CS982: Big Data Technologies
CS989: Big Data Fundamentals

Lab 3 - Visualisation

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 some box plots for numeric values in the dataset

4. Plot a bar chart showing the number of survivors and fatalities. Include a title on the chart

5. Produce a horizontal bar chart showing all passenger classes, ordered with smallest number at top and largest at bottom

6. Produce a density plot for number of siblings (SibSp)

7. Produce a stacked bar chart showing the number of each gender in each passenger class

8. Produce a heatmap showing the correlation between each numerical variable. What shows a strong correlation?

9. 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

10. Play around with different visualisations of this dataset

11. Using your interesting dataset from last week explore the dataset starting with some techniques that we explored in the lecture and this lab