Subject: Re: Comma seperators
Posted by Paul van Delst on Thu, 18 May 2000 07:00:00 GMT
View Forum Message <> Reply to Message

Simon de Vet wrote:
>
> I am reading in data that looks like the following:
>
> CHATHAM ISLAND - NEW ZEALAND (DOE),,,,,,,,,
> 43.92ï¿½S,176.50ï¿½W,,,,,,,,
> 16-Sep-1983,11-Oct-1996,,,,,,,,
> Mon,Stat,Cl,NO3,SO4,Na ,SeaSalt,nssSO4,MSA,Dust,NH4
> of,Param,Air,Air,Air,Air,Air,Air,Air,Air
> Yr,*,ï¿½g/m3,ï¿½g/m3,ï¿½g/m3,ï¿½g/m3,ï¿½g/m3,
ï¿½g/m3,ï¿½g/m3,ï¿½g/m3,ï¿½g/m3
> Jan,N,58,58,58,58,58,57,0,0,58
> Jan,Mean,7.330,0.120,1.572,4.233,13.766,0.508,#N/A,#N/A,0.10 3
> Jan,StdDev,2.788,0.055,0.412,1.479,4.811,0.249,#N/A,#N/A,0.0 51
>
> Which continues untill the end of the year, and then another observation
> station follows the fame general format.
>
> I want to be able to read in the data into an array. I can already take
> out the header, but I cannot read in the data.

What do you consider the header?

> By default, IDL is
> treating each line as one entry, not recognizing the commas as entry
> seperators. I've read the help extensively, but as a non-fortran user,
> the input format documentation makes my brane hurt.

Let's say you have:

Jan,N,58,58,58,58,58,57,0,0,58
Jan,Mean,7.330,0.120,1.572,4.233,13.766,0.508,#N/A,#N/A,0.10 3
Jan,StdDev,2.788,0.055,0.412,1.479,4.811,0.249,#N/A,#N/A,0.0 51
Feb,N,58,58,58,58,58,57,0,0,58
Feb,Mean,7.330,0.120,1.572,4.233,13.766,0.508,#N/A,#N/A,0.10 3
Feb,StdDev,2.788,0.055,0.412,1.479,4.811,0.249,#N/A,#N/A,0.0 51
..etc..

How about:

```
char_buffer = ' '

REPEAT BEGIN
  READF, lun, char_buffer
```

```
  input_data = STR_SEP( char_buffer, ',' )

  ....here split up the data how you want by, say, testing
     input_data[0] == month (Jan, Feb, Mar, ....
     input_data[1] == data type (N, Mean, StdDev)
  ....and checking for invalid data, e.g. the #N/A thingoes

ENDREP UNTIL EOF( lun )
```

paulv
--
Paul van Delst Ph: (301) 763-8000 x7274
CIMSS @ NOAA/NCEP Fax: (301) 763-8545
Rm.202, 5200 Auth Rd. Email: pvandelst@ncep.noaa.gov
Camp Springs MD 20746

Subject: Re: Comma seperators
Posted by Martin Schultz on Mon, 22 May 2000 07:00:00 GMT
View Forum Message <> Reply to Message

Who ever put out the word that MS Excel could be used by sane scientists should
be hanged, quartered, stoned, etc. (or, to be a little more friendly: at least put
into a different state of mind ;-)

As to the solution of your problem, Paul is giving the right clue: read as a string
first, then do the processing. Depending on the file size and the number of times you will need to access this sort of output, it will eventually pay off to add a few error handling capacities such as CATCH... and test for empty strings in your STR_SEP result.
Now, do you realy want to attempt analyzing the date output as well ??

If I receive this kind of data, most often I prefer to start up this old moloch and clunky memory hog (I mean Excel) and attempt to put the stuff in a more ASCII friendly
order and format before writing an IDL reader. Largest trouble I have with this piece of creamware is that seldomly two spreadsheets look alike because columns or rows are shifted etc. Oh well, this world ain't perfect (but on average certainly better than MS software)...

Enough rummaging,
Martin

Simon de Vet wrote:
>
> I am reading in data that looks like the following:
>
> CHATHAM ISLAND - NEW ZEALAND (DOE),,,,,,,,,
> 43.92ï¿½S,176.50ï¿½W,,,,,,,,
> 16-Sep-1983,11-Oct-1996,,,,,,,,
> Mon,Stat,Cl,NO3,SO4,Na ,SeaSalt,nssSO4,MSA,Dust,NH4
> of,Param,Air,Air,Air,Air,Air,Air,Air,Air,Air
> Yr,*,ï¿½g/m3,ï¿½g/m3,ï¿½g/m3,ï¿½g/m3,ï¿½g/m3,
ï¿½g/m3,ï¿½g/m3,ï¿½g/m3,ï¿½g/m3
> Jan,N,58,58,58,58,58,57,0,0,58
> Jan,Mean,7.330,0.120,1.572,4.233,13.766,0.508,#N/A,#N/A,0.10 3
> Jan,StdDev,2.788,0.055,0.412,1.479,4.811,0.249,#N/A,#N/A,0.0 51
>
> Which continues untill the end of the year, and then another observation
> station follows the fame general format.
>
> I want to be able to read in the data into an array. I can already take
> out the header, but I cannot read in the data. By default, IDL is
> treating each line as one entry, not recognizing the commas as entry
> seperators. I've read the help extensively, but as a non-fortran user,
> the input format documentation makes my brane hurt.
>
> If I can read in the data, I think that I can manipulate it without too
> many problems...
>
> Thanks!
>
> Simon

--
[[[[[[[[[[
[[Dr. Martin Schultz Max-Planck-Institut fuer Meteorologie [[
[[Bundesstr. 55, 20146 Hamburg [[
[[phone: +49 40 41173-308 [[
[[fax: +49 40 41173-298 [[
[[martin.schultz@dkrz.de [[
[[[[[[[[[[

Subject: Re: Comma seperators
Posted by Ben Tupper on Mon, 22 May 2000 07:00:00 GMT
View Forum Message <> Reply to Message

Paul van Delst wrote:

> Simon de Vet wrote:
>>
>> I am reading in data that looks like the following:
>>
>> CHATHAM ISLAND - NEW ZEALAND (DOE),,,,,,,,,
>> 43.92ï¿½S,176.50ï¿½W,,,,,,,,
>> 16-Sep-1983,11-Oct-1996,,,,,,,,
>> Mon,Stat,Cl,NO3,SO4,Na ,SeaSalt,nssSO4,MSA,Dust,NH4
>> of,Param,Air,Air,Air,Air,Air,Air,Air,Air,Air
>> Yr,*,ï¿½g/m3,ï¿½g/m3,ï¿½g/m3,ï¿½g/m3,ï¿½g/m3,
ï¿½g/m3,ï¿½g/m3,ï¿½g/m3,ï¿½g/m3
>> Jan,N,58,58,58,58,58,57,0,0,58
>> Jan,Mean,7.330,0.120,1.572,4.233,13.766,0.508,#N/A,#N/A,0.10 3
>> Jan,StdDev,2.788,0.055,0.412,1.479,4.811,0.249,#N/A,#N/A,0.0 51
>>
>> Which continues untill the end of the year, and then another observation
>> station follows the fame general format.
>>
>> I want to be able to read in the data into an array. I can already take
>> out the header, but I cannot read in the data.
>
> What do you consider the header?
>
>> By default, IDL is
>> treating each line as one entry, not recognizing the commas as entry
>> seperators. I've read the help extensively, but as a non-fortran user,
>> the input format documentation makes my brane hurt.
>
> Let's say you have:
>
> Jan,N,58,58,58,58,58,57,0,0,58
> Jan,Mean,7.330,0.120,1.572,4.233,13.766,0.508,#N/A,#N/A,0.10 3
> Jan,StdDev,2.788,0.055,0.412,1.479,4.811,0.249,#N/A,#N/A,0.0 51
> Feb,N,58,58,58,58,58,57,0,0,58
> Feb,Mean,7.330,0.120,1.572,4.233,13.766,0.508,#N/A,#N/A,0.10 3
> Feb,StdDev,2.788,0.055,0.412,1.479,4.811,0.249,#N/A,#N/A,0.0 51
> ..etc..
>
> How about:
>
> char_buffer = ' '
>
> REPEAT BEGIN
> READF, lun, char_buffer
>
> input_data = STR_SEP(char_buffer, ',')
>

> ....here split up the data how you want by, say, testing
> input_data[0] == month (Jan, Feb, Mar, ...
> input_data[1] == data type (N, Mean, StdDev)
> ....and checking for invalid data, e.g. the #N/A thingoes
>
> ENDREP UNTIL EOF(lun)
>
>

Hello,

I'ld like to add that on occasion, I have found it useful to add the /TRIM keyword to the STR_SEP() function.
Once in a while the last element in input_data will become something unexpected, such as the expected value padded with blanks. I think the problem is in how the file was written, not in how it is read by IDL.

Ben

--
Ben Tupper

Bigelow Laboratory for Ocean Science
tupper@seadas.bigelow.org

pemaquidriver@tidewater.net