# 6 - Data Visualisation - Lab Exercises

Create a new cell for each question.

## Setup

If starting a new notebook, ensure you have run the following code.

```
import seaborn as sns
titanic_df = sns.load_dataset('titanic')
titanic_df['pclass'] = titanic_df['pclass'].astype('category')
titanic_df['alive'] = titanic_df['alive'].astype('category')
titanic_df['n_family'] = titanic_df['parch'] + titanic_df['sibsp']
```

## Section 1

### Exercise 1.1

Using Seaborn and your `titanic_df` dataframe, create a `count` plot that shows the number of passengers that embarked at each `embark_town`. Where did the majority of passengers embark?

### Exercise 1.2

Make the same plot as above, but split the data so that it is coloured by `pclass`. What was the majority class of the passengers that embarked in Queenstown?

### Exercise 1.3

Let's examine passenger `fare`'s, how much the passenger paid to travel. Create a `strip` plot that has the passenger class as the x axis variable, and the passenger fare as the y axis variable. What stands out to you about the fares that passengers paid?

### Exercise 1.4

Reproduce the plot above but adjust the necessary argument to make a...

- Box plot
- Boxen plot
- Violin plot

How do the different plot types add to or shape your interpretation of the same data? Is there a plot type that you find more intuitive or more useful in some way? Are some plots obscuring inferences?

### Exercise 1.5

The titanic dataset has a number of categorical columns that you could use to split your data.

- survived
- pclass
- deck
- embark_town
- alone and a few columns with numerical ranges
- age
- fare
- n_family

Remember you can split the data by using the `x` axis, using colour (`hue`) and by using the `col` argument to create a different column per group. Experiment with different combinations of categorical and numerical variables and types of plot to see what insights you can gain.