Lab 18 *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_Lab18.sas`).

Notes: (1) For this lab you do not need to do any statistical tests or interpret any p-values - I am only asking for descriptive statistics. (2) You will need to research different kinds of SAS informats to complete this lab. (3) You are not allowed to modify the raw data files!

1. Store the computer location of your data sets as a macro variable.
2. In 1983 New York city presented evidence in court that Brink's (a collections contractor) employees had been stealing parking meter money - delivering to the city less than the total collections. The data consists of the amount of money collected each month by Brinks and a control sample that was taken by the city.

   `Data_parking.dat`
   Date
   $ collected by state of NY
   $ collected by city of NY
   Brinks (1) or other contractor (0)

   (a) Read the data into a temporary data set in SAS, using the macro variable you created in (1).

   (b) Use a SAS procedure to verify that all variables are numeric (including date).

   (c) Visually check that your date read in correctly by verifying that the first observation has a SAS date value of 6330. (No SAS code/comment required.)

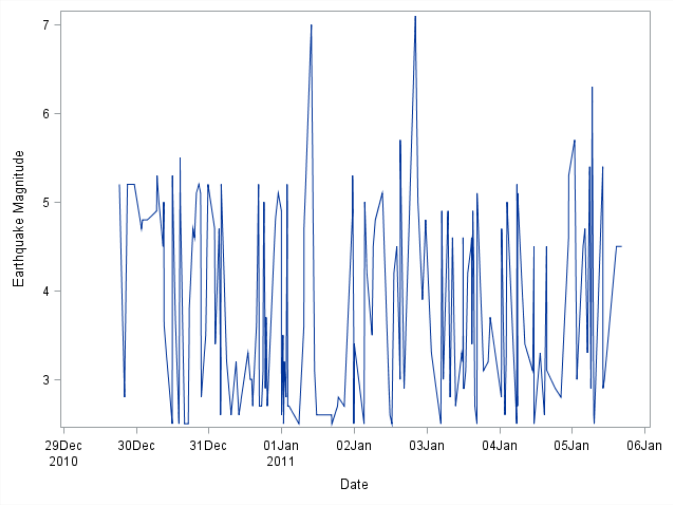
   (d) In one PROC, calculate the average amount of money collected by the city of NY among Brink's contractors and among other contractors. Note the averages in a comment in your SAS code and identify which is higher.
3. We have information on some earthquakes from December 2010 and January 2011.

   `Earthquake.dat`
   Map name
   Magnitude
   Date and time
   Latitude
   Longitude
   Depth
   Location

(a) Read the data into a temporary data set in SAS, using the macro variable you created in (1). *Hint: In SAS, dates and times can be read into one variable - a date/time variable.*

(b) Use a SAS procedure to verify that only map name and location are character variables - all other variables (including date and time) should be numeric.

(c) Use the `series` statement in `PROC SGPLOT` to re-create the figure below exactly, which shows a time series of the magnitude of the earthquakes.



(d) Recall that the `where` statement in SAS can use additional operators compared to the `if` statement (Lecture 6). Using a `PROC` with a `where` statement, identify how many earthquakes occurred in (1) Mexico, and (2) Alaska. (This can be two separate PROCs, one for each location.) State your findings in a comment in your SAS code.

4. The Rose Bowl is a college tradition that dates back to the early 1900's.

`rosebowl.dat`

Game date, [c] Team 1, Team 1 score, [c] Team 2, Team 2 score

Note: the two variables with the [c] *symbol should be read in as character variables.*

(a) Open the `rosebowl.dat` in a text editor such as notepad. Describe the features that you see in the data that you will need to address when reading it in to SAS. Note your findings as a comment in your SAS code.

(b) Read the data into SAS using the macro variable you created in question (1).

(c) Use a SAS procedure to verify that your variables read in correctly as character or numeric.

(d) Create a new variable that indicates the following possible game outcomes: "team 1 wins", "team 2 wins", and "tie".

(e) Use a SAS procedure to print the games which resulted in a tie. Apply a format to the game date variable so that it displays as a calendar date rather than a SAS date value.

(f) Use any PROCs and/or DATA steps necessary to create a variable that represents the cumulative total number of games won by team 1. Do not include ties in this cumulative sum.

(g) Create a data set which only has one observation per team 1 and that only keeps the variables team 1 name and cumulative total number of wins. Print this data set. So that you may check your work, the first three observations are displayed below.

first 3 observations

Obs	team1	num_wins
1	Alabama	4
2	Arizona State	1
3	California	2

first 3 observations

5. Ben & Jerry's is a popular ice cream manufacturer that was founded in 1978.

`BenAndJerrys.dat`

[c] flavor name, portion size (g), calories, calories from fat, fat (g), saturated fat (g), trans fat (g), cholesterol (mg), sodium (mg), total carbohydrate (g), [c] dietary fiber (g), sugars (g), protein (g), [c] year introduced, [c] year retired, [c] content description, [c] notes

Note: the six variables with the [c] symbol should be read in as character variables.

(a) Open the `BenAndJerrys.dat` in a text editor such as notepad. Describe the features that you see in the data that you will need to address when reading it in to SAS. Note your findings as a comment in your SAS code.

(b) Read the data into SAS using the macro variable you created in question (1). Use the `INFILE` statement

```
INFILE "&path.BenAndJerrys.dat" DLM="," DSD ;
```

which contains options discussed in the extra material (not required) of Lecture 18.

(c) Use a SAS procedure to verify that your variables read in correctly as character or numeric.

(d) Subset the data to keep only flavors that that can be purchased at the grocery store. A flavor that is "Scoop Shop Exclusive" can only be purchased in the Ben & Jerry's ice cream store; otherwise, the flavor can be purchased at the grocery store.

(e) Recreate the following plot. Your axis labels do not need to be displayed exactly as seen below (the figure below just uses the variable names for axis labels).