



Journal of Statistical Software

January 2017, Volume 76, Book Review 1.

doi: [10.18637/jss.v076.b01](https://doi.org/10.18637/jss.v076.b01)

Reviewer: Virgilio Gómez-Rubio
Universidad de Castilla-La Mancha

Working with the American Community Survey in R: A Guide to Using the `acs` Package

Ezra Haber Glenn
Springer-Verlag, Cham, 2016.
ISBN 978-3-319-45772-7. 53 pp. EUR 49.99.
<http://www.springer.com/9783319457710>

Government agencies are increasingly making a number of socioeconomic and demographic datasets available for researchers and policy makers to use. This book is an introduction to the R package `acs`, that provides an interface to access the data in the American Community Survey (ACS, henceforth) and other census data. Very briefly, the ACS contains estimates of several variables from individual information collected in surveys at different administrative levels. The analysis of these survey data will be of interest to anyone working in the USA and involved in urban planning, demography and social sciences research in a broad sense. For example, the ACS can be used to obtain an estimate of the number of high school graduates in Orange county (New York) and its associated standard error. Estimates of a large number of socioeconomic variables (and their standard errors) are available in the ACS for download. In the first chapter of the book, an introduction to the ACS and the challenges of working with these data are briefly addressed, such as the importance of adjusting for inflation or the complexity of aggregating estimates at different administrative levels. At the end of the chapter, a brief case study on the use of this package is presented.

Chapter 2 deals with the installation of the package in R. I believe that this part could have been omitted as the package can be easily installed from CRAN on all platforms and this will be known by most R users.

Chapter 3 describes the main functions in the `acs` package. It starts with a description of the different geographical entities covered by the ACS and how these can be defined using some of the functions in the package. Then several functions to fetch data are described, but these require the exact names of the variables in the ACS to gather the data. At the end of the chapter, several functions to find particular variable names using partial matching are described. In my opinion, this would have been a better start for the chapter as it will help users to find out what data are actually available in the ACS, and then follow with a description on how to define geographical regions and then how to download the actual data.

Chapters 4 and 5 cover exporting data and additional resources, respectively. However, exporting data only covers how to save the data in `csv` format and the list of additional

resources is so short that it could have been included elsewhere. Finally, the appendix provides a worked example on the extraction and analysis of the ACS data.

Although the book describes the installation and use of the `acs` package I find a few important issues missing. First of all, I would have liked the example mentioned in Chapter 1 to have been developed throughout the book as this would have been a way to link all the parts of the book together. Also, the book focuses on data extraction of ACS data but it does not cover plotting methods and the case studies included in the book do not include a discussion on the implications of the estimates and standard errors of the variables downloaded from the ACS survey for policy making.

The `acs` package provides a number of functions to plot point estimates and associated errors from the ACS and I believe that this would have been an interesting part of the book. In addition, I miss some guidelines on how to represent results using maps (as claimed at the Springer website of the book!). The US Census Bureau provides boundaries at different administrative levels and these could be used together with the ACS for urban planning, for example. Furthermore, maps are often an important part in policy making as they help to communicate policies and actions to the tax payers. The author mentions a couple of improvements to the package that will be added later, and perhaps these should have already been implemented before releasing the book.

This book does not describe the American Community Survey in detail nor how variable estimates are computed from the original survey data. I have also missed a link to other R packages for the analysis of survey data. The ACS provides valuable data that could be analysed with the myriad of packages available in R for modelling, small area estimation and official statistics.

Finally, the Springer Briefs in Statistics series aims at providing short reports that are the state-of-the-art on a particular subject. However, titles in this series seem, in general, a bit overpriced considering, for example, that the titles in the `useR!` series provide a more comprehensive approach to topics at a lesser price. R has a number of excellent ways to include short descriptions on the usage of a package (such as the one given in this book) that may be more appropriate as they can be easily kept up to date (for example, vignettes) and include the R code from the examples.

Reviewer:

Virgilio Gómez-Rubio
Department of Mathematics
Universidad de Castilla-La Mancha
Avda. España s/n
02071 Albacete, Spain
E-mail: Virgilio.Gomez@uclm.es
URL: <http://www.uclm.es/profesorado/vgomez/>

Journal of Statistical Software

published by the Foundation for Open Access Statistics

January 2017, Volume 76, Book Review 1

doi:[10.18637/jss.v076.b01](https://doi.org/10.18637/jss.v076.b01)

<http://www.jstatsoft.org/>

<http://www.foastat.org/>

Published: 2017-01-12
