

3B

Dafnie Jayne Benajiba

2024-10-29

#a

```
df <- data.frame(
  Respondents = 1:20,
  Sex = c(2, 2, 1, 2, 2, 2, 2, 1, 2, 1, 1, 2, 1, 2, 1, 2, 2, 1, 2),
  Fathers_Occupation = c(1, 3, 3, 3, 1, 2, 3, 1, 1, 3, 3, 2, 1, 3, 1, 2, 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, 3, 1, 2, 1, 5, 3, 1, 2, 4, 2, 5, 2, 3, 1, 2, 5, 3, 2),
  Types_of_Houses = c(1, 2, 3, 1, 1, 1, 3, 1, 2, 3, 3, 2, 2, 2, 3, 1, 3, 3, 2, 2)
)

print(df)
```

##	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	3
## 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	1	1	7	3
## 9	9	2	1	8	1
## 10	10	1	3	4	2
## 11	11	1	3	7	4
## 12	12	2	2	5	2
## 13	13	1	1	4	5
## 14	14	2	3	7	2
## 15	15	1	1	8	3
## 16	16	2	2	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	1				

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

#b The dataset includes 20 respondents, mostly female, with varied father occupations (farmer, driver, or other). Households typically have 5-8 members, with several siblings in school (1-6). Housing types range from wooden to concrete, with concrete being the most common.

#b

```
str(df)
```

```
## '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 1 2 1 ...
## $ Fathers_Occupation: num 1 3 3 3 1 2 3 1 1 3 ...
## $ Persons_at_Home : num 5 7 3 8 5 9 6 7 8 4 ...
## $ Siblings_at_School: num 6 4 3 1 2 1 5 3 1 2 ...
## $ Types_of_Houses : num 1 2 3 1 1 1 3 1 2 3 ...
```

```
summary(df)
```

```
## Respondents      Sex      Fathers_Occupation Persons_at_Home
## Min.   : 1.00   Min.   :1.00   Min.   :1      Min.   : 3.0
## 1st Qu.: 5.75   1st Qu.:1.00   1st Qu.:1      1st Qu.: 5.0
## Median :10.50   Median :2.00   Median :2      Median : 7.0
## Mean   :10.50   Mean   :1.65   Mean   :2      Mean   : 6.4
## 3rd Qu.:15.25   3rd Qu.:2.00   3rd Qu.:3      3rd Qu.: 8.0
## Max.   :20.00   Max.   :2.00   Max.   :3      Max.   :11.0
## Siblings_at_School Types_of_Houses
## Min.   :1.00      Min.   :1.00
## 1st Qu.:2.00      1st Qu.:1.00
## Median :2.50      Median :2.00
## Mean   :2.85      Mean   :2.05
## 3rd Qu.:4.00      3rd Qu.:3.00
## Max.   :6.00      Max.   :3.00
```

#c

```
mean_siblings <- mean(df$Siblings_at_School)
mean_siblings
```

```
## [1] 2.85
```

```
#d
```

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

```
## 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
```

```
extracted_rows <- df[c(3, 5), c(2, 4)]  
extracted_rows
```

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

```
#F
```

```
types_houses <- df$Types_of_Houses  
types_houses
```

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

```
#G
```

```
males_farmers <- df[df$Sex == 1 & df$Fathers_Occupation == 1, ]  
males_farmers
```

```
## Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School  
## 8          8 1              1              7              3  
## 13         13 1              1              4              5  
## 15         15 1              1              8              3  
## Types_of_Houses  
## 8          1  
## 13         2  
## 15         3
```

```
#H
```

```
females_with_siblings <- df[df$Sex == 2 & df$Siblings_at_School >= 5, ]  
females_with_siblings
```

```
##      Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1           1    2              1              5              6
## 7           7    2              3              6              5
## 18          18    2              1             11              5
##      Types_of_Houses
## 1              1
## 7              3
## 18             3
```

#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:"
```

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

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

#3-B

```
respondents_data <- read.csv("~/RBasics/CS101_DataScience/worksheet3b/HouseholdData.csv")
respondents_data
```

```
##      Respondents      Sex Father.s.Occupation Person.at.Home Sibling.at.Home
## 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 Female              3              6              2
##      Types.of.House
## 1              Wood
## 2             Congrete
## 3             Congrete
## 4              Wood
```

```
## 5    Semi-concrete
## 6    Semi-concrete
## 7           Wood
## 8    Semi-concrete
## 9    Semi-concrete
## 10   Concrete
```

#3-B

```
respondents_data$Sex <- factor(respondents_data$Sex, levels = c("Male", "Female"), labels = c(1, 2))
respondents_data$Sex
```

```
## [1] 1 2 2 1 1 2 2 1 2 2
## Levels: 1 2
```

#3-C

```
respondents_data$Types.of.House <- factor(respondents_data$Types.of.House, levels = c("Wood", "Concrete", "Other"))
respondents_data$Types.of.House
```

```
## [1] 1    <NA> <NA> 1    <NA> <NA> 1    <NA> <NA> <NA>
## Levels: 1 2 3
```

#3-D

```
respondents_data$Father.s.Occupation <- factor(respondents_data$Father.s.Occupation, levels = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10))
respondents_data$Father.s.Occupation <- as.integer(respondents_data$Father.s.Occupation)
respondents_data$Father.s.Occupation
```

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

#3-E.

```
females_driver <- respondents_data[respondents_data$Sex == 2 & respondents_data$Father.s.Occupation == 1, ]
females_driver
```

```
##   Respondents Sex Father.s.Occupation Person.at.Home Sibling.at.Home
## 2           2   2                   2              7              3
## 6           6   2                   2              4              3
## 7           7   2                   2              4              1
##   Types.of.House
## 2           <NA>
## 6           <NA>
## 7           1
```

#3-F

```
siblings <- respondents_data[respondents_data$Sibling.at.Home >= 5, ]
siblings
```

##	Respondents	Sex	Father.s.Occupation	Person.at.Home	Sibling.at.Home
## 4	4	1		3	8
## 9	9	2		1	11
##	Types.of.House				
## 4		1			
## 9		<NA>			

#4 The graph shows that from July 14 to July 21, 2020, negative tweets are consistently the highest, peaking on July 15 and 21. Positive tweets peak on July 18, while neutral tweets fluctuate, generally falling between negative and positive counts.