RWorksheet_Sison#4.

Kathrina Casandra Sison

2024-10-14

```
Shoe_Size <- c(6.5, 9.0, 8.5, 8.5, 10.5, 7.0, 9.5, 9.0, 13.0, 7.5, 10.5, 8.5, 12.0, 10.5, 13.0, 11.5, 8
Height <- c(66.0, 68.0, 64.5, 65.0, 70.0, 64.0, 70.0, 71.0, 72.0, 64.0, 74.75, 67.0, 71.0, 71.0, 77.0, 
Gender <- c("F", "F", "F", "F", "M", "F", "F", "F", "M", "F", "M", "F", "M", "M", "M", "M", "F", "F", "
shoe_data1 <- data.frame(Shoe_Size,Height, Gender)
```

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

#a. Describe the data. The data is consist of different shoe sizes along with its height and gender of the user.

#b. Create a subset by males and females with their corresponding shoe size and height. What its result? Show the R scripts.

```
str(shoe_data1)
```

```
## 'data.frame':    28 obs. of  3 variables:
##  $ Shoe_Size: num  6.5 9 8.5 8.5 10.5 7 9.5 9 13 7.5 ...
##  $ Height   : num  66 68 64.5 65 70 64 70 71 72 64 ...
##  $ Gender   : chr  "F" "F" "F" "F" ...
```

```
male_subset <- shoe_data1[shoe_data1$Gender == "M",]
female <- shoe_data1[shoe_data1$Gender == "F",]
print("Male_subset")
```

```
## [1] "Male_subset"
```

```
print(male_subset)
```

```
##    Shoe_Size Height Gender
## 5       10.5  70.00      M
## 9       13.0  72.00      M
## 11      10.5  74.75      M
## 13      12.0  71.00      M
## 14      10.5  71.00      M
## 15      13.0  77.00      M
## 16      11.5  72.00      M
## 19      10.0  72.00      M
## 22       8.5  67.00      M
## 23      10.5  73.00      M
## 25      10.5  72.00      M
## 26      11.0  70.00      M
## 27       9.0  69.00      M
## 28      13.0  70.00      M
```

```
print("Female Susbet")
```

```
## [1] "Female Susbet"
```

```
female
```

```
##    Shoe_Size Height Gender
## 1        6.5   66.0      F
## 2        9.0   68.0      F
## 3        8.5   64.5      F
## 4        8.5   65.0      F
## 6        7.0   64.0      F
## 7        9.5   70.0      F
## 8        9.0   71.0      F
## 10       7.5   64.0      F
## 12       8.5   67.0      F
## 17       8.5   59.0      F
## 18       5.0   62.0      F
## 20       6.5   66.0      F
## 21       7.5   64.0      F
## 24       8.5   69.0      F
```

c. Find the mean of shoe size and height of the respondents. Write the R scripts and its result.

```
Shoe_Size <- c(6.5, 9.0, 8.5, 8.5, 10.5, 7.0, 9.5, 9.0, 13.0, 7.5, 10.5, 8.5, 12.0, 10.5, 13.0, 11.5, 8
Height <- c(66.0, 68.0, 64.5, 65.0, 70.0, 64.0, 70.0, 71.0, 72.0, 64.0, 74.75, 67.0, 71.0, 71.0, 77.0, 
mean_shoe_size <- mean(Shoe_Size)
mean_height <- mean(Height)
mean_shoe_size
```

```
## [1] 9.410714
```

```
mean_height
```

```
## [1] 68.58036
```

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

```
correlation <- cor(shoe_data1$Shoe_Size, shoe_data1$Height, use = "complete.obs")
print(paste("Correlation between Shoe Size and Height:", correlation))
```

```
## [1] "Correlation between Shoe Size and Height: 0.775187654599107"
```

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","N April")

```
months <- 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)

print(factor_months_vector)
```

```
##  [1] March     April     January   November  January   September October
##  [8] September November  August    January   November  November  February
## [15] May       August    July      December  August    August    September
```

```
## [22] November  February  April
## 11 Levels: April August December February January July March May ... September
```

```
print(levels(factor_months_vector))
```

```
##  [1] "April"     "August"    "December"  "February"  "January"   "July"
##  [7] "March"     "May"       "November"  "October"   "September"
```

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

```
factor_months <- factor(months)
summary(months)
```

```
##    Length     Class      Mode
##        24 character character
```

```
summary(factor_months)
```

```
##     April    August  December  February   January      July     March       May
##         2         4         1         2         3         1         1         1
##  November   October September
##         5         1         3
```

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

```
direction <- c("East", "West", "North")
frequency <- c(1, 4, 3)
factor_data <- factor(direction, levels = c("East", "West", "North"))
print(factor_data)
```

```
## [1] East  West  North
## Levels: East West North
```

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

```
## [1] East  West  North
## Levels: East West North
```

5.

```
write.csv("import_march.csv", row.names = FALSE)
```

```
## "x"
## "import_march.csv"
```

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