Worksheet-4 in R

Worksheet for R Programming

Instructions:

- Use RStudio or the RStudio Cloud accomplish this worksheet.

- Save the R script as *RWorksheet_lastname#4.R*.
- On your own *GitHub repository*, push the R script, the Rmd file, as well as this pdf worksheet to the repo you have created before.
- Do not forget to comment your Git repo on our VLE
- Accomplish this worksheet by answering the questions being asked and writing the code manually.

1. The table below shows the data about shoe size and height. Create a data frame..

Shoe size	Height	Gender	Shoe size	Height	Gender
6.5	66.0	F	13.0	77.0	M
9.0	68.0	F	11.5	72.0	M
8.5	64.5	F	8.5	59.0	F
8.5	65.0	F	5.0	62.0	F
10.5	70.0	M	10.0	72.0	M
7.0	64.0	F	6.5	66.0	F
9.5	70.0	F	7.5	64.0	F
9.0	71.0	F	8.5	67.0	M
13.0	72.0	M	10.5	73.0	M
7.5	64.0	F	8.5	69.0	F
10.5	74.5	M	10.5	72.0	M
8.5	67.0	F	11.0	70.0	M
12.0	71.0	M	9.0	69.0	M
10.5	71.0	M	13.0	70.0	M

a. Describe the data.

Data for estimating shoe sizes was collected from a combination of male and female respondents. I added a two to the end of the names of the other three variables because I believed it might confuse the R program because there are three variables with the same name. I made an effort to adhere to the variable names of the provided data, but the end of the other three variable names all had a 1 in them.

b. Find the mean of shoe size and height of the respondents.
Copy the codes and results.

```
#Gender Male Shoe_size and Height mean.

data1 <- subset(data_frame[1:14, 1:3])
data1
male_only <- data1[data_frame$Gender == 'M',]
male_only
mean_male <- mean(male_only$Shoe_size)
mean_male
height_male <- mean(male_only$Height)
height_male

#Gender Male Shoe_size2 and Height2 mean.
data2 <- subset(data_frame[1:14, 4:6])
data2
male_only2 <- data2[data_frame$Gender2 == 'M',]
male_only2
mean_male2 <- mean(male_only2$Shoe_size2)
mean_male2
height_male2 <- mean(male_only2$Height2)
height_male2

#Gender Female Shoe_size and Height mean.
data3 <- subset(data_frame[1:14, 1:3])
data3
female_only3 <- data3[data_frame$Gender == 'F',]
female_only3
mean_female3 <- mean(female_only3$Shoe_size)
mean_female3
height_female3 <- mean(female_only3$Height)
height_female3

#Gender Female Shoe_size2 and Height2 mean
data4 <- subset(data_frame[1:14, 4:6])
data4
female_only4 <- data4[data_frame$Gender2 == 'F',]
female_only4

mean_female4 <- mean(female_only4$Shoe_size2)
mean_female4
height_female4 <- mean(female_only4$Height2)
height_female4
```

c. Is there a relationship between shoe size and height? Why?

The first three columns, the average shoe size for male respondents is 11.3 and the height is 71.7. For the female respondents the average shoe size is 8.222222 and the height is 66.61111. For the last three columns, the average shoe size for male respondents is 10.77778 and the height is 71.33333. For the female respondents the average shoe size is 7.2 and the height is 64. The relationship of shoe size and height for the first three columns is that the male respondents are mostly tall and they have a larger feet for the female they have smaller feet and short in height. I could still say the same about the last three columns the male respondents have larger feet and tall in height. The female are short in height and have smaller feet.

Factors

A nominal variable is a categorical variable without an implied order. This means that it is impossible to say that *'one is worth more than the other'*. In contrast, ordinal variables do have a natural ordering.

Example:

```
Gender <- c("M","F","F","M")
factor_Gender <- factor(Gender)
factor_Gender
```

```
## [1] M F F M
## Levels: F M
```

2. Construct character vector months to a factor with factor() and assign the result to factor_months_vector. Print out factor_months_vector and assert that R prints out the factor levels below the actual values.

Consider data consisting of the names of months:
"March","April","January","November","January",
"September","October","September","November","August",
"January","November","November","February","May","August",
"July","December","August","August","September","November","February","April")

```
months_vector <- c("March","April","January","November","January", "September",
        "October","September","November","August", "January","November",
        "November","February","May","August", "July","December","August",
        "August","September","November","February","April")
factor_months_vector <- factor(months_vector)
print(factor_months_vector)
```

3. Then check the summary() of the months_vector and factor_months_vector. | Interpret the results of both vectors. Are they both equally useful in this case?

```
summary(months_vector)
summary(factor_months_vector)
```

4. Create a vector and factor for the table below.

Direction	Frequency
East	1
West	4
North	3

Note: Apply the factor function with required order of the level.

```
new_order_data <- factor(factor_data,levels = c("East","West","North"))
print(new_order_data)
```

```
Direction <- c("East", "West", "North")
Frequency <- c(1, 4, 3)
x1 <- factor(Direction)
x2 <- factor(Frequency)
print(x1)
print(x2)
```

5. Enter the data below in Excel with file name = *import_march.csv*

Students	Strategy 1	Strategy 2	Strategy 3
Male	8	10	8
	4	8	6
	0	6	4
Female	14	4	15
	10	2	12
	6	0	9

Figure 1: Excel File

a. Import the excel file into the *Environment Pane* using read.table() function. Write the code.

```
getwd()
a<- read.table("import_march.csv", header= TRUE, sep= "," )
a
```

b. View the dataset. Write the code and its

```
View(a)
```