HW 6

QBA200a

January 17, 2017

In this homework we will analyze a Titanic data set from Kaggle to see if we can start to identify which variables would be helpful in predicting whether a passenger survived or not. To see a description of the columns visit

https://www.kaggle.com/c/titanic/data.

The data set is contained in the file "Titanic.csv", which has been uploaded to Blackboard. All questions are worth 4 points.

1. Read in the data into R. Convert the Pclass column to a factor and the Name column to a character.2.
2. What is the overall survival percentage? What is the survival percentage broken down by sex? By Pclass? By both sex and Pclass?
3. What is the average age of passengers with the Miss title in their name? Mrs. title?
4. Create four histograms (All in one plot) showing the distribution of the ages of people with each title. The four titles are Miss, Mr., Mrs., and Master.
5. For each of these four histograms, further partition the data by whether the passengers survived or not. So you have four plots (one for each title) that each have two histograms; one for the distribution of ages that survived and one for the distribution of ages that did not survive.