STAT2005 Programming Languages for Statistics
Assignment 3
Due: 4 April 2025

1. Write a function named `questionnaire()` and ask each respondent three questions, namely.

Question 1: What's your name?

Question 2: What's your age?

Question 3: What's your gender?

Your function should be interactive, i.e., you should display these questions on screen and ask the users to input their responds. Input verifications are required for question 2 and 3. In particular, we require the user to input integer value for question 2 and either "M" or "F" for question 3 ("m" or "f" are also accepted). If the inputs from the user are not acceptable, display an appropriate message and ask that question again. All answers should be stored in a data frame with three variables

```
Name Age Gender
```

as an output of this function. When a set of three questions is completed, the program should display a message

```
New entry? <Y/N>:
```

If the input is "Y", a new respondent is expected to answer those three questions with their answers stored as a new observation of the output data frame. If the input is "N", the function is terminated with the following tasks.

i. Summarize all the inputs by displaying a message with the following information.

```
 The average age is 23.5 from 3 respondents.
```

ii. Produce a pie chart to show the gender distribution.

A detailed example of the operations of the function `questionnaire()` is shown below.

```
> data <- questionnaire()
Please respond to the following three questions.

Question 1: What's your name? : Tom
Question 2: What's your age? <positive integer required>: 20
Question 3: What's your gender? <M/F>: M
New entry? <Y/N>: y
Question 1: What's your name? : May
Question 2: What's your age? <positive integer required>: 23.5
Please input positive integer.
```

Question 2: What's your age? <positive integer required>: 23

Question 3: What's your gender? <M/F>: f

New entry? <Y/N>: Y

Question 1: What's your name? : John

Question 2: What's your age? <positive integer required>: eighteen

Please input positive integer.

Question 2: What's your age? <positive integer required>: -18

Please input positive integer.

Question 2: What's your age? <positive integer required>: 18

Question 3: What's your gender? <M/F>: male

Please input M or F.

Question 3: What's your gender? <M/F>: m

New entry? <Y/N>: No

Please input Y or N.

New entry? <Y/N>: n> data
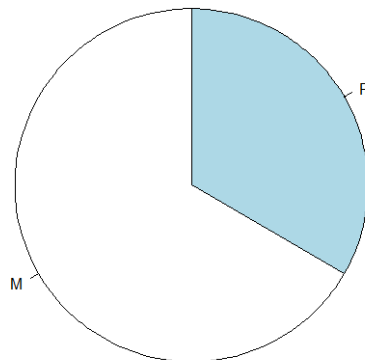
	Name	Age	Gender
1	Tom	20	M
2	May	23	F
3	John	18	M

2. Suppose that there are two TV channels competing with each other in a city. Each of the individual TV set could take three possible statuses, namely (i) Channel 1 (state 1), (ii) Channel 2 (state 2), and (iii) Power-off (state 3). Every hour, each TV set remains in the same status or switches to a different status according to the transition probability matrix

$$T = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.3 & 0.3 \end{bmatrix}.$$

(a) Compute the 2-step transition probability matrix.

(b) If a TV is on Channel 1 at 9am, what is the probability that it will be on Channel 1 at 3pm on the same day?

(c) Consider two mutually exclusive promotion strategies for Channel 2. The first strategy could permanently increase the probability of switching from Channel 1 to Channel 2 from 0.2 to 0.25, while keeping the probability of switching off the TV unchanged. The second strategy could permanently increase the probability of switching on the TV to watch Channel 2 from 0.3 to 0.35, while keeping the probability of switching on to watch Channel 1 unchanged. Suppose that the costs of promotion are the same, which strategy is more preferable for Channel 2?

3. Which of the following SAS statements are incorrect? For each incorrect statement, point out the mistake(s) and suggest a correction.

```
a) CARDS3;

b) DATA;

c) DATA IN.NEW.RECORD;

d) LENGTH X 1 A B $ 12;

e) ***  COMMENT ***;

f) INPUT _A C2 - C10 $2 L 5-9. */ L IS NUMERIC /*;

g) INPUT @5 (X2 - X5) COMMA4 ;

h) INPUT X X X;

i) INFILE C:\DATA.TXT

j) LIBNAME A.B 'C:\";
```

4. Below are three personnel records of a company.

Staff ID	Birthday	Monthly Salary
A123	4 March 1989	8,600
A037	23 June 1957	21,450
M015	19 September 1977	17,500

Staff ID number has the format that it starts with a character followed by three digits. The first character is the abbreviation of the department ("A" for Accounting, and "M" for Marketing, etc.).

Suppose the data are as given in (i) and (ii) below. Write SAS programs to read the data in (i) and (ii) to create a temporary SAS data file, called `PERSONNEL`, which contains five variables, namely `ID`, `DEPT`, `BIRTHDAY`, `YEAR` and `SALARY`. `YEAR` is the year of birth. Variables `ID` and `DEPT` are character variables and all other variables are numeric. The first record should have `ID = "A123", DEPT = "A", BIRTHDAY = 10655, YEAR = 1989, SALARY = 8600`. You are not allowed to use assignment statement to create any of the above five variables.

(i)

```
A123    4Mar1989     8,60000

A037   23Jun1957    21,45000

M015   19Sep1977    17,50000
```


(ii)

```
A123   4Mar1989   8,6,00

* * * * * * * * * * * * * *

    A037 23Jun1957   21,450

* * * * * * * * * * * * *

 M015 19Sep1977$17,500

* * * * * * * * * *
```


You should submit two files `asg3.r` and `asg3.sas` via Blackboard, which contains respectively all the R and SAS codes you use to finish this assignment. The codes should be commented as clearly as possible. Written work (if any) should also be attached.