Problem Set 1 - Early Childhood Development
Due by Second Lecture.

In this problem set, we will look at the effects of The Abecedarian Project, a preschool intervention for children from low-income families implemented in North Carolina the 1960s. Before beginning this problem set, you should review Lecture 1 (slide deck), the Mentor Session's Lab 1 (slide deck, handout), and the supplemental **R** scripts functions_R.R and sapply_review.R.

1. Use `readRDS` to import the dataset, `earlytrainingproject_clean.rds`.[1]

2. The data set contains many outcomes, including outcomes from after the children have become teenagers or even adults. However, for this problem set we will use five outcome variables from when the subjects were still children: `iq5` (IQ at age 5), `iq6` (IQ at age 6), `iq12` (IQ at age 12), `retn12` (whether or not the child was retained at age 12), and `iep12` (whether or not the child was receiving special education at age 12). Create a vector `vars` that contain the names of the corresponding variables. Create a vector `varlabels` that contains the corresponding labels for each variable for use in making tables.

3. Create a new data frame containing only the variables that we will use in the rest of our analysis: `SUBJECT` (subject ID number), `DC_TRT` (treatment indicator), and the outcome variables in `vars`.

4. Load package `Hmisc`. (you will have to install it first if you haven't already done so). Use `describe` to explore the data set. How many observations are in your data frame? Notice that there are missing values for some variables.

5. Use `na.omit` to create a new data frame containing only those observations with no missing values. Use `describe` to explore the resulting data set. How many observations were dropped by removing the missing values? Verify that there are no more missing values.

6. What values does the variable `DC_TRT` take?

7. Create a binary variable called `treat` (using `DC_TRT`) that is 1 if treated and 0 if control.

[1] You can download the data to your own computer and then load using `readRDS`. Alternatively, you can read in the data directly from the web using `readRDS(url("https://edward-vytlacil.github.io/Data/earlytrainingproject_clean.rds"))`.

8. Use `Stargazer` to make tables of descriptive statistics for the variables in `vars`, separately for controls and treated observations.

9. Using `sapply` and `tapply`, create a table compute the mean of each variable in `vars` for the control observations, the corresponding means for treatment observations, and the difference in those means as estimates of the average treatment effects. Use `Stargazer` to create a table of the estimated average treatment effects. Briefly discuss the results. Does the early childhood intervention seem to work to improve outcomes for the participants?