Assignment 2 - Shell Scripting

For this assignment you can use **sed, grep** and **awk** commands in a shell script to do the desired tasks. You can use **man** pages to know more about these commands. You will also need to use output redirection ($>$, $>>$, or some equivalent method) to write to a file.

[1] Write a shell script and name it as **lab.sh**, which will take exactly two filenames as arguments. If exactly two names are not provided, the script prints a usage message and exits, otherwise the script do what is described next. Usage message tells the user how to use the script.

[2] Till this step, you have provided two filenames to the script. The first file name is an input file, and the second one is an output file. You need to check whether the input file exists or not. If the input file does not exist (How do you test that in bash?), the script generates an appropriate error message and exits. If the input file exist then proceed to next step.

[3] The input file is CSV (Comma Separated Values) file representing a dataset of colleges. The data is divided into different columns:

`Name`, `PredominantDegree`, `HighestDegree`, `FundingModel`, `Region`, `Geography`, `AdmissionRate`, `ACTMedian`, `SATAverage`, `AverageCost`, `Expenditure`, `AverageFacultySalary`, `MedianDebt`, `AverageAgeofEntry`, `MedianFamilyIncome`, `MedianEarnings`.

We have provided a sample file college.csv for use. You can assume that there are NO errors in the input file if it exists (See [2] for the case when the file does not exist). The input file name can be other than college.csv, actual data can be different, but the data format (columns) will remain the same.
Now, your script should print the following fields in ALL lines in the input file to output file (the second argument) by replacing ","(comma) to " "(space).

`Name`, `PredominantDegree`, `HighestDegree`, `Region`, `Geography`, `AdmissionRate`, `AverageCost`, `Expenditure`.

If any data already exists in the output file then **it should be overwritten.**

[4] Next the script should save the Name of the college whose `HighestDegree` is **Bachelor's** to the output file provided as second argument. Be careful that the script **should not overwrite** the output file,

[5] The script should compute the average admission rate by grouping the colleges with respect to `Geography` Column of the dataset and save it to the output file(second argument). Be careful that the script **should not overwrite** the output file. The output should be saved as follows:

Geography: Average Admission Rate
Mid-size City: XYZ
Small City: ABC
.....

Note that XYZ,ABC etc should be replaced by actual output. This way you have to save average admission rate for all geographical area.

[6] The script should save the top five colleges to the output file who have maximum `MedianEarnings`. Be careful that the script **should not overwrite** the output file.

You can create one or more **awk** script files, or any other auxiliary file if needed. Submit all the source files in the zipped format. There is no need to submit college.csv or any generated output file. The directory must contain lab.sh, any auxiliary awk or other script files you used.