

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))
```

Este resulta 0.567, que es menor a la correlacion entre las matrices de distancia calculada anteriormente. **por que?**

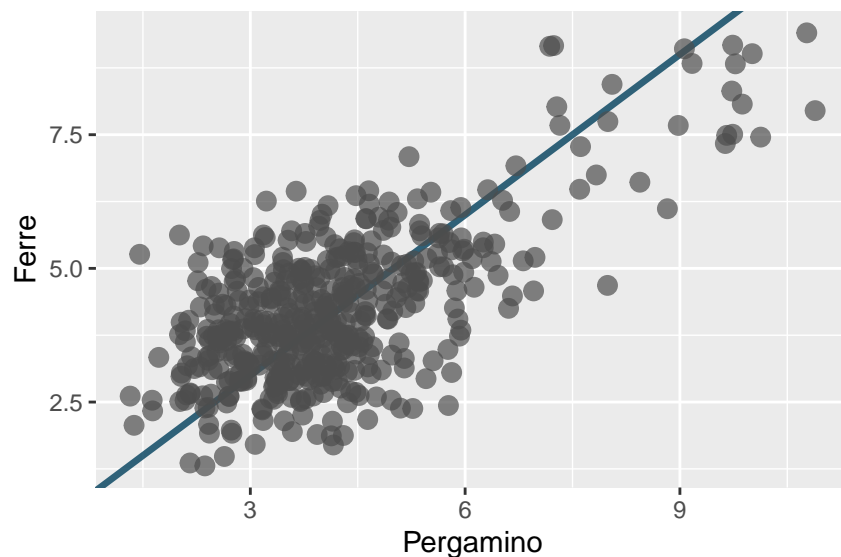


Figure 1: Distancia entre poblaciones para los ambientes Pergamino y Ferre. La linea azul representa a la recta identidad.

- 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.

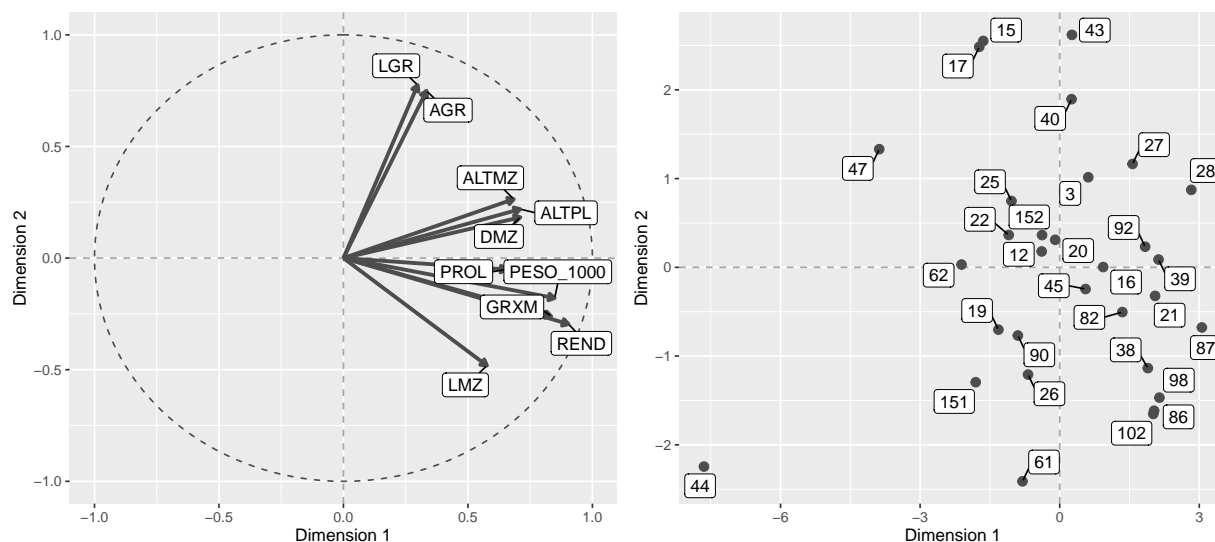


Figure 2: Caracterizacion de los individuos y las variables en el plano principal del ACP para el ambiente Pergamino. 64% variabilidad explicada.

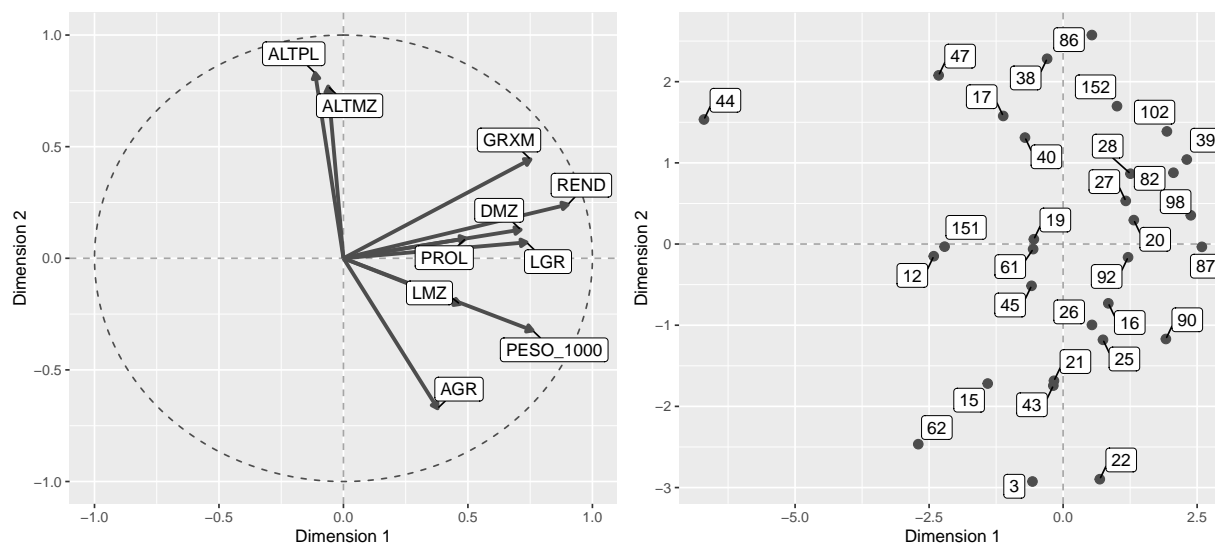


Figure 3: Caracterizacion de los individuos y las variables en el plano principal del ACP para el ambiente Ferre. 57% variabilidad explicada.

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 ademas el ancho del grano tiene una correlacion negativa cercana a -0.7.

El angulo entre los vectores asociados a las cargas nos brinda informacion sobre la relacion entre las variables, y podemos ver si estas relaciones varian segun 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 mas largas se asocian a rindes mayores. Sin embargo, esta asociacion no sucede en el ambiente Ferre, donde vemos por el angulo entre los vectores, que el rendimiento de la planta no se asocia a estas variables de altura. Otro ejemplo similar ocurre con la asociacion entre largo y ancho de grano, que en el ambiente Pergamino resultan altamente dependientes, mientras que en el ambiente Ferre su correlacion es casi nula.

En cuanto al grafico de los individuos, lo primero que se observa es que la poblacion **44** presenta en ambos ambientes un comportamiento muy diferente al resto. En ambos ambientes tiene valores muy bajos en la primer componente, lo que indica que se trata de una poblacion 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 correlacion entre ambas matrices.

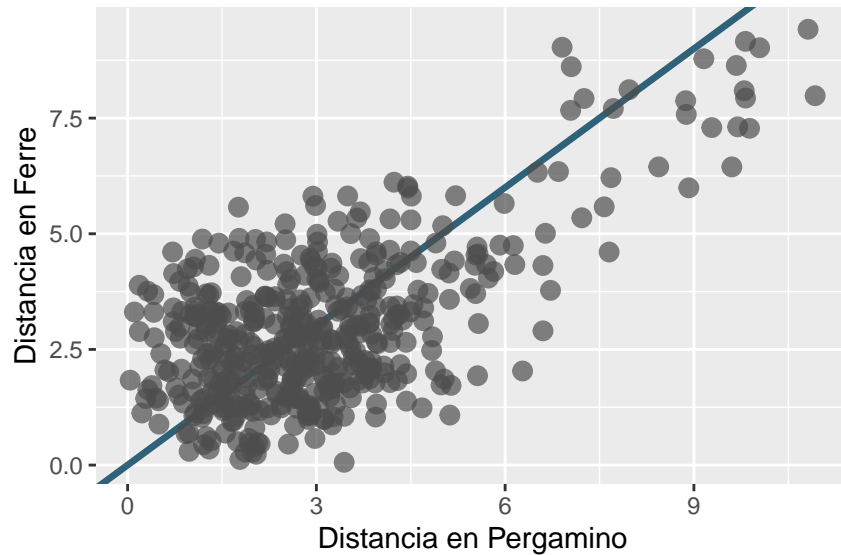


Figure 4: Distancia en Pergamino vs. distancia en Ferre para cada poblacion de maiz.

La correlacion 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 condfiguraciones de las poblaciones en ambos ambientes.

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 interaccion genotipo-ambiente tenemos a la poblacion 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)
```

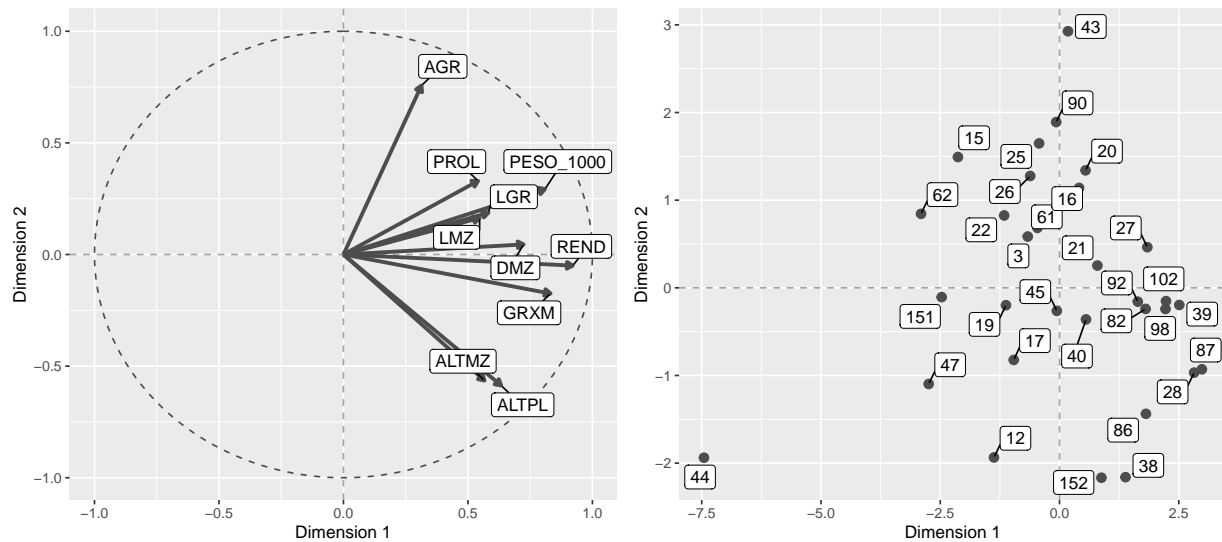


Figure 5: Caracterización de los individuos y las variables en el plano principal del ACP para el promedio de ambientes. 60% variabilidad explicada.

No estoy seguro sobre que hacer aca. Comparo con lo anterior? Repito la manera de interpretar? Me parece un embole.

- 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)
```

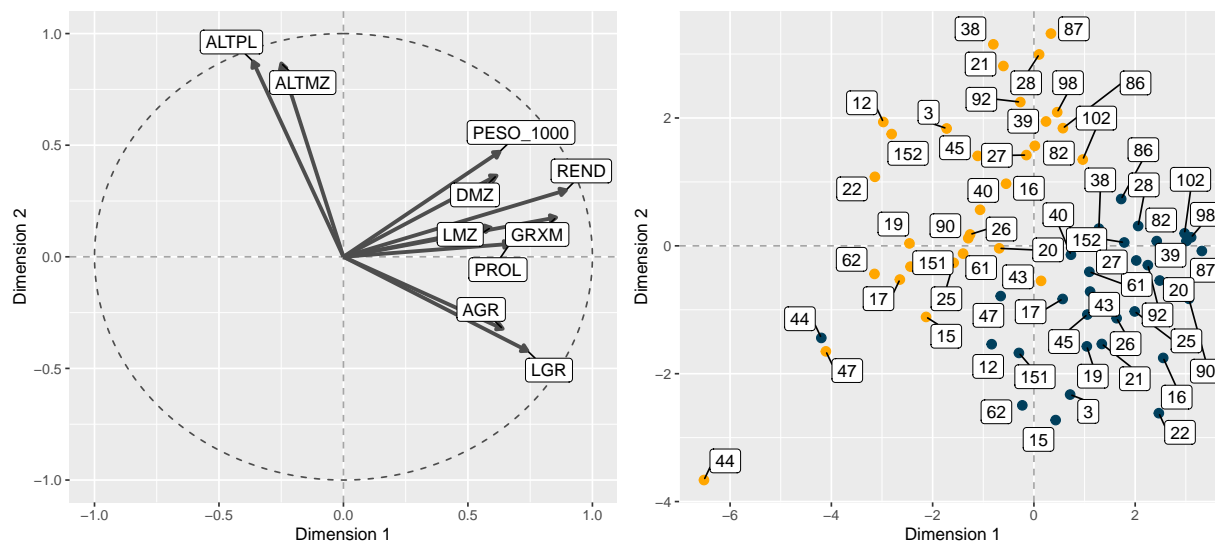


Figure 6: Caracterización de los individuos y las variables en el plano principal del ACP al concatenar filas. 65% variabilidad explicada. Ferre en amarillo, Pergamino en azul.

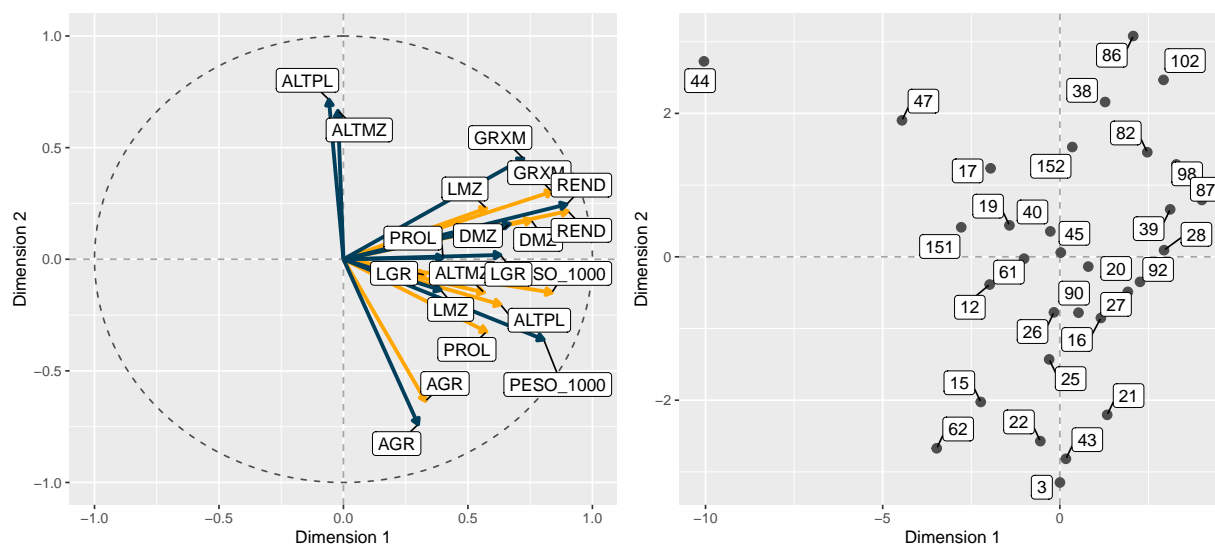


Figure 7: Caracterización de los individuos y las variables en el plano principal del ACP para al concatenar columnas. 51% variabilidad explicada. Ferre en amarillo, Pergamino en azul.

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

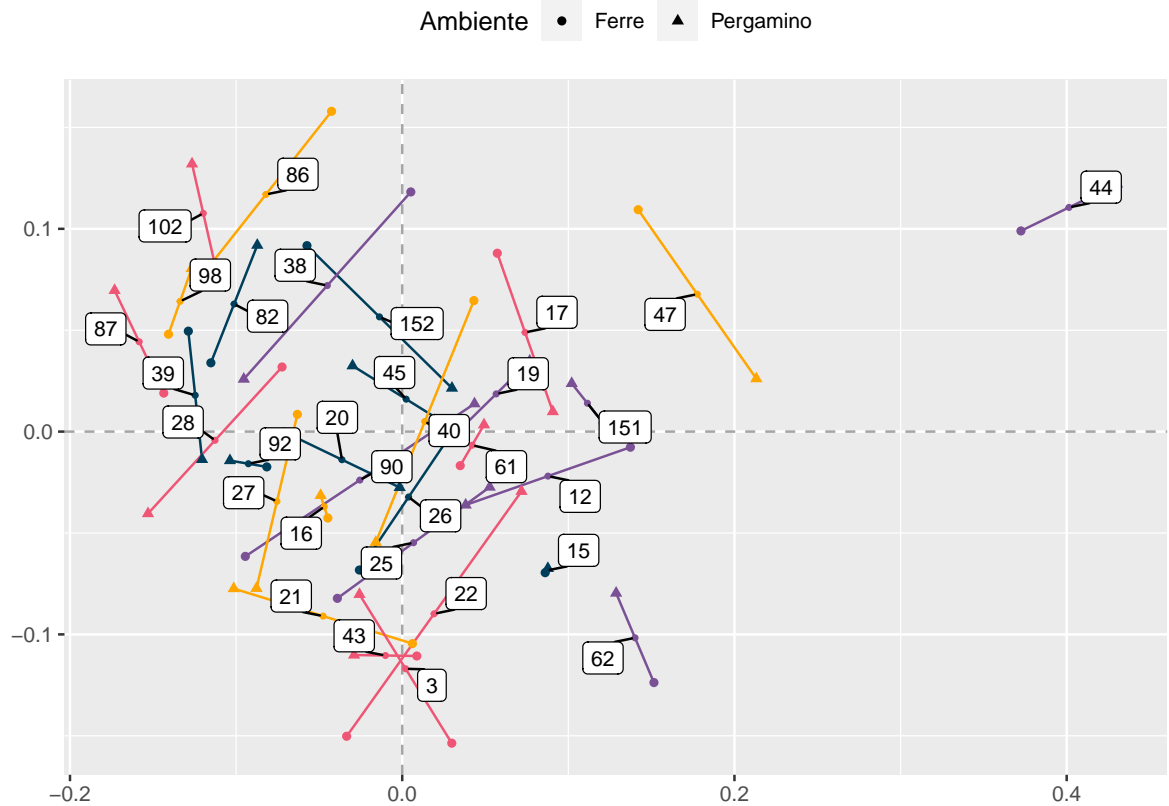
Primero obtenemos ACP para ambos ambientes, reteniendo todas las dimensiones, y creamos un data frame con las coordenadas de los individuos de ambos ambientes.

```
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 gráfico 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.

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

	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 según 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 90 y 22 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.