

Making an Energy Histogram Using Pierre Auger Observatory Surface Detector Data

Histogram

A histogram is essentially a graphical representation of the frequency or distribution of a variable over specific intervals known as “bins”. In other words is a graph that shows how often a specific number or range of numbers comes up in a data set. This manual focuses on building a histogram that shows the frequency at which the Pierre Auger Observatory Surface Detector has recorded cosmic ray showers of certain energies.

Before you start you should know that:

- This set of instructions only assumes a very minimal level of proficiency with computers and Microsoft Excel. It is written mostly with teachers in mind not just as a learning tool but as a document that could be adapted for student use.
- These instructions are written using Microsoft Excel 2007 for PC. Some screens may look different in other versions.
- In order to complete the following steps you may need to install Data Analysis Tool Pack to Excel. If so you may use Microsoft Office Help options or visit this website and follow their directions.

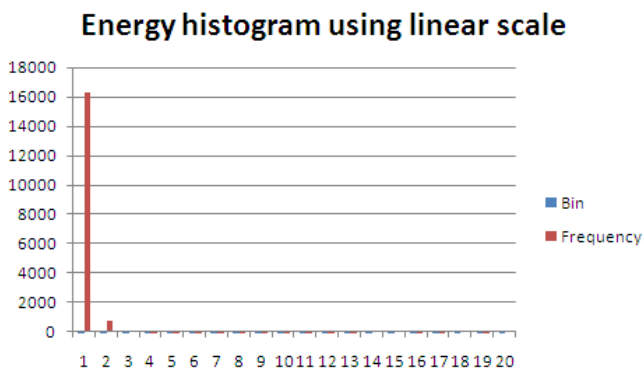
<http://www.dummies.com/how-to/content/how-to-install-the-excel-2007-analysis-toolpak.html>

1) Download and organize the data

- Use the instructions on the Auger website to download the data into an Excel file:
http://www.auger.org/cosmic_rays/guides/downloading_events.html
- Go to the “Sheet 1” and “Sheet 2” tabs and rename them “Data” and “Energy Histogram” respectively. This manual will refer to these as spreadsheets. These spreadsheets are all part of the same document and will be saved as such.



First things first! In order to generate a histogram that shows anything meaningful the data must be presented in a log₁₀ scale. Otherwise the histogram looks like this:



This is not very informative. The axes need to be scaled appropriately so the data get better spread out

among more bins. Our data will have a power distribution so using a log scale will work just fine.

2) Calculate the log₁₀ of the energy data

- Label a new column for your log data

H	I	J
Lat	log Energy	
9.74423		

- Type the log₁₀ function in your new column. Apply the calculation to all the energy data

E	F	G	H	I
Energy	Time	Long	Lat	log Energy
1.5952	1.07E+09	-117.861	9.74423	=LOG10(E2)
0.355838	1.07E+09	-115.096	-2.20735	

Tip: the traditional way of applying a calculation to all cells usually involves some kind of “copy/paste” or “drag down” process that due to the large amount of data in your spreadsheet may be inconvenient. Instead try the following trick:

- Highlight the third cell of any of the other columns. For this example I will chose to use the energy column which is in E3.

E
Energy
1.5952
0.355838
0.613309
0.750369

- Press CTRL+SHIFT+DOWN on your keyboard. This will select every number in that column from E3 down.
- Hit CTRL+C to copy these numbers

E
Energy
1.5952
0.355838
0.613309
0.750369
0.283068
0.333249
0.352402

- Select the second cell down in your “log Energy” column which is I3 in this example. Then press CTRL+V to paste the copied data in this column

I
log Energy
0.202815
0.355838
0.613309
0.750369
0.283068
0.333249
0.352402

- Select the first number in your “log Energy” column. This is the number that was actually calculated in Step 2.

I
log Energy
0.202815
0.355838

- Press CTRL+C to copy the log function in this cell then press CTRL+SHIFT+DOWN to select every number in this column then press CTRL+V to paste the log function in every cell in this column. PRESTO you are now ready to create your histogram.

I
log Energy
0.202815
-0.44875
-0.21232
-0.12473
-0.54811
-0.47723
-0.45296
0.50170

Note: Aside from speeding up the calculation process this also allows you to perform the calculations on only cells that have numbers in them.

- 3) Determine the minimum and maximum values in your “log Energy” column – this will help you determine your “Bin” values later

- Label two new columns in Min and Max

I	J	K
log Energy	Min	Max
0.202815		

- Under the “Min” label type the min function then select the first value in the “log Energy” column

I	J	K	L
log Energy	Min	Max	
0.202815	=min(I2		
-0.44875	MIN(number1, [number2], ...)		

- Don’t close parenthesis yet! Now press CTRL+SHIFT+DOWN to select every value in this column. Press Enter.

J
Min
-0.95713

- Repeat this process in the “Max” Column only this time use the Max function instead

I	J	K	L
log Energy	Min	Max	
0.202815	-0.95713	1.613522	
-0.44875			

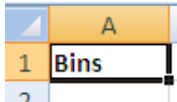
- 4) Create your “Bins”

The “Bins” are essentially placeholders in the horizontal axis in which Excel will graph a specific range of numbers. For example this manual uses a “Bin Size” of 0.1. This means that any number between 0.1 and 0.2 will be graphed in the same column.

- For organizational purposes this guide starts the histogram the “Energy Histogram” spreadsheet. Go to this spreadsheet by clicking on the “Energy Histogram” tab created at the beginning of this manual.



- In your new sheet label a column “Bins”.



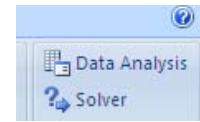
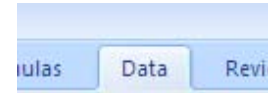
- Fill the “Bins” column with a range of numbers with reasonable increments. Since the Min and Max in this Manual’s example are -0.95713 and 1.61352 respectively we will use bin values starting at -0.9 and ending at 1.6 with 0.1 increments

Note: Choosing the right bins can be tricky depending on the minimum and maximum values of your data and the kind of data distribution desired. Experimenting with the bin values to observe differences in data distribution is recommended.

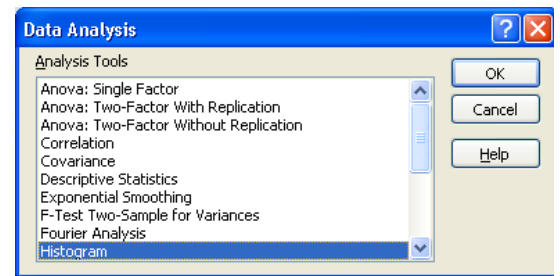
	A
1	Bins
2	-0.9
3	-0.8
4	-0.7
5	-0.6
6	-0.5
7	-0.4
8	-0.3
9	-0.2
10	-0.1
11	0
12	0.1
13	0.2
14	0.3
15	0.4
16	0.5
17	0.6
18	0.7
19	0.8
20	0.9
21	1
22	1.1
23	1.2
24	1.3
25	1.4
26	1.5
27	1.6


5) Find the Frequency of Occurrences

- Click on the “Data” tab in the Excel menu then click on “Data Analysis”



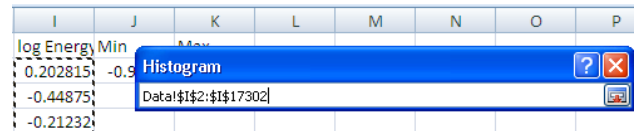
- From the pop up menu choose “Histogram” then click “OK”




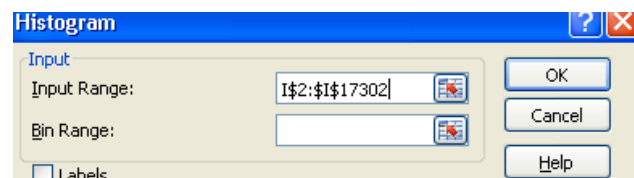
- Click on the “Input Range” collapse box . The window will become smaller and will be awaiting your input.




- Go to the “Data” spreadsheet and select the “log Energy” data by selecting the first cell and using pressing CTRL+SHIFT+DOWN

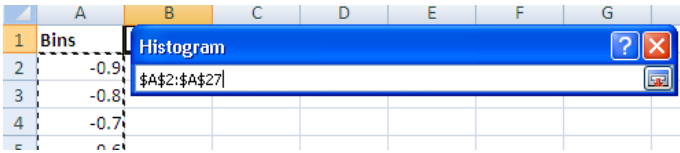


- Click the collapse box  to return to the input window.

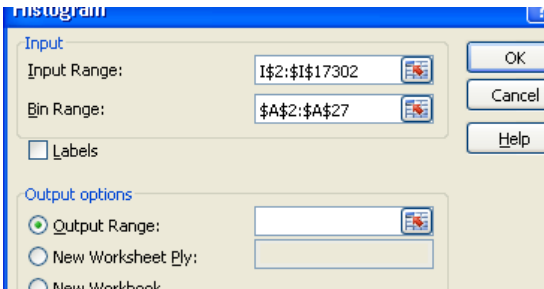



- Click the the “Bin Range” Collapse box . This process will be similar to the Input range process.

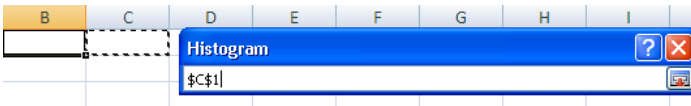
- Go to the “Energy Histogram” spreadsheet and select the “Bin Range” numbers created in step 4.


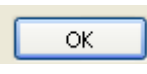


- Click the collapse box  to return to the input window.



- Click the “Output Range” Collapse box . With this you will select the cells you wish to generate in.
- Select any reasonable cell in your “Energy Histogram” spreadsheet.

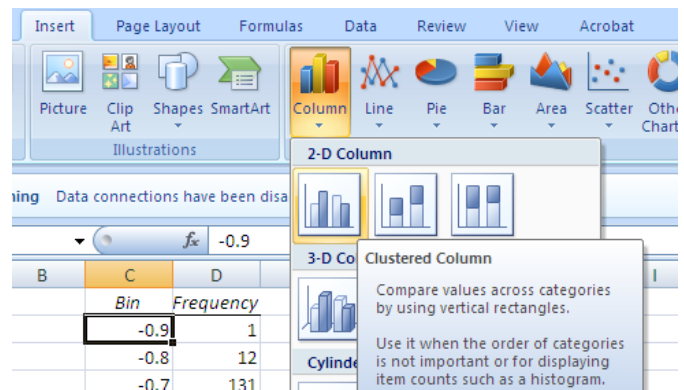


- Click the collapse box  to return to the input window.
- Click “OK” . Your Frequency Range and bins should appear. This is the data that will be graphed in your histogram.

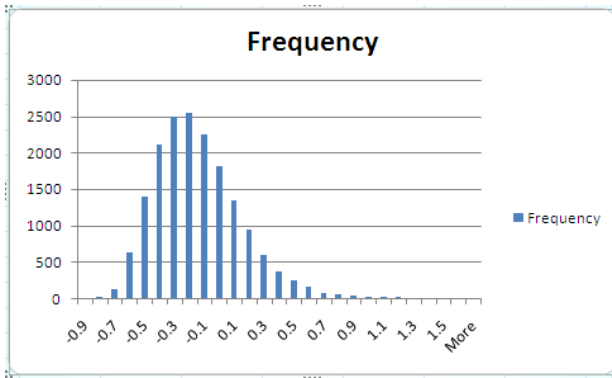
Bin	Frequency
-0.9	1
-0.8	12
-0.7	131
-0.6	636
-0.5	1404
-0.4	2110
-0.3	2503
-0.2	2555
-0.1	2257
0	1819
0.1	1348
0.2	945
0.3	589
0.4	367
0.5	242
0.6	152
0.7	79
0.8	49
0.9	34
1	21
1.1	18
1.2	12
1.3	6
1.4	3
1.5	3
1.6	4
More	1

6) Generate Graphical Histogram

- Select the first value in the “Bins” column. DO NOT SELECT BOTH COLUMNS; this will cause the X axis values to be displayed incorrectly.
- Go to the “Insert” tab in Excel and Click on “Column” in the graph menu, then choose “Clustered Column” from this list. Excel should automatically generate a table.

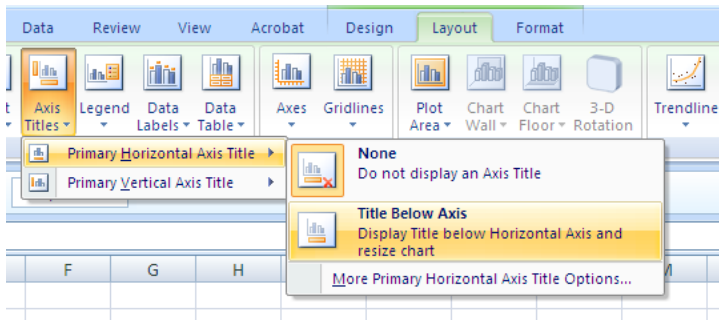


- Your graph should look something like this



7) Insert Titles and Labels

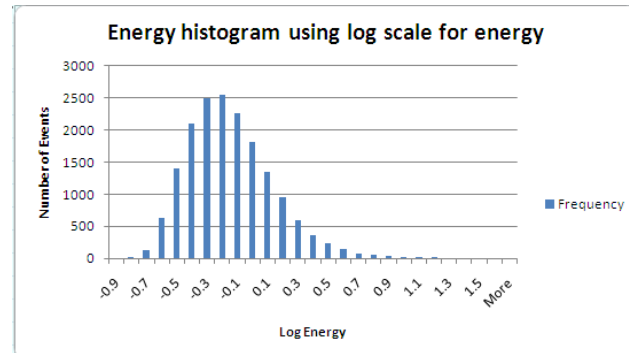
- Click your graph to select it. Go to the "Layout" tab.
- Go to "Axis Titles", select "Primary Horizontal Axis Title" and click on "Title Below Axis"



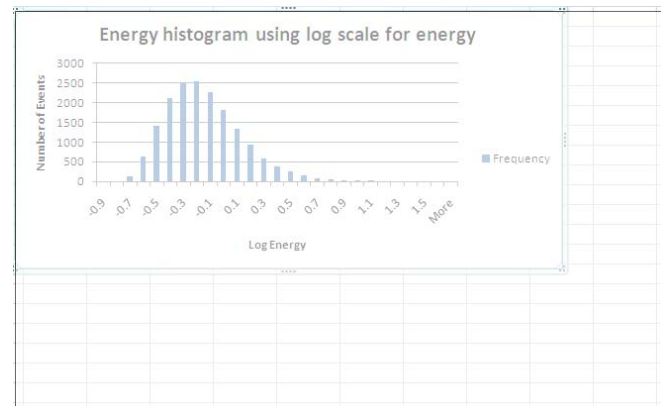
- Type in an appropriate name for the Horizontal Axis
- Repeat this process for the Vertical Axis
- The Title Can be typed in directly into the graph

Sensitivity of the Observatory

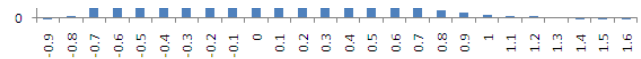
When observing the histogram, you'll notice that the graph seems to almost grow linearly and then decreases exponentially. This is due to the sensitivity range of the detector. As the energy of the cosmic ray showers increases the detector becomes more sensitive until it reaches a maximum. When reporting their results scientists only report using the data at which the detector is 100% reliable. The circled section of the graph below is usually ignored when reporting data.



- The amount of detail that can be observed in each axis can also be edited and it mainly depends on the size of the graph. The larger the graph the more detail can be seen. To increase the size of the graph, simply click on the bottom right corner of the graph and while holding the mouse button "drag" the corner of the graph until desired size is obtained.



- You'll notice that you now see a larger number of "Bins" on the x axis.



Energy histogram using log scale for energy

