IST3110 - Week 4 Lab Session 4

- The data file is named `mammo.txt` and contains the following variables related to mammography results. Variables, type of variables and range of the variables are:
  - BI-RADS (integer, [0,6])
  - Age (integer, [18,96])
  - Shape (integer, [1,4])
  - Margin (integer, [1,5])
  - Density (integer, [1,4])
  - Severity (integer, [0, 1])

IST3110 - Week 4 Lab Session 4

1. Read the data set from the file `mammo.txt`, look at the first few lines of the data, structure of the data and summarize it.

2. Make frequency tables by using the functions `apply` and `table`.

3. With which values did the missing values in the data is coded? Turn them into NA.

4. Replace missing values for `Age` with mean of age and replace the missing values of other variables with the median of each variable. (**Hint:** You can use `ifelse()` function for each variable)

IST3110 - Week 4 Lab Session 4

5. Convert all the variables except `Age` to factors (categorical variables).

6. Make a new variable called `AgeGroups` by using the `cut()` function on `Age` and group this variable as (17, 40], (40, 60], (60, 96].

7. Make a crosstab and a proportional crosstab to reveal the relation between age groups and severity.