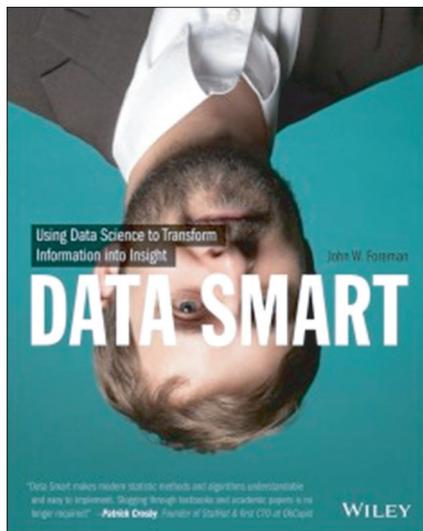


Data Smart: Using Data Science to Transform Information into Insight

Mona Choi, PhD, RN

College of Nursing, Yonsei University, Seoul, Korea
monachoi@yuhs.ac



Author: John W. Foreman

Year: November 4, 2013

Publisher: Wiley

ISBN: 978-1118661468

As big data increasingly becomes a buzzword, Health Informatics Research is diligently following the trends in this regard; book reviews of recent issues have introduced several books on ways to deal with and analyze data to enhance business [1-3]. Globally, big data initiatives have demonstrated interest in and attention towards the effective and efficient use of big data. Examples are the European Union's EUDAT, a major collaborative data infrastructure project in Europe [4], and the National Institutes of Health's Big Data to Knowledge (BD2K) initiative for biomedical big data [5]. Another movement is National Consortium for Data Science (NCDS) in the United States that seeks to advance the application of data to solve challenging problems, create jobs, protect national security, and improve quality of life [6]. Especially, BD2K appears to be an interesting attempt as it would enable scientists to take advantage of the big data being generated by research communities in the biomedical fields [5].

Indeed, we live in a world where the academic society has experienced a cultural shift from claiming 'my-own data' to actively sharing data and publication [5]. Hence, I want to add another book on data science, entitled, *Data Smart: Using Data Science to Transform Information into Insight*. As the subtitle itself indicates, this book places great emphasis on data science. This resource is not a theory- and code-based heavy reading; rather, it can help readers utilize data as critical insights for decision making. Whether you see yourself as book smart or street smart, this book will help you become data smart.

The author, John W. Foreman, is the Chief Data Scientist for MailChimp.com, an email service powering subscriptions

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2014 The Korean Society of Medical Informatics

for marketing campaigns. He has also worked with various organizations, such as the FBI, Department of Defense, Coca-Cola, and Intercontinental Hotels Group. Based on his background, Foreman uses examples and concepts from business; however, professionals working in healthcare will be able to apply this book to their fields as well.

Foreman defined data science as “the transformation of data using mathematics and statistics into valuable insights, decisions, and products.” *Harvard Business Review* published the article “Data Scientist: The Sexiest Job of the 21st Century,” which claimed that data scientists are a new kind of breed [7]. In fact, the term ‘data scientist’ was introduced first in 2008. Data scientists continue to be in great need in this big data era. According to the article, “if ‘sexy’ means having rare qualities that are much in demand, data scientists are already there.”

The author of *Data Smart* aims to provide an introduction to the practice of data science in a comfortable and conversational manner, and I think that he has been successful. He wants his readers to replace their anxiety of data science with excitement and ideas on how to use data to the next level for business. This book does not talk about health data at all, but I certainly insist that readers will feel more confident about data science after reading up to the last page of this book.

The first chapter is a short tutorial for the spreadsheet program, Microsoft Excel. Concepts and techniques are provided with the familiar Excel for most of readers. After readers learn these techniques with Excel for hands-on exercises, the last chapter talks about the use of the programming language R, which is appropriate for data science aiming at scalability. Foreman provides sample analyses in R with the same datasets and problems in previous chapters, thereby expanding the reader’s understanding of how the earlier techniques work in R environment, which focuses on analytics compared with Excel. He also provides a list of reference books on R at the end of this chapter for someone who wants to learn more about R.

This book consists of ten chapters that delve into the following topics:

- ◆ Cluster analysis
- ◆ Nut graphs
- ◆ k-means
- ◆ Artificial intelligence
- ◆ Regression
- ◆ Ensemble models
- ◆ Forecasting
- ◆ Outlier detection

These topics are introduced with eye-catching chapter titles, for example, “Naïve Bayes and the Incredible Lightness of Being an Idiot.” In the book, Foreman associates data science with other terms, such as business analytics, operations research, business intelligence, competitive intelligence, data analysis and modeling, and knowledge extraction, and those techniques show a glimpse of data science. In addition, each chapter offers pertinent datasets that readers can use in their hands-on exercise. Graphics and screen captures are presented to help the reader keep up with the concepts and exercises that are introduced.

As Foreman intended for this book to be an “introduction to the practice of data science in a comfortable and conversational way,” with particular attention given to clarity over mathematical correctness, readers can even attempt this book as enjoyable reading while on vacation this summer. His Twitter handle is @John4man, and readers might want to follow him. And please do not forget to visit the publisher’s website to listen to his introduction of this book and download datasets corresponding to the chapters at <http://www.wiley.com/go/datasmart>.

References

1. Ryu S. Book review: Predictive analytics: the power to predict who will click, buy, lie or die. *Healthc Inform Res* 2013;19(1):63-5.
2. An JY. Book review: Healthcare analytics for quality and performance improvement. *Healthc Inform Res* 2013;19(4):324-5.
3. Ryu S. Book Review: Big data management, technologies, and applications. *Healthc Inform Res* 2014;20(1):76-8.
4. EUDAT. European data infrastructure [Internet]. Espoo, Finland: EUDAT; c2014 [cited at 2014 Jul 22]. Available from: <http://www.eudat.eu/>.
5. Margolis R, Derr L, Dunn M, Huerta M, Larkin J, Sheehan J, et al. The National Institutes of Health’s Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data. *J Am Med Inform Assoc* 2014 Jul 9 [Epub]. <http://dx.doi.org/10.1136/amiajnl-2014-002974>.
6. Ahalt SC. Establishing a national consortium for data science [Internet]. [place unknown: publisher unknown]; 2012 [cited at 2014 Jul 22]. Available from: http://data2discovery.org/dev/wp-content/uploads/2012/09/NCDS-Consortium-Roadmap_July.pdf.
7. Davenport TH, Patil DJ. Data scientist: the sexiest job of the 21st century. *Harv Bus Rev* 2012;90:70-6.