Birla Institute of Technology and Science-Pilani
First Semester 2023-2024
Comprehensive Exam
Course No CS F441 Course Title: Selected Topics From Computer Science
Type: Closed Book Time: 180 min Total Marks:60 (30%) Date: 10/12/2023
Name:_____ ID:_____

(Note: Questions are of MCQ type. A couple of the questions has multiple answers. You will be alerted about it in the question. Please note that for such questions you will receive credit only when you have correctly chosen all the correct answers.)
There is a total of 60 questions. Answer all questions. No negative marking.

Q01. What type of data is the ranking of players in a competition (first, second, third, etc.)?

    a. Quantitative

    b. Ordinal

    c. Nominal

Q02. What type of data is the GDP of countries?

    a. Quantitative

    b. Ordinal

    c. Nominal

Q03 What type of data is the gender of students in a class?

    a. Quantitative

    b. Ordinal

    c. Nominal

Q04. Which of the following <u>cannot</u> be inferred (in general) directly from the box plot?

    a. IQR (Inter-Quartile Range)

    b. Median

    c. Mean

    d. First quartile

Q05. What is IQR (interquartile range) of a 1D data set?

    a. Difference between 1st quartile and 2nd quartile values.

    b. Difference between 1st quartile and 3rd quartile values.

c. Difference between 2nd quartile and 3th quartile values.

d. Difference between maximum and minimum values.

Q06. What does the outline of the violin plot represent?

a. The density curve of the data.

b. Histogram of the data.

c. A smooth curve with two peaks.

Q07. Which of the following <u>does not represent</u> an aesthetics component in data visualization?

a. Position

b. Area

c. Color

d. Data value

Q08. If a student's z-score on an assessment is -1, the population's mean is 50 and the standard deviation is 10, what was the student's raw score? Note: *Z-score computation is a normalization technique that is often used for data normalization*.

a. 30

b. 40

c. 50

d. 60

Q09. Which of the following visual encoding (aesthetics) is the most preferred visualization of all types of data?

a. Area

b. Color

c. Length

d. Position

Q10. Which of the following visual encodings (aesthetics) is the most effective for visual communication of quantitative data?

a. Height

b. Area

c. Color Saturation

d. Color Hue

Q11. Which of the following visual encodings (aesthetics) is the most effective for visual communication of nominal data?

    a. Height

    b. Area

    c. Color Saturation

    d. Color Hue

Q12. In Line plot for time series data, what type of data is plotted against time?

    a. Ordinal

    b. Quantitative

    c. Nominal

Q13. What is the height of a bar in the histogram plot related to?

    a. Sum of the data values in the interval corresponding to the bar.

    b. The corresponding data interval of the bar.

    c. The number of data points in the interval corresponding to the bar.

Q14. CandleStick charts are commonly used time-series plots in financial sectors. Which of the following information a candlestick in the chart does not show?

    a. Maximum market value.

    b. Closing market value.

    c. Mean market value.

    d. Minimum market value.

Q15. In the multivariate time series line plots, which of the following is the commonly used aesthetic to distinguish the line plots corresponding to multiple variables?

    a. Length

    b. Position

    c. Color

    d. Area

Q16. In Bar plots, the independent variable is mapped to which of the following aesthetics (visual representation)?

    a. Position

    b. Area

    c. Length

Q17. In Bar plots, the dependent variable is mapped to which of the following aesthetics (visual representation)?

    a. Position

    b. Area

    c. Length

    d. Color

Q18. Which of the following plots is generally not used to show part-in-whole relationship?

    a. Line plot

    b. Bar plot

    c. Pie chart

    d. Area chart

Q19. What is the best plot for visual analysis of correlation between two data dimensions?

    a. Line plot

    b. Bar plot

    c. Scatter plot

    d. Histogram plot

Q20. Which of the following is not a measure of spread in a data dimension?

    a. Standard deviation

    b. Mean

    c. Variance

Q21. What is the range of covariance values between two data dimensions of a data table (or data frame)?

    a. -1 to +1

    b. 0 to 1

    c. 0 to $+\infty$

    d. $-\infty$ to $+\infty$

Q22. One cannot determine the correlation between two data dimensions from the best fit line between them.

    a. TRUE

b. FALSE

Q23. What does *k* in *k*-Cluster technique stand for?

a. Number of data dimensions

b. Number of clusters

c. Number of iterations.

Q24. For a data-frame with M data dimensions (or M data columns) each of which is N-elements long, what is the size of the covariance matrix?

a. M x N
b. N x N
c. M x M

Q25. In PCA (Principal Component Analysis), the goal is to linearly transform the data dimensions to remove any correlation between them, and then choose a couple of data dimensions from the transformed set. What can you say about the covariance matrix of the transformed data set?

a. It is a diagonal matrix.

b. It is an upper triangular matrix.

c. It is a lower triangular matrix.

d. It is an identity matrix

Q26. When using PCA for dimension reduction, the principal components are chosen from the transformed dimensions according to which of the following criteria?

a. In random order

b. Highest variance first

c. Lowest variance first

Q27. Star charts (a.k.a. radar chart, spider chart) are used for multi-dimensional data visualization in 2D. How is each multidimensional observation (i.e. each data row) visualized in this chart?

a. A polygon

b. A polar area section

c. A polar bar

Q28. A bubble chart is an extended scatter plot. Whereas scatter plot visualizes 2 dimensions of the data, the bubble chart can effectively visualize up to 4 dimensions. Which of the following aesthetics pairs can the additional 2 dimensions be mapped to?

    a. (area, length)
    b. (position, length)
    c. (area, hue)
    d. (position, hue)

Q29. What is the most preferred color?

    a. Red
    b. Black
    c. Blue
    d. Brown

Q30. Which of the following are *cool* colors? Note: *There may be multiple answers to this question. Choose all that apply*.

    a. Yellow
    b. Blue
    c. Green
    d. Red
    e. Orange

Q31. Which of the following are *warm* colors? Note: *There may be multiple answers to this question. Choose all that apply.*

    a. Yellow
    b. Blue
    c. Green
    d. Red
    e. Orange

Q32: What are the three primary colors in Color Wheel?

    a. Yellow, Red, Blue
    b. Red, Green, Blue
    c. Cyan, Magenta, Yellow

Q33: Qualitative color scheme is the most preferred color scheme for which type of data?

    a. Quantitative

    b. Ordinal

    c. Nominal data

Q34: Diverging color scheme chooses a pair of sequential colors that start with hues that are:

    a. Analogous colors

    b. Cool colors

    c. Complimentary colors

    d. Warm colors

Q35. Our visual system is better at estimating quantitative data when it is mapped to which of the following geometrical properties?

    a. Length

    b. Angle

    c. Area

Q36. Which of the Gestalt principles explain the visual grouping of points into cluster/clusters (if they exist) in a scatter plot?

    a. Closure

    b. Similarity

    c. Continuity

    d. Proximity

Q37. In a vertical bar chart, quantitative data values 400 and 500 are shown with bars of height 2.5cm and 5 cm respectively, and Y-axis minimum starts at 300. What is Tufty's Lie factor in this chart?

    a. 1

    b. 2

    c. 3

    d. 4

    e. 5

Q38. Which of the following is an example of the application of Tufty's principle of "small multiples to increase data density" in a plot?

    a. Line plot

    b. Pie chart

c. Scatter plot matrix
d. Bar plot

Q39. Which of the following answers is true about the major color deficiency problem?

a. It affects males and females equally.

b. It is a genetic problem.

c. 20% of the Indian population are affected by this problem.

Q40. Which of the following color deficiencies is the most common among the Indian population?

a. Red-Green

b. Green-Blue

c. Blue-Red

Q41. What type of geo-thematic data can be shown in Choropleth?

a. Nominal data.

b. Ordinal data.

c. Quantitative data.

d. All of the above.

Q42. Bubble map is best used for what type of geo-thematic data?
a. Nominal data.
b. Ordinal data.
c. Quantitative data.

Q43. Which visual property (aesthetics) Choropleth uses to show data related to a geographical region?

a. Area

b. Length

c. Color

Q44. Where is (0,0) of (longitude, latitude) located?

a. At South Pole

b. On Prime Meridian

c. At North Pole

Q45. What are the signs of longitude, latitude values of Indian sub-continent locations?

a. Negative longitude, Positive latitude

b. Negative longitude, Negative latitude

c. Positive longitude, Positive latitude

d. Positive longitude, Negative latitude

Q46. What is the range of Latitude values of Earth locations?

a. 0 to +360
b. 0 to +180
c. -180 to +180
d. 0 to +90
e. -90 to +90

Q47. What is the range of Longitude values of Earth locations?

a. 0 to +180
b. -180 to +180
c. 0 to +90
d. 0 to +360
e. -90 to +90

Q48. How does Cartogram visually represent quantitative data related to geographical regions?

a. By adjusting their area
b. By using sequential color scheme
c. By using qualitative color scheme

Q49. Which of the following is useful for visually displaying the inter-relationships between data in a matrix?

a. Sunburst chart
b. Cluster layout
c. Chord Diagram

Q50. What type of data can be visualized using the Sunburst chart?

a. Social Network
b. Hierarchical Data
c. Flow Network

Q51. Circle packing may serve as an alternative for which of the following?

a. Cluster layout
b. Treemap chart
c. Chord diagram

d. Bubble chart

Q52. Which shape is used to represent nodes in Sankey diagram?

a. Rectangular

b. Circular

c. Fragment of outer part of a circle

Q53. Contours are used to visualize scalar functions of the type Z = f(X, Y). For example, elevation of the earth at each Latitude and Longitude points. What can you say about the statement: Contour lines corresponding to different Z values can cross each other?

a. TRUE

b. FALSE

Q54. A word cloud is a visual representation of a text. What does it visually represent?

a. Length of text

b. importance of words

c. Length of words

Q55. Which of the following is not essential to interactive visualization?

a. Filtering

b. Zoom and Pan

c. Details on demand

d. Color

e. Pointing and Selection

Q56. What is *lasso* tool's use in interactive visualization?

a. Selecting hidden data

b. Selecting data points in rectangular region

c. Selecting data points in non-rectangular region

Q57. Which of the following may negatively impact text visualization using Word Cloud?

a. Long text

b. complex text

c. Long words

Q58. What are the various components of data storytelling?

a. Narration

b. Visualization

c. Data

d. All of the above

Q59. What type of method is *brushing* in interactive data visualization?

a. Selection

b. Coloring

c. Zooming

Q60. Tooltips in interactive visualization are useful for providing:

a. Overview.

b. Zoom capability.

c. Filtering.

d. Detail on Demand.