Please provide code used for all your answers (in R if possible). There are 3 questions.

## QUESTION 1

Write a function that groups observations in a data frame and returns a subset of the observations. The function must first group the observations based on one or more columns (for clarity this will be referred to as ***G\_1***). Within each ***G\_1*** it must then identify each unique combination of values based on one or more columns (this will be referred to as ***g\_2***; the columns can include any column, including those used for ***G\_1***). Finally, it must return the first ***n*** observations for every ***g\_2***.

For this question you are expected to use the provided mtcars-x.csv file.

For example, this is what the function should return for the model with the lowest 'drat' for each ‘gear’ value, grouped by 'cyl':

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **mpg** | **cyl** | **disp** | **hp** | **drat** | **wt** | **qsec** | **Vs** | **am** | **gear** | **carb** |
| *Merc 240D* | 24 | 4 | 147 | 62 | 3.69 | 3.2 | 20 | 1 | 0 | 4 | 2 |
| *Toyota Corona* | 22 | 4 | 120 | 97 | 3.70 | 2.5 | 20 | 1 | 0 | 3 | 1 |
| *Lotus Europa* | 30 | 4 | 95 | 113 | 3.77 | 1.5 | 17 | 1 | 1 | 5 | 2 |
| *Valiant* | 18 | 6 | 225 | 105 | 2.76 | 3.5 | 20 | 1 | 0 | 3 | 1 |
| *Ferrari Dino* | 20 | 6 | 145 | 175 | 3.62 | 2.8 | 16 | 0 | 1 | 5 | 6 |
| *Mazda RX4* | 21 | 6 | 160 | 110 | 3.90 | 2.6 | 16 | 0 | 1 | 4 | 4 |
| *Dodge Challenger* | 16 | 8 | 318 | 150 | 2.76 | 3.5 | 17 | 0 | 0 | 3 | 2 |
| *Maserati Bora* | 15 | 8 | 301 | 335 | 3.54 | 3.6 | 15 | 0 | 1 | 5 | 8 |

Or, if we remove all unused columns:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **drat** | **cyl** | **gear** |
| *Merc 240D* | 3.69 | 4 | 4 |
| *Toyota Corona* | 3.70 | 4 | 3 |
| *Lotus Europa* | 3.77 | 4 | 5 |
| *Valiant* | 2.76 | 6 | 3 |
| *Ferrari Dino* | 3.62 | 6 | 5 |
| *Mazda RX4* | 3.90 | 6 | 4 |
| *Dodge Challenger* | 2.76 | 8 | 3 |
| *Maserati Bora* | 3.54 | 8 | 5 |

The function must at least take the following parameters:

* the dataframe to subset (the provided mtcars-x.csv)
* which column(s) to group by (‘cyl’ in the example above)
* which column(s) to use within each group to determine the unique combinations (‘gear’ in the example above)
* the number of records to return for each unique combination (1 in the example above)

Use your function to subset the provided file to get 2 models with the lowest ‘mpg’ for each ‘gear’ and ‘carb’ combination for cars with 8 cylinders, i.e. your answer must have at most 2 records for each ‘gear’ and ‘carb’ combination.

## QUESTION 2

Load the provided Lorem Ipsum text and answer the following questions.

1. How many paragraphs are in the text?
2. How many sentences is it comprised of?
3. Which paragraph has the most sentences?
4. How many words are in the text?
5. Which are the 5 most and least common words, with how many occurrences?
6. Which are the longest words in the text?

## QUESTION 3

Write efficient R code to evaluate the following when n = 250.