Basic R Cheat Sheet

getting started:
install.packages("lname")
update.packages()
getwd() & setwd()
library(lname)

loading data:
read.table("filename", sep="\t", header=T)
read.table(pipe("any unix cmd"), header=F)
f<-file(filename,"r"); readLines(f,n=1)
attach(x)
view data: search, head, summary, table
view features: class, attributes, str, length, dim, labels, names
get objects: save, load, get, ls

functions:
?func # help
fname <- function(v1=NULL, v2=NULL) {}
ifelse(condition, if_true, if_false)
conditionals: any(), all(), all.equal(x, y, tolerance=0.0001)
which(x==val) # vector of indices
fix(fname)
do.call("fname", list(v1=NULL, v2=NULL))
source("R code file name")

simple objects:
logical, numeric, complex, character, list, NULL, function, formula
rep(range,numtimes)
seq(start,end,by=interval,length=num)

data structures:
vector: single column of the same type
matrix: columns must have same type & length
array: matrix that can have 2+ dimensions
data frame: matrix where columns can have different types
access with x[["name"]], x[,"name"], or x$name
list: collection of objects (eg c(name="Fred", age=12)
access with list[["name"]] or list[[index]]
factor: divides a list of strings into levels
remap using factor(data,levels=c()) or levels(fdata) = c()

data modifiers:
combine: cbind, rbind, c, append
binning: cut(x, breaks=quantile(x,c(0,.25,.5,.75,1)), labels=c("1","2","3","4"), include.lowest=T)
objects: objects, rm(list=ls())
type: as.[data_type]
order: rev, sort, unique
names: names(vector/list), dimnames(matrix/array)
strings: strsplit, substring, paste, regexpr(regex, string vector), gsub, sub (replace 1st only)

data transformations:
sample(a,size=n,prob=c(..),replace=T)
t(), diag()
reshape(a,timevar="new_cols",id_var="new_rows",v.names="remaining_cols",direction="wide")
table(), table(x > 5) #contingency table
set ops: union, intersect, setdiff (order matters), setequal, %in%

applying a function:
matrix: apply(x,[1=row,2=col],func)
list/vector: sapply(x, FUN=fname)
vector broken into groups by 2nd vector:
aggregate(vector1, vector2, func)
multiple structures: mapply(func, range1, range2)
by(df,df$group,func)
outer(x,y,func) # func(x[i],y[j]) for all i & j

removing elements:
remove<-c("Upstream1","Upstream2")
subset(x,V2 %in% remove)
x[x[,2] %in% remove,]
remove<-c(FALSE,FALSE,TRUE)
x[remove] # returns every third element
x[,-n]
filters: is.na, is.null

output:
plots: hist, boxplot, barplot, plot, pie, pairs (all data combinations)
dev.new()
dev.copy(device=pdf, file="name.pdf")
par(mfrow=c(2,2))
sink("filename", append=T) ... sink()
write.table(a,filename,row.names=F,col.names=F)
cat(output1,output2,…,file=fname,append=T)

other details:
reserved words: c, q, s, t, C, D, F, I, and T
the following are equivalent: apply(x, 2, sum), colSums(x), rep(1, nrow(x)) %*% x
use & and | for booleans on vectors
always passes by value
everything is an object
most operations are vectorized

resources:
http://www.nceas.ucsb.edu/files/scicomp/Dloads/RProgramming/BestFirstRTutorial.pdf
solutions to common errors:
www.burns-stat.com/pages/Tutor/R_inferno.pdf p44-114
actual statistics:
www.stat.berkeley.edu/classes/s133/Rcourse.pdf p38-50
optimizing/debugging programs:
heather.cs.ucdavis.edu/~matloff/132/NSPpart.pdf p129-167
R course: https://www.coursera.org/course/compdata

Basic ggplot2 Cheat Sheet

aes()
args:x, y, color, fill, shape, size, linetype, weight, alpha, factor, group, order, x/yend, xmin/max, ymin/max

geom_*
*=abline, area, bar, bin2d, boxplot, density, density2d, dotplot, errorbar(h), hex, histogram, line, path, point, pointrange, raster, rect, ribbon, smooth, text, violin
args: mapping, data, stat, position, na.rm, position
position="stack", "dodge", "fill", "identity", "jitter"

stat_*
*=function, identity, sum, summary, summary2d, unique
args: bins, na.rm, mapping, data, geom, position
summary: fun.y = var, fun.data = "mean_c1_normal", geom = c("bar","errorbar")

scale_[color/fill]_*
*=brewer, continuous, discrete, gradient(2|n), grey, identity, manual
scale_[shape/linetype/size/alpha]_*
*=continuous, discrete, identity, manual

scale_x_*
*=log10, sqrt, date, datetime, discrete, continuous, reverse
args: palette, name, breaks, limits, labels, colours, guide, na.value

facet
facet_wrap(~group, nrow, ncol) #one plot each
facet_grid(discrete_x~group, nrow, ncol)

coord_cartesian, flip, polar, trans
cut_interval, number
guide(s)_legend, colorbar
labs(ggtitle, xlab, ylab)

notes:

- stat_smooth(geom="point") same as geom_point(stat="smooth")
- *_smooth(): loess smooth for <1000 points; generalized additive model with penalized cubic regression splines otherwise
- pass "show_guide = F" to geom to prevent addition to legend
- use ..count.. to sum into bins

fancy example:

```
fdw <- read.delim("http://bit.ly/fdw_2005")
fdw <- ddply(fdw, .(Child), transform, prob =
value/sum(value))
ggplot(fdw, aes(Age/12, fill = variable)) +
  geom_density(aes(weight = prob, y = ..count..),
position = "fill")+
  facet_wrap(~Sex)+
  scale_fill_brewer(name = "variant", palette =
"Set1")+
  theme_bw()
```

resources:
documentation: http://docs.ggplot2.org/current/
guide:
http://www.ling.up/6enn.edu/~joseff/avml2012/
cheat sheets:
http://rpubs.com/woobe/ggplot2_ref_part01
http://rpubs.com/woobe/ggplot2_ref_part02