
Contents

1	Introduction	1
1.1	Overview	1
1.2	Why Learn R?	2
1.3	Is R Accurate?	3
1.4	What About Tech Support?	4
1.5	Getting Started Quickly	5
1.6	The Five Main Parts of SAS and SPSS	5
1.7	Our Practice Data Sets	7
1.8	Programming Conventions	8
1.9	Typographic Conventions	9
2	Installing and Updating R	11
2.1	Installing Add-on Packages	11
2.2	Loading an Add-on Package	13
2.3	Updating Your Installation	15
2.4	Uninstalling R	17
2.5	Uninstalling a Package	17
2.6	Choosing Repositories	18
2.7	Accessing Data in Packages	18
3	Running R	21
3.1	Running R Interactively on Windows	21
3.2	Running R Interactively on Macintosh	24
3.3	Running R Interactively on Linux or UNIX	26
3.4	Running Programs That Include Other Programs	28
3.5	Running R in Batch Mode	29
3.6	Running R in SAS and WPS	30
3.6.1	SAS/IML Studio	30
3.6.2	A Bridge to R	31
3.6.3	The SAS X Command	31
3.6.4	Running SAS and R Sequentially	32

3.6.5	Example Program Running R from Within SAS	32
3.7	Running R in SPSS	33
3.7.1	Example Program Running R from Within SPSS	37
3.8	Running R in Excel	37
3.9	Running R from Within Text Editors	39
3.10	Integrated Development Environments	40
3.10.1	Eclipse	40
3.10.2	JGR	41
3.10.3	RStudio	42
3.11	Graphical User Interfaces	42
3.11.1	Deducer	43
3.11.2	R Commander	46
3.11.3	rattle	48
3.11.4	Red-R	51
4	Help and Documentation	53
4.1	Starting Help	53
4.2	Examples in Help Files	55
4.3	Help for Functions That Call Other Functions	57
4.4	Help for Packages	57
4.5	Help for Data Sets	58
4.6	Books and Manuals	58
4.7	E-mail Lists	58
4.8	Searching the Web	59
4.9	Vignettes	60
4.10	Demonstrations	60
5	Programming Language Basics	61
5.1	Introduction	61
5.2	Simple Calculations	62
5.3	Data Structures	63
5.3.1	Vectors	63
5.3.2	Factors	68
5.3.3	Data Frames	74
5.3.4	Matrices	78
5.3.5	Arrays	82
5.3.6	Lists	83
5.4	Saving Your Work	88
5.5	Comments to Document Your Programs	90
5.6	Comments to Document Your Objects	91
5.7	Controlling Functions (Procedures)	92
5.7.1	Controlling Functions with Arguments	92
5.7.2	Controlling Functions with Objects	95
5.7.3	Controlling Functions with Formulas	96
5.7.4	Controlling Functions with an Object's Class	96

5.7.5	Controlling Functions with Extractor Functions	99
5.8	How Much Output There?	100
5.9	Writing Your Own Functions (Macros)	105
5.10	Controlling Program Flow	107
5.11	R Program Demonstrating Programming Basics	108
6	Data Acquisition	115
6.1	Manual Data Entry Using the R Data Editor	115
6.2	Reading Delimited Text Files	117
6.2.1	Reading Comma-Delimited Text Files	118
6.2.2	Reading Tab-Delimited Text Files	120
6.2.3	Reading Text from a Web Site	121
6.2.4	Reading Text from the Clipboard	122
6.2.5	Missing Values for Character Variables	122
6.2.6	Trouble with Tabs	124
6.2.7	Skipping Variables in Delimited Text Files	125
6.2.8	Reading Character Strings	126
6.2.9	Example Programs for Reading Delimited Text Files	126
6.3	Reading Text Data Within a Program	129
6.3.1	The Easy Approach	130
6.3.2	The More General Approach	131
6.3.3	Example Programs for Reading Text Data Within a Program	132
6.4	Reading Multiple Observations per Line	134
6.4.1	Example Programs for Reading Multiple Observations per Line	136
6.5	Reading Data from the Keyboard	138
6.6	Reading Fixed-Width Text Files, One Record per Case	138
6.6.1	Reading Data Using Macro Substitution	141
6.6.2	Example Programs for Reading Fixed-Width Text Files, One Record per Case	142
6.7	Reading Fixed-Width Text Files, Two or More Records per Case	143
6.7.1	Example Programs to Read Fixed-Width Text Files with Two Records per Case	145
6.8	Reading Excel Files	146
6.8.1	Example Programs for Reading Excel Files	147
6.9	Reading from Relational Databases	148
6.10	Reading Data from SAS	149
6.10.1	Example Programs to Write Data from SAS and Read It into R	150
6.11	Reading Data from SPSS	151
6.11.1	Example Programs for Reading Data from SPSS	152

6.12	Writing Delimited Text Files	153
6.12.1	Example Programs for Writing Delimited Text Files .	154
6.13	Viewing a Text File	156
6.14	Writing Excel Files	156
6.14.1	Example Programs for Writing Excel Files	157
6.15	Writing to Relational Databases	158
6.16	Writing Data to SAS and SPSS	158
6.16.1	Example Programs to Write Data to SAS and SPSS .	159
7	Selecting Variables	161
7.1	Selecting Variables in SAS and SPSS	161
7.2	Subscripting	162
7.3	Selecting Variables by Index Number	163
7.4	Selecting Variables by Column Name	166
7.5	Selecting Variables Using Logic	167
7.6	Selecting Variables by String Search (varname: or varname1-varnameN)	169
7.7	Selecting Variables Using \$ Notation	172
7.8	Selecting Variables by Simple Name	172
7.8.1	The <code>attach</code> Function	173
7.8.2	The <code>with</code> Function	174
7.8.3	Using Short Variable Names in Formulas	174
7.9	Selecting Variables with the <code>subset</code> Function	175
7.10	Selecting Variables by List Subscript	176
7.11	Generating Indices A to Z from Two Variable Names	176
7.11.1	Selecting Numeric or Character Variables	177
7.12	Saving Selected Variables to a New Data Set	180
7.13	Example Programs for Variable Selection	180
7.13.1	SAS Program to Select Variables	181
7.13.2	SPSS Program to Select Variables	181
7.13.3	R Program to Select Variables	182
8	Selecting Observations	187
8.1	Selecting Observations in SAS and SPSS	187
8.2	Selecting All Observations	188
8.3	Selecting Observations by Index Number	189
8.4	Selecting Observations Using Random Sampling	191
8.5	Selecting Observations by Row Name	193
8.6	Selecting Observations Using Logic	194
8.7	Selecting Observations by String Search	198
8.8	Selecting Observations with the <code>subset</code> Function	200
8.9	Generating Indices A to Z from Two Row Names	200
8.10	Variable Selection Methods with No Counterpart for Selecting Observations	201

8.11	Saving Selected Observations to a New Data Frame	201
8.12	Example Programs for Selecting Observations	202
8.12.1	SAS Program to Select Observations	202
8.12.2	SPSS Program to Select Observations	203
8.12.3	R Program to Select Observations	203
9	Selecting Variables and Observations	209
9.1	The <code>subset</code> Function	209
9.2	Subscripting with Logical Selections and Variable Names	211
9.3	Using Names to Select Both Observations and Variables	212
9.4	Using Numeric Index Values to Select Both Observations and Variables	213
9.5	Using Logic to Select Both Observations and Variables	213
9.6	Saving and Loading Subsets	214
9.7	Example Programs for Selecting Variables and Observations	215
9.7.1	SAS Program for Selecting Variables and Observations	215
9.7.2	SPSS Program for Selecting Variables and Observations	215
9.7.3	R Program for Selecting Variables and Observations	216
10	Data Management	219
10.1	Transforming Variables	219
10.1.1	Example Programs for Transforming Variables	223
10.2	Procedures or Functions? The <code>apply</code> Function Decides	225
10.2.1	Applying the <code>mean</code> Function	225
10.2.2	Finding N or NVALID	229
10.2.3	Standardizing and Ranking Variables	231
10.2.4	Applying Your Own Functions	233
10.2.5	Example Programs for Applying Statistical Functions	234
10.3	Conditional Transformations	237
10.3.1	The <code>ifelse</code> Function	237
10.3.2	Cutting Functions	241
10.3.3	Example Programs for Conditional Transformations	242
10.4	Multiple Conditional Transformations	246
10.4.1	Example Programs for Multiple Conditional Transformations	248
10.5	Missing Values	250
10.5.1	Substituting Means for Missing Values	252
10.5.2	Finding Complete Observations	253
10.5.3	When “99” Has Meaning	254
10.5.4	Example Programs to Assign Missing Values	255
10.6	Renaming Variables (and Observations)	258
10.6.1	Advanced Renaming Examples	260
10.6.2	Renaming by Index	261

10.6.3	Renaming by Column Name	262
10.6.4	Renaming Many Sequentially Numbered Variable Names	263
10.6.5	Renaming Observations	264
10.6.6	Example Programs for Renaming Variables	264
10.7	Recoding Variables	268
10.7.1	Recoding a Few Variables	269
10.7.2	Recoding Many Variables	269
10.7.3	Example Programs for Recoding Variables	272
10.8	Indicator or Dummy Variables	274
10.8.1	Example Programs for Indicator or Dummy Variables	277
10.9	Keeping and Dropping Variables	279
10.9.1	Example Programs for Keeping and Dropping Variables	280
10.10	Stacking/Concatenating/Adding Data Sets	281
10.10.1	Example Programs for Stacking/Concatenating/Adding Data Sets	283
10.11	Joining/Merging Data Sets	285
10.11.1	Example Programs for Joining/Merging Data Sets	288
10.12	Creating Summarized or Aggregated Data Sets	290
10.12.1	The <code>aggregate</code> Function	290
10.12.2	The <code>tapply</code> Function	292
10.12.3	Merging Aggregates with Original Data	294
10.12.4	Tabular Aggregation	296
10.12.5	The <code>plyr</code> and <code>reshape2</code> Packages	298
10.12.6	Comparing Summarization Methods	298
10.12.7	Example Programs for Aggregating/Summarizing Data	299
10.13	By or Split-File Processing	302
10.13.1	Example Programs for By or Split-File Processing	306
10.14	Removing Duplicate Observations	308
10.14.1	Completely Duplicate Observations	308
10.14.2	Duplicate Keys	311
10.14.3	Example Programs for Removing Duplicates	311
10.15	Selecting First or Last Observations per Group	314
10.15.1	Example Programs for Selecting Last Observation per Group	317
10.16	Transposing or Flipping Data Sets	319
10.16.1	Example Programs for Transposing or Flipping Data Sets	322
10.17	Reshaping Variables to Observations and Back	324
10.17.1	Summarizing/Aggregating Data Using <code>reshape2</code>	328
10.17.2	Example Programs for Reshaping Variables to Observations and Back	330
10.18	Sorting Data Frames	333

10.18.1	Example Programs for Sorting Data Sets	336
10.19	Converting Data Structures	338
10.19.1	Converting from Logical to Numeric Index and Back	341
10.20	Character String Manipulations	342
10.20.1	Example Programs for Character String Manipulation	349
10.21	Dates and Times	354
10.21.1	Calculating Durations	358
10.21.2	Adding Durations to Date–Time Variables	362
10.21.3	Accessing Date–Time Elements	362
10.21.4	Creating Date–Time Variables from Elements	363
10.21.5	Logical Comparisons with Date–Time Variables	364
10.21.6	Formatting Date–Time Output	364
10.21.7	Two-Digit Years	365
10.21.8	Date–Time Conclusion	366
10.21.9	Example Programs for Dates and Times	366
11	Enhancing Your Output	375
11.1	Value Labels or Formats (and Measurement Level)	375
11.1.1	Character Factors	376
11.1.2	Numeric Factors	378
11.1.3	Making Factors of Many Variables	380
11.1.4	Converting Factors to Numeric or Character Variables	383
11.1.5	Dropping Factor Levels	384
11.1.6	Example Programs for Value Labels	385
11.1.7	R Program to Assign Value Labels and Factor Status	386
11.2	Variable Labels	389
11.2.1	Other Packages That Support Variable Labels	393
11.2.2	Example Programs for Variable Labels	393
11.3	Output for Word Processing and Web Pages	395
11.3.1	The <code>xtable</code> Package	396
11.3.2	Other Options for Formatting Output	398
11.3.3	Example Program for Formatting Output	398
12	Generating Data	401
12.1	Generating Numeric Sequences	402
12.2	Generating Factors	403
12.3	Generating Repetitious Patterns (Not Factors)	404
12.4	Generating Values for Reading Fixed-Width Files	405
12.5	Generating Integer Measures	406
12.6	Generating Continuous Measures	408
12.7	Generating a Data Frame	409
12.8	Example Programs for Generating Data	411
12.8.1	SAS Program for Generating Data	411
12.8.2	SPSS Program for Generating Data	412
12.8.3	R Program for Generating Data	413

13 Managing Your Files and Workspace	417
13.1 Loading and Listing Objects	417
13.2 Understanding Your Search Path	421
13.3 Attaching Data Frames	422
13.4 Loading Packages	424
13.5 Attaching Files	426
13.6 Removing Objects from Your Workspace	427
13.7 Minimizing Your Workspace	430
13.8 Setting Your Working Directory	430
13.9 Saving Your Workspace	431
13.9.1 Saving Your Workspace Manually	431
13.9.2 Saving Your Workspace Automatically	431
13.9.3 Getting Operating Systems to Show You .RData Files ..	432
13.9.4 Organizing Projects with Windows Shortcuts	432
13.10 Saving Your Programs and Output	433
13.11 Saving Your History	433
13.12 Large Data Set Considerations	435
13.13 Example R Program for Managing Files and Workspace	435
14 Graphics Overview	441
14.1 Dynamic Visualization	441
14.2 SAS/GRAPH	442
14.3 SPSS Graphics	442
14.4 R Graphics	443
14.5 The Grammar of Graphics	444
14.6 Other Graphics Packages	445
14.7 Graphics Archives	445
14.8 Graphics Demonstrations	445
14.9 Graphics Procedures and Graphics Systems	447
14.10 Graphics Devices	448
15 Traditional Graphics	451
15.1 The <code>plot</code> Function	451
15.2 Bar Plots	453
15.2.1 Bar Plots of Counts	453
15.2.2 Bar Plots for Subgroups of Counts	457
15.2.3 Bar Plots of Means	458
15.3 Adding Titles, Labels, Colors, and Legends	459
15.4 Graphics Parameters and Multiple Plots on a Page	462
15.5 Pie Charts	465
15.6 Dot Charts	466
15.7 Histograms	466
15.7.1 Basic Histograms	467
15.7.2 Histograms Stacked	469

15.7.3	Histograms Overlaid	470
15.8	Normal QQ Plots	475
15.9	Strip Charts	476
15.10	Scatter and Line Plots	480
15.10.1	Scatter Plots with Jitter	483
15.10.2	Scatter Plots with Large Data Sets	483
15.10.3	Scatter Plots with Lines	486
15.10.4	Scatter Plots with Linear Fit by Group	487
15.10.5	Scatter Plots by Group or Level (Coplots)	489
15.10.6	Scatter Plots with Confidence Ellipse	489
15.10.7	Scatter Plots with Confidence and Prediction Intervals ..	490
15.10.8	Plotting Labels Instead of Points	496
15.10.9	Scatter Plot Matrices	498
15.11	Dual-Axis Plots	500
15.12	Box Plots	502
15.13	Error Bar Plots	505
15.14	Interaction Plots	505
15.15	Adding Equations and Symbols to Graphs	505
15.16	Summary of Graphics Elements and Parameters	507
15.17	Plot Demonstrating Many Modifications	507
15.18	Example Traditional Graphics Programs	508
15.18.1	SAS Program for Traditional Graphics	510
15.18.2	SPSS Program for Traditional Graphics	510
15.18.3	R Program for Traditional Graphics	511
16	Graphics with ggplot2	521
16.1	Introduction	521
16.1.1	Overview of <code>qplot</code> and <code>ggplot</code>	522
16.1.2	Missing Values	524
16.1.3	Typographic Conventions	525
16.2	Bar Plots	526
16.3	Pie Charts	528
16.4	Bar Plots for Groups	530
16.5	Plots by Group or Level	531
16.6	Presummarized Data	532
16.7	Dot Charts	534
16.8	Adding Titles and Labels	535
16.9	Histograms and Density Plots	536
16.9.1	Histograms	536
16.9.2	Density Plots	537
16.9.3	Histograms with Density Overlaid	538
16.9.4	Histograms for Groups, Stacked	539
16.9.5	Histograms for Groups, Overlaid	540
16.10	Normal QQ Plots	540
16.11	Strip Plots	541

16.12	Scatter Plots and Line Plots	544
16.12.1	Scatter Plots with Jitter	547
16.12.2	Scatter Plots for Large Data Sets	548
16.12.3	Scatter Plots with Fit Lines	553
16.12.4	Scatter Plots with Reference Lines	555
16.12.5	Scatter Plots with Labels Instead of Points	557
16.12.6	Changing Plot Symbols	559
16.12.7	Scatter Plot with Linear Fits by Group	560
16.12.8	Scatter Plots Faceted by Groups	561
16.12.9	Scatter Plot Matrix	562
16.13	Box Plots	564
16.14	Error Bar Plots	567
16.15	Geographic Maps	568
16.15.1	Finding and Converting Maps	573
16.16	Logarithmic Axes	574
16.17	Aspect Ratio	575
16.18	Multiple Plots on a Page	575
16.19	Saving <code>ggplot2</code> Graphs to a File	577
16.20	An Example Specifying All Defaults	578
16.21	Summary of Graphics Elements and Parameters	579
16.22	Example Programs for Grammar of Graphics	580
16.22.1	SPSS Program for Graphics Production Language	580
16.22.2	R Program for <code>ggplot2</code>	583
17	Statistics	599
17.1	Scientific Notation	599
17.2	Descriptive Statistics	600
17.2.1	The <code>Deducer</code> <code>frequencies</code> Function	600
17.2.2	The <code>Hmisc</code> <code>describe</code> Function	601
17.2.3	The <code>summary</code> Function	603
17.2.4	The <code>table</code> Function and Its Relatives	604
17.2.5	The <code>mean</code> Function and Its Relatives	606
17.3	Cross-Tabulation	607
17.3.1	The <code>CrossTable</code> Function	607
17.3.2	The <code>table</code> and <code>chisq.test</code> Functions	608
17.4	Correlation	612
17.4.1	The <code>cor</code> Function	614
17.5	Linear Regression	616
17.5.1	Plotting Diagnostics	620
17.5.2	Comparing Models	621
17.5.3	Making Predictions with New Data	622
17.6	t-Test: Independent Groups	622
17.7	Equality of Variance	624
17.8	t-Test: Paired or Repeated Measures	625

17.9	Wilcoxon–Mann–Whitney Rank Sum: Independent Groups	626
17.10	Wilcoxon Signed-Rank Test: Paired Groups	627
17.11	Sign Test: Paired Groups	628
17.12	Analysis of Variance	630
17.13	Sums of Squares	633
17.14	The Kruskal–Wallis Test	635
17.15	Example Programs for Statistical Tests	637
	17.15.1 SAS Program for Statistical Tests	637
	17.15.2 SPSS Program for Statistical Tests	639
	17.15.3 R Program for Statistical Tests	641
18	Conclusion	647
	References	663
	Index	669



<http://www.springer.com/978-1-4614-0684-6>

R for SAS and SPSS Users

Muenchen, R.A.

2011, XXVIII, 686 p. 118 illus., 32 illus. in color.,

Hardcover

ISBN: 978-1-4614-0684-6