Geographic Information System (GIS)

IS 454

Lecture 5: Spatial Analysis

Professor Dr. Safa A. Najim
Computer IS department
College of CS and IT

Spatial Analysis

- What is spatial analysis?
- Vector GIS analysis capabilities
- Raster GIS analysis capabilities

A statement

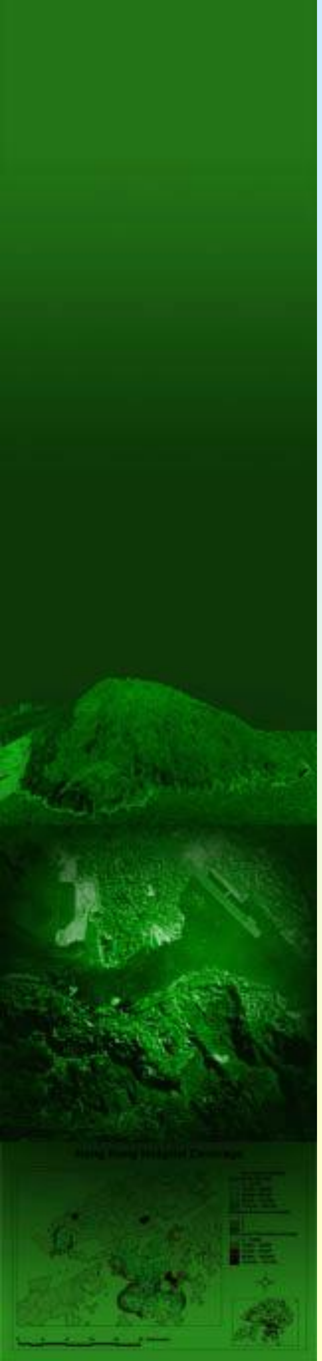
The distinguishes of a GIS from other types of information systems are its spatial analysis functions. These functions use the spatial and non-spatial attribute data in the GIS database to answer questions about the real world.

What is “analysis”?

- GIS offers capabilities for transforming the original spatial data to answer user’s questions.
- Such transformations are often referred to as “data analysis” capabilities of GIS.

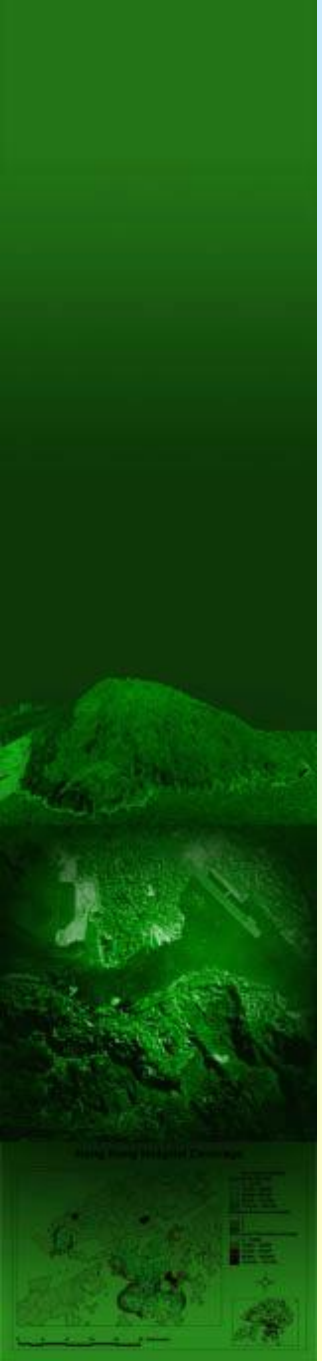
The definition of “analysis”

- “Analysis” is the process to resolve and separate the reference system into its parts to illuminate their nature and interrelationships, and to determine general principles of behavior.
- e.g. calculating the average income of a specific city population is an analysis
  - discovers the central tendency of income distributions of all residents in that city.



What is spatial analysis?

- Spatial analysis is a set of methods whose results change when the locations of the objects being analysed change.
  - Spatial dependency
- e.g. the calculation of an average income for a group of people is not spatial analysis
  - it does not depend on the locations of the people.
- e.g. the calculation of the centre of US population is spatial analysis
  - the results depend on knowing where all US residents are located.



Nature of spatial analysis

- Spatial analysis can reveal things that might otherwise be invisible – it can make what is implicit explicit.
- Spatial analysis adds value to geographical data and of turning data into useful information.
- Effective spatial analysis requires an intelligent user, not just a powerful computer.

Organising geographical data for analysis

- Data layers
  - A data layer consists of a set of logically related geographical features and their attributes
- Representations of a data layer
  - Raster — grid, overlay (grid cells)
  - Vector — coverage (point, line, polygon)

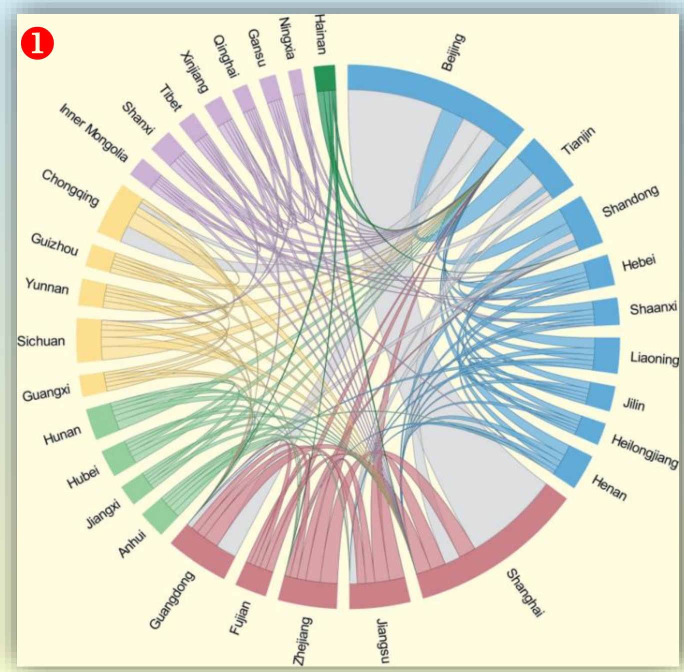
Vector analysis functions

- Analysis of attribute tables
- Topological overlay
- Buffering

Analysis of attribute tables

- Attributes of locational data are organized as tables.
- Normal database management functions can be applied
  - e.g. relational join
- Statistical analysis can be performed on those tabular data
  - The results can later be visualized in map through the links between tables and map features.
- The tabular analytical functions can be further applied to Big Data (e.g. volunteered geographical information)
  - e.g. textual documents, especially webpages

Toponym co-occurrences retrieved from webpages



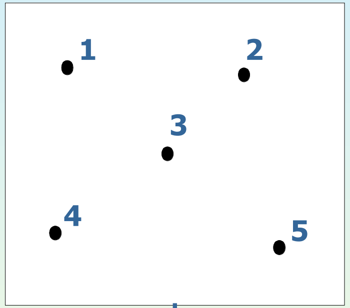
Relatedness of provinces of China, measured using toponym co-occurrences in web documents. ❶ Chord graph depicting the relatedness between provinces. ❷ Regionalization based on relatedness between provinces (Longley et al. 2016, pp 218).

Topological overlay

- When two layers are combined, the result must have more information as well.
- New intersection must be calculated and created wherever two lines cross and a line across an area object will create two new area objects.
- When topological overlay occurs, spatial relationships between objects area updated for the new, combined map.

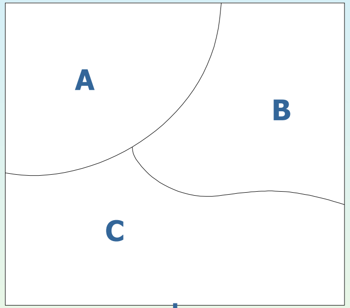
Point in polygon

Fast food restaurant

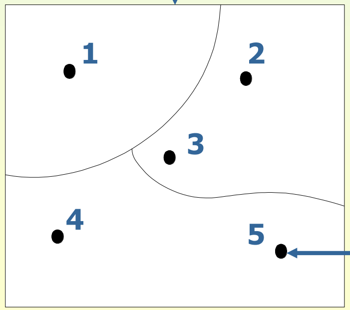


ID	*Restaurant*
1	McDonald
2	Pizza Hut
3	KFC
4	McDonald
5	Berger King

Towns



ID	*Town*
A	Shi Qi
B	Gang Kou
C	San Jiao

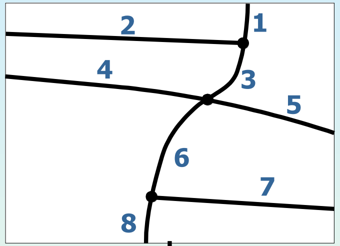


ID	*Town*	*Restaurant*
1	Shi Qi	McDonald
2	Gang Kou	Pizza Hut
3	Gang Kou	KFC
4	San Jiao	McDonald
5	San Jiao	Berger King

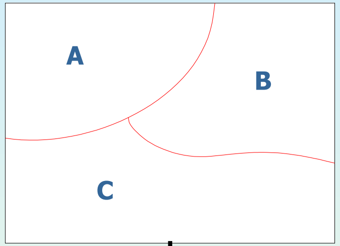
Line on polygon

Roads

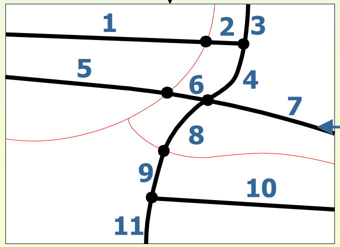
ID	Road No.
1	35
2	22
3	35
4	60
5	60
6	35
7	82
8	35



Geology

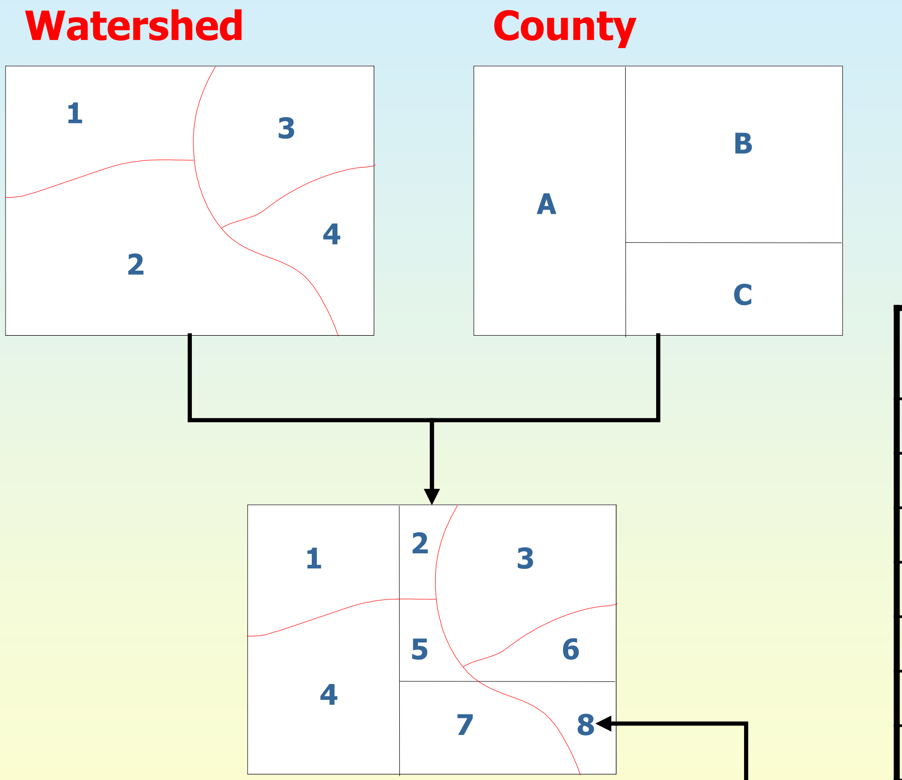


ID	Geology
A	Granite
B	Sandstone
C	Sand



ID	Original	Road No.	Geology
1	2	22	Granite
2	2	22	Sandstone
3	1	35	Sandstone
4	3	35	Sandstone
5	4	60	Granite
6	4	60	Sandstone
7	5	60	Sandstone
8	6	35	Sandstone
9	6	35	Sand
10	7	82	Sand
11	8	35	Sand

Polygon on polygon

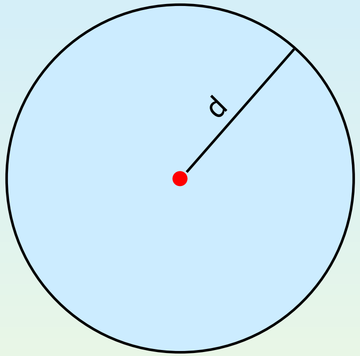


ID	Water shed	County
1	1	A
2	1	B
3	3	B
4	2	A
5	2	B
6	4	B
7	2	C
8	4	C

Buffering

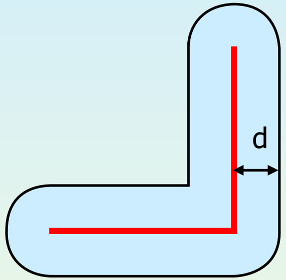
- A buffer can be constructed around a point, line or area.
- Buffering algorithm creates a new area enclosing the buffered object.
- The application of this buffering algorithms fundamentally addresses the creation of zones around the target.
  - e.g. protected zone around lakes, or streams
  - zone of noise pollution around highways or airports
  - service zone around bus route
  - groundwater pollution zone around waste site

Buffering on point, line and area



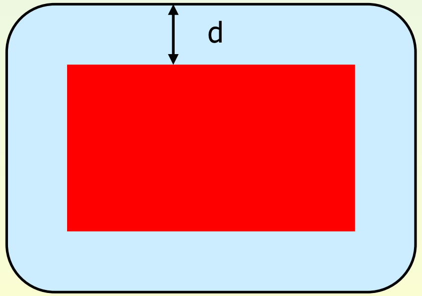
Buffering a point

e.g. area within 1km to a hospital.



Buffering a line

e.g. area within 100m to a road.



Buffering an area

e.g. area within 100m to a building.

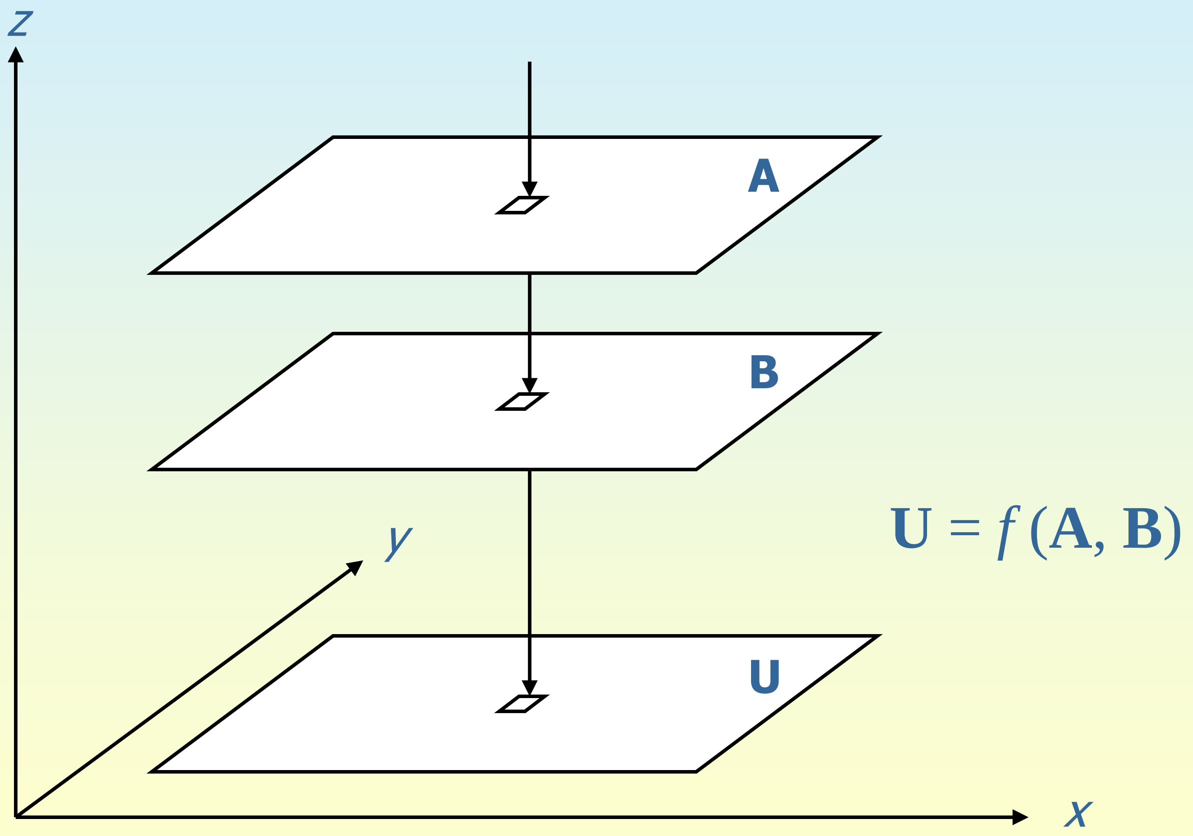
Raster analysis functions

- Local functions (point functions)
- Neighbourhood functions
- Global functions

Local functions

- Local functions examine rasters cell by cell, examining the value in a cell in one data layer and perhaps comparing it with the values in the same cell in other layers. The operations are independent of the effects of attribute values from neighbouring cells.
- A local function results in a new grid as a function of one or more input grids.

Spatial context of local functions



Merge

1	1	1	1
1	1	4	2
1	4	3	3
4	3	3	3

= merge (

1	1	1	N
1	1	N	N
1	N	3	3
N	N	3	3

,

5	5	4	1
3	5	4	2
4	4	2	2
4	3	3	1

)

If **A** == NODATA then
U = **B**
else
U = **A**

e.g. **U** = merge(**A**, **B**)

Maximising

5	5	4	2
3	5	4	2
4	4	3	3
4	3	3	3

= max (

1	1	1	2
1	1	2	2
1	2	3	3
2	2	3	3

,

5	5	4	1
3	5	4	2
4	4	2	2
4	3	3	1

)

U = max (**A**, **B**, …)

e.g. **U** = max (**A**, **B**)

Minimising

1	1	1	1
1	1	2	2
1	2	2	2
2	2	3	1

= min (

1	1	1	2
1	1	2	2
1	2	3	3
2	2	3	3

,

5	5	4	1
3	5	4	2
4	4	2	2
4	3	3	1

)

U = min (**A**, **B**, ...)

e.g. **U** = min (**A**, **B**)

Reclassification – arbitrary

1	1	1	2
1	1	2	2
1	2	3	3
2	2	3	3

Lookup table

A	**U**
1	5
2	4
3	2

5	5	5	4
5	5	4	4
5	4	2	2
4	4	2	2

Given a_1, a_2, …, a_n
and b_1, b_2, …, b_n

if **A** == a_1 then
    U = b_1
else if **A** == a_2 then
    U = b_2
…
else
    U = **A**

e.g. If (**A** == 1)
    U = 5
  else if (**A** == 2)
    U = 4
  else if (**A** == 3)
    U = 2
  endif

or **U** = con(**A** == 1, 5, con(**A** == 2, 4, 2))

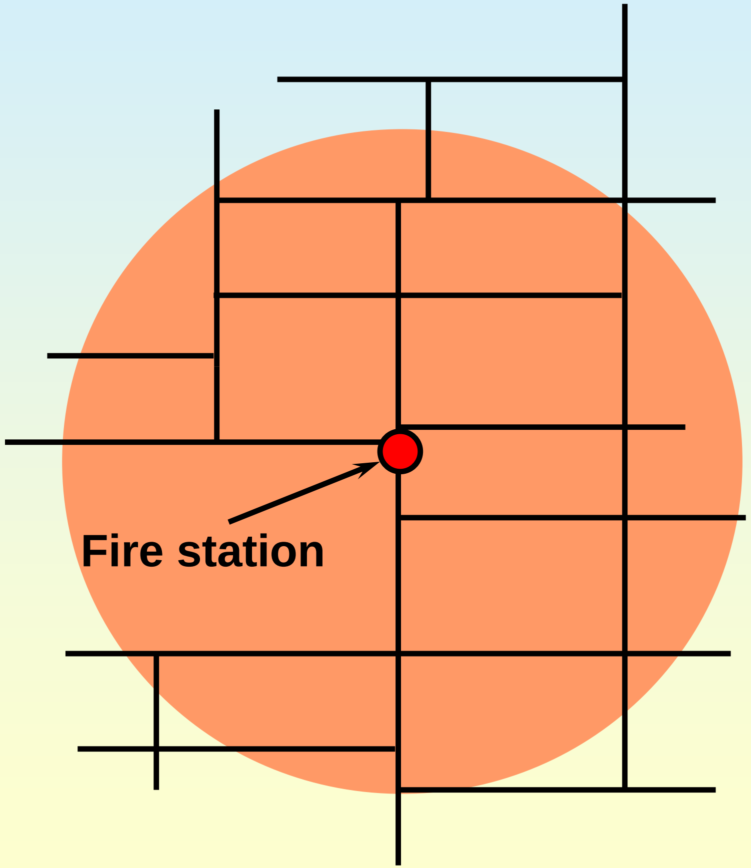
The Neighbourhood and global functions

- The neighbourhood functions compare the value in each cell with the values in its neighbouring cells.
- Global functions produce results that are true of the entire layer, e.g. its mean value.
- These are functions that explicitly make use of some kind of spatial associations in order to determine the value for the locations on the new output grid.

Neighbourhood function parameters

- Every neighbourhood function requires at least three basic parameters:
  - Target location(s) (neighbourhood focus)
  - A specification of the neighbourhood around each target
  - A function to be performed on the elements within the neighbourhood

Problem addressed by neighbourhood functions



Question: What is the number of residential buildings within 5km to the given fire station?

Target: fire station

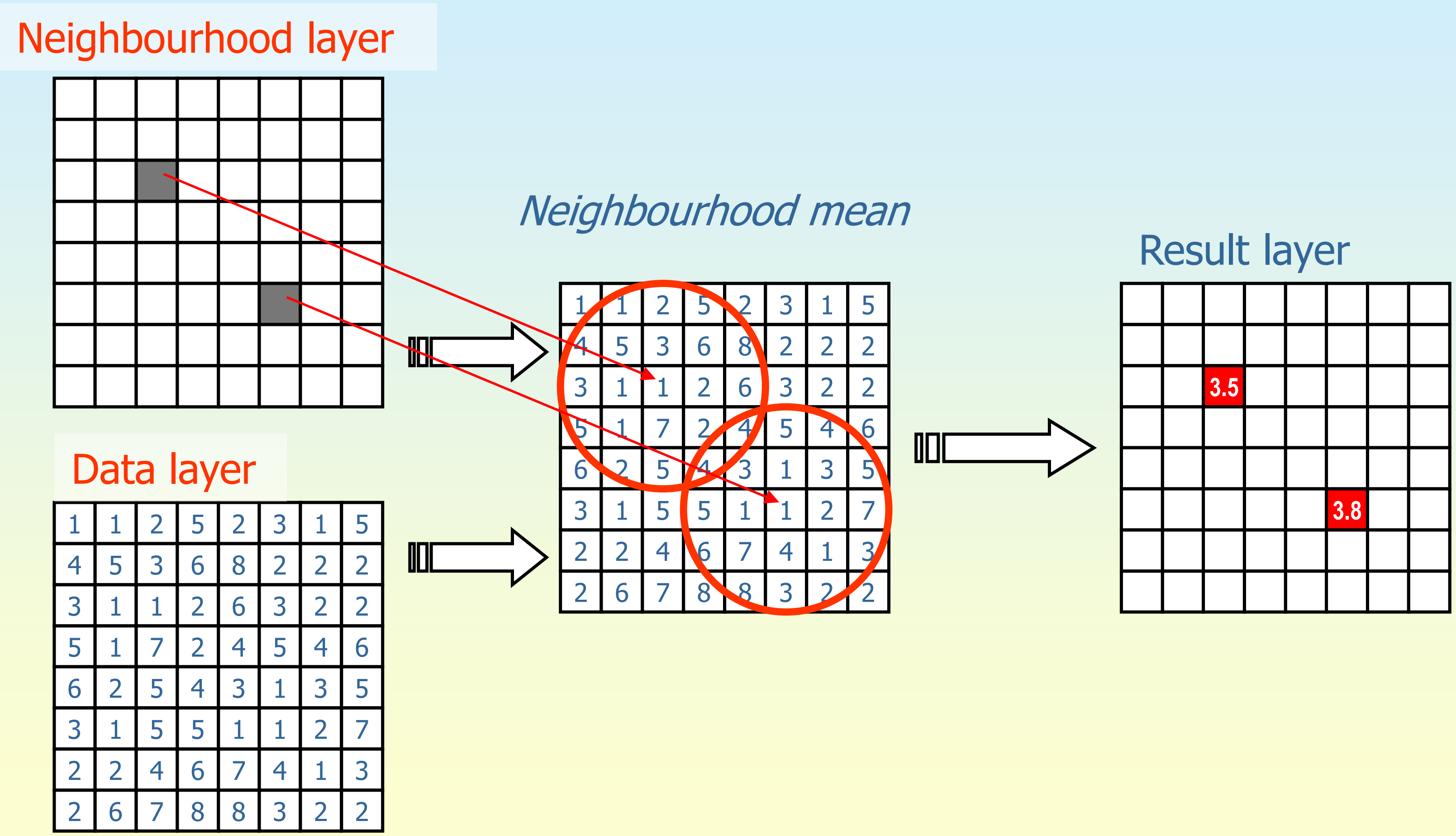
Neighbourhood: the area within 5km radius

Function: count the number of residential buildings

Spatial search

- Compute an attribute value for each target cell as a function of attribute values of its neighbourhood in an existing grid.
- Target: target cell(s) on focal grid
- Functions: sum, mean, standard deviation, etc.
- Neighbourhood: circular, square or “ring-shape”

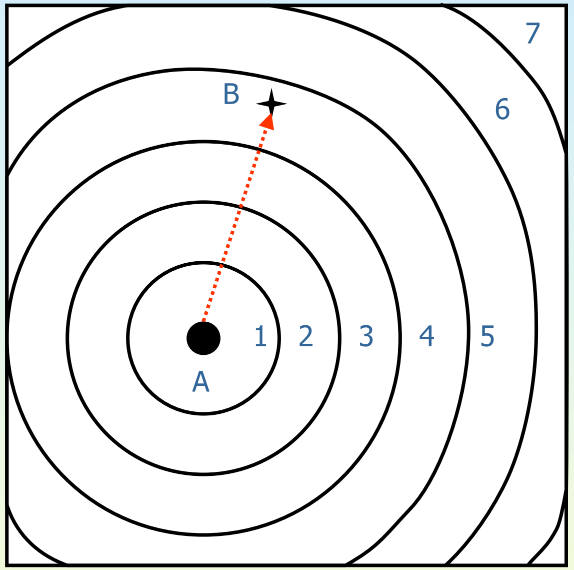
Neighbourhood statistics



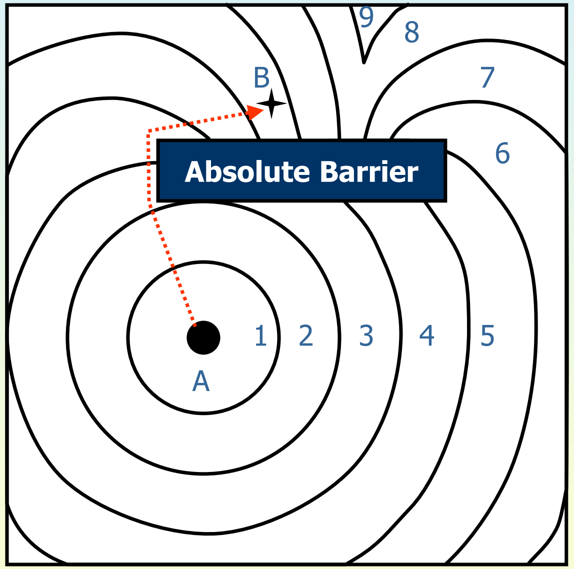
Proximity

- Compute an attribute value for each cell according to the length of the shortest path between that cell and the target location or area.
- The distance can be measured in Euclidean distance or “cost distance” (or “weighted distance”).
- The least-cost path is the route between two targets where the cost distance is the minimum.
- In many cases, the cost distance is different from the Euclidean distance.

Proximity analysis variables

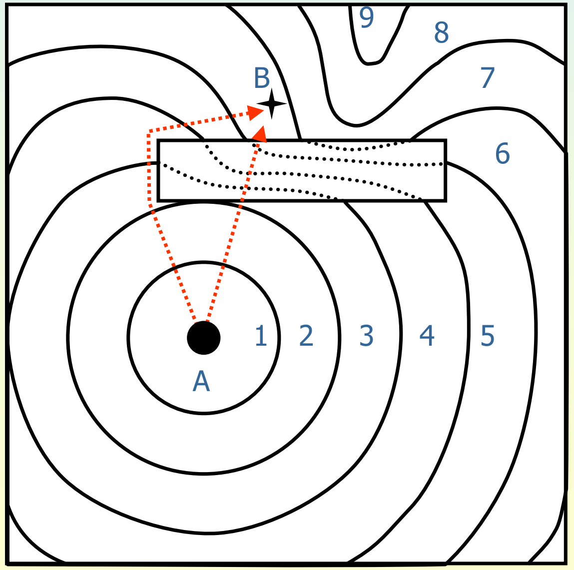


Travel zones defined by Euclidean distance



The effect of an absolute barrier on travel zones

The effect of a partial barrier (friction) on travel zones



Summary

- One most significant advantage for GIS is the capability for geographical analysis.
- GIS analytical capabilities are closely related to its data model.
- Vector data analysis functions include, e.g., geographical query, manipulation, topological overlay, buffering.
- Raster data analysis functions include, e.g., local, neighbourhood and global functions.