
Contents

Preface	xi
1 Introduction	1
1.1 What are good graphics?	3
1.2 Graphics examples	4
1.3 Writing about graphics	5
1.4 Reading about graphics	5
1.5 Software	8
1.6 Structure of the book	8
In a nutshell (tl;dr)	10
I Graphics in Action	11
2 Graphics and Gapminder	13
2.1 Gapminder and Hans Rosling	13
2.2 What else can be seen in Gapminder's plots?	16
2.3 Analysis by regions of the world	18
2.4 Checking what is in the data	23
2.5 The importance of data definitions	24
2.6 Discussion	24
3 Looking at the movies	27
3.1 Trailer	27
3.2 Movies in IMDb over the years	29
3.3 What might IMDb code numbers mean?	31
3.4 What ratings do people give films?	33
3.5 Genres from Rom-Com to Film Noir	36
3.6 Comparing Raging Bull with Jane Austen	40
4 Voting 46 times to choose a Presidential candidate	43
4.1 A first look at the 1912 Democratic Convention	43
4.2 The USA in 1912 and 2020	44
4.3 How the nomination contest unfolded	49
4.4 When did the ballots actually take place?	52

5	Measuring the speed of light	55
5.1	Michelson, the Master of Light, and his data	55
5.2	Newcomb, self-educated polymath, and his data	57
5.3	Comparing the data of Michelson and Newcomb	61
6	The modern Olympic Games in numbers	63
6.1	How good are the data?	63
6.2	How much have performances improved?	70
6.3	How many countries and individuals competed?	74
7	Re-viewing Bertin's main example	79
7.1	Bertin's classic book	79
7.2	Bertin's many graphics of the French workforce data	80
7.3	Viewing the data today	82
7.4	Are there other geographic patterns?	84
7.5	Density of the workforce by area	87
7.6	Comments on Bertin (with hindsight)	88
8	Comparing chess players across the board	91
8.1	FIDE ratings of chess players	91
8.2	Comparing chess players by age and sex	93
8.3	Which nations have the best chess players?	97
8.4	Ratings change over time	99
8.5	Grandmasters and Masters	106
9	Results from surveys on gay rights	109
9.1	Survey results by age, sex, and race	109
9.2	Mapping the survey results	112
9.3	Comparing responses to related, but different questions	115
10	Who went up in space for how long?	121
10.1	What human spaceflights have there been?	121
10.2	Mission times in space	124
10.3	Checking the data with graphics	128
11	Data in the sky at night	131
11.1	Data on diamonds for sale	131
11.2	Checking for outliers and unusual values	134
11.3	Intriguing features	138
12	Psoriasis and the Quality of Life	141
12.1	Comparing treatment and placebo groups	141
12.2	Examining responses to individual questions	143

13 Charging electric cars	145
13.1 How was the study organised?	145
13.2 How long does charging cars take?	152
13.3 Charging patterns through the week	154
14 Darwin's Finches	157
14.1 How do Darwin's finches differ?	157
14.2 Do species vary across the Galápagos islands?	162
15 Home or away: where do soccer players play?	167
15.1 The European Football Championship since 1960	167
15.2 Which leagues do international players play in?	170
15.3 How have countries changed?	174
16 Watching soccer—the English leagues	177
16.1 History of the English football leagues since 1888	177
16.2 How have numbers of draws changed?	178
16.3 How big is home advantage?	179
16.4 Have the numbers of goals scored gone down?	181
16.5 How well have teams performed over the years?	182
16.6 Wormcharts show league progress over a season	184
17 Fuel efficiency of cars in the USA	189
17.1 How efficient are different types of car?	189
17.2 Small (?) luxury cars	193
17.3 How many car models have been evaluated over the years?	194
17.4 Comparing fuel efficiency of individual models over time	196
18 Differences amongst the Palmer penguins	199
18.1 Getting an overview of the penguin data	199
18.2 Looking at several variables in one plot	204
18.3 Using raw data to get more information	208
19 Comparing tests for malaria	211
19.1 How much agreement is there between the four tests?	211
19.2 Additional analyses	213
20 Are swimmers swimming faster?	217
20.1 Best times for different strokes	217
20.2 How good are the best swimmers across multiple events?	223
20.3 Comparing events graphically	224
20.4 How do relay races compare with individual races?	227
20.5 How do countries compare?	228

21 Over 90 years running 90 km	231
21.1 How many run in the Comrades (Ultra)Marathon?	231
21.2 Numbers by age and sex	233
21.3 How long do runners take to finish?	234
21.4 Comparing times for men and women	237
21.5 How many runners get medals?	239
21.6 How different are the “up” and “down” races?	240
21.7 How do individual runners perform over the years?	242
22 Comparing software for facial recognition	245
22.1 How well do the softwares do?	245
22.2 Are positive predictive values (PPV) a good criterion?	248
23 Distinguishing shearwaters	251
23.1 Distinguishing shearwaters and puffins	251
23.2 Comparing two shearwater species (Audubon and Galápagos)	252
23.3 Can the species be clustered?	254
24 When do road accidents with deer happen in Bavaria?	259
24.1 Vehicle accidents involving deer	259
24.2 What about vehicle accidents without deer?	264
25 The Titanic Disaster	267
25.1 Who travelled on the Titanic?	267
25.2 Which groups had high survival rates?	272
26 German Election 2021—what happened?	277
26.1 Who won the 2021 election?	277
26.2 Are there geographic patterns?	282
26.3 Who won the first-past-the-post seats?	288
26.4 How do first-past-the-post and party votes compare?	291
II Advice on Graphics	295
27 Provenance and quality	297
27.1 What does provenance mean?	297
27.2 What problems arise with datasets?	298
27.3 The provenance of datasets used in the book	301
27.4 How do invented, simulated, and real datasets compare?	312
28 Wrangling	315
28.1 Data wrangling: what and why	315
28.2 Data cleaning and transforming	315

28.3 Restructuring: subsetting, grouping, aggregating	319
28.4 Naming, renaming and reclassifying	321
28.5 Wrangling for layered graphics	321
28.6 Putting datasets together and reproducibility	328
29 Colour	331
29.1 Theory and implementation	331
29.2 Perception of colour	332
29.3 Value of colour	332
29.4 Choice of colour	338
30 Setting the scene	343
30.1 Scaling	344
30.2 Formatting	348
30.3 Making space	353
30.4 Layout	356
30.5 Comparisons and graphics	360
30.6 Defaults	363
31 Ordering, sorting, and arranging	365
31.1 What gets ordered, sorted, arranged?	365
31.2 Ordering and sorting within graphics	369
31.3 Numbers of orderings and sortings	375
31.4 Arranging graphics	376
32 What affects interpretation?	379
32.1 Attention	379
32.2 Understanding	386
32.3 Checking	390
33 Varieties of plots, developing plots	397
33.1 Varying various plots	397
33.2 Ensembles of plots	400
33.3 Differences between early drafts and later graphics	402
33.4 Classifying graphics	414
34 Conclusions and recommendations	417
References	419
Index	429