API-231 / GIS-PubPol
Meeting 14 (Remote Sensing and Satellite Data)

Yuri M. Zhukov
Visiting Associate Professor of Public Policy
Harvard Kennedy School

March 19, 2024

Remote sensing and public policy

Definitions

What is remote sensing?

Information obtained through long-range observation (e.g. from satellites, aircraft)

1. Collection of raw imagery from the surface of the Earth
   a. *passive sensors*: collect information on emitted light/radiation
      - examples: photography, infrared
   b. *active sensors*: emit energy, collect information on reflected light/radiation
      - examples: radar, LiDAR
2. Image processing
   a. raw images are *georeferenced* to ground control points
   b. emitted/reflected light matched to *specific spectral signatures*
      - examples: types of vegetation, land cover, CO_2 emissions
   c. processed data are stored as pixels in raster datasets

Advantages:

- remote sensing is sometimes cheaper and safer than direct observation (e.g. hard-to-reach areas, conflict zones)
- measurement is consistent across regional, national borders

Remote sensing = raster data

- not all raster data are derived from remote sensing imagery
- but all remote sensing imagery originates as raster data

Raster data structure ≠ vector data structure

- vectors store information in "attribute tables" (N features × K fields)
- rasters store information in a grid of pixels (N_R rows × N_C columns)
  - pixels are of constant size, shape, area
  - each pixel represents a unique location
  - each pixel contains just one value (e.g. precipitation, land use)
  - size of pixels determines resolution (e.g. 1 meter, 1 km, 1 degree)
- rasters usually have larger file size than vectors, but not necessarily more precision

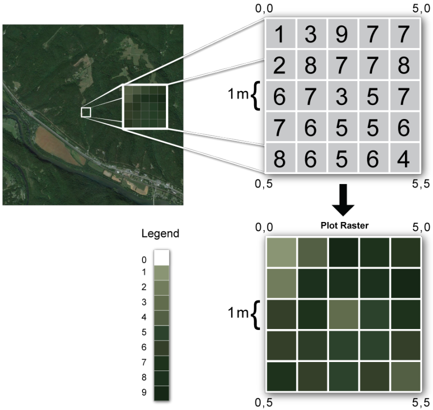

Figure 1: Raster data structure

Applications

Many **variables of interest** to public policy originate as remote sensing imagery

- weather (precipitation, temperature)
- climate model forecasts
- flooding depth and risk
- active fires
- night light emissions
- elevation, slope, line of sight
- pollution and air quality
- cloud cover
- vegetation indices
- soil quality, fertility
- land use and land cover (LULC)
- built-up areas
- population density (derived from above)

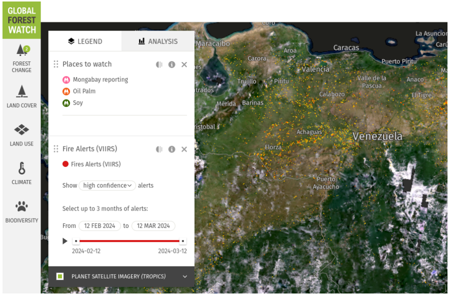

Figure 2: Example use case

But raster data were not (originally) built for social science and public policy applications

- original policy purpose: military reconnaissance, damage assessment
- original scientific purpose: natural sciences (e.g. geology, ecology, biology)
- no sensor systems, spectral measurements were designed for dedicated monitoring of social, economic processes
- reliance on indirect/proxy measures

Divergent data structures, approaches

- social science: "vector view" of world (e.g. organize data into discrete units)
- natural science: "raster view" of world (e.g. organize data into regular lattice)
- integrating raster and vector data requires *interdisciplinary* cooperation

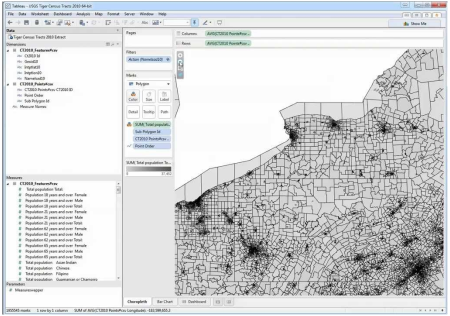

Figure 3: Social science prefers vectors

Raster data analysis

Rasterization and Vectorization

In social science and public policy, raster data integration requires that we either

1. Rasterize the vector data
   - convert discrete features into continuous field
   - examples:
     a. frequency/density of features
     b. presence/absence of feature
     c. distance to features
     d. assignment to feature
2. Vectorize the raster data
   - summarize values of continuous field at each feature
   - examples:
     a. zonal statistics (e.g. mean, max cell values)
     b. image tracing (e.g. of georeferenced maps – we covered this earlier)

Point-to-raster: suppose points are locations of 100 events (e.g. wolf attacks)

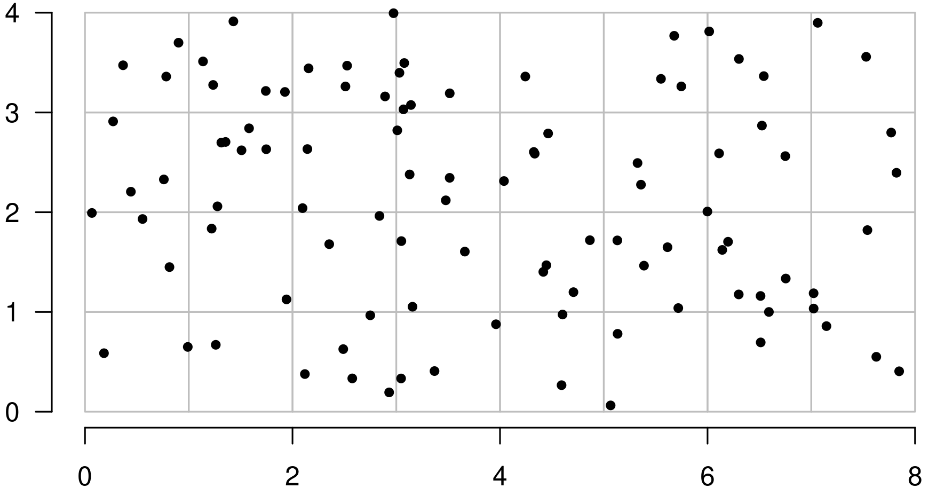

Figure 4: Point geometries

Option 1: count number of features in each raster pixel/cell

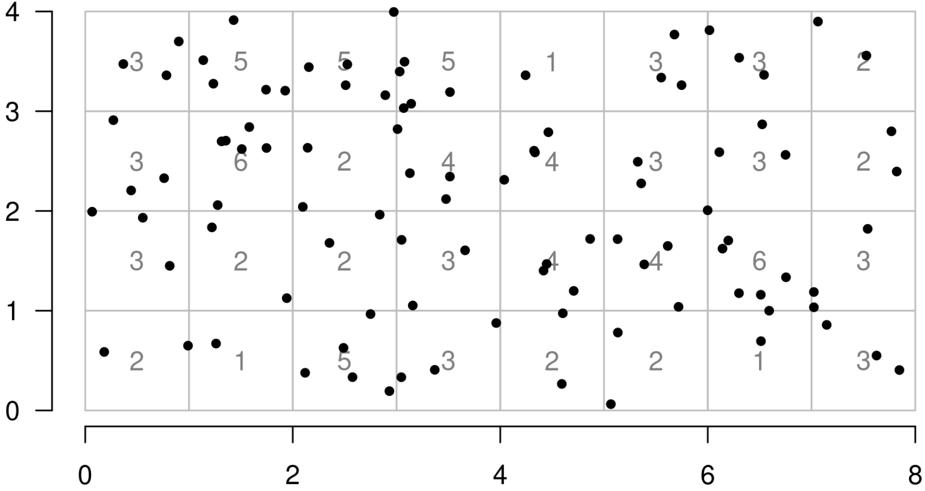

Figure 5: Point counts per cell

Pixels values are local frequency (number of points) or point density (number/area)

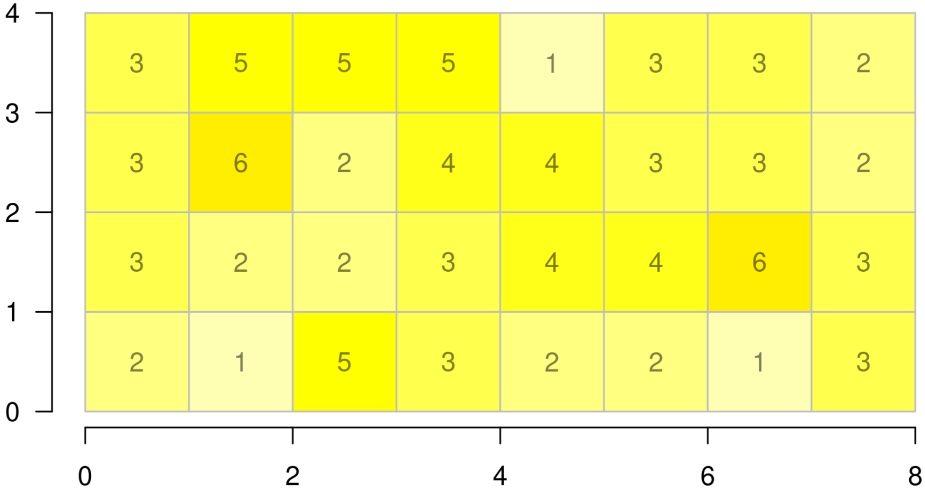

Figure 6: Local point frequency

Line-to-raster: suppose this line is an infrastructural object (e.g. road, power line)

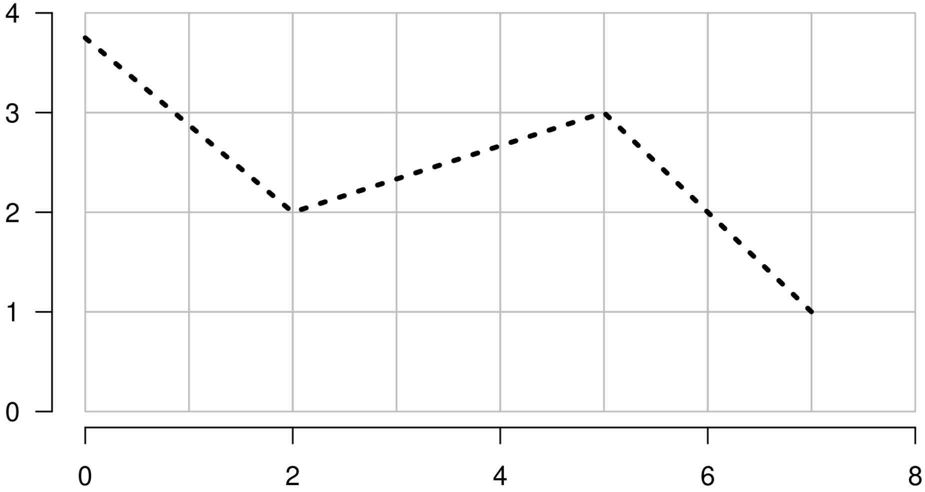

Figure 7: Line geometries

Option 2: presence/absence of features at each raster pixel/cell

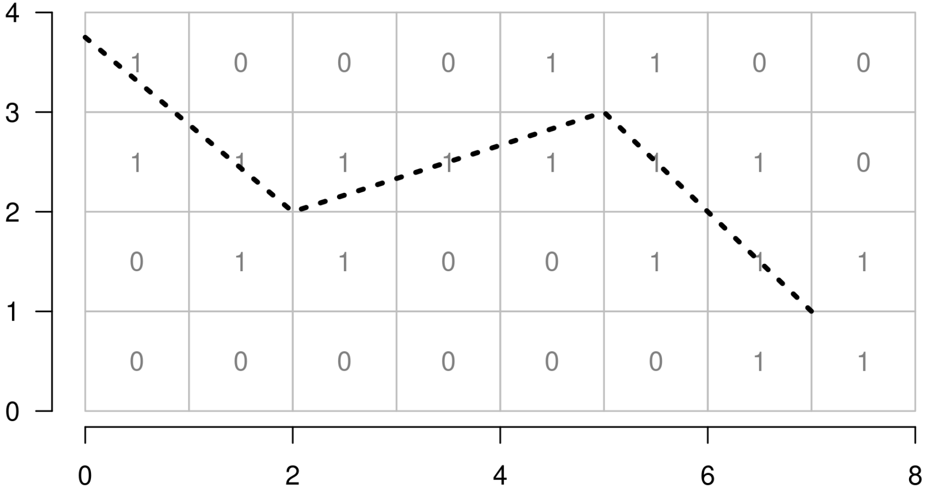

Figure 8: Line presence/absence per cell

Pixels values are indicators of whether an object is locally present/accessible

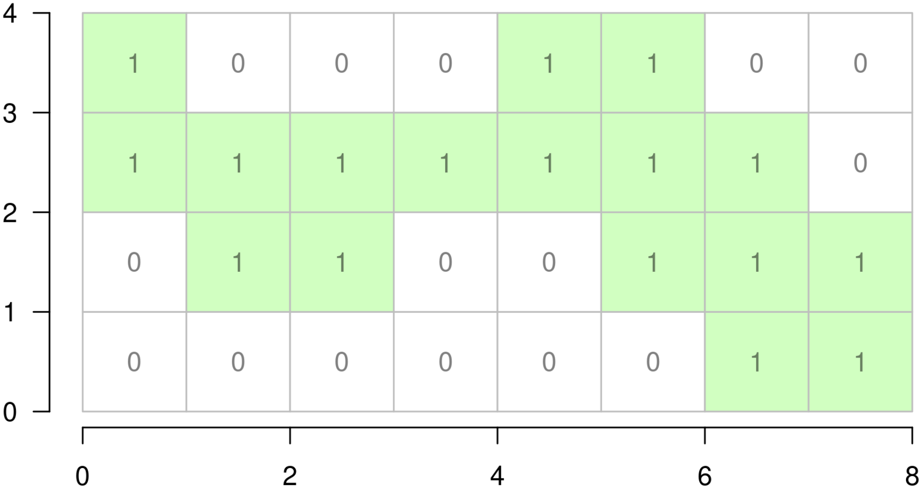

Figure 9: Local line access

Option 3: distance from feature to each raster pixel/cell

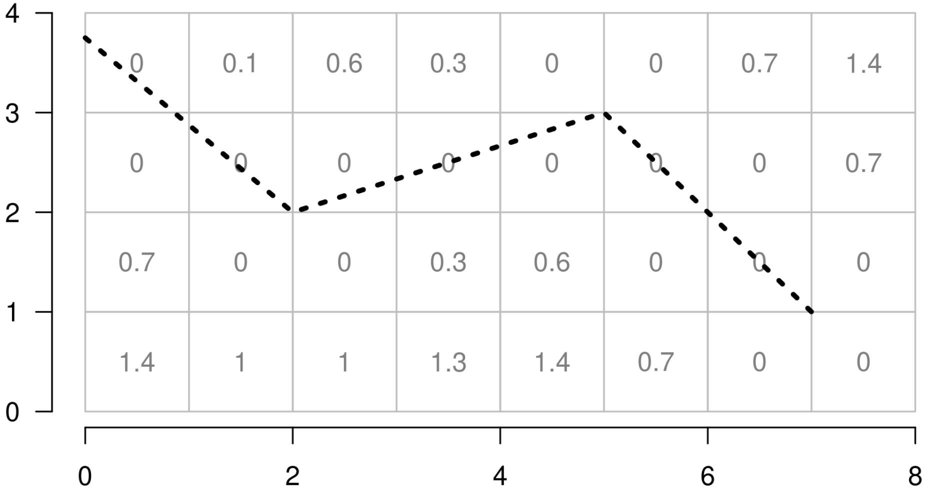

Figure 10: Distance from line to cell

Pixels values represent proximity

Figure 11: Local distance

Polygon-to-raster: suppose polygons are 4 administrative areas (e.g. districts)

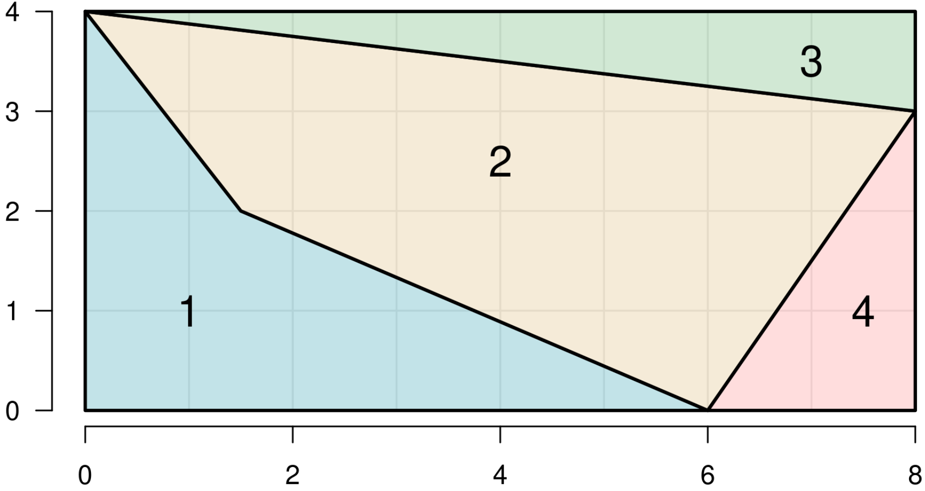

Figure 12: Polygon geometries

Option 4: assign pixels to overlapping features (e.g. by center of cell)

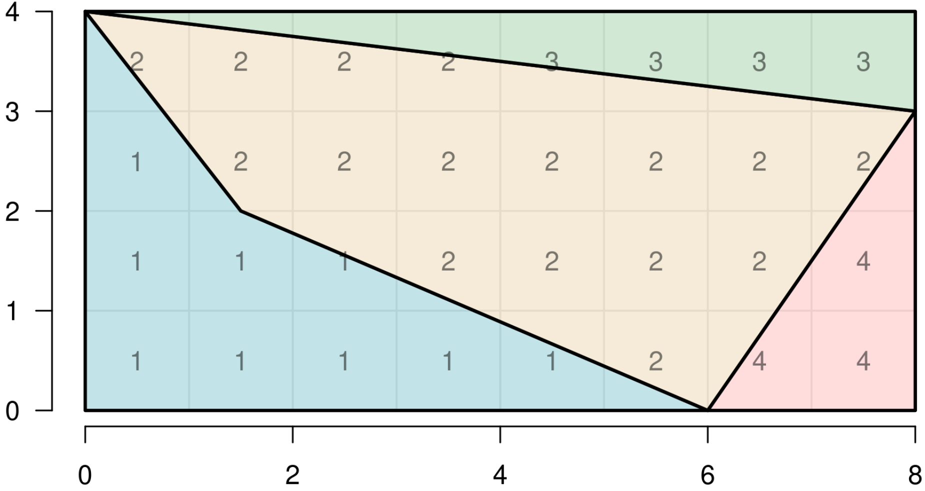

Figure 13: Polygon assignment by centroid

Pixel values are polygon labels or attributes (e.g. assumed constant)

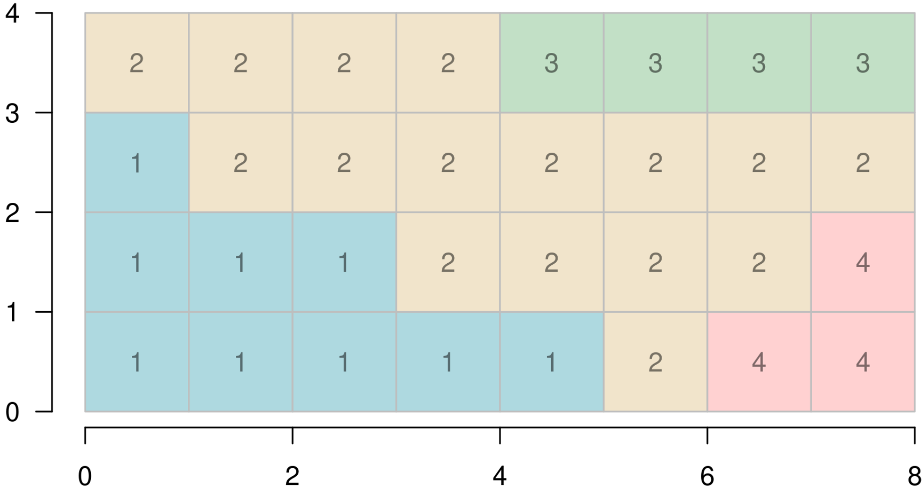

Figure 14: Local polygon assignment

Rasterization overview

These operations can be done on *all types of vector data*

1. count/density of points/lines/polygons
2. presence/absence of points/lines/polygons
3. distance to points/lines/polygons
4. assignment to points/lines/polygons (with tie-breaking rule)

But problem: why do this?

- pixels are not meaningful spatial units for public policy
- policymakers don't think of the world as a "continuous field"
- policy is made in discrete geographic jurisdictions, with well-defined borders
- more common approach to analysis: *convert raster to vector*

Raster-to-polygon: suppose raster represents a continuous variable (e.g. elevation)

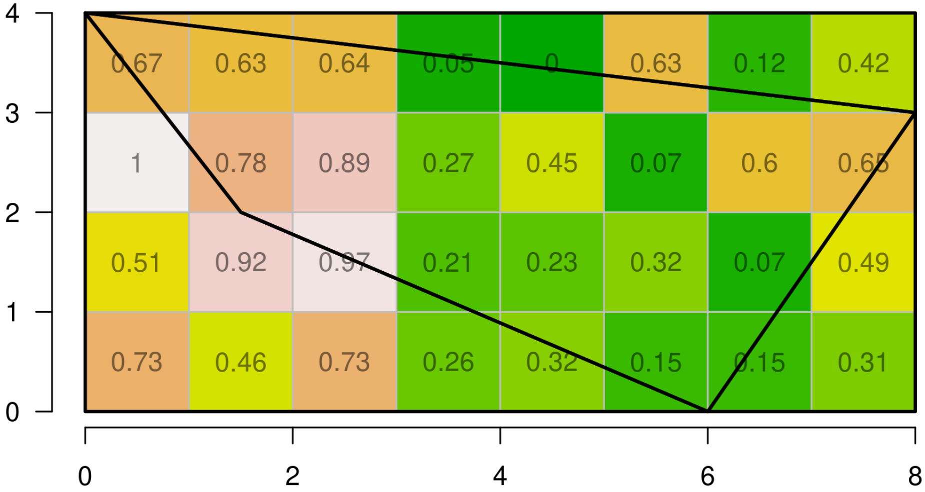

Figure 15: Raster cell values

Option 1: calculate zonal statistics (e.g. average cell values) for each polygon

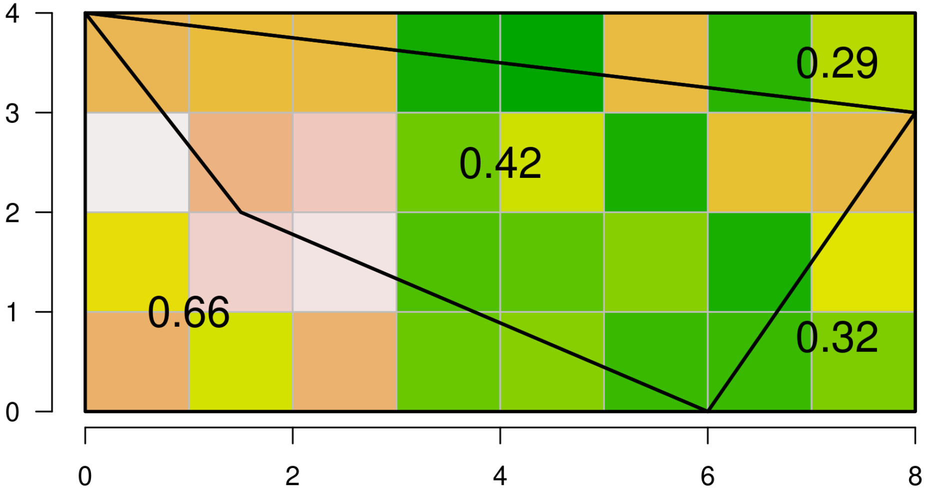

Figure 16: Zonal statistics: mean cell values

Average cell values are added to attribute table for polygons

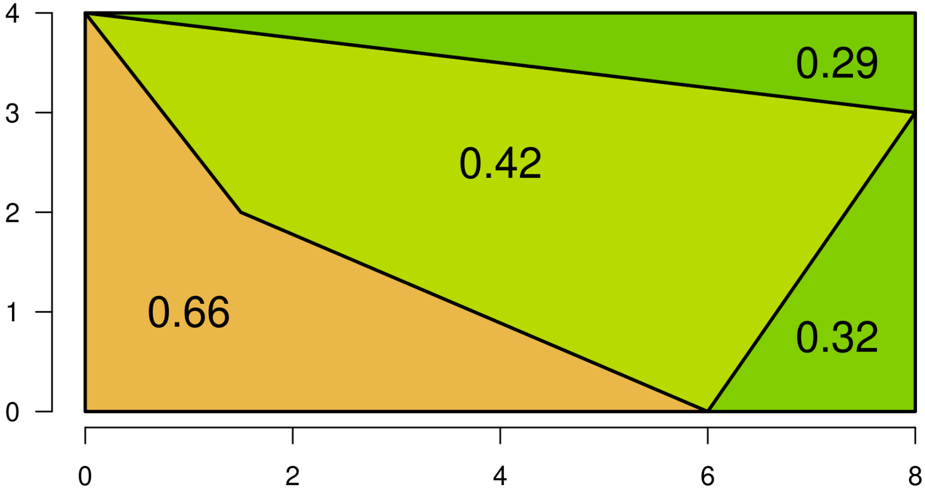

Figure 17: Mean cell values for each polygon

Same operation could be used to obtain maximum cell values...

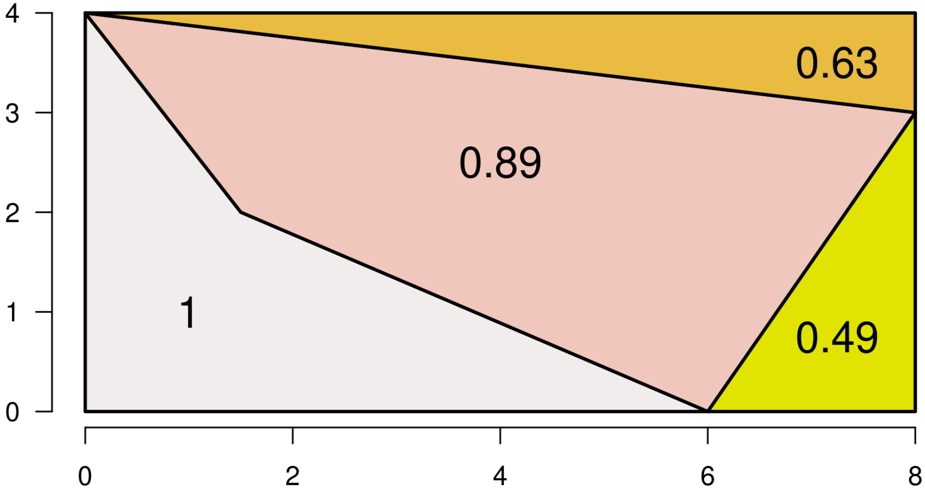

Figure 18: Maximum cell values for each polygon

. . . or minimum values (or any other summary statistic)

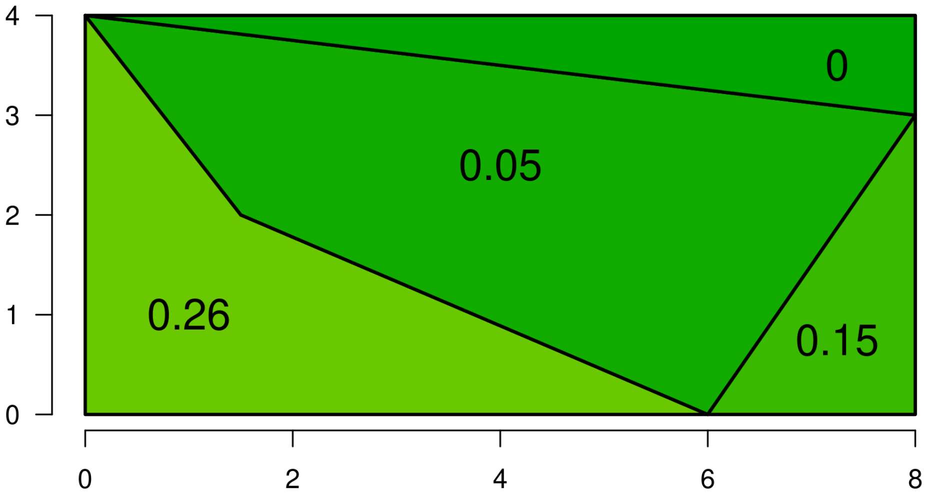

Figure 19: Minimum cell values for each polygon

But what if raster represents a categorical variable (e.g. land use)?

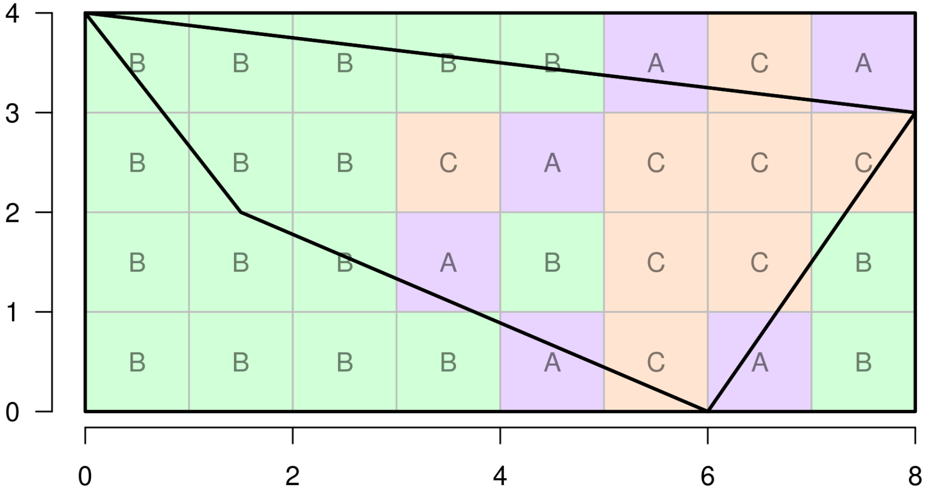

Figure 20: Raster cell values

Option 2: reclassify raster to binary (e.g. 1 if land use "A", 0 otherwise)

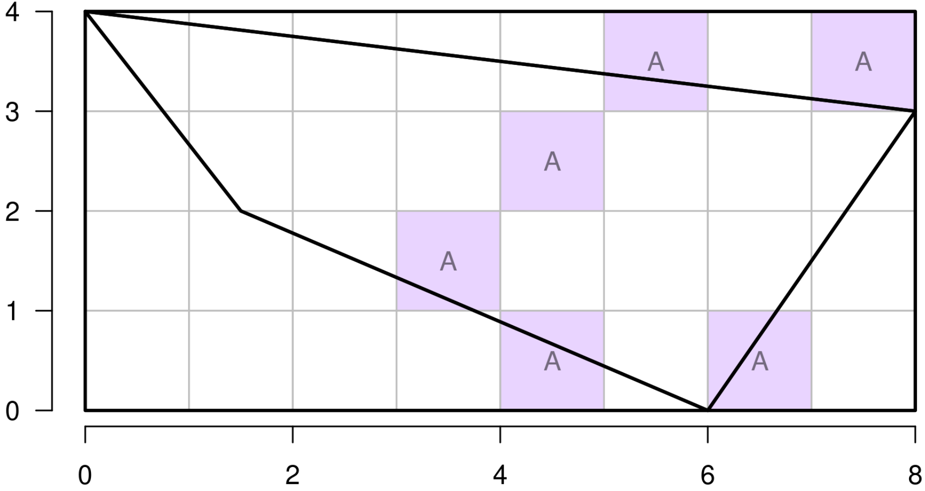

Figure 21: Reclassified raster

Calculate zonal statistics: percent of each polygon with cell values of "A"

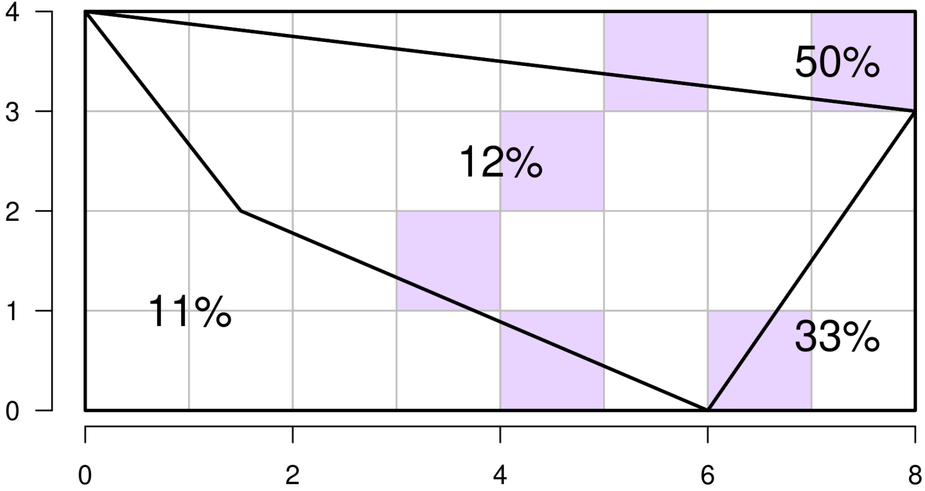

Figure 22: Zonal statistics: value "A" as percent of overlapping cells

Percentages are added to attribute table for polygons

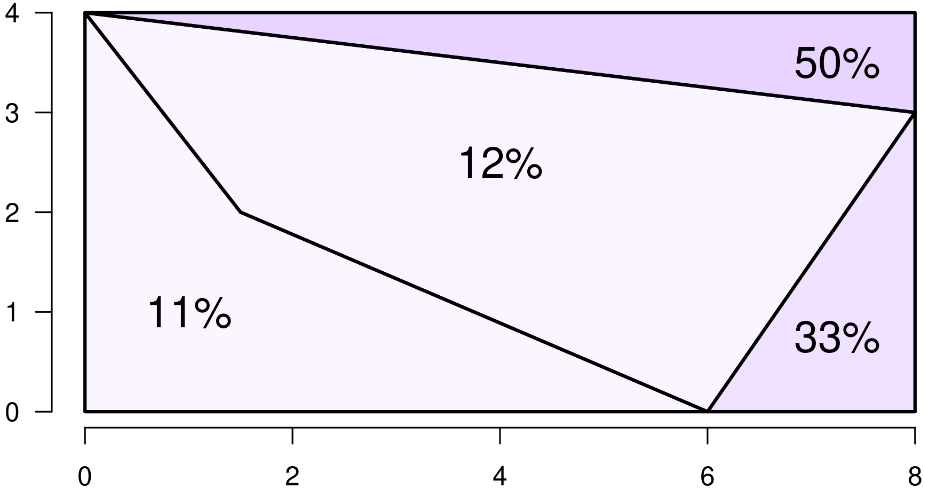

Figure 23: Percent “A” per polygon

Scale-dependence

Scale-Pattern-Process

1. Scale of analysis (spatial, temporal) impacts which patterns are observable
    - these observations shape inferences we draw about underlying social processes
2. Processes drive patterns whose observation is scale-dependent
    - some research questions require high spatial resolution:
        a. urban/neighborhood policy
        b. bomb damage assessment
    - some research questions require high temporal resolution:
        a. emergency response
        b. weather forecasting
    - some questions can be answered at low resolution (e.g. long-term, large-scale)
        a. economic development
        b. deforestation, changes in land use

Trade-offs

- lower resolution (large pixels) = more information loss
- higher resolution (small pixels) = higher collection, storage, computation costs

How scale impacts rasterization and vectorization

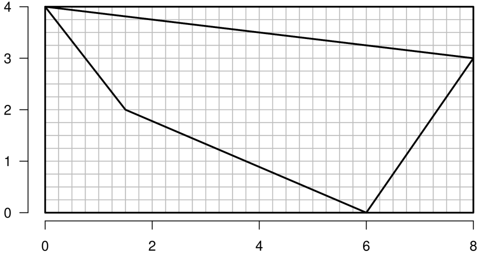

Figure 24: High resolution (small pixels)

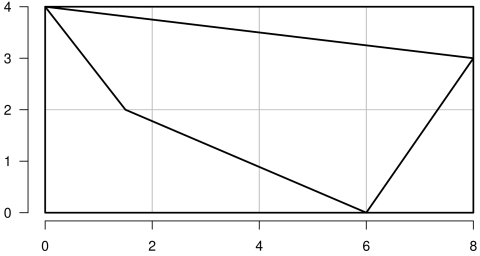

Figure 25: Low resolution (large pixels)

Point-to-raster: same underlying point pattern, two very different rasters

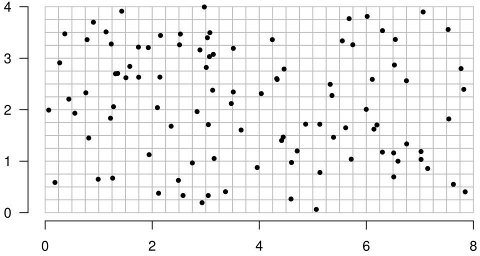

Figure 26: High resolution (small pixels)

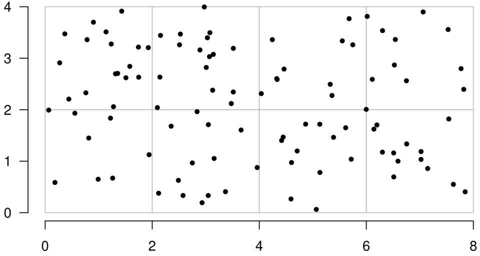

Figure 27: Low resolution (large pixels)

Counts, densities will appear sparser (more intense) in high-(low-)resolution rasters

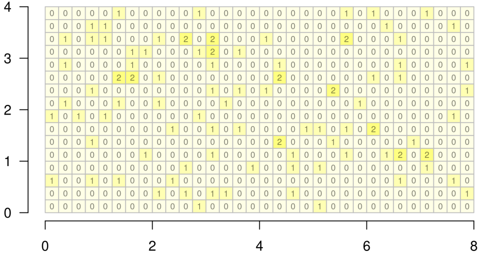

Figure 28: High resolution (small pixels)

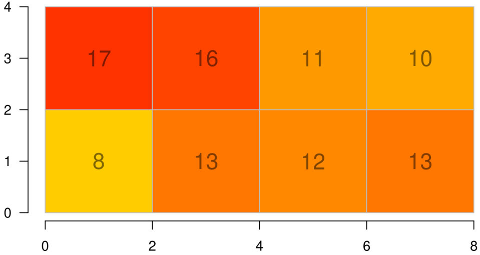

Figure 29: Low resolution (large pixels)

Line-to-raster: same line features, two very different rasters

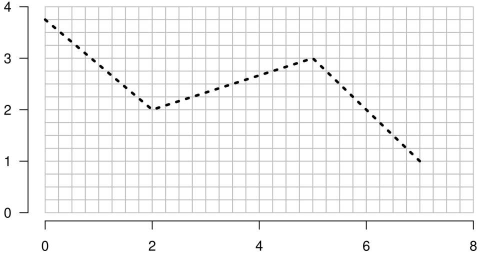

Figure 30: High resolution (small pixels)

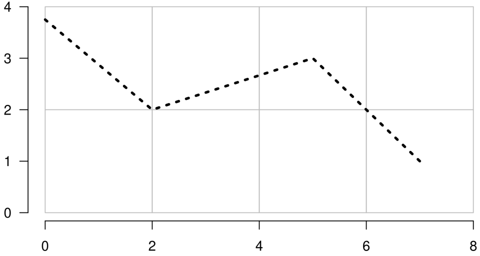

Figure 31: Low resolution (large pixels)

Absence/presence measures are more (less) variable in high-(low-)resolution rasters

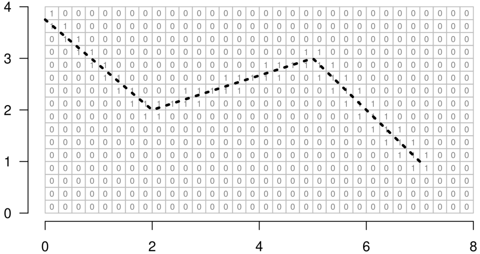

Figure 32: High resolution (small pixels)

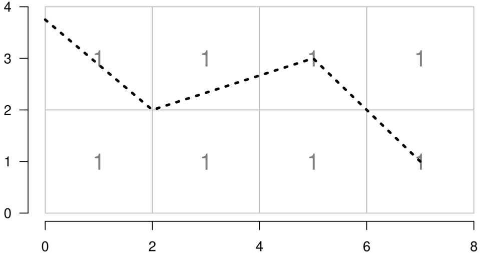

Figure 33: Low resolution (large pixels)

High-(low-)resolution rasters more (less) precisely reflect shape of vector geometries

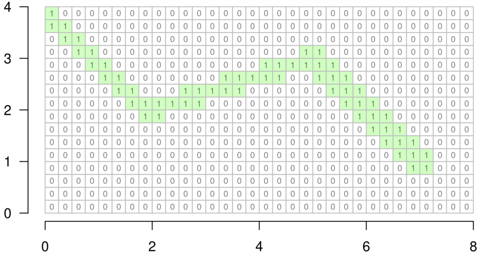

Figure 34: High resolution (small pixels)

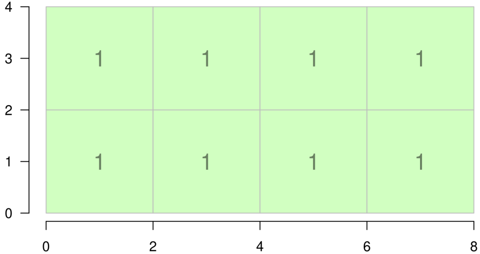

Figure 35: Low resolution (large pixels)

Distance measures also have more (less) variation in high-(low-)resolution rasters

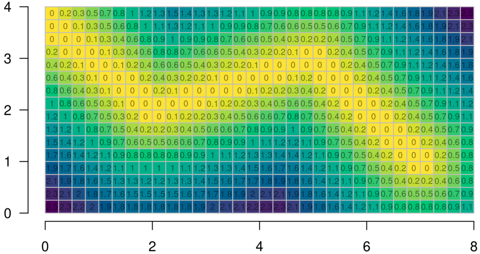

Figure 36: High resolution (small pixels)

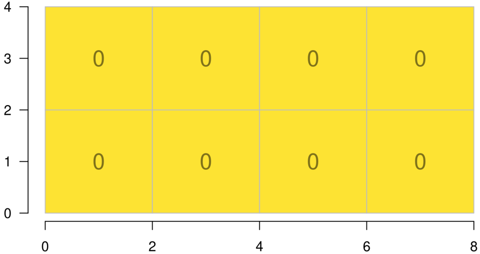

Figure 37: Low resolution (large pixels)

Polygon-to-raster: same polygon features, two very different rasters

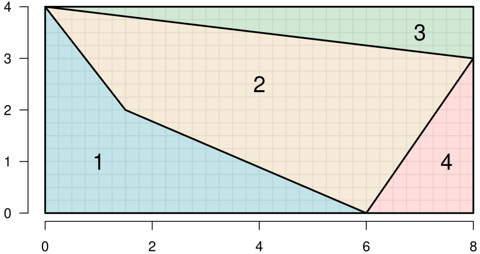

Figure 38: High resolution (small pixels)

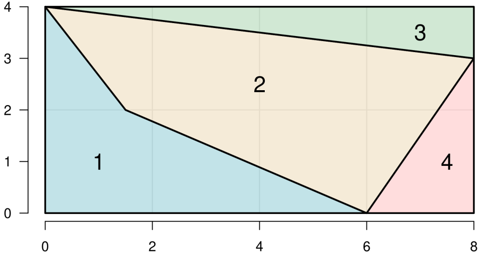

Figure 39: Low resolution (large pixels)

Assignment operations are more (less) coarse in low-(high-)resolution rasters

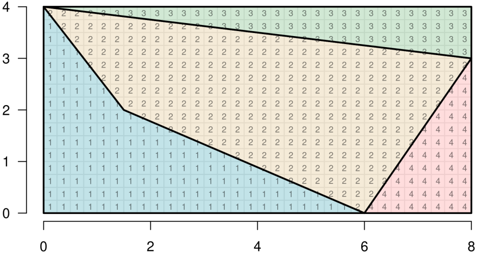

Figure 40: High resolution (small pixels)

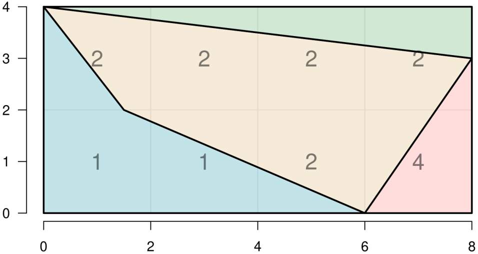

Figure 41: Low resolution (large pixels)

Some polygon features may disappear entirely in low-resolution rasters

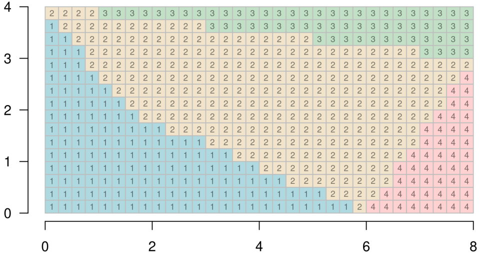

Figure 42: High resolution (small pixels)

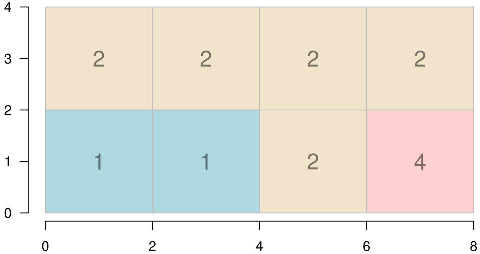

Figure 43: Low resolution (large pixels)

Raster-to-polygon: suppose we have two rasters with same underlying data

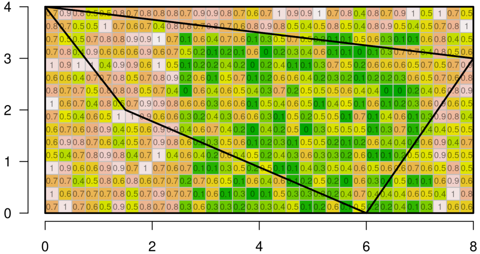

Figure 44: High resolution (small pixels)

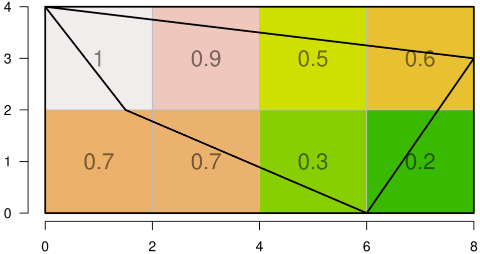

Figure 45: Low resolution (large pixels)

Zonal statistics on the high-resolution raster will be more precise

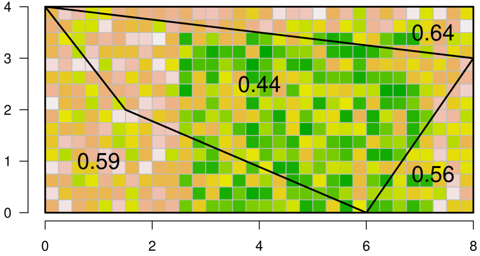

Figure 46: High resolution (small pixels)

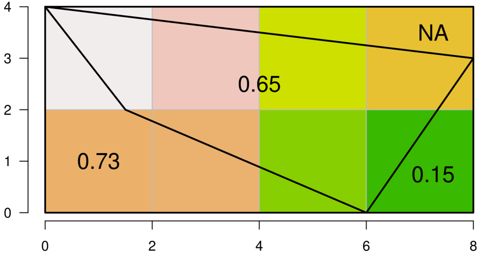

Figure 47: Low resolution (large pixels)

Low-resolution raster is more likely to generate missing values in polygon features

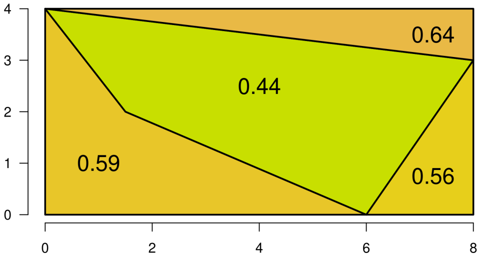

Figure 48: High resolution (small pixels)

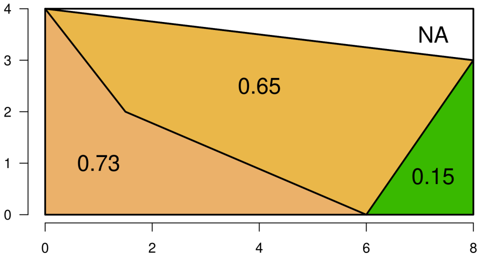

Figure 49: Low resolution (large pixels)

Similar problems arise with zonal statistics on rasters with categorical variables

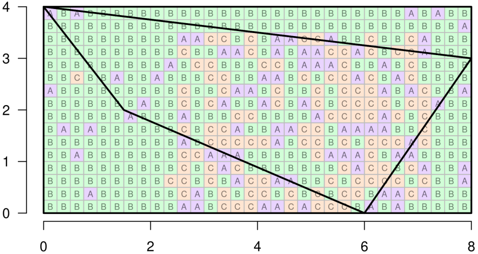

Figure 50: High resolution (small pixels)

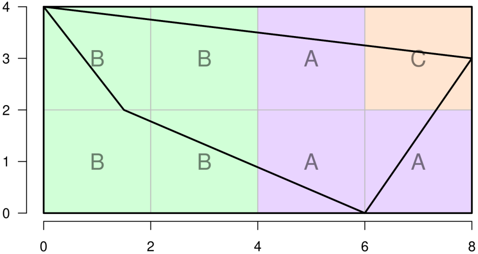

Figure 51: Low resolution (large pixels)

Lower resolution → fewer raster cells to calculate statistics over, less precision

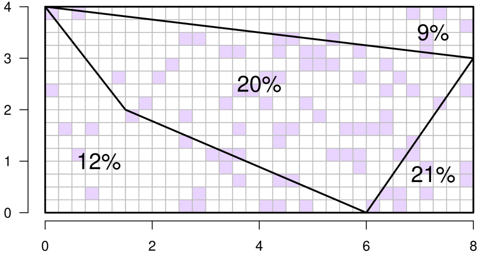

Figure 52: High resolution (small pixels)

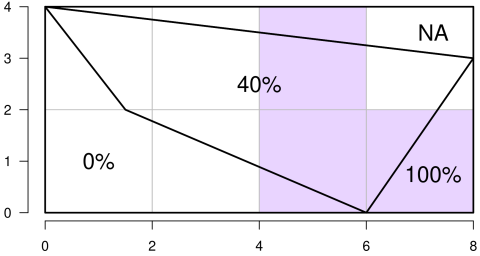

Figure 53: Low resolution (large pixels)

Low resolution rasters may sometimes also exaggerate amount of variation

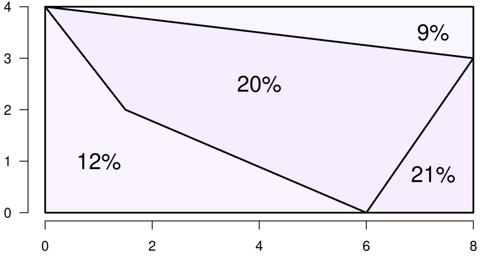

Figure 54: High resolution (small pixels)

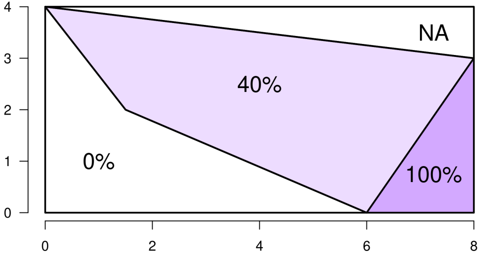

Figure 55: Low resolution (large pixels)

Why not always use highest-resolution raster data?

- high-res data may not exist (due to orbital requirements, low user demand)
- high-res satellite data are sometimes inaccessible (classified, proprietary)
- high-res data are expensive to collect, transmit, store (terabytes, petabytes)
- high-res data take up *a lot* of memory, need high-performance computing
- high-res data may not be needed to answer research question
  (don't need 1-meter resolution to study regional, national, global phenomena)

Figure 56: Which scale is right for me?