HED 612

Homework #4

**Directions**:

* Write your name on this document
* If the questions below ask you to execute R commands, then copy all R syntax (indicated via Courier Font) into the R script
* The R script should have a #comment indicating what number question the R syntax refers to for this assignment
* *Submit your answers file along with your R script to the D2L Dropbox*

**Before you begin**:

*Download the CA Data [If you did not complete this during class]*:

* Create a new data folder called “ca”
  + hed612 >>> data >>> ca
* Download the California Dataset from D2L (under Datasets)
  + Place the “caschool-v2” dataset into the “ca” folder you created in the previous step

*Create a new R Script for this homework assignment*

* Open the RProject you created last week (should be in your main hed612 folder)
* Once the RStudio window opens, within the R project session, open a new R Script
  + files >>> New File… >>> R Script
* Save the file as HW4\_lastname.R within lecture2 subfolder

*About the data*

The CA Schools dataset contains data on test performance, school characteristics and student demographic backgrounds. The data used here are from all 420 K-6 and K-8 districts in California with data available for 1998 and 1999. Each observation indicates one school district.

# Questions related to caschools-v2.dta

Our research question for this homework assignment is: What is the effect of teacher-student ratio on student test scores?

1. Within the R script created above: load the haven, tidyverse, labelled, and ggplot2 libraries; check that your directory is set to the R project; open the California data.

caschool<- read\_dta("data/ca/caschool-v2.dta")

1. Describe the dependent variable (testscr) and independent variable of interest (str) by running summary statistics and interpreting the min, max, and mean of each variable. *Note: For testscr, higher scores indicates “better” outcomes and is an average test scores for all students within each district. For str, the value is the number of students per one teacher at the district-level.*

summary(caschool$testscr)

summary(caschool$str)

1. In R, generate a scatter plot where X= str and Y= testscr. Interpret the plot in your own words. (Is it a positive or negative relationship? Just try your best on this!)

caschool %>% ggplot(aes(x= str, y= testscr)) + geom\_point()

1. We want to test the effect of student-teacher ratio (X) on average student test scores (Y). Write out the population regression model, label each element of the regression model.
2. Write out the OLS prediction line without estimate values for student-teacher ratio (X) on student test scores (Y).
3. Run a linear regression in R with X= str and Y= testscr. See lecture4.Rmd if you need help with syntax.
4. Write out the OLS prediction line with estimate values [rounding to nearest whole number is fine).
5. Interpret the value of in words. Is this “logical” given our X is student teacher ratio?
6. Interpret the value of in words.
7. What is the predicted average student test score for districts that have a student-teacher ratio of 25? Show your calculation using the OLS prediction line.