Subject: Histogram shift
Posted by Michael Wallace on Tue, 22 Feb 2005 01:11:52 GMT
View Forum Message <> Reply to Message

Okay, another object graphics question for everyone. Is there an easy way to shift a histogram over such that the bars are left justified with the data point rather than having the bars centered on the data point?

I realize that I could shift the data itself and then draw the histogram, but there's a voice in the back of my mind that tells me "you're gonna shoot yourself in the foot, kid." Is there some other way to left justify the plot?

-Mike

Subject: Re: Histogram
Posted by James Kuyper on Tue, 28 Jun 2005 18:29:22 GMT
View Forum Message <> Reply to Message

James wrote:
> Hi guys !
>
> I have a question regarding histograms in general. Is there a "right"
> criteria (e.g. strict matmematical rule, etc..) of chosing the bin size ?
> I'm playing with some data and obviously the histogram looks differently
> with different bin sizes. Any help and references would be extremely
> helpful !

About 3 decades ago I read an argument, based upon the binomial distribution, that the maximum number of bins you should use was 1 more than the logarithm base 2 of the number of data points you're analyzing, rounded upward. Sorry - I can't give you a reference. I've no idea how reasonable that argument is in general.

My impression, trying to use this criteria over many years, is that it works best for data with a distribution that has a wide dynamic range, like the binomial distribution. For data that is more nearly uniform in distribution, a larger number of bins seems to give better results. I've never seen it give more bins than I thought was justified, but it often gives fewer.

Subject: Re: Histogram
Posted by Michael Wallace on Tue, 28 Jun 2005 18:45:49 GMT
View Forum Message <> Reply to Message

> I have a question regarding histograms in general. Is there a "right"
> criteria (e.g. strict matmematical rule, etc..) of chosing the bin size ?
> I'm playing with some data and obviously the histogram looks differently
> with different bin sizes. Any help and references would be extremely
> helpful !

Well, it all depends on what you want to figure out about your data.
There is no "right" or "wrong" mathematical rule when it comes to this.
Choosing the right bin size is an art, but I link to think of it as
the larger the bin size, the more generalized your results will be and
the finer the bin size, the more specific your results will be. Whether
you want more general or more specific is totally dependent on what
you're trying to see in your data. For example, let's say you have some
data with noise in it. If you wanted to create a histogram such that
the noise had less effect (so you could concentrate on the actual data),
you'd use a larger bin size. If you wanted to study the noise itself
and trying to figure out where it is coming from, you'd want a smaller
bin size. That's what I mean by "it all depends."

Some data sets will come with natural boundaries already in place and
those may make good bin sizes. For example, say you are acquiring data
through an instrument and every minute the position of the instrument
changes. In this case, it would be logical to bin the data on minute
intervals so that you can easily drawn correlations between angle and
the data.

Don't know if that helps your thinking or not...

Mike
