
Subject: Re: histogram and binsize problems

Posted by [David Fanning](#) on Thu, 27 Feb 2003 17:11:59 GMT

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

Chad Bender (cbender@mail.astro.sunysb.edu) writes:

> I'm encountering some confusion using the histogram function when
> specifying the min, max, and binsize keywords.

Oh, right. I'm glad I'm not the only one this drives
crazy.

I spend several chapters in a book talking about this oddity
in Histogram, but I have never gotten around to writing a
wrapper for the command. (Which seems really odd to me.)

In any case, here is the relevant code from the program in
the book:

```
; Calculate the histogram.
```

```
histdata = Histogram(image, Binsize=binsize, Max=Max(image), $  
  Min=Min(image))
```

```
; Have to fudge the bins and histdata variables to get the  
; histogram plot to make sense.
```

```
npts = N_Elements(histdata)  
halfbinsize = binsize / 2.0  
bins = Findgen(N_Elements(histdata)) * binsize + Min(image)  
binsToPlot = [bins[0], bins + halfbinsize, bins[npts-1] + binsize]  
histdataToPlot = [histdata[0], histdata, histdata[npts-1]]  
xrange = [Min(binsToPlot), Max(binsToPlot)]
```

```
; Plot the histogram of the display image. Axes first.
```

```
Plot, binsToPlot, histdataToPlot, $ ; Fudged histogram and bin data.
```

```
  Background=backcolor, $ ; Background color of the display.
```

```
  Charsize=thisCharsize, $ ; Character size.
```

```
  Color=axiscolor, $ ; The color of the axes.
```

```
  Max_Value=max_value, $ ; The maximum value of the plot.
```

```
  NoData=1, $ ; Draw the axes only. No data.
```

```
  Position=histoPos, $ ; Position of the plot.
```

```
  Title='Image Histogram', $ ; The title of the plot.
```

```
  XRange=xrange, $ ; The X data range.
```

```
  XStyle=1, $ ; Exact axis scaling.
```

```
  XTickformat='(I6)', $ ; Format of the X axis annotations.
```

```
  XTitle='Image Value', $ ; The title of the X axis.
```

```
YMinor=1, $           ; One minor tick mark on X axis.  
YRange=[0,max_value], $       ; The Y data range.  
YStyle=1, $           ; Exact axis scaling.  
YTickformat='(I6)', $        ; Format of the Y axis annotations.  
YTitle='Pixel Density', $     ; The title of the Y axis.  
_Extra=extra           ; Pass any extra PLOT keywords.
```

```
; Overplot the histogram data in the data color.
```

```
OPlot, binsToPlot, histdataToPlot, PSym=10, Color=dataColor
```

```
; Make histogram boxes by drawing lines in data color.
```

```
FOR j=1L,N_Elements(bins)-2 DO BEGIN  
  PlotS, [bins[j], bins[j]], [!Y.CRange[0], histdata[j] < max_value], $  
    Color=dataColor  
ENDFOR
```

Cheers,

David

--

David W. Fanning, Ph.D.

Fanning Software Consulting, Inc.

Phone: 970-221-0438, E-mail: david@dfanning.com

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

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

Subject: Re: histogram and binsize problems

Posted by [Liam E. Gumley](#) on Thu, 27 Feb 2003 18:36:30 GMT

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

"Chad Bender" <cbender@mail.astro.sunysb.edu> wrote in message
news:Pine.LNX.4.33.0302271124480.4043-100000@hapuna.ess.suny sb.edu...

> I'm encountering some confusion using the histogram function when
> specifying the min, max, and binsize keywords. I looked at JD Smith's
> tutorial. But I couldn't find the answer to my question there, as all I
> really want is a plain old histogram, not some fancy array manipulation.
[stuff deleted]

You may wish to check out the HIST_PLOT procedure I wrote as an example
program for my book (Chapter 6, pp 260-262):

http://www.gumley.com/PIP/About_Book.html

You should be able to do something like this:

hist_plot, data, min=0.0, max=0.15, binsize=0.005

Cheers,

Liam.

Practical IDL Programming

<http://www.gumley.com/>

Subject: Re: histogram and binsize problems

Posted by [Chad Bender](#) on Thu, 27 Feb 2003 19:26:35 GMT

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

> You may wish to check out the HIST_PLOT procedure I wrote as an example
> program for my book (Chapter 6, pp 260-262):

Thanks for the suggestions. Liam, your routine gives me the same result
as I was getting with just:

```
plot, findgen(31), histogram(data,min=0.0,max=0.15,binsize=0.005), psym=10
```

I looked through it briefly, and it is doing enough strange things to
convince me that how histogram sets up its bins is not trivial. However,
using it does not get rid of the bins with value 0.

I tried a few other things to see if I could figure out something of what
histogram is doing. It's obvious from something like the following that
histogram has rounding issues at the bin boundaries.

```
IDL> test=findgen(31)*0.005
```

```
IDL> plot, test, histogram(test, min=0.0, max=0.15, binsize=0.005),  
psym=10
```

```
IDL> plot, test, histogram(test, binsize=0.005), psym=10
```

But from expanding the min and max by $\pm \text{binsize}/2$, I get almost the
expected result

```
IDL> hist=histogram(test, min=-0.0025, max=0.1525, binsize=0.005)
```

```
IDL> help, hist
```

```
HIST      LONG      = Array[32]
```

```
IDL> plot, findgen(32)*0.005, hist, psym=10, ystyle=3, xstyle=3
```

From this I can see that the last bin is empty. It
falls outside the range I specified by min and max (ie. it covers
0.1525-0.1575). So why does IDL create it at all?

Chad

--

#####

Chad Bender

Subject: Re: histogram and binsize problems
Posted by [K. Bowman](#) on Thu, 27 Feb 2003 19:31:18 GMT
[View Forum Message](#) <> [Reply to Message](#)

In article

<Pine.LNX.4.33.0302271407160.4043-100000@hapuna.ess.sunysb.edu>,
Chad Bender <cbender@mail.astro.sunysb.edu> wrote:

> I tried a few other things to see if I could figure out something of what
> histogram is doing. It's obvious from something like the following that
> histogram has rounding issues at the bin boundaries.

One option is to do your own scaling and conversion to integer, then use
histogram with a binsize of 1. Then you can control how the rounding
occurs.

Ken Bowman

Subject: Re: histogram and binsize problems
Posted by [David Fanning](#) on Thu, 27 Feb 2003 20:00:34 GMT
[View Forum Message](#) <> [Reply to Message](#)

Chad Bender (cbender@mail.astro.sunysb.edu) writes:

> I tried a few other things to see if I could figure out something of what
> histogram is doing. It's obvious from something like the following that
> histogram has rounding issues at the bin boundaries.
> IDL> test=findgen(31)*0.005
> IDL> plot, test, histogram(test, min=0.0, max=0.15, binsize=0.005),
> psym=10
> IDL> plot, test, histogram(test, binsize=0.005), psym=10

Humm. I'm not sure that's what I would conclude from
this test.

This program seems to work correctly:

PRO TEST

data = randomu(-3L, 500)

data = scale_vector(data, 0.00, 0.14999999)

test = data

```

binsize = 0.005
histdata = histogram(test, min=min(data), $
  max=max(data), binsize=binsize)
TVLCT, 0, 255, 0, !D.Table_Size-2
DEVICE, Decomposed=0
color = !D.Table_Size-2
npts = N_Elements(histdata)
halfbinsize = binsize / 2.0
bins = Findgen(N_Elements(histdata)) * binsize + Min(test)
binsToPlot = [bins[0], bins + halfbinsize, $
  bins[npts-1] + binsize]
histdataToPlot = [histdata[0], histdata, histdata[npts-1]]
xrange = [Min(binsToPlot), Max(binsToPlot)]
Plot, binsToPlot, histdataToPlot, PSYM=10, /NoData
OPlot, binsToPlot, histdataToPlot, Color=color, PSYM=10
END

```

You will need SCALE_VECTOR from my web page to run it:

http://www.dfanning.com/programs/scale_vector.pro

I think if a value is greater than or equal to the lower bin range, it goes in that bin. Seems reasonable to me.

Cheers,

David

--

David W. Fanning, Ph.D.

Fanning Software Consulting, Inc.

Phone: 970-221-0438, E-mail: david@dfanning.com

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

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

Subject: Re: histogram and binsize problems

Posted by [JD Smith](#) on Thu, 27 Feb 2003 21:12:36 GMT

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

On Thu, 27 Feb 2003 09:52:33 -0700, Chad Bender wrote:

> Hi-

>

> I'm encountering some confusion using the histogram function when
 > specifying the min, max, and binsize keywords. I looked at JD Smith's
 > tutorial. But I couldn't find the answer to my question there, as all I
 > really want is a plain old histogram, not some fancy array manipulation.

I think what you are after is in the tutorial, just up front in the part people skip over:

NBINS

A relative of BINSIZE, NBINS is something of a misnomer. The relation HISTOGRAM uses to compute the bin size if passed NBINS is:

$$\text{BINSIZE} = (\text{MAX} - \text{MIN}) / (\text{NBINS} - 1)$$

and if NBINS is specified, MAX is changed to be (independent of any value passed as MAX):

$$\text{MAX} = \text{NBINS} * \text{BINSIZE} + \text{MIN}$$

As such, it's probably better to avoid NBINS, if you care about the MAX value staying put. A better relation which would leave MAX as is and give you exactly NBINS bins between MIN and MAX:

$$\text{BINSIZE} = (\text{MAX} - \text{MIN}) / \text{NBINS}$$

Good luck,

JD
