

# The Puzzle of Falling Birthrates in USA

Venkatesh Kannan      Email: kannan.venkatesh@stud.hs-fresenius.de  
ID: 400370901

## 1 1 Abstract

The purpose of this project is to replicate and analyse the trends in the birth rates of the United States as represented in a recent report about demographic and societal shifts in fertility patterns. The objective will be to recreate the figures and analyses showing the decline in U.S. birth rates from 2001 to 2019 by age group, demographic characteristics, educational attainment, marital status, birth order, and state-level variations. Using R and high-level data visualization, we move closer to the exact replica of the report's visualizations and findings.

## 2 2 Data Retrieval and Reading

The data for this study was extracted from the publicly available datasets made available by the article "The Puzzle of Falling U.S. Birth Rates Since the Great Recession" \* by Kearney et al., published in the Journal of Economic Perspectives, and accessible at <https://www.aeaweb.org/articles?id=10.1257/jep.36.1.151> These include the US birth rates disaggregated by age groups, demographic characteristics, education levels, marital statuses, and birth orders, along with state variations.

The analysis was done in RStudio, where the processing were done using tools from the packages `dplyr` and `tidyr`. Key trends and patterns were visualized using the `ggplot2` package; in many cases, datasets had to be reshaped into long formats to be able to plot them effectively. Also, customizations to legends and visual aesthetics have been done to adapt the plots to the original report format. Library `usmap` provided an easy way to map state-to-state variation in birth rates, while R Markdown was used for integrating all visualizations and results into a reproducible report format.

## 3 3 Graphical Representation

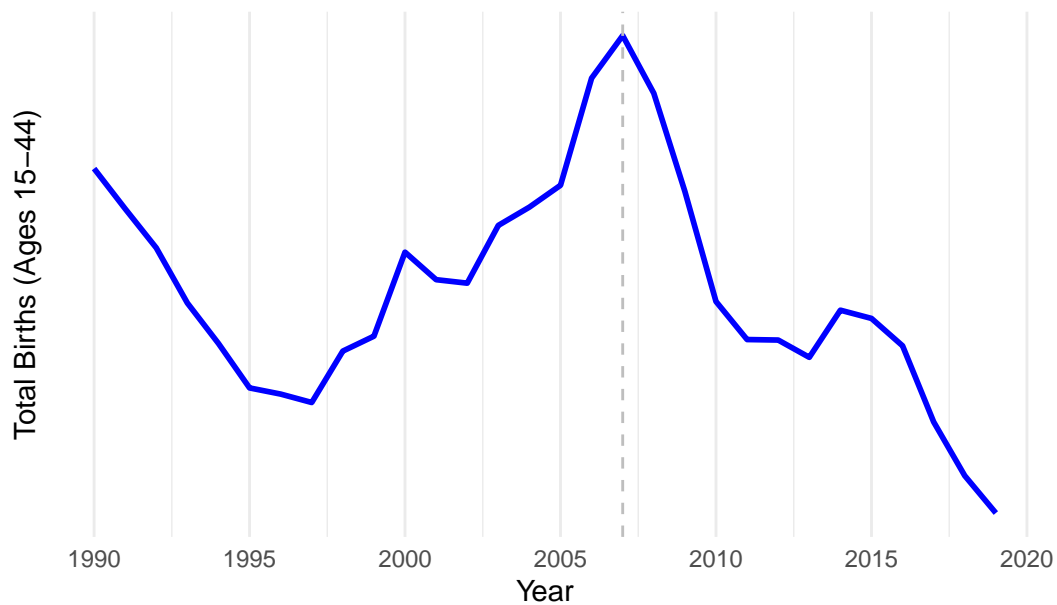
```
library(ggplot2)
library(dplyr)
library(sf)           # For spatial data
library(tidyr)        # For data manipulation (pivot_wider)
library(haven)
```

### 3.1 3.1 Figure 1: Trend in US Birth Rates (2001-2019)

```
# Load the dataset
birth_data <- read.csv("numbirths_2001_2019.csv")
#data manipulation
birth_trend <- birth_data %>%
  group_by(year) %>%
  summarise(total_births = sum(numbirth1544))%>%
  mutate(birth_rate = total_births / 1000) # Assuming birth rate per 1,000 women

# Plot the trend in US birth rates
ggplot(birth_trend, aes(x = year, y = total_births)) +
  geom_line(color = "blue", size = 1) +
  geom_vline(xintercept = 2007, linetype = "dashed", color = "grey") +
  scale_x_continuous(breaks = seq(1980, 2020, 5)) +
  scale_y_continuous(breaks = seq(55, 80, 5)) +
  labs(title = "Fig:1 Trend in US Birth Rates (1990-2019)",
       x = "Year",
       y = "Total Births (Ages 15-44)") +
  theme_minimal()
```

Fig:1 Trend in US Birth Rates (1990–2019)



### 3.2 3.2 Figure 2c: Trends in Hispanic Birth Rates by Origin

```
data_2c <- read.csv("fig_2c.csv")

# Rename columns to match the desired legend labels
colnames(data_2c) <- c("year",
  "brate_hispforb_other" = "Foreign-born Other",
  "brate_hispforb_mex" = "Foreign-born Mexican",
  "brate_hispnativeb_nonmex" = "Native-born Non-Mexican",
  "brate_hispnativeb_mex" = "Native-born Mexican")

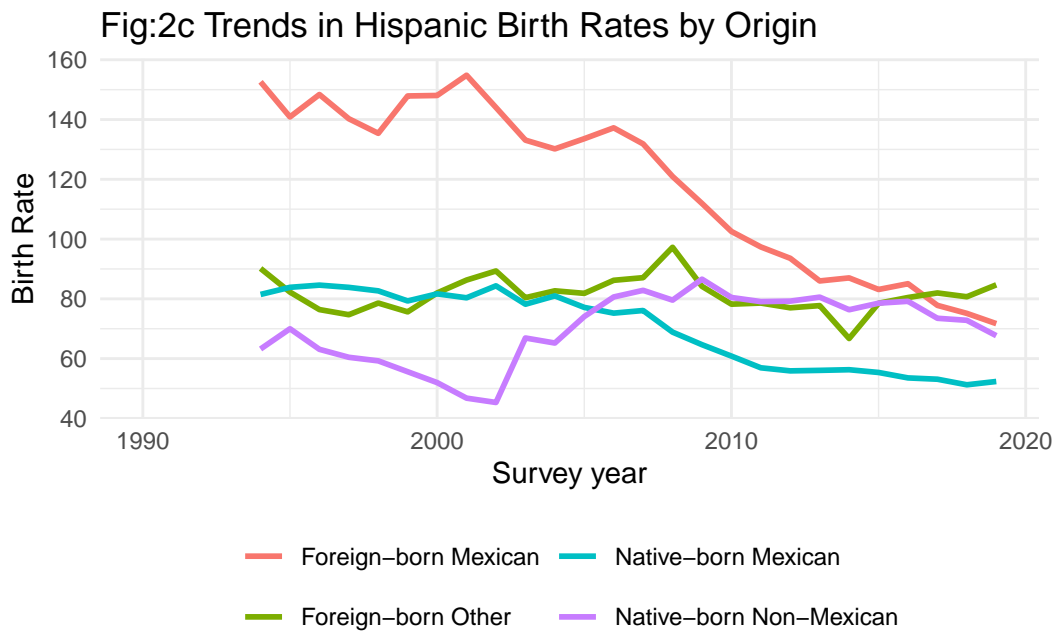
# Reshape the dataset to long format
long_data_2c <- pivot_longer(data_2c, cols = -year,
  names_to = "Hispanic_Group",
  values_to = "Birth_Rate")

# Plot
ggplot(long_data_2c, aes(x = year, y = Birth_Rate, color = Hispanic_Group)) +
  geom_line(size = 1) +
  labs(title = "Fig:2c Trends in Hispanic Birth Rates by Origin",
    x = "Survey year",
```

```

y = "Birth Rate",
color = "") +
theme_minimal() +
theme(legend.position = "bottom", legend.title = element_blank()) +
guides(color = guide_legend(ncol = 2)) # Display legend in two columns

```



### 3.3 3.3 Fig 2d: Trends in Birth Rates by Education Level

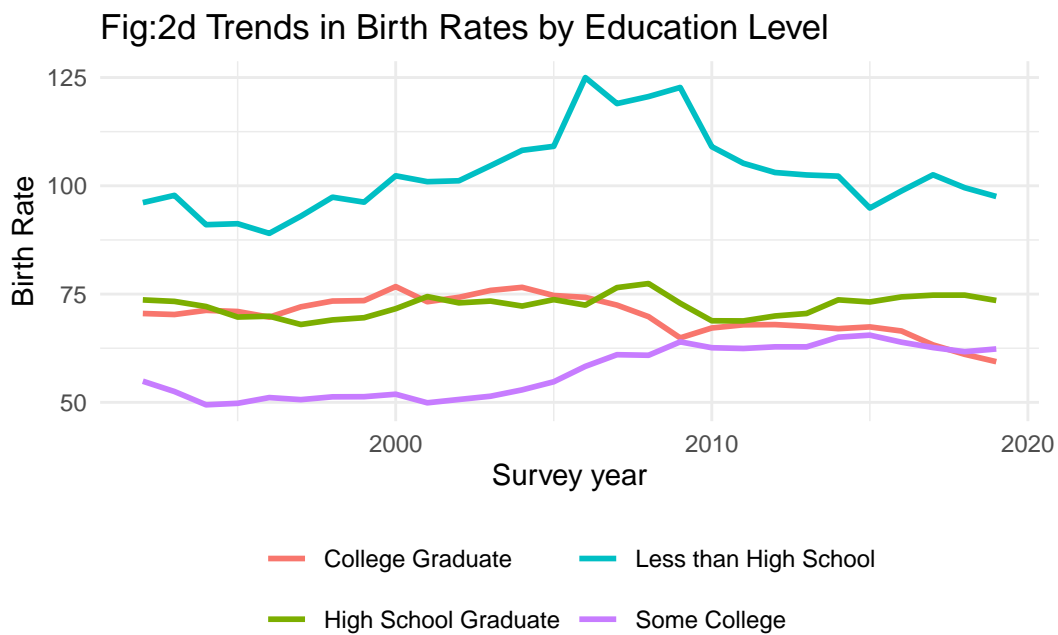
```

data_2d <- read.csv("fig_2d.csv")
colnames(data_2d) <- c("year",
  "brate_lesschs" = "Less than High School",
  "brate_hsgrad" = "High School Graduate",
  "brate_somcoll" = "Some College",
  "brate_colgrad" = "College Graduate")

# Reshape the dataset to long format
long_data_2d <- pivot_longer(data_2d, cols = -year,
  names_to = "Education_Level",
  values_to = "Birth_Rate")

```

```
# Plot
ggplot(long_data_2d, aes(x = year, y = Birth_Rate, color = Education_Level)) +
  geom_line(size = 1) +
  labs(title = "Fig:2d Trends in Birth Rates by Education Level",
       x = "Survey year",
       y = "Birth Rate",
       color = "") +
  theme_minimal() +
  theme(legend.position = "bottom", legend.title = element_blank()) +
  guides(color = guide_legend(ncol = 2)) # Display legend in two columns
```



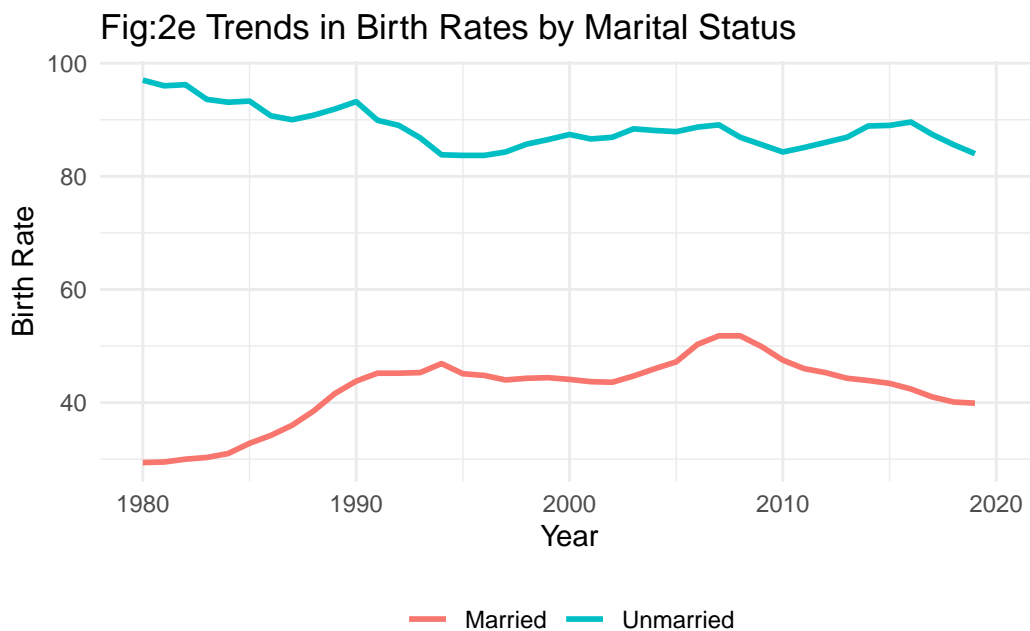
### 3.4 3.4 Fig 2e: Trends in Birth Rates by Marital Status

```
data_2e <- read.csv("fig_2e.csv")

# Rename columns to match the desired legend labels
colnames(data_2e) <- c("year",
                      "brate_married" = "Married",
                      "brate_unmarried" = "Unmarried")
```

```
# Reshape the dataset to long format
long_data_2e <- pivot_longer(data_2e, cols = -year,
                             names_to = "Marital_Status",
                             values_to = "Birth_Rate")

# Plot
ggplot(long_data_2e, aes(x = year, y = Birth_Rate, color = Marital_Status)) +
  geom_line(size = 1) +
  labs(title = "Fig:2e Trends in Birth Rates by Marital Status",
       x = "Year",
       y = "Birth Rate",
       color = "") +
  theme_minimal() +
  theme(legend.position = "bottom", legend.title = element_blank()) +
  guides(color = guide_legend(ncol = 2))
```



### 3.5 3.5 Fig 2f: Birth Rates by Parity, 15-44

```
data_2f <- read.csv("fig_2f.csv")
```

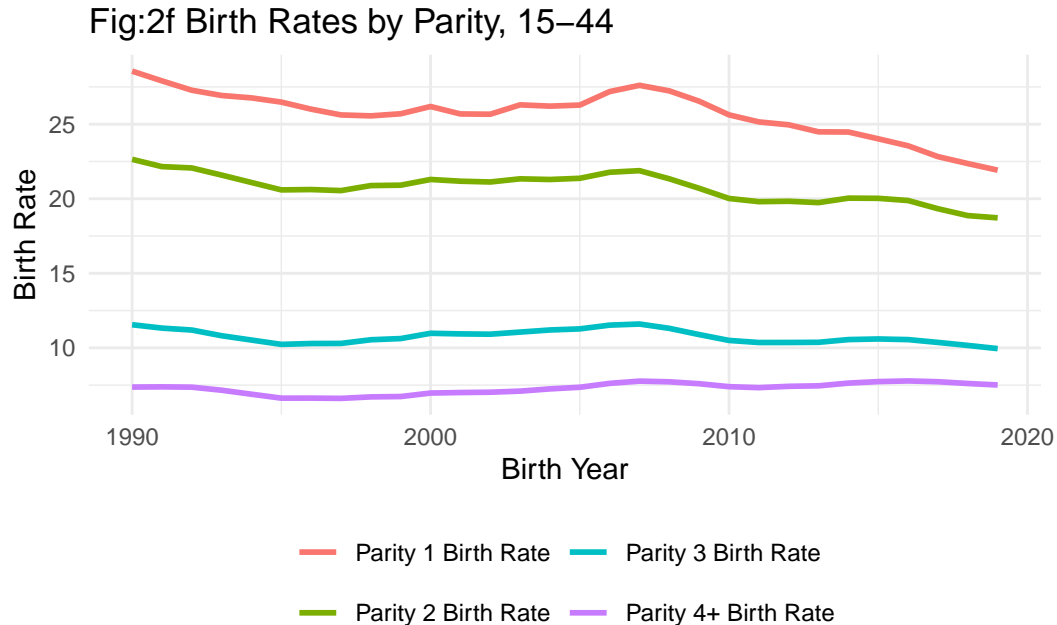
```

# Rename columns to match the desired legend labels
colnames(data_2f) <- c("year",
                      "brate_firstbirth" = "Parity 1 Birth Rate",
                      "brate_secondbirth" = "Parity 2 Birth Rate",
                      "brate_thirdbirth" = "Parity 3 Birth Rate",
                      "brate_fourthplusbirth" = "Parity 4+ Birth Rate")

# Reshape the dataset to long format
long_data_2f <- pivot_longer(data_2f, cols = -year,
                             names_to = "Parity",
                             values_to = "Birth_Rate")

# Plot
ggplot(long_data_2f, aes(x = year, y = Birth_Rate, color = Parity)) +
  geom_line(size = 1) +
  labs(title = "Fig:2f Birth Rates by Parity, 15-44",
       x = "Birth Year",
       y = "Birth Rate",
       color = "") +
  theme_minimal() +
  theme(legend.position = "bottom", legend.title = element_blank()) +
  guides(color = guide_legend(ncol = 2))

```



## 4 4 Implementation :

The code makes use of several libraries in processing and visualizing U.S. birth rate data: `dplyr` for manipulating data, like grouping and summarizing; `tidyr`, which reshapes datasets into long format for compatibility with `ggplot2`; and `ggplot2`, the main library in making line plots illustrating trends across age groups, demographics, marital status, and education levels. Besides that, the `usmap` package is used to map state-level variations in birth rate changes geographically. These libraries make data cleaning, reshaping, and visualization easier for thorough analysis.

## 5 5 Challenges faced:

The names of most of the columns had to be assigned manually as the column names in the provided data set was confusing to understand.

The `usmap` library was not readily available and therefore had to be installed as a separate package.

A do file with instructions was provided for replication of the file which could not directly be imported into R studio.

There are missing values in Figure 2a, 2b, 2c and 3 which need to be rectified before plotting the figure.

## 6 6 Further Steps for the Project

Further Steps for the Project - Construction of Figures 4 and 5: Preprocess the data sets necessary to reproduce Figures 4 and 5, using the same style and observations in the original report. Repeat the work of the other figures: cleaning the data, reshaping to a long format, and using `ggplot2` or any other tool, showing trends or geographic pattern visualizations.

-Writing of the Replicated Paper: Prepare a comprehensive written report that replicates the original study, incorporating all figures (1–5) with detailed narratives.