**Project Name:**

Titanic - Who was likely to survive?

**Team**

Gassan Soukaev (soukaev2)

Hanna Loboda (hloboda2)

Hao Tang (haot3)

You Li (youl3)

**Description of dataset**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Definition** | **Key** |
| survival | Survival | 0 = No, 1 = Yes |
| pclass | Ticket class | 1 = 1st, 2 = 2nd, 3 = 3rd |
| sex | Sex |  |
| Age | Age in years |  |
| sibsp | # of siblings / spouses aboard the Titanic |  |
| parch | # of parents / children aboard the Titanic |  |
| ticket | Ticket number |  |
| fare | Passenger fare |  |
| cabin | Cabin number |  |
| embarked | Port of Embarkation | C = Cherbourg, Q = Queenstown, S = Southampton |

**Background information**

“The sinking of the RMS Titanic is one of the most infamous shipwrecks in history.  On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships.

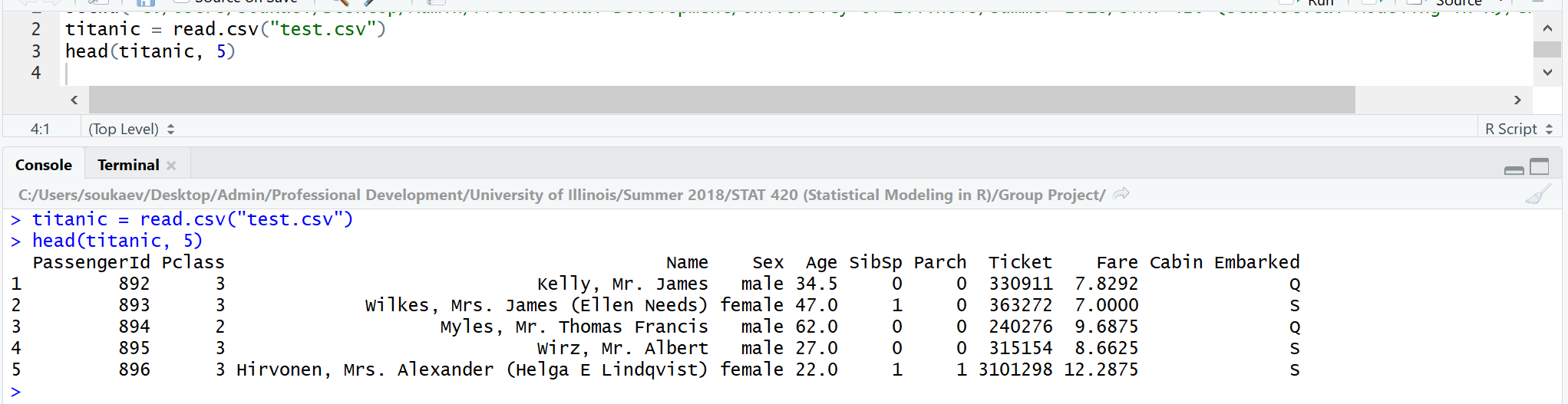
One of the reasons that the shipwreck led to such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others, such as women, children, and the upper-class.”

This dataset is available at [kaggle](https://www.kaggle.com/c/titanic)

**Personal interest**

We think this data provides a good mix of categorical (such as sex, ticket class) and numeric data (such as age or ticket price) which will allow us to leverage linear / logistic regression to determine who was more likely to survive this tragedy.

**R data load**

****