



CARTO

The State of Spatial Data Science in Enterprise 2020

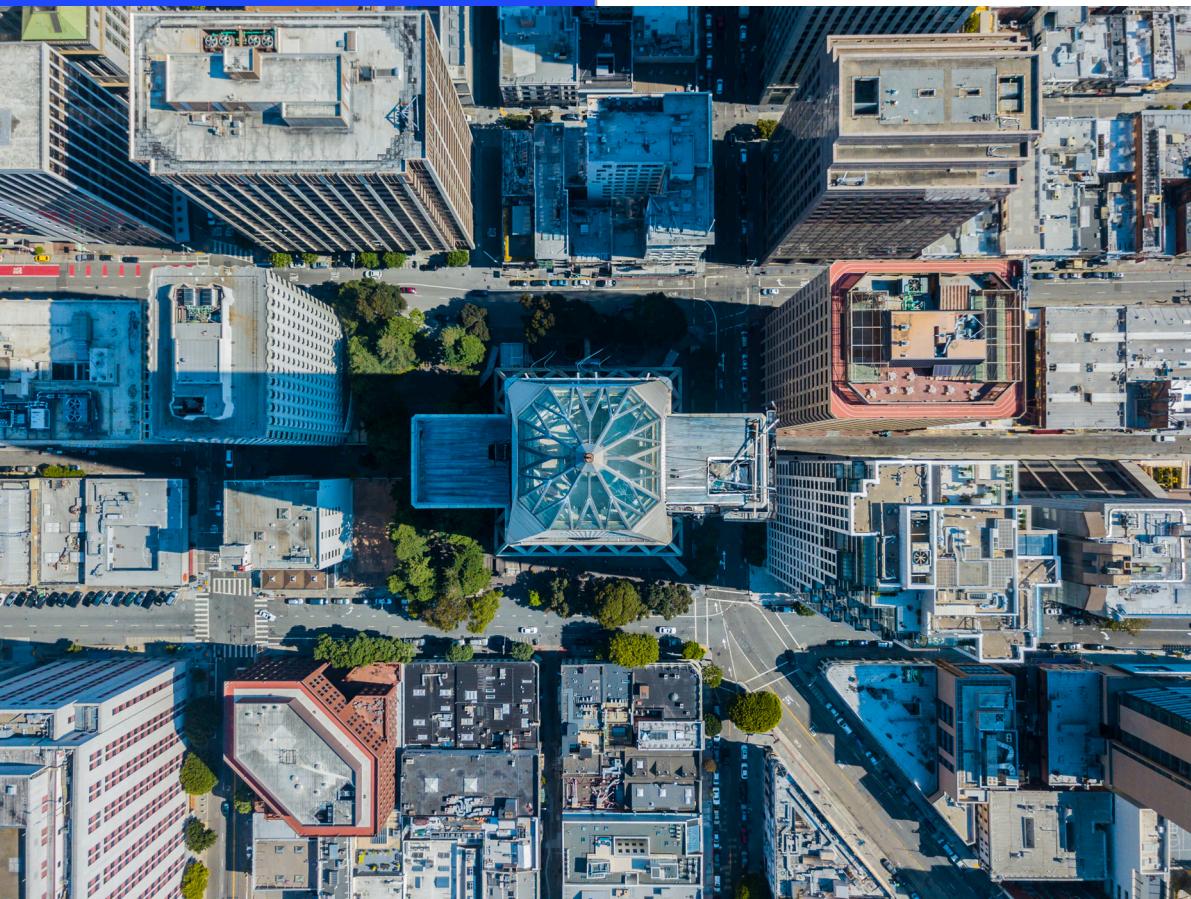
New insights on the status
of Spatial Data Science and
what we can expect in 2020

Introduction

INTRODUCTION

The Spatial Data Science Discipline

Spatial Data Science, or Geographic Data Science, is a **rapidly growing discipline** across a wide range of industries in both the private and public sector, as well as academia.



Nestled between Data Science and GIS, Spatial Data Science profiles are in high demand in a wide range of verticals including **Financial Services, Real Estate, Cities & Government, Management Consulting, Retail, Utilities, Telecommunications**, and many more.

In order to understand trends and shifts for Spatial Data Science across these industries we invited a group of thought leaders from enterprise organizations across the globe to participate in the **State of Spatial Data Science in Enterprise 2020 Survey** to gain new insights on the status quo and to explore what we can expect in 2020.

What is Spatial Data Science?

Before we discover the results, we should first explore how thought leaders in the space define it:

“Spatial data science treats location, distance, and spatial interaction as core aspects of the data”

"Spatial data science can be viewed as a subset of generic "data science" that focuses on the special characteristics of spatial data, i.e., the importance of "where." Data science is often referred to as the science of extracting meaningful information from data. In this context, it is useful to stress the difference between standard (i.e., non-spatial) data science applied to spatial data on the one hand and spatial data science on the other. The former treats spatial information, such as the latitude and longitude of data points as simply an additional variable, but otherwise does not adjust analytical methods or software tools. In contrast, "true" spatial data science treats location, distance, and spatial interaction as core aspects of the data and employs specialized methods and software to store, retrieve, explore, analyze, visualize and learn from such data. In this sense, spatial data science relates to data science as spatial statistics to statistics, spatial databases to databases, and geocomputation to computation."

Luc Anselin PhD, Center for Spatial Data Science Senior Fellow, University of Chicago



INTRODUCTION

How is Spatial Data Science different from GIS?



Before we can fully investigate the users and platforms emerging in our industry, we first need to understand what GIS is and what it means to different users.

GIS or Geographic Information Systems refer to varied types of information systems - they may be a website, an application, or a database that handles and stores different types of spatial data.

For most GIS professionals, these information systems are a means to a more sophisticated end - such as carrying out a location-based analysis, ranging from simple reports, to intersecting data, to more complex spatial models.

GIS applies to a wide range of users all with very different use cases, **yet is one of those strange anomalies that, despite its value spanning many industries, has remained a niche field.** From Retail and Real Estate, to Biodiversity and Banking - **having a GIS department and community is commonplace.**

INTRODUCTION

A clear theme throughout this report is that most GIS today is done by people who don't know that it is GIS.

After 10 years of working with some of the most “spatially aware” companies, we’ve observed 4 types of personas doing GIS today.

Many would argue that the purist GIS community is not actually growing too much, but the reality is that GIS is expressing itself beyond its own traditional community. And Spatial Data Science is growing rapidly as a result.

Reflecting on these different personas, it is clear that we are observing millions of users here - and all three of the groups are growing rapidly.



	GIS professionals 200.000	Data Scientists 1 million	Developers 20 million	Data Analysts 40 million
Do they know what GIS is?	They know GIS inside out and have been trained on it.	Few of them know what GIS is, they think of location as another dimension of their data, and consider potentially enriching their data with demographics.	Very few of them know what GIS is. They think it is simply data visualization or plotting points on Google Maps.	Some of them know what GIS is, but for most of them location is just one more dimension of their data.
What tools do they use?	They use GIS specific software, 40% ESRI, the rest open source.	They live in R and Python Notebooks. 1 to 5% of them are aware of the specifics of Spatial Analytics, (we call them Spatial Data Scientists).	Google Maps, deck.gl, CARTO, Mapbox.	They live in Excel and BI tools like Tableau, or in data blending solutions like Alteryx. They are keen to get their hands on more data.
Background	Geography, environmental science, statistics, civil engineering.	All sorts of fields, including statistics, physics and computer science. A subset may be defined as data engineers coming mostly from Computer Science backgrounds. When the Data Scientist comes from a science background they tend to have more spatial analytics knowledge.	Computer Science. Engineering. Math.	Many places, they enter as Business Analysts, Sales analytics, fraud analytics, insurance.
Where do they work?	GIS departments in large corporations, consulting firms, public sector, or academia / non profit.	Now touching every industry. Tech companies tend to have more resources focused on Machine Learning and AI.	Companies in every industry have them. Many work in SaaS.	Every sector has them, from hedge funds, to utilities, to retail and beyond. Lots of this skill set in consulting firms. As more organizations, including public sector, are becoming more data-driven they appear everywhere.
What type of work do they do?	Reports, analysis, sometimes they become developers for GIS web apps, and a few of them dabble in spatial models.	Apart from the tech world of ML and AI, most of them spend their time doing advanced analytics for their organizations to understand their customer base, operations, churn etc. The Spatial Data Scientist normally focuses on process optimization and predictions for market analysis.	They build websites, mobile apps, sometimes they may touch data pipelines, making them more like Data Engineers.	Reports, typically dashboards or Excel spreadsheets. More advanced analysts will perform simulations or scenarios.

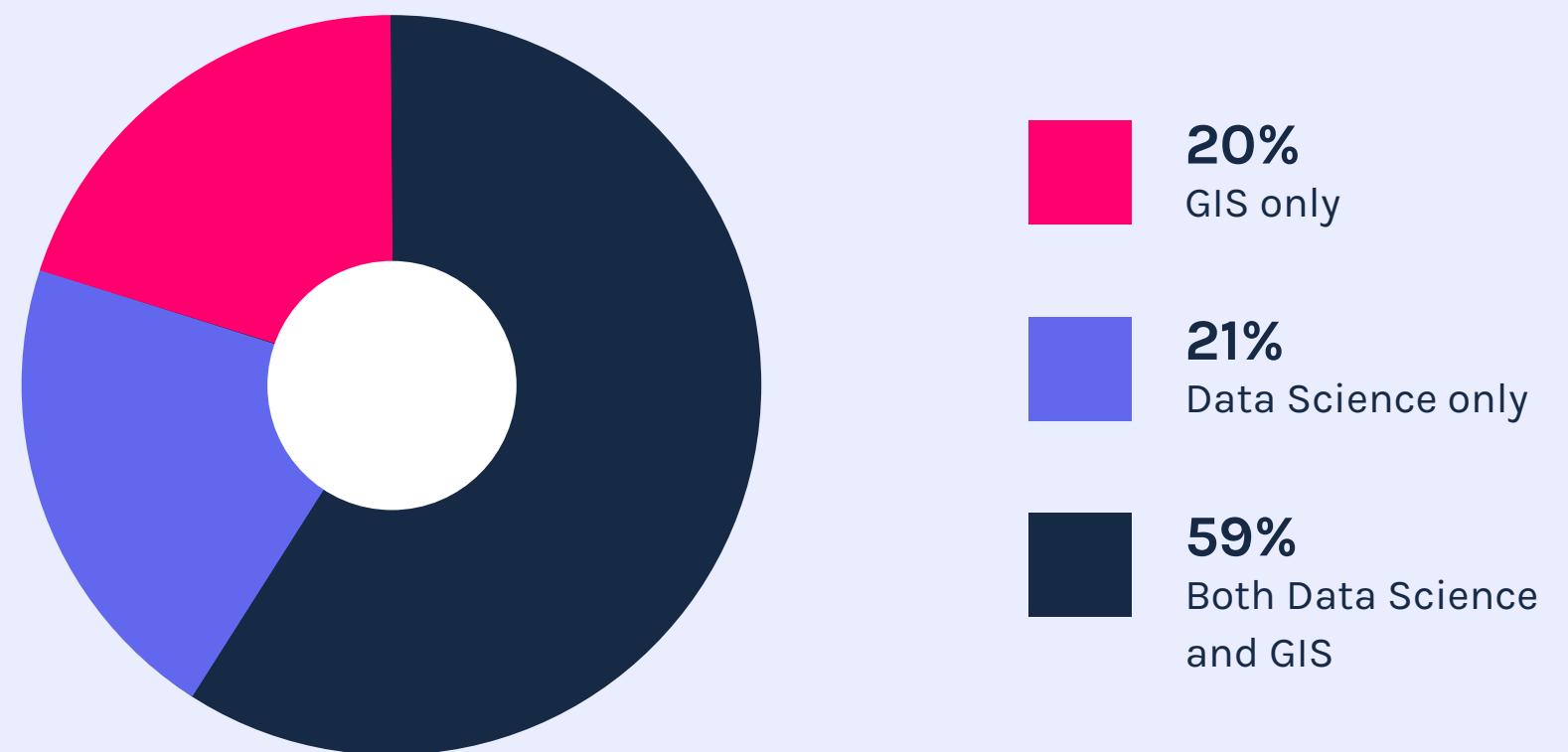
Results

RESULTS

Data Science and GIS teams at organizations

59% of participants stated that their organizations have both Data Science and GIS teams - whilst the percentage who had just a Data Science or GIS team only was equal (20% for each).

Do you have Data Science and / or GIS teams in your organization?



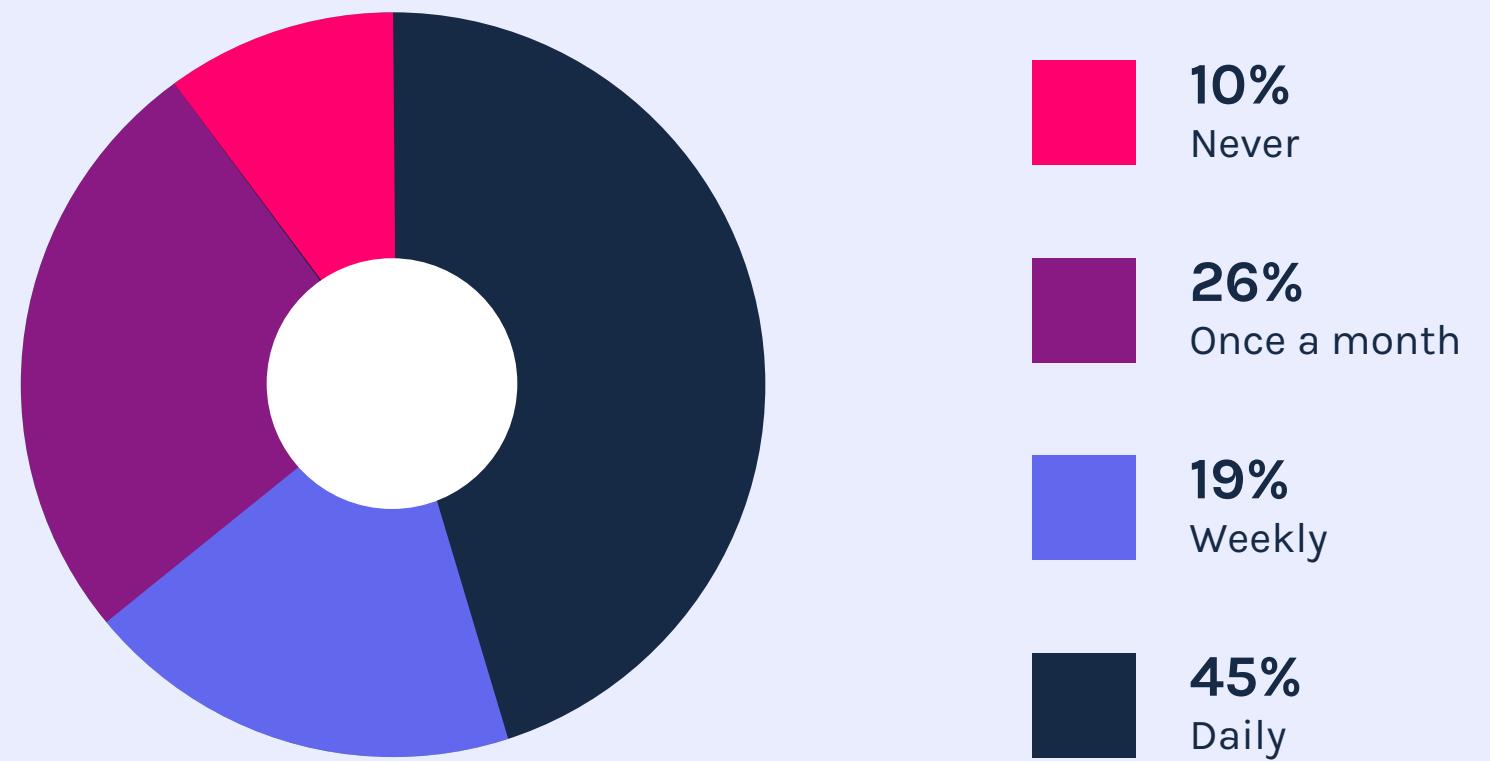
59% of organizations now have both Data Science & GIS teams

RESULTS

Frequency of collaboration between Data Science & GIS teams

For the organizations that have both Data Science and GIS teams, we asked them how often they collaborate:

How often do your Data Science and GIS teams collaborate?



45%

of Data Science & GIS teams collaborate on a daily basis



Interestingly, we saw that 45% of Data Science and GIS teams are collaborating on a daily basis

- which reinforces just how relevant the Spatial Data Science business function is becoming due to the spatial nature of the business problems and challenges these teams are encountering.



With 25% of participants sharing that their teams only collaborate once a month, and 10% stating that they never work together

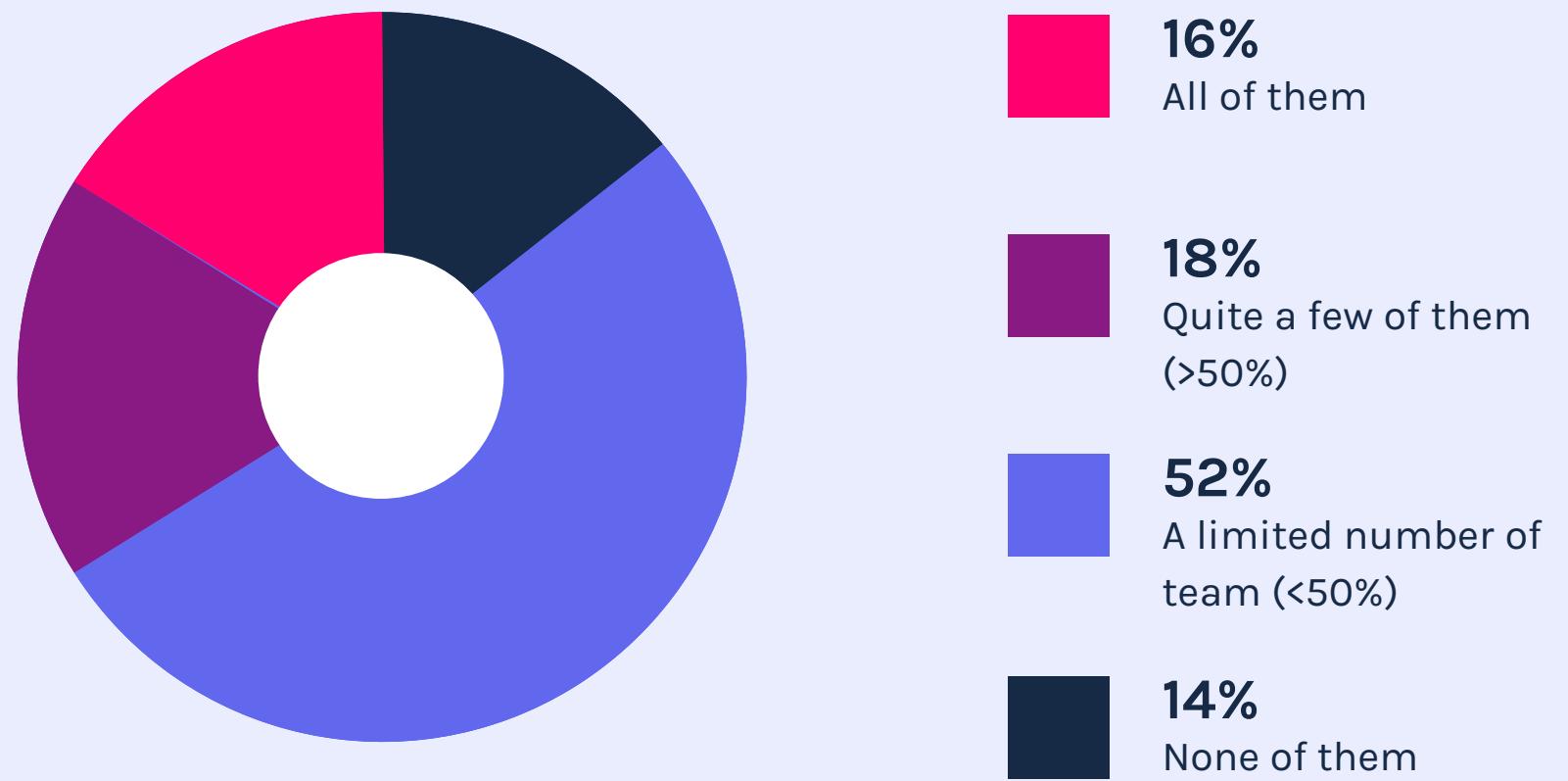
- there is also a trend where teams being located in very separate departments or even different territories can lead to silos that do not help organizations to advance their Spatial Data Science strategy and look at spatial problems more holistically.

RESULTS

Spatial Analysis experience of Data Science teams

As part of the survey, we also asked participants how many of their Data Science team have significant experience in Spatial Analysis. A staggering 52% said that only a limited number of their Data Science team have such experience, while 14% said that none of their Data Scientists had significant experience.

How many of your Data Science team have significant experience in Spatial Analysis?



1 in 3

Data Scientists have significant expertise in spatial techniques.



Interestingly, 34% of participants shared that quite a few of their Data Science team members or all of their Data Science team members did have significant experience in spatial analysis - meaning that **only 1 in 3 Data Scientists actually have significant expertise in spatial techniques.**

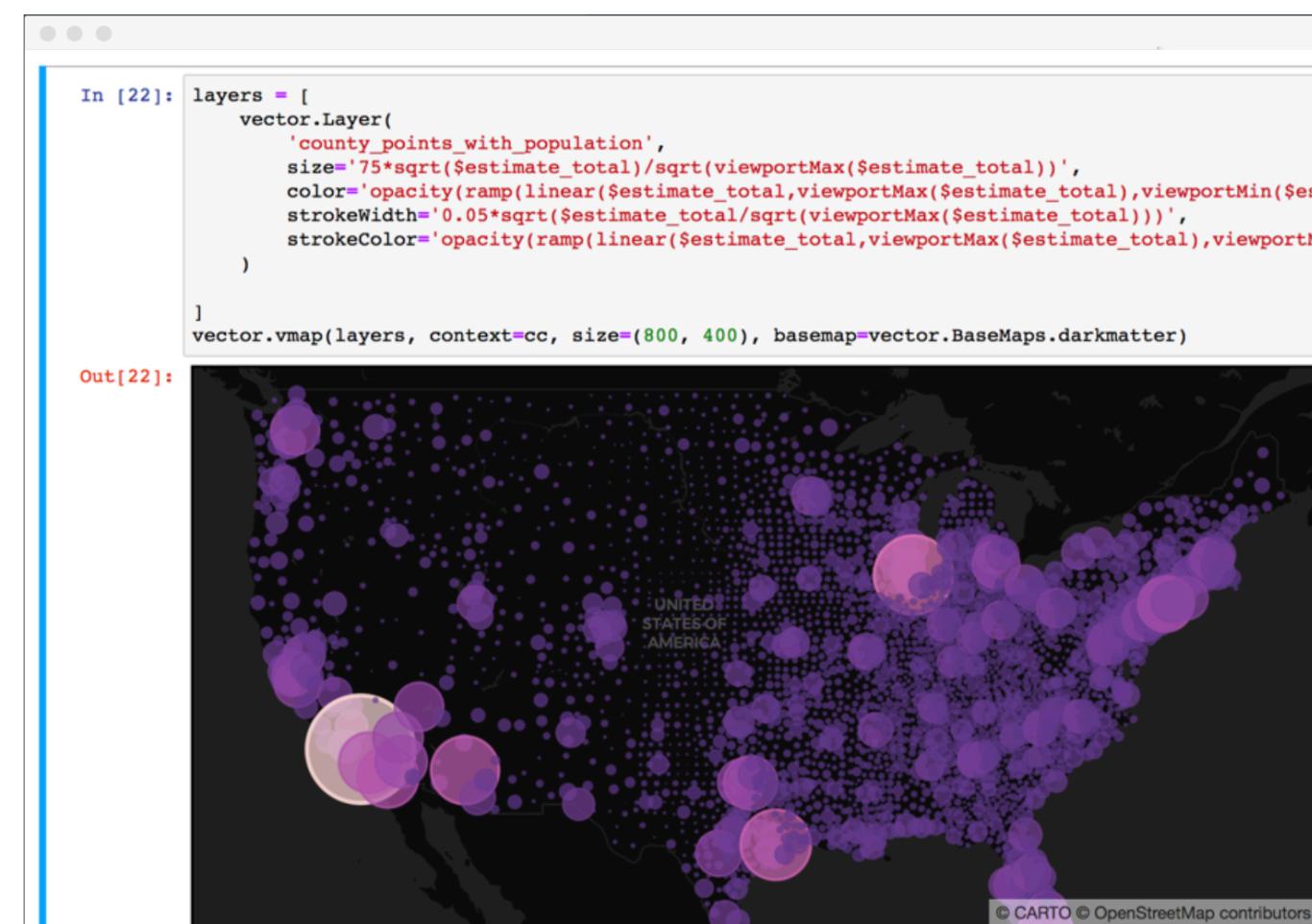
RESULTS

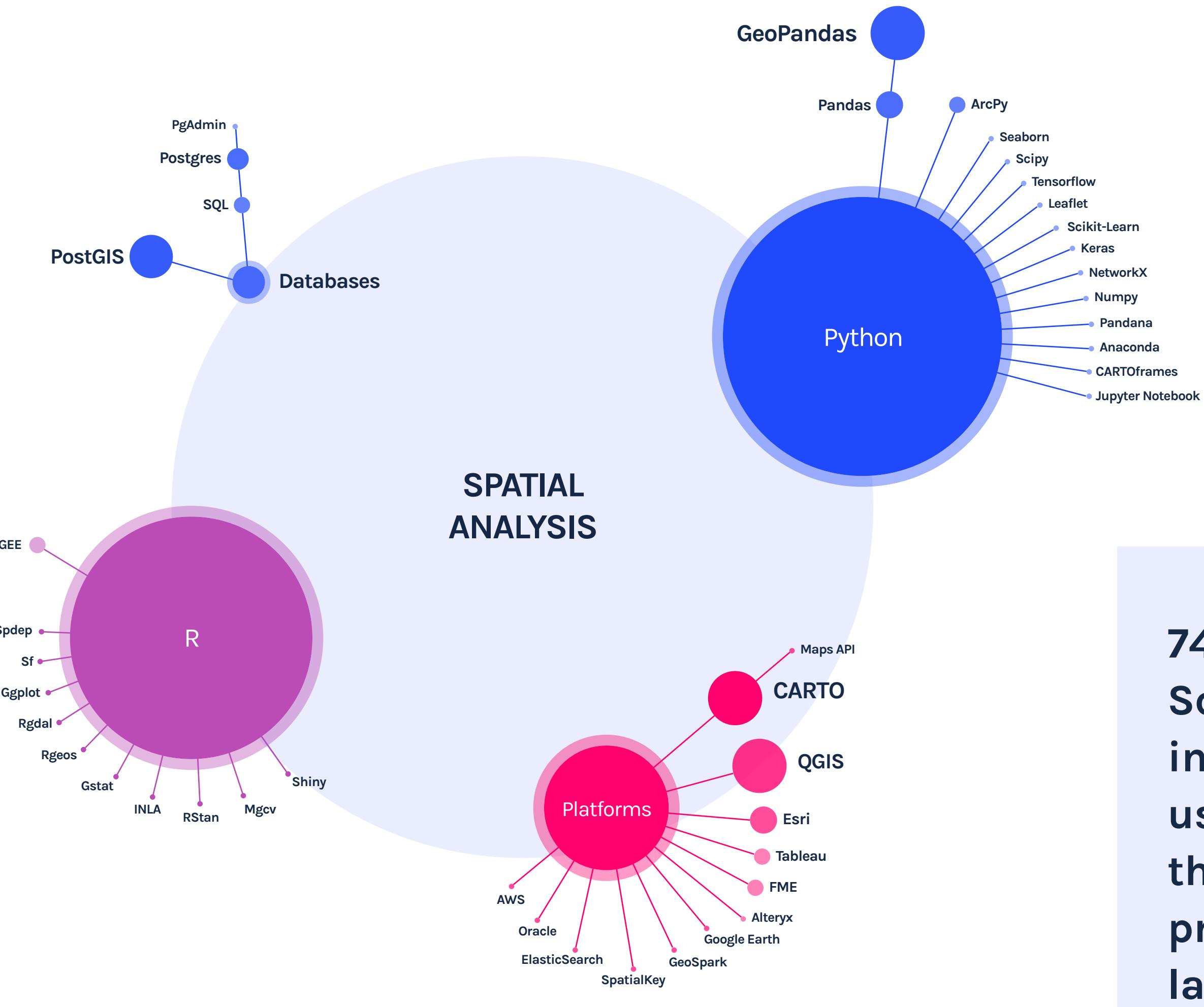
Python edged ahead of R as the most common language used in Spatial Data Science with 10% more teams using it.

We can clearly see more and more of the community favor Python due to its more unified approach

With Python's main libraries for Data Science typically noted for being better centralized and organized. However, there was recurring feedback that R still includes a more complete offer in terms of specific libraries for Spatial Data Science (vs Data Science more generally).

Other key tools included Geopandas, QGIS, PostGIS, CARTO and Esri - providing solutions for analysis and visualization of spatial models in a range of verticals.



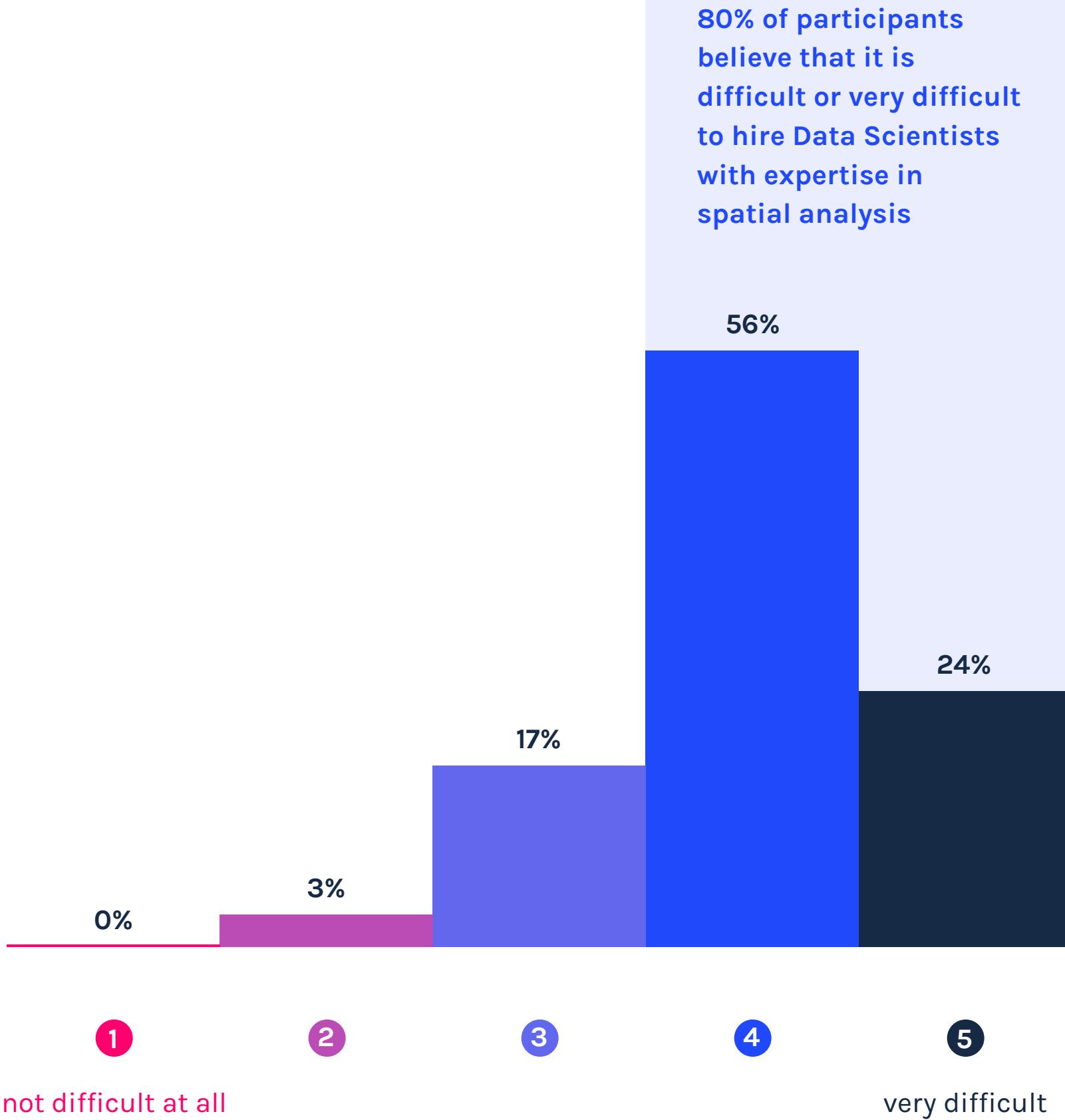


74% of Data Science teams interviewed use Python as their principal programming language

RESULTS

How difficult is it to hire Data Scientists with expertise in spatial analysis?

We also explored the challenges around hiring relevant talent in this highly competitive field - asking how difficult it really was to find individuals to join their organizations.





An optimal Spatial Data Science profile requires skills such as:

- ① Strong background in statistics
- ② Extensive experience in coding skills relating to Data Science (Spark, SQL, Python, R, Tensorflow, Pytorch)
- ③ Experience developing production-quality data products using the results of quantitative research
- ④ Extensive experience in data visualization (in Python and R or other applications)
- ⑤ Effective application of Data Science workflows to business problems, and the ability to storytell around results
- ⑥ Familiarity with data pipelines and ETL practices (Airflow, scheduled notebooks, Google DataFlow, etc.)
- ⑦ Familiarity with neural networks and deep learning (e.g., Tensorflow, PyTorch)
- ⑧ Experience working with distributed computing systems like Spark or Google BigQuery
- ⑨ Experience working with GIS software such as CARTO, QGIS, or ArcGIS

A growing number of leading management consulting firms are scrambling to identify such profiles in the market - especially at a time when multiple (and high performing) verticals are all looking for similar skill sets, such as Private Equity, Real Estate, and Tech.



Most individuals that fit such criteria have limited professional experience (1 to 6 years), having graduated from programs in Urban Planning, Math, Physics, or Computer Science.*

The difficulty that organizations are facing in finding candidates is leading to regular delays in spatial modeling initiatives moving forward within enterprise organizations, and an added demand for management consulting firms who need to ramp up their Advanced Analytics and Data Science teams fast enough to serve a growing number of spatially focused projects - particularly in site planning, logistics optimization and territory management.

Organizations who have the added challenge of trying to hire the above profile, with industry-specific expertise (e.g. in Retail or Insurance) are facing even further delays and difficulties in sourcing candidates in the market.

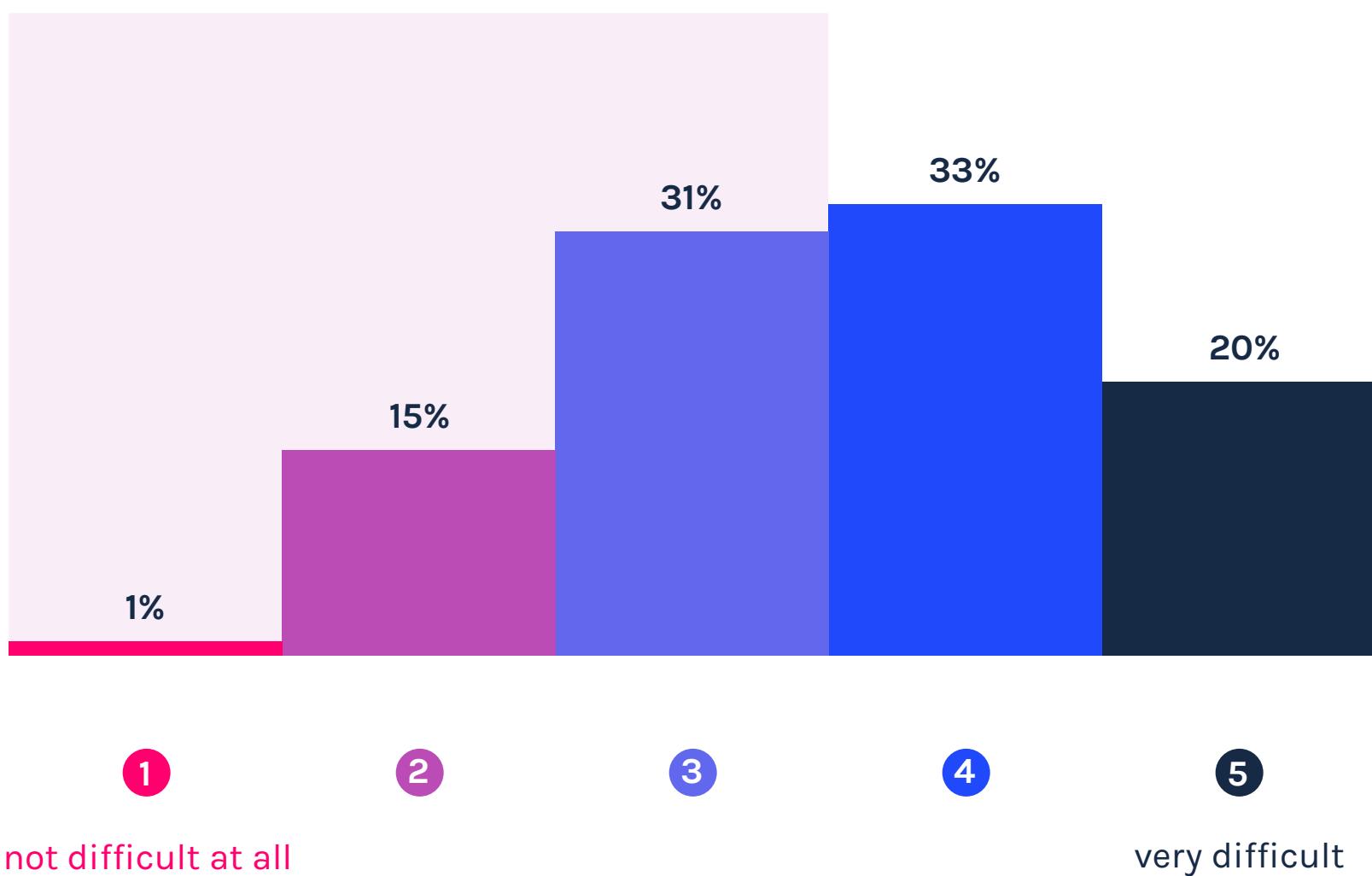
* However, a growing number of academic institutions are starting to offer specialized masters and PhD programs (as outlined in the [Becoming a Spatial Data Scientist ebook](#)) in Spatial Data Science.

RESULTS

How difficult is it to find the right software & data to support Spatial Data Science projects?

Although it might be challenging to find the right people to fill Spatial Data Science roles, most participants seemed to agree that there are plenty of datasets and software solutions available for them to be successful once hired.

47% of participants do not find it challenging to identify the right software & data to support Spatial Data Science projects

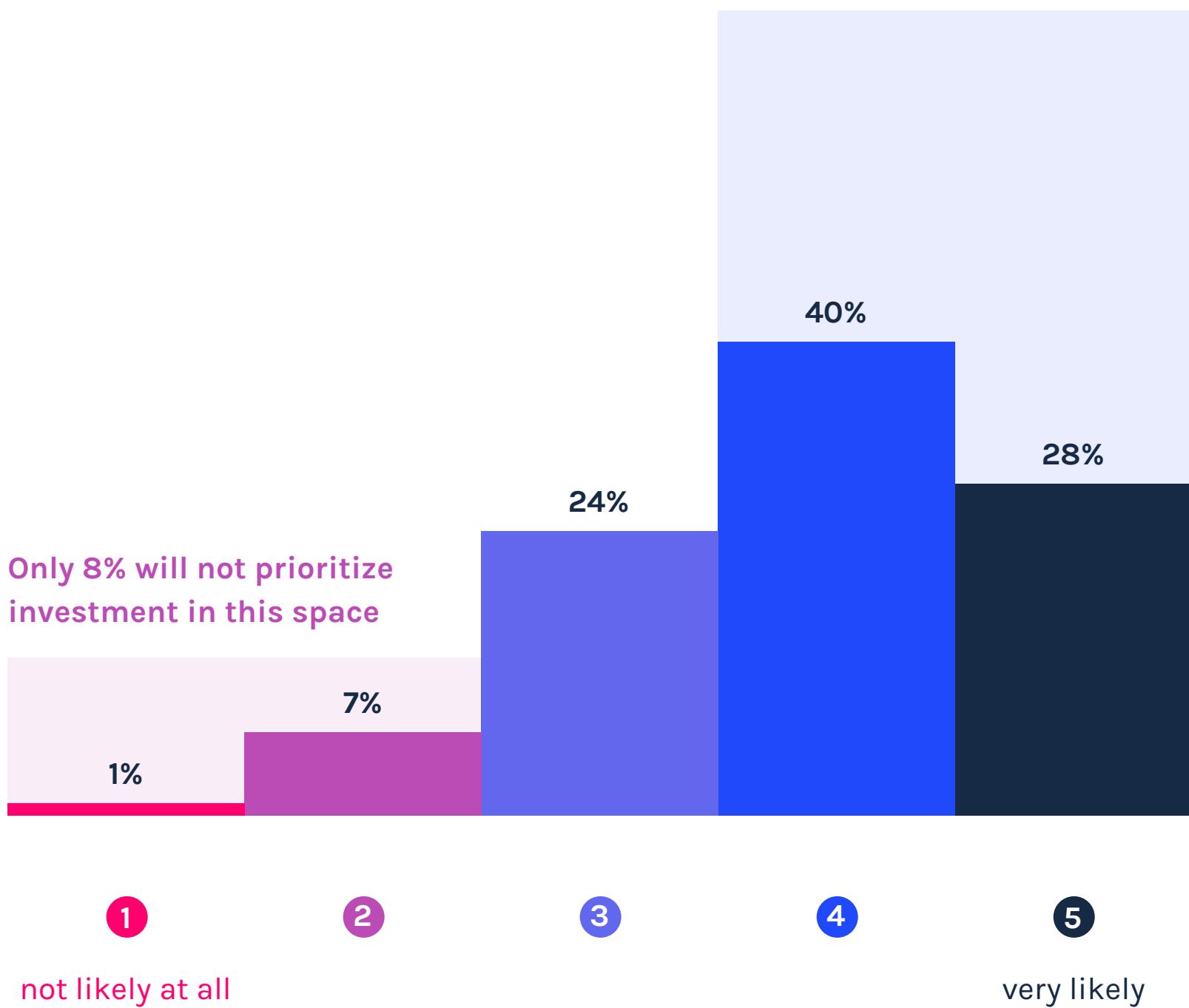


RESULTS

How likely are you to increase your investment in Spatial Data Science in the next two years?

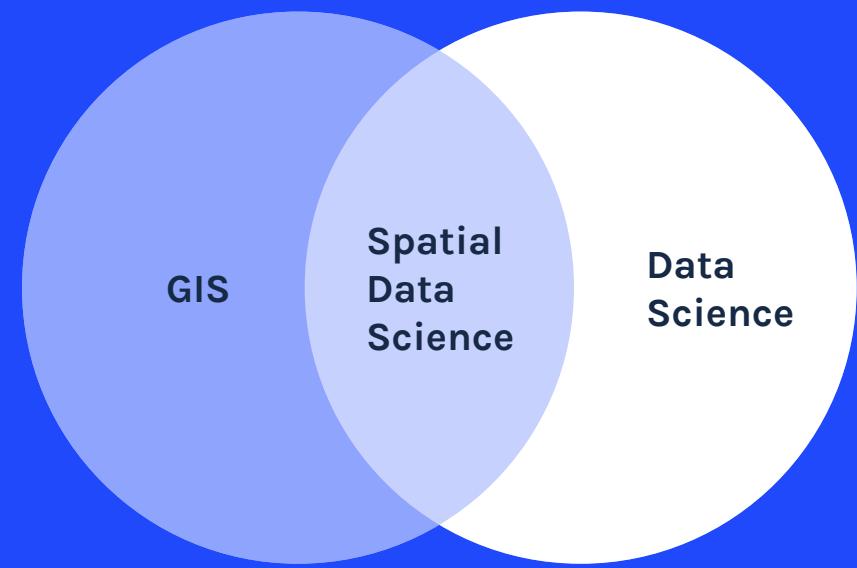
The large majority of organizations will be prioritizing investment in this space in the next 2 years. This could be due to the fact that these organizations are still early in their adoption of Spatial Data Science, or they are already making significant investments and do not need to continue to invest at the same rate as before.

68% of organizations are likely to increase their investment in Spatial Data Science in the next 2 years



Conclusion

Nestled between Data Science and GIS, **Spatial Data Science** is the future of GIS and professionals skilled in both are in high demand.



Based on the results of this survey, it is evident that Spatial Data Science is a field set to flourish across dozens of sectors

With enterprises beginning to de-silo GIS functions and encouraging collaboration with Data Science teams, we can expect increased leveraging of spatial solutions and techniques amongst early adopters, and rising demand for talent as other companies start to play catch-up.

And while there seems to be commitment to significant investment in Spatial Data Science in the very near-term, executives need to be prepared for a significant portion of that budget to be earmarked for hiring. With low supply and high demand for the complex Spatial Data Scientist persona, hiring managers should expect competition for talent to be intense.

That being said, the pool of Spatial Data Scientists should expand to meet that demand fairly rapidly, **due to three changes:**

- ① The global proliferation of Spatial Data Science programs, will provide training at the undergrad, post-grad, and doctoral levels.
- ② The de-siloing of GIS will allow traditional GIS professionals to up-skill on Python, R, and other skills that will enable them to more effectively perform Spatial Data Science functions.
- ③ The increasing accessibility of tools and resources will allow traditional Data Scientists to integrate Spatial into their workflows.

Given its growing importance as a driver of critical insight for nearly every industry, those enterprises looking to stay ahead should be making Spatial Data Science a high priority from the start of this new decade.

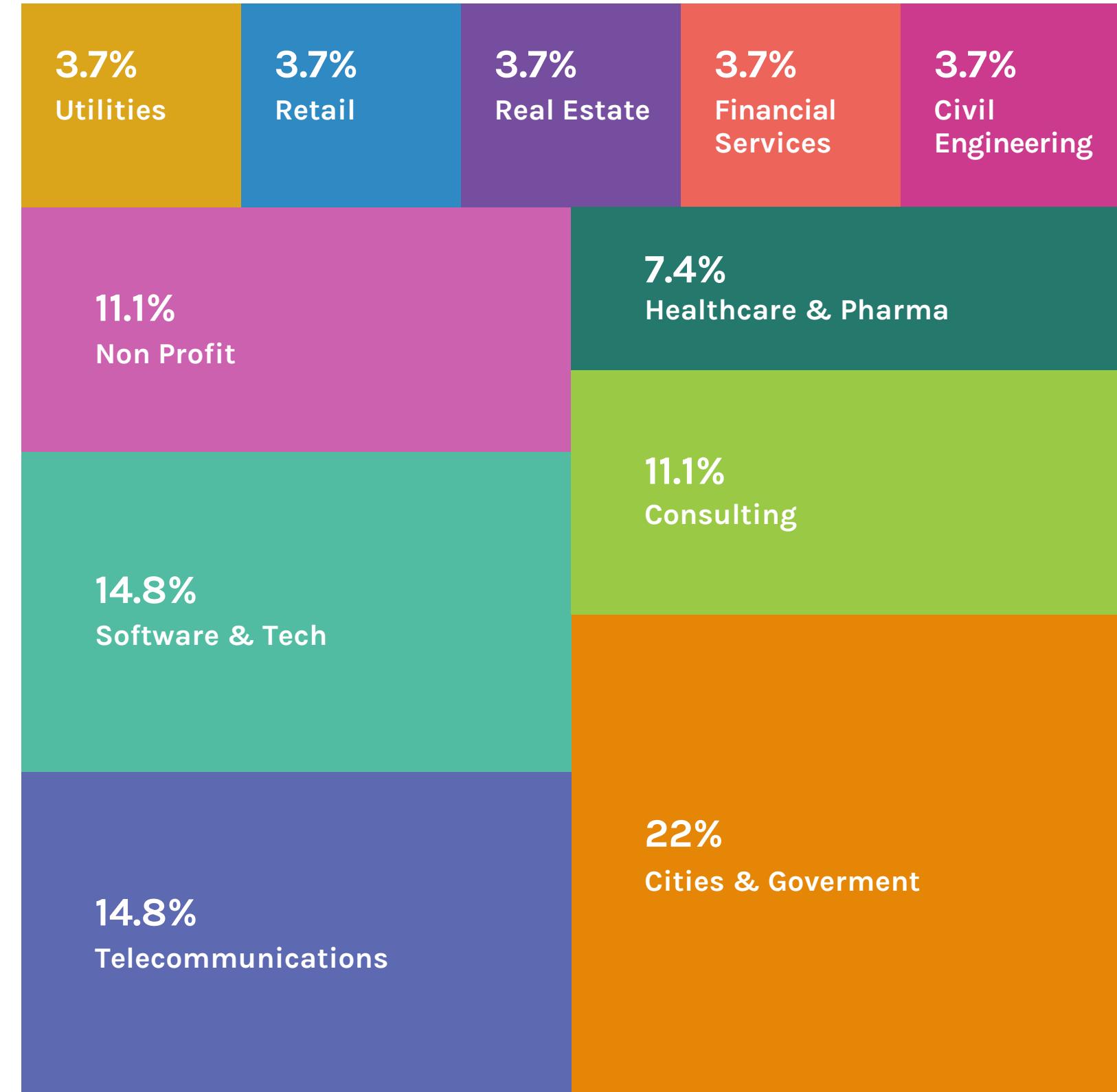
Participants

PARTICIPANTS

Industry participants

We asked more than 150 thought leaders from enterprise organizations worldwide to share their perspective. As you can see below, we chose to include Non Profit organizations as there are many experts in Spatial Data Science working in that space - using sophisticated tools and hiring top talent to use spatial analysis to have a positive impact on our society.

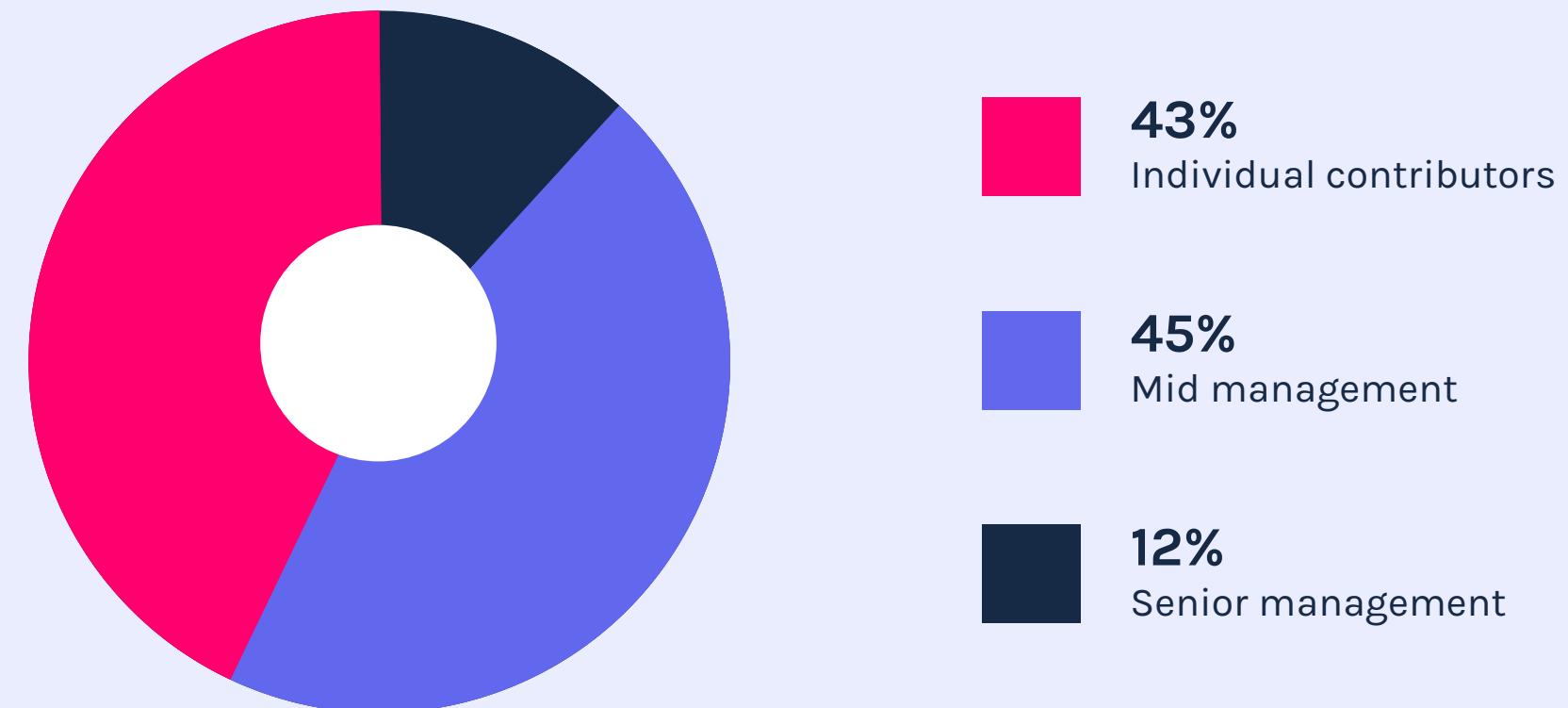
As you can see, the other top 4 segments included **Cities & Government**, **Telecommunications**, **Software & Tech**, and **Consulting**:



PARTICIPANTS

Seniority

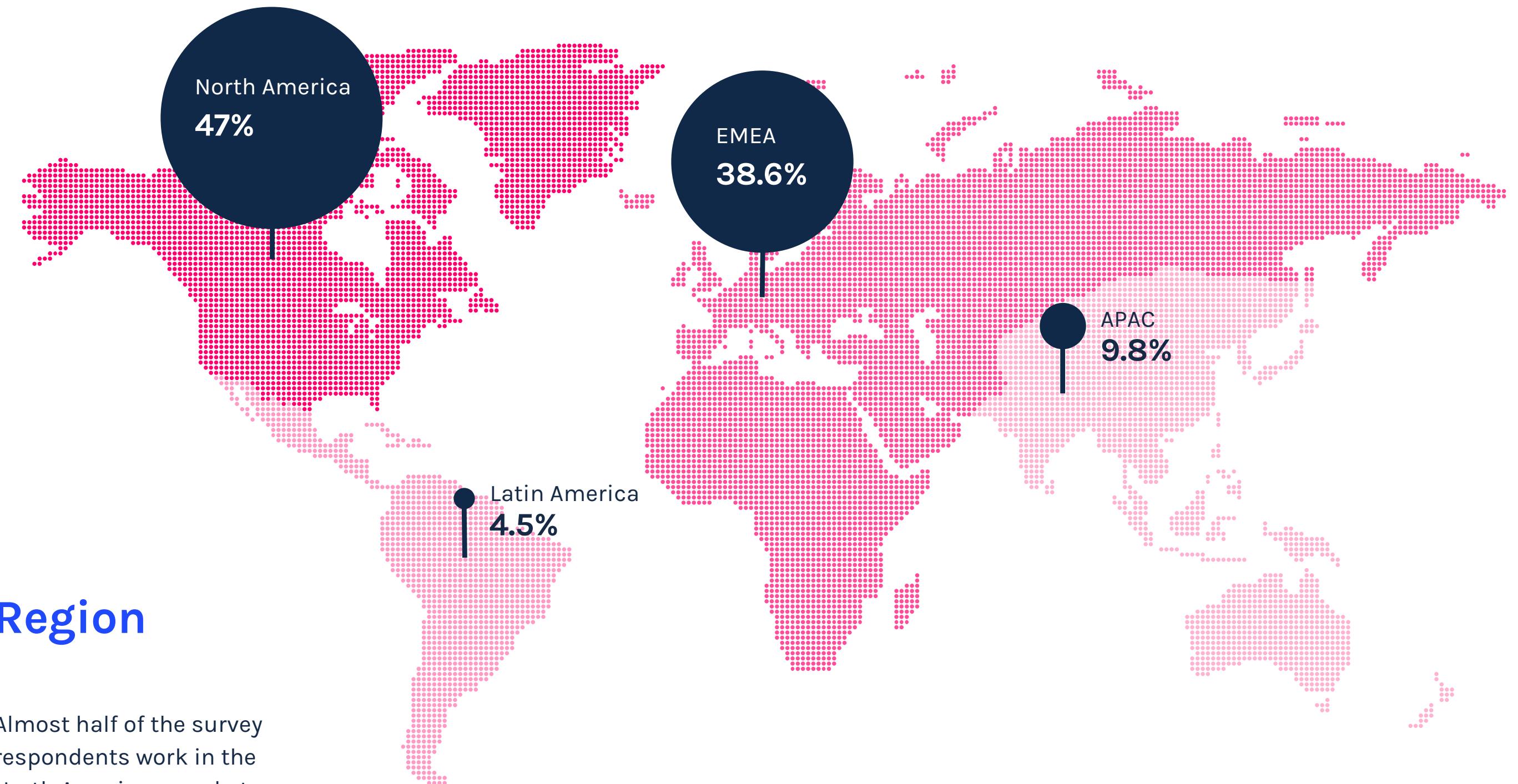
In terms of the level of respondents, we saw 3 tiers of participants.



Individual contributors in Data Science, Development, GIS, or Analysis roles - who may not manage teams, but manage highly complex and important analytical projects within their organizations.

Mid management were typically VP level - leading Data Science, GIS, Advanced Analytics, or Business Intelligence teams, with multiple technical stakeholders reporting to them.

Senior management included executives responsible for large business areas with many different teams reporting to them. This ranged from Chief Data Officers to Chief Information Officers and SVPs for areas such as operations, technology, and marketing - depending on the vertical in question.



Almost half of the survey respondents work in the North American market, closely followed by the EMEA region with 38% of responses. APAC and Latin America were also well represented with 9.8% and 4.5% of responses respectively:

If you have any questions
about the content of this
survey, please contact
hello@carto.com

To learn more about Spatial
Data Science, why not join us
at one of this year's Spatial
Data Science Conferences in
London or NYC?

[Visit website](#)