Visualizing and Working with Spatial Data

MPIDR Workshop
Sean Reid
07/07/2022

Outline

- Types of data - points, lines, and polygons
- Datums, coordinate systems, and projections
- Trusting your data
- Modifiable areal unit problem (MAUP)
- Thinking like a geographer

Types of Data

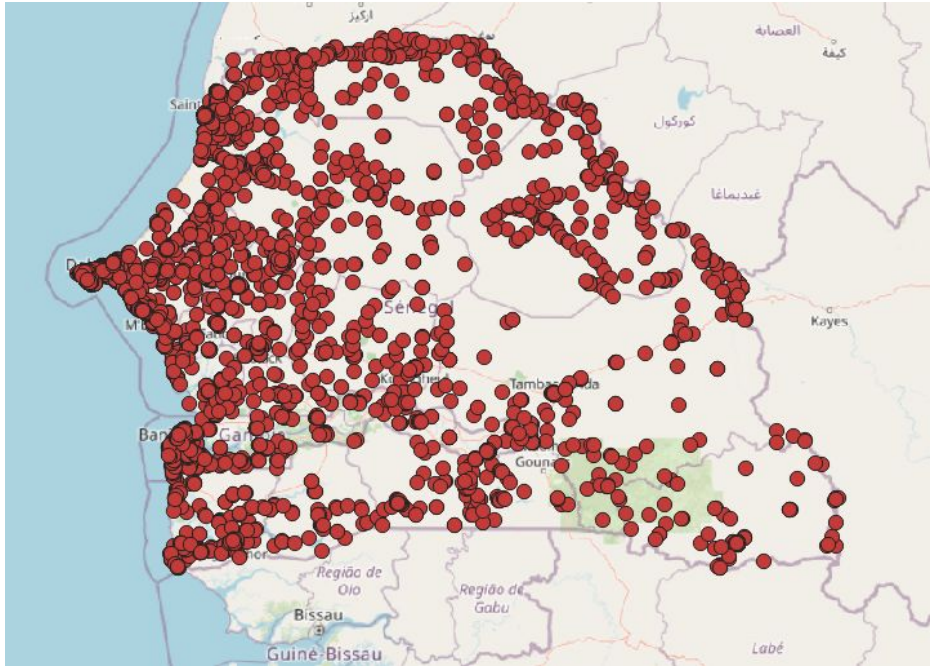
Smart people have come together and decided the best way to represent spatial data:

"*OGC standards are developed by members to make location information and services FAIR – Findable, Accessible, Interoperable and Reusable. They are used by software developers to build open interfaces and encodings into their products and services.*"

There are standards for data types, processes, APIs, etc. Really more information than you will ever need. Links for documentation can be found in the Jupyter Notebook.

Point Data

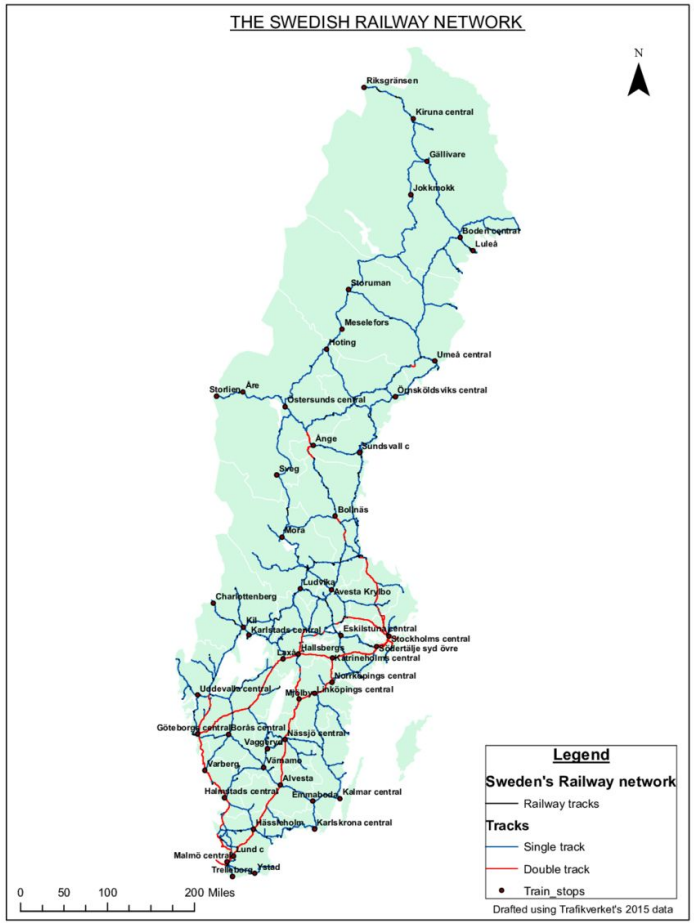
- Discrete location
- Has no area or length
- Can have nearly unlimited attributes
- Examples (not exhaustive)
  - Point of Interest
  - Restaurant
  - Home location
  - Weather station
  - Mountain tops
  - Medical providers
- When to use point data
  - Precision is important
  - Area or access to the point is not important



Medical Facilities in Senegal

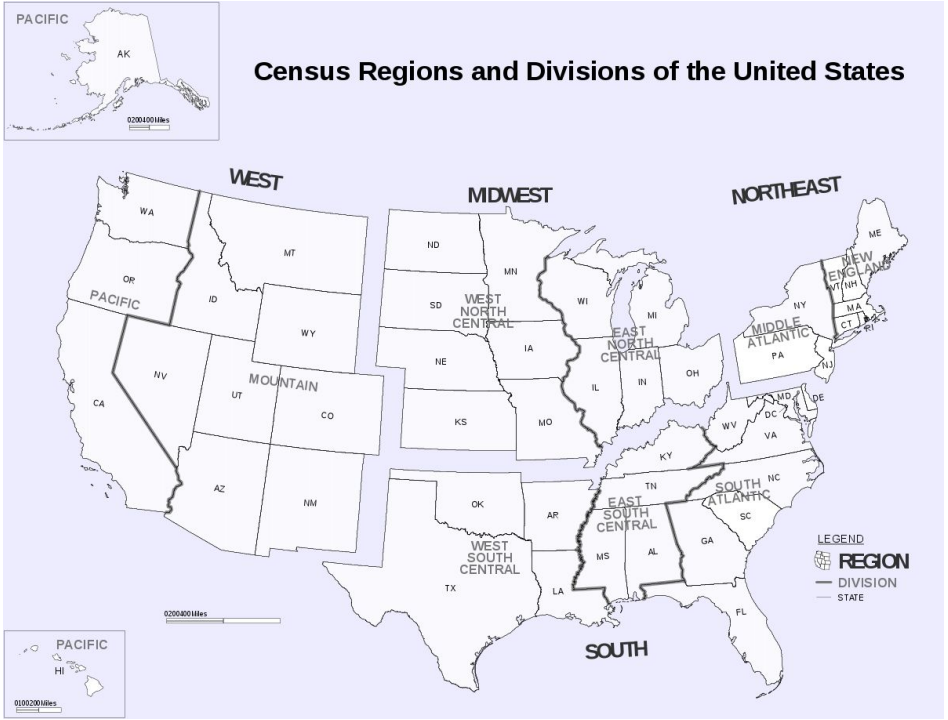
Line (Polyline) Data

- A collection of points connected by a line
- Has no area but does have a length
- Nearly unlimited attributes
- Examples (not exhaustive)
  - Roads
  - Rivers
  - Railways
  - Travel history
  - Movement patterns
- When to use line data
  - Representing a physical network like roads
  - Connection between points in important
  - Area is not important



Polygon Data

- A collection of points connected by lines and is closed into a shape
- Has area and length (even volume)
- Nearly unlimited attributes
- Examples
  - Administrative units
  - Boundaries
    - Schools
    - Parks
    - Neighborhoods
- When to use polygon data
  - When area is important
  - Representation of the extent is important



Datums, coordinate systems, and projections

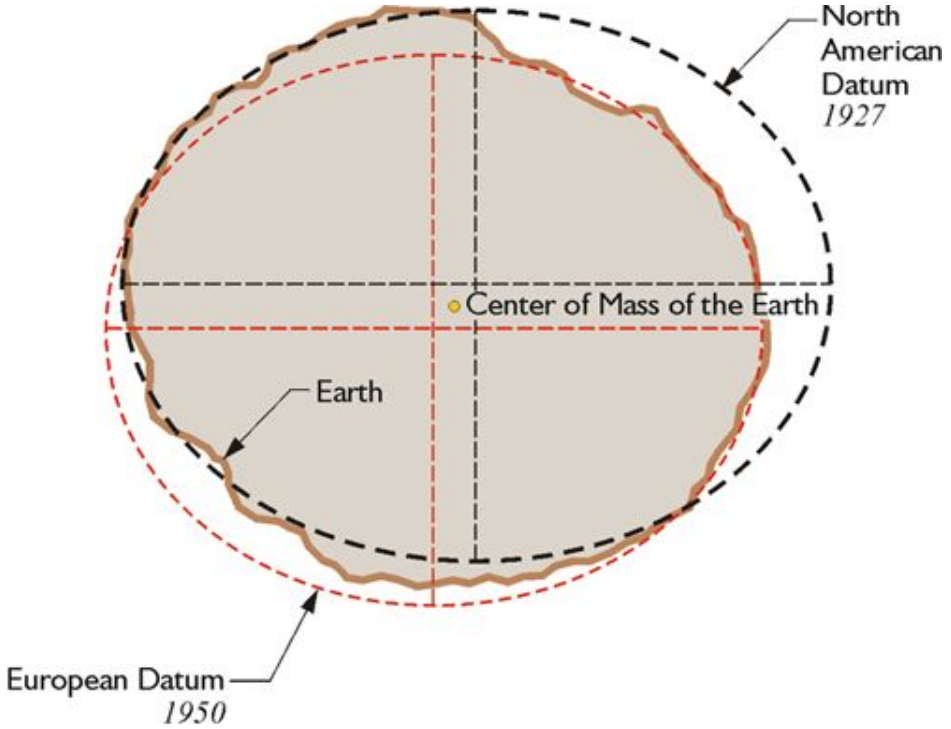
All of these types of data are great but if we can’t tie them to locations then they aren't super useful.

Definitions:

- Datum - A reference shape of the earth used for mapping
- Coordinate System - Units that define the location (meters, lat/lon, etc)
- Projection - 2D representation of the 3D world

Datum

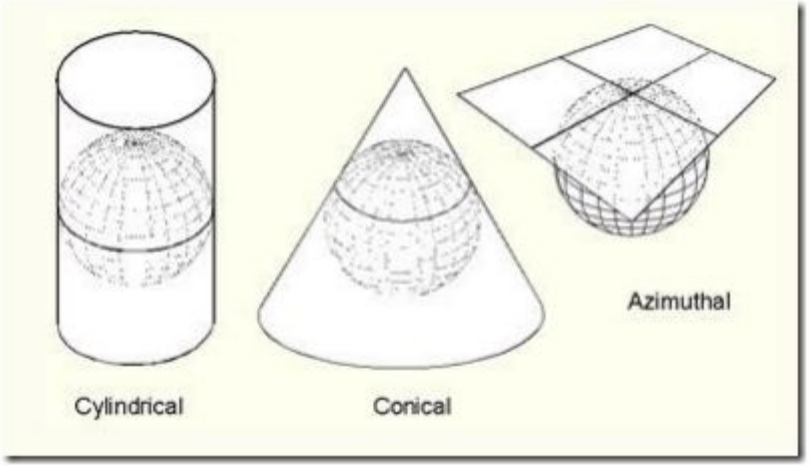
- Datums can be shifted to better approximate the shape of the earth in different places.
- Geodetic points are “tie” points where the ellipsoid touches the earth.
- WGS 84 is the most commonly used geodetic system
    - Google, Apple, OSM, etc



Projections

Different types:

- Conical
- Cylindrical
- Azimuthal



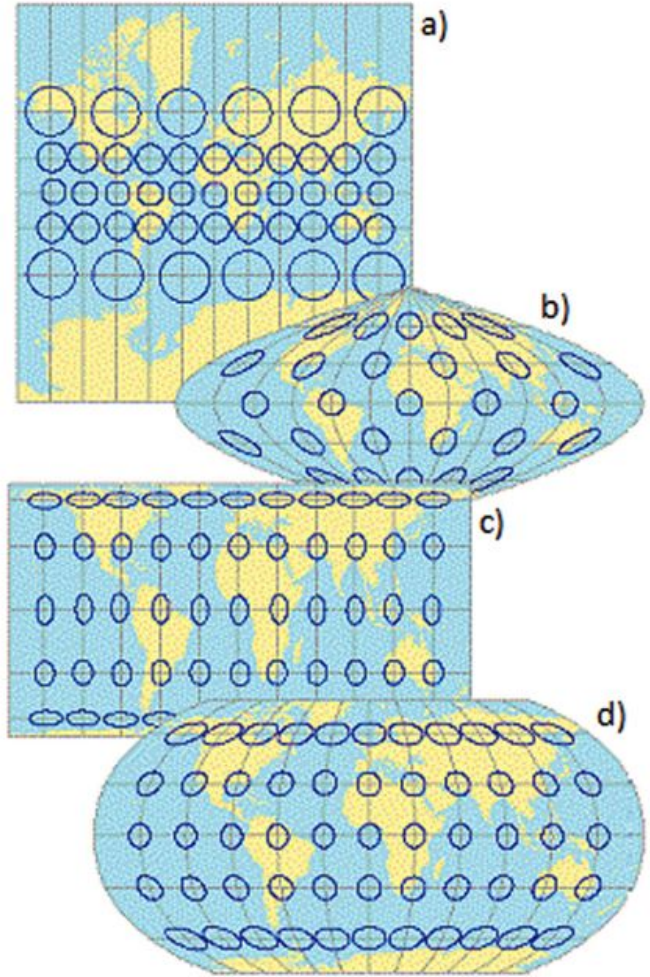
Projections

Different types:

- Conical
- Cylindrical
- Azimuthal

Important differences:

- Preserve shape
- Preserve area/distance
- Preserves some of both



Choosing the right projection

There are so many and it can be overwhelming. Honestly, just google it!

If you are working in a small geographic area (like a city) you should be able to find a good projection. If you are working in a broad area (like multiple countries) you need to be more careful and think about strengths and weaknesses between the projections.

Should I trust my spatial data?

It depends……

Most times you can use the spatial data as is but you should always think about how and why you are using it. You need to figure out if it makes sense.

Let’s think about state level population on a map as an example.

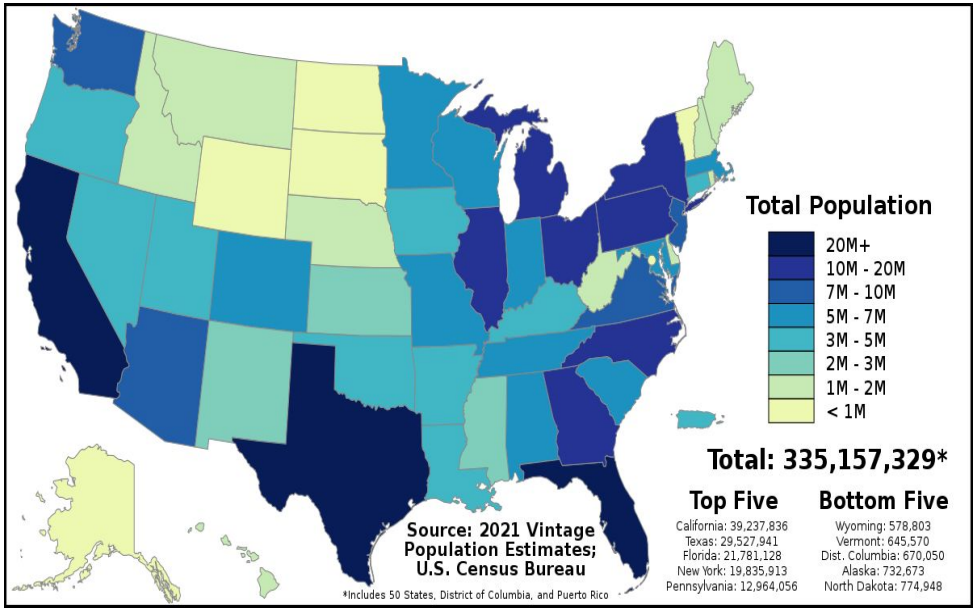
Should I trust my spatial data?

It depends……

Most times you can use the spatial data as is but you should always think about how and why you are using it. You need to figure out if it makes sense.

Let’s think about state level population on a map as an example.

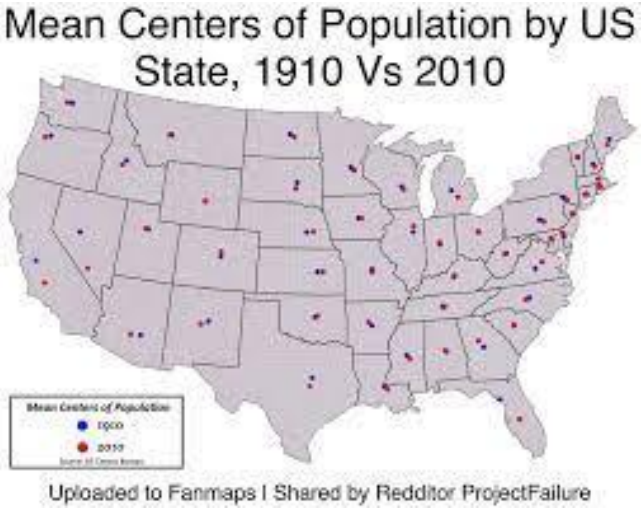
Most people are used to seeing something like this but it’s an abstraction of where people actually live.



Should I trust my spatial data?

Let’s change it up now and think about population represented as points.

We are still dealing with population but now it’s mean centers (a discrete location). The information is largely the same but the use case is different.



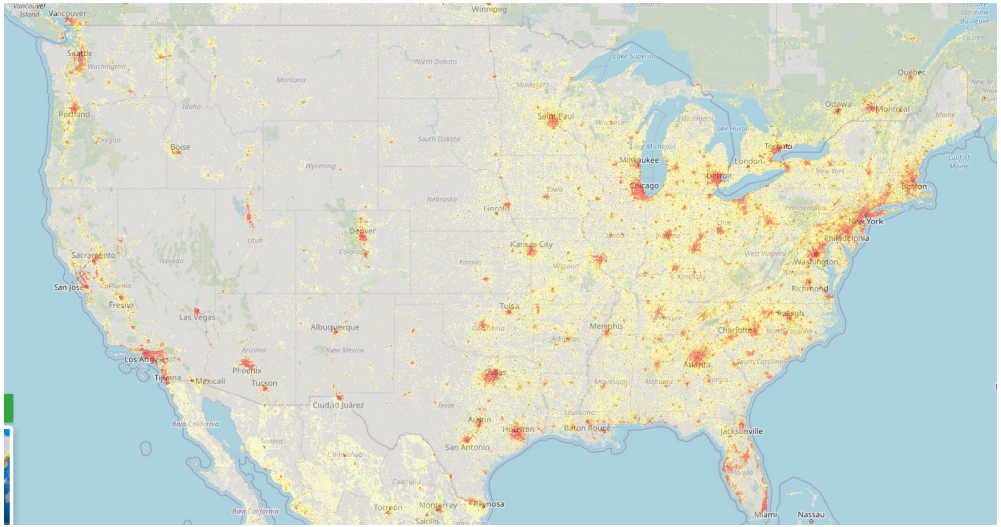
Should I trust my spatial data?

Let's change it up now and think about population represented as points.

We are still dealing with population but now it's mean centers (a discrete location). The information is largely the same but the use case is different.

If we really care about where populations are located we should use something like this.

Again, the underlying data hasn't changed much but what we are asking from the data drives how we present it.



Not all data is created equal

This has been the bane of my existence at times

Data from different sources report showing the same information but in practice they are different……. And it can ruin everything :(

Borders are particularly troublesome in this regard. So when you are mixing and matching data sources make sure you pay special attention to these types of issues.

MAUP

So far we've talked about:

- Spatial data types
- Datums, projections, and coordinate systems
- Data representation and when to use different types of spatial data

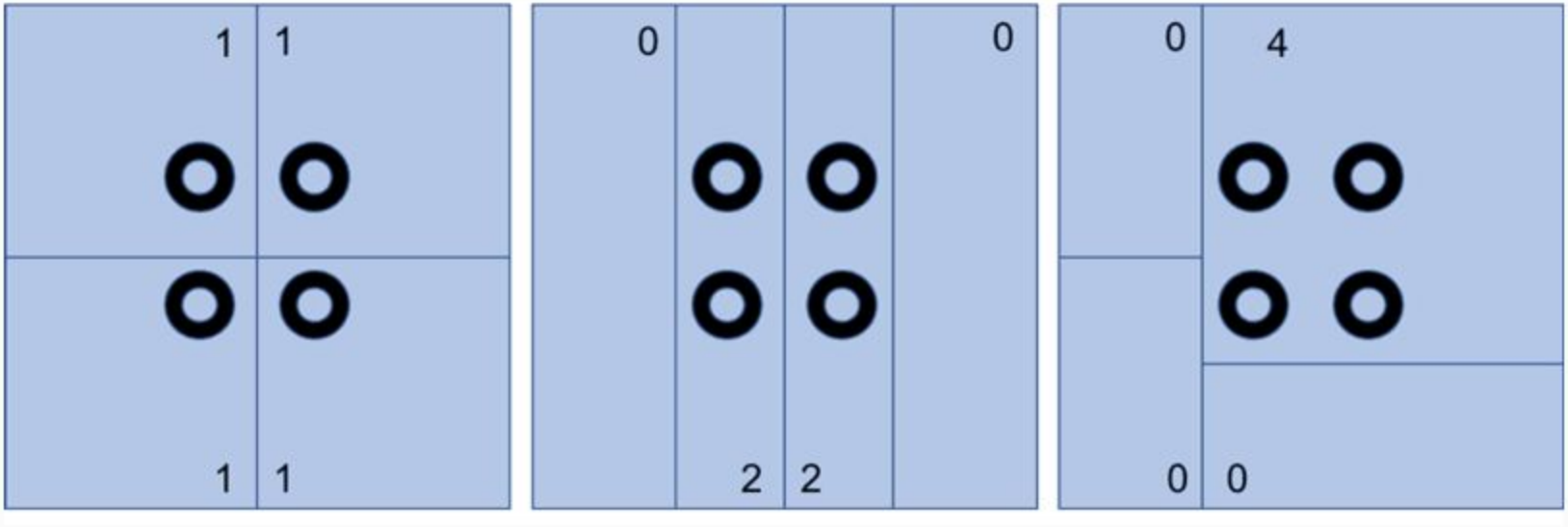
Now let's talk about scale!

MAUP

MAUP (Modifiable Areal Unit Problem) definition:

"*The modifiable areal unit problem (MAUP) is a statistical biasing effect when samples in a given area are used to represent information such as density in a given area. The area defined by an analyst is often arbitrary, thus measurement such as density could be deceptive because that density measure could have widely different results based on shape and scale chosen for analysis.*" - GIS Lounge, link

MAUP



How areas are aggregated can have an impact on the analysis of the results. Graphic: GISLounge

MAUP

Scale is important and should be considered in the context of the problem you are working on.

Sometimes the data available to you isn't at the ideal scale for privacy reasons. This can limit your analysis but is important for interpretation. Health data is an area where this is particularly problematic.

The framework for you analysis may be changed based on scale. For example I wrote an agent based model for migration behavior. The parameters in my models would change if I was modeling migration between countries, states, or counties.

Think Like a Geographer

There is more to spatial data than just information attached to points, lines, and polygons. There are geographic rules and laws that govern relationships between places. Thinking about these things can push the boundaries of your analysis.

Let's start with Tobler's first law of geography:

"*Everything is related to everything else. But near things are more related than distant things.*"

Seems simple but it can get complicated

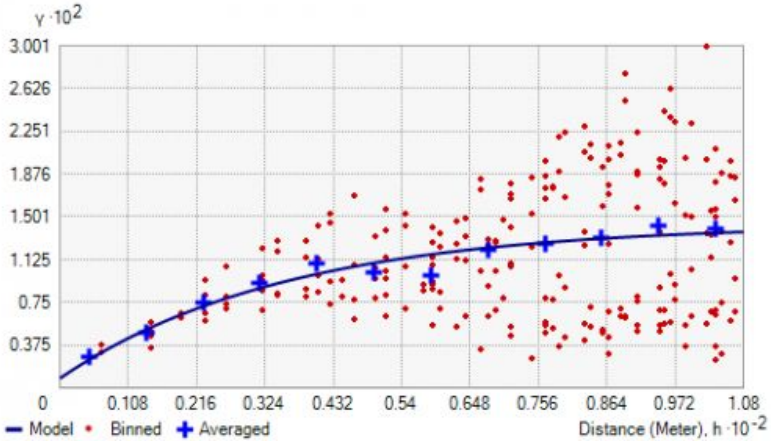
Think Like a Geographer

This law can be visualized several different ways but I think semivariograms are best.

We can see that as distance increases from a particular point things because less similar. Greater dispersion in our data.

Closely associated with distance decay and become muddied based on context.

Let’s look at grocery stores as an example



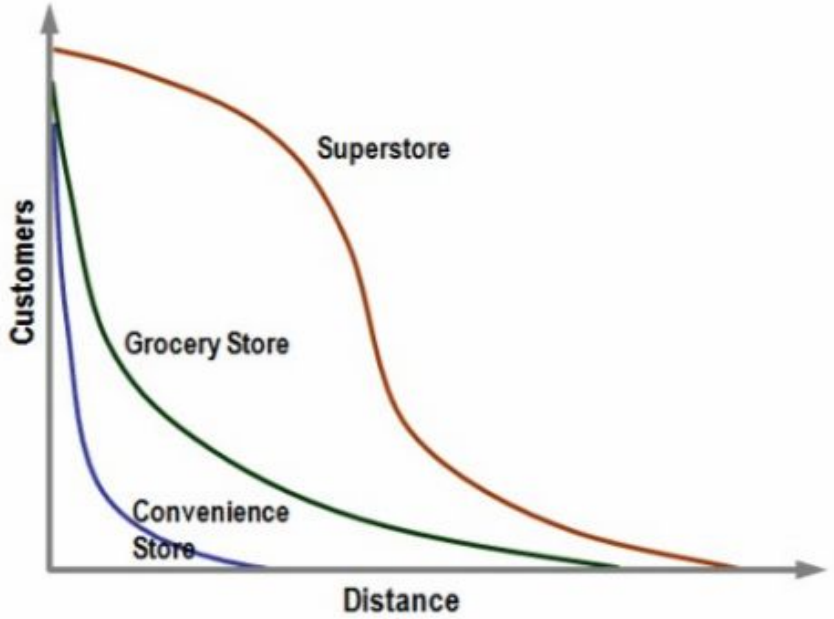
Think Like a Geographer

Huff's Gravity Model is a good example of how the attributes of a place can shift distance decay.

If a location like a grocery store has everything you could ever want (like a Walmart) then people will be willing to travel further to shop there.

Smaller corner markets have limited items and people will not travel to distant locations because the pay off isn't worth it.

This influences the spatial structure of these locations. Fewer Walmarts and more corner stores.

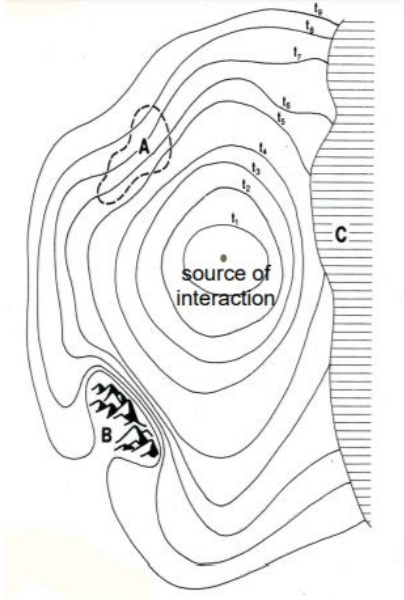


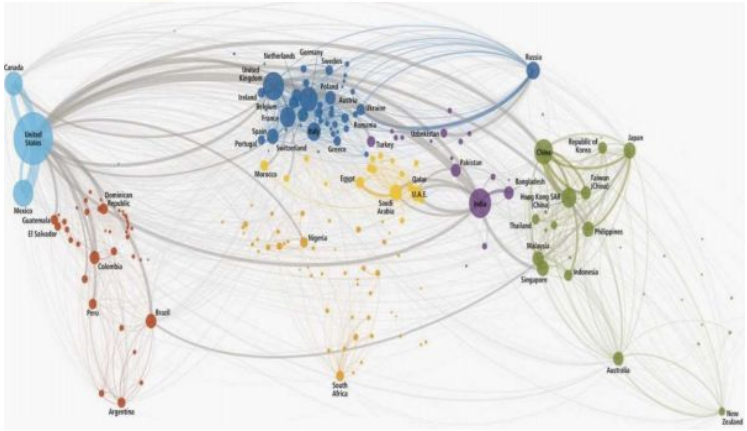
Think Like a Geographer

In our globalized world we have to think about space differently. Places are more connected than ever through digital and transportation technology.

This has collapsed geographic space and made physical distance less important than it used to be.

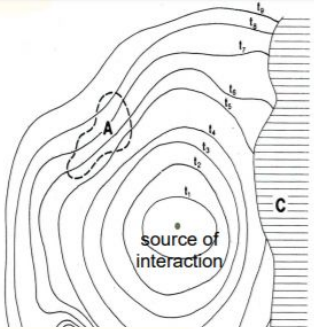
For example, New York City is more similar to London than the more proximate Syracuse. There are barriers that can limit geographic interaction: physical, socio-cultural, and psychological.





Distance decay or modified distance decay?

Think Like a Geographer



In our globalized world we have to think about space di
ever thr
technol

This has
physical
be.

For exar
London
are barr
physical

Distance decay or modified distance decay?

Break Time :)

Let’s take a break and then move on to some code where we work with spatial data and explore spatial relationships