All In the Family?

An Analysis on the 2019 Annual Business Survey

Group 2: Gavan VanOver, James Miller, Norman Morris, Yiqing Guo

The US Census Bureau is constantly collecting vast amounts of data from the country over all sorts of topics. The topic of interest for this project is the Annual Business Survey for 2019 which, according to census.gov, “combines data results from survey respondents and administrative records to produce data on business ownership”. The survey itself had four different APIs: “Company Summary”, “Characteristics of Business”, “Characteristics of Business Owners”, and “Technology Characteristics of Businesses”. Each of these APIs contained different data but we decided to focus specifically on data relating to family-owned businesses and therefore we only used data from two of the APIs.

Data

All data was provided by the US Census Bureau. The data was gathered specifically from the Annual Business Surveys APIs for 2019. Each API included the data on race, age, ethnicity, and veteran status of the owners.

Business Characteristics

api.census.gov/data/2018/abscb.html

This API provided data on businesses in the United States during 2018. The data includes the number years in business, employment size, revenue, types of customers, if the business had ceased operations during 2018, whether or not the business was family owned, how many owners there were, if spouses/partners owned the business,

Business Owner Characteristics

api.census.gov/data/2018/abscbo.html

This API provided more in-depth information on the business owners in the United States in during 2018. The data includes whether the owner is disabled, education level of the owner, owner’s age, the status of the owner’s previous business, whether this business was the primary source of income for the owner, why the owner had the business, if the owner was born in the US, if the owner was a US citizen, military service questions, and year of acquisition of the business.

Questions

- What can we learn about business owners of family-owned businesses vs non-family-owned businesses?
    - How do the owners of family-owned businesses vs non-family-owned businesses compare by sex?

    - How do the owners of family-owned businesses vs non-family-owned businesses compare by race?
    - How do owners rate the importance of carrying on the family business?
- What differences can we learn about the size and performance of family-owned businesses vs non-family-owned businesses?
    - How does firm revenue size category break down between family owned and non-family-owned businesses?
    - Do family-owned businesses have more employees?
    - How does the number of employees break down between family-owned and non-family-owned businesses?
    - How long have family-owned businesses been operating?
- What else can we learn about family-owned businesses?
    - Are family-owned businesses more concentrated in certain industries?
    - Which states have the most family-owned businesses?
    - Which states have the highest percentage of family-owned businesses?
    - Which states have the most employees per family-owned business?
    - How does the average annual payroll of family-owned businesses compare per state?

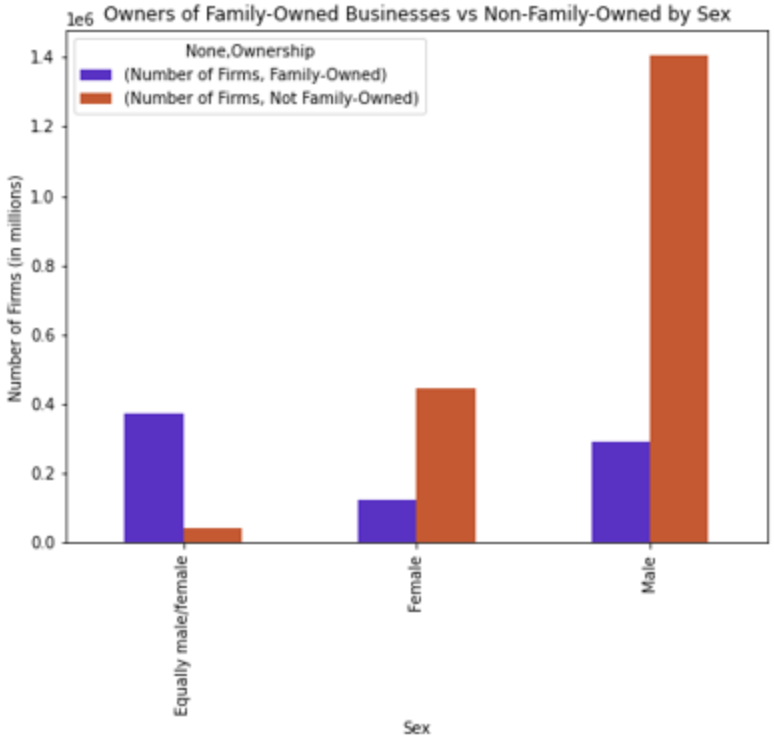
Process and Results

After deciding what to focus on and what questions were going to be asked, the first step of this project was to be able to consume the APIs and access the data with Python. By creating a link containing the variables, data can be read into a pandas DataFrame using the pandas read_json() function. This process is detailed in the ETL report. From there, the data is adjusted and then used to plot visualizations. A combination of pandas, matplotlib, plotly, and seaborn were used to create each of the following visualizations.

By Question:

- **How do the owners of family-owned businesses vs non-family-owned businesses compare by sex?**

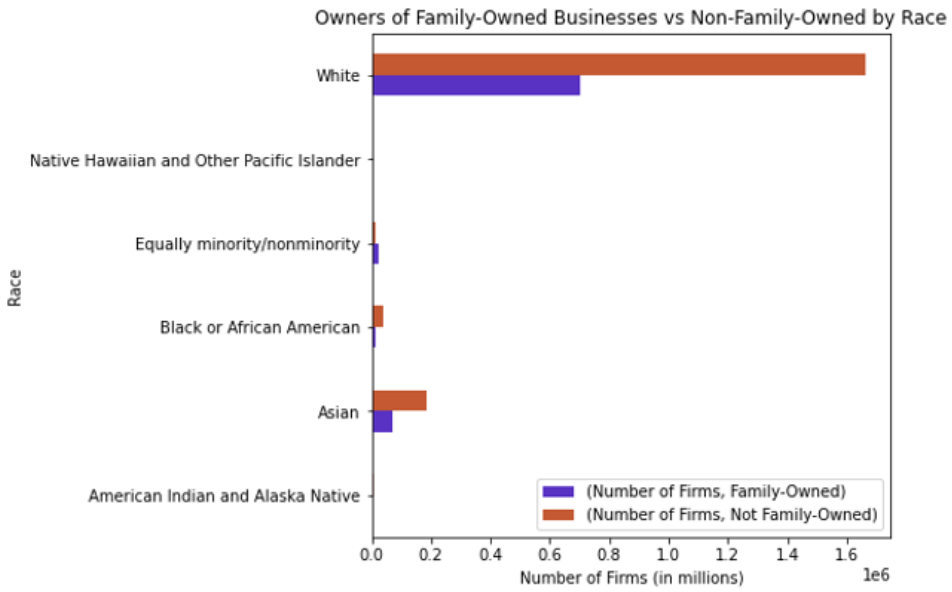
This question required the use of the Characteristics of Businesses API and specifically wanted to look at three things: whether the business was family-owned or not, the different sex categories (male, female, equally male and female), and the number of owners and that fell into each category. Sex was used as the x-axis, number of owners was the y-axis, and the data was grouped by if the business was family-owned or not.



Looking at the resulting bar graph there are a few things that are immediately apparent. First off, there are far more non-family-owned companies owned by men than any other category. Next, there are more family-owned businesses owned equally owned by men and women than there are by just men or just women. This is likely due to family-owned companies being run by spouses or even siblings.

- **How do the owners of family-owned businesses vs non-family-owned businesses compare by race?**

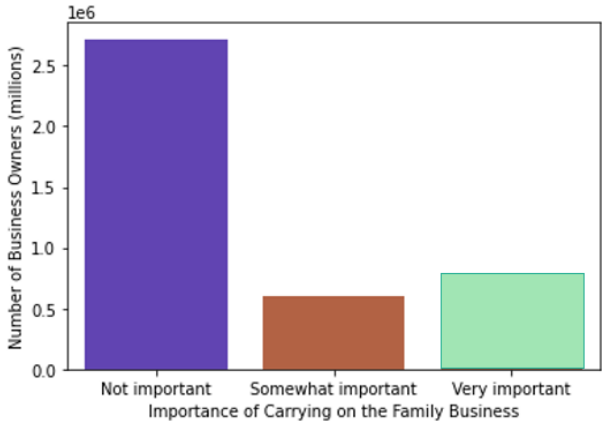
The process for this question was the same as the previous one with sex exchanged for race group. So the three variables called on were whether the business was family-owned or not, the different race groups(white, Native Hawaiian or Pacific Islander, black or African American, Asian, American Indian and Alaska Native, or equally minority/nonminority), and the number of owners and that fell into each category.



This horizontal bar graph paints a clear picture of the disparity between the number of businesses owned by the various races. The large majority of businesses in both family-owned and non-family-owned are owned by white Americans. Asian is the distant second highest category. From there the values are almost miniscule with none of the other categories even breaking 100,000 businesses between both the family-owned and non-family-owned categories combined.

- **How do owners rate the importance of carrying on the family business?**

This question used the Characteristics of Business Owners API, and the data was filtered to only be for family-owned businesses. From there the only two fields considered were the level of importance and the number of business owners in each category.



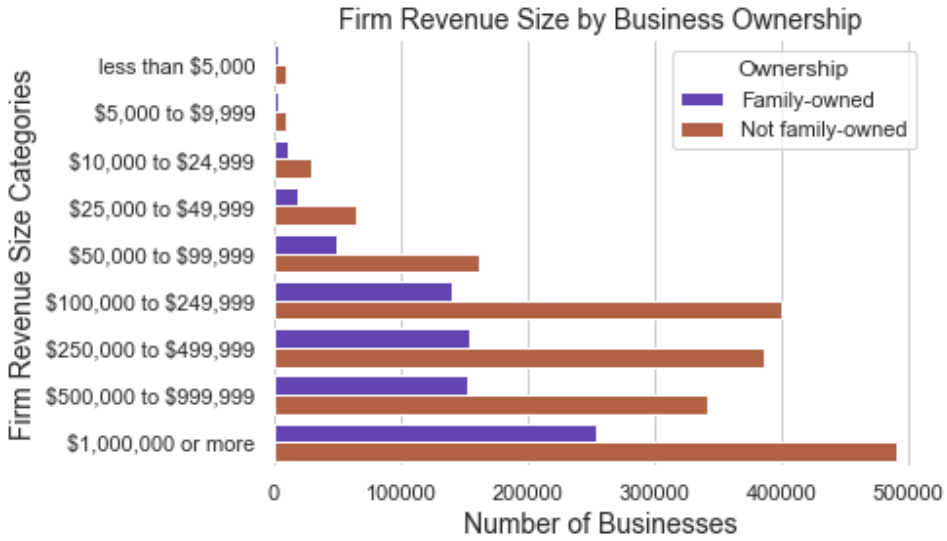
According to this bar graph over 2.5 million business owners don't find carrying on the family business important and only about 700,000 found it very important. Based on this

census data it would be easy to say that most business owners aren't concerned with the family business being carried on.

- **How does firm revenue size category break down between family owned and non-family-owned businesses?**

This question used the Characteristics of Businesses dataset and pulled firm numbers (FIRMPDEMP) alongside the data field (RCPSZFI) which represents the sales, receipts, and revenue sizes of all firms. The survey question "In 2018, did two or more members of one family own the majority of this business?" was also filtered for using the QDESC code 'B02' and the BUSCHAR codes 'BM' and 'BN', which stand for Family-owned businesses and Not family-owned businesses respectively.

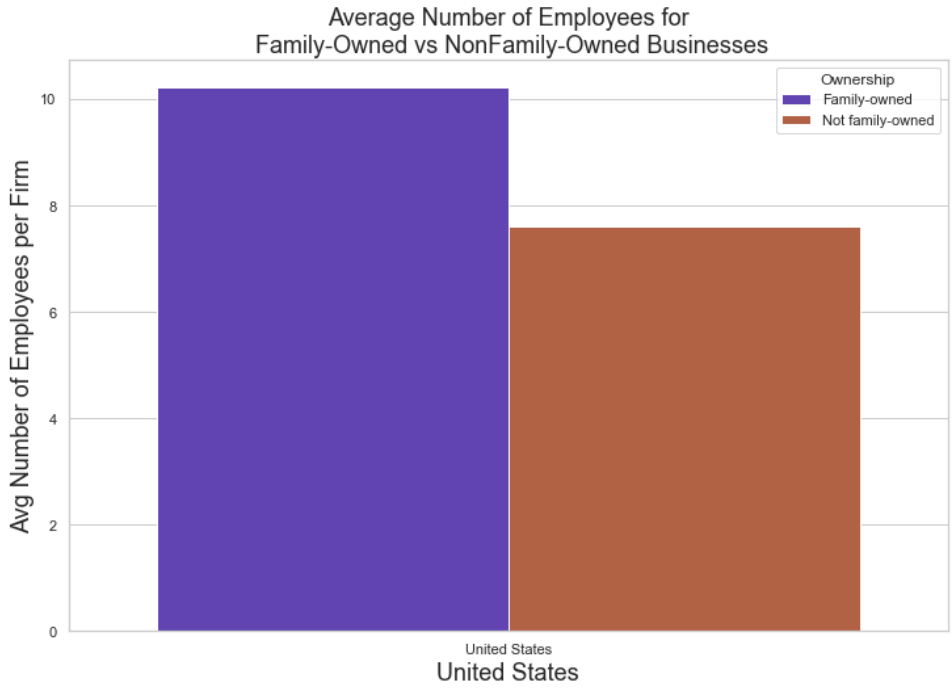
Using the seaborn plotting package, the total number of businesses was plotted against the revenue size category of the firms. The hue attribute is used to further differentiate between family owned and non-family-owned firms per revenue size category.



As shown, the number of businesses per firm size steadily increases for both family-owned and non-family-owned as the firm revenue increases. When observing the trend for family-owned and non-family-owned businesses, the graph shows that the two ownership categories have the same upward trend. However, for the higher revenue size categories, the number of non-family-owned businesses far exceeds the family-owned.

- **Do family-owned businesses have more employees?**

This question uses the Characteristics of Business API. We created a column in the dataframe to store the average number of employees, 'avg_emp_per_firm', deriving its values by dividing the values of EMP by the values of FIRMPDEMP (as FIRM_COUNT). The bar chart plots the country on the x axis and the average number of employees on the y axis. The chart was created using the seaborn library.



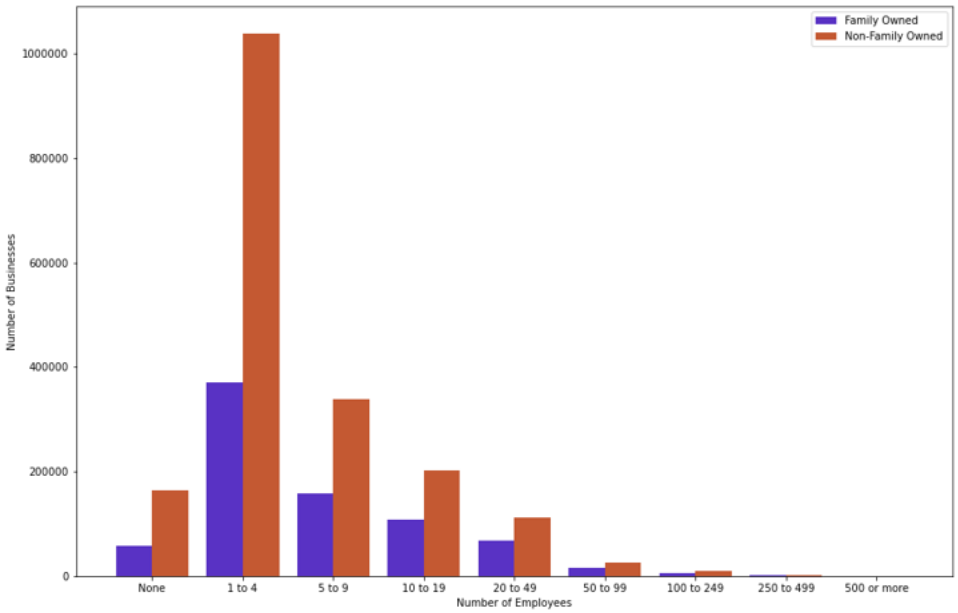
We hypothesized that family-owned businesses would be more likely to have fewer employees. To answer this question, we compared a simple mean average of employees per business between family-owned and non-family-owned businesses. Initially, this was plotted on a bar chart by state. You could clearly see that family-owned businesses in most states actually had more employees on average than non-family-owned businesses. Among other factors, we considered after the fact that a business is only considered family owned if two or more members of one family own a majority of the business. This means that a sole proprietorship, for example, would not be considered a family-owned business. Based on the findings, we decided that this was better visually represented by looking at the country as a whole for the two segments.

- **How does the number of employees break down between family-owned and non-family-owned businesses?**

This question used the Characteristics of Businesses dataset and pulled firm numbers (FIRMPDEMP) alongside the data field (EMPSZFI) which represents the employee sizes of all firms. The survey question "In 2018, did two or more members of one family own the majority of this business?" was also filtered for using the QDESC code ‘B02’ and the BUSCHAR codes ‘BM’ and ‘BN’, which stand for Family-owned businesses and Not family-owned businesses respectively.

Using the matplotlib plotting package, the total number of businesses was plotted against the employee size category of the firms. The hue attribute is used to further

differentiate between family owned and non-family-owned firms per revenue size category.

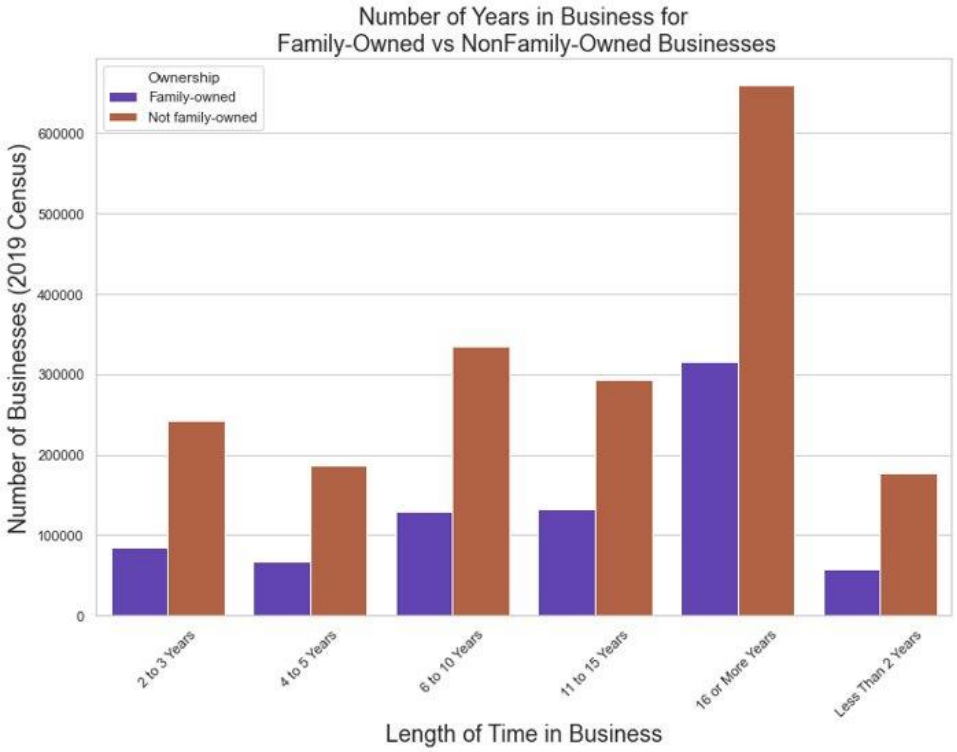


Surprisingly, the vast majority of firms retain just a few employees, for both family and non-family-owned businesses. Furthermore, the trends for both family-owned and non-family-owned businesses are very similar. This shows that there is not a very big difference between family owned and non-family-owned businesses in this respect. There are simply more firms with less employees than with more employees.

- **How long have family-owned businesses been operating?**

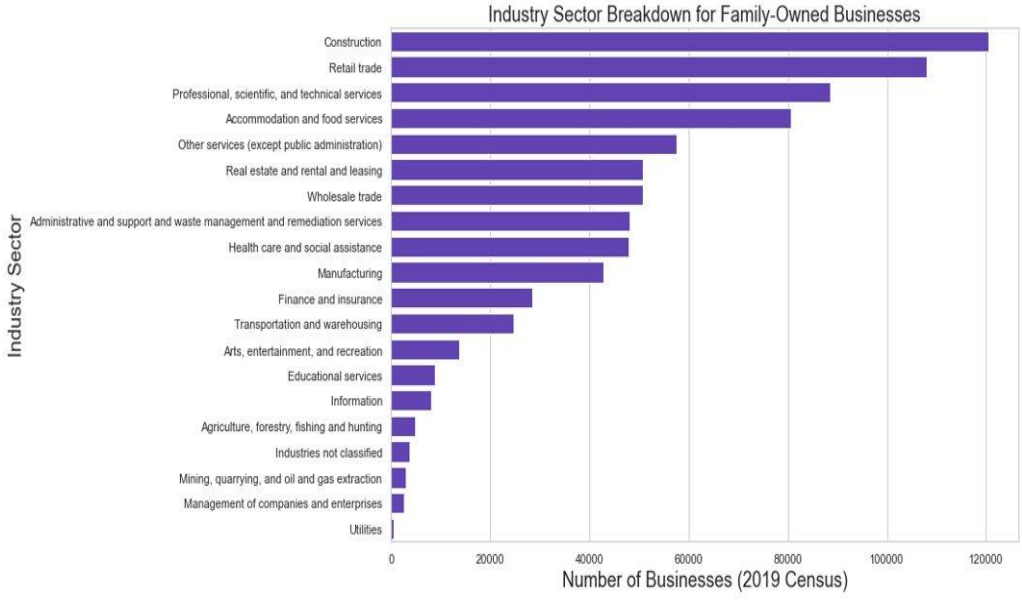
This question uses the Characteristics of Business API, filtered on the BUSCHAR variable where it equals BM (for family-owned) and BN (for non-family-owned). The chart was created using the seaborn library and plots the time buckets on the x axis and the number of businesses on the y axis.

We wondered about the survivability of family-owned vs non-family-owned businesses. While we were unable to include whether businesses were still operating or not, we were able to find that most family-owned and non-family-owned businesses have been operating for 16 or more years. The count of businesses in each time bucket for both segments generally decrease as the time decreases.



- **Are family-owned businesses more concentrated in certain industries?**

This question required data from the Characteristics of Businesses API, was filtered to only be from family-owned businesses, and then the fields were limited to the industry sector labels (NAICS2017_LABEL) and the number of businesses within each industry (FIRMPDEMP).



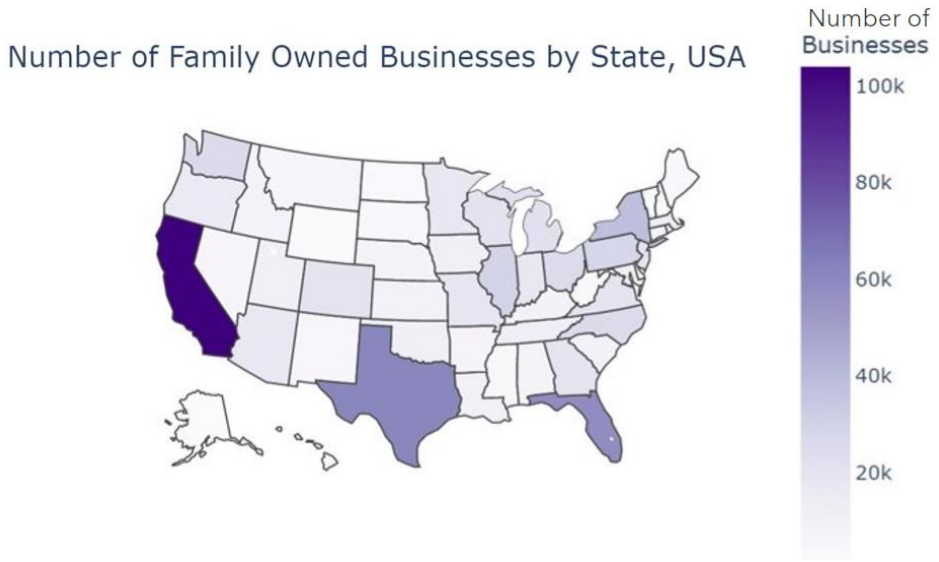
The resulting horizontal bar graph shows that there is definitely a difference between the number of family-owned businesses within each industry. The top three industries

appear to be construction, retail trade, and professional/scientific/technical services. The bottom three appear to be utilities, management of companies or enterprises, and mining.

- **Which states have the most family-owned businesses?**

This data in this map comes from the Characteristics of Businesses dataset. Geolocation data is at the state level and was merged with a table containing state abbreviations for compatibility with the plotly Choropeth map in Python. The variable BUSCHAR_LABEL in the dataset was filtered for “Family-owned” businesses.

This particular map shows the total number of businesses (FIRMPDEMP) per state. At a glance, the number of family businesses is greatest in California, followed by Texas and Florida. A possible explanation for the large number of family businesses in these states is that they have more population than lighter colored states. However, some states such as New York have comparable populations to states like California and are not as dark. In addition to a high population, states with a high number of family businesses could have policies that are more welcoming to family businesses.



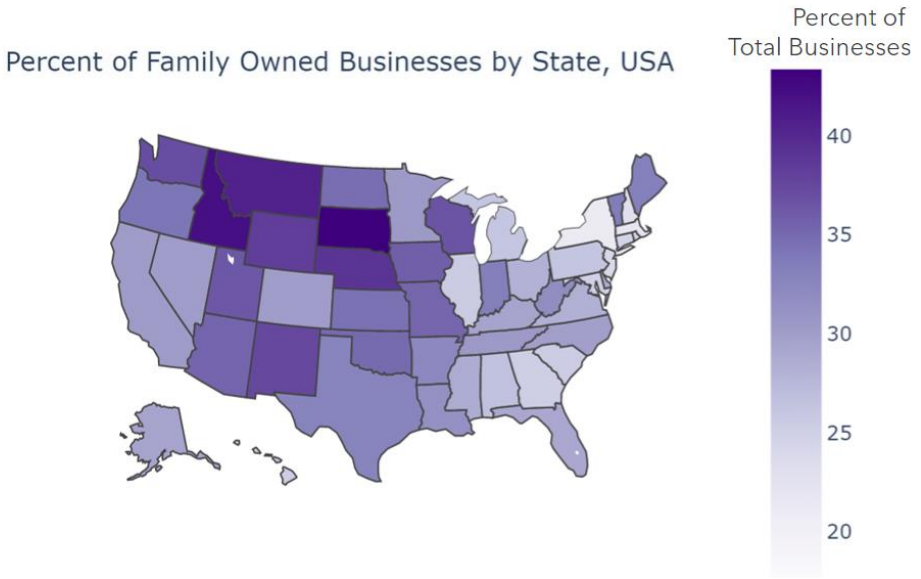
- **Which states have the highest percentage of family-owned businesses?**

This data in this map comes from the Characteristics of Businesses dataset. Geolocation data is at the state level and was merged with a table containing state abbreviations for compatibility with the plotly Choropleth map in Python. The variable BUSCHAR_LABEL in the dataset was filtered for “Family-owned” businesses.

This particular map shows the percentage of family business per state. This value is calculated by dividing the number of family businesses (FIRMPDEMP) per state by the total number of businesses per state. To make this calculation, a separate pandas DataFrame was created as above, except the BUSCHAR_LABEL in the dataset was filtered for “Total reporting” for the family ownership question (FAMOWN in QDESC).

This “total reporting” dataset was merged with the first only “family owned” dataset by state. A new column, the percentage of family businesses per state, was created by dividing the number of family-owned businesses by the total number of businesses in the state.

At a glance, the states with the highest percentage of family-owned businesses are in the Northwest. When compared with the map visualization of total number of family-owned businesses by state, this map seems like an inversion. The states with smaller populations are the states with the highest proportions of family-owned businesses.

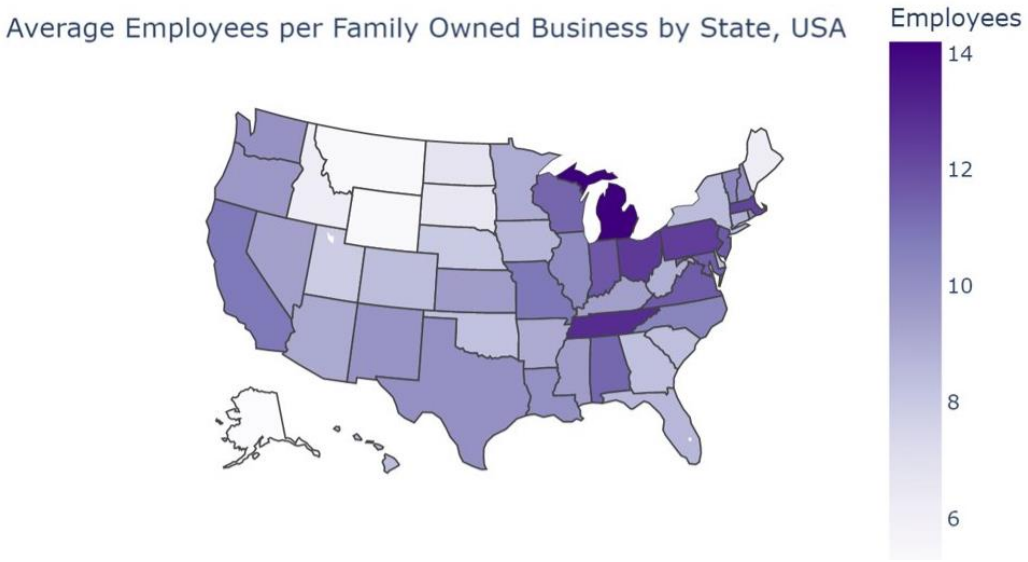


- **Which states have the most employees per family-owned business?**

This data in this map comes from the Characteristics of Businesses dataset. Geolocation data is at the state level and was merged with a table containing state abbreviations for compatibility with the plotly Choropeth map in Python. The variable BUSCHAR_LABEL in the dataset was filtered for “Family-owned” businesses.

This particular map shows the average number of employees in a family business per state. This value is calculated by dividing the total number of employees (EMP) in family-owned businesses per state by the total number of family-owned businesses (FIRMPDEMP) per state. To make this calculation, a new column, the average number of employees, was created by dividing the number of employees by the number of businesses.

At a glance, the states with the highest number of employees per business are in the Northeast, the highest of which being Michigan. The states with the lowest number of employees per business are in the Northwest. When compared to the percentage of family-owned businesses by state map visualization, it seems that the more family-owned businesses are in a state, the less employees there are in that state.



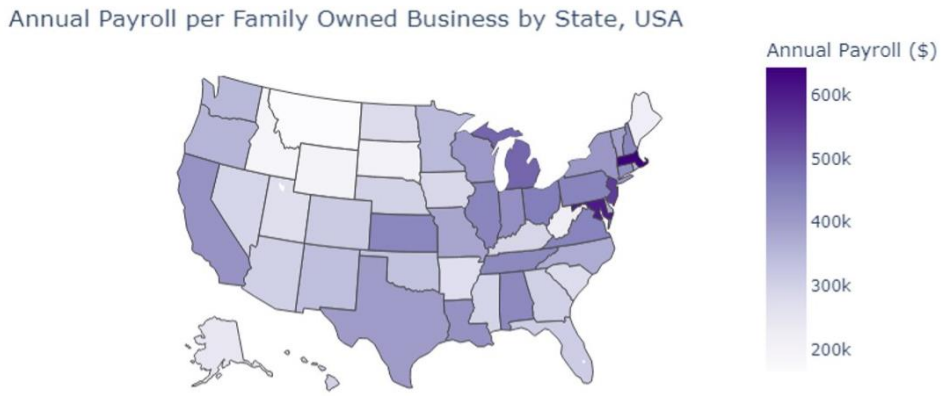
- **How does the average annual payroll of family-owned businesses compare per state?**

This data in this map comes from the Characteristics of Businesses dataset. Geolocation data is at the state level and was merged with a table containing state abbreviations for compatibility with the plotly Choropeth map in Python. The variable BUSCHAR_LABEL in the dataset was filtered for “Family-owned” businesses.

This particular map shows the average annual payroll of a family business per state. This value is calculated by dividing the total annual payroll (PAYANN) in family-owned businesses per state by the total number of family-owned businesses (FIRMPDEMP) per state. Furthermore, the original PAYANN variable is in units of $1000. In this map, the annual payroll is represented in terms of dollars and not thousands of dollars.

At a glance, the states with the highest number of employees per business are in the Northeast, the highest of which are Maryland, Connecticut, and New Jersey. The states with the lowest number of employees per business are in the Northwest. This map looks remarkably similar to the average employees per firm map above. This makes sense, as with more employees, the amount paid to employees should increase. The minor differences in shades between the maps could possibly be due to increased pay in states along the East Coast as compared to states in the Midwest.

Furthermore, when compared to the map of percent of family-owned businesses by state, it seems that when there is very little pay, there are also more family-owned businesses. It’s possible that the family-owned businesses in states in the Northwest could differ from the ones in states in more populated regions in function and industry.



Conclusion

The dataset on the Annual Business Survey contains a lot of information on businesses, business owners, and technical capabilities. This report focuses on family-owned businesses, asking questions about business owners, employee size, revenue size, years in business, and providing industry and geographic breakdowns for family-owned businesses.

Some restrictions in the dataset are how all the tables are only commonly sorted by the owner's sex, ethnicity, veteran status, and race. Because this report focuses primarily on family-owned businesses, it was hard to merge tables and filter for information specifically concerning family-owned businesses over multiple tables. Despite merging tables together on the owner's sex, ethnicity, veteran status, and race, most of the information still came from the Characteristics of Business Owners dataset.

More research can be done by comparing the state data with the industry data. Considering the geographic distribution of different statistics (number of family-owned businesses, annual payroll, and employee count), it would be interesting to see how they break down by each industry in each state. Some states and metro areas are hubs of activity for certain industries, and it would be interesting to see how family-owned businesses fit into these industries in different areas.

Citation

Census Data API: /data/2018/abscb. (2019). [Dataset]. U.S. Census Bureau.

https://api.census.gov/data/2018/abscb.html