

RWorksheet 3b

Porras, Dorlyne Fate

2022-11-22

1. Create a data frame using the table below.

a. Write the codes.

```
Respondents <- c(seq(1,20))
Sex <- c(2,2,1,2,2,2,2,2,2,2,1,2,2,2,2,2,2,1,2)
FathersOccupation <- c(1,3,3,3,1,2,3,1,1,1,3,2,1,3,3,1,3,1,2,1)
Personsathome <- c(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6)
Siblingsatschool <- c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2)
Typesofhouses <- c(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2)

dataf <- data.frame(Respondents,Sex,FathersOccupation,Personsathome,
                    Siblingsatschool,Typesofhouses)
```

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

```
summary(dataf)
```

##	Respondents	Sex	FathersOccupation	Personsathome
##	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
##	Siblingsatschool	Typesofhouses		
##	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?

```
mean(Siblingsatschool)
```

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

Answer: NO

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

```
rows2 <- subset(dataf[c(3,5),c(2,4)])
rows2
```

```
## Sex Personsathome
```

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

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

```
rows3 <- subset(dataf[c(3,5),c(2,4)])
rows3
```

```
## Sex Personsathome
## 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.

```
rows4 <- dataf[c(6)]
type_houses <- rows4
```

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

```
Males <- subset(dataf[c(3,11),c(2,3)])
Males
```

```
## Sex FathersOccupation
## 3 1 3
## 11 1 3
```

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.

```
rows5 <- subset(dataf[c(1:20), c(2,5)])
Female <- rows5[dataf$Siblingsatschool >= 5,]
Female
```

```
## Sex Siblingsatschool
## 1 2 6
## 7 2 5
## 13 2 5
## 14 2 5
## 18 2 5
```

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:") print(str(df))

a. Describe the results. The result is the print of Structure of the empty dataframe is printed. The data.frame have 0 object but have a 5 variables. The variables are Ints or int, Doubles or num, Characters that is chr, Logicals or logi and Factors w/0 levels and the result is NULL.

3. Interpret the graph. The graph is all about The Sentiments Tweets per Day - Donal Trump. In the graph there have a three sentiments, the Negative- The red one, the Neutral- The yellow, and lastly, the Positive- The bulue one. So the graph shows the tweet per day It starts in July 14,2000 to July 21, 2000. The graph shows that the day July 15 and 21 have the more positive in that day.