**Student name: Trevor Cardoza SMC username: TMC12**

Instructions: Fill out your full name and SMC username above. Answer each question in its respective following space on this answer sheet. **Give a sufficient yet succinct answer for each question, and show figures / tables / graphs on this answer sheet when appropriate or explicitly requested.** After completion, rename and save this file as **“Assignment 2\_BUSAD 040\_*your full name*” in either MS Word (.docx or .doc) or PDF (.pdf) format**, then upload and submit it on Moodle by the due date. No other file formats will be accepted. Late submission will not be accepted.

Note: This is the **only file (answer sheet)** you need to submit on Moodle after completion. No other accompanying files are needed.

**Total possible points: 70 points**

1. **Data structures and subsetting in R (35 points in total)**

For the following questions, you may work out and run the R commands / codes in RStudio either in a R Script file, or in the console. You should **copy and paste the original R commands as well as the run results from the RStudio console for each question** in its respective following space on this answer sheet.

1.1 Create three vectors as follows using the **c( )** combine function. (12 points in total, 4 points each)

a. The first vector is a numeric vector containing the following five real numbers. Assign the vector to a new variable named “**numbers**”. Type the variable name and run it to get the vector content.

**10, 20, 30, 40, 50**

numbers <- c(10,20,30,40,50)

OUTPUT:

[1] 10 20 30 40 50

b. The second vector is a character vector containing the following five character strings. Assign the vector to a new variable named “**strings**”. Type the variable name and run it to get the vector content.

**"ten", "twenty", "thirty", "forty", "fifty"**

strings <- c(‘ten’,’twenty’,’thirty’,’forty’,’fifty’)

OUTPUT:

[1] "ten" "twenty" "thirty" "forty" "fifty"

c. The third vector is a character vector containing the following five logical values. Assign the vector to a new variable named “**logical**”. Type the variable name and run it to get the vector content.

**TRUE, FALSE, TRUE, FALSE, TRUE**

logical <- c(TRUE,FALSE,TRUE,FALSE,TRUE)

OUTPUT:

[1] TRUE FALSE TRUE FALSE TRUE

1.2 Combine the three vectors into a data frame using the **data.frame( )** function and assign it to the new variable named “**mydata**”. Type the variable name and run it to get the data frame content. (8 points)

mydata <- data.frame(numbers,strings,logical)

OUTPUT:

numbers strings logical

1 10 ten TRUE

2 20 twenty FALSE

3 30 thirty TRUE

4 40 forty FALSE

5 50 fifty TRUE

1.3 Select a smaller portion as data frame or content of a row/column from the data frame **mydata** using brackets [ ] and indexes/variable names, or the $ operator. (10 points in total)

a. Select every element in the **2nd column** of the data frame and display the result. (2 points)

mydata[2]

OUTPUT:

strings

1 ten

2 twenty

3 thirty

4 forty

5 fifty

b. Select every element from the **1st row** (inclusive) **through 4th row** (inclusive) of the data frame and display the result as a smaller data frame. (3 points)

mydata[1:4,]

OUTPUT:

numbers strings logical

1 10 ten TRUE

2 20 twenty FALSE

3 30 thirty TRUE

4 40 forty FALSE

c. Select the element at the **intersection** of the **5th row** and the **2nd column** of the data frame and display the result. (3 points)

mydata[5,2]

OUTPUT:

[1] fifty

d. Retrieve the **content** of the **3rd column** directly as a logical vector using the $ operator and the column variable name. (2 points)

c(mydata[c(3)])

OUTPUT:

$logical

[1] TRUE FALSE TRUE FALSE TRUE

1.4 Run the structure function **str( )** on the data frame **mydata**. According to the function result, what is the data structure for **each of the three individual variables/columns** of the data frame? Among the three variables, which variable(s) is/are factor(s) (i.e., categorical variable(s))? (5 points)

str(mydata)

OUTPUT:

'data.frame': 5 obs. of 3 variables:

$ numbers: num 10 20 30 40 50

$ strings: Factor w/ 5 levels "fifty","forty",..: 3 5 4 2 1

$ logical: logi TRUE FALSE TRUE FALSE TRUE

Data Structures are int, string and Boolean/Logical respectively

All three are Categorical Variables

Strings is a Factor

1. **R tables and graphs (35 points in total)**

Use the accompanying **2012Networks.CSV** data file for Assignment 2 to complete the following questions.

Nielsen Media Research provided the list of the 25 top-rated single shows in television history (The World Almanac, 2012). The data in 2012Networks.CSV show the television network that produced each of these 25 top-rated shows.

2.1 Import/read the provided 2012Networks.CSV data into a variable named “**networks**” in RStudio using either the **read.csv( )** or **read.table( )** function. Remember to specify the arguments for **header** and **sep** correctly in the function. Type in and run the variable name **networks** to show the imported result in console. **Copy, paste and show both the R commands and imported result in the following space.** (5 points)

networks <- read.csv('2012Networks.CSV',header = TRUE,sep = ',')

OUTPUT:

Network

1 CBS

2 CBS

3 ABC

4 CBS

5 NBC

6 CBS

7 NBC

8 NBC

9 NBC

10 CBS

11 NBC

12 NBC

13 ABC

14 CBS

15 CBS

16 FOX

17 NBC

18 ABC

19 ABC

20 ABC

21 CBS

22 ABC

23 NBC

24 NBC

25 CBS

2.2 Check the data structure of the variable **networks** using the **str( )** function. **Show the R command and result in the following space.** What is the data structure for this **networks** variable? And what is the data structure for the “**Network**” **column/variable** in the **networks** variable? (5 points)

str(networks)

OUTPUT:

'data.frame': 25 obs. of 1 variable:

$ Network: Factor w/ 4 levels "ABC","CBS","FOX",..: 2 2 1 2 4 2 4 4 4 2 .

Data structure is String

2.3 Construct a frequency distribution using the **table( )** function and assign it to a variable named “**freq.networks**”. **Show the R commands and table result in the following space.** (5 points)

freq.networks <- table(networks)

OUTPUT:

networks

ABC CBS FOX NBC

6 9 1 9

2.4 Construct a **percent** (not relative) frequency distribution using the **prop.table( )** function and assign it to a variable named “**percent.freq.networks**”. **Show the R commands and table result in the following space.** (5 points)

percent.freq.networks <- prop.table(freq.networks)\*100

OUTPUT:

networks

ABC CBS FOX NBC

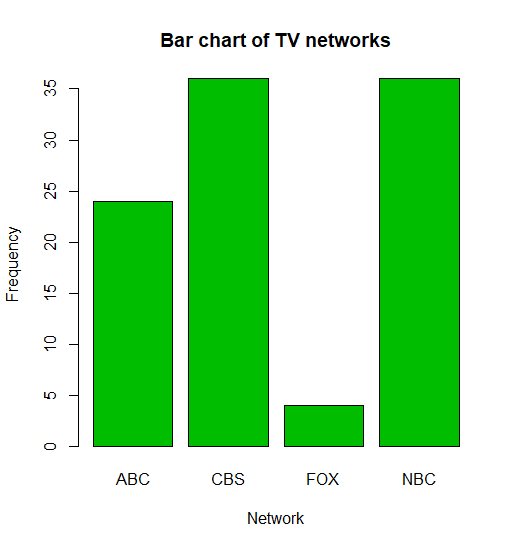
24 36 4 36

2.5 Construct a bar chart for the frequency distribution using the **barplot( )** function. In the bar chart, make the main title as “Bar chart of TV networks”, x axis label as “Network”, y axis label as “Frequency”, and the color for the bars as green. **Show the R command and bar chart in the following space.** (10 points)

(Hint: to save the bar chart image, in the bottom right pane in RStudio, go to the Plots tab – Export – Save as Image, choose the image format as PNG or JPEG, give a file name, choose a directory/folder to save the image, then click the Save button. Then copy and paste the saved image in the following space.)

barplot(percent.freq.networks,main = "Bar chart of TV networks", col = '#00bd00', xlab = 'Network', ylab = 'Frequency')

OUTPUT:



2.6 Which network or networks have done the best in terms of presenting top-rated television shows? Compare the performance of ABC, CBS, and NBC. (5 points)

CBS and NBC are tied at 36% with ABC in second with 24% and FOX last at 4%