ECON 121, Applied Econometrics and Data Analysis **Prof. Tom Vogl**

Fall 2025 **University of California San Diego**

PROBLEM SET 1: NAMING TRENDS IN THE UNITED STATES

DUE BY 11:59 PM PDT ON THURSDAY 10/9

In this problem set, you will set up your computing environment for the quarter and then explore baby naming trends in the United States. If you need to install R and RStudio, go to the RStudio website and complete the steps. Open RStudio and enter `install.packages(c("tidyverse", "fixest", "marginaleffects"))` at the command line in the Console window (bottom left). Then download the Problem Set 1 template from the GitHub folder and open it in RStudio. The template will automatically load the Social Security Administration Baby Names dataset as well as the `tidyverse`. Modify the name at the top, and then answer the questions using only `tidyverse` functions and base R.

1. List your group members.
2. Summarize the data, with 1-3 sentences explaining the summary statistics.
3. What were the most popular boy names and girl names over the sample period? How many babies were given these names? Explain in 1-2 sentences.
4. In cultures with more nonconformity, parents may be more likely to choose uncommon names, leading to a greater number of names. How has the number of unique boy names and unique girl names changed over time in the United States? For each sex separately, plot the number of unique names over time in a graph. Interpret the graph in 2-3 sentences.
5. Changes in the number of unique names may reflect changes in the number of babies rather than changes in conformity among parents. Redo the graph from question 4 to address this issue. For each sex and year, calculate the ratio of the number of unique names to the total number of babies. For each sex, plot the ratio over time in a graph. Did the number of unique names rise faster than the number of babies? Are the patterns consistent with increasing or decreasing conformity? Explain in 2-3 sentences.
6. Choose 1-3 names that interest you. State the names and explain why you chose them in 1-2 sentences.
7. Combining across all years, how common are your chosen names relative to the most popular names you found in question 3? Explain in 1-2 sentences.
8. Plot the time series of frequencies for all 1-3 names in one graph. Interpret your results in 3-4 sentences.

When you finish writing, click **knit**. R Markdown will produce an HTML file, which you should open in a web browser and save to PDF. If you have LaTeX, you can skip this step by changing `html_document` to `pdf_document` at the top, so that R Markdown produces a PDF file. Upload your PDF to Gradescope.