Mixed data types

Another example might be in attempting to reconcile two different types of data in the same matrix, e.g. counts of motile organisms and area cover of colonial species. These cases can be problematic. One solution is to use a similarity measure such as the Gower coefficient, which scales the range of each species across samples to be identical, but this generally performs badly because very rare species are given the same weight as very common ones. A preferable alternative is to use Bray-Curtis similarity as usual, but prior to that **Weight Variables** to convert counts into approximate area cover, species by species, or both counts and area cover into a rough estimate of biomass, or even just to balance the two sets of variables against each other in some arbitrary way, e.g. give the cover numbers 10 times as much weight, or 10 times less weight, keeping the counts unchanged, and see what difference it makes to the analysis. (See also the discussion on p5-19 of CiMC.)

Revision #1
Created 22 May 2024 01:03:29 by Arden
Updated 22 May 2024 01:05:16 by Arden