

Work Sheet 4

Wendy Nalaza

2022-11-22

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

```
library(dplyr) library(readr) library(data.table)
```

a. Describe the data.

##	Shoesize	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
## 5	10.5	70.0	M
## 6	7.0	64.0	F
## 7	9.5	70.0	M
## 8	9.0	71.0	F
## 9	13.0	72.0	M
## 10	7.5	64.0	M
## 11	10.5	74.5	M
## 12	8.5	67.0	F
## 13	12.0	71.0	M
## 14	10.5	71.0	M
## 15	13.0	77.0	M
## 16	11.5	72.0	M
## 17	8.5	59.0	F
## 18	5.0	62.0	F
## 19	10.0	72.0	M
## 20	6.5	66.0	F
## 21	7.5	64.0	M
## 22	8.5	67.0	M
## 23	10.5	73.0	M
## 24	8.5	69.0	F
## 25	10.5	72.0	M
## 26	11.0	70.0	M
## 27	9.0	69.0	M
## 28	13.0	70.0	M

Answer: The output will show a data base on what we put on each rows within the dataframe

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

- Male

```
Boy <- subset(data_frame, Gender == "M")
mean(Boy$Shoesize)
```

```
## [1] 10.47059
```

```
mean(Boy$Height)
```

```
## [1] 70.5
```

- Female

```
Girl <- subset(data_frame, Gender == "F")
mean(Girl$Shoesize)
```

```
## [1] 7.772727
```

```
mean(Girl$Height)
```

```
## [1] 65.59091
```

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

Yes, the higher the value of height, the greater the value of the shoe size, and the factor levels below are the actual values.

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.

```
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_Month_Vector <- factor(Months_Vector)
Factor_Month_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
```

3. Then check the *summary()* of the *Months_Vector_vector* and *Factor_Month_Vector_vector*. Interpret the results of both vectors. Are they both equally useful in this case?

```
summary(Months_Vector)
```

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

```
summary(Factor_Month_Vector)
```

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

Answer: Yes, as for the months_vector, it shows the number of months and their class and mode, while in the factor_months_vector, the month has been factored by level and alphabetically, and it shows the number of each month.

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

```
factorData <- data.frame(
  Direction = c("East", "West", "North"),
  Frequency = c(1, 4, 3)
)
factorData
```

```
##   Direction Frequency
## 1      East          1
## 2      West          4
## 3      North          3
```

```
newOrderData <- factor(factorData, levels = c("East", "West", "North"))
print(newOrderData)
```

```
## Direction Frequency
##      <NA>      <NA>
## Levels: East West North
```

5. Enter the data below in Excel with file name = import_march.csv a. Import the excel file into the Environment Pane using read.table() function.

```
import_march.csv <- read_excel("iris_sample.xls") import_march.csv
```

```
table1 <- read.table("C:/CS101_DATA_SCIENCE/Nalaza_Repo/Worksheet#4/import_march.csv", header = TRUE, sep = ";")
table1
```

```
##   Students Strategy.1 Strategy.2 Strategy.3
## 1      Male          8          10          8
## 2          4          8          6
## 3          0          6          4
## 4      Female         14          4         15
## 5          10          2         12
## 6          16          0          9
```

b. View the dataset. Write the code and its result.

```
read.csv("C:/CS101_DATA_SCIENCE/Nalaza_Repo/Worksheet#4/import_march.csv")
```

##	Students	Strategy.1	Strategy.2	Strategy.3
## 1	Male	8	10	8
## 2		4	8	6
## 3		0	6	4
## 4	Female	14	4	15
## 5		10	2	12
## 6		16	0	9

import_march.csv