
Income Inequality Visualization

Project Proposal and Process Book

Basic Info

Team Members

Jared Nielson	jarednielsonn@gmail.com	u0495206
Brian Park	bapark92@gmail.com	u0735732

Repository

https://github.com/jarednielson/income_vis

Website

https://jarednielson.github.io/income_vis

Background and Motivation

Income inequality has steadily grown since the mid 1970's. In 1975 the top 1% of earners captured 8.9% of new income created, their lowest share in the past 100 years. Since then the percentage of total income for top earners (both in the top 10% and top 1%) has increased, peaking at 50% and 22% respectively just prior to the 2008 financial crisis (Saez, 2016). Total income for top earners dropped post 2008, but quickly recovered. From 2008 to present the top 1% of earners have more than made up for losses in the crisis increasing their income from 2008 to 2012 by 34.7% while the bottom 99% saw incomes increase by a modest 0.8% during the same time period (Saez, 2016).

The gap between top earners and the rest of the country is at heights not seen since the before great depression when top earners captured a 50/25% split for top 10% and 1% respectively (Saez, 2016). The economic policies enacted by the Roosevelt administration helped mitigate income inequality and restore balance between top and bottom earners from the 1940s to the 1970s. At which point more regressive policies were enacted that have led to the state we see today. The 1940s and 50s saw the highest marginal tax rates peaking at 92% in 1951 and 1952. While the rates dropped in the 1970s the lowest marginal tax rate for top earners was during the Reagan Administration in the late 1980s at 28%. Rates have gone up since then but are still only 1/3 of their peak in the 1950s (Tax Policy Center, 2017).

Why is this concerning? Income inequality isn't just about money, it is about quality of life. The gap in life expectancy between those in the top 1% and those in the bottom 1% is 14.6 years (Chetty, Stepner and Abraham, 2016). Millions of Americans cannot afford health insurance and lack access to health care, affordable housing, and even food. This is a life and death issue.

In December of 2017 congress passed and President Trump signed into law tax cuts for the wealthiest of Americans. Among many, especially those at the top there seems to be an unwillingness to accept the data at its face value, that as the richest prosper the middle and lower class struggle. One of these reasons may be that most people can't internally visualize just how skewed the income distribution is. Most papers on income inequality present graphics such as line and bar charts, but fail to create compelling and interactive visualizations that drive home the reality of income inequality. Our goal is to create a visualization that will drive home both the reality and the scale of income inequality.

Project Objectives

1. Prior to seeing the data, users will be given an opportunity to create two income distributions. The first will be an idealized version that they think might be considered fair and the second what they predict the distribution actually looks like. These distributions will then be saved and shown to the user as a comparison to present and historical income distributions.
2. Allow users to explore historical income distributions as well as historical marginal tax rates, and other interesting historical features, like average home price per year to allow the data to be put in real context.
3. Allow users to explore income distributions per state (as data is available).
4. Create a visualization for top earners in the United States (Jeff Bezos, Mark Zuckerberg, etc.) that shows how many average or low income earners it takes to make up the wealth of these individuals.

Data

Historical income data will be obtained from the Census Bureau. They have both a json API and xml files available:

<https://www.census.gov/data/developers/updates/new-discovery-tool.html>

<https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-income-households.html>

Historical Tax rates will be obtained from the Tax Policy Center:

<https://www.taxpolicycenter.org/statistics/historical-highest-marginal-income-tax-rates>

Data Processing

Because of the wealth of data that exists from these sources as well as the preformatted nature of them as well, including already constructed json, the need for advanced data processing is low. From the census api it is easy to obtain almost directly usable json representations of the information that we desire to compare.

Most if not all of the data processing that will need to be done can be accomplished by simply applying filters and using built in tools in d3 and JavaScript. These subsets of the data can then be directly visualized or further manipulated programmatically.

Visualization Design

There are potentially a large number of visualizations that are able to encode the data that is acquired through the census, however, we will specifically be looking at several key features of this data that we must encode. Groups based on income ranges, the number of people in each of these subgroups, and chronology / time are the main elements that we will need to encode in our visualization.

The first prototype uses a map to encode information about individual states and wealth distribution in these states. When a state is selected it shows a small tooltip displaying extra information about the state, and a time slider that shows the change in income disparity over time. The second prototype is a line chart that allows the user to 'guess the data' line chart and then reveals to them the actual distribution and could have chronology options as well based on year this will show users how common it is to misjudge the problem. The last prototype is the simplest, it is a histogram with the income groups and shows their distribution over time.

To combine all three and present the story of the data, we would like to possibly implement some simpler form of all three, and create a multiphase visualization similar to scrolly telling or the new york times' multiphase 3d visualization. Beginning with requesting the users guess at distribution first, showing the actual distribution with the histogram, and then letting them explore the data with the map and year slider. This provides a hook introduction, shows useful easy to digest data, and then allows for user interaction to promote memorability.

Other topics that we will look at include chronological information / events in relation to large tax cuts and its effects on the distinct income groups. Our objective is to show not only income disparity, but other topics around this including but not limited to political events, recessions and how these influence the 'wealth gap'. To show this we can add steps to the visualization that show snapshots of info, potentially before or after the map. It may be difficult to show these relationships faithfully without misconstruing the data, so the project will err on the side of caution with hard to prove areas and trends.

Must-Have Features

- Visualizations demonstrating the wealth gap and income disparities between different economic classes
- Helpful interactivity components that help demonstrate the story of the data, eg background knowledge on important events
- Visual components showing tax rates of these different economic groups
- Interactivity combining tax rates and income disparities

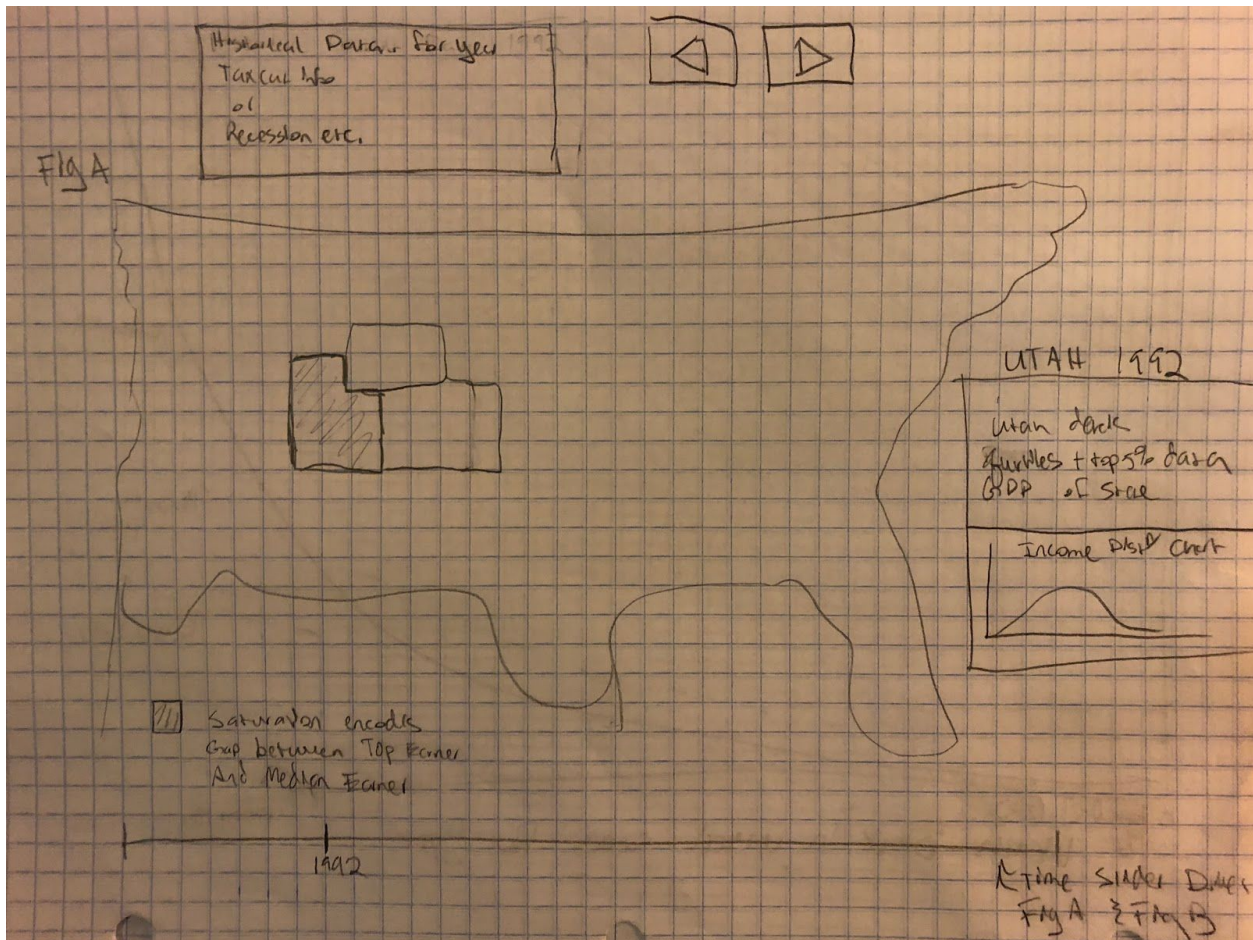
Optional Features

- Multiphase / scrollytelling interactivity and changes
- Highlight & explain historical events such as the dot com bubble, and the 2008 recession
- Allowing user customizability of what data is being displayed allowing for user data exploration

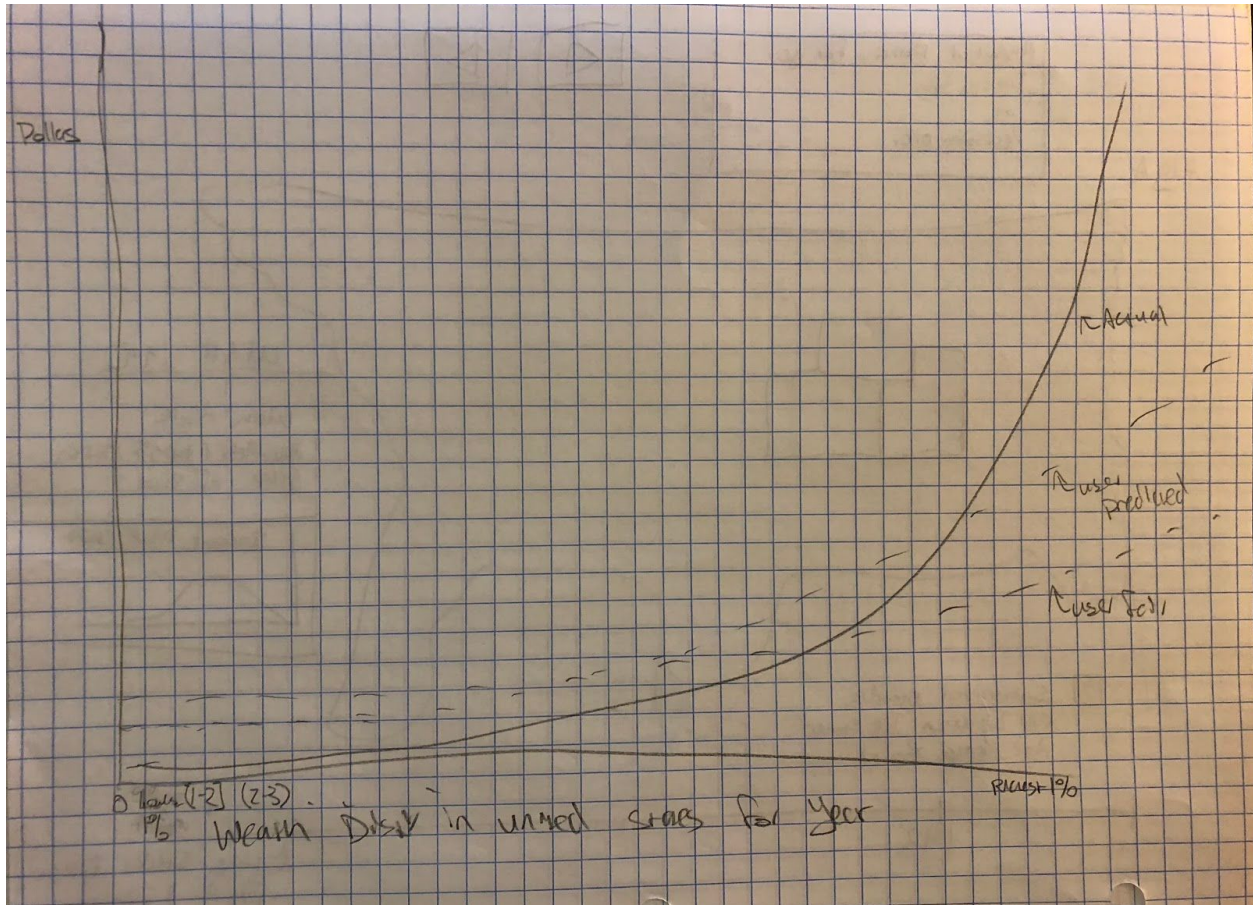
Project Schedule

Week 11
<ul style="list-style-type: none">• Data acquisition and formatting done• Final design settled on
Week 12
<ul style="list-style-type: none">• Data storage & formatting done• First basic visualization elements done• Line chart with basic features finished
Week 13
<ul style="list-style-type: none">• Advanced Visualization components finished• Line chart advanced features done• Histogram begun
Week 14
<ul style="list-style-type: none">• Advanced interactivity finished• Scrollytelling features started• Histogram done• Map basic features begun
Week 15
<ul style="list-style-type: none">• Polish visual design• Finished scrollytelling features• Map implementation finished

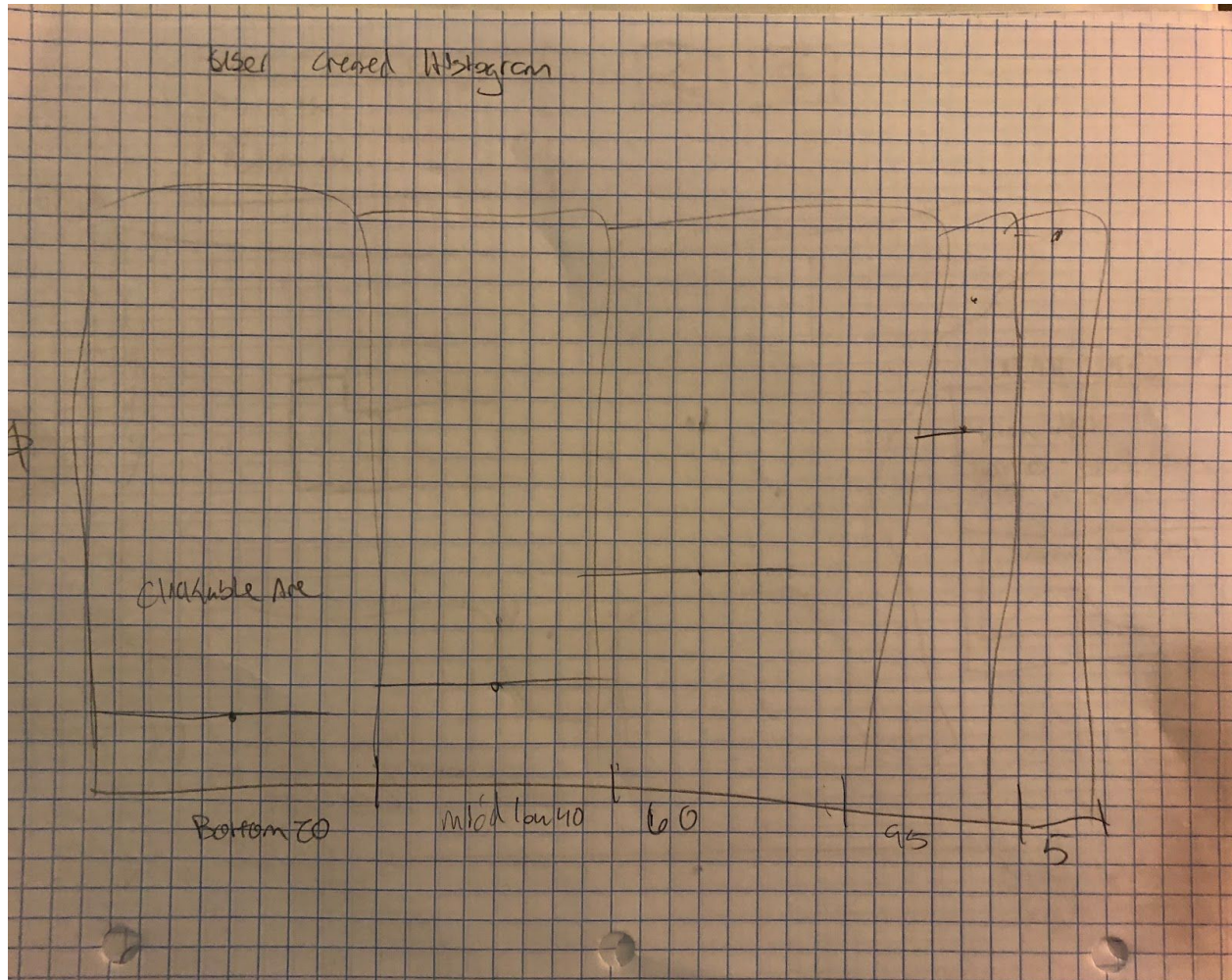
Design 1



Design 2



Design 3



Process Book

Overview and Motivation

Income inequality has steadily grown since the mid 1970's. In 1975 the top 1% of earners captured 8.9% of new income created, their lowest share in the past 100 years. Since then the percentage of total income for top earners (both in the top 10% and top 1%) has increased, peaking at 50% and 22% respectively just prior to the 2008 financial crisis (Saez, 2016). Total income for top earners dropped post 2008, but quickly recovered. From 2008 to present the top 1% of earners have more than made up for losses in the crisis increasing their income from 2008 to 2012 by 34.7% while the bottom 99% saw incomes increase by a modest 0.8% during the same time period (Saez, 2016).

Why is this concerning? Income inequality isn't just about money, it is about quality of life. The gap in life expectancy between those in the top 1% and those in the bottom 1% is 14.6 years (Chetty, Stepner and Abraham, 2016). Millions of Americans cannot afford health insurance and lack access to health care, affordable housing, and even food. This is a life and death issue.

In December of 2017 congress passed and President Trump signed into law tax cuts for the wealthiest of Americans. Among many, especially those at the top there seems to be an unwillingness to accept the data at its face value, that as the richest prosper the middle and lower class struggle. One of these reasons may be that most people can't internally visualize just how skewed the income distribution is. Most papers on income inequality present graphics such as line and bar charts, but fail to create compelling and interactive visualizations that drive home the reality of income inequality. Our goal is to create a visualization that will drive home both the reality and the scale of income inequality.

Related Work

In the past, we have learned about disparities that are present in both class as well as race from schooling as well as politics. It is clear that a large percentage of our societal burdens are caused by factors that are a direct result of unfair income differences. In class we have also watched visualizations about the distribution of wealth in the USA and just how mind boggling it is.

Questions

When we started this project our main questions we wanted to answer:

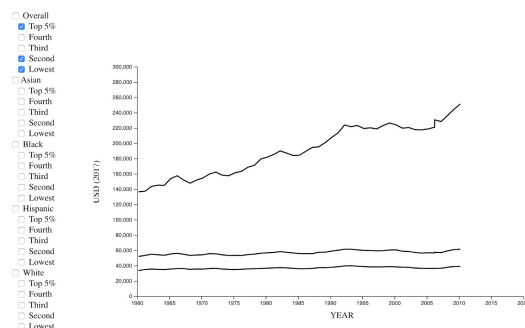
1. What does the income distribution in the United States look like.
2. Do people have an understanding of what that distribution is before seeing it?
3. Are there areas of the country where economic inequality is greater than others?

1. How has income distribution changed over time in the USA? And which groups have made the most economic gains?
2. Just how large is the gap in income between various groups and can it be effectively visualized?
3. How does race factor into income in the USA?
4. What events have influenced wealth distribution in America?

Data

Exploratory Data Analysis

One of the first visualizations we did (which we include in a more polished form in the final project) was a simple line chart shown below. This visualization allowed us to explore the data in various combinations and really highlighted income disparity in the United States.

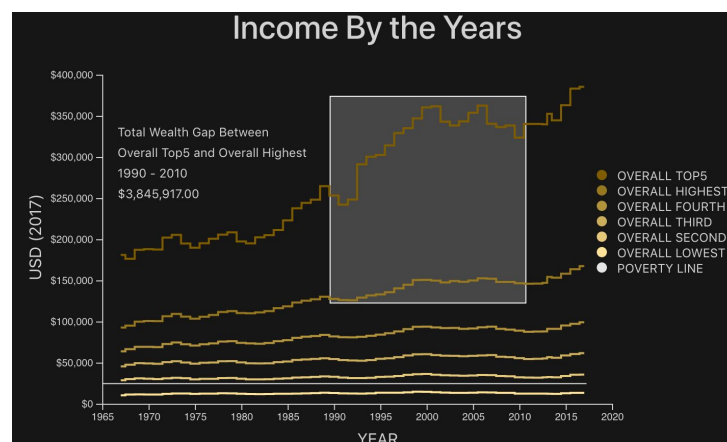


Design Evolution

Although wealth disparity is more of a problem in the USA compared to income disparity, wealth visualizations have been done frequently, for example the video about wealth that we watched in class is a great example. We decided to focus more on income and tax rates, with some highlights about wealth. Our main data focus however, changed to income, and with the ease of access racial disparities in income.

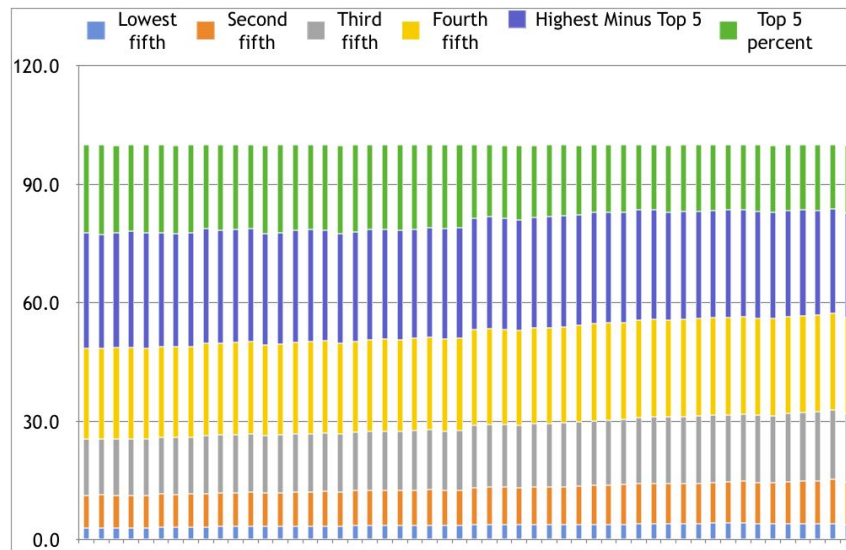
The data we were able to get from the census bureau's website we felt was best visualized with a line chart. However, there were many different racial categories for income (Overall, White, Hispanic, Black, Asian) and within each category 4 quartiles and the top 5%. This led to 25 different sets of historical income data. An interested user might want to explore or compare any number of these data sets, for example the top 5% overall vs the top 5% of black earners. Rather than clutter our line chart with 25 different lines we decided to implement selectors so users can select which sets of data they want to see. Color encodes the category while y position encodes income and x position encodes time. Unwanted data is hidden away.

The first prototype of the line chart is shown above. The final version is much more polished. We decided to use a step line generator since that made more sense. We also encoded category with hue and pentile with saturation added a legend, brushing to show wealth gaps and the poverty line for 2017. The controls from the original design are not shown but are used to link this chart with the aggregate income chart.

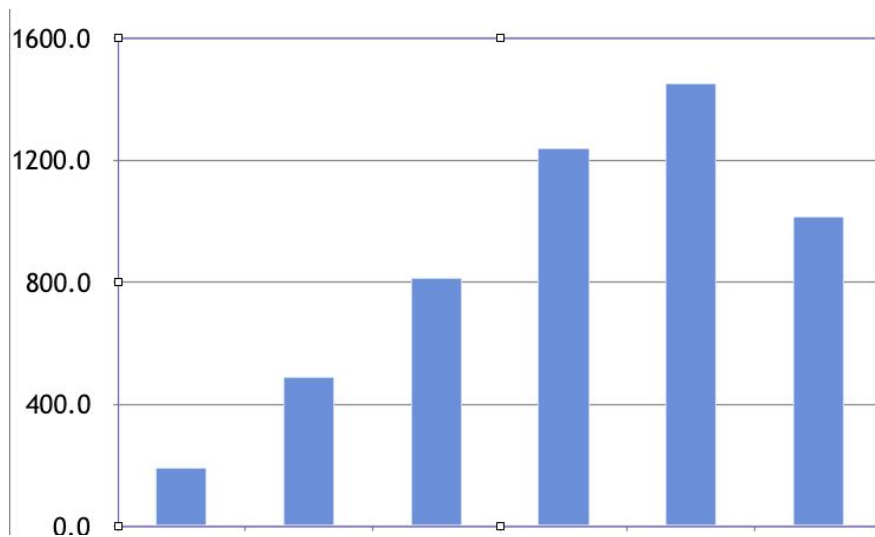


The Census Bureau had similarly formatted data for total income shares that we wanted to use as well. It made sense to link another view to our checkbox controller since the demographic categories were identical. Originally we wanted to do a stacked bar chart (where each demographic category makes up a percentage of the whole), but that ended up mirroring the line chart data and being uninteresting. We thought it might be more interesting to show total income shares by demographic instead. Below are our mockups for this process and the final result.

Mockup 1

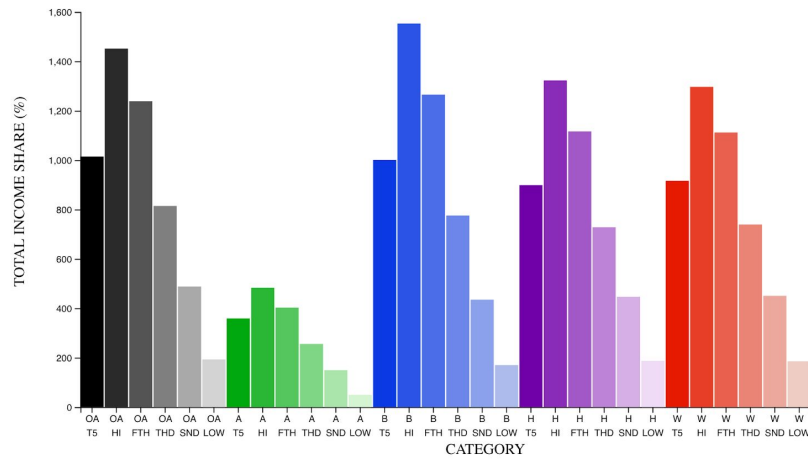


Mockup 2



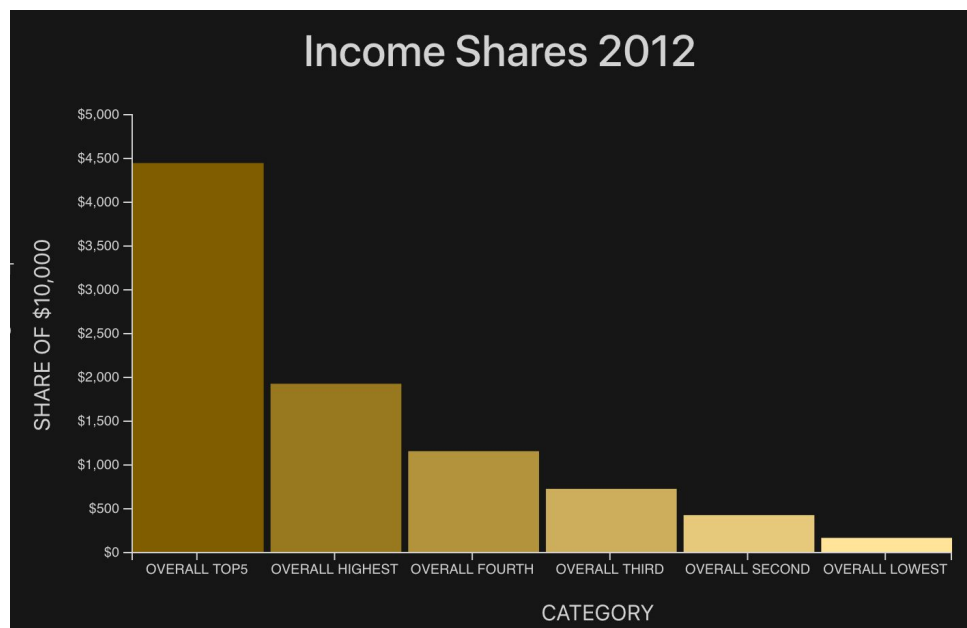
First Pass

Aggregated Income Shares



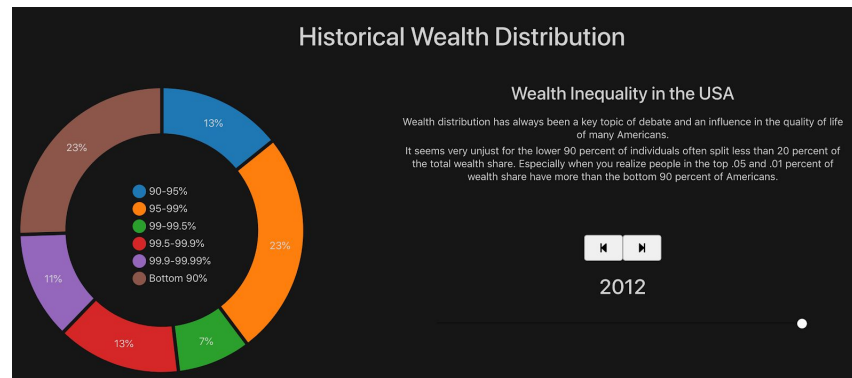
*Highest contains the 80th-94th percentiles

Final

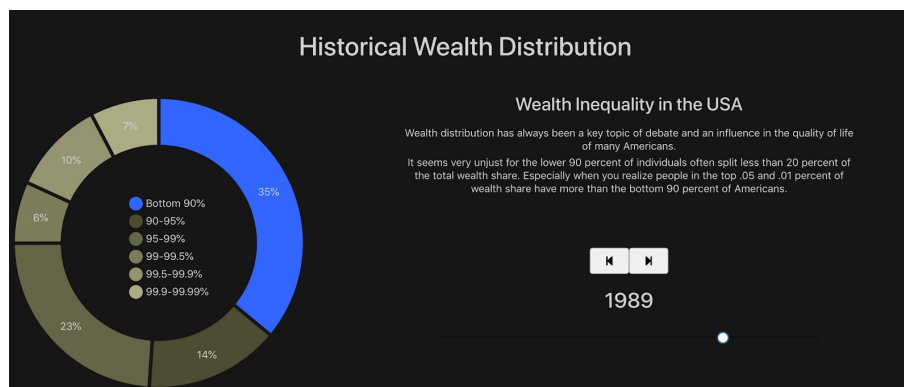


The final version is a different take on the problem. We manipulated the data to show how much an individual in each pentile and the top 5% would earn for every \$10,000 of income generated. This fixed a few problems. First it normalized the data so Asian earners didn't look under represented because we were missing historical data. Second it really drives home the stark difference between top earners and the rest of the country. We were able to include this in our income story.

To visualize wealth shares we wanted to easily show the proportion. While angle isn't the greatest visual encoding, it seemed appropriate for this case as the biggest comparison we wanted to show was between the groups in the top 10 and the bottom 90. Below is our first version of this chart.



Notice how difficult it is to tell the bottom 90 from the top 10. We thought it would be better if we changed how we used our color encoding for category. We decided to use one hue for all groups in the 10% and another for the bottom 90. Our final version is below.



With this new color encoding it becomes much easier to see what we want to point out (that the bottom 90% has much less than the top 10%).

For our story we wanted to take the martini glass approach. We have two stories to tell, one about income, the other about wealth. In each case users can step through the story and then explore the data themselves.

Our final design definitely deviated from our proposal. Originally we had planned to use a cargogram to show income inequality by state. However our final project remains true to the ideals that we had set when we first started. We wanted to highlight income inequality in America.

Implementation

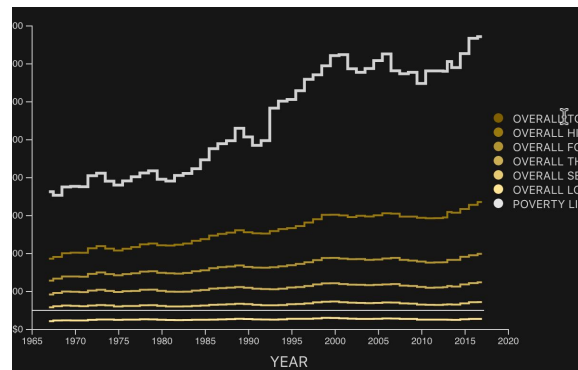
For our main view we wanted to make selecting categories easy. There are 25 different possible data sets to select. We implemented a way to quickly select demographics. These checkboxes drive both the Income Year Chart and the Income Shares Chart. Because both charts had handlers for the change it actually made implementing our story logic fairly easy. More on that later.

<input checked="" type="checkbox"/> Overall	<input type="checkbox"/> Asian	<input type="checkbox"/> Black	<input type="checkbox"/> Hispanic	<input type="checkbox"/> White
<input checked="" type="checkbox"/> Top 5%	<input type="checkbox"/> Top 5%	<input type="checkbox"/> Top 5%	<input type="checkbox"/> Top 5%	<input type="checkbox"/> Top 5%
<input checked="" type="checkbox"/> Highest*	<input type="checkbox"/> Highest*	<input type="checkbox"/> Highest*	<input type="checkbox"/> Highest*	<input type="checkbox"/> Highest*
<input checked="" type="checkbox"/> Fourth	<input type="checkbox"/> Fourth	<input type="checkbox"/> Fourth	<input type="checkbox"/> Fourth	<input type="checkbox"/> Fourth
<input checked="" type="checkbox"/> Third	<input type="checkbox"/> Third	<input type="checkbox"/> Third	<input type="checkbox"/> Third	<input type="checkbox"/> Third
<input checked="" type="checkbox"/> Second	<input type="checkbox"/> Second	<input type="checkbox"/> Second	<input type="checkbox"/> Second	<input type="checkbox"/> Second
<input checked="" type="checkbox"/> Lowest	<input type="checkbox"/> Lowest	<input type="checkbox"/> Lowest	<input type="checkbox"/> Lowest	<input type="checkbox"/> Lowest

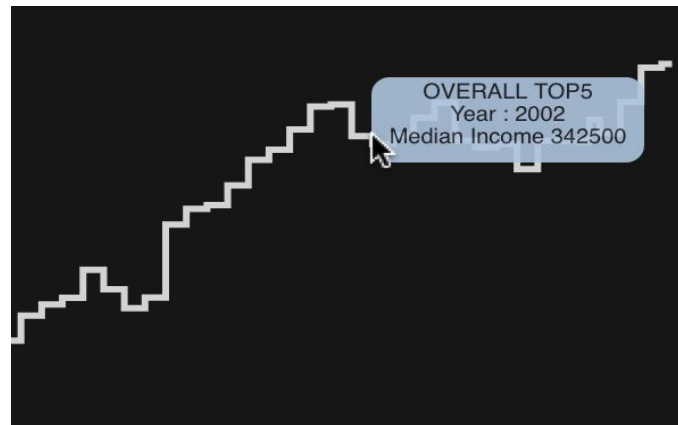
Each view was implemented in it's own class (We have an IncomeYearChart class, AggregateIncome class and a Wealth class) to drive its own display. We then have an init.js file that loads all of the data, does any mangling and then instantiates each view and sets any necessary callbacks. This way each view is basically encapsulated by itself and we just have one global driver function.

For the Income chart, in addition to filtering with the check boxes there are several interactive features:

[Hover highlighting for both lines and legend](#)



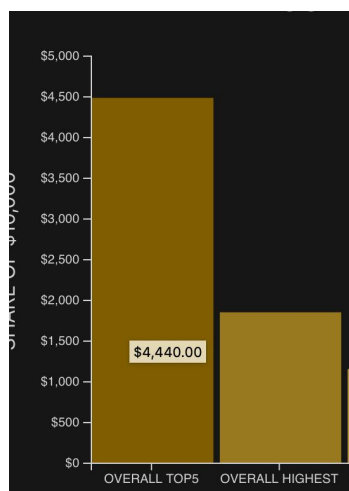
[Hover tooltip on lines](#)



[Brush for income gap.](#)

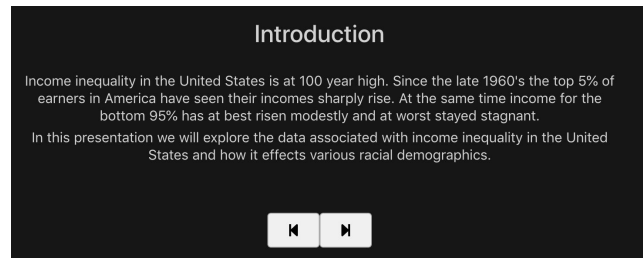


The Income Shares Chart has both a time slider and tool tip.



The wealth share chart has a timeline to allow scrubbing and seeing how wealth distribution has changed over time.

The stories were implemented using a Story class that is reusable (we use different instances of the class for each story). The class takes selectors for elements to modify as well as a side effects function that makes necessary changes on the page. Because of this design we were able to add a second story with only a few additional lines of code. Below is an example of what a story card looks like.



One challenge for the Income Shares Chart was getting the binned axis labels to behave well. We were able to set a condition to reformat the labels themselves if the number of bins were too high, and to abbreviate them if they were even higher. This way we were able to display every category without getting overlap in our labels.

The brush implementation on the income chart was pretty tricky. We had to find a way to calculate the area between two brushed curves. Ultimately we decided to only use the highest and lowest curves caught by the left end of the brush. From there we could loop through the data and calculate the wealth gap.

Evaluation

Both of us are well aware of income and wealth inequality in the United States. One of the reasons we chose this topic was because we both see it as an issue. Neither of us knew the extent of the problem though. Both of us were extremely surprised by the racial disparity in income. In fact we made it one of the key parts of our story. One thing that stood out as well was that a person in the top 5% will earn nearly 30x as much, that is staggering.

Questions 1-3 were answered with our first two visualizations (Income Chart and Income Share Chart). From the Income Chart it's very easy to see which groups are making the most economic gains and the disparity between economic and racial groups. The brushing on the chart as well as the Income Share chart show just how big the gap is. Question 4 was answered with our wealth chart.

Our visualization works very well. I feel it presents the data in a fair and accurate way. It also paints a picture of the problems we face in America. If we were to improve it we might add transitions to the Income charts. It would also be nice to allow for re-ordering of the bars in the Income Shares chart.

Works Cited

Chetty, R., PhD, Stepner, M., BA, & Abraham, S., BA. (2016). The Association Between Income and Life Expectancy in the United States, 2001-2014. *JAMA*, 315(16), 1750-1766. doi:10.1001

Saez, E. (2016). Striking it Richer: The Evolution of Top Incomes in the United States. Retrieved October 20, 2018, from <https://eml.berkeley.edu/~saez/saez-UStopincomes-2015.pdf>

Tax Policy Center. (2017, March 23). Historical Highest Marginal Income Tax Rates. Retrieved from <https://www.taxpolicycenter.org/statistics/historical-highest-marginal-income-tax-rates>