Assignment 2 Question 1

The data for this assignment comes from a Kaggle project which collected data on heart disease from five different sources (https://www.kaggle.com/fedesoriano/heart-failure-prediction/version/1 (https://www.kaggle.com/fedesoriano/heart-failure-prediction/version/1)). The data can be downloaded at the link below. You can load the data with the following code:

```
## Note that the heart.csv file must be in your working directory

library(tidyverse)
heart_tbl<-read_csv("heart.csv")
```

a. Using the plot(s) of your choice, assess whether there is an association between the sex of the patient and their resting heart rates, i.e. is there a difference in distribution of the resting heart rates across the sexes? Explain your answer.

b. Produce a stacked barplot showing the distribution of Chest Pain Type for each level of RestingECG.

c. Produce a summary table containing counts and proportions of RestingECG category for each sex/ChestPainType factor combination.

d. Create a summary table that finds the mean, median and IQR of RestingBP, Cholesterol, FastingBS, and MaxHR for each of the Chest Pain Types and report those results in a tibble where the columns are the levels of Chest Pain Types and the summary statistics are in the rows.

e. Using the plot(s) of your choice, explain which of the following measurements seem most strongly associated with Heart Disease (heart disease vs. normal) : RestingBP, Cholesterol, FastingBS, and MaxHR.

f. Create both a 2-d histogram and a 2-d contour plot to assess the association betwen RestingBP and MaxHR. Describe this association and also explain which plot you think shows the association most clearly (or explain why they are about the same).

g. Using the plot(s) of your choice, determine whether the association in (f) depends on either the Chest Pain Type or the Heart Disease status (or both).