

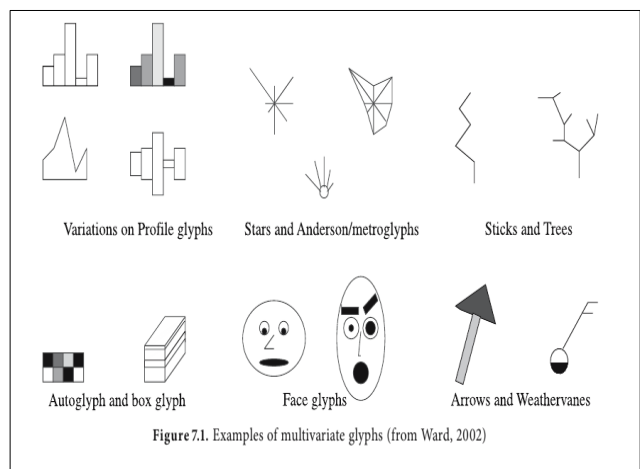
# Gráficos multivariados

**Glifo: Objeto gráfico para visualizar datos multivariados (representar simultáneamente varias variables)**

Base de datos  
(matriz de datos)

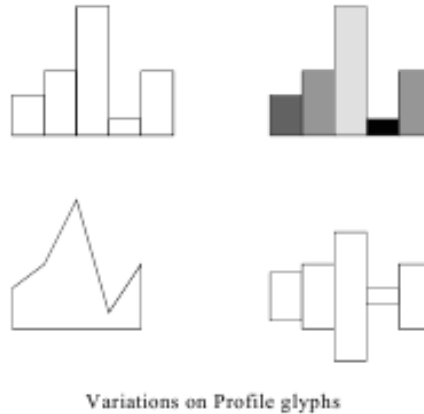
individuos	Variables				
	X1	X2	X3		
1					
2					
3					

Atención: Los individuos a representar están en las filas



En ocasiones estos gráficos ayudan a esclarecer relaciones en los problemas de las pruebas de hipótesis de independencia (Aunque la técnica es más general, obviamente)

### Gráficos de perfiles



## Descripción de los valores medios utilizando diagramas de perfiles



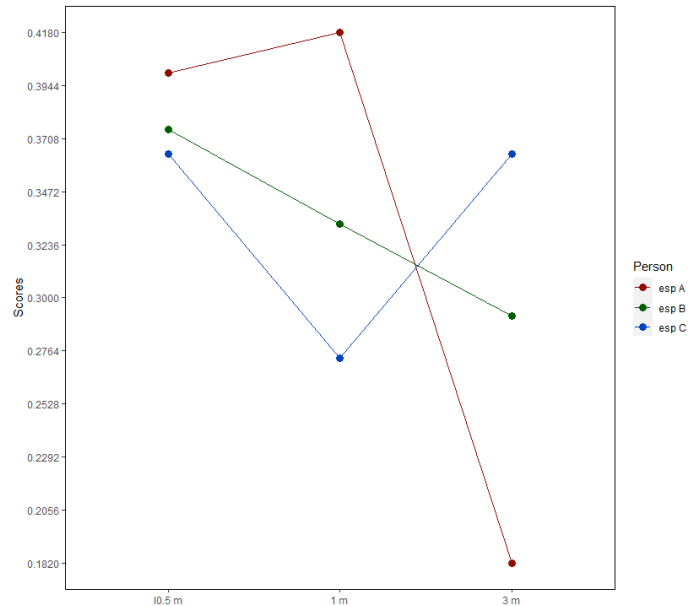
Trabajaremos perfiles con proporciones

En un estudio sobre las moscas de arena se utilizaron unas trampas que fueron colocadas a tres alturas respecto al suelo del bosque (0.5, 1 y 3 metros). Después de concluida la recolección de los especímenes para el estudio a los investigadores les pareció que la especie de las moscas capturadas dependía de la altura de la trampa

Especie	altura		
	0.5	1	3
A	110	115	50
B	90	80	70
C	20	15	20
total	220	210	140

Construya un diagrama perfiles de líneas (utilizando segmentos) para representar las alturas a las que vuelan las especies

Haga el gráfico como ejercicio (a continuación veremos los aspectos computacionales). Ahora interprete los resultados



El grupo de investigación de mercado de una importante compañía productora de cervezas realizó un estudio de mercado. Al analizar los segmentos de mercado de cada uno de los tipos de cerveza que producen estaban interesados en determinar si las preferencias por un tipo específico de cerveza dependía del género del consumidor. En caso de que las preferencias fueran independientes del género del consumidor, iniciarían una campaña publicitaria para todas las cervezas de la compañía. Pero, si las preferencias por los distintos tipos de cerveza dependían del género del consumidor, la empresa ajustaría sus promociones a los mercados. Para analizar este problema tomaron una muestra de consumidores con los resultados siguientes:

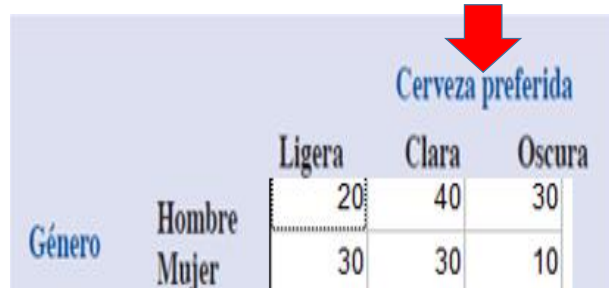
Explore utilizando los perfiles (de hombres y mujeres) si se observan diferencias en las preferencias entre los sexos

		Cerveza preferida		
		Ligera	Clara	Oscura
Género	Hombre	20	40	30
	Mujer	30	30	10

Recuerde: Las proporciones son promedios

## Datos procesados

Variables dicotómicas (3)  
Procesadas con sus totales



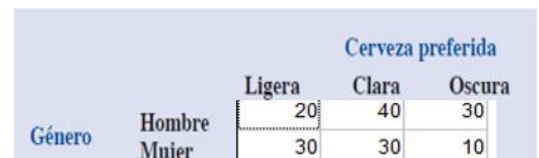
		Cerveza preferida		
		Ligera	Clara	Oscura
Género	Hombre	20	40	30
	Mujer	30	30	10

Grupos de individuos

Los totales deben estandarizarse:  
como porcentajes (promedios)

Se caracterizan hombres / mujeres para describir los grupos que deseo comparar

**Macro:** Paso 1: digitar los datos  
líneas Paso 2: Los estandarizo, en este caso como proporciones (por columna)  
cerveza



		Cerveza preferida		
		Ligera	Clara	Oscura
Género	Hombre	20	40	30
	Mujer	30	30	10

```
#perfiles con grafico de lineas (poligonal)

#M matriz con los datos cerveza
#digitar los datos.
M=matrix(nr=2 , nc= 3 )
data.entry(M)

#nombres
rownames(M)=c("hombres","mujeres")
colnames(M)=c("ligera","clara","oscura")

#calculo proporciones por columna (valores por sexo)
T1=round(prop.table(M,2),3)

#al escribir prop.table(M,2) indico que las
#proporciones son por columna
#por fila seria prop.table(M,1)
#datos a representar
T1[1,]
T1[2,]
```

```
> M
      ligera clara oscura
hombres   20   40   30
mujeres   30   30   10
> T1
      ligera clara oscura
hombres  0.4 0.571 0.75
mujeres  0.6 0.429 0.25
```

Base de  
datos  
cerveza

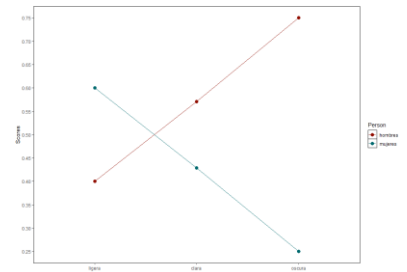
**Macro:**  
líneas  
cerveza

### Paso 3: Construir el gráfico

```
#para los perfiles con líneas:
#cargar paquete profileR y activarlo:
library(profileR)

#el comando con matrices
T1=as.matrix(T1)

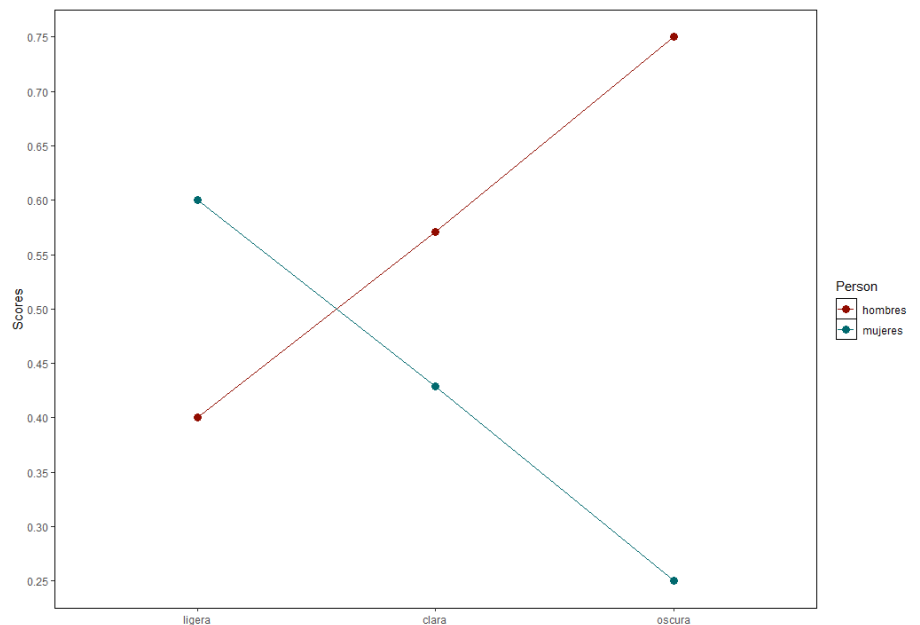
# el gráfico.
# Atención: No se estandariza porque los datos ya lo están,
# si no hay habría que hacerlo
profileplot(T1, person.id=rownames(T1), standardize=FALSE)
```



		Cerveza preferida		
		Ligera	Clara	Oscura
Género	Hombre	20	40	30
	Mujer	30	30	10

```
> T1
      ligera clara oscura
hombres  0.4 0.571  0.75
mujeres  0.6 0.429  0.25
```

Caracterice  
los gustos.  
Identifique  
las  
diferencias  
si las hay.



Se realiza un estudio para ver cómo los niveles de dióxido de azufre en la atmósfera influyen sobre el número de cloroplastos por célula en las hojas de los árboles.

Para el estudio se muestrean hojas de árboles tomadas en zonas (3) con niveles de  $\text{SO}_2$  que se clasifican como alto, normales o bajos siguiendo los criterios de las autoridades ambientales del país

Tabla 12.20. Datos utilizados para contrastar asociación entre nivel de cloroplasto y exposición al dióxido de azufre

Nivel de $\text{SO}_2$	Nivel de cloroplastos		
	Alto	Normal	Bajo
Alto	3	4	13
Normal	5	10	5
Bajo	7	11	2

**Construya un diagrama de perfiles (de líneas) para representar los niveles medios de cloroplastos según los niveles de  $\text{SO}_2$  de la zona**

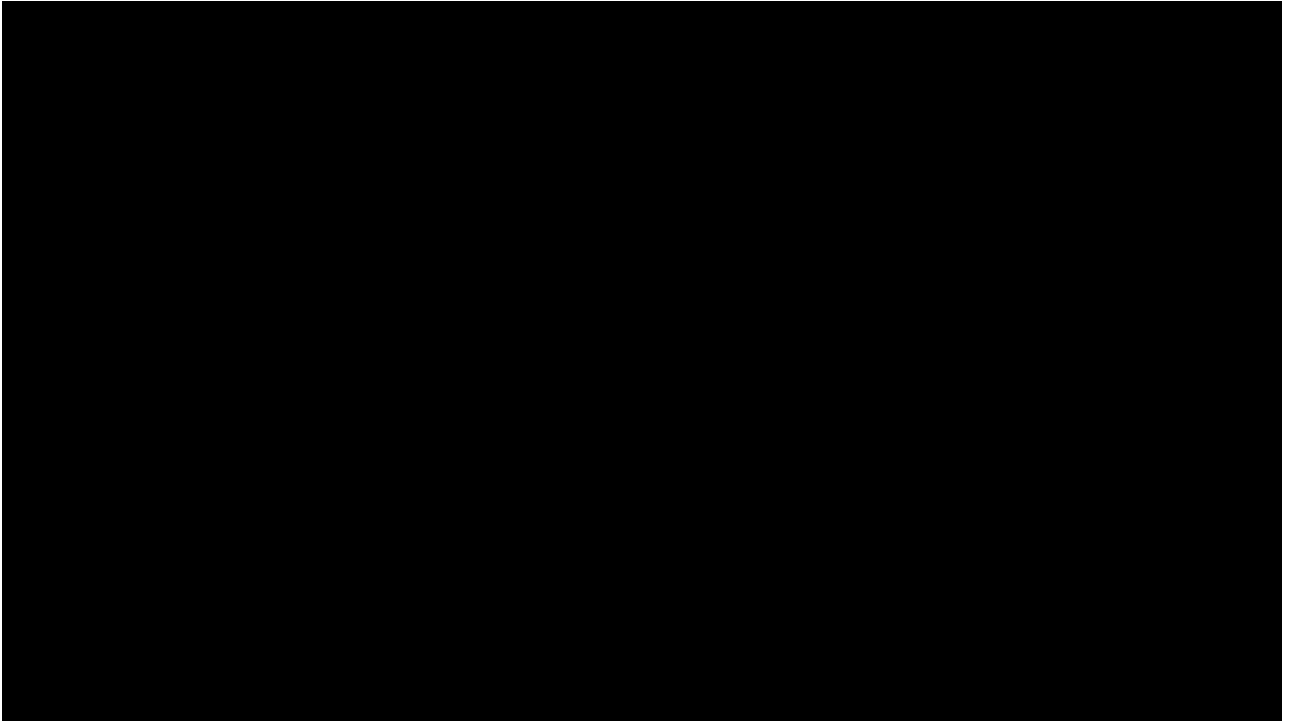
## Paso 1: digite los datos

Tabla 12.20. Datos utilizados para contrastar asociación entre nivel de cloroplasto y exposición al dióxido de azufre

Nivel de $\text{SO}_2$	Nivel de cloroplastos		
	Alto	Normal	Bajo
Alto	3	4	13
Normal	5	10	5
Bajo	7	11	2

Muestra de individuos agrupada  
... muestrean hojas de árboles tomadas en zonas (3)

... A cada árbol se le evalúan las 3 variables dicotómicas que aparecen en la tabla....los resultados son totales en cada caso



### Datos: exámenes

Problema de rendimiento académico

Estudio de resultados docentes en un curso de estadística

Interés del estudio: ver si se observan diferencias en los resultados en dos parciales en dependencia de si la nota final fue superior a 80 puntos o no

Variables

grupo: 1 (nota final  $> 80$ )

2 (nota final  $\leq 80$ )

P1: nota parcial 1

P2: nota parcial 2

Paso 1 debe resumir los resultados (promedios) en los grupos

Grupos	Variables	
	P1	P2
1	75.3	72.7
2	87.1	89.8

```
> attach(res)
The following objects
  P1, P2

> head(res)
  grupo P1 P2
1     1 78 73
2     1 74 76
3     2 82 79
4     2 90 96
5     2 87 90
6     2 90 92
> mean(P1[grupo==1])
[1] 75.33333
> mean(P1[grupo==2])
[1] 87.07692
> mean(P2[grupo==1])
[1] 72.77778
> mean(P2[grupo==2])
[1] 89.84615
```

Paso 3 se hace una matriz con los resultados (para los perfiles de los grupos)

```
> M
      Parcial 1 Parcial 2
grupo 1      75.3      72.7
grupo 2      87.1      89.8
```

```
M=matrix(nr=2 , nc= 2 )
data.entry(M)

rownames(M)=c("grupo 1","grupo 2")
colnames(M)=c("Parcial 1","Parcial 2")
```

```
> attach(res)
The following objects
  P1, P2

> head(res)
  grupo P1 P2
1     1 78 73
2     1 74 76
3     2 82 79
4     2 90 96
5     2 87 90
6     2 90 92
> mean(P1[grupo==1])
[1] 75.33333
> mean(P1[grupo==2])
[1] 87.07692
> mean(P2[grupo==1])
[1] 72.77778
> mean(P2[grupo==2])
[1] 89.84615
```



```
#el grafico
library(profileR)

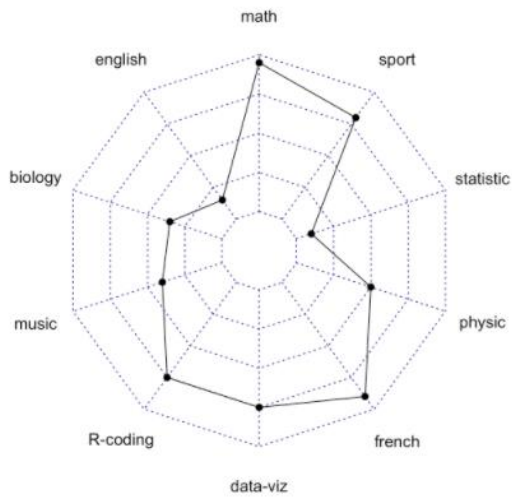
profileplot(M, person.id=rownames(M), standardize=FALSE)
```

No es necesario porque las dos variables se miden en la misma escala. Si las escalas de medición fueran diferentes (como ocurre usualmente) hay que estandarizar las variables.

$$0 \leq \frac{X - \text{Min}}{\text{Max} - \text{Min}} \leq 1$$

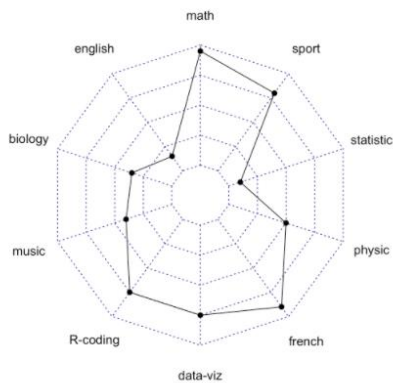
$$0 \leq \frac{X - \bar{X}}{\text{Max} - \text{Min}} \leq 1$$

Otra forma de  
representar los datos  
multivariados:  
Los diagramas de radar



## Base de datos (matriz de datos)

individuos	Variables				
	X1	X2	X3		
1					
2					
3					



## Estructura de los datos

Nombre de fila	Variables		
	X1	X2	
Max			
Min			
1			
2			
3			
.....			
n			

		Cerveza preferida		
		Ligera	Clara	Oscura
Género	Hombre	20	40	30
	Mujer	30	30	10

## Archivo gráficos de radar plantilla

## Preparación de la base de datos

```

1 #M matriz del ejemplo
2 M=matrix(c(20,40,30,30,30,10),nr=2,nc=3,byrow=TRUE)
3 rownames(M)=c("hombres","mujeres")
4 colnames(M)=c("ligera","clara","oscura")
5 M1=round(prop.table(M,2),3)
6
7 #desde acá es el procedimiento:
8 #M1 en este caso es proporciones
9 #en general debe ser una matriz normalizada
10
11 #defino máximos y mínimos, en el caso de las cervezas
12 # como son proporciones
13 #los pongo a 0,1. utilizo el comando de repetir
14 f1=rep(1,3)
15 f2=rep(0,3)
16 #datos finales
17 MF=rbind(f1,f2,M1)
18 #lo convierto a marco de datos
19 D=as.data.frame(MF)
20
21 #le pongo nombre a las filas que añadí
22 rownames(D)[1]="max"
23 rownames(D)[2]="min"

```

Nombre de fila	Variables		
	ligera	clara	oscura
Max			
Min			
hombres			
mujeres			

```

> D
      ligera clara oscura
max      1.0 1.000  1.00
min      0.0 0.000  0.00
hombres  0.4 0.571  0.75
mujeres  0.6 0.429  0.25

```

```

#cargo el paquete
library(fmsb)

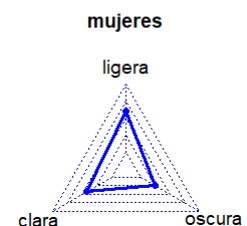
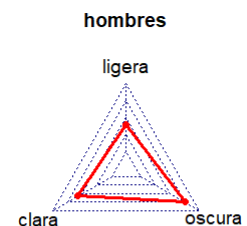
#para poner los graficos juntos (decido la posición)
par(mfrow=c(2,1))

#grafico de los hombres (filas 1,2 y 3)
radarchart(D[1:3,],pcol="red",plwd=3,vlcex = 1.2,title="hombres")

#grafico de las mujeres (filas 1,2,4)
x=c(1,2,4)
radarchart(D[x,],pcol="blue",plwd=3,vlcex = 1.2,title="mujeres")

#significado
#pcol: color de la línea
#plwd: ancho de la línea
#vlcex: tamaño fuente etiqueta

```



La tabla siguiente muestra los resultados de promedios de 4 grupos de estudiantes de una misma escuela en 5 exámenes finales

Estudiante	A1	A2	A3	A4	A5
a	5	23	4	75	2.1
b	2	50	1.2	94	4.8
f	1.5	48	1.9	92	4.4
j	0.5	7	0.4	20	1.2

**Problema:** Construya un diagrama de radar para representar el desempeño de los grupos y analice semejanzas y deferencias.

Los exámenes A1, A3 y A5 se califican de 0 a 5

El examen A2 se califica de 0 a 50

El examen A4 se califica de 0 a 100

Los datos  
deben  
estandarizarse

```
#digitar
M=matrix(nr=4,nc=5)
data.entry(M)

#nombres
rownames(M)=c("a", "b","c","d")
colnames(M)=c("A1","A2","A3","A4","A5")

#los datos deben estandarizarse
library(BBmisc)

#los estandarizo
M1=normalize(M,method="range",range=c(0.5,1.5),margin=2)
```

```
> M
  A1 A2  A3 A4  A5
a 5.0 23 4.0 75 2.1
b 2.0 50 1.2 94 4.8
c 1.5 48 1.9 92 4.4
d 0.5  7 0.4 20 1.2
```

Trabajando con los  
datos estandarizados  
continuar la  
construcción del  
gráfico

```
> #los estandarizo
> normalize(M,method="range",range=c(0.5,1.5),margin=2)
  A1      A2      A3      A4      A5
a 1.500000 0.872093 1.500000 1.243243 0.750000
b 0.833333 1.500000 0.722222 1.500000 1.500000
c 0.722222 1.453488 0.916667 1.472973 1.388889
d 0.500000 0.500000 0.500000 0.500000 0.500000
```