

CS989 / CS982: Big Data Fundamentals / Techniques

Lab 3 – Visualisation

Part 2 - Visualisations with more than one variable

1. Import the necessary libraries (numpy, pandas, matplotlib and seaborn)
2. Import the Titanic dataset available at <https://www.kaggle.com/c/titanic/data> (it is train.csv that you want)
3. Produce a stacked bar chart showing the number of each gender in each passenger class
4. Produce a heatmap showing the correlation between each numerical variable. What shows a strong correlation?
5. Produce a single scatter plot showing age and passenger class as well as age and number of siblings, different symbols should be used to represent the two different comparisons
6. Play around with different visualisations of this dataset