Practical Data Science
(Explorative Data Analysis)

Write R Scripts or use R to perform any mathematical operations while solving the following problems.

Problem1: Exploring Religious Terror Attacks

Go through the following problem of kaggle:
https://www.kaggle.com/argolof/predicting-terrorism
Do the following tasks:

a. Download the dataset from the following link:
b. Load the dataset and do the required type conversions
c. Explore all the attributes individually using univariate numeric and graphics
d. Do the following:
    a. Find top-10 countries with most attacks, most injured and most killed respectively and show them with plots
    b. Find top-10 cities with most attacks, most injured and most killed respectively and show them with plots
    c. Draw a plot that shows the relationship between killed and top-4 countries
e. Clean the description column using following steps:
    a. Normalize text: convert entire text to lower case
    b. Remove numbers
    c. Remove whitespaces
    d. Remove punctuation symbols
    e. Remove stop words
    f. Stem the words
f. Find out the most frequent words across all the attacks from the cleaned description column
g. Draw the wordcloud showing the frequencies of the words used across attacks and also show them via barplot
h. Find the correlations among each pair of words and cluster the words based on correlation

Problem 2: Explore Car dataset

Download the car.data from following link:
https://github.com/algorithmica-repository/datascience/tree/master/datasets
Here are the descriptions of the attributes of the car dataset:

buying: vhigh, high, med, low.
maint: vhigh, high, med, low.
doors: 2, 3, 4, 5more.

persons: 2, 4, more.
lug_boot: small, med, big.
safety: low, med, high.

The output class attribute can take one of the following values:

unacc, acc, good, vgood

Do the following tasks:

a. Load the dataset into frame and convert all the attributes to factor type.
b. Explore all the attributes individually using univariate numerics and graphics.
c. Explore all the bivariate relationships numerically and graphically.
d. What features do you recommend for predicting class category and why?
e. What kind of patterns have you discovered with the above explorations?

Problem 3: Exploring Kidney data

Download the chronic_kidney_data.txt from following link:

https://github.com/algorithmica-repository/datascience/tree/master/datasets/

The description of the dataset can be found at following link:

http://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease

Do the following tasks:

a. Load the dataset into frame and convert all the attributes to factor type.
b. Explore all the attributes individually using univariate numerics and graphics.
c. Explore all the bivariate relationships numerically and graphically.
d. What features do you recommend for predicting the disease is chronic or not and why?
e. What kind of patterns have you discovered with the above explorations?