

RWorksheet_SABIO#4a.Rmd

2023-10-25

```
install.packages("readxl")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

```
library(readxl)
```

```
HouseholdData <- read_excel("Household Data.xlsx")
```

```
## New names:
```

```
## * `Height` -> `Height...2`
```

```
## * `Gender` -> `Gender...3`
```

```
## * `Height` -> `Height...5`
```

```
## * `Gender` -> `Gender...6`
```

```
# View(HouseholdData)
```

```
HouseholdData
```

```
## # A tibble: 14 x 6
```

```
##   `Shoe Size` Height...2 Gender...3 `Shoe size` Height...5 Gender...6
```

```
##           <dbl>      <dbl> <chr>           <dbl>      <dbl> <chr>
```

```
## 1           6.5        66   F              13         77   M
```

```
## 2           9         68   F             11.5        72   M
```

```
## 3          8.5       64.5   F              8.5         59   F
```

```
## 4          8.5        65   F              5          62   F
```

```
## 5         10.5       70    M              10         72   M
```

```
## 6           7         64   F              6.5         66   F
```

```
## 7          9.5       70   F              7.5         64   F
```

```
## 8           9         71   F              8.5         67   M
```

```
## 9          13        72    M             10.5        73   M
```

```
## 10          7.5       64   F              8.5         69   F
```

```
## 11         10.5       74    M             10.5        72   M
```

```
## 12          8.5       67   F              11         70   M
```

```
## 13          12       71    M              9          69   M
```

```
## 14         10.5       71    M              13         70   M
```

```
#describe the data
```

```
mean(HouseholdData$`Shoe Size`)
```

```
## [1] 9.321429
```

```
mean(HouseholdData$Height...2)
```

```
## [1] 68.39286
```

```
#yes,because a person with a taller height may have a larger shoe size on average.
```

```
months_vector <- c("March","April","January","November","January",
```

```
"September","October","September","November","August",
```

```
"January","November","November","February","May","August","July","December","August","August","September")
```

```

"April")
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

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

# the result of summary of months_vector it provides information on
#the counts of each month in the character vector it is useful when
#understanding the distribution of month names in original data. In
#factor_months_vector gives you the same information but represented
#as counts of factor levels it is useful when you want to work with
#categorical data and perform operation depend on factor levels.

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

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

```