

Ejercicio 8

Ejercicio 8: El archivo LACAR contiene la caracterización de 29 sitios de la cuenca binacional Lacar - Hua Hum (San Martín de los Andes, Provincia de Neuquén) en función de la abundancia (en frecuencia) de 36 familias macroinvertebrados que sirven como bioindicadores de contaminación orgánica. Por otra parte, se registraron 9 variables fisicoquímicas in situ y de hábitat, para determinar la calidad ambiental del lugar.

- a) Realizar un Análisis de Correspondencia para caracterizar los sitios en función de la abundancia de Lacar, macroinvertebrados y de Componentes Principales para ver la misma caracterización en función de variables ambientales. A través de la configuración de ambas caracterizaciones puede ver algún tipo de asociación entre las familias de macroinvertebrados y las variables ambientales ?

A partir de una análisis de componentes principales sobre las variables ambientales se pueden establecer cuatro grupos de clusters, fuertemente asociados con la primera dimensión que absorbe más de a mitad de la varianza.

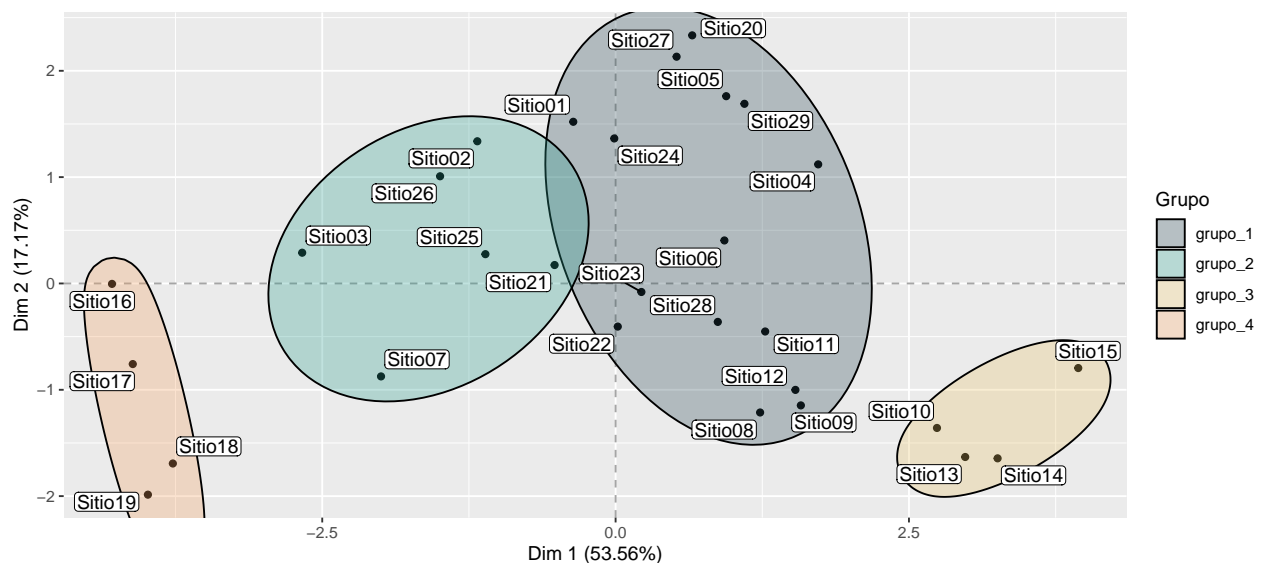


Figure 1: PCA de Var. Ambientales

Mediante un análisis de correspondencia sobre las variables de abundancia, se pueden obtener dos grupos bien diferenciados de sitios, siendo suficiente el primer componente para explicar la fuente diferenciación. El grupo de la izquierda está conformado por el grupo 1 y 3 descubierto a partir de las variables ambientales y el de la derecha posee los grupos 2, 3 y 4.

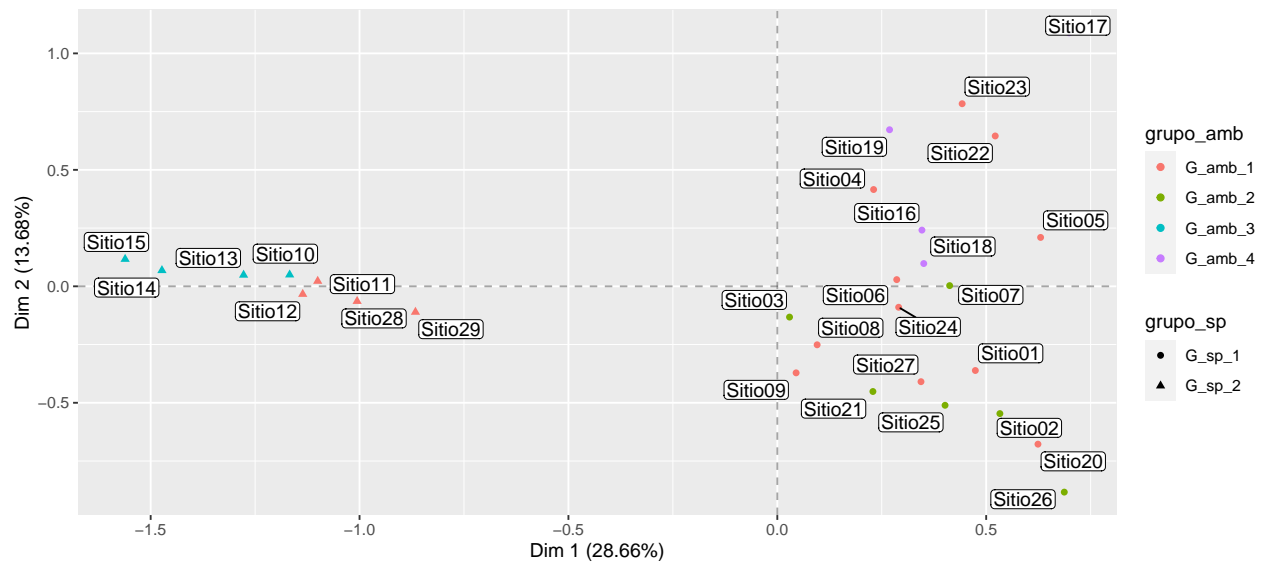


Figure 2: AC de Abundancia

Al analizar las variables de frecuencias de macroinvertebrados, es fácil ver que un grupo específicas de ellas están fuertemente relacionadas con los grupos de sitios de Abundancia anteriormente descritos.

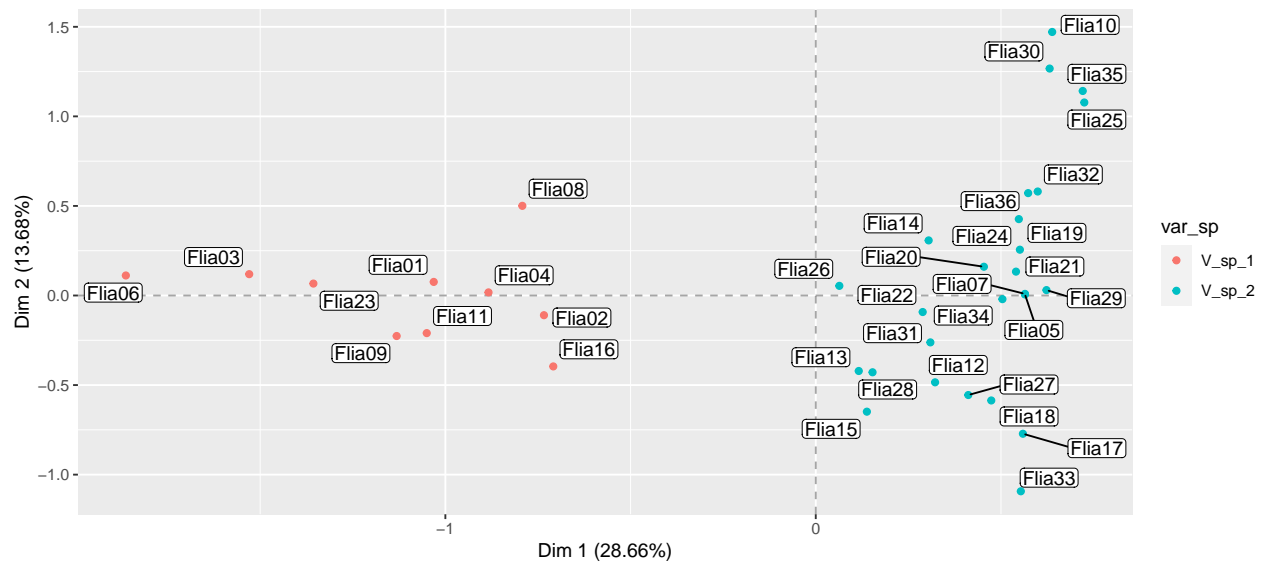


Figure 3: AC de Abundancia

b) Cuantificar la relación entre los planos principales de las dos configuraciones halladas en (a)

La correlación entre ambas distancias es

```
cor(dist_pca_x, dist_ca_y)
```

```
## [1] 0.3869169
```

Aunque dicha correlación no es despreciable, distan de ser equivalentes como se observa en el gráfico:

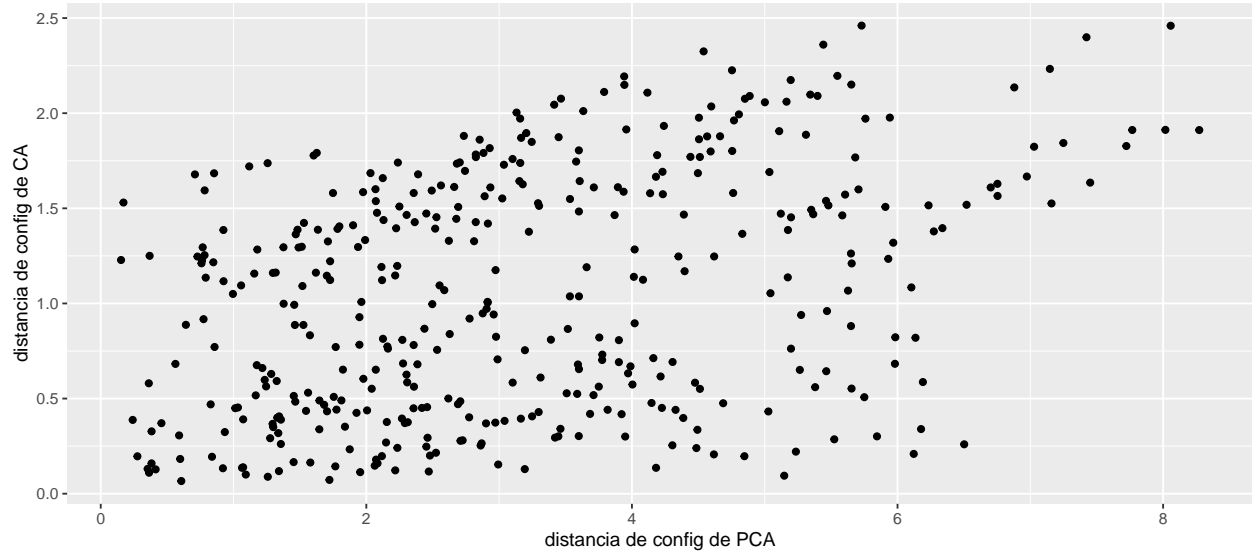


Figure 4: Relación entre los planos principales

- c) Recurrir a AFM para caracterizar brevemente los sitios en función de la abundancia de macroinvertebrados y gradientes de las variables ambientales

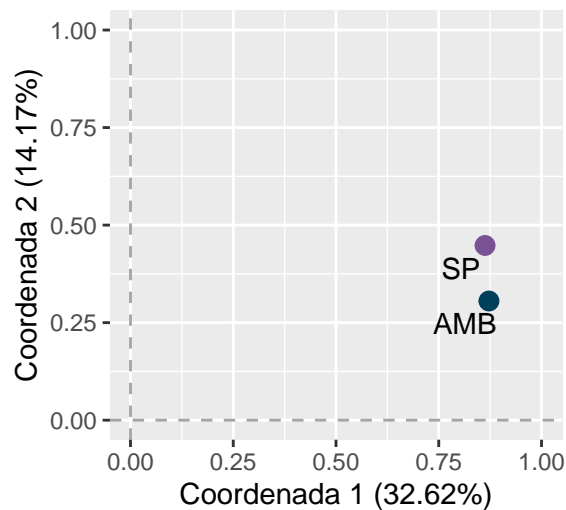
Al analizar los datos con MFA, lo primero que se nota es que la relación entre las variables ambientales y las variables de abundancia poseen una relación mayor que .5 en términos de RV.

Table 1: Coeficiente RV

	SP	AMB	MFA
SP	1.000	0.537	0.897
AMB	0.537	1.000	0.854
MFA	0.897	0.854	1.000

Ambos grupos de variables están fuertemente relacionadas con la primer dimensión la cual explica un 32.6% y un tanto menos con la segunda, la cual explica un 14.7%

Representacion de los Grupos



Como puede observarse el bajo valor explicativo de las dimensiones latentes tiene que ver con la poca congruencia entre las dimensiones individuales de abundancia y ambiente.

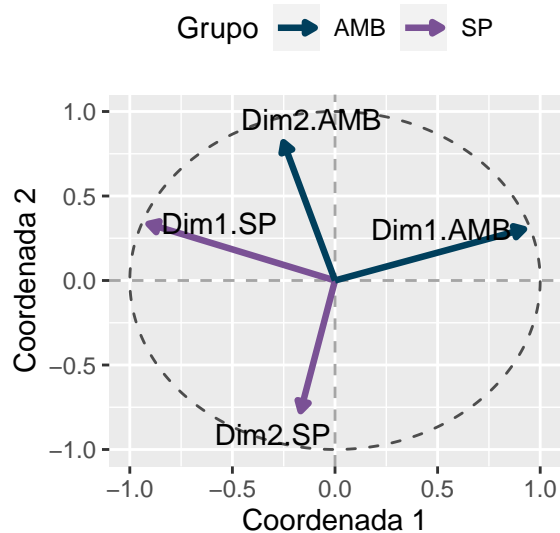


Figure 5: Ejes Parciales

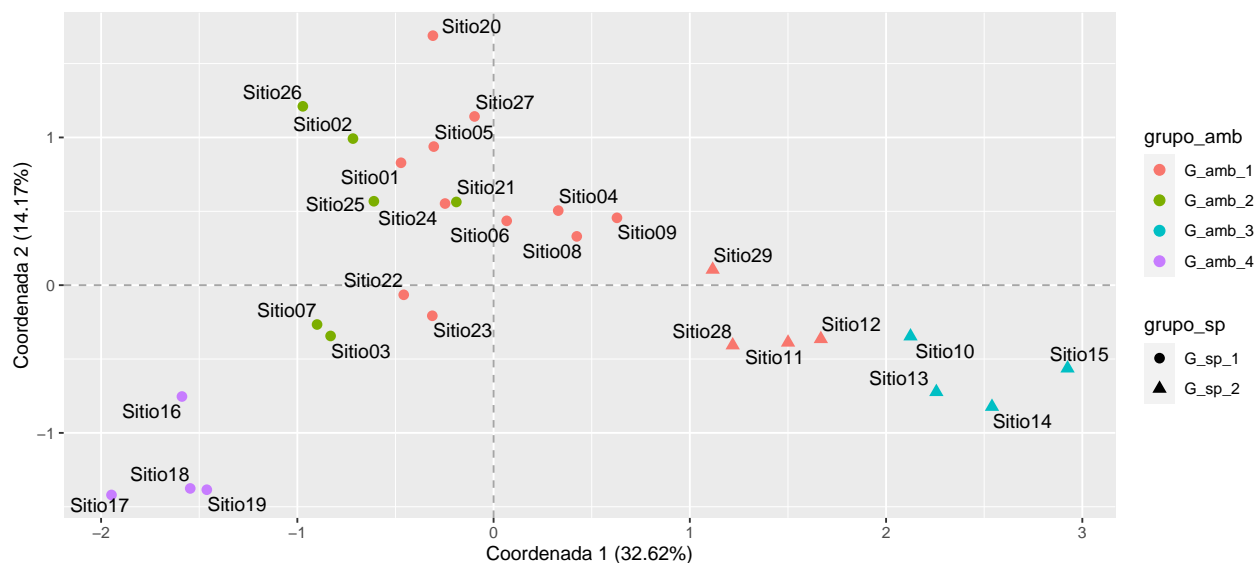


Figure 6: Configuración de Concenso

A pesar de no absorber mucha variabilidad, la primera dimensión alcanza para explicar los resultados observados en el punto a) en términos de abundancia y condiciones ambientales de los sitios (y por tanto la diferencia de frecuencias). En dicha dimensión se pueden separar los 4 grupos de sitios ambientales al mismo tiempo que extraer los dos grupos de abundancia.

- d) Con Análisis de Redundancia profundizar sobre el estudio de asociación entre abundancia de familias de macroinvertebrados y variables ambientales. Relacionar (si es posible) con los grupos de sitios que venía estableciendo.

```
## Call: rda(formula = stdY ~ CE + PH + ALT + PRS + FE + SI + DU + ALC +
```

```
## PEH, data = stdX)
##
##              Inertia Proportion Rank
## Total          0.3743      1.0000
## Constrained    0.2063      0.5512    9
## Unconstrained  0.1680      0.4488   19
## Inertia is variance
##
## Eigenvalues for constrained axes:
##   RDA1   RDA2   RDA3   RDA4   RDA5   RDA6   RDA7   RDA8   RDA9
## 0.13013 0.02661 0.01420 0.01137 0.00952 0.00539 0.00404 0.00300 0.00207
##
## Eigenvalues for unconstrained axes:
##   PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8
## 0.04772 0.02617 0.01760 0.01634 0.01449 0.00876 0.00792 0.00597
## (Showing 8 of 19 unconstrained eigenvalues)
## $r.squared
## [1] 0.5511786
##
## $adj.r.squared
## [1] 0.338579
```

El porcentaje de varianza explicado de las variables de abundancia por la variables ambientales es del 55% lo cual coincide contra el coeficiente RV anteriormente descripto. Si se ajusta el coeficiente, la relación baja a 33%.

```
## Permutation test for rda under reduced model
## Permutation: free
## Number of permutations: 1000
##
## Model: rda(formula = stdY ~ CE + PH + ALT + PRS + FE + SI + DU + ALC + PEH, data = stdX)
##           Df Variance      F    Pr(>F)
## Model      9  0.20632 2.5926 0.000999 ***
## Residual  19  0.16801
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Si se realiza un anova basado en test de permutaciones, se ve que el modelo es significativo.

Si en cambio se testea que variable dentro de las ambientales en el modelo restringido son significativas, el resultado es que solo una de ellas lo es: PRS: fósforo reactivo soluble. Se puede suponer que PRS tiene una alta correlación para con el resto de las variables relevantes y basta con ella sola para explicar las abundancias.

```
## Permutation test for rda under reduced model
## Marginal effects of terms
## Permutation: free
```

```

## Number of permutations: 1000
##
## Model: rda(formula = stdY ~ CE + PH + ALT + PRS + FE + SI + DU + ALC + PEH, data = stdX)
##           Df Variance      F Pr(>F)
## CE          1 0.011095 1.2547 0.22777
## PH          1 0.009211 1.0416 0.32767
## ALT         1 0.010249 1.1591 0.24575
## PRS         1 0.024760 2.8001 0.02498 *
## FE          1 0.007745 0.8759 0.44755
## SI          1 0.009718 1.0990 0.31169
## DU          1 0.012664 1.4322 0.16484
## ALC         1 0.013667 1.5456 0.14386
## PEH         1 0.012544 1.4186 0.15884
## Residual 19 0.168007
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

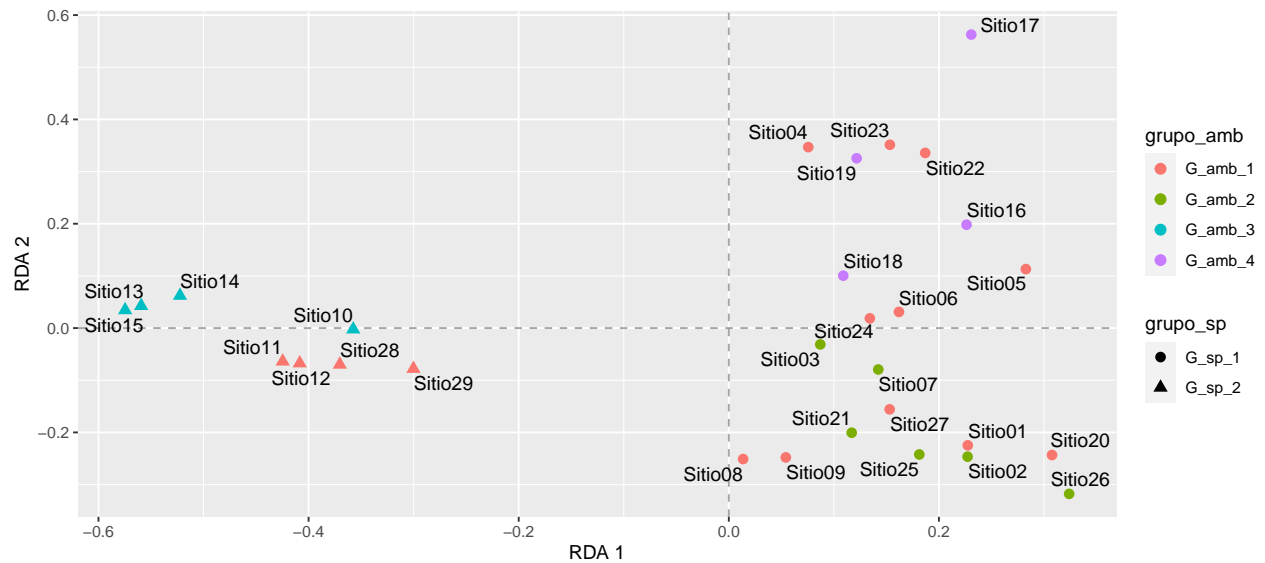
Finalmente, se ve que con una sola dimensión es posible explicar los datos, siendo solo una de ellas significativas.

```

## Permutation test for rda under reduced model
## Forward tests for axes
## Permutation: free
## Number of permutations: 1000
##
## Model: rda(formula = stdY ~ CE + PH + ALT + PRS + FE + SI + DU + ALC + PEH, data = stdX)
##           Df Variance      F Pr(>F)
## RDA1        1 0.130129 14.7164 0.000999 ***
## RDA2        1 0.026607  3.0090 0.264735
## RDA3        1 0.014204  1.6063 0.928072
## RDA4        1 0.011366  1.2853 0.985015
## RDA5        1 0.009520  1.0766 0.988012
## RDA6        1 0.005389  0.6094 1.000000
## RDA7        1 0.004039  0.4567 1.000000
## RDA8        1 0.003003  0.3396 0.999001
## RDA9        1 0.002068  0.2338 0.998002
## Residual 19 0.168007
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Respecto a los grupos, se observa que la configuración resultante expresa muy bien la diferencia que existe entre los grupos de abundancias en los sitios. Sin embargo, se pierde calidad de la representación de las variables ambientales; a diferencia de del AFM, ya no es posible distinguir los cuatro grupos con claridad.



e) Y en cuanto a la subcuenca a la que pertenecen los sitios, encuentra algún patrón ?

En base a la configuración hallada en el análisis de RDA, no pareciera haber relación alguna para la subcuenca.

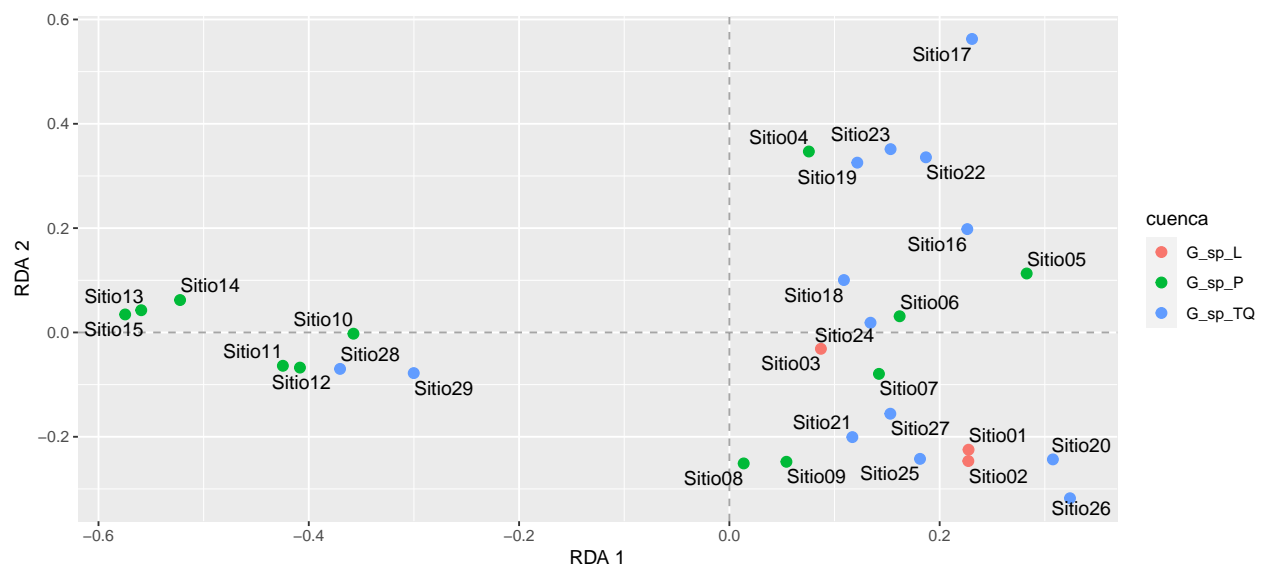


Figure 7: Impacto de la Subcuenca

f) Verificar, en base a la longitud del gradiente implícito en Y, si fue adecuado usar RDA

El eje 1 tiene una longitud de 2.6 desvíos estándares por lo cual, de a cuerdo lo visto en clase, al estar debajo de 3 un RDA es aceptable.

```
##
## Call:
## decorana(veg = Y)
##
## Detrended correspondence analysis with 26 segments.
## Rescaling of axes with 4 iterations.
##
##           DCA1    DCA2    DCA3    DCA4
```

```
## Eigenvalues      0.5298 0.2505 0.1371 0.08308
## Decorana values 0.5334 0.2426 0.1183 0.07009
## Axis lengths    2.6124 1.9641 1.7352 1.43967
```

g) Repita lo realizado en (d) recurriendo a Análisis Canónico de Correspondencia y compare resultados

```
## Call: cca(formula = Y ~ CE + PH + ALT + PRS + FE + SI + DU + ALC + PEH,
## data = X)
##
##              Inertia Proportion Rank
## Total          1.8613      1.0000
## Constrained    0.9252      0.4971    9
## Unconstrained  0.9361      0.5029   19
## Inertia is scaled Chi-square
##
## Eigenvalues for constrained axes:
##   CCA1   CCA2   CCA3   CCA4   CCA5   CCA6   CCA7   CCA8   CCA9
## 0.4400 0.1741 0.0966 0.0726 0.0528 0.0332 0.0251 0.0187 0.0121
##
## Eigenvalues for unconstrained axes:
##   CA1    CA2    CA3    CA4    CA5    CA6    CA7    CA8
## 0.19554 0.13389 0.11277 0.09666 0.08861 0.06391 0.05747 0.04902
## (Showing 8 of 19 unconstrained eigenvalues)
## $r.squared
## [1] 0.4970576
##
## $adj.r.squared
## [1] 0.2591736
```

El porcentaje de varianza explicado de las variables de abundancia por la variables ambientales es un poco menor que el RDA, un 50%. Si se ajusta el coeficiente, la relación baja más que con RDA a 26%.

```
## Permutation test for cca under reduced model
## Permutation: free
## Number of permutations: 1000
##
## Model: cca(formula = Y ~ CE + PH + ALT + PRS + FE + SI + DU + ALC + PEH, data = X)
##           Df ChiSquare      F    Pr(>F)
## Model      9   0.92515 2.0864 0.000999 ***
## Residual  19   0.93611
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Si se realiza un anova basado en test de permutaciones, se ve que el modelo de CCA también es significativo.

Si en cambio se testea que variable dentro de las ambientales en el modelo restringido son significativas, el resultado una de ellas lo es al 95%: PRS: fósforo reactivo soluble al igual que el modelo de RDA.


```

## Permutation test for cca under reduced model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 1000
##
## Model: cca(formula = Y ~ CE + PH + ALT + PRS + FE + SI + DU + ALC + PEH, data = X)
##           Df ChiSquare      F Pr(>F)
## CE         1    0.06237 1.2659 0.24176
## PH         1    0.04762 0.9665 0.44855
## ALT        1    0.06752 1.3705 0.17782
## PRS        1    0.09369 1.9017 0.04695 *
## FE         1    0.04260 0.8647 0.54046
## SI         1    0.05478 1.1118 0.33167
## DU         1    0.07267 1.4749 0.12887
## ALC        1    0.06404 1.2998 0.19980
## PEH        1    0.07085 1.4381 0.14386
## Residual 19    0.93611
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

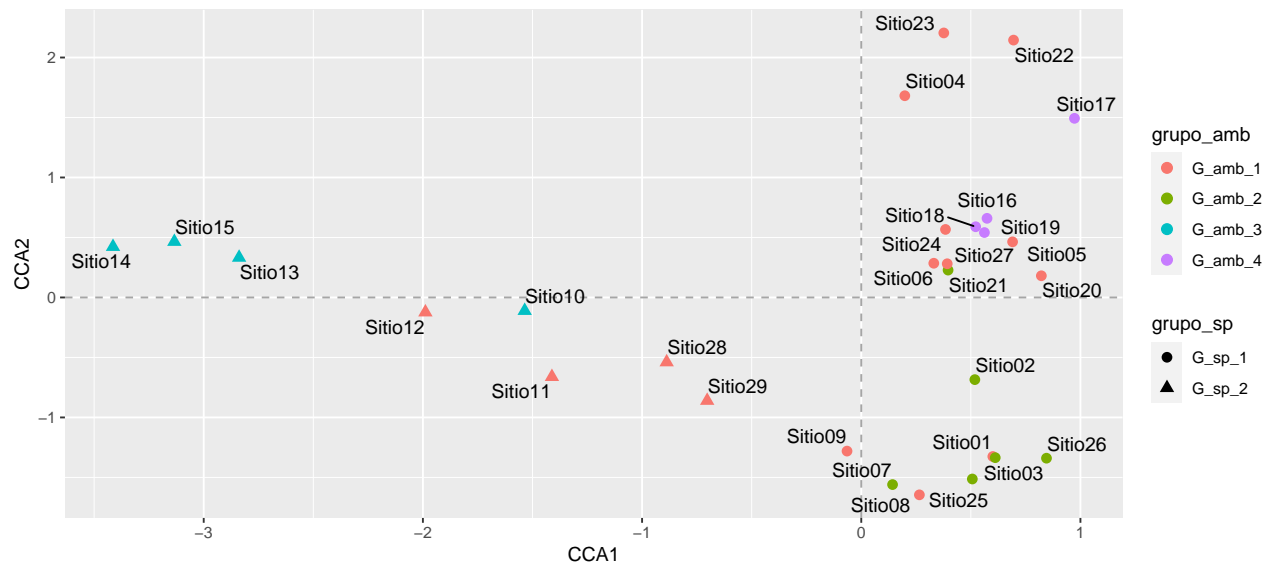
```

Finalmente, la primera dimensión es muy significativa y a diferencia del RDA, la segunda dimensión el significativa al 95%.

```

## Permutation test for cca under reduced model
## Forward tests for axes
## Permutation: free
## Number of permutations: 1000
##
## Model: cca(formula = Y ~ CE + PH + ALT + PRS + FE + SI + DU + ALC + PEH, data = X)
##           Df ChiSquare      F Pr(>F)
## CCA1        1    0.44000 8.9307 0.000999 ***
## CCA2        1    0.17409 3.5334 0.016983 *
## CCA3        1    0.09655 1.9597 0.700300
## CCA4        1    0.07262 1.4739 0.919081
## CCA5        1    0.05279 1.0714 0.992008
## CCA6        1    0.03321 0.6741 1.000000
## CCA7        1    0.02510 0.5094 0.999001
## CCA8        1    0.01873 0.3801 0.998002
## CCA9        1    0.01207 0.2450 0.998002
## Residual 19    0.93611
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```



En conclusión, se puede decir que CCA presenta una peor representación de los datos tanto en términos visuales como en términos numéricos.