
Subject: Comma seperators

Posted by [Simon de Vet](#) on Thu, 18 May 2000 07:00:00 GMT

[View Forum Message](#) <> [Reply to Message](#)

I am reading in data that looks like the following:

```
CHATHAM ISLAND - NEW ZEALAND (DOE),,,,,,,,,,
43.92i½S,176.50i½W,,,,,,,,,
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,*,i½g/m3,i½g/m3,i½g/m3,i½g/m3,i½g/m3,
i½g/m3,i½g/m3,i½g/m3,i½g/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

Subject: Re: Comma seperators

Posted by [John-David T. Smith](#) on Mon, 22 May 2000 07:00:00 GMT

[View Forum Message](#) <> [Reply to Message](#)

Paul van Delst wrote:

>

> Ben Tupper wrote:

>>

>> Paul van Delst wrote:

>>

>>> Simon de Vet wrote:

>>>>

```

>>>> I am reading in data that looks like the following:
>>>>
>>>> CHATHAM ISLAND - NEW ZEALAND (DOE),,,,,,,,,,
>>>> 43.92°S,176.50°W,,,,,,,,,
>>>> 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,*,µg/m3,µg/m3,µg/m3,µg/m3,µg/m3,
µg/m3,µg/m3,µg/m3,µg/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'd 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.
>
> You know, the same thought occurred to me when I used this method to
> read *space*-separated data - I kept getting extra "fields" at the
> beginning of my string. I stuck the /TRIM keyword in the STRSEP call and
> nothing changed!!?? Weird.
>
> So instead of doing a
>
> result = STRSEP( string, ' ', /TRIM )
>
> I do a
>
> result = STRSEP( STRTRIM( string, 2 ), ' ' )
>
> Mind you this was one of those cases where something didn't work
> straight up and I spent precisely 0.1seconds figuring out why not before
> going on to something else.. :o)
>
> BTW, is there some sequence of layered string function calls one can use
> to trim and "collapse" a string with multiple delimiters between items
> to a single delimiter? e.g. to convert
>
> ,,this,,is,,,a,,,multiple,,,,delimited,,,,,,string,,,,
>
> to
>
> this,is,a,multiple,delimited,string
>
> I wrote a function to do it but it has a loop in it and a bunch of logic
> checking that looks horrendous. It does the job, but no reason why it
> can't look pretty....right?
>

```

res=strsplit(str,',',/EXTRACT)

will do it. The reason is null-length fields are **not** returned unless you use PRESERVE_NULL. You can also split on regular expressions. So, e.g. if you could be delimited by one or more spaces or commas, you could use:

```
res=strsplit(str,[' ,',/REGEX,/EXTRACT)
```

This is mostly v5.3 specific.

JD

--

J.D. Smith |*| WORK: (607) 255-5842
Cornell University Dept. of Astronomy |*| (607) 255-6263
304 Space Sciences Bldg. |*| FAX: (607) 255-5875
Ithaca, NY 14853 |*|

Subject: Re: Comma seperators
Posted by [davidf](#) on Mon, 22 May 2000 07:00:00 GMT
[View Forum Message](#) <> [Reply to Message](#)

Paul van Delst (pvandelst@ncep.noaa.gov) writes:

> Mind you on the scale of smart, I have been fortunate in that all
> my supervisors have been way off the scale (in the positive sense) - you
> know, the "think outside of the box" type of bods.

Paul, this cheap ploy to get more time to work on IDL newsgroup articles is not likely to work, in my experience. :-)

Cheers,

David

--

David Fanning, Ph.D.
Fanning Software Consulting
Phone: 970-221-0438 E-Mail: davidf@dfanning.com
Coyote's Guide to IDL Programming: <http://www.dfanning.com/>
Toll-Free IDL Book Orders: 1-888-461-0155

Subject: Re: Comma seperators
Posted by [davidf](#) on Mon, 22 May 2000 07:00:00 GMT
[View Forum Message](#) <> [Reply to Message](#)

Martin Schultz (martin.schultz@dkrz.de) writes:

- > 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 ;-)

Martin,

Take the rest of the afternoon off and go have a couple of brewskies. You deserve it for the wonderful advice you have rendered today. :-)

Cheers,

David

P.S. Let's just say I've heard those Germans have a hard time lightening up. :-)

--

David Fanning, Ph.D.

Fanning Software Consulting

Phone: 970-221-0438 E-Mail: davidf@dfanning.com

Coyote's Guide to IDL Programming: <http://www.dfanning.com/>

Toll-Free IDL Book Orders: 1-888-461-0155

Subject: Re: Comma seperators

Posted by [Paul van Delst](#) on Mon, 22 May 2000 07:00:00 GMT

[View Forum Message](#) <> [Reply to Message](#)

Martin Schultz wrote:

- >
- > 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 ;-)

I'm not disagreeing with you, but I have seen one of my bosses do things with Excel in a night that took me weeks to replicate in IDL. My result was a lot more flexible, but we still ended up with the same numbers and in the intervening 2 weeks he had furthered the science much, *much* more. Mind you on the scale of smart, I have been fortunate in that all my supervisors have been way off the scale (in the positive sense) - you know, the "think outside of the box" type of bods.

> order and format before writing an IDL reader. Largest trouble I have
> with this piece of creamware is that seldomly two spreadsheets look

~~~~~

Ha ha. Nice description! If I was drinking milk it would be coming out  
my nose. hee hee.

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 [Paul van Delst](#) on Mon, 22 May 2000 07:00:00 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Ben Tupper wrote:

>

> Paul van Delst wrote:

>

>> Simon de Vet wrote:

>>>

>>> I am reading in data that looks like the following:

>>>

>>> CHATHAM ISLAND - NEW ZEALAND (DOE),,,,,,,,,,

>>> 43.92°S,176.50°W,,,,,,,,,

>>> 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,\*,°g/m3,°g/m3,°g/m3,°g/m3,°g/m3,

°g/m3,°g/m3,°g/m3,°g/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'd 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.

```

You know, the same thought occurred to me when I used this method to read \*space\*-separated data - I kept getting extra "fields" at the beginning of my string. I stuck the /TRIM keyword in the STRSEP call and nothing changed!!?? Weird.

So instead of doing a

```
result = STRSEP( string, ' ', /TRIM )
```

I do a

```
result = STRSEP( STRTRIM( string, 2 ), ' ' )
```

Mind you this was one of those cases where something didn't work straight up and I spent precisely 0.1seconds figuring out why not before going on to something else.. :o)

BTW, is there some sequence of layered string function calls one can use to trim and "collapse" a string with multiple delimiters between items to a single delimiter? e.g. to convert

```
,,,this,,,is,,,a,,,multiple,,,,delimited,,,,,,string,,,,
```

to

```
this,is,a,multiple,delimited,string
```

I wrote a function to do it but it has a loop in it and a bunch of logic checking that looks horrendous. It does the job, but no reason why it can't look pretty....right?

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 [promashkin](#) on Tue, 23 May 2000 07:00:00 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

While sharing some points on MS products, I would argue that Excel is very good for some things, separating string record being among them. I don't know it well enough to beat IDL development with Excel, but I also use it to save time when working with unfriendly ASCII data. After that, other programs can be used. Excel, IMHO, has pathetic graphics. But again, you tend to use a screwdriver for driving screws, and the fact that it is no good for driving nails does not make it useless. Morale of the story - each application has its upsides, even Excel :-)

Cheers,  
Pavel

Martin Schultz wrote:

>

> 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 ;-)  
> -----  
> 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)....

---

---

Subject: Re: Comma seperators

Posted by [Martin Schultz](#) on Wed, 24 May 2000 07:00:00 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Ok, ok... I back off. I even admit that I have used this "thing" once a while for screwing (pun intended ;-). It's just that I have seen too many people (especially in the experimental world) relying heavily on Excel, and I have also seen many bad data sets which contained errors that were introduced by the spreadsheet program's "intelligence". That whole spreadsheet approach supports unorganized thinking in my view, because you can just add a parameter here or there and apply it to only a few cells or a few more, and once you exceed 1000 rows or so, it becomes almost impossible to track down such things. Then you create a second data set and you say: "huraah, I've got a template from the first data set", so you simply copy the new data into the old spread sheet. And suddenly you apply wrong calibration factors etc. And if only from a purely educational perspective: one should not allow students to use this sort of program for scientific data analysis! It's about as bad as writing 3d model code with hardcoded dimensions...

In summary:

Surgeon general's warning: Use of this software may endanger the health of your data especially under stress conditions as during field experiments. One piece of software contains 10 mg good stuff and 250 mg bad ballast.

Cheers,  
Martin

Pavel Romashkin wrote:

>

> While sharing some points on MS products, I would argue that Excel is  
> very good for some things, separating string record being among them. I  
> don't know it well enough to beat IDL development with Excel, but I also  
> use it to save time when working with unfriendly ASCII data. After that,  
> other programs can be used. Excel, IMHO, has pathetic graphics. But  
> again, you tend to use a screwdriver for driving screws, and the fact  
> that it is no good for driving nails does not make it useless. Morale of  
> the story - each application has its upsides, even Excel :-)  
> Cheers,  
> Pavel  
>  
> Martin Schultz wrote:  
>>  
>> 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 ;-)  
>> -----  
>> 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)....

--

```

[[ 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          [[
[[

```

---