Subject: Match Histogram Binsize with Data Type
Posted by David Fanning on Tue, 05 Mar 2013 16:57:02 GMT

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

--
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.")

Paul van Delst writes:

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

Well, it *does* make me feel better. :-)

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

I was on a forum web page not too long ago and the folks there were
going off on how horrible IDL was to do anything sensible and how it
confounded expectations, was impossible for new people to learn, etc.
Then some guy jumps in to say, 'Wait, there is a web site out there
where the guy tries to fix all the things that are wrong with IDL.'

It stopped me. Is that what I've been doing all these years? Fixing IDL?

I suppose it is. And then I got totally depressed. Why do I do this
work? And why do I do it for free, for God's sake!? I've no money in the
bank, a job that pays very little, and I choose to spend my time fixing
IDL. It's a joke.

That's when I decided to go trekking in Patagonia. :-)

Before I leave, though, I wrote my own histogram routine, cgHistogram,
which I am just about to add to the Coyote Library. It fixes the bug
with byte arrays (not fixed officially until IDL 8.2), checks to be sure
the binsize data type matches the data type of the data you are binning,
so you always get the correct results, allows for missing values in the
histogram, will smooth the data if needed, and will return the relative
frequency, instead of the histogram count, if you want that. It is a
HELL of a lot better than the Histogram command, as always. Maybe one or
two people will use it. ;-)

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.")

On 03/07/13 13:12, David Fanning wrote:
> Paul van Delst writes:
>
>> If it makes you feel any better, I lay awake at night worrying about
>> plausibly incorrect results. In any language.
>
> Well, it *does* make me feel better. :-)
>
>> But, I don't really use IDL to produce quantitative results (computing
>> means and stddevs don't count)
>
> I was on a forum web page not too long ago and the folks there were
> going off on how horrible IDL was to do anything sensible and how it
> confounded expectations, was impossible for new people to learn, etc.
> Then some guy jumps in to say, 'Wait, there is a web site out there
> where the guy tries to fix all the things that are wrong with IDL.'
>
> It stopped me. Is that what I've been doing all these years? Fixing IDL?
>
> I suppose it is. And then I got totally depressed. Why do I do this
> work? And why do I do it for free, for God's sake!? I've no money in the
> bank, a job that pays very little, and I choose to spend my time fixing
> IDL. It's a joke.

Maybe. But I (personally) don't think so.

You've chosen not to "pivot". :o)

 http://pragprog.com/magazines/2013-03/what-if-you-dont-want- to-pivot

And, based purely on this forum (and the one day when we met and shared
a beer), you certainly have lived a much more interesting life than most
people (especially if we restrict the criteria to the more, uh, senior
years of one's life! ha! :o).

I also wouldn't be too quick to discount the countless hours of
hair-pulling frustration you have saved many people via your
webpage/posts/library/etc. Karmic currency, you know? Doesn't help you
fiscally, but still...

> That's when I decided to go trekking in Patagonia. :-)

Perhaps that's the normal response for people caught in existential
cul-de-sacs? In the same vein, I go cycling in mountains somewhere (I
recommend Andalusia by the way...).

> Before I leave, though, I wrote my own histogram routine, cgHistogram,
> which I am just about to add to the Coyote Library. It fixes the bug
> with byte arrays (not fixed officially until IDL 8.2), checks to be sure
> the binsize data type matches the data type of the data you are binning,
> so you always get the correct results, allows for missing values in the
> histogram, will smooth the data if needed, and will return the relative
> frequency, instead of the histogram count, if you want that. It is a
> HELL of a lot better than the Histogram command, as always. Maybe one or
> two people will use it. ;-)

It won't put $$ in your account but: good one.

Enjoy your trek.

cheers,

paulv

Subject: Re: Match Histogram Binsize with Data Type
Posted by Craig Markwardt on Fri, 08 Mar 2013 00:37:32 GMT
View Forum Message <> Reply to Message

On Thursday, March 7, 2013 1:12:35 PM UTC-5, David Fanning wrote:
> That's when I decided to go trekking in Patagonia. :-)

I'm confused: how will we send you our IDL questions while you are trekking? :-)

Subject: Re: Match Histogram Binsize with Data Type
Posted by David Fanning on Fri, 08 Mar 2013 00:42:53 GMT
View Forum Message <> Reply to Message

Paul van Delst writes:

> You've chosen not to "pivot". :o)
>
>   http://pragprog.com/magazines/2013-03/what-if-you-dont-want- to-pivot

When the student is ready, the teacher appears.

If I read the War of Art, recommended in that article, while I am in
Patagonia (and I have already downloaded it onto my Kindle), I think
there is an excellent chance I will have written my last IDL program
before I leave. ;-)

> I also wouldn't be too quick to discount the countless hours of
> hair-pulling frustration you have saved many people via your
> webpage/posts/library/etc. Karmic currency, you know? Doesn't
> help you fiscally, but still...

I went to lunch after I wrote that note this morning, and when I got
back there was an e-mail from a student from China I had helped the
other day by waving my hands at an answer for him. He had found a better
answer, and wrote to tell me about it. But, more significant to me, he
bought a $2 consulting contract from my store, to thank me. It's not all
about Karminc currency, but it's surprising how much of it you can buy
for just two dollars.

I appreciate your thoughts, as always.

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 Fri, 08 Mar 2013 00:46:43 GMT

View Forum Message <> Reply to Message

Craig Markwardt writes:

> I'm confused: how will we send you our IDL questions while you are trekking? :-)

Don't know. Too far south for satellite reception. Even my trusty Spot
transmitter can't summon the rescue squad from down there. But, when
they unpack me from the ice 5000 years from now, I'll probably still be
wearing my IDL T-shirt. ;-)

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.")

On Thursday, March 7, 2013 11:12:35 AM UTC-7, David Fanning wrote:
> Paul van Delst writes:
>
>
>
>> If it makes you feel any better, I lay awake at night worrying about
>
>> plausibly incorrect results. In any language.
>
>
>
> Well, it *does* make me feel better. :-)
>
>
>
>> But, I don't really use IDL to produce quantitative results (computing
>
>> means and stddevs don't count)
>
>
>
> I was on a forum web page not too long ago and the folks there were
>
> going off on how horrible IDL was to do anything sensible and how it
>
> confounded expectations, was impossible for new people to learn, etc.
>
> Then some guy jumps in to say, 'Wait, there is a web site out there
>
> where the guy tries to fix all the things that are wrong with IDL.'
>
>
>
> It stopped me. Is that what I've been doing all these years? Fixing IDL?
>
>
>
> I suppose it is. And then I got totally depressed. Why do I do this
>
> work? And why do I do it for free, for God's sake!? I've no money in the
>
> bank, a job that pays very little, and I choose to spend my time fixing
>
> IDL. It's a joke.
>

>
>
> That's when I decided to go trekking in Patagonia. :-)
>
>
>
> Before I leave, though, I wrote my own histogram routine, cgHistogram,
>
> which I am just about to add to the Coyote Library. It fixes the bug
>
> with byte arrays (not fixed officially until IDL 8.2), checks to be sure
>
> the binsize data type matches the data type of the data you are binning,
>
> so you always get the correct results, allows for missing values in the
>
> histogram, will smooth the data if needed, and will return the relative
>
> frequency, instead of the histogram count, if you want that. It is a
>
> HELL of a lot better than the Histogram command, as always. Maybe one or
>
> two people will use it. ;-)
>
>
>
> 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.")

It always amazed me IDL never did these things with the histogram function. It seems so obvious. I just imagined that every user has their own histogram wrapper function that does return the % rather than the count, and returns the actual x-axis of the histogram, etc.

cheers,
bob