LA.3.R

Sara (local)

2021-02-16

```
## LA.3: Examining Data
## Student Name

## Load packages
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------ tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```
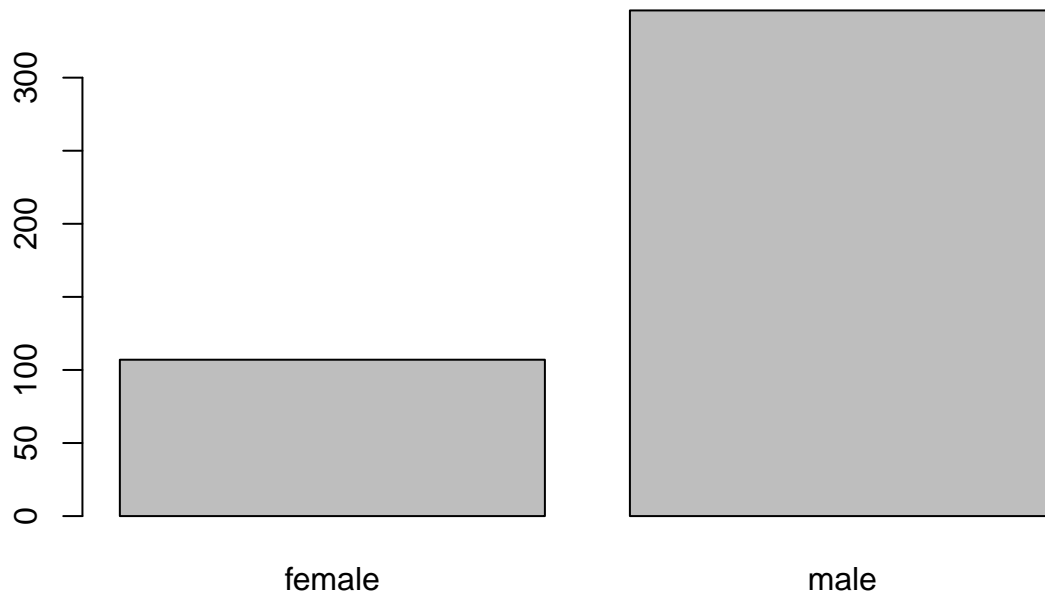
```
## -- Conflicts --------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(descr)

## Set working directory
setwd("C:/Users/SaraK/Downloads/sarakyeo.github.io/COMM-3710/LA")

## 1. Load dataset
hdata <- read.csv("HELP.csv") # load HELP.csv
        # loaded data of 453 observations of 27 variables

## 2. Make frequency tables for sex and d1
freq(hdata$sex) # freq table for sex
```
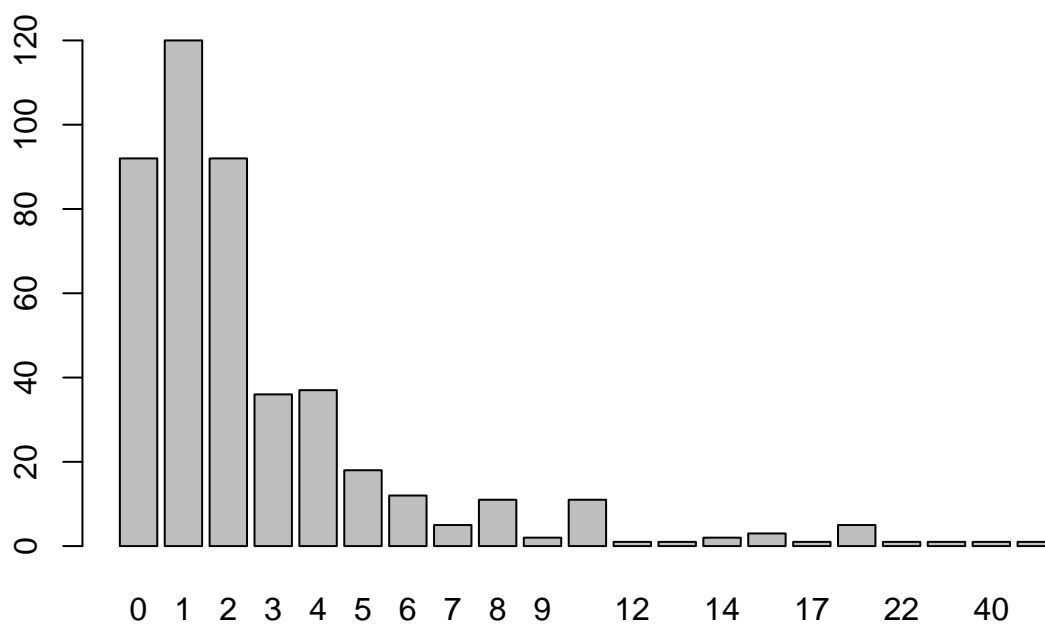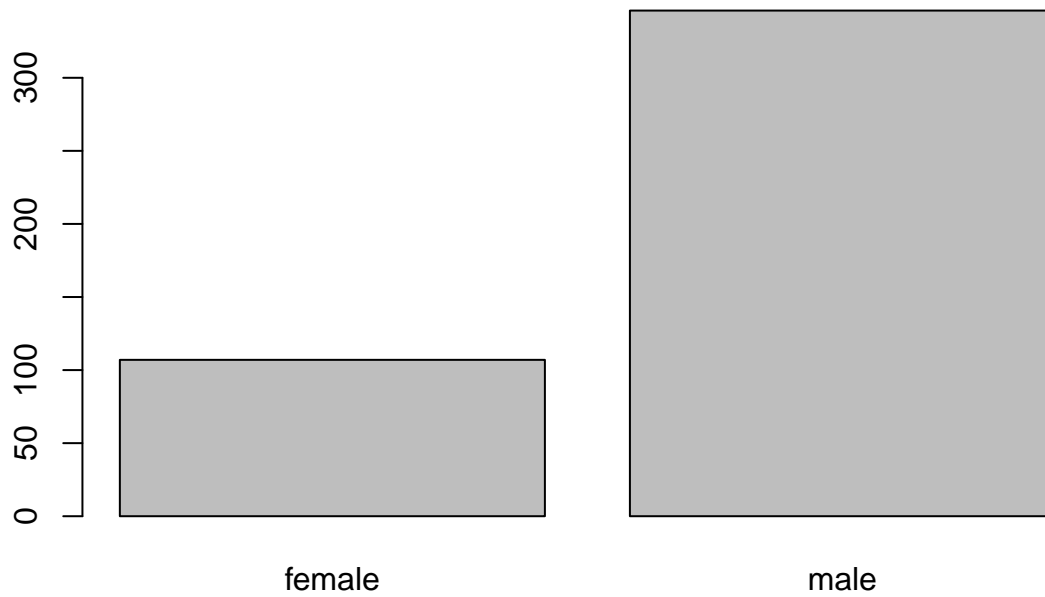
```
## hdata$sex
##        Frequency Percent
## female       107   23.62
## male         346   76.38
## Total        453  100.00
```

```
freq(hdata$d1) # freq table for d1
```

```
## hdata$d1
##       Frequency  Percent
## 0           92  20.3091
## 1          120  26.4901
## 2           92  20.3091
## 3           36   7.9470
## 4           37   8.1678
## 5           18   3.9735
## 6           12   2.6490
## 7            5   1.1038
## 8           11   2.4283
## 9            2   0.4415
## 10          11   2.4283
## 12           1   0.2208
## 13           1   0.2208
## 14           2   0.4415
## 15           3   0.6623
## 17           1   0.2208
## 20           5   1.1038
## 22           1   0.2208
## 36           1   0.2208
## 40           1   0.2208
## 100          1   0.2208
## Total      453 100.0000
```
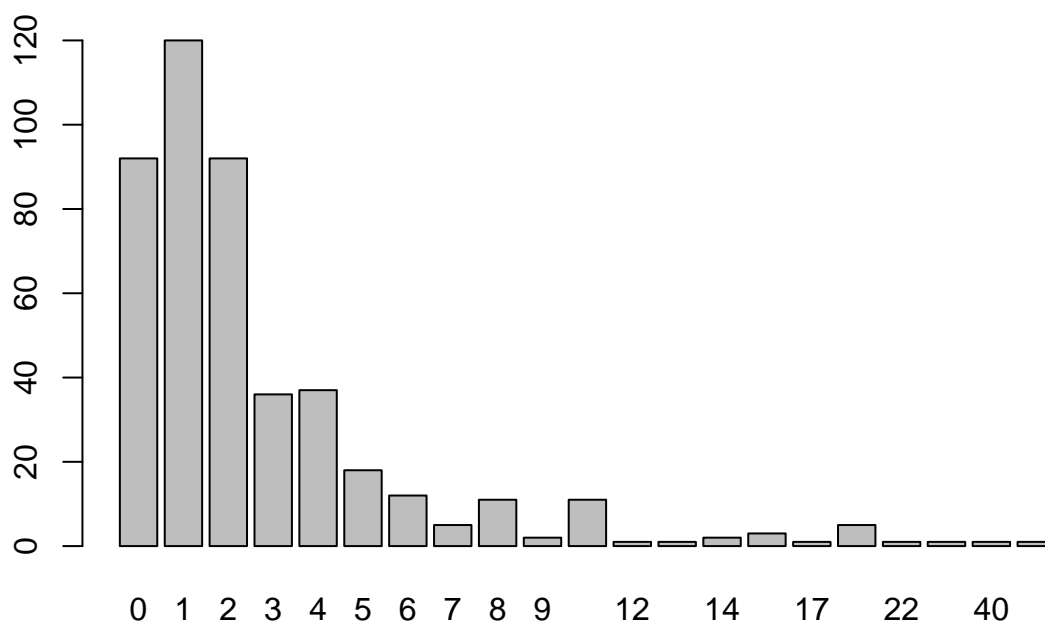
```
## 2a. How many patients in the study are female?
freq(hdata$sex) # 107 female patients
```



```
## hdata$sex
##         Frequency Percent
## female        107   23.62
## male          346   76.38
## Total         453  100.00
```

```
## 2b. How many have never been hospitalized for medical problems?
freq(hdata$d1) # 92 have never been hospitalized for medical problems
```

```
## hdata$d1
##       Frequency  Percent
## 0            92  20.3091
## 1           120  26.4901
## 2            92  20.3091
## 3            36   7.9470
## 4            37   8.1678
## 5            18   3.9735
## 6            12   2.6490
## 7             5   1.1038
## 8            11   2.4283
## 9             2   0.4415
## 10           11   2.4283
## 12            1   0.2208
## 13            1   0.2208
## 14            2   0.4415
## 15            3   0.6623
## 17            1   0.2208
## 20            5   1.1038
## 22            1   0.2208
## 36            1   0.2208
## 40            1   0.2208
## 100           1   0.2208
## Total       453 100.0000
```
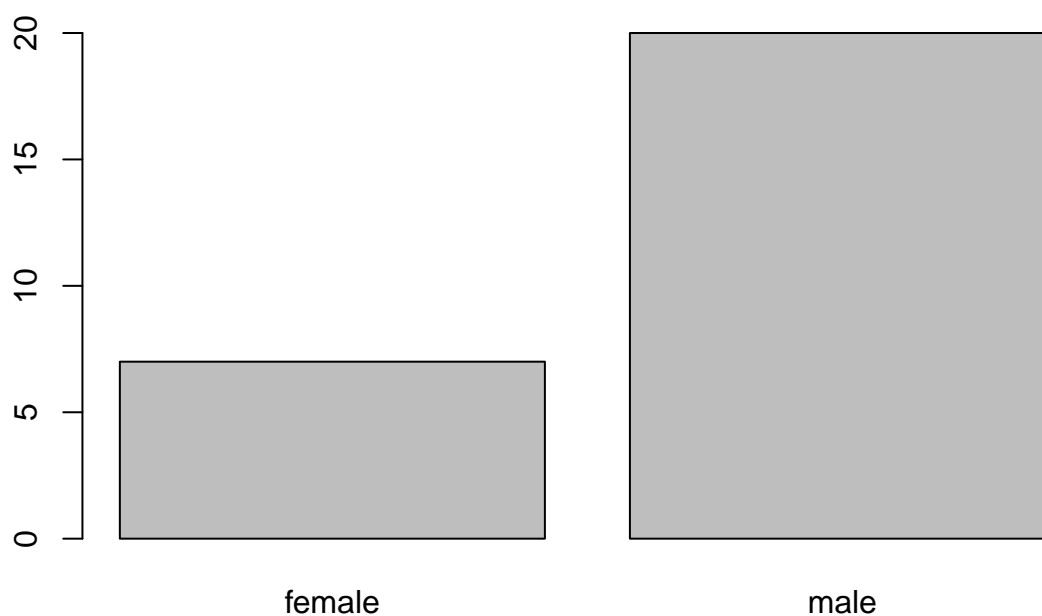
```
## 2c. What percentage of patients in the study have been hospitalized fewer
##     than 5 times?
lessfive <- (92 + 120 + 92 + 36 + 37) / 453
lessfive # 83.2%
```

```
## [1] 0.8322296
```

```
## 3. Subset data to incl substance == cocaine and over 40 years old
hsub <- subset(hdata, substance == "cocaine" & age >= 40) # subset data
        # contains 27 obs of 27 var

## 4. Make freq table based on subset
freq(hsub$sex) # freq table for sex (subset of data)
```



```
## hsub$sex
##          Frequency Percent
## female           7   25.93
## male            20   74.07
## Total           27  100.00
```

```
## 4a. 27 patients are at least 40 years old and have cocaine listed as his/her
##     primary abuse substance.

## 4b. 74.07% patients are at least 40 years old, have cocaine listed as his/her
##     primary abuse substance, and are male.
```