Cheatography

R Programming Cheat Sheet

by raeshmi via cheatography.com/67112/cs/16792/

Util functions

getwd()	
setwd('C://file/path')	
rm(variable_name)	
str(variable_name)	
help.start()	
install.packages("ade4")	
library(ade4)	
detach(package:ade4)	
history()	

DataFrame

d=data.frame(subjectID=1:3,gender=c("M","F","F"),score=c(8,3,6))	A list where all elements are the same length.
rbind(a_data_frame, another_data_frame)	Bind rows
cbind(a_data_frame, another_data_frame)	Bind columns

Strings

x <- (1:15) ^ 2	toString(x)
toupper("I'm Shouting")	tolower("I'm Shouting")
tolower("I'm Shouting")	strsplit(woodchuck, " ", fixed = TRUE)

Data.table

library(data.table)	class(flights)
head(flights)	flights[, .(.N), by = .(origin)]
flights[, head(.SD, 2), by = month]	flights[1:5,sum(arr_delay, dep_delay),]

Vectors

t(a)	transpose
5 * a	scalar multiplication
a+b	summing vector
c(1,0)	unit vectors

Matrices

matrix(1:6,2,3)	
m2=matrix(1:3)	

Vectors

y<-c(5,7,7,8,2,5,6,4)	Numeric vector
x <- c("one","two","three")	Character vector
z <- c(TRUE,TRUE,FALSE)	Logical vector

Lists

cars<-list(c("Toyota", "Nissan", "Honda"), c(150,180,50))	Collection of elements which can be of different types.
cars[[1]]	first row of the list

Descriptive Statistics

summary(mydat)	describe(mydat)
str(mydat)	names(mydat)
par(mfrow=c(2,2))	plot(density(female_dat$science_score))

Functions

hypotenuse(3, 4)	formalArgs(hypotenuse)
normalize(c(1, 3, 6, 10, NA))	f(sqrt(5))

Hypothesis Testing

t.test(x, y)	t-test - difference between means.
prop.test	Test for difference between proportions.
pairwise.t.test	t-test for paired data.
cor.test(sample1,sample2)	Correlation
wilcox.test(data3)	Alternate hypothesis is proved
chisq.test(marks1)	Chi square test
shapiro.test(vnor)	Distribution is normal
aov	ANOVA - Analysis of Variance

Arrays & Matrices

```
(two_d_array <- array( 1:12,
dim = c(4, 3),
dimnames = list(
c("one", "two", "three", "four"),
c("c1", "c2", "c3"))))
dim(two_d_array)
nrow(two_d_array)
ncol(two_d_array)
length(two_d_array)
```

Visualization

barplot(Species)	ggplot(mydata1,aes(x = subject, fill = subject)) + geom_bar()
hist(Sepal.Length)	plot(Sepal.Width)
qqnorm(Sepal.Width)	library(ggplot2)
pie(table(Species))	library(learningr)

Probability

Uniform	u <- runif(2000)
Normal or Gaussian	u <- rnorm(2000,mean=50,sd=3)
Exponential	u <- rexp(2000)
Binomial Distribution	mybinom(k,n,p) * 1000
Poisson Distribution	mypois(lambda, 2)

Matrix Manipulation

det(matrix(c(1,0,0,1),2))	Determinant
solve(m1) %*% m1	Inverse
library(MASS)	ginv(m1)

Statistics - Algorithms

predict(model3,mydat)	Regression
table(predict.glm(modelg,newdata=mydat,type="response")>0.5)	Classification
cl$cluster	Clustering