Spatial Data, GIS, and Remote Sensing

Jared Edgerton

Why Space Matters

Many social processes vary across space:

- Economic development
- Conflict and violence
- Urbanization
- Environmental exposure

Spatial structure is not noise.

What Is Spatial Data

Spatial data link observations to location:

- Points
- Lines
- Polygons
- Rasters

Location is part of the measurement.

Vector vs Raster Data

Two core spatial data types:

- Vector data (points, lines, polygons)
- Raster data (gridded surfaces)

Each supports different kinds of questions.

Spatial Joins

Spatial joins combine data using geometry:

- Point-in-polygon
- Distance-based joins
- Overlap and containment

Joins encode spatial assumptions.

Coordinate Reference Systems (CRS)

CRS define:

- How locations map to the Earth
- Units of distance and area
- Projection distortions

CRS mismatches silently break analyses.

Reading Spatial Data (Python)

```
import geopandas as gpd

gdf = gpd.read_file('shapefile.shp')
gdf.head()
```

Reading Spatial Data (R)

```r
library(sf)

gdf ← st_read('shapefile.shp')
head(gdf)
```

Reprojection

Before analysis:

- Ensure common CRS
- Reproject explicitly
- Verify units

```python
gdf = gdf.to_crs(epsg=3857)
```

```r
gdf ← st_transform(gdf, 3857)
```

Mapping Is Not Analysis

Maps are:

- Descriptive
- Persuasive
- Sensitive to design choices

Visualizations can mislead.

Basic Spatial Features

Common spatial features include:

- Distances
- Buffers
- Neighborhood counts
- Exposure measures

Feature choice shapes inference.

Distance Example (Conceptual)

```
gdf['dist'] = gdf.geometry.distance(reference_point)
```

```
gdf$dist ← st_distance(gdf, reference_point)
```

Raster Data and Remote Sensing

Raster data represent:

- Satellite imagery
- Climate surfaces
- Land use and cover

They enable consistent global measurement.

Working with Rasters

```
import rasterio
```

```
library(terra)
```

Raster resolution trades detail for coverage.

Measurement and Bias

Spatial data raise issues of:

- Modifiable areal unit problem (MAUP)
- Boundary definitions
- Cloud cover and missingness
- Strategic behavior

Space interacts with incentives.

Scalability Considerations

Spatial analysis can be expensive:

- Geometry operations are slow
- Rasters are large
- Indexing matters

Efficiency is part of design.

Documentation and Reproducibility

Spatial workflows should record:

- Data sources
- CRS choices
- Join logic
- Resolution and aggregation

Spatial results are sensitive to preprocessing.

What We Emphasize in Practice

- Treat location as measurement
- Inspect geometry explicitly
- Be explicit about CRS
- Validate spatial assumptions

Discussion

- Where does spatial aggregation matter most?
- Which CRS choices feel consequential?
- How does space interact with incentives?