Chandra X-ray Center

URL: http://cxc.harvard.edu/ciao3.4/dmgroup.html
Last modified: December 2006

AHELP for CIAO 3.4

dmgroup

Context: tools

Jump to: Description Examples Parameters CHANGES IN CIAO 3.3 CHANGES IN CIAO 3.0 Bugs See Also

Synopsis

Group a specified column in a table with various options

Syntax

```
dmgroup infile outfile grouptype grouptypeval binspec xcolumn ycolumn
[tabspec] [tabcolumn] [stopspec] [stopcolumn] [errcolumn] [maxlength]
[clobber] [verbose]
```

Description

'dmgroup' takes as input a FITS table, and generates a new table by grouping elements of the column 'xcolumn' in the user specified column 'ycolumn'. Elements in 'xcolumn' are grouped according to the grouping method chosen (NONE, BIN, SNR, NUM_BINS, NUM_CTS, ADAPTIVE, ADAPTIVE_SNR, BIN_WIDTH, MIN_SLOPE, MAX_SLOPE, BIN_FILE). The various grouping methods each have required and optional parameters. With the exception of the NONE method, the tabspec, tabcolumn, stopspec, and stopcolum parameters may be used to restrict and refine the grouping using any method. Note that input files may also be stacks ('@stackfile.txt') provided the number of input files matches the number of output files in both infile and outfile stacks.

With all grouping methods, grouping−specific columns will be created if they do not yet exist in the input file. The GROUPING column's values designate the groups. The QUALITY column denotes the quality of each group. The GRP_NUM column enumerates the groups. The CHANS_PER_GROUP column counts the number of channels within each group. The GRP_DATA column gives the total number of 'counts' in each group. And the GRP_STAT_ERR shows the total statistical error of each group.

GROUPTYPE OPTIONS:

- NONE: The data are grouped using maximum resolution. This means that each channel in the GROUPING column will be reset to '1' (Group begin) and each channel in the QUALITY column will be reset to '0' (Good quality). This option may be used to reset the grouping scheme of a given file.
- BIN: The user may explicitly state the binning boundaries according to the given binspec. The binspec applies to the column given with the 'xcolumn' parameter (i.e. time, energy, channels, etc.). Certain 'xcolumn' elements may be excluded by using the optional tabspec and tabcolun parameters. NOTE: If binspec is unspecified but grouptypeval is set, then grouptypeval will be used as the binning specification.

Also, ranges may be grouped at maximum resolution using the stopspec and stopcolumn parameters. The parameters maxlength, errcolumn, and grouptypeval are ignored.

- SNR: Data in 'ycolumn' are grouped until the square root of the number of counts in each group exceeds the given signal−to−noise value specified in the grouptypeval parameter. It assumes that Poisson statistics apply to the grouped column. Group sizes will not exceed maxlength if given.
- NUM_BINS: Data in 'ycolumn' are grouped into a number of equal−width groups starting with the first element of 'xcolumn'. The widths of the groups depend on the number of elements of 'xcolumn' and the number of desired bins given in grouptypeval.
- NUM_CTS: Data in 'ycolumn' are grouped until the number of counts in each group exceeds the minimum number of counts specified in grouptypeval. Group sizes will not exceed maxlength if given.
- ADAPTIVE: Keeps bright features ungrouped, while grouping low SNR regions. Data in 'ycolumn' are first grouped into bins comprised of single 'xcolumn' elements, whose counts exceed the number of counts specified in grouptypeval. From the remaining ungrouped elements of 'xcolumn', data are then grouped into bins comprised of 2 contiguous elements in which the resulting bin would have counts exceeding the minimum value chosen in grouptypeval. The remaining ungrouped data are then similarly grouped into groups of size 3, and so forth. The 'xcolumn' elements that remain ungrouped after this process are given a quality flag=2. Note that this option may greatly increase the run−time, especially when the grouping is performed on large tables (e.g. grating pha1 spectra, with ~8000 rows). Group sizes will not exceed maxlength if given.
- ADAPTIVE_SNR: Works similarly to the ADAPTIVE method, but instead of using a count threshold to determine group cutoffs, the specified signal−to−noise ratio is used. The ratio is given in grouptypeval, and applies to the column 'ycolumn'. Group sizes will not exceed maxlength if given.
- BIN_WIDTH: Similar to the NUM_BINS method, will create equal−width groups beginning with the first element of 'xcolumn'. Instead of specifying the number of bins, the user designates the desired width of each bin using the grouptypeval parameter.
- MIN_SLOPE: Groups are formed when the absolute value of the slope of the data in 'ycolumn' is less than the user−specified minimum slope threshold given in grouptypeval. The slope is calculated as: delta[ycolumn]/delta[xcolumn], by stepping through and using the data of the respective columns. The maxlength parameter can be used to limit group size.
- MAX_SLOPE: Identical to MIN_SLOPE calculation, except that grouptypeval refers to a maximum slope threshold. A group is formed if its average slope exceeds this value.
- BIN_FILE: Data are grouped as they were in a previously grouped file. The column to use to determine the grouping must be specified with the xcolumn parameter. The file to mimic is given using binspec='/full/path/to/file.fits'. The input file and binfile need not be the same length in channels. BIN_FILE will determine the data value ranges in the specified column, and will try to reconstruct the original binspec and tabspec. However, the binfile must be of PHA Type I format. Note that the quality column is not transferred to the output, as the original grouping method cannot be determined. Note also that the upper bound of a group is always taken to be the last data value included in the group, regardless of whether the data column is of type integer or real. While this is appropriate for integer data columns, for real data columns the value of upper bound in the original group method was somewhere between the value in last included row and the following row. Since the exact value cannot be determined, the last data value in the original group is used as the best approximation to the upper bound for real data columns.

INCOMPLETE GROUPS:

For each grouping option, a group may be left incomplete due to insufficient counts or an insufficient number of rows. This may occur because of a tab interval, the end of the dataset, or the presence of a previously grouped row. All rows in such an incomplete group will be given a quality flag=2. For example, for the option NUM_BINS with the grouptypeval=4 and an input file with 26 rows, the resulting 4 groups would be (in the case of no tab intervals): group 1 (rows 1−6), group 2 (rows 7−12), group 3 (rows 13−18), and group 4 (rows 19−24), with rows 25 and 26 ungrouped and given quality flags=2.

BIN, TAB, AND STOP SPECIFICATIONS:

The binning specification, tab specification, and stop specification all share the same syntax, and are now consistent with the dmbinning syntax. The binspec, tabspec, and stopspec are applied respectively to the 'xcolumn', 'tabcolumn', and 'stopcolumn' column data when used. The syntax is given by spec='min:max:step,min:max:#bins,...'. One or two of the three values may be omitted (ex: spec='1::10'). Any omitted min, max, or step values will be determined from the file if possible using the TLMINn, TLMAXn, and TDBIN keywords if available. The values given in these specifications must be within the data value range of the table.

Tabs are used to label intervals of data to not be binned over (e.g. instrumental features). They can be used for any grouping option except NONE. Each row appearing in a tab specification is not grouped and given a quality flag=5. If the column name given in 'tabcolumn' is a valid column in the input file, the tab specification(s) will pertain to data in this column. If 'tabcolumn' is left blank, the tab specification(s) will apply to the row numbers in the input table.

Stops are used to group specified ranges of data at maximum resolution. That is, each channel within the stopspec will be in its own group, and will be marked with quality=0. Stops can be used for all grouping methods except NONE.

OUTPUT TABLE ADDITIONS:

The output table contains all the columns present in the input table, plus the following additional columns:

- GRP_NUM: The number of the group in which that row falls.
- CHANS_PER_GRP: The total number of rows in the given group.
- GRP_DATA: The total number of counts in the given group.
- GROUPING: Equal to 1 if the row is the first in a group; equal to −1 otherwise.
- QUALITY: The quality flag for a given group (same conventions as for the ftool GRPPHA): 0 for good (grouped) data; 5 for data labeled as bad by the user (within a tab), and 2 for data labeled as questionable by dmgroup (incomplete groups, etc.).
- GRP_STAT_ERR: The statistical error for the group. If the error column is supplied, it is the value of these errors added in quadrature. If no error column is supplied, it is sqrt(grp_data).

Example 1

```
dmgroup in.fits out.fits grouptype=BIN xcolumn=pha grouptypeval=""
binspec="10:40:5" ycolumn=counts tabspec="25:32:#1" tabcolumn=pha
```

Group the data in column 'counts' of the input file in.fits using the BIN method. Begin grouping where the pha column=10, creating new groups whenever pha increases by 5 until reaching 40. Exclude channels where pha is between 25 and 32 inclusive, marking them with bad quality.

Example 2

```
dmgroup @instack.txt @outstack.txt grouptype=SNR xcolumn=channel
grouptypeval=30 binspec="" ycolumn=counts
```

Group the data in column 'counts' of each of the input files listed in instack.txt, mapping the output to the files listed in outstack.txt. Construct groups with a minimum signal−to−noise ratio of 30.

Example 3

```
dmgroup in.fits out.fits grouptype=ADAPTIVE xcolumn=chipx
grouptypeval=100 binspec="" ycolumn=counts
tabspec="55710000:55760000:#1" tabcolumn=time
```

Group the data in column 'counts' of in.fits creating groups with a minimum of 100 events each. The rows specified by the tab specification (rows in which the time has values 55710000 to 55760000 inclusive) will remain ungrouped and flagged bad. Note that although an xcolumn is given ('chipx'), it is not used in this example.

Example 4

```
dmgroup in.fits out.fits grouptype=NUM_BINS xcolumn=channel
grouptypeval=20 binspec="" ycolumn=counts stopspec="50:150:"
stopcolumn="channel"
```

Group the data in column 'counts' of in.fits, making 20 equal−sized (if possible) groups of 'channel'. Group channels 50−150 inclusive at the highest resolution.

Example 5

```
dmgroup in.fits out.fits grouptype=NUM_CTS xcolumn=pi grouptypeval=270
binspec="" ycolumn=counts
```

Group the data in column 'counts' of in.fits creating groups containing a minimum of 270 events.

Example 6

```
dmgroup in.fits out.fits grouptype=NONE xcolumn=time grouptypeval=0
binspec="" ycolumn=counts
```

Reset the grouping, grouping everything at the highest resolution. Group each channel by itself and give it a good quality (0).

Example 7

```
dmgroup in.fits out.fits grouptype=MAX_SLOPE xcolumn=channel
grouptypeval=100 binspec="" ycolumn=counts
```

Group those features in in.fits that increase or decrease at a rate higher than 100 counts/channel.

Example 8

```
dmgroup in.fits out.fits grouptype=BIN_FILE xcolumn=channel
grouptypeval=0 binspec="/data/pha/grouped.fits" ycolumn=counts
```

Group in.fits the same way /data/pha/grouped.fits has been grouped. Reconstruct the original binspec using the 'channel' column.

Parameters

name	type	ftype	def	min	max	reqd
infile	string	input				yes
outfile	string	output				yes
grouptype	string		NONE			yes
grouptypeval	real		0			yes
binspec	string					yes
xcolumn	string					yes
ycolumn	string					yes
tabspec	string					no
tabcolumn	string					no
stopspec	string					no
stopcolumn	string					no
errcolumn	string					no
clobber	boolean		no			no
verbose	integer		0	0	5	no
maxlength	real		0	0		no

Detailed Parameter Descriptions

Parameter=infile (`string required filetype=input`)

The input file. Can be a fits table or a stack of tables.

Parameter=outfile (`string required filetype=output`)

The output file. Can be a file name for fits output or a stack of output filenames.

Parameter=grouptype (string required default=NONE)

Type of grouping: one of NONE, BIN, SNR, NUM_BINS, NUM_CTS, ADAPTIVE, ADAPTIVE_SNR, BIN_WIDTH, MIN_SLOPE, MAX_SLOPE, BIN_FILE, and must be in upper case.

Parameter=grouptypeval (real required default=0)

The numerical value used for the methods SNR, NUM_BINS, NUM_CTS, ADAPTIVE, ADAPTIVE_SNR, MIN_SLOPE, MAX_SLOPE, BIN, and BIN_WIDTH. Ignored otherwise.

Parameter=binspec (string required)

The binning specification having syntax consistent with the Data Model dmbinning syntax. For grouptype=BIN, if binspec is unspecified but grouptypeval is set, then grouptypeval will be used as the binning specification.

Parameter=xcolumn (string required)

The x−axis of the data. Only used in the following grouping method: BIN, BIN_FILE, BIN_WIDTH, MIN_SLOPE, and MAX_SLOPE. Typically values depending on file type might be 'channel', 'pha', 'pi', 'time', 'energy', although any column could be used. It must contain monotonically increasing data when the BIN method is selected.

Parameter=ycolumn (string required)

The name of the column of the input table containing the data to be counted and used to compute the GRP_DATA column. Typically values might be something like 'counts', 'rate', or 'surface_brightness' depending on the input file type although any value can be used.

Parameter=tabspec (string not required)

The tab specification having syntax consistent with the Data Model dmbinning syntax. Denotes the channels not to be grouped, but instead marked as bad.

Parameter=tabcolumn (string not required)

The name of the column of the input table containing the data that have to be used by the tabspec. It must contain monotonically increasing data.

Parameter=stopspec (string not required)

The stop specification having syntax consistent with the Data Model dmbinning syntax. Denotes the channels to be grouped at the highest resolution.

Parameter=stopcolumn (string not required)

The name of the column of the input table containing the data that have to be used by the stopspec. It must contain monotonically increasing data.

Parameter=errcolumn `(string not required)`

The name of the column of the input table containing the data that should be used to calculate the GRP_STAT_ERROR values and the signal−to−noise when using the SNR and ADAPTIVE_SNR methods. If no value is given, sqrt(grp_data) is used.

Parameter=clobber `(boolean not required default=no)`

Specifies if an existing output file should be overwritten.

Parameter=verbose `(integer not required default=0 min=0 max=5)`

Specifies the level of verbosity in displaying diagnostic messages.

Parameter=maxlength `(real not required default=0 min=0 max=)`

Specifies the maximum group size of groups. If set to 0, then no maximum is specified. Only applies to methods BIN, SNR, NUM_CTS, ADAPTIVE, ADAPTIVE_SNR, MIN_SLOPE, MAX_SLOPE, and is ignored otherwise.

CHANGES IN CIAO 3.3

Changes to the error calculations in GRP_STAT_ERR column

If the "errcolumn" parameter is left blank (the default), the GRP_STAT_ERR column is calculated assuming SQRT(GRP_DATA).

CHANGES IN CIAO 3.0

New 'column' Parameters

In 3.0 the parameters 'bincolumn' and 'column' have been removed, and the parameters 'xcolumn' and 'ycolumn' have been added. Note that there is no direct correlation between old and new parameters.

'ycolumn' is the column that contains the data to be grouped (i.e. counts in a spectrum), and must always be specified. 'xcolumn' is the column over which the grouping scheme is defined. 'xcolumn' is only used when grouptype=BIN, BIN_FILE, BIN_WIDTH, MIN_SLOPE, or MAX_SLOPE. If the grouping option BIN is chosen, then 'xcolumn' must contain monotonically increasing data. If the xcolumn is needed and not specified (e.g. you cannot use one of the SLOPEs without specifying both the X and Y column), then the row number is used.

One can therefore say, for example:

```
  unix% dmgroup in.fits out.fits ycolumn='counts'
        grouptype='BIN_WIDTH' grouptypeval=1.e−3 binspec=""
        xcolumn="ENERG_LO"
```

to group a grating spectrum in bins with width of 1 eV.

'stop'−Columns

In addition to the parameters 'tabcolumn' and 'tabspec', the two parameters 'stopcolumn' and stopspec' have been added. The two 'stop'−parameters are complementary to the two 'tab'−parameters, and produce essentially the same effect: i.e. leaving the bins in between two consecutive 'tab' or 'stop' specifications, ungrouped (tab) or 'grouped' at full resolution (stop). The difference, then, is in the 'quality' flag of these bins: ungrouped channels (i.e. tabs) are given quality=5 (i.e. 'bad' data, e.g. instrumental features), while full−resolution−grouped channels (i.e. stops) are given quality flag=0 (i.e. good, 'grouped' channels).

Errors

The parameter 'errcolumn' has also been added. This parameter defines the column containing the data to be added in quadrature to compute the GRP_STAT_ERR column. If 'errcolumn' is left blank (the default), errors on grouped data (i.e. data in the output column GRP_DATA), are assumed to be equal to:

```
        GRP_STAT_ERR = [1 + SQRT(ycolumn + 0.75)]
```

If, instead, 'errcolumn' is defined, then the column GRP_STAT_ERR will be filled with the sum in quadrature of the data contained in each new group of the column 'errcolumn'.

Maximum Length of a Grouped Channel

The user can specify the maximum length of a new grouped channel, using the new parameter 'maxlength'. If 'maxlength=0' (or it is left blank), no maximum length criterion is applied. Otherwise data are grouped according to the chosen grouping method ('grouptype') and scheme ('xcolumn' and 'grouptypeval'), but not to exceed the maximum number of rows defined in 'maxlength'. This new parameter is ignored with the grouping options BIN, BIN_WIDTH and BIN_FILE.

Bugs

See the bugs page for this tool on the CIAO website for an up−to−date listing of known bugs.

See Also

dm

dmbinning, dmcols, dmfiltering, dmopt

tools

dmappend, dmarfadd, dmcopy, dmextract, dmfilth, dmgti, dmimghist, dmjoin, dmmerge, dmpaste, dmregrid, dmsort, dmtcalc, dmtype2split, fullgarf, get_sky_limits, mkacisrmf, mkarf, mkexpmap, mkgarf, mkgrmf, mkinstmap, mkpsf, mkrmf, mkwarf, tgextract

The Chandra X−Ray Center (CXC) is operated for NASA by the Smithsonian
Astrophysical Observatory.
60 Garden Street, Cambridge, MA 02138 USA.

URL:
http://cxc.harvard.edu/ciao3.4/dmgroup.html
Last modified: December 2006