RWorksheet_Magallanes#3b

Killy Magallanes

2024-10-03

1. Create a data frame using the table below.

a. write the codes

```
data <- data.frame( respondents = c(1:20), Sex = c(2,2,1,2,2,2,2,2,2,2,1,2,2,2,2,2,2,2,1,2), fathers_oc
print(data)
```

```
##    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

```
summary(data)
```

```
##   respondents         Sex       fathers_occupation persons_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 types_of_houses
##  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
```

c. Is the mean number of siblings attending is 5? No, the mean number of siblings is 2.95

d. extract the 1st two rows and then all the columns using the subsetting functions. write the codes and its input

```
subset_data <- data[1:2,]
subset_data
```

```
##   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

```
subset_d <- data[c(3,5), c(2,4)]
subset_d
```

```
##   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 <- data$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

```
male_farmers <- data[data$Sex == 1 & data$fathers_occupation == 1,]
print(male_farmers)
```

```
## [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 output

```
female_siblings <- data[data$Sex == 2 & data$siblings_at_school >= 5, ]

female_siblings
```

```
##    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 followig 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
```

Ints=integer() will store integers. Doubles = double() will store numeric values. Characters = character() will store character strings. Logicals = logical(): will store logical (boolean) values, i.e., TRUE or FALSE. Factors = factor() will store factors stringsAsFactors = FALSE: This ensures that strings are not automatically converted to factors

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

a. Import the csv file into the R environment. write the codes.

```
Household_data <- read.csv("HouseholdData.csv")
print(Household_data)
```

```
##    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         congrete
## 3         congrete
## 4             wood
## 5    semi-congrete
## 6    semi-congrete
## 7             wood
## 8    semi-congrete
## 9    semi-congrete
## 10        congrete
```

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.

c. 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.

```
Household_data$Sex <- factor(Household_data$Sex, levels = c("male", "female"), labels = c(1, 2))

Household_data$Types.of.houses <- factor(Household_data$Types.of.houses, levels = c("wood", "congrete",
print(Household_data)
```

```
##    Respondents Sex Fathers.Occupation persons.at.home Siblings.at.school
## 1            1   1                  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

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

```
##    Respondents Sex Fathers.Occupation persons.at.home Siblings.at.school
## 1            1   1             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

```
female_drivers <- Household_data[Household_data$Sex == 2 & Household_data$Fathers.Occupation == "Driver
print(female_drivers)
```

```
##   Respondents Sex Fathers.Occupation persons.at.home Siblings.at.school
## 2           2   2             Driver               7                  3
## 6           6   2             Driver               4                  3
## 7           7   2             Driver               4                  1
##   Types.of.houses
## 2               2
## 6               3
## 7               1
```

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

```
respondents_with_siblings <- subset(Household_data, Siblings.at.school >= 5)
respondents_with_siblings
```

```
##   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 graph visualizes the changes in sentiment among tweets over a specific week. The spikes in negative sentiment, especially on July 15 and July 20, may indicate reactions to particular events or news stories, while the peak in positive sentiment on July 18 suggests a shift in public mood or response to certain developments.