

Final exam practice questions

1. Which clustering algorithm is known as the top-down approach?

- Divisive

Correct! Top-down clustering requires a method for splitting a cluster that contains all of the data and proceeds by splitting clusters recursively until each individual datum has been split into singleton clusters.

- Agglomerative

Incorrect. The agglomerative algorithm is known as the bottom-up approach. Each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

- Single link

Incorrect. Single link is a form of agglomerative clustering, which is a bottom up approach.

- Complete link

Incorrect. Complete link is a form of agglomerative clustering, which is a bottom up approach.

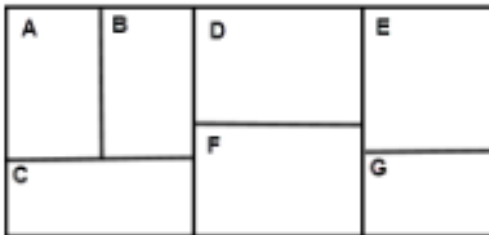
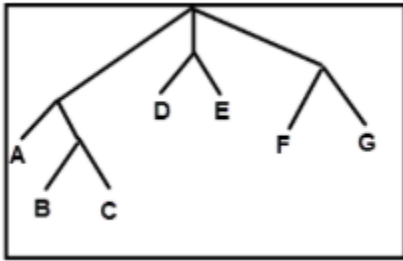


Figure 1

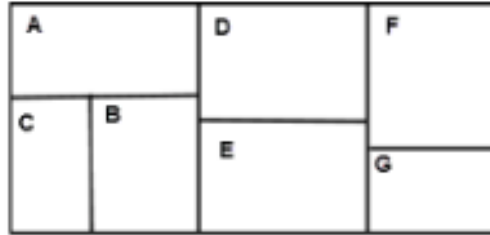


Figure 2

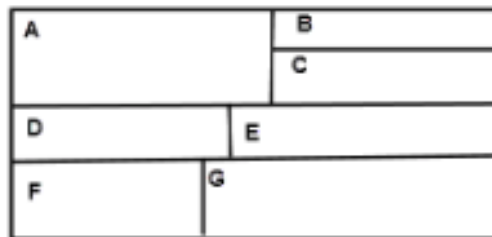


Figure 3

2. Which figure in Image 2 represents the treemap for the tree in Image 1 if the initial cut is vertical?

- Figure 1

Incorrect. This is not the corresponding treemap.

- Figure 2

Correct!

- Figure 3

Incorrect. This is not the corresponding treemap.

	A	B	C	D	E
A	1				
B	0.5	1			
C	0.6	0.3	1		
D	0.8	0.1	0.9 5	1	
E	0.9	0.2	0.3	0.5	1

3. For the similarity matrix between five objects given in Image 1, which two objects are *farthest* from each other?

- B and D

Correct! This pair of objects is the least similar, so it is the farthest pair in this matrix.

- A and C

Incorrect. This is not the farthest pair of objects.

- A and E

Incorrect. This is not the farthest pair of objects.

- D and E

Incorrect. This is not the farthest pair of objects.

4. Which visualization can be used to get a quick overview of the structure of the data?

- Treemap

Correct! Treemaps give a compact view of the part to a whole, and hence they provide a quick overview.

- Sunburst chart

Incorrect. Sunburst charts are more appropriate for a hierarchy with layers or categories within it, so they are best for more complex visualization.

- Cartogram

Incorrect. Cartograms are used for thematic mapping, such as travel time, population, etc.

- Time series

Incorrect. A time series is a sequence taken at successive equally spaced points in time, so it is appropriate for identifying changes over time.

5. What describes the set of tools and techniques that are used to analyze data with a geographic component?

- Geovisualization

Correct!

- Cartographics principles

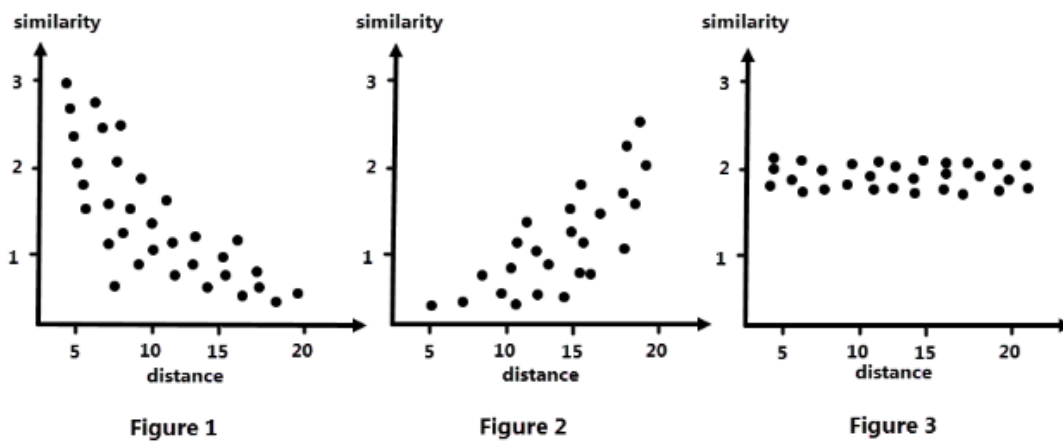
Incorrect. Cartographics principles are not used for data analysis.

- Isopleth map

Incorrect. Isopleth maps are used to show changes over space.

- Tobler's first law of geography

Incorrect. Tobler's first law of geography explains the relationship between similarity and distance.



6. Which graph can be used to explain Tobler's First Law of Geography?

- Figure 1

Correct!

- Figure 2

Incorrect. Similarity and distance are not necessarily positively correlated.

- Figure 3

Incorrect. There is a correlation between similarity and distance, and similarity is not necessarily constant.

7. Which type of map should be used to display the amount of a particular crop grown in a specific region?

- Thematic

Correct!

- General-purpose

Incorrect. General-purpose maps do not necessarily show statistics.

- Political

Incorrect. Political maps show country borders.

- Physical

Incorrect. Physical maps show the physical properties of region.

8. What is the *first step* in successful communication of map information?

- Determine the purpose of the map and the intended audience.

Correct! Determining the purpose of the map and the intended audience is the first step in successful communication of map information.

- Collect data appropriate for the map's purpose.

Incorrect. Appropriate data collection is not the first step in communicating map information.

- Consider only how the designer wants to visualize.

Incorrect. Considering only the designer's view is not the first step in communicating map information.

- Contrast the real world distribution of the phenomenon.

Incorrect. Considering the real world phenomenon is one of the first steps in communicating map information.

9. In geographic visualization, which components help in determining the initial point of a coordinate system?

- The projection, ellipse model, and datum

Correct! The initial point of the coordinate system is determined by the projection, ellipse model, and datum.

- The projection only

Incorrect. The purpose of projection is to provide a common basis for communication about a particular place or area on the earth's surface. On its own, it cannot help to determine the initial point of a coordinate system.

- The datum only

Incorrect. A datum is a specific, known point on or in the Earth that is used for reference. On its own, it cannot help in determining the initial point of a coordinate system.

- The projection and datum

Incorrect. A datum is a specific, known point on or in the Earth that is used for reference. Projection uses a datum as a reference point. The purpose of projection is to provide a common basis for communication about a particular place or area on the earth's surface. Without an ellipse model, these two alone cannot help in determining the initial point of a coordinate system.

10. Which type of value is a quantity and allows simple forms of computation?

- Internal data value

Correct! Internal data values are quantities and allow simple forms of computations (they do not support multiplication or division).

- Nominal value

Incorrect. Nominal data values are also known as categorical data, which can be sorted according to some categories, and they can be used for computations.

- Ordinal value

Incorrect. Ordinal data values can be put in some natural sequence but they do not allow any other type of computation

- Ratio data value

Incorrect. Ratio data values allow all forms of arithmetic computations, multiplication and division.

11. Which statement about line graphs is *most accurate*?

- Line graphs are useful in simple representation of time series data.

Correct! Line graphs show a simple linear representation of data that varies over time

- Line graphs cannot represent temporal data.

Incorrect. Line graphs are used to represent temporal data, which is data that varies over time

- Line graphs are only used to show time series data.

Incorrect. Line graphs are not limited to only showing time series data.

12. Which term describes a long run of increase or decrease in a dataset?

- Trend variation

Correct! Trend variations show a basic tendency of components to grow or decline over a period of time. It includes steady movements over a long time and excludes short-range oscillations.

- Seasonal variation

Incorrect. Seasonal variation refers to the phenomenon where the production and plan of a product changes based on a certain seasonal trend, depending on the characteristics of the product.

- Cyclic variation

Incorrect. Cyclic variation refers to any change in economic activity that is due to some regular and/or recurring cause.

- Random/irregular variation

Incorrect. Random variations are caused by erratic and irregular actions that are the result of random chance.

13. What is an example of a time series model?

- Naive approach

Correct! The Naive approach is a method used in time series models.

- Principal component analysis

Incorrect. Principal component analysis is a method used in unsupervised learning.

- Scatter plot

Incorrect. Scatter plots are used in multivariate analysis.

- Bar graph

Incorrect. Bar graphs are used to relate two variables.

14. Which pairs of strings have an edit distance of 3? Select all that apply.

- Saturday, Sunday

Correct! This string has an edit distance of 3.

- Sitting, Kitten

Correct! This string has an edit distance of 3.

- Stall, Table

Correct! This string has an edit distance of 3.

- Food, Money

Incorrect. This pair of strings does not have an edit distance of 3.

15. Which of the following is used to visualize thematic changes in large document collections?

- Theme river

Correct!

- Sparkline chart

Incorrect. Sparkline charts are not the preferred way to visualize thematic changes in large document collections.

- Bar chart

Incorrect. Bar charts are not the preferred way to visualize thematic changes in large document collections.

- Histogram

Incorrect. Histograms are not the preferred way to visualize thematic changes in large document collections.