RWorksheet_Pabriaga#3b

John Martin S. Pabriaga

2024-09-30

```
# 1.a
df <- data.frame(
  Respondents = 1:20,
  Sex = c(1, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 1, 2, 2, 1, 2),
  Fathers_Occupation = c(1, 3, 3, 3, 1, 2, 3, 1, 1, 1, 3, 2, 1, 3, 3, 1, 3, 2, 2, 1),
  Persons_at_Home = c(5, 7, 3, 8, 5, 9, 6, 7, 8, 4, 7, 5, 4, 7, 8, 1, 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)
)

df
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   1                  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   2                  3               7                  3
## 12          12   2                  2               5                  2
## 13          13   1                  1               4                  5
## 14          14   2                  3               7                  5
## 15          15   2                  3               8                  2
## 16          16   1                  1               1                  1
## 17          17   2                  3               3                  2
## 18          18   2                  2              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
```

```r
# 1.b
str(df)
```

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

```r
summary(df)
```

```
##   Respondents         Sex        Fathers_Occupation Persons_at_Home
##  Min.   : 1.00   Min.   :1.00   Min.   :1          Min.   : 1.00
##  1st Qu.: 5.75   1st Qu.:1.75   1st Qu.:1          1st Qu.: 4.75
##  Median :10.50   Median :2.00   Median :2          Median : 6.50
##  Mean   :10.50   Mean   :1.75   Mean   :2          Mean   : 6.05
##  3rd Qu.:15.25   3rd Qu.:2.00   3rd Qu.:3          3rd Qu.: 7.25
##  Max.   :20.00   Max.   :2.00   Max.   :3          Max.   :11.00
##  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
```

```r
# 1.c
mean(df$Siblings_at_School)
```

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

```r
# 1.d
df[1:2, ]
```

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

```r
# 1.e
df[c(3, 5), c(2, 4)]
```

```
##   Sex Persons_at_Home
## 3   1              3
## 5   2              5
```

```r
# 1.f
types_houses <- df$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
```

```r
# 1.g
subset(df, Sex == 1 & Fathers_Occupation == 1)
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   1                  1               5                  6
## 13          13   1                  1               4                  5
## 16          16   1                  1               1                  1
##    Types_of_Houses
## 1                1
## 13               2
## 16               3
```

```r
# 1.h
subset(df, Sex == 2 & Siblings_at_School >= 5)
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 7            7   2                  3               6                  5
## 14          14   2                  3               7                  5
## 18          18   2                  2              11                  5
##    Types_of_Houses
## 7                3
## 14               2
## 18               3
```

```r
# 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:"
```

```r
print(str(df))
```

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

```r
# 3
household_data <- data.frame(
  Respondents = 1:10,
  Sex = c("Male", "Female", "Female", "Male", "Male", "Female",
          "Female", "Male", "Female", "Male"),
  Fathers_Occupation = c(1, 2, 3, 3, 1, 2, 2, 1, 3, 1),
  Persons_at_Home = c(5, 7, 3, 8, 6, 9, 6, 7, 1, 6),
  Siblings_at_School = c(2, 3, 0, 5, 2, 3, 2, 1, 6, 3),
  Types_of_Houses = c("Wood", "Concrete", "Concrete", "Wood", "Semi-concrete",
                      "Semi-concrete", "Wood", "Semi-concrete", "Semi-concrete", "Concrete")
)

write.csv(household_data, "HouseholdData.csv", row.names = FALSE)

# 3.a
imported_data <- read.csv("HouseholdData.csv")
print(imported_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               9                  3
## 7            7 Female                  2               6                  2
## 8            8   Male                  1               7                  1
## 9            9 Female                  3               1                  6
## 10          10   Male                  1               6                  3
##    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
```

```r
# 3.b
imported_data$Sex <- factor(imported_data$Sex, levels = c("Male", "Female"), labels = c(1, 2))
print(imported_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               9                  3
## 7            7   2                  2               6                  2
## 8            8   1                  1               7                  1
## 9            9   2                  3               1                  6
```

```
## 10           10   1                    1                 6                    3
##     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
```

```r
# 3.c
imported_data$Types_of_Houses <- factor(imported_data$Types_of_Houses,
                                        levels = c("Wood", "Concrete", "Semi-concrete"),
                                        labels = c(1, 2, 3))
print(imported_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               9                  3
## 7             7   2                  2               6                  2
## 8             8   1                  1               7                  1
## 9             9   2                  3               1                  6
## 10           10   1                  1               6                  3
##     Types_of_Houses
## 1                 1
## 2                 2
## 3                 2
## 4                 1
## 5                 3
## 6                 3
## 7                 1
## 8                 3
## 9                 3
## 10                2
```

```r
# 3.d
imported_data$Fathers_Occupation <- factor(imported_data$Fathers_Occupation,
                                           levels = c(1, 2, 3),
                                           labels = c("Farmer", "Driver", "Others"))
print(imported_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               9                  3
## 7             7   2             Driver               6                  2
```

```
## 8               8    1          Farmer                7                    1
## 9               9    2          Others                1                    6
## 10             10    1          Farmer                6                    3
##    Types_of_Houses
## 1                1
## 2                2
## 3                2
## 4                1
## 5                3
## 6                3
## 7                1
## 8                3
## 9                3
## 10               2
```

```r
# 3.e
female_driver <- subset(imported_data, Sex == 2 & Fathers_Occupation == "Driver")
print(female_driver)
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 2            2   2             Driver               7                  3
## 6            6   2             Driver               9                  3
## 7            7   2             Driver               6                  2
##    Types_of_Houses
## 2                2
## 6                3
## 7                1
```

```r
# 3.f
siblings_school <- subset(imported_data, Siblings_at_School >= 5)
print(siblings_school)
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 4            4   1             Others               8                  5
## 9            9   2             Others               1                  6
##    Types_of_Houses
## 4                1
## 9                3
```

```r
# 4.
# The graph shows the daily distribution of tweet sentiments (Negative, Neutral, Positive)
# from July 14 to July 21, 2020. Negative sentiments (red) consistently dominate across all
# days, with notable peaks on July 15 and July 21. Positive sentiments (blue) are the second
# most frequent, showing a steady rise, particularly on July 21. Neutral sentiments (yellow)
# are generally the least represented and fluctuate more than the other categories. The chart
# indicates that during this period, the majority of tweets had negative sentiments, with
# positive sentiment increasing toward the end of the timeframe.
```