Stat 6021: Guided Question Set 2

For this excercise, 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 the data are:

- `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 2 alcoholic drinks
- `StudyHrs`: number of hours spent studying per week
- `PartyAnimal`: whether the student parties more than 8 days a 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" if GPA at least 3.5.

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`?
2. Produce a bar chart that summarizes the number of students in each level of `GPA.cat`. Be sure to add appropriate labels and titles so that the bar chart conveys its message clearly to the reader. Be sure to remove the bar corresponding to the missing values.
3. Create a similar bar chart as you did in part 2, but with proportions instead of counts. Be sure to remove the bar corresponding to the missing values.

4. Produce a frequency table for the number of female and male students and the GPA category.

5. Produce a table for the percentage of GPA category for each gender. For the percentages, round to 2 decimal places. Comment on the relationship between gender and GPA category.

6. Create a bar chart to explore the proportion of GPA categories for female and male students. Be sure to remove the bar corresponding to the missing values.

7. Create a similar bar chart similar to the bar chart in part 6, but split by smoking status. Comment on this bar chart.

8. Create a scatterplot of GPA against the amount of hours spent studying a week. How would you describe the relationship between GPA and amount of time spent studying?

9. Edit the scatterplot from part 8 to include information about the number of days the student parties in a month.

10. Edit the scatterplot from part 9 to include information about whether the student smokes or not.