14. Create ordered groups

14.1 Overview

There is a new tool in PRIMER 8 that permits the end-user to *create a set of ordered groups*, based on the numerical values of a given variable.

Rationale

Below we provide a few examples of this tool's potential utility. There are many more!

Binning and consolidating information

A case where we might want 'categories', but where there may be (initially) continuous values, occurs when we have a broad-scale dataset with a *lot of replicates* that span a very *large area* or domain (in terms of latitude and longitude). For example, consider the Continuous Plankton Recorder (CPR Survey. Many other ocenographic, landscape-level and environmental datasets have these characteristics. We might want to *consolidate* that information in some way, perhaps in order to make comparisons with more discrete (less continuous) datasets or variables (e.g., dis-continuous biotic samples at particular sites or regions). We could consider making a 'grid' over the sampling extent (i.e., by creating 1-degree sized bins for each of latitude and longitude) and then get the average values of the variable(s) within those grid cells. Indeed, to understand spatial relationships and patterns, it may be essential to bin samples into latitude (and/or longitude) groupings, so that structural changes through space can be more easily summarised, tested, visualised in plots, etc.

Classification based on an ordination axis

Let's consider another scenario. Perhaps you have just done a principal component analysis (PCA) on a set of morphometric variables (e.g., for a set of individuals belonging to a given species). Suppose the first PC axis explains a lot of the multivariate variation and the position of any individual on that PC axis can essentially be interpreted as a measure of its overall relative size. The values of the individuals on the PC axis are continuous, but maybe you want to classify the individuals into (say) five roughly equal-sized groups along that axis (from small to large). In the past, this task could have taken some considerable time, sorting and data wrangling to get your groups based on the PC axis. Our new tool in PRIMER 8 will create equal sample-sized groups along your chosen continuous variable (e.g., a PC axis, or any other meaningful ordination axis of choice) in a flash.

Setting known break points in the data

Yet another use-case for this tool is the situation where you want to identify a set of samples that fall within a certain range of values for a continuous variable; for example, suppose you want to separate out samples occuring at sites that are below a certain temperature (e.g., 20°C). You can

use the tool to specify the value of 20 as a 'break point', and get a factor that groups samples either side of that break point. Multiple such breaks points can also, optionally, be constructed all at once. For example, suppose you have the continuous variable of depth (in metres); you might want to 'bin' samples into several categories according to their depths, using break points of 10, 20, 30, 40, etc. (We shall see an example of this in the next section.)

Finding natural peaks or gaps to define groupings

You may have a continuous variable that is not evenly distributed across its full range. For example, a histogram may show a multimodal distribution pattern. You may want to define groups that coincide in some natural way with that uneven distribution, such that each group will capture an individual 'mode' or 'peak'. You may not be certain about where the break-points in-between those modes should be placed. In this case, you would want the tool to find suitable break points for you by (say) minimising the within-group sum of squares for the variable, given the number of groups (peaks). If the modes are asymmetric, then sums of distances-to-median (rather than the within-group sum of squares) might be more appropriate to use here. Or you could go even more non-parametric in your criterion and choose groups so as to minimise only the rank inter-point Euclidean distances within each group.

Specifying groups of replicates for dispersion weighting

Many species abundance variables (e.g., counts) display intrinsic mean-variance relationships. In other words, the variance increases with the mean. This relationship is particularly strong and important to consider in the context of species that tend to aggregate, e.g., such as fish that school, birds that flock, mammals that travel in herds, insects that cluster together in swarms, etc. We may wish to accommodate this by applying a pre-treatment to our data called *dispersion weighting* (see Clarke *et al.* (2006a) for details). This pre-treatment option effectively transforms our data to consider the abundances (counts) of *clusters* of organisms (for those species that do indeed demonstrate significant clustering of individuals), rather than analysing the raw abundance values. Of course, the dispersion weights that may be needed here will vary according to the individual species' distribution of counts across the sampling units.

To characterise mean-variance relationships, hence to apply dispersion-weighting, we need *groups* (levels of some factor) across which we can calculate and measure means and variances. In cases where we do not have any groups to begin with (e.g., whenever our sampling units occur along a continuous gradient, in a simple spatial array or regular samples through time, etc.), then there will be no way to use dispersion weighting at all. Usnig this new tool, we can create groups along one (or more) continuous variables which can then form the basis of a dispersion-weighting pre-treatment option.

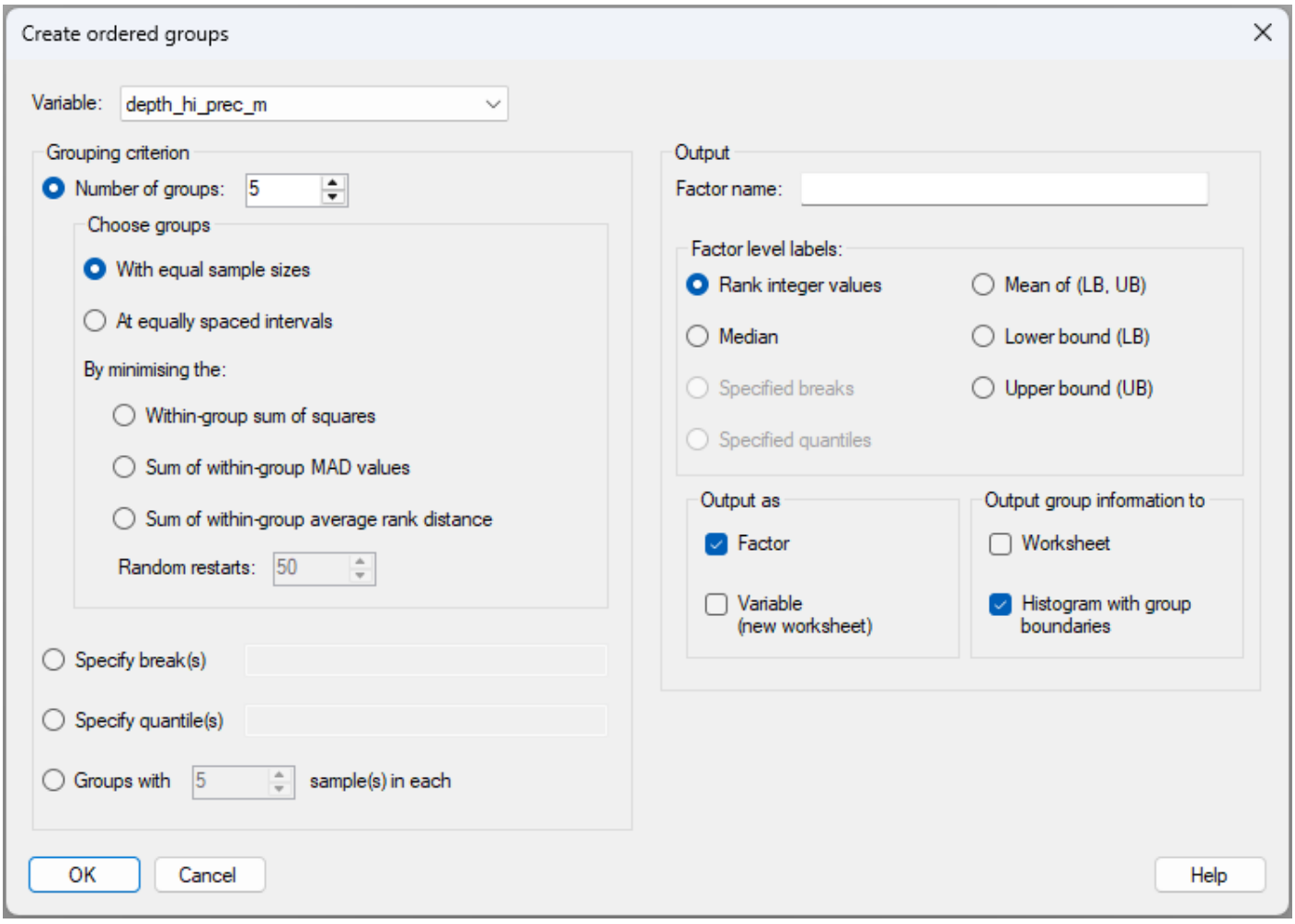
Plotting symbols and colours

It can be extremely useful to create groups from a continuous variable for rather simple practical or logistic reasons. For example, suppose we want to use different colours or symbols for different levels of a continuous variable, such as temperature. We may not want to have a different colour

for every single temperature, but rather we might prefer to have a colour (and/or symbol) for all cases where temperature values occur within some specified range, for each of a set of specified temperature groups (eg., low, medium and high). The use of well-defined symbols and colours like this can make interpretations of patterns in ordinations, especially along gradients, much easier to visualise and communicate to others.

Implementation

This tool is accessed by clicking **Tools** > **Create Ordered Groups...**. The dialog window is shown below.



There is clearly a large diversity of options to choose from to create ordered groups based on a chosen continuous variable, as shown in the dialog above. The various methods that can be used to create groups are itemised and outlined, correspondingly, in more detail below.

Methods for creating groups

Consider an individual variable Y, which has a set of observed sample values y and which our algorithm is going to split into g ordered groups using some criterion. Let the number of sample values in each group be denoted by n_i for $i = 1, \ldots, g$ and the total number of sample values is $\sum_i^g n_i = N$. Also, let y_{ij} denote the j^th^ observed value of Y in the i^th^

group for $j = 1, \ldots, n_i$ and $i = 1, \ldots, g$. In what follows, we shall let $\bar{y}_ i = \sum_{j=1}^{n_i} {y_{ij}} / {n_i} $ be the sample average of the variable for group i, and m_i be the median value for the set of samples in group i.

The methods for choosing ordered groups based on the variable Y include:

- (1). Specify the *number of groups*, g, and, given that number, generate groups:
  - (a). with *equal sample sizes* ($n_1 = n_2 = \cdots = n_g$); or
  - (b). at *equally spaced intervals*, so as to *minimise*:
    - (i). *the within-group sum of squares*, i.e., $\sum_{i=1}^{g}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_ i)^2$; or
    - (ii). *the sum of within-group mean absolute deviations (MAD) from the median*, i.e., $\sum_{i=1}^{g}\sum_{j=1}^{n_i} |y_{ij} - m_i|$; or
    - (iii). *the sum of within-group average rank inter-point Euclidean distances*.
- (2). Specify *break(s)* manually, i.e., a list of specific values of the variable that will serve as break-points between consecutive ordered groups (e.g., 10, 20, 50, 100); or
- (3). Specify the *quantile(s)* of the variable's empirical distribution where you want break-points to be (e.g., 0.25, 0.5, 0.75); or
- (4). Specify groups by nominating a certain value for the *number of samples in each group*, i.e. $n_i = n$ for all i.[¶]

[†]*Dispersion weighting (Clarke et al. (2006a)) is achieved easily and directly in PRIMER 8 by clicking Pre-treatment > Dispersion Weighting....*

[¶]*Note that if there are any 'remainders', given the total sample size, these will be placed in a final group, corresponding to the largest values of the variable, but with reduced n. For example, if you have 10 samples and you ask for groups having a sample size of n = 3, the tool will create 3 ordered groups {1,2,3}, {4,5,6}, {7,8,9}, and one remainder sample will spill into a fourth group on its own {10}.*

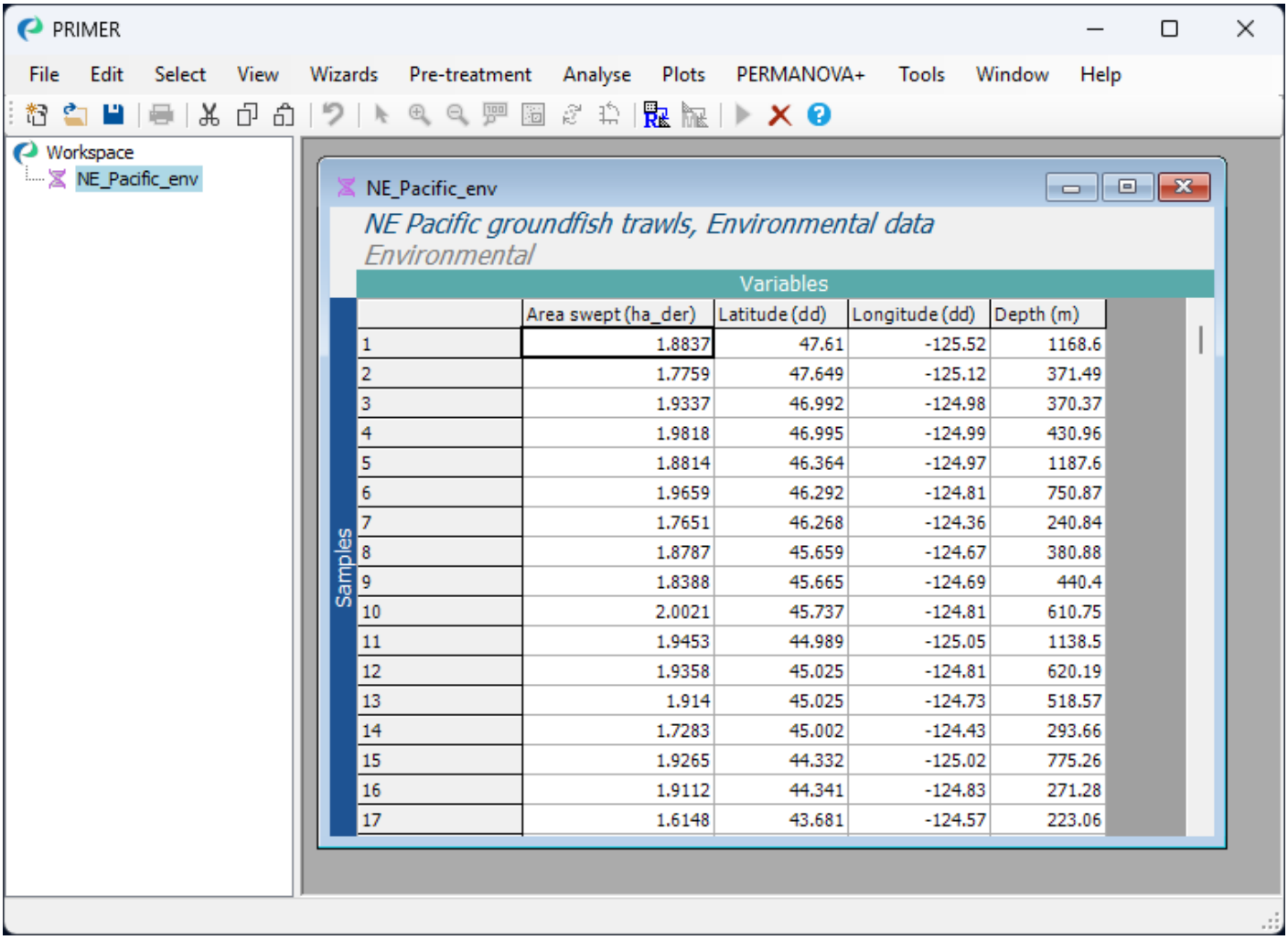
14.2 Example: NE Pacific groundfish vs depth

To demonstrate the creation and use of ordered groups from a continuous variable, we will look at data comprised of an excerpt from the West Coast Groundfish Bottom Trawl (Slope and Shelf Combination) Survey, conducted annually by the National Oceanic and Atmospheric Association (NOAA)'s Northwest Fisheries Science Center¶ and available online (https://www.nwfsc.noaa.gov/data/map). This specific excerpt was created and used by Anderson *et al.* (2022) and is also available on Dryad. Data are provided from trawl surveys for the years from 1999-2018 inclusive. There are two data sheets referred to here (both are found in the Examples_P8 > NE_Pacific_groundfish folder), namely: (1) counts of abundances of p = 310 species of groundfish obtained in each trawl (in the file called 'NE_Pacific_groundfish.pri') and (2) values for latitude, longitude, depth (in m) and the area swept per trawl (in the file called ' NE_Pacific_env.pri').

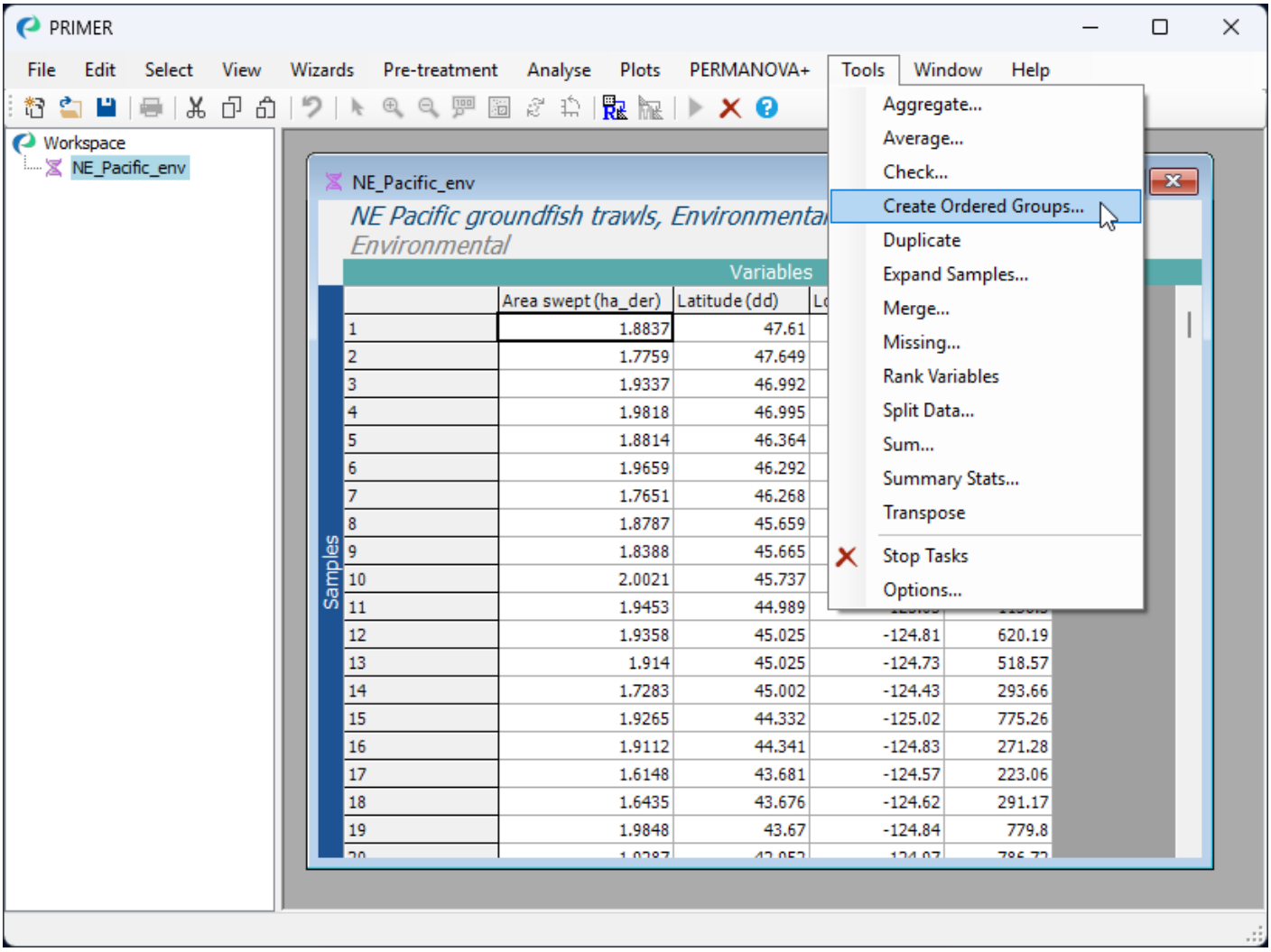
In this example, there are N = 6002 (!) rows of data (sample trawls), with depth values from 24 m - 1,428 m, and latitudes from 32°N to 48°N. Constructing a nMDS ordination plot of these fish assemblages for all 6002 samples is unlikely to be helpful here! To make some sense of these data that span such large latitudinal and depth ranges, it would be helpful to create some groups of samples occuring at similar depths (and also groups of samples occurring at similar latitudes, e.g., see Anderson *et al.* (2013)). Here, we shall focus purely on creating ordered groups of samples from the variable of *depth*. More specifically, our interest lies in uncovering potential changes in the structure of fish assemblages along the depth gradient.

Open file and create ordered depth groupings

1. Open up the file called 'NE_Pacific_env.pri' in PRIMER 8. It will look like this:



2. From the NE_Pacific_env data sheet, click **Tools** > **Create Ordered Groups...**, like so:

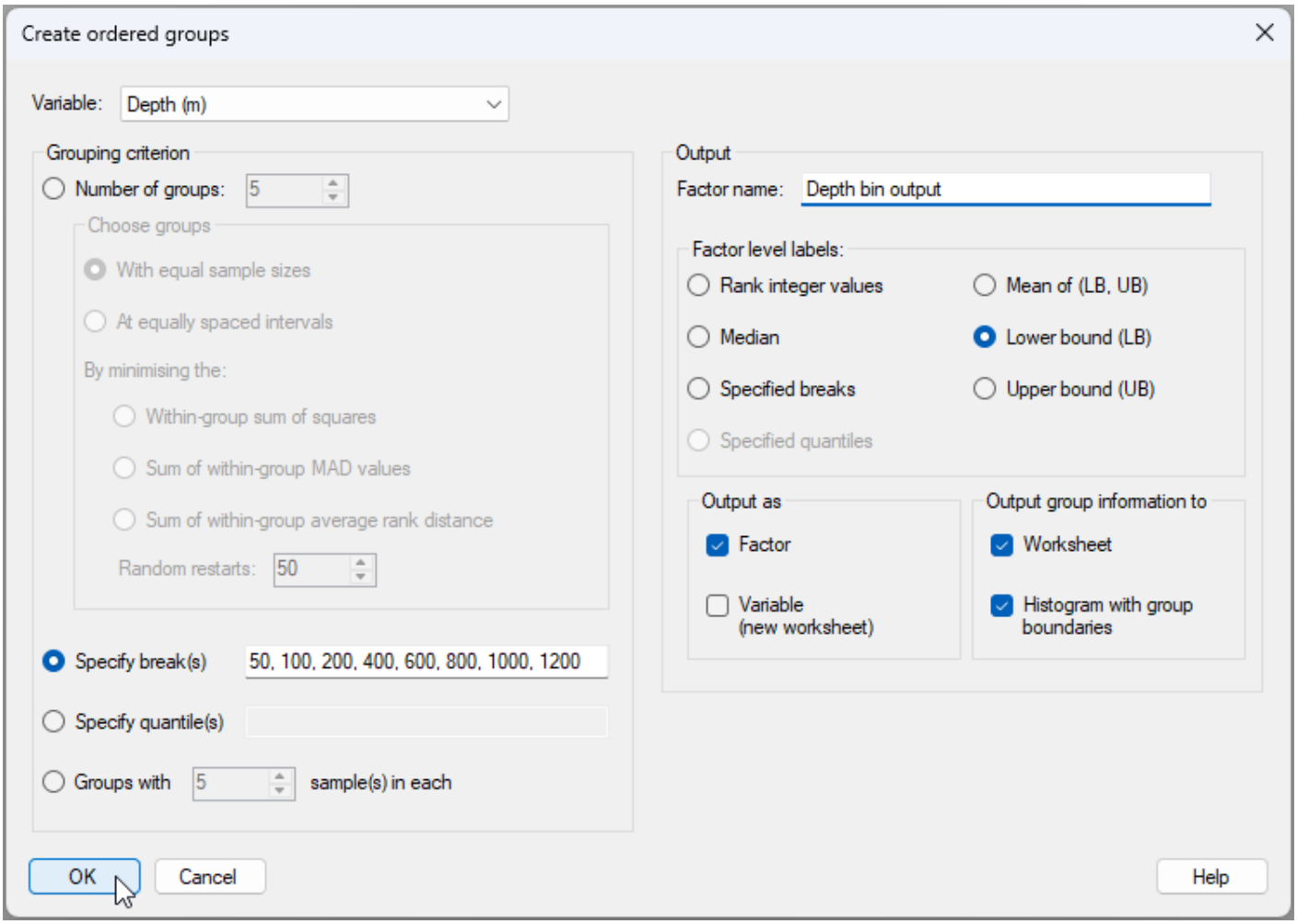


	Area swept (ha_der)	Latitude (dd)		
1	1.8837	47.61		
2	1.7759	47.649		
3	1.9337	46.992		
4	1.9818	46.995		
5	1.8814	46.364		
6	1.9659	46.292		
7	1.7651	46.268		
8	1.8787	45.659		
9	1.8388	45.665		
10	2.0021	45.737		
11	1.9453	44.989		
12	1.9358	45.025	-124.81	620.19
13	1.914	45.025	-124.73	518.57
14	1.7283	45.002	-124.43	293.66
15	1.9265	44.332	-125.02	775.26
16	1.9112	44.341	-124.83	271.28
17	1.6148	43.681	-124.57	223.06
18	1.6435	43.676	-124.62	291.17
19	1.9848	43.67	-124.84	779.8

3. In the resulting dialog, choose:

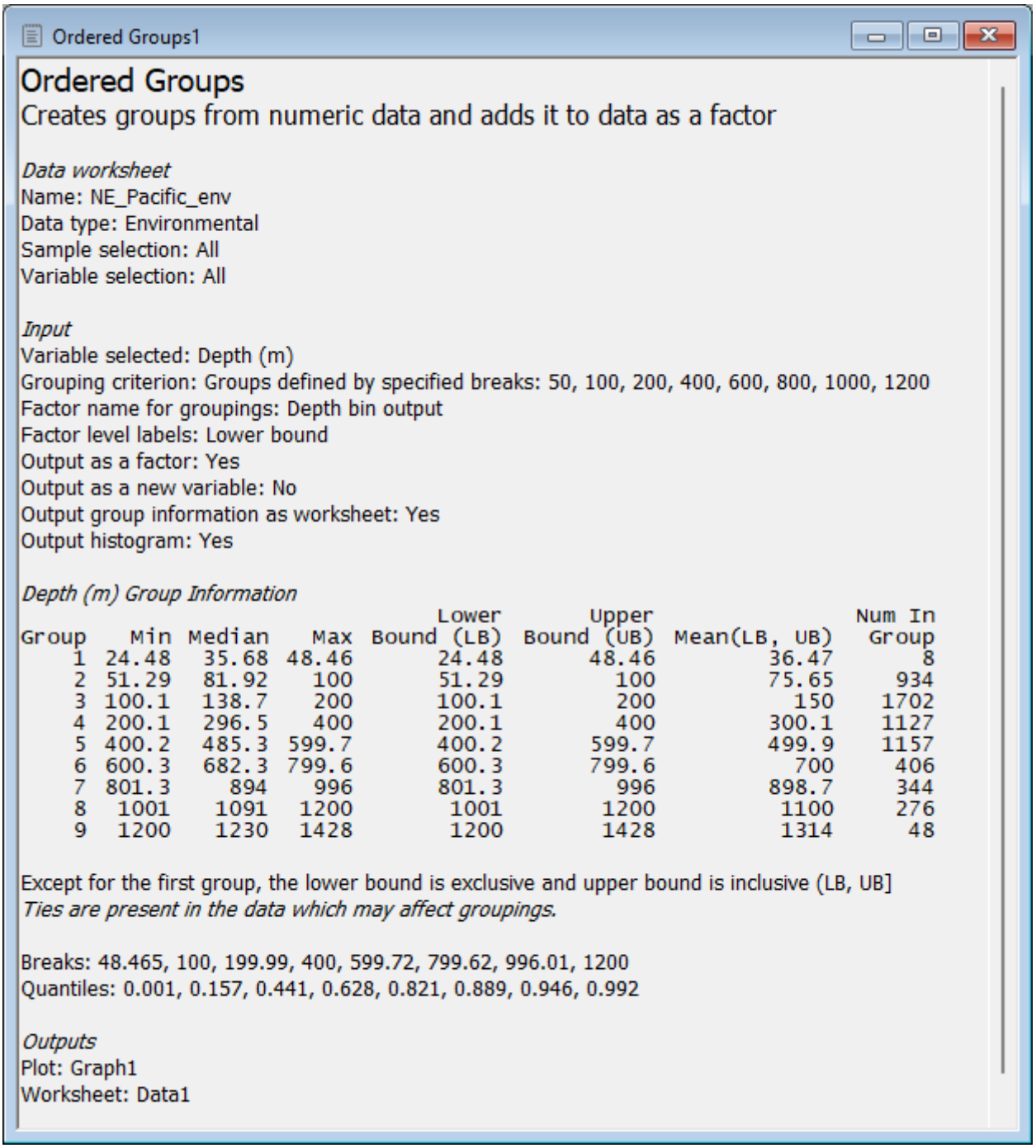
- Variable: Depth (m)
- Grouping criterion > $\bullet$Specify break(s) 50, 100, 200, 400, 600, 800, 1000, 1200
- Output >
  - Factor name: Depth bin output
  - Factor level labels: $\bullet$Lower bound (LB)
  - Output as $\checkmark$Factor
  - Output group information to ($\checkmark$Worksheet) & ($\checkmark$Histogram with group boundaries)

The dialog with these choices will look like this:



Output from the 'Create Ordered Groups...' tool

The output file ( ) called 'Ordered Groups1' in the Explorer tree, has some useful information about how the groups were created. For each group we can see: the minimum, median, maximum, lower bound (LB) upper bound (UB), the mean of (LB) and (UB), and the number of samples that fell into each group. This output file is shown below:



Ordered Groups1

Ordered Groups

Creates groups from numeric data and adds it to data as a factor

Data worksheet
Name: NE_Pacific_env
Data type: Environmental
Sample selection: All
Variable selection: All

Input
Variable selected: Depth (m)
Grouping criterion: Groups defined by specified breaks: 50, 100, 200, 400, 600, 800, 1000, 1200
Factor name for groupings: Depth bin output
Factor level labels: Lower bound
Output as a factor: Yes
Output as a new variable: No
Output group information as worksheet: Yes
Output histogram: Yes

Depth (m) Group Information

Group	Min	Median	Max	Lower Bound (LB)	Upper Bound (UB)	Mean(LB, UB)	Num In Group
1	24.48	35.68	48.46	24.48	48.46	36.47	8
2	51.29	81.92	100	51.29	100	75.65	934
3	100.1	138.7	200	100.1	200	150	1702
4	200.1	296.5	400	200.1	400	300.1	1127
5	400.2	485.3	599.7	400.2	599.7	499.9	1157
6	600.3	682.3	799.6	600.3	799.6	700	406
7	801.3	894	996	801.3	996	898.7	344
8	1001	1091	1200	1001	1200	1100	276
9	1200	1230	1428	1200	1428	1314	48

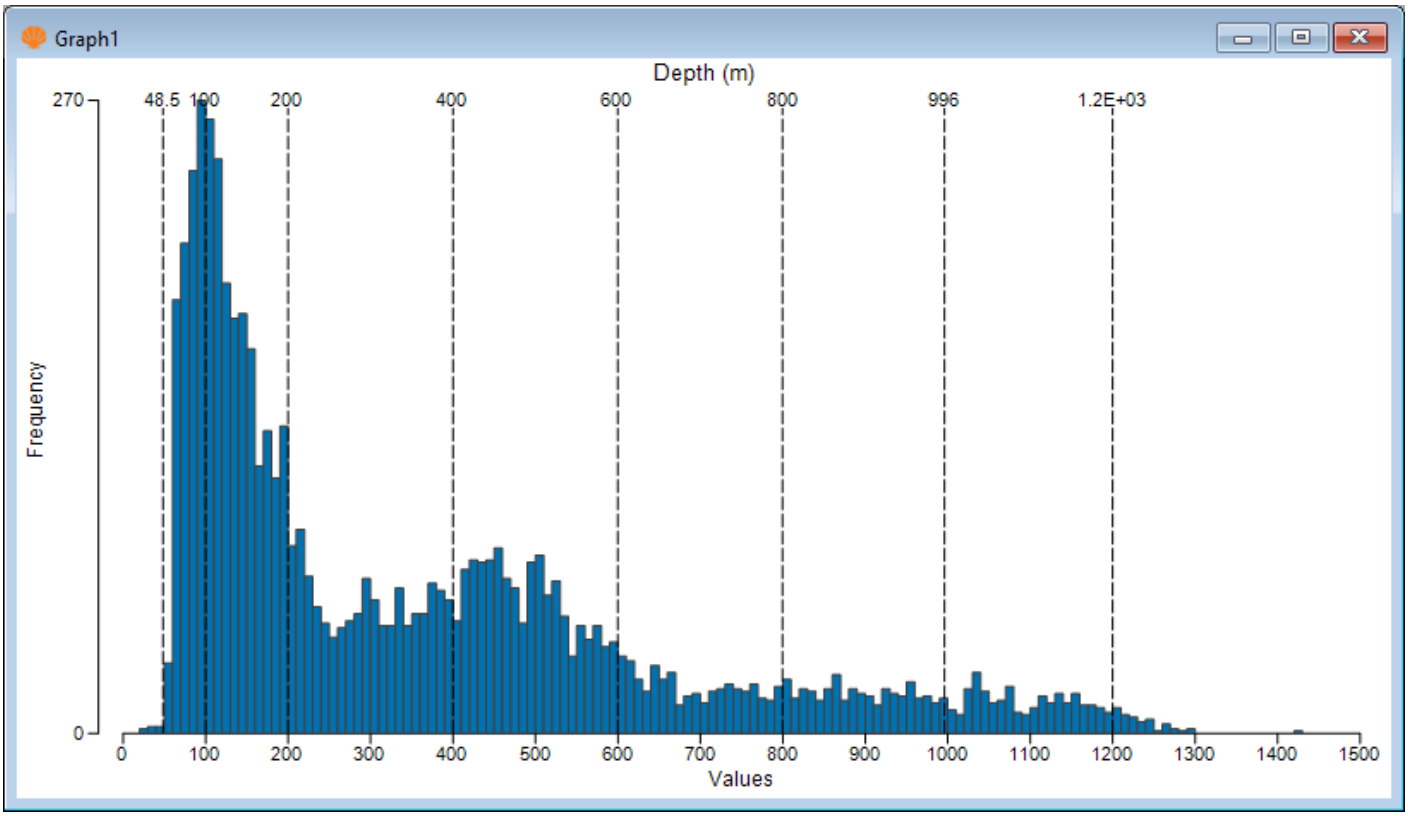
Except for the first group, the lower bound is exclusive and upper bound is inclusive (LB, UB]
Ties are present in the data which may affect groupings.

Breaks: 48.465, 100, 199.99, 400, 599.72, 799.62, 996.01, 1200
Quantiles: 0.001, 0.157, 0.441, 0.628, 0.821, 0.889, 0.946, 0.992

Outputs
Plot: Graph1
Worksheet: Data1

(As an aside, note that we opted in this example also to output this information to a separate data sheet as well, which is called 'Data1' in the Explorer tree.)

Another important part of our output (called 'Graph1' in the Explorer tree) is a *histogram* of the ' Depth (m)' variable, with the values for the breaks separating the groups shown as vertical dotted lines (as shown below). This is a really helpful way to see the groupings we obtained by reference to the full distribution of sample values for the variable of interest.



Another thing we get (and perhaps the most useful 'handle' for subsequent analyses we may want to do) is a new *factor* that is now associated with our original data file. To see this factor, click on the NE_Pacific_env data sheet in the Explorer tree, then click **Edit** > **Factors...**. In the 'Factors' window, you will now see this new factor, called 'Depth bin output', provided as the last column (on the right), like this:

Factors

Edit Fill

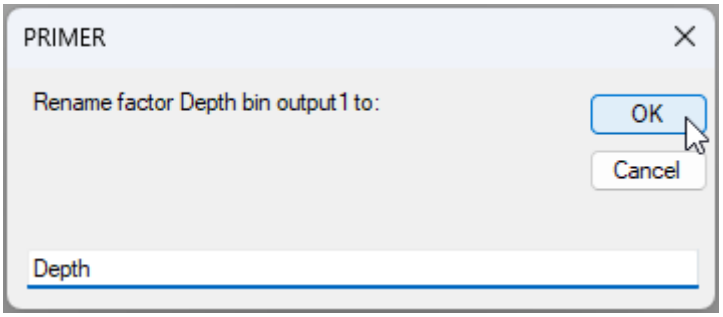
Add... Duplicate Combine... Rename... Rename Levels... Reorder... Delete... Key... Import... OK Cancel

Label	Trawl_id	Project	Year	Depth bin output
1	199901001002	Groundfish Slope Survey	1999	1000.6
2	199901001006	Groundfish Slope Survey	1999	200.13
3	199901001008	Groundfish Slope Survey	1999	200.13
4	199901001010	Groundfish Slope Survey	1999	400.17
5	199901001014	Groundfish Slope Survey	1999	1000.6
6	199901001016	Groundfish Slope Survey	1999	600.3
7	199901001018	Groundfish Slope Survey	1999	200.13
8	199901001019	Groundfish Slope Survey	1999	200.13
9	199901001020	Groundfish Slope Survey	1999	400.17
10	199901001021	Groundfish Slope Survey	1999	600.3
11	199901001024	Groundfish Slope Survey	1999	1000.6
12	199901001026	Groundfish Slope Survey	1999	600.3
13	199901001027	Groundfish Slope Survey	1999	400.17
14	199901001028	Groundfish Slope Survey	1999	200.13
15	199901001030	Groundfish Slope Survey	1999	600.3
16	199901001032	Groundfish Slope Survey	1999	200.13
17	199901001034	Groundfish Slope Survey	1999	200.13
18	199901001035	Groundfish Slope Survey	1999	200.13

Tweak the factor level names (optional)

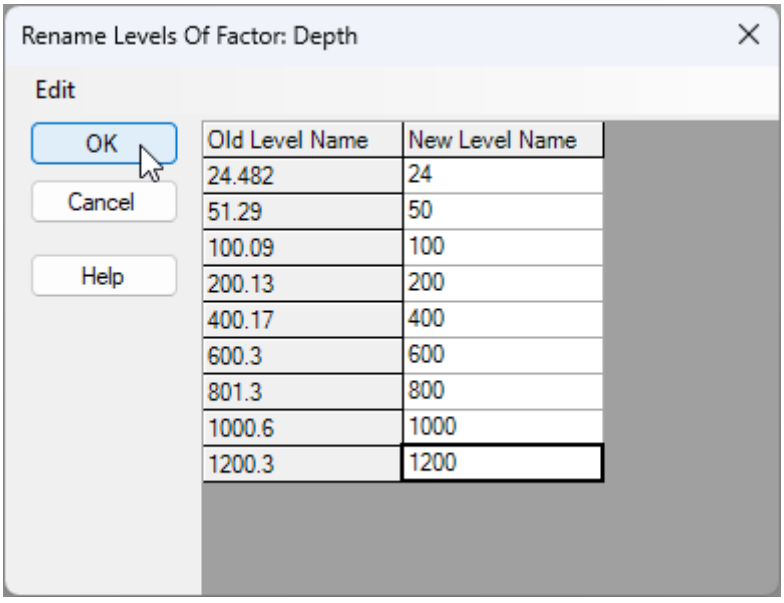
We asked for the lower bound value of each group to be made the factor level names of our groups, but we can see that these are not necessarily whole numbers. It might be nice to round these to the appropriate whole number, in each case.

4. First, let's duplicate the newly created factor. From the NE_Pacific_env data sheet, click **Edit** > **Factors...**. In the 'Factors' window, click anywhere in the column named 'Depth bin output', then click the 'Duplicate' button (Duplicate).
5. You will get a new column with a duplicate factor called 'Depth bin output1'. Click anywhere in this new column, then click the 'Rename...' button (Rename...). Rename this factor simply as 'Depth', then click '**OK**'.



6. Now we are going to rename the levels of this factor called 'Depth' to whole numbers. Click the 'Rename Levels...' button (Rename Levels...).

In the 'Rename Factor Levels' dialog (new to PRIMER8!), you will see two columns: the 'Existing Level Name' on the left and the 'New Level Name' on the right. Change the values in the 'New Level Name' to the desired values , and click'**OK**', as shown below.



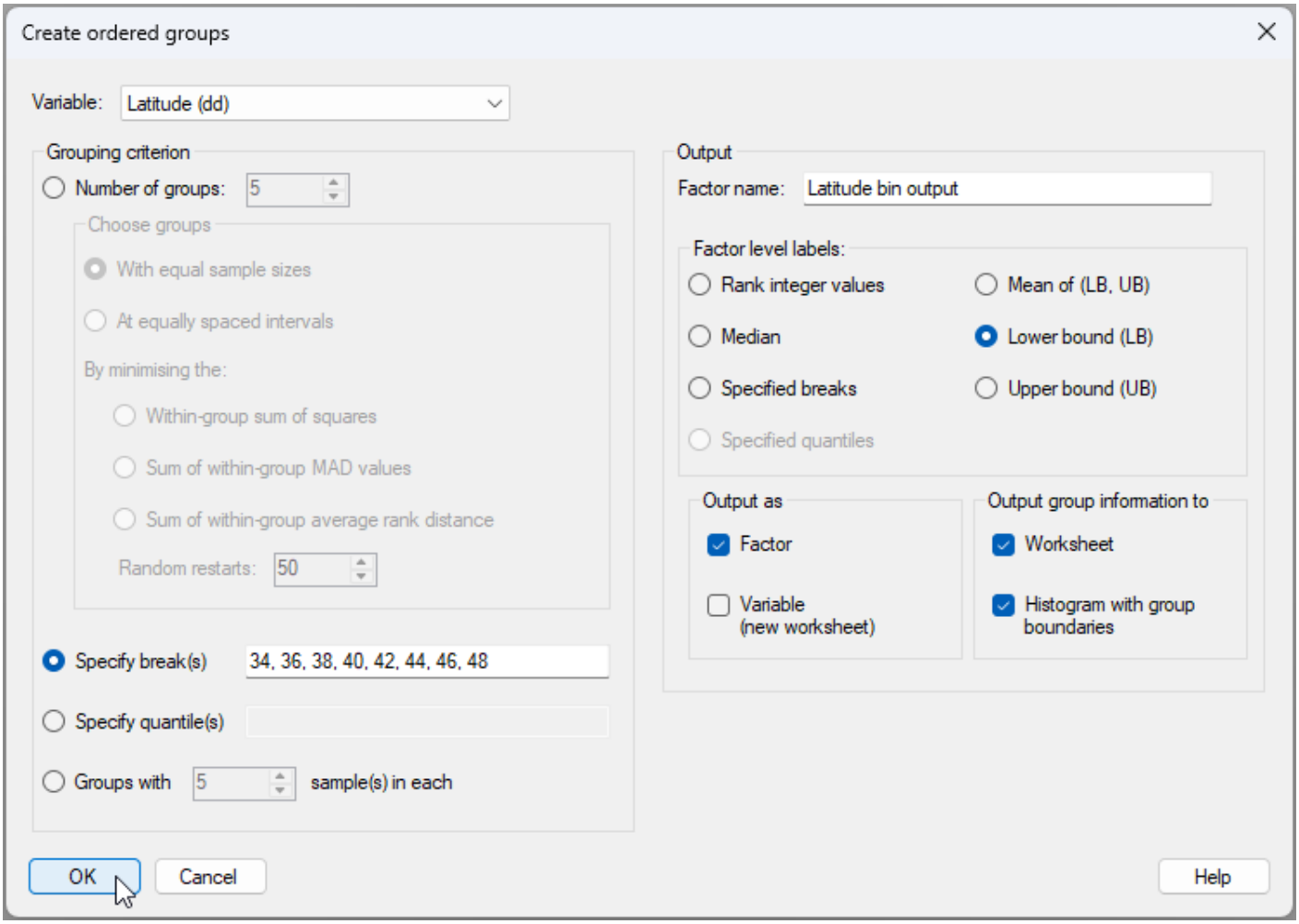
In our case, we have specified new factor level names that are still *numeric* and consist of whole numbers that will make sense for us in this example for plots/symbols, etc. But note that this tool can be used to re-define the factor level names to anything we wish (not necessarily numbers). Note also, for this example, that we have to be careful not to do too much 'rounding'. We need to

stay true to what we know about the data and the bounds of the groups we have created. Bear in mind that we could have changed the names of these factor levels to ranges of depths (e.g., such as 50-100m), which may be better (or more accurate). However, if we change these names to ranges in this way, then we no longer have a strictly numeric factor. Factor levels that are numeric can be really useful in PRIMER, because they allow us to do things like treat the factor as *ordered* in an ANOSIM, or superimpose *trajectories* to connect consecutive depths on an ordination plot, etc. In this example, given the new names we have chosen, whenever we describe this factor we will have to be clear what the labels mean; specifically, that the group that we have named '50' here corresponds to samples that occurred between 50 m and 100 m in depth, and so on.

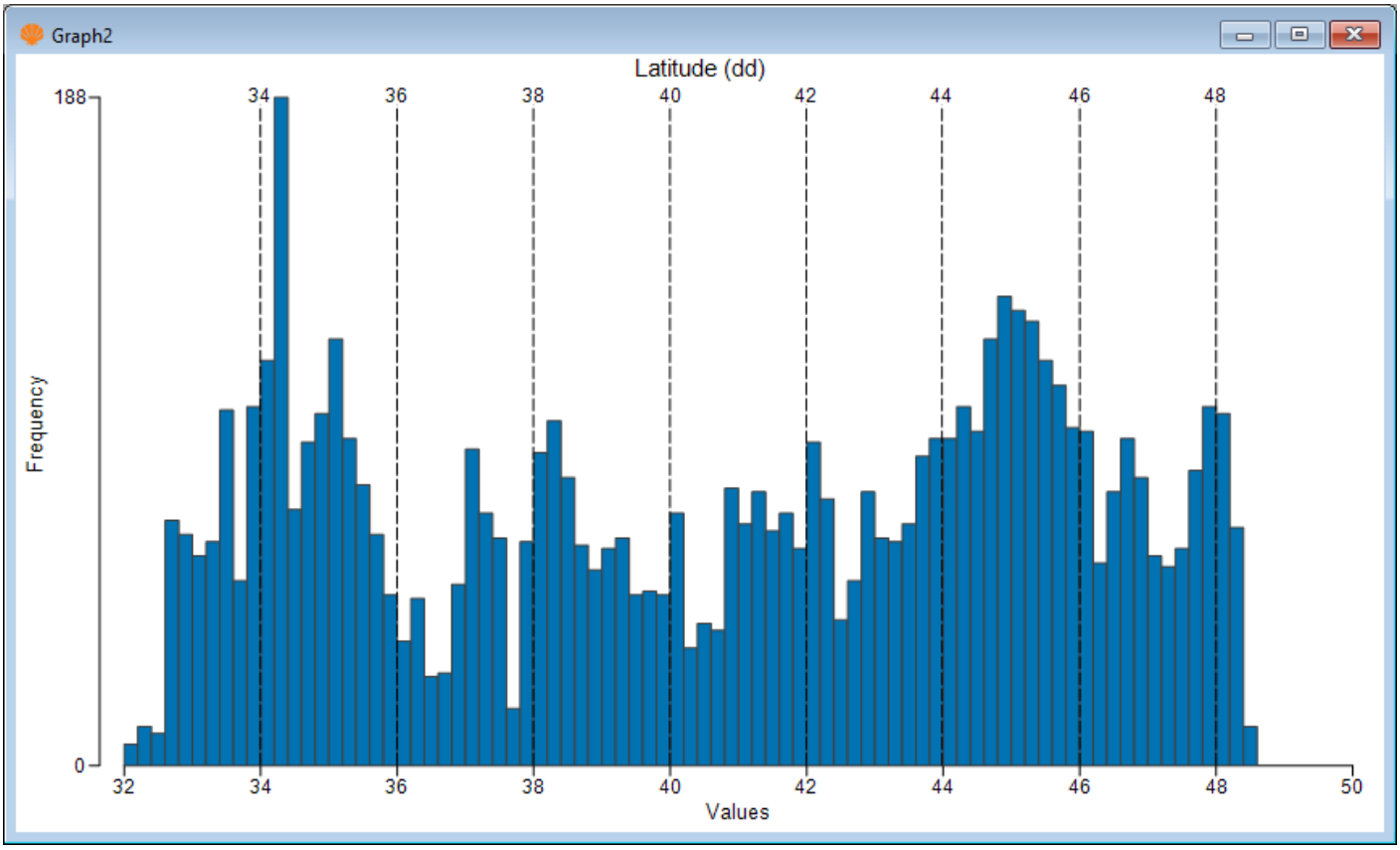
Create a factor for Latitude

We can repeat the above steps for another important spatial factor here: namely, latitude.

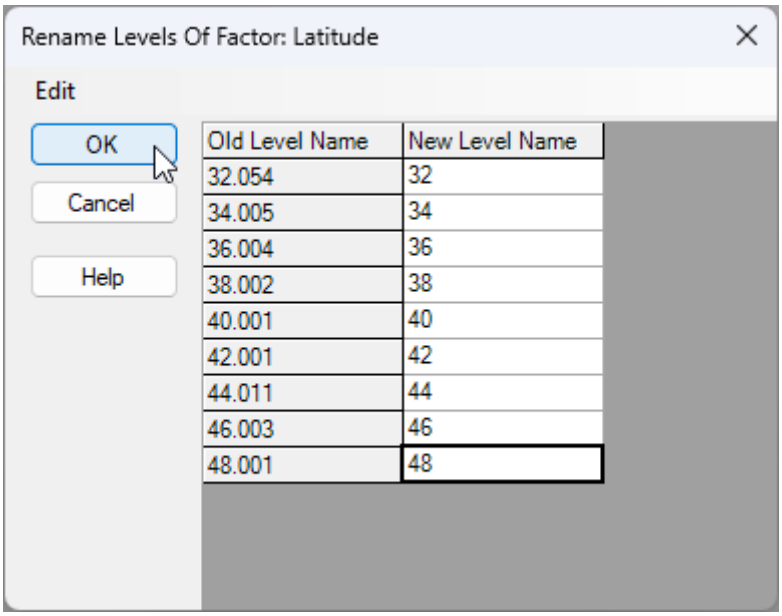
7. From the 'NE_Pacific_env' data sheet in the Explorer tree, click **Tools** > **Create Ordered Groups...**, then choose to create ordered groups from the variable of 'Latitude (dd)', and specify the breaks to occur in 2-degree increments: {34, 36, 38, 40, 42, 44, 46, 48}, as shown below.



A histogram showing the break-points for latitude that we have chosen is shown below ('Graph2' in the Explorer tree).



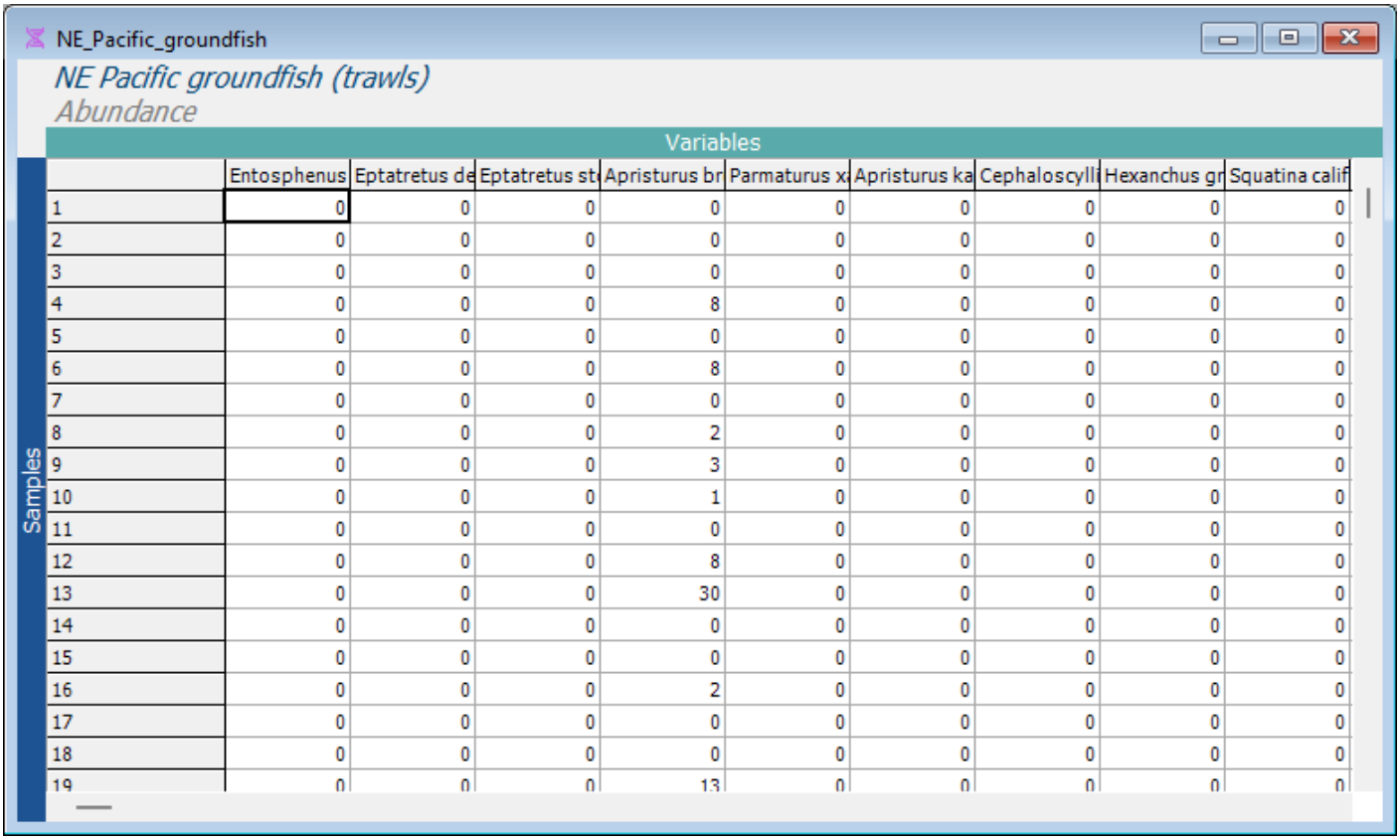
8. You will want to tweak this new factor for Latitude (just as we did for the Depth factor before). From the NE_Pacific_env data sheet, click **Edit** > **Factors...**, then proceed to do the following:
   - *Duplicate the factor* of 'Latitude bin ouput' to get 'Latitude bin ouput1' (Duplicate).
   - *Rename the factor* 'Latitude bin ouput1' to call it 'Latitude' (Rename...).
   - *Rename the levels of the factor* 'Latitude' to whole numbers (Rename Levels...). An image of this last operation is shown below.



Open the groundfish data and import the new factors

Now that we have the factors we want, it would be great to use this to our advantage in analyses of the groundfish data. First we will get the fish data into the workspace, then we will import the new factors (currently associated wtih the environmental data sheet) over to the groundfish data sheet.

9. In the same PRIMER workspace, click **File** > **Open...** and open the file called ' NE_Pacific_groundfish.pri'. It will look like this:


NE_Pacific_groundfish

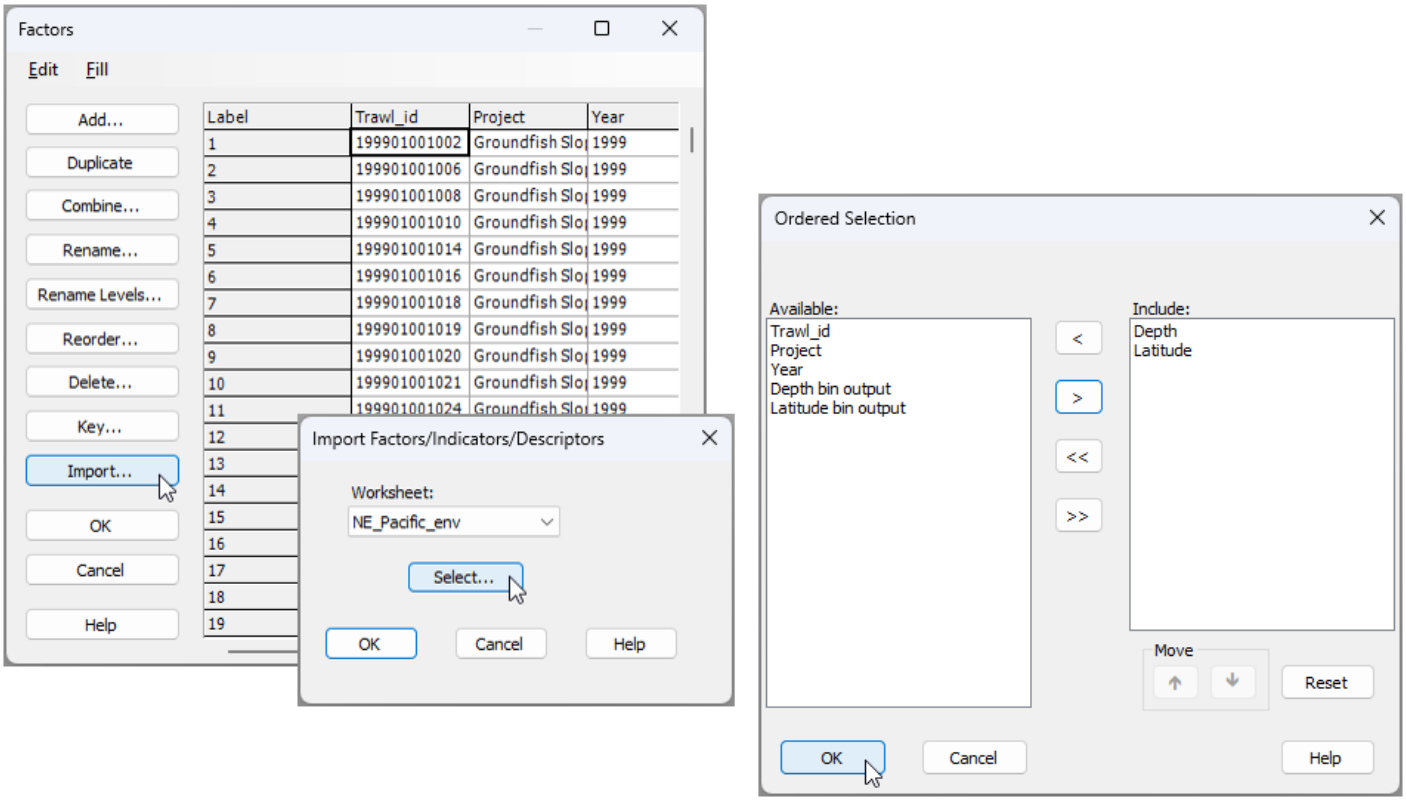
NE Pacific groundfish (trawls)

Abundance

Variables

Samples	Entosphenus	Eptatretus de	Eptatretus st	Apristurus br	Parmaturus x	Apristurus ka	Cephaloscylli	Hexanchus gr	Squatina calif
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	8	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0	8	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0
8	0	0	0	2	0	0	0	0	0
9	0	0	0	3	0	0	0	0	0
10	0	0	0	1	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0
12	0	0	0	8	0	0	0	0	0
13	0	0	0	30	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0
16	0	0	0	2	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0
19	0	0	0	13	0	0	0	0	0

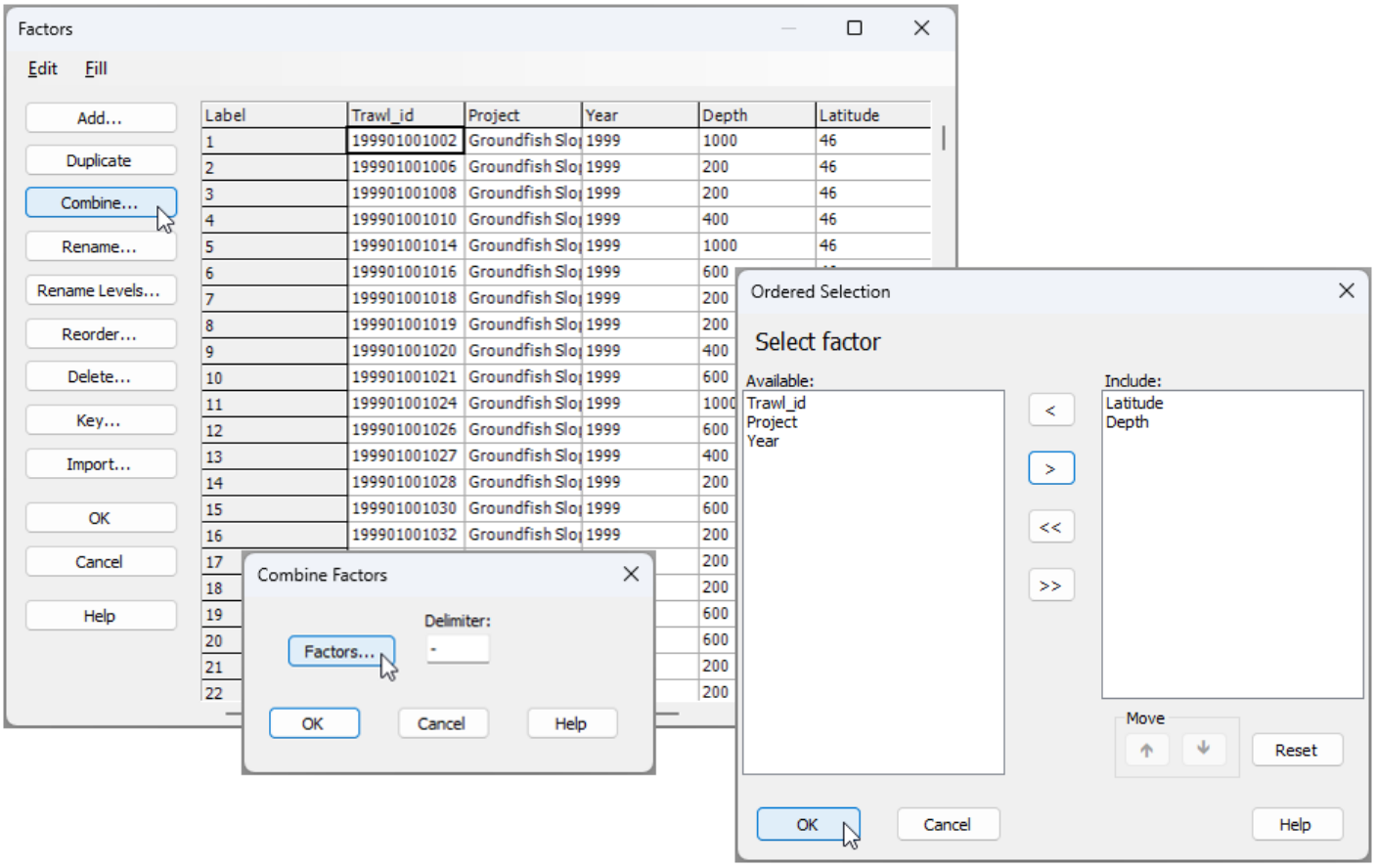
10. From the 'NE_Pacific_groundfish' data sheet in the Explorer tree, click **Edit** > **Factors...** and click the 'Import' button (Import...). Choose to import from the 'NE_Pacific_env' worksheet, then click the 'Select' button (Select...). In the selection dialog, pick only the two factors named 'Depth' and 'Latitude' to include in the import, then click '**OK**', like so:



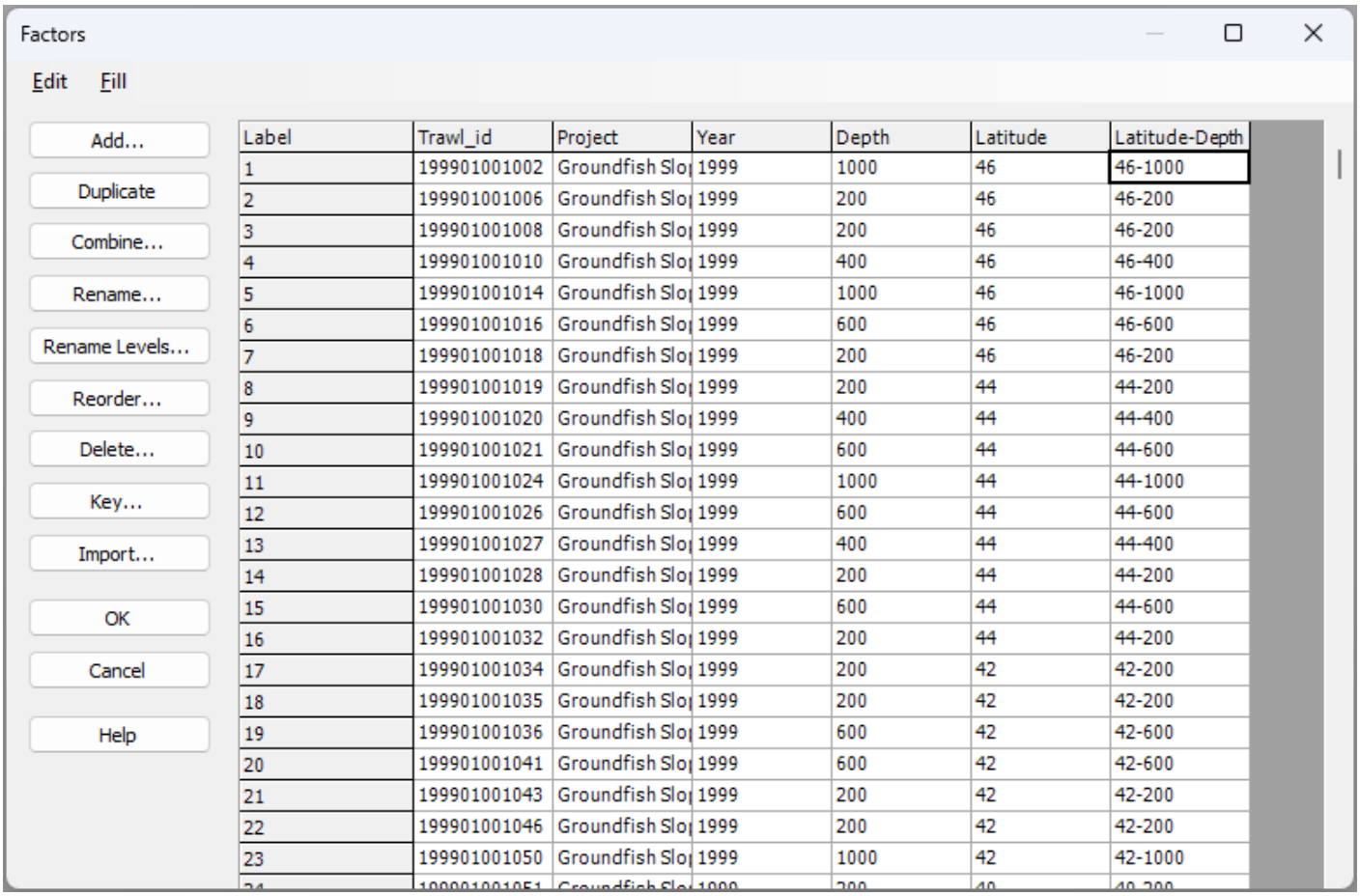
You can confirm that the groundfish sheet now has the 'Depth' and 'Latitude' factors, imported from the environmental data sheet.

Create a combined factor of depth-by-latitude

11. For our analysis and plots, we will want now to create a factor that corresponds to the *combination* of all depth-by-latitude bins. From the 'NE_Pacific_groundfish', click **Edit** > **Factors...**, then click the 'Combine' button (Combine...). In the 'Combine Factors' dialog, click on the 'Factors...' button (Factors...), then in the 'Ordered Selection' dialog, choose to include just Latitude and Depth, as shown below, then click '**OK**' (3 separate times for the three windows).



We now have a factor that identifies groups of samples with similar latitude and depth. This new combined factor of 'Latitude-Depth' effectively corresponds to spatial 'cells' of practical interest in our study design, and it will serve us very well for subsequent analyses. (It is the final column on the right in the image below):



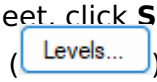
Label	Trawl_id	Project	Year	Depth	Latitude	Latitude-Depth
1	199901001002	Groundfish Slo	1999	1000	46	46-1000
2	199901001006	Groundfish Slo	1999	200	46	46-200
3	199901001008	Groundfish Slo	1999	200	46	46-200
4	199901001010	Groundfish Slo	1999	400	46	46-400
5	199901001014	Groundfish Slo	1999	1000	46	46-1000
6	199901001016	Groundfish Slo	1999	600	46	46-600
7	199901001018	Groundfish Slo	1999	200	46	46-200
8	199901001019	Groundfish Slo	1999	200	44	44-200
9	199901001020	Groundfish Slo	1999	400	44	44-400
10	199901001021	Groundfish Slo	1999	600	44	44-600
11	199901001024	Groundfish Slo	1999	1000	44	44-1000
12	199901001026	Groundfish Slo	1999	600	44	44-600
13	199901001027	Groundfish Slo	1999	400	44	44-400
14	199901001028	Groundfish Slo	1999	200	44	44-200
15	199901001030	Groundfish Slo	1999	600	44	44-600
16	199901001032	Groundfish Slo	1999	200	44	44-200
17	199901001034	Groundfish Slo	1999	200	42	42-200
18	199901001035	Groundfish Slo	1999	200	42	42-200
19	199901001036	Groundfish Slo	1999	600	42	42-600
20	199901001041	Groundfish Slo	1999	600	42	42-600
21	199901001043	Groundfish Slo	1999	200	42	42-200
22	199901001046	Groundfish Slo	1999	200	42	42-200
23	199901001050	Groundfish Slo	1999	1000	42	42-1000

Analyse changes in groundfish assemblages *vs* depth and latitude

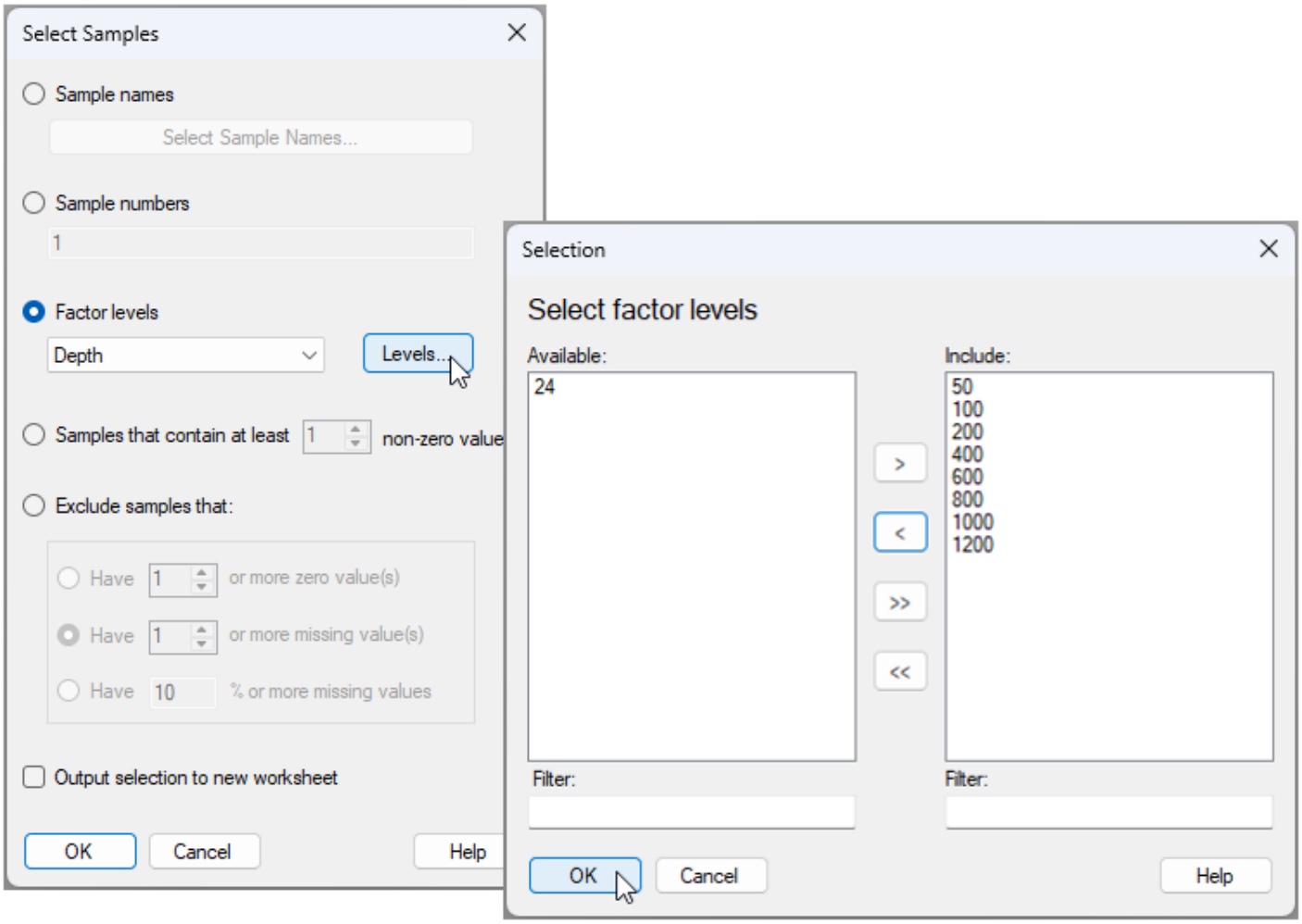
We now wish to analyse potential changes in groundfish assemblages with shifts in depth and latitude. Our plan will be to apply a suitable pre-treatment to the data, calculate averages wtihin each latitude-by-depth cell, proceed with calculating a square-root transformation of the data and Bray-Curtis resemblances among these cells, followed by an ordination (nMDS) and tests of hypotheses (ANOSIM) on that resemblance matrix. Note that the averaging step is really important here. We would have no hope of seeing any sensible patterns if we were to 'throw' the full dataset of over 6000 replicates into a single nMDS plot!

Pre-treatment

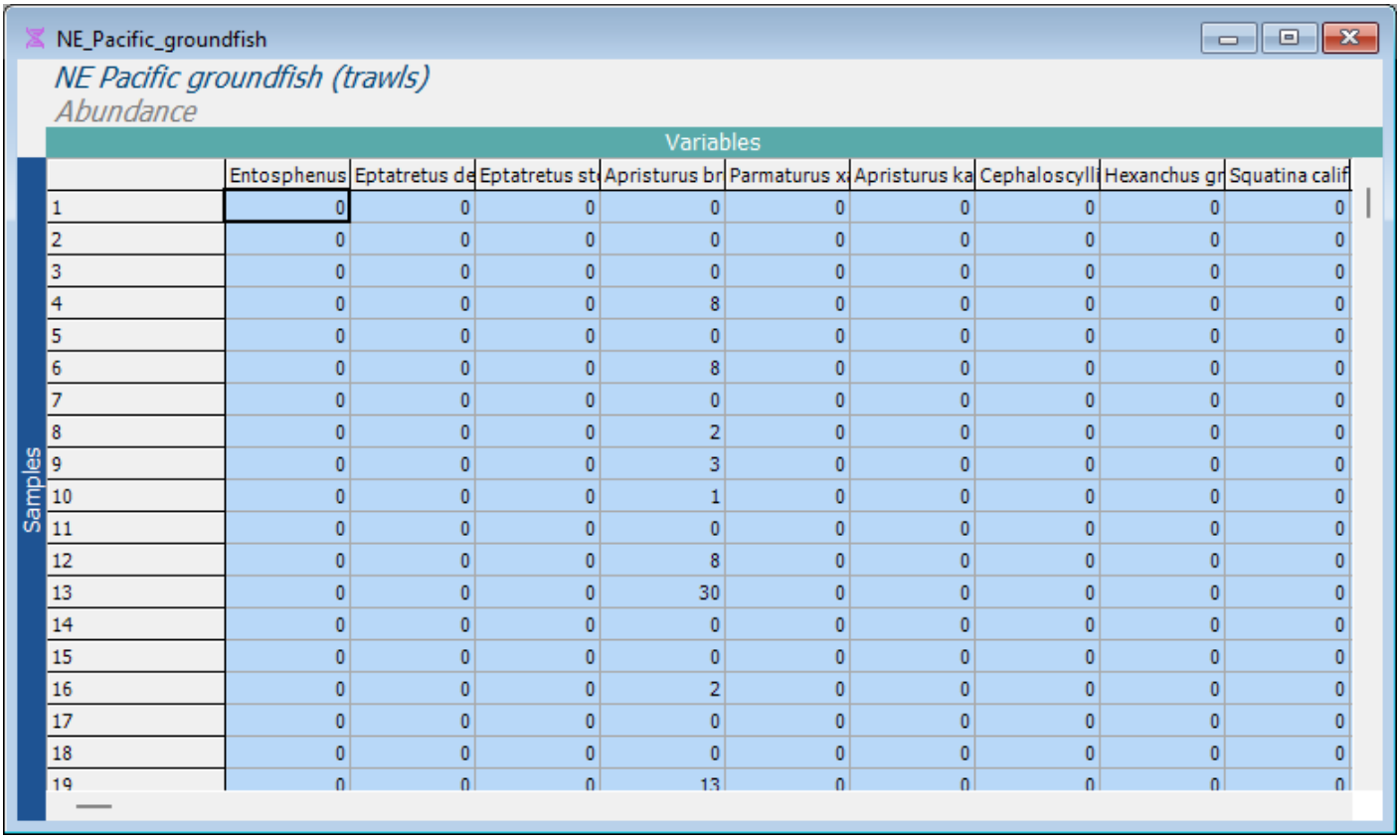
To analyse the groundfish data, it is appropriate to consider that many fish species occur in clusters or aggregations of individuals. Hence, a pre-treatment option such as *dispersion weighting* ({{2954#bkmrk-clarkeetal2006a}}) would likely be a really appropriate option here. We shall consider the replicate trawls within each latitude-by-depth cell as fairly natural groupings to use in order to apply this pre-treatment option. We noted that there were only 8 replicate trawls from depths less than 50 m, so we will omit those replicates in what follows.

12. From the 'NE_Pacific_groundfish' data sheet, click **Select** > **Samples...** > ($\bullet$Factor levels > Depth), click the 'Levels' button (

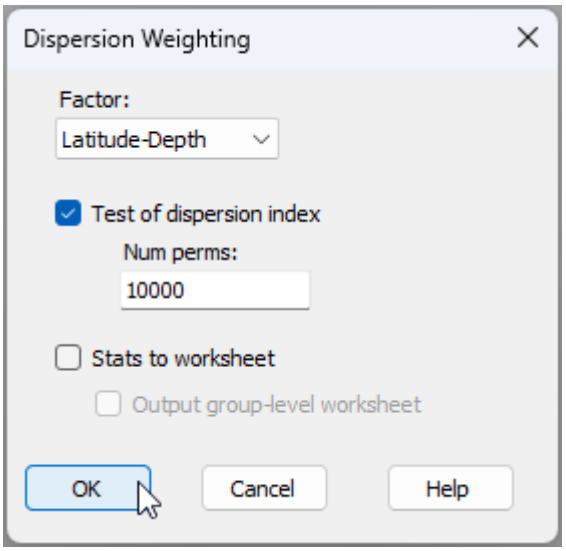
), then choose to retain all depth groups except '24', and click '**OK**', as shown below:



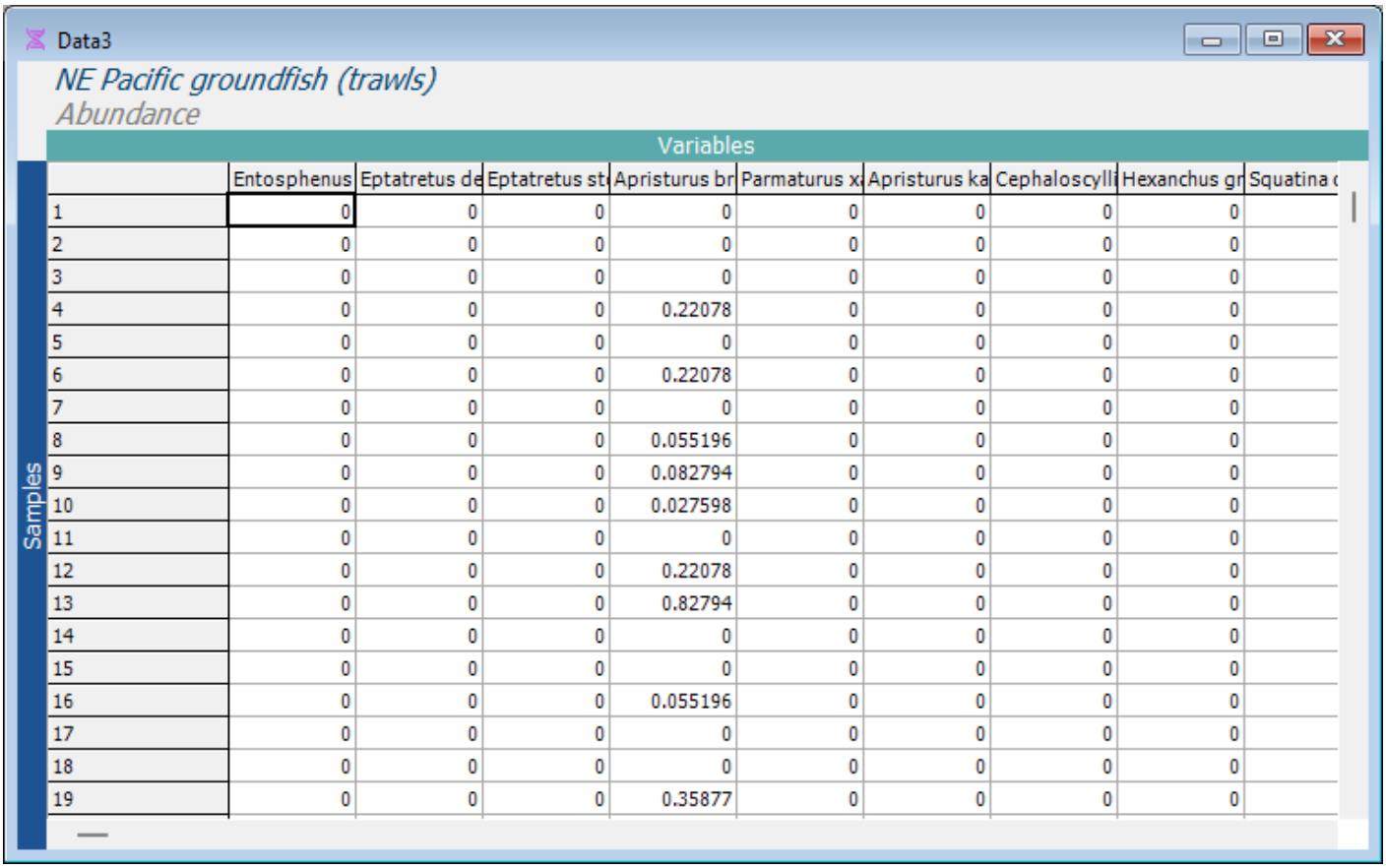
This will turn the worksheet cells blue, and this indicates that a subset of the data has been selected.



13. From this subset-selected groundfish data sheet, click **Pre-treatment** > **Dispersion Weighting...** and choose to do this on the basis of the factor 'Latitude-Depth', then click '**OK**', like so:



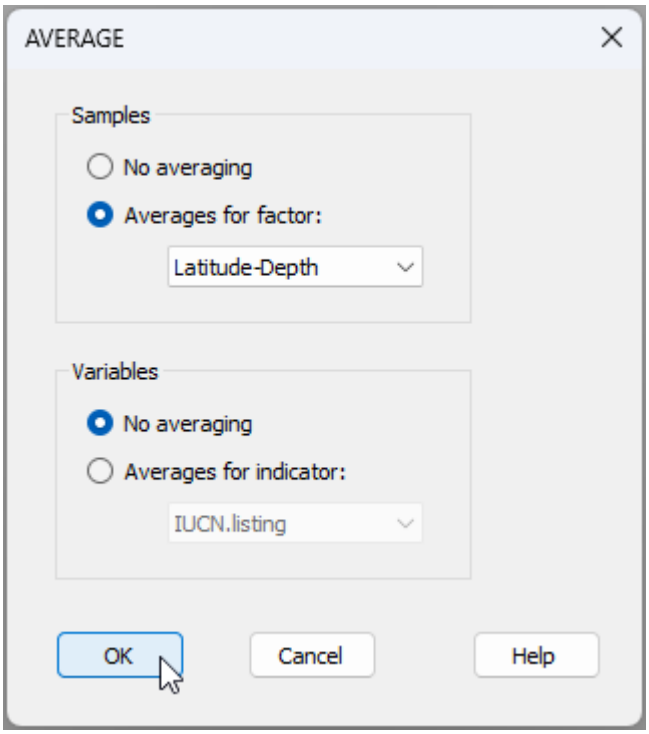
This operation will take some considerable time, simply because of the sheer size of the dataset. The randomization test (for which the individuals of each species are randomly re-assigned to replicates wtihin each of the latitude-depth cells), which is done indpendently for every species, is computationally demanding. However, the results file from the dispersion-weighting pre-treatment (called 'Dispersion weighting1') demonstrates very clearly that *many* of these fish species show significant clustering (i.e., wherever the value in the 'Divisor' column is greater than 1), hence should sensibly be pre-treated in this way. The dispersion-weighted data is called 'Data3' in the Explorer tree and will look like this:



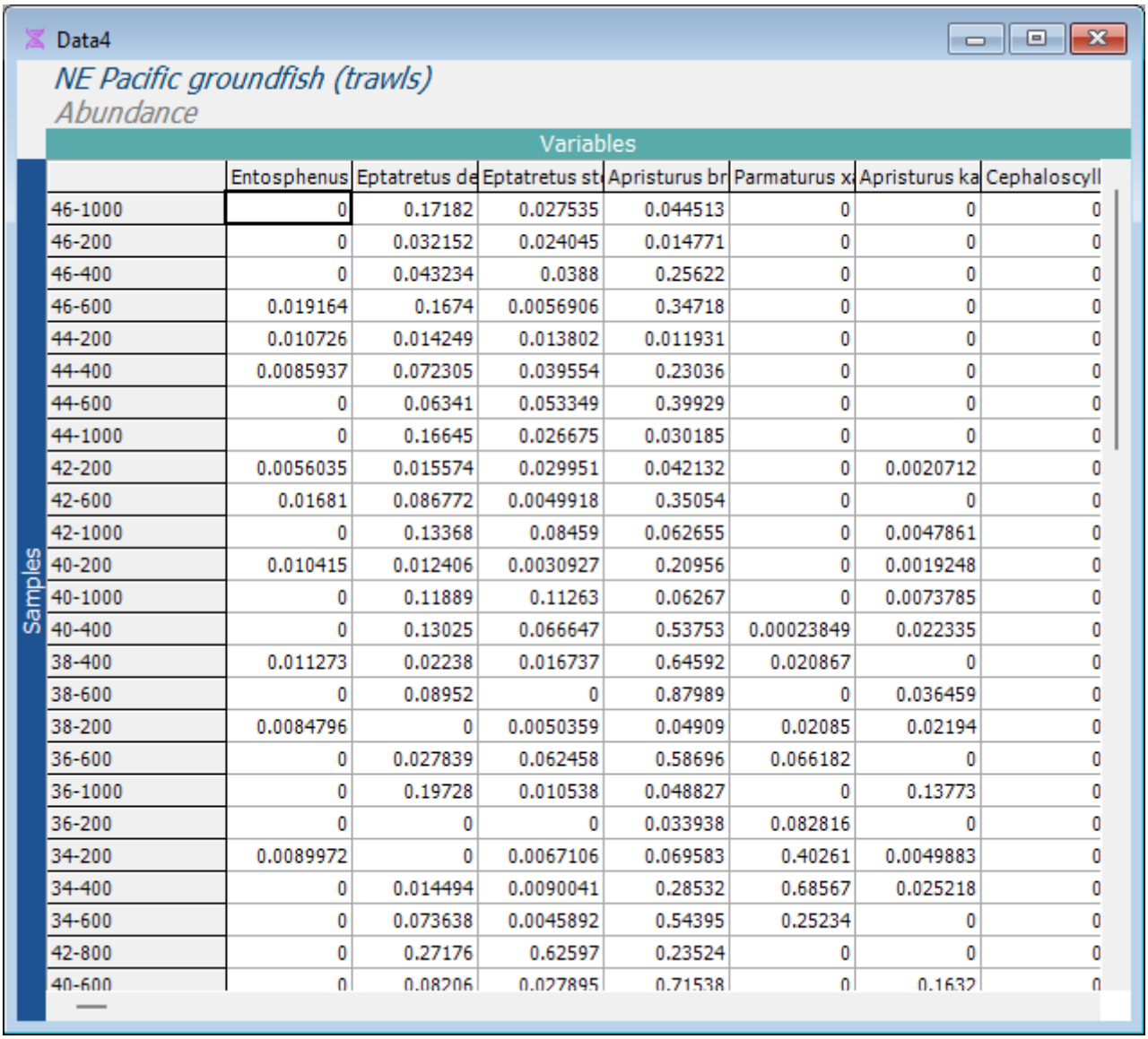
	Entosphenus	Eptatretus de	Eptatretus st	Apristurus br	Parmaturus x	Apristurus ka	Cephaloscylli	Hexanchus gr	Squatina
1	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	
4	0	0	0	0.22078	0	0	0	0	
5	0	0	0	0	0	0	0	0	
6	0	0	0	0.22078	0	0	0	0	
7	0	0	0	0	0	0	0	0	
8	0	0	0	0.055196	0	0	0	0	
9	0	0	0	0.082794	0	0	0	0	
10	0	0	0	0.027598	0	0	0	0	
11	0	0	0	0	0	0	0	0	
12	0	0	0	0.22078	0	0	0	0	
13	0	0	0	0.82794	0	0	0	0	
14	0	0	0	0	0	0	0	0	
15	0	0	0	0	0	0	0	0	
16	0	0	0	0.055196	0	0	0	0	
17	0	0	0	0	0	0	0	0	
18	0	0	0	0	0	0	0	0	
19	0	0	0	0.35877	0	0	0	0	

Averaging

14. From the dispersion-weighted data ('Data3'), click **Tools** > **Average...** and choose to average the samples by the factor of 'Latitude-Depth', like so:



In the resulting data sheet (called 'Data4'), we now have average values for each species in each latitude-by-depth cell (rows), as shown below:


Data4

NE Pacific groundfish (trawls)

Abundance

Variables

Samples

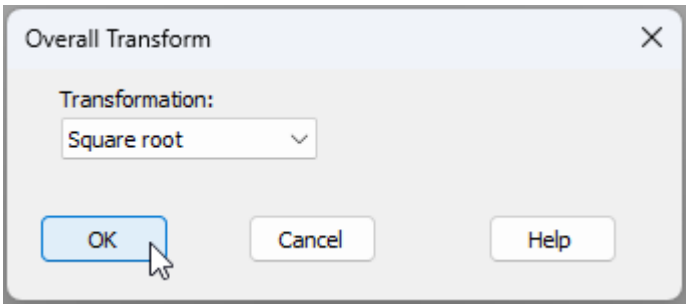
	Entosphenus	Eptatretus de	Eptatretus st	Apristurus br	Parmaturus x	Apristurus ka	Cephaloscyll
46-1000	0	0.17182	0.027535	0.044513	0	0	0
46-200	0	0.032152	0.024045	0.014771	0	0	0
46-400	0	0.043234	0.0388	0.25622	0	0	0
46-600	0.019164	0.1674	0.0056906	0.34718	0	0	0
44-200	0.010726	0.014249	0.013802	0.011931	0	0	0
44-400	0.0085937	0.072305	0.039554	0.23036	0	0	0
44-600	0	0.06341	0.053349	0.39929	0	0	0
44-1000	0	0.16645	0.026675	0.030185	0	0	0
42-200	0.0056035	0.015574	0.029951	0.042132	0	0.0020712	0
42-600	0.01681	0.086772	0.0049918	0.35054	0	0	0
42-1000	0	0.13368	0.08459	0.062655	0	0.0047861	0
40-200	0.010415	0.012406	0.0030927	0.20956	0	0.0019248	0
40-1000	0	0.11889	0.11263	0.06267	0	0.0073785	0
40-400	0	0.13025	0.066647	0.53753	0.00023849	0.022335	0
38-400	0.011273	0.02238	0.016737	0.64592	0.020867	0	0
38-600	0	0.08952	0	0.87989	0	0.036459	0
38-200	0.0084796	0	0.0050359	0.04909	0.02085	0.02194	0
36-600	0	0.027839	0.062458	0.58696	0.066182	0	0
36-1000	0	0.19728	0.010538	0.048827	0	0.13773	0
36-200	0	0	0	0.033938	0.082816	0	0
34-200	0.0089972	0	0.0067106	0.069583	0.40261	0.0049883	0
34-400	0	0.014494	0.0090041	0.28532	0.68567	0.025218	0
34-600	0	0.073638	0.0045892	0.54395	0.25234	0	0
42-800	0	0.27176	0.62597	0.23524	0	0	0
40-600	0	0.08206	0.027895	0.71538	0	0.1632	0

Note that the names of the samples in the averaged data now correspond to the latitude-by-depth bin combinations. If you click **Edit** > **Properties...**, you will see that this sheet now has 71 rows. We have consolidated these data in a very useful way across our study design, while maintaining the integrity of the underlying information.

Transformation & resemblance

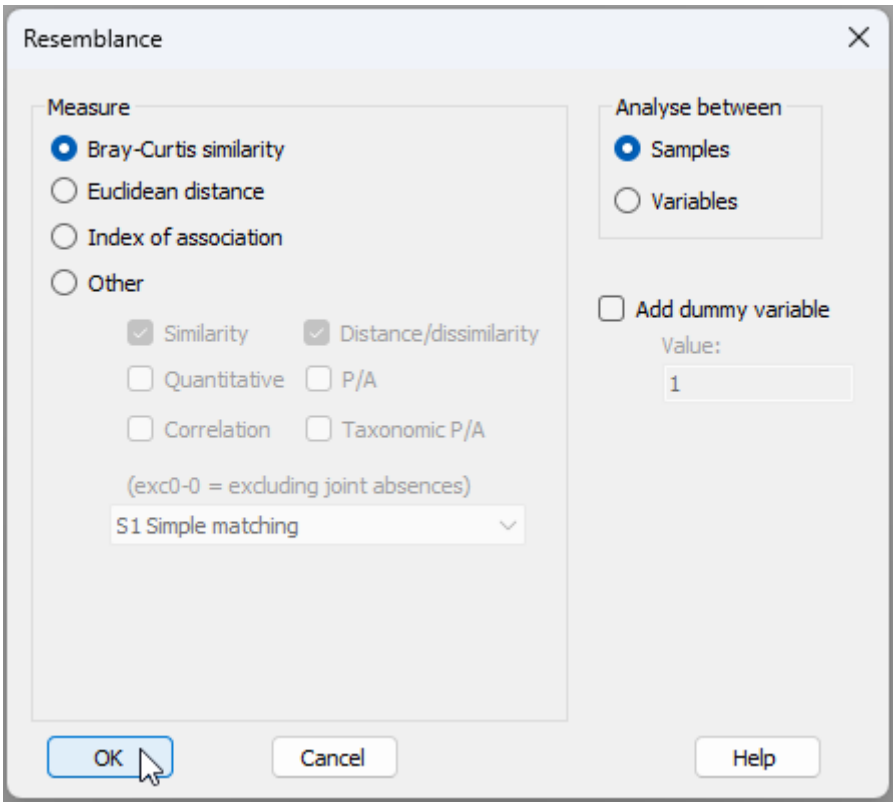
If we look at a shade plot of the averaged data (you can do this by clicking on **Plots** > **Shade Plot...** from 'Data4'), we can see that, even after dispersion-weighting and averaging, these data still look pretty sparse. We will therefore do a (mild) overall transformation to square roots, then calculate the Bray-Curtis resemblance measure.

15. From 'Data4', click **Pre-treatment** > **Transform(overall)...** > Transformation: Square root, **OK**.

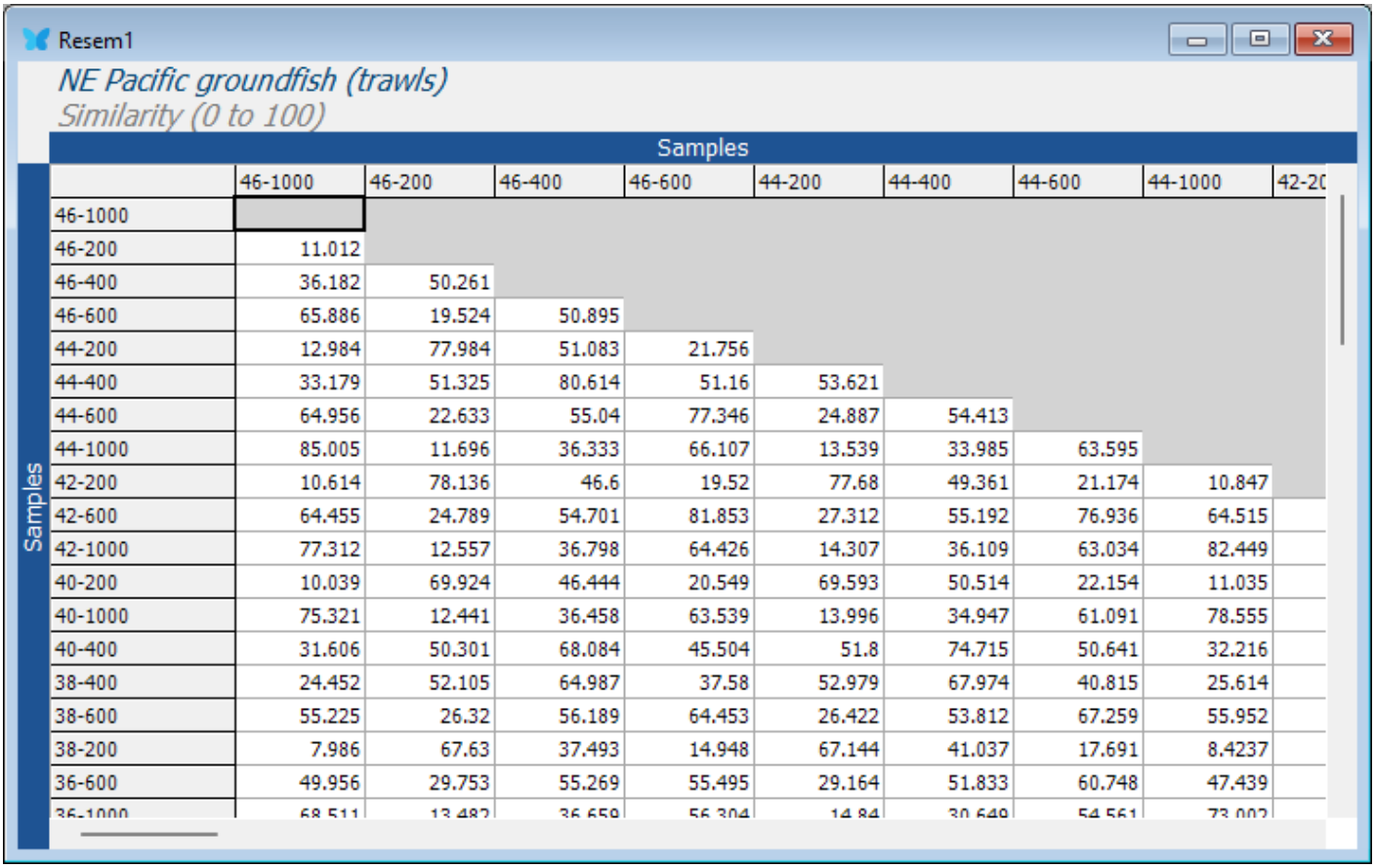


The resulting square-root transformed data matrx will be called 'Data5'.

16. From 'Data5', click **Analyse** > **Resemblance...** > (Measure $\bullet$ Bray-Curtis similarity) & (Analyse between $\bullet$Samples), '**OK**'.



The resulting resemblance matrix will be called 'Resem1', and will look like this:



	46-1000	46-200	46-400	46-600	44-200	44-400	44-600	44-1000	42-20
46-1000									
46-200	11.012								
46-400	36.182	50.261							
46-600	65.886	19.524	50.895						
44-200	12.984	77.984	51.083	21.756					
44-400	33.179	51.325	80.614	51.16	53.621				
44-600	64.956	22.633	55.04	77.346	24.887	54.413			
44-1000	85.005	11.696	36.333	66.107	13.539	33.985	63.595		
42-200	10.614	78.136	46.6	19.52	77.68	49.361	21.174	10.847	
42-600	64.455	24.789	54.701	81.853	27.312	55.192	76.936	64.515	
42-1000	77.312	12.557	36.798	64.426	14.307	36.109	63.034	82.449	
40-200	10.039	69.924	46.444	20.549	69.593	50.514	22.154	11.035	
40-1000	75.321	12.441	36.458	63.539	13.996	34.947	61.091	78.555	
40-400	31.606	50.301	68.084	45.504	51.8	74.715	50.641	32.216	
38-400	24.452	52.105	64.987	37.58	52.979	67.974	40.815	25.614	
38-600	55.225	26.32	56.189	64.453	26.422	53.812	67.259	55.952	
38-200	7.986	67.63	37.493	14.948	67.144	41.037	17.691	8.4237	
36-600	49.956	29.753	55.269	55.495	29.164	51.833	60.748	47.439	
36-1000	68.511	13.482	36.659	56.304	14.84	30.649	54.561	73.002	

Ordination *via* nMDS

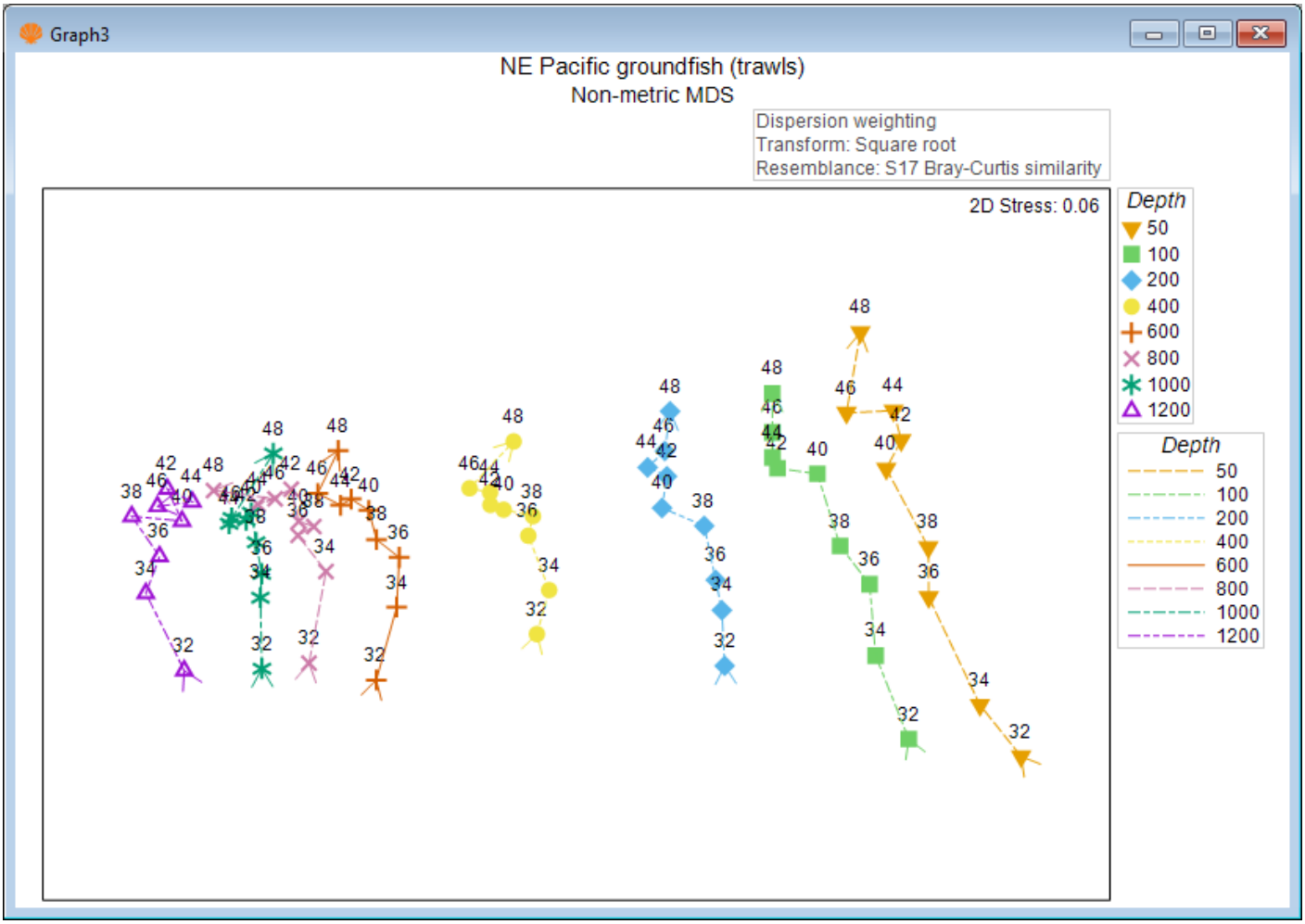
17. From the 'Resem1' matrix, click **Analyse** > **MDS** > **Non-metric MDS (nMDS)...**, retain all of the defaults and click '**OK**'. The resulting best 2d solution for the nMDS ordination plot is called 'Graph3' in the Explorer tree, and it has a nice low stress of 0.06.

We can consider looking at two different 'views' of this ordination:

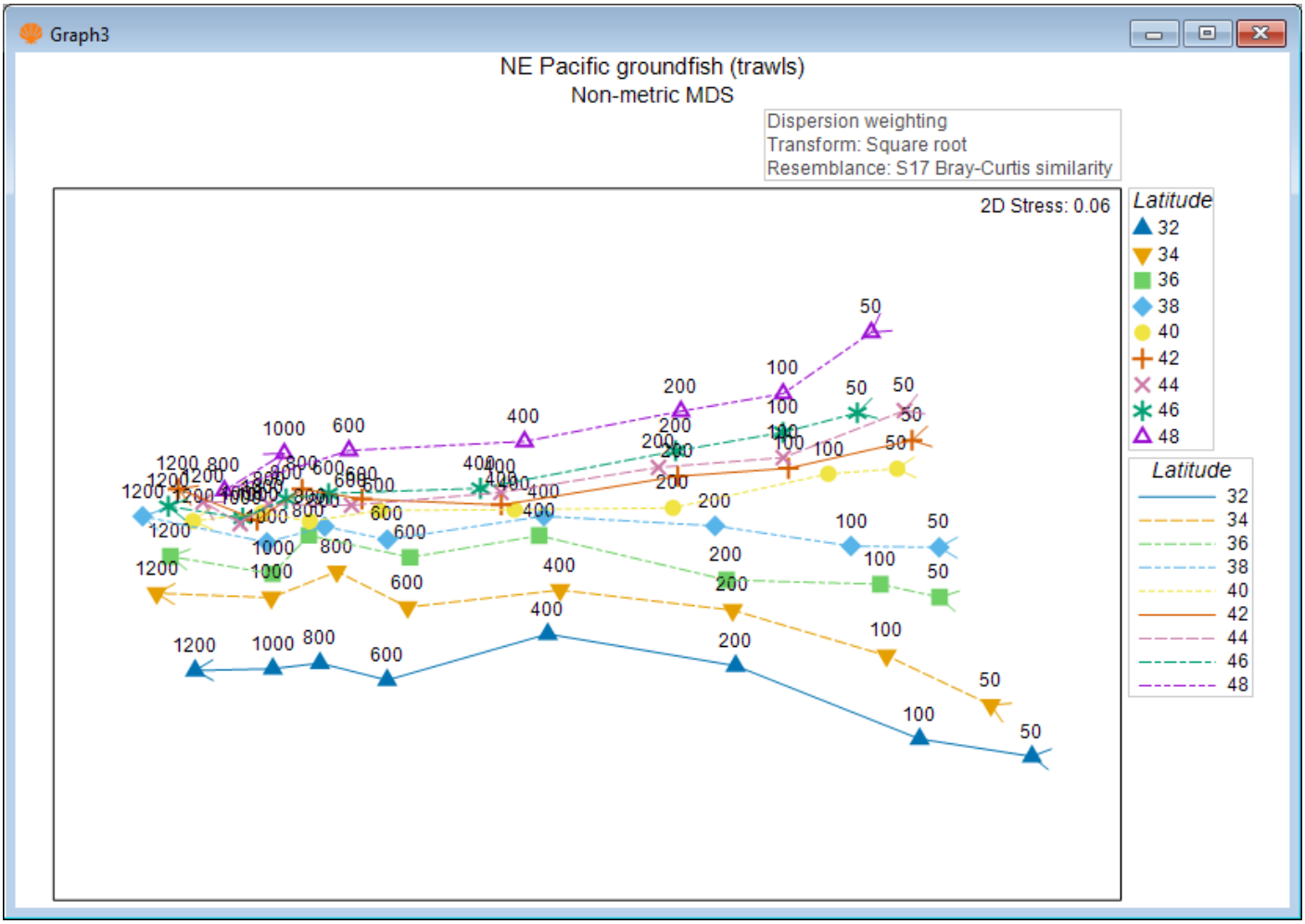
- **(a)** With symbols corresponding to 'Depth' and Labels correspond to 'Latitude' (optionally with trajectories for 'Latitude', split by 'Depth' groups); or
- **(b)** With symbols corresponding to 'Latitude' and Labels correspond to 'Depth' (optionally with trajectories for 'Depth', split by 'Latitude' groups).

To obtain (a): From 'Graph3', click **Graph** > **Sample Labels & Symbols...** (Labels > $\checkmark$Plot > $\checkmark$By factor Latitude) & (Symbols > $\checkmark$Plot > $\checkmark$By factor Depth). Get the trajectories by clicking **Graph** > **Special**, click the 'Overlays' tab then choose: Trajectory > $\checkmark$Overlay trajectory > Trajectory numeric factor: Latitude > $\checkmark$Split trajectory Depth.

The result looks like this:



To obtain (b): Simply swap the role of 'Latitude' and 'Depth' factors in the above instructions for (a). The result looks like this:

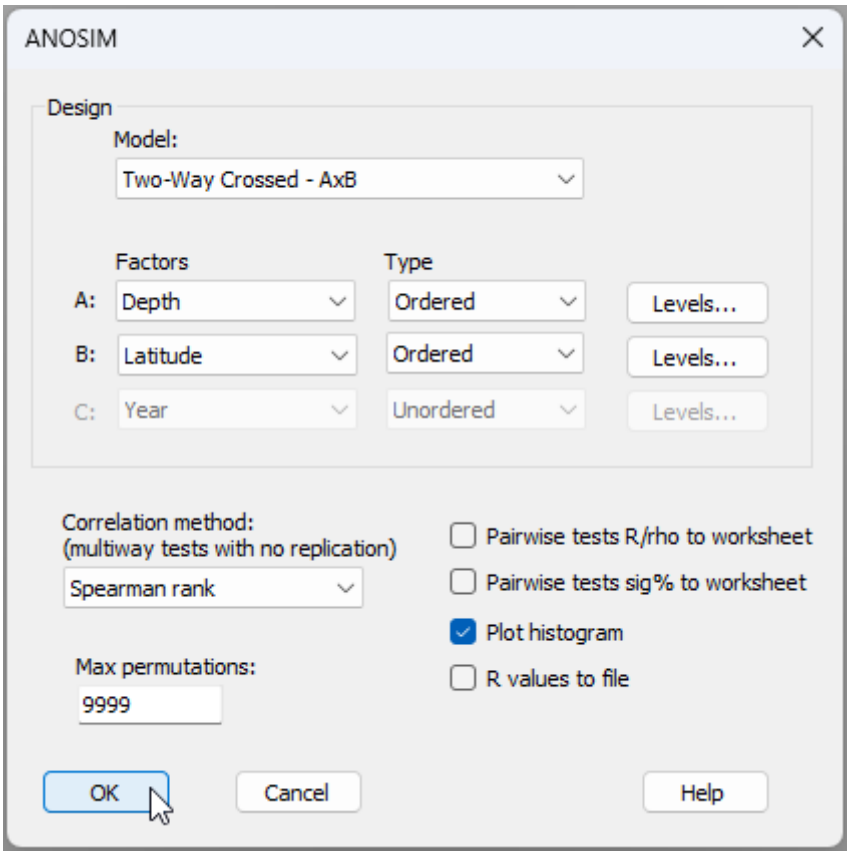


These ordinations show a very highly spatially structured ecological system, with marked gradual changes in fish assemblages with both latitude and depth. It is clear that latitudinal turnover in fish asesmblages is more marked at shallower depths than at deeper depths. In addition, turnover in fish assemblages with depth becomes less marked after about 600 m, particularly at higher latitudes.

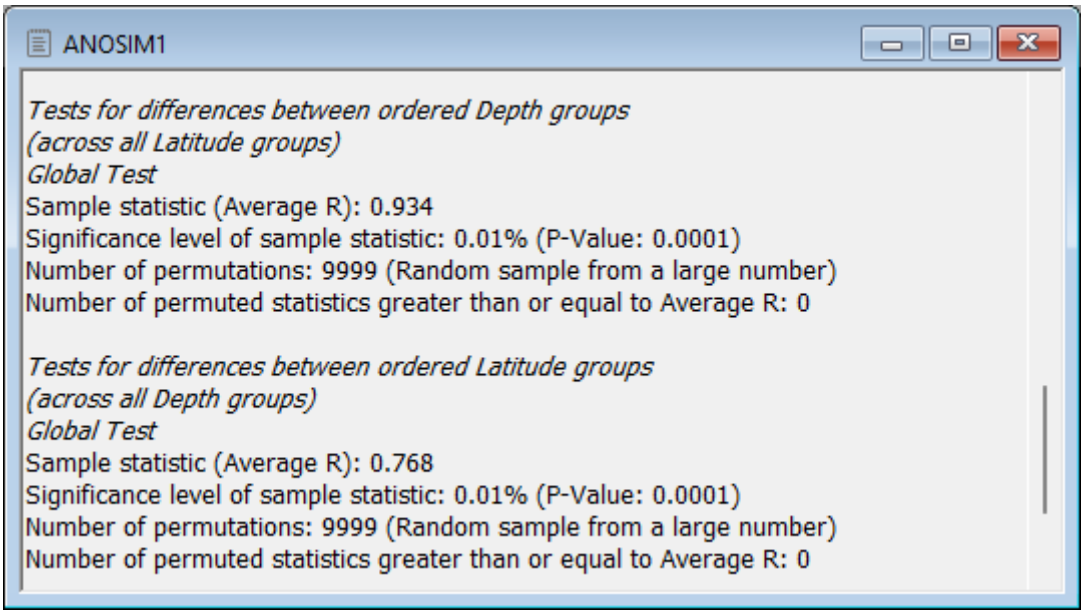
Testing ordered factors *via* ANOSIM

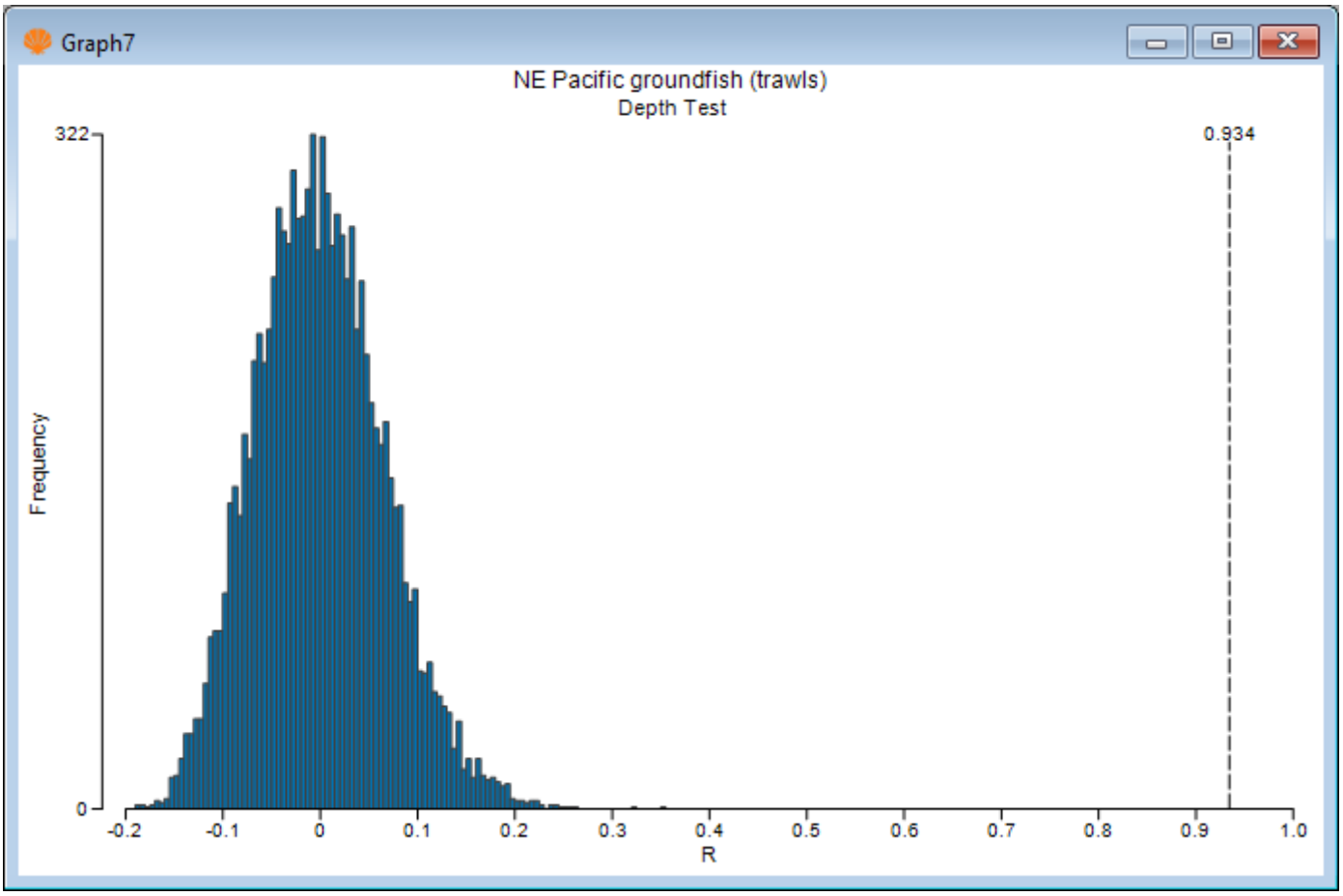
We can treat each of these factors as *ordered factors* in a *two-way ANOSIM* (see Somerfield *et al.* (2021a) and Somerfield *et al.* (2021b)) to test and quantify these effects in a non-parametric (rank-resemblance) framework.

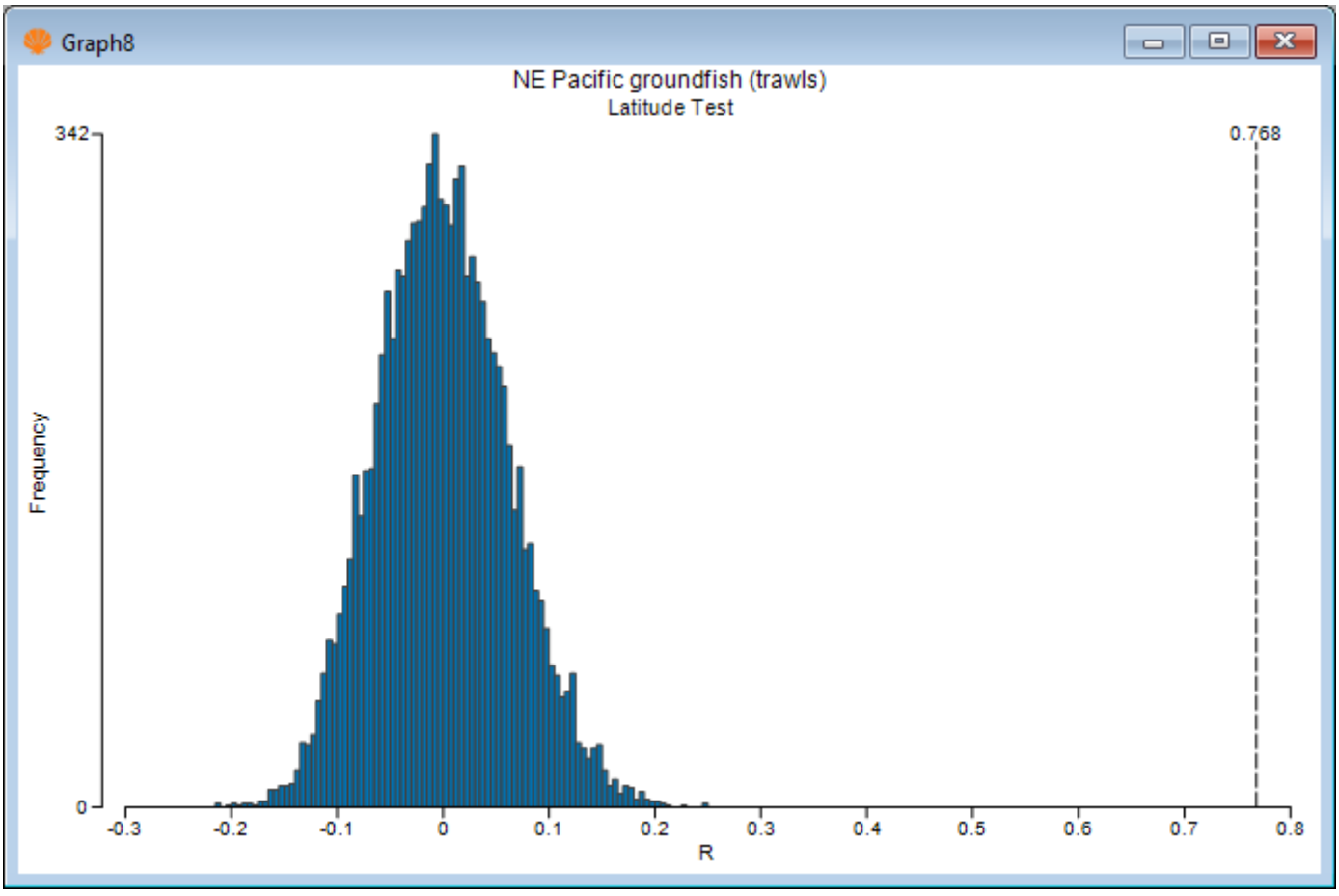
18. From the 'Resem1' matrix, click **Analyse** > **ANOSIM...**, and specify the two factors of 'Depth' and 'Latitude' as ordered factors in a two-way crossed design, as shown in the dialog below:



The results (not suprisingly) are uber clear (see the output file 'ANOSIM1' in the Explorer tree). There is a highly significant ordered effect of depth generating sequential turnover in groundfish assemblages from 50 m to 1200 m on the NE Pacific coast (ANOSIM, R^O = 0.934, P = 0.0001). There are also significant sequential changes (i.e., turnover) in groundfish assemblages along the latitudinal gradient, from 32°N to 48°N (ANOSIM, R^O = 0.768, P = 0.0001).







¶*NOAA Fisheries, NWFSC/FRAM, 2725 Montlake Blvd. East, Seattle, WA 98112, USA*