

Final Project

Alan Qiao, Jacqueline Dinh

Read in the data

```
library(readxl)
library(tidyverse)
library(broom)
library(Stat2Data)
library(tidymodels)

nc_school <- read_excel("NC_School_Performance.xlsx")
```

Introduction

Through our project, we hope to predict grade level proficiency of a school across all EOG subjects based on demographics of the school using predictors such as total number of female, male, American Indian, Asian, Black, Hispanic, two or more races, white, economically disadvantaged, limited English proficient, and academically or intellectually gifted students, along with the total number of students and the number of students with disabilities. We intend to create a linear regression model and will calculate the predictors by manipulating the dataset.

Our EDA will include a plot of percentages of each demographic that is proficient to illustrate disparities and box plots categorized by school size.

Our hypothesis is that schools with higher percentages of white students will, on average, have a greater percentage of students that are grade level proficient. In addition, we hypothesize that a larger school will, on average, have a greater variability in students that are grade level proficient.