TUHH Summer Course Manipulating Data Frames

Load the file '**Cars93**' from the **MASS** package into your R session.

One way:

```
library("MASS") # load MASS package (might have to install it first)
attach(Cars93) # move Cars93 to #2 position in search space
```

Note: Is good to be aware what the **attach()** function (or command) does….and how the use of **attach()** at this point will affect your approach (syntax) in the questions that follow…..

1) Issue the following commands, one at a time:

```
Cars93
head(Cars93)
names(Cars93)
str(Cars93)
nrow(Cars93)
ncol(Cars93)
```

How many different data types can you identify?

How many rows of data are there in Cars93?

How many columns of data?

2) Print out (**note:** 'print' means print to the screen or list out) the first 8 columns of data in Cars93, but only for the first 28 rows. Look at the data. Does the Cars93 data frame appear to be sorted? By what variable? In what order?

3) Create a separate data frame for the first 8 columns of data in Cars93, but only for the first 28 rows. Name the data frame "my.car.data". Do not attach the data set. We will use this data frame later in the exercises.

4) Create separate data frames for the following Manufacturers and assign them to their own data frame named: Chevrolet; Dodge; and Nissan. How many records are there in each?

5) What Make (Make is same as Manufacturer) and Model of car gets the best gas mileage on the Highway? In the city?

6) Print out the entire 'Max.price' column from Cars93. Why do you think it lists out horizontally instead of vertically?

7) What is the object type (that is, 'class') of this 'Max.price' column? What is the object type (that is, 'class') of the 6th row (or observation) of Cars93? How do you explain this?

Use the 'my.car.data' data frame to answer questions #8 to #13.

Detach Cars93 and attach my.car.data before proceeding

8) Print out 'my.car.data' but only include rows 1, 3, 5 and 6 and the price columns (#4, 5 and 6).

9) Do it again, but show the rows in their reverse order (6, then 5, then 3, then 1) and show the columns switched left to right (6 then 5 then 4) in order.

10) Print out exactly the first ten records of 'my.car.data' ***in reverse order***, and ***randomly select and rearrange any 6 (but exactly 6 only) of the 8 total columns***. Do it ten times while you watch the results appear in R. How many times does the 'Manufacturer' column appear as the first column? Try to explain what is going on here.

11) Print out 'my.car.data' sorted Z to A (e.g. high to low) by Manufacturer, but sorted 'worst to best' (e.g. low to high) in terms of MPG.city. Which Chevrolet Make and Model gets the **worst** MPG in the city?

12) Select all of the car records (e.g. rows) in my.car.data that have an MPG in the city greater than at least 20 MPG. Of only these exclusive records, which one (Make and model) has the **highest** maximum price (MAX.price)?

13) Which Make and Model cars in my.car.data have an MPG on the highway greater than 26 MPG but which cost more than the mean of the Price of all of the cars in my.car.data?

14) With your 'my.car.data' data frame still attached, issue the following command:

```
attach(my.car.data)
```

What happens? Why do you think R responds this way? Look up the attach command by issuing this command:

```
?attach
```