University Master in Statistics and Operational Research (MESIO)
Statistical Software: R and SAS 2021-2022
Laboratory with R (Grupo B)

`IMPORTANT:` Last day to deliver the practice is on 12-31-2021 (23.59h) through the ATENEA campus. The practice should be done in groups of 2-3 people (maximum 3 people and minimum 1 person). You must send a report in word or pdf with the results, graphs and results of the study and on the other hand you must send the script in R

`PROBLEM STATEMENT:` The Scottish city of Glasgow hosted last COP26, the international summit on climate change, between 31 October and 12 December. These are the keys to an event organized under the umbrella of the United Nations and which seeks to put the fight against global warming on track. Many countries are committing to achieving zero net emissions by the middle of the century. This means that then they will only be able to emit the gases that can be captured by the sewers, both natural (for example, forests) and artificial (through capture and storage techniques that are currently in the experimental phase). To date, some 75 countries have set zero emissions targets by the middle of the century, most in 2050, although some also in 2060. During the Glasgow summit it is possible that more countries will do so. However, international organizations warn that there is a problem of coherence between many of these goals and the austerity plans that countries have and that do not lead to this neutrality.

One of the conclusions reached is about the importance analyze the existing real data of: CO2 and Greenhouse Gas Emissions to see the evolution of the same and the control of the policies that do the different states for his decrease. There are many repositories with such data and we will use one of them(https://github.com/owid/co2-data)

In the file *data frame* https://github.com/owid/co2-data/raw/master/owid-co2-data.xlsx you can find a complete Our World in Data CO2 and Greenhouse Gas Emissions dataset. The variable definition and description can be found in: https://github.com/owid/co2-data/blob/master/owid-co2-codebook.csv

The objective of your practical laboratory is contribute to the study of climate change, especially that produced in the last 30 years, through the use of real data.

What is requested is to carry out a study as complete as possible (descriptive, graphical, functions, bookstore of R) on the CO2 and Greenhouse Gas Emissions by country and their comparison with the previous periods.

Different functions should be performed as described in the following different sections.

Exercise 1 (4 puntos)

a) Import the study data *data frame* https://github.com/owid/co2-data/raw/master/owid-co2-data.xlsx and place them in a *data frame* with name `Greenhouse_Gas_Emissions`. Name the variables according to the specifications in the data set and place labels describing units and what each variable and case is (years, months, co2, etc.). You should place it properly according to the format of the

data.

If you can, import the data directly from the data file in the link indicated above, without having to save it previously to file.

Make appropriate comments on the results obtained in each section

b) Use the function `label` of package `Hmisc` to tag variables (using of data frame `Greenhouse_Gas_Emissions`).

c) Indicate the dimension of the data-frame and make a original description of each variables.

d) Represent graphically co2, methane and nitrous_oxide together by country and year. Can be possible all countries or a selection of a country, according to the request made.

e) Aggregate data-frame by decade and select data only by 3 last decades (1990-2000, 2000-2010, 2010-2020)

f) Calculate the mean, median, standard deviation and the interquartile range for each of the previous groups (or other statistics if necessary) for all quantitative variables (co2, co2_per-capita, ...). You can present other statistics suitable for this type of data, which you think are convenient.

g) Properly represent the data to be able to see the decade variations of the average co2,methane and nitrous_oxide in the total countries and country by country (maybe a selection of ten with more co2production).

h) Export *data frame* with the new aggregate variables (by country) to a new file sheet, for example in `agg_co2.xlsx`.

Exercise 2 (2 puntos)

Make two or more nice graphs with the data from Exercise 1 using functions from two of the following packages:

gplots `http://cran.r-project.org/web/packages/gplots/index.html`,

plotrix `http://cran.r-project.org/web/packages/plotrix/index.html`,

vcd `http://cran.r-project.org/web/packages/vcd/index.html`.

For (at least) one of the two graphs you should use the study group variable (country, decade) and for the other one a couple of the numerical variables. Present the syntax and interpret both graphs.

Schedule a function that does the following,

Exercise 3 (4 puntos)

a) Calculate the percentage increase between one decade and the following decade (eg. between 1990-2000 and 2000-2010) for co2, methane and nitrous_oxide together by each required country and decade, offer a graph per country required and the average of all years studied. You should be able to indicate what decades or countries it should represent.

b) Try to calculate the projection of co2, methane and nitrous_oxide by country requiered. Use a linear model to this computation (lm): For this, supposed that decade 1990-2000 has the value 1, 2000-2010 the value of 2 and so consecutively so it can be transformed into a regression model. You should be able to indicate what country it should represent. Use this example of prediction: `https://stat.ethz.ch/R-manual/R-devel/library/stats/html/predict.lm.html`

c) Comment on the results found and if you think they have effects on climate change. Optionally, you can build a library to project the future of co2, methane and nitrous_oxide and calculate statistics

by country (countries where CO2 increases, percentage increase for the next decade, variance of variation, etc).

d) Optionally, Look for other data (public repositories, other sources), if you think it is necessary and cross them with those already available, especially if they help to better explain the possible climate change that is taking place

e) Optionally, create a library in R to store the previous functions, the source data and place it in a public repository such as Github (see at: https://github.com/)