## Group Project – Instructions

If you use your own laptop:

* You just need to put the train.csv and test.csv into under the work directory
* Then run the code:
  + Install all packages
  + Load training set and testing set and combine to new dataset - full
  + Plot the impact of Pclass
  + Extract the title and plot the impact of passenger title
  + Plot the impact of sex and further investigate the effect by combining Pclass with gender
  + Draw graphs of the influences of number of siblings and parents
  + Create a new variable – family size
  + Plot the impacts of cabin and ticket account
  + Predict the missing values of embarked, fare and age; then explore their impacts
  + Create new variable mother and investigate its effect
  + Split full dataset back into training set and testing set
  + Investigate correlations between attributes
  + Test accuracy of different algorithms
  + Improve the result by applying a Box-Cox
  + Tuning the SVM algorithm
  + Choose Support Vector Machines (SVM) to build the model
* A new file named “submission.csv” will be generated under the same work directory

If you use the PC in Computer Lab 1:

* You need to download the R 3.2.5 version, this is because the school’s PC did not support “car” package. Here is the link: <https://cran.rstudio.com/bin/windows/base/old/3.2.5/>
* The rest instruction is the same as you run in your own laptop.