Question Number 1

	1	2	3
0	123	158	528
1	200	119	181

Table 1: Two-way table of the number of passengers that survived and did not survive for different classes of passengers for the passengers dataset. This two-way table shows that the 0 row specifies the number of passengers that did not survive for each column of the different passenger classes.

A)

i. The total number of first class passengers who survived is 200.
ii. The total number of third class passengers who survived is 181.
iii. The proportion of all the passengers that survived is 0.382.
Calculation: (500) / (1309) = 0.382
iv. The proportion of all the passengers that were in the third class is 0.543.
Calculation: (709) / (1309) = 0.543
v. The proportion of the first class passengers that survived is 0.619.
Calculation: (200) / (323) = 0.619
vi. The proportion of the third class passengers that survived is 0.255.
Calculation: (181) / (709) = 0.255

B)

i.

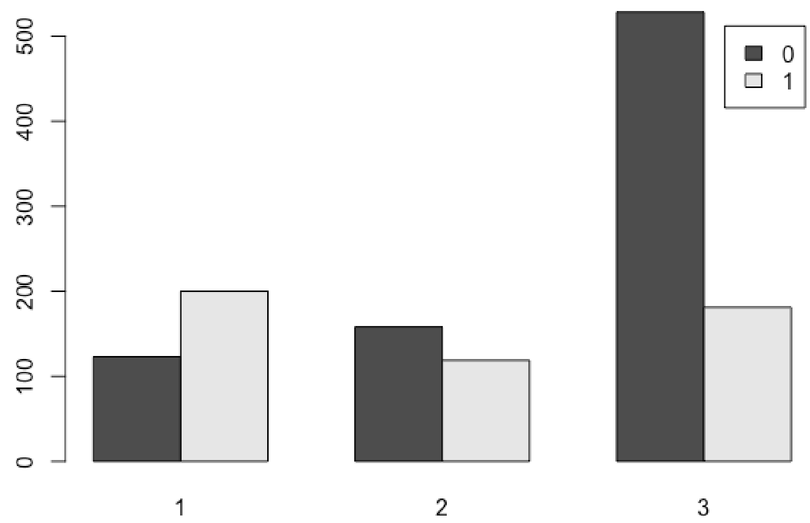

Figure 1: Side-by-side barplot of the number of passengers that survived and did not survive for the different classes of passengers for the passengers dataset. This barplot shows that the number of first class passengers that did not survive is considerably lower than that of the third class passengers.

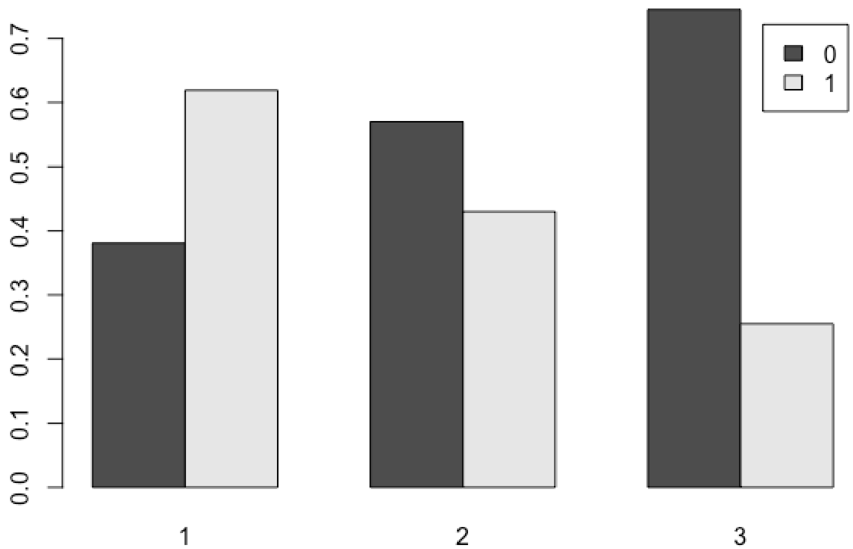

Figure 2: Side-by-side proportion barplot of the number of passengers that survived and did not survive for the different classes of passengers for the passengers dataset. This proportion barplot shows that the proportion of first class passengers that survived is much higher than that of the third class passengers.

ii. The class with the highest survival proportion appears to be the first class passengers and the class with the lowest survival proportion appears to be the third class passengers. Although, the number of passengers who survived in both the first class and third class is quite similar in numbers, however, the number of passengers who did not survive in the third class is 3 to 4 times of that of the number of passengers who did not survive in the first class. Due to the great disparity in the total number of passengers in each of the two classes, the survival proportion of the third class appears to be much lower than that of the first class.

<u>Question Number 2</u>

A)

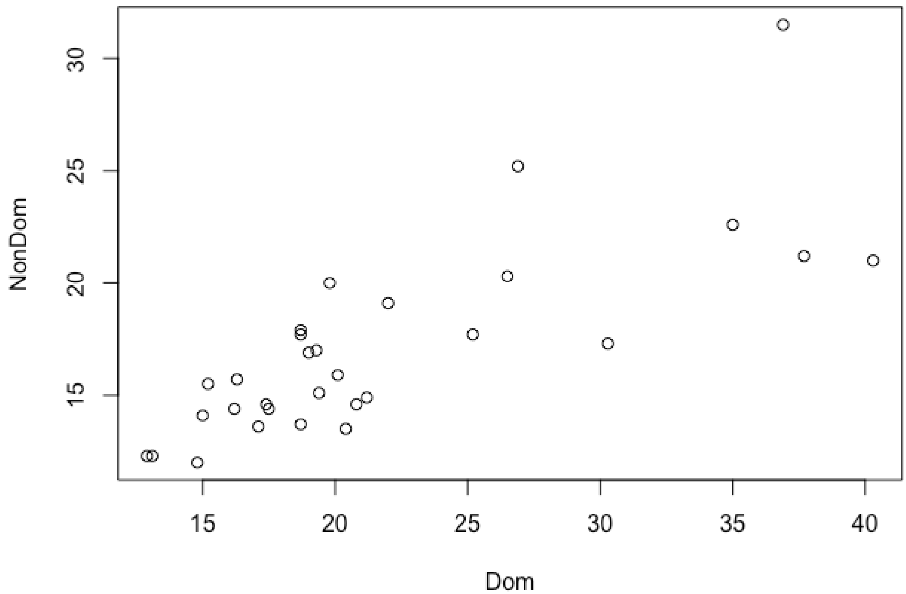

Figure 3: Scatterplot of bone strength for the dominant and non-dominant arms for the bone strength dataset. This scatterplot shows that the dominant and non-dominant arms have a positive relationship as their data move upwards into the positive quadrant.

Comment : The dominant and non-dominant arms have a positive relationship and their data forms a linear best fit line. The strength of their data is moderately strong because some of the data points are scatter

B)

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.33146    1.52264   4.815 4.60e-05 ***
Dom          0.44766    0.06637   6.745 2.53e-07 ***
```

Table 2: Liner model coefficients of bone strength for the dominant and non-dominant arms for the bone strength dataset. This linear model coefficients shows the estimate for the intercept(non-dominant bone strength) and slope(amount by which the non-dominant bone strength will change for each one-unit increase of dominant bone strength) of the least-squared regression line of the model.

Intercept: 7.331

Slope of the least-squared regression line: 0.448

Least-squared regression line: $Y = 7.331 + 0.448*X$

a. Y : The response variable (non-dominant bone strength)
b. 7.331 : The y-intercept is the value for when x = 0 (non-dominant bone strength for when the dominant bone strength is 0)
c. 0.448 : slope that states how much the non-dominant bone strength changes with each one-unit increase of the dominant bone strength.
d. X : The explanatory variable (dominant bone strength)

C) X = 20 cm^4/1000, Y = 7.331 + 0.448*X

= 7.331 + 0.448(20)

= 7.331 + 8.96

= 16.291 cm^4/1000

D) This value is beyond the boundary of the dominant arm bone strength dataset.

Question Number 3

(a) Control group – compare the effect of treatment between a treatment group and control group. The reason for this is to be able to verify whether or not the effect of the treatment is indeed caused by the treatment and not other factors.

Randomisation – two groups of individuals are as similar as possible in all aspects to prevent selection bias.

Replication – same number of subjects in each treatment group to reduce variation in the result.

(b) Randomisation can prevent selection bias and can also mitigate any lurking variables that may affect the result of the experiment.
(c) It is a treatment that tastes, smells and looks like the actual treatment but without any active ingredients.
(d) Both subjects and researchers have no information on which treatments the subjects are receiving.