Exploring Income Inequality in US after and before financial crises: Analysis*

Insights from 2002 to 2019

Dong Jun Yoon

April 18, 2024

This paper investigates income inequality in the United States from 2002 to 2019 using data from the United States Census and the World Inequality Database (WID). Throughout the paper, we will explore trends in income inequality, examining whether it is increasing or decreasing over time, and how the Great Recession of 2007-2009 affected income inequality. We will utilize data analysis and regression modeling to investigate these trends. Investigating these trends is significant, as it could assist policymakers or the government in developing strategies that could decrease the inequalities occurring within the country. As a result, this report may be used to help promote social and economic equity in the United States.

1 Introduction

Income inequality is a pressing issue not only in the United States but also globally. It pertains to the uneven distribution of income across a population, spotlighting disparities in financial well-being. Unlike wealth, which accounts for assets such as homes and stocks, income inequality focuses on earnings from various sources—wages, salaries, interest from savings, stock dividends, rental income, and profits from selling assets at a higher price than their purchase cost. This distinction is crucial for understanding the different facets of economic inequality. Therefore understanding inequality may provide insights for policymakers or researchers in contributing the problem.

This paper aims to investigate the dynamics of income inequality in the United States, especially focusing on the period from 2002 to 2019. Using the dataset provided by United States Census and World Inequality Database (WID), we will fully investigate income trends across the social class, top 1%, top 10% and bottom 50% of income inequality. Moreover, will be

*Code and data are available at: https://github.com/DJY1231/Exploring-Income-Inequality-in-US.git

considering variables such as year, social class, and race to find out the correlation between these factors.

Even though researchers has studied on income inequality in the United States, there are still remaining income gap between individual firms who are living in the United States. To efficiently investigate this income gap, our paper would focus on exploring datasets and multiple linear regression modeling to find out the full details (trends, patterns) about income inequality that is occurring in the United States.

The paper is structured as following: Section 2 introduces the data used for analysis and findings, including the variables of interest. Section 3 introduces the linear regression model to examine the correlation between time (year) and social class to investigate income inequality in the United States. Section 4 introduces the full analysis of the table which will include the summary. Section 5 may discuss about the findings that we investigated and also talk about the weaknesses of this paper and its next steps.

1.1 Estimand

The study's goal is to evaluate the relationship between time (year) and income inequality in the United States, focusing on how changes over the years affect income disparity. It aims to assess the impact of each year on income inequality, considering other constant variables, and to examine trends from 2002 to 2019.

Additionally, the study explores the influence of income inequality and its interaction with time, race, and social class. By examining the coefficients such as year and income in our regression model, we aim to understand the differences in income inequality between different race and social class, and see how it changes over time.

Thus revealing the complex interplay between time, race, social class, and income disparity in the United States.

2 Data

2.1 Data Source

The first data is gathered from World Inequality Database (WID) and second data is gathered from the United States Census Bureau, where the first dataset introduces the income share by each social class over time: 2002 to 2019 in the United States. The second data illustrates the median household income based on race and year.

2.2 Data Measurement

The dataset provides information that offers insights into the income trends in the United States. The y-axis represents the income share, measured in dollars (%). The x-axis corresponds to the years from 2002 to 2019, representing the total income for each year.

Data cleaning and analysis were conducted using the base statistical programming language R (R Core Team 2023), and the core components such as tidyverse (Wickham et al. 2019), ggplot2 (Wickham 2016), dplyr (Wickham et al. 2023), readr (Wickham, Hester, and Bryan 2024), tibble (Müller and Wickham 2023),stringr (Wickham 2023), haven (Wickham, Miller, and Smith 2023), janitor (Firke 2023), knitr (Xie 2023).

2.3 Variables of Interest

Table 1: First 40 rows of Income Share by Social Class Over Time: 2002 to 2019 in the United States.

Country	Social_Class	Year	Share_of_total_income
USA	Top 1% income	2002	0.1610
USA	Top 1% income	2003	0.1632
USA	Top 1% income	2004	0.1706
USA	Top 1% income	2005	0.1806
USA	Top 1% income	2006	0.1852
USA	Top 1% income	2007	0.1838
USA	Top 1% income	2008	0.1793
USA	Top 1% income	2009	0.1670
USA	Top 1% income	2010	0.1789
USA	Top 1% income	2011	0.1808
USA	Top 1% income	2012	0.1949
USA	Top 1% income	2013	0.1846
USA	Top 1% income	2014	0.1897
USA	Top 1% income	2015	0.1889
USA	Top 1% income	2016	0.1866
USA	Top 1% income	2017	0.1905
USA	Top 1% income	2018	0.1924
USA	Top 1% income	2019	0.1904
USA	Bottom 50% income	2002	0.1476
USA	Bottom 50% income	2003	0.1446
USA	Bottom 50% income	2004	0.1416
USA	Bottom 50% income	2005	0.1372
USA	Bottom 50% income	2006	0.1349
USA	Bottom 50% income	2007	0.1371

Country	Social_Class	Year	Share_of_total_income
USA	Bottom 50% income	2008	0.1369
USA	Bottom 50% income	2009	0.1364
USA	Bottom 50% income	2010	0.1328
USA	Bottom 50% income	2011	0.1297
USA	Bottom 50% income	2012	0.1259
USA	Bottom 50% income	2013	0.1288
USA	Bottom 50% income	2014	0.1258
USA	Bottom 50% income	2015	0.1265
USA	Bottom 50% income	2016	0.1239
USA	Bottom 50% income	2017	0.1291
USA	Bottom 50% income	2018	0.1277
USA	Bottom 50% income	2019	0.1301
USA	Top 10% income	2002	0.4149
USA	Top 10% income	2003	0.4164
USA	Top 10% income	2004	0.4242
USA	Top 10% income	2005	0.4359

Table 1: Extracting the first 40 rows of Income Share by social class over time: 2002 to 2019 in the United States.

1. Location: The model uses all the dataset collected across the country of the United States. This geographic coverage enables us to access different social class with different income level.

2. Social Class: The dataset contains information on income levels for each social class, including Top 1% income, Bottom 50% income, and Top 10% income. These income levels provide detailed insights into income distribution across different years and shares of total income in the United States.

3. Year: The dataset displays years from 2002 to 2021, covering a range that allows for the exploration of long-term income trends and changes over time. This period aids in analyzing income dynamics before and after the Great Recession, offering insights into patterns of income distribution.

4. Share of total income: The dataset likely represent the share of total income that the specified social class (e.g., "Top 1% income") earns in the United States for the given year. For instance, a value of 0.1610 for "Top 1% income" in the USA for 2002 could mean that the top 1% of earners received 16.10% of the total income generated in the USA that year.This data can be used to analyze trends in income distribution and inequality within the United States over time, examining how the income shares of different segments (like the top 1%, top 10%, and bottom 50%) of the population change.

2.4 Data Visualization

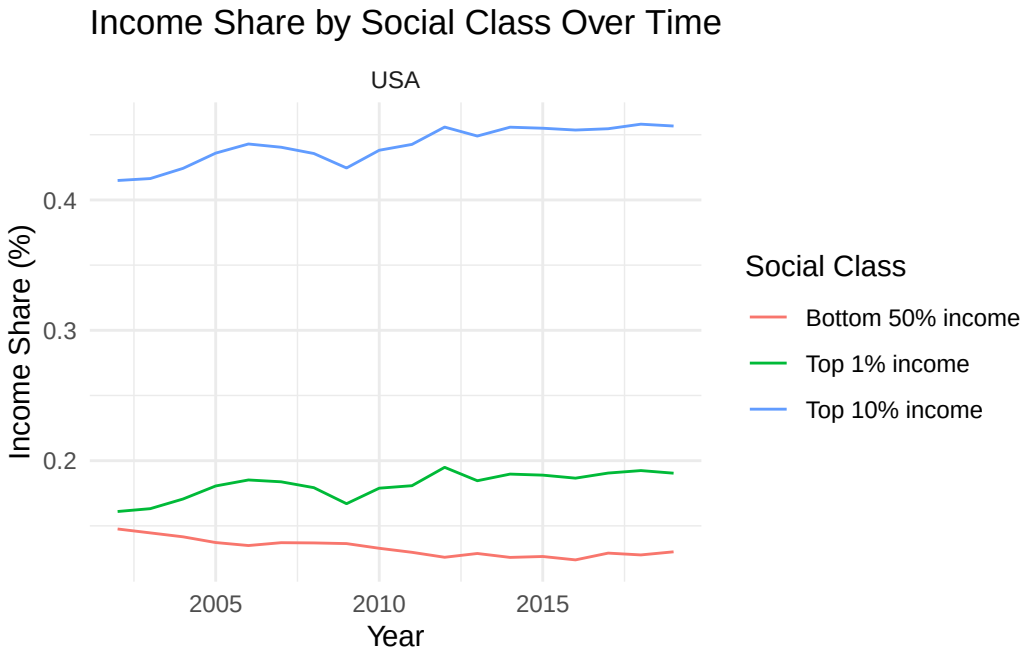

Figure 1: Income Share by Social Class Over Time: 2002 to 2019 in the United States.

Figure 1 illustrates the changing dynamics of income distribution across different social classes in the United States from 2002 to 2019. Specifically, the data show an upward trend in the income share for the top 1% and top 10% of earners, suggesting that these groups have been capturing a progressively larger portion of the total income pie over the examined period. In contrast, the bottom 50% of earners experienced a decline in their income share, indicating that their relative economic position has decreased during the same time frame. The consistent increase in income share for the top 1% and top 10% suggests a concentration of wealth at the top, a classic indicator of growing income inequality. This trend raises concerns about the economic well-being of the majority, as the bottom 50%'s shrinking share implies that a significant portion of the population is benefiting less from overall economic growth.

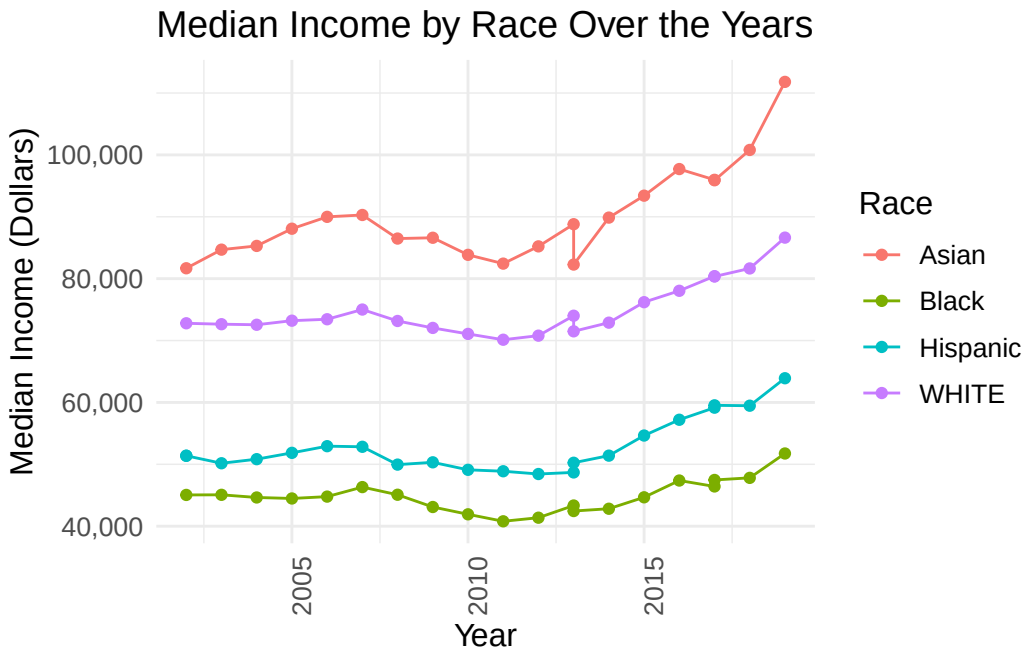

Figure 2: Trend of Median Income by Race over the year: 2002 to 2019 in the United States.

The second set of data, sourced from the United States Census Bureau, displays the median household income segmented by race and year. As depicted in Figure 2, the trend of median income by race from 2002 to 2019 generally shows an upward trajectory. However, starting from 2007, a downward trend is observable across all racial groups, likely influenced by the financial crisis spanning December 2007 to June 2009. This period witnessed a marked decrease in median income across most racial categories, extending over a five-year span or longer. Notably, Asian and White households recovered more rapidly than their Hispanic and Black counterparts, indicating a potential exacerbation of income inequality among races.

For instance, between 2007 and 2019, the median income for Asians increased by 23.8%, from \$90,290 to \$111,800. In contrast, Black households saw a more modest increase of 11.7%, from \$46,320 to \$51,750. Examining the five years post-recession (2007 to 2011), Asian income decreased by 8.7%, whereas Black income declined by 11.9%. This comparative analysis reveals a pronounced disparity: Asian households experienced a smaller reduction in income post-recession and achieved a quicker and greater recovery compared to Black households.

Such trends underscore a notable income inequality between racial groups, influenced by various factors including workplace and environmental contexts. While individual circumstances vary, the overarching data suggests persistent disparities in income recovery and growth among different races, highlighting a critical area for policy intervention and societal attention.

This information provides additional insight into income inequality among different racial groups. While Figure 1 focuses solely on social classes, this could offer a more detailed analysis of income disparities.

3 Model

In this section, we will discuss the multiple linear regression model that will be used in the analysis.

3.1 Model set-up

The multiple linear regression model is formulated as follows:

$$

$$\begin{aligned}\text{Share\_of\_total\_income}(Y) = \beta_0 + \beta_1 \cdot \text{Year} + \beta_2 \cdot \text{Social\_Class} \\ + \beta_3 \cdot \text{Period} + \beta_4 \cdot (\text{Year} \times \text{Social\_Class}) \\ + \beta_5 \cdot (\text{Social\_Class} \times \text{Period}) + \varepsilon\end{aligned}$$

$$

We run the model in R (R Core Team 2023) using the rstanarm package of Goodrich et al. (2022). We use the default priors from rstanarm.

3.1.1 Model justification

We expect a capture the relationship between the response variable, Y, representing the share of income based on each social class. - β_0 is the overall intercept. - β_1 represents the trend of income share over time. - β_2 represents the differences in income share across different social classes. - β_3 represents the impact of the economic period on income share. - β_4 and β_5 are interaction terms that capture the nuanced effects of time and economic periods on different social classes. - ε is the error term.

The choice of priors for the coefficients, β_0, β_1, and β_3, β_4, β_5 as Normal distributions with mean 0 and a standard deviation of 2.5 reflects a weak prior assumption. This indicates that there are no strong preconceptions regarding the size or direction of their impacts.

4 Results

The regression results are as following:

```
Loading required package: stargazer
```

```
Please cite as:

 Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

 R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

Loading required package: performance
```

	Dependent variable:
	Share_of_total_income
Year	-0.001
	(0.001)
Social_ClassTop 1% income	-4.644***
	(1.594)
Social_ClassTop 10% income	-5.830***
	(1.594)
PeriodPost-Recession	-0.005
	(0.006)
PeriodPre-Recession	0.001
	(0.005)
Year:Social_ClassTop 1% income	0.002***
	(0.001)
Year:Social_ClassTop 10% income	0.003***
	(0.001)
Social_ClassTop 1% income:PeriodPost-Recession	0.009
	(0.008)
Social_ClassTop 10% income:PeriodPost-Recession	0.012
	(0.008)

Social_ClassTop 1% income:PeriodPre-Recession	0.006
	(0.007)
Social_ClassTop 10% income:PeriodPre-Recession	0.007
	(0.007)
Constant	1.364
	(1.127)
Num. Obs.	54
RMSE	0.005
AIC	-394
BIC	-368
Observations	54
R2	0.999
Adjusted R2	0.998
Residual Std. Error	0.006 (df = 42)
F Statistic	2,863.266*** (df = 11; 42)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2: Summary results for the share of total income model

This table summarize the output of a regression model that estimates the share of total income based on different predictors: year, social class (top 1% income and top 10% income), periods (pre- and post-recession), and interaction terms between these predictors.

Social Class (Top 1% income and Top 10% income): Both coefficients are negative and statistically significant (as indicated by the three asterisks, typically denoting $p < 0.001$). This suggests that compared to the reference social class (presumably the rest of the population), the top 1% and the top 10% have a significantly lower share of the total income.

PeriodPost-Recession and PeriodPre-Recession: The coefficients for both are very close to zero and are likely not statistically significant, indicating no change in the share of total income during these periods compared to the reference period (possibly the recession period).

Interaction terms: The positive and significant interaction terms for Year:Social_ClassTop 1% income and Year:Social_ClassTop 10% income suggest that the share of total income for these social classes is increasing annually at a rate faster than the rest of the population.

Constant: The constant term is positive but not statistically significant, suggesting that for the base category (which is not detailed in the summary), the share of total income is not significantly different from zero.

5 Discussion

5.1 Findings

This paper examined income inequality in United States from 2002 to 2019, especially analyzing the income inequality before and after the great recession (2007-2009). Throughout the data analysis and regression modeling, we found that income inequality has potentially increased over time and the income gap between each social classes has also increased.

Comparing income inequality trends pre and post recession:

Before the recession: the gap was widening, as evidenced by significant gains for the top earners and losses for the bottom 50%.

After the Recession: The trend of widening inequality continued and possibly accelerated, with the top groups seeing substantial increases in their income shares while the bottom 50% experienced a decline.

Throughout the discussion part, we will now provide detailed points of our results in the following sections.

Additionally we could observe income inequality between different races. In result, Asian and White tend to have higher income comparison to Black and Hispanic.

5.2 Comparing income inequality trends pre, during, and post recession in the United States:

Based on the analysis of Figure 1, we can find out the details about the changing dynamics of income distribution across different social classes in the United States from 2002 to 2019. Specifically, we can examine the trends that has occurred in income inequality before, during, and after the great recession in the United States. The data shows an upward trend in the income share for the top 1% and top 10% of earners, and a downward trend for the bottom 50%. This indicates that the top 1% and the top 10% group has more portion of the total income pie over the examined period and the bottom 50% group has less portion of the total income pie over the same time frame.

This trend could be related to the financial crisis, where the economic recession, officially beginning in December 2007 and ending in June 2009. During this period, many businesses faced significant challenges, including reduced consumer spending, tightened credit conditions, and a decline in investments. During the recession, there was significant amount of job losses where the loss of jobs and income contributed to the decrease in personal income.

To assess the impact of the Great Recession, which spanned from late 2007 to mid-2009, we can divide the analysis into three phases: Pre-Recession (2002-2007), during the Recession (2007-2009), and Post-Recession (2009-2019). By comparing these periods, we aim to understand

how this significant economic downturn influenced the distribution of income among these groups.

Pre-Recession (2002-2007): During this period, the economy was generally expanding. The data show that the income share for the top earners (1% and 10%) was already on the rise, this indicates that the trends towards inequality began before the economic downturn. The income share of the top 1% increased by 14.2% during this time period, from 0.1610 to 0.1838. Furthermore, the income share of the top 10% has increased by 6.14% during this time period, from 0.4149 to 0.4404. On the other hand where the income share for the bottom 50% was on a decrease by 7.11%, from 0.1476 to 0.1371. This suggest that this group of people are less benefiting from overall economic growth.

During the Recession (2007-2009): Typically, recessions can disproportionately affect lower-income groups, but the specific impacts can vary. In this case, the income share of the top 1% decreased by 9.14% during the recession, from 0.1838 to 0.1670. This significant reduction suggests that the highest earners were not immune to the economic downturn, likely reflecting declines in investment income, real estate values, and high-paying jobs, which are common income sources for this group. The top 10% experienced a 3.61% decrease in their income share, from 0.4404 to 0.4245. While this group also felt the effects of the recession, the impact was less severe compared to the top 1%. This difference could be due to a broader diversification of income sources within this group or a slightly more resilient income structure compared to the very top earners. The bottom 50% saw a relatively minor decrease in their income share of 0.51%, from 0.1371 to 0.1364. This smaller change could indicate that the income sources for this group, often more reliant on wages rather than investments, did not fluctuate as dramatically as those for higher income groups. However, it's important to note that for individuals with lower income, even a small decrease can have a substantial impact on their living standards. The more significant decrease in income share for the top 1% temporarily narrowed the income inequality gap. However, the critical question for understanding long-term trends is how these groups' income shares rebounded in the post-recession recovery period.

Post-Recession (2009-2019): The recovery phase is crucial for understanding whether the income distribution trends were a temporary response to the recession or part of a longer-term pattern. The income share of the top 1% earners continued to grow from 0.1670 to 0.1904, which represents a 14% increase in total income share. The top 10% earners also observed their income share increase from 0.4245 to 0.4567, a 7.58% rise in total income share. However, the bottom 50% experienced a decrease in their income share by 4.62%, from 0.1364 to 0.1301. While the income shares of top earners continued to grow, the bottom 50%'s share did not recover or even continued to fall, indicating a worsening of income inequality post-recession.The top 1% and top 10% of earners saw their share of total income increase significantly between 2009 and 2019. The top 1% experienced a 14% increase, and the top 10% saw a 7.58% rise. This growth suggests that these groups were able to recover from the recession and even enhance their economic standing, likely benefiting from the recovery in investment markets, real estate, and high-income employment sectors. However, the bottom 50% saw their share of total income decrease by 4.62%. This decline indicates that this group

did not benefit from the economic recovery to the same extent as the higher income groups. In conclusion, increased income share for the top earners and decreased share for the bottom half suggest that income inequality worsened in the post-recession period. Where it indicates that the gap between the rich and the poor got widened.

In conclusion, the depicted trends in Figure 1 suggest that income inequality has been widening over the years, with significant gains concentrated among the top earners and losses by the bottom half of the distribution. Comparing income inequality trends pre and post recession: Before the recession: the gap was widening, as evidenced by significant gains for the top earners and losses for the bottom 50%. After the Recession: The trend of widening inequality continued and possibly accelerated, with the top groups seeing substantial increases in their income shares while the bottom 50% experienced a decline. By comparing these trends, it's clear that the gap between the rich and the poor has indeed grown larger post-recession. The top earners not only recovered their losses but also gained a larger portion of the income share, while the bottom half saw a continual decline.

5.3 Asian and White are likely to earn more income compared to Black and Hispanic people.

Based on the analysis of Figure 2, we can find out the details about the trend of median income by race over the year: 2002 to 2019 in the United States. Specifically, we can examine the income inequality among different racial groups that has occurred in the period of before, during, and after the great recession in the United States. As similar to figure 1, the income has increased before the great recession, however decrease during the recession, and covered after the recession over time.

In overall, figure 2 traces the median household income by race from 2002 to 2019, showing an overall increase but with a dip during the 2007 financial crisis, affecting all racial groups for at least five years. Asian and White households bounced back quicker than Hispanic and Black ones, suggesting growing racial income inequality. From 2007 to 2019, Asian median income rose by 23.8%, while Black households saw an 11.7% increase. Post-recession, Asian income fell by 8.7%, and Black income by 11.9%, revealing Asian households had a smaller decline and a faster recovery than Black households. These trends indicate persistent income disparities across racial lines, pointing to the need for targeted policy measures and societal engagement to address these inequalities.

5.4 Weaknesses and next steps

Multiple linear regression show weaknesses in these factors: 1. Model Complexity: Including multiple interaction terms can lead to a model that is too complex, potentially over fitting the data. This can make the model sensitive to small variations in the data and can reduce its predictive performance on new data. 2. Sample Size: With multiple linear models, having

a smaller sample size can be particularly problematic because the model's estimates can be heavily influenced by the priors. This could result in less reliable posterior distributions if the data isn't sufficient to overwhelm the prior beliefs.

Within the overall income analysis, there could be potential biases that could occur. For example for sample size for our model, the data may still have unobserved factors that may influence income inequality. These unobserved information could potentially change the estimates and results we have goten.

6 References

R Core Team. 2023. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Xie, Yihui. 2023. Knitr: A General-Purpose Package for Dynamic Report Generation in r. https://yihui.org/knitr/.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. Readr: Read Rectangular Text Data. 27. https://readr.tidyverse.org.

Wickham, Hadley. 2016. Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. Dplyr: A Grammar of Data Manipulation. https://dplyr.tidyverse.org.

Wickham, Hadley, and Dana Seidel. 2022. scales: Scale Functions for Visualization. https://CRAN.R-project.org/package=scales.

Bureau, U. C. (2023, September 12). Historical income tables: Income inequality. Census.gov. https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-income-inequality.html

Data - wid - world inequality database. WID. (n.d.). https://wid.world/data/