Question Paper

Exam Date & Time: 29-Nov-2018 (10:00 AM - 01:00 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

SCHOOL OF INFORMATION SCIENCES

FIRST SEMESTER MASTER OF ENGINEERING - ME (BIG DATA AND DATA VISUALIZATION)

Principles of Data Visualization [BDA 615.3]

Marks: 100 **Duration: 180 mins.**

END SEMESTER DEGREE EXAMINATION NOVEMBER 2018

Answer all the questions.

1) Give any five reasons for why visualization is a better tool for communicating information. (10)

2) Show the contexts in which you would choose the following Visualization techniques: Q-Q Plots, Stacked graphs, Choropleth Maps
Give an example for each to justify your response. (10)

3) Provide the formal definition and an example for each of the following operators found in the algebra of Grammar of Graphics: Cross operator, Blend operator (10)

4) With an example for each, state the rationale behind the following rules for producing better visualization: Do not trust the defaults, Use colors effectively, Avoid Chartjunk (10)

5) Briefly describe the three ingredients for successful visualization. (10)

6) What are scales in D3? List different scales you can create in D3. With an example show that how to create scales in D3. (10)

7) Write D3 script to create a scatter plot. (10)

8) Assume that "Vehicals.csv" contains information about different types of vehicles. Fields in csv are "Type of fuel", "Company", "Model", "Type of Vehicle" and "Cost".Write the script using pandas, to perform following tasks.
List different types of vehicles
List companies that manufacture "CAR" as vehicle type (10)

Give the costliest "model" among "CAR"

List the types of Vehicles company "TATA" produces.

9) Write a python code to create a pie chart. Generate random data. Visualization should include legend, user defined colors for each part of the pie chart. (10)

10) Consider the following Python dictionary data and Python list labels: (10)

data = {
'animal': ['cat', 'cat', 'snake', 'dog', 'dog', 'cat', 'snake', 'cat', 'dog', 'dog'],
'age': [2.5, 3, 0.5, np.nan, 5, 2, 4.5, np.nan, 7, 3],
'visits': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],
'priority':
['yes', 'yes', 'no', 'yes', 'no', 'no', 'no', 'yes', 'no', 'no']
}
labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']
Create a df from dictionary "**data**" which has the **index** labels.
Return the first 3 rows of the DataFrame df.
Select the data in rows [3, 4, 8] and in columns ['animal', 'age'].
Select the rows where the animal is a cat and the age is less than 3.

-----End-----