

CONTENTS

About the Author	ix
Preface	xi
1 Introduction	1
1.1 Objects, variables, and workspaces	4
1.2 Working with external files: data files, scripts, file paths, and working directories	8
1.2.1 Saving variables and objects in data files and restoring them from files	8
1.2.2 R scripts	9
1.2.3 Files, paths, and working directories	11
2 Building Blocks of Data	15
2.1 Basic data types	16
2.1.1 Numeric vectors	16
2.1.2 Logical vectors	19
2.1.3 Character vectors	22
2.1.4 Factors	24
2.2 Basic data manipulations	29
2.2.1 Extracting and replacing elements of vectors	29
2.2.2 Reordering the elements of a vector	31
2.2.3 Regular sequences and repetitions	33
2.2.4 Sampling from vector elements	34

2.3 The construction of more complex data types: lists, attributes, and classes	36
2.3.1 Lists	37
2.3.2 Attributes	40
2.3.3 Generic functions, classes, and methods	41
3 Rectangles of Variables and Observations: Data Frames and Their Management	45
3.1 The structure of data frames	46
3.2 Accessing and changing variables in data frames	51
3.3 Manipulating data frames	55
3.3.1 Subsetting	55
3.3.2 Combining data on different variables about the same observations: merging data frames	56
3.3.3 Appending	63
3.3.4 Reshaping	66
3.3.5 Sorting	72
3.4 Aggregating data frames	74
3.5 Groupwise computations within data frames	77
3.6 Importing data into data frames	79
3.6.1 Importing data from text files	79
3.6.2 Importing data from other statistics packages	82
4 Data Tables and the Tidyverse	87
4.1 Data tables	88
4.2 Tidyverse	93
4.3 A comparison of approaches	98
4.4 Tidying up data using <i>tidyverse</i>	100
5 Handling Data from Social Science Surveys	109
5.1 The structure of data from opinion polls and social and political surveys	110
5.2 Importing survey data from SPSS and Stata into R	118
5.3 Recoding and similar transformations of survey items	126
6 Managing Data from Complex Samples	131
6.1 Survey design objects	133
6.2 Survey replicate weights	141
6.3 Post-stratification, raking, and calibration	145
7 Dates, Times, and Time Series	153
7.1 Dates and times	154
7.2 Time series	164
8 Spatial/Geographical Data	185
8.1 The structure of spatial data in R	186
8.2 Spatial relations and operations	196

8.3 Cartographic projections	208
8.4 Geographical data files	210
9 Text as Data	215
9.1 Manipulation of character strings	216
9.2 Working with text corpora with the <i>tm</i> package	223
9.3 Improvements provided by the <i>quanteda</i> package	233
Bibliography	249
Index	253