

Index of R Functions

The round brackets following the function name enclose, either the name of the package that has the function, or the function used to create the object that was supplied as argument. Selected appearances only are cited for the very common functions.

- abbreviate (*base*), 11, 81, 82
- abline (*graphics*), 44, . . . , 85
- abs (*base*), 85
- acf (*stats*), 84, 136, 171, 243, 295, 300, 302
- add1 (*stats*), 176, 178
- addControl (*inline code*), 446, 447
- addmargins (*stats*), 24
- adjustcolor (*grDevices*), 86, 110, 310, 494
- aggregate (*stats*), 11, 26, 86, 329, 476, 492
- AIC (*stats*), 58–62, 66, 86, 107, 112, 155, 176–179, 207, 223, 226, 231, 241, 242, 288, 301, 347, 452, 468
- AICc (*AICcmodavg*), 59, 60, 176–179, 241
- ampute (*mice*), 458
- anova (glm()), 260
- anova (lm()), 106, 123, 125, 146, 176, 178, 205, 207, 218, 241, 243
- anova (lme()), 212
- anova (lmer()), 347
- any (*base*), 484
- aov (*stats*), 123, 209, 210, 212–214, 216
- apply (*base*), 52, 82, 84, 87, 121, 138, 177, 178, 249, 264, 459, 464, 465, 479
- ar (*stats*), 293, 297, 315
- Arima (*forecast*), 299, 307
- arima (*stats*), 297, 299, 307, 316
- arima.sim (*stats*), 243, 302, 303, 315, 317
- array (*base*), 475
- as.character (*base*), 476
- as.data.frame (*base*), 63, 64, 408, 411, 459, 481
- as.Date (*base*), 17, 474
- as.DNAbin (*ape*), 465
- as.integer (*base*), 142
- as.layer (*latticeExtra*), 275, 491, 492
- as.matrix (*base*), 409, 410, 459, 465
- as.numeric (*base*), 249, 474, 476
- as.vector (*base*), 23, 91
- as_tibble (*dplyr*), 481
- auto.arima (*forecast*), 293, 299–306, 308, 309, 311, 312, 316, 367
- avPlots (*car*), 197, 198
- bestsetNoise (*DAAG*), 185
- betabinomial (*glmmTMB*), 351
- BIC (*stats*), 58, 62, 65, 107, 133, 134, 176–179, 189, 224, 226, 231, 299, 301, 307, 309, 347, 420, 452, 468
- BIC (gam(), gamm()), 309
- BIC (glm(), gam()), 452
- binomial (glm), 355
- binomial (*stats*), 424, 452
- boot (*boot*), 68, 70, 119, 120, 139
- boot.ci (*boot*), 69, 70, 139
- box (*graphics*), 97
- Box.test (*stats*), 232, 243, 305, 306
- boxcox (*MASS*), 114
- boxplot (*graphics*), 410, 435, 490
- brewer.pal (*RColorBrewer*), 414, 494
- bsnVaryNvar (*DAAG*), 186
- bwplot (*lattice*), 492
- c (*base*), 20, 34, . . . , 470, 471, 474, 475, 483
- capture.output (*utils*), 260
- cat (*base*), 427, 470, 484
- cbind (*base*), 308, 309, 317, 339, 351, 355, 357, 363, 404, 408, 425, 427, 452, 454, 460, 464
- cc (*mice*), 459
- chisq.test (*stats*), 101, 102
- cloud (*lattice*), 401
- cmdscale (*stats*), 406–408, 465
- coef (*stats*), 411
- col (*base*), 427, 464
- colnames (*base*), 425, 427, 430, 431, 437, 464, 478
- colSums (*base*), 430

- compareTreecalcs (*DAAG*), 389, 393
- complete (*mice*), 459
- complete.cases (*stats*), 80, 404, 406, 410
- confint (*lme4*), 327
- contr.SAS (*stats*), 212
- contr.sum (*stats*), 211, 212
- contr.treatment (*stats*), 212
- contrasts (*stats*), 211, 212, 366
- cor (*stats*), 30, 70, 81, 94, 361, 464
- cor.test (*stats*), 30, 95
- corARMA (*nlme*), 309
- corCAR1 (*nlme*), 367
- cox.zph (*survival*), 286–288
- coxph (*survival*), 288
- crPlots (*car*), 152, 163
- cumulative (*VGAM*), 279
- curve (*graphics*), 82, 140, 418
- cut (*base*), 465
- cutree (*stats*), 414
- CVbinary (*DAAG*), 256, 257, 414, 425
- cvdisc (*hddplot*), 439–441
- CVlm (*DAAG*), 117, 118
- cvscores (*hddplot*), 442
- data (*utils*), 187, 316, 392, 434, 464, 489, 490
- dbinom (*stats*), 31, 96
- decideTests (*limma*), 432, 433, 443
- defectiveCVdisc (*hddplot*), 440
- density (*stats*), 398, 447, 465
- densityplot (*lattice*), 13
- deparse (*base*), 485
- deviance (*VGAM*), 280
- dfbetas (*stats*), 171, 175, 176
- diag (*base*), 413
- diff (*base*), 16, 29
- diff.ts (*stats*), 299, 316
- dim (*base*), 303, 397, 398, 440
- dimnames (*base*), 184, 260, 303, 445
- dist (*stats*), 406, 408, 414, 415, 465
- dist.dna (*ape*), 407, 465
- dnorm (*stats*), 34, 418
- dotplot (*lattice*), 492
- dpois (*stats*), 32
- drop1 (*stats*), 178, 305
- droplevels (*base*), 366, 367
- droplevels (*stats*), 239, 264
- dummy.coef (*stats*), 212
- eBayes (*limma*), 431, 442, 443
- eqscplot (*MASS*), 465
- errorsINx (*DAAG*), 192, 193, 206
- ets (*forecast*), 293, 303
- exp (*base*), 17, 39, 63, 100, 113, 134, 139
- expand.grid (*base*), 87
- expression (*base*), 84, 307
- extractLT (*qra*), 355
- factor (*base*), 25, 52, 81, 85, 120, 122, 406, 430, 447, 450, 474–477
- fitted (glm()), 255
- fitted (aov()), 210
- fitted (gam()), 227, 307
- fitted (gamm()), 317
- fitted (lm()), 223
- fitted (lmer()), 326, 327, 342, 343
- forecast (*forecast*), 301, 316
- format (*base*), 17
- ftable (*stats*), 24
- gam (*mgcv*), 208, 219, 220, 222, 225–227, 229, 231, 233–237, 240, 242–245, 274, 275, 304–306, 309–311, 317, 449, 452, 453
- gam.check (*mgcv*), 226, 229, 242
- gamlss (*gamlss*), 97, 98, 140, 268, 269, 271, 274, 357
- gamm (*mgcv*), 220, 233, 304, 308, 309
- garch (*tseries*), 313, 314
- ggplot (*ggplot2*), 454
- glm (*stats*), 99, 225, 233, 245, 247, 249, 253, 259, 260, 262, 263, 265, 266, 271, 278, 290, 357, 424, 425, 451, 452, 455, 457
- glmer (*lme4*), 277, 357
- glmmTMB (*glmmTMB*), 349, 351, 355
- grid.arrange (*gridExtra*), 454
- gsub (*base*), 350
- hatvalues (*stats*), 174, 277
- hclust (*stats*), 414, 415
- head (*utils*), 432, 484
- histogram (*lattice*), 35
- HoltWinters (*stats*), 314
- ic (*mice*), 459
- ifelse (*base*), 447
- importance (*randomForest*), 391, 399
- interaction.plot (*stats*), 129, 347
- IQR (*stats*), 86
- is.factor (*base*), 484
- is.na (*base*), 397, 445, 447, 484, 488
- isoMDS (*MASS*), 407, 408
- kmeans (*stats*), 416
- layer (*latticeExtra*), 52, 61, 492
- layout (*graphics*), 136
- lda (*MASS*), 422, 424–429, 455, 456, 465, 466
- legend (*graphics*), 491
- length (*base*), 27, 50, 52, 54, 60, 84, 87, 91, 120, 127, 134, 142, . . . , 406, 465, 479, 484
- levels (*base*), 430, 431, 434, 437, 443, 474–476
- levels (*stats*), 283
- library (*base*), 6, 39, 68, 87, 91, 119, 123, 135, 178, 186, 235, 239, 240, 243, 244, 268, 271, 272, 275, 279, 300, 303, 305, 314, 316, 317, 325, 328, 350, 351, 357, 367, 376, 390, 398, 489–491, 494
- lines (*graphics*), 490
- list (*base*), 482, 485, 491, 492

- lm (*stats*), 43–47, 60, 61, 105–107, 110, 115, 119, 131, 134, 135, 140, . . . , 486
- lmBF (*BayesFactor*), 178
- lme (*nlme*), 131, 367, 372
- lmer (*lme4*), 131, 318, 320, 325, 326, 328, 330, 331, 334, 335, 342, 346, 361, 362, 364, 365, 372
- lmFit (*limma*), 431, 437, 443
- lmList (*lme4*), 361–363
- lmrob (*robustbase*), 108
- loess (*stats*), 105, 106, 220, 237, 380
- log (*base*), 16–18, . . . , 471, 486, 487, 489
- log10 (*base*), 15, 422
- log2 (*base*), 431
- logLik {*stats*}, 60
- love.plot (*cobalt*), 454
- lowess (*stats*), 44, 220, 237
- lqs (*MASS*), 108, 172, 173, 206

- mad (*stats*), 28, 86
- makeClust (*inline code*), 413, 416
- makeContrasts (*limma*), 431, 443
- margin.table (*base*), 20, 89, 90
- match (*base*), 326, 334, 348, 363, 447, 451, 454, 491
- match.data (*matchIt*), 451, 453, 456
- matchit (*matchIt*), 451–453, 455, 456
- matrix (*base*), 52, 54, 102, 103, 134, 136, 172, 184, 189, 441, 445, 487
- max (*base*), 403, 491
- MCMCregress (*MCMCpack*), 134
- md.pattern (*mice*), 309, 458, 459
- mean (*base*), 6, 11, 25, 26, 28, 39, . . . , 478, 479, 484, 486–489
- mean.and.sd (*inline code*), 485
- median (*stats*), 68, 478, 484, 488
- mice (*mice*), 459, 460
- min (*base*), 403
- mixed (*afex*), 325, 332, 366
- model.matrix (*stats*), 47, 147, 174, 210, 242, 430, 437
- modifyList (*utils*), 404
- monthplot (*stats*), 303
- mqgam (*qgam*), 238, 239
- mtext (*graphics*), 302, 414, 420, 490

- na.omit (*stats*), 332, 403, 409, 420, 427
- names (*base*), 26, 45, 51, 85, 294, 326, 329, 334, 363, 397, 478, 480, 483
- NBI (*gamlss.dist*), 357
- ncol (*base*), 459
- nrow (*base*), 60, 81, 142, 478
- ns (*splines*), 225, 452
- numeric (*base*), 61, 63, 142, 257, 314, 406, 490

- onewayPlot (*DAAG*), 123
- options (*base*), 35, 209, 211, 332, 366, 472, 484, 488
- order (*base*), 105, 239, 339, 441

- ordered (*base*), 279
- outer (*base*), 82, 134
- overimpute (*micemd*), 460, 461

- p.adjust (*stats*), 124, 127, 432
- pairs (*graphics*), 154, 157, 187
- palette (*grDevices*), 494
- panel.abline (*lattice*), 52
- panel.dotplot (*lattice*), 13, 25, 492
- par (*graphics*), 148, 262, 414, 421, 493, 494
- paste (*base*), 15, 63, 84, 85, 127, 134, 152, 177, 184, 280, 302, 303, 339
- paste0 (*base*), 27, 52, 54, 61, 116, 134, 153, 166, 181, 350, 386, 387, 408, 431, 460
- pbinom (*stats*), 31, 95
- plot (*base, generic*), 12, 15, . . . , 471, 472, 485, 490, 491
- plotCI (*gplots, plotrix*), 489
- plotcp (*rpart*), 385
- plotMDS (*limma*), 431, 437
- plotMix2 (*inline code*), 418
- plotQQuif (*DHARMA*), 353
- plotResiduals (*DHARMA*), 353
- plotSimDiags (*DAAG*), 181
- plowerTransform (*car*), 163
- pnorm (*stats*), 34
- points (*graphics*), 15, 46, . . . , 463, 490, 491
- poisson (*stats*), 155, 357
- poly (*stats*), 222, 225, 241, 242, 259, 288, 308
- pool (*mice*), 459
- posterior (*BayesFactor*), 62
- power.t.test (*stats*), 54
- powerTolog (*inline code*), 272
- powerTransform (*car*), 84, 113, 114, 140, 164, 165, 168, 169, 310
- ppois (*stats*), 32, 95
- prcomp (*stats*), 141, 403, 404, 406, 409
- predict (aov()), 211
- predict (gam()), 275
- predict (glm()), 255, 262, 266
- predict (lm()), 45, 109, 120, 121, 181, 188, 205, 272
- predict (nls()), 201
- predict (qgam()), 240, 272
- predict (vglm()), 280
- pretty (*base*), 17, 34, 334
- printcp (*rpart*), 376, 385, 388, 398
- prod (*base*), 440
- profile (lmer()), 327, 328, 332
- prop.table (*base*), 89
- prune (*rpart*), 386, 388, 398
- pt (*stats*), 50–52, 84, 85

- qbinom (*stats*), 31
- qdo (*qgam*), 240
- qgam (*qgam*), 225, 238–240, 272
- qnorm (*stats*), 34
- qpois (*stats*), 32
- qqmath (*lattice*), 84
- qqnorm (*stats*), 236, 306

- preference (*DAAG*), 36, 46
 qt (*stats*), 41, 63, 65, 87, 109, 134
 qtukey (*stats*), 211
 quasipoisson (*stats*), 357
 quickplot (*ggplot2*), 406, 464, 465

 rainbow (*grDevices*), 416
 randomForest (*randomForest*), 390–399, 450, 455, 464
 ranef (lmer()), 326, 334, 342, 344
 range (*base*), 16, 17, 46, 227, 240, 266, 363, 479, 482, 484, 488
 rbind (*base*), 23, 61, 65, 81, 96, 98, 101, 134, 327, 399, 413, 459
 rbinom (*stats*), 31
 refit (lmer()), 344
 regsubsets (*leaps*), 185, 189
 releve (*stats*), 122, 209, 357
 reorder (*stats*), 319
 rep (*base*), 3, 35, 52, ..., 471, 475–477, 487
 rePCA (lmer()), 365
 residuals (gam()), 232, 235, 243, 305, 306, 309
 residuals (gamlss()), 271
 residuals (glm()), 262, 263, 276
 residuals (lm()), 46, 105, 197
 residuals (lmer()), 334, 343
 rev (*base*), 484
 rlm (*MASS*), 108, 140
 rm (*base*), 472
 rmvnorm (*mvtnorm*), 413
 rnorm (*stats*), 34–36, 39, 52, 61, 82–84, 86, 184, 206, 314, 440, 484, 486
 round (*base*), 11, 23, 27, 28, ..., 478
 row (*base*), 303, 427, 441, 464
 row.names (*base*), 175, 181, 401, 463, 465
 rownames (*base*), 82, 97, 274, 332, 355, 357, 363, 425, 427, 464, 478
 rowSums (*base*), 430
 rpart (*rpart*), 376, 378, 380–383, 385–387, 389, 390, 393, 394, 396, 398
 rpart.control (*rpart*), 384
 rpart.plot (*rpart.plot*), 376, 378, 381, 386, 388
 rpois (*stats*), 32, 83
 rt (*stats*), 83

 sammon (*MASS*), 407, 408, 465
 sampdist (*DAAG*), 39
 sample (*base*), 3, 82, 116, 142, 164, 257, 372, 374, 397–399, 406, 435
 supply (*base*), 16, 17, 25, 27, 28, 91, 126, 127, 177, 179, 216, 241, 263, 264, 326, 338, 363, 367, 397, 398, 403, 445, 459, 479, 492
 scale (*base*), 11, 86, 344, 350, 351, 415
 scam (*scam*), 225
 scatter.smooth (*stats*), 106, 207, 380
 scoreplot (*hddplot*), 436, 442
 sd (*stats*), 27, 28, 50, 52, 84–86, 91, 121, 478, 479, 484, 488
 search (*base*), 489, 490

 seq (*base*), 3, 17, 62, 142, 339, 374, 484, 487
 set.seed (*base*), 3, 32, 33, 35, 52, 67, 68, 70, 116, 119, 257
 setNames (*stats*), 27, 29, 34, 49, 50, 63, 65, 85, 91, 96, 98, 127, 153, 178, 179, 239, 241, 266, 408, 460, 486
 signif (*base*), 261
 simulate (*stats*), 46, 344
 simulateLinear (*DAAG*), 125
 simulateResiduals (*DHARMa*), 352
 sort (*base*), 85, 350, 362, 399, 410, 432, 484
 split (*base*), 25, 27, 263, 326, 338, 344, 363, 367, 479, 492
 splom (*lattice*), 401
 spm (*car*), 149, 310, 398, 460
 sqrt (*base*), 50, 52, 62, 63, 81, 84, 87, ..., 470, 472
 str (*utils*), 80
 strip.custom (*lattice*), 52, 307, 491
 stripplot (*lattice*), 123
 StructTS (*stats*), 314
 subset (*base*), 13, 68, 85, 101, 110, 181, 207, 239, 264, 265, 268, 275, 366, 367, 445, 447, 454, 458, 487
 substitute (*base*), 447, 485
 substring (*base*), 357, 363, 431
 sum (*base*), 24, ..., 427, 445, 465, 476, 488
 summary (generic), 489
 suppressMessages (*base*), 350, 376
 suppressPackageStartupMessages (*base*), 135, 178, 303, 305, 314, 357, 390
 Surv (*survival*), 283, 285–288
 survfit (*survival*), 283, 285

 t (*base*), 43, 86, 274, 399, 436, 484
 t.test (*stats*), 50, 63, 83, 92, 93, 363, 483
 table (*base*), 23, ..., 474, 476, 488
 tapply (*base*), 334
 teigen (*teigen*), 420, 422
 termplot (aov()), 216
 termplot (glm()), 256, 259
 termplot (lm()), 163, 164, 201, 225
 termplot (rlm()), 207
 text (*graphics*), 82, 463, 465
 time (*stats*), 293
 title (*graphics*), 490
 topTable (*limma*), 432, 443
 ts (*stats*), 15, 294, 304
 ttest.tstat (*BayesFactor*), 63–65, 85, 87, 133, 134
 ttestBF (*BayesFactor*), 62–64, 87
 tuneRF (*randomForest*), 390

 unclass (*base*), 13, 334, 437, 465
 unique (*base*), 25, 81, 262, 348, 350, 484
 unit (*grid*), 454
 unlist (*base*), 17, 425, 478, 479, 483
 unname (*base*), 96, 174
 update (generic), 16, 17, 36, 52, 176, 180, 304, 311, 319, 348, 351, 401, 491, 492

var (<i>stats</i>), 83, 124, 264, 315, 334	white.test (<i>tseries</i>), 313
VarCorr (lmer()), 326, 330, 331, 334, 372	window (<i>stats</i>), 15, 293, 303, 316
vif (<i>DAAG</i>), 189	with (<i>base</i>), 478, 479, 483, 484, 489, 491, 492
voom (<i>limma</i>), 430, 431, 437	within (<i>base</i>), 149, 156, 165, 227, 251, 319, 350, 362, 363, 478
vooma (<i>limma</i>), 437, 442	
WDI (<i>WDI</i>), 239	xtabs (<i>stats</i>), 23, 24
weighted.mean (<i>stats</i>), 96, 98	xyplot (<i>lattice</i>), 404, 408, 490, 491
which (<i>base</i>), 484	
which.max (<i>base</i>), 410, 484	yjPower (<i>car</i>), 168, 169, 236, 253, 272, 274
which.min (<i>base</i>), 398, 484	

Index of Terms

- abline (*graphics*), 15
- Akaike information criterion (AIC and AICc),
 see model selection
- allometric growth, 88, 114, 115
- analysis of variance
 - linear regression, 179
 - multilevel experimental design, 335
- analysis of variance, leave one out
 GAM model, 305
- analysis of variance, sequential
 - balanced incomplete block design, 214
 - linear regression, 146
 - multiple lines, 218, 241
 - randomized block design, 213
- argument, *see* function

- Bayes Factor, 58, 61
 - complement or replace p -value, 78
 - null interval, 64
 - prior odds, 63
 - scale of evidence, 65
 - Sellke et al. upper limit, 63, 132
 - vs. BIC, 176
 - vs. p -value
 - sample size differences, 64
- Bayesian estimation
 - normal priors and likelihood, 132
- Bayesian inference
 - commentary, 131
 - discrete prior probabilities, 48
 - false discovery rate, 55
 - Jeffreys Unit Information (JUI) family of priors, 65
 - Jeffreys–Zellner–Siow (JZS) family of priors, 61
 - Markov Chain Monte Carlo (MCMC), 134
 - Monty Hall problem (prize behind one of three doors), 143
 - positive predictive value (PPV), 55, 56, 139
 - posterior, 49, 55, 61–63, 131, 132, 134, 136, 425, 427
 - prior, 49, 51, 53–55, 57, 58, 61–63, 65, 78, 87, 132–134, 391, 392, 424, 426
- Bayesian Information Criterion (BIC), *see* model selection
- binomial models
 - leverage, 277
- binomial-like and count data
 - deviance, 276
 - dispersion (aka dispersion index) estimates, 276
- bootstrap, *see* resampling methods
- bootstrap estimate
 - confidence interval, 69
 - confidence interval for median, 69
 - correlation coefficient, 69
 - SE of median, 67
 - when to use, 70

- C_p statistic (Mallows), *see* model selection
- case studies, 202
 - birthweight paradox, 160
 - Boeing 737 Max 8 2019 crashes, 73
 - deaths from Atlantic hurricanes, 168
 - dietary effects on health, 191
 - Google flu trends, 74
 - hurricane deaths data, 270
 - IQ tests, 367
 - left- vs. right-hander cricketer lifespans, 274, 288
 - lifestyle effects on health, 159
 - medical screening programs, 104
 - moth habitats, 263
 - spread of cholera, 257, 259
 - UCB admissions data, 260
- causal inference, 7, 202, 259
 - balance and matching, 443
 - Bradford Hill criteria, 158
 - coarsened exact matching, 455
 - confounders, 161
 - directed acyclic graphs, 159
 - graphical summary of differences from balance, 453
 - interpretable regression coefficients, 158
 - labor training data, complications, 448
 - labor training data, regression comparisons, 449

- causal inference (*cont.*)
 - propensity scores, 444
 - scores replace regression covariates, 450
 - use GAM fit to derive propensity scores, 452
 - use multiple evidence sources, 158
 - use randomForest proximities to compare groups, 450
- censoring, *see* survival analysis
- classification, *see* discriminant analysis
- cluster analysis
 - k*-means, 416
 - k*-means vs. mixture model, 422
 - agglomerative clustering, dendrogram display, 413
 - hierarchical clustering, 412
 - mixture model-based clustering, 417
 - BIC is –BIC as used elsewhere, 420
 - EM algorithm, 419
 - single, average, and complete linkage, 413
- comparisons, multiple
 - false discovery rate (FDR), 429
 - least significant difference (LSD) vs. Tukey's honest significant difference (HSD), 122
 - multiple range tests, 124
 - regression vs. qualitative comparisons, 124
- comparisons, one-way
 - analysis of variance, 123
- contingency table
 - chi-squared residuals, 101
 - chi-squared test for no association, 101
 - combining data across tables, 138
 - consistency of pattern, 103
 - Fisher exact test, 102
 - interpretation issues, 104
 - Mantel–Haenzel test, 138
 - mechanics of the chi-squared test, 101
- correlation
 - alternative measures, 29
 - Fisher *z*-transformation, 94, 139
 - Kendall, 95
 - Pearson or product–moment, 94
 - spatial autocorrelation, 314
 - Spearman, 95
- cross-validation, *see* resampling methods
- data
 - acupuncture vs. sham, 137
 - comparison of inhaler instructions, 278
 - cuckoo eggs, 4, 12
 - drawings of dreams, 102
 - email spam (complete data), 396
 - guarding time vs. size of group (meerkats), 242
 - inhibition of current flow vs. protein concentration, 290
 - input and output from a file, 480
 - kidney stone surgery, 20
 - labor training data, available datasets, 444
 - machinist accident counts, 98
 - morphometric, 88, 114
 - painkiller experiment, 196
 - rare plant habitats, 103
 - weight gain, toxaphene in rat diet, 139
- data analysis
 - advance planning, 9
 - challenges to results, 9
 - failures of human intuition, 75
 - organization and management of work, 70
 - risks of naive and misleading analyses, 75
 - the benefits of shared insights, 10
 - two-sample *t*-test, 91
- data collection and analysis
 - planning, 6
- data, *compositions*
 - Coxite, 187
- data, *DAAG*
 - ACF1 (rat colon growths), 82, 261
 - ais (measures on athletes), 81
 - allbacks (book dimensions and weight), 144, 195, 204
 - anesthetic, 248
 - ant111b (corn yield measurements), 319
 - appletaste, 215
 - bomregions2021 (annual weather indices, Murray–Darling Basin, Australia), 304, 315
 - cities, 203
 - coralPval (coral gene expression comparison), 125
 - cricketer (left- vs. right-hander cricketer lifespans), 288, 291
 - cuckoohosts, 81
 - cuckoos, 84, 241, 479, 492
 - dewpoint, 234
 - frogs, 250
 - frostedflakes, 311, 315
 - fruitohms (resistance vs. juice content), 227, 244
 - geophones (seismic timing), 242
 - headInjury, 290, 396
 - hills2000, 172, 180, 204
 - houseprices, 119
 - humanpower1 (oxygen input vs. power output), 360
 - hurricNamed (deaths from Atlantic hurricanes), 168, 236, 270, 289
 - ironslag, 104
 - jobs, 17, 293, 479
 - kiwishade (kiwifruit vine yields), 25, 335
 - LakeHuron, 315
 - leafshape, 141, 464
 - leafshape17, 424
 - leaftemp, 218
 - litters (mouse brain vs. body weight), 47, 157
 - mdbAVtJtoD (monthly average temperatures, Murray–Darling Basin, Australia), 303
 - measles, 14
 - mifem, 290

- mifem (female heart attack patient mortality), 384, 396
- mignonette (crossed vs. self-fertilized plants), 92
- moths, 264, 290, 356
- nassCDS (road accident fatality data), 23
- nihills, 148, 180, 197, 204
- nswdemo (labor training evaluation data), 82, 100
- nswpsid1 (labor training evaluation data), 22, 396, 465
- oddbooks, 154, 204
- orings (1986 space shuttle disaster), 80
- ozone, 243
- pair65 (stretch of elastic bands), 35, 140
- possum, 12, 13, 80, 81, 139, 401, 427, 463
- repPsych (psychology replication study), 207
- rice, 128, 213
- rockArt, 465
- roller, 43
- science (student attitudes to science), 7, 329
- seedrates, 220, 242
- socsupport, 7, 408
- sorption, 372
- spam7 (email spam), 374, 387, 390
- sugar, 209
- tinting, 344, 491
- toycars, 241
- two65 (elastic band stretch), 28, 91
- wages1833, 242
- worldRecords, 243
- data, *DAAGbio*
 - plantStressCounts (mRNA gene expression counts), 430
 - primateDNA (bases at 232 mitochondrial locations, for 14 primates), 465
- data, *datasets*
 - airquality, 309, 317, 460
 - cars (stopping distance vs. speed), 85
 - chickwts, 476
 - EuStockMarkets, 316
 - LakeHuron, 293
 - pressure (vapor pressure vs. temperature), 140
 - UCBAdmissions, 89, 138, 260, 479
- data frame, 81, 477
 - as database, 478
 - vs. matrix, 478
 - vs. tibble, 481
- data, *gamclass*
 - FARS (road accident fatality data), 24
- data, *gclus*
 - wine (Italian wines, three cultivars), 464
- data, *ggplot2*
 - diamonds (diamond prices), 86, 398
- data, *hddplot*
 - Golub (cancer gene expression indices), 433
- data, *HistData*
 - Cholera (deaths, London 1848–49), 259
 - GaltonFamilies, 110
- data manipulation, 487
 - tidyverse packages, 481
- data, *MASS*
 - Aids2, 282
 - Animals, 16, 81
 - biopsy (breast cancer biopsies), 397
 - cement (heat in setting), 204
 - cpus (performance, 209 CPUs), 204
 - galaxies (speeds of 82 galaxies), 82
 - painters (assessments of 54 classical painters), 463
 - Pima.tr (diabetes in Pima Indian women), 391, 398
 - whiteside (gas consumption vs. temperature), 231
- data measurement issues, 7
 - Beck Depression Inventory (BDI), 8
 - DALYs and QALYs, 8
- data, *MEMSS*
 - ergostool, 372
 - Gun, 372
 - MatchAchieve, 372
 - Orthodont, 362
- data, *mice*
 - Boys (data on 748 Dutch boys), 458
- data, *mlbench*
 - Vehicle, 464
- data, *MPV*
 - radon (radon release in showers), 85
 - table.b1 (NFL team performance data), 207
 - table.b3 (gas mileage etc. for 32 cars), 205
- data, *qra*
 - HawCon, 350
 - malesINfirst12, 96, 140
- data, *rpart*
 - car.test.frame (mileage vs. weight), 381
- data structures
 - matrices and arrays, 475
- data, *tseries*
 - ice.river, 316
- database, 490
- datasets
 - in R packages, 489
- dates
 - functions for use with, 474
- dependence
 - Markov chain, 141
 - temporal or spatial, 6
- deviance, 347, 357, 378–380
 - generalized additive model, 227–229, 232

- deviance (*cont.*)
 - generalized linear model, 247–249, 253, 260, 262, 268, 276, 277
 - shape constrained additive model (SCAM), 231
 - vglm model, 280
- discriminant analysis
 - linear discriminant analysis (LDA), 435
 - logistic discrimination vs. linear discriminant analysis (LDA), 422
 - misclassification costs, 424
- distribution, 30
 - betabinomial, 96, 247
 - betabinomial vs. binomial, 96
 - binomial, 31
 - Cauchy, 61, 62, 65
 - changes in variability, 21
 - clustering, 21
 - dispersion index vs. dispersion parameter, 99, 247, 349
 - distributional asymmetry, 21
 - double binomial, 140
 - exponential, 34
 - extra-binomial and extra-Poisson, 95
 - extra-Poisson variation, 98
 - fitted vs. empirical centiles, 273
 - Jeffreys, 65
 - measures of variation, 27
 - negative binomial (NBI and NBII), 98, 247, 267
 - normal ('Gaussian'), 33, 34
 - normal approximation to binomial, 93
 - Poisson, 32, 262
 - quasibinomial adjustment, 247, 271
 - quasipoisson adjustment, 261, 264
 - t -, 40
 - technical notes on modeling extra-binomial or Poisson variation, 99
 - time trends, 22
 - uniform, 34
- effect size, 29, 53
- equation
 - Claudius–Capeyron, 140
- errors in x
 - classical vs. Berkson error model, 194
 - reliability ratio, 192
 - simulations, 192, 193
 - two explanatory variables, 193
- experimental design, 6
 - multi-site, 344
- factor
 - alternative choices of contrasts, 211
 - functions for working with, 475
 - missing values, 477
 - ordered factor, 476
- false discovery rate (FDR), 125, 126
- fitted values
 - calculation, 47
 - calculation from columns of model matrix, 210
- frequentist inference
 - confidence intervals, 50
 - hypothesis testing and p -values, 50
 - p -value, 50, 56
 - p -value decision process, 51
 - p -value simulation, 51
 - power, 52
 - true and false positives, 53
- function
 - anonymous, 486
 - argument, 404, 472, 473, 475, 477, 478, 481, 484, 486, 491, 492, 494
 - common useful built-in functions, 483
 - default argument, 485
 - generic (tune action to class of object), 485
 - in R packages, 489
 - operator, 486
 - return value, 485
 - user-written, 484
- generalized additive mixed model (GAMM)
 - correlated errors, 308
- generalized additive model (GAM), 225, 317
 - abilities in mgcv and related packages, 233
 - binomial errors, 275
 - check residuals for ARIMA errors, 305
 - choice of penalty, 226
 - diagnostic checks, 229
 - monotone curves, 230
 - multiple smoothing terms, 233
 - smooth for each factor level, 231
 - smooth surface, 235
- generalized additive model, shape constrained (SCAM), 230
 - diagnostic plots, 231
- generalized linear mixed model
 - betabinomial error, 349
 - diagnostic plots, 353
 - dispersion index, 352
 - intra-class correlation, 349, 352
 - lethal time confidence interval, Fieller's formula, 353
 - lethal time estimate, 353, 355
 - link function, 350
 - observation-level random effects, 356
 - scale correction, 351
- generalized linear model
 - choice of link function and reference level, 265
 - common link functions, 246
 - complementary log-log link, 276
 - count data, 261
 - diagnostic checks, 269
 - fitted values, 255
 - linear predictor, 246
 - log(odds) scale with logit link, 247
 - logistic multiple regression, 250
 - logit link, 246

- probit link, 276
- square root link, 269
- types of residuals, 276
- variance inflation factor, 252
- graphics
 - added variable plot, 197, 198
 - added variable plot, algebraic details, 199
 - boxplot, 5
 - boxplot, compared with density curve, histogram, and rug plot, 12
 - categorical data, 19
 - classification tree (email spam), 376
 - component+residual plot, 151, 163
 - component+residual plot or termplot, 163
 - density contours of fitted mixture model, 421
 - diagnostic checks for GAM model, 229
 - dotplot, 4
 - dotplot, with overlaid group means, 264
 - dynamic graphics, 174, 492
 - flawed graphs (variable selection bias), 435
 - influence plot, 176
 - interaction plot, 347
 - join separately created lattice objects, 492
 - lattice package, 491
 - leverage vs. fitted proportion, 278
 - linear discriminant analysis, scatterplot matrix for canonical variates, 428
 - logarithmic scale(s), 14, 15
 - logarithmic scale(s), labeling, 19
 - love plot (summarizes differences from balance), 454, 456
 - mean-variance relationship for gene expression counts, 431
 - mosaic plots, 20
 - normal quantile–quantile plot, 35, 36
 - plot characters, symbols, line types, and colors, 493
 - regression tree, 381
 - same vs. sliced logarithmic scale, 17, 18
 - scatterplot matrix, 149
 - scatterplot matrix, suggested power transformations, 169, 254
 - simulated diagnostic plots, 181
 - termplot, 151, 153, 164, 182, 197, 201, 225, 227, 255, 258, 305, 307
 - termplot equivalent for GAM model, 236
 - termplot equivalent for smooth terms in GAM model, 306
 - three-dimensional perspective plot, 402
 - time series autocorrelation and partial autocorrelation plots, 295
 - time series lag plots, 295
 - time series predictions, 301
 - time series seasonal effects, 303
 - time series trace plot, 294
 - univariate, breakdown by one or more factors, 13
 - worm plot, 98, 140
- graphics packages
 - graphics, grid, lattice, latticeExtra, ggplot2, 490
- graphics, generalized linear model, diagnostic plots, 269
- imputation, *see* missing values
- likelihood and log-likelihood, 42, 58
 - maximum likelihood estimation, 42
- likelihood and maximum likelihood, 301
- likelihood ratio test, 268
- linear model, *see* regression
- list, 482
 - as opposed to “atomic” vector, 482
- machine learning
 - algorithmic motivations, 72
 - limits of current systems, 73
- missing values, 243, 282, 309, 394, 487
 - alternative approaches, 74
 - counting and identifying, 488
 - in modeling functions, 488
 - missing at random (MAR), 457
 - missing completely at random (MCAR), 457
 - multiple imputation, 315, 457
- model
 - allometric growth, 115
 - assumptions, 88
 - limitations, 90
 - outliers, 21
- model matrix, 210
- model object
 - extractor functions, 45
- model predictions
 - fitted values vs. new observations, 109, 166
- model selection
 - AIC, AICc, and BIC, 58, 176, 207, 300
 - BIC vs. Bayes Factor, 176, 178
 - drop1() and add1(), 178
 - Mallows C_p , 178, 207
 - over-fitting, 184–186
 - R-squared and adjusted R-squared, 176, 241
 - sampling properties of AIC differences, 60
- multidimensional scaling (MDS), 406
 - binary data, 408
 - distance measures, 406
- multilevel model
 - analysis of variance sums of squares, 339
 - balance, 318
 - best linear unbiased predictor (BLUP), 326
 - history, 369
 - instructive faulty analyses, 334
 - intra-class correlation, 324, 332
 - Kenward–Roger approximation, 325
 - key ideas, 318
 - levels of variation, 320
 - maximum likelihood (ML) vs. restricted maximum likelihood (REML) criterion, 346

- multilevel model (*cont.*)
 - meta-analysis, 370
 - model selection, 346
 - nested factors, 322, 330
 - prediction levels, 322, 324
 - profile likelihood, 327, 332
 - random effects, 320
 - random vs. fixed effects, 322
 - simplistic error structure, 368
 - variance components vs. analysis of variance
 - mean squares, 323, 338, 341
 - variance components, interpretation, 324
 - within and between block effects, 342
 - within and between class effects, 330
 - within and between site effects, 322, 344
 - within and between subject effects, 344
- multiple imputation, *see also* missing values
- overimputation graphs, 461
- time series cross-sectional data, 460
- multiple *p*-values
 - false discovery rate (FDR) estimates, 432, 442
- multiway table
 - loglinear model, 281
- observational data
 - inference, *see* causal inference
- ordination, *see* multidimensional scaling (MDS)
- outliers, *see* regression
- p*-value
 - adjusted, 127
 - threshold, 78
- package
 - attach or load?, 489
- permutation, 3, 66, 67
- permutation tests, *see* resampling methods
- pipe
 - using the `|>` operator, 485
- positive predictive values, 54
- posterior density or probability, *see* Bayesian inference
- predictive accuracy, 116, 121, 148, 176, 256, 400, 422, 423, 425, 426, 429, 440, 462, 468
- principal components
 - rotations ('loadings'), 404
 - stability relative to statistical variation, 405
- prior density or probability, *see* Bayesian inference
- publication
 - peer review of study plans, 78
 - reporting standards, 57
- published experimental studies
 - independent replication, 57
 - power issues, 55
- quantile regression, 238
 - smoother with 2SE bounds added, 238
- R package task views, 469
- R session
 - image file, 471
 - search list, 478, 490
 - working directory, 32, 70, 71
 - workspace, 33, 71, 471, 472, 477, 478, 484, 489, 490
- random forest
 - boosting and bagging approaches, 396
 - bootstrap samples, 390
 - data values that are not independent, 392
 - importance measures, 391
 - out-of-bag (OOB) error, 390
 - prior probabilities, 391
 - proximity measure, 392
 - tree size, 390
 - tuning parameters, 390
- random numbers, 35, 118
 - generation, 32
 - seed, 3, 32, 33
- random sample, *see* sample, random
- regression, *see also* generalized additive model (GAM), generalized additive mixed model (GAMM), generalized linear model,
 - analysis of variance, 106
 - confusion between explanatory and response variables, 195
 - contributions of basis terms to regression
 - spline fit, 228
 - Cook's distance, 108, 148, 173
 - correlations between parameter estimates, 145
 - criteria for model choice, 162
 - cubic regression spline bases, 225
 - diagnostic checks, 171
 - diagnostic plots, 104, 147, 152, 170
 - diagnostic vs. simulated plots, 181
 - different parameterizations serve different purposes, 151
 - errors in *x*, 191, 206
 - factor contrasts, 209
 - fits to simulated data, 46
 - fitting multiple lines, 216
 - GLM vs. normal theory linear model, 248
 - hat matrix, 174
 - influence on coefficients (dfbetas), 175
 - influential point(s), 148
 - leverage, 173, 277
 - `lm()` vs. `aov()`, 209
 - locally weighted regression smoothers, 237
 - low order polynomial, 114
 - many explanatory variables, 183
 - missing explanatory variables, 195
 - model fitting strategy, 161, 170
 - model formula and model objects, 44
 - model matrix, 47, 220
 - multicollinearity, 186
 - multicollinearity, remedies, 191
 - nonlinear methods, 200

- omission of explanatory variables, 205
- ordered categorical or categorical response, 278
- orthogonal polynomials, 222
- outliers and influence, 107, 173
- polynomial, 220
- polynomial bases, 220
- polynomial fit vs. loglinear fit, 222
- proportional odds logistic, 278
- residuals and residual plots, 45, 46
- resistant, 171–173, 202, 206, 237
- ridge, 205
- robust and resistant, 108, 140, 171–173, 202, 206, 207
- roughness penalty methods, 220
- SE of slope, 111
- selection effects in observational data, 156
- simulate variable selection procedures, 184
- spline bases, 220, 223
- structural equation models, 203
- summary information, 144
- test for main effect in presence of interaction, 212
- thin plate regression bases, 224
- variable selection penalty strategies, 186
- variance inflation factor, 189, 252
- y vs. x , or x vs. y , or functional relationship, 112
- regression, model matrix
 - choice of factor contrasts, 211
- repeated measures, 357
 - alternative approaches, 359
 - correlation structure, 359
 - random coefficients model, 361, 364
 - sequential correlation structure, 367
 - theory, 358
- replicability
 - implying independent replication of a total study, 75
 - laboratory economics studies, 76
 - multiple replicates internationally of one study, 77
 - preclinical cancer research, 77
 - psychology, 76
 - studies where most published results fail, 76
- replication (“reproducibility”) studies, 56
- reproducible reporting
 - gains from sharing data and code, 72
 - markup tools (mix code and text), 71
- resampling methods, 66
 - assumptions, 121
 - bootstrap, 118
 - cross-validation, 116, 180
 - cross-validation, binary response, 256
 - one-sample permutation test, 66
 - permutation tests, 88
 - training/test, 116, 180
 - two-sample permutation test, 67
- residuals, *see also* regression
 - randomized quantile (models with discrete error distributions), 95, 352
 - randomized quantile, simulation, 352, 353
 - robust and resistant methods, *see* regression
- sample, cluster, 3
- sample, nonresponse, 7
 - multiple vs. single imputation, 7
- sample, random, 3
 - bootstrap, 4
 - with replacement, 3
 - without replacement, 3
- sampling bias, 5
- sampling distribution, 37
 - central limit theorem, 37, 38
 - of t statistic, 41
 - of mean (standard error of mean, or SEM), 38, 39
 - of median (SE of median), 39
 - of standard deviation, 39
- search list, *see* R session
- selection bias, 56, 146, 163, 182, 184–186, 191, 203
- simulation
 - logistic regression model, 290
 - Poisson regression model, 291
- standard deviation, 27
 - compared with inter-quartile range (IQR) and mean absolute deviation (MAD), 28
 - degrees of freedom, 27
 - pooled estimate, 28
- summary measures
 - analysis using aggregated data, issues, 26
 - bias from omission of conditioning variable, 90
 - dengue status example, 26
 - R-squared and adjusted R-squared, 106, 176
 - tabulation, 23
 - uses and potential for distortions, 22
 - when to use aggregated data for analysis, 26
- survival analysis
 - censoring, 282
 - Cox proportional hazards, 286
 - data collection process, 282
 - hazard rate, 283, 285
 - left- vs. right-hander cricketer lifespans, 288
 - reasons for right censoring, 284
- survival analysis, Cox model
 - alternative tests for coefficient(s), 287
 - diagnostics, 287
 - time-dependent coefficients, 287
- time sequence
 - test for sequential correlation, 232, 243, 305
- time series
 - AR, ARMA, ARIMA, ETS, ARCH, GARCH, 292
 - AR(1) model, standard error of mean, 297
 - ARCH and GARCH models, 313

- time series (*cont.*)
 - ARIMA (ARMA with differencing) model, 299
 - ARIMA errors, 304
 - ARIMA errors, drift, 299
 - autocorrelation and partial autocorrelation, 295
 - automatic model selection, 299, 307, 312
 - autoregressive (AR) models, 296
 - autoregressive model of order $p > 1$, 297
 - autoregressive moving average (ARMA) model, 299
 - calibration, 311
 - financial and economic, 313
 - innovations, 297
 - moving average (MA) processes, 298
 - residuals for ARIMA models, 308
 - seasonal effects, 303
 - simulation, 315, 317
 - stationarity, 296
 - univariate, 14
- time series objects
 - functions, 293
- time series predictions
 - ARMA, ARIMA, and ETS, 301
- transformation
 - Box–Cox and Yeo–Johnson power, 112, 113, 164, 168, 252, 310
 - logarithmic, 112, 150
 - proportions and counts, 288
 - square or cube root, 304
 - to linearity, 21
- tree
 - cost-complexity parameter, 383
 - event rate and survival data, 395
 - one-standard-deviation rule, 386
 - predictive accuracy, and the cost-complexity tradeoff, 382
 - predictive accuracy, cross-validation, 382
 - printed information on each split, 386
 - pruning, 383
 - splitting criteria, 378, 380
 - toy example, 378
- tree, classification
 - relative vs. cross-validation error rate, 377
 - splitting criteria, 380
- tree, regression
 - vs. loess regression smoothing, 380
- tree-based methods
 - appropriate use, 374
 - as complement to other methods, 374, 395
 - history, 373
 - single-tree vs. random forests, 373, 393
 - vs. other approaches, 394
 - see also* random forest
- variance
 - Fligner–Killeen test for homogeneity, 263
 - heterogeneity, 105, 106, 166, 346, 348
- variance inflation factor, 189
- working directory, *see* R session
- workspace, *see* R session