Welcome, Philip | My Account | Log Out

White Papers | Web Seminars | Newsletters | eBooks

- Big Data & Analytics
- Data Management
- MDM & Data Governance
- Infrastructure
- Info Strategy & Leadership
- BI & Data Discovery
- Mobility
- web seminars & white papers
- resource center

Blug

Data Science And Statistics: Colleges Must Evolve

by Steve Miller
JAN 7, 2015 4:32pm ET

🖶 Print
✉ Email
▤ Reprints
💬 Comment (1)
Twitter
LinkedIn
Facebook
Google+

Inquidia recently completed our seventh season of college recruiting, interviewing 30 candidates from three top universities, ultimately hiring two Spring 2015 grads. The newbies include a natural sciences major and a joint econ/computer science student.

We look for expertise in the scientific method, statistics/machine learning, data management, and computation – which we consider the foundations of both business intelligence and data science. Ideally, students will be at an intermediate undergraduate level of statistical knowledge that includes experimental design, regression and forecasting. Data skills include SQL and wrangling/munging with scripting languages like Python and Ruby. And computation can include programming expertise in Java/C++, Python/Ruby and R/SAS. The acquired expertise can either be course or self-taught. A capstone project or internship combining stats, data and programming is ideal.

I'm often asked what majors produce these candidates. The good and bad news: There seems to be no "perfect" curricula for Inquidia's needs. Inquidia hires computer science majors, engineers, science majors, business, math, econ, actuary and statistics students – as well as some humanities.

We generally find that specific majors aren't a perfect fit – and are often pleasantly surprised by "non-predictor" candidates who connect the academic/commerce dots in the interview process. Often it's the research and internship work that seals the deal. Computer science and math majors can be great but can also be overly theoretical, while students without pertinent work experience may not have the requisite data and computation skills. Several recent physical/natural science majors with strong research backgrounds have become capable apprentices.

The Statistics Challenge

A statistics major back in the day, I naturally gravitate towards the discipline when first looking at resumes. After all, exploratory data analysis and predictive analytics/modeling have statistical foundations. Alas, I'm pretty disappointed with most current statustics curricula, which seem to have changed little from my time, emphasizing high-end math to the detriment of computation, utilizing curated, well-behaved data to exercise the methods.

Indeed, I corroborate the observations of statistics PhD data scientist Rachel Schutt, who noted on starting her first post-graduation job at Google, "It was clear to me pretty quickly that the stuff I was working on at Google was different than anything I had learned in school when I got my PhD in statistics...But there were many skills I had to acquire on the job at Google...(I) picked up more computation, coding, and visualization skills, as well as domain expertise while at Google."

Schutt's Columbia advisor, eminent statistician Andrew Gelman, piles on in distinguishing academic statistics from commercial data science. "There's so much that goes on with data that is about computing, not statistics....statistics is the fun part and the part that, in this context, is new. The tech industry has always had to deal with databases and coding; that stuff is a necessity. The statistical part of data science is more of an option. To put it another way: you can do tech without statistics but you can't do it without coding and databases."

All's not gloom with undergraduate statistics, fortunately, as a recently-published paper, "Data Science in the Statistics Curricula: Preparing Students to 'Think with Data'", attests. The authors first embellish the frustrations reported by Inquidia, Schutt, and Gelman:

"A growing number of students are completing undergraduate degrees in statistics and entering the workforce as data analysts. In these positions, they are expected to understand how to utilize databases and other data warehouses, scrape data from Internet sources, program solutions to complex problems in multiple languages, and think algorithmically as well as statistically. These data science topics have not traditionally been a major component of undergraduate programs in statistics. Consequently, a curricular shift is needed to address additional learning outcomes....To be relevant and able to tackle the data-rich problems of today, as well as the ever-increasing data challenges for the future, our students' statistical problem solving skills must be enhanced to incorporate practical computational aspects.... computational thinking involve[s]... involved in formulating problems and their solutions so that the solutions are rep[resented]... effectively be carried out by an information-processing agent."

The paper continues: "As statisticians our first inclination is to focus on the infer[ence]...

computational aspects. However, as we encounter more and more complicated questions, data structures, and algorithms, the computational and algorithmic aspects become an integral part of arriving at a principled and statistically sound solution....(D)ata science is the computational aspect of carrying out a complete data analysis, including acquisition, management, and analysis of data (Johnstone & Roberts, 2014). We would also add data munging and manipulating, visualization, and statistical machine learning to the NSF list of data analysis tasks."

The curriculum prototypes proffered by the authors offer encouragement to those organizations seeking budding data scientists. All seven appear to have merit, but two courses/programs, one jointly between UC's Berkeley and Irvine, the other from Johns Hopkins University, are noteworthy.

UC's Approach

The California course, Concepts in Computing, is required of upper division students, and focuses "on the computational aspects of the data analysis cycle, from data acquisition and cleaning to data organization and analysis, and reporting.... To handle the data, students learn various tools and technologies including shell commands, regular expressions, structured query language (SQL) for relational databases, JavaScript for developing interactive Web pages, and R." The focus on computation emphasizes "generating random numbers and carrying out simulation studies (that) help students understand the concept of a random variable and its properties. Students use exploratory data analysis(EDA) to debug code; cross-validation and bootstrapping to assess models and variability; and presentation graphics to summarize findings from advanced statistical analyses." Tellingly, "Berkeley reduced the number of mathematics requirements from 5 to 4 in order to add this new requirement." Computation crowding out pure math.

Johns Hopkins' Courses

In contrast, Data Science Specialization, developed by Johns Hopkins professors Roger Peng, Jeff Leek, and Brian Caffo, expands the university curriculum the trio developed to a massive online open course (MOOC) platform in Coursera. The nine segments (each four weeks) cover the breadth of data science, from (1) The Data Scientist's Toolbox; (2) R Programming; (3) Getting and Cleaning Data; (4) Exploratory Data Analysis; (5) Reproducible Research (6) Statistical Inference; (7)Regression Models; (8) Practical Machine Learning; and (9) Developing Data Products.

Inquidia consultants routinely partake in the Johns Hopkins program and without exception claim it worthwhile. My nephew, a biochemistry PhD now self-training in Bioinformatics, has completed six of the segments and is a believer. And I know someone who recently completed an MS in Analytics program who swears the Data Science Specialization is a more rigorous curriculum.

Bottom Line

The long-overdue expansion of college statistics curricula into data and computation is nothing but good for the analytics world. If I were just beginning my college days, I can think of no major I'd prefer to data science-complected statistics. I'd encourage undergrads seeking entry into data-driven business to give it serious consideration.

Open Thoughts on Analytics

JOIN THE DISCUSSION

(1) [Comment](#)

SEE MORE IN

[Big Data/Analytics](#)

RELATED TAGS

[Big Data](),
[Analytics]()
Comments (1)
Glad to see you mention the "Data Science Specialization" track offered by Johns Hopkins, available via Coursera. This is quite helpful to people like me - mature, employed full time but also a 'lifelong learner' - the world and workplace are changing, we all need to change with it. Having previously taken the "Tackling the Challenges of Big Data" offered by MIT via their edX platform, I am thus far pleased to say the Data Science Specialization (a series of 9 courses + capstone project) is far more pragmatic and hands-on.

Thanks for your thoughtful insights into a ever evolving area.
Posted by Gary B | Thursday, January 08 2015 at 10:27AM ET

Add Your Comments:

Add your comments here.

☐ Notify me when other readers comment on this article.
Click here to receive notifications without commenting

POST ▶

CANCEL

Videos

[60 Seconds Smarter: E&M Guidelines Explained](#)

Most Read
Most Emailed

- [10 Internet of Things (IoT) Trends & Considerations](#)
- [Data Acceleration: Turning Technology Into Solutions](#)
- [10 Tech Trends Through 2015](#)
- [Next Shift: From Big Data to Deep Data](#)
- [Data Science And Statistics: Colleges Must Evolve](#)

Featured Topics

- [BIG DATA](#)
- [BUSINESS INTELLIGENCE](#)
- [DATA INTEGRATION](#)
- [DATA MGMT](#)
- [DATA MODELING](#)
- [DATA QUALITY](#)
- [FINANCIAL SERVICES](#)
- [GOVERNANCE, RISK & COMPLIANCE](#)
- [HEALTH CARE](#)

[Slideshows](#)

[10 Big Data Career Killers](#)

[Gartner's Top 10 Cloud Myths](#)

[Gartner's Top 10 Strategic Tech Trends for 2015](#)

[8 Principles for Adopting Casework for Data Governance](#)

[Debunking 5 of the Biggest Data Myths](#)

[Five Habits of Data-Driven People](#)

[More Slide Shows »](#)

Current Buzz on Analytics

Twitter

Facebook

LinkedIn

- [about 2 days ago](#) we were checking out Get Ready for BI Change [bit.ly/14MbHk3](#)
- [about 2 days ago](#) we were checking out Cloud vs. Enterprise CRM Software Price War Coming [bit.ly/1xzw0Ik](#)
- [about 2 days ago](#) we were checking out Obama Touts Well-Wired Iowa Town as Model for City-Run Web [bit.ly/14LVone](#)
- [about 3 days ago](#) we were checking out Obama Cybersecurity Plan Could Change Healthcare

Processes [bit.ly/159Q0uL](bit.ly/159Q0uL)

- about 3 days ago we were checking out Enterprise Marketing's Mobile Mind Shift bit.ly/14FVxsd

- about 4 days ago we were checking out Cloud Security Concerns Shift From IT to Boardroom bit.ly/1FWbSun

- about 5 days ago we were checking out Large Enterprises Embrace Disaster Recovery as a Service (DRaaS) [bit.ly/1ya8LtE](bit.ly/1ya8LtE)

- about 6 days ago we were checking out Data Acceleration: Turning Technology Into Solutions bit.ly/14SDAXX

- about 6 days ago we were checking out Next User Experience Battleground: Mobile Customer Experience [bit.ly/1FNn4t4](bit.ly/1FNn4t4)

- about 9 days ago we were checking out CNA Using Big Data to Fight Insurance Fraud [bit.ly/1y0PSI3](bit.ly/1y0PSI3)

- about 10 days ago we were checking out Five Social Data Marketing Resolutions bit.ly/1FvHZ3M

- about 10 days ago we were checking out Personal Healthcare Data: Patients Will Take Control bit.ly/1AMBkhj

- about 11 days ago we were checking out Next Shift: From Big Data to Deep Data [bit.ly/1xT9YUz](bit.ly/1xT9YUz)

- about 11 days ago we were checking out Global Tech Market Looking Better For 2015 [bit.ly/143GKY1](bit.ly/143GKY1)

- about 11 days ago we were checking out Data Science And Statistics: Colleges Must Evolve bit.ly/1FpdFrs

- about 11 days ago we were checking out Flash Storage In the Data Center Catching On [bit.ly/14qb1B4](bit.ly/14qb1B4)

- about 11 days ago we were checking out CNA Using Big Data to Fight Insurance Fraud [bit.ly/14pO6Wl](bit.ly/14pO6Wl)

- about 12 days ago we were checking out Patient Monitoring Market: $5 Billion by 2020 bit.ly/1FgARbm

- [about 13 days ago](#) we were checking out Why Once-Successful Companies Fail [bit.ly/1vYZoHn](#)
- [about 13 days ago](#) we were checking out Average Mobile App Earns "D" Grade From Customers [bit.ly/1vTsrMv](#)

[Follow @infomgmt] [9,942 followers]

Analytics

- [Hail to the Chiefs! Meet the CDO, CAO and CSO](#)
- [IBM: New Mainframe Boasts Improved Ability to Perform Analytics](#)
- [Data Acceleration: Turning Technology Into Solutions](#)
- [CNA Using Big Data to Fight Insurance Fraud](#)

Business Intelligence

- [Get Ready for BI Change](#)
- [UPS ORION Advanced Analytics Case Study](#)
- [Business Intelligence Analysts as Architects](#)
- [The Three C's of Self-Service BI](#)

Customer Experience

- [Insurance Executives See IT Role In Customer Satisfaction](#)
- [Cloud Security Concerns Shift From IT to Boardroom](#)
- [Next User Experience Battleground: Mobile Customer Experience](#)
- [Average Mobile App Earns "D" Grade From Customers](#)

Open Source

- [CNA Using Big Data to Fight Insurance Fraud](#)
- [Hadoop Isn't for Everyone](#)
- [China Plans Foreign Technology Purge, Shift to Homegrown Suppliers](#)
- [8 Hadoop Big Data Predictions for 2015](#)

Predictive Analytics

- [Know More, Sell More With Data Science in 2015](#)
- [Internet of People, Not Things, Is Next Frontier](#)
- [The MIT Analytics Imperative](#)
- [Google's Big Data to Help Auction.com Predict Homebuying Trends](#)

Data Governance

- [CIO Priorities: Strategic Architecture vs Quick Project Wins](#)

- Obamacare Dental Heathcare Data Doesn't Add Up
- Target Wants Data Breach Bank Claims Dismissed
- Creating the Data Governance Killer App

Data Integration

- Big Data Requires Integration Technology
- Actuate, Oracle, Red Hat, Others Announce New or Updated Products
- How Healthcare CIOs View the Data Analytics Landscape
- Headache Number Two: Remedies for Data Migration Pain

Data Management

- Large Enterprises Embrace Disaster Recovery as a Service (DRaaS)
- Personal Healthcare Data: Patients Will Take Control
- Why Once-Successful Companies Fail
- Top Healthcare Technology Stories of 2014

- HOME
- About Us
- Contact Us
- Content Licensing
- Advertise with Us
- Customer Service
- Feedback
- My Account
- Site Map
- Privacy Policy
- Editorial Submissions

sourcemedia
corporate site

banking

- American Banker
- Bank Technology News
- American Banker Magazine
- Credit Union Journal

MORTGAGES

- National Mortgage News

PAYMENTS

- PaymentsSource
- Collections & Credit Risk

- ISO & Agent

capital markets

- Mergers & Acquisitions
- Asset Securitization Report
- Leveraged Finance News
- Private Placement Letter
- Traders Magazine
- Securities Technology Monitor

MUNICIPAL FINANCE

- The Bond Buyer

accounting

- Accounting Today
- Tax Pro Today

HEALTHCARE & BENEFITS

- Employee Benefit News
- Employee Benefit Adviser
- Health Data Management
- Insurance Networking News
- Information Management

INVESTMENT ADVISORY

- Financial Planning
- On Wall Street
- Bank Investment Consultant
- Money Management Executive
- Investment Mgt. Mandate Pipeline