Sara Nasab
Snasab1@jh.edu
Module 3: Limitations of our study

There are several limitations of the study stem from two sources that are worth highlighting here: (1) anonymous user-submitted data and (2) web-scraping techniques:

First, the anonymity of the Grad Café applicant data makes it impossible to link multiple submissions from the same person. Additionally, since the data is anonymous, it also makes it impossible to follow up with applicants when the data is incorrectly entered. For example, if the GRE score (which is expected to be in the range of 260 to 340) is entered as 28 instead of 280, unless one can follow up with that specific user – that data point is unusable and is omitted from our statistics. Third, the applicants who chose to submit the data may not reflect a representative applicant pool population. For example, the applicants who were international students were more than 50%; that is not representative of a typical applicant pool. Another example is that 66.67% of the Fall 2024 were accepted into their respective graduate programs. That's much higher than most (if not all) graduate school programs. The typical average GRE score (according to BestColleges.com in 2024) was about 309; the average GRE scores were much higher for our applicant pool. These examples all suggest that the applicants submitting this data are most likely more academically driven and high-achieving than your typical applicant pool.

Second, although the web-scraping techniques employed for Module 2 were sufficient – there are more improvements that can be made. In Module 3, I made corrections to the data entered: for example, making sure that the data entered for GRE scores fell within the valid score range. However, the scraper makes no attempt to correct the data if entered incorrectly. For example, if the user enters 5.0 for their total GRE score, because it does not fall within the GRE range, it is automatically replaced with `None` in the code. However, that score is most likely a GRE AW score with a range of 0 to 6.