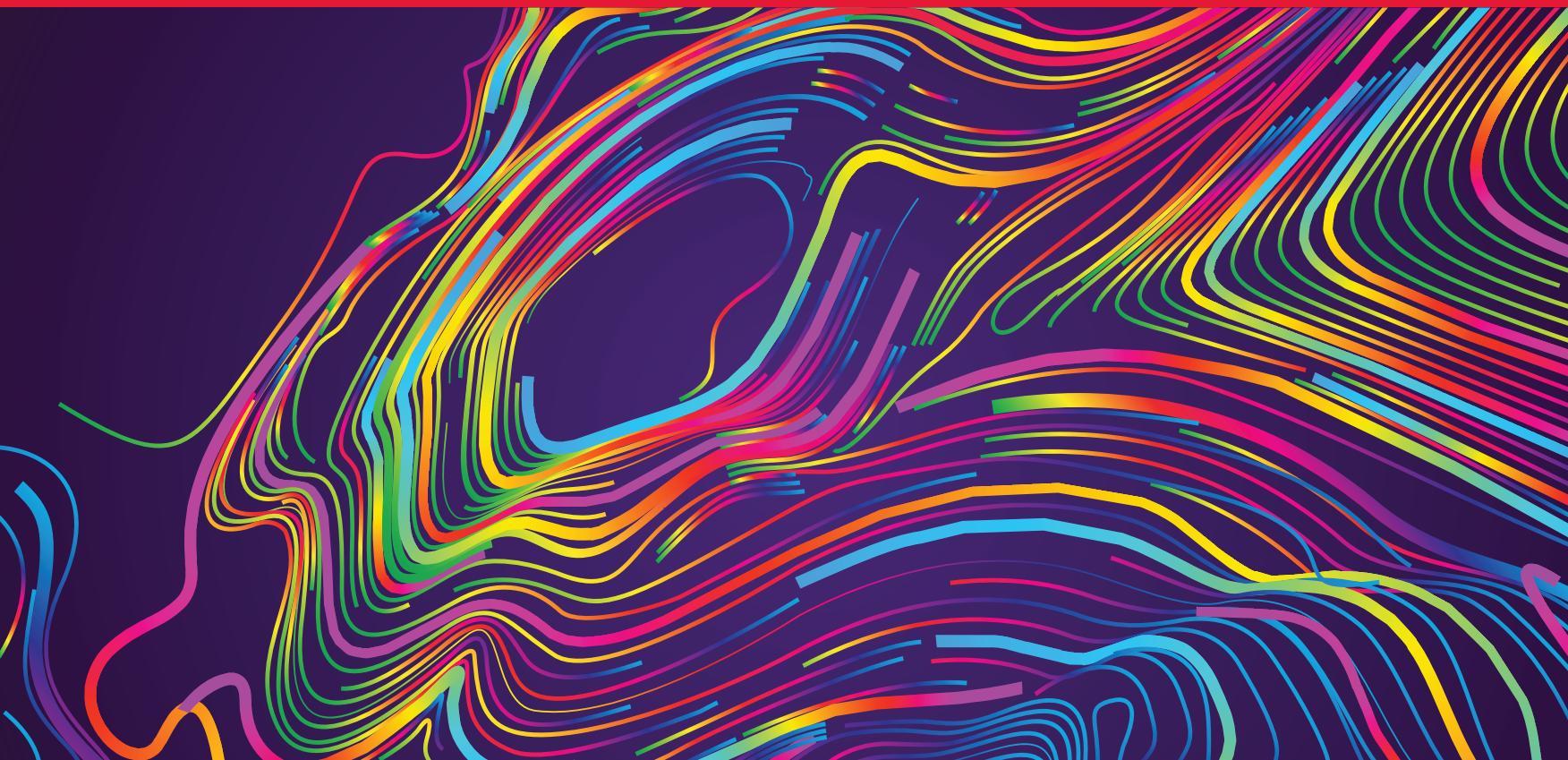


2017 Data Science Salary Survey

Tools, Trends, Titles, What Pays (and What Doesn't) for Data Professionals



Brian Suda

Get a clear picture of the salaries and bonuses data science professionals around the world receive, as well as the tools and cloud providers they use, the tasks they perform, and how interpersonal ("soft") skills might affect their pay. The fifth edition of O'Reilly's online Data Science Salary Survey provides complete results from nearly 800 participants from 69 countries, 42 US states, and Washington, DC.

With five years of data, the survey's results are consistent enough to reliably identify changes and trends. The survey asked specific questions about industry, team, and company size, but also posed questions such as, "How easy is it to move to another position?" or "What is your next career step?" You can plug in your own data points to the survey model and see how you compare to other data science professionals in your industry.

With this report, you'll learn:

- **Where data scientists make the highest salaries—by country and by US state**
- **Tools that respondents most commonly use on the job, and tools that contribute most to salary**
- **Activities that contribute to higher earnings**
- **How gender and bargaining skills affect salaries when all other factors are equal**
- **Salary differences between those using open source tools versus those using proprietary tools**
- **How the increase in respondents outside of the US signal a rise in international companies starting and growing data organizations**



Participate in the 2018 Survey:

Spend just 5 to 10 minutes and take the anonymous salary survey here: <https://www.oreilly.com/ideas/take-the-data-science-salary-survey>.



Take the Data Science Salary Survey

As data analysts and engineers—as professionals who like nothing better than petabytes of rich data—we find ourselves in a strange spot: we know very little about ourselves. But that's changing. This salary and tools survey is the fifth in an annual series. To keep the insights flowing, we need one thing: **PEOPLE LIKE YOU TO TAKE THE SURVEY.**

Anonymous and secure, the survey will continue to provide insight into the demographics, work environments, tools, and compensation of practitioners in our field. We hope you'll consider it a civic service and participate today.

[Take the Survey](#) 

2017 Data Science Salary Survey

Tools, Trends, What Pays (and What Doesn't) for Data Professionals

Brian Suda

O'REILLY

2017 DATA SCIENCE SALARY SURVEY

by Brian Suda

Editor: Colleen Torporek

Designer: Ellie Volckhausen

Production Editor: Shiny Kalapurakkal

Copyright © 2017 O'Reilly Media, Inc. All rights reserved.

Printed in Canada.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North,
Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (<http://safaribooksonline.com>). For more information, contact our corporate/institutional sales department: 800-998-9938 or corporate@oreilly.com.

2017-08-25. First Edition

ISBN: 978-1-491-97750-7

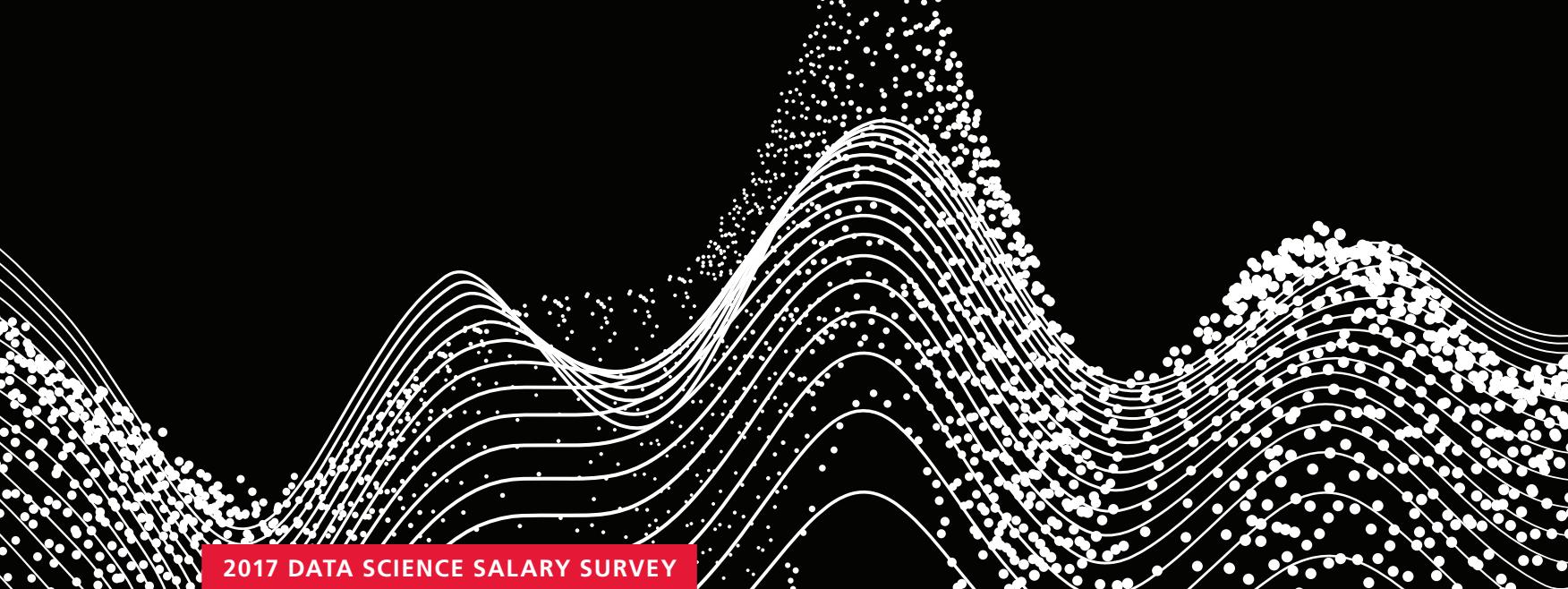
REVISION HISTORY FOR THE FIRST EDITION

2017-08-25: First Release

While the publisher and the author have used good faith efforts to ensure that the information and instructions contained in this work are accurate, the publisher and author disclaim all responsibility for errors or omissions, including without limitation responsibility for damages resulting from the use of or reliance on this work. Use of the information and instructions contained in this work is at your own risk. If any code samples or other technology this work contains or describes is subject to open source licenses or the intellectual property rights of others, it is your responsibility to ensure that your use thereof complies with such licenses and/or rights.

Table of Contents

2017 Data Science Salary Survey	V	Ease of Finding a New Role.....	26
Executive Summary.....	1	Self-Assessed Bargaining Skills	26
Introduction	2	Advance Your Career	28
Salary.....	5	Work Setup and Tools	30
By World Region.....	5	Operating Systems	30
By State	7	Programming Languages	30
By Gender.....	9	Relational Databases	33
Age	9	Hadoop	33
Experience.....	11	Search	37
Industry	13	Big Data Platforms.....	37
Education	16	Business Intelligence and Reporting	37
Company Age.....	18	Machine Learning.....	42
Company Size	18	Viz Tools	42
Job Titles.....	20	Importance of Tasks	42
Time Spent in Meetings	22	Conclusion	49
Time Spent Coding.....	22	Model	50
Length of Work Week.....	24		



2017 DATA SCIENCE SALARY SURVEY

HERE WE TAKE A DEEP
DIVE INTO THE RESULTS
FROM RESPONDENTS,
EXPLORING CAREER DETAILS
AND FACTORS THAT
INFLUENCE SALARY

YOU CAN PRESS ACTUAL BUTTONS (and earn our sincere gratitude) by taking the 2018 survey—it only takes about 5 to 10 minutes, and is essential for us to continue to provide this kind of research.

oreilly.com/ideas/take-the-data-science-salary-survey

Executive Summary

IN THIS FIFTH EDITION OF the O'Reilly *Data Science Salary Survey*, we analyze input from nearly 800 participants from 69 countries, 42 US states, and Washington, DC. We explore everything from salaries and bonuses to tools, cloud providers, and reporting. We also investigate how interpersonal skills—aka soft skills—might be affecting salaries.

Key findings include the following:

- We tie the drop in share of US respondents to a rise in international companies starting and growing their data organizations.
- Those who self-assess as having the best bargaining skills make substantially more than others.
- The larger the company, the higher the salary.

We analyze input from nearly 800 participants from 69 countries, 42 US states, and Washington, DC.

- Python usage is up: 63% from last year's 58%.
- Although two-thirds of respondents use Windows to accomplish at least some of their work, that's down from 74% last year.
- Spark and Spark MLlib are gaining in popularity, and worth keeping an eye on.
- Global median salary is \$90,000 (USD).

With five years of data, our results are consistent enough to reliably identify change and trends. When we see an increase in, for example, the popularity of a programming language, we can recognize a real change in the data ecosystem, and one to which it's worth paying attention. There are a few surprises this year, but most of the data is consistent with past results.

Introduction

THIS IS THE FIFTH YEAR for the *Data Science Salary Survey*, and we certainly see some trends over that time. With nearly 800 participants taking this online, self-reported survey, we can use this data to get a better picture of what tools data scientists are using, where the industry is heading, and most important, get an overview of salaries for the data community. The respondents came from 69 countries and 42 US states. This gives us a good geographic dispersion when we look at the trends.

The survey asked specific questions about salary, industry, team, and company size, but it also asked questions such as, "How easy is it to move to another position?" or "What is your next career step?" When all of these questions are put together, a better picture of the overall landscape comes into focus when looking at data in various industries.

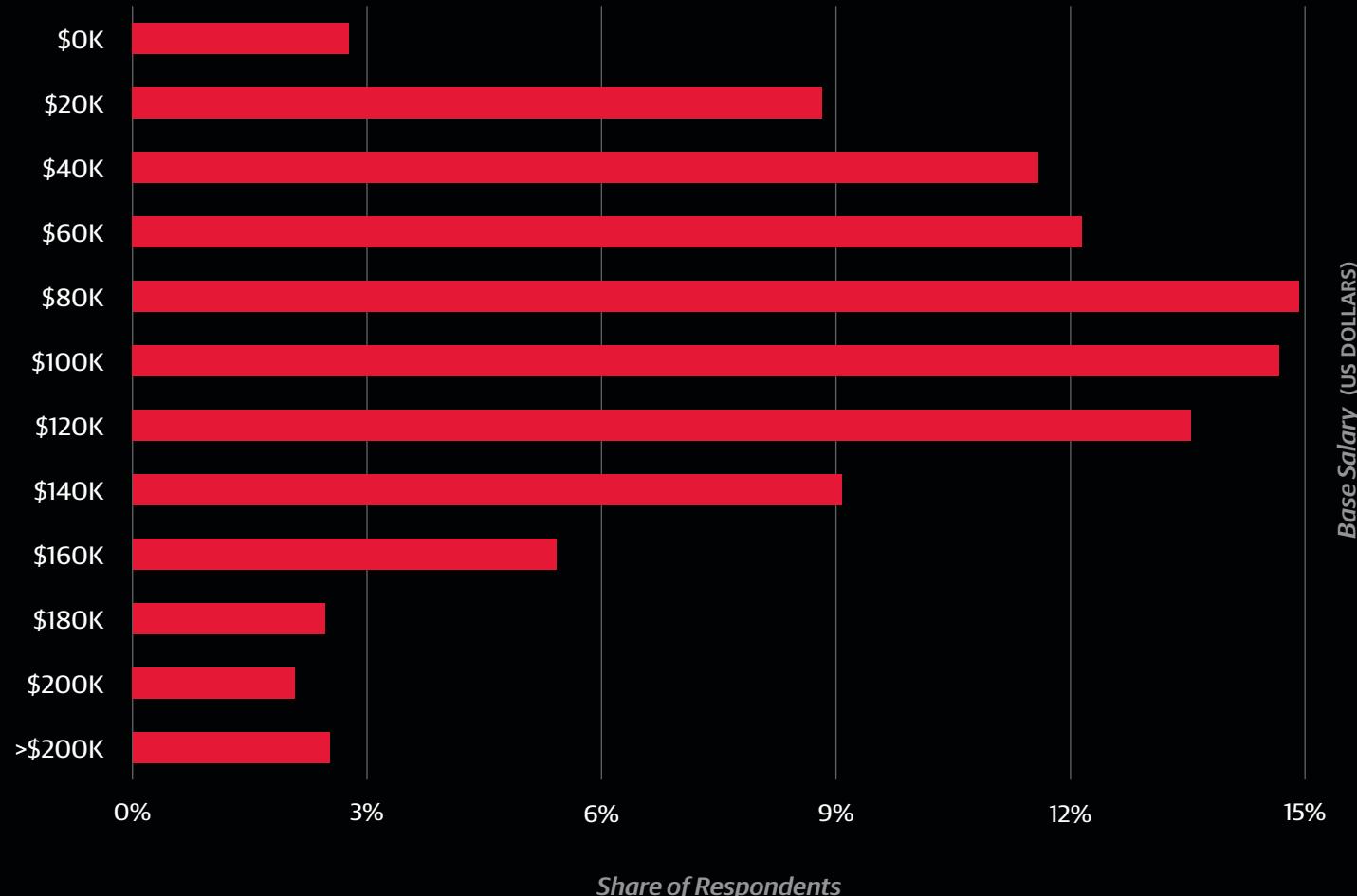
The data and model are best used to start a larger discussion about, for instance, how you compare to your peers and the industry as a whole, and the soft and hard skills you might think about acquiring in order to stay competitive and up-to-date.

In the horizontal bar charts throughout this report, we include the interquartile range (IQR) to show the middle 50% of respondents' answers to questions such as salary. One quarter of the respondents have a salary below the displayed range, and one quarter have a salary above the displayed range.

The IQRs are represented by colored, horizontal bars. On each of these colored bars, the white vertical band represents the median value.

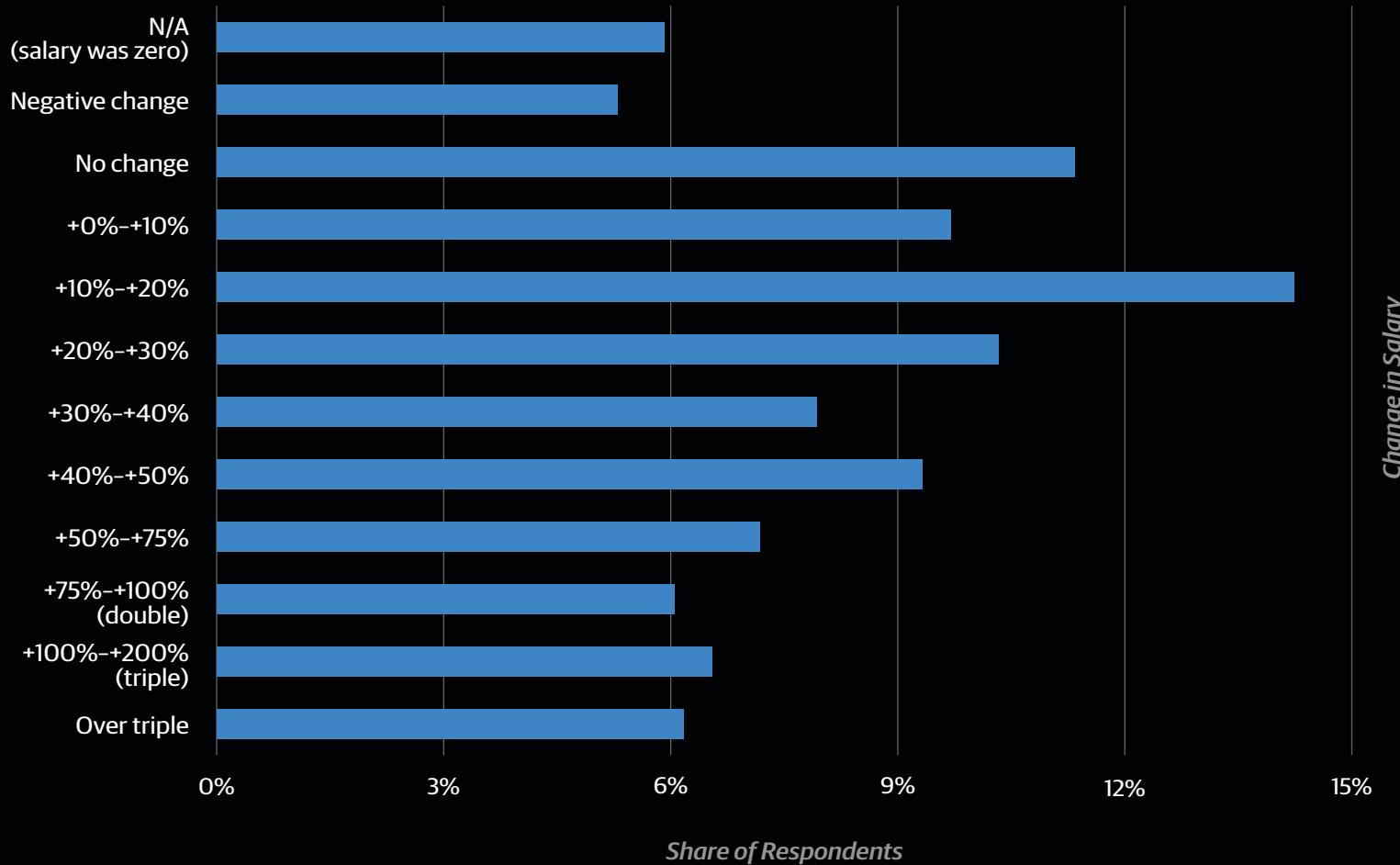
TOTAL SALARY (US DOLLARS)

SHARE OF RESPONDENTS



PERCENTAGE CHANGE IN SALARY OVER LAST THREE YEARS

SHARE OF RESPONDENTS



Salary

THE DISTRIBUTION OF SALARIES SKEWS TO THE RIGHT; that is, compared to a symmetric distribution, there are more people making extreme amounts on the high end of the scale. To compensate for that skew, we use median income as the best overall salary measure. For the 2017 survey, we find a median of \$90,000, which is up \$5,000 compared to last year's median income of \$85,000.

By World Region

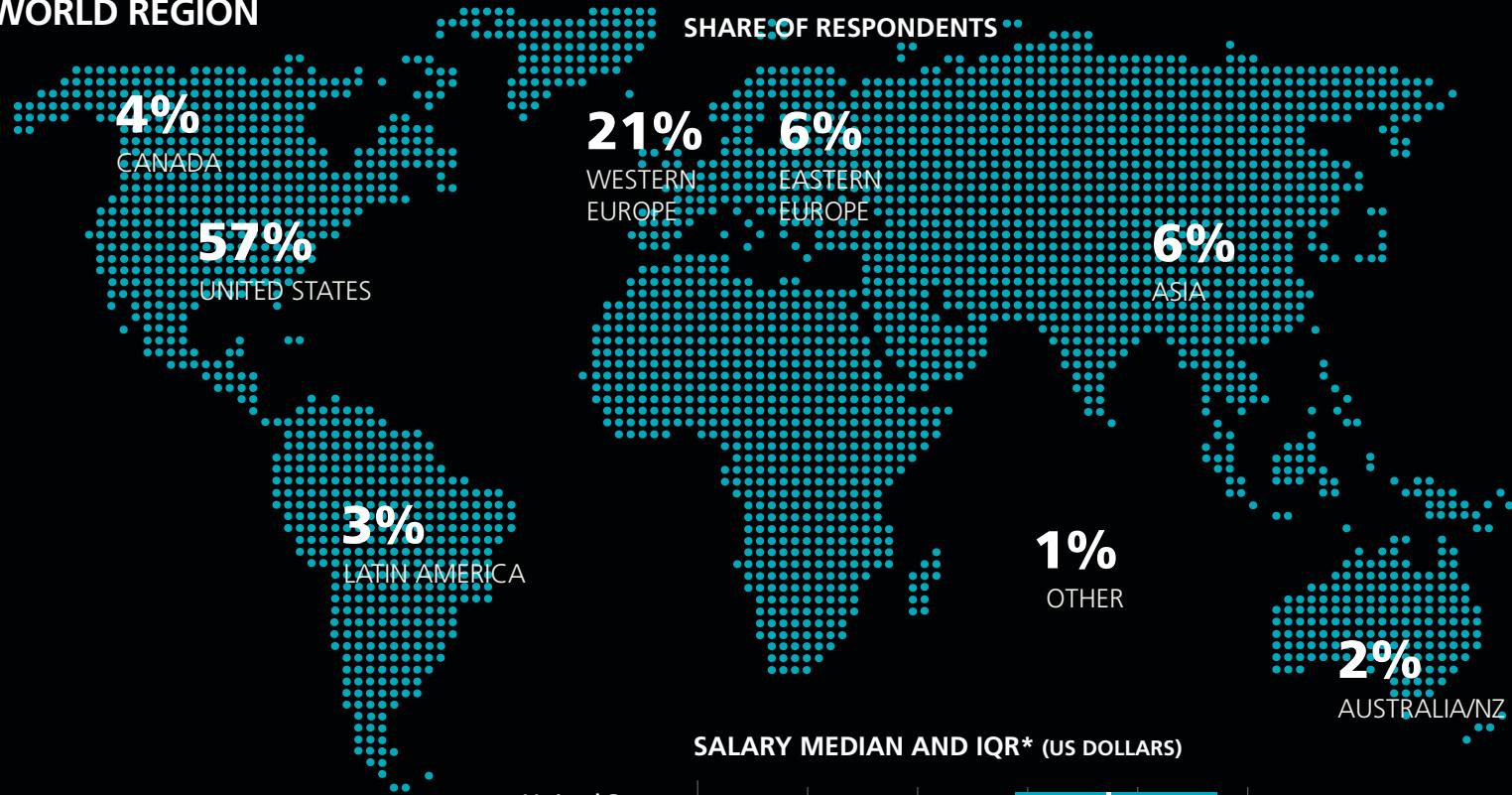
It is no surprise that the US has the highest median salaries of any region, coming in at \$112,000 (up 6.7% over last year), with 57% of

respondents. That salary is nearly double that of the Western European average of \$57,000. This phenomenon might be due to several factors; for instance, the value of the UK pound has nosedived compared to the dollar this year, the value of the Euro has also declined, and some respondents might report their salary in local currency rather than converting to US dollars.

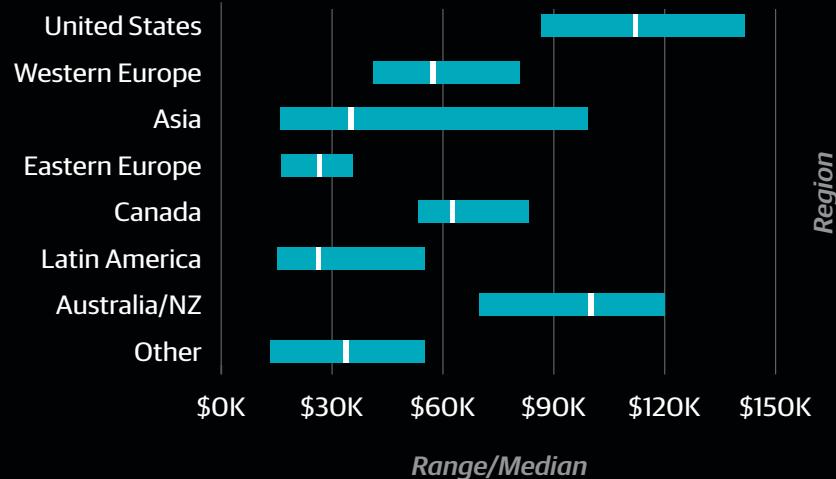
Australia and New Zealand have a healthy data culture; the two countries are second highest in pay, with a \$100,000 median salary

Australia and New Zealand have a healthy data culture; the two countries are second highest in pay, with a \$100,000 median salary. Eastern Europe shows the lowest median salary, \$27,000, but only 5.8% respondents.

WORLD REGION



SALARY MEDIAN AND IQR* (US DOLLARS)



By State

When we break down the US respondents by regions, we see California with the highest median salary, \$134,000, and the highest share of respondents, 19% (down from 22% in 2016).

This result likely reflects the large concentration of software and data-oriented companies in the San Francisco/Silicon Valley area, we also suspect O'Reilly's local presence in the Bay Area might attract more respondents. California salaries are up slightly, just under 5%, compared to 2016's \$128,000 median—in line with the overall US trend.

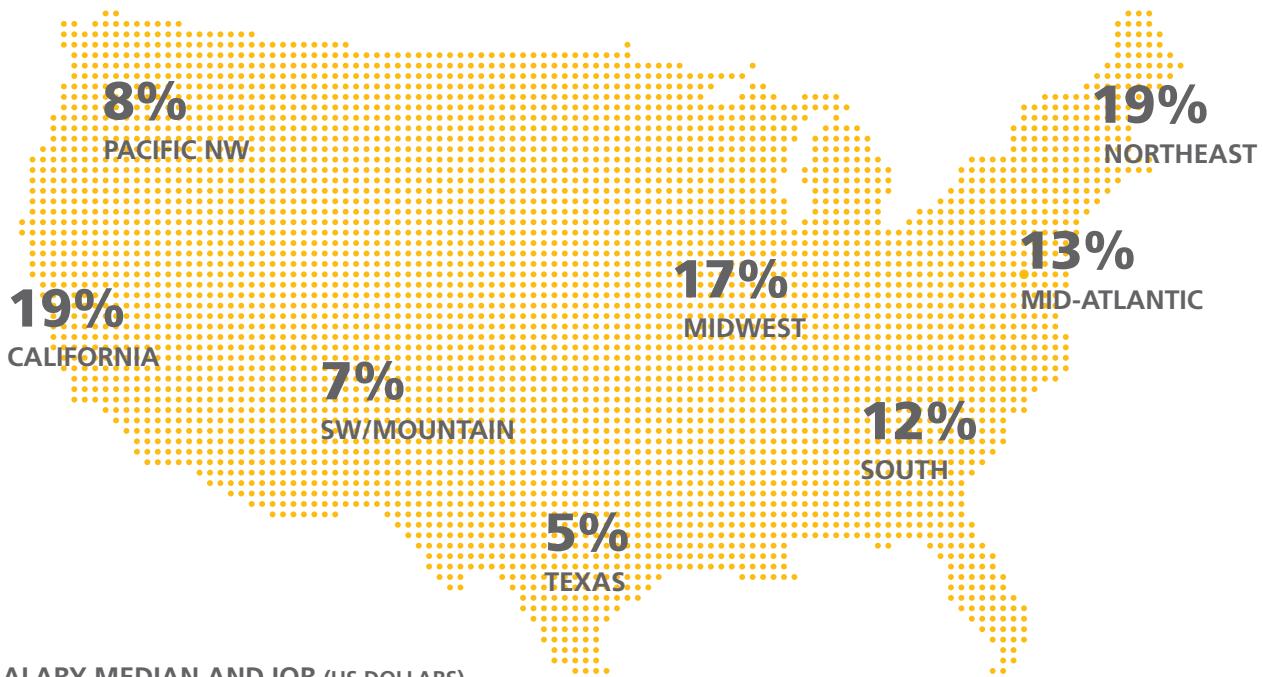
The Northeast is the next largest group of respondents (18.5%) as well as the next best paid, at \$119,000 median salary.

The regions with the smaller share of respondents, Texas (5%) and the Midwest (17%), have a lower cost of living and different mix of industries, which might explain their lower \$97,000 median salary

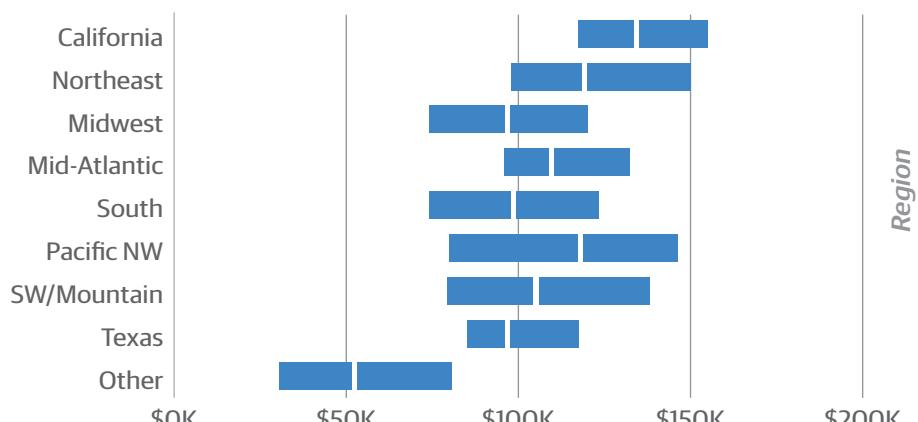
The regions with the smaller share of respondents, Texas (5%) and the Midwest (17%), have a lower cost of living and a mix of industries, which might explain their lower \$97,000 median salary—a salary still above the \$90,000 median for all respondents.

US REGION

SHARE OF RESPONDENTS



SALARY MEDIAN AND IQR (US DOLLARS)



2017 DATA SCIENCE SALARY SURVEY

By Gender

This year, we are seeing a similar number of female respondents as last year (21%). Women's salaries have stayed about the same since last year, rising from \$82,000 to \$84,000, whereas men's salaries have increased from \$88,000 to \$93,000.

The percentage of women participating in data science is still more than double that of other salary surveys O'Reilly runs, including programming and operations.

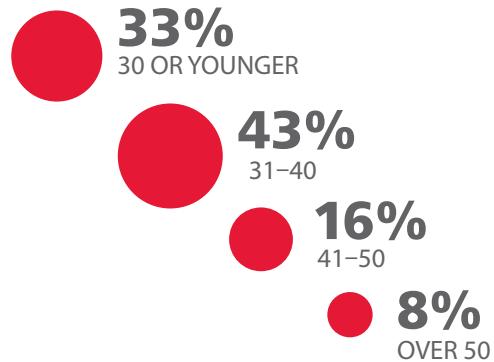
By Age

The age range of people who responded to our *Data Science Salary Survey* certainly skews youngish. More than 75% were younger than 40, and 43% were between 31 and 40.

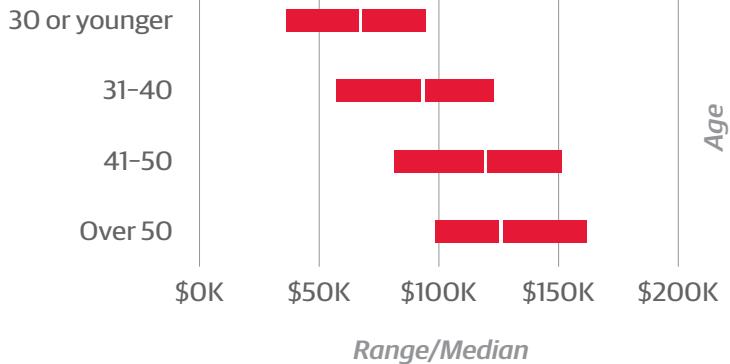
Only 24% of the respondents were older than 40, but they have the highest median salaries, with 41- to 50-year-olds making \$119,000, and those older than 50 were making \$126,000—nearly double than respondents younger than 30, who report a \$67,000 median salary.

AGE

SHARE OF RESPONDENTS

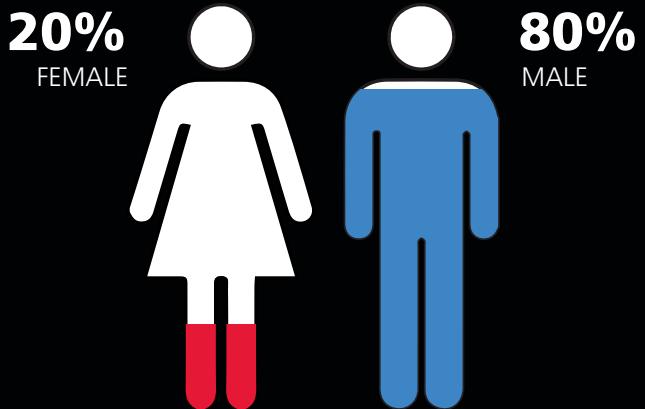


SALARY MEDIAN AND IQR (US DOLLARS)

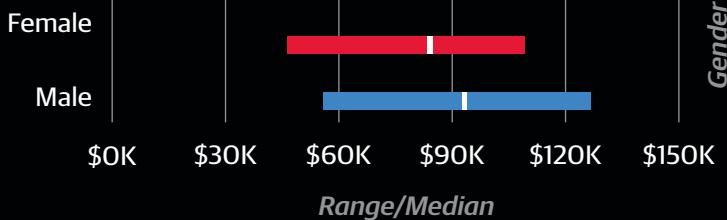


GENDER

SHARE OF RESPONDENTS



SALARY MEDIAN AND IQR (US DOLLARS)



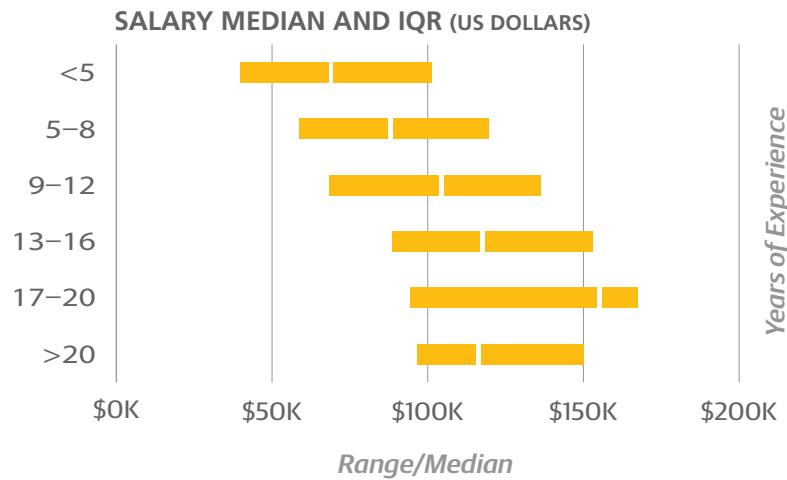
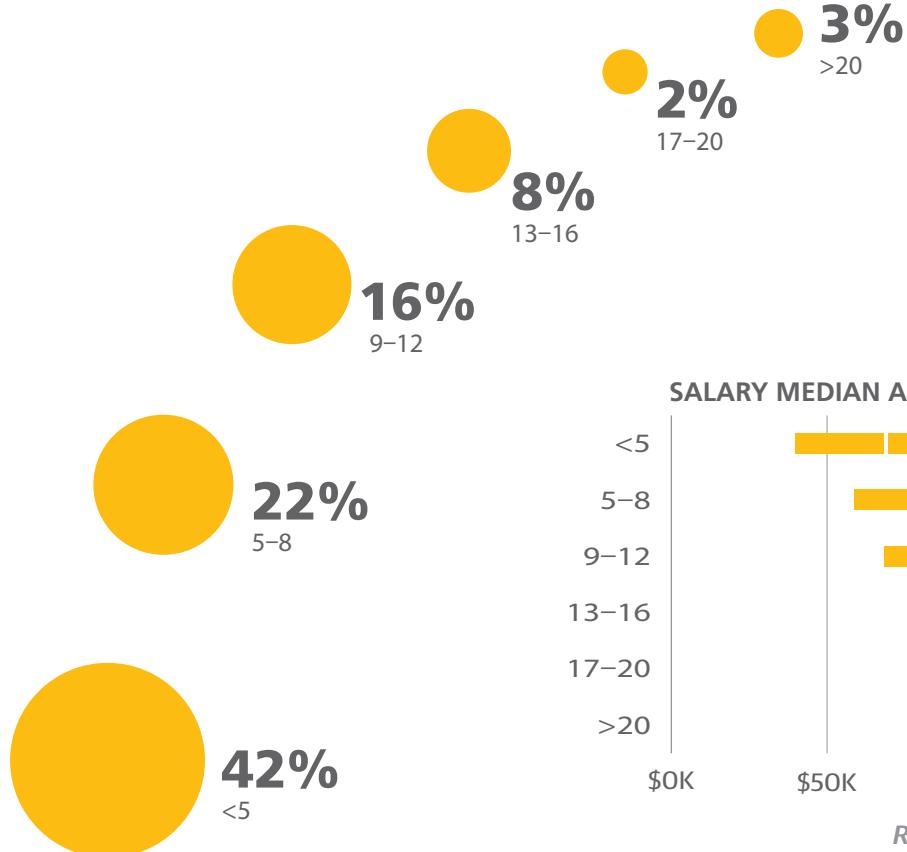
2017 DATA SCIENCE SALARY SURVEY

By Experience

With experience, the more years you have, the greater your median pay—with one exception. The group with the highest level of experience (more than 20 years) had a significantly lower pay rate than those with 17 to 20 years' experience; a drop from \$155,000 to only \$116,000. This data might be explained by the low response rates of those with so many years' experience. Fewer than 3% of respondents are in the "Over 20 years" bucket.

YEARS OF EXPERIENCE (IN YOUR FIELD)

SHARE OF RESPONDENTS



Industry

Software, consulting, and banking/finance are the top three industries in which our respondents work, at 21%, 15%, and 8%, respectively.

In 2016, the software industry was still in top place, with 17% of our respondents. As a percentage of respondents, software has grown 4% over the past year. Consulting was still number two at 15%, and banking/finance was at 8%. The downside of software taking a larger slice of the pie is that it now represents more types of workers. The median salary for someone in the software industry has dropped from \$98,000 in 2016 to \$93,000 in 2017.

There are several industries that do employ small numbers of data scientists and seem to pay them very well. Media/entertainment had only 3.3% of respondents but paid

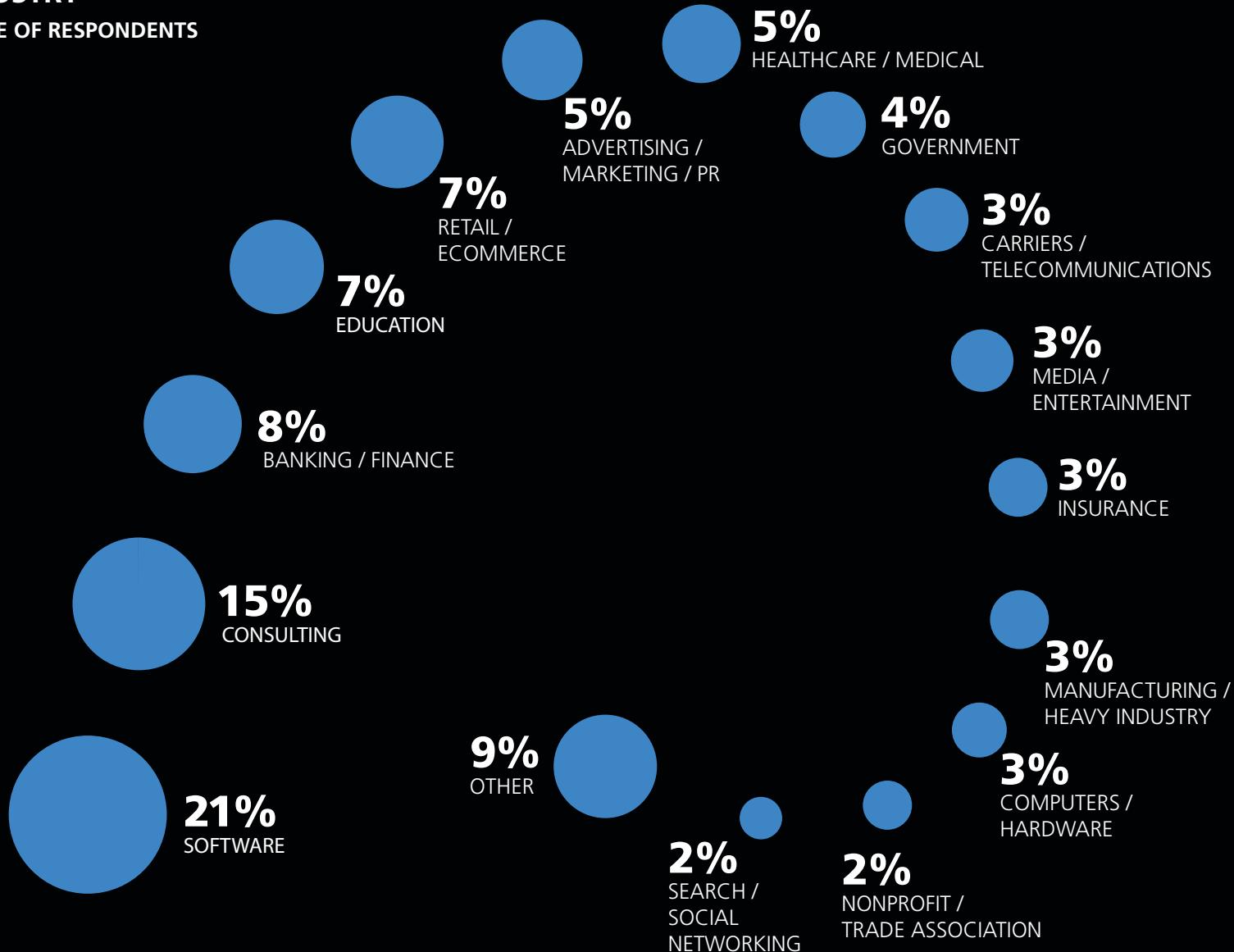
The median salary for someone in the software industry has dropped from \$98,000 in 2016 to \$93,000 in 2017.

them a \$103,000 median salary. Search/social networking had only 1.5% of the respondents, but paid a healthy \$118,000 median salary. Although these are high salaries, they could simply be outliers in the datasets. If more respondents had answered from those industries, the median average might regress more to the mean.

The lowest paid industry was nonprofit/trade association, with only a \$60,000 median salary, but if we look at the third quartile, the median salary was \$101,000, which is much closer to other industries. That said, banking/finance, which was the third most popular industry with our respondents, pays a median salary of only \$79,000—less than the global median of \$90,000.

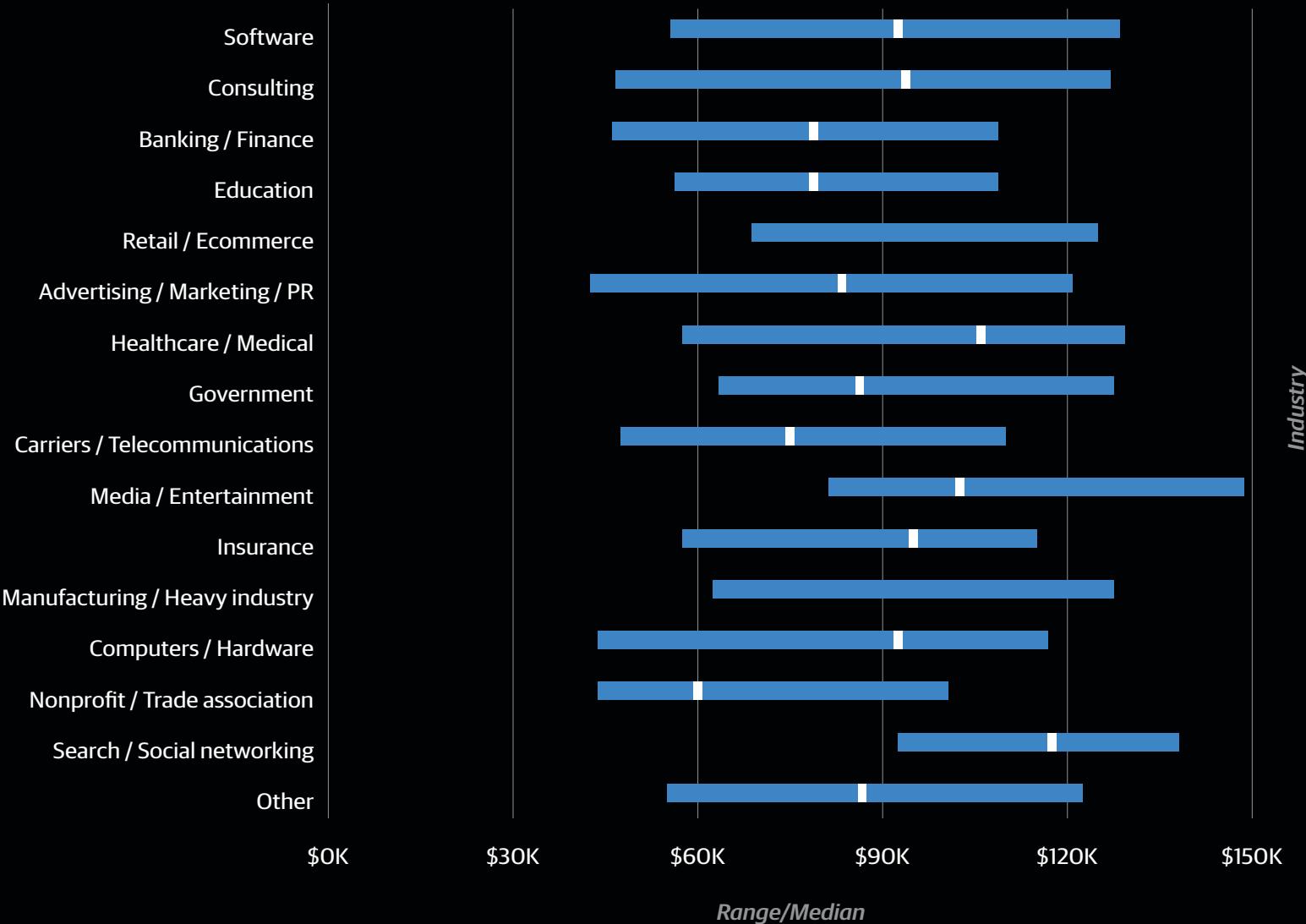
INDUSTRY

SHARE OF RESPONDENTS



INDUSTRY

SALARY MEDIAN AND IQR*



Education

More than 75% of respondents have a graduate degree, 56% have a master's, and 26% have a doctorate.

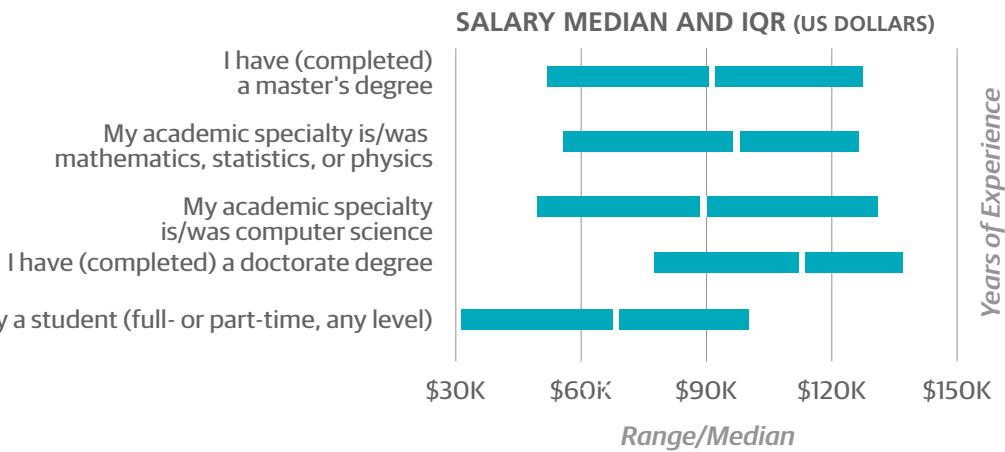
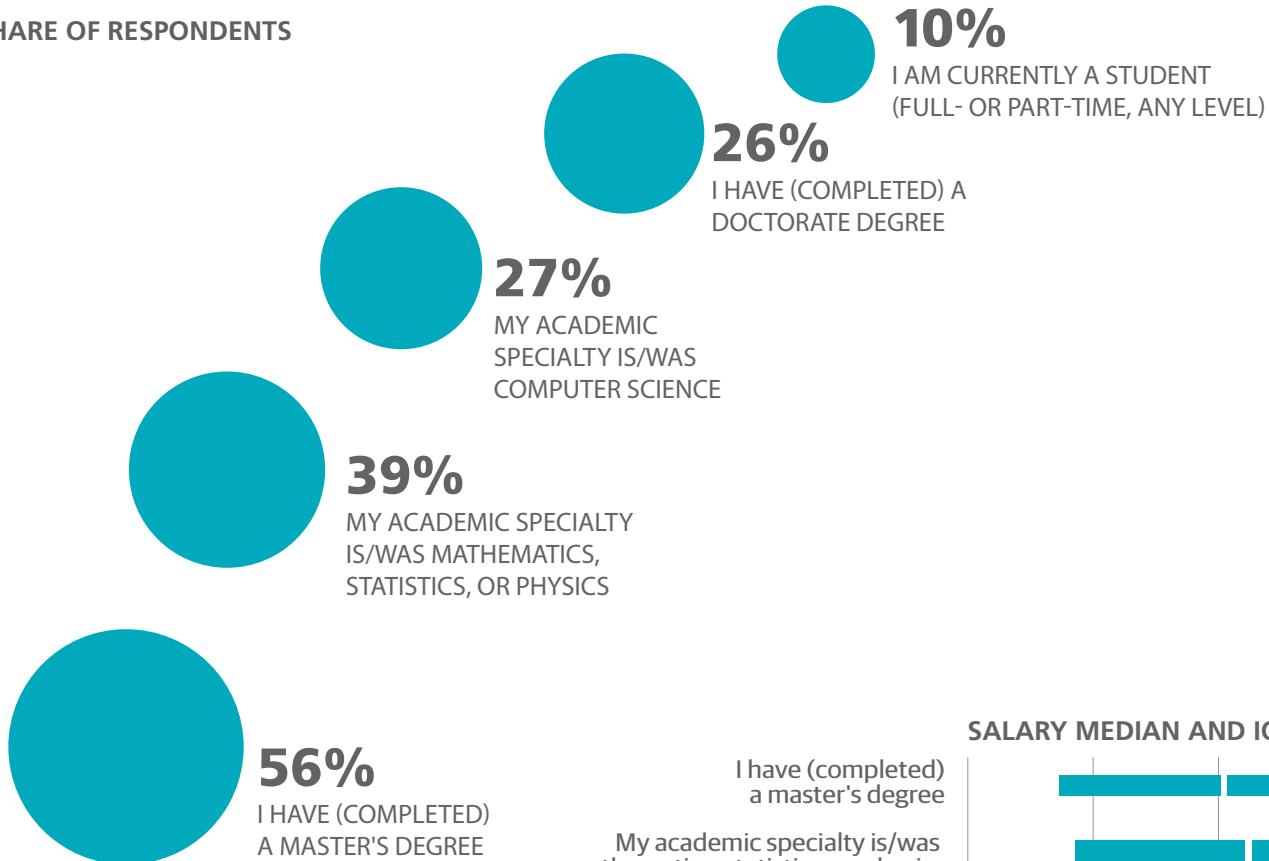
There is definitely an increase in salary as your degree increases. Students have a median salary of \$68,000, whereas computer science majors with a degree have a salary of \$89,000; those with a master's earn \$91,000, and doctorates receive \$113,000. You should keep in mind that just because you have a higher degree, that doesn't automatically mean that you can expect a higher wage. Having a deeper knowledge on one specific, niche topic might be in high demand, or the types of companies needing those skills might pay better, or the tasks might just be more complex and require more experience and expertise, and therefore are rewarded with higher pay.

A doctorate degree has a wage increase of around \$15,000, but not entering the workforce three years earlier sets you back nearly \$270,000 in lost salary, plus school tuition.

How many more years would you need to work if you got an annual a \$15,000 bonus to pay off that debt?

EDUCATION

SHARE OF RESPONDENTS



Company Age

The salary based on company age is in a pretty narrow band between \$85,500 and \$92,500, with one outlier: companies less than two years old (a small sector with 4% of respondents) pay a median salary of \$43,000—much lower than the global median. When we dig into those respondents, we see a few students and others with very low salaries, causing the average to drop. With only 4% of respondents, these outliers have a large impact.

Young companies (2 to 5 years old) are paying out the best median salary at \$93,000 per year. They represent 14% of the respondents, similar to 6- to 10-year old and 11- to 20-year-old companies. This might be the early startups fighting to hire away good talent from established companies.

47% of our respondents work for companies with more than 20 years' experience. It seems that although everyone understands the value of big data and the need for talented people to analyze it, older companies are more in the position to do something about it.

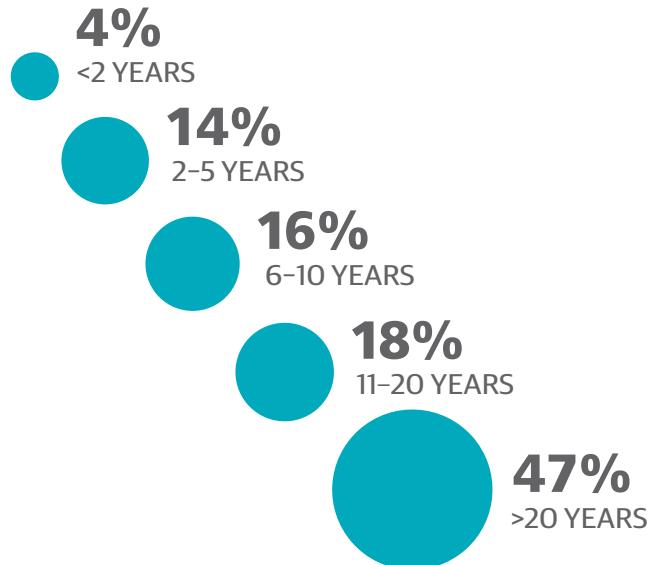
Company Size

Our respondents were split almost perfectly into four equal buckets for the various company sizes, creating nice representative groups for each of the four company size options in the survey. It seems that as the company size grows, so does the salary: from \$73,000 for small 2- to 100-person companies to \$100,000 for companies with 1,001 to 10,000 and 10,000 or more employees.

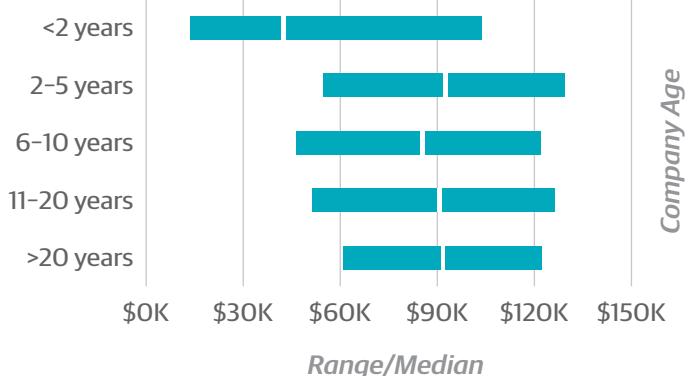
Young companies (2 to 5 years old) are paying out the best median salary at \$93,000 per year. This might be the early startups fighting to hire away good talent from established companies.

COMPANY AGE

SHARE OF RESPONDENTS



SALARY MEDIAN AND IQR (US DOLLARS)

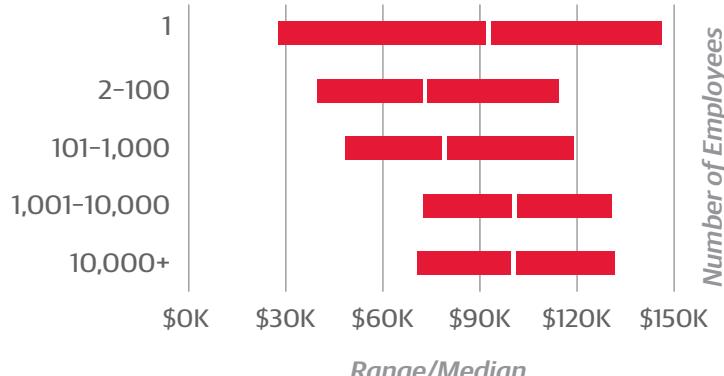


COMPANY SIZE

SHARE OF RESPONDENTS



SALARY MEDIAN AND IQR (US DOLLARS)



2017 DATA SCIENCE SALARY SURVEY

Job Titles

Although this is a Data Science Salary Survey, we do see that people who work in this field have different titles. By far, the most common title is “data scientist/analyst,” at 52% of the respondents. Their median salary was \$87,000.

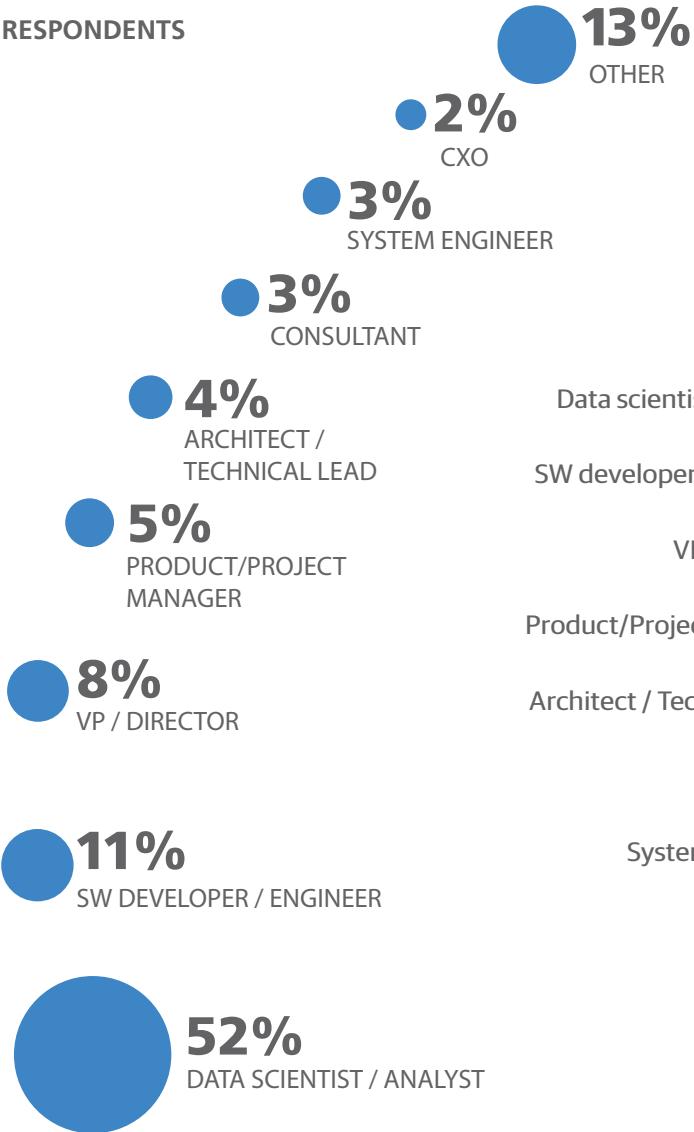
The next largest group of respondents drops to 11%, with a median salary of only \$80,000. These are the folks who consider themselves software developers or engineers.

They might have slipped into the role of data analytics or support the data team.

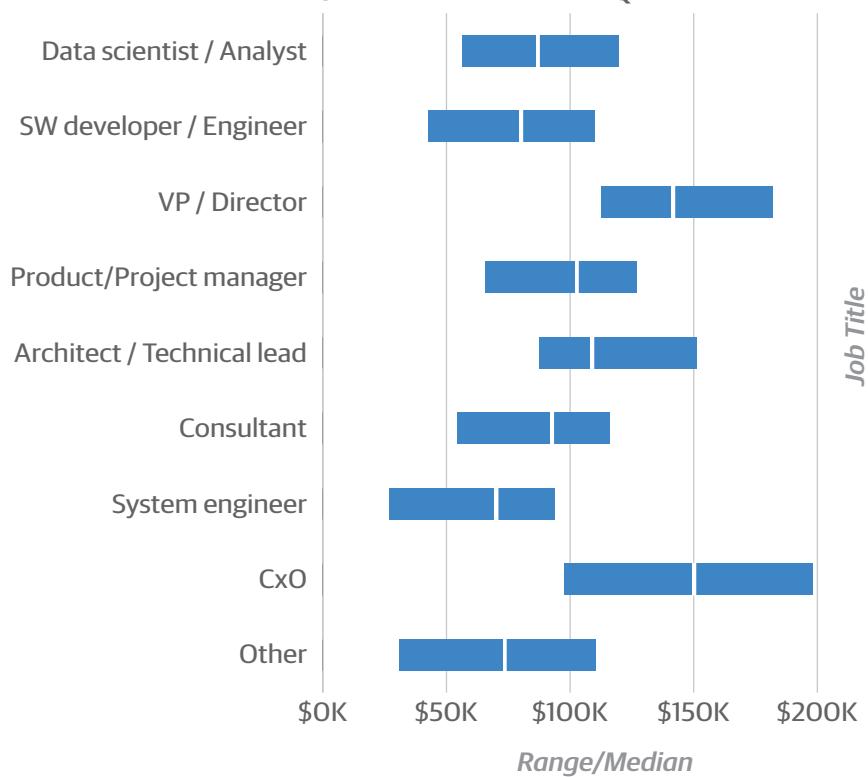
The third most common title was VP/director, at 7.5% of our respondents. These folks garnered the highest median salaries at \$142,000, well above data scientists/analysts and above the global average. The only title to do better was CxO, with a median salary of \$150,000, but they represent only 1.7% of our respondents.

JOB TITLE

SHARE OF RESPONDENTS



SALARY MEDIAN AND IQR*



Time Spent in Meetings

When we look at the time respondents are spending in meetings, we see that as the meetings increase, so does salary. But simply booking yourself into every optional meeting won't necessarily increase your salary. We've found that meetings tend to be a good proxy for responsibility. The more responsibility you have, the more meetings you attend, but you are also rewarded for that responsibility with a higher salary. It isn't the meetings that cause the higher salaries; it's the responsibility that causes more meetings.

2% of our respondents did not attend any meetings. Their salaries are only \$33,000, well below everyone else's. These can be considered outliers, given that most team members will meet, even for short intervals, if they are practicing Agile and have 10- to 15-minute daily stand-ups.

For those who attend one to three hours of meetings per week, the median salary is around \$68,000, representing 23% of our respondents.

The majority (43%) spend between 4 and 8 hours each week in meetings. That's between 45 minutes to 1.5 hours per day in meetings. These respondents had a median salary of \$89,000, much closer to our global average.

The 26% of respondents spending 9 to 20 hours per week in meetings were making \$111,000 median salary. The 6% of respondents who spend more than half their work week (20-plus hours) in meetings have a median salary of \$130,000.

Time Spent Coding

When we look at how much time our respondents spend on coding, we see that their median salaries are not following a regular pattern. There are probably other responsibilities that are affecting the salaries. For instance, if you don't spend any time coding, you have a salary of \$80,000, the same as someone who codes between four to eight hours per week.

More than 60% of our respondents are spending nine or more hours coding each week.

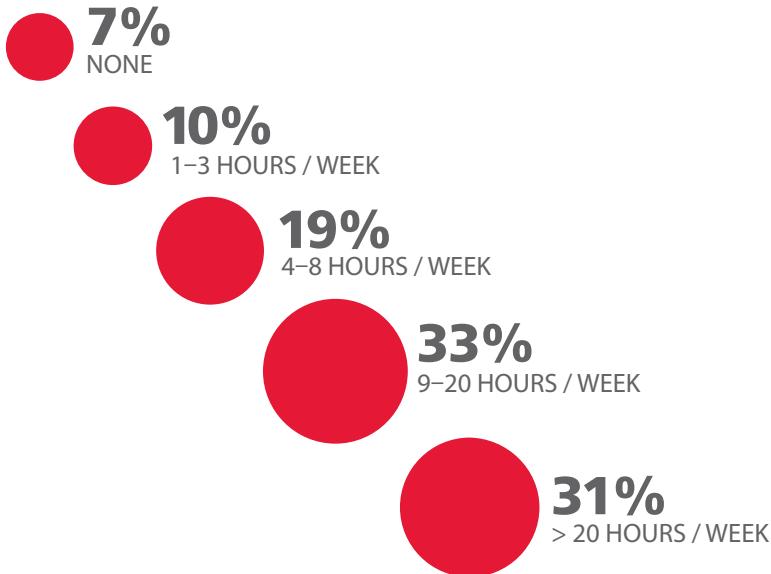
There are about 10% of our respondents who code between one to three hours per week (less than an hour a day). They have a median salary of \$88,000.

The 33% of respondents who code between 9 and 20 hours each week have a median salary of \$97,000, whereas the 31% who code more than 20 hours per week have a median salary of \$91,000.

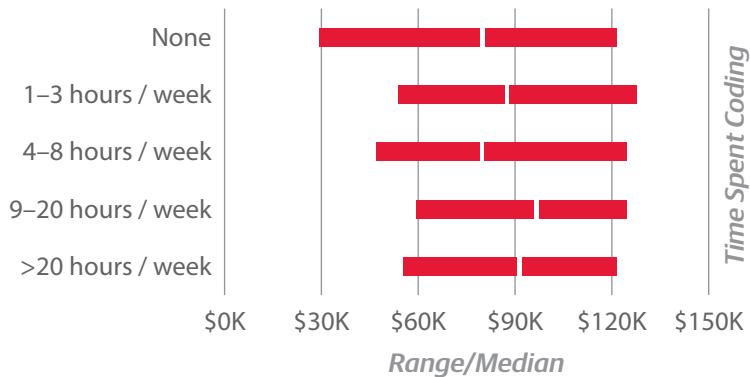
More than 60% of our respondents are spending nine or more hours coding each week. That's a minimum of two hours a day.

TIME SPENT CODING (HOURS PER WEEK)

SHARE OF RESPONDENTS

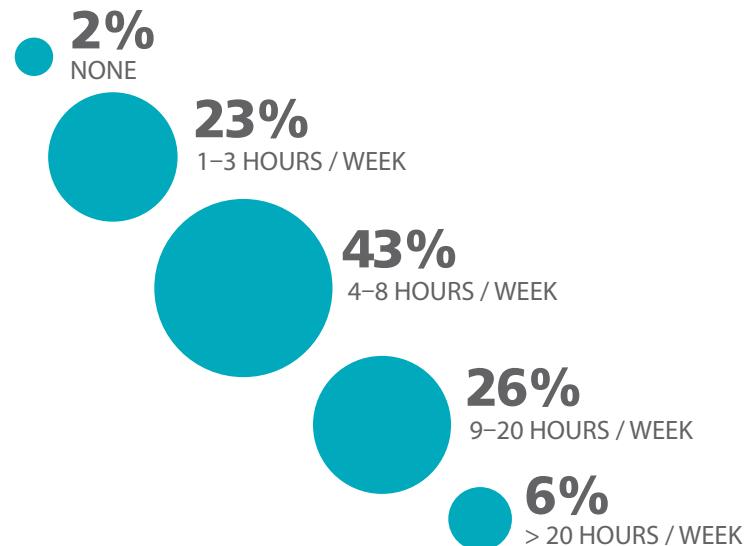


SALARY MEDIAN AND IQR (US DOLLARS)

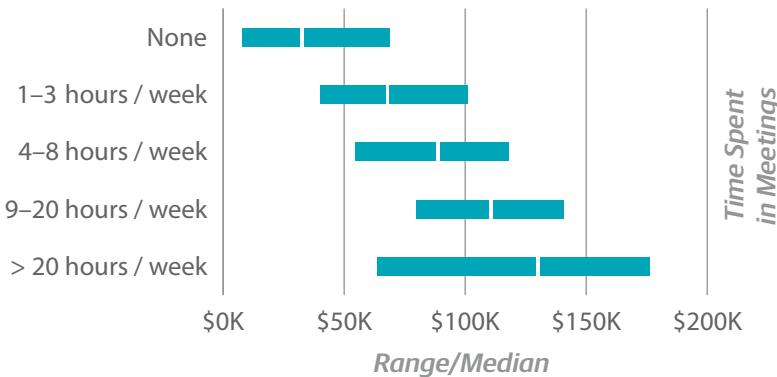


TIME SPENT IN MEETINGS (HOURS PER WEEK)

SHARE OF RESPONDENTS



SALARY MEDIAN AND IQR (US DOLLARS)

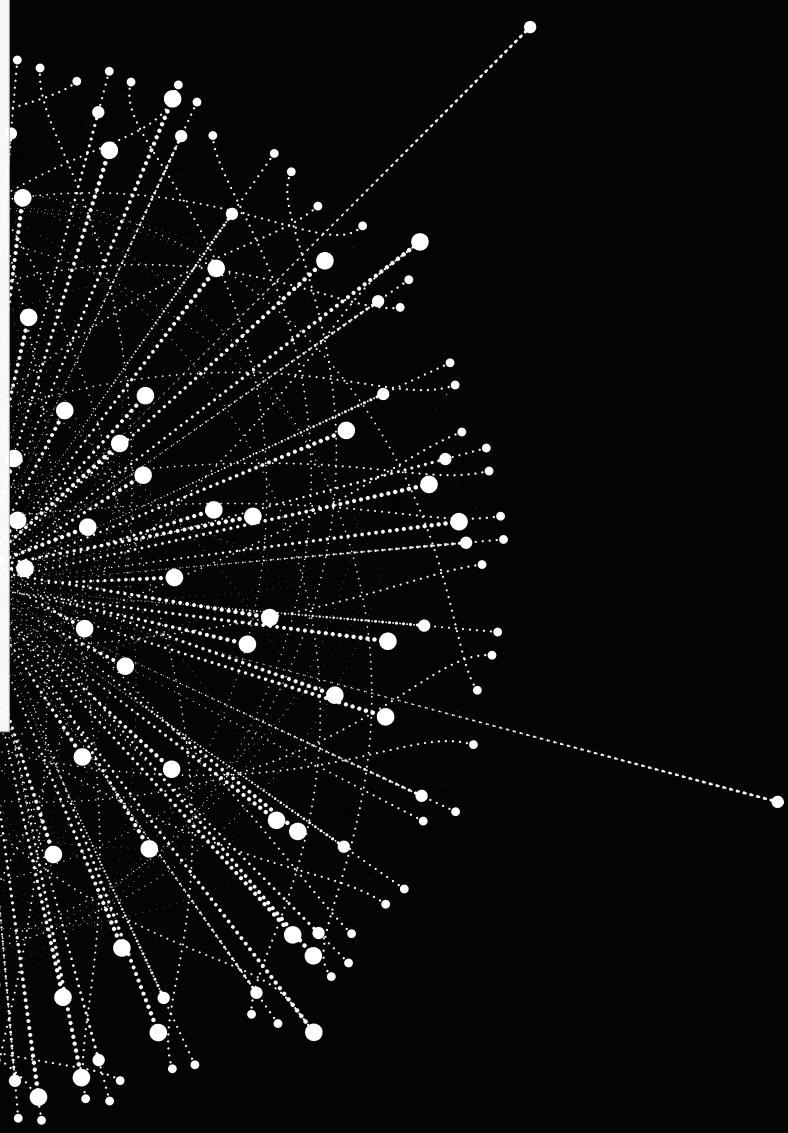


Length of Work Week

When we look at the number of hours our respondents are working each week, 36% are reporting that they work 40 hours. Half of our respondents are working more than 40 hours, and 2% said they work 60-plus hours each week.

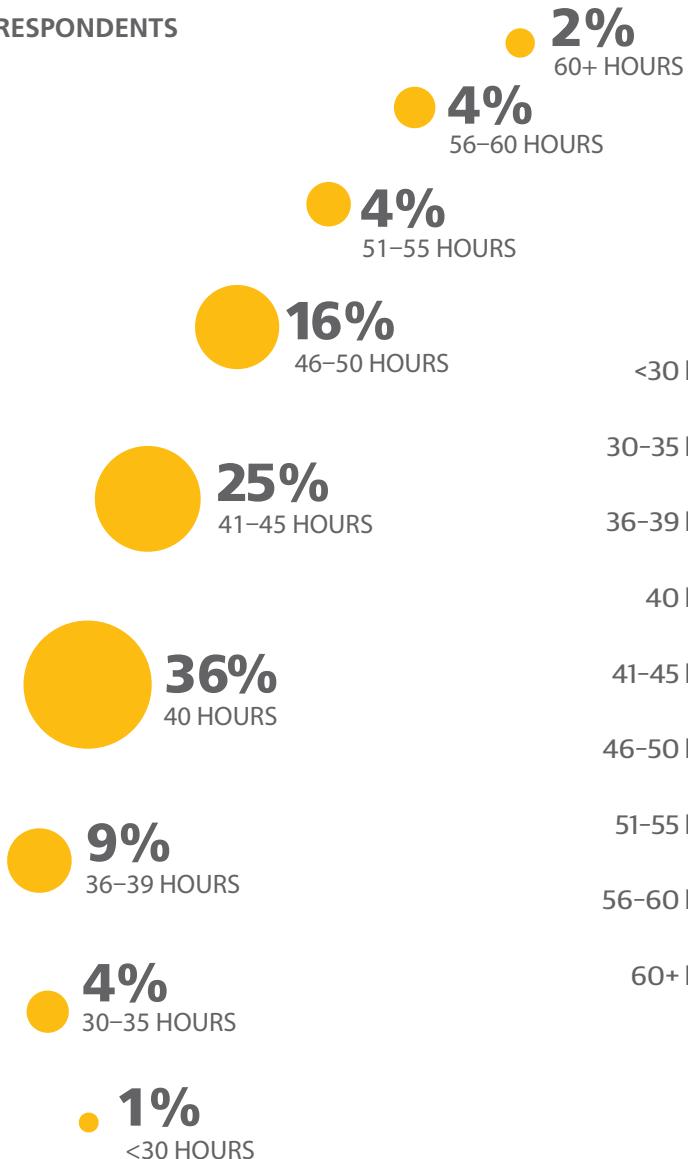
That leaves 14% working fewer than 40 hours. 9% are between 36 and 39 hours per week, which is more common in Europe. 4% of respondents are working between 30 and 35 hours, and less than 1% are working fewer than 30 hours per week.

It seems that there is a lot of work to be done around data and our respondents are putting in extra hours. This could be due to the size of the workload or the lack of team members.

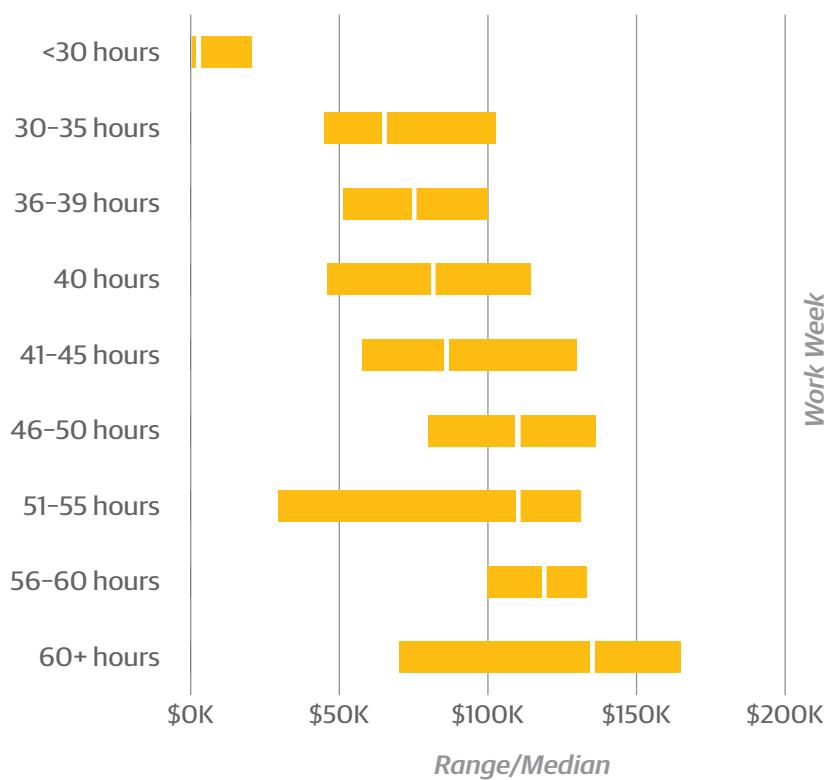


WORK WEEK

SHARE OF RESPONDENTS



SALARY MEDIAN AND IQR*



Ease of Finding a New Role

One of the ways to test the demand in the market for people with data skills is to ask them how easy it is to find a new role. In other words, how easy do they feel it would be to move horizontally in the organization or to another role?

We ask our respondents on a scale of one to five—one being difficult and five being easy—what was the ease of finding a new role.

Only 9% of respondents thought it would be difficult, rating it a one or two. 26% of respondents rated it a three. The median salaries for those rating the ease of finding a new role one, two, or three have median salaries of \$78,000, \$79,000, and \$78,000, respectively.

38% rated the ease of finding a new role a four, and 27% rated it a five. That's 65% of respondents not worrying about finding a new role. We can also see that those who are very confident in their job abilities also have much higher salaries: \$95,000 and \$103,000 median salaries, respectively.

65% of respondents are not worried about finding a new role.

Self-Assessed Bargaining Skills

One of the questions asked of our respondents was to rate their perceived bargaining skills on a scale from one to five—one being the lowest and five the highest. Although this is a self-assessment of how good you think you are at bargaining, it does actually show an increase in salary. Maybe those who report themselves as excellent bargainers are actually good, or maybe it doesn't matter if you're good or not, just that you think you are.

At the low end, respondents did not think they were good bargainers and the lowest median salaries. The 5% who rated themselves with a one have a median salary of only \$70,000. 19% of respondents rated themselves with

a two, which increased their median salary to \$79,000, still below the global average.

The third option was the largest bucket of respondents, at 35%. Their median salary was \$86,000.

For the 40% who think they are good bargainers, they were certainly rewarded for their beliefs. Those who rated themselves a four had a median salary of \$98,000, and those with a five had a median salary of \$115,000; that's a \$17,000 higher salary between a good bargainer and excellent one.

EASE OF FINDING A NEW ROLE ON A SCALE FROM 1-5

SHARE OF RESPONDENTS

Very Difficult - 1 ● 3%

2 ● 6%

3

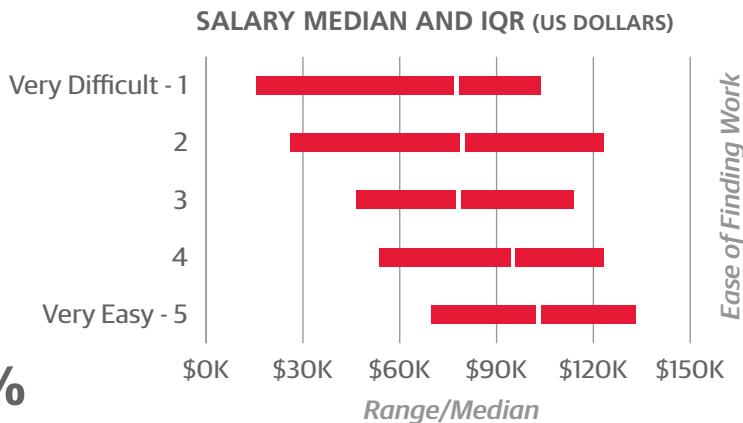
26%

4

38%

Very Easy - 5

27%



SELF-ASSESSED BARGAINING SKILLS ON A SCALE FROM 1-5

SHARE OF RESPONDENTS

Poor - 1 ● 5%

2 ● 19%

3

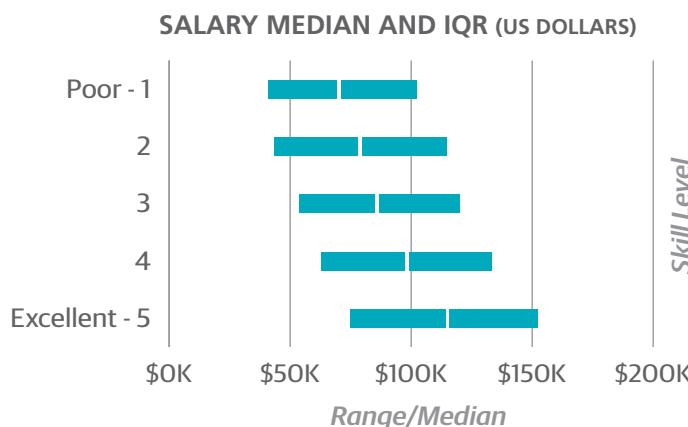
35%

4

31%

Excellent - 5

● 10%



Advance Your Career

Respondents were then asked what steps they were likely to take to advance their careers. The list included these goals:

- Learn new technologies/skills
- Work on more interesting/important projects
- Move into leadership roles, switch companies
- Start your own company
- Other

Only 3% of respondents selected other.

When we order the median salaries from lowest to highest, those with the lowest, at \$76,000, wanted to

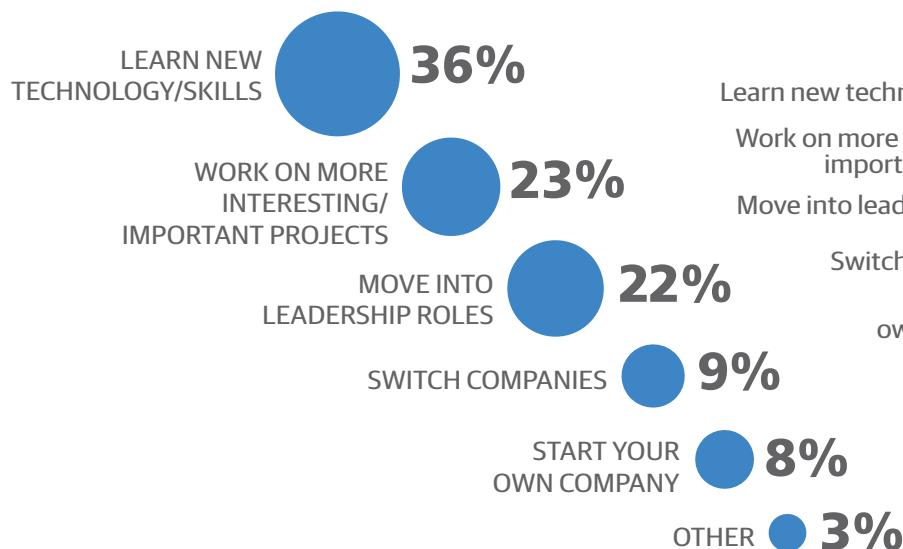
learn new technologies and skills. Respondents with a median salary of \$82,000 wanted to work on more interesting and important projects. Respondents in the \$86,000 bucket wanted to switch companies; at \$101,000 they were thinking more about starting their own company; and at \$114,000 they were moving into leadership roles.

Only 9% see themselves switching companies, and 8% see themselves starting a new company.

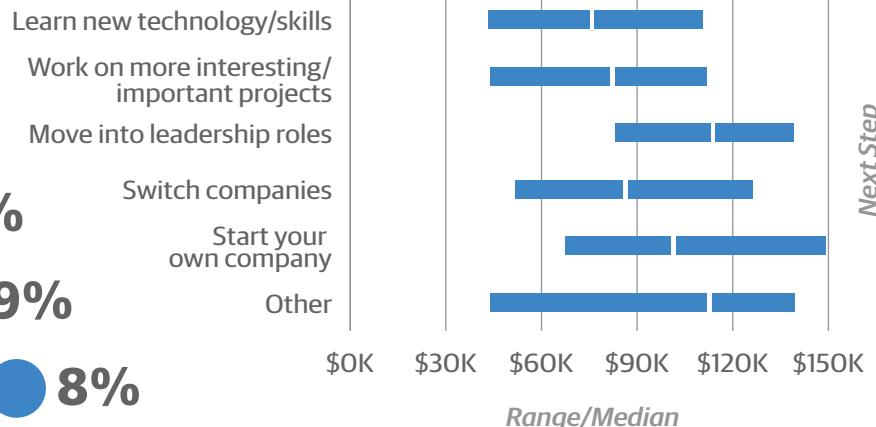
Even though the ease of finding a new role is high, it seems that many are happy with where they are. In fact, 21% of the respondents saw their next move as leadership roles.

WHICH OF THE FOLLOWING MOST ACCURATELY DESCRIBES THE NEXT STEP YOU WOULD TAKE TO ADVANCE YOUR CAREER?

SHARE OF RESPONDENTS

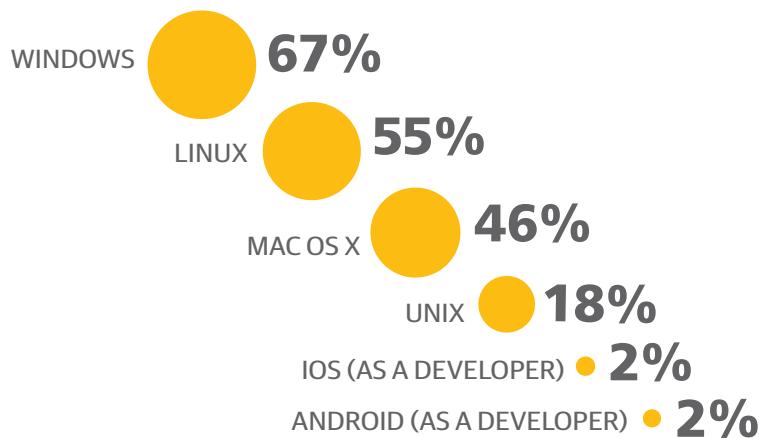


SALARY MEDIAN AND IQR (US DOLLARS)

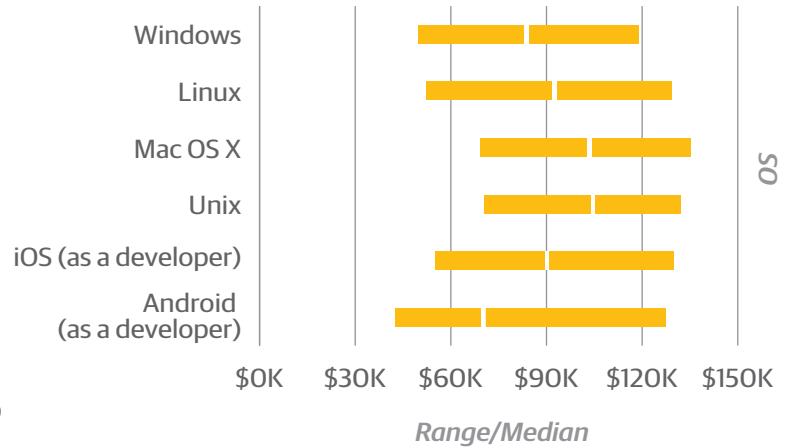


OPERATING SYSTEMS

SHARE OF RESPONDENTS



SALARY MEDIAN AND IQR (US DOLLARS)



Work Setup and Tools

EVERYONE HAS DIFFERENT TASKS, NEEDS, AND ROLES, but it is good to have a peek at what others are using to ensure that you are staying on top of new trends, in addition to justifying the tools you might already use.

Operating Systems

When it comes to processing data, we see a mix of different operating systems in use. 67% of our respondents are using Windows at some point in their work. 55% are using Linux, whereas only 18% use Unix. MacOS has around 46% use by our respondents.

When it comes to mobile operating systems, only 2% are using iOS, and 2% are using Android for development.

Programming Languages

When asked about programming languages, SQL was on top with 64% of our respondents saying they are using it. 63% are using Python, and 54% use R.

Then we begin to get into the long tail of other languages. Bash has a strong following at 33%, Javascript at 20%, Java at 18%, and Scala at 13%.

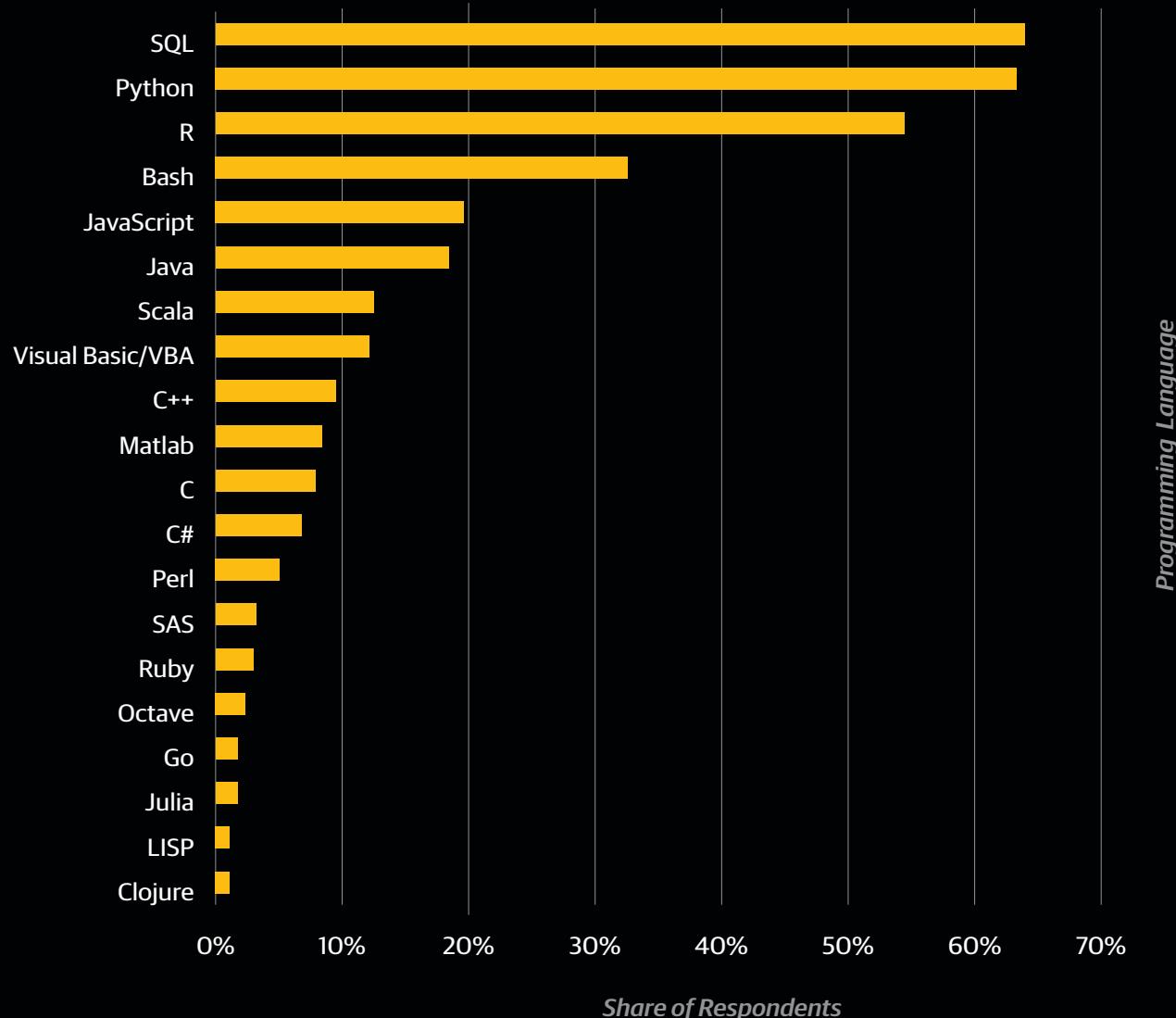
C++, C, and C# are used by 9%, 8%, and 7%, respectively.

Some programming languages certainly equate to higher salaries than others. For instance, Visual Basic/VBA is used by around 13% of our respondents, but the median salary is \$69,000, followed by C# at \$78,000. Perl is the language with the highest median salary at \$109,000, but it was used by only 6% of our respondents.

When we look back at the responses from 2016, we can see which programming languages are gaining in adoptions and which are declining. SQL has dropped from 75% in 2016 to only 64% in 2017. Maybe more data scientists are using GUI tools or working with other parts of the workflow than data retrieval? The other big surprise is that Python jumped from 58% in 2016 to 63% this year. Bash saw a big jump from only 26% of people using it in 2016 to 33% in 2017.

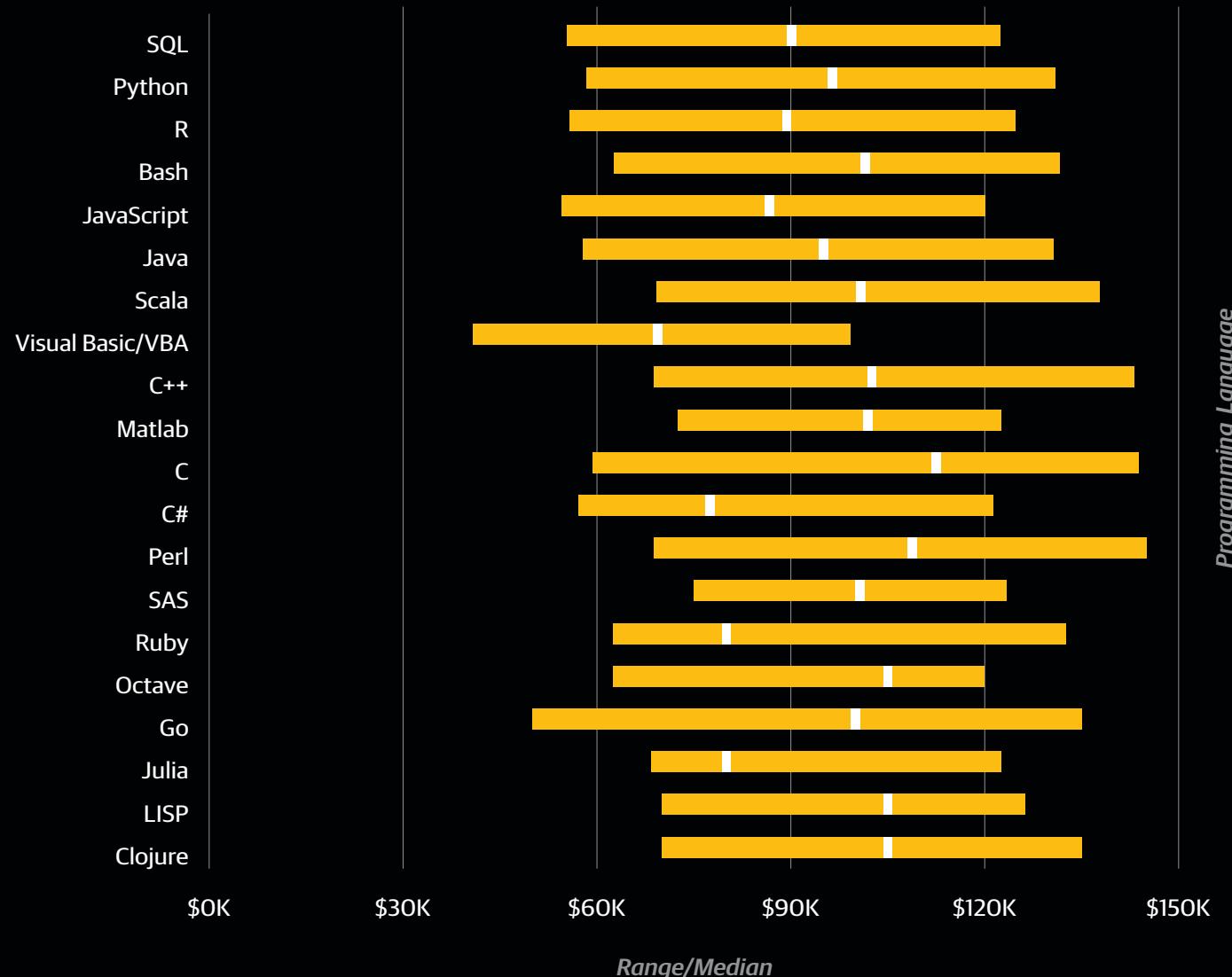
PROGRAMMING LANGUAGES

SHARE OF RESPONDENTS



PROGRAMMING LANGUAGES

SALARY MEDIAN AND IQR*



Relational Databases

A big part of any data management is databases, so it is interesting to see what is popular and being used.

The top relational database—used by 37% of our respondents—is MySQL. That is followed by Microsoft’s SQL Server, used by 30% of our respondents. PostgreSQL appeared quickly on their heels at 28%. Oracle takes the fourth spot, with 20%.

Then we get into the long tail of other databases, from SQLite at 12% to EMC/Greenplum with only 1% at the bottom of the table.

The top five spots all have a median salary between \$83,000 and \$96,000. It seems that knowing the most popular databases isn’t a great differentiator when it comes to salary.

Hadoop

Hadoop is the industry-standard way to do large map-reduce queries on standard hardware, and it has moved to cloud services, as well. 18% of respondents reported using Apache Hadoop (on their own infrastructure).

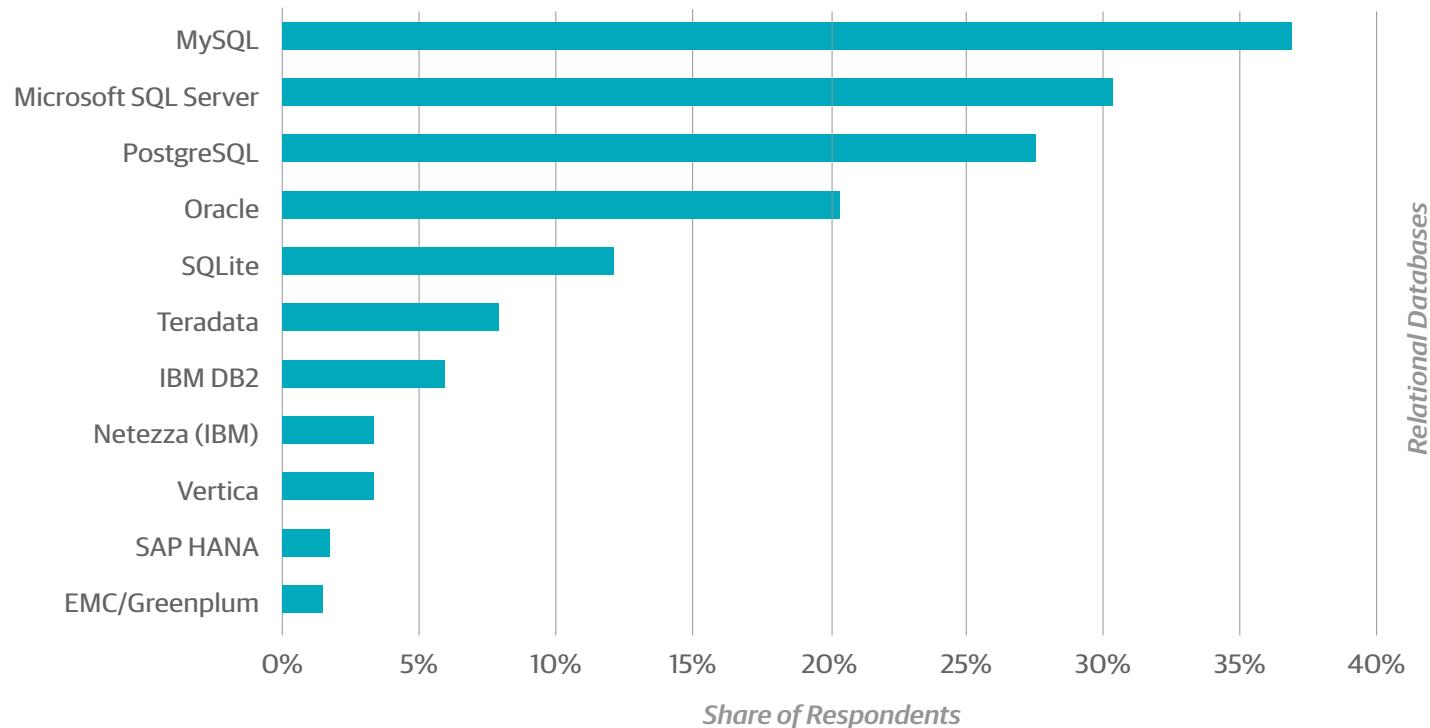
The top relational database—
used by 37% of our
respondents—is MySQL.

The most popular cloud-based Hadoop solution was Cloudera with 12% of our respondents having used it. Following quickly behind was Amazon’s Elastic MapReduce (ERM), with 10%. Hortonworks was the next most popular with 8%, and then MapR with 3%, IBM with 2%, and Oracle with 1%.

Even though some of these solutions might have small responses, we need to take into consideration the number of database instances by that vendor in general. Also, some of these services are cloud-based, whereas others are dedicated datacenters. That factor will also affect their popularity.

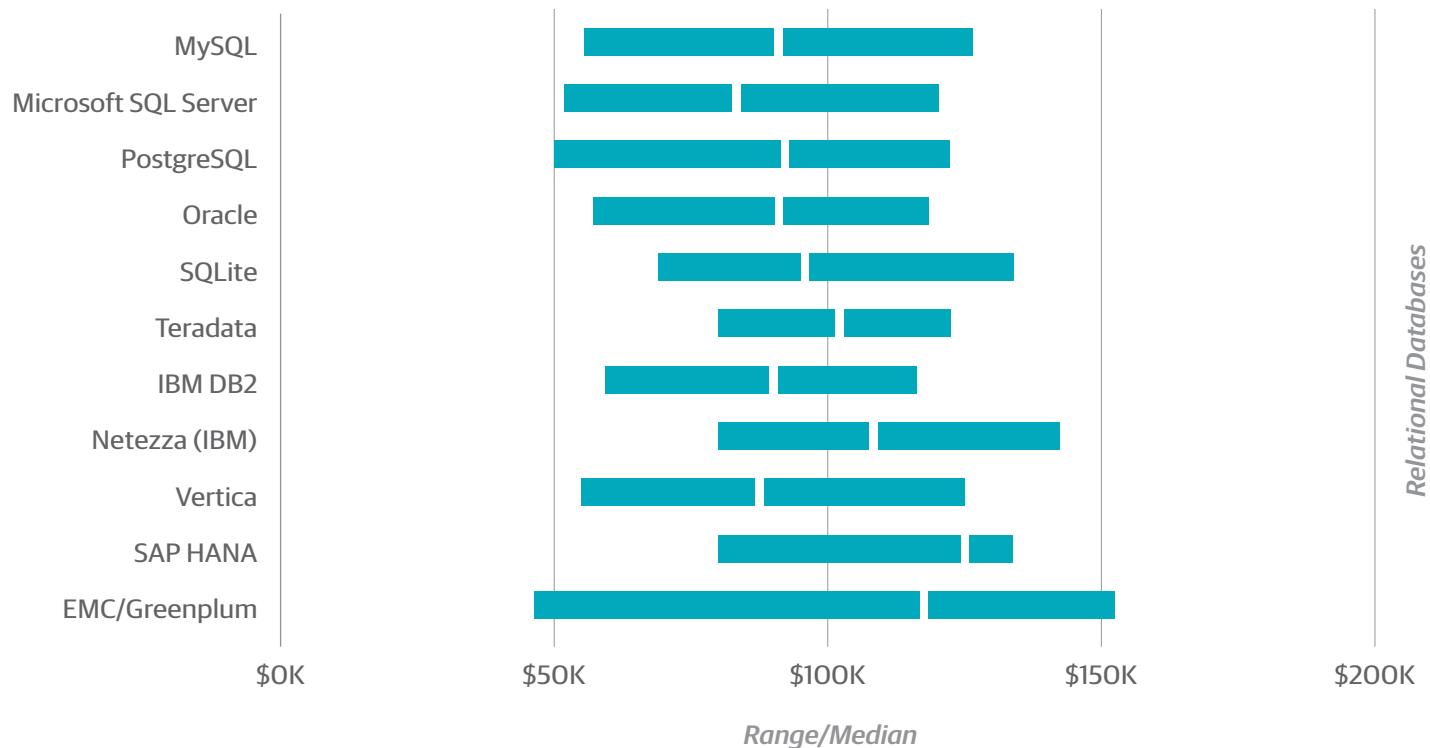
RELATIONAL DATABASES

SHARE OF RESPONDENTS



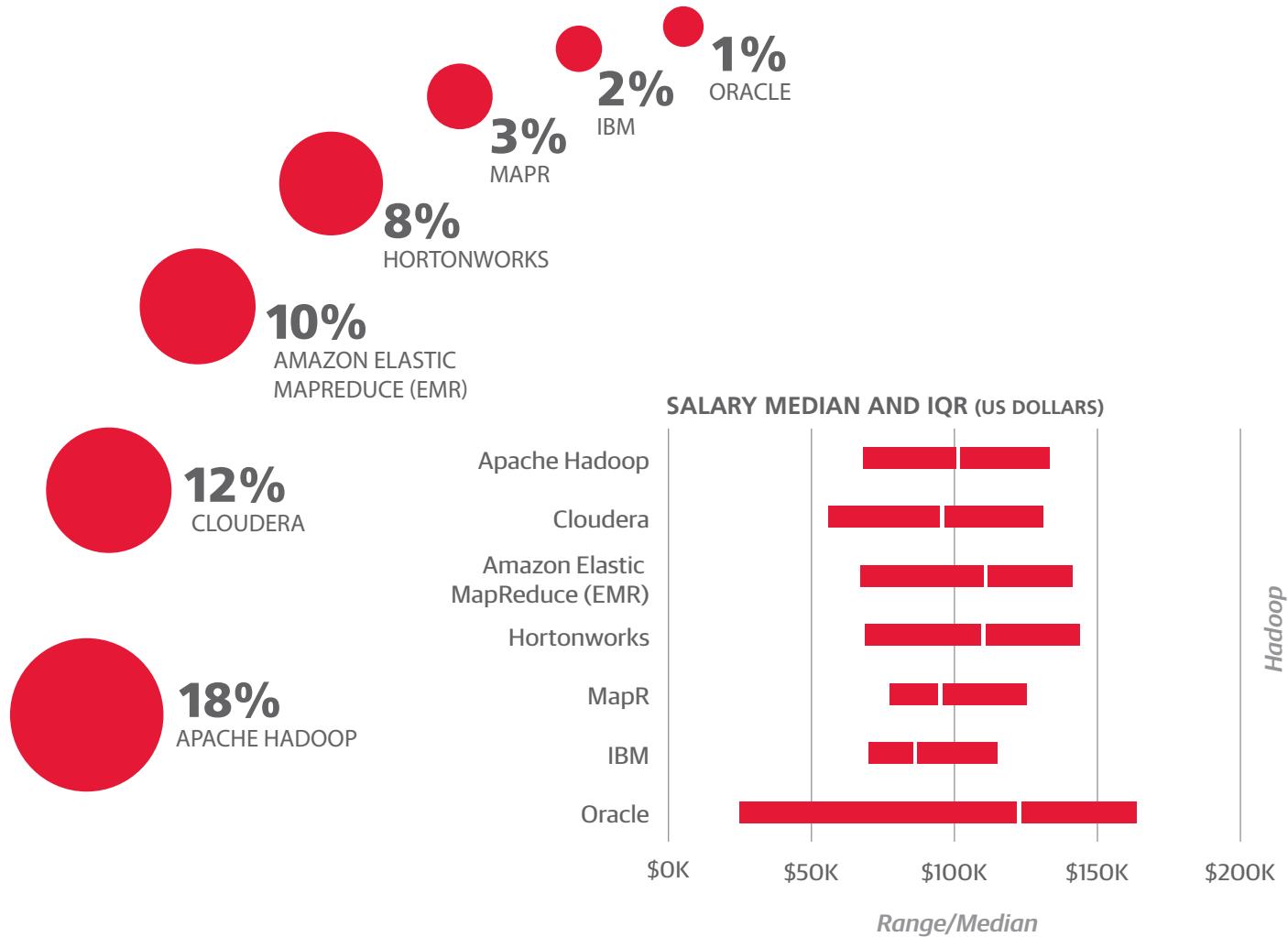
RELATIONAL DATABASES

SALARY MEDIAN AND IQR*



HADOOP

SHARE OF RESPONDENTS



Search

Search and retrieval are also important parts of data collection and storage, and it seems like only a small amount of our respondents are using these tools.

The most popular, Elasticsearch, is used by only 15% of our respondents, followed by Solr at 6%, and Lucene at 4%.

That leaves about 75% of our respondents not using any search tools in their daily workflow.

Big Data Platforms

When the respondents were asked about various platforms used within their company, we saw a wide range of languages, architectures, and vendors.

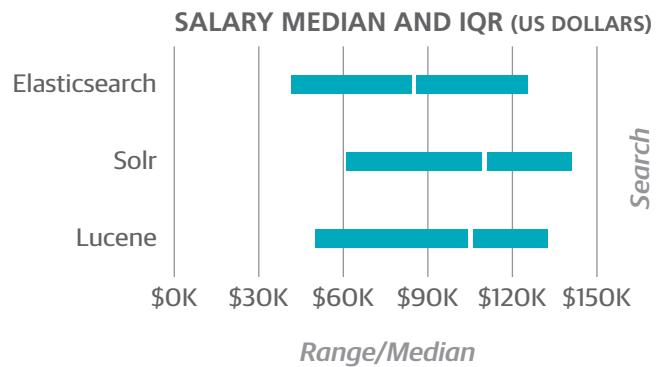
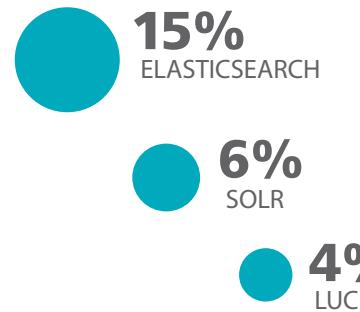
Spark is the current favorite, with 27% of our respondents using it in some way. That is quickly followed by 18% using Hive, 13% using MongoDB, and 12% using Amazon Redshift.

After that, there is a long tail of other technologies, all below 10%.

Business Intelligence and Reporting

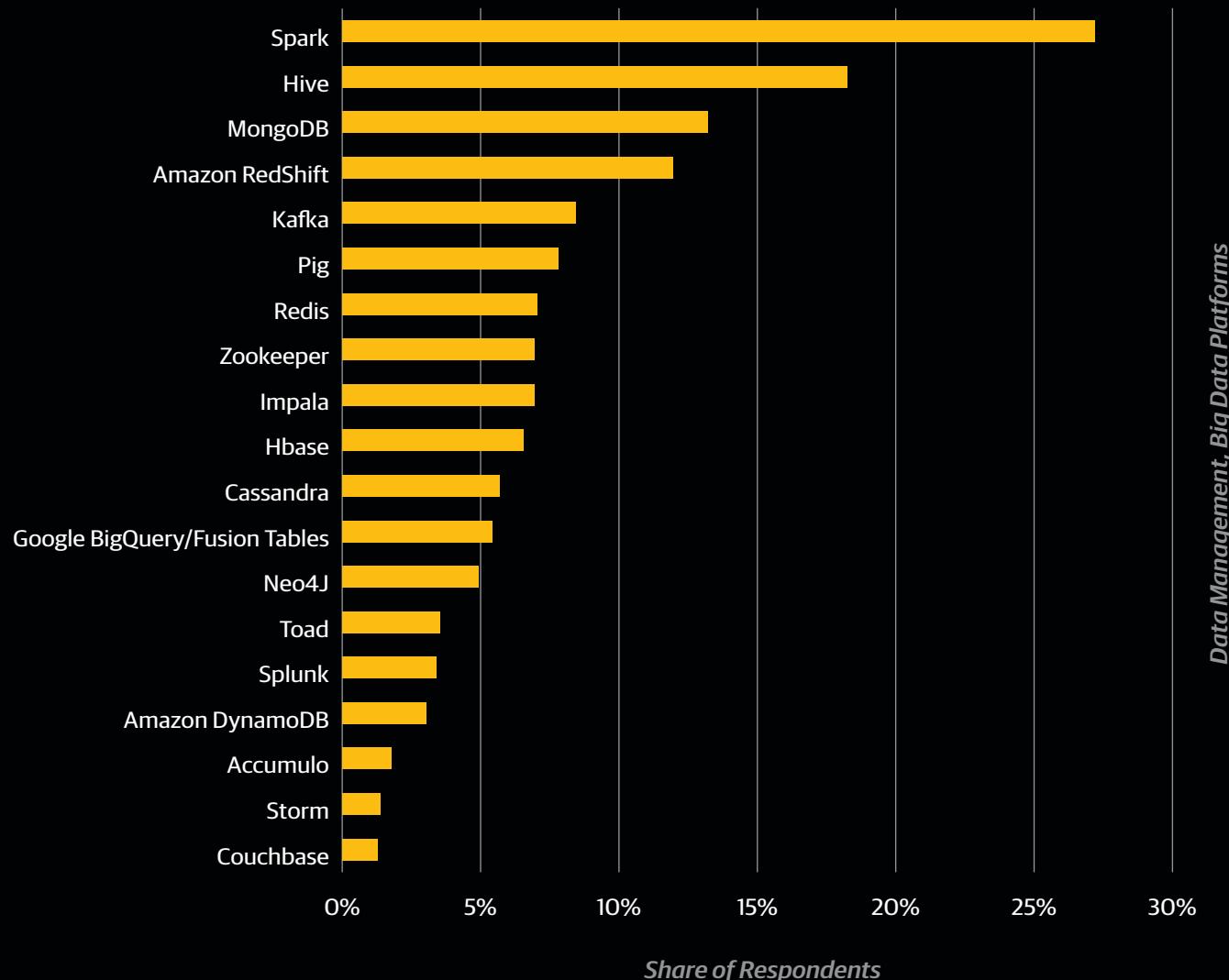
When asked about spreadsheets, business intelligence (BI) tools, and reporting, there was no contest: two-thirds of the respondents said Microsoft Excel. Excel is a completely ubiquitous tool for reporting. The next closest competitor is Microsoft Power BI, with less than 10% of the respondents.

SEARCH SHARE OF RESPONDENTS



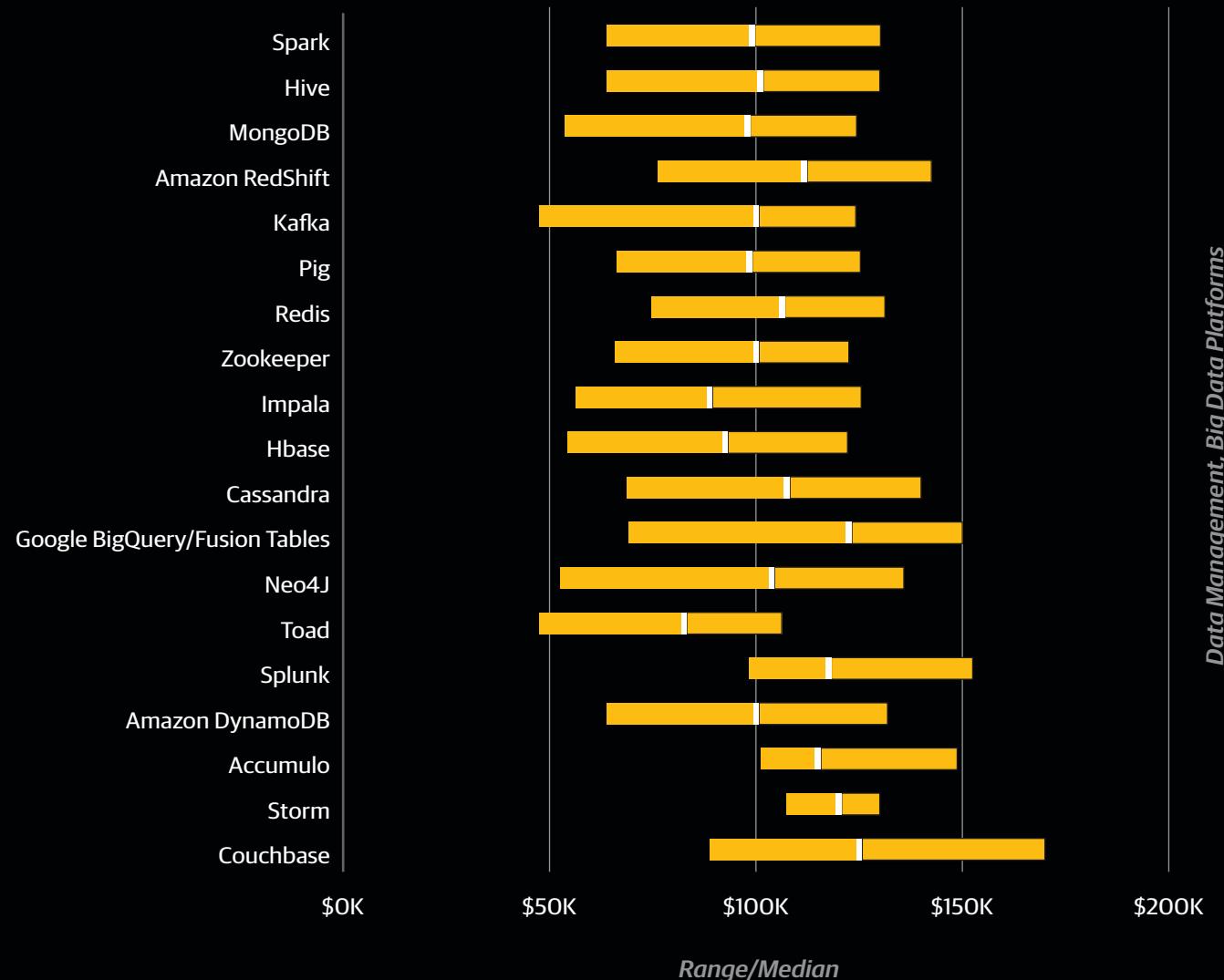
DATA MANAGEMENT, BIG DATA PLATFORM

SHARE OF RESPONDENTS



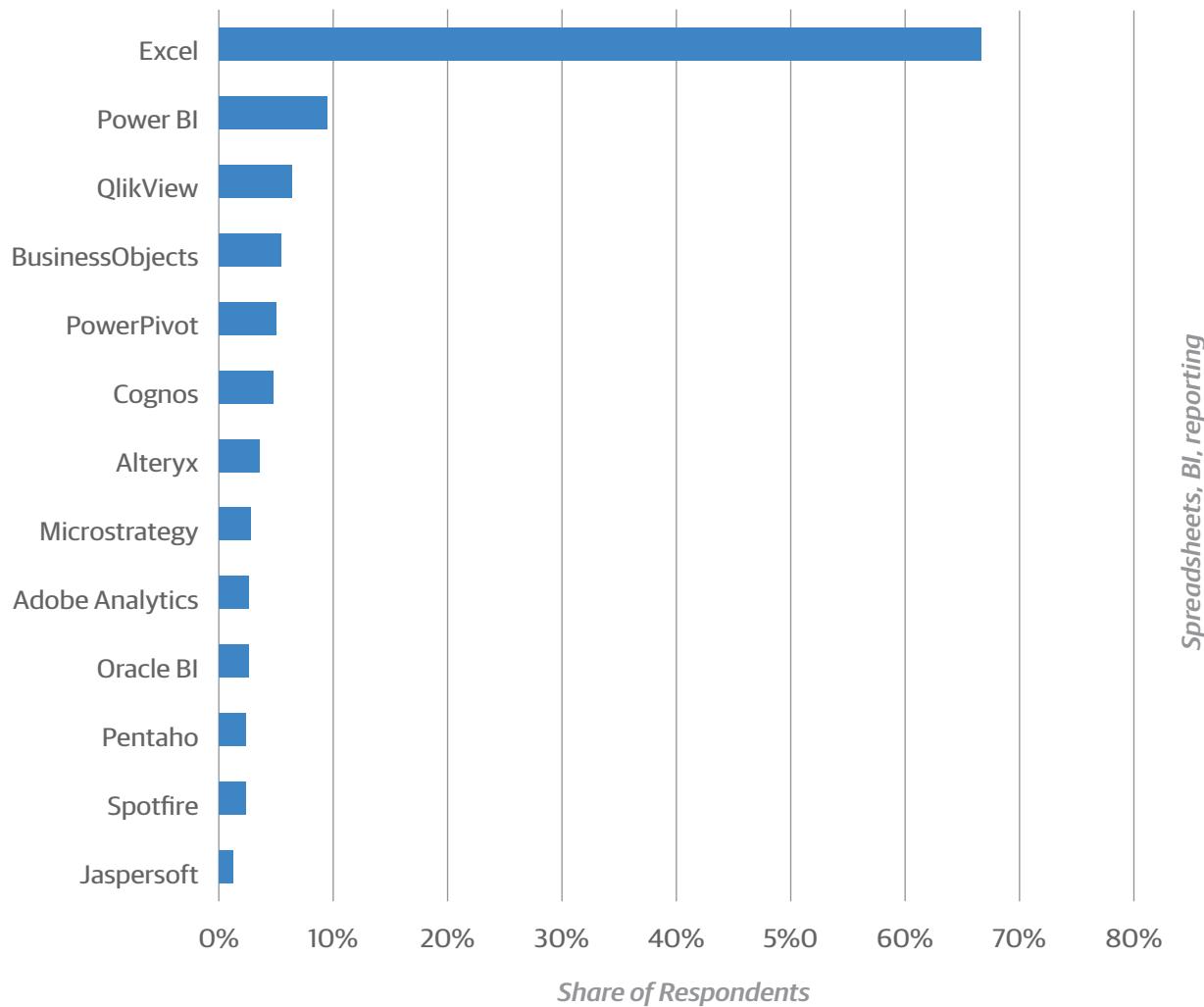
DATA MANAGEMENT, BIG DATA PLATFORM

SALARY MEDIAN AND IQR*



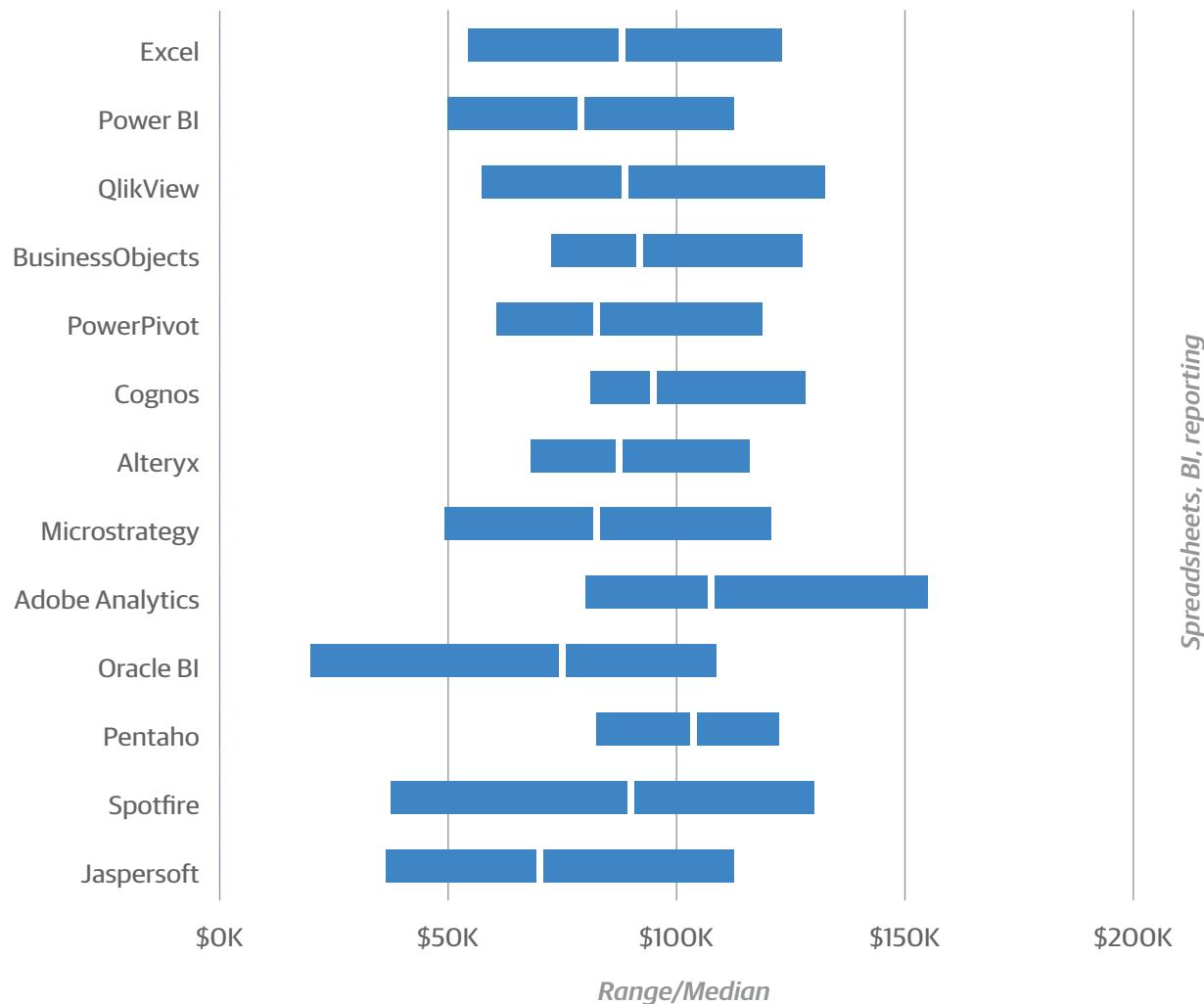
SPREADSHEETS, BI, REPORTING

SHARE OF RESPONDENTS



SPREADSHEETS, BI, REPORTING

SALARY MEDIAN AND IQR*



That's a 57% gap between the most popular and second most popular.

There is a long list of other BI tools, but they trail off in popularity pretty quickly. Some of these might be legacy tools, others might be the exact right tool for the job, so just because only 3% of respondents are using Oracle BI, it might be the perfect tool if you use Oracle DB.

Machine Learning

Machine learning is a very hot topic. With more and more vendors entering into the arena and attempting to make it easier to use, we'll see an explosion in what is considered machine learning as well as a very long tail of potential software packages.

Our respondents seem to have chosen a few popular software solutions, but it is still a diverse choice in the tail. 37% of our respondents use Scikit-learn, and 16% Spark MLlib. Given that 27% of our respondents are using Spark in their big data platform, Spark MLlib makes sense.

H2O, ML as a service, is used by 8%, the Java-based Weka by 7%, and then we drop to 4% and below for the rest of the options.

Viz Tools

Our respondents were asked about which data visualization tools they are using. There is a good mix of different tools, with no single one dominating the group. ggplot which is used in R, Python, and Jupyter Notebooks, is used by 43% of our respondents. 34% have used Matplotlib, 32% Tableau, and 21% Shiny (another R tool).

At 18% is D3, an open source JavaScript library used for visualization. Hosted Google Charts has around 10% usage, and then the percentage drops from there: Bokeh, 7%; Processing, 2%; and Processing.js, 1%.

These tools can serve different purposes. Using something like D3 means that you are focusing on HTML output, whereas ggplot might be more for screens and reports.

Importance of Tasks

We asked our respondents about various tasks and whether they had major, minor, or no involvement in those tasks. When just looking at how they rate themselves in major involvement, we get a good picture of what it is to be considered a data scientist.

67% of our respondents said they have major involvement in “Basic exploratory data analysis.” 61% said they “conduct data analysis to answer research questions.” These are the most popular tasks, and they both deal directly with the datasets. No big surprise there.

The third most popular task was to “communicate findings to business decision-makers.” This is interesting, because beyond just crunching the data, this role is expected to be a communicator: finding the story in the data and expose that to those in charge.

53% have major involvement in “data cleaning”: checking for outliers or missing data, reformatting values, and so on. This role is also probably one of the longest and most tedious tasks, calling to mind the old quote attributed to Abe Lincoln, “Give me four hours to chop down a tree and I’ll spend the first three sharpening my axe.”

The rest of the tasks are all less than 50% major involvement, but that’s not to say they aren’t important; rather, they are

taken on by other members of the team or company. For instance, “create data visualizations” involved only 47% of our respondents. This number could simply be a result of companies hiring a dedicated illustrator or design team, with raw data sent to them for processing.

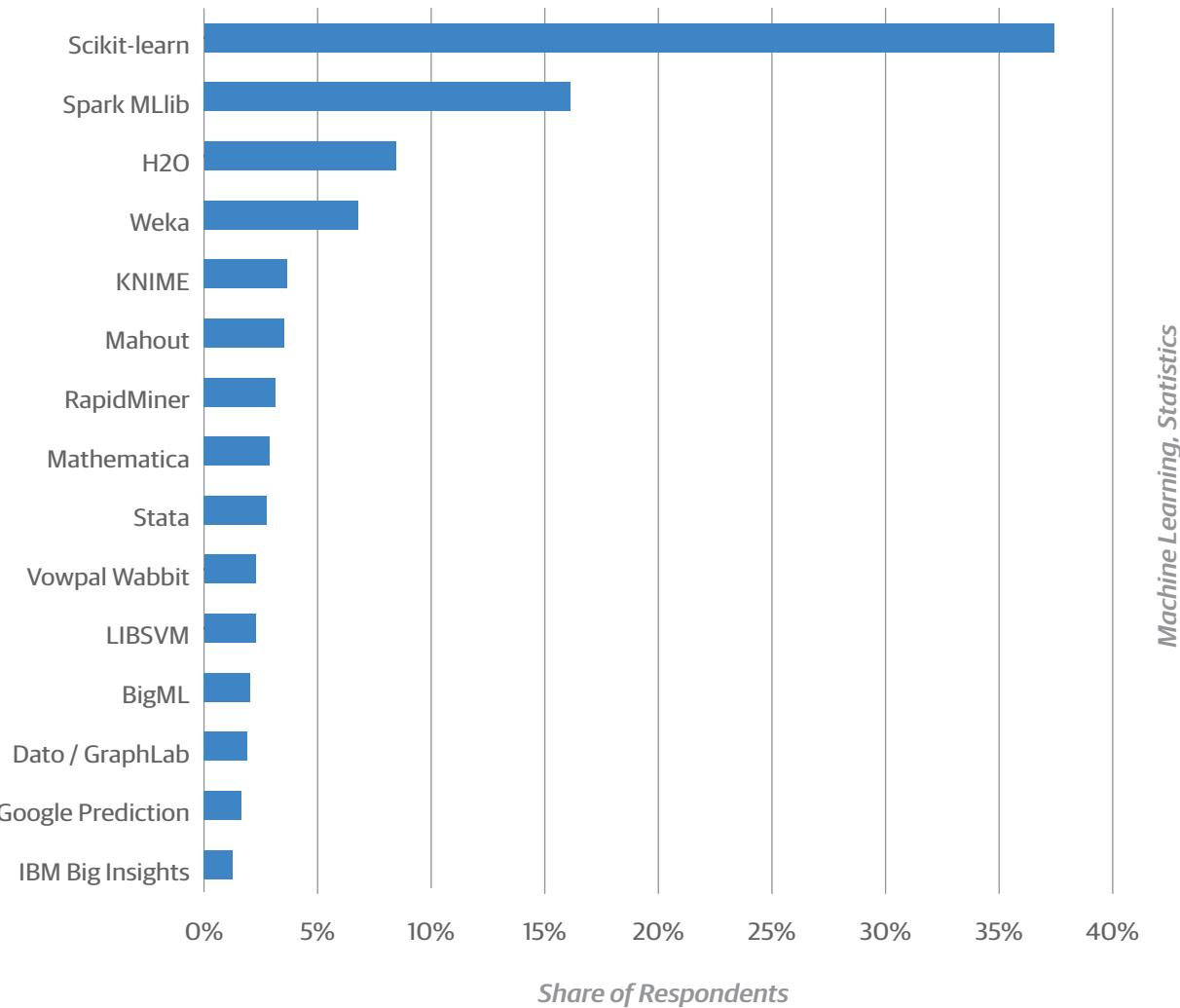
Extract, transform, and load (ETL) is an important part of working with data, but according to this survey, only 30% of our respondents are working on ETL pipelines as one of their major tasks. Maybe this role is shifting to a dedicated person or a different team. It is worth watching this in the future.

The bottom three tasks were to “develop products that depend on real-time data analytics” at 18%; “use dashboards and spreadsheets (made by others) to make decisions” at 15%; and “develop hardware (or work on software projects that require expert knowledge of hardware)” at 4%. Although these tasks might be important for some data scientists, they do not seem to be central to the field.

The third most popular task
was to “communicate findings
to business decision-makers.”

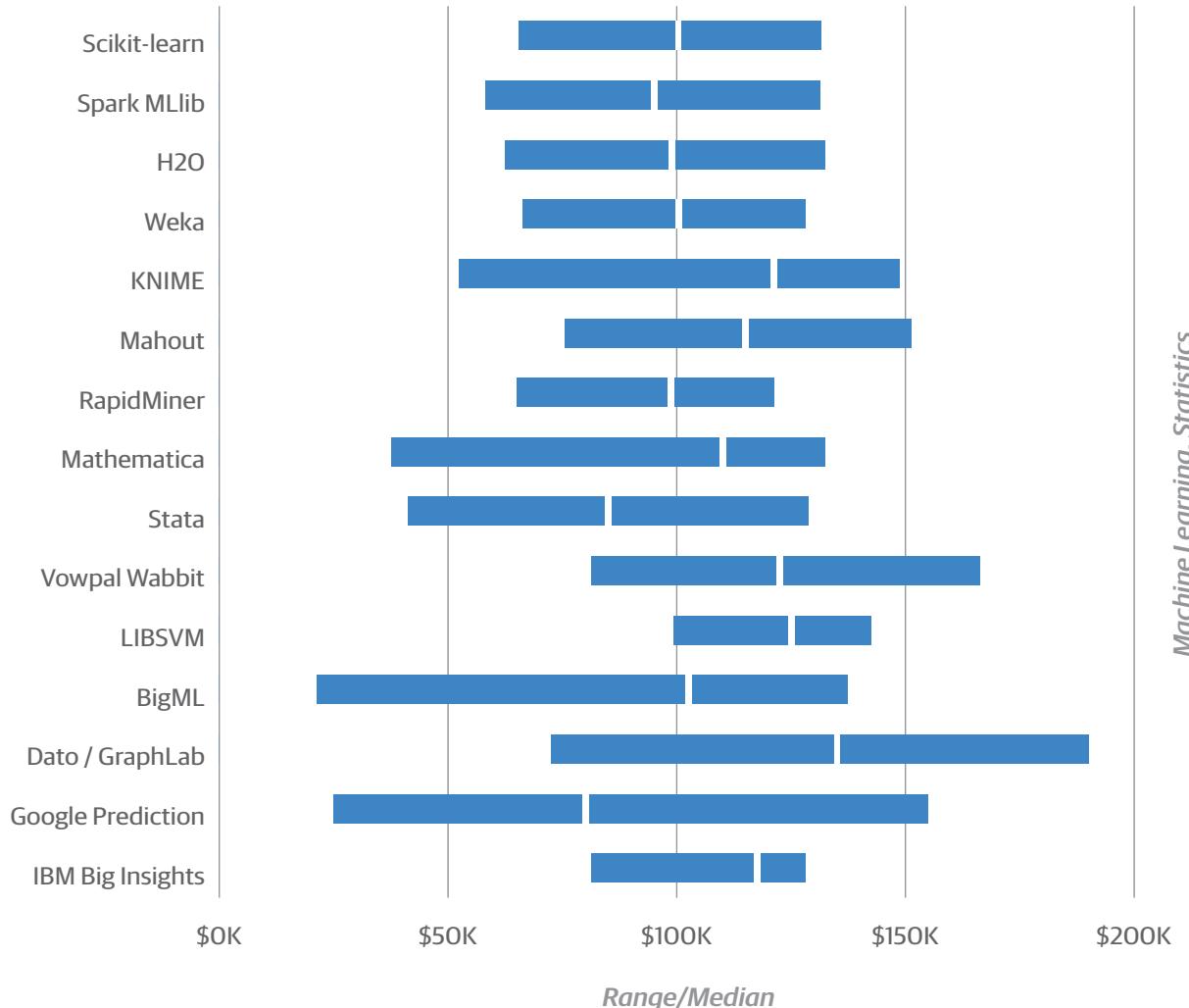
MACHINE LEARNING, STATISTICS

SHARE OF RESPONDENTS



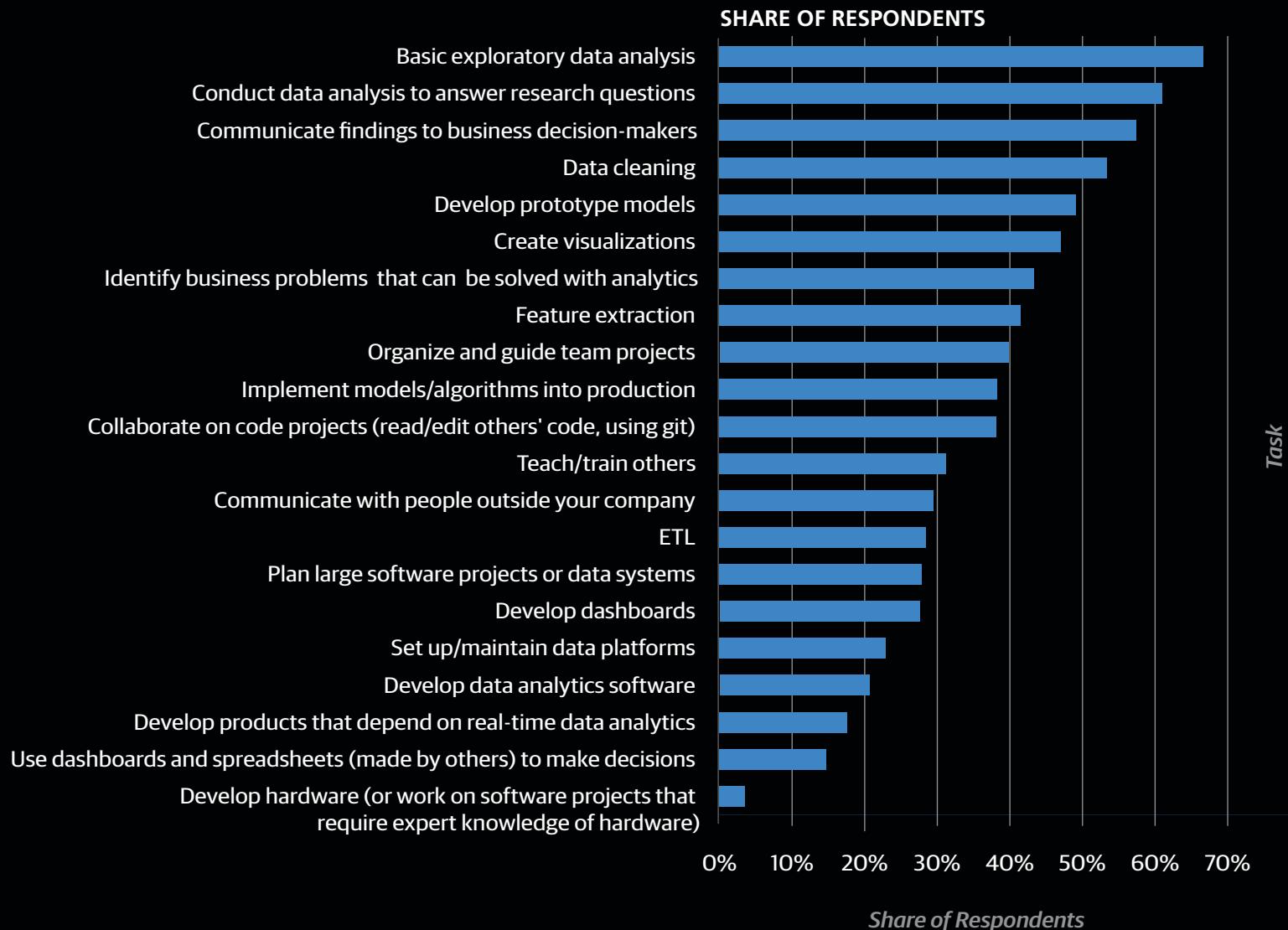
MACHINE LEARNING, STATISTICS

SALARY MEDIAN AND IQR*



TASKS (RESPONDENTS COUNTED IF THEY SAID THEY HAVE "MAJOR INVOLVEMENT" IN THIS TASK)

SHARE OF RESPONDENTS



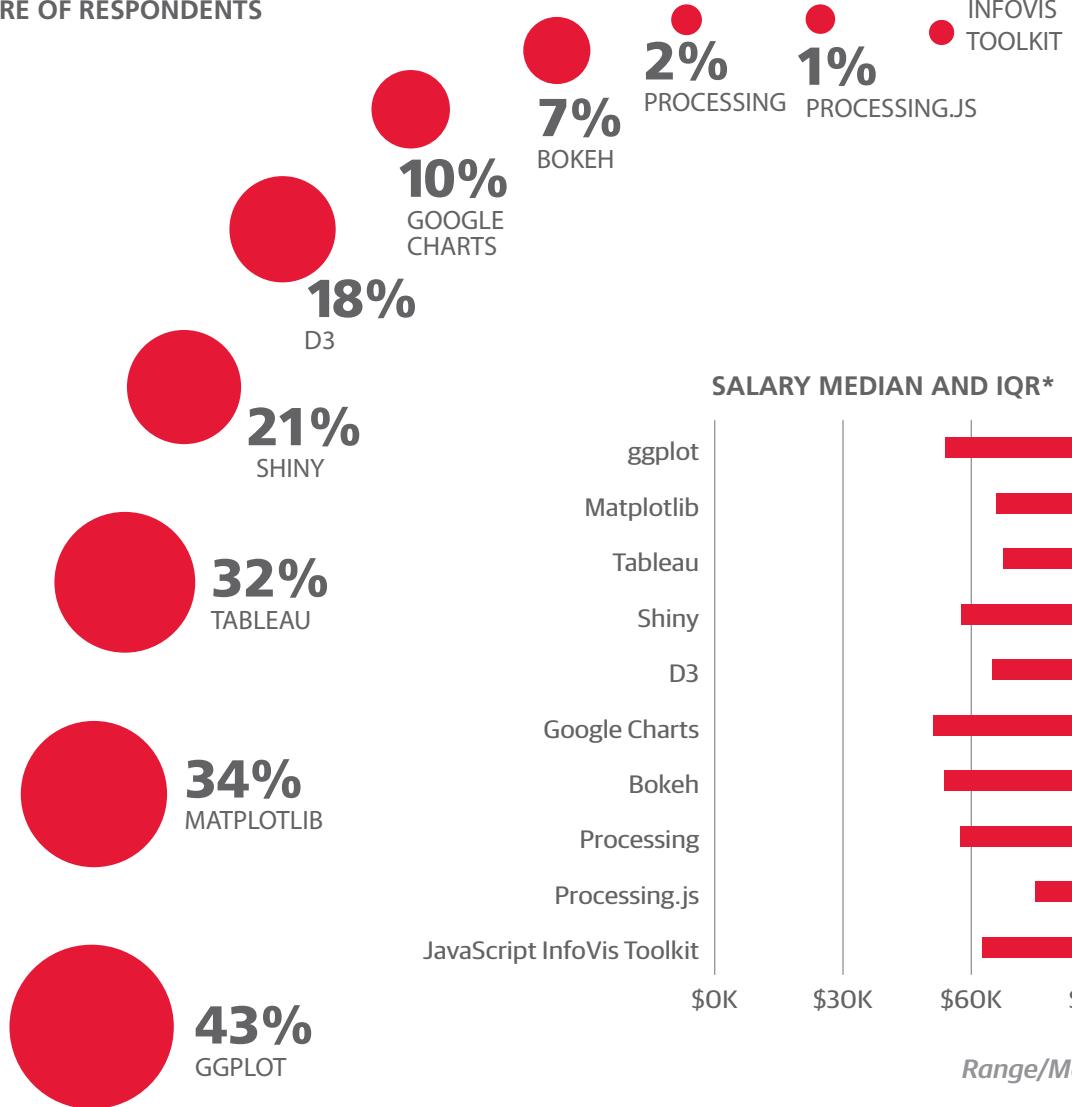
TASKS (RESPONDENTS COUNTED IF THEY SAID THEY HAVE "MAJOR INVOLVEMENT" IN THIS TASK)

SALARY MEDIAN AND IQR*

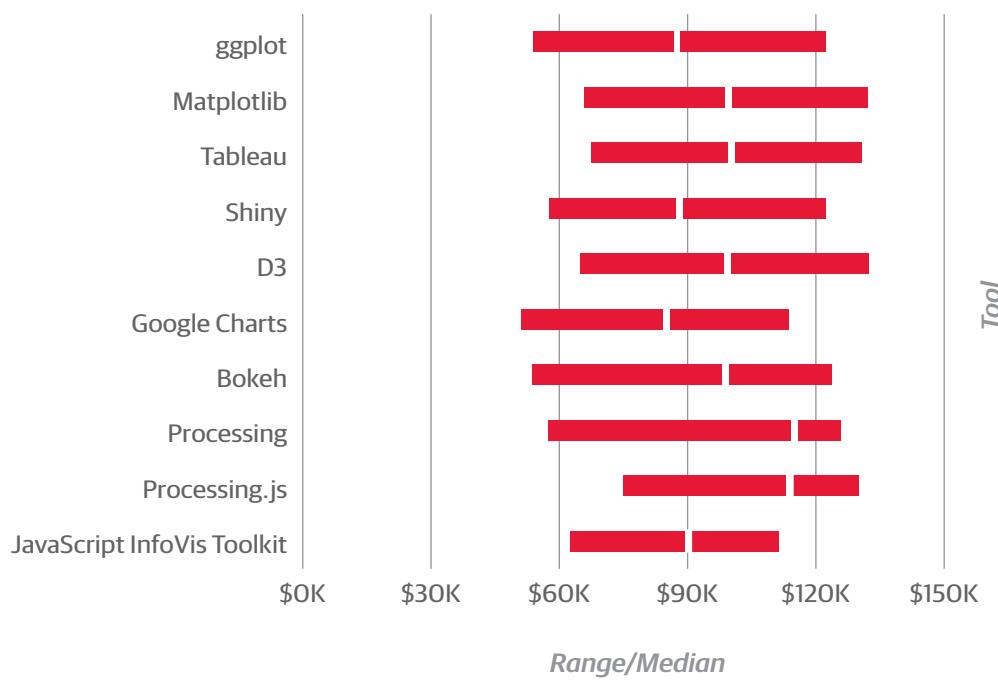


VISUALIZATION TOOLS

SHARE OF RESPONDENTS



SALARY MEDIAN AND IQR*



Conclusion

THE DATA SCIENTIST ROLE CONTINUES TO GROW

globally as more nonsoftware companies understand the need for resources to analyze and report on data using modern tools. Overall, results are similar to what we found in 2016, helping confirm the reliability of the survey data. Most of the salary data shows stable trends, with a few sectors showing increases—what we expect as the years pass.

There were a few surprises, such as the shuffle of programming languages and relative drop in US representation. This change might be due to the types of companies responding to the survey or other trends in the market. Some software releases garner intense attention, creating a rush to try and learn a new programming language in order to make use of new features or libraries.

Low usage rates for tools (e.g., languages and databases) doesn't imply some inherent deficiency. These tools might address niche functionality, legacy systems, or a functional area on the cusp of more widespread adoption.

This report is our best guide to what is happening in the industry surrounding data. Use the report to start conversations with your team and company regarding tools and processes and help map out the elements that make up the data landscape and how that relates to your company's technology infrastructure and business model. Look at what soft and hard skills you should consider in order to stay competitive and relevant in the data ecosystem. With the guidance of producing five years of reports, we look forward to continuing to conduct and share the salary surveys with the data community.

Model

The model has an R-squared of 0.60: this means the model explains approximately 60% of the variation in the sample salaries. Geography is used as the Y-axis intercept of the model. Select the appropriate location and then proceed through the coefficients, adding or subtracting the ones associated with a feature that applies to you. After you sum up the coefficients, you will obtain the model's estimate for your annual total salary in US dollars.

<i>World Region</i>	<i>Experience</i>	<i>Job Title</i>
Canada: \$43,991	+\$2,464 (per year)	Data scientist/analyst: +\$14,227
Western Europe: \$38,113		CxO: +\$42,209
Asia: \$23,542	<i>Gender</i>	VP/Director: +\$45,678
Eastern Europe: \$6,180	Male: +\$8,071	Product/Project manager: +\$10,345
Latin America: \$5,559	<i>Company Size</i>	Architect/Technical lead: +\$24,001
Australia/New Zealand: \$63,067	1: -\$19,402	Marketing: +\$11,092
Africa: \$17,145	2-25: -\$21,442	System engineer: -\$13,872
<i>US Region</i>	26-100: -\$13,581	<i>Industry</i>
Midwest: \$68,087	101-500: -\$9,753	Healthcare/Medical: -\$8,505
California: \$101,834	501-1,000: -\$8,484	Consulting: +\$4,474
Texas: \$73,048	1,001-2,500: +\$13,951	Retail/Ecommerce: +\$12,594
Mid/Atlantic: \$84,487	2,501-10,000: -\$5,708	Government: -\$10,050
Southwest/Mountain: \$73,327	<i>Education</i>	Nonprofit/Trade association: -\$21,545
Pacific/Northwest: \$79,525	PhD: +\$5,376	Logistics: +\$18,171
Northeast: \$87,944		Search/Social networking: -\$11,885

We need *your* data.

To stay up to date on this research, your participation is critical. The survey is now open for the 2018 report, and if you can spare just 10 minutes of your time, we encourage you to take the survey.

oreilly.com/ideas/take-the-data-science-salary-survey

Take the Survey 