

22S:166
Computing in Statistics
Data Preparation Using SAS

Lecture 16
 Oct. 29, 2007

Kate Cowles
 374 SH, 335-0727
 kcowles@stat.uiowa.edu

- We wish to report the following for all the NADP/NTN sites in Colorado:
 1. site id code
 2. elevation
 3. sulfuric acid deposition in the year 2000 in kg/ha
 4. actual number of days measurements contributing to year 2000 total
 5. scatterplot of sulfuric acid deposition vs. elevation
- We go to NADP/NTN data access at <http://nadp.sws.uiuc.edu/>
- We download year 2000 annual data.

Example: Acid rain deposition in Colorado

The National Atmospheric Deposition Program/National Trends Network (NADP/NTN) is a nationwide network of precipitation monitoring sites. The network is a cooperative effort between many different groups, including the State Agricultural Experiment Stations, U.S. Geological Survey, U.S. Department of Agriculture, and numerous other governmental and private entities. For a full list of contributors, see the collaborating agencies page. The NADP/NTN has grown from 22 stations at the end of 1978, our first year, to over 200 sites spanning the continental United States, Alaska, and Puerto Rico, and the Virgin Islands.

The purpose of the network is to collect data on the chemistry of precipitation for monitoring of geographical and temporal long-term trends. The precipitation at each station is collected weekly according to strict clean-handling procedures. It is then sent to the Central Analytical Laboratory where it is analyzed for hydrogen (acidity as pH), sulfate, nitrate, ammonium, chloride, and base cations (such as calcium, magnesium, potassium and sodium). Our excellent quality assurance programs ensure that the data remain accurate and precise.

NATIONAL ATMOSPHERIC DEPOSITION PROGRAM/NTN

Site ID: 4313 Date Range: 2000 to 2000 Report Date: 9/12/2001 3:11:54 PM

"SiteID"	"Summ.Per."	"Year"	"Crit1"	"Crit2"	"Crit3"	"Crit4"	"Ca"
"C000"	"Annual"	2000	82.0	93.0	90.0	78.0	0.34
"C001"	"Annual"	2000	90.0	100.0	86.0	98.0	1.07
"C002"	"Annual"	2000	77.0	100.0	69.0	21.0	9.72
"C008"	"Annual"	2000	91.0	100.0	99.0	83.0	0.97
"C010"	"Annual"	2000	81.0	100.0	85.0	84.0	0.47
"C015"	"Annual"	2000	70.0	100.0	80.0	77.0	0.73
"C019"	"Annual"	2000	83.0	100.0	94.0	91.0	0.81
"C021"	"Annual"	2000	85.0	100.0	94.0	106.0	0.69
"C022"	"Annual"	2000	85.0	100.0	90.0	87.0	0.74
"C091"	"Annual"	2000	71.0	100.0	74.0	65.0	2.22
"C092"	"Annual"	2000	89.0	100.0	90.0	88.0	1.16
"C093"	"Annual"	2000	79.0	100.0	77.0	58.0	1.09
"C094"	"Annual"	2000	94.0	100.0	98.0	94.0	0.74
"C096"	"Annual"	2000	73.0	100.0	78.0	83.0	1.22
"C097"	"Annual"	2000	77.0	100.0	85.0	44.0	1.58
"C098"	"Annual"	2000	68.0	100.0	69.0	62.0	1.66
"C099"	"Annual"	2000	84.0	100.0	85.0	103.0	1.13

- Note:
 - elevation is not in file
 - header info in different format from actual records

Two ways to run SAS

- batch mode
 - use text editor to write a file containing SAS code
 - extension “.sas”; e.g. filename might be “conadp.sas”
 - run it in batch mode by entering “sas <filename>”; e.g. “sas conadp”
 - SAS produces two new output files
 - * log file with .log extension
 - * output file with .lst extension
- interactive mode
 - just enter “sas”
 - program editor, log, and output are three windows
 - online help available

The DATA step: Creating a SAS dataset

Before it can process data, SAS must read in the data in the form of a table with

- a row for each *observation*
- a column for each *variable*

You must choose a name for the entire dataset and a name for each variable.

SAS distinguishes between two types of variables:

- *numeric variables*, which contain only digits and decimal points and with which arithmetic operations may be done
- *character variables* (all other kinds of data).

How SAS programs and commands are organized

- Use a *DATA step* to organize your data by creating a SAS dataset.
- Then use *PROC steps* to analyze your data using SAS procedures.
 - Once you have created a SAS dataset, you may apply any SAS procedures to it during the current SAS session without recreating the dataset.
- DATA and PROC steps consist of SAS *statements*.
 - Each statement must end with a semicolon.
 - Most statements include one or more *keywords* that must be spelled exactly as shown.

Using options before the data step to control printed output

Put this line at the beginning of every SAS program if you want output to print correctly on 8-1/2 by 11 inch paper:

```
options linesize = 75 ;
```

Use additional options to suppress dates and page numbers at the beginning of each page.

```
options linesize = 75 pagesize = 60 nodate nonumber;
```

SAS data step

- *data* statement
 - must begin the data step
 - gives name by which to refer to dataset during this SAS run
- *infile* statement
 - tells SAS in what physical file to find data
 - may optionally give SAS information about file and how to read it

SAS code to read in this file

```
options linesize = 75 pagesize = 60 nodate ;

data depo ;
infile 'depoRepCO.asp' firstobs = 6 ;
input SiteID $ Per $ Year Crit1 Crit2 Crit3 Crit4 Ca Mg K Na
NH4 NO3 InorgN Cl S04 HLab HField Svol Ppt Pct ValidF ValidL
Days Date1 $ Date2 $ ;
run;

proc print data = depo (obs=5) ;
run ;
```

- *input* statement
 - gives names to all variables
 - variables are assumed to be numeric unless you put a \$ after name in input statement
 - may contain informats (goes *after* variable name to which it applies)
 - may contain @ <column-number> (goes *before* variable name to which it applies)
- may contain additional statements
 - formats for printing values of variables
 - calculating new variables
 - etc.
- ends with *run* statement

Log file

```
NOTE: SAS initialization used:
      real time          0.55 seconds
      cpu time           0.11 seconds

1      options linesize = 75 pagesize = 60 ;
2
3      data depo ;
4      infile 'depoRepCO.asp' firstobs = 6 ;
5      input SiteID $ Per $ Year Crit1 Crit2 Crit3 Crit4 Ca Mg K Na
6      ! NH4 NO3
7      ! InorgN Cl S04 HLab HField Svol Ppt Pct ValidF ValidL Days
8      ! Date1 $
9      Date2 $ ;
10     run;
```

```
NOTE: The infile 'depoRepCO.asp' is:
      File Name=/tmp_mnt/space/kcowles/166/lectures/lect1mkc/depoRepCO.asp,
      Owner Name=kcowles,Group Name=faculty,
      Access Permission=rw-----,
      File Size (bytes)=4300

NOTE: 17 records were read from the infile 'depoRepCO.asp'.
      The minimum record length was 228.
      The maximum record length was 228.
      One or more lines were truncated.

NOTE: The data set WORK.DEPO has 17 observations and 26 variables.
NOTE: DATA statement used:
      real time          0.10 seconds
      cpu time           0.05 seconds
```

```
10 proc print data = depo (obs=5) ;
11 run ;
```

```
NOTE: There were 5 observations read from the dataset WORK.DEPO.
NOTE: The PROCEDURE PRINT printed page 1.
NOTE: PROCEDURE PRINT used:
      real time          0.04 seconds
      cpu time           0.03 seconds
```

Output of print

The SAS System 1
09:34 Thursday, September 13, 2001

Obs	SiteID	Per	Year	Crit1	Crit2	Crit3	Crit4	Ca	Mg
1	"C000"	"Annual"	2000	82	93	90	78	0.34	0.029
2	"C001"	"Annual"	2000	90	100	86	98	1.07	0.076
3	"C002"	"Annual"	2000	77	100	69	21	9.72	0.886
4	"C008"	"Annual"	2000	91	100	99	83	0.97	0.098
5	"C010"	"Annual"	2000	81	100	85	84	0.47	0.048

Obs	K	Na	NH4	NO3	Inorg		S04	HLab	HField	Svol
					N	Cl				
1	0.099	0.119	0.51	1.79	0.80	0.17	1.18	0.01	0.02	6849.6
2	0.103	0.138	1.36	3.09	1.76	0.19	2.06	0.00	0.01	13377.7
3	1.020	1.879	6.55	33.60	12.68	2.44	19.32	0.22	0.34	26264.1
4	0.128	0.152	0.57	3.93	1.33	0.26	2.13	0.03	0.04	28189.3
5	0.091	0.140	0.48	3.94	1.26	0.29	1.85	0.05	0.05	26739.0

Obs	Ppt	Pct	Valid		Days	Date1	Date2
			F	L			
1	13.79	51	22	11	367	"12/28/1"	"12/05/2"
2	23.85	83	27	22	371	"12/28/1"	"01/02/2"
3	268.38	62	38	32	371	"12/28/1"	"01/02/2"
4	49.05	70	43	29	371	"12/28/1"	"01/02/2"
5	53.82	11	36	6	370	"12/29/1"	"01/02/2"

Fixing the truncation and misreading of last two variables

```
data depo ;
infile 'depoRepCO.asp' firstobs = 6 ;
input SiteID $ Per $ Year Crit1 Crit2 Crit3 Crit4 Ca Mg K Na NH4
NO3 InorgN Cl S04 HLab HField Svol Ppt Pct ValidF ValidL Days
@201 Date1 $12. @217 Date2 $12. ;
run;
```

Obs	K	Na	NH4	NO3	Inorg		S04	HLab	HField	Svol
					N	Cl				
1	0.099	0.119	0.51	1.79	0.80	0.17	1.18	0.01	0.02	6849.6
2	0.103	0.138	1.36	3.09	1.76	0.19	2.06	0.00	0.01	13377.7
3	1.020	1.879	6.55	33.60	12.68	2.44	19.32	0.22	0.34	26264.1
4	0.128	0.152	0.57	3.93	1.33	0.26	2.13	0.03	0.04	28189.3
5	0.091	0.140	0.48	3.94	1.26	0.29	1.85	0.05	0.05	26739.0

Obs	Ppt	Pct	Valid		Days	Date1	Date2
			F	L			
1	13.79	51	22	11	367	"12/28/1999"	"12/05/2000"
2	23.85	83	27	22	371	"12/28/1999"	"01/02/2001"
3	268.38	62	38	32	371	"12/28/1999"	"01/02/2001"
4	49.05	70	43	29	371	"12/28/1999"	"01/02/2001"
5	53.82	11	36	6	370	"12/29/1999"	"01/02/2001"

SAS Informats and Dropping Variables

- We can use the “drop” statement to eliminate unneeded variables from SAS’s internal dataset.
 - does *not* affect external file
 - saves memory for SAS processing
- We can use “informat” to read in data in different formats from how it was stored in the external file.
- Learning about SAS informats from on-line help
 - SAS system help — Help on SAS software products — Base SAS — Using Base SAS — SAS Language — SAS Formats and Informats

```

data depo ;
infile 'depoRepCO.asp' firstobs = 6 ;
input SiteID $QUOTE6. @9 Per $QUOTE8. Year Crit1 Crit2 Crit3 Crit4
Ca Mg K Na NH4 NO3 InorgN Cl SO4 HLab HField Svoll Ppt Pct
ValidF ValidL Days @201 Date1 $QUOTE12. @217 Date2 $QUOTE12. ;
drop Per Year Crit1 - Crit4 ;
run;

```

Obs	Site			Inorg								
	ID	Ca	Mg	K	Na	NH4	NO3	N	Cl	SO4	HLab	HField
1	C000	0.34	0.029	0.099	0.119	0.51	1.79	0.80	0.17	1.18	0.01	0.02
2	C001	1.07	0.076	0.103	0.138	1.36	3.09	1.76	0.19	2.06	0.00	0.01
3	C002	9.72	0.886	1.020	1.879	6.55	33.60	12.68	2.44	19.32	0.22	0.34
4	C008	0.97	0.098	0.128	0.152	0.57	3.93	1.33	0.26	2.13	0.03	0.04
5	C010	0.47	0.048	0.091	0.140	0.48	3.94	1.26	0.29	1.85	0.05	0.05

Obs	Svoll	Ppt	Pct	Valid		Days	Date1	Date2
				F	L			
1	6849.6	13.79	51	22	11	367	12/28/1999	12/05/2000
2	13377.7	23.85	83	27	22	371	12/28/1999	01/02/2001
3	26264.1	268.38	62	38	32	371	12/28/1999	01/02/2001
4	28189.3	49.05	70	43	29	371	12/28/1999	01/02/2001
5	26739.0	53.82	11	36	6	370	12/29/1999	01/02/2001

```

data depo ;
infile 'depoRepCO.asp' firstobs = 6 ;
input SiteID $QUOTE6. @9 Per $QUOTE8. Year Crit1 Crit2 Crit3 Crit4 Ca Mg
K Na NH4 NO3 InorgN Cl SO4 HLab HField Svoll Ppt Pct ValidF ValidL Da
drop Per Year Crit1-Crit4 ;
sdate = input(Date1, mmddyy10.) ; * convert from char var to date var ;
edate = input(Date2, mmddyy10.) ;
daysop = edate - sdate ;
format sdate edate date8. ;
run ;

```

Here is the output for the later variables in the dataset when format statement is not used.

Obs	S	v	P	P	i	i	a	t	e	e	t	t	o	s	l	t	t	F	L	s	1	2	e	e	p		
																										V	V
1	6849.6	13.79	51	22	11	367	12/28/1999	12/05/2000	14606	14949	343																
2	13377.7	23.85	83	27	22	371	12/28/1999	01/02/2001	14606	14977	371																
3	26264.1	268.38	62	38	32	371	12/28/1999	01/02/2001	14606	14977	371																
4	28189.3	49.05	70	43	29	371	12/28/1999	01/02/2001	14606	14977	371																
5	26739.0	53.82	11	36	6	370	12/29/1999	01/02/2001	14607	14977	370																

Computing the days: SAS date arithmetic

- SAS stores dates as numeric variables so it can compute number of days between different dates
- SAS uses “informats” for reading dates that appear in data files in different standard formats.
- Our data file stored dates as character variables. We will use the “input” function and the “mmddyy10.” informat to convert the character variables into a date variables and copy the values into new variables.
- Then we need to use a “format” to *print* the new variables as a meaningful date.

Output when format statement is used.

Obs	Date2	sdate	edate	daysop
1	12/05/2000	12/28/1999	12/05/2000	343
2	01/02/2001	12/28/1999	01/02/2001	371
3	01/02/2001	12/28/1999	01/02/2001	371
4	01/02/2001	12/28/1999	01/02/2001	371
5	01/02/2001	12/29/1999	01/02/2001	370

Other document

[Image] Home AIRMoN MDN Search
 Data Maps QA Sponsors Overview Contacts

[Image] NADP/NTN Sites in CO

Click on either a dot (site location) or site label in Colorado to access data or other site-specific information.
 Active sites are indicated by a solid black dot.
 Inactive sites have a hollow dot and a slightly smaller site label.
 Create a customized list for multiple-site data retrievals
 Access data for all sites in Colorado
 [Image]

Site ID	Site Name	Start Date	End Date	Elevation (meters)
Active Sites				
C000	Alamosa	4/22/1980		2298
C001	Las Animas Fish Hatchery	10/4/1983		1213
C002	Niwot Saddle	6/5/1984		3520
C008	Four Mile Park	12/29/1987		2502
C010	Gothic	2/2/1999		2926
C015	Sand Spring	3/20/1979		1998
C019	Rocky Mtn National Park-Beaver Meadows	5/29/1980		2490

C021	Manitou	10/17/1978		2362
C022	Pawnee	5/22/1979		1641
C091	Wolf Creek Pass	5/26/1992		3292
C092	Sunlight Peak	1/13/1988		3206
C093	Dry Lake	10/14/1986		2527
C094	Sugarloaf	11/4/1986		2524
C096	Molas Pass	7/29/1986		3249
C097	Buffalo Pass	2/7/1984		3234
C098	Rocky Mtn National Park-Loch Vale	8/16/1983		3159
C099	Mesa Verde National Park	4/28/1981		2172
Inactive Sites				
C095	Engineer Mountain Guard Station	7/29/1986	1/2/1990	2758

 Your Comments and Suggestions are always Welcome
 Return to :[NADP Home] [AIRMoN] [MDN] [Search]

First stab at reading in these data

```
data sites ;
infile 'stateCO.asp' firstobs = 19 ;
input @13 SiteID $ @20 sitename $18. @40 strtdate mmdyy10.
@53 stopdate mmdyy10. @68 elev ;
run ;
```

```
proc print data = sites ;
run ;
```

From log file

```
19 data sites ;
20 infile 'stateCO.asp' firstobs = 19 ;
21 input @13 SiteID $ @20 sitename $18. @40 strtdate mmdyy10. @53
21 ! stopdate mmdyy10. @68 elev ;
22 run ;
```

NOTE: The infile 'stateCO.asp' is:
 File Name=/tmp_mnt/space/kcowles/166/lectures/lect1mkc/stateCO.asp,
 Owner Name=kcowles,Group Name=faculty,
 Access Permission=rw-----,
 File Size (bytes)=2493

```
NOTE: Invalid data for elev in line 59 13-16.
RULE: ----+----1-----2-----3-----4-----5-----6-----
59 C095 Engineer Mountain 7/29/1986 1/2/1990
66 2758 71
SiteID=Park sitename= strtdate=. stopdate=. elev=. _ERROR_=1 _N_=18
10:51 Thursday, September 13, 2001
```

NOTE: Invalid data for strtdate in line 61 40-49.

```
NOTE: Invalid data for stopdate in line 61 53-62.
NOTE: Invalid data for elev in line 61 68-77.
RULE: ----+----1-----2-----3-----4-----5-----6-----
61 -----
66 ----- 77
SiteID=Guard sitename=----- strtdate=. stopdate=. elev=. _ERROR_=1
_N_=19
NOTE: Invalid data for strtdate in line 62 40-49.
NOTE: Invalid data for stopdate in line 63 1-10.
NOTE: LOST CARD.
SiteID=Comments sitename=ts and Suggestions strtdate=. stopdate=. elev=.
_ERROR_=1 _N_=20
NOTE: 45 records were read from the infile 'stateCO.asp'.
The minimum record length was 0.
The maximum record length was 77.
NOTE: SAS went to a new line when INPUT statement reached past the end of
a line.
NOTE: The data set WORK.SITES has 19 observations and 5 variables.
NOTE: DATA statement used:
real time 2.19 seconds
cpu time 0.09 seconds
```

Output

Obs	SiteID	sitename	strtdte	stopdate	elev
1	C000	Alamosa	7417	.	2298
2	C001	Las Animas Fish	8677	.	1213
3	Hatchery		.	.	3520
4	C008	Four Mile Park	10224	.	2502
5	C010	Gothic	14277	.	2926
6	C015	Sand Spring	7018	.	1998
7	C019	Rocky Mtn National	7454	.	2490
8	Park-Bea	Park-Beaver Meadow	.	.	2362
9	C022	Pawnee	7081	.	1641
10	C091	Wolf Creek Pass	11834	.	3292
11	C092	Sunlight Peak	10239	.	3206
12	C093	Dry Lake	9783	.	2527
13	C094	Sugarloaf	9804	.	2524
14	C096	Molas Pass	9706	.	3249
15	C097	Buffalo Pass	8803	.	3234
16	C098	Rocky Mtn National	8628	.	3159
17	Park-Loc		.	.	2172
18	Park		.	.	.
19	Guard	-----	.	.	.

Combining the two datasets to produce the scatterplot

- Note that observations are in exactly the same order in the two datasets.
 - i.e. we want to match the first “site” record with the first “depo” record, the second with the second, etc.
- In this case only, we can use a one-to-one merge.

```
data combined ;
merge depo sites ;
run ;
```

```
proc print ;
run ;
```

Using “misover” and the “subsetting if” in data steps

```
data sites ;
infile 'stateCO.asp' firstobs = 19 misover ;
input @13 SiteID $ @20 sitename $18. @40 strtdte mmdyy10. @53 stopdate mmddyy10. ;
if strtdte ne . ; * subsetting if: exclude observations meeting condition
format strtdte stopdate date8. ;
run ;
```

Obs	Site ID	sitename	strtdte	stopdate	elev
1	C000	Alamosa	22APR80	.	2298
2	C001	Las Animas Fish	04OCT83	.	1213
3	C002	Niwot Saddle	05JUN84	.	3520
4	C008	Four Mile Park	29DEC87	.	2502
5	C010	Gothic	02FEB99	.	2926
6	C015	Sand Spring	20MAR79	.	1998
7	C019	Rocky Mtn National	29MAY80	.	2490
8	C021	Manitou	17OCT78	.	2362
9	C022	Pawnee	22MAY79	.	1641
10	C091	Wolf Creek Pass	26MAY92	.	3292
11	C092	Sunlight Peak	13JAN88	.	3206
12	C093	Dry Lake	14OCT86	.	2527
13	C094	Sugarloaf	04NOV86	.	2524
14	C096	Molas Pass	29JUL86	.	3249
15	C097	Buffalo Pass	07FEB84	.	3234
16	C098	Rocky Mtn National	16AUG83	.	3159
17	C099	Mesa Verde Nationa	28APR81	.	2172
18	C095	Engineer Mountain	29JUL86	02JAN90	2758

Log file

```
27 data combined ;
28 merge depo sites ;
29 run ;
```

```
NOTE: There were 17 observations read from the dataset WORK.DEPO.
NOTE: There were 18 observations read from the dataset WORK.SITES.
NOTE: The data set WORK.COMBINED has 18 observations and 27 variables.
NOTE: DATA statement used:
```

```
real time          0.08 seconds
cpu time           0.02 seconds
```

```
30
31 proc print ;
32 run ;
```

```
NOTE: There were 18 observations read from the dataset WORK.COMBINED.
NOTE: The PROCEDURE PRINT printed pages 1-2.
NOTE: PROCEDURE PRINT used:
```

```
real time          1.57 seconds
cpu time           0.07 seconds
```

Output from print

Obs	Site											
	ID	Ca	Mg	K	Na	NH4	NO3	Inorg				
							N	Cl	S04	Hlab	HField	
1	C000	0.34	0.029	0.099	0.119	0.51	1.79	0.80	0.17	1.18	0.01	0.02
2	C001	1.07	0.076	0.103	0.138	1.36	3.09	1.76	0.19	2.06	0.00	0.01
3	C002	9.72	0.886	1.020	1.879	6.55	33.60	12.68	2.44	19.32	0.22	0.34
4	C008	0.97	0.098	0.128	0.152	0.57	3.93	1.33	0.26	2.13	0.03	0.04
5	C010	0.47	0.048	0.091	0.140	0.48	3.94	1.26	0.29	1.85	0.05	0.05
6	C015	0.73	0.081	0.036	0.150	0.45	3.35	1.11	0.22	1.80	0.03	-9.00
7	C019	0.81	0.076	0.094	0.121	1.11	3.79	1.72	0.21	2.00	0.02	0.03
8	C021	0.69	0.075	0.106	0.120	0.92	4.78	1.80	0.23	2.69	0.05	0.06
9	C022	0.74	0.070	0.125	0.105	1.72	4.00	2.24	0.16	2.11	0.01	0.01
10	C091	2.22	0.168	0.180	0.385	1.02	10.63	3.20	0.58	5.77	0.13	0.16
11	C092	1.16	0.118	0.111	0.221	0.75	5.55	1.84	0.34	2.99	0.06	0.06
12	C093	1.09	0.124	0.115	0.214	0.95	7.66	2.47	0.35	5.29	0.11	0.13
13	C094	0.74	0.074	0.099	0.103	1.23	4.69	2.01	0.20	2.79	0.04	0.05
14	C096	1.22	0.095	0.126	0.229	0.56	6.27	1.85	0.36	3.25	0.09	0.08
15	C097	1.58	0.163	0.136	0.313	1.63	11.55	3.88	0.52	8.62	0.18	0.20

Obs	Svol	Ppt	Pct	Valid		Days	Date1	Date2	sdate
				F	L				
1	6849.6	13.79	51	22	11	367	12/28/1999	12/05/2000	28DEC99
2	13377.7	23.85	83	27	22	371	12/28/1999	01/02/2001	28DEC99
3	26264.1	268.38	62	38	32	371	12/28/1999	01/02/2001	28DEC99
4	28189.3	49.05	70	43	29	371	12/28/1999	01/02/2001	28DEC99
5	26739.0	53.82	11	36	6	370	12/29/1999	01/02/2001	29DEC99
6	14207.8	32.56	0	32	0	371	12/28/1999	01/02/2001	28DEC99
7	20265.0	34.66	61	36	23	371	12/28/1999	01/02/2001	28DEC99
8	21956.6	34.24	79	30	22	371	12/28/1999	01/02/2001	28DEC99
9	11922.4	21.92	88	25	23	371	12/28/1999	01/02/2001	28DEC99
10	40791.5	120.24	45	37	24	366	01/02/2000	01/02/2001	02JAN00

16	28482.8	97.44	53	30	23	371	12/28/1999	01/02/2001	28DEC99
17	18208.4	30.54	52	30	18	371	12/28/1999	01/02/2001	28DEC99
18

Obs	edate	daysop	sitename	strtdate	stopdate	elev
16	02JAN01	371	Rocky Mtn National	16AUG83	.	3159
17	02JAN01	371	Mesa Verde Nationa	28APR81	.	2172
18	.	.	Engineer Mountain	29JUL86	02JAN90	2758

11	35244.8	69.20	63	42	27	371	12/28/1999	01/02/2001	28DEC99
12	25042.0	82.50	64	34	26	371	12/28/1999	01/02/2001	28DEC99
13	25903.0	41.33	97	40	35	371	12/28/1999	01/02/2001	28DEC99
14	33234.0	78.98	64	33	26	371	12/28/1999	01/02/2001	28DEC99
15	34487.0	135.99	71	36	26	371	12/28/1999	01/02/2001	28DEC99

Obs	edate	daysop	sitename	strtdate	stopdate	elev
1	05DEC00	343	Alamosa	22APR80	.	2298
2	02JAN01	371	Las Animas Fish	04OCT83	.	1213
3	02JAN01	371	Niwot Saddle	05JUN84	.	3520
4	02JAN01	371	Four Mile Park	29DEC87	.	2502
5	02JAN01	370	Gothic	02FEB99	.	2926
6	02JAN01	371	Sand Spring	20MAR79	.	1998
7	02JAN01	371	Rocky Mtn National	29MAY80	.	2490
8	02JAN01	371	Manitou	17OCT78	.	2362
9	02JAN01	371	Pawnee	22MAY79	.	1641
10	02JAN01	366	Wolf Creek Pass	26MAY92	.	3292
11	02JAN01	371	Sunlight Peak	13JAN88	.	3206
12	02JAN01	371	Dry Lake	14OCT86	.	2527
13	02JAN01	371	Sugarloaf	04NOV86	.	2524
14	02JAN01	371	Molas Pass	29JUL86	.	3249
15	02JAN01	371	Buffalo Pass	07FEB84	.	3234

Obs	Site											
	ID	Ca	Mg	K	Na	NH4	NO3	Inorg				
							N	Cl	S04	Hlab	HField	
16	C098	1.66	0.166	0.185	0.263	1.92	8.16	3.33	0.61	5.05	0.07	0.09
17	C099	1.13	0.098	0.107	0.189	0.54	4.36	1.40	0.24	2.70	0.04	0.05
18	C095

Obs	Svol	Ppt	Pct	Valid		Days	Date1	Date2	sdate
				F	L				
16	6849.6	13.79	51	22	11	367	12/28/1999	12/05/2000	28DEC99
17	13377.7	23.85	83	27	22	371	12/28/1999	01/02/2001	28DEC99
18	26264.1	268.38	62	38	32	371	12/28/1999	01/02/2001	28DEC99

Producing the scatterplot

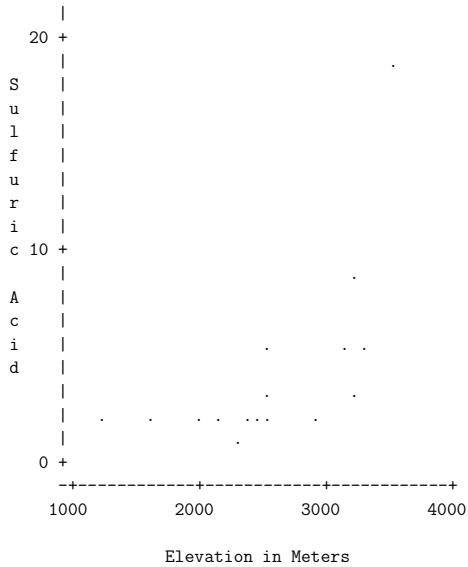
- SAS has two styles of plots
 - text plots (ugly as sin but easy to print)
 - SAS/Graph plots (pretty but a little harder to work with)
 - * run these in interactive mode
 - * when plot appears in window, use “File — Export as image” to save to file
 - * then print file or embed into Latex file

```
proc plot data = combined ; * text plot ;
plot S04 * elev = '.' / hpos = 40 vpos = 20 ;
* syntax: y-var * x-var = 'plotting symbol', hpos and vpos control size ;
title1 'Sulfuric Acid vs. Elevation' ;
title2 'Colorado NADP sites, Annual Data for 2000' ;
footnote 'Sulfuric Acid in kg/ha' ;
label S04 = 'Sulfuric Acid' elev = 'Elevation in Meters' ;
run ;
```

```
proc gplot data = combined ; * SAS/Graph plot ;
plot S04 * elev = '.' ; * syntax: y-var * x-var = 'plotting symbol' ;
title1 'Sulfuric Acid vs. Elevation' ;
title2 'Colorado NADP sites, Annual Data for 2000' ;
footnote 'Sulfuric Acid in kg/ha' ;
label S04 = 'Sulfuric Acid' elev = 'Elevation in Meters' ;
run ;
```

Sulfuric Acid vs. Elevation
Colorado NADP sites, Annual Data for 2000

Plot of S04*elev. Symbol used is '.'.



= 6 in so4elev.eps

Two other miscellaneous things you need to know for homework

- SAS expects data files to be delimited with spaces
- use “delimiter = ” option on infile statement to override

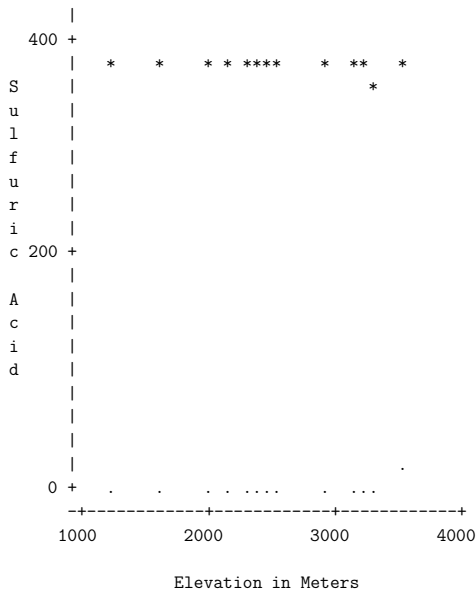
```
infile '/group/ftp/pub/kcowles/datasets/whatever' delimiter = ',' ;
infile '/group/ftp/pub/kcowles/datasets/whatever' delimiter = '\09'x
```

- overlaying two scatterplots on the same axis
- use “overlay” option on plot or gplot statement

```
proc plot data = combined ; * text plot ;
plot S04 * elev = '.' days * elev = '*' / hpos = 40 vpos = 20 overlay ;
* syntax: y-var * x-var = 'plotting symbol' ;
title1 'Sulfuric Acid vs. Elevation' ;
title2 'Colorado NADP sites, Annual Data for 2000' ;
footnote 'Sulfuric Acid in kg/ha' ;
label S04 = 'Sulfuric Acid' elev = 'Elevation in Meters' ;
run ;
```

Sulfuric Acid vs. Elevation
Colorado NADP sites, Annual Data for 2000

Plot of S04*elev. Symbol used is '.'.
Plot of Days*elev. Symbol used is '*'.



NOTE: 2 obs had missing values. 8 obs hidden.

- this works ok in *gplot* and when you run *plot* from within an interactive session
- characters in text plot may not line up correctly in overlay graph produced in *plot* in batch mode