

RWorksheet_Subosa#4a.Rmd

Gian Adree Subosa

2024-10-18

#1. Data for shoe size and height

```
# Create the data frame
```

```
shoe_size <- c(6.5, 9.0, 8.5, 8.5, 10.5, 7.0, 9.5, 9.0, 13.0, 7.5, 8.5, 10.5, 12.0, 10.5, 13.0, 11.5, 8
```

```
height <- c(66.0, 68.0, 64.5, 65.0, 70.0, 64.0, 69.0, 71.0, 72.0, 64.0, 67.0, 74.5, 71.0, 71.0, 77.0, 71.0)
```

```
gender <- c('F', 'F', 'F', 'F', 'M', 'F', 'F', 'M', 'M', 'F', 'F', 'M', 'M', 'M', 'M', 'M', 'F', 'F', 'F')
```

```
df <- data.frame(Shoe_Size = shoe_size, Height = height, Gender = gender)
```

```
summary(df)
```

```
##      Shoe_Size      Height      Gender
##  Min.   : 5.000   Min.   :59.00   Length:26
##  1st Qu.: 8.500   1st Qu.:65.25   Class :character
##  Median : 9.000   Median :69.00   Mode  :character
##  Mean    : 9.442   Mean    :68.35
##  3rd Qu.:10.500   3rd Qu.:71.00
##  Max.    :13.000   Max.    :77.00
```

#a.) Describe the data.

The data contains information about the shoe size, height and gender. The data indicates that m

#b.) Create a subset

```
male_subset <- subset(df, Gender == 'M')
```

```
female_subset <- subset(df, Gender == 'F')
```

```
print(male_subset)
```

##	Shoe_Size	Height	Gender
## 5	10.5	70.0	M
## 8	9.0	71.0	M
## 9	13.0	72.0	M
## 12	10.5	74.5	M
## 13	12.0	71.0	M
## 14	10.5	71.0	M
## 15	13.0	77.0	M
## 16	11.5	72.0	M
## 19	10.0	72.0	M
## 22	8.5	67.0	M
## 23	10.5	69.0	M
## 24	11.0	73.0	M
## 25	9.0	69.0	M
## 26	13.0	70.0	M

```
print(female_subset)
```

```
##      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    69.0    F
## 10     7.5    64.0    F
## 11     8.5    67.0    F
## 17     8.5    59.0    F
## 18     5.0    62.0    F
## 20     6.5    66.0    F
## 21     8.5    64.0    F
```

#c.) Find the mean of shoe size and height of the respondents

```
mean_shoe_size <- mean(df$Shoe_Size)
mean_height <- mean(df$Height)
print(mean_shoe_size)
```

```
## [1] 9.442308
```

```
print(mean_height)
```

```
## [1] 68.34615
```

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

```
correlation_data <- cor(df$Shoe_Size, df$Height)
print(correlation_data)
```

```
## [1] 0.7750242
```

A correlation of 0.7750242 indicates a strong positive relationship between shoe size and height. This suggests that as shoe size increases, height also tends to increase.

#2. Construct character vector months to a factor with factor() and assign the result to factor_months_vector

```
months_vector <- c("March", "April", "January", "November", "January", "September", "October", "September")
factor_months_vector <- factor(months_vector)
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
```

#3. Check the summary() of the months_vector and factor_months_vector

```
summary(months_vector)
```

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

```
summary(factor_months_vector)
```

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

Are they both equally useful in this case?

The factor vector is more beneficial in this scenario because it provides a clear split of the data into categories.

#4. Create a vector for direction and frequency and factor below

```
direction <- c("East", "West", "North")
frequency <- c(1, 4, 3)

direction_data <- data.frame(Direction = direction, Frequency = frequency)

factor_data <- factor(direction_data$Direction)

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

```
## [1] East West North
## 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
data_table <- read.table("import_march.csv", header = TRUE, sep = ",", stringsAsFactors = FALSE)
print(data_table)
```

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