

*Christopher Gandrud*

---

# *Reproducible Research with R and RStudio*

*Chapman & Hall/CRC Press*

*Expected Publication: August 2013*

<b>I</b>	<b>Getting Started</b>	<b>1</b>
<b>1</b>	<b>Introducing Reproducible Research</b>	<b>3</b>
	<i>and</i>	
1.1	What is reproducible research? . . . . .	3
1.2	Why should research be reproducible? . . . . .	5
1.2.1	For Science . . . . .	5
1.2.2	For You . . . . .	6
1.3	Who should read this book? . . . . .	7
1.3.1	Academic Researchers . . . . .	8
1.3.2	Students . . . . .	8
1.3.3	Instructors . . . . .	8
1.3.4	Editors . . . . .	9
1.3.5	Private sector researchers . . . . .	9
1.4	The Tools of Reproducible Research . . . . .	9
1.5	Why use R, knitr, and RStudio for reproducible research? . . . . .	10
1.5.1	Installing the main software . . . . .	12
1.6	Book overview . . . . .	14
1.6.1	How to read this book . . . . .	15
1.6.2	Reproduce this book . . . . .	15
1.6.3	Contents overview . . . . .	15
<b>2</b>	<b>Getting Started with Reproducible Research</b>	<b>17</b>
2.1	The Big Picture: A workflow for reproducible research . . . . .	17
2.1.1	Reproducible Theory . . . . .	18
2.2	Practical tips for reproducible research . . . . .	20
2.2.1	Document everything! . . . . .	20
2.2.2	Everything is a (text) file . . . . .	21
2.2.3	All files should be human readable . . . . .	22
2.2.4	Explicitly tie your files together . . . . .	24
2.2.5	Have a plan to organize, store, & make your files available . . . . .	26
<b>3</b>	<b>Getting Started with R, RStudio, and knitr</b>	<b>27</b>
3.1	Using R: the basics . . . . .	27
3.1.1	Objects . . . . .	28
3.1.2	Component Selection . . . . .	34
3.1.3	Subscripts . . . . .	36
3.1.4	Functions and commands . . . . .	37
3.1.5	Arguments . . . . .	38
3.1.6	The Workspace & History . . . . .	39
3.1.7	Global R options . . . . .	41
3.1.8	Installing new packages and loading commands . . . . .	42

3.2	Using RStudio . . . . .	43
3.3	Using knitr: the basics . . . . .	44
3.3.1	What <i>knitr</i> does . . . . .	45
3.3.2	File extensions . . . . .	45
3.3.3	Code Chunks . . . . .	47
3.3.4	Global chunk options . . . . .	49
3.3.5	knitr package options . . . . .	51
3.3.6	Hooks . . . . .	52
3.3.7	<i>knitr</i> & RStudio . . . . .	52
3.3.8	<i>knitr</i> & R . . . . .	55
<b>4</b>	<b>Getting Started with File Management</b>	<b>59</b>
4.1	File paths & naming conventions . . . . .	60
4.1.1	Root directories . . . . .	60
4.1.2	Subdirectories & parent directories . . . . .	60
4.1.3	Spaces in directory & file names . . . . .	61
4.1.4	Working directories . . . . .	61
4.2	Organizing your research project . . . . .	63
4.3	Setting directories as RStudio Projects . . . . .	64
4.4	R file manipulation commands . . . . .	64
4.5	Unix-like shell commands for file management . . . . .	67
4.6	File navigation in RStudio . . . . .	72
<b>II</b>	<b>Data Gathering and Storage</b>	<b>73</b>
<b>5</b>	<b>Storing, Collaborating, Accessing Files, Versioning</b>	<b>75</b>
5.1	Saving data in reproducible formats . . . . .	76
5.2	Storing your files in the cloud: Dropbox . . . . .	77
5.2.1	Storage . . . . .	78
5.2.2	Accessing Data . . . . .	78
5.2.3	Collaboration . . . . .	79
5.2.4	Version control . . . . .	80
5.3	Storing your files in the cloud: GitHub . . . . .	80
5.3.1	Setting up GitHub: Basic . . . . .	83
5.3.2	Version Control with Git . . . . .	83
5.3.3	Remote Storage on GitHub . . . . .	91
5.3.4	Accessing on GitHub . . . . .	93
5.3.4.1	Collaboration with GitHub . . . . .	93
5.3.5	Summing up the GitHub workflow . . . . .	94
5.4	RStudio & GitHub . . . . .	94
5.4.1	Setting Up Git/GitHub with Projects . . . . .	95
5.4.2	Using Git in RStudio projects . . . . .	96

<b>6</b>	<b>Gathering Data with R</b>	<b>99</b>
6.1	Organize your data gathering: makefiles . . . . .	99
6.1.1	R Make-like files . . . . .	100
6.1.2	GNU Make . . . . .	101
6.1.2.1	Example Makefile . . . . .	102
6.1.2.2	Makefiles and RStudio Projects . . . . .	106
6.1.2.3	Other information about Makefiles . . . . .	107
6.2	Importing locally stored data sets . . . . .	107
6.3	Importing data sets from the internet . . . . .	108
6.3.1	Data from non-secure ( <b>http</b> ) URLs . . . . .	108
6.3.2	Data from secure ( <b>https</b> ) URLs . . . . .	109
6.3.3	Compressed data stored online . . . . .	111
6.3.4	Data APIs & feeds . . . . .	112
6.4	Advanced Automatic Data Gathering: web scraping . . . . .	114
<b>7</b>	<b>Preparing Data for Analysis</b>	<b>117</b>
7.1	Cleaning data for merging . . . . .	117
7.1.1	Get a handle on your data . . . . .	117
7.1.2	Reshaping Data . . . . .	118
7.1.3	Renaming variables . . . . .	121
7.1.4	Ordering data . . . . .	122
7.1.5	Subsetting data . . . . .	123
7.1.6	Recoding string/numeric variables . . . . .	125
7.1.7	Creating new variables from old . . . . .	126
7.1.8	Changing variables types . . . . .	129
7.2	Merging data sets . . . . .	129
7.2.1	Binding . . . . .	129
7.2.2	The merge command . . . . .	130
7.2.3	Duplicate values . . . . .	131
7.2.4	Duplicate columns . . . . .	132
<b>III</b>	<b>Analysis and Results</b>	<b>137</b>
<b>8</b>	<b>Statistical Modelling and knitr</b>	<b>139</b>
8.1	Incorporating analyses into the markup . . . . .	140
8.1.1	Full code chunks . . . . .	140
8.1.2	Showing code & results inline . . . . .	142
8.1.2.1	LaTeX . . . . .	142
8.1.2.2	Markdown . . . . .	143
8.1.3	Dynamically including non-R code in code chunks . . . . .	144
8.2	Dynamically including modular analysis files . . . . .	145
8.2.1	Source from a local file . . . . .	146
8.2.2	Source from a non-secure URL ( <b>http</b> ) . . . . .	147
8.2.3	Source from a secure URL ( <b>https</b> ) . . . . .	148
8.3	Computationally intensive analyses . . . . .	149

<b>9</b>	<b>Showing Results with Tables</b>	<b>151</b>
9.0.1	Basic <i>knitr</i> syntax for tables . . . . .	152
9.1	Table Basics . . . . .	152
9.1.1	Tables in LaTeX . . . . .	152
9.1.2	Tables in Markdown/HTML . . . . .	157
9.2	Creating tables from R objects . . . . .	161
9.2.1	<i>xtable</i> & <i>apsrtable</i> basics with supported class objects	161
9.2.1.1	<i>apsrtable</i> for LaTeX . . . . .	164
9.2.2	<i>xtable</i> with non-supported class objects . . . . .	167
9.2.3	Creating variable description documents with <i>xtable</i> .	170
<b>10</b>	<b>Showing Results with Figures</b>	<b>173</b>
10.1	Including non-knitted graphics . . . . .	173
10.1.1	Including graphics in LaTeX . . . . .	174
10.1.2	Including graphics in Markdown/HTML . . . . .	175
10.2	Basic <i>knitr</i> figure options . . . . .	177
10.2.1	Chunk options . . . . .	177
10.2.2	Global options . . . . .	179
10.3	Knitting R's default graphics . . . . .	179
10.4	Including <i>ggplot2</i> graphics . . . . .	182
10.4.1	Showing Regression Results with Caterpillar Plots . .	186
10.5	JavaScript graphs with <i>googleVis</i> . . . . .	190
<b>IV</b>	<b>Presentation Documents</b>	<b>195</b>
<b>11</b>	<b>Presenting with LaTeX</b>	<b>197</b>
11.1	The Basics . . . . .	197
11.1.1	Getting Started with LaTeX Editors . . . . .	197
11.1.2	Basic LaTeX command syntax . . . . .	198
11.1.3	The LaTeX preamble & body . . . . .	198
11.1.4	Headings . . . . .	202
11.1.5	Paragraphs & spacing . . . . .	202
11.1.6	Horizontal lines . . . . .	203
11.1.7	Text formatting . . . . .	203
11.1.8	Math . . . . .	204
11.1.9	Lists . . . . .	205
11.1.10	Footnotes . . . . .	206
11.1.11	Cross-references . . . . .	206
11.2	Bibliographies with BibTeX . . . . .	206
11.2.1	The <i>.bib</i> file . . . . .	207
11.2.2	Including citations in a LaTeX document . . . . .	208
11.2.3	Generating a BibTeX file of R packages . . . . .	209
11.3	Presentations with LaTeX Beamer . . . . .	211
11.3.1	Beamer basics . . . . .	212
11.3.2	<i>knitr</i> with LaTeX slideshows . . . . .	215

<b>12 Large LaTeX Documents: Theses, Books, &amp; Batch Reports</b>	<b>217</b>
12.1 Planning large documents . . . . .	217
12.2 Large documents with traditional LaTeX . . . . .	218
12.2.1 Inputting/including children . . . . .	219
12.2.2 Other common features of large documents . . . . .	220
12.3 <i>knitr</i> and large documents . . . . .	221
12.3.1 The parent document . . . . .	221
12.3.2 Knitting child documents . . . . .	222
12.4 Child documents in a different markup language . . . . .	223
12.5 Creating batch reports . . . . .	224
<b>13 Presenting on the Web with Markdown</b>	<b>229</b>
13.1 The Basics . . . . .	229
13.1.1 Getting Started with Markdown Editors . . . . .	230
13.1.2 Preamble and document structure . . . . .	230
13.1.3 Headers . . . . .	233
13.1.4 Horizontal Lines . . . . .	233
13.1.5 Paragraphs and new lines . . . . .	233
13.1.6 Italics and bold . . . . .	234
13.1.7 Links . . . . .	234
13.1.8 Special characters and font customization . . . . .	234
13.1.9 Lists . . . . .	234
13.1.10 Escape characters . . . . .	235
13.1.11 Math with MathJax . . . . .	235
13.2 Markdown with Pandoc and Custom CSS . . . . .	236
13.2.1 Pandoc . . . . .	236
13.2.2 CSS style files and Markdown . . . . .	239
13.3 Presentations with <i>slidify</i> . . . . .	242
13.4 Publishing Markdown Documents . . . . .	247
13.4.1 Stand alone HTML files . . . . .	247
13.4.2 Hosting webpages with Dropbox . . . . .	247
13.4.3 GitHub Pages . . . . .	247
<b>14 Conclusion</b>	<b>249</b>
14.1 Citing reproducible research . . . . .	249
14.2 Licensing your reproducible research . . . . .	250
14.3 Sharing your code in packages . . . . .	251
14.4 Is it possible to completely future proof your research? . . . .	252