

# Contents

<b>Figures</b>	<b>ix</b>
<b>Tables</b>	<b>xi</b>
<b>Boxes</b>	<b>xiii</b>
<b>Preface</b>	<b>xv</b>
<b>1 What is data science?</b>	<b>1</b>
Data, information, knowledge, wisdom	1
Data everywhere	3
The data deserts	4
Data science	5
The potential of data science	8
From research data services to data science in libraries	10
Programming in libraries	14
Programming in this book	15
The structure of this book	16
<b>2 Little data, big data</b>	<b>17</b>
Big data	17
Data formats	19
Standalone files	20
Application programming interfaces	21
Unstructured data	25
Data sources	27
Data licences	37
<b>3 The process of data science</b>	<b>39</b>
Modelling the data science process	39
Frame the problem	41
Collect data	44
Transform and clean data	46
Analyse data	48

	Visualise and communicate data	50
	Frame a new problem	54
<b>4</b>	<b>Tools for data analysis</b>	<b>55</b>
	Finding tools	55
	Software for data science	56
	Programming for data science	69
<b>5</b>	<b>Clustering and social network analysis</b>	<b>77</b>
	Network graphs	77
	Graph terminology	79
	Network matrix	80
	Visualisation	82
	Network analysis	85
<b>6</b>	<b>Predictions and forecasts</b>	<b>97</b>
	Predictions and forecasts beyond data science	97
	Predictions in a world of (limited) data	99
	Predicting and forecasting for information professionals	101
	Statistical methodologies	102
<b>7</b>	<b>Text analysis and mining</b>	<b>113</b>
	Text analysis and mining, and information professionals	113
	Natural language processing	115
	Keywords and n-grams	125
<b>8</b>	<b>The future of data science and information professionals</b>	<b>133</b>
	Eight challenges to data science	134
	Ten steps to data science librarianship	139
	The final word: play	144
	<b>References</b>	<b>147</b>
	<b>Appendix – Programming concepts for data science</b>	<b>165</b>
	Variables, data types and other classes	165
	Import libraries	167
	Functions and methods	168
	Loops and conditionals	170
	Final words of advice	171
	Further reading	172
	<b>Index</b>	<b>173</b>