##### IS 470 Lab 1--------------------------------------------------------------

#### Overview of R and RStudio--------------------------------------------------

#R is a language and environment for statistical computing and graphics.

#R provides a wide variety of statistical methods for data mining and modeling,

#such as linear and nonlinear regression, classical statistical tests,

#time-series analysis, classification, clustering and association rule mining.

#RStudio is an integrated development environment (IDE) for R. It includes a

#console, syntax-highlighting editor that supports direct code execution, as

#well as tools for plotting, history, debugging and workspace management.

#### Basic operators and data structures in R-------------------------------------

# You are required to complete all the tasks by following the lab video.

# Please name your R script as Lab1\_FirstLastName.R, and submit on BeachBoard.

### ------------------------------------------------------------------------------

# 1. Assignment operator

x=2

print(x)

#or

x<- 1

x

x="hello"

x

# 2. Use the colon (:) operator to create integer sequences.

x=1:20

x

#Vector: The fundamental R data structure is the vector, which stores an ordered

#set of values called elements. A vector can contain any number of elements, but

#all of the elements must be of the same type of values. For instance, a vector

#cannot contain both numbers and text. vectors can be created by using the c()

#combine function.

# 3. Create vectors

subject\_name = c("John Doe","Jane Doe", "Steve Grave")

subject\_name

temperature = c(98.1, 98.6, 101.4)

temperature

flu\_status = c(FALSE, FALSE, TRUE)

flu\_status

# 4. Obtain the body temperature for patient Jane Doe

temperature[2]

# 5. A range of values can be obtained using the (:) colon operator

temperature[2:3]

# 6. Exclude Jane Doe's temperature data

temperature[-2]

#Factor: A factor is a special case of vector that is solely used to represent

#categorical variables.

# 7. Create a factor from a character vector

gender = factor(c("Male","Female","Male"))

gender

#factor is useful categorical variable;Categorical variables are stored as factors

# 8. Create a factor for the blood type

blood\_type = factor (c("O","AB","A"),levels = c("A","B","AB","O"))

blood\_type

#Data frames: A structure analogous to a spreadsheet or database, since it has

#both rows and columns of data.

# 9. Create a data frame for our patient dataset, by using the patient data vectors we created.

pt\_data = data.frame(subject\_name,temperature, flu\_status,gender,blood\_type)

pt\_data

# 10. Obtain the subject\_name vector from the created data frame.

pt\_data$subject\_name

pt\_data[,1]#this is an alternative, the first value is row, second is column.

# 11. Extract the first and second columns from the data frame.

pt\_data[,c("subject\_name","temperature")]#method 1

#method 2

pt\_data[,c(1,2)]

# 12. Extract the value in the first row and second column of the patient data frame.

pt\_data[1,2]

# 13. Extract the first column from data frame.

pt\_data[,1]

# 14. Extract the first row from data frame.

pt\_data[1,]

# 15. Extract everything from data frame.

pt\_data

#or

pt\_data[,]

# 16. Exclude the first column from data frame.

pt\_data[,-1]

# 17. Exclude the first row from data frame.

pt\_data[-1,]