**Project Report**

Greg Wagner, Chris Nash, Sawyer Tucker, Danny Brickman

January 16, 2021

For this project, we were tasked with reviewing US Census data, calling to the corresponding APIs for this data, and making conclusions based on the data with which we worked. The APIs, called the Annual Business Survey (ABS) APIs, focused on information regarding businesses in the United States, and from this information, we wanted to answer several questions:

1. What is the demographic breakdown of business owners?
2. Which demographic has the most employees, and how do the demographics compare to each other?
3. Do the differences in demographics lead into any disparities in metrics such as payroll?
4. What are the differences in technology uses between different demographics of businesses?

To answer these questions, we first extracted and transformed the data from the US Census website to better understand the kinds of visualizations we could create. The data we used came from four different surveys[[1]](#footnote-510), listed below:

* Company Summary
* Characteristics of Businesses
* Characteristics of Business Owners
* Technology Characteristics of Businesses

The data comes from the 2019 version of the ABS and the link to the APIs we used can be found in the footer below.

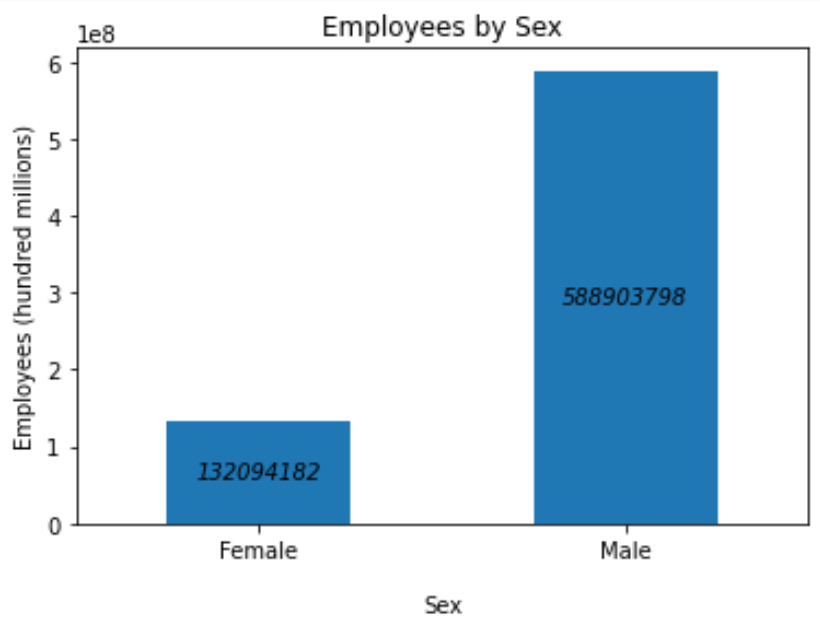
As stated above, the data had to be extracted and transformed before creating the visualizations to answer these questions. The main transformation done on these datasets was removing columns that were unnecessary to answer the questions. The data was loaded into Python using the pandas library. The visualizations along with the conclusions drawn from them are shown below.

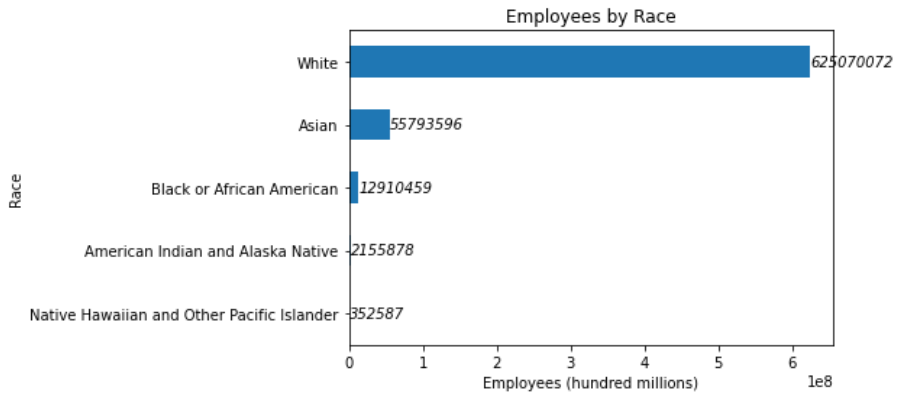
**Visuals**

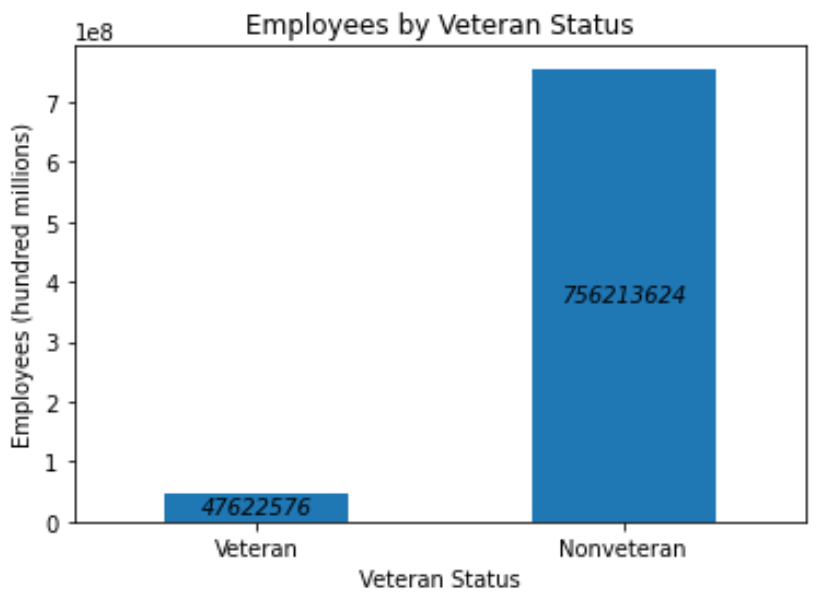
The first three visualizations were created in order to answer the question of which demographic has the most employees in the United States. The first thing that we noticed examining this data is that the number of employees, summed over all variations, is much greater than the total population of the United States. Each row represents a single company's information. Our hypothesis that the number of employees is so high because a person can work/affiliate with multiple companies. This data may also include contract and freelance work, which may allow one person to be listed with many companies.

Another thing to address is that the graphs below received feedback that allowed revision. The three graphs now have an increased font size on the axis for readability, and while the feedback to include bar labels was directed at the second visualization, they are helpful, and we extended this to cover all three tables. Finally, the third graph was originally quite unhelpful, and was changed to better answer the question posed.

The three main demographic groups reported in the census were sex, race, and veteran status. Graphs depicting the number of employees in each group are below.







As we see in the above visualizations, most employees by gender are Male, with 588,903,798 employees, in comparison to Female with 132,094,182 employees in the United States. This is a ratio of 82/18 Male/Female, or 82% of employees in America are reported male.

According to the second table, the Race table, most employees are white, with 625,070,072 out of 695,965,292 people, or 89%. This seems egregiously high, even considering the hypothesis we posited earlier, but without conducting further research, we cannot say for sure that this is incorrect.

Finally, we have the veteran status table. In this table, we can see that the workforce are mostly nonveterans, with 756,213,624 employees out of 803, 836,200, or 94% of the reported employees being nonveterans. While this is the most lopsided of the three graphs, it is more in line with the distribution we were expecting.

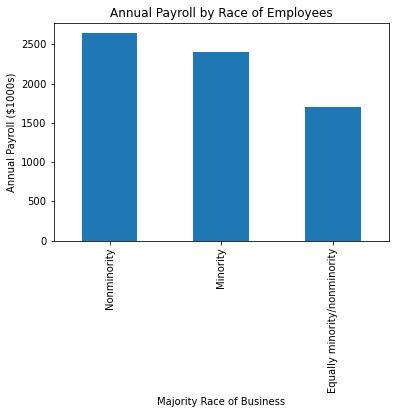
These three graphs show us the most usual form of employee in the country; a white male who is not a veteran. While there are too many combinations of employees between these three graphs, we can easily approximate the likelihood and distribution of employee demographics by combining the three tables above.

Visual 4 (Characteristics of Businesses)



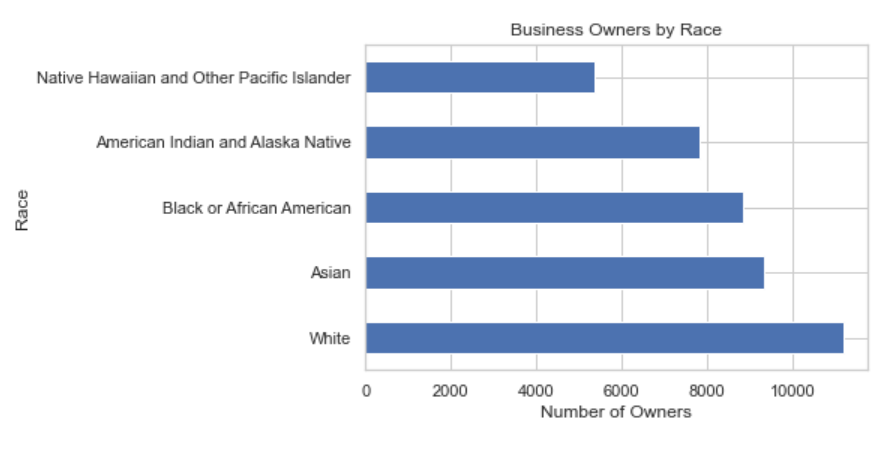
Before explaining the above visualization, it is important to note how Characteristics of Businesses categorizes businesses by sex. This survey asks businesses which sex makes up most of the employees in said business. For instance, the “Equally male/female” category signifies a business in which the demographic makeup of this is almost exactly even between male and female employees. The above graphic shows a marked difference between businesses with majority male employees and majority female employees. The businesses with majority male employees have a higher annual payroll than businesses with majority female employees. This suggests businesses with more male employees spend more on employees than businesses with more female employees. There could be several reasons for this: one possibility is businesses with more male employees may have more employees total which would necessitate a higher budget for payroll. Alternatively, businesses simply may pay more if they have more male employees. Further research is needed to explain this disparity. Interestingly, businesses with an equal split between male and female employees have payrolls more like businesses with majority male employees. This seems to suggest that the presence of male employees increases the payroll of a business. However, research on correlations on these datasets would need to be run to check this relationship.

Visual 5 (Characteristics of Businesses)



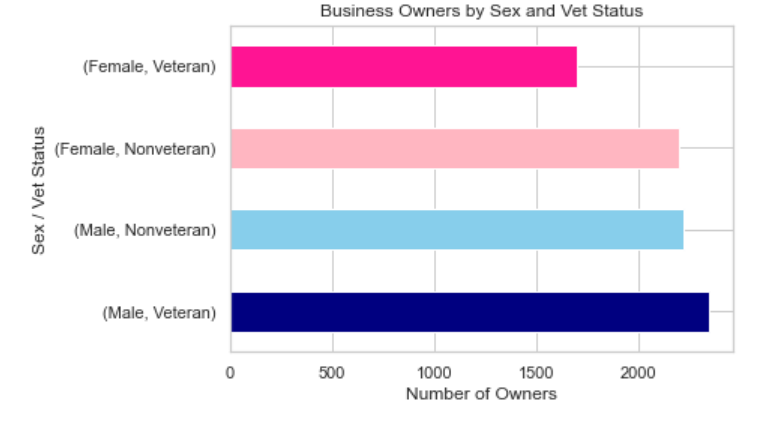
Visual 5 is like Visual 4 in that it is a breakdown of annual payrolls of businesses, but in this visual the comparison is with the race of employees. For this visualization, it is important to note that “Nonminority” refers to anyone that identifies as white and conversely, “Minority” refers to anyone that identifies as non-white. The difference in this graphic is much starker than the last graphic: businesses with more nonminority employees have overall higher payrolls than businesses with more minority employees. Again, we could cynically conclude that businesses pay more when there are more nonminority employees in the business. However, we could also conclude that businesses with more nonminority employees simply have more employees to pay, thus would have higher payrolls. This seems to be supported by the “Equally minority/nonminority” bar, as one would anticipate businesses with an equality in that metric would have fewer employees to pay as achieving this equality is easier to do with less employees. With that said, it would still behoove the analysis to do further research as this data does not sufficiently explain the difference shown in the graph.

Visual 6 (Characteristics of Business Owners)



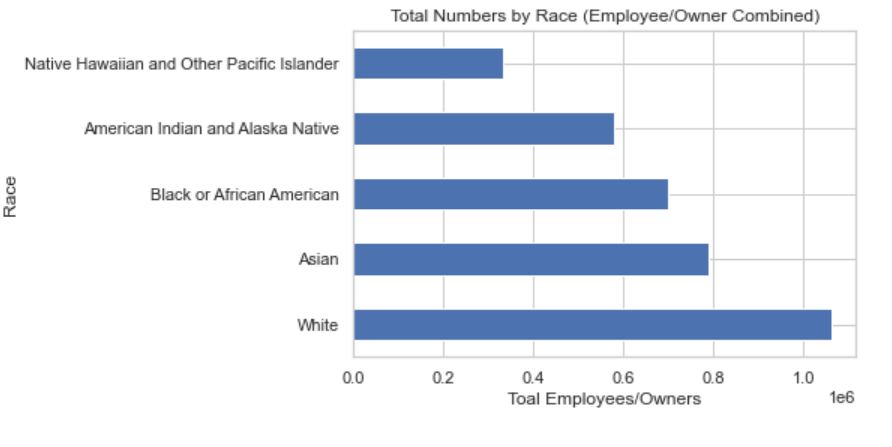
Visual 6 begins to look at the Business Owners dataset by exploring the race breakdown for business owners that responded to this part of the Census. With any demographic question, it is important to realize that the data presented is only as good as the data received. Every question is optional and not everyone answers every question. With that being said, from the above graph it appears that a greater proportion of business owners identify as White than any other race, with Asian in second followed closely by Black or African American. The order of this graph matches with the order of the employee race graph above, however the proportion of each race is different. This is most likely a result of us pulling the data from 2 different data sets and any calculation/removal of data along the way might have affected the numbers presented in the graphs. But even with the numbers and proportions differing, it is still important to reiterate that the order matches and is the main thing we can take away from this graph.

Visual 7 (Characteristics of Business Owners)



Visual 7 was also created from the Business Owners dataset; this time taking the sex of business owners (male/female) and combining it with their veteran status (vet/non-vet). In the graph, it appears that business owners who identify as male veterans are the most common, followed closely by males who are nonveterans. There is a drop off for female nonveterans and female veterans in the data. What I find most interesting about the data is the reverse order that the groups are in; with make veterans over male nonveterans but female nonveterans over female veterans. What is also interesting about this graph is that both male categories have more business owners than the female categories. Overall, there are more male business owners than females, and more non-veteran business owners than veterans

Visual 8 (Characteristics of Business Owners and Businesses)

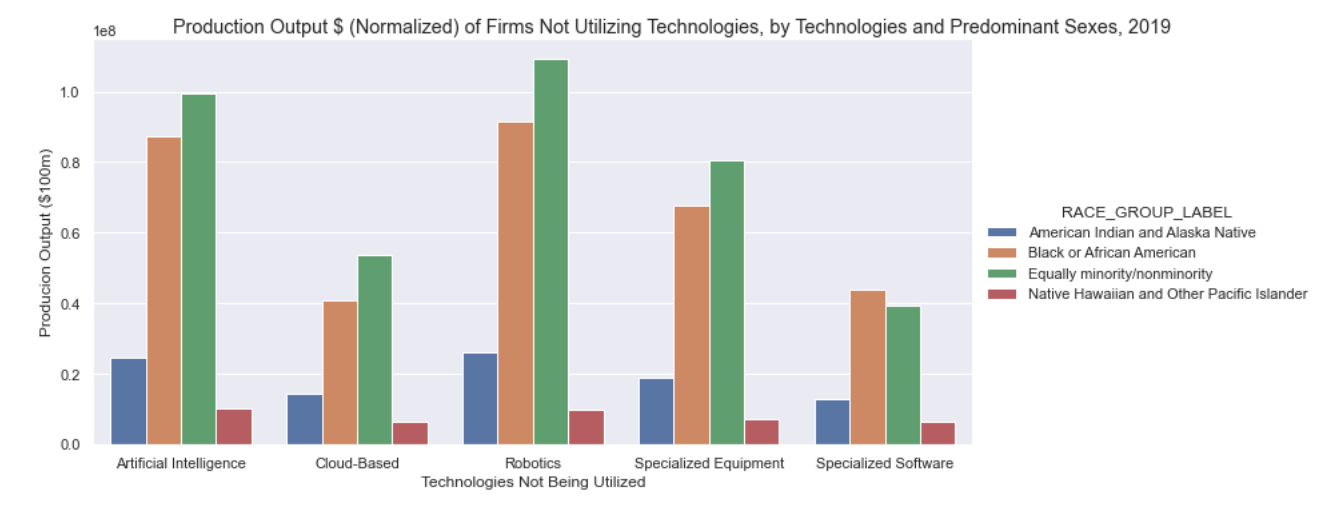


For visual 9, we merged the data from a business employee dataset and the business owner dataset and looked at the breakdown of races. As mentioned in visual 7, the order of the races matched up in both the owner's dataset and the employee dataset. By merging the data, that order gets exaggerated, with Whites clearly having more people working than any other race, followed by Asians then closely by Black or African Americans. It also exaggerates the other end where it is clearer that the Native Hawaiian and Other Pacific Islander race is not as populated in the US Census as the other races and thus doesn’t have as many employees/business owners. For all three graphs that looked at business owners, it would be interesting to explore the dataset further for 2021 during the pandemic and then a couple years later when hopefully we are recovered and more businesses are opening up instead of shutting down.

Technology in Business Survey

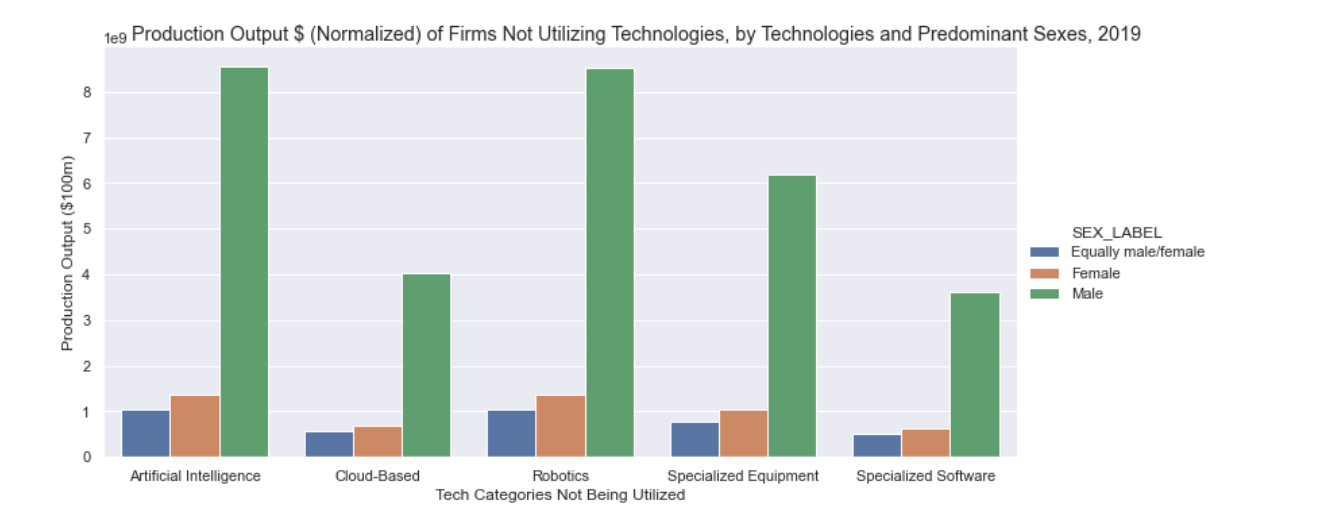
Analyzing the important technologies being utilized by businesses today is crucial to finding the direction the economy is heading. We wanted to analyze some of the demographic patterns affecting the production output of businesses, specifically finding what technologies each group are not using. From there, we want to find the most marginalized technology and see some of the business impacts surrounding the technology when being implemented.

First, Fig. 1 breaks down the production outputs from companies not implementing certain techs by race:

Fig. 1

The racial breakdowns are very similar throughout each technology. Cloud-Based and Specialized technologies seem to be the major laggards, while the more niche Robotics leads the way. The one interesting change-up is that Specialized Software’s absence is benefitting African American businesses the most. More analysis on this could be insightful in the future.

Next, we look at the same breakdown but by sex in Fig 2:

Fig. 2

The Males dominate the production throughout each technology in a very proportional way, so we can’t draw much from sex. AI and Robotics seem to have a tandem relationship, where the absence of these technologies is dominated by males (meaning more traditional technologies).

A reason we hypothesized as to why this might be the case is based on firm size. Bigger firms might have the resources to employ more diverse employees, in terms of both demographics and skills. Therefore, things like Robotics would be used less in smaller firms. We compared what technologies weren’t being used in smaller vs larger firms in the following 2 graphs:

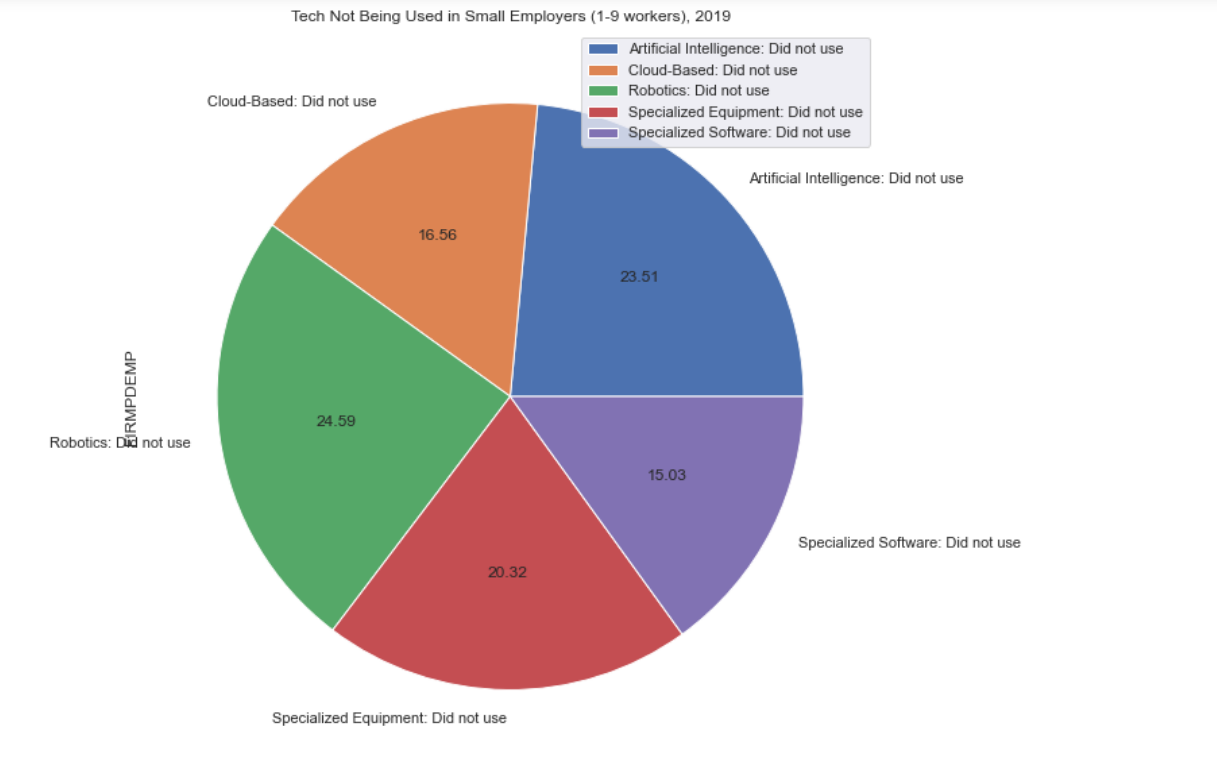
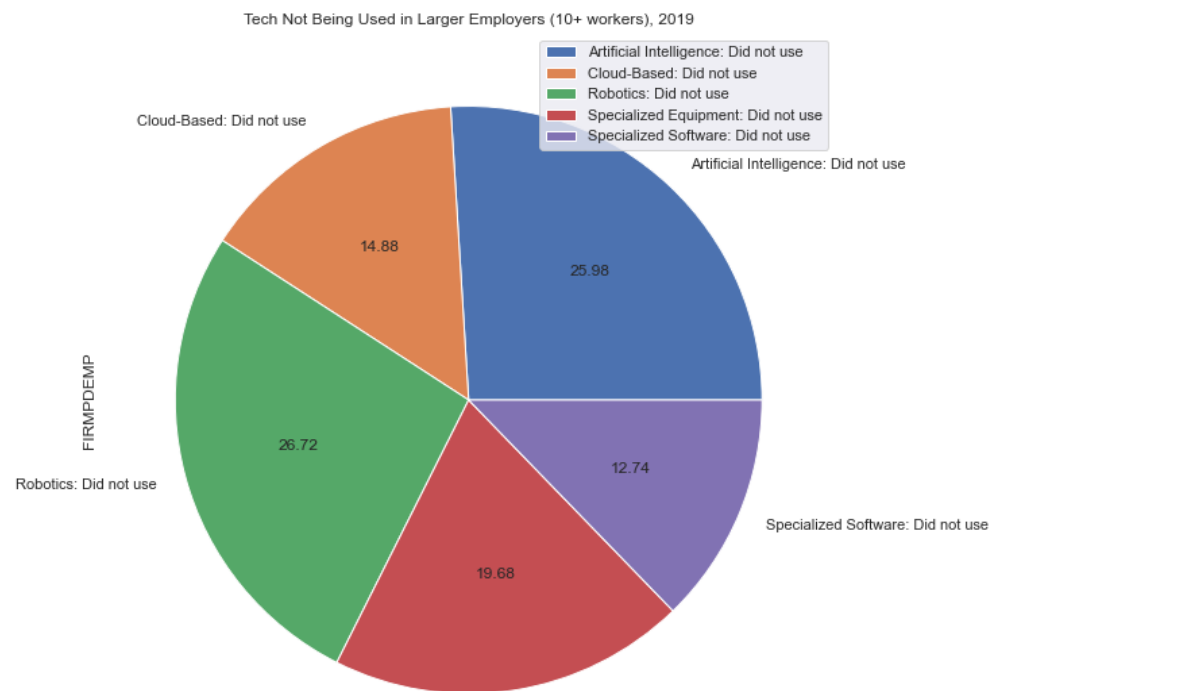
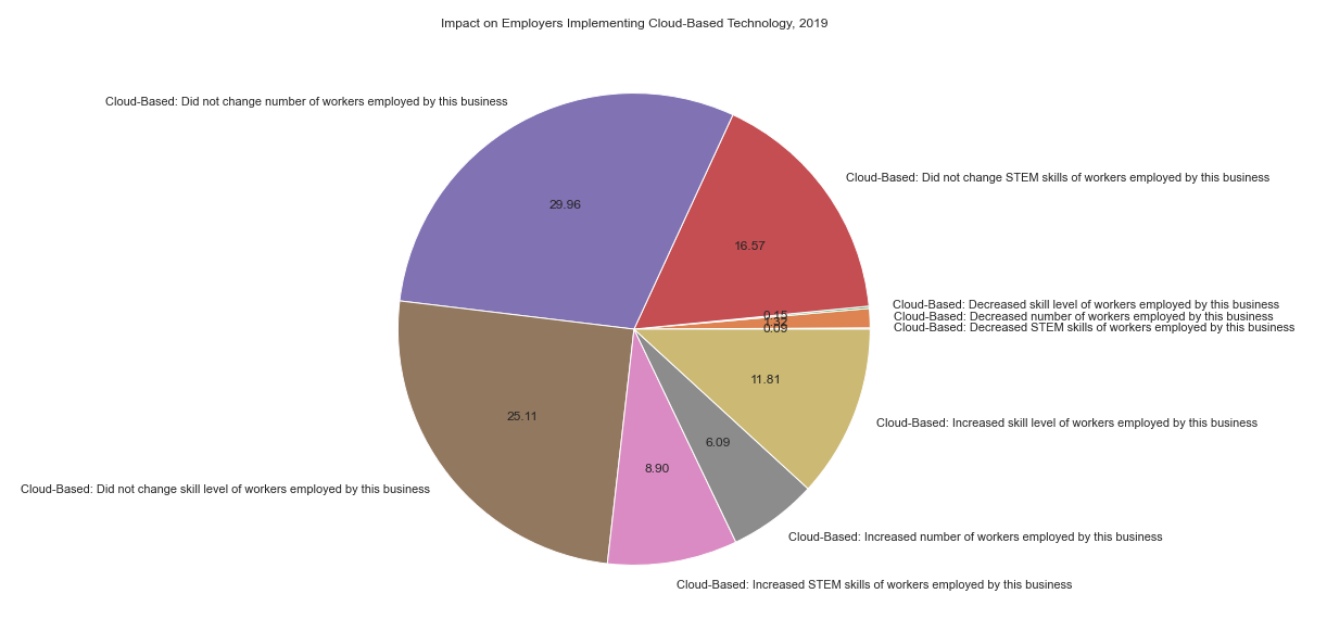


Fig. 3

Fig. 4

There are certainly some differences between small and larger businesses, but the change in percentages are marginal at best. However, it seems like we got our hypothesis backwards. The smaller companies seem to use the more niche technologies (AI and Robotics). These differences are only around 2%, with Robotics moving from 24% not used to 26%. A glaring issue we had with this data is on the small parameters of how the survey defined business sized. Many different companies have employees of over 10+ employees. For example, it would be quite silly to compare Apple to some local niche software company serving a small market. It is by no means a good way to split business size up.

For the last analysis, we wanted to pick a technology and break down some of the common impacts it has had on their employees and hiring practices. We picked cloud-based technologies as we have noticed that it is becoming almost essential in today’s world in one way or another. The next Figure looks at the proportions of impacts it has had on businesses:

Fig. 5

The vast majority of businesses haven’t had to upgrade to more employees, nor have they had to increase their current employee’s skill level. This might boil down to how each business defines ‘cloud-based’ technologies. Things like Office 365 utilize software that most office workers have had to master one way or another. However, it could even include more advanced technologies like DevOps tracking and Data Engineering; things that aren’t necessary for a lot of businesses.

Conclusion

While significant conclusions were able to be drawn based on the visualizations above, it is important to recognize that this can only be seen as a cursory glance at the state of businesses in the United States. In other words, these visualizations can only serve as a snapshot of the demographics of businesses. Further research is required to explain all disparities shown in this report as the US Census data does nothing to help determine the causation of any of these disparities. Research into, but not limited to, United States history and policy of business throughout its lifetime would be needed to better understand why the visualizations look the way they do. However, that is not to say these visualizations are not without merit altogether. An introductory look into the demographics of American businesses is valuable regardless of complete understanding, and this report has done a sufficient job in providing that.

1. <https://www.census.gov/data/developers/data-sets/abs.2019.html> [↑](#footnote-ref-510)