

## References and Reference Cards

There are oceans of literature about statistics, about R and about both. Below is a small selection of publications which are either mentioned in the text, or could be really useful (as we think) to readers of this book.

2em-2em1ex

Cleveland W. S. 1985. The elements of graphing data. Wandsworth Advanced Books and Software. 323 p.

Crawley M. 2007. R Book. John Wiley & Sons. 942 p.

Dalgaard P. 2008. Introductory statistics with R. 2 ed. Springer Science Business Media. 363 p.

Efron B. 1979. Bootstrap Methods: Another Look at the Jackknife. Ann. Statist. 7(1): 1–26.

Gonick L., Smith W. 1993. The cartoon guide to statistics. HarperCollins. 230 p.

Kaufman L., Rousseeuw P. J. 1990. Finding groups in data: an introduction to cluster analysis. Wiley-Interscience. 355 p.

Kimble G. A. 1978. How to use (and misuse) statistics. Prentice Hall. 290 p.

Li Ray. Top 10 data mining algorithms in plain English. URL: <http://rayli.net/blog/data/top-10-data-mining-algorithms-in-plain-english/>

Li Ray. Top 10 data mining algorithms in plain R. URL: <http://rayli.net/blog/data/top-10-data-mining-algorithms-in-plain-r/>

Marriott F. H. C. 1974. The interpretation of multiple observations. Academic Press. 117 p.

McKillup S. 2011. Statistics explained. An introductory guide for life scientists. Cambridge University Press. 403 p.

Murrell P. 2006. R Graphics. Chapman & Hall/CRC. 293 p.

Petrie A., Sabin C. 2005. Medical statistics at a glance. John Wiley & Sons. 157 p.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rowntree D. 2000. Statistics without tears. Clays. 195 p.

Sokal R. R., Rolf F. J. 2012. Biometry. The principles and practice of statistics in biological research. W.H. Freeman and Company. 937 p.

Sprent P. 1977. Statistics in Action. Penguin Books. 240 p.

Tukey J. W. 1977. Exploratory Data Analysis. Pearson. 688 p.

Venables W. N., Ripley B. D. 2002. Modern applied statistics with S. 4th ed. Springer. 495 p.

*Happy Data Analysis!*

And just a reminder: if you use R and like it, do not forget to *cite it*. Run `citation()` command to see how.

Reference cards are attached to the very end of the book. They have a different page format, more suitable for printing. The first one was is actually one-page “cheatsheet”; we recommend to print is and use while you learn R.

- 
1. <https://xkcd.com/thing-explainer/>↵
  2. There is however the `SOAR` package which overrides this behavior.↵
  3. If you do not use these managers or centers, it is recommended to regularly *update* your R, at least once a year.↵
  4. There is command `xpager()` in the `asmisc.r` collection of commands, it allows to see help in the separate window even if you work in terminal.↵
  5. Within parentheses immediately after example, we are going to provide comments.↵
  6. By the way, on Linux systems you may exit R also with `Ctrl+D` key, and on Windows with `Ctrl+Z` key.↵
  7. Usually, small exercises are **boldfaced**.↵
  8. By the way, if you want the Euler number,  $e$ , type `exp(1)` .↵
  9. And also like editor which is embedded into R for Windows or into RmacOS GUI, or the editor from `rite` R package, but **not** office software like MS Word or Excel!↵

10. Yet another possibility is to set working directory in preferences (this is quite different between operating systems) but this is not the best solution because you might (and likely will) want different working directories for different tasks.↵
11. There is `rio` package which can determine the structure of data.↵
12. Again, download it from Internet to `data` subdirectory first. Alternatively, replace subdirectory with URL and load it into R directly—of course, after you check the structure.↵
13. On macOS, type `Enter` twice.↵
14. With commands `dput()` and `dget()`, R also saves and loads textual representations of objects.↵
15. This is a bit similar to the joke about mathematician who, in order to boil the kettle full with water, would empty it first and therefore *reduce the problem to one which was already solved!*↵
16. If, by chance, it started and you have no idea how to quit, press uppercase `ZQ`.↵
17. Within nano, use `Ctrl+O` to save your edits and `Ctrl+X` to exit.↵
18. Does not work on graphical macOS.↵
19. Under graphical macOS, this command is not accessible, and you need to use application menu.↵
20. You can also use `savehistory()` command to make a “starter” script.↵
21. On Windows and macOS, this will open internal editor; on Linux, it is better to set `editor` option manually, e.g.,  
`file.edit("hello.r", editor="geany")`.↵
22. The better term is *generic command*.↵
23. Cleveland W. S., McGill R. 1985. Graphical perception and graphical methods for analyzing scientific data. *Science*. 229(4716): 828–833.↵
24. `lattice` came out of later ideas of W.S. Cleveland, *trellis* (conditional) plots (see below for more examples).↵
25. `ggplot2` is now most fashionable R graphic system. Note, however, that it is based on the different “ideology” which related more with SYSTAT visual statistic software and therefore is alien to R.↵
26. By the way, both PDF and SVG could be opened and edited with the freely available vector editor Inkscape.↵
27. Collection `gmoon.r` has game-like command `Miney()`, based on `locator()`; it partly imitates the famous “minesweeper” game.↵
28. In the case of our `eggs` data frame, the command of second style would be `plot(eggs[, 1:2])` or `plot(eggs$V1, eggs$V2)`, see more explanations in the next chapter.↵
29. Another variant is to use high-level `scatter.smooth()` function which replaces `plot()`. Third alternative is a cubic smoother `smooth.spline()` which calculates numbers to use with `lines()`.↵
30. Discrete measurement data are in fact more handy to computers: as you might know, processors are based on 0/1 logic and do not readily understand non-integral, floating numbers.↵
31. For unfamiliar words, please refer to the glossary in the end of book.↵
32. By default, `Ls()` does not output functions. If required, this behavior could be changed with `Ls(exclude="none")`.↵
33. In fact, columns of data frames might be also matrices or other data frames, but this feature is rarely useful.↵
34. There is also `hexbin` package which used hexagonal shapes and color shading.↵
35. Package `DescTools` has the handy `Mode()` function to calculate mode.↵
36. While it is possible to run here a cycle using `for` operator, `apply`-like functions are always preferable.↵
37. In the book, we include minimum and maximum into quartiles.↵
38. Note that these options must be set *a priori*, before you run the test. It is not allowed to change alternatives in order to find a better p-values.↵
39. Look also into the end of this chapter.↵
40. There is a workaround though, *robust rank order test*, look for the function `Rro.test()` in the `asmisc.r`.↵
41. Bennett C.M., Wolford G.L., Miller M.B. 2009. The principled control of false positives in neuroimaging. *Social cognitive and affective neuroscience* 4(4): 417–422, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2799957/>↵
42. Like it is implemented in the `ARTool` package; there also possible to use multi-way nonparametric designs.↵
43. Fisher R.A. 1971. *The design of experiments*. 9th ed. P. 11.↵
44. Mendel G. 1866. Versuche über Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereines in Brünn*. Bd. 4, Abhandlungen: 12. <http://biodiversitylibrary.org/page/40164750>↵
45. Yates F. 1934. Contingency tables involving small numbers and the  $\chi^2$  test. *Journal of the Royal Statistical Society*. 1(2): 217–235.↵
46. There are, however, advanced techniques with the goal to understand the difference between causation and correlation: for example, those implemented in `bnlearn` package.↵

47. Function `Cladd()` is applicable only to simple linear models. If you want confidence bands in more complex cases, check the `Cladd()` code to see what it does exactly.↵
48. Fisher R.A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*. 7(2): 179–188.↵
49. Package `Boruta` is especially god for all relevant feature selection.↵
50. For example, “Encyclopedia of Distances” (2009) mentions about 1,500!↵
51. Emphasis mine.↵
52. With command `source("ashipunov.info/r/gmoon.r")` .↵
53. To know which symbols are available, run `demo(Hershey)` .↵
54. Linux users might want to add option `editor=` .↵
55. Package `lintr` contains `lint()` command which checks R scripts.↵
56. There is, by the way, a life-hack for lazy reader: all plots which you need to make yourself are actually present in the output PDF file.↵
57. Among text editors, Geany is one of the most universal, fast, free and works on most operation systems.↵
58. Thompson D. W. 1945. *On growth and form*. Cambridge, New York. 1140 pp.↵
59. Rohlf F.J. tpsDig. Department of Ecology and Evolution, State University of New York at Stony Brook. Freely available at [life.bio.sunysb.edu/morph/](http://life.bio.sunysb.edu/morph/)↵
60. Actually, `geomorph` package is capable to digitize images with `digitize2d()` function but it works only with JPEG images.↵