Subject: Re: name vs score 2D plot
Posted by nbehrnd on Wed, 08 Mar 2023 20:32:22 GMT
View Forum Message <> Reply to Message

Hi sublimeuser,

a set of 300k indeed might a bit large (for DW). In such a case I recommend to have a look into AWK. Though initially written for text processing, this «Swiss pocket knive» understands some mathematics you can use to filter by thresholds. Based on the assumption your raw data file is organized as two-column ASCII like `test_tsv.txt` with docking score in the second column, it can be used e.g.

+ to report only the data with an entry in the second column higher than 0.2:

```shell
$ awk '{if ($2 > 0.2) print}' test_tsv.txt
```

+ to report only the data where the second column's entries are in the in the interval between 0.2 and 0.8:

```shell
$ awk '{if ($2 > 0.2 && $2 < 0.8) print}' test_tsv.txt
```

which you can redirect into a permanent record either by overwriting the old content (`>`), or by append (`>>`). In case of access to an installation of Linux, you can combine this with a line count (`wc -l`) you either can run on the newly written record

```shell
$ awk '{if ($2 > 0.2 && $2 < 0.8) print}' test_tsv.txt > records.txt && wc -l records.txt
2 records.txt
```

or pipe

```shell
$ awk '{if ($2 > 0.2 && $2 < 0.8) print}' test_tsv.txt | wc -l
2
```

Though this equally removes the headers, the newly written (filtered) record files should be less resource hungry and hence accessible to DW (Edit -> Paste Special -> Paste Without a Header Row).

With regards,

Norwid