```
# UCLA Extension - Introduction to Data Science
#
# Homework #1 Solutions
#
#
# (c) Copyright 2016-2019 - AMULET Analytics
# ---------------------------------------------------------------------

# ---------------------------------------------------------------------
# Question 1
# ---------------------------------------------------------------------

mat <- matrix(1:12, nrow=3, ncol=4, byrow=TRUE)

mat <- rbind(c(9,9,9,9), mat)

mat <- cbind(c(8,8,8,8), mat)

mat
#     [,1] [,2] [,3] [,4] [,5]
#[1,]    8    9    9    9    9
#[2,]    8    1    2    3    4
#[3,]    8    5    6    7    8
#[4,]    8    9   10   11   12

# ---------------------------------------------------------------------
# Question 2
# ---------------------------------------------------------------------

lst <- list(names=c("Ellen", "Catherine", "Stephen"), grades=c(90L, 95L, 99L), attendance=matrix(c(TRUE, TRUE, TRUE,
FALSE, FALSE, TRUE), nrow=2))
class(lst$names)           # character
class(lst$grades)          # integer
class(lst$attendance)      # matrix

lst$names             # Display students in class
#[1] "Ellen"     "Catherine" "Stephen"

lst$grades[3]         # Display Stephen's grade
#[1] 99

lst$attendance[,2]
#[1]  TRUE FALSE

# ---------------------------------------------------------------------
# Question 3
# ---------------------------------------------------------------------

# variable gender with 20 "male" entries and
# 30 "female" entries
gender <- c(rep("male",20), rep("female", 30))
gender <- factor(gender)
# stores gender as 20 1s and 30 2s and associates
# 1=female, 2=male internally (alphabetically)
# R now treats gender as a nominal variable
summary(gender)
#female   male
#    30     20

# ---------------------------------------------------------------------
# Question 4
# ---------------------------------------------------------------------

data(airquality)
nrow(subset(airquality, is.na(Ozone)))
#[1] 37              # Answer: B

# ---------------------------------------------------------------------
# Question 5
# ---------------------------------------------------------------------

# subset() removes the NAs
b <- subset(airquality, Ozone>31 & Temp>90)
mean(b$Solar.R)
#[1] 212.8           # Answer: D

# ---------------------------------------------------------------------
# Question 6
# ---------------------------------------------------------------------

# Data frames

aq <- airquality

aq$hotcold <- ifelse(aq$Temp > median(aq$Temp), "hot", "cold")
```

```
head(aq)
#  Ozone Solar.R Wind Temp Month Day hotcold
#1    41     190  7.4   67     5   1    cold
#2    36     118  8.0   72     5   2    cold
#3    12     149 12.6   74     5   3    cold
#4    18     313 11.5   62     5   4    cold
#5    NA      NA 14.3   56     5   5    cold
#6    28      NA 14.9   66     5   6    cold

tail(aq)
#    Ozone Solar.R Wind Temp Month Day hotcold
#148    14      20 16.6   63     9  25    cold
#149    30     193  6.9   70     9  26    cold
#150    NA     145 13.2   77     9  27    cold
#151    14     191 14.3   75     9  28    cold
#152    18     131  8.0   76     9  29    cold
#153    20     223 11.5   68     9  30    cold

# ---------------------------------------------------------------
# Question 7
# ---------------------------------------------------------------

# Fizz Buzz

# FizzBuzz test solution

#Based on a traditional English children's game
#Print the numbers 1..100
#For multiples of 3, print "Fizz" instead of the number
#For multiples of 5, print "Buzz" instead of the number
#For multiples of 3 and 5, print "FizzBuzz" instead of the number


# Solution #1 -----------------------------------------------------

fizzbuzz = function(i, multiples = c(3,5), text = c("Fizz", "Buzz")) {
  words = text[i %% multiples == 0]
  if (length(words) == 0)
    as.character(i)
  else
    paste(words, collapse = "")
}

sapply(1:100,fizzbuzz)


# Solution #2 -----------------------------------------------------
# A little fun extending the problem adding 7 using same fcn

sapply(1:200,function(x) fizzbuzz(x, c(3,5,7), c("Fizz","Buzz","Bang")))


# Solution #3 -----------------------------------------------------
f = "Fizz"
b = "Buzz"
for (i in c(1:100))
  if (i %% 15 == 0) {
    print(paste(f, b))
  } else if (i %% 5 == 0) {
    print(b)
  } else if (i %% 3 == 0) {
    print(f)
  } else {
    print(i)
  }


# Solution #4 -----------------------------------------------------

lapply(c(1:100), function(x){
  if(x%%15==0) {
    print('FizzBuzz')
  }
  else if(x%%3==0) {
    print('Fizz')
  }
  else if(x%%5==0) {
    print('Buzz')
  }
  else {
    print(x)
  }
})
```

```
# Solution #5 ------------------------------------------------------

for(i in 1:100){
  if(i%%3==0) if(i%%5==0) print("FizzBuzz") else print("Fizz") else
    if(i%%5==0) print("Buzz") else
      print(i)
}


# Solution #6 ------------------------------------------------------

ec <- 1:100
ec[which(ec%%5==0)][which(ec%%3==0)] <- "fizzbuzz"
ec[which(as.numeric(ec)%%3==0)] <- "fizz"
ec[which(as.numeric(ec)%%5==0)] <- "buzz"
print(ec)


# ------------------------------------------------------------------
# Question 8
# ------------------------------------------------------------------

mat1 <- matrix(rep(seq(4), 4), ncol = 4)
mat1
#[,1] [,2] [,3] [,4]
#[1,]    1    1    1    1
#[2,]    2    2    2    2
#[3,]    3    3    3    3
#[4,]    4    4    4    4

# Sum of rows (not part of answer)
apply(mat1, 1, sum)
#[1]  4  8 12 16

# METHOD 1
#using a user defined function
sum.plus.2 <- function(x){
  sum(x) + 2
}

# METHOD 2
# using the sum.plus.2 function on the rows of mat1
apply(mat1, 1, sum.plus.2)
#[1]  6 10 14 18

# Using a anonymous function
# the function can be defined inside the apply function
# note the lack of curly brackets
apply(mat1, 1, function(x) sum(x) + 2)
#[1]  6 10 14 18

# ------------------------------------------------------------------
# Question 9
# ------------------------------------------------------------------

set.seed(314)
subset_states <- sample(state.name,10)
subset_states
#[1] "California"  "Indiana"     "Oregon"      "Hawaii"      "Georgia"     "Wisconsin"
#[7] "Washington"  "Kansas"      "Mississippi" "New Mexico"

subset_states <- subset_states[order(subset_states)]
subset_states
#[1] "California"  "Georgia"     "Hawaii"      "Indiana"     "Kansas"      "Mississippi"
#[7] "New Mexico"  "Oregon"      "Washington"  "Wisconsin"

# ------------------------------------------------------------------
# Question 10
# ------------------------------------------------------------------

xct <- as.POSIXct("1969-07-20 20:18", tz="UTC")

as.numeric((Sys.time() - xct)/365)
#[1] 46.59523
```