```
# 6306_Doing_Data_Science
# Fall 2017 - Sept, 2017
# Assignment 1
# ENCODING: ISO-8859-1 (System Default)
# Bruce Granger
#------------------------------------------------------------------------------
# SETUP
#------------------------------------------------------------------------------
# REDIRECT SCREEN OUTPUT TO FILE, sink() WILL BE UTILIZED
sink("N:/SMU/6306_Doing_Data_Science_Fall2017/Assignments/Assignment_1_SWIRL_2.txt")
#------------------------------------------------------------------------------
# DEFAULT WORKING DIRECTORY
# "N:/SMU/6306_Doing_Data_Science_Fall2017"
working_directory_startup<-getwd()

# ASSIGNED WORKING DIRECTOR
assigned_working_directory <- "N:/SMU/6306_Doing_Data_Science_Fall2017"
setwd(assigned_working_directory)
getwd()
#------------------------------------------------------------------------------
# ENVIRONMENT VARIABLES
#------------------------------------------------------------------------------
n<-64
#------------------------------------------------------------------------------
# 1) BACIS MATH
#------------------------------------------------------------------------------
q1a<- "The log of a positive number?"
q1a
q1a_log<-log(n)
q1a_log
#------------------------------------------------------------------------------
q1b<- "What is the default base for the log function? Calculate the log of your previous number with a different base."
q1b
q1b1<- "The default base for the log function in R is the natural log"
q1b2<-log(n,10)
q1b2
q1b3<-log10(n)
q1b3
#------------------------------------------------------------------------------
q1c<- "The log of a negative number?"
q1c
q1c1<- log(-1)
q1c2<- "This produces a return of NaN, which means 'Not a Number'."
#------------------------------------------------------------------------------
q1c<- "The square-root of a positive number."
q1c
q1c1<- sqrt(n)
q1c1
#------------------------------------------------------------------------------
# 2) RANDOM NUMBER GENERATION
#------------------------------------------------------------------------------
q2a<-"Create a vector of 15 standard normal random variables. Calculate its mean and SD"
q2a
```

```
q2a1<- rnorm(15)
q2a1
q2a1_mean<-mean(q2a1)
q2a1_mean
q2a1_sd<-sd(q2a1)
q2a1_sd
#---------------------------------------------------------------------------
q2b<-"Change the mean to 10 and the SD to 2 and recalculate the vector of 15 random normal variables. Calculate its mean and SD."
q2b
q2b1<-rnorm(15, mean= 10, sd=2)
q2b1
q2b1_mean<-mean(q2b1)
q2b1_mean
q2b1_sd<-sd(q2b1)
q2b1_sd
#---------------------------------------------------------------------------
q2c<-"Why are the means and SD not exactly the same as the means and SDs specified in the function?"
q2c1<-"Since rnorm was used to select the fifteen random normal variables, the small selection sample doesn't mean the random sample will match the values"
#---------------------------------------------------------------------------
# 3) VECTOR OPERATIONS
#---------------------------------------------------------------------------
q3a<-"The weights of 6 individuals in kg are 60, 72, 57, 90, 95, 72."
q3a
q3b<-"Their heights (in m) are 1.80, 1.85, 1.72, 1.90, 1.74, 1.91."
q3b
q3c<-"Enter these vectors into R."
q3c_weight_kg<- c(60, 72, 57, 90, 95, 72)
q3c_weight_kg
q3c_hight_m<-c(1.80, 1.85, 1.72, 1.90, 1.74, 1.91)
q3c_hight_m
q3d<-"Create a scatterplot of weight vs. height. Interpret the scatterplot."
q3d
plot(q3c_weight_kg,q3c_hight_m, main='Individual Weight to Height',
    xlab = 'Weight', ylab='Height', pch=19)
q3d_interpretation<- "Generally, height and weight are positively correlated, meaning as one increases, so does the other."
q3e<-"Calculate the BMI for each individual."
q3e
q3c_bmi<-q3c_weight_kg/((q3c_hight_m)^2)
q3c_bmi
q3f<-"Calculate the mean for weight."
q3c_weight_mean<-mean(q3c_weight_kg)
q3c_weight_mean
q3g<-"Subtract the mean from each value of weight."
q3g_variance<-q3c_weight_kg-q3c_weight_mean
q3g_variance
q3h<-"Sum the result."
q3h_sum_of_variance<-sum(q3g_variance)
q3h_sum_of_variance
#---------------------------------------------------------------------------
# 4) DATA SCIENCE PROFILE
#---------------------------------------------------------------------------
```

```
data_science_skill<-c('Computer Programming', "Math", "Statistics", "Machine Learning", "Domain Expertise",
"Communications", "Presentation Skills", "Data Vizulation")
data_science_skill
data_science_skill_rank<-c(3.5, 2.5, 2, 1,4,4,3,3)
data_science_skill_rank
data_science_profile<-data.frame(data_science_skill,data_science_skill_rank, stringsAsFactors = FALSE)
data_science_profile
#---------------------------------------------------------------------------
# CLOSEOUT SINK
#---------------------------------------------------------------------------
sink()
#---------------------------------------------------------------------------
```

```
[1] "N:/SMU/6306_Doing_Data_Science_Fall2017"
[1] "The log of a positive number?"
[1] 4.158883
[1] "What is the default base for the log function? Calculate the log of your previous number with a different base."
[1] 1.80618
[1] 1.80618
[1] "The log of a negative number?"
[1] "The square-root of a positive number."
[1] 8
[1] "Create a vector of 15 standard normal random variables. Calculate its mean and SD"
 [1] -0.375514937  0.233937702  1.100984021 -0.531557625 -2.530671644  1.597177354 -2.025902046  0.108743774
-0.421409765  0.025294698
[11] -1.609565147 -1.147868147 -0.003189261 -1.544647659 -0.779116314
[1] -0.526887
[1] 1.12368
[1] "Change the mean to 10 and the SD to 2 and recalculate the vector of 15 random normal variables. Calculate its
mean and SD."
 [1]  8.433871 10.741288 10.105947 12.516105  9.623386  7.190125  8.983868  9.289397  7.782870 10.859058
11.057104  7.859663  7.147472
[14] 13.286241  9.443670
[1] 9.621338
[1] 1.839374
[1] "The weights of 6 individuals in kg are 60, 72, 57, 90, 95, 72."
[1] "Their heights (in m) are 1.80, 1.85, 1.72, 1.90, 1.74, 1.91."
[1] 60 72 57 90 95 72
[1] 1.80 1.85 1.72 1.90 1.74 1.91
[1] "Create a scatterplot of weight vs. height. Interpret the scatterplot."
[1] "Calculate the BMI for each individual."
[1] 18.51852 21.03725 19.26717 24.93075 31.37799 19.73630
[1] 74.33333
[1] -14.333333  -2.333333 -17.333333  15.666667  20.666667  -2.333333
[1] 2.842171e-14
[1] "Computer Programming" "Math"                 "Statistics"           "Machine Learning"     "Domain Expertise"
[6] "Communications"       "Presentation Skills"  "Data Vizulation"
[1] 3.5 2.5 2.0 1.0 4.0 4.0 3.0 3.0
   data_science_skill data_science_skill_rank
1 Computer Programming                    3.5
2                Math                    2.5
3          Statistics                    2.0
4    Machine Learning                    1.0
5    Domain Expertise                    4.0
6      Communications                    4.0
7  Presentation Skills                   3.0
8     Data Vizulation                    3.0
```

```
# 6306_Doing_Data_Science
# Fall 2017 - Sept, 2017
# Assignment 1 - SWIRL
# Bruce Granger
#-------------------------------------------------------------------------------
# REFERENCES
# http://swirlstats.com/scn/
# https://github.com/swirldev/swirl_courses
#-------------------------------------------------------------------------------
# REDIRECT SCREEN OUTPUT TO FILE, sink() WILL BE UTILIZED
sink("N:/SMU/6306_Doing_Data_Science_Fall2017/Assignments/Assignment_1_SWIRL_2.txt")
#-------------------------------------------------------------------------------
# INSTALL SWRIL PACKAGE
#install.packages("swirl")
library("swirl")
swirl()
#-------------------------------------------------------------------------------
# MODULES
# 1: Basic Building Blocks
#-------------------------------------------------------------------------------
# TURN OFF SCREEN REDIRECT OUTPUT TO FILE
sink()
```

1: R Programming Basic Building Blocks
2: No. Let me start something new.

Enter an item from the menu, or 0 to exit

1: R Programming
2: Take me to the swirl course repository!

Enter an item from the menu, or 0 to exit
Enter an item from the menu, or 0 to exit

1: Basic Building Blocks 2: Workspace and Files 3: Sequences of Numbers 4: Vectors
5: Missing Values 6: Subsetting Vectors 7: Matrices and Data Frames 8: Logic
9: Functions 10: lapply and sapply 11: vapply and tapply 12: Looking at Data
13: Simulation 14: Dates and Times 15: Base Graphics

1: R Programming
2: Take me to the swirl course repository!

1: R Programming Basic Building Blocks
2: No. Let me start something new.

Enter an item from the menu, or 0 to exit

[1] 12

|
|========== | 8%

|
|============= | 11%

|
|================= | 13%

|
|==================== | 16%

|
|======================== | 18%

|
|============================ | 21%

|
|================================ | 24%[1] 12

```
  |
  |==================================                                    | 26%

  |
  |====================================                                  | 29%[1]
9



  |
  |======================================                                | 32%

  |
  |========================================                              |
34%

  |
  |==========================================                            |
37%

  |
  |============================================
|  39%[1] 1.10 9.00 3.14


  |
  |==============================================
|  42%[1]   1.10   9.00   3.14 555.00   1.10   9.00   3.14


  |
  |================================================
|  45%[1] 102.20 118.00 106.28


  |
  |==================================================
|  47%

  |
  |====================================================
|  50%

  |
  |======================================================
|  53%

  |
  |========================================================
|  55%
1: a vector of length 3
2: a single number (i.e a vector of length 1)
3: a vector of length 0 (i.e. an empty vector)

Enter an item from the menu, or 0 to exit
```

Enter an item from the menu, or 0 to exit

```
  |
  |=======================================================================
|  58%[1] 0.3162278 2.8284271 1.4628739


  |
  |=======================================================================
|  61%

  |
  |=======================================================================
|  63%

  |
  |=======================================================================
|  66%
```

1: The first element of my_div is equal to the first element of z divided by the first element of my_sqrt, and so on...
2: my_div is a single number (i.e a vector of length 1)
3: my_div is undefined

Enter an item from the menu, or 0 to exit

```
  |

|=======================================================================
===                     |  68%[1] 3.478505 3.181981 2.146460


  |

|=======================================================================
======                      |  71%

  |

|=======================================================================
=========                      |  74%

  |

|=======================================================================
=============                      |  76%

  |

|=======================================================================
================                      |  79%[1]  1 12  3 14


  |
```

|==========| 82%

|

|==========| 84%[1] 1 12 103 4

|

|==========| 87%

|

|==========| 89%

|

|==========| 92%[1] 1002.20 1018.00 1006.28

|

|==========| 95%

|

|==========| 97%[1] 3.478505 3.181981 2.146460

|

|==========|100%
1: No
2: Yes

1: No
2: Yes