Data Analysis with R Cheat Sheet

by jbjoker via cheatography.com/181883/cs/37815/

Phases of Data Analysis

Ask	Define the problem you are trying to solve.
Prepare	What data do I need to solve this problem? Do I have access to obtain it?
Process	Clean the data of errors and inaccuracies.
Analyze	Perform calculations to tell a data story. Exploratory Analysis, Statistical modelling
Share	Clear visuals of the data and solution. This includes the reproducible code.
Act	Provide recommendations based on data.

File Manipulation

Get Working Directory	getwd()
Set Working Directory	setwd()
See Directory Contents	dir()
Create Folder	dir.create("tFolder")
Create File	file.create("test.csv")
Copy File	file.copy("test.csv", "-tFolder")
Edit File	myedit(test.R)
Delete File	unlink("test.csv")

Structure & Dimensions

Structure	str(data)
Get # of Rows & Columns	dim(data)
Return # of Rows	nrow(data)
Return # of Cols	ncol(data)
Return 1st 6 Rows	head(data)
Get Class Type	class(data)

Importing Data

Web Scraping	con = url("http://google.com")
	htmlCode = readlines(con)
	close(con)
Remote File	fileUrl <- "https://website.com/data.csv"
	download.file(fileUrl, destfile = "./myData.csv", method = "curl")
Import Data as Table	inData <- read.table("data.csv", sep = " ", header = TRUE)

Applying Functions

Apply a function over an array	**apply**(data,Margin,Function) #1=Rows 2=Cols
Apply a function to each element of list, vector, or DF and return a list	**lapply**(data, Function)
Same as lapply, but returns a vector instead	**sapply**(data, Function)
Apply a function to a subset specified by the FactorList	**tapply**(vector, factorList, Function)

Clean & Test Data

Check for NAs	colSums(is.na(data))
Logical NA Test	all(colSums(is.na(data)) == 0)
Trim Whitespace	trimws(charVector)
Verify Data Type	class(data) **or** str(data)
Find Specific	test[test$someCol %in% c("abcdefg", "hello"),]

String Manipulation

Uppercase	toupper(names(charVector))
Lowercase	tolower(names(charVector))

String Manipulation (cont)

String Split	strsplit(names(charVector), "\\.")
Find & Replace 1st	sub("_", "", names(charVector))
Find & Replace All	gsub("_", "", names(charVector))
Get Location of Value	grep("F", LETTERS)
Get Value from location	grep("F", LETTERS, value=TRUE)
Table Count Instances	table(grepl("F", LETTERS))
Get Substring	substr(charData, 1, 7)
Paste with Space	paste("Test", "Message")
Paste Without Space	paste0("Test", "Message")

Statistics

Statistical Summary	summary(data)
Mean	mean(data)
Standard Deviation	sd(vector)
Variance	var(vector)
Range	range(vector)
Normal Distribution	rnorm(n, mean, sd)
Binomial Distribution	rbinom(n, size, prob)
Poisson Distribution	rpois(n, size)
Uniform Distribution	runif(n, min=0, max=10)
Exponential Distribution	rexp(n)
K-Means Clustering	kmeans(data, centers = 3)
Hierarchical Clustering	hclust(dist(data))

Sponsored by **Readable.com**
Measure your website readability!
https://readable.com