**Level 2**

*Help yourself with the documentation and google! :)*

Import the data in 2 formats:

1) import each variable individually.

2) import a matrix with all the variables

Remember to import only the values and not the text (for the matrix)

and remember to delete the text question in the second row, when you import the individual variables.

Import the following variables from the data set provided (COVID\_emotion\_EXCEL):

1) Go to "Liver Editor" in the menu --> Open--> open the excel file with the data. a window will open.

2) in the upper left corner of the matrix, there is an empty space bewteen the 1st column and 1st row (A,1). if you click it, you can select the whole matrix

3) press "ctrl" and de-select the first two rows of the matrix. you don't want "strings" in the matrix at this point.

4) in the menu go to import--> output data. select "numeric matrix". then go on the green V and press "import selection"

5) repeat the process changing the output types. you need to import the data also as "vectors".

now you should see all your variables in the workspace on the bottom right.

in the main matlab menu, go to "home"--> save workspace.

**It's very important to save your variables, otherwise you have to re-import them every time and it is VERY annoying!**

Use the excel version of the data to orient yourself with the labels.

**Exploratory data analysis**

Plot a histogram of the sample data with a normal density fit for each variable (excluding gender, age, and political orientation)

for %...

%histfit()

end

Plot the age of participants

%hist or normal plot

and exclude people <18 and >65 yo. You can write a for loop.

for i = % to the ength of the participants

 if %...

 ...

 end

end

**Demographics**

Calculate the demographics of the data set.

how many males, females, and non binary participants there are?

female = 1; male = 2; non binary = 3

what is the political orientation of people?

 %question Q2.8 (political orientation: 1 = extremely liberal, 7 = extremely conservative)

What is the median age? and the mean? what is the standard deviation?

% for gender, and later for political believes, check out this command:

%summary(categorical(A)) %% where "A" is an array.

Plot the age by using a box plot.

Plot the 3 gender with 3 bar chart in the same plot.

check this out! you can plot multiple things on the X- axis

[Visualize summary statistics with box plot - MATLAB boxplot (mathworks.com)](https://www.mathworks.com/help/stats/boxplot.html)

%boxplot()

%...

%you can also use "hold on" to plot multiple things in the same figure.

*If Else...*

write an IF / ELSE statement for the "age".

% We want to devide participants into 3 age groups.

%from 18 to 33 --> young

% 34 to 49 --> young-middle

% 50 to 65 --> middle- older

% Write a statement that says "if the participant's age is

% less than 33, display "young", if it's between 34 and

% 49, then display 'young-middle', else display middle-older.

%In addition, plot the count of the 3 age groups, with a bar-chart.

%HINT: wrap the if else statement in a for loop, to loop through each

%participant.

Write a while loop (focus on age variable).

While the participant number is > 210, break the loop.

while

 break

end

**You can try to visualize the relations among variable with a multivariate matrix figure**

Check gplotmatrix() in the documentation

or this general link: [Visualizing Multivariate Data - MATLAB & Simulink Example (mathworks.com)](https://www.mathworks.com/help/stats/visualizing-multivariate-data.html)

**Fisher transformation and outliers**

1) normalize the scores (transform all the scores to Z-scores, so that they have an average of 0 and a standard deviation of 1). This is useful so you can compare variables that are measured on different ways, with different scales

To do it, use the "zscore()" function.

% e.g., "Z\_data = ... " and use the zscore() function, which take the

% variable that you want to tranform as argument. you can tranform the

% whole matrix (check documentation)

2) exclude the values that are above 3 standard deviation from the mean. you can use the function removeoutlier, or whatever works for you.

[Detect and remove outliers in data - MATLAB rmoutliers (mathworks.com)](https://www.mathworks.com/help/matlab/ref/rmoutliers.html)

Once the outliers are removed, go back to the raw data and calculate mean, max, min, and standard deviation for every variable in the data set. To do that, write a for loop that loops through each column and calculate these stats, then print the results. Use "fprintf()" to print the statement (e.g., 'the mean for column "i" is "mean"...).

BTW, check the documentation (e.g., for the mean)... it is possible to calculate these things for the whole data-set at once!

% for

% mean()

% ...

% fprintf([...])

% end

**hypothesis (chose one)**

Political orientation influences the way we think of emotions.

Age group influences the way we think of emotions.

Gender influences the way we think of emotions.

e.g., Do conservative people attribute to "fear"/"anger" more negative (or positive) attributes than liberal people?

USE THE Z-SCORES.

1) divide political orientation/gender/age in 2 or 3 groups based on scores.

2) calculate the mean for the variables related to the attributes for fear and anger

Ps: did you know that you can calculate summary stats organized by group?

[Summary statistics organized by group - MATLAB grpstats (mathworks.com)](https://www.mathworks.com/help/stats/grpstats.html)

%

Plot a figure for FEAR. the figure is composed by 6 subplots (for each attributes). each plot shows the scores for conservative vs liberal (or gender or age) (plot them with 2 different colors). You can chose the plot that you prefer (e.g., bar-chart etc.)

%FEAR

Figure

%SEARCH "SUBPLOT" IN THE DOCUMENTATION"

%use "hold on" to plot multiple things in the same plot.

%... then do the same for anger

Test if there is a statistical difference in mean between conservative vs liberal (or gender or age) regarding fear attributes

(you can use Ttests or ANOVAs. check the documentation!).

Then do the same for anger.

**Correlations / simple linear regressions**

Do the way that people think of fear and anger (e.g., want feel more/less) shapes how much they actually feel these emotions, in a covid context (Questions Q.92\_1 and Q.92\_2)?

use pearson correlation (r) (check documentation for "corrcoef" or correlation in general)

OR use simple linear regression (check documentation for simple linear model OR Least squares fitting)

Plot the figures (you can use scatter(). ). Remember to plot the regression (or correlation line). try to find out how to do it with the documentation or google.

If you want to plot a correlation matrix, try out **heatmap(corrcoef(X))**, where X is your data matrix.

**At this point, we can write a poster together and submit to a conference (e.g., RISE, SAS). You are very much encouraged to do so, because it's a lot of fun, it's rewarding, and it is very valuable in the CV (for example if you want to apply for grad school).**

**So... let me know :)**