

Contents

Part I Basics

1 Set-Up . . . . . 3

1.1 Introduction . . . . . 3

1.2 Structure of This Book . . . . . 4

1.3 Obtaining R . . . . . 5

1.4 Supplemental Materials . . . . . 5

1.5 Getting Help with R . . . . . 5

References . . . . . 6

2 A Short Introduction to R . . . . . 7

2.1 Introduction . . . . . 7

2.2 Calculator and Objects . . . . . 7

2.3 Numeric Vectors . . . . . 9

2.4 Logical Vectors . . . . . 11

2.5 Subsetting . . . . . 12

2.6 Character Vectors . . . . . 14

2.7 Matrices and Data Frames . . . . . 16

2.8 Data I/O . . . . . 19

2.9 Advanced Subsetting . . . . . 22

References . . . . . 24

3 EDA I: Continuous and Categorical Data . . . . . 25

3.1 Introduction . . . . . 25

3.2 Tables . . . . . 26

3.3	Histogram . . . . .	29
3.4	Quantiles . . . . .	31
3.5	Binning . . . . .	35
3.6	Control Flow . . . . .	37
3.7	Combining Plots . . . . .	40
3.8	Aggregation . . . . .	42
3.9	Applying Functions . . . . .	44
	References . . . . .	46
<b>4</b>	<b>EDA II: Multivariate Analysis . . . . .</b>	<b>47</b>
4.1	Introduction . . . . .	47
4.2	Scatter Plots . . . . .	47
4.3	Text . . . . .	50
4.4	Points . . . . .	53
4.5	Line Plots . . . . .	54
4.6	Scatter Plot Matrix . . . . .	58
4.7	Correlation Matrix . . . . .	60
<b>5</b>	<b>EDA III: Advanced Graphics . . . . .</b>	<b>63</b>
5.1	Introduction . . . . .	63
5.2	Output Formats . . . . .	63
5.3	Color . . . . .	65
5.4	Legends . . . . .	70
5.5	Randomness . . . . .	71
5.6	Additional Parameters . . . . .	76
5.7	Alternative Methods . . . . .	77
	References . . . . .	78

---

## Part II Humanities Data Types

---

<b>6</b>	<b>Networks . . . . .</b>	<b>81</b>
6.1	Introduction . . . . .	81
6.2	A Basic Graph . . . . .	81
6.3	Citation Networks . . . . .	84
6.4	Graph Centrality . . . . .	87
6.5	Graph Communities . . . . .	90
6.6	Further Extensions . . . . .	92
	References . . . . .	93
<b>7</b>	<b>Geospatial Data . . . . .</b>	<b>95</b>
7.1	Introduction . . . . .	95
7.2	From Scatter Plots to Maps . . . . .	96
7.3	Map Projections and Input Formats . . . . .	100
7.4	Enriching Tabular Data with Geospatial Data . . . . .	105
7.5	Enriching Geospatial Data with Tabular Data . . . . .	107



7.6 Further Extensions . . . . . 110

References . . . . . 110

**8 Image Data . . . . . 113**

8.1 Introduction . . . . . 113

8.2 Basic Image I/O . . . . . 113

8.3 Day/Night Photographic Corpus . . . . . 117

8.4 Principal Component Analysis . . . . . 120

8.5 K-Means . . . . . 123

8.6 Scatter Plot of Raster Graphics . . . . . 126

8.7 Extensions . . . . . 127

References . . . . . 129

**9 Natural Language Processing . . . . . 131**

9.1 Introduction . . . . . 131

9.2 Tokenization and Sentence Splitting . . . . . 132

9.3 Lemmatization and Part of Speech Tagging . . . . . 134

9.4 Dependencies . . . . . 138

9.5 Named Entity Recognition . . . . . 143

9.6 Coreference . . . . . 145

9.7 Case Study: Sherlock Holmes Main Characters . . . . . 148

9.8 Other Languages . . . . . 150

9.9 Conclusions and Extensions . . . . . 152

References . . . . . 153

**10 Text Analysis . . . . . 157**

10.1 Introduction . . . . . 157

10.2 Term Frequency: Inverse Document Frequency . . . . . 157

10.3 Topic Models . . . . . 162

10.4 Stylometric Analysis . . . . . 167

10.5 Further Methods and Extensions . . . . . 174

References . . . . . 175

---

**Part III Appendix**

---

**11 R Packages . . . . . 179**

11.1 Installing from Within R . . . . . 179

11.2 rJava . . . . . 181

11.3 coreNLP . . . . . 181

11.4 sessionInfo . . . . . 182

**12 100 Basic Programming Exercises . . . . . 183**

**13 100 Basic Programming Solutions . . . . . 193**