,000 \text{ to } \$74,999 \\ \$75,000 \text{ to } \$99,999 \\ \$100,000 \text{ to } \$149,999 \end{array} \right. \quad \text{Upper} \left\{ \begin{array}{l} \$150,000 \text{ to } \$199,999 \\ \$200,000 \text{ and over} \end{array} \right. \end{array}$$

This result conforms with policy institutes who define income tiers:

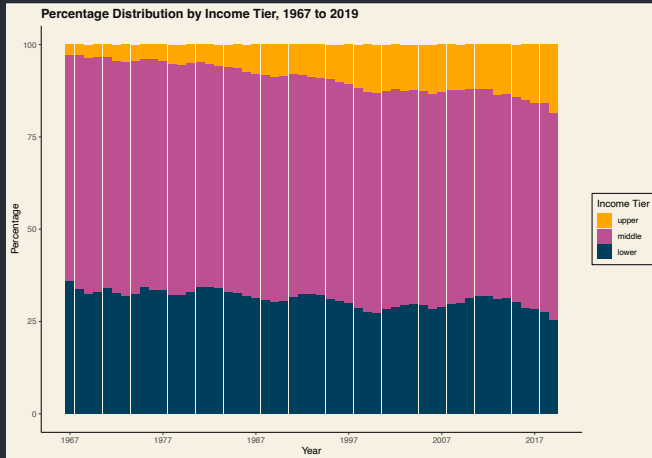
- ▶ Lower - less than $\frac{2}{3}$ the median
- ▶ Middle - $\frac{2}{3}$ to 2 times the median
- ▶ Upper - more than 2 times the median

5.3 Income Tiers



5.4 Percentage Distribution

There is only slight increase in the distribution of upper income tier, but the lower income appears to be about the same.



6. Conclusion

It is apparent there is a differences in income for the income groups, the rate of change of the highest income group is more rapid than others.

I offer two remedies:

- ▶ 1. Investing in schooling, especially for younger children to promote economic mobility.
- ▶ 2. Lower interest rate on home mortgage to not only create economic opportunity, but promote intergenerational wealth.

Limitation

We do not have count for the various groups in `income_bracket`, and `income_quintile`, for this reason we cannot make a conclusion on income stagnation. Even then, we would need unique identifiers for count, otherwise we cannot differentiate if, for example, equal count of people joined middle income tier as those that left it.

Acknowledgment

Thank you,

Dr. Xiaoyin Li!!