1. Create a data frame using the table below.

a. Write the codes.

```
myData <- data.frame(
  Respondents = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20),
  Sex = c(2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2, 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),
  Types_of_Houses = c(1, 2, 3, 1, 1, 3, 3, 1, 2, 3, 2, 3, 2, 2, 3, 3, 3, 3, 3, 2)
)
myData
```

```
##    Respondents Sex Fathers_Occupation Persons_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
##    Types_of_Houses
## 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(myData)
```

```
## 'data.frame':    20 obs. of  6 variables:
##  $ Respondents       : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ Sex               : num  2 2 1 2 2 2 2 2 2 2 ...
##  $ Fathers_Occupation: num  1 3 3 3 1 2 3 1 1 1 ...
##  $ Persons_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 ...
##  $ Types_of_Houses   : num  1 2 3 1 1 3 3 1 2 3 ...
```

The data has 20 observations from 6 variables. The variables are Respondents, Sex, Fathers Occupation, Person at home, Siblings at school, and Types of houses. c. Is the mean number of siblings attending is 5?

```
mean(myData$Siblings_at_School)
```

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

No, the mean is 2.95

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

```
first2 <- myData[c(1,2), ]
first2
```

```
##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1           1   2                  1               5                  6
## 2           2   2                  3               7                  4
##   Types_of_Houses
## 1               1
## 2               2
```

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

```
thirdand5 <- myData[c(3, 5), c(2, 4)]
thirdand5
```

```
##   Sex Persons_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 codes.

```
types_houses <- myData$Types_of_Houses
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 output.

```
allMalesandFarmer <- myData[myData$Sex == 1 & myData$Fathers_Occupation == 1, ]

allMalesandFarmer <- subset(myData, Sex == 1 & Fathers_Occupation == 1)
allMalesandFarmer
```

```
## [1] Respondents        Sex                Fathers_Occupation Persons_at_Home
## [5] Siblings_at_School Types_of_Houses
```

```
## <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 school. Write the codes and its outputs.

```
allFemalesGreater5 <- myData[myData$Sex == 2 & myData$Siblings_at_School >= 5, ]
allFemalesGreater5 <- subset(myData, Sex == 2 & Siblings_at_School >= 5)
allFemalesGreater5
```

```
##    Respondents Sex Fathers_Occupation Persons_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
##    Types_of_Houses
## 1                1
## 7                3
## 13               2
## 14               2
## 18               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
```

a. Describe the results. The output is an empty dataframe with 0 observations and 5 variables.

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

```
householdData <- read.csv("HouseholdData.csv")
householdData
```

```
##    Respondents    Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1            1   Male                  1               5                  2
## 2            2 Female                  2               7                  3
## 3            3 Female                  3               3                  0
## 4            4   Male                  3               8                  5
## 5            5   Male                  1               6                  2
## 6            6 Female                  2               4                  3
```

```
## 7            7 Female                 2              4                 1
## 8            8   Male                 3              2                 2
## 9            9 Female                 1             11                 6
## 10          10   Male                 3              6                 2
##    Types.of.Houses
## 1             Wood
## 2         Concrete
## 3         Concrete
## 4             Wood
## 5    Semi-Concrete
## 6    Semi-Concrete
## 7             Wood
## 8    Semi-Concrete
## 9    Semi-Concrete
## 10        Concrete
```

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

```
householdData$Sex <- factor(householdData$Sex, levels = c("Male", "Female"), labels = c(1, 2))
householdData$Sex <- as.integer(householdData$Sex)
householdData
```

```
##    Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1            1  NA                 1              5                 2
## 2            2   2                 2              7                 3
## 3            3   2                 3              3                 0
## 4            4   1                 3              8                 5
## 5            5   1                 1              6                 2
## 6            6   2                 2              4                 3
## 7            7   2                 2              4                 1
## 8            8   1                 3              2                 2
## 9            9   2                 1             11                 6
## 10          10   1                 3              6                 2
##    Types.of.Houses
## 1             Wood
## 2         Concrete
## 3         Concrete
## 4             Wood
## 5    Semi-Concrete
## 6    Semi-Concrete
## 7             Wood
## 8    Semi-Concrete
## 9    Semi-Concrete
## 10        Concrete
```

c. Convert the Type of Houses into factor and change it into integer. [Legend: Wood = 1; Congrete = 2; Semi-Congrete = 3]. Write the R codes and its output.

```
householdData$Types.of.Houses <- factor(householdData$Types.of.Houses, levels =c("Wood", "Concrete", "S
householdData$Types.of.Houses<- as.integer(householdData$Types.of.Houses)
householdData
```

```
##    Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1            1  NA                 1              5                 2
## 2            2   2                 2              7                 3
## 3            3   2                 3              3                 0
```

```
## 4            4   1                 3               8                 5
## 5            5   1                 1               6                 2
## 6            6   2                 2               4                 3
## 7            7   2                 2               4                 1
## 8            8   1                 3               2                 2
## 9            9   2                 1              11                 6
## 10          10   1                 3               6                 2
##    Types.of.Houses
## 1                1
## 2                2
## 3                2
## 4                1
## 5                3
## 6                3
## 7                1
## 8                3
## 9                3
## 10               2
```

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

```
householdData$Fathers.Occupation <- factor(householdData$Fathers.Occupation, levels =c(1, 2, 3), labels
householdData
```

```
##    Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1            1  NA             Farmer               5                 2
## 2            2   2             Driver               7                 3
## 3            3   2             Others               3                 0
## 4            4   1             Others               8                 5
## 5            5   1             Farmer               6                 2
## 6            6   2             Driver               4                 3
## 7            7   2             Driver               4                 1
## 8            8   1             Others               2                 2
## 9            9   2             Farmer              11                 6
## 10          10   1             Others               6                 2
##    Types.of.Houses
## 1                1
## 2                2
## 3                2
## 4                1
## 5                3
## 6                3
## 7                1
## 8                3
## 9                3
## 10               2
```

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

```
allFemalesandDriver <- householdData[householdData$Sex == "Female" & householdData$Fathers.Occupation ==
allFemalesandDriver <- subset(householdData, Sex == "Female" & Fathers.Occupation == 2)
allFemalesandDriver
```

```
## [1] Respondents        Sex                Fathers.Occupation Persons.at.Home
## [5] Siblings.at.School Types.of.Houses
```

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

f. Select the respondents that have greater than or equal to 5 number of siblings attending school. Write the codes and its output.

```
respGreater5 <- householdData[householdData$Siblings.at.School >= 5, ]
respGreater5 <- subset(householdData, Siblings.at.School >= 5)
respGreater5
```

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 4           4   1             Others               8                  5
## 9           9   2             Farmer              11                  6
##   Types.of.Houses
## 4               1
## 9               3
```

4. Interpret the graph. The x-axis shows several dates between July 14 and July 21, 2020, and within each date, the sentiment categories are shown Negative, Neutral, Positive. Red bars represent Negative sentiment tweets. Blue bars represent Positive sentiment tweets. Yellow bars represent Neutral sentiment tweets. The y-axis represents the count of tweets, indicating how many tweets fall under each sentiment category for each date. Across the dates, there are fluctuations in the sentiment of tweets, with some days experiencing more negative sentiment. The sentiment distribution also varies daily, with the number of neutral tweets being prominent on some days.