

Diary Entry

Loy Yee Keen

2023-10-25

Week 9

(1) What is the topic that you have finalised? (Answer in 1 or 2 sentences).

The topic that I have finalised is “Names”, specifically “Is there a rise in the gender-neutral names given to babies born in America?”

(2) What are the data sources that you have curated so far? (Answer 1 or 2 sentences).

I have curated a dataset consisting of a list of names given to babies born in the US each year, the gender of the babies, and the count of each name per gender. The data sources span from the years 1880 to 2022. These data sources are extracted from the website of the United States Social Security Administration, an independent agency of the U.S. federal government.

Week 10

(1) What is the question that you are going to answer? (Answer: One sentence that ends with a question mark that could act like the title of your data story),

Is there a rise in the gender-neutral names given to babies born in America?

(2) Why is this an important question? (Answer: 3 sentences, each of which has some evidence, e.g., “According to the United Nations...” to justify why the question you have chosen is important)

The Council of Europe underscores the role of gender in shaping power dynamics and opportunities in society. The popularity of gender-neutral names reflects a broader shift toward inclusivity and the challenge of traditional gender roles.

Source: <https://www.coe.int/en/web/gender-matters/exploring-gender-and-gender-identity#:~:text=Gender%20is%20of%20key%20importance,equality%20and%20freedom%20from%20discrimination>

Additionally, gender-neutral names empower girls and women by challenging gender stereotypes. According to a New York Times article, some parents opt for these names to counter biases and promote strength for their daughters.

Source: <https://nypost.com/2018/03/21/why-gender-neutral-baby-names-are-on-the-rise/>

This trend aligns with the United Nations' Sustainable Development Goal 5, which aims to achieve gender equality and empower women and girls.

Source: <https://sdgs.un.org/goals/goal5>

(3) Which rows and columns of the dataset will be used to answer this question? (Answer: Actual names of the variables in the dataset that you plan to use).

I will use multiple datasets to answer the chosen question. All the datasets have the same format; each dataset represents a specific year spanning from 1882 to 2022.

In each dataset, there are 3 columns, corresponding to the name of the baby, the gender of the baby and the count of babies with that name and gender (The original dataset did not define the names of the variables, so I will redefine the variables as Name, Gender and Count).

The number of rows, each corresponding to the observation for each name, differ for every year.

Every row and column of the datasets will be used to answer my chosen question as all of them are relevant in comparing the shifts in naming trends over the years.

I will use rbind to combine the datasets into a single dataset.

(4) Challenges and errors that you faced and how you overcame them.

I encountered difficulties when I read the files using read_csv because the datasets did not have column names. Consequently, the output assigned the first value in each cell of the respective columns as the column names. This approach was erroneous, as the data in the first row represented observations, not variables. To resolve this issue, I tried to look up the answer in the textbook reading (<https://r4ds.hadley.nz/data-import>) provided in the Lecture 9 slides.