

Ejercicio 7

Sobre 15 poblaciones locales de maíz (*Zea mays*) de las provincias de Buenos Aires y Santa Fe, más tres variedades mejoradas de polinización abierta (VM1, VM2 y VM3 las cuales se utilizaron como testigos) y una línea endocriada (LE), se realizó una caracterización molecular por la técnica de Microsatélites. Cada población estuvo representada por 25 individuos y se registró la frecuencia alélica absoluta para 6 microsatélites. En el archivo MICROSATELITES se encuentra la información donde en las filas la misma letra indica las alternativas alélicas para cada uno de los loci microsatélites. Este archivo se encuentra con el formato que fue entregado por el propietario de los datos.

a) Adecuar el archivo MICROSATELITES para poder trabajar sobre el con la sentencia `dist.genpop`

```
X <- read.csv(
  here::here("data", "raw", "MICROSATELITES.csv"),
  row.names = 1
)

Y <- as.data.frame(t(as.matrix(X)))

colnames(Y) <- paste(
  substr(colnames(Y),1,1),
  substr(colnames(Y),2,3),
  sep="."
)
```

b) Hallar la distancia genética de Prevosti entre poblaciones con `dist.genpop`.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	VM1	VM2	VM3	LE
M1	0.00	0.17	0.40	0.33	0.37	0.21	0.43	0.39	0.47	0.54	0.37	0.41	0.29	0.31	0.56	0.59	0.45	0.49	0.83
M2	0.17	0.00	0.40	0.33	0.38	0.28	0.43	0.40	0.44	0.53	0.41	0.48	0.34	0.35	0.55	0.63	0.45	0.48	0.81
M3	0.40	0.40	0.00	0.53	0.42	0.35	0.48	0.53	0.61	0.49	0.39	0.66	0.43	0.49	0.63	0.67	0.53	0.51	0.79
M4	0.33	0.33	0.53	0.00	0.46	0.37	0.49	0.49	0.53	0.62	0.39	0.49	0.38	0.39	0.49	0.66	0.49	0.54	0.81
M5	0.37	0.38	0.42	0.46	0.00	0.28	0.40	0.45	0.47	0.45	0.43	0.50	0.35	0.45	0.58	0.57	0.45	0.46	0.79
M6	0.21	0.28	0.35	0.37	0.28	0.00	0.45	0.47	0.49	0.51	0.39	0.46	0.33	0.38	0.65	0.68	0.48	0.53	0.86
M7	0.43	0.43	0.48	0.49	0.40	0.45	0.00	0.45	0.40	0.47	0.37	0.51	0.37	0.45	0.44	0.57	0.41	0.36	0.83
M8	0.39	0.40	0.53	0.49	0.45	0.47	0.45	0.00	0.36	0.49	0.37	0.52	0.50	0.43	0.52	0.60	0.41	0.40	0.75
M9	0.47	0.44	0.61	0.53	0.47	0.49	0.40	0.36	0.00	0.43	0.50	0.56	0.44	0.41	0.49	0.63	0.38	0.39	0.80
M10	0.54	0.53	0.49	0.62	0.45	0.51	0.47	0.49	0.43	0.00	0.56	0.71	0.56	0.57	0.63	0.71	0.51	0.43	0.93
M11	0.37	0.41	0.39	0.39	0.43	0.39	0.37	0.37	0.50	0.56	0.00	0.44	0.37	0.43	0.60	0.63	0.45	0.46	0.79
M12	0.41	0.48	0.66	0.49	0.50	0.46	0.51	0.52	0.56	0.71	0.44	0.00	0.35	0.45	0.54	0.58	0.58	0.53	0.73
M13	0.29	0.34	0.43	0.38	0.35	0.33	0.37	0.50	0.44	0.56	0.37	0.35	0.00	0.36	0.53	0.59	0.49	0.47	0.79
M14	0.31	0.35	0.49	0.39	0.45	0.38	0.45	0.43	0.41	0.57	0.43	0.45	0.36	0.00	0.41	0.62	0.51	0.45	0.78
M15	0.56	0.55	0.63	0.49	0.58	0.65	0.44	0.52	0.49	0.63	0.60	0.54	0.53	0.41	0.00	0.63	0.53	0.43	0.75
VM1	0.59	0.63	0.67	0.66	0.57	0.68	0.57	0.60	0.63	0.71	0.63	0.58	0.59	0.62	0.63	0.00	0.53	0.53	0.41
VM2	0.45	0.45	0.53	0.49	0.45	0.48	0.41	0.41	0.38	0.51	0.45	0.58	0.49	0.51	0.53	0.53	0.00	0.44	0.79
VM3	0.49	0.48	0.51	0.54	0.46	0.53	0.36	0.40	0.39	0.43	0.46	0.53	0.47	0.45	0.43	0.53	0.44	0.00	0.79
LE	0.83	0.81	0.79	0.81	0.79	0.86	0.83	0.75	0.80	0.93	0.79	0.73	0.79	0.78	0.75	0.41	0.79	0.79	0.00

c) Sobre la matriz de distancia hallada aplique Análisis de Coordenadas Principales para conformar grupos de poblaciones de Maíz según su semejanza (puede ayudarse haciendo un clúster UPGMA sobre las distancias euclídeas observadas en el plano principal).

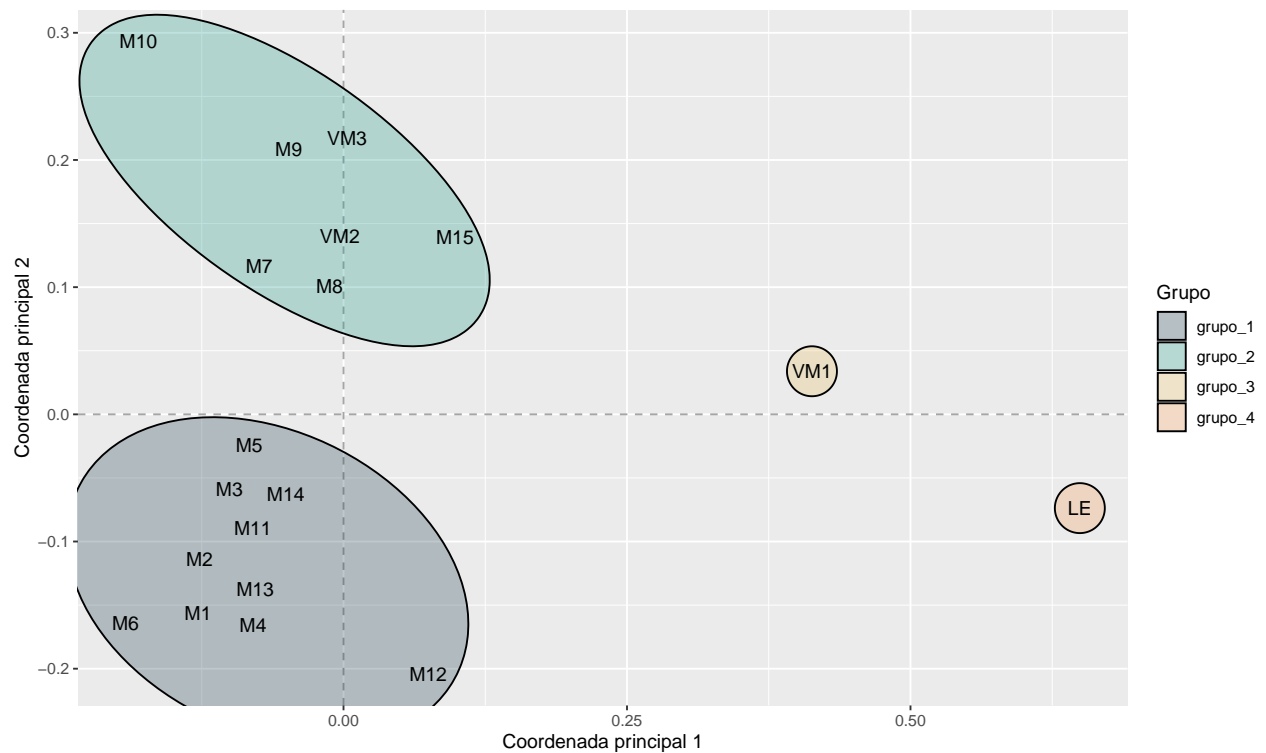


Figure 1: Caracterizacion de ls datos vis prevosti

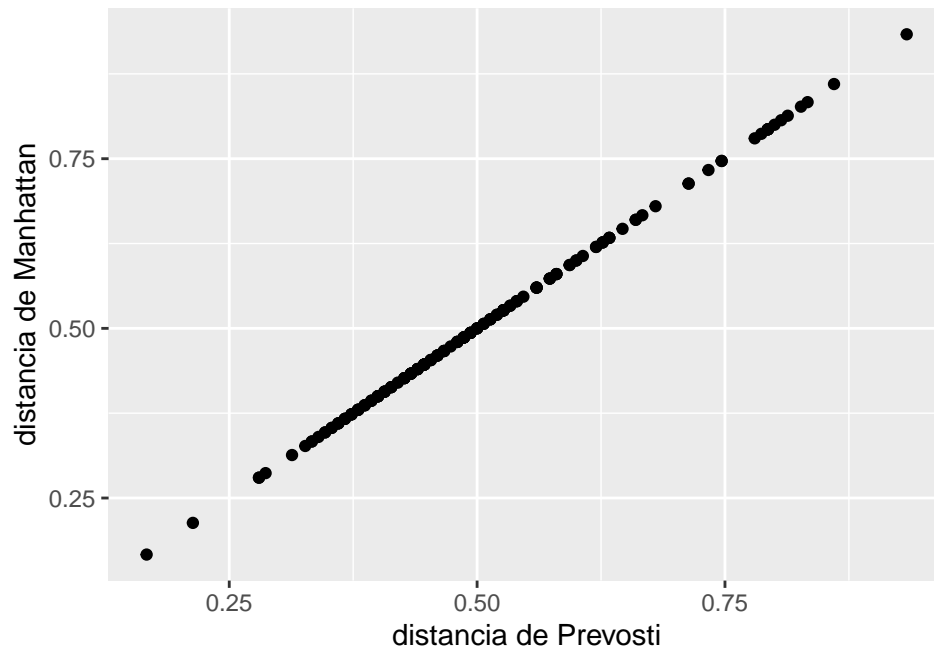
d) Encontrar la distancia de Manhattan acotada y comparar con lo hallado en (b)

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	VM1	VM2	VM3	LE
M1	0.00	0.17	0.40	0.33	0.37	0.21	0.43	0.39	0.47	0.54	0.37	0.41	0.29	0.31	0.56	0.59	0.45	0.49	0.83
M2	0.17	0.00	0.40	0.33	0.38	0.28	0.43	0.40	0.44	0.53	0.41	0.48	0.34	0.35	0.55	0.63	0.45	0.48	0.81
M3	0.40	0.40	0.00	0.53	0.42	0.35	0.48	0.53	0.61	0.49	0.39	0.66	0.43	0.49	0.63	0.67	0.53	0.51	0.79
M4	0.33	0.33	0.53	0.00	0.46	0.37	0.49	0.49	0.53	0.62	0.39	0.49	0.38	0.39	0.49	0.66	0.49	0.54	0.81
M5	0.37	0.38	0.42	0.46	0.00	0.28	0.40	0.45	0.47	0.45	0.43	0.50	0.35	0.45	0.58	0.57	0.45	0.46	0.79
M6	0.21	0.28	0.35	0.37	0.28	0.00	0.45	0.47	0.49	0.51	0.39	0.46	0.33	0.38	0.65	0.68	0.48	0.53	0.86
M7	0.43	0.43	0.48	0.49	0.40	0.45	0.00	0.45	0.40	0.47	0.37	0.51	0.37	0.45	0.44	0.57	0.41	0.36	0.83
M8	0.39	0.40	0.53	0.49	0.45	0.47	0.45	0.00	0.36	0.49	0.37	0.52	0.50	0.43	0.52	0.60	0.41	0.40	0.75
M9	0.47	0.44	0.61	0.53	0.47	0.49	0.40	0.36	0.00	0.43	0.50	0.56	0.44	0.41	0.49	0.63	0.38	0.39	0.80
M10	0.54	0.53	0.49	0.62	0.45	0.51	0.47	0.49	0.43	0.00	0.56	0.71	0.56	0.57	0.63	0.71	0.51	0.43	0.93
M11	0.37	0.41	0.39	0.39	0.43	0.39	0.37	0.37	0.50	0.56	0.00	0.44	0.37	0.43	0.60	0.63	0.45	0.46	0.79
M12	0.41	0.48	0.66	0.49	0.50	0.46	0.51	0.52	0.56	0.71	0.44	0.00	0.35	0.45	0.54	0.58	0.58	0.53	0.73
M13	0.29	0.34	0.43	0.38	0.35	0.33	0.37	0.50	0.44	0.56	0.37	0.35	0.00	0.36	0.53	0.59	0.49	0.47	0.79
M14	0.31	0.35	0.49	0.39	0.45	0.38	0.45	0.43	0.41	0.57	0.43	0.45	0.36	0.00	0.41	0.62	0.51	0.45	0.78
M15	0.56	0.55	0.63	0.49	0.58	0.65	0.44	0.52	0.49	0.63	0.60	0.54	0.53	0.41	0.00	0.63	0.53	0.43	0.75
VM1	0.59	0.63	0.67	0.66	0.57	0.68	0.57	0.60	0.63	0.71	0.63	0.58	0.59	0.62	0.63	0.00	0.53	0.53	0.41
VM2	0.45	0.45	0.53	0.49	0.45	0.48	0.41	0.41	0.38	0.51	0.45	0.58	0.49	0.51	0.53	0.53	0.00	0.44	0.79
VM3	0.49	0.48	0.51	0.54	0.46	0.53	0.36	0.40	0.39	0.43	0.46	0.53	0.47	0.45	0.43	0.53	0.44	0.00	0.79
LE	0.83	0.81	0.79	0.81	0.79	0.86	0.83	0.75	0.80	0.93	0.79	0.73	0.79	0.78	0.75	0.41	0.79	0.79	0.00

La correlación entre ambas distancias es

```
cor(mnhtdist, prevosti)
```

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



e) Podría aplicar el coeficiente de similitud SM en este problema ? Porque ?

Para marcadores moleculares, si el interés es analizar presencia y ausencia, se deberá utilizar un coeficiente de similitud. En cambio si el interés es estudiar frecuencias alélicas, se deberá contemplar distancias genéticas. En el caso del juego de datos de Maíz que utiliza microsatélites (simple sequence) o SSR, tenemos un problema de factores codominantes multialélicos, por lo cual es mejor considerar una distancia genética.