This is the script I wrote to run studies through a “multiverse analysis”: looping a number of datasets through a number of different analyses to understand how sensitive the study conclusions are to the choice of analytic method (a more detailed project description is provided below for interested readers). The original analysis involved 19 datasets; however, many of these were obtained from researchers at other institutions who may or may not be comfortable with their data being shared online, so the .csv file here includes only five datasets collected in my own lab.

## Project Overview:

Within the field of experimental psychology, a subfield has sprung up using simulated use-of-force situations to measure racial bias in the kinds of decisions police officers make. But experimental and analytic methods vary widely within this subfield, demonstrating a lack of consensus as to best practices in this type of research. This could lead to the perpetuation of poor practice, which could waste research resources and lead to the publication of inaccurate findings.

To better inform researchers’ methodological decisions, I built a multiverse analysis program in R. The multiverse analysis method involves looping data through every possible analysis to examine how results change in response to analytic decisions. I persuaded researchers at a variety of institutions to share their data, ultimately accumulating 19 datasets that had measured both accuracy and reaction time in shooting simulations. My R program ran each dataset through each of 25 different analyses commonly used in this subfield (including error rate ANOVAs, linear regressions, logistic regressions, and various multilevel modeling specifications of the regressions) and compiled the relevant output in a csv file. The output revealed the sensitivity of study conclusions to various analytic decisions, and how this varied depending on certain differences in study methods.

It was necessary to interpret results in light of statistical theory: for example, when there was a difference in the results of logistic regressions of binary error data versus ANOVAs of proportional error rate data, statistical knowledge was required to understand whether one technique was causing false positive results or the other was causing false negative results. However, after visualizing the results in ggplot2 and interpreting them appropriately, I produced a written report of the multiverse analysis which demonstrated the relative impact of various decisions and made specific recommendations regarding which analyses should be used, how experimental tasks should be designed depending on the response variable of interest, and how many experimental stimuli (actors in the simulations) are required for sufficient statistical power when employing multilevel modeling methods.