
Contents

Preface	xi
1 Setting the Scene	1
1.1 Graphics in action	1
1.2 Introduction	3
1.3 What is Graphical Data Analysis (GDA)?	5
1.4 Using this book, the R code in it, and the book's webpage	14
2 Brief Review of the Literature and Background Materials	19
2.1 Literature review	19
2.2 Interactive graphics	21
2.3 Other graphics software	21
2.4 Websites	22
2.5 Datasets	23
2.6 Statistical texts	25
3 Examining Continuous Variables	27
3.1 Introduction	27
3.2 What features might continuous variables have?	29
3.3 Looking for features	30
3.4 Comparing distributions by subgroups	44
3.5 What plots are there for individual continuous variables?	46
3.6 Plot options	47
3.7 Modelling and testing for continuous variables	48
4 Displaying Categorical Data	53
4.1 Introduction	53
4.2 What features might categorical variables have?	56
4.3 Nominal data—no fixed category order	57
4.4 Ordinal data—fixed category order	62
4.5 Discrete data—counts and integers	66
4.6 Formats, factors, estimates, and barcharts	70
4.7 Modelling and testing for categorical variables	71

5	Looking for Structure: Dependency Relationships and Associations	75
5.1	Introduction	75
5.2	What features might be visible in scatterplots?	77
5.3	Looking at pairs of continuous variables	78
5.4	Adding models: lines and smooths	83
5.5	Comparing groups within scatterplots	86
5.6	Scatterplot matrices for looking at many pairs of variables	88
5.7	Scatterplot options	92
5.8	Modelling and testing for relationships between variables	94
6	Investigating Multivariate Continuous Data	99
6.1	Introduction	99
6.2	What is a parallel coordinate plot (pcp)?	100
6.3	Features you can see with parallel coordinate plots	102
6.4	Interpreting clustering results	106
6.5	Parallel coordinate plots and time series	108
6.6	Parallel coordinate plots for indices	112
6.7	Options for parallel coordinate plots	115
6.8	Modelling and testing for multivariate continuous data	127
6.9	Parallel coordinate plots and comparing model results	127
7	Studying Multivariate Categorical Data	131
7.1	Introduction	131
7.2	Data on the sinking of the Titanic	132
7.3	What is a mosaicplot?	133
7.4	Different mosaicplots for different questions of interest	140
7.5	Which mosaicplot is the right one?	148
7.6	Additional options	149
7.7	Modelling and testing for multivariate categorical data	151
8	Getting an Overview	155
8.1	Introduction	155
8.2	Many individual displays	159
8.3	Multivariate overviews	162
8.4	Multivariate overviews for categorical variables	168
8.5	Graphics by group	170
8.6	Modelling and testing for overviews	174

9 Graphics and Data Quality: How Good Are the Data?	177
9.1 Introduction	177
9.2 Missing values	178
9.3 Outliers	183
9.4 Modelling and testing for data quality	193
10 Comparisons, Comparisons, Comparisons	197
10.1 Introduction	197
10.2 Making comparisons	199
10.3 Making visual comparisons	202
10.4 Comparing group effects graphically	208
10.5 Comparing rates visually	212
10.6 Graphics for comparing many subsets	214
10.7 Graphics principles for comparisons	216
10.8 Modelling and testing for comparisons	218
11 Graphics for Time Series	223
11.1 Introduction	223
11.2 Graphics for a single time series	224
11.3 Multiple series	226
11.4 Special features of time series	233
11.5 Alternative graphics for time series	237
11.6 R classes and packages for time series	238
11.7 Modelling and testing time series	238
12 Ensemble Graphics and Case Studies	243
12.1 Introduction	243
12.2 What is an ensemble of graphics?	246
12.3 Combining different views—a case study example	248
12.4 Case studies	252
13 Some Notes on Graphics with R	257
13.1 Graphics systems in R	257
13.2 Loading datasets and packages for graphical analysis	258
13.3 Graphics conventions in statistics	258
13.4 What is a graphic anyway?	259
13.5 Options for all graphics	261
13.6 Some R graphics advice and coding tips	264
13.7 Other graphics	271
13.8 Large datasets	272
13.9 Perfecting graphics	273

14 Summary 275

14.1 Data analysis and graphics	275
14.2 Key features of GDA	276
14.3 Strengths and weaknesses of GDA	276
14.4 Recommendations for GDA	277

References 279

General index	291
Datasets index	295