Categorical Data

1 Introduction

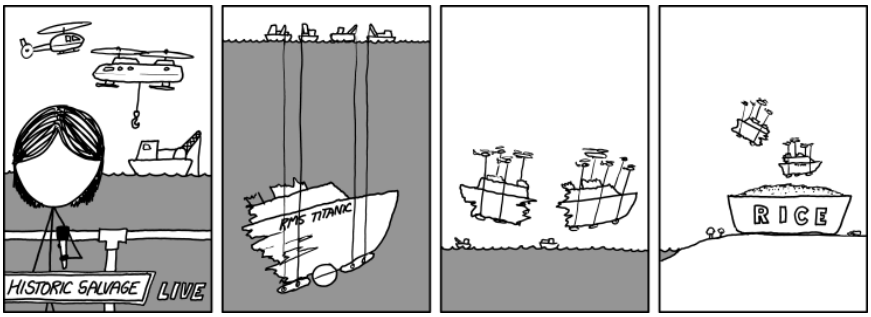

Figure 1: xkcd.com

In today's session, you'll take a look at a publicly available data set containing data on the Titanic and its passengers - the Titanic data set. We want to visualize the data describing the people aboard the vessel and find out more about their fates. To do this, we propose you rely on jupyter notebooks.

Column	Description
PassengerId	Passenger Id
Survived	Survival (0 = No; 1 = Yes)
Pclass	Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)
Name	Passenger Name
Sex	Passenger Sex
Age	Passenger Age
SibSp	Number of Siblings/Spouses Aboard
Parch	Number of Parents/Children Aboard
Ticket	Ticket Information
Fare	Passenger Fare (Pound sterling - GBP)
Cabin	Cabin Information
Embarked	Point of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)

Table 1: Meta Information for Titanic data set

simon.schindler@fh-salzburg.ac.at
maximilian.schirl@fh-salzburg.ac.at
martin.uray@fh-salzburg.ac.at

2 Lab Assignments

Data Loading and Selection Methods ✓

To complete the first assignment, you are asked to apply your knowledge in Python and pandas to the Titanic data set. Table 1 contains some meta information on the data set, providing some additional context for you. After you got to know more about the Titanic data set, answer the following questions using pandas data selection methods:

- How many passengers have embarked in Southampton?
- What was the price of the most expensive ticket? Find out the name(s) of the passenger(s) who bought them.
- Who were the youngest passengers? Find out their name(s) and age(s).
- Who was/were the oldest passenger(s) who died during the trip? Find out their name(s) and age(s).

Visualizing Categorical Data

Familiarize yourself with matplotlib and plotly to get to know some of the graphing options these libraries provide. The three **lab assignments** below are to be answered using plots. Don't forget to set appropriate values for title and axis labels and consider customizing the legend where applicable.

1. Create a plot that gives us an idea how likely it was for passengers of different classes (1st, 2nd, 3rd) to survive. ✓
2. Visualizing passengers' journeys using their embarking locations and their survival status (think voter-flow analysis). ✓
3. Make up one additional question yourself and come up with a fitting plot to answer it! ✓

If you aren't sure how to choose an appropriate type of plot answering the question posed, you may find data-to-viz.com helpful!

3 Homework

At the beginning of the next lab session, you will get a chance to indicate which of the **lab assignments** (✓) you completed and some of you will be asked to present their solution in class. Each student will be asked to present their solutions **at least twice** over the course of the semester. Other than that, no additional upload is needed. Please see the Course Syllabus for details on how this marking of assignments and their presentation affects your grade.

simon.schindler@fh-salzburg.ac.at
maximilian.schirl@fh-salzburg.ac.at
martin.uray@fh-salzburg.ac.at