ESDA 1: Distance

Class #16 | Geog 215

Introduction to Spatial Data Science

Spring 2020

> *“Everything is related to everything else, but near things are more related than distant things"* - Waldo Tobler

Quick Recap from last class

Exploratory (Spatial) Data Analysis

Two questions to consider:

- What type of geographic variation occurs *within* my variables?
- What type of geographic variation occurs *between* my variables?

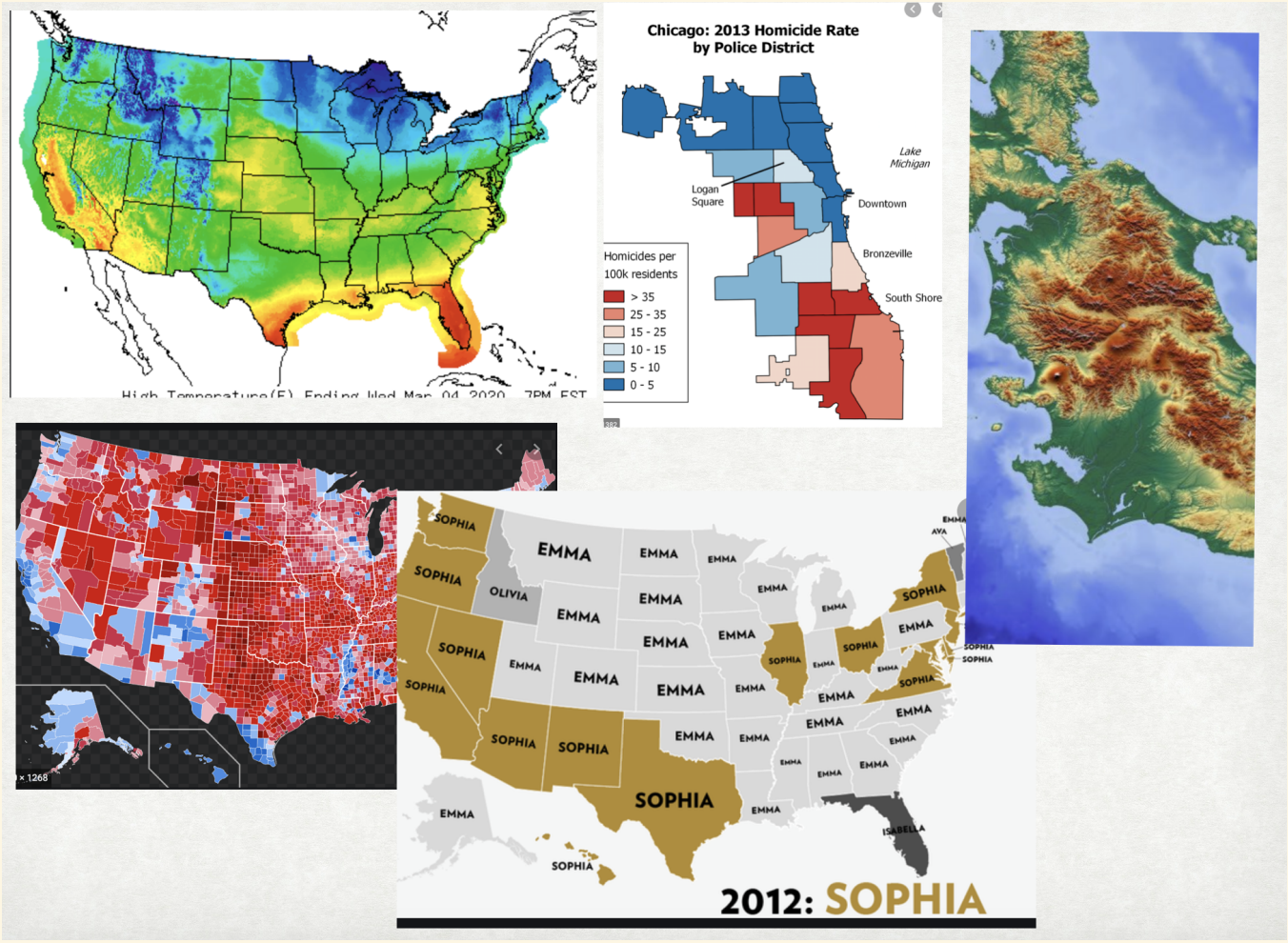
Why do we care?

Tobler's First Law of Geography (TFL)

"All things are related, but nearby things are more related than distant things"

- Spatial Dependence
  - What happens at one location is dependent on what happens at another location
- Underlines almost every physical and social phenomena

Tobler's First Law of Geography (TFL) in Action



Tobler's First Law of Geography (TFL)

"All things are related, but nearby things are more related than distant things"

- **Spatial Dependence**
  - What happens at one location is dependent on what happens at another location
- Underlines almost every physical and social phenomena
- If TFL werent' true
  - **GIS would be impossible**
  - **Life would be impossible**
  - **I will be out of work... :/**

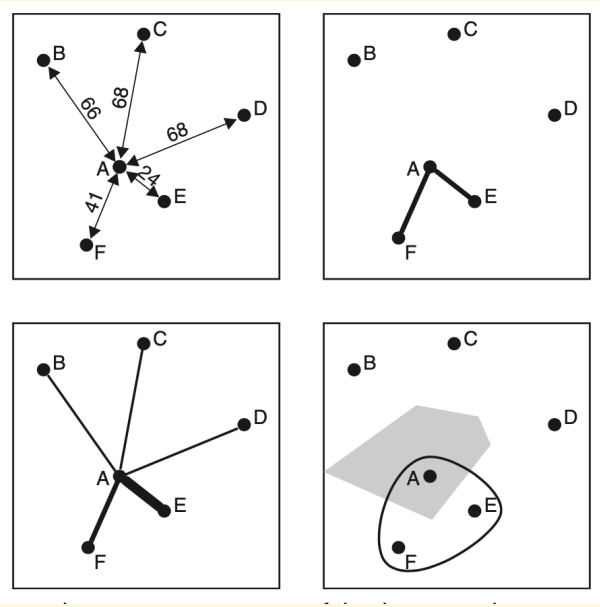
In Exploratory Spatial Data Analysis

We try to learn about spatial dependence/trends in our data

- What is the geographic variation within my variables
  - **At what spatial scale does my data look similar/dissimilar**
  - eg. Till what distance from a cafo do I see similar cancer rates ?
  - eg. Where are the outliers my in my data situated?
  - Can i predict missing data?
- What is the geographic variation between my variables
  - **Are 2 variables spatially correlated?**
  - eg - Do places with higher bednet coverage have lower rates of malaria? are there any places with high coverage of bednets, but still high malaria?

Quantifying Spatial dependence and relationships

- **Distance**
- Adjacency
- Interaction
- Neighborhood



Distance

- A fundamental geographic concept
  - **Affects relationships and interactions among things and places**
- Physical distance
  - **Separation in either space or time**
  - **Essential to understanding “space”**
  - **As an effect and/or as an outcome!**

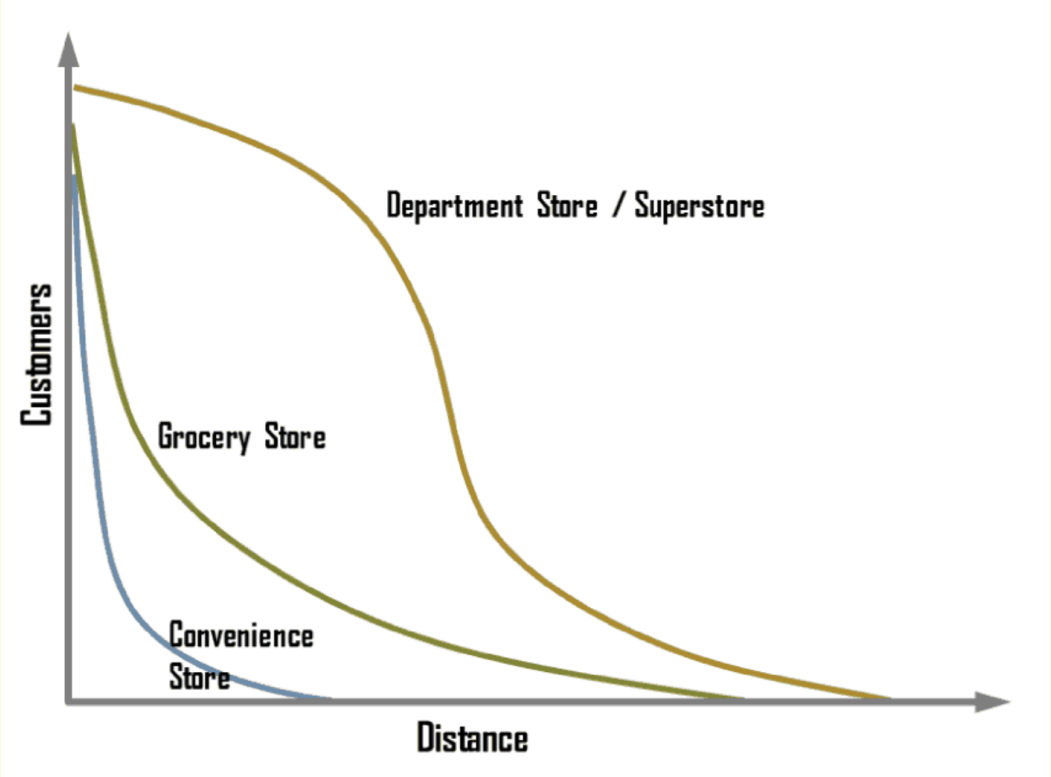
Distance

- Social Distance
  - **Connection and isolation**
  - **Formation of social networks**
  - Are some subgroups more likely to interact or not?
  - Information dissemination

Distance Decay

- Important concept in Geography
- Interaction (actual or potential) or relationship among phenomena decreases with increasing distance
    - **Distance has “friction”**

Distance Decay



Can you think of other examples?

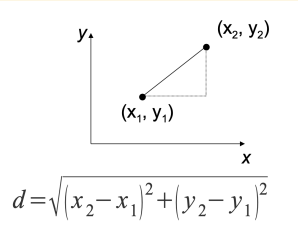
Why Distance

Distance is a fundamental geographic concept that affects relationships and interactions among phenomena

- Sometime we want to explain it
- Sometimes we want to measure it for input to other calculations
- Sometimes we want to map it

Euclidean Distance

- Based on **Pythogoras Theorem**
  - $a^2 + b^2 = c^2$
- Assumes:
  - A planar Earth (Project CS)
  - No differences in elevation
- Does not require a GIS
  - Simple Math Calculation
  - **Should not be used with latitude and longitude!!!**



Euclidean Distance

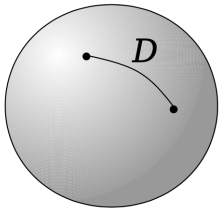
Limitations

- Assumes flat earth
  - Especially limiting over large distances
- Assumes no slope in terrain
  - Especially limiting in mountainous terrain
- Assumes no barriers to movement
  - Barriers can include anything that's not a road (in the case of vehicular travel) or anything on the landscape that stands in the way

Great Circle Distance

Great circle distance is the shortest distance between two points on a sphere

- **Used for points stored in spherical coordinates**
- **Required to calculate distance in spherical coordinates, especially if points are far apart**

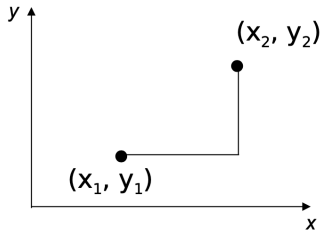


Point 1: (λ_1, Φ_1) ← longitude, latitude
Point 2: (λ_2, Φ_2) ← longitude, latitude
R: radius of the Earth (~6378km)

$$D = R \cdot \cos^{-1}((\sin\phi_1 \cdot \sin\phi_2) + (\cos\phi_1 \cdot \cos\phi_2 \cdot |\lambda_1 - \lambda_2|))$$

Manhattan Distance

- One way to account for barriers is to recognize that travel can be limited to transportation routes that follow cardinal directions (N, S, E, W)
  - #### In cities with a grid system
    - Eg. in Manhattan where most streets are at right angles



$$d = |x_2 - x_1| + |y_2 - y_2|$$

Minkowski Distance

- Falls between Euclidean and Manhattan distance
- Distance of a curved arc connecting two points

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_2)^2} \quad \textit{Euclidean}$$

$$d = |x_2 - x_1| + |y_2 - y_2| \quad \textit{Manhattan}$$

$$d = \sqrt[p]{(x_2 - x_1)^p + (y_2 - y_2)^p}$$

where $1 < p < 2$

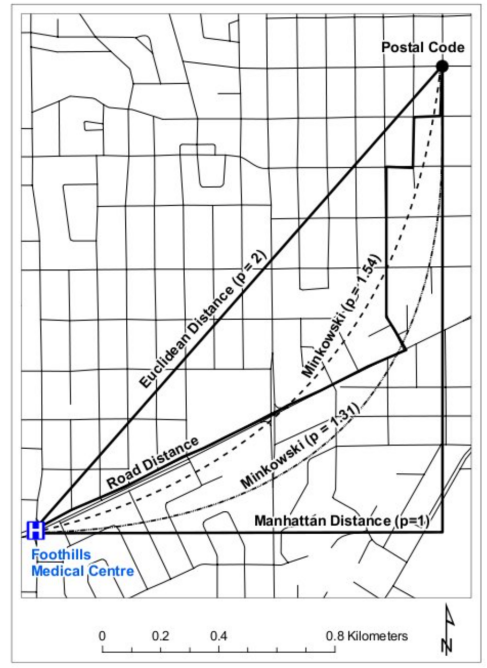


Figure 1
Visual illustration of road distance and distance metrics.

Network Distance

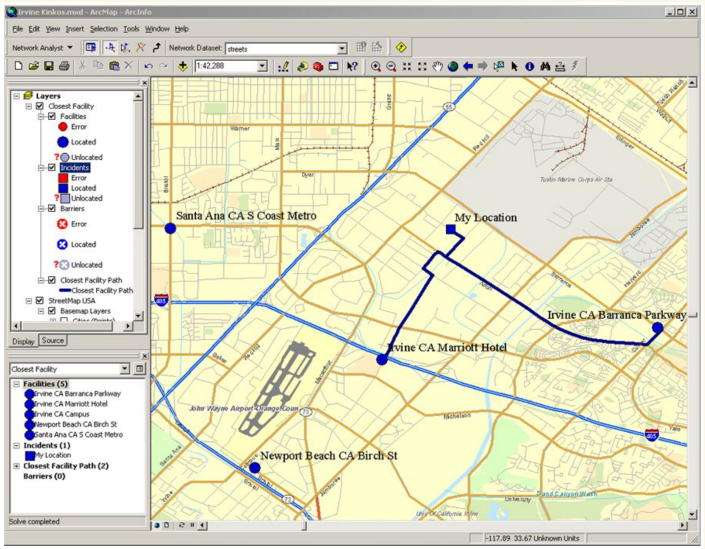
- Restricts travel along

  existing arcs (lines)

  - **e.g., on city streets**

- Distance is the sum of the lengths of arcs that make up the route

- Available as an API from ESRI, open street maps etc. !



Travel Time

More accurate portrayal of human movement

- Requires travel network or travel surface
  - **Available as an API from ESRI**
  - **Can be computationally intensive**
- Vehicular-based : US studies
  - **Augmented using alternative travel modes**
  - e.g., public transit, walking

In what situation is Euclidean Distance Appropriate?

POLL EVERYWHERE

Networks and Distance

In many cases, human movement is constrained to a network or network-like features

- **Point**: True "separation" among places should not be measured as straight lines, as this is not how people "overcome" distance
- **Counterpoint**: in some (many?) cases, Euclidean distances and network-based distances/times are highly correlated; hence Euclidean distances are a reasonable proxy

Measuring Distance

Basic approaches to measuring distance among things (for comparative purposes)

- Absolute distance
  - **Based upon the actual distance separating them**
- Relative distance
  - Nearest feature
  - **The nearest feature is considered a neighbor, or**
  - **Or, nearest k features**
- Topology-based
- **Connecting features are considered neighbors**

FINISH MIDTERM (Quiz closes indefinitely on Friday 11:55 pm)

--

Start thinking about project proposals

Enjoy Spring-Break

WASH YOUR HANDS, OFTEN!

Distance based queries

- Distance based operations may include:
  - **Spatial query**
  - Select geographic features based on distance
    - Example, distance to health care facility
  - **Buffer**
  - Use distance to define area around a point, line, or polygon feature
  - Number of grocery stores within walking distance
  - **Identify neighbors**
  - Identify geographic features within a neighborhood
  - Whether a Cafo exists in my neighborhood