Census Business API Report

Adam Brewer, Anthony Rondos, Shannon Bayless, Stephanie Leiva
September 6, 2022

Data Sources

- State Populations Census File:
    - Contains national, regional, and state population data from 2010 to 2019.
    - US Census Bureau. (2021, October 8). 2019 National and State Population Estimates. Census.gov. Retrieved September 2, 2022, from https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-total.html
- State Code File
    - Contains a state name, and state code column used for making the map with Plotly.
    - List of State Abbreviations (Download CSV, JSON). Retrieved September 3, 2022, from https://worldpopulationreview.com/states/state-abbreviations
- Census Business Data Tables
    - Contains four tables with surveyed data related to business operations within the U.S:
        - Table 1: Company Summary
        - Table 2: Characteristics of Businesses
        - Table 3: Characteristics of Business Owners
        - Table 4: Technology Characteristics of businesses
    - US Census Bureau. (2021, October 14). Annual Business Survey (ABS) APIs. Census.gov. Retrieved September 2, 2022, from https://www.census.gov/data/developers/data-sets/abs.2019.html

Initial Inquiries

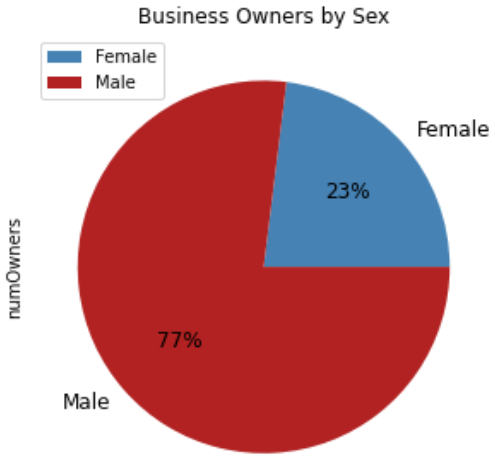
- Which groups are over/underrepresented in business ownership?
- What do business demographics look like across the U.S?
- What are the demographics of the most diverse state?
- Which Industries have the most disparity between groups?
- Where in the U.S do business owners tend to reside?
- Which state has the most firms relative to its population?
    - What are the demographics of that state?
    - Which technology was used in that state?
        - Which industry is that technology most associated with?

Analytical Processes

Which groups are over/underrepresented in business ownership?

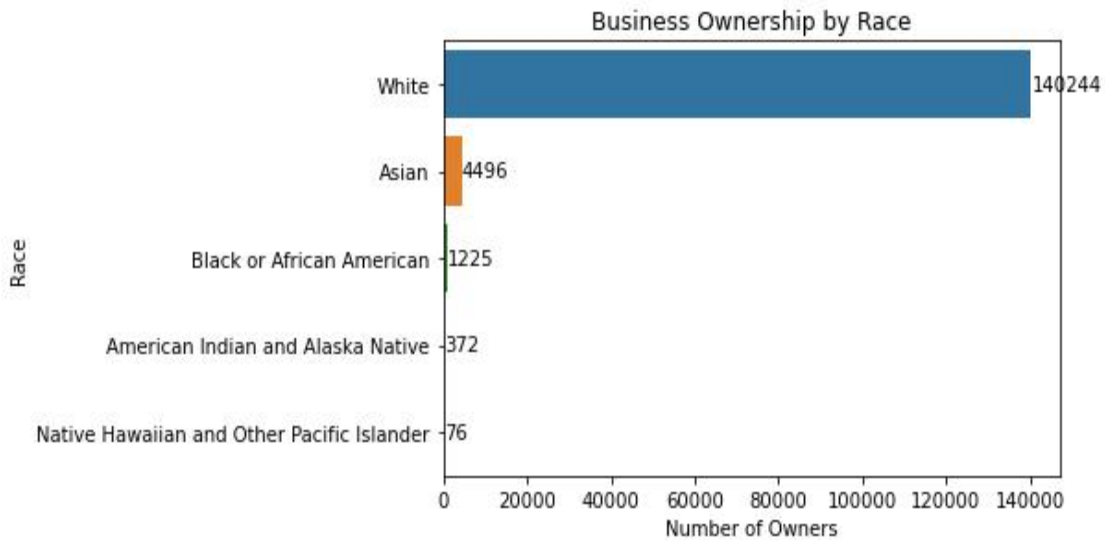
Business ownership is an important metric that can help measure how well a demographic group is doing in comparison to the population at large. Businesses are often seen as the premier method to generate wealth and climb the socio-econinomic ladder, so making sure all societal groups have equal opportunities to found an enterprise is important for generating a fair society.

To compare business ownership between the sexes, we summed the number of respondants to the Census Business Ownership Survey grouped by by sex and created the following pie chart:



Historically, women have often faced societal discirmination both formally and informally that have limited their ability to start businesses. This discrimination still exists to this day as represented by their significantly lower business ownership rates. Women own only 23 percent of business, compared to 77 percent which are male owned.

Minority racial groups also often face discrimination in the business world, and to better understand the degree of this discirmination, we summed the number of respondants to the Census Business Ownership Survey grouped by by race and created the following bar graph:

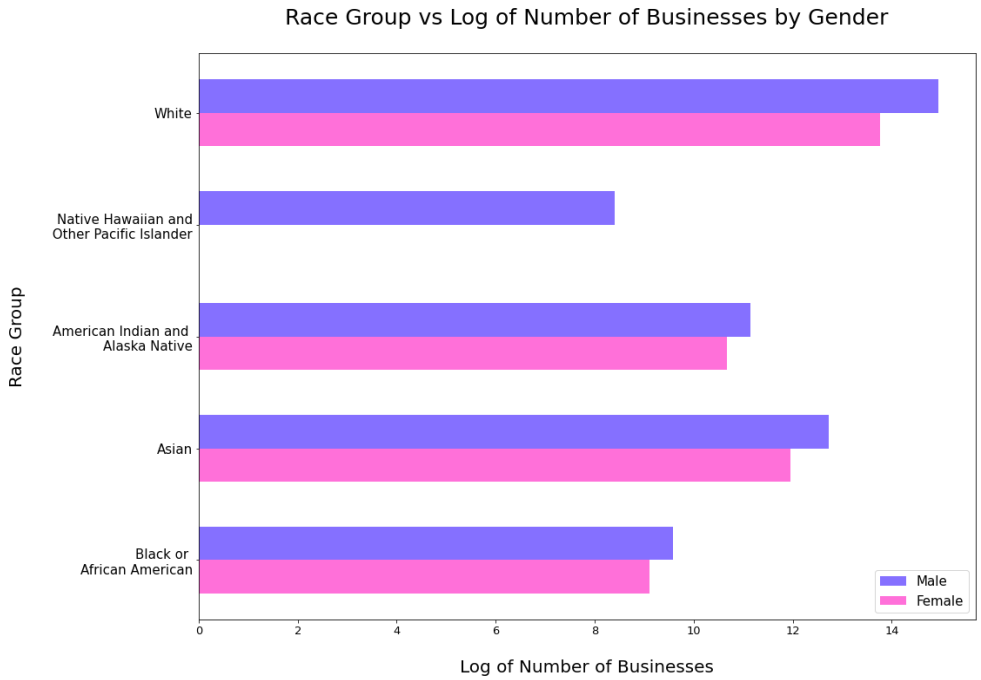


Firstly, the census collects data via survey, which is a notoriously inconsistent way to collect accurate data. Non-response bias can make it so that certain groups are less likely to reply to a survey, and end up having their data underrepresented. It is entirely possible that the underrepresented racial groups own business at a higher rate than it appears, yet are less likely to submit or receive census surveys. However, even in light of this, minority groups are extremely underrepresented in business ownership, and it is hard to imagine that any form of bias can solely explain this gap.

To get a clearer image of the demographics of business ownership, we could adjust the number of owners of each racial group in respect to their proportion of the total U.S population. Then, we could look at business owners per state or even per city, to get a more specialized view of diversity in a more specific region. This could give us insight on where the diversity or lack thereof is really coming from.

What do business demographics look like across the U.S?

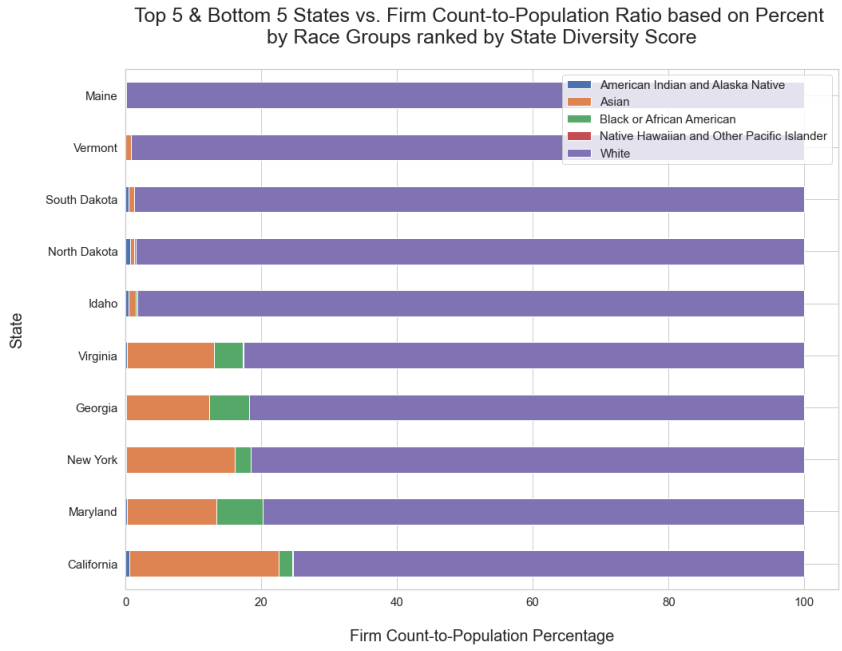
Moving on from the demographics of business owners, we then were interested to see the demographics of their employees. We created a bar graph to compare racial and gender groups based on the racial and gender breakdown of respondent businesses. We used a log function to represent the x-axis, as otherwise nonwhite groups would barely even show up on the graph as seen in the business ownership by race graph.



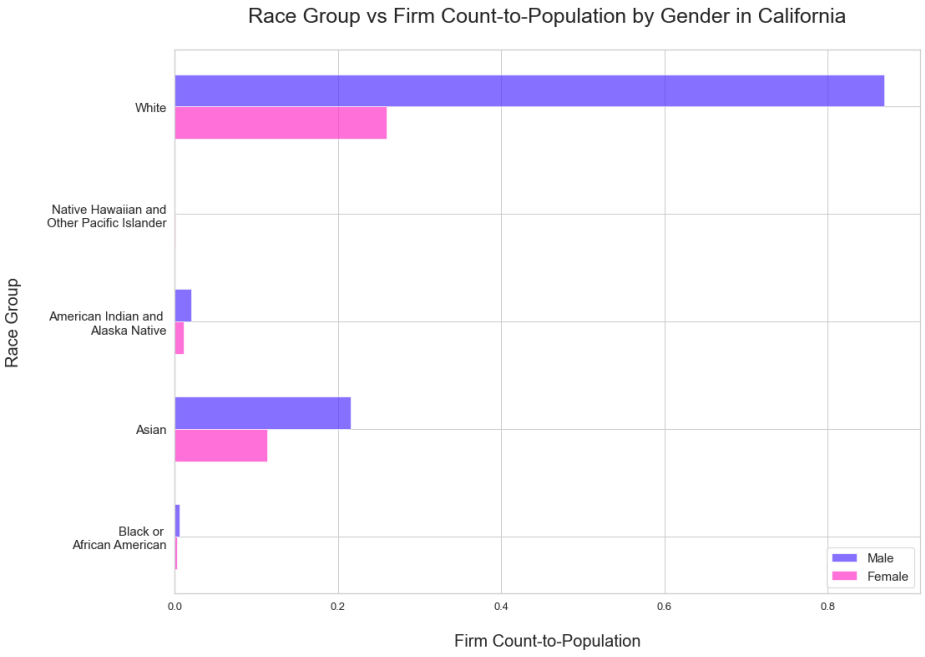
Once again whites are extremely overrepresented in the data. Obviously, this is in part due to the majority of the U.S population being white, and also factors such as non-response bias can be at play yet again, but even so minority groups are shockingly underrepresented once again. Contrary to business ownership however, women seem to generally have more equal representation in this dataset, although still less than their male counterparts.

What are the demographics of the most diverse state?

By adding up to .2 for each of the five listed races in respect to their portion of the population and subtracting the excess, we created an artificial diversity score that ranges from -.8 to 1, -.8 being the least diverse and 1 being the most. We can then create the following stacked bar chart to represent our results:



According to this, Maine is the least diverse state, so much so that it is nearly impossible to discern any minority population from the stacked bar graph alone. California is a hub for immigration for South America and Asia, and therefore is unsurprisingly the most diverse state in the U.S. We then can look deeper into the demographics of California respective to their total population and see how they compare to the U.S as a whole.



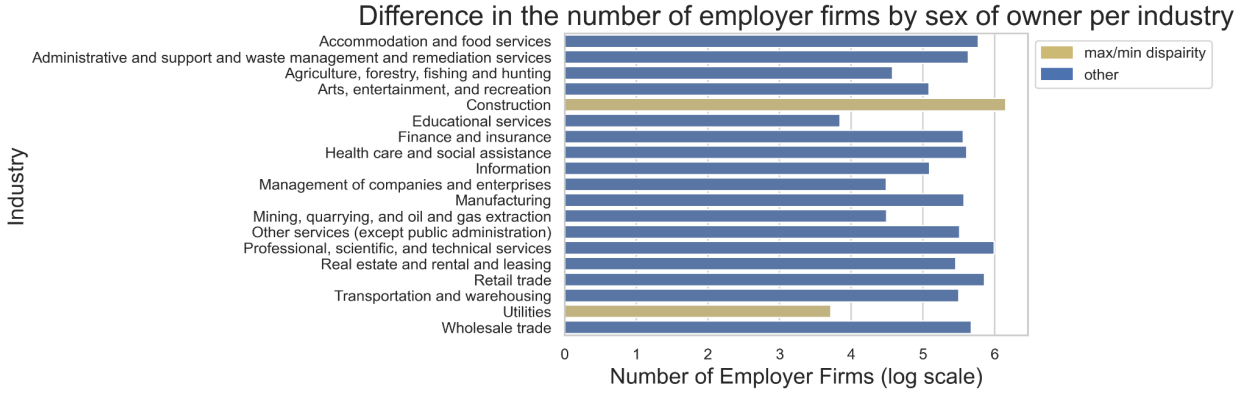
While the diversity at respondent firms is certainly better in California, it still is concerningly low considering that this is the most diverse state. Asian Americans seem to be better represented within this state, but other minority groups seem to be similarly unrepresented to the U.S as a whole. It is important to note that often hispanics are categorized as white in federalized demographic polls such as the census. To better understand diversity in California, and the U.S as a whole, we could do further research separating hispanics as their own racial group.

Which Industries have the most disparity between groups?

Through exploring this data, it is clear that there are business ownership disparities among many demographic categories such as sex, ethnicity, and race. In order to examine these disparities, we can look at both numeric differences in the number of businesses owned and a ratio of businesses owned by a majority group compared to those owned by a non-majority group. Using a ratio allows us to see disparities that are not skewed by the relative number of businesses in any given industry.

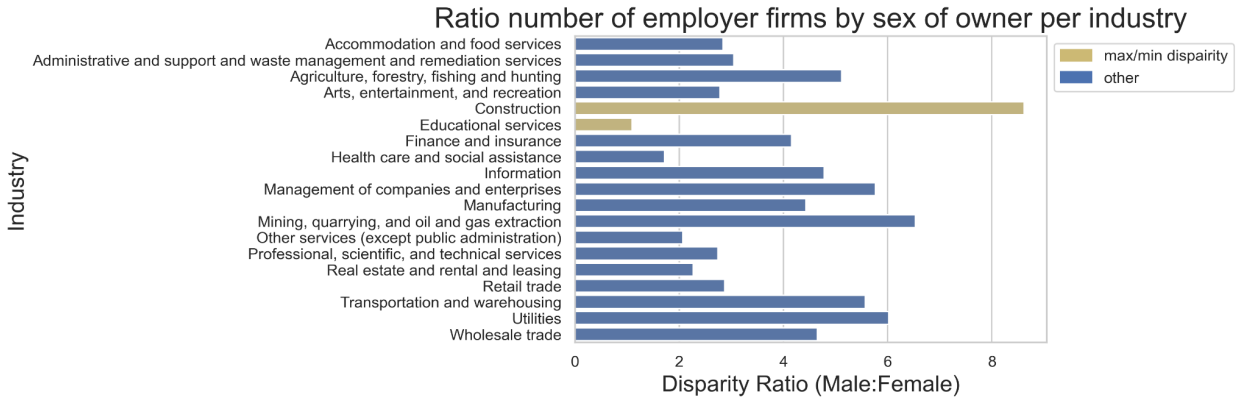
<u>Sex</u>

We explored the two sex groups of 'Male' and 'Female.' When looking purely at the difference between the number of firms owned by men and women, we can see that the largest disparity is in the Construction industry with men owning 1,921,683 more firms than women. The smallest numeric disparity is in Utilities, where men only own 6950 more firms than women.



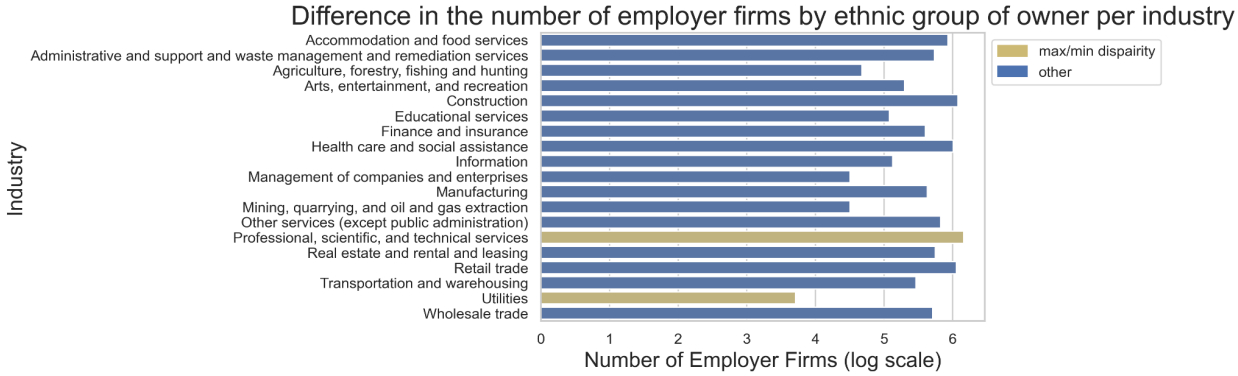
However, we can see more true disparity when looking at the number of businesses that men own per a single business that a woman owns. Below, we can see that the largest disparity still exists in the Construction industry, with men owning 8.6 businesses for every 1 business a

woman owns. The smallest disparity is in Educational services, with men owning about 1.1 businesses for every 1 business a woman owns.

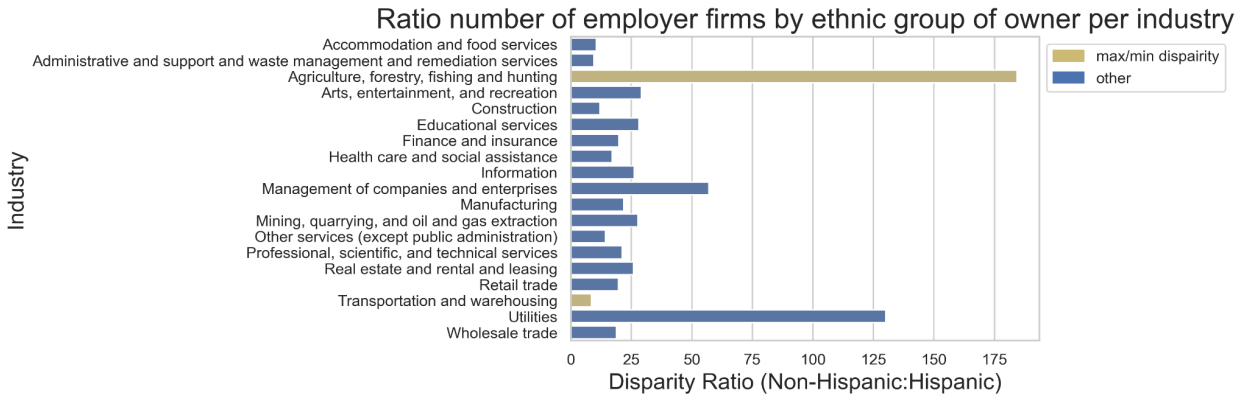


Ethnic Group

We explored ‘Hispanic’ and ‘Non-Hispanic’ when looking at the ethnic disparities in business ownership. When looking purely at the difference between the number of firms with non-Hispanic and Hispanic owners, we can see that the largest disparity is in the Professional, scientific, and technical services industry with non-Hispanic business owners owning 2,165,902 more firms than Hispanic business owners. The smallest numeric disparity is in Utilities, where non-Hispanic business owners only own 8016 more firms than Hispanic business owners.

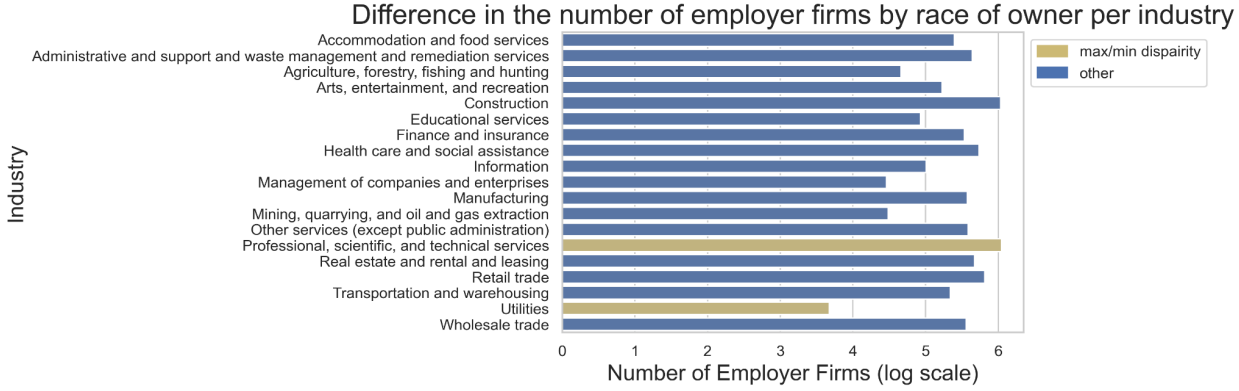


Again, we can see more true disparity when looking at the number of businesses owned by a non-Hispanic business owner per a single business owned by a Hispanic business owner. Below, we can see that the largest disparity is actually in the Agriculture, forestry, fishing and hunting industry, with non-Hispanic business owners owning 269.9 businesses for every 1 business a Hispanic business owner owns. The smallest disparity is in Transportation and warehousing, with non-Hispanic business owners owning abou 8.6 businesses for every 1 business a Hispanic business owner owns.

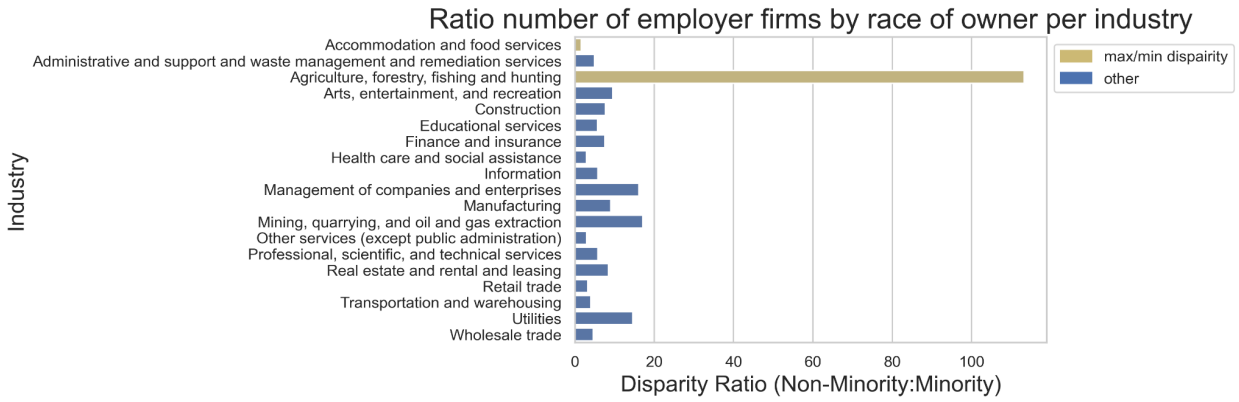


Race

We explored 'Minority' and 'Non-Minority' owned businesses when looking at the racial disparities in business ownership. When looking purely at the difference between the number of firms with non-minority and minority owners, we can see that the largest disparity is in the Professional, scientific, and technical services industry with non-minority business owners owning 1,661,851 more firms than minority business owners. The smallest numeric disparity is in Utilities, where non-minority businesses only own 7335 more firms than minority business owners. Again, we can see more true disparity when looking at the number of businesses owned by a non-minority business owner per a single business owned by a minority business owner.

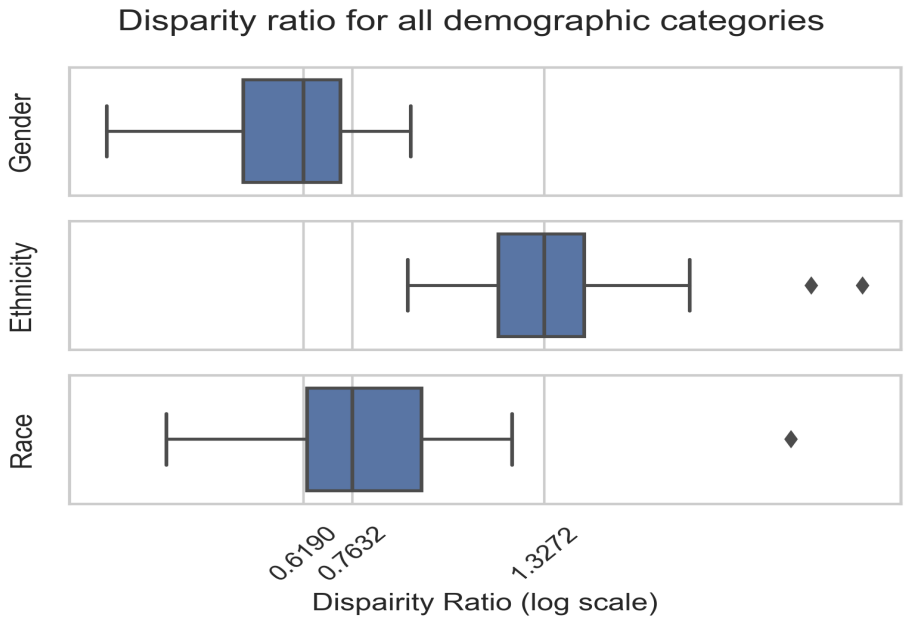


Below, we can see that the largest disparity is actually in the Agriculture, forestry, fishing and hunting industry, with non-minority business owners owning 166.1 businesses for every 1 business a minority business owner owns. The smallest disparity is in Accommodation and food services, with non-minority business owners owning 1.6 businesses for every 1 business a minority business owner owns.



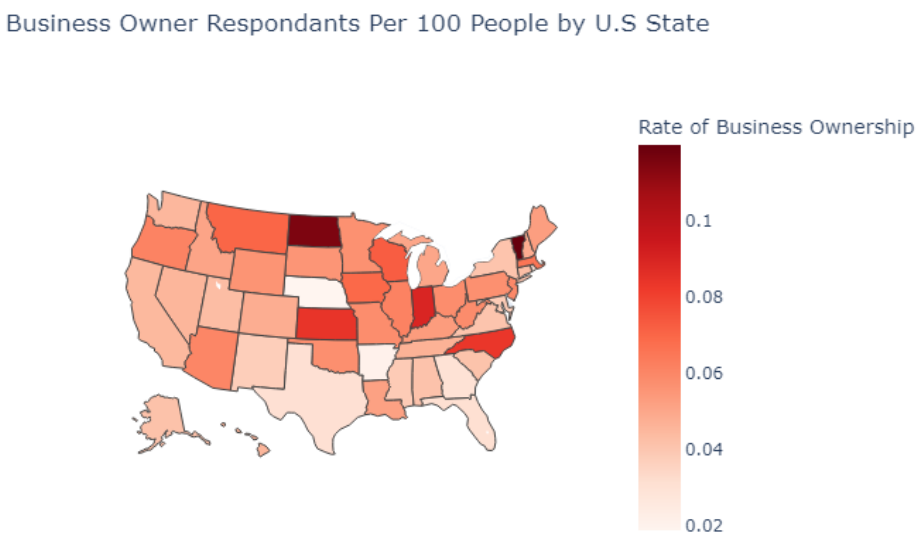
Below, we show a comparison of the disparity ratios among the 3 demographic groups (sex, ethnicity, and race). Note that the highest disparity ratio is in the ethnicity group, with a median ratio of about 21.3 (1.3 on a log scale) businesses with non-hispanic owners per every 1 business with a hispanic owner. The median disparity ratios for sex and race are about 4.2 (0.6 on a log scale) and 5.8 (0.76 on a log scale) respectively.

Overall, it seems that the Agriculture, forestry, fishing and hunting industry sees the most disparities when it comes to ethnic groups and races. The lowest disparity we see is between men and women who own businesses in the Educational services industry.



Where in the U.S do business owners tend to reside?

The U.S is a diverse country with many different ways of life from person to person and place to place. We considered that some states may have a more entrepreneurial aspect than others, and that would be represented by business ownership by state. To get data that better represented the average person in each state and avoid the more populous states obscuring the results, we merged the business data with a state population csv file. We then plotted the number of responding owners per 100 people on a map of the U.S:



Surprisingly, Vermont had the highest rate of business ownership, closely followed by North Dakota. Nebraska and Arkansas had the lowest rate of business ownership with the 46 other states and DC falling in between. However, much like possible bias in the racial data, the same could be said about possible Non-response bias and underreporting in certain states.

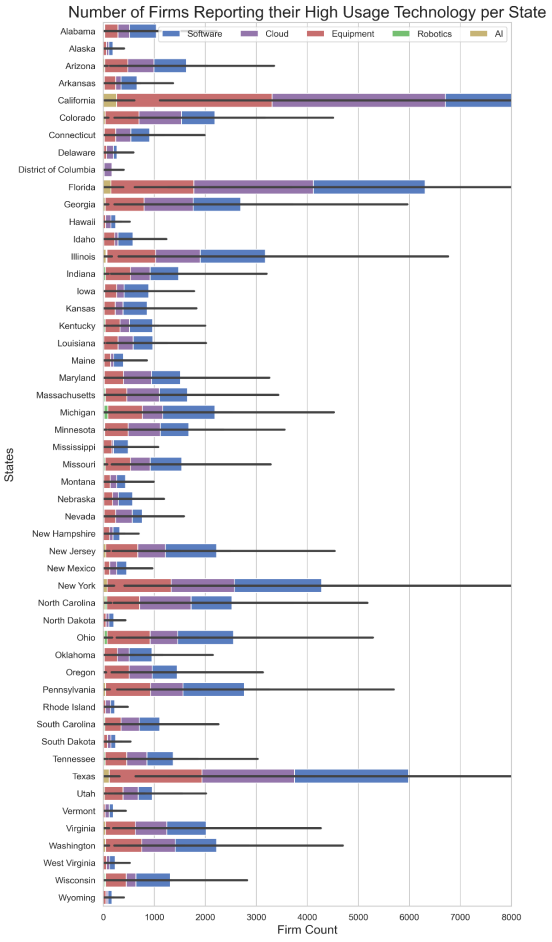
Which state has the most firms relative to its population? What are the demographics of that state?

There are many directions we could've gone in with analyzing the technology data. We could've looked at which technology was least used, or which technology was in total use in a certain state, or maybe zooming out to a nationwide scoop. We chose to structure our technology data analysis as a continuous zoom in on the top.

Before analyzing which state was to be our top state for analysis, we want to explain *why* we chose to zoom in on a state vs. looking nationwide and *how* we chose the top state to analyze.

Why we chose to zoom in on a state vs. looking nationwide

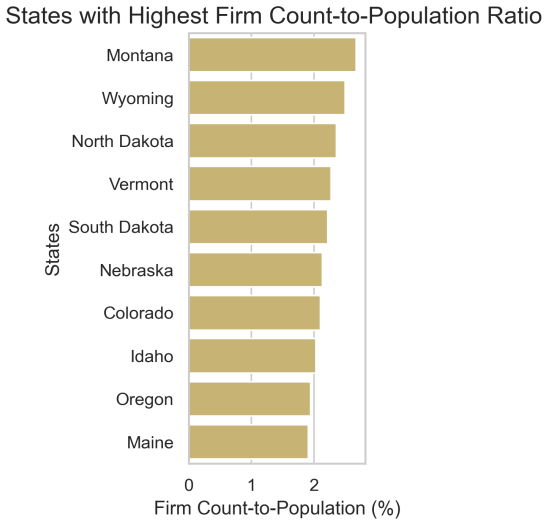
If we looked nationwide or looked at all states, it's quite overwhelming. The layered bar chart to the right shows how many firms categorized each technology as being highly used. We can see which technology is mostly used in each state but there's too much going on and we wouldn't have a clear direction on where to go next.

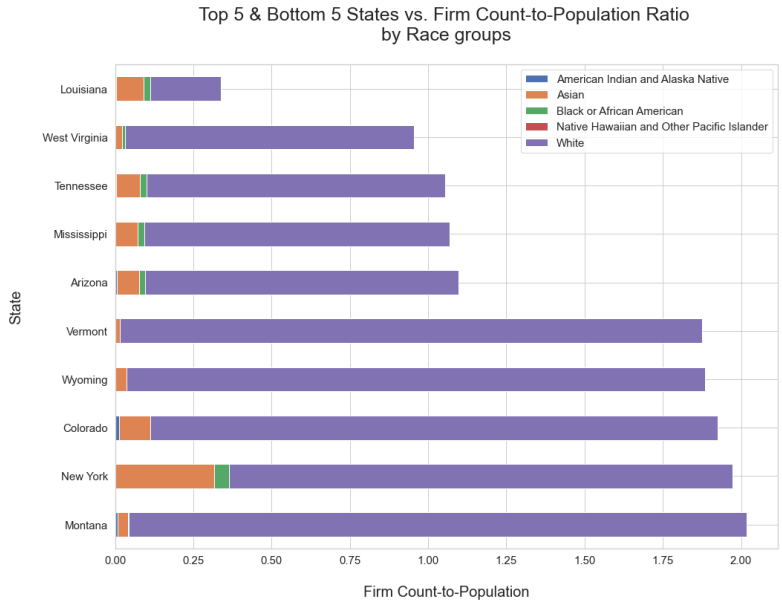


How we chose the top state to analyze.

We did not want to simply go off of the largest population of 2018 or the largest firm counts in the data. If we did, it would always skew to California. We wanted to focus on the state that is best represented with that state's dataset, and if it happened to be California, then that's that's justified!

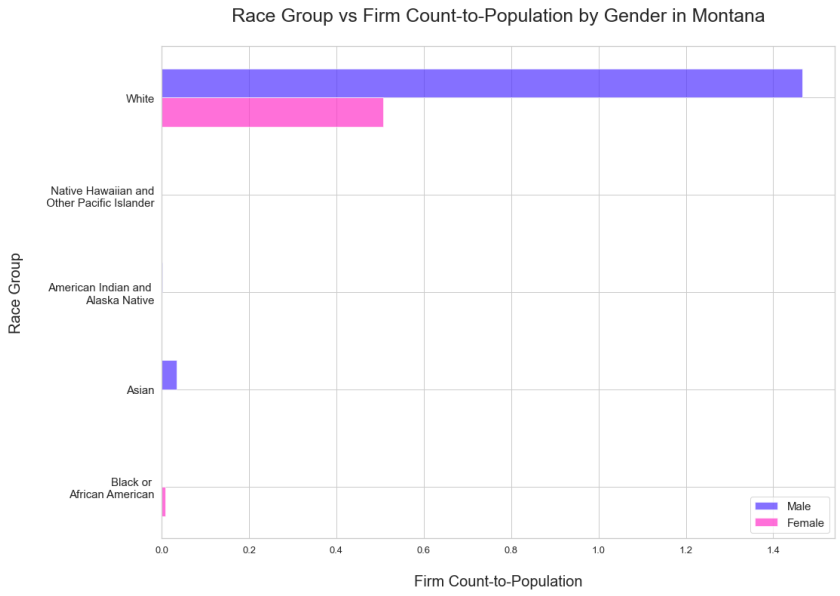
We found the ratio/percentage of each state's firm count-to-population. The bar chart over to the right shows the top 10 states with the highest firm count-to-population ratio. Surprisingly, we see that Montana's firm count to population metric is the highest in the nation, with more than a 2 percent ratio. Below is a graph of the same metric, broken down by racial group with the top and bottom 5 states.





There are slightly different results for the states coming after Montana due to the Census using aggregated data causing relative differences when filtering the data in these two visualizations.

To better understand Montana, we can take a closer look at demographic data of Montana firms.



Montana's businesses clearly are extremely nondiverse. Nearly all the respondents are white, and minority groups barely make it onto the graph at all, with small representation from Asian men and Black women. The gender disparity is also much larger in Montana firms with a male to female ratio much larger than previously seen when looking at national data.

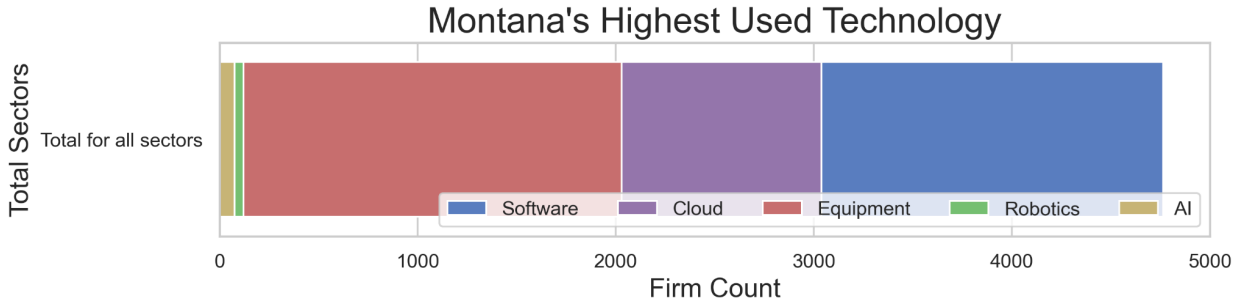
In Montana, which technology was in high use?

It's important to note how the survey was taken.

> "During the three years 2016 to 2018, to what extent did this business use the following technologies in production processes for goods or services?" - Census Bureau's 2018 Annual Business Survey - Technology Characteristics of Businesses Technical Information

The data is broken down further into counting the number of firms that reported the level of use for each technology. Ex: how many firms reported moderate use for cloud, no use for robotics, high use for software etc.

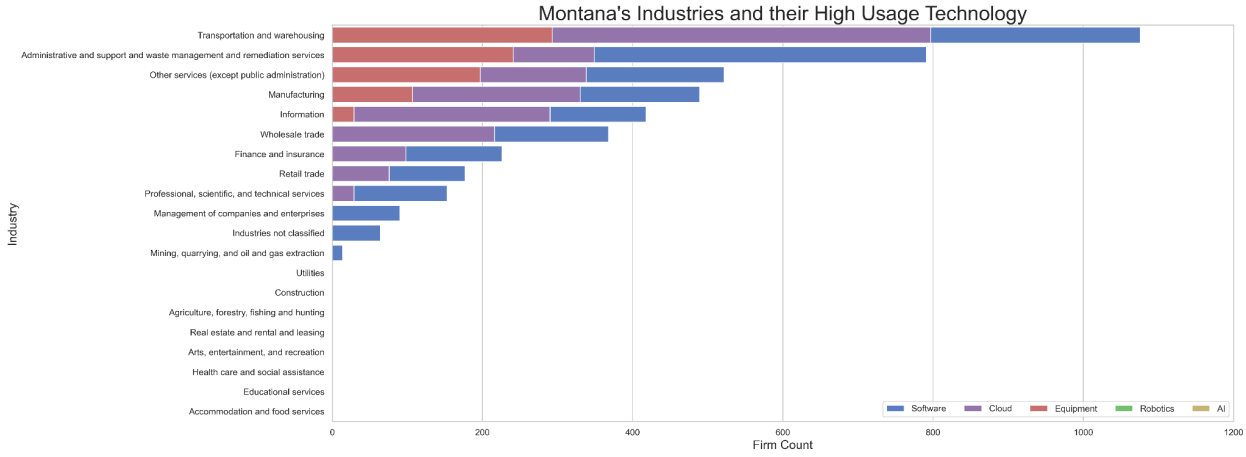
Our graph below shows the sum for all industries in Montana and the count of how many firms classified that technology as 'high usage'.



Keep in mind that the bar is not *stacked*, this is *layered.* **We see here that more firms classified specialized software over the others as their highest use technology.** The census's documentation does not have any concrete examples of what they would classify as specialized software but given the other options, we can infer their differences.

Which Industry Used Software the Most?

For our final zoom in, we ask the question: which industry is using specialized software the most in Montana? Our graph below shows Montana's industries and the number of firms reporting their high usage technology.



The transportation and warehousing industry seems to be the industry that uses specialized software the most. But it could also be that most of the survey entries just happen to come from that industry; the survey documentation doesn't give much clarification in this area. Even so, it's great to see that the highest used tech is consistent across industries. Consistently, specialized software is the highest used, then cloud, then specialized equipment.In summary, it's safe to assume that Montana transportation and warehouse workers use specialized software the most often!

Conclusion

In general respondent businesses seem to have low gender equity and racial diversity. This disparity is exacerbated in industries such as construction for gender and agriculture, farming, fishing, and hunting for race. States such as Maine and Montana also have particularly low amounts of diversity across the board. Montana for instance had the most firms per capita and its respondents for instance were nearly 100 percent white with great gender inequity. Montana used a lot of software, which is most common within transportation and warehousing, an industry shown to have great gender and racial disparities.