Subject: Re: Duplicates - a new twist
Posted by Chris[1] on Tue, 18 May 2004 11:42:32 GMT
View Forum Message <> Reply to Message

The easiest way is to take copies of the lats & lons, reduce them to the resolution you think is sufficient that the same station has the same coordinate (independent of which country reports it); sort on one (say latitude), and then look for unique values, using uniq() on the sorted values. Then use the output from uniq() to look at whether points with the same latitude also have the same longitude. It's an exercise in indexing :)

If you want a more robust technique - one that doesn't fall apart near the poles or the dateline, for example - use a spherical to Cartesian coordinate conversion, and do similarly, except now with the three coordinates.

Cheers;
Chris

"Martin Doyle" <m.doyle@uea.ac.uk> wrote in message
news:d33d6a4b.0405171324.1272c4e0@posting.google.com...
> Hello all,
>
> I have a problem which I've searched everywhere to try and
> solve...many posters on this newsgroup have had _similar_ problems but
> the resolutions didn't help me...anyway, here goes;
>
> I have a dataset which consists of 3 columns: longitude, latitude and
> a value for an emission of an air pollutant. European countries report
> the emission of this pollutant for the latitude longitude coordinates
> which are within their countries. However, some of the latitude,
> longitude coordinates lie on the borders of countries and therefore an
> emission is sometimes reported by 2 or more countries for the same
> coordinate (i,e. There are multiple instances of the same coordinate
> within the dataset).
>
> What I need to do is to look through the dataset and sum the emissions
> when the coordinate is the same, resulting in a dataset with unique
> coordinates and a total emission for each grid point.
>
> Does anyone have any ideas about how to go about this? I've seen posts
> on this newsgroup which have had problems with duplicate values in one
> column of data, but I'm unsure about how to go about it when there are
> 2 columns which need to be examined.
>
> Thanks guys...
>

> All the best,
>
> Martin
