RWorksheet_Malayas#4a.Rmd

Andrew Miguel M. Malayas BSIT2A

2024-10-14

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

```
df <- data.frame(
shoesize = c(6.5, 9.0, 8.5, 8.5, 7.0, 9.0, 9.5, 13.0, 7.5, 10.5, 10.5, 12.0, 10.5,
             13.0, 11.5, 8.5, 5.0, 10.0, 6.5, 7.5, 8.5, 10.5, 8.5, 10.5, 11.0, 9.0, 13.0),
height = c(66.0, 68.0, 65.0, 65.0, 64.0, 71.0, 72.0, 72.0, 74.5, 67.0, 74.5, 71.0, 71.0,
           77.0, 72.0, 59.0, 62.0, 72.0, 66.0, 64.0, 67.0, 73.0, 69.0, 72.0, 70.0, 69.0, 70.0),
gender = c("F", "F", "F", "F", "F", "F", "F", "M", "F", "M", "M", "M", "M",
           "M", "M", "F", "F", "M", "F", "F", "F", "M", "F", "M", "M", "M", "M")
)

df
```

```
##    shoesize height gender
## 1       6.5   66.0      F
## 2       9.0   68.0      F
## 3       8.5   65.0      F
## 4       8.5   65.0      F
## 5       7.0   64.0      F
## 6       9.0   71.0      F
## 7       9.5   72.0      F
## 8      13.0   72.0      M
## 9       7.5   74.5      F
## 10     10.5   67.0      M
## 11     10.5   74.5      M
## 12     12.0   71.0      M
## 13     10.5   71.0      M
## 14     13.0   77.0      M
## 15     11.5   72.0      M
## 16      8.5   59.0      F
## 17      5.0   62.0      F
## 18     10.0   72.0      M
## 19      6.5   66.0      F
## 20      7.5   64.0      F
## 21      8.5   67.0      F
## 22     10.5   73.0      M
## 23      8.5   69.0      F
## 24     10.5   72.0      M
## 25     11.0   70.0      M
## 26      9.0   69.0      M
## 27     13.0   70.0      M
```

a. Describe the data.

- The dataset contains 27 entries with information on shoe size, height, and gender. Shoe sizes range from 5.0 to 13.0, while heights vary from 59 to 77 inches. Gender is categorized as either "F" for female or "M" for male.

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

- The data displays the shoe size and height for each male and female included in the dataset.

```
male <- subset(df, gender == "M", select = c(shoesize, height))


female <- subset(df, gender == "F", select = c(shoesize, height))


male
```

```
##    shoesize height
## 8      13.0   72.0
## 10     10.5   67.0
## 11     10.5   74.5
## 12     12.0   71.0
## 13     10.5   71.0
## 14     13.0   77.0
## 15     11.5   72.0
## 18     10.0   72.0
## 22     10.5   73.0
## 24     10.5   72.0
## 25     11.0   70.0
## 26      9.0   69.0
## 27     13.0   70.0
```

```
female
```

```
##    shoesize height
## 1       6.5   66.0
## 2       9.0   68.0
## 3       8.5   65.0
## 4       8.5   65.0
## 5       7.0   64.0
## 6       9.0   71.0
## 7       9.5   72.0
## 9       7.5   74.5
## 16      8.5   59.0
## 17      5.0   62.0
## 19      6.5   66.0
## 20      7.5   64.0
## 21      8.5   67.0
## 23      8.5   69.0
```

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

```r
meanS <- mean(df$shoesize)


meanH <- mean(df$height)


meanS
```

```
## [1] 9.444444
```

```r
meanH
```

```
## [1] 69
```

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

- Yes, as taller individuals often have larger feet, there is frequently a correlation between shoe size and height. This can vary, though, and the strength of this association may be influenced by things like heredity.

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.

```r
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")

factormonths <- factor(months)

print(factormonths)
```

```
##  [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
```

```r
levels(factormonths)
```

```
##  [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? -The result of months_vector is that is states the Length, Class and Mode. While the factor_months_vector states how many months in the data for example April, April has 2. -I think they are both useful because it is easy for me to understand and determine how many types of data from the raw data itself.

```r
summary(months)
```

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

```r
summary(factormonths)
```

```
##     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.

```r
direction <- c("East", "West", "North", "West", "North", "West", "North", "West")

factordata <- factor(direction)

newdata <- factor(factordata,levels = c("East","West","North"))
newdata
```

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