

RWorksheet_Huervana3b

2023-10-11

#1. A Create a data frame using the table below.

```
resp <- c(1:20)
sex <- c(2,2,1,2,2,2,2,2,2,1,2,2,2,2,2,2,1,2)
occu <- c(1,3,3,3,1,2,3,1,1,1,3,2,1,3,3,1,3,1,2,1)
persHome <- c(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6)
sib <- c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2)
houseType <- c(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2)
```

```
householdData <- data.frame(
  Respondents = resp,
  Sex = sex,
  FatherOccupation = occu,
  PersonAtHome = persHome,
  SiblingsAtSchool = sib,
  HouseType = houseType
)
```

householdData

##	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 B

```
str(householdData)
```

```
## '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(householdData)
```

```
## 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
```

```
#You have a table with 20 rows and 6 columns. The columns are:
```

```
#Respondents: A number that identifies each person.
#Sex: Tells if the person is a male (1) or female (2).
#Father's Occupation: indicates the father's occupation .
#Persons at Home: Indicates how many people are in the person's home.
#Siblings at School: Tells you how many siblings of the person are going to school.
#Type of House: describes the type of house
```

```
#1.C Is the mean number of siblings attending is 5?
sibsMean <- mean(householdData$SiblingsAtSchool)
```

```
sibsMean
```

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

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

```
#1.D Extract the 1st two rows and then all the columns using the subsetting functions.
firstTwoRows <- householdData[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 Extract 3rd and 5th row with 2nd and 4th column.
thirdAndFifthRows <- householdData[c(3,5),c(2,4)]
thirdAndFifthRows
```

```
## Sex PersonAtHome
## 3 1 3
## 5 2 5
```

```
#1.F Select the variable types of houses then store the vector that results as types_houses.
types_houses <- householdData$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 Select only all Males respondent that their father occupation was farmer.
```

```
maleFarmer <- householdData[householdData$Sex == 1 & householdData$FatherOccupation == 1,]
```

```
maleFarmer
```

```
## [1] Respondents      Sex      FatherOccupation PersonAtHome
```

```
## [5] SiblingsAtSchool HouseType
```

```
## <0 rows> (or 0-length row.names)
```

```
# there are no observations
```

```
#1.H Select only all females respondent that have greater than or equal to 5 number of siblings attending school
```

```
femaleResp <- householdData[householdData$SiblingsAtSchool >= 5,]
```

```
femaleResp
```

```
##      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
```

```
# It shows 5 observations
```

```
# -----
```

```
# 2 Write a R program to create an empty data frame.
```

```
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
```

```
#You have an empty data table with no rows and 5 columns.
```

```

#These columns have different data types:
#ints: They can hold whole numbers (integers).
#doubles: They can store decimal numbers.
#characters: These columns can hold text or words.
#logicals: They can store either "true" or "false" values.
#factors: These are used to categorize data but currently have no categories (0 levels), so they're essentially empty.
#This table is like a template and can be filled with data later.

```

```

# 3 Create a .csv file of this. Save it as HouseholdData.csv

```

```

newResp <- c(1:10)
newSex <- c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male")
newOcc <- c(1,2,3,3,1,2,2,3,1,3)
newPersonsAtHome <- c(5,7,3,8,6,4,4,2,11,6)
newSib <- c(2,3,0,5,2,3,1,2,6,2)
newType <- c("Wood", "Congrete", "Congrete", "Wood", "Semi-concrete", "Semi-concrete", "Wood", "Semi-concrete", "Wood", "Semi-concrete")

```

```

HouseholdData <- data.frame(
  Respondents = newResp,
  Sex = newSex,
  FatherOccupation = newOcc,
  PersonAtHome = newPersonsAtHome,
  SiblingsAtSchool = newSib,
  HouseType = newType
)

```

```

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

```

```

#3 A Import the csv file into the R environment.

```

```

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
```

#3 B Convert the Sex into factor using factor() function and change it into integer. [Legend: Male = 1 and Female = 2]

```
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
```

#3 C Convert the Type of Houses into factor and change it into integer. [Legend: Wood = 1; Congrete = 2; Semi-concrete = 3]

```
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
```

#3 D On father's occupation, factor it as Farmer = 1; Driver = 2; and Others = 3.

```
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
```

#3 E Select only all females respondent that has a father whose occupation is driver.

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

```
## 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
```

#3 F Select the respondents that have greater than or equal to 5 number of siblings attending school.

```
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 Interpret the graph.

Based on this data, we can infer that public sentiment is influenced by external events or circumstances.