1.	Record Nr.	UNINA9910813793203321
	Autore	Buttrey Samuel
	Titolo	A data scientist's guide to acquiring, cleaning and managing data in R / / by Samuel Buttrey, Naval Postgraduate School, California, United States, Lyn R. Whitaker, Naval Postgraduate School, California, United States
	Pubbl/distr/stampa	Hoboken, New Jersey : , : Wiley, , 2017 ©2017
	ISBN	1-119-08007-X 1-119-08006-1 1-119-08005-3
	Edizione	[1st edition]
	Descrizione fisica	1 online resource (284 pages)
	Collana	THEi Wiley ebooks
	Disciplina	005.74/3
	Soggetti	Database design - Computer programs Data editing - Computer programs Input design, Computer R (Computer program language)
	Lingua di pubblicazione	Inglese
	Formato	Materiale a stampa
	Livello bibliografico	Monografia
	Nota di bibliografia	Includes bibliographical references and index.
	Sommario/riassunto	The only how-to guide offering a unified, systemic approach to acquiring, cleaning, and managing data in R Every experienced practitioner knows that preparing data for modeling is a painstaking, time-consuming process. Adding to the difficulty is that most modelers learn the steps involved in cleaning and managing data piecemeal, often on the fly, or they develop their own ad hoc methods. This book helps simplify their task by providing a unified, systematic approach to acquiring, modeling, manipulating, cleaning, and maintaining data in R. Starting with the very basics, data scientists Samuel E. Buttrey and Lyn R. Whitaker walk readers through the entire process. From what data looks like and what it should look like, they progress through all the steps involved in getting data ready for modeling. They describe best practices for acquiring data from numerous sources; explore key issues in data handling, including text/regular expressions, big data, parallel

processing, merging, matching, and checking for duplicates; and outline highly efficient and reliable techniques for documenting data and recordkeeping, including audit trails, getting data back out of R, and more. The only single-source guide to R data and its preparation, it describes best practices for acquiring, manipulating, cleaning, and maintaining data Begins with the basics and walks readers through all the steps necessary to get data ready for the modeling process Provides expert guidance on how to document the processes described so that they are reproducible Written by seasoned professionals, it provides both introductory and advanced techniques Features case studies with supporting data and R code, hosted on a companion website A Data Scientist's Guide to Acquiring, Cleaning and Managing Data in R is a valuable working resource/bench manual for practitioners who collect and analyze data, lab scientists and research associates of all levels of experience, and graduate-level data mining students.