Subject: Re: Using Set Value Range option
Posted by thomas on Sat, 31 Dec 2016 16:05:46 GMT
View Forum Message <> Reply to Message

The implicit value range of a numerical column is the span from the lowest existing to the highest existing value. It is mainly used to define the maximum visible range on axes in views. The value range option allows to extend this range, but not to reduce it, because a reduction would cause that some values would never be shown. On the other hand visible ranges on views can be achieved easily by the individual view sliders.

The selectivity score that is used is the Gini score described in Wikipedia and for instance used by P.P Graczyk as a compounds selectivity measure against a spectrum of kinases. The code used to calculate the score is simple (see below). If you have less than two columns containing valud numerical values the score is a NaN.

For every row the following is done:
A 'value' array contains the row values of the selected columns, a 'sum' array (same size) is provided:

```
Arrays.sort(value);  // sort the value array in ascending order
int count = 0;
float area = 0;  // to determine area of cumulative value sums

if (!Float.isNaN(value[0])) {  // we skip not-a-number values
  sum[0] = value[0];
  count++;
  while (count<columnCount && !Float.isNaN(value[count])) {
    area += sum[count-1];
    sum[count] = sum[count - 1] + value[count];
    count++;
  }
}

return (count <= 1) ? Double.NaN : 1.0 - 2.0 * area / (sum[count-1] * count);
```

Hope this explains it. I will describe it a little more detailled in the manual.