

# Midterm Project 1

Brady Lamson

March 14, 2022

## Data Work

To accomplish this project a large amount of data wrangling was necessary. Thankfully this wrangling was *largely* consistent across visualizations, with some slight differences depending on the specific analysis being done. The first step I took was combining the two provided csv files. Since the four year and two year files shared the exact same format it was trivial to combine them using `dplyr::bind_rows()` and saved me a great deal of trouble in the long run.

Next I began creating a couple simple helper functions, one to calculate average differences for a data frame and the other to create a simple table utilizing the kableExtra package. The first of these two is used in nearly every single visual made and is the meat of the data wrangling being done. The function selects out only the columns starting with “*dif*”, computes a summarized mean for every one of those columns and then renames all of them to be a more presentable format for visualizations. Finally it converts this new data frame to tidy format using `tidyr::pivot_longer()` to allow the ggplot2 package to work its magic. Any other data wrangling was largely smaller in scale, typically involving `dplyr::group_by()` to prep the data frame before passing it into the difference calculator function I wrote.

---

## Research Question 1

Overall, to what degree do college racial and ethnic compositions differ from the racial and ethnic compositions of the institutions' geographic "markets"?

