Introducing Spatial Statistics

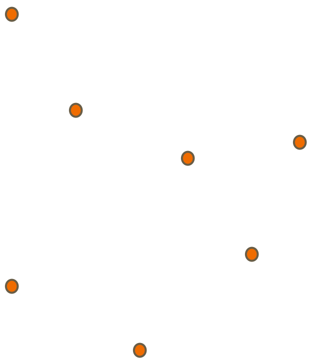
Lecture 11

Xiaofei Shi

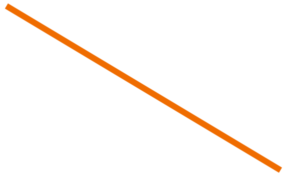
Learning Objectives

- Understand different types of spatial datasets
- Crash course on spatial statistics

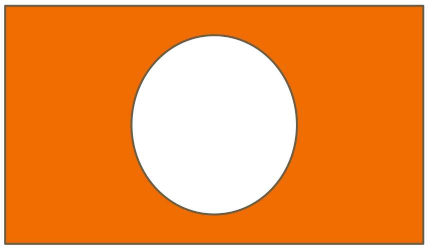
Types of spatial data



Points



Lines



Polygons

Points example - weather station measurements

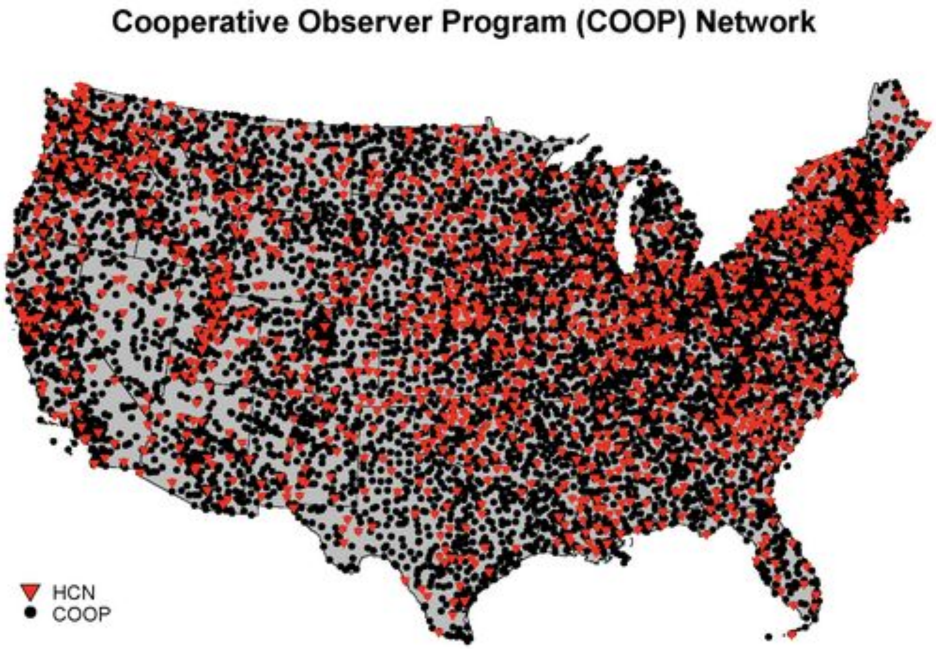
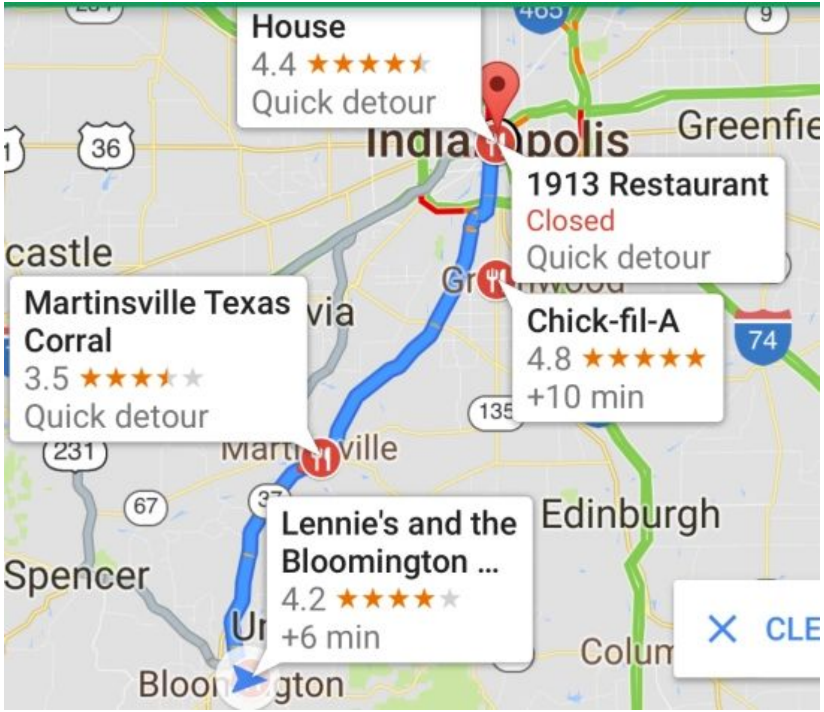
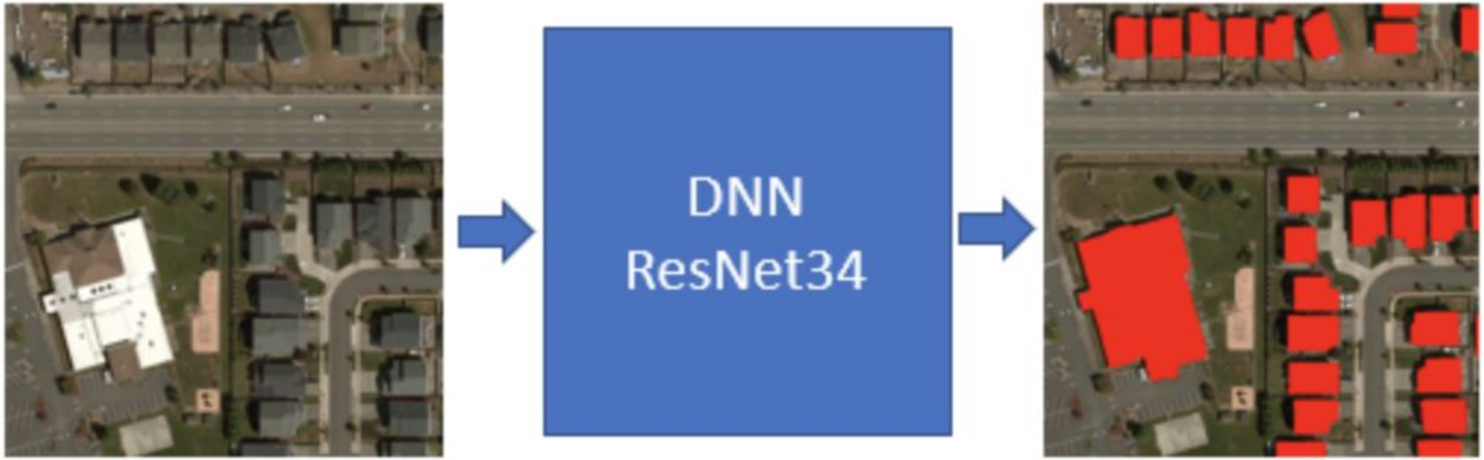


Photo credit from [NOAA USHCN website]

Line example - routes



Polyon Example - US Building Footprint



Thanks to Microsoft for the BuildingFootprint dataset!

Why are there different types of data?

We often talk about records being (X_1, X_2, X_3, ..., X_p)

- What is the distance between X_1 and X_2 ?
- Now what if one is a line and the other is a point?
- Now what if both are polygons?

Quick note on data on Earth - projections

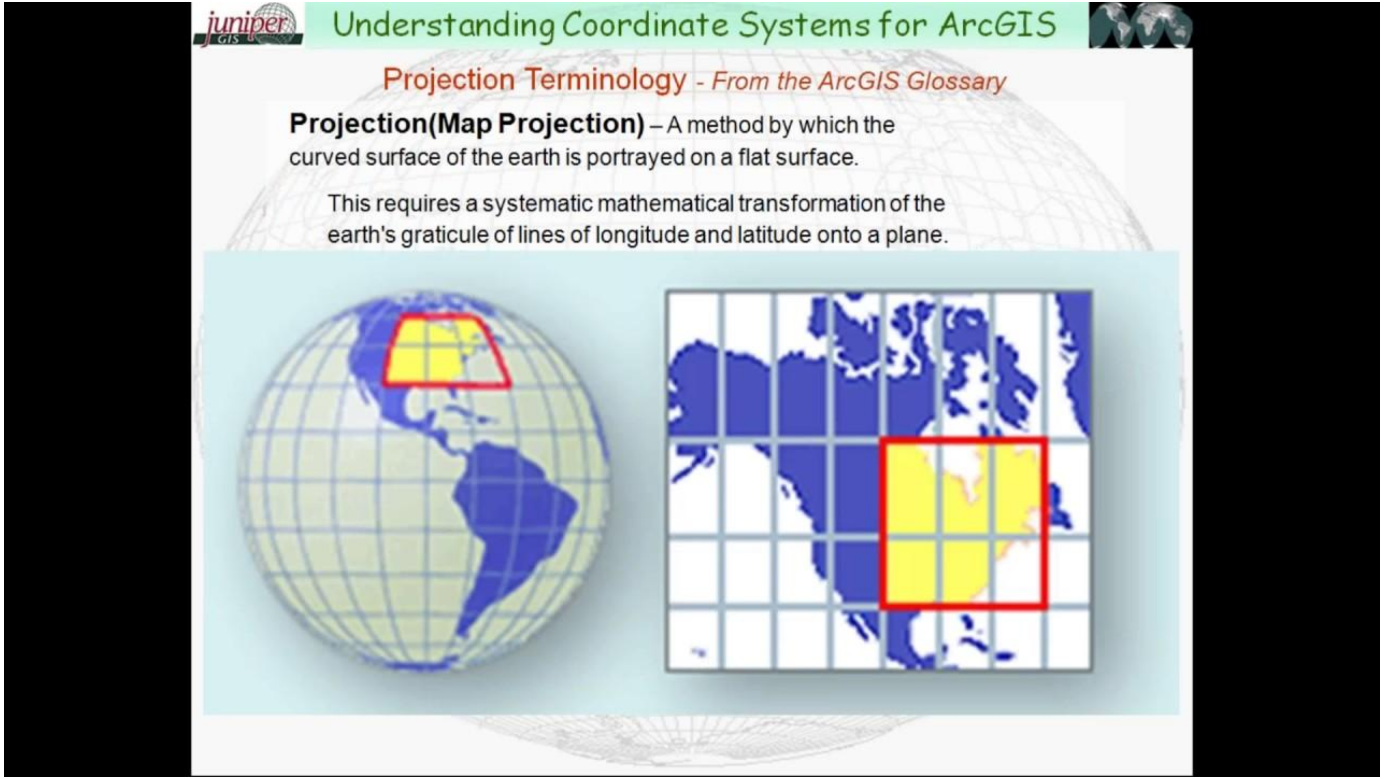
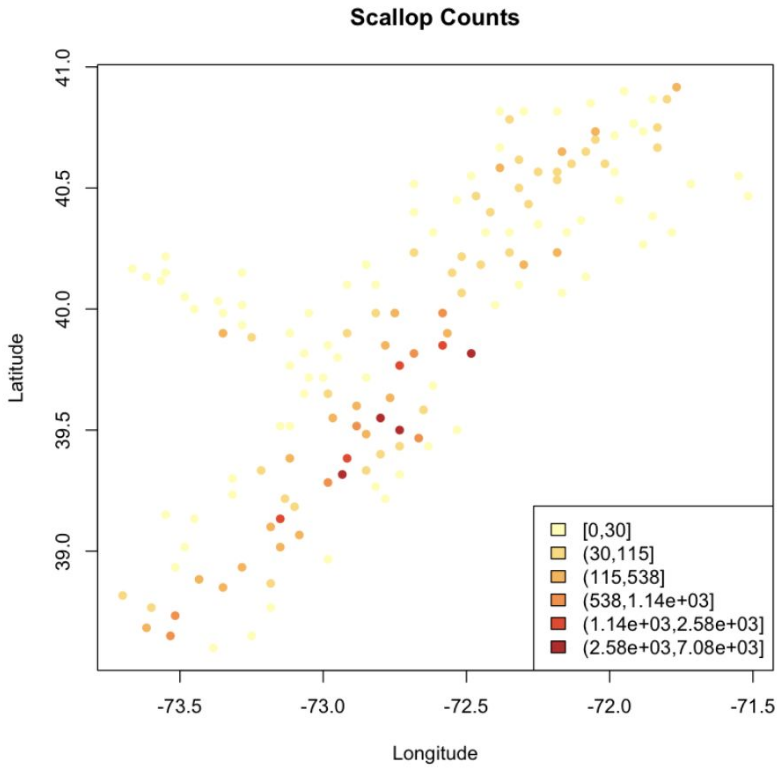
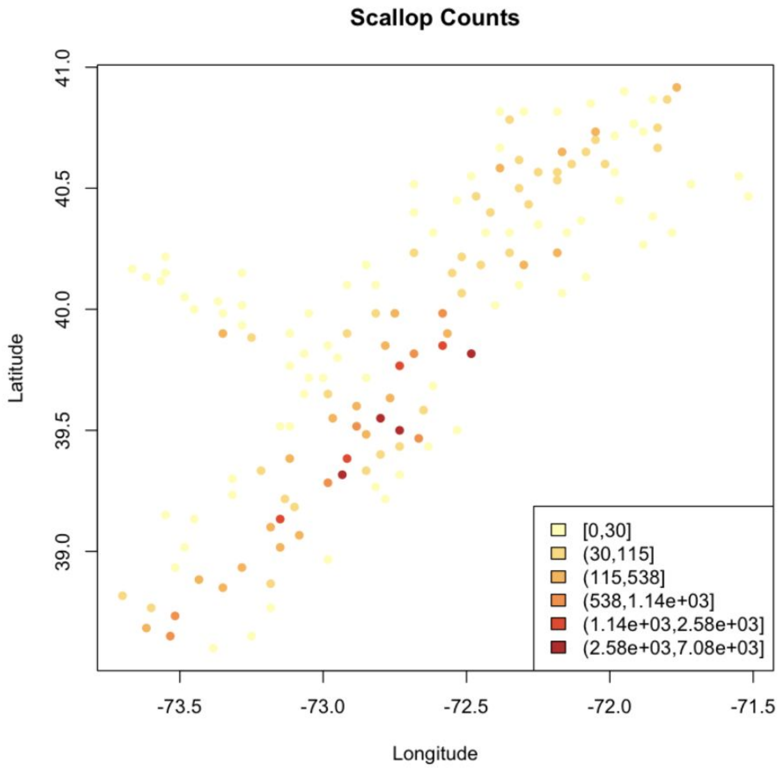


Image from YouTube

Spatial Statistics - modeling

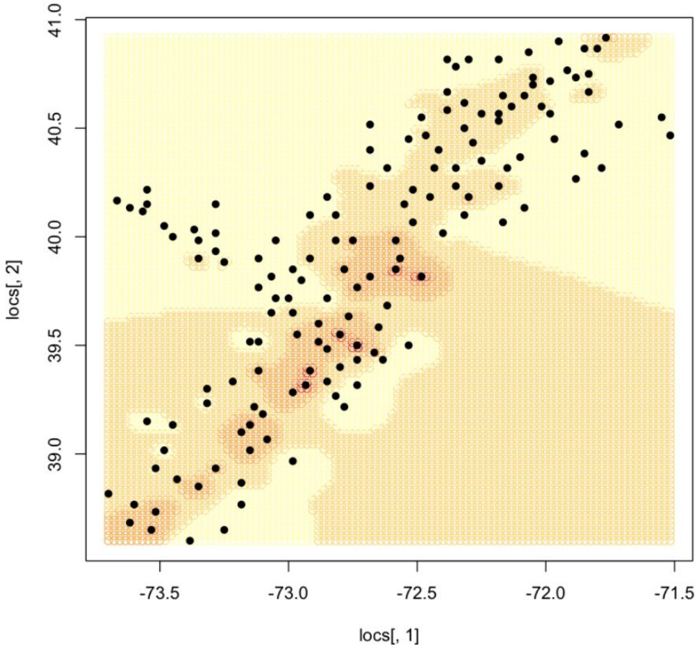


Spatial Statistics - modeling



- Our belief about the data at s_1 is influenced by our expectation at that location
- Deviations from the expectations is where the neighboring data points help us understand

Model - spatial data is a noisy observation of a surface



$$Y(s_1) = m(s_1) + \epsilon(s_1)$$
$$\epsilon(s_1)|\epsilon(s_2) = N(0, Cov(s_1, s_2))$$

Kriging - the OLS of spatial modeling

$$E(Y(s_1)|Y(s_2)) = m(s_1) + Cov(s_1, s_2)Cov(s_2, s_2)^{-1}(Y(s_2) - m(s_2))$$

Kriging formula explained

$$E(Y(s_1)|Y(s_2)) = m(s_1) + Cov(s_1, s_2)Cov(s_2, s_2)^{-1}\boxed{(Y(s_2) - m(s_2))}$$

- $Y(s_2) - m(s_2)$ is the error between our expectation and our training data

Kriging formula explained

$$E(Y(s_1)|Y(s_2)) = m(s_1) + Cov(s_1, s_2)\boxed{Cov(s_2, s_2)^{-1}}(Y(s_2) - m(s_2))$$

- $Y(s_2) - m(s_2)$ is the error between our expectation and our training data
- $Cov(s_2, s_2)^{-1}$ discounts the correlation between our training data points.

Kriging formula explained

$$E(Y(s_1)|Y(s_2)) = m(s_1) + \boxed{Cov(s_1, s_2)}Cov(s_2, s_2)^{-1}(Y(s_2) - m(s_2))$$

- $Y(s_2) - m(s_2)$ is the error between our expectation and our training data
- $Cov(s_2, s_2)^{-1}$ discounts the correlation between our training data points.
- $Cov(s_1, s_2)$ picks up covariance between the training locations and our locations of interest

Kriging formula explained

$$E(Y(s_1)|Y(s_2)) = \boxed{m(s_1)} + Cov(s_1, s_2)Cov(s_2, s_2)^{-1}(Y(s_2) - m(s_2))$$

- $Y(s_2) - m(s_2)$ is the error between our expectation and our training data
- $Cov(s_2, s_2)^{-1}$ discounts the correlation between our training data points.
- $Cov(s_1, s_2)$ picks up covariance between the training locations and our locations of interest
- $m(s_1)$ is what we expected at location s_1

Choices for kriging - what exactly are these values?

$$E(Y(s_1)|Y(s_2)) = m(s_1) + Cov(s_1, s_2)Cov(s_2, s_2)^{-1}(Y(s_2) - m(s_2))$$

Choices for kriging - mean is a vector

$$E(Y(s_1)|Y(s_2)) = m(s_1) + Cov(s_1, s_2)Cov(s_2, s_2)^{-1}(Y(s_2) - m(s_2))$$

$$m(s) = 0$$

Choices for kriging - covariance is a matrix

$$E(Y(s_1)|Y(s_2)) = m(s_1) + Cov(s_1, s_2)Cov(s_2, s_2)^{-1}(Y(s_2) - m(s_2))$$

$$m(s) = 0$$

$$Cov(s_1, s_2) = \sigma^2 \exp(-\frac{d(s_1,s_2)}{\theta})$$

Choices for kriging - distance is a matrix

$$E(Y(s_1)|Y(s_2)) = m(s_1) + Cov(s_1, s_2)Cov(s_2, s_2)^{-1}(Y(s_2) - m(s_2))$$

$$m(s) = 0$$

$$Cov(s_1, s_2) = \sigma^2 \exp(-\frac{d(s_1,s_2)}{\theta})$$

$d(s_1, s_2)$ can be the great circle distance