Assignment-2

Shraddha Ganesh

2023-09-24

Question-1

```
install.packages("readr", repos = "http://cran.us.r-project.org")

## package 'readr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\KREASU\AppData\Local\Temp\RtmpEvWton\downloaded_packages

library(readr)
getwd()

## [1] "C:/Users/KREASU/Desktop"

setwd("C:/Users/KREASU/Downloads/")
expenditures <-  read_csv("expenditures.csv")
expenditures

## # A tibble: 11 × 7
##    `narendra moDi`  `50`  `30`  `20`   `0` `2021` `09171950`
##    <chr>           <dbl> <dbl> <dbl> <dbl>  <dbl> <chr>
##  1 RAhul gandhi       60    40    30     0   2021 06191970
##  2 AmiL Shah          40    50    40     0   2021 10221964
##  3 Sonia Gyandhi      45    35    25     1   2021 12091946
##  4 mamata Banerjee 55     45    35     1   2021 01051955
##  5 MayawaTi Das       35    25    15     1   2021 01151956
##  6 narendra moDi      40    25    10     0   2022 09171950
##  7 RAhul gandhi       50    35    20     0   2022 06191970
##  8 AmiL Shah          45    30    15     0   2022 10221964
##  9 Sonia Gyandhi      30    20     5     1   2022 12091946
## 10 mamata Banerjee 60     40    25     1   2022 01051955
## 11 MayawaTi Das       25    15    10     1   2022 01151956
```

Question-2

```
colnames(expenditures) <-  c("person","food","clothing","entertainment","gender",
                             "year","date_of_birth")
df <-  data.frame(list(person = "narendra moDi", food = 50,
                       clothing = 30,entertainment = 20,
                       gender = 0,year = 2021,
                       date_of_birth = "09171950"))
expenditures <-  rbind(df,expenditures)
expenditures
```

```
##             person food clothing entertainment gender year date_of_birth
## 1    narendra moDi   50       30            20      0 2021      09171950
## 2     RAhul gandhi    60       40            30      0 2021      06191970
## 3        AmiL Shah     40       50            40      0 2021      10221964
## 4    Sonia Gyandhi   45       35            25      1 2021      12091946
## 5  mamata Banerjee 55       45            35      1 2021      01051955
## 6     MayawaTi Das   35       25            15      1 2021      01151956
## 7    narendra moDi    40       25            10      0 2022      09171950
## 8     RAhul gandhi     50       35            20      0 2022      06191970
## 9        AmiL Shah      45       30            15      0 2022      10221964
## 10   Sonia Gyandhi   30       20             5      1 2022      12091946
## 11 mamata Banerjee  60       40            25      1 2022      01051955
## 12    MayawaTi Das   25       15            10      1 2022      01151956
```

Question-3

```
install.packages(c("dplyr","tidyr","tidyverse"), repos = "http://cran.us.r-project.org")

## package 'dplyr' successfully unpacked and MD5 sums checked
## package 'tidyr' successfully unpacked and MD5 sums checked
## package 'tidyverse' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\KREASU\AppData\Local\Temp\RtmpEvWton\downloaded_packages

library(dplyr)
library(tidyr)
library(tidyverse)
new_names <-  strsplit(expenditures$person, " ")
unlisted <-  unlist(new_names)
matrix <-  matrix(unlisted,
              nrow = length(unlisted)/2,
              ncol = 2,
              byrow = TRUE)
colnames(matrix) <- c("first_name","last_name")
final <-  as.data.frame(matrix)
expenditures <-  expenditures %>% select(-person)
expenditures <-  cbind(final,expenditures)
expenditures

##    first_name last_name food clothing entertainment gender year date_of_birth
## 1    narendra      moDi   50       30            20      0 2021      09171950
## 2       RAhul    gandhi    60       40            30      0 2021      06191970
## 3       AmiL      Shah     40       50            40      0 2021      10221964
## 4       Sonia   Gyandhi   45       35            25      1 2021      12091946
## 5      mamata  Banerjee 55       45            35      1 2021      01051955
## 6    MayawaTi       Das   35       25            15      1 2021      01151956
## 7    narendra      moDi    40       25            10      0 2022      09171950
## 8       RAhul    gandhi     50       35            20      0 2022      06191970
## 9       AmiL      Shah      45       30            15      0 2022      10221964
## 10      Sonia   Gyandhi     30       20             5      1 2022      12091946
```

```
## 11     mamata  Banerjee   60       40           25      1 2022      01051955
## 12   MayawaTi       Das   25       15           10      1 2022      01151956
```

Question-4

```
expenditures$first_name <-  str_to_title(expenditures$first_name)
expenditures$last_name <-  str_to_title(expenditures$last_name)
expenditures

##    first_name last_name food clothing entertainment gender year date_of_birth
## 1    Narendra      Modi   50       30           20      0 2021      09171950
## 2       Rahul    Gandhi   60       40           30      0 2021      06191970
## 3        Amil      Shah     40       50           40      0 2021      10221964
## 4       Sonia   Gyandhi   45       35           25      1 2021      12091946
## 5     Mamata  Banerjee   55       45           35      1 2021      01051955
## 6    Mayawati       Das   35       25           15      1 2021      01151956
## 7    Narendra      Modi   40       25           10      0 2022      09171950
## 8       Rahul    Gandhi   50       35           20      0 2022      06191970
## 9        Amil      Shah   45       30           15      0 2022      10221964
## 10      Sonia   Gyandhi   30       20            5      1 2022      12091946
## 11    Mamata  Banerjee   60       40           25      1 2022      01051955
## 12   Mayawati       Das   25       15           10      1 2022      01151956
```

Question-5

```
expenditures$first_name[grep("Amil",expenditures$first_name)] = "Amit"
expenditures$last_name[grep("Gyandhi",expenditures$last_name)] = "Gandhi"
expenditures

##    first_name last_name food clothing entertainment gender year date_of_birth
## 1    Narendra      Modi   50       30           20      0 2021      09171950
## 2       Rahul    Gandhi   60       40           30      0 2021      06191970
## 3        Amit      Shah   40       50           40      0 2021      10221964
## 4       Sonia    Gandhi   45       35           25      1 2021      12091946
## 5     Mamata  Banerjee   55       45           35      1 2021      01051955
## 6    Mayawati       Das   35       25           15      1 2021      01151956
## 7    Narendra      Modi   40       25           10      0 2022      09171950
## 8       Rahul    Gandhi   50       35           20      0 2022      06191970
## 9        Amit      Shah   45       30           15      0 2022      10221964
## 10      Sonia    Gandhi   30       20            5      1 2022      12091946
## 11    Mamata  Banerjee   60       40           25      1 2022      01051955
## 12   Mayawati       Das   25       15           10      1 2022      01151956
```

Question-6

```
person <-  paste(expenditures$first_name,expenditures$last_name,sep = " ")
expenditures <-  expenditures %>% select(-first_name,-last_name)
expenditures <-  cbind(person,expenditures)
expenditures

##             person food clothing entertainment gender year date_of_birth
## 1    Narendra Modi   50       30           20      0 2021      09171950
## 2     Rahul Gandhi   60       40           30      0 2021      06191970
```

```
## 3        Amit Shah   40       50           40      0 2021      10221964
## 4     Sonia Gandhi   45       35           25      1 2021      12091946
## 5  Mamata Banerjee   55       45           35      1 2021      01051955
## 6     Mayawati Das   35       25           15      1 2021      01151956
## 7    Narendra Modi   40       25           10      0 2022      09171950
## 8     Rahul Gandhi   50       35           20      0 2022      06191970
## 9        Amit Shah   45       30           15      0 2022      10221964
## 10    Sonia Gandhi   30       20           5       1 2022      12091946
## 11 Mamata Banerjee   60       40           25      1 2022      01051955
## 12    Mayawati Das   25       15           10      1 2022      01151956
```

Question-7

```
date_of_birth <-  as.Date(expenditures$date_of_birth, format = "%m%d%Y")
expenditures <-  expenditures %>% select(-date_of_birth)
expenditures <-  cbind(expenditures,date_of_birth)
expenditures
```

```
##            person food clothing entertainment gender year date_of_birth
## 1    Narendra Modi   50       30           20      0 2021    1950-09-17
## 2     Rahul Gandhi   60       40           30      0 2021    1970-06-19
## 3        Amit Shah   40       50           40      0 2021    1964-10-22
## 4     Sonia Gandhi   45       35           25      1 2021    1946-12-09
## 5  Mamata Banerjee   55       45           35      1 2021    1955-01-05
## 6     Mayawati Das   35       25           15      1 2021    1956-01-15
## 7    Narendra Modi   40       25           10      0 2022    1950-09-17
## 8     Rahul Gandhi   50       35           20      0 2022    1970-06-19
## 9        Amit Shah   45       30           15      0 2022    1964-10-22
## 10    Sonia Gandhi   30       20           5       1 2022    1946-12-09
## 11 Mamata Banerjee   60       40           25      1 2022    1955-01-05
## 12    Mayawati Das   25       15           10      1 2022    1956-01-15
```

Question-8

```
expenditure_2021 <-  expenditures %>% filter(year == "2021")
expenditure_2022 <-  expenditures %>% filter(year == "2022")
write.csv(expenditure_2021,file = "expenditure_2021.csv")
write.csv(expenditure_2022,file = "expenditure_2022.csv")
expenditure_2021
```

```
##            person food clothing entertainment gender year date_of_birth
## 1    Narendra Modi   50       30           20      0 2021    1950-09-17
## 2     Rahul Gandhi   60       40           30      0 2021    1970-06-19
## 3        Amit Shah   40       50           40      0 2021    1964-10-22
## 4     Sonia Gandhi   45       35           25      1 2021    1946-12-09
## 5  Mamata Banerjee   55       45           35      1 2021    1955-01-05
## 6     Mayawati Das   35       25           15      1 2021    1956-01-15
```

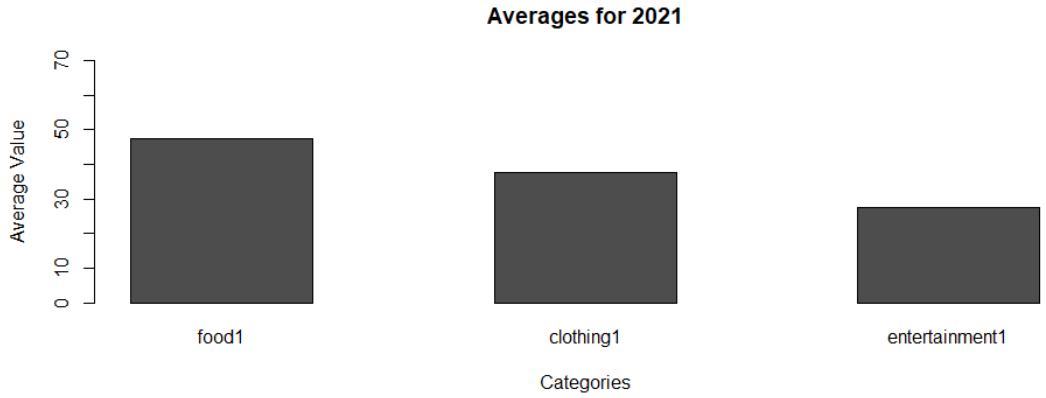
```
expenditure_2022
```

```
##            person food clothing entertainment gender year date_of_birth
## 1    Narendra Modi   40       25           10      0 2022    1950-09-17
```

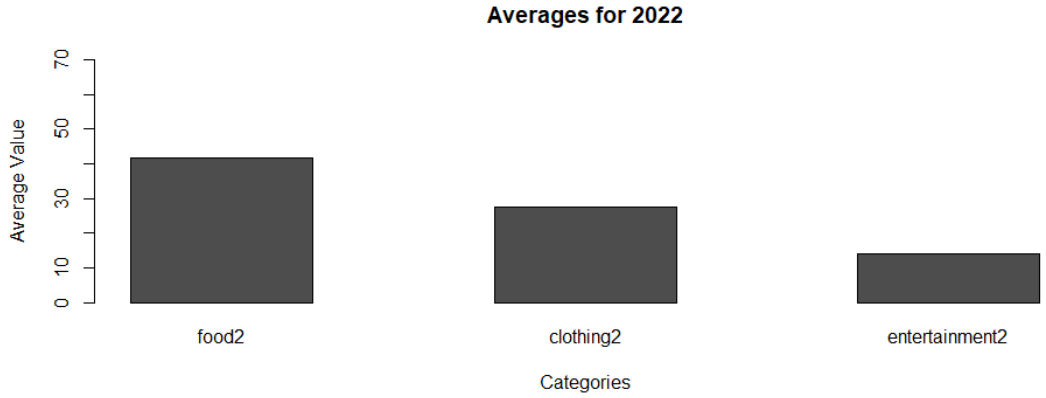
```
## 2    Rahul Gandhi   50       35            20      0 2022    1970-06-19
## 3       Amit Shah   45       30            15      0 2022    1964-10-22
## 4    Sonia Gandhi   30       20             5      1 2022    1946-12-09
## 5 Mamata Banerjee   60       40            25      1 2022    1955-01-05
## 6    Mayawati Das   25       15            10      1 2022    1956-01-15
```

Question-9

```
#Averages for 2021
food1 <-  mean(expenditure_2021$food)
food1

## [1] 47.5

clothing1 <-  mean(expenditure_2021$clothing)
clothing1

## [1] 37.5

entertainment1 <-  mean(expenditure_2021$entertainment)
entertainment1

## [1] 27.5

#Averages for 2022
food2 <-  round(mean(expenditure_2022$food),2)
food2

## [1] 41.67

clothing2 <-  mean(expenditure_2022$clothing)
clothing2

## [1] 27.5

entertainment2 <-  round(mean(expenditure_2022$entertainment),2)
entertainment2

## [1] 14.17

#Plot for 2021
combined1 <-  cbind(food1,clothing1,entertainment1)
barplot(combined1,beside = TRUE,
        ylim=c(0,70),
        main = "Averages for 2021",
        xlab = "Categories",
        ylab = "Average Value")
```



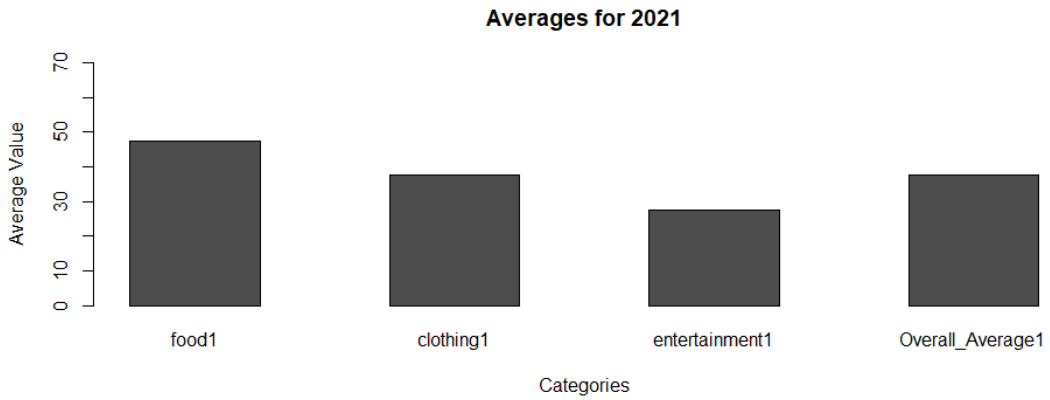

```
#Plot for 2022
combined2 <-  cbind(food2,clothing2,entertainment2)
barplot(combined2,beside = TRUE,
        ylim=c(0,70),
        main = "Averages for 2022",
        xlab = "Categories",
        ylab = "Average Value")
```



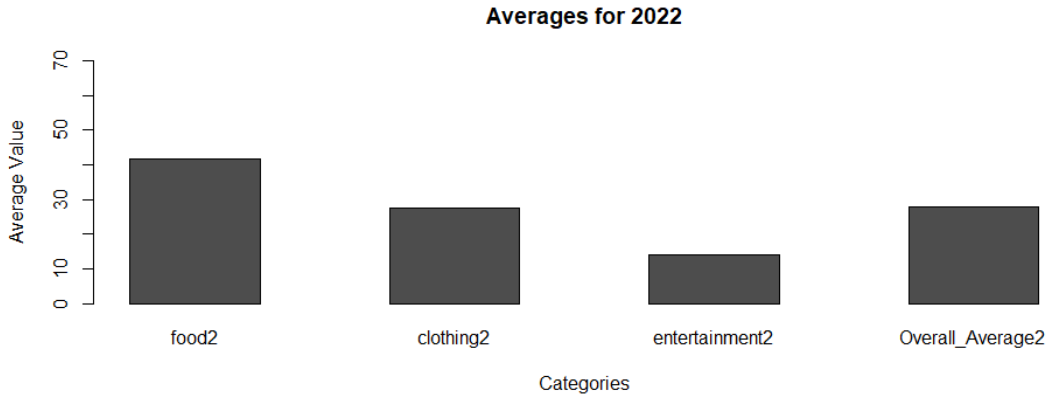
Question-10

```
Overall_Average1 <-  mean(combined1)
Overall_Average1

## [1] 37.5

Overall_Average2 <-  mean(combined2)
Overall_Average2

## [1] 27.78

#Plot for 2021
c1 <-  cbind(food1,clothing1,entertainment1,Overall_Average1)
barplot(c1,beside = TRUE,
        ylim=c(0,70),
```

```
      main = "Averages for 2021",
      xlab = "Categories",
      ylab = "Average Value")
```




```
#Plot for 2022
c2 <-  cbind(food2,clothing2,entertainment2,Overall_Average2)
barplot(c2,beside = TRUE,
      ylim=c(0,70),
      main = "Averages for 2022",
      xlab = "Categories",
      ylab = "Average Value")
```



Question-11

```
#Total consumption in 2022
Narendra_Modi <-  sum(expenditure_2022[1,c(2:4)])
Narendra_Modi

## [1] 75

Rahul_Gandhi <-  sum(expenditure_2022[2,c(2:4)])
Rahul_Gandhi

## [1] 105
```

```
Amit_Shah <-  sum(expenditure_2022[3,c(2:4)])
Amit_Shah
```

```
## [1] 90
```

```
Sonia_Gandhi <-  sum(expenditure_2022[4,c(2:4)])
Sonia_Gandhi
```

```
## [1] 55
```

```
Mamata_Banarjee <-  sum(expenditure_2022[5,c(2:4)])
Mamata_Banarjee
```

```
## [1] 125
```

```
Mayawati_Das <-  sum(expenditure_2022[6,c(2:4)])
Mayawati_Das
```

```
## [1] 50
```

```
total_consumption <-  cbind(Narendra_Modi,Rahul_Gandhi,Amit_Shah,Sonia_Gandhi,
                      Mamata_Banarjee,Mayawati_Das)

#Plot for total consumption
barplot(total_consumption, beside = TRUE,
        ylim = c(0,150),
        main = "Total Consumption (2022)",
        xlab = "Names",
        ylab = "Consumption",
        xlim = c(0,12))
```