IC152 - Lab 5

Data Visualization & Recursion

Lab Submission

Questions 1 and 2 require you to submit two kind of files:
1. Python code
2. Output images of the plots. These an be created using any technique, for instance, screen/window capture.

One single .zip file named `rollnumber_lab5.zip` containing all the above files is to be uploaded on moodle.
Your code for Question 3 and 4 is to be sumitted on HackerRank. Nothing is to be submitted for Question 0.

Question 0

Import and print the Data.csv file[1] using the following snipped of code. Download Data.csv first into your working directory on your laptop.

```
import pandas as pd
# Use pandas to read the CSV
csvdata = pd.read_csv('Data.csv',sep=',')
# Convert dataframe into a numpy array
foo = csvdata[['State','A','B','C','D','E','F','G','H','I','J','K']].to_num
# Covert numpy array into a list (of lists)
data = foo.tolist()
# Access values as usual from data
# For eg. data[0][3] is 25
print(data[0][3])
```

The column names given as `A, B, C, D, E, F, G, H, I, J, K` in the `Data.csv` file have the following meaning:

[1] For more background information and the source of the given data, see

Climate Vulnerability Assessment for the Himalayan Region Using a Common Framework, https://dst.gov.in/sites/default/files/. For West Bengal, only the hilly region of West Bengal has been considered.

A Percentage crop area insured under all Insurance Schemes (2013-15)
B Percentage farmers taking loans (2015-16)
C Average person days per household under MGNREGA (2006-2016)
D Average Percentage area with $>$ 30% slope
E Road Density
F Population density (2011): person/sq. km
G Percentage of marginal farmers (2011-12)
H Livestock to human ratio (2017-18)
I Per Capita Income (2014-15)
J Number of Primary Health Centres per 100,000 households (2017-18)
K Percentage of women in the overall workforce (2011)

Question 1

For this question use `data` of Question 0.

1. Write functions working on `data` which provide information on (a) population density, (b) percentage of marginal farmers and (c) percentage of women in the overall workforce by giving:

   - *Highest* = the state with the highest value
   - *Lowest* = the state with the lowest value
   - *Median* = the median
   - *Average* = the average value

   Don't use predefined pandas methods for these functions. Print the information in a way you find best represented.

2. For the given order of states create a single bar chart which takes as variables the

   (a) the percentage area with slope $>$ 30%

   (b) the road density

   For each state the two values should be displayed next to each other.

3. Create a bar chart for the states ordered by increasing percentage area with slope $>$ 30% showing the road density.

Question 2

Consider the following functions:

$$f(n) = \begin{cases} 1 & \text{if} \quad n < 2 \\ 1.65\, f(n-1) & \text{if} \quad n \geq 2 \end{cases}$$

$$g(n) = \begin{cases} 1 & \text{if} \quad n < 2 \\ g(n-1) + g(n-2) & \text{if} \quad n \geq 2 \end{cases}$$

$$h(n) = \begin{cases} 2 & \text{if} \quad n < 2 \\ 2\, h(n-2) & \text{if} \quad n \geq 2 \end{cases}$$

$$k(n) = \begin{cases} 3 & \text{if} \quad n < 3 \\ k(n-1) + k(n-3) & \text{if} \quad n \geq 3 \end{cases}$$

1. Implement f, g, h and k as recursive Python functions.
2. Create a scatter plot for each of the functions. The values of the x-axis should range from 0 to 9.
3. Display the curves of f, g, h and k within one plot. Add the legend that is, for every curve indicate which of the functions it represents. Again, for the range of the x-axis use the range from 0 to 9. After that try larger ranges as well. What are your observations?

Question 3

Write a recursive function `cStr` which prefixes a string with its captitalized reversal separated by an arrow (with a blank before and after the arrow):

```
cStr("Holi-to-come") = "EMOC-OT-ILOH -> Holi-to-come"
```

Question 4

Write a recursive function `scSS` which takes as input two strings, *w* and *s* and yields `"yes"` if *w* occurs in *s* as a scattered substrings and `"false"` otherwise. A string *w* occurs scattered in a string *s* if it may be obtained by cancelling some of the letters of *s*. For example, *abb* occurs scattered in *cadbeb* (cancel *c*, *d*, *e* and one of the occurrences of *b*).

If you want to do more...

1. Question 1: Do more analyses concerning the Indian Himalayan states with the data provided in the `Data.csv` file. For example, you might want to know whether a higher percentage of women in the overall workforce leads to a higher per capita income.

2. Question 2: Choose your own functions like the square function, exponential functions etc and plot them.

3. Question 4: Refine your solution by counting the number of scattered occurrences of w in s.