

Titanic Dataset

Sociology 312, University of Oregon
Prof. Gullickson

Overview

The titanic data contain information on all 1,309 passengers aboard the Titanic. The data do not include information about the crew. The data primarily come from the online database, Encyclopedia Titanica.

Table 1 provides summary statistics for all the quantitative variables in the movie data.

Table 1: Summary statistics for all quantitative variables in the titanic dataset

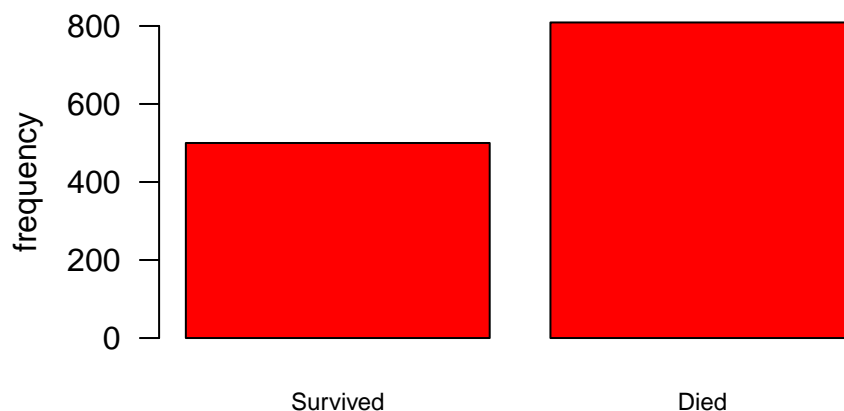
Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
age	1309	29.738	13.886	0.167	21.000	28.000	38.000	80.000
fare	1309	33.276	51.744	0.000	7.896	14.454	31.275	512.329
family	1309	0.884	1.584	0	0	0	1	10

Variable Descriptions

survival

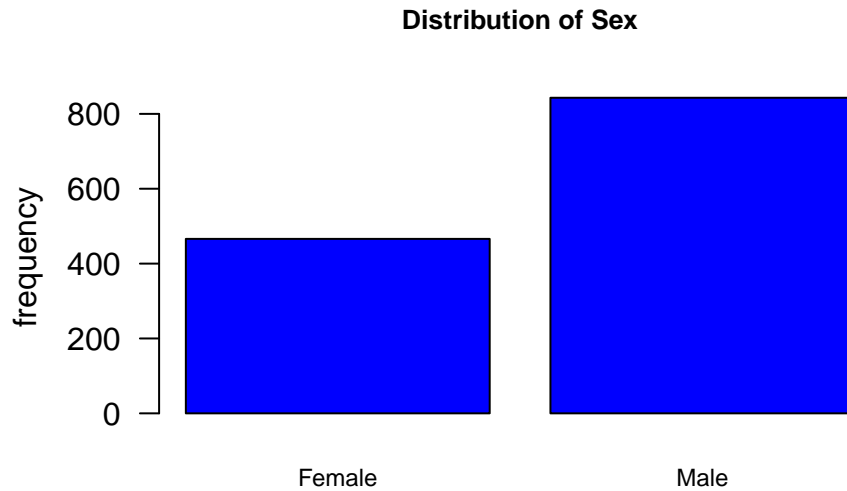
Whether or not the passenger survived. The original entry in the Encyclopedia Titanica, goes as far as giving the unique number id assigned to each body (for those who died) and the boat number (of those who survived). We will leave out those morbid details and just focus on whether a person survived or died.

Distribution of Survival



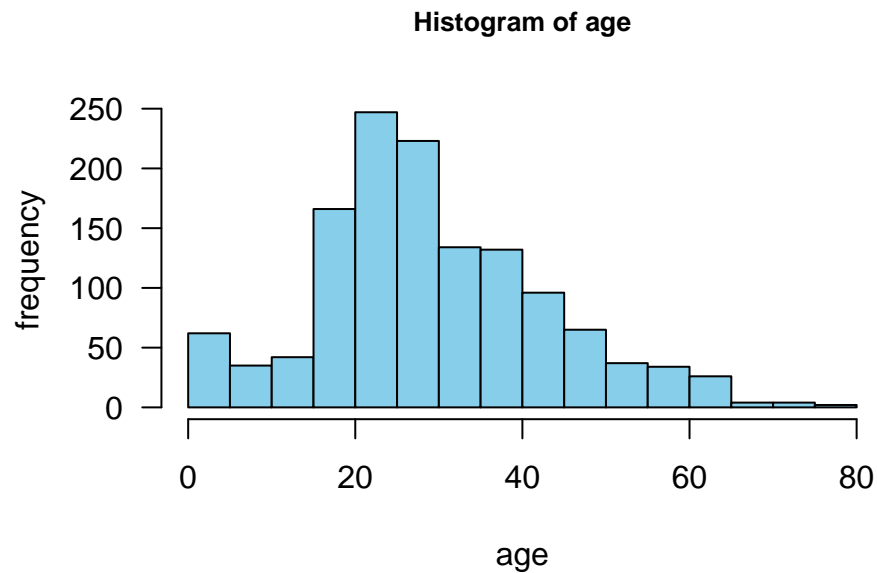
sex

The reported sex of the passenger. This is a straightforward nominal variable. There were a lot more male passengers than female passengers, although most of this discrepancy is among third class passengers who were largely male labor migrants.



age

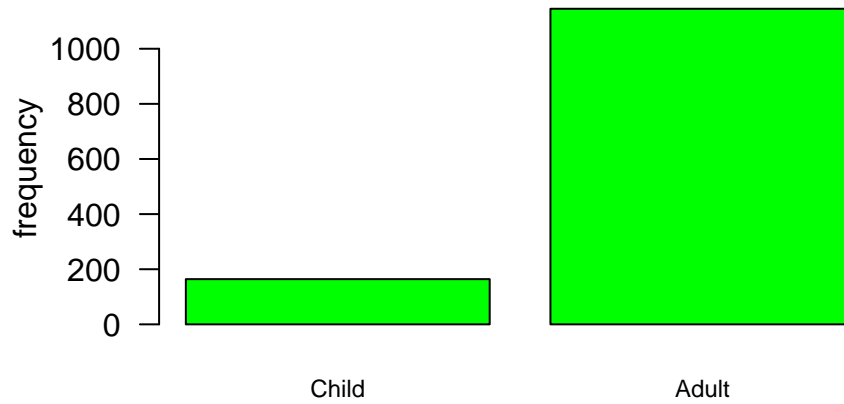
The age of the person. This variable is reported in whole numbers for those over one year old and as a decimal (based on months of age) for infants under a year of age. The age of passengers is not known in all cases, age is missing or about 20% of passengers. I have used advanced techniques to impute valid values for all missing values.



agegroup

A categorical variable indicating whether the person was an adult or a child. I have constructed this variable from the age variable. The cutoff for adults is sixteen years of age.

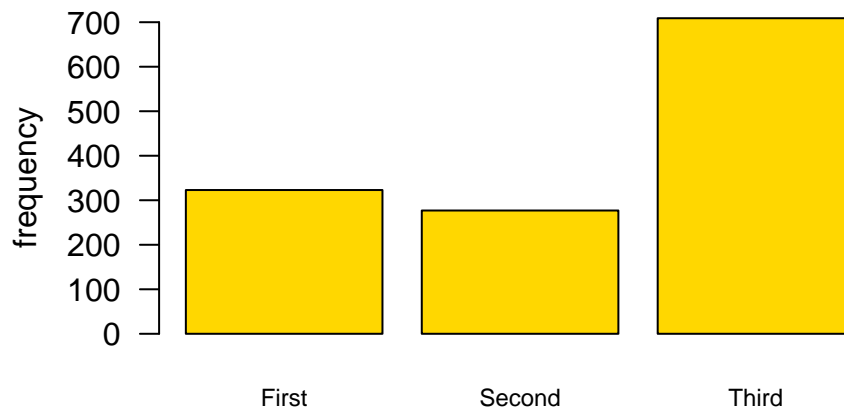
Distribution of age groups



pclass

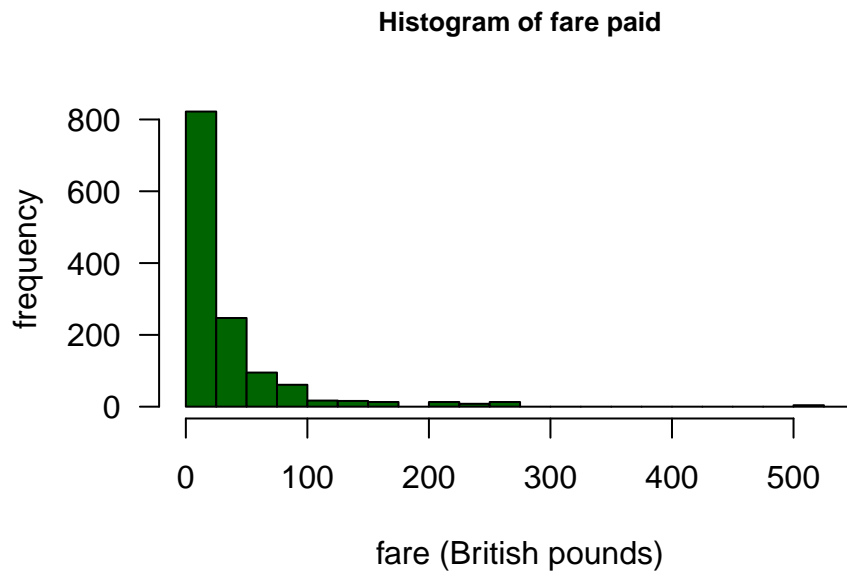
The passenger class of the passenger. There were three passenger classes: First, second, and third (also known as steerage). To give some pop culture references, Rose was first class, and Jack was third class. Most of the passengers were in third class.

Distribution of passenger class



fare

The fare paid for this passenger's ticket. This is measured in British pounds. Obviously, this will vary substantially by passenger class, but also varies within each class. This variable is heavily right skewed.



family

The number of family members traveling with the passenger. These family members can either be parents, spouses, siblings, or children.

