

The Myth of the Data Scientist

The Importance of Teams in Providing Data Science Support

John Cocklin
Dartmouth College Library





“Data Scientist: Sexiest Job of the 21st Century”

Davenport and Patil, 2012

Skills

Data scientists are the people who understand how to fish out answers to important business questions from today’s tsunami of unstructured information. They have both statistical AND computational expertise.

Skills Gap

As companies rush to capitalize on the potential of big data, the largest constraint many face is the scarcity of this special talent.



THE WALL STREET JOURNAL.

“It’s Already Time to Kill the ‘Data Scientist’ Title”

Davenport, 2014

The data scientist term, which originally stood for quantitative experts who could also do a lot of the computational wizardry involved in analyzing unstructured Big Data, has come to mean almost anything.

I think the only answer is to be much more specific about the type of worker you want to be or hire. The specificity could include:

- The type of data that needs to be analyzed (and not just “big”).
- What the job is going to do with the data.



Google Trends

Interest over time 



data science



Shifting to Data Savvy: The Future of Data Science In Libraries

Barriers

- formal LIS education
- drive-by workshops – *“The typical three-hour to multi-day workshop is often not enough to yield substantive expertise in data savviness.”*
- information overload
- the brick wall - *“I learn new skills, but I still need to do my old job.”*
- Leadership
- incentive structures
- branding



Shifting to Data Savvy: The Future of Data Science In Libraries

Exemplars

- the Carpentry suite
- training programs such as Data Scientist Training for Librarians (DST4L)
- new trends such as reproducibility librarianship



Shifting to Data Savvy: The Future of Data Science In Libraries

Recommendations

- highlighting success stories
- collaborating with leadership institutes
- utilizing physical learning spaces
- advocating for the Carpentries
- leveraging existing educational resources
- repositioning the Masters in Library and Information Science (MLIS)



Shifting to Data Savvy: The Future of Data Science In Libraries

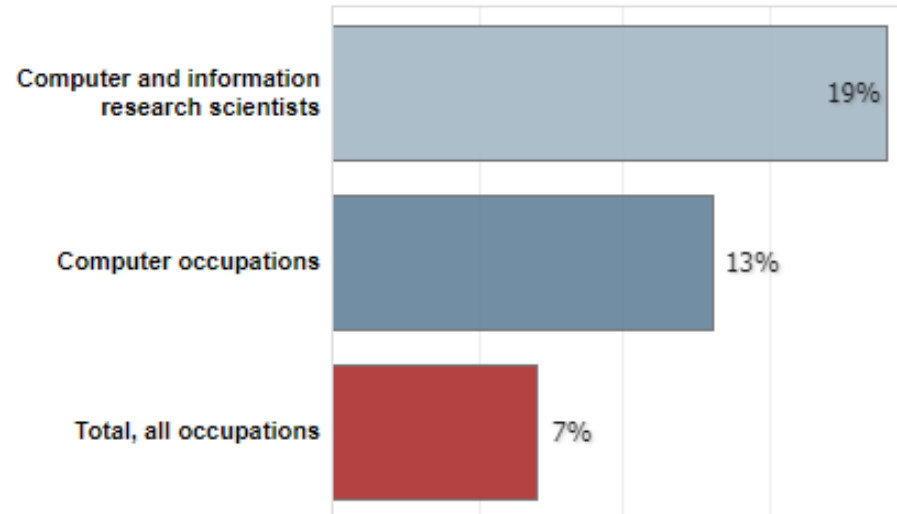
While librarians developing the same skills as scientists and researchers is not entirely realistic, librarians gaining exposure to these skills and developing a spectrum of skills, across a **team**, or even across the various business units at a university, is a more viable alternative.





Computer and Information Research Scientists

Percent change in employment, projected 2016-26

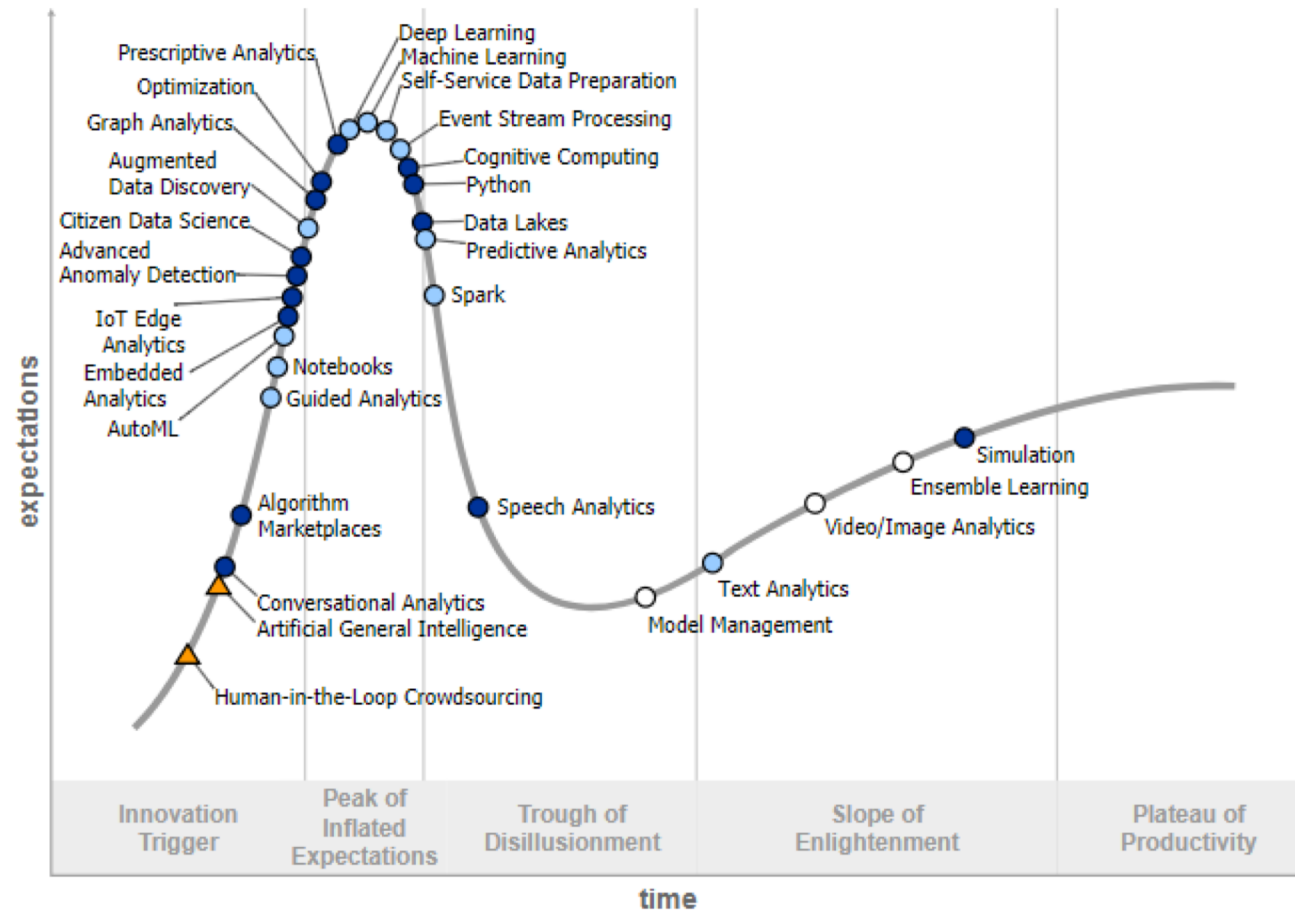


Note: All Occupations includes all occupations in the U.S. Economy.

Source: U.S. Bureau of Labor Statistics, Employment Projections program

“... because it is a small occupation, the fast growth will result in only about 5,400 new jobs over the 10-year period.” Some indications that demand will peak in 8-9 years.







Hype Cycle for Data Science and Machine Learning, 2017

- Citizen Data Science – Transformational, 5 to 10 years to mainstream adoption. An emerging set of capabilities and practices that allows users to extract advanced insights from data while not requiring them to be highly skilled. Central to citizen data science are rapidly progressing tools.
- Self-Service Data Preparation - Self-service data preparation tools enable users to reduce the time and complexity of interactively accessing, cataloging, harmonizing, transforming and modeling data for analytics in an agile manner with metadata and lineage support.
- R - Has reached the Plateau of Productivity, and for the first time this year is off the Hype Cycle.





Leading Upskilling Initiatives in Data Science and Machine Learning

Strategic Planning Assumptions

- By 2019, citizen data scientists will surpass highly skilled data scientists in terms of the amount of advanced analysis they produce.
- By 2020, more than 40% of data science tasks will be automated, resulting in increased productivity and broader usage by citizen data scientists.
- By 2020, predictive and prescriptive analytics will attract 40% of enterprises' net new investment in analytics and BI (Business Intelligence).





Leading Upskilling Initiatives in Data Science and Machine Learning

Recommendations

- Raise overall data science and machine-learning awareness and literacy by showcasing existing use cases and success stories.
- Identify citizen data science (CDS) candidates in your organization by creating an inventory of in-house skills.
- Match upskilling paths to the backgrounds and aspirations of CDS candidates.
- Design upskilling roadmaps for both the average consumer of analytics and your CDS candidates.





Staffing Data Science Teams

The Team

- Data Scientists
- Citizen Data Scientists
- Data Engineers
- Subject Experts
- Source System Experts
- Software Engineers

We also identified two variations of the data scientist role, as they are the source of many misconceptions.

- Quant Geeks - Excel in a specific range of quantitative skills. In certain situations they are a "nice-to-have," in rare situations a "must-have."
- Unicorns - Data scientists that are extremely well-versed in the whole range of skills — they are those "know-it-alls," that are romanticized every now and then in the literature. They are super rare.





Staffing Data Science Teams

Recommendations

- Have the appropriate cast of characters accessible and collaborating. Some can be a permanent part of the team, while some can be sourced in on an ad hoc basis according to the project's requirements.
- Understand that finding the right balance depends significantly on the actual and ongoing project portfolio. There are no established rules as yet.
- Avoid making just one job role (data scientists, for example) responsible for too many things. This will not only be a cause of frustrations, but is also unproductive.
- Build data science teams gradually from the bottom and middle by outsourcing, partnering, or hiring junior/mid-level data scientists. Starting a team from the top (with a chief data scientist) can be very risky.



Ideas

- Be specific about who people will be supporting, what people are going to be doing, and what skills, tools, and resources they will require to do it.
- Build teams rather than look for unicorns. Unicorns are “super rare.”
- Insure the team includes subject expertise along with statistical and computational expertise.
- Find and encourage the potential citizen data scientists in your organization.
- Develop an upskilling road map for each citizen data scientist.
- Purchase or develop AI and other software tools to aid citizen data scientists.
- Realize that the increase of data science in subject curriculums will have a double impact on support services.

