Topography of Mt. Tamalpais, Marin County, USA

A Need-Finding Study with Users of Geospatial Data


Parker Ziegler


Sarah E. Chasins

Check out the CHI '23 paper

EPIC Data Lab Retreat • April 17, 2023

Berkeley
UNIVERSITY OF CALIFORNIA

EPIC DATA lab

Ok, but hold up, Parker. What is geospatial data?

(And why study how domain experts work with it?)

Background

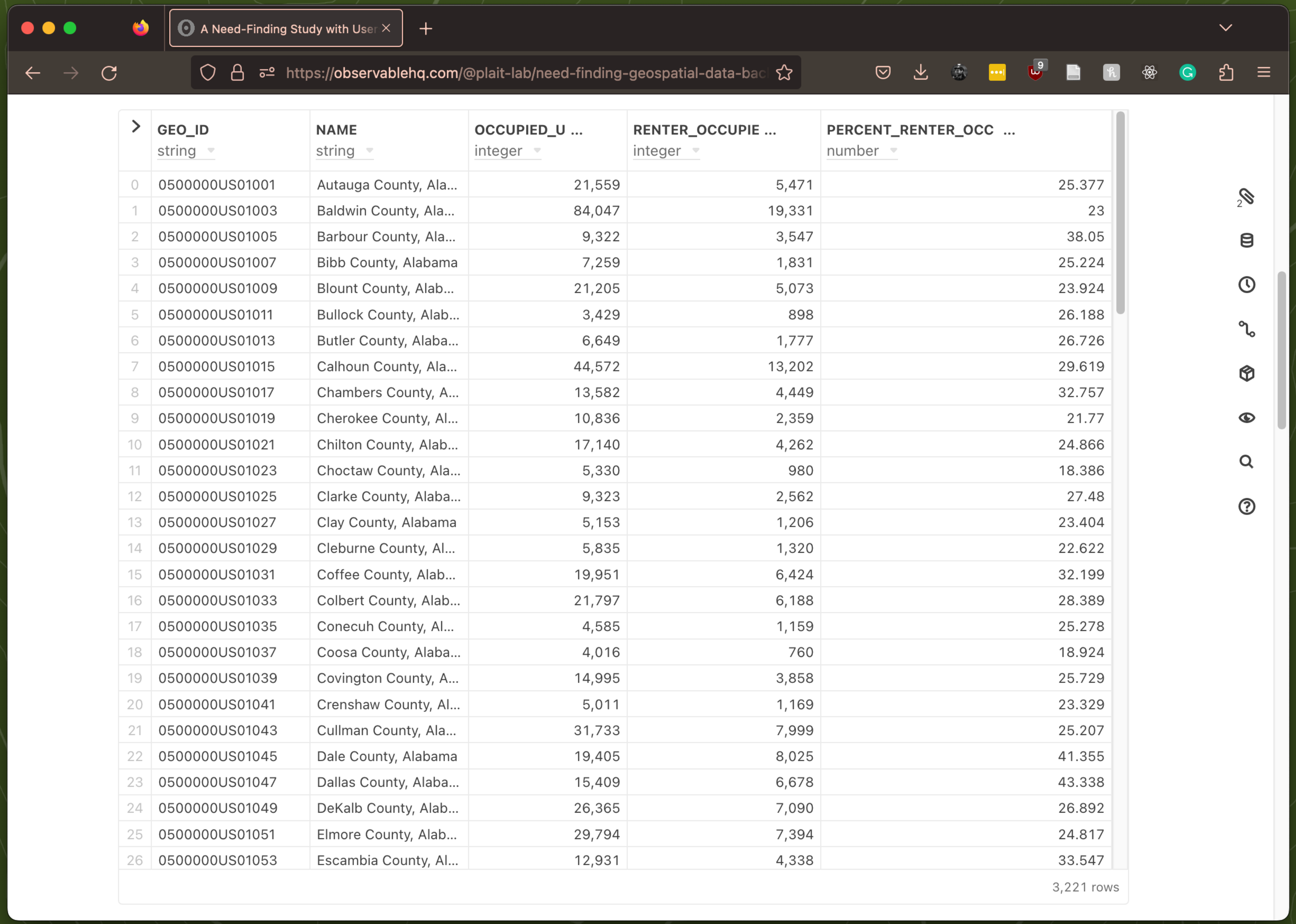
Geospatial Data

Geospatial data describes the **location** and **attributes** of phenomena on the Earth's surface.

Background

Geospatial Data

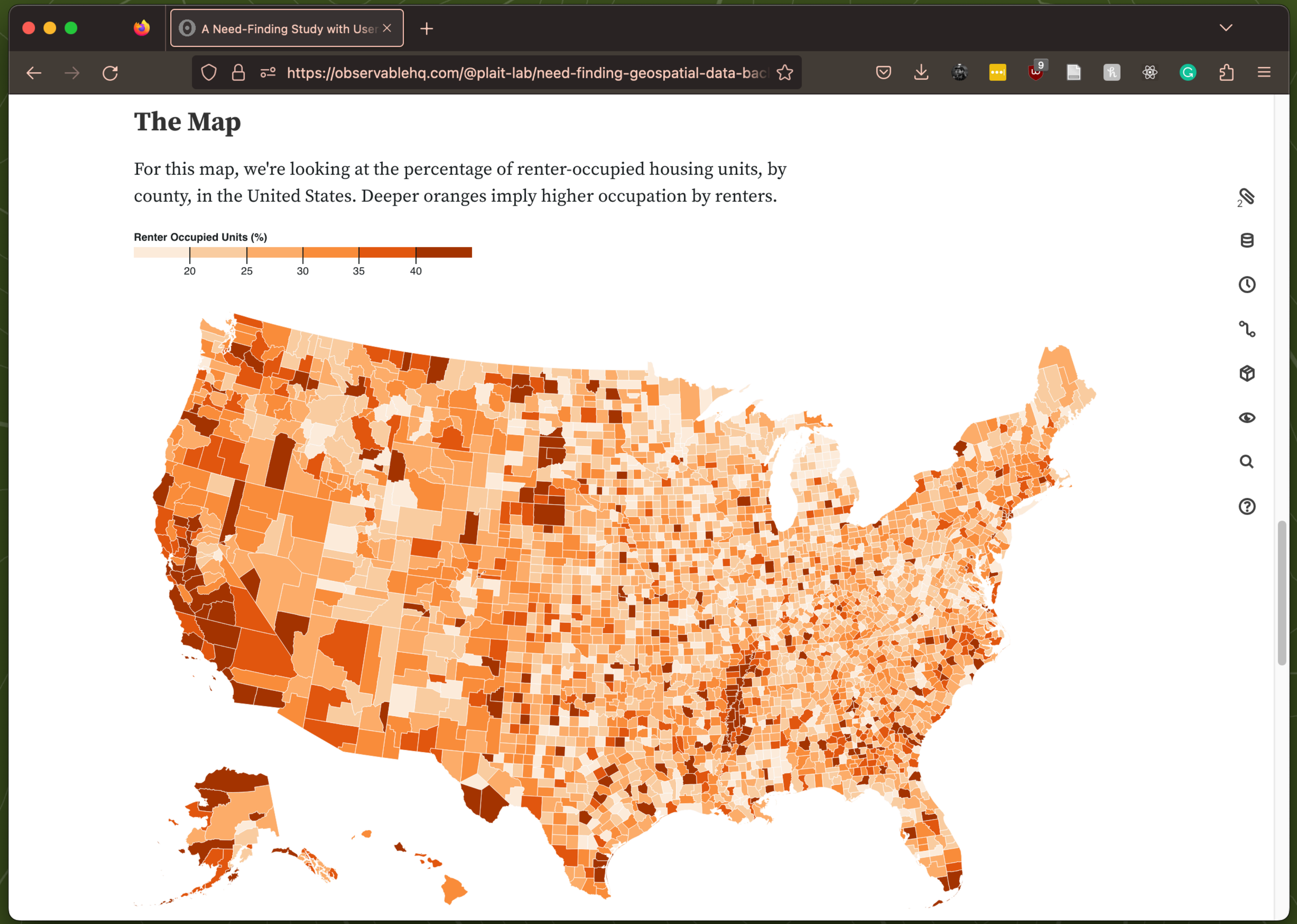
attributes



	GEO_ID string	NAME string	OCCUPIED_U ... integer	RENTER_OCCUPIE ... integer	PERCENT_RENTER_OCC ... number
0	0500000US01001	Autauga County, Ala...	21,559	5,471	25.377
1	0500000US01003	Baldwin County, Ala...	84,047	19,331	23
2	0500000US01005	Barbour County, Ala...	9,322	3,547	38.05
3	0500000US01007	Bibb County, Alabama	7,259	1,831	25.224
4	0500000US01009	Blount County, Alab...	21,205	5,073	23.924
5	0500000US01011	Bullock County, Alab...	3,429	898	26.188
6	0500000US01013	Butler County, Alaba...	6,649	1,777	26.726
7	0500000US01015	Calhoun County, Ala...	44,572	13,202	29.619
8	0500000US01017	Chambers County, A...	13,582	4,449	32.757
9	0500000US01019	Cherokee County, Al...	10,836	2,359	21.77
10	0500000US01021	Chilton County, Alab...	17,140	4,262	24.866
11	0500000US01023	Choctaw County, Ala...	5,330	980	18.386
12	0500000US01025	Clarke County, Alaba...	9,323	2,562	27.48
13	0500000US01027	Clay County, Alabama	5,153	1,206	23.404
14	0500000US01029	Cleburne County, Al...	5,835	1,320	22.622
15	0500000US01031	Coffee County, Alab...	19,951	6,424	32.199
16	0500000US01033	Colbert County, Alab...	21,797	6,188	28.389
17	0500000US01035	Conecuh County, Al...	4,585	1,159	25.278
18	0500000US01037	Coosa County, Alaba...	4,016	760	18.924
19	0500000US01039	Covington County, A...	14,995	3,858	25.729
20	0500000US01041	Crenshaw County, Al...	5,011	1,169	23.329
21	0500000US01043	Cullman County, Ala...	31,733	7,999	25.207
22	0500000US01045	Dale County, Alabama	19,405	8,025	41.355
23	0500000US01047	Dallas County, Alaba...	15,409	6,678	43.338
24	0500000US01049	DeKalb County, Alab...	26,365	7,090	26.892
25	0500000US01051	Elmore County, Alab...	29,794	7,394	24.817
26	0500000US01053	Escambia County, Al...	12,931	4,338	33.547

3,221 rows

location

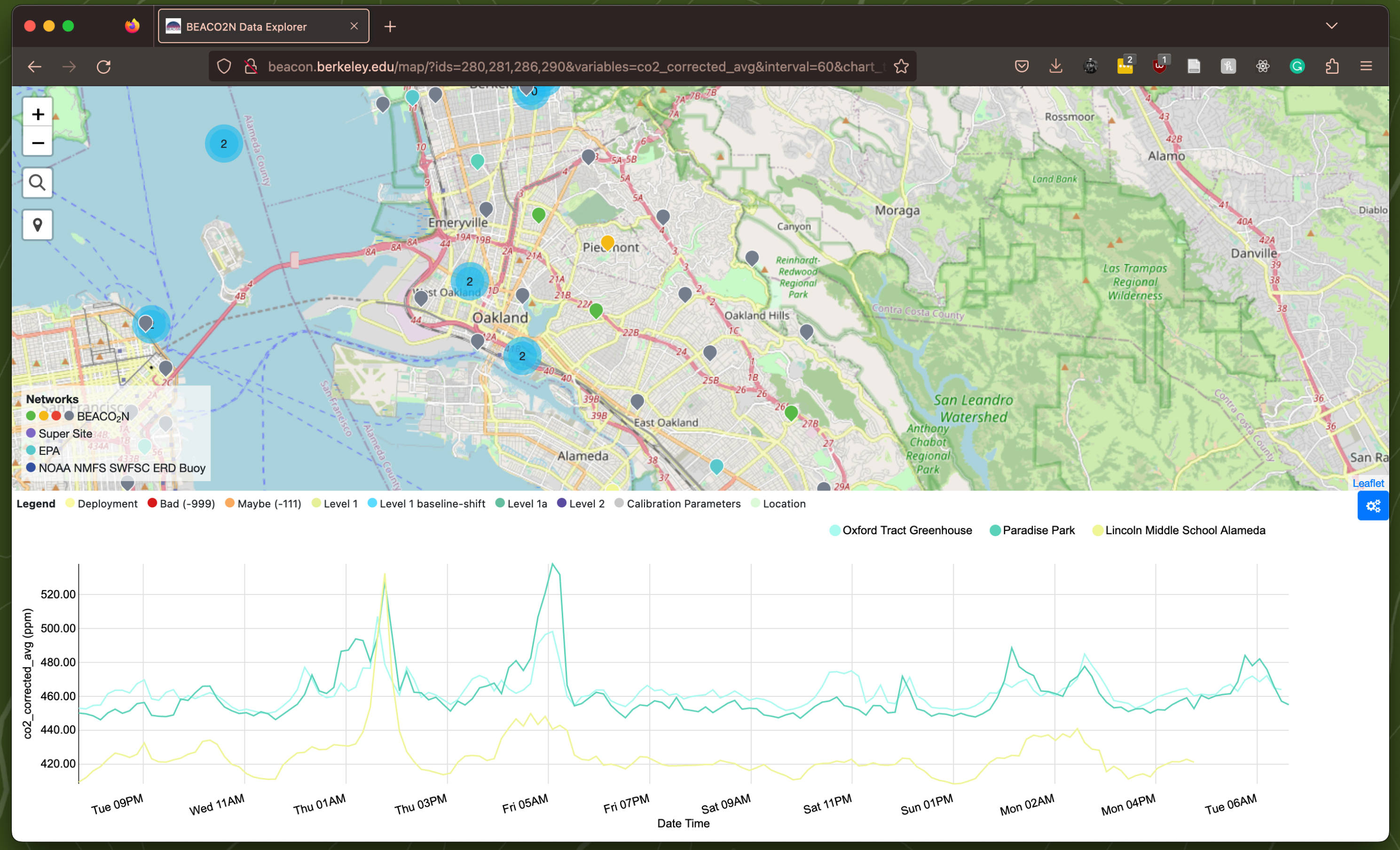


Background

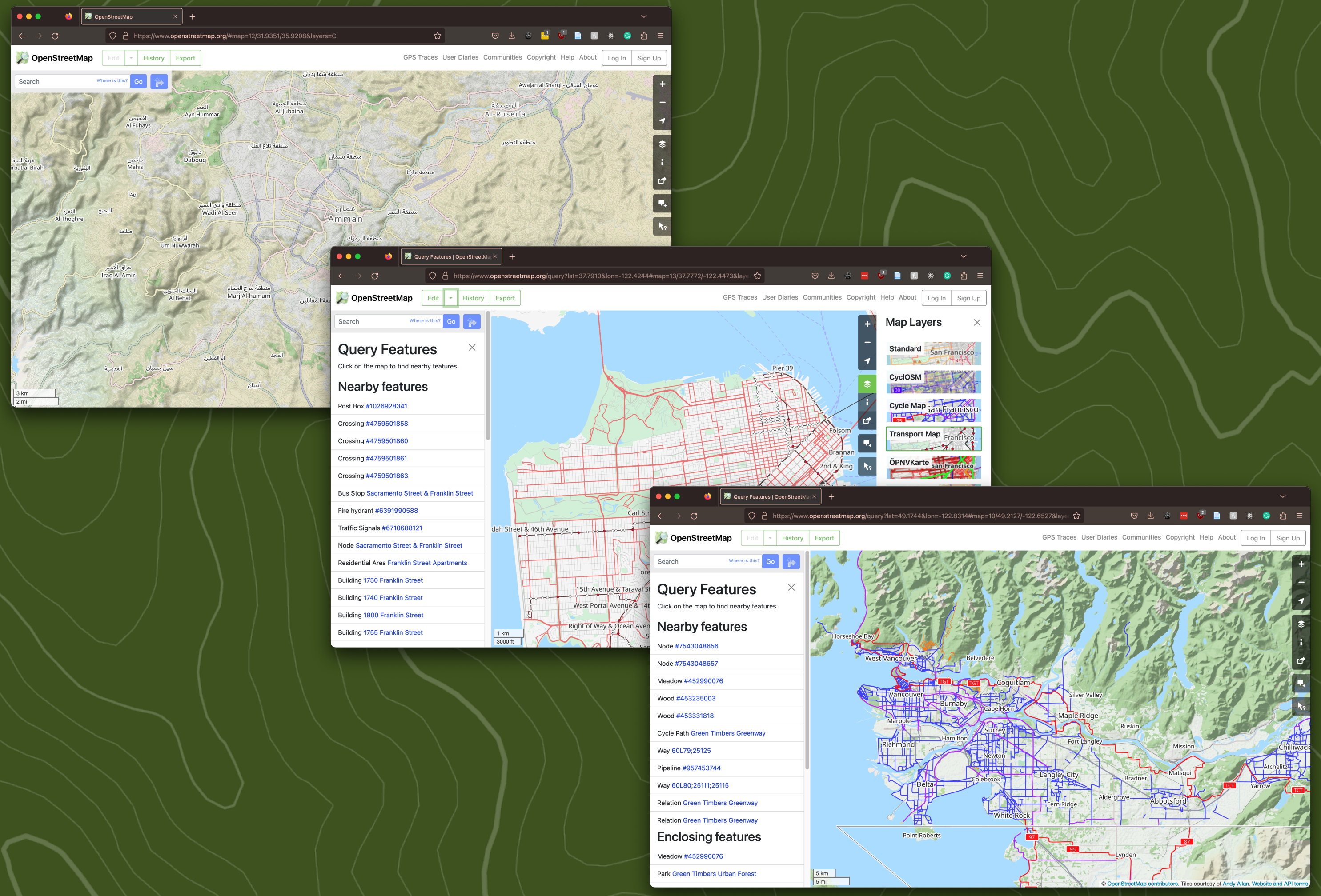
Geospatial Data

Geospatial data is **everywhere** today.

Satellite Imagery



Environmental Sensor Networks



OpenStreetMap

Background

Domain Experts and Geospatial Data

Earth and Climate Science

Social Sciences

Data Journalism

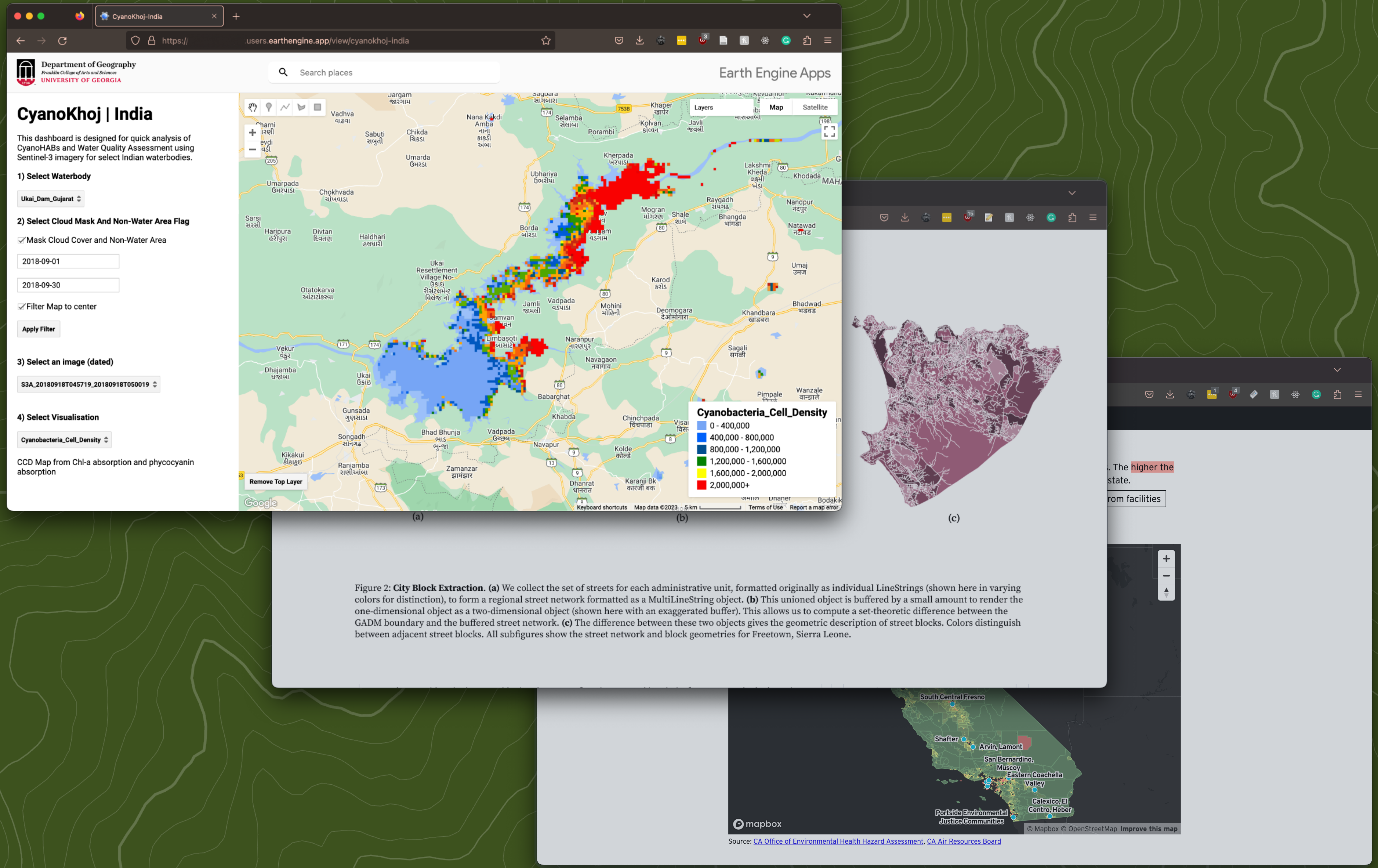
Background

Domain Experts and Geospatial Data


Earth and Climate Science


Social Sciences


Data Journalism



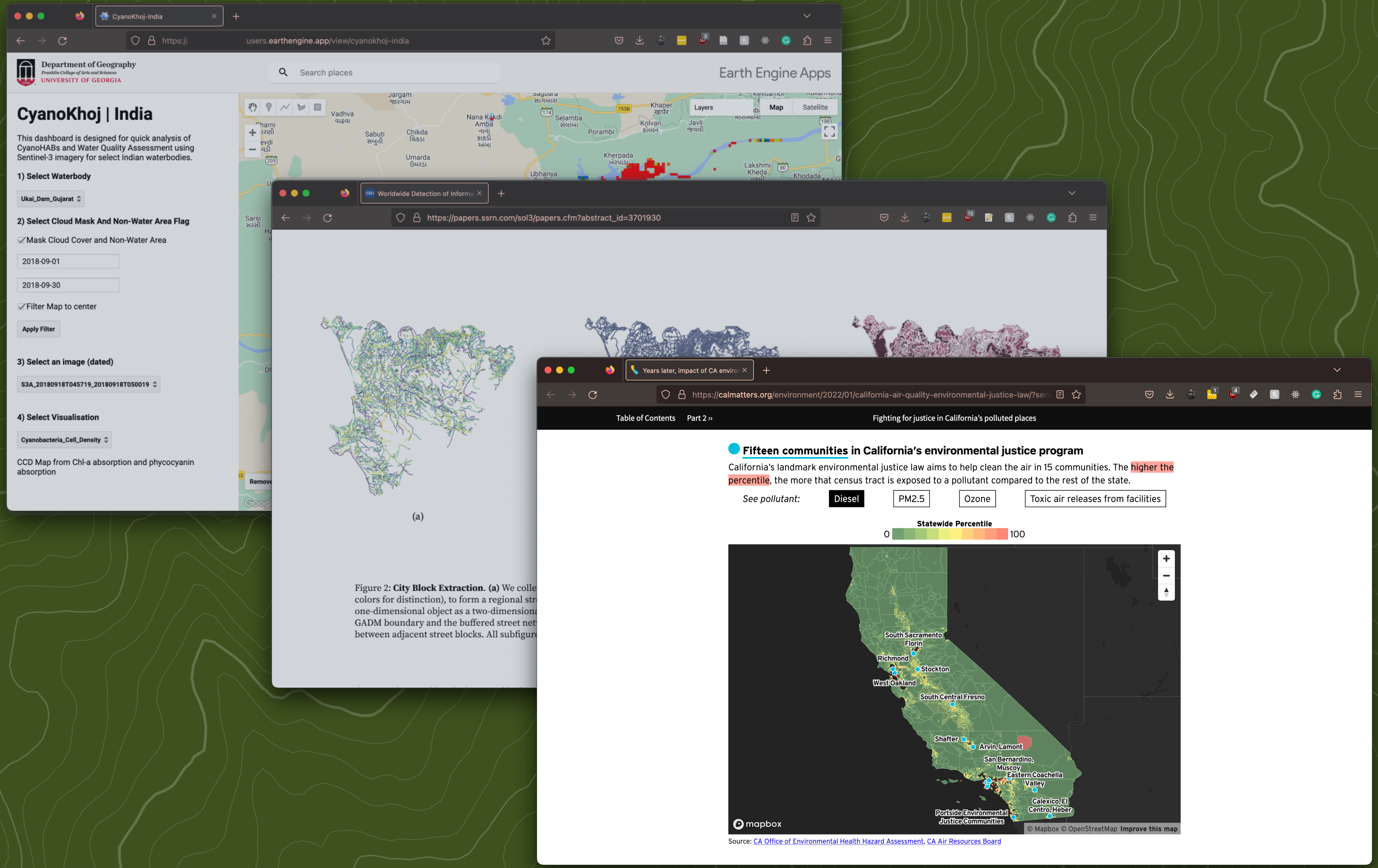
Background

Domain Experts and Geospatial Data

Earth and Climate Science

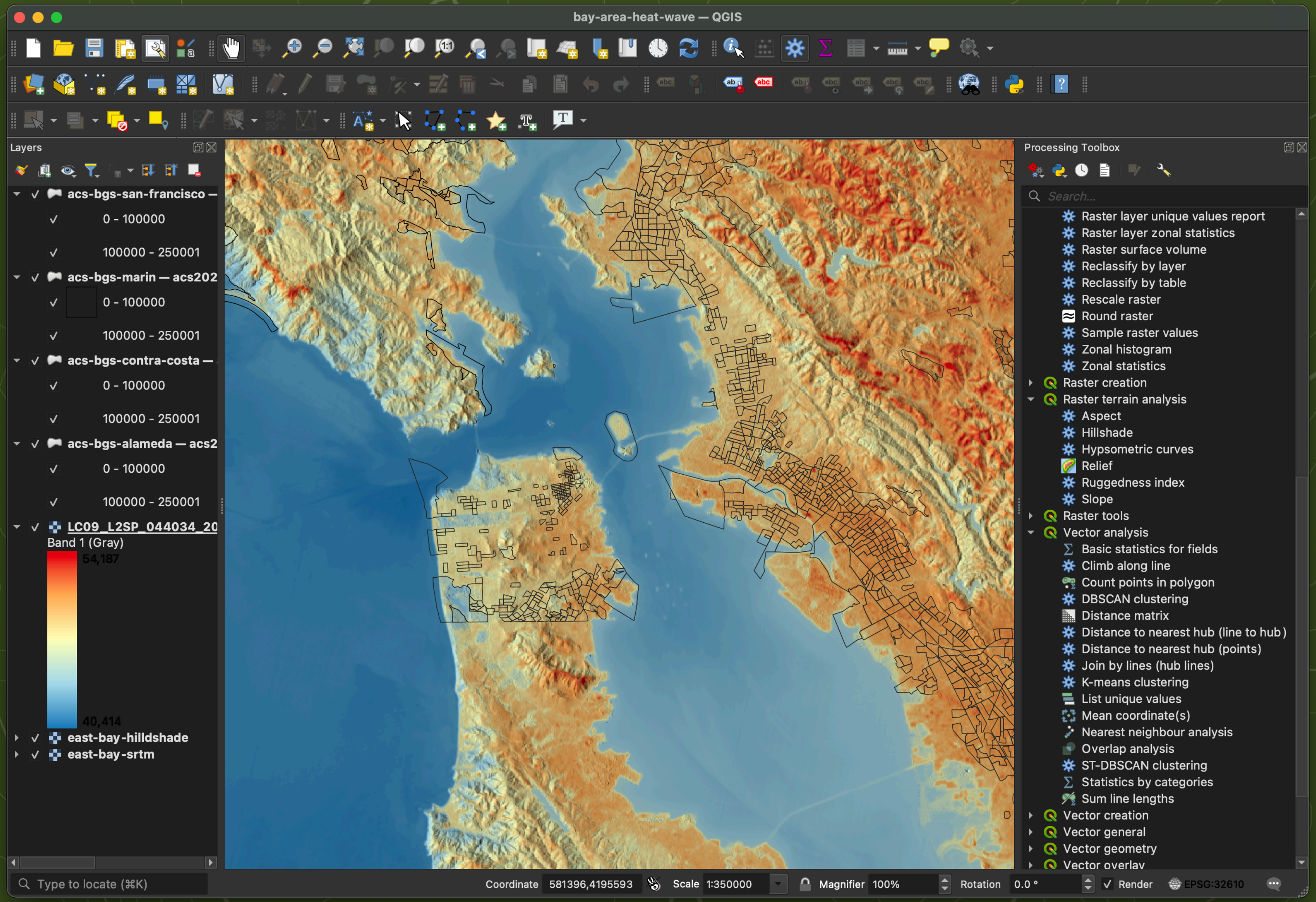
Social Sciences

Data Journalism



Barriers to working with geospatial data are high.

Barriers to working with geospatial data are high.



Example QGIS

Geographic Information Systems

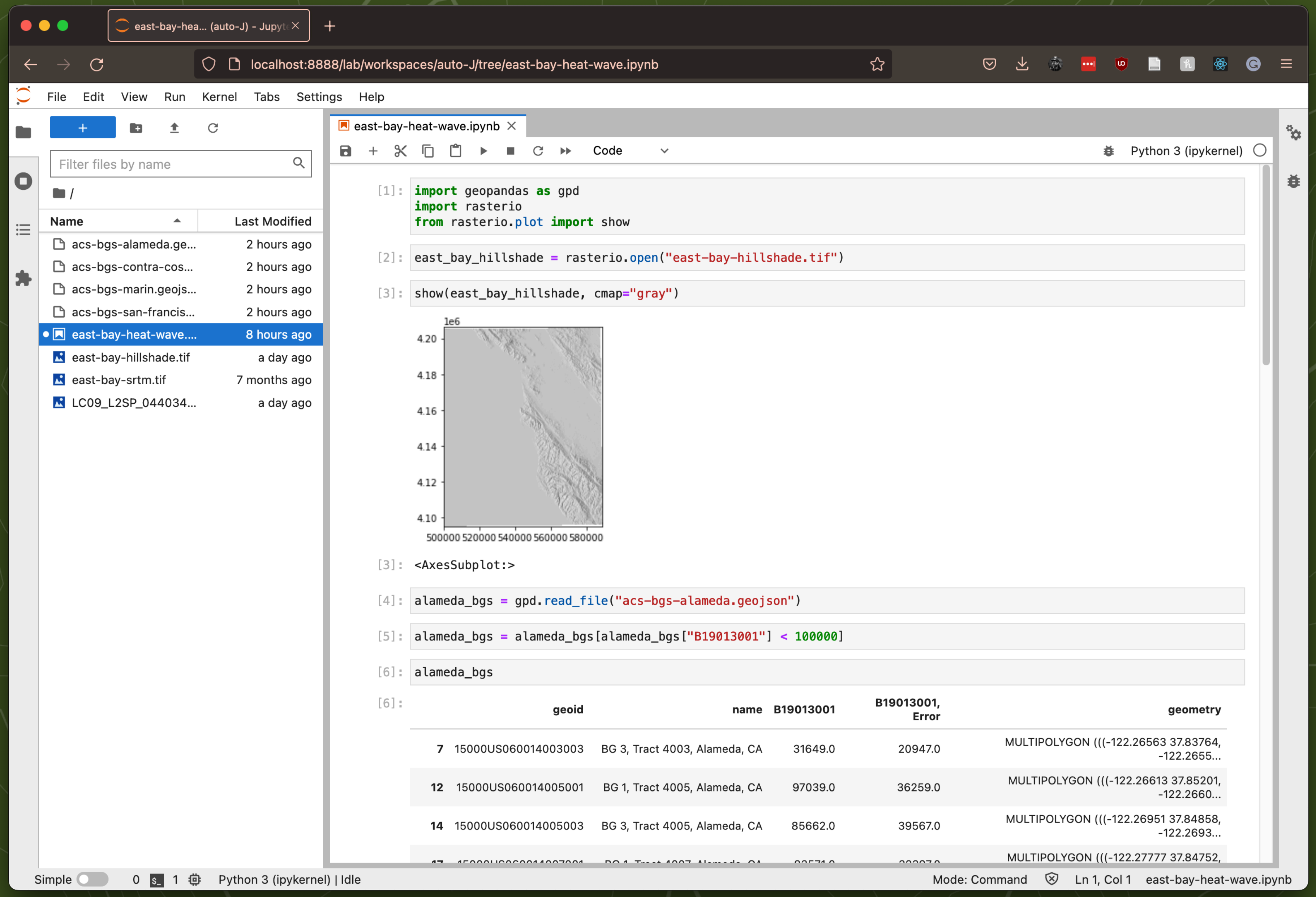
- Require significant background in **geospatial data theory**

Geography   Cartography   Databases   Statistics

  - HCI research[1, 2, 3] has shown that GISs are especially difficult for non-geographers to learn and use.

1. Traynor, C. and Williams, M.G. Why are geographic information systems hard to use? *Conference Companion on Human Factors in Computing Systems* (1995).
2. Traynor, C. & Williams, M. G. End users and GIS: a demonstration is worth a thousand words. in *Your wish is my command: programming by example* 115–134 (Morgan Kaufmann Publishers Inc., 2001).
3. Haklay, M. (Muki) & Skarlatidou, A. Human-Computer Interaction and Geospatial Technologies – Context. in *Interacting with Geospatial Technologies* 1–18 (John Wiley & Sons, Ltd, 2010). doi:10.1002/9780470689813.ch1.

Barriers to working with geospatial data are high.



Example *Jupyter Notebooks*

Programming Systems

- **Geospatial programming abstractions** are increasingly common in Python, R, and JavaScript

mapbox

geopandas

sf

- Must develop proficiency with **programming languages** and **environments**

Background

Research has yet to explore the specific obstacles **domain experts** face in their work with geospatial data.

ArcGIS

GIS Usability

Computational Notebooks

Design Software

mapbox

Geospatial Analysis and Visualization Libraries

Analysis
Visualization

Data Discovery
Data Transformation
Analysis Representation

Background

Contribution

The goal of this research is to **identify the computing needs of domain expert geospatial data users**.

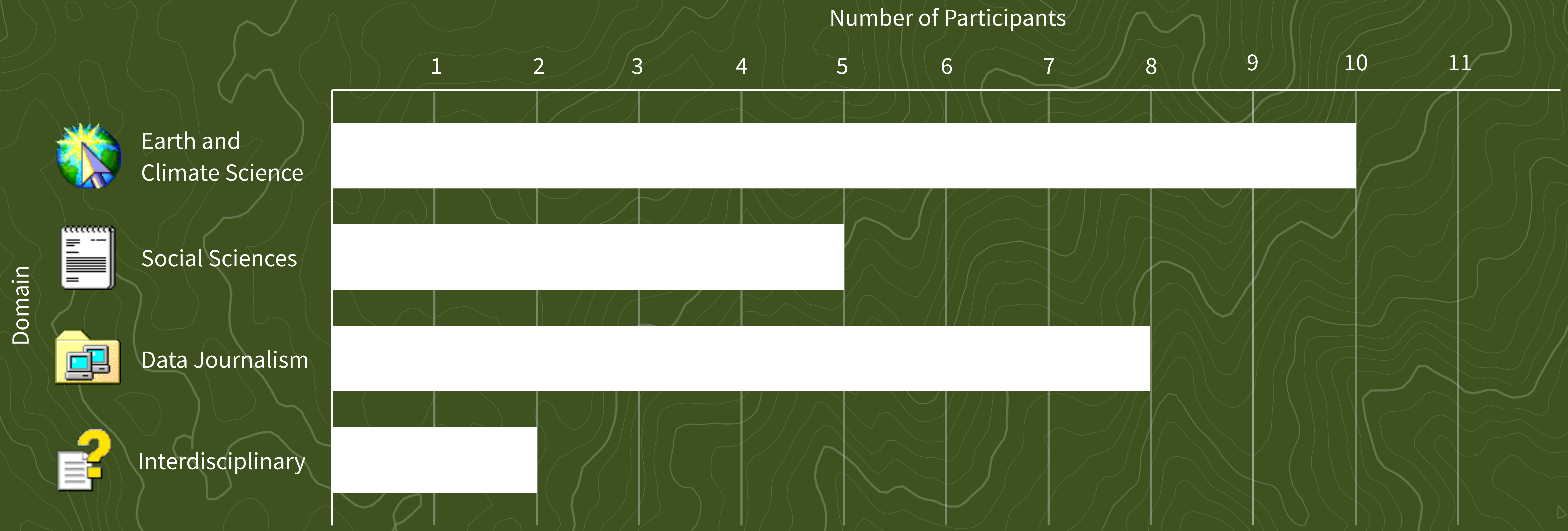
Roadmap


1. Study Design
2. Findings
3. Design Opportunities
4. Introducing `cartokit`

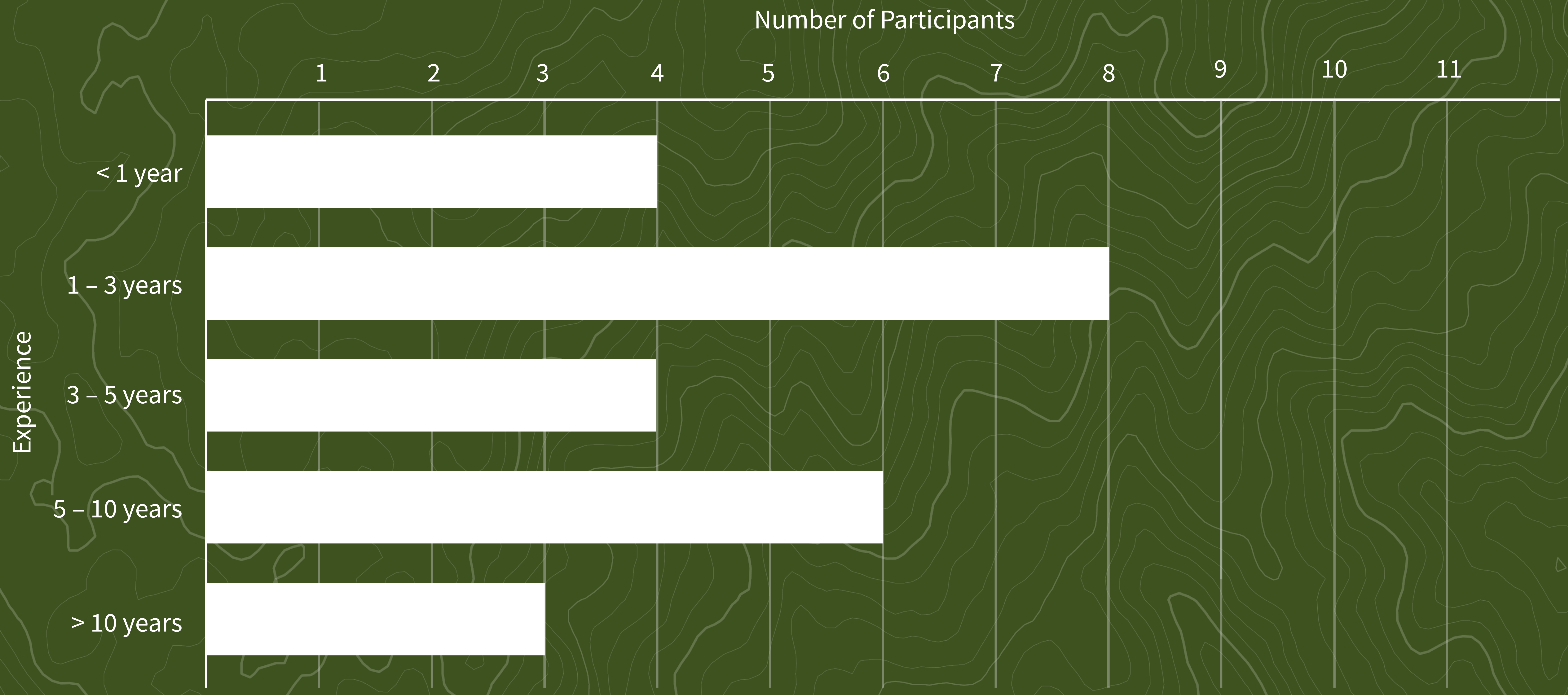
Roadmap

1. **Study Design**
2. Findings
3. Design Opportunities
4. Introducing `cartokit`

We conducted a contextual inquiry study with 25 participants.



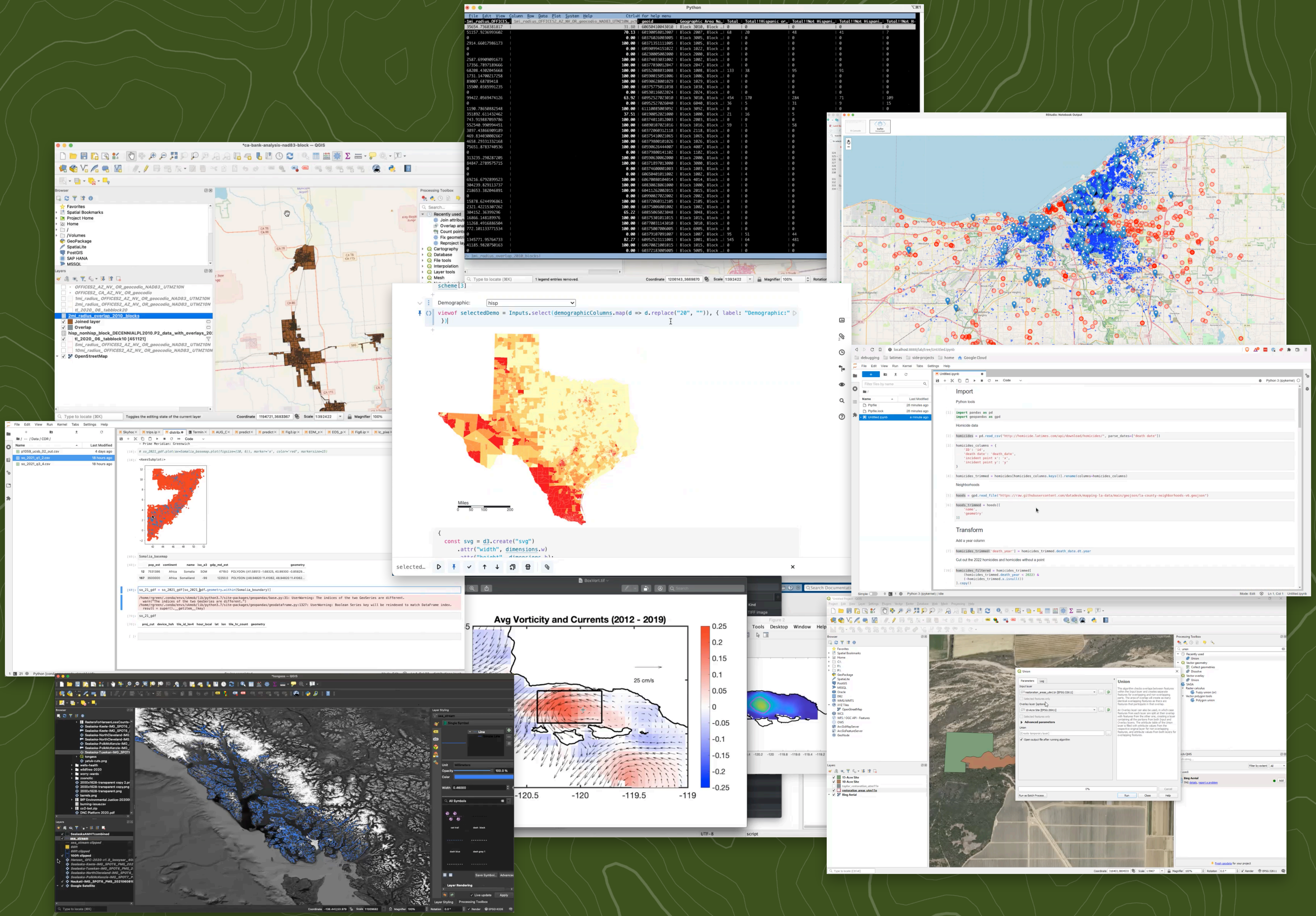
We conducted a contextual inquiry study with 25 participants.



Study Design

Session Structure and Analysis

- 50–70 minute open-task **observations**
  - Followed by **semi-structured post-interviews**
- Inductive **thematic analysis**



Roadmap


1. Study Design
2. Findings
3. Design Opportunities
4. Introducing `cartokit`

Roadmap

1. Study Design

2. Findings

3. Design Opportunities

4. Introducing `cartokit`

Findings

We identified **12 challenges** across **five phases** of participants' work with geospatial data.

Data Discovery

Solving Geospatial Data Constraints

Data Transformation

Aligning Geospatial Datasets

Topological Errors

Reducing Resolution to Improve Performance

Data Subsetting and Caching

Analysis

Identifying Geospatial Operators

Understanding Geospatial Operator Semantics

Visibility of Geometry in Programming Environments

Analysis Representation

Reproducing Geospatial Analyses

Creating Informal Program Representations

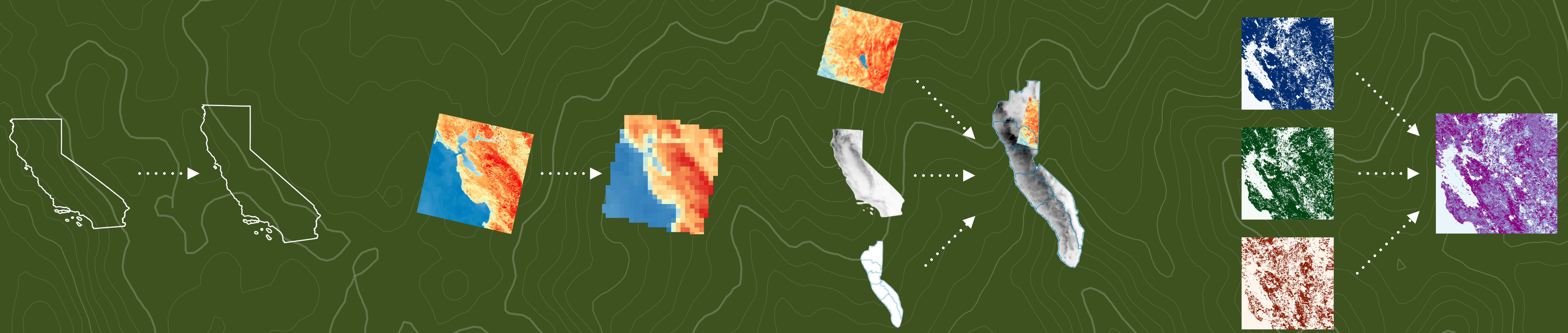
Visualization

Sketching Cartographic Variants

Geospatial Information in Design Software

Aligning Geospatial Datasets

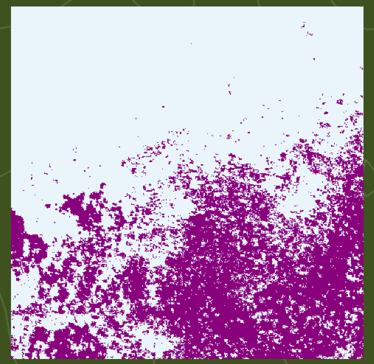
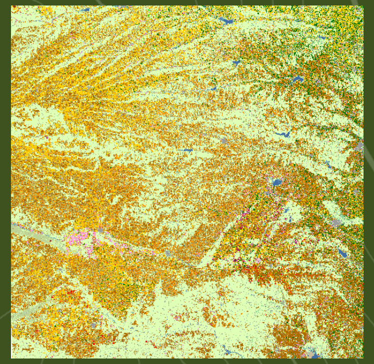
Participants needed to transform datasets to a **shared spatial and temporal reference** for analysis, but alignment required **complex preprocessing**.



Reprojection *Resampling* *Clipping* *Temporal Aggregation*

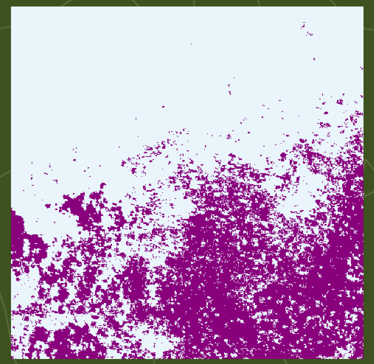
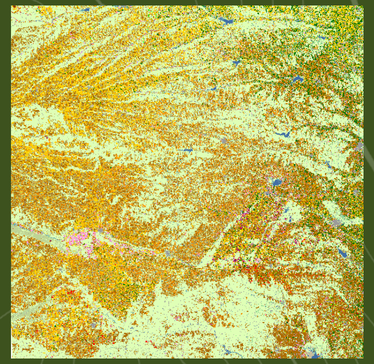
Aligning Geospatial Datasets

PE2's Task. Develop a model to predict groundwater withdrawal.

	Spatial Resolution	*Temporal Interval*	*Geographic Extent*
 *MOD16*	**500m**	**8-Day**	**Global**
 *PRISM*	**4638.3m**	**Monthly**	**Conterminous U.S.**
 *USDA-NASS*	**30m**	**Yearly**	**Conterminous U.S.**

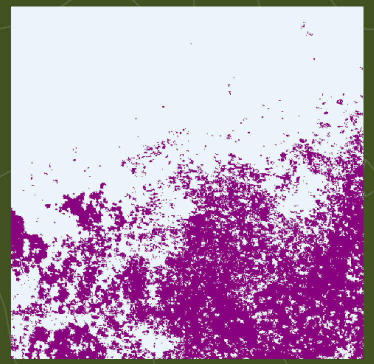
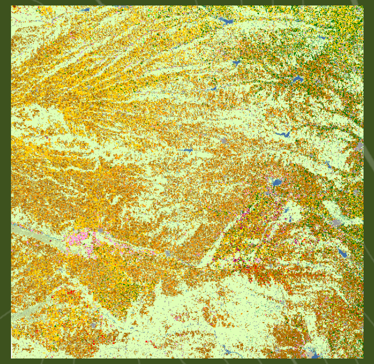
Aligning Geospatial Datasets

PE2's Task. Develop a model to predict groundwater withdrawal.

	Spatial Resolution	*Temporal Interval*	*Geographic Extent*
 *MOD16*	500m 4638.3m Resample	**8-Day**	**Global**
 *PRISM*	4638.3m	**Monthly**	**Conterminous U.S.**
 *USDA-NASS*	30m 4638.3m Resample	**Yearly**	**Conterminous U.S.**

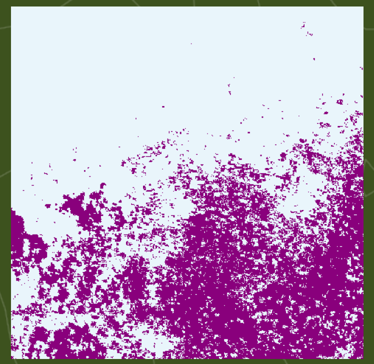
Aligning Geospatial Datasets

PE2's Task. Develop a model to predict groundwater withdrawal.

	Spatial Resolution	*Temporal Interval*	*Geographic Extent*
 *MOD16*	500m → Resample → 4638.3m	8-Day → Accumulate → Yearly	**Global**
 *PRISM*	4638.3m	Monthly → Accumulate → Yearly	**Conterminous U.S.**
 *USDA-NASS*	30m → Resample → 4638.3m	Yearly	**Conterminous U.S.**

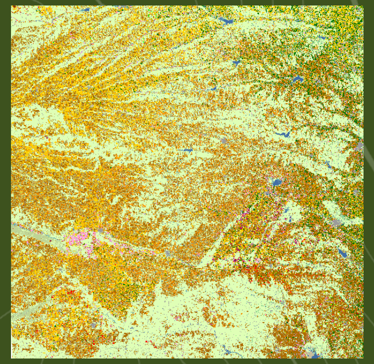
Aligning Geospatial Datasets

PE2's Task. Develop a model to predict groundwater withdrawal.



MOD16

PRISM



USDA-NASS

Dataset	*Spatial Resolution*	*Temporal Interval*	*Geographic Extent*
MOD16	**500m** Resample **4638.3m**	**8-Day** Accumulate **Yearly**	**Global** Clip **Kansas**
PRISM	**4638.3m**	**Monthly** Accumulate **Yearly**	**Conterminous U.S.** Clip **Kansas**
USDA-NASS	**30m** Resample **4638.3m**	**Yearly**	**Conterminous U.S.** Clip **Kansas**

Aligning Geospatial Datasets

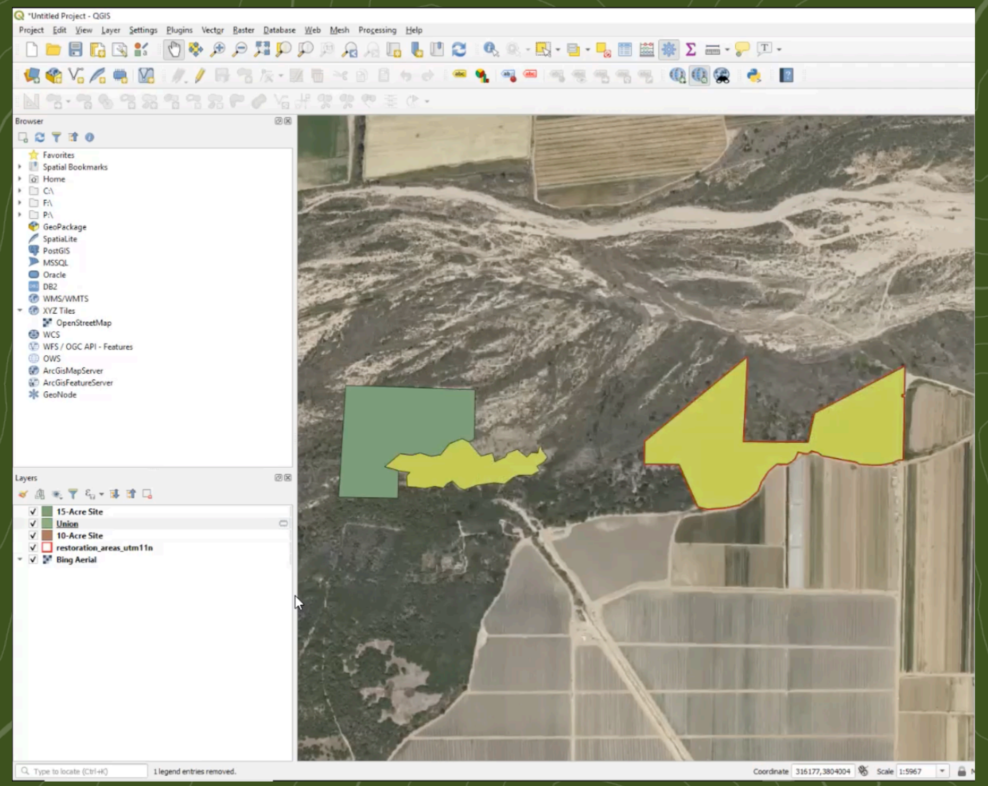
Aligning geospatial datasets required participants to have significant **fluency in geospatial data theory** in addition to **contextual information** about the datasets themselves.

Server Toolbox
Ready to Use Toolbox
Spatial Analyst Toolbox
Spatial Statistics Toolbox
… ***+35 More***

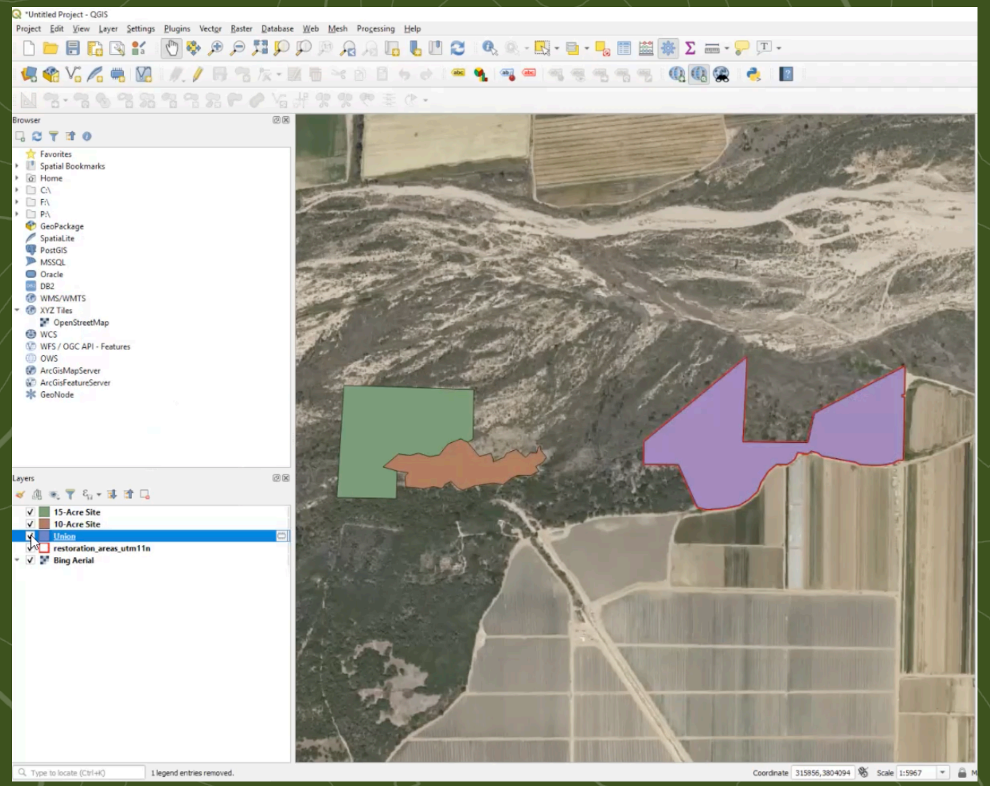
Bitwise Left Shift
Kriging
Raster Calculator
Iso Cluster Unsupervised
Fuzzy Overlay
Zonal Histogram
Darcy Flow
… ***+200 More***

Identify the correct **sequence** of **transformations** among **hundreds of operators**

Expected



Actual



Determine when selected **transformations** produced **undesirable results**

Findings

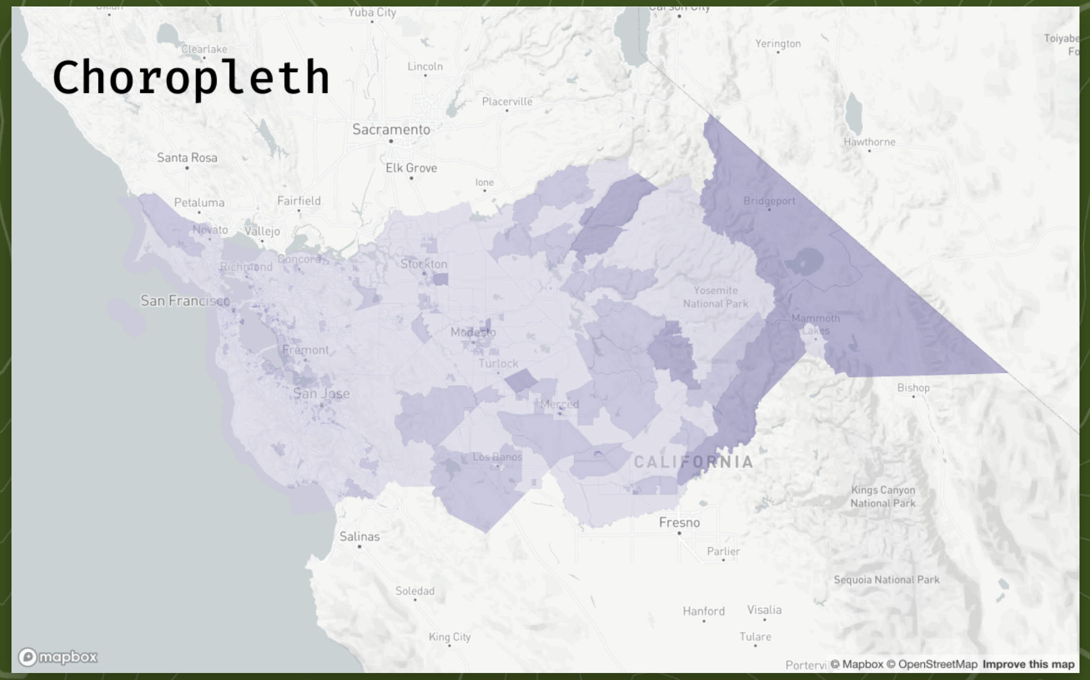
We identified **12 challenges** across **five phases** of participants' work with geospatial data.

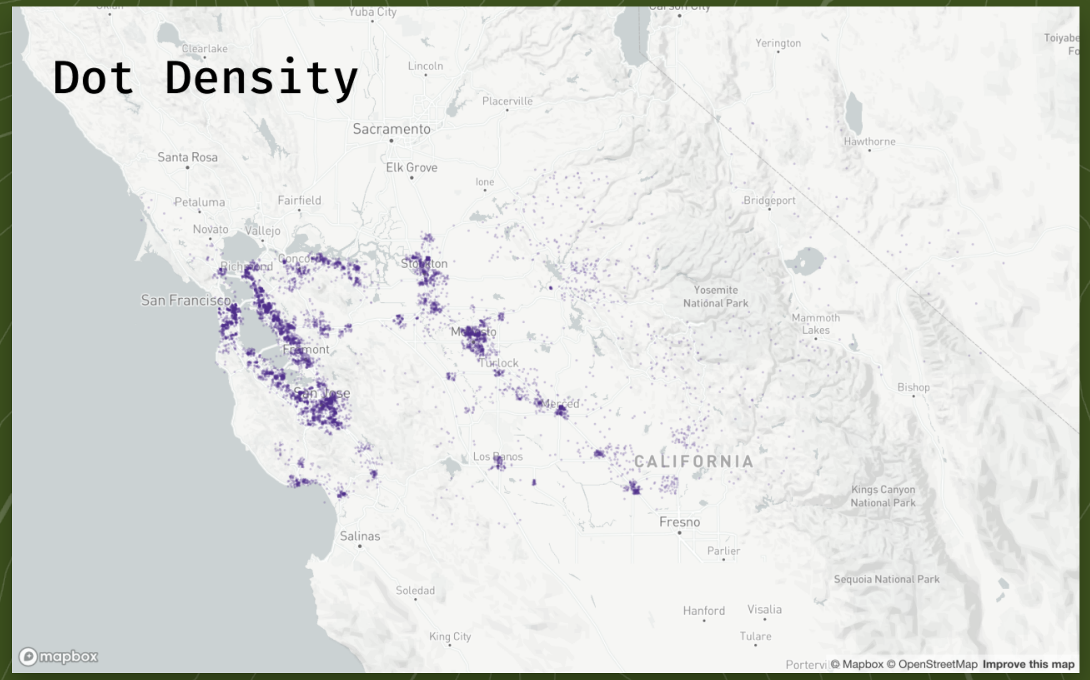
Data Discovery	Data Transformation	Analysis	Analysis Representation	Visualization
Solving Geospatial Data Constraints	*Aligning Geospatial Datasets*	*Identifying Geospatial Operators*	*Reproducing Geospatial Analyses*	*Sketching Cartographic Variants*
	Topological Errors	*Understanding Geospatial Operator Semantics*	*Creating Informal Program Representations*	*Geospatial Information in Design Software*
	Reducing Resolution to Improve Performance	*Visibility of Geometry in Programming Environments*		
	Data Subsetting and Caching			

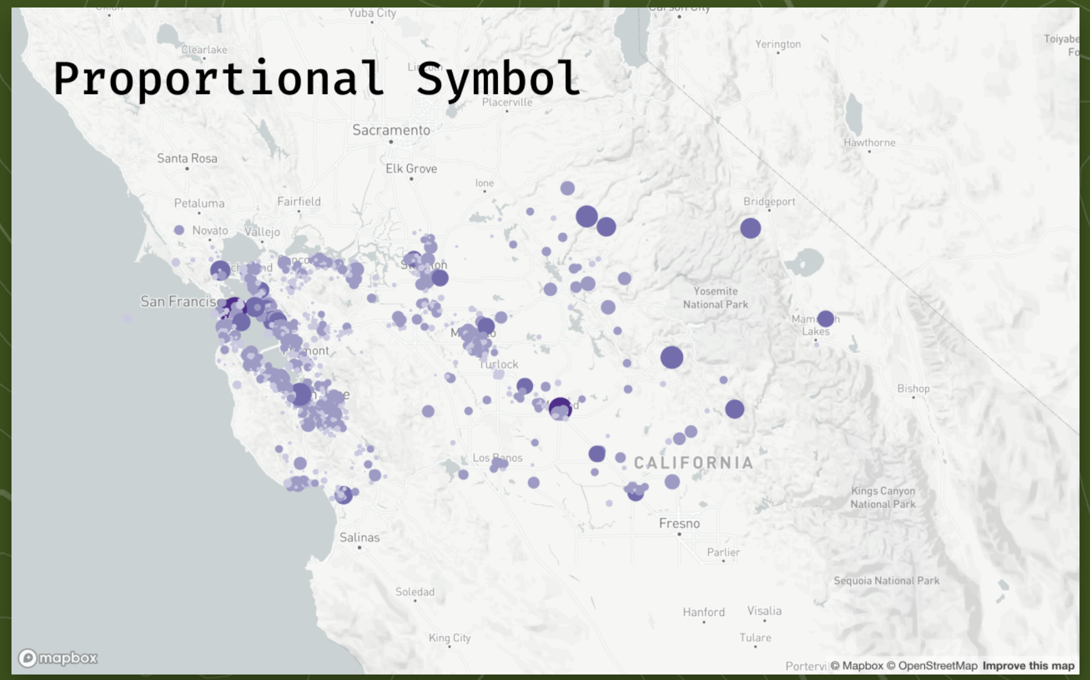
Sketching Cartographic Variants

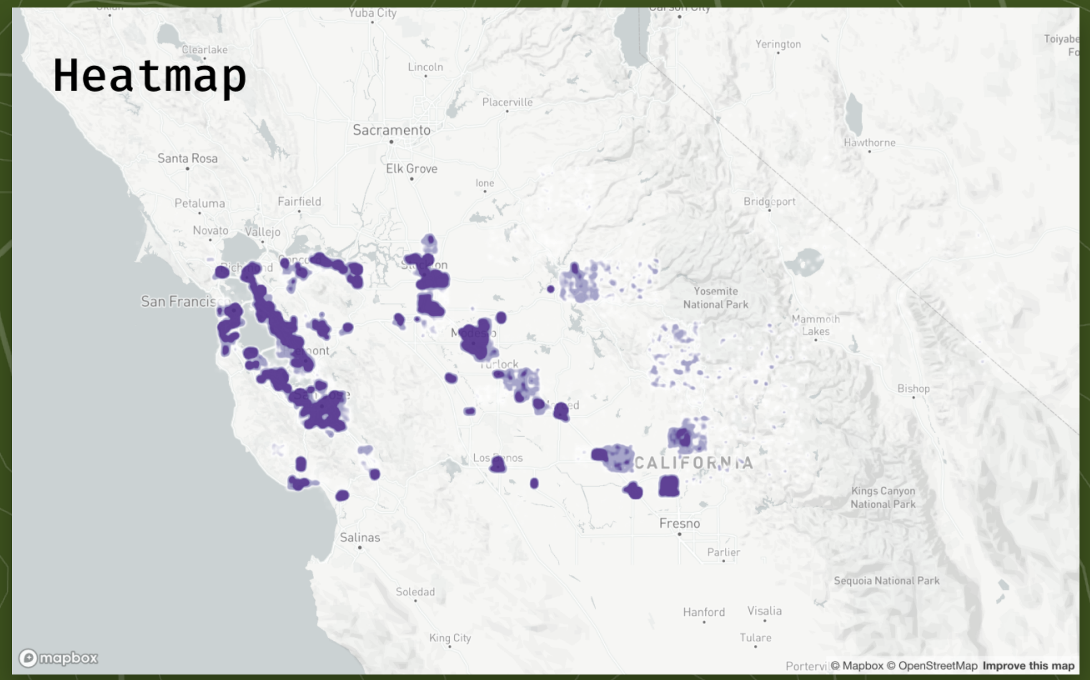
Participants wanted to visualize their data using many different **cartographic representations** to:

- Identify the **map type** that represented their data most effectively
- Produce **tangible artifacts** for collaborators to evaluate



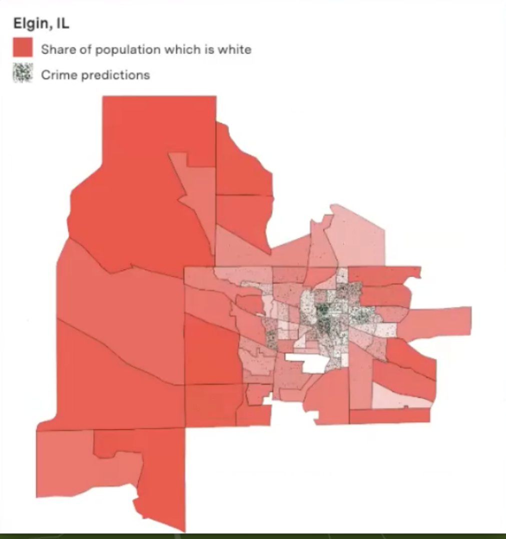


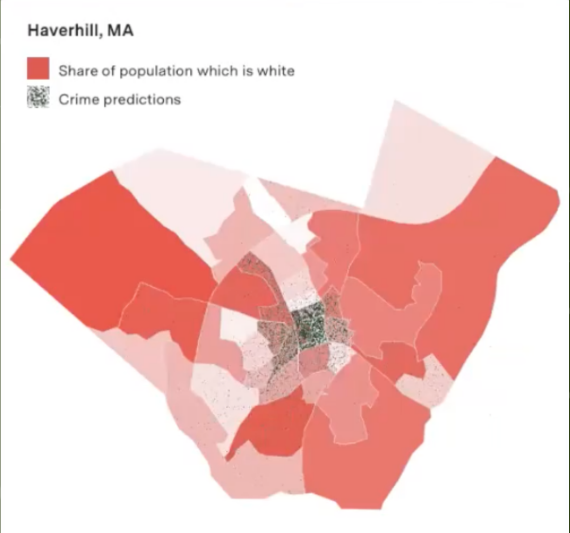




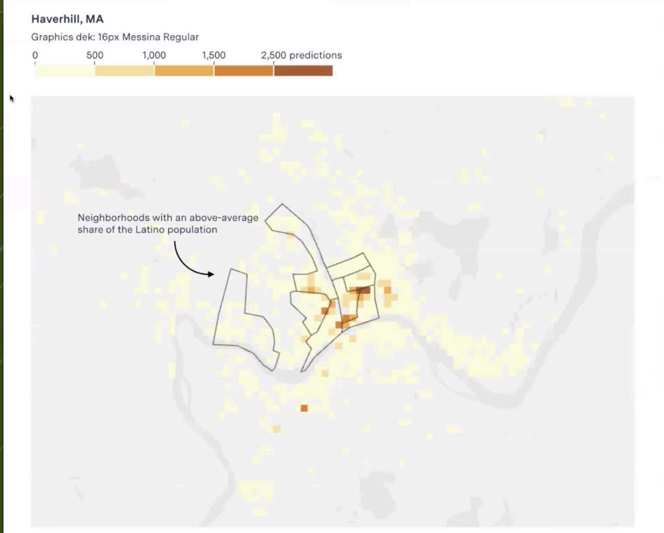
Sketching Cartographic Variants

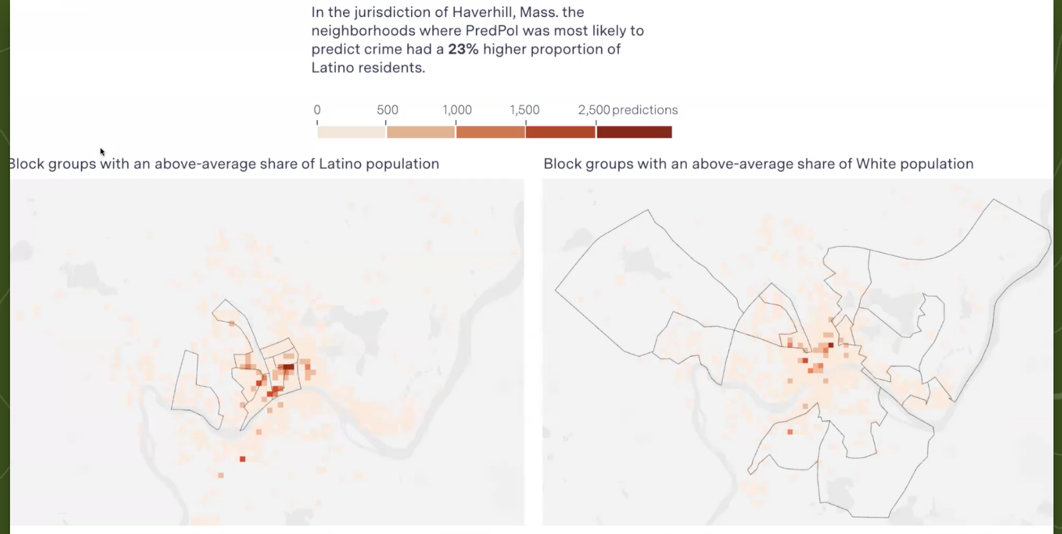
PJ5 created over **20 draft maps** for a story on biased predictive policing algorithms.



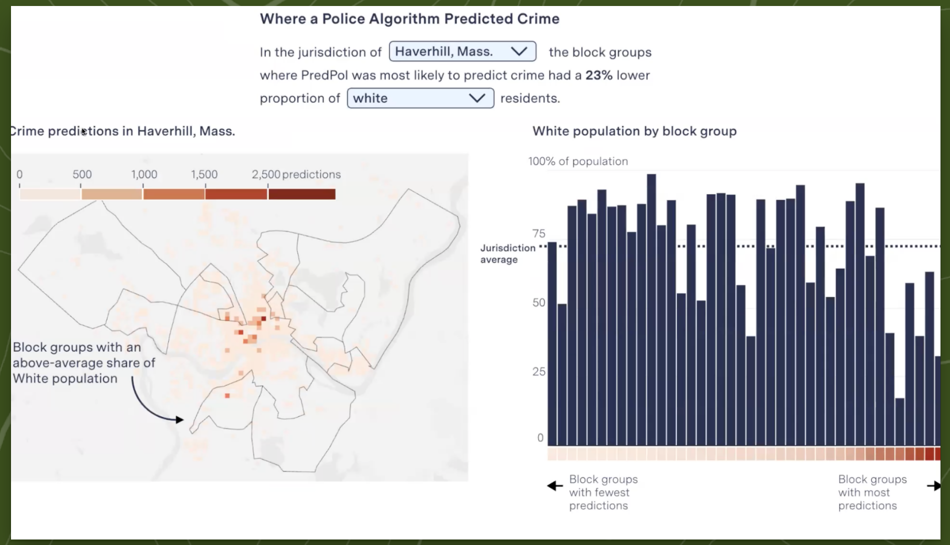


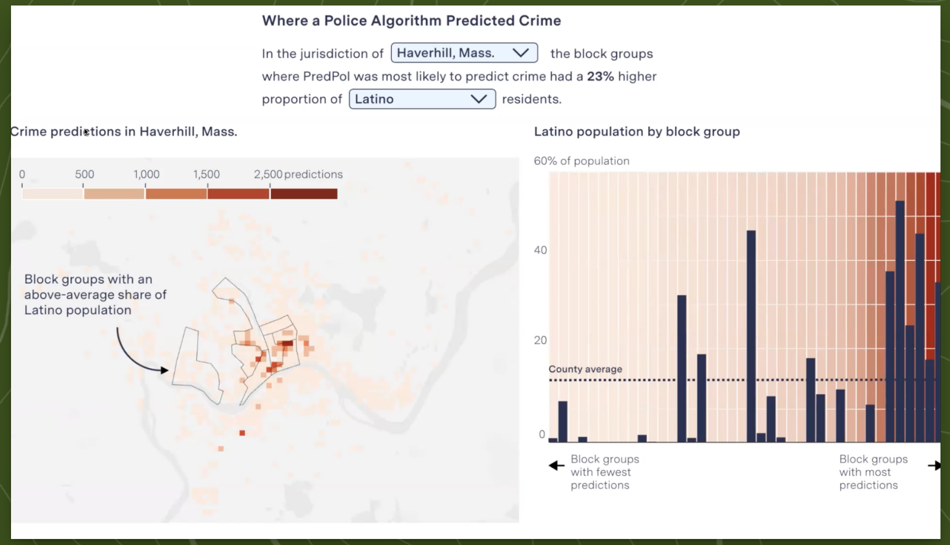
Choropleth and Dot Density





Gridded Heat Map

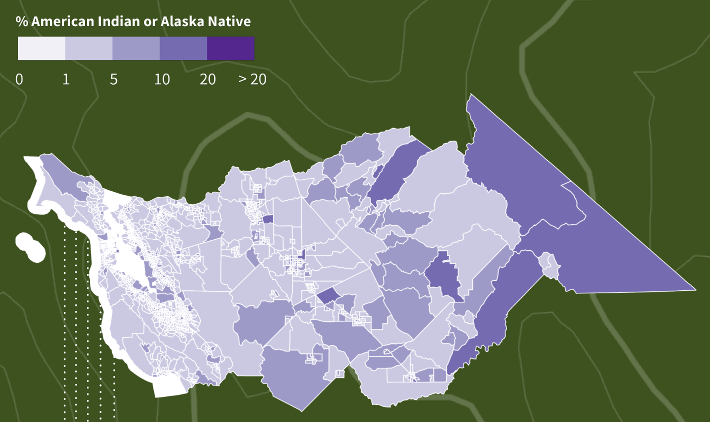




Gridded Heat Map with Bar Charts

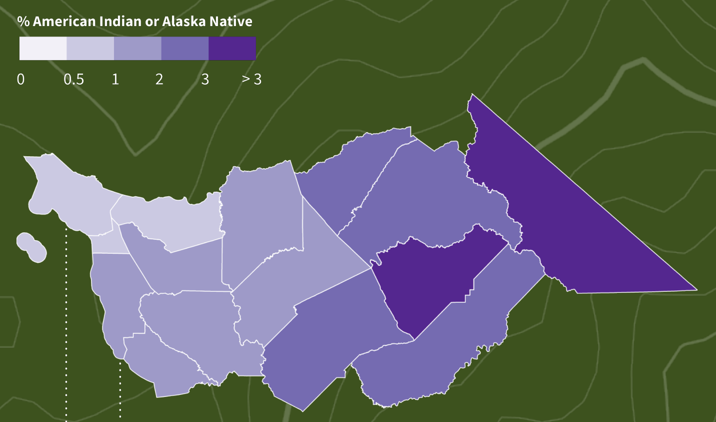
Sketching Cartographic Variants

Producing most map variants required going through **the entire analysis and visualization pipeline**.



GEO_ID	NAME	COUNTY	TOTAL_POP	AIAN_POP
14000US06081613800	Census Tract 6138, San Mateo, CA	081	4363	2
14000US06081602900	Census Tract 6029, San Mateo, CA	081	3727	0
14000US06041131100	Census Tract 1311, Marin, CA	041	1280	1
14000US06041132100	Census Tract 1321, Marin, CA	041	1926	0
14000US06041133000	Census Tract 1330, Marin, CA	041	2870	57
…	…	…	…	…

Census Tracts



COUNTY	TOTAL_POP	AIAN_POP
081	737888	9397
041	259441	673
…	…	…

Counties

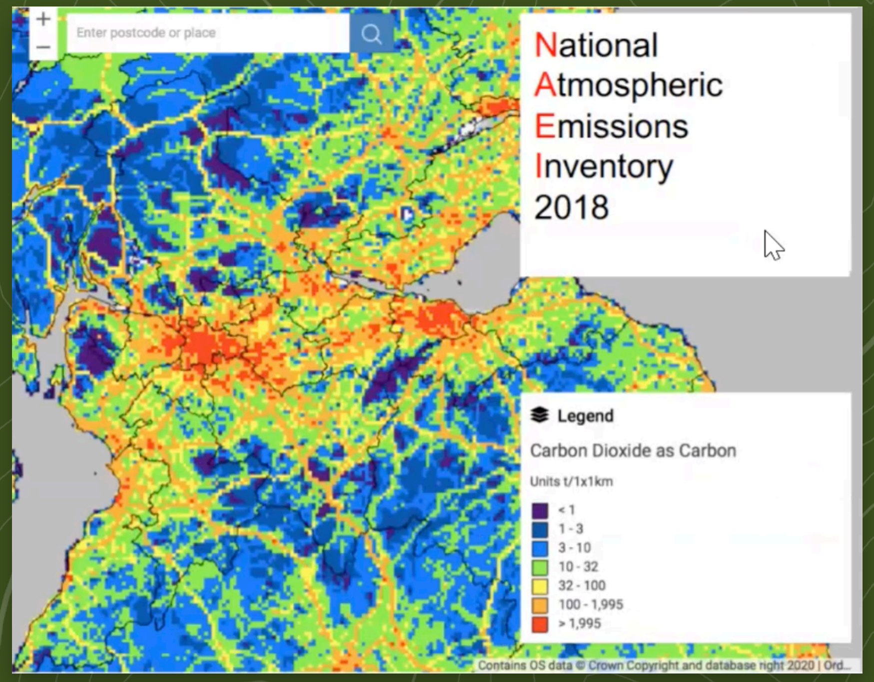
Additional Data Transformation

Across Multiple Tools

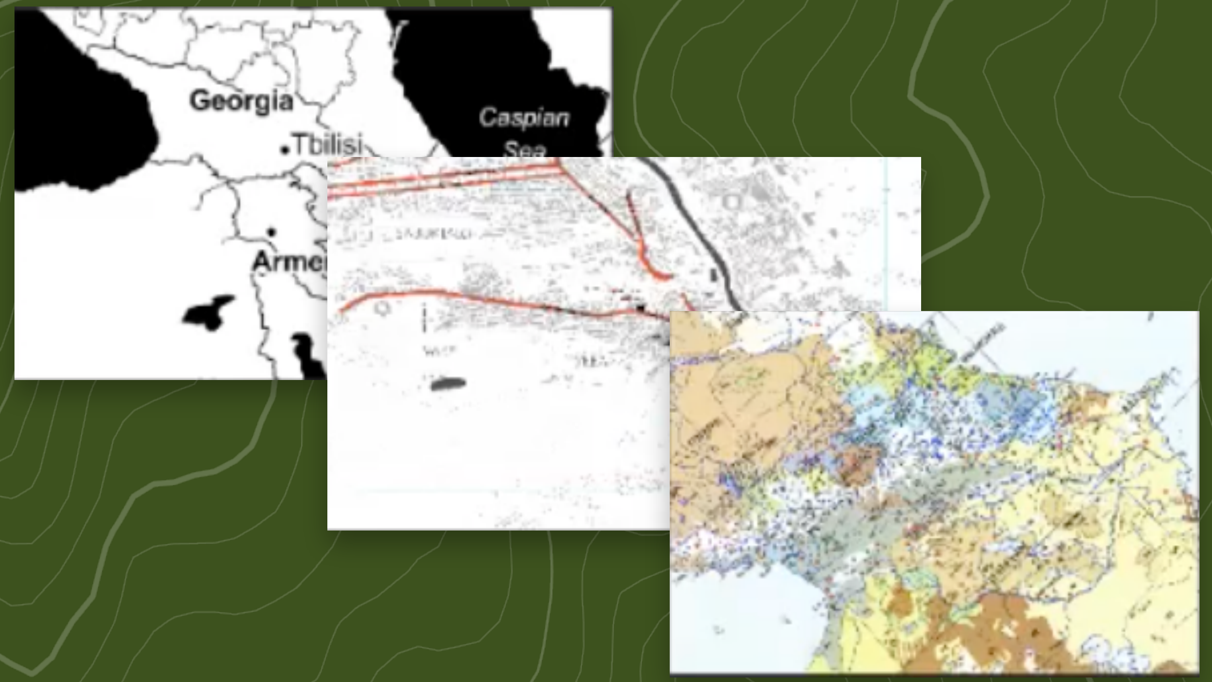
Sketching Cartographic Variants

Participants tried to **speed up** the drafting process in creative ways. One common technique involved **screenshotting in-progress maps**.

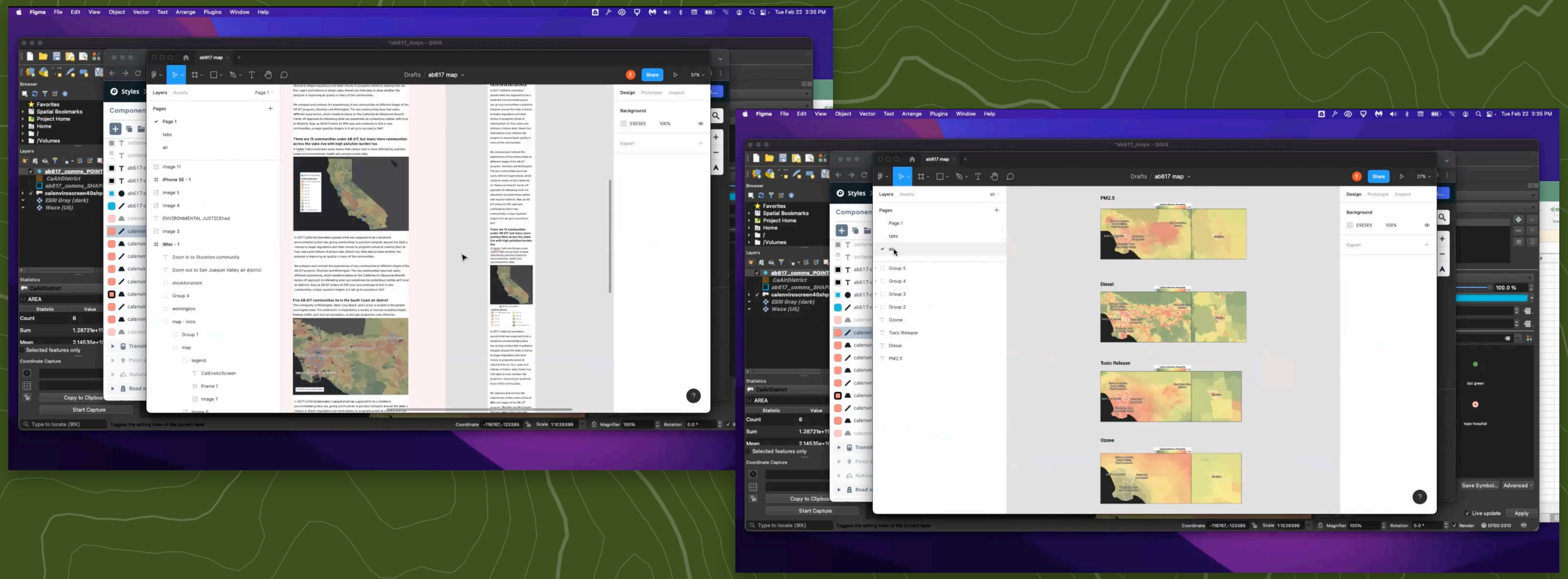
Participant E5



Participant S2



Participant J6

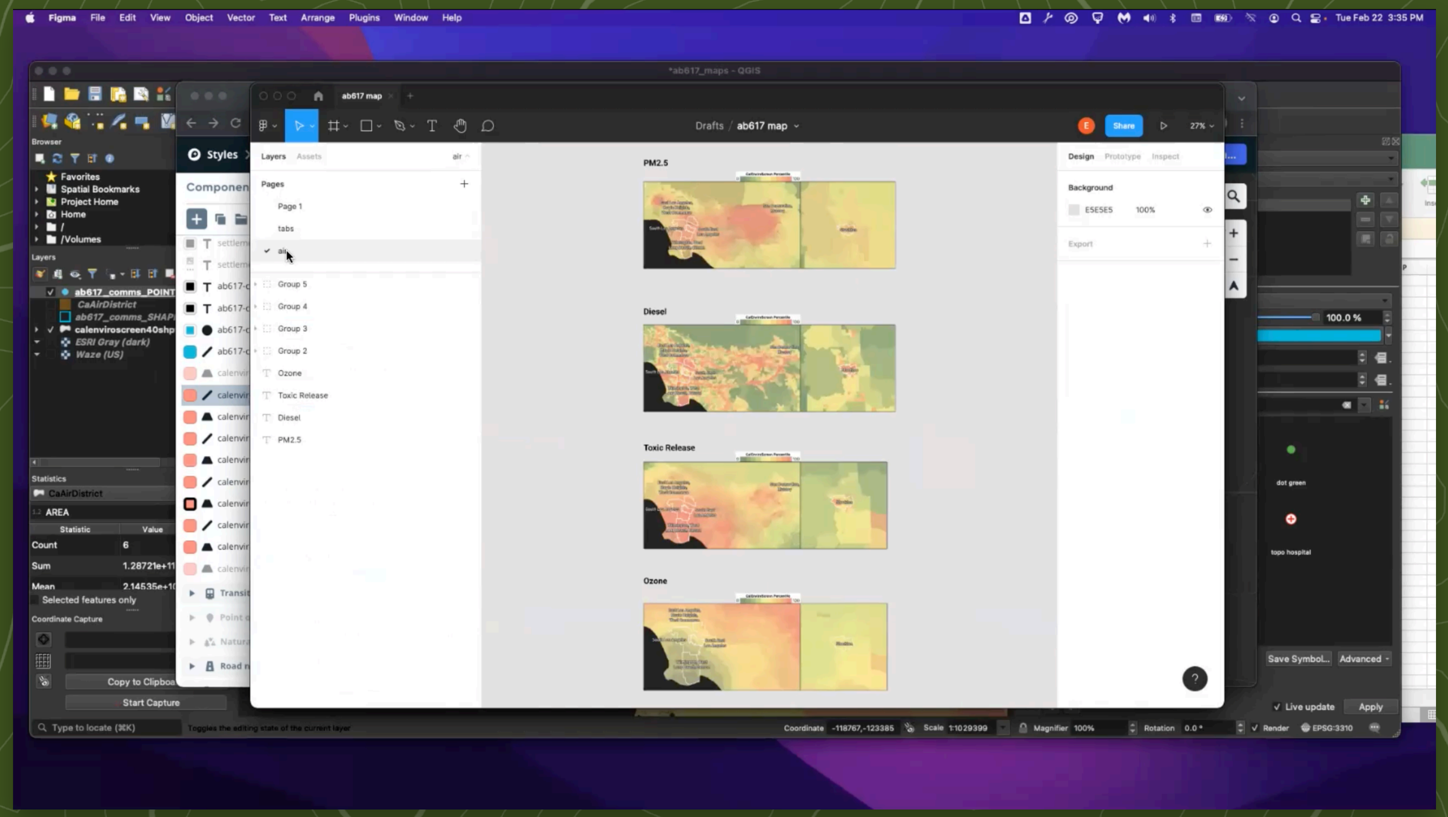


Sketching Cartographic Variants

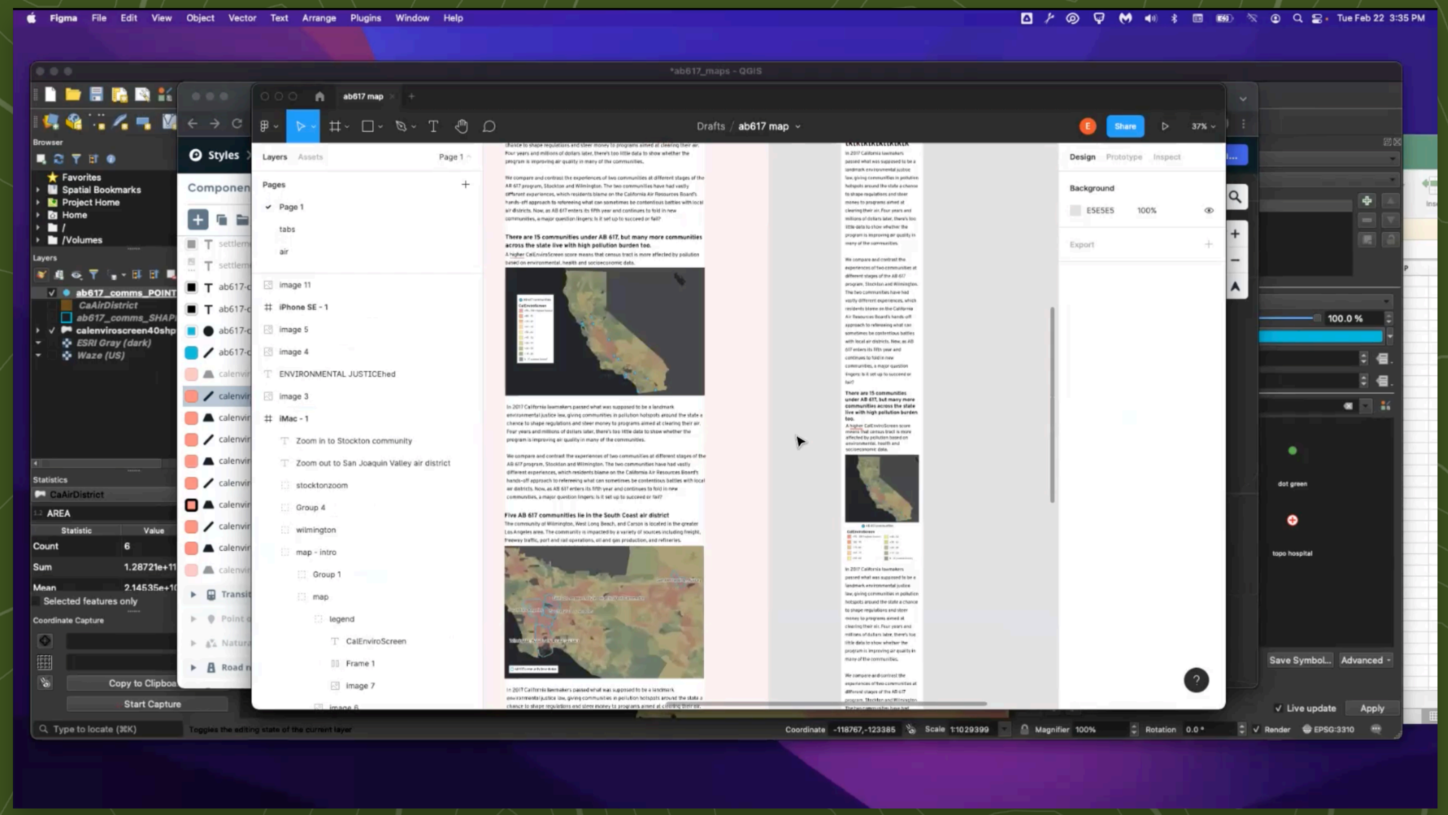
Participants tried to **speed up** the drafting process in creative ways. One common technique involved **screenshotting in-progress maps**.

Participant J6

Screenshots



Layouts



Screenshotting came with limitations.

1. Only allowed users to capture cartographic changes **within** a map type rather than **across** map types
2. Once a final map design was chosen, participants had to **reproduce the selected draft in code**

Roadmap


1. Study Design
2. Findings
3. Design Opportunities
4. Introducing cartokit

Roadmap

1. Study Design
2. Findings
3. **Design Opportunities**
4. Introducing `cartokit`

Design Opportunities

We synthesized **six design opportunities** for designers and developers of geospatial analysis and visualization systems.

Solving Geospatial Data Constraints

Opportunity 1. Participants struggled to find geospatial data satisfying complex spatial and temporal constraints (Section 5.1). While many could describe their constraints succinctly, satisfying them involved constructing bespoke workflows to combine, align, and simplify their raw datasets (Section 5.2). These challenges suggest an opportunity for tools that (1) offer alternative programming abstractions to express data constraints and (2) infer geospatial data queries and transformations from constraints.

Assistive Tools for Constructing Geospatial Analysis Pipelines

Opportunity 2. Participants could describe the target outputs of their geospatial analyses but struggled to construct pipelines to produce them (Section 5.3). This suggests an opportunity for tools that (1) accept non-code specifications of analysis intent, (2) synthesize analysis programs that satisfy specifications, and (3) support users in editing programs.

Opportunity 3. Participants relied on running operators and manually inspecting outputs to understand operator semantics (Section 5.3.2). This was computationally expensive and time-consuming, suggesting an opportunity for tools that surface information on operator semantics without requiring execution across entire inputs.

Reproducible, Shareable Geospatial Workflows

Opportunity 4. Participants using GISs struggled to create reproducible, shareable geospatial workflows (Section 5.4.2). Limitations in existing history interfaces made it difficult to recover information on the current analysis state or revisit past analysis decisions (Section 5.4.1). These struggles suggest opportunities for tools that (1) support efficient search through system history and (2) distill history into a portable and executable representation.

Exploring the Cartographic Design Space

Opportunity 5. Participants wanted to visualize their geospatial data using multiple cartographic representations, but transitioning between representations required engineering each one from scratch (Section 5.5.1). This suggests an opportunity for cartographic design tools that reduce the viscosity [8] of switching between map types.

Opportunity 6. Many participants used direct manipulation design software to visualize geospatial data. These tools discard all geographic information, making it difficult to refactor an analysis once visualization work has begun (Section 5.5.2). This suggests an opportunity for tools that (1) bridge geospatial analysis and cartographic design and (2) maintain the underlying geospatial data representation of graphical elements while supporting direct manipulation.

Design Opportunities

Opportunity. Cartographic design tools could focus on **reducing the "viscosity"** of **map type transitions**.

vega-lite

plot

ggplot2

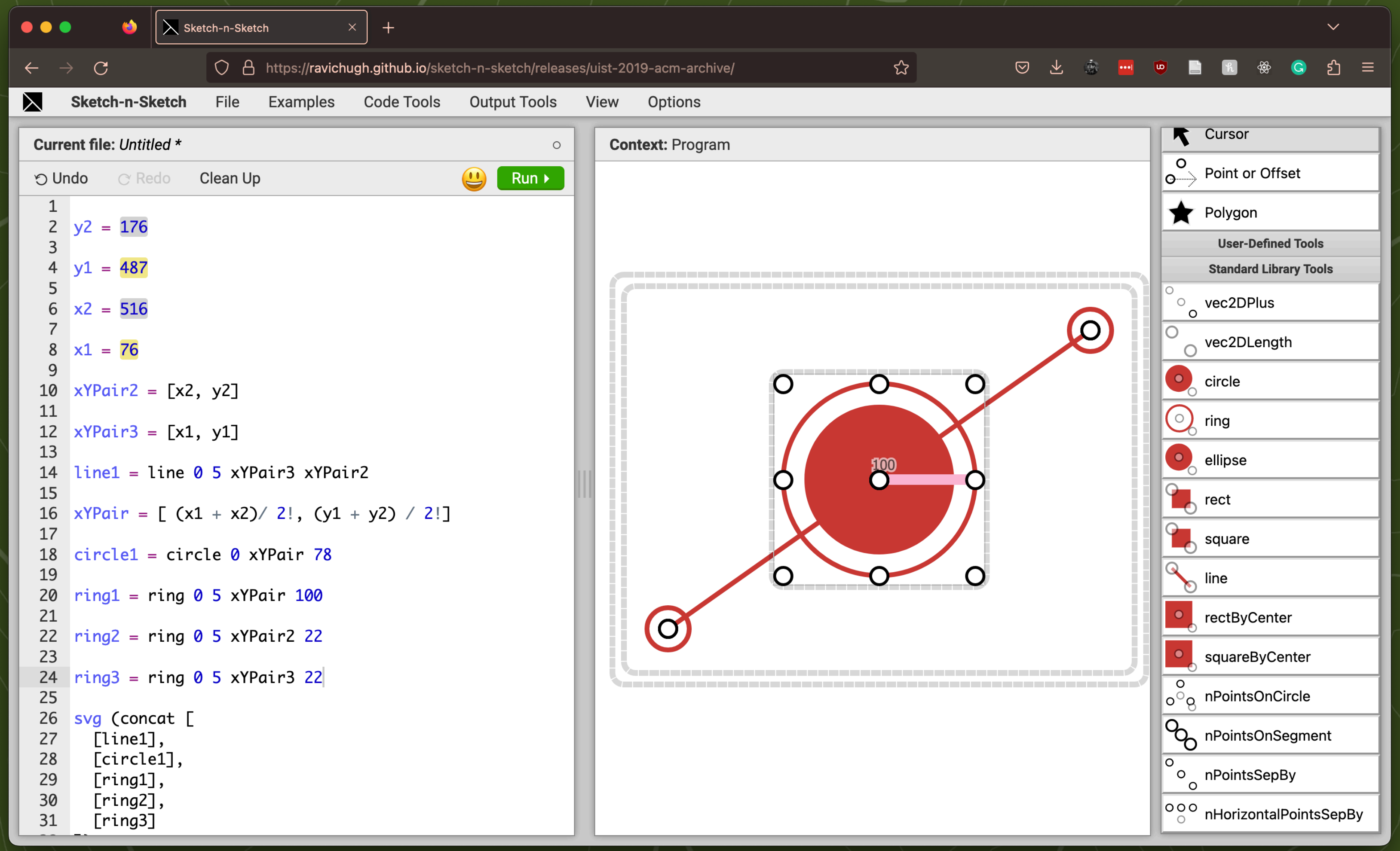
- Restrict **geospatial file formats**, **data models**, and **map types**

⇒ Could not express many of the maps participants made

Possible Solution. *Grammar of Graphics*

Design Opportunities

Opportunity. Cartographic design tools could **pair programmatic and direct manipulation paradigms** for map construction.



Sketch-n-Sketch

- Edit **source** or **output** and propagate edits **bidirectionally**

⇒ Design maps using **direct manipulation** while giving access to **program representations**

Roadmap

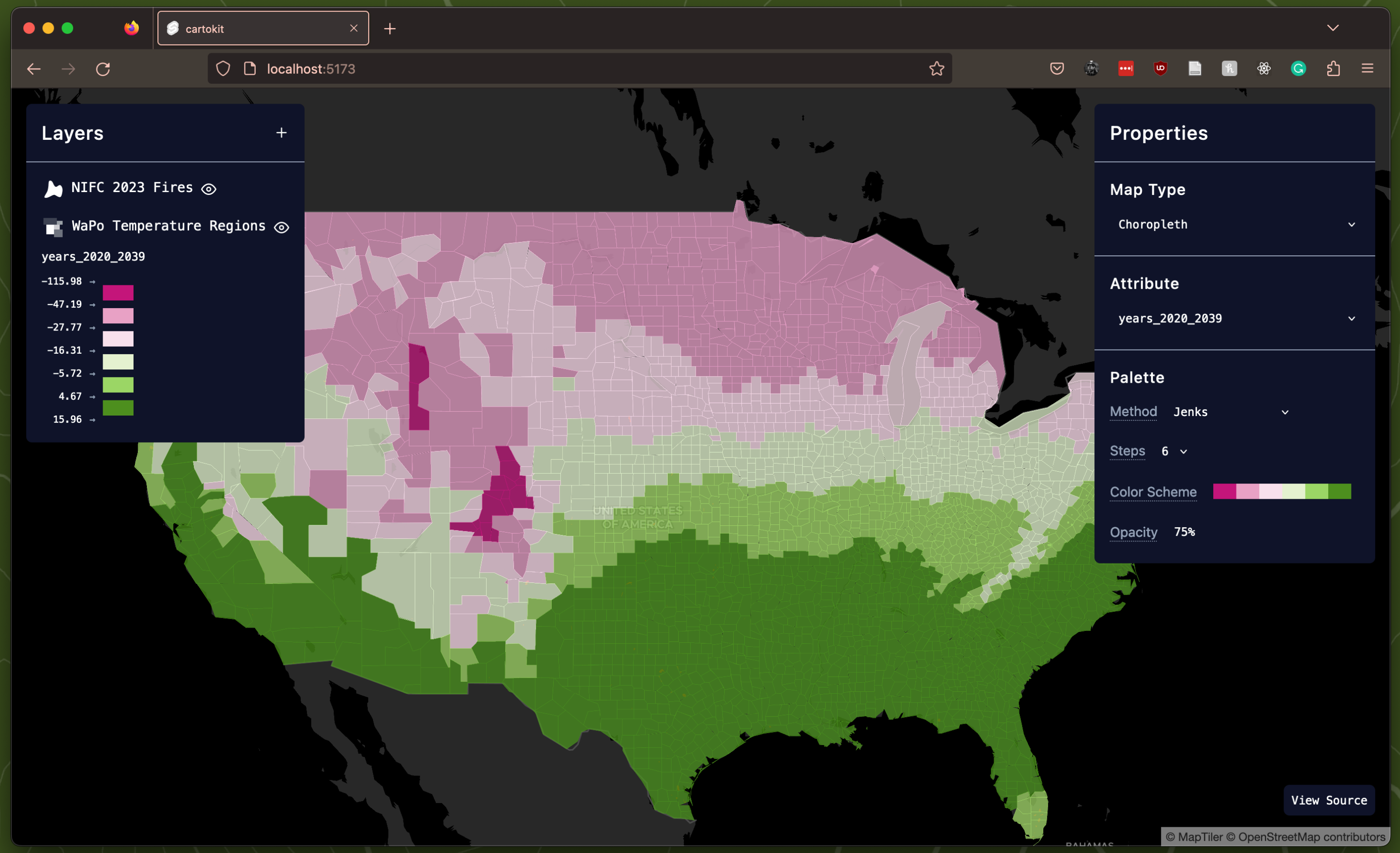
1. Study Design
2. Findings
3. **Design Opportunities**
4. Introducing `cartokit`

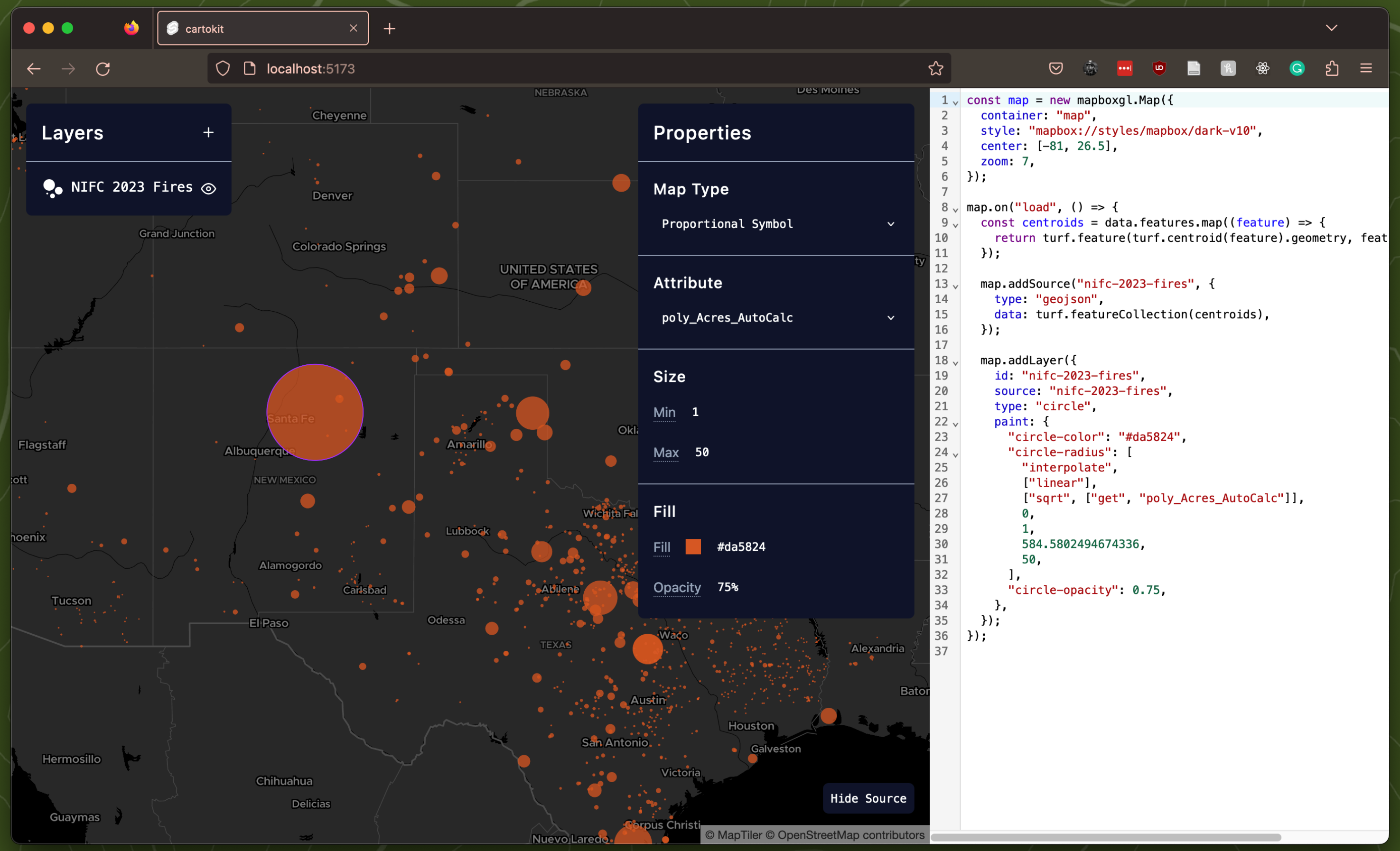
Roadmap

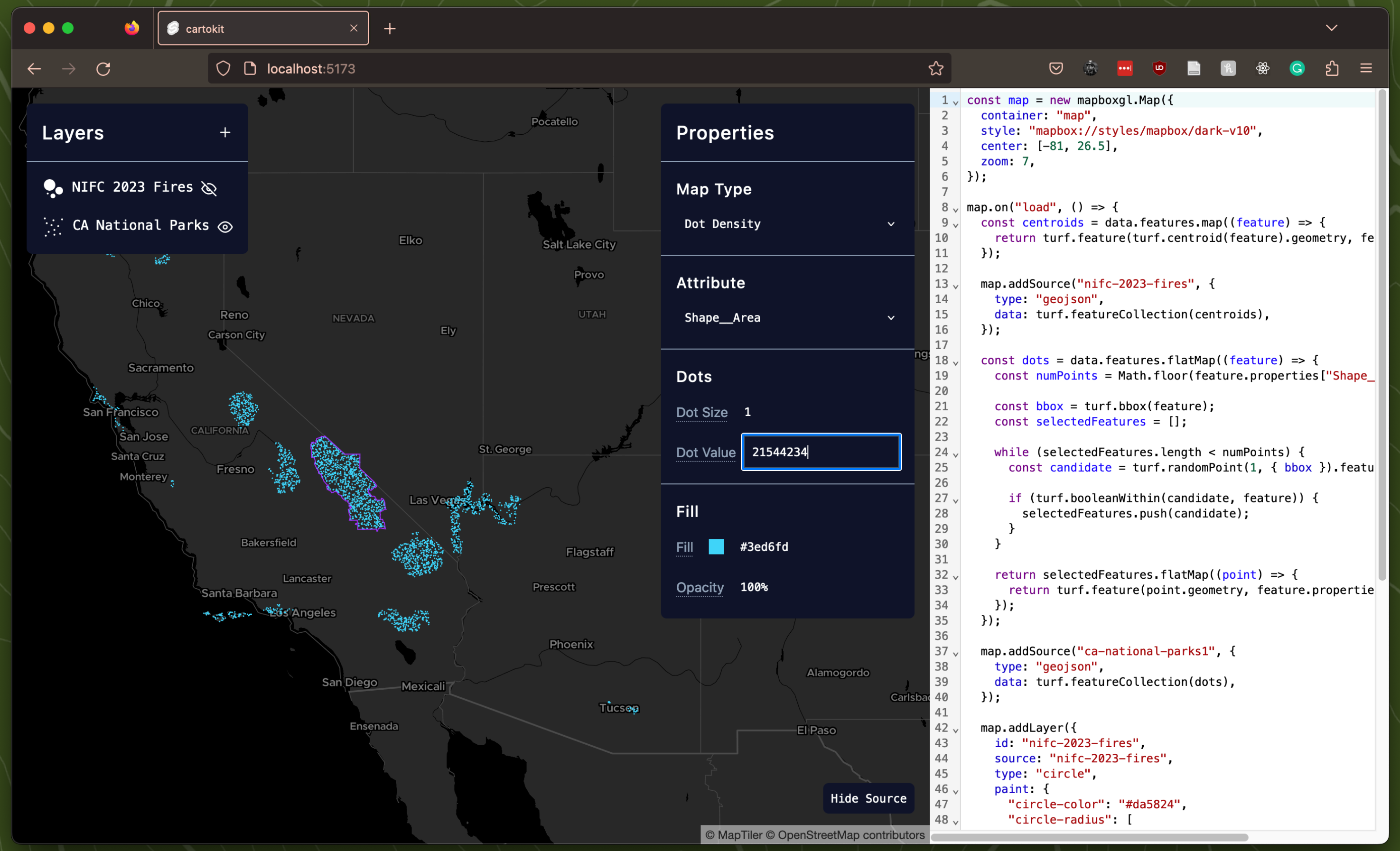
1. Study Design
2. Findings
3. Design Opportunities
4. **Introducing** `cartokit`

Introducing `cartokit`

A **direct manipulation programming system** for **interactive cartography** on the web.

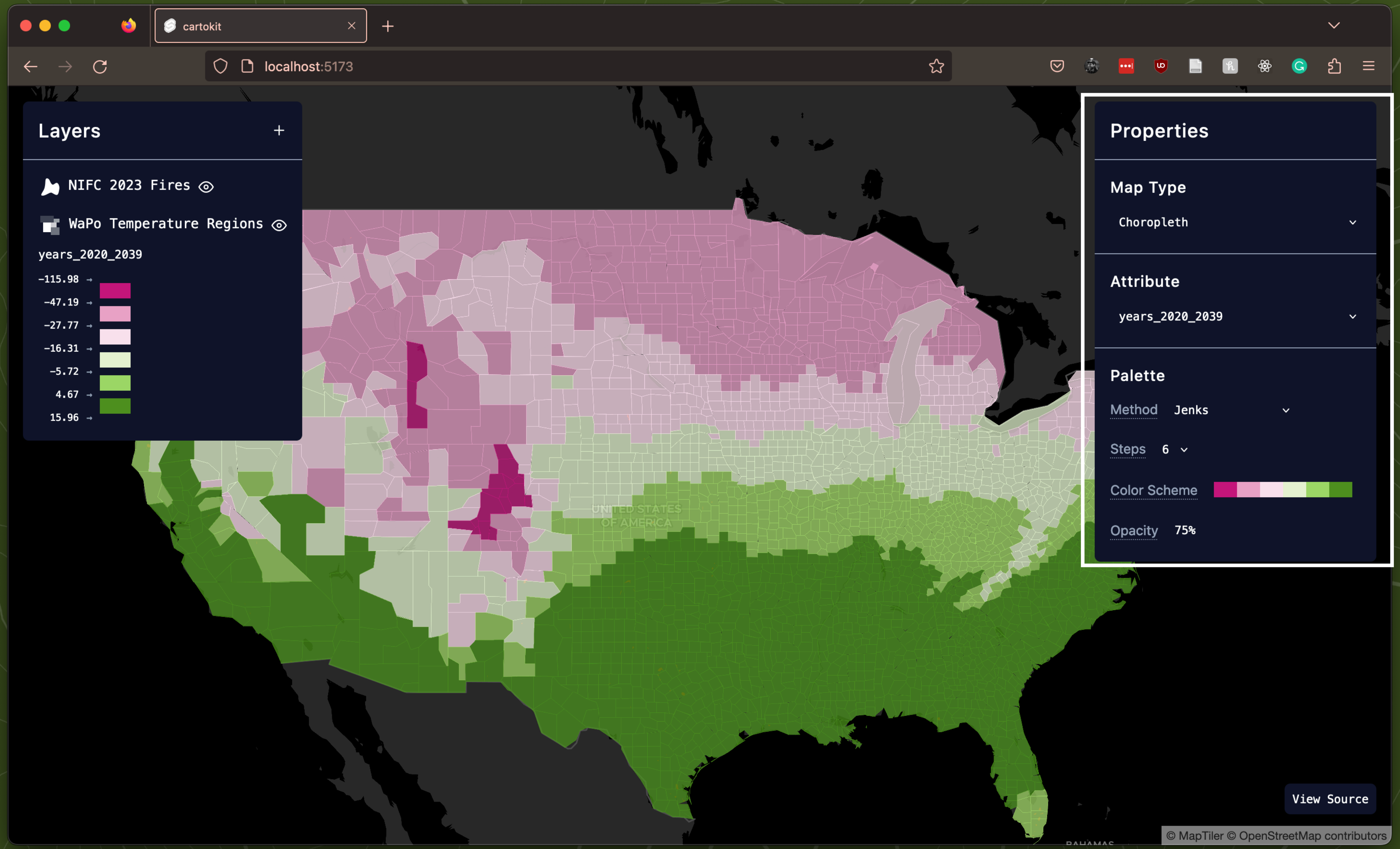




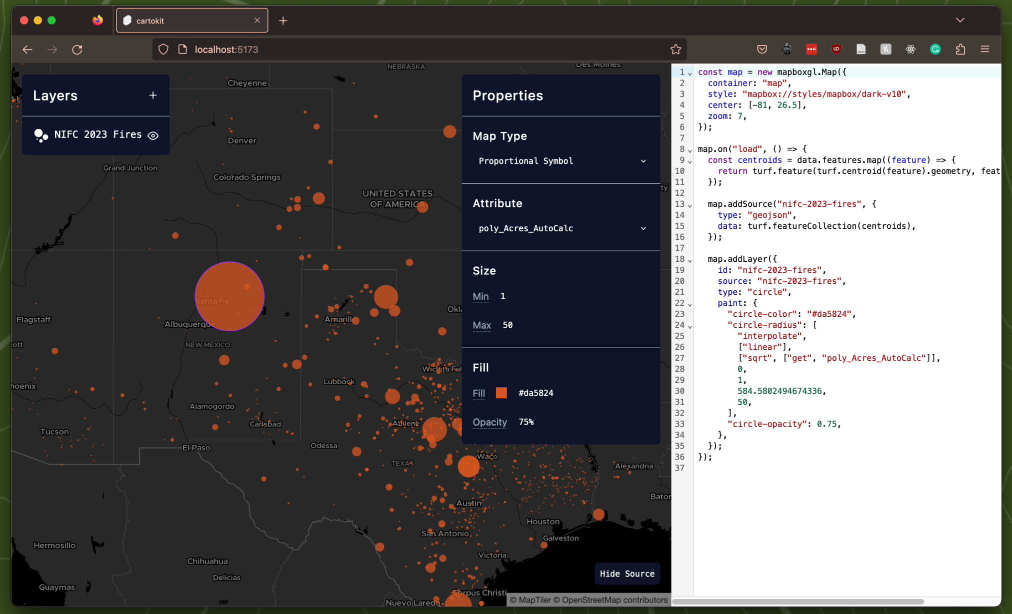


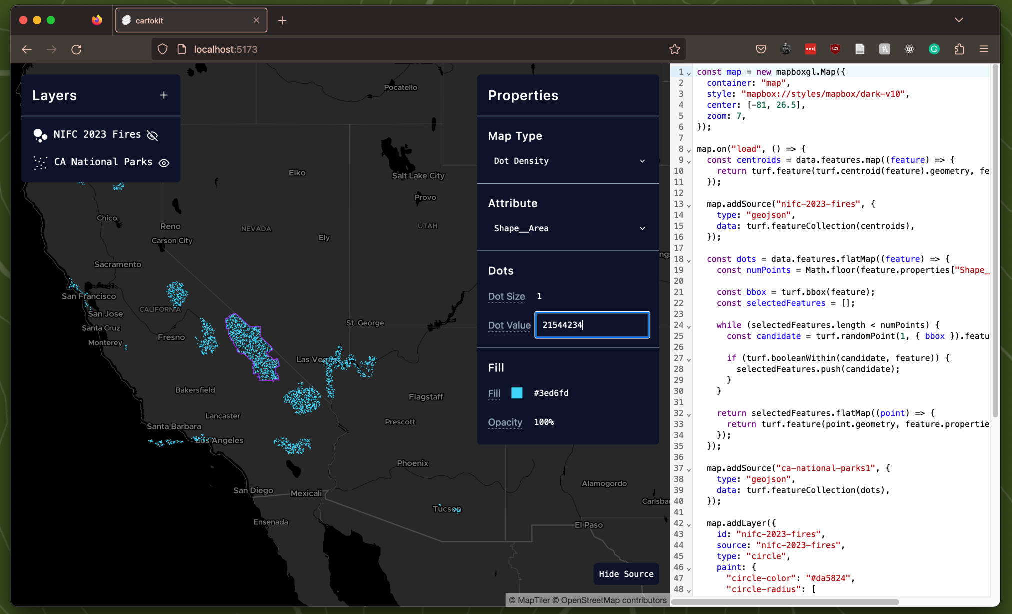
Introducing `cartokit`

A **direct manipulation programming system** for **interactive cartography** on the web.



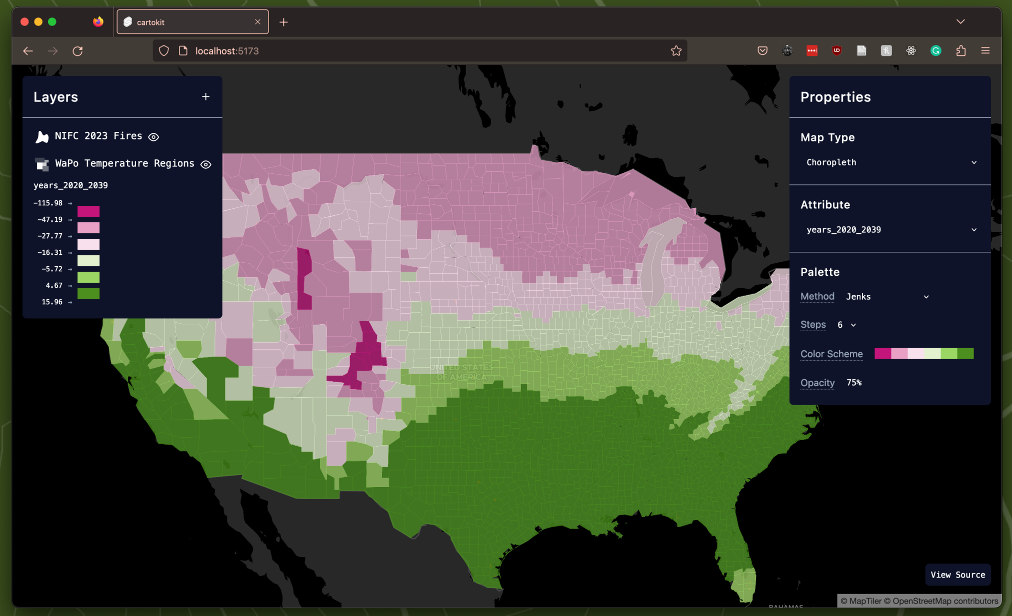
Edit programs through direct manipulation of the output

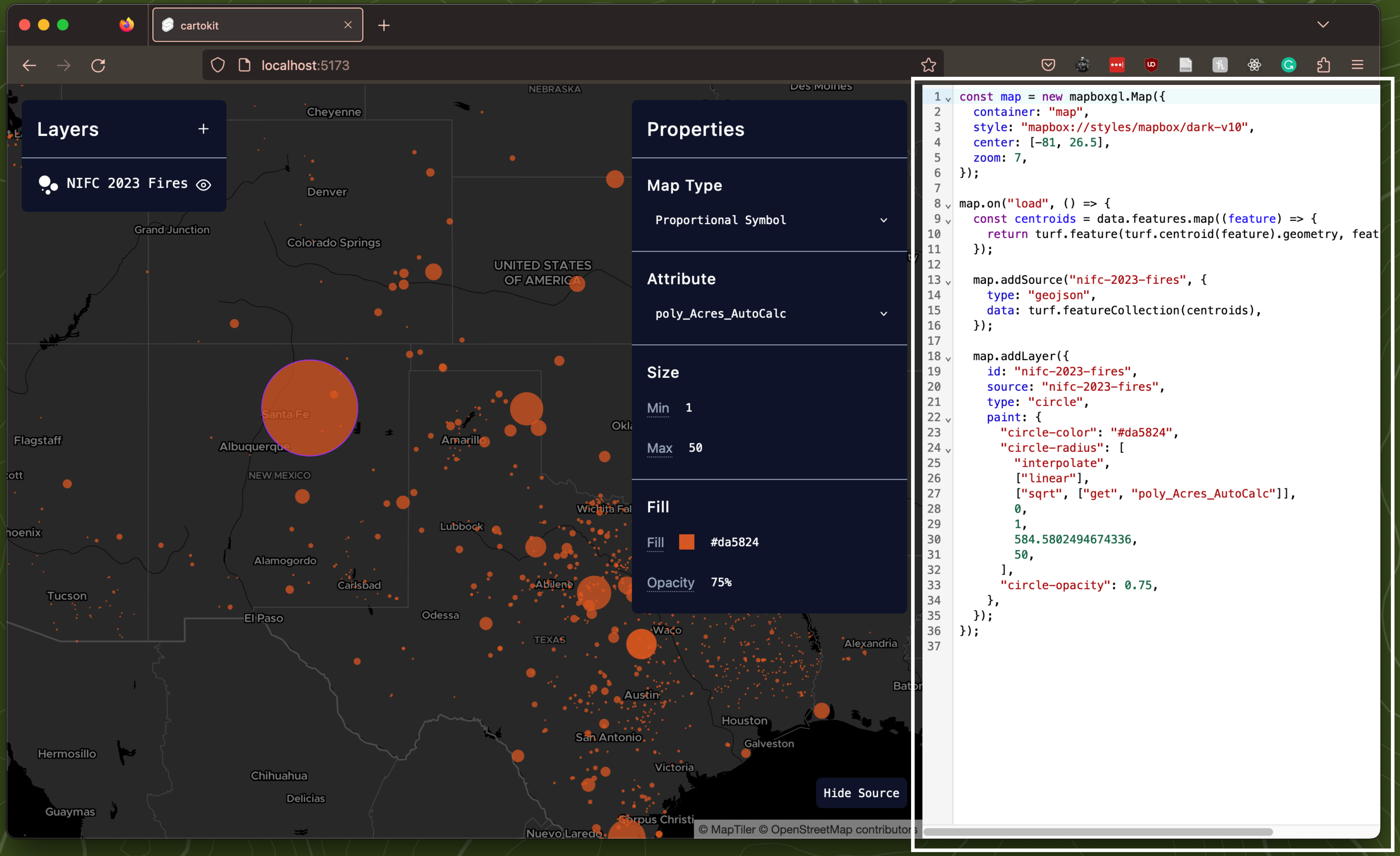




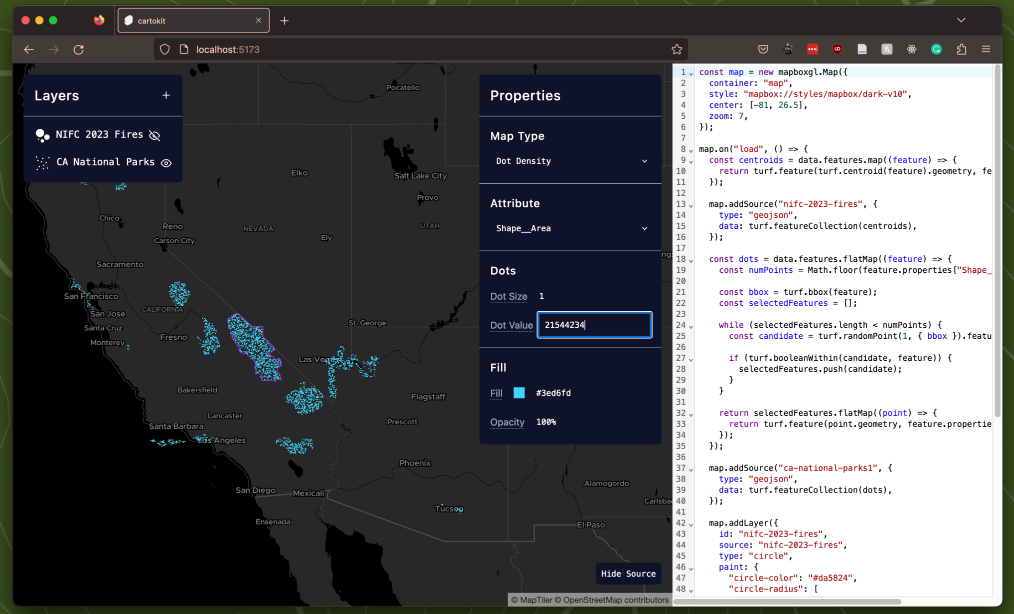
Introducing cartokit

A **direct manipulation programming system** for **interactive cartography** on the web.



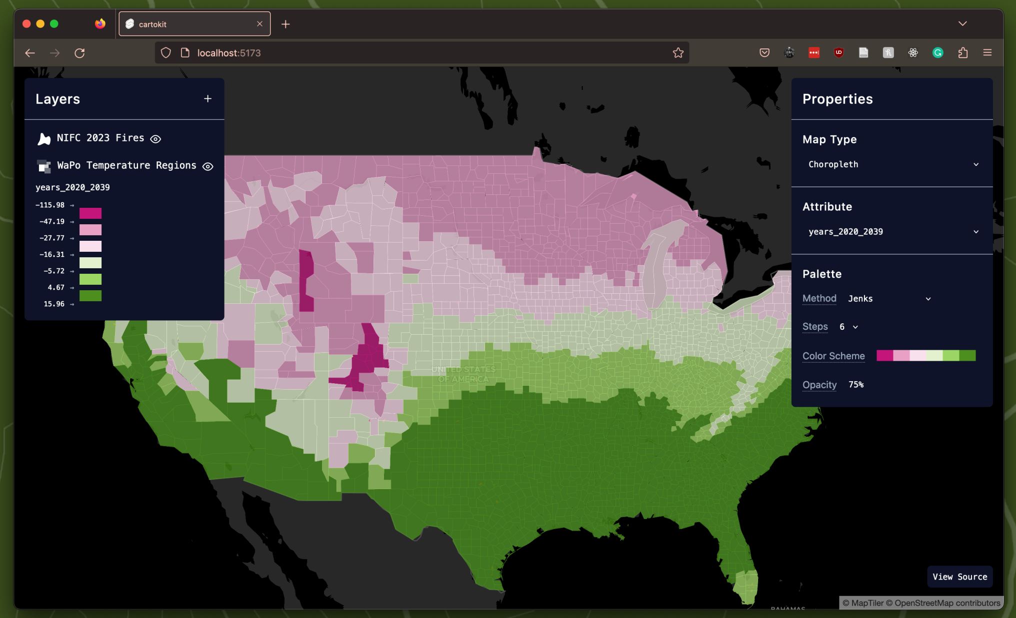


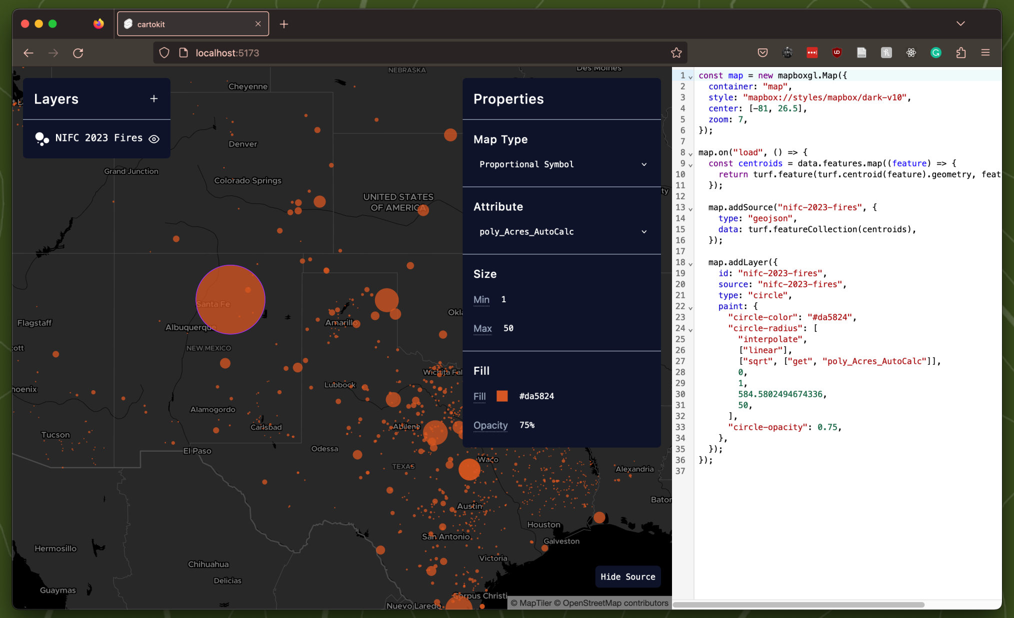
Synthesize the program to produce the output map



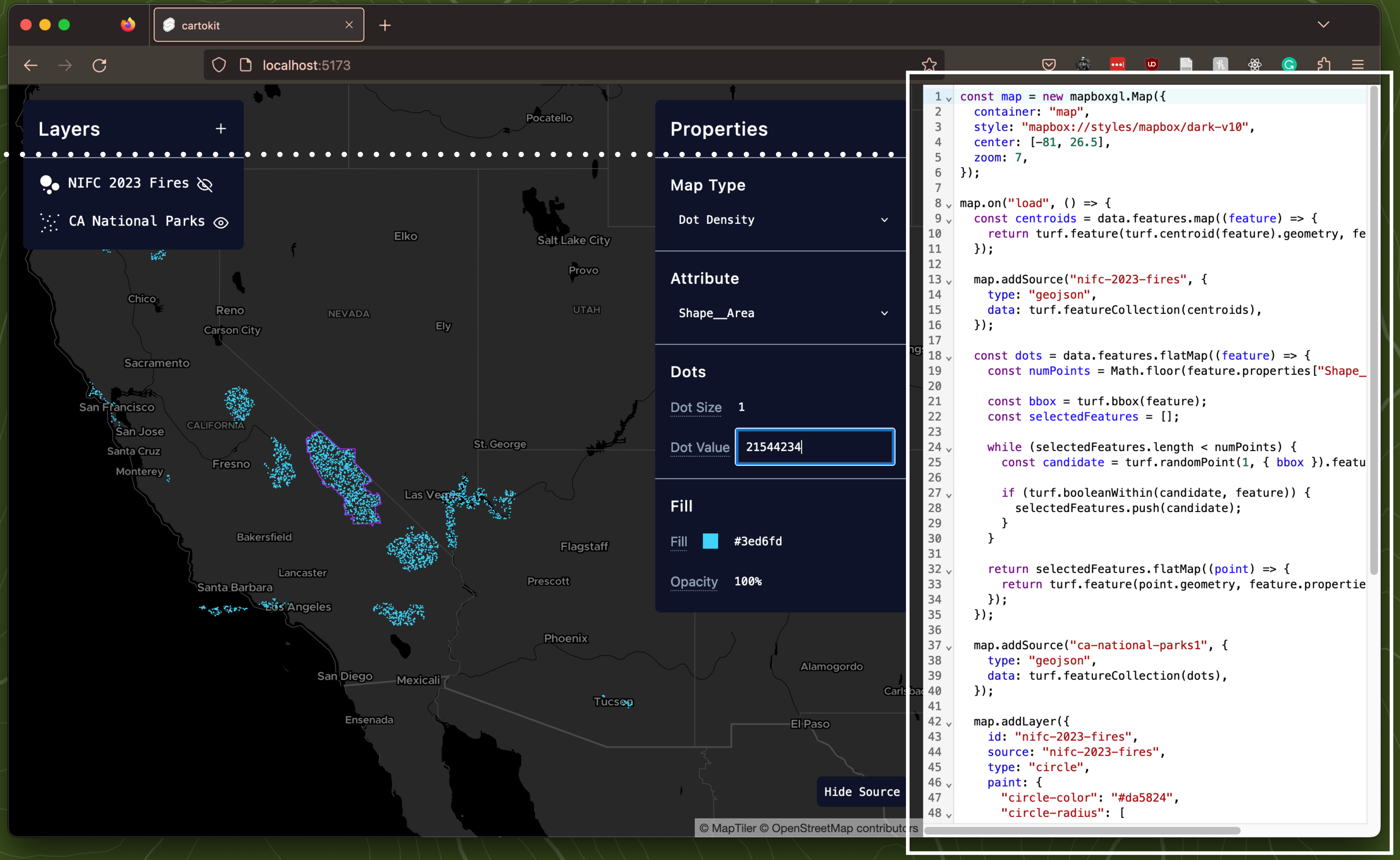
Introducing cartokit

A **direct manipulation programming system** for **interactive cartography** on the web.





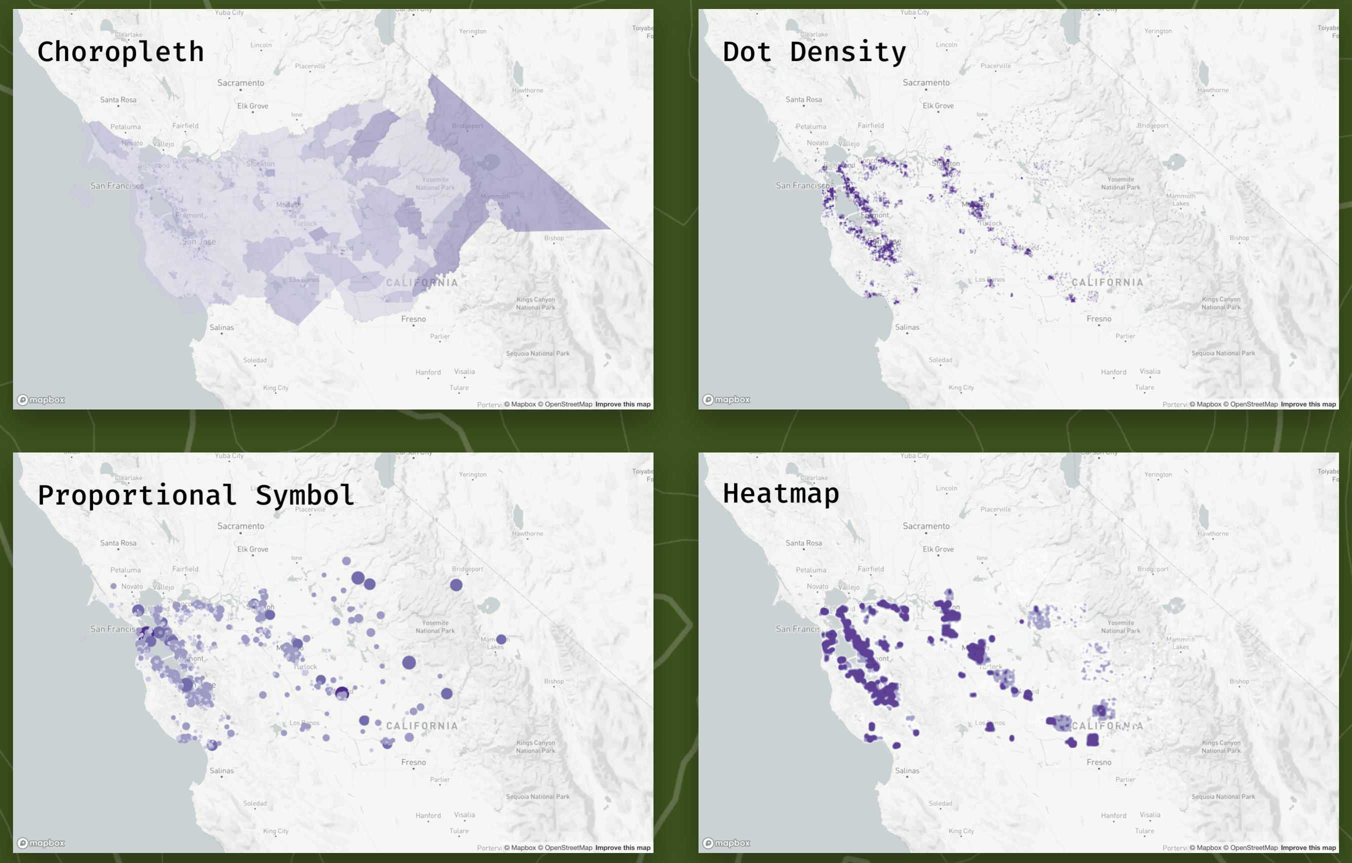
Support sophisticated geospatial data transformations

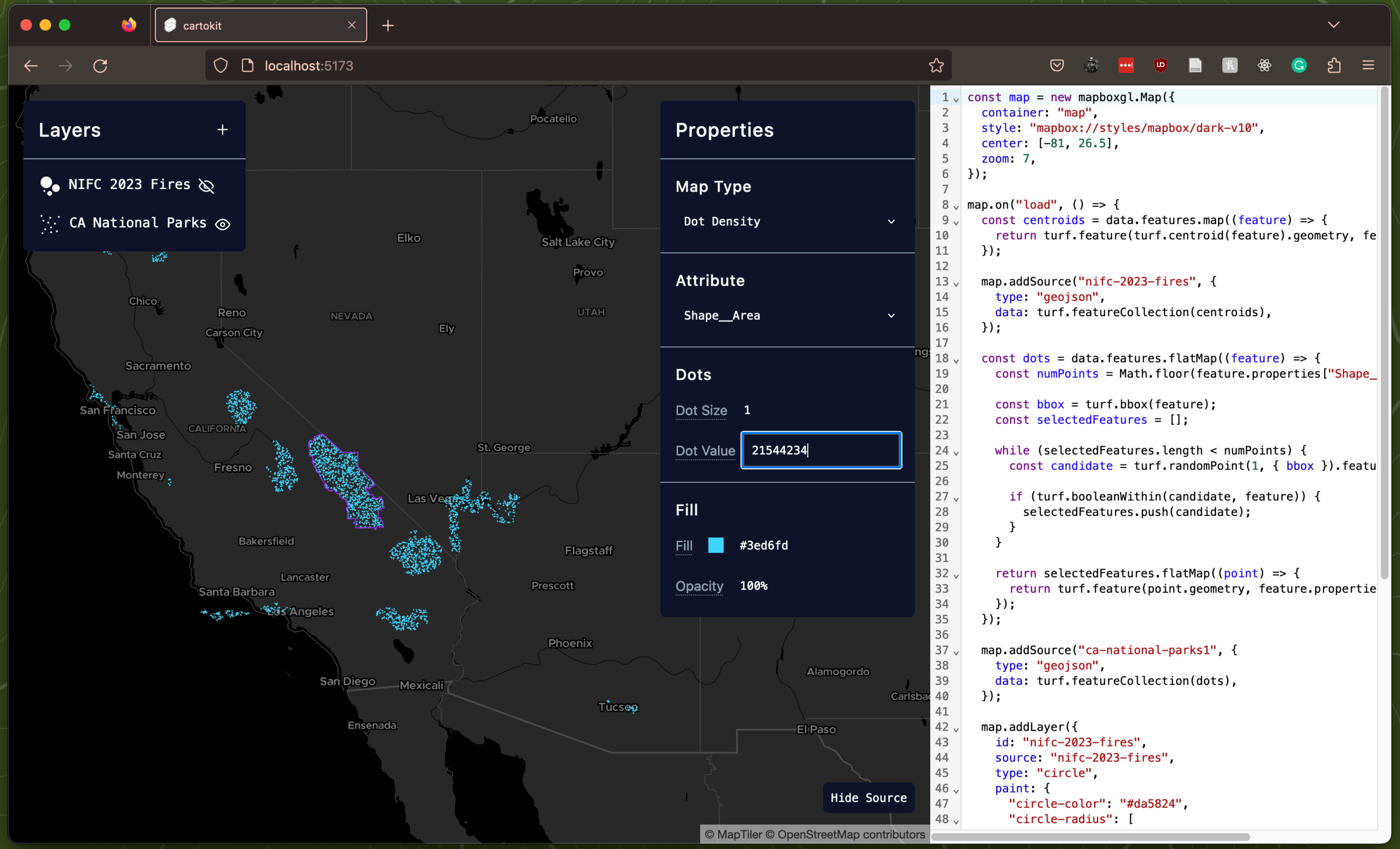


Introducing `cartokit`

A **direct manipulation programming system** for **interactive cartography** on the web.

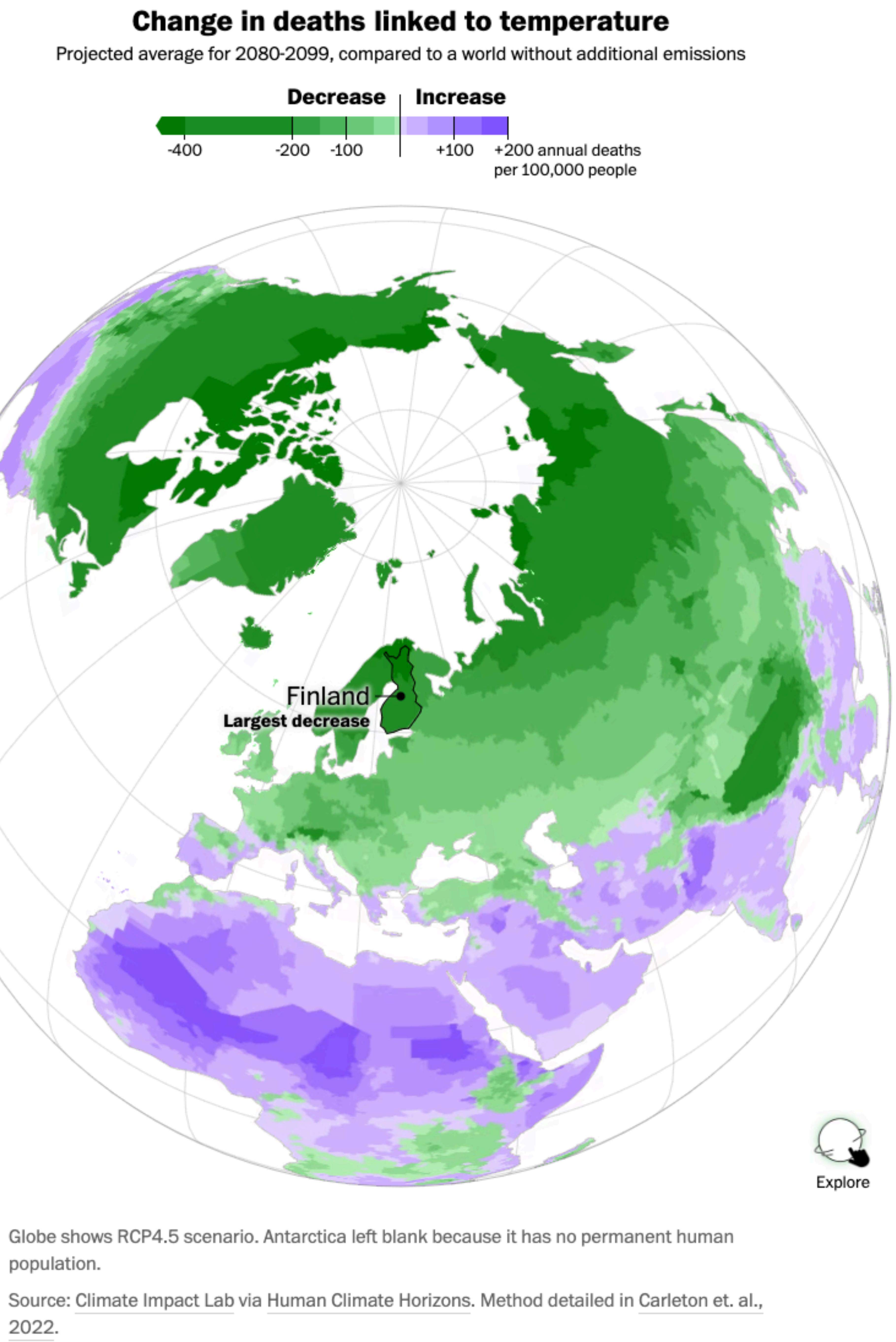
Core Challenge. Transforming data to achieve different cartographic representations.





Demo

Can we reproduce this graphic[1] from the Washington Post using `cartokit`?



1. **Will global warming make temperature less deadly?** Washington Post https://www.washingtonpost.com/climate-environment/interactive/2023/hot-cold-extreme-temperature-deaths/.