MTH208: Worksheet 8

Introduction to dplyr

In this worksheet, we will go through some basics of the `dplyr` R package. The `dplyr` package provides a rich framework to do data manipulations and wrangling.

1. The following website has the ODI rankings of cricket batters:

   https://www.icc-cricket.com/rankings/womens/player-rankings/odi/batting

   Using `html_table()`, extract the dataset of the 100 batters. Make sure the dataset is clean — including the first column of the dataset.

2. We know that datasets are saved as `data.frame` in `R`. However, the `tidyverse` and `dplyr` R packages use a `tibble` to save a dataset. Save the dataset scraped in Question 1 in a tibble object called `batting`.

3. `dplyr` packages allow easy handling of tasks such as subsetting etc in R. Go through this website to understand the details of the `dplyr` package and these functions:

   https://cran.r-project.org/web/packages/dplyr/vignettes/dplyr.html

   Repeat the commands in this document on the `starwars` dataset on your own and test them out.

4. The `mtcars` dataset in R contains data extracted from the 1974 *Motor Trend* US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

5. Using the `group_by()` and `summarise()` functions, create a table of average displacement and average horsepower of cars with various cylinders.

6. For the `batting` dataset:

   a. find the rankings of all the Indian Players.

   b. find the number of players in each team

   c. find the average ranking of each team

d. reorganize the final count of the average ranking of each team from highest ranking to lowest ranking.

7. For the `batting` dataset, using `group_by`, create a grouped tibble according to whether the player is from Asia or not from Asia.

   a. How many Asian vs Non-Asian players?

   b. Find the average rating of Asian players versus non-Asian players