R Cheatsheet

Copyright 2014, Bruce Beauchamp

Contents

[R Cheatsheet 1](#_Toc409940188)

[Control statements 3](#_Toc409940189)

[Iterating through a vector 3](#_Toc409940190)

[data.frames 4](#_Toc409940191)

[Append a row to a data.frame 4](#_Toc409940192)

[Create a data.frame from scratch 4](#_Toc409940193)

[Edit a data.frame in table form 4](#_Toc409940194)

[Get a cell of data from a data.frame 5](#_Toc409940195)

[Get the column names of a data.frame 5](#_Toc409940196)

[Get the number of rows in a data.frame 5](#_Toc409940197)

[Get a Row of Data from a data.frame 5](#_Toc409940198)

[Select Rows which meet criteria 6](#_Toc409940199)

[Select Columns from a data.frame 6](#_Toc409940200)

[Data Types 7](#_Toc409940201)

[Get the type of an object 7](#_Toc409940202)

[Environment 8](#_Toc409940203)

[Change the Working Directory 8](#_Toc409940204)

[Get the Working Directory 8](#_Toc409940205)

[List the Files in the Working Directory 8](#_Toc409940206)

[Load a Function into the Environment 8](#_Toc409940207)

[List the Functions Available in the Environment 8](#_Toc409940208)

[Files and Directories 8](#_Toc409940209)

[Create a directory 8](#_Toc409940210)

[Download a file from the web 8](#_Toc409940211)

[See if a directory or file exists 8](#_Toc409940212)

[Linear Regression 9](#_Toc409940213)

[Confidence interval for a regression line fit 9](#_Toc409940214)

[Generate the Regression line 9](#_Toc409940215)

[Get more information about the linear regression 9](#_Toc409940216)

[Plot a regression line on a plot 9](#_Toc409940217)

[Matrices 10](#_Toc409940218)

[Create a matrix 10](#_Toc409940219)

[Misc 10](#_Toc409940220)

[Printing a variable to the console 10](#_Toc409940221)

[Plotting data 11](#_Toc409940222)

[Scatter plot of y vs x 11](#_Toc409940223)

[Reading Data 12](#_Toc409940224)

[Read a CSV file into a data.frame 12](#_Toc409940225)

[Read from an excel spreadsheet 12](#_Toc409940226)

[Strings 12](#_Toc409940227)

[Concatenate two strings 12](#_Toc409940228)

[Padding a string with zeros 12](#_Toc409940229)

[Vectors 13](#_Toc409940230)

[Creating a vector 13](#_Toc409940231)

[Iterating over a vector 13](#_Toc409940232)

[Length of a vector 13](#_Toc409940233)

[Return valid elements of a vector 13](#_Toc409940234)

# Control statements

## Iterating through a vector

id = 1:3

for (i in id) {

cat(i, "\n")

}

|  |
| --- |
| 1  2  3 |
|  |
| |  | | --- | |  | |

# data.frames

## Append a row to a data.frame

> df

id nobs

1 1 2

> df <- rbind(df, data.frame(id=17, nobs=34))

> df

id nobs

1 1 2

2 17 34

## Create a data.frame from scratch

> df <- data.frame(id=integer(), nobs=integer())

> df <- rbind(df, data.frame(id=1, nobs=2))

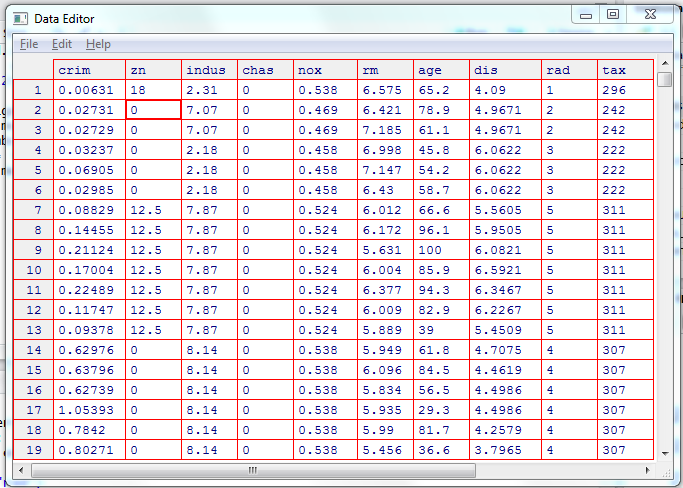
> df

id nobs

1 1 2

## Edit a data.frame in table form

> fix(Boston)



## Get a cell of data from a data.frame

> pollutantData[1,]

Date sulfate nitrate ID

1 2003-01-01 NA NA 1

> pollutantData[1,"ID"]

[1] 1

## Get the column names of a data.frame

> colnames(outcomeData)

## Get the number of rows in a data.frame

|  |
| --- |
| > nrow(data1)  [1] 153 |
|  |
| |  | | --- | |  | |

## Get a Row of Data from a data.frame

|  |
| --- |
| data1[1,]  Ozone Solar.R Wind Temp Month Day  1 41 190 7.4 67 5 1 |
|  |
| |  | | --- | |  | |

## Select Rows which meet criteria

> test <- iris[iris$Species=="virginica", ]

> test

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

101 6.3 3.3 6.0 2.5 virginica

102 5.8 2.7 5.1 1.9 virginica

103 7.1 3.0 5.9 2.1 virginica

## Select Columns from a data.frame

outcomeData <- outcomeData[, c(2, 7, column)]

Here, columns 2, 7, and one additional column are retained

# Data Types

## Get the type of an object

|  |
| --- |
| > x <- 4  > class(x)  [1] "numeric" |
|  |
|  |

# Environment

## Change the Working Directory

> setwd("c:/r/Prog3")

> getwd()

[1] "c:/r/Prog3"

## Get the Working Directory

> getwd()

[1] "C:/Users/Bruce/Documents"

>

## List the Files in the Working Directory

dir()

## Load a Function into the Environment

source(“myCode.R”)

## List the Functions Available in the Environment

ls()

# Files and Directories

## Create a directory

> dir.create("data")

## Download a file from the web

download.file()

## See if a directory or file exists

> file.exists("no such dir")

[1] FALSE

# Linear Regression

## Confidence interval for a regression line fit

> confint(fit)

2.5 % 97.5 %

(Intercept) 33.448457 35.6592247

lstat -1.026148 -0.8739505

## Generate the Regression line

> fit = lm(medv~lstat, Boston)

> fit

Coefficients:

(Intercept) lstat

34.55 -0.95

Note generates the regression line for mdev as a function of lstat

## Get more information about the linear regression

> summary(fit)

Residuals:

Min 1Q Median 3Q Max

-15.168 -3.990 -1.318 2.034 24.500

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 34.55384 0.56263 61.41 <2e-16 \*\*\*

lstat -0.95005 0.03873 -24.53 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

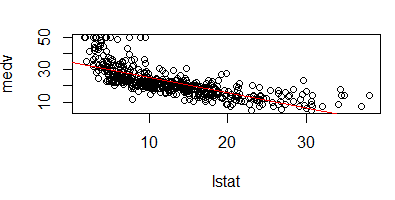
Residual standard error: 6.216 on 504 degrees of freedom

Multiple R-squared: 0.5441, Adjusted R-squared: 0.5432

F-statistic: 601.6 on 1 and 504 DF, p-value: < 2.2e-16

## Plot a regression line on a plot

> abline(fit, col="red")



# Matrices

## Create a matrix

x <- matrix(c(1,2,3,4), 2,2)

> x

[,1] [,2]

[1,] 1 3

[2,] 2 4

# Misc

## Printing a variable to the console

for (i in id) {

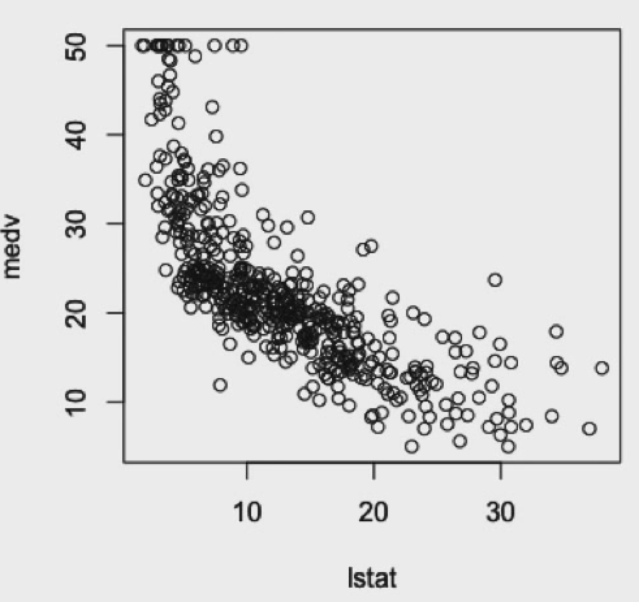
**cat(i, "\n")**

}

# Plotting data

## Scatter plot of y vs x

plot(medv~lstat, Boston)



# Reading Data

## Read a CSV file into a data.frame

> data1 <- read.csv("hw1\_data.csv")

> data1

Ozone Solar.R Wind Temp Month Day

1 41 190 7.4 67 5 1

2 36 118 8.0 72 5 2

3 12 149 12.6 74 5 3

## Read from an excel spreadsheet

library(xlsx)

read.xlsx()

# Strings

## Concatenate two strings

|  |
| --- |
| > paste("a","b", sep = "")  [1] "ab" |
|  |
| |  | | --- | | > |   > paste0("1","2")  [1] "12" |

## Padding a string with zeros

> sprintf("%03d", 17)

[1] "017"

# Vectors

## Creating a vector

> a = 1:2

> a

[1] 1 2

|  |
| --- |
| > x <- c(1,3,2,5)  > x  [1] 1 3 2 5 |
|  |
| |  | | --- | |  | |

## Iterating over a vector

for (i in id) {

writeLines(paste(i))

}

|  |
| --- |
| 1  2 |
|  |
| |  | | --- | |  | |

## Length of a vector

> bad <- ozone\_col[is.na(ozone\_col)]

> length(bad)

[1] 37

## Return valid elements of a vector

|  |
| --- |
| > a  [1] 1 2 NA 4  > a[complete.cases(a)]  [1] 1 2 4 |
|  |
| |  | | --- | |  | |