## CS544 Module 3 Assignment

### General Rules for Homework Assignments

- You must work on your assignments individually. You are not allowed to copy the answers from the others.

- Each assignment has a strict deadline. If there is a delay, you must be in touch with the instructor. Late submissions without reasons will result in a grade deduction.

- When the term lastName is referenced in an assignment, please replace it with your last name.

### Part 1) 30 points

Read in data from the attached file, Aids2.csv. The data describe patients diagnosed with AIDS in Australia before 1 July 1991.Here is the description of some of the data columns:

state: Grouped state of origin: "NSW "includes ACT and "other" is WA, SA, NT and TAS
Sex: Sex of patient.
diag: date of diagnosis.
death: date of death or end of observation.
status: "A" (alive) or "D" (dead) at the end of observation.
T.categ: Reported transmission category.
age: Age (years) at diagnosis.

Do not use explicit loops for any calculations. Do not hard code in the solution. The solution should work for any denominations.
  a) How many female and male patients are there in the data? Visualize it.
  b) How many percentages of patients were alive and dead at the end of observation? Visualize it.
  c) How many percentages of male and female patients were alive and dead at the end of observation?
  d) What are the female and male patients' median age at diagnosis? Draw two histograms to show the distributions of age at diagnosis for females and males.

### Part 2) 30 points

Suppose that a group of people voted on whether to support a policy. Consider the two-way summarized data shown below showing the voting result of men and women.

| Gender | Voting Result | | |
|--------|-----|-----|---------|
|        | Yes | No  | Abstain |
| Men    | 36  | 10  | 4       |
| Women  | 24  | 30  | 6       |

a) Create a matrix for the above data. Set row names and column names for the matrix.

b) Add the dimension variables Gender and Vote to the data. Show the marginal distributions for the Gender and the Voting Result. Show the result of adding margins to the data.

c) Show the proportional data separately for Gender and Voting result. Interpret the results.

d) Using appropriate colors, show the mosaic plot for the data. Add legend to the plot.

e) Show the barplot for Gender and Voting Result separately with the bars side by side. Add legend to the plot.

**Part 3) 20 points**
Use the midsize (UsingR) dataset.
a) Show the pair wise plots for all the variables.
b) Provide at least 4 interpretations of the results.

**Part 4) 20 points**
Use the MLBattend (UsingR) dataset.
a) Extract the wins for the teams BAL, BOS, DET, LA, PHI into the respective vectors.
b) Create a data frame of five columns using these vectors. Use the team names for the columns
c) Show the boxplot of the data frame.
d) Provide at least 5 interpretations of the results.

**Submission:**
Upload your result file to the Assignments section of Blackboard.

Provide all R code in a single file, CS544_A3_lastName.R. Clearly mark each subpart of each question and add appropriate comments.

If you need to submit more than one files, create a folder, CS544_A3_lastName and place all files in this folder. Archive the folder (CS544_A3_lastName.zip).