MINING SPATIAL DATA

STQD6414 PERLOMBONGAN DATA

Assoc. Prof. Dr. Nurulkamal Masseran

Department of Mathematical Sciences

Universiti Kebangsaan Malaysia

INTRODUCTION:

- Spatial data is a data that involves space, location and geography.
- Spatial data includes "spatial objects" or "spatial fields".

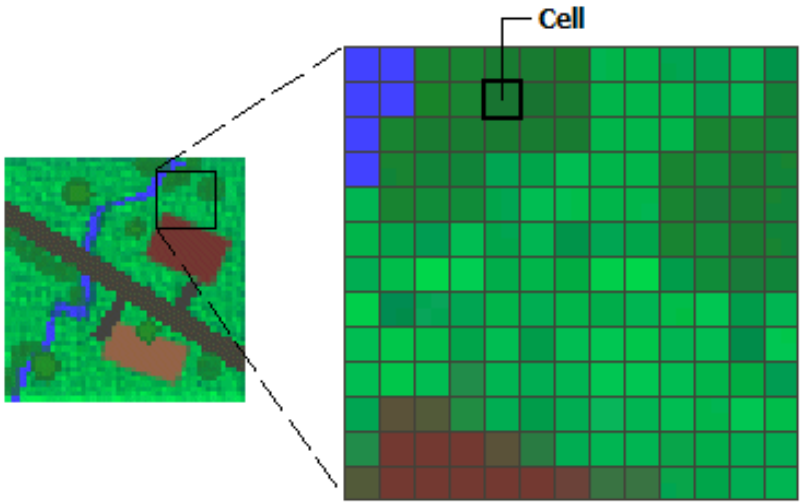
i) Spatial Objects: objects with a borders. Example: river, road, country, city, etc.

ii) Spatial Field: objects with no border. Example: altitude, temperature, air quality, etc.

- Spatial objects are usually represented in the form of vector data.
- Vector data consists of information such as “geometry” or “shape” related to a location, as well as information about the variables to be analyzed.
- Example: a set of vector data might describe the information about the border of countries in the world altogether with information about population sizes.

INTRODUCTION:

- While, spatial fields are commonly shown in a raster form.
- A raster consist a grid of equally sized rectangles (referred to as cells or pixels)
- Thus, raster data are determined based on the length and width of the image in units of pixels or the number of bits per pixel.
- Raster data is commonly used to represent spatially continuous phenomena such as elevation.

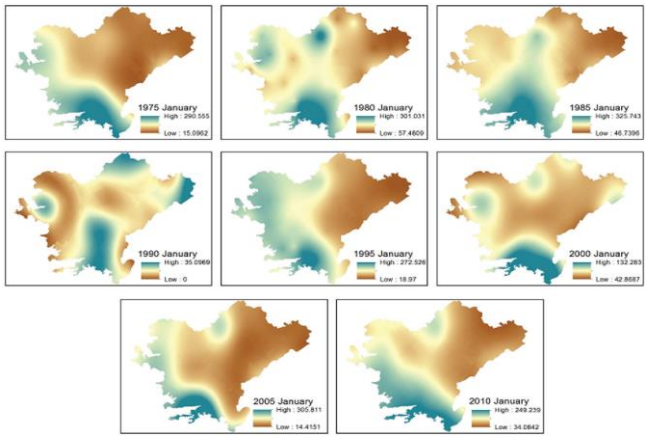


APPLICATIONS OF SPATIAL ANALYSIS:

- There are various applications of spatial analysis involving meteorological data, earth science, image analysis and etc.
- If spatial data is collected together with time information, it is known as space-time (spatio-temporal).

1. Meteorological Data:

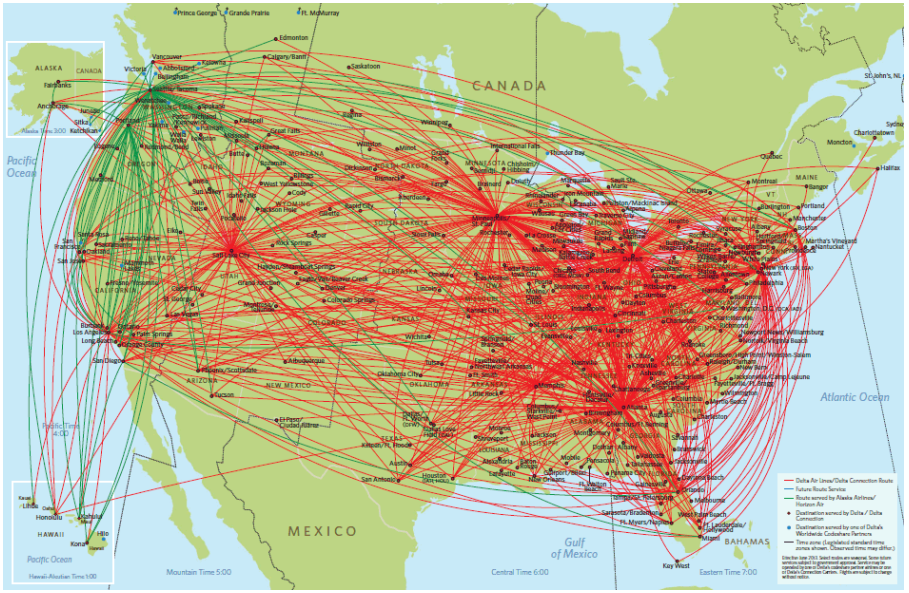
- Weather-related analysis, such as temperature, rainfall and pressure.
- Generally, this type of data is observed in different geographical locations.
- It needs to be analyzed simultaneously to extract useful information related to the country weather.



APPLICATIONS OF SPATIAL ANALYSIS:

2. Moving Objects Data:

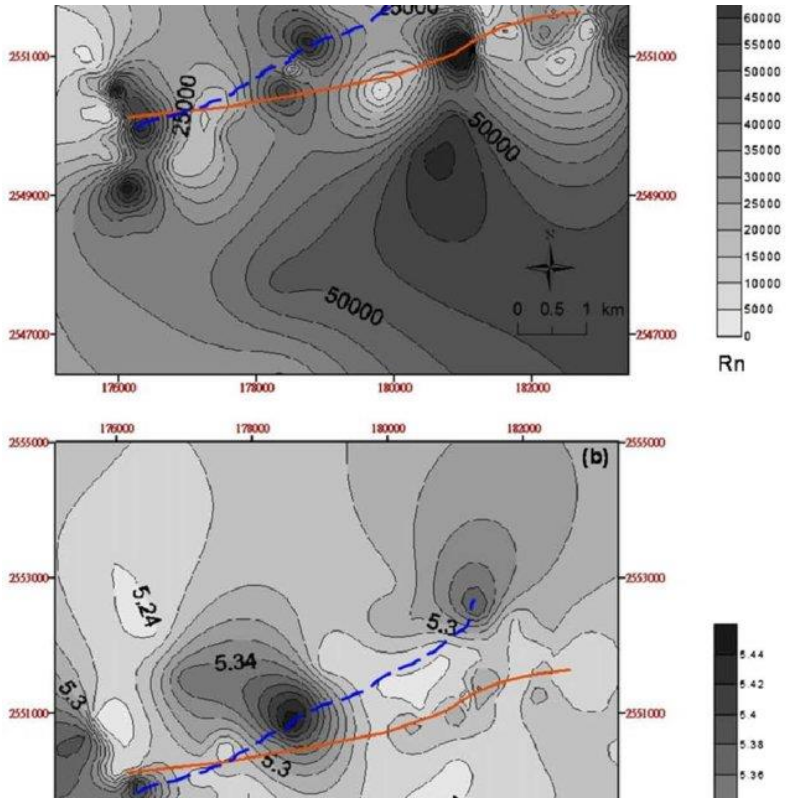
- Moving objects data can be obtain form trajectories path.
- Example: Flight trajectories, Waze, GPS and etc.
- These trajectories data can be analyzed to unearth information such as feature trends, or determining anomalous object paths.



APPLICATIONS OF SPATIAL ANALYSIS:

3. Earth Science Data:

- Soils in different spatial locations indicate different properties.
- Example: land altitude, forest density, and etc.

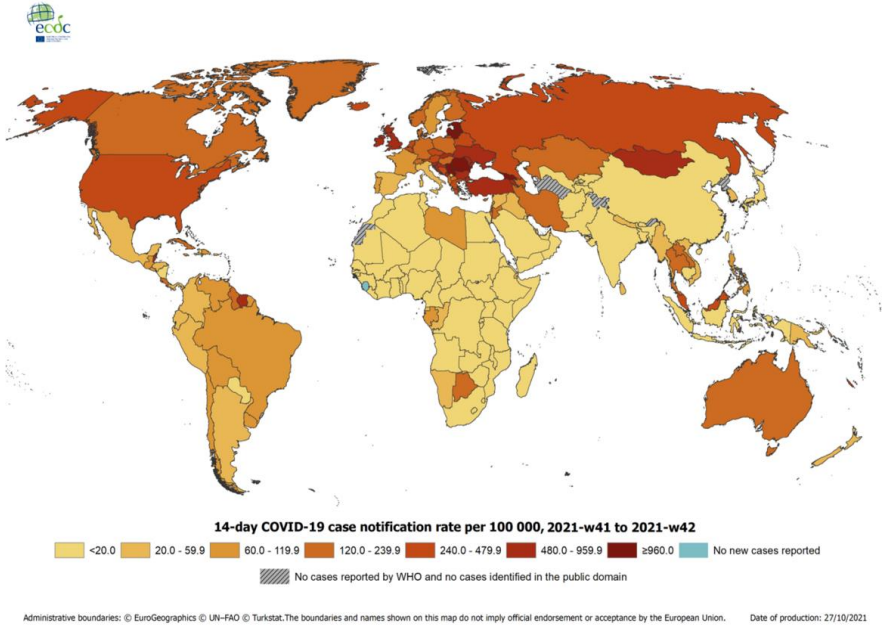


- Anomalies in Earth Science data show anomalous trends in human activities.
- Example: deforestation activities, anomalous vegetation trends, and etc.

APPLICATIONS OF SPATIAL ANALYSIS:

4. Disease Outbreak Data:

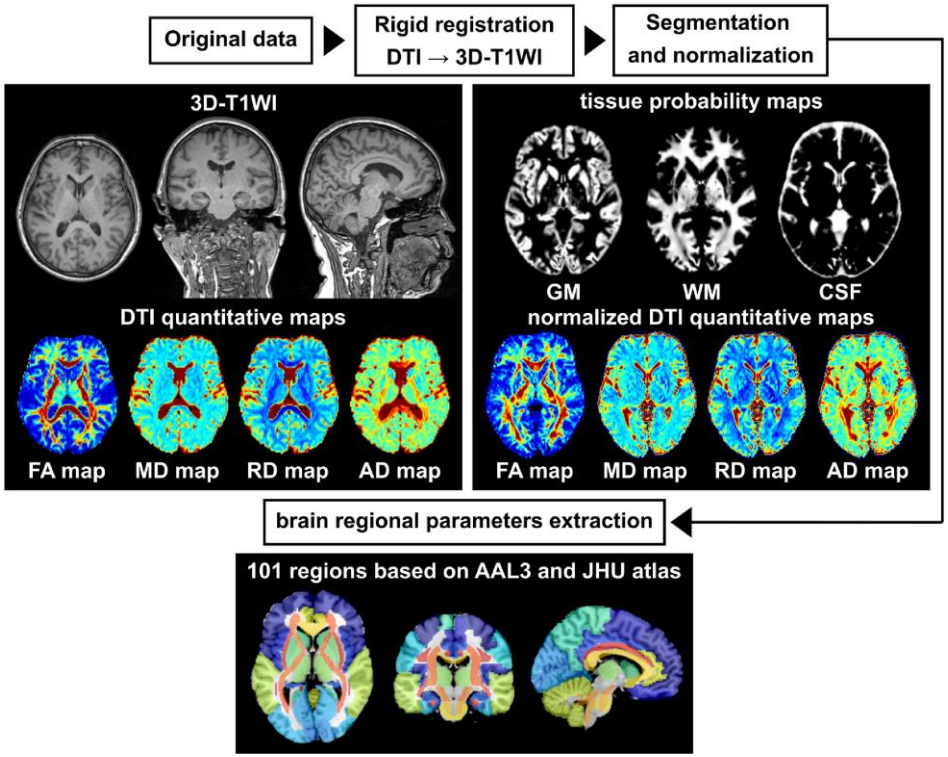
- Data on disease outbreaks is often aggregated by spatial location such as ZIP code, district and state.
- Analysis of trends in this kind of data provide information about the origins of the epidemic.
- Thus, it help the authorities in managing the risk in reducing the disease spreading.



APPLICATIONS OF SPATIAL ANALYSIS :

5. Medical Diagnostics Data:

- Magnetic resonance imaging (MRI) data are spatial data in 2 or 3 dimensions.
- The detection of unusual anomalies in these data can be helpful in detecting diseases such as brain tumors, onset of Alzheimer's disease, etc.

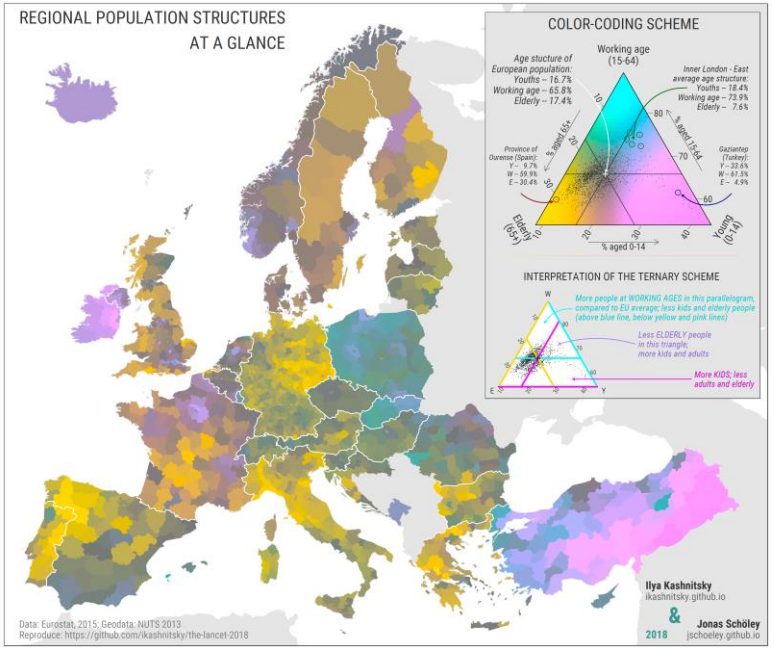


APPLICATIONS OF SPATIAL ANALYSIS :

6. Demographic Data:

- Demographic attributes such as age, gender, race and salary correspond to their spatial attributes provide information about the distribution of demographic patterns.

- This information is important for the government to develop a reasonable policy.
- For business, this information is important for marketing and business planning.

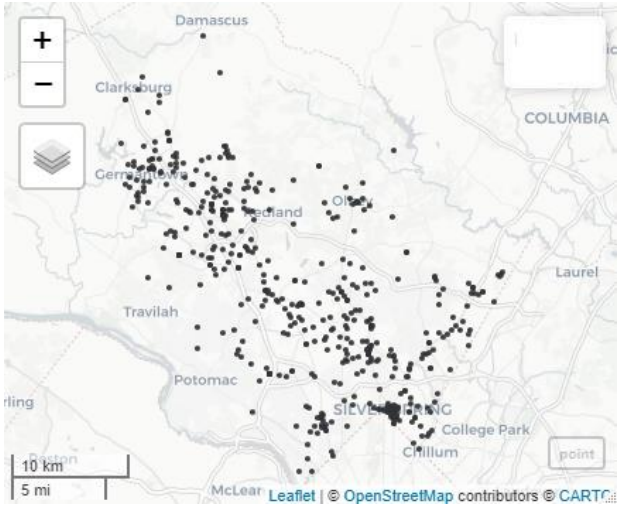


SPATIAL DATA: VECTOR FORM

- There are three main types of vector data in spatial analysis.

i) Point Data:

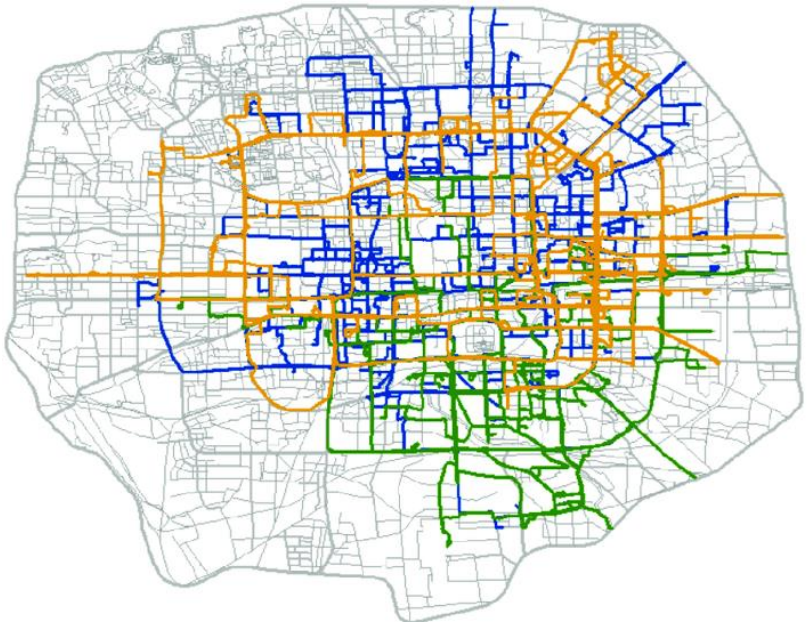
- Each point has a pair of coordinates, and their associated variables.
- Example: a dot may represent a location where an outbreak cases was detected. Other characteristics may include; the date it was detected, the information about the patient, such as; gender, race, social status and etc.
- This data can also consist a combination of several points in a multi-point structure, with a single attribute record.
- Example: all coffee shops in the city will be considered a combination of several points.



SPATIAL DATA: VECTOR FORM

ii) Lines Data:

- Lines data refer to one or more poly-lines or connected line segments.
- Example: rivers, roads and all tributaries are lines.



- Lines are shown as ordered sets of coordinates (nodes).
- Actual line segments can be determined and drawn on a map by connecting the dots.

SPATIAL DATA: VECTOR FORM

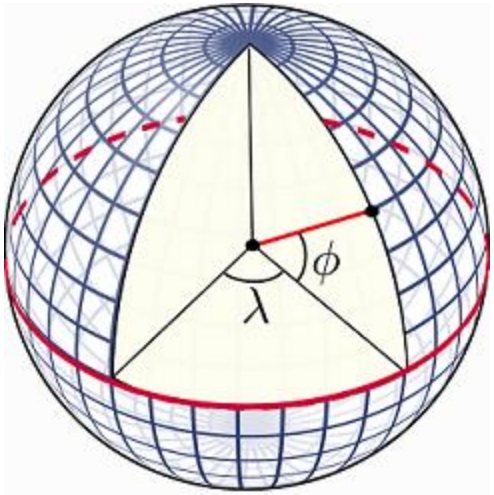
iii) Polygon data:

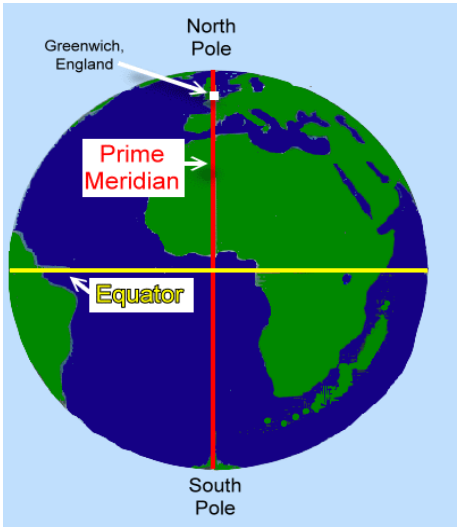
- Polygons refer to sets of closed polylines.
- The last pair of coordinates is equal to their the first pair of coordinates.
- Several polygons can be combined in a single geometry.
- Example: Malaysia is form by a combination of several states.

COORDINATE REFERENCE SYSTEMS(CRS):

i) Angular and Planar Coordinates:

- The CRS for geographic data is referred to longitude and latitude system.
- This CRS is known as the Angular system.
- Latitude is determined by the angle between the equatorial line and the perpendicular line that passes through a point from the center of the Earth.
- Longitude is the angle from the reference meridian line to the meridian line for the location.





COORDINATE REFERENCE SYSTEMS(CRS):

- The determination of the coordinates is made through the pair of angles and also the reference datum (model for the shape of the earth).
- The most commonly used datum is type WGS84 (World Geodesic System 1984).
- Projections for a 3-dimensional to 2-dimensional Angular System are referred to as planar systems ("Cartesian").

ii) Notation:

- The CRS planar system is defined from projection information, datum and some set of parameters.
- In R, commonly the notation of PROJ.4 is used of vector data and the notation of utm is used of raster data.

SPATIAL VECTOR DATA IN R:

- In R software, the sp class indicates a spatial data object.
- Among the important packages are sp, raster, spatial, rgdal, and many more.

- The three basic forms of spatial vector data in R are:

i) SpatialPoints

ii) SpatialLines

iii) SpatialPolygons

- On these basic of vector data, the value of new attributes can be added.

Example:

- SpatialPointsDataFrame and SpatialPolygonsDataFrame

SPATIAL RASTER DATA IN R:

- The three basic forms of Raster data in R are RasterLayer, RasterBrick, RasterStack.

i) RasterLayer:

- RasterLayer objects represent single-layer raster data.
- It describes basic parameters of raster data such as the number of columns, rows and the Coordinate Reference System (CRS).
- Attribute-related information can also be stored in the RasterLayer.

ii) RasterStack and RasterBrick:

- RasterStack and RasterBrick can be used to form a raster data from multi layers of data.

DATA MANIPULATION FOR SPATIAL VECTOR:

- Data manipulation is the process of converting data into a more organized form.
- The process depends on the objectives of the analysis.
- Some spatial vector data manipulation techniques:

i) Presenting spatial data as a data.frame format.

ii) Extract specific attributes.

iii) Add new attribute information.

iv) Delete any attribute.

v) Data integration.

vi) Select a subset of data.

vii) Data Aggregation.

viii) Map Manipulation.

ix) And many more.

DATA MANIPULATION FOR SPATIAL RASTER:

- Some techniques for raster data manipulation:

i) Extracts a single RasterLayer object from a RasterBrick or RasterStack object.

ii) Algebra in raster data.

iii) Add new values in the cell.

iv) Add new attribute information.

v) Crop and merge raster data.

vi) Descriptive functions.

vii) And many more.

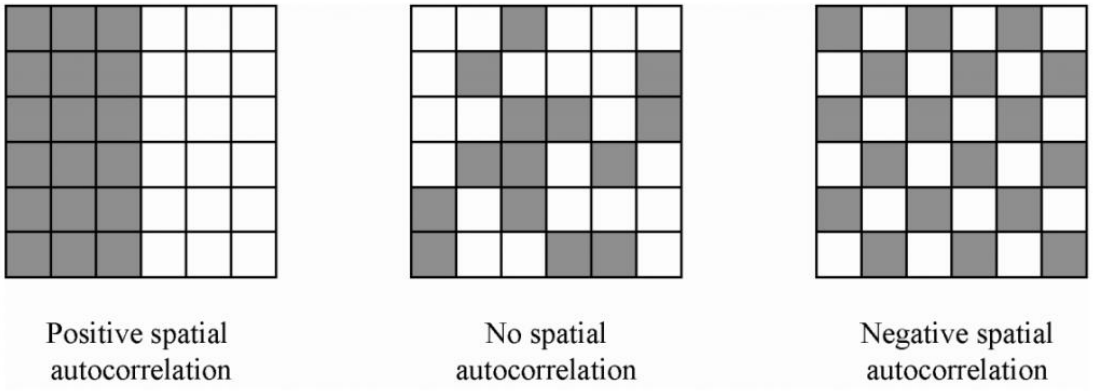
SPATIAL AUTOCORRELATION:

- Spatial autocorrelation is a measure of similarity between nearby observations.
- A positive spatial autocorrelation value indicates that a nearby observations have similar properties and vice versa for a negative value.
- A commonly used statistics that measure spatial autocorrelation are:

i) Moran's-I statistic.

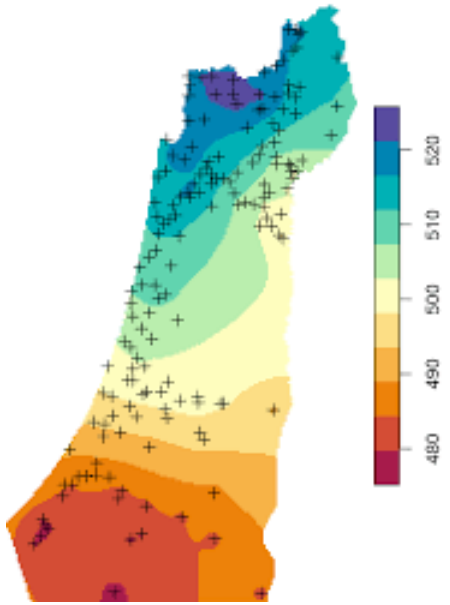
ii) Geary's-C statistic (for binary data).

iii) Semi-variogram.



SPATIAL INTERPOLATION:

- This technique is useful to predict values at locations where no measurements have been made.
- Based on spatial autocorrelation, values in nearby locations tend to be similar.
- Example: Spatial interpolation can estimate the temperatures at locations without recorded data by using known temperature readings at nearby weather stations.



- Among the methods of spatial interpolation:

i) Proximity polygons.

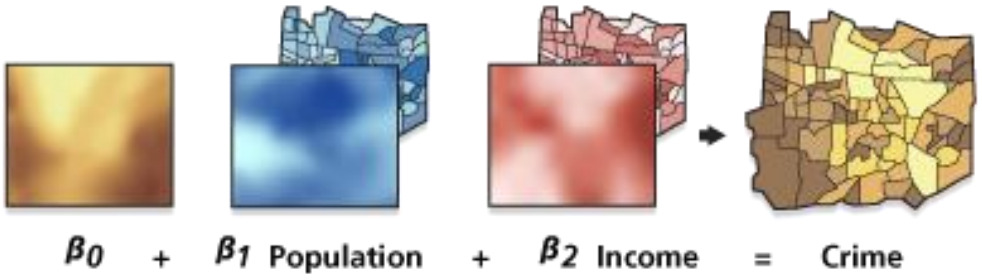
ii) Nearest neighbour interpolation.

iii) Inverse distance weighted.

iv) Kriging.

LOCAL SPATIAL REGRESSION:

- In spatial context, local refer to location.
- Instead of fitting single regression model to spatial data, it is better to fit several regression models with each one referring to a specific location.
- This technique known as geographically weighted regression, (GWR).
- GWR takes into account non-stationary variables (Example: demographic factors; characteristics of the physical environment) and models the local relationship between the predictor variable and the response variable.
- GWR is a technique that is used to get an insight about the changes in importance of different variables over space.



REFERENCES:

- Bivand, R.S., Pebesma, E., Gómez-Rubio, V. (2013). *Applied Spatial Data Analysis with R*. New York: Springer.
- Fischer, M. M., Wang, J. (2011). *Spatial Data Analysis: Models, Methods And Techniques*. Berlin: Springer.
- Lovelace R, Nowosad J, Muenchow J. (2019). *Geocomputation with R*. Chapman and Hall/CRC
- Onaya, T.J. (2021). *Spatial Analysis with R: Statistics, Visualization, and Computational Methods. Second Edition*. Boca Raton: CRC Press.
- Schabenberger, O., Gotway, C. A. (2004). *Statistical Methods For Spatial Data Analysis*. Chapman and Hall/CRC.
- *Spatial Data Science with R*. Available at: https://rspatial.org/raster/index.html#

NEXT TOPIC:

Mining Graph Data