11-2022

Teaching Income Inequality with Data-Driven Visualization

Humberto Barreto
DePauw University, hbarreto@depauw.edu

Sang T. Truong
Stanford University, sttruong@cs.stanford.edu

Recommended Citation

Truong, S. T., & Barreto, H. (2023). Teaching Income Inequality with Data-Driven Visualization. The American Economist, 68(1), 140–155. https://doi.org/10.1177/05694345221129022

Teaching Income Inequality with Data-driven Visualization

Sang T. Truong[1,2,3], **Humberto Barreto**[3]
sttruong@cs.stanford.edu, hbarreto@depauw.edu
[1]Department of Computer Science, Stanford University
[2]Graduate School of Business, Stanford University
[3]Department of Economics and Management, DePauw University

Abstract

The distribution of household income is a central concern in economics due to its strong influence on society's well-being and social cohesion. Yet, non-expert audiences face serious obstacles in understanding conventional measures of inequality. To effectively communicate the extent of income inequality in the United States, we have developed a novel technique for visualizing income distribution and its dispersion over time by using U.S. household income microdata from the Current Population Survey. The result is a striking dynamic animation of income distribution over time, drawing public attention and encouraging further investigation of income inequality. Detailed implementation is available at `github.com/sangttruong/incomevis`. An interactive demonstration of our project is available at `research.depauw.edu/econ/incomevis`.

1 Introduction

According to the most recent quinquennial survey of teaching methods in both introductory courses (Asarta et al. 2021) and across the curriculum (Harter and Asarta 2022), economics pedagogy remains largely unchanged: "the dominance of instructor-centric methods is compelling, and the continued lack of teaching methods that address diversity, inclusion, and gender issues is striking" (Asarta et al. 2021, page 26). This paper offers data and graphical resources to support courses that cover income distribution and inequality in the United States. Our contribution is suitable for use across the curriculum, from introductory to more advanced courses.

Income inequality has emerged as a critical issue in mainstream discussions of economics and politics (Piketty 2014; Heathcote et al. 2010; Alderson et al. 2005). Non-experts are aware that income is not evenly distributed, but they frequently underestimate magnitudes and changes in income inequality (Norton 2011). Conventional means of measuring, communicating, and teaching about income distribution make it difficult to understand income inequality. The most common approach is via the Gini coefficient (G) (Dorfman 1979; Hartmann et al. 2017). For a discrete income distribution with n entities, the Gini coefficient is expressed as:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2\sum_{i=1}^{n}\sum_{j=1}^{n}x_j} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n^2\bar{x}} \tag{1}$$

where x_i is the income of entity i and $\bar{x}$ is the average income. Since G depends on the area between Lorenz curve and a diagonal line (Figure 9 in Supplementary Material), G generates a value between 0 and 1 that varies non-linearly: $G = 0$ indicates a perfectly egalitarian distribution where everyone has the same income, and $G = 1$ indicates one person receives all of the income. Unlike the conventional interpretation of other common metrics with a linear scale, $G = 0.5$ does not indicate an "average" degree of income inequality but rather a high one. In addition, a Gini decrease of 0.1 from $G = 0.3$

is not the same if it occurs from 0.6. While the two extreme values of Gini coefficient are relatively easy for non-experts to understand, the values in between are more difficult to interpret.

To overcome the obstacle of communicating information about how income is distributed to non-experts, we propose a framework that visualizes income inequality over time to audiences without relying on mathematical knowledge. Our visualization conveys information about income inequality over time without advanced prerequisites, such as familiarity with the Gini coefficient. We leverage rich sources of microdata as well as modern visualization techniques to produce graphs that convey information about the distribution of household income in the United States.

2 Related Works

The inspiration for our visualization is (Sutcliffe 2001) whose 3D chart of global income inequality is reproduced in Figure 1. For a more recent rendering, see (The Core ECON Project 2020). The 3D chart provides an easy way to make connections, such as the fact that Chinese income inequality has risen as the country itself has gotten richer (Han 2011). We use this same 3D chart type with U.S. data, displaying income deciles for each state (and the District of Columbia).

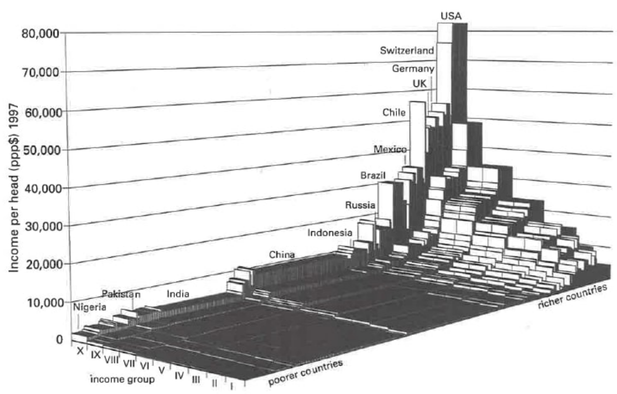


Figure 1: (Sutcliffe 2001) shows the world distribution of income as a city, where the height of each block reflects income deciles and population of each country determines the width. Most prominently, in the far corner, the "skyscraper plot" represents the income of the wealthiest 10^{th} percent of the world population (Sutcliffe 2001, p. 17). The visualization excels at displaying information about income inequality not only across different countries, but also within a specific country.

3 Methodology

To create our visualization, we need to be able to (1) meaningfully compare household income, (2) construct an empirical distribution of income for a given state, and (3) compare these income distributions across states. We explain how to achieve these goals in the sections below.

3.1 Data Sources

Our data are collected from Integrated Public Use Microdata Series - Current Population Survey (IPUMS-CPS), Integrated Public Use Microdata Series - United States of America (IPUMS USA), and the Bureau of Economic Analysis (BEA), from 1977 to 2020 (Flood et al. 2020; BEA 2020; Ruggles et al. 2020). Table 1 (Supplementary Material) presents summary statistics for the data. Each variable has a total of 2.82 million observations in the studied period. We begin our analysis with the 1977 survey because this is the first year that the CPS included a state geographical variable. Prior to this, some states were grouped together. Since CPS income data collected in year t is for the previous year, $t-1$, our analysis is for household income in the 1976-2019 period.

3.2 Adjustments to Income

Before comparing household incomes of two families, we want to make sure that our comparison is sensible. For instance, because of inflation, it does not make much sense to compare nominal income 50 years ago and nominal income today. As another example, since living in California is much more expensive than living in Indiana, it is also not sensible to compare nominal income of a family in California with the one in Indiana. We employ various "normalizers" to adjust the income data before comparing them.

There are many ways to adjust household income, but we address three of them: inflation over time, price differences across states, and household size. There are many other adjustments we could include, such as taxes and transfers. Unfortunately, the information needed is frequently unreported and there is no consensus on exactly which taxes and transfers should be included. Our three adjustments enable a base-level comparison of income for a standardized household size across states and time.

Adjusting for Inflation over Time Adjustment for inflation is crucial for proper comparison of income over time. We use `CPI99`, provided by IPUMS, to hold the price level constant: "`CPI99` provides the CPI-U multiplier (available from the Bureau of Labor Statistics) to convert dollar figures to constant 1999 dollars (This corresponds to the dollar amounts in the 2000 CPS, which inquired about income in 1999)" [1]. The CPI, C_t^0, gives a measure of the overall price level each year. At time period t, C_t^0 is the ratio of the market basket price m_t to the base period m_0: $C_t^0 = m_t/m_0$. Using the CPI to compute real household income allows us to adjust nominal household income H for inflation. We also convert the base period of the IPUMS-CPS CPI variable from 1999 to survey year 2020, which is 2019 dollars: $C_t^{2019} = m_t/m_{2019} = (m_t/m_{1999})/(m_{2019}/m_{1999}) = C_t^{1999}/C_{2019}^{1999}$.

Adjusting for Difference Price Level across States We adjust for differences in price levels across states using the regional price level, R, from the BEA (BEA 2020). This allows meaningful comparison of incomes across states with varying price levels. For example, incomes are much higher in California than Alabama, but it is much more expensive to live in California. Although the BEA database is the most complete source for regional price parity in the U.S. that we have access to, it is only available annually from 2008. One way to deal with data unavailability is to ignore about 40 years worth of data before 2008 (about 80% of our data), implicitly assuming equal price levels across states. Another way is to fill in the missing data by an "educated guess" (e.g. using a statistical model to infer missing data). We recognize that this is an imperfect solution, but believe that the informational benefits of adding 40 years worth of data exceed the informational costs derived from their inaccuracy.

Housing cost is the key variable in variation in state cost of living. We assume prices of all other goods the same across states and set housing cost (including utilities) as 44% of a household's budget: "for families at the 35th percentile of the distribution of spending on food, housing and clothing, housing represented 44 percent of total expenditures assuming miscellaneous expenditures are set at 15 percent of the food, housing and clothing amount" (Renwick 2011, page 4). Therefore, we employ the following regression model to fill in the missing data:

$$R = \alpha + \beta_1 r + \beta_2 F(r) + \beta_3 F(R) + \beta_{4,i} FE_i + \epsilon \tag{2}$$

where $F(\cdot)$ is the forward-shift operator (i.e. lead operator); FE_i is the binary fixed effect for every state i; and r is the gross rent. A panel model with additional explanatory variable (instead of just a time-series model with only regional price level) is employed because the time series of regional price level is completely unobserved before 2008 and the series does not have simple behavior (such as linear trend). Hence, traditional methods for dealing with missing data in univariate time-series, such as linear or spline interpolation, is not suitable here (Friedman 1962).

Our explanatory variable, gross rent, also has a missing value issue. Indeed, before 2000, we only have access to gross rent data once every 10 years. However, the issue of missing data in gross rent is much easier to work with relative to regional price because of two reasons. First, the data are not completely missing before 2000, instead, we just observe data less frequently. Second, the gross rent series is relatively well-behaved with linear trend (see Figure 10a in Supplementary Material). Therefore spline interpolation for univariate time-series is suitable for missing data in this situation.

[1] See `cps.ipums.org/cps-action/variables/CPI99` for more information

Data about actual and predicted regional price parity of some example states is shown in Figure 10b (Supplementary Material). We observe a moderate growth in regional price level in all states. The value of fixed-effect coefficients for each state is presented in Figure 10c (Supplementary Material). The result agrees with our intuition, as CA, DC, and NY are the states with the highest costs of living.

Adjusting for Difference Size between Households Household size, S, should also be taken into account since it is more costly to take care of more people. For example, a household comprised of two adults with total income of $100,000 is more affluent than a six-person family with the same income. This adjustment is needed because households have gotten smaller over time and there are differences across states. For example, Figure 10d (Supplementary Material) shows that effective household size is much higher in California than in Washington, DC over the last 50 years. We adjust for household size using the "square root equivalent scale" (Johnson et al. 2005, page 13): "Adjusting resources in this manner yields 'equivalent resources per person,' and provides us with a population of individuals whose resources are given by the equivalent resources of their consumer unit." We do not simply divide H by the total number of household members since some expenses do not scale up linearly. For example, a couple usually does not pay twice the amount of rent of a single person.

Adjusting for Different Age Distribution across States and over Time Controlling for age distribution differences is an example of another way we can normalize household income data. Since age correlates strongly with income, we explored adjusting for different age distributions (Almas and Mogstad 2012; Formby and Seaks 1980; Murphy and Welch 1990). For example, California's population is aging faster than in Washington, DC. In 1976, the mode of age distribution in both areas is around 25, but the mode of CA shifts toward 50 as time progresses (see Figure 11c) while the DC distribution is relatively more stable (see Figure 11d). To adjust for difference in age distribution, we apply bootstrap resampling with constraint that the resulting samples have similar age distribution.

We conduct an ablation study to understand the effect of different combinations or adjustments in California and DC (Figure 12a and 12b in Supplementary Material). Note that $\hat{H}$ is the household income adjusted by inflation, regional price, household size, and age: $\hat{H} = H_{\text{age}}/CRS$. Adjusting for household size, S, has the biggest impact. Age matters, but not a great deal, as shown by the closeness of H/CRS and $\hat{H}$ for both California and DC in 1976 and 2019.

3.3 Income Segmentation

After adjusting household income, we sort and compute deciles and percentiles for each state for each year. We construct an empirical distribution for each state by segmenting the adjusted household income into group B_k from 5 percentile to 95 percentile [2]. To figure out which group a household belongs to, we compute a percentile for each household according to their household weight (which is obtained from IPUMS-CPS). Concretely, for household h in state s with total household H at year y, the household percentile $P_{h,s,y}$ is computed according to its cumulative household weight $w_{i,s,y}$ as:

$$P_{h,s,y} = \frac{\sum_{i=0}^{h} w_{i,s,y}}{\sum_{i=0}^{H} w_{i,s,y}} \tag{3}$$

The adjusted household income is then sorted by ascending decile groups or percentile groups [3]:

$$\begin{aligned} \text{decile: } & k \in \{05\%, 15\%, ..., 45\%, 50\%, 55\%, ..., 95\%\} \\ \text{percentile: } & k \in \{05\%, 06\%, ..., 49\%, 50\%, 51\%, ..., 95\%\} \end{aligned} \tag{4}$$

In other words, given a k percentile, a state s, and a year y, a group $b_{k,s,y}$ is a set of adjusted household income defining as

[2]We exclude the bottom 5 percentiles to avoid negative values of household income since some income sources can be negative, such as business income. We also remove the top 5 percentiles to avoid invalid high-income measurement due to the disclosure avoidance measure of IPUMS-CPS survey (Flood et al. 2020).

[3]The percentile grouping is more fine-grained than the decide grouping, which is helpful in the situation where we need a high-resolution view of the population. We will demonstrate this in a case study later in the paper.

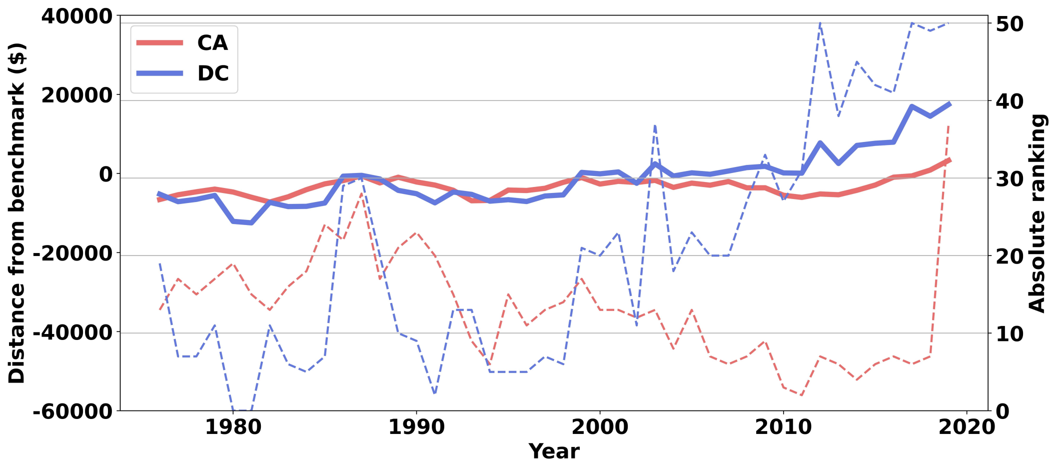


Figure 2: State ranking according to median income(dashed line) and distance to national benchmark (solid line) of CA and DC over time. Ranking according to distance to national benchmark is much more stable over time.

$$b_{k,s,y} = \{\hat{H}_{h,s,y} | k - 1 \leq P_{h,s,y} \leq k\} \tag{5}$$

where $\hat{H}_{h,s,y}$ is the adjusted household income of the given percentile, state, and year. We choose the maximum of the group as its summary statistic.

3.4 Representation of Population Size

In our 3D visualizations, the width of the bar representing state s in year y with h households is proportional to its normalized population $\bar{p}$:

$$\bar{p}_s = \frac{\sum_{i=0}^{h} w_i}{\min(p_s)} \tag{6}$$

where w_i is the weight of household i in the sample.

3.5 Choice of Household Income Benchmark

We need to have an order for states to display them on the horizontal axis (i.e. the$x$-axis) in the 3D plot. One way is to sort the states according to their median adjusted income (using the 50 percentile). As Figure 2 shows, this ranking system is sensitive to the small movement of other states' income. An increase in rank in this system does not necessarily imply economic growth. Another way to order the states is to use their distance to the 2019 national income benchmark. Each state's position on the x-axis is the difference between their median income and the national median income in the final year 2019. This approach is more stable (Figure 2) and the movement along the horizontal axis in our visualization matches better with the economic development.

Figure 3 summarizes our overall data processing framework. Note that although we focus on household income in this study, other variables (e.g. personal income or cumulative wealth) can be analyzed with this workflow as well.

3.6 Sampling Variability

The information communicated through our visualization (For example, in Figure 4) should be seen through a lens that includes sampling variability due to its empirical nature. The height of each block in Figure 4 is a sample estimate of adjusted household income percentile for a state, which has a standard error (SE). We note that the need for SE is not unique to our method, but any that involves data-driven statistics (such as the popular Gini coefficient estimated from the empirical income distribution). We understand that the introduction of sampling variability can confuse readers

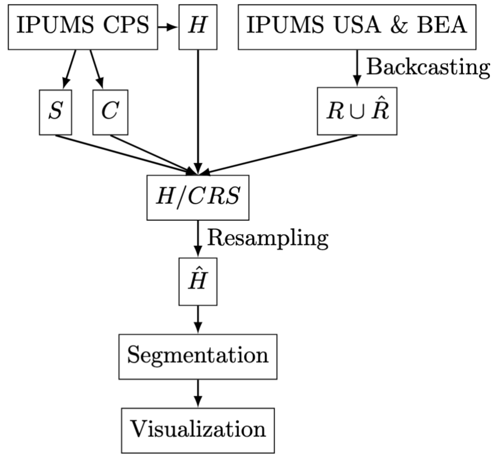


Figure 3: A summary of our workflow. We first adjusted the household income data with inflation, state-price difference, and household size. We then bootstrap resample to adjust for age distribution difference. Finally, we construct an empirical household distribution for each state by segmenting them according to their percentile. The data is then ready to be displayed.

with less statistics experience, hence we advise educators to introduce this concept with care. We include the analysis of standard error in this paper for completeness.

To estimate the SE for each block's height depicted in the charts, we employ bootstrap resampling. We demonstrate sampling variability with CA and DC because they have the highest and lowest sample size, respectively. Within a given state, the column heights vary more as we go from front to back because the sample size decreases as the level of income increases: the sample size of the high income group is much smaller than that of the low income group. Indeed, in Figure 4, the back wall in 2019 is not only taller but more jagged than it was in 1976. The same conclusion can be obtained by examining Figure 18 in the Supplementary Material. The SE of highly populated states is several thousand dollars for the 95^{th} percentile, but for states with thin slides (e.g. DC and SD), the bounce is in the tens of thousands of dollars. If the Bureau of Labor Statistics had carried out a second survey for 1976, we would see the 95^{th} percentile of real household income in DC vary by plus or minus roughly $13,500.

4 Results and Discussion

4.1 3D Interactive Visualization

Figure 4 is the simplest 3D visualization we offer, where household income is only adjusted by inflation. An interactive graph for all years from 1976 to 2019 as well as animation of the graph over time are available at `research.depauw.edu/econ/incomevis`. This interactive visualization is an excellent starting point for discussion of how household incomes have evolved over time and across states. Since the colors are held constant at the initial year, it is easy to see that poor states (the online version allows them to be identified easily by hovering over them with the mouse) have mostly remained poor. There has, however, been some mixing of colors, suggesting that some states have grown faster than others. The online version of the chart invites comparison and generates questions.

Figure 4 is a direct analog to Figure 1. Instead of using countries, sorted by lowest to highest incomes with width indicating population, we use states. Mississippi is the poorest state (at the extreme left) and remains so, while California is the widest state and its light-green color indicates incomes are relatively higher. But if you use the online version to adjust for state price differences, California falls back in the rankings.

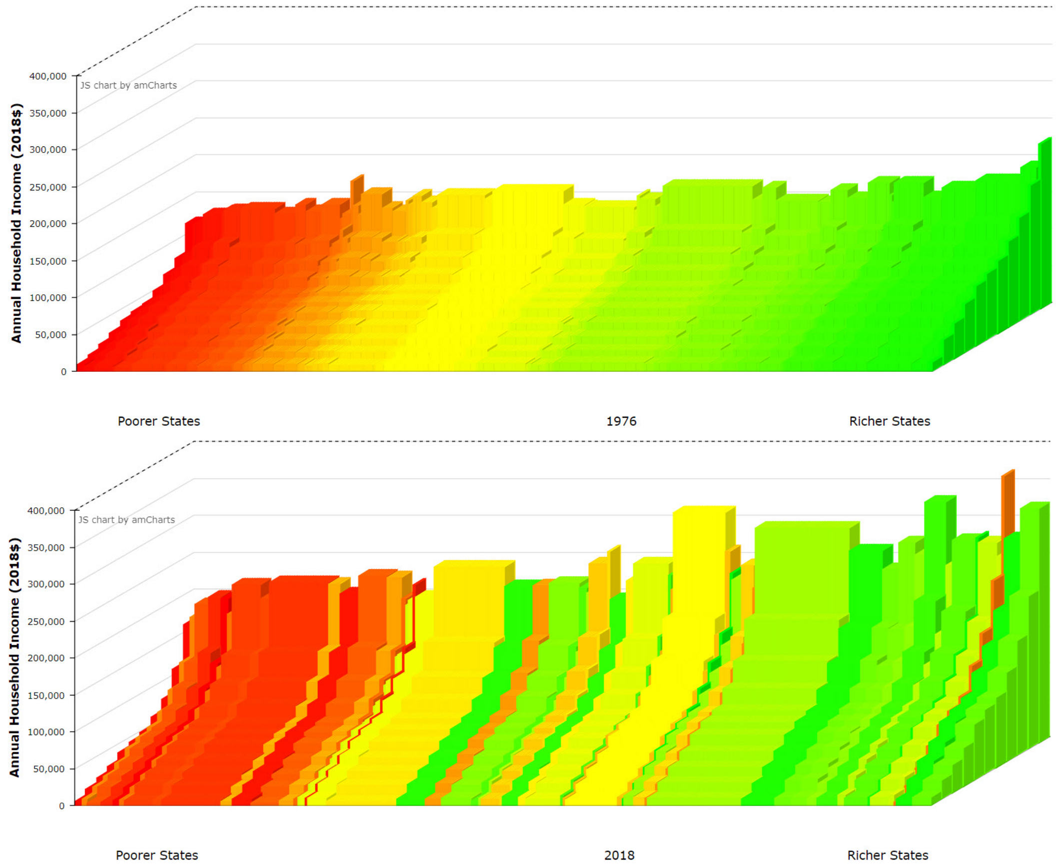


Figure 4: 1976 (top) and 2019 (bottom) household distribution adjusted only by inflation

This visualization is implemented with `JavaScript`'s `amCharts` library and is executable on any browser. Users can hover their mouse over different parts of the chart to get a popup with information. The down arrow in the screen's top-right corner allows the user to download the chart's data for other income computations. For teaching or exploration purposes, users can get a customized interactive graph with various options from our website. For example, they can choose to display the label of a particular state to track its position over time, or they can choose to apply different combinations of income adjustment as described in the Methodology section (such as inflation, state price differences, and household size).

4.2 A Case Study: Perspectives on the Progression of Income Inequality in the US from 1976 to 2019 with `incomevis`

We promote data visualization as a better communication channel that allows non-expert audiences to better engage in discussion around economic inequality relative to numerical representation. Nonetheless, we also note that the visualization has limitations. Indeed, the story presented by the visualization is only as good as the underlying data and may be compromised by the data preprocessing method. In a presentation to non-experts on economic inequality, it is important to look at the issue from multiple visualization perspectives. Our tool is well-positioned for this purpose.

In this section, we utilize our visualization methods to present seven visualization perspectives of income distribution data in the US from 1976 to 2019 (Figures 6, 7, 8 in the main text and Figures 13, 14, 15, 16 in Supplementary Material). Figure 5 explains how to read these graphs. We have data from 1976 to 2019, but we only display three years (1976, 1998, and 2019) in this paper for demonstration. Data from each year are presented in a 3D graph, and the sequence of three 3D graphs visualizes the progression of economic activity over time. Each 3D graph typically contains a collection of 51 "slides," each representing a state unless specified otherwise. Each slide is typically convex (see Figure 5), communicating the economic inequality within that state. A state's position on

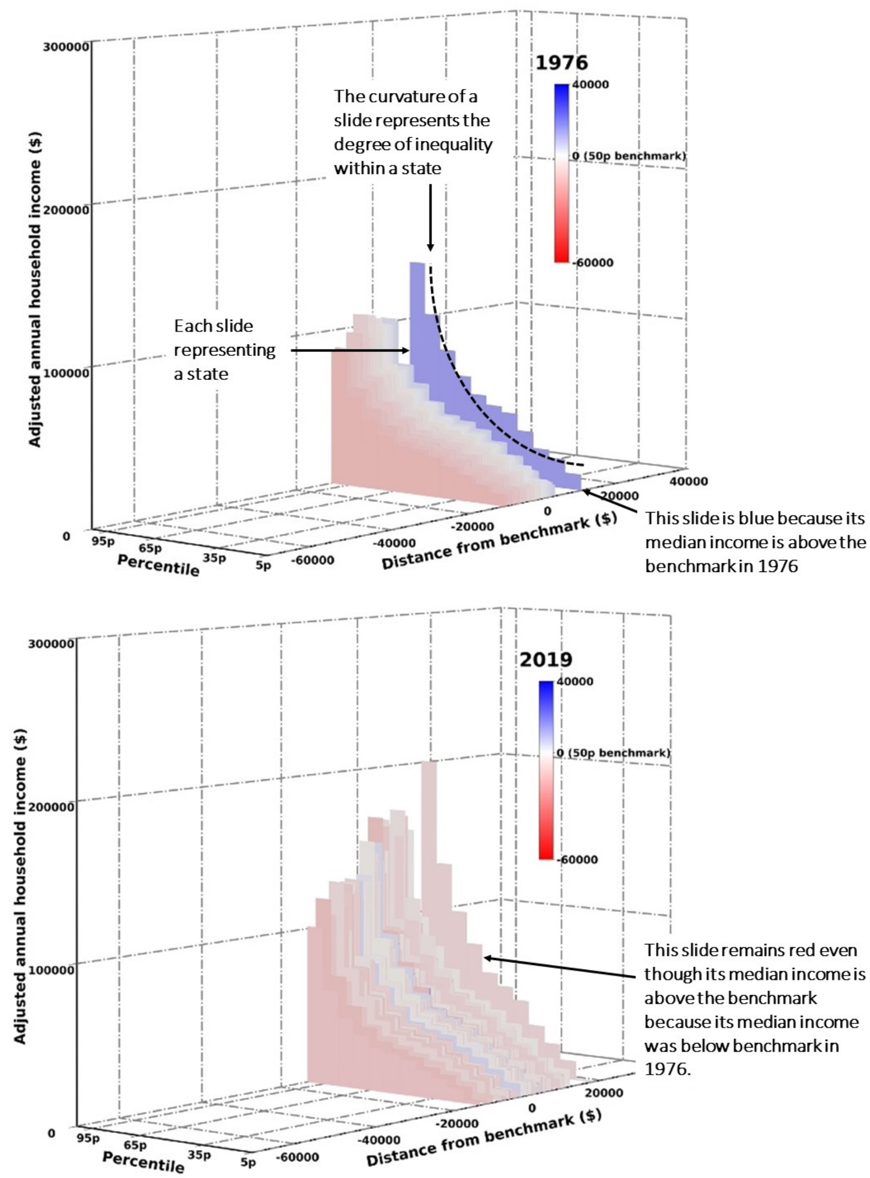


Figure 5: Instruction on how to read the graph

the horizontal axis (“Distance from benchmark ($)”) is determined by its relative distance to national average household income in 2019.

The color of a state is chosen according to its initial position in 1976 and this color is fixed for the entire 1976-2019 period (see Figure 5 for more details). This choice of coloring gives viewers an overall sense of position-changing patterns over time. For example, if a state is red, we know that it was very poor at the beginning in comparison to the 2019 benchmark. In addition, if a red state has a large positive distance from the benchmark in a given year, we know that incomes have grown significantly. The thickness of each slide is proportional to its state population, but this information might not be as visible as in Figure 4 due to a different in view perspective as well as the limitation of our underlying Python visualization engine (`matplotlib`). The ranges of all axes are fixed and carefully aligned to facilitate the comparison between graphs. We offer seven perspectives to highlight hidden insights from the raw data.

First, we visualize unadjusted data. Figure 6 suggests extremely rapid growth as the income range moves from –\$55,000 to –\$40,000 in 1976 to –\$20,000 to \$25,000 in 2019. All of the states are red because they were all below benchmark (median national income in 2019). Because they have shifted right, closer to the benchmark (0), and the blocks are taller as time goes by, we conclude that incomes have grown. Figure 6 also suggests that inequality has risen drastically because the 95th percentile grows much faster than the 5th percentile.

Note that the income data displayed in Figure 6 is without adjustment for inflation, household size, and state price levels, hence these conclusions about growth and inequality are misleading. Household incomes have not really grown this fast, once changing price levels and household size are taken into account. This motivates the need for the second perspective in Figure 7.

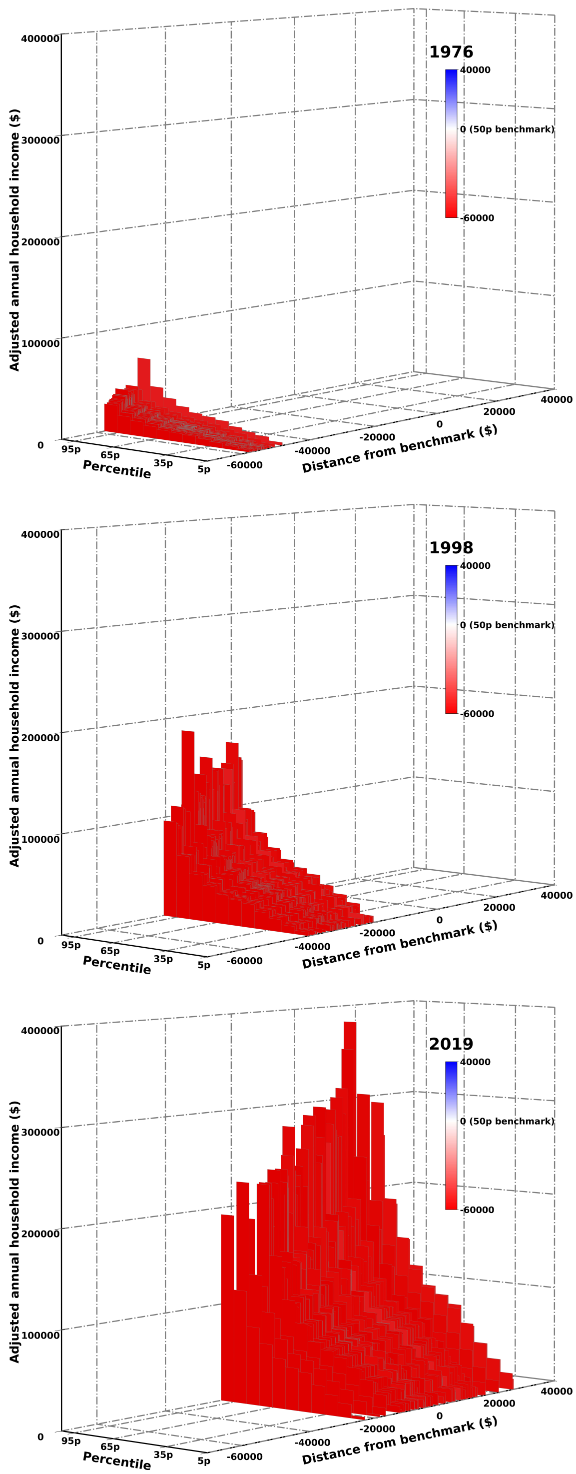


Figure 6: The 1st perspective presents unadjusted income.

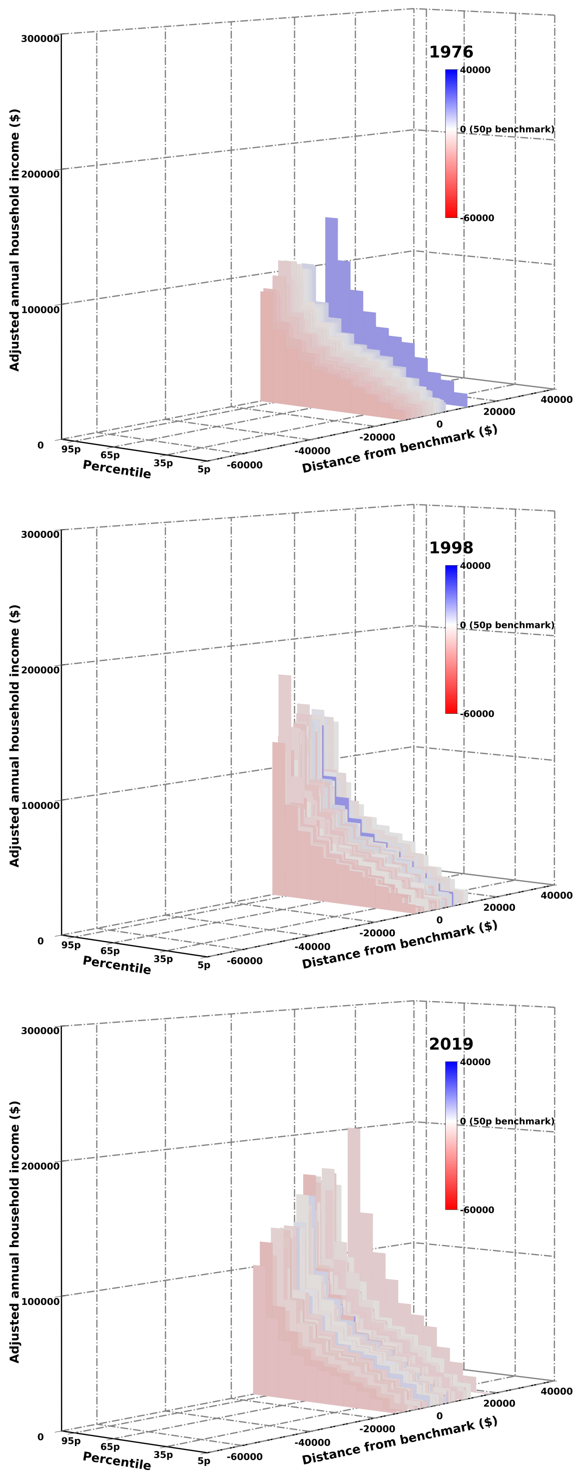


Figure 7: The 2nd perspective presents adjusted income ($\hat{H}$)

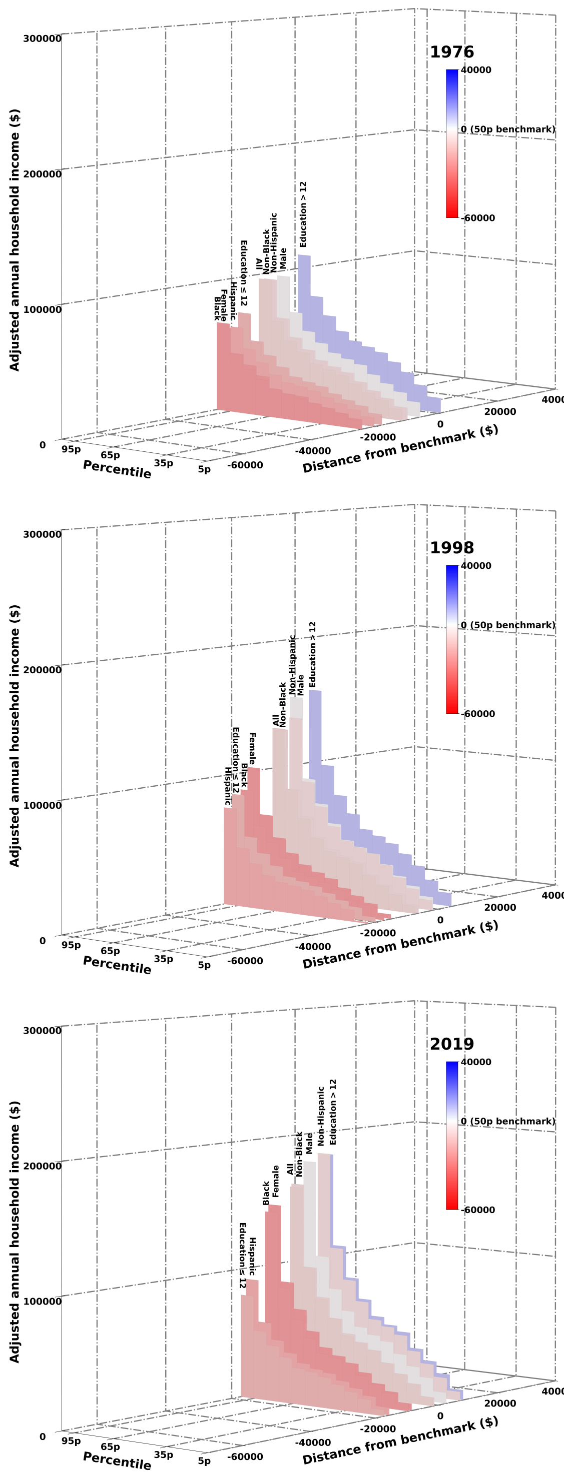


Figure 8: The 5th perspective is the adjusted income of some sub-population in CA

To tell a more accurate and complete story about the evolution of household income, Figure 7 displays incomes adjusted for inflation, household size, and state price differences. As we previously stated, by no means do we believe that these adjustments are all that we need – there can be many more normalizers. What we want to demonstrate here is that a few normalizers can drastically change the story. After adjustment, we do not see as much growth. The grayish colors in Figure 7 mean that most states were not as poor as suggested by the dominant red color in Figure 6.

We previously mentioned that the curvature in each slide can communicate information about the level of inequality. The closer the transition from low to high percentile is to linear, the more equality we have. However, it is not always easy to recognize the degree of curvature with deciles. Therefore we provide a third perspective in Figure 13 to demonstrate a higher resolution by presenting percentiles instead of deciles along the z-axis (shortest one). Although Figure 13 has more information than Figure 7, generating high-resolution visualization introduces higher computational cost.

Next, we focus on specific states. We suppress the color of every state into transparent green but CA and DC in the fourth perspective as an example. Interesting insight emerges immediately as a result: Both CA and DC were relatively poor and had (relatively) low income inequality in 1976, but DC quickly rises to the top with higher median income as well as a larger degree of inequality.

Our visualization tool not only allows us to study inequality among states but also between sub-populations within a state or sub-populations across states (Figure 8, 15, 16). We examine several subgroups [4] for CA in Figure 8: Black, Non-Black, Hispanic, Non-Hispanic, Education with or without high school diploma (Education ≤ 12), Education with higher degrees after high school (Education > 12), Male, and Female. Unlike in previous perspectives, each slide in Figure 8 represents a sub-population in CA. This picture reveals inequality between groups: households that are headed by male, people with more than 12 years of education, and people that are neither Black nor Hispanic are all above the state average in almost all years. Households with highly educated head have the highest income, higher than race, gender, and ethnicity groups, demonstrating the strong relationship between income and education. We also observe that the gender household income gap has shrunk as female-headed household income rises from below to above the benchmark. This trend is further highlighted in Figure 17 (Supplementary Material).

The gender income gap is not unique to CA. We further study this phenomenon by visualizing gender subgroups across states in Figure 15 and 16. In 1976, the dominant color for male-headed household is blue and for female-headed household it is red, indicating that female-headed households tend to start below the benchmark. Similar to CA, the gender income gap across states has also decreased over time.

5 Conclusion and Further Research

We introduce a framework for visualizing income distribution in the United States based on a similar "skyscraper plot" for global income. Consumer price index, regional price parity, and household size are the primary adjustment variables for our household income data. Our fundamental takeaway is that a 3D visualization communicates more information to a non-expert audience than conventional measures (such as the Gini coefficient). Our data visualization is eye-catching and requires little background technical knowledge. Thus, it provides a gentle introduction to the study of income distribution and inequality.

As with any other empirical, long-term study, our analysis suffers from some underlying limitations. For example, the CPS ASEC survey is not the same every year. Questions change, and so do data collection methods (e.g., see a discussion of majority changes in 2017 at (Rothbaum and Edwards 2019)). Our skyscraper plot for household income in the U.S. offers an excellent starting point into exploration of income inequality because it lays bare the facts, captures attention, and stimulates many questions about causes and remedies.

6 Acknowledgements

This project is supported by the Hewlett Mellon Presidential Fund for undergraduate research at DePauw University. We thank Frank Howland, Jonah Barreto, Jarod Hunt, and Bu Tran for their support with this project.

[4] The household is classified according to the household head

7 Author Biographies

Sang T. Truong received Bachelor of Arts in Economics, Computer Science, and Computational Chemistry from DePauw University, USA. He is currently pursuing a Doctor of Philosophy degree in Computer Science at Stanford University (USA).

Humberto Barreto, PhD, is Professor of Economics and Management at DePauw University. He is interested in using computers (especially Microsoft Excel) to improve the teaching of economics and has published papers and books on pedagogy, including (with Frank M. Howland) Introductory Econometrics using Monte Carlo Simulation with Microsoft Excel (Cambridge University Press, 2006), Intermediate Microeconomics with Microsoft Excel (Cambridge University Press, 2009) – now freely open-access – and Teaching Macroeconomics with Microsoft Excel (Cambridge Univeristy Press, 2016). He has been a Fulbright Scholar, won several teaching awards, and regularly offers teaching workshops: `depauw.edu/learn/econexcel/`. His ORCID is `orcid.org/0000-0003-4822-038X`.

8 Supplementary Material

Table 1: Summary statistics.

Variables	**Description**	μ	σ	**min**	**max**
Household income (H)	Total nominal income of all household members	52343.81	65094.54	-37040	3299997
Consumer price index (C)	Price level differences over time	1.18	0.55	0.65	2.92
Regional price parity (R)	Price level differences across states in a year	97.53	8.54	84.8	119.2
Gross rent (r)	Monthly rental cost of housing, with utilities	841.98	201.19	512.0	1600.0
Effective household size (S)	Square root of number in household	1.58	0.45	1	5.10
Household weight (w)	CPS-provided sampling weight	1,538.89	920.93	0.00	17957.53

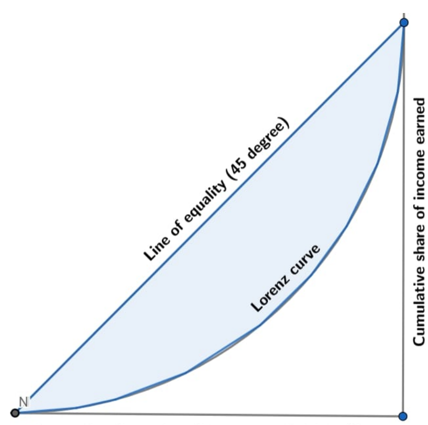
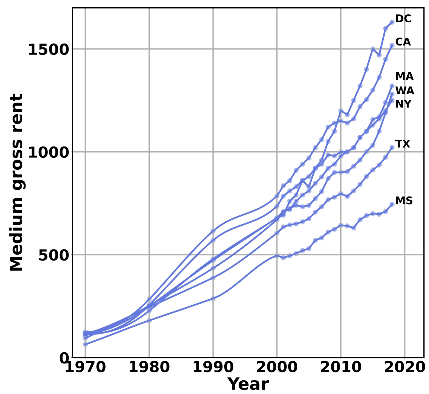
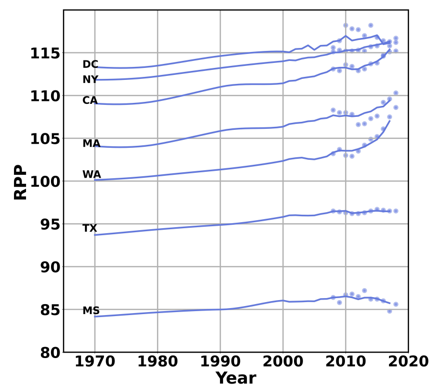


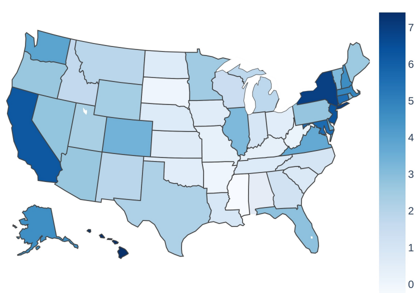
Figure 9: The Lorrenz curve (Aaberge 2001).



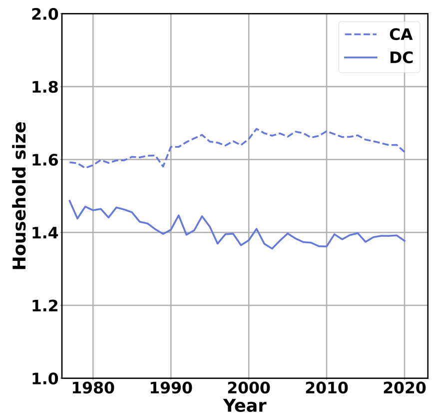
(a) Gross rent (r and $\hat{r}$)



(b) Regional price parity (R and $\hat{R}$)



(c) RPP's state fixed effect



(d) Average effective household size S

Figure 10: Estimation of regional price parity for each state from 1976 to 2019 applies equation 2. Gross rent has a linear trend in the majority of states. The fixed effect coefficients in the backcasting model show that DC and CA are among the states with the highest living expense.

References

[Aaberge 2001] Aaberge, R. Axiomatic Characterization of the Gini Coefficient and Lorenz Curve Orderings. *Journal Of Economic Theory*. **101**, 115-132 (2001), EconPapers.repec.org/RePEc:eee:jetheo:v:101:y:2001:i:1:p:115-132

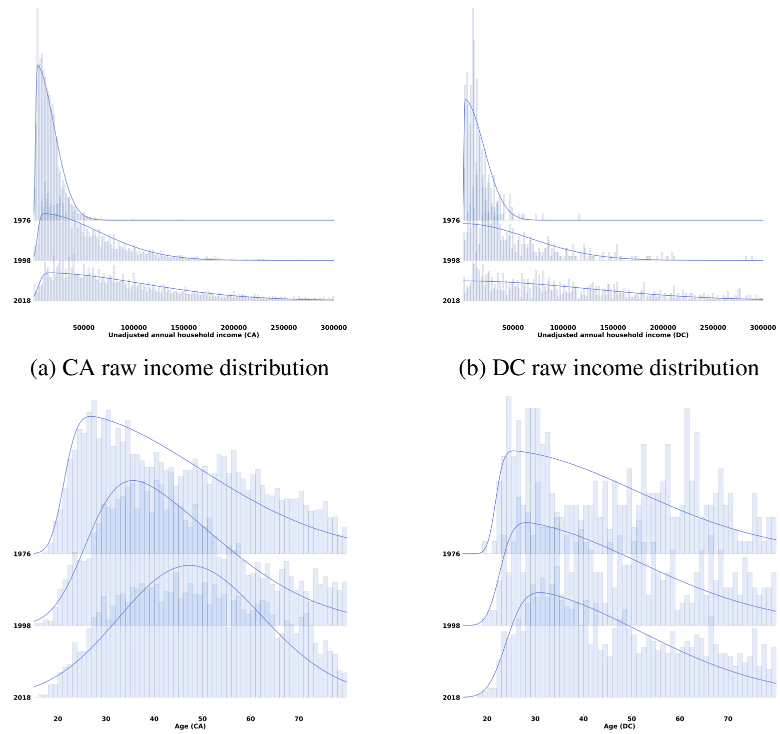
[Asarta et al. 2021] Asarta, C., Chambers, R. & Harter, C. Teaching Methods in Undergraduate Introductory Economics Courses: Results From a Sixth National Quinquennial Survey. *The American Economist*. **66**, 18-28 (2021), https://doi.org/10.1177/0569434520974658

[Almas and Mogstad 2012] Almas, I. & Mogstad, M. Older or Wealthier? The Impact of Age Adjustment on Wealth Inequality. *The Scandinavian Journal Of Economics*. **114**, 24-54 (2012), onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9442.2011.01662.x

[Alderson et al. 2005] Alderson , A., Beckfield, J. & Nielsen, F. Exactly How Has Income Inequality Changed?: Patterns of Distributional Change in Core Societies. *International Journal Of Comparative Sociology*. **46**, 405-423 (2005), doi.org/10.1177/0020715205059208

[BEA 2020] Bureau of Economic Analysis Regional Price Parities by State and Metro Area. (www.bea.gov/data/prices-inflation/regional-price-parities-state-and-metro-area,2020), Accessed: 2021-01-31

[Dorfman 1979] Dorfman, R. A Formula for the Gini Coefficient. *The Review Of Economics And Statistics*. **61**, 146-149 (1979), http://www.jstor.org/stable/1924845



(a) CA raw income distribution

(b) DC raw income distribution

(c) CA age distribution

(d) DC age distribution

Figure 11: The income distribution of both CA and DC has been increasingly right-skewed. The household size of both states is relatively stable over time. CA has a larger average household size than DC.

[Friedman 1962] Friedman, M. The Interpolation of Time Series by Related Series. *Journal Of The American Statistical Association*. **57**, 729-757 (1962), https://www.tandfonline.com/doi/abs/10.1080/01621459.1962.10500812

[Flood et al. 2020] Flood, S., King, M., Rodgers, R., Ruggles, S. & Warren, J. Integrated Public Use Microdata Series, Current Population Survey: Version 8.0 [dataset]. (doi.org/10.18128/D030.V8.0,2020), Accessed: 2021-01-31

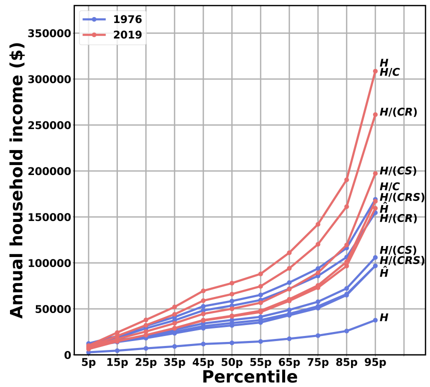
[Formby and Seaks 1980] Formby, J. and Seaks, T. Paglin's Gini Measure of Inequality: A Modification. *The American Economic Review*. **70**, 479-482 (1980), http://www.jstor.org/stable/1805239

[Heathcote et al. 2010] Heathcote, J., Perri, F. & Violante, G. Unequal we stand: An empirical analysis of economic inequality in the United States, 1967-2006. *Review Of Economic Dynamics*. **13**, 15 - 51 (2010), http://www.sciencedirect.com/science/article/pii/S1094202509000659, Special issue: Cross-Sectional Facts for Macroeconomists

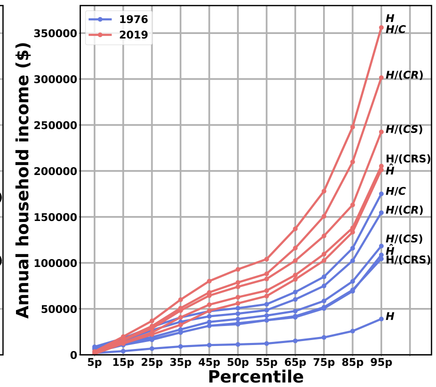
[Han 2011] Han, J. The analysis of opportunity inequality, income-gap and public policy. *2011 International Conference On Computer Science And Service System (CSSS)*. pp. 2665-2668 (2011)

[Harter and Asarta 2022] Harter, C. & Asarta, C. Teaching Methods in Undergraduate Intermediate Theory, Statistics and Econometrics, and Other Upper-Division Economics Courses: Results From a Sixth National Quinquennial Survey. *The American Economist*. **67**, 132-146 (2022), https://doi.org/10.1177/05694345211037904

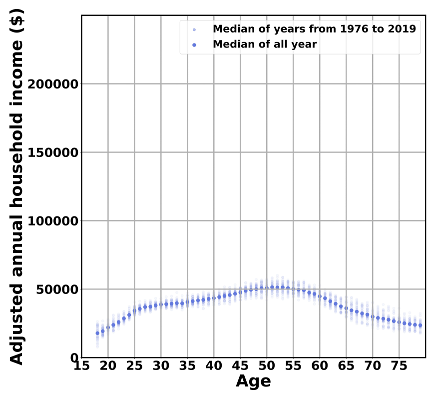
[Hartmann et al. 2017] Hartmann, D., Guevara, M., Jara-Figueroa, C., Aristarán, M. & Hidalgo, C. Linking Economic Complexity, Institutions, and Income Inequality. *World Development*. **93** pp. 75 - 93 (2017), http://www.sciencedirect.com/science/article/pii/S0305750X15309876



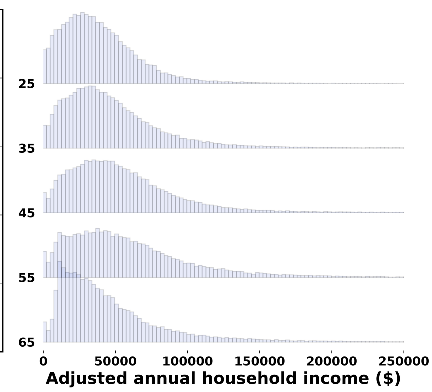
(a) Various normalizers for CA



(b) Various normalizers for CA



(c) Median $\hat{H}$ by age (1976-2019)



(d) $\hat{H}$ distribution by age (1976-2019)

Figure 12: The age distributions differ across states. DC has a relatively younger population than CA. The age distribution of each state may not be stable over time. Age and income imply a strong relationship in the U.S.

[Johnson et al. 2005] Johnson, D., Smeeding, T. & Torrey, B. Economic inequality through the prisms of income and consumption. *Monthly Labor Review*. pp. 11-24 (2005), www.bls.gov/opub/mlr/2005/04/art2full.pdf

[Murphy and Welch 1990] Murphy, K. & Welch, F. Empirical Age-Earnings Profiles. *Journal Of Labor Economics*. **8** (1990,4)

[Norton 2011] Norton, M. Rising Wealth Inequality: Should We Care?. *The New York Times*. (2011,5), Accessed: 2021-01-31

[Piketty 2014] Piketty, T. Capital in the Twenty First Century. (Boston: Belknap Press,2014)

[Rothbaum and Edwards 2019] Rothbaum, J. & Edwards, A. Survey Redesigns Make Comparisons to Years Before 2017 Difficult. (2019), Accessed: 2021-01-31

[Ruggles et al. 2020] Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J. & Sobek, M. IPUMS USA: Version 10.0 [dataset]. (doi.org/10.18128/D010.V10.0,2020), Accessed: 2021-01-31

[Renwick 2011] Renwick, T. Geographic Adjustments of Supplemental Poverty Measure Thresholds: Using the American Community Survey Five-Year Data on Housing Costs. *SEHSD Working Paper*. , (2011), https://www.census.gov/content/dam/Census/library/working-papers/2011/demo/SEHSD-WP2011-21.pdf

[Sutcliffe 2001] Sutcliffe, B. 100 ways of seeing an unequal world. (London: Zed Books,2001)

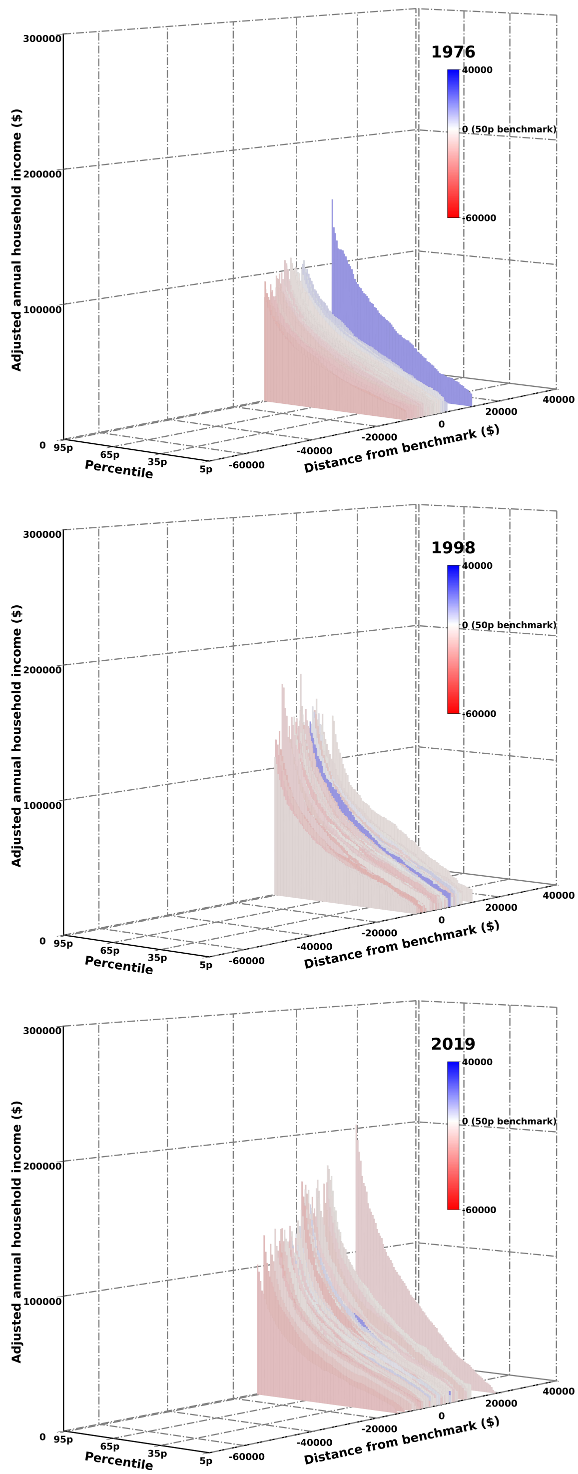


Figure 13: The 3rd perspective is a high resolution view of the 2nd one.

[The Core ECON Project 2020] The CORE Team The Economy. (core-econ.org/the-economy/book/text/19.html,2020), Accessed: 2021-01-31

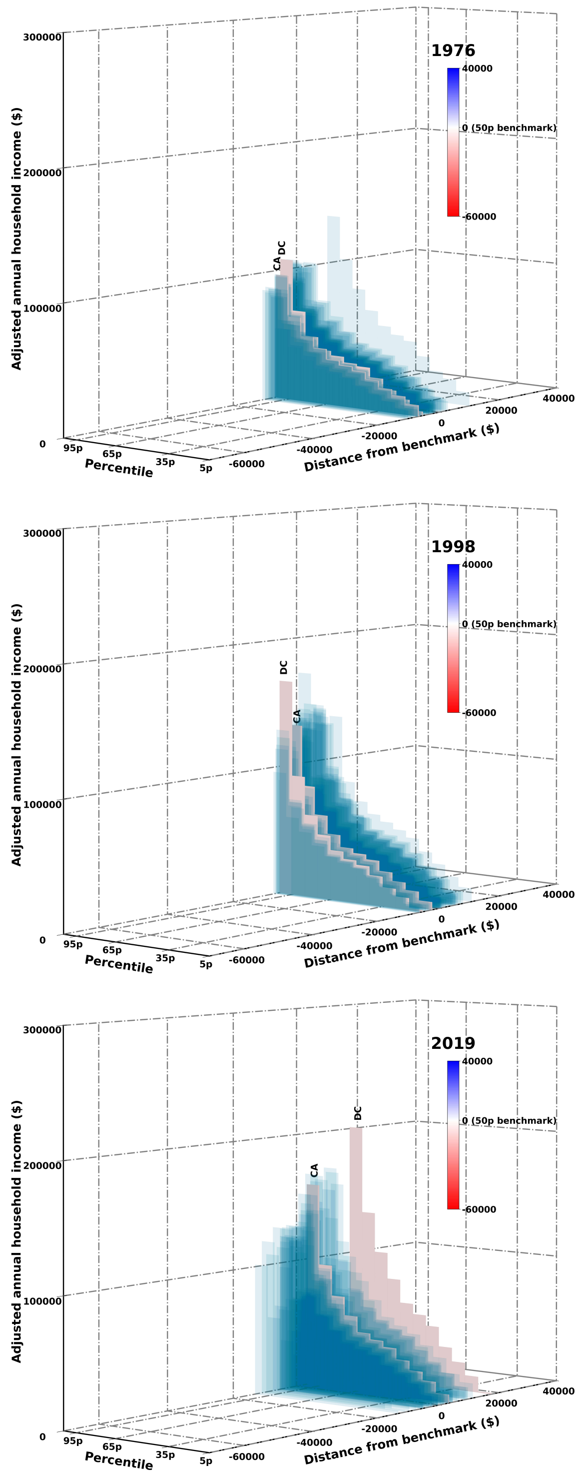


Figure 14: The 4th perspective focuses on adjusted income of CA and DC

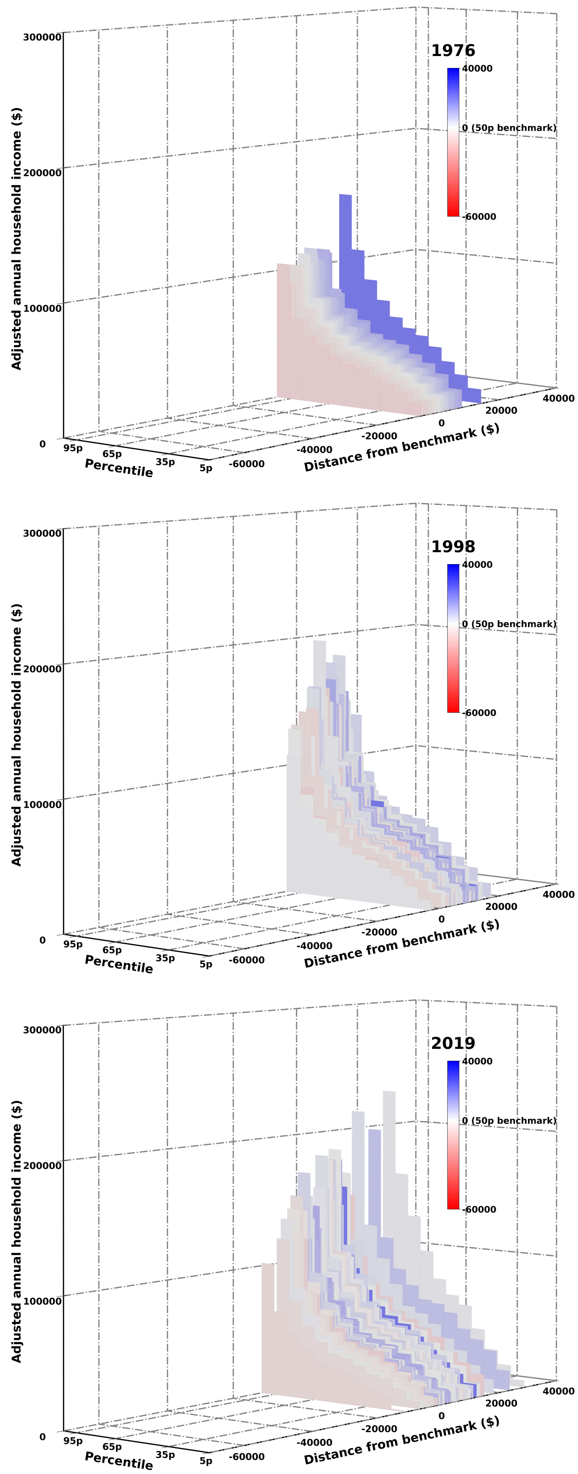


Figure 15: The 6th perspective shows adjusted income of male-headed household

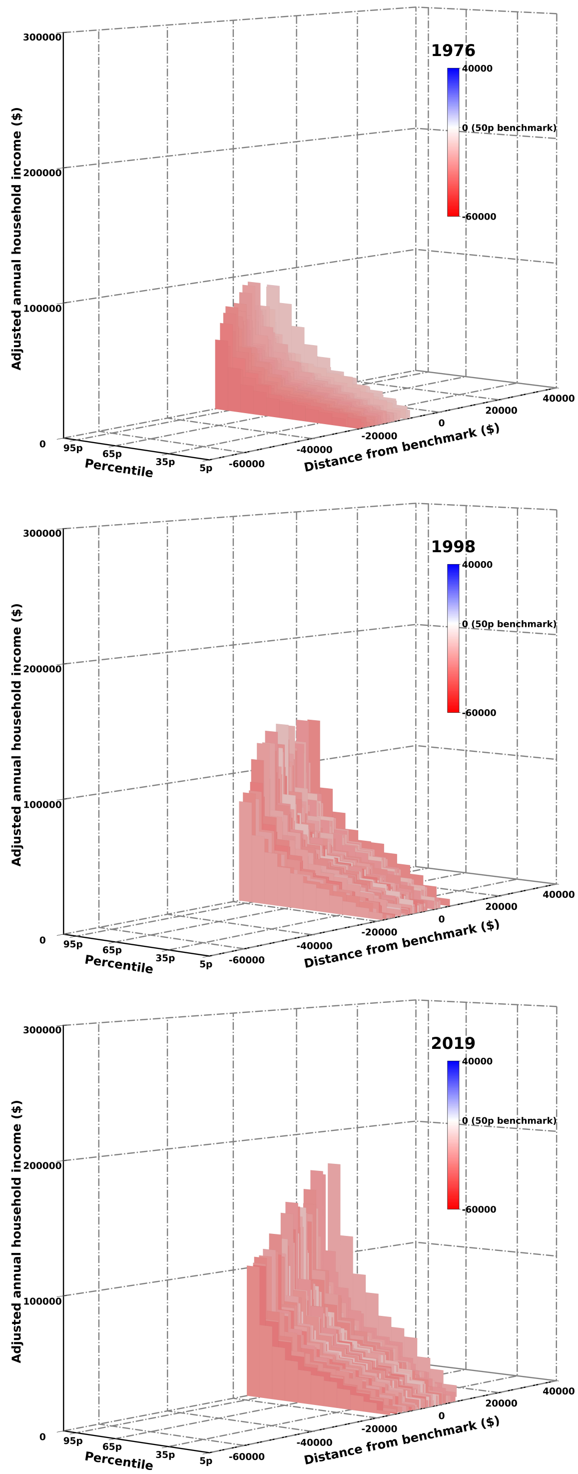


Figure 16: The 7th perspective shows adjusted income of female-headed household

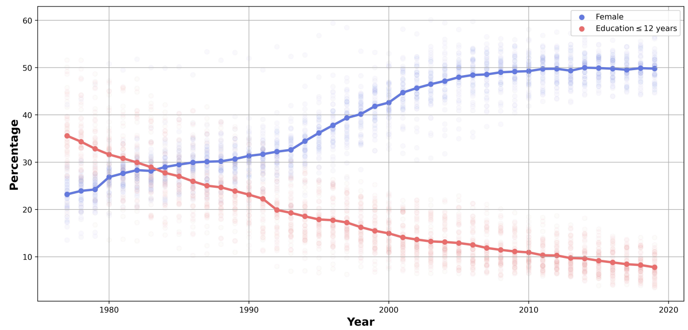


Figure 17: The percentage of households with high school or below education has declined drastically over time. Conversely, the percentage of the female-headed household approaches 50%.

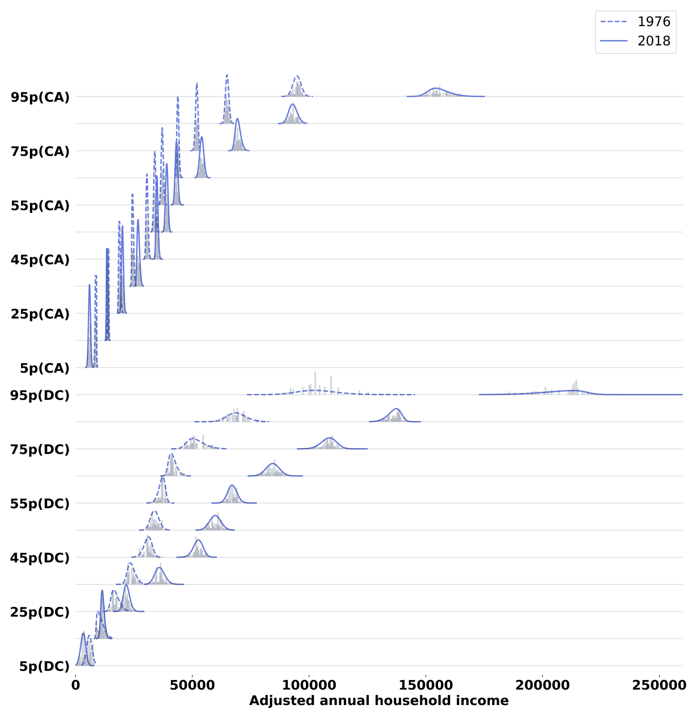


Figure 18: Adjusted annual household income distribution in CA and DC in 1976 and 2019.