Computing in Statistics

Data Preparation Using SAS

Lecture 18
Oct. 19, 2009

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

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.

● 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.
SAS code to read in this file

```sas
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 Inorg N Cl SO4 HLab HField Svol Ppt Pct ValidF ValidL Days Date1 $ Date2 $;
run;
proc print data = depo (obs=5);
run;
```

Log file

```
NOTE: SAS initialization used:
real time 0.55 seconds
cpu time 0.11 seconds
1 options linesize = 75 pagesize = 60;
data depo;
infile 'depoRepCO.asp' firstobs = 6;
input SiteID $ Per $ Year Crit1 Crit2 Crit3 Crit4 Ca Mg K Na NH4 NO3 Inorg N Cl SO4 HLab HField Svol Ppt Pct ValidF ValidL Days Date1 $ Date2 $;
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
```

Output of print

```
The SAS System
9:34 Thursday, September 13, 2001
Obs SiteID Per Year Crit1 Crit2 Crit3 Crit4 Ca Mg K Na NH4 NO3 N Cl SO4 HLab HField Svol Ppt Pct ValidF ValidL Days Date1 Date2
1 "CO00" "Annual" 2000 82 93 90 78 0.34 0.029
2 "CO01" "Annual" 2000 90 100 86 98 1.07 0.076
3 "CO02" "Annual" 2000 77 100 69 21 9.72 0.886
4 "CO08" "Annual" 2000 91 100 99 83 0.97 0.098
5 "CO10" "Annual" 2000 81 100 85 84 0.47 0.048

Inorg
Obs K Na NH4 NO3 N Cl SO4 HLab Svol
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 3377.7
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

Valid
Obs Ppt Pct F L Days Date1 Date2
1 13.79 51 22 11 367 "12/28/1" "12/28/2"
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

```sas
data depo;
  infile 'depoRepCO.asp' firstobs = 6;
  input SiteID $ Per $ Year Crit1 Crit2 Crit3 Crit4 Ca Mg K Na NH4 NO3 InorgN Cl SO4 HLab HField Svol Ppt Pct ValidF ValidL Days @201 Date1 $12. @217 Date2 $12.;
run;
```

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 “informats” 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

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 “mmdddy10.” 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.
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 Svol Ppt Pct ValidF ValidL Do;
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 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
```

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

<table>
<thead>
<tr>
<th>Site ID</th>
<th>Site Name</th>
<th>Start Date</th>
<th>End Date</th>
<th>Elevation (meters)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Active Sites

- CD00 Alamosa 4/22/1980 2298
- CD01 Las Animas Fish Hatchery 10/4/1983 1213
- CD02 Niwot Saddle 6/5/1984 3520
- CD08 Four Mile Park 12/29/1987 2502
- CD10 Gothic 2/2/1999 2926

Inactive Sites

- CD21 Manitou 10/17/1978 2362
- CD22 Pawnee 5/22/1979 1641
- CD91 Wolf Creek Pass 5/26/1992 3292
- CD92 Sunlight Peak 1/13/1988 3206
- CD93 Dry Lake 10/14/1986 2527
- CD94 Sugarloaf 11/4/1986 2524
- CD96 Molas Pass 7/29/1986 3249
- CD97 Buffalo Pass 2/7/1984 3234
- CD98 Rocky Mt National Park-Loch Vale 8/16/1983 3159
- CD99 Mesa Verde National Park 4/28/1981 2172

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 mmddyy10. @53 stopdate mmddyy10. @68 elev;
run;

proc print data = sites;
run;

From log file

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  CO95  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: LOST CARD.
SiteID=Comments sitename=ts and Suggestions strtdate=. stopdate=. elev=.
_ERROR_=1 _N_=19
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

<table>
<thead>
<tr>
<th>Obs</th>
<th>SiteID</th>
<th>sitename</th>
<th>strtdate</th>
<th>stopdate</th>
<th>elev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CO00</td>
<td>Alamosa</td>
<td>7417</td>
<td></td>
<td>2298</td>
</tr>
<tr>
<td>2</td>
<td>CO01</td>
<td>Las Animas Fish</td>
<td>8677</td>
<td></td>
<td>1213</td>
</tr>
<tr>
<td>3</td>
<td>Hatchery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>CO08</td>
<td>Four Mile Park</td>
<td>10224</td>
<td></td>
<td>2502</td>
</tr>
<tr>
<td>5</td>
<td>CO10</td>
<td>Gothic</td>
<td>14277</td>
<td></td>
<td>2926</td>
</tr>
<tr>
<td>6</td>
<td>CO15</td>
<td>Sand Spring</td>
<td>7018</td>
<td>1998</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>CO19</td>
<td>Rocky Mtn National</td>
<td>7454</td>
<td></td>
<td>2490</td>
</tr>
<tr>
<td>8</td>
<td>Park-Bea</td>
<td>Park-Beaver Meadow</td>
<td></td>
<td>2362</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>CO22</td>
<td>Paseene</td>
<td>7081</td>
<td></td>
<td>1641</td>
</tr>
<tr>
<td>10</td>
<td>CO91</td>
<td>Wolf Creek Pass</td>
<td>11834</td>
<td></td>
<td>3292</td>
</tr>
<tr>
<td>11</td>
<td>CO92</td>
<td>Sunlight Peak</td>
<td>10239</td>
<td></td>
<td>3206</td>
</tr>
<tr>
<td>12</td>
<td>CO93</td>
<td>Dry Lake</td>
<td>9783</td>
<td></td>
<td>2527</td>
</tr>
<tr>
<td>13</td>
<td>CO94</td>
<td>Sugarloaf</td>
<td>9804</td>
<td></td>
<td>2524</td>
</tr>
<tr>
<td>14</td>
<td>CO95</td>
<td>Molas Pass</td>
<td>9706</td>
<td></td>
<td>3249</td>
</tr>
<tr>
<td>15</td>
<td>CO97</td>
<td>Buffalo Pass</td>
<td>8803</td>
<td></td>
<td>3234</td>
</tr>
<tr>
<td>16</td>
<td>CO98</td>
<td>Rocky Mtn National</td>
<td>8628</td>
<td></td>
<td>3159</td>
</tr>
<tr>
<td>17</td>
<td>Park-Loc</td>
<td></td>
<td></td>
<td></td>
<td>2172</td>
</tr>
<tr>
<td>18</td>
<td>Park</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Guard</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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

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  18 18
SiteID=Guard sitename= strtdate=. stopdate=. elev=.
_ERROR_=1 _N_=19
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

<table>
<thead>
<tr>
<th>Obs</th>
<th>Site ID</th>
<th>sitename</th>
<th>strtdate</th>
<th>stopdate</th>
<th>elev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CO00</td>
<td>Alamosa</td>
<td>22APR80</td>
<td>.</td>
<td>2298</td>
</tr>
<tr>
<td>2</td>
<td>CO01</td>
<td>Las Animas Fish</td>
<td>04OCT83</td>
<td>.</td>
<td>1213</td>
</tr>
<tr>
<td>3</td>
<td>Hatcheny</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>CO08</td>
<td>Four Mile Park</td>
<td>29DEC87</td>
<td>.</td>
<td>2926</td>
</tr>
<tr>
<td>5</td>
<td>CO10</td>
<td>Gothic</td>
<td>12FEB99</td>
<td>29DEC87</td>
<td>2502</td>
</tr>
<tr>
<td>6</td>
<td>CO15</td>
<td>Sand Spring</td>
<td>20MAR97</td>
<td>1998</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>CO19</td>
<td>Rocky Mtn National</td>
<td>29MAY80</td>
<td>.</td>
<td>2490</td>
</tr>
<tr>
<td>8</td>
<td>Park-Bea</td>
<td>Park-Beaver Meadow</td>
<td>.</td>
<td>2362</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>CO22</td>
<td>Paseene</td>
<td>22MAY92</td>
<td>.</td>
<td>1641</td>
</tr>
<tr>
<td>10</td>
<td>CO91</td>
<td>Wolf Creek Pass</td>
<td>11834</td>
<td>.</td>
<td>3292</td>
</tr>
<tr>
<td>11</td>
<td>CO92</td>
<td>Sunlight Peak</td>
<td>10239</td>
<td>.</td>
<td>3206</td>
</tr>
<tr>
<td>12</td>
<td>CO93</td>
<td>Dry Lake</td>
<td>9783</td>
<td>.</td>
<td>2527</td>
</tr>
<tr>
<td>13</td>
<td>CO94</td>
<td>Sugarloaf</td>
<td>9804</td>
<td>.</td>
<td>2524</td>
</tr>
<tr>
<td>14</td>
<td>CO95</td>
<td>Molas Pass</td>
<td>9706</td>
<td>.</td>
<td>3249</td>
</tr>
<tr>
<td>15</td>
<td>CO97</td>
<td>Buffalo Pass</td>
<td>8803</td>
<td>.</td>
<td>3234</td>
</tr>
<tr>
<td>16</td>
<td>CO98</td>
<td>Rocky Mtn National</td>
<td>8628</td>
<td>.</td>
<td>3159</td>
</tr>
<tr>
<td>17</td>
<td>Park-Loc</td>
<td></td>
<td></td>
<td></td>
<td>2172</td>
</tr>
<tr>
<td>18</td>
<td>Park</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Guard</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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;

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 proc print;
31 run;

NOTE: There were 18 observations read from the dataset WORK.COMBINED.
NOTE: PROCEDURE PRINT printed pages 1-2.
NOTE: PROCEDURE PRINT used:
real time 1.57 seconds
cpu time 0.07 seconds

Output from print

<table>
<thead>
<tr>
<th>Site</th>
<th>Inorg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs ID Ca Mg K Na NH4 NO3 N Cl SO4 HLab HField</td>
<td></td>
</tr>
<tr>
<td>1 C000 0.34 0.029 0.099 0.119 0.51 1.79 0.80 0.17 1.18 0.01 0.03 0.02</td>
<td></td>
</tr>
<tr>
<td>2 C001 1.07 0.076 0.103 0.138 3.09 1.76 0.19 2.06 0.00 0.01 0.01</td>
<td></td>
</tr>
<tr>
<td>3 C002 9.72 0.886 1.020 1.879 33.60 12.68 2.44 19.32 0.22 0.34</td>
<td></td>
</tr>
<tr>
<td>4 C008 0.97 0.098 0.128 0.152 0.57 3.93 1.33 0.26 2.13 0.03 0.00 0.04</td>
<td></td>
</tr>
<tr>
<td>5 C010 0.47 0.048 0.091 0.140 3.94 1.26 0.29 1.85 0.05 0.00 0.05</td>
<td></td>
</tr>
<tr>
<td>6 C015 0.73 0.081 0.036 0.150 3.35 1.11 0.22 1.80 0.03 9.90</td>
<td></td>
</tr>
<tr>
<td>7 C019 0.81 0.076 0.121 0.111 3.79 1.72 0.21 2.06 0.00 0.00 0.03</td>
<td></td>
</tr>
<tr>
<td>8 C021 0.69 0.075 0.106 0.120 4.78 1.80 0.23 2.69 0.05 0.00 0.06</td>
<td></td>
</tr>
<tr>
<td>9 C022 0.74 0.070 0.125 0.105 4.00 2.24 0.16 2.11 0.01 0.00 0.01</td>
<td></td>
</tr>
<tr>
<td>10 C091 2.22 0.168 0.180 0.385 1.02 10.63 3.20 0.58 5.77 0.13 0.16</td>
<td></td>
</tr>
<tr>
<td>11 C092 1.16 0.118 0.111 0.221 0.75 5.55 1.84 0.34 2.99 0.06 0.06</td>
<td></td>
</tr>
<tr>
<td>12 C093 1.09 0.124 0.115 0.214 3.55 1.84 0.34 2.99 0.06 0.06</td>
<td></td>
</tr>
<tr>
<td>13 C094 0.74 0.074 0.099 0.103 4.69 2.01 0.20 2.79 0.04 0.05 0.05</td>
<td></td>
</tr>
<tr>
<td>14 C096 1.22 0.095 0.126 0.229 6.27 1.85 0.36 3.25 0.09 0.08</td>
<td></td>
</tr>
<tr>
<td>15 C097 1.58 0.163 0.136 0.313 1.63 11.55 3.88 0.52 8.62 0.18 0.20</td>
<td></td>
</tr>
</tbody>
</table>

Valid Valid

<table>
<thead>
<tr>
<th>Obs</th>
<th>Svol</th>
<th>Ppt</th>
<th>Pct</th>
<th>F</th>
<th>L</th>
<th>Days</th>
<th>Date1</th>
<th>Date2</th>
<th>sdate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6849.6</td>
<td>13.79</td>
<td>51</td>
<td>22</td>
<td>11</td>
<td>367</td>
<td>12/28/1999</td>
<td>01/02/2000</td>
<td>28DEC99</td>
</tr>
<tr>
<td>2</td>
<td>13377.7</td>
<td>23.85</td>
<td>83</td>
<td>27</td>
<td>22</td>
<td>371</td>
<td>12/28/1999</td>
<td>01/02/2000</td>
<td>02DEC99</td>
</tr>
<tr>
<td>3</td>
<td>26264.1</td>
<td>268.38</td>
<td>62</td>
<td>38</td>
<td>32</td>
<td>371</td>
<td>12/28/1999</td>
<td>01/02/2000</td>
<td>28DEC99</td>
</tr>
<tr>
<td>4</td>
<td>28189.3</td>
<td>49.05</td>
<td>70</td>
<td>43</td>
<td>29</td>
<td>371</td>
<td>12/28/1999</td>
<td>01/02/2000</td>
<td>28DEC99</td>
</tr>
<tr>
<td>5</td>
<td>26739.0</td>
<td>53.82</td>
<td>11</td>
<td>36</td>
<td>6</td>
<td>370</td>
<td>12/29/1999</td>
<td>01/02/2001</td>
<td>29DEC99</td>
</tr>
<tr>
<td>6</td>
<td>14207.8</td>
<td>32.56</td>
<td>0</td>
<td>32</td>
<td>6</td>
<td>371</td>
<td>12/28/1999</td>
<td>01/02/2001</td>
<td>28DEC99</td>
</tr>
<tr>
<td>7</td>
<td>20265.0</td>
<td>34.66</td>
<td>61</td>
<td>36</td>
<td>23</td>
<td>371</td>
<td>12/28/1999</td>
<td>01/02/2001</td>
<td>28DEC99</td>
</tr>
<tr>
<td>8</td>
<td>21956.6</td>
<td>34.24</td>
<td>80</td>
<td>30</td>
<td>22</td>
<td>371</td>
<td>12/28/1999</td>
<td>01/02/2001</td>
<td>28DEC99</td>
</tr>
</tbody>
</table>

Valid Valid

<table>
<thead>
<tr>
<th>Obs</th>
<th>Svol</th>
<th>Ppt</th>
<th>Pct</th>
<th>F</th>
<th>L</th>
<th>Days</th>
<th>Date1</th>
<th>Date2</th>
<th>sdate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35244.8</td>
<td>69.20</td>
<td>63</td>
<td>27</td>
<td>371</td>
<td>12/28/1999</td>
<td>01/02/2001</td>
<td>28DEC99</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>25042.0</td>
<td>82.50</td>
<td>64</td>
<td>26</td>
<td>371</td>
<td>12/28/1999</td>
<td>01/02/2001</td>
<td>28DEC99</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>28503.0</td>
<td>41.33</td>
<td>40</td>
<td>35</td>
<td>371</td>
<td>12/28/1999</td>
<td>01/02/2001</td>
<td>28DEC99</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>33234.0</td>
<td>78.98</td>
<td>64</td>
<td>33</td>
<td>26</td>
<td>371</td>
<td>12/28/1999</td>
<td>01/02/2001</td>
<td>28DEC99</td>
</tr>
</tbody>
</table>

Valid Valid

<table>
<thead>
<tr>
<th>Obs</th>
<th>Svol</th>
<th>Ppt</th>
<th>Pct</th>
<th>F</th>
<th>L</th>
<th>Days</th>
<th>Date1</th>
<th>Date2</th>
<th>sdate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>05DEC00</td>
<td>343</td>
<td>Alamoa</td>
<td>22APR80</td>
<td>00DEC99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>02JAN01</td>
<td>371</td>
<td>Las Animas Fish</td>
<td>04OCT83</td>
<td>05DEC99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>02JAN01</td>
<td>317</td>
<td>Niwot Saddle</td>
<td>05JUN84</td>
<td>06DEC99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>02JAN01</td>
<td>371</td>
<td>Four Mile Park</td>
<td>29DEC87</td>
<td>30DEC99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>02JAN01</td>
<td>370</td>
<td>Gothic</td>
<td>02FEB99</td>
<td>30MAR99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>02JAN01</td>
<td>371</td>
<td>Sand Spring</td>
<td>20MAY99</td>
<td>31MAY99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>02JAN01</td>
<td>371</td>
<td>Rocky Mt National</td>
<td>29MAR80</td>
<td>30MAR80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>02JAN01</td>
<td>371</td>
<td>Manitou</td>
<td>17OCT77</td>
<td>31MAY79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>02JAN01</td>
<td>371</td>
<td>Pawnee</td>
<td>22MAY79</td>
<td>31MAY79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>02JAN01</td>
<td>366</td>
<td>Wolf Creek Pass</td>
<td>26MAY92</td>
<td>31MAY92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>02JAN01</td>
<td>371</td>
<td>Sunlight Peak</td>
<td>13JAN88</td>
<td>31MAY92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>02JAN01</td>
<td>371</td>
<td>Dry Lake</td>
<td>14OCT86</td>
<td>31MAY92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>02JAN01</td>
<td>371</td>
<td>Sugarloaf</td>
<td>04NOV86</td>
<td>31MAY92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>02JAN01</td>
<td>371</td>
<td>Molas Pass</td>
<td>29JUL86</td>
<td>31MAY92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>02JAN01</td>
<td>371</td>
<td>Buffalo Pass</td>
<td>07FEB84</td>
<td>31MAY92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

```sas
proc plot data = combined ; * text plot ;
plot SO4 * elev = '.' / hpos = 40 vpos = 20 ;
title1 'Sulfuric Acid vs. Elevation' ;
title2 'Colorado NADP sites, Annual Data for 2000' ;
footnote 'Sulfuric Acid in kg/ha' ;
label SO4 = 'Sulfuric Acid' elev = 'Elevation in Meters' ;
run ;
```

```sas
proc gplot data = combined ; * SAS/Graph plot ;
plot SO4 * 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 SO4 = 'Sulfuric Acid' elev = 'Elevation in Meters' ;
run ;
```

Reading delimited files

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

```sas
infile '/group/ftp/pub/kcowles/datasets/whatever' delimiter = ',' ;
```

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