EDUC 641 Assignment 01 Key

1. Read in the dataset (1 point)

Open your RStudio, create a project and save it. Go to the root directory of the project and create folders named: “Code”, “Data”, “Figures” and “Tables.” Download the [cat.csv dataset](https://daviddliebowitz.github.io/EDUC641_22F/assignments/data/cat.csv) and store it in the folder “Data”. Create an R script (or .Rmd) file in the Code folder. Read the data into your R environment.

See code.

2. Understand the structure of the data (4 points)

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.). (3 points)

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

2.2. Write your own code to transform the variables, *treat*, *absenteeism*, and *cgender* into factor variables and label them consistently with the data background information above. (1 point)

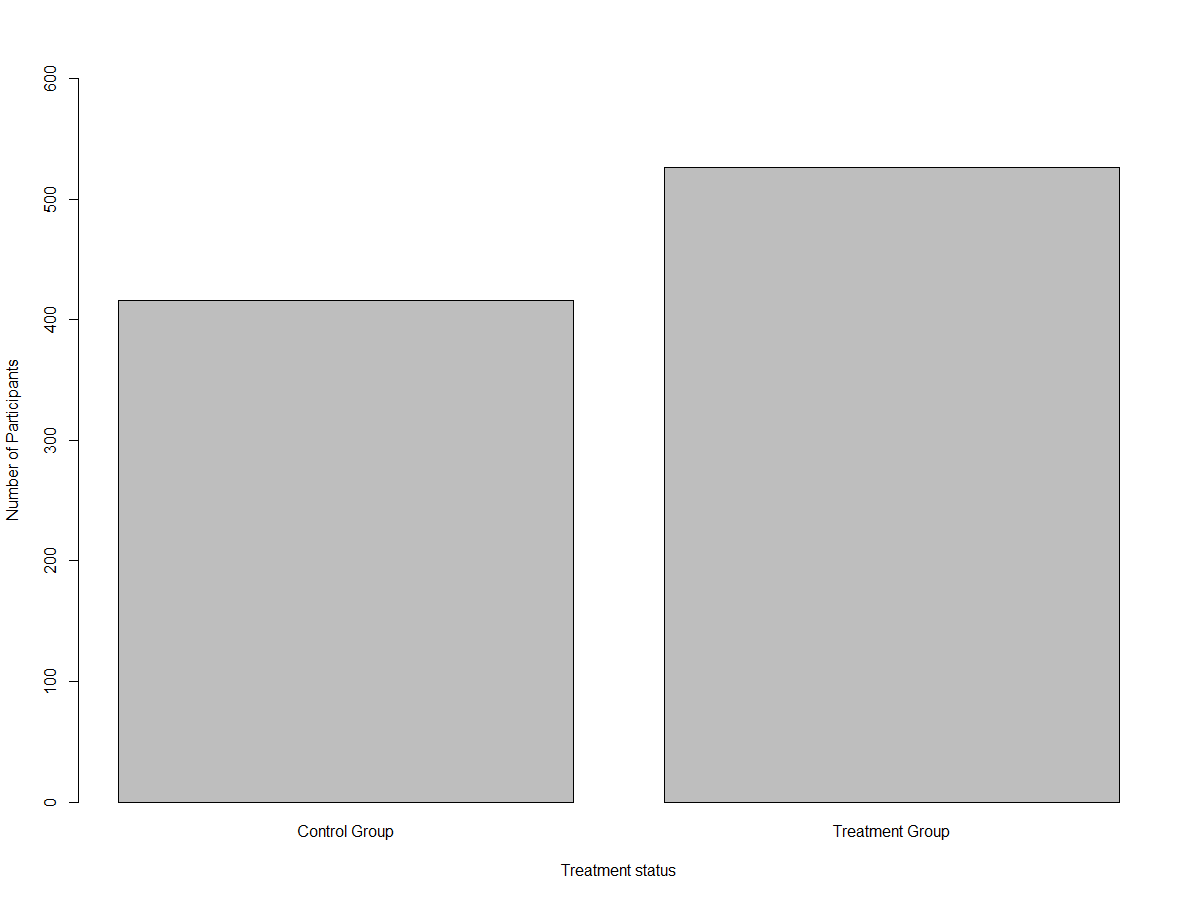
See code.

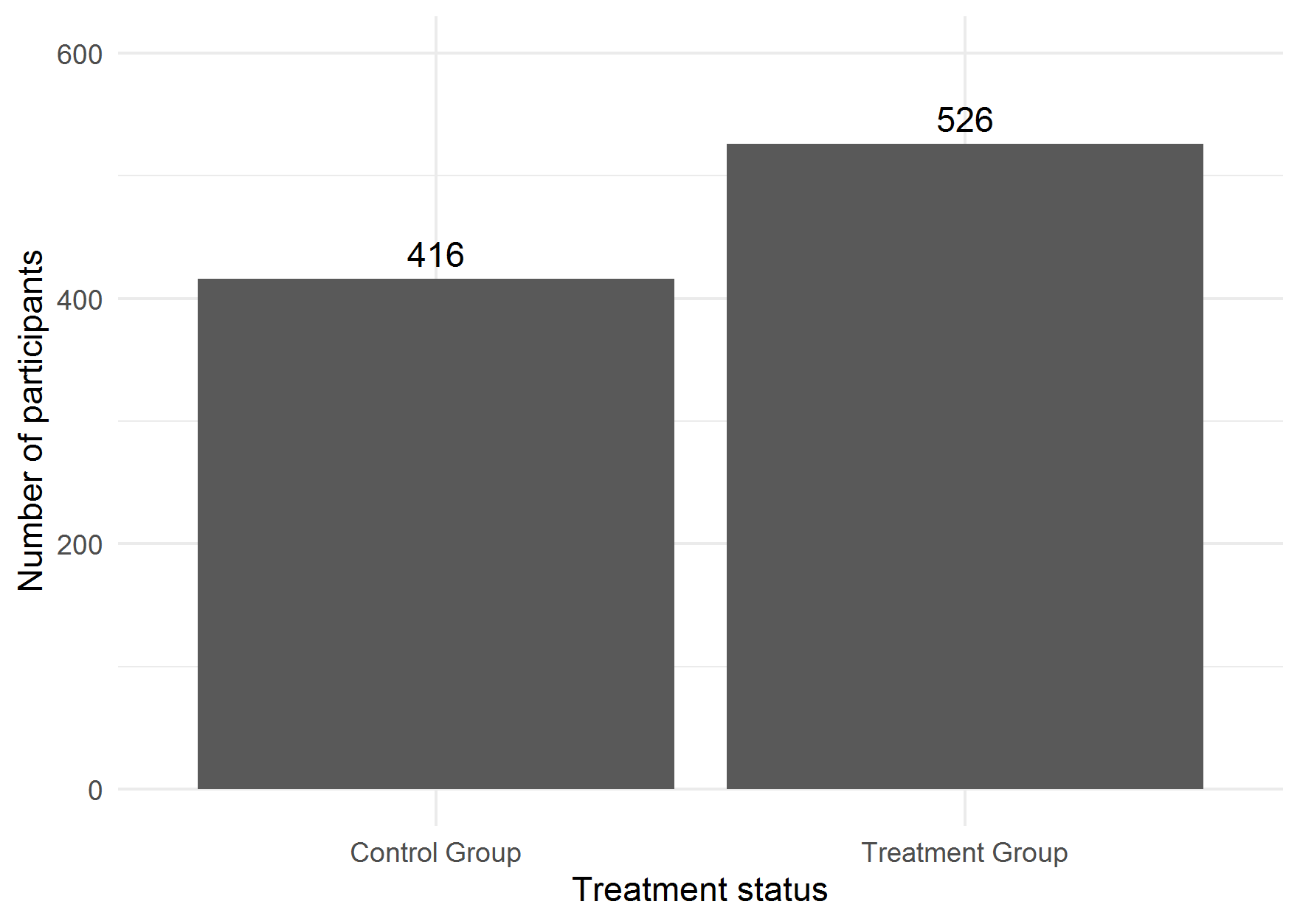
3. Describe and summarize the two key variables (10 points)

3.1. Write one sentence that states how many students were in the treatment and control groups. Provide your response and create a table AND a figure to demonstrate your answer. Make sure to label all parts of your figure! (5 points)

There are 526 and 416 students in the treatment and control groups, respectively. We present the results below in table form, and two representations of the same figure.







3.2. Write one sentence that states what proportion of students were chronically absent? Provide your response and create a table AND a figure to demonstrate your answer. Write one sentence indicating whether this seems like a large or small proportion to you. Make sure to label all parts of your figure! **(5 points)**

Approximately 18 percent of the sample students were chronically absent. This seems like a relatively high proportion of students to be chronically absent, though perhaps for pre-K children it might not be that high.



