
Contents

SECTION I The basics

CHAPTER 1 ■ Why visualize? 3

- 1.1 BECAUSE OUR BRAINS ARE WIRED THAT WAY 3
- 1.2 TO HELP THE ANALYST AVOID PROBLEMS 5
- 1.3 TO WIN OVER THE AUDIENCE 8
- 1.4 WORKING ON DATA VISUALIZATION 12
- 1.5 A TOOLBOX 14
- 1.6 BE PREPARED TO SKETCH AND DISCARD 15

CHAPTER 2 ■ Translating numbers to images 21

- 2.1 LEAVES ON THE LINE: AN EXAMPLE OF VISUAL ENCODING 21
- 2.2 CHOOSING VISUAL PARAMETERS 27
- 2.3 UNDERSTANDING COLOR 30
- 2.4 THE LIMITATIONS OF AREAS 31
- 2.5 ANNOTATION 37
- 2.6 USER TESTING 37

SECTION II Statistical building blocks

CHAPTER 3 ■ Continuous and discrete numbers 41

- 3.1 ONE VARIABLE AT A TIME 41

viii ■ Contents

3.2	COMPARING UNMATCHED DATA	45
3.3	COMPARING MATCHED DATA	49
3.4	ASSOCIATIONS	51
CHAPTER 4 ■ Percentages and risks		59
<hr/>		
4.1	SHOWING ONE VARIABLE AT A TIME	59
4.2	COMPARING UNMATCHED DATA	62
4.3	COMPARING MATCHED DATA	63
4.4	CATEGORIES WITHIN CATEGORIES	65
4.5	ASSOCIATIONS	67
4.6	RISKS, RATES AND ODDS	70
CHAPTER 5 ■ Showing data or statistics		71
<hr/>		
5.1	CHOOSING DATA OR STATISTICS	71
5.2	THE STANDARD DEVIATION	73
5.3	QUANTILES AND OTHER ROBUST STATISTICS	75
5.4	SMOOTHING	77
CHAPTER 6 ■ Differences, ratios, correlations		81
<hr/>		
6.1	DIFFERENCE OR RATIO?	81
6.2	ODDS AND THE ODDS RATIO	85
6.3	CHOOSING A BASELINE	88
6.4	CHERRY-PICKING	89
6.5	CORRELATIONS	89
SECTION III Specific tasks		
CHAPTER 7 ■ Visual perception and the brain		97
<hr/>		
7.1	ATTENTION AND CLARITY	98

7.2	CULTURAL ASSUMPTIONS	102
7.3	LEARNING FROM OPTICAL ILLUSIONS	103
CHAPTER 8 ■ Showing uncertainty		105
8.1	THE BOOTSTRAP	105
8.2	CONFIDENCE REGIONS	107
8.3	OTHER SOURCES OF UNCERTAINTY	109
CHAPTER 9 ■ Time trends		111
9.1	MORE FORMATS AND ENCODINGS FOR TIME	111
9.2	STATISTICAL CONSIDERATIONS	116
9.3	UNCERTAINTY OVER TIME	119
9.4	STATISTICAL TIME SERIES MODELS	119
CHAPTER 10 ■ Statistical predictive models		123
10.1	LINEAR REGRESSION MODELS	124
10.2	LOGISTIC REGRESSION MODELS	127
10.3	SEMI- AND NON-PARAMETRIC MODELS	131
10.4	HOW GOOD IS THAT MODEL?	135
10.5	USES OF COMPUTER SIMULATION	135
CHAPTER 11 ■ Machine learning techniques		137
11.1	ENSEMBLES, BAGGING AND BOOSTING	138
11.2	CLASSIFICATION AND REGRESSION TREES	138
11.3	RANDOM FORESTS	142
11.4	SUPPORT VECTOR MACHINES	144
11.5	NEURAL NETWORKS AND DEEP LEARNING	145
11.6	IMAGE DATA	146
CHAPTER 12 ■ Many variables		149

x ■ Contents

12.1	SMALL MULTIPLES	149
12.2	IMPRESSIONS OF 3-D	150
12.3	DISTANCES	151
12.4	PROJECTIONS INTO TWO DIMENSIONS	155
12.5	CLUSTER ANALYSIS	158
12.6	OTHER APPROACHES	161
CHAPTER 13 ■ Maps and networks		163
<hr/>		
13.1	MAPPING BASICS	163
13.2	DATA ON TOP OF MAPS	167
13.3	SPATIAL MODELS AND UNCERTAINTY	171
13.4	NETWORKS	173
CHAPTER 14 ■ Interactivity		177
<hr/>		
14.1	WEB PAGES AND JAVASCRIPT	177
14.2	FORMS OF INTERACTIVITY	178
14.3	METHODVIZ	181
14.4	RUNNING THE ANALYSIS TOO	181
14.5	SECURITY AND CONFIDENTIALITY	182
CHAPTER 15 ■ Big data		185
<hr/>		
15.1	TOO BIG	186
15.2	TOO FAST	189
CHAPTER 16 ■ Visualization as part of a bigger pack- age		191
<hr/>		
16.1	THINK ABOUT IT	191
16.2	MAKE IT	192

Contents ■ xi

16.3 TALK ABOUT IT 196

SECTION IV *Closing remarks*

CHAPTER 17 ■ Some overarching ideas 203

Index 205
