

Census Assessment

Radha Avudaiappan, Dustin Fife, Michael Zheng

Introduction

The purpose of this study is to analyze owner, employee and technology characteristics of businesses across the U.S., so that we can gain an understanding of: 1.) Who are our business owners? 2.) Who are our business employees? 3.) What problems are businesses having with technology? By asking these three questions, we hope to gather a holistic summary of the state of businesses across the U.S. The data for this study was retrieved from the 2019 Annual Business Survey from the United States Census¹. The data was extracted using the following APIs provided: Company Summary, Characteristics of Businesses, Characteristics of Business Owners, and Technology Characteristics of Businesses. The data was narrowed down to businesses that provided Professional, Scientific, and Technical Services (NAICS Label). Visualizations for this data were developed using Matplotlib, Plotly, and Word Cloud.

Characteristic of Business Owner Visualizations

Breakdown of Business Owners By Sex

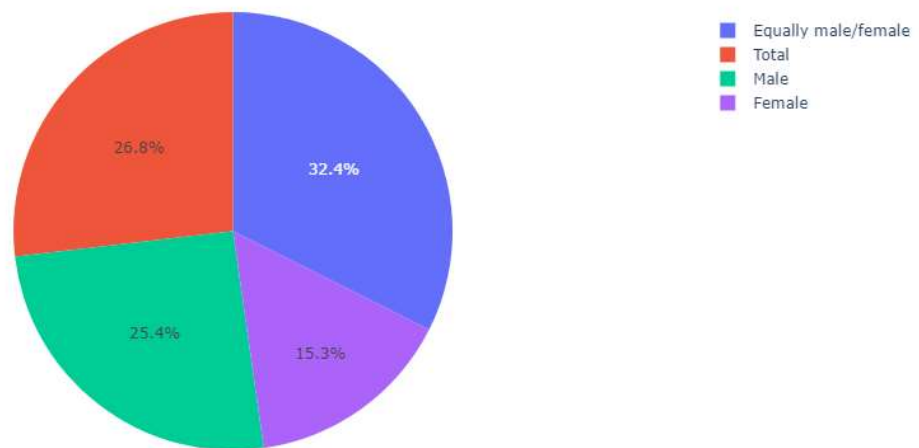


Figure 1: Pie Chart Breakdown of Gender of Owners in Firms

This pie chart was developed using the Characteristics of Business API. The data was filtered down to firms that provide Professional, Scientific, and Technical Services. This data specified genders of owners where any one person owns more than 10% of the company's shares. In addition, firms that reported their gender breakdown as Classified or Unclassified were filtered out since there was no context for what those labels referred to. This graph was generated using Plotly. The "Total" value refers to businesses that were aggregated in a way that could not

specify the gender of their owners. Around a third of businesses have equally male and female owners with 15% of businesses having a majority of female owners.

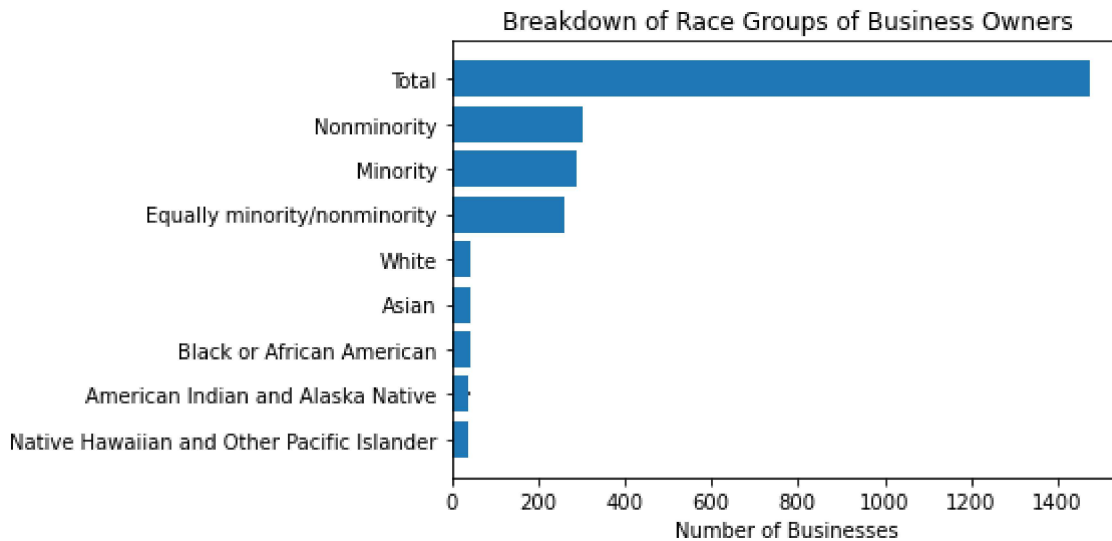


Figure 2: Horizontal Bar Graph Breakdown of Race Groups of Owners in Firms

The chart was developed using the Characteristics of Business API. The data was filtered down to firms that provide Professional, Scientific, and Technical Services. This data specified races of owners where any one person owns more than 10% of the company's shares. In addition, firms that reported their race breakdown as Classified or Unclassified were filtered out since there was no context for what those labels referred to. It is important to note that users of this survey had the option to select more than 1 option for this question. This graph was generated using Matplotlib. The "Total" value refers to businesses that were aggregated in a way that could not specify the race of their owners. The majority of businesses are not owned solely by one race group, rather a mixture. However, more businesses do characterize their owners as part of a nonminority race group.

Number of Business Firms in Each State

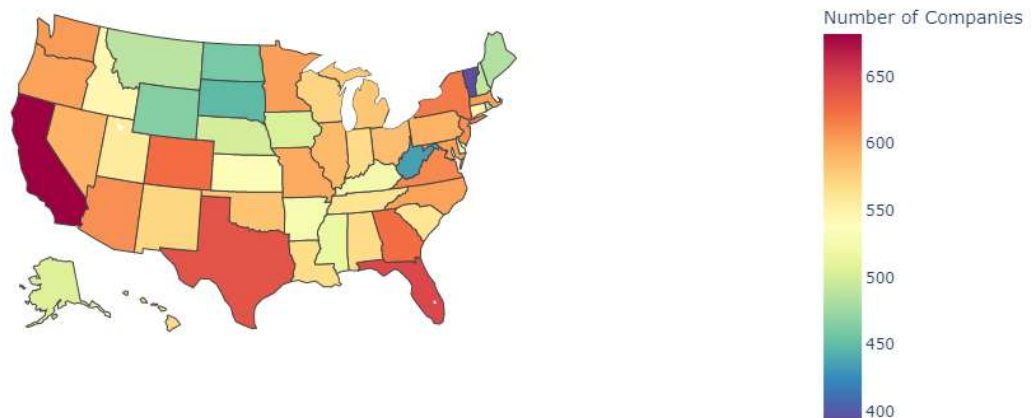


Figure 3: Breakdown of Number of Business Firms in Each State (Choropleth)

This diagram was developed using the Company Summary API. The data was filtered down to firms that provide Professional, Scientific, and Technical Services. In order to map to businesses by each state (instead of FIPs Code) a separate dictionary was created in order to map each state name to the state abbreviation. This graph was generated using Plotly.

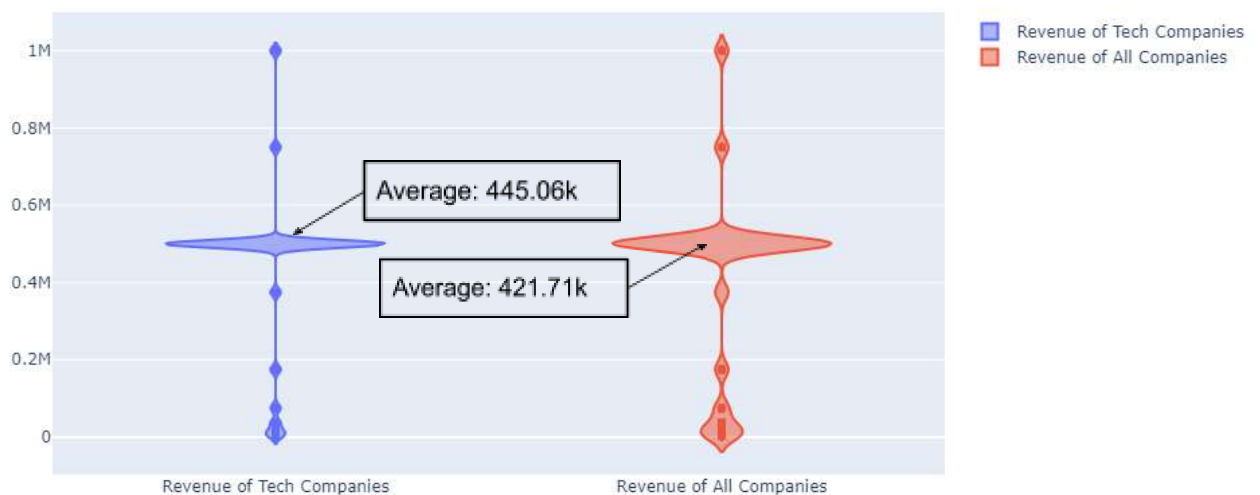


Figure 4: Violin Plot Comparing Tech Companies with All Companies

This diagram was developed using the Company Summary API to show the spread of Revenue for the two groups. The blue violin diagram was filtered down to firms that provide Professional, Scientific, and Technical Services. The red violin diagram included firms from all sectors. The revenue was provided in ranges. These instances were converted to numerical values of the average of their range. (Ex. Firms with sales/receipts between \$5,000-\$9,999: The revenue for was reported as \$7500). For the instances where data was aggregated into 'All Firms' the reported revenue was labeled as \$500,000 since this was the average of the overall range that the Census analyzed. This graph was generated using Plotly. It is important to note that this is an estimate of the revenue of these two groups since a lot of assumptions had to be made. The average for firms in the tech sector was slightly larger than the overall average revenue for all firms (around 20k).

Employee and Employer Data Visualizations:

The visuals in this section detail some of our observations for the employee and employers data shown within the census dataset. The data was obtained from the Company Summary API, and the charts were generated using Plotly.

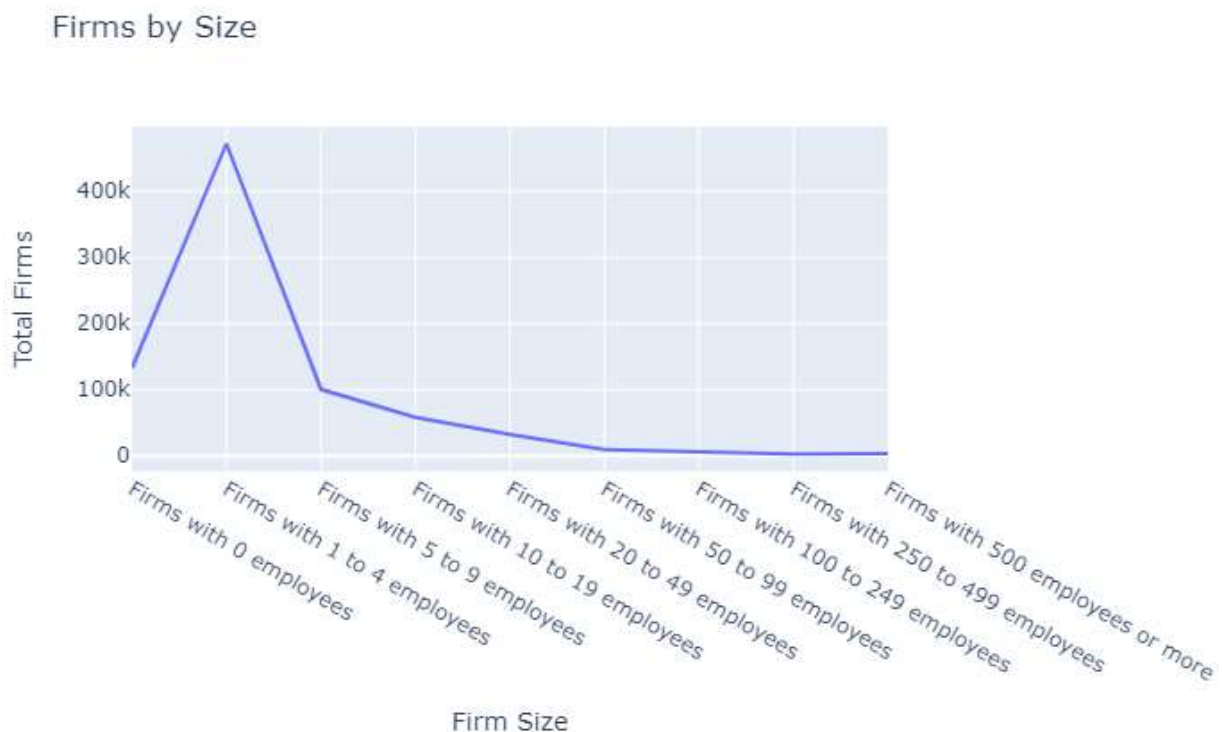


Figure 5: Line Chart Comparing Tech Firms by Size

This model observes the number of firms present within the Professional, Scientific, and Technical Services industry of the USA sorted by size. Based on the information from the chart,

you can see that there is a large number of firms with 4 or less employees, and the number of firms that employ over 100 people is fairly limited. The distribution is very skewed towards small businesses, so in terms of overall quantity of businesses there are many more small firms than large.

To check if this trend continues across multiple factors, we decided to check the total annual payroll by the firm's size. Despite the large number of businesses with less than 10 employees, the vast majority of annual payroll is paid by the firms with 500 employees or more, as shown below:



Figure 6: Line Chart Comparing Firm Size with Annual Salary

So, while the total number of large firms is quite small, this does not seem to impact their powerful financial position within the industry. This makes sense, considering the existence of the big-tech powerhouses such as Google and Microsoft. Now that we have established a bit of the landscape regarding firm size and shape, employee characteristics were observed next. Looking at the overall contrast of sex groups within the workforce we get the distribution below.

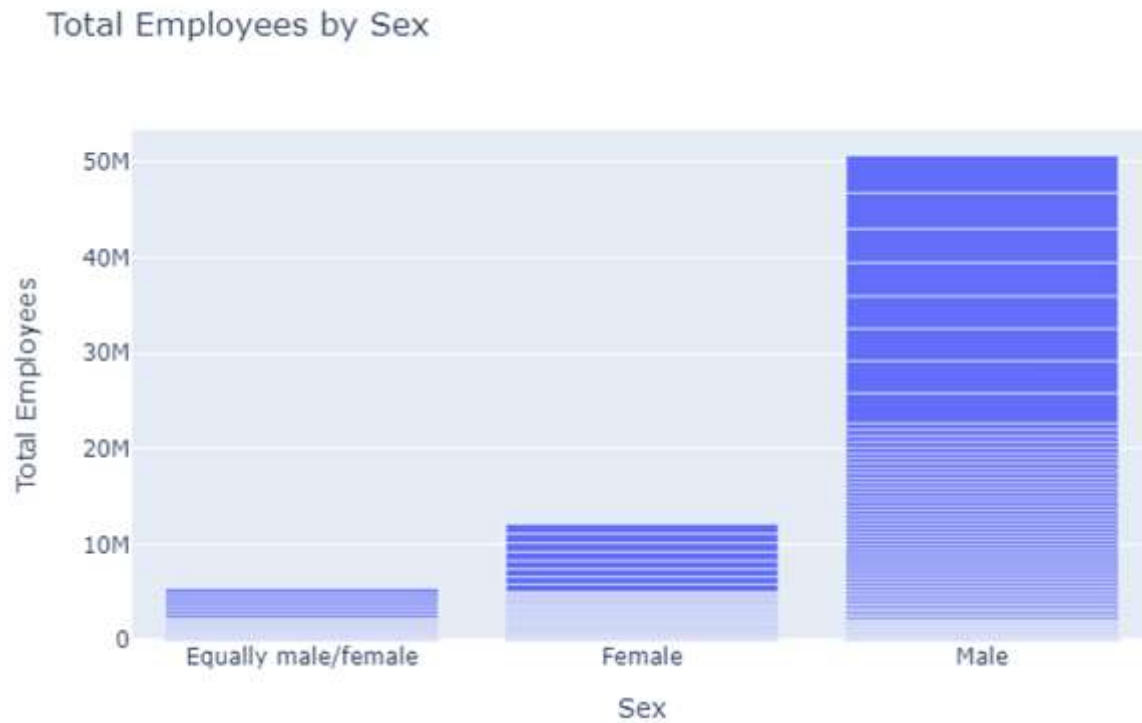


Figure 7: Column Chart Comparing Total Employees by Sex

It became visible that within the professional and technical service industry males are vastly over-represented. Over 50 million instances within our dataset were male, and female employees only represented about 20% of the overall population, with equally male/female coming last at less than 10%.

Finally, the last factor we analyzed was the total amount of salary being paid to employees based on race. It turns out that the majority of wages is paid to white or non-minority workers, while minority and Asian and well below these two. This is possibly because of the distribution of races within the industry which is heavily white and non-minority. Outside of these four groups, there was very little to no representation within our dataset for the last set of backgrounds.

Annual Payroll by Race

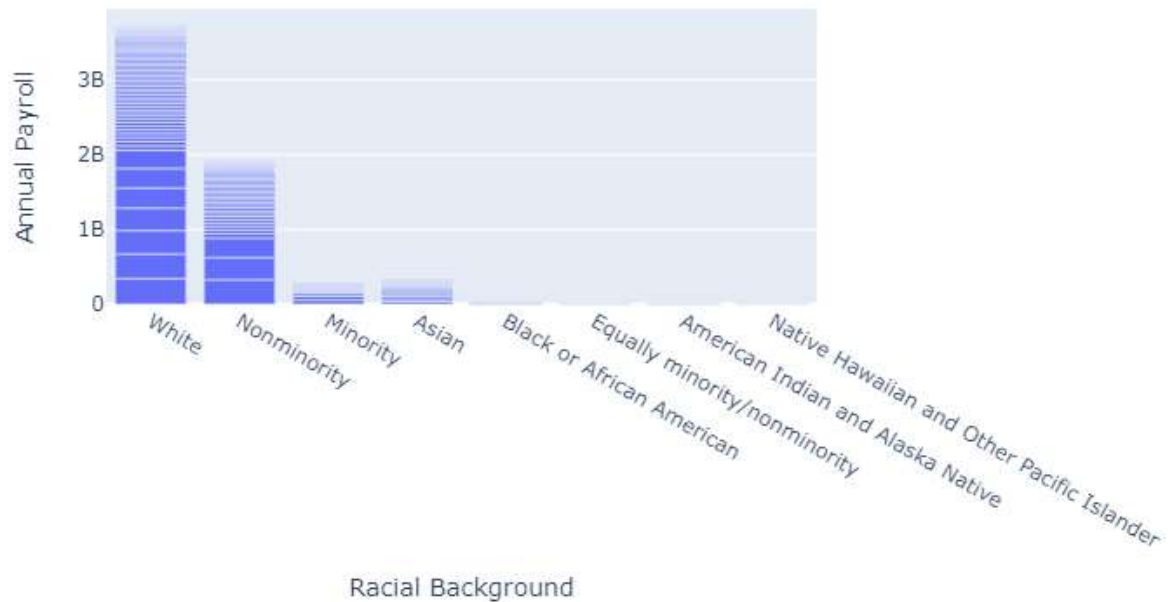


Figure 8: Column Chart, Annual Salary by Race

Characteristics of Business Technology Visualizations:



Figure 9: Word Map: Problems Adversely Affecting Technology Production

The word map in figure 9 was produced using the Characteristics of Business Technology data set and gives us a visual representation of what problems are affecting companies in the U.S.. The variable “FACTORS_P_LABEL” from the data set gives us information about problems affecting different technologie’s productions across all U.S. companies, and is of the form “Technology : Problem Description”. Once the endpoint data for “FACTORS_P_LABEL” was extracted using the appropriate API URL, the column was split into two separate columns, “Technology” and “Problem Description”. After getting the set of all unique problem descriptions, the word map in figure 9 can be produced by inputting the set of words into a word cloud.

Artificial Intelligence



Figure 10: Tree Map: Problems w/ AI across all US Firms

Cloud-Based



Figure 11: Tree Map: Problems w/ Cloud across all US Firms

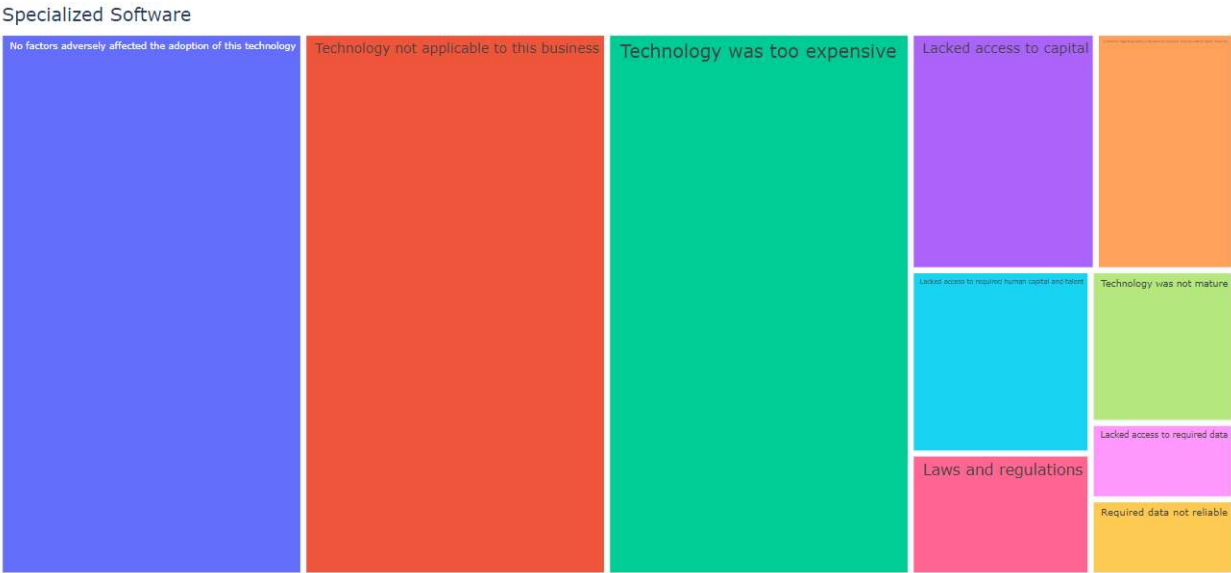


Figure 12: Tree Map: Problems w/ Specialized Software across all US Firms

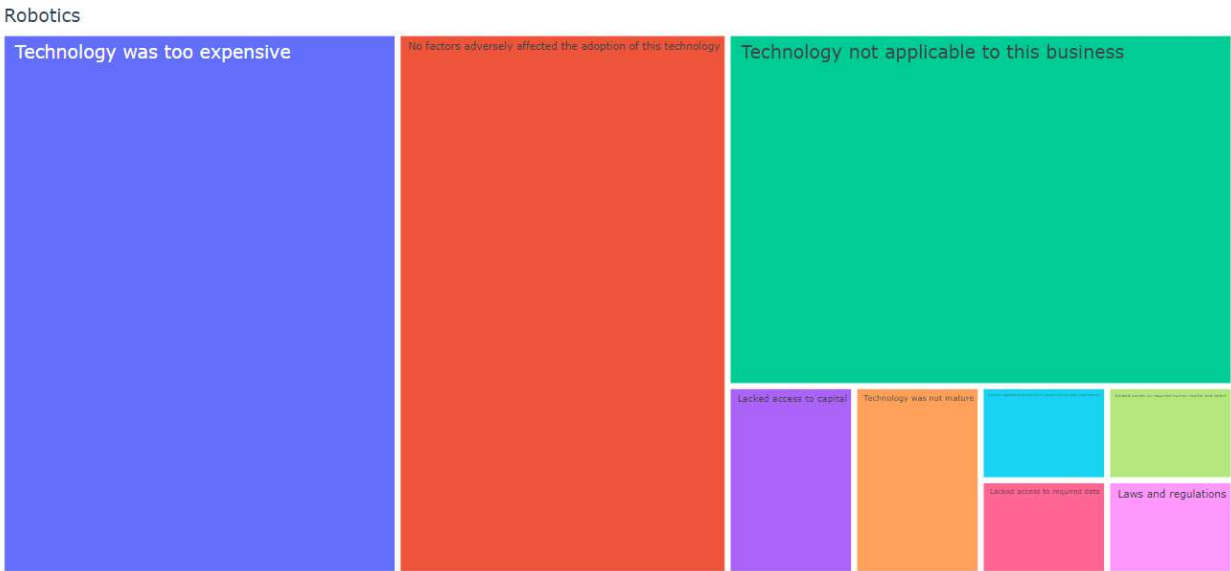


Figure 13: Tree Map: Problems w/ Robotics across all US Firms

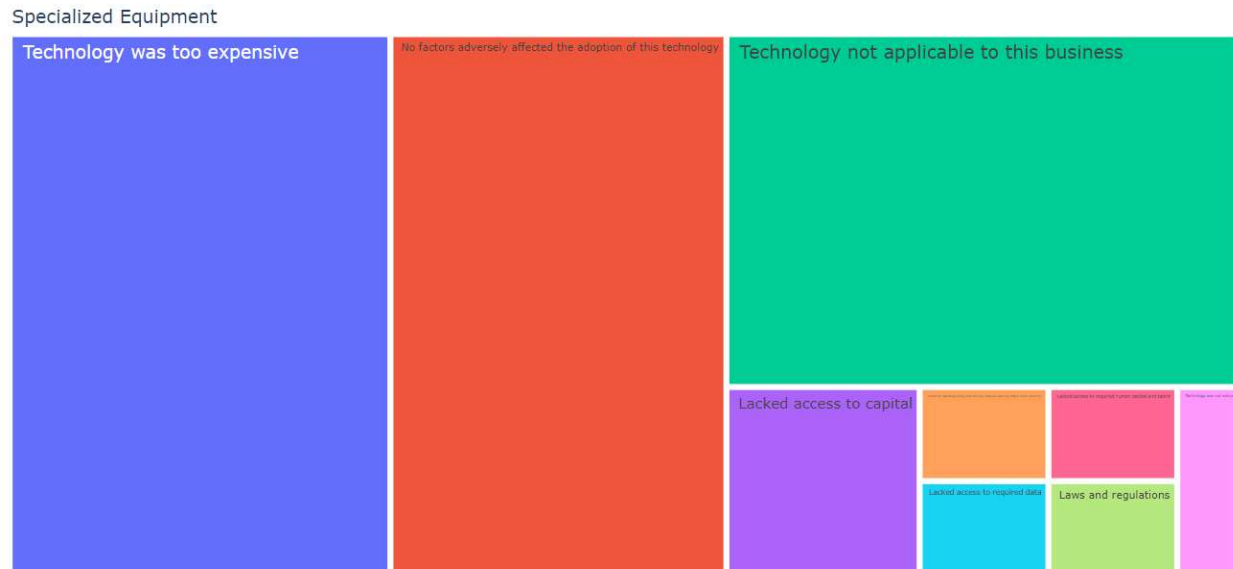


Figure 14: Tree Map: Problems w/ Specialized Equipment across all US Firms

The tree maps in figure 10-14 were produced using the Characteristics of Business Technology data set and gives us a visual representation of how severely different problems are affecting companies in the U.S. for five different technology categories. According to figures 9-14, “technology not applicable to this business”, “technology was too expensive”, and “lacked access to capital” were the dominant problems for business across all five technologies. In addition to the split variable “FACTORS_P_LABEL” used in figure 9, the variable “FIRMPDEMP” from the Characteristics of Business Technology data set was used to generate all five tree maps; this variable gives the count of firms across the U.S. that reported each problem for each technology.

Conclusion

With respect to the question, “Who are our business owners?”, we discovered that around a third of businesses have equally male and female owners with 15% of businesses having a majority of female owners and more businesses do characterize their owners as part of a nonminority race group.

With respect to the question, “Who are our business employees?”, we discovered that the average for firms in the technology sector was slightly larger than the overall average revenue for all firms, males are overly represented in technology, and the majority of workers ,and thus the payroll, are composed of non-minorities.

With respect to the question, “What problems are businesses having with technology?”, we discovered that the main problems adversely affecting technology production today are applicability, cost, and capital.

References

¹Bureau, U., 2021. *Annual Business Survey (ABS) APIs*. [online] The United States Census Bureau. Available at: <<https://www.census.gov/data/developers/data-sets/abs.html>> [Accessed 27 June 2021].