

# Contents

<b>Foreword</b>	<b>v</b>
<b>About the Author</b>	<b>vii</b>
<b>Acknowledgements</b>	<b>ix</b>
<b>Preface</b>	<b>xi</b>
<b>Contents</b>	<b>xix</b>
<b>I Introduction</b>	<b>1</b>
1 The Big Picture with Kondratiev and Kardashev	3
2 The Scientific Method and Data	9
3 Conventions	15
<b>II Starting with R and Elements of Statistics</b>	<b>25</b>
4 The Basics of R	27
4.1 Getting Started with R . . . . .	30
4.2 Variables . . . . .	34
4.3 Data Types . . . . .	36
4.3.1 The Elementary Types . . . . .	36
4.3.2 Vectors . . . . .	38
4.3.2.1 Creating Vectors . . . . .	38
4.3.3 Accessing Data from a Vector . . . . .	38

## CONTENTS

---

4.3.3.1	Vector Arithmetic . . . . .	39
4.3.3.2	Vector Recycling . . . . .	39
4.3.3.3	Reordering and Sorting . . . . .	40
4.3.4	Matrices . . . . .	41
4.3.4.1	Creating Matrices . . . . .	41
4.3.4.2	Naming Rows and Columns . . . . .	43
4.3.4.3	Access Subsets of a Matrix . . . . .	43
4.3.4.4	Matrix Arithmetic . . . . .	44
4.3.5	Arrays . . . . .	50
4.3.5.1	Creating and Accessing Arrays . . . . .	50
4.3.5.2	Naming Elements of Arrays . . . . .	51
4.3.5.3	Manipulating Arrays . . . . .	51
4.3.5.4	Applying Functions over Arrays . . . . .	52
4.3.6	Lists . . . . .	53
4.3.6.1	Creating Lists . . . . .	53
4.3.6.2	Naming Elements of Lists . . . . .	54
4.3.6.3	List Manipulations . . . . .	55
4.3.7	Factors . . . . .	60
4.3.7.1	Creating Factors . . . . .	60
4.3.7.2	Ordering Factors . . . . .	62
4.3.8	Data Frames . . . . .	64
4.3.8.1	Introduction to Data Frames . . . . .	64
4.3.8.2	Accessing Information from a Data Frame . . . . .	65
4.3.8.3	Editing Data in a Data Frame . . . . .	67
4.3.8.4	Modifying Data Frames . . . . .	68
4.3.9	Strings or the Character-type . . . . .	72
4.4	Operators . . . . .	75
4.4.1	Arithmetic Operators . . . . .	75
4.4.2	Relational Operators . . . . .	76
4.4.3	Logical Operators . . . . .	76
4.4.4	Assignment Operators . . . . .	78
4.4.5	Other Operators . . . . .	80
4.5	Flow Control Statements . . . . .	83
4.5.1	Choices . . . . .	83
4.5.1.1	The if-Statement . . . . .	83
4.5.1.2	The Vectorised If-statement . . . . .	84
4.5.1.3	The Switch-statement . . . . .	85
4.5.2	Loops . . . . .	85
4.5.2.1	The For Loop . . . . .	86
4.5.2.2	Repeat . . . . .	87
4.5.2.3	While . . . . .	88

---

4.5.2.4	Loop Control Statements . . . . .	88
4.6	Functions . . . . .	91
4.6.1	Built-in Functions . . . . .	91
4.6.2	Help with Functions . . . . .	91
4.6.3	User-defined Functions . . . . .	92
4.6.4	Changing Functions . . . . .	93
4.6.5	Creating Function with Default Arguments . . . . .	94
4.7	Packages . . . . .	96
4.7.1	Discovering Packages in R . . . . .	96
4.7.2	Managing Packages in R . . . . .	98
4.8	Selected Data Interfaces . . . . .	100
4.8.1	CSV Files . . . . .	100
4.8.2	Excel Files . . . . .	105
4.8.3	Databases . . . . .	105
<b>5</b>	<b>Lexical Scoping and Environments</b>	<b>109</b>
5.1	Environments in R . . . . .	110
5.2	Lexical Scoping in R . . . . .	112
<b>6</b>	<b>The Implementation of OO</b>	<b>117</b>
6.1	Base Types . . . . .	120
6.2	S3 Objects . . . . .	122
6.2.1	Creating S3 Objects . . . . .	126
6.2.2	Creating Generic Methods . . . . .	128
6.2.3	Method Dispatch . . . . .	130
6.2.4	Group Generic Functions . . . . .	131
6.3	S4 Objects . . . . .	134
6.3.1	Creating S4 Objects . . . . .	134
6.3.2	Using S4 Objects . . . . .	136
6.3.3	Validation of Input . . . . .	140
6.3.4	Constructor functions . . . . .	143
6.3.5	The .Data slot . . . . .	144
6.3.6	Recognising Objects, Generic Functions, and Methods . . .	145
6.3.7	Creating S4 Generics . . . . .	147
6.3.8	Method Dispatch . . . . .	149
6.4	The Reference Class, reclass, RC or R5 Model . . . . .	151
6.4.1	Creating RC Objects . . . . .	151
6.4.2	Important Methods and Attributes . . . . .	157
6.5	Conclusions about the OO Implementation . . . . .	159

## CONTENTS

---

<b>7</b>	<b>Tidy R with the Tidyverse</b>	<b>161</b>
7.1	The Philosophy of the Tidyverse . . . . .	162
7.2	Packages in the Tidyverse . . . . .	166
7.2.1	The Core Tidyverse . . . . .	167
7.2.2	The Non-core Tidyverse . . . . .	168
7.3	Working with the Tidyverse . . . . .	170
7.3.1	Tibbles . . . . .	170
7.3.2	Piping with R . . . . .	176
7.3.3	Attention Points When Using the Pipe . . . . .	177
7.3.4	Advanced Piping . . . . .	179
7.3.4.1	The Dollar Pipe . . . . .	179
7.3.4.2	The T-Pipe . . . . .	181
7.3.4.3	The Assignment Pipe . . . . .	181
7.3.5	Conclusion . . . . .	183
<b>8</b>	<b>Elements of Descriptive Statistics</b>	<b>185</b>
8.1	Measures of Central Tendency . . . . .	186
8.1.1	Mean . . . . .	186
8.1.1.1	The Arithmetic Mean . . . . .	186
8.1.1.2	Generalised Means . . . . .	188
8.1.2	The Median . . . . .	190
8.1.3	The Mode . . . . .	191
8.2	Measures of Variation or Spread . . . . .	194
8.3	Measures of Covariation . . . . .	196
8.3.1	The Pearson Correlation . . . . .	196
8.3.2	The Spearman Correlation . . . . .	197
8.3.3	Chi-square Tests . . . . .	198
8.4	Distributions . . . . .	200
8.4.1	Normal Distribution . . . . .	200
8.4.2	Binomial Distribution . . . . .	204
8.5	Creating an Overview of Data Characteristics . . . . .	207
<b>9</b>	<b>Visualisation Methods</b>	<b>211</b>
9.1	Scatterplots . . . . .	213
9.2	Line Graphs . . . . .	216
9.3	Pie Charts . . . . .	218
9.4	Bar Charts . . . . .	220
9.5	Boxplots . . . . .	224
9.6	Violin Plots . . . . .	226
9.7	Histograms . . . . .	230
9.8	Plotting Functions . . . . .	232

---

9.9	Maps and Contour Plots . . . . .	233
9.10	Heat-maps . . . . .	235
9.11	Text Mining . . . . .	238
9.11.1	Word Clouds . . . . .	238
9.11.2	Word Associations . . . . .	244
9.12	Colours in R . . . . .	247
<b>10</b>	<b>Time Series Analysis</b>	<b>255</b>
10.1	Time Series in R . . . . .	256
10.1.1	The Basics of Time Series in R . . . . .	256
10.1.1.1	The Function <code>ts()</code> . . . . .	256
10.1.1.2	Multiple Time Series in one Object . . . . .	257
10.2	Forecasting . . . . .	260
10.2.1	Moving Average . . . . .	261
10.2.1.1	The Moving Average in R . . . . .	261
10.2.1.2	Testing the Accuracy of the Forecasts . . . . .	262
10.2.1.3	Basic Exponential Smoothing . . . . .	266
10.2.1.4	Holt-Winters Exponential Smoothing . . . . .	266
10.2.2	Seasonal Decomposition . . . . .	268
<b>11</b>	<b>Further, Reading</b>	<b>273</b>
<b>III</b>	<b>Data Import</b>	<b>277</b>
<b>12</b>	<b>A Short History of Modern Database Systems</b>	<b>279</b>
<b>13</b>	<b>RDBMS</b>	<b>285</b>
<b>14</b>	<b>SQL</b>	<b>291</b>
14.1	Designing the Database . . . . .	292
14.2	Building the Database Structure . . . . .	295
14.2.1	Installing a RDBMS . . . . .	295
14.2.2	Creating the Database . . . . .	298
14.2.3	Creating the Tables and Relations . . . . .	299
14.3	Adding Data to the Database . . . . .	306
14.4	Querying the Database . . . . .	310
14.4.1	The Basic Select Query . . . . .	310
14.4.2	More complex queries . . . . .	312
14.5	Modifying the Database Structure . . . . .	317
14.6	Selected Features of SQL . . . . .	323
14.6.1	Changing data . . . . .	323

## CONTENTS

---

14.6.2	Functions in SQL . . . . .	323
<b>15</b>	<b>Connecting R to an SQL Database</b>	<b>327</b>
<b>IV</b>	<b>Data Wrangling</b>	<b>333</b>
<b>16</b>	<b>Anonymous Data</b>	<b>339</b>
<b>17</b>	<b>Data Wrangling in the tidyverse</b>	<b>343</b>
17.1	Importing the Data . . . . .	344
17.1.1	Importing from an SQL RDBMS . . . . .	344
17.1.2	Importing Flat Files in the Tidyverse . . . . .	346
17.1.2.1	CSV Files . . . . .	349
17.1.2.2	Making Sense of Fixed-width Files . . . . .	351
17.2	Tidy Data . . . . .	356
17.3	Tidying Up Data with tidyr . . . . .	358
17.3.1	Splitting Tables . . . . .	359
17.3.2	Convert Headers to Data . . . . .	364
17.3.3	Spreading One Column Over Many . . . . .	366
17.3.4	Split One Columns into Many . . . . .	369
17.3.5	Merge Multiple Columns Into One . . . . .	370
17.3.6	Wrong Data . . . . .	371
17.4	SQL-like Functionality via dplyr . . . . .	372
17.4.1	Selecting Columns . . . . .	372
17.4.2	Filtering Rows . . . . .	373
17.4.3	Joining . . . . .	374
17.4.4	Mutating Data . . . . .	378
17.4.5	Set Operations . . . . .	383
17.5	String Manipulation in the tidyverse . . . . .	386
17.5.1	Basic String Manipulation . . . . .	387
17.5.2	Pattern Matching with Regular Expressions . . . . .	390
17.5.2.1	Regular Expressions . . . . .	392
17.5.2.2	Functions Using Regex . . . . .	399
17.6	Dates with lubridate . . . . .	407
17.6.1	ISO 8601 Format . . . . .	408
17.6.2	Time-zones . . . . .	411
17.6.3	Extract Date and Time Components . . . . .	412
17.6.4	Calculating with Date-times . . . . .	414
17.6.4.1	Durations . . . . .	415
17.6.4.2	Periods . . . . .	417

17.6.4.3	Intervals . . . . .	419
17.6.4.4	Rounding . . . . .	420
17.7	Factors with forcats . . . . .	422
<b>18</b>	<b>Dealing with Missing Data</b>	<b>431</b>
18.1	Reasons for Data to be Missing . . . . .	433
18.2	Methods to Handle Missing Data . . . . .	436
18.2.1	Alternative Solutions to Missing Data . . . . .	436
18.2.2	Predictive Mean Matching (PMM) . . . . .	439
18.3	R Packages to Deal with Missing Data . . . . .	441
18.3.1	mice . . . . .	441
18.3.2	missForest . . . . .	442
18.3.3	Hmisc . . . . .	443
<b>19</b>	<b>Data Binning</b>	<b>447</b>
19.1	What is Binning and Why Use It . . . . .	448
19.2	Tuning the Binning Procedure . . . . .	452
19.3	More complex Cases: Matrix Binning . . . . .	459
19.4	Weight of Evidence and Information Value . . . . .	467
19.4.1	Weight of Evidence (WOE) . . . . .	467
19.4.2	Information Value (IV) . . . . .	467
19.4.3	WOE and IV in R . . . . .	467
<b>20</b>	<b>Factoring Analysis and Principle Components</b>	<b>471</b>
20.1	Principle Components Analysis (PCA) . . . . .	472
20.2	Factor Analysis . . . . .	477
<b>V</b>	<b>Modelling</b>	<b>481</b>
<b>21</b>	<b>Regression Models</b>	<b>483</b>
21.1	Linear Regression . . . . .	484
21.2	Multiple Linear Regression . . . . .	487
21.2.1	Poisson Regression . . . . .	488
21.2.2	Non-linear Regression . . . . .	491
21.3	Performance of Regression Models . . . . .	494
21.3.1	Mean Square Error (MSE) . . . . .	494
21.3.2	<i>R</i> -Squared . . . . .	494
21.3.3	Mean Average Deviation (MAD) . . . . .	496

<b>22 Classification Models</b>	<b>497</b>
22.1 Logistic Regression . . . . .	498
22.2 Performance of Binary Classification Models . . . . .	500
22.2.1 The Confusion Matrix and Related Measures . . . . .	502
22.2.2 ROC . . . . .	505
22.2.3 The AUC . . . . .	507
22.2.4 The Gini Coefficient . . . . .	510
22.2.5 Kolmogorov-Smirnov (KS) for Logistic Regression . . . . .	511
22.2.6 Finding an Optimal Cut-off . . . . .	513
<b>23 Learning Machines</b>	<b>519</b>
23.1 Decision Tree . . . . .	522
23.1.1 Essential Background . . . . .	522
23.1.1.1 The Linear Additive Decision Tree . . . . .	522
23.1.1.2 The CART method . . . . .	523
23.1.1.3 Tree Pruning . . . . .	524
23.1.1.4 Classification Trees . . . . .	525
23.1.1.5 Binary Classification Trees . . . . .	526
23.1.2 Important Considerations . . . . .	528
23.1.2.1 Broadening the Scope . . . . .	528
23.1.2.2 Selected Issues . . . . .	530
23.1.3 Growing Trees with the Package rpart . . . . .	531
23.1.3.1 Getting Started with the Function rpart() . . . . .	531
23.1.3.2 Example of a Classification Tree with rpart . . . . .	533
23.1.3.3 Visualising a Decision Tree with rpart.plot . . . . .	535
23.1.3.4 Example of a Regression Tree with rpart . . . . .	538
23.1.4 Evaluating the Performance of a Decision Tree . . . . .	542
23.1.4.1 The Performance of the Regression Tree . . . . .	544
23.1.4.2 The Performance of the Classification Tree . . . . .	544
23.2 Random Forest . . . . .	548
23.3 Artificial Neural Networks (ANNs) . . . . .	555
23.3.1 The Basics of ANNs in R . . . . .	555
23.3.2 Neural Networks in R . . . . .	558
23.3.3 The Work-flow to for Fitting a NN . . . . .	560
23.3.3.1 Cross Validate the NN . . . . .	565
23.4 Support Vector Machine . . . . .	570
23.4.1 Fitting a SVM in R . . . . .	571
23.4.2 Optimizing the SVM . . . . .	573
23.5 Unsupervised Learning and Clustering . . . . .	575
23.5.1 k-Means Clustering . . . . .	576
23.5.1.1 k-Means Clustering in R . . . . .	576



---

23.5.1.2	PCA before clustering . . . . .	581
23.5.1.3	On the Relation Between PCA and k-Means . . . . .	588
23.5.2	Visualizing Clusters in Three Dimensions . . . . .	589
23.5.3	Fuzzy Clustering . . . . .	591
23.5.4	Hierarchical Clustering . . . . .	594
23.5.5	Other Clustering Methods . . . . .	596
<b>24</b>	<b>Towards a Tidy Modelling Cycle with modelr</b>	<b>599</b>
24.1	Adding predictions . . . . .	601
24.2	Adding Residuals . . . . .	602
24.3	Bootstrapping Data . . . . .	603
24.4	Other Functions of modelr . . . . .	604
<b>25</b>	<b>Model Validation</b>	<b>605</b>
25.1	Model Quality Measures . . . . .	607
25.2	Predictions and Residuals . . . . .	609
25.3	Bootstrapping . . . . .	610
25.3.1	Bootstrapping in base R . . . . .	611
25.3.2	Bootstrapping in the tidyverse with modelr . . . . .	612
25.4	Cross-Validation . . . . .	615
25.4.1	Elementary Cross Validation . . . . .	615
25.4.2	Monte Carlo cross validation . . . . .	619
25.4.3	<i>k</i> -Fold Cross Validation . . . . .	622
25.4.4	Comparing Cross Validation Methods . . . . .	624
25.5	Validation in a Broader Perspective . . . . .	626
<b>26</b>	<b>Labs</b>	<b>631</b>
26.1	Financial Analysis with quantmod . . . . .	632
26.1.1	The Basics of quantmod . . . . .	632
26.1.2	Types of Data Available in quantmod . . . . .	633
26.1.3	Plotting with quantmod . . . . .	634
26.1.4	The quantmod Data Structure . . . . .	637
26.1.4.1	Sub-setting by Time and Date . . . . .	638
26.1.4.2	Switching Time Scales . . . . .	639
26.1.4.3	Apply by Period . . . . .	640
26.1.5	Support Functions Supplied by quantmod . . . . .	641
26.1.6	Financial Modelling in quantmod . . . . .	642
26.1.6.1	Financial Models in quantmod . . . . .	642
26.1.6.2	A Simple Model with quantmod . . . . .	643
26.1.6.3	Testing the model robustness . . . . .	646

## CONTENTS

---

<b>27 Multi Criteria Decision Analysis (MCDA)</b>	<b>651</b>
27.1 What and Why . . . . .	652
27.2 General Work-flow . . . . .	654
27.3 Identify the Issue at Hand: Steps 1 and 2 . . . . .	658
27.4 Step 3: the Decision Matrix . . . . .	660
27.4.1 Construct a Decision Matrix . . . . .	660
27.4.2 Normalize the Decision Matrix . . . . .	662
27.5 Step 4: Delete Inefficient and Unacceptable Alternatives . . . . .	664
27.5.1 Unacceptable Alternatives . . . . .	664
27.5.2 Dominance – Inefficient Alternatives . . . . .	664
27.6 Printing Preference Relationships . . . . .	668
27.7 Step 5: MCDA Methods . . . . .	670
27.7.1 Examples of Non-compensatory Methods . . . . .	670
27.7.1.1 The MaxMin Method . . . . .	670
27.7.1.2 The MaxMax Method . . . . .	671
27.7.2 The Weighted Sum Method (WSM) . . . . .	671
27.7.3 Weighted Product Method (WPM) . . . . .	675
27.7.4 ELECTRE . . . . .	676
27.7.4.1 ELECTRE I . . . . .	677
27.7.4.2 ELECTRE II . . . . .	685
27.7.4.3 Conclusions ELECTRE . . . . .	687
27.7.5 PROMethEE . . . . .	687
27.7.5.1 PROMethEE I . . . . .	690
27.7.5.2 PROMethEE II . . . . .	698
27.7.6 PCA (Gaia) . . . . .	703
27.7.7 Outranking Methods . . . . .	709
27.7.8 Goal Programming . . . . .	710
27.8 Summary MCDA . . . . .	712
<b>VI Introduction to Companies</b>	<b>713</b>
<b>28 Financial Accounting (FA)</b>	<b>717</b>
28.1 The Statements of Accounts . . . . .	718
28.1.1 Income Statement . . . . .	718
28.1.2 Net Income: The P&L statement . . . . .	718
28.1.3 Balance Sheet . . . . .	719
28.2 The Value Chain . . . . .	722
28.3 Further, Terminology . . . . .	724
28.4 Selected Financial Ratios . . . . .	726

<b>29 Management Accounting</b>	<b>735</b>
29.1 Introduction	736
29.1.1 Definition of Management Accounting (MA)	736
29.1.2 Management Information Systems (MIS)	737
29.2 Selected Methods in MA	739
29.2.1 Cost Accounting	739
29.2.2 Selected Cost Types	742
29.3 Selected Use Cases of MA	745
29.3.1 Balanced Scorecard	745
29.3.2 Key Performance Indicators (KPIs)	747
29.3.2.1 Lagging Indicators	747
29.3.2.2 Leading Indicators	748
29.3.2.3 Selected Useful KPIs	749
<b>30 Asset Valuation Basics</b>	<b>753</b>
30.1 Time Value of Money	754
30.1.1 Interest Basics	754
30.1.2 Specific Interest Rate Concepts	754
30.1.3 Discounting	756
30.2 Cash	758
30.3 Bonds	759
30.3.1 Features of a Bond	759
30.3.2 Valuation of Bonds	762
30.3.3 Duration	765
30.3.3.1 Macaulay Duration	765
30.3.3.2 Modified Duration	766
30.4 The Capital Asset Pricing Model (CAPM)	769
30.4.1 The CAPM Framework	769
30.4.2 The CAPM and Risk	771
30.4.3 Limitations and Shortcomings of the CAPM	772
30.5 Equities	774
30.5.1 Definition	774
30.5.2 Short History	774
30.5.3 Valuation of Equities	775
30.5.4 Absolute Value Models	777
30.5.4.1 Dividend Discount Model (DDM)	777
30.5.4.2 Free Cash Flow (FCF)	782
30.5.4.3 Discounted Cash Flow Model	784
30.5.4.4 Discounted Abnormal Operating Earnings Model	786
30.5.4.5 Net Asset Value Method or Cost Method	786
30.5.4.6 Excess Earnings Method	788

## CONTENTS

---

30.5.5	Relative Value Models . . . . .	789
30.5.5.1	The Concept of Relative Value Models . . . . .	789
30.5.5.2	The Price Earnings Ratio (PE) . . . . .	790
30.5.5.3	Pitfalls when using PE Analysis . . . . .	791
30.5.5.4	Other Company Value Ratios . . . . .	791
30.5.6	Selection of Valuation Methods . . . . .	796
30.5.7	Pitfalls in Company Valuation . . . . .	796
30.5.7.1	Forecasting Performance . . . . .	796
30.5.7.2	Results and Sensitivity . . . . .	797
30.6	Forwards and Futures . . . . .	805
30.7	Options . . . . .	807
30.7.1	Definitions . . . . .	807
30.7.2	Commercial Aspects . . . . .	809
30.7.3	Short History . . . . .	811
30.7.4	Valuation of Options at Maturity . . . . .	813
30.7.4.1	A Long Call at Maturity . . . . .	813
30.7.4.2	A Short Call at Maturity . . . . .	814
30.7.4.3	Long and short put . . . . .	815
30.7.4.4	The Put-Call Parity . . . . .	816
30.7.5	The Black and Scholes Model . . . . .	818
30.7.5.1	Pricing of Options Before Maturity . . . . .	819
30.7.5.2	Apply the Black and Scholes formula . . . . .	820
30.7.5.3	The Limits of the Black and Scholes Model . . . . .	824
30.7.6	The Binomial Model . . . . .	824
30.7.6.1	Risk Neutral Method . . . . .	826
30.7.6.2	The Equivalent Portfolio Binomial Model . . . . .	830
30.7.6.3	Summary Binomial Model . . . . .	832
30.7.7	Dependencies of the Option Price . . . . .	833
30.7.7.1	Dependencies in a Long Call Option . . . . .	834
30.7.7.2	Dependencies in a Long Put Option . . . . .	836
30.7.7.3	Summary of Findings . . . . .	836
30.7.8	The Greeks . . . . .	838
30.7.9	Delta Hedging . . . . .	840
30.7.10	Linear Option Strategies . . . . .	843
30.7.10.1	Plotting a Portfolio of Options . . . . .	843
30.7.10.2	Single Option Strategies . . . . .	847
30.7.10.3	Composite Option Strategies . . . . .	848
30.7.11	Integrated Option Strategies . . . . .	855
30.7.11.1	The Covered Call . . . . .	855
30.7.11.2	The Married Put . . . . .	856
30.7.11.3	The Collar . . . . .	858

30.7.12 Exotic Options . . . . .	858
30.7.13 Capital Protected Structures . . . . .	861
<b>VII Report</b>	<b>863</b>
<b>31 A Grammar of Graphics with ggplot2</b>	<b>867</b>
31.0.1 The Basics of ggplot2 . . . . .	868
31.0.2 Over-plotting . . . . .	871
31.0.3 Case Study for ggplot2 . . . . .	876
<b>32 R Markdown</b>	<b>879</b>
<b>33 knitr and L<sup>A</sup>T<sub>E</sub>X</b>	<b>885</b>
<b>34 An Automated Development Cycle</b>	<b>891</b>
<b>35 Writing and Communication Skills</b>	<b>893</b>
<b>36 Interactive Apps</b>	<b>899</b>
36.1 Shiny . . . . .	902
36.2 Browser born data visualization . . . . .	907
36.2.1 HTML-widgets . . . . .	907
36.2.2 Interactive Maps with leaflet . . . . .	908
36.2.3 Interactive Data Visualisation with ggvis . . . . .	909
36.2.3.1 Getting Started in R with ggvis . . . . .	910
36.2.3.2 Combining the Power of ggvis and Shiny . . . . .	911
36.2.4 googleVis . . . . .	912
36.3 Dashboards . . . . .	915
36.3.1 The Business Case: a Diversity Dashboard . . . . .	916
36.3.2 A Dashboard with flexdashboard . . . . .	923
36.3.2.1 A Static Dashboard . . . . .	923
36.3.2.2 Interactive Dashboards with flexdashboard . . . . .	928
36.3.3 A Dashboard with shinydashboard . . . . .	930
<b>VIII Bigger and Faster R</b>	<b>935</b>
<b>37 Parallel computing</b>	<b>937</b>
37.1 Combine foreach and doParallel . . . . .	940
37.2 Distribute Calculations over LAN with snow . . . . .	943
37.3 Using the GPU . . . . .	948
37.3.1 Getting Started with gpuR . . . . .	950

## CONTENTS

---

37.3.2	On the Importance of Memory use . . . . .	954
37.3.3	Conclusions for GPU programming . . . . .	957
<b>38</b>	<b>R and Big Data</b>	<b>959</b>
38.1	Use a Powerful Server . . . . .	962
38.1.1	Use R on a Server . . . . .	962
38.1.2	Let the Database Server do the Heavy Lifting . . . . .	962
38.2	Using more memory than we have RAM . . . . .	964
<b>39</b>	<b>Parallelism for Big Data</b>	<b>965</b>
39.1	Apache Hadoop . . . . .	967
39.2	Apache Spark . . . . .	969
39.2.1	Installing Spark . . . . .	969
39.2.2	Running Spark . . . . .	972
39.2.3	SparkR . . . . .	975
39.2.3.1	A User Defined Function on Spark . . . . .	980
39.2.3.2	Some Other Functions of SparkR . . . . .	985
39.2.3.3	Machine learning with SparkR . . . . .	987
39.2.4	sparklyr . . . . .	990
39.2.5	SparkR or sparklyr . . . . .	995
<b>40</b>	<b>The Need for Speed</b>	<b>997</b>
40.1	Benchmarking . . . . .	998
40.2	Optimize Code . . . . .	1001
40.2.1	Avoid Repeating the Same . . . . .	1001
40.2.2	Use Vectorisation where Appropriate . . . . .	1001
40.2.3	Pre-allocating memory . . . . .	1003
40.2.4	Use the Fastest Function . . . . .	1006
40.2.5	Use the Fastest Package . . . . .	1006
40.2.6	Be Mindful about Details . . . . .	1007
40.2.7	Compile Functions . . . . .	1010
40.2.8	Use C or C++ Code in R . . . . .	1012
40.2.9	Using a C++ Source File in R . . . . .	1017
40.2.10	Call Compiled C++ Functions in R . . . . .	1019
40.3	Profiling code . . . . .	1020
40.3.1	The Package profr . . . . .	1022
40.3.2	The Package proftools . . . . .	1022
40.4	Optimize Your Computer . . . . .	1025

---

<b>IX</b>	<b>Appendices</b>	<b>1029</b>
<b>A</b>	<b>Create your own R package</b>	<b>1031</b>
A.1	Creating the package in the R console . . . . .	1033
A.2	Update the package description . . . . .	1036
A.3	Documenting the functions . . . . .	1037
A.4	Loading the Package . . . . .	1038
A.5	Further Steps . . . . .	1039
<b>B</b>	<b>Levels of Measurement</b>	<b>1041</b>
B.1	Nominal Scale . . . . .	1042
B.2	Ordinal Scale . . . . .	1043
B.3	Interval Scale . . . . .	1044
B.4	Ratio Scale . . . . .	1045
<b>C</b>	<b>Trademark Notices</b>	<b>1047</b>
C.1	General Trademark Notices . . . . .	1048
C.2	R-Related Notices . . . . .	1049
C.2.1	Crediting Developers of R packages . . . . .	1049
C.2.2	The R-packages used in this book . . . . .	1050
<b>D</b>	<b>Code Not Shown in the Body of the Book</b>	<b>1053</b>
<b>E</b>	<b>Answers to Selected Questions</b>	<b>1061</b>
	<b>Bibliography</b>	<b>1077</b>
	<b>Index</b>	<b>1089</b>
	<b>Nomenclature</b>	<b>1105</b>