

Laboratory Session : March 30, 2023

Exercises due on : April 16, 2023

## Exercise 1 - American Airlines Employees

- The following repository <https://drive.google.com/drive/folders/1MBXTCbUJbX6kRo4fnjsb0Gv0ZZ7SQFkV?usp=sharing> contains few files with the number of part-time and full-time employees of the major airline companies in the United States:
    - `americal_airline_empl.txt`
    - `delta_airline_empl.txt`
    - `federal_express_empl.txt`
    - `united_airline_empl.txt`
- ~~1)~~ read the data and import them in a `data.frame` or `tibble` structure
  - ~~2)~~ merge the four data tibble in a common `tibble`
  - ~~3)~~ produce a plot of the behaviour of the employees as a function of time for all four companies, separately for the number of full-time and part-time employees
  - ~~4)~~ when did each company reach the minimum and maximum number of employess ?
  - ~~5)~~ plot the fraction of part-time worker over the total employess as a function of time
  - 6) did the COVID-19 pandemic have any influence in the employed workers of the airline companies ? Can you see a trend in the years 2019-2023 ?

## Exercise 2 - Data Frames and Tibble

- the `nycflights13` R package [1] contains data on all flights departing from New York City airports in 2013. All available data is organized into four tibbles [1] :
  - `airlines`: contains metadata on airlines names and corresponding carrier codes
  - `airports`: contains metadata on all airports connected to NYC
  - `flights`: has data of all flights departing from the three NYC airports (JFK, LGA and EWR) in 2013
  - `planes`: Plane metadata for all plane numbers found in the FAA aircraft registry.

further descriptions can be found here [1].

- ~~1.1)~~ Plot the total number of flights departed from each of the three NYC airports as a function of time (one entry for each of the 365 days of the year).
- 1.2) Plot the average number of flights computed over the first five working days of each week as a function of the week number of the year. Produce the same plot for the flights departing over the weekend (Saturdays and Sundays).
- 2.1) For each flight in the data frame, compute the departure delay<sup>1</sup> and extract the following pieces of information (separately for each NYC airport):
  - min, max and average delay for each day of the year (show the data in corresponding plots)
- 3) assuming the distance flew by the plane is, at first approximation, the distance between the two connecting airports (as given in the data frame), compute the average speed of each plane. Produce a plot of the average plane speed as a function of departure day of the year
- 4) analyze the flights offered by each airline company and determine:
  - the airline companies offering the largest two numbers of flights per day and per week;
  - the airline company offering the smallest number of flight per month;
  - the airline company offering the longest distance flight per month.

(you can produce plots, if you like, to visualize the results of the analysis)

Hint: use the `tidyverse` packages to manipulate the data frame and produce the visualization plots (i.e. `dplyr`, `ggplot2`, ...)

## References

[1] <https://cran.r-project.org/web/packages/nycflights13/nycflights13.pdf>

---

<sup>1</sup>a positive numbers indicates the flight had a delay, while a negative number is for a flight departing ahead of time.