Subject: missing data
Posted by caron on Wed, 15 Jun 1994 17:29:37 GMT
View Forum Message <> Reply to Message

1) efficiency of where():

Can anyone comment on the efficiency of using:

  max = max(my_arr(where(my_arr ne missing_data)))

to find the maximum value of an array, but excluding
missing values?

the array here has size 63000 elements, and only a few (0 - 30)
are "missing". I am imagining that the where() command returns
an array of indices, which then is used to grab just those indices
out of my_arr.

In a higher level language I could make a single pass through my_arr
to find the max value, excluding the missing data as I go.

Is there a more efficient way to do this in IDL, or is this just
the price we pay for its generality?

2) general ramblings on missing data and contour():

Can anyone comment on "good" ways to handle missing data? I map missing
data to a very large number, so that I can exclude it from plots using
the max_value keyword. Currently I believe that countour() does not actually
work correctly with max_value. Contour lines seem to stop at regions of
missing data, and contour fills get really messed up. I suppose you could
claim this is reasonable, but RSI hasn't gotten back to me as to whether they
consider this behavior a feature or a bug (hello?).

What should contour do in the presence of missing data? Well, it could
ignore it, and draw contours based on the good data. Since the contour
algorithm probably assumes a regular grid, you could "fill in" the
missing data using simple interpolation. This is equivilent to assuming
that the data represents a sampling of an underlying continuous physical
value, which is pretty much what contouring assumes.

On the other hand, you could argue that you wouldnt want to show the user
data thats " not really there", so contours must not go through regions
of missing data. If thats the assumption contour() makes, I can replace
my missing data myself with interpolated values if I want continuous contours.
Can RSI verify thats the assumption they make with no plans to fix this
behavior or add an option to change it? Does RSI have any highly technical
notes describing various assumptions like this that we could have access to?

How about starting one?

Another variant would be to fill the areas of missing data with some color,
or outline it. You cant do that now with contour because youll get every
contour line between the missing data value and the adjacent good data value.

3) On bloodsucking support fees:

    $200/year is a pretty low price for service. Any vertical niche company
like RSI needs to make customer service self supporting. I would like for RSI
to make avaliable things like known bugs, and other tech documents, that cost
them little and can save users lots. Also an e-mail address to RSI to report
possible bugs. Also a language summary on a single page. Also a "see also"
section for each routine in the reference manual. Also hypertext help. Ok Ill
stop now.....

Subject: Re: Missing Data
Posted by Andy Loughe on Thu, 18 Jul 1996 07:00:00 GMT
View Forum Message <> Reply to Message

Erica Mushovic wrote:
>
> Hi,
> I have to plot data gaps. I know that there is a plotm routine in the
> user library that will plot missing data as long as there is a default
> value, however, my data has not default values. I know where the
> missing data occurs. For example, if I have an array of x(10) I know
> that I do not want to connect the points between x(3:4) and x(7:8). Is
> there any way to plot that with the existing functions or with simple
> modifications to a user routine? I can think of a couple of ways to do
> that but those ways won't be practical with large data sets.
>
> Thank you.
> --
> Erica Mushovic

Does this example work for you?

```
x = indgen(10)           ; Create a data array.
x(2) = -99.  &  x(7) = -99.  ; Assign missing data values.
x (where (x lt -50)) = 999.  ; Change missing data points to large values.
plot, x, max=500.           ; Use the max_value keyword.
```


--

Andrew F. Loughe   afl@cdc.noaa.gov
University of Colorado, CIRES   http://cdc.noaa.gov/~afl
Campus Box 449   phn:(303)492-0707   fax:(303)497-7013
Boulder, CO 80309-0449   "He who laughs last thinks slowest!"

Subject: Re: Missing Data
Posted by Peter Mason on Thu, 18 Jul 1996 07:00:00 GMT
View Forum Message <> Reply to Message

```
>> I have to plot data gaps.  I know that there is a plotm routine in the
>> user library that will plot missing data as long as there is a default
>> value, however, my data has not default values.  I know where the
>> missing data occurs.  For example, if I have an array of x(10) I know
>> that I do not want to connect the points between x(3:4) and x(7:8).  Is
>> there any way to plot that with the existing functions or with simple
>> modifications to a user routine?  I can think of a couple of ways to do
>> that but those ways won't be practical with large data sets.
>
>
> Does this example work for you?
>
> x = indgen(10)            ; Create a data array.
> x(2) = -99.  &  x(7) = -99.  ; Assign missing data values.
> x (where (x lt -50)) = 999.  ; Change missing data points to large values.
> plot, x, max=500.          ; Use the max_value keyword.
```

Alternatively, you could use the special NaN ("not-a-number") functionality
which was introduced (for most platforms?) in IDL version 4.   A number of
IDL routines, including PLOT and PLOTS, automatically ignore data with NaN
"values".   Also, NaN is not specific to IDL; it's "value" is defined in
the IEEE floating point standard, and so can be recognised by other
applications which are aware of IEEE "denormals" (NaN and infinity).

You can get at NaN via the !VALUES system variable:
 !VALUES.F_NAN for float
 !VALUES.D_NAN for double

If DAT is a FLOAT data array and INX is an index into DAT giving the bad
values, you can set the bad values to NaN with:
 DAT(INX) = !VALUES.F_NAN
Or, considering the latest posting on "Problems with the IDL TIME_TEST"
(=> system variables seem to be accessed slower than general ones), it
might be quicker to use:
 NAN = !VALUES.F_NAN & DAT(INX) = NAN
if you have a lot of bad data.

Peter Mason

Subject: Re: Missing Data
Posted by hahn on Fri, 19 Jul 1996 07:00:00 GMT
View Forum Message <> Reply to Message

Erica Mushovic <mushovic@v2ma26.gsfc.nasa.gov> wrote:

> Hi,
> I have to plot data gaps. I know that there is a plotm routine in the
> user library that will plot missing data as long as there is a default
> value, however, my data has not default values. I know where the
> missing data occurs. For example, if I have an array of x(10) I know
> that I do not want to connect the points between x(3:4) and x(7:8).

If you have large arrays with contiguous data you may not want
to concept of NaN or MAXVALUE because you have to *insert*
data into an existing large arry. But as you have the intervals of
good data there is a simple solution. Given your numbers
you have to

  Plot, x(0:3), y(0:3)
  Oplot, x(4:7), y(4:7)
  Oplat, x(8:9), y(8:9)

> Is
> there any way to plot that with the existing functions or with simple
> modifications to a user routine? I can think of a couple of ways to do
> that but those ways won't be practical with large data sets.

All you need is a loop which calls Plot for the first segment of
data and oplot for the remaining segments of data.

> Thank you.
> --
> Erica Mushovic

Hope this helps
Norbert