Preparing your data set for further analysis

100xp

Your data still looks a bit messy so it's time to clean it up with data manipulation techniques. You will do this using `dplyr`, and R package that gives you access to the most important data manipulation tools and makes them easy to use.

`Dplyr` makes use of the pipe operator `%>%` from the `magrittr` package. Pipes take the output from one function and feed it to the first argument of the next function:

```
head(AC_Survey_Subset,20)

# Equivalent piped version
AC_Survey_Subset %>% head(20)
```

You should read it as "Take the AC_Survey_Subset dataset and then apply the `head()` function to it with the optional argument 20". By using this pipe operator `%>%` you can also chain operations:

```
tail(head(AC_Survey_Subset,20),5)

# Equivalent piped version
AC_Survey_Subset %>% head(20) %>% tail(5)
```

This reads as: "Take the AC_Survey_Subset dataset, then apply the `head()` function to it with the optional argument 20, and finally apply the `tail()` function to it with the optional argument 5".

The `AC_Survey_Subset` dataset you imported in the previous exercise is already available.

Instructions

- Load in the `dplyr` package and convert `AC_Survey_Subset` to a `tbl` with `tbl_df()`.
- Next, use a chain of piping operators to:
  - Remove observations that have `NA` values from `AC_Survey_Subset` with `na.omit()`;
  - Retain observations for which `SCHL` is in `c(21, 22, 24)`, corresponding to Bachelors, Masters and PhDs, using `dplyr`'s `filter()` function;
  - Group according to `SCHL` with `dplyr`'s `group_by()` function;
- Assign the final result from the second instruction to `AC_Survey_Subset_Cleaned`.

Take Hint (-30xp)