Stat 6021: Addressing Guided Question Set 2

Tom Lever

09/02/22

For this exercise, use the `.csv` data file that you created at the end of the previous guided question set, `new_students.csv`. As a reminder, the dataset contains information on students taking an introductory statistics class at a large public university. The columns of data are:

```
* `Student`: ID number on survey
* `Gender`: gender of student (male / female)
* `Smoke`: whether the student smokes (yes / no)
* `Marijuan`: whether the student smokes marijuana (yes / no)
* `DrivDrnk`: whether the student has ever driven while drunk (yes / no)
* `GPA`: student's current GPA
* `PartyNum`: number of days per month the student parties
* `DaysBeer`: number of days per month the student has at least two alcoholic drinks
* `StudyHrs`: number of hours spent studying per week
* `PartyAnimal`: whether the students parties more than 8 days per month (yes / no)
* `GPA.cat`: "low" if GPA is less than 3.0, "moderate" if GPA is at least 3.0 and less than 3.5, "high"
```

```
students_dataframe <- read.csv(
    file = "../../Module_1--Data_Wrangling/Guided_Question_Set/new_students.csv"
)
head(students_dataframe, n = 3)
```

```
##   Student Gender Smoke Marijuan DrivDrnk  GPA PartyNum DaysBeer StudyHrs
## 1       1 female    No      Yes      Yes 3.40        4        6        7
## 2       2 female    No       No       No 3.45        4        0       20
## 3       3   male    No       No      Yes 3.89        9        4       30
##   PartyAnimal  GPA.cat
## 1          no moderate
## 2          no moderate
## 3         yes     high
```

1. Produce a frequency table of the number of students in each level of `GPA.cat`. If needed, be sure to arrange the order of the output appropriately. How many students are in each level of `GPA.cat`?

   ```
   table(factor(students_dataframe$GPA.cat, labels = c("low", "moderate", "high")))
   ```

   ```
   ##
   ##      low moderate     high
   ##       70       87       85
   ```