

Ejercicio 7

- A) Cuantifique la concordancia de la configuración de poblaciones nativas de maíz en ambos ambientes en el espacio original mediante el coeficiente de correlación de Pearson entre matrices de distancias euclídeas estandarizadas entre individuos, y utilizando el coeficiente Rv.

Lo primero que hacemos es calcular las matrices de distancia entre las poblaciones de maíz a partir de las variables estandarizadas, utilizando la distancia euclídea.

```
dist_m1 <- dist(scale(datos_m1), method = "euclidean")
dist_m2 <- dist(scale(datos_m2), method = "euclidean")
correlacion <- cor(dist_m1, dist_m2)
```

La correlacion entre las matrices de distancia es igual a 0.676. Esto indica que existe una concordancia media-alta entre la configuracion de las poblaciones de maíz en ambos ambientes.

Por otro lado, tambien calculamos el coeficiente RV.

```
coef_rv <- coeffRV(scale(datos_m1), scale(datos_m2))
```

que resulta 0.567.

Mientras que la correlacion entre las matrices de distancia mide la similaridad entre las posiciones relativas de las poblaciones de maíz, en terminos de las variables medidas, el coeficiente RV mide directamente la correlacion entre los valores de estas variables para ambos ambientes.

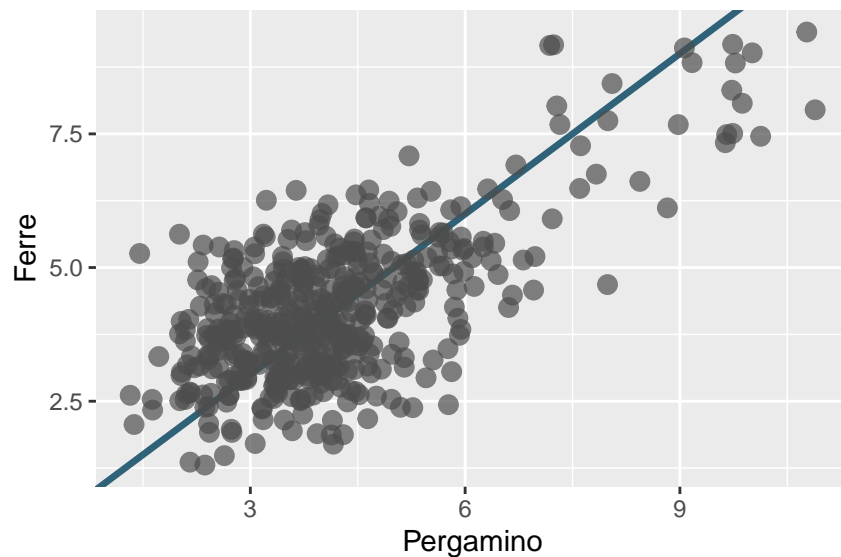


Figure 1: Distancia entre poblaciones para los ambientes Pergamino y Ferre. La linea azul representa a la recta identidad.

- B) Realice un ACP para cada ambiente, compare semejanzas y diferencias entre ambas caracterizaciones tanto para individuos como para variables.

Utilizamos la sentencia `PCA()` de la libreria **FactoMineR**.

```
acp_m1 <- PCA(datos_m1, ncp = 2, graph = FALSE)
acp_m2 <- PCA(datos_m2, ncp = 2, graph = FALSE)
```

Y luego obtenemos los graficos para los individuos y para las variables en el plano principal.

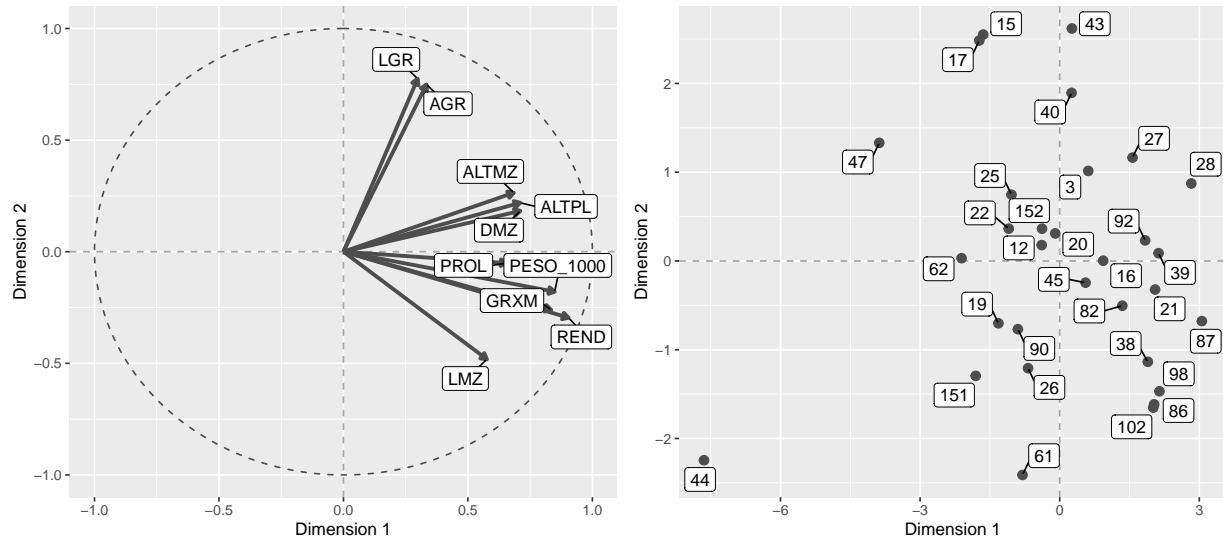


Figure 2: Caracterizacion de los individuos y las variables en el plano principal del ACP para el ambiente Pergamino. 64% variabilidad explicada.

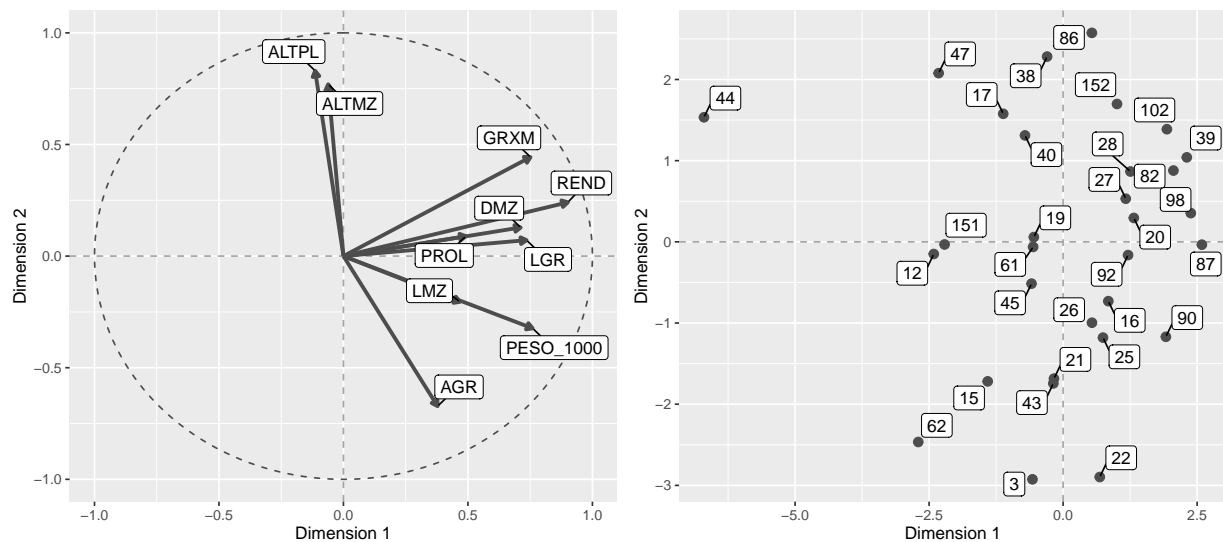


Figure 3: Caracterizacion de los individuos y las variables en el plano principal del ACP para el ambiente Ferre. 57% variabilidad explicada.

Respecto al grafico de las variables, vemos que en ambos ambientes se da que **rendimiento**, **peso cada 1000 granos**, y **granos por metro** son las que mas contribuyen a la primera dimension. En cambio, las variables que mas contribuyen al segundo eje dependen del ambiente. En el ambiente Pergamino se trata de las variables asociadas al grano, **largo** y **ancho**, mientras que en el ambiente Ferre se trata

de la **altura de la planta** y la **altura de la mazorca**, donde además el ancho del grano tiene una correlación negativa cercana a -0.7.

El ángulo entre los vectores asociados a las cargas nos brinda información sobre la relación entre las variables, y podemos ver si estas relaciones varían según el ambiente. Por ejemplo, en Pergamino se da que la altura de la planta y de la mazorca se relacionan positivamente con el rinde, indicando que plantas con mayor altura y mazorcas más largas se asocian a rindes mayores. Sin embargo, esta asociación no sucede en el ambiente Ferre, donde vemos por el ángulo entre los vectores, que el rendimiento de la planta no se asocia a estas variables de altura. Otro ejemplo similar ocurre con la asociación entre largo y ancho de grano, que en el ambiente Pergamino resultan altamente dependientes, mientras que en el ambiente Ferre su correlación es casi nula.

En cuanto al gráfico de los individuos, lo primero que se observa es que la población **44** presenta en ambos ambientes un comportamiento muy diferente al resto. En ambos ambientes tiene valores muy bajos en la primera componente, lo que indica que se trata de una población con un rendimiento y un peso de grano muy inferior al resto de las poblaciones. Sin embargo, en el ambiente Pergamino se corresponde con plantas de poca altura con mazorcas cortas, mientras que en el ambiente Ferre se presenta plantas altas con mazorcas altas.

C) Cuantifique la relación de las dos configuraciones en el plano principal originado por los ACP

Para cada ambiente obtenemos la matriz de distancia entre los puntos en el plano principal y calculamos la correlación entre ambas matrices.

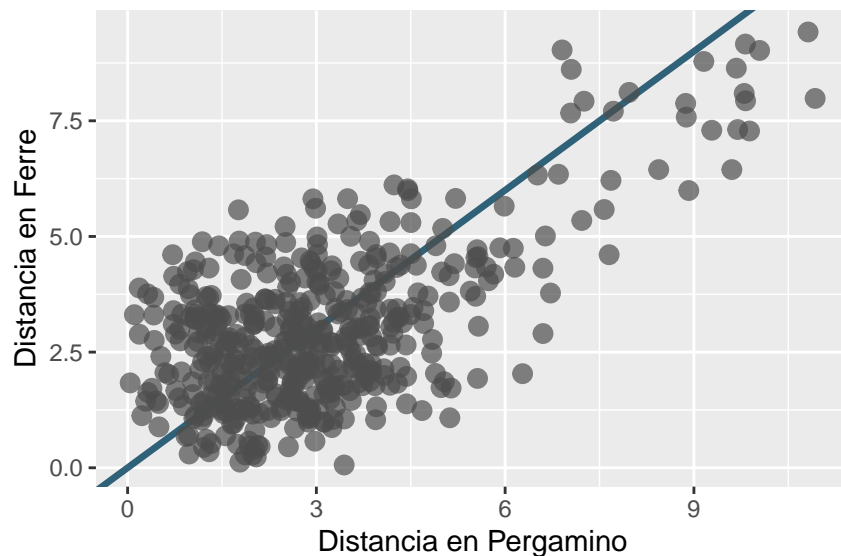


Figure 4: Distancia en Pergamino vs. distancia en Ferre para cada población de maíz.

La correlación entre ambas matrices de distancia en el plano principal es de 0.642. A partir de este valor podemos decir que hay una concordancia media o media-alta entre las configuraciones de las poblaciones en ambos ambientes. Esto significa que el comportamiento de las poblaciones de maíz es en general similar en ambos ambientes, pero también presenta características distintas en cada uno de ellos.

D) Encuentra indicios de interacción tanto genotipo-ambiente como variable-ambiente ?

Si, en ambos casos, por todo lo explicado en el inciso **b**. Como ejemplo de la interacción genotipo-ambiente tenemos a la población 44, que en el ambiente Pergamino se corresponde con plantas bajas y mazorcas cortas, mientras que en Ferre son plantas altas y mazorcas largas.

Y como ejemplo de la interaccion variable-ambiente tenemos que en Pergamino se da que la altura de la planta y de la mazorca se relacionan positivamente con el rinde, indicando que plantas con mayor altura y mazorcas mas largas se asocian a rindes mayores. Sin embargo, esta asociacion no sucede en Ferre, donde vemos que el rendimiento de la planta no se asocia a estas variables de altura.

- E) Como se quiere encontrar una caracterización ‘media’ o ‘promedio’ para las 31 poblaciones en función de la información dada en ambos ambientes proceda a realizar una ACP sobre el promedio de las variables para ambos ambientes

Obtenemos un nuevo data frame que representa al promedio entre ambos ambientes, y realizamos el ACP como lo hicimos anteriormente.

```
datos_media <- (datos_m1 + datos_m2) / 2
acp_media <- PCA(datos_media, ncp = 2, graph = FALSE)
```

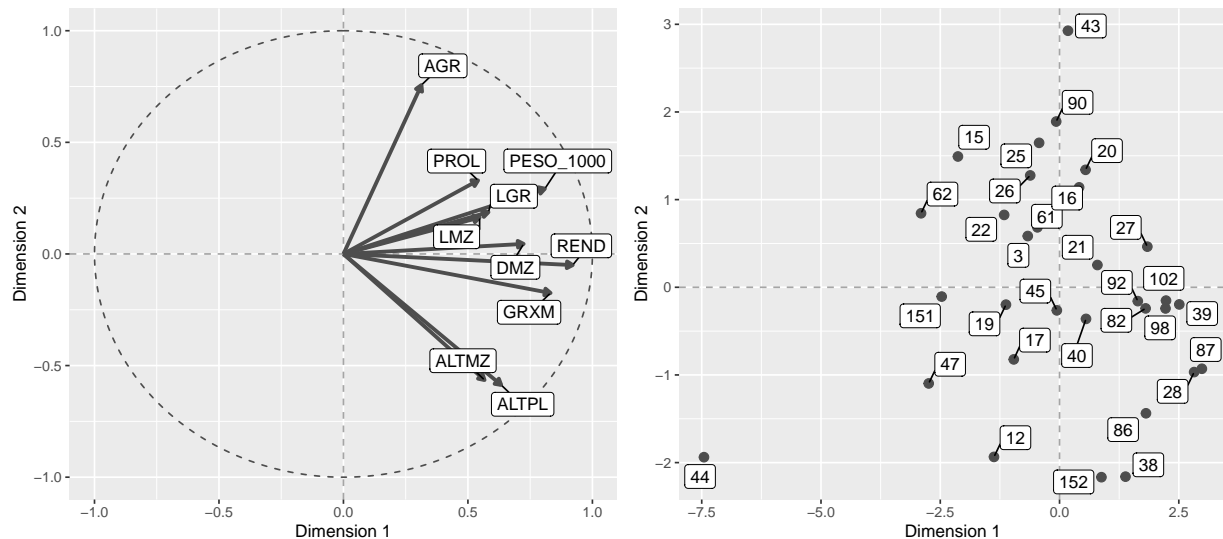


Figure 5: Caracterización de los individuos y las variables en el plano principal del ACP para el promedio de ambientes. 60% variabilidad explicada.

En el panel izquierdo de la Figura 5 vemos que el ACP realizado sobre el promedio de las variables en ambos ambientes presenta una primera coordenada que está muy correlacionada positivamente con el rendimiento, el peso cada 1000 granos, la cantidad de granos por metro, la altura de la planta, y la altura de la mazorca. Por otro lado, la segunda dimensión está relacionada, principalmente, con el ancho del grano.

Cuando el comportamiento de la población es consistente en ambos ambientes, vemos que esa información también se ve reflejada aquí. Por ejemplo, la variedad 43 presenta rendimientos medios y granos anchos y largos tanto en Ferre como Pergamino, que se corresponde con lo que se observa en este ACP basado en el promedio.

Sin embargo, observando el ángulo que se forma entre los vectores que representan a las variables comenzamos a notar ciertos problemas con esta representación. Por ejemplo, el ángulo entre ancho de grano (AGR) y altura de mazorca (ALTMZ) y altura de planta (ALTPL) es de un poco más de 90 grados, indicando casi una total independencia entre AGR y el resto. Sin embargo, este comportamiento *promedio*, no representa lo que sucede en ninguno de los dos ambientes. En el ambiente Pergamino AGR tiene una correlación media con ALTMZ y ALTPL (ángulo aproximado 45 grados) y en el ambiente Ferre tiene una correlación negativa muy fuerte (ángulo aproximado 180 grados).

F) Ahora concatene ambos archivos por filas y columnas y realice un ACP para ambas situaciones, encuentra respuesta para las interacciones planteadas en el inciso b ?

```
datos_long <- rbind(datos_m1, datos_m2)
acp_long <- PCA(datos_long, ncp = 2, graph = FALSE)
```

```
datos_wide <- cbind(datos_m1, datos_m2)
acp_wide <- PCA(datos_wide, ncp = 2, graph = FALSE)
```

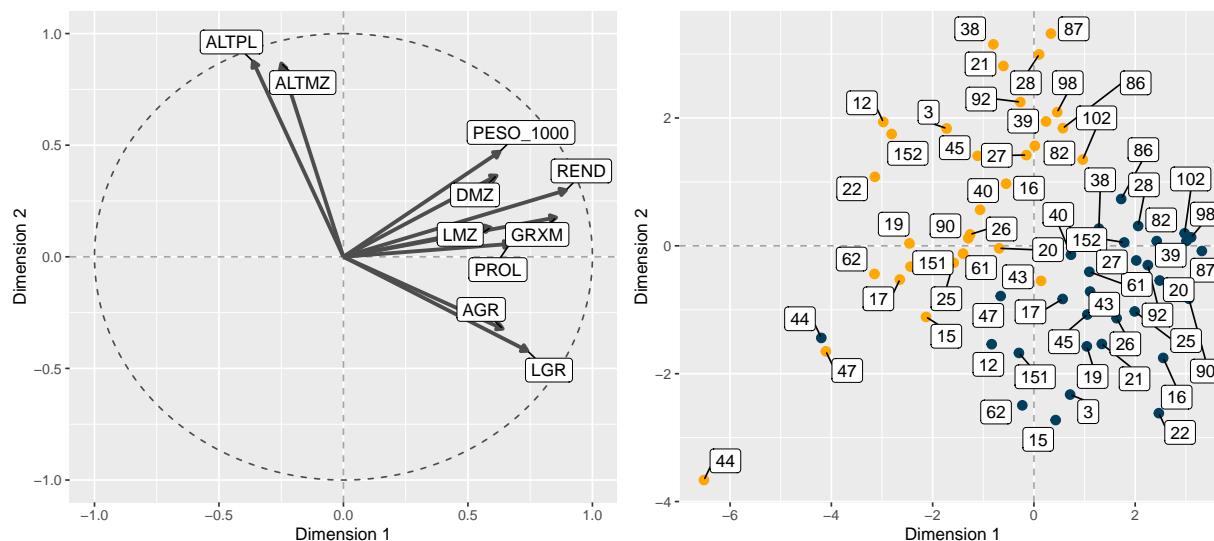


Figure 6: Caracterizacion de los individuos y las variables en el plano principal del ACP al concatenar filas . 65% variabilidad explicada. Ferre en azul, Pergamino en amarillo.

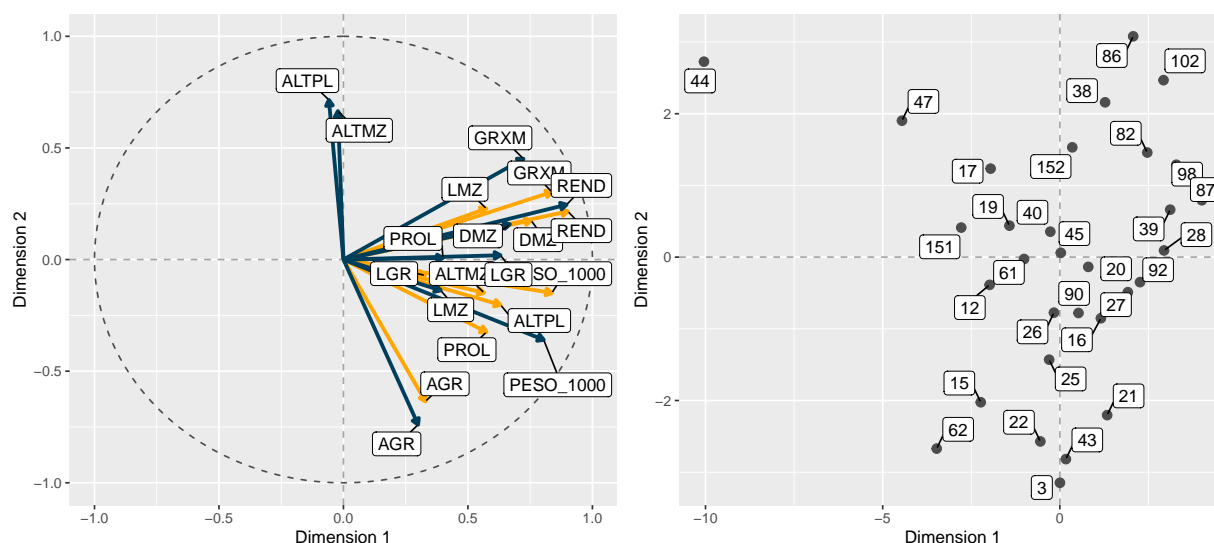


Figure 7: Caracterizacion de los individuos y las variables en el plano principal del ACP para al concatenar columnas. 51% variabilidad explicada. Ferre en azul, Pergamino en amarillo.

De la Figura 6 podemos apreciar que la primera dimension esta muy relacionada con el rinde de la

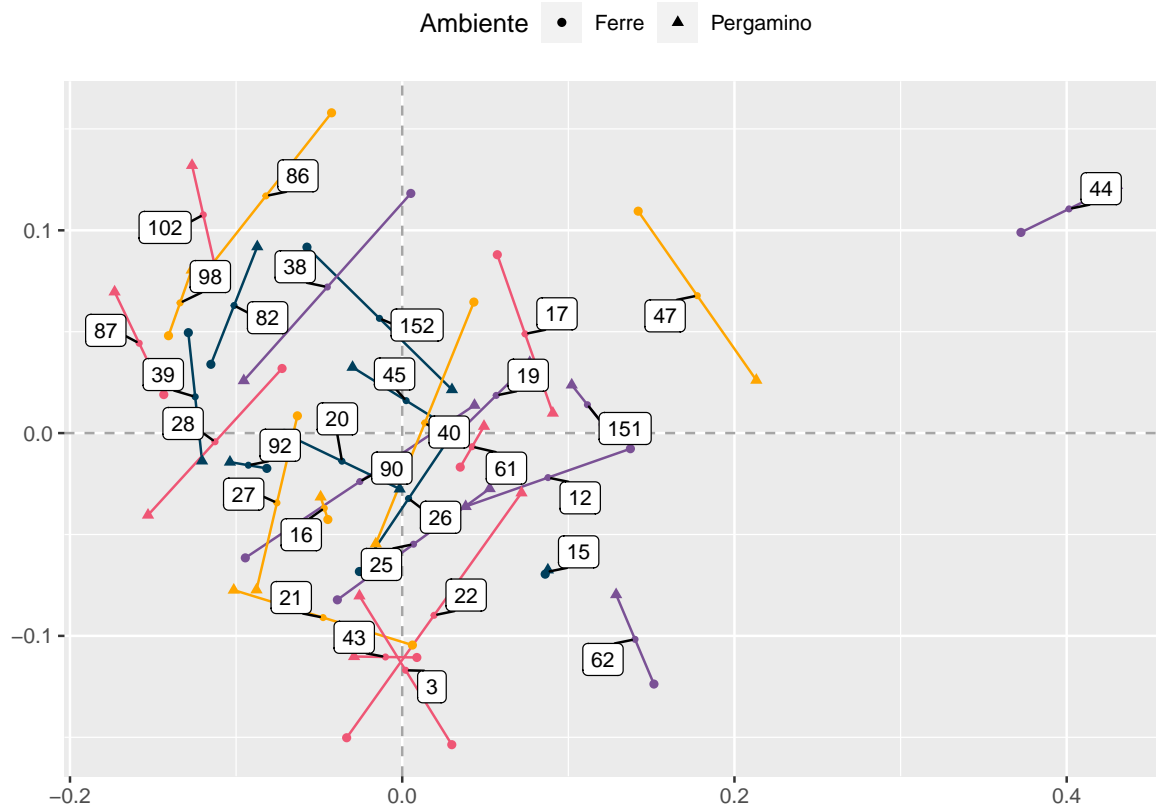
cosecha, la cantidad de granos por metro y el ancho del grano. La segunda dimension, en cambio, representa variables de altura, como son la altura de la planta y de la mazorca. Al observar el panel derecho de esta figura, podemos ver que los cultivos en Ferre presentaron en general mejores rindes, pero menores tamanos de planta, mientras que las plantas de Pergamino fueron mas grandes, pero con menor rinde. En otras palabras, en Pergamino se presentaron mejores características en terminos de la planta, pero en Ferre se presentaron mejores características en cuanto al grano y el rendimiento de la cosecha.

En el panel izquierdo de la Figura 7 podemos ver, por ejemplo, que el primer eje esta asociado al rendimiento, en ambos ambientes. Por lo tanto, las poblaciones que se encuentren hacia la derecha en el panel derecho presentaron rindes altos tanto en Pergamino como en Ferre. En cambio, el segundo eje representa a la altura de planta y mazorca solo para el ambiente Ferre. Allí podemos ver que la variedad 44 en el ambiente Ferre resulto ser de plantas altas, con mazorcas altas, pero rinde muy por debajo del promedio. Por otro lado, la variedad 102 tambien se caracterizo por su altura, pero tambien presento rindes elevados. Finalmente, notamos que las poblaciones 44 y 47 se destacan por haber tenido rindes muy pobres en ambos ambientes.

- G) Aplicar APG para tener otra visualización de la interacción genotipo-ambiente (retenga todas las dimensiones de ambas configuraciones).

```
acp_m1 <- PCA(datos_m1, ncp = 10, graph = FALSE)
acp_m2 <- PCA(datos_m2, ncp = 10, graph = FALSE)
df <- data.frame(cbind(acp_m1$ind$coord, acp_m2$ind$coord))
gpa <- GPA(df, group = c(10, 10), name.group = c("Pergamino", "Ferre"),
           graph = FALSE)
```

- H) Visualizar del gráfico correspondiente las 3 poblaciones con mayor efecto ambiente y las 3 con menor efecto



A la hora de observar el grafico hay que tener en cuenta que la escala de los ejes no es la misma. El eje horizontal va desde -0.2 a 0.45 y el eje vertical va desde -0.175 a 0.175 aproximadamente.

Las tres poblaciones con mayor efecto ambiente son 22, 90 y 38. Por otro lado, las tres poblaciones con menor efecto ambiente son la 15, 16 y 98.

Tambien podemos obtener la longitud de los segmentos y obtener a los que presentan mayor y menor efecto ambiente de manera numerica.

	Longitud del segmento		Longitud del segmento
15	0.00001	38	0.01863
16	0.00014	90	0.02470
92	0.00050	22	0.02567

I) Comparar lo encontrado en h) con el ANOVA correspondiente. Realice comentarios

Imprimimos las 3 poblaciones con mayor y menor efecto ambiente segun la suma de cuadrados residual.

	SSresidual		SSresidual
61	0.14821	17	1.10177
102	0.20504	90	1.13240
92	0.21264	22	1.30115

Alli vemos que la informacion que obtenemos en el plano de dos dimensiones no es exactamente la misma que la que se obtiene al calcular los residuos utilizando las ubicaciones de los puntos en el espacio original.

Aca podemos observar que los ambientes 22 y 90 siguen siendo los que mas efecto ambiente presentan, pero aparece el ambiente 17 en tercer lugar. Por otro lado, los ambientes 61 y 102 no aparecieron cuando buscamos los ambientes con menor efecto, y ahora figuran como los dos con menor residuo.

Estas diferencias se dan porque en el grafico estamos analizando una proyeccion de los datos en un plano de dos dimensiones, cuando en realidad se encuentran en un espacio de 10 dimensiones.