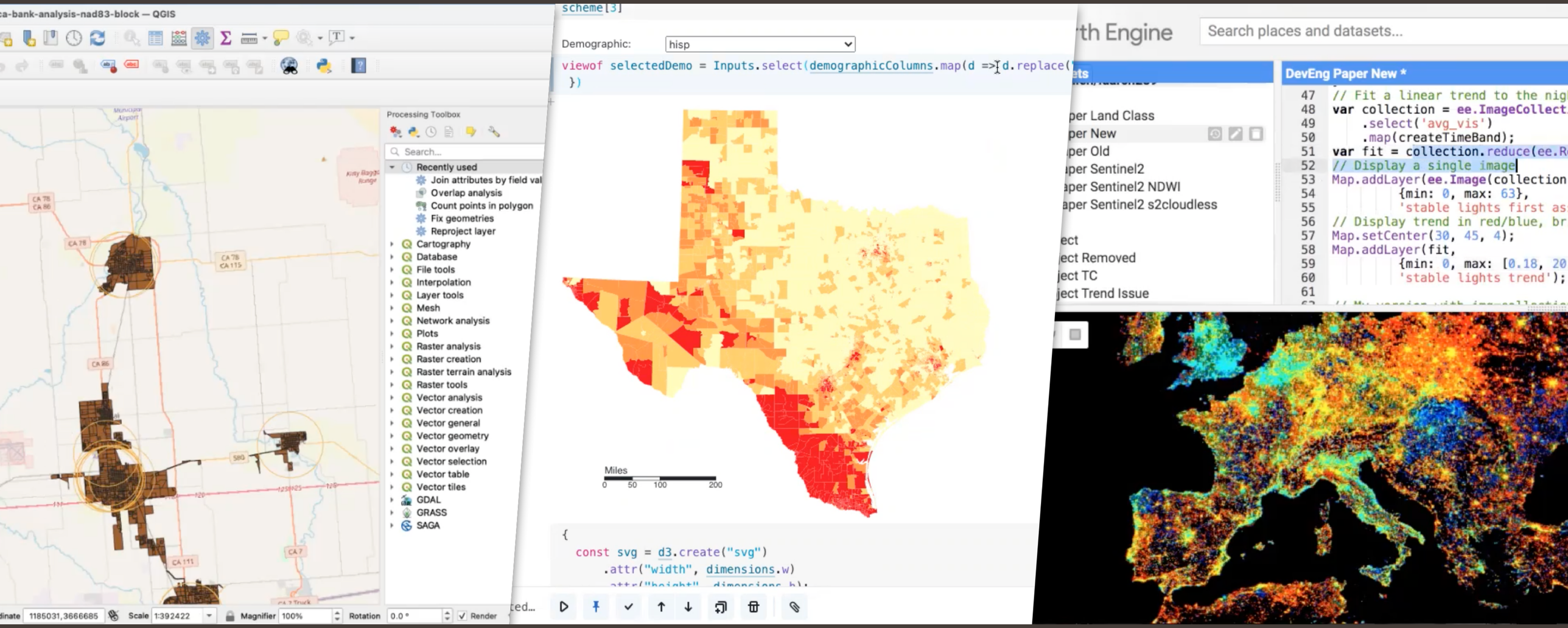


Early findings from a cross-domain need-finding study with users of geospatial data

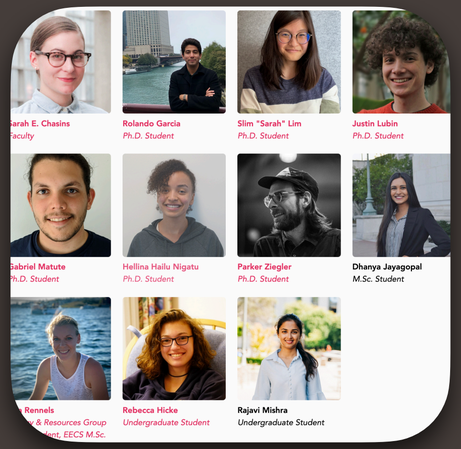
Programming Systems Seminar // April 7, 2022
UC Berkeley EECS

Parker Ziegler // @parkie-doo

Acknowledgments

Sarah E. Chasins
Ph.D. Advisor,
Principal Investigator

Rachel Leven
CDSS Media
Communications
Specialist



PLAIT Lab
Kind, thoughtful friends who let me rant about maps, KEXP, and the Red Sox

My Participants
Computing tools are built for those who participate!

A Note on Feedback

1. Put on your **{CHI, UIST, OOPSLA, PLDI}** reviewer hat!

2. Pay special attention to the **Motivation** section

What's this talk about?

Sharing **preliminary findings** from a **need-finding study** with users of geospatial data.

25

participants

3

domains of interest

30hrs **16**mins

of video from observational interviews

Roadmap


1. What is geospatial data? What is GIS?
2. Motivation
3. Research Questions
4. Study Design
5. Findings and Discussion

Roadmap


1. What is geospatial data? What is GIS?
2. Motivation
3. Research Questions
4. Study Design
5. Findings and Discussion

What is geospatial data?

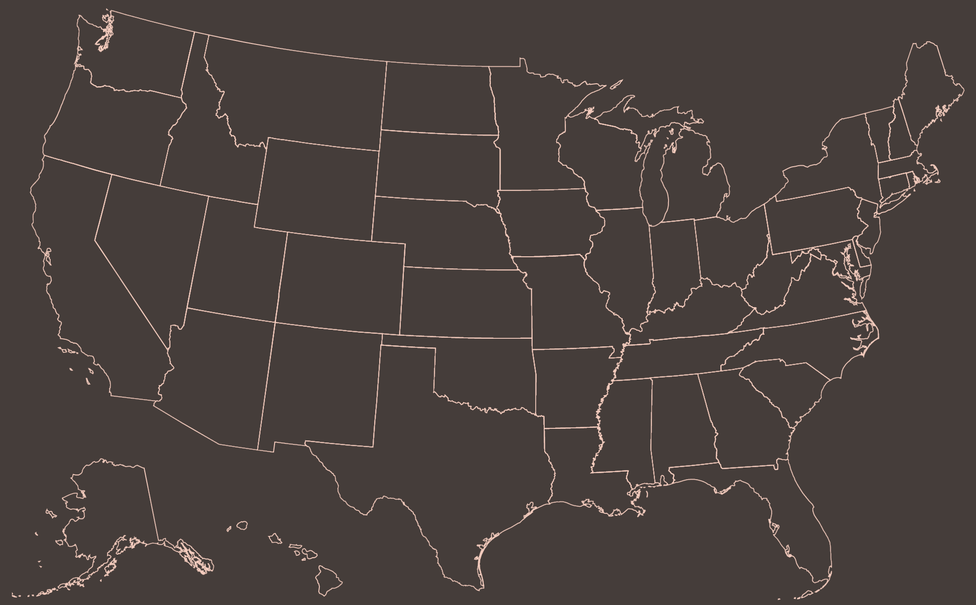
GEOID	STATE	TOTAL_POP	NUM_SHOOTINGS	TOTAL_AREA
05	AR	3,011,524	27	134,770.0
27	MN	5,706,494	37	206,233.0
32	NV	3,104,614	34	284,331.5
09	CT	3,605,944	9	12,540.7
51	VA	8,631,393	41	102,278.6
22	LA	4,657,757	49	111,897.8
41	OR	4,237,256	34	248,607.8
18	IN	6,785,528	41	92,788.9
13	GA	10,214,860	69	148,958.0
39	OH	11,799,448	64	105,829.5
02	AK	733,391	13	1,477,953.4
49	UT	3,271,616	18	212,819.3
33	NH	1,377,529	5	23,188.2
21	KY	4,505,836	41	102,268.3
35	NM	2,117,522	43	314,160.4

Tabular Data

What is geospatial data?

Data in which **location** is encoded alongside **attributes**

GEOID	STATE	TOTAL_POP	NUM_SHOOTINGS	TOTAL_AREA
05	AR	3,011,524	27	134,770.0
27	MN	5,706,494	37	206,233.0
32	NV	3,104,614	34	284,331.5
09	CT	3,605,944	9	12,540.7
51	VA	8,631,393	41	102,278.6
22	LA	4,657,757	49	111,897.8
41	OR	4,237,256	34	248,607.8
18	IN	6,785,528	41	92,788.9
13	GA	10,214,860	69	148,958.0
39	OH	11,799,448	64	105,829.5
02	AK	733,391	13	1,477,953.4
49	UT	3,271,616	18	212,819.3
33	NH	1,377,529	5	23,188.2
21	KY	4,505,836	41	102,268.3
35	NM	2,117,522	43	314,160.4



Attributes

Geometry

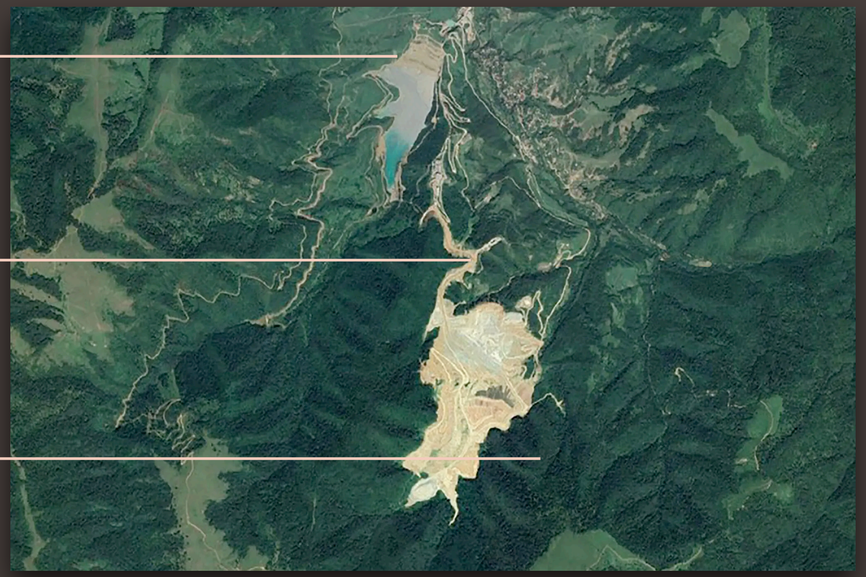
What is geospatial data?

Data in which **location** is encoded alongside **attributes**

B4: 106, B3: 121, B2: 102

B4: 193, B3: 181, B2: 154

B4: 40, B3: 60, B2: 58



```
bbox: [
  [44.44, 40.49],
  [44.62, 40.45],
  [44.80, 40.49],
  [44.62, 40.53]
]
```

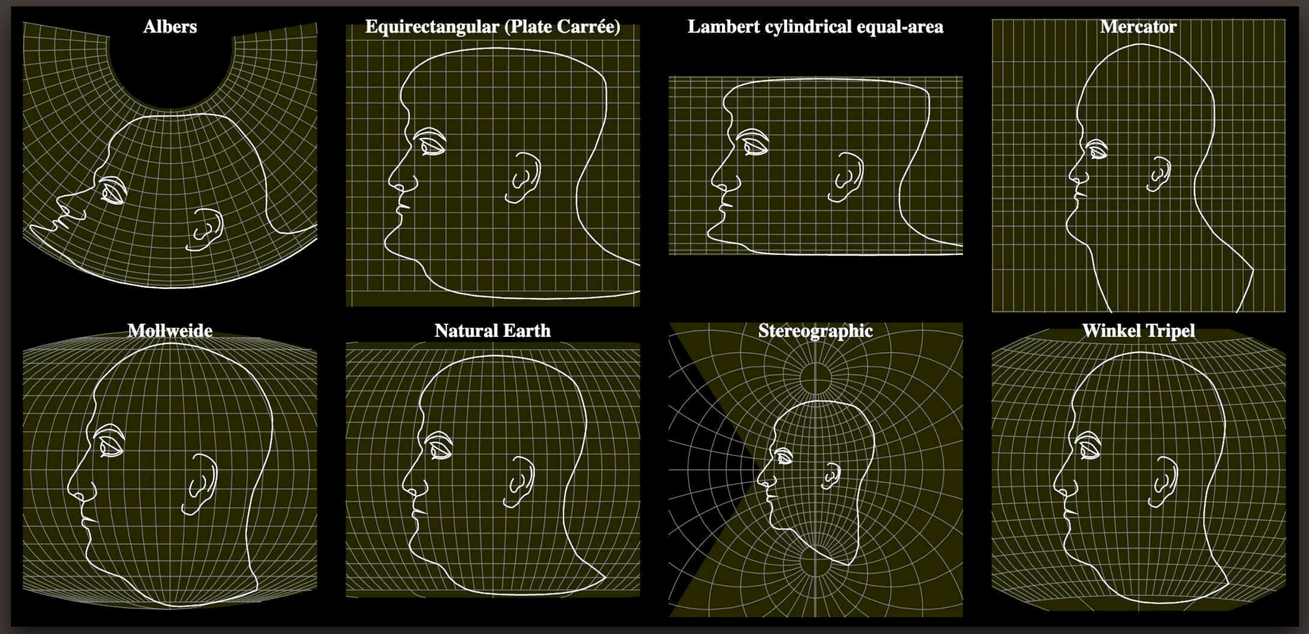
Attributes

Geometry

What is geospatial data?

Location is encoded in reference to a **coordinate system**

All coordinate systems trade-off between **size**, **shape**, and **distance**



Andy Woodruff's projected heads, based off Charles Deetz's 1921 book
Elements of Map Projection with Applications to Map and Chart Construction

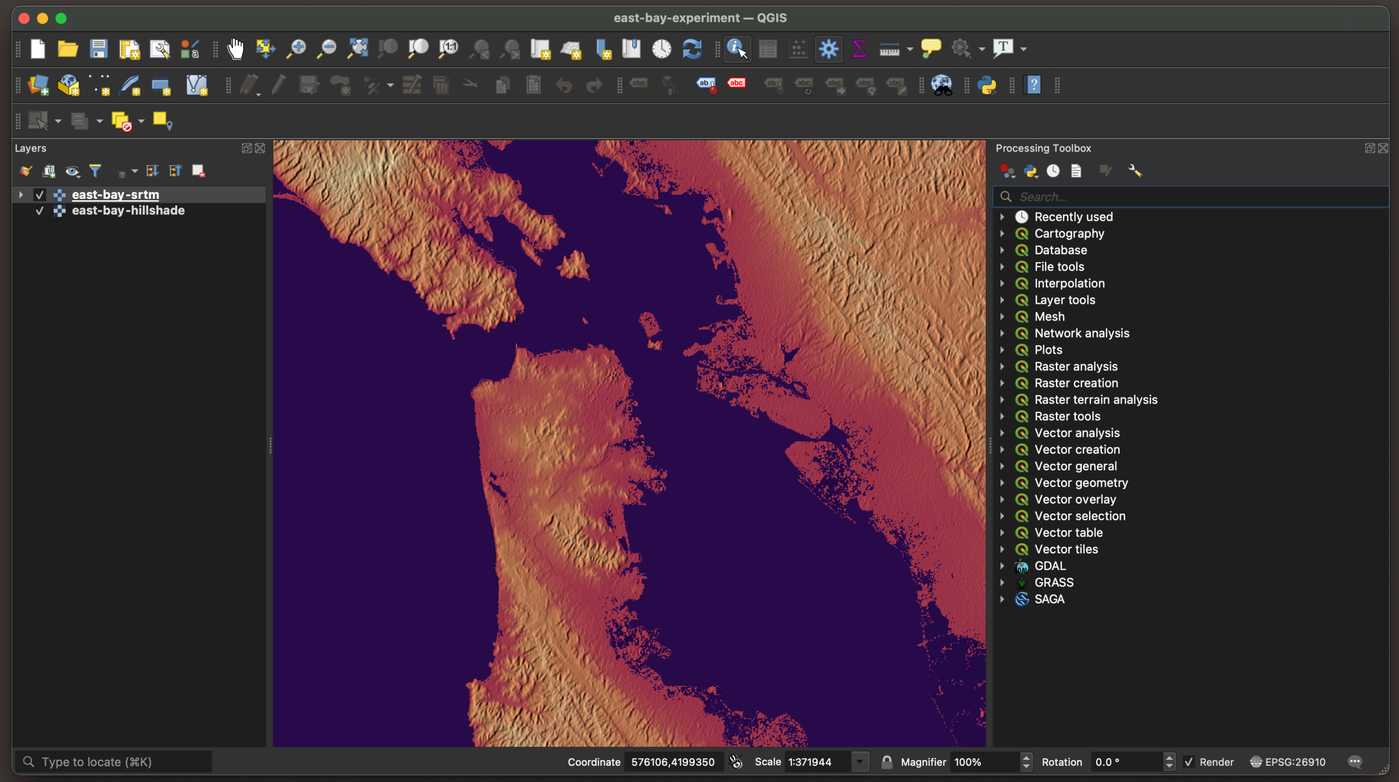
What is GIS?

Geographic
Information
System

What is GIS?

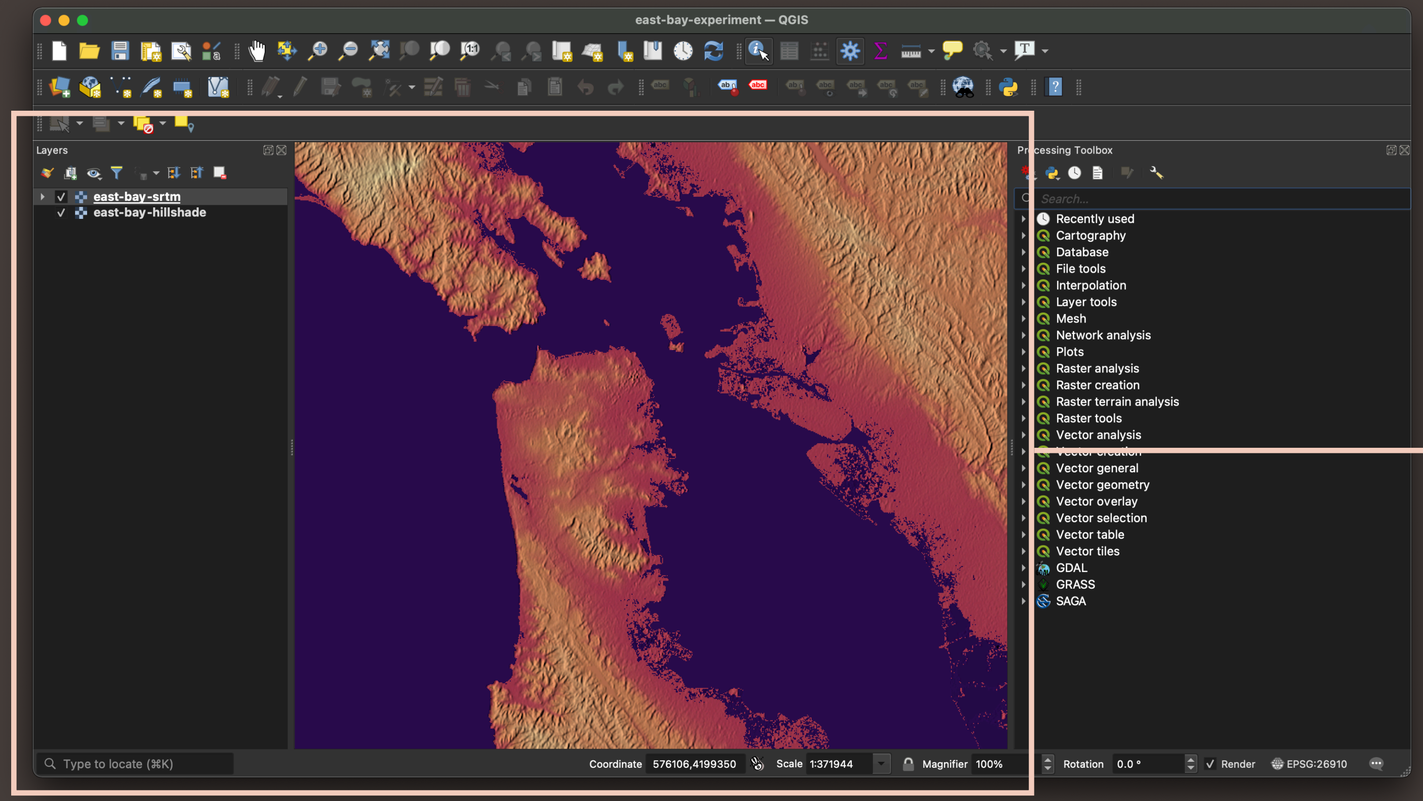
Geographic
Information
System

Software for **viewing**, **operating on**, and **managing** geospatial data.



QGIS

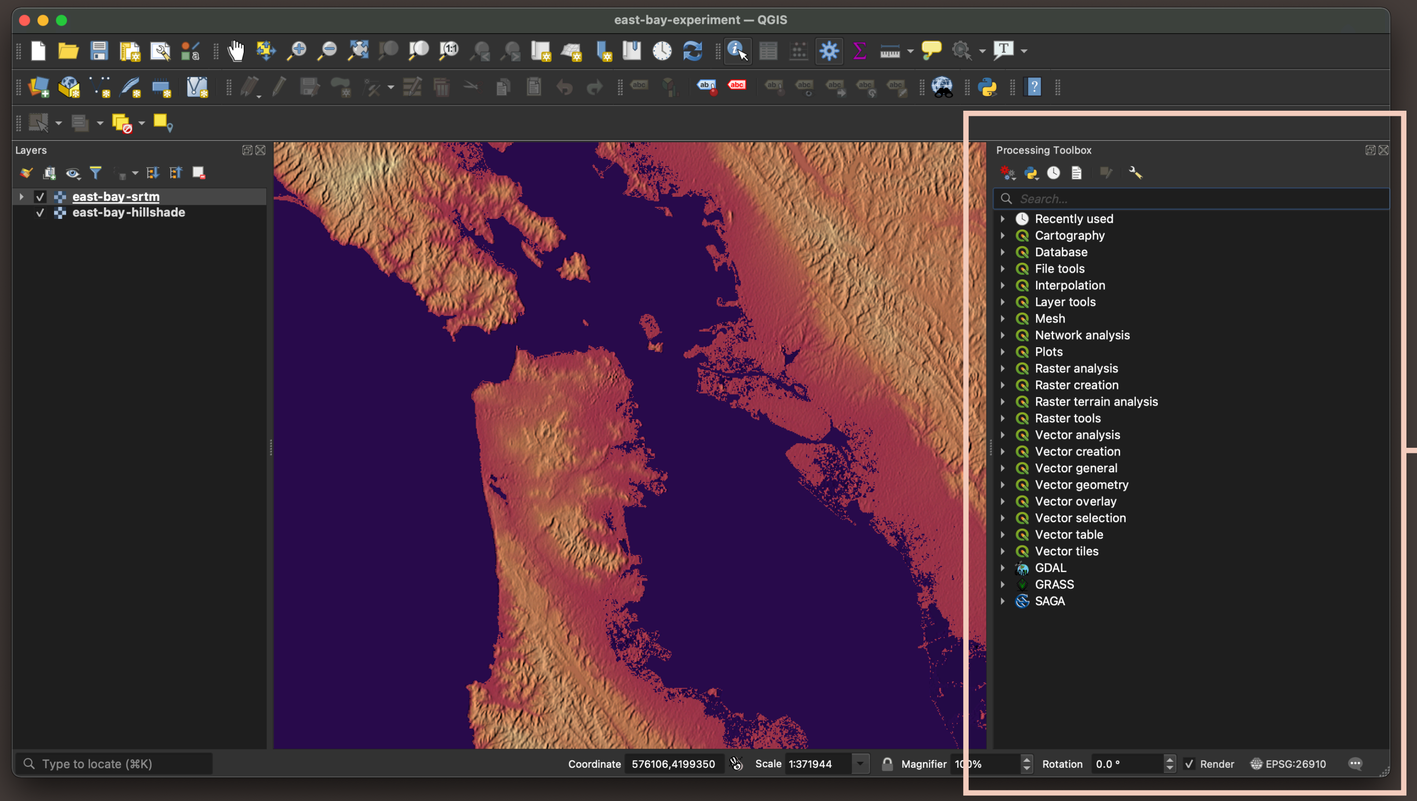
What is GIS?



Zoomable, **pan-able canvas** for viewing geospatial data as **layers**

QGIS

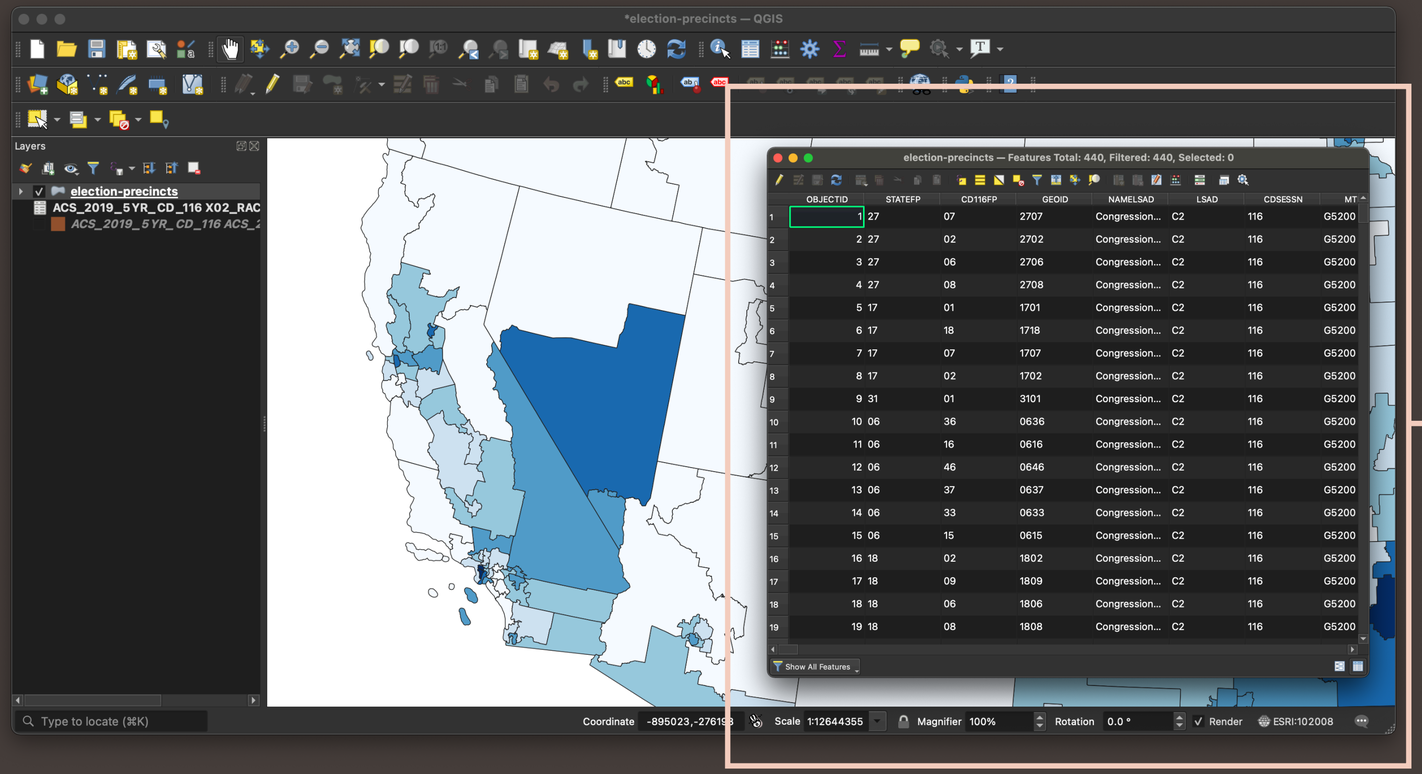
What is GIS?



Processing tools for operating on geospatial data to produce new layers

QGIS

What is GIS?



Attribute tables for connecting tabular data to geometry

QGIS

Roadmap


1. What is geospatial data? What is GIS?
2. Motivation
3. Research Questions
4. Study Design
5. Findings and Discussion

Roadmap


1. What is geospatial data? What is GIS?
2. Motivation
3. Research Questions
4. Study Design
5. Findings and Discussion

Motivation

Why study the challenges and needs of geospatial data users?

1. There has been **relatively little focus** on tools for working with geospatial data in PL and HCI circles.

2. There is a **growing gap** between the **amount** of geospatial data and the **experts** capable of analyzing it.

3. Geospatial data is fundamental to understanding **climate change**, **public health**, **election integrity**, **racial and economic inequity**, and much more.

Motivation

Why study the challenges and needs of geospatial data users?

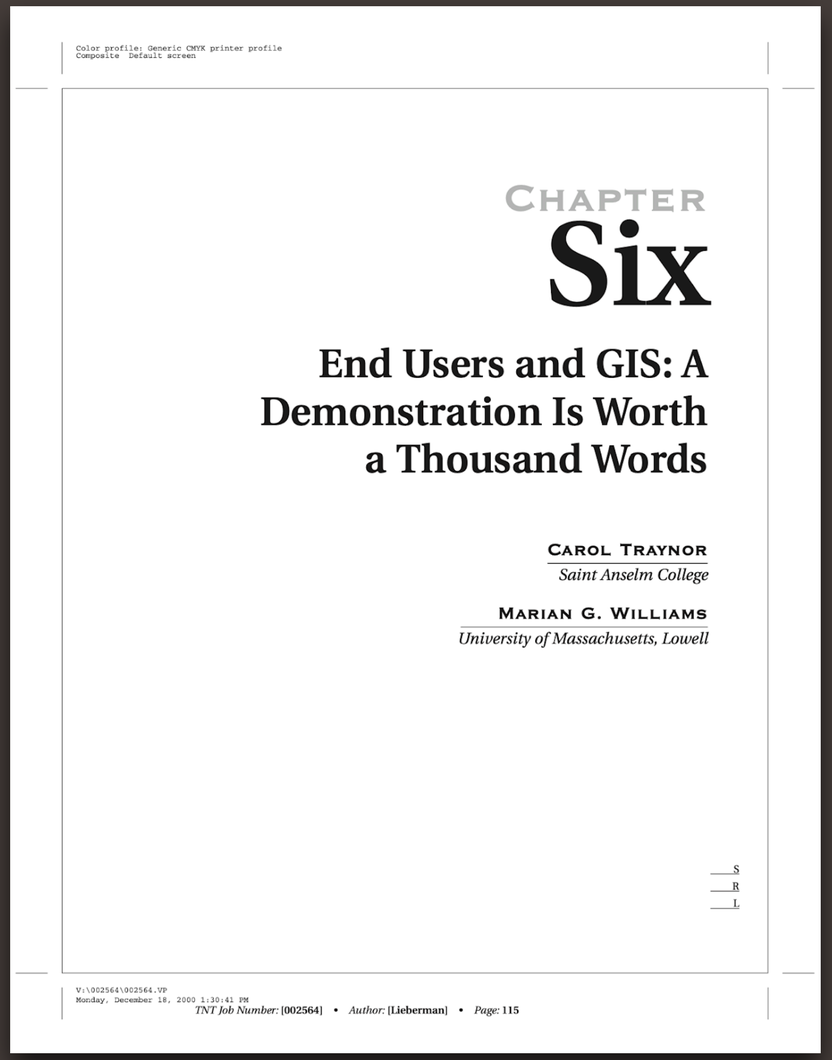
1. There has been **relatively little focus** on tools for working with geospatial data in PL and HCI circles.

Geospatial Data in HCI

“Why Are Geographic Information Systems Hard To Use?”
Traynor and Williams, CHI Short Papers, 1995

“End Users and GIS: A Demonstration Is Worth a Thousand Words”
Traynor and Williams, Your Wish is My Command: *Programming By Example*, 2001

“Human-computer interaction and geospatial technologies – context”
Mordechai Hackley, *Interacting with Geospatial Technologies*, 2010



Color profile: Generic CMYK printer profile
Composite Default screen

CHAPTER
Six

End Users and GIS: A Demonstration Is Worth a Thousand Words

CAROL TRAYNOR
Saint Anselm College

MARIAN G. WILLIAMS
University of Massachusetts, Lowell

S
R
L

V:\002564\002564.VP
Monday, December 18, 2000 1:30:41 PM
TNT Job Number: [002564] • *Author:* [Lieberman] • *Page:* 115

Geospatial Data in HCI

The Many Skillset Problem

Your Domain Knowledge

Geography

Cartography

Databases

Programming

Statistics

Geospatial Data in HCI

The Disconnected Toolbox Problem

ArcGIS

Server Toolbox

Ready to Use Toolbox

Spatial Analyst Toolbox

Spatial Statistics Toolbox

... +35 More

Bitwise Left Shift

Kriging

Raster Calculator

Iso Cluster Unsupervised

Fuzzy Overlay

Zonal Histogram

Darcy Flow

... +200 More

Geospatial Data in HCI

What does this work miss?

This work relies (mainly) on **experts analyzing GIS software** rather than **directly observing GIS users**.

This work hasn't been revisited to address **modern geospatial tools** beyond desktop GIS software.

This work doesn't consider **domain-specific needs** of geospatial data users.

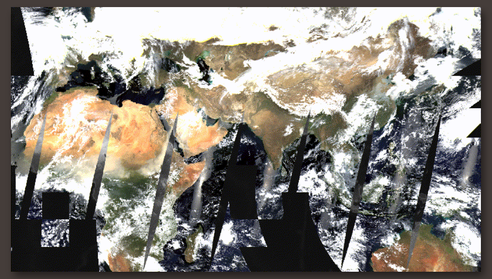
Motivation

Why study the challenges and needs of geospatial data users?

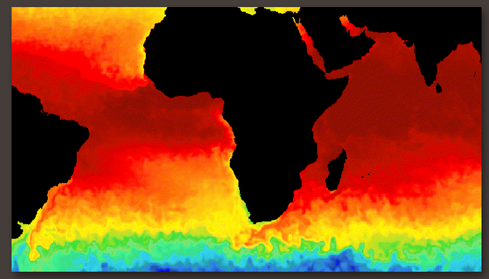
2. There is a **growing gap** between the **amount** of geospatial data and the **experts** capable of analyzing it.

An Abundance of Geospatial Data

Rolf et al. estimate that global imagery data increases by ***80TB per day.***



NASA MODIS

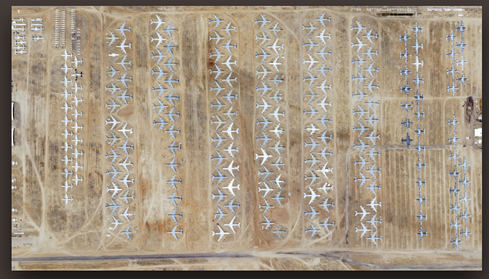


NASA ASTER

ESA Sentinel-2

NASA / USGS Landsat

NASA SRTM



USDA NAIP

Rolf, E., Proctor, J., Carleton, T. et al. A generalizable and accessible approach to machine learning with global satellite imagery. *Nat Commun* 12, 4392 (2021). https://doi.org/10.1038/s41467-021-24638-z

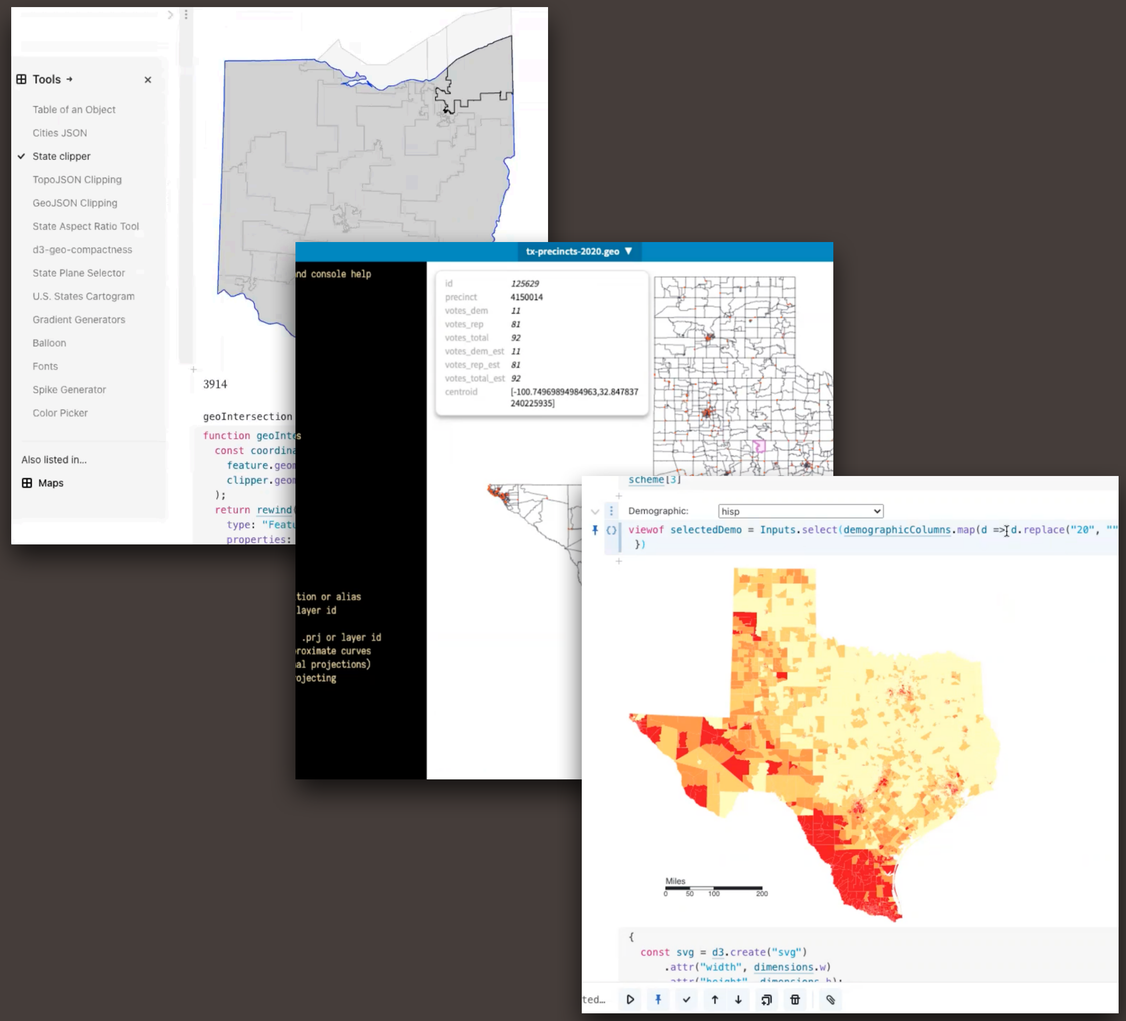
Motivation

Why study the challenges and needs of geospatial data users?

3. Geospatial data is fundamental to understanding **climate change**, **public health**, **election integrity**, **racial and economic inequity**, and much more.

The Importance of Geospatial Data

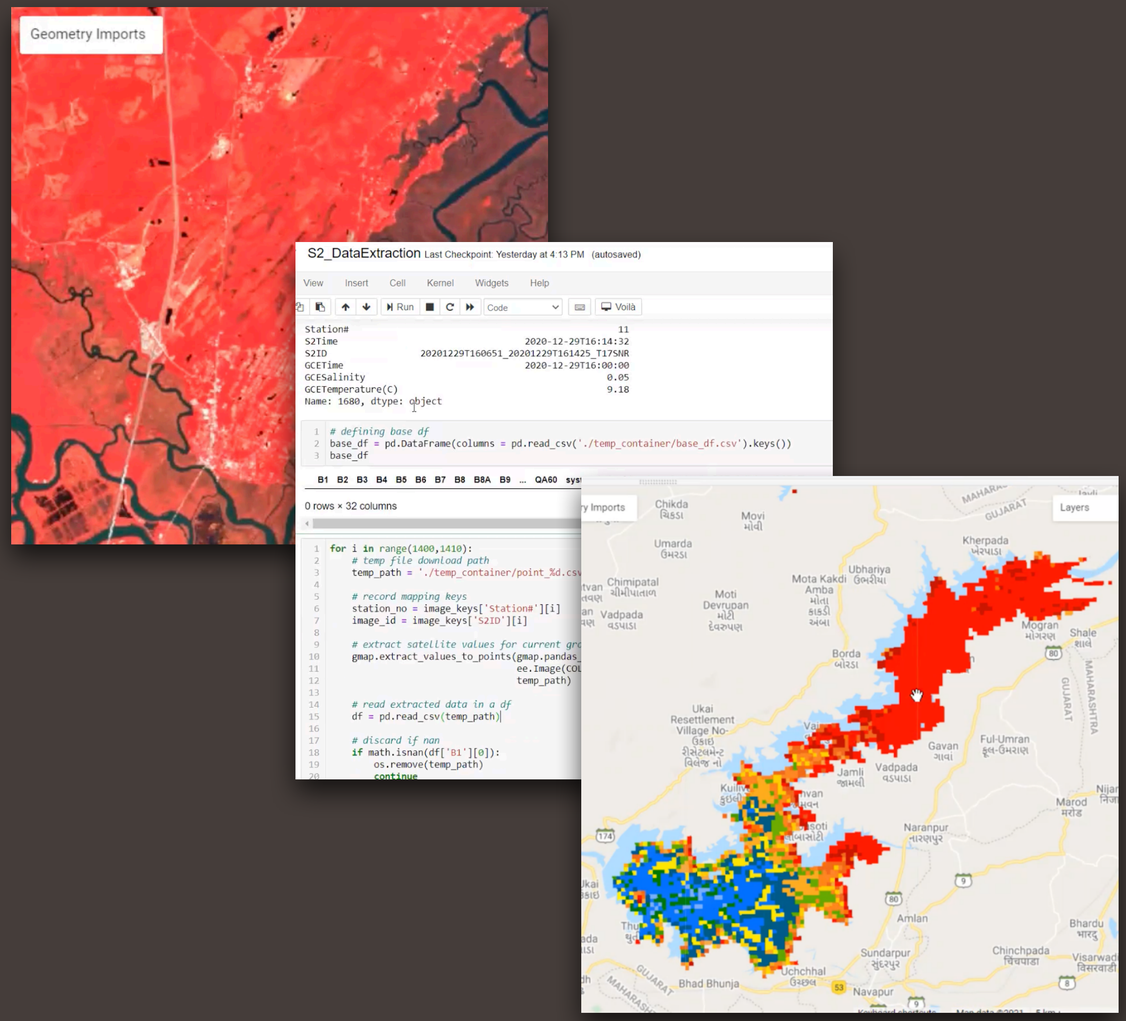
Participant 13 is a **data journalist** exploring whether Texas' new electoral precincts **"pack" or "crack" racial minority groups**.



Observable JavaScript

The Importance of Geospatial Data

Participant 8 is a **climate scientist** building **ML models** to predict development of Cyanobacterial harmful algal blooms (CyanoHABs) from satellite imagery.



Google Earth Engine

Jupyter

Python

Roadmap


1. What is geospatial data? What is GIS?
2. Motivation
3. Research Questions
4. Study Design
5. Findings and Discussion

Roadmap


1. What is geospatial data? What is GIS?
2. Motivation
3. Research Questions
4. Study Design
5. Findings and Discussion

Research Questions

1.

What are the **challenges** users face when attempting to **gather**, **analyze**, or **visualize** geospatial data?

2.

Which of these challenges are **shared** between **experts** and **non-experts** alike? Which are **unique** to one group?

3.

Which of these challenges are **shared** across our **domains of interest?** Which are **unique** to one group?

Roadmap


1. What is geospatial data? What is GIS?
2. Motivation
3. Research Questions
4. Study Design
5. Findings and Discussion

Roadmap


1. What is geospatial data? What is GIS?
2. Motivation
3. Research Questions
4. Study Design
5. Findings and Discussion

Study Design

25

participants

Domain

9

Earth and climate scientists

8

Data journalists

6

Social scientists

2

“Miscellaneous”

Tool Usage

14

Programming Environments

8

GUI-based Software

3

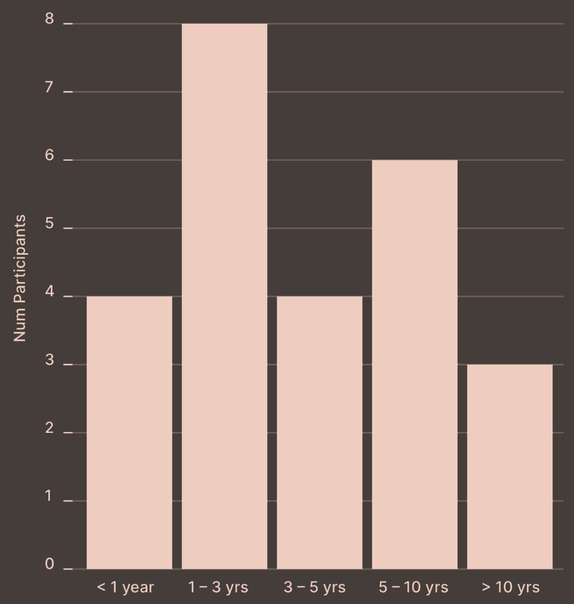
Both

Study Design

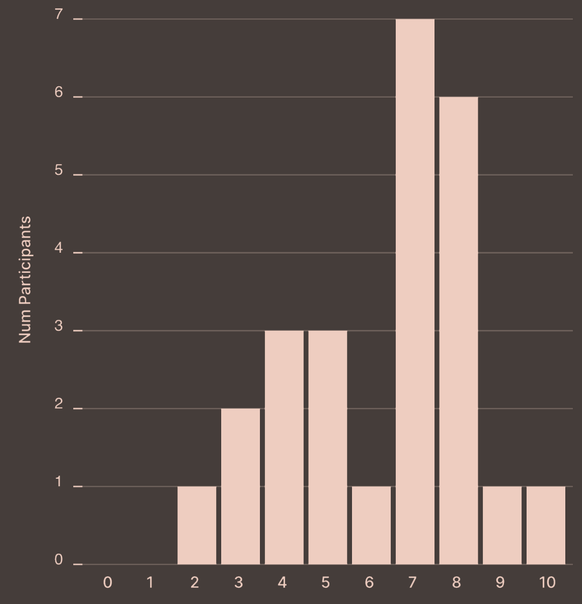
25

participants

Expertise



How long have you been working with geospatial data?



How would you assess your skill level in working with geospatial data?

Methods

Contextual Inquiry

> “…go where the [user] works, observe the [user] as he or she works, and talk to the [user] about the work.”

Karen Holtzblatt and Hugh Beyer. 2017. 3 - Principles of Contextual Inquiry. In *Contextual Design (Second Edition)*.

Methods

Thematic Analysis

> "...a method for identifying, analysing and reporting patterns (themes) within data. It minimally organizes and describes your data set in (rich) detail."

Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. Qualitative Research in Psychology 3, 2 (April 2006), 77–101.

Roadmap

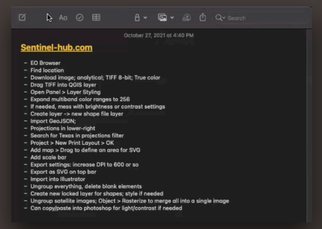

1. What is geospatial data? What is GIS?
2. Motivation
3. Research Questions
4. Study Design
5. Findings and Discussion

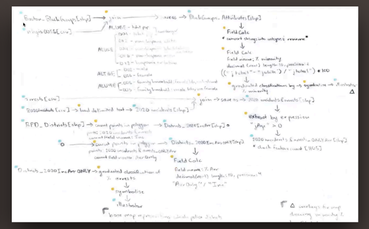
Roadmap


1. What is geospatial data? What is GIS?
2. Motivation
3. Research Questions
4. Study Design
5. Findings and Discussion

Findings

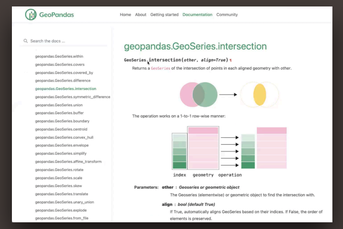
⚠️ These findings are **preliminary**.

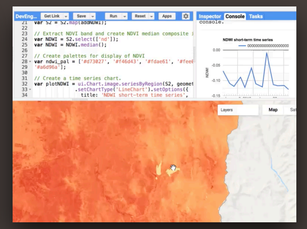




Creating informal program representations

GIS Software

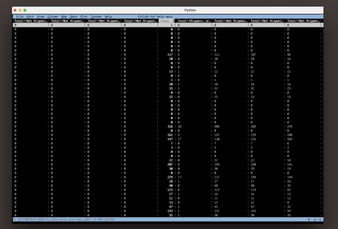


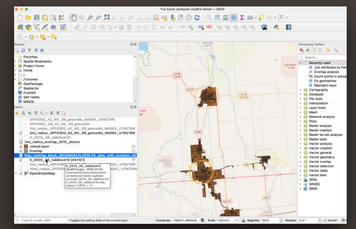


Reasoning about geospatial operator behavior

GIS Software

Programming Environments

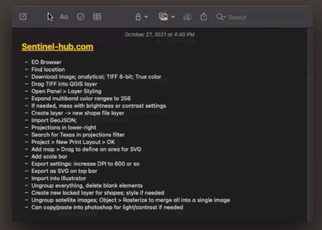


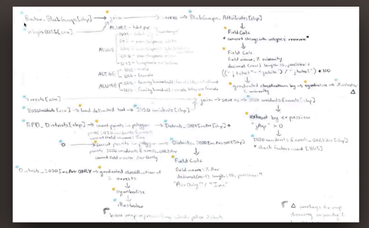


Lack of spatial visibility in code-based tools

Programming Environments

Findings





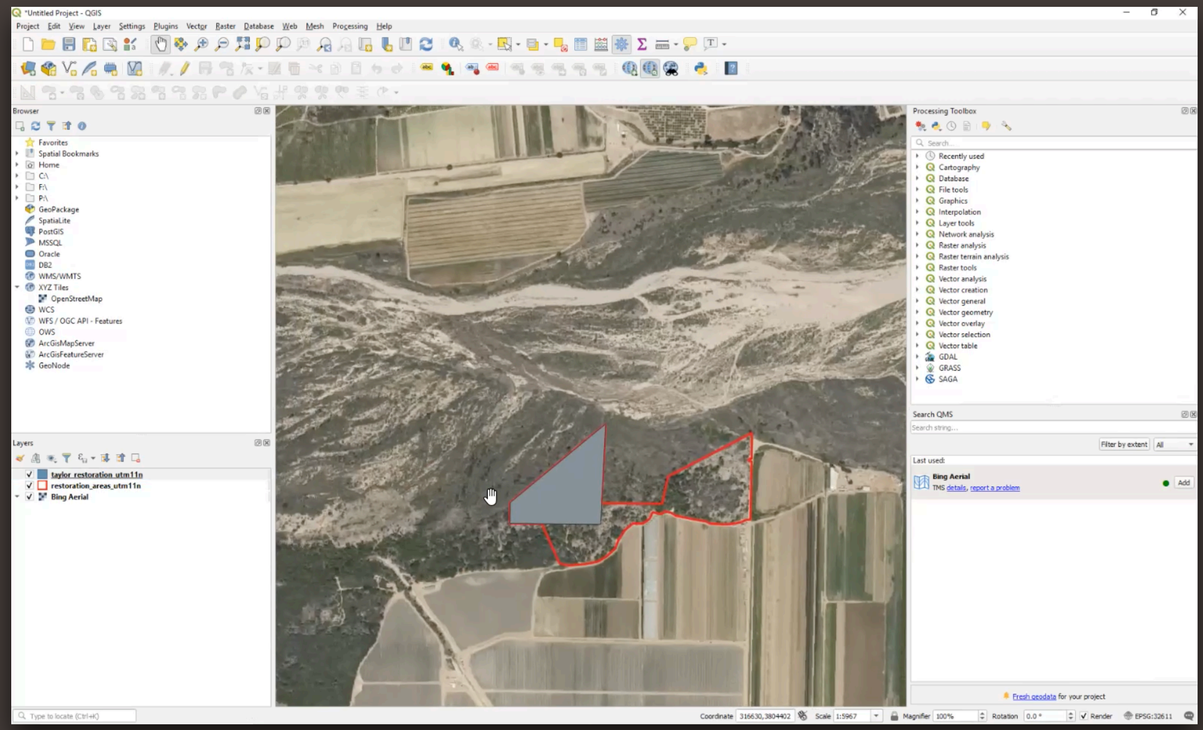
Creating informal program representations

GIS Software

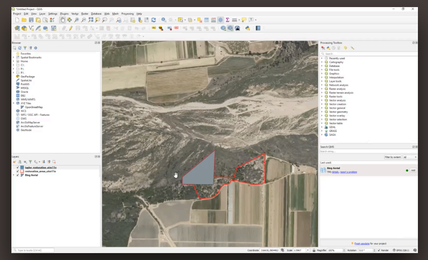
GIS software maintains only a minimal record of steps taken by a user in the interface, **but users want detailed process information**.

Creating informal program representations

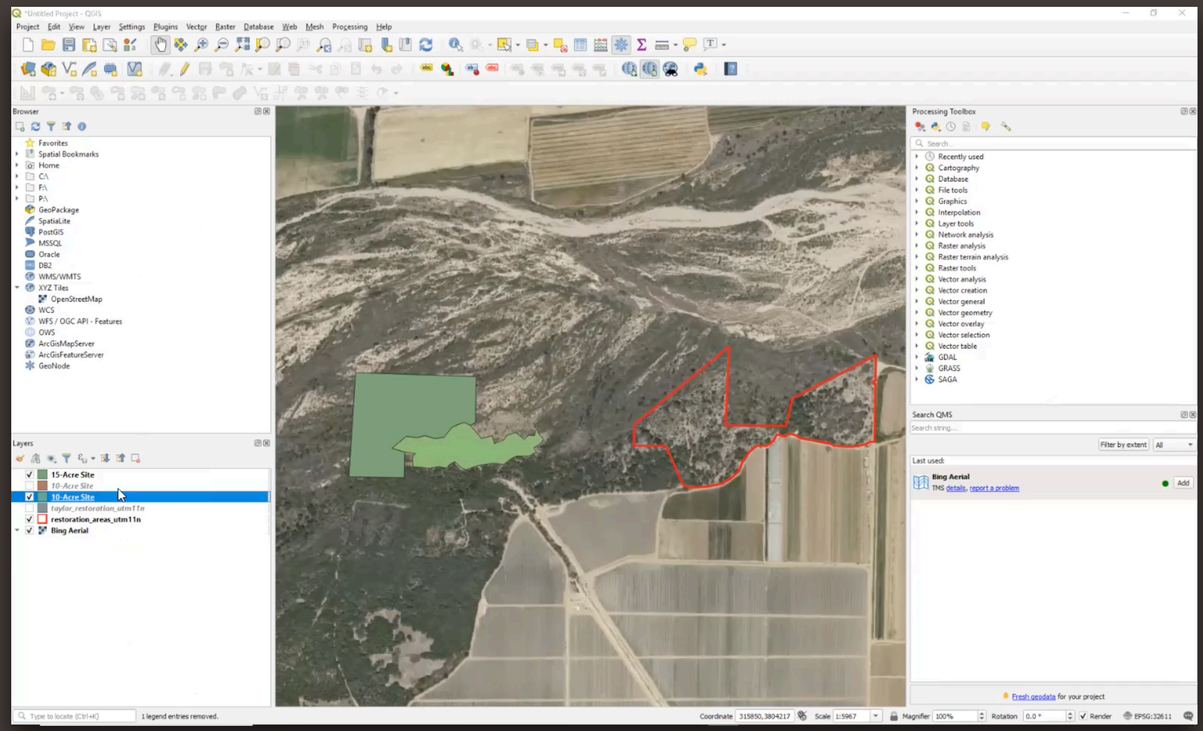
1. **Load two shapefiles** into a new QGIS project. **Change Layer Symbology** to use red outline with no fill.



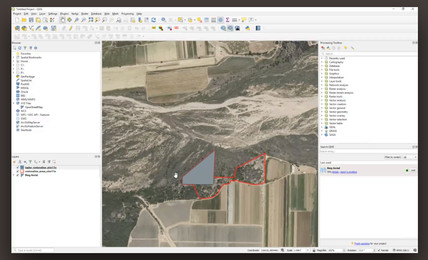
Creating informal program representations

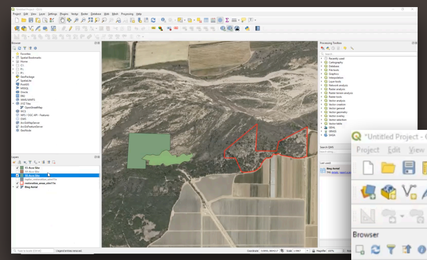
1. 

2. **Load two additional shapefiles** into the QGIS project. **Toggle layer visibility** of one previous layer.

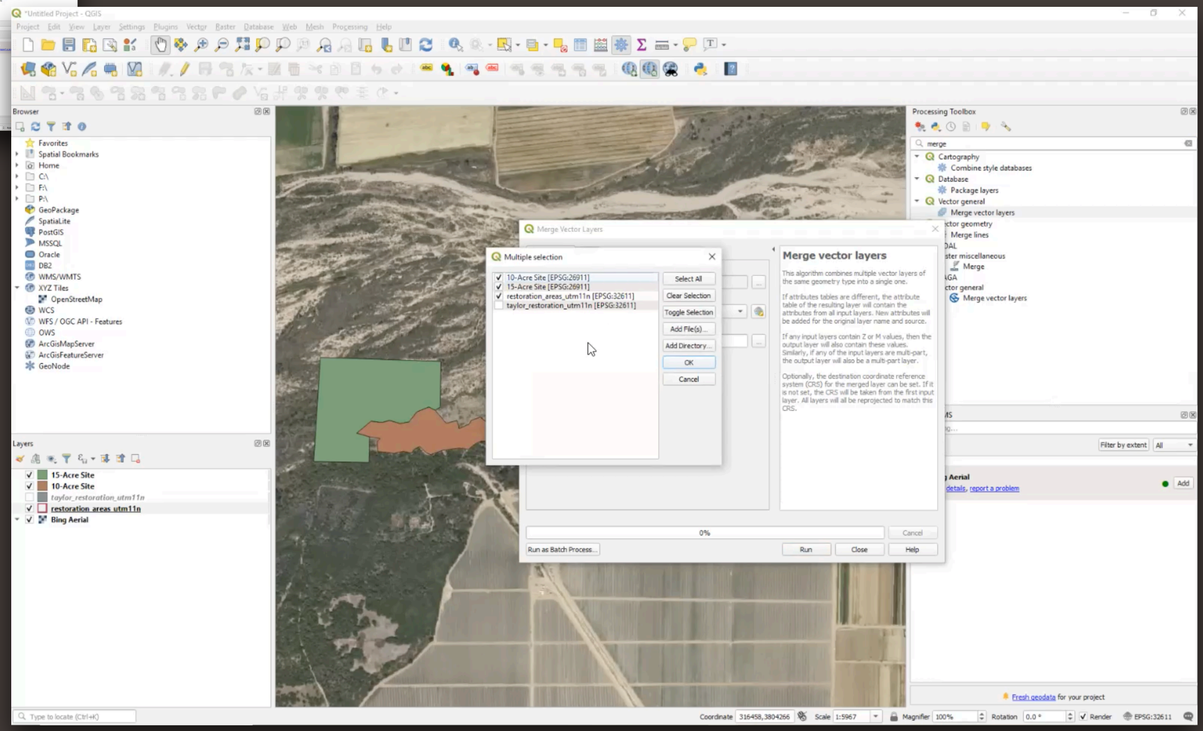


Creating informal program representations

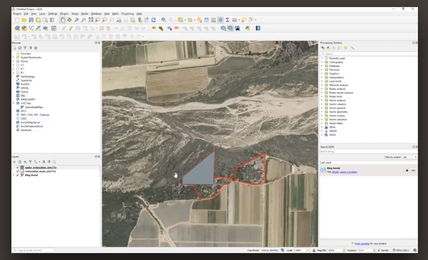
1. 

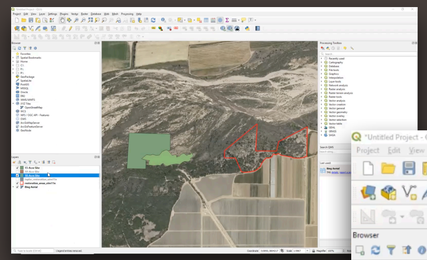
2. 

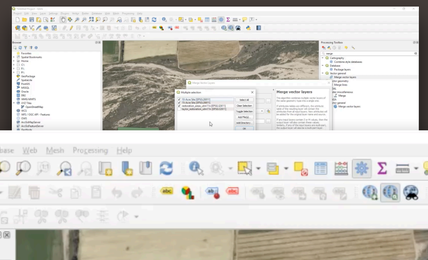
3. **Attempt a `Merge` operation** on the three visible shapefiles. **Create a temporary layer** for the result.



Creating informal program representations

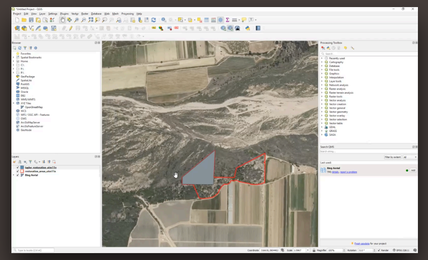
1. 

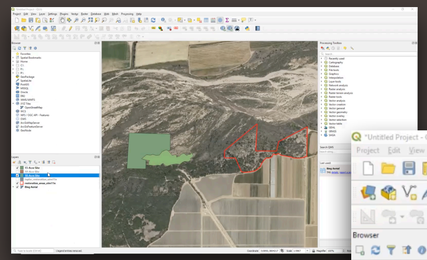
2. 

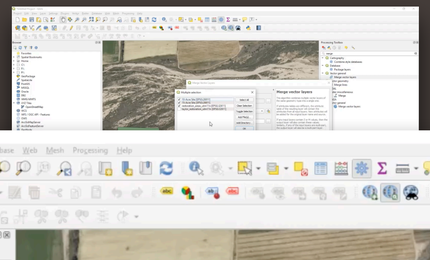
3. 

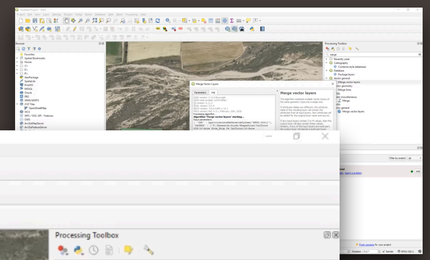
4. `Merge` **operation fails** due to mismatching data types in attribute tables of source layers.

Creating informal program representations

1. 

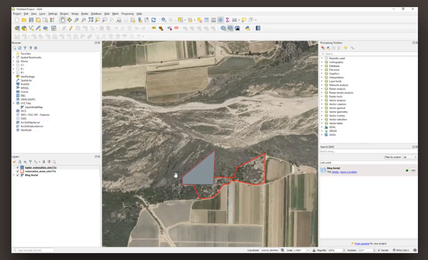
2. 

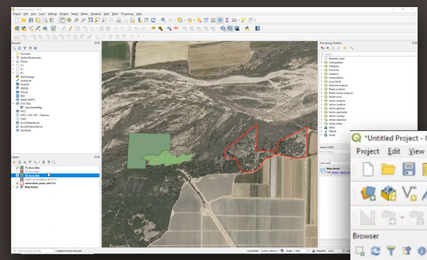
3. 

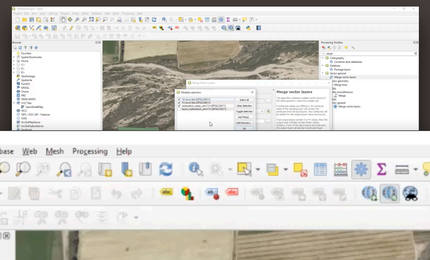
4. 

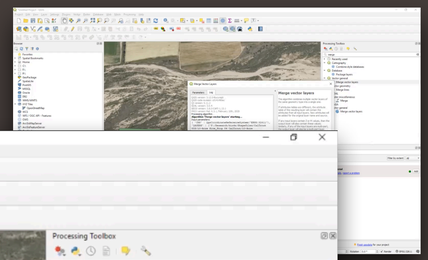
5. **Attempt** `Union` **operation** of just two shapefiles.

Creating informal program representations

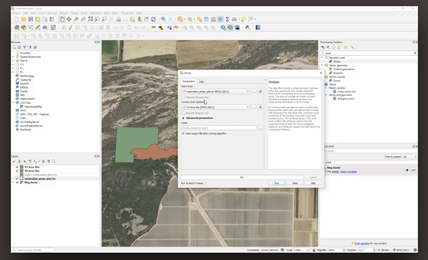
1. 

2. 

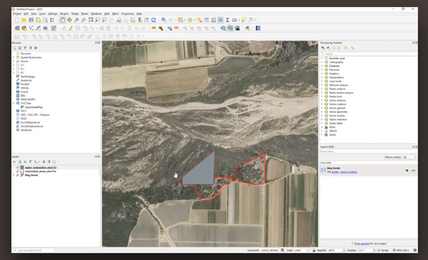
3. 

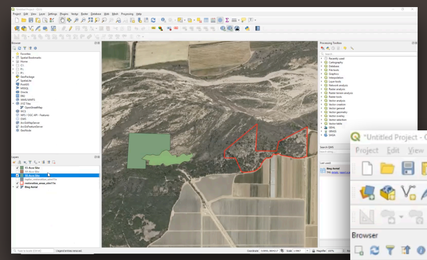
4. 

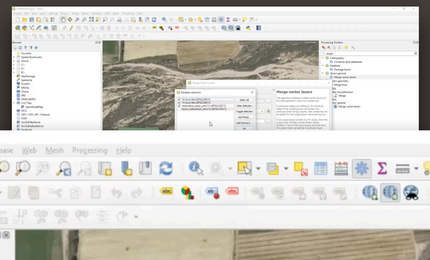
6. `Union` **operation doesn't error** but produces an **unexpected output**.

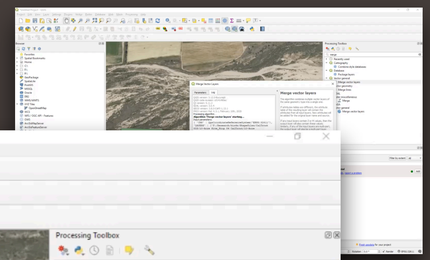
5. 

Creating informal program representations

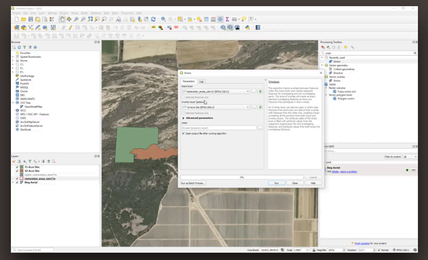
1. 

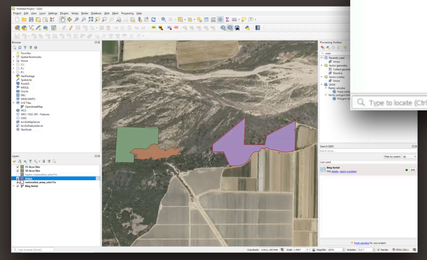
2. 

3. 

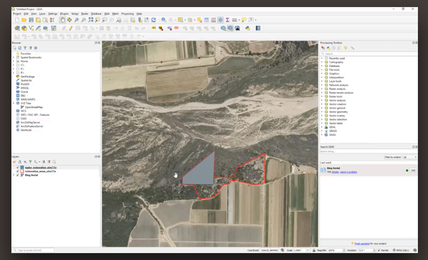
4. 

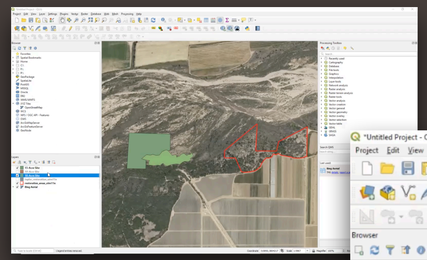
7. **Retry `Union` operation** but flip layer parameter order.

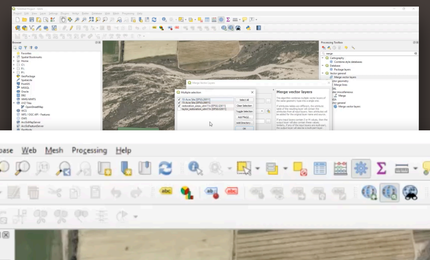
5. 

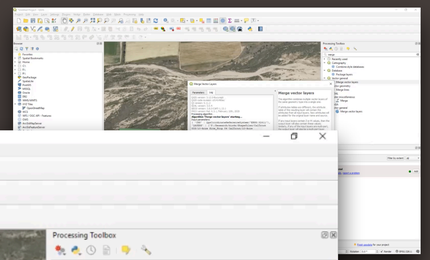
6. 

Creating informal program representations

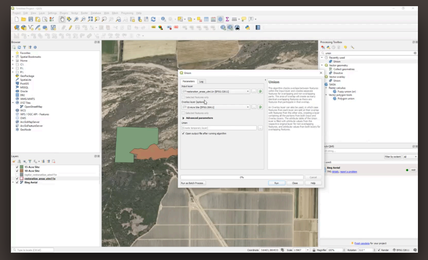
1. 

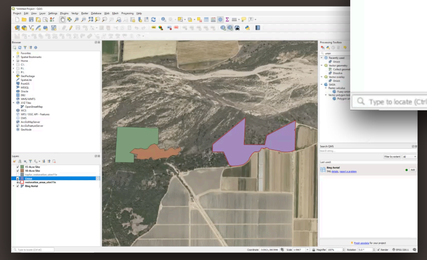
2. 

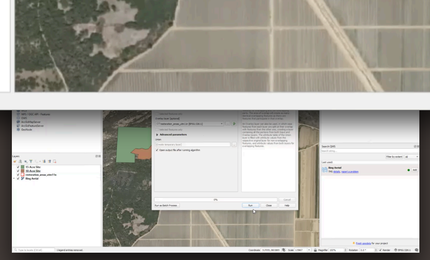
3. 

4. 

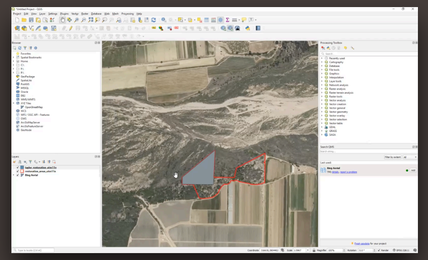
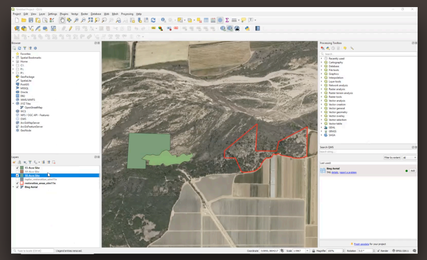
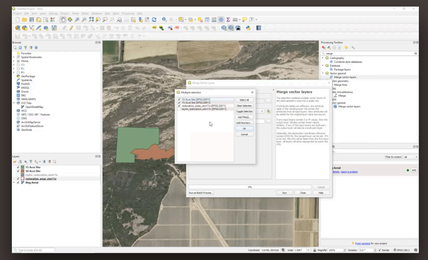
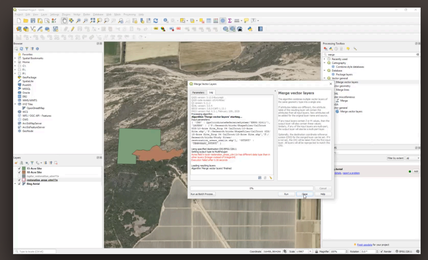
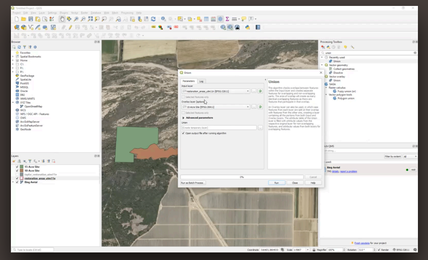
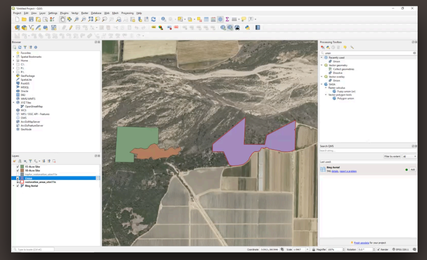
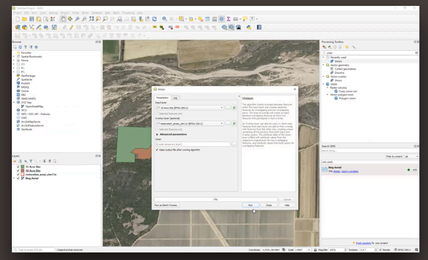
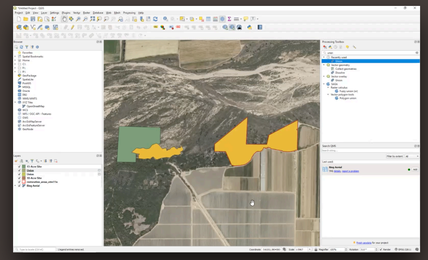
8. `Union` **operation** succeeds and gives expected output.

5. 

6. 

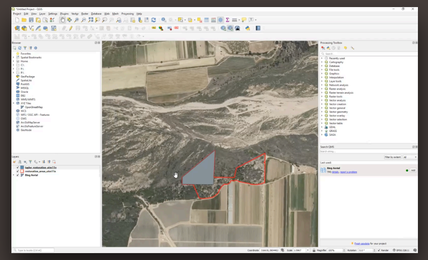
7. 

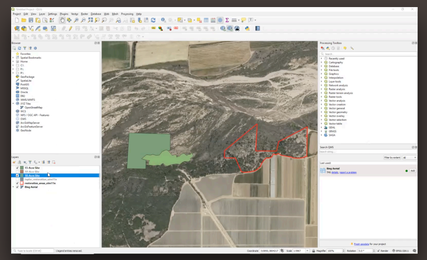
Creating informal program representations

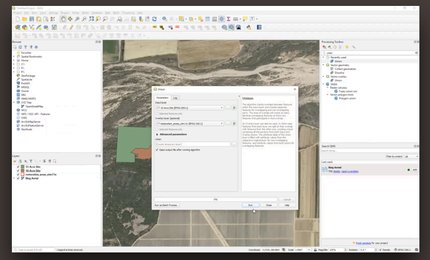
1. 
2. 
3. 
4. 
5. 
6. 
7. 
8. 

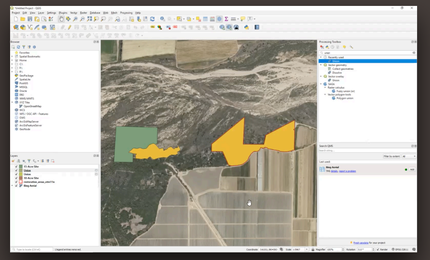
Full trace of user actions

Creating informal program representations

1. 

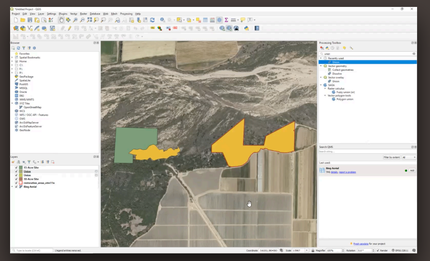
2. 

7. 

8. 

Full trace of desired program

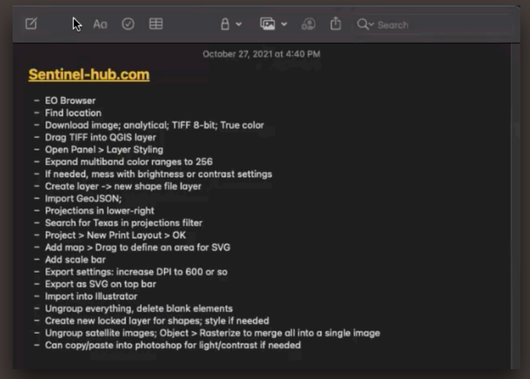
Creating informal program representations

8. 

What QGIS actually gives us

Creating informal program representations

Participants using GIS software persisted process information about their analysis **outside of GIS software**.

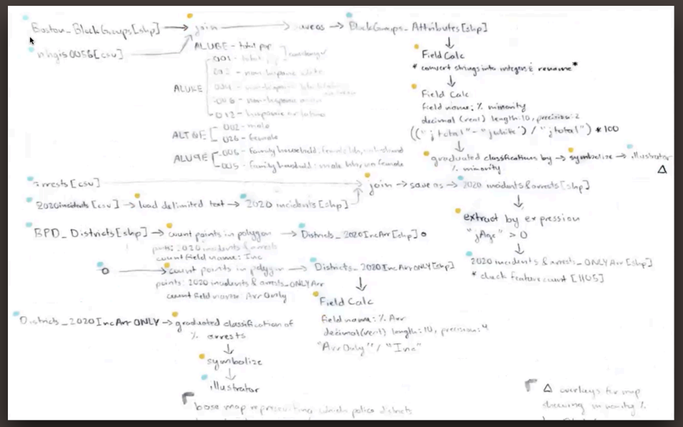


October 27, 2021 at 4:40 PM

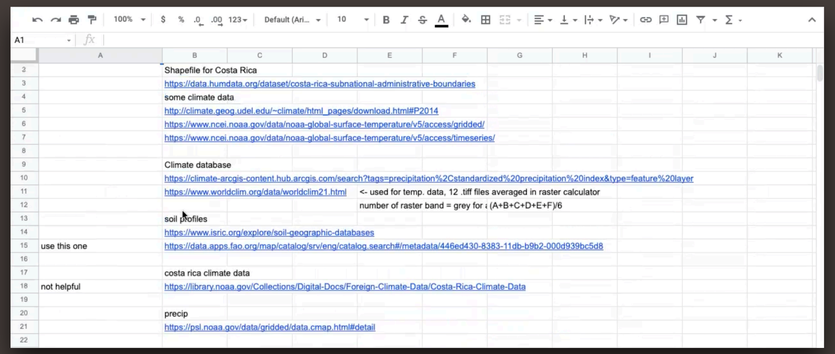
Sentinel-hub.com

- EO Browser
- Find location
- Download image; analytical; TIFF 8-bit; True color
- Drag TIFF into QGIS layer
- Open Panel > Layer Styling
- Expand multiband color ranges to 256
- If needed, mess with brightness or contrast settings
- Create layer -> new shape file layer
- Import GeoJSON;
- Projections in lower-right
- Search for Texas in projections filter
- Project > New Print Layout > OK
- Add map > Drag to define an area for SVG
- Add scale bar
- Export settings: increase DPI to 600 or so
- Export as SVG on top bar
- Import into Illustrator
- Ungroup everything, delete blank elements
- Create new locked layer for shapes; style if needed
- Ungroup satellite images; Object > Rasterize to merge all into a single image
- Can copy/paste into photoshop for light/contrast if needed

Participant 19 used macOS Notes.



Participant 4 used a hand-drawn diagram, with color distinguishing layers vs. operators.

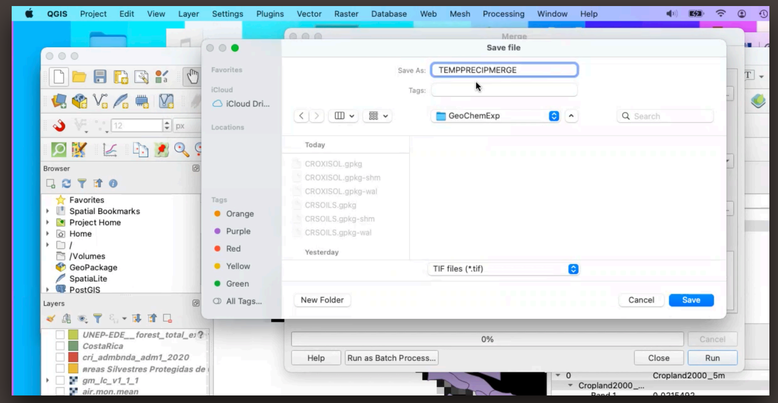


2		Shapefile for Costa Rica
3		https://data.humdata.org/dataset/costa-rica-subnational-administrative-boundaries
4		some climate data
5		http://climate.geog.udel.edu/~climate/html_pages/download.html#P2014
6		https://www.ncei.noaa.gov/data/noaa-global-surface-temperature/v5/access/gridded/
7		https://www.ncei.noaa.gov/data/noaa-global-surface-temperature/v5/access/timeseries/
8		
9		Climate database
10		https://climate-arcgis-content.hub.arcgis.com/search?tags=precipitation%2Catandardized%20precipitation%20index&type=feature%20layer
11		https://www.worldclim.org/data/worldclim21.html <- used for temp. data, 12 .tiff files averaged in raster calculator
12		number of raster band = grey for ((A+B+C+D+E+F)/6
13		soil profiles
14		https://www.isric.org/explore/soil-geographic-databases
15	use this one	https://data.apps.fao.org/map/catalog/srv/eng/catalog.search#/metadata/446ed430-8383-11db-b9b2-000d939bc5d8
16		
17		costa rica climate data
18	not helpful	https://library.noaa.gov/Collections/Digital-Docs/Foreign-Climate-Data/Costa-Rica-Climate-Data
19		
20		precip
21		https://psl.noaa.gov/data/gridded/data.cmap.html#detail
22		

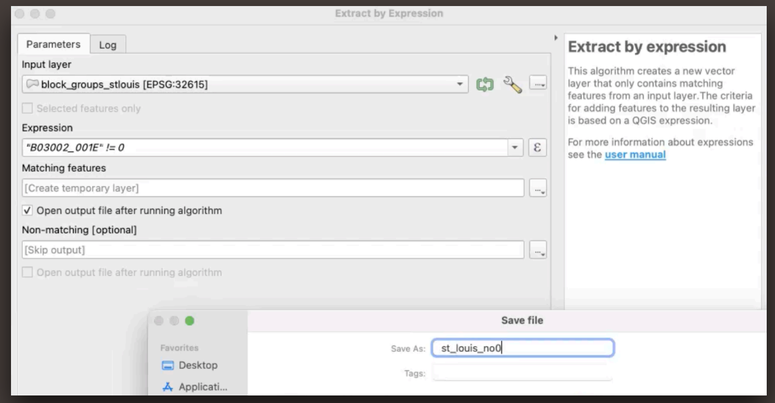
Participant 3 used Google Sheets, adding annotations on data source quality and expressions to use in the `Raster Calculator` tool in QGIS.

Creating informal program representations

Participants using GIS software left hints to their analysis steps in **layer names**.



Participant 3 added the suffix "`MERGE`" to the layer generated by merging temperature and precipitation layers.



Participant 4 added the suffix "`_no0`" when filtering out Census block groups with no population from his dataset.

Creating informal program representations

Why do GIS software users create informal representations of their processes?

Assists in **recall** of how the current program state was achieved

Assists in **manual reproduction** of the full analysis at a future time

Creating informal program representations

What are the **pitfalls** of working from these informal program representations?

Difficult to record a **sufficient level of detail** for precise reproduction

Not **independently verifiable** by other researchers or journalists

Creating informal program representations

> “I don’t do any of my processing in [QGIS], mainly because **I like that you can track what you did, the traceability of doing it in Python**. Versus there’s like **none of that if you do it in QGIS**. It’s like you use a plugin or a function, but there’s **no track record of it**.”

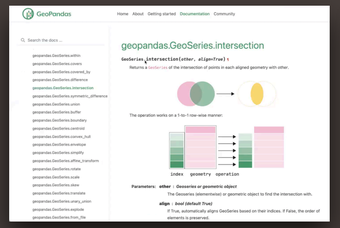
— Participant 17

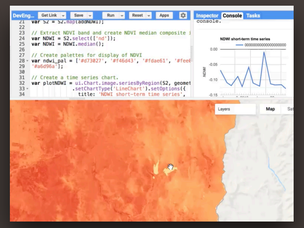
Creating informal program representations

> “If someone wants to go back and look at my code — ‘Oh, he got this shapefile from here and this shapefile from here and he's pushing them together.’ Whereas if you do that in Arc **you can't really replicate that workflow in the same way**.”

— Participant 15

Findings





Reasoning about geospatial operator behavior

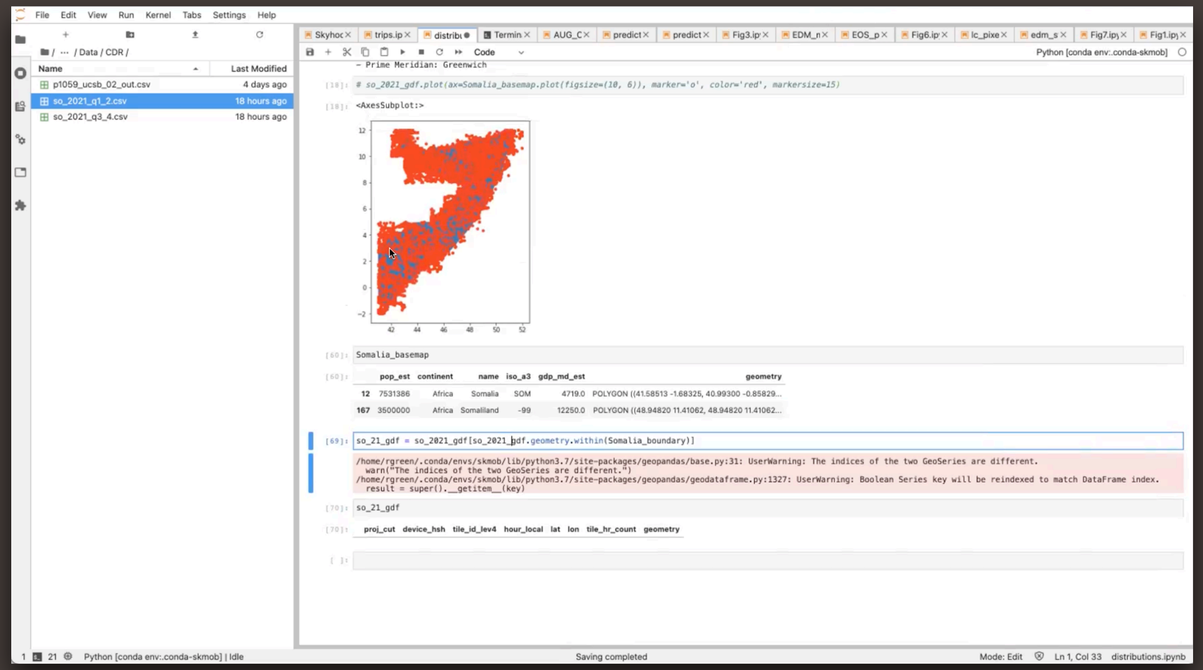
GIS Software

Programming Environments

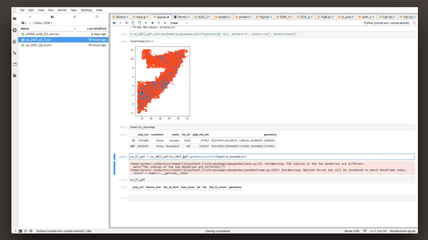
Geospatial tools have hundreds of operators, and **finding the right operator for an analysis context is difficult.**

Reasoning about geospatial operator behavior

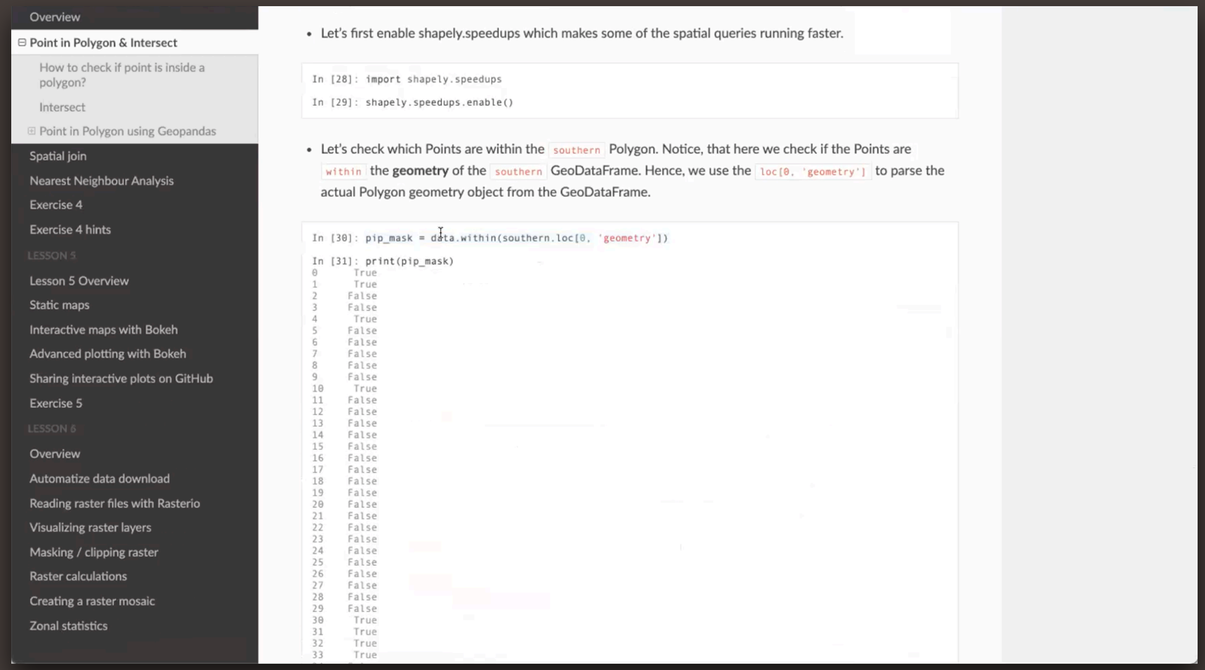
1. “I want to subset to just the points that are within the polygon.”



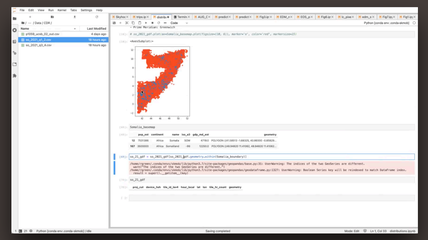
Reasoning about geospatial operator behavior

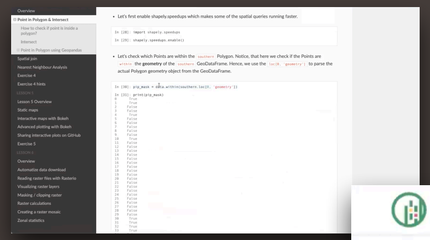
1. 

2. Copy `.within` example code from a blog post.

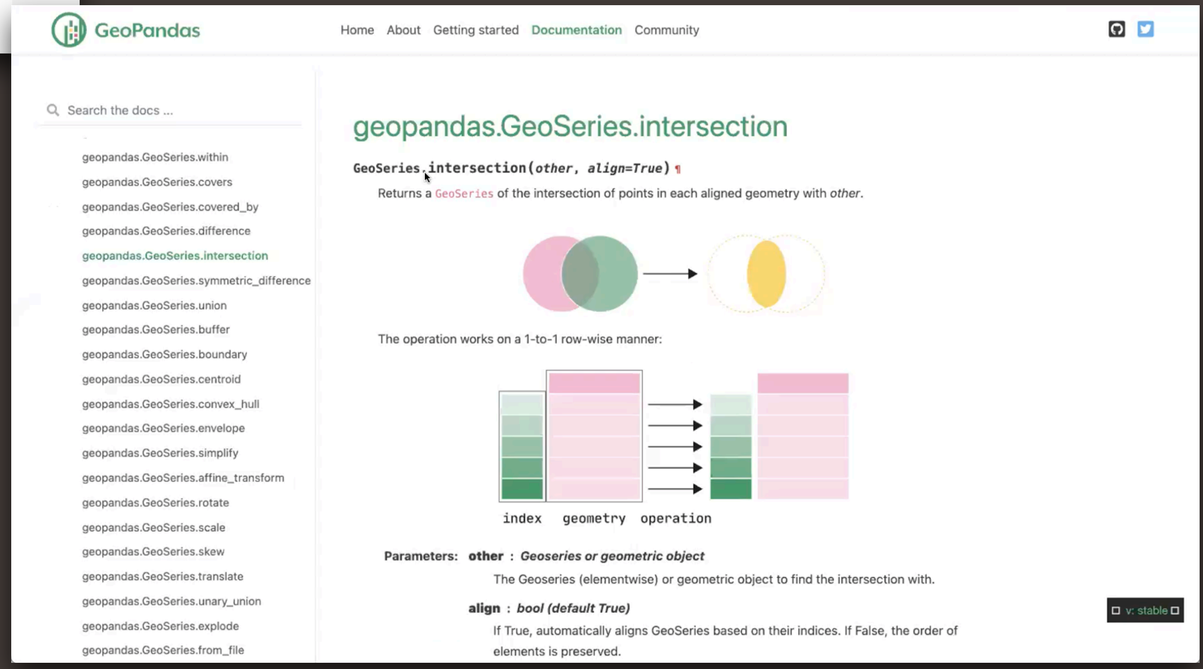


Reasoning about geospatial operator behavior

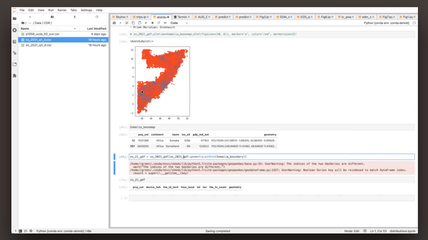
1. 

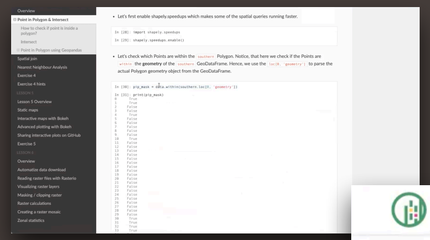
2. 

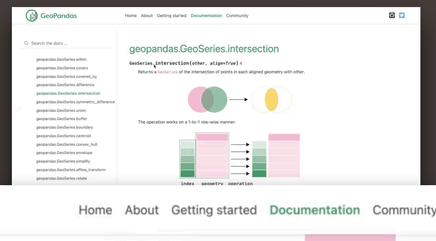
3. When that fails, check the `geopandas` API documentation for `.intersection`.



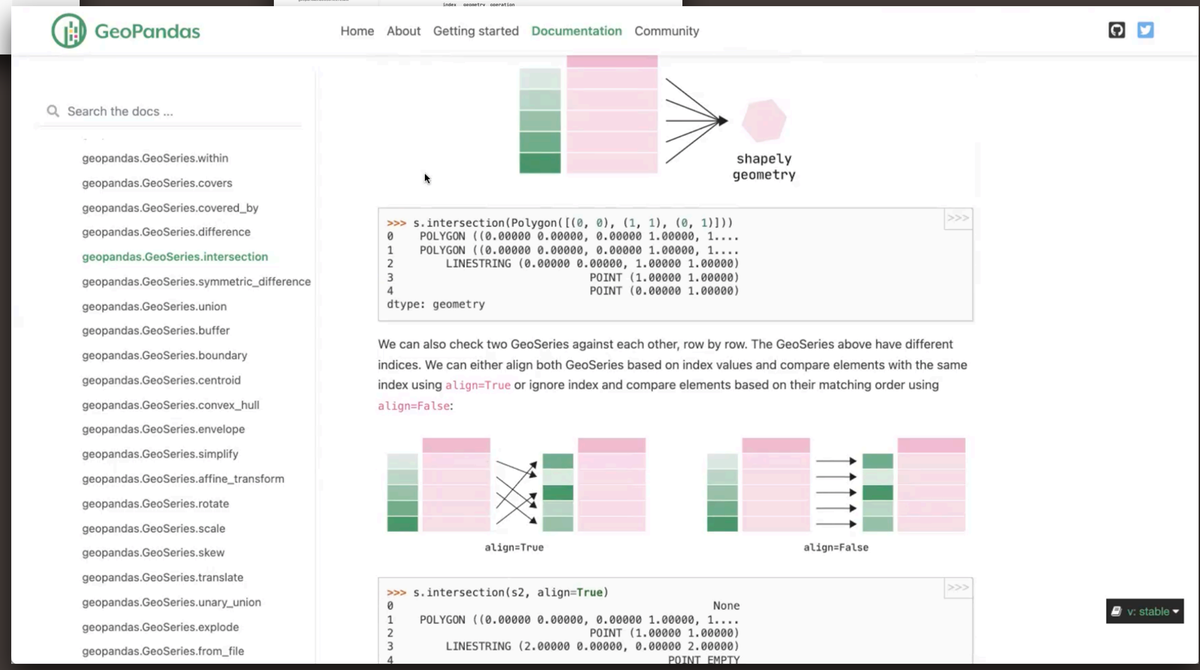
Reasoning about geospatial operator behavior

1. 

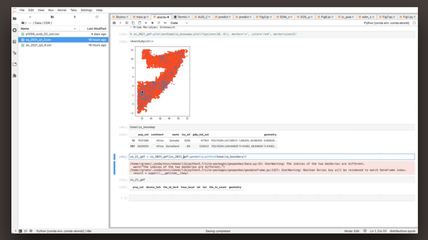
2. 

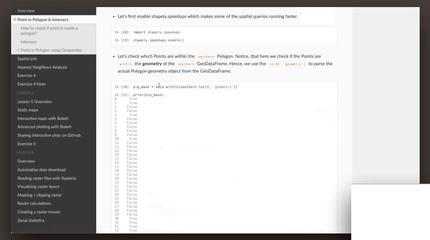
3. 

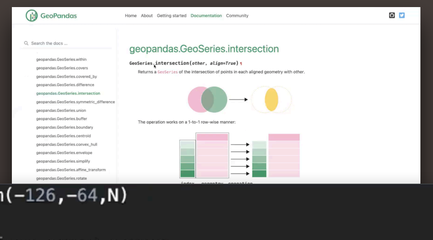
4. Get confused by the **geopandas** API docs' use of tables in lieu of geometries.

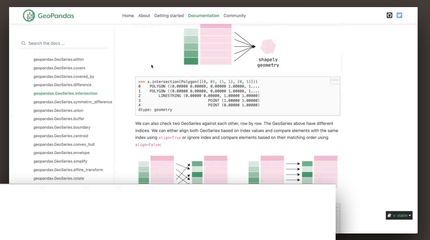


Reasoning about geospatial operator behavior

1. 

2. 

3. 

4. 

5. Copy `.sjoin` example code from a different blog post.

```
o.random.uniform(-126,-64,N)

e geodataframe from numpy arrays
.DataFrame({'lon':lon, 'lat':lat})
rds'] = list(zip(df['lon'],df['lat']))
rds'] = df['coords'].apply(Point)
= gpd.GeoDataFrame(df, geometry='coords', crs=counties.crs)

rm spatial join to match points and polygons
Polys = gpd.tools.sjoin(points, counties, op="within", how='left')
```

That's it, we now have matched points and polygons. For example use in how to retrieve the points from a specific polygon, we can continue as follows:

```
# Example use: get points in Los Angeles, CA.
pnt_LA = points[pointInPolys.NAME=='Los Angeles']

# Plot map with points in LA in red
base = counties.boundary.plot(linewidth=1, edgecolor="black")
points.plot(ax=base, linewidth=1, color="blue", markersize=1)
pnt_LA.plot(ax=base, linewidth=1, color="red", markersize=8)
plt.show()
```

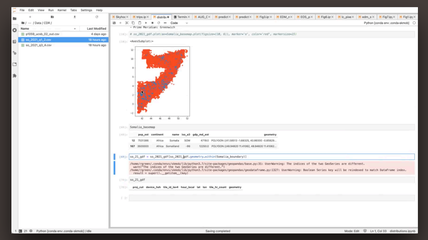
This will produce the following image:

On this page:

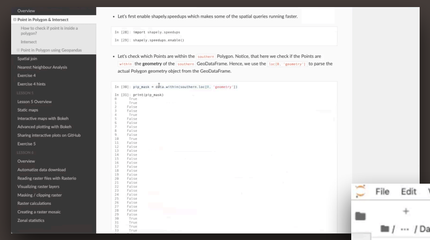
- How to check if a point is inside a polygon in Python
- How to get a list of points inside a polygon in Python
- Speeding up Geopandas point-in-polygon tests
  - Use Rtree
  - Use the optimized PyGEOS library
- Final remarks

Reasoning about geospatial operator behavior

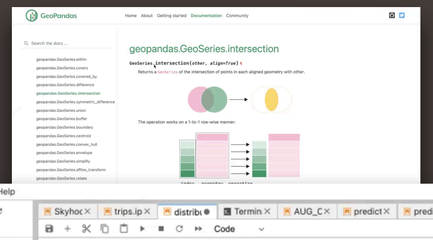
1.



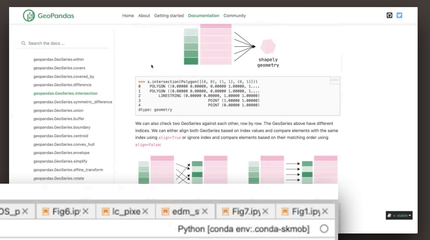
2.



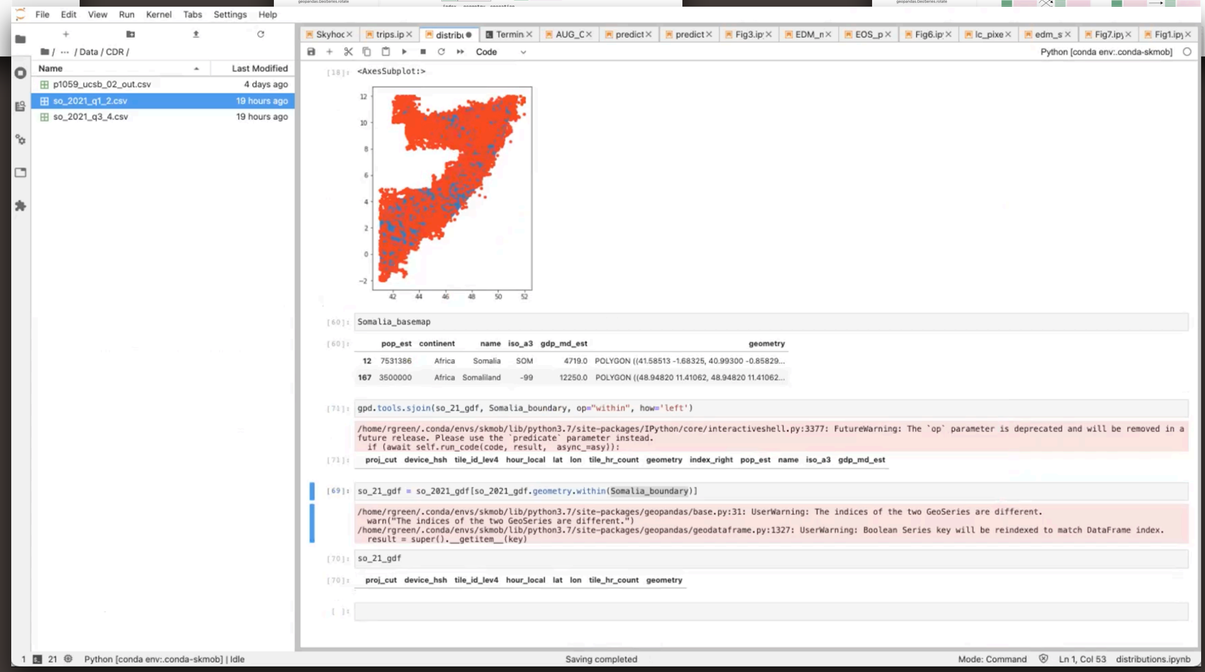
3.



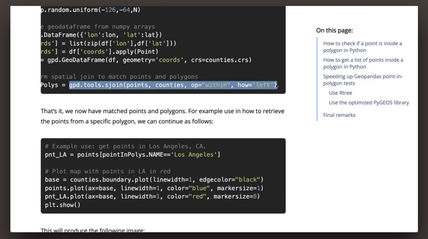
4.



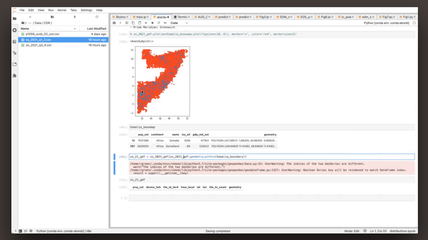
6. Adapt copied `.sjoin` for the notebook's context. Notice a runtime warning.

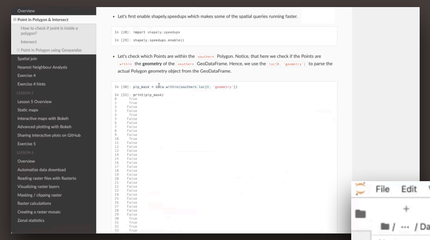


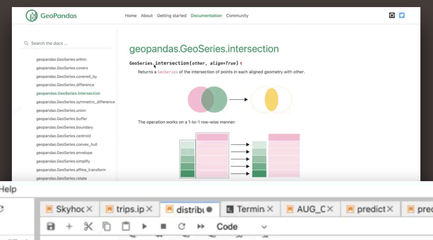
5.

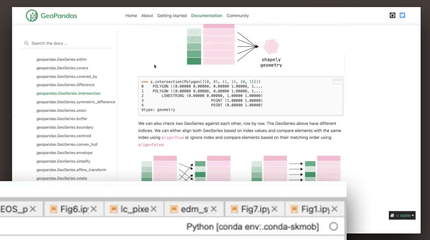


Reasoning about geospatial operator behavior

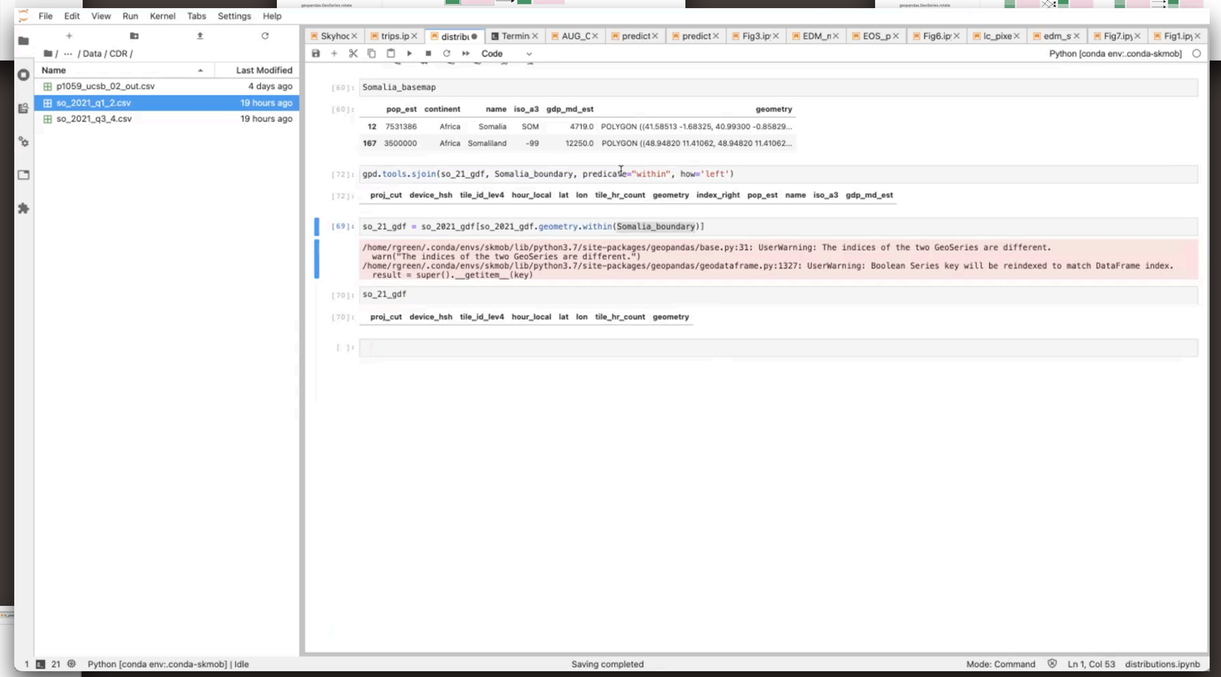
1. 

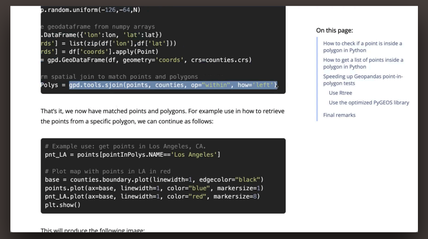
2. 

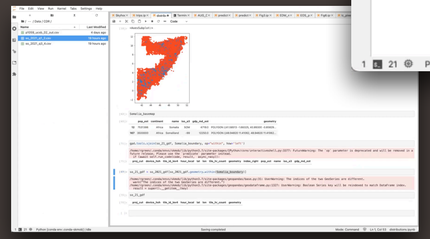
3. 

4. 

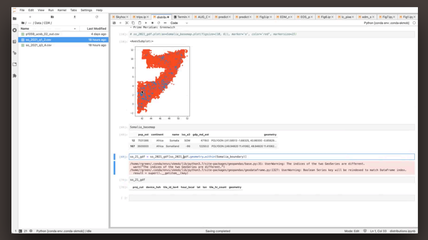
7. Change `op` to `predicate` to fix warning. The result is an empty `GeoDataFrame` 😱.



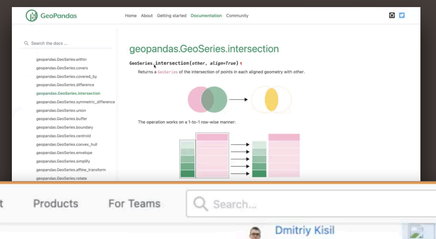
5. 

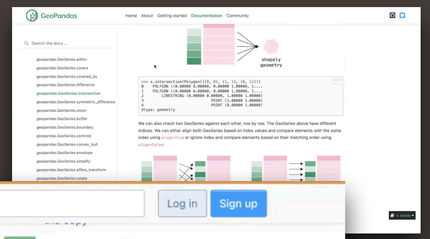
6. 

Reasoning about geospatial operator behavior

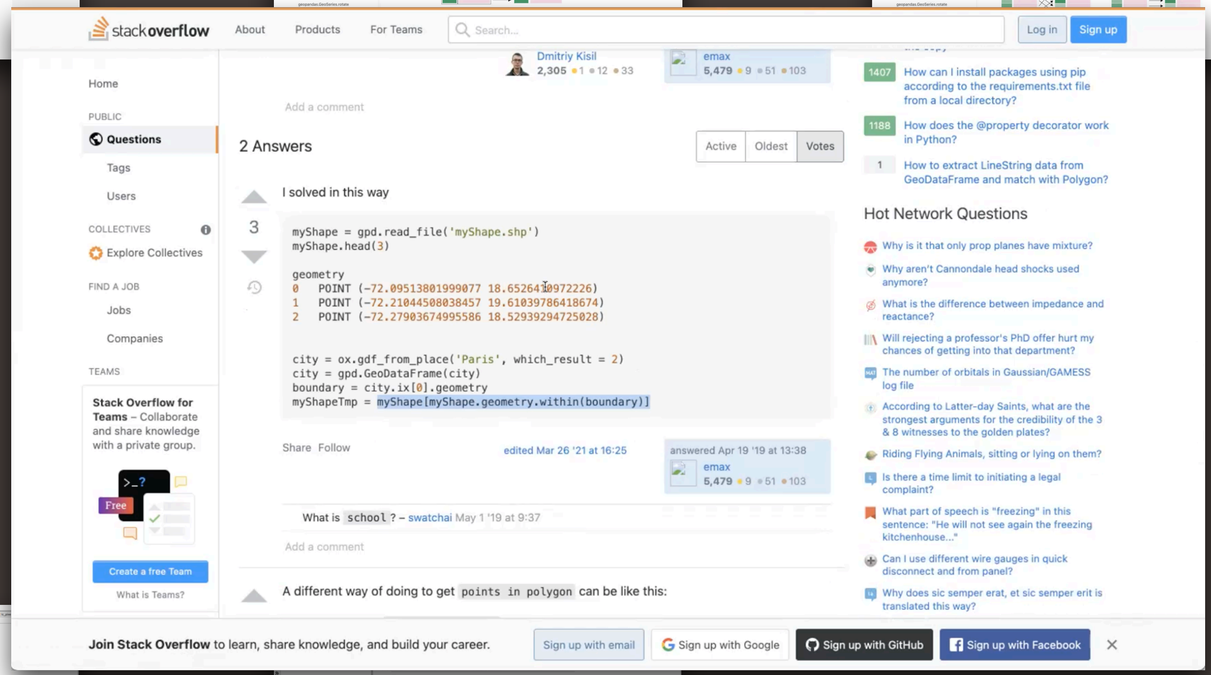
1. 

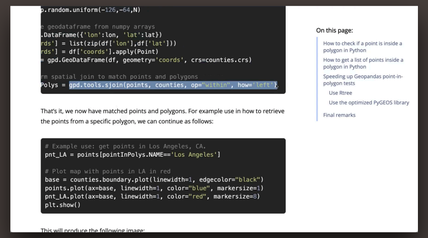
2. 

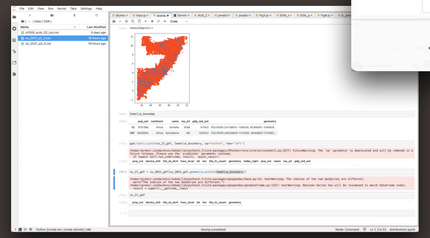
3. 

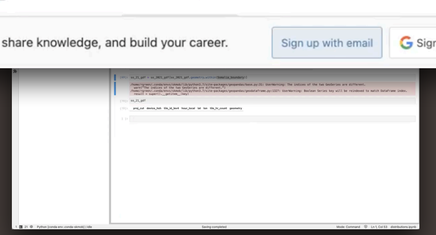
4. 

8. Go to StackOverflow post. “I feel really silly. I waste a lot of time doing things like this.”

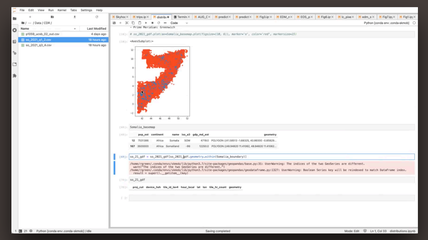


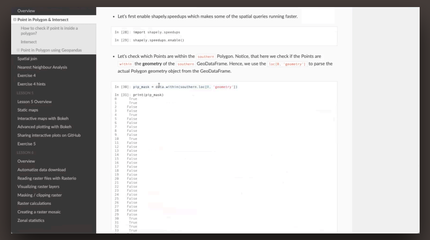
5. 

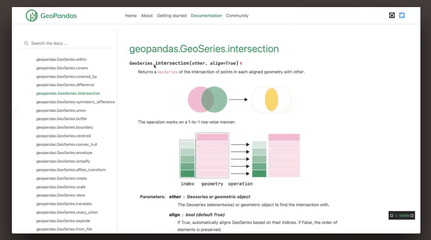
6. 

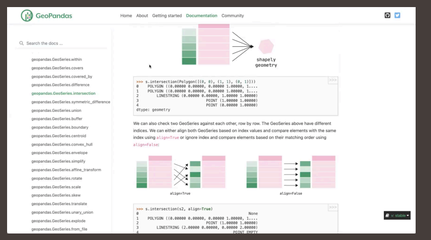
7. 

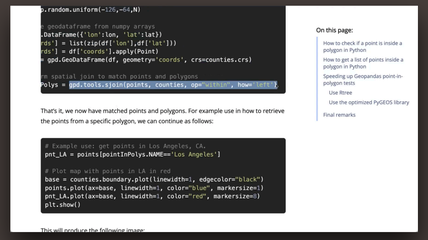
Reasoning about geospatial operator behavior

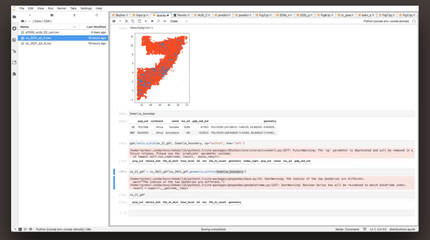
1. 

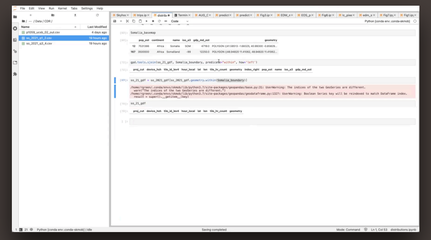
2. 

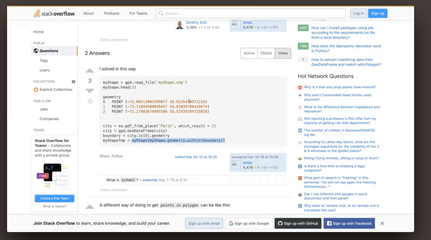
3. 

4. 

5. 

6. 

7. 

8. 

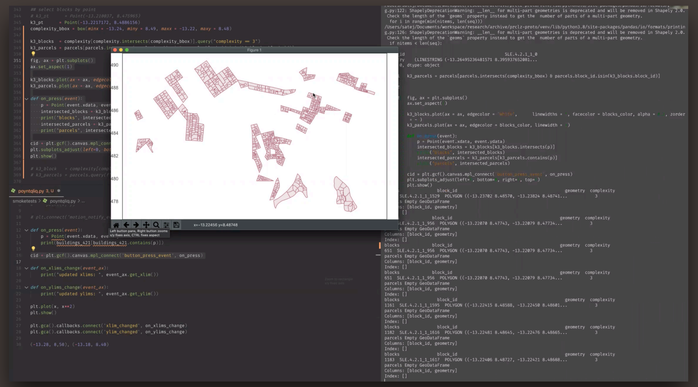
The above sequence took **6 minutes**.

We continued for **another 12 minutes** before arriving at the desired output.

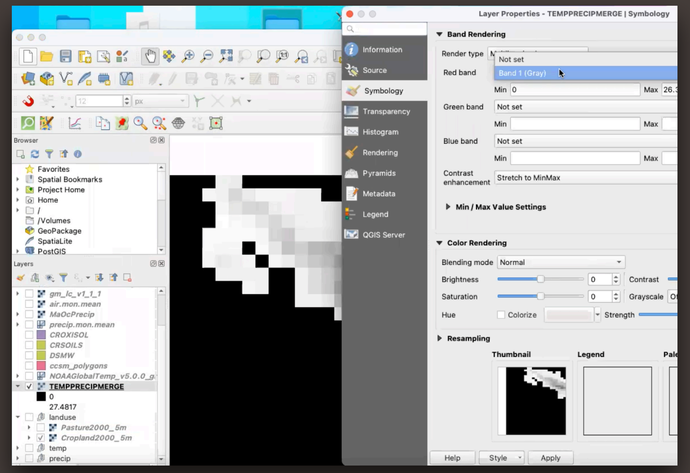
```
so_2021_gdf[so_2021_gdf.geometry.within(Somalia_boundary.loc[0, 'geometry'])]
```

Reasoning about geospatial operator behavior

Participants understood an operator's behavior **by running it and inspecting the output**.



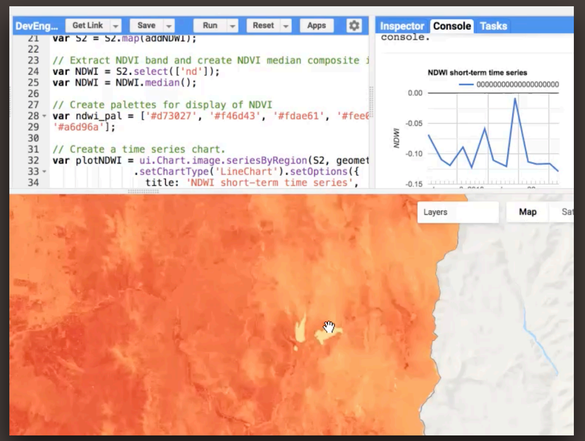
Participant 11 tried using `geopandas'.contains` API to see if it would work on point-within-closed-line operations.



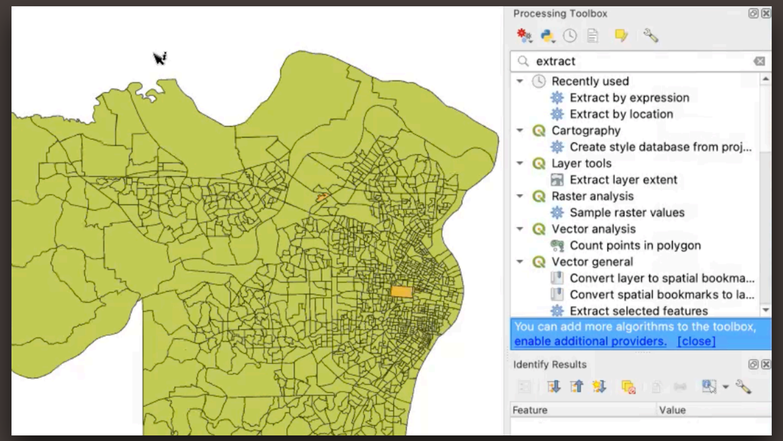
Participant 3 executed a `Merge` on their temperature and precipitation rasters and learned that the operator, by default, combines all input rasters into a single band rather than preserving multiple bands.

Reasoning about geospatial operator behavior

Participants use specific features or pixels to **validate operator semantics**.



Participant 1 checked the pixel values of a lake in their normalized difference water index (NDWI) raster to validate the algorithm's results.



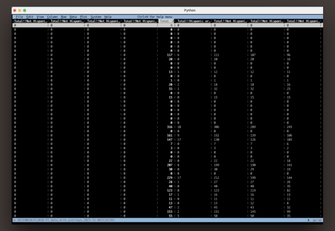
Participant 4 used color to validate that two features that should have been removed by `Extract by Expression` were indeed removed.

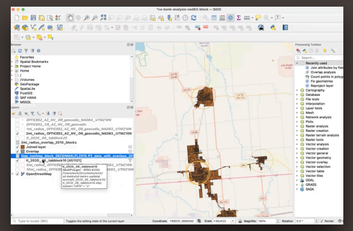
Reasoning about geospatial operator behavior

> “I can never remember the vector operations. There’s like Union and Merge. Combine! **I can never remember exactly what they do**. I know exactly what the output should look like in the end, **I’m just trying to figure out the tool that gets me that output**.”

— Participant 15

Findings





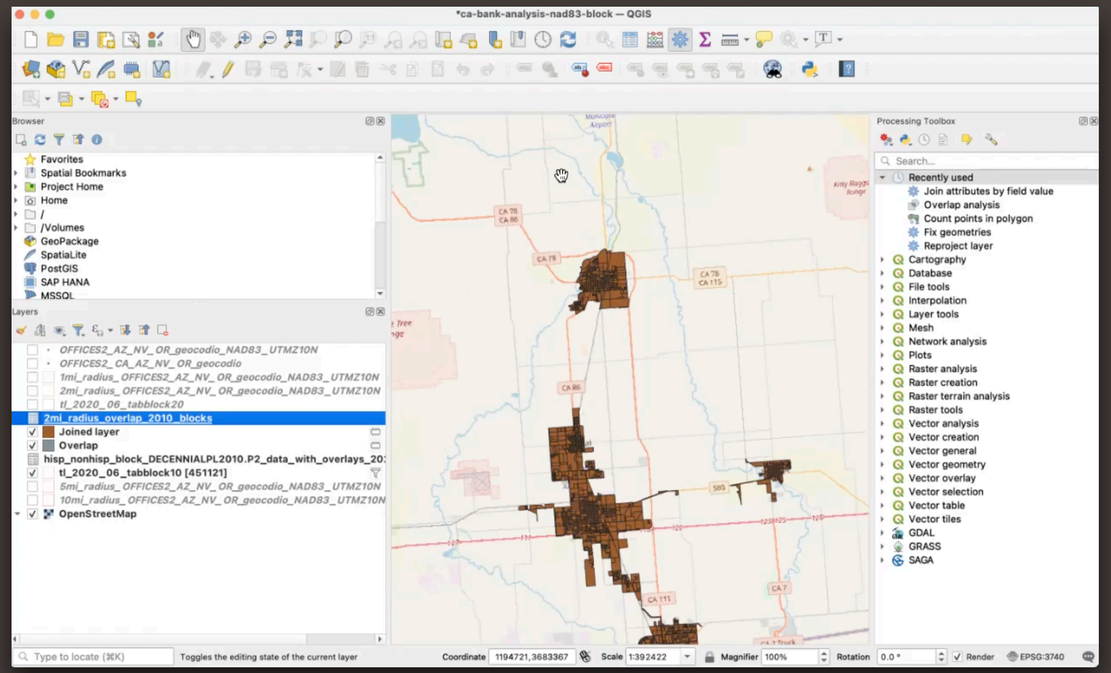
Lack of spatial visibility in code-based tools

Programming Environments

Code-based tools represent geospatial data as tables, **but users rely heavily on visual spatial reasoning during data exploration.**

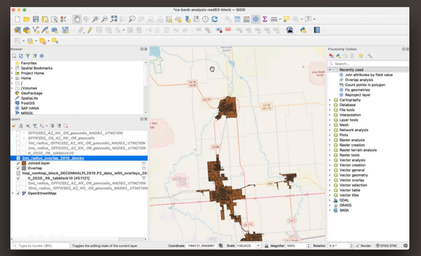
Lack of spatial visibility in code-based tools

1. Load TIGER/Line shapefile of urban Census block groups in Imperial County into QGIS.

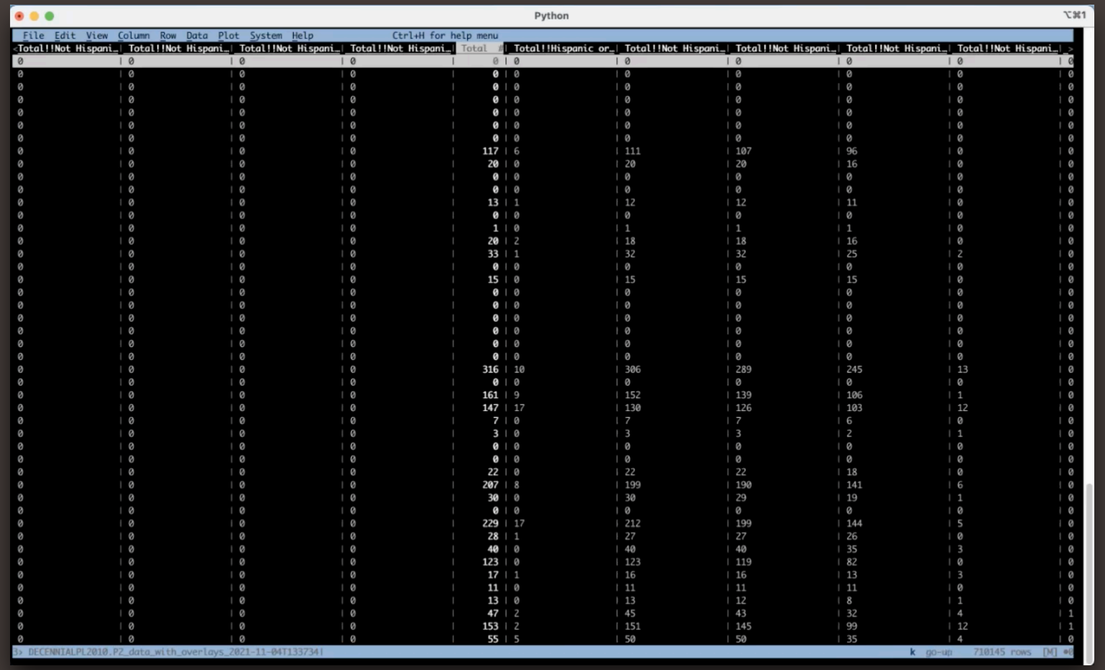


Lack of spatial visibility in code-based tools

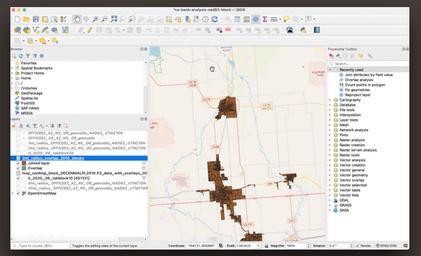
1.

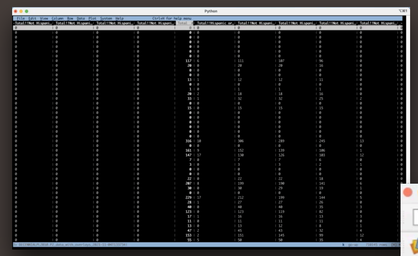


2. Move to **`VisiData`**, a CLI tool for data wrangling, to **`Join`** 2010 Census data to block groups.

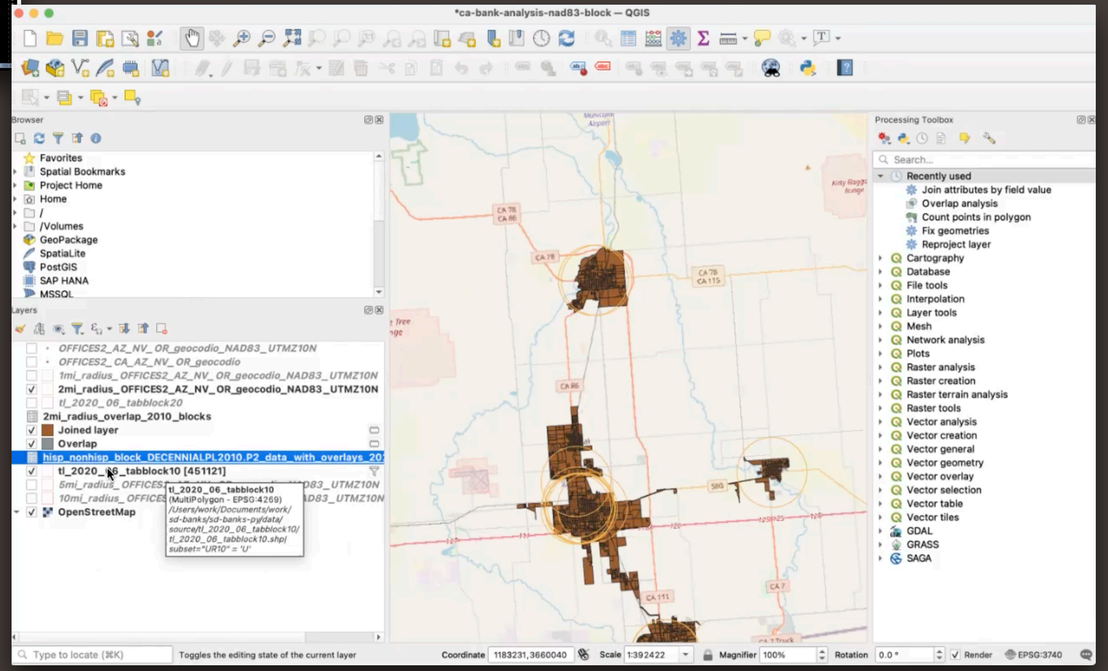


Lack of spatial visibility in code-based tools

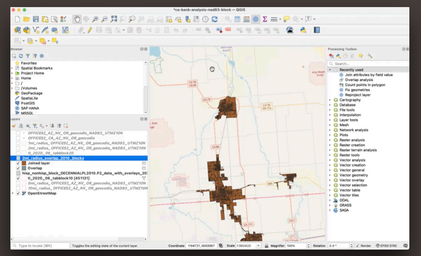
1. 

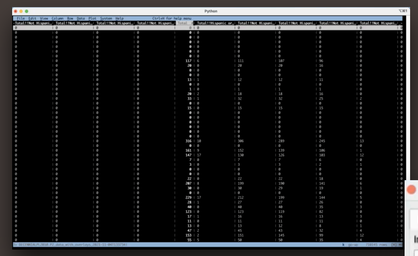
2. 

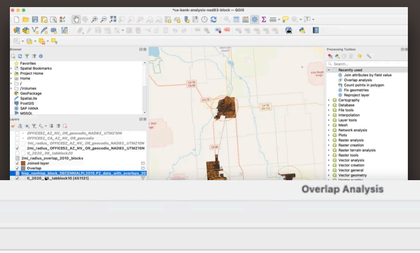
3. Move back to QGIS and generate 2-mile buffers around the point locations of banks.



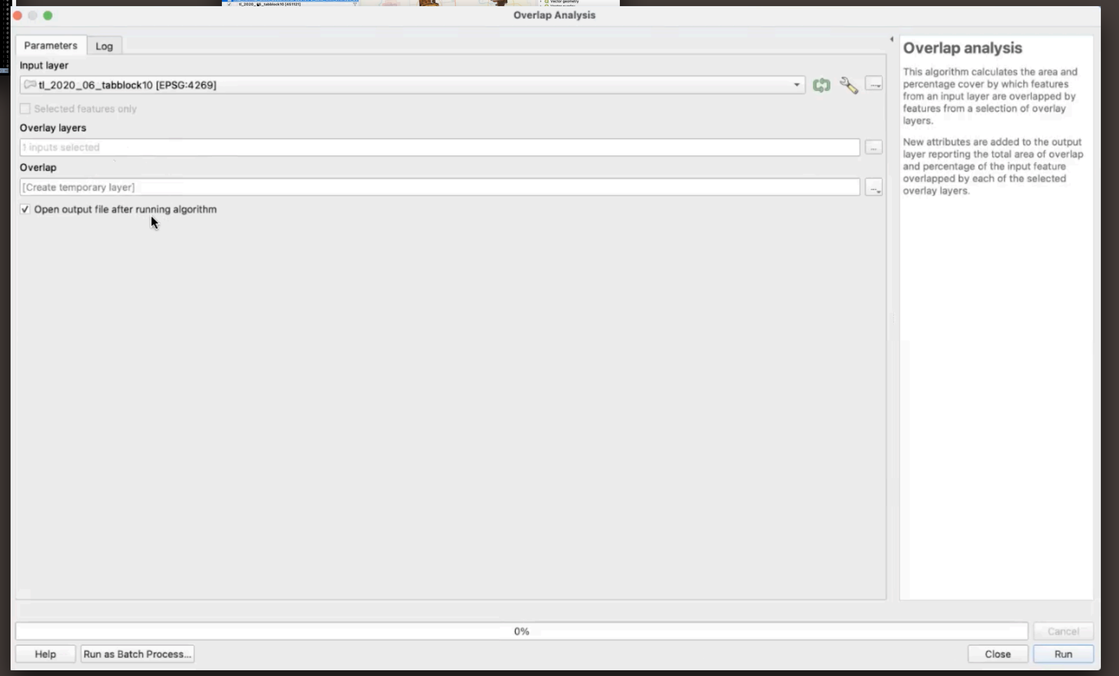
Lack of spatial visibility in code-based tools

1. 

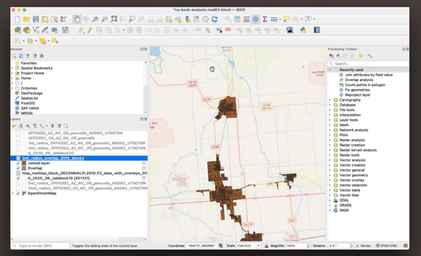
2. 

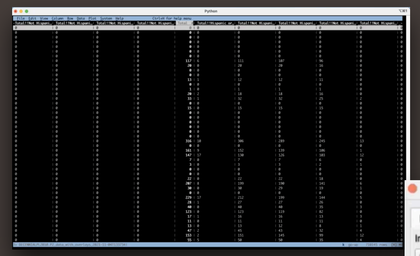
3. 

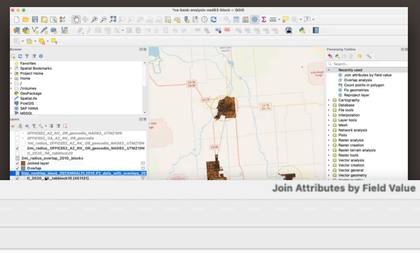
4. Use the `Overlap analysis` operator to compute the area overlap between each block group and any buffer.

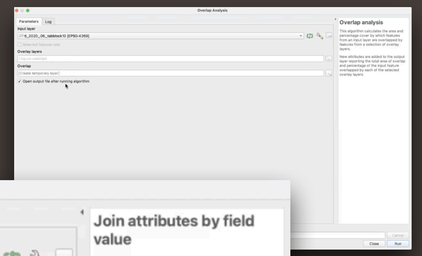


Lack of spatial visibility in code-based tools

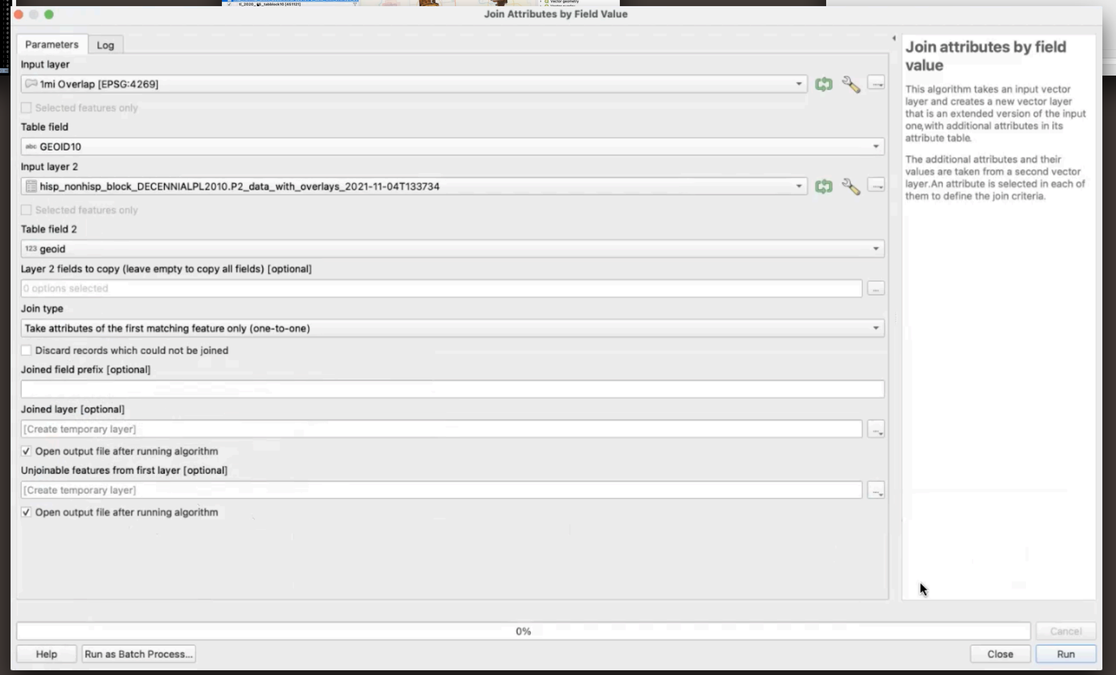
1. 

2. 

3. 

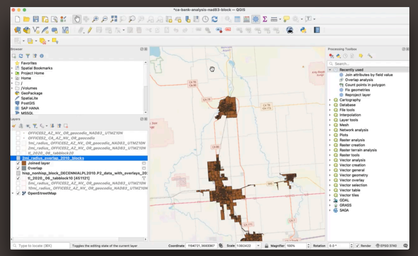
4. 

5. Use the `Join Attributes By Field Value` operator to join the areal overlay back to the 2010 Census data.

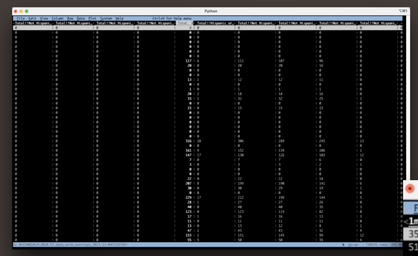


Lack of spatial visibility in code-based tools

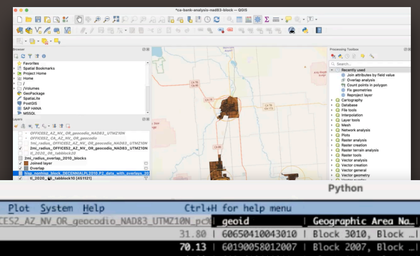
1.



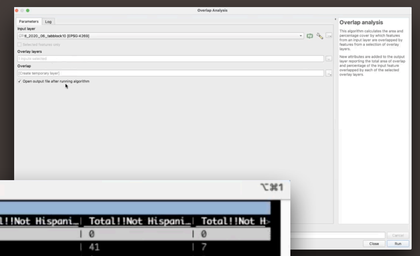
2.



3.

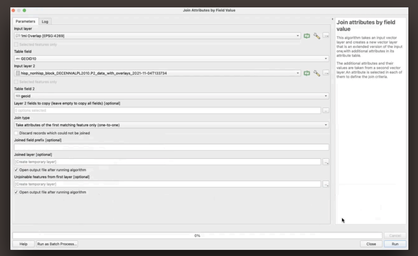


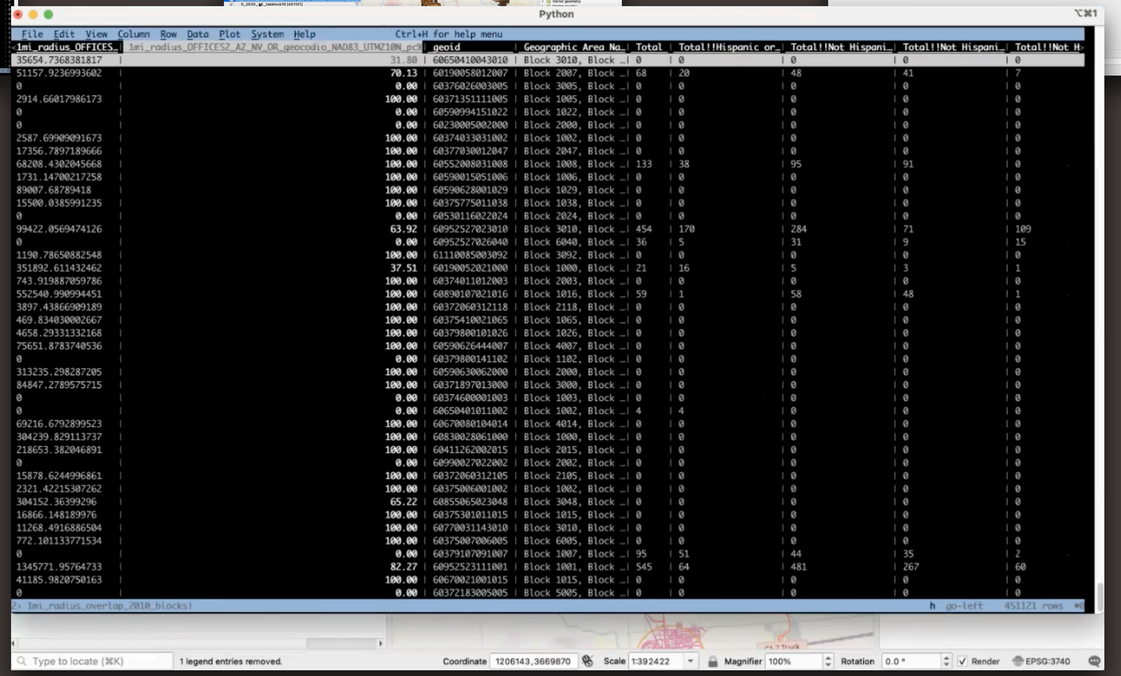
4.



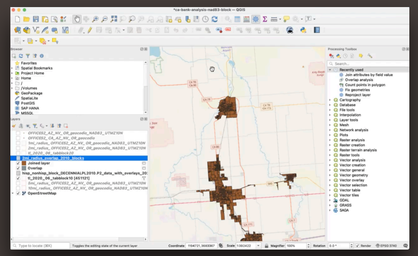
6. Move back to `VisiData` to multiply % overlay with bank buffers by total population for each block group.

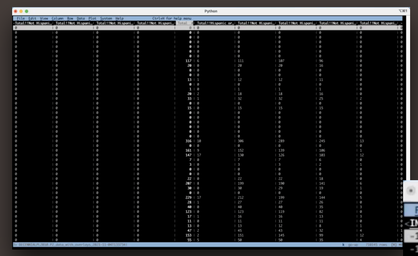
5.

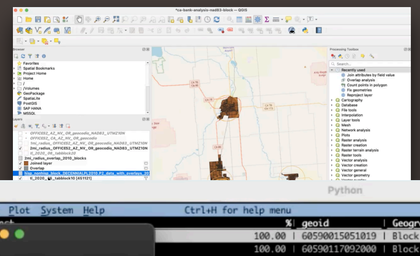


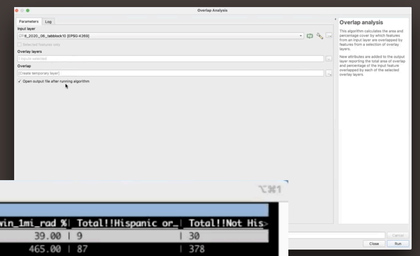


Lack of spatial visibility in code-based tools

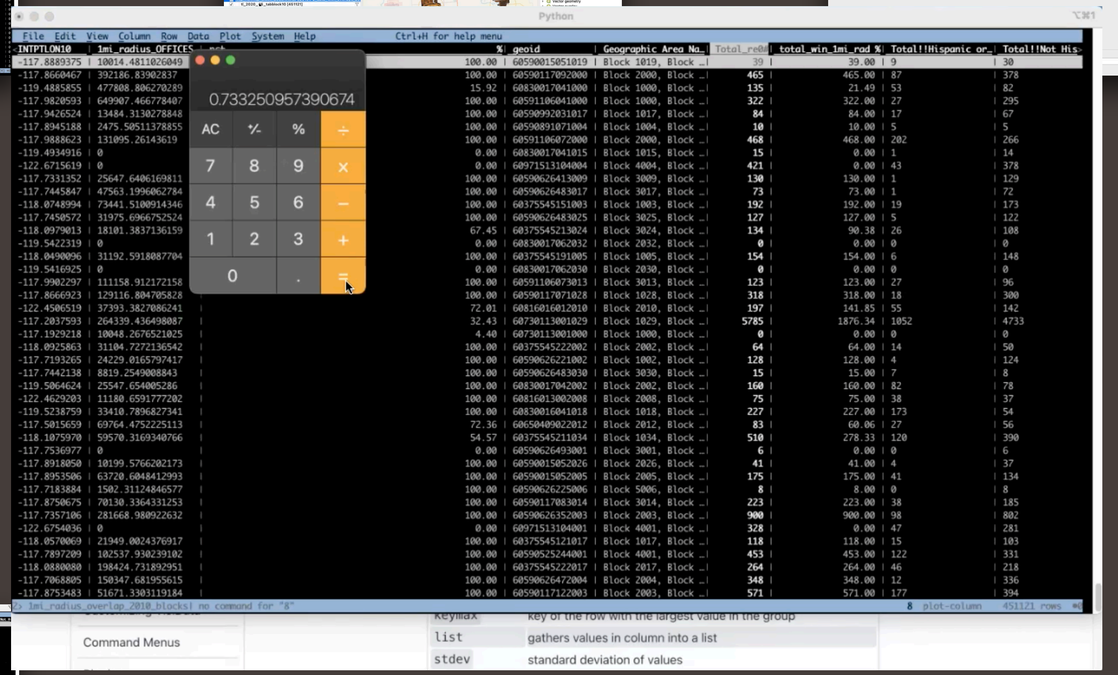
1. 

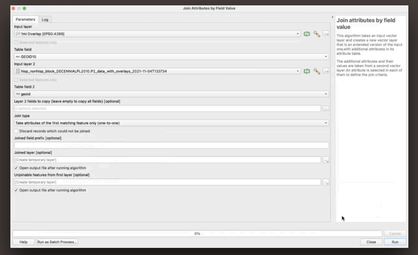
2. 

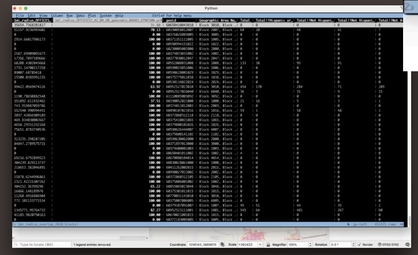
3. 

4. 

7. Use the macOS Calculator to compute the percentage of Imperial County residents living in a banking desert.

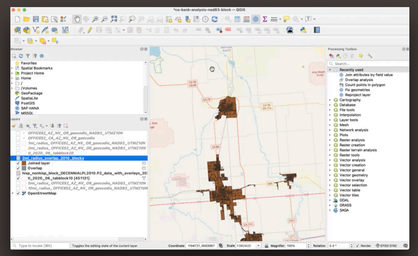


5. 

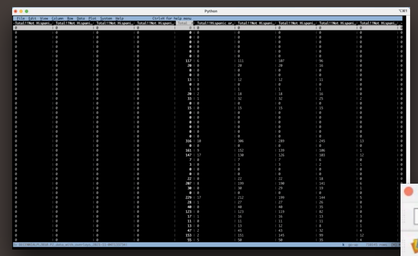
6. 

Lack of spatial visibility in code-based tools

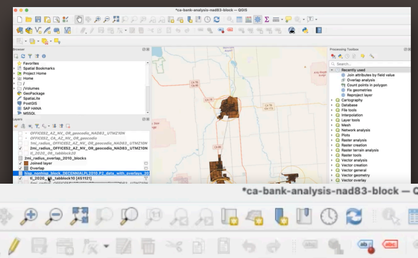
1.



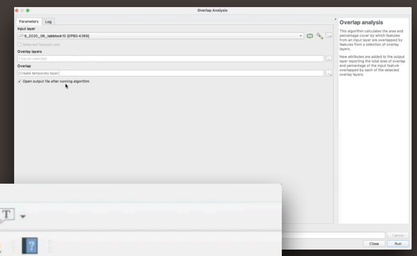
2.



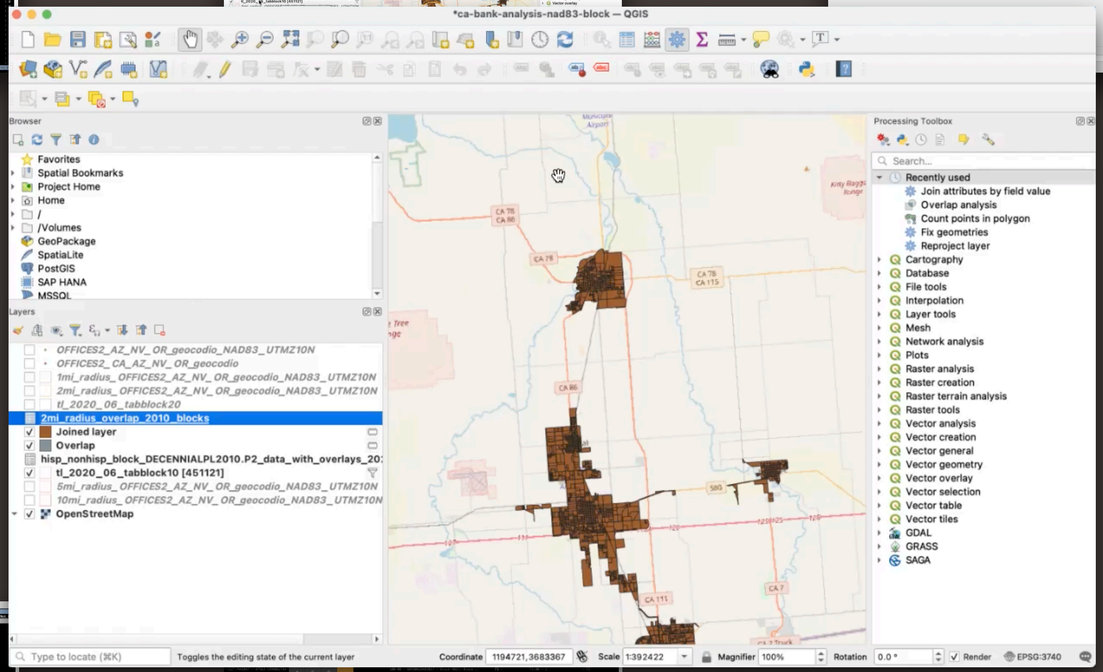
3.



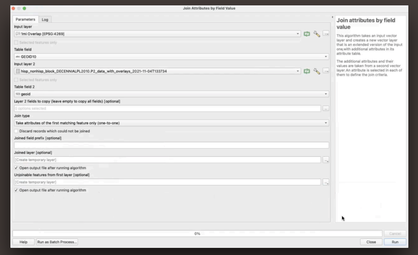
4.



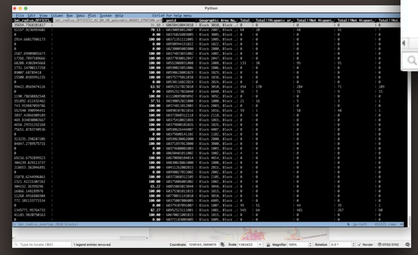
8. Move back to QGIS to repeat the analysis using 1-mile buffers.



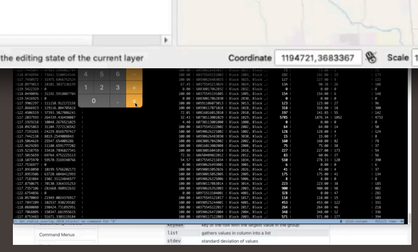
5.



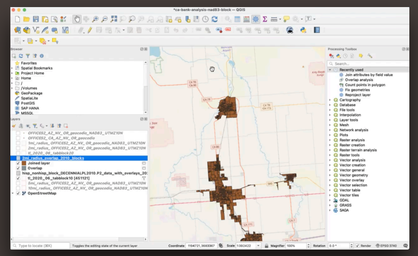
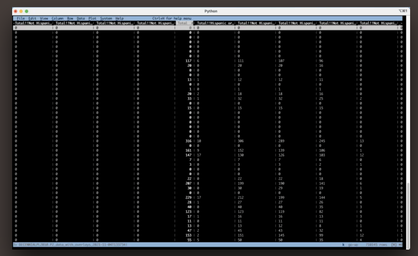
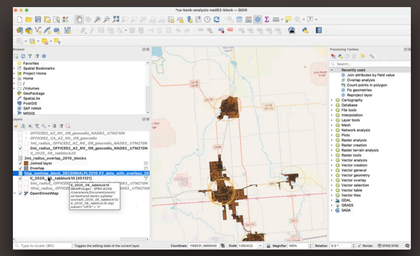
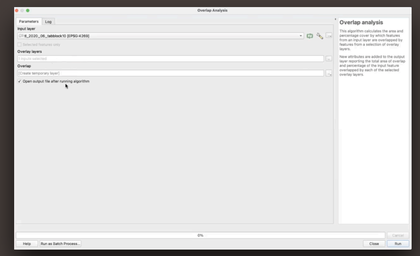
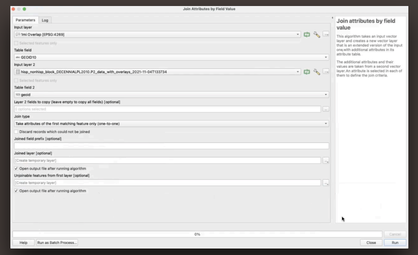
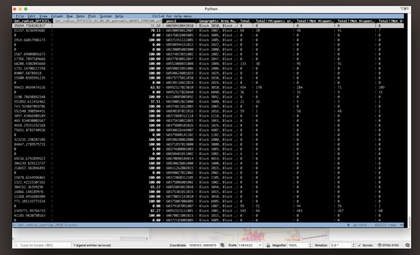
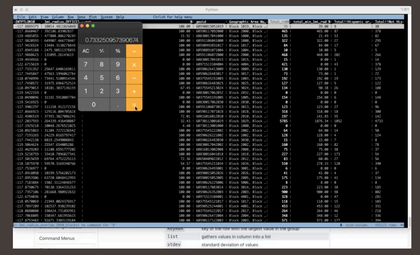
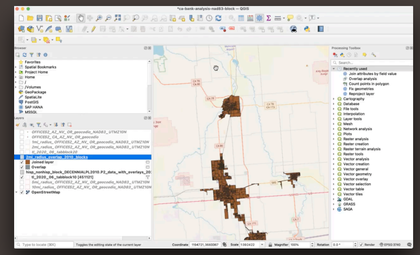
6.



7.



Lack of spatial visibility in code-based tools

1. 
2. 
3. 
4. 
5. 
6. 
7. 
8. 

Participant 9 oscillates between a programming environment and GIS software, **moving to QGIS for all geometric operations**.

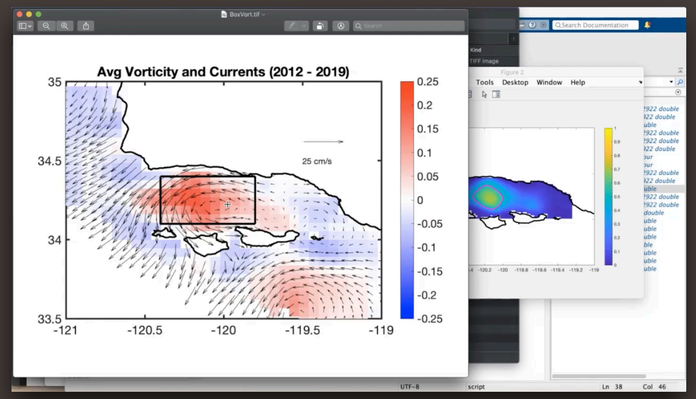
Lack of spatial visibility in code-based tools

> “I'm working in QGIS. I know that it’s slower than it would be to do it in PostGIS or maybe even `geopandas` and so I've considered switching to that. But I'm still sort of new enough that **I need to kind of ‘see’** to make sure my projections are right and stuff like that.”

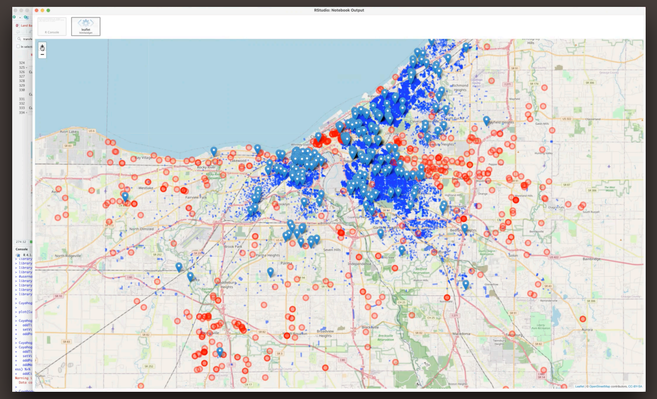
— Participant 9

Lack of spatial visibility in code-based tools

Participants rely heavily on **visual overlays of layers** to investigate spatial correlations, and **recreate this functionality in programming environments**.



Participant 16 compared an image of a plot against his Matlab figure to “eyeball” the spatial correlation between two variables of interest.



Participant 6 wrote an 8-line function in R to render three layers on a Leaflet map, allowing her to inspect the spatial overlap between them.

Lack of spatial visibility in code-based tools

> “Now I *could* try to visualize it here with `matplotlib` and `geopandas`, but I know those things are shitty, shitty, shitty and **not interactive** and so I'm like, **‘I gotta take this to QGIS.’**”

— Participant 18

Roadmap


1. What is geospatial data? What is GIS?
2. Motivation
3. Research Questions
4. Study Design
5. Findings and Discussion

Roadmap

Highlights

1. The importance of **informal program representations** to the working process of GIS software users
2. The challenge of **reasoning about the behavior of geospatial operators**
3. The **lack of spatial visibility in code-based tools**