Lab 4 *Some of these problems may be more challenging than others. Please feel free to work with others, attend office hours, or post on the course discussion forum if you need help. While collaboration with other students is encouraged, each student is responsible for submitting his or her own work. This assignment should be submitted in one well-commented SAS program. For any questions that require a written answer, do so in the SAS comments. Be sure to re-name the uploaded SAS scripts according to the naming convention* `LastnameFirstinitial_Lab#.sas` (*e.g.,* `PileggiS_Lab4.sas`).

This data set comes from a STAT 217 course survey. Questions marked with MC indicates it was a multiple choice question. All other questions were free response.

`Stat217Survey.csv`:

`Q02`	MC: What year are you?
`Q03a`	MC: Have you taken a statistics course previously? Mark all that apply. *This is my first statistics course.* 1 = yes, 0 = no
`Q03b`	MC: Have you taken a statistics course previously? Mark all that apply. *I took AP Statistics in high school.* 1 = yes, 0 = no
`Q03c`	MC: Have you taken a statistics course previously? Mark all that apply. *I have previously taken STAT 130 at Cal Poly.* 1 = yes, 0 = no
`Q03d`	MC: Have you taken a statistics course previously? Mark all that apply. *I have previously taken STAT 217 at Cal Poly.* 1 = yes, 0 = no
`Q03e`	MC: Have you taken a statistics course previously? Mark all that apply. *I have taken some other statistics course.* 1 = yes, 0 = no
`Q04`	What is your GPA? (If this happens to be your first quarter, and you don't yet have a GPA, enter 9.99).
`Q05`	MC: Do you consider yourself to be tech-savvy and handy with your computer?
`Q06`	MC: Do you own a laptop?
`Q07`	MC: What is your targeted grade in this course?
`Q08`	MC: What kind of note taker are you?
`Q09`	Please tell me something about yourself. This can be involvement in extracurricular activities, interests outside of school, or something unique about yourself.
`Q10`	To the nearest inch, how tall are you? (I am 5 feet 4 inches, so I would enter 64.)
`Q11`	MC: What is your gender pronoun?
`Q12`	How much money did you spend on your last haircut (including tip)? (Enter the dollar amount without the dollar sign.)
`Q13`	How many siblings do you have?
`Q14`	What is the length in months of your longest serious relationship?
`Q15`	MC: Are you currently in a serious relationship?
`Q16`	MC: Was Cal Poly your first choice?

`Q17`	How many colleges did you apply to?
`Q18`	About how many text messages do you send in a day?
`Q19`	MC: Rate your opinion on the value of statistics in society on a numerical scale of 1 (completely useless) to 9 (incredibly important).

1. Create a SAS library reference called `flash` that points a location on your flash drive or computer where you want to save your SAS data set.
2. Identify the `Stat217Survey.csv` file from either PolyLearn or the shared drive. Save the file to a location on your computer or your flash drive.
3. Import the `Stat217Survey.csv` to a temporary SAS data set named `work.survey`. *Hint: Use GUESSINGROWS = 35 to be able to read in the data set without error.* Apply the CONTENTS procedure to `work.survey` to get an overview of the data set, with the variables listed in order of appearance rather than alphabetical. Lastly, apply the PRINT procedure so that you can review the data.

For the remaining items: All DATA step operations should be completed in **ONE** DATA step. You should contribute pieces of code in small chunks, verify that it worked correctly, and then continue to add more to the DATA step. Some items may require a PROC to verify output. Please include the PROCs sequentially after your DATA step, commented with the corresponding question number and objective. (The first PROC after the data step should correspond to question 5, the last PROC in your program should correspond to question 16.)

4. Create a permanent SAS data set called `survey` in your `flash` library. This data set should be a copy of the `work.survey` data set.
5. *Under the data step, using your permanent SAS data set:* Using the procedure of your choice, identify any unusual values of GPA. Explain why this(these) values are present. Again, using the procedure of your choice, identify how many students submitted this(these) value(s) and note your findings as a comment in your SAS code.
6. *In your data step:* Create a new variable called `GPA_clean` that is a copy of the GPA variable. Re-code the unusual values that you identified in the previous question to missing.
7. *Under the data step, using your permanent SAS data set:* Apply PROC MEANS to the `GPA_clean` variable to verify that the re-coding worked correctly. Your output should match the output shown below.

Question 7 Output

The MEANS Procedure

Analysis Variable : GPA_clean				
N	Mean	Std Dev	Minimum	Maximum
26	3.0624231	0.4726480	2.0000000	3.9000000

8. *In your data step:* There are a series of question 3 items that corresponded to a single "Mark all that apply" type question. Utilize these responses to create a new variable called `prev_stats` which has a value of `yes` if students have previous experience with statistics and a value of `no` if the student does not have previous experience with statistics. *Hint: it may be helpful to discuss a 'game plan' with neighboring students or the instructor prior to coding.*

9. *Under the data step, using your permanent SAS data set:* Apply PROC FREQ to the `prev_stats` variable to determine how many students had previous experience with statistics. Your output should match the output shown below.

Question 9 Output

The FREQ Procedure

prev_stats	Frequency	Percent	Cumulative Frequency	Cumulative Percent
no	11	31.43	11	31.43
yes	24	68.57	35	100.00

10. *Under the data step, using your permanent SAS data set:* Utilize a procedure that summarizes all of the values regarding year at Cal Poly.

11. *In your data step:* Create a new variable called `class` that classifies students as "lower" class (first years and second years) and "upper" class (third years, fourth years, etc.).

12. *Under the data step, using your permanent SAS data set:* Apply PROC FREQ to the `class` variable to determine how many lower and upper class students are enrolled. Your output should match the output shown below.

Question 11 Output

The FREQ Procedure

class	Frequency	Percent	Cumulative Frequency	Cumulative Percent
lower	26	74.29	26	74.29
upper	9	25.71	35	100.00

13. *In your data step:* According to the registrar's office, graduation with honors are awarded as follows:

    - *Summa cum laude:* cumulative Cal Poly GPA of 3.85 or higher
    - *Magna cum laude:* cumulative Cal Poly GPA of 3.70 to 3.84
    - *Cum laude:* cumulative Cal Poly GPA of 3.50 to 3.69

    Use the `GPA_clean` variable to create a new variable called `honors` that classifies students according to their current GPA; students who do not yet achieve honors should be classified as "none". Students with a missing value for `GPA_clean` should also have a missing value for `honors`.

14. *Under the data step, using your permanent SAS data set:* Apply PROC FREQ to the `honors` variable to determine the distribution of honors. Your output should match the output shown below.

Question 14 Output

The FREQ Procedure

honors	Frequency	Percent	Cumulative Frequency	Cumulative Percent
	9	25.71	9	25.71
Cum laude	2	5.71	11	31.43
Magna cum laude	2	5.71	13	37.14
Summa cum laude	1	2.86	14	40.00
none	21	60.00	35	100.00

15. When I ask students to tell me something about themselves, many say that they enjoy the outdoors. Identify a SAS function that can look for phrases in a character string. Use this function to create a new variable called `outdoors` which should have a value of "yes" if their statement includes the word "outdoors" and a value of "no" otherwise.

16. *Under the data step, using your permanent SAS data set:* Apply PROC FREQ to

the `outdoors` variable to determine how many students mentioned the outdoors in their statement. Your output should match the output shown below.

Question 16 Output

The FREQ Procedure

outdoors	Frequency	Percent	Cumulative Frequency	Cumulative Percent
no	32	91.43	32	91.43
yes	3	8.57	35	100.00

17. *After your last PROC:* Create a new, temporary SAS data set called `shortlist` that contains only the five students with honors and only two variables: honors and how they take notes. Apply a PROC PRINT to this data set to view it. Your output should match the output shown below.

Question 17 Output

Obs	Q08	honors
1	I take ok notes.	Cum laude
2	I take very detailed notes.	Magna cum laude
3	I take very detailed notes.	Magna cum laude
4	I take very detailed notes.	Summa cum laude
5	I take very detailed notes.	Cum laude