# From Arrival to Assimilation: Exploring Immigration and the American Dream for Chinese Immigrants in the San Francisco Bay Area

Nyah Mattison and Naitian Zhou

INFO 247: Information Visualization and Presentation, Spring 2023

Dr. Marti Hearst

May 8, 2023

# **Table of Contents**

Table of Contents	2	
1. Project Goals	3	
2. Discussion of Related Work		
3. Visualization Walkthrough		
Introduction to the topic	4	
Education	6	
Occupation	9	
Income	10	
Healthcare	11	
Gradient of detail	12	
4. Data	13	
5. Tools	14	
6. Results	14	
Usability Testing Study	14	
Usability Test Results	15	
Revisions to the Design	16	
7. Link to Code and Visualization	23	
8. Team Contributions	24	
Appendix		
References	26	

### 1. Project Goals

Our goal for this project was to interrogate the validity of the American Dream for Chinese Immigrants using a compelling narrative format. Following the blueprint of organizations that have done similar narrative work such as the New York Times and Vox, we built a website to contain a visual story that walks viewers through the component parts of the American Dream that we have defined and interrogated using Census data from 2022 and explains generational differences that occur as immigrant families move towards assimilation, compared to the general population of the San Francisco Bay Area. Our target viewer demographic is a general audience of educated readers who are interested in social issues.

We had two main goals for our visualization:

- Help users understand that, for Chinese immigrant families, the ability to reach the American Dream is variable between generations, with the first and second generation performing the best, while the third generation at times underperforms when compared to the non-Chinese immigrant population.
- Provide an empathetic and engaging viewer experience that facilitates learning and understanding of the immigrant experience and journey

### 2. Discussion of Related Work

The American Dream is a throughline in the United States mythos: the idea that, regardless of your social, geographical, or economic background, hard work and perseverance will lead you to a better life in America. However, millennials and subsequent generations of American citizens, have begun to cast doubt on if the American Dream still exists despite ever-increasing class stratification.

Research by Livingston and Kahn. analyzing the social mobility of Mexican-American citizens in 2002 shows that, when controlling for social capital, effective wages decrease across immigrant generations for Mexican-American men, and stay stagnant or marginally decrease for Mexican-American women. The U.S. Department of Commerce came to a similar conclusion in 2016, in a report analyzing the 2013 Census. In this report, they concluded that current data supports the conclusion that often, high first- and second-generation immigrant achievement is not followed up by further upward mobility. However, research conducted by the Pew Research Center in 2013, on adult children of immigrants (or second generation immigrants) contained more optimistic findings: across a host of metrics like educational attainment and income, second generation adults fared better than first generation immigrant adults.

Work by Abramitzky et al, as well as Chetty et al, adds additional nuance to these earlier findings. Abramitzky et al analyzed millions of father-son pairs across 100 years, drawing from Census data and tax returns. Their research on intergenerational mobility finds that immigrants have a distinct economic advantage when it comes to upward class mobility for their children than children of US-born citizens. Chetty et al directly builds on this research, referencing the work of Abramitzky et al, in addition to comparing the Census and tax return data to the US Gross Domestic Product (GDP) to estimate absolute income mobility. However, Chetty et al differ from Abramitzky in their conclusion that, while it is true that generally immigrant children in the second generation earn more than their parents, this upward trend fades in third and higher generations, and children are more likely to remain in the same economic class as their parents or earn less.

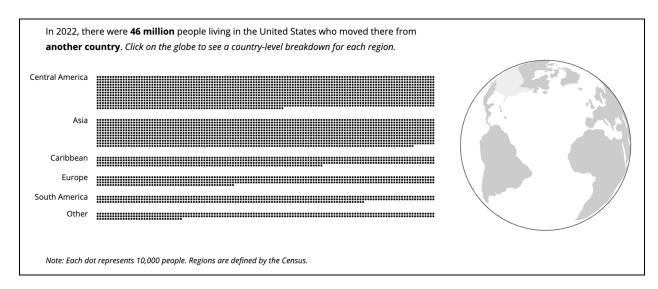
With our final project, we iterate on this previous work, as well as present the summations of our findings in a visual and accessible format. These previous works were targeted towards an audience of academics or policy-makers. We present compelling visualizations that allow the non-expert reader to better engage with the data. We also provide an updated analysis with more recent Census data.

## 3. Visualization Walkthrough

Link: https://naitian.github.io/ngen-website

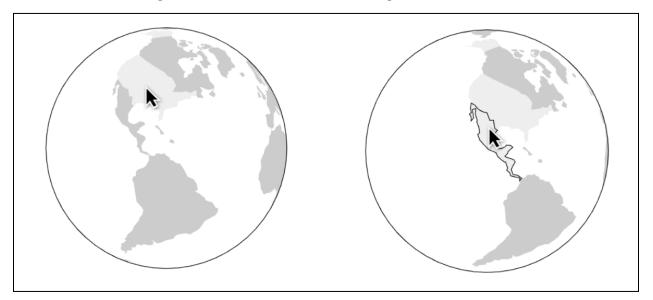
### Introduction to the topic

The first visualization serves as an introduction to the general topic of immigration to the United States. We use it to give readers an idea of the extent to which immigrants comprise the U.S. population before drilling down into the specific population that we are covering. The globe on the right allows users to select global regions and see a country-level breakdown of the number of immigrants from each country. The top regions are displayed by default, and clicking on a region displays the top 10 countries within that region, with the rest aggregated into an "Other" field so as to maintain the layout of the page.



The default view of the globe with no regions selected.

The United States is grayed-out on this map to emphasize that it cannot be selected as an option, since this is a visualization of people moving to the U.S. from other countries. To further hint to users that this is the case, there is no hover interaction for the U.S., whereas the other regions will receive a stroke and a lighter fill on hover.



Hovering over the United States (left) compared to hovering over Central America (right).

We combine this visualization with a dynamic caption; when a region is selected, the caption updates with numbers for the specific region. The dynamic sections of the caption are bolded to draw attention to the fact that they can change.

	nere were <b>15 million</b> people living in the United States who moved there from on the globe to see a country-level breakdown for each region.	
India China Philippines Vietnam Korea Japan Bangladesh Iran Pakistan South Korea		
Note: Each d	ot represents 10,000 people. Regions are defined by the U.S. Census Bureau.	

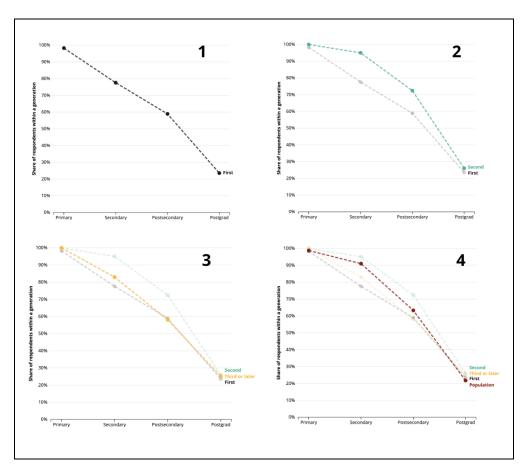
The visualization after the user clicks on a region

This serves to better contextualize the data and provide a discrete number in addition to the rough estimate provided by the visualization.

### Education

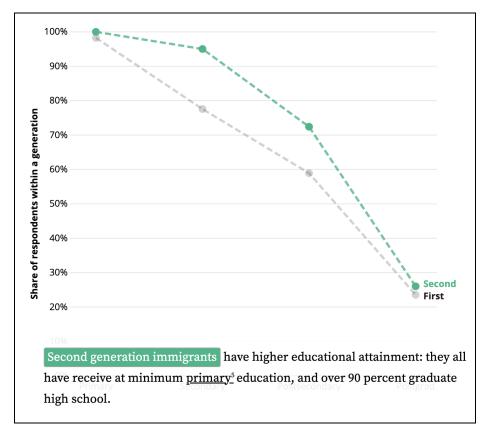
The section exploring educational attainment works with scrollytelling. We chose to use scrollytelling so that we can introduce the data one generation at a time. This way, we can point out trends within each generation that are notable, as well as compare between generations in a guided way. When we introduce a new group, we lower the opacity of the previous ones to highlight the current generation being discussed.

Lines are drawn between points as a method of gestalt to group generations together, making trends that are present within each generation easier to compare across generations. We present the levels of educational attainment as an ordinal variable, since someone with a high school diploma necessarily also has a primary education. However, keeping in mind that this is still a categorical variable, we draw dashed lines so as to discourage interpolation.



The four steps of the visualization, stepping through first, second, third-and-later generations, as well as the Bay Area average, labeled 1-4 by the order in which the frames occur.

The generations are encoded in the color channel, and the inline labels share this coloring to emphasize the association. Further, the copy in the scrolly boxes are highlighted with the same color-scheme, again serving to disambiguate which line corresponds to which generation.



Text in the scrollytelling boxes are also highlighted according to the generational color scheme.

The lines are labeled by text at the end of each line, allowing readers to easily tell which line corresponds to which generation without shifting focus to a legend. This also allows us to change line opacity without having to include multiple versions of the same color in the legend. However, naively placing the labels by the line will result in overlapping text that's difficult to read. We use D3 force simulations to calculate the positions of text labels that are in close proximity to the lines and preserve ordering, but do not overlap.

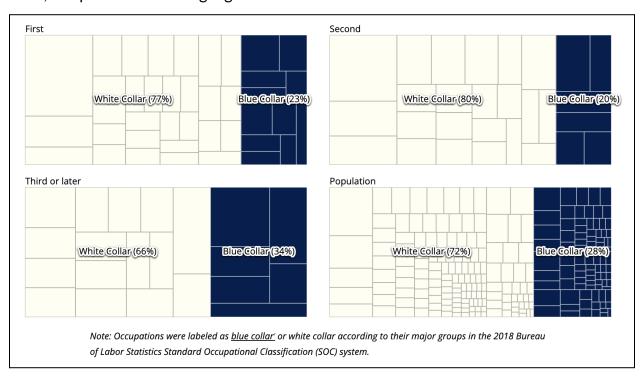


Text labels at the end of the line with hard constraints (left) and with relaxed constraints (right).

### Occupation

We explore occupation class in the context of blue and white collar jobs. This meant there was aggressive grouping, since the raw data has the specific occupations. In order to allow readers to get the main idea of class distribution while not totally obfuscating the distribution of actual jobs, we chose a hierarchical data visualization in the form of a tree map. It first shows the distribution of occupation classes within a generation, then the distribution of the jobs that fall under each class.

The colors for the first level of the hierarchy correspond literally to the "white" and "blue" collar classifications. To further drive home this point, we label the areas on the chart, and provide a color highlight in the text.

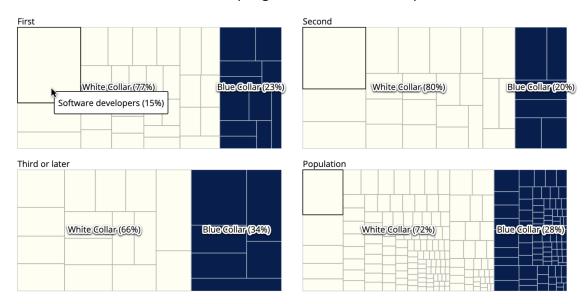


The occupation visualization without any interaction.

on Chinese immigrants have a larger proportions working in white collar industries, that from the seneral Bay Area pop

"White collar" is highlighted in the text with the same color as visualization.

We chose this small multiple set-up to allow readers to easily compare between generations. The colored areas correspond to proportions within each group. We also encode this information as a percentage in the label. The rectangular shape accentuates differences in widths, and invokes a "progress bar" visual metaphor.



Hover interaction for the occupation visualization

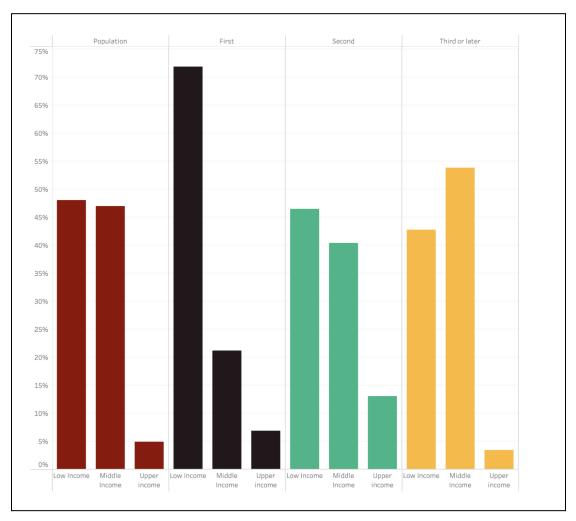
When a user hovers over one of the small multiples, a tooltip appears showing the specific job that is being represented by a given square, as well as the proportion of people within that group who have this job. The corresponding squares in the other groups are also highlighted. This allows readers to compare how prevalent a given occupation is between generations.

### Income

For this visualization, we wanted to compare Chinese immigrant generations against each other and the general population of the San Francisco Bay Area based on the percentage of each population that earns wages within the low, middle, and upper income classes. This data is presented as a bar chart and grouped by the colors previously established for each generation in order to increase the overall cohesiveness of the design. Each class represents a combined field of raw salary data from our dataset, that was then grouped according to the Pew Research Center's classifications for low, middle, and upper classes:

- Low Income
  - Less than \$52,200 combined household income
- Middle Income
  - \$52,200 \$156,600
- Upper Income

### More than \$156,600



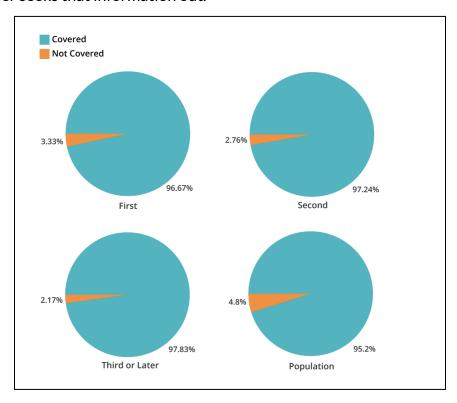
Static Income Visualization.

### Healthcare

With only two variables to compare when it comes to healthcare coverage—if a household is covered or not covered—we took a simple approach to representing data in this section. Our goal with this visualization was not necessarily to show a more granular view, compared to the other sections of the article, but a big picture takeaway that could easily be spotted and digested by the reader without much thought or effort.

Specifically, we wanted to draw attention to the fact that all Chinese immigrant generations and the general Bay Area population are well-covered, with minute differences between the four groups. However, as the reader moves their eyes across the charts, they can also see the slice of people who do not have some kind of healthcare coverage shrinks between subsequent generations, and increases as we get back to the

general population. We also included percentages to match with the other visualizations within the article and allow for granular comparison between these groups, if a more interested reader seeks that information out.



Static Income Visualization.

### Gradient of detail

The topic we chose to explore is relatively technical, with many precise definitions around which we've operationalized our analyses. Discerning readers may want to know the specifics of the analysis, but to introduce every definition in the text would make it less accessible to our intended audience of general readers. As a compromise, we incorporated an interactive element into the text: certain concepts are underlined and marked with a footnote. Hovering over this section of the text brings up a tooltip that includes more information about the term.

In this article, we focus on exploring the  $\underline{\text{American Dream}^1}$  for one small subset of immigrants: those who moved from China to the  $\underline{\text{San Francisco Bay Area}^2}$ . We

follow these immigrants, as well as later generation
United States to both immigrant and native-born pa

Within the scope of this project we are defining the Bay Areas as the following counties: Alameda, Marin, San Francisco, Santa Clara, San Mateo, Sonoma, Solano, Napa, and Contra Costa

Using Census data from March 2022—the most rece

collection— this project only represents a small, static slice of the Chinese

Tooltip appears when hovering over an underlined section.

At the end of the article there is also a "nerd box" that includes more detail about the dataset that we used.

**About the data:** We used data from the 2022 Annual Social and Economic Supplement of the Current Population Survey from the US Census Bureau. This is a nationally representative survey of the US population taken in March of every year.

### 4. Data

We used the Annual Social and Economic Supplement (ASEC) for the Current Population Survey for Social, Economic and Health Research (CPS) from the Census Bureau, because it has survey data on a variety of socioeconomic variables as well as foreign-born and parental birthplace information, which we can use to study immigrant generations.

The CPS is collected monthly, but the ASEC supplement is collected annually every March. We are using the March 2022 dataset, which is the most recent data for the ASEC. The Census provides both tabulations of the data and microdata, which contain responses from individuals, with *disclosure avoidance* techniques that prevent the identification of specific individuals. We use the microdata as prepared by the IPUMS project at the University of Minnesota, which processes datasets from a variety of national agencies and organizations and provides an interface to request subsets with specific variables.

To narrow the scope of the project, almost all of our analysis focuses on a smaller slice of the ASEC: Chinese immigrants to the San Francisco Bay Area. We chose this subset because of our familiarity with the immigrant population as well as the regional

relevance of the immigrant destination, but with the awareness that this subset of immigrant stories is not representative of all immigrant stories (nor is that a feasible goal to pursue.)

### 5. Tools

We used Python and Tableau for data preparation and exploratory data analysis, and we used Figma to create wireframes and mock-ups of the website. For creating the final visualizations, we used a combination of Tableau, Adobe Illustrator, and D3. We used HTML, CSS and Javascript to create the website, and we deployed it to Github pages.

### 6. Results

We developed a website that combines visualizations, narrative, and research-driven informational text. As part of the development process, we conducted usability tests on a preliminary draft of the website to understand what edits we should make.

### **Usability Testing Study**

The purpose of the usability study was to test three aspects of the website:

- Understanding: Does the viewer understand both the history and theory surrounding the issue of immigration and the American Dream, as well as our ultimate conclusion? Is it easy to understand for a variety of viewer demographics?
- Narrative: Does the narrative serve as a useful and impactful guide for viewers to understand our conclusions?
- Design: Is the design engaging, impactful, and supportive of the narrative?

In order to measure this our usability test consisted of first an unguided viewing of the visualization, with the viewer giving their first impression, understandings, and feedback as they walked through each part of the design, and a post-test that asked participants to summarize their takeaways about how the American Dream was defined, as well as how immigrant generations and education, occupation, income, and class are related. Finally, we asked participants to evaluate the success of the design, narrative, and glossary, as well as to note anything particularly engaging, educational, or surprising.

We recruited participants for testing who are young professionals within the San Francisco Bay Area. We wanted to survey educated readers who were not experts in immigration policy. We also wanted to test how our visualization came across to participants with and without immigrant backgrounds, so we selected two participants with immigrant ties and one whose family did not immigrate to the United States within recent history.

### **Usability Test Results**

### **Reader First Impressions:**

In the introduction, multiple participants were uncertain about what the dots represented in the visualization, since it is not contextualized until they scroll further down. Further, two participants did not realize the globe was interactive; the third participant noticed the interactivity, but was not aware of all of the available actions. For example, they did not know they could deselect a region after clicking on it.

The scrolly-telling in the introductory graphic was not clear for some participants; one participant said it was "overwhelming", and the graphic in the background distracted from the text. She found herself "looking between the two things."

The education chart was more successful, especially with respect to the scrolly-telling. One participant found that tying the findings to the scroll allowed her to digest the information better. Other participants found the lack of labels in the chart to be confusing.

The occupation chart was generally the most successful; participants understood the chart and used the hover interaction to explore the data more deeply. One participant used the chart to draw new conclusions from the data, pointing out that the *kinds* of white collar jobs were qualitatively different between the generations. One participant noted that it was unclear why the orientation of the white/blue collar subdivisions was different for 3+ generations compared to the others.

The income section was most confusing for participants. One participant noted that she wasn't sure whether the two charts were displaying the same information, and the relationship between the two was "a little nebulous". It was also unclear which chart the participants preferred: two preferred the bar chart while the other preferred the line chart.

Participants had few issues understanding the healthcare section because there was a small but clear trend and bar charts are a conventional way of representing the data. Multiple participants noted that the generational effect was quite small.

### Reader Takeaways

For each section, we asked the participants what their main takeaways were. For education, all three participants drew accurate conclusions. Two of the three participants explicitly stated the main point we had intended to demonstrate: that the third and later generations have a lower educational attainment than the first two generations.

For the occupation section, all three participants had accurate conclusions that matched the conclusion we intended to demonstrate: that the first two generations have higher proportions of white collar workers.

For the income section, the results were more mixed; one participant relied exclusively on the bar charts, while two others were unable to draw a conclusion about income from the visualizations. One participant mentioned that she relied on the surrounding text to a greater extent when viewing this section.

In the healthcare section, all three participants drew accurate conclusions that matched the conclusion we intended to demonstrate. Further, they all remarked that the size of the trend was small, demonstrating that they understood not only the relation between immigrant generation and healthcare coverage, but also the extent to which they are related.

### Other Findings

We asked users to what extent the narrative was successful in aiding their understanding, on a scale of (not successful, somewhat successful, no impact, successful, very successful). Two participants said somewhat successful, while one said successful, but qualified her statement. When asked to explain, participants mentioned that the family might be introduced too early, and it might be preferable to elaborate more on the narrative.

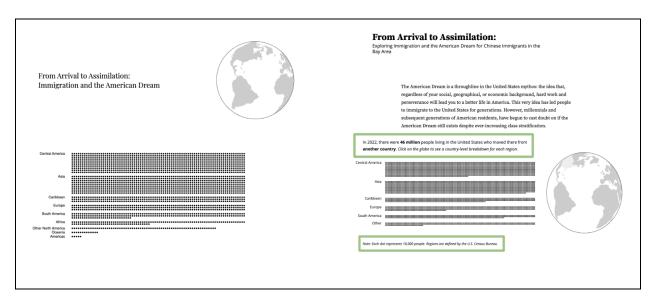
We also asked users whether they found the glossary to be successful in aiding understanding. Two participants said the glossary had no impact, while a third said that they liked the glossary because "it was nice to see how things were categorized and defined." Multiple participants mentioned that the glossary would have been more effective if it was not pushed to the very end.

All participants said they had learned something new from the article, and that they liked the spinning globe as a design element.

### Revisions to the Design

Based on the feedback we received for this draft, we made many revisions for the final draft. While many participants understood the key takeaways of the visualizations, they found some visualizations unintuitive or unclear. We cleaned up the visualizations, made style changes to increase legibility, and added more labels and interaction cues.

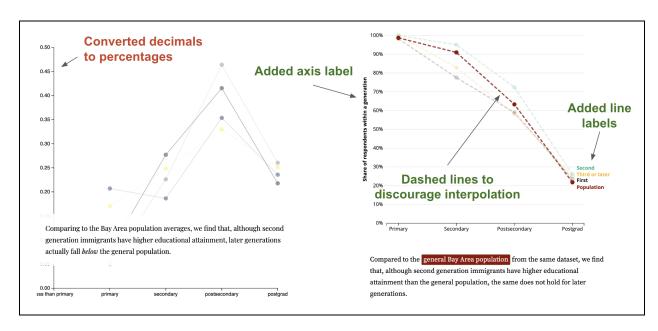
The introduction section was originally also a scrollytelly section, with the globe and dots taking up most of the initial screen. This was confusing to readers because the rest of the article focused on Chinese American immigrants in the Bay Area, but the big introduction was centered around national immigration patterns more generally. Some readers were also frustrated by the scrolly, remarking that the dots in the background were distracting from the text and didn't seem particularly relevant. Multiple readers were also unaware that visualization was interactive.



Before and after, with green boxes highlighting new elements in the visualization

In our revisions, we condensed the graphic so that it was more embedded in the text of the introduction. We removed the scrolling, choosing to instead make the interactive graph a static section of the page. We also added the caption to contextualize the data. One feedback we received was that the grouping of the countries seemed confusing or arbitrary. We added a note to the visualization that these groups were defined in line with the Census Bureau.

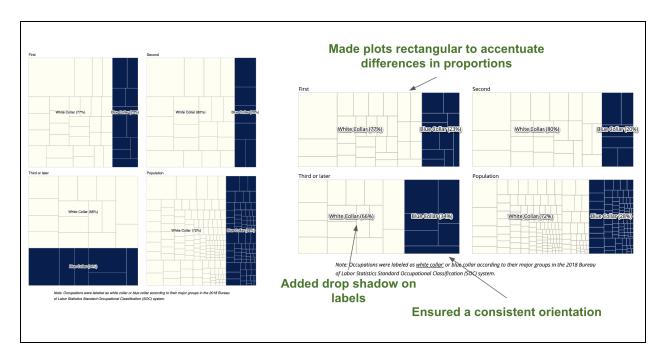
In the education chart, we received feedback that the scrollytelling was useful in highlighting the trends, but the colors were difficult to read and by the end of the chart, it was difficult to remember which lines corresponded with which generation. We also received feedback that it was unclear how the variables were encoded, and why lines were being used to link between levels of categorical data.



Before and after, with significant changes labeled

We revised the education chart by adding labels for the y-axis and the lines; we also made the lines thicker to be more legible and dashed to discourage interpolation. We also changed the colors and added a y-grid to improve legibility. Finally, instead of reporting the percentage of each generation with a given highest educational attainment, we report the percentage of each generation who had achieved a given level of education. The monotonically decreasing lines give a better picture of how educational attainment changes between generations.

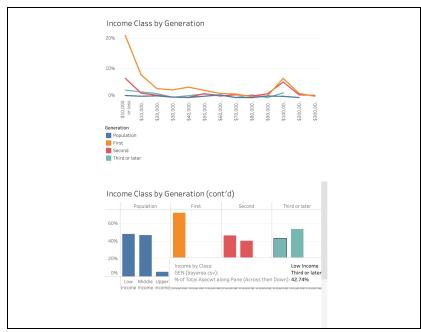
The occupation visualization was the most successful, but there were still points of confusion for readers. First and foremost, they were unsure why the third-generation plot was a different orientation (this was an artifact from using d3-treemap on a square plot). We also received feedback that the text labels were difficult to read, and it was hard to compare the areas of the charts.

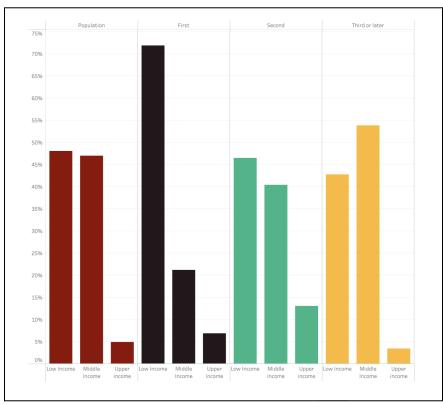


Before and after, with significant changes labeled

Addressing these changes, we made the treemaps rectangular. This served two purposes. First, this change fixed the orientation issue automatically. Second, and more importantly, it makes the proportions easier to compare. Because the chart is shorter, the differences in widths of the colors become accentuated. The rectangular shape also evokes the visual metaphor of a progress bar, and it is easy to see how far each rectangle is from being filled up. We also added a drop shadow to the labels to make them easier to read.

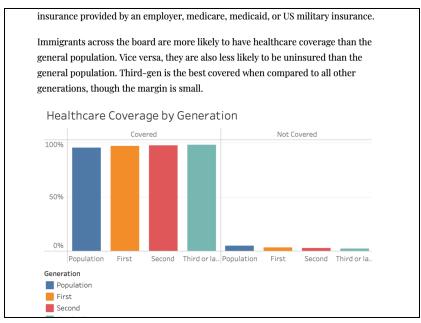
Additional feedback on our visualizations also commented on the confusing nature of including both a line chart and bar chart to show differences in income class across generations. We removed the line chart altogether and focused on the bar chart for showing the percentage of each generation that belonged to an income class. To aid in presenting this narrative in a more straight-forward manner, the colors were also adjusted to match the colors for the different generations introduced with the first visualization on the page.

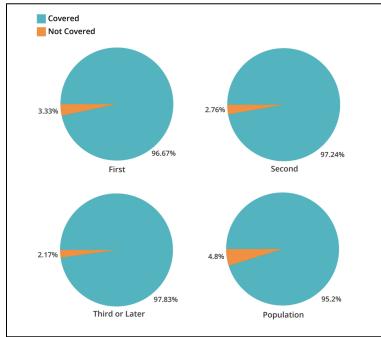




Income Visualizations: Before and After

For the section on Healthcare, we received feedback that the differences between the generations was hard to perceive in bar chart form since they were so small. We changed the original bar charts to pie charts since we were visualizing only two discrete states, as well as added the percentages to the side for readability and understanding.

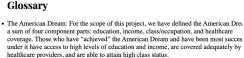




Healthcare Visualization: Before and After

Readers also appreciated the inclusion of more details in a glossary at the end, but remarked that it wasn't particularly useful because the definitions weren't available when

they were needed most: when the terms were introduced. To address this, we moved the glossary definitions into an interactive tooltip that appears when the user hovers over a term.



Generations: In this project we define four main categories of immigrant generation: 1, and 3+. The first generation are the original immigrants to the United States. 2nd-gen or of their children whether native-born or born outside the US. The 3rd generation is con of the grandchildren of the 1st generation, born within the US and any subsequent child

Bay Area: Within the scope of this project we are defining the Bay Areas as the followin counties: Alameda, Marin, San Francisco, Santa Clara, San Mateo, Sonoma, Solano, Na and Contra Costa.

- Primary: Elementary and Middle School
- Secondary: High school
   Post-Secondary: Associate's Degree, Bachelor's Degree
- Post-graduate: Professional Degree, Doctorate

In this article, we focus on exploring the <u>American Dream</u>? for one small subset of immigrants: those who moved from China to the Ann Francisco Ray Area? We follow

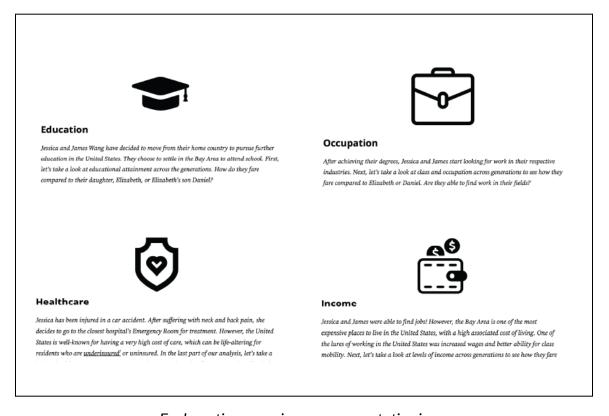
these immigrants, as well as later <u>generations</u> of it Bay Areas as the following counties: Alameda, Marin, States to both immigrant and native-born parents.

San Francisco, Santa Clara, San Mateo, Sonoma, Solano, Napa, and Contra Costa

sing Census data from March 2022—the most recent available at the time of data collection—this project only represents a small, static slice of the Chinese Immigrants who reside in the Bay Area instead of across the history of the United States. However, we aim to showcase the larger trends in the data and analyze what that means for an average Chinese immigrant household today and their ability to achieve socio-economic success.

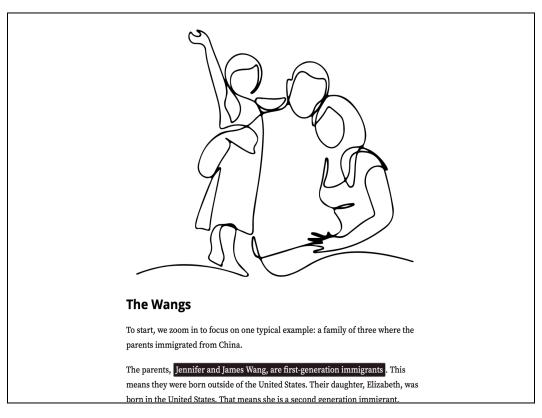
The glossary at the end is replaced by a showing the definition in a tooltip.

In order to increase engagement and visual interest, as well as support the flow of the article, icons were also added for each main section.



Each section was given a representative icon.

Additionally, all copy on the site was edited and/or rewritten for both clarity and engagement. One of our recurring comments from participants was that the narrative of the Wang family did not connect with them, or they forgot that there was a narrative serving as a throughline throughout the article. We added more copy to directly relate to that narrative so that it book-ended each section of the American Dream discussed, and added a section specifically to introduce the narrative, outside of just the introductory paragraph.



A new section was added to specifically support the narrative.

### 7. Link to Code and Visualization

Repository: https://github.com/naitian/ngen-website

Visualization: https://naitian.org/ngen-website/

Jupyter notebooks:

https://drive.google.com/drive/folders/1n3uMOtYycA7TC0ocGoIFX5rsmjawQUSw?usp=sharing

# Tableau workbooks:

https://drive.google.com/drive/folders/1QOCN2AVTx2LSUabw3gfmd8XWUsQIhZkL?usp=share\_link

# 8. Team Contributions

Category	Task	Nyah Mattison	Naitian Zhou
Preparation			
	Background Research	50%	50%
	Data Collection and Processing	30%	70%
Visualizations			
	Education	0%	100%
	Occupation	0%	100%
	Income	100%	0%
	Healthcare	100%	0%
User Testing			
	Test Design	50%	50%
	Testing	50%	50%
	Findings Analysis	50%	50%
Website			
	Development	40%	60%

	Copywriting	75%	25%
Final Report Writing		50%	50%
Average Contribution		50%	50%

# Appendix

Icons on the site from  $\underline{\text{The Noun Project}}$ 

### References

- Abramitzky, Ran et al. Intergenerational Mobility of Immigrants in the US over Two Centuries.

  National Bureau of Economic Research, 2019.
- Livingston, G., Kahn, J.R., 2002. An American Dream Unfulfilled: The Limited Mobility of Mexican Americans. Social Science Quarterly 83, 1003–1012. https://doi.org/10.1111/1540-6237.00129
- Raj Chetty, et al. "The fading American dream: Trends in absolute income mobility since 1940". *Science* 356. 6336(2017): 398-406.
- "Second-Generation Americans: A Portrait of the Adult Children of Immigrants." Pew Research Center's Social & Demographic Trends Project, Pew Research Center, 20 Aug. 2020,

  https://www.pewresearch.org/social-trends/2013/02/07/second-generation-amer icans/.
- Tan, T.X., 2016. Emotional and Behavioral Disorders in 1.5th Generation, 2nd Generation Immigrant Children, and Foreign Adoptees. J Immigrant Minority Health 18, 957–965. <a href="https://doi.org/10.1007/s10903-016-0388-0">https://doi.org/10.1007/s10903-016-0388-0</a>.
- Trevelyan, E., Gambino, C., Gryn, T., Larsen, L., Acosta, Y., Grieco, E., Harris, D., Walters, N. Characteristics of the U.S. Population by Generational Status: 2013, Current Population Survey Reports, US Census Bureau, 2016.