ASGN 2
Due by 4/25 10:30 PM
The assignment has two parts. We will start the work on PART A on 4/11 and on PART B on 4/18. Both parts are due on 4/25.

Part A:

Data for this assignment represent a very small sample (100 observations total, extracted for demonstration purposes) from a couple of surveys on family finance and financial consumer characteristics carried out for the Board of Governors of the Federal Reserve System in 1963. All data files are in asgn2 folder and P:\QAC\qac156\data\ ; the unit of observation is the household head

family63A_tab.txt is a tab delimited file with 80 observations and the following variables (names in row 1) id fs edu age exp m race region (id unique identifier of household head, fs is family size, educ is the number of years the household head has received, is exp is the labor market experience of the household head, m is number of months the household head worked the last year, for region 1 is Northeast, 2 is North Central 3 is South and 4 is West.)

family63B_FreeForm.txt is an unformatted ascci file with the following variables
id fs edu age exp m race region
for 20 observations

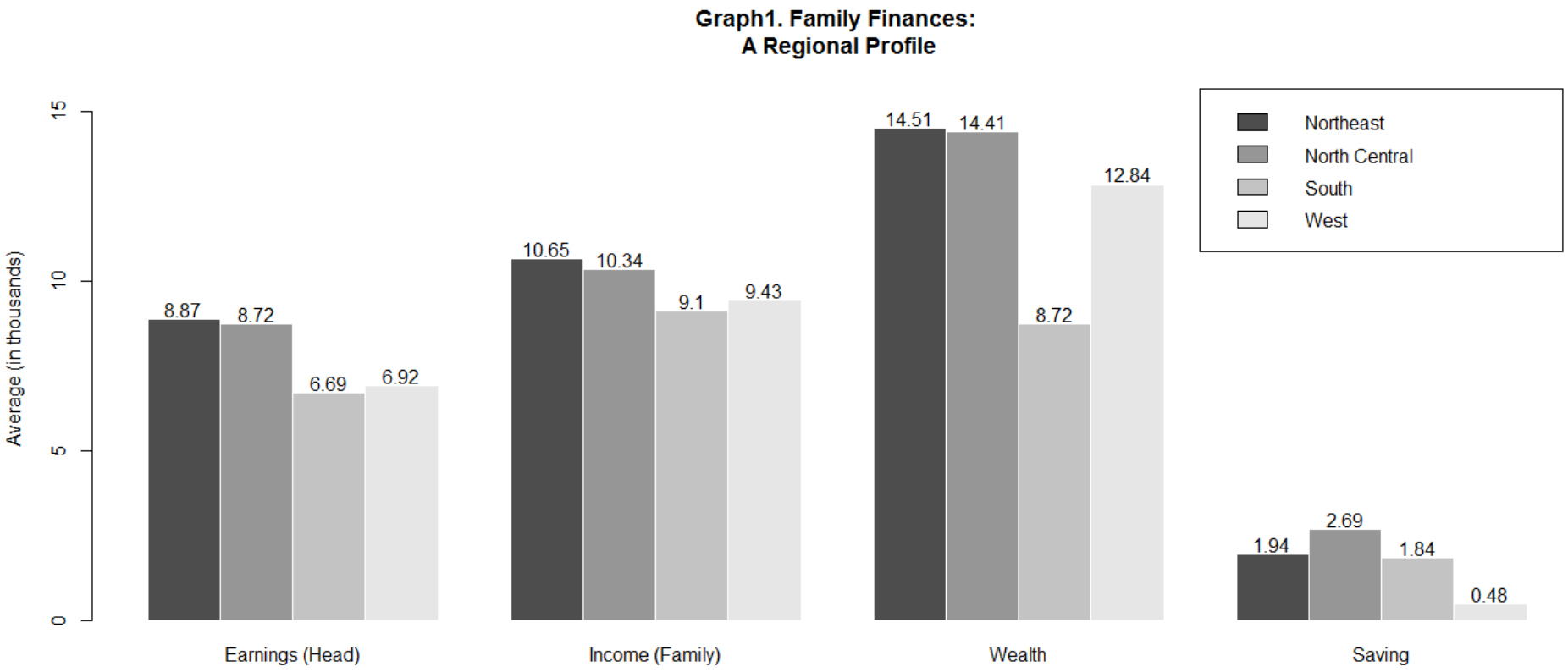
family63C.xls has all 100 observations and 4 additional variables in addition to the unique identifier
E is the wage and salary earning of the head expressed in thousands of dollars
I is the family total income, in thousands
W is Family wealth, in thousands
S is the family saving (flow), in thousand

1. import family63A_tab.txt (tab delimited data) as A.data
2. import family63B_FreeForm.txt (variable names: id fs edu age exp m race region) as B.data
3. import family63C.xls as C.data with variable names set to lower case
4. Merge, append as appropriate to create a data file with all observation and variables that you can use to
    a. calculate appropriate summary statistics for edu, race, region , e, i, w, s

b. use appropriate graphs to show the distributions of region and e. (Just include relevant code in your R script file. You don't need to add these graphs to the pdf file along with the graphs of (c).
c. to produce the following three graphs
(If you run into any problems in the previous steps (putting all the data together) you can use family63.RData (in the course's data directory) to create the graphs)

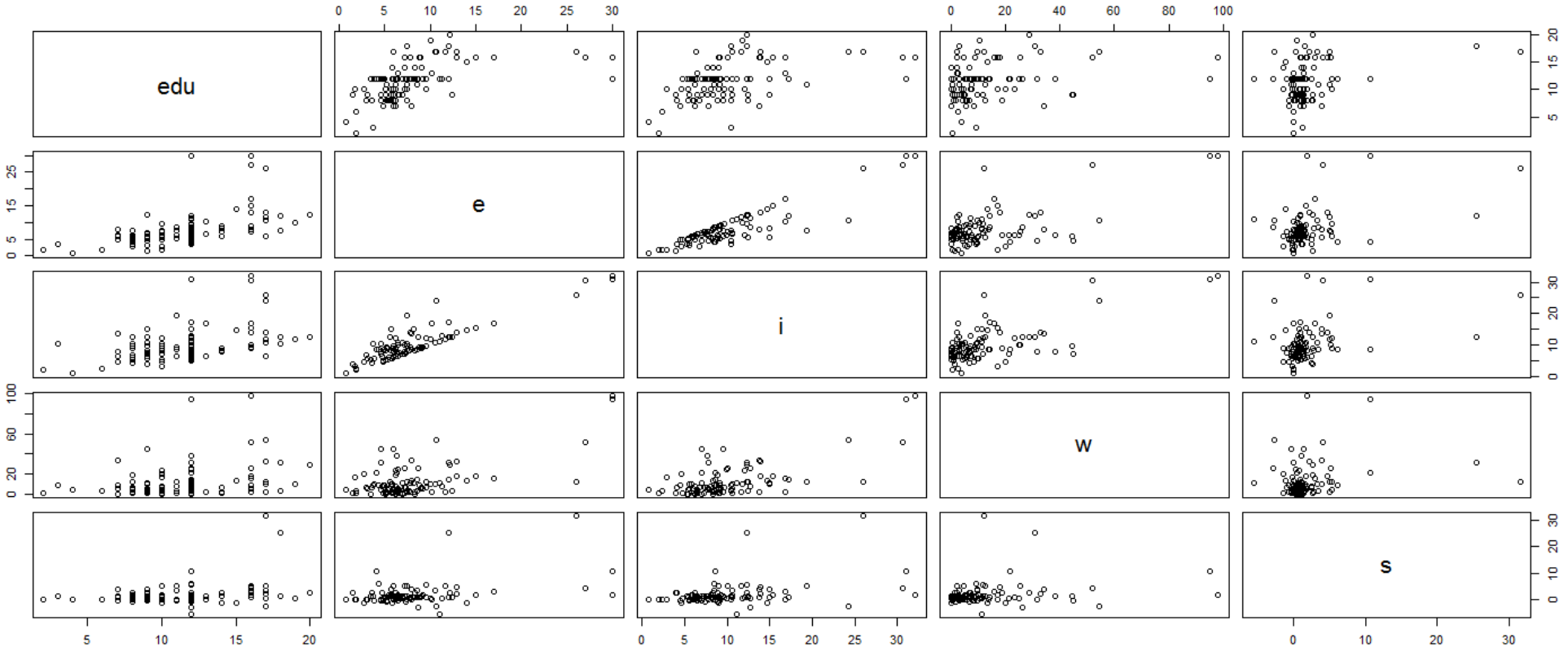
For PART A please submit, via Moodle, two files: a) Your R script file (name it yourusernameASGN2A.R) with appropriate comments for all steps of the assignment; and b) a pdf file (name it yourusernameASGN2A.pdf) with the three graphs of part 4c. Graph 3 and 4 are optional (will add 10 bonus points). Again both file names should start with your username (e.g. mkaparakis_asgn2A.R, mkaparakis_asgn2A.pdf)

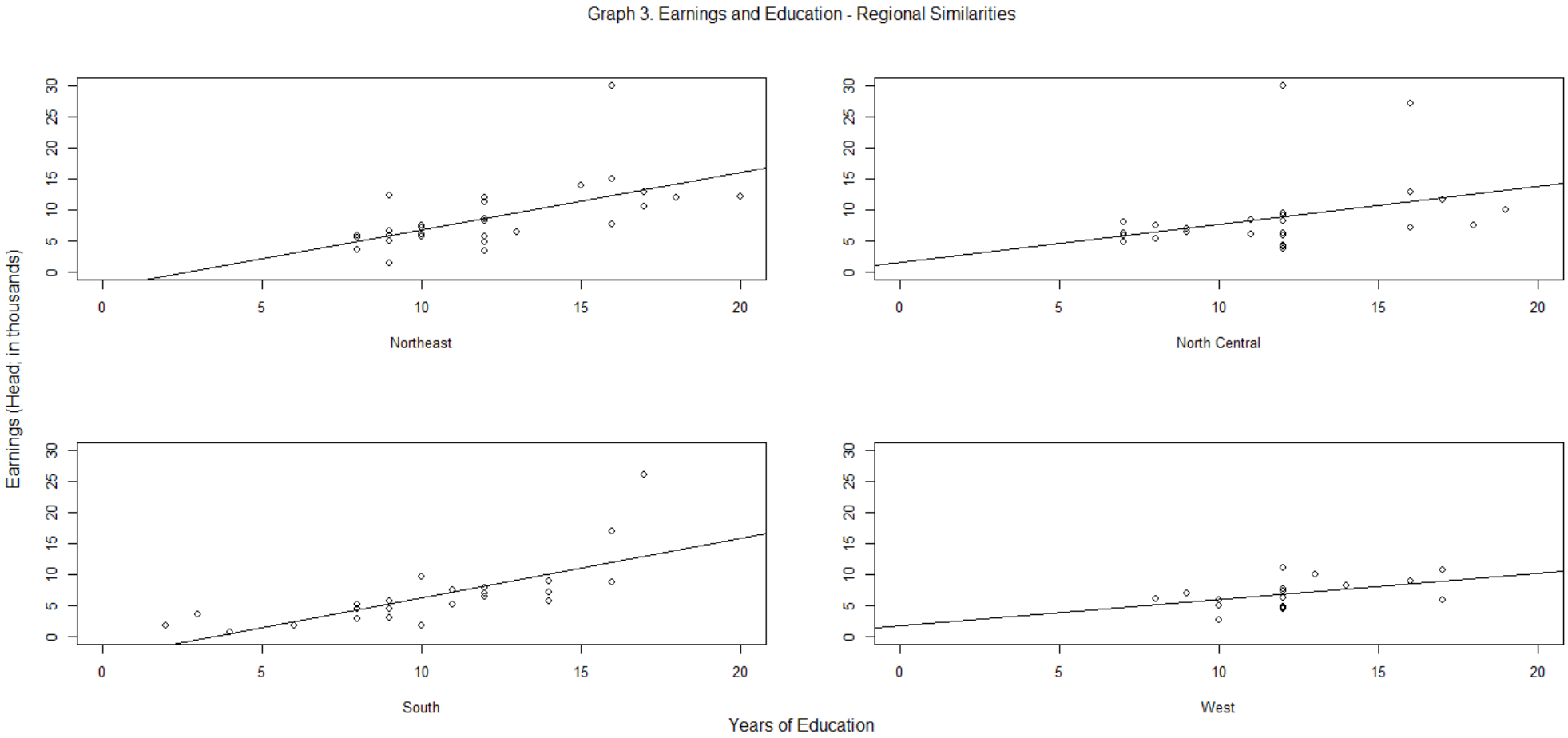
You may discuss parts of the assignment with the QAC tutors and course TAs, but I will again ask that you include (you must include) comments next to your script lines indicating that you received help in that particular part of the script.



You will need to get the means of the variables by region (aggregate (collapse) your data) before you create the bars

Graph 2. Exploring Simple Associations





functions to consider:

par

plot

abline

mtext