

RWorksheet_Olivo#3b

2023-10-11

```
#1.a
resp_no <- c(1:20)
sex <- c(2,2,1,2,2,2,2,2,2,1,2,2,2,2,2,2,1,2)
occ <- c(1,3,3,3,1,2,3,1,1,1,3,2,1,3,3,1,3,1,2,1)
pers_at_home <- c(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6)
sibs <- c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2)
h_type <- c(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2)
```

```
household_data <- data.frame(
  Respondents = resp_no,
  Sex = sex,
  FatherOccupation = occ,
  PersonAtHome = pers_at_home,
  SiblingsAtSchool = sibs,
  HouseType = h_type
)
household_data
```

##	Respondents	Sex	FatherOccupation	PersonAtHome	SiblingsAtSchool	HouseType
## 1	1	2	1	5	6	1
## 2	2	2	3	7	4	2
## 3	3	1	3	3	4	3
## 4	4	2	3	8	1	1
## 5	5	2	1	5	2	1
## 6	6	2	2	9	1	3
## 7	7	2	3	6	5	3
## 8	8	2	1	7	3	1
## 9	9	2	1	8	1	2
## 10	10	2	1	4	2	3
## 11	11	1	3	7	3	2
## 12	12	2	2	5	2	3
## 13	13	2	1	4	5	2
## 14	14	2	3	7	5	2
## 15	15	2	3	8	2	3
## 16	16	2	1	8	1	3
## 17	17	2	3	3	2	3
## 18	18	2	1	11	5	3
## 19	19	1	2	7	3	3
## 20	20	2	1	6	2	2

```
#1.2
```

```
str(household_data)
```

```
## 'data.frame': 20 obs. of 6 variables:
```

```
## $ Respondents      : int  1 2 3 4 5 6 7 8 9 10 ...
## $ Sex              : num  2 2 1 2 2 2 2 2 2 2 ...
## $ FatherOccupation: num  1 3 3 3 1 2 3 1 1 1 ...
## $ PersonAtHome     : num  5 7 3 8 5 9 6 7 8 4 ...
## $ SiblingsAtSchool: num  6 4 4 1 2 1 5 3 1 2 ...
## $ HouseType        : num  1 2 3 1 1 3 3 1 2 3 ...
```

```
summary(household_data)
```

```
##   Respondents      Sex      FatherOccupation  PersonAtHome
##   Min.   : 1.00   Min.   :1.00   Min.   :1.00   Min.   : 3.0
##   1st Qu.: 5.75   1st Qu.:2.00   1st Qu.:1.00   1st Qu.: 5.0
##   Median :10.50   Median :2.00   Median :2.00   Median : 7.0
##   Mean   :10.50   Mean   :1.85   Mean   :1.95   Mean   : 6.4
##   3rd Qu.:15.25   3rd Qu.:2.00   3rd Qu.:3.00   3rd Qu.: 8.0
##   Max.   :20.00   Max.   :2.00   Max.   :3.00   Max.   :11.0
##   SiblingsAtSchool  HouseType
##   Min.   :1.00   Min.   :1.0
##   1st Qu.:2.00   1st Qu.:2.0
##   Median :2.50   Median :2.5
##   Mean   :2.95   Mean   :2.3
##   3rd Qu.:4.25   3rd Qu.:3.0
##   Max.   :6.00   Max.   :3.0
```

```
# the data frame consists of 20 observations(rows) and 6 variables (columns)
```

```
# the variables are:
```

```
# respondents - which contains a numeric identifier for each respondent
```

```
# sex - represents the gender of the respondent (1 for male, 2 for female)
```

```
# father's occupation - indicates the father's occupation (1 for farmer, 2 for driver, 3 for others)
```

```
# persons at home - represents the number of people at home
```

```
# siblings at school - indicates the number of siblings attending school
```

```
# type of house - describes the type of house (1 for wood, 2 for semi-concrete, 3 for concrete)
```

```
#1.c
```

```
sibs_mean <- mean(household_data$SiblingsAtSchool)
```

```
sibs_mean
```

```
## [1] 2.95
```

```
# the mean of the number of siblings at school is 2.95, which is not 5
```

```
#1.d
```

```
firstTwoRows <- household_data[1:2,]
```

```
firstTwoRows
```

```
##   Respondents Sex FatherOccupation PersonAtHome SiblingsAtSchool HouseType
## 1           1  2              1             5             6             1
## 2           2  2              3             7             4             2
```

```
#1.e
```

```
thirdAndFifthRows <- household_data[c(3,5),c(2,4)]
```

```
thirdAndFifthRows
```

```
##   Sex PersonAtHome
## 3   1             3
```

```
## 5      2      5

#1.f
types_houses <- household_data$HouseType
types_houses

## [1] 1 2 3 1 1 3 3 1 2 3 2 3 2 2 3 3 3 3 2

#1.g
male_farmer <- household_data[household_data$Sex == 1 & household_data$FatherOccupation == 1,]

male_farmer

## [1] Respondents      Sex      FatherOccupation PersonAtHome
## [5] SiblingsAtSchool HouseType
## <0 rows> (or 0-length row.names)

# there are no observations

#1.h

female_resp <- household_data[household_data$SiblingsAtSchool >= 5,]

female_resp

##      Respondents Sex FatherOccupation PersonAtHome SiblingsAtSchool HouseType
## 1              1  2              1              5              6          1
## 7              7  2              3              6              5          3
## 13             13  2              1              4              5          2
## 14             14  2              3              7              5          2
## 18             18  2              1             11              5          3

# there are five observations

# -----

# 2
df = data.frame(Ints=integer(),
Doubles=double(), Characters=character(),
Logicals=logical(),
Factors=factor(),
stringsAsFactors=FALSE
)

print("Structure of the empty dataframe:")

## [1] "Structure of the empty dataframe:"

print(str(df))

## 'data.frame':  0 obs. of  5 variables:
##  $ Ints      : int
##  $ Doubles   : num
##  $ Characters: chr
##  $ Logicals  : logi
##  $ Factors   : Factor w/ 0 levels:
## NULL
```

```

# df is an empty data frame created with 0 rows and 5 columns
# the columns has the following data type:
# ints = integer
# doubles = double
# characters = character
# logicals = logical
# factors = factor (0 levels which means empty)

# serves as a template and can be populated with data

# -----

# 3

new_resp <- c(1:10)
new_sex <- c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male")
new_occ <- c(1,2,3,3,1,2,2,3,1,3)
new_personsAtHome <- c(5,7,3,8,6,4,4,2,11,6)
new_sibs <- c(2,3,0,5,2,3,1,2,6,2)
new_type <- c("Wood", "Congrete", "Congrete", "Wood", "Semi-concrete", "Semi-concrete", "Wood", "Semi-c

HouseholdData <- data.frame(
  Respondents = new_resp,
  Sex = new_sex,
  FatherOccupation = new_occ,
  PersonAtHome = new_personsAtHome,
  SiblingsAtSchool = new_sibs,
  HouseType = new_type
)

write.csv(HouseholdData, file = "HouseholdData.csv")

#3a

imported <- read.csv("HouseholdData.csv")
imported

```

##	X	Respondents	Sex	FatherOccupation	PersonAtHome	SiblingsAtSchool
## 1	1	1	Male	1	5	2
## 2	2	2	Female	2	7	3
## 3	3	3	Female	3	3	0
## 4	4	4	Male	3	8	5
## 5	5	5	Male	1	6	2
## 6	6	6	Female	2	4	3
## 7	7	7	Female	2	4	1
## 8	8	8	Male	3	2	2
## 9	9	9	Female	1	11	6
## 10	10	10	Male	3	6	2
##						
##			HouseType			
## 1			Wood			
## 2			Congrete			
## 3			Congrete			
## 4			Wood			
## 5			Semi-concrete			

```
## 6 Semi-concrete
## 7 Wood
## 8 Semi-concrete
## 9 Semi-concrete
## 10 Congrete
```

#3b

```
imported$Sex <- factor(imported$Sex, levels = c("Male", "Female"))
imported$Sex <- as.integer(imported$Sex)

imported$Sex
```

```
## [1] 1 2 2 1 1 2 2 1 2 1
```

#3c

```
imported$HouseType <- factor(imported$HouseType, levels = c("Wood", "Congrete", "Semi-concrete"))
imported$HouseType <- as.integer(imported$HouseType)

imported$HouseType
```

```
## [1] 1 2 2 1 3 3 1 3 3 2
```

#3d

```
imported$FatherOccupation <- factor(imported$FatherOccupation, levels = c(1,2,3), labels = c("Farmer", "Driver", "Others"))
imported$FatherOccupation
```

```
## [1] Farmer Driver Others Others Farmer Driver Driver Others Farmer Others
## Levels: Farmer Driver Others
```

#3e

```
female_driver <- imported[imported$Sex == 2 & imported$FatherOccupation == "Driver",]
female_driver
```

```
## X Respondents Sex FatherOccupation PersonAtHome SiblingsAtSchool HouseType
## 2 2 2 2 Driver 7 3 2
## 6 6 6 2 Driver 4 3 3
## 7 7 7 2 Driver 4 1 1
```

#3f

```
greaterFive <- imported[imported$SiblingsAtSchool >= 5,]
greaterFive
```

```
## X Respondents Sex FatherOccupation PersonAtHome SiblingsAtSchool HouseType
## 4 4 4 1 Others 8 5 1
## 9 9 9 2 Farmer 11 6 3
```

#4

On July 14, there were more negative sentiments compared to the other sentiments. This could indicate

On July 15, all sentiments increased, with the negative sentiment as the highest. This could imply that

On July 17 and July 18, the negative sentiments stayed high and the neutral and positive sentiments

On July 20, all sentiments got to their lowest with but there were still more negative sentiments than
On July 21, experienced an increase in all sentiments, with the negative being the highest. This could
From this data, we could assume that public sentiment is responsive to external factors and it also