1.

In a the range of values of a numeric variable of interest is usually laid out on the horizontal scale (x-axis). The scale is divided into sections called class. The vertical scale (y-axis) shows how many observations fall into each class. (1 mark)

Scatter plot

Bubble chart

~~Line chart~~

Histogram

None of the options.

Which of the following codes adds a legend at the top right of a clustered bar chart? (1 mark)

None of the options

legend("bottomright", MS, cex=0.8, fill=colors)

legend("top=right", MS, cex=0.8, fill=colors)

legend("topright", MS, cex=0.8, fill=colors)

legend("top&right", MS, cex=0.8, fill=colors)

A graphical depiction of a frequency distribution for numerical data in the form of a column chart is called:

Scatter plot

Line chart

Bubble chart

None of the options

Histogram

Which of the following parameters allows you to create a clustered bar chart? (1 mark)

~~None of the options~~

beside=TRUE

beside=FALSE

beside=OPTIONAL

beside=UNKNOWN

Horizontal and vertical bar plots are useful for the following except?(1 mark)

Illustrating differences between sets of values

Showing proportions or percentages of a whole

Comparing categorical or ordinal data

Displaying data over time

Which of the following is useful for displaying data over time? (1 mark)

None of the options

Vertical bar plots

Line charts

Stacked bar charts

Horizontal bar charts

What does the output of this code: quantile(cars$mpg) mean? (1 mark)

None of the options.

It breaks the data into four parts. The 25th percentile is called the first quartile, Q1; the 50th percentile is called the second quartile, Q2; the 75th percentile is called the third quartile, Q3; and the 100th percentile is the fourth quartile.

Create summary statistics for a given column in the data frame such as finding the mean.

Add new columns to the data frame.

Split the data frame by some variable, apply a function to the individual data frames and then combine the output.

In creating histograms in R using the 'hist' function, the parameter is used to specify the width of each bar.

width

break

xlab

xlim

ylim

When using the barplot function, what does the parameter 'cex' mean? (1 mark)

It is a number indicating the amount by which plotting text and symbols should be scaled relative to the default. 1=default, 1.5 is 50% larger, 0.5 is 50% smaller, etc.

It is a number indicating color for the axis

None of the options

It is a number indicating the line type

It is a number indicating the plot background color

Which of the following is true about the 'names.arg' parameter when using the barplot function?(1 mark)

~~None of the options~~

~~names.arg=(character vector) to specify the name of the data variables to plot~~

names.arg=(character vector) to label the bars

names.arg=(character vector) to label the title of the bar charts

names.arg=(list) to color the bars

What does the parameter 'ylim' mean when using the plot function in R?

ylim is the limits of the values of y used for plotting

ylim indicates whether both axes should be drawn on the plot

ylim is the limits of the values of x used for plotting

None of the options

ylim indicates the label in the vertical axis

A represents the proportion of the total number of observations that fall at or below the upper limit of each group.(1 mark)

Histogram

Cumulative relative frequency

~~Frequency distribution~~

~~Relative frequency distribution~~

~~Bubble chart~~

legend(x,y=NULL, legend, fill, col, bg). What does x and y represent? (1 mark)

x and y show the text of the legend

x and y are titles

x and y are labels for the axis

x and y are coordinates to be used to position the legend

None of the options

Histogram represents the frequency distribution of \_\_\_\_\_\_\_\_\_\_\_variables. Conversely, a bar graph is a diagrammatic comparison of \_\_\_\_\_\_\_ variables. Histogram presents numerical data whereas bar graph shows categorical data.

discrete; discrete

None of the options.

continuous; continuous

discrete, continuous

continuous; discrete

We may express the frequencies as a fraction, or proportion, of the total; this is called the: (1 mark)

Descriptive statistics

Frequency distribution

Exploratory machine learning

None of the options

Relative frequency

A tabular summary of cumulative relative frequencies is called a: (1 mark)

Frequency

Frequency distribution

Cumulative relative frequency distribution

Relative frequency

Relative frequency distribution

Bar charts are useful for comparing:(1 mark)

Categorical or ordinal data

None of the options

Incomplete data

Time series data

Ratio data

When creating a bar chart, it is useful to include the 'xlab' parameter. What does this parameter indicate?

x-axis limits

y-axis label

x-axis label

None of the options

x coordinate of the legend

type= " " in line charts can take the following values except:

type="Z"

type="o"

type="p"

type="n"

type="h"

What does the 'table' function accomplish?

Splits the data frame by some variable, apply a function to the individual data frames and then combine the output.

Adds new columns to the data frame.

None of the options

Uses the cross-classifying factors to build a contingency table of the counts at each combination of factor levels.

Creates summary statistics for a given column in the data frame such as finding the mean.

Which of the following is true about a stacked bar chart?(1 mark)

To create a stacked chart, the 'beside' parameter does not need to be included because the 'beside' parameter is FALSE by default

To create a stacked chart, 'stacked' parameter should be set to TRUE

None of the options

~~To create a stacked chart, 'beside' parameter MUST be set to FALSE since 'beside' is TRUE by default~~

To create a stacked chart, 'beside' parameter should be set to TRUE

Which of the following is true about contingency tables?

i. They are one of most basic statistical tool for summarizing categorical data

ii. They are a tabular method that displays number of observations in a data set for different subcategories of two or more categorical variables.

iii. Contingency tables can accept numerical variables but grouping variable must be categorical.

iv. Subcategories of variables must be mutually exclusive and exhaustive (i.e. each observation can be classified into only one subcategory, and, taken together over all subcategories, they must constitute the complete data set)

~~ii,iv~~

iii,iv

i, iii, iv

i,ii,iii,iv

~~ii, iii, iv~~

Which of the following statements about Pareto Analysis is not TRUE?

~~Pareto analysis can be used to identify inequalities in distribution of wealth.~~

Calculating cumulative proportions is part of Pareto Analysis.

~~Pareto analysis can be used to help businesses identify main sources of problems.~~

It is not necessary to sort data in increasing order to do Pareto Analysis.

Pareto analysis can be used to help businesses identify main sources of sales revenue.