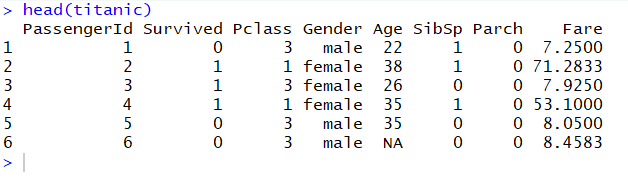
**Titanic Ship dataset**

* **Read the data into Rstudio**



* **A Glimpse of the data**



Here “Pclass” refers to passenger’s class, 1 is used for 1st, 2 for 2nd and 3 for 3rd respectively. Feature “SibSp” refers to number of siblings and spouses onboard and “Parch” refers to number of parent and child of the passenger are on board.

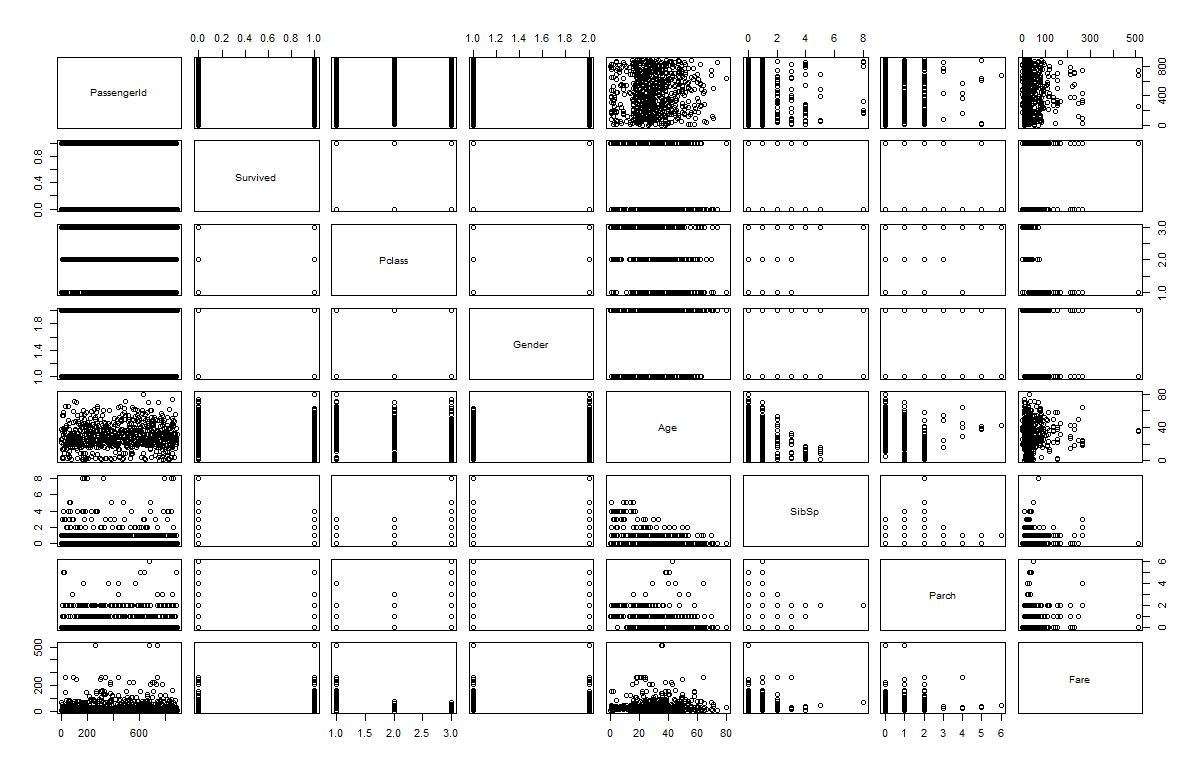
* **Intuitive prediction**

Intuitively one could infer that larger the family more will be the fare or higher the class higher the prizes in return more will be fares.

It means predictors “Pclass”, “SibSp” and “Parch” are most relevant for predicting the fare.

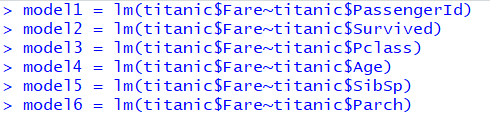
* **Visualizing with Pairs**

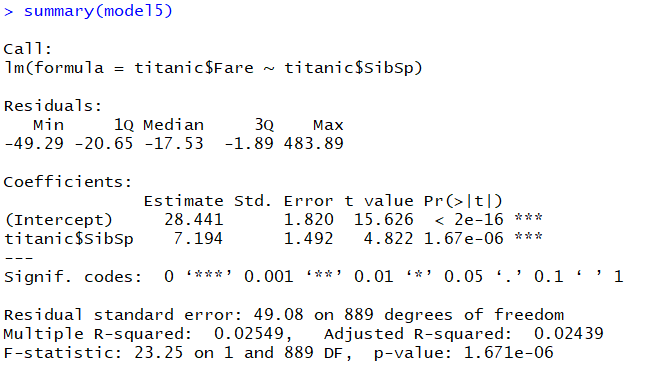




No such trends could be found by simple visualization so we now move on to creating models and correlation matrixes to extract meaningful inference.

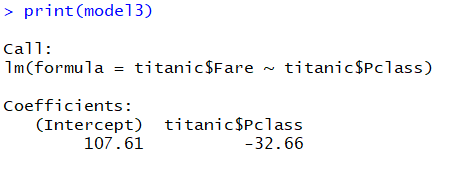
* **Simple regression models**



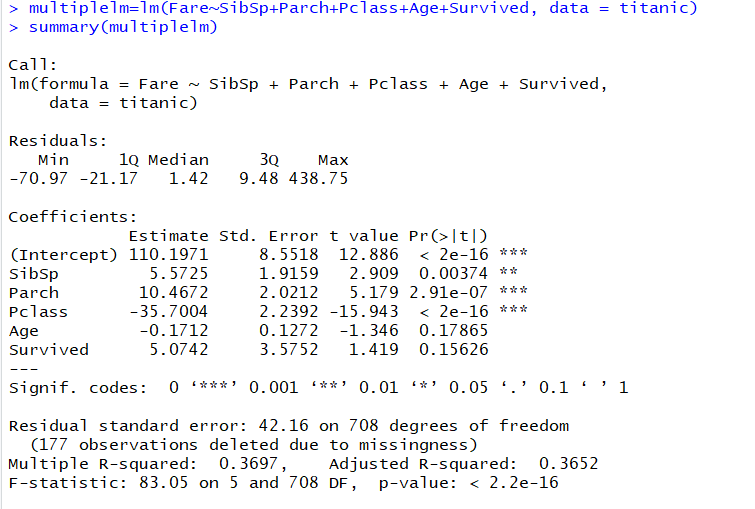


Now after regressing Fare on each of its predictor separately, some relation can be inferred by generating summaries of these regressed models.

The p-values of “SibSp” and “Parch” show some relationship as inferred intuitively. Apart from that an unexpected relation is seen as “Pclass” as a negative relation on Fare which is a little hard to digest.

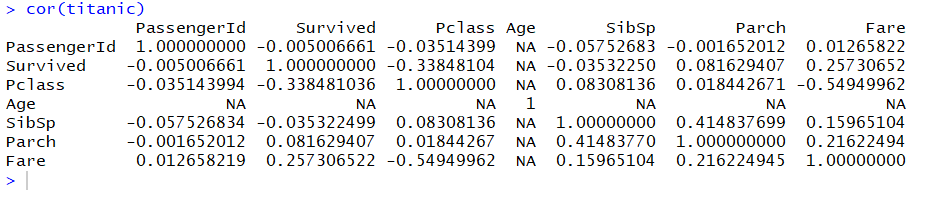


* **Multiple linear regression**



“SibSp” is not that significant anymore as we can see from its p-value. But how is it possible? “SibSp” had a clear significance in simple regression model. This “how” can be answered through correlation matrix. So let’s find out which other significant predictor is correlated to “SibSp”.

* **Correlation Matrix**



* **Conclusion**

Finally from correlation matrix it is inferred that predictors “Survived” and “Parch” are the most helpful for predicting “Fare” as they have high enough correlations with “Fare”. The significance of “SibSp” predictor in simple regression model was due to its high correlation “Parch”.