A Need-Finding Study with Users of Geospatial Data

Parker Ziegler
peziegler@cs.berkeley.edu
https://parkie-doo.sh/

Sarah E. Chasins
schasins@cs.berkeley.edu

Check out the paper
Ok, but hold up, Parker. What is geospatial data? (And why should we 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

location
Geospatial data is everywhere today.
Background

Domain Experts and Geospatial Data
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.

Geographic Information Systems

- Require significant background in geospatial data theory

- HCI research\(^1\), \(^2\), \(^3\) has shown that GISs are especially difficult for non-geographers to learn and use.

Barriers to working with geospatial data are high.

Programming Systems

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

  mapbox

  geopandas

  sf

• Must develop proficiency with programming languages and environments
Research has yet to explore the specific obstacles *domain experts* face in their work with geospatial data.
The goal of this research is to identify the computing needs of domain expert geospatial data users.
Roadmap

1. Background
2. Study Design
3. Findings
4. Design Opportunities
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 on the 29 hours of video recordings
Roadmap

1. Background
2. Study Design
3. Findings
4. Design Opportunities
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.

<table>
<thead>
<tr>
<th>Spatial Resolution</th>
<th>Temporal Interval</th>
<th>Geographic Extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOD16</td>
<td>500m</td>
<td>Global</td>
</tr>
<tr>
<td>PRISM</td>
<td>4638.3m</td>
<td>Conterminous U.S.</td>
</tr>
<tr>
<td>USDA-NASS</td>
<td>30m</td>
<td>Conterminous U.S.</td>
</tr>
<tr>
<td></td>
<td>8-Day</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monthly</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yearly</td>
<td></td>
</tr>
</tbody>
</table>
Aligning Geospatial Datasets

**PE2’s Task.** Develop a model to predict groundwater withdrawal.

<table>
<thead>
<tr>
<th>Spatial Resolution</th>
<th>Temporal Interval</th>
<th>Geographic Extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>500m</td>
<td>8-Day</td>
<td>Global</td>
</tr>
<tr>
<td>4638.3m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4638.3m</td>
<td>Monthly</td>
<td>Conterminous U.S.</td>
</tr>
<tr>
<td>30m</td>
<td>Yearly</td>
<td>Conterminous U.S.</td>
</tr>
</tbody>
</table>

MOD16
PRISM
USDA-NASS

Next, consider the resampling for these datasets.
Aligning Geospatial Datasets

**PE2’s Task.** Develop a model to predict groundwater withdrawal.
Aligning Geospatial Datasets

PE2’s Task. Develop a model to predict groundwater withdrawal.
Aligning geospatial datasets required participants to have significant fluency in geospatial data theory in addition to contextual information about the datasets themselves.

Identify the correct sequence of transformations among hundreds of operators

Determine when selected transformations produced undesirable results
## Findings

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

<table>
<thead>
<tr>
<th>Data Discovery</th>
<th>Data Transformation</th>
<th>Analysis</th>
<th>Analysis Representation</th>
<th>Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solving Geospatial Data Constraints</td>
<td>Aligning Geospatial Datasets</td>
<td>Identifying Geospatial Operators</td>
<td>Reproducing Geospatial Analyses</td>
<td>Sketching Cartographic Variants</td>
</tr>
<tr>
<td></td>
<td>Topological Errors</td>
<td>Understanding Geospatial Operator Semantics</td>
<td>Creating Informal Program Representations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reducing Resolution to Improve Performance</td>
<td>Visibility of Geometry in Programming Environments</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data Subsetting and Caching</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We identified **12 challenges** across **five phases** of participants’ work with geospatial data.
Sketching Cartographic Variants

Participants wanted to visualize their data using many different cartographic representations.

• 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.
Sketching Cartographic Variants

Producing most map variants required going through the entire analysis and visualization pipeline.
Participants tried to speed up the drafting process in creative ways. One common technique involved screenshotting in-progress maps.
Sketching Cartographic Variants

Screenshots
Sketching Cartographic Variants

Screenshots ● Layouts

Allowed PJ6 to compare cartographic choices “before I code anything.”
Screenshotning 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. Background
2. Study Design
3. Findings
4. Design Opportunities
Roadmap

1. Background
2. Study Design
3. Findings
4. Design Opportunities
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 5.2). While many could describe their constraints audibly, 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 analyses 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 QGIS 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.

Opportunity 5. Participants wanted to visualize their geospatial data using multiple cartographic representations, but transitioning between representations required engineering each one from scratch (Section 5.3.3). This suggests an opportunity for cartographic design tools that reduce the vexancy of switching between map types.

Exploring the Cartographic Design Space

Opportunity 6. Many participants used direct manipulation design software to visualize geospatial data. These tools discarded all geographic information, making it difficult to relate an analysis once visualization work has begun (Section 5.3.2). This suggests an opportunity for tools that (1) bridge geospatial analytics 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.

• 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.

- Edit source or output and propagate edits bidirectionally
  ⇒ Design maps using direct manipulation while giving access to program representations
A Need-Finding Study with Users of Geospatial Data

Learn about all 12 challenges, all six design opportunities, and hear from our participants in the paper.

Parker Ziegler
peziegler@cs.berkeley.edu
https://parkie-doo.sh/

Sarah E. Chasins
schasins@cs.berkeley.edu

Topography of Mt. Tamalpais, Marin County, USA