

In these labs you should work through the problems on the sheet and type your answers into a Microsoft Word document. When you have completed the tasks, print the document and hand it in.

Your answers can benefit from having small sections of R output and graphs copied and pasted into your word document. It can be useful to limit the width of output produced by R so that it fits into your document. You can change the width of output produced by R with a command like: `options(width=50)`.

The labs will count for 1 mark in the current assignment. They will be graded on a 0-1 basis. Don't forget to put your name and student ID on the document.

Today we'll have a look at some categorical data analysis. To do this we'll use the data set on hair colour, eye colour and gender mentioned in class. The data is stored in a data set available from the class web page (it's called *students.txt*). Rather than downloading the data and then reading it into R we can read it directly into R over the network. You can do this as follows:

```
x = read.table("http://www.stat.auckland.ac.nz/~ihaka/120/Data/students.txt",
              header = TRUE)
```

Have a look at the form of the data by printing the first 10 rows of the data set.

When working with categorical observations, it is common to summarise the data by reducing it to tables of counts. Do this for the student data by producing tables of counts for each variable. For example, a table giving the number of number of people with each hair colour can be produced with:

```
table(x$Hair)
```

Produce a bar plot for each variable from the summary tables.

The next task is to create *contingency table* of counts by *cross-tabulating* the data. A cross tabulation of hair colour and gender can be produced as follows.

```
table(x$Hair, x$Sex)
```

Produce a summary table by cross tabulating hair colour and eye colour and then produce a side-by-side bar plot from the table. Do the same with the order of the variables reversed.

From the contingency table for hair colour and eye colour produce a mosaic plot (make eye colour the first variable). Now do the same for the three way contingency table for gender, eye colour and hair colour. Are there gender differences in the hair/eye pattern?