Homework 2: Reading & Writing Data with R

In this assignment use different R packages to import and export data into RStudio.

Once completed, submit two total documents via Blackboard: an .Rmd R Markdown file and the corresponding HTML document.

Fig. 1: Hex logos for the readr and data.table packages. Image obtained from https://github.com/rstudio/hex-stickers and https://slowkow.com/notes/data-table-aggregate/.

1.

a) First, create a new R Markdown document with File > New File > R Markdown… Knit it by clicking the Knit button (top left).

b) In the `setup` chunk, update the `knitr::opts_chunk$set(echo = TRUE)` code to `knitr::opts_chunk$set(echo = TRUE, error = TRUE)`. The `error` global chunk option being set to `TRUE` still renders / compiles the document even if an error occurs.

2.

a) Import data on Taylor Swift songs directly from the URL https://raw.githubusercontent.com/dilernia/STA418-518/main/Data/swiftSongs.csv into R using the `read.csv()` function creating an object called `swiftSongs`.

b) Print a single table of the `swiftSongs` data containing the last 13 rows of the data using the `slice_tail()` function and the 30th through 34th columns using the `select()` function from the `dplyr` package.

c) Import data on Taylor Swift songs directly from the URL using the `read.csv()`, `read_csv()`, and the `fread()` functions, comparing the read times using the `mark` function from the `bench` package, storing the results of the `mark` function in an object called `readTimes`. Specify a minimum of 5 iterations in the `mark` function.

d) Create a violin plot to display the varying speeds of the three different functions for importing the data.

e) Based on the violin plot, which function was typically the fastest?

3.

a) Export data on Taylor Swift songs in the `swiftSongs` data set using the `write.csv()`, `write_csv()`, and the `fwrite()` functions, comparing the write times using the `mark` function from the `bench` package, storing the results of the `mark` function in an object called `writeTimes`. Specify a minimum of 5 iterations in the `mark` function.

b) Create a violin plot to display the varying speeds of the three different functions for exporting the data.

c) Based on the violin plot, which function was typically the fastest?