FactSheet:: Titanic Data

M	Topic & Assignment

M2

Titanic data mining analysis

A. **Background and overviews**

- https://www.rdocumentation.org/packages/titanic/versions/0.1.0
- https://www.kaggle.com/competitions/titanic/overview
- https://www.encyclopedia-titanica.org/

The Titanic DataFrames describe the survival status of individual Titanic passengers, not the crew, with ages for ~half the passengers. One of the original sources is Eaton & Haas (1994) Titanic: Triumph and Tragedy, Patrick Stephens Ltd includes a passenger list created by many researchers and edited by Michael A. Findlay [1].

B. **Interesting models - built in R code for display convenience**

```
> data <- read.csv('titanic.csv')
```

- # Linear regression model
- model <- lm(survived ~ age + sex + pclass + sibsp + parch, data = data)

- **Binomial** Predicting survival based on age, sex, and passenger class
- model <- glm(survived ~ age + sex + pclass, data = titanic, family = binomial)

- **Poisson** - Predicting the count of siblings/spouses based on passenger age
- model <- glm(sibsp ~ age, data = titanic, family = poisson) summary(model)

- **Neg.Binomial** - Predict count of parents/children by passenger age and sex
- model <- glm.nb(parch ~ age + sex, data = titanic) summary(model)

C. **Data <class.github>**

- raw data; unsplit and preprocessed [source: https://hbiostat.org/data/ <titanic.3>
- train, test; from kaggle

D. **Data dictionary**

passengerid	sequential unique id
survived	0=no, 1=yes
pclass	1,2,3:passenger class (1st, 2nd, 3rd); proxy for socio-economic class
name	Christian name
sex	male, female
age	00, NA, blank. in years; some infants w fractional values
sibsp	number of siblings and spouses aboard
parch	<parent.child> #parents or chil
ticket	alpha, numeric, character
fare	0.0000 decimals
cabin	C#, blank,
embarked	C, Q, S <Cherbourg, Southampton, and Queenstown>

References:

1. Harrell Jr, F.E.,(2002). Titanic data, Vanderbuilt biostatistics datasets. Vanderbilt University. Retrieved from: https://hbiostat.org/data/repo/titanic.html. Retrieved on 05.15.2023.