

## ***MULTIVARIATE DATA ANALYSIS – ITS APPROACH, EVOLUTION, AND IMPACT***

William Black (Emeritus)  
Louisiana State University, USA  
[wcblack@lsu.edu](mailto:wcblack@lsu.edu), (225) 578-8684

Barry J. Babin  
Max P. Watson Jr. Professor of Business, Louisiana Tech University, USA  
[bbabin@latech.edu](mailto:bbabin@latech.edu), (318) 257-4012

### **ABSTRACT**

The present chapter provides a brief chronology of the best known and most widely cited of many “Hair et al.” references, *Multivariate Data Analysis*. Along with providing a description of each edition of the book, brief citation analyses allow the chapter to convey the scale of impact resulting from the book. The near countless citations place *Multivariate Data Analysis* among the truly elite scholarly reference books across a myriad of disciplines, languages, countries, and years. In fact, an increasing trend in the number of citations suggest a long life ahead for this approachable statistics book that has helped so many in the past.

### **Introduction**

In the 40 years since a cadre of newly-minted PhD students from the University of Florida first decided to write a “*a simple introduction to multivariate analysis with emphasis on the practical use of these valuable tools*” (Preface, first edition, 1978), their text *Multivariate Data Analysis (MDA)* has emerged as one of the most widely used methodology texts within the social sciences. The primary driving force behind this effort was Joseph F. Hair, Jr., who has acted as a constant visionary in expanding the text over the years to incorporate emerging techniques (e.g., structural equation modeling) while still maintaining the initial, practical guiding principle – users first. Joe has become a global MDA evangelist through seminars and workshops he participates in worldwide. Through these efforts Joe has had, and continues to have, a profound influence on the widespread application of multivariate techniques across a number of academic domains. While Joe would be the first to credit the efforts of his co-authors in the development of the text over the years, we all can attest to his primary role as leader of the author team in continuing to make *Multivariate Data Analysis* a staple among users of statistics within both the academic and practitioner communities.

In trying to quantify the impact of *Multivariate Data Analysis*, all of the co-authors can recount innumerable anecdotes where researchers have expounded on their admiration for the book and its

influence on their use of these techniques in their research. Beyond these testimonials, the large numbers of journal articles that would not have been possible with *MDA* speak to its impact. Later, we provide a more quantitative assessment of that impact.

### **The Evolution of *Multivariate Data Analysis***

Before beginning the citation analysis, we thought a brief review of the development of *Multivariate Data Analysis* over the eight editions would provide some context for the subsequent analyses. As shown in Table 1, the text has slowly expanded in both content and co-authors. The first edition included the basic crux of multivariate methods and built off a regression framework. Cluster analysis was added in the second edition, with much of the rest of the book remaining the same in structure. The third edition, added an introductory chapter on SEM, but it also was the first edition to provide complete software syntax (IBM-SPSS) and outputs for all of the chapters. While other books have subsequently incorporated this feature, at that time it was a notable advance and somewhat of a challenge to provide given the download capabilities of the day.

Usefulness was a criterion in all changes across all editions. The Examining Your Data feature was added in the fourth edition. Another feature introduced in the fourth edition, which has been a tremendous success, is the six-step decision process used in the discussion of each technique. In this manner the researcher has a consistent framework that he/she can apply to any research situation, and the six-steps provide a basic “checklist” to ensure that all requisite issues have been considered. The fifth edition, which was 8 years (1998-2006) without revision, acknowledge advanced topics and expanded the introductory chapter on SEM. By this time, a shift to using factor analysis as the introductory technique was complete leading the way for perhaps the biggest change to the table of contents in the book’s history. The sixth edition, expanded coverage of SEM to three full chapters. Thus, the book provided a user-friendly approach to a technique not considered user-friendly at the time. The sixth edition also saw a formalization of the Rules of Thumb in each topic area, providing a set of guidelines to assist the researcher in both practical and statistical decisions. A new data set, more appropriate for SEM, HBAT-SEM, was created to illustrate the technique. By the seventh edition, published in 2010, a fourth chapter on SEM appears to provide descriptions of more advanced techniques.

--- insert Table 1 about here ---

In 2018 came the eighth edition, by now with Cengage International. The addition of a chapter for PLS-SEM, which was originally introduced in the sixth edition, provides the biggest change to the TOC. However, the edition also integrates data science - *Big Data* - principles into each chapter, providing an expanded perspective on the applicability of the techniques not only to classical statistical problems, but also the emerging situation of analyses involving very large datasets in terms of both cases and variables. The coverage of multi-level types of analyses also is expanded.

The overarching objective in all of these developments has been to provide the researcher, whether in the academic or practitioner domain, with a “practical” guidebook to (a) identifying the appropriate multivariate technique, (b) applying it in the correct manner and then (b) interpreting the results for both statistical and managerial significance. The steady influence of Joe throughout this process, even while content and co-authors have changed, has kept *Multivariate Data Analysis* true to its original purpose.

## **The Impact on Social Science Methodology**

Perhaps the most direct evidence of impact is through citation analysis, which long has been used as an objective approach to assessing influence and impact upon the body of published research (Narin 1976, White and McCain 1998, Harzing and Van der Wal 2008, Hirsch 2005). Over time, several methods have been developed for compiling citations (e.g., Google Scholar, Web of Science and SCOPUS), each with its own advantages and limitations (Falagas, et al 2008). A number of efforts have been made to compare among the various citation sources, and while identifying issues with each approach, their caveats for use point to a similarity among the methods (MacRoberts and MacRoberts 1989, Leydesdorff 2008). For our purposes, we will rely on results from several of these sources as each provides some unique perspectives on the impact of *Multivariate Data Analysis* in the academic literature.

The growing interest in citation analysis has provided analysts with a number of perspectives on the influence of any specific article, book or even researcher. We will employ several of those perspectives to better understand objectively the contributions of *Multivariate Data Analysis*. First is a comparison among the most cited articles and texts in the social science as compiled by Green (2016). We then provide a longitudinal view of the citation count to understand its growth and the prospects for continued influence. A third approach illustrates the citation count of *Multivariate Data Analysis* among the basic research domains, focusing on the top 15 domains based upon citation count. Finally, we take a global view by cumulating citation count by country and provide evidence of the truly global impact of the text.

### Ranking Among Social Science Methodology Texts

The most direct measure of impact for *Multivariate Data Analysis* is the total number of citations, which using Google Scholar now totals over 100,000. While that number by itself is quite amazing, it becomes even more impressive when compared to other methodology texts within the social sciences. Green (2016) conducted a citation analysis of the most cited papers and texts within the social sciences. While looking at the most influential books and articles, he also put a specific focus on methodology texts, which have an especially high citation count due to their applicability across many types of research settings. Within this set of texts, *Multivariate Data Analysis* was the fifth most cited

methodology text and updated figures for 2018 indicate it moved into fourth place. Moreover, if we consider just quantitative analysis, it is the third most cited text.

--- insert Table 2 about here ---

Green's work also identified the 25 most cited books in the social sciences (excluding methodology texts) and in that list *Multivariate Data Analysis* would rank fourth, just behind Thomas Kuhn's (1962) *The Structure of Scientific Revolutions*, Everitt Rogers' (1962) *Diffusion of Innovations* and Paulo Freire's (1968/70) *Pedagogy of the Oppressed* and ahead of Michael Porter's (1980) *Competitive Strategy*. This comparison demonstrates the profound impact *Multivariate Data Analysis* has had within the social sciences, comparable to the most seminal works within the substantive research disciplines of Philosophy, Sociology, Education and Economics.

### An Increasing Impact Over Time

As the total citation count for *Multivariate Data Analysis* demonstrates, it has had a profound impact in the social sciences as a guide for quantitative analysis. When we view the citation counts over the past 20 years, we see an ever increasing trend upwards, basically tripling in the last decade. We sense that this trend reflects the increased use of quantitative analysis within these research areas. As for *Multivariate Data Analysis*, we feel it provides further evidence of the text's significant role in the methodological traditions within the social sciences, its relevance today, and an expectation that it will remain relevant into the future.

--- insert Figure 1 about here ---

### The Span of Research Domains

While the previous citation analyses have demonstrated to significant influence of *Multivariate Data Analysis* in the area of quantitative methodology, there is still an unresolved question: How concentrated among research domains is this influence? As any methodologist knows, disciplines are quite parochial in their perspectives and uses of specific techniques, so can a text like *Multivariate Data Analysis* transcend these barriers and gain a widespread acceptance across research domains?

Web of Science, in its citation analyses, provides a breakdown of the citations of any article or text by major research domain. Figure 2 provides a distribution of citations across the 15 research domains with the highest number of citations. As would be expected, given the primary discipline of Marketing for all of *MDA*'s co-authors, Business Economics accounts for 37% of the citations, followed by Social Sciences Other Topics and Psychology, which account for about 10% of citations. Engineering, Computer Science, Environmental Sciences, and Educational Research follow and each accounts for at least 5% of *MDA*'s citations. The remaining citations appear primarily in other areas (from the top 15) within the social sciences.

--- insert Figure 2 about here ---

## The Global Reach of *Multivariate Data Analysis*

Our final perspective is on the global influence of *Multivariate Data Analysis*. Foreign language versions of *MDA* exist with perhaps the most prominent being in Portuguese. Over 7,000 sources cite the Portuguese edition according to Google Scholar. Further, considering only the English version of *MDA*, we find citations in dozens of foreign language journals. For example, 1,000 citations appear in Spanish and Chinese, hundreds in German and in French, and dozens in Italian and Hebrew, just to name a few. A Web of Science breakdown provides citations by country, with Figure 3 providing a distribution of citations across the 15 countries with the highest number of citations. Here, from the perspective of coauthor, we find perhaps the most interesting results. Joe has always been an avid promoter of multivariate analysis and of *MDA*, doing seminars globally for many years, with a particular emphasis on the Pacific Rim. As can be seen in these results, three Pacific Rim countries - Taiwan, China, and Australia - account for about the same number of citations as does the body of work authored in the United States. Moreover, the Web of Science results indicate citations from authors in a total of 155 countries, clear evidence of the global impact matched by few other publications.

--- insert Figure 3 about here ---

## **The Future?**

*Multivariate Data Analysis*, under the leadership of Joe Hair, continues to influence the practice of quantitative analysis profoundly across the social sciences globally. What has been its appeal? As noted several times, the *practical* nature of the text makes it appealing to a vast number of analysts, both academic and practitioner. Joe learned how to use multivariate data analysis out of necessity more than choice. Lacking a dominant foundation in advanced mathematics, he came to understand the statistics more conceptually than mathematically. Given that many users of statistics likewise lack experience in advanced mathematics, readers took (and continue to take) comfort in a pedagogical style that makes advanced topics more approachable. In fact, Joe often brags that, at least in the first two editions, *MDA* contained only one equation! A sparsity of “formulas and squiggles,” as we many times say in jest, represents an explicit attempt to provide an approachable and unimposing perspective of statistics, far more accessible than one based primarily on mathematical notation. While the addition of coauthors and other topics led to the introduction of a few more equations, the pedagogy still relies on illustration more than manual computation. Moreover, the common six-step decision framework across all the techniques, coupled with common databases for all the analyses, provides a consistent perspective for most analytical contexts. With the expansion of content in the eighth edition the author team is hopeful *Multivariate Data Analysis* is positioned for the needs of today’s researchers, which have changed so dramatically over the four decades of the text.

While the original and current author teams face the ever-advancing challenges of aging, we all look at amazement at Joe as he represents the “Energizer Bunny™” of our group; always moving forward

with new ideas and topics for the book, while seemingly constantly stepping onto another plane, off to another seminar. His boundless energy has served us well and will provide a continuing motivation for improvements in the next edition of *Multivariate Data Analysis*. Indeed, the traditional textbook market itself is in flux and holds an uncertain future. But, *MDA* has never been a traditional textbook. As long as users need a user-friendly ready reference on factor rotations, multicollinearity, cluster selection, construct validity or chi-square goodness of fit, *MDA* will remain relevant and impactful. At its heart, *Multivariate Data Analysis* is a very approachable book born out of the labor of a very approachable guy. Thousands of researchers worldwide are grateful for the work of Dr. Joseph F. Hair, Jr.

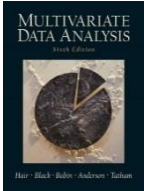

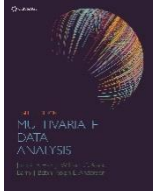
## References

- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). "Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses." *The FASEB Journal*, 22(2), 338-342.
- Green, Elliott D. (2016) "What are the most-cited publications in the social sciences (according to Google Scholar)? " *Impact of Social Sciences Blog*, London School of Economics and Political Science.
- Harzing, A. W. K., & Van der Wal, R. (2008). "Google Scholar as a new source for citation analysis." *Ethics in science and environmental politics*, 8(1), 61-73.
- Hirsch, J. E. (2005). "An index to quantify an individual's scientific research output." *Proceedings of the National Academy of Sciences*, 102(46), 16569-16572.
- Leydesdorff, L. (2008). "Caveats for the use of citation indicators in research and journal evaluations." *Journal of the American Society for Information Science and Technology*, 59(2), 278-287.
- Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity* (pp. 206-219). Cherry Hill, NJ: Computer Horizons.
- MacRoberts, M. H., & MacRoberts, B. R. (1989). "Problems of citation analysis: A critical review." *Journal of the American Society for Information Science*, 40(5), 342-349.
- White, H. D., & McCain, K. W. (1998). "Visualizing a discipline: An author co-citation analysis of information science, 1972–1995." *Journal of the American Society for Information Science*, 49(4), 327-355.

**Table 1: The Evolution of *Multivariate Data Analysis* Since 1978**

Edition (Year)	Publisher/Author Changes	Additions and Changes
<p><b>First edition</b> (1978)</p> 	<ul style="list-style-type: none"> <li>• Petroleum Publishing Company (PPE), Tulsa, Oklahoma</li> <li>• J.F. Hair, Jr., Rolph E. Anderson, Ronald L. Taham, Bernie J. Grabiowsky</li> </ul>	<ul style="list-style-type: none"> <li>• Seven chapters (Introduction, Multiple Regression, Discriminant Analysis, MANOVA, Canonical Correlation, Factor Analysis, MDS/Conjoint Analysis)</li> <li>• Readings for each chapter</li> <li>• Introduction of HATCO database</li> </ul>
<p><b>Second edition</b> (1987)</p> 	<ul style="list-style-type: none"> <li>• MacMillan Publishing</li> <li>• J.F. Hair, Jr., Rolph E. Anderson, Ronald L. Taham,</li> </ul>	<ul style="list-style-type: none"> <li>• Addition of cluster analysis chapter</li> <li>• Separate MDS and conjoint into separate chapters</li> </ul>
<p><b>Third edition</b> (1992)</p> 	<ul style="list-style-type: none"> <li>• MacMillan Publishing</li> <li>• Added William Black as co-author</li> </ul>	<ul style="list-style-type: none"> <li>• Addition of SEM chapter</li> <li>• Addition of syntax and outputs for all analyses</li> </ul>
<p><b>Fourth edition</b> (1995)</p> 	<ul style="list-style-type: none"> <li>• Prentice-Hall</li> <li>• Same as third edition</li> </ul>	<ul style="list-style-type: none"> <li>• Addition of <i>Examining Your Data</i> chapter</li> <li>• First translations/international editions</li> </ul>
<p><b>Fifth edition</b> (1998)</p> 	<ul style="list-style-type: none"> <li>• Prentice-Hall</li> <li>• Same as fourth edition</li> </ul>	<ul style="list-style-type: none"> <li>• Addition of <i>Advanced and Emerging Techniques</i> chapter</li> <li>• Expansion of SEM chapter</li> </ul>



<p><b>Sixth edition</b> (2006)</p> 	<ul style="list-style-type: none"> <li>• Prentice-Hall</li> <li>• Addition of Barry J. Babin as co-author</li> </ul>	<ul style="list-style-type: none"> <li>• Expansion of SEM into three chapters (<i>Introduction to SEM, Confirmatory Factor Analysis, Structural Models in SEM</i>)</li> <li>• Formalized <u>Rules of Thumb</u> in each topic area</li> <li>• Revised and updated database to HBAT</li> </ul>
<p><b>Seventh edition</b> (2010)</p> 	<ul style="list-style-type: none"> <li>• Prentice-Hall</li> <li>• Hair, Black, Babin and Anderson</li> </ul>	<ul style="list-style-type: none"> <li>• Addition of fourth chapter on SEM</li> <li>• Removal of <i>Canonical Correlation</i> chapter in domestic edition</li> <li>• Introduction of global edition</li> </ul>
<p><b>Eighth edition</b> (2018)</p> 	<ul style="list-style-type: none"> <li>• Cengage Europe</li> <li>• Same as seventh edition</li> </ul>	<ul style="list-style-type: none"> <li>• Addition of separate PLS chapter in SEM section</li> <li>• Addition of data science perspective in multivariate chapters</li> <li>• Move chapters on <i>Canonical Correlation, Multidimensional Scaling</i> and <i>Conjoint Analysis</i> to online resources</li> </ul>

**Table 2 The 10 Most Cited Methodology Books in the Social Sciences<sup>a</sup>**

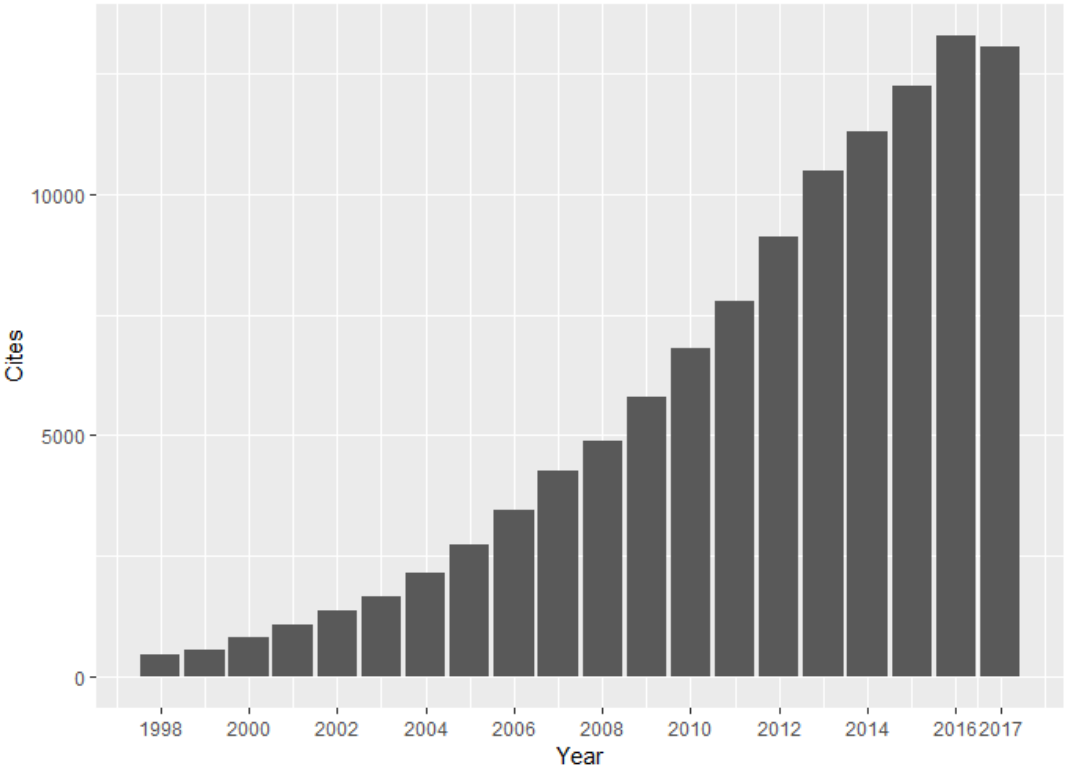
Book	Author(s)	Citations		
		Date	Article	Updated <sup>b</sup>
<i>Applied Multiple Regression/Correlation Analysis For the Behavioral Sciences</i>	J. Cohen, P. Cohen, S. West and L. Aiken	1975	131,033	180,063
<i>Case Study Research: Designs and Methods</i>	Robert Yin	1984	107,931	160,755
<i>Psychometric Theory</i>	Jim Nunnally	1967	80,196	111,474
<i>The Discovery of Grounded Theory: Strategies for Qualitative Research</i>	Barney Glaser and Anselm Strauss	1967	78,385	103,424
<i>Multivariate Data Analysis</i>	J F Hair, R. E. Anderson and R.L. Tatham	1979	70,700	105,114
<i>Qualitative Data Analysis</i>	Mathew Miles and A. Michael Huberman	1985	58,829	84,347
<i>Using Multivariate Statistics</i>	Barbara Tabachnick and Linda Fidell	1989	57,324	81,553
<i>Econometric Analysis</i>	William Greene	1990	54,424	68,583
<i>An Introduction to Probability Theory and its Applications</i>	William Feller	1950	51,825	52,158
<i>Naturalistic Inquiry</i>	Yvonna Lincoln and Egon Guba	1985	51,169	76,924

<sup>a</sup> Includes Anthropology, Economics, Education, Geography, Linguistics, Management, Philosophy, Political Science and Psychology

<sup>b</sup> Source: Google Scholar, August, 2018

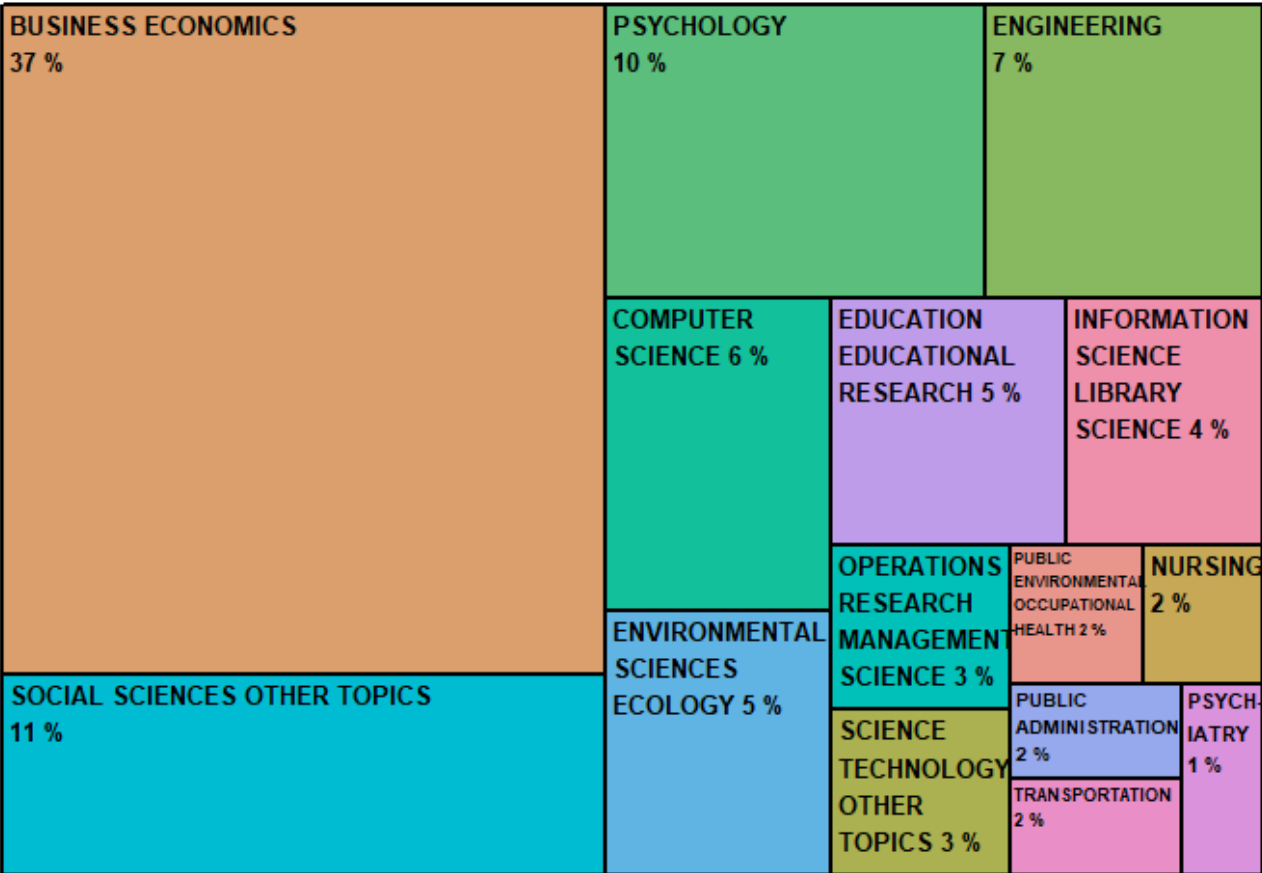
Source: Green, Elliott D. (2016) "What are the most-cited publications in the social sciences (according to Google Scholar)?" *Impact of Social Sciences Blog*, London School of Economics and Political Science.

Figure 1 Longitudinal Pattern of Citations for *Multivariate Data Analysis* Over the Past 20 Years



Source: Google Scholar

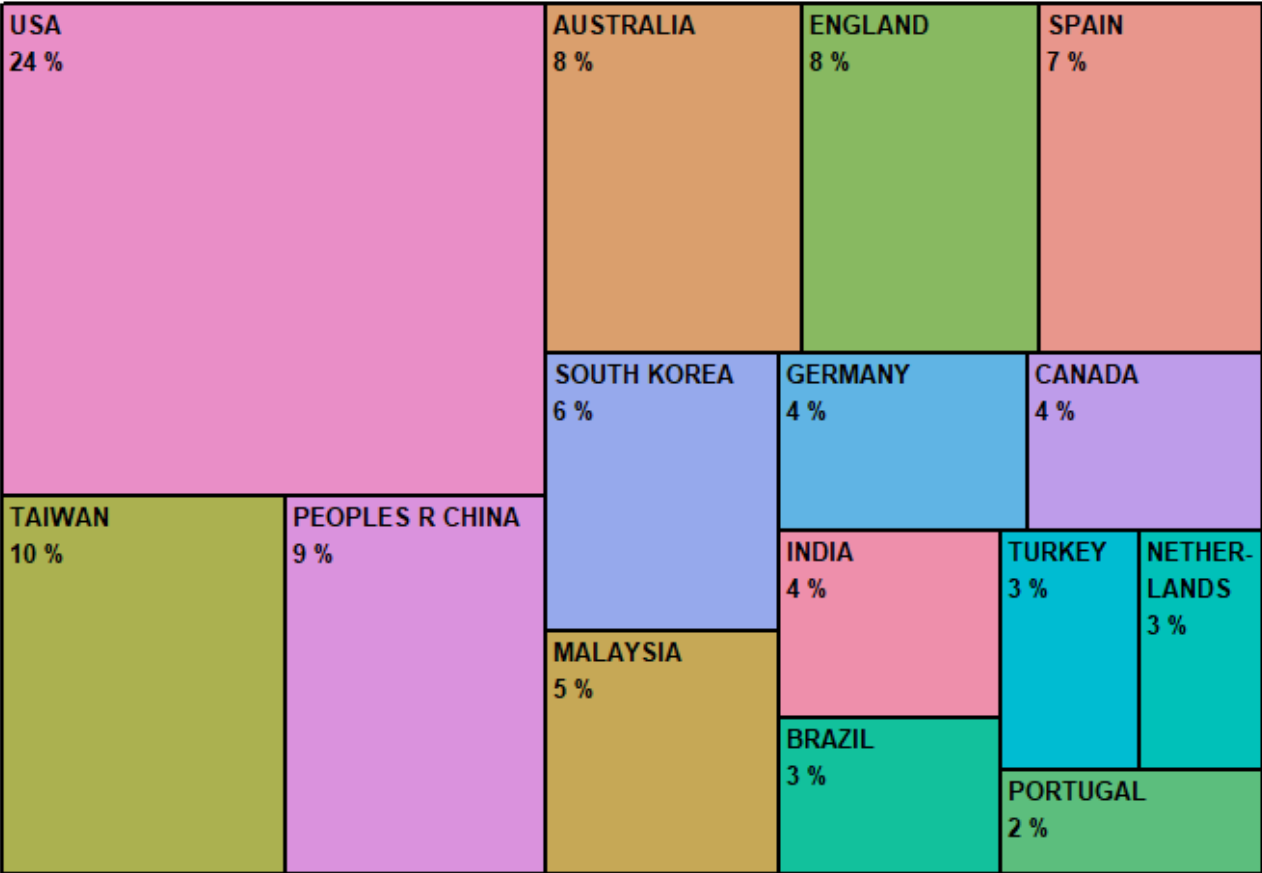
Figure 2 Distribution of Citations for *Multivariate Data Analysis* Across Research Domains<sup>a</sup>



a Includes the top 15 research domains based on citation counts

Source: Web of Science

Figure 3 Distribution of Citations for *Multivariate Data Analysis* Across Countries<sup>a</sup>



<sup>a</sup> Includes the top 15 countries based on citations

Source: Web of Science