Subject: Re: Match Histogram Binsize with Data Type
Posted by Fabzi on Tue, 05 Mar 2013 18:37:43 GMT
View Forum Message <> Reply to Message

Hi David,

Regardless of the probably wrong things that have been published and
will be published in the future, I don't see the point of using a .5
binsize with integer data, and I don't think the behaviour of histogram
is defined in this case.

Following code is doing ok for example:

```
d = Fix(Scale_Vector(RandomU(-3L, 1000), 0, 360))
h1 = Histogram(d, Min=0, Max=360, BINSIZE=22.)
h2 = Histogram(Float(d), Min=0.0, Max=360.0, BINSIZE=22)
cgPlot, h1
cgPlot, h2, Color='red', /overplot
```

Cheers,

Fabien

On 03/05/2013 05:57 PM, David Fanning wrote:
> Folks,
>
> I just spent an uncomfortable and depressing couple of hours either (1)
> thinking I was going crazy or (2) convinced the IDL Histogram command
> had a bug of such monumental proportions that any thinking person would
> ..., etc.
>
> Boiled down, it amounted to me using a floating point binsize with
> integer data. A BIG no-no when using the Histogram command. (I was
> actually using HIST_2D, which provides no such warning in its
> documentation.)
>
> I can't stress this enough. You get INCORRECT values if you mismatch the
> binsize and the data type. Let me say it again, you get INCORRECT
> answers!
>
> I'm just guessing, but it wouldn't surprise me to learn that the
> Histogram command produces incorrect values 50% of the time, simply
> because people don't realize the consequences of their thoughtless use
> of the command. (Guess arrived at by personal experience.)
>
> Wouldn't it be nice if there could be a warning about this somewhere?
> Like, say, in the Histogram command itself.
>

> Here is what I mean:
>
> d = Fix(Scale_Vector(RandomU(-3L, 1000), 0, 360))
> h1 = Histogram(d, Min=0, Max=360, BINSIZE=22.5)
> h2 = Histogram(Float(d), Min=0.0, Max=360.0, BINSIZE=22.5)
> cgPlot, h1
> cgPlot, h2, Color='red', /overplot
>
> Cheers,
>
> David
>

Subject: Re: Match Histogram Binsize with Data Type
Posted by David Fanning on Tue, 05 Mar 2013 18:58:34 GMT
View Forum Message <> Reply to Message

Fabien writes:

> Regardless of the probably wrong things that have been published and
> will be published in the future, I don't see the point of using a .5
> binsize with integer data, and I don't think the behaviour of histogram
> is defined in this case.

Well, the 22.5 represents the 16 directions around a compass where I wish to draw something (360/16=22.5). The number 22 just ain't gonna get it done in this case. :-)

Plus, I didn't *know* the data was integer type. It was a variable that came from somewhere else.

I'm saying, you have to do a hell of a lot of checking to get things right. I know, because I do the checking in cgHistoplot. In fact, I've written code to do the job (Convert_To_Type, for example).

It seems to me Histogram ought to do a little bit of checking on its own if it is going to be so damn obstinate about returning incorrect values, just because you didn't read the small print and know what the rules were.

Histogram is great. It just isn't user friendly. And, for such an important routine in IDL, it ought to be.

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.idlcoyote.com/
Sepore ma de ni thue. ("Perhaps thou speakest truth.")

Subject: Re: Match Histogram Binsize with Data Type
Posted by Jeremy Bailin on Wed, 06 Mar 2013 05:32:02 GMT
View Forum Message <> Reply to Message

On 3/5/13 12:58 PM, David Fanning wrote:
> Fabien writes:
>
>> Regardless of the probably wrong things that have been published and
>> will be published in the future, I don't see the point of using a .5
>> binsize with integer data, and I don't think the behaviour of histogram
>> is defined in this case.
>
> Well, the 22.5 represents the 16 directions around a compass where I
> wish to draw something (360/16=22.5). The number 22 just ain't gonna get
> it done in this case. :-)
>
> Plus, I didn't *know* the data was integer type. It was a variable that
> came from somewhere else.
>
> I'm saying, you have to do a hell of a lot of checking to get things
> right. I know, because I do the checking in cgHistoplot. In fact, I've
> written code to do the job (Convert_To_Type, for example).
>
> It seems to me Histogram ought to do a little bit of checking on its own
> if it is going to be so damn obstinate about returning incorrect values,
> just because you didn't read the small print and know what the rules
> were.
>
> Histogram is great. It just isn't user friendly. And, for such an
> important routine in IDL, it ought to be.
>
> Cheers,
>
> David

This might make me unpopular, but I think it's doing exactly what it's supposed to be doing in this case. The Sky Is Falling, etc.

-Jeremy.

Subject: Re: Match Histogram Binsize with Data Type
Posted by David Fanning on Wed, 06 Mar 2013 12:25:11 GMT
View Forum Message <> Reply to Message

Jeremy Bailin writes:

> This might make me unpopular, but I think it's doing exactly what it's
> supposed to be doing in this case. The Sky Is Falling, etc.

Yes, of course. I'm not saying it is wrong. I'm saying it is needlessly confusing and leads people to inadvertently get the wrong values without a word of warning that it is doing so. Histogram is in many ways the heart of soul of IDL. If it has special requirements to make it work correctly, then either it should (1) enforce those requirements itself by checking to see that the requirements are met, or (2) warn users that they are using the program inappropriately and will get incorrect results. People will change their habits soon enough. Although God knows there will be a few weeks where 90 percent of the IDL programming public has to fix their damn programs!

Silently going about your business and producing incorrect results is irresponsible at best and certainly not what you would expect from a software company looking out for user's interest. I realize we use the program "as is and without warranty," but this is like if the TV command just decided to arbitrarily truncate your data to bytes before displaying it. Oh, wait. Bad example. Uh, well, you probably understand what I mean. People don't expect to use a function that blithely returns incorrect results that are EXTREMELY hard to distinguish from the correct results. So hard, I'm sure 99.9 percent of us don't know we are being bamboozled.

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.idlcoyote.com/
Sepore ma de ni thue. ("Perhaps thou speakest truth.")

Subject: Re: Match Histogram Binsize with Data Type
Posted by David Fanning on Wed, 06 Mar 2013 13:08:08 GMT
View Forum Message <> Reply to Message

David Fanning writes:

> People don't expect to use a function that blithely returns
> incorrect results that are EXTREMELY hard to distinguish from the
> correct results. So hard, I'm sure 99.9 percent of us don't know we are
> being bamboozled.

In my case, I wouldn't have known except that I was writing a program that depended on these results and I kept noticing that my results didn't agree with the results of another well-known program that did the same thing. My program wasn't terribly different. Just different enough that it was obvious to me that it wasn't the same. Investigating this discrepancy is what put me onto this.

Had I not had the other program to check my work, I would have certainly felt comfortable publishing my results. The program "works", of course. And, if you just eyeball it, the results "make sense". But, the results are just ever so much wrong. That bothers me. I don't know if it bothers anyone else. :-)

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.idlcoyote.com/
Sepore ma de ni thue. ("Perhaps thou speakest truth.")

Subject: Re: Match Histogram Binsize with Data Type
Posted by Paul Van Delst[1] on Thu, 07 Mar 2013 17:51:20 GMT
View Forum Message <> Reply to Message

If it makes you feel any better, I lay awake at night worrying about plausibly incorrect results. In any language.

But, I don't really use IDL to produce quantitative results (computing means and stddevs don't count)

:o)

cheers,

paulv

p.s. In my youth I subscribed to the GIGO concept. Not anymore. :o\

On 03/06/13 08:08, David Fanning wrote:
> David Fanning writes:
>
>> People don't expect to use a function that blithely returns
>> incorrect results that are EXTREMELY hard to distinguish from the
>> correct results. So hard, I'm sure 99.9 percent of us don't know we are
>> being bamboozled.
>
> In my case, I wouldn't have known except that I was writing a program
> that depended on these results and I kept noticing that my results
> didn't agree with the results of another well-known program that did the
> same thing. My program wasn't terribly different. Just different enough
> that it was obvious to me that it wasn't the same. Investigating this
> discrepancy is what put me onto this.
>
> Had I not had the other program to check my work, I would have certainly
> felt comfortable publishing my results. The program "works", of course.
> And, if you just eyeball it, the results "make sense". But, the results
> are just ever so much wrong. That bothers me. I don't know if it bothers
> anyone else. :-)
>
> Cheers,
>
> David
>
>