

Five Reasons Every Data Science Team Needs a Geographer

By Lauren Bennett and Trisalyn Nelson

The authors, Lauren Bennett and Trisalyn Nelson, are both trained geographers. Bennett is the program manager for spatial analysis and data science at Esri. Nelson holds the Jack and Laura Dangermond Endowed Chair of Geography at the University of California, Santa Barbara. The authors combine their perspectives from industry and academia to consider the role of geography and GIS in data science. They hope that data scientists and geographers alike will gain perspectives about the unique contributions a geographic approach brings to the field of data science.

There is little doubt we are in the age of data science. While there was a time when saying we studied spatial statistics made us unpopular party guests, today being a spatial data scientist carries cachet. People are interested. There is an intuitive understanding that society is generating massive amounts of data that—in the right hands—has power.

Recently, we were panelists for a spatial data science webinar. An audience member asked us to comment if GIS was still relevant in the age of data science. Another asked if data science should be taught in geography programs or left to computer science and statistics programs. We found these questions both surprising and not surprising.

Despite the success of GIS, geographers wrestle with their identities. This makes sense. Geography can be difficult to define. In many ways, geography is everywhere—in our cars and smartphones and in the media we consume. If it is all around us, what makes geography—and geographers—special? How many times have we explained we are not geologists or been diminished as scientists because don't all the rivers already have names?

As both geographers and spatial data scientists, we have wrestled with our identities. Our skill sets sit at the intersection of a Venn diagram of the fields of geography, statistics, and computer science. When we are struggling with spatial analyses and our code keeps crashing, it is easy to wonder why we didn't learn more computer science. We can envy the deep skills of colleagues in computer science and statistics. Yet, as we have advanced in our careers, our geographic training has been a powerful asset.

Geographers are trained in interdisciplinary thinking. They are trained to identify questions and pose solutions. In addition to bringing a spatial perspective to data science, they also bring the ability to link methods and solutions to applications. They are the people who connect all the threads, using geography's

interdisciplinary approach to link problems to solutions and data science to decisions.

Interestingly, we have found the benefits of our geographic training have amplified over the course of our careers. As we have moved into roles leading teams, our geographer's secret sauce has been critical in helping us accelerate science and pose solutions to pressing environmental and societal problems. This article highlights five reasons why data science teams should include a geographer.

1 Data science is (mostly) spatial science.

The current era of big data is an era of big spatial data. Eighty percent of big data is spatial. It contains geographic coordinates that can be mapped, according to *Big spatial data for urban and environmental sustainability*, a 2020 paper by Bo Huang and Jionghua Wang.

The importance of spatial data is evidenced by the growing prevalence of maps in every form of communication, from media outlets to public health briefings to social media. The spatial dimension of data provides additional information on how phenomena vary across space and relate to location.

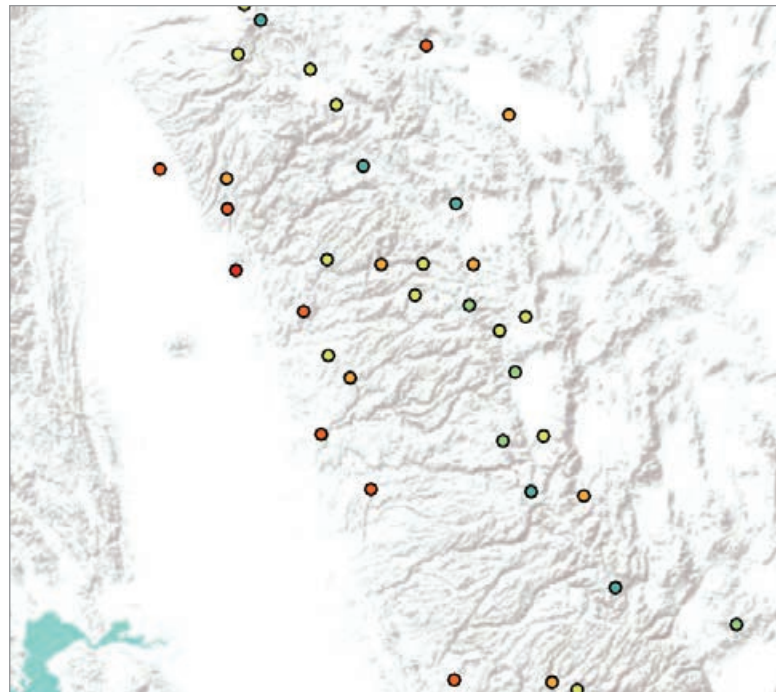
This means that if you're analyzing data, you could already be doing spatial data science. If you're not already taking advantage of the spatial characteristics of your data, you may be missing key information. But to take advantage of spatial information, you need to treat your data right.

If you are not a geographer, you should call one before you dig into the time-consuming, persnickety task of data formatting to make sure you are handling the spatial aspects of the data correctly. Geographers can make informed decisions about whether to use raster or vector representations and which resolution, extents, and spatial units to employ.

2 Spatial relationships can inexpensively provide information.

Mapping data can yield useful information that is hidden in spatial relationships. By leveraging Waldo Tobler's first law of geography, "All things are related and near things are more related than distant things," geographers can use spatial queries and analysis to uncover gems of information buried in the data.

Temperature Readings			
Field: Add Calculate Select			
	OBJECTID *	Shape *	temperature
1	2	Point Z	99
2	3	Point Z	96
3	4	Point Z	94
4	5	Point Z	95
5	6	Point Z	83
6	9	Point Z	92
7	10	Point Z	96
8	11	Point Z	82
9	14	Point Z	76
10	15	Point Z	82
11	16	Point Z	83
12	17	Point Z	89
13	18	Point Z	80
14	20	Point Z	87
15	22	Point Z	90
16	24	Point Z	95



← These sampled temperature values in a table provide limited information.

← Mapping the same temperature values based on where samples were acquired and symbolizing by temperature range provides more context, but GIS tools can provide more information.

As trained sleuths, geographers can uncover spatial relationships to better understand natural and human processes. Predictive models can pull in hundreds, if not thousands, of variables to try to make sense of complex phenomena but ignore the value that spatial relationships can bring to the table. Often those relationships incorporate the characteristics of places that are difficult to measure or quantify but can dramatically improve models. For example, if you are modeling data at the ZIP code level in your state, you know that ZIP codes that are closer together are more related than ZIP codes that are more distant from each other.

Geographers can model this simple truth and it will pay huge dividends. For a model that ignores geography and is given enough data, a relationship between two nearby ZIP codes may be uncovered based on their similarity. Including spatial relationships can cut down on the number of variables required and still get to the same conclusion more quickly and efficiently. Data can be expensive to purchase or time-consuming to organize. Often models can be improved when reality is more accurately represented. Why not include a geographer to make the most of your spatial information?

3 Geographers know how to communicate with maps.

There are few tools as effective as maps at communicating complex concepts in tangible, approachable ways. Maps can make even the most abstract academic concept evident and illustrate its importance. For example, it's much harder to ignore the results of analysis if they show your house will be directly affected.

However, it is easy to make cartographic mistakes, intentionally

or unintentionally. Books have been written about how to lie with maps. If you don't produce correct results from your analysis, it can mean losing money or resources or credibility. Sometimes, it can even be a matter of life or death.

You might want to call a geographer if you want to display your results on a map. Better yet, get help from a geographer who specializes in the art and science of mapmaking: a cartographer. While anyone can display data on a map, a cartographer is trained in map communication and will help you communicate the right message, using appropriate colors, shades, symbols, labels, and scales. ArcGIS Dashboards and ArcGIS StoryMaps can bring your maps and message to life.

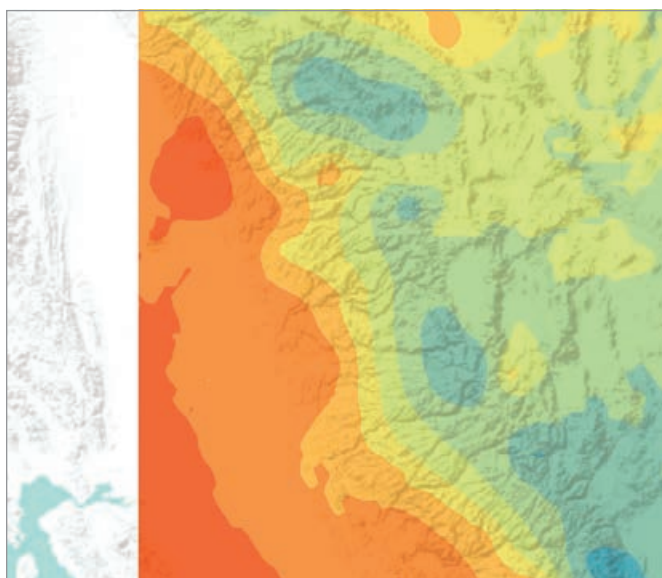
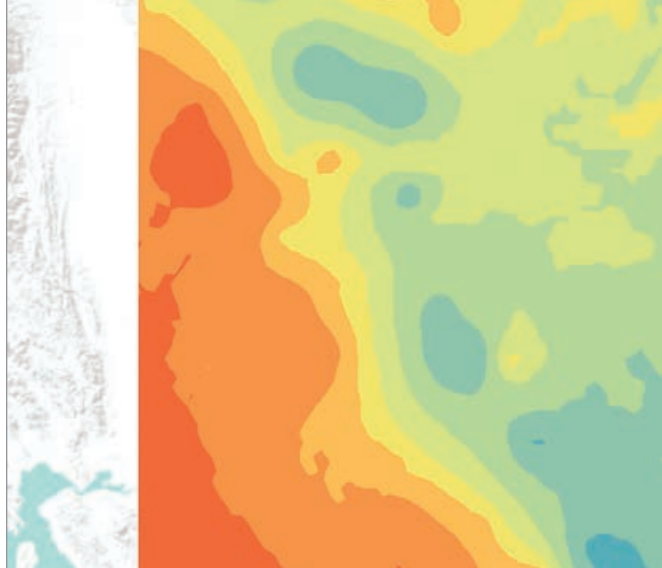
4 Geographers link science to the real world.

While data science seems to be resonating broadly with society, there is a growing subset of society that does not trust science. Effective scientific communication is critical as more decisions are made using the results of increasingly complex algorithms.

Communication needs to reach even those who are unfamiliar with data science. Because data analytics is used to manipulate perspectives and increase sales as well as create knowledge, the public can be suspicious when the same methods are used to do things that are ethically questionable or just plain wrong.

Geographers have a critical role in science communication because they are trained in linking science to real-world problems. They know how to work with data, but they also understand the nuances of the phenomena that are represented by the data.

Geographers can make a good analysis great by adding the



↑↑ Using geostatistical tools to interpolate a surface from the temperature values and relate them spatially provides more information.

↑ Combining the interpolated temperature surface with elevation data can yield useful information hidden in spatial relationships.

perspective of real-world problems and making it actionable. They can bridge the gap between analysis and action. As a result, geographers should be able to turn data into a truthful narrative and answer the critical questions required to build trust in their findings.

5 Geographers are interdisciplinary.

If your data science team is having trouble finding common ground or struggling to communicate with subject matter experts, you may have forgotten to add a geographer.

Very few geographers arrive at university knowing they want to be a geographer. Geography is often called a “found” major. Students from a variety of fields stumble across geography and find a fit because it overlaps a huge variety of fields, from the

humanities to engineering. People who like to connect problems and solutions often find a home in geography.

Some geographers are purely qualitative and others are purely quantitative, but a good geographer can work with teams that include various perspectives, approaches, and kinds of expertise. The interdisciplinary perspective of geography is a strength when it comes to building and leading teams because geographers are trained to integrate perspectives, ideas, and skills. Geographers learn that complex problems are rarely solved using approaches from a single discipline or field. Success is the product of carefully examining the problem and combining complementary and, at times, even seemingly contradictory approaches.

Call to Action

The problems that we are trying to tackle—with the best and brightest minds in data science—are formidable. Solutions addressing climate change, racial equity, and environmental justice are complex. It is important to remember that effective data science solutions need to resonate with decision-makers, politicians, and public opinion.

That doesn’t mean that everyone has to like your results. Rather, people need the best information to make the best decisions. Geographers can help get the best information out of spatial data and create pathways to communicate their findings through maps and by translating science for decision-makers and the public.

Although there have been times in our careers when we wished for deeper knowledge in one field, as we have progressed into leadership roles, we have been grateful for the breadth and approach of our training.

We are grateful that data scientists have deep and diverse skills. The geographic approach of spatial scientists includes the ability to integrate diverse disciplines, connect problems and action, and lead in developing solutions to address many of our world’s greatest challenges.

If you’re a geographer, lean into your unique skill set. If you’re leading a data science team or project and haven’t yet brought a geographer into the mix, what are you waiting for?

About the Authors

Dr. Lauren Bennett leads the spatial analysis and data science product engineering team at Esri. In this role, she oversees research and development of the ArcGIS geoprocessing framework, which includes spatial and spatiotemporal statistics, raster and multidimensional analysis, machine learning, and big data analytics. Bennett is passionate about the power of spatial data science to make an impact on the world in many areas including racial equity, social justice, public health, and climate change.

Dr. Trisalyn Nelson holds the Jack and Laura Dangermond Endowed Chair of Geography at University of California, Santa Barbara, and is the chair of the Department of Geography. Nelson and her team develop and apply spatial and spatiotemporal analyses to address applied questions in a wide range of fields ranging from ecology to health. Currently, her research focuses on active transportation and the use of big data and analytics to better plan cities.