

RWorksheet#3b_Parrenas

Erikka Jane Parrenas

2023-10-19

#1. Create a data frame using the table below

#a Write the codes

```
respondent <- c(1:20)
sex <- c(2,2,1,2,2,2,2,2,2,1,2,2,2,2,2,2,1,2)
fathers_occupation <- c(1,3,3,3,1,2,3,1,1,1,3,2,1,3,3,1,3,1,2,1)
persons_at_home <- c(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6)
siblings_at_school <- c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2)
type_of_house <- 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 = respondent,
  Sex = sex,
  Father_Occupation = fathers_occupation,
  Person_At_Home = persons_at_home,
  Siblings_At_School = siblings_at_school,
  House_Type = type_of_house
)
household_data
```

##	Respondents	Sex	Father_Occupation	Person_At_Home	Siblings_At_School
## 1	1	2	1	5	6
## 2	2	2	3	7	4
## 3	3	1	3	3	4
## 4	4	2	3	8	1
## 5	5	2	1	5	2
## 6	6	2	2	9	1
## 7	7	2	3	6	5
## 8	8	2	1	7	3
## 9	9	2	1	8	1
## 10	10	2	1	4	2
## 11	11	1	3	7	3
## 12	12	2	2	5	2
## 13	13	2	1	4	5
## 14	14	2	3	7	5
## 15	15	2	3	8	2
## 16	16	2	1	8	1
## 17	17	2	3	3	2
## 18	18	2	1	11	5
## 19	19	1	2	7	3
## 20	20	2	1	6	2
##	House_Type				

```
## 1      1
## 2      2
## 3      3
## 4      1
## 5      1
## 6      3
## 7      3
## 8      1
## 9      2
## 10     3
## 11     2
## 12     3
## 13     2
## 14     2
## 15     3
## 16     3
## 17     3
## 18     3
## 19     3
## 20     2
```

#b Describe the data. Get the structure or the summary of the data

```
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 ...
## $ Father_Occupation : num  1 3 3 3 1 2 3 1 1 1 ...
## $ Person_At_Home    : num  5 7 3 8 5 9 6 7 8 4 ...
## $ Siblings_At_School: num  6 4 4 1 2 1 5 3 1 2 ...
## $ House_Type        : num  1 2 3 1 1 3 3 1 2 3 ...
```

```
summary(household_data)
```

```
## Respondents      Sex      Father_Occupation Person_At_Home
## 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
## Siblings_At_School House_Type
## 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
```

There are 20 observations in the data frame, and there are 6 variables. There are 20 respondents in the data. In Sex: A numeric variable with the values 1 and 2 designating the genders of men and women. #Father_Occupation: A numeric variable with the values 1, 2, and 3 may indicate the occupation level. #Person_At_Home is a numeric variable that indicates how many people live in each respondent's household. #Siblings_At_School is a numeric variable that indicates how many siblings the respondent has enrolled. #House_Type: A numeric variable with the values 1, 2, and 3 designates the various types of homes.

#c Is the mean number of siblings attending is 5?

```
mean_num_sib <- mean(household_data$Siblings_At_School)
mean_num_sib
```

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

#no, the mean number of siblings attending is 2.95

#d Extract the 1st two rows and then all the columns using the subsetting functions. Write the codes and

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

```
## Respondents Sex Father_Occupation Person_At_Home Siblings_At_School
## 1          1  2              1              5              6
## 2          2  2              3              7              4
## House_Type
## 1          1
## 2          2
```

#output

#Respondents Sex Father_Occupation Person_At_Home

```
#1          1  2              1              5
#2          2  2              3              7
```

#Siblings_At_School House_Type

```
#1          6          1
#2          4          2
```

#e Extract 3rd and 5th row with 2nd and 4th column. Write the codes and its result.

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

```
## Sex Person_At_Home
## 3  1              3
## 5  2              5
```

#result

Sex Person_At_Home

```
#3  1              3
#5  2              5
```

#f Select the variable types of houses then store the vector that results as types_houses. Write the cod

```
types_houses <- household_data$House_Type
types_houses
```

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

#g Select only all Males respondent that their father occupation was farmer. Write the codes and its ou

```
household_data[household_data$Sex == 1 & household_data$Father_Occupation == "farmer", ]
```

```
## [1] Respondents      Sex              Father_Occupation  Person_At_Home
## [5] Siblings_At_School House_Type
## <0 rows> (or 0-length row.names)
```

```
household_data$Father_Occupation
```

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

```
#output
```

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

```
#h Select only all females respondent that have greater than or equal to 5 number of siblings attending
```

```
female <- household_data[household_data$Siblings_At_School >= 5,]
```

```
female
```

```
## Respondents Sex Father_Occupation Person_At_Home Siblings_At_School
```

```
## 1 1 2 1 5 6
```

```
## 7 7 2 3 6 5
```

```
## 13 13 2 1 4 5
```

```
## 14 14 2 3 7 5
```

```
## 18 18 2 1 11 5
```

```
## House_Type
```

```
## 1 1
```

```
## 7 3
```

```
## 13 2
```

```
## 14 2
```

```
## 18 3
```

```
#output
```

```
# Respondents Sex Father_Occupation Person_At_Home
```

```
#1 1 2 1 5
```

```
#7 7 2 3 6
```

```
#13 13 2 1 4
```

```
#14 14 2 3 7
```

```
#18 18 2 1 11
```

```
# Siblings_At_School House_Type
```

```
#1 6 1
```

```
#7 5 3
```

```
#13 5 2
```

```
#14 5 2
```

```
#18 5 3
```

```
#2 Write a R program to create an empty data frame. Using the following codes:
```

```
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

#result
# the data frame is empty, this part of the report tells you that your data frame has 5 variables and
# Integer data type, column.
# Doubles; double data type for the column.
# Characters; character data type; column.
# Logicals; the data type of the column is logical (boolean).
#There are currently no unique levels in the column a factor variable # Factors with 0 levels.

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

#a Import the csv file into the R environment. Write the codes.
ot_respondent <- c(1:10)
  ot_sex <- c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male")
  ot_fathers_occupation <- c(1,2,3,3,1,2,2,3,1,3)
  ot_persons_at_home <- c(5,7,3,8,6,4,4,2,11,6)
  ot_siblings_at_school <- c(2,3,0,5,2,3,1,2,6,2)
  ot_type_of_house <- c("Wood", "Concrete", "Concrete", "Wood", "Semi-concrete", "Semi-concrete", "Semi-concrete", "Semi-concrete", "Semi-concrete", "Semi-concrete")

HouseholdData <- data.frame(
  Respondents = ot_respondent,
  Sex = ot_sex,
  FatherOccupation = ot_fathers_occupation,
  PersonAtHome = ot_persons_at_home,
  SiblingsAtSchool = ot_siblings_at_school,
  HouseType = ot_type_of_house
)

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

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

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

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

#output
#[1] 1 2 2 1 1 2 2 1 2 1

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

HouseholdData$HouseType <- factor(HouseholdData$HouseType, levels = c("Wood", "Concrete", "Semi-concrete"))
HouseholdData$HouseType <- as.integer(HouseholdData$HouseType)
HouseholdData$HouseType

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

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

#d. On father's occupation, factor it as Farmer = 1; Driver = 2; and Others = 3. What is the R code and

```
HouseholdData$FatherOccupation <- factor(HouseholdData$FatherOccupation, levels = c(1,2,3), labels = c(
HouseholdData$FatherOccupation
```

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

#output

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

#e Select only all females respondent that has a father whose occupation is driver. Write the codes and

```
female <- HouseholdData[HouseholdData$Sex == 2 & HouseholdData$FatherOccupation == "Driver",]
female
```

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

#output

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

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

```
five_res <- HouseholdData[HouseholdData$SiblingsAtSchool >= 5,]
five_res
```

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

#output

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

#4 Interpret the graph

#Sentiment Analysis 2020

On July 14, there were 2500 negative, 1500 neutral, and between 1500 and 2000 pleasant attitudes sta
Compared to the day before, there were 4000 more unfavorable tweets on July 15. Positive tweets witne
On July 17, the proportion of negative sentiments fell to a range of 3000-3500, that of neutral senti
On July 18, there were the same number of unfavorable comments as the day before, ranging between 300
On July 20, there were around 2500 fewer unfavorable tweets than there were the day before. Positive
From the previous day, on July 20, there were an additional 4100 tweets with negative sentiment on Ju
#Additional context and analysis of the tweets are required to determine the precise causes of these se