1 MAT013 - Example Sheet

1.1 Chapter 5

1. Using sql, create a copy of the MMM and JJJ data sets, including all the variables.

2. Using sql, create the previous copies selecting just the variables, Name, Age, Sex, Random_Number, as well as the bmi of the observations.

3. For the following data set:

```
Var1, Var2, Var3, Var4, Var5
A, 1, A, 2, B
A, 1, A, 2, B
B, 1, A, 1, C
C, 2, B, 2, D
C, 2, C, 1, E
```

    1. Create a copy of the data set removing complete duplicate rows.
    2. Create a copy of the data set removing duplicates of Var2.
    3. Create a copy of the data set removing duplicates of Var3 and Var4.
    4. Create a copy of the data set selecting only observations where Var2 $>$ Var4.
    5. Create a copy of the data set ordering by Var1.
    6. Create a data set containing the mean, std, max, min and variance of Var4 and Var2 by Var1.

4. Download the data sets dogs.csv and cats.csv use sql to:

    1. create an inner join.
    2. a left outer join.
    3. a right outer join (you won't be able to use sql for this in R).
    4. a full outer join (you won't be able to use sql for this in R).

5. Create a histogram for the Height of people in the JJJ data set.

6. Modify the above plot to be a density plot with your own legends labels and title.

7. Obtain a scatter plot of weight against height for people in the JJJ data set.

8. Modify the above plot so that the points are proportional to the age.

9. Obtain a box plot for the Height of people in the JJJ data set by sex.

10. Obtain a scatter plot of Weight against Height with a smoothed trend line.
11. Obtain histograms of height against weight, compartmentalised by sex.
12. Create a scatter plot with a fitted linear model for Height against Weight for all people in MMM and JJJ. Compartmentalise your graphs based on the data set and the sex.
13. Save the above graph to file.
14. Get all of todays trends on twitter.
15. Search for all tweets with the `orms` hashtag.
16. Find the tweets from INFORMS.

The relevant data can be found here:

- dogs.csv
- cats.csv