ORIGIN ≡ 1

CANONICAL CORRESPONDENCE ANALYSIS

Canonical Correspondence Analysis (CCA) is a very popular technique especially in Ecology where one wishes to relate a table X of species occurrences among localites with a matrix Y of environmental data for each locality. As such, this method is an extension of standard Correspondence Analysis (CA) that has only table X, or Principal Components Analysis (PCA) with only a matrix X typically comprised of measurement data. CCA also differs from Canonical *Correlation* Analysis (CCA) where two sets of measurement data, matrix X & matrix Y respectively, are compared. CCA shares with Multivariate Multiple Regression (MMR) the sense that table X comprises a "dependent" set of variables, whereas table Y include "independent" variables used to "explain" X. In terms of computation, all of these methods are interrelated with several commonly implemented together by software functions, including those in R. The most common use of CCA in Ecology and other fields is "ordination", that is, graphical display of complex multivariate relationships in as few dimensions as possible. However, like regression, the method may also be used to compare "models" involving different sets of variables from matrix Y.
Although perhaps the preferred method among multivariate ordination procedures, I found worked examples of CCA suitable for a prototype surprisingly hard to find, perhaps because the method is considered "computationally intense". Useful references for CCA are linked to the CCA page on the Toolkit website. Both Greenacre (G) and de Leeuw & Mair (dLM) provide fine algebraic introductions to both CA and CCA. However, I was unable to implement their formulae directly into this prototype. The matrix formulation in (G) seems useful in showing the relationship between CA and CCA and how "projection" onto a space defined by matrix Y is accomplished. Both show clearly the use of scaling of the projection by means of an inverse square root matrix. However, to study calculations for ordination in detail, I followed the script for function anacor() in the R package {anacor}. This script with my annotations (lines marked by # to indicate steps in the calcuations) is also linked to the CCA page on the Toolkit website.
The output of CCA ordinations typically involve calculation of row and column "scores" in the restricted space of matrix Y along with various diagnostic measures indicating "variance explained" based on eigenvalue (or singular value) decomposition of an appropriate matrix (here matrix z). There are also measures intended to show how much variance is "explained" by matrix Y compared to the unrestricted CA problem that ignores Y entirely. Neither these measures, nor tests looking at "significance" of variables in Y, have been prototyped here. The scores are commonly plotted in the manner of biplots (triplots?) simultaneously showing points for rows (localities) and columns (species) as well as vectors indicating directions of variables in matrix Y. Interpreting these graphs, especially given multiple ways to scale each of these three elements, requires some practice. Discussions occurs in the linked papers by ter Braak & Verdonshot (tBV) and Oksanen et al. (O) and are highly recommended as a place to start.

Read in Data:

Localities (rows 1-12) in which species 1-10 (columns 2-11) are found. Environmental data for localities (columns 12-14).

M := READPRN("c:\DATA\Multivariate\CCAExample1.txt")

M =

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	15	26	20	9	5	7	32	47	19	17	460	72	350
2	2	19	46	39	15	5	7	36	44	42	18	465	80	350
3	3	29	13	9	12	1	26	29	31	37	27	465	65	340
4	4	47	34	46	14	3	25	7	27	46	49	470	69	320
5	5	18	37	43	20	0	10	22	19	48	35	460	75	310
6	6	32	44	28	24	39	2	2	11	10	0	455	79	300
7	7	8	47	9	33	7	7	23	3	21	17	450	82	300
8	8	31	24	12	42	11	17	46	9	8	35	450	82	280
9	9	32	20	48	46	25	18	6	4	35	38	450	81	280
10	10	19	43	26	47	24	7	23	10	1	27	445	83	200
11	11	12	40	13	30	3	8	5	15	5	18	440	85	120
12	12	11	35	8	31	28	12	6	22	19	19	440	85	100

Creating Submatricies:

$X := \text{submatrix}(M, 1, 12, 2, 11)$

$Y := \text{submatrix}(M, 1, 12, 12, 14)$

$r := \text{rows}(X) \qquad r = 12$

$c := \text{cols}(X) \qquad c = 10$

$i := 1 .. r \qquad j := 1 .. c$

$l_{r_i} := 1 \qquad l_{c_j} := 1$

$$T := \sum_i \sum_j X_{i,j} \qquad T = 2673$$

^ total occurrences

$$X = \begin{pmatrix} 15 & 26 & 20 & 9 & 5 & 7 & 32 & 47 & 19 & 17 \\ 19 & 46 & 39 & 15 & 5 & 7 & 36 & 44 & 42 & 18 \\ 29 & 13 & 9 & 12 & 1 & 26 & 29 & 31 & 37 & 27 \\ 47 & 34 & 46 & 14 & 3 & 25 & 7 & 27 & 46 & 49 \\ 18 & 37 & 43 & 20 & 0 & 10 & 22 & 19 & 48 & 35 \\ 32 & 44 & 28 & 24 & 39 & 2 & 2 & 11 & 10 & 0 \\ 8 & 47 & 9 & 33 & 7 & 7 & 23 & 3 & 21 & 17 \\ 31 & 24 & 12 & 42 & 11 & 17 & 46 & 9 & 8 & 35 \\ 32 & 20 & 48 & 46 & 25 & 18 & 6 & 4 & 35 & 38 \\ 19 & 43 & 26 & 47 & 24 & 7 & 23 & 10 & 1 & 27 \\ 12 & 40 & 13 & 30 & 3 & 8 & 5 & 15 & 5 & 18 \\ 11 & 35 & 8 & 31 & 28 & 12 & 6 & 22 & 19 & 19 \end{pmatrix}$$

^ Raw occurrence matrix

$$Y = \begin{pmatrix} 460 & 72 & 350 \\ 465 & 80 & 350 \\ 465 & 65 & 340 \\ 470 & 69 & 320 \\ 460 & 75 & 310 \\ 455 & 79 & 300 \\ 450 & 82 & 300 \\ 450 & 82 & 280 \\ 450 & 81 & 280 \\ 445 & 83 & 200 \\ 440 & 85 & 120 \\ 440 & 85 & 100 \end{pmatrix}$$

^ Environmental matrix

Matrix P of occurrence frequencies:

$$P := \frac{1}{T} X$$

^ Matrix P is called the "correspondence matrix"

$$P = \begin{pmatrix} 0.0056 & 0.0097 & 0.0075 & 0.0034 & 0.0019 & 0.0026 & 0.012 & 0.0176 & 0.0071 & 0.0064 \\ 0.0071 & 0.0172 & 0.0146 & 0.0056 & 0.0019 & 0.0026 & 0.0135 & 0.0165 & 0.0157 & 0.0067 \\ 0.0108 & 0.0049 & 0.0034 & 0.0045 & 0.0004 & 0.0097 & 0.0108 & 0.0116 & 0.0138 & 0.0101 \\ 0.0176 & 0.0127 & 0.0172 & 0.0052 & 0.0011 & 0.0094 & 0.0026 & 0.0101 & 0.0172 & 0.0183 \\ 0.0067 & 0.0138 & 0.0161 & 0.0075 & 0 & 0.0037 & 0.0082 & 0.0071 & 0.018 & 0.0131 \\ 0.012 & 0.0165 & 0.0105 & 0.009 & 0.0146 & 0.0007 & 0.0007 & 0.0041 & 0.0037 & 0 \\ 0.003 & 0.0176 & 0.0034 & 0.0123 & 0.0026 & 0.0026 & 0.0086 & 0.0011 & 0.0079 & 0.0064 \\ 0.0116 & 0.009 & 0.0045 & 0.0157 & 0.0041 & 0.0064 & 0.0172 & 0.0034 & 0.003 & 0.0131 \\ 0.012 & 0.0075 & 0.018 & 0.0172 & 0.0094 & 0.0067 & 0.0022 & 0.0015 & 0.0131 & 0.0142 \\ 0.0071 & 0.0161 & 0.0097 & 0.0176 & 0.009 & 0.0026 & 0.0086 & 0.0037 & 0.0004 & 0.0101 \\ 0.0045 & 0.015 & 0.0049 & 0.0112 & 0.0011 & 0.003 & 0.0019 & 0.0056 & 0.0019 & 0.0067 \\ 0.0041 & 0.0131 & 0.003 & 0.0116 & 0.0105 & 0.0045 & 0.0022 & 0.0082 & 0.0071 & 0.0071 \end{pmatrix}$$

Matrix P_s of Centered & Scaled frequencies:

$R := X \cdot l_c$

$C := X^T \cdot l_r$

Row (R) and Column (C) vectors of sums of matrix P:

$$IsD_r := \text{diag}\left(\frac{1}{\sqrt{R}}\right)$$

$$IsD_c := \text{diag}\left(\frac{1}{\sqrt{C}}\right)$$

< Inverse square-root matrices with values along the main diagonal

$$R = \begin{pmatrix} 197 \\ 271 \\ 214 \\ 298 \\ 252 \\ 192 \\ 175 \\ 235 \\ 272 \\ 227 \\ 149 \\ 191 \end{pmatrix} \qquad C = \begin{pmatrix} 273 \\ 409 \\ 301 \\ 323 \\ 151 \\ 146 \\ 237 \\ 242 \\ 291 \\ 300 \end{pmatrix}$$

$P_s := IsD_r \cdot (X) \cdot IsD_c$ **< centered & scaled correspondence matrix.**

$$P_s = \begin{pmatrix} 0.06468 & 0.0916 & 0.08213 & 0.03568 & 0.02899 & 0.04128 & 0.1481 & 0.21526 & 0.07935 & 0.06993 \\ 0.06985 & 0.13817 & 0.13655 & 0.0507 & 0.02472 & 0.03519 & 0.14205 & 0.17181 & 0.14956 & 0.06313 \\ 0.11998 & 0.04394 & 0.03546 & 0.04564 & 0.00556 & 0.14709 & 0.12877 & 0.13622 & 0.14827 & 0.10656 \\ 0.16478 & 0.09739 & 0.15359 & 0.04513 & 0.01414 & 0.11985 & 0.02634 & 0.10054 & 0.15621 & 0.16388 \\ 0.06863 & 0.11525 & 0.15613 & 0.0701 & 0 & 0.05213 & 0.09002 & 0.07694 & 0.17725 & 0.12729 \\ 0.13977 & 0.15701 & 0.11647 & 0.09637 & 0.22905 & 0.01195 & 0.00938 & 0.05103 & 0.04231 & 0 \\ 0.0366 & 0.17568 & 0.03921 & 0.1388 & 0.04306 & 0.04379 & 0.11294 & 0.01458 & 0.09306 & 0.07419 \\ 0.12239 & 0.07741 & 0.04512 & 0.15245 & 0.05839 & 0.09178 & 0.19492 & 0.03774 & 0.03059 & 0.13182 \\ 0.11743 & 0.05996 & 0.16775 & 0.15519 & 0.12336 & 0.09033 & 0.02363 & 0.01559 & 0.1244 & 0.13303 \\ 0.07632 & 0.14112 & 0.09947 & 0.17357 & 0.12963 & 0.03845 & 0.09916 & 0.04267 & 0.00389 & 0.10346 \\ 0.0595 & 0.16203 & 0.06139 & 0.13675 & 0.02 & 0.05424 & 0.02661 & 0.07899 & 0.02401 & 0.08514 \\ 0.04817 & 0.12522 & 0.03336 & 0.12481 & 0.16487 & 0.07186 & 0.0282 & 0.10233 & 0.08059 & 0.07937 \end{pmatrix}$$

Square Environmental Matrix weighted by Row Occurrences:

$one12_i := 1$

$xr := augment(one12, Y)$

$cx := xr^T \cdot diag(R) \cdot xr$

Matrix xr appears as a design matrix in linear regression >

$$xr = \begin{pmatrix} 1 & 460 & 72 & 350 \\ 1 & 465 & 80 & 350 \\ 1 & 465 & 65 & 340 \\ 1 & 470 & 69 & 320 \\ 1 & 460 & 75 & 310 \\ 1 & 455 & 79 & 300 \\ 1 & 450 & 82 & 300 \\ 1 & 450 & 82 & 280 \\ 1 & 450 & 81 & 280 \\ 1 & 445 & 83 & 200 \\ 1 & 440 & 85 & 120 \\ 1 & 440 & 85 & 100 \end{pmatrix}$$

$$cx = \begin{pmatrix} 2673 & 1217000 & 207797 & 744480 \\ 1217000 & 554335200 & 94477215 & 340599250 \\ 207797 & 94477215 & 16256083 & 57052100 \\ 744480 & 340599250 & 57052100 & 222715200 \end{pmatrix}$$

^ weighted squared matrix of environmental factors - similar to a variance/covariance matrix...

Finding Inverse square root of matrix cx:

$k := 1 .. cols(xr)$

$\Lambda := reverse(sort(eigenvals(cx)))$

$E^{\langle k \rangle} := eigenvec(cx, \Lambda_k)$

$\Psi_{k,k} := \dfrac{1}{\sqrt{\Lambda_k}}$

$or := E \cdot \Psi \cdot E^T$

^ inverse square root of matrix cx

$$\Lambda = \begin{pmatrix} 783221050.965 \\ 10012084.378 \\ 76020.518 \\ 0.139 \end{pmatrix}$$

$$E = \begin{pmatrix} -0.00184 & -0.00138 & 0.00123 & -1 \\ -0.83938 & -0.5008 & -0.21129 & 0.00198 \\ -0.14242 & -0.1725 & 0.97466 & 0.0017 \\ -0.52456 & 0.8482 & 0.07347 & -0.00012 \end{pmatrix}$$

Matrix square root method involves singular value decomposition (SVD) to find singular values Λ, taking the square root, and then reconsituting matrix or.

$$\Psi = \begin{pmatrix} 0.000035732 & 0 & 0 & 0 \\ 0 & 0.000316037 & 0 & 0 \\ 0 & 0 & 0.003626892 & 0 \\ 0 & 0 & 0 & 2.685744921 \end{pmatrix}$$

$$or = \begin{pmatrix} 2.685726598 & -0.00531142 & -0.004569787 & 0.000315808 \\ -0.00531142 & 0.000276854 & -0.000706283 & -0.00017544 \\ -0.004569787 & -0.000706283 & 0.003463312 & 0.000215611 \\ 0.000315808 & -0.00017544 & 0.000215611 & 0.000256817 \end{pmatrix}$$

< inverse square root of cx

Constructing Matrix z for SVD in CCA:

$z := or \cdot \left(xr^T \cdot X\right)$

$$z = \begin{pmatrix} 0.112272 & -0.513947 & -0.995178 & 1.132256 & 0.45206 & 0.382377 & 0.849082 & -0.428341 & -0.718296 & 0.20134 \\ 4.538194 & 7.281128 & 4.709606 & 6.054265 & 2.926787 & 2.663088 & 3.324914 & 4.025894 & 4.497886 & 5.33714 \\ 0.072093 & 2.649151 & 1.028394 & 3.033526 & 1.637145 & -0.493411 & 1.337509 & -0.564837 & -0.451562 & 0.17452 \\ 2.849762 & 2.428665 & 3.393851 & 0.9557 & 0.181769 & 1.250612 & 3.193749 & 2.757958 & 3.905465 & 2.41498 \end{pmatrix}$$

$one4_k := 1$

$ff := one4 \cdot \sqrt{C}^T$

$$ff = \begin{pmatrix} 16.523 & 20.224 & 17.349 & 17.972 & 12.288 & 12.083 & 15.395 & 15.556 & 17.059 & 17.321 \\ 16.523 & 20.224 & 17.349 & 17.972 & 12.288 & 12.083 & 15.395 & 15.556 & 17.059 & 17.321 \\ 16.523 & 20.224 & 17.349 & 17.972 & 12.288 & 12.083 & 15.395 & 15.556 & 17.059 & 17.321 \\ 16.523 & 20.224 & 17.349 & 17.972 & 12.288 & 12.083 & 15.395 & 15.556 & 17.059 & 17.321 \end{pmatrix}$$

$z := \overrightarrow{\left(\frac{z}{ff}\right)}$ **< scalar division by equivalent position in z and ff above**

$$z = \begin{pmatrix} 0.006795 & -0.025413 & -0.057361 & 0.063 & 0.036788 & 0.031646 & 0.055154 & -0.027535 & -0.042107 & 0.0116 \\ 0.274664 & 0.360029 & 0.271457 & 0.336868 & 0.238179 & 0.220399 & 0.215976 & 0.258794 & 0.263671 & 0.308 \\ 0.004363 & 0.130992 & 0.059276 & 0.16879 & 0.133229 & -0.040835 & 0.086881 & -0.036309 & -0.026471 & 0.0100 \\ 0.172475 & 0.12009 & 0.195618 & 0.053177 & 0.014792 & 0.103501 & 0.207456 & 0.177288 & 0.228942 & 0.1394 \end{pmatrix}$$

^ Matrix z is contructed in two stages following R script for anacor()

Singular Value Decomposition of matrix z:

$\gamma_u := reverse\left(sort\left(eigenvals\left(z \cdot z^T\right)\right)\right)$

$\gamma_v := reverse\left(sort\left(eigenvals\left(z^T \cdot z\right)\right)\right)$

$U^{\langle k \rangle} := eigenvec\left(z \cdot z^T, \gamma_{u_k}\right)$

$V^{\langle k \rangle} := eigenvec\left(z^T \cdot z, \gamma_{v_k}\right)$

$\lambda := \sqrt{\gamma_u}$

$$\gamma_u = \begin{pmatrix} 1 \\ 0.084927 \\ 0.016151 \\ 0.011351 \end{pmatrix} \qquad \gamma_v = \begin{pmatrix} 1 \\ 0.08493 \\ 0.01615 \\ 0.01135 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

^ eigenvalues >

$$\lambda = \begin{pmatrix} 1 \\ 0.291422 \\ 0.127088 \\ 0.106543 \end{pmatrix}$$

^ singular values

$$U = \begin{pmatrix} 0.009160934 & 0.249040733 & -0.059317914 & -0.966631355 \\ 0.877330237 & 0.180138436 & -0.437253011 & 0.081557368 \\ 0.162908334 & 0.724613243 & 0.653028082 & 0.14815818 \\ 0.451296585 & -0.616818048 & 0.615504276 & -0.192409391 \end{pmatrix}$$

$UU := augment\left(U^{\langle 2 \rangle}, U^{\langle 3 \rangle}, U^{\langle 4 \rangle}\right)$

$VV := augment\left(V^{\langle 2 \rangle}, V^{\langle 3 \rangle}, V^{\langle 4 \rangle}\right)$

^ removing first column defining general level in matrix U & V.

$$V = \begin{pmatrix} 0.31958166 & -0.17862277 & -0.09042518 & 0.15680894 \\ 0.39116697 & 0.27235862 & 0.02786347 & -0.47144665 \\ 0.33557049 & -0.14787647 & 0.34479687 & -0.45737316 \\ 0.34761762 & 0.56920887 & -0.06356814 & 0.17502971 \\ 0.23767801 & 0.47862723 & -0.0804131 & -0.00710954 \\ 0.23370981 & -0.1573245 & -0.481618 & 0.3621022 \\ 0.29776572 & -0.04243601 & 0.6823425 & 0.58890386 \\ 0.30089031 & -0.32908732 & -0.20548258 & -0.07725725 \\ 0.32994915 & -0.42339438 & 0.0852604 & -0.13359832 \\ 0.33501261 & -0.06965196 & -0.33854873 & 0.10737744 \end{pmatrix}$$

Calculating CCV Scores:

Raw Scores:

$x := xr \cdot or \cdot UU$

$y := \mathrm{diag}\left(\frac{1}{\sqrt{C}}\right) \cdot VV$

$$x = \begin{pmatrix} -0.0110604 & 0.0066725 & -0.0278851 \\ -0.0105525 & 0.0285905 & 0.0361949 \\ -0.0276703 & -0.0195917 & -0.0329045 \\ -0.0325049 & -0.0167078 & 0.0185912 \\ -0.0095214 & 0.0013212 & 0.0000908 \\ 0.0028573 & 0.0115257 & -0.0026428 \\ 0.0145734 & 0.0223244 & -0.0135782 \\ 0.013676 & 0.0151874 & -0.0068372 \\ 0.0125647 & 0.0122131 & -0.0116686 \\ 0.0195799 & -0.0085104 & -0.0004714 \\ 0.0265951 & -0.029234 & 0.0107257 \\ 0.0256976 & -0.036371 & 0.0174666 \end{pmatrix}$$

^ rows (sites)

$$y = \begin{pmatrix} -0.0108107 & -0.0054728 & 0.0094905 \\ 0.0134673 & 0.0013778 & -0.0233115 \\ -0.0085235 & 0.0198738 & -0.0263626 \\ 0.0316716 & -0.003537 & 0.0097389 \\ 0.0389501 & -0.0065439 & -0.0005786 \\ -0.0130203 & -0.039859 & 0.0299678 \\ -0.0027565 & 0.0443229 & 0.0382534 \\ -0.0211545 & -0.0132089 & -0.0049663 \\ -0.0248198 & 0.0049981 & -0.0078317 \\ -0.0040214 & -0.0195461 & 0.0061994 \end{pmatrix}$$

^ columns (species)

"Standard" Scores:

$kk := 2..4$

$\lambda\lambda_{kk-1} := \lambda_{kk}$ **< singular values**

$\gamma\gamma_{kk-1} := \gamma u_{kk}$ **< eigenvalues**

$$\lambda\lambda = \begin{pmatrix} 0.291 \\ 0.127 \\ 0.107 \end{pmatrix} \qquad \gamma\gamma = \begin{pmatrix} 0.085 \\ 0.016 \\ 0.011 \end{pmatrix}$$

$x_s := x \cdot \sqrt{T}$

$y_s := y \cdot \sqrt{T}$

Scores with "standard" scaling {anacor}
Scores(CCA, scaling=1) {vegan}

$$x_s = \begin{pmatrix} -0.5718352 & 0.3449751 & -1.441689 \\ -0.5455741 & 1.4781568 & 1.8713169 \\ -1.430586 & -1.0129127 & -1.7011987 \\ -1.6805389 & -0.8638087 & 0.9611827 \\ -0.4922669 & 0.0683074 & 0.0046956 \\ 0.1477241 & 0.5958892 & -0.1366344 \\ 0.7534594 & 1.154195 & -0.7020067 \\ 0.7070622 & 0.7852037 & -0.3534917 \\ 0.649608 & 0.6314322 & -0.6032765 \\ 1.0123004 & -0.4399986 & -0.0243737 \\ 1.3749929 & -1.5114294 & 0.5545291 \\ 1.3285957 & -1.8804207 & 0.903044 \end{pmatrix}$$

$$'_s = \begin{pmatrix} -0.5589269 & -0.2829486 & 0.4906694 \\ 0.696272 & 0.0712317 & -1.2052312 \\ -0.4406718 & 1.0274946 & -1.3629719 \\ 1.6374569 & -0.182868 & 0.5035122 \\ 2.0137632 & -0.3383279 & -0.0299125 \\ -0.6731617 & -2.0607522 & 1.5493667 \\ -0.1425148 & 2.2915415 & 1.9777423 \\ -1.0937119 & -0.6829153 & -0.2567622 \\ -1.283211 & 0.2584047 & -0.4049058 \\ -0.2079085 & -1.0105552 & 0.3205176 \end{pmatrix}$$

"Centroid" Scores:

$gg := one12 \cdot \lambda\lambda^T$

$x_c := \overrightarrow{(x_s \cdot gg)}$

$one10_j := 1$

$hh := one10 \cdot \lambda\lambda^T$

$y_c := \overrightarrow{(y_s \cdot hh)}$

Scores with "standard" scaling {anacor}

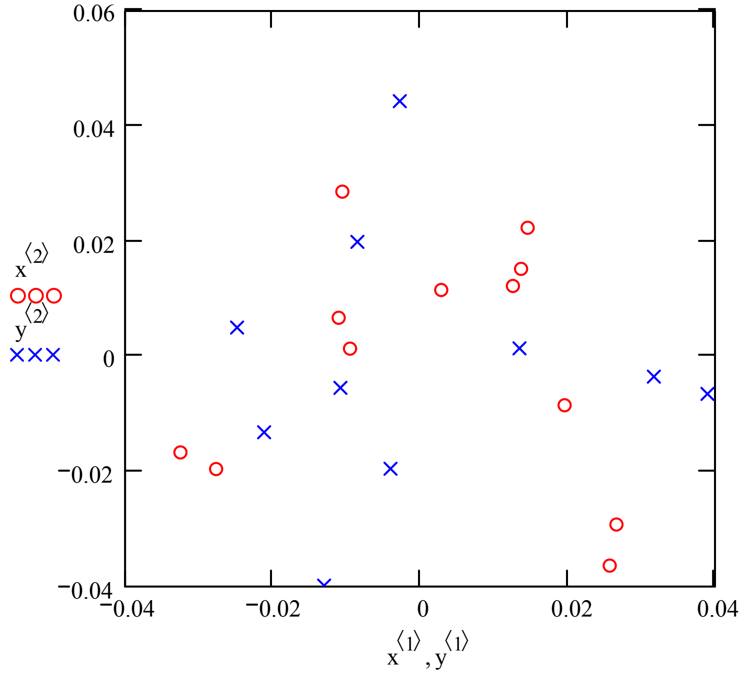
$$x_c = \begin{pmatrix} -0.1666451 & 0.0438423 & -0.1536014 \\ -0.158992 & 0.1878565 & 0.1993751 \\ -0.4169036 & -0.1287294 & -0.1812503 \\ -0.4897453 & -0.10978 & 0.102407 \\ -0.1434572 & 0.0086811 & 0.0005003 \\ 0.04305 & 0.0757305 & -0.0145574 \\ 0.2195743 & 0.1466847 & -0.0747937 \\ 0.2060532 & 0.0997902 & -0.0376619 \\ 0.1893098 & 0.0802476 & -0.0642747 \\ 0.2950062 & -0.0559187 & -0.0025968 \\ 0.4007026 & -0.192085 & 0.059081 \\ 0.3871814 & -0.2389795 & 0.0962127 \end{pmatrix}$$

$$'_c = \begin{pmatrix} -0.1628833 & -0.0359595 & 0.0522772 \\ 0.2029087 & 0.0090527 & -0.1284086 \\ -0.1284213 & 0.1305826 & -0.1452147 \\ 0.4771903 & -0.0232404 & 0.0536455 \\ 0.586854 & -0.0429975 & -0.003187 \\ -0.1961738 & -0.2618975 & 0.1650737 \\ -0.0415319 & 0.2912281 & 0.210714 \\ -0.3187312 & -0.0867905 & -0.0273561 \\ -0.3739554 & 0.0328402 & -0.0431397 \\ -0.060589 & -0.1284298 & 0.0341488 \end{pmatrix}$$

Biplots of Scores:

"Standard" scaling:

CCA 2

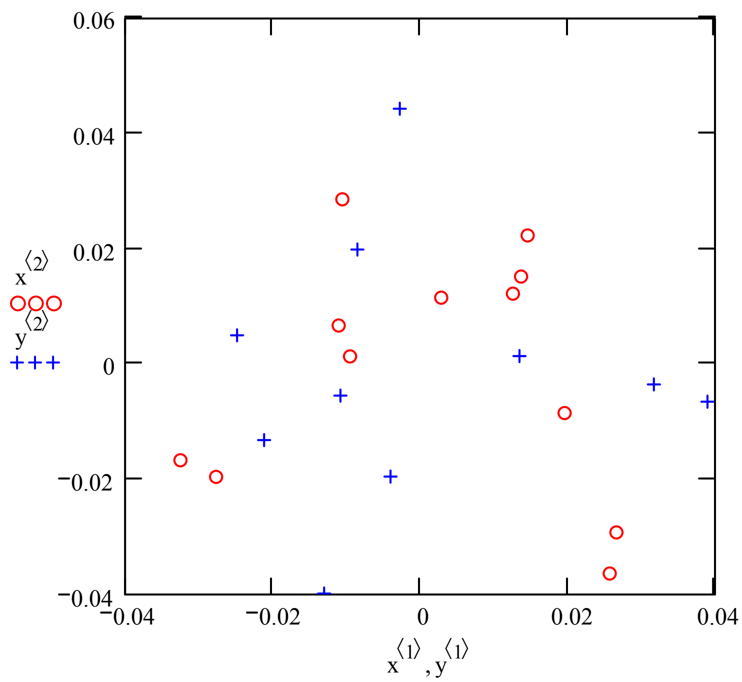


Red = rows (sites)
Blue = cols (species)

CCA 1

"Centroid" Scaling:

CCA 2



CCA 1

Prototype in R:

```
#CANONICAL CORRESPONDENCE ANALYSIS
#EXAMPLE DATA FROM:
#http://www.xlstat.com/en/learning-center/tutorials/running-a-canonical-correspondence-analysis-cca-with-xlstat-ada.html
M=read.table("c:/DATA/Multivariate/CCAExample1.txt")
M
X=data.frame(M[,2:11],row.names=M$Site)
X
Y=data.frame(M[,12:14],row.names=M$Site)
Y
```

> X

	Insect1	Insect2	Insect3	Insect4	Insect5	Insect6	Insect7	Insect8	Insect9	Insect10
Site1	15	26	20	9	5	7	32	47	19	17
Site2	19	46	39	15	5	7	36	44	42	18
Site3	29	13	9	12	1	26	29	31	37	27
Site4	47	34	46	14	3	25	7	27	46	49
Site5	18	37	43	20	0	10	22	19	48	35
Site6	32	44	28	24	39	2	2	11	10	0
Site7	8	47	9	33	7	7	23	3	21	17
Site8	31	24	12	42	11	17	46	9	8	35
Site9	32	20	48	46	25	18	6	4	35	38
Site10	19	43	26	47	24	7	23	10	1	27
Site11	12	40	13	30	3	8	5	15	5	18
Site12	11	35	8	31	28	12	6	22	19	19

```
library(anacor)
library(vegan)

#USING {anacor}:
CCAas=anacor(X,row.covariates=Y,ndim=3,scaling=c("standard","standard"))
CCAas
```

> Y

	Altitude	Humidity	Disttolake
Site1	460	72	350
Site2	465	80	350
Site3	465	65	340
Site4	470	69	320
Site5	460	75	310
Site6	455	79	300
Site7	450	82	300
Site8	450	82	280
Site9	450	81	280
Site10	445	83	200
Site11	440	85	120
Site12	440	85	100

$$\gamma\gamma = \begin{pmatrix} 0.085 \\ 0.016 \\ 0.011 \end{pmatrix}$$

$$\sum \gamma\gamma = 0.1124293$$

> CCAas

```
CA fit:
Sum of eigenvalues:  0.1124293

Total chi-square value: 300.524

Chi-Square decomposition:
              Chisq Proportion Cumulative Proportion
Component 1 227.009      0.279                 0.279
Component 2  43.173      0.053                 0.332
Component 3  30.342      0.037                 0.369
```

summary(CCAas)

$$\lambda\lambda = \begin{pmatrix} 0.29142 \\ 0.12709 \\ 0.10654 \end{pmatrix}$$

< singular values

"Standard" Scores:

$$x_S = \begin{pmatrix} -0.57184 & 0.34498 & -1.44169 \\ -0.54557 & 1.47816 & 1.87132 \\ -1.43059 & -1.01291 & -1.7012 \\ -1.68054 & -0.86381 & 0.96118 \\ -0.49227 & 0.06831 & 0.0047 \\ 0.14772 & 0.59589 & -0.13663 \\ 0.75346 & 1.15419 & -0.70201 \\ 0.70706 & 0.7852 & -0.35349 \\ 0.64961 & 0.63143 & -0.60328 \\ 1.0123 & -0.44 & -0.02437 \\ 1.37499 & -1.51143 & 0.55453 \\ 1.3286 & -1.88042 & 0.90304 \end{pmatrix}$$

$$y_S = \begin{pmatrix} -0.55893 & -0.28295 & 0.49067 \\ 0.69627 & 0.07123 & -1.20523 \\ -0.44067 & 1.02749 & -1.36297 \\ 1.63746 & -0.18287 & 0.50351 \\ 2.01376 & -0.33833 & -0.02991 \\ -0.67316 & -2.06075 & 1.54937 \\ -0.14251 & 2.29154 & 1.97774 \\ -1.09371 & -0.68292 & -0.25676 \\ -1.28321 & 0.2584 & -0.40491 \\ -0.20791 & -1.01056 & 0.32052 \end{pmatrix}$$

```
> summary(CCAas)
z-test for singular values:
   Singular Values Asymptotical SE p-value
D1          0.2914          0.0171       0
D2          0.1271          0.0174       0
D3          0.1065          0.0180       0

Row scores:
              D1       D2       D3
Site1   -0.57184  0.34498 -1.44169
Site2   -0.54557  1.47816  1.87132
Site3   -1.43059 -1.01291 -1.70120
Site4   -1.68054 -0.86381  0.96118
Site5   -0.49227  0.06831  0.00470
Site6    0.14772  0.59589 -0.13663
Site7    0.75346  1.15419 -0.70201
Site8    0.70706  0.78520 -0.35349
Site9    0.64961  0.63143 -0.60328
Site10   1.01230 -0.44000 -0.02437
Site11   1.37499 -1.51143  0.55453
Site12   1.32860 -1.88042  0.90304

Column scores:
              D1       D2       D3
Insect1  -0.55893 -0.28295 -0.49067
Insect2   0.69627  0.07123  1.20523
Insect3  -0.44067  1.02749  1.36297
Insect4   1.63746 -0.18287 -0.50351
Insect5   2.01376 -0.33833  0.02991
Insect6  -0.67316 -2.06075 -1.54937
Insect7  -0.14251  2.29154 -1.97774
Insect8  -1.09371 -0.68292  0.25676
Insect9  -1.28321  0.25840  0.40491
Insect10 -0.20791 -1.01056 -0.32052
```

"Standard" scores match.

#USING {vegan}:
CCAv=cca(X,Y,scale=F)
CCAv

```
> CCAv
Call: cca(X = X, Y = Y, scale = F)

               Inertia Proportion Rank
Total           0.3047     1.0000
Constrained     0.1124     0.3689    3
Unconstrained   0.1923     0.6311    8
Inertia is mean squared contingency coefficient

Eigenvalues for constrained axes:
   CCA1    CCA2    CCA3
0.08493 0.01615 0.01135

Eigenvalues for unconstrained axes:
      CA1       CA2       CA3       CA4       CA5       CA6       CA7       CA8
0.0751634 0.0508831 0.0273051 0.0191076 0.0128508 0.0051541 0.0011804 0.0006632
```

$$\gamma\gamma = \begin{pmatrix} 0.08493 \\ 0.01615 \\ 0.01135 \end{pmatrix}$$

^ eigenvalues

summary(CCAv)

> summary(CCAv)

```
Call:
cca(X = X, Y = Y, scale = F)
Partitioning of mean squared contingency coefficient:
              Inertia Proportion
Total          0.3047     1.0000
Constrained    0.1124     0.3689
Unconstrained  0.1923     0.6311
Eigenvalues, and their contribution to the mean squared contingency coefficient
Importance of components:
                         CCA1    CCA2    CCA3     CA1     CA2     CA3     CA4     CA5      CA6     CA7       CA8
Eigenvalue            0.08493 0.01615 0.01135 0.07516 0.05088 0.02731 0.01911 0.01285 0.005154 0.00118 0.0006632
Proportion Explained  0.27869 0.05300 0.03725 0.24665 0.16697 0.08960 0.06270 0.04217 0.016910 0.00387 0.0021800
Cumulative Proportion 0.27869 0.33169 0.36894 0.61559 0.78256 0.87216 0.93487 0.97704 0.993950 0.99782 1.0000000
Accumulated constrained eigenvalues
Importance of components:
                         CCA1    CCA2    CCA3
Eigenvalue            0.08493 0.01615 0.01135
Proportion Explained  0.75538 0.14366 0.10096
Cumulative Proportion 0.75538 0.89904 1.00000
Scaling 2 for species and site scores
* Species are scaled proportional to eigenvalues
* Sites are unscaled: weighted dispersion equal on all dimensions
...
```

scores(CCAv,scaling=1)

$$
y_S = \begin{pmatrix}
-0.5589269 & -0.2829486 & 0.4906694 \\
0.696272 & 0.0712317 & -1.2052312 \\
-0.4406718 & 1.0274946 & -1.3629719 \\
1.6374569 & -0.182868 & 0.5035122 \\
2.0137632 & -0.3383279 & -0.0299125 \\
-0.6731617 & -2.0607522 & 1.5493667 \\
-0.1425148 & 2.2915415 & 1.9777423 \\
-1.0937119 & -0.6829153 & -0.2567622 \\
-1.283211 & 0.2584047 & -0.4049058 \\
-0.2079085 & -1.0105552 & 0.3205176
\end{pmatrix}
$$

^"Standard" scaling for insects (columns) only.

> scores(CCAv,scaling=1)

```
$species
               CCA1        CCA2
Insect1  -0.5589269  0.28294860
Insect2   0.6962720 -0.07123166
Insect3  -0.4406718 -1.02749460
Insect4   1.6374569  0.18286802
Insect5   2.0137632  0.33832788
Insect6  -0.6731617  2.06075216
Insect7  -0.1425148 -2.29154150
Insect8  -1.0937119  0.68291525
Insect9  -1.2832110 -0.25840465
Insect10 -0.2079085  1.01055520

$sites
             CCA1         CCA2
Site1  -0.3191922 -0.149022999
Site2  -0.2832097 -0.236986061
Site3  -0.4583770  0.124224459
Site4  -0.3707168  0.197095329
Site5  -0.2778443 -0.126715861
Site6   0.4778799 -0.004164712
Site7   0.2895244 -0.150898936
Site8   0.1686596 -0.105513682
Site9   0.1117459  0.112547767
Site10  0.4728175 -0.053354991
Site11  0.2544837  0.173555168
Site12  0.3168404  0.250451475
```

"Centroid" Scores:

$$x_c = \begin{pmatrix} -0.16665 & 0.04384 & -0.1536 \\ -0.15899 & 0.18786 & 0.19938 \\ -0.4169 & -0.12873 & -0.18125 \\ -0.48975 & -0.10978 & 0.10241 \\ -0.14346 & 0.00868 & 0.0005 \\ 0.04305 & 0.07573 & -0.01456 \\ 0.21957 & 0.14668 & -0.07479 \\ 0.20605 & 0.09979 & -0.03766 \\ 0.18931 & 0.08025 & -0.06427 \\ 0.29501 & -0.05592 & -0.0026 \\ 0.4007 & -0.19209 & 0.05908 \\ 0.38718 & -0.23898 & 0.09621 \end{pmatrix}$$

$$y_c = \begin{pmatrix} -0.16288 & -0.03596 & 0.05228 \\ 0.20291 & 0.00905 & -0.12841 \\ -0.12842 & 0.13058 & -0.14521 \\ 0.47719 & -0.02324 & 0.05365 \\ 0.58685 & -0.043 & -0.00319 \\ -0.19617 & -0.2619 & 0.16507 \\ -0.04153 & 0.29123 & 0.21071 \\ -0.31873 & -0.08679 & -0.02736 \\ -0.37396 & 0.03284 & -0.04314 \\ -0.06059 & -0.12843 & 0.03415 \end{pmatrix}$$

```
#"CENTROID" SCORES:
#USING {anacor}:
CCAac=anacor(X,row.covariates=Y,ndim=3,scaling=c("centroid","centroid"))
CCAac
summary(CCAac)
#USING {vegan}
scores(CCAv,scaling=2)
```

```
> summary(CCAac)
...
Row scores:
             D1       D2       D3
Site1  -0.16665  0.04384 -0.15360
Site2  -0.15899  0.18786  0.19938
Site3  -0.41690 -0.12873 -0.18125
Site4  -0.48975 -0.10978  0.10241
Site5  -0.14346  0.00868  0.00050
Site6   0.04305  0.07573 -0.01456
Site7   0.21957  0.14668 -0.07479
Site8   0.20605  0.09979 -0.03766
Site9   0.18931  0.08025 -0.06427
Site10  0.29501 -0.05592 -0.00260
Site11  0.40070 -0.19209  0.05908
Site12  0.38718 -0.23898  0.09621

Column scores:
              D1       D2       D3
Insect1  -0.16288 -0.03596 -0.05228
Insect2   0.20291  0.00905  0.12841
Insect3  -0.12842  0.13058  0.14521
Insect4   0.47719 -0.02324 -0.05365
Insect5   0.58685 -0.04300  0.00319
Insect6  -0.19617 -0.26190 -0.16507
Insect7  -0.04153  0.29123 -0.21071
Insect8  -0.31873 -0.08679  0.02736
Insect9  -0.37396  0.03284  0.04314
Insect10 -0.06059 -0.12843 -0.03415
```

"Centroid" scaling for insects only >

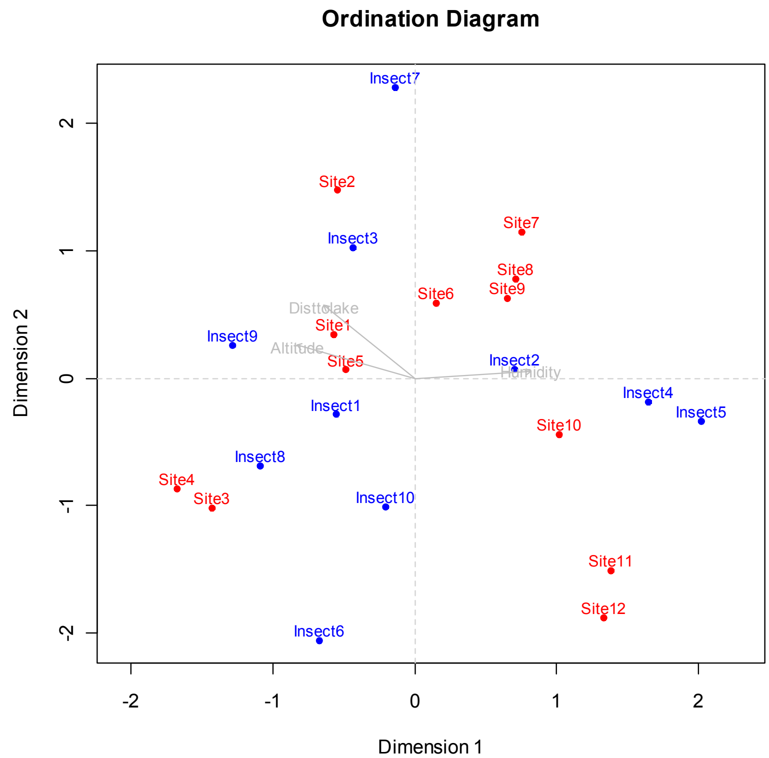
```
> scores(CCAv,scaling=2)
$species
                CCA1         CCA2
Insect1  -0.16288334  0.035959460
Insect2   0.20290869 -0.009052712
Insect3  -0.12842128 -0.130582556
Insect4   0.47719026  0.023240388
Insect5   0.58685404  0.042997519
Insect6  -0.19617384  0.261897516
Insect7  -0.04153188 -0.291228144
Insect8  -0.31873124  0.086790547
Insect9  -0.37395536 -0.032840211
Insect10 -0.06058901  0.128429756

$sites
             CCA1        CCA2
Site1  -1.0952936 -1.17259404
Site2  -0.9718214 -1.86473528
Site3  -1.5729001  0.97746564
Site4  -1.2720982  1.55085330
Site5  -0.9534102 -0.99706935
Site6   1.6398233 -0.03277022
Site7   0.9934898 -1.18735494
Site8   0.5787477 -0.83023906
Site9   0.3834509  0.88558707
Site10  1.6224521 -0.41982610
Site11  0.8732493  1.36562650
Site12  1.0872237  1.97068848
```

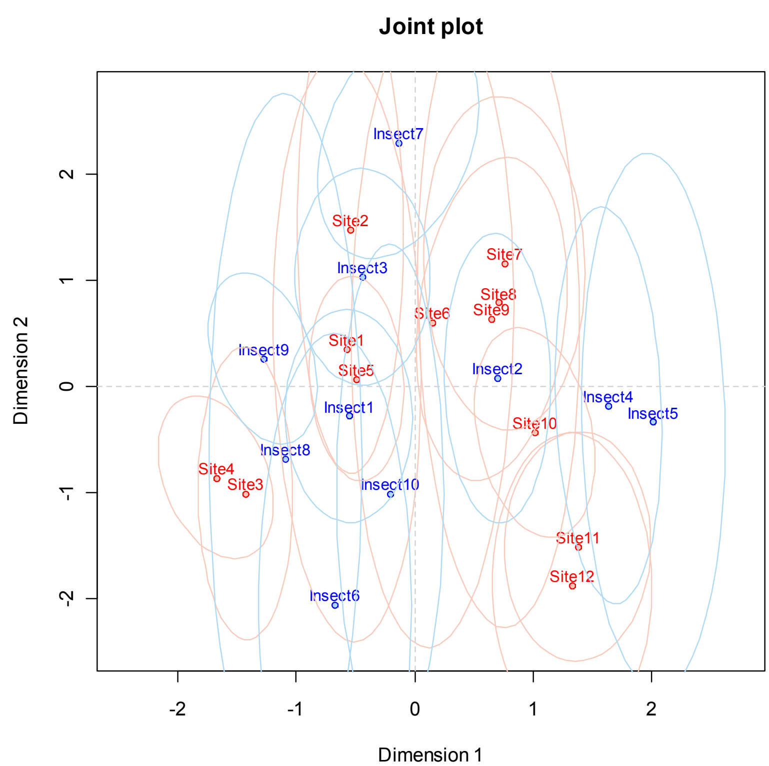
CCA Plots:

```
#PLOTTING WITH {anacor}:
CCAas=anacor(X,row.covariates=Y,ndim=2,scaling=c("standard","standard"))
plot(CCAas,plot.type="orddiag")
```

"Standard" scaling:

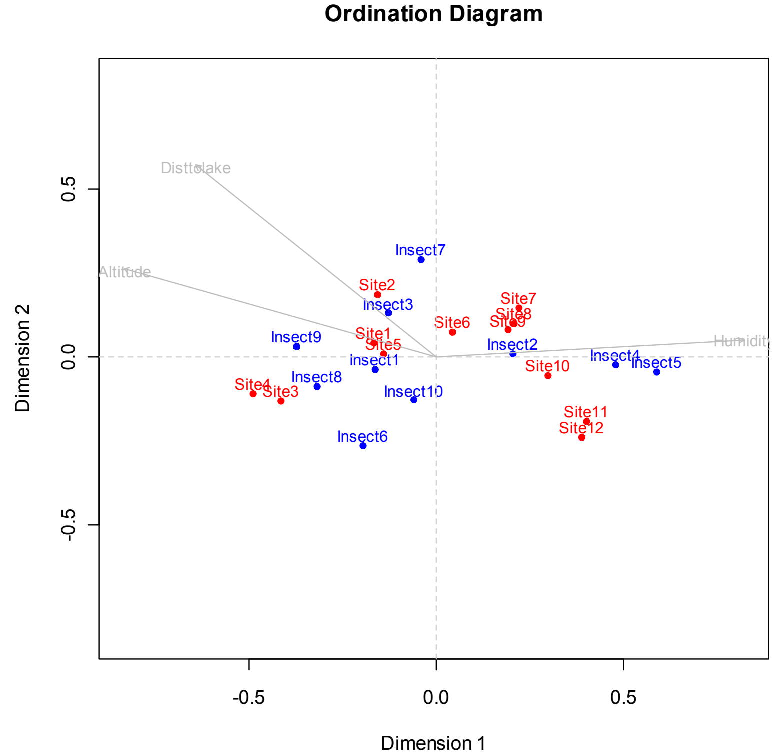


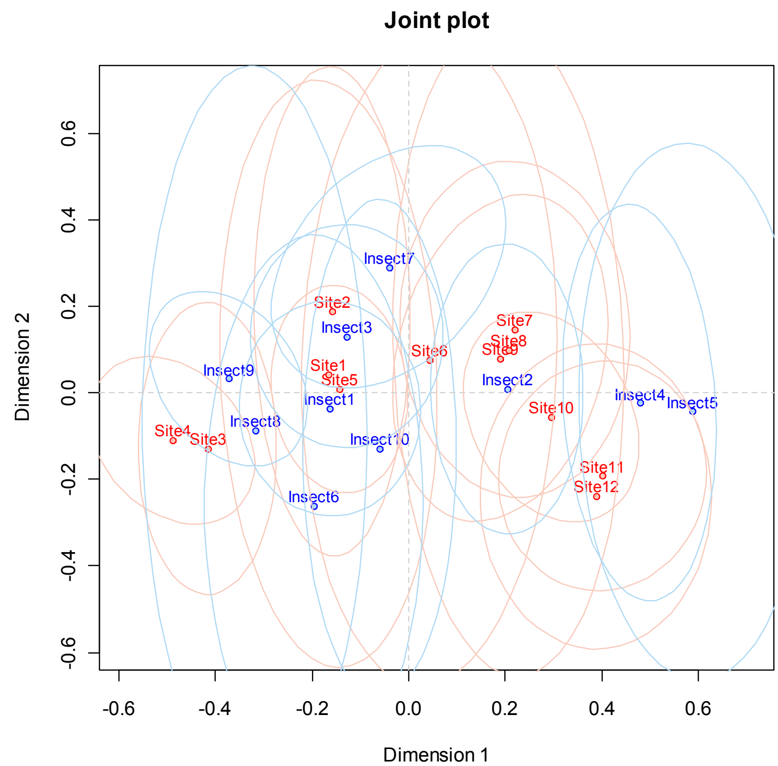

```
plot(CCAas)
```



```
CCAac=anacor(X,row.covariates=Y,ndim=2,scaling=c("centroid","centroid"))
plot(CCAac,plot.type="orddiag")
plot(CCAac)
```

"Centroid" Scaling:





```
#PLOTTING WITH {vegan}:
plot(CCAv,scaling=1)
plot(CCAv,scaling=2)
plot(CCAv,scaling=3)
```