

JOHN W. TUKEY

*Princeton University and
Bell Telephone Laboratories*

Exploratory Data Analysis



ADDISON-WESLEY PUBLISHING COMPANY

Reading, Massachusetts • Menlo Park, California

London • Amsterdam • Don Mills, Ontario • Sydney ■

Contents

1	SCRATCHING DOWN NUMBERS (stem-and-leaf)	1
	Comments about the index page	2
	1A Quantitative detective work	1
	1B Practical arithmetic	3
	1C Scratching down numbers	6
	1D Doing better with stem-and-leaf	7
	1E Using the right number of stems	11
	1F How to count by tallying	16
	1G What does it mean to "feel what the data are like"?	19
	1H How far have we come?	20
	1K How to use stem-and-leaf to pick up additional information (optional technique)	23
	1P Additional problems	25
2	SCHEMATIC SUMMARIES (pictures and numbers)	27
	2A Extremes and median	29
	2B Hinges and 5-number summaries	32
	2C Box-and-whisker plots	39
	2D Fences, and outside values	43
	2E Schematic plots	47
	2F Pros and cons; the Rayleigh example	49
	2G Eighths, sixteenths, etc.	53
	2H How far have we come?	55
3	EASY RE-EXPRESSION	57
	3A Logarithms = logs	59
	3B Quick logs	61
	3C Comparisons of two batches	64
	3D Quick roots and quick reciprocals	69
	3E Looking quickly	79
	3F Counted data	83
	3G Relation among powers and logs (optional)	86
	3H How far have we come?	92
	3K How to think about logs (background)	93
	3P Additional problems	93
4	EFFECTIVE COMPARISON (including well-chosen expression)	97
	4A Alternative forms of display of summaries	99
	4B Comparing several batches (continued)	102
	4C A more extensive example	105

4D	The meaning of comparison	110
4E	Adjustments, rough and exact	110
4F	Residuals	113
4H	How far have we come?	115
4P	Additional problems	116
5	PLOTS OF RELATIONSHIP	125
5A	How to plot y against x	126
5B	Looking at subtraction	131
5C	Subtracting straight lines	135
5D	Plotting the population of the U.S.A.	141
5E	Plotting the ratio of births to deaths	148
5F	Untilting defines "tilt"	154
5H	How far have we come?	156
5P	Additional problems	157
6	STRAIGHTENING OUT PLOTS (using three points)	169
6A	Looking at three points	171
6B	Re-expressing y alone	172
6C	Re-expressing x alone	175
6D	A braking example	181
6E	The vapor pressure of H_2O	187
6F	Re-expressing the second variable	191
6G	Wise change of origin as a preliminary	193
6H	How far have we come?	197
6P	Additional problems	199
7	SMOOTHING SEQUENCES	205
7A	Medians of 3	210
7B	Eye resmoothing	214
7C	Looking ahead	216
7D	Copying-on--and more, usually.	221
7E	Blurring the smooth--and setting the fences	223
7F	Splitting peaks and valleys	227
7G	Hanning	231
7H	How far have we come?	235
7+	OPTIONAL SECTIONS FOR CHAPTER 7	237
7I	Breaking a smooth	237
7J	Choice of expression	247
7K	A two-section example	259
7M	How much more may we have learned?	264
8	PARALLEL AND WANDERING SCHEMATIC PLOTS	265
8A	Parallel schematic plots	265
8B	Smoothing the cross-medians	274
8C	Smoothing broken hinges	276
8D	Dealing with the two questions	279

8E	Wandering schematic plots	283
8F	A more demanding example: Governor's salary and bank deposits	287
8G	Further questions/analysis in the example	298
8H	How far have we come?	306
8I	The need to smooth both coordinates (optional)	307
9	DELINEATIONS OF BATCHES OF POINTS	309
9A	E-traces and D-traces	309
9B	Simple delineation--Twin Rivers again	311
9C	Reduced and schematic delineations	313
9D	What our schematic plots and delineations have missed	319
9E	Three variables at once--or more	321
9H	How far have we come?	329
10	USING TWO-WAY ANALYSES	331
10A	Two-way residuals; row-PLUS-column analysis	332
10B	The row-PLUS-column fit	337
10C	Some points of technique	343
10D	Row-TIMES-column analysis	344
10E	Looking at row-PLUS-column fits and their residuals	349
10F	Fitting one more constant	352
10G	Converting PLUS to TIMES; re-expression	358
10H	How far have we come?	360
11	MAKING TWO-WAY ANALYSES	362
11A	Taking medians out	363
11B	Alternative organizations of the arithmetic	372
11C	Making the core of a two-way plot	374
11D	Going on with the residuals	378
11E	Coding residuals; condensing fits and residuals	382
11F	We can combine!	390
11G	Guidance for expression	396
11H	How far have we come?	399
11+	OPTIONAL SECTIONS FOR CHAPTERS 10 AND 11	401
11I	Exploring beyond PLUS-one (extends Chapter 10)	401
11J	Taking out any summary	404
11K	An example of re-expression--city killings	408
11L	An unusual fit	415
11M	How much more may we have learned?	419
12	ADVANCED FITS	420
12A	PLUS-one fits	421
12B	Pictures for "-PLUS-one" fits	424
12C	Making those pictures	428
12D	Sometimes we can have parallel-line plots, still	431
12E	More extended fits	433
12F	Simplification is sometimes possible	438
12H	How far have we come?	441

13	THREE-WAY FITS	443
	13A Three- and more-way analyses: Arrangement and tagging	443
	13B An analysis of the psychological example	448
	13C Making three-way analyses	452
	13D Three-way re-expression	458
	13E More about the example	462
	13H How far have we come?	465
14	LOOKING IN TWO OR MORE WAYS AT BATCHES OF POINTS	466
	14A Coordinates and level traces	467
	14B Different middle traces for the same slices	470
	14C An explanation	475
	14D Changing the slicing coordinate	476
	14E What matters?	481
	14F Rematching and strength of relationship	482
	14H How far have we come?	491
	14I The ubiquity of medians (optional section)	492
15	COUNTED FRACTIONS	494
	15A Started counts and counted fractions	496
	15B Three matched scales for counted fractions	498
	15C Quicker calculation	502
	15D Examples where careful expression clearly pays off	508
	15E Double folding--the 2×2 case	513
	15F Double folding--larger cases	516
	15G Easy froots and flogs with a slide rule (optional)	520
	15H How far have we come?	522
16	BETTER SMOOTHING	523
	16A Reroughing	523
	16B Some examples	526
	16C If we want things still smoother	531
	16D Further possibilities	534
	16H How far have we come?	542
17	COUNTS in BIN after BIN	543
	17A Root smooth and root rough	543
	17B Counts of basic counts	550
	17C Fitting to smoothed roots	555
	17D Corn borers, wheat prices, and Student's simulations	561
	17E Bins of unequal width	570
	17F Double roots	576
	17G Cautionary examples	582
	17H How far have we come?	587
18	PRODUCT-RATIO PLOTS	588
	18A Sizes and counts	589
	18B Product-ratio analysis	594
	18C Forcing the unusual to be noticed	598

18D	Comparisons between collections	602
18E	Looking at the smallest basic count	604
18F	When zeros are counted	605
18G	Under the microscope	608
18H	How far have we come?	612
19	SHAPES OF DISTRIBUTION	614
19A	Looking at shapes of distribution	616
19B	The Gaussian reference	623
19C	Using letter values to look at shapes of distribution	626
19D	Pushback technique (optional section)	637
19H	How far have we come?	644
20	MATHEMATICAL DISTRIBUTIONS	646
20A	Binnings vs. distributions	648
20B	Densities for distributions vs. densities for binnings	651
20C	Tables and pictures comparing two sets of shapes of distributions	654
20H	How far have we come?	661
21	POSTSCRIPT	662
21A	Our relationship to the computer	663
21B	What has been omitted?	664
21C	How should the past chapters look different?	665
21D	What have we been introduced to?	666
	GLOSSARY	667
	INDEX TO REFERENCE TABLES	677
	ALPHABETICAL INDEX	677
	FRONTPAPERS	
	1. Break table for two-decimal logs	
	2. Break table for (square) roots	
	3. Main break table--digits of negative reciprocals	
	REARPAPERS	
	4. Pluralities, folded roots, folded logarithms	
	5. Values of $\log_e \sqrt{\text{count} + \frac{1}{8}}$	
	6. Values of $\sqrt{\text{count} + \frac{1}{8}}$	