Gander: the Preliminary Design and Evaluation of an AR+Tablet System for Geospatial Analysis

Sathaporn Hu*
Dalhousie University
Algoma University

Muhammad Raza†
Dalhousie University

Derek Reilly‡
Dalhousie University

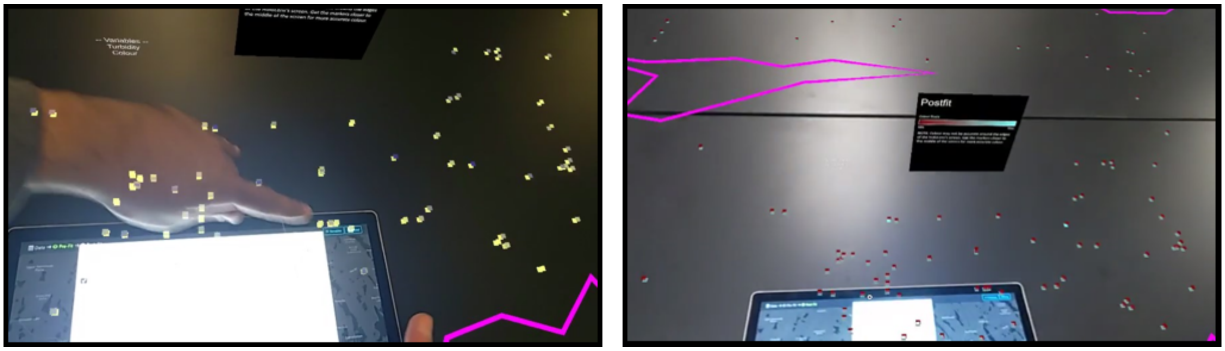


Figure 1: Screenshots from the study which also showcase Gander. **LEFT:** An investigator is guiding the participant through the study in the pre-fit stage. The purple line represents the map boundary. All gyphs are contained within it. **RIGHT:** Post-fit stage. The glyph colours have changed.

ABSTRACT

We explore the combination of headworn augmented reality (AR) displays and handheld tablet devices to support geospatial analysis. In this paper, we present the design of an AR+tablet prototype named Gander. Gander supports the selection of attributes when building a predictive model for some geospatial phenomenon, such as water pollution is correlated to the proximity of large cities, and the comparison of these models. We conducted a walkthrough evaluation with five experts in geospatial analysis and GIS tools. The experts identified strengths and weaknesses in Gander's design. We propose new design changes such as the pancake plot, rank-based visualization, and reducing steps on the tablet interface.

Index Terms: Augmented reality, tablet, immersive analytics, geospatial analysis.

1 INTRODUCTION

Whitlock et al. [17] argue that there is a need to make geospatial analysis more mobile–so that researchers can analyze data where they are collected. We designed and evaluated Gander, an AR+tablet application for geospatial data analysis. Gander (Fig. 1) is a high-fidelity prototype with enough functionalities to perform tests. We conducted walkthrough demonstration (WD) study involving five expert geospatial practitioners. In summary, the study was useful to provide a wide range of feedback on the design of Gander. Some design choices (e.g., provide overview information) have been suggested by the study results. However, we also identify additional design changes and future topic for condiserations.

2 RELATED WORK

Whitlock et al. [17] developed an AR app to be used with a mobile phone for in-situ data collection and immersive analytics. We designed Gander, an AR+tablet. Unlike Whitlock et al.'s app, we focused on understanding multiple linear regression. We previously conducted and presented two studies to evaluate certain aspects of Gander. The first study [6] involved the participants a restricted "kiosk-mode" version of Gander. The goal was to analyze the participants' trajectory data. The second study [7] was focused on how the participants understood the main glyph-based visualization used in Gander. The study was also extremely restricted: the task only involved the participant judging values of the glyphs and indicating them. This was similar to the one conducted by Jankun-Kelly et al [8]. In their study, each participant judged a 3D glyph per trial. The WD study described here complements [6, 7] by involving experts.

3 DESIGN OF GANDER

Gander is an AR+tablet prototype designed to be easily testable–i.e. research questions could be easily generated and aspects of the interface could be evaluated. The design of Gander has the following requirements: (R1) demonstrable, (R2) testable, and (R3) thin. R1 means we must be able to perform a walkthrough using the prototype. R2 represents that we could perform some scientific studies with the prototype. Lastly, R3 means we focus on the core features, then we use the results of [6, 7] and the WD study to inform future design updates.

3.1 Gander Workflow

The user of Gander begins by selecting a data in a CSV format on the tablet. After the selection, the participants advance to the pre-fit stage (i.e. data exploration). Then, through a tablet-based dialog box on the tablet named "The Variable Picker" (Fig. 2, the user selects the variables that they want to view. Glyphs are displayed in AR (Section 3.2). The user pans the tablet using swipe gestures to move the glyphs and the map background on the tablet. The user can also move their bodies or walk around to check the glyphs. Panning the tablet also moves the AR content.

---

*e-mail: hus@algomau.ca
†e-mail: m.raza@dal.ca
‡e-mail: reilly@cs.dal.ca

Based on multiple works in multiple regression analyses and model selection [5, 1, 2, 18, 4, 9, 13], we argue that a user's goals of a multiple linear regression (MLR) analysis are:

- **Parsimony.** Does a model have too many independent variables (IVs) for a single depedent variable (DV)? If multiple IVs are highly correlated, some should be removed. Otherwise, a fitted model will not be parsimonious and can have issues like multicollinearity which means one independent variable can linearly predict another [2].
- **Multiplicativity.** Considering if there is any multiplicavity or in other words interaction effect between the IVs, or if each IV's effect is independent. Excluding multiplicative effects can produce misleading models as certain effects can only be observed as multiplicative effects [1, 5].
- **Correlation.** The IVs must be able to explain the variances of observed DV values. Example effect sizes include R^2 and adjusted R^2 to quantify this [12]. In visualization, if there is a correlation between two or more variables, a change in one variable should also be observable in another. When the values of one variable are observed to be increasing, the values of another variable could be increasing or decreasing. For instance, if we try to water pollution and we believe that pollution is correlated to the proximity of large cities, we expect the levels of polluting chemicals to be higher nearer to large cities.
- **Spatial Autocorrelation.** Data may differ based on their locations, which may complicate fitting; for instance, the provincial average levels of lake chemical statistics may be different in different counties [13].

Through the three studies, we aim to observe and understand how a user may perform the aforementioned tasks.

The user can launch another touch-friendly dialogue box to adjust a MLR model called "The Equation Modeller" (Fig. 3). They can then fit a regression model using a drag-and-drop touch interface. After fitting, the user advances to the post-fit stage (i.e. model checking) to compare two models using the same glyph-based visualization as described in Section 3.2. Although we have not implemented the functionality, we envision the user could then compare different models.

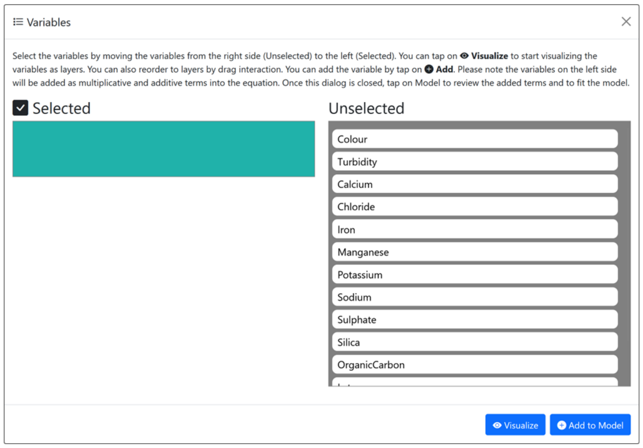
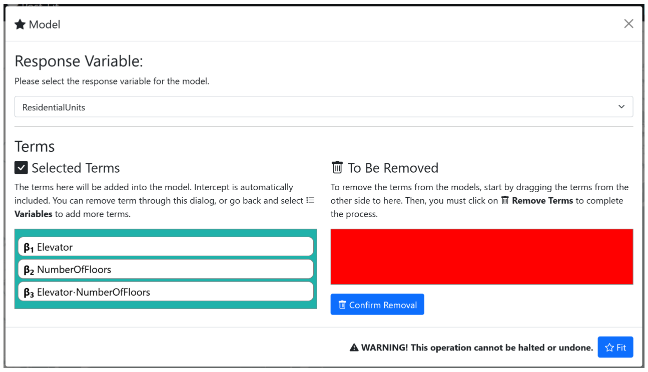

Select the variables by moving the variables from the right side (Unselected) to the left (Selected). You can tap on **Visualize** to start visualizing the variables as layers. You can also reorder to layers by drag interaction. You can add the variable by tap on **Add**. Please note the variables on the left side will be added as multiplicative and additive terms into the equation. Once this dialog is closed, tap on Model to review the added terms and to fit the model.

Figure 2: Variable Picker


Please select the response variable for the model.

The terms here will be added into the model. Intercept is automatically included. You can remove term through this dialog, or go back and select **Variables** to add more terms.

To remove the terms from the models, start by dragging the terms from the other side to here. Then, you must click on **Remove Terms** to complete the process.

WARNING! This operation cannot be halted or undone.

Figure 3: Equation Modeller

3.2 Glyph-based Visualizations

Gander supports a 3D glyph-based visualization. During the pre-fit stage (Fig. 1-LEFT), assuming v_i is the value of a variable and x_i is a datum belonging to x, then the value assignment follows: $v_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$. During the post-fit stage (Fig. 1-RIGHT) where the user is examining the likelihood, the likelihood of an individual datum relative to the models follows the equation based on [10]: $E_L = \frac{p_i}{\max\{p,q\}}$ where p_i is a likelihood of a model, and p,q are likelihoods of the two models being compared.

4 WALKTHROUGH DEMONSTRATION STUDY

We interviewed the participants to better understand their requirements, and evaluated the prototype through walkthrough demonstrations (WD). A WD involves the investigator guiding the user through the interface [11]. They developed software that allows VR designers to create ergonomic VR software. They invited experts and guided them through a tool instead of allowing them to explore the tool themselves since the tool has some learning curves. A WD is appropriate here because Gander has a linear workflow with a specific end goal.

4.1 Participants

The participants were: (1) two geospatial analysts working with a governmental agency, (2) two usability experts, and (3) two lecturers working at rural universities. The first group of participants worked with maritime geospatial data. We assigned the participants in this group the following ID: A1, A2. The second group was not geospatial analysts; however, they evaluated mixed-reality geospatial analytics software used by the first group. We assigned them: B1, B2. B2, though not a geosaptial analysts, had spatial analysis experience through working with eye-tracker data. The third group of participants possessed the most technical backgrounds. Not only they advised people on conducting research with geospatial data, they also taught advanced geospatial analysis classes. They received these IDs: C1 and C2.

4.2 Apparatus

For the AR headset, we used Microsoft HoloLens v2. due to its mobility. While most participants (A1, A2, B1, B2) performed the task in a blacked out laboratory, we also anticipated our study to be conducted elsewhere. For instance, C2's session was in his own office. As such, an AR device with a high mobility was necessary. The tablet was Microsoft Surface Book 3 with a 15-in. We also used a voice recorder to record the participants' interview answers and Microsoft Azure to transcribe them.

4.3 Protocol

In the first interview, we performed a semi-structured interview where we asked the participants about their backgrounds: (1) their technical background, (2) their process for geospatial analysis, (3) their challenges, (4) how to communicate information, and (5) familiarity with mixed reality technologies. Before the interview and the WD activity, the participants except C1 watched the video of the training. C1 did not perform any in-person activity due to his prior commitments.

The data used in the WD were modified from Nova Scotia Lake Chemistry data [16]. The investigator launched Gander and selected the data. The participant then interacted with the system. As the participant interacted with the prototype, the investigator guided the participant into creating a single MLR model and fitting it. The participant chose the concentration of a chemical to be the DV, and other variables (e.g., other chemicals, lake turbidity) to be the IVs. Once the fitting process was complete, the participant analyzed the likelihood in the post-fit process. In the exit interview, we asked the participants for their feedback. Generally, the questions pertained to: (1) the benefit of the interface, (2) the comprehensibility of the interface, and (3) where could Gander be improved.

Given that each participant had different availability, tailoring the procedure to suit their schedule was necessary. A1, A2, B1, and B2 each attended a 30-minute session that included a background interview plus a WD and an exit feedback interview. The 30-minute session was hybrid—i.e. the participant interacted with the prototype in person; however, they were interviewed by a remote interviewer through a computer. A1 agreed to an additional round of remote interviews. In this extra session, we asked A1 questions on the communication of his results, the stakeholders involved in the process, and the post-fit stage.

Before his interview, C1 requested for the training to be made available to him. Then, he attended a 45-minute session of a background interview plus the WD with an exit interview. C1 was unable to trial Gander in person. Therefore, the WD involved us and C1 reviewing the video, and critiquing the interface. Although C1 had more time than A1, A2, B1, and B2, the questions did not change based on the extra time. Instead, C1 received more time to answer the questions. We performed a 60-minute with C2 in person at his office. His procedure was the same with A1, A2, B1, and B2. Like C1, the additional time did not result in a significant change of the procedure. He simply had more time to respond and to interact with Gander.

4.4 Analysis

To analyze the interview and the demonstration walkthrough, we first generated thick descriptions and analyzed their details afterward (available as supplementary materials). We use the qualitative analysis method outlined by Reilly & MacKay [15]. In their work, they interviewed biologists to understand how they annotated ecological data collected from fieldwork. Before interviewing the participants, they performed a first principle analysis–that is, trying to understand what constitutes the "standard practice" in the existing literature.

4.5 Results

We present our results as two parts: the Workflow, and the WD. The former is based on the interview data on the participants' backgrounds. It involves identifying the steps and procedures deployed by the participants in their work, this is similar to the reporting in Reilly & MacKay [15]. The latter refers to the direct feedback pertaining to the interface of Gander.

4.5.1 The Workflow.

Each step in Gander was supposed to represent how a user operates based on prior literature in regression analyses. However, the interviews reveal that Gander may be missing certain steps. First, some experts identify the end goal before beginning any work. C1 indicated that geospatial analysts may modify their procedures based on the end goal. He stated: "If I'm teaching the lower-level GIS courses or doing some casual research without any publication goal or anything, I would [perform] simple statistics." Secondly, the user may involve other stakeholders in the decision-making process. For instance, A1 stated that he deferred to his colleagues for interpreting his model's outputs. He stated: "[It is] someone else's job to figure out how to [interpret the risk]. All I had to do was figure out how to assign risk to locations in space." This highlights the need for collaboration between the direct and indirect users.

Geospatial software packages can vary in terms of functionalities. Some software, while more limited in terms of features, is more automated. This can affect the workflow, as C2 stated: "From my perspective, [online GIS] has very limited functionalities, but in some ways, they are very smart; you just fill it with data and the app starts think about the type of data you have and then provide you with some options or solutions. In contrast, [to use] the desktop version, you still need to know more about the different types of maps, and more about how to create [them]." C2 further argued that online GIS software could democratize geospatial analysis to those without prior technical experience.

We did not hear much from the experts about the post-fit stage after the WDs. From listening to the participants, we found the post-fit tasks are usually performed only with aggregated overview information (i.e. test statistics). Furthermore, some experts would not benefit from the post-fit stage at all; for instance, A2 and B2 only worked with descriptive statistics, and would never fit a mathematical model.

4.5.2 Interface

After using the system, we interviewed the experts. Although the sample was small and saturation was not reached, summarizing the participants' thoughts and feedback could help us to better reflect on the interview results of the mixed-method study. The summary of the interviews are as follows.

- **Overall Feedback.** In general, the participants deemed the interface as good due to its general ease of use, and streamlined appearance–as A2 stated: "[Gander] was clear, and uncluttered." B1 said commented that Gander was "pretty intuitive and straightforward." However, both participants also had some criticisms: A2 thought Gander was too general for his tasks, and B1 commented that the AR interface could be replaced with a desktop-, or a laptop-based one.

- **Improvement for Tablet.** A1, A2, B1 wanted an improved tablet interface. Namely, they wanted to reduce the numbers of dialogue box opening. B1 summarized the experience as: "needing to go back [to the tablet] and reset."

- **Improvement for AR.** C1 thought the AR map's background should not have been transparent. He explained that he must "try to ignore what is in the background and just focus on the data." He added that AR content should be mounted on other surfaces (e.g., walls) rather than on the tablet. A1 and C2 indicated that the ability to zoom in and out is also important. C2 stated: "I did not get to see the whole map at one time." C2 pointed out that walking away from the map and glyphs could simulate zooming, because they would appear smaller–i.e. change in angular size. However, the ability to change the zoom level via the interface would be better. Lastly, A1 wished that there could be an alternative to the glyph-based visualization, as he was working with shape-based data instead of point-like ones.

- **Overview Information.** C1 and C2 stated that aggregated information and statistics were vital for statistical inference. An example of aggregated statistics is a statistical table outlining a MLR model's test results. C2 stated: "[Gander should] just show some [overview] results. Or [it] may show some charts or figures."

- **Design Appropriation.** C2 suggested that Gander could be used for pedagogical purposes. He said that we should be "introducing your app or your tool or incorporating your tool in teaching ... either remotely or in class." By proposing a novel and unintended use of Gander, C2 introduced the concept of design appropriate–using a design to accomplish a task not originally intended [3].

5 LIMITATION

We acknowledge several limitations. First, our prototype followed the vertical slice principle. According to Ratner et al. [14], a vertical slice is a prototype that is functional but lacks many functionality. The goal is to create a prototype that the user can complete the main tasks from the beginning until the end. Other auxiliary tasks are excluded. In our WD study, the participants could only perform multiple linear regression. Furthermore, Microsoft HoloLens v2 has difficulty tracking the tablet. Therefore, we designed the study so that the tablet would be fixed once the AR headset detected its initial position.

6 FUTURE WORK

6.0.1 To Zoom or not to Zoom

Although zooming is a common functionality in most map-based software, we did not include it in the vertical slice because there are multiple designs that must first be considered. First, zooming could simply keep glyphs at the same size. This means that the distances between the glyphs increase; therefore, the user can zoom in to resolve glyph overdrawing. Secondly, zooming, in addition to increasing the space between the glyphs, can also increase the glyph sizes. While this type of zoom cannot resolve overdrawing, it makes each more visible to the user. In this case, finer details of glyphs (e.g., the overlapping areas of Radial glyphs) become more discernible. Lastly, zooming could be used for aggregation. For instance, when the user zooms out, multiple glyphs disappear and in their places, a larger glyph appears showing summary statistics of the disappeared glyphs. This type of zoom requires the designer to pick a good summary statistic. For example, an aggregate glyph that represents means may not be appropriate if the values are not normally distributed. Another example is a glyph representing the mode of the glyphs might be better if the user cares more about the most frequently appeared value. The three zooming behaviours could also affect inference as well. For instance, increasing spaces between the glyphs can lead to less visual comparisons between the glyphs. Consequently, the user starts to conflate values of the fewer glyphs as a detectable trend. This is a type of atomic fallacy.

6.0.2 Overview+Detail

We observed that in S1, many participants did not use the tablet to glean additional visual information. Instead, they must rely on AR visualizations and the tablet as a sole input device. This inspires us to implement an overview+detail version of Gander. Instead of the tablet providing additional details, it provides the overview visualization to the user. It also allows for the tablet-based input to be "amplified." Input amplification allows the user to annotate (e.g., creating a circle to highlight an area) and see the annotation appearing on the AR map. Without input implication, the user may need to walk around the visualization in order to annotate which may not be as convenient. Input implication also improves panning. Instead of swiping 1m on the tablet to pan the AR content for 1m, the user can simply scroll a few centimetres instead. This does not mean that we are eschewing the Focus+Context paradigm. There are still areas where additional resolution must be implemented. For instance, some details like terrains are still best being rendered on the tablet itself.

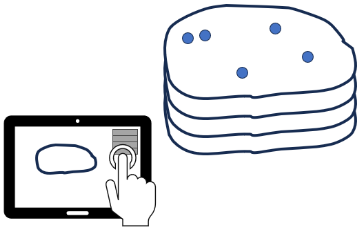


Figure 4: An overview+detail mode of Gander. The user can use touch to select an area and create the pancake plot. The dots are "chips" which represent the extreme values.

6.0.3 Selection Tool and Pancake Plot

We propose that in addition to glyphs, we can also provide an aggregated visualization. Using the tablet, when the user selects an area in AR, the user can create a stack of cartograms which represent aggregated information of multiple variables. Swiping a certain on the tablet itself also supports the user to seamlessly switch between the cartogram. We call the stacked cartograms, the pancake plot. Furthermore, the plot can have chips, or glyphs that represent outliers on them. Fig. 4 shows a prototype of how the pancake plot may operate.

6.0.4 Akaike Information Criterion-based Post-Fit and Rank-based Visualization

Based on feedback from A1 in the study, we believe that the post-fit stage should move away from using likelihood-based effect sizes. Instead, Akaike Information Criterion (AIC) should be adopted as it allows for multiple models to be compared at once. However, by using AIC, the post-fit glyphs must now convey different information. Furthermore, AIC-based inferences rely on rank-based inferences. For instance, when comparing multiple nested models, we select the one with the lowest rank (i.e. the smallest AIC value). This necessitates additional work in how ranks can be displayed as aggregates (e.g., using the pancake plot), and individual points (e.g., as chips).

6.0.5 Collaboration

Some participants suggested that Gander could be used for pedagogical purposes. This means we should explore expanding the interface to support multiple users with each user playing different role. Furthermore, we believe that generative AI and virtual agents might be incorporated to help with inferences.

7 CONCLUSION

We designed and prototyped Gander using thin prototyping. The prototype supports a small set of features at a high-fidelity which allowed us to conduct the WD study. The future work includes expanding the prototype by incorporating new techniques such as the pancake plot. We further plan to develop new research based on our previous studies [6, 7] and this one.

ACKNOWLEDGMENTS

The work was partially funded by Mitacs (Grant ID: IT16687) and Ericsson. We also would like to thank Miguel Garcia-Ruiz, Calyssa Skeggs, and the anonymous participants for their support.

REFERENCES

[1] B. F. Braumoeller. Hypothesis testing and multiplicative interaction terms. *International Organization*, 58(4):807–820, 2004. doi: 10.1017/S0020818304040251 2

[2] J. I. Daoud. Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*, 949, Dec 2017. doi: 10.1088/1742-6596/949/1/012009 2

[3] A. Dix. Designing for appropriation. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but Not as We Know It - Volume 2*, BCS-HCI '07, p. 27–30. BCS Learning & Development Ltd., Swindon, GBR, 2007. 4

[4] J. J. Dudley and P. O. Kristensson. A Review of User Interface Design for Interactive Machine Learning. *ACM Transactions on Interactive Intelligent Systems*, 8(2), June 2018. doi: 10.1145/3185517 2

[5] R. J. Friedrich. In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science*, 26(4):797–833, 1982. 2

[6] S. Hu and D. Reilly. Comparative Glyph-Field Trajectory Analyses with an AR+Tablet Hybrid User Interface for Geospatial Analysis Tasks. In J.-M. Normand, M. Sugimoto, and V. Sundstedt, eds., *ICAT-EGVE 2023 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*. The Eurographics Association, 2023. doi: 10.2312/egve.20231320 1, 4

[7] S. Hu and D. Reilly. Parallax-based Glyph Composition Technique with Colour-Blending Glyphs. In A. Campbell, C. Krogmeier, and G. Young, eds., *ICAT-EGVE 2023 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments - Posters and Demos*. The Eurographics Association, 2023. doi: 10.2312/egve.20231342 1, 4

[8] T. Jankun-Kelly, Y. Lanka, and J. Swan II. An Evaluation of Glyph Perception for Real Symmetric Traceless Tensor Properties. *Computer Graphics Forum*, 29(3):1133–1142, 2010. doi: 10.1111/j.1467-8659.2009.01711.x 1

[9] H. Jenkins-Smith, J. Ripberger, G. Copeland, M. Nowlin, T. Hughes, A. Fister, and W. Wehde. Quantitative Research Methods for Political Science, Public Policy and Public Administration: 4th Edition With Applications in R. chap. 13, pp. 127–133. University of Oklahoma, Norman, OK, 2021. 2

[10] J. E. Johnston, K. J. Berry, and J. Paul W. Mielke. Measures of Effect Size for Chi-Squared and Likelihood-Ratio Goodness-of-Fit Tests. *Perceptual and Motor Skills*, 103(2):412–414, 2006. doi: 10.2466/pms.103.2.412-414 2

[11] D. Ledo, S. Houben, J. Vermeulen, N. Marquardt, L. Oehlberg, and S. Greenberg. Evaluation strategies for hci toolkit research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–17. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173610 2

[12] M. S. Lewis-Beck and A. Skalaban. The R-Squared: Some Straight Talk. *Political Analysis*, 2:153–171, 1990. doi: 10.1093/pan/2.1.153 2

[13] V. Peña-Araya, E. Pietriga, and A. Bezerianos. A comparison of visualizations for identifying correlation over space and time. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):375–385, 2020. doi: 10.1109/TVCG.2019.2934807 2

[14] I. M. Ratner and J. Harvey. Vertical Slicing: Smaller is Better. *2011 Agile Conference*, pp. 240–245, 2011. doi: 10.1109/AGILE.2011.46 4

[15] D. Reilly and B. MacKay. Annotating Ecology: Looking to Biological Fieldwork for Mobile Spatial Annotation Workflows. *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services*, p. 35–44, 2013. doi: 10.1145/2493190.2493245 3

[16] The Government of Nova Scotia. Nova Scotia Lake Chemistry Data, 2021. Data Set. 3

[17] M. Whitlock, K. Wu, and D. A. Szafir. Designing for mobile and immersive visual analytics in the field. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):503–513, 2020. doi: 10.1109/TVCG.2019.2934282 1

[18] L. Zhang, Z. Ma, and L. Guo. An Evaluation of Spatial Autocorrelation and Heterogeneity in the Residuals of Six Regression Models. *Forest Science*, 55(6):533–548, 12 2009. doi: 10.1093/forestscience/55.6.533 2