**IST 687 PREP EXERCISE 09**

**Student Name: Bhavish Kumar**

**Prep Exercise No: 09**

**Date Due: 30th October 2019**

1. **Getting Ready: Loading and Verifying the Titanic Dataset**
   1. An R dataset containing the titanic data is available on the Blackboard site. Download it to your computer and use the *load()* command to bring it into your RStudio environment.

**load("titanic.raw.rdata")**

* 1. If you are having trouble using the *load()* command, make sure you downloaded the titanic data to your working directory. You can check what is your working directory in R using the code below:

**getwd()**

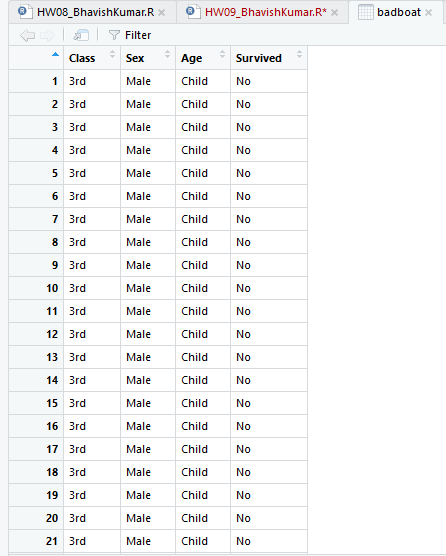
### Directory was set to working directory before loading

* 1. Load the imported data into a dataframe called “badboat”.

**badboat <- data.frame(titanic.raw)**

* 1. Run the *View()* command to verify that the dataframe has been loaded correctly and place a screenshot of the dataset below (do not worry about screenshotting the entire dataframe, a portion will do).

View(badboat)



1. **Understanding Terminology that will be used in this PE and HW.**
   1. In a paragraph or two explain the concept of a sparse matrix and how to identify whether or not data is a sparse matrix.

A matrix is a data structure that has rows and columns like a datframe. A matrix is called sparse if only very few columns have occur frequently in all the rows, i.e. only few columns have a value 1 under a large proportion of rows or have high support. In other words if there are more than 50% 0s in the entire matrix then it is called a Sparse Matrix.

* 1. In a few sentences explain what a dense matrix is and its defining feature.

A dense matrix is the opposite of a sparse matrix where the number of 1s in the entire matrix is more than the number of 0s, i.e. a large number of columns have value 1 under them.

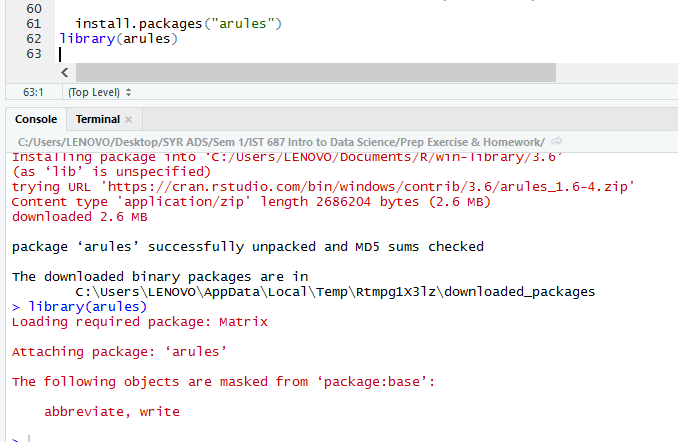
* 1. Explain contingency tables and how you would go about creating one in RStudio (if needed, you can google “contingency table in r”, there are many resources on the web). How would you go about reading /understanding a contingency table?

A contingency table is a confusion matrix that shows the Joint Distribution of categorical variables. The table() function with 2 arguments can be used to create a contingency table in R. For example table(cars$type, cars$origin) will yield a contingency table with the first argument cars$type being the row variable and second argument cars$origin will be the column variable.

1. **Loading Necessary Packages.**
   1. In the homework portion of this week’s lesson you will need two packages; *arules* and *arulesViz*.
   2. Install both packages and check/update them using the *library()* command to ensure the packages exist. Paste a screenshot of the code and respective output below.

**install.packages("arules")**

**library(arules)**



**install.packages("arulesViz")**

**library(arulesViz)**

