



# **Review of geospatial analysis**

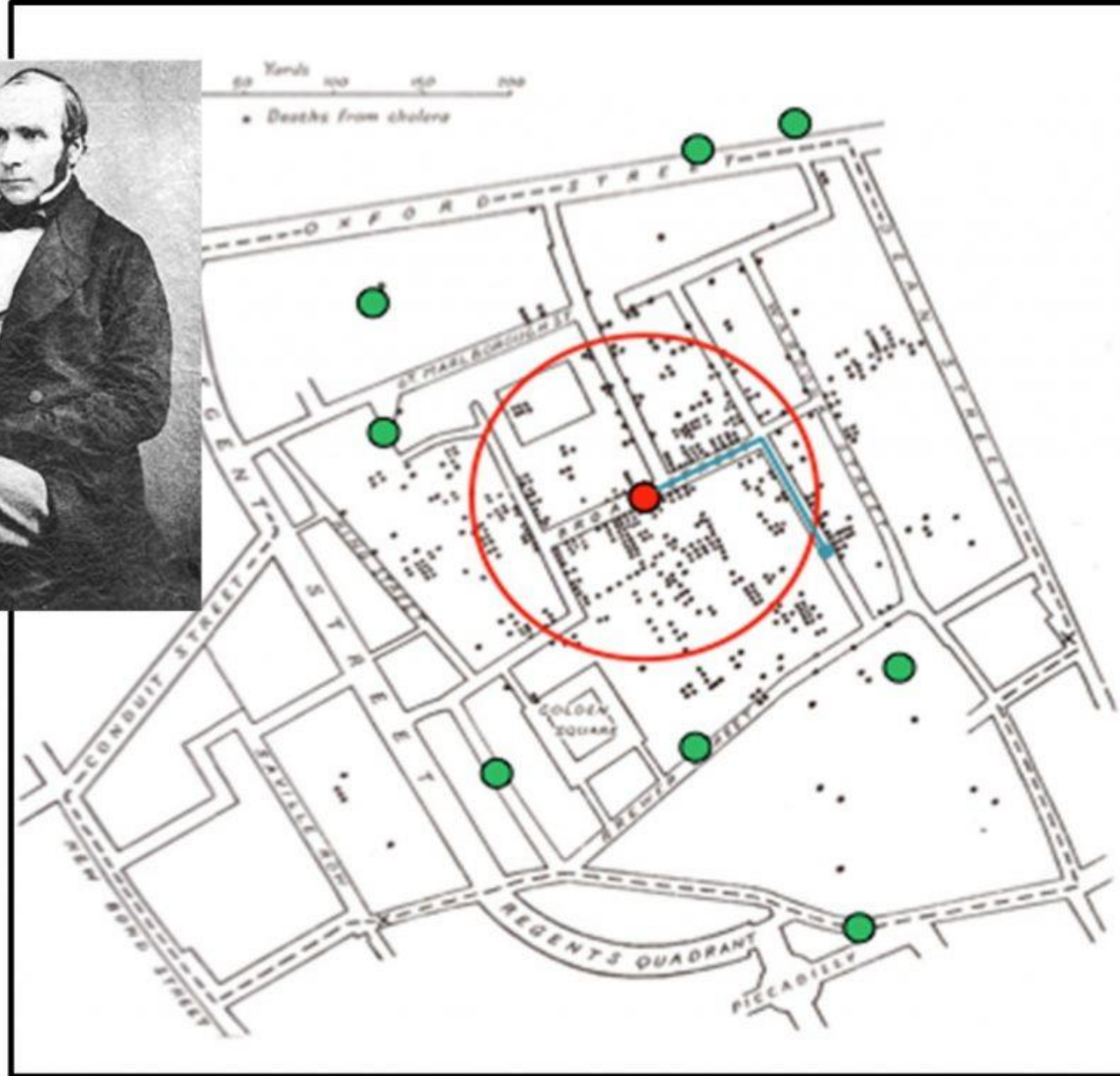
**Daniel Fuller**  
**Department of Community Health and Epidemiology**

# Time and Space

- Asking questions that relate to how things change over time and in space...
  - a) How has disease X change over time AND over space?
    - Similar thinking to an interaction analysis
  - b) Often see questions like, how has disease X change over time OR over space?

# Space?!

- Absolute Space
  - a) Where a person/thing is on the globe
  - b) Measured with a coordinate system
- Relative Space
  - a) How two things relate together in space
  - b) More on this later



# Coordinates

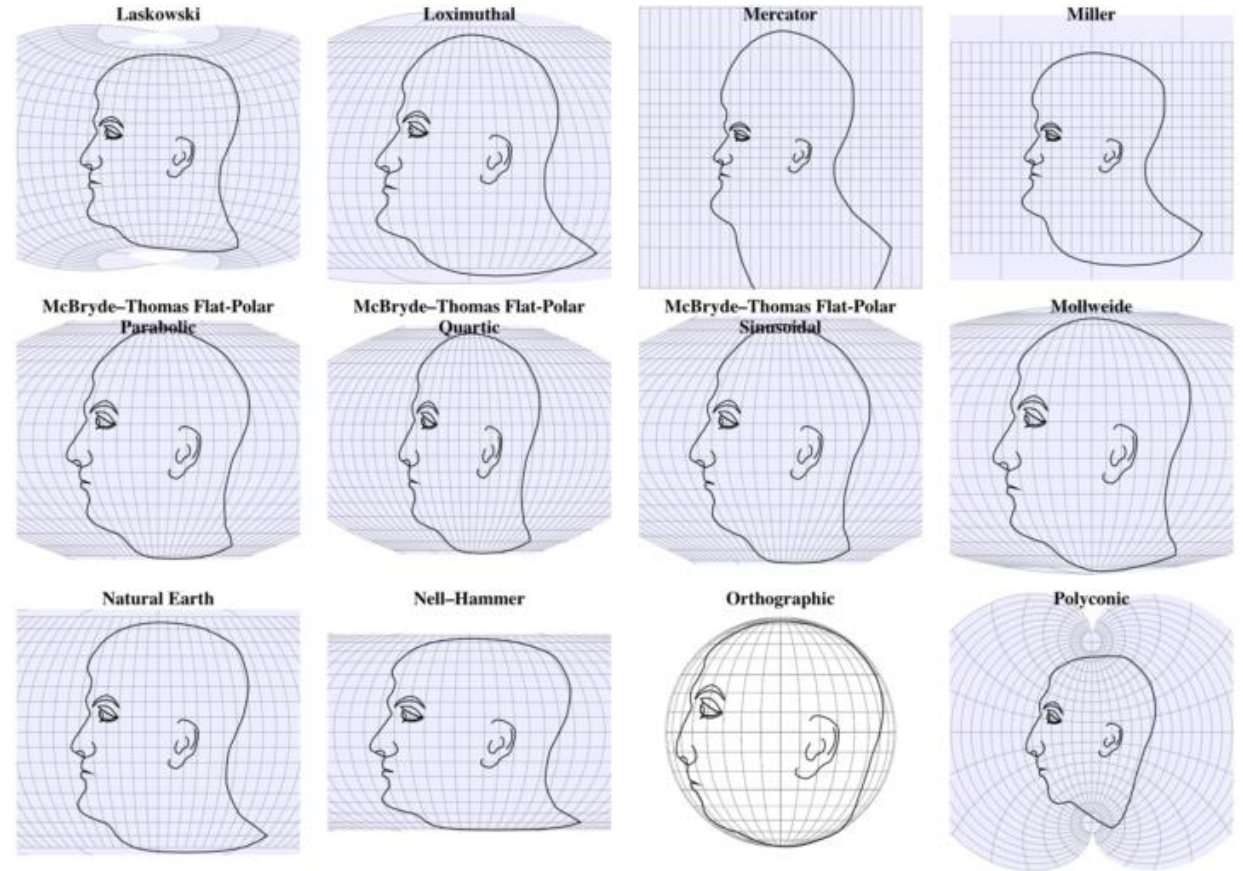
- A **spatial reference system (SRS)** or **coordinate reference system (CRS)** is a framework used to precisely measure locations on the surface of the Earth as coordinates. It is thus the application of the abstract mathematics of coordinate systems and analytic geometry to geographic space. (Wiki)
  - a) Geographic Coordinate System (the Earth is round, 3D)
  - b) Projected Coordinate System (the earth is modeled as flat, 2D)

# Geographic Coordinate Systems

- The **geographic coordinate system (GCS)** is a spherical or ellipsoidal coordinate system for measuring and communicating positions directly on the Earth as latitude and longitude. It is the simplest, oldest and most widely used of the thousands of spatial reference systems that are in use, and forms the basis for most others. Although latitude and longitude form a coordinate tuple like a cartesian coordinate system, the geographic coordinate system is not cartesian because the measurements are angles and are not on a planar surface.

# Projected Coordinate Systems

- Modelling a round surface on a flat map
- Mercator or Pseudo-Mercator are most common (All online maps)
- Statistics Canada uses Lambert Conformal Conic Projection



<http://bl.ocks.org/vlandham/raw/9216751/>

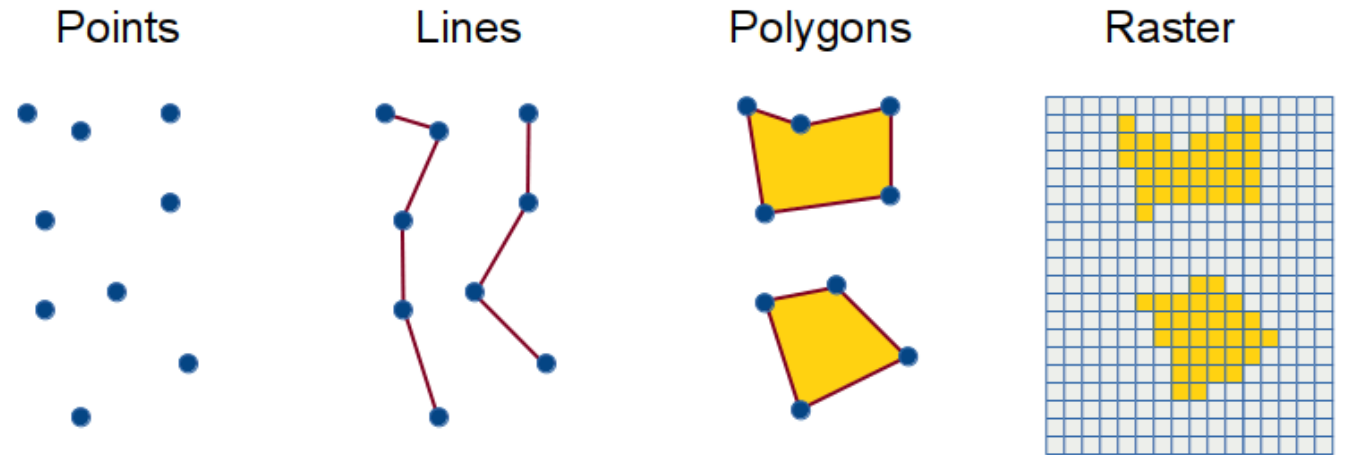


# Spatial Data

## ■ Types of spatial data

### a) Vector Data

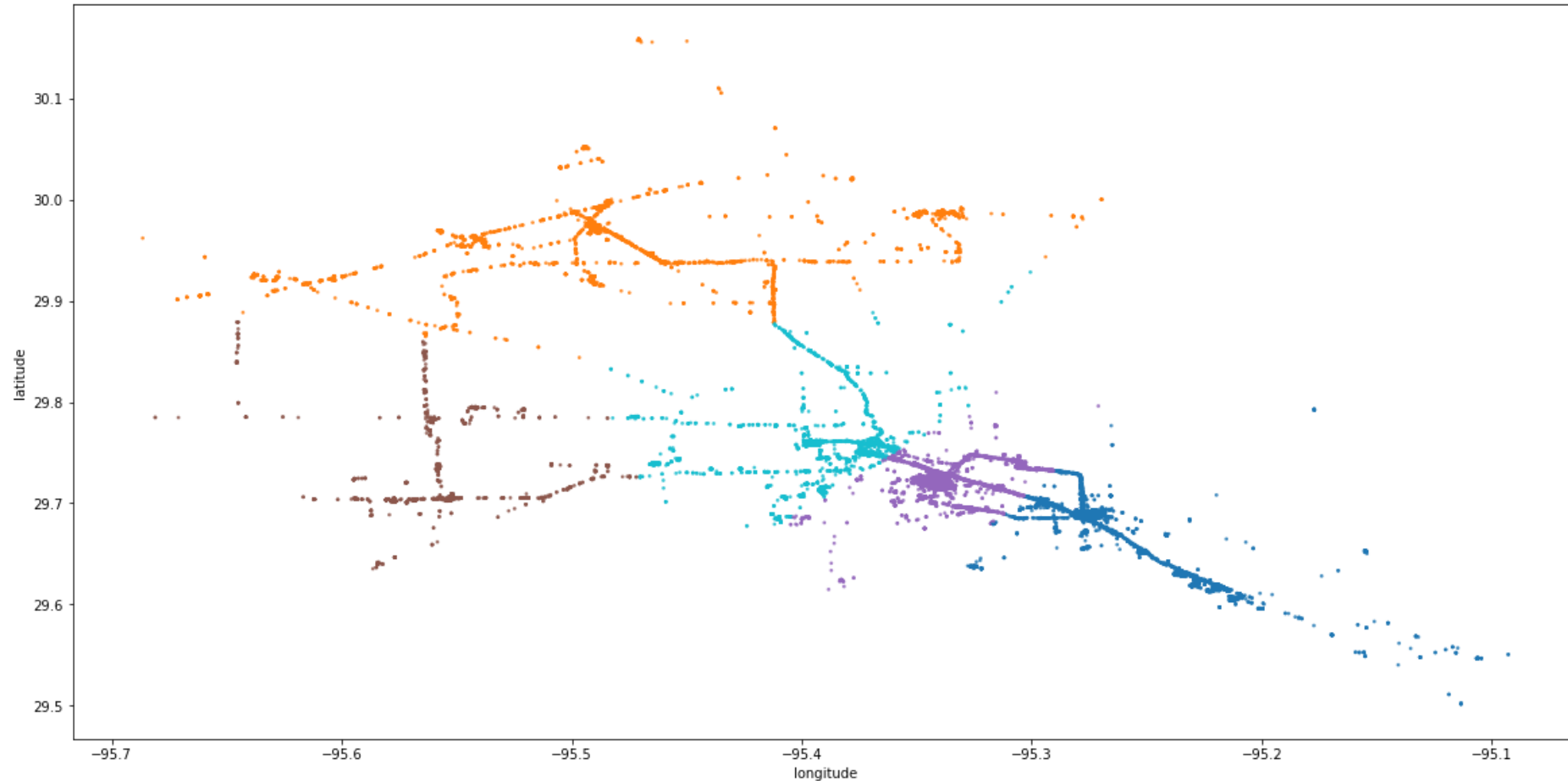
- Point data
  - GPS data, case locations
- Line data
  - Roads, rivers
- Polygon data
  - Census tracts



<https://michaelminn.net/tutorials/arcgis-pro-terrain/2019-points-lines-polygons.png>



# Point Data



[https://www.reddit.com/r/houston/comments/92g626/my\\_2year\\_worth\\_of\\_gps\\_data\\_points\\_from\\_google/](https://www.reddit.com/r/houston/comments/92g626/my_2year_worth_of_gps_data_points_from_google/)

# Line Data



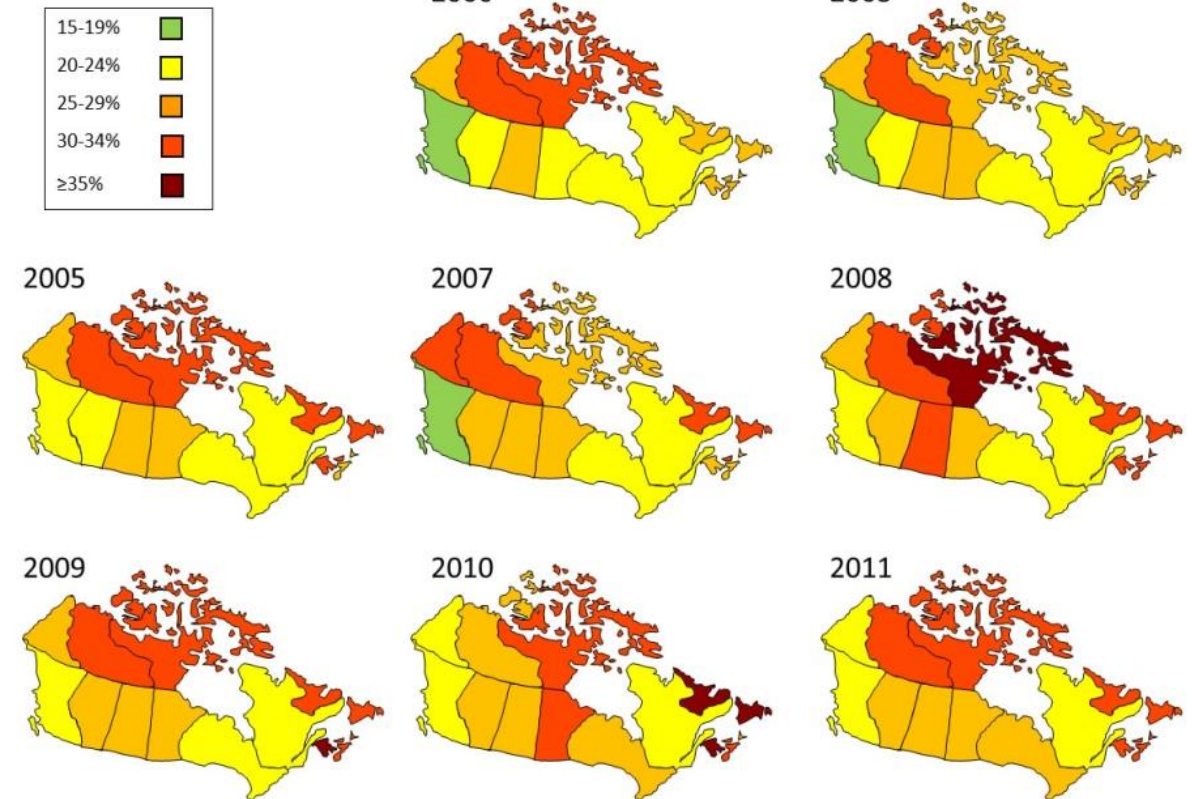
<https://aewinc.com/geographic-information-system-gis-project-centerline/>

# Polygon Data

## ■ Spatial Data Operations

- a) Counting
  - Points in polygons
- b) Simplification
- c) Centroids
- d) Buffers
- e) Clipping

Figure 1



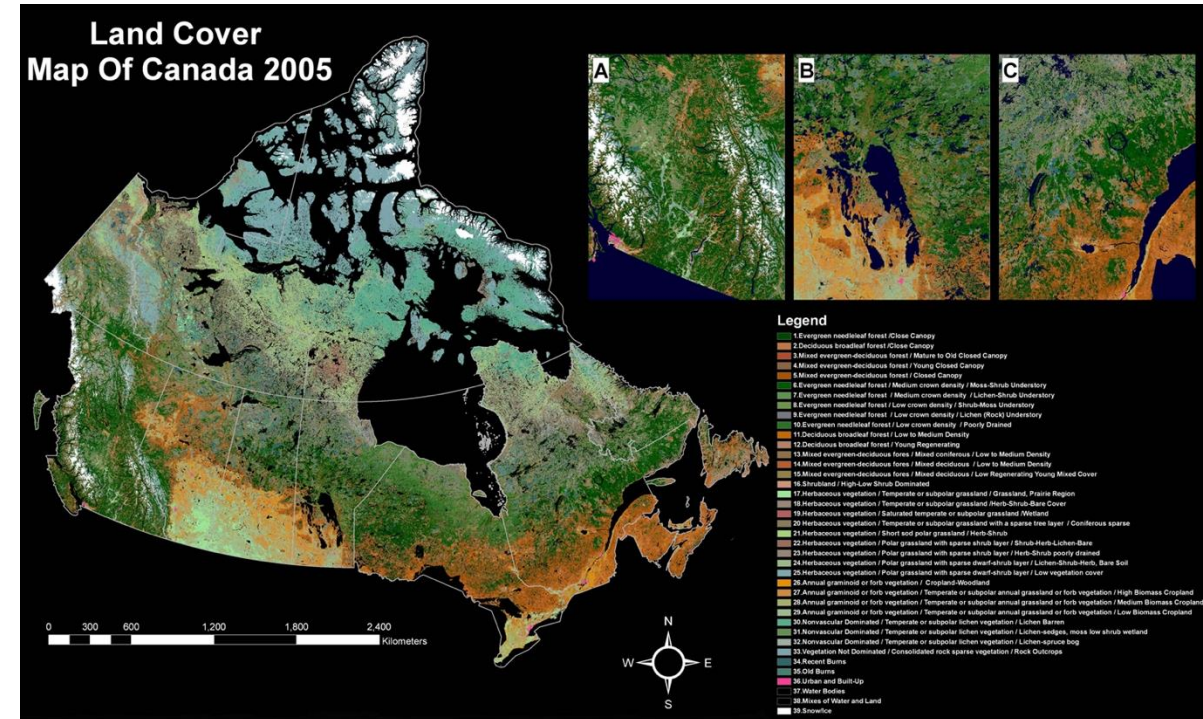
<https://www.ctvnews.ca/health/health-headlines/adult-obesity-rates-at-historic-high-ubc-study-says-1.1174630>

# Spatial Data

- Types of spatial data

- a) Raster data (Not really covering this)

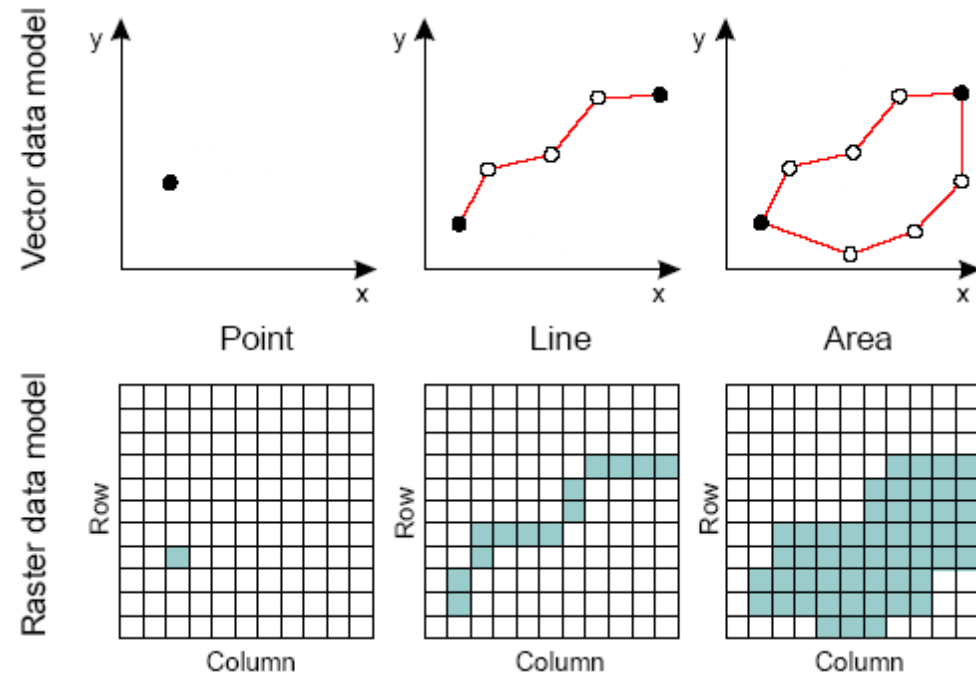
- Gridded surfaces
      - Green space (NDVI)
      - Forest cover



[https://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/files/earthsciences/images/optical/images/nlcc\\_fig1\\_e\\_.jpg](https://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/files/earthsciences/images/optical/images/nlcc_fig1_e_.jpg)

# Spatial Data

- Relationship between vector and raster data



[https://firenorth.org.au/nafi3/views/help/Maps\\_and\\_Fire\\_help4.htm](https://firenorth.org.au/nafi3/views/help/Maps_and_Fire_help4.htm)

# Spatial Data Operations

- Spatial Data Operations
  - a) Counting
    - Points in polygons
  - b) Simplification
  - c) Centroids
  - d) Buffers
  - e) Clipping



# Spatial Data Operations

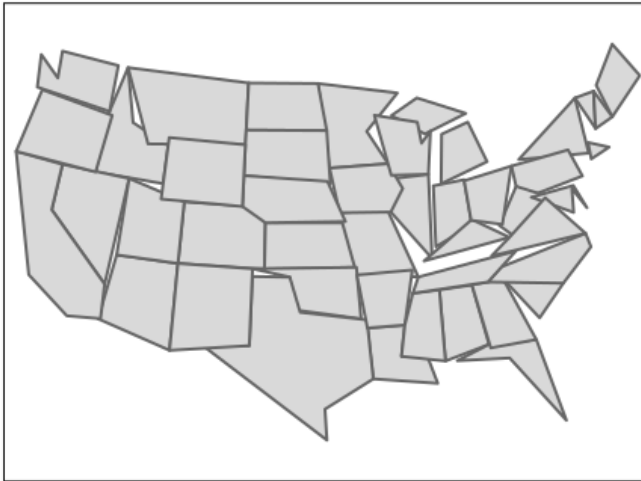
- Simplification

- a) Balance between data representation and speed
- b) Common problem with online maps.

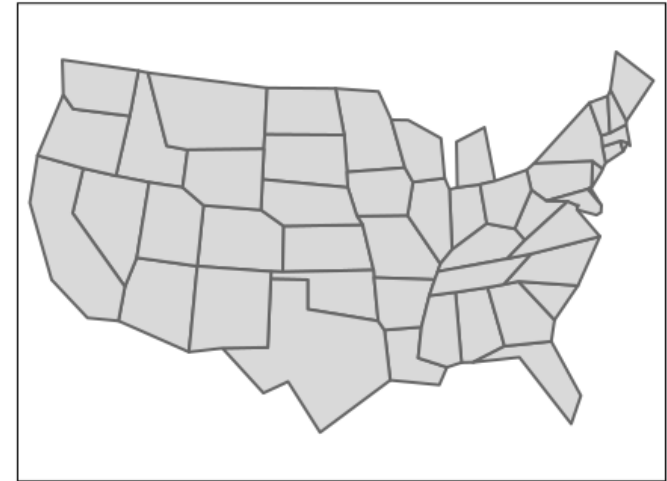
Original data



st\_simplify



ms\_simplify





# Spatial Data Operations

- Centroids

- a) Find the middle of something
- b) Weighted by population or other factors?
- c) Can be for different types of spatial data



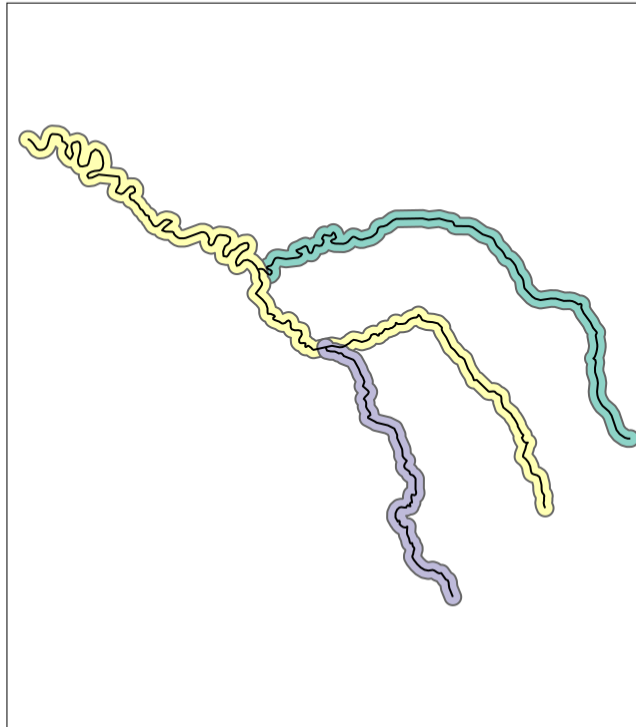
<https://geocompr.robinlovelace.net/spatial-operations.html>

# Spatial Data Operations

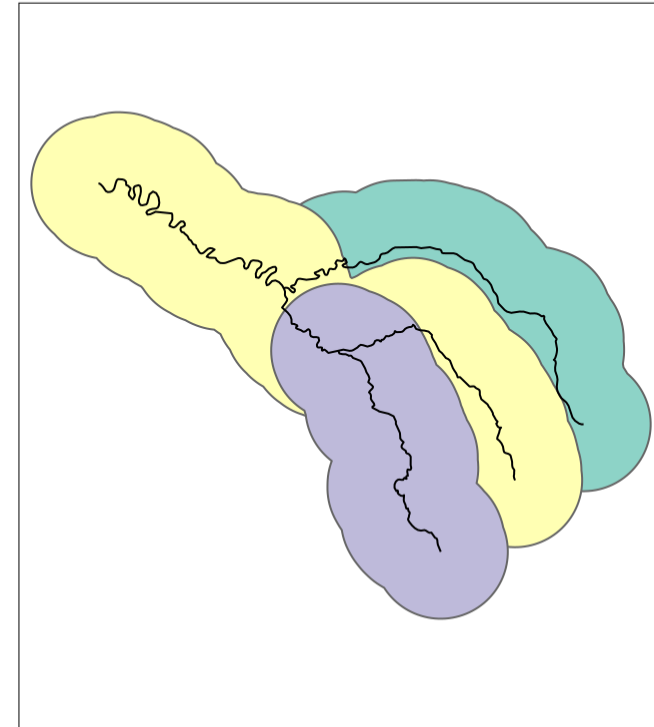
## ■ Buffers

- a) Extend the spatial extent
- b) Think about exposure?
- c) Types
  - Network based
  - Crow flies (Euclidean)
  - Different shapes (SD Ellipse)

5 km buffer



50 km buffer

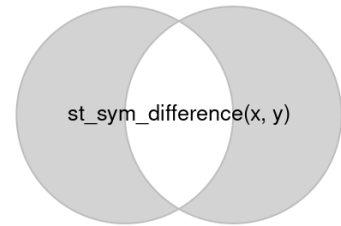
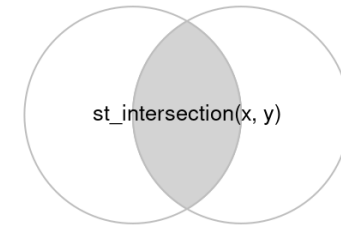
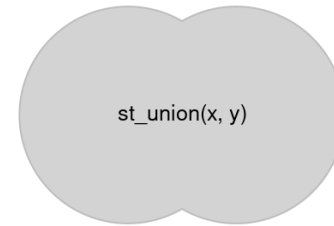
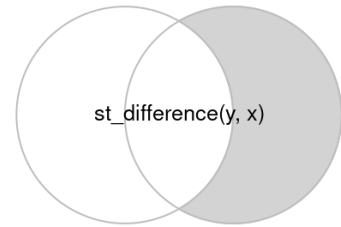
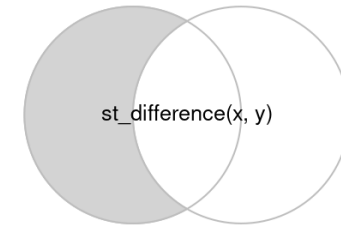
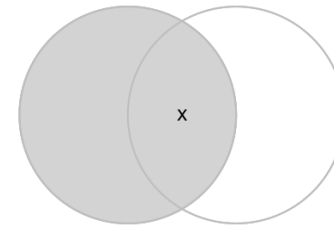


<https://geocompr.robinlovelace.net/spatial-operations.html>

# Spatial Data Operations

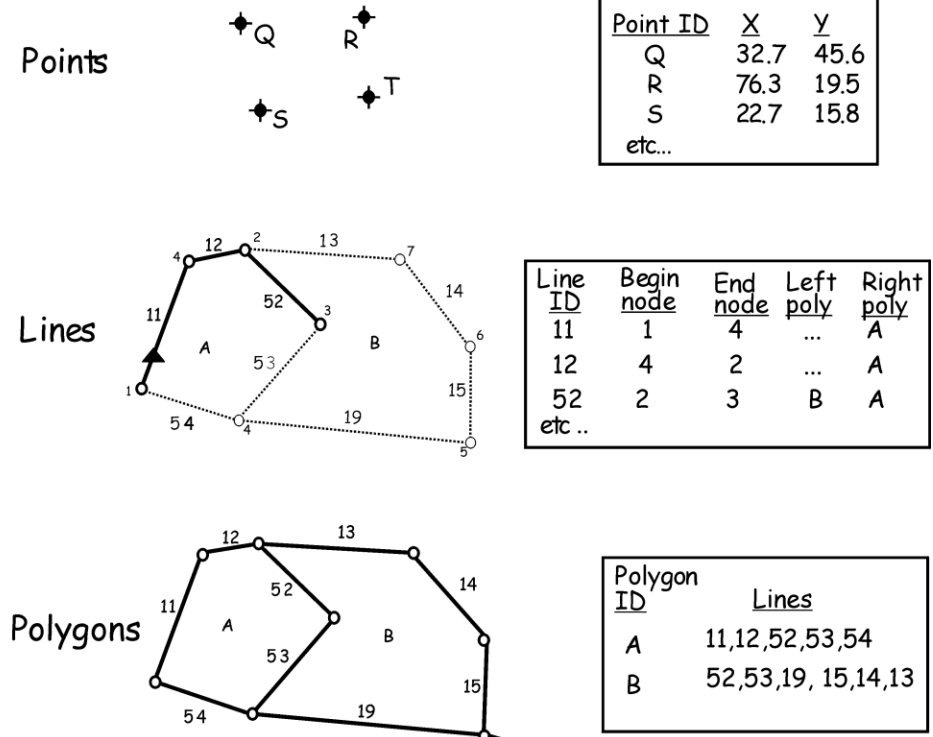
- Clipping

- a) Similar to topological relationships
- b) Way to subset the data
- c) Can count within different clippings



# Spatial Data Operations

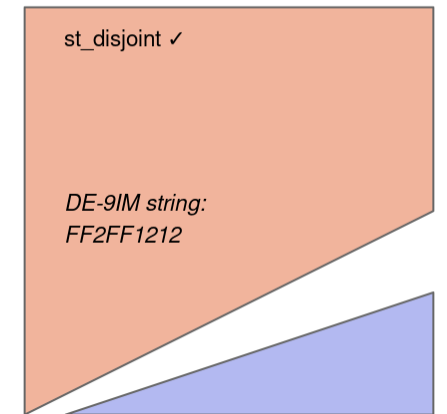
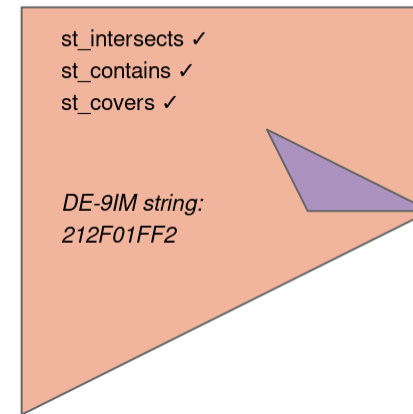
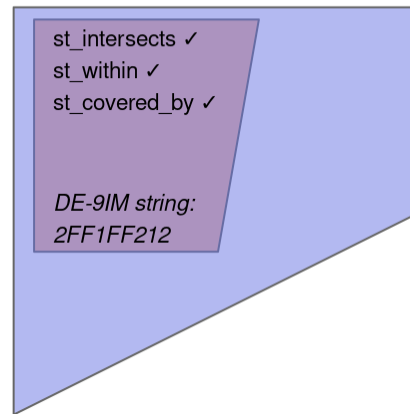
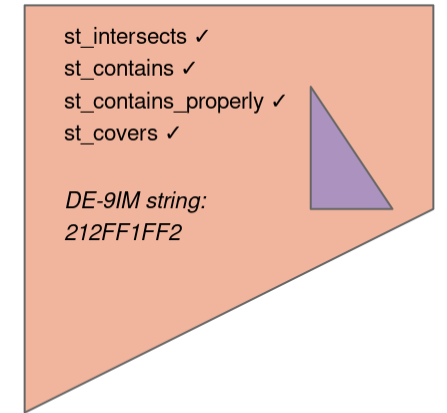
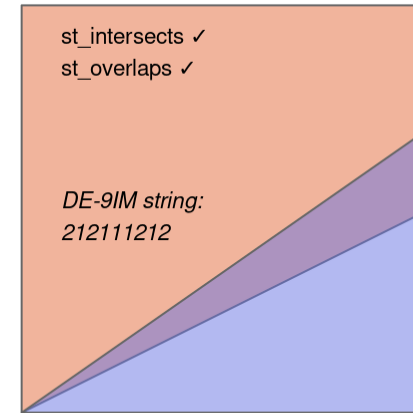
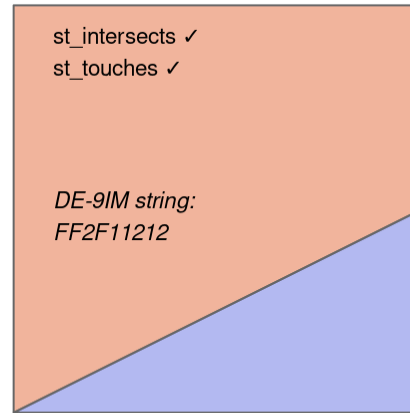
- Topological conversions
  - a) Convert vector/raster data types
    - Points to lines



# Spatial Data Operations

## ■ Topological relations

- a) Equals
- b) Intersects
- c) Crosses
- d) Touches
- e) Overlaps



<https://geocompr.robinlovelace.net/spatial-operations.html>

# Spatial Data Analysis

- Types of spatial data analysis
  - a) Overall Measures
  - b) Local Measures
  - c) Geographically weighted regression
  - d) Scan based measures

# Spatial Relationships

- How do we define what is a spatially related?

- a) Contiguity

- Contiguity means that two spatial units share a common border of non-zero length. Operationally, we can further distinguish between a *rook* and a *queen* contiguity, in analogy to the moves allowed for chess pieces.
    - Useful for polygon/line data



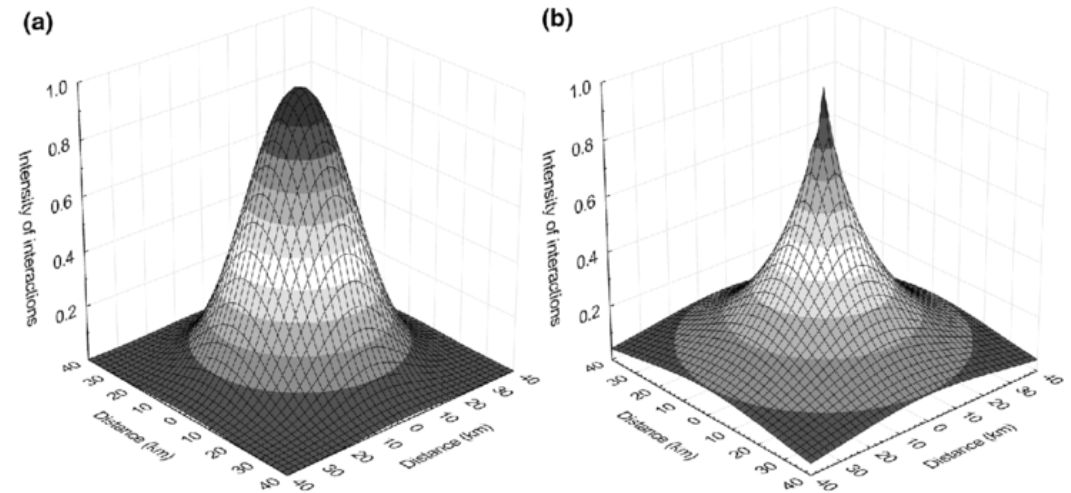


# Spatial Relationships

- How do we define what is a spatially related?

- a) Distance based weights

- Search radius
      - How far to look around
    - Decay function
      - Does the relationship change over space

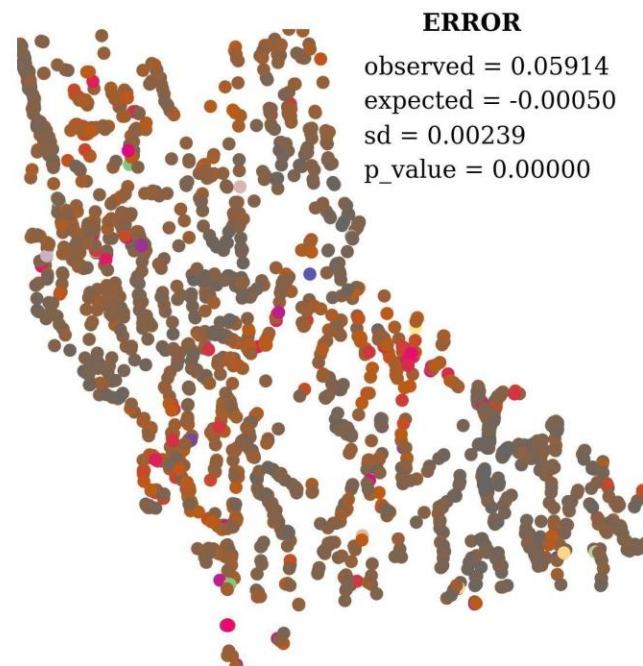
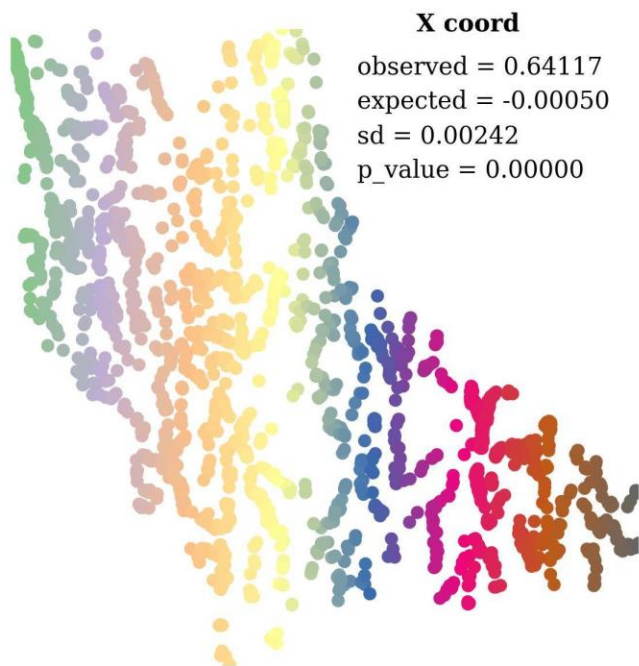
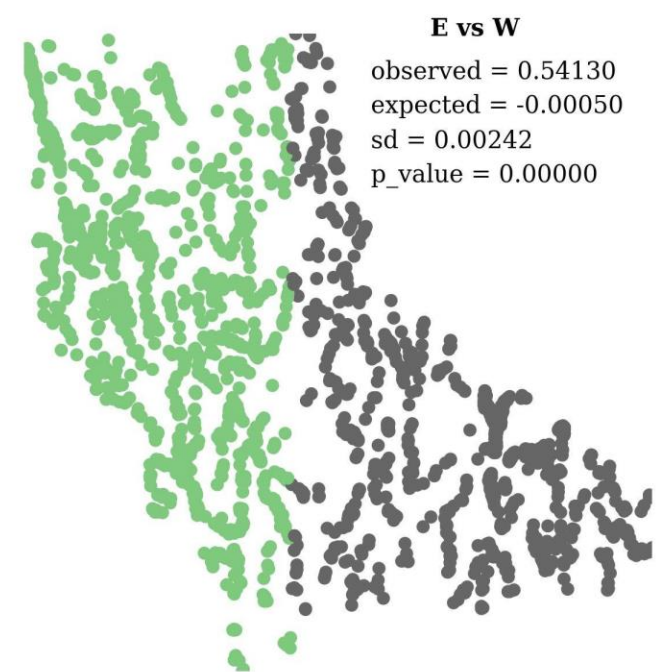
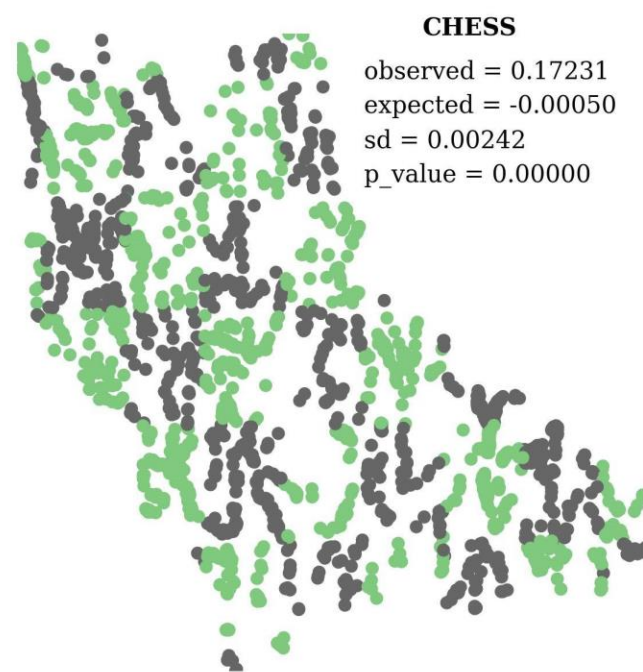
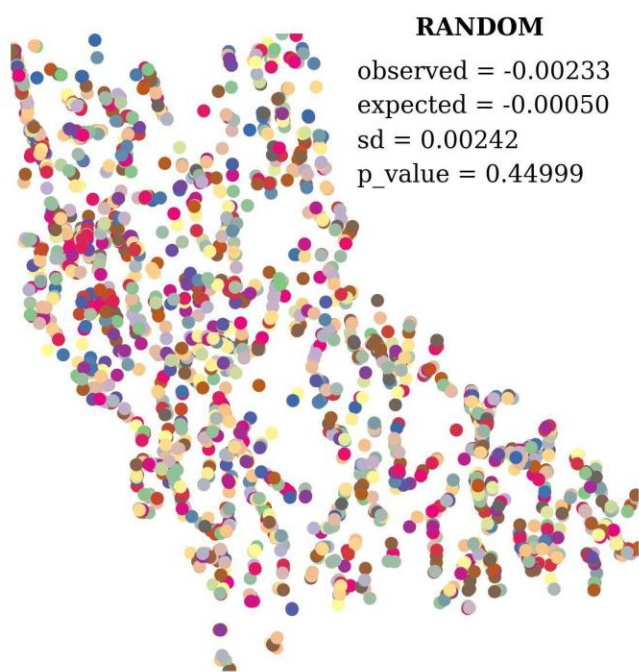


[https://www.researchgate.net/profile/Marian\\_Halas2/publication/272622119/figure/download/fig2/AS:613871756783616@1523369840018/D-model-for-distance-decay-functions-for-daily-travel-to-work-flows-to-Nitra-a-and.png](https://www.researchgate.net/profile/Marian_Halas2/publication/272622119/figure/download/fig2/AS:613871756783616@1523369840018/D-model-for-distance-decay-functions-for-daily-travel-to-work-flows-to-Nitra-a-and.png)

# Spatial Data Analysis

- Want to understand how things are clustered in space
- Want to know how two things are related when accounting for their relative spatial relationship
- Want to know how much spatial auto-correlation influences results of regression models
- ....







# Overall Measures

- Are points correlated in space?
  - a) Moran's I statistic, the null hypothesis states that the attribute being analyzed is randomly distributed among the features in your study area.
  - b) Can be used for point, polygon, or line data
  - c) Only tells us if there is a correlation somewhere in the data
  - d) ANOVA for spatial analysis
- Go to Toronto Public Health maps
  - a) Spatial Patterns?

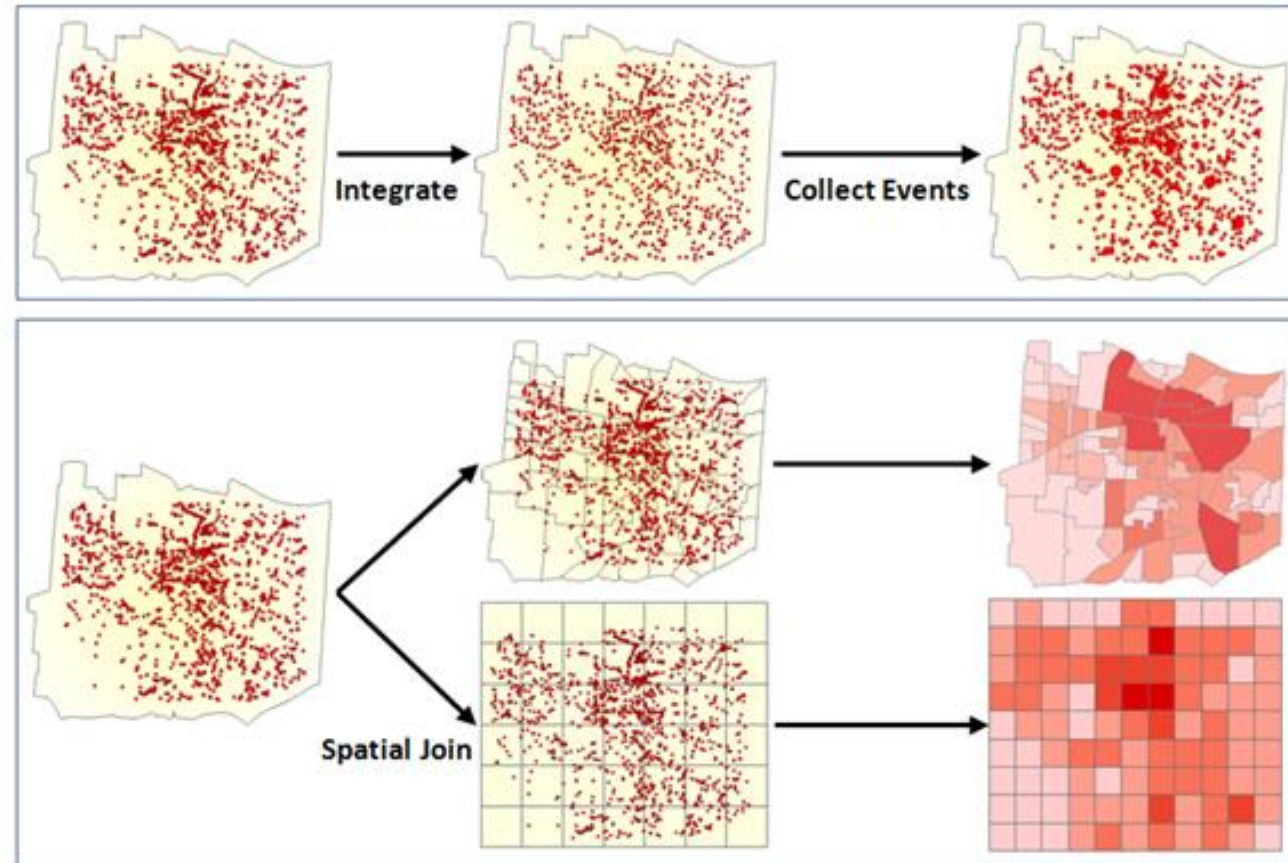
<https://public.tableau.com/app/profile/tphseu/viz/COVID-19-PublicFacingMap/MapDashboard>

# Local Measures

- Getis and Ord  $G^*$
- Getis and General Ord  $G$
  
- Where are points clustered in space?
  - a) Where are clusters in space?
  - b) Are cluster significantly related to each other?
  - c) Are clusters significantly different from other clusters?

# Getis and Ord G\*

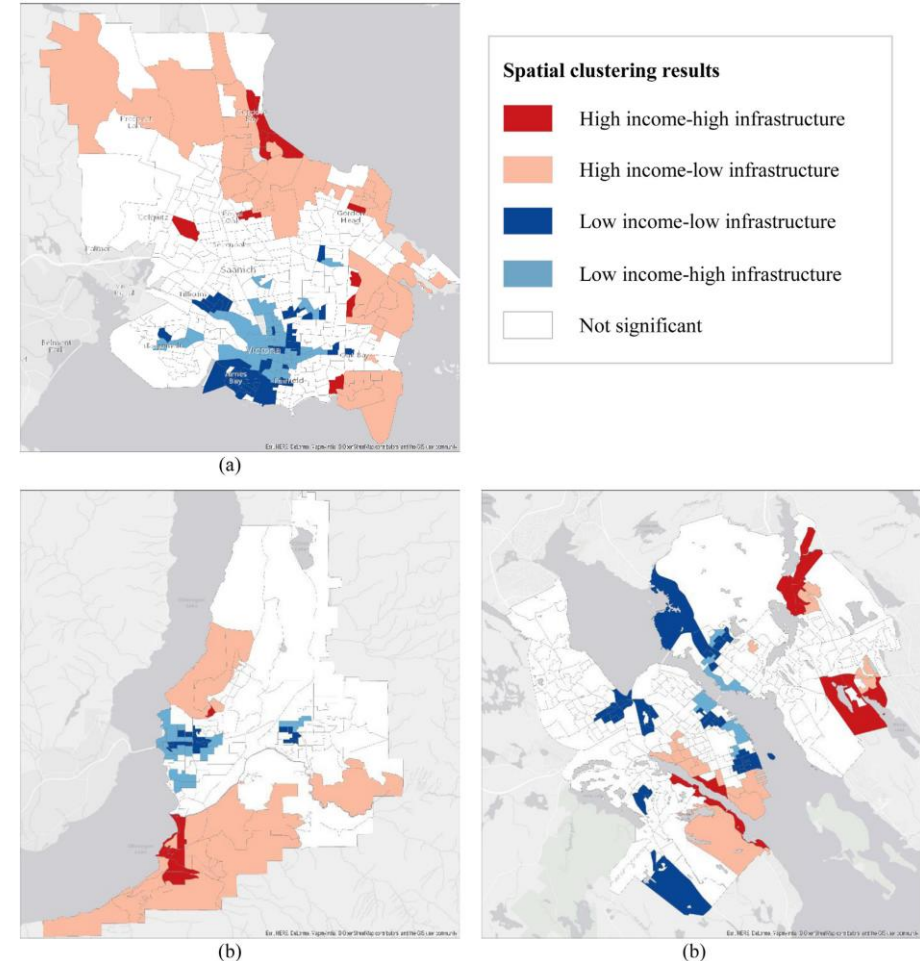
- Getis and Ord G
- Where are points clustered in space?
  - a) Where are clusters in space?
  - b) Are cluster significantly related to each other?
  - c) Are clusters significantly different from other clusters?





# Getis and General Ord G

- Getis and General Ord G
- Where are points clustered in space?
  - a) Are their high and low clusters in space that might be important?
  - b) Where are they?



# Geographically weight regression

- A regression where we include a spatial weights matrix (contiguity or distance based) to account for the spatial associations between data points (or polygons or lines).
- Can be done with many flavours of regression
  - a) Linear, logistic, multilevel models, count data models

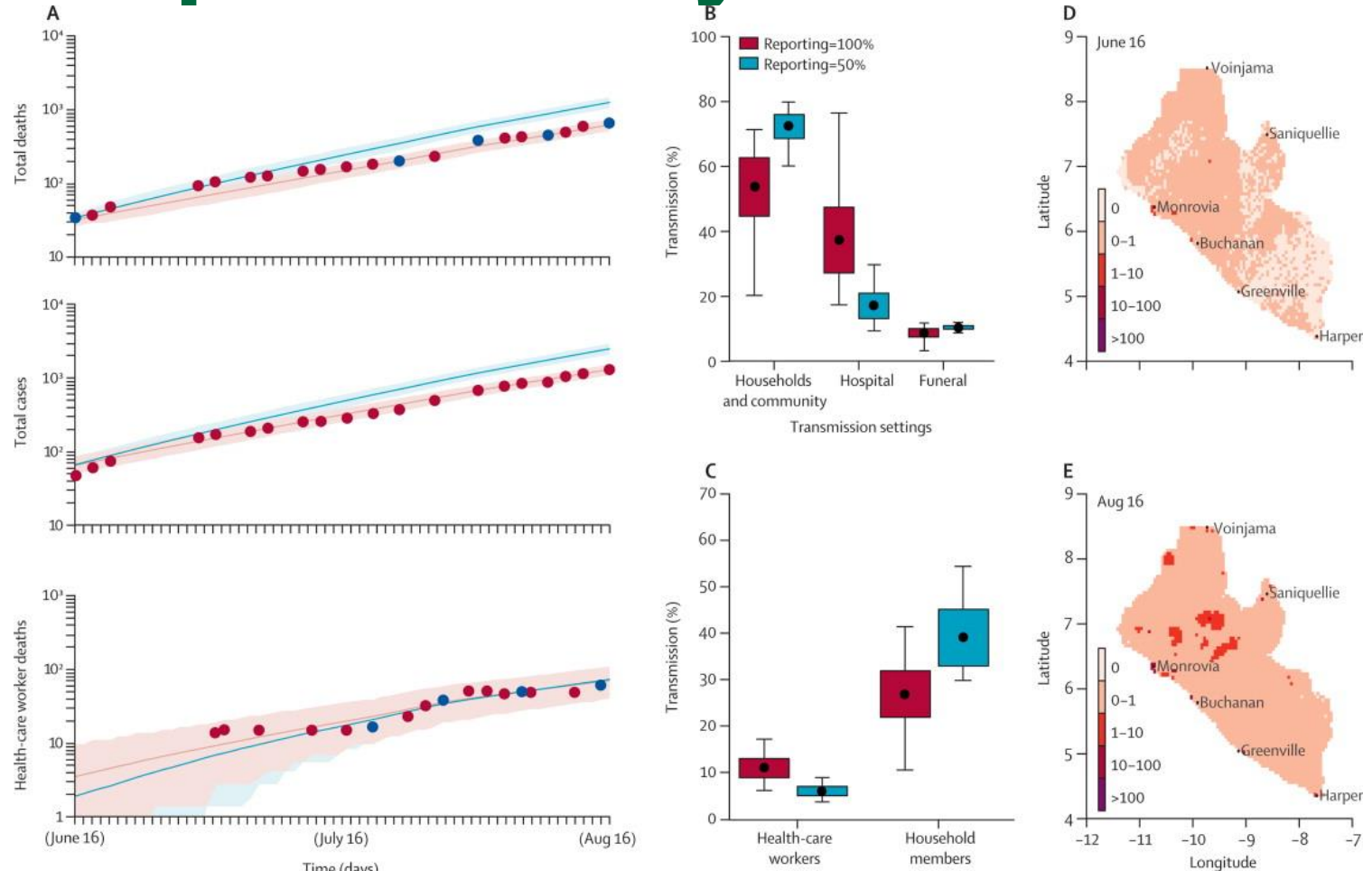
# Geographically weight regression

- Hailegebreal S, Haile F, Haile Y, Simegn AE, Enyew EB (2023) Using geographically weighted regression analysis to assess predictors of home birth hot spots in Ethiopia. PLOS ONE 18(6): e0286704.
- <https://doi.org/10.1371/journal.pone.0286704>

# Spatio-Temporal Analysis

- Spatio-Temporal Scan Statistics for cluster detection and change over time

a) Pretty Fancy – Not always necessary



# How to do this?

- Geographic Information Systems
  - a) A **geographic information system (GIS)** is a type of database containing geographic data.
- ArcGIS
  - a) Industry leader in GIS analysis. \$\$\$. License through USASK.
  - b) Python under the hood
- QGIS
  - a) Open source GIS analysis tool
  - b) Python under the hood

# How to do this?

- Programming
  - a) Integration with SPSS, SAS, and Stata is relatively poor (Don't do this)
  - b) R ([Geocomputation with R](#))
  - c) Python ([GeoPandas](#))
- Specific tools ([Geoda](#), [SaTScan](#))

# Cartography is Important

- <https://www.axismaps.com/guide/using-colors-on-maps>
- <https://colorbrewer2.org/>
- <https://storymaps.arcgis.com/stories/87d862191c1047d7be050cd6b77ffeae>



# Go find a map!

- Go find a map