Chapter 61
Comparison of Graduates' and Academics' Perceptions of the Skills Required for Big Data Analysis:
Statistics Education in the Age of Big Data

Busayasachee Puang-Ngern
Macquarie University, Australia

Ayse A. Bilgin
Macquarie University, Australia

Timothy J. Kyng
Macquarie University, Australia

ABSTRACT

There is currently a shortage of graduates with the necessary skills for jobs in data analytics and "Big Data". Recently many new university degrees have been created to address the skills gap, but they are mostly computer science based with little coverage of statistics. In this chapter, the perceptions of graduates and academics about the types of expertise and the types of software skills required for this field are documented based on two online surveys in Australia and New Zealand. The results showed that Statistical Analysis and Statistical Software Skills were the most necessary type of expertise required. Graduates in industry identified SQL as the most necessary software skill while academics teaching in relevant disciplines identified R programming as the most necessary software skill for Big Data analysis. The authors recommend multidisciplinary degrees where the appropriate combination of skills in statistics and computing can be provided for future graduates.

INTRODUCTION

Nobody can imagine a part of the developed world where more than half of the population is illiterate. In the developing world there might still be small pockets of the population who would benefit from learning to read and write, although they would be a minority. What about numerical and statistical literacy? Unfortunately, even in the developed world, we cannot be so confident as to say that people with numerical skills and statistical literacy are the majority (OECD, 2016, p.13). The need to educate people on statistical skills and literacy has been becoming more urgent with advent of the "Big Data" era, especially where most of the educators of the "Big Data" users and analysts are from computer science departments.

Around the world, there is an explosion of "data science" and "Big Data" degrees which were initiated by computer science departments who may lack the expertise to provide statistical literacy and statistical knowledge to future data scientists. Most Australian universities offer degree programs at both undergraduate and postgraduate level in mathematics, statistics, and computer science, however there does not seem to be a good example of collaboration between these three disciplines to educate future data scientists who would be dealing with "Big Data" analysis (Kyng, Bilgin, & Puang-Ngern, 2016; Puang-Ngern, 2015). Unfortunately the need to collaborate is not obvious to either the majority of the computer science academics or to the majority of the statistics academics. The authors argue that computer science academics believe that data science is all about computing. The structure of currently available data science degrees (which have subjects mainly offered by computer science departments) is a manifestation of what they think data science is, such that these degrees do not have as much statistics as needed for data scientists. The authors also argue that the majority of the statistics academics prefer to ignore data science with the expectation that "data science" is a temporary fad and a rebadging of statistical science by IT professionals which will go away soon. The miniscule number of presentations related to data science in the international statistics education conferences since 2010 is an indication of this ignorance. There were zero in ICOTS 8 (2010) and in IASE satellite (2011 and 2013); one in IASE roundtable (Finzer, 2012); five in ICOTS 9 (Kudo et al., 2014; Ridgway et al., 2014; McNamara & Hansen, 2014; Kaplan et al., 2014; Horton et al., 2014); one in IASE satellite (Kyng et al., 2015) and three in the last IASE roundtable (Schwab-McCoy, 2016; Gould, 2016; Bergen, 2016). These conferences are well attended by the statistics education community and influence the ideas of many attendees. The limited number of presentations during these conferences on data science/Big Data is an indication of the lack of interest in data science by statistics educators.

For a recent talk (10th March 2015) to the Statistics Society of New South Wales (Australia), Professor Thomas Lumley of the University of Auckland (New Zealand) wrote "Mainstream statistics ignored computing for many years ….. Practical estimation of conditional probabilities and conditional distributions in large data sets was often left to computer science and informatics. Although statistics started behind, we are catching up: many individual statisticians and some statistics departments are taking computing seriously. More importantly, applied statistics has a long tradition of understanding how to formulate questions: large-scale empirical data can tell you a lot of things, but not what your question is. Big Data are not only Big but Complex, Messy, Badly Sampled, and Creepy. These are problems that statistics has thought about for some time, so we have the opportunity to take all the shiny computing technology that other people have developed and use it to re-establish statistics at the centre of data science". Glance (2013) also pointed out that a combination of computational, statistical and

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/comparison-of-graduates-and-academics-perceptions-of-the-skills-required-for-big-data-analysis/217882

Related Content

Modern Diffusion of Products with Complex Network Models
Atsushi Tanaka (2011). *E-Activity and Intelligent Web Construction: Effects of Social Design (pp. 119-133).*
www.irma-international.org/chapter/modern-diffusion-products-complex-network/53279

Knowledge Representation Technologies Using Semantic Web
Vudattu Kiran Kumar (2019). *Web Services: Concepts, Methodologies, Tools, and Applications (pp. 1068-1076).*
www.irma-international.org/chapter/knowledge-representation-technologies-using-semantic-web/217876

A Web Application-Based Secured Image Retrieval System With an IoT-Cloud Network
Shikha Bhardwaj, Gitanjali Pandoveand Pawan Kumar Dahiya (2021). *International Journal of Web Services Research (pp. 1-20).*
www.irma-international.org/article/a-web-application-based-secured-image-retrieval-system-with-an-iot-cloud-network/271725

Big Data Security Framework for Distributed Cloud Data Centers
Chandu Thota, Gunasekaran Manogaran, Daphne Lopezand Vijayakumar V. (2019). *Web Services: Concepts, Methodologies, Tools, and Applications (pp. 240-257).*
www.irma-international.org/chapter/big-data-security-framework-for-distributed-cloud-data-centers/217834

Adoption of Web Services in Digital Libraries: An Exploratory Study
Fatih Oguz (2009). *Managing Web Service Quality: Measuring Outcomes and Effectiveness (pp. 307-320).*
www.irma-international.org/chapter/adoption-web-services-digital-libraries/26085