

# Contents

## Part I Introduction to Visual Analytics in Data Science

<b>1</b>	<b>Introduction to Visual Analytics by an Example</b>	<b>3</b>
1.1	What is visual analytics? (A brief summary)	3
1.2	A motivating example: Investigating an epidemic outbreak	5
1.2.1	Data and task description	5
1.2.2	Data properties	6
1.2.3	Data preparation	7
1.2.4	Analysing the temporal distribution	10
1.2.5	Analysing the spatial distribution	11
1.2.6	Transforming data to verify observed patterns	12
1.2.7	Exploring the spatio-temporal distribution	14
1.2.8	Revealing the patterns of the disease spread	14
1.2.9	Identifying the mechanisms of the disease transmission	16
1.2.10	Identifying the epidemic development trend	18
1.2.11	Summary: The story reconstructed	19
1.3	Discussion: How visual analytics has helped us	20
1.4	General definition of visual analytics	22
<b>2</b>	<b>General Concepts</b>	<b>27</b>
2.1	Subjects of analysis	27
2.2	Structure of an analysis subject	30
2.3	Using data to understand a subject	32
2.3.1	Distribution	33
2.3.2	Patterns and outliers	35
2.3.3	Patterns in different kinds of distributions	36
2.3.4	Co-distributions	41
2.3.5	Spatialisation	45
2.4	Concluding remarks	48

<b>3 Principles of Interactive Visualisation</b>	51
3.1 Preliminary notes	51
3.2 Visualisation	52
3.2.1 A motivating example	52
3.2.2 Visualisation theory in a nutshell	55
3.2.3 The use of display space	61
3.2.4 Commonly used visualisations	63
3.2.5 General principles of visualisation	68
3.2.6 Benefits of visualisation	70
3.2.7 Limitations of visualisation	72
3.3 Interaction	75
3.3.1 Interaction for changing data representation	75
3.3.2 Interaction for focusing and getting details	79
3.3.3 Interaction for data transformation	80
3.3.4 Interaction for data selection and filtering	83
3.3.5 Relating multiple graphical views	85
3.3.6 Limitations and disadvantages of interaction	87
3.4 Concluding remarks	88
<b>4 Computational Techniques in Visual Analytics</b>	89
4.1 Preliminary notes	89
4.1.1 Visualisation for supporting computations	90
4.1.2 Computations for supporting visual analysis	92
4.2 Distance functions	95
4.2.1 Multiple numeric attributes	95
4.2.2 Distributions	99
4.2.3 Numeric time series	100
4.2.4 Categorical attributes and mixed data	102
4.2.5 Sets	103
4.2.6 Sequences	103
4.2.7 Graphs	104
4.2.8 Distances in space and time	105
4.2.9 Data normalisation and standardisation	106
4.3 Feature selection	108
4.4 Data embedding	111
4.4.1 Embedding space	111
4.4.2 Representing strengths of relationships by distances	112
4.4.3 Distinctions between data embedding methods	116
4.4.4 Interpreting data embeddings	117
4.5 Clustering	122
4.5.1 Types of clustering methods	122
4.5.2 Interpreting clusters	125
4.5.3 Clustering process	130
4.5.4 Assigning colours to clusters along the clustering process	138
4.5.5 Generalisation to other computational methods	141

4.6	Topic modelling .....	141
4.6.1	General ideas and properties of methods .....	142
4.6.2	How many topics? .....	143
4.6.3	Topic modelling versus clustering .....	144
4.6.4	Use of topic modelling in data analysis .....	145
4.7	Conclusion .....	147

## Part II Visual Analytics along the Data Science Workflow

<b>5</b>	<b>Visual Analytics for Investigating and Processing Data</b> .....	151
5.1	Examples of data properties that may affect data analysis .....	151
5.2	Investigating data properties .....	153
5.2.1	Overall view of a distribution .....	154
5.2.2	Outliers .....	156
5.2.3	Missing data .....	160
5.2.4	Impacts of data collection and integration procedures .....	166
5.3	Processing data .....	170
5.3.1	Data cleaning .....	170
5.3.2	Modelling for data preparation .....	171
5.3.3	Transformation of data elements .....	173
5.3.4	Synthesis of data components .....	173
5.3.5	Data integration .....	175
5.3.6	Transformation of data structure .....	176
5.3.7	Data reduction and selection .....	178
5.4	Concluding remarks .....	179
5.5	Questions and exercises .....	180
<b>6</b>	<b>Visual Analytics for Understanding Multiple Attributes</b> .....	181
6.1	Motivating example .....	181
6.2	Specifics of multivariate data .....	184
6.3	Analytical Goals and Tasks .....	186
6.4	Visual Analytics Techniques .....	186
6.4.1	Analysing characteristics of multiple attributes .....	187
6.4.2	Analysing Multivariate Relations .....	188
6.4.3	Analysing higher-order relations and local structures .....	190
6.5	Further Examples .....	193
6.5.1	Exploring projections through interactive probing .....	193
6.5.2	Manually crafting projections through “Explainers” .....	195
6.6	Concluding remarks .....	199
6.7	Questions and exercises .....	200
<b>7</b>	<b>Visual Analytics for Understanding Relationships between Entities</b> ..	201
7.1	Motivating example .....	201
7.1.1	Extracting relationships .....	202
7.1.2	Visualising relationships .....	204
7.1.3	Exploring relationships .....	207

7.1.4	Main takeaways from the example . . . . .	208
7.2	Graphs as a mathematical concept . . . . .	209
7.2.1	Definition . . . . .	209
7.2.2	Graph-theoretic metrics . . . . .	210
7.3	Specifics of this kind of phenomena/data . . . . .	211
7.4	Graph/network visualisation techniques . . . . .	212
7.5	Common tasks in graph/network analysis . . . . .	219
7.6	Further Examples . . . . .	220
7.6.1	Analysis of graphs with multiple connected components . . . . .	220
7.6.2	Analysis of dynamic graphs . . . . .	223
7.7	Concluding remarks . . . . .	227
7.8	Questions and exercises . . . . .	228
<b>8</b>	<b>Visual Analytics for Understanding Temporal Distributions and Variations . . . . .</b>	<b>229</b>
8.1	Motivating example . . . . .	229
8.2	Specifics of temporal phenomena and temporal data . . . . .	234
8.3	Transformations of temporal data . . . . .	236
8.4	Temporal filtering . . . . .	239
8.5	Analysing temporal data with visual analytics . . . . .	240
8.5.1	Events . . . . .	240
8.5.2	Univariate time series . . . . .	246
8.5.3	Time series of complex states . . . . .	253
8.6	Questions . . . . .	258
8.7	Exercises . . . . .	260
<b>9</b>	<b>Visual Analytics for Understanding Spatial Distributions and Spatial Variation . . . . .</b>	<b>261</b>
9.1	Motivating example . . . . .	261
9.2	How spatial phenomena are represented by data . . . . .	263
9.2.1	Forms of spatial data . . . . .	263
9.2.2	Georeferencing . . . . .	264
9.2.3	Spatial joining . . . . .	265
9.2.4	Coordinate systems and cartographic projections . . . . .	266
9.3	Specifics of this kind of phenomena . . . . .	268
9.3.1	Spatial dependence and interdependence . . . . .	268
9.3.2	Spatial precision and accuracy . . . . .	270
9.3.3	Spatial scale of analysis . . . . .	270
9.3.4	Spatial partitioning . . . . .	272
9.4	Transformations of spatial data . . . . .	274
9.4.1	Coordinate transformations . . . . .	274
9.4.2	Aggregation . . . . .	277
9.5	Analysis tasks and visual analytics techniques . . . . .	283
9.6	An example of a spatial analysis workflow . . . . .	286
9.7	Conclusion . . . . .	294

9.8	Questions and exercises .....	294
<b>10</b>	<b>Visual Analytics for Understanding Phenomena in Space and Time .....</b>	<b>297</b>
10.1	Motivating example .....	297
10.2	Specifics of this kind of phenomena/data .....	304
10.2.1	Data structures and transformations .....	304
10.2.2	General properties .....	310
10.2.3	Possible data quality issues .....	313
10.3	Visual Analytics Techniques .....	319
10.3.1	Spatio-temporal distribution of spatial events .....	319
10.3.2	Analysis of spatial time series .....	323
10.3.3	Analysis of trajectories .....	328
10.4	Analysis example: Understanding approach schemes in aviation ..	334
10.5	Concluding remarks .....	338
10.6	Questions and exercises .....	340
<b>11</b>	<b>Visual Analytics for Understanding Texts .....</b>	<b>341</b>
11.1	Motivating example .....	341
11.2	Specifics of this kind of phenomena/data .....	344
11.3	Analysis tasks .....	344
11.4	Computational processing of textual data .....	345
11.5	Visualisation of structured data derived from texts .....	346
11.5.1	Numeric attributes .....	346
11.5.2	Significant items with numeric measures .....	346
11.5.3	Named entities and relationships .....	351
11.6	Analysing word occurrences and their contexts .....	352
11.7	Texts in geographic space .....	352
11.8	Texts over time .....	355
11.9	Concluding remarks .....	358
11.10	Questions and exercises .....	359
<b>12</b>	<b>Visual Analytics for Understanding Images and Video .....</b>	<b>361</b>
12.1	Motivating example .....	361
12.2	Specifics of this kind of phenomena/data .....	363
12.3	Analysis tasks .....	365
12.4	Visual Analytics techniques .....	365
12.4.1	Spatialisation of image collections or video frames .....	366
12.4.2	Detection of relevant objects and analysis of their changes ..	368
12.4.3	Analysis of object movements .....	371
12.5	General scheme for visual analysis of unstructured data .....	372
12.6	Questions and exercises .....	374
<b>13</b>	<b>Computational Modelling with Visual Analytics .....</b>	<b>375</b>
13.1	Basic concepts .....	375
13.2	Motivating example .....	378
13.2.1	Problem statement .....	378

13.2.2 Understanding relationships among variables .....	379
13.2.3 Iterative construction of a model .....	381
13.2.4 Main takeaways from the example .....	386
13.3 General tasks in model building .....	387
13.4 Doing modelling tasks with visual analytics .....	388
13.5 Further examples:	
Evaluation and refinement of classification models .....	390
13.5.1 Assessment of the quality of a classifier .....	391
13.5.2 Example: improving a binary classifier .....	394
13.5.3 Example: analysing and comparing performances of multi-class classifiers .....	396
13.5.4 General notes concerning classification models .....	400
13.6 Visual analytics in modelling time series .....	400
13.7 Explaining model behaviour .....	403
13.8 General principles of thoughtful model building with visual analytics .....	405
13.9 Questions .....	407
<b>14 Conclusion .....</b>	<b>409</b>
14.1 What you have learned about visual analytics .....	409
14.2 Visual analytics way of thinking .....	410
14.3 Examples in this book .....	410
14.4 Example: devising an analytical workflow for understanding team tactics in football .....	411
14.4.1 Data description and problem statement .....	412
14.4.2 Devising the approach .....	412
14.4.3 Choosing methods and tools .....	415
14.4.4 Implementing the analysis plan .....	418
14.4.5 Conclusion .....	421
14.5 Final remarks .....	422
<b>Glossary .....</b>	<b>423</b>
References .....	426
<b>Index .....</b>	<b>435</b>