**IST687 – aRULES Lab: The Titanic**

Download the titanic dataset from blackboard.

Note is it not a cvs file, but rather, an R dataset. So, to load the data, one would do:

load("….whatever…./Downloads/titanic.raw.rdata")

#You need to look at titanic.raw (the name of the R dataset)

t <- titanic.raw

Now that you have the datafile, do some descriptive statistics, getting some extra practice using R.

**Step 1: Descriptive Stats**

1. Compute the percentage of people that survived.
2. Compute the percentage of people that were children
3. Compute the percentage of people that were female
4. Finally, compute the percentage of people that were in first class

**Step 2: More Descriptive Stats**

1. What percentage of children survived?
2. What percentage of female survived?
3. What percentage of first class people survived?
4. What percentage of 3rd class people survived?

**Step 3: Writing a Function**

1. Write a function that returns the a new dataframe of people that satisfy the specified criteria of sex, age, class and survived as parameters
2. Write a function, using the previous function, that calculates the percentage (who lives, who dies) for a specified (parameters) of age, class and sex.
3. Use the function to compare age & 3rd class male survival rates
4. Use the function to compare age & 1st class female survival rates

**Step 4: Use aRules**

1. Use arules to calculate some rules (clusters) for the titanic dataset
2. Visualize the results
3. Pick the 3 most interesting & useful rules.
4. How does this compare to the descriptive analysis we did on the same dataset?