STATS 2DA3 Winter 2023

ASSIGNMENT 1

Submit through Avenue to Learn.

Due before 11pm on Friday, January 27th.

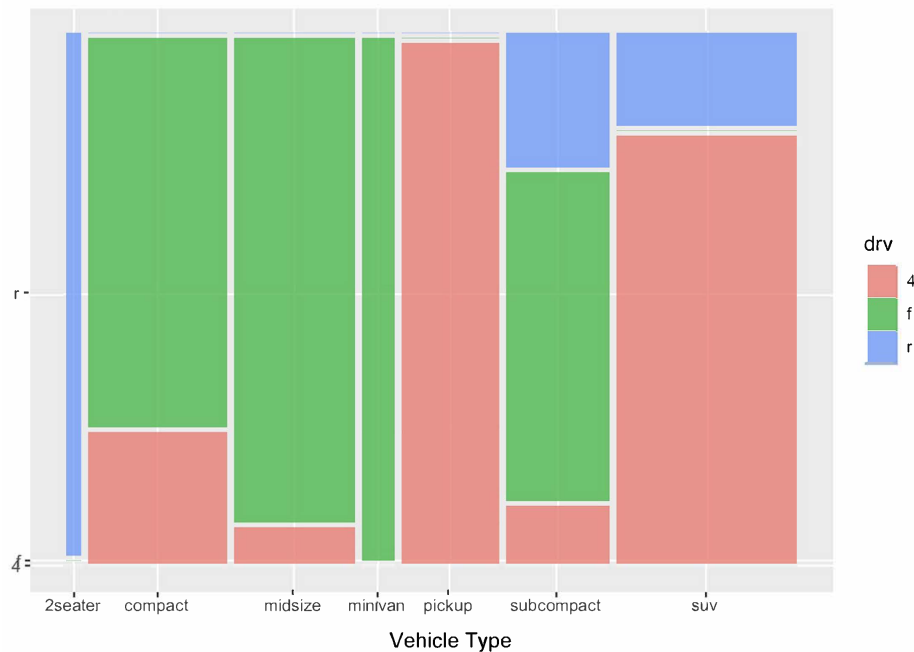**Your assignment must conform to the Assignment Standards listed below.**

Assignments submitted up to 24 hours late will incur a 30% penalty.
Assignments submitted more than 24 hours late will receive a zero grade.

Answer all 4 questions. Not all questions carry equal marks.
All graphs must be labelled (including axes).

1. **(7 MARKS)** Using the **mpg** dataset in R, answer the following questions:

   (a) How many rows and columns are in the dataset?

   (b) What does the "displ" variable describe?

   (c) Which variables are categorical and which are continuous?

   (d) Using the **ggplot** function, make a scatterplot of "displ" against "hwy" (putting "displ" on the x-axis). Make the data points green.

2. **(5 MARKS)** Using the **Galton** dataset in HistData library, carry out the following tasks:

   (a) Make a histogram displaying "child" height.

   (b) Make a histogram displaying "parent" height.

   (c) Display both graphs in one image (hint: you could use the grid.arrange function, but you need to know what library contains this function)

3. **(3 MARKS)** Consider the plot below; it displays information on Vehicle Type and on the associated drive train. There are 3 different types of drivetrain :
   4 = four wheel drive, f = front wheel drive, r = rear wheel drive.

   (a) Which Vehicle Type has the most observations associated with it in the dataset?

   (b) For "subcompact" vehicles, what is the majority drive train type?

   (c) For "midsize" vehicles, which of the 3 drive train types occurs least often?

4. **(5 MARKS)** For the **Titanic** dataset in R, perform the following tasks:

   (a) Create a Double Decker plot, displaying "Age" as a function of "Class" and "Sex" (Class and Sex should be on the x-axis).

   (b) List 2 interesting facts regarding the observations in the "Crew" Class that the graph shows.

   (c) The majority of people on board came from which "Class"?

Assignment Standards

- LATEX is strongly recommended but not strictly required. The use of Markdown in R studio is also recommended.

- Submit your assignment as one **.pdf document**. **All R code should be included and organized either at the end of the assignment or inline (if using R Markdown).**

- Eleven-point font (times or similar) must be used with 1.5 line spacing and margins of at least 1 inch all around.

- Do not include a title page. The title and your name should be printed at the top of the first page.

- Various tools, including publicly available internet tools, may be used by the instructor to check the originality of submitted work.