EDUC 641 Assignment 01 Key

1. Read in the dataset (20% point)

Open your RStudio, create a project and save it. Go to the root directory of the project and create a folder named “data”. Download the [cat.csv dataset](https://uo-educ-quant.github.io/641/assignments/data/cat.csv) and store it in the folder “data”. Create an Rmd/R file in the project root directory. Read the data into your R environment.

2. Understand the structure of the data (40% point)

2.1. Write your own code to view the dataset and write 3-4 sentences about the structure of the data (how many variables are there, what is the current type of each variable and what the type should be, how many rows/observations, etc.).

* There are four variables in the dataset, which are childid, treat, absenteeism, and cgender
* Currently, all variables are double, but they should all be factor
* There are 942 student-level observations

2.2. Write your own code to transform the variables, *treat*, *absenteeism*, and *cgender* into factor and label them consistently with the data background information above.

3. Describe and summarize the two key variables (40% point)

3.1. How many students were in the treatment group? Control group? Provide your response and create a table AND a figure to demonstrate your answer. *Make sure to label all parts of your figure*

* There are 526 and 416 students in the treatment and control group, respectively.

3.2. What proportion of students were chronically absent? Provide your response and create a table AND a figure to demonstrate your answer. Does this seem like a large or small proportion? *Make sure to label all parts of your figure*

* About 18% of the sample students were chronically absent.
* I would say this is a normal proportion of chronic absenteeism for pre-K children (e.g., 12% for Head Start children in [a national study](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5685942/))