Module 4 MCQ with Answers

- **What is the main purpose of data visualization?**

  A) To store data
  B) To make data understandable
  C) To collect data
  D) To delete data
  Answer: B

- **Which Python library is primarily used for data analysis?**

  - A) NumPy
  - B) pandas
  - C) Matplotlib
  - D) seaborn
  - **Answer: B**

- **What is the first step in the data wrangling process?**

  - A) Analyzing data
  - B) Cleaning data
  - C) Importing data as a DataFrame
  - D) Transforming data into graphs
  - **Answer: C**

- **Which measure of central tendency is calculated by summing all measurements and dividing by the number of observations?**

  - A) Mean
  - B) Median
  - C) Mode
  - D) Range
  - **Answer: A**

- **What type of data visualization shows the relationship between body mass and maximum longevity of various animals?**

  - A) Line chart
  - B) Bar chart
  - C) Scatter plot
  - D) Pie chart
  - **Answer: C**

- **Which of the following is a measure of dispersion?**

  - A) Mean

- B) Median
- C) Mode
- D) Standard deviation
- **Answer: D**

- **In a positive correlation, how do the variables move?**

  - A) In opposite directions
  - B) Independently
  - C) In the same direction
  - D) Randomly
  - **Answer: C**

- **Which library is NOT mentioned as useful for data visualization in Python?**

  - A) Matplotlib
  - B) Bokeh
  - C) Tableau
  - D) seaborn
  - **Answer: C**

- **What is the main advantage of visualizing data?**

  - A) Easier storage
  - B) Better understanding
  - C) Faster transmission
  - D) Enhanced security
  - **Answer: B**

- **What does the interquartile range measure?**

  - A) Difference between the highest and lowest values
  - B) Average of all values
  - C) Difference between the 75th and 25th percentiles
  - D) Most frequent value
  - **Answer: C**

- **What is the purpose of data wrangling?**

  - A) To clean and transform data for analysis
  - B) To collect raw data
  - C) To visualize data
  - D) To store data securely
  - **Answer: A**

- **Which of the following is an example of discrete data?**

  - A) Height of students
  - B) Temperature readings
  - C) Number of books

- D) Time taken to run a race
- **Answer: C**

- **Which plot is ideal for visualizing trends over time?**

  - A) Bar chart
  - B) Line chart
  - C) Pie chart
  - D) Histogram
  - **Answer: B**

- **What does a probability distribution function provide?**

  - A) The sum of all possible values
  - B) The average of all possible values
  - C) The probability of each possible event
  - D) The range of all possible values
  - **Answer: C**

- **Which of the following is NOT a measure of central tendency?**

  - A) Mean
  - B) Mode
  - C) Median
  - D) Variance
  - **Answer: D**

- **What is the median value in a dataset?**

  - A) The most frequent value
  - B) The middle value when data is ordered
  - C) The average value
  - D) The difference between the highest and lowest values
  - **Answer: B**

- **What does a scatter plot show?**

  - A) Comparison of multiple variables over time
  - B) Distribution of a single variable
  - C) Relationship between two variables
  - D) Proportion of parts to a whole
  - **Answer: C**

- **What is the square root of the variance called?**

  - A) Mean
  - B) Median
  - C) Standard deviation
  - D) Range
  - **Answer: C**

- **Which type of correlation indicates that variables move in opposite directions?**

  - A) Positive correlation
  - B) Negative correlation
  - C) Zero correlation
  - D) None of the above
  - **Answer: B**

- **What is a key benefit of using Python for data visualization?**

  - A) High cost
  - B) Limited libraries
  - C) Ease of use and speed
  - D) Complex syntax
  - **Answer: C**

- **Which statistical measure is most affected by outliers?**

  - A) Mean
  - B) Median
  - C) Mode
  - D) Interquartile range
  - **Answer: A**

- **What is the process of transforming raw data into a suitable representation called?**

  - A) Data cleaning
  - B) Data wrangling
  - C) Data visualization
  - D) Data mining
  - **Answer: B**

- **Which plot type is best for showing the distribution of a dataset?**

  - A) Line chart
  - B) Bar chart
  - C) Histogram
  - D) Pie chart
  - **Answer: C**

- **What does a high standard deviation indicate?**

  - A) Data points are close to the mean
  - B) Data points are spread out from the mean
  - C) Data points are evenly distributed
  - D) Data points are identical
  - **Answer: B**

- **What is the range in a dataset?**

- A) The average value
- B) The middle value
- C) The most frequent value
- D) The difference between the highest and lowest values
- **Answer: D**

- **Which type of chart is useful for comparing multiple variables at a single time point?**

  - A) Line chart
  - B) Bar chart
  - C) Scatter plot
  - D) Pie chart
  - **Answer: B**

- **What is the main use of a pie chart?**

  - A) Showing trends over time
  - B) Comparing numerical values
  - C) Showing proportions of a whole
  - D) Displaying relationships between variables
  - **Answer: C**

- **What kind of data is best visualized using a radar chart?**

  - A) Temporal data
  - B) Geospatial data
  - C) Multiple variables for multiple groups
  - D) Single variable over time
  - **Answer: C**

- **What does the term "data exploration" refer to?**

  - A) The initial investigation of data to discover patterns
  - B) The process of cleaning data
  - C) The visualization of data
  - D) The storage of data
  - **Answer: A**

- **What type of visualization is best for showing relationships among multiple variables?**

  - A) Line chart
  - B) Scatter plot
  - C) Bubble plot
  - D) Bar chart
  - **Answer: C**

- **Which measure of central tendency is least affected by outliers?**

  - A) Mean

- B) Median
- C) Mode
- D) Standard deviation
- **Answer: B**

- **Which tool is NOT mentioned for creating data visualizations in Python?**

  - A) NumPy
  - B) pandas
  - C) Matplotlib
  - D) Excel
  - **Answer: D**

- **What is the main characteristic of continuous data?**

  - A) It can only have certain values
  - B) It can take any value within a range
  - C) It is categorical
  - D) It is qualitative
  - **Answer: B**

- **Which of the following is NOT a type of probability distribution?**

  - A) Discrete
  - B) Continuous
  - C) Categorical
  - D) All are types of probability distribution
  - **Answer: C**

- **What does a correlogram show?**

  - A) The correlation between variable pairs
  - B) The distribution of a single variable
  - C) The relationship between three variables
  - D) The proportion of parts to a whole
  - **Answer: A**

- **Which plot type is best for visualizing geospatial data?**

  - A) Line chart
  - B) Scatter plot
  - C) Dot map
  - D) Bar chart
  - **Answer: C**

- **Which library is used for creating animated visualizations in Python?**

  - A) NumPy
  - B) pandas
  - C) Bokeh

- D) seaborn
- **Answer: C**

- **What type of chart would be best to visualize the frequency distribution of a single variable?**

  - A) Pie chart
  - B) Histogram
  - C) Line chart
  - D) Scatter plot
  - **Answer: B**

- **Which data visualization library in Python is known for its simplicity and ability to create static, animated, and interactive visualizations?**

  - A) pandas
  - B) Matplotlib
  - C) seaborn
  - D) Plotly
  - **Answer: D**

- **What does the term "data mining" refer to?**

  - A) The process of cleaning data
  - B) The exploration and analysis of large datasets to discover patterns
  - C) The visualization of data
  - D) The storage of data
  - **Answer: B**

- **Which Python library provides a high-level interface for drawing attractive and informative statistical graphics?**

  - A) pandas
  - B) Matplotlib
  - C) seaborn
  - D) NumPy
  - **Answer: C**

- **Which visualization would you use to compare the distribution of a variable across several categories?**

  - A) Bar chart
  - B) Box plot
  - C) Pie chart
  - D) Line chart
  - **Answer: B**

- **In data visualization, what is the purpose of using a color gradient?**

  - A) To enhance visual appeal

- B) To represent different categories
- C) To show variations in data values
- D) To separate different plots
- **Answer: C**

- **Which of the following is not a step in the data wrangling process?**

  - A) Data collection
  - B) Data cleaning
  - C) Data transformation
  - D) Data storage
  - **Answer: D**

- **What type of data is best represented by a heatmap?**

  - A) Temporal data
  - B) Categorical data
  - C) Numerical data across two dimensions
  - D) Text data
  - **Answer: C**

- **Which Python library is specifically designed for working with large datasets and is an extension of pandas?**

  - A) NumPy
  - B) Dask
  - C) Matplotlib
  - D) seaborn
  - **Answer: B**

- **Which plot is useful for visualizing the relationship between three variables, where two variables are represented by the x and y axes, and the third variable is represented by the size of the markers?**

  - A) Line chart
  - B) Scatter plot
  - C) Bubble plot
  - D) Histogram
  - **Answer: C**

- **What is a key characteristic of a box plot?**

  - A) It shows the relationship between two variables
  - B) It represents the distribution of a dataset
  - C) It displays the frequency of data points
  - D) It shows proportions of a whole
  - **Answer: B**

- **Which type of visualization is best for showing changes in a variable over time for multiple categories?**

- A) Bar chart
- B) Line chart
- C) Pie chart
- D) Scatter plot
- **Answer: B**

- **Which statistical measure can be visualized using a box plot?**

  A) Mean
  B) Standard deviation
  C) Interquartile range
  D) Mode
  E) **Answer: C**