Problem Set 3

Due by 3/2

**Generate a script file for your answers and you will submit a print of your script only. Before working on this PS, take a look at the third set of slides on Python.*

A researcher is interested in New York State Assessment results provided by the *New York City Department of Education* for grades three to five, from 2006 to 2017. Specifically, she is focused on the percentage of the student population with in a given DBN at each grade level that earned a passing test score of three or four on New York State wide math and reading exams. Here, a DBN (district, borough, number) identifies a specific school. She uses this percentage of students passing state assessments as a representation of school quality. She begins with grade three because that is when students are required to begin to take state examinations. She also focuses specifically on grades three to five because in New York City students are required to go to their zoned school, which is the school in their designated school district assigned based on place of residence, until Middle School which can begin at grade six.

She collects the following four data sets.

(1) `Math_2006_2012_All.csv`
(2) `Math_2013_2017_All.csv`
(3) `English_2006_2012_All.csv`
(4) `English_2013_2017_All.csv`

Note again that the data sets contain percentage of the student population with a passing grade broken down to four levels. Also, note that the column names accross the English score data sets are not unified.

(1) Import all files to Python.

(2) Change the measurements of percent passing variables to ratios by dividing them by 100.

(3) Append the English score data sets and name it `English_2006_2017_all`. Similarly append the Math score data sets and name it `Math_2006_2017_all`.

(4) Merge `English_2006_2017_all.csv` and `Math_2006_2017_all.csv` on `DBN`, `Year` and `Grade`. Name this data set `ME_2006_2017_all`.

(5) Generate a new string variable, `borough`, by slicing the third character in the `DBN` column in `ME_2006_2017_all`.

(6) Generate a table of descriptive statistics for `ME_2006_2017_all`.

(7) Grouping by `borough`, `year` and `grade`, obtain *mean*, *standard deviation*, *min* and *max* of percent passing scores for both types of exam.

(8) Generate a new numeric variable, `district`, by slicing the first two characters in the `DBN` column in `ME_2006_2017_all`.

(9) Regress percent passing for English scores, on an intercept, year dummies, borough dummies and district dummies.

(10) Regress percent passing for Math scores, on an intercept, year dummies, borough dummies and district dummies.