I. TITLE

i. title

This article is a summary of a collection of videos produced by the Rstudio team discussing the data manipulation tools dplyr and tidyr. These are my own notes on these videos that I am publishing for reference for myself and others.

Ways in which we can access information

1. Extracting existing variables: select()

2. Extract existing observations: filter()

3. Drive new variables from existing ones: mutate()

4. Change the unit of analysis: summarise()

select()
This function allows for the subsection of a smaller number of columns from a larger dataframe.
Dropping a column.
Obtaining a range of variables.
filter()
Filter out observations using logical tests.
mutate()
Mutate() allows you to generate new variables from existing ones.
summarise()
Calculates summary statistics from a dataframe.

arrange()
Arranges the order of the rows within a dataframe based on the values within a column.

Adding additional variables into the arrange function will use those as tiebreakers between observations that have the same value.
Pipe Operator

group_by() + summarise()