

Contents

Foreword x

Acknowledgements xii

Author Biographies xiv

Chapter 1 Introduction 1

The Origins of Data Journalism 5

Technology: The Tools of Data and How Journalists Use Them 7

PART I GETTING DATA 11

Chapter 2 Online and Open Data 12

What You Will Learn 12

Principles 17

The History of Open Government and Open Data 19

What Kind of Data Can You Get? 21

Chicago's Crime Database 22

Toronto Payment Card Data 23

Road Accidents in the United Kingdom 24

Vancouver Food Vendors 25

Mine Accidents 25

Data Quality Concerns 27

Steps to Take When Working with Any Dataset 28

Study Questions and Exercises 30

Chapter 3 When Data Is Not Conveniently Available 32

What You Will Learn 32

Informing Yourself 33

Asking 36

Negotiating for Data 38

Demanding 41

Making an Effective Request for Data 42

After You Make the Request 43

The Answer 45

Study Questions and Exercises 45

PART II ANALYZING DATA 47

Chapter 4 Spreadsheets: The Basic Tool 48

What You Will Learn	48
So What Exactly Is a Spreadsheet?	49
Fundamentals of Spreadsheets	50
How Journalists Use Spreadsheets	51
Getting Data into a Spreadsheet	52
Getting to Work	55
Sorting and Simple Analysis	57
Filtering	58
Doing the Math	60
Functions Big and Small	61
Calculating Percentages	62
Calculating Rates	64
Dealing with Errors	64
Working with Dates	66
Concatenation	67
Summarizing Information with Pivot Tables	69
Logical Functions and IF Statements	73
Using Paste Special to Convert Formulas to Values	75
Chronologies	75
Conclusion	76
Study Questions and Exercises	77

Chapter 5 Working with Databases 78

What You Will Learn	78
Working with a Database	80
Building Relationships	83
Desktop and Server Database Programs	85
Making Tables and Adding Data	87
The Main Course: Querying your Data	89
The Language of Queries	90
Writing Math Queries	97
The GROUP BY Clause	98
The HAVING Clause	99
Aliases	100
Joining Tables	100
Some Notes about Joining	102
More Advanced Queries	103

Outer Joins	103
Union Queries	104
Subqueries	105
Using Calculated Fields and String Functions	105
Creating Views (MySQL)	108
Improving Query Performance by Adding Indexes	108
Queries to Alter your Data	109
Exporting Query Results	110
Building Your Own Database	110
Conclusion	112
Study Questions and Exercises	112

Chapter 6 Introduction to Maps in Journalism 114

What You Will Learn	114
Maps Are Not Reality, but Representations of Reality	115
Web Mapping Services	118
Getting Your Data onto a Web Map	120
Data Not Already in a Map Format	123
Your Data Is Ready: How to Use It?	124
What Kinds of Maps Can You Make?	126
Boundary Maps	128
Some Important Design Principles	129
A Choice of Styles	129
Choosing Appropriate Colours	131
Setting Breakpoints	133
Conclusion	134
Study Questions and Exercises	134

Chapter 7 Working with GIS Programs 135

What You Will Learn	135
The Inside Story	136
Ellipsoids and Datums: Modelling the World	138
Projections	140
Four Properties: Direction, Area, Shape, and Distance	141
How Datums and Projections are Incorporated in a GIS	145
Basics of a GIS	146
How Spatial Data Is Structured	147
Metadata	147
Geometries	148
Attributes	148
Working with Layers	149

How Journalists Use GIS Programs	150
Joining Non-geographic Data with Geographic Data	153
Selecting Features That Meet Certain Criteria	153
Buffering	155
Joining Data Based on Geographical Location	155
Making a Choropleth Map	156
Geocoding	156
Common Problems and Solutions	157
Projecting and Re-projecting a Map Layer	158
Converting File Formats	158
Simplifying Polygons	159
Combining Layers into a Single Layer	160
Combining Features within a Single Layer	160
Special Considerations Relative to Coordinate Systems	160
Using a Map Layer That Has Only a Geographic Coordinate System	161
Using Maps Based on Different Geographic Coordinate Systems	162
Using an Inappropriate Projected Coordinate System	162
Conclusion	163
Study Questions and Exercises	163
Further Reading	164
Chapter 8 Visualizing Data	165
What You Will Learn	165
A History of Visualization: From William Playfair to the Present Day	165
What Is Meant by Visualization?	167
Basic Dos and Don'ts	168
Content Is King	178
Choosing the Right Chart	175
The Pie Chart	175
The Bar Chart	178
The Line Chart	179
How Journalists Use "Data Viz" Tools	183
One Producer's View	185
Conclusion	187
Study Questions and Exercises	187
PART III ADVANCED TOPICS	189
Chapter 9 Web Scraping	190
What You Will Learn	190
The Underpinnings of Scraping	192

Options for Scraping	196
First Steps, Thinking through Your Scrape	199
Coding Basics	201
Getting Ready to Code	203
Fetching the Page	204
Scraping More than One Page	213
More Complicated Scrapes	213
The Ethics of Web Scraping	214
APIs	217
Conclusion	219
Study Questions and Exercises	219
Further Resources	219
Chapter 10 Web Development	220
What You Will Learn	220
State of Developers in Newsrooms	221
Core Languages Used by Newsroom Developers: An Introduction to the Work Environment	225
The Server Side: Writing for Your Own Machine	226
The Client Side: Writing for Everyone Else's Machines	227
Case Study: Using freeDive	234
Working with JavaScript Libraries Such as jQuery and D3	235
Conclusion	242
Study Questions and Exercises	242
Further Reading	243
PART IV TELLING STORIES WITH DATA	245
Chapter 11 Incorporating Data Journalism into Traditional Reporting	246
What You Will Learn	246
Why Data is Just the Beginning	246
Finding Outliers that Lead to Human Stories	249
Testing Your Data in the Real World	250
Connecting Data Patterns with Real-Life Patterns	252
Building Powerful Interviews and Writing the Story	253
Conclusion	254
Study Questions and Exercises	254
Glossary	257
Index	263