M08 Final Report

Census Data: Annual Business Survey 2019

Jed Dryer, Jake Thompson, Ryan Lippe, Marjea Mckoy
Submitted: Monday, April 25th, 2022

Introduction

For this project, an API was used to access US census data for the 2019 Annual Business Survey (ABS). Python was used to query this API and the pandas library was used to clean and transform the data. After initial exploration, the members of the group decided on questions to research. The questions are listed below. To answer these questions, visualizations were generated using python.

Jed researched these questions:

- *When considering Finance vs. education (or some other low paying industry,) what are the motivations for having a business?*
- *How is average pay distributed across the country when considering the race or gender of the company owner as well as industry?*
- *What are some of the most popular tech solutions in Minnesota?*

Jake researched these questions:

- *What trends are there in the ownership of firms by gender?*
- *Is it significantly different across industries?*
- *Similarly, what are the trends for ownership of firms by racial group?*
- *How do these trends compare to the demographic breakdown of the US?*

Ryan researched these questions:

- *Is there significant variance in new businesses across the country, and if so, which states have a higher portion of new businesses?*

- *For the states with the highest contrast in portion of new businesses, which categories with years in business are making up the difference?*
- *Which industries have the highest portion of new businesses, and does this help explain location variance?*

Marjea researched these questions:

- *Of the questions in the ABS, which were business owners most likely to respond to?*
- *Is there a difference in the gender of business owners based on race?*

Data Sources

The 4 primary datasets used for this report were from the US Census 2019 Annual Business Survey.[1] These datasets give information on businesses and business owners by sex, ethnicity, race, and veteran status. The 4 datasets include company summary, characteristics of businesses, characteristics of business owners, and technology characteristics of businesses.

The second source of information for this report was from the 2020 Census, it contains information on racial and ethnic demographics in the United States.[2]

[1] US Census Bureau. (2021b, October 14). *Annual Business Survey (ABS) APIs*. Census.Gov. Retrieved April 22, 2022, from https://www.census.gov/data/developers/data-sets/abs.2019.html

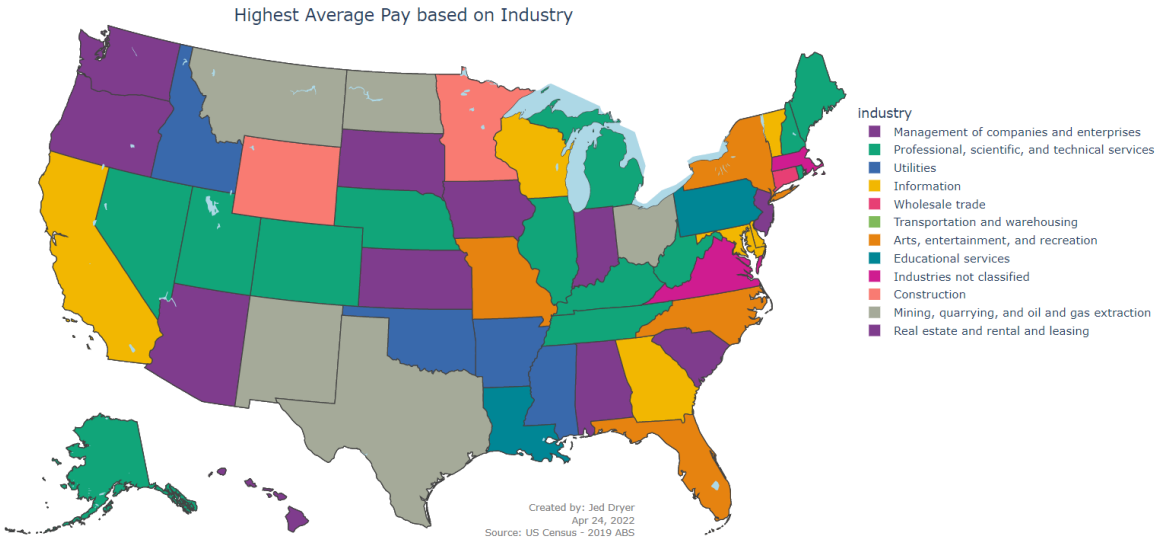
[2] US Census Bureau. (2022, February 5). *Race and Ethnicity in the United States: 2010 Census and 2020 Census*. Census.Gov. Retrieved April 22, 2022, from https://www.census.gov/library/visualizations/interactive/race-and-ethnicity-in-the-united-state-2010-and-2020-census.html

Results and Conclusions by Group Member

Jed's Results

As mentioned during the introduction, one of the questions that I wished to answer was "*how is average pay distributed across the country when considering the industry, or the gender and race of the company owner?"* In order to inform my answers to this question, I took a subset of the 'Company Summary" that was transformed to include the average annual pay for each of the target categories, and then picked the maximum of those for the map. To view the results as intended, click the figure and follow the link.

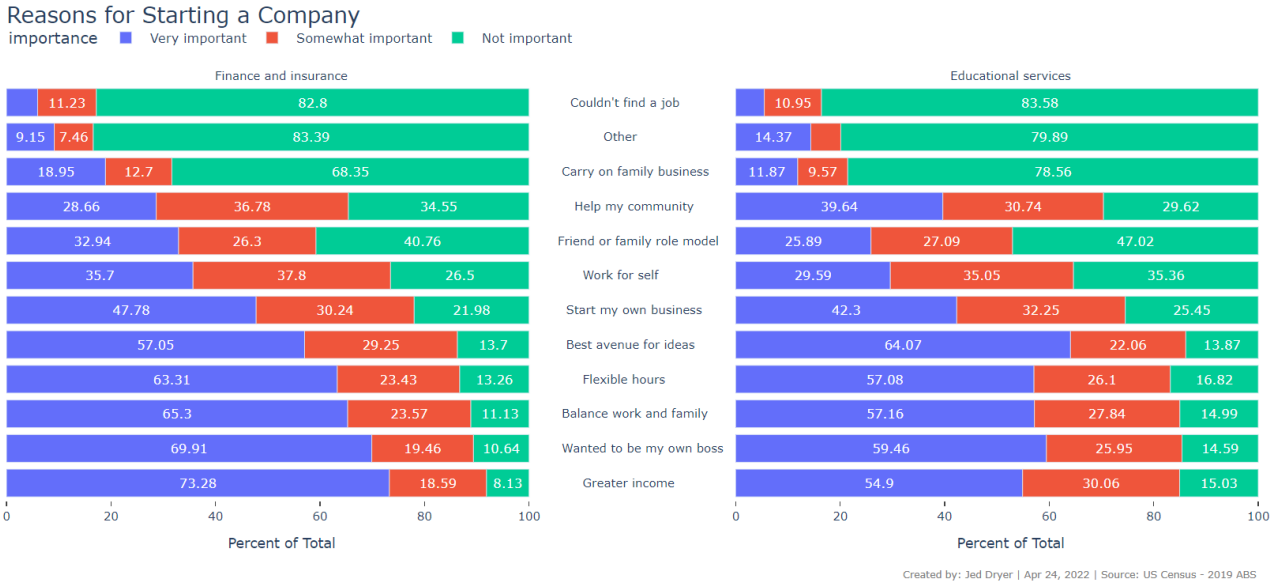
Figure 1: (Click the Figure and follow the link to interact with it.)



In regards to gender, I found that the majority of the country reported the highest average pay for employees was from companies owned by males. In regards to race, the states had greater diversity of pay based on ownership. Industry was probably the most interesting to me, and validated many of the questions that I had about the accuracy of the analysis. High averages for resource extraction were found in Texas, Montana, North Dakota, all of which are states where resource extraction is a major portion of the economy. Other aspects of this map can also be easily corroborated such as information being highlighted in California and entertainment and recreation in Florida.

Another of the questions that I wished to answer was what some of the motivations for starting a company were for people in the Finance industry versus those in the Education industry. For this analysis, I took the “Characteristics of Business Owners” dataset and the ratios for the different responses to the question. I then plotted the two industries side by side to visualize the differences.

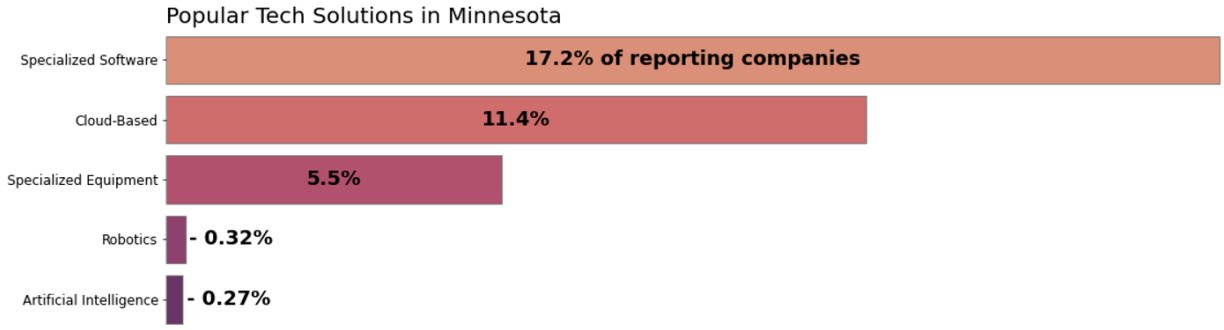
Figure 2: (Click the Figure and follow the link to interact with it.)



I found the results to be quite fascinating. Those in the finance industry were primarily motivated by a desire for greater income and placed a very low importance on helping their community. Those in Education services were primarily motivated by the opportunity being an avenue for ideas and placed a relatively equal importance on a number of other aspects.

My final question was “What are some of the most popular tech solutions for Minnesota companies?”. For this question, I looked at the “Technology Characteristics of Business” dataset. I calculated the ratio of companies that reported a high use of technologies and plotted the resulting data on a horizontal bar chart.

Figure 3: (Click the Figure and follow the link to view a larger version.)



I was surprised to note that Artificial Intelligence has such a low usage rate, and specialized software has such a high rate. This can be explained by the relative newness of Artificial Intelligence, and the possibility that many companies use custom software solutions for their business needs.

Jed's Conclusions

Based on my analysis, I found that the best industries to work for can vary based on where you live. The motivations of people to start businesses can have a very interesting spectrum of responses. My final conclusion is that artificial intelligence has not yet peaked and will continue to rise in usage in Minnesota. I believe more importantly, I found that working with the Census Data can be extremely complicated and it is very easy to accidentally manipulate the data which can lead to inaccurate reporting. In a world where accurate data reporting is vital, data clarity is key to good data management.

Jake's Results

The first thing explored was what trends exist in the majority ownership of US firms by industry and minority status, seen in **Figure 4.**

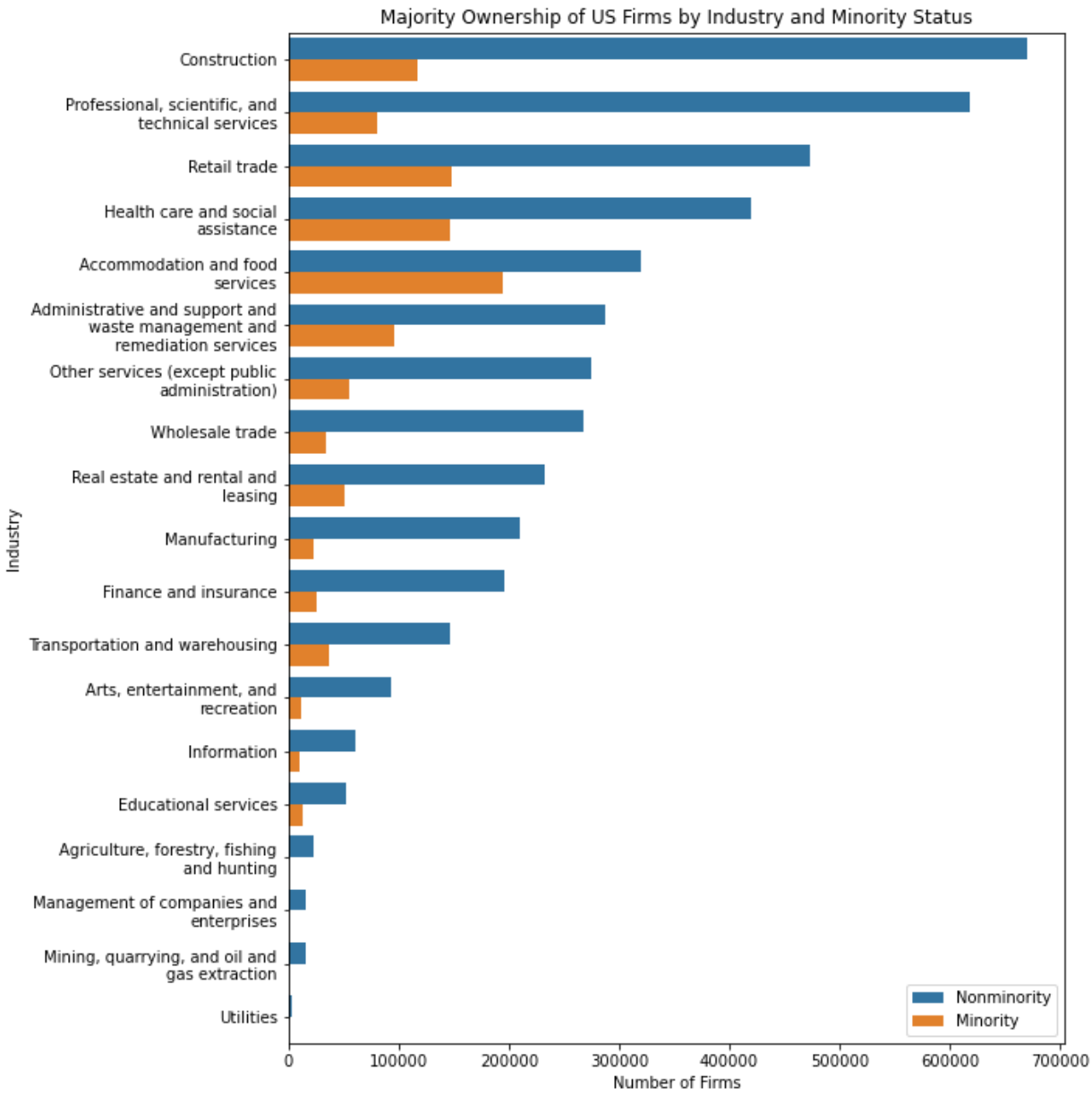


Figure 4.

It was seen that the industry with the highest proportion of minority ownership was accommodation and food services, which was 62.1% nonminority owned. Looking at 2020 US census data, 61.6% of the population is white, which closely matches this value.[2] For every other industry, however, the proportion of minority ownership was lower, which implies that minorities are underrepresented in firm ownership across nearly every industry.

The next question considered was what trends might exist in the majority ownership of US firms by industry and gender, visualized in **Figure 5**.

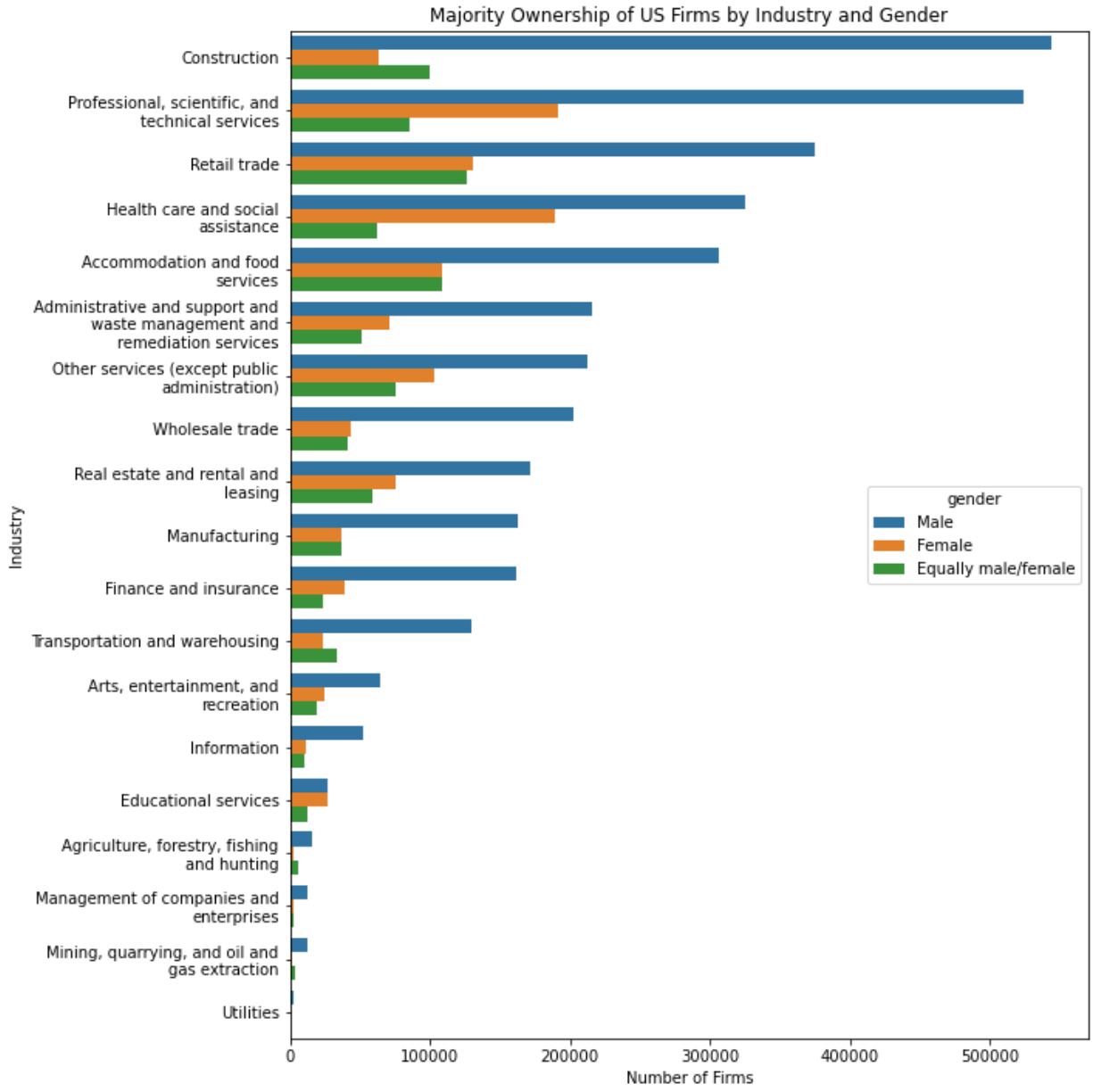


Figure 5.

It was seen that women are underrepresented in the ownership of nearly every industry tracked, considering that they make up approximately half of the US population. The educational services industry was the exception, with nearly equal ownership across gender. The industry with the greatest disparity was seen to be construction, so

this was investigated further to see if this disparity varied geographically, visualized in **Figure 6.**

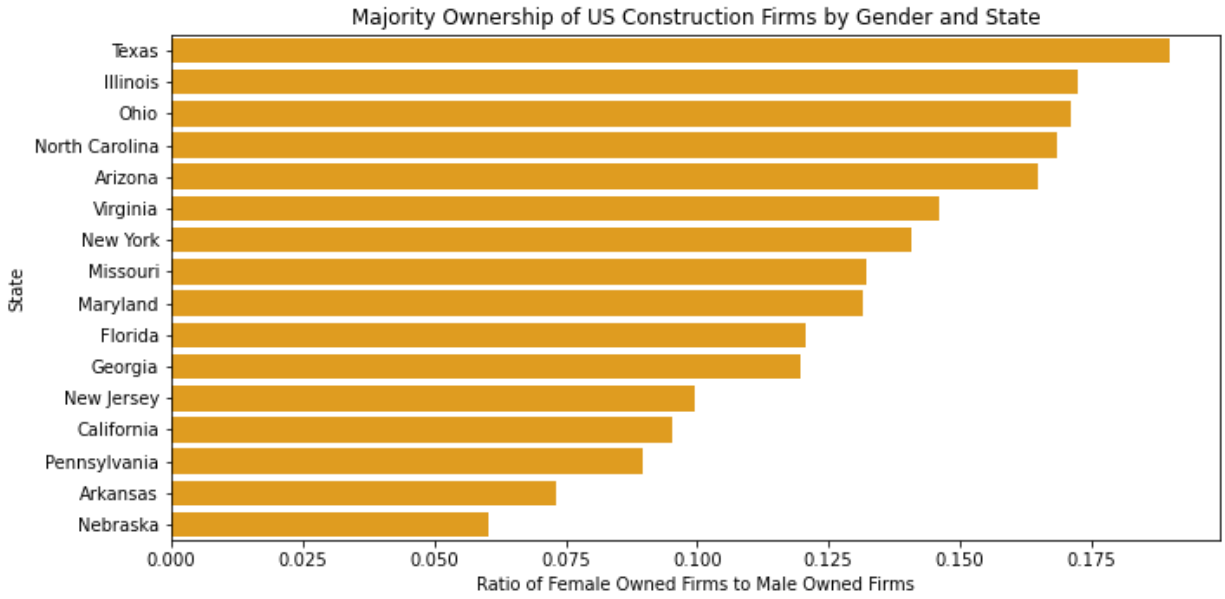


Figure 6.

The ratio of firms with majority female ownership to firms with majority male ownership was calculated for every state. **Figure 6** only contains 16 states because for almost all of the others, there were no firms tracked with majority female ownership, resulting in a ratio of 0. The 2 exceptions were Montana and South Dakota, which had no male-owned firms, resulting in a ratio of infinity. Note that for the data from the ABS, firms that couldn't be classified by sex, race, ethnicity, and veteran status weren't included in this state-level, so these values aren't exhaustive;[1] this means that it's difficult to draw conclusions from this data at a state level.

Jake's Conclusions

It was seen that for every industry other than accommodation and food services, minorities were underrepresented in business owners. Similarly for gender, women were underrepresented in every industry except for education services, with construction having the greatest disparity. It was also seen that a significant number of firms that can't be classified by race, ethnicity, gender, and veteran status were

excluded from the data, so further research is necessary to see if the excluded firms change these results.

Ryan's Results

The initial data looked into was the percentage of businesses less than two years old by state, which can be seen below.

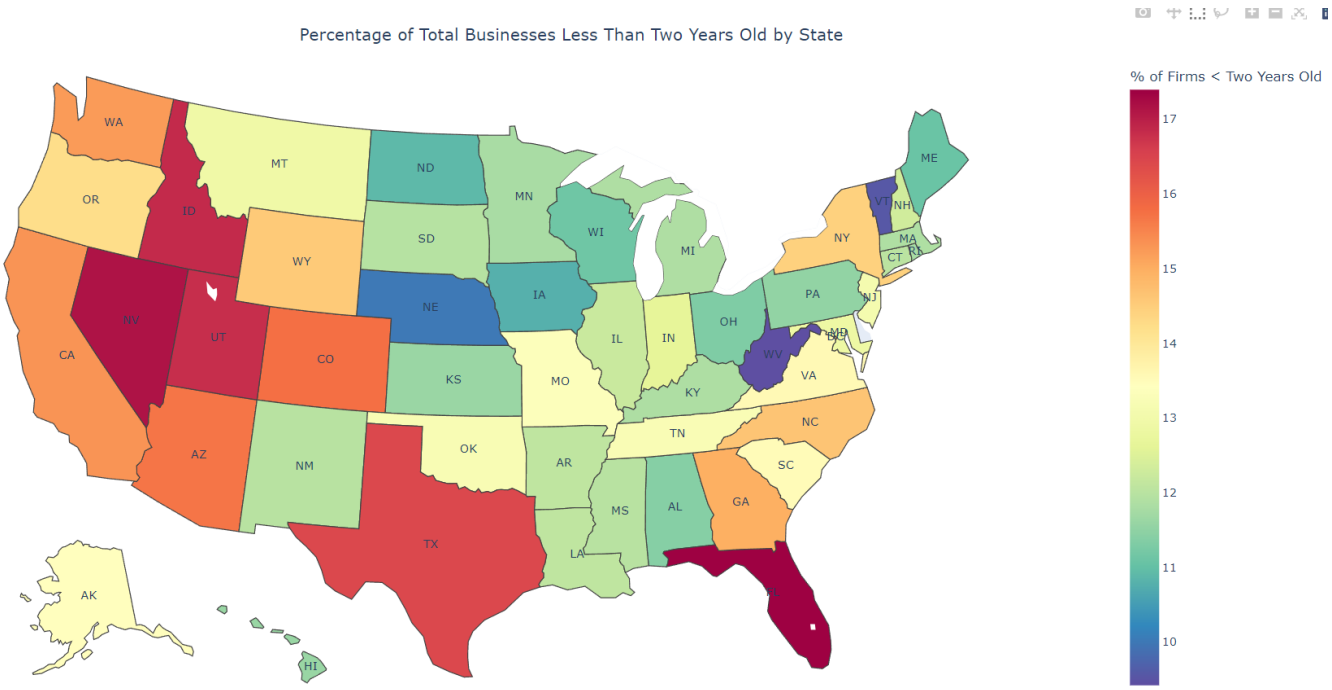


Figure 7.

As can be seen in **Figure 7**, there is significant variance by state, with some states having as little as 9% of businesses created within the last two years, while others have as much as 17%. For the states with a higher portion, this could mean greater economic instability and/or greater economic opportunity. The top states for new business were Florida, Texas, Nevada, Utah, Idaho.

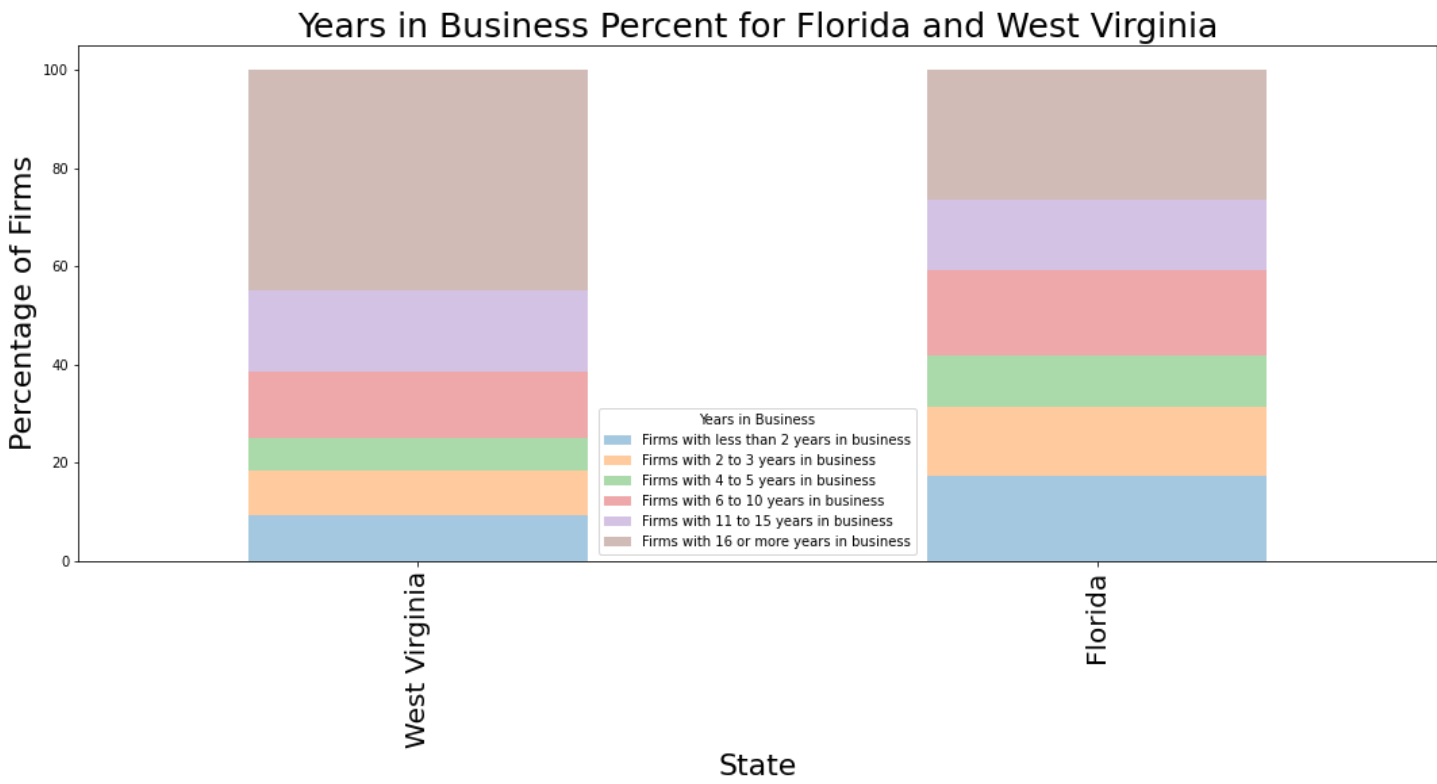


Figure 8.

As seen in **Figure 8**, a majority of the difference in proportion of new firms between West Virginia and Florida can be explained by the firms with 16 or more years in business category, which made up ~42% of total businesses in West Virginia, compared to just ~23% for Florida. Some potential causes for this in West Virginia could be lack of growth, or simply a more stable and predictable economy.

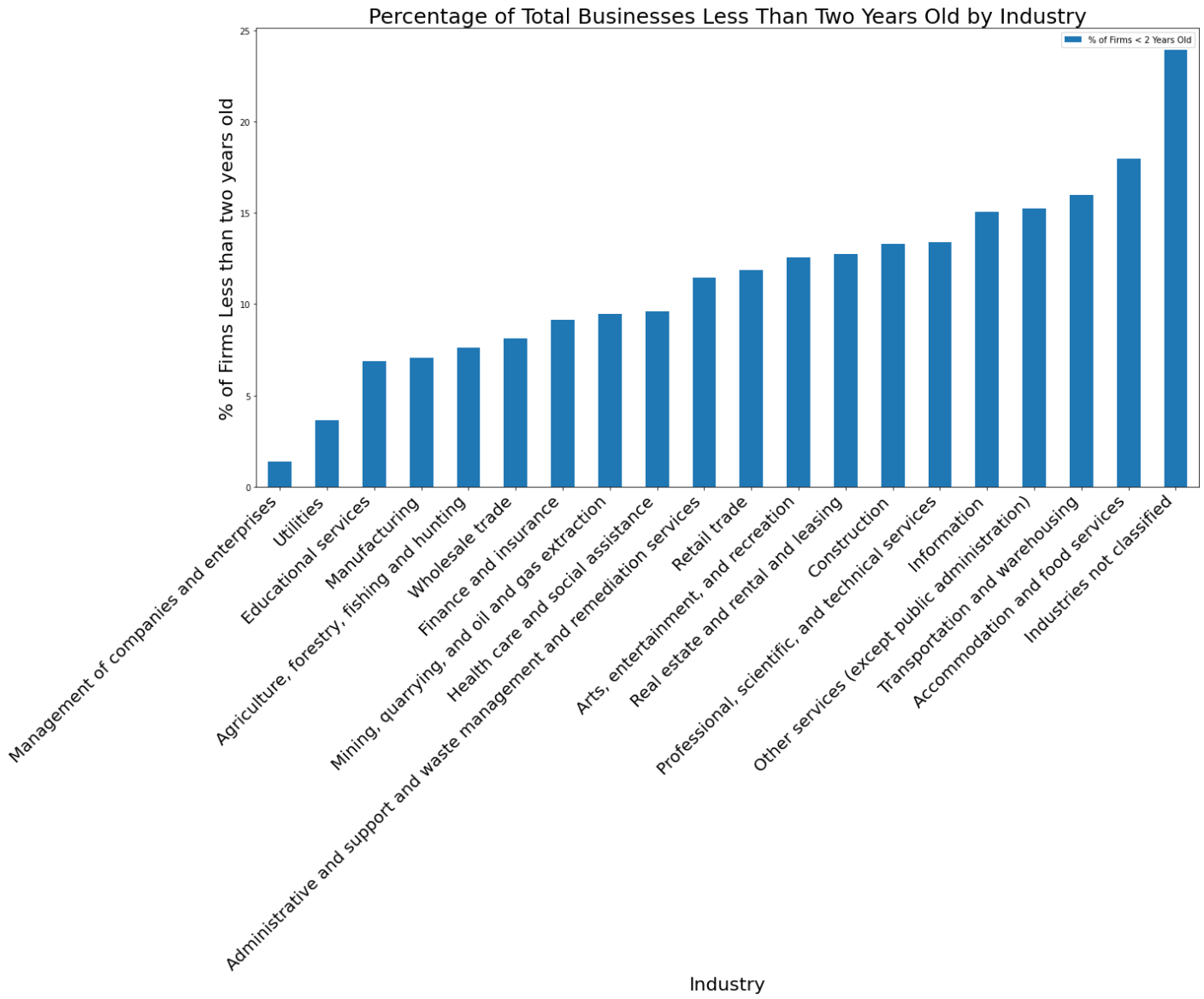


Figure 9.

In **Figure 9**, we see that the industries with the highest portion of new business are the industries not classified by this survey, which is very interesting. It could suggest the creation of new jobs that haven't existed in the past at a higher rate than any pre-existing industry. There was also significant new business within Transportation and Warehousing, as well as Accommodation and Food Services. This could be due to a high turnover rate in the food business, as well as the growth of uber, lift, amazon, and doordash. This does not however explain the location variance.

Ryan's Conclusions

This analysis shows that there is high variance in the creation of new businesses at both state and industry levels. Potential causes for these variances include high turnover for businesses due to market instability, as well as potential gaps in the market creating more opportunity in some places and industries than others. Our research also found that there are more new businesses in unclassified industries, which could be showing a high rate of creation for jobs that have not existed in the past.

Marjea's Results

The first question considered was "Of the questions in the ABS, which are business owners most likely to respond to?" To answer this, **Figure 10** was generated.

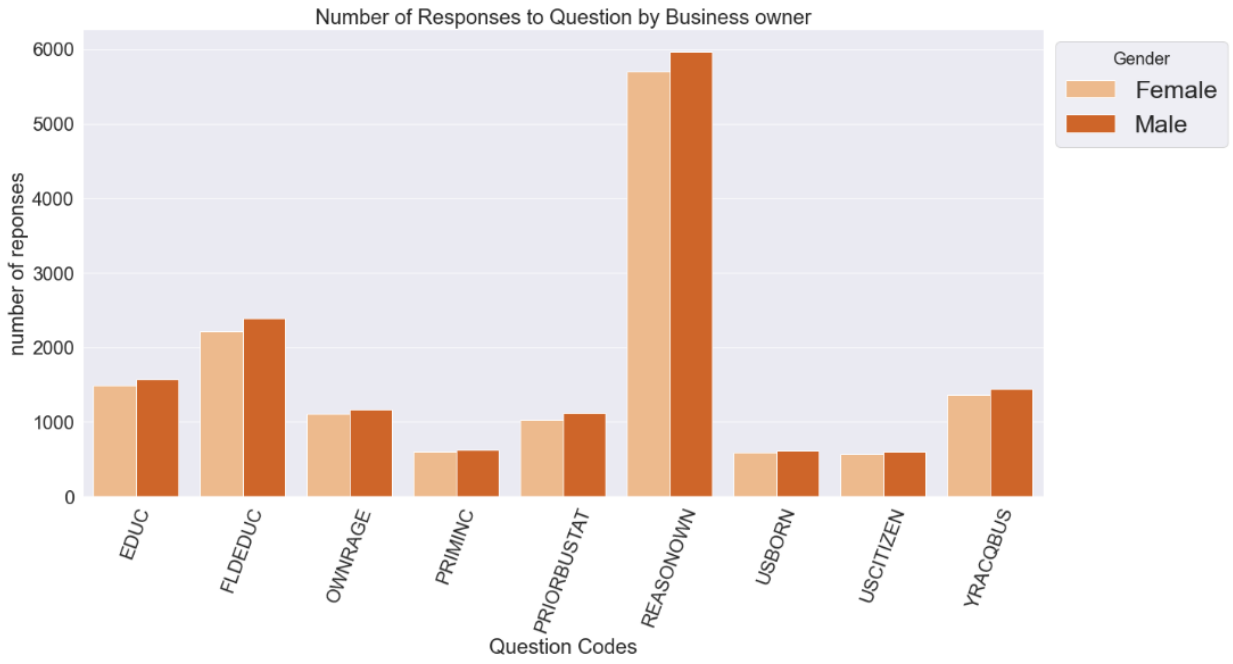


Figure 10.

It was seen that REASONOWN, the owner's reason for owning a business, was the most commonly answered question. There wasn't a significant gender gap in the number of responses. **Figure 11** shows the percentage for the frequency of answers for each of the ABS questions.

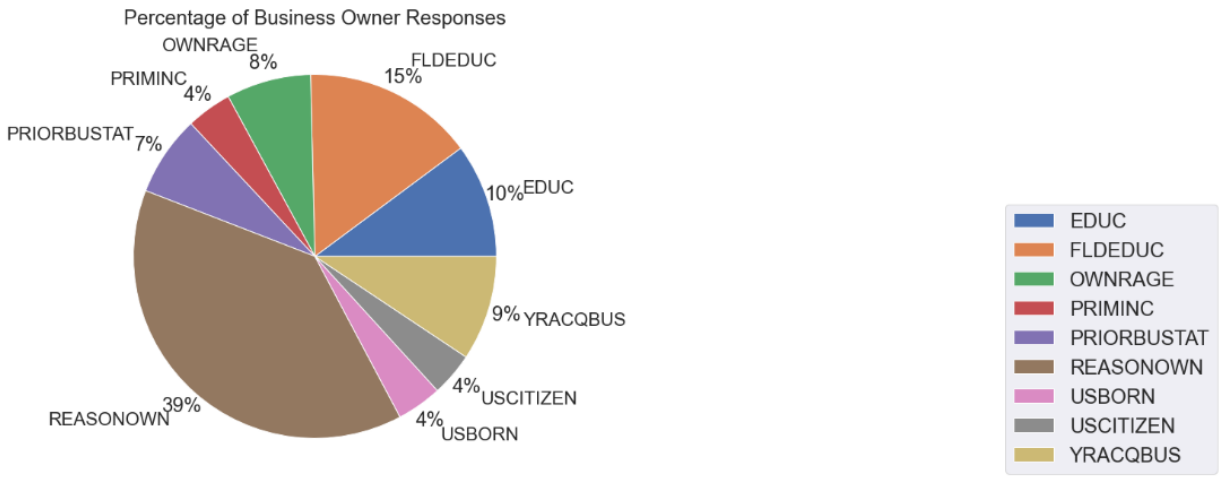


Figure 11.

REASONOWN was the most commonly answered question, and FLDEDUC (field of the owner's highest finished degree before establishing, purchasing, or acquiring the business) was the next most commonly answered question.

The last question considered was whether there was a gender gap among business owners based on race, visualized in **Figure 12.**

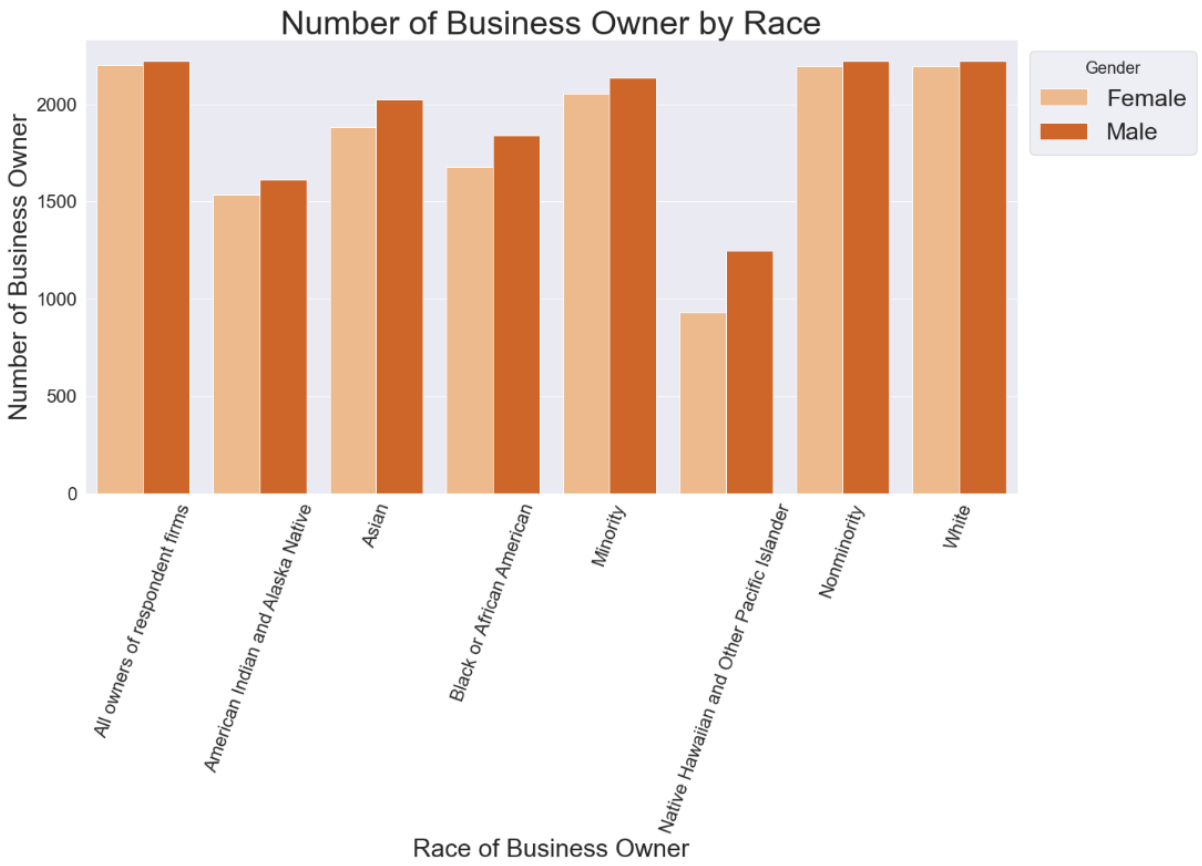


Figure 12.

This showed that there was a significant gender gap between the Native and Pacific Islander business owners.

Marjea's Conclusions

This research showed that business owners' most frequently answered question in the ABS was the owner's reason for owning a business. No significant gender gap was seen. In the number of business owners by race, the biggest gender gap was for the Native Hawaiian and Other Pacific Islanders category.