Question Paper

MANIPAL UNIVERSITY

SCHOOL OF INFORMATION SCIENCES
FIRST SEMESTER MASTER OF ENGINEERING - ME (Big Data and Data Analytics)
DEGREE EXAMINATION - NOVEMBER 2017
DATE : Tuesday, November 21, 2017
TIME : 10:00AM - 1:00PM
Principles of Data Visualization [BDA 615.3]

Marks: 100   Duration: 180 mins.

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 example show we can create scales in D3. (10)

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

8) Assume that "power.csv" contains information about power production. Fields in csv are "Type of fuel", "Country", "Place", "Mega Watt" and "Cost per Unit".Write the script using pandas, to perform following tasks.
List different types of fuels used.
Which country generates more power using 'Coal' as fuel used.
Which country generates maximum power(including all types of fuels).
List the place name which produces maximum power. (10)

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 DataFrame df from this dictionary data which has the index labels.
Change the age in row 'f' to 1.5.
Select the rows the age is between 2 and 4 (inclusive)
Count the number of each type of animal in df.

-----End-----