

A Closer Look at Year 2000 Birth Data

STAT184 Final Project

Kaitlyn Forister, Harshi Chinnam, Varun Kaveti

2024-12-17

This study of human birth patterns, using sample data from the year 2000, helps us gain valuable insights into observable reproductive trends. Identifying seasonal and temporal trends in birth data can improve resource allocation in healthcare settings. The key takeaway of our research answers the key question “During which times of the year will hospitals have the greatest demand for resources in the labor and delivery unit?”. By analyzing birth data, hospitals can improve staffing needs and provide more consistent care for both the mother and child. This research also shows patterns in human behavior, and begs the question of whether mothers prefer giving birth during a certain time of the year or if these trends are driven by subconscious factors.

Literature Analysis

Literature on seasonal and economic birth patterns reveal significant trends that can help us stay prepared in uptick seasons and predict low birth rate years before they even occur. Lund et al. (2013) found that, “The overall seasonal pattern of birth was characterized by a peak in August–September and troughs in April–May and November–January”. Factors like high summer temperatures, which may reduce conception due to decreased sperm quality or reduced coital frequency, contribute to these patterns (Lund et al., 2013). These findings help us identify peak seasons and allocate greater resources to labor and delivery professionals during these times.

Economic factors, particularly the Great Recession, have also influenced birth rates. Schochet (2014) notes that between 2008 and 2013, there were 2.3 million fewer births than expected, despite an increase in women of childbearing age. Fertility rates of women dropped significantly and unexpectedly during this period. This data suggests that financial strain greatly determines a family's willingness and ability to have a child.

Our research focuses on discovering more birth trends that will reveal information that can help L&D units stay prepared for high and low birth seasons.

Methodology

For this research, we're using a dataset from the GitHub FiveThirtyEight data library called "births." Each entry in the dataset represents the number of births on a specific day of the year. The dataset was put together by Jay Boice, Dhruvil Mehta, and Andrew Flowers, using data from the Social Security Administration, which keeps detailed records of U.S. births.

The dataset covers the years 2000 to 2014, but for simplicity, we're focusing on the year 2000 for most of the analysis. We're also using data from 2008 to make a comparison and observe any changes in birth rates over time. The dataset includes information like the year, month (1-12), date (1-31), and day of the week (1 = Monday, 7 = Sunday).

By analyzing this data, we hope to spot patterns in daily, weekly, and seasonal birth rates. These patterns will help answer our research questions about when hospitals might need more staff and whether there are trends in human reproductive behavior.

This data meets FAIR principles:

Findable: It is sourced directly from the US Social Security Administration and contains a large breadth of fortified birth data from the years 2000-2014. It is titled "birth" in the github fivethirtyeight data library, which is clear, concise, and easy to find.

Accessible: the dataset is open, free, and public use. It can be found by anyone at the link [data/births/US_births_2000-2014_SSA.csv at master · fivethirtyeight/data · GitHub](#).

Interoperable: The data is clear with all value meanings. It is stated that for days of the week, 1 represents Monday, and 7 represents Sunday, with the days in between assigned in increasing order. The months and years are clearly stated through their number representations (12 for the month of December, 2000 for the year Two-Thousand)

Reusable: This data is easy to manipulate and handle for many different uses. It is high quality and easy to be used in the future by anyone who downloads the data set off github.

Data Exploration

To make sense of the birth data, we've used a mix of data visualizations like bar graphs, dot plots, pie charts, and line graphs. These graphs help us spot trends and make the data easier to understand. Each type of visualization was used to answer different questions and highlight specific parts of the data. For example, bar graphs show comparisons between months or days, while line graphs track changes over time. Dot plots and pie charts help visualize distributions and proportions. In the next sections, we'll dive into how each of these visuals answers key questions about birth trends and what they tell us about the birth data.

What was the most common month for births in the year 2000?

This bar chart was chosen because it clearly displays the total number of births for each month in the year 2000, sorted in descending order, allowing for quick and easy comparisons. The use of distinct colors for each month enhances clarity and readability. From the visualization, August stands out as the month with the highest number of births, as indicated by its tallest bar. July and September follow closely behind with slightly lower birth totals, while April and February have the fewest births, as shown by the shorter bars on the far right. In conclusion, the most common month to be born in the year 2000 was August, as it had the highest total number of births compared to all other months.

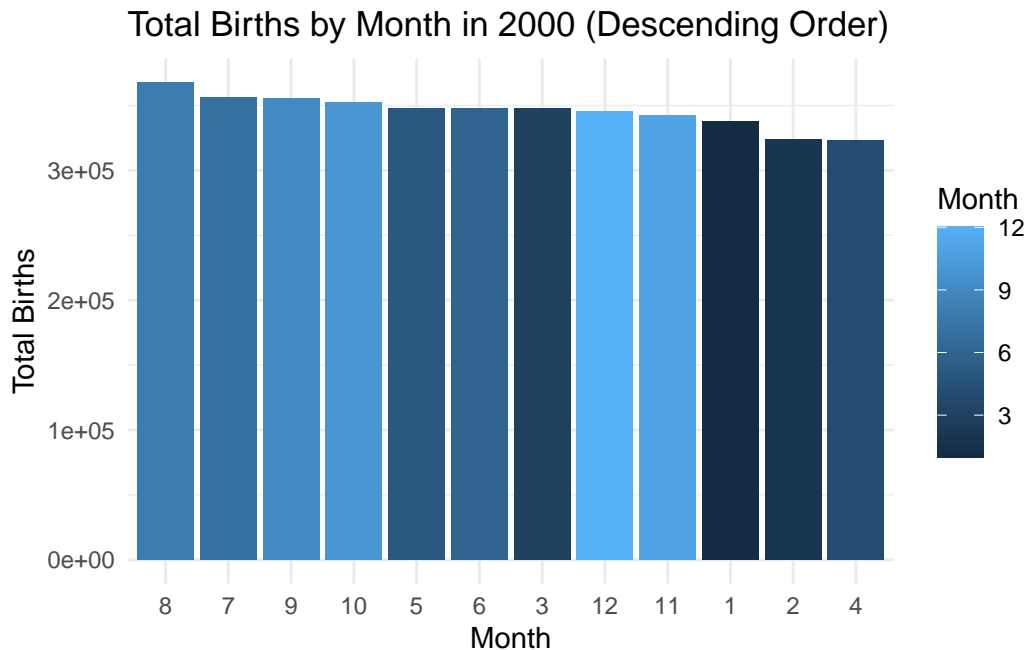


Figure 1: Bar Chart

What is the most common date of the month (1-31) to be born, and how can we effectively visualize the differences among the top dates?

We used a dot plot to visualize the top 5 most popular dates of the month to be born because it clearly displays individual data points (birth counts per date) and allows for easy comparison of differences in total births. By presenting this in a landscape format (horizontal layout), we emphasize the disparity along the x-axis (Total Births), making the differences between birth counts across dates more apparent. From the visualization, the 20th of the month stands out as the most common date for births, as it has the highest total births, represented by its position at the far right of the graph. Other popular dates, including the 21st, 18th, 14th, and 12th, follow in descending order of total births.

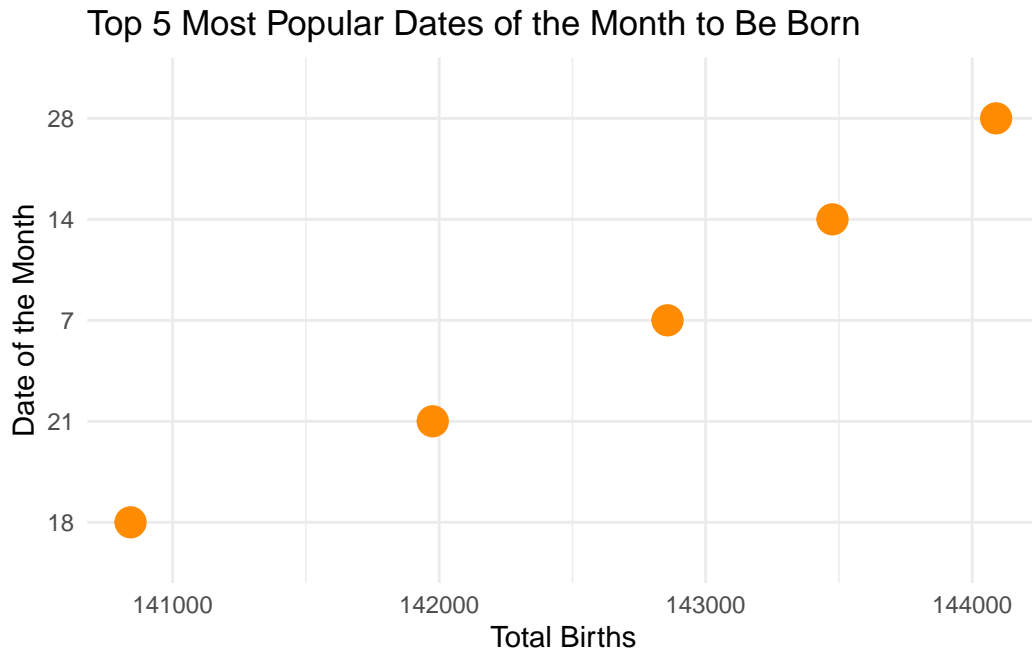


Figure 2: Dot Plot

How did total births change between 2000 and 2009, and was the birth rate in 2009 smaller than in 2000 following the recession?

From the visualization, total births started at approximately 4.12 million in 2000 and steadily increased after 2002, peaking dramatically around 2007 at nearly 4.38 million. However, following this peak, births declined sharply from 2008 to 2009, falling to approximately 4.01 million by 2009. Comparing the two years directly, total births declined by about 110,000 between 2000 and 2009. This steep drop aligns with the economic recession that began in late 2007, suggesting that economic uncertainty may have led families to delay having children due to financial concerns. To determine if the birth rate declined, we would need to compare total births to the overall US population for both years. However, given the likely growth in the US population over this period, the decline in total births indicates a more significant drop in the birth rate. Overall, yes, the total births in 2009 were smaller than in 2000, decreasing from approximately 4.12 million to 4.01 million. The sharp decline after 2007 suggests that the economic recession may have contributed to this reduction, as families postponed childbirth due to financial instability.

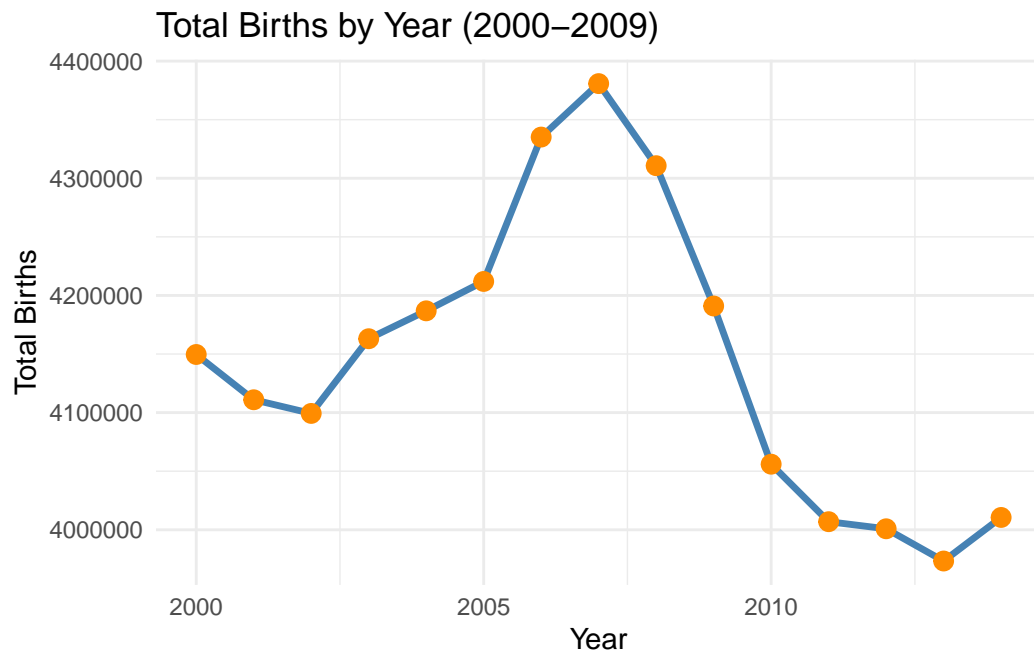


Figure 3: Line Chart

What was the most common day of the week to be born in 2000?

Births by Day of the Week in 2000

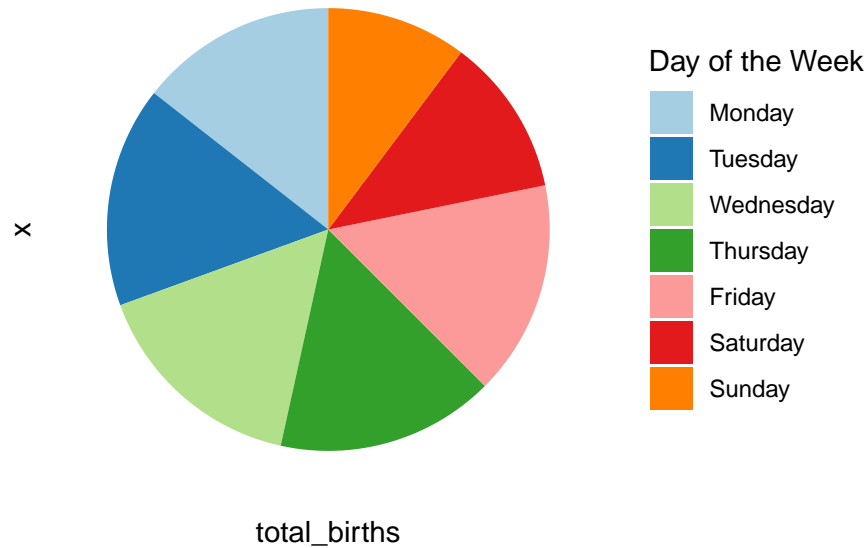


Figure 4: Pie Chart

While the data shows no dramatic bias towards a certain day of the week, there are subtle patterns revealed from the pie chart. Weekend days including Saturday and Sunday see slightly lower birth rates than Tuesday, Wednesday, and Thursday, which fall in the middle of the week. This could suggest that there is a slight preference for mothers and medical professionals to schedule non-emergency births on weekdays rather than weekends.

What was the most common holiday to be born on?

This bar chart highlights an interesting trend: women are the least likely to give birth on two major US holidays, Thanksgiving and Christmas. In particular, Christmas stands out with a strikingly low 6719 births compared to a less relevant holiday, like Columbus Day, which saw 11343 births. Interestingly, the day after Christmas, December 26th, 2000, had a large spike with 10395 births, a 54% increase. and the number of births on the day directly before Thanksgiving was 52% higher than Thanksgiving Day with 12420 births compared to 8144. The data suggests a deliberate effort from families and medical professionals to schedule births for the day before or after major holidays if possible.

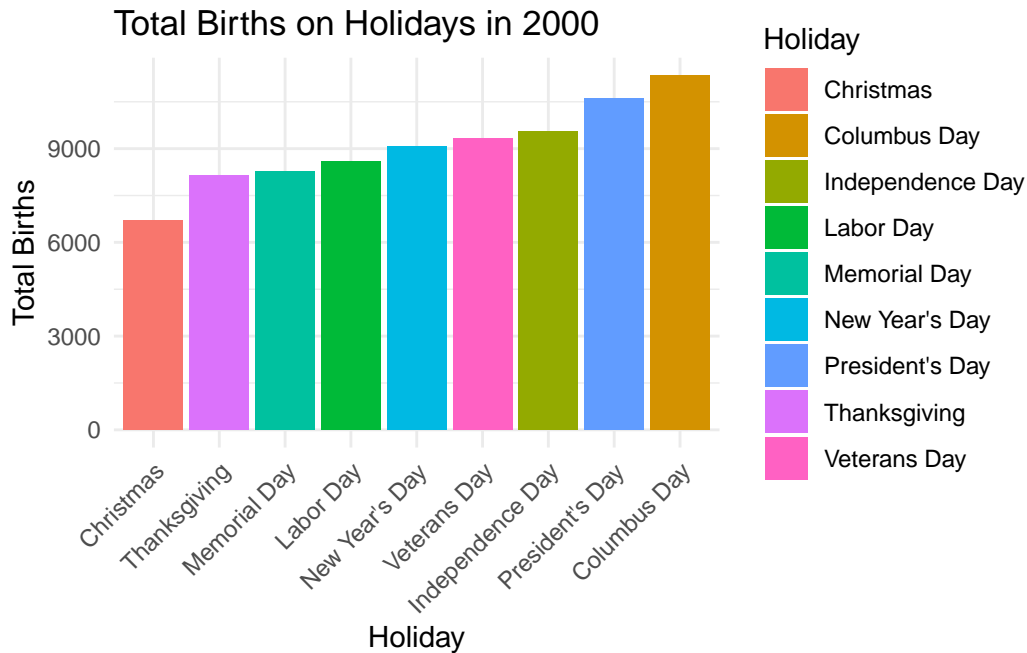


Figure 5: Bar Chart

Most common date to be born in 2000?

```
# A tibble: 1 x 2
  date      total_births
  <date>      <dbl>
1 2000-11-21    13991
```

The most common date to be born in the year 2000 was November 21st with 13,991 births.

Results and Findings

Visualizing data from the year 2000 gave us a look into when mothers are most likely to have their children. After breaking down the data, we were able to find that July, August, and September are the busiest birth months, while February and April are the slowest. We also found that the third week of the month, around the 18th, 20th, and 21st were the most common days to give birth, and the least common days of the week were Saturday and Sunday. Unsurprisingly, Christmas and Thanksgiving saw significantly decreased births with an uptick on the days before and after.

We also saw how the Great Recession significantly decreased the birth rate in the years following, which suggested that people consciously chose to not give birth when faced with financial hardship.

To continue this research, it would be interesting to see how human birth trends compare to other mammals. This would help us understand what trends are specifically of human design versus biological factors.

References

Lund, J. R., et al. (2013). *Seasonal patterns of birth: Review of the literature and implications for healthcare and family planning*. National Library of Medicine. Retrieved from <https://pmc.ncbi.nlm.nih.gov/articles/PMC3743217>.

Schochet, P. Z. (2014). *The long-term impact of the Great Recession on U.S. births*. Carsey School of Public Policy, University of New Hampshire. Retrieved from <https://scholars.unh.edu/cgi/viewcontent.cgi?article=1230&context=carsey>.

Code Appendix

```
library(readr)
library(dplyr)
library(ggplot2)

# Load data
birthsData <- read_csv(url("https://raw.githubusercontent.com/fivethirtyeight/data/refs/heads/master/data/births.csv"))
births_2000 <- birthsData %>%
  filter(year == 2000)

# Summarize births by month
most_common_month <- births_2000 %>%
  group_by(month) %>%
  summarise(total_births = sum(births)) %>%
  arrange(desc(total_births))

# Bar chart sorted in descending order with month names
ggplot(most_common_month, aes(x = reorder(month, -total_births), y = total_births, fill = month)) +
  geom_bar(stat = "identity") +
  labs(
    title = "Total Births by Month in 2000 (Descending Order)",
```



```

    x = "Month",
    y = "Total Births",
    fill = "Month"
  ) +
  theme_minimal()

# Top dates
top_dates <- births_2000 %>%
  group_by(date_of_month) %>%
  summarise(total_births = sum(births)) %>%
  arrange(desc(total_births)) %>%
  head(5)

# Dot plot
ggplot(top_dates, aes(x = total_births, y = reorder(factor(date_of_month), total_births))) +
  geom_point(size = 5, color = "darkorange") +
  labs(
    title = "Top 5 Most Popular Dates of the Month to Be Born",
    x = "Total Births",
    y = "Date of the Month"
  ) +
  theme_minimal()

# Summarize total births by year
annual_births <- births_2000 %>%
  group_by(year) %>%
  summarise(total_births = sum(births)) %>%
  arrange(year)

# Create a line chart
ggplot(annual_births, aes(x = year, y = total_births)) +
  geom_line(color = "steelblue", size = 1.2) + # Add line
  geom_point(color = "darkorange", size = 3) + # Highlight points
  labs(
    title = "Total Births by Year (2000-2009)",
    x = "Year",
    y = "Total Births"
  ) +
  theme_minimal()

births_by_day <- births_2000 %>%
  group_by(day_of_week) %>%

```

```

    summarize(total_births = sum(births)) %>%
    ungroup()

# Add labels for days of the week
day_labels <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday")
births_by_day <- births_by_day %>%
  mutate(day_label = factor(day_of_week, levels = 1:7, labels = day_labels))

# Calculate percentage for each day
births_by_day <- births_by_day %>%
  mutate(percentage = total_births / sum(total_births) * 100,
         label = paste0(day_label, " (", round(percentage, 1), "%)"))

# Create a pie chart with labels
ggplot(births_by_day, aes(x = "", y = total_births, fill = day_label)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar(theta = "y") +
  geom_text(aes(label = label), position = position_stack(vjust = 0.5)) +
  labs(title = "Births by Day of the Week in 2000", fill = "Day of the Week") +
  theme_minimal() +
  theme(axis.text = element_blank(),
        axis.ticks = element_blank(),
        panel.grid = element_blank())

holidays <- data.frame(
  month = c(1, 2, 5, 7, 9, 10, 11, 11, 12),
  date_of_month = c(1, 21, 29, 4, 4, 9, 11, 23, 25),
  holiday_name = c("New Year's Day", "President's Day", "Memorial Day",
                   "Independence Day", "Labor Day", "Columbus Day",
                   "Veterans Day", "Thanksgiving", "Christmas")
)

# Filter the birthsData for the year 2000 and the specific holidays
filtered_data <- birthsData %>%
  filter(year == 2000) %>%
  inner_join(holidays, by = c("month", "date_of_month"))

# Create the bar chart
ggplot(filtered_data, aes(x = reorder(holiday_name, births), y = births, fill = holiday_name)) +
  geom_bar(stat = "identity") +
  labs(title = "Total Births on Holidays in 2000",

```

```

    x = "Holiday", y = "Total Births") +
  scale_fill_discrete(name = "Holiday") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

births_2000 <- births_2000 %>%
  mutate(date = as.Date(paste(year, month, date_of_month, sep = "-")))
births_by_date <- births_2000 %>%
  group_by(date) %>%
  summarize(total_births = sum(births)) %>%
  ungroup()
most_common_date <- births_by_date %>%
  filter(total_births == max(total_births))
print(most_common_date)

```