Analysis on the birth rate decline in the US[*]

: focusing on mother's age group, race, and ethnicity

Yewon(Chloe) Han

01 March 2022

Abstract

This paper is the replication of The Puzzle of Falling US Birth Rates Since the Great Recession (Melissa S. Kearney, Phillip B. Levine, and Luke Pardue, 2022). The US birth rate has drastically decreased since 2007, and this replication focuses on three variables expected to impact the birth rate decline in the US: age group of childbearing women, race and ethnicity. To be specific, the birth rate in teens and twenties age groups has dramatically shrank, and that of the hispanic demographic subgroup has plummeted as well. The latter illustrates the discrepancy of birth rates between the states where hispanic accounts for a significant amount of population and the others. This paper will further analyze the significant factors related to US birth rates decline and suggest next steps to improve the original paper.

1 Introduction

Birth rates are significant in that they are critical determinants of two socioeconomic factors: population growth and the age structure of the population(Cleland 2008). They also provide a measure in reviewing and rearranging the related policies. Therefore, it is important to track the birth rate changes over years and analyze the trend with a variety of factors affecting it.

In the US, there is a remarkable decline in birth rates as of the Great recession in 2007. The Puzzle of Falling US Birth Rates since the Great Recession(Melissa Schettini Kearney 2021) decomposes it in diverse aspects including ethnicity, age group of childbearing women, economic and education status of women, etc. Especially in terms of race and ethnicity variables, it shows how the decline aspects differ by each state plotting the birth rate change dataset on the US map.

---

[*]Code and data are available at: https://github.com/OnedayOneyeah/Analysis-on-the-birth-rate-decline-in-the-US

This paper is a replication of the report, but it is differentiated from the original in that it focuses on three specific factors: age group of childbearing women, race and ethnicity. Moreover, it critically analyzes the weakness of the datasets and gives a few insights to improve them. It also demonstrates whether the figures and tables in the report are reproducible with provided datasets in the package.

Specifically, the birth rates in teens and twenties age groups have dramatically shrank while the counterpart of over thirties has increased overall. In terms of race and ethnicity, the birth rates of all the demographic groups including white, black and hispanic have decreased, but this tendency was the most dominant in hispanic group. It describes the reason why there is a discrepancy in birth rates between the states which have a high percentage of hispanic population and those which have not in comparison.

Although there is a limitation that this paper only focuses on specific factors, this replication still has its meaning in that it includes further analyses of the significant factors related to US birth rates decline. Furthermore, it suggests the next steps to improve the original paper especially regarding the comprehensiveness of the datasets and the reproducibility of the paper.

2 Data

The original paper basically uses the birth data as its primary dataset. It is collected and aggregated from the National Center for Health Statistics(NCHS) restricted-use microdata(2022d). The paper also uses Vital Statistics Final Births reports for 2015, 2019(NCHS 2022), and the 2020 Provisional Birth Report(Brady E. Hamilton 2021) to disaggregate the childbearing women group whose ages are between 15 to 44 and further decompose the groups by age, Hispanic origin, marital status, and state. To disaggregate the childbearing women group by other variables such as mother's education level, nativity, birth order, and cohort, public birth microdata across 1980-1989 from the NBER Natality Database and NCHS microdata from 1990-2019(2015, 2019, 2021) were manipulated.

Another data the original paper deals with is the population data. In particular, the CDC SEER database(2022b) was used, and the paper covers the single-age population counts among all races from 1969-2019 and by race, and Hispanic origin from 1990-2019.

Lastly it also includes other datasets related to various factors. They are long-term changes, female-male median wage ratio, LARC(long-acting reversible contraceptives) usage, student loan debt per capita, average gross rent, child care expenses, and religious importance data.

This replication paper, however, focuses on the general birth rate trend and three main variables which significantly impacted the trend: age group, race, and ethnicity. Hence, only the birth rate and population data are used in this paper.

The raw and cleaned datasets are both provided in a project package on OPENICPSR(2022c). Since the datasets are all Stata Data File, generally manipulated in STATA environment(*Stata Statistical Software* 2021), a general-purpose statistical software package for data manipulation, visualization, and automated reporting, the files cannot be opened in

the R environment. However, the project package also includes some of the csv files, a more general format which can be manipulated with R. The list of datasets that this package offers are listed in Table 1.

File Name of dataset (.csv)	File Path
fig_1	144981-V1/data
fig_2e	144981-V1/data
fig_5	144981-V1/data
fig_6	144981-V1/data
figs_1_2_3	144981-V1/data
igs_2a_2b	144981-V1/data
numberbirths_2001_2019	144981-V1/data/annual_policy
births_educ_race_age6	144981-V1/data/decomp
numbirths_educ_race_age6	144981-V1/data/decomp
numbirths_educ_2044	144981-V1/data/educ
numbirths_hisp_nativity_mexican	144981-V1/data/hisp

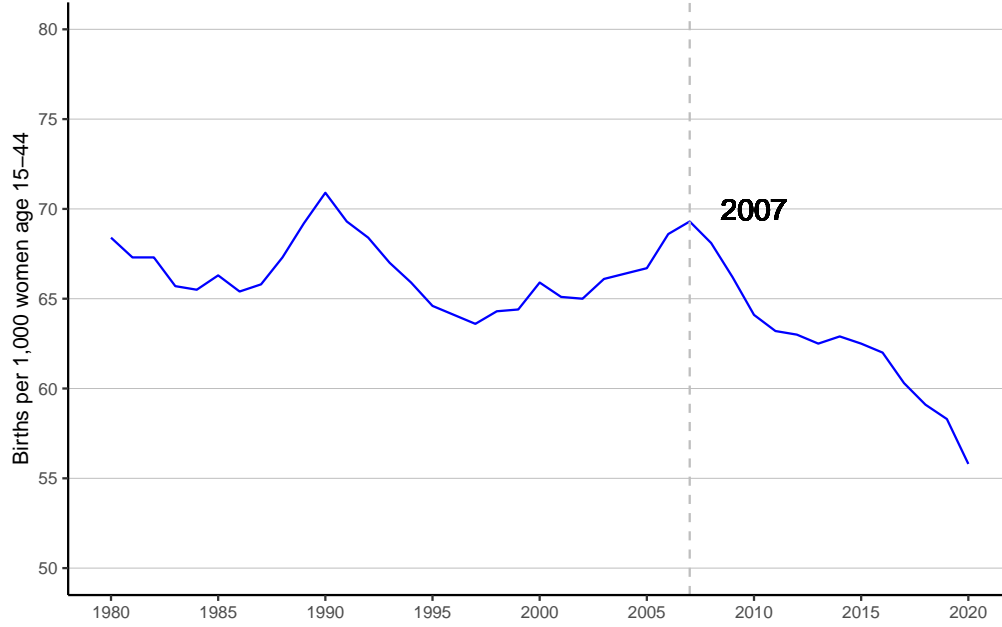Table 1: The List of datasets with File Path

Among these datasets, this replication paper uses the 'fig_1.csv' file for tracking the annual trend in US birth rates from 1980 to 2020. Also, 'figs_2a_2b.csv' file is used to see the trends in birth rates by five-year age group, race and ethnicity. Lastly the 'fig_7.csv' is manipulated for mapping the change in birth rates by state between two periods 2004-2008 to 2015-2019.

Besides the datasets from the project package, this replication paper also includes those from other sources for further analysis. Most of all, World population review dataset(2022a) is used to figure out if the states where the hispanic race group accounts for a large portion of the US population is. This dataset was downloaded in csv format from the World Population Review website and was added to the R Project folder. It was further analyzed to demonstrate whether the portion of hispanic population of each state affects the birth rate change between two different periods. With the dataset, it is possible to calculate the total rate of hispanic population of each state, combining all race and ethnicity groups including white hispanic, black hispanic, Asian hispanic, Hawaiian hispanic, Indian hispanic and other hispanic. It was manipulated as a table with the ratio of total hispanic population out of total population by state.

Moreover, we refered Kerney and Levine's article published in Brookings journal(Phillip Levine 2021) to show the number of children ever born changes depending on the mother's birth cohort. Also a dataset from Pew Research Center(Livingston 2019) is used to verify whether the birth rate of the hispanic group is the highest among foreign born as well as among US born as of 2020, which supports that the hispanic race group strongly affects the total US birth rate. Lastly, the datasets from United States Census Bureau(2021) and Iman Ghosh(Ghosh 2020) are used to calculate the portion of Asian population by each state.

3   Results

Figure 1: Trend in US Birth Rates

| Period	Decrease Rate (%)
1990-1997	-10.29619
2007-2020	-19.48052

Table 2: Changes in Births per 1,000 women age 15-44 from 1990-1997 to 2007-2020

As it is projected on Figure 1, the birth rate of childbearing women has dropped remarkably as of 2007 albeit it has been fluctuating over time before then. This tendency becomes much clearer when we compare two periods: 1990-1997 and 2007-2020. Apparently there seemed to be a huge decline in the birth rate between 1990-1997 as well. However, the decrease rate of the second period is almost twice lower than that of 1990-1997(Table 2). This result shows the decline in the birth rate from 2007 is an unprecedented phenomenon which should be further analyzed in a variety of aspects to discover attributable factors.

The first significant factor is related to age groups of childbearing women. While the birth rates of 'Age 15-19,' 'Age 20-24,' and 'Age 25-29' have drastically decreased since 2007, those of the other older age groups tended to increase overall besides 'Age 30-34' subgroup, which slightly increased from 2007 to 2015, but ended up decline from then to 2020.

Figure 2: Trends in Birth Rates for Five-year Age Group
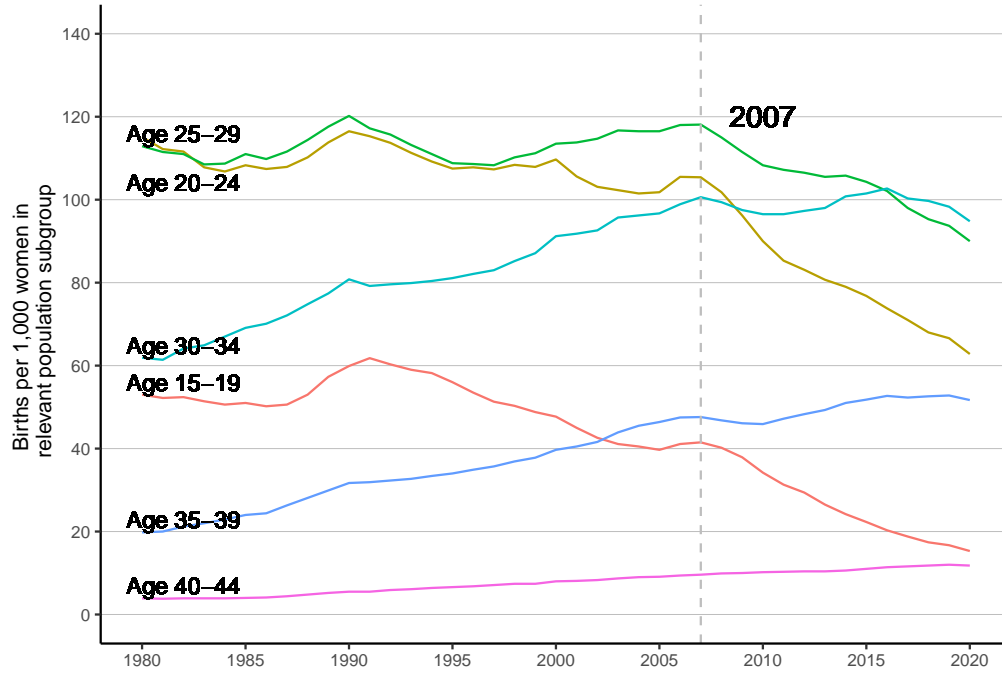
| Age Group	Change Rate (%)
Age 1519	-63.132530
Age 2024	-40.417457
Age 2529	-23.793395
Age 3034	-5.765408
Age 3539	8.613445
Age 4044	22.916667

Table 3: Changes in Births per 1,000 women age 15-44 from 2007 to 2020 by age group

Specifically, the rate changes in proportion to the age. As it is described in Table 3, the birth rates have drastically declined in two youngest age groups and especially in 'Age 15-19,' the birth rate decreased more than 60% around in 10 years. On the other hand, the birth rate in older age groups tended to increase. For instance, the birth rate in the 'Age 40-44' age group increased more than 20% in a given period.

This result supports the notion that the maternal age is getting older, which means the childbearing women are delaying their childbearing. However, Kearney and Levine(Phillip Levine 2021) suggests it is not merely reduced to it. Their research on children ever born by mother's birth cohort projects how the number of children ever born changes depending on mother's birth cohort. As the mother's birth cohort is younger, the number of children ever born at the same age is lower. This means more and more women end up having fewer children over generations. Therefore, it is crucial to figure out what makes women delay their childbearing age, and have fewer children. The original paper suggests multiple possi-

ble variables such as enhanced women's socioeconomic status, popularization of long-acting reversible contraceptives (LARC) usage, student loan debt, etc. Besides these quantitative factors, it seems it should also include qualitative research which might help to portray the concrete context of the decision of childbearing women.

Figure 3: Trends in Birth Rates by Race and Ethnicity (Ages 15-44)
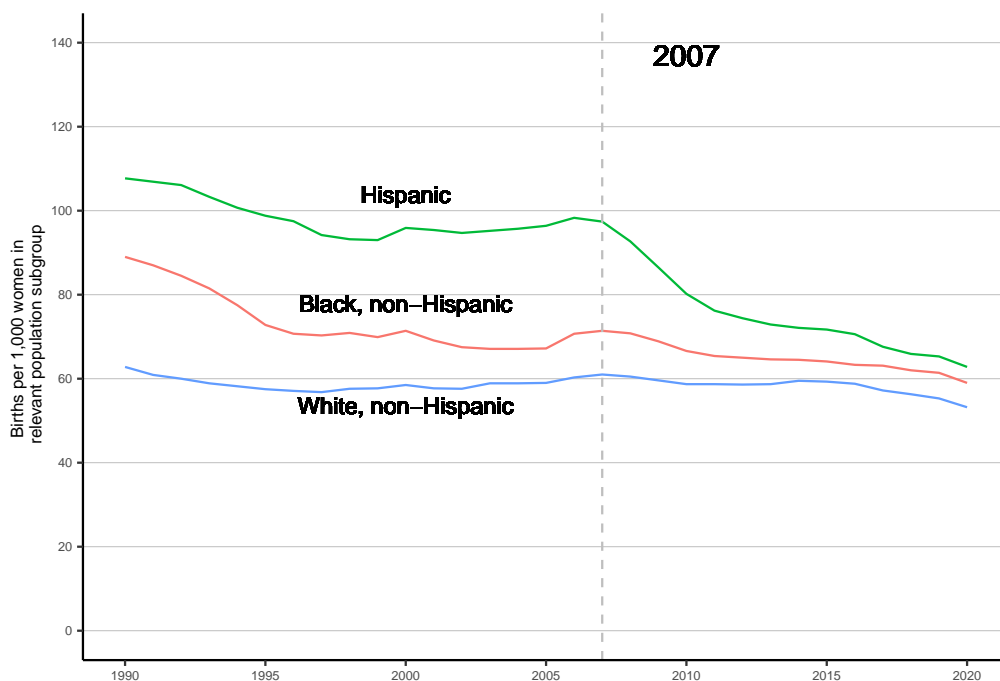
Another variable affecting the US birth rate is race and ethnicity. Figure 3 displays how the birth rate differs depending on three different race groups: hispanic, black non-hispanic, and white non-hispanic. As of 2007, the birth rate started to decrease gradually in all race groups, and this tendency was particularly dominant in the hispanic group. In 2007, the birth rate of the hispanic group was 97.4. In 2020, it ended up plummeting to 62.8. Considering the birth rate of the hispanic group is the highest among foreign born as well as among US born as of 2020(Livingston 2019), it is possible to say that the hispanic group strongly affects the total US birth rate as well.

To prove the decline in US birth rate is not merely due to the decrease of the number of hispanic women, the absolute population of the hispanic childbearing women should be reviewed. According to the calculations from the American Community Survey, the share of Hispanic women of childbearing age who are native born has increased from 49.3 percent in 2007 to 61.7 percent in 2018. Therefore, it is hard to say that the decline is mainly attributed to the decrease of hispanic women population. It is more convincing to say that hispanic women started to have fewer children from 2007.

Figure 4: Change in Birth Rates by State, 2004-2008 to 2015-2019

Change in Birth Rate

> 0
0 to –5
–5 to –10
< –10

Source: Birth data are from NCHS Vital Statistics.
Population data are from CDC sruveillance, Epidemiology, and End Results (SEER) program.
Note: Birth rates are calculated among women aged 15 to 44

State Name	Hispanic Percentage (%)
Puerto Rico	98.73
New Mexico	48.78
Texas	39.33
California	39.03
Arizona	31.32

Table 4: The percentage of hispanic population by each state

This tendency can be plotted in a geological dimension as well. Figure 44 portrays the change in birth rates by state between two periods 2004-2008 to 2015-2019. The changes in birth rate in Southwest states were more dramatic compared to the others, reaching down to -10% difference. According to World population review(2022a), Puerto Rico, New Mexico, Texas, California, and Arizona are the top 5 states(Table 4)where the Hispanic demographic accounts for a sizable portion of the population, which are colored red. This consistency makes it possible to conclude the birth rate decline in the hispanic group impacted on the changes by each state.

4   Discussion

4.1   Comprehensiveness of the race and ethnicity birth rate dataset

One of the fatal weaknesses of the original paper is that they are using the dataset of US birth rate by race and ethnicity which basically doesn't include Asian population. It only consists of white, black, and hispanic.

However, according to the population census completed on April 1, 2020 by the United States Census Bureau(2021), Asian population reaches 5.9% of the whole US total population, which is not neglectable. In particular, Asians account for 15% of California population, and 9% of Washington D.C as of 2020(Ghosh 2020). Furthermore, it is forecasted to account for 8.9% of the US total population in 2060. Therefore, encompassing more various demographic groups including Asian will be the next step for enhancing comprehensiveness in terms of race and ethnicity.

4.2   Weakness of reproducibility

The other weakness of this paper is in reproductivility. To replicate the Figure 4, we manipulated two provided csv datasets: 'numbirths_2001_2009.csv' and 'fig_7.csv'(Table 1). The first one consists of three columns: year, state name, and the number of births among 15 to 44 women. We filtered the only years that we need to replicate the figure: 2004-2008, 2015-2019. Then, we calculated the average number of births of each period and made a new dataset of the changes between two periods. The problem occurred when we plotted the dataset on map; some states were mis-colored. Hence, we decided to use another dataset file 'fig_7.csv,' instead of 'numbirths_2001_2009.csv.' The dataset provides the calculated change rates between two periods. However, the dataset plot was still mis-colored albeit the discrepancy was reduced compared to the previous dataset.

Figure 5.1. Generated with 'numbirths_2001_2009.csv'
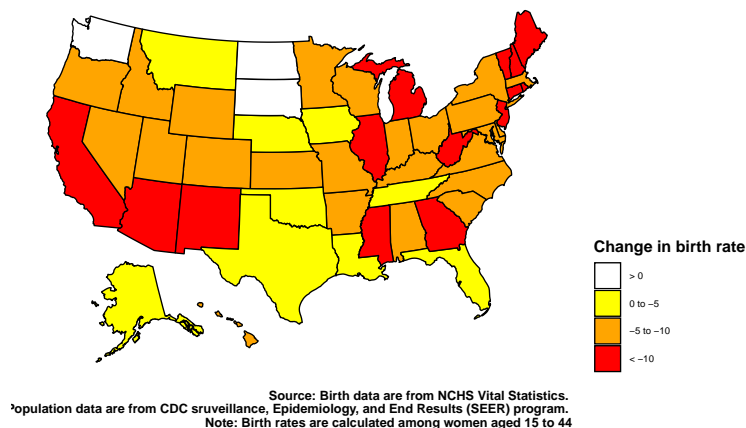**Change in Birth Rates by State, 2004–2008, 2015–2019**

Source: Birth data are from NCHS Vital Statistics.
Population data are from CDC sruveillance, Epidemiology, and End Results (SEER) program.
Note: Birth rates are calculated among women aged 15 to 44

Figure 5.2.: Generated with 'fig_6.csv'
Change in Birth Rates by State, 2004–2008 to 2015–2019

Change in birth rate
> 0
0 to –5
–5 to –10
< –10

Source: Birth data are from NCHS Vital Statistics.
Population data are from CDC sruveillance, Epidemiology, and End Results (SEER) program.
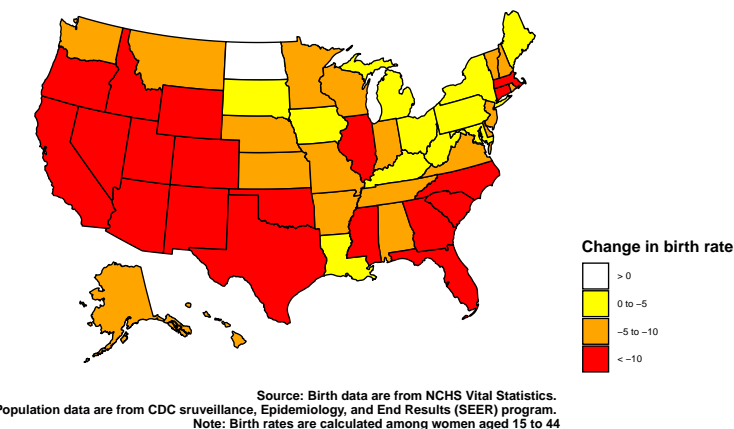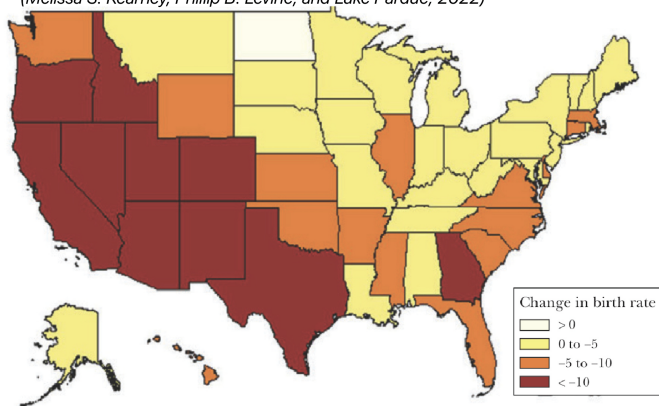**Note: Birth rates are calculated among women aged 15 to 44**

Figure 5.3.: Original Figure in The Puzzle of Falling US Birth Rates Since the Great Recession
(Melissa S. Kearney, Phillip B. Levine, and Luke Pardue, 2022)

Change in birth rate
> 0
0 to –5
–5 to –10
< –10

Source: Birth data are from NCHS Vital Statistics. Population data are from CDC Surveillance, Epidemiology, and End Results (SEER) program.
Note: Birth rates are calculated among women aged 15 to 44.

In conclusion, it was impossible to successfully reproduce the same figure as in the original paper. The birth rate changes calculated with the raw data 'numbirths_2001_2009.csv' were not consistent with the provided dataset 'fig_7.csv, and moreover the 'fig_7.csv' turned out impossible to be plotted in the same way with that in the original paper. Inconsistency between two datasets, and the discrepancy between the datasets with the figure are fatal in terms of reproducibility of a paper.

4.3   Conclusion

This paper analyzed why the US birth rates have dramatically shrank since the Great Recession in 2007, focusing on three variables: age group of childbearing women, race, and ethnicity. All the datasets and their sources used in this paper were listed mainly birth rate data, and population data. The replications of tables and figures were suggested as well, with their weaknesses and next steps for further study.

In this replication paper, we demonstrated the US birth rate has drastically decreased since

2007, and especially the decline was remarkable in younger childbearing women age group among teens and twenties. It means more and more women end up having fewer children as well as childbearing women are delaying their childbearing over generation. The decline of US birth rate can be also decomposed in race and ethnicity dimension. The decrease rate was the highest in hispanic race group, and the US states where hispanic population accounts for a sizable portion of the whole showed more prominent decrease rate, particularly compared to the counterparts.

It is meaningful that this paper scrutinizes a variety of factors affecting the decline in US birth rate, and suggest next steps for further study. To further improve the limitation of the original paper in terms of comprehensiveness and reproducibility, it is critical to include more various demographic groups into datasets, and to provide more consistent datasets which can be reproduced in every working environment.

References

2015, 2019, 2021. https://www.nber.org/research/data/vital-statistics-natality-birth-data.

2021. https://www.census.gov/quickfacts/fact/table/US/PST045221.

2022a. https://worldpopulationreview.com/states/states-by-race.

2022b. https://seer.cancer.gov/popdata/download.html.

2022c. https://www.openicpsr.org/openicpsr/project/144981/version/V1/view.

2022d. https://www.cdc.gov/nchs/nvss/nvss-restricted-data.htm.

Brady E. Hamilton, Joyee A. Martin & Michelle J. K. Osterman. 2021. "Vital Statistics Rapid Release." *NVSS*, no. 012 (May).

Cleland, J. G. 2008. "Trends in Human Fertility." *International Encyclopedia of Public Health*, 364–71.

Ghosh, Iman. 2020. "Visualizing the u.s. Population." *Visual Capitalist*, December. https://www.visualcapitalist.com/visualizing-u-s-population-by-race/.

Livingston, Gretchen. 2019. "Hispanic Women No Longer Account for the Majority of Immigrant Births in the u.s." August. https://www.pewresearch.org/fact-tank/2019/08/08/hispanic-women-no-longer-account-for-the-majority-of-immigrant-births-in-the-u-s/.

Melissa Schettini Kearney, Phillip B. Levine & Luke W. Pardue. 2021. "The Puzzle of Falling US Birth Rates Since the Great Recession." *National Bureau of Economic Research*, no. 29286.

NCHS. 2022. "Birth Data." https://www.cdc.gov/nchs/nvss/births.htm.

Phillip Levine, Melissa S. Kearney &. 2021. "Will Births in the US Rebound? Probably Not." *Brookings*, March.

Stata Statistical Software. 2021. 17th ed. College Station, TX: StataCorp LLC.