Master of Science in Analytics Program
2015-2016 Assessment Report

1 Identifying Information

Name of Program: Analytics
Type of Program: Graduate Program
College of Arts and Sciences Division: Sciences
Submitter: David Uminsky, Program Director
Submitter Email Address: `duminsky@usfca.edu`

Additional feedback concerning this report can be directed to Kirsten Keihl (Program Manager), or Mindi Mysliwiec (Director of Operations, Data Institute).

2 Program Vision and Mission

Program Vision. Our vision is to become a national leader in training the next generation of technically-competent and career-ready professionals who are fully engaged in, and continuously advance, the data science revolution in the Bay Area and beyond.

Program Mission. The mission of our program is to produce graduates who possess a theoretical and practical understanding of many classical and modern statistical modeling and machine learning techniques; who use contemporary programming languages and technologies to scrape, clean, organize, query, summarize, visualize, and model large volumes and varieties of data; and who use their knowledge and skills to successfully solve real-world data-driven business problems and to communicate those solutions effectively.

Notes. The MSAN program's vision and mission statements were ratified by its faculty during a January 2016 vote over email.

3 Program Goals

Faculty members affiliated with the MSAN program have identified a total of ten long-term goals for the program. These goals include, but are are broader than, student satisfaction or student learning outcomes. Some of these goals are tangible; others are not.

1. **Student satisfaction:** In exit surveys, a super-majority of graduates will indicate a high degree of satisfaction with quality of the faculty's teaching, the practicum component of the MSAN curriculum, and the program's value proposition.
2. **A strong value proposition:** In a recent exit survey, a graduate of the program indicated that "It was worth it, quitting my job and making [this] investment." Another said "I learned things that I didn't know about or how to do. I got a job in the career that I wanted." Still another said "What I know now compared to a year ago—the

change is insane." We seek to transform people with a wide variety of backgrounds and aptitudes into highly skilled and in demand data science positions within 12 months.

3. **Practitioner engagement and recognition:** Each year, the program will meaningfully engage with data-driven Bay Area companies. This engagement will extend to a number of companies approximately equal to one-third of the program's current enrollment. It will also be measured by practitioner participation in the Analytics Seminar Series, which is jointly sponsored by the MSAN program and the Data Institute.
4. **Increased national and international visibility:** Over time, the program's faculty will increasingly be well-known for making theoretical and applied contributions to fields such as computer science, statistics, and business. The program will routinely be identified with other top-tier analytics programs (e.g., Columbia University, Georgia Tech, Northwestern University, Georgetown University, Johns Hopkins University, New York University, etc.).
5. **A community of scholars:** Faculty members will actively and regularly pursue scholarly activities, including the publication of high quality research papers and books, grant-writing and submission, and dissemination of their work at important conferences and colloquia in their respective fields. The faculty will share their interdisciplinary expertise with one another and build a culture of research collaboration with each other as well as with external collaborators.
6. **A community of teachers:** Faculty members will establish a culture of helping one another to become more successful teachers (e.g., open and constructive and voluntary discussion of BLUE results, voluntary classroom visits to secure feedback and suggestions for improvement, pedagogical innovations, etc.).
7. **A culture of service to the program:** Faculty members will participate in recruitment and yield events, conduct technical admissions interviews, support the new Data Institute, and vigorously engage in deliberations related to the program's curriculum.

Notes: These program goals were ratified by the program's faculty during a May 2016 vote over email.

4 Program Learning Outcomes (PLOs)

Upon successfully completing the the Master of Science in Analytics (MSAN) program, our graduates will:

(1) Possess a solid theoretical understanding of—and ability to apply—classical statistical models (e.g., generalized linear models, linear time series models, etc.), machine learning techniques (e.g., boosting, random forests, neutral networks, etc.), and classification techniques (e.g., naive Bayes, k-means, spectral clustering, etc.).

(2) Effectively use modern programming languages (e.g., R, Python, SQL, etc.) and technologies (AWS, Hive, Spark, Hadoop, etc.) to scrape, clean, organize, query, summarize, visualize, and model large volumes and varieties of data.

(3) Be prepared for careers as analytics professionals by integrating with the Bay Area data science community, solving real-world data-driven business problems with members of this

community, developing professional presentation and interviewing skills, and obtaining some domain expertise.

Notes: Please refer to the attached curriculum map relating program learning outcomes to courses in the MSAN curriculum.

5 Academic Program Review

The Master of Science in Analytics Program was founded in academic year 2012-2013 and has, therefore, not yet had its first formal program review. That program review is scheduled for academic year 2019-2020.

6 2015-2016 Assessment Plan

In academic year 2015-2016, the program's faculty decided to investigate the following two related curricular questions: (1) Is there evidence that our students successfully learn basic SAS programming skills in MSAN 690 (Introduction to Programming in SAS)? and (2) Should this course be retained or replaced by a more appropriate alternative? To answer these two questions, we will pursue two different lines of investigation with two different methodologies.

Note that both the first question ("Is there evidence that our students successfully learn basic SAS programming skills?") and the second question ("Should the SAS programming course be retained?") relate to the program's second learning outcome.

6.1 The First Question.

To answer the first question, the MSAN program faculty will rely on the results of each cohort[1] on the Base SAS Certification Examination. The Base SAS certification is the first in a family of SAS Certification credentials[2] which"are globally recognized as the premier means to validate SAS knowledge." The exam is administered by SAS and Pearson VUE locally at USF for MSAN students and consists of 60-65 multiple-choice and short-answer questions for which the students are given 110 minutes to complete the exam. Passing this exam consists of getting a score of 70% or higher.

To pass MSAN 690, students must pass the Base SAS Certification Examination. The course is only offered on a pass/fail basis. The faculty will use the results from each cohort's first collective attempt to pass the examination as a direct method of assessing the effectiveness of the course at introducing students to the basic principles (and idiosyncrasies) of programming in SAS.

Results. The pass rates were as follows: for cohort one, 8 out of 12 students; for cohort two, 23 out of 24 students; for cohort three, 32 out of 33; for cohort four, 25 out of 35. With

[1] Four cohorts of MSAN students have graduated so far.

[2] `http://support.sas.com/certify/`

the exception of cohort four—the cohort that, in fact, lobbied the faculty to remove SAS from the curriculum—pass rates were generally high. So, our students were learning the SAS programming language. Moreover, while we have decided to measure initial pass rates, final pass rates (i.e., pass rates after two or more attempts at the exam) approach 100% over the program's four cohorts.

Perhaps the more relevant questions become: Were our students using SAS in their practicum experiences? Do our alumni use SAS at their current jobs? Do Bay Area employers want to see data scientists with SAS programming skills?

6.2 The Second Question.

Methods. The MSAN program faculty used several methods to ascertain the importance of SAS programming skills to the Bay Area data science community. First, in the middle of the 2015-2016 academic year, we surveyed then-current students to assess the technologies they use in their practicum projects with Bay Area companies. We also separately surveyed this same group of students about a variety of both curricular and non-curricular issues as they exited the program. Finally, we informally surveyed our alumni to determine if any of them are using SAS at their current jobs. Because our alumni are largely employed in the Bay Area and our practicum partners are Bay Area employers, we expect that these indirect methods of assessment will give us better insights into the value of having our students learn the SAS programming language.

Second, we undertook a brief quantitative textual analysis of a sample of job listings (for data scientists, data analysts, business intelligence professionals, etc.) in the Bay Area. The intensity with which we see words like "R" and "Python" used in these job listings—as opposed to "Tableau" or "SAS"—was expected to provide the program's faculty with additional evidence of what prospective employers find valuable in the MSAN curriculum. We expect that the results of this indirect method will provide confirmation of the results of the various surveys we used. We note that our data gathering methodology complements the report by RJMetrics (as described below) in that our approach was to aggregate our text analysis on job listings as opposed to aggregating on LinkedIn profiles.

Results. In a 2016 study of skills associated with LinkedIn profiles by RJmetrics[3] (and also reported on by *Forbes*), "data analysis" was the skill most frequently listed by self-identified data scientists. Of the specific software skills listed, R was the most common, followed closely by Python. See, for example, the chart from RJMetrics in Appendix A or the complete report in Appendix H. SAS programming skills, while rated one of the top twenty skills for a data scientist to possess, is nonetheless not very high on the list. The results of this study are confirmed by our informal email survey of our alumni (see Appendix D), very few of whom indicated that they use SAS in their current jobs.

In our own textual analysis of 6,500 LinkedIn job postings for data scientists or data analysts (see Appendix B), we found that only 28% of the postings mentioned SAS. In fact, SAS was the 122$^{\text{nd}}$ highest-ranked n-gram, getting trounced by machine learning (mentioned

[3] `https://www.stitchdata.com/resources/reports/the-state-of-data-science/?thanks=true`

in 75% of job postings), big data (68%), statistics (67%), R (63%), Python (55%), and SQL (44%). Hadoop, a somewhat obscure open-source software framework for distributed storage and distributed processing of very large data sets on computer clusters built from commodity hardware, was mentioned in 44% of job postings. But Hadoop does not receive its own course in the MSAN curriculum. Neither does MATLAB, which was mentioned in 28% of job postings. Hence, the evidence from job postings would seem to suggest that SAS is overemphasized in the current MSAN curriculum.

Finally, our own surveys of our current students at the time (i.e., the fourth cohort) suggested that SAS was viewed as not very useful (see Appendix C), particularly for practicum projects (see Appendix E).

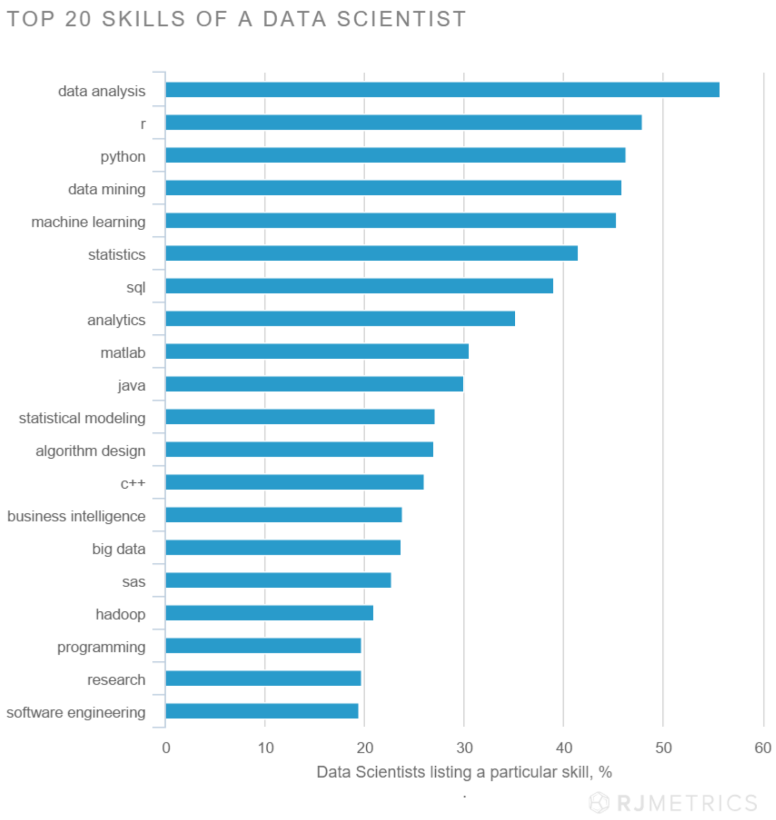
7 Closing the Loop

At its January 2016 faculty meeting—and after a thorough discussion—the faculty unanimously voted to remove MSAN 690 (Introduction to SAS Programming) as a curricular requirement (see a syllabus for the course in Appendix G). MSAN 690 is typically offered during the winter interession term. The distributed computing course (MSAN 694) was moved into the winter intersession term and upgraded from a one-unit course to a two-unit course. In its place, the faculty voted to create a new course, Application Development for Data Science (MSAN 698), which will be a new graduation requirement in academic year 2016-2017.

A Top Skills for Data Scientists

A recent study conducted by RJMetrics[4]—a Bay Area firm that "builds data infrastructure and analytics software to help businesses make smarter decisions with their data"—showed that the cumulative number of data scientists has grown at an exponential rate since 1995. Data scientists are most highly concentrated at firms like Microsoft, IBM, Booz Allen Hamilton, General Electric, LinkedIn, Hewlett Packard, Twitter, Google, Oracle, and AT&T.

RJMetrics also analyzed 254,000 profiles of data scientists on LinkedIn and ranked each skill by the number of people listing it on their profile. In addition to broad categories like "data analysis," "data mining," and "analytics," the top skills are R, Python, machine learning, statistics, SQL, MATLAB, Java, statistical modeling, and C++. Hadoop (20.9%) is at the bottom of the list of top twenty skills—just behind SAS, which was identifed in 22.78% of profiles.



[4] https://www.stitchdata.com/resources/reports/the-state-of-data-science/?thanks=true

B Text Mining of Job Descriptions

We downloaded 6,500 Bay Area job postings associated with the phrases "data scientist" and "data analyst" and tokenized (i.e., stripped away white space, punctuation, etc.) the words in the job postings. We then constructed all n-grams from the corpus and computed their relative frequencies. Selected results from the top 250 results are shown below. We eliminated from display below various n-grams unrelated to the curriculum, e.g., words such as "great," "ideal," "over," "experience," "fun," etc.

SAS was ranked 122nd among the n-grams. It ranked below MATLAB (which is not featured in the MSAN curriculum) and Java (which is only occasionally taught in the data visualization course). In contrast, R and Python—which are used intensively throughout the curriculum—were ranked 16th and 24th, respectively. The faculty's conclusion is that while SAS may be an important analytical tool in some industries, or with some employers, or in some regions of the country, it is not of critical importance for our students (who generally secure Bay Area jobs) to learn.

Rank	Word/Phrase	Proportion	Rank	Word/Phrase	Proportion
1	data	1.00	88	predictive	0.33
2	team	0.83	101	visualization	0.32
4	science	0.80	104	Java	0.32
8	machine learning	0.75	105	physics	0.31
10	business	0.68	118	data analysis	0.28
12	big data	0.68	119	MATLAB	0.28
13	statistics	0.67	121	engineers	0.28
14	models	0.67	122	SAS	0.28
15	analysis	0.64	131	MapReduce	0.27
16	R	0.63	158	Spark	0.24
17	mining	0.63	159	Ph.D.	0.24
20	data scientist	0.61	160	improve	0.24
22	data mining	0.59	162	C++	0.24
24	Python	0.56	180	scalable	0.20
27	computer	0.55	199	predictive analytics	0.20
29	analytics	0.53	199	strategy	0.20
32	computer science	0.51	211	Hive	0.20
34	problems	0.48	218	statistical analysis	0.20
35	research	0.48	219	code	0.20
37	engineering	0.48	223	actionable insights	0.20
42	data science	0.45	224	multivariate	0.20
44	SQL	0.44	225	decision	0.20
48	Hadoop	0.44	240	validate	0.19
49	algorithms	0.44	244	technologies	0.19
66	mathematics	0.40	250	data science team	0.19

C Exit Survey from Cohort Four

Since its inception, the MSAN program has issued exit surveys to all of its graduates. These surveys have played a critical role in informing faculty decisions about the curriculum. The program's most recent cohort of graduates were asked an open-ended question: "What class or topic would you remove from the curriculum?" The tabulation of the results is as follows:

Course or Topic	Count
SAS (MSAN 690)	15
Business Communications (MSAN 610)	6
Geographic Information Systems (an elective, MSAN 631)	4
Multivariate Statistical Analysis (MSAN 623)	2
Interviewing Skills (MSAN 696)	2
Business Strategies for Big Data (MSAN 603)	1
Distributed Computing (MSAN 694)	1
Web Analytics (MSAN 695)	1
Javascript (a topic)	1

It is natural to think that MSAN students, who are predisposed to favor technical material, might resist a more qualitative course like Business Communications or Interviewing Skills or Business Strategies for Big Data. But the Introduction to SAS Programming (MSAN 690) course stands out among these results—15 out of 33 respondents drop the SAS programming course from the curriculum.

In another sequence of open-ended questions, graduates of the fourth cohort were asked the following: "Thinking back over the year, please list the three topics or technologies that were presented that you found LEAST useful." The responses are organized as follows:

Response	Count
SAS (MSAN 690)	19
Business Communications (MSAN 610)	9
Geographic Information Systems (or Google Earth Engine) (MSAN 631)	7
Business Strategies for Big Data (MSAN 603)	6
Nothing. Everything was useful.	5
Javascript or D3	5
Hive	4
Multivariate Statistical Analysis (MSAN 623)	3
Interviewing Skills (MSAN 696)	3
Distributed Computing (MSAN 694)	3
Web Analytics (MSAN 695)	3
Data Visualization (MSAN 622)	2
Data Acquisition (MSAN 692)	2
NoSQL (MSAN 697)	2
Too many theoretical homework problems in Time Series Analysis (MSAN 604).	1
Using algebra to solve problems.	1
Linear regression theory.	1
Too much homework in general.	1

While the intention of this survey question was to focus on topics within courses (e.g., Box-Jenkins models, principal component analysis, deep learning, etc.) or technologies (e.g., Hive, Javascript, Python, R, SAS, etc.), students often responded with the names of particular courses. Regardless of whether or not you believe that a student writing the word "SAS" is referring to the *course* or the *technology*, the conclusion seems clear: out of the 77 responses, 19 of them (or 25%) would seem to recommend eliminating the course or topic from the program.

D Informal Email Survey of Current Alumni

The faculty were concerned that in spite of evidence from industry (or job descriptions), and in spite of our *current* students not feeling that learning SAS was useful, it might be the case that *our* alumni do use SAS (or are exposed to SAS) at work. We emailed our alumni list, which contains approximately 100 persons. Only four of the 60 respondents indicated that they use SAS in some capacity in any of the posititions they have taken since completing the MSAN program.

E Technology Survey

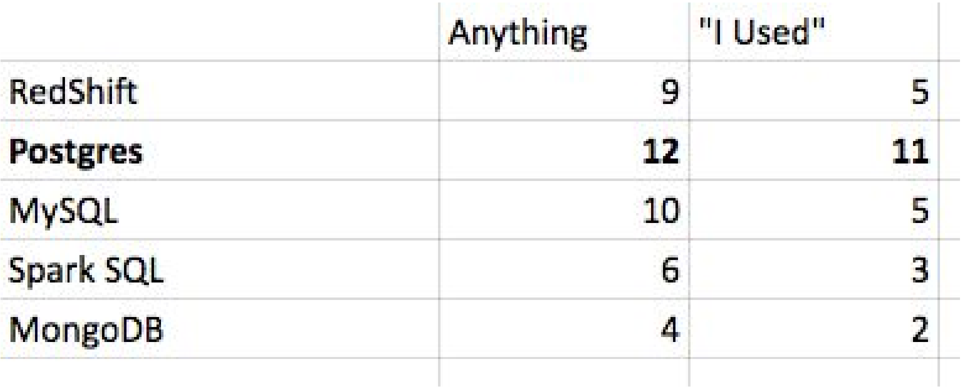
Nicholas Ross, an MSAN faculty member, conducted a survey during the middle of the 2015-2016 academic year. The purpose of this survey was to determine *which* technologies our students were being asked to use for their practicum projects. While the results of this survey were used to have a general discussion (at the faculty's annual winter meeting) about the state of the program's curriculum, we can also view the results of a survey as an opportunity to affirm the importance of teaching SAS in the MSAN curriculum. However, *none* of our students reported using SAS during the course of their practicum projects. The presentation that Nick Ross made to the faculty is included below.

Practicum Technology Survey 2016

Info

- 27 Students Responded of 42 (65%)
- In future:
    - Be broader with definition of data management. Students were putting “other” all over the place
    - Better Venn Diagram w.r.t. companies use vs. students use.
    - Add more packages (ggplot, etc.) to the choices, as students will write them in anyway.
- There were 3 choices for each technology:
    - I used it frequently for the practicum
    - I used it occasionally for the practicum
    - Technology used by the practicum company, but I did not use it personally
- On Summary, first two were combined
- Thanks Kirsten!

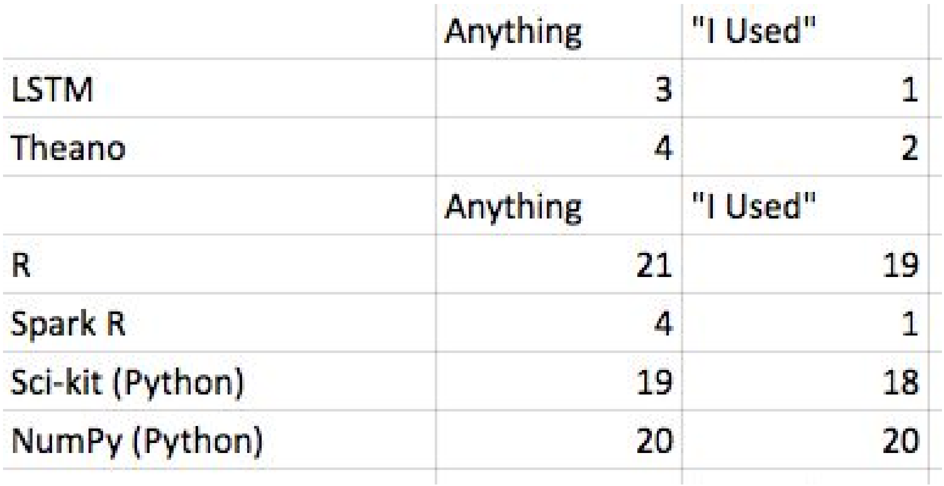
SQL / NoSQL



	Anything	"I Used"
RedShift	9	5
Postgres	**12**	**11**
MySQL	10	5
Spark SQL	6	3
MongoDB	4	2

Others Mentioned:
2 MSFT, 1 Oracle, 1 Dynamo and 1 Impala

Deep Learning / Statistical Tools

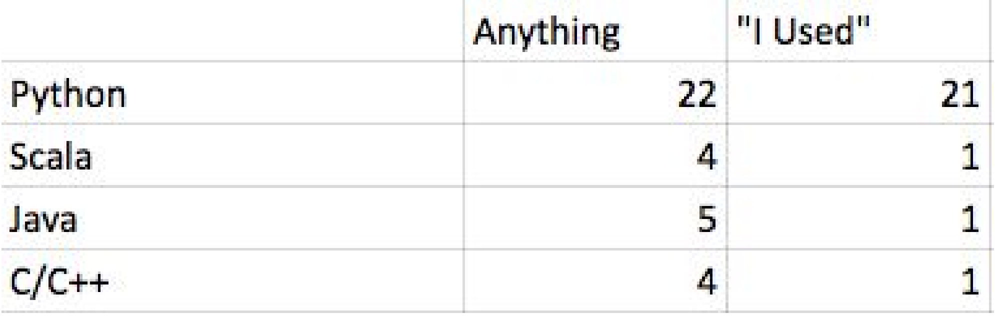


	Anything	"I Used"
LSTM	3	1
Theano	4	2
	Anything	"I Used"
R	21	19
Spark R	4	1
Sci-kit (Python)	19	18
NumPy (Python)	20	20

Others Mentioned:
None for Deep Learning

Statistical Tools:
Pandas x3, NLTK

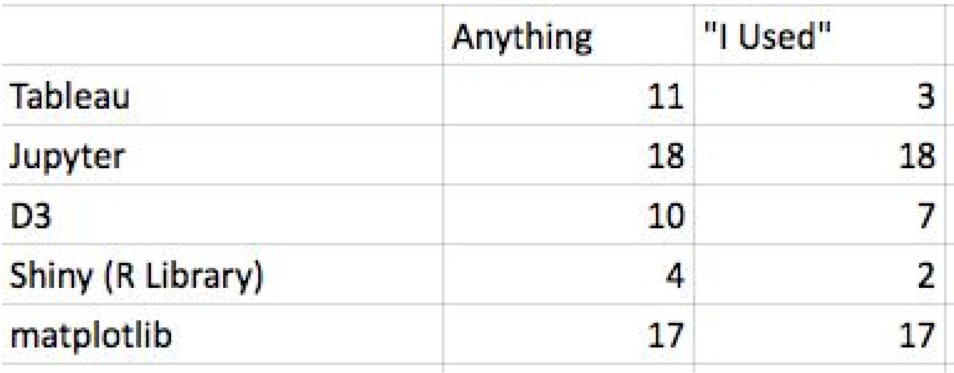
General Programming



	Anything	"I Used"
Python	22	21
Scala	4	1
Java	5	1
C/C++	4	1

Others Mentioned:
Shell, JavaScript 3x,
HTML and CSS

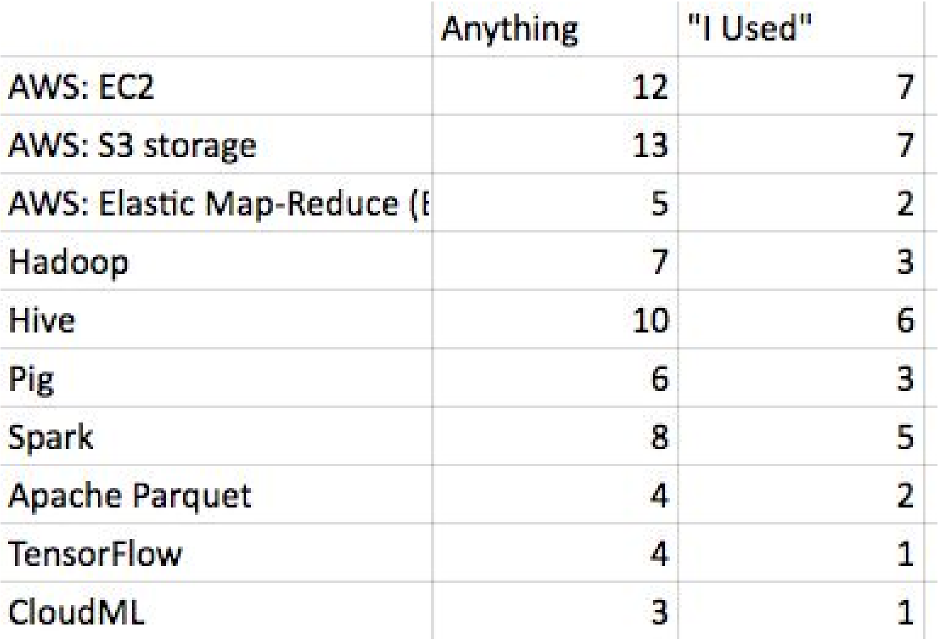
Visualization



	Anything	"I Used"
Tableau	11	3
Jupyter	18	18
D3	10	7
Shiny (R Library)	4	2
matplotlib	17	17

Others Mentioned:
Ggplot 5x, plotly, bokeh
and highchart.js

Other



	Anything	"I Used"
AWS: EC2	12	7
AWS: S3 storage	13	7
AWS: Elastic Map-Reduce (E	5	2
Hadoop	7	3
Hive	10	6
Pig	6	3
Spark	8	5
Apache Parquet	4	2
TensorFlow	4	1
CloudML	3	1

Others Mentioned:
Qubole

F Minutes of MSAN Faculty Meetings

The MSAN faculty typically meets 2-4 times per year. At nearly every one of these meetings, the program's curriculum is discussed. We attach, as evidence of our ongoing and vigorous discussions of the curriculum, or minutes from our winter 2016 and fall 2016 meetings.

Faculty Meeting Minutes 1/22/2016

Quality of applicants is much higher this year
Interview questions are harder
Accepts are trending away from India and toward China
More Americans applying this year
Over 30 deposited as of today
13 internationals, 16 women
Some of our best applicants come in at the March 1 deadline

We are aiming at 60 admits this year
We are killing discussion sections
We are running all courses twice - instructors do double lectures. This will keep classes relatively small.
This way students can take advantage of office hours as well.
35 is too many students for Bus Comm. Maybe split into 3 sections.

Next year:
Bus Comm will be one unit
Distributed Computing will be two units and move into Intersession

One idea for intersession:
Students get Hadoop certification and/or Spark certification instead of SAS.
Students need to learn this earlier than spring for their practicums.

Need to replace the one unit class!
Idea 1: Mike's Data Science as a Service
Students want linux mastery>>could be in data acquisition.
Rebranding NoSQL course - same basic content
Rename distributed computing Hadoop and Spark: Voted...DONE
Idea 2: Kaggle competition class
Idea3: optimization class

Voting: New 1 unit course is:
Data Science as a Service (DSaaS)...or something….
Application Development for Data Science

This is in addition to NoSQL course.

Aiming for 75 deposits, and then melt to 60 after bootcamp.
MSAN will make a large space acquisition this coming year.
Large offices on 5th floor and both 5th floor classrooms.

Practicum:
23 projects at 22 companies
Beebell is being shut down
City of Hope is over
Netbase - lost our contact

Need to put logos on the website company logos and school logos that they come from

10 of 22 companies pay students
median salary is around $20/hour
left 8 companies with no students
Might start BART in March
ChannelMeter needs our students badly

Currently soliciting 49ers, Disney, Kaiser, Upwork, Wikia

Mentorships next year: we won't know until we hire the new faculty
Also BSDS hire will want to be a mentor

Analytics Institute
in 2012 we were 1 of 5 programs, now we are 1 in 100. We can continue to grab market share by developing an institute.

1. Introducing visiting positions, grants, post-docs, a chair, and a USF Analytics Conference.
2. Strong and lasting corporate partnerships, certifications, exec ed, and in-house partners, building adjunct pool and building strong relationships with companies.
3. Developing courses for companies
4. Training future adjuncts
5. Developing online masters program for AT&T etc.
6. Specializations
7. Industry partners pay $25,000 per year for membership
8. large companies pay $10,000 for a practicum team
9. exec ed, corporate bootcamps and certifications generating $$
10. A named benefactor will be sought by 2019 (better a person than a company)

However, University has killed 4 online masters programs over the last year. Online could hurt our reputation.

G Syllabus for MSAN 690

We include a syllabus from MSAN 690 (Introduction to SAS Programming), the course that was eliminated from the curriculum as a result of this year's assessment exercise.

MSAN 690 — Intro. to SAS Programming
Instructor: Jeff Hamrick
Course Syllabus
Winter Intersession 2016

SUMMARY INFORMATION
Instructor: Jeff Hamrick, Ph.D., CFA, FRM
Office: Lone Mountain Rossi (LMR) 422
Office Hours: By appointment.
Cell Phone: 617/943-4619
Office Phone: 415/422-6810
Email Address: `jhamrick@usfca.edu`

Class Location: 101 Howard Street, Room 529
Class Times: 9:00 a.m. - 5:00 p.m., Monday through Friday

ON COURSE OBJECTIVES. Any student who successfully completes this course should be able to:

- Navigate the SAS windowing environment;
- Navigate the SAS Enterprise Guide programming environment;
- Read various types of data into SAS data sets;
- Create SAS variables and subset data;
- Combine SAS data sets;
- Create and enhance listing and summary reports;
- Validate SAS data sets;
- Control SAS data set input and output;
- Combine SAS data sets;
- Summarize, read, and write different types of data;
- Perform DO loop and SAS array processing;
- Transform character, numeric, and date variables;
- Identify and correct data, syntax, and programming logic errors in SAS; and
- Prepare for the furst SAS Certification Examination.

ABOUT YOU. You should be hard-working and enthusiastic about learning and, in most cases, you are a candidate for the Master of Science in Analytics at the University of San Francisco. See me immediately if you are not!

ABOUT US. We will meet to talk about programming in SAS from Monday, January 11, 2016 to Wednesday, January 20, 2016. We will meet at the University of San Francisco's downtown campus at 101 Howard Street in room 529. We will use selected the third edition of the SAS Certification Prep Guide (Base Programming for SAS 9). The ISBN for this book is 978-1-60764-924-3. It is on reserve and is available for minimal cost through online book vendors like Amazon Prime.

ON SAS. SAS is a software suite that can mine, alter, manage and retrieve data from a variety of sources and perform statistical analysis on it. SAS programs have a DATA step, which retrieves and manipulates data, usually creating a SAS data set, and a PROC step, which analyzes the data. SAS is developed and maintained by the SAS Institute. It was originally developed at North Carolina State University from 1966 through 1976. Today, SAS is widely used for statistical analysis and predictive analytics, particularly in industries like finance, healthcare, and logistics. SAS holds the largest market share in "advanced analytics" with about 1/3 of the market as of 2016.

ON ATTENDANCE. This course is an intensive winter intersession course, and you are expected to pass the Base SAS Programming Examination at the end of this two-week period of time. Consequently, you may only miss class under the most dire of circumstances. These circumstances should be both unusual **and** documentable. For example, having a bad cold is documentable but not unusual. On the other hand, being kidnapped by aliens is unusual but is most likely not documentable.

ON HOMEWORK. This class will not have homework in the traditional sense. Instead, selected problems will be given to you at the end of class each day so that you can review the material we just learned. As we near the end of the course, you will be expected to work on practice examinations on your own time. You'll be asked to grade these practice examinations on your own (i.e., to self-assess), but doing so is a critical part of preparing for the Base SAS Programming Examination.

ON QUIZZES. We will take at least one practice quiz in class each day. These practice quizzes are intended to help you review the material that we learn each day. These quizzes will be self-graded.

ON THE FINAL EXAMINATION. There will be no final written comprehensive examination in this course. However, we will finish lectures on Wednesday, January 20, 2016. On Thursday, January 21, 2016, you are expected to spend the day studying for the Base SAS Certification Examination, reviewing SAS syntax, taking practice examinations, etc. On Friday, January 22, 2016, we will meet in the afternoon on the main campus at a computer laboratory and you will take the Base SAS Certificate Examination for the first (and hopefully last) time. Professor Terence Parr and I will proctor the examination together.

ON GRADING. This course is offered on a pass/fail basis only. You pass this course if, and only if, you pass the Base SAS Certification Examination after the course ends. You will have a total of three attempts before your grade in this course is set to failing. The Master of Science in Analytics program will pay for your first seating of the Base SAS Certification Examination. You will have to pay for the second and third seatings—should you require them. With my approval, and the approval of the MSAN program director, you can attempt the examination more than three times and then, when you pass, have your failing grade changed to a passing grade.

ON CHEATING. As a Jesuit institution committed to *cura personalis*—the care and education of the whole person—the University of San Francisco has an obligation to embody and foster the values of honesty and integrity. The university upholds standards of honesty and integrity from all members of the academic community, including faculty, students, and staff. All students are expected to know and to adhere to the university's honor code. You can find the full text of the code online at `http://www.usfca.edu/catalog/policies/honor/`.

Specific examples of academic dishonesty include, but are not limited to, the following:

1. While you *may* be required to work in groups with other students on homework assignments or class projects, you should not allow your name to be placed on a group write-up if it does not reflect your own understanding of the material and if you have not made an honest, equitable contribution to the group effort.
2. Copying answers from other students or other sources during a quiz or examination is a violation of the university's honor code and will be treated as such.
3. Plagiarism consists of copying material from any source and passing off that material as your own original work. Plagiarism is plagiarism: it does not matter if the source being copied is on the Internet, from a book or textbook, or from quizzes or problem sets written up by other students.

All incidents of cheating will be reported to the director of the MSAN program. The policy of the MSAN program is the following:

1. The first observed incident of cheating will result in a zero on the quiz or the assignment (for example). It will be reported to both the MSAN program director and the MSAN program manager for tracking.
2. The second observed incident of cheating **in any course** after the initial incident will result in a failing grade for the course. A student receiving a failing grade in **any** MSAN course cannot proceed to the next module of the program. These students will leave the program.

If you ever have questions about what constitutes plagiarism, cheating, or academic dishonesty in my course, please feel free to ask me. I'm happy to discuss the issue in a forthright manner.

ON DISABILITIES. If you are a student with a disability or disabling condition, or if you think you may have a disability, please contact USF Student Disability Services (SDS) at 415/422-2613 within the first week of class, or immediately upon onset of the disability, to speak with a disability specialist. If you are determined eligible for reasonable accommodations, please meet with your disability specialist so they can arrange to have your accommodation letter sent to me, and we will discuss your needs for this course. For more information, please visit `http://www.usfca.edu/sds/` or call 415/422-2613.

ON BEHAVIORAL EXPECTATIONS. All students are expected to behave in accordance with the University's Student Conduct Code, as well as other University policies. Open discussion and disagreement are encouraged when done respectfully and in the spirit of academic discourse. There are also a variety of behaviors that, while not against a specific University policy, may create disruption in this course. Students whose behavior is disruptive or who fail to comply with the instructor, may be dismissed from this class for the remainder of the class period and may need to meet with the instructor, program director, or associate dean prior to returning for the next class lecture. If necessary, please see `http://www.usfca.edu/fogcutter/student-conduct`, or refer to the MSAN Program Policies that you agreed to adhere to at the start of the program.

ON COUNSELING AND PSYCHOLOGICAL SERVICES. The University's diverse staff offers brief individual, couple, and group counseling to student members of our community. CAPS services are confidential and free of charge. Call 415/422-6352 for an initial consultation appointment. Telephone consultation through CAPS After Hours is available between the hours of 5:00 p.m. and 8:30 a.m. Call the above number and press two.

ON SEXUAL HARASSMENT AND ASSAULT. As an instructor, one of my responsibilities is to help create a safe learning environment. I also have a *mandatory reporting responsibility* related to my role as a faculty member. I am required to share information about sexual misconduct or information about a crime that may have occurred on the University's campus with the University. Here are other resources:

- To report any sexual misconduct, students may visit the Title IX Coordinator on the 5th floor of the University Center.
- You can also visit `https://myusf.usfca.edu/title-ix` for more information.
- Students may speak to someone confidentially, or report a sexual assault confidentially, by contacting Counseling and Psychological Services at 415/422-6352.
- To find out more about reporting a sexual assault at USF, visit the University's Callisto website at `https://usfca.callistocampus.org`.
- For an off-campus resource, contact San Francisco Women Against Rape (SFWAR) at 415/647-7273 (`www.sfwar.org`).

ON YOUR STUDENT ACCOUNT. Students who wish to have tuition charges reversed on their student account should withdraw from this course by the end of the business day on the last day to withdraw with tuition credit (census date). Please note that the last day to withdraw with tuition credit may vary by course. The MSAN program's policy for courses in winter intersession 2016 is that students must drop classes by January 15, 2016 in order to receive a full tuition refund. You should consult with the MSAN program director and staff in order to do so.

ON LAPTOPS. In general, you should have a laptop in class and you should have SAS installed on that laptop before the course begins. You will be expected to use SAS in class. I would ask you to be respectful of me and your classmates and to refrain from surfing the web, checking out Facebook, tweeting people your various tweets, etc., during the middle of my lectures. It is absolutely forbidden to use instant messaging programs, email, etc. during class lectures or quizzes.

H Complete report from RJMetrics

The State of Data Science

Like 0 Tweet Share 0 G+1 0

Key Findings

We found 11,400 data scientists currently employed by companies known to LinkedIn. This number is a conservative estimate of data science professionals and includes only those who explicitly identify themselves as data scientists on LinkedIn. We analyzed in detail 60,200 associated records of professional experiences, 27,700 records of education, and 254,600 records of skills. We also analyzed information about 6,200 unique companies that employed self-identified data scientists as of June 1, 2015.

Growth

The number of data scientists has doubled over the last 4 years.

Top Industry

The Information Technology and Services industry employs the largest number of data scientists.

Top Skills

The top five skills listed by data scientists are: Data Analysis, R, Python, Data Mining, and Machine Learning.

Education Level

Over 79% of data scientists that list their education have earned a graduate degree, and 38% have earned a PhD.

Top Backgrounds

The majority of data scientists come from STEM graduate-level backgrounds, with Computer Science, Statistics, Mathematics and Physics leading the way. However, there are significant differences at the Master's and PhD levels.

Whereabouts

6,300, or 55%, of identified data scientists are located in the United States, with the United Kingdom, India, France, Canada, the Netherlands, Germany, Spain, Australia and Israel closing out the top 10.

Executive Summary

Few people are more responsible for the rise of the modern data scientist than Jeffrey Hammerbacher. Hammerbacher is the Founder and Chief Scientist at Cloudera, yet he i known for his role in building the original data team at Facebook. Forbes credits

Hammerbacher, along with DJ Patil — the first ever Chief Data Scientist at the White House — with coining the term “data scientist.”

People on Facebook’s data team were originally given one of two job titles: data analyst or research scientist. This was primarily based on academic background: if you had a PhD, then you were a research scientist. Yet, as Hammerbacher describes in his 2009 essay *Information Platforms and the Rise of the Data Scientist*, the work at both levels soon started overlapping and became increasingly varied:

> On any given day a team member might author a multistage processing pipeline in Python, design a hypothesis test, perform a regression analysis over data samples with R, design and implement an algorithm for some data-intensive product or service in Hadoop, or communicate the results of an analysis to other members of the organization in a clear and concise fashion.

The work at Facebook called for a mash-up of skills combining computer science, business, social science, statistics, and more. **It was out of the need to accomplish this growing “multitude of tasks” that the role of the data scientist at Facebook was born**.

And the rest is history. So to speak.

In his 2012 lecture course on the evolution of data science Hammerbacher was quick to dispel the myth that he coined the term "data scientist," instead providing a comprehensive history of the discipline, and singling out the Bell Labs researcher John Tukey as “the first data scientist.”

50 years after Tukey, there continues to be debate about the precise definition of data science. We find that this debate generally revolves around too many opinions and, surprisingly, too little data.

Read on to learn what data says about the state of data science in 2015.

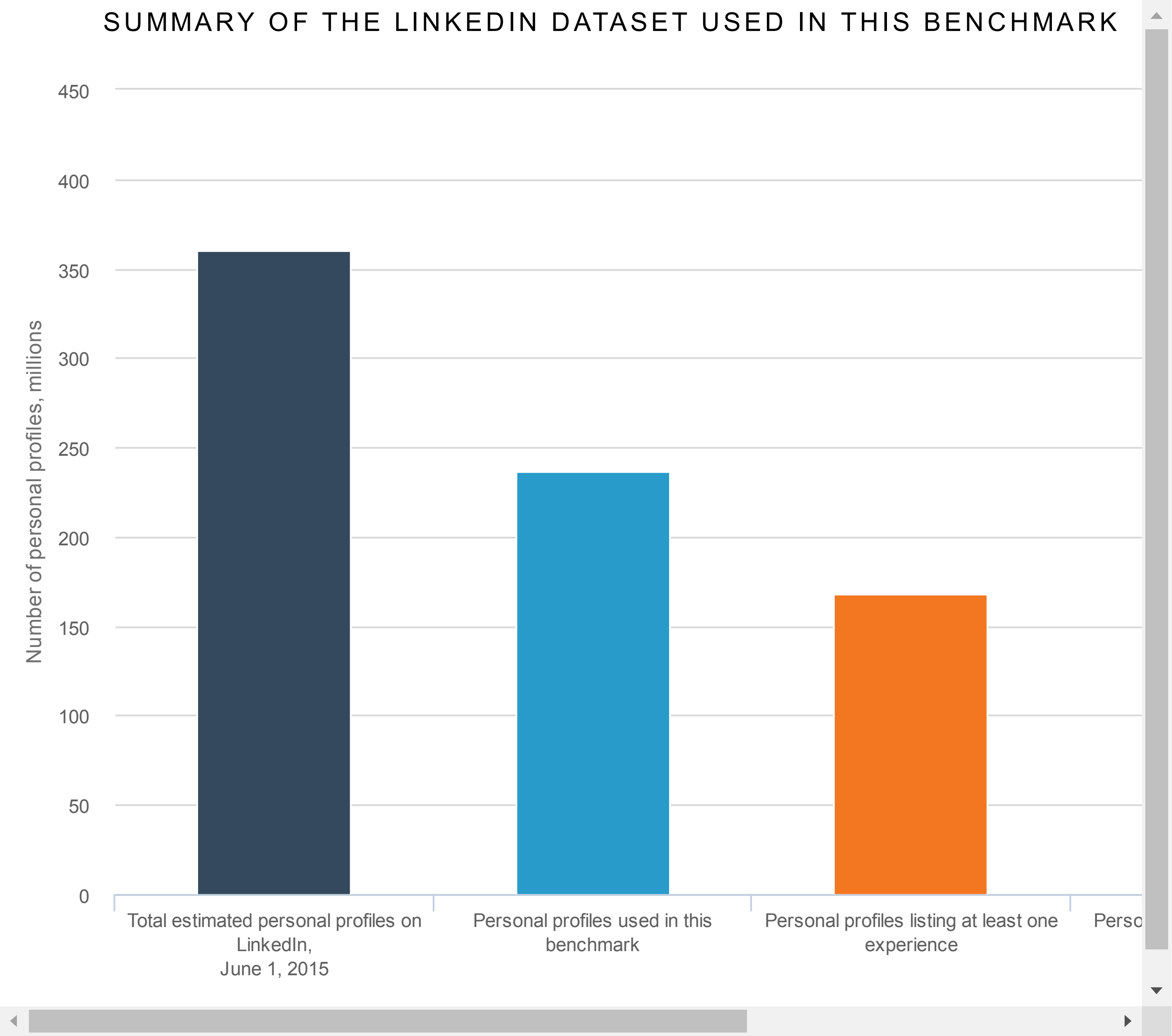
Methodology

Data

This report is based on self-reported information from Linkedin, including all publicly visible personal and company profiles, skills, professional experiences, and education.

We identified data scientists based on their professional headline and current title. We only included data scientists associated with companies we could identify in our sample. We considered the possibility that those listing “data scientist” in their profile without an association with an actual company may only have aspirations about a career in data science, so we did not include those profiles in our analysis.

A summary of the dataset is provided in the chart below.



FOR THE DATA SCIENTISTS IDENTIFIED, WE ANALYZED IN DETAIL:

60k	27k	254k	6.2k
Professional experience records	Education records	Skills records	Unique companies

Analysis Tools

The analysis was carried out in Python, SQL, and an open-source computing platform Jupyter. Python packages charts and python-highcharts were used to create interactive visualizations in HighCharts and HighMaps. Data was stored and processed using Amazon Redshift.

Data Scientists In the World

How many data scientists are there?

Rather than getting lost in the “What is a data scientist?” debate, as so many have done before, we chose to let data scientists speak for themselves. We certainly could have identified data scientists using a complex machine learning algorithm, employing skills, education, keywords, or other identifying characteristics, but the best solution is often the simplest. Similar to how LinkedIn simply asks its users, “does Joe know about Python?” we asked “Does Joe say he is a data scientist?” Specifically, **we looked at people who actually state “data scientist” either in their title or in their professional headline**.

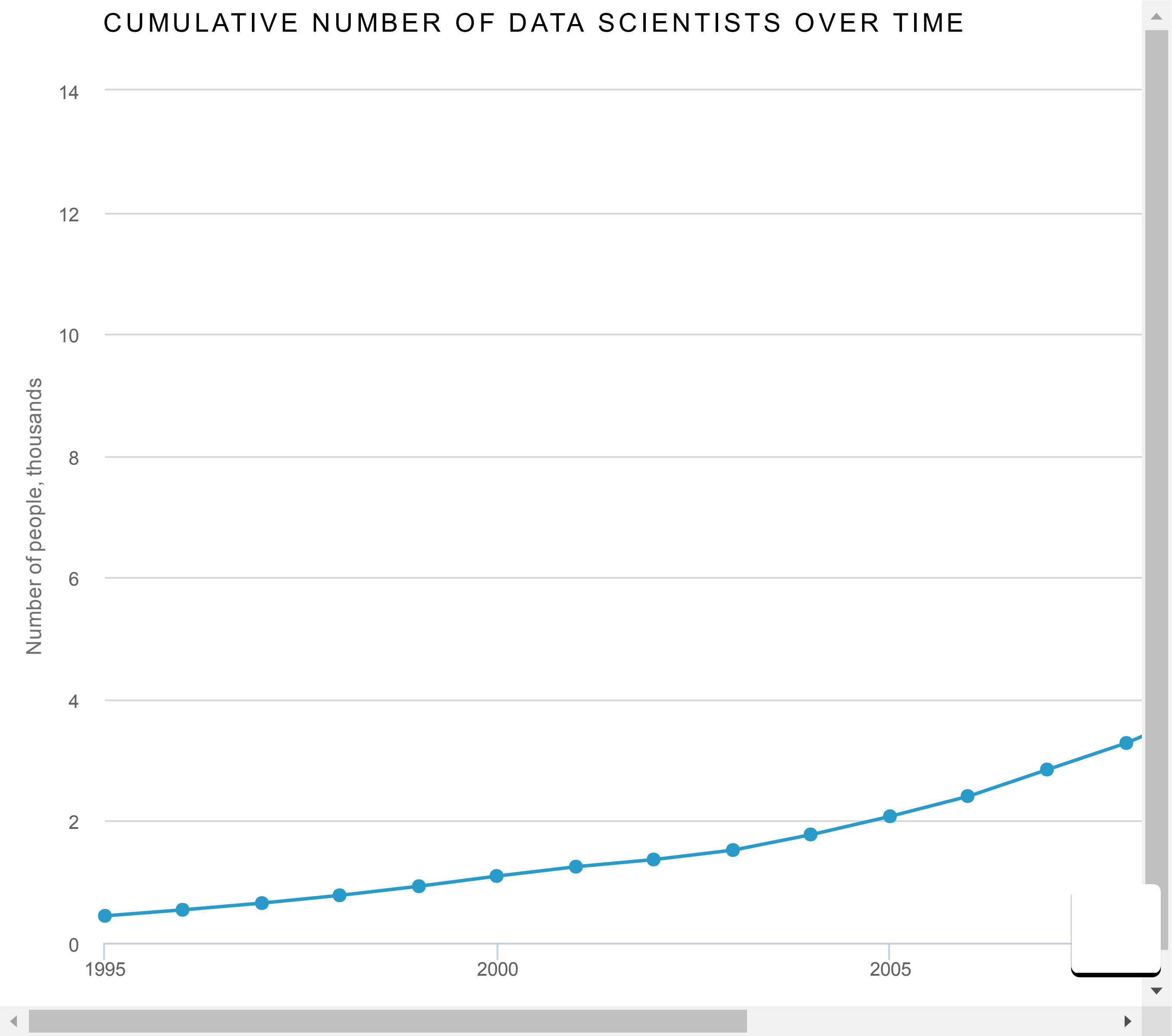
A direct consequence of our approach is that it does not necessarily capture everyone doing data science. Today, many companies employ data analysts, business intelligence analysts, quantitative analysts, or simply scientists who may very well be doing the same work as someone with a data scientist title at another company. However, many companies also employ analysts who do very little with data beyond working with it in Excel. We intentionally avoid including this great diversity of titles so as not to contaminate our sample, and instead consider a very small list of permutations around the phrase “data scientist.”

In addition to searching for data scientists in English, we translated data science titles into eight other languages on LinkedIn: French, Spanish, Italian, Portuguese, German, Swedish, Dutch, and Turkish. If you are curious to see precisely how we identified someone as a scientist, feel free to take a look at the final query that we ran on our Redshift cluster.

All in all, we found only 11,400 data scientists worldwide. While this number seems low at first glance, it is in line with the analysis by LinkedIn's own Senior Data Scientist Peter Skomoroch, who shared his insights in this Quora answer in March of 2014. Taking Skomoroch's estimate of 6,900 self-identified data scientists and factoring in both (a) the 15 months of growth in LinkedIn's user base experienced during that time and (b) the 22% change in the number of new data scientists added between 2014 and 2015 (see the year-over-year chart below), we get 8,900 data scientists. This number is smaller than our estimate, most likely due to the fact that Skomoroch's search excluded other variations of the phrase "data scientist," and was conducted only in English.

How has the number of data scientists changed over time?

Our analysis revealed impressive growth in the number of data scientists over time. In fact, at least 52% of all data scientists have earned that title within the past 4 years.

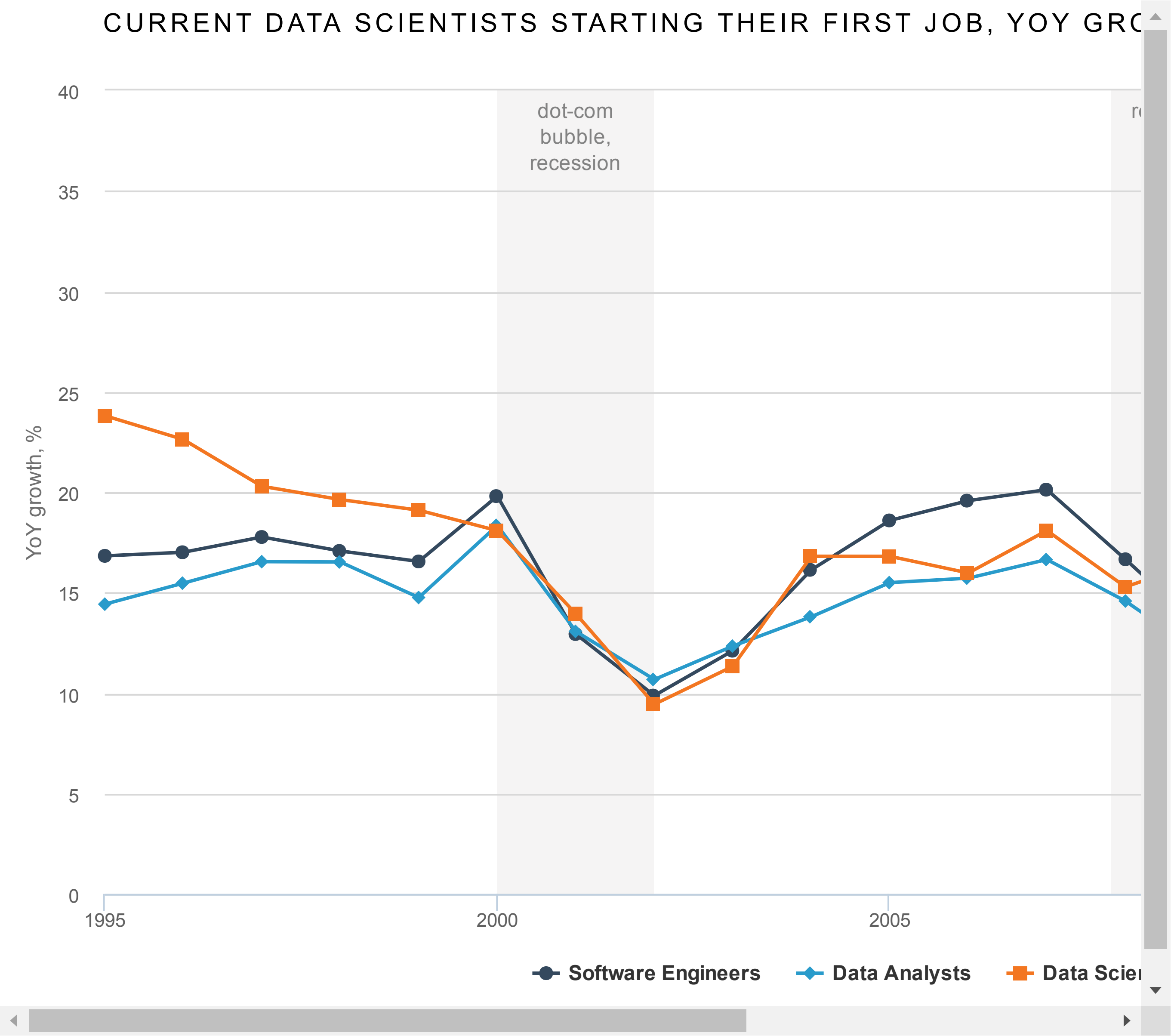


CUMULATIVE NUMBER OF DATA SCIENTISTS OVER TIME

In the chart above, the cumulative number of data scientists in any given year corresponds to the number of present-day data scientists who started their first job that year. Since the first job of someone who is a data scientist today may not have been data-science related, the curve underestimates the growth in the total number of data scientists. One advantage of this approach is that we can see how the number of data scientists grew before LinkedIn was founded, because people list both their professional experience and education prior to 2003.

Note that while LinkedIn has certainly exhibited impressive growth ever since it was founded, our analysis in this chart does not depend on this growth. Specifically, what is important for this type of analysis is how many data scientists have profiles on LinkedIn today, and not how the number of LinkedIn profiles has changed over time.

While this growth in the total number of data scientists is impressive, how does it compare to other technical fields? To answer this question, we looked at the change in the number of people having data science skills starting their first job, and compared that number to two other disciplines: software engineering and data analysis.

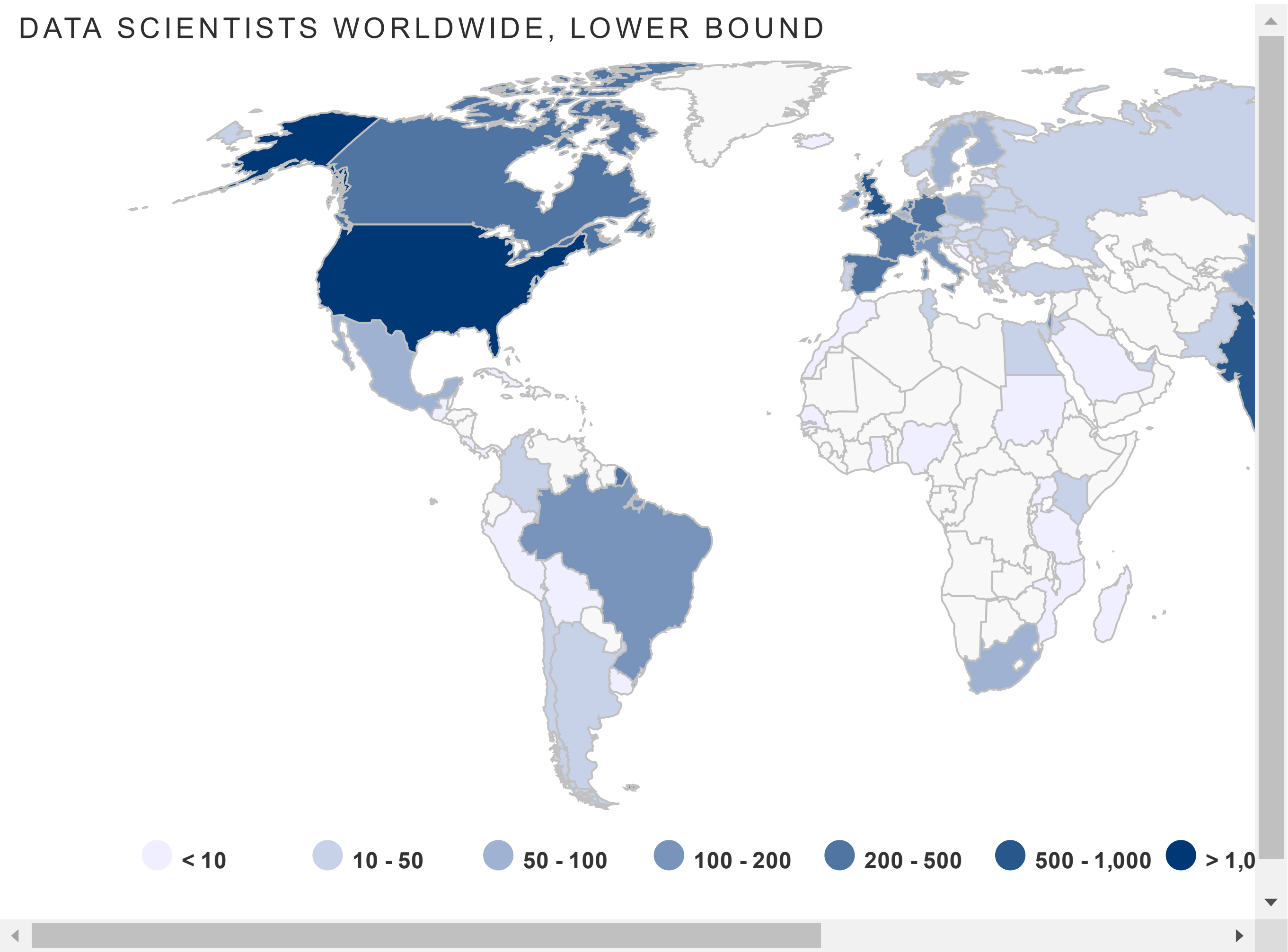


Over time, all three disciplines exhibited similar behavior, including contractions in the number of new people added during the dot-com bubble and the most recent recession. However, **since 2012 the number of data scientists starting their first job has increased at a rate that is consistently 50% higher than that for software engineers and data analysts**.

Note that the year-over-year change shown in this chart is different from the year-over-year growth of the field in general, as we are looking only at the number of people added to the field. Unfortunately, there is no way for us to determine when and how many people have left the field, as they would not be identified as present-day data scientists. However, given how young the field is, we speculate that very few people have left. If one were to take the outflow of people into account, the difference between data scientists and software engineers/data analysts would be even more pronounced.

Where are data scientists located?

55% of all the data scientists on LinkedIn are located in the United States. This makes sense, given that data science originated in the US, and that the US has arguably the highest concentration of high tech companies in the world. However, we were surprised to see hotbeds of data scientists in countries like India, the Netherlands, and Israel. All three made it into the top 10 countries, ranked by the absolute number of data scientists.

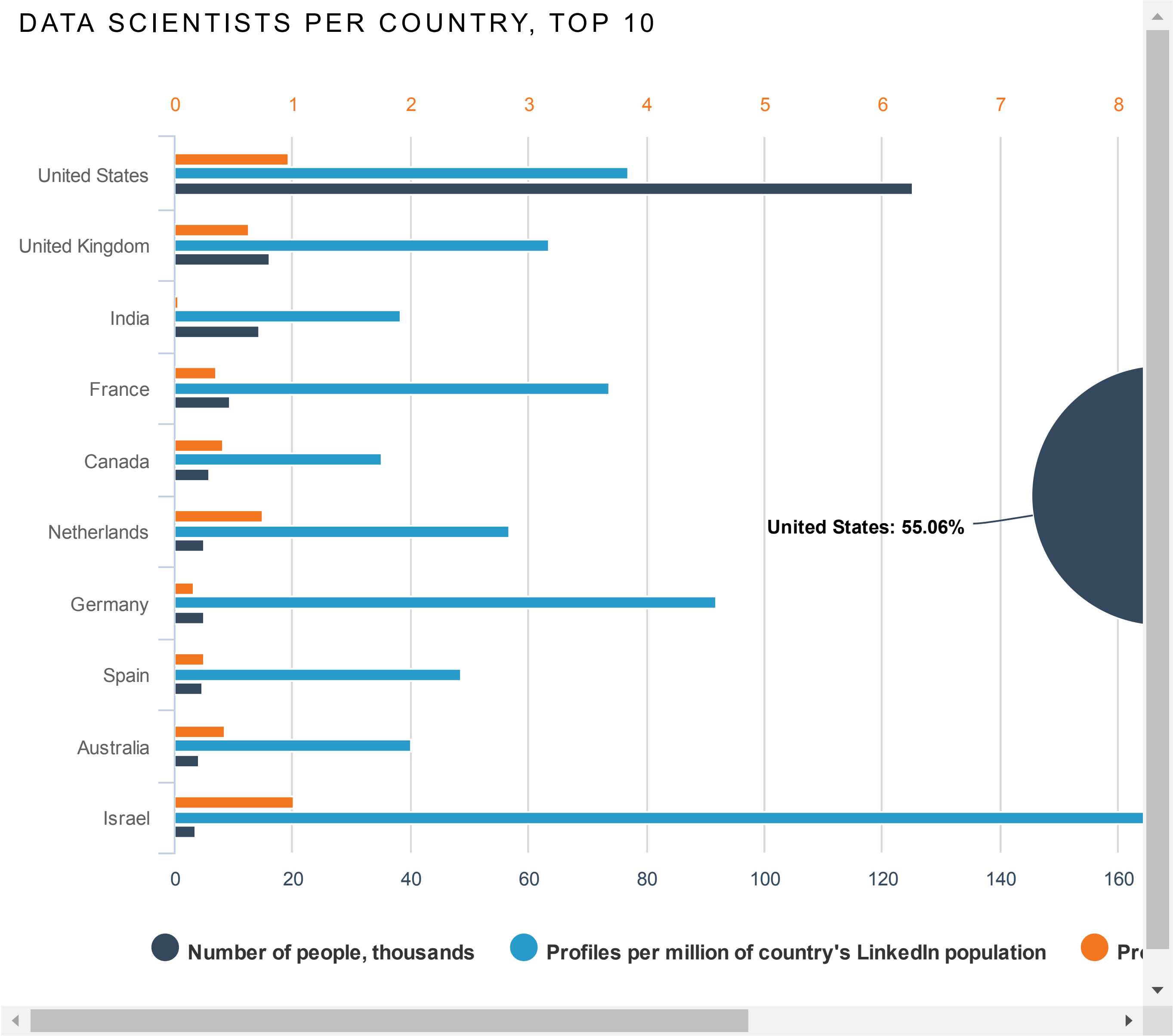


DATA SCIENTISTS WORLDWIDE, LOWER BOUND

There are of course many factors contributing to our estimates of the total number of data scientists in each country.

For example, LinkedIn adoption rates vary greatly on a country-by-country basis, and people in some countries do not use LinkedIn at all. However, given the education level of data scientists, and their propensity towards all things technological, we speculate that the vast majority of them are in fact on LinkedIn.

To show how both of these considerations affect the country ranking, we normalized th number of data scientists for each country by both the total LinkedIn membership and country's most recent census data. Both normalizations paint a very different picture.



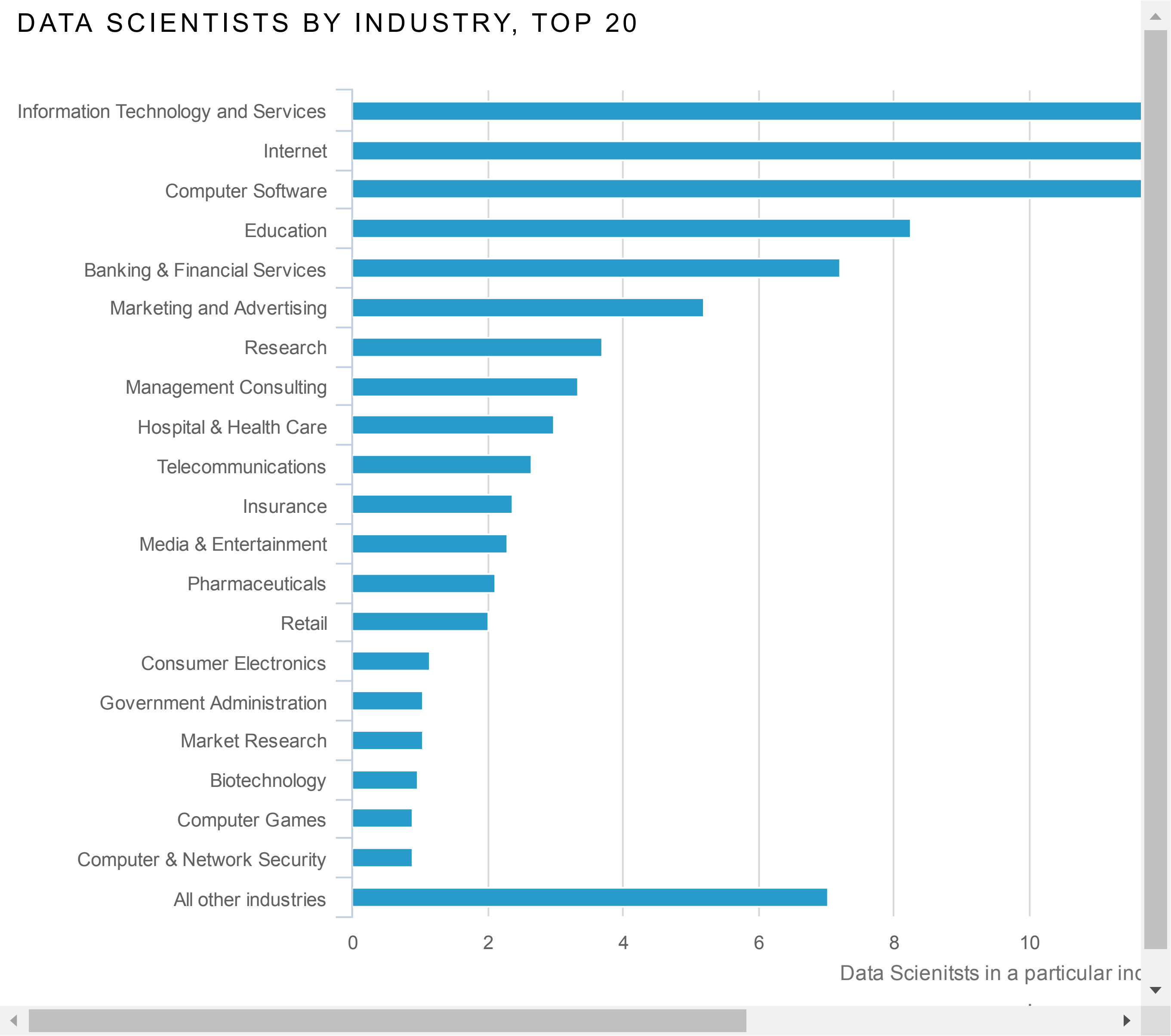
Specifically, LinkedIn adoption appears to be very poor in India, Israel, and Germany, yet India is high up on the list of the total number of data scientists.

At the same time, the "density" of data scientists, or the number of data scientists per unit of country's population, is the highest in Israel, followed by the United States and the Netherlands. While it is not surprising to see Israel, long known as the startup nation with Silicon Wadi as its own Silicon Valley, it is interesting to see such a high concentration of data scientists in the Netherlands.

Data Scientists at Work

What industries employ the largest number of data scientists?

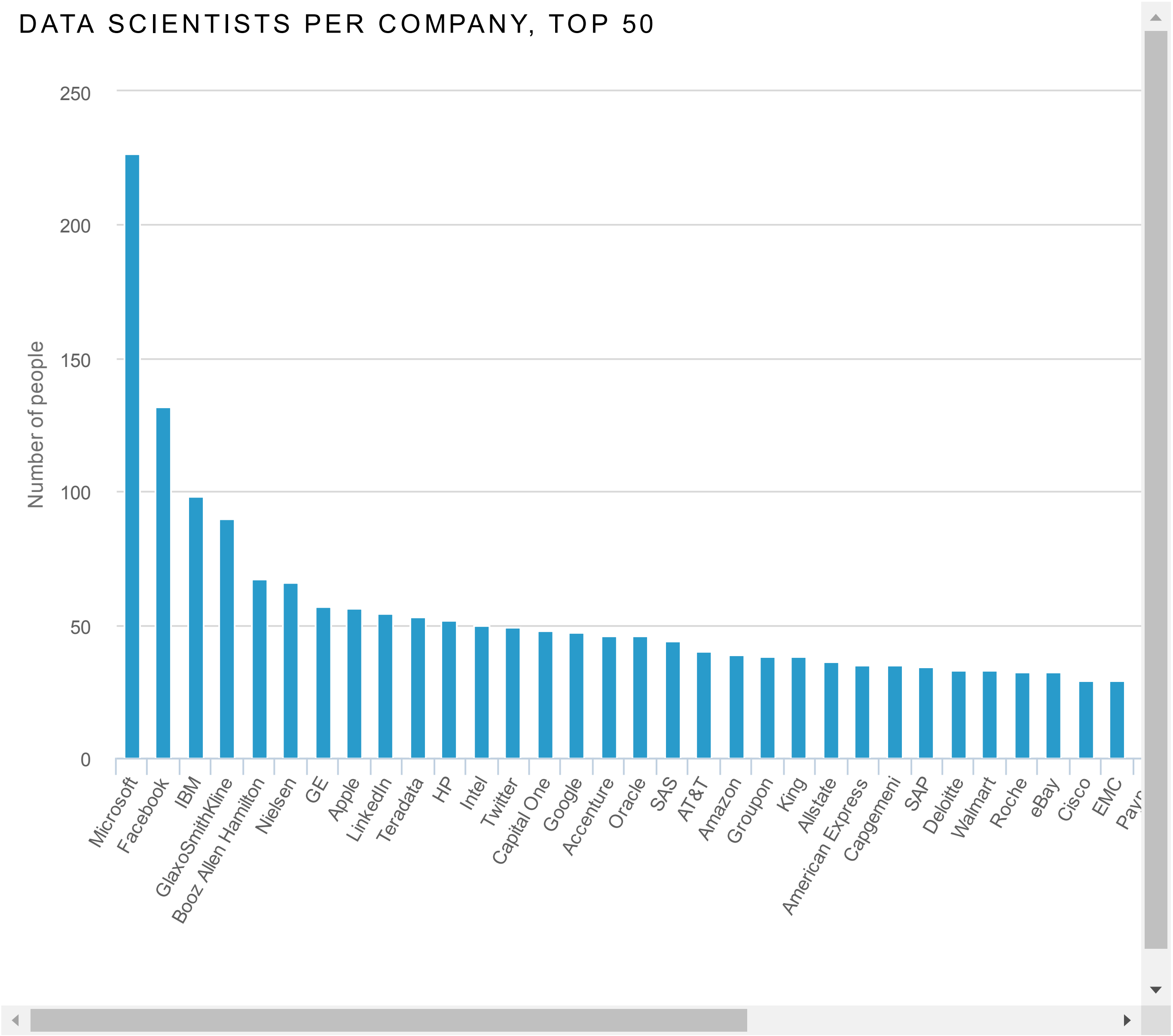
Today, the Information Technology and Services industry employs the largest number of data scientists, followed by Internet and Computer Software.



Both are noteworthy when examined in the context of Marc Andreessen's "software is eating the world" hypothesis. For example, both Airbnb and Uber are listed as Internet companies, yet Airbnb is disrupting the Hotel Industry, and Uber is taking on incumbents in both the Transportation and Shipping industries with approaches that, in large part, are fueled by data science. Traditional businesses will need to quickly adopt best data science practices or risk having data-driven competitors simply out-innovate them.

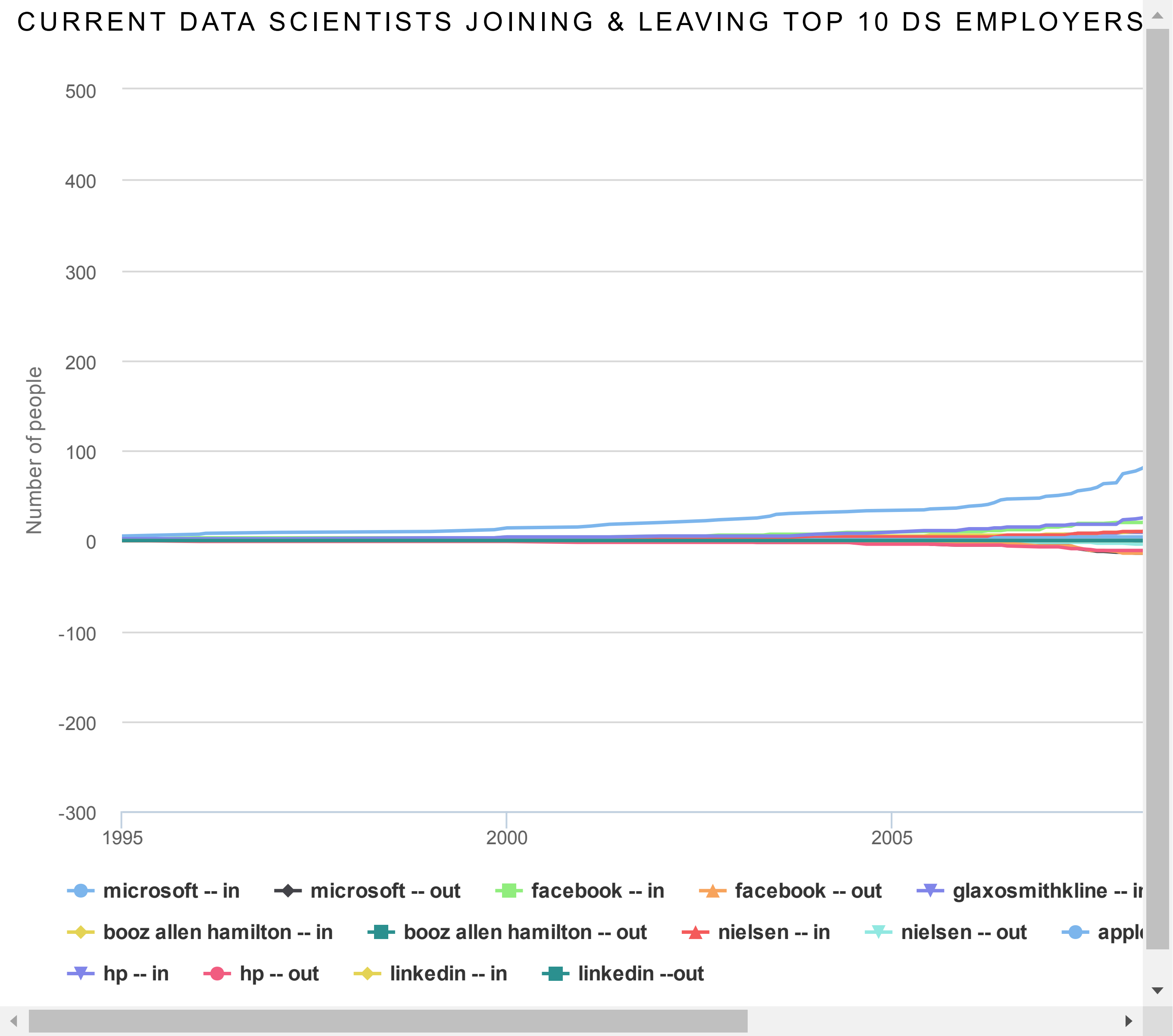
What companies employ the most data scientists?

There is a very healthy mix of new companies and more established businesses employing data scientists. Facebook, LinkedIn, and Twitter are high on the list, along with Apple, Microsoft, IBM, GlaxoSmithKline, and GE. Facebook is the only young tech company to crack the top five, second to Microsoft, which employs almost twice as many data scientists in total.


DATA SCIENTISTS PER COMPANY, TOP 50

It is important to note that these numbers account for all divisions within each company world-wide, so the figures should be considered with this context in mind.

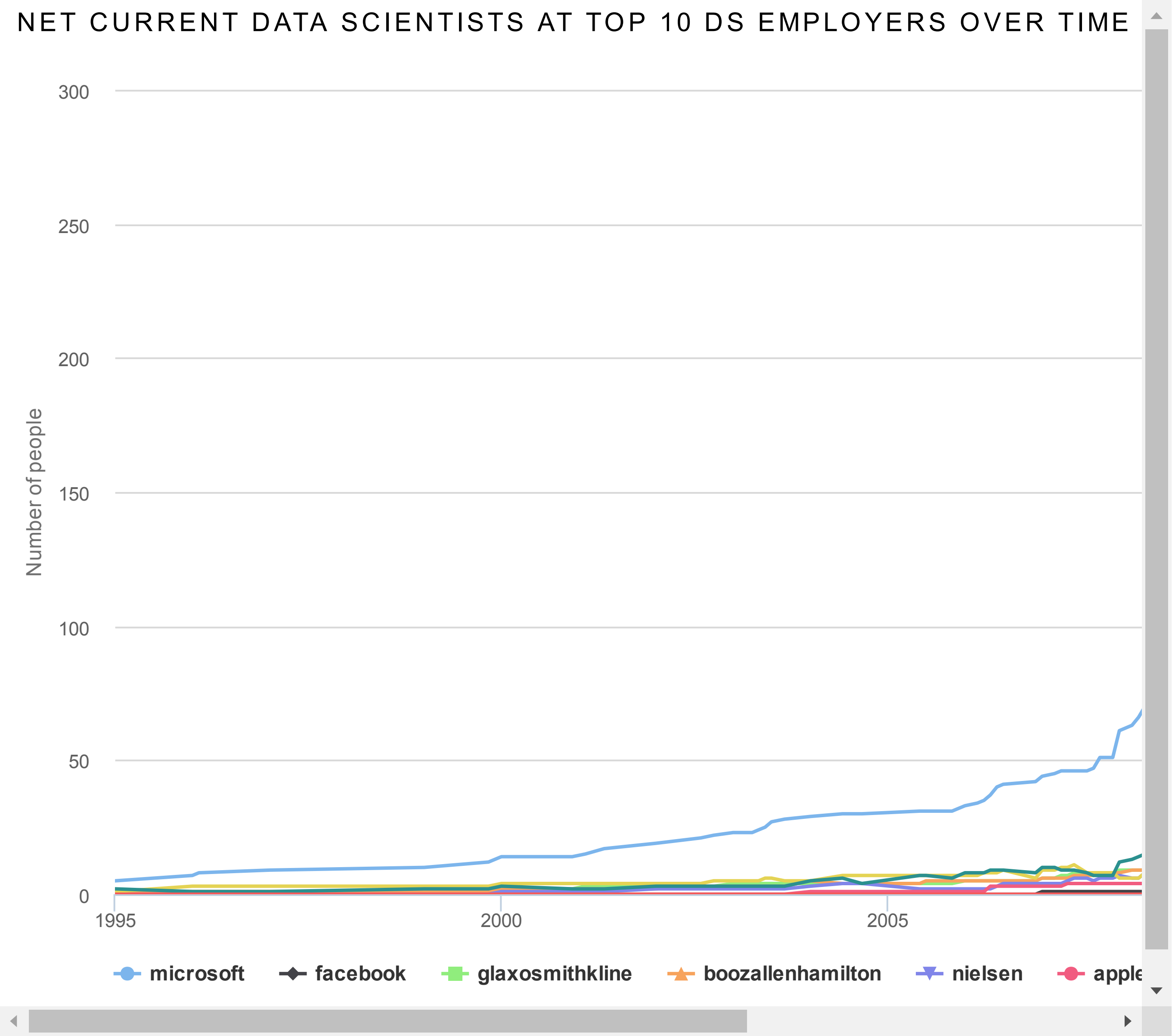
While the above chart is interesting, it is just a snapshot of the state of data science at company. Given how rapidly the field is evolving, we also wanted to see how businesse their data science teams. To do this, we used an approach similar to that described in our analysis of the number of data scientists over time. Specifically, we looked at the work experience of current data scientists to see when was the last time they joined and left any of the top 10 employers of data scientists. Since a data scientist's last employment with any of these companies may not have been in a data science role, the trends shown in this chart are underestimating the actual efforts of these companies to grow their data science teams.



Two companies stand out in particular: Microsoft and Facebook. Both Microsoft and Facebook appear to be on a hiring spree, accelerating their data scientist recruiting during the 2014 calendar year by at least 151% and 39%, respectively, when compared to 2013 (Microsoft went from at least 49 to 123 people hired, and Facebook from 43 to 60).

While Microsoft appears to be bringing the largest number of new people with data sci skills on board, it seems to be losing the largest number of data scientists as well. Note that our estimate of the number of people leaving each company excludes anybody who no longer identifies themselves as a data scientist on LinkedIn. Thus, the actual number of people with data science skills leaving each company is in reality larger.

Attrition and divisions aside, Microsoft still has an impressive lead both in the number of data scientists it has on staff and the pace of hiring.



Notably, Google, which is probably the largest designer of algorithms on the planet, does not appear on the top 10 list. This in no way suggests that Google is not doing data science. They simply do not call all of their data science practitioners data scientists.

The DNA of a Data Scientist

What are the primary skills of a data scientist?

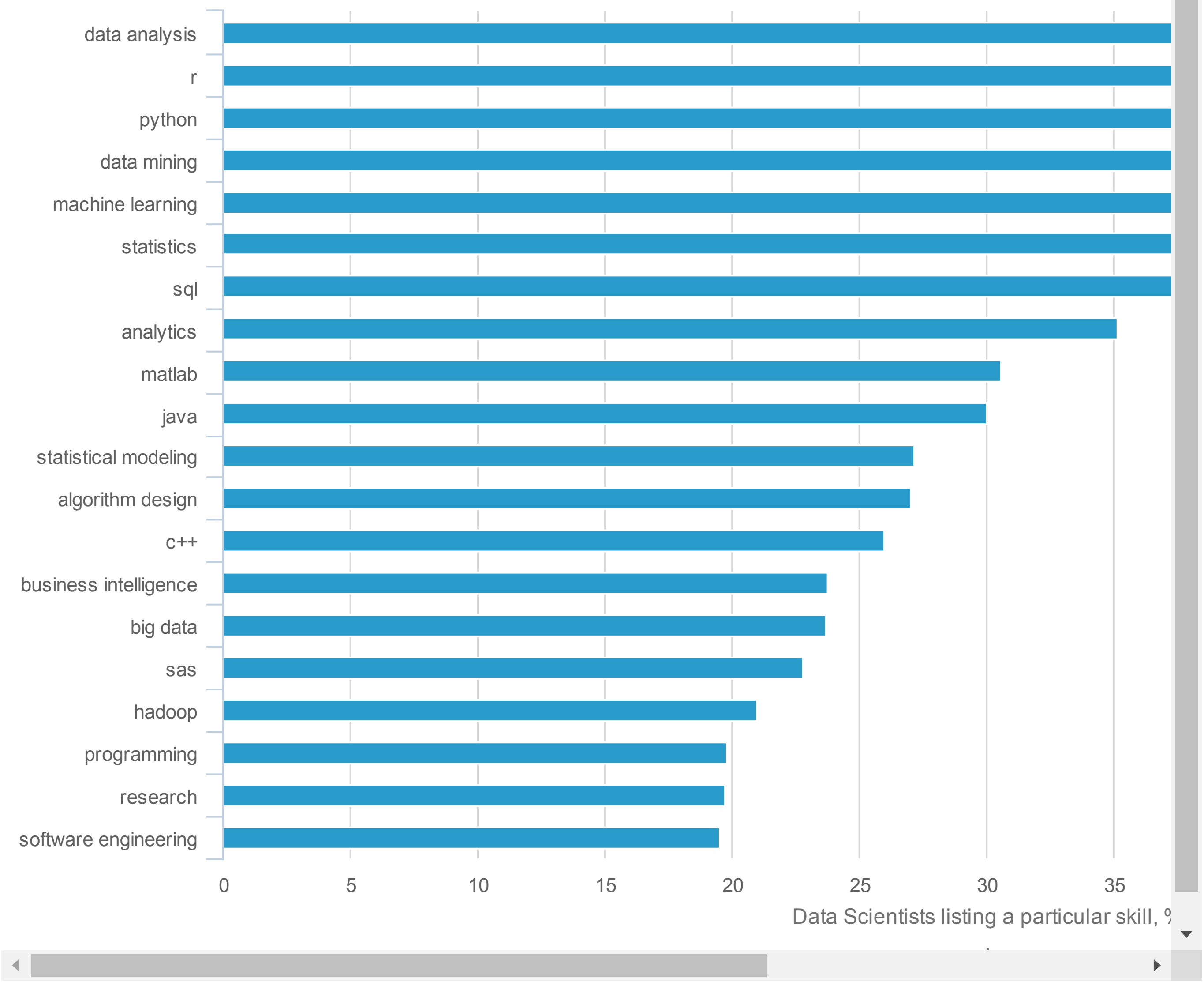
In today's world, a data scientist is expected to be a jack of all trades; a self-learner who has a solid quantitative foundation, an aptitude for programming, infinite intellectual curiosity, and great communication skills.

Instead of relying on personal and professional biases, we wanted to let the data speak for itself. We analyzed 254,000 skill records of self-identified data scientists and ranked each skill by the number of people listing it on their profile.

While "big data" and "hadoop" might still be buzzwords in some circles, they are not even in the top 10 actual skills employed by a garden-variety data scientist. Instead, generic "data analysis," R, Python, and machine learning lead the way, followed by statistics, SQL, analytics, MATLAB, and Java.

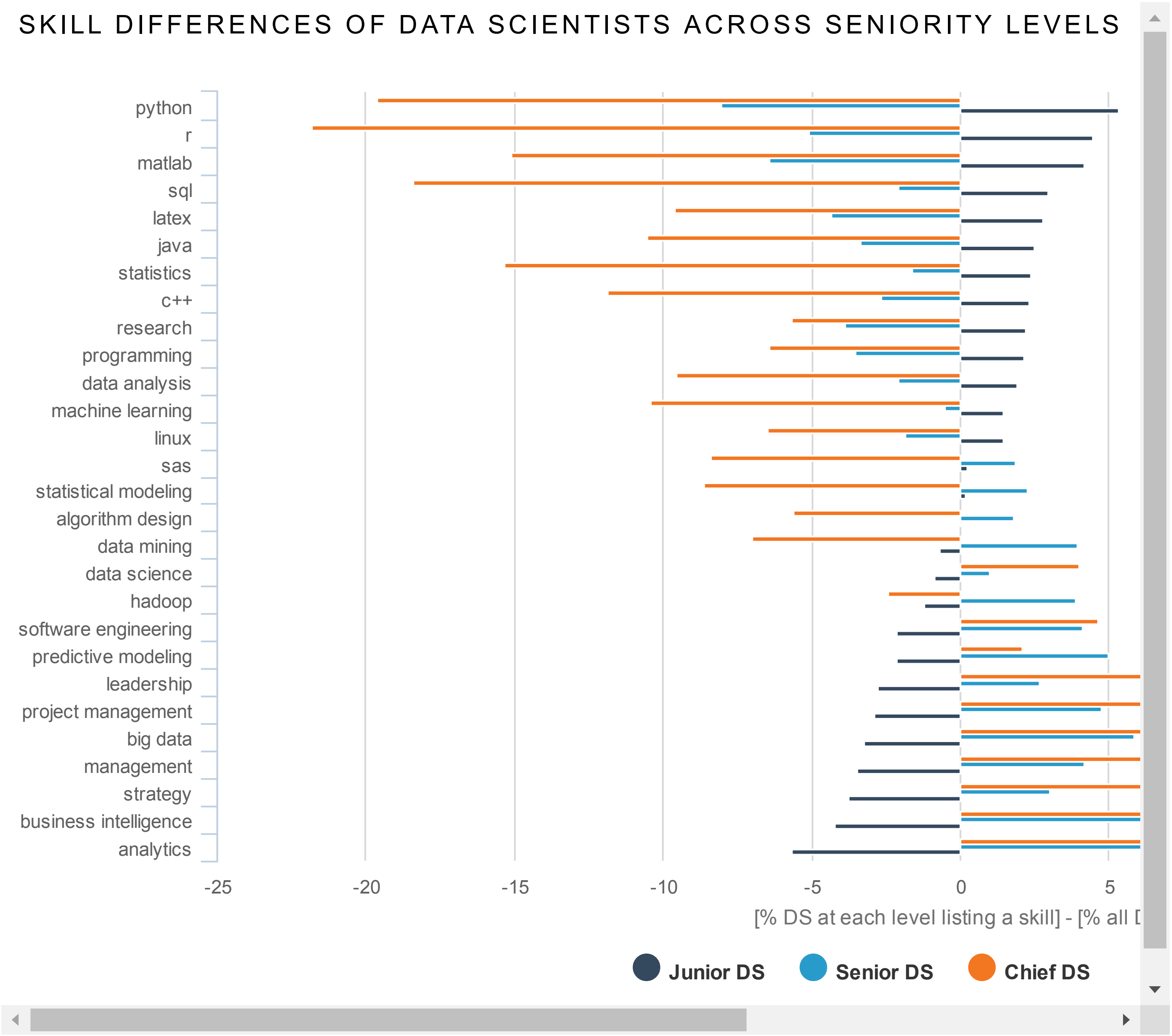
Note that data analytics differs from data analysis in that it is a broader term, generally implying an understanding of techniques and methods as opposed to just familiarity with tools for exploring and analyzing data (see this article).

TOP 20 SKILLS OF A DATA SCIENTIST



While this ranking does show the most prominent skills in the data science community as a whole, it averages out important hierarchical differences. We took a closer look at LinkedIn data and compared top skills across three different seniority levels: chief, senior, and junior.

The chief data scientist group included people in the C-suite, as well as founders, co-founders, owners and vice-presidents. The senior group consisted of directors of data science, managers, heads of data science, data science leads, principal and senior data scientists. Finally, the junior group included everybody not already captured by the chief and senior groups.


SKILL DIFFERENCES OF DATA SCIENTISTS ACROSS SENIORITY LEVELS

To highlight differences across seniority levels and make these differences easier to digest, we compared each level to the same common denominator: the average data scientist.

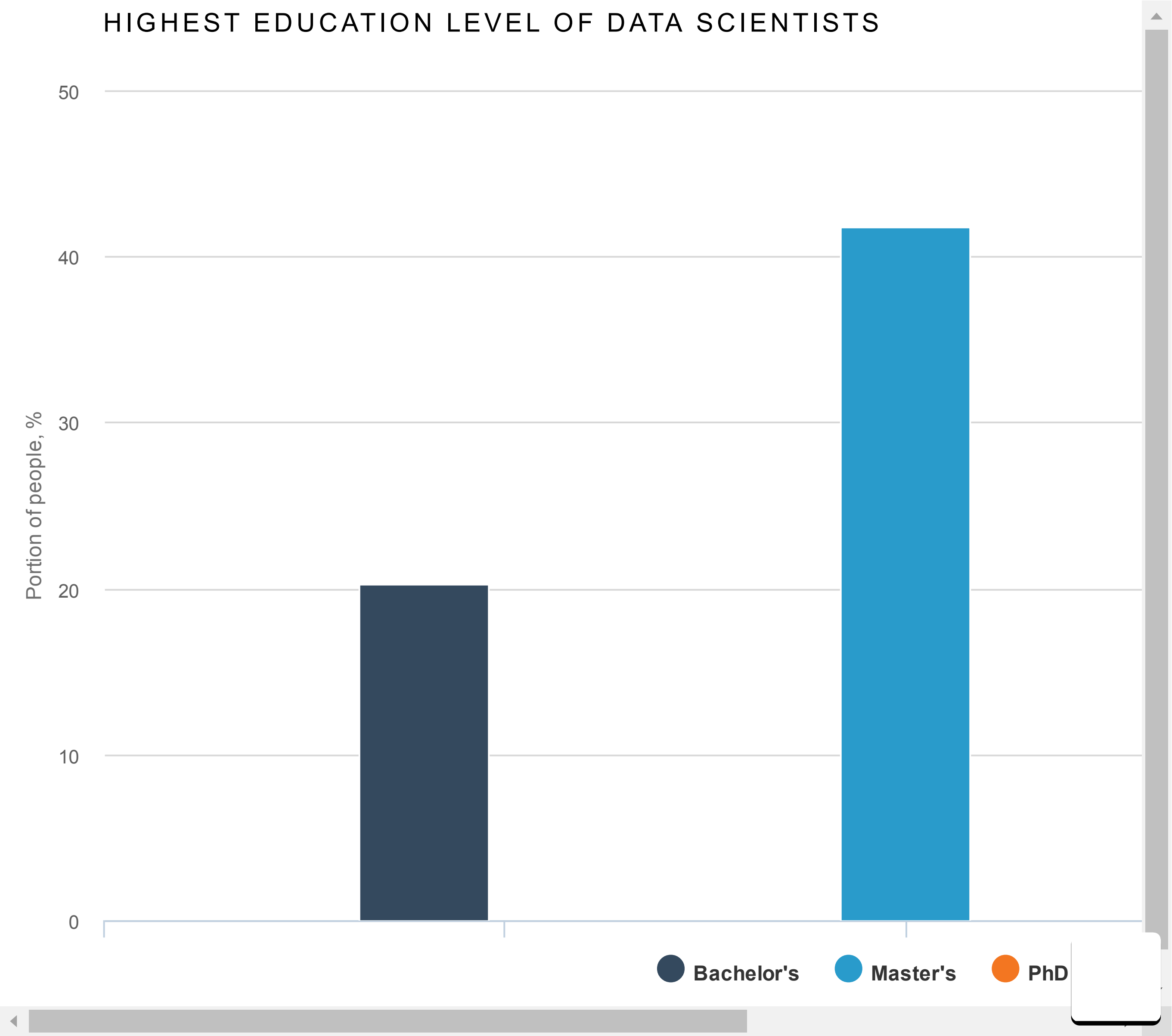
Interestingly, chief data scientists were significantly more likely to list business intelligence, analytics, leadership, strategy and management among their skills than both junior and senior data scientists. At the same time, today's chief data scientists appear to be less technical on average: only 27% and 26% listed Python and R, respectively. Compare this to the corresponding 52% and 53% of junior data scientists, along with 38% and 43% of senior practitioners.

While it is certainly true that chief data scientists may be simply emphasizing skills that are more relevant to their position within the company, we also speculate that many chief scientists assumed these roles by virtue of being in the field longer or having additiona qualifications, such as a business degree. Therefore, it is also possible that some chief data scientists never actually learned many of the skills listed by more junior people.

Similarly to chief data scientists, senior data scientists de-emphasized data analysis, and instead were more likely to emphasize data analytics when compared to junior data scientists: more than 45% of senior data scientists listed that skill vs. only 30% who did so at the junior level.

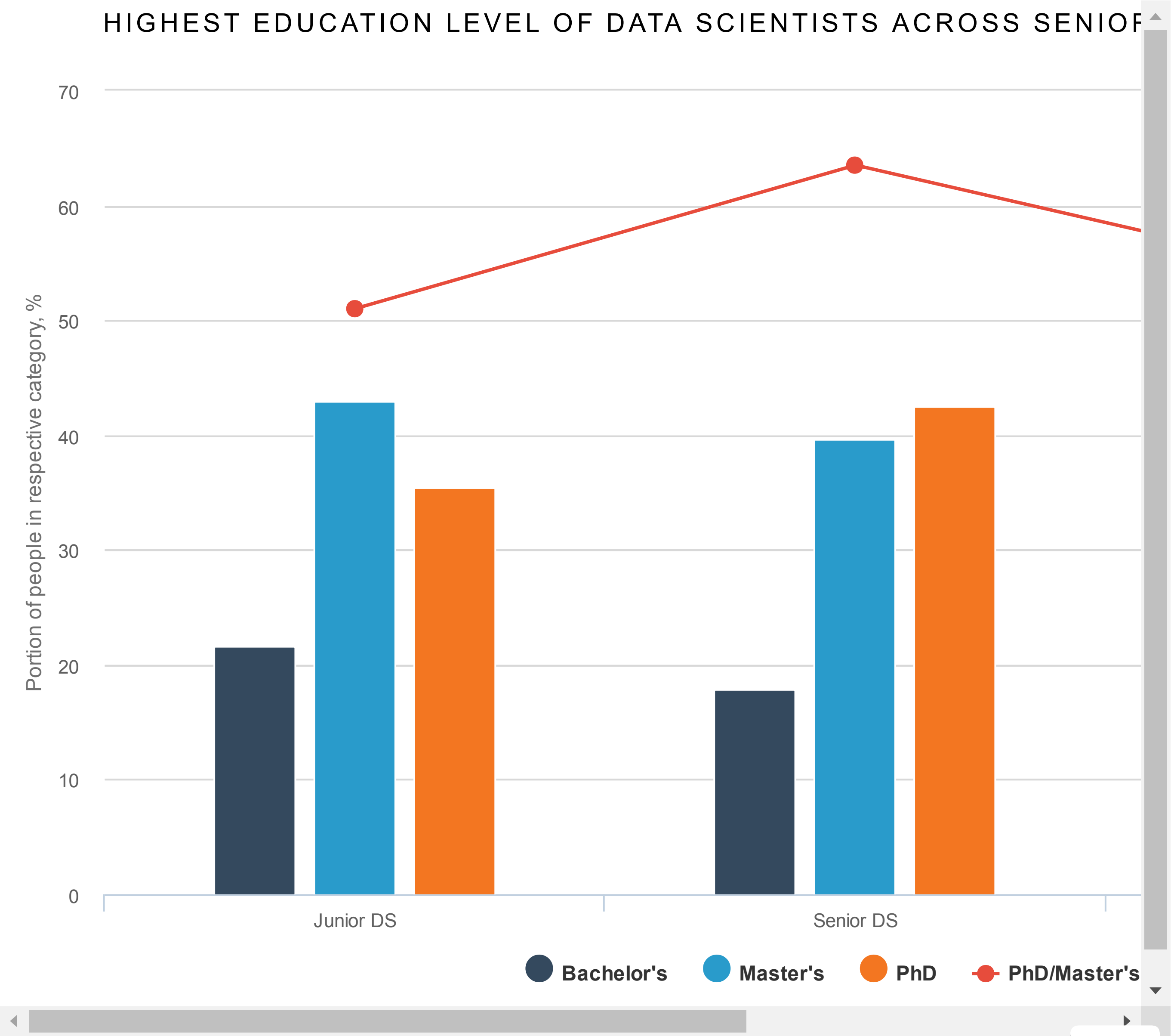
What is a data scientist's level of education?

We analyzed 27,000 education records to evaluate what percentage of data scientists hold advanced degrees and what fields of academic speciality they come from. This is shown as a percentage of all distinct bachelor's, master's, and doctorate degrees listed by data scientists (there are typically multiple degrees per person). 12% of all self-identified data scientists did not list any degrees.



Over 79% of data scientists listing their education have earned a graduate degree, with 38% of all data scientists who had an education record earning a PhD, and close to 42% listing a Master's degree as the highest degree attained. This shockingly large percentage of data scientists with graduate degrees is indicative of the increasing demand for specialists and a desire for advanced training in general. This trend is echoed by many of today's data science initiatives that build on research backgrounds of PhDs by helping them learn the tools and the technology stack most commonly used in the industry. This allows them to quickly get up to speed and become productive members of any data science team.

As with our analysis of skills, we saw significant differences in education across seniority levels.



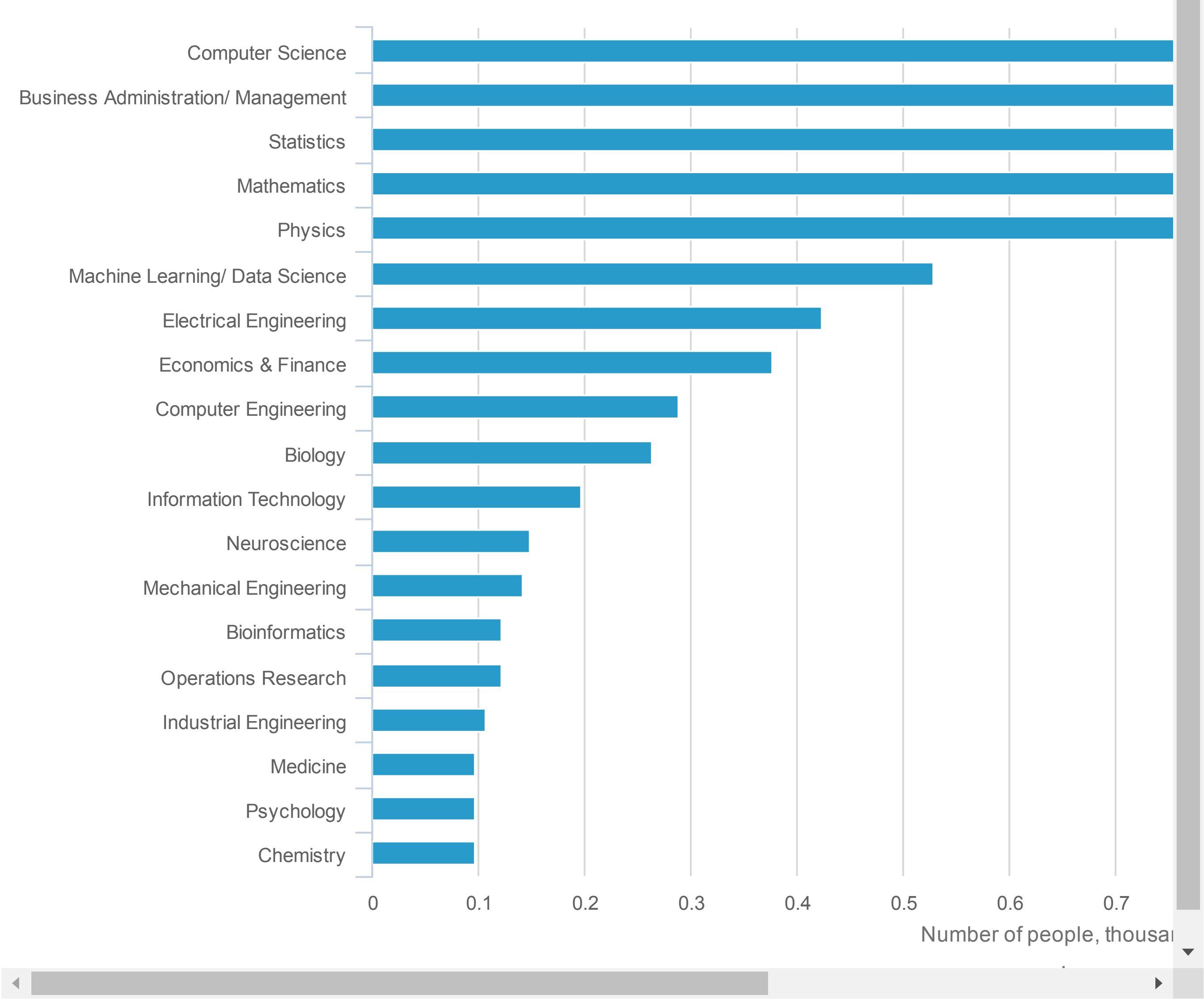
The ratio of data scientists with a PhD to data scientists with only a Master's degree is the highest at the senior level. In fact, it is almost 31% higher for senior data scientists when compared to junior data scientists. This indicates that in today's market, having a PhD helps data scientists climb the corporate ladder. We also noticed that fewer data scientists had a PhD at the chief data scientist level than at the senior level (35% vs 43%). Again, we speculate that this is largely due to the fact that people in more senior positions have been in the field longer and/or have other credentials that may be more relevant to their position.

What are the top academic backgrounds of data scientists?

Overall, Computer Science is the dominant field of study among data scientists. This supports what we found in our analysis of the skills listed, and what DJ Patil and Hilary Mason expressed in their book *Data Driven: Creating a Data Culture*. According to these two data science pioneers, "a data scientist who lacks the tools to get data from a database into an analysis package and back out again will become a second-class citizen in the technical organization."

That being said, we speculate that another reason Computer Science is so prominent is that many more people graduate with backgrounds in Computer Science than with backgrounds in Biology, Neuroscience, Bioinformatics or Psychology. Furthermore, today's Computer Science majors are arguably much more likely to work in the technology sector compared to any other graduate.

TOP 20 BACKGROUNDS OF DATA SCIENTISTS WITH A GRADUATE DEGRE



It is also interesting to note the differences between master's and doctoral degrees:

Top 10 Backgrounds for People with Master's and Doctoral Degrees

Master's

Rank	% of people	Field
1	12.86%	Computer Science
2	12.49%	Business Administration/ Management
3	10.98%	Statistics
4	10.20%	Mathematics
5	8.54%	Physics
6	5.25%	Machine Learning/ Data Science
7	4.50%	Electrical Engineering
8	4.21%	Economics & Finance
9	2.85%	Computer Engineering
10	2.48%	Biology
	16.54%	All other fields
	9.10%	Not provided

PhD

Rank	% of people	Field
1	14.74%	Physics
2	14.46%	Computer Science
3	10.83	Mathematics
4	8.24%	Statistics
5	4.77%	Electrical Engineering
6	4.08%	Biology
7	4.06%	Machine Learning/Data Science
8	3.25%	Computer Engineering
9	3.09%	Neuroscience
10	2.74%	Economics & Finance
	20.32%	All other fields
	9.41%	Not provided

First, some programs, such as programs in Business Administration are virtually not offered at the PhD level. This explains why Business Administration/Management does not appear in the list of top 10 PhD fields, but is ranked second on the list of Master's fields, with over 12% of data scientists listing an MBA as their highest and most recent level of education.

Second, graduates across fields face different job prospects upon graduation. For example, Physics majors arguably have fewer career options in both industry and academia than people with a computer science, electrical engineering, computer engineering or a statistics background. For that very reason, Physics majors historically have applied their experti other fields post-graduation. We speculate that this is the same reason they are likely overrepresented in data science.

That being said, there does appear to be a strong connection between data science and the mindset that Physics encourages. Kevin Novak, Head of Data Science Platform at Uber, notes that Physics helps “you become very good at understanding how to approximate, as well as when and why it’s appropriate.” Physics, and other disciplines like Biology, Neuroscience, and Electrical Engineering all involve experimentation, problem solving, and working with empirical data. Of course, empiricism inevitably gets messy and practical, encouraging exactly the mindset one needs to be a great data scientist.

Closing Thoughts from Lillian Pierson

Lillian Pierson
Data Mania

In 2012, Thomas H. Davenport and DJ Patil reported on the growing need for data scientists in a world where the amount of data is growing exponentially. In an article titled, Data Scientist: The Sexiest Job of the 21st Century, they likened the new data scientist to the “Wall Street ‘quants’ of the 1980’s and 1990’s”:

> In those days people with backgrounds in physics and math streamed to investment banks and hedge funds, where they could devise entirely new algorithms and data strategies. Then a variety of universities developed master’s programs in financial engineering, which churned out a second generation of talent that was more accessible to mainstream firms. The pattern was repeated later in the 1990s with search engineers, whose rarefied skills soon came to be taught in computer science programs.

Three years later, this report reveals the future that Davenport and Patil envisioned. Universities, bootcamps, and training programs have sprung up to bridge the skills gap. Simultaneously, organizations are clarifying and shaping the data scientist’s role. They are recognizing that the distributed tasks formerly carried out by a variety of roles can be most effectively executed when condensed under one title.

This report shows consolidation among the skills of data scientists, coupled with a growth in people with this title. I expect to see this upward trend continue as pioneers realize that the blend of machine learning, Python, and deep domain expertise that they have already mastered is actually data science, and as they inspire others to acquire these same skills.

This is exactly how my own career has played out. I spent years in technical roles in engineering and analytics, but there was limited opportunity for me to use the breadth skills in a single job. The growth in demand for data scientists led me to round out my own skillset and pursue a role within the field.

My personal belief is that the next four years of growth in the profession will come from those who work in adjacent fields and now just need to sharpen their skills. Now that I am a data scientist, I am excited to have the chance to help others who are just beginning this same journey.

Your data is scattered, and we can help.

Stitch is a simple, powerful ETL service built for developers. Stitch connects to all your data sources – from databases like MongoDB and MySQL, to SaaS tools like Salesforce and Zendesk – and replicates that data to your warehouse. With Stitch, developers can provision data to analysts and other team members in minutes, not weeks.

Type your email address

TRY STITCH TODAY →

Set up in minutes
Unlimited data volume during trial
5 million rows of data free, forever

About
Contact
Resources
Blog
Press
Changelog

Terms
Privacy
Status