# Predicting Survival On the Titanic Using R and Machine Learning Algorithm

*Linfeng Zhou*

*October 13, 2015*

Group Member: Tianyi Gu, Xiaoge Wu, Yi Zhang

## Problem Addressed

The sinking of the RMS Titanic is one of the most infamous shipwrecks in history. Our team are planning to figure out what sorts of people were likely to survive. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others, such as women, children, and the upper-class.

## Description of Dataset

Our dataset is from Kaggle Competition:

```
setwd("/Users/luke/dropbox/kaggle/titanic/")
train<-read.csv("train.csv",header=TRUE)
names(train)
```

```
##  [1] "PassengerId" "Survived"    "Pclass"      "Name"        "Sex"
##  [6] "Age"         "SibSp"       "Parch"       "Ticket"      "Fare"
## [11] "Cabin"       "Embarked"
```

The name of features are:

- survival: Survival(0 = No; 1 = Yes)
- pclass:Passenger Class(1 = 1st; 2 = 2nd; 3 = 3rd)
- name:Name
- sex:Sex
- age:Age
- sibsp: Number of Siblings/Spouses Aboard
- parch: Number of Parents/Children Aboard
- ticket: Ticket Number
- fare: Passenger Fare
- cabin: Cabin
- embarked Port of Embarkation(C = Cherbourg; Q = Queenstown; S = Southampton)

## Exploration Process

Our final project will include following six steps, to make the analysis exhausted.

- Getting and Cleaning Data

- Exploratory Data Analysis
- Features Selection
- Model Training
- Model Validation
- Conclusion