

Gráficos multivariantes

¿Qué son?

En ocasiones queremos herramientas gráficas que nos ayuden a identificar **relaciones entre variables**, **grupos**, y a detectar posibles **datos atípicos**.

Estos gráficos pueden responder a:

- ¿qué **variables son las dominantes** en cada observación?
- ¿se pueden establecer **similitudes entre estas unidades**?
- ¿se pueden establecer **clusters**?
- ¿existen **valores atípicos**?

Gráfico de estrella

Este gráfico nos permite **ver relaciones** entre las unidades de investigación y **agruparlas** según la similitud de la estrella.

Asignan a cada observación una estrella con **tantos rayos o ejes (igualmente espaciados) como variables** queramos representar.

Las **longitudes de los rayos son proporcionales** a los valores de las variables en la observación asociada a la estrella.

Los **extremos de los rayos se conectan con segmentos** de líneas rectas para formar una estrella.

Gráfico de estrella

En todas las estrellas se usa siempre **el mismo rayo** o eje para representar **la misma variable**.

El eje j en una estrella i **depende de x_{ij}** .

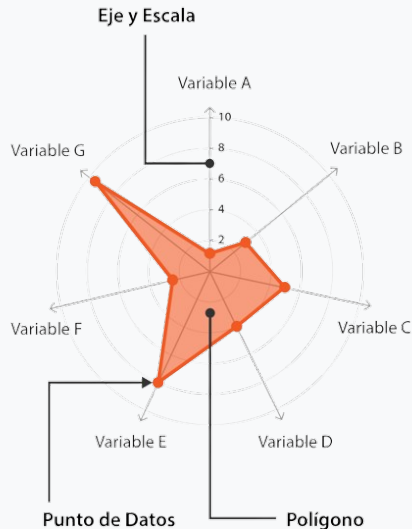


Gráfico de estrella

¿Cómo se hace?

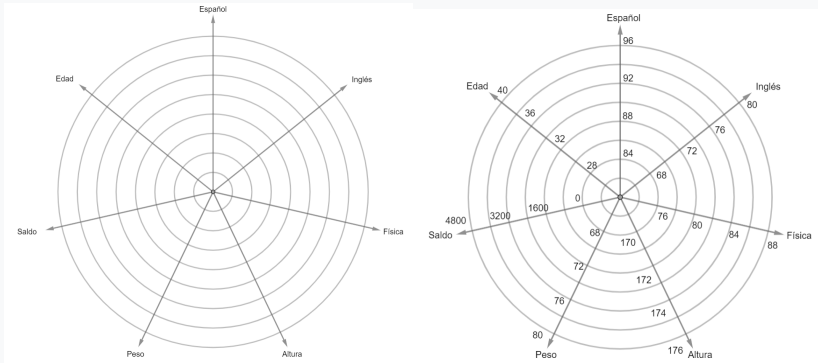
Contexto (hipotetico)

Supongamos que tenemos 7 variables en nuestro estudio: puntajes de una prueba de español, inglés y física, la edad, la altura, el saldo en la cuenta bancaria y el peso de una persona. Se generaron observaciones **hipotéticas**, obteniendo lo siguiente:

Español	Inglés	Física	Salario	Edad	Altura	Peso
88	72	80	\$3200	32	168	72
96	80	88	\$4800	40	176	80
90	76	80	-\$1600	36	173	68
88	72	86	-\$800	38	176	74

Gráfico de estrella

¿Cómo se hace?

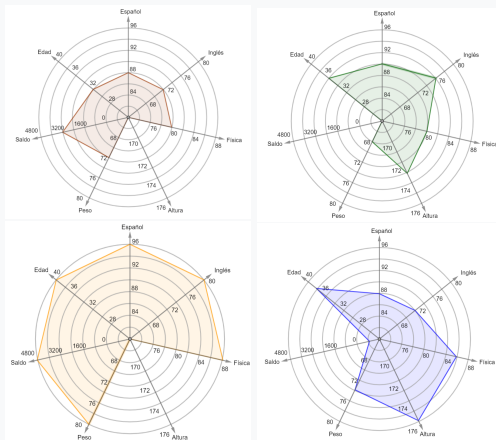


Paso 1: ubicar tantos ejes como variables. Deben estar igualmente espaciados.

Paso 2: definir la escala de cada eje con el máximo y el mínimo de cada variable.

Gráfico de estrella

¿Cómo se hace?



Paso 3: Ubicar los puntos en cada observación y unirlos para formar la estrella.

Gráfico de estrella

¿Cómo se hace?

El resultado anterior se puede obtener **escalando** los datos por medio del metodo *min-max* y siguiendo los mismos pasos.

$$X_{i_{transformado}} = \frac{X_i - X_{min}}{X_{max} - X_{min}}$$

Con esto tendremos que **todos los ejes tendrán la misma escala [0,1]**.

Es recomendable escalar los datos, pues hay algunos softwares que hacen la escala de los ejes con el menor y el mayor dato de **todas** las variables.

Gráfico de estrella

Función en R

En R utilizamos el comando ***stars()*** para realizar este gráfico. Los parametros más relevantes son;

- **x**: la matriz de datos. Los datos faltantes serán tratados como si fueran 0.
- **full**: argumento logico. Si es *TRUE*, los gráficos ocuparán un círculo completo. De lo contrario, solo ocuparán un semicírculo superior.
- **scale**: argumento logico. Si es *TRUE*, los datos serán escalados con el metodo *min-max*.
- **labels**: vector de caracteres para etiquetar cada gráfico.

Gráfico de estrella

Función en R

- **key.loc:** vector con coordenadas x e y para ubicar la estrella que da las claves que identifican las variables en todas las estrellas.
- **key.labels:** vector de caracteres para etiquetar los segmentos de la unidad de claves que identifica las variables en todas las estrellas.
- **draw.segments:** argumento lógico. Si es *TRUE* dibuja un diagrama de segmento.

Gráfico de estrella

Ejemplo

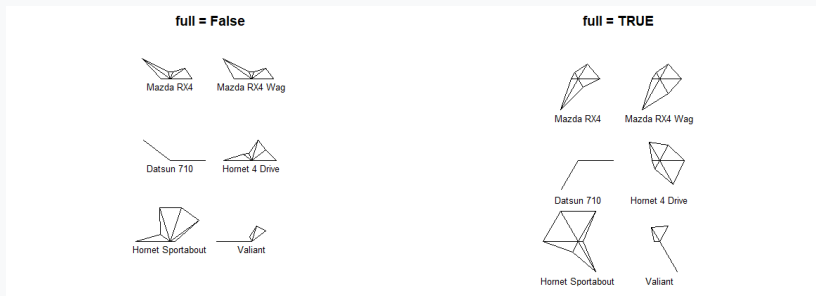
Ejemplo

Los siguientes datos muestran información relevante sobre diferentes modelos de automóviles, incluyendo la eficiencia de combustible (MPG), el número de cilindros (Cyl), el desplazamiento del motor (Disp), la potencia (HP), la relación de compresión (Drat), y el peso (Wt) de cada vehículo.

Modelo	MPG	Cyl	Disp	HP	Drat	Wt
Mazda RX4	21.0	6	160	110	3.90	2.620
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875
Datsun 710	22.8	4	108	93	3.85	2.320
Hornet 4 Drive	21.4	6	258	110	3.08	3.215
Hornet Sportabout	18.7	8	360	175	3.15	3.440
Valiant	18.1	6	225	105	2.76	3.460

Gráfico de estrella

Ejemplo - full

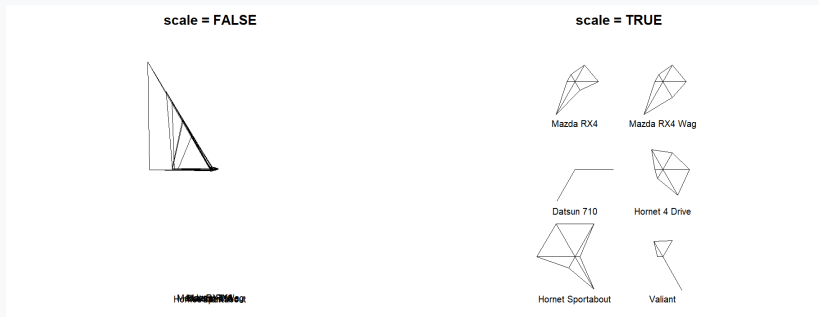


```
> stars(mtcars[1:6, 1:6], main = "full = False", full = FALSE)
```

```
> stars(mtcars[1:6, 1:6], main = "full = TRUE", full = TRUE)
```

Gráfico de estrella

Ejemplo - scale

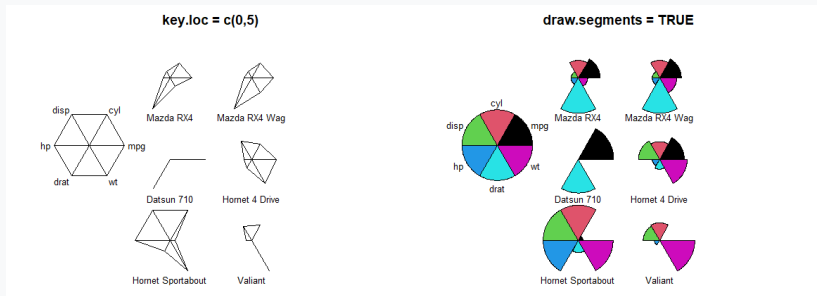


```
> stars(mtcars[1:6, 1:6], main = "scale = FALSE", full = TRUE,  
scale = FALSE)
```

```
> stars(mtcars[1:6, 1:6], main = "scale = FALSE", full = TRUE,  
scale = TRUE)
```

Gráfico de estrella

Ejemplo - key.loc/draw.segments

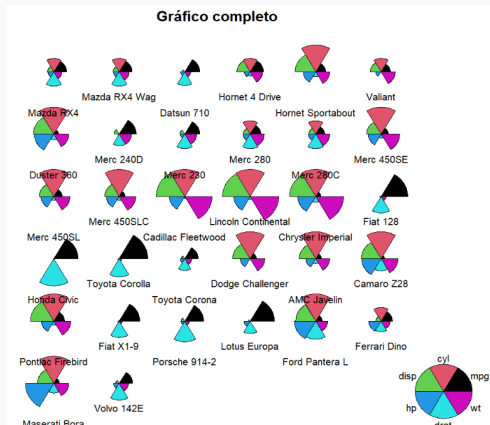


```
> stars(mtcars[1:6, 1:6], main = "key.loc = c(0,5)", full = TRUE,  
scale = TRUE, key.loc = c(0,5))
```

```
> stars(mtcars[1:6, 1:6], main = "draw.segments = TRUE", full =  
TRUE, scale = TRUE, key.loc = c(0,5), draw.segments = TRUE)
```

Gráfico de estrella

Ejemplo - completo



Este tipo de gráficos se vuelve **difícil de interpretar cuando hay muchas variables o muchas observaciones.**

Caras de Chernoff

Introducción

Las caras de Chernoff, introducidas por Herman Chernoff en 1973 en el *Journal of the American Statistical Association*, constituyen un método de visualización k-dimensional que transforma las observaciones de un conjunto de datos cuantitativos en representaciones faciales.

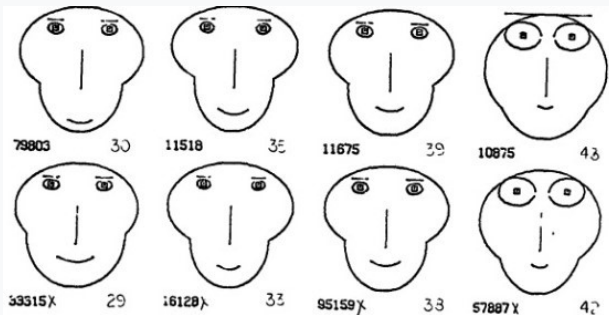
Inicialmente se limitó a representar hasta 18 variables, pero las versiones actuales han superado esta limitación, permitiendo la visualización de un número mayor de variables y ampliando así su utilidad en la exploración y comprensión de datos multivariados.

Este método es reconocido por la facilidad de comprensión, que aprovecha la capacidad humana desarrollada de **reconocer diferencias y similitudes en los rostros**.

Caras de Chernoff

Idea Principal

Las caras de Chernoff grafican cada unidad muestral \mathbf{X}_j como una cara simétrica, donde cada variable \mathbf{X}_{ji} representa una característica, como ojos, cejas, pupilas, boca, entre otras. La dimensionalidad de cada rasgo varía de acuerdo a la importancia, asignando un peso al valor \mathbf{X}_{ji} según la relevancia que tiene en la variable \mathbf{X}_j .



Caras de Chernoff

Pasos

Supongamos que tenemos una muestra aleatoria \mathbf{X} de tamaño $n = 5$, extraída de una Población con $p = 5$ variables.

ID	X_1	X_2	X_3	X_4	X_5
1	1	0	2.4	2.25	100
2	0	10	1.2	1.5	20
3	0.25	7.5	1.5	0.75	100
4	0.2	7	0.3	1.8	25
5	0.1	2	0.45	1.2	40

Paso 1 : Identificar variables discriminante y la cantidad de filas a gráficar.

En nuestro caso, disponemos de una muestra pequeña con un número limitado de variables, por lo que optaremos por utilizar la muestra tal como está.

Paso 2 : Indentificar los valores extremos para cada variable \mathbf{X}_i , donde:

$$\min(X_i) = \min(x_{1i}, x_{2i}, \dots, x_{ni})$$

$$\max(X_i) = \max(x_{1i}, x_{2i}, \dots, x_{ni})$$

ID	\mathbf{X}_1	\mathbf{X}_2	\mathbf{X}_3	\mathbf{X}_4	\mathbf{X}_5
1	1	0	2.4	2.25	100
2	0	10	1.2	1.5	20
3	0.25	7.5	1.5	0.75	100
4	0.2	7	0.3	1.8	25
5	0.1	2	0.45	1.2	40

Paso 3 : Diseñar el esquema de representación de las caras de Chernoff, asignando valores de variables a características faciales de acuerdo con los pesos asignados en función de los valores extremos y los valores intermedios de las variables.

Paso 4 : Generar las caras de Chernoff utilizando el diseño establecido en el Paso 3, representando cada observación de datos como una cara única.

Los pasos 2, 3 y 4 suelen ser ejecutados mediante funciones en un entorno de trabajo como R o Python. Sin embargo, el primer paso es de suma importancia, ya que implica la asignación de las variables más discriminantes o informativas en el contexto a los rasgos faciales destacados, como los ojos, las cejas o el cabello.

Chernoff recomienda explorar diversas permutaciones de las variables hasta encontrar aquella que visualmente represente de manera óptima las similitudes y diferencias entre las observaciones.

Caras de Chernoff

Función en R

Para graficar caras de Chernoff se usa la función **faces** del paquete **aplpack**.

```
> install.packages("ggplot2")
```

```
> require (aplpack)
```

```
> faces(df)
```

effect of variables:

"height of face " "Var1"

"width of face " "Var2"

"structure of face" "Var3"

"height of mouth " "Var4"

"width of mouth " "Var5"

"smiling " "Var1"

"height of eyes " "Var2"

"width of eyes " "Var3"

"height of hair " "Var4"

"width of hair " "Var5"

"style of hair " "Var1"

"height of nose " "Var2"

"width of nose " "Var3"

"width of ear " "Var4"

"height of ear " "Var5"

Caras de Chernoff

Hay parámetros que resultan importantes de manipulación, como:

- `face.type`

`face.type = 0`



`face.type = 1`



`face.type = 2`



- `nrow.plot`, `ncol.plotx`
- `fill`
- `scale`

Caras de Chernoff

Hay que tener en cuenta que la función **faces()** sólo utiliza las siguientes 15 características faciales:

Índice de columna	Rasgo facial que representa	Extremo inferior	Extremo superior
1	Altura de cara	Cara aplastada	Cara alta
2	Anchura de cara	Cara angosta	Cara ancha
3	Forma de cara (redonda o delgada)	Cara redonda	Cara delgada
4	Altura de boca	Boca fina	Boca gruesa
5	Anchura de boca	Boca angosta	Boca ancha
6	Curva de sonrisa	Expresión triste	Sonrisa
7	Altura de ojos	Ojos rasgados	Ojos muy abiertos
8	Anchura de ojos	Ojos angostos	Ojos anchos
9	Altura de pelo	Pelo aplastado	Pelo alto y estirado
10	Anchura de pelo	Pelo angosto	Pelo ancho
11	Estilo de pelo	Pelo caído	Pelo levantado
12	Altura de nariz	Nariz achatada	Nariz larga
13	Anchura de nariz	Nariz fina	Nariz ancha
14	Altura de orejas	Orejas achatadas	Orejas altas
15	Anchura de orejas	Orejas pegadas a la cara	Orejas abiertas

Caras de Chernoff

Implementación

Caso hipotético

Se tienen datos de cinco estudiantes en un estudio que analiza su rendimiento académico. Estos datos incluyen medidas relacionadas con aspectos como la precisión en exámenes, el tiempo dedicado al estudio semanalmente, las calificaciones obtenidas en diferentes proyectos, y el gasto en material de estudio durante el semestre.

ID	X_1	X_2	X_3	X_4	X_5
1	1	0	2.4	2.25	100
2	0	10	1.2	1.5	20
3	0.25	7.5	1.5	0.75	100
4	0.2	7	0.3	1.8	25
5	0.1	2	0.45	1.2	40

Caras de Chernoff

Las caras que representan el rendimiento de los estudiantes sugieren que el estudiante 1 es excepcional, mientras que los estudiantes 2, 3 y 4 muestran un rendimiento similar, aunque con diferencias notables entre ellos.

Estudiante 1



Estudiante 2



Estudiante 3



Estudiante 4



Estudiante 5



Caras de Chernoff

Desventajas

- Compresión de las pupilas depende de la dimensionalidad de los ojos.
- Reducción de variabilidad mediante normalización, puede involucrar pérdida de la información.
- Alta sensibilidad a valores atípicos.
- Los rasgos no siguen un orden.
- La interpretación depende de experiencias pasadas.



Caras de Chernoff

Ventajas

- Ser usada como una mnemotécnica para detectar y comprender fenómenos importantes.
- Realizar filtros visuales para enfocar variables discriminantes y descartar las demás.

Usos principales

- Análisis de cluster
- Detección de cambios de comportamiento durante un periodo de tiempo.
- Análisis discriminante
- Detección de outliers