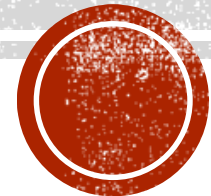


PERLOMBONGAN DATA RERUANG

STQD6414 PERLOMBONGAN DATA



Prof. Madya Dr. Nurulkamal Masseran

Jabatan Sains Matematik

Universiti Kebangsaan Malaysia

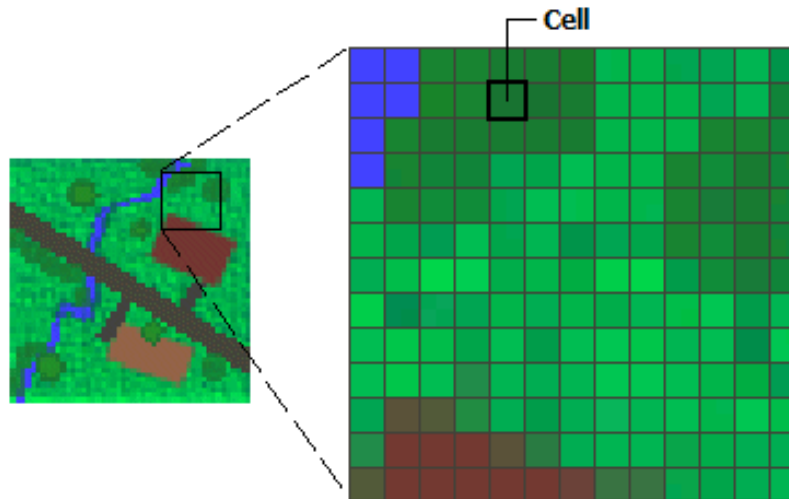
PENGENALAN:

- Data reruang ialah data yang melibatkan ruang, lokasi dan geografi.
- Data reruang merangkumi objek reruang (*spatial objects*) atau medan reruang (*spatial fields*).
 - i) **Objek reruang:** objek dengan sempadan (*borders*). **Contoh:** sungai, jalan, sempadan negara, bandar, dan lain-lain.
 - ii) **Medan reruang:** objek tanpa sempadan. **Contoh:** altitud, suhu, kualiti udara, dan lain-lain.
- Objek reruang umumnya diwakili dalam bentuk data vektor.
- Data vektor mengandungi maklumat “geometri” atau “bentuk” berkaitan lokasi, bersama dengan maklumat p/ubah lain yang dikaji.
- **Contoh:** set data vektor mungkin menerangkan maklumat tentang sempadan negara di dunia bersama-sama dengan maklumat tentang saiz populasi.



PENGENALAN:

- Manakala, medan reruang umumnya ditunjukkan dalam bentuk data raster.
- Raster ialah struktur data yang mewakili grid piksel (*pixel*) segi empat tepat atau juga dikenali sebagai sel.
- Data raster ditentukan berdasarkan panjang dan lebar imej dalam unit piksel atau bilangan bit setiap piksel.
- Data raster biasanya digunakan untuk mewakili fenomena ruang selanjar seperti ketinggian.

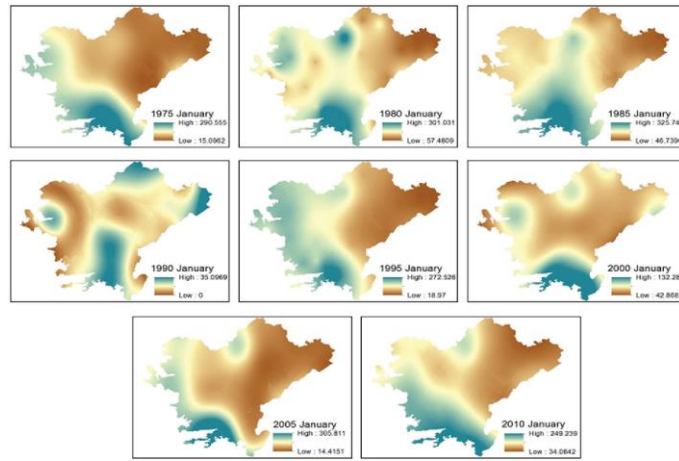


APLIKASI ANALISIS RERUANG:

- Terdapat pelbagai aplikasi analisis reruang yang melibatkan data meteorologi, sains bumi, analisis imej dan lain-lain.
- Jika data reruang dicerap bersama dengan maklumat masa, ianya dikenali sebagai analisis ruang-masa (*spatio-temporal analysis*).

1. Data Meteorologi:

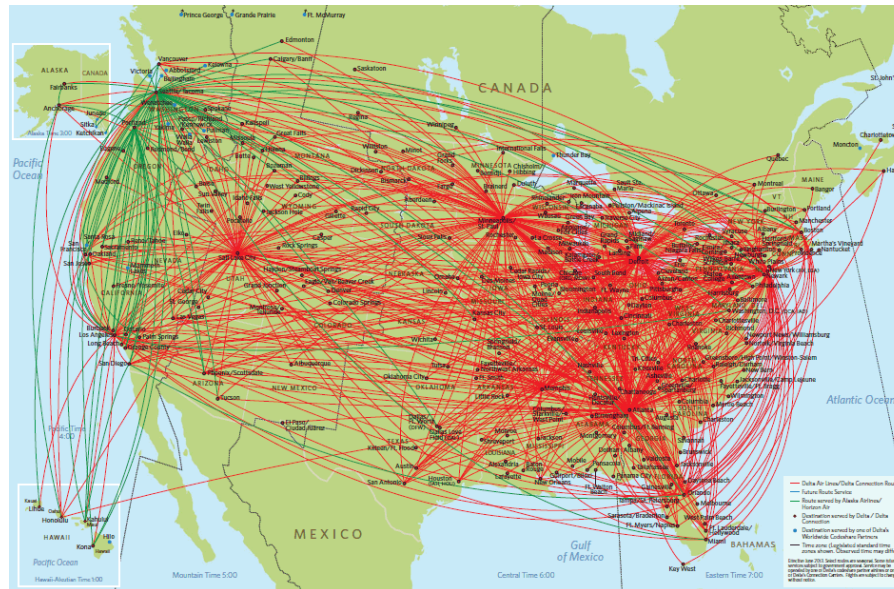
- Analisis berkaitan cuaca, seperti suhu, hujan dan tekanan.
- Umumnya, data jenis ini dicerap di lokasi geografi yang berbeza.
- Ia perlu dianalisis secara serentak untuk mengeluarkan maklumat berguna berkaitan cuaca di sesebuah negara.



APLIKASI ANALISIS RERUANG:

2. Data Objek Bergerak:

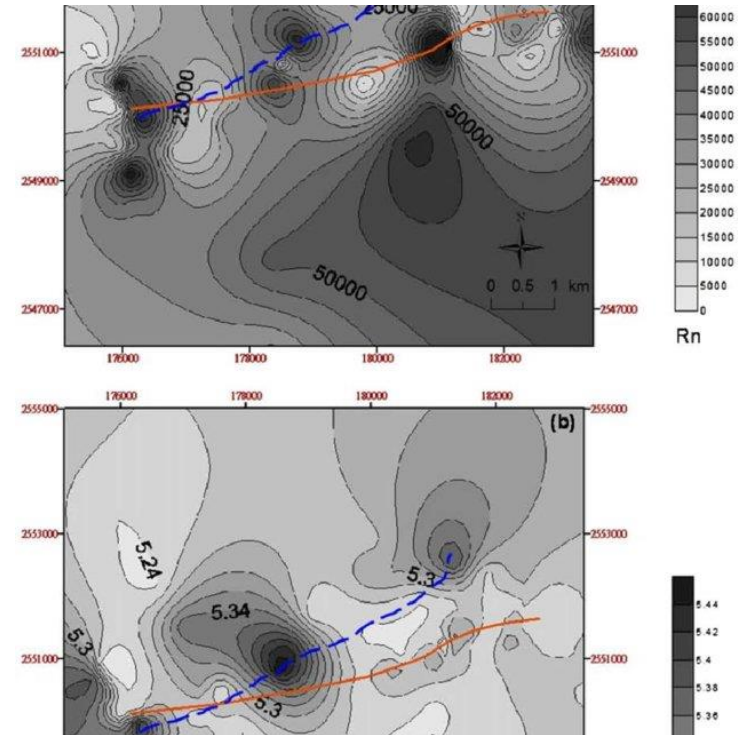
- Data objek bergerak menerangkan maklumat berkaitan laluan trajektori (*trajectories path*).
- **Contoh:** Trajektori penerbangan pesawat, Waze, GPS dan lain-lain.
- Data trajektori ini boleh dianalisis untuk mencungkil maklumat seperti ciri arah aliran atau mengenalpasti laluan anomali.



APLIKASI ANALISIS RERUANG:

3. Data Sains Bumi:

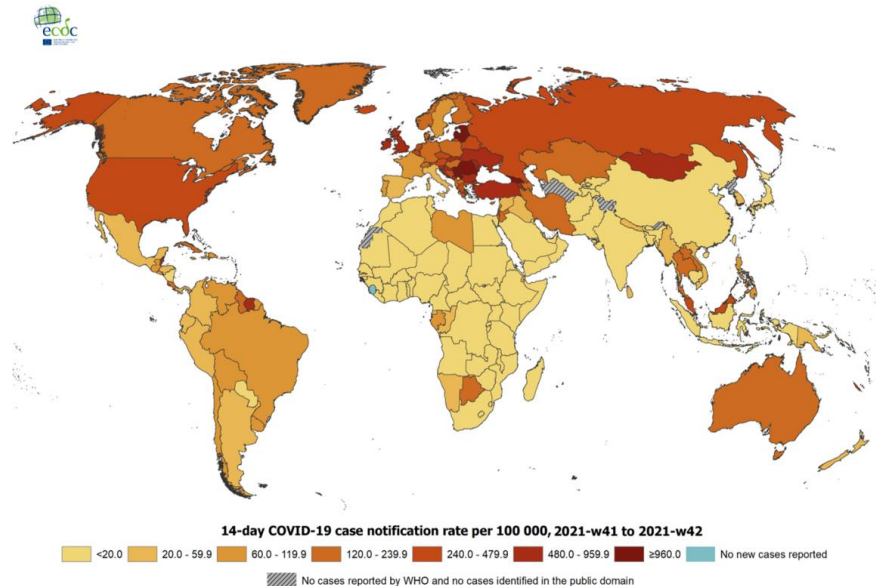
- Ciri geografi di lokasi yang berbeza umumnya menunjukkan sifat yang berbeza.
- **Contoh:** ketinggian tanah, kepadatan hutan, dan lain-lain.
- Pengesanan anomali dalam data Sains Bumi menunjukkan tentang trend anomali dalam aktiviti manusia.
- **Contoh:** penebangan hutan berlebihan, kewujudan tumbuhan anomali, dan lain-lain.



APLIKASI ANALISIS RERUANG:

4. Data Wabak Penyakit:

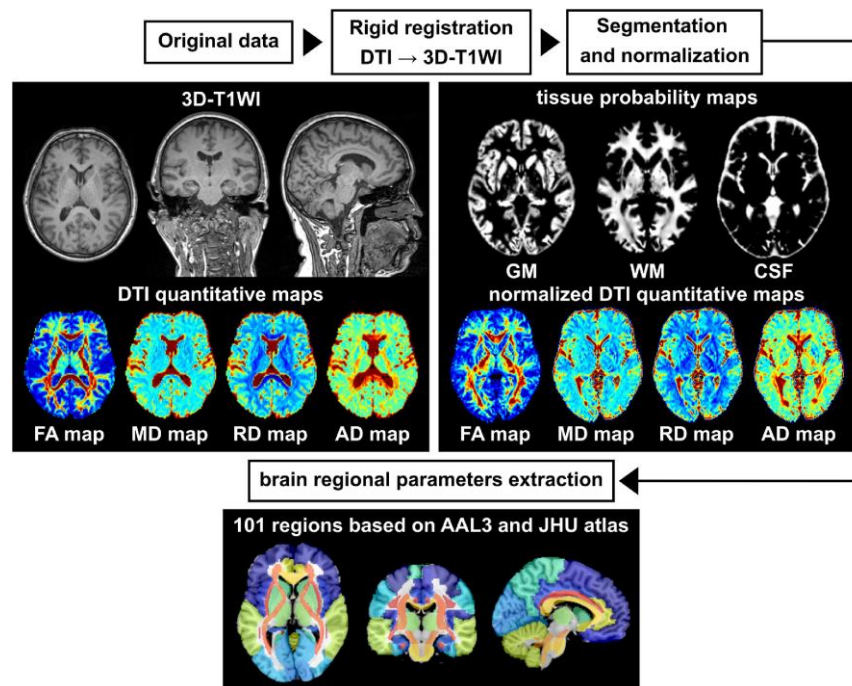
- Data mengenai wabak penyakit selalunya diagregatkan mengikut lokasi reruang seperti poskod, daerah dan negeri.
- Analisis data jenis ini memberikan maklumat tentang asal-usul wabak tersebut.
- Justeru, ia membantu pihak berkuasa dalam menguruskan risiko penyebaran wabak penyakit itu.



APLIKASI ANALISIS RERUANG:

5. Data Diagnostik Perubatan:

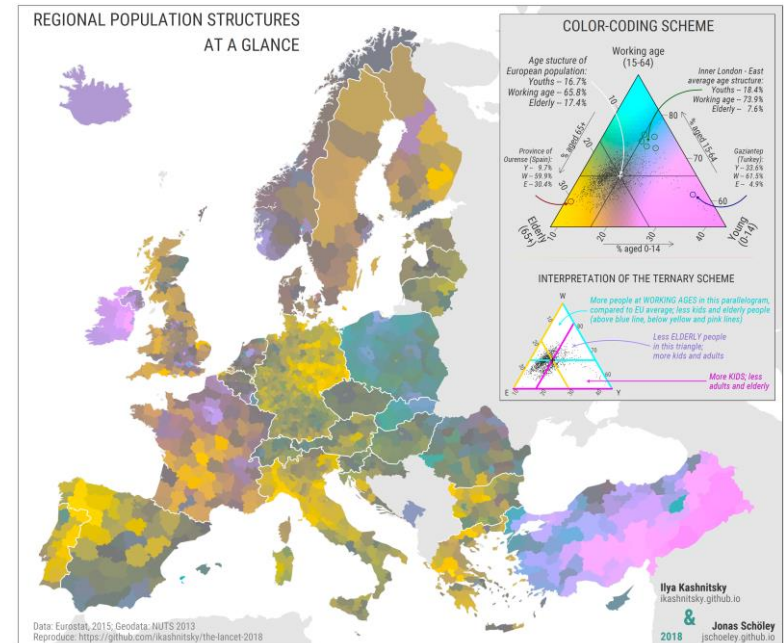
- Data pengimbas pengimejan resonans magnetik (MRI) ialah data reruang dalam 2 atau 3 dimensi
- Pengesanan anomali luar biasa dalam data ini boleh membantu dalam mengesan penyakit seperti tumor otak, permulaan penyakit Alzheimer, dan lain-lain.



APLIKASI ANALISIS RERUANG:

6. Data Demografi:

- Atribut demografi seperti umur, jantina, bangsa dan gaji sepadan dengan atribut reruang bagi setiap individu memberikan maklumat tentang corak taburan demografi.
- Maklumat ini penting untuk kerajaan membangunkan dasar yang munasabah.
- Untuk perniagaan, maklumat ini penting untuk pemasaran dan perancangan perniagaan.

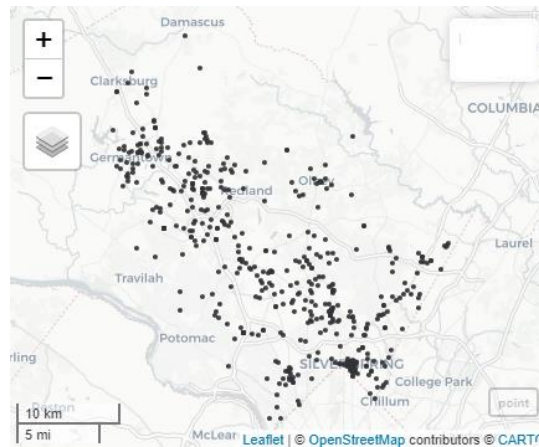


DATA RERUANG: BENTUK VEKTOR

- Terdapat tiga jenis data vektor yang utama dalam analisis reruang.

i) Data titik (*Point*):

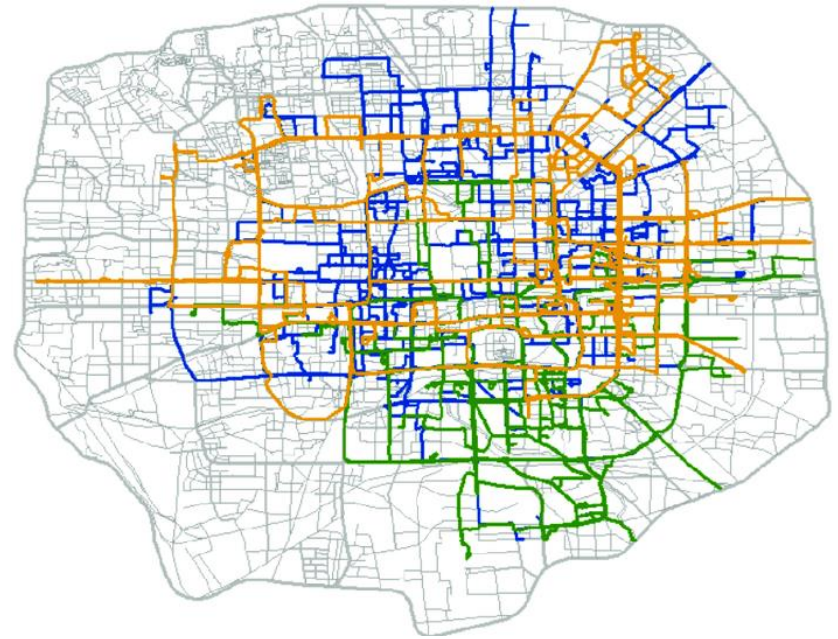
- Setiap titik mempunyai sepasang koordinat, dan pembolehubah yang berkaitan.
- **Contoh:** satu titik boleh mewakili lokasi di mana kes wabak dikesan. Ciri-ciri lain termasuk; tarikh ia dikesan, maklumat tentang pesakit, seperti; jantina, bangsa, status sosial dan sebagainya.
- Data ini juga boleh terdiri daripada gabungan beberapa titik dalam struktur berbilang titik, dengan rekod atribut tunggal.
- **Contoh:** semua kedai kopi di suatu bandar dianggap sebagai gabungan beberapa titik.



DATA RERUANG: BENTUK VEKTOR

ii) Data Garis (*Lines*):

- Data garis merujuk kepada satu atau lebih garisan poli atau segmen garisan yang disambungkan
- **Contoh:** sungai-sungai dan jalan-jalan adalah data garis.
- Garisan ditunjukkan sebagai set koordinat (nod) yang tersusun.
- Segmen garisan diperolehi dengan menyambungkan titik-titik.



DATA RERUANG: BENTUK VEKTOR

iii) Data Poligon (Polygon):

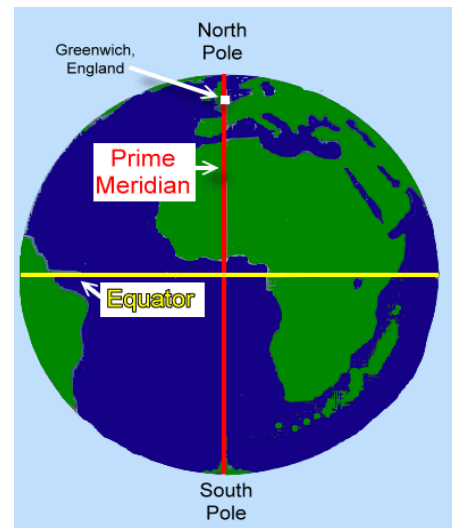
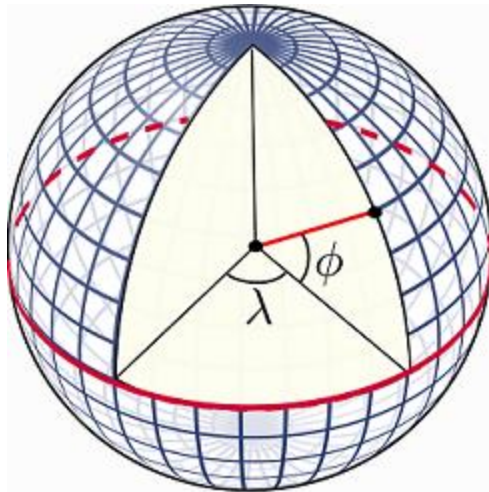
- Poligon merujuk kepada set garis-poli tertutup (*closed polylines*) .
- Iaitu, pasangan koordinat terakhir adalah sama dengan pasangan koordinat pertama mereka.
- Beberapa poligon boleh digabungkan dalam satu geometri.
- **Contoh:** Malaysia adalah terbentuk daripada gabungan beberapa negeri.



SISTEM KOORDINAT RUJUKAN (CRS):

i) Koordinat Angular (Sudut) dan Planar (Satah):

- CRS untuk data geografi dirujuk kepada sistem longitud dan latitud.
- CRS ini dikenali sebagai sistem Angular.
- Latitud ditentukan oleh sudut antara garis khatulistiwa dan garis seranjang yang melalui titik dari pusat Bumi.
- Longitud ialah sudut dari garis meridian rujukan ke garis meridian bagi lokasi.



SISTEM KOORDINAT RUJUKAN (CRS):

- Penentuan koordinat dibuat melalui pasangan sudut dan juga datum rujukan (model bagi bentuk muka bumi).
- Datum yang paling biasa digunakan ialah jenis WGS84 (World Geodesic System 1984).
- Unjuran sistem sudut 3-dimensi kepada sistem sudut 2-dimensi dirujuk sebagai sistem satah (planar).

ii) Tatatanda:

- Sistem planar CRS ditakrifkan daripada maklumat unjuran (*projection information*), model datum dan beberapa set parameter.
- Dalam R kita menggunakan tatatanda yang dikenali sebagai PROJ.4



DATA VEKTOR RERUANG DALAM R:

- Dalam perisian R, kelas `sp` menunjukkan objek data reruang.
- Antara pakej penting ialah `sp`, `raster`, `spatial`, `rgdal`, dan banyak lagi.
- Tiga bentuk asas data vektor reruang dalam R ialah:
 - i) `SpatialPoints`
 - ii) `SpatialLines`
 - iii) `SpatialPolygons`
- Pada data vektor asas tersebut, nilai atribut baharu boleh ditambah.
- **Contoh:** `SpatialPointsDataFrame` dan `SpatialPolygonsDataFrame`



DATA RASTER RERUANG DALAM R:

- Tiga bentuk asas data raster dalam R ialah RasterLayer, RasterBrick dan RasterStack.

i) RasterLayer:

- Objek RasterLayer mewakili data raster satu lapisan (*single-layer*).
- Objek RasterLayer menerangkan maklumat parameter asas data raster seperti bilangan lajur, baris dan system rujukan koordinat (CRS).
- Maklumat berkaitan atribut juga boleh disimpan dalam RasterLayer.

ii) RasterStack dan RasterBrick:

- RasterBrick dan RasterStack digunakan untuk membentuk data raster lapisan berganda.



MANIPULASI DATA VEKTOR RERUANG:

- Manipulasi data ialah proses menukar data kepada bentuk yang lebih teratur.
- Proses ini bergantung kepada objektif penganalisis data.
- Beberapa teknik manipulasi data vektor reruang:
 - i) Mempersembahkan data reruang sebagai format data.frame.
 - ii) Mengekstrak atribut tertentu.
 - iii) Tambah maklumat atribut baharu.
 - iv) Pilih subset data.
 - v) Integrasi Data.
 - vi) Manipulasi peta.
 - vii) Dan banyak lagi.



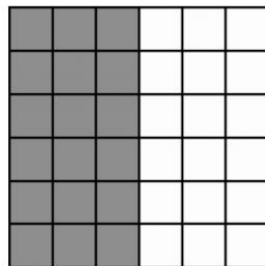
MANIPULASI DATA RASTER RERUANG:

- Beberapa teknik manipulasi data Raster:
 - i) Mengekstrak objek tunggal RasterLayer daripada objek RasterBrick atau RasterStack.
 - ii) Algebra dalam data raster.
 - iii) Tambah nilai baharu dalam sel.
 - iv) Tambah maklumat atribut baharu.
 - v) Pangkas dan gabungkan data raster.
 - vi) Fungsi deskriptif.
 - vii) Dan banyak lagi.

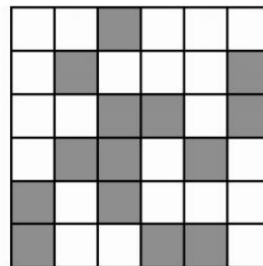


AUTOKORELASI RERUANG:

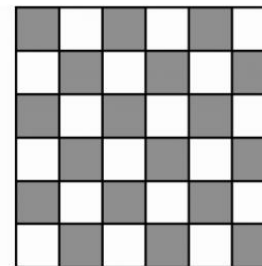
- Autokorelasi reruang ialah ukuran ciri kesamaan antara cerapan berdekatan.
- Nilai autokorelasi reruang positif menunjukkan cerapan yang berdekatan mempunyai ciri yang hampir sama dan sebaliknya untuk nilai negatif.
- Statistik yang biasa digunakan yang mengukur autokorelasi reruang ialah:
 - i) Statistik Moran's-I.
 - ii) Statistik Geary's-C (untuk data binari).
 - iii) Semivariogram.



Positive spatial
autocorrelation



No spatial
autocorrelation

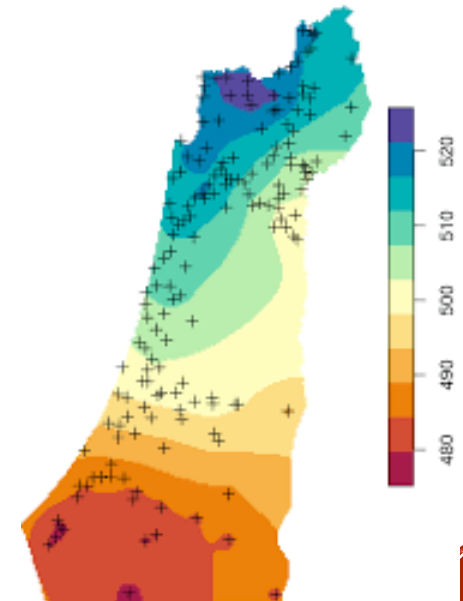


Negative spatial
autocorrelation



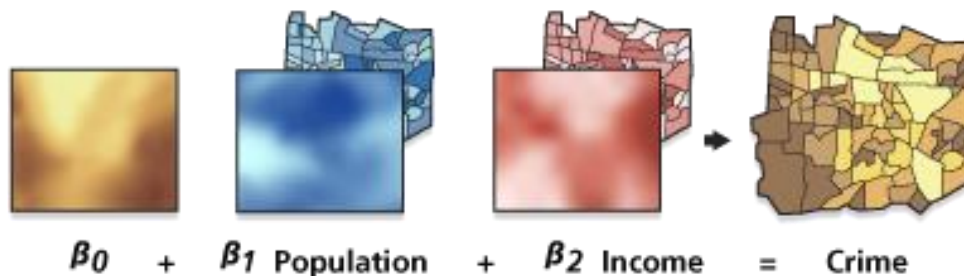
INTERPOLASI RERUANG:

- Teknik ini berguna untuk menganggar nilai data reruang di lokasi yang tiada cerapan dibuat.
- Berdasarkan autokorelasi reruang, nilai dari lokasi berdekatan cenderung untuk mempunyai nilai yang hampir sama.
- **Contoh:** Interpolasi reruang boleh menganggarkan suhu di lokasi tanpa data direkodkan dengan menggunakan maklumat suhu yang diketahui di stesen cuaca berdekatan.
- Antara kaedah interpolasi reruang:
 - i) Poligon Hampiran (*Proximity polygons*).
 - ii) Jiran terdekat (*Nearest neighbour*).
 - iii) Jarak songsang berpemberat (IDW)
 - iv) Kriging.



REGRESI RERUANG SETEMPAT:

- Dalam konteks reruang, setempat (*local*) merujuk kepada lokasi tertentu.
- Berbanding penyuaian model regresi tunggal, adalah lebih baik untuk menyuai beberapa model regresi dengan setiap satu model merujuk kepada lokasi tertentu.
- Teknik ini juga dikenali sebagai regresi geografi berwajaran (*geographically weighted regression, GWR*).
- GWR mengambil kira pembolehubah tidak pegun (**Contoh:** faktor demografi; ciri persekitaran fizikal) dan memodelkan hubungan setempat antara p/ubah peramal dengan p/ubah sambutan.
- GWR membolehkan kita memahami perubahan parameter pembolehubah yang berbeza secara reruang



RUJUKAN:

- 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#>



TOPIK SETERUSNYA:

Perlombongan Data Teks

