Subject: Using histogram (or cghistogram) to get top X percent of distribution?
Posted by rjp23 on Mon, 25 Jan 2016 21:48:20 GMT
View Forum Message <> Reply to Message

I'm not sure if this is just me not quite understanding how histogram works.

I have a large array containing data. I want to identify the array indices of the top 10% (for example) of the data. I thought I should be able to use histogram to do this quickly but can't figure out how.

Any help much appreciated as always :-)

Subject: Re: Using histogram (or cghistogram) to get top X percent of distribution?
Posted by Helder Marchetto on Tue, 26 Jan 2016 14:37:20 GMT
View Forum Message <> Reply to Message

On Monday, January 25, 2016 at 9:48:23 PM UTC, rj...@le.ac.uk wrote:
> I'm not sure if this is just me not quite understanding how histogram works.
>
> I have a large array containing data. I want to identify the array indices of the top 10% (for example) of the data. I thought I should be able to use histogram to do this quickly but can't figure out how.
>
> Any help much appreciated as always :-)

Hi,
I normally do something very similar to scale images: I clip the histogram around a few percent from the lowest and highest side. This can be a good reference to find those pixels...
Here is the histogram to perform the histogram percentage clipping:
http://idl.marchetto.de/percentage-of-histogram/

It's not very clean, but works fine for me.

Cheers,
Helder

PS: previously I searched the cumulative histogram until I hit the percentage I wanted, but using value_locate is more efficient.

Subject: Re: Using histogram (or cghistogram) to get top X percent of distribution?
Posted by Med Bennett on Tue, 26 Jan 2016 15:38:12 GMT
View Forum Message <> Reply to Message

On Monday, January 25, 2016 at 2:48:23 PM UTC-7, rj...@le.ac.uk wrote:
> I'm not sure if this is just me not quite understanding how histogram works.

>
> I have a large array containing data. I want to identify the array indices of the top 10% (for example) of the data. I thought I should be able to use histogram to do this quickly but can't figure out how.
>
> Any help much appreciated as always :-)

If all you need are the indices of the top 10% of the data, you can also just do

IDL> s = sort(data)
IDL> idx = s[round(0.9*n_elements(s)):*]

Subject: Re: Using histogram (or cghistogram) to get top X percent of distribution?
Posted by Med Bennett on Tue, 26 Jan 2016 15:46:28 GMT
View Forum Message <> Reply to Message

On Monday, January 25, 2016 at 2:48:23 PM UTC-7, rj...@le.ac.uk wrote:
> I'm not sure if this is just me not quite understanding how histogram works.
>
> I have a large array containing data. I want to identify the array indices of the top 10% (for example) of the data. I thought I should be able to use histogram to do this quickly but can't figure out how.
>
> Any help much appreciated as always :-)

If all you need are the indices of the top 10% of the data, you can also just do

IDL> s = sort(data)
IDL> idx = s[round(0.9*n_elements(s))+1:*]

Subject: Re: Using histogram (or cghistogram) to get top X percent of distribution?
Posted by penteado on Wed, 27 Jan 2016 14:26:08 GMT
View Forum Message <> Reply to Message

I have a function for that:

http://www.ppenteado.net/idl/pp_lib/doc/pp_quartile.html

On Monday, January 25, 2016 at 1:48:23 PM UTC-8, rj...@le.ac.uk wrote:
> I'm not sure if this is just me not quite understanding how histogram works.
>
> I have a large array containing data. I want to identify the array indices of the top 10% (for example) of the data. I thought I should be able to use histogram to do this quickly but can't figure out how.
>

> Any help much appreciated as always :-)

Subject: Re: Using histogram (or cghistogram) to get top X percent of distribution?
Posted by pfp on Wed, 27 Jan 2016 14:51:08 GMT
View Forum Message <> Reply to Message

On Wednesday, January 27, 2016 at 7:26:12 AM UTC-7, Paulo Penteado wrote:
> I have a function for that:
>
> http://www.ppenteado.net/idl/pp_lib/doc/pp_quartile.html

Ex:

```
IDL> a=reverse(dindgen(20))
IDL> print,pp_quartile(a,0.9d0,index=ind,sort=s)
      18.000000
IDL> print,a[s[ind:-1]]
      18.000000       19.000000
```
