

Why Birth Rates Keep falling in the US, even now.*

Reproduction of ‘The Puzzle of Falling US Birth Rates since the Great Recession’

Sampson Zhao

27 February 2022

Abstract

Through the past 40 years, we have seen a huge decrease in the Birth rates in the US, which was exasperated further within the past 10 years. We have seen birth rates at 65-70 per 1000 people in 1980s to less than 50 per 1000 people in 2020. These birth rates have looked at multiple different factors as reasons, such as the cost of living being too high for some families to bare, to effectively the lack of wanting a child. These factors play a large role in the continued development of the future population which has lead to large concerns towards the further continuation of this decreasing trend of birth rates and effectively the downstream effects of this continuing. This paper is a reproduction of the original paper ‘The Puzzle of Falling US Birth Rates since the Great Recession.’

Contents

1	Introduction	2
2	Data	2
2.1	Acknowledgement	2
2.2	Where this data comes from	3
2.3	Disclosure	3
2.4	Understanding the Data	3
3	Discussion	6
3.1	Problems with the provided data	6
3.2	Data bias of ‘The Puzzle of Falling US Birth Rates since the Great Recession’	7
4	Conclusion	8
	References	9

*Code and data are available at: <https://github.com/Sampsonzhao/bikesharetorontodec2021>. Reproduction under SSPR is available at: <https://doi.org/10.48152/ssrp-bqjp-n865>

1 Introduction

Throughout the past 40 years, we have seen a steady decrease in the number of births within the US, looking specifically at the number of births per 1000 people, we see that birth rates go from 65-70 births per 1000 people in 1980 to less than 60 births per 1000 people in 2008. This steady decrease has been propagated further in more recent years, as the average birth rate in 2020 was found to be closer to 55 births per 1000 people. This steady decrease can be seen as a possible reaction to many different social factors, such as the decrease in the “demand for children” and increase in the “monetary cost” of raising a child. These factors are not only just an overview of the possibilities that define why there is a decrease in the birth rates but are also a trend that aggregate due to the increase in social possibilities for both men and women. ‘The Puzzle of Falling US Birth Rates since the Great Recession’ (Kearney, Levine, and Pardue 2022) goes on to investigate the multi-factoral issues that lie with the decreasing birth rates in the US, and explores the impact some of them have towards the current situation.

Overall, this paper (Kearney, Levine, and Pardue 2022) goes into looking at the possible factors that affect how the change in time is related to the decrease in birth rates, more specifically making estimations to the possible reason for the decrease. The paper goes on to claim that there is no one reason that answered the question of “Why the Birth Rates in the US are decreasing since the Great Recession.” They undergo a thorough investigation over some of the possible reasons of the dip in birth rates during the 1980s till 2008, citing the huge impact of the Great Recession, but there was no specific link between policies made by the government, significant economic factors, and the declining birth rates. They then investigated probable differences between the US and other similarly developed countries and found that the decrease in birth rates were seen similar across the board in relatively high-income countries. They concluded that both time and location are not factors that significantly alter the rates of birth, through their previous investigations, and concluded that the societal consequences of the US, primarily the decrease in growth of the general productivity of the country, as well as the decrease in support of the programs provided to both elderly and new parents, play some role in the effects of the decrease in birth rates, even though the relation between these factors are not as clear cut.

This paper replicates ‘The Puzzle of Falling US Birth Rates since the Great Recession,’ while investigating some of the pros and cons of the paper. This analysis will discuss the granularity of the data that was used for the initial data set, and address how this data would transform from the raw data to the graphs and tables shown in the paper. The paper will then discuss some of the more pressing concerns of the paper, such as the use of proprietary software and misrepresenting the data as a whole.

2 Data

2.1 Acknowledgement

This reproduction uses the ‘R’ programming language (R Core Team 2020), and the ‘R-Studio’ Integrated development environment (RStudio Team 2020) as the primary source of data processing and the generation of the paper. The ‘Tidyverse’ package was used for manipulation of the datasets (Wickham et al. 2019). The ‘Here’ package was used to locate files produced from R scripts (Müller 2020). ‘KableExtra’ package was used to enhance the overall appearance of tables (Zhu 2021). ‘knitr’ was used to tie all of the information on this document into one digestible PDF (Xie 2021)(Xie 2015)(Xie 2014). ‘ggplot2’ was used to create the bar plots in this document (Wickham 2016). ‘reshape2’ was used to modify data sets to better create the wanted plots, while simplifying the code needed for those graphs (Wickham 2007).

2.2 Where this data comes from

Based on the original paper (Kearney, Levine, and Pardue 2022), they cited that the data originated from the National Center for Health Statistics' (NCHS), where they were able to collect 'restricted-use' data, and obtained the birth rate data for women from 15-44 years old. This data would be segregated as followed:

- Age
- Marital Status
- Ethnicity
- State where the birth record took place

Though not all information in recent years were available through the NCHS, other pieces of information would be retrieved through publically available reports, such as the Vital Statistics Final Births report and the 2020 Provisional Birth Report provided by the CDC (Martin et al. 2017)(Martin et al. 2021)(Hamilton, Martin, and Osterman 2021). From these specific data sets, they were able to add on to the information from the NCHS, more specifically, they were able to find data related to:

- Education Level of the Mother
- Nativity
- Order of Birth of the Child at the time of Birth

Though not specifically related to the reproduction of this paper, other sources were used to add flavour to the data sets, like level of student debt that the mother has, use of contraceptives, child care expenses and average rent costs. These factors heavily impact the birth rate, but to simplify the data set, these data sources were not used in this reproduction to focus on the change in population parameters, instead of lifestyle factors.

2.3 Disclosure

As a parameter to use the data properly and have a reference back to the data set used by Kearney, Levine, and Pardue 2022, the processed data that was provided by the authors were used, as the original source of data restricted access to certain parts of the data that was used initially. Hence this reproduction relies on some of the processed data in the public repository to generate some of the charts. While this goes against the nature of a reproduction/replication, it is due to the nature of the original data set, as well as the lack of information to properly decipher (as the original authors used proprietary software to clean and decipher the data) the raw data to the csv trimmed data, that this was not done.

2.4 Understanding the Data

As an overview of the original hypothesis of Kearney, Levine, and Pardue 2022, they understood that many different factors led to the decrease in number of birth per year in the US. This is seen in Figure 1, where we can see that the overall trend of the data points go lower and lower each year. This effect is exasperated after 2007, when there was the economic crash, deeming that year to be the Great Depression. Due to the events here, we can correlate that even though before 2008, the birth rate in the US was increasing year before and directly after the economic crises, as depicted by the dotted line in left graph of Figure 1, we see an immediate decrease in birth rates. We can assume that because of the economic crisis, families began to be less willing to raise a child, effectively linking the fact that the economy had a role in controlling the birth rate. We can also conclude that the economics of a family play a large role to the family deciding to have children, as money and the economic welfare of the family is highly correlated to the family being able to raise a child. As we can see in Figure 1 (right), we can see that by 2007, the cost of food products like flour, began to rise significantly. This is representative of the economy after the Economic crash in 2007, where a

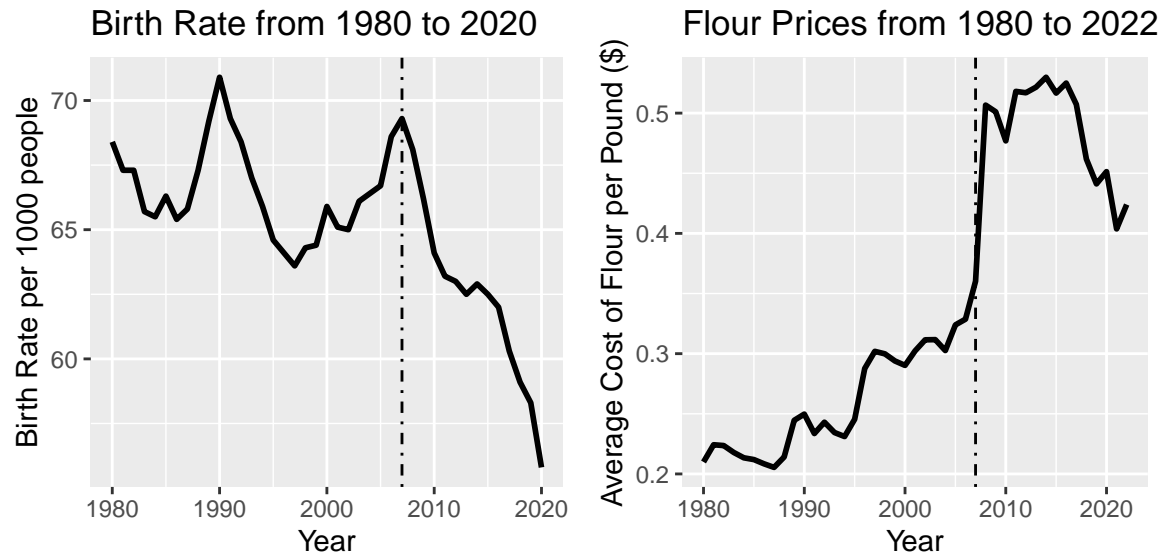


Figure 1: Overall Birth Rate in US (Left) and Cost of White Flour (Right) in the years 1980 - 2020

sharp increase in the cost of food items effectively represents the cost of living. More bluntly, if we compared the cost of living to the Birth Rates, we can see that as the cost of living increases, thus families are more hesitant in raising a child during those times, leading to a significant decrease in the birth rate. Though this is not the only factor that controls birth rate, this is a significant factor to many modern families, as many young couples choose to not raise a child, especially in recent years, as they are financially unable to feed, nurture and protect a child.

Now, if we look at Figure 2, we can create some granularities, and show the birth rates in the US, separated by the age of the mother. Here we can still see the overall trend that was shown in 1 (left), where the birth rates of all age ranges were increased, to some extent, until the stock market crash in 2007. But, here we can see a difference between the age ranges. Younger mothers, or to some extent, younger couples, were seen to have a significant decrease in the birth rates, which is representative of the times, where younger couples who have just entered the work force, were less willing to give birth to children, as they were financially unable to raise a child. This is seen in the age ranges from 15-29. But if we look at the slightly older generation, we can see that the birth rates in those populations did not decrease significantly, but instead saw marginal increases in birth rates. Though many factors can attribute to this increase, typically, we can assume that it is only due to these families being less affected by the economic crash in 2007, allowing them to have financial ability to keep up with the cost of living, as well as to raise children. Though this is not representative of the entire picture, it can be one possible explanation as to why we don't see a decrease in birth rates compared to younger couples.

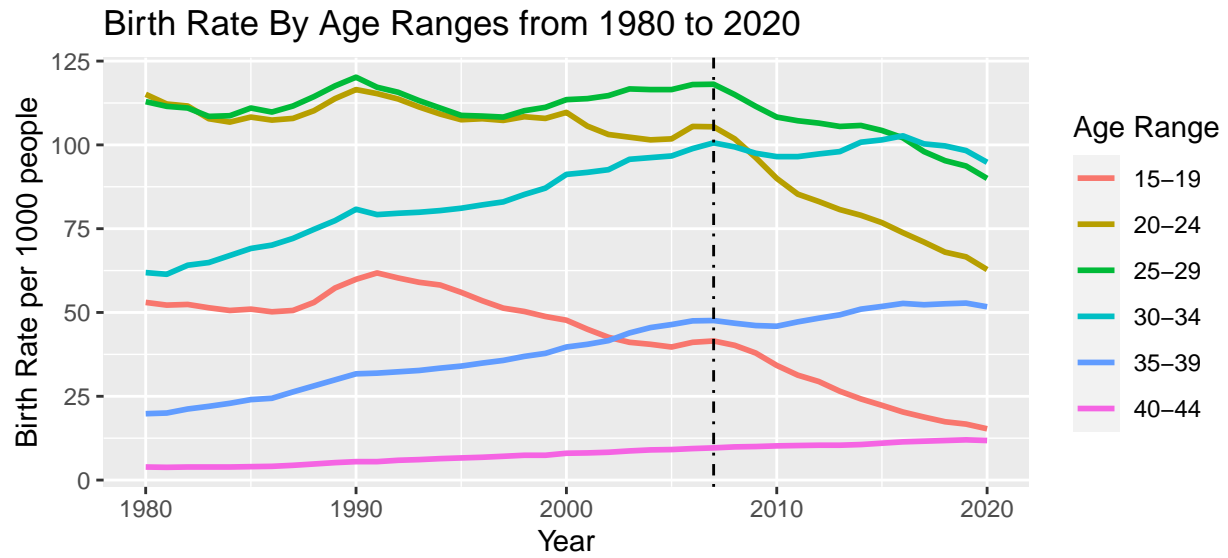


Figure 2: Birth Rates separated by Age Groups In the years 1980 - 2020

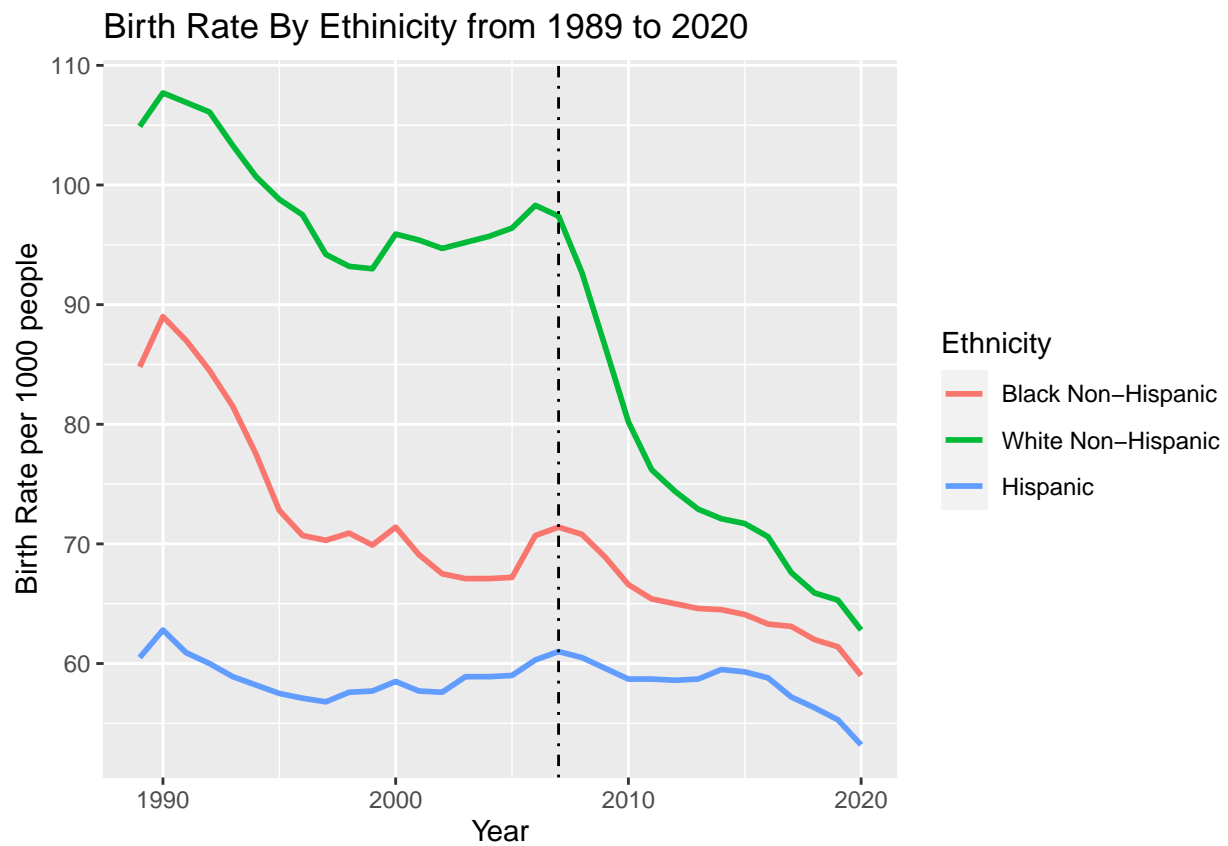


Figure 3: Birth Rates separated by race In the years 1989 - 2020

Table 1: Demographic Distribution of Decline in Birth Rate

Group	2007 Dis- tribution of population	2007 Birth rate	2019 Birth Rate	2007-2019 Change
Age 15–19, Hispanic	3.1%	82.19994	24.74746	-57.45247
Age 15–19, White Non-Hispanic	11.0%	27.10618	10.99167	-16.11451
Age 25–29, White NH, College Grad	4.2%	101.12422	65.10260	-36.02162
Age 15–19, Black NH	2.8%	65.11736	25.30435	-39.81300
Age 20–24, White NH, HS Grad	2.8%	139.92938	105.32291	-34.60648
Age 20–24, White NH, Some College	5.5%	54.15648	37.68721	-16.46927
Age 30–34, White NH, College Grad	4.2%	131.74390	112.04236	-19.70154
Age 20–24, Hispanic, Less than HS	0.8%	295.10907	206.81287	-88.29620
Total	34.3%	76.45805	50.52254	-25.93550

Finally, if we looking into separating the overall birth rates into the ethnicity of the mother, we can interpolate information that would be supportive of the hypothesis of how economic driven the birth rate is. Here we can see that looking at the greater picture of the ethnicity of the mother, we can understand that this is not as big of a factor on the decreasing birth rates, though it is still an impact. Looking at both figure 3 and table 1, we can see the impact of each ethnicity that the authors of the paper have deemed as the most representative of the US population. Though through figure 3, we see that although each ethnicity has seen significant decrease throughout the 40 years, if we compare the % change between the Hispanic population and the rest of the population, we can see that their relative decrease in Hispanic population tends to be less compared to the white and black non-Hispanic population. This may be representative of the population itself as well as the economic wealth of these families, as typically, Hispanic families tend to earn less on average compared to other ethnicities in the US (Connley 2020), hence this would support the previously mentioned hypothesis of economics playing a role in affecting these decisions. In the case of Hispanic families, because the average wage for Hispanic families are considerably lower for more then the past 40 years, we see that the average birth rate of Hispanic women typically will be lower then white or black families, ultimately explaining why they were impacted less during the economic crisis in 2007.

As for table 1, similar claims can be seen when comparing Hispanic and Non-Hispanic families, where we can see that from 2007 to 2019, though the delta change between 2007 and 2019 for Hispanic Birth rates seem significantly higher compared to non-Hispanic mothers, that is most likely due to the fact that they are part of a smaller population compared to Non-Hispanic families, hence the impact of the change in birth rates may be compacted by low population, leading to a significant difference in Birth rate per 1000 people. As this table only shows the age groups that have seen increased loss of birth rate (looking at Figure 2), we can also consider the age of these populations being a considerable factor on how different the Delta change is. We could assume factors like increased amount of education being a factor that drives down the birth rate at younger ages, as seen in all ethnicities at the age of 15-19.

3 Discussion

3.1 Problems with the provided data

The paper that was used for the replication (Kearney, Levine, and Pardue 2022) went into a thorough investigation to understand some of the more traceable factors that may have affected the decrease in birth rate, as they go step by step, looking into multiple different data sets, and were able to extract the relevant data through the use of the Stata software. But because of the lack of access to this software, the

supplementing code that was provided for the replication that was done through Social Science Reproduction Platform (SSPR) to generate the figures were undecipherable, leading to a blind walk through to decipher and generate the charts through R script. This may have generated some differences between the charts that were shown on the paper and the graphs and tables shown in this replication, leading to possible flaws in the reproduction of the graphs themselves.

This issue extended not only to the generation of the graphs here on this paper, but also the scoping and Assessment steps of the Accelerating Computational Reproducibility in the Social Sciences Guide (ACRe) created by the Berkeley Initiative for Transparency in the Social Sciences (Team 2021). Though this issue seemed to be isolated in the assessment part of the guide, this essentially extended further throughout the entire assessment, leading to a haphazard attempt of a replication through SSPR. The largest issue is the use of Stata as a baseline, as seen throughout the guide as well as throughout the SSPR tutorial of conducting a replication, which contradicts the overall point of accessibility and their mentioning of using open-source software so that anyone with some knowledge would be able to replicate their thought process and overall graph and table generation.

The paper itself also claims that all of the needed programs are available within the data set that was shared on openICPRS, as dictated through one of the README files they have included in their project data set, but this master program file was not found at all, as it was removed before the publication of the data set, as seen on the openICPRS link for the project (Link To Project Data Set). Note that only one version of the dataset was ever uploaded, hence this was not available for the overall replication and reproduction of this paper. I would assume that with the proper knowledge of how to use the Stata software, as well as open access to this software, would allow for a better understanding of the generation of the graphs, and ultimately being able to understand the workflow tree that would have been generated in the replication done through SSPR. Though this may not be a limitation to institutions who have free access to this software, as well as common knowledge of how to use the software, this is a limitation that exceeds beyond the goal of the ACRe, to allow for anyone to have access to being able to reproduce data that were processed by the original authors. This is also representative of the ACRe as many of the examples that were shared throughout the replication process were assuming that the author of the replication has thorough knowledge of the Stata software, as well as assuming that Stata would be the primary mode of data processing in the field. Though this software is popular, it is locked behind a pay wall, making it prohibitive to freelance workers and other people in this field. I hope that both the ACRe and the authors of this paper can reflect on this further in future papers and iterations of the replication tutorial, making sure that what they are doing is respective of their mission.

3.2 Data bias of ‘The Puzzle of Falling US Birth Rates since the Great Recession’

Though the original paper by Kearney, Levine, and Pardue 2022 highlights some of the main reasons towards factors that affect the general decrease in birth rates in the US, they have chosen to exclude some of the other ethnicities throughout the entire study. They selected data that would exclude ethnicities like East Asian, Indian, and other significant populations, especially in more recent years, where the proportion of these minority populations have increase in select states. Though they may have seen these minority populations to have less of an impact on the overall decrease, we cannot excluded these populations without going into specific reasons why these data sets would not be considered, assuming that this data is available at the time. Because these populations were excluded from this study, there could be an argument to make, where this subset of the US population may outweigh the decrease in birth rate, in which case, the entire premise of the hypothesis that Kearney, Levine, and Pardue 2022 made would be incorrect. This bias in data selection may prove to be a critical downfall towards their investigation on the decrease in birth rate, as if there was the consideration for other ethnicities, other then white, black and Hispanic, the impact of certain values that were shown in most of the figures of this reproduction as well as the table. For example, if we take the birth rate of Asian mothers, we can see that in 2019, it is slightly lower then the White population, but this was also separated by age group, we would see a much smaller population of Asian teens within the 15-19 age range. This would most likely be due to stringent education of Asian children, as well as overall differences

in culture, leading to the decrease in Asian youth birth.

4 Conclusion

Although the authors of the paper ‘The Puzzle of Falling US Birth Rates since the Great Recession’ demonstrated some issues within their investigation, the paper overall dives deep into investigating the reasons behind the decrease birth rates in the US, citing the economic condition of families as well as the passage of time as the main driver behind this. They looked into the importance of money when considering raising a child, especially after economic crises, where many families are deterred in raising children due to the increased cost of living, and the decrease flow of money. Many other factors play a role in how families may choose to not have a child, but overall, the decrease in birth rate from the 1980s till now, may be a concern towards future planning, as the decrease in younger populations may impact the older generations later on.

References

- Connley, Courtney. 2020. *CNBC*. CNBC. <https://www.cnbc.com/2020/10/29/latinas-face-an-ongoing-pay-gap-that-has-barely-moved-in-30-years.html>.
- Hamilton, Brady E., Joyce A. Martin, and Michelle J. K. Osterman. 2021. “Births : Provisional Data for 2020,” May. <https://stacks.cdc.gov/view/cdc/104993>.
- Kearney, Melissa S., Phillip B. Levine, and Luke Pardue. 2022. “The Puzzle of Falling US Birth Rates Since the Great Recession.” *Journal of Economic Perspectives* 36 (1): 151–76. <https://doi.org/10.1257/jep.36.1.151>.
- Martin, Joyce A., Brady E. Hamilton, Osterman. Michelle J. K., and Anne K. Driscoll. 2021. “Births: Final Data for 2019” 70 (2). <https://stacks.cdc.gov/view/cdc/100472>.
- Martin, Joyce A., Brady E. Hamilton, Michelle J. K. Osterman, Anne K. Driscoll, and T. J. Mathews. 2017. “National Vital Statistics Reports - Centers for Disease ...” *National Vital Statistics Reports, Volume 66, Number 1, January 5, 2017*, January. https://www.cdc.gov/nchs/data/nvsr/nvsr66/nvsr66_01.pdf.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- RStudio Team. 2020. *RStudio: Integrated Development Environment for r*. Boston, MA: RStudio, PBC. <http://www.rstudio.com/>.
- Team, ACRE. 2021. “Guide for Accelerating Computational Reproducibility in the Social Sciences: Guide for Accelerating Computational Reproducibility in the Social Sciences.” *Guide for Accelerating Computational Reproducibility in the Social Sciences | Guide for Accelerating Computational Reproducibility in the Social Sciences*. <https://bitss.github.io/ACRE/>.
- Wickham, Hadley. 2007. “Reshaping Data with the reshape Package.” *Journal of Statistical Software* 21 (12): 1–20. <http://www.jstatsoft.org/v21/i12/>.
- . 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Xie, Yihui. 2014. “Knitr: A Comprehensive Tool for Reproducible Research in R.” In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. <http://www.crcpress.com/product/isbn/9781466561595>.
- . 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.
- . 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.
- Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.