Data Science - Practice 2 (R Basic I-2)

Make sure you not only just “write down” the R code but also “explain the answer with your own language”. All answers without explanation will not be accepted.

Problem

Import 'country_set.RData’. It is a DataFrame, which contains various national indicators and continental information.

Variable	Description
code	country's code
country_name	country's name
GDP	Gross Domestic Product per capita
life_expect	Life expectancy
Population	Population
CO2	CO2 emission quantity (estimated)
battle_death	a death in battle (per 100,000)
child.per.women	Number of children per woman
programmable.aid	National program aid

< Question 1 >

(1) Get first 3 rows of the country_set data & (2) Get 10th, 11th, 12th,13th row of the country_set data.

Expected Result

(1)

```
  code country_name     continent   GDP life_expect population    CO2 battle_death
1  afg  Afghanistan          Asia  1757       61.22   35400000   8660          9.45
2  alb      Albania        Europe 11357       78.12    2890000   4540          0.13
3  arg    Argentina South America 18645       76.54   43500000 200000          0.00
  child.per.woman programmable.aid
1            4.64       3663.25163
2            1.71        277.18911
3            2.29         59.06856
```

(2)

```
   code country_name continent   GDP life_expect population   CO2 battle_death child.per.woman programmable.aid
10  bfa Burkina Faso    Africa  1642       61.18   1.86e+07  3380       0.0000            5.35         970.8594
11  bgd   Bangladesh      Asia  3424       72.89   1.58e+08 76100       0.1650            2.10        3088.7784
12  bgr     Bulgaria    Europe 17793       74.86   7.15e+06 45300       0.0286            1.56         850.5670
13  bhr      Bahrain      Asia 43732       79.42   1.43e+06 31500       0.0000            2.03         850.5670
```

< Question 2 >

(1) Get variables names of the country_set data & (2) Get the name of last variable of country_set.

Expected Result

(1)

```
 [1] "code"           "country_name"     "continent"        "GDP"              "life_expect"      "population"
 [7] "CO2"            "battle_death"     "child.per.woman"  "programmable.aid"
```

(2)

```
 [1] "programmable.aid"
```

< Question 3 >

Write down a R code that returns the structure of the country_set data and (2) the summary of the data set. If you find anything strange either from structure or summary, write them down.

Expected Result

(1)

```
'data.frame':   126 obs. of  10 variables:
 $ code            : chr  "afg" "alb" "dza" "arg" ...
 $ country_name    : chr  "Afghanistan" "Albania" "Algeria" "Argentina" ...
 $ continent       : chr  "Asia" "Europe" "Africa" "South America" ...
 $ GDP             : int  1757 11357 13940 18645 8159 44606 44671 16132 43732 3
424 ...
 $ life_expect     : num  61.2 78.1 77.4 76.5 75.4 ...
 $ population      : int  35400000 2890000 40600000 43500000 2940000 24300000 8
750000 9740000 1430000 158000000 ...
 $ CO2             : num  8660 4540 148000 200000 5180 413000 67400 37200 31500
 76100 ...
 $ battle_death    : num  9.45 0.13 3.41 0 0 0 0 0.0726 0 0.165 ...
 $ child.per.woman : num  4.64 1.71 2.78 2.29 1.63 1.85 1.49 2.08 2.03 2.1 ...
 $ programmable.aid: num  3663.3 277.2 108.3 59.1 373.1 ...
```

(2)

```
    code          country_name        continent             GDP         life_expect      population
Length:127         Length:127         Length:127         Min.   :    0   Min.   : 0.00   Min.   :0.000e+00
Class :character   Class :character   Class :character   1st Qu.: 5449   1st Qu.:69.52   1st Qu.:4.345e+06
Mode  :character   Mode  :character   Mode  :character   Median :14200   Median :75.32   Median :1.130e+07
                                                         Mean   :19942   Mean   :73.42   Mean   :5.250e+07
                                                         3rd Qu.:29239   3rd Qu.:79.61   3rd Qu.:3.590e+07
                                                         Max.   :93941   Max.   :84.69   Max.   :1.410e+09
      CO2           battle_death      child.per.woman programmable.aid
Min.   :      0   Min.   : 0.0000   Min.   :0.000   Min.   :   0.0
1st Qu.:   6875   1st Qu.: 0.0000   1st Qu.:1.665   1st Qu.: 275.4
Median :  25300   Median : 0.0000   Median :2.050   Median : 850.6
Mean   : 256279   Mean   : 1.8796   Mean   :2.547   Mean   : 843.9
3rd Qu.: 133500   3rd Qu.: 0.2225   3rd Qu.:2.970   3rd Qu.: 850.6
Max.   :9710000   Max.   :90.9000   Max.   :7.240   Max.   :4710.9
```

< Question 4 >

(1) Check whether there is a missing value in “code” variable. If so, (2) find them and (3) remove them if necessary.

< Question 5 >

(1) Find the unique list of continent variable & (2) find the frequency of continent variable (meaning the number of data observation correspond to each unique element).

Expected Result

(1)

```
[1] "Asia"          "Europe"        "Africa"        "South America"
[5] "Oceania"       "North America"
```

(2)

```
        Africa           Asia         Europe North America        Oceania
            31             34             35             13              4
South America
             9
```

< Question 6 >

What are the countries with the longest and shortest life expectancy (life_expect)? Write down a R code that produces expected result as shown below.

Expected Result

```
    code             country_name continent   GDP life_expect population   CO2 battle_death child.per.woman
104  sgp                Singapore      Asia 84704       84.69    5650000 37500          0.0             1.25
19   caf Central African Republic    Africa   731       51.68    4540000   297         29.9             4.87
    programmable.aid
104          850.567
19           273.657
```

< Question 7 >

(1) Measure the average of GDP, life_expect, population, child.per.woman & (2) Convert it into DataFrame shape (Hint: use data.frame() function).

Expected Result

(1)

```
          GDP     life_expect      population child.per.woman
 2.010002e+04    7.400183e+01    5.291611e+07    2.566905e+00
```

(2)

```
                average.values
GDP                2.010002e+04
life_expect        7.400183e+01
population         5.291611e+07
child.per.woman    2.566905e+00
```

< Question 8 >

Find the list of Asian countries with the value of 'battle_death' equals to 0.

Expected Result

```
 [1] "Armenia"     "Bahrain"     "Brunei"      "China"       "Cyprus"      "Iran"        "Japan"
 [8] "Kazakhstan"  "Kuwait"      "Mongolia"    "Malaysia"    "Oman"        "Singapore"   "Timor-Leste"
[15] "Turkey"
```

< Question 9 >

Create a list called 'country_list' which contains each continent's information only. (don't worry about the order of list. Make sure that 'country_list' has 6 list elements).

Expected Result

```
> country_list
[[1]]
    code country_name continent   GDP life_expect population     CO2 battle_death child.per.woman programmable.aid
1    afg  Afghanistan      Asia  1757       61.22   35400000    8660       9.4500             4.64       3663.25163
5    arm      Armenia      Asia  8159       75.37    2940000    5180       0.0000             1.63        373.09101
8    aze   Azerbaijan      Asia 16132       70.62    9740000   37200       0.0726             2.08        182.79669
9    bhr      Bahrain      Asia 43732       79.42    1430000   31500       0.0000             2.03        850.56700
10   bgd   Bangladesh      Asia  3424       72.89  158000000   76100       0.1650             2.10       3088.77836
```

< Question 10 >

Create names for the list elements of ‘country_list’. List name should be the name of continent.

Expected Result

```
> names(country_list)
[1] "Asia"          "Europe"        "Africa"        "South America" "Oceania"       "North America"
> country_list
$Asia
    code country_name continent   GDP life_expect population     CO2 battle_death child.per.woman programmable.aid
1    afg  Afghanistan      Asia  1757       61.22   35400000    8660       9.4500            4.64        3663.25163
5    arm      Armenia      Asia  8159       75.37    2940000    5180       0.0000            1.63         373.09101
8    aze   Azerbaijan      Asia 16132       70.62    9740000   37200       0.0726            2.08         182.79669
9    bhr      Bahrain      Asia 43732       79.42    1430000   31500       0.0000            2.03         850.56700
```