Exercises 5

Maria Cuellar

2024-09-20

Exercises 5

```r
# install.packages("tidyverse")
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

1. Load the data called domestic violence

```r
dat <- read_csv("data/domestic_violence.csv") # make sure data is in the right folder
```

```
## Rows: 347 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (4): Education, Employment, Marital status, Violence
## dbl (3): SL. No, Age, Income
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Answer: Loaded the data.

2. What type of stat variable is Employment?

```r
dat$Employment # base R version
```

```
##   [1] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##   [5] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##   [9] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [13] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [17] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [21] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [25] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [29] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [33] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [37] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [41] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [45] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [49] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [53] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [57] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [61] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [65] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [69] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [73] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [77] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [81] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [85] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [89] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [93] "unemployed"    "unemployed"    "unemployed"    "unemployed"
##  [97] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [101] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [105] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [109] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [113] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [117] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [121] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [125] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [129] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [133] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [137] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [141] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [145] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [149] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [153] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [157] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [161] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [165] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [169] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [173] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [177] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [181] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [185] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [189] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [193] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [197] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [201] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [205] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [209] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [213] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [217] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [221] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [225] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [229] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [233] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [237] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [241] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [245] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [249] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [253] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [257] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [261] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [265] "unemployed"    "unemployed"    "unemployed"    "unemployed"
## [269] "unemployed"    "unemployed"    "semi employed" "unemployed"
## [273] "semi employed" "semi employed" "semi employed" "semi employed"
## [277] "semi employed" "semi employed" "employed"      "semi employed"
## [281] "semi employed" "semi employed" "semi employed" "employed"
## [285] "semi employed" "semi employed" "semi employed" "semi employed"
## [289] "semi employed" "semi employed" "employed"      "semi employed"
## [293] "semi employed" "semi employed" "semi employed" "semi employed"
## [297] "semi employed" "employed"      "semi employed" "semi employed"
## [301] "semi employed" "semi employed" "semi employed" "semi employed"
## [305] "employed"      "semi employed" "semi employed" "semi employed"
## [309] "semi employed" "semi employed" "semi employed" "semi employed"
## [313] "employed"      "semi employed" "semi employed" "semi employed"
## [317] "semi employed" "semi employed" "semi employed" "semi employed"
## [321] "employed"      "employed"      "semi employed" "semi employed"
## [325] "employed"      "employed"      "employed"      "semi employed"
## [329] "employed"      "employed"      "employed"      "employed"
## [333] "employed"      "employed"      "employed"      "employed"
## [337] "employed"      "employed"      "employed"      "employed"
## [341] "employed"      "employed"      "semi employed" "employed"
## [345] "unemployed"    "unemployed"    "unemployed"
```

```r
dat %>% select(Employment) # tidyverse version, pick out the variable Employment to look at it
```

```
## # A tibble: 347 x 1
##    Employment
##    <chr>
##  1 unemployed
##  2 unemployed
##  3 unemployed
##  4 unemployed
##  5 unemployed
##  6 unemployed
##  7 unemployed
##  8 unemployed
##  9 unemployed
## 10 unemployed
## # i 337 more rows
```

Answer: Employment is a categorical variable.

3. What categories does it have?

```r
levels(as.factor(dat$Employment)) # base R version
```

```
## [1] "employed"      "semi employed" "unemployed"
```

```
dat <- dat %>% mutate(Employment=as.factor(Employment)) # tidyverse version, make Employment a factor

levels(dat$Employment) # read the levels of the factor
```

```
## [1] "employed"      "semi employed" "unemployed"
```

Answer: Employed, semi employed, and unemployed.

4. Do quantitative EDA for Employment.

```
table(dat$Employment)
```

```
##
##      employed semi employed    unemployed
##            26            47           274
```

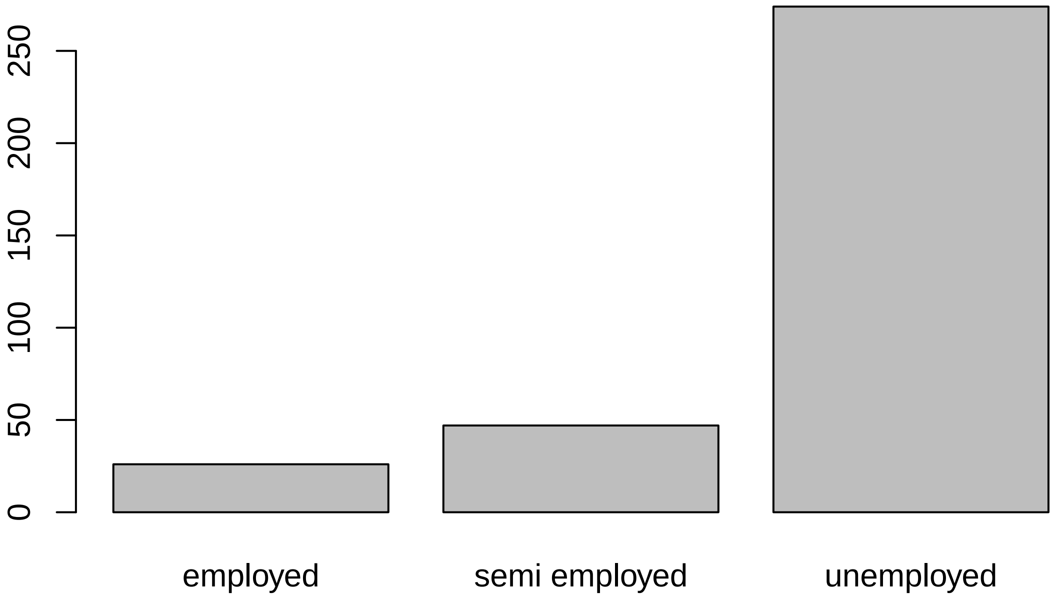
```
dat %>%
  count(Employment) %>%
  mutate(prop = prop.table(n)) # make a table of counts, this one includes proportions.
```

```
## # A tibble: 3 x 3
##   Employment        n   prop
##   <fct>         <int>  <dbl>
## 1 employed         26 0.0749
## 2 semi employed    47 0.135
## 3 unemployed      274 0.790
```

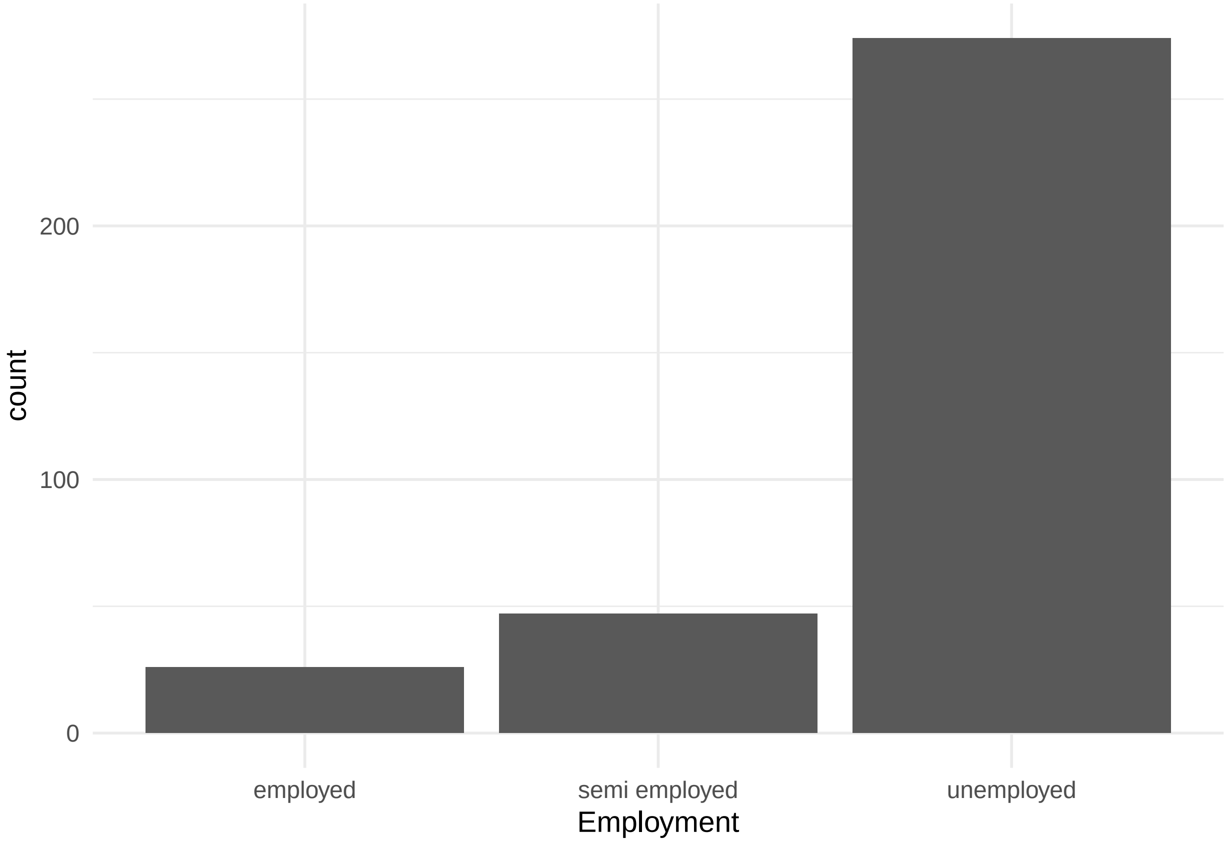
Answer: I made a table of Employment, and I can see the three categories and their respective counts, as well as their proportions.

5. Do visual EDA for Employment.

```
barplot(table(dat$Employment)) # base R
```



```r
dat %>%
  ggplot(aes(x=Employment)) +
  geom_bar() +
  theme_minimal()
```



Answer: Made a barplot.

6. What kind of variables are `Marital status` (and why does it have single quotes around it) and Violence?

Categorical, and `Marital status` has quotes because it has a space in the name.

7. Make a contingency table of both Marital status and Violence

```r
addmargins(table(dat$`Marital status`, dat$Violence)) # use base-R to make this. Is there a better way?
```

```
##
##              no yes Sum
##   married   217  83 300
##   unmarred   44   3  47
##   Sum       261  86 347
```

Answer: Made a contingency table of the two categorical variables, with margins.