**CSC/CIS 435 LAB 01 Points: 20**

Basic R & Visualization

Your Team Member names:

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1993 New Car Data

The 93CARS dataset contains information on 93 new cars for the 1993 model year. Measures given include price, mpg ratings, engine size, body size, and indicators of features. The 17 variables in the dataset offer sufficient variety to illustrate a broad range of statistical techniques.

Each data case starts with the COMPANY (e.g., Acura, Audi, BMW, Buick,...), MODEL (e.g., Caprice, 90, Accord,...), and a TYPE (Small, Sporty, Compact, Midsize, Large, Van). The other 15 variables are all numeric (**not necessarily all continuous type data**). The EPA fuel efficiency ratings are given as both CITY and HIGHWAY miles per gallon (MPG).

Several measures reflect relative size and power of the standard engine. These include the number of CYLINDERS, engine displacement SIZE (in liters), and maximum HORSEPOWER.

Other variables note the presence of standard AIR BAGS (driver or passenger), the type of DRIVETRAIN (front-wheel, rear-wheel, or all-wheel), and an option for a MANUAL transmission. A final variable categorizes the manufacturer as domestic (U.S.) or foreign, although this distinction is becoming less and less clear.

Purpose:

This lab is designed for you to perform the basic R tasks and to demonstrate the Visualization. After completing the tasks in this lab, you should be able to:

* Identify the data types (Nominal, Ordinal, or Continuous)
* Change the data types of input variables
* Recode variables
* Create new columns
* Explore variables
* Create summary table
* Analyze the distribution of a continuous variable
* Create various graphs
* Analyze graphs, interpret the results, and provide your insight

Perform the following tasks

**Note: You have to create a new R script called Lab01. All your code has to be written in this script with appropriate comment for each step.**

1. Download the Lab01 from BB.
2. Read the dataset “Lab1\_93CARS Data.csv”.

Look at the data types currently present for the columns given in the table below in the R file, if you feel they need to be changed to another datatype, do so. Fill-in the appropriate data types for columns in the Table below (Categorical (Nominal, Ordinal), Continuous).

|  |  |  |  |
| --- | --- | --- | --- |
| Column | Details | Value/(Units) | Data Type |
| B | MODEL | Model name |  |
| C | TYPE | Car Type: Small, Sporty, Compact, Midsize, Large |  |
| D | MID-PRICE | $1,000s |  |
| E | CITY-MPG | Miles per gallon |  |
| G | AIRBAGS | 0 = none; 1 = driver only; 2 = driver and passenger |  |
| H | DRIVETRAIN | 0 = rear wheel drive 1 = front wheel drive 2 = all wheel drive |  |
| L | TRANS | Manual transmission available - 0 = No, 1 = Yes |  |
| M | FUEL-CAP | Fuel tank capacity in gallons |  |
| O | LENGTH | In inches |  |
| P | WEIGHT | In pounds |  |
| Q | MANUFACTURER | 0 = Foreign; 1 = Domestic |  |

1. Now use the data types listed in Table1 and make necessary changes to the data types to input variables in R as per Table1.
2. Recode the following variables: **AIRBAGS**, **DRIVETRAIN** and **TRANS** with the value information given in Table**.** Eg. **Airbags** should have the values **none**/**driver only**/**driver and passenger** depending on the code for the airbags. **Copy & paste a screen shot of the result.**

**Hint: use recode() function, this function is in the dplyr package. For more information see the following link:**

<https://dplyr.tidyverse.org/reference/recode.html>

1. Create a new column named **“Fuel Economy”** with values **“High”** and **“Low to Medium”** using the following formula; If HGW-MPG > 30 Then “High” else “Low to Medium”. After creating the new column, answer the following question.

How many cars have High Fuel Economy? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Create a summary table with **Mean** of **MID-PRICE**, **CITY-MPG**, and **HORSEPWR** grouped by **Car Type**. Fill in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
|  | MID-PRICE | CITY-MPG | HORSEPWR |
| TYPE | Mean | Mean | Mean |
| Compact |  |  |  |
| Large |  |  |  |
| Midsize |  |  |  |
| Small |  |  |  |
| Sporty |  |  |  |
| Van |  |  |  |

1. Create a pie graph to find the percentage of cars that do not have **Airbags**? **Copy & paste your chart below and interpret the graph.**
2. Create an appropriate chart to answer the following question. What is the average **Fuel Capacity** for a **Rear wheel drive (DRIVETRAIN)**? **Copy & paste your chart below and interpret the chart**
3. Create an appropriate chart to explain the relationship between the **Engine Size** and **Horsepower**. **Copy and** **Paste the chart below and interpret the results.**
4. Create a graph to identify outliers in **Length** variable. Identify the record ID and the value of the outliers. **Copy and Paste your chart below**
5. Create Stem and Leaf plot for **Fuel Capacity** variable. Identify the number of cars that have fuel capacity of 23 gallons? **Copy & Paste your chart below**
6. Create a Mosaic plot to understand the relation between variables **Airbags** and **Type** of car. Interpret the graph and list three unique observations you can make. (For Example: A high percentage of small cars (about 75%) do not have any airbags.)
7. For the final graph, create a chart of your choice with a minimum of three variables. **Copy and paste the chart below and interpret the chart.**