Wrangling Spatial Data

GEOG 215 - March 25, 2020

Today's Agenda

- Spatial Data Operations
  - Vector Data
    - Spatial Subsetting
    - Topological Operations
    - Spatial Join
  - Raster Data
    - Map Algebra
    - Local Operations
    - Focal Operations
    - Zonal Operations
- Raster-vector interactions

Announcements

YOU ARE INVITED TO
Geography's Pre-registration & More Virtual Party!
Friday, March 27
3:30-5:00 PM

We can't have our usual pizza & registration event, but we can still hang out.

Come chat with Geography instructors and students & pre-register for Fall 2020 courses!

ZOOM LINK:

https://unc.zoom.us/j/863601568

Also, updates on Hurston Lounge!

Recommendation

- Fully available online at https://geocompr.robinlovelace.net/
- Up-to date using latest R packages and combines both theory, applications and R code really well
- Great resource for your practice and your final projects - especially for reading, writing, wrangling vector or raster data
- Your lab 4 is based on exercises from the book
- This lecture too, borrows heavily from chapter 4-5 of the book.

Geocomputation with R

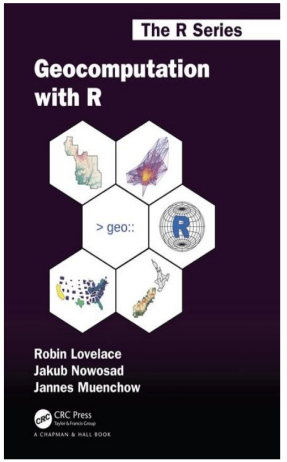
Robin Lovelace, Jakub Nowosad, Jannes Muenchow

2020-02-16

Welcome

This is the online home of *Geocomputation with R*, a book on geographic data analysis, visualization and modeling.

Note: This book has now been published by CRC Press in the R Series. You can buy the book from CRC Press, Wordery, or Amazon.



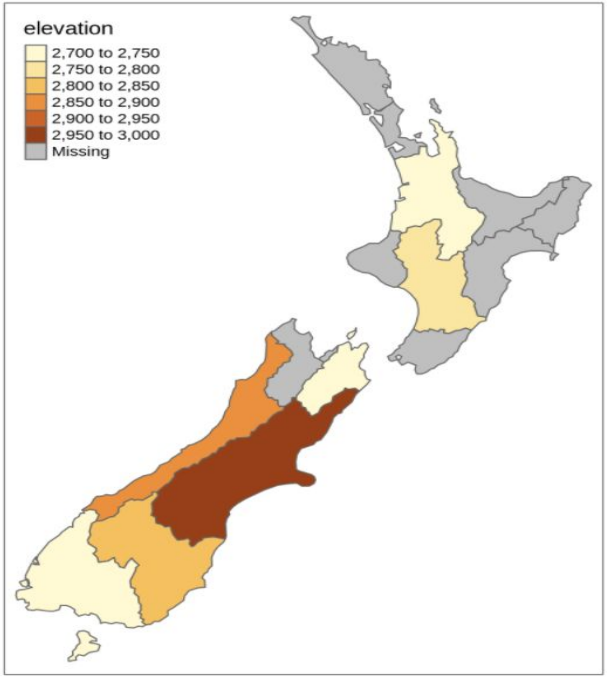
Inspired by **bookdown** and the Free and Open Source Software for Geospatial (FOSS4G) movement, this book is open source. This ensures its contents are reproducible and publicly accessible for people worldwide.

The online version of the book is hosted at geocompr.robinlovelace.net and kept up-to-date by Travis, which provides information on its 'build status' as follows:

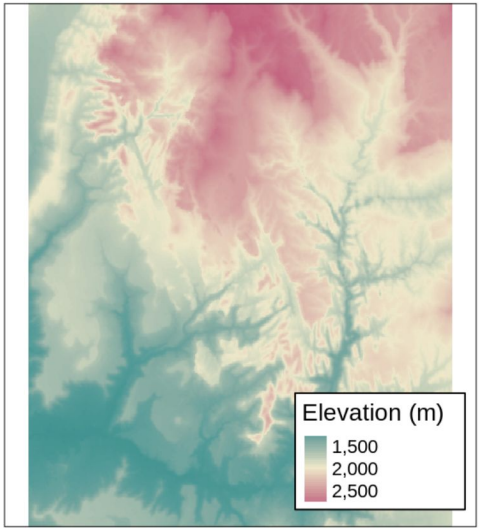
build error

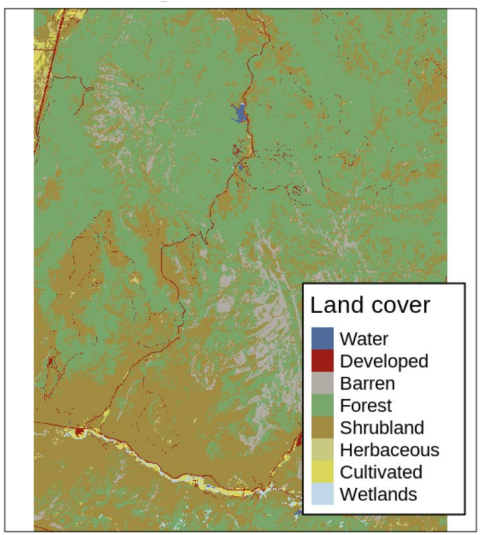
How to contribute?

Recap



VECTOR





RASTER

Recap

```
> print(australia_provinces[1:3])
Simple feature collection with 11 features and 3 fields
geometry type:  MULTIPOLYGON
dimension:      XY
bbox:           xmin: 112.9194 ymin: -54.75042 xmax: 159.1065 ymax: -9.240167
epsg (SRID):    4326
proj4string:    +proj=longlat +datum=WGS84 +no_defs
```

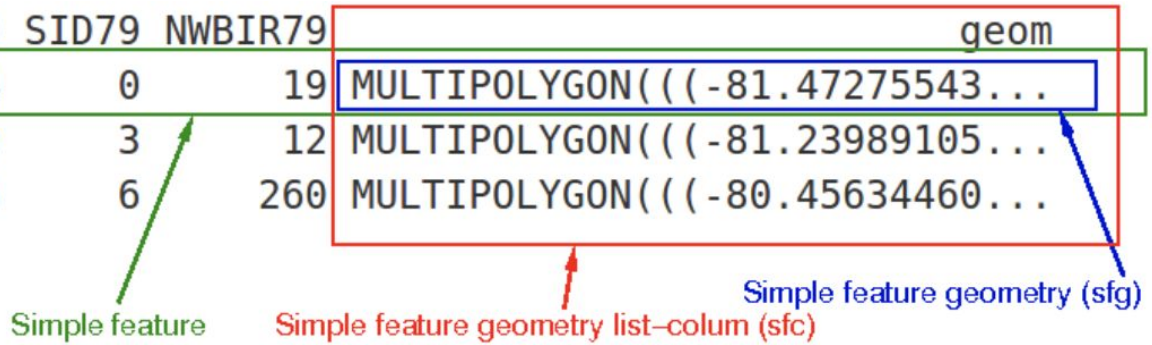
VECTOR

```
## dimensions  : 5060, 4299, 21752940  (nrow, ncol, ncell)
## resolution  : 1, 1  (x, y)
## extent      : 254570, 258869, 4107302, 4112362  (xmin, xmax, ymin,
ymax)
## coord. ref. : +proj=utm +zone=11 +datum=WGS84 +units=m +no_defs
+ellps=WGS84 +towgs84=0,0,0
```

RASTER

Recap

	BIR74	SID74	NWBIR74	BIR79	SID79	NWBIR79	geom
1	1091	1	10	1364	0	19	MULTIPOLYGON(((-81.47275543...
2	487	0	10	542	3	12	MULTIPOLYGON(((-81.23989105...
3	3188	5	208	3616	6	260	MULTIPOLYGON(((-80.45634460...



- A vector file in the sf package has both a *dataframe (or a tibble)* and an *sf* class.
- You can perform all data operations as you did for other non-spatial dataframes.
- These operations that do not need spatial data *(geom)* column are called *attribute* operations

Attribute Data Operations

- **You can perform all attribute data operations such as filter, select, mutate, arrange, joins on sf objects just as you would for data frames**
    - However, aggregating (group_by, summarize) is different: why?
        - Different spatial scales
        - Need a combination of aggregation and geometry operations (discussed later)
- Raster data is essentially a matrix - you can perform most matrix operations on a raster dataset
- Read Chapter 3: https://geocompr.robinlovelace.net/attr.html
- Raster data attribute operations discussed in lab 4

Spatial Operations on Vector Data

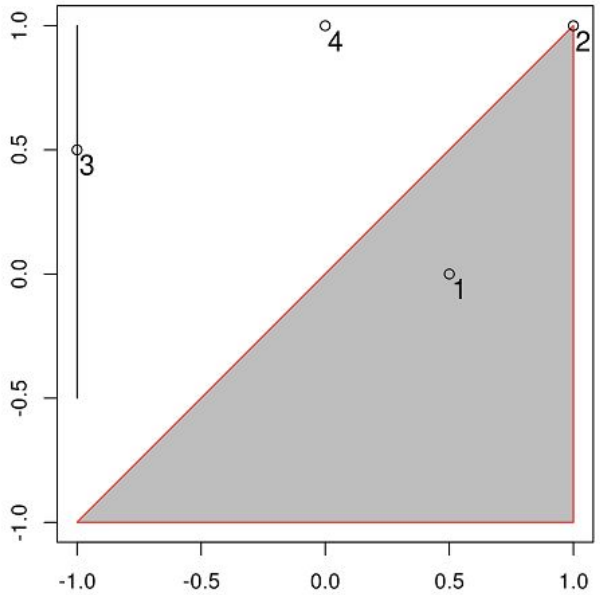
Spatial Sub-setting

Sub-setting (filtering) features of a *target* vector file

- Using its own attributes - filter() or traditional data frame subsetting commands
- Using spatial features of a *source* spatial file
  - `x [y , ]`
  - Here, you subset features of a *target* `x` using contents of a *source* `y`
  - Both files should be an sf object
  - By default, subsetting is done based on ***intersection,*** however, other topological operations are possible too (next slide)
- Read Chapter 4.2.1

Topological Operations

- Workhorse of the sf library
- Topology = spatial relationships
- Topological describe the “spatial relationships” between object.
    - By using spatial *predicates* -
        - Unary - needs 1 input
            - Is_longlat, is_empty, is_valid
        - Binary - needs 2 input
            - Intersects, disjoint, within, touches, is_within_distance, overlaps, crosses
- Good explanation in assigned datacamp exercise: **Conducting spatial analysis with the sf and raster packages**
- https://keen-swartz-3146c4.netlify.com/geommanip.html (Section 5.1)



Topological Operations

The binary predicates provided by package `sf` are

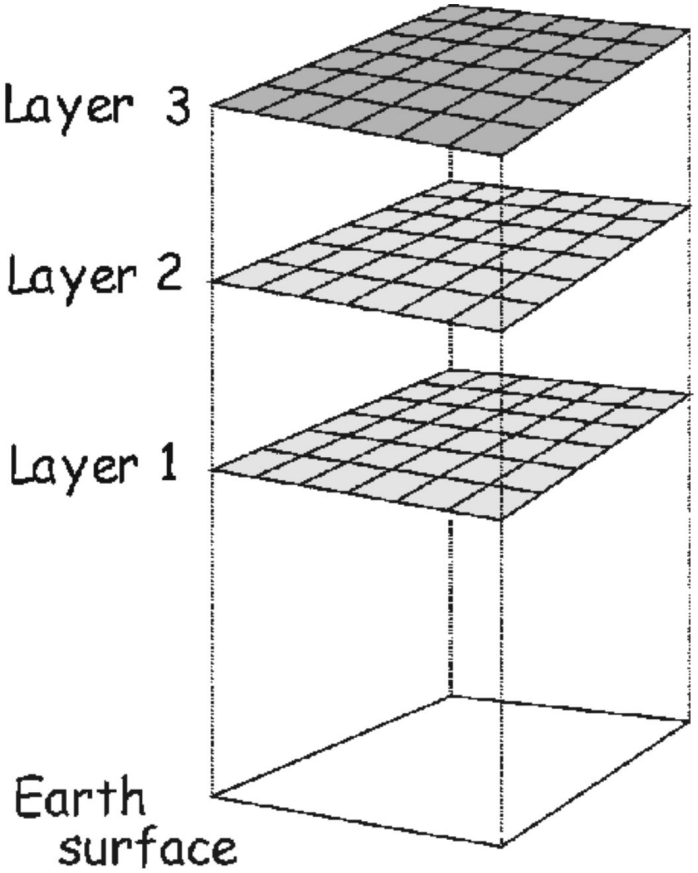
predicate	value	inverse of
`st_contains`	None of the points of A are outside B	`st_within`
`st_contains_properly`	A contains B and B has no points in common with the boundary of A	
`st_covers`	No points of B lie in the exterior of A	`st_covered_by`
`st_covered_by`	inverse of `st_covers`	
`st_crosses`	A and B have some but not all interior points in common	
`st_disjoint`	A and B have no points in common	`st_intersects`
`st_equals`	A and B are geometrically equal; node order number of nodes may differ; identical to A contains B AND A within B	
`st_equals_exact`	A and B are geometrically equal, and have identical node order	
`st_intersects`	A and B are not disjoint	`st_disjoint`
`st_is_within_distance`	A is closer to B than a given distance	
`st_within`	None of the points of B are outside A	`st_contains`
`st_touches`	A and B have at least one boundary point in common, but no interior points	
`st_overlaps`	A and B have some points in common; the dimension of these is identical to that of A and B	
`st_relate`	given a pattern, returns whether A and B adhere to this pattern	

Start thinking about applications of these to your final projects

Spatial Join

Each layer represents a single “theme”

- Spatial join leverages the “location” information in each layer



https://geocompr.robinlovelace.net/spatial-operations.html (4.2.3)

Spatial Join

Spatial Join is a spatial overlay operation that transfers attributes from one spatial data layer to another based on overlapping location

- How attributes are transferred is dependent on..
  - The two types of input features (points, lines, polygons)
  - The selected approach of combining attributes from multiple features
  - Selected approach for overlapping
- Read about non-overlapping joins - very useful
  - https://geocompr.robinlovelace.net/spatial-operations.html (Section 4.2.4)

Spatial Data Aggregation

A way of condensing or grouping data

- Moving information across multiple spatial scales
    - Eg. Grouping point data at polygon level
        - Taking individual crime locations in each county to estimate crimes per capita at the the county level

Distance Relations

Unlike topological relations (True or False), distance relations are continuous

- Ability to return distance matrices
    - Distances between all combinations of x and y sf objects
- Can you think of examples where distance matrices might be useful?

Spatial Operations on Raster Data

Spatial Sub-setting

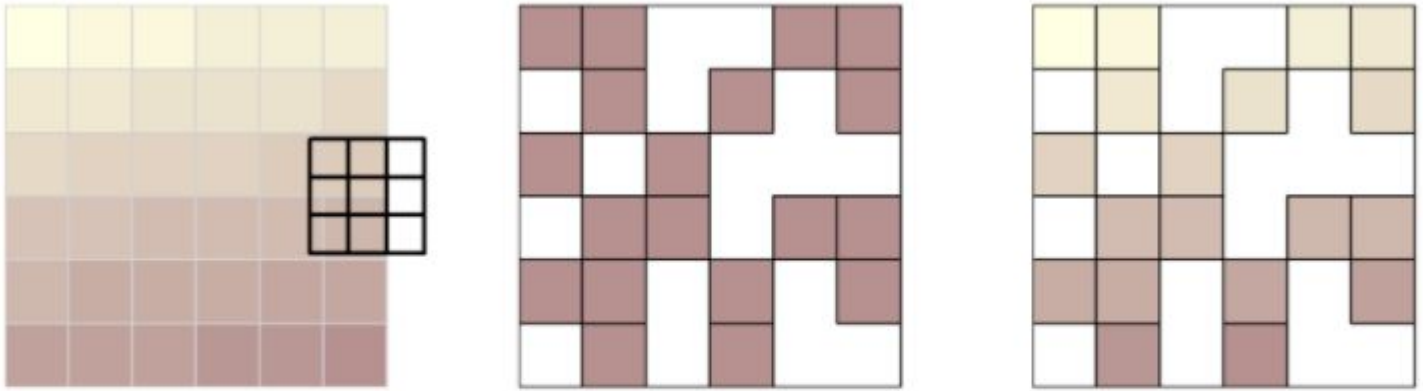


FIGURE 4.7: Subsetting raster values with the help of another raster (left). Raster mask (middle). Output of masking a raster (right).

Map Algebra

Workhorse of raster objects

- Helps perform calculations on raster(s) efficiently and extremely fast
  - Takes advantage of the matrix data structure of raster
- Mainly 4 types of operations
  - Local
    - Per-cell operations
  - Focal
    - Neighborhood operations
  - Zonal
    - Similar to focal but the grid has irregular sizes and shapes
  - Global
    - Per-raster operations, cell may derive values from *one or more entire rasters*

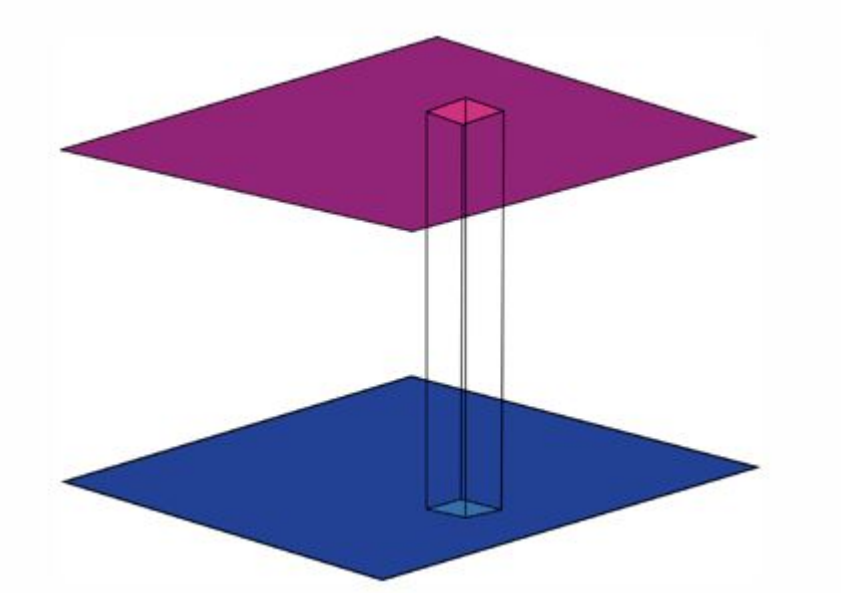
https://desktop.arcgis.com/en/arcmap/latest/extensions/spatial-analyst/performing-analysis/the-types-of-operations-in-spatial-analyst.htm

Local Operations

Allows you to perform cell-by-cell operations in one or multiple raster layers

- Perform any arithmetic operations
  - Addition, subtraction, multiplication, division, and many more
    - Eg: adding rasters with male population and female population per pixel to get total population
  - Average, sum, max, min etc.
    - Calculating average monthly temperature for each cell from a raster of daily temperature
  - Reclassification
    - Classifying rasters with % of trees to different forest classifications

Local Operations

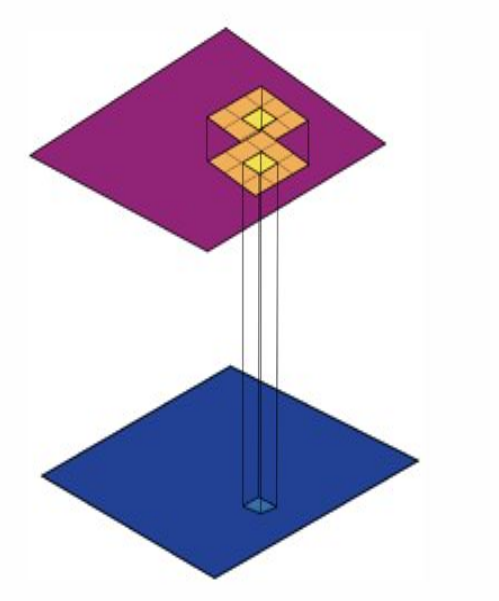


Focal Operations

Computes an output raster where the value for each output cell is a function of the values of all the input cells that are in a specified neighborhood around that location.

- The function performed on the input is a statistic, such as the maximum, average, or sum of all values encountered in that neighborhood.
- Neighborhood (is typically 3 by 3) can be of any shape (not necessarily rectangular) or size as defined by the user
- Also referred to as *spatial filtering* and *convolution*
- Widely used in image-processing
    - Smoothing uncertain estimates
    - Creating terrain maps (ironing out small variations in slopes, elevation etc)

Focal Operations



0	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36

NA	NA	NA	NA	NA	NA
NA	0	2	3	4	NA
NA	7	8	9	10	NA
NA	13	14	15	16	NA
NA	19	20	21	22	NA
NA	NA	NA	NA	NA	NA

Zonal Operations

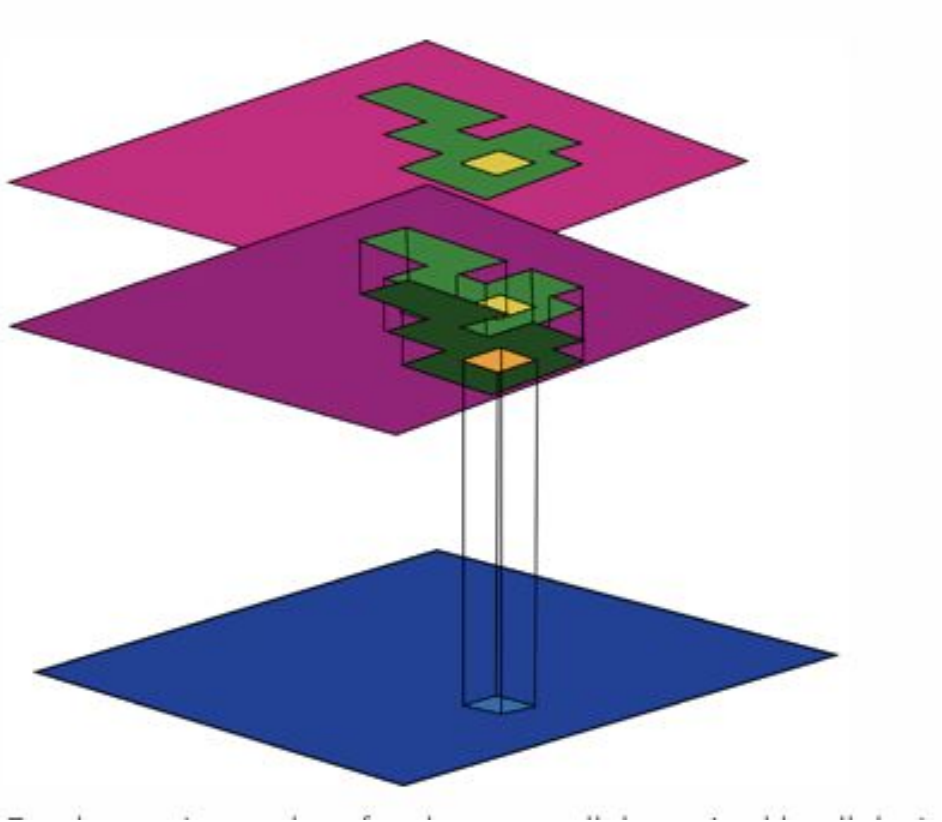
Computes an output raster where the output value for each location depends on the value of

- Value of the cell at the location, and
- The association that location has within a cartographic zone

- Similar to focal operations but zones can be irregularly sized or shaped (not constant)

- Zone could be an administrative unit (county, census tract) a natural unit such as a forest patch, lake, national park boundary

Zonal Operations



Global Operations

Special case of zonal operations where the entire raster dataset represents a *single* zone.

- Example descriptive statistics such as min, max for the entire raster dataset
- Also useful to calculate distance rasters
    - Euclidean distance operations
        - Assign to each cell in the output raster dataset its distance from the closest source cell
            - Eg, Source cell could be the location from which to start a new road
    - Weighted (non-euclidean) distance operations
        - Accounts for friction/cost in addition to euclidean distance
            - Eg, there is more cost to crossing a swamp or a hill to reach the new road compared to flat land
- You need to know the entire surface (cannot have missing values in a raster)

Combining Vectors and Rasters

There are many applications where you might need to combine vectors and rasters

- Eg, extracting temperature values for each point event of malaria to see whether temperature plays a role in where someone is likely to get malaria
- Calculating % deer habitats in each county to estimate chances of lyme disease.
- Moving from an object to field view or vice versa
- Cropping a raster using a vector geometry to show better maps
- And many more - can you think of some?
    - https://geocompr.robinlovelace.net/geometric-operations.html (Section 5.4)

Combining Vectors and Rasters

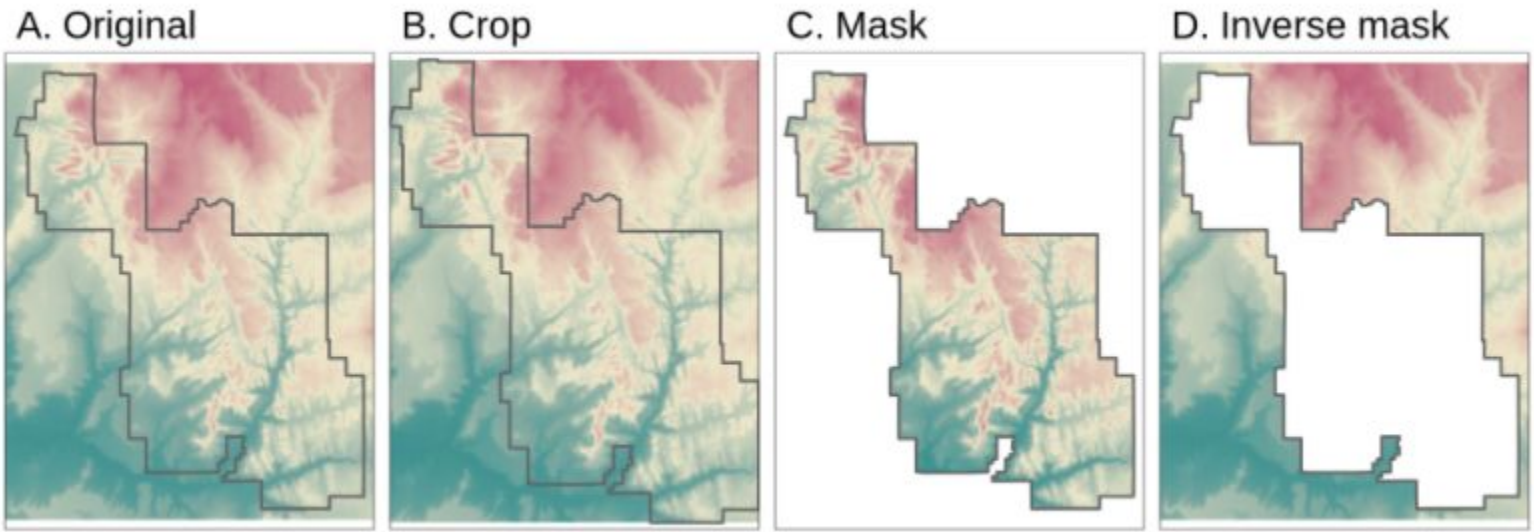


FIGURE 5.17: Illustration of raster cropping and raster masking.

Combining Vectors and Rasters

Rasterization - Vector to Raster

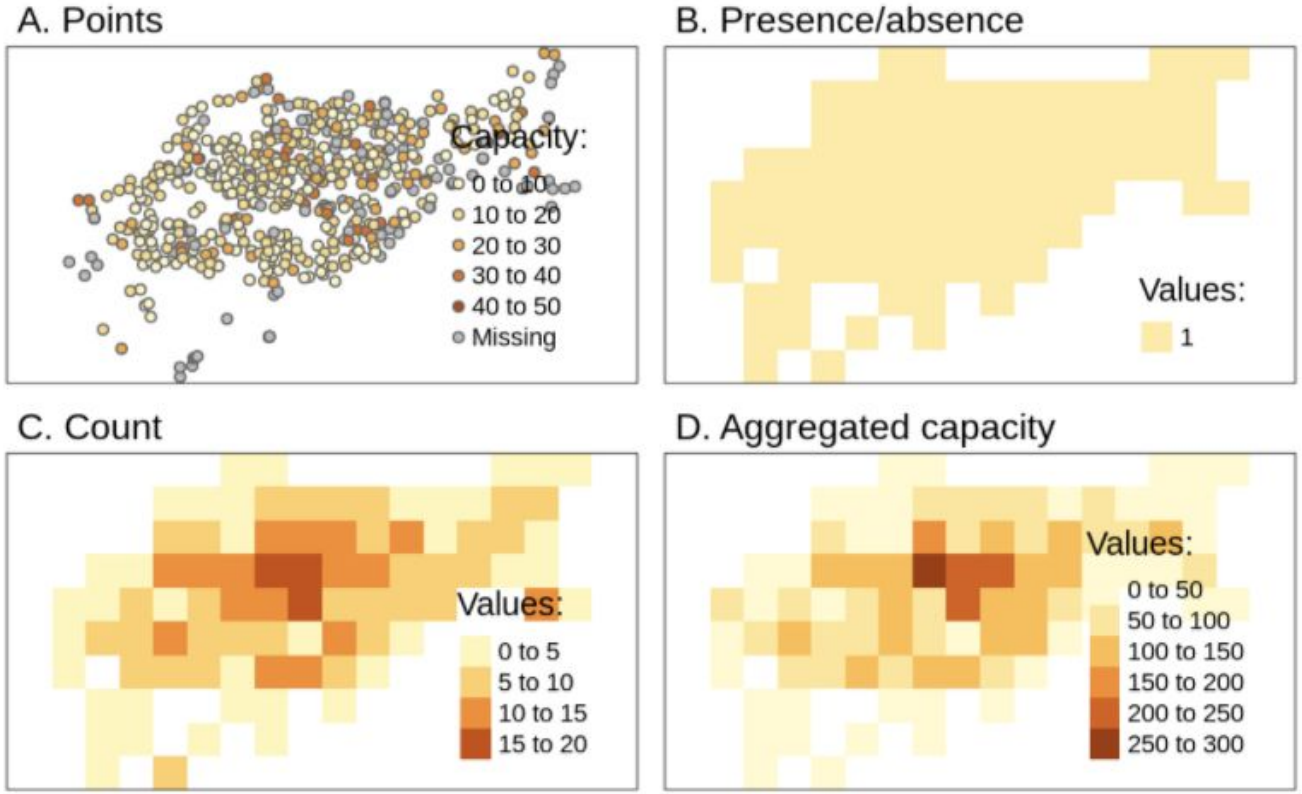


FIGURE 5.21: Examples of point rasterization.

Combining Vectors and Rasters

Vectorization - raster to vector

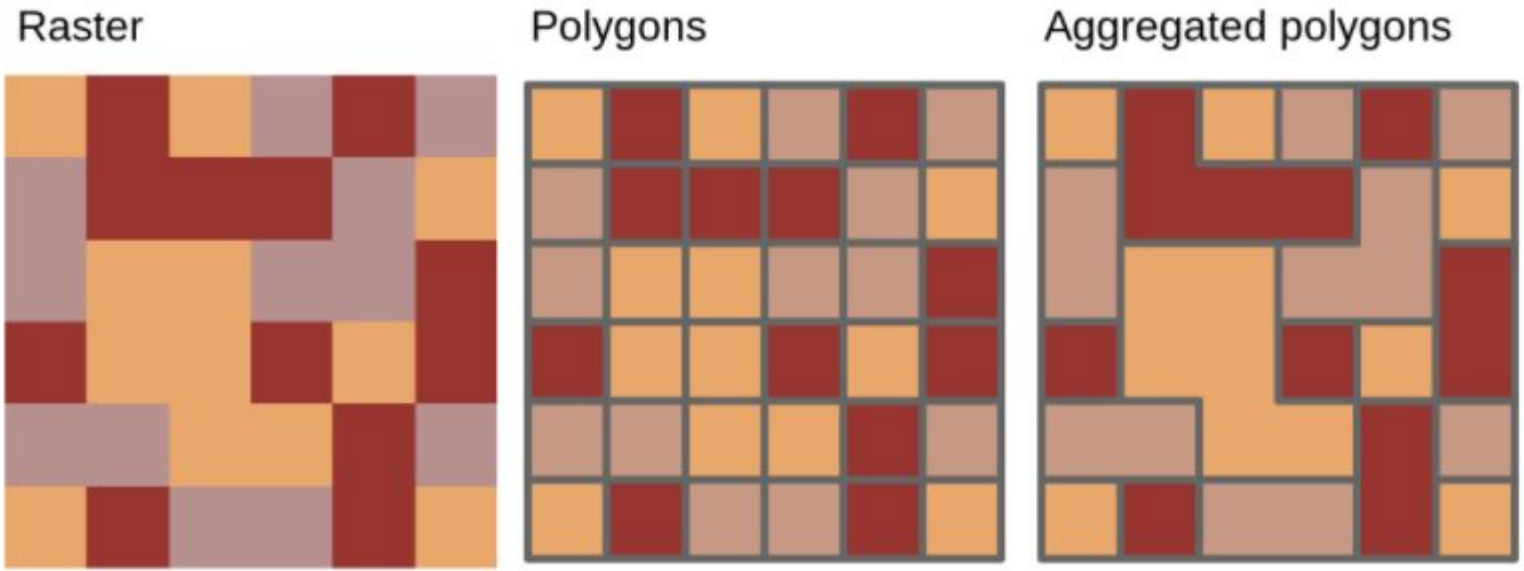


FIGURE 5.25: Illustration of vectorization of raster (left) into polygon (center) and polygon aggregation (right).

Next Class

- Spatial Neighborhoods
  - Geometric Operations
  - Creating different kinds of spatial neighborhoods
    - Foundation to any spatial analysis

- Submit project proposal to get timely detailed feedback
- Think about topological operations to answer your project questions
- Schedule consultation hours with me/TA
- Read assigned readings (including for today's lecture and next week's)
- Practice Practice Practice (Datacamp, lab, readings)