The Labor Force Through Time*

Trends in the workplace

Bo Phi

20 March 2022

Abstract

With a pandemic hopefully coming to a close in the near future, many people have started to return to the workforce. However, while many people are returning to in-person work, many newer jobs can be done remotely. Combined with the increase of jobs in the technology sector, the ways in which people work and the demographics of people who work are drastically different compared to times of the past. This statistical report uses data gathered by the University of Chicago from the General Social Survey (GSS) to (1) examine the ways in which working demographics have been recorded and changed through different decades, and (2), discuss the ways in which the COVID-19 pandemic could have implications on this dataset. With many states in the US removing vaccine and mask mandates, more and more people are returning to the workforce but as a result of the pandemic, many small businesses were forced to close. Although the true numbers are hard to attain, analyzing the past data can give us insight as to what the pandemic data will look like.

1 Introduction

Since the beginning of 2020, the world has been struck by a global pandemic. In the United States, it took a while for politicians to shut down the economy, and many employees got laid off as a result. Many small businesses were forced to close and lay off their employees causing many people to be out of jobs. The government offered assistance to many businesses in an effort to keep businesses afloat but despite those efforts, the economy still suffered. As a person who is prospectively looking into finding work opportunities during the tail end of the pandemic, it would be important for me to understand the trends and demographics of the working class. Many people were forced into retirement with many of the older working-class opting to stop work for health reasons. However, many people were also forced to work in hazardous conditions because they did not have any other options. Most working-class individuals do not have the ability to pick their own schedules and do not have the ability to take lots of time off because they do not own their own businesses. Looking at the previous data can allow us to make better inferences on what the pandemic data will look like when it is released.

Now being at what seems like the end of the pandemic, the restrictions are beginning to be lifted. In the United States, mask and vaccine mandates are beginning to be lifted and work offices are beginning to be repopulated with in-person workers. With the broader population preparing to go back to their normal lives, it is safe to assume that the working-class will look a lot different than it has in the past. Being able to have insight as to what the working-class statistics will look like compared to previous years will allow us to infer the implications of other events that could take place in the future to a similar data set. There are also issues with the current data set that we can analyze as to why it is not being a strong representation of what is happening in reality.

The importance of being able to predict data trends allows governments and people to plan for the future. Not many people could have predicted the COVID-19 pandemic and the implications it caused but now, being on the tail end of the pandemic, we can see how data has changed throughout the pandemic to better

*Code and data are available at: https://github.com/bophiwork/paper2work.

understand future events and how they would affect the workforce data. I will be using data from the General Social Survey (GSS) to analyze the ways in which the workforce demographics has changed over time and discuss the shortcomings of the data and how the dataset could be improved. The data will be processed and analyzed in R (R Core Team 2020) using the 'tidyverse' (Wickham et al. 2019) amd 'dplyr' (Wickham et al. 2021) packages. The packages 'knitr' (Xie 2021a), 'bookdown' (Xie 2020), and 'tinytex' (Xie 2021b) are also used to generate the final markdown report.

This report uses data collected annually from the General Social Survey obtained by the University of Chicago (Chicago 2022). The University of Chicago has collected this annual data for many years and is meant to gain general information about the American people. This data contains demographic, geographic, and family relation data from all survey responders. This data is supposed to be collected annually but there are some years where there is no data recorded so for some examples, I have chosen specifically to pick out notable years to analyze.

In this particular dataset, there are over 6000 different variables and pieces of metadata that can be associated with each survey participant. Many data variables were very obscure such as "college major" or "father's highest degree." While these points were interesting to look at, many of these more obscure pieces of data could not be observed for all the years I was interested in looking at. As a result, I chose to examine fairly easy to obtain pieces of data such as age and employment status among others to develop an understanding of the workforce demographics of the time. While this data can serve as a segment of a much larger population, I do find that the sample size of some years is rather small so it is important to note that this data could potentially not be truthfully representative of the broader American population.

2 Data

The next 4 plots are of the age of workers throughout different decades (Figure 1).

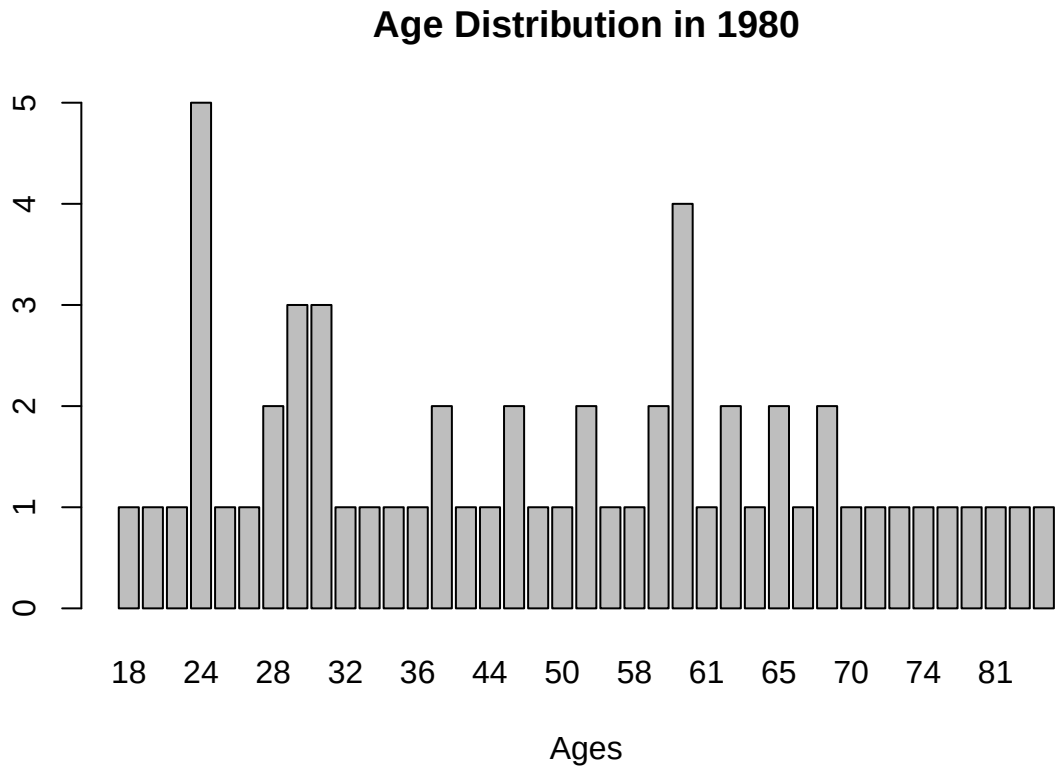

Figure 1: Age of Workers Through Different Decades

These plots shows us the age distribution amongst the working class during different decades. As seen in the data, we find that the trends in 1980 and 1990 are pretty similar. The distribution shows a skew towards a young work force with not nearly as many senior working citizens (ages 65+). In 2000 we find an interesting shift with there being marginally more working senior citizens. In 2010 we find that the skew of the distribution goes back towards favoring a younger workforce.

Figure 2: Age of Workers Through Different Decades

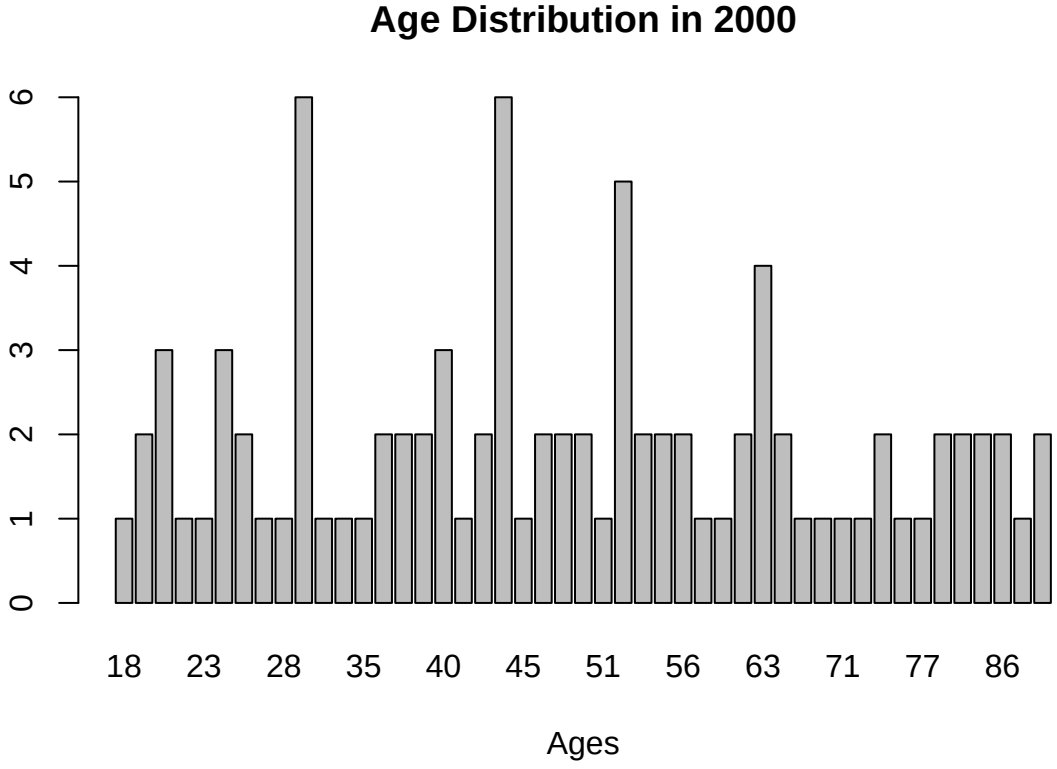

Figure 3: Age of Workers Through Different Decades

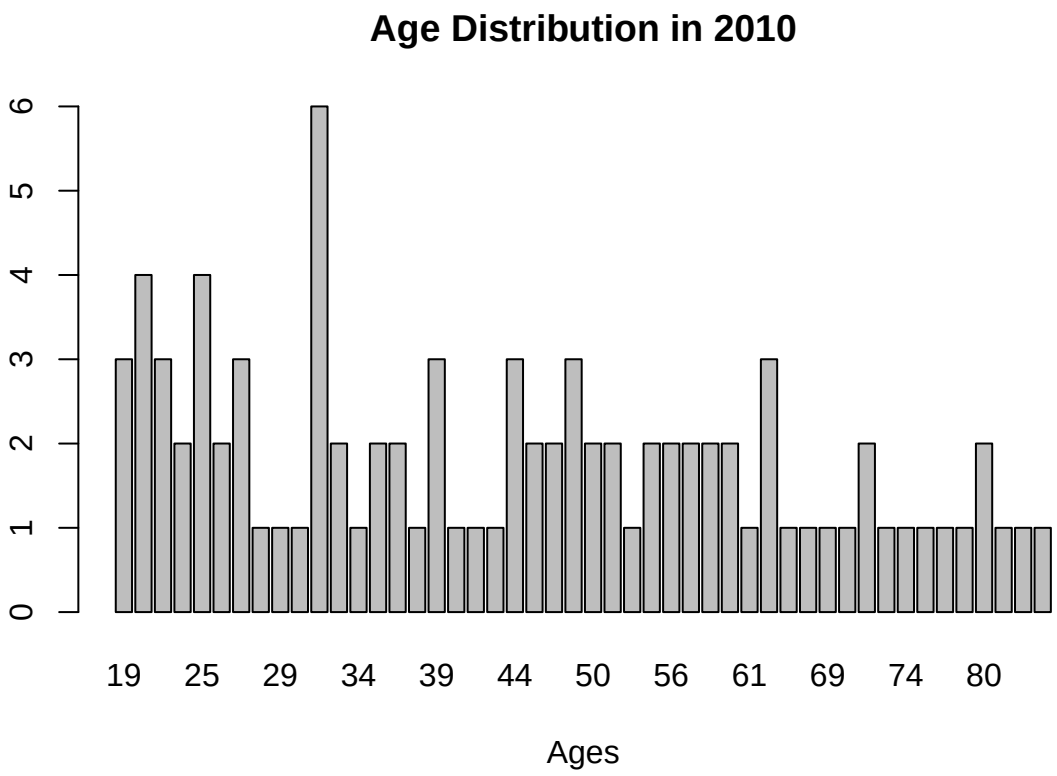

Figure 4: Age of Workers Through Different Decades

This plot shows us the work status of everyone who participated in the survey. We find that one of the biggest groups is those who are working full time. Part-time workers, retired people, and homekeepers all had similar amounts of representation as well. Notably one of the groups with the lowest amount of representation was people in school.

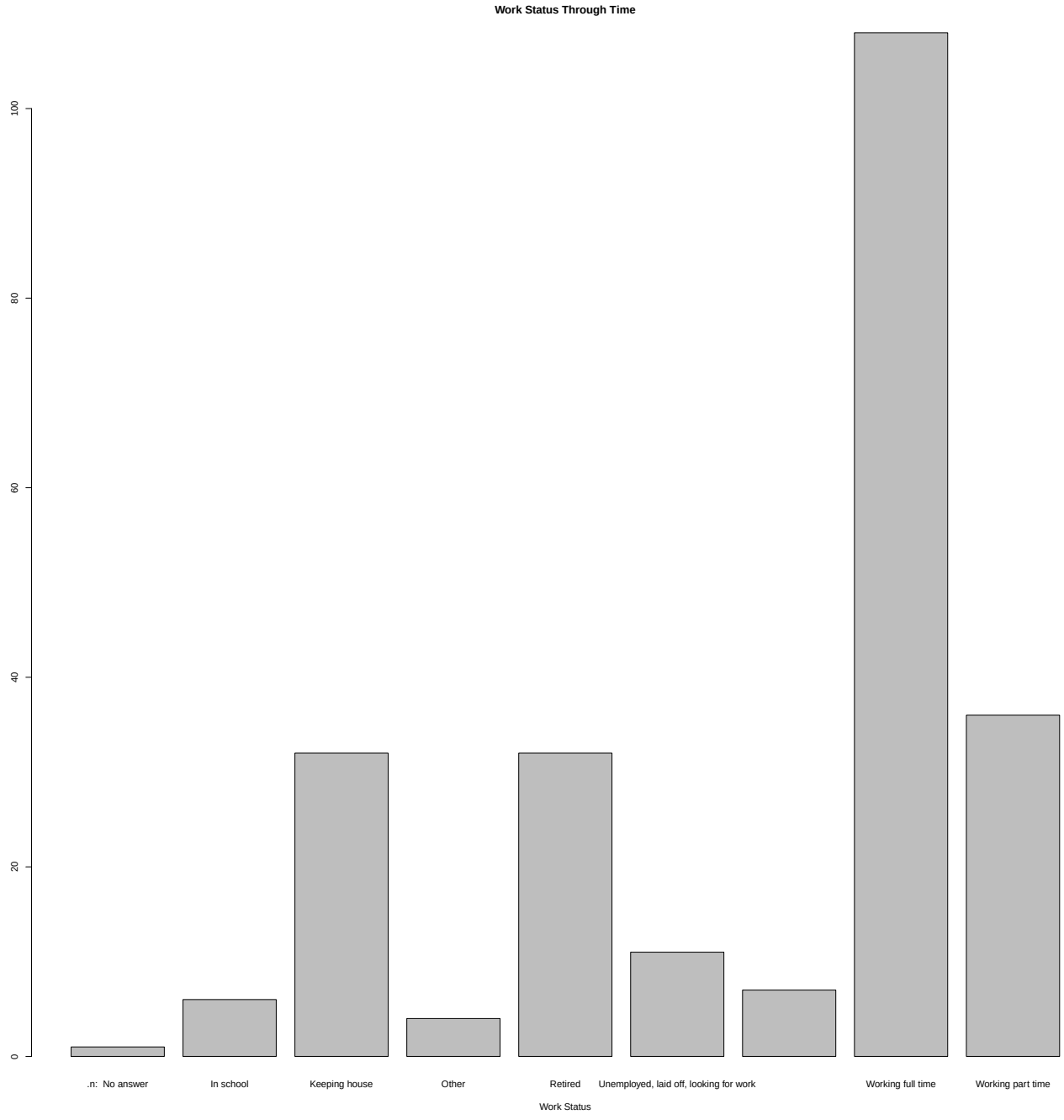

Figure 5: Comparison

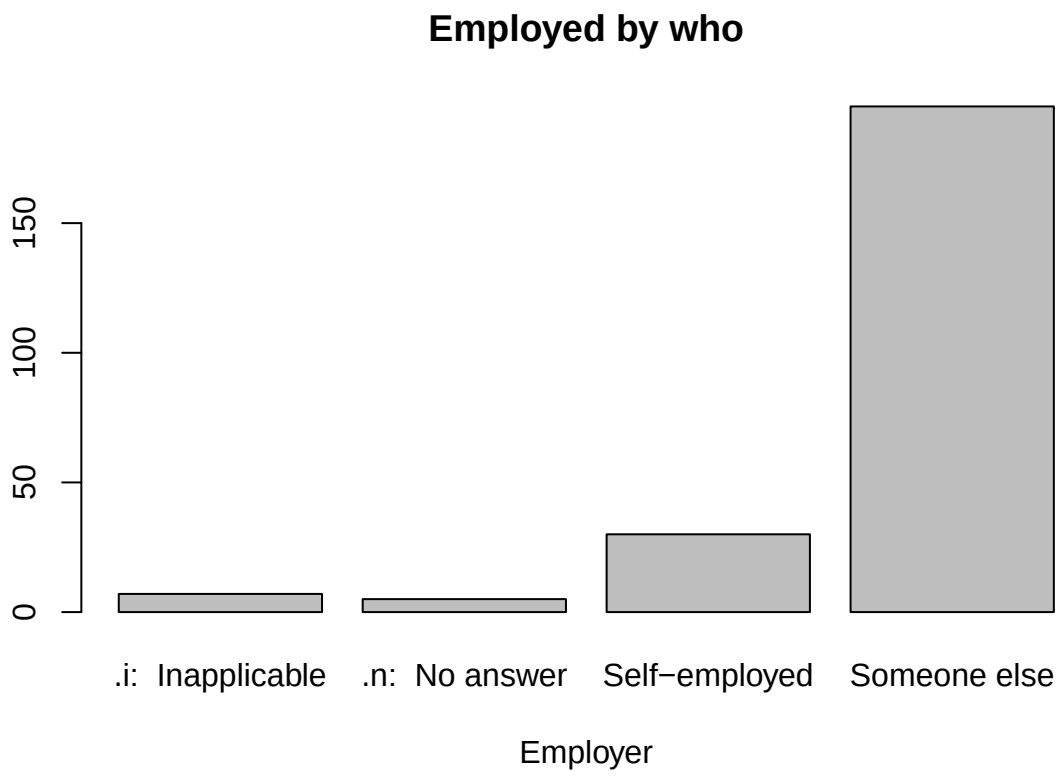

Figure 6: Comparison

This plot shows us the employed by status of everyone who participated in the survey. We find that the vast majority of people are employed by someone else and the minority of people are self-employed. Those without answers are likely not employed by anyone which would reflect the low numbers in the plot we examined earlier.

3 Results

All of the data serves to inform us on how the data will change as a result of the COVID-19 pandemic. I would predict that the age distribution will maintain its skew or even skew further towards the direction of young people because of senior working people being less likely to work because they are at higher risk of harm because of the pandemic. I will be discussing the various forms of shortcomings with the dataset particularly pertaining to its lack of accurate representation and potential biases.

4 Discussion

Looking at the data and what we were able to produce with it, it seems largely incomplete and does not do a good job of telling the whole story of American citizens. The first thing I would like to point out about this particular dataset is the sample size. Many of the years selected had less than 100 entries/participants. If the data collected is truly meant to be an accurate representation of the American population, the dataset should have reflected that. What this tells us is that this survey is not being taken by everyone in the country. During my research, I was under the assumption that this survey was released in a similar manner to the country's census. What this small sample size tells me is that the GSS has a different method of obtaining data. Upon further investigation, I found that the GSS survey asks for participants across the country that are meant to represent "alike" households. This means that the researchers of the GSS are conducting selective sampling in their studies and purposefully choosing to ask for participants that fit specific criteria. While this method of obtaining data is likely cheaper and easier for the University of Chicago to analyze, it leads to

many shortcomings. Even if researchers were able to get every American citizen to complete the survey, there would still be holes in the data with the possibility of people filling out the survey incorrectly or missing participants entirely. Using a sample population to represent a broader population is a common practice in statistics and data gathering but I still would have wished to see a larger sample size because it would have led to a more accurate representation of the American people and led to a more accurate conclusion.

4.1 Weaknesses and next steps

Another weakness of the dataset lies in the nature of the data collection. Participants were asked if they would like to participate and that can cause unintended biases in the trends we observe. Of course, the data will not account for those who were not asked but it will also not account for those who were asked but did not wish to participate. Because of the selective sampling, the asked participants were specifically chosen to represent a target demographic. If they do not wish to participate in the survey, however, then the entire segment of the population they are meant to represent would not be represented in the data we find. As it stands, this dataset is fairly weak. Not many real conclusions can be drawn from this dataset because of its inability to properly represent the American population. Moving forwards, it would have been more beneficial to have a larger sample size that would have led to a more accurate dataset we could actually draw meaningful conclusions from. Additionally, it would have been interesting to be able to see which state each participant was living in. That information would have allowed us to see if there was a relationship between different states and the working population and trends within those states. Additionally, it would have been nice to look at if any significant events in the world had any effects on the years I decided to examine. Similarly to how we are using this dataset to see how the pandemic data might be different when it is reported, it would be interesting to see if anything of not happened during the years we outlined to see if how it affected that year's data.

Appendix

A Additional details

References

Chicago, University of. 2022. *GSS Data Explorer*. https://gssdataexplorer.norc.org/home.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2020. *Bookdown: Authoring Books and Technical Documents with r Markdown*. https://github.com/rstudio/bookdown.

———. 2021a. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

———. 2021b. *Tinytex: Helper Functions to Install and Maintain TeX Live, and Compile LaTeX Documents*. https://github.com/yihui/tinytex.