#Jonathan Marin

#Doing Data Science

#

#1

#a) Log of Positive Number

log(7)

#b)

?log

#Answer log 10

#C)

log(-1)

#Warning message:

# In log(-1) : NaNs produced Log of a negative number is undefined

#d)

sqrt(64)

#2

#a)

vector <- runif(15, min=0, max=100)

mean <- mean(vector)

sd <- sd(vector)

#b)

mean <- 10

sd <- 2

vector2 <- runif(15, min=0, max=100)

mean2 <- mean(vector2)

sd2 <- sd(vector2)

#c)

# These were random samples of 15 between 1-100 (decimcal).

#3

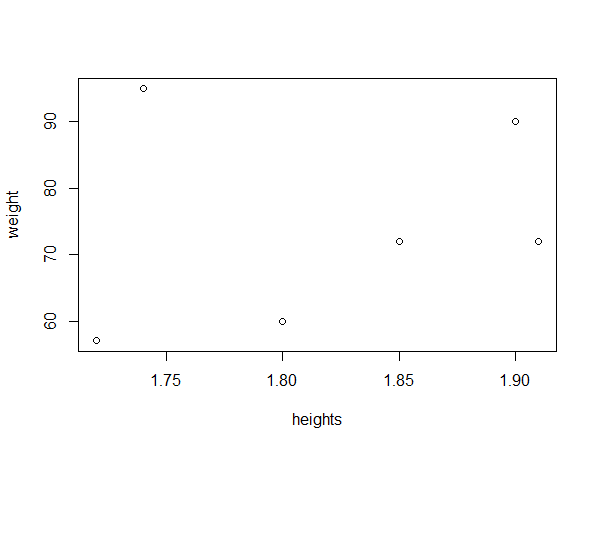
#a) - #c)

weight <- c(60,72,57,90,95,72)

heights <- c(1.8,1.85,1.72,1.90,1.74,1.91)

#d

plot(weight ~ heights)



# As people get taller, you can see that weight increases. There is one person that is short that weighs more than others that seems to be an outlier.

#e

bmi <- weight/(heights \* heights)

#f

meanweight <- mean(weight)

meanweight

#g

meanweighterror <- meanweight - weight

#h

sum(meanweighterror)

#4)

profile <- data.frame("Category" = "Computer Programming" , "Ranking" = 3, stringsAsFactors = FALSE)

profile <- rbind(profile, c("Math",3))

profile <- rbind(profile, c("Statistics", 2))

profile <- rbind(profile, c("Machine Learning",1))

profile <- rbind(profile, c("Domain Expertise",5))

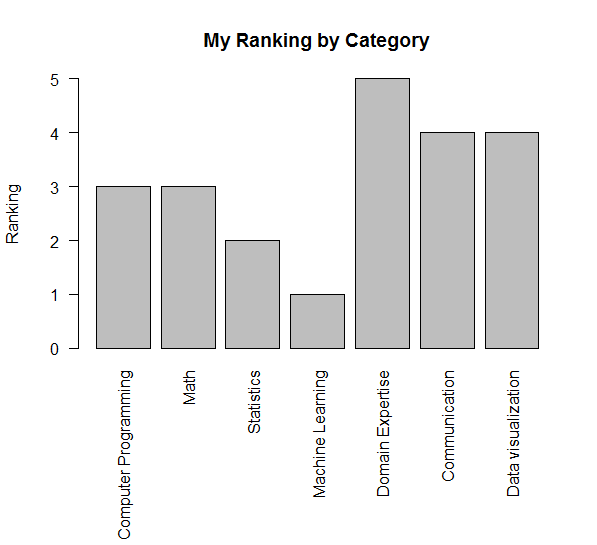
profile <- rbind(profile, c("Communication", 4))

profile <- rbind(profile, c("Data visualization", 4))

profile$Ranking <- as.numeric(profile$Ranking)

op <- par(mar = c(10,4,4,2) + 0.1)

barplot(profile$Ranking, names = profile$Category, horiz = F, ylab = "Ranking", las = 2, main = "My Ranking by Category" )



#5)

library(swirl)

install\_from\_swirl("R Programming")

swirl()

#Module 1

5+7

x <- 5 + 7

x

y <- x - 3

y

c(1.1, 9, 3.14)

z <- c(1.1, 9, 3.14)

?c

z

c(z,555,z)

z \* 2 + 100

my\_sqrt <- sqrt(z-1)

my\_sqrt

my\_div <- z

my\_div

c(1,2,3,4)+c(0,10)

c(1, 2, 3, 4) + c(0, 10, 100)

z \* 2 + 1000

my\_div

#Module 2

getwd()

ls()

x <- 9

ls()

list.files()

?list.files

args()

args(list.files())

args(list.files)

old.dir <- getwd()

dir.create("testdir")

setwd(testdir)

setwd("testdir")

file.create()

file.create("mytest.R")

list.files()

file.exists("mytest.R")

info("mytest.R")

file.info("mytest.R")

file.info("mytest.R")

file.rename("mytest.R","mytest2.R")

file.copy("mytest2.R","mytest3.R")

file.path("mytest3.R")

file.path

file.path("folder1","folder2")

?dir.create

dir.create()

dir.create("testdir2", file.path("testdir3"))

dir.create(file.path('testdir2', 'testdir3'), recursive = TRUE)

setwd(old.dir)

#Module3

1:20

pi:10

15:1

?`:`

seq(1, 20)

seq(0, 10, by=0.5)

seq(5, 10, length=30)

my\_seq <- seq(5, 10, length=30)

length(my\_seq)

1:length(my\_seq)

seq(along.with = my\_seq)

seq\_along(my\_seq)

rep(0, times=40)

rep(c(0, 1, 2), times = 10)

rep(c(0, 1, 2), each = 10)

#Module 4

num\_vect <- c(0.5, 55, -10, 6)

tf <- (num\_vect < 1)

tf <- num\_vect < 1

tf

num\_vect = 6

my\_char <- c("My", "name", "is")

my\_char

paste(my\_char, collapse = " ")

my\_name <- c(my\_char, "Johnny")

my\_name

paste(my\_name, collapse = " ")

paste("Hello", "world!", sep = " ")

paste(c("X", "Y", "Z"), sep = " ")

paste(1:3, c("X", "Y", "Z"), sep = " ")

paste(1:3, c("X", "Y", "Z"), sep = "")

paste(LETTERS, 1:4, sep = "-")

#Module 5

x <- c(44, NA, 5, NA)

x \* 3

y <- rnorm(1000)

z <- rep(NA, 1000)

my\_data <- sample(c(y, z), 100)

my\_na <- is.na(my\_data)

my\_na

my\_data == is.na

sum(my\_na)

my\_data

0 / 0

Inf - Inf

#Module 6

x

x[1:10]

x[is.na(x)]

y <- x[!is.na(x)]

y

c(x[3], x[5], x[7])

x[c(3, 5, 7)]

x[0]

x[3000]

x[c(-2, -10)]

play()

x[c(2, 10)]

nxt()

x[-c(2, 10)]

vect <- c(foo = 11, bar = 2, norf = NA)

vect

names(vect)

vect2 <- c(11, 2, NA)

names(vect2) <- c("foo", "bar", "norf")

identical(vect, vect2)

vect["bar"]

vect[c("foo", "bar")]

#Module 7

my\_vector <- c(1:20)

my\_vector <- 1:20

my\_vector

dim(my\_vector)

length(my\_vector)

dim(my\_vector) <- c(4, 5)

dim(my\_vector)

attributes(my\_vector)

my\_vector

class(my\_vector)

my\_matrix <- my\_vector

?matrix

my\_matrix2 <- as.matrix(1:20)

play()

my\_matrix2

matrix(1:20, nrow = 4, ncol = 5, byrows = TRUE)

matrix(1:20, nrow = 4, ncol = 5, byrow = TRUE)

matrix(1:20, nrow = 4, ncol = 5, byrow = FALSE)

nxt

nxt()

matrix(1:20, nrow = 4, ncol = 5, byrow = FALSE)

my\_matrix2 <- matrix(1:20, nrow = 4, ncol = 5, byrow = FALSE)

identical(my\_matrix, my\_matrix2)

patients <- c("Bill", "Gina", "Kelly", "Sean")

cbind(patients, my\_matrix)

my\_data <- data.frame(patients, my\_matrix)

my\_data

class(my\_data)

cnames <- c("patient", "age", "weight", "bp", "rating", "test")

colnames(my\_data)

colnames(my\_data) <- cnames

my\_data