STAT 435
Spring 2024

Name: William Franzen ____________________________

Lab Assignment #3 (*15 points*) – Due by 1:00 PM on Friday, 2/2/2024

Notes: ~~~~~~

- If you need help on an R function/command, type *?functionname* or *?commandname* and help for this function/command will appear in the Help window.
- Create a new folder on your desktop and name it **LA3_*your name***. Set this as your working directory in RStudio.
- In RStudio, open a blank source file (R Script) to work in, and make sure all History entries are cleared before you start your work on the following questions.
- In the R Script, add *Lab Assignment-3 (by **your name**)* as a comment line, and specify the question number.

~~~~~~

In the following steps, you will analyze the box office numbers of the Star Wars franchise. May the force be with you!

```
# Star Wars box office in millions (!)
box.office <- c(460.998, 314.400, 290.475, 247.900, 309.306, 165.800)
```

The above vector `box.office` is defined that represents the box office numbers from the first three Star Wars movies. The first, third and fifth element correspond to the US box office revenue for the movies, the second, fourth and sixth element represent the non-US box office revenue.

1) Construct a matrix with one row for each movie. The first column is for the US box office revenue, and the second column for the non-US box office revenue. Name the matrix `star_wars_matrix`. Print the matrix.

2) Instead of as a single vector, the box office numbers for the three Star Wars movies are represented as three vectors.

```
New_hope <- c(460.998, 314.400)
Empire_strikes <- c(290.475, 247.900)
Return_jedi <- c(309.306, 165.800)
```

   Remember the R function that can be used to paste together different vectors as if they were rows of a matrix? Try it out on these astronomical numbers! Again, construct the matrix `star_wars_matrix` with one row for each movie.

3) Give the columns of `star_wars_matrix` the names `"US"` and `"non-US"`, respectively. Give the rows of the matrix `star_wars_matrix` the names of the three movies: "A New Hope", "The Empire Strikes Back" and "Return of the Jedi". After naming the rows and columns of your matrix, print it.
~~~~~~

4) The single most important thing for a movie in order to become an instant legend in Tinseltown is its worldwide box office figures. To calculate the total box office revenue for the three Star Wars movies, you have to take the sum of the US revenue column and the non-US revenue column.

   Calculate the worldwide box office figures for the three movies and put these in the vector named `worldwide_vector`. Print the vector.

5) In the previous step, you calculated the vector that contained the worldwide box office receipt for each of the three Star Wars movies. However, this vector is not yet part of `star_wars_matrix`.

   Add `worldwide_vector` as a new column at the end of the `star_wars_matrix` matrix and assign the result to `star_wars_ext`. Print the new matrix.

6) Earlier you have stored the data for the first trilogy in `star_wars_matrix`. The box office numbers on the second trilogy are given below:

	US	Non-US
The Phantom Menace	474.5	552.5
Attack of the Clones	310.7	338.7
Revenge of the Sith	380.3	468.5

   Create another matrix with this data using one line of code (including naming of the rows and columns) using `dimnames` argument and name it `star_wars_matrix2`. Print the matrix

7) Assign to `all_wars_matrix` a new matrix with `star_wars_matrix` in the first three rows and `star_wars_matrix2` in the next three rows. Print `all_wars_matrix`.

8) Calculate the total revenue for the US and the non-US region and assign it to `total_revenue_vector`. Then, print the result.

9) You'll continue working on `star_wars_matrix`, which is still a matrix containing both US and non-US box office figures for the first three movies.

   No need to assign the following elements to new variables; simply write the command to print them.

   - Select the US box office figure for "Return of the Jedi".
   - Select the non-US box office number for "The Empire Strikes Back".
   - Select all non-US box office revenue from `star_wars_matrix`.
   - Extract the revenue information for "A New Hope".

10) Using `star_wars_matrix`, calculate the average Non-US revenue for all movies. Assign this to the `non_us_all` variable. In other words, take the average of all elements of the second column. Report the result here: 242.7

    Same question, but now only for the first two Star Wars movies. Assign the result to `non_us_some`. Report the result here: 281.15

11) Select all revenue figures (both US and non-US) for "A New Hope" and "Return of the Jedi" from `star_wars_matrix`.

12) Just as with vectors, you can also subset matrices using names and logical vectors. Of course, you can only subset by name if the matrix you're working with actually has names associated with it.

- Using names, select the US revenues for "A New Hope" and "The Empire Strikes Back".
- Using logical vectors, select the last two rows and both columns from `star_wars_matrix`.
- Finally, select the non-US revenue for "The Empire Strikes Back" with whatever technique you like.

13) After all these exercises on Star Wars, you start a conversation about it with your friends. Before you realize it, a big discussion has started about who is the biggest fan, with three people claiming to be true adepts. To sort this out, you decide to collect data by asking everyone how often they have seen each of the movies. To keep things structured, the information (the view counts) are stored in two parts:

```
view_count_1
view_count_2
```

- Go to the Moodle Course page. From **Data, Script, and other R Files** folder under **Course Documents**, download the R script file *view_counts.R* and save it into your *LA3_your name* directory. Open this R script in your RStudio. It should appear under a new tab in your application. Add the two lines of code from *view_counts.R* to your *Lab Assignment-3* R script and execute it.
- Print `view_count_1` and `view_count_2`, and take a good look at these two matrices.
- Combine the matrices `view_count_1` and `view_count_2` in a new matrix, `view_count_all`. Then, print this new matrix.
- From `view_count_all`, subset the view counts for the three loudest debaters (Mark, Sheldon, and Monica). You can use subsetting by indices, names or a logical vector here. Store the result in the matrix `view_count_loud`, and print it.
- Calculate the total number of times the three loud debaters have seen any of the Star Wars movies. Store the result in a vector in `total_views_loud` and print it.
- Who has the highest number? Mark . With how many times? 19

Save your RScript naming it **StarWars_*your name***. Then, email this file and this lab assignment file (with the blanks filled in and saved) to the professor at sgazioglu@mtech.edu.

Have '**Stat435 – LA3**' in the subject line of the e-mail.