

Appendix A

Basic String Manipulations in R

In this Appendix we summarize some of the basic string manipulation capabilities of R. For more powerful and complex capabilities, especially regular expressions, consult Chapter 5 in Gentleman [26].

Elementary string manipulations

First, use a simple test string to show the `substring` function.

```
> ts="abcde"
> substring(ts,1,1) # What is the first character in ts?
[1] "a"
> substring(ts,1,1)="X" # Change the first character to "X"
> ts
[1] "Xbcde"
> substring(ts,1,3) # What are the first three characters in ts?
[1] "Xbc"
```

`nchar` counts the number of characters in a string, which is different from the `length` (number of elements) of the string vector.

```
> nchar(ts) # Number of characters in ts
[1] 5
> length(ts) # Length of the vector ts
[1] 1
```

As a more relevant biological example, we examine the beginning of the lambda bacteriophage genome. We start with the first 70 bases.

```
lam70="GGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTTCCGGTT
      TAAGGCGTTTCCGTTCTTCTTCG"
> nchar(lam70)
[1] 70
# Keep the first 60 bases in lam60
```

```

> lam60 = substr(lam70,1,60)
> lam60
[1] "GGGCGGCGACCTCGCGGGTTTTTCGCTATTTATGAAAATTTTCCGG
    TTTAAGGCGTTTCCG"
> nchar(lam60)
[1] 60

```

Now we break lam60 into 20 triplets with substring.

```

> lam60.triplets = substring(lam60, seq(1,58,3),
    seq(3,60,3))
> lam60.triplets
[1] "GGG" "CGG" "CGA" "CCT" "CGC" "GGG" "TTT"
    "TCG" "CTA" "TTT" "ATG" "AAA" "ATT" "TTC"
[15] "CGG" "TTT" "AAG" "GCG" "TTT" "CCG"

```

The related function `substr` would yield just the first triplet.

The reverse of `substring` is `paste`.

```

paste(lam60.triplets,sep="")
> paste(lam60.triplets,sep="", collapse="")
[1] "GGGCGGCGACCTCGCGGGTTTTTCGCTATTTATGAAAATTTTCCGGTTT
    AAGGCGTTTCCG"

```

The R base installation has functions `toupper` and `tolower` to convert strings between upper and lower case. The function `chartr` translates specified characters, and is useful to convert from DNA to RNA.

```

> tolower("ABCDE")
[1] "abcde"
> toupper("abcde")
[1] "ABCDE"
> chartr("T","U",lam60)
[1] "GGGCGGCGACCU CGCGGGUUUUCGCUAUUUUAUGAAAAUUUUCGGUUU
    AAGGCGUUUCCG"

```

`chartr` can change more than one character at a time.

```

> chartr("ab","12","aabb")
[1] "1122"

```

We can thus use `chartr` to obtain the complement of the sequence.

```

> complam60 = chartr("ACGT", "TGCA", lam60)
> paste(rev(unlist(strsplit(complam60,split=""))),
    collapse="")
[1] "CGGAAACGCCTTAAACCGGAAAATTTTCATAAATAGCGAAAACCCGC
    GAGGTCGCCGCC"

```

By combining `chartr` to convert the bases to their complements, `strsplit` to split the complement string into its components (remember to convert the resulting

list into a vector with `unlist`), `rev` to reverse the elements of the vector, and `paste` to recombine the vector elements into a string, one can write a function `revcomp` to obtain the reverse complement of the original DNA string.

```
> revcomp = function(DNAstr) {
  step1 = chartr("ACGT", "TGCA", DNAstr)
  step2 = unlist(strsplit(step1, split=""))
  step3 = rev(step2)
  step4 = paste(step3, collapse="")
  return(step4)
}
> revcomp(lam60)
[1] "CGGAAACGCCTTAAACCGGAAAATTTTCATAAATAGCGAAAACCCGC
    GAGGTCGCCGCC"
```

Genetic code translation

The Bioconductor package `Biostrings` contains the genetic code to convert triplets of bases to amino acids.

```
> source("http://bioconductor.org/biocLite.R")
biocLite("Biostrings")

> library(Biostrings)

Attaching package: 'Biostrings'

> paste(GENETIC_CODE[lam60.triplets], collapse="")
[1] "GRRPRGFSLFMKIFRFKAFF"
> strsplit("GRRPRGFSLFMKIFRFKAFF", split="")
[[1]]
[1] "G" "R" "R" "P" "R" "G" "F" "S" "L" "F" "M" "K"
    "I" "F" "R" "F" "K" "A" "F" "P"
```

`Biostrings` contains the method `AMINO_ACID_CODE` to translate one-letter to three-letter amino acid codes. It selects the array element corresponding to the one-letter code, and replaces it with the three-letter equivalent. We apply it to the amino acid sequence above.

```
> AMINO_ACID_CODE[strsplit(as.character("GRRPRGFSLFMK
  IFRFKAFF"), NULL)[[1]]]
      G      R      R      P      R      G      F      S      L
"Gly" "Arg" "Arg" "Pro" "Arg" "Gly" "Phe" "Ser" "Leu"
      F      M      K      I      F
"Phe" "Met" "Lys" "Ile" "Phe"
```

```
      R      F      K      A      F      P  
"Arg" "Phe" "Lys" "Ala" "Phe" "Pro"
```

Two things should be noted about this code. `strsplit` returns a list, and `[[1]]` picks out the first (and only) element of the list, which is a vector. The outer square brackets then denote the element of the vector to which `AMINO_ACID_CODE` is applied.

`Biostrings` also has functions to reverse or complement nucleic acid sequences: `reverse`, `complement`, and `reverseComplement` operate on DNA and RNA sequences, which must be declared as such. See `help(reverse)` for details.

References

1. Adler, F.: Modeling the Dynamics of Life: Calculus and Probability for Life Scientists, 2nd edn. Brooks/Cole (2005)
2. Allman, E., Rhodes, J.: Mathematical Models in Biology: An Introduction. Cambridge University Press (2004)
3. Alon, U.: An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman & Hall/CRC, Boca Raton, FL (2007)
4. Altman, D.: Practical Statistics for Medical Research. Chapman & Hall (1991)
5. Barkai, N., Leibler, S.: Robustness in simple biochemical networks. *Nature* **387**, 913–917 (1997)
6. Beard, D., Qian, H.: Chemical Biophysics: Quantitative Analysis of Cellular Systems. Cambridge University Press (2008)
7. Berg, H.: *E. coli in Motion*. Springer (2003)
8. Bloomfield, V., Crothers, D., Tinoco Jr., I.: Nucleic Acids: Structures, Properties, and Functions. University Science Books (2000)
9. Boroujerdi, M.: Pharmacokinetics: Principles and Applications. McGraw-Hill (2002)
10. Braun, M.: Differential Equations and Their Applications. An Introduction to Applied Mathematics, 4th edn. Springer (1993)
11. Cicirelli, M., Smith, L.: Cyclic amp levels during the maturation of xenopus oocytes. *Dev. Biol.* **108**, 254–258 (1985)
12. Council, N.R.: BIO 2010: Transforming Undergraduate Education for Future Research Biologists. The National Academies Press (2003)
13. Crank, J.: The Mathematics of Diffusion, 2nd edn. Oxford University Press (1975)
14. Cropper, W.: Mathematica Computer Programs for Physical Chemistry. Springer (1998)
15. Dalggaard, P.: Introductory Statistics with R. Springer (2002)
16. von Dassow, G., Meir, E., Munro, E., Odell, G.: The segment polarity network is a robust developmental module. *Nature* **406**, 188–192 (2000)
17. von Dassow, G., Odell, G.: Design and constraints of the *Drosophila* segment polarity module: Robust spatial patterning emerges from intertwined cell state switches. *J. Exp. Zool. Mol. Dev. Evol.* **294**, 179–215 (2002)
18. Deonier, R., Tavaré, S., Waterman, M.: Computational Genome Analysis: An Introduction. Springer (2005)
19. Diekmann, O., Heesterbeek, J.: Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. John Wiley & Sons (2000)
20. Eldar, A., Dorfman, R., Weiss, D., Ashe, H., Shilo, B.Z., Barkai, N.: Robustness of the bmp morphogen gradient in *Drosophila* embryonic patterning. *Nature* **419**, 304–308 (2002)
21. Eldar, A., Rosin, D., Shilo, B.Z., Barkai, N.: Self-enhanced ligand degradation underlies robustness of morphogen gradients. *Dev. Cell* **5**, 635–646 (2003)

22. Eldar, A., Shilo, B.Z., Barkai, N.: Elucidating mechanisms underlying robustness of morphogen gradients. *Curr. Opin. Genet. Dev.* **14**, 436–439 (2004)
23. Everitt, B., Hothorn, T.: *A Handbook of Statistical Analyses Using R*. Chapman & Hall/CRC (2006)
24. Fall, C., Marland, E., Wagner, J., Tyson, J. (eds.): *Computational Cell Biology*. Springer (2002)
25. Fell, D.: *Understanding the Control of Metabolism*. Portland Press (1997)
26. Gentleman, R.: *R Programming for Bioinformatics*. Chapman & Hall/CRC (2008)
27. Gentleman, R., Carey, V., Huber, W., Irizarry, R., Dudoit, S. (eds.): *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer (2005)
28. Gillespie, D.: Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem* **81**, 2340 (1977)
29. Gillespie, D.: Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58**, 35 (2007)
30. Groen, A., Westerhoff, H.: Modern control theories: A consumers' test. In: A. Cornish-Bowden, M. Cardenas (eds.) *Control of Metabolic Processes*, chap. 6, pp. 101–118. Plenum Press (1990)
31. Hahne, F., Huber, W., Gentleman, R., Falcon, S.: *Bioconductor Case Studies*. Springer (2008)
32. Heinrich, R., Rapoport, T.: A linear steady-state treatment of enzymatic chains: general properties, control and effector strength. *Eur. J. Biochem.* **42**, 89–95 (1974)
33. Heinrich, R., Schuster, S.: *The Regulation of Cellular Systems*. Chapman & Hall (1996)
34. Holling, C.: The components of predation as revealed by a study of small mammal predation of the european pine sawfly. *Can. Entomol.* **91**, 293–320 (1959)
35. Ingolia, N.: Topology and robustness in the *Drosophila* segment polarity network. *PLoS Biol.* **2**(6), 805–815 (2004)
36. Jones, O., Maillardet, R., Robinson, A.: *Introduction to Scientific Programming and Simulation Using R*. Chapman & Hall/CRC (2009)
37. Kacser, H., Burns, J.: The control of flux. *Symp. Soc. Exp. Biol.* **27**, 65–104 (1973)
38. Keeling, M., Rohani, P.: *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press (2008)
39. Keen, R., Spain, J.: *Computer Simulation in Biology: A BASIC Introduction*. Wiley-Liss (1992)
40. Kermack, W., McKendrick, A.: A contribution to the mathematical theory of epidemics. *Proc. R. Soc. London, Ser. A* **115**, 700–721 (1927)
41. Leslie, P., Gower, J.: Properties of a stochastic model for the predator-prey type of interaction between two species. *Biometrika* **47**, 219–301 (1960)
42. Macheras, P., Iliadis, A.: *Modeling in Biopharmaceutics, Pharmacokinetics, and Pharmacodynamics: Homogeneous and Heterogeneous Approaches*. Springer (2006)
43. Maindonald, J., Braun, J.: *Data Analysis and Graphics Using R: An Example-Based Approach*, 2nd edn. Cambridge University Press (2007)
44. Mangan, S., Alon, U.: Structure and function of the feed-forward loop network motif. *PNAS* **100**(21), 11,980–11,985 (2003)
45. Milanese, M., Molino, G.: Structural identifiability of compartmental models and pathophysiological information from the kinetics of drugs. *Math. Biosci.* **26**, 175–190 (1975)
46. Molino, G., Milanese, M.: Structured analysis of compartmental models for the hepatic kinetics of drugs. *J. Lab. Clin. Med.* **85**, 865–878 (1975)
47. Murrell, P.: *R Graphics*. Chapman & Hall/CRC (2006)
48. Nowak, M., May, R.: *Virus Dynamics: Mathematical Principles of Immunology and Virology*. Oxford University Press (2000)
49. Paradis, E.: *Analysis of Phylogenetics and Evolution with R*. Springer (2006)
50. Pearl, R.: The growth of populations. *Q. Rev. Biol.* **2**, 532–548 (1927)
51. Powers, W., Canale, R.: Some applications of optimization techniques to water quality modeling and control. *IEEE Trans. Syst., Man, Cybernetics* **5**, 312–321 (1975)
52. Samuels, M., Witmer, J.: *Statistics for the Life Sciences*, 3rd edn. Pearson Education (2003)

53. SantaLucia Jr., J., Allawi, H., Seneviratne, P.: Improved nearest-neighbor parameters for predicting dna duplex stability. *Biochemistry* **35**, 3555–3562 (1996)
54. Savageau, M.: *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Addison-Wesley (1976)
55. Seefeld, K., Linder, E.: *Statistics Using R with Biological Examples*. http://cran.r-project.org/doc/contrib/Seefeld_StatsRBio.pdf (2007)
56. Simon, R., Korn, E., McShane, L., Radmacher, M., Wright, G., Zhao, Y.: *Design and Analysis of DNA Microarray Investigations*. Springer (2003)
57. Smith, H., Waltman, P.: *The Theory of the Chemostat: Dynamics of Microbial Competition*. Cambridge University Press (2008)
58. Staub, J.F., Brezillon, P., Perault-Staub, A., Milhaud, G.: Nonlinear modeling of calcium metabolism—first attempt in the calcium-deficient rat. *Trans. Inst. Measur. Control (London)* **3**, 89–97 (1981)
59. Steinmetz, C., Larter, R.: The quasiperiodic route to chaos in a model of the peroxidase-oxidase reaction. *J. Chem. Phys.* **94**(2), 1388–1396 (1991)
60. Stekel, D.: *Microarray Bioinformatics*. Cambridge University Press (2003)
61. Tanner, J.: The stability and intrinsic growth rates of predator and prey populations. *Ecology* **56**, 855–867 (1975)
62. Tinoco Jr., I., Sauer, K., Wang, J., Puglisi, J.: *Physical Chemistry: Principles and Applications in Biological Sciences*, 4th edn. Prentice-Hall (2002)
63. Torres, N., Voit, E.: *Pathway Analysis and Optimization in Metabolic Engineering*. Cambridge University Press (2002)
64. Venables, W., Ripley, B.: *Modern Applied Statistics with S*, 4th edn. Springer (2002)
65. Verzani, J.: *Using R for Introductory Statistics*. Chapman & Hall/CRC (2005)
66. Voit, E.: *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*. Cambridge University Press (2000)
67. Wilkinson, D.: *Stochastic Modelling for Systems Biology*. Chapman & Hall/CRC (2006)
68. Wolpert, L.: Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25**, 1–47 (1969)
69. Xia, T., SantaLucia Jr., J., Burkard, M., Kierzek, R., Schroeder, S., Jiao, X., Cox, C., Turner, D.: Thermodynamic parameters for an expanded nearest-neighbor model for formation of rna duplexes with watson-crick base pairs. *Biochemistry* **37**, 14,719–14,735 (1998)

Index

- < -, assignment operator, 6
- >, prompt symbol, 5
- #, comment symbol, 5

- abline, 32

- barplot, 25
- Bioconductor, 82, 281, *see also* microarray
 - Biostrings, 311
 - mass spectrometry
 - PROcess, 273
- Brownian motion, 167

- calculations
 - arithmetic, 5
 - complex numbers, 5
 - rounding, 8
 - vector, 8
- cbind, 40, 258
- character vectors, 11
- Chi-squared statistic, 239
- command window, 5
- compartmental models, 169
 - calcium metabolism, 172
 - liver function
 - multi-compartment, 171
 - two-compartment, 170
 - periodic dosage, 170
- conditional execution
 - if, 56
- constants, built-in
 - months, letters, pi, 45
- CRAN, 3
- curve, 26

- data
 - attach, 257
 - protein, 76
 - reading from file, 70
 - spreadsheet files, reading and writing, 71
 - tables, reading and writing, 71
 - writing to file, 70
- data frame, 19, 73
- decimal places, 5
- demonstrations, 3
- diagram, constructing, 37
 - expression, 37
 - plot.new, 37
 - plot.window, 37
 - plotmath, 37
 - segments, 37
- differential equations, 113
 - analytical solution, 113
 - limited growth, 117, 141
 - logistic, 141
 - second-order, 133
 - stochastic, 133
- differential equations, systems of
 - 4th order Runge-Kutta, 124
 - Euler method, 121
 - improved Euler method, 122
 - lsoda, 128
 - odesolve, 124
 - stability of, 130
- diffusion, 159
 - Brownian motion, 167
 - bacterial chemotaxis, 193
 - countercurrent, 166
 - Fick's first law, 159
 - Fick's second law, 160
 - Crank's stability criterion, 164
 - numerical solution, 163
 - spherical and cylindrical geometries, 165
- directory, changing, 69

- electrophoresis, 166
- enzyme kinetics, 175
 - competitive inhibition, 108
 - Eadie-Hofstee plot, 104, 106
 - Lineweaver-Burk plot, 23, 28, 103, 104
 - Michaelis-Menten mechanism, 23, 30, 31, 42, 103, 136
- epidemics, 149
 - gonorrhea, 154
 - SI model of fatal infections, 153
 - SIR model, 150
 - with births and deaths, 152
 - SIS model of infection without immunity, 153
- equilibrium, chemical, 85
 - equilibrium constant
 - experimental determination, 96
 - temperature dependence, 97
 - ligand binding, 85
 - cooperative, 90
 - Hill plot, 94
 - multiple sites, 89
 - oxygen binding by hemoglobin, 92
 - Scatchard plot, 87
 - nucleic acids, 98
 - thermodynamic principles, 98
- exporting data, 69
- factor, 19, 75
- for loop, 56
- Fourier transform, 246
 - fft, 246
 - periodogram, 247
 - spectrum, 247
- functions
 - built-in, 49
 - Gaussian, 52
 - Lorentzian, 65
 - mathematical, 7
 - matrix, 19
 - string, *see* string manipulation
 - user-defined, 52
 - vector, 10
 - rev, 311
 - unlist, 311
- genetic code translation, 311
- genomics, *see* microarray
- graph, 23
 - 2d
 - contour, 27
 - image, 27, 284
 - 3d
 - persp, 27
- arrows, 36
- axes, 34
- axis, 35
- axis labels, 31
- axis limits, 34
- barplot
 - cross-hatching, 44
- boxplot, 286
- colors, 31
- curve, 26
- data, 23, 25
 - with line, 29
- data, superimposing two or more series, 39
- different sized plots in a figure, 44
- digitizing
 - identify, 45
 - locator, 45
- error bars, 44
- function, 26
- histogram, 26, 252
 - stacked, 258
- line thickness, 31
- line type, 30
- line width, 30
- math expressions, adding, 36
- matplotlib, 40
- placing several plots in a figure, 42
- plot characters
 - size, cex, 29
- plotting points and lines, 40
- setting parameters with par, 35
- straight lines with abline, 32
- superimposing curves, 41
- text, adding, 35
- ticks, 34
- title, 31
- graphi
 - boxplot, 284
- if, 56
 - ifelse, 56
- importing data, 69
- isoelectric point, 54
- kinetics, 113
 - bimolecular reaction, 117
 - chemical interconversion, 115
 - oscillating reaction, 128, 136, 138
 - stability of, 132
 - T-jump method, 137
 - temperature-dependence, 137
- layout, 44
- least-squares

- non-linear, 109
- LETTERS, 45
- letters, 45
- list, 20
- logical
 - TRUE, FALSE, 13
- Lotka-Volterra model, 125, 145
 - stability of, 131
- Markov chain, 214
 - diffusion, 214
 - dinucleotide frequencies, 240
- matrix, 14
 - arithmetic operations, 15
 - determinant, 16
 - diagonal, 17
 - eigenvalues, 18
 - inverse, 17
 - multiplication, 16
 - other functions, 19
 - power, 242
 - transpose, 17
- metabolism, regulation of, 175
 - biochemical systems theory, 184
 - feedback inhibition, 185
 - disequilibrium ratio, 178, 180
 - metabolic control analysis, 181
 - connectivity theorem, 184
 - elasticities, 183
 - flux control coefficients, 182
 - stability, 186
 - steady-state flux, 180
 - successive reactions, 175, 178
- microarray, 279
 - Affymetrix, 280
 - Bioconductor
 - affy package, 281
 - AffyBatch class, 282
 - ExpressionSet, 284, 291
 - exprs, 292
 - importing data in nonstandard form, 284
 - preprocessing with *expresso*, 291
 - detecting groups of genes, 297
 - clustering, 300
 - correlation, 298
 - principal component analysis, 302
 - differential gene expression, 293
 - paired samples, 293
 - identifying highly expressed genes, 292
 - preprocessing, 280, 284, 286
 - background correction, 288
 - Dilution dataset, 286
 - MAplot, 287
 - normalization, 289
 - probe mismatch correction, 290
 - summarization, 290
 - two-color spot, 280
 - Monte Carlo method, 210
 - month.abb, 45
 - month.name, 45
- nls, 109, 267
 - residuals, 270
 - summary, 269
- nucleic acids
 - helix-coil transition, 98
 - calculation of melting temperature, 99
 - parameters for DNA, 99
 - parameters for RNA, 99
- numerical analysis, 58
 - differentiation, 64
 - integration, 63
 - smoothing noisy data, 67
- numerical differentiation, 64, 67
 - diff, 64
- numerical integration, 63
 - rectangular, 66
 - Simpson's rule, 67
 - trapezoidal, 66
- optim, 62, 85
- order, 50
- package, 72
 - datasets, 72
 - MASS, 82
 - Package Manager, 82
 - seqinr, 76
- par, 35
 - compressing space with, 43
 - mfrow and mfcpl, 42
- periodogram, 247
- pH titration curves, 54
- phase plot, 145
- plot characters, 28
- population dynamics, 141, *see also* differential
 - equations, systems of
- epidemics, 149
- logistic, 141
 - variable carrying capacity, 143
- Lotka-Volterra model, 145
- microbial growth, 146
 - batch culture, 147
 - chemostat, 147
 - competition for limiting nutrients, 148
 - Monod model, 147
 - multiple limiting nutrients, 148
 - toxic inhibition of, 149

- stochastic models, 210
 - population growth, 211
 - radioactive decay, 213
 - two-species competition, 146
- population genetics, 228
 - effect of selection, 229
 - Hardy-Weinberg principle, 228
 - mutation, 230
 - selection for heterozygotes, 229
 - sex-linked recessive genes, 230
- probability, 207
 - basic ideas, 207
 - Bayes' theorem, 209
 - conditional, 208
 - law of total probability, 208
- probability distribution, 216
 - binomial, 222
 - choose, 223
 - sample, 223, 237
 - chi-square, 228
 - exponential, 227
 - F, 228
 - geometric, 226
 - lognormal, 228
 - normal, 219
 - Poisson, 224, 244
 - Student's t, 228
 - uniform, 217
- programming, 55
- proteomics
 - mass spectrometry, 272
- quantitative biology
 - BIO 2010, viii
- R
 - advantages, ix
 - demos, 3
 - downloading, 3
 - getting help, 4
 - installing, 3
 - useful references, ix
- regulation
 - bacterial chemotaxis, 193
 - robust mechanism, 195
 - metabolism, *see* metabolism, regulation of
 - morphogenesis, 197
 - exponential morphogen gradient, 198
 - patterning in dorsal region of *Drosophila*, 200
 - robustness, 197
 - self-enhanced morphogen degradation, 199
 - transcription, 191
 - feed-forward loops, 191
 - repeat loop, 58
 - root-finding, 59
 - for polynomial, polyroot, 60
 - Newton's method, 65
- sample, 52
- seqinr, 76
- sequence analysis, DNA, 233
 - counting bases, 237
 - dinucleotide frequencies, 239
 - Entrez, 233
 - periodicity, 246
 - restriction fragment lengths, 245
 - restriction sites, 244
 - simulating distribution of bases, 238
 - simulating random sequence, 237
- sequences, 11
 - of random numbers, 12
- simultaneous equations
 - linear, 61
 - nonlinear, 62
- sort, 50
- spectrum, 247
- splines, 51
- split.screen, 44
- statistics
 - analysis of variance (anova), 296
 - bootstrapping, 296
 - clustering, 300
 - correlation, 298
 - distance measures, 298
 - differential gene expression, *see* microarray
 - principal component analysis, 302
 - biplot, 303
 - type I and type II errors, 265
- statistics, analysis of noisy spectra, 267
 - baseline subtraction, 271, 274
 - fitting to sum of decaying exponentials, 267
 - fitting to sum of Gaussians, 269
 - mass spectrometry, 272
 - peak identification, 275
 - residuals, 268
- statistics, single sample, 251
 - bins, 252
 - boxplot, 255
 - cumulative distribution, 253
 - grouped data, 257
 - histogram, 252
 - mean, 251
 - median, 251
 - power, 306
 - quantile, 252
 - quartile, 252

- standard deviation sd, 251
- summary, 252
- tapply, 258
- test of normal distribution, qqnorm, 254
- variance var, 251
- statistics, two-sample comparison, 258
 - power, 265, 305
 - power.t.test, 265
 - t statistic, 260
 - t test, 258
 - assumptions for use of, 260
 - one-sample, 260
 - two-sample, 262, 295
 - two-sample paired, 264, 294
 - Welch test with unequal variances, 263
 - Wilcoxon (Mann-Whitney), 264, 294
- steady state, 85
 - enzyme kinetics, 103
- stochastic differential equations, 133
 - Gillespie algorithm, 133
- stochastic modeling, 210
- string manipulation
 - as.numeric, 234
 - chartr, 310
 - collapse, 238
 - gsub, 234
 - nchar, 309
 - paste, 238, 310, 311
 - sep, 238
 - strsplit, 234, 311
 - substr, 310
 - substring, 309
 - tolower, 310
 - toupper, 310
- text editing, 4
- unit conversions, 6
- variables, assigning, 6
- vectors, 8
 - character, 11
 - logical, 13
- which, 58
- while loop, 57