

Project 3: What's in a name?

MTH448

Due: 3/25/2025

1 Introduction

The United States Social Security Administration maintains records of babies born in the USA each year. Historical data of first names of babies is available on the Social Security website (also attached to this project). It provides the number of babies registered with a given first name in a given year. Another data set provides the number of babies registered with a given first name in each US state.

2 Project Objectives

Download the files with baby names data and analyze some facets of this data. You can use one of the data sets (national or state data), or both.

Some ideas to get you started:

- **Name diversity:** How many names were used each year? How many babies shared the same name on average? How did this change over the years? How does this differ among various states? How does this differ depending on the gender of babies?
- **Name popularity:** Which names are most and least popular? How has the popularity of various names been changing over the years? Are names popular in one state also popular in other states? What fraction of babies are given, say, one of the one hundred most popular names, and what fraction are given very unique names? Does this depend on the gender of babies? Are there sudden spikes of the popularity of some names?
- **Sex specificity:** Some names are given only to male babies, some only to female babies, and some are given to both males and females. What fraction of names belong to each of these groups? Does this differ depending on the year and state? For names given to both males and females, how can you measure if a name is predominantly male or female? Does such a measure vary by state? Are there names that over the years evolved from being predominantly male to predominantly female and vice versa?

- **Name history:** What was the first year that a given name appeared in the data? Did some names appear in some states much later than in other states? What names have been used for many years and what names were in use only for a few years? What fraction of names fall into each category? Does this depend on state and gender? Are there names that disappear for several years and then appear again?
- Anything else you find interesting.

3 Grading

The report grade will have the following distribution:

Element	Weight	What will be graded
Introduction Section	10%	Quality of narrative.
Conclusion Section	10%	Quality of narrative.
Report Content	30%	Work done on developing the project. Your analysis, insights, observations, and interpretations. Quality of the narrative of the report.
Python Code	30%	Quality of the Python code included in the report. Relevance of the code to the project. Code organization and readability. Documentation of code by code comments.
Presentation	20%	Organization of the report. Text formatting. Use of LaTeX to typeset mathematics. Formatting of code output. Quality of graphs and plots.

3.1 Introduction

The introduction is the first section of your report. It describes the project (the underlying math) and the goals of the report. It should be written in a way that is engaging and understandable to a student who has some background in math and coding but does not take this course. The introduction **SHOULD NOT** be in list form. While some of the concepts in the introduction will be repeated from the project description, you should state the concepts in your own words and not copy from the project description.

3.2 Conclusion

This section should summarize your results and major findings. It can also include potential future extensions of the project. This should be the last section of the report.

3.3 Content and presentation

Outside of the two sections mentioned your report should be broken up into sections and subsections in such a way as to maximize readability and compre-

hensibility. Where appropriate use Latex to format math.

Please note, projects ARE NOT just sets of coding exercises. They are reports in which you explore a particular math problem or phenomenon. Projects should be readable and interesting for a person not taking this class (and thus not having access to the project description) but familiar with math and python.

3.4 Code

Your code should be readable and understandable. Code should be broken up into small snippets each of which gets its own cell. Words should be used to describe what the code is doing and what the logic of your approach is. DO NOT put all of your code into a single section with no words or discussion.

Code should be readable with understandable variable names and comments included to explain what is not clear.

All python code included in the report should be written in such a way that it can be executed sequentially. For example, a function should never be used prior to its definition.

All code included must work. Do not submit code with errors.

Code output must serve the narrative of your report and should be formatted for easy reading and understanding.

3.5 Use of external resources

You are allowed but not at all required to consult resources outside of what is presented in the course. If you use an external resource in a significant way (not just googling an error message or a command that you forgot) please include a citation in your report. You are welcome to use features of python that were not shown in this class but you must understand fully what the feature does. The instructor reserves the right to ask you to explain any fragment of your code. An inability to do so may result in significant grade reduction or even more extreme consequences.

3.6 Collaboration

While students are allowed (and even encouraged) to collaborate and discuss projects the final submission must be the work of solely the submitting individual. Any assignment that is suspected of not being the student's own work will receive a zero and further disciplinary actions may also follow if they are deemed necessary. Any use of generative AI (e.g., ChatGPT) is prohibited in this class and will be considered a violation of UB's academic integrity policy.