Contents

1 Introduction 1
1.1 What Is Data Analysis? 1
1.2 Objectives of the Book 5
1.3 Outline of the Book 6
1.3.1 Data 6
1.3.2 Tasks 8
1.3.3 Tools 10
1.3.4 General Principles 14
References 16

2 Data 17
Abstract 17
2.1 Structure of Data 18
2.1.1 Functional View of Data Structure 21
2.1.2 Other Approaches 25
2.2 Properties of Data 27
2.2.1 Other Approaches 31
2.3 Examples of Data 34
2.3.1 Portuguese Census 34
2.3.2 Forests in Europe 36
2.3.3 Earthquakes in Turkey 36
2.3.4 Migration of White Storks 38
2.3.5 Weather in Germany 40
2.3.6 Crime in the USA 41
2.3.7 Forest Management Scenarios 42
Summary 44
References 45

3 Tasks 47
Abstract 47
3.1 Jacques Bertin's View of Tasks 49
3.2 General View of a Task 53

3.3 Elementary Tasks 60
3.3.1 Lookup and Comparison 61
3.3.2 Relation-Seeking 69
3.3.3 Recap: Elementary Tasks 75
3.4 Synoptic Tasks 81
3.4.1 General Notes 81
3.4.2 Behaviour and Pattern 83
3.4.3 Types of Patterns 91
3.4.3.1 Association Patterns 91
3.4.3.2 Differentiation Patterns 93
3.4.3.3 Arrangement Patterns 94
3.4.3.4 Distribution Summary 95
3.4.3.5 General Notes 96
3.4.4 Behaviours over Multidimensional Reference Sets 98
3.4.5 Pattern Search and Comparison 107
3.4.6 Inverse Comparison 112
3.4.7 Relation-Seeking 115
3.4.8 Recap: Synoptic Tasks 119
3.5 Connection Discovery 124
3.5.1 General Notes 124
3.5.2 Properties and Formalisation 127
3.5.3 Relation to the Former Categories 134
3.6 Completeness of the Framework 139
3.7 Relating Behaviours: a Cognitive-Psychology Perspective 143
3.8 Why Tasks? 148
3.9 Other Approaches 151
Summary 158
References 159

4 Tools 163
Abstract 163
4.1 A Few Introductory Notes 165
4.2 The Value of Visualisation 166
4.3 Visualisation in a Nutshell 171
4.3.1 Bertin's Theory and Its Extensions 171
4.3.2 Dimensions and Variables of Visualisation 182
4.3.3 Basic Principles of Visualisation 189
4.3.4 Example Visualisations 196
4.4 Display Manipulation 207
4.4.1 Ordering 207
4.4.2 Eliminating Excessive Detail 214
4.4.3 Classification 217

4.4.4 Zooming and Focusing 231
4.4.5 Substitution of the Encoding Function 241
4.4.6 Visual Comparison 248
4.4.7 Recap: Display Manipulation 257
4.5 Data Manipulation 259
4.5.1 Attribute Transformation 261
4.5.1.1 "Relativisation" 261
4.5.1.2 Computing Changes 263
4.5.1.3 Accumulation 268
4.5.1.4 Neighbourhood-Based Attribute Transformations 269
4.5.2 Attribute Integration 276
4.5.2.1 An Example of Integration 278
4.5.2.2 Dynamic Integration of Attributes 279
4.5.3 Value Interpolation 288
4.5.4 Data Aggregation 293
4.5.4.1 Grouping Methods 294
4.5.4.2 Characterising Aggregates 297
4.5.4.3 Visualisation of Aggregate Sizes 300
4.5.4.4 Sizes Are Not Only Counts 312
4.5.4.5 Visualisation and Use of Positional Measures 316
4.5.4.6 Spatial Aggregation and Reaggregation 327
4.5.4.7 A Few Words About OLAP 332
4.5.4.8 Data Aggregation: a Few Concluding Remarks 333
4.5.5 Recap: Data Manipulation 335
4.6 Querying 336
4.6.1 Asking Questions 337
4.6.1.1 Spatial Queries 341
4.6.1.2 Temporal Queries 346
4.6.1.3 Asking Questions: Summary 349
4.6.2 Answering Questions 351
4.6.2.1 Filtering 353
4.6.2.2 Marking 363
4.6.2.3 Marking Versus Filtering 371
4.6.2.4 Relations as Query Results 373
4.6.3 Non-Elementary Queries 381
4.6.4 Recap: Querying 393
4.7 Computational Tools 395
4.7.1 A Few Words About Statistical Analysis 397
4.7.2 A Few Words About Data Mining 401
4.7.3 The General Paradigm for Using Computational Tools 406
4.7.4 Example: Clustering 407
4.7.5 Example: Classification 415

4.7.6 Example: Data Preparation 423
4.7.7 Recap: Computational Tools 425
4.8 Tool Combination and Coordination 428
4.8.1 Sequential Tool Combination 429
4.8.2 Concurrent Tool Combination 434
4.8.3 Recap: Tool Combination 447
4.9 Exploratory Tools and Technological Progress 450
Summary 453
References 454

5 Principles 461
Abstract 461
5.1 Motivation 463
5.2 Components of the Exploratory Process 465
5.3 Some Examples of Exploration 467
5.4 General Principles of Selection of the Methods and Tools 480
5.4.1 Principle 1: See the Whole 481
5.4.1.1 Completeness 483
5.4.1.2 Unification 494
5.4.2 Principle 2: Simplify and Abstract 506
5.4.3 Principle 3: Divide and Group 509
5.4.4 Principle 4: See in Relation 518
5.4.5 Principle 5: Look for Recognisable 530
5.4.6 Principle 6: Zoom and Focus 540
5.4.7 Principle 7: Attend to Particulars 544
5.4.8 Principle 8: Establish Linkages 552
5.4.9 Principle 9: Establish Structure 572
5.4.10 Principle 10: Involve Domain Knowledge 579
5.5 General Scheme of Data Exploration: Tasks, Principles, and Tools 584
5.5.1 Case 1: Single Referrer, Holistic View Possible 587
5.5.1.1 Subcase 1.1: a Homogeneous Behaviour 588
5.5.1.2 Subcase 1.2: a Heterogeneous Behaviour 590
5.5.2 Case 2: Multiple Referrers 593
5.5.2.1 Subcase 2.1: Holistic View Possible 595
5.5.2.2 Subcase 2.2: Behaviour Explored by Slices and Aspects 598
5.5.3 Case 3: Multiple Attributes 602
5.5.4 Case 4: Large Data Volume 606
5.5.5 Final Remarks 611
5.6 Applying the Scheme (an Example) 613
Summary 630

References 632

6 Conclusion 635

Appendix I: Major Definitions 639
I.1 Data 639
I.2 Tasks 643
I.3 Tools 647

Appendix II: A Guide to Our Major Publications Relevant to This Book 651
References 653

Appendix III: Tools for Visual Analysis of Spatio-Temporal Data Developed at the AIS Fraunhofer Institute 657
References 658

Index 659