Data Programming with R.
Lab 5: factors and tables.

1. Load in the `MASS` library into `R`, and load the `quine` data set using the `attach` function. What does the `quine` data set record? Have a look at the structure of the data set using `str`.

2. Give the columns of the data set more informative names.

3. Use the `table` function on the ethnic background variable. What ethnicity is most prevalent in the data set?

4. Use the `factor` function to change the age and learner variables into (sensibly!) ordered factors using the extra argument `ordered = TRUE`. Check the structure of the data set to ensure what you think has happened to these variables, actually has.

5. Use the `table` function to compare learner ability and ethnicity in a 2-way table. How many children in the study are slow learners? What other tables of interest can you create from this data set?

6. Use `tapply` to see if the mean number of days absent is different for Aboriginal and non-Aboriginal students. What about the standard deviation?

7. Use `aggregate` to calculate the mean days absent by learner status and sex. What about the standard deviation?

8. Use the cut function on the days absent variable to divide it up into 3 categories: `few` (0-20), `medium` (20-60) and `many` (60-100). Have a look at the output. Is there anything strange going on here? If so, how can you fix it (use the help file!). How many students fall into the `many` category?