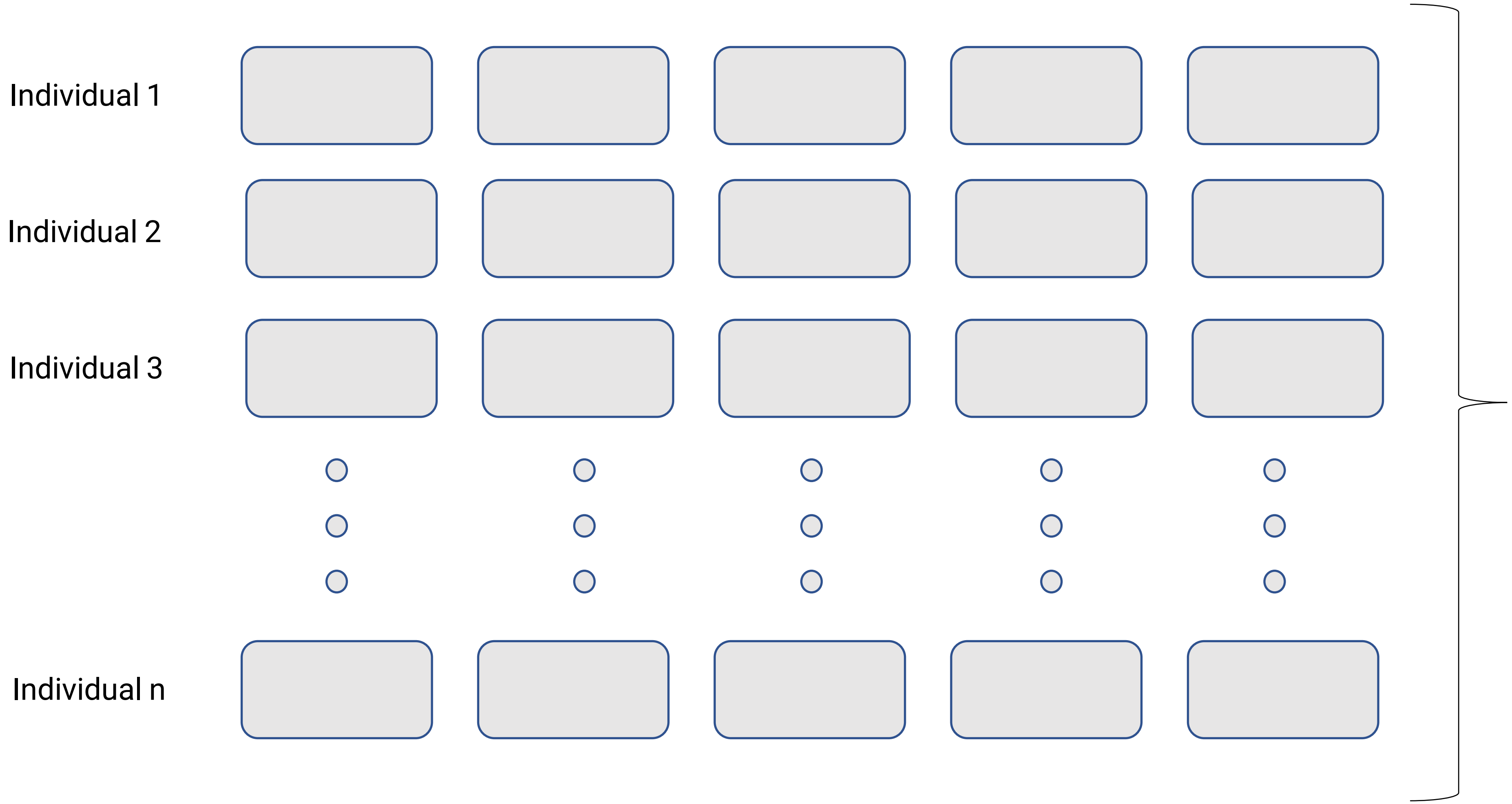
Spatial regression

Fifth Session

Orlando Sabogal-Cardona
PhD researcher
University College London UCL

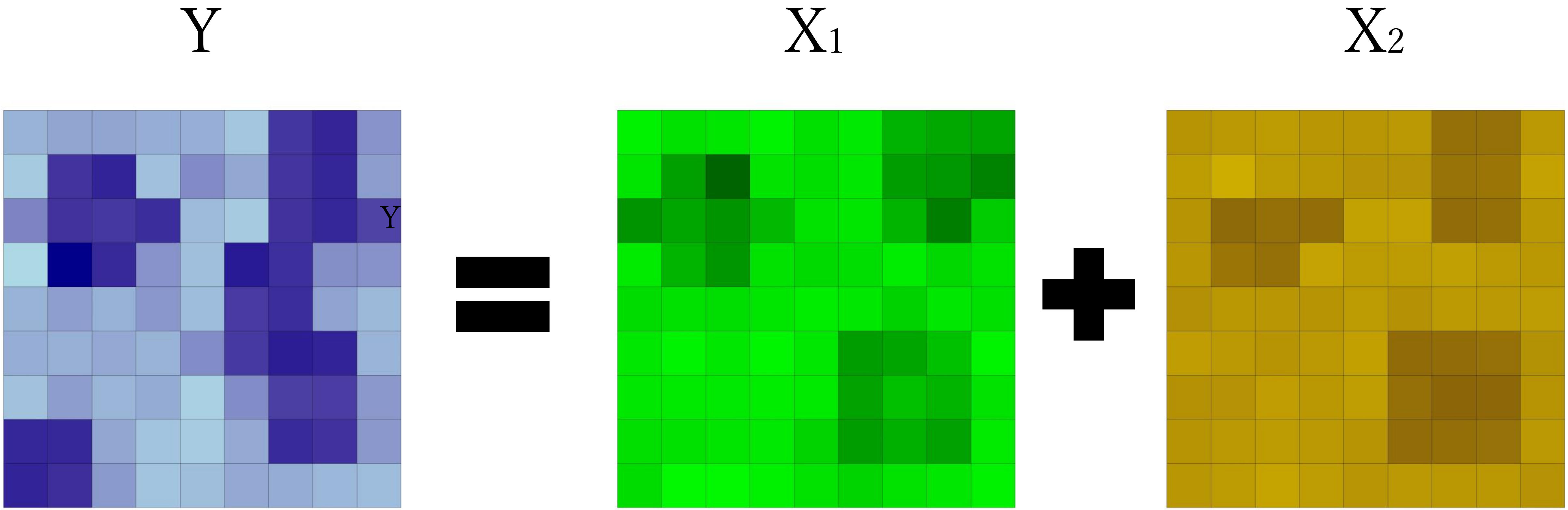
Individuals can have a “spatial” dimension



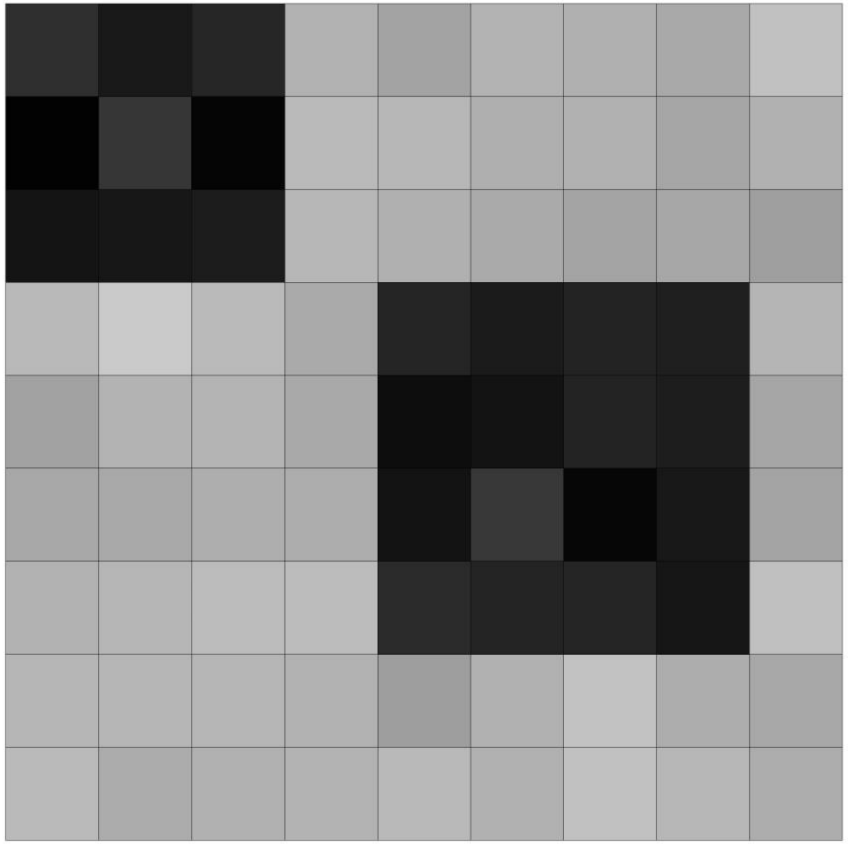
They can be cities, neighborhoods, TAZ, counties.

In short:
geolocation + database

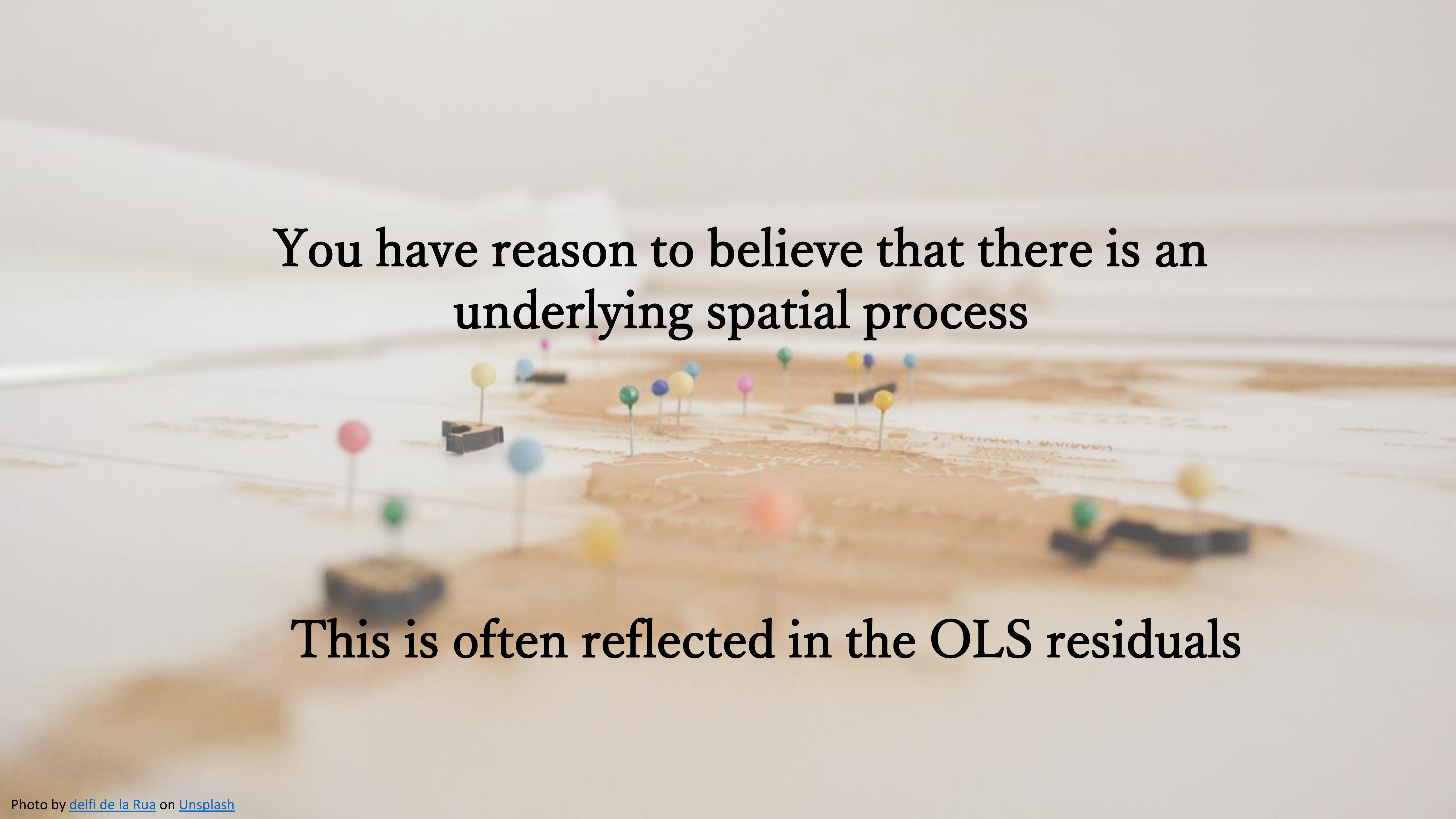
You can run a linear regression:



What about this residual map?



You can run a Moran's I on the residuals



You have reason to believe that there is an underlying spatial process

This is often reflected in the OLS residuals

Photo by delfi de la Rua on Unsplash

Photo by Steven Weeks on Unsplash

How to account for geography?

Photo by Honey Yanibel Minaya Cruz on Unsplash

In simple terms, there are three main alternatives:

On the dependent variable (Y)

On independent variable (X)

On the error term (e)

In simple terms, there are three main alternatives:

On the dependent variable (Y) → The outcome at a specific location is influenced by the values of the same variable at neighboring locations

On independent variable (X) → The outcome is influenced by the characteristics of neighbours

On the error term (e) → Focuses on the spatial structure of the error term. The error term represents the unobserved factors that affect the dependent variable but are not accounted for by the independent variables

In simple terms, there are three main alternatives:

On the dependent variable (Y) → Spatial Autoregressive SAR model
Spatial Lag
Spatially Lagged Y

On independent variable (X) → Spatially Lagged X SLX
Spatially Lagged Explanatory Variables

On the error term (e) → Spatial Error Model SEM
Spatial Autoregressive Error Model
Spatially Lagged Error

Spatial Autoregressive SAR model

$$Y = \rho WY + X\beta + \epsilon$$

W: The spatial weights matrix

ρ: The spatial autoregressive parameter, which represents the strength of spatial dependence. It indicates how much the values of the dependent variable in neighboring locations influence the value of the variable in the focal location

ρWY: The spatial lag term. This term captures the weighted average of the dependent variable's values in neighboring locations, reflecting the spatial autocorrelation. It represents the spatial influence of nearby observations on the focal observation.

Spatially Lagged X

$$Y = X\beta + WX\theta + \epsilon$$

WX: The "lag" (average?) of the characteristics of the neighbors

θ: Estimates for the influence of WX. Indirect effects

Spatial Error Model SEM

$$Y = X\beta + u, \quad u = \lambda W u + e$$

On the dependent variable (Y)

On independent variable (X)

On the error term (e)

You can have some combinations

Spatial Durbin Model SDM*

$$Y = \rho WY + X\beta + WX\theta + \epsilon$$

Spatial Durbin Error Model

$$Y = X\beta + WX\theta + u, \quad u = \lambda Wu + e$$

“Manski” Model

$$Y = \rho WY + X\beta + WX\theta + u,$$

$$u = \lambda Wu + e$$

Spatial models

Ordinary Least Square OLS model $Y = X\beta + \epsilon$

Models are compared based on Lagrange multiplier and likelihood ratio tests. Theoretical reasons are also considered.

In the presence of spatial autocorrelation can be expanded in three fundamental ways:

"Outcome (Y variable) of my neighbors affects me"
Spillover effect

"Regressors (X variables) of my neighbors affects me"

"Error of my neighbors affects me"

Lag – Y model / SAR

$$Y = \rho W y + X\beta + \epsilon$$

Lag – X model

$$Y = X\beta + WX\Theta + \epsilon$$

Spatial Error Model SEM

$$Y = X\beta + u,$$
$$u = \lambda W u + \epsilon$$

Spatial Durbin Model SDM

$$Y = \rho W y + X\beta + WX\Theta + \epsilon$$

Spatial Durbin Error Model SDEM

$$Y = X\beta + WX\Theta + u,$$
$$u = \lambda W u + \epsilon$$

Thank you

Orlando Sabogal-Cardona
PhD researcher
University College London UCL