Assignment_4

Evan Z-G

2023-10-26

A common task is to take a set of data that has multiple categorical variables and create a table of the number of cases for each combination. An introductory statistics textbook contains a dataset summarizing student surveys from several sections of an intro class. The two variables of interest for us are Gender and Year which are the students gender and year in college.

Download the dataset and correctly order the Year variable using the following:

Survey <- read.csv('https://www.lock5stat.com/datasets3e/StudentSurvey.csv', na.strings=c(',' ')) Using some combination of dplyr functions, produce a data set with eight rows that contains the number of responses for each gender:year combination. Make sure your table orders the Year variable in the correct order of First Year, Sophmore, Junior, and then Senior. You might want to look at the following functions: dplyr::count and dplyr::drop_na.

Using tidyr commands, produce a table of the number of responses in the following form:

Gender First Year Sophmore Junior Senior Female
Male

```
Survey <- read.csv('https://www.lock5stat.com/datasets3e/StudentSurvey.csv', na.strings=c('',' '))

head(Survey)
```

```
##          Year Sex Smoke   Award HigherSAT Exercise TV Height Weight Siblings
## 1    Senior   M    No Olympic      Math       10  1     71    180        4
## 2 Sophomore   F   Yes Academy      Math        4  7     66    120        2
## 3 FirstYear   M    No   Nobel      Math       14  5     72    208        2
## 4    Junior   M    No   Nobel      Math        3  1     63    110        1
## 5 Sophomore   F    No   Nobel    Verbal        3  3     65    150        1
## 6 Sophomore   F    No   Nobel    Verbal        5  4     65    114        2
##   BirthOrder VerbalSAT MathSAT  SAT  GPA Pulse Piercings
## 1          4       540     670 1210 3.13    54         0
## 2          2       520     630 1150 2.50    66         3
## 3          1       550     560 1110 2.55   130         0
## 4          1       490     630 1120 3.10    78         0
## 5          1       720     450 1170 2.70    40         6
## 6          2       600     550 1150 3.20    80         4
```

```
finaltable <- Survey %>%
  drop_na() %>%
  mutate(Year = factor(Year, levels=c('FirstYear', 'Sophomore', 'Junior', 'Senior')) ) %>% count(Year, S
finaltable
```

```
##          Year Sex Year and Sex
## 1 FirstYear    F           36
## 2 FirstYear    M           43
## 3 Sophomore    F           90
```

```
## 4 Sophomore   M           89
## 5    Junior   F           15
## 6    Junior   M           16
## 7    Senior   F           10
## 8    Senior   M           26
```

```
Temps <- read.csv('https://raw.githubusercontent.com/dereksonderegger/444/master/data-raw/FlagMaxTemp.c
```

```
Temps2 <- Temps %>% select(-X) %>% filter(Year == 2005) %>% pivot_longer(!Year:Month, names_to = "Dates"
Temps2
```

```
## # A tibble: 372 x 4
##     Year Month Dates Temp
##    <int> <int> <chr> <chr>
##  1  2005     1 X1    37.94
##  2  2005     1 X2    37.94
##  3  2005     1 X3    39.02
##  4  2005     1 X4    33.98
##  5  2005     1 X5    35.06
##  6  2005     1 X6    26.96
##  7  2005     1 X7    33.08
##  8  2005     1 X8    28.04
##  9  2005     1 X9    35.96
## 10  2005     1 X10   44.06
## # i 362 more rows
```

```
Temps2
```

```
## # A tibble: 372 x 4
##     Year Month Dates Temp
##    <int> <int> <chr> <chr>
##  1  2005     1 X1    37.94
##  2  2005     1 X2    37.94
##  3  2005     1 X3    39.02
##  4  2005     1 X4    33.98
##  5  2005     1 X5    35.06
##  6  2005     1 X6    26.96
##  7  2005     1 X7    33.08
##  8  2005     1 X8    28.04
##  9  2005     1 X9    35.96
## 10  2005     1 X10   44.06
## # i 362 more rows
```