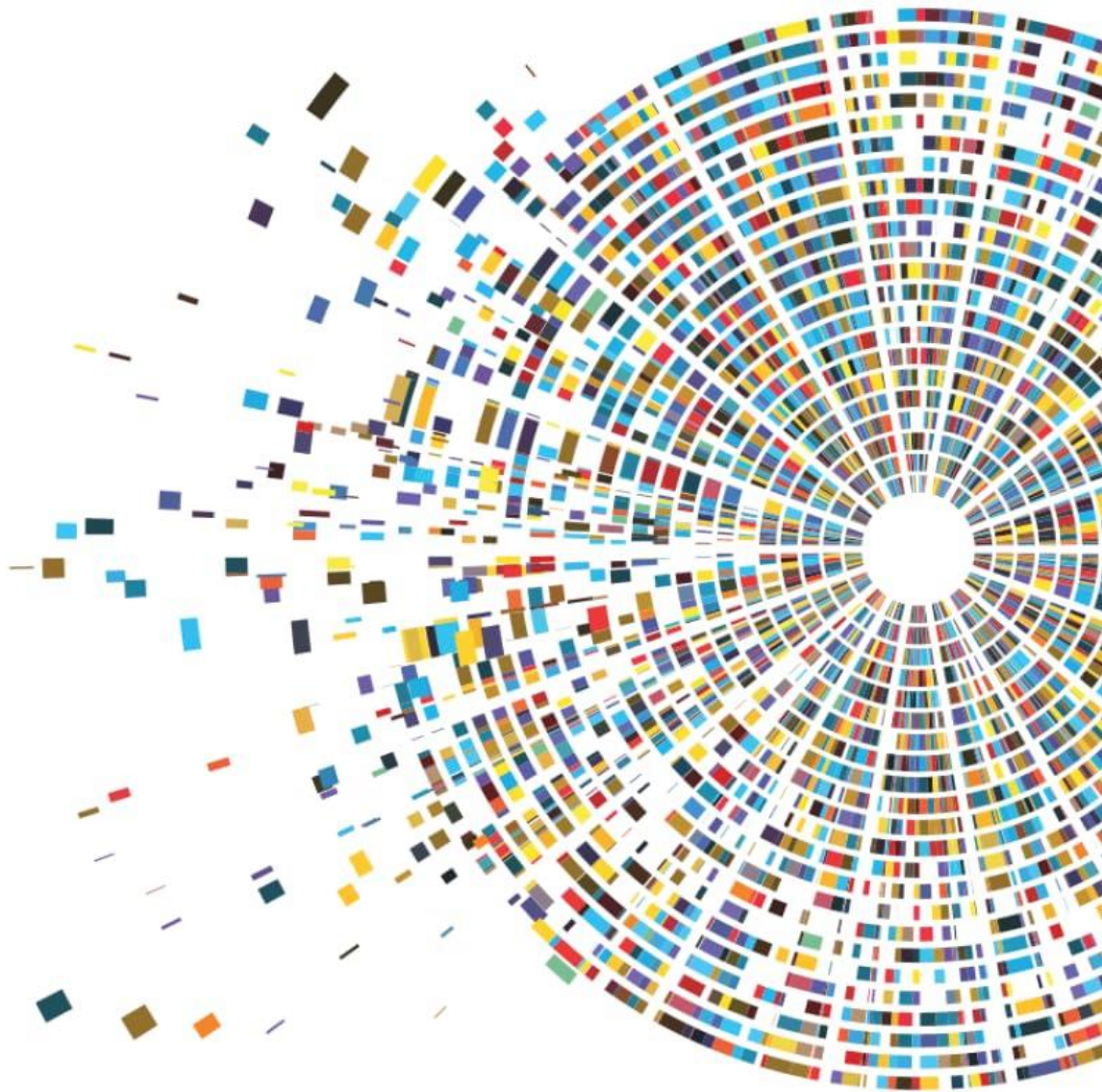


R Crash Course for Biologists

An introduction to R
for bioinformatics
and biostatistics



Robert I. Colautti
Queen's University

Contents

Preface	9
0.1 Is this book for you?	10
0.2 Why R?	12
0.3 Advice	13
0.4 Learn By Doing!	16
0.5 What to Expect	17
0.6 Translational Coding	18
 Setup	 23
0.7 R	23
0.8 R Studio	23
0.9 Console	24
0.10 Packages	25
 1 R Fundamentals	 27
1.1 Overview	27
1.2 R Basics	28

1.3	Use ? for HELP	37
1.4	Random Numbers	39
1.5	Repeat, Replicate & Sample	42
1.6	set.seed()	44
1.7	Combining objects	46
1.8	Sequence	47
1.9	Rows and Columns	48
1.10	Data Types	50
1.11	Objects & Variables	52
1.12	Matrix Algebra	60
1.13	PCA	67
1.14	Lists	69
1.15	print() and paste()	73
1.16	External Files	75
1.17	Other Functions	86
1.18	Tidyverse	90
1.19	Save	91
1.20	Packages	92
1.21	Readable code	95

2 Flow Control 97

2.1	Overview	97
2.2	if(){}	98

CONTENTS	3
2.3 ifelse()	98
2.4 nested if	99
2.5 for loop	100
2.6 Nested Loops	103
2.7 while loop	104
2.8 Modulo	106
2.9 Faster loops	107
3 Quick Visualizations	111
3.1 Overview	111
3.2 Graphical Concepts	113
3.3 Getting Started	119
3.4 Basic Graphs	122
4 Basic Customization	133
4.1 Overview	133
4.2 Setup	133
4.3 binwidth	134
4.4 size	135
4.5 alpha	138
4.6 colour (or color)	138
4.7 Colour with rgb()	140
4.8 fill	143
4.9 position	144

4.10	shape	144
4.11	lab, xlab, and ylab	147
4.12	labs	148
4.13	Themes and Geoms	149
4.14	Basic Multi-Plot Graphs	156
4.15	Graph output	158
4.16	Practice	160
5	Advanced Visualizations	161
5.1	Overview	161
5.2	Getting Started	162
5.3	Rules of thumb	162
5.4	Example	169
5.5	Measuring Selection	173
5.6	Distribution Plots	176
5.7	Full ggplot	184
5.8	Bootstrap	185
5.9	Plot data	187
6	Multi-plot Graphs	193
6.1	Overview	193
6.2	Setup	193
6.3	facets	194
6.4	gridExtra package	196

6.5	grid package	198
6.6	Further Reading	201
7	Regular Expressions	203
7.1	Overview	203
7.2	Functions	205
7.3	Wildcards	208
7.4	New Lines	212
7.5	Special characters	213
7.6	[] (set)	219
7.7	^ (start or negate) and \$ (end)	220
7.8	() (capture)	221
7.9	Scraping	222
7.10	Examples	225
7.11	Transect Data	225
7.12	Genbank	226
7.13	Solutions	229
7.14	More Exercises	231
8	Data Science	233
8.1	Overview	233
8.2	Setup	235
8.3	2D Data Wrangling	236
8.4	Join datasets	252

8.5	<i>Wide vs Long data</i>	255
8.6	Missing Data	257
8.7	Naughty Data	261
8.8	Dates	263

9 R Markdown **273**

9.1	Overview	273
9.2	Setup	274
9.3	Cheat Sheet	274
9.4	Create	274
9.5	YAML Header	276
9.6	Markdown Elements	276
9.7	Basic elements:	277
9.8	Headers	278
9.9	Other Elements	278
9.10	Links	278
9.11	Images	279
9.12	Lists	280
9.13	Tables	282
9.14	Tables	282
9.15	Embed R Code	283
9.16	Dynamic tables	285
9.17	Embed Graphs	286

9.18 Content as tabs	287
9.19 Equations	289

10 Custom Functions **293**

10.1 Overview	293
10.2 General form:	294
10.3 Example function	295
10.4 Local vs Global	296
10.5 Run custom functions	297
10.6 Annotation	299
10.7 Verbose parameter	301
10.8 External files	302

11 Conclusion **305**

11.1 You made it!	305
11.2 What next?	308
11.3 Support Open & Accessible Science	310
11.4 Picture a Coder	311