

Discovering the Unexpected & Explicable: Scientific Reasoning and Research Design for Spatial Data Analysis

Koschinsky, Julia^a

^aUniversity of Chicago, Center for Spatial Data Science, jkoschinsky@uchicago.edu

Keywords: scientific reasoning, exploratory spatial data analysis

Methods for exploratory data analysis and exploratory spatial data analysis (or ESDA) are useful to identify outliers, clusters, skewed distributions and correlations (see Figure 1 for examples implemented in our GeoDa software). Researchers routinely use these methods to find insights.

However, what motivated this project is that, by default, it is easier to find insights that confirm the expected. Often, in fields like geography, statistics or computer science, available software and data drive the choice of research questions and the process of how we explore data. Typical insights gained this way are descriptive, like where a cluster is located or whether variables are correlated.

To be sure, expected insights are reassuring. And descriptive insights are important. But instead of stopping here we want to build on them to also find insights that are new and relevant — we want to discover the unexpected (Anselin, 1998; Kielman & May, 2009). And, while we do need to know where clusters and outliers are — as researchers, we also want to go further and explain why these patterns exist (Good, 1983). In this project we presume that there are ways for structuring the process of data exploration that make it more likely to discover unexpected and explanatory insights (Platt, 1964).

This presentation summarizes results from a summer 2020 lab where we started experimenting with how to do this using our Center’s GeoDa software. The summer lab was directed by Julia Koschinsky of the University of Chicago’s Center for Spatial Data Science. Marcos Falcone helped mentor five young University of Chicago and high school students for 7-10 weeks (majoring in statistics, computation, geography, political science, and economics).

Our approach was to draw on philosophy of science and scientific reasoning to understand how the discovery of unexpected and explicable insights can work. We then tried to translate this to research designs for ESDA. Finally, we implemented the designs in replicable prototype examples for teaching and learning spatial research at the undergraduate or high school level.

For instance, in terms of scientific reasoning, classic work on causal explanations (Mill, 1843) augments the typical current focus on correlations by also highlighting the need to assess the plausibility of your own explanation versus alternatives. This requires a mindset and practice of rigorously testing how our explanations might be wrong (Popper, 1959) rather than confirming that they're right (Nuzzo, 2015; Kahneman, 2011). To do this requires an iterative exchange between data and explanations — referred to as abductive reasoning, as it combines inductive and deductive approaches (Peirce, 1878; Heckman and Singer, 2017). We used Sherlock Holmes stories and the famous John Snow cholera case to illustrate the structure of these scientific reasoning concepts for a high school context (Coleman, 2019; cf Konnikova, 2013; Vinten-Johansen, 2020).

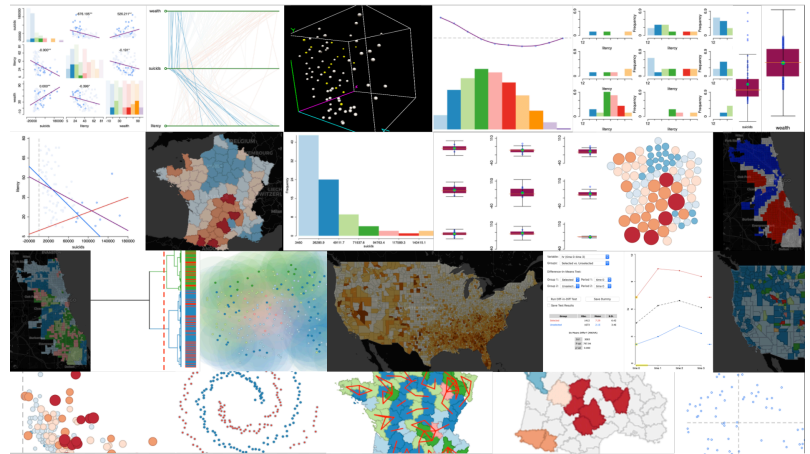


Figure 1. EDA and ESDA methods in GeoDa.



Figure 2. Four resources to use ESDA for discovering the unexpected and explicable

Scientific reasoning goes back hundreds of years. Our challenge this summer and from here on has been to translate this reasoning to research designs that are applicable to modern interactive ESDA tools. Each of us developed four prototype resources for teaching and learning ESDA in GeoDa that we will develop further (Fig. 2): 1) protocols for how this could be done; 2) case examples to apply and revise the protocol; 3) GeoDa demo scripts to make the examples replicable; and 4) cleaned data and documentation. These resources will be released as part of a GeoDa Cookbook in the near future.

Fig. 3 illustrates one of the protocols that differs from how ESDA is typically navigated. The starting point is the exploration of patterns in the outcome variable of interest. Next is the formulation of alternative explanations whose patterns plausibly match those of the outcome variable. Then we draw on quasi-experimental research designs to structure the testing of this match (Shadish et al., 2002). Finally, data about the hypothesized explanations are analyzed with ESDA and regressions to test or reformulate the hypotheses as part of an abductive process.

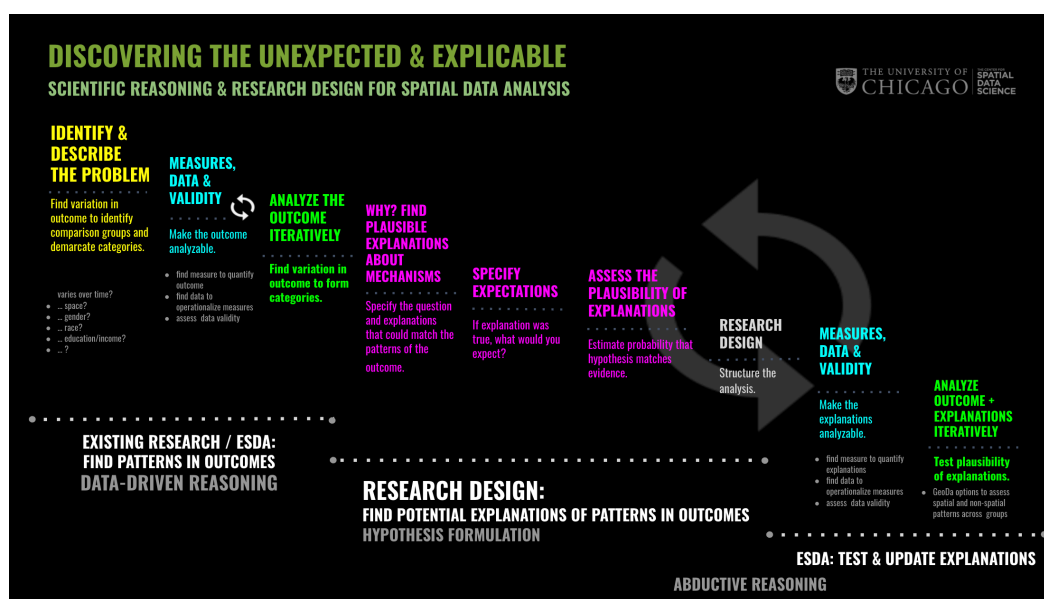


Figure 3. Example protocol for navigating ESDA

Acknowledgements

The contributions of Marcos Falcon, R.E. Stern, Atman Mehta, Mark Baker, Jizhou Wang, Felix Farb, Luc Anselin, Thomas Coleman, and Peter-Vinten Johansen are gratefully acknowledged.

References

- Anselin, L. (1998). Exploratory Spatial Data Analysis in a Geocomputational Environment. *Geocomputation: A Primer*. Chichester New York: Wiley.
- Coleman, T. (2019). Causality in the Time of Cholera: John Snow as a Prototype for Causal Inference. Available at SSRN 3262234.
- Good, I. J. (1983). The Philosophy of Exploratory Data Analysis. *Philosophy of Science*, 50(2), 283-295.
- Heckman, J. J., & Singer, B. (2017). Abducting Economics. *American Economic Review*, 107(5), 298-302.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan.
- Kielman, J., J. Thomas, and R. May. 2009. Foundations and Frontiers in Visual Analytics. *Information Visualization* 8:239-46.
- Konnikova, M. (2013). *Mastermind: How to Think like Sherlock Holmes*. Penguin.
- Mill, John Stuart (1843). *A System of Logic*. Vol. 1. John W. Parker
- Nuzzo, R. (2015). Fooling Ourselves. *Nature*, 526(7572), 182.
- Peirce, C. (1878). Deduction, Induction, and Hypothesis. *Popular Science Monthly*, 13, 470-482.
- Platt, J. R. (1964). Strong Inference. *Science*, 146(3642), 347-353.
- Popper, K. (1959). *The Logic of Scientific Discovery*. Taylor & Francis, New York.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Vinten-Johansen, P. (Ed.). (2020). *Investigating Cholera in Broad Street: A History in Documents*. Broadview Press.