Data Story: Parker Milestones

In my last few years of teaching Calculus I have noticed an increase in the D,W, F rates amongst students not only in Algebra sequences, but also Calculus. Students are not as prepared and sometimes have weak background of Algebra entering into the course. Calculus 1 occupies a unique position as a gateway course to science, technology, engineering, and mathematics (STEM) degrees. Almost all STEM majors need to take at least the first course in a traditional Calculus sequence. Hence for many students, this first course in the Calculus sequence is either an obstacle that they cannot overcome or a discouragement to continue in their current degree path. Many students may have felt that they were strong in mathematics in high school, but after their first college course in Calculus, they become discouraged in their abilities to continue from the unexpected rigor of the course.

Students that enroll in CMAT 111( Calculus 1), can be categorized into two groups after completion of the course; those that are successful and able to proceed to the next course by earning a grade of “C” or better (pass) and those that are not successful and unable to proceed to the next course by earning below a “C” (fail). The Mathematics Association of America (MAA) has reported the national average of unsuccessful Calculus 1 students to be 25%.

Meeting with my mentor several times and my colleague Dr. Lewis, I saw an interests to team with Dr. Lewis in exploring propensity score matching to predict/ show the students readiness for the Calculus courses. I will also take a look into if students perform better in Fall versus Spring Semesters.

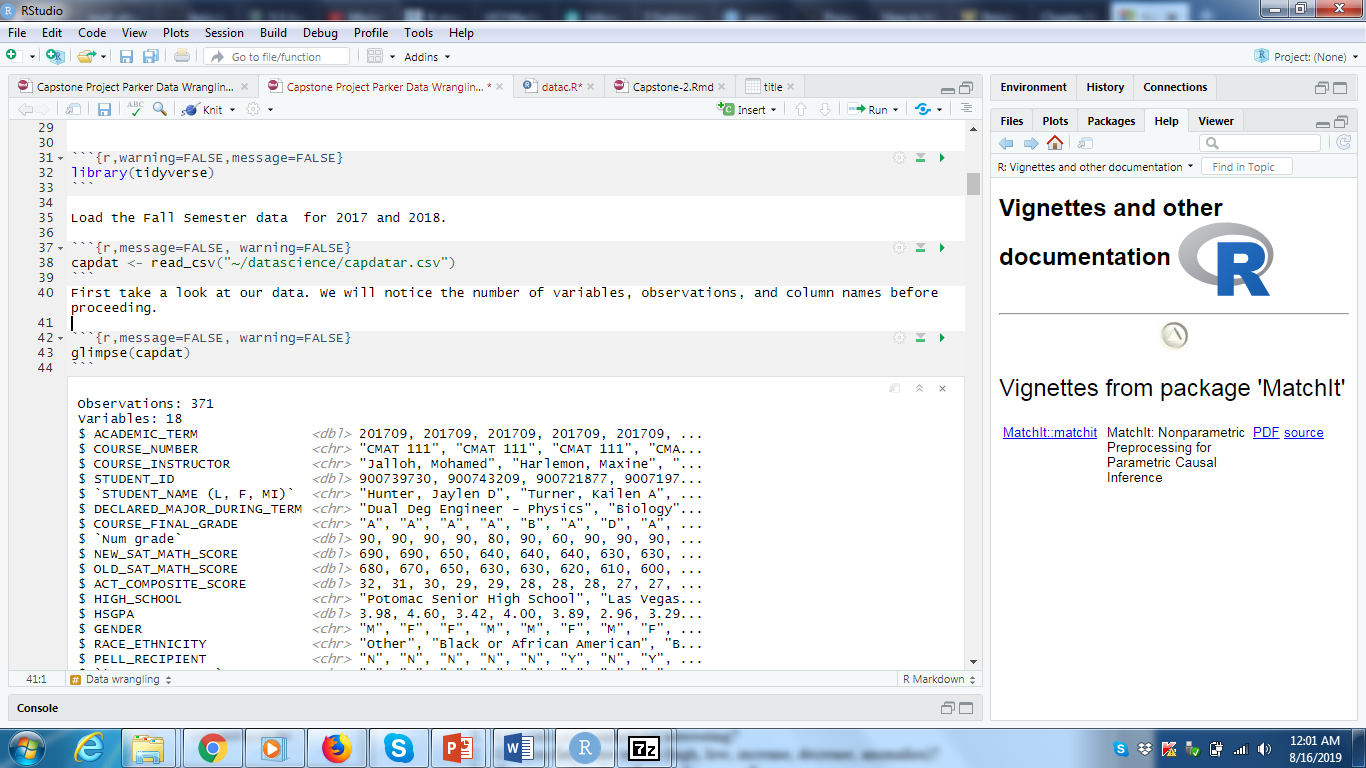
In my initial observation of the data the following was given; grades, majors, names, instructors, Pell grant, 1st generation college students, Old Sat scores, New Sat Scores, ACT, High School GPA, High School, Major, and Intervention. All of the categories are great but the most important information gathered will pertain to the Major, instructor, grades, ACT and SAT score, and intervention. After looking at the structure of the data, it was evident that all categories did not have any data. Not having all the data for the SAT and ACT scores, meant I cannot test the readiness students are for Calculus 1. I started with 371 observations and 18 variables. Out of the 371 variables only 117 of them had SAT scores. This approach would make me lose nearly half of my data set. However, this does give to opportunity in dive into a better understanding of propensity score matching.

Propensity Score matching is a matching technique that attempts to estimate the effectiveness of a treatment group versus a control group through observational statistics. This method also reduces the biasness due to cofounding variable.

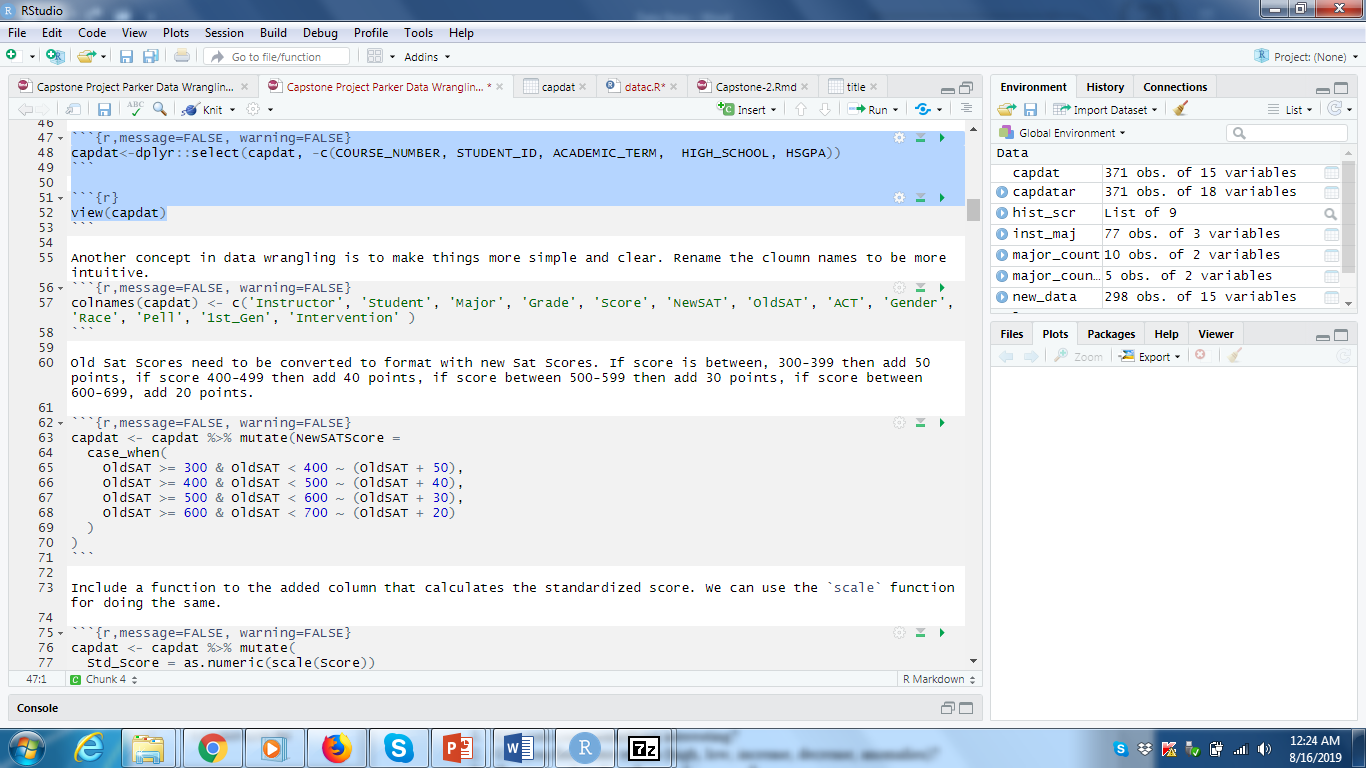
Now, looking at the data set, I will see if the intervention in the calculus I course is effective. The intervention at Clark Atlanta University began Spring 2018 with a course redesign using the adaptive learning; Assessment and LEarning in Knowledge Spaces (ALEKS). The goal of the effort was to improve students’ mastery and the use of mathematical concepts through course redesign, assessment, and implementation using ALEKS which will enhance student mastery of learning outcomes, retention, and persistence rates in the undergraduate STEM degree programs at CAU. The ALEKS course product used was “Prep for Calculus,” a course that is designed to help students to develop the prerequisite foundation needed to learn Calculus I. Calculus I students in the course were divided into two groups. One section of the course was exposed (treatment group) to the ‘intervention”, while the other students learned Calculus I with previous teaching methods (control group).

Data wrangling is necessary to "clean up" data in order to analyze your information. The packages below were necessary for this part of the Data Wrangling project. Load the necessary packages. I had to install a few packages before loading.

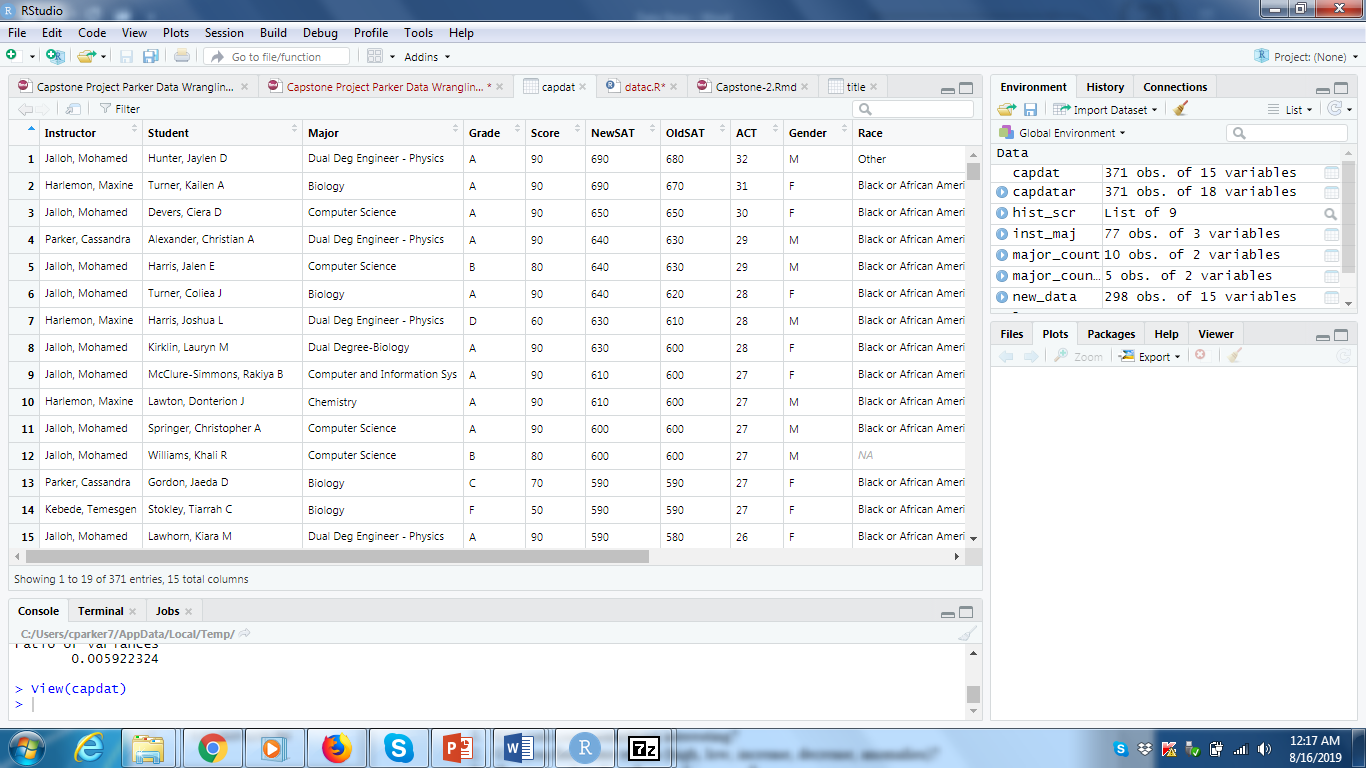
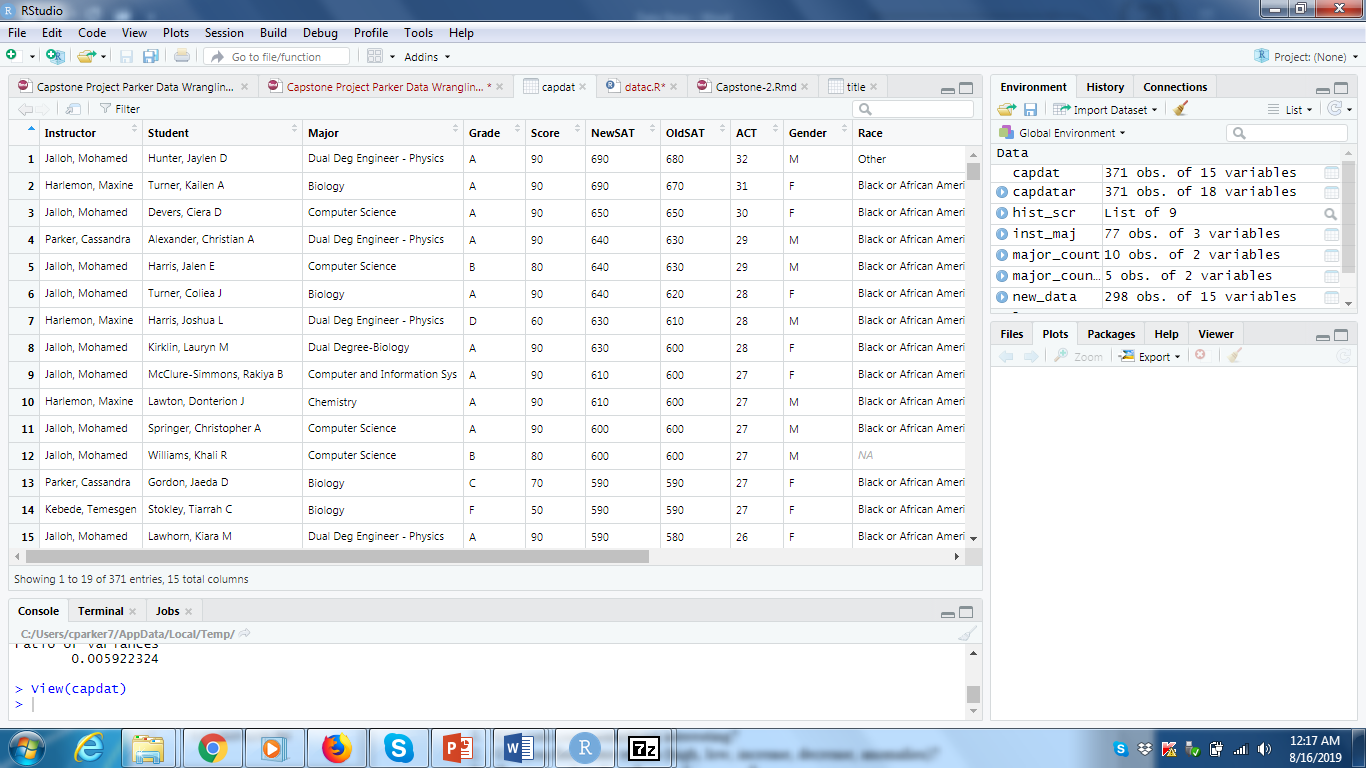
Data set that was provided from Clark Atlanta University has been cleaned. The techniques used were necessary in removing variables and changing names of columns, adding variables and few calculations.



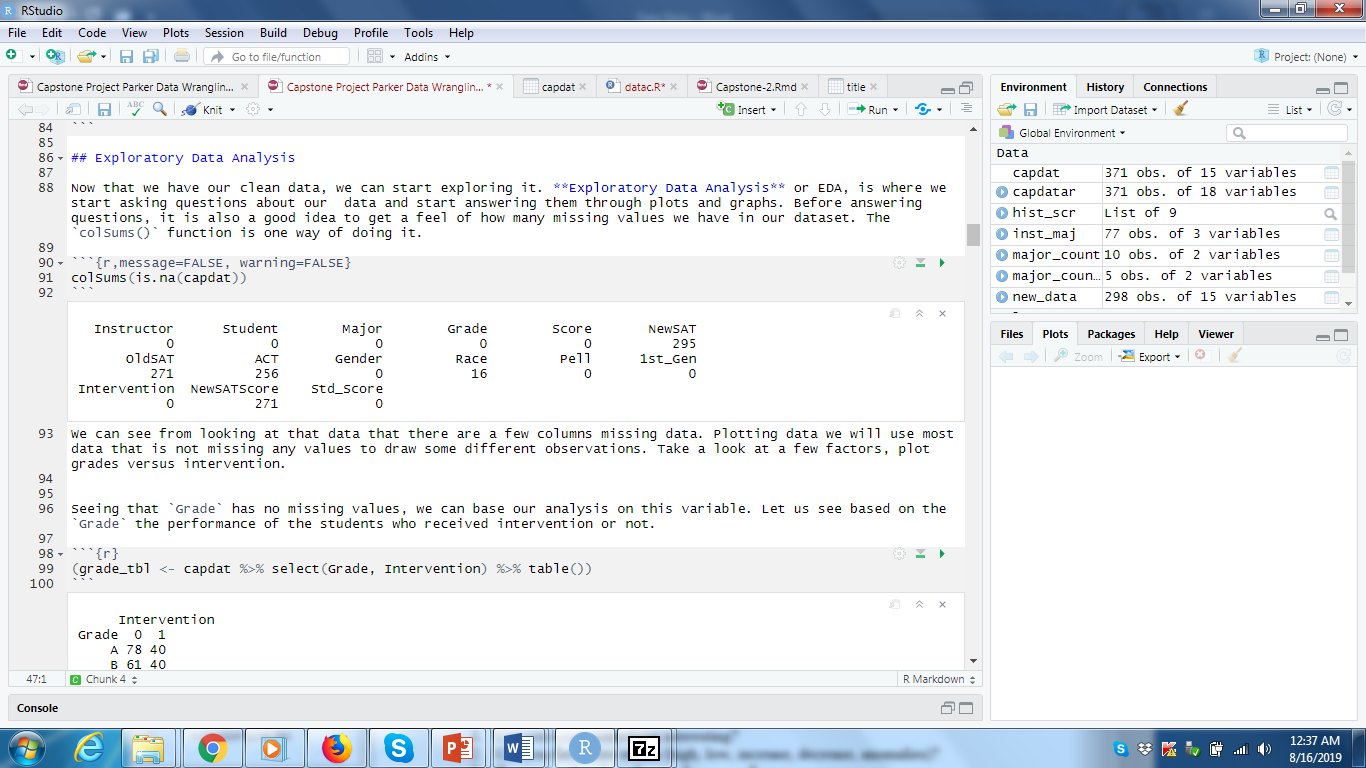
It is best to select the variables of interest. Therefore, remove the columns that are not necessary. Since all data is already Calculus courses from Fall semesters 2017 (control) and 2018 (intervention), we can eliminate those columns along with a few other columns, such as high school and HSGPA. The dplyr package was used in removing those columns. To make the columns more intuitive, I used colnames.

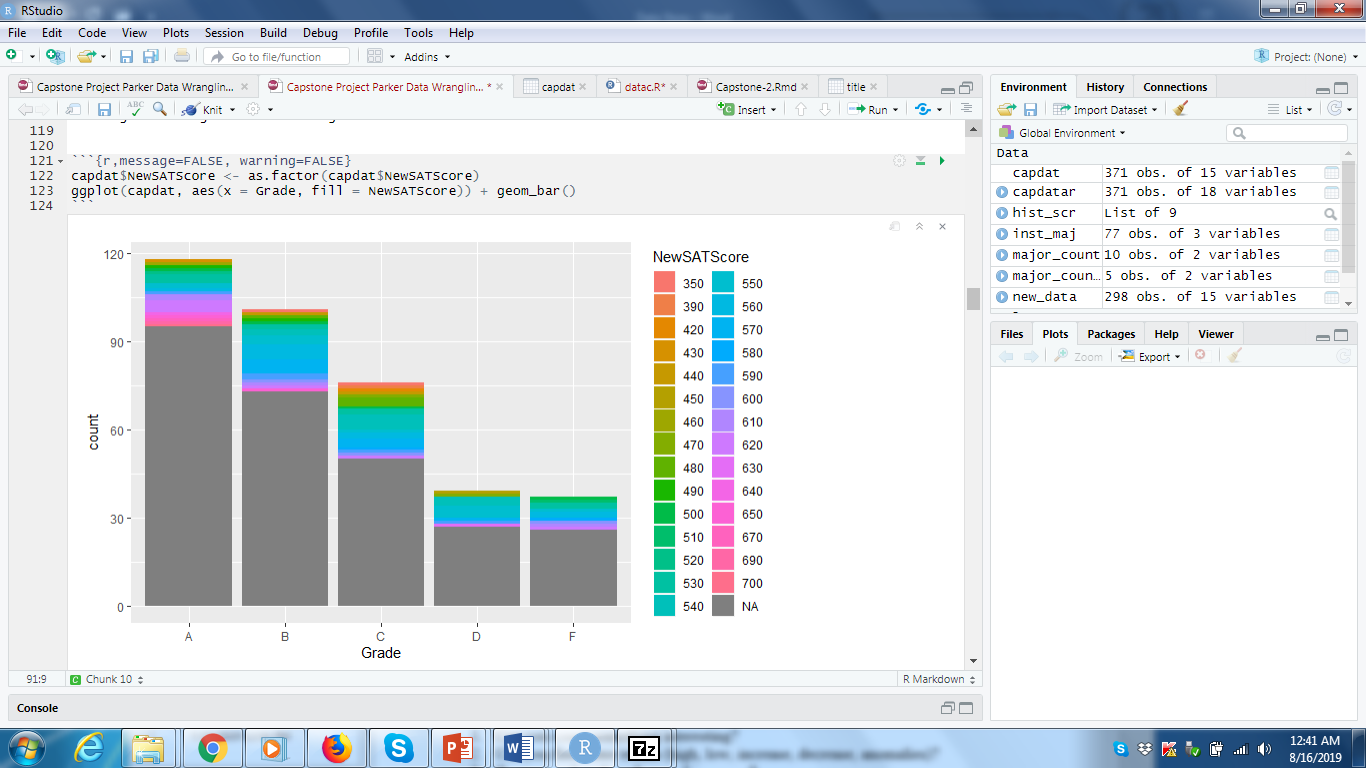


In 2016 SAT changed their testing scores. So to make all the SAT scores in the newly “scores”, I used and interesting function mutate. Mutate allowed me to add to multiple ranges and add a new column to replace the old values seen below. This view function is a nice feature to look at your entire data. Capdat as my new data set now have 15 variables and 371 observations.



During the exploratory data analysis is where I presumed I could not use the SAT scores. Here in the columns sums that have NA. SAT, ACT and Old SAT were the only columns that several observations missing for those variables. As well as using ggplot. Again too much Not Available (NA) was plotted as plotted on grey.





Also during the exploratory analysis I found a few interesting plots in relation to the instructors teaching in terms of the intervention. And just the overall grade in terms of the intervention allowed me to draw the conclusion that intervention is working. 