
Subject: Re: setting histogram bin sizes?

Posted by [Juggernaut](#) on Tue, 22 Jul 2008 17:18:11 GMT

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

On Jul 22, 12:59 pm, Bennett <juggernaut...@gmail.com> wrote:

> On Jul 22, 11:52 am, "Jeff N." <jeffnettles4...@gmail.com> wrote:

>

>

>

>> Hi folks,

>

>> I'm looking for suggestions for a way to set bin sizes for a histogram
>> when I don't know much about the data before calculating the
>> histogram. Here's my situation: I'm putting together some code that
>> takes a hyperspectral image cube and extracts a series of one-band
>> parameters from the cube (band depth at a certain wavelength, etc.).
>> In trying to assess which of these parameters is most useful for our
>> particular application i thought about calculating a histogram for
>> each parameter. The problem is that these parameter images (one band,
>> floating point images per parameter) will not necessarily fall into
>> the same range. Many have possible values of 0 - 1, but they won't
>> necessarily take up that entire range. Some however, will not have
>> possible values of 0 - 1, but could instead have numbers in the 10s or
>> even hundreds. Some parameters have values that are actually in log
>> space.

>

>> I know that I could simply set the NBINS keyword to HISTOGRAM(), but
>> then the question would become how many bins to use? I did some quick
>> searching, and there are a few attempts at calculating bin sizes or
>> the number of bins on Wikipedia (<http://en.wikipedia.org/wiki/Histogram>). Short of any other information, i am going to use an
>> equation from that page that is at least based on the standard
>> deviation of the data. But, since I don't have a lot to go on, I
>> would very much like to have input from anyone on this newsgroup who
>> might have any suggestions for me.

>

>> Thanks,

>> Jeff

>

> I'd have to say if I was going to approach this and this is definitely
> not a very fun problem is to do the following.
> nels = n_elements(data)
> range = max(data) - min(data)
> IF range/median(data) GT 10 THEN nbins = 10 ELSE nbins = round(range/
> median(data))
> nbins = nbins < nels/10 ;- Make sure you don't have way too many bins
> This is just off the top of my head and working with some random
> data...

- > I'm sure there are special cases that require some serious thought.
- >
- > Hope it helps a bit....somehow....someday

The GT should be a LT by the way...sorry for the confusion
