GEOG 491/891: Special Topics - Spatial Analysis in R

Week 7.02: summaries, spatial joins, and the practical effects of topological relations

Dr. Bitterman

Today's schedule

- Open discussion
- Spatial data operations
- For next class

Anything to discuss? Questions?

today's setup

```
library(tidyverse)
library(sf)
library(tmap)
```

new data in the repository

A story from my Vermont research

- Nutrient runoff (primarily phosphorus) causes harmful cyanobacteria blooms in Lake Champlain
- TMDL regulations "limit" the amount of criteria pollutants allowed to enter waterbodies that flow to the Lake
- To regulate, and to improve, we need to measure
- So EPA has broken-down the Lake Champlain Basin into units of analysis called **NHDPlus segments**

But management is a different issue

- We don't commonly manage water-related problems using watershed-dervied jurisdictions
- We should, but we don't
- In Vermont, they'd doing something a bit different
- Tactical Basins, CWSPs, and RPCs

Let's estimate the load in each RPC

Start with the NHDs

```r
# NHDs
nhds <- sf::read_sf("./data/nhdplus_loads.shp") %>% sf::st_make_valid()
glimpse(nhds)
tm_shape(nhds) + tm_polygons("Baseline_L", n = 10)
```

`Baseline_L` is phosphorus load in kg/year

And now the RPCs

```r
# RPCs
rpcs <- sf::read_sf("./data/gn_vt_rpcs.shp") %>% sf::st_make_valid()
glimpse(rpcs)
tm_shape(rpcs) + tm_polygons(col = "INITIALS")
```

Overlay them

```r
tm_shape(rpcs) + tm_borders(col = "red") +
  tm_shape(nhds) + tm_polygons(col = "Baseline_L", n = 7) +
  tm_shape(rpcs) + tm_borders(col = "red")
```

Why am I calling `tm_shape(rpcs)` twice? Any ideas?

Our objective:

Calculate the total phosphorus load in each RPC

Ideas for how to accomplish that task?

A spatial join

```
# do the join
nhd_rpcs <- st_join(nhds, rpcs, join = st_intersects)

# look at it/confirm it worked
glimpse(nhd_rpcs)

# plot it
tm_shape(nhd_rpcs) + tm_polygons(col = "RPC")
```

and now to summarize...

Suggestions?

The "tidy way"

```
# the "tidy way"
nhd_rpcs %>%
  group_by(RPC) %>%
  summarize(totalLoad = sum(Baseline_L))
```

What was the output?

Add a line to plot it

```
nhd_rpcs %>%
  group_by(RPC) %>%
  summarize(totalLoad = sum(Baseline_L)) %>%
  tm_shape(.) + tm_polygons(col = "totalLoad") # <- this line is new
```

the utility of pipes

An alternative method

from your book...

- works a bit differently
- uses `sf` features to do the aggregation

```
# using aggregate instead
aggregate(x = nhds, by = rpcs, FUN = sum) # throws an error... what's the problem?
```

Error message:

```
Error in FUN(X[[i]], ...) : invalid 'type' (character) of argument
```

What's the problem?

Fixing the problem

```
glimpse(nhds) # How many character attribues?

# fix the problem
nhds %>% dplyr::select(-SOURCEFC, -NHDPlus_Ca, -Tactical_B) %>%
  aggregate(x = ., by = rpcs, FUN = sum)
```

Do it again, but assign it to a variable

```r
# same function, but assign it to a variable
agg.rpcs <- nhds %>% dplyr::select(-SOURCEFC, -NHDPlus_Ca, -Tactical_B) %>%
  aggregate(x = ., by = rpcs, FUN = sum)
```

Plot it

```r
# plot it... why and how is it different?
tm_shape(agg.rpcs) + tm_polygons(col = "Baseline_L")
```

how/why is it different than our earlier plot?

So, have we completed our task?

Any remaining issues?

Let's check our spatial relationships

Break down the code

```
# issues with overlap
nhd_rpcs %>% group_by(NHDPlus_ID) %>% summarise(count = n()) %>%
  arrange(desc(count))
```

What's the result?

Area-weighted interpolation

(to the whiteboard!)

```r
# area-weighted interpolation
interp.loads <- nhds %>% dplyr::select(Baseline_L, geometry) %>%
  st_interpolate_aw(., rpcs, extensive = T)

tm_shape(interp.loads) + tm_polygons(col = "Baseline_L")
```

Same? Different? How can we compare?

Do a join

(note, I'm cheating a bit here and relying on the RPC geometry being identical)

```r
# do a join
comparison <- st_join(agg.rpcs, interp.loads, st_equals)

# calculate the error, then map it
tmap_mode("view")

comparison %>% mutate(diff = Baseline_L.x - Baseline_L.y) %>%
  tm_shape(.) + tm_polygons(col = "diff") +
  tm_shape(nhds) + tm_borders(col = "blue")
```

How different were the methods?

If there's time, return to the code reading activity…

For next week

- Wildcard Friday is no-class Friday
- Chapter 5 from Lovelace (https://geocompr.robinlovelace.net/geometric-operations.html)
- Practice, practice, practice
- Lab 02 - keep working
- Work on your projects (in-class updates on 10/29)