Workforce Demographics

Prepared by:

Emily Atkinson, Elsa Carlson, Ryan Gamilo, and Nolan Thomas

September 6th, 2022

Part I: Introduction to the Data

We took data from the Annual Business Survey 2019. Respondents answer general questions about their business, from the number of employees to the technology utilized.

Although respondents take one survey, the Census Bureau categorizes the data into 4 different sections, namely:

1. Company Summary - Provides data for employer businesses by sector, sex, ethnicity, race, veteran status, years in business, receipts size of firm, and employment size of firm for the U.S., states, and metro areas. Data for counties and economic places are available for 2018.

2. Characteristics of Businesses - Provides data for respondent employer firms by sector, sex, ethnicity, race, veteran status, years in business, receipts size of firm, and employment size of firm for the U.S., states, and metro areas, including detailed business characteristics. Data for counties and economic places are available for 2018.

3. Characteristics of Business Owners - Provides data for owners of respondent employer firms by sector, sex, ethnicity, race, and veteran status for the U.S., states, and metro areas, including detailed owner characteristics. Data for counties and economic places are available for 2018.

4. Technology Characteristics of Businesses - Provides data on technology use and production for Artificial Intelligence, Cloud-Based Computing, Specialized Software, Robotics, and Specialized Equipment technologies data at the U.S and State level for 2018.

Part II: Research Questions

We centered our focus around four main questions, all regarding the demographics of the employees and owners of different businesses. We decided to each explore a question within a different dataset, allowing us to understand this question from multiple perspectives.

1. Is there a relationship between the race of owners and the race of employees?
2. Does the demographic makeup for businesses opened before/after 1980s look different?
3. What does the demographic makeup of employees look like in relation to the longevity of the business?
4. What is the relationship between the demographic composition of a business and the business's specialized technology use?

Part III: Our Findings

In order to answer these questions, we all queried the Census API and cleaned the data into an appropriate format for our report. The ETL process can be found in our ETL report. Our findings are below.

Question 1: Is there a relationship between the race of owners and the race of employees?

To answer this question, we considered looking at the number of employees vs the number of owners grouped by race. Since the data from the API call comes to us as aggregated data, we needed to make extra rows to justify using a correlation coefficient and a scatterplot. To do this, we chose to include the industry in our API call and subsequently merged on that. The resulting graph is shown below:
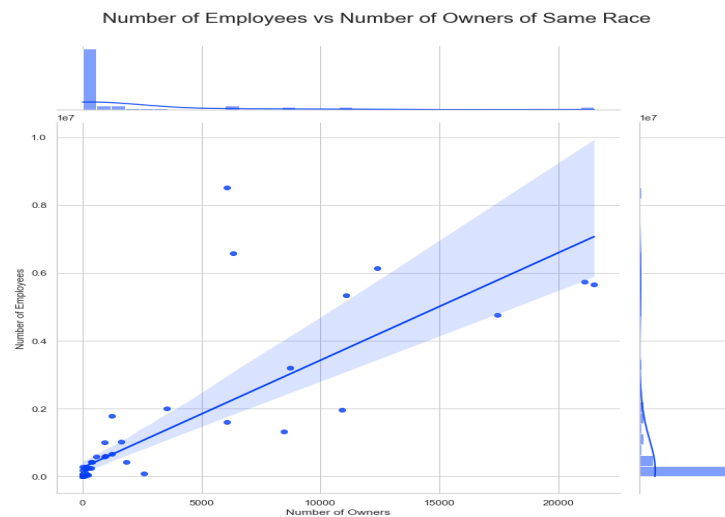

It appears that there is a strong positive correlation between the number of owners of a certain race and the number of employees of the same race. In fact, the correlation coefficient of this data is 0.81! However, if we look at the distributions of each variable on the top and right, we notice that many of the points are concentrated on the low end, with much of the the x-values (number of owners) actually being in the hundreds and their associated y-values (number of employees) actually being in the tens of thousands (note that our y-scale is in tens of millions!). When we color our points based on race, we get a revealing insight:

Number of Employees vs Number of Owners of Same Race
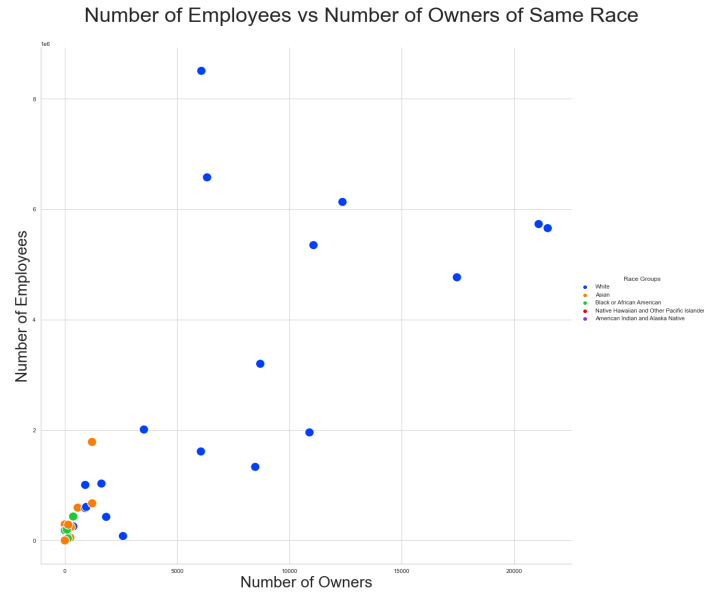
The heavy concentration of points in the lower left are actually made up of non-White races, and the points with White races and employees are actually outliers! This complicates our analysis as the data is now skewed. While we could split hairs by filtering out White, we were interested in how this affected the results on an industry and state level. As in, how does the distribution of race appear across industries and states? Here are the results across industry:
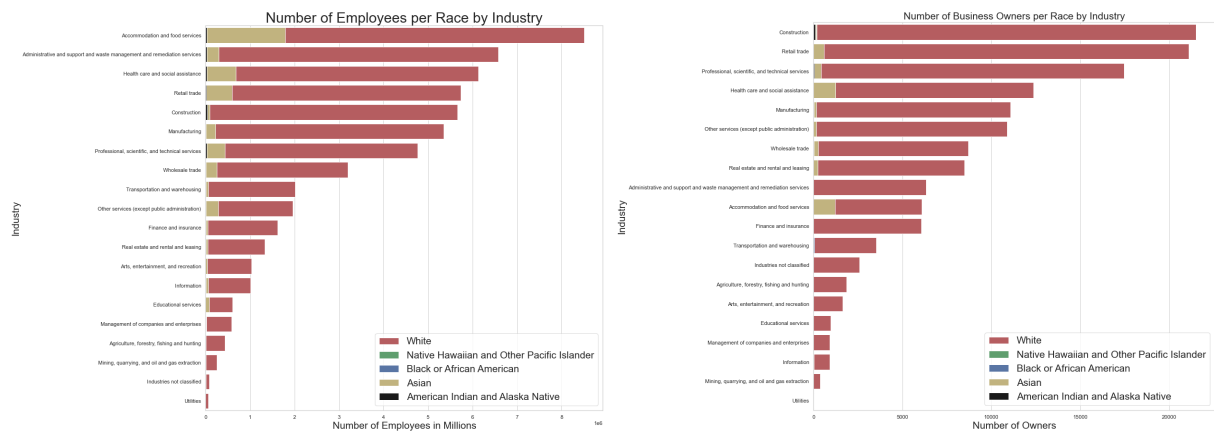
The lack of colors on these charts is not the result of a plotting error: it actually shows that there is an overwhelming majority of White people, whether owners or employees, across every industry. In addition, when we look at the look at the distribution of race across state:

Number of Employees per Race by State
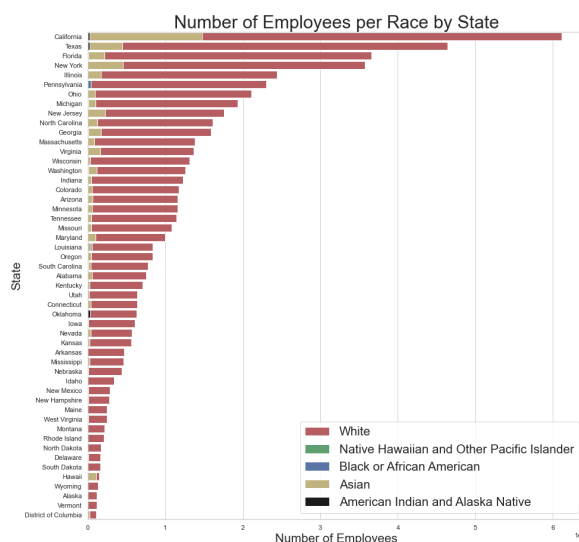
(Note: The census does not appear to aggregate Owner data by State.)

We see the same thing. In regards to why this is, we cannot use only the Annual Business Survey datasets to answer this. All we know is that there is no data for owners and employees of other races besides what is taken from the API call.

So is there a relationship between the race of owners and the race of employees? We can only claim with our current datasets that there *seems* to be a correlation between the two. However this analysis is complicated by the overwhelming majority of White owners and employees, which in itself becomes a difficult issue of whether to keep that data or not, as well as the fact that there is no variable conveniently linking a business owner to the race of their employees. Thus, we require more research with different datasets to sufficiently answer this question.

Question 2: Does the demographic makeup for business owners look different depending on when the owners acquired their business?

We answered this question by first looking at the overall makeup of the business owners, by year of acquisition and then by race. Unsurprisingly, there are more business owners who acquired their business around 10-20 years ago than those who have owned their business for either 1 or 30+ years ago.

Distribution of Business Owners by Business Start Dates
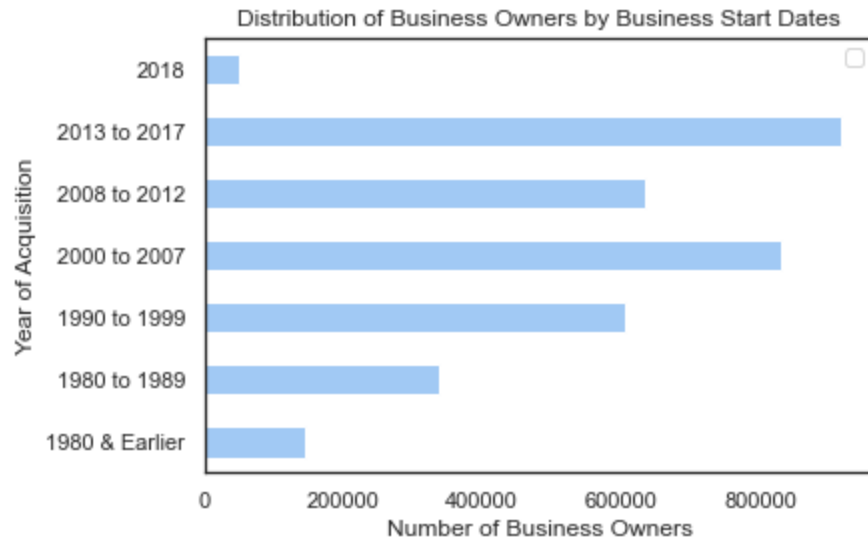
In looking at racial demographics, we observed that the census data divides race into categories of "Minority" and "Nonminority" as well as by specific race. We concluded that the "nonminority" category is distinct from "white" insofar as it excludes people of mixed-race heritage; this indicates that the numbers in any given racial category can include mixed-race business owners. In looking at the overall racial breakdown, we initially excluded the minority/nonminority category. Unsurprisingly, white business owners severely outnumber those of other races.



Distribution of Business Owners' Race
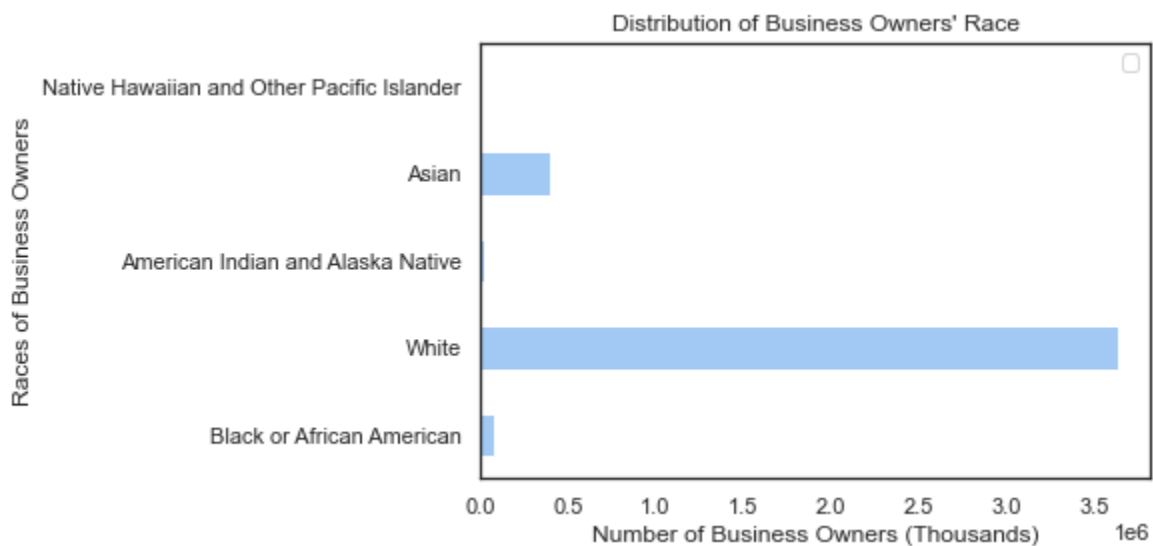
Our next question was how this changed for business owners who have owned their business. First, we used the minority/nonminority categories to see a simple graph that shows the year the owners acquired their business compared to their minority status. This date represents when the owner first acquired this business, no matter what–there are no details about what this means specifically. It could

represent either the date the business was founded, purchased, or inherited, for any given business. Once again, nonminority business owners significantly outnumber minority business owners. Notably, minority business owners have, on the whole, owned businesses for a much shorter time than nonminority business owners.
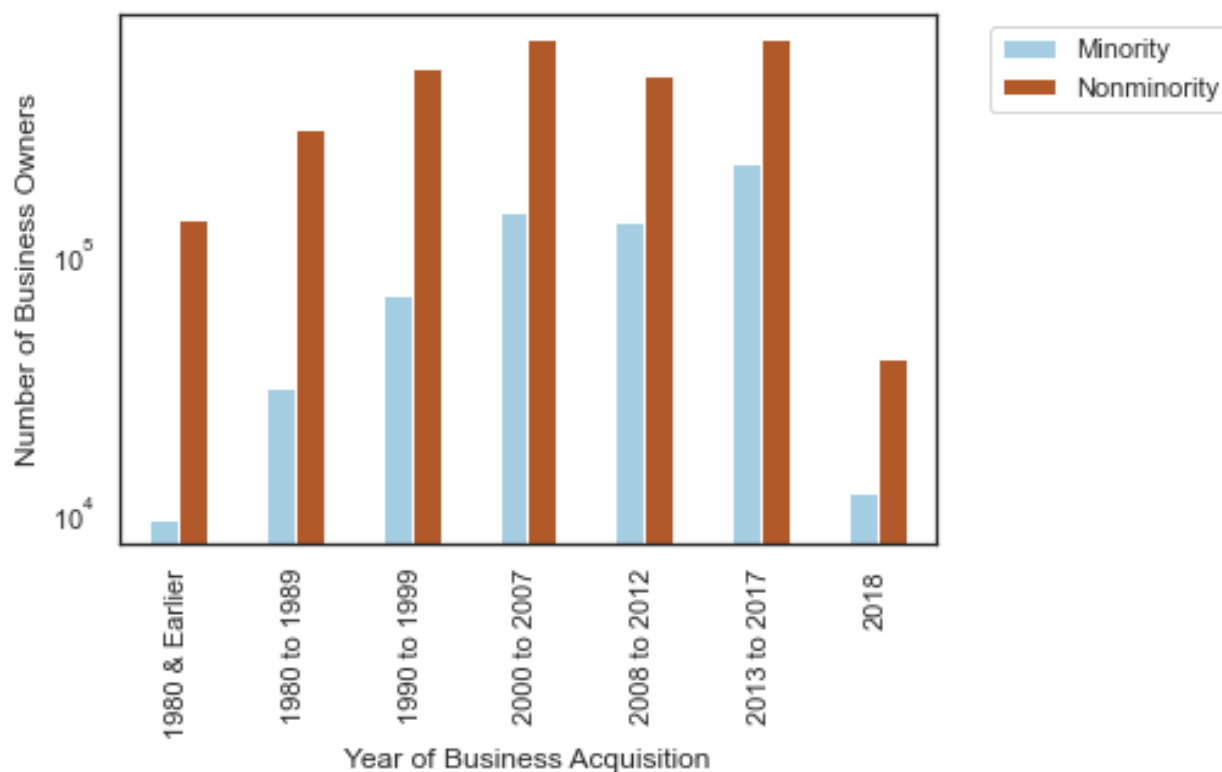

Finally, we created a visualization showing the breakdown of race more specifically by year of business acquisition. Once again, there is a significantly higher number of white business owners than any other race, followed by Asian business owners. The discrepancy between white and nonwhite business owners, while still significant, isn't quite as high as in the previous visualization–possibly due to mixed-race business owners being included in the category of "white." The context from the previous visualizations reinforces that while it is natural for business ownership to change over the years, white business owners have historically acquired more businesses and maintained ownership for longer. Our next step would be to investigate the factors behind the underrepresentation of minority business owners and how to ensure that nonwhite entrepreneurs have equal opportunity to pursue their goals.
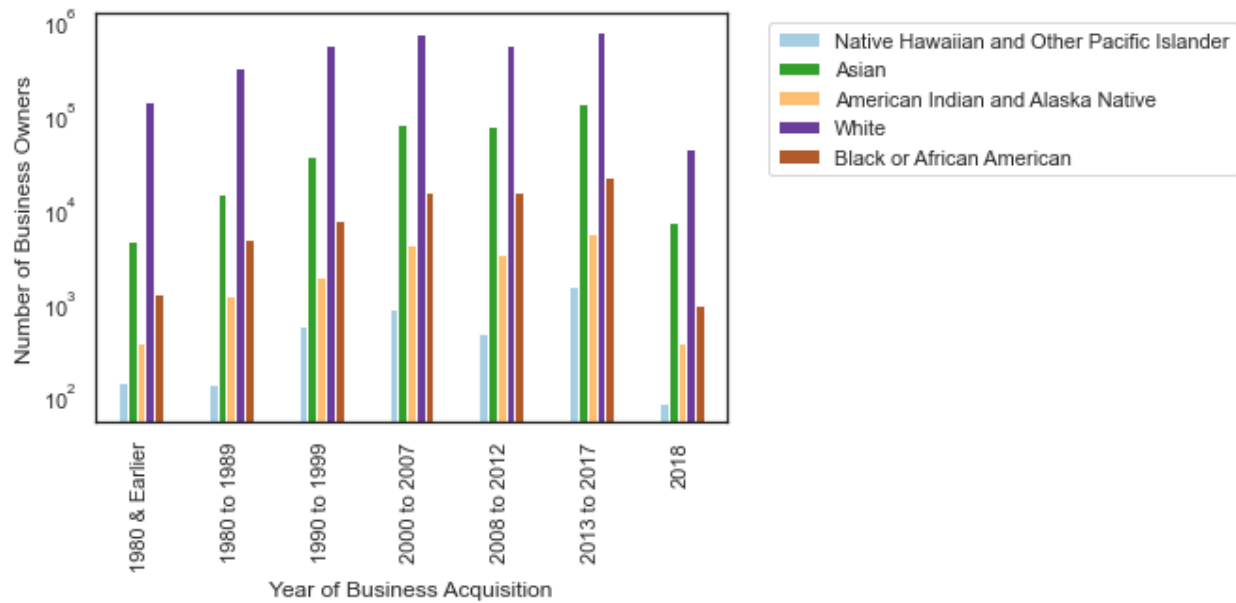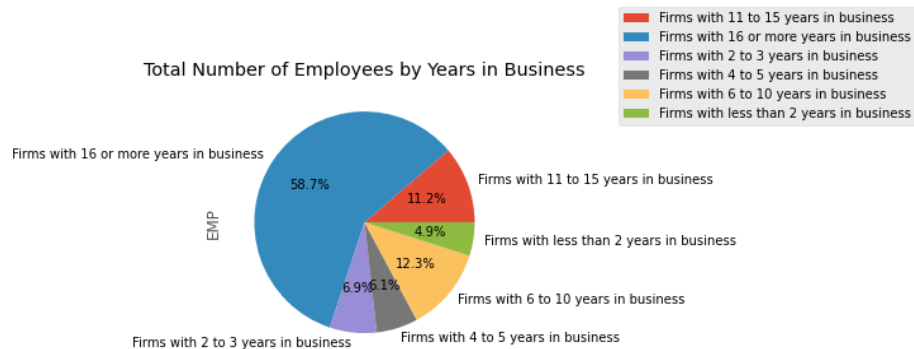
Question 3: What does the demographic makeup of employees look like in relation to the longevity of the business?

In order to answer this question, I started by running some basic information on the dataset. The first thing I looked at was the general breakdown of numbers of firms by years in business. The composition is shown below.
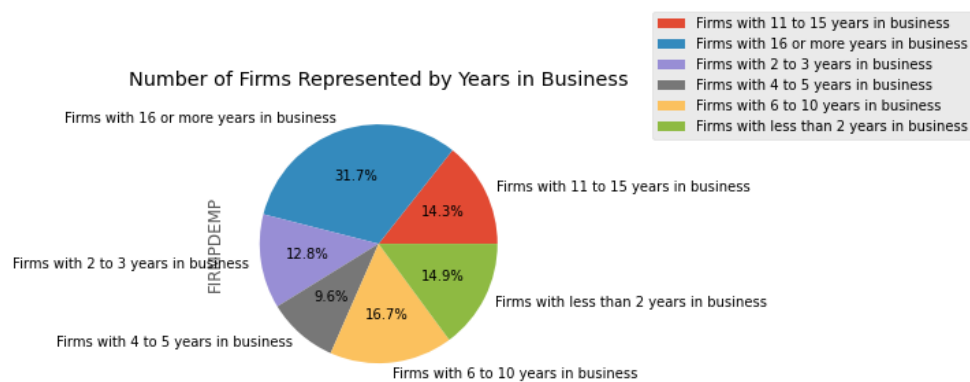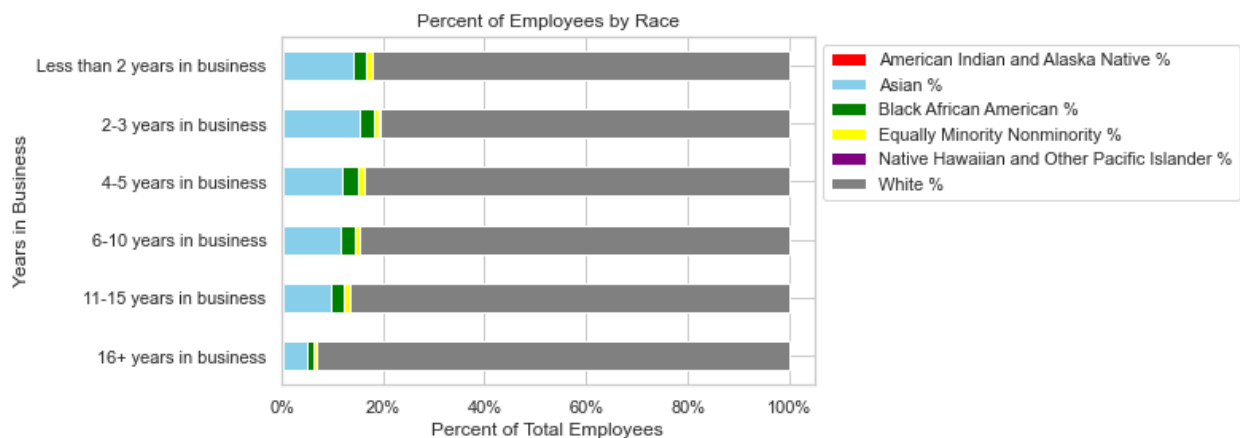

I was surprised to see that firms with 16 or more years in business make up the largest share of firms represented.
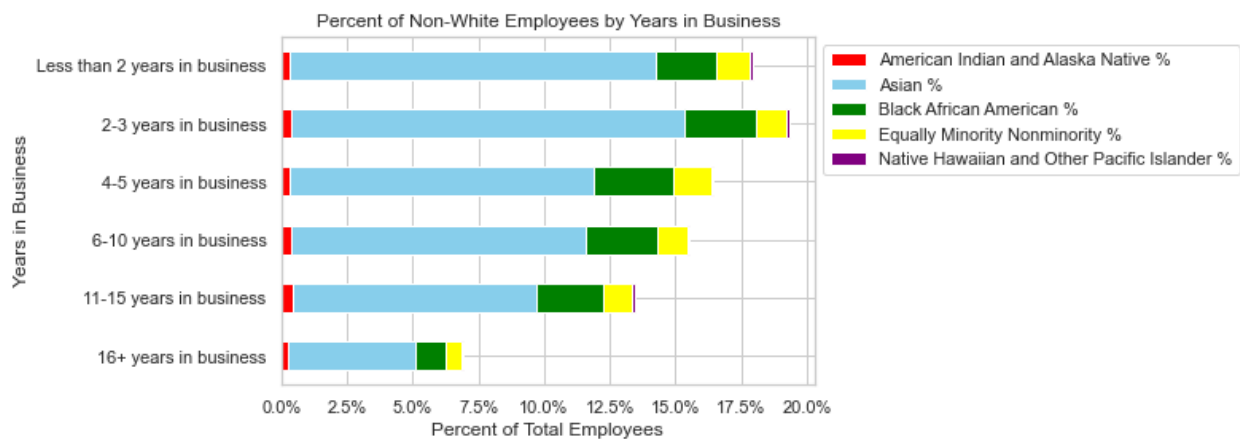
The largest number of employees also work for firms that have been in business for 16 or more years.

Total Number of Employees by Years in Business

Firms with 11 to 15 years in business
Firms with 16 or more years in business
Firms with 2 to 3 years in business
Firms with 4 to 5 years in business
Firms with 6 to 10 years in business
Firms with less than 2 years in business

Firms with 16 or more years in business 58.7%

Firms with 11 to 15 years in business 11.2%

Firms with less than 2 years in business 4.9%

Firms with 6 to 10 years in business 12.3%

Firms with 4 to 5 years in business 6.1%

Firms with 2 to 3 years in business 6.9%

EMP

After looking at the general breakdown, I decided to zoom in on representation of race by years in business. I started with a stacked bar chart that shows the percent of employees broken down by race and years in business.

Percent of Employees by Race

American Indian and Alaska Native %
Asian %
Black African American %
Equally Minority Nonminority %
Native Hawaiian and Other Pacific Islander %
White %

Years in Business

Less than 2 years in business
2-3 years in business
4-5 years in business
6-10 years in business
11-15 years in business
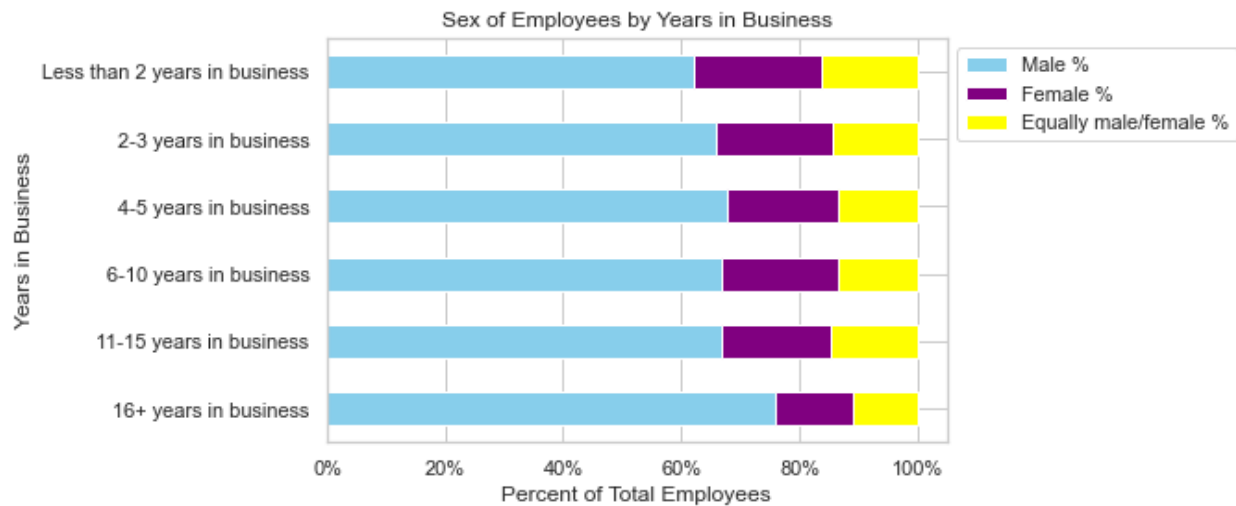16+ years in business

Percent of Total Employees

All businesses have majority white employees, making the breakdown of non-white employees harder to visualize. I created a second stacked bar chart with White employees filtered out.

Percent of Non-White Employees by Years in Business

American Indian and Alaska Native %
Asian %
Black African American %
Equally Minority Nonminority %
Native Hawaiian and Other Pacific Islander %

Years in Business

Less than 2 years in business
2-3 years in business
4-5 years in business
6-10 years in business
11-15 years in business
16+ years in business

Percent of Total Employees

The general trend, as shown above, is that the longer a business has been around, the less diverse its employees are. While longevity doesn't necessarily mean success, it is certainly a meaningful indicator

for a business. It is surprising to see the clear decline in percentage of minority employees for businesses that have been open longer.

I also wanted to see if this phenomenon persists across breakdown by sex.

While following a less stark pattern than the breakdown by race, business open for 16+ years are still the least diverse amongst all the businesses. There are certainly many factors that contribute to this, and more research is needed to delve into why companies that have been open longer are the least representative of the diversity of America's population.

Question 4: What is the relationship between the demographic composition of a business and the business's specialized technology use?

To answer this question, we began by getting the baseline information of total number of employees in the Technology Characteristics of Businesses database based on the Race Groups of White, Aisan, Black/African American, Native American/Alaska Native, Hawaiian Native/Pacific Islander, and Equally Minority/Non-Minority. The below chart represents total employee counts included in questions regarding specialized technology use:
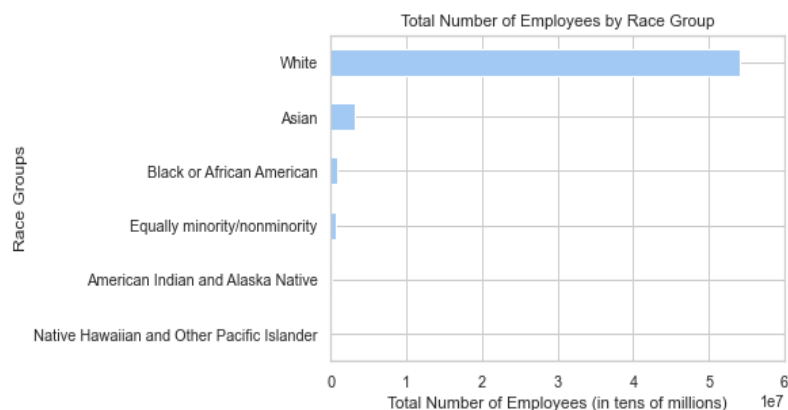
With the large gap in workforce representation between the White race group, and all others, we created a second visual for a closer look at the categories. Note that this second visual represents the employee population in millions, rather than tens of millions like the previous visual.
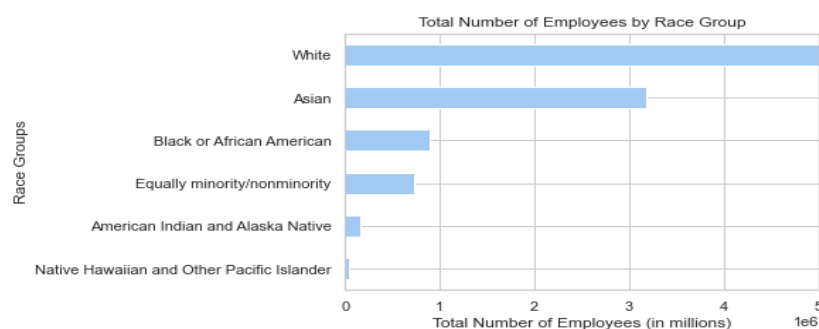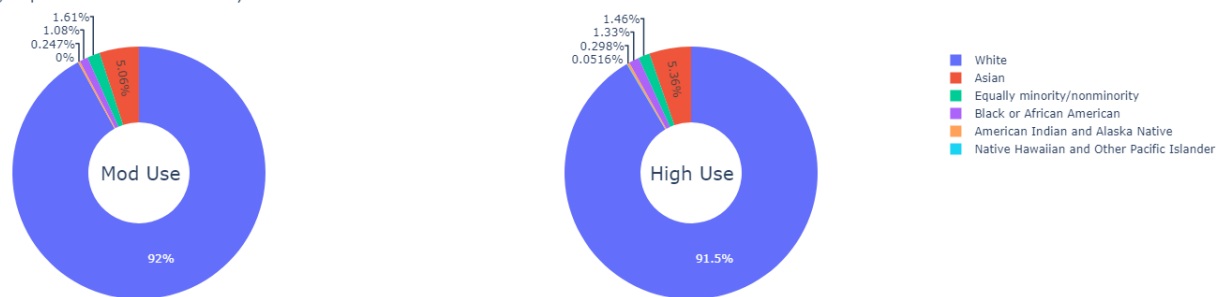

With the baseline information set, we started to dig a little deeper into the specialized software statistics of the dataset. In looking at the Impact of technology use on the workforce, we have shown the increased skill level of employees in using specialized software by race:

As with the previous visuals, due to the large gap in number of White employees versus all the other reported race groups, we created a second visual to more closely show representation of all of the other races. Please note that the graph on the left is represented in millions, while the graph on the right is represented in hundreds of thousands.

Now that we see the amount of increased skilled workers due to specialized software, the last visual that we wanted to create was a chart to see the percentage of employees who report moderate or high use of specialized software. As with the previous charts, we see an overwhelming percentage of employees under the white race group. To get a closer look at the other races outside of the white race group, we reset the charts to reflect the percentages of employees with the white category removed.

As with the previous charts, we see an overwhelming percentage of employees under the white race group. To get a closer look at all the other races outside of the white race group, we reset the charts to reflect t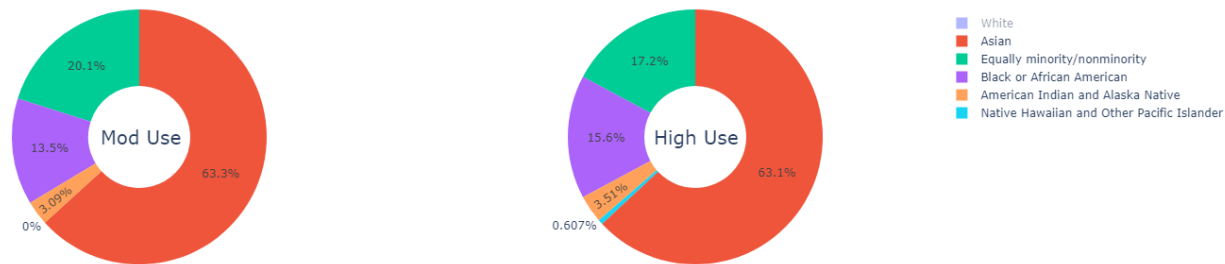he percentages of employees with the white category removed. This graph below represents the breakdown of the remaining 8-8.5% of employees who report moderate to high use of specialized

software.

Moderate & High Specialized Software Use by Race

As shown above in all the figures the White race group is overrepresented, but also the Native Hawaiian and Pacific Islander population is grossly underrepresented in not only the workforce as a whole, but also when looking specifically at specialized technology use and improvement.

Part IV: Conclusions

Our project showed that across the board, white employees & white business owners are overrepresented in all sectors and in businesses of all longevities. The white population in business so exceeds that of any other race that it is difficult to see any meaningful information about employees or business owners of other races. Our data does not answer the question of why these disparities exist. However, it does give us some ideas of where to start, including: the lack of diversity in older businesses, high-use technology jobs, and resources that white vs nonwhite entrepreneurs & other business owners have access to. As a next step, we would investigate the reasons behind white overrepresentation and what factors lead to more equal representation in a given business.

References

US Census Bureau. (2022, July 25). *About the Annual Business Survey (ABS)*. Census.gov. Retrieved

September 5, 2022, from https://www.census.gov/programs-surveys/abs/about.html

United States Census. (2019). *Annual Business Survey* [Dataset].

https://www.census.gov/data/developers/data-sets/abs.2019.html