Lab 10 *Some of these problems may be more challenging than others. Please feel free to work with others, attend office hours, or post on the course discussion forum if you need help. While collaboration with other students is encouraged, each student is responsible for submitting his or her own work. This assignment should be submitted in one well-commented SAS program. For any questions that require a written answer, do so in the SAS comments. Be sure to re-name the uploaded SAS scripts according to the naming convention* `LastnameFirstinitial_Lab#.sas` (*e.g.,* `PileggiS_Lab10.sas`).

Many people incur debt at some point in their life, and for some it starts with student loans. The `studentloans` data is for students at some of the most and least expensive private institutions.

`Loan`	the total amount that the student took out in loans - the loan does not accrue interest until after the student graduates
`Interest`	the rate is the *annual* rate which the amount to be paid back increases after graduation, and is compounded *monthly* (so the monthly interest is `Interest/12`)
`College_Start`	the year the student started college
`Years`	the total number of years the student spent in school
`Salary`	the current salary of these former students
`Payment`	the *monthly* payment the student is making towards the loan, after graduation

1. Create a library reference called `mylib` to access the `studentloans` and `adni` SAS data sets.
2. Assuming fixed income and payments, determine when each person pays off their loan by completing the following tasks with the `studentloans` data set.
   (a) Write a loop that identifies how many months until the loan is paid off (i.e., the balance is ≤ 0). Here is the idea behind the calculations:
   - initialize the values of
     - `months` to 1
     - `balance` to `loan`
     - `monthly_interest` to `interest/12`
   - each month, the monthly payment is deducted from the loan balance
   - each month, the loan balance increases by the interest rate
   - increment months by one

- continue until the loan balance is less than zero

(b) Print your output with an appropriate title. Your results should look something like this (only select variables shown):

Obs	name	loan	interest	months
1	Hank	45000	0.0590	63
2	Sarah	40000	0.0580	51
3	Steve	204000	0.0600	376
4	Chris	180000	0.0550	480
5	Emily	120000	0.0580	196
6	Jessica	6000	0.0500	9
7	Mark	38000	0.0575	52

(c) In a comment in your SAS code, briefly describe how SAS executed this data step.

Recall the `adni` data set from the Alzheimer's Disease Neuroimaging Initiative (ADNI) longitudinal study that began in 2005. This research is designed to track AD biomarkers, identify at-risk patients, and evaluate the efficacy of novel treatments.

`DX`	Alzheimer's disease diagnosis 1 - Normal cognitive function 2 - Mild cognitive impairment (MCI) 3 - Alzheimer's disease (AD)
`AGE`	Age (years)
`APOE4`	Type of APOE4 variant (genetics) 0 - No copies of the ApoE4 allele 1 - One copy of the ApoE4 allele 2 - Two copies of the ApoE4 allele
`GENDER`	Patient gender
`MMSE`	Mini Mental State Exam (score out of 30, lower scores indicate more cognitive impairment)
`ADAS`	Alzheimer's Disease Assessment Scale (larger scores indicate greater dysfuction)
`WholeBrain`	Brain volume (mm^3)

Example plots to verify your work are on the last pages of this lab.

3. Create a format for the `dx` variable that displays values as `Normal`, `MCI`, and `AD`.

4. Create a temporary data set which copies `adni` data set. In this temporary data set, apply the format to `dx` and also apply a label to `dx` so that it displays "Diagnosis". Use this temporary data set for the remaining exercises.

5. Bar plots

   (a) Create a stacked bar plot that displays the frequency of `dx` separated by `APOE4`. *(Note: `APOE4` is spelled with an "oh" and not a zero.)*

   (b) Create a side by side bar plot that displays the frequency of `dx` separated by `APOE4`. *(Note: this is not an option we discussed in class! You'll need to use the help file for sgplot to figure this out.)*

6. Box plots

   Create a side by side box plot that shows the distribution of `MMSE` by `dx` such that the `dx` categories are displayed in order from `normal` to `MCI` to `AD`. *(Note: in order for boxes to be displayed in order, you need to sort your data set by `dx` using `PROC SORT`.)*

7. Histograms

   (a) Create overlaid histograms that shows the distribution of the `MMSE` and `ADAS` scores. Be sure to adjust the transparency so that the two distributions are visible.

   (b) Create overlaid histograms that shows the distribution of the `MMSE` among males and females. Be sure to adjust the transparency so that the two distributions are visible.

8. Scatter plot

   Create a scatter plot showing the relationship between `WholeBrain` (y) and `AGE` (x) with solid circles.

9. Modify your scatter plot code from the previous question to turn it into a macro.

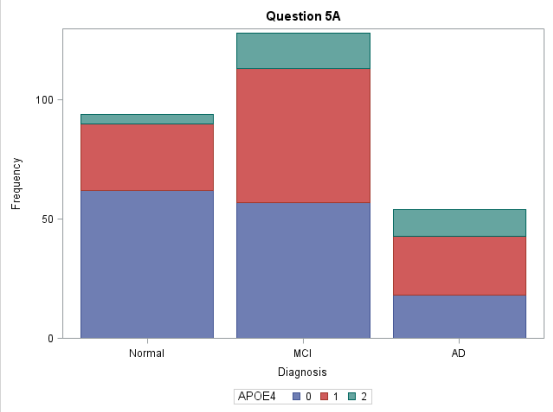
   (a) Use the same base scatter plot, but modify it to plot points in different colors by the values of a categorical variable.

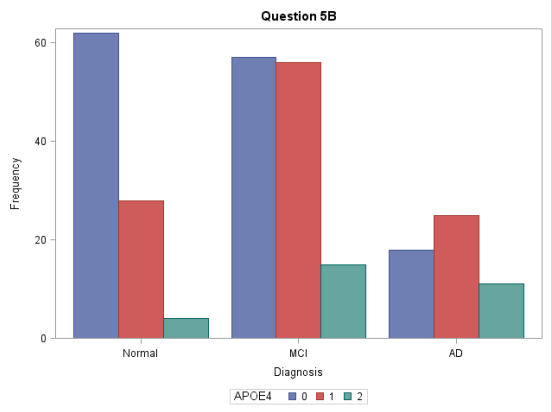
   (b) Your macro should take one parameter, which is the categorical variable that defines the colors.

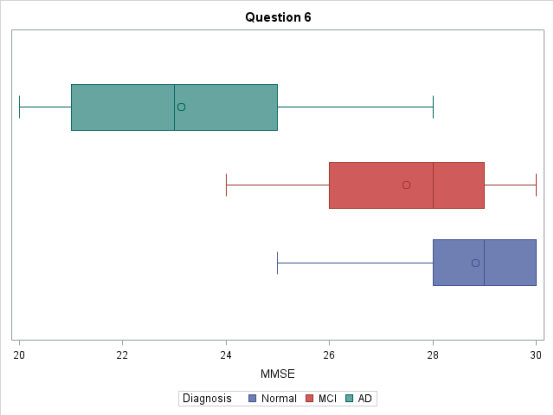
   (c) Make sure the title of your plot is modified to identify the categorical variable used.

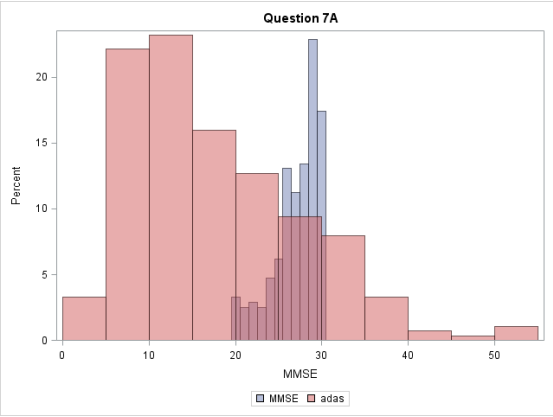
   (d) Test your macro using `gender`, `APOE4`, and `dx`.

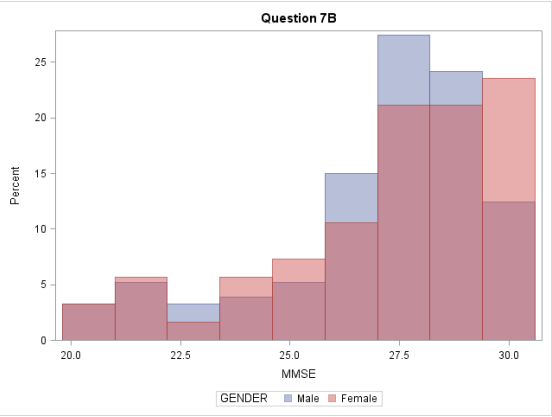
Example plots to verify your work:

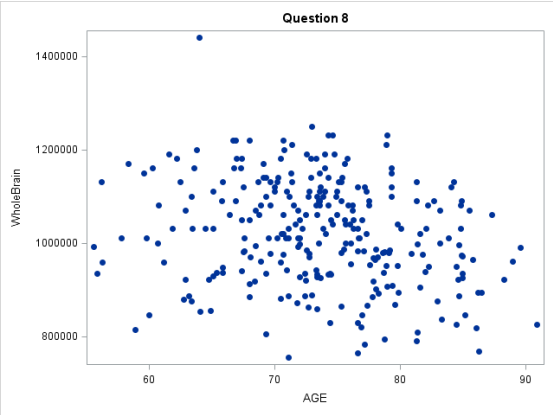


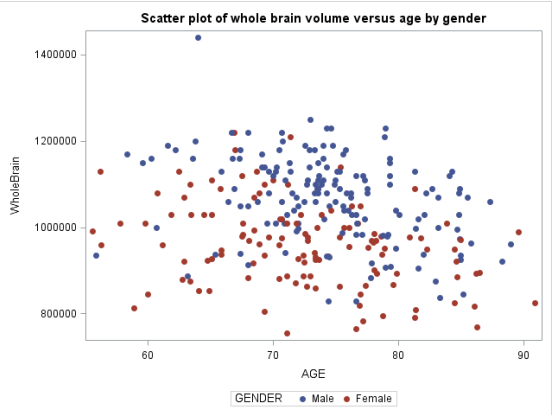


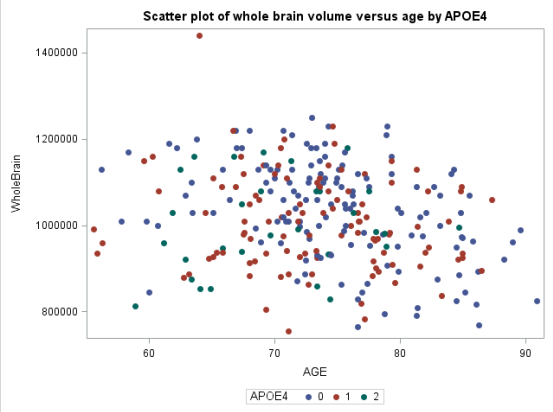


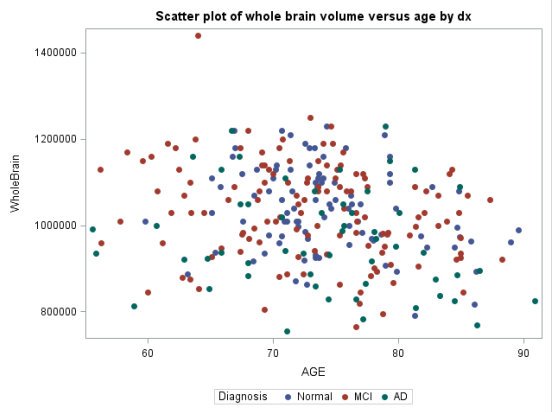





Question 8
WholeBrain
1400000
1200000
1000000
800000
60
70
80
90
AGE


Scatter plot of whole brain volume versus age by gender
WholeBrain
1400000
1200000
1000000
800000
60
70
80
90
AGE
GENDER Male Female


Scatter plot of whole brain volume versus age by APOE4
WholeBrain
1400000
1200000
1000000
800000
60
70
80
90
AGE
APOE4 0 1 2


Scatter plot of whole brain volume versus age by dx
WholeBrain
1400000
1200000
1000000
800000
60
70
80
90
AGE
Diagnosis Normal MCI AD