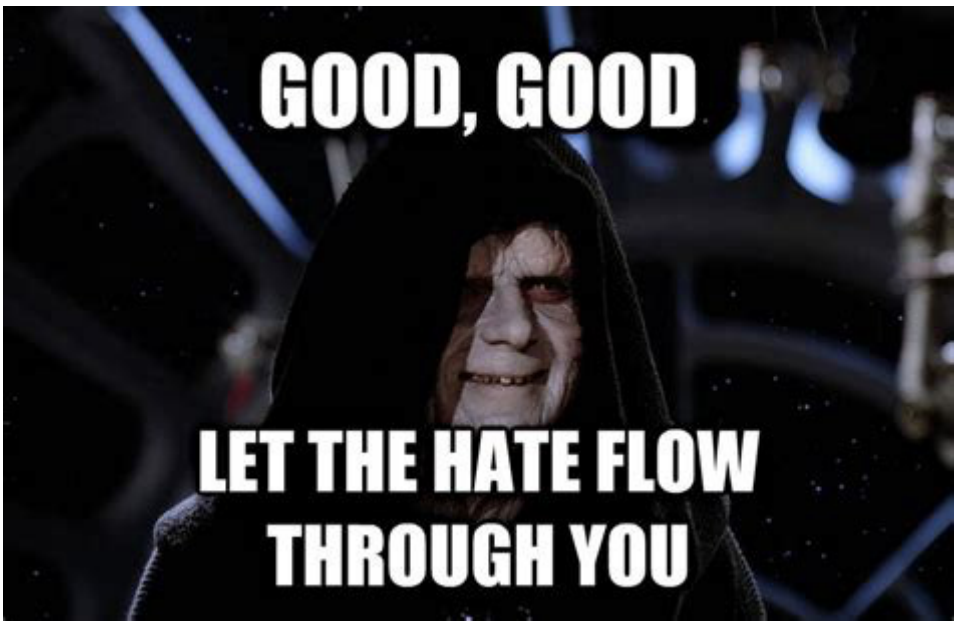
CA176 R Data Analysis Group Mini Project
Dr. Brian Davis & Dr. Michael Scriney

The submission box will open at **11:59PM on the 9th April 2023**. Late submissions will be accepted up until **11:59PM on the 16th of April 2023**



You *must* submit a **.zip** file containing all your commented rscripts, exported graphs and Rdata file. You may also choose to submit a word (.docX,doc) or PDF file documenting your code if you wish.

> **Required Reading:**
>
> - Students should familiarise themselves with Chapters 6-7 of R for Everyone (available on loop)
> - As well as Lecture - Reading data, manipulating data & visualising data in R.
> - Students should read up on the `ggplot2` library (Chapter 6-7 R for everyone)
> - Additional documentation + examples can be found here: https://ggplot2.tidyverse.org/

The Assignment

1. Install and setup
    - You will be using the `starwars` dataset from the `dplyr` package. If you do not have the package installed you should install it.
2. Scatterplot
    - Create a scatterplot with `mass` on the y-axis and `height` on the x-axis, save this as a picture (anyformat i.e png, jpeg).
3. Barchart
    - Provide a barchart of the `skin_color` of the characters in star wars, save this as a picture (any format i.e png, jpeg).

4. Histogram

    - Create a histogram of the `mass` of each starwars character, save this as a picture (any format i.e png, jpeg).

5. Height & Eye Colour

    - Create a dataframe called `not_droid` listing all characters who are **not a** `Droid`. Using this dataframe, create a violin plot comparing `height` to `eye_color`, save this as a picture (any format).

    - > **Note** to create a subset of a dataframe use the `subset` function.
      > To create a dataframe of everyone from the planet Alderaan we type
      > `from_alderaan<-subset(starwars, homeworld=="Alderaan")`
      > **Note 2**: A violin plot can be made from the `ggplot2` library, see Ch 7 - R for everyone.

6. Weight over time

    - Create a line chart contrasting the `mass` of each star wars character over their `birth_year`. Save this as a picture (any format).

7. Which film?

    Create a function called `films` which takes in **a character name** and prints out every film that character appear in *if only if the character appeared in more than one film* . Print each film on a separate line: A sample output can be seen below

```
> films("Luke Skywalker")
[1] "Character Luke Skywalker appears in films The Empire Strikes Back"
[2] "Character Luke Skywalker appears in films Revenge of the Sith"
[3] "Character Luke Skywalker appears in films Return of the Jedi"
[] "Character Luke Skywalker appears in films A New Hope"
[7] "Character Luke Skywalker appears in films The Force Awakens"
```

> **Note:** You can select what columns you want returned from the subset command, the code `as.list(blue_eyes<-subset(starwars, eye_color=="blue", select=c(name)))` would return a list of the *names* of all characters who have blue eyes.

8. Bounty Hunter?

    - One notable absence of the dataset is the character's allegiance Take the csv file "allegiance.csv" from loop and merge it with the starwars dataset. Save this as an *.rdata* file called `starwars_updated.rdata`

9. "There will be substantial reward for the one who finds the Millennium Falcon"

Add a column called `will_find` to the dataset where `TRUE` indicates if the character would collect the reward for the *Millennium Falcon* and `FALSE` otherwise.

> **Hint:** Add a column called `will_find` which is `TRUE` if the characters `Affiliation` is `"Bounty Hunter"`

10. The tragedy of Darth Plagueis the Wise is "not a story the Jedi would tell you".

    Add a column called `would_tell` to the dataset where `TRUE` indicates if the character would tell you the tragedy and `FALSE` otherwise.

> **Hint:** Add a column called `would_tell` which is `TRUE` if the characters `Affiliation` is `Sith`