R Programming: Worksheet 7

1. **Writing a report about our analysis of movie data using tidyverse**

   In Worksheet 4, we performed a series of analyses of some movie data, using tidyverse. Here, we're going to turn our analyses into a literate programming report!

   We're going to write an `.rnw` file, then knit it using `knitr` to make a `.tex`, then turn it into a PDF using `pdflatex` (or using built in RStudio commands). We'll show some, but not all, of our our code commands. We'll mostly hide the code, to make our report pretty, but will occasionally print out exemplary commands, for example some code to perform linear regression (and you can choose to add more if you'd like).

   Note that when using `knitr`, in `.rnw` files, outside of code chunks *i.e.* in the middle of a text paragraph, we can use the command `\Sexpr{}` to print the result of running one line of code, which could be the value of some variable defined in a previous code chunk to the screen. This is useful for including numbers in a paragraph.

   Also, note that if you are using RStudio, your default compilation method might be `Sweave` and not `knitr`. You can change this by going to RStudio then Preferences, and then within "Sweave", selecting to Weave Rnw files using knitr. If you use `Sweave` instead of `knitr`, you might need to set options differently to replicate the results as compared to the `movie_report.rnw` included on the course website.

   (a) Get started by making a minimal `.rnw` file. Include some trivial text and code. Make sure you can compile this example.

   (b) Now start modifying your minimal working example to do something useful. Using the Latex commands `author`, `title` and `maketitle`, give your report a title. Make a code block that you do not print the results of using `echo = FALSE`, and use it to load the movie data into memory. Make sure nothing from this code block is printed to the screen. *Consider setting `echo = FALSE`, `message = FALSE` and `cache = TRUE` in this code chunk, and make sure you understand what they do.*

   (c) Migrate the content of Worksheet 4 Question 2 over to your knitr document, with some basic summaries of the dataset. For example, how many rows are there, and how many different MPAA ratings there are. Use `section` to break your report into relevant sections.

   (d) As part of migrating Question 2 over, add a table of the `movies_description` that we saw in Worksheet 4, as a table, using a code block, using `kable`. This table will be very wide. Either figure out how to make the table wrap, or just make the long entry shorter (change the contents of the entry to `"action, ..."`).

   (e) As part of migrating Question 2 over, include a table summarizing how many times each of the different MPAA ratings are seen in the `mpaa` column. Note that you can pipe results to `kable` when making your table.

   (f) Migrate Worksheet 4 Question 3 over to your knitr document, about budget and ratings.

   (g) Migrate Worksheet 4 Question 4 over to your knitr document, where we queried the data to learn about movies with large budgets and the movies represented the most in the dataset. Optionally, put `options(scipen = 999)` in your initial code chunk do disable rounding of large numbers into scientific notation, when printing tables to the document using `kable`

(h) Finally, migrate Worksheet 4 Question 5 over to your knitr document, where we looked at the relationship between budget and movie type

(i) *Optional*, migrate over Worksehet 4 Question 6