For Today’s Graduate, Just One Word: Statistics

By STEVE LOHR

MOUNTAIN VIEW, Calif. — At Harvard, Carrie Grimes majored in anthropology and archaeology and ventured to places like Honduras, where she studied Mayan settlement patterns by mapping where artifacts were found. But she was drawn to what she calls “all the computer and math stuff” that was part of the job.

“People think of field archaeology as Indiana Jones, but much of what you really do is data analysis,” she said.

Now Ms. Grimes does a different kind of digging. She works at Google, where she uses statistical analysis of mounds of data to come up with ways to improve its search engine.

Ms. Grimes is an Internet-age statistician, one of many who are changing the image of the profession as a place for dromish number nerds. They are finding themselves increasingly in demand — and even cool.

“I keep saying that the sexy job in the next 10 years will be statisticians,” said Hal Varian, chief economist at Google. “And I’m not kidding.”
What Are the Odds That Stats Would Be This Popular?

By QUENTIN HARDY

“Most of my life I went to parties and heard a little groan when people heard what I did,” says Robert Tibshirani, a statistics professor at Stanford University. “Now they’re all excited to meet me.”

It’s not because of a new after-shave. Arcane statistical analysis, the business of making sense of our growing data mountains, has become high tech’s hottest calling. There are billions of bytes generated daily, not just from the Internet but also from sciences like genetics and astronomy. Companies like Google and Facebook, as well as product marketers, risk analysts, spies, natural philosophers and gamblers are all scouring the info, desperate to find a new angle on what makes us and the world tick. Computing has become cheap and available enough to process any number of formulas.

What no one has are enough people to figure out the valuable patterns that lie inside the data.
In the year ending Monday, we saw some gains in statistical savvy: Data-crunching pollsters accurately forecast the outcome of the presidential election; the Memphis Grizzlies hired a vice president of basketball operations for his statistical expertise; and folks grew comfortable with the phrase "big data," to describe the billions of billions of bytes generated daily by information technology.

The growing importance of statistical analysis is set to be a theme of next year, too, with more than 150 professional organizations worldwide, including the American Statistical Association, designating 2013 as the International Year of Statistics.
Data Crunchers Now the Cool Kids on Campus

On campuses, at cocktail parties and in American corporations, statisticians are walking a bit taller these days.

The explosive growth in data available to businesses and researchers has brought a surge in demand for people able to interpret and apply the vast new swaths of information, from the analysis of high-resolution medical images to improving the results of Internet search engines.

Schools have rushed to keep pace, offering college-level courses to high-school students, while colleges are teaching intro stats in packed lecture halls and expanding statistics departments when the budget allows.

Number of Students (thousands)

Year
the signal and the noise and the noise and the noise and the noise and the noise and the noise and the noise

why so many predictions fail—but some don’t

nate silver
"Niet meer het buisvak van vroeger"

DIEPENBEEK - In de jaren '80 waren beurshandelaars sexy. In de jaren '90 waren dat de computer nerds en tegenwoordig zijn dat de statistici. Zo zien ze dat toch in de VS waar statstici Nate Silver de presidentsverkiezingen van elke staat correct voorspelde. Vrijdag heeft de Universiteit Hasselt de internationale dag van de statistiek ingezaagd met Marie Davidson, voorzitter van de American Statistical Association.

Statistiek staat belang als een bijvlek, om vak werknemers interessante werk te bieden. Het kan statistiek als een dilemna zien van ontwikkeling. Dat het uiteindelijk al te zwaar zou worden voor weinig mensen die dat nu nog doen. Vroeger nam dat precies zo, nu staat dat precies andersom. Volgens professor Guert Molenberghs, directeur van het Internationaal Instituut voor Statistiek en statistische Bio-informatica (IBIS) van de Universiteit Hasselt en de "Natuuristentag op de Carpetten", is statistiek gezien als noodzakelijk. "Het is statistiek nodig als noodzakelijk" voor Marie Davidson: "Het is statistiek nodig om de dingen te verklaren. We hebben dat nodig om te begrijpen hoe we kunnen ontwikkelen." Volgens professor Guert Molenberghs, "Als we het gebruiken in de praktijk, dan hebben we een betere voorspelling van de toekomst. We kunnen dan beter beslissingen nemen."

Het belang van statistiek voor de economie is groot. "Statistiek is de sleutel voor het begrijpen van de economie". Bijvoorbeeld, hoe zullen we de volgende periode houden? "Als we dat weten, kunnen we beter beslissingen nemen." Volgens professor Guert Molenberghs, "Statistiek is de sleutel voor het begrijpen van de economie." Volgens professor Guert Molenberghs, "Als we dat weten, kunnen we beter beslissingen nemen." Volgens professor Guert Molenberghs, "Statistiek is de sleutel voor het begrijpen van de economie." Volgens professor Guert Molenberghs, "Als we dat weten, kunnen we beter beslissingen nemen."

Ambtenaren om hun prestaties te bewaken

BRUGGE - De voedselcontrole uitgevoerd door de BRDG, de belastinginメリchedering van de NVB, de belastinginメリchedering van de NVB, de belastinginメリchedering van de NVB, de belastinginメリchedering van de NVB,

16 procent minder asielaanvragen

BRUGGE - Het aantal asielaanvragen in ons land is in 2017 met 16,6 procent geminderd vergeleken met het jaar ervoor. Dit is mede te danken aan het feit dat de situatie in enkele oorsprongstaten beter is geworden. De meerderheid van de asielaanvragen komt uit het Midden-Oosten en Noord-Afrika. Het aantal asielaanvragen is daarmee aanzienlijk afgenomen. Het aantal asielaanvragen is daarmee aanzienlijk afgenomen. Het aantal asielaanvragen is daarmee aanzienlijk afgenomen. Het aantal asielaanvragen is daarmee aanzienlijk afgenomen.

Heel veel innoveren bedrijven


Criminaliteit

Niet alleen de politie maar ook de bewoners merken een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit. "De criminaliteit is in de voormalige verenigde staten van Amerika gedaald, en die trend is ook vooral in Europa merkbaar. Nu is er een daling in de criminaliteit.
David Alexander’s job didn’t exist ten years ago. He works for Pacific Biosciences in Menlo Park, California, writing software that can analyse the data generated by DNA polymerase enzymes, which sequence DNA in real time. A decade ago, it took scientists weeks to sequence DNA, one base at a time, using a seemingly endless series of reactions. Back then, they also thought that they would be able to find the causes of major diseases just by identifying the common genetic variants shared by affected individuals.

Both the technology and the hypotheses have changed greatly since then. In the mid- to late 2000s, while Alexander was working towards his PhD, scientists were using genome-wide association studies (GWAS)—searching genomes for known genetic variants that are shared by people with a particular disease or trait. But by the time he graduated, last time, GWAS had mostly been superseded by techniques that sequence entire genomes. The machines designed to do this sequencing are pouring out huge amounts of data, thereby creating a huge need for mathematicians and statistics experts. So Alexander, and many others working on statistical genetics, now have many more opportunities. Scientifically, there are much richer questions to ask, and there are still a lot of deep discoveries to be made. It’s an interesting time, he says. His career track reveals just how much opportunities in the field have changed.

CAREER VARIATION

It was not for a lack of trying that GWAS didn’t pan out. The completion of the Human Genome Project in 2003 spared major funders from around the world to invest millions of dollars to build an international haplotype map, a catalogue of all the common human variants at single bases, called single nucleotide polymorphisms (SNPs), to be used in GWAS. The SNP map should have helped researchers to identify genes that are associated with disease. But instead, it showed that SNPs don’t account for much of the heritability of disease.

Researchers now think that many rare variants play a part in causing disease, but rare variants are much harder to find than the common SNPs. As a result, statistical geneticists are now mining sequence data for directly...
BIG DATA
A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK

VIKTOR MAYER-SCHÖNBERGER
KENNETH CUKIER
What is data scientist

About data scientists
Rising alongside the relatively new technology of big data is the new job title data scientist. While not tied exclusively to big data projects, the data scientist role does complement them because of the increased breadth and depth of data being examined, as compared to traditional roles.

So what does a data scientist do?
A data scientist represents an evolution from the business or data analyst role. The formal training is similar, with a solid foundation typically in computer science and applications, modeling, statistics, analytics and math. What sets the data scientist apart is strong business acumen, coupled with the ability to communicate findings to both business and IT leaders in a way that can influence how an organization approaches a business challenge. Good data scientists will not just address business problems, they will pick the right problems that have the most value to the organization.

The data scientist role has been described as “part analyst, part artist.” Anjul Bhambri, vice president of big data products at IBM, says, “A data scientist is somebody who is inquisitive, who can stare at data and spot trends. It’s almost like a Renaissance individual who really wants to learn and bring change to an organization.”

Whereas a traditional data analyst may look only at data from a single source – a CRM system, for example – a data scientist will most likely explore and examine data from multiple disparate sources. The data scientist will sift through all incoming data with the goal of discovering a previously hidden insight, which in turn can provide a competitive advantage or address a pressing business problem. A data scientist does not simply collect and report on data, but also looks at it from many angles, determines what it means, then recommends ways to apply the data.
Data Science Is Multidisciplinary

Source: Brendan Tierney, Oralytics Blog
Frontiers in Massive Data Analysis

Committee on the Analysis of Massive Data
Committee on Applied and Theoretical Statistics
Board on Mathematical Sciences and Their Applications
Division on Engineering and Physical Sciences
NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

PREPUBLICATION DRAFT – Subject to Further Editorial Correction
Data Science: The Numbers of Our Lives

By CLAIRE CAIN MILLER

HARVARD BUSINESS REVIEW calls data science “the sexiest job in the 21st century,” and by most accounts this hot new field promises to revolutionize industries from business to government, health care to academia.

The field has been spawned by the enormous amounts of data that modern technologies create — be it the online behavior of Facebook users, tissue samples of cancer patients, purchasing habits of grocery shoppers or crime statistics of cities. Data scientists are the magicians of the Big Data era. They crunch the data, use mathematical models to analyze it and create narratives or visualizations to explain it, then suggest how to use the information to make decisions.

In the last few years, dozens of programs under a variety of names have sprung up in response to the excitement about Big Data, not to mention the six-figure salaries for some recent graduates.

In the fall, Columbia will offer new master’s and certificate programs heavy on data. The University of San Francisco will soon graduate its charter class of students with a master's in analytics. Other institutions teaching data science include New York University, Stanford, Northwestern, George Mason, Syracuse, University of California at Irvine and Indiana University.
Data Science: The End of Statistics?

As I see newspapers and blogs filled with talk of “Data Science” and “Big Data” I find myself filled with a mixture of optimism and dread. Optimism, because it means statistics is finally a sexy field. Dread, because statistics is being left on the sidelines.

The very fact that people can talk about data science without even realizing there is a field already devoted to the analysis of data — a field called statistics — is alarming, I like what Karl Broman says:

*When physicists do mathematics, they don’t say they’re doing “number science”. They’re doing math.*

*If you’re analyzing data, you’re doing statistics. You can call it data science or informatics or analytics or whatever, but it’s still statistics.*

Well put.

Maybe I am just pessimistic and am just imagining that statistics is getting left out. Perhaps, but I don’t think so. It’s my impression that the attention and resources are going mainly to Computer Science. Not that I have anything against CS of course, but it is a tragedy if Statistics gets left out of this data revolution.

“The very fact that people can talk about data science without even realizing there is a field already devoted to the analysis of data – a field called statistics – is alarming”

— Larry Wasserman
FOR IMMEDIATE RELEASE
March 29, 2012

OBAMA ADMINISTRATION UNVEILS “BIG DATA” INITIATIVE:
ANNOUNCES $200 MILLION IN NEW R&D INVESTMENTS

Aiming to make the most of the fast-growing volume of digital data, the Obama Administration today announced a “Big Data Research and Development Initiative.” By improving our ability to extract knowledge and insights from large and complex collections of digital data, the initiative promises to help solve some the Nation’s most pressing challenges.

To launch the initiative, six Federal departments and agencies today announced more than $200 million in new commitments that, together, promise to greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data.
The mission of the NIH Big Data to Knowledge (BD2K) initiative is to enable biomedical scientists to capitalize more fully on the Big Data being generated by those research communities. With advances in technologies, these investigators are increasingly generating and using large, complex, and diverse datasets. Consequently, the biomedical research enterprise is increasingly becoming data-intensive and data-driven. However, the ability of researchers to locate, analyze, and use Big Data (and more generally all biomedical and behavioral data) is often limited for reasons related to access to relevant software and tools, expertise, and other factors. BD2K aims to develop the new approaches, standards, methods, tools, software, and competencies that will enhance the use of biomedical Big Data by supporting research, implementation, and training in data science and other relevant fields that will lead to:  

Read more
“My impression is that many scientists view statistics not so much as a science but as a ‘bag of tools.’” – Alan Leshner, CEO, AAAS
2011 McKinsey Global Institute Report

**Big Data: The Next Frontier for Innovation, Competition, and Productivity**

“A significant constraint will be a shortage of... people with deep expertise in statistics and data mining...a talent gap of 140K – 190K positions in 2018.”
statistics
“We’re spending $70 per person to fill this out. That’s just not cost effective, especially since in the end this is not a scientific survey. It’s a random survey.” – Rep. Daniel Webster, The New York Times, May 19, 2012.
Promoting the Practice and Profession of Statistics
ASA Strategic Plan Themes

• The ASA as “The Big Tent for Statistics”

• Increasing the Visibility of the Profession
Steve Pierson, ASA Director of Science Policy
Why Statistics?

Popular media and science publications sound the drum: “BIG DATA” WILL DRIVE OUR FUTURE, from translating genomic information into new therapies, to harnessing the Web to untangle complex social interactions, to detecting infectious disease outbreaks. Statistics is the science of learning from data, and of measuring, controlling, and communicating uncertainty; and it thereby provides the navigation essential for controlling the course of scientific and societal advances. This field will become ever more critical as academia, businesses, and governments rely increasingly on data-driven decisions, expanding the demand for statistics expertise.

The melding of science and statistics has often propelled major breakthroughs. Last year’s Nobel Prize in Physics was awarded for the discovery of the accelerating expansion of the universe. That discovery was facilitated by sophisticated statistical methods, establishing that the finding was not an artifact of imprecise measurement or miscalculations. Statistical methods also allowed the trial demonstrating that zidovudine reduces the risk of HIV transmission from infected pregnant women to their infants to be stopped early, benefitting countless children. Statistical principles have been the foundation for field trials that have improved agricultural quality and for the randomized clinical trial, the gold standard for comparing treatments and the backbone of the drug regulatory system.

Statistics often informs policy development. For example, in the United States, billions of dollars are allocated to school districts based on county-specific estimates of income and poverty, derived by combining data using statistical methods. In evaluating pollutants, statistical modeling isolates true associations with illnesses and deaths. Big data payoffs can be enormous, but there are many pitfalls. Take the promise of personalized medicine: Achieving this goal will require the integration of vast landscapes of genomic, clinical, and related data from legions of patients. The potential for false discovery looms large.

New statistical methods will be needed to address some of these issues. Similar challenges arise from the haystacks of information on social network, time-use, economic, and other activities that can be mined to benefit science, business, and society. Close collaboration with statisticians is the best way to ensure that critical issues are identified and solutions found.

A dramatic increase in the number of statisticians is required to fill the nation’s needs for expertise in data science. A 2011 report by a private consulting firm projected a necessary increase of nearly 200,000 professionals (a 50% increase) by 2018.* Graduates specializing in statistics are equipped with skills that allow them to pursue diverse careers, and there has been a surge in applications for graduate education in these fields. But available places are limited; for example, the ratio of qualified applicants to slots in the Johns Hopkins Biostatistics program exceeds 10 to 1. Resources must be found to expand the number and size of graduate programs.

No amount of statistical intervention can circumvent flawed subject-matter models or salvage valid conclusions from poorly designed studies, and even sound statistical analysis may fail to yield straightforward answers. The future demands that scientists, policy-makers, and the public be able to interpret increasingly complex information and recognize both the benefits and pitfalls of statistical analysis. It is a good sign that the new U.S. Common Core K-12 Mathematics Standards† introduce statistics as a key component in precollege education, requiring that students be skilled in describing data, developing statistical models, making inferences, and evaluating the consequences of decisions. Embedding statistics in science and society will pave the route to a data-informed future, and statisticians must lead this charge.

— Marie Davidian and Thomas A. Louis

10.1126/science.1231645

*www.mckinsey.com/insights/…/Research_Technology_and_Innovation_Big_Data_The_Next_Frontier_for_Innovation;
†www.corestandards.org

Marie Davidian is William Neil Reynolds Professor of Statistics at North Carolina State University, president-elect of the American Statistical Association, and executive editor of Biometrics. E-mail: davidian@ncsu.edu.

Thomas A. Louis is professor of Biostatistics at the Johns Hopkins Bloomberg School of Public Health, retiring chair of the ASA Statistics Section OG, former president of the International Biometric Society, and former editor of Journal of the American Statistical Association and Biometrics. E-mail: toluis@jhsph.edu.
Jeff Myers, ASA Public Relations Coordinator
2013: The International Year of... Statistics

Posted: 02/12/2013 1:18 pm

It's a scene we statisticians know all too well. The passenger next to us asks, "So, what do you do?" "I'm a statistician," we reply, bracing for impact. "That was the worst course I ever took," our seatmate exclaims. "I had to take it for my major, but I didn't understand any of it." It's a recurrent narrative we've experienced for decades.

But times have changed. From Google chief economist Hal Varian's well-circulated 2009 quote in The New York Times that "the sexy job in the next 10 years will be statisticians" to the assertions of numerous blog posts and articles, statistics is "hot."

And the dead-on predictions of the 2012 presidential election by not only statistician Nate Silver, but also Huffington Post's Mark Blumenthal and Simon Jackson, political scientist Drew Linzer, and others, have captured the attention of the public and media about the power of statistics like never before. Now, the reactions of our seatmates have transformed from groans into eager curiosity, due in large part to this watershed event.

What is statistics? To many, statistics is the class they took in college or figures on the sports pages. But statistics is so much more. Statistics is the science of learning from data and measuring, controlling, and communicating uncertainty. Statisticians do this by developing models to describe data. These models help us design methods to collect data, draw conclusions from data, and characterize the uncertainty in the findings.
Statistical Thinking: The Bedrock of Data Science

By Joel B. Greenhouse

Thanks to Google Chief Economist Hal Varian's 2009 prediction that "the sexy job in the next 10 years will be statisticians," it is now OK to self-identify as a statistician. This calls for some explanation. For many, their first experience with statistics was anything but pleasant, so telling someone you are a statistician has often been a conversation stopper -- or worse. Recently, a taxi driver told me, somewhat accusingly, that his wife had had to change her major because she was failing her required statistics class. I was sympathetic, knowing that all too often the first statistics class is divorced from real-world applications and emphasizes methods but not how to apply them or interpret their results -- or even tell why one would care. That's a far cry from sexy. What's changed?
The International Year of Statistics (Statistics2013)

What is Statistics?

When many people hear the word “statistics,” they think of either sports-related numbers or the college class they took and barely passed. While statistics can be thought about in these terms, there is more to the relationship between you and statistics than you probably imagine.

Read More →

New Research on Why More Maths and Stats Are Needed for Jobs

JULY 12, 2013

New research ‘The Employment Equation: Why our young people need more Maths for today’s jobs’ commissioned by the Sutton Trust, makes it clear that having a good grade in GCSE Mathematics doesn’t mean you have the practical mathematical skills needed in today’s workplace.

Read More →

getstats is a campaign of the Royal Statistical Society.

www.statistics2013.org
2011: The International Year of Chemistry
2011: The International Year of Chemistry
2011: The International Year of Chemistry
2013: The International Year of Statistics

- ASA
- Institute of Mathematical Statistics (IMS)
- International Biometric Society (IBS)
- International Statistical Institute (ISI) and Bernoulli Society
- Royal Statistical Society (RSS)
2013: The International Year of Statistics

- ASA
- Institute of Mathematical Statistics (IMS)
- International Biometric Society (IBS)
- International Statistical Institute (ISI) and Bernoulli Society
- Royal Statistical Society (RSS)
S. RES. 150

To designate the year 2013 as the “International Year of Statistics”.

IN THE SENATE OF THE UNITED STATES

MAY 21, 2013

Mrs. HAGAN submitted the following resolution; which was referred to the Committee on the Judiciary

RESOLUTION

To designate the year 2013 as the “International Year of Statistics”.
The ASA and Big Data

Nathaniel Schenker, Marie Davidian, and Robert Rodriguez

This month’s column is a team effort. President-elect Nat Schenker and Past President Bob Rodriguez join me in announcing a strategic initiative for the ASA.

As Bob discussed in his June 2012 column (http://bit.ly/16q1inA), Big Data is a Big Topic. It is almost impossible to avoid the daily barrage of media accounts (http://on.w3.org/zawR8D), conference announcements (www.aawite.org/conferences/bigdata/2013), and events such as the recent Big Data Week (http://bigdatawork.com/) focused on Big Data. Last year, President Obama announced a major Big Data research and development initiative (www.whitehouse.gov/blog/2012/03/29/big-data-big-deal) and, last month, the White House hosted a Big Data workshop. The National Institutes of Health created the position of associate director for data science (www.nih.gov/news/health/Jan2013/Mod-104.cfm), and a new book—Big Data: A Revolution That Will Transform How We Live, Work, and Think—(http://n.pr/154wUac), which explores the explosion of digital information, has received extensive press coverage.

Big Data are data on a massive scale in terms of volume, intensity, and complexity, and their promise for transforming business, health care, scientific discovery, public policy, and a host of other areas has been proclaimed widely. But, despite the enormous potential for contributions by statisticians, our profession and the ASA have not been very involved in Big Data activities. We are often missing from Big Data discussions in the media.

There are three reasons for this disconnect. First, the media and public lack a general understanding of what statisticians contribute to society (the issue that motivated the International Year of Statistics, www.statistic2013.org). Second, few statisticians are engaged in Big Data projects or have the special skills necessary to handle Big Data challenges.

Third, the statistical community is disconnected from the new (and vaguely defined) community of data scientists, who are completely identified with Big Data in the eyes of the media and policymakers. Data science (http://nyti.ms/2jRFb0) is frequently described as an amalgam of computer science, mathematics, data visualization, machine learning, distributed data management—and statistics. Data scientists must be innovative modelers and programmers; they also must be exceptional communicators who have a deep understanding of the problem domain and can formulate key questions, uncover novel insights, and use this information to guide high-impact decision making. Other disciplines have been quick to identify themselves with data science and are routinely featured in media accounts. Although statistics is mentioned in passing, statisticians are nearly invisible.

Ideally, statistics and statisticians should be the leaders of the Big Data and data science movement. Realistically, we must take a different view. While our discipline is certainly central to any data analysis context, the scope of Big Data and data science goes far beyond our traditional activities. As Bob noted in his column, the sheer scale and velocity of the data being
Promote the importance of statistics
• To your collaborators and leaders
• To the public and your local media
• Within the AAAS (join!)

Seize the opportunity presented by the positive stories in the media
INTRODUCTION TO DATA SCIENCE, COLUMBIA UNIVERSITY

Blog to document and reflect on Columbia Data Science Class

RACHEL SCHUJT

Dr. Rachel Schutt is a Senior Research Scientist at Johnson Research Labs. Prior to that, she was a Senior Statistician at Google Research in the New York office. She is also an Adjunct Assistant Professor in Columbia’s Statistics Department, and is a founding member of the Education Committee for the Institute for Data Sciences and Engineering at Columbia. Rachel is co-authoring a book (with Cathy O’Neil) called “Doing Data Science” to be published by O’Reilly in 2019.

Her interests include statistical modeling, exploratory data analysis, machine learning algorithms, and social networks, as well as the ethical dimensions of Data Science, and using Data Science to do good.
Promote the importance of statistics
• To your collaborators and leaders
• To the public and your local media
• Within the AAAS (join!)
• To businesses and research organizations with data challenges
• To colleagues in “data science” disciplines

Engage with those in other fields doing data-driven work
Data Science Meetup Groups

Related topics: Big Data Analytics, Data Analytics, Data Visualization, Predictive Analytics, Data Mining, Big Data, Machine Learning, Business Intelligence, Hadoop, R Users Group

Largest Meetup Groups

1. **SF Data Mining**
   - **San Francisco, CA**
   - The SF Data Mining meetup focuses on all aspects of the data pipeline—from data acquisition and big data storage to machine learning and data visualization.
   - **4,093** Data Scientists
   - **Meetup topics:** Data Mining, Data Science, Big Data, Machine Learning, Data Analytics, Data Visualization, Artificial Intelligence, A Project for Statistical Computing, Applied Statistics, Programming, Show more...

2. **IGTCloud**
   - **Tel Aviv-Yafo, Israel**
   - The IGTCloud – the Israeli Cloud Organization, is a non-profit organization of leading industry companies, startups, ISVs, customers, vendors, VCs and academia, focused on innovation, knowledge sharing and networking for developing Cloud computing/SA...
   - **4,075** Members
   - **Meetup topics:** Big Data, Data Science, Data Analytics, Cloud Computing, Web Technology, Technology User Groups, Cyber Security, OpenStack, Machine Learning, Internet of Things, Show more...

3. **NYC Predictive Analytics**
   - **New York, NY**
   - NYC Predictive Analytics is a non-profit professional group that meets monthly to discuss diverse topics in predictive analytics and applied machine learning. We are a group 1000+ members strong comprised of analysts, computer scientists, engineers....
   - **3,200** Members
   - **Meetup topics:** Predictive Analytics, Machine Learning, Big Data, Data Mining, Startup Businesses, Web Startups, Web Technology, Data Science, Business Intelligence, Internet & Technology, Show more...

4. **Data Science DC**
   - **Washington, DC**
   - Data Science DC is a non-profit professional group that meets monthly to discuss diverse topics in predictive analytics, applied machine learning, statistical modeling, open data, and data visualization. Our members are professionals, students, and....
   - **2,452** Data Scientists
   - **Meetup topics:** Predictive Analytics, Machine Learning, Applied Statistics, Data Mining, Data Visualization, Open Data, Data Analytics, Data Science