**问题2：关于生物群落数据的探索性分析，主要包括针对样地/样点的分析（Q-mode），以及针对物种/栖息地环境的分析（R-mode），请回答如下问题：**

**1）针对样点的分析主要是聚类分析，聚类分析是基于距离、相关系数或协方差？ 有哪几种主要聚类方法？**

聚类分析的本质是判别样点间的相似程度，也就是样点在数据空间内距离彼此的远近，因此是基于距离的，需构建距离矩阵。

常用的聚类方法包括：

* 层次聚类，计算样本间的距离，每次将距离最近的样本合并到同一个类，根据计算距离方法的不同分为单链接聚类、全链接聚类、ward最小方差聚类等；
* K均值聚类：预先指定要划分的簇数K，迭代寻找能使簇内平方误差（SSE）最小化的聚类方案；
* 此外在生态学中有较多应用的还有基于密度的聚类、谱聚类等。

**2）针对物种或环境的分析主要是排序分析，即识别物种按照特定的环境梯度变化的规律，排序是基于距离还是相关系数、协方差？如何选择排序模型（单峰或线性模型）？限制性排序与非限制性排序主要区别是什么？**

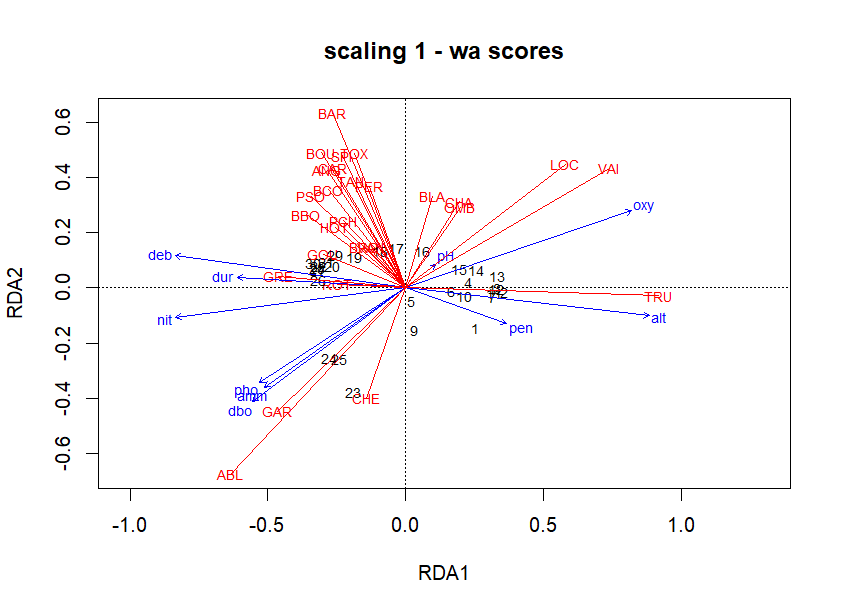
* 排序分析基于相关系数和协方差，基于协方差矩阵分析样本间的变异，或以相关系数表征物种和环境因子之间的相关性。
* 单峰模型假设物种丰度随环境梯度呈钟形曲线变化，在某一环境值附近出现最大值；线性模型假设物种丰度随环境梯度线性变化。选择方法是使用DCA进行梯度长度测量，若梯度长度第一轴即DCA1 > 4，选择单峰模型；当3 < DCA1 < 4，选择线性模型；当DCA1 < 3时，选用线性模型可能得到更合理结果。
* 主要的非限制性排序方法包括PCA和CA，限制性排序方法包括RDA和CCA，主要的区别在于是否增加环境变量约束（自变量-因变量关系约束）。非限制性排序只基于各个变量（例如各个物种）的共同变化模式进行，没有环境变量的约束限制；而限制性排序强调自变量受因变量影响下的变化情况，例如根据物种数据在环境变量约束下的变化进行排序。

**3）****对双序图/三序图的解释要看标度，当scaling=1，该图表示的意思是什么？当scaling=2，图中矢量长度和夹角分别表示的是什么？**

（以doubs河样点、鱼类、环境变量为例）

标度scaling=1时，关注样点之间的距离和结构。**样点（黑色数字）之间的欧氏距离**尽可能反映其在原始高维空间的差异和分布关系。

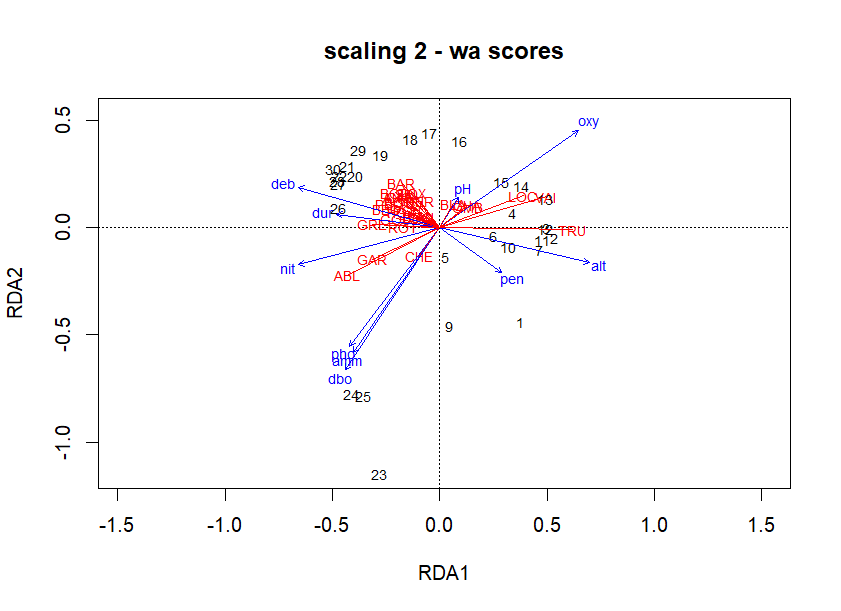
对自变量和因变量的关系进行定性解释。下图中红色矢量为物种，蓝色矢量为环境变量。各矢量的方向定性解释其对应变量在这个排序空间中的分布趋势；各个代表物种丰度的红色矢量与代表环境变量的蓝色矢量夹角反映一个物种的丰度与一个环境变量之间的相关性。

****

标度scaling=2时，关注的是变量之间的关系。图中的矢量长度包括物种矢量长度和环境变量的矢量长度。

其中，物种矢量长度代表该物种在排序轴上的方差大小，即物种在各个样方中的分布中的变异性，矢量越长，物种分布的变异性越大，对环境变量梯度的响应越明显；而环境矢量长度则代表环境变量在主排序轴上的解释能力大小和权重，矢量越长，该环境变量对排序结果就有更强的解释力。

至于矢量夹角，物种-环境变量之间的矢量夹角代表物种丰度与该环境变量间的相关性大小及其方向，可用于判断物种对环境条件的偏好；环境-环境变量之间的矢量夹角代表两环境变量间相关性大小和方向，可查看变量间共线性的状况；而物种-物种间矢量夹角可用于解释两个物种对环境变量的偏好、分布状况的相近程度。

****