Opening GIScience: A Process Based Approach

Jerry Shannon, University of Georgia
Kyle Walker, Texas Christian University

This is an Accepted Manuscript of an article published by Taylor & Francis in the International Journal of Geographic Information Science in May 2018, available online:
https://www.tandfonline.com/eprint/xxFcFXA8Dh57hvE9YZKJ/full

Abstract

Many scholars have demonstrated growing interest in GIScience in recent years, including use of open data portals, shared code, and options for open access publication. These practices have made research and data more transparent and accessible for a broad audience. For populations without expertise in the technology and methods undergirding these data, however, this research may be open only in a limited sense. Based on two case studies using RStudio's Shiny web platform, we argue that a process based approach focusing on how analysis is opened throughout the research process provides a supplementary way to define and reflect upon public facing geographic research. Reflecting upon decisions we made at key points in each case study project, we identify four key tensions inherent to work in open GIScience: standardized vs. flexible tools, expert vs. community led design, single vs. multiple audiences, and established vs. emerging metrics.

Keywords

Open source; Public Participation GIS; Census data; Internet GIS

1. Introduction

Open GIScience is an emerging alternative to conventional research methods. As described by Singleton et al. (2016), open GIScience is characterized by open source software, sharable datasets and code, and open access publication. Open GIScience is supported by the development of improved technologies for sharing and analyzing geospatial data, including open source tools (e.g., R, QGIS) and visualization tools (e.g., Leaflet) as well as code sharing sites (e.g., GitHub). These tools and practices allow for broader collaborations and analysis that is more transparent.

The growth of open GIScience provides significant opportunities for collaborative research models, but thus far the focus of this work has been on the analysis stage of research, primarily on best practices for sharing data and reproducible methods. Open GIScience often makes use of software and analytical methods that often assume significant training and resources for users (Gurstein 2011, Kitchin 2014), including familiarity with code and often complex forms of data visualization. This can complicate efforts to engage with a broader public lacking in technical expertise. Incorporating a stronger public engagement into open GIScience has at least the potential to provide multiple benefits, informing data analysis and interpretation or facilitating use of data by policy makers and key stakeholders. This is evident in ongoing work in participatory geospatial research and citizen science (Sieber *et al.* 2016). It is thus important to consider how open GIScience can be combined with strategies for public engagement.

In this paper, we build on past work to develop a process based approach for defining and assessing an engaged, open GIScience. Our analysis is based on two case studies, both using the open source data dashboard software Shiny, developed for the R statistical package. In one case, Shiny facilitated interaction with community generated housing data. In the other, Shiny provided an interface to explore a demographic dataset used in a traditional research paper. Reflection on our process shows the complexity of open GIScience in practice when working with public stakeholders. Our conclusion identifies four key tensions inherent to work in open GIScience, ones that must be negotiated rather than resolved. By focusing on the process of *opening* GIScience, we argue that researchers, policy makers, and other stakeholders can better pursue the goal of more transparent, engaged, and collaborative research.

2. Open GIScience

In recent years, many authors have emphasized the need for open and reproducible research practices, often referred to as an open science framework (OSF, 2018). Munafo et al. (2017) argue that the lack of open scientific practices harms the credibility of scientific research by limiting transparency and reproducibility. As a solution, these authors provide a list of best practices, including improved public access to data (where appropriate), independent oversight, and greater use of preprints and replication studies. In the journal *Science*, Nosek et al. (2015) present eight standards for journals seeking to encourage open research practices, including transparency of data and research materials, study preregistration, and code transparency. New forums for sharing data and ongoing research, such as the collaborative coding service Github (github.com) or preprint services such as SocArXiv (https://osf.io/preprints/socarxiv) provide additional tools to implement these suggestions.

Still, open science practices are rare. A metastudy published in *PLoS Biology* found that of 268 sampled empirical articles, none had access to the underlying raw data, and only three provided information about how these data could be obtained. Less than two percent of articles were explicitly framed as replication studies (Iqbal *et al.* 2016). Additionally, a survey of 1,576 scientific researchers conducted by the journal *Nature* and published in 2016 found that 90% felt there was a significant or slight crisis in study replicability (Baker & Penny 2016).

While the majority of published work on open science has taken place in the hard sciences, some authors have suggested practices for an open GIScience. This shift is part of a broader “digital turn” in geography that includes related work with volunteered geographic information (VGI) and crowdsourcing (Johnson and Sieber 2013, Ash *et al.* 2015, Brovelli *et al.* 2015a). Singleton et al. (2016) outline one vision for open GIScience, advocating for the use of open source software and scripting languages, repositories for both data and code that make research “scrutable,” and journal practices requiring greater transparency from authors (see also Rey 2014). Open access publishing specifically provides new forums for both vetting and sharing scholarly research (Van Noorden 2013).

Open GIScience can also incorporate projects outside of scholarly research. For example, many large municipalities have championed smart city initiatives that include open data portals and data dashboards

(Townsend 2013, Kitchin 2014, Roche 2015). Other initiatives involved residents and stakeholders in the data collection process through mobile applications (Brovelli 2015b). By allowing “civic hacking” (Townsend 2013) by residents and local organizations, proponents of these approaches argue that they make governance more transparent, data driven, and democratic.

Certainly, open GIScience and open data initiatives allow for new flows of data and tools that would have been impractical prior to the rapid growth of digital technologies and the geoweb, the “suite of geographically related services and locationally aware devices” (Sieber *et al.* 2016). Yet as several authors have noted, the new networks of knowledge production and consumption enabled through these efforts have their own boundaries and exclusions. Drawing from this work, we identify two limitations common to open data projects.

First, open GIScience, whether academic or public, primarily engages individuals with significant levels of technological and social capital. Kitchin (2014), for example, cites the reinscription of technocractic modes of governance as an effect of smart cities initiatives, privileging those who can read, analyze, and tell stories about available civic data and reinforcing the panoptic power of the state. Even when engaging with a diverse public, open GIScience practices are not inherently liberatory (Leszczynski 2014). Online and digital initiatives may blur boundaries between experts and non-experts, but this can lead to highly divergent outcomes, from neoliberal cooptation of citizen activism to the transformation of state policies and institutions (Haklay 2013, Brandusescu *et al.* 2015). Recognizing these limitations, Gurstein (2011) argues for a focus on the “effective use” of open data which would “include such factors as the cost and availability of Internet access, the language in which the data is presented, the technical or professional requirements for interpreting and making use of the data, and the availability of training in data use and visualization” (p. 3). Increasing reliance on VGI and citizen science initiatives for data production and analysis prioritizes those who have “the luxury of volunteering” (Sieber *et al.* 2016) and can depend heavily on contextual factors such as familiarity with technology or urgency of the research (Brovelli *et al.* 2015a).

Second, rather than treating software and algorithms as neutral tools, multiple authors have urged attention to the context and purposes that shape coding and software deployment. As Dalton (2015) writes, research on open data and digital geographies more broadly should understand how “social relations are materially condensed in a technology, optimizing it for a limited set of purposes” (p. 1032). Similarly, Haklay (2013) describes four different levels of technological hacking, ranging from making use of an existing application for new data collection to recoding applications themselves to create entirely new platforms. The latter, he argues, allows for “deep democratization,” where community members’ concerns and perspectives can be designed into the system. Following Kitchin and Dodge’s (2014) argument that digital technologies are productive of new spaces and social orders, deep hacking has greater potential for transformative social change. At the same time, it also requires high levels of technical expertise, making the balance between technological sophistication and community engagement a delicate one.

With these two problematics in mind, it is perhaps helpful to define open GIScience not just through a set of specific *characteristics*—open code and publicly available data, for example—but also as an ever evolving *process* through which researchers seek out methods that are actively transparent and inclusive for a range of audiences. Doing so means using "open" as both verb and adjective, critically examining how GIScience is opened as well as defining its characteristics. This focus on process follows earlier work in public participatory GIS (Sieber 2006), moving away from definitions of geospatial research that focus on a "set of topics and propositions" and toward a more reflexive critical GIS praxis (Elwood and Wilson 2017, p. 4). A process focused approach requires consistent attention to the practices and tools through which various publics are brought into the research process. It views data collection, analysis, and sharing as always socially situated, embedded in processes that both reflect and reinscribe social norms and structures.

In this paper, we model a process based approach by reflecting critically on the tradeoffs and choices made in two open GIScience projects, shaped by the data we used, the software coding process, and interactions with diverse audiences. Our projects each have an explicitly public focus, unlike open science approaches outlined above oriented around academic knowledge production and in line with previous work in PPGIS. We focus on three key moments in the research process: creating tools, opening data, and evaluating impact. At each stage, we draw from two case studies to examine how and where data was opened to a broader public. The first is a project to collect and analyze data on housing conditions in several small cities in Georgia. The second is a data dashboard providing users the chance to interact with data presented in a published peer-reviewed journal. Through examining how these tools were constructed, we identify key tensions and limitations we faced in this process.

3. Case studies: Opening GIScience with Shiny

In this section, we provide background on the Shiny platform (RStudio 2016), the software tool used in both of our projects. We also provide background on our two case studies. We chose these based on their contrasting purpose and audiences, which helps triangulate our findings. The housing application (section 3.2) facilitates data exploration for community generated data within specific local communities. In contrast, the urban diversity application displays the results of analysis using secondary data, supplementing a published research article.

3.1 Shiny

Both case studies described in this article make use of the Shiny software package (RStudio 2016). Originally created by the private company RStudio, Shiny's code is open source, allowing for community input and improvements, though the majority of edits have been made by just six individuals (RStudio 2017a). Shiny utilizes the open source R software package to provide online, interactive data visualization. It does so by drawing on R's existing tools for visualization, including a variety of HTML widgets for graphing libraries such as Plotly or the mapping library Leaflet. By using a variety of input tools such as checkboxes, dropdown menus, and slider bars, users can select and subset variables for visualization. Shiny thus allows end users with minimal programming experience to interact with complex, large datasets, extending R's ability to facilitate exploratory data analysis (Tukey 1977, Wickham and Grolemund 2016). Figure 1 shows a simple example of a Shiny application for viewing different types of data distributions.

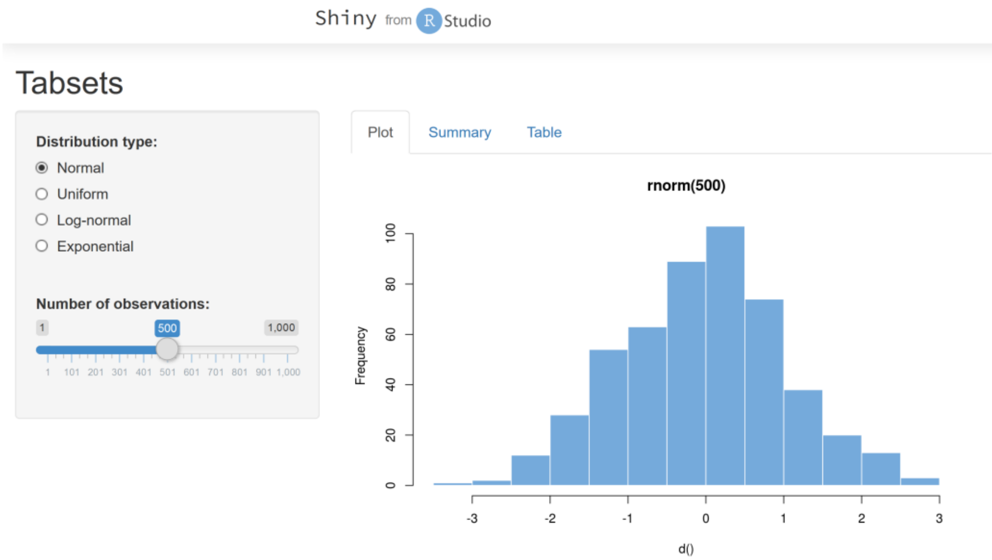

Figure 1: A simple Shiny application allowing users to explore different types of data distributions. (https://shiny.rstudio.com/gallery/tabsets.html)

Shiny is similar to other data dashboard tools such as Tableau (www.tableau.com) or Microsoft's Power BI (powerbi.microsoft.com). Yet Shiny is code based, unlike the point and click GUI environments in those tools. At the same time, compared to pure code environments such as D3 (https://d3js.org), Shiny's reliance on external libraries means that it requires lower levels of technical expertise. It thus falls somewhere in the middle of the continuum of tools for data exploration and analysis. In addition, compared to other tools for geovisualization such as Google Maps, Shiny's reliance on code and its customizability essentially hard wires hacking into the design process (Dalton 2015). RStudio's decision to host Shiny's code completely on GitHub reflects a more open stance when compared to Tableau or PowerBI, whose GitHub presence is limited to tutorials or program plugins. Lastly, Shiny is largely available free of charge, with the costs restricted to server hosting, unlike Tableau or PowerBI, which have high fees for full software licenses along with less powerful public versions. Our interest in using Shiny stems from the combination of its relative accessibility and its inherent hackability.

3.2 Georgia Initiative for Community Housing

Our first case study involves research done with the Georgia Initiative for Community Housing (GICH) (Tinsley 2017). Run through the University of Georgia, with assistance from the state Department of Community Affairs and private companies, GICH annually enrolls several communities around the state into a three-year program focused on the development of affordable housing. Most GICH communities are quite small, many with 5,000 or fewer residents, and often located in rural areas of the state. Housing teams, comprised of a variety of local stakeholders, attend biannual retreats during their time in the GICH program while also developing plans for refurbishment or redevelopment of their housing.

Preliminary conversations with GICH communities showed widespread interest in tools for conducting digitized community housing assessments. Supported through grant funding from USDA, one of this paper's authors (Shannon) helped digitize an existing paper housing survey for these communities. Adopting a similar process to Brovelli, et al. (2015b), we used the open source survey tool OpenDataKit for teams to collect data on existing housing conditions. These data were then visualized through a Shiny application that allowed team members to identify properties with multiple problems, hot spots of specific types of issues (e.g., rotting wood, damaged roofs), or correlation between housing conditions and underlying census variables. Figure 2 shows an example of this tool, and both the code and sample application are available online.[1]

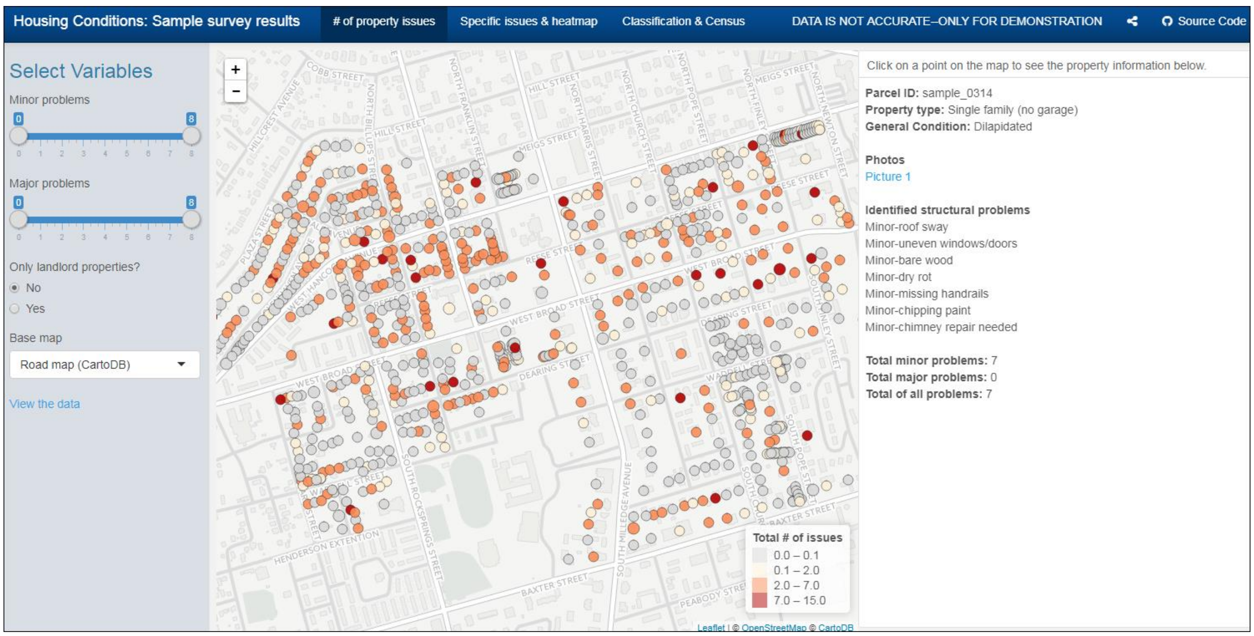

Figure 2: Shiny application for viewing collected data on housing conditions

The goals of this process were to provide communities with low cost ways to both collect and analyze data on community housing. In this case, the Shiny application replaced traditional desktop GIS software, which was either too expensive and/or technically demanding for teams to use. This is an ongoing project, but most communities plan to use these tools to support applications for governmental grants such as HUD's Community Development Block Grant.

This case study draws from conversations with five small communities that are, as of this writing, halfway through the GICH program. They range in population from 2,000 to 14,000 people. We provide pseudonyms for both communities and the housing team members. Spencer, in central Georgia, is a small rural town located next to a historic rail line. Birchwood, near Alabama, benefits from tourism related to nearby state parks. Ashland and Murphy are both at the southern end of Appalachia in the northwest

[1] Available at https://comapuga.shinyapps.io/sampledata_flexdash

corner of the state and are primarily manufacturing towns. Lexington is a small town just outside the edge of Atlanta's suburbs.

3.3 Mapping Urban Diversity

Our second case study concerns a Shiny application developed by the paper's other author (Walker) in support of a research project on the geography of racial and ethnic diversity in US metropolitan areas. The application was designed in support of an article published in that the journal *Urban Studies* that examines the shift of highly-diverse neighborhoods to the suburbs from central cities from 1990 to 2010 (Walker 2016). In turn, the application allows visitors to interactively explore the published research results in more depth.[2]

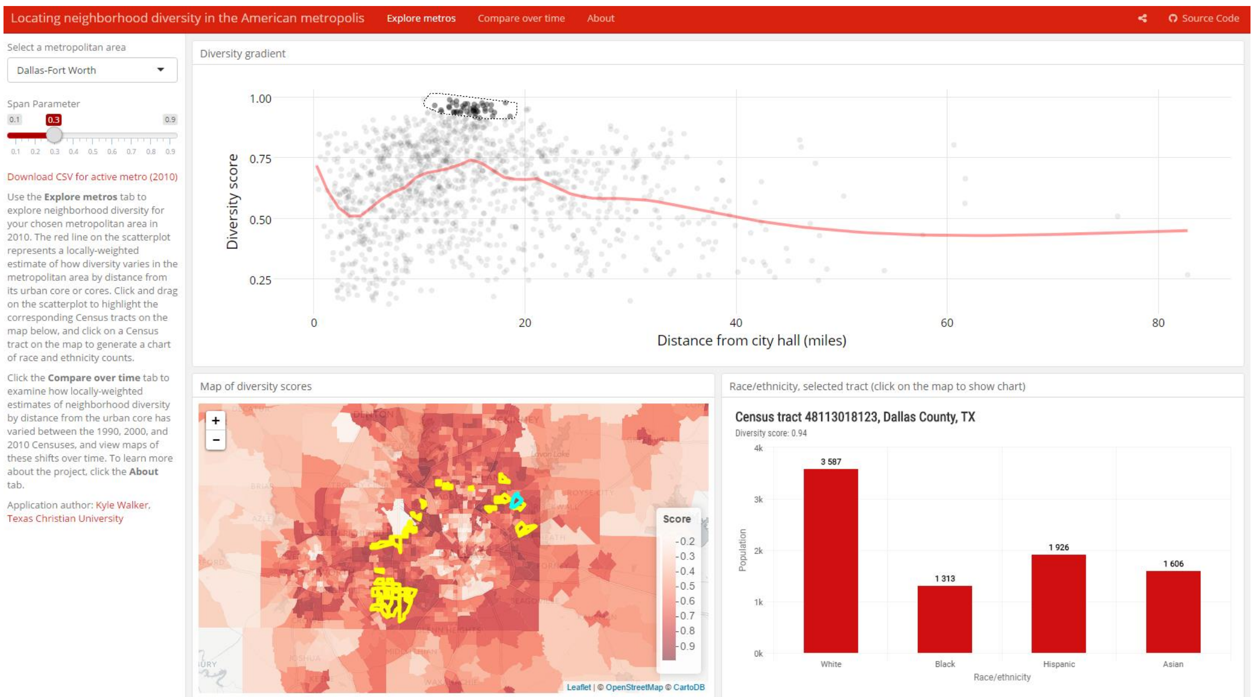

Figure 3: *Locating neighborhood diversity in the American metropolis* Shiny application, showing the Dallas-Fort Worth, Texas metropolitan area.

The application has several components. In the journal article, the main methodological tool is the *diversity gradient*, which uses LOESS smoothing to illustrate how a tract-level racial and ethnic diversity index varies by distance from the urban core in a given metropolitan area. In the Shiny application, the diversity gradient is made interactive. The top panel of the application view shown in Figure 3 illustrates the diversity gradient for the Dallas-Fort Worth, Texas metropolitan area. Application viewers can click and drag on the scatterplot to select individual Census tracts. Upon doing so, those Census tracts are shown on the map below. Viewers can then click on any Census tract on the map to generate a bar chart that breaks down the tract population by race and ethnicity. As Figure 3 illustrates, the highest-diversity

---

[2] Available at https://walkerke.shinyapps.io/neighborhood_diversity/

Census tracts in Dallas-Fort Worth in 2010 were located in inner-ring suburbs surrounding Dallas. A second tab allows users to explore historical trends within selected metropolitan areas.

Due to space constraints, the original article focuses on the Chicago and Dallas-Fort Worth metropolitan areas as case studies. The accompanying interactive application extends the study results to 44 other metropolitan areas, making the research results more broadly applicable. Additionally, the original study is published in a closed-access academic journal, which has a relatively limited circulation relative to the number of individuals who might be interested in the topic. Sharing study results in a freely accessible Shiny application thus allows a larger audience to engage and interact with the study's findings. In fact, media coverage of this project focused more on the Shiny application than the published article (Misra 2016).

4. Opening GIScience

Multiple factors shaped the coding and design decisions we made in creating our Shiny applications. In the sections below, we describe how we navigated tensions in the technical construction, design, and evaluation of our Shiny applications.

4.1 Creating tools

Both of our case studies involved complex coding challenges. For the housing application, this included the creation and display of interactive text panels as users clicked on specific points or the ability to switch basemaps or toggle map layers. For the urban diversity application, Shiny's reactive programming model allowed for connections between multiple HTML widgets, meaning that clicking on one tract on the Leaflet created map would pull up a Highcharts graph showing demographics of that tract in another window. The technical expertise required to create and maintain these tools was considerable. For the housing application, we have a small research team (one faculty and one undergraduate research assistant) and are working with multiple communities—seven at the time we write this article. While each of these communities have their own concerns, in order to keep workloads manageable, there is in inevitable tendency toward standardization: creating set surveys with limited customization options, allowing little attention to communities' specific concerns.

This method of construction is also fragile. The use of multiple, independently developed packages in our applications increases the risk of future incompatibility. In the case of the housing application, we benefitted from an API and package created by Ona, the company hosting our survey data. This made import into R relatively easy. Yet a change in that package or to the terms of service for Ona (who currently hosts for free) would complicate our current and future projects. A centrally designed application such as ESRI's Collector application or Fulcrum would remove some of this risk, but comes with a greater financial cost. Additionally, the rapidly changing nature of open-source software means that packages used in application development – and the dependencies of those packages – might change shortly after the deployment. This can limit reproducibility, as published application code may no longer work if package authors have introduced breaking changes. In turn, this requires testing the application code regularly against updated versions of the packages to ensure continued functionality.

Lastly, we also made decisions about what data to show and what should be left off the map. For the housing applications, these were often in consultation with communities. Lexington was interested in the ability to filter properties by the most prominent landlords in town, for example, or Spencer wanted the ability to identify an urban redevelopment zone. Several communities were interested in mapping crime patterns or in viewing census data. To help identify patterns in housing problems, we created a layer of tessellated hexagons to sum the number of properties with specific issues in different sections of the community. Decisions about how to contextualize neighborhood problems can be crucial. Hot spots of crime can be targets for increased policing or investment in social programs, two quite divergent outcomes. In the case of the housing application, we were wary of stigmatizing low-income households or providing more efficient ways for local government to penalize them for code violations. We continue to navigate this tension, in part through creating contextual materials and training that emphasizes community involvement at all phases of the research.

4.2 Opening data and code

The two case studies we use in this article pursue differing avenues for opening data to a broader public. As Cairo (2013) suggests, we attempted to create multidimensional tools that can work for multiple audiences by providing both quick summaries and the opportunity to explore data in greater detail. Through our Shiny interfaces, users with minimal training or access to GIS software to visualize these data and understand spatial and statistical patterns within them. In addition, we hope that the tools we created can be adapted and reused by other communities by hosting a sample version on GitHub or similar code sharing services.

Shiny lends itself to exploratory data analysis based on the variety of visualization tools provided by the application. This style of analysis can be difficult for users unaccustomed to it. Many communities expressed real excitement at seeing the application for the first time in their communities: as one official wrote in an email, “This is really cool.” Making use of the interface to draw meaningful conclusions takes more time. A team member in Birchwood, for example, said that it has been used primarily to guide data collection: “I haven't used it a lot. I've looked at it when I've gone through to find out what we finished and what we didn't finish. And *I've just looked at it*” (emphasis added). In line with the goals of EDA, these applications often create more questions than they answer, with some communities wondering whether additional datasets could be added.

While this shows that Shiny has value for generating questions, it remains unclear if communities will be able to move from exploration to actionable conclusions. We could assist in this process by providing more narrative structure around these conclusions. Using Rmarkdown code, for example, we could create reports that systematically summarize patterns for community members. While this would ease the difficulty of interpreting complex visualizations—potentially allowing communities to make stronger conclusions and claims—it would also potentially lessen community input on how to interpret and contextualize data. Given our own limited resources, we also lack the ability to work intensively with communities on this issue. Thus, opening these data—making them available to housing teams in ways that allow these teams to understand and interpret them—involves several significant tradeoffs related to community agency and the role of researcher expertise.

The ability to share our tools and code online is one other sensitive issue. Online hosting sites such as Github allow us to share data and code publicly. For the diversity application, this allowed all data used in the analysis to be hosted online, and the source code made available on GitHub allowing other developers and researchers to learn from and build upon the application. The application itself includes a "Source Code" icon and link, directing visitors to the R Markdown and Shiny source code for the application. This functionality is built-in to RStudio's shiny dashboard templates like flexdashboard, encouraging developers to share their application code with the community. In the housing project, however, things were more complicated. The community collected housing data is technically public—all responses are completed based on what is visible from the street. Given the potentially stigmatizing nature of the data, however, household level data are best kept from wide public view. The most straightforward solution to this is the creation of synthetic data for a different community, but as Singleton et al. (Singleton *et al.* 2016) note this can be time intensive, especially when the data include multi-media such as property photos. In this case, we matched survey data with non-corresponding address points and used this for an example application[3]. We also made our application code available on a linked Github repository[4].

The neighborhood diversity application serves as an attempt to “open up” aspects of a peer-reviewed research project to the public, even though the entirety of the project (e.g. the published article) is not open. In Singleton et al’s (2016, p. 1518) framework for open GIScience, their fifth tenet reads, “Where full reproducibility is not possible… researchers should aim to adopt aspects of an open GISc framework attainable within their particular circumstances.” Many top-ranked journals which scholars target for publishing their research, and which may be privileged by promotion and tenure committees, are not open access by default, and open access publishing fees for these journals may be prohibitive for scholars without external grant funding. As such, a companion interactive Shiny app can represent this “partial openness” proposed by Singleton et al.

4.3 Evaluating impact

As Sieber (2006) noted more than a decade ago, “few concrete measurement strategies” are in place for publicly engaged geospatial research, an assertion that remains accurate. Unlike more conventional research, which has both clear goals (research questions/hypotheses) and measurable outcomes (statistically significant results/citation counts), public research is more multidimensional. Goals can include not only answering research questions but also building networks and capacity in communities, obtaining external funding, creating reusable code/software tools, or facilitating the empowerment of underrepresented groups. Our experiences emphasized the need to be creative and attentive in assessing research impact.

[3] Available at https://comapuga.shinyapps.io/sampledata_flexdash/
[4] Available at https://github.com/jshannon75/HousingSurvey_ShinyFlexdash

For the housing application, research results have often been surprising and unpredictable. The town of Spencer made use of both community members and students at a local college for their housing assessment. According to a member of the housing team, simply collecting data transformed many residents' perceptions of low-income sections of their community. The large number of properties with tires in third yards stood out to one college volunteer. Such tires can be a public health threat, in part as a breeding ground for mosquitoes carrying the Zika virus (CDC 2016). Through documenting the prevalence of this issue, this student, in partnership with local government, secured funding from Georgia's Department of Public Health to create a free tire cleanup initiative. While not an intended focus of this project, collecting data produced a significant impact for the community before the first map was even drawn.

In other communities, however, the limits of mapping applications have been clearly apparent. In Birchwood, repeated efforts to do outreach with African-American homeowners for programs to improve housing issues have yielded little interest. One white member of the housing team perceived the causes for this lack of response in this way: "I think people don't trust it. They've been played all their life. People say they're going to do this for them or that for them, and then they don't. So I don't think they really trust it." Birchwood, like many Georgia communities, has a long and difficult racial history. While our Shiny application may help identify some of the effects of this history, engaging historically marginalized communities in addressing issues likely requires a longer engagement that has little to do with mapping.

For the diversity application, the use of open source tools and Shiny have allowed other developers to learn from and build upon its source code. The application is a featured example on RStudio's flexdashboard website (RStudio 2017b) and is a recommended example for the British Columbia Government's Data Visualization Tech Challenge (British Columbia Innovation Council 2017). Additionally, the application is active for 100+ hours every month, its source code website on GitHub commonly receives over 10 hits per day, and the application source code has been forked 30 times by users who want to adapt the code for their own applications.

5. Implications

Our experiences show how a process focused analysis of open GIScience projects reveals the tensions and tradeoffs researchers must navigate. Certainly, it is helpful to have set criteria (e.g., freely available data, shared code, open publication) to define open GIScience. In practice, however, we have found that the process is often more logistically and conceptually complicated than these set criteria suggest. In our concluding section, rather than focusing on revising these guidelines, we instead describe four key tensions we experience as practitioners of open GIScience. These tensions are not dialectical. That is, we do not argue that they require resolution. Rather, they are inherent to work in GIScience and context dependent, meaning that no two projects may balance them in quite the same way. Figure 4 outlines these tensions, which on the left are generally characterized by stable, defined, and specific approaches and on the right by new, emergent, and flexible ones.

	Settled		Unsettled
Tools:	Standardized	⟷	Flexible
Design:	Expert driven	⟷	Community driven
Audience:	Single	⟷	Multiple
Metrics:	Established	⟷	Emerging

Figure 4: Tensions we navigated in opening GIScience

5.1 Tools: Standardized vs. Flexible

Use of open source tools such as R and Shiny is common practice within open GIScience, and the rapid growth and adoption of these customizable tools means they can better serve the research needs of any given group. A code-based tool such as Shiny provides significant flexibility on decisions including color schemes, choice of basemaps, and graphing methods. This flexibility is based upon R's diffuse and growing system of developer created packages. The housing tool, for example, loads 14 packages explicitly (several as part of the tidyverse) and another 59 packages as dependencies. This flexibility comes with a cost, however, as updates to any of these packages can cause problems within the application. While privately developed software such as Tableau or PowerBI can also exhibit problems with backward compatibility, the largely non-centralized structure of development for R and similar open source tools makes this risk especially acute.

As one solution, RStudio has promoted increased standardization through packages such as packrat, which facilitates use of archived package versions, and tidyverse, a suite of tools for analysis and visualization developed to insure current and future interoperability (RStudio 2017c, Wickham 2017). Docker (https://www.docker.com) provides another solution, providing a way to reproduce the software environment used for specific analyses or product development. Nevertheless, navigating the tension between customizability and standardization with regard to application design and software tools remains a key aspect of opening GIScience.

5.2 Design: Expert vs. community driven

Our applications both used data visualization to engage with an array of end users. Paradoxically, though, creating online dashboards that made data more easily accessible to non-experts required difficult and highly technical work on our part. The ability to do "deep technical hacking" in Haklay's terms (2013) allows for greater customization. At the same time, as Brandusescu et al. (2015) note, this process inherently empowers technical experts rather than those outside the academy. Combining our own technical expertise with still soliciting and responding to user feedback was thus a key tension in our projects.

For the diversity application, the design process was mainly expert driven. While Walker (who developed that application) would have benefitted from user input, the audience for this application was rather broad and had little invested in its use. This input came sometimes indirectly from observing how others used and talked about this application online. Yet despite this "closed" method of development, Walker's online application received significant press and public attention, showing that there can be multiple paths to opening data to the public. For example, Walker was originally contacted by *The Atlantic*'s CityLab to discuss the findings of the published article (Misra 2016). The interview quickly turned from the article's content to the interactive Shiny application, and the published summary primarily covers the online interface. In contrast, Shannon developed the housing survey primarily for community stakeholders, not a general public. Like other forms of participatory research, application development was built around an iterative process of conversation including several rounds of input from stakeholders. Our varied contexts and goals meant that balancing our expertise with end users' feedback and goals played out differently for each of us.

5.3 Audience: Single vs. multiple

One primarily benefit of using Shiny for applications was its customizability, allowing us to create dashboards tailored to the needs of a specific audience. In our experience, though, we quickly became aware of the multiple audiences we were addressing through these applications. This included our end users with a range of skills and backgrounds, but it also included other communities and programmers outside our specific projects interested in adapting our code for their purposes. Part of our research was an active consideration of how to address these multiple audiences. For the housing application, Shannon has created standardized code and survey questions that can be used in multiple communities, as well as synthetic data for a public version of the Shiny application. In a different situation, working with just one or two communities over a longer period of time, applications may have been more tailored to specific communities' interests.

For the diversity application, Walker sought to ensure fidelity to the scientific content of the scholarly article while also making this content comprehensible to a broader, yet technically literate, audience. This involved the inclusion of the diversity gradient visualization, a concept visually comprehensible to a lay audience, but the omission of the other methodological element of the scholarly article, exploratory spatial data analysis of neighborhood diversity with local indicators of spatial association (LISA) (Anselin 1995). Instead of including the LISA maps on the application, Walker opted for simple choropleth maps of the entropy indices in which users could click on Census tracts to retrieve the racial and ethnic breakdowns of their populations. The LISA maps in the article are arguably more effective at illustrating the spatial patterns of diversity within metropolitan areas; however, choropleth maps also are comprehensible without the need for additional methodological or technical explanation.

This question of the intended audiences of application design is significant. The original purpose of the neighborhood diversity application was to share scientific research results with technical and academic communities through social media and the *Urban Studies* journal blog. Broader media coverage of the application came later and was not directly sought by the author. Because of this, scientific concepts such entropy indices and LOESS smoothing remain in the application, but these might have been presented differently if the original intent of the application was to advertise to a lay audience. As noted above, finding better ways to provide multi-dimensional, layered interfaces that combine broad summaries of data along with the ability to explore results in detail remains a challenge in projects such as ours.

5.4 Metrics: Established vs. Emerging

Academically, many open GIScience projects can be assessed using similar metrics as other kinds of academic work: citation counts and publication in high impact journals. As our experiences demonstrate, however, these metrics may fail to capture the full impact of the research. In some cases, this is a matter of how impacts are defined: not just access to data but long-term impact on community involvement and policy formation. In others, evaluating impacts involves creating new metrics within emerging tools. While Github provides some metrics, such as forks, that can be used to evaluate reuse of our code, long-term evaluation of this impact is still time intensive and difficult to measure. Finding ways to creatively measure the impact of our research will remain an ongoing challenge given the rapidly shifting technological landscape.

This tension between established and emerging metrics has implications for scholars engaged in open GIScience, however. Tenure and promotion committees still view peer-reviewed scholarship as the principal metric by which academics are judged. Further, the difficulty in measuring "impact" of public-facing applications can limit their influence on tenure and promotion decisions. Certainly, sharing of data and/or code through public-facing applications could draw more public attention to a scholar's research and improve its citation rate (Piwowar *et al.* 2007). However, a graduate student or junior scholar may have to assess trade-offs between the extra effort involved in open research and using that time to author another article.

6. Conclusion

Researchers using open GIScience often argue that their approach creates new opportunities for accountability, innovation, and public engagement. We enthusiastically support these goals. The tools and methods common in open GIScience have primarily been used to promote reproducible and transparent research among experts, but our work demonstrates the capacity of code based, open source tools such as Shiny to create interfaces for non-expert public users to explore and analyze research data.

Our reflection in this article has several limitations. While using a common software platform (R/Shiny) allowed us to compare the experience of building two differently situated platforms, future research could extend this work to comparison across software environments. This paper reflects on our experiences as developers in just modest data projects, and the dynamics in larger initiatives such as citywide open data initiatives may differ. Lastly, both case studies used in this article facilitate data analysis by a public audience. Future work could consider tensions related to traditional models of academic research, such as alternatives to traditional peer review or mechanisms to promote collaboration and study replication.

Still, we navigated multiple audiences for both the content we analyzed and the code we wrote, combined our own technical expertise with community engagement, and assessed the impact of our work based on both established and emerging metrics. Based on these experiences, we argue that the development of open GIScience should focus on *process* as well as by set *products*. As Elwood and Wilson (2017) argue, this focus on the praxis of open GIScience requires practitioners to combine both expertise in emergent software and techniques with sensitivity to the social and political dynamics at play within multiple communities of interest. By examining the key tensions that shape such engagements, researchers may be better equipped to present data and analysis to a broader public.

Acknowledgements

The authors thanks the editor and three anonymous reviewers whose comments spurred helpful improvements in this article. We are also grateful for Dr. Kim Skobba and Dr. Karen Tinsley, who provided assistance in developing the housing application as co-investigators. The Georgia Initiative for Community Housing also provided logistical support for community outreach.

References

Anselin, L., 1995. Local Indicators of Spatial Association-LISA. *Geographical Analysis*, 27 (2), 93-115.

Ash, J., Kitchin, R., and Leszczynski, A., 2016. Digital turn, digital geographies? *Progress in Human Geography*, 1-19.

Baker, M., & Penny, D. (2016). Is there a reproducibility crisis? *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452A

Brandusescu, A., Sieber, R.E., and Jochems, S., 2015. Confronting the hype: The use of crisis mapping for community development. *Convergence: The International Journal of Research into New Media Technologies*, 22 (6), 616–632.

British Columbia Innovation Council, 2017. $75,000 available to solve government's needs for data visualization [online]. *BCIC.* Available from: http://bcic.ca/news/blog/75000-available-solve-provincial-governments-data-visualization-challenges [Accessed 21 July 2017].

Brovelli, M.A., Minghini, M., and Zamboni, G., 2015a. Public participation in GIS via mobile applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 306-315.

Brovelli, M.A., Minghini, M., and Zamboni, G., 2015b. Public Participation GIS: a FOSS architecture enabling field-data collection. *International Journal of Digital Earth*, 8 (5), 345–363.

Cairo, A. (2013). The Functional Art: An introduction to information graphics and visualization. Berkeley: New Riders.

CDC, 2016. Help Control Mosquitoes that Spread Dengue, Chikungunya, and Zika Viruses. *CDC.* Available from: https://www.cdc.gov/zika/pdfs/control_mosquitoes_chikv_denv_zika.pdf [Accessed 11 April 2017].

Dalton, C.M., 2015. For fun and profit: the limits and possibilities of Google-Maps-based geoweb applications. *Environment & Planning A*, 47 (5), 1029–1046.

Elwood, S. and Wilson, M., 2017. Critical GIS pedagogies beyond 'Week 10: Ethics'. *International Journal of Geographical Information Science*, 1–19.

Gurstein, M., 2011. Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16 (2).

Haklay, M., 2013. Neogeography and the delusion of democratisation. *Environment and Planning A*, 45 (1), 55–69.

Iqbal, S. A., Wallach, J. D., Khoury, M. J., Schully, S. D., & Ioannidis, J. P. A. (2016). Reproducible Research Practices and Transparency across the Biomedical Literature. *PLoS Biology*, *14*(1), 1–13. https://doi.org/10.1371/journal.pbio.1002333

Johnson, P.A. and Sieber, R.E., 2013. Situating the Adoption of VGI by Government. *In*: D. Sui, S. Elwood, and M.F. Goodchild, eds. *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. 65–81.

Kitchin, R., 2014. The real-time city? Big data and smart urbanism. *GeoJournal*, 79, 1–14.

Kitchin, R. and Dodge, M., 2014. *Code/space: software and everyday life*. Cambridge: MIT Press.
Leszczynski, A., 2014. On the Neo in Neogeography. *Annals of the Association of American Geographers*, 104 (1), 60–79.
Misra, T. (2016). Where are the most diverse neighborhoods?
Munafò, M.R., Nosek, B.A., Bishop, D.V.M., Button, K.S., Chambers, C.D., Percie Du Sert, N., Simonsohn, U., Wagenmakers, E.J., Ware, J.J., Ioannidis, J.P.A., 2017. A manifesto for reproducible science. *Nature Human Behaviour*.
Nosek, B.A., 2015. Promoting an open research culture. *Science*, 348, 1422-1425.
OSF, 2018, Feb. 12. Retrieved from https://osf.io.
Piwowar, H.A., Day, R.S., and Fridsma, D.B., 2007. Sharing detailed research data is associated with increased citation rate. *PLoS ONE*, 2 (3).
Rey, S.J., 2014. Open regional science. *Annals of Regional Science*, 52 (3), 825–837.
Roche, S., 2015. Geographic information science II: Less space, more places in smart cities. *Progress in Human Geography*.
RStudio, 2016. Shiny [online]. Available from: https://shiny.rstudio.com/ [Accessed 9 May 2017].
RStudio, 2017a. Shiny [online]. *GitHub Repository*. Available from: https://github.com/rstudio/shiny [Accessed 9 May 2017].
RStudio, 2017b. flexdashboard: Easy interactive dashboards for R [online]. Available from: http://rmarkdown.rstudio.com/flexdashboard [Accessed 14 March 2017].
RStudio, 2017c. Packrat: Reproducible package management for R [online]. Available from: https://rstudio.github.io/packrat [Accessed 19 Jun 2017].
Sieber, R.E., 2006. Public participation geographic information systems: A literature review and framework. *Annals of the Association of American Geographers*, 96 (3), 491–507.
Sieber, R.E., Robinson, P.J., Johnson, P.A., and Corbett, J.M., 2016. Doing Public Participation on the Geospatial Web. *Annals of the American Association of Geographers*, 4452, 1–17.
Singleton, A.D., Spielman, S., and Brunsdon, C., 2016. Establishing a framework for Open Geographic Information science. *International Journal of Geographical Information Science*, 30 (8), 1–15.
Tinsley, K., 2017. Georgia Initiative for Community Housing [online]. Available from: http://www.fcs.uga.edu/fhce/gich/3 [Accessed 9 May 2017].
Townsend, A.M., 2013. *Smart cities: big data, civic hackers, and the quest for a new utopia*. New York: Norton.
Tukey, J.W., 1977. *Exploratory data analysis*. Reading, Mass.: Addison-Wesley Pub. Co.
Van Noorden, R., 2013. The true cost of science publishing. *Nature*, 495, 426–429.
Walker, K. (2016). Locating neighbourhood diversity in the American metropolis. *Urban Studies*. https://doi.org/10.1177/0042098016643481
Wickham, H., 2017. tidyverse [online]. Available from: http://tidyverse.org/ [Accessed 19 Jun 2017].
Wickham, H. and Grolemund, G., 2016. *R for data science : import, tidy, transform, visualize, and model data*. Sebastopol, CA: O'Reilly.