



Paola Zuccolotto and Marica Manisera (2020): *Basketball Data Science: With Applications in R*, CRC Press, 243 pp., £80.50 (Hardcover), ISBN: 978-1-138-60079-9

Andreas Groll¹ · Carsten Jentsch¹

Received: 25 November 2021 / Revised: 6 December 2021 / Published online: 10 April 2022
© The Author(s) 2022

How can data science help for sports analytics in general and basketball in particular? As expressed by legendary basketball coach Ettore Messina in the foreword of the book, the goal is to “bring together the world of rigorous and thoughtful data analysis with things in the hand of a basketball coach, i.e., that are difficult to measure and rely on feelings, characters of a player, communication on the field etc.”

Situated within “big data analytics in sports”, the authors wrote the book “to achieve the ambitious objective of addressing a range of different audiences” (from the preface). The book is perfect for anyone interested in learning and applying data analytics in basketball. The interested reader should have basic knowledge in statistics and previous experience using the statistical software R. The intended audience consists of: graduate students in data or sports sciences; technicians, sports coaches and analysts; and data scientists interested in working in sports analytics.

The book is one of the first to present statistical and data mining methods for the steadily emerging field of basketball analytics. It provides a huge variety of tools and visualization techniques and covers plenty of real world data and applications, with a special focus on one of the most successful NBA teams in recent years, the Golden State Warriors. The motivation for this is that “the reader can imagine to be interested in a given team, perform a huge set of analyses about it and, in the end, build a deep awareness from a lot of different perspectives by summing up all the results in mind” (from the preface). Finally, source code and data are provided consistently throughout the book. Along with the availability of the R package *BasketballAnalyzeR* (Sandri et al., 2020), which was co-developed by the authors, the readers are given the ability to reproduce all the presented analyses and to perform their own analyses on other basketball teams and players.

✉ Andreas Groll
groll@statistik.uni-dortmund.de

Carsten Jentsch
jentsch@statistik.uni-dortmund.de

¹ TU Dortmund University, Dortmund, Germany

In general, the book is structured into three major parts. In Part I, the general peculiarities of basketball data are introduced along with introductory level, mostly descriptive statistical techniques. Part II presents a comprehensive overview of more advanced methods for basketball analytics. Each of the chapters concludes with a “focus section” covering specific applications of data analysis tools in basketball more deeply. The final Part III focuses on computational insights and gives more details on the R package `BasketballAnalyzeR`, which accompanies this book.

In the introductory Chapter 1 of Part I, the authors start with a thoughtful discussion on the potential benefit of data analytics in basketball, which has to be understood as “a tool for decision and not a substitute for human intelligence”. They give intuition on what data science *can* and what it *cannot* do in this context, and on how sports experts and data scientists should cooperate to gain from one another. Chapter 2 is devoted to the nature of basketball data and covers descriptive statistics and data visualization methods suitable for an explorative data analysis.

Motivated nicely in the context of basketball data analytics, the chapters in Part II cover more advanced statistical methods suitable for pattern discovery (Chapter 3), for finding groups (Chapter 4), and for modeling relationships in the data (Chapter 5). In particular, Chapter 3 discusses different kinds of variables, such as discrete, continuous and categorical variables, and proposes suitable statistical tools to deal with them as e.g. suitable dependence measures. As for finding groups in the data, Chapter 4 gives an overview of k-means and hierarchical clustering techniques and discusses how they can be productively applied for different data analysis tasks. With the specific goal of prediction in mind, Chapter 5 deals with statistical modelling using linear and non-linear as well as parametric and non-parametric models.

Finally, Part III consists of Chapter 6 devoted to the R package `BasketballAnalyzeR` and those little things that are inevitable when analyzing data, but that are not necessarily embraced by the term data science: data pre-processing and preparation or customizing plots and building interactive graphics.

In total, the book has the potential to become the new standard reference for data analytics in basketball. This is because it succeeds not only in presenting a whole variety of statistical methods suitable for data collected in the sport of basketball, but also because it provides a distinguished discussion of what an expert on sports can actually expect from data science, and, even more importantly, what data science may not be suitable for. Altogether, this results in a book that we believe can be of value for quite different audiences coming from various disciplines and having different backgrounds and pre-knowledge in statistics. So if you are either a basketball enthusiast or are passionate about data science applications in sports, you’ll surely enjoy this book!

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line

to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Reference

Sandri M, Zuccolotto P, Manisera M (2020) BasketballAnalyzeR: analysis and visualization of basketball data. R package version 0.5.0. <https://CRAN.R-project.org/package=BasketballAnalyzeR>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.