Subject: Re: HISTOGRAM and string data
Posted by mchinand on Thu, 03 May 2007 06:11:25 GMT
View Forum Message <> Reply to Message

In article <1178153706.474934.309640@y80g2000hsf.googlegroups.com>,
Ed Hyer <ejhyer@gmail.com> wrote:
> Hello IDL Wizards,
>
> I did a search on the group for this, and found a post whose _subject_
> was my problem exactly, but the poster actually wanted something
> completely different (and was instantly satisfied).
>
> I have an application where I have an array of STRING_DATA, and I need
> to calculate stats of FLOAT_DATA based on the value of STRING_DATA. If
> HISTOGRAM worked on strings, this would be as simple as:
>
> hstr=histogram(STRING_DATA,reverse_indices=ristr)
> answer=hstr * 0.0
> for i=0l,n_elements(hstr) do if(hstr[i] gt 0) then answer[i] =
> f(FLOAT_DATA[ristr[ristr[i]:(ristr[i+1]-1)]])
>
> I thought UNIQ might help me, but it depends on doing a SORT, and
> sorting DATA is something I'd like to avoid if possible.
>
> One approach is to convert the STRING_DATA into some form of number,
> like longword integers. Any suggestions on how to do that without
> creating a very sparse field (if the resulting histogram has 1e8
> elements, that isn't necessarily going to work)?
>
> Oh, and feel free to bring on the slow solutions, this is not a time-
> dependent problem ;)
>
> --Edward H.
>

This doesn't generate the reverse indices but it's a start. It finds the
unique strings in the array and the number of occurences of each string.

Hope this helps,

--Mike


```
;===============================
pro str_hist, array

hist=intarr(1)
```

```
hist[0]=1
uniqstrings=strarr(1)
uniqstrings[0]=array[0]
for i=1, n_elements(array)-1 do begin
   idx=where(uniqstrings eq array[i])
   if (idx eq -1) then begin ; found new string
      uniqstrings=[uniqstrings,array[i]]
      hist=[hist,1]
   endif else begin
      hist[idx]++
   endelse
endfor


print, hist
print, uniqstrings
end
```

--
Michael Chinander
m-chinander@uchicago.edu
Department of Radiology
University of Chicago

Subject: Re: HISTOGRAM and string data
Posted by MarioIncandenza on Thu, 03 May 2007 16:08:56 GMT
View Forum Message <> Reply to Message

Caught up in doing things "the IDL way", my brain just quit on me.

"The Klunky Way":

```
IDL> unique_i = uniq(STRING_DATA,sort(STRING_DATA))
IDL> unique_strings = STRING_DATA[unique_i]
IDL> STRING_CODE = intarr(n_elements(STRING_DATA))
IDL> for i=0l,n_elements(unique_strings)-1 do $
IDL>    STRING_CODE[where(STRING_DATA eq unique_strings[i])] = i
```

All the fancy HISTOGRAM magic then works like a charm on STRING_CODE.

Subject: Re: HISTOGRAM and string data
Posted by Jean H. on Thu, 03 May 2007 16:19:54 GMT
View Forum Message <> Reply to Message

Do you know the max number of character per string?

you could try something like this, which works well if you have only 1
character and would need a bit of work if you have more:

```
IDL> a= ['a','b','b','c','a']
IDL> b = byte(a)
IDL> print,b
   97
   98
   98
   99
   97

IDL> print, histogram(b, min = 97, max = 99)
          2       2        1
```

Jean