

## Day 2 – Task 2

Open the High School and Beyond xls data using the R starter script provided (see the resources folder for the data and information about the contents).

1. Look at all the commands in the start script and add comments to explain what they do.
2. Now work in pairs to create an R notebook file that shows examples of how to use some functions from the `dplyr` package, demonstrating how they can be used for the High School data set.

To use this you may need to install `dplyr`

```
install.packages("dplyr")
```

Once you have installed `dplyr` you will need to load it into the R environment:

```
library(dplyr)
```

You can look into the help files (e.g. `?select`) or search (e.g. `dplyr` tutorial) to find out how to use the commands:

A.

- `select()` picks variables based on their names.
- `filter()` picks cases based on their values.
- `mutate()` adds new variables that are functions of existing variables

B.

- `arrange()` changes the ordering of the rows.
- `group_by()` which allows you to perform operations “by group”
- `summarise()` reduces multiple values down to a single summary.
- Can you work out how to produce the following summary information of grouped data: mean, count, max, min, standard deviation.