

Data Science vs. Statistics: What's the Difference?

Ron Fricker
August 10, 2015

What is Statistics?



- ❧ *“Statistics is about gathering data and working out what the numbers can tell us.”*¹
- ❧ *“Statistics as a discipline is the development and application of methods to collect, analyze, interpret, and learn from data.”*²
- ❧ *“Statistics is the discipline concerned with the study of variability, with the study of uncertainty and with the study of decision-making in the face of uncertainty.”*³
- ❧ *“Statistics is the science of learning from data, including collecting, organizing, analyzing, interpreting and presenting data, often with a particular focus on measuring, controlling, and communicating the uncertainty inherent in the data, associated analyses, and final results or conclusions.”*⁵
- ❧ *“Definitions of statistics abound but many fail to capture adequately the essential interplay of data and theory.”*⁴
- ❧ *“There can be no statistics without data and no statistics with data alone.”*

1 Champkin, J. (2013). Timeline of Statistics, *Significance*, 10, 23-26.

2 Virginia Tech website: www.stat.vt.edu/undergraduate/index.html.

3 Lindsay, B.G., et al. (2004). A Report on the Future of Statistics, *Statistical Science*, 19, p. 388.

4 Fricker, R.D., Jr. (2015). *Introduction to Probability and Statistics for Data Scientists*, draft textbook.

5 Bartholomew, D.J. (1995). What is Statistics? *JRSS-A*, 158, 1-20.

The Future Looks Good – At Least According to Google

☞ *“I keep saying that the sexy job in the next 10 years will be statisticians. And I’m not kidding.”¹*

Hal Varian, Google Chief Economist

☞ *“We are in the era of big data, and big data needs statisticians to make sense of it. The democratization of data means that those who can analyze it well will win. Data is the sword of the twenty-first century, those who wield it well, the samurai.”²*

Eric Schmidt, Google Chairman

¹ Lohr, S. (2009). “For Today’s Graduate, Just One Word: Statistics,” *New York Times*, August 6, p. A1.

² “Statisticians: The Data Samurai?” *Significance*, vol. 11, p. 3, 2014.

But It's Not Just Google

☞ *“The statistical sciences are very healthy across all subareas in the United States, which is the clear world leader.”*¹

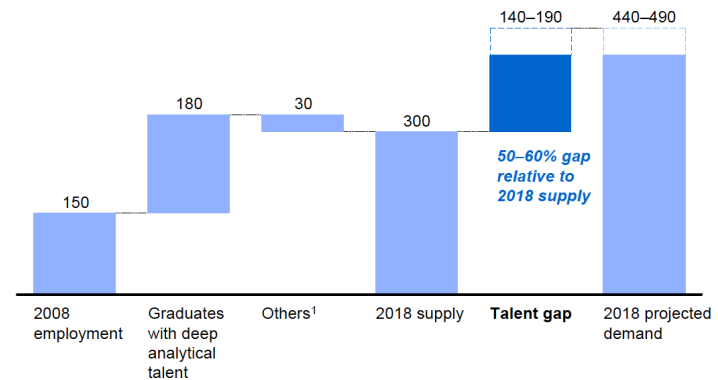
☞ BLS projecting 27% growth (“*much faster than average*”) in statistics occupations from 2012 to 2022²

☞ McKinsey Report: “...by 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of Big Data to make effective decisions.”³

Exhibit 4

Demand for deep analytical talent in the United States could be 50 to 60 percent greater than its projected supply by 2018

Supply and demand of deep analytical talent by 2018
Thousand people



¹ Other supply drivers include attrition (-), immigration (+), and reemploying previously unemployed deep analytical talent (+).
SOURCE: US Bureau of Labor Statistics; US Census; Dun & Bradstreet; company interviews; McKinsey Global Institute analysis

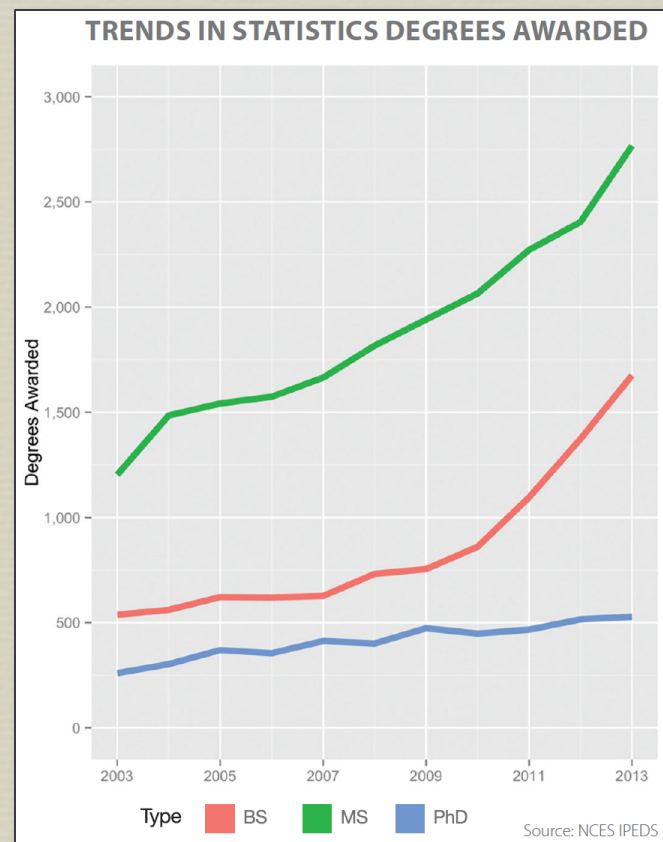
¹ Odom, W.E. (1998). Report of the Senior Assessment Panel for the International Assessment of the U.S. Mathematical Sciences, National Academy of Sciences, p. 55.

² Bureau of Labor Statistics (2014). *Occupational Outlook Handbook*, United States Department of Labor, January 8.

³ Manyika, J., et al., (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute.

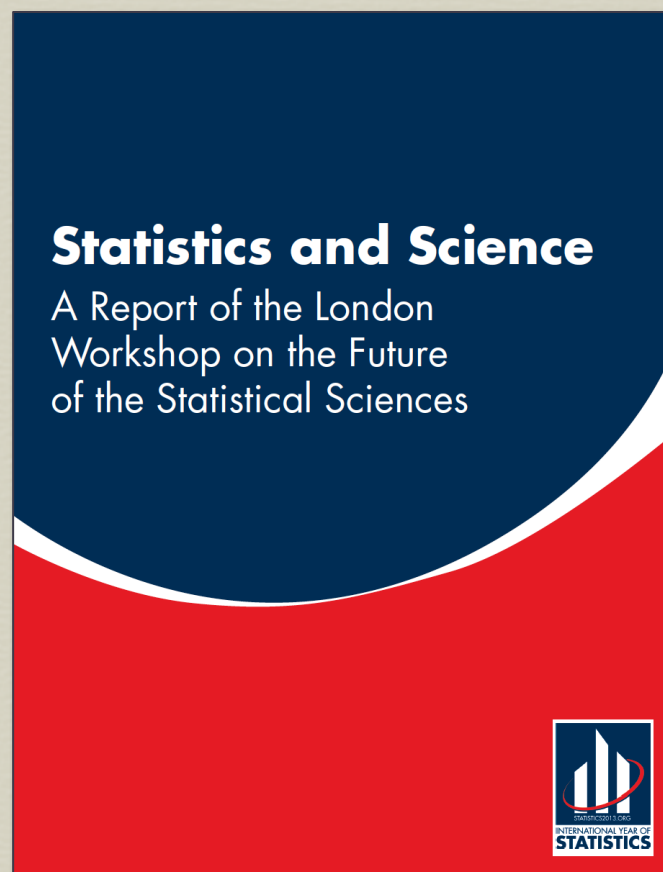
Reaction to the Demand

- Number of high school students taking the AP test in statistics has tripled in last decade
- Number of students studying statistics at two-year colleges greater than 137,000
 - Larger than total enrollment in calculus classes
 - In 1960, ratio was 10 to 1 for calculus to statistics students
- Phil Stark's statistics MOOC had more than 50,000 students sign up
- Number of statistics bachelors degrees awarded has roughly tripled since 2003
 - Up 21% from 2012 to 2013



Preparing Future Statisticians

- ❧ *“...statistics departments need to do a better job of preparing them for the jobs that are actually available and not necessarily to become carbon copies of the professors” (p. 34)*
- ❧ Suggestions include:
 - ❧ Improved communication skills
 - ❧ Working on team projects, especially with non-statisticians
 - ❧ Leadership training
 - ❧ Strong training in an application field
 - ❧ More exposure to “live” data
 - ❧ More exposure to Big Data
 - ❧ More emphasis on computer algorithms



In Summary: “SWOT”

Analysis of the Profession

∞ Strengths

- ∞ A “core” theory (set of theories) that unites the profession
- ∞ Diversity of well-developed methods founded on theory
- ∞ Professional focus on disseminating methods to non-statisticians

∞ Weaknesses

- ∞ Methods oriented towards well-structured, numeric, experimental data
- ∞ Much of today’s data is unstructured, non-numeric, and observational

∞ Opportunities

- ∞ Data, data everywhere!
- ∞ Strong demand signal for analysts & professional reputation never stronger

∞ Threats

- ∞ Lack of diversity of potential analysts / hesitancy to engage on non-traditional problems / difficulty communicating to non-statisticians
- ∞ Lack of knowledge among those who hire and engage analysts / others who are willing to “dive in” and work to solve problems

What is Data Science?

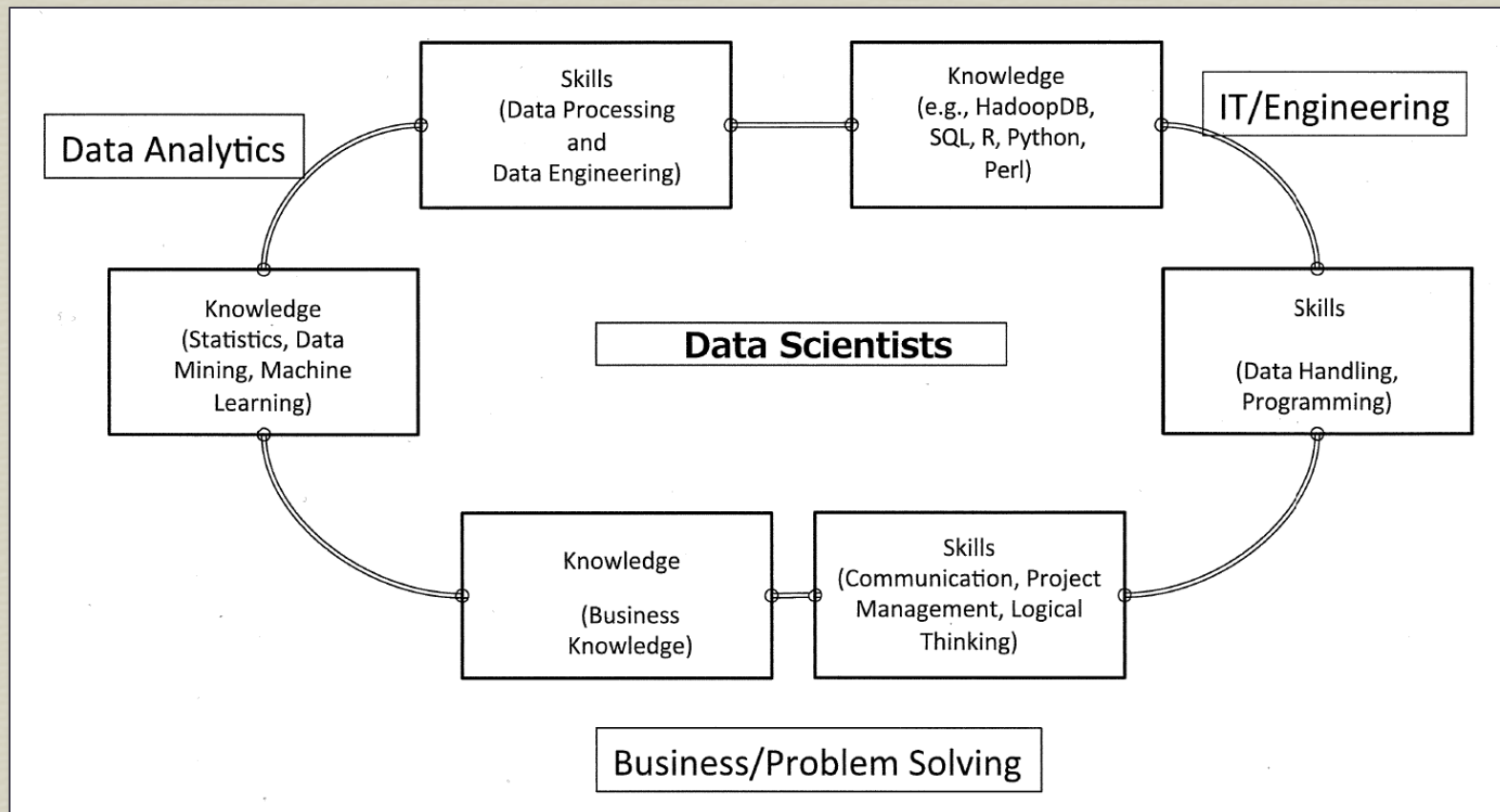
- ❧ *“Science of planning for, acquisition, management, analysis of, and inference from data.”*¹
- ❧ *“Data scientists’ most basic, universal skill is the ability to write code.”*²
 - ❧ *“A quantitative analyst can be great at analyzing data but not at subduing a mass of unstructured data and getting it into a form in which it can be analyzed.”*
 - ❧ Data scientists want to be *“on the bridge.”*
- ❧ *“Data science is the study of how to extract useful information from data using quantitative methods and theories from many fields, including statistics, operations research, computer science, and various engineering disciplines. Data science often focuses on large data sets not originally designed or collected to address the question of interest.”*³

1 Johnstone, I. and F. Roberts (2014). Data Science at NSF, draft report of StatSNSF committee, April.

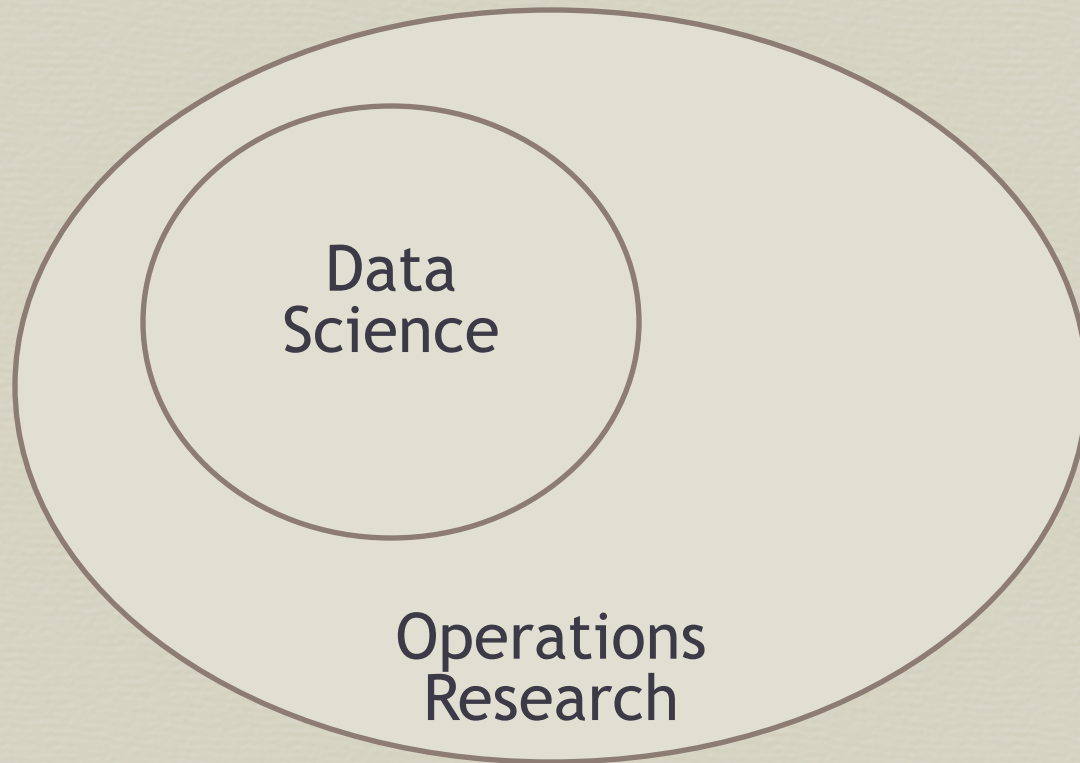
2 Davenport, T.H. and D.J. Patil (2012). Data Scientist: The Sexiest Job of the 21st Century, *Harvard Business Review*, October.

3 Fricker, R.D., Jr. (2015). *Introduction to Probability and Statistics for Data Scientists*, draft textbook.

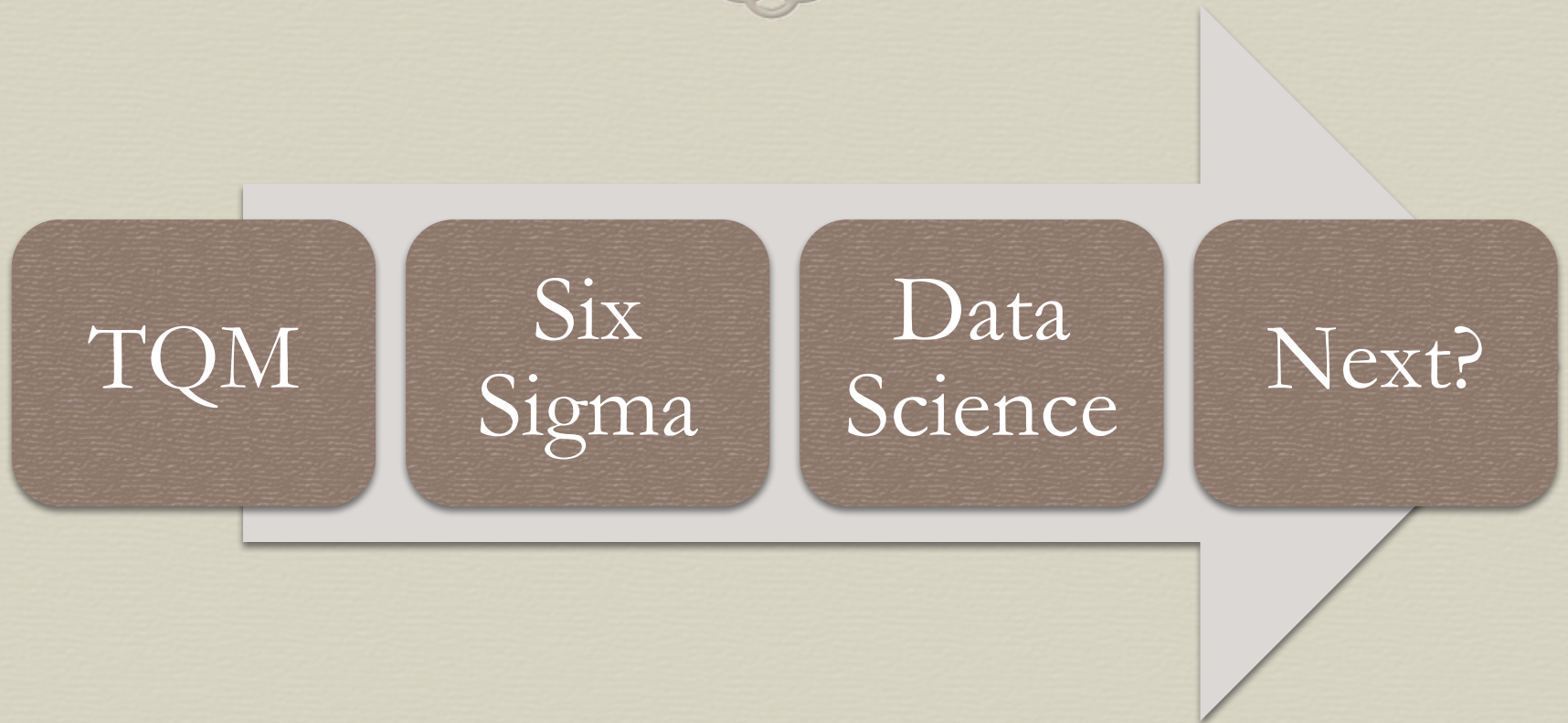
Data Science Skill Sets: One View



Another View

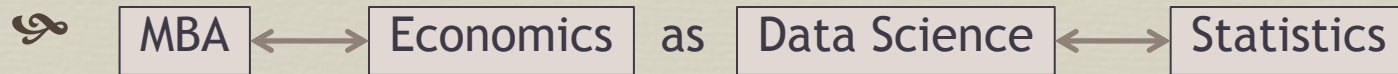


A Third View



Business fad of the month?

Data Science: My View



↪ Of course, I (we?) think it should be

$$\boxed{\text{Data Science}} = \boxed{\text{Statistics}}$$

↪ But really:

