

# INDEX

## Symbols and Numbers

- & (element-wise AND operator), 65, 319
- && (single comparison AND operator), 65
- <- (arrow) notation, for assigning values, 22, 28
- < (less than) operator, 61, 386
- <= (less than or equal to) operator, 61
- > (greater than) operator, 61, 386
- >= (greater than or equal to) operator, 61
- \* (asterisk)
  - as cross-factor notation, 517, 520, 524
  - as multiplication operator, 17, 49
- \ (backslash), 74, 76
- \\ escape sequence, for backslash, 76
- \" escape sequence, for double quotation mark, 76
- \b escape sequence, for backspace, 76
- \n escape sequence, for newline, 76
- \t escape sequence, for tab, 76
- { } (braces), 666
- ^ (exponent operator), 17
- : (colon)
  - for creating numeric sequences, 24
  - for grouping variables, 446, 516, 524
- :: (double colon), for specifying function version, 253
- := operator, 624–625
- \$ (dollar) operator, 92, 256, 461
- ~. notation, 534, 535, 539
- " (double quotation mark), 7, 73
- = (equal sign), 21–22
- == (equal to) operator, 61
- ! (factorial operator), 333
- ! (NOT operator), 65, 107
- != (not equal to) operator, 61
- / (forward slash)
  - for division, 17
  - for file paths, 7
- # (hash mark)
  - for comments, 6
  - for hexadecimal color codes, 632
- ∞ (infinity), 104–106
- (subtraction operator), 17, 49
- (matrix multiplication), 488
- ≠ (not equal sign), 386, 459
- () (parentheses), 17, 19, 601
- %>% (pipe) operator, 624
- %\*% (matrix product operator), 50
- | (element-wise OR operator), 65, 452
- || (single comparison OR operator), 65
- + (addition operator), 17
- ?? command, 10
- ; (semicolon), 666
- [] (square brackets). *See* square brackets ([])
- ~ (tilde), 299, 445, 490
- 2D circle, 721–723
- 3D cone, 725
- 3D cylinder, 723–725
- 3D plots. *See* 3D scatterplots; interactive 3D plots
- 3D scatterplots, 649–653
  - basic syntax, 649–650
  - visual enhancements, 650–653

## A

- $\alpha$  (significance level), 387
- abline function, 134, 137, 456, 525
- add1 function, 533, 571
- add\_axis function, 626, 628, 629, 630
- addition, 17, 18
  - of matrices, 49–50
- additive effect, 468
- add\_legend function, 626, 628, 629, 630
- add.smooth argument, 552
- adjust argument, 628
- adjustcolor command, 612–613, 643–645, 665
- adjusted measure, 460
- Adjusted R-squared (coefficient of determination), 460, 548
- advanced plot customization.  
*See* customizing plots; plotting
- aesthetic mapping, with geoms, 143–146
- aggregate function, 444, 446–447, 450
- AIC (Akaike’s Information Criterion), 541–548
- AIC function, 542
- airquality data frame, 614, 615, 617, 653, 676, 698
- Akaike’s Information Criterion (AIC), 541–548
- algorithms, model selection, 529–548
  - backward selection, 537–541
  - forward selection, 533–537
  - nested comparisons, 529–532
  - stepwise AIC selection, 541–548
- all function, 63, 64
- alpha (significance level), 426
- alpha argument, 643, 644, 701, 714
- alpha.f argument, 613, 643
- alternative argument, 391, 405
- alternative hypothesis, 386
- American Standard Code for Information Interchange (ASCII), 160
- analysis of variance. *See* ANOVA
- AND operator, 65
- Anderson-Darling test, 438
- angle argument, 370
- anonymous functions, 236
- anorexia data set, 401
- ANOVA (analysis of variance), 435–450
  - Kruskal-Wallis test, 447–450
  - one-way, 435–442
    - ANOVA table construction, 439–440
    - building ANOVA tables with aov function, 440–442
    - equivalence with, 481–483
    - hypotheses and diagnostic checking, 436–439
  - two-way, 443–447
    - main effects and interactions, 444–447
    - suite of hypotheses, 443–444
- anova function, 531, 571
- any value, 63, 64
- aov function, 440–442, 481
- appearance constants, setting with geoms, 141–143
- apply function, 204–209, 236, 723
- args.legend argument, 292
- arguments, 8, 222–232
  - defaults, setting, 225–227
  - ellipses and, 176–177, 228–233
  - lazy evaluation, 222–225
  - matching, 172–177
    - ellipses and, 176–177
    - exact, 172–173
    - mixed, 175–176
    - partial, 173–174
    - positional, 174–175
  - missing, checking for, 227–228
- Arguments section, of help files, 10
- arithmetic, 18–19. *See also* math
- arithmetic operators, displaying in plots, 601
- array function, 53, 58
- arrays, multidimensional, 52–58
- arr.ind argument, 71
- arrows function, 134, 138, 353, 585

ASCII (American Standard Code for Information Interchange), 160

as.dot functions, 121, 248

as.matrix function, 124

as.numeric function, 278, 438, 462, 469, 638

asp argument, 705

aspect argument, 695, 705

aspect ratio, of bivariate surfaces, 704–708

assigning objects, 21–22

as.vector function, 123–124

asymmetric distribution, 326

attach function, 256–257

attributes, 114–116

attributes function, 114–115, 117

automatic plot types, 129–130

average squared distance, 277

axes, 577

- customizing, 594–596
- labeling, 130–131, 299
- spacing, 595
- tick marks, 606

axis function, 594–595, 604

## B

\b (backspace) escape sequence, 76

$\beta$  (Type II error), 424, 426

$\beta_0$  (intercept), 454

$\beta_1$  (slope), 454

backslash (\) symbol, 74, 76

backward selection, 537–541

barplot function, 290–291, 316

barplots, 289–293

base distribution, 739

base packages, 253, 739

base R graphics, 139

baseline hypothesis, 386, 421

bases, of logarithms, 19

Bayesian Information Criterion (BIC), 548

Bayesian probability, 310

Bernoulli distribution, 332–334

between-group variability, 440

bg3d function, 693

bgplot3 function, 696, 706

bias, 377

BIC (Bayesian Information Criterion), 548

bimodal distribution, 326

binary variables, 332, 468–472

- linear regression model of, 470–471
- predictions from, 471–472

binding, matrices, 41–42

binomial distribution, 333–337

- dbinom function, 335–336
- pbinom function, 336
- qbinom function, 337
- rbinom function, 337–338

binomial random variable, 339

bivariate surfaces, 699–708

- adding points, 701
- adding surfaces, 701–703
- basic perspective surface, 700
- coloring by  $z$  value, 703–704
- dealing with aspect ratio, 704–708

body code, of a function, 216

Bonferroni correction, 423

boot package, 500, 523, 687, 740

border argument, 682

Boston data frame, 733

box function, 583

box-and-whisker plots. *See* boxplots

boxes, customizing, 593–594

boxplot function, 298–299, 606

boxplots, 298–300, 469–470

- side-by-side, 299–300, 437, 448–449
- stand-alone, 298–299

braced area, of a function, 216

braces, for enclosing commands, 666

breaks argument, 85, 295

bty parameter, 593

built-in data sets, 148–149

built-in palettes, 635–636

byrow argument, 40

## C

c function, 58, 83

calculations, controlling order of, 17

- calling functions, 165–177
  - argument matching, 172–177
  - ellipses and, 176–177
  - exact, 172–173
  - mixed, 175–176
  - partial, 173–174
  - positional, 174–175
- scoping, 165–171
  - environments, 166–168
  - reserved and protected names, 170–172
  - search path, 168–170
- car package, 254, 628, 646, 741
- Cars93 data frame, 450
- case-sensitivity, 7
- cat command, 74–75, 219, 244
- categorical predictors, 468–483
  - binary variables, 468–472
  - changing reference level, 477–478
  - equivalence with one-way ANOVA, 481–483
- interactions
  - between two, 519–521
  - with continuous predictors, 515–519
  - higher-order, 523–526
- multilevel variables, 472–477
  - dummy coding, 472–474
  - linear regression model of, 474–476
- treating categorical variables as numeric, 478–480
- categorical variables, 262–263, 490, 515
  - facets mapped to, 619–623
  - testing, 410–420
    - single categorical variable, 410–415
    - two categorical variables, 415–420
- categories
  - exhaustive, 410
  - mutually exclusive, 410
- categorization, indexing
  - continuums using color palettes via, 637–639
- cbind function, 41–42, 98
- central limit theorem (CLT), 369, 401
- centrality, 267–270
- cex (character expansion)
  - parameter, 129, 693, 707
- cex.lab argument, 599
- char argument, 250
- characters, 72–79
  - concatenation of, 74–76
  - escape sequences, 76–77
  - expansion, 129
  - matching, 77–79
  - strings of, 73–74
  - substrings and matching, 77–79
- $\chi$  (chi) symbol, 410
- ChickWeight data set, 148
- chickwts data frame, 263, 269, 279
- chisq.test function, 414–415, 419
- chi-squared test
  - of distribution, 411–414
  - of independence, 416–419
- chorley data set, 678
- circles, 2D, 721–723
- CIs. *See* confidence intervals (CIs)
- citation command, 748
- citing R packages, 748–749
- class function, 117, 118, 119
- classes. *See* object class
- classical probability, 310
- clmfires data set, 707
- close function, 249–250
- CLT (central limit theorem), 369, 401
- cm.colors function, 635
- coef argument, 518
- coef function, 457
- coefficient of determination, 458, 460
- coefficients component, 457
- coercion, 120–126
- col (color) parameter, 129, 605, 682, 693
- col2rgb function, 632–634
- collective fashion, 477
- collinearity, 549, 569–572
  - correlated predictors, example of, 569–572
  - potential warning signs, 569

- color, 631–648. *See also* color palettes
  - coloring facets, 682–686
  - controlling multiple levels in
    - isosurfaces with, 714–715
  - including color legend, 641–643
  - opacity, 643–645
  - on plots, 131–132
  - RGB alternatives, 645–648
  - RGB hexadecimal color codes, 632–635
- color argument, 715
- color palettes
  - built-in palettes, 635–636
  - custom palettes, 636–637
  - using to index continuum, 637–641
    - via categorization, 637–639
    - via normalization, 639–641
- colorlegend function, 641–643, 644, 647, 648, 651, 669, 675, 690
- color.palette argument, 663
- colorRamp function, 639, 640, 646, 690
- colorRampPalette function, 636–637, 639, 645–646, 664
- colors function, 131, 632
- cols argument, 633
- colSums function, 417
- columns
  - binding together matrices as, 41–42
  - extracting from matrices, 43–44
- combination plots, 333, 563
- commands
  - built-in, 7
  - scrolling through, 5
- commas, in code, 13
- comma-separated values (.csv), 150
- comments, 6
- compiler package, 739
- complement of an event, 312–313
- completion time, 250–252
- complexity vs. goodness-of-fit, 527–528
  - general guidelines, 528–529
  - principle of parsimony, 528
- Comprehensive R Archive Network. *See* CRAN (Comprehensive R Archive Network)
- compute.cont argument, 717
- concatenation, of characters, 74–76
- conditional probability, 311
- conditions. *See* if statements
- cones, 3D, 725
- confidence bands, 507
- confidence intervals (CIs), 378–384, 461–462
  - interpretation of, 382–384
  - for mean, 378–381
  - for proportion, 381–382
- confint function, 460, 495
- conf.level argument, 405, 410
- confounding, 486
- console pane, 5–6
- continuous predictors, 493, 515–519, 521–523
- continuous random variables, 318–326
  - cumulative probability distributions of, 323–324
  - mean and variance of, 326–329
- continuums, indexing using color palettes, 637–641
  - via categorization, 637–639
  - via normalization, 639–641
- contour function, 657–658, 661–663
- contour plots, 657–668
  - color-filled contours, 663–668
  - drawing contour lines, 657–663
  - nonparametric bivariate density estimate example, 660–663
  - parametric response surface, 659–660
  - topographical map example, 657–658
- contour3d function, 713, 719
- contributed data sets, 149–150
- control flow mechanisms, 209–214. *See also* loops
  - declaring break or next, 209–211
  - repeat statement, 211–214
- cook.levels argument, 565

- Cook's distance
    - graphically combining
      - with residuals and leverage, 563–568
    - overview, 559–563
  - `cooks.distance` function, 561
  - `coord_flip` function, 292–293
  - `cor` function, 281–285, 453
  - correct argument, 405
  - correlation, 280–285
  - `cos` function, 722
  - cosine function, 720
  - counts, 271–273
  - `cov` function, 281–284
  - covariance, 280–285
  - covariance matrices, 732
  - CRAN (Comprehensive R Archive Network)
    - finding packages on, 742
    - mirror link, 737
    - obtaining and installing R from, 3
    - switching CRAN mirror, 747–748
  - cross-factor notation (\*), 517, 520, 524
  - .csv* (comma-separated values), 150
  - `cumsum` function, 316
  - custom color palettes, 636–637
  - customizing plots, 575–608
    - axes, 594–596
    - boxes, 593–594
    - default styles, suppressing, 592–593
    - fonts, 597–598
    - graphics devices, handling, 576–582
      - closing, 578
      - manually opening new, 576–608
      - multiple plots in one, 578–582
      - switching between, 577–578
    - Greek symbols, 598–599
    - mathematical expressions, displaying, 599–601
  - point-and-click coordinate interaction, 586–591
    - ad hoc annotation, 588–591
    - retrieving coordinates silently, 586–587
    - visualizing selected coordinates, 587–588
  - regions and margins, 582–586
    - clipping, 584–586
    - custom spacing, 583–584
    - default spacing, 582–583
  - scatterplot, fully annotated, 601–608
  - `cut` function, 638, 705
  - cylinders, 3D, 723–725
- D**
- .dat* file format, 155
  - data frames, 95–102
    - adding data columns, 98–100
    - combining, 98–100
    - creating, 96–98
    - logical record subsets, 100–102
  - `data` function, 147
  - data sets, 147–150, 156–157
    - built-in, 148–149
    - contributed, 149–150
  - data visualization, 289–308
    - barplots, 289–293
    - box-and-whisker plots, 298–300
      - side-by-side boxplots, 299–300
      - stand-alone boxplots, 298–299
    - histograms, 294–298
    - pie charts, 293–294
    - scatterplots, 300–308
      - matrix of plots, 303–308
      - single plot, 301–302
  - `data.frame` class, 120
  - `data.frame` function, 96, 97
  - `datasets` package, 148, 263, 739
  - `dbeta` function, 363
  - `dbinom` function, 334, 335–336
  - `dchisq` function, 362, 412
  - debugging, 244

- decimal places (d.p.), 13
- decreasing argument, 26, 27, 59, 60
- degrees of freedom
  - for chi-squared distribution, 362
  - for *F*-distribution, 363
  - for *t*-distribution, 357–358
- deleting
  - elements from matrices, 44–46
  - elements from vectors, 29
- delimiter, in table-format files, 150
- demo(plotmath) demonstration, 601
- density command, 614
- density functions, 342–363
  - exponential, 359–362
    - dexp function, 359–360
    - pexp function, 360–361
    - qexp function, 361–362
  - normal, 348–357
    - dnorm function, 350
    - example, 356–357
    - pnorm function, 350–353
    - qnorm function, 353–354
    - rnorm function, 355–356
  - Student's *t*-distribution, 357–359
  - uniform, 343–347
    - dunif function, 344–346
    - punif function, 346
    - qunif function, 346–347
    - runif function, 347
- dependencies argument, 743
- Description section, of help files, 10
- detach function, 255, 257, 740
- Details section, of help files, 10
- dev.new function, 576, 579–580
- dev.off function, 157, 578
- dev.set function, 577, 578
- dexp function, 359–360
- df argument, 362, 412
- df function, 363
- df1 argument, 363
- df2 argument, 363
- d-function, 342, 344
- dgamma function, 363
- dgeom function, 342
- dget command, 160, 161
- dhyper function, 342
- diabetes data frame, 516, 531, 620–621, 652
- diag command, 46, 48
- diagnostic checks, 438, 550
- diagonal element extraction, 558
- dichotomous variables, 332
- digit argument, 643
- digits argument, 272
- dim function, 42, 49, 97, 114, 115, 657
- dimensions, of matrices, 42
- dimnames argument, 115–116, 413
- direct-access function, 457
- discrete random variables, 315–318
  - cumulative probability distributions of, 315–317
    - mean and variance of, 317–318
- discrete numeric variables, 262
- disposable functions, 236
- distributions, chi-squared test
  - of, 411–414. *See also* probability distributions; sampling distributions
- division, 17, 18
- dmultinom function, 342
- dmvnorm function, 711, 713
- dnbinom function, 342
- dnorm function, 350, 648, 711
- dodged barplot, 290
- dollar (\$) operator, 92, 256, 461
- double colon (::), for specifying
  - function version, 253
- double quotation marks ("), 7, 73
  - escape sequence (\"), 76
- download from source prompt, 744
- downloading R, 737–739
- d.p. (decimal places), 13
- dpois function, 340–341
- dput command, 160, 161
- drop1 function, 537, 571
- dt function, 357
- dummy functions, 220
- Duncan data set, 420
- Dunedin temperatures example, 369–373
- dunif function, 344–346

## E

- e* (Euler's number), 19
- each argument, 25–26, 417
- earthquake data example, 660–663
- editor, in RStudio, 752–753
- editor pane, 5–6
- element-wise checks, 184–186
- ellipses, arguments and, 176–177, 228–233
- else statements, 183–184
- empty parentheses, 216
- empty pixels, 671–679
- end argument, 635, 672
- e-notation, 20–21
- environments, 166–168
  - global, 166
  - local, 167–168
  - package environments and namespaces, 166–167
- equal sign (=), 21–22
- errors. *See also* exception handling
  - overview, 420–421
  - Type I errors, 421–423
    - Bonferroni correction, 423
    - simulating, 421–423
  - Type II errors, 424–428
    - other influences on error rate, 426–428
    - simulating, 425–426
- escape sequences, 76–77
- Euclidian space, 720
- Euler's number (*e*), 19
- evaluation grid, constructing, 654–655
- events, probability and, 310
  - complement of an event, 312–313
  - intersection of two events, 311–312
  - union of two events, 312
- exact argument matching, 172–173
- Examples section, of help files, 10
- Excel, file format for, 153
- exception handling, 241–248
  - errors and warnings, 242–244
  - try statements, 244–248
  - suppressing warning messages, 246–248

- using in body of function, 245–246
- exp function, 19
- expand argument, 682, 688, 705
- expand.grid function, 654–655, 659, 675, 678, 706, 709, 724
- explanatory variables, 451, 453, 485, 490, 528, 589
- explicit attributes, 114, 115
- exponential distribution, 359–362
  - dexp function, 359–360
  - vs. exponential function, 361
  - pexp function, 360–361
  - qexp function, 361–362
- exponential function, 19–20
- exponents, 17
- expression function, 598–599
- externally defined helper functions, 234–235
- extractAIC function, 542
- extraction
  - of elements in vectors
    - using indexes, 28–32
    - using logicals, 68–72
  - from matrices, 43–44
- extraneous variables, 486
- extrapolation, 466

## F

- F (abbreviation for FALSE), 60
- F* (cumulative distribution function), 323
- facet\_grid command, 619–620
- facets
  - coloring, 682–686
  - multiple plots using, 616–623
    - facets mapped to categorical variable, 619–623
    - independent plots, 616–618
- facet\_wrap command, 619–620
- factor function, 80
- factor variable, 435
- factors, 79–87
  - combining and cutting, 83–86
  - defining and ordering levels, 82–83
  - identifying categories, 79–81



- faithful data frame, 667
- FALSE value, 27, 60
- family parameter, 597
- faraway package, 513, 516, 526, 531, 620–621, 652, 741
- F*-distribution, 363
- Fibonacci sequence, 216–218
- figure margins, 582
- figure margins too large error, 584
- figure region, of a plot, 582
- file formats, for plots, 157
- file paths, 7
- file.choose command, 152, 156
- filled.contour function, 663, 666, 690
- fitted function, 457, 560
- fitted models, 518, 529
- fitted values, 458, 462, 551
- fitted.values component, 457
- fitting linear models, 454
- five-number summary, 274–275
- flag vectors, logical, 68
- floating-point numbers, 117
- floor operation, 394
- font parameter, 597
- fonts, 597–598
- for loops, 193–200
  - looping via index or value, 194–197
  - nesting, 197–200
- forecasting, 498
- foreign package, 156, 740
- formula argument, 535
- forward selection, 533–537
- forward slashes (/), 7
- frac operator, 601
- frequentist probability, 310
- from argument, 24
- F*-test, partial, 529–532
- FUN argument, 270, 272, 444
- function command, 215–222, 236
  - creating functions, 218–219
  - return statement and, 220–222
- functions, documentation for, 8–10.
  - See also* calling functions;
  - writing functions

## G

- geom argument, 296
- geom\_bar function, 292–293
- geom\_density function, 614–615
- geometric distribution, 342
- geometric mean, 238
- geom\_histogram function, 297, 611
- geom\_hline function, 145
- geom\_line function, 142, 145
- geom\_point function, 142, 145
- geoms (geometric modifiers)
  - aesthetic mapping with, 143–146
  - setting appearance constants with, 141–143
- geom\_segment function, 145
- geom\_smooth function, 612, 613, 621, 630
- geom\_vline function, 297, 610
- getOption function, 747
- getwd function, 7
- GGally package, 304, 741
- ggpairs function, 304–306
- ggplot function, 609–611
- ggplot2 package
  - aesthetic mapping with geoms, 143–146
  - multiple plots and variable-mapped facets, 616–623
    - facets mapped to categorical variable, 619–623
    - independent plots, 616–618
  - qplot function vs. ggplot function, 609–611
  - setting appearance constants with geoms, 141–143
  - smoothing and shading, 611–615
    - adding LOESS trends, 611–614
    - constructing smooth density estimates, with KDE, 614–615
- ggsave function, 159
- ggtitle function, 610, 615
- ggvis package, 623–630, 741

- global environment, 166
- `.GlobalEnv`, 253
- goodness-of-fit vs. complexity, 527–528
  - general guidelines, 528–529
  - principle of parsimony, 528
- grammar of graphics, 139
- graphical parameters, 129–134
- graphical user interface. *See*
  - GUI (graphical user interface)
- graphics devices, 576–582
  - closing, 578
  - manually opening new, 576–608
  - multiple plots in one, 578–582
    - defining particular layout, 580–582
    - setting `mfrow` parameter, 579–580
  - switching between, 577–578
- graphics files, 157–159
- graphics package, 739
- graphs. *See* plotting
- gray palette, 635, 639, 677
- `gray.colors` palette, 635, 636, 677, 690
- `grDevices` package, 739
- greater than operator (`>`), 61, 386
- greater than or equal to operator (`>=`), 61
- greater-than statement, 386
- Greek symbols, 598–599
- `grep` command, 78
- `grid` function, 605
- `grid` package, 739
- `grid3d` function, 694–696
- `grid.arrange` function, 616–618
- `gridExtra` package, 616, 617, 741
- `gridsize` argument, 717
- `group_by` function, 628
- `gsub` function, 78
- GUI (graphical user interface)
  - installing packages with, 744–745
  - overview, 4–6

## H

- `HairEyeColor` data set, 419
- hash mark (`#`)
  - for comments, 6
  - for hexadecimal color codes, 632
- `hatvalues` function, 558, 561, 571
- HCL (hue-chroma-luminance), 645
- `hcl` function, 645
- header, in table-format files, 150
- `heat.colors` palette, 635, 638, 678, 689, 708
- height argument, 579, 603
- help files, 8–10
- helper functions, 233–236
  - externally defined, 234–235
  - internally defined, 235–236
- heteroscedasticity, 550, 551
- hexadecimal color codes, 632
- higher-dimensional probability densities, 710–711
- higher-order interactions, 523–526
- high-level programming language, R as, 3
- `hist` function, 294–297
- histograms, 294–298, 577
- homoscedasticity, 550
- HSV (hue-saturation-value), 645
- `hsv` function, 645
- hue-chroma-luminance (HCL), 645
- hue-saturation-value (HSV), 645
- hypergeometric distribution, 342
- hypothesis testing, 385–433
  - categorical variables, 410–420
    - single categorical variable, 410–415
    - two categorical variables, 415–420
  - components of, 385–388
    - hypotheses, 386
    - p*-value, 387
    - significance level, 387–388
    - test statistic, 387
- criticisms of, 388
- errors, 420–421
  - Type I errors, 421–423
  - Type II errors, 424–428

- means, 388–402
    - single mean, 389–392
    - two means, 392–402
  - proportions, 402–410
    - single proportion, 402–405
    - two proportions, 405–410
  - statistical power, 428–433
    - power curves, 431–433
    - simulating power, 429–431
- I**
- I function, 505
  - ice.river data set, 148
  - identity matrix, 48
  - IDEs (integrated development environments), 751
  - id.n argument, 552
  - if statements, 179–193
    - else statements and, 183–184
    - nesting and stacking, 186–189
    - stand-alone, 180–183
    - switch function and, 189–193
  - ifelse function, using for element-wise checks, 184–186
  - image function, 668, 672, 673, 682–683
  - implicit attributes, 114
  - implicit looping, with apply function, 204–209
  - include.lowest argument, 85
  - independence, chi-squared test of, 416–419
  - INDEX argument, 270, 444
  - indexes
    - of lists, 90–91
    - looping via, 194–197
    - of vector elements, 28–32
    - vectors of, 30
  - Inf function, 104, 107
  - infinity ( $\infty$ ), 104–106
  - influence, 555–557
  - inheritance, 116
  - input\_ functions, 624, 627
  - input\_checkbox function, 626
  - input\_numeric function, 626
  - input\_radiobuttons function, 626
  - input\_select function, 626
  - input\_slider function, 624–625, 627
  - InsectSprays data set, 273, 306
  - inside.owin function, 674, 675, 678, 690
  - installing R, 737–750
    - from CRAN, 3
    - downloading, 737–739
    - packages, 739–746
      - base packages, 739
      - contributed packages, 740–746
      - finding on CRAN, 742
      - installing at prompt, 742–744
      - installing from GUI, 744–745
      - installing using local files, 745–755
      - recommended packages, 740
      - updating, 746–747
    - using other mirrors and repositories, 747–748
    - other package repositories, 748
    - switching CRAN mirror, 747–748
  - install.packages command, 8, 742, 756
  - integers, 117–118
  - integrated development environments (IDEs), 751
  - intensities (of RGB colors), 632
  - interaction.plot function, 446
  - interactions with plot coordinates,
    - point-and-click, 586–591
      - ad hoc annotation, 588–591
      - retrieving coordinates silently, 586–587
      - visualizing selected coordinates, 587–588
  - interactive 3D plots, 691–735
    - bivariate surfaces, 699–708
      - adding points, 701
      - adding surfaces, 701–703
      - basic perspective surface, 700
      - coloring by z value, 703–704
      - dealing with aspect ratio, 704–708

- interactive 3D plots, *continued*
    - parametric equations, 720–735
      - mathematical abstractions, 725–735
      - simple loci, 720–725
    - point clouds, 691–699
      - adding further 3D
        - components, 694–699
      - basic 3D cloud, 692–693
      - visual enhancements and legends, 693–694
    - trivariate surfaces, 709–719
      - evaluation coordinates in 3D, 709–710
      - isosurfaces, 710–715
        - nonparametric trivariate density example, 715–720
  - interactive effects, 443, 514–515
    - between categorical and continuous predictors, 515–519
    - between two categorical variables, 519–521
    - between two continuous predictors, 521–523
    - higher-order, 523–526
  - interactive.arrow function, 590, 608
  - intercept parameter, estimating, 454–455
  - intercept-only model, 542
  - internally defined helper functions, 235–236
  - interpolation, 480, 498
  - interquartile range (IQR), 277, 287
  - intersection of two events, 311–312
  - interval argument, 463
  - inversion, of matrices, 51–52
  - involutory function, 603
  - inward-facing tick marks, 596
  - IQR (interquartile range), 277, 287
  - IQR function, 278–279
  - iris data frame, 649–650, 692
  - is-dot functions, 119–120
  - is.factor(survey\$Smoke) function, 474
  - is.finite function, 105
  - is.infinite function, 105
  - is.integer function, 125
  - is.matrix function, 223
  - is.na function, 109
  - is.nan function, 107
  - is.null function, 111
  - is.numeric function, 125
  - isosurfaces, 710–715
    - basic one-level isosurface, 712–714
    - controlling multiple levels
      - with color and opacity, 714–715
    - higher-dimensional probability densities, 710–711
  - italic function, 600–601
  - italicized alphabetic variables, 601
  - iterative looping, 237
- J**
- jitter function, 276
  - jittering, 276
  - .jpeg files, 157–158
- K**
- k* categories, 410, 637
  - KDE (kernel density estimation), 614, 660
  - kde function, 717, 719
  - kde2d function, 660, 671, 674, 678, 690
  - kernel density estimation (KDE), 614, 660
  - KernSmooth package, 740
  - keystroke shortcuts, 6
  - key.title argument, 664
  - knitr package, 757–758
  - kruskal.test function, 448
  - Kruskal-Wallis test, 447–450
  - ks package, 741
- L**
- label notation, specialized, 597–601
    - font, 597–598
    - Greek symbols, 598–599
    - mathematical expressions, 599–601

- labels argument, 85, 557
- labs function, 297
- lambda parameter, 341
- lapply function, 238
- las parameter, 595–596
- lattice package, 740
- layer\_densities function, 628
- layer\_histograms function, 624, 628, 630
- layer\_points function, 626
- layers, 53, 139
- layer\_smooths function, 627, 628, 630
- layout function, 580, 664
- layout.show function, 581
- lazy evaluation, 222–225
- least-squares regression, 455, 456
- left argument, 645
- legend function, 134, 138, 232, 605
- legend3d function, 693
- legends, on interactive 3D plots, 693–694
- legend.text argument, 292
- length function, 27, 80
- length.out value, 24, 25
- less than operator (<), 61, 386
- less than or equal to operator (<=), 61
- less-than statement, 386
- letters object, 655
- levels function, 80, 115, 661
- leverage, 550
  - calculating, 555–558
  - combining with residuals and Cook’s distance, 563–568
  - illustrating, 555–557
- libraries. *See* packages
- library command, 7–8, 148
- line of best fit, 458, 487
- line type (lty) parameter, 129, 133, 605, 659
- line width (lwd) parameter, 129, 133, 659
- linear model selection and diagnostics
  - goodness-of-fit vs. complexity, 527–528
  - general guidelines, 528–529
  - principle of parsimony, 528
- linear regression. *See* multiple linear regression; simple linear regression
- lines
  - adding to plots, 134–139
  - contour lines, drawing, 657–663
    - nonparametric bivariate density estimate example, 660–663
    - parametric response surface, 659–660
    - topographical map example, 657–658
  - on plots, 133
- lines function, 134
- lines3d function, 694, 708
- list command, 92, 230
- list slicing, 91
- list.files function, 151
- lists of objects, 89–95
  - definition and component access, 89–91
  - naming, 91–93
  - nesting, 93–95
- lm command, fitting linear models with, 455–456
- load command, 11
- model selection algorithms, 529–548
  - backward selection, 537–541
  - forward selection, 533–537
  - nested comparisons, 529–532
  - stepwise AIC selection, 541–548
- residual diagnostics, 548–568
  - assessing normality, 554–555
  - calculating leverage, 555–558
  - Cook’s distance, 559–563
    - graphically combining residuals, leverage, and Cook’s distance, 563–568
  - illustrating outliers, leverage, and influence, 555–557
  - inspecting and interpreting residuals, 549–554

- loading
  - packages, 7–8
  - workspace image files, 11–12
- local environments, 167–168
- locally weighted scatterplot
  - smoothing (LOESS), 611–614
- locator function, 586–587, 608, 633–634
- loci, 720–725
  - 2D circle, 721–723
  - 3D cone, 725
  - 3D cylinder, 723–725
- LOESS (locally weighted scatterplot
  - smoothing), 611–614
- loess function, 612
- log function, 20
- logarithmic scale, 509
- logarithms, 19–20, 508–512
  - fitting log transformation, 509–511
  - plotting log transformation fit, 511–512
- logical flag vector, 547
- logical operators, 64–67
- logical values, 26–27, 59–72
  - logical operators, 64–67
  - logical subsetting and extraction, 68–72
  - numbers as, 67–68
  - relational operators, 60–64
  - TRUE or FALSE, 60
- log-likelihood, 542
- loops, 193–209
  - for loops, 193–200
    - looping via index or value, 194–197
    - nesting, 197–200
  - implicit looping with `apply` function, 204–209
  - rotating with, 686–690
  - while loops, 200–204
- lower quartile, 274
- lower-tail probability, 351
- lower-tailed test, 386, 394
- ls command, 22

- lty (line type) parameter, 129, 133, 605, 659
- lty.hpplot argument, 650
- lurking variable, 486
- lwd (line width) parameter, 129, 133, 659

## M

- main (plot title) parameter, 129
- Mann-Whitney U test, 401
- mar parameter, 582, 594
- margins, plotting, 582–586
  - clipping, 584–586
  - custom spacing, 583–584
  - default spacing, 582–583
- markup tools, 601
- masking, 252–258
  - data frame variable distinction, 255–258
  - function and object distinction, 252–255
- mass functions, 332–342, 344
  - Bernoulli distribution, 332–334
  - binomial distribution, 333–337
    - `dbinom` function, 335–336
    - `pbinom` function, 336
    - `qbinom` function, 337
    - `rbinom` function, 337–338
  - Poisson distribution, 338–342
    - `dpois` function, 340–341
    - `ppois` function, 340–341
    - `qpois` function, 341
    - `rpois` function, 341–342
- MASS package, 254–255, 401, 450, 474, 611, 622, 629, 637, 660, 686, 697, 740
- matching, of characters in strings, 77–79
- math, 17–21, 599–601. *See also*
  - matrices; vectors
  - arithmetic, 18–19
  - e-notation, 20–21
  - exponentials, 19–20
  - logarithms, 19–20
- mathematical abstractions, 725–735
  - Möbius strip, 726–729
  - torus, 729–735

- matrices, 303–308
  - addition of, 49–50
  - algebra and, 47–52
  - binding together, 41–42
  - defining, 39–42
  - deleting elements from, 44–47
  - dimensions of, 42
  - extracting elements from, 43–44
  - filling direction, 40–41
  - identity matrix, 48
  - inversion of, 51–52
  - multiplication of, 50–51
  - omitting and overwriting
    - elements from, 44–47
  - operations and algebra, 47–52
  - scalar multiple of, 49
  - subsetting elements from, 42–47
  - subtraction of, 49–50
  - transpose of, 47–48
- matrix command, 40, 58
- max function, 268–269
- maxColorValue argument, 633, 643
- mean
  - of continuous random variable, 326–329
  - of discrete random variable, 317–318
- mean function, 9, 268–270, 453
- mean square (MS), 439, 440
- mean squared error (MSE)
  - effect, 440
- mean squared group (MSG)
  - effect, 440
- means, hypothesis tests of, 388–402
  - single mean, 389–392
  - two means, 392–402
    - paired/dependent samples, 398–402
    - pooled variance, 396–398
    - unpooled variances, 393–395
- median function, 268–269
- member references, 90
- mfrow parameter, 579–580, 581, 616, 666, 688
- mgp parameter, 595–596
- Microsoft Office Excel, file format for, 153
- min function, 268
- mirrors, installing packages using, 747–748
- misc3d package, 713, 718, 741
- missing function, 227
- missing value, in table-format files, 150
- Möbius strip, 726–729
- modality, 326
- model diagnostics, 548
- model selection, 492, 525, 528
- model-based inference, 458
- MS (mean square), 439, 440
- MSE (mean squared error)
  - effect, 440
- MSG (mean squared group)
  - effect, 440
- mtcars data frame, 287, 290, 478, 503, 514, 521, 526, 653, 664, 669, 702
- mtext function, 584
- multicollinearity. *See* collinearity
- multidimensional arrays, 52–58
  - definition, 53–55
  - subsets, extractions, and replacements, 55–58
- multilevel variables, 472–477
  - dummy coding, 472–474
  - linear regression model of, 474–476
  - predictions from, 476–477
- multinomial distribution, 342
- multiple linear regression, 485–526
  - implementing in R and interpreting, 490–501
    - additional predictors, 490–493
    - finding confidence intervals, 495
    - interpreting marginal effects, 493–494
    - omnibus *F*-test, 496–498
    - predicting from multiple linear model, 498–501
    - visualizing multiple linear model, 494–495

- multiple linear regression, *continued*
    - interactive terms, 514–526
      - concept and motivation, 514–515
      - higher-order interactions, 523–526
      - one categorical, one continuous, 515–519
      - two categorical, 519–521
      - two continuous, 521–523
    - terminology, 486
    - theory, 486–489
      - basic example, 488–489
      - estimating in matrix form, 487–488
      - extending simple model to multiple model, 487
    - transforming numeric variables, 501–514
      - logarithmic, 508–512
      - polynomial, 502–508
  - multiple plots, using facets, 616–623
    - facets mapped to categorical variable, 619–623
    - independent plots, 616–618
  - Multiple R-squared (coefficient of determination), 460
  - multiple testing problem, 423
  - multiple-factor ANOVA, 443
  - multiplication of matrices, 50–51
  - multivariate data, 264–265
  - multivariate model, 494
  - mvtnorm package, 710, 741
- N**
- \n (newline) escape sequence, 76
  - NA (Not Available) value, 108–110
  - names function, 93, 115
  - namespaces, 166–167
  - naming lists of objects, 91–93
  - NaN (Not a Number) value, 106–108
  - na.omit function, 110, 638, 653, 676, 697
  - na.rm argument, 270, 469, 640, 684
  - natural log, 19
  - ncol argument, 40, 168, 620
  - ncol function, 42, 97
  - negative binomial distribution, 342
  - negative coefficient
    - of categorical predictors, 468–476
    - of continuous predictors, 454, 522
  - negative power, 20
  - negative skew, 326
  - nesting
    - if statements, 186–189
    - lists of objects, 93–95
    - for loops, 197–200
  - newdata argument, 463, 471, 511, 659
  - newline escape sequence (\n), 76
  - news function, 746
  - no-change hypothesis, 386, 421
  - nominal variables, 262. *See also*
    - categorical variables
  - non-numeric values. *See* characters; factors; logical values
  - nonparametric bivariate density
    - estimate example, 660–663
  - nonparametric smoothing, 611
  - nonparametric trivariate density
    - example, 715–720
    - calculating 3D estimate, 717
    - isosurface level selection, 717–720
  - normal distribution, 348–357
    - dnorm function, 350
    - example, 356–357
    - pnorm function, 350–353
    - qnorm function, 353–354
    - rnorm function, 355–356
  - normalization, indexing continuums
    - using color palettes via, 639–641
  - normalize function, 640, 644, 646, 651, 677
  - Not a Number (NaN) value, 106–108
  - Not Available (NA) value, 108–110
  - NOT operator, 65
  - Notes section, of help files, 10
  - nrow argument, 40, 620
  - nrow function, 97, 101, 271, 675, 723



nuclear data frame, 500, 533, 546, 666, 687  
 nuisance variables, 486  
 null device, 576, 580  
 null hypothesis, 386, 421  
 null model, 496, 542  
 NULL value, 110–114  
 numbers, as logical values, 67–68  
 numeric variables, 262, 515  
   transforming, 501–514  
     logarithmic, 508–512  
     polynomial, 502–508  
 numerical simulation, 421

## O

object class, 116–119  
   multiple classes, 119  
   other data structures, 118–119  
   stand-alone vectors, 117–118  
 object-oriented programming  
   language, R as, 116  
 objects, assigning, 21–22  
 observed count, 411, 416  
 oma parameter, 582  
 omitting elements  
   from matrices, 44–47  
   from vectors, 28–33, 68–72  
 omnibus  $F$ -test, 482, 496–498  
 one-factor analysis, 435  
 one-sided statement, 386  
 one-to-one transformations, 514  
 one-way ANOVA, 443  
 opacity, 643–645, 714–715  
 OR operator, 65  
 order of operations, 18  
 ordinal variables, 262. *See also*  
   categorical variables  
 orthogonal contrasts, 477  
 outer function, 724  
 outer region, of a plot, 582  
 outliers, 285–288, 555–557  
 overwriting elements  
   from matrices, 44–47  
   from vectors, 28–33, 68–72  
 owin class, 672

## P

package environments, 166–167  
 packages, 7  
   citing, 748–749  
   finding on CRAN, 742  
   installer and updater in  
     RStudio, 755–756  
   installing, 8  
     with graphical user interface  
       (GUI), 744–745  
     at prompt, 742–744  
     using local files, 745–755  
   loading, 7–8  
   masking, 254–255  
   updating, 8  
   writing, 749–750  
 padding space, on a plot, 592  
 pairs function, 303  
 palette function, 632  
 palettes. *See* color palettes  
 panes, in RStudio, customizing,  
   753–754  
 par function, 579–580, 666, 688  
 parameters, 265–266  
 parametric equations, 720–735  
   mathematical abstractions,  
     725–735  
     Möbius strip, 726–729  
     torus, 729–735  
   simple loci, 720–725  
     2D circle, 721–723  
     3D cone, 725  
     3D cylinder, 723–725  
 parametric response surface,  
   659–660  
 parentheses ( $()$ ), 17, 19, 601  
 parsimony, 528, 548  
 partial argument matching,  
   173–174  
 partial  $F$ -tests, 529–531  
 paste function, 74–75, 601  
 pattern matching, 77–79  
 pbeta function, 363  
 pbinom function, 334, 336  
 pch (point character) parameter,  
   129, 133, 647, 651, 707

- pchisq function, 362, 414
- pcol function, 633, 634
- .pdf files, 157–158
- PEMDAS (order of operations), 18
- percentages, 271–273
- percentiles, 274–275
- persp function, 680, 682–683, 686
- persp3d function, 700–706, 724–731
- perspective plots, 679–690
  - basic plots and angle
    - adjustment, 679–682
    - coloring facets, 682–686
    - rotating with loops, 686–690
- perspplot function, 686–687, 690
- pexp function, 360–361
- pf function, 363
- p-function, 331, 337, 342, 346, 361
- pgamma function, 363
- pgeom function, 342
- phyper function, 342
- PI (prediction interval), 462
- $\pi$  (pi) symbol, 373, 403
- pi object, 722
- pie charts, 293–294
- pie function, 293
- pipe operator (%%), 624
- pixel images, 668–679
  - one grid point = one pixel, 668–671
  - surface truncation and empty pixels, 671–679
- PlantGrowth data set, 401
- plot function, 128, 129, 134, 230, 469, 551, 571, 610, 642, 662, 672
- plot3d function, 692
- plot.axes argument, 665
- plot.new function, 696
- plotrix package, 643
- plotting, 127–146. *See also*
  - customizing plots;
  - interactive 3D plots;
  - scatterplots
  - 3D scatterplots, 649–653
    - basic syntax, 649–650
    - visual enhancements, 650–653
  - adding points, lines, and text, 134–139
  - automatic plot types, 129–130
  - barplots, 289–293
  - boxplots, 298–300, 469–470
    - side-by-side, 299–300, 437, 448–449
    - stand-alone, 298–299
  - color, 131–132
  - contour plots, 657–668
    - color-filled contours, 663–668
    - drawing contour lines, 657–663
  - graphical parameters, 129–134
  - line and point appearances, 133
  - perspective plots, 679–690
    - basic plots and angle
      - adjustment, 679–682
      - coloring facets, 682–686
      - rotating with loops, 686–690
  - pie charts, 293–294
  - pixel images, 668–679
    - one grid point = one pixel, 668–671
    - surface truncation and empty pixels, 671–679
  - preparing surface for, 653–657
    - conceptualizing z-matrix, 656–657
    - constructing evaluation grid, 654–655
    - constructing z-matrix, 655–656
  - region limits, 133–134
  - title and axis labels, 130–131
  - using plot with coordinate vectors, 127–129
  - writing plots to file, 157–159
- pnbinom function, 342
- .png files, 157, 159
- pnorm function, 350–353, 376
- point character (pch) parameter, 129, 133, 647, 651, 707
- point clouds, 691–699
  - adding further components, 694–699
  - basic 3D cloud, 692–693

- visual enhancements and legends, 693–694
- point-and-click coordinate interaction, 586–591
- ad hoc annotation, 588–591
- retrieving coordinates silently, 586–587
- visualizing selected coordinates, 587–588
- points
  - adding to bivariate surfaces, 701
  - adding to plots, 134–139
    - on plots, 133
  - points function, 134, 635, 661
  - points3d function, 694, 698, 701
  - Poisson distribution, 338–342, 360
    - dpois function, 340–341
    - ppois function, 340–341
    - qpois function, 341
    - rpois function, 341–342
  - polygon function, 322, 345, 352–353, 370
  - polynomials, 502–508
    - fitting polynomial transformation, 503–506
    - pitfalls of, 508
    - plotting polynomial fit, 506–508
  - pooled variance, 396–398
  - positional argument matching, 174–175
  - positive coefficient
    - of categorical predictors, 468–476
    - of continuous predictors, 454, 522
  - positive skew, 295, 326
  - power. *See* statistical power
  - ppois function, 340–341
  - ppp object, 671
  - predict function, 463, 659, 706
  - prediction, 461–468. *See also*
    - categorical predictors
    - confidence interval or prediction interval, 461–462
    - interpolation vs. extrapolation, 466–468
    - interpreting intervals, 462–464
      - confidence intervals for mean heights, 463
      - prediction intervals for individual observations, 463–464
      - plotting intervals, 464–466
    - prediction interval (PI), 462
    - pretty function, 594
    - print command, 244
    - prob argument, 274
    - probability, 309–329. *See also*
      - probability distributions
        - complement of an event, 312–313
        - conditional probability, 311
        - events and, 310
        - intersection of two events, 311–312
      - overview, 309–310
      - random variables, 313–329
        - continuous random variables, 318–326
        - discrete random variables, 315–318
        - realizations, 314
        - union of two events, 312
    - probability densities, higher-dimensional, 710–711
    - probability distributions, 331–363
      - density functions, 342–363
        - exponential, 359–362
        - normal, 348–357
        - Student's *t*-distribution, 357–359
        - uniform, 343–347
      - mass functions, 332–342
        - Bernoulli distribution, 332–334
        - binomial distribution, 333–337
        - other mass functions, 342
        - Poisson distribution, 338–342
    - proc.time function, 251
    - progress bars, 249–250

- prompt
  - customizing, 5
  - installing packages at, 742–744
- proportions, 271–273
  - testing, 402–410
    - single proportion, 402–405
    - two proportions, 405–410
- prop.test function, 405, 407–410
- protected names, 170–172
- pt function, 357, 380
- punif function, 346
- p*-value, 386, 387, 390, 412

## Q

- q function, 12
- qbeta function, 363
- qbinom function, 334, 337
- qchisq function, 362
- qexp function, 361
- qf function, 363
- q-function, 331, 337, 341, 342, 346, 353
- qgamma function, 363
- qgeom function, 342
- qhyper function, 342
- qbinom function, 342
- qnorm function, 353–354, 425
- qplot function, 140–141, 144, 145, 296, 609–611
- qpois function, 341
- QQ (quantile-quantile) plot, 353, 438, 554
- qqline command, 354, 438
- qqnorm function, 354, 438
- qt function, 357–359
- quadratic equation, 232
- quakes data frame, 264, 275, 392, 576, 644, 660
- quantile function, 274, 278, 347, 718
- quantile-quantile (QQ) plot, 353, 438, 554
- quantiles, 274–275
- Quartz window system, 576
- qunif function, 346–347

## R

- .R extension, 5
- R markdown, 758–759
- R programming language. *See also* installing R
  - comments, 6
  - console and editor panes, 5–6
  - help files and function
    - documentation, 8–10
  - obtaining and installing from CRAN, 3
  - project website, 3
  - saving work and exiting, 11–12
  - third-party editors, 11
  - updating, 746–747
  - working directory, 7
- R> prompt, 5, 13
- radians, 720
- rainbow palette, 635, 643, 684, 704
- random variables, 313–329
  - continuous, 318–326
    - cumulative probability distributions of, 323–324
    - mean and variance of, 326–329
  - discrete, 315–318
    - cumulative probability distributions of, 315–317
    - mean and variance of, 317–318
  - realizations, 314
- range function, 268, 673
- raw data, describing, 261–266
  - categorical variables, 262–263
  - numeric variables, 262
  - parameters, 265–266
  - univariate and multivariate data, 264–265
- rbenchmark package, 251
- rbeta function, 363
- rbind function, 41–42, 98, 99, 322
- rbinom function, 334, 337
- rchisq function, 362
- RColorBrewer package, 645
- .RData file, 11, 12
- read.csv function, 154

- reading and writing files, 147–162
  - reading in external data files, 150–156
  - R-ready data sets, 147–150
  - writing out data files and plots, 156–159
- read.table function, 151, 152, 154, 155
- real roots, 233
- records, in data frames, 96
- recursive functions, 237–240
- red-green-blue. *See* RGB (red-green-blue)
- reference level, 468, 473, 477–478
- References section, of help files, 10
- regions, plotting, 582–586
  - clipping, 584–586
  - custom spacing, 583–584
  - default spacing, 582–583
  - limits or regions, 133–134
- regression coefficients, 454
- regular expressions, 72
- relational operators, 60–64
- relative device coordinates, 642
- relevel function, 477
- rep function, 25–26
- repeat statement, 211–214, 218, 219
- repeating values, 25–26
- replacement argument, 78
- repos argument, 743
- repositories, installing packages
  - using, 748
- rescaling variables, 348
- reserved names, 170–172
- resid function, 457
- residual diagnostics, 548–568
  - assessing normality, 554–555
  - calculating leverage, 555–558
  - Cook’s distance, 559–563
  - graphically combining residuals, leverage, and Cook’s distance, 563–568
  - illustrating outliers, leverage, and influence, 555–557
  - inspecting and interpreting residuals, 549–554
- residual standard error, 481, 491
- residuals, illustrating, 456–458
- resolution, 668
- response-predictor data, 563
- return statement, 216, 220–222, 223
- rev function, 613
- rexp function, 361
- rf function, 363
- r-function, 342, 347
- rgamma function, 363
- RGB (red-green-blue)
  - alternatives to, 645–648
  - hexadecimal color codes, 632–635
- rgb function, 633, 634, 640, 643, 644, 709
- rgeom function, 342
- rgl package, 691, 699, 718, 741
- rhyper function, 342
- right skew, 295, 326
- ring torus, 729
- RJDBC package, 156
- rmarkdown package, 758
- rmultinom function, 342
- rmvnorm function, 710–711
- RMySQL package, 156
- rnbinom function, 342
- rnorm function, 355–356
- .Rnw extension, 757
- RODBC package, 156
- rotating, with loops, 686–690
- round function, 272, 335
- rownames function, 702
- rows
  - binding together matrices as, 41–42
  - extracting from matrices, 43–44
- rowSums function, 417
- rpois function, 341
- .Rproj file, 754
- R-squared value, 476, 491
- rstandard function, 571
- RStudio, 11, 751–779
  - auxiliary tools, 754–759
  - markup, document, and graphics tools, 756–759
  - package installer and updater, 755–756
  - projects, 754–755
  - support for debugging, 756

RStudio, *continued*  
  basic layout and usage, 752–754  
  customizing panes, 753–754  
  editor features and  
    appearance options,  
    752–753  
rt function, 357  
runif function, 347

## S

S3 classing structure, 116  
Salaries data frame, 622–623,  
  628, 646  
sample function, 730–731  
sampling distributions, 367–377  
  distribution for sample mean,  
  368–373  
    Dunedin temperatures  
    example, 369–373  
    standard deviation  
    known, 369  
    standard deviation  
    unknown, 369  
  distribution for sample  
  proportion, 373–377  
sapply function, 223  
saturations, 632  
save.image command, 11  
saving  
  scripts, 12  
  workspace image files, 11–12  
scalar multiple, of matrices, 49  
scale\_fill\_grey function, 292–293  
scale\_linetype\_manual function, 297  
scale\_shape\_manual function, 304  
scale\_x\_discrete function, 292–293  
scale\_y\_continuous function,  
  292–293  
scatterplot3d function, 649, 652  
scatterplot3d package, 741  
scatterplots, 300–308  
  matrix of plots, 303–308  
  single plot, 301–302  
scatter.smooth function, 612  
scope argument, 533, 539, 542

scoping, 165–171  
  environments, 166–168  
    global environment, 166  
    local environments, 167–168  
  package environments and  
  namespaces, 166–167  
  reserved and protected names,  
  170–172  
  search path, 168–170  
scripts  
  overview, 5  
  saving, 12  
scrolling through commands, 5  
sd (standard deviation), 277, 278  
search function, 168–169, 252  
search path, 168–170  
segments function, 134, 137, 345  
segments3d function, 694–696, 702  
semicolon, 666  
seq function, 24–26, 169, 350,  
  638, 654  
sequences, creating numeric, 24–26  
setTxtProgressBar function, 249–250  
setwd function, 7, 151  
shade argument, 682, 700  
shading. *See* smoothing and shading  
shape package, 644, 647, 648, 652,  
  653, 669, 705, 741  
shapiro.test function, 554, 571  
Shapiro-Wilk test, 438, 554  
shiny package, 759  
short-form logical operator, 452  
side-by-side boxplots, 299–300  
sigma argument, 350, 422, 560  
significant digits, 643  
silent argument, 244, 246  
simple linear regression, 451–483  
  categorical predictors, 468–483  
    binary variables, 468–472  
    changing reference level,  
    477–478  
    equivalence with one-way  
    ANOVA, 481–483  
  multilevel variables, 472–477  
  treating categorical variables  
  as numeric, 478–480

- example of linear relationship, 451–453
- general concepts, 453–458
  - definition of model, 453–454
  - estimating intercept and slope parameters, 454–455
  - fitting linear models with `lm`, 455–456
  - illustrating residuals, 456–458
- prediction, 461–468
  - confidence interval or prediction interval, 461–462
  - interpolation vs. extrapolation, 466–468
  - interpreting intervals, 462–464
  - plotting intervals, 464–466
  - statistical inference, 458–461
    - coefficient of determination, 460
    - other summary output, 460–461
    - regression coefficient significance tests, 459–460
    - summarizing fitted model, 458–459
- `sin` function, 722
- sine function, 720
- single categorical predictor, 481–482
- singular matrices, 51
- `skip` argument, 155–156
- slicing, of lists, 91
- slope parameter, estimating, 454–455
- smoothing and shading, 611–615
  - adding LOESS trends, 611–614
  - constructing smooth density estimates, 614–615
- `solve` function, 247
- `sort` function, 26–27, 59–60
- sorting, vectors, 26–27
- `spatstat` package, 254, 671, 674, 678, 689, 741
- special values, 103–114
  - infinity (`Inf`), 104–106
  - Not a Number (`NaN`), 106–108
  - Not Available (`NA`), 108–110
  - `NULL`, 110–114
- specialized functions, 233–240
  - disposable functions, 236
  - helper functions, 233–236
    - externally defined, 234–235
    - internally defined, 235–236
  - recursive functions, 237–240
- `split.screen` function, 582
- `spread`, 275–280. *See also* variance
- spreadsheet workbooks, reading, 153–154
- `sqrt` command, 18, 278, 318
- square brackets (`[]`)
  - double, for referencing list members, 90–91
  - for inclusion of values in intervals, 85
  - for indexes in vectors, 28–31, 42–43, 44
  - for list slicing, 91
- square root, 18
- `SS` (sum-of-squares), 439, 440
- stacking, if statements, 186–189
- stand-alone boxplots, 298–299
- stand-alone if statements, 180–183
- standard deviation (`sd`), 277, 278
- `state.abb` object, 306
- statistical power, 428–433
  - power curves, 431–433
  - simulating power, 429–431
- statistics, 261–288
  - describing raw data, 261–266
    - categorical variables, 262–263
    - numeric variables, 262
    - parameters, 265–266
    - univariate and multivariate data, 264–265
  - summary statistics, 267–288
    - centrality, 267–270
    - correlation, 280–285

- statistics, *continued*
  - summary statistics, *continued*
    - counts, 271–273
    - covariance, 280–285
    - five-number summary, 274–275
    - outliers, 285–288
    - percentages, 271–273
    - percentiles, 274–275
    - proportions, 271–273
    - quantiles, 274–275
    - spread, 275–280
- step function, 542–543
- stop function, 242–243
- stopping condition, 237
- strings
  - of characters, 73–74
  - format of, 72
- stringsAsFactors argument, 97
- Student's *t*-distribution, 357–359
- style argument, 250
- subclasses, 119
- subsetting elements
  - from matrices, 42–47
  - in vectors
    - using indexes, 28–33
    - using logicals, 68–72
- substr function, 77
- substrings and matching, of characters, 77–79
- sum function, 253
- summary function, 275, 279, 441, 458, 461, 497
- sum-of-squares (SS), 439, 440
- superscripts, 601
- suppressWarnings function, 247
- surface truncation, 671–679
- survey data frame, 448, 451, 497, 611, 697
- survival package, 740
- Sweave markup language, 757
- switch function, 189–193
- Sys.sleep function, 249–250
- system.time function, 251
- Sys.time function, 250–251

## T

- T (abbreviation for TRUE), 60
- \t (tab) escape sequence, 76
- table function, 268–269, 271, 436
- tables, ANOVA
  - building with aov function, 440–442
  - construction of, 439–440
- tapply function, 270, 272, 279, 436, 444
- tcl parameter, 595–596, 603
- t*-distribution, 369, 372
- terrain.colors function, 635, 647, 688
- testing. *See* hypothesis testing
- text, adding to plots, 134–139
  - font, 597–598
  - Greek symbols, 598–599
  - mathematical expressions, 599–601
- text function, 134, 353, 585, 702
- text3d function, 702
- textual progress bars, 249–250
- theme\_bw function, 292–293
- third-party editors, 11
- ticktype argument, 681
- times argument, 25–26
- timing, 250–252
- title function, 598–599
- title labels, on plots, 130–131
- to argument, 24
- topo.colors palette, 635, 653, 665, 677, 699, 706
- topographical map example, 657–658
- tori, 729–735
- trace.factor argument, 446
- trace.label argument, 446
- traditional R graphics, 139
- transpose, of matrices, 47–48
- trees data frame, 516, 526
- trimodal distribution, 326
- triplet, 632
- trivariate functions, 709, 712



- trivariate surfaces, 709–719
  - evaluation coordinates in 3D, 709–710
- isosurfaces, 710–715
  - basic one-level isosurface, 712–714
  - controlling multiple levels with color and opacity, 714–715
  - higher-dimensional probability densities, 710–711
- nonparametric trivariate density
  - example, 715–720
  - calculating 3D estimate, 717
  - isosurface level selection, 717–720
  - raw data, 716
- TRUE value, 27, 60
- try statements, 244–248
  - suppressing warning messages, 246–248
  - using in body of function, 245–246
- tryCatch function, 246
- "try-error" class, 244
- tseries package, 149, 741
- t.test function, 391–392, 397–398
- t*-tests
  - two-sample, 393
  - Welch's, 394
- .txt extension, 150
- txtProgressBar function, 249, 250, 258
- Type I errors, 421–423
  - Bonferroni correction, 423
  - simulating, 421–423
- Type II errors, 424–428
  - influences on error rate, 426–428
  - simulating, 425–426
- type graphical parameter, 129
- typeof function, 119

## U

- umbilic torus, 734–735
- uniform distribution, 343–347
  - dunif function, 344–346
  - punif function, 346
  - qunif function, 346–347
  - runif function, 347
- union of two events, 312
- units argument, 157
- univariate data, 264–265
- unpooled variances, 393–395
- update function, 533, 571
- update.packages function, 8, 747, 756
- updating, R and installed packages, 746–747
- Usage section, of help files, 10
- USArrests, 306
- UScereal data frame, 622–623, 629

## V

- values. *See also* logical values
  - looping via, 194–197
  - non-numeric. *See* characters; factors; logical values
- var function, 278
- var.equal argument, 397, 402
- variable-mapped facets, and multiple plots, 616–623
  - facets mapped to categorical variable, 619–623
  - independent plots, 616–618
- variance, 275–280
  - of continuous random variable, 326–329
  - of discrete random variable, 317–318
- vector-oriented behavior, of R, 33–37
- vectors, 23–37
  - creating, 23
  - extracting elements
    - using indexes, 28–33
    - using logicals, 68–72

- vectors, *continued*
  - of indexes, 30
  - length of, 27
  - repeating values, 25–26
  - sequences, 24–25
  - sorting, 26–27
  - stand-alone, 117–118
  - subsetting
    - using indexes, 28–33
    - using logicals, 68–72
- viewing angle, of perspective plots, 679–682
- visualization of data. *See* data visualization
- volcano data set, 657, 668

**W**

- warning function, 242
- warning messages
  - overview, 242–244
  - suppressing, 246–248
- Warnings section, of help files, 10
- warpbreaks data frame, 443, 444, 519
- web-based files, reading, 154–155
- Welch's *t*-test, 394
- which function, 69, 70, 71, 80, 431, 551
- while loops, 200–204
- width argument, 579, 603, 625, 628
- Wilcoxon rank-sum test, 401
- wilcox.test function, 401
- Wilson score interval, 405
- wireframe, 679
- working directory, 7
- workspace image files, saving and loading, 11–12
- write.csv function, 156
- write.table function, 156
- writing files. *See* reading and writing files
- writing functions, 215–240
  - arguments, 222–232
  - defaults, setting, 225–227
  - ellipses and, 228–233
  - lazy evaluation, 222–225
  - missing, checking for, 227–228
  - function command, 215–222
    - creating functions, 218–219
    - return statement, 220–222
  - specialized functions, 233–240
    - disposable functions, 236
    - helper functions, 233–236
    - recursive functions, 237–240
- writing R packages, 749–750

**X**

- xaxis parameter, 592
- xaxt parameter, 593
- x.factor argument, 446–447
- xlab (axis label) parameter, 129, 130
- XLConnect function, 153
- xlim (plotting region limit)
  - parameter, 129, 133, 642, 673
- xpd parameter, 584

**Y**

- yaxis parameter, 592
- yaxt parameter, 593
- ylab (axis label) parameter, 129, 130
- ylim (plotting region limit)
  - parameter, 129, 133, 673

**Z**

- z value, coloring bivariate surfaces by, 703–704
- zlim argument, 641, 647, 648, 688
- z-matrix, 654, 655
  - conceptualizing, 656–657
  - constructing, 655–656
- Z.test function, 409
- Z-tests, 402
- zval argument, 641, 642, 648