

"Trials and Tribulations (and Triumphs) of data acquisition and wrangling"

We faced a lot of trials during our data wrangling journey for this project. We at first tried to map covid cases based on Zip Codes in our data set, but realized that there was no easy way to map based on Zip, which meant we had to somehow merge those zip code rows with longitude and latitude data. This allowed us to create the first map. Also, before we learned how to scrape data from websites we added in MANY race proportion values from a pdf we found relevant, which was a pain, but paid off a bunch later in the project and allowed us to compare covid case proportions by race to total population proportions by race. Lastly, a lot of the column names we had gotten from our various datasets were lengthy and filled with spaces. This made it a pain to call them for ggplots, so we were forced to rename about fifteen variables to have more coder friendly names.

-Paul Martinuzzi & Lucienne Devitt

MSCS 264 B