Professional, Scientific and Technical Businesses: Demographics and Technology Use

Abby Bilger, Linsey Stokes, and Nick Verbus

Introduction

The technology industry is the fastest growing industry in the United States and is constantly creating and utilizing new technology to serve businesses and individuals. This industry has far reaching influence and plays a role in the advancement of almost every other industry in the United States. In addition, due to the fact that this industry focuses on and works extensively with technology, businesses within this industry are more likely to come into contact with and utilize new technologies. Due to these factors, our report seeks to look deeper at several aspects of this industry. As a general overview, we examine the overall demographics of the industry, focusing on race and gender. We also attempt to gain a deeper understanding of how businesses in this industry utilize technology, particularly developing and growing technologies such as Robotics, Artificial Intelligence, and Cloud-Based services.

Data

All the data used in this analysis was taken from the United States Census Bureau's APIs for the 2019 Annual Business Survey. This API can be found [here](#). The Annual Business Survey data includes information about United States businesses such as industry classification, employee and employer demographics, characteristics, and statistics such as number of employees and annual pay. This data can be broken down to examine states, metropolitan areas, or the entire United States. During our analysis, we looked at the whole United States as opposed to individual states or metropolitan areas because it better represents the current condition of the technology industry.

To explore the datasets, we began by individually requesting API keys and loading the APIs to a Jupyter notebook starting with the example queries found at the link above. To get a sense of the data we would be working with, we looked at the column information, the census documentation, and explored case studies that utilized the census data using Python. Individually, we explored the datasets using pandas' built in functions to observe the sizes, columns, and other information found in each dataset. Following our initial exploration, we each created basic charts to observe the information available. Individually we analyzed the data to explore our own interests, both, in demographics such as race and gender or, in characteristics of businesses related to technology.

The data on technology use was specifically taken from the Technology Characteristics of Businesses dataset available from the API. To obtain this data, we used the Python Requests

library to query the API, getting most of the available variables in the response, to minimize the amount of aggregation done by the API behind the scenes. The response from the API was loaded into a pandas dataframe, and filtered afterwards.

First, all columns not relevant to the analysis were dropped from the dataframe. This included several columns with data not being examined. In the case of data that was represented across multiple columns with a code column and a label column, such as information on the business owners' sex, only the label columns were kept to ensure that the data was easily readable. The remaining columns included several columns of numeric data that were, at the time, typed as objects, so the next step was to convert the columns containing the count of responses, the total annual payroll, and the total annual revenue into integers, so that we could later use the data in calculations.

The next task was to filter down to the specific data we wanted to include in our analysis. In this case, we were interested in looking at specifically the businesses related to technology, so we filtered only to rows that were classified under the North American Industry Classification System (NAICS) as "Professional, scientific, and technical services". Because we wanted to focus our analysis mainly on this category, we also filtered the data to ensure that other varaibles, such as the business owners' sex, race and veteran status, only included rows for the total across all categories of that variable. The penultimate step was to filter out the rows containing the total reporting for the questions that we wished to analyze, so that we could focus on individual categories of responses. From there, the final step was calculating columns for the average payroll and the average business revenue. These columns were created by dividing the columns containing the annual payroll and total sales, receipts or revenue, respectively, by the number of firms.

Analysis

Business Demographics

One area we analyzed is the demographics of business owners by minority status and gender. As shown in Figure 1, many technology business owners are nonminority men, with men owning most technology businesses in general. A much smaller portion are owned by women with even less owned by minority women.
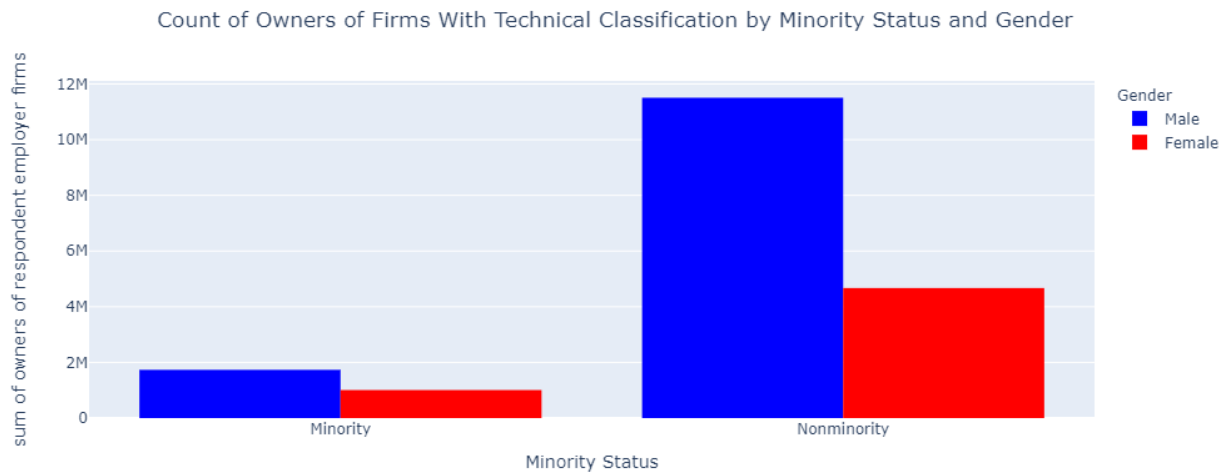
Figure 1. The number of technology related businesses owned by men and women separated by minority and nonminority status.

To expand upon the difference in ownership among different races, we further separated the groups of minority and nonminority to be more granular over different races. As shown in Figure 2, we can see that an overwhelming majority of technology businesses are owned by white men, many are owned by white women, some are owned by asian men and asian women, and almost none are owned by any other racial minority.
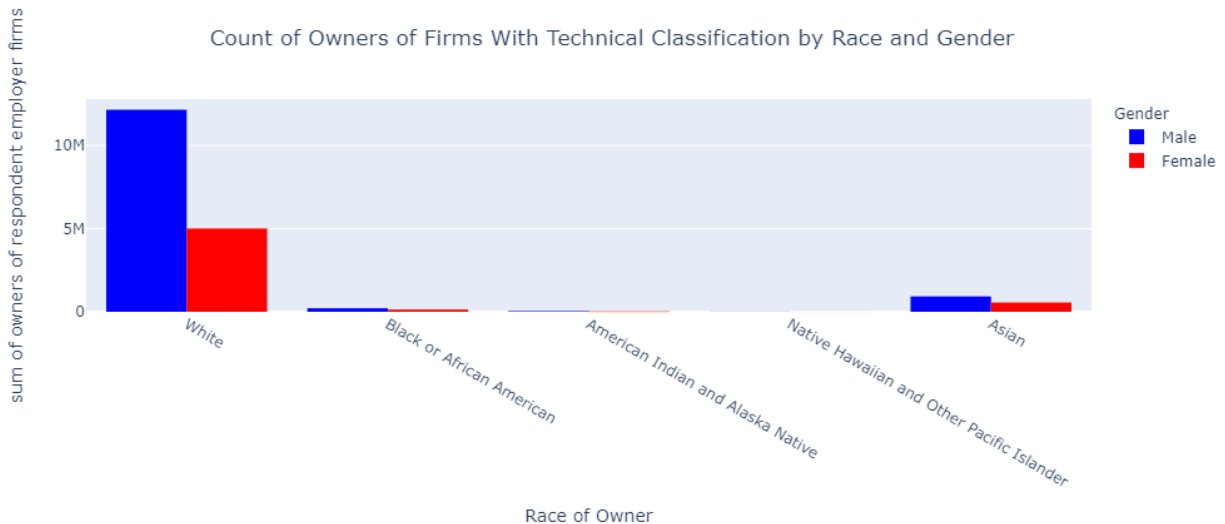

Figure 2. The number of technology related businesses owned by men and women separated by racial group.

When observing the number of employees that work in a technology centered business, we see a similar pattern of demographics as seen in Figure 2 looking at ownership. The most

notable difference in Figure 3 shows almost double the amount of white men employees compared to business owners while all other groups remain nearly the same across ownership and employment.
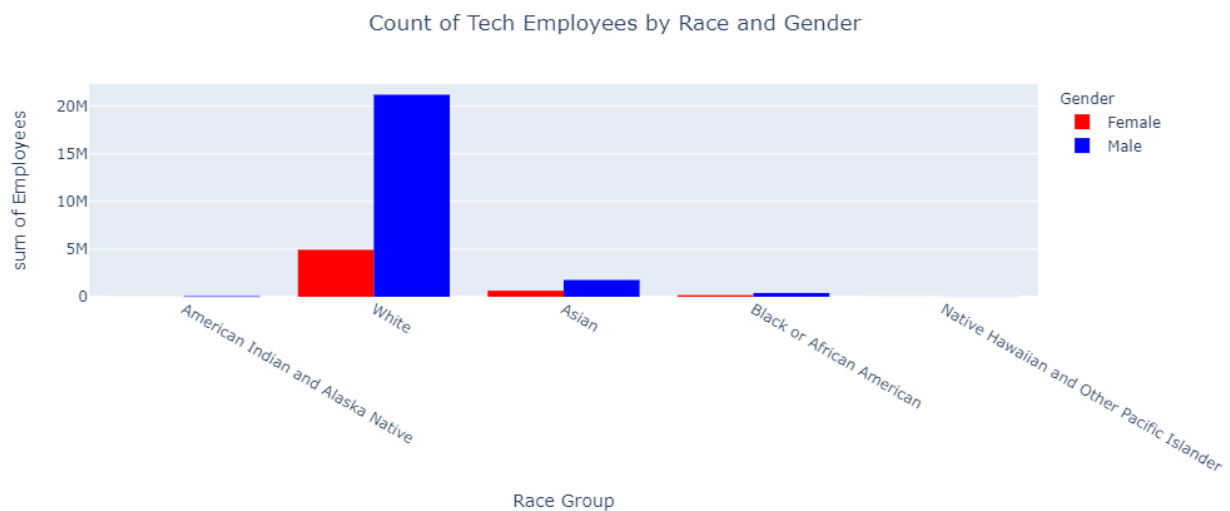

Figure 3. The number of employees at technology related businesses by gender and racial group.

Technology Use

Beyond demographics, we also wanted to examine the technology use of these companies. In particular, we focused our analysis on the responses to two questions: the extent of technology use and the factors adversely affecting technology use. Within these questions, we primarily looked at three of the technologies: Artificial Intelligence, Cloud-Based services, and Robotics.

To begin, we investigated the responses regarding the extent of usage of each of these technologies. In the case of Robotics and Artificial Intelligence, the vast majority of businesses did not use these technologies at all, with only a small proportion reporting any level of usage of these technologies. For cloud-based technologies, the distribution was very different. Overall, more businesses utilized cloud-based technologies than did not, and the proportion of businesses using and not using cloud-based technologies was much closer than for the other technologies.
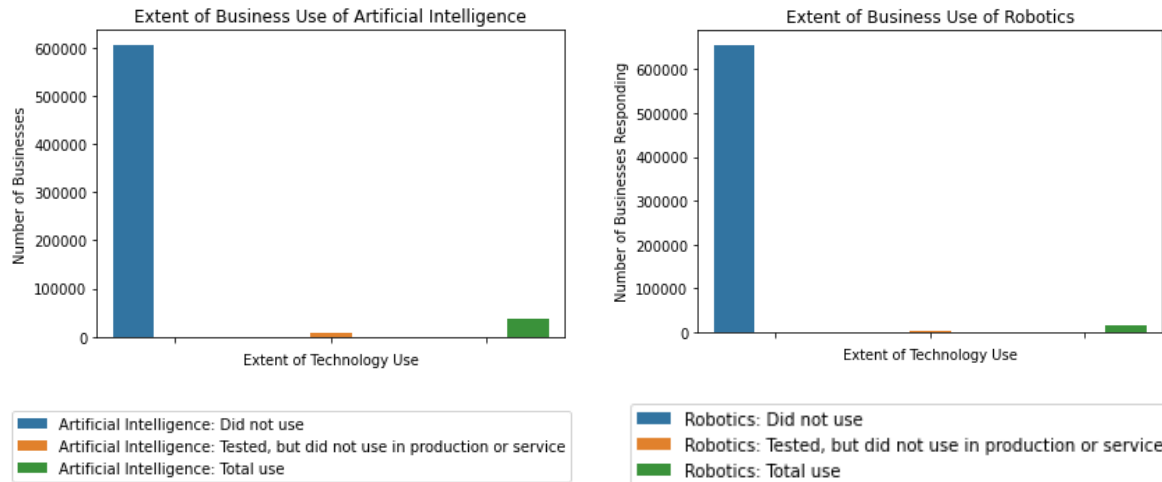
Figures 4 and 5. Results for usage of Artificial Intelligence and Robotics technologies. In both of these cases, businesses that utilized the technology were significantly in the minority.
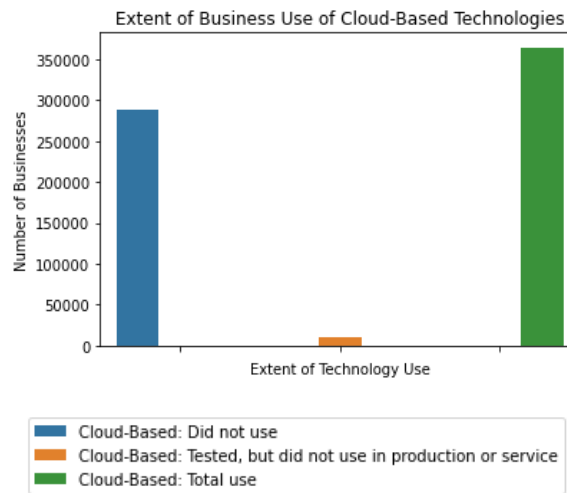

Figure 6. Results for usage of Cloud-Based technologies. The majority of technology businesses utilized cloud based services, though proportionally there are still many that did not.

In light of the extremely low usage of Artificial Intelligence and Robotics, we wanted to dive deeper into the motivations behind the lack of adoption of these technologies. Additionally, we wanted to compare these responses to the responses for Cloud-Based technologies, which seem to be much more accepted in this section of industry. To do so, we analyzed the responses to the question on factors adversely affecting the usage of these technologies, to see if there were any significant factors, or trends across technologies. When visualizing the responses, the significant majority of responses for all technology categories were either that there were no factors adversely affecting use of the technology, or the technology was not relevant to the business. In order to get a more detailed look at the factors

affecting the use of these technologies, we excluded these two response categories in our visualizations below.
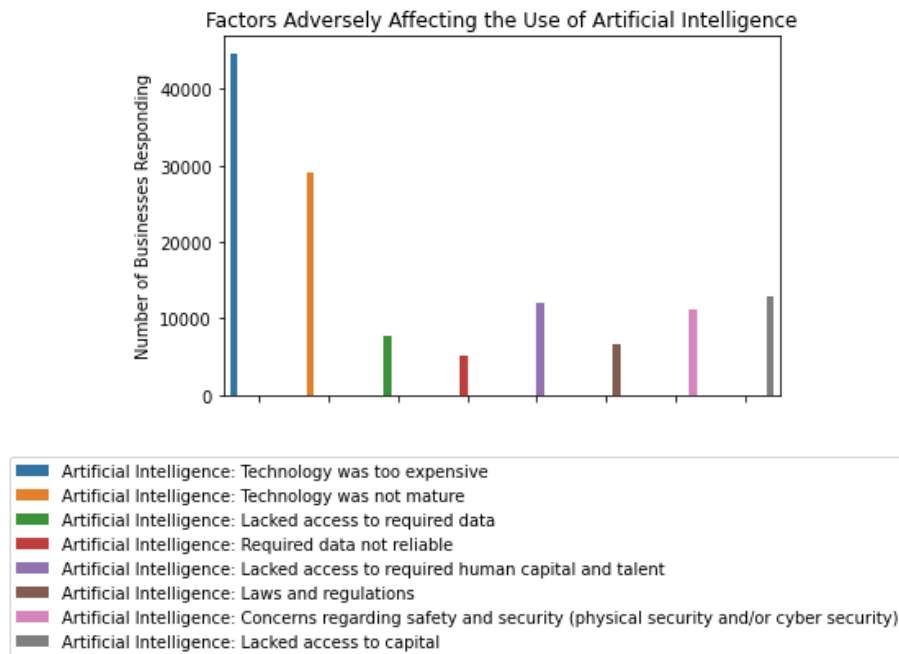

Figure 7. Results for factors affecting adoption of Artificial Intelligence technologies. The primary factor was cost, followed by the level of maturity of the technology.
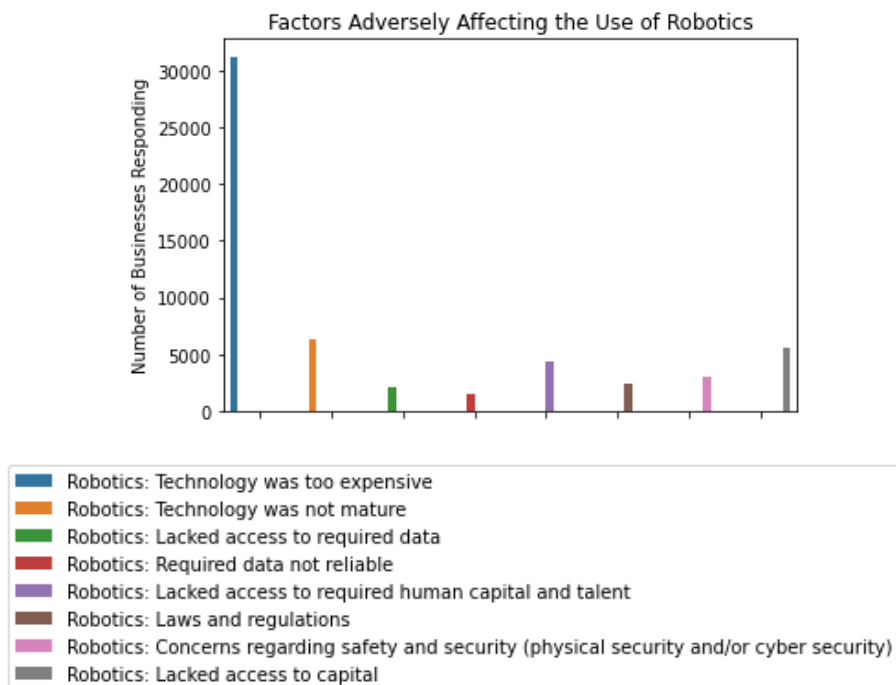

Figure 8. Results for factors affecting adoption of Robotics technologies. Cost was by far the most significant factor.

For both Artificial Intelligence and Robotics, the most cited factors discouraging businesses from utilizing these technologies was cost. In both cases, lack of access to capital, a factor also related to the price of the technology, was one of the top factors as well. This suggests that financial concerns are a primary factor affecting the adoption of these newer, more experimental technologies. In the case of Artificial Intelligence, there was a secondary concern that the technology was not yet mature enough to use. It may be the case that Artificial Intelligence technologies are more expensive, as they are still in relatively early stages of development, or perhaps, businesses may be less willing to pay money for a technology that they see as not yet mature, leading to a lower threshold for what is considered "too expensive" to utilize. Another thing worth noting from these graphs is that the overall number of responses for Artificial Intelligence is higher than that for Robotics, suggesting that there are more businesses who are interested in the former technology.
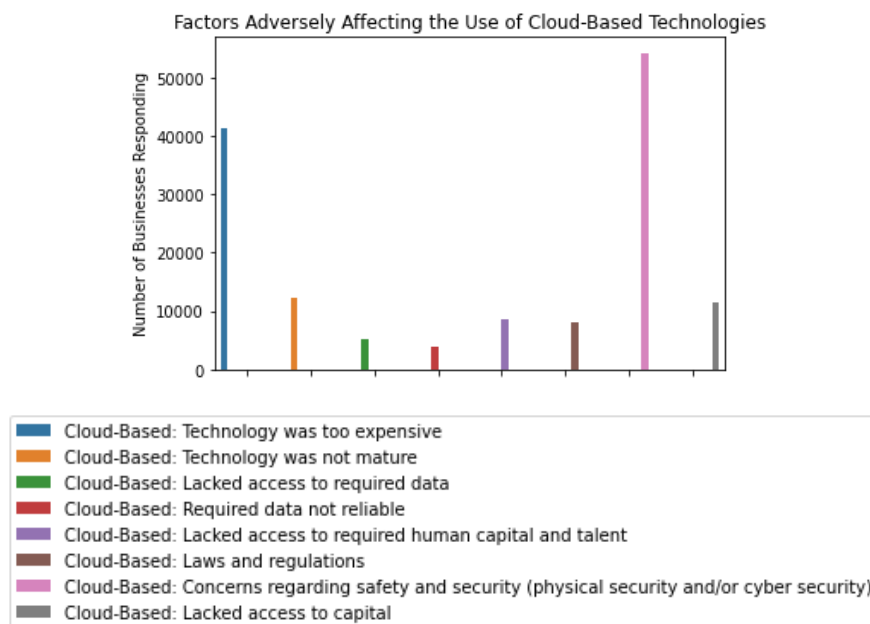

Figure 9. Results for factors affecting adoption of Cloud-Based technologies. There is much more variety in factors than for the previous two technologies.

Compared to Artificial Intelligence and Robotics, the responses for Cloud-Based technologies showed a different primary concern, and a greater variety of overall factors. In this case, the most significant factor was the safety and security of the technology, though the cost was still a significant factor as the second most common response. It is worth noting, again, that the overall number of responses here is higher than for the previous technologies, which suggests again that cloud-based technology is the subject of more interest. This is supported by the fact that it is much more widely adopted than the previous technologies.

Although there was some variety in factors for each individual technology, across all categories, cost was clearly the primary factor that adversely affected businesses choosing to utilize new technologies. Given this, we wanted to look deeper into the relationship between a business's financial situation and their adoption and use of various technologies. We hoped to be able to utilize overall business revenue in this analysis as an approximate measure of the resources a business may have to spend on new technologies, however revenue data was not available for the professional, scientific, and technical services businesses within our dataset. Instead, we utilized annual payroll data for this purpose, as our earlier exploration found that there was a strong relationship between revenue and payroll in general. As seen below, the annual payroll of the businesses who deemed these technologies "Too Expensive", on average, fell approximately within the range of $1.2-$1.6 million.
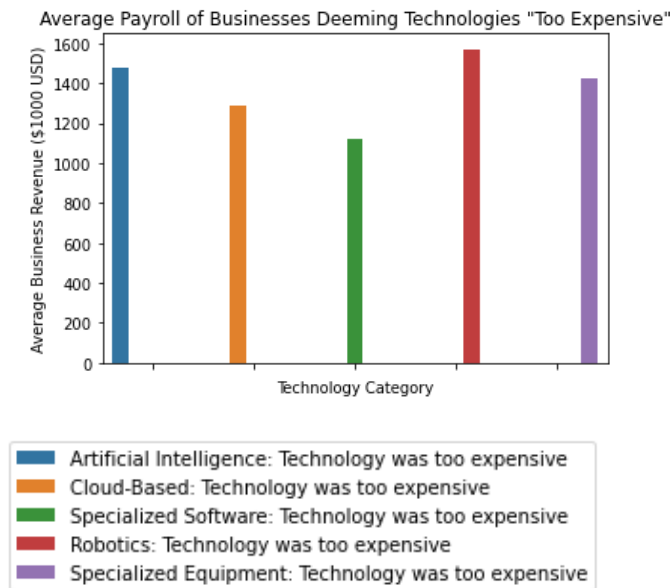

Figure 10. Across multiple technologies, the average payroll of businesses that considered the technologies "Too Expensive" was fairly consistent.

Although this data provides some information, its usefulness is limited without the additional context of payroll data from businesses that do utilize these technologies. To investigate further, we returned to the first question, regarding the current usage of technologies, to compare the average payroll of companies currently using the technology with those that do not. This analysis showed a clear difference across every category of technology, between businesses that do and do not use a given technology. In every case, the businesses using the technology had a higher payroll, often by a very large margin. This difference was most extreme for the least utilized technologies: Artificial Intelligence and Robotics, with businesses using robotics having the highest overall average payroll. Without additional information, we can guess, but not conclude with certainty, that the higher payroll (and,

presumably, overall higher level of financial resources) enables companies to utilize new and advanced technologies, as it may also be the case that using these technologies makes a business much more profitable. However, it is clear from the data that there is a strong relationship between technology use and the financial status of a business.
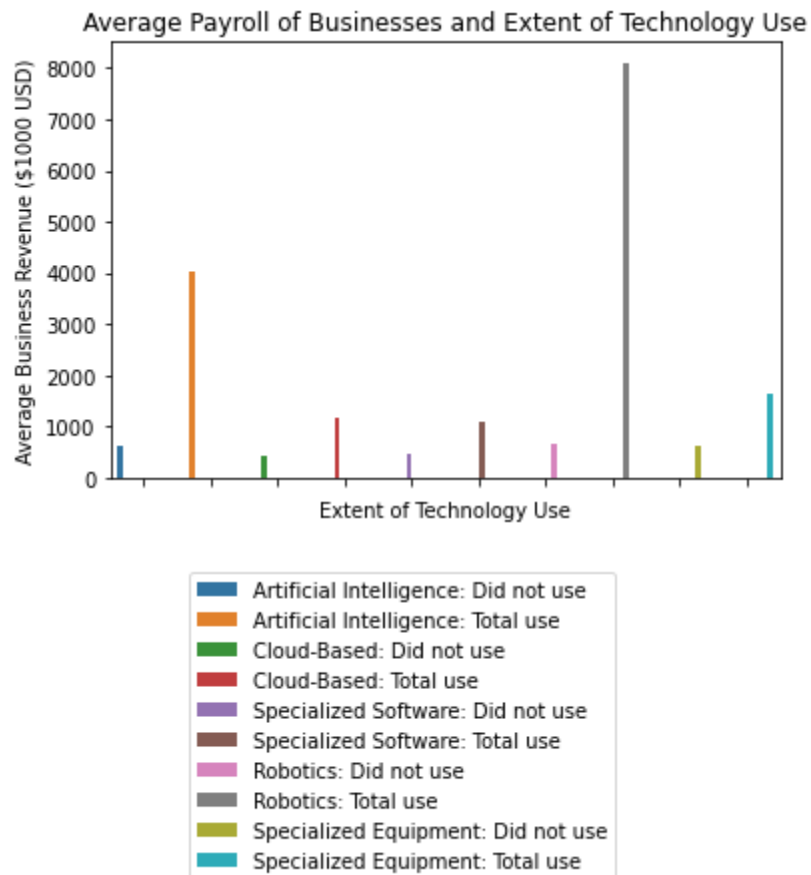

Figure 10. For all technologies, the average payroll of businesses that used the technology was significantly higher than the average payroll of those that did not. This difference was especially extreme for Robotics and Artificial Intelligence.

Statistics

Finally, we analyzed the numeric data to determine relevant statistics such as number of employees, annual payroll, and number of employer firms within the industry. Figure 12 shows that there is a positive correlation among all of these properties with the number of employees and annual payroll being almost completely linear with very little variation. This shows that there is a trend between an increase in payroll correlating with an increase in employees and vice versa.

Figure 12. Scatter plot matrix of number of employer firms, employees, and annual payroll (in $1000).

Conclusion

The technology industry is far-reaching in its influence on United States business. It is primarily composed of white men and needs to do more to address the lack of diversity in employment and ownership. In addition to discrepancies in numerical representation across sex and racial groups, there are also large discrepencies in pay. As for the adoption of technology, it seems that while Cloud-Based services have gained wide acceptance and are close to becoming an industry norm, other major technologies such as Robotics and Artificial Intelligence are still developing and are in the early stages of being adopted, with only a few businesses utilizing them. The largest factor stopping businesses from using these and other new technologies is cost, as they are currently too expensive for many business owners. This can also be seen in the large difference in annual pay between businesses that utilize these technologies and businesses that do not. Further research would need to be done to better understand the nature of this connection, and understand how to address the differing access to these technologies.