



**Universidad Nacional Autónoma de México**  
Facultad de Ingeniería

# Selección de características

**Guillermo Molero-Castillo**

guillermo.molero@ingenieria.unam.edu

Noviembre, 2020

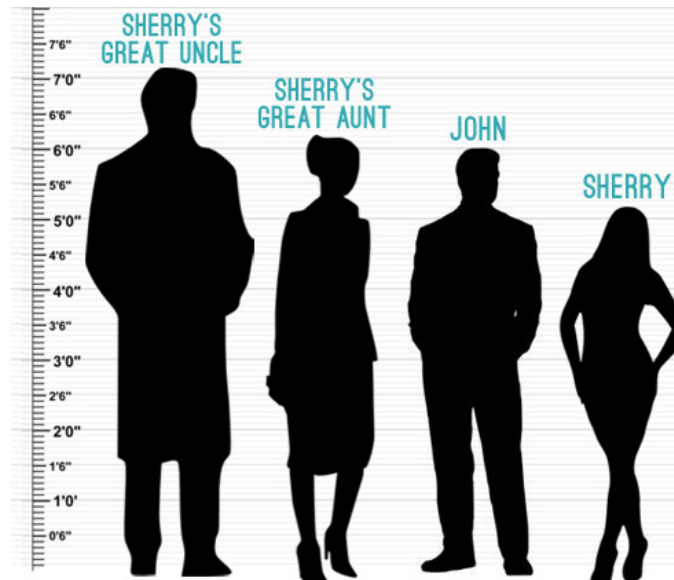
# Selección de características

- Existe una gran cantidad de factores que se debe tener en cuenta a la hora de seleccionar un algoritmo en **Machine Learning**.
- En realidad es casi un arte, y entre lo más importantes se encuentra decidir qué datos de entrada va a recibir el sistema.

**Ejemplo:** Clasificar personas por su estatura y sexo a través de imágenes de cámaras de seguridad.

- Estatura
- Color de piel
- Joroba
- Tipo de peinado
- Tamaño del cabello
- Tipo de vestimenta
- Tipo de zapatos
- Sombrero

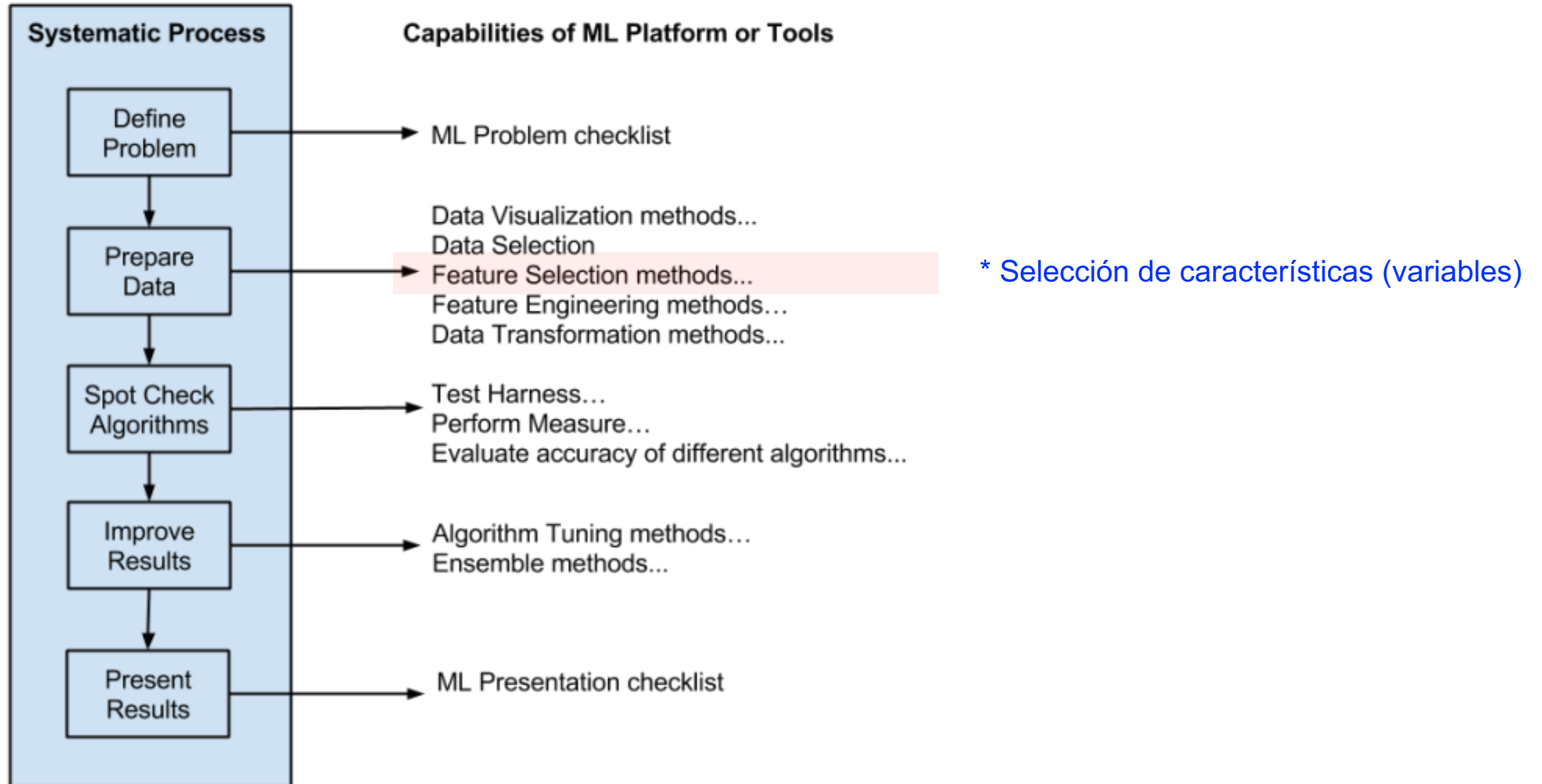
**Tamaño = 8 dimensiones**



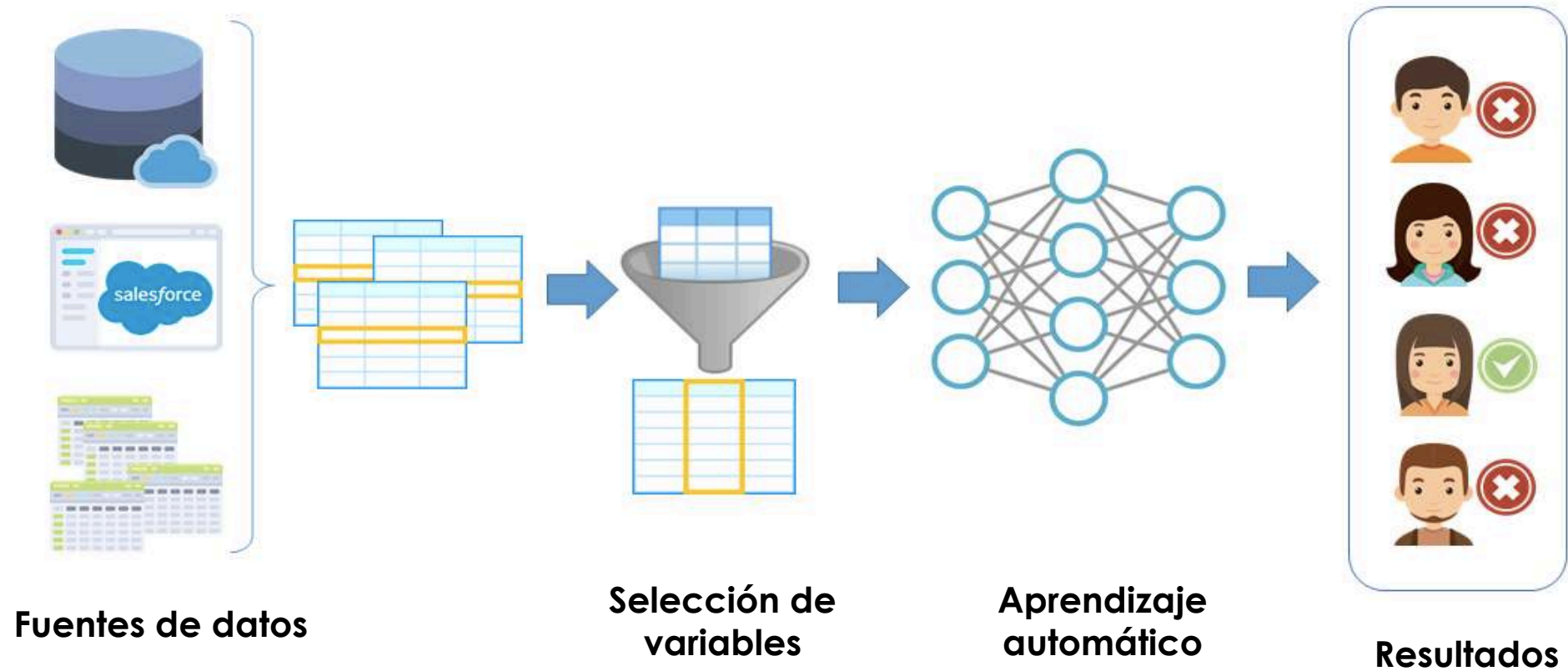
- Estatura
- Tipo de vestimenta
- Tipo de zapatos

**Tamaño = 3 dimensiones**

# Selección de características

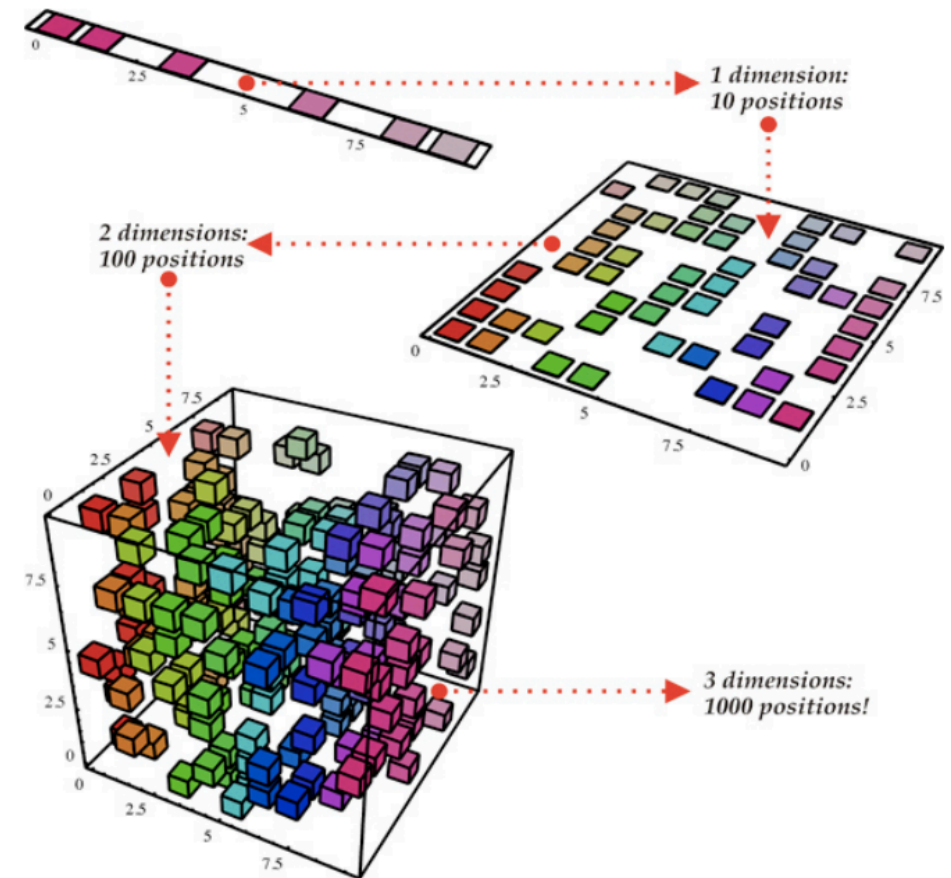


# Selección de características



# Selección de características

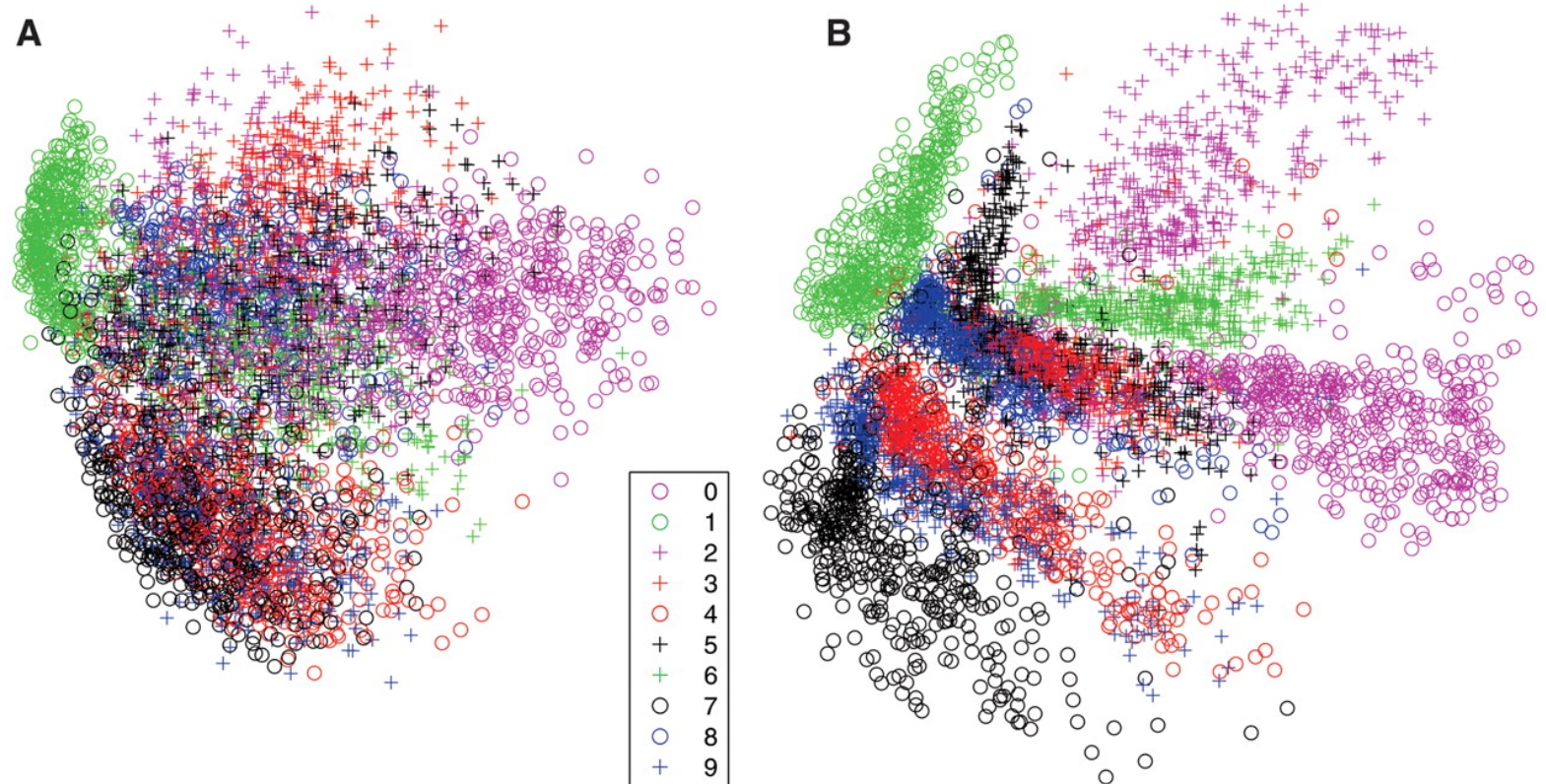
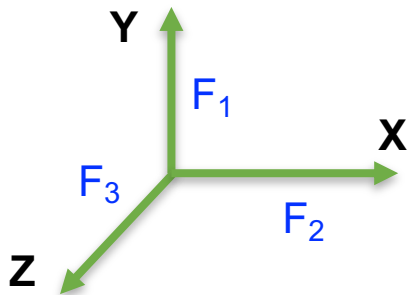
- A menudo hay demasiadas **variables**, en función de las cuales se condiciona el resultado final de un modelo.
- Cuanto mayor es el número de variables, más complejo es visualizar los datos y luego más complejo también trabajar con éstos.
- No es ilógico pensar que entre más variables (atributos) se tenga en cuenta será mejor.
- Sin embargo, esto es un **error común** que no se debe cometer.



# Selección de características

**En resumen**, a medida que aumenta la dimensionalidad de datos, incrementa también el nivel de complejidad de un proyecto de aprendizaje automático.

	$F_1$	$F_2$	$F_3$
$E_1$	3	5	0
$E_2$	2	1	1
$E_3$	5	2	0



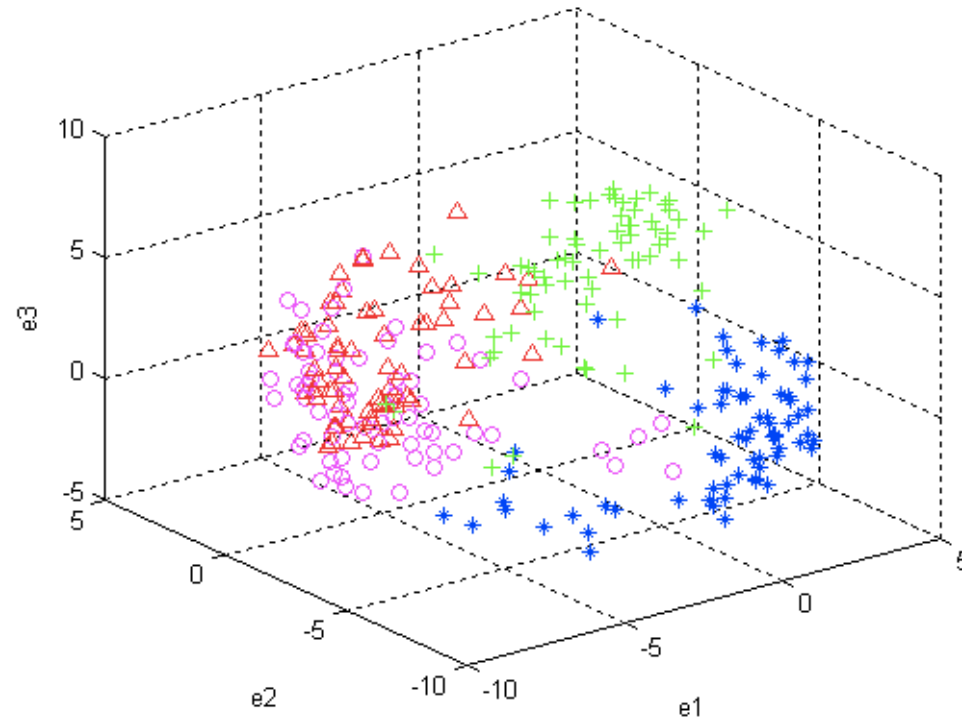
Se nota una superposición entre clases

# **Dimensionalidad de datos**



# Maldición de la dimensionalidad

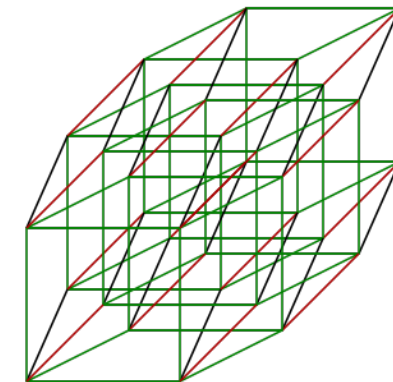
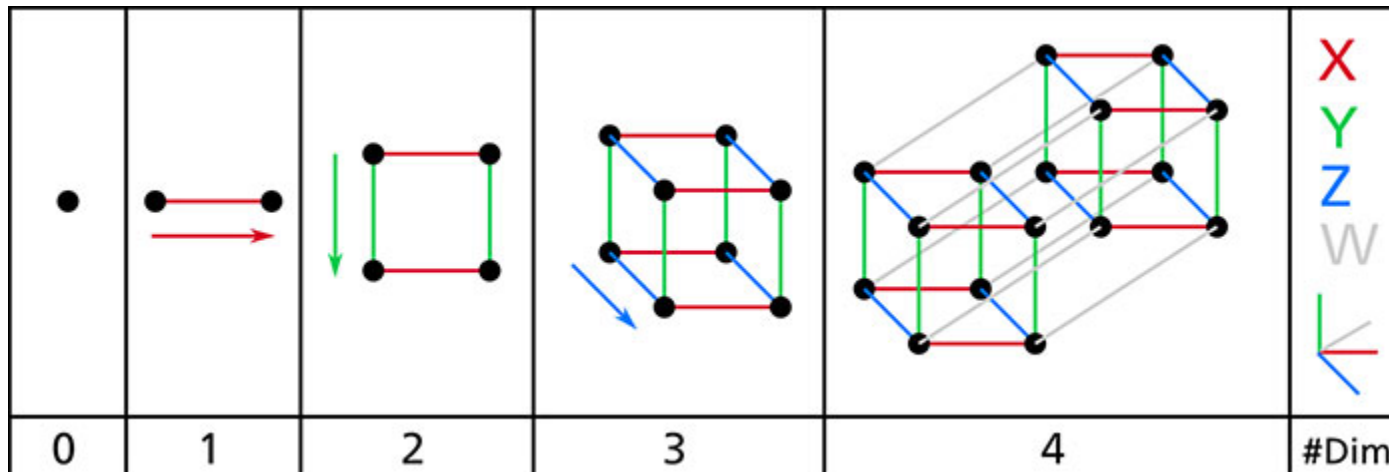
La **maldición de la dimensionalidad de datos** es un problema que se puede presentar si se quiere tener en cuenta todas las características (variables) posibles en un sistema.





# Maldición de la dimensionalidad

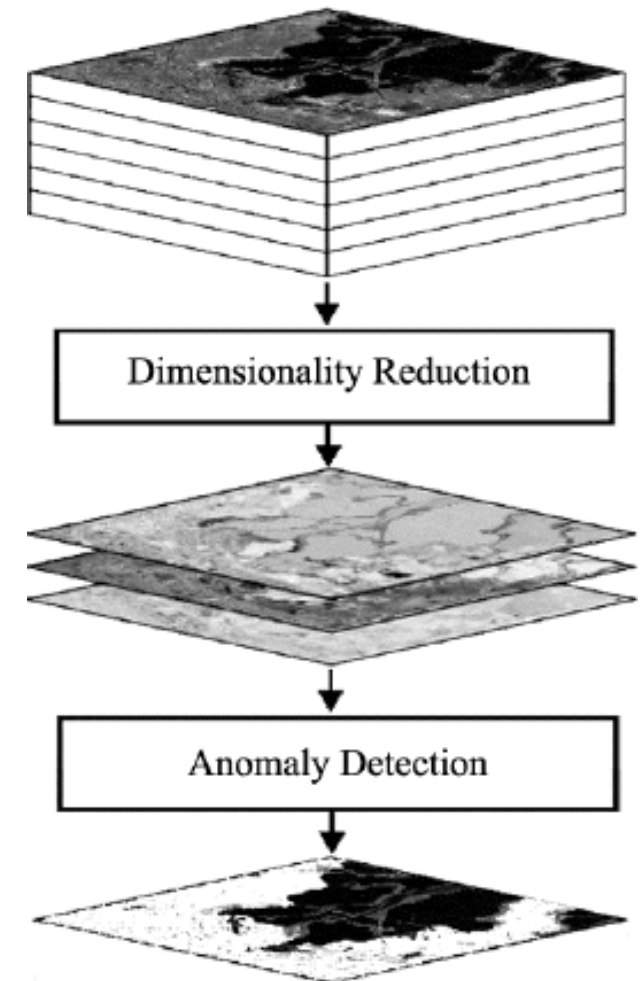
Esta **maldición** hace referencia al aumento exponencial de la dimensionalidad de datos.



5 Dimensiones

# Maldición de la dimensionalidad

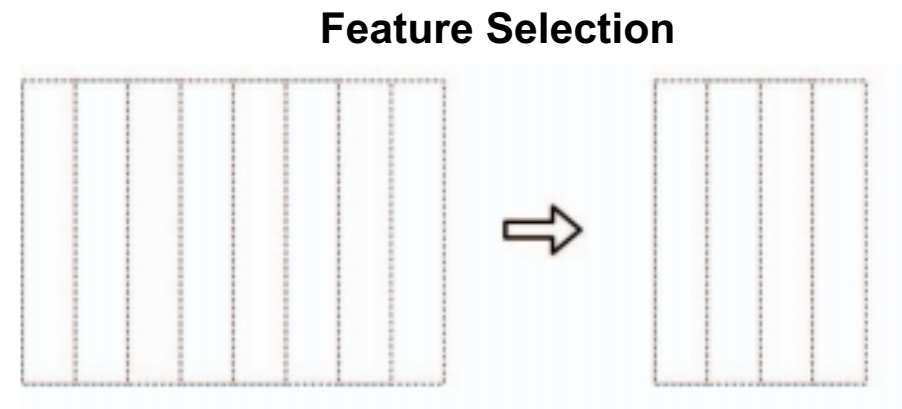
- En general, la mayoría de estas **variables están correlacionadas** y, por lo tanto, son redundantes.
- Aquí es donde entran en juego los algoritmos de reducción de dimensionalidad de datos.
- Esta **reducción de la dimensionalidad** es el proceso de reducir el número de variables mediante la obtención de alguna función de puntuación, que generalmente mide la relevancia de la característica.



# Maldición de la dimensionalidad

## Feature Selection

- Es el proceso de ordenar las variables por el valor de alguna función de puntuación.
- Para reducir la maldición de la dimensionalidad existen algunas estrategias:
- Discriminación manual, pero tiene limitaciones.
- Análisis correlacional de datos (Correlational Data Analysis, **CDA**)
- Análisis de componentes principales (Principal Component Analysis, **PCA**)
- Análisis discriminante lineal (Linear Discriminant Analysis, **LDA**)
- Análisis discriminante generalizado (Generalized Discriminant Analysis, **GDA**)



# Maldición de la dimensionalidad

## Ventajas de la reducción de dimensionalidad

- Ayuda en la compresión de datos y, por lo tanto, reduce el espacio de almacenamiento.
- Reduce el tiempo de cálculo.
- Ayuda a eliminar variables redundantes, si las hay.

## Desventaja de la reducción de dimensionalidad

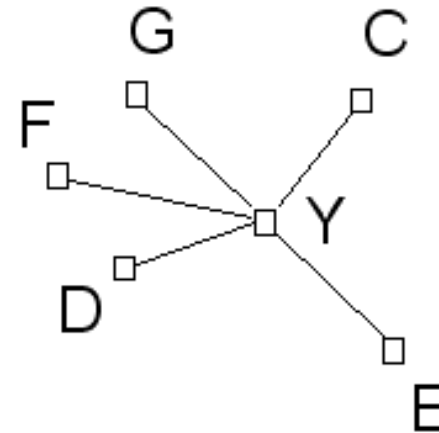
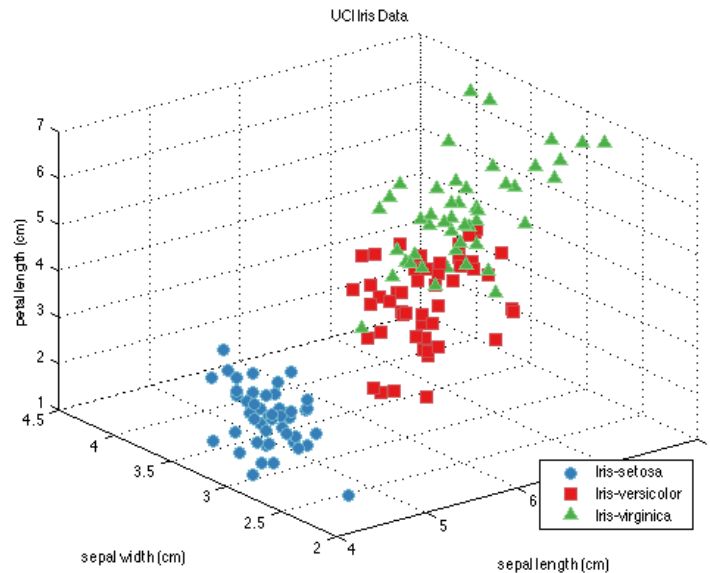
- Si no se hace un análisis cuidadoso, puede provocar pérdida de datos valiosos.

# **1. Análisis correlacional de datos**

# 1. Análisis correlacional de datos

## Correlaciones

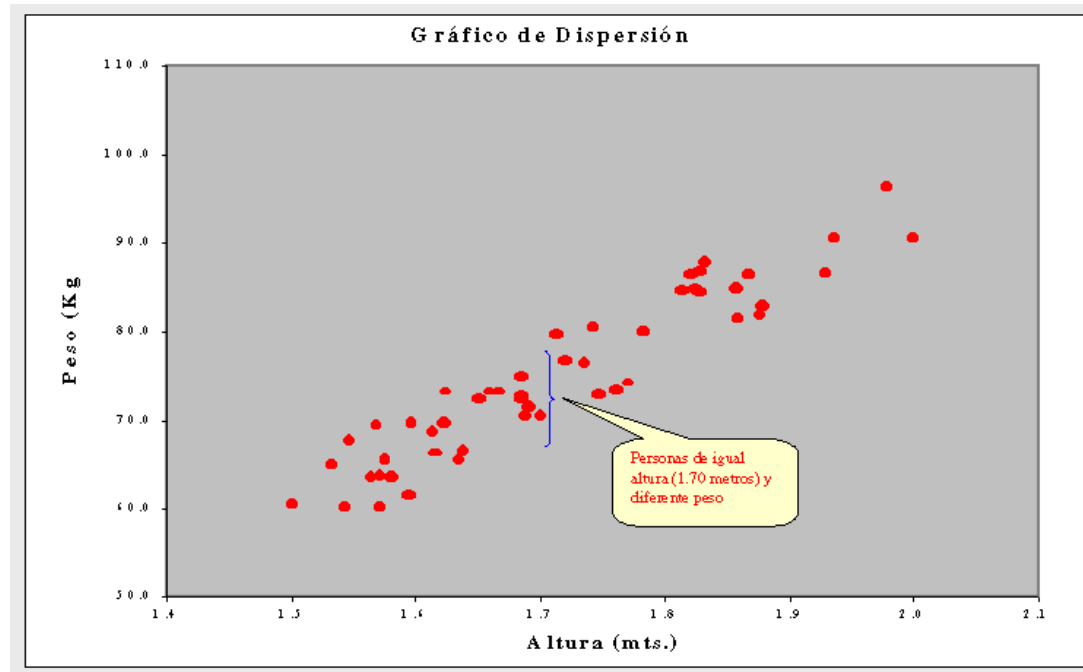
- El **CDA (ACD)** es útil para reducir el número de variables.
- De un espacio de alta dimensión, a en uno de un menor número de dimensiones.
- Esto se logra a través de la identificación de variables significativas.



# 1. Análisis correlacional de datos

## Correlaciones

- Estas identificaciones de correlaciones son utilizadas para determinar el **grado de similitud** (relevancia/irrelevancia) de los valores de dos variables numéricas.
- Existe correlación en 2 variables (**X e Y**) si al aumentar los valores de **X** también los hacen de **Y**, y viceversa.

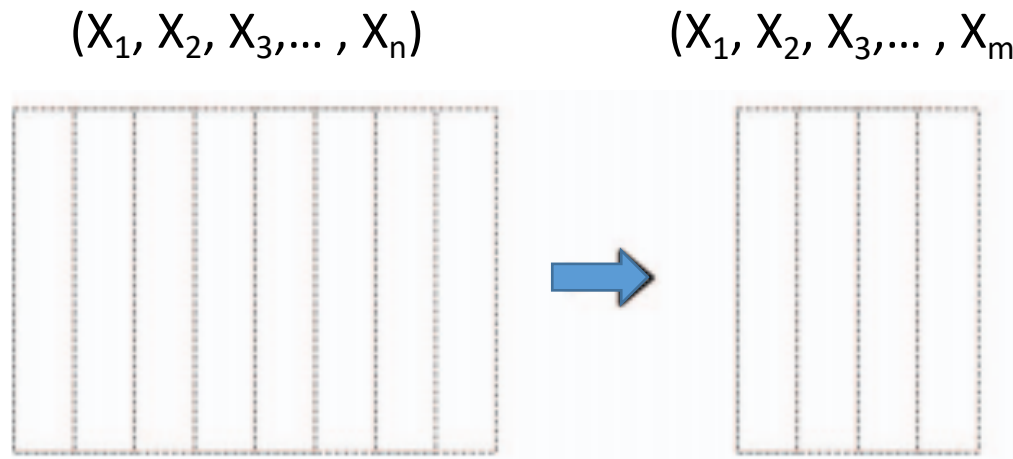




# 1. Análisis correlacional de datos

## Correlaciones

- La reducción consiste en que a partir de un conjunto de **variables originales**:  $X_1, X_2, X_3, \dots, X_n$
- Se obtiene otro subconjunto de **variables relevantes** :  $X_1, X_2, X_3, \dots, X_m$ , donde  $m < n$ .



# 1. Análisis correlacional de datos

## Coeficiente de correlación

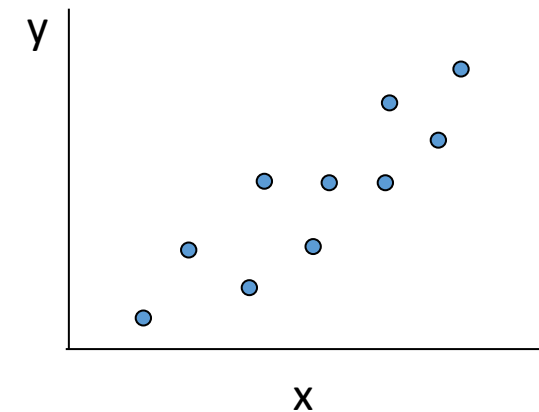
Los valores de correlación, conocidos como coeficiente de correlación de Pearson (su creador, Karl Pearson, 1857-1936), se define como:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

*$\bar{x}, \bar{y}$  son las medias aritméticas de  $x$  e  $y$ .*

**Covarianza**

**Desviaciones**



Los valores de correlación, en este caso **R**, pueden variar entre **-1 y 1**.

# 1. Análisis correlacional de datos

## Coeficiente de correlación

- Cuanto **más cerca está R de 1 o -1**, más fuerte es la correlación.
- Si **R es cercano a -1** las variables están correlacionadas negativamente.
- Si **R es cero** no existe correlación.

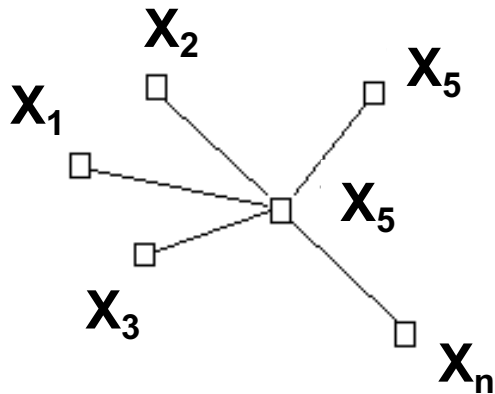
Intervalos para la identificación de correlaciones:

- De **-1.0 a -0.67** y **0.67 a 1.0** se conocen como correlaciones **fuertes o altas**.
- De **-0.66 a -0.34** y **0.34 a 0.66** se conocen como correlaciones **moderadas o medias**.
- De **-0.33 a 0.0** y **0.0 a 0.33** se conocen como correlaciones **débiles o bajas**.

# 1. Análisis correlacional de datos

## Matriz de correlaciones

- Consiste en crear una **matriz** que aporta información sobre la relación entre pares de variables.
- El objetivo es obtener un subconjunto de variables representativas que no tengan dependencia entre sí.



1	$r(X_1, X_2)$	$r(X_1, X_3)$	$r(X_1, X_4)$	$r(X_1, X_5)$	$r(X_1, X_6)$	$r(X_1, X_7)$	$r(X_1, X_8)$
$r(X_2, X_1)$	1	$r(X_2, X_3)$	$r(X_2, X_4)$	$r(X_2, X_5)$	$r(X_2, X_6)$	$r(X_2, X_7)$	$r(X_2, X_8)$
$r(X_3, X_1)$	$r(X_3, X_2)$	1					
$r(X_4, X_1)$	$r(X_4, X_2)$		1				
$r(X_5, X_1)$	$r(X_5, X_2)$			1			
$r(X_6, X_1)$	$r(X_6, X_2)$				1		
$r(X_7, X_1)$	$r(X_7, X_2)$					1	
$r(X_8, X_1)$	$r(X_8, X_2)$						1

# 1. Análisis correlacional de datos

## Matriz de correlaciones

	ABSOLUTOS																			MÁXIMOS
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		
1		0.035	0.043	0.061	0.250	0.230	0.340	0.155	0.090	0.132	0.148	0.071	0.080	0.130	0.083	0.181	0.149	0.102		0.340
2			0.022	0.014	0.006	0.085	0.056	0.043	0.031	0.044	0.060	0.030	0.021	0.052	0.038	0.072	0.060	0.039		0.085
3				0.107	0.013	0.035	0.037	0.024	0.010	0.042	0.011	0.010	0.017	0.035	0.039	0.039	0.033	0.014		0.107
4					0.018	0.007	0.016	0.111	0.068	0.115	0.048	0.057	0.112	0.097	0.096	0.130	0.109	0.066		0.130
5						0.311	0.330	0.090	0.066	0.091	0.133	0.051	0.054	0.099	0.064	0.136	0.118	0.082		0.330
6							0.175	0.213	0.149	0.192	0.275	0.103	0.099	0.216	0.112	0.277	0.225	0.142		0.277
7								0.126	0.072	0.088	0.126	0.055	0.087	0.094	0.075	0.134	0.115	0.077		0.134
8									0.133	0.239	0.210	0.118	0.109	0.216	0.117	0.293	0.239	0.161		0.293
9										0.276	0.384	0.147	0.186	0.302	0.138	0.404	0.339	0.223		0.404
10											0.395	0.194	0.153	0.458	0.255	0.592	0.502	0.295		0.592
11												0.245	0.306	0.561	0.117	0.756	0.583	0.452		0.756
12													0.105	0.193	0.117	0.272	0.209	0.154		0.272
13														0.176	0.201	0.221	0.162	0.156		0.221
14															0.227	0.623	0.518	0.316		0.623
15																0.303	0.375	0.123		0.375
16																	0.685	0.423		0.685
17																		0.345		0.345
18																				0.000

Variable 11

CDCIAGRU

75.6%

Variable 16

CDCANALD

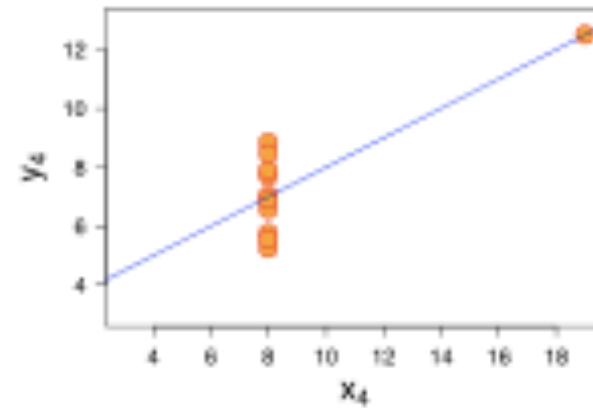
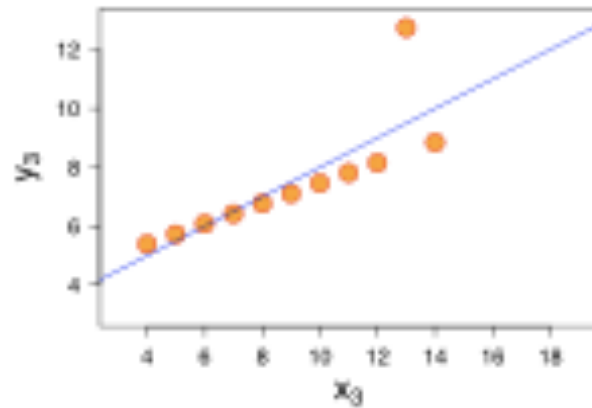
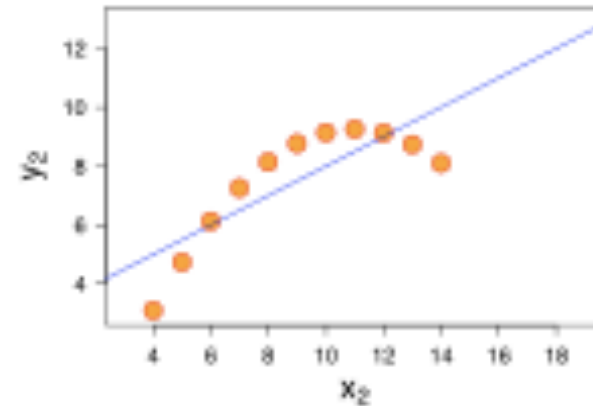
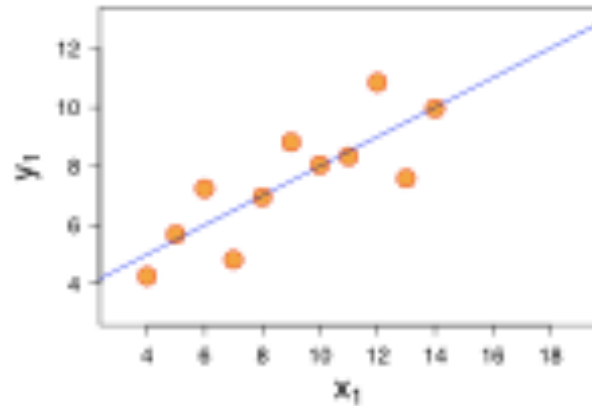
68.5%

Variable 17

CDANULA

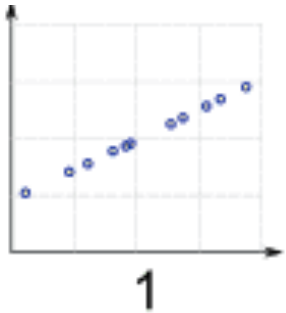
# 1. Análisis correlacional de datos

Es importante una evaluación visual de los datos.

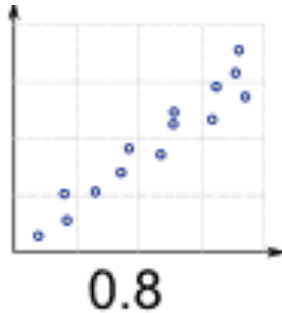


# 1. Análisis correlacional de datos

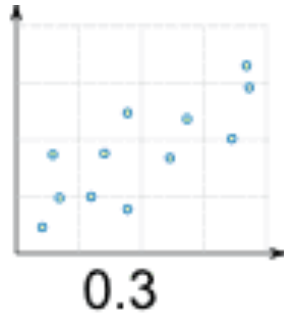
**Correlación perfecta positiva**



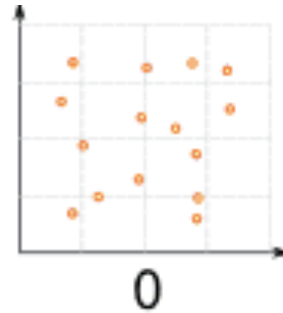
**Alta correlación positiva**



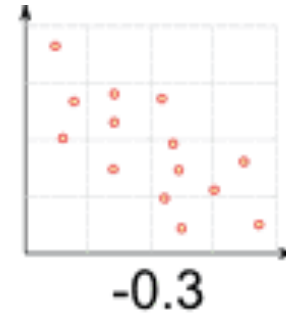
**Baja correlación positiva**



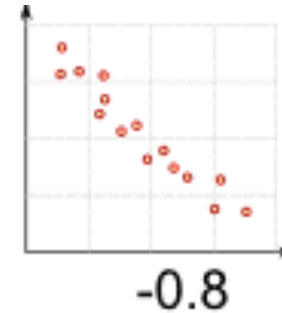
**Sin correlación**



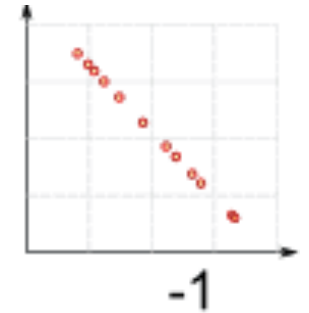
**Baja correlación negativa**



**Alta correlación negativa**



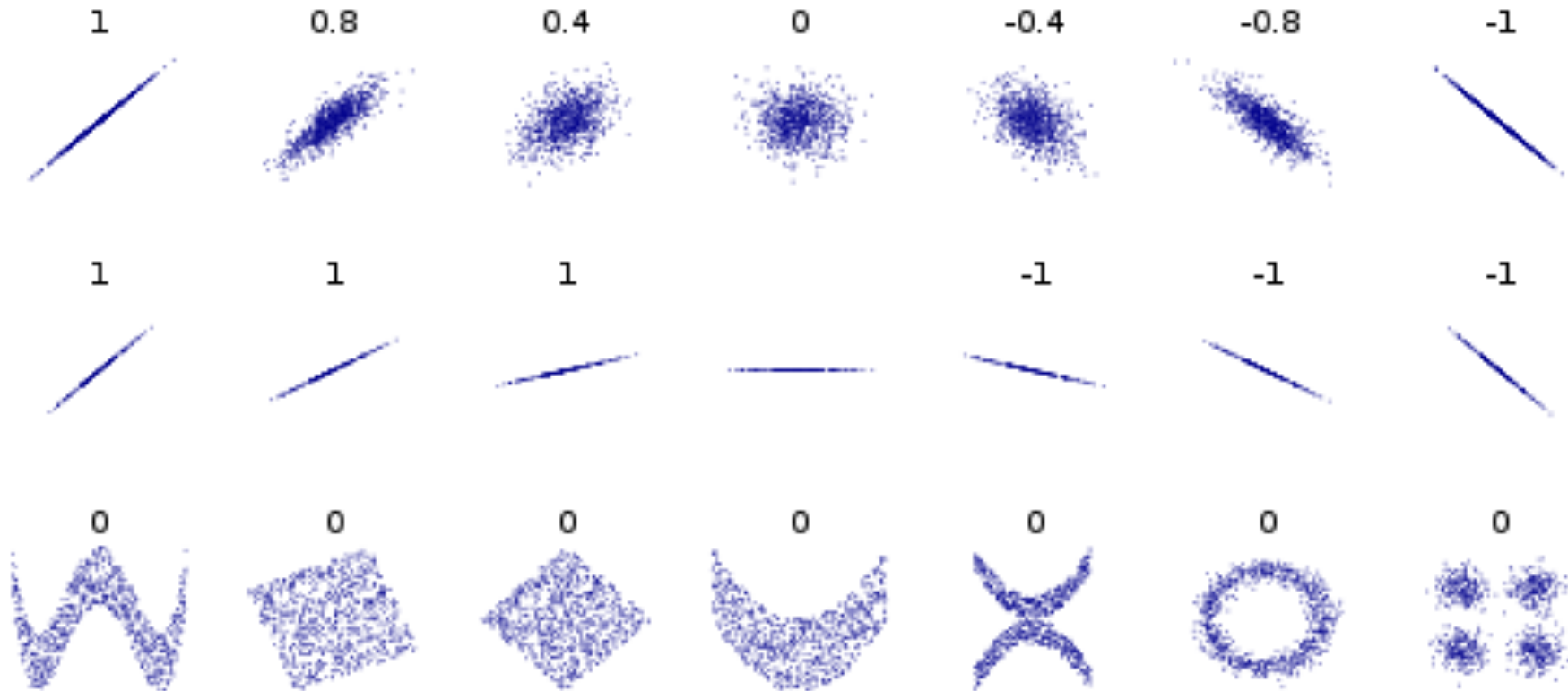
**Correlación perfecta negativa**





# 1. Análisis correlacional de datos

Correlaciones diversas



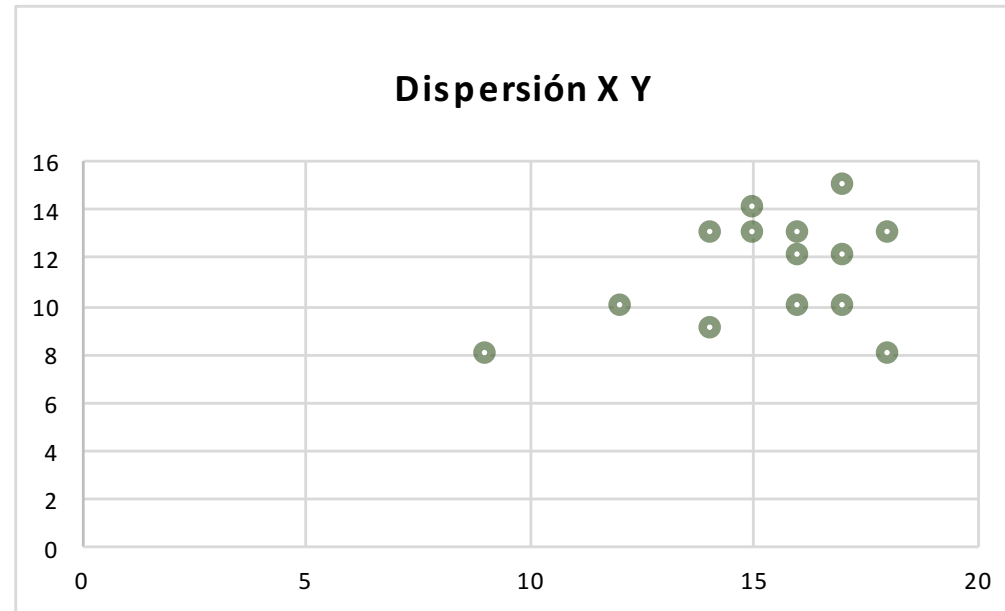
## **Ejemplo ilustrativo**

## Ejemplo ilustrativo

Sean las temperaturas de dos ciudades (**X** –Ciudad de México–, **Y** –Puebla–), determinar el coeficiente de correlación de Pearson:

Día	X	Y
Día 1	18	13
Día 2	17	15
Día 3	15	14
Día 4	16	13
Día 5	14	9
Día 6	12	10
Día 7	9	8
Día 8	15	13
Día 9	16	12
Día 10	14	13
Día 11	16	10
Día 12	18	8
Día 13	17	10
Día 14	17	12

Diagrama de dispersión



## Ejemplo ilustrativo

Sean las temperaturas de dos ciudades (**X** –Ciudad de México–, **Y** –Puebla–), determinar el coeficiente de correlación de Pearson:

Día	X	Y	x = X-X'	y = Y-Y'	x <sup>2</sup>	y <sup>2</sup>	xy
Día 1	18	13	2.71	1.57	7.37	2.47	4.27
Día 2	17	15	1.71	3.57	2.94	12.76	6.12
Día 3	15	14	-0.29	2.57	0.08	6.61	-0.73
Día 4	16	13	0.71	1.57	0.51	2.47	1.12
Día 5	14	9	-1.29	-2.43	1.65	5.90	3.12
Día 6	12	10	-3.29	-1.43	10.80	2.04	4.69
Día 7	9	8	-6.29	-3.43	39.51	11.76	21.55
Día 8	15	13	-0.29	1.57	0.08	2.47	-0.45
Día 9	16	12	0.71	0.57	0.51	0.33	0.41
Día 10	14	13	-1.29	1.57	1.65	2.47	-2.02
Día 11	16	10	0.71	-1.43	0.51	2.04	-1.02
Día 12	18	8	2.71	-3.43	7.37	11.76	-9.31
Día 13	17	10	1.71	-1.43	2.94	2.04	-2.45
Día 14	17	12	1.71	0.57	2.94	0.33	0.98
<b>Total</b>	<b>214</b>	<b>160</b>			<b>78.86</b>	<b>65.43</b>	<b>26.29</b>
<b>Media (X')</b>	<b>15.29</b>						
<b>Meda (Y')</b>	<b>11.43</b>						

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$

$$r = \frac{26.29}{\sqrt{78.86 * 65.43}} = \frac{26.29}{71.83} = 0.36$$

**¿Qué pasa con variables cualitativas?**

# Variables cualitativas

## En el caso de variables cualitativas

Pacientes, 7 variables:

- **Capacidad.** Capacidad del paciente para acudir a una consulta. (1-10)
- **Necesidad.** Importancia que le da el paciente a la consulta médica. (1-10)
- **Transporte.** Disponibilidad de transporte del paciente. (1-10)
- **Cuidado.** Disponibilidad para tener el cuidado de los niños. (1-10)
- **Permiso.** En caso de trabajar, facilidad para solicitar permisos médicos. (1-10)
- **Satisfacción.** Satisfacción del cliente con la atención médica. (1-10)
- **Facilidad.** Facilidad para obtener una cita y eficiencia de la misma. (1-10)
- **Visita.** Visita del paciente durante el último año (0 - no visitó, 1 - si visitó)

## Variables cualitativas

### En el caso de variables cualitativas

	Capacidad	Importancia	Transporte	Cuidado	Permiso	Satisfacción	Facilidad	Visita
Capacidad	1							
Importancia	-0.737	1						
Transporte	0.312	-0.104	1					
Cuidado	0.312	-0.104	0.379	1				
Permiso	0.277	0.060	0.623	0.623	1			
Satisfacción	0.220	-0.134	0.654	0.654	0.626	1		
Facilidad	0.389	-0.033	0.650	0.650	0.659	0.896	1	
Visita	0.396	-0.542	-0.503	-0.503	-0.425	-0.399	-0.328	1

- **R1.** Existe una relación fuerte (negativa) entre la **capacidad** que tiene el paciente para acudir a una consulta y la **importancia** que le da el paciente a la consulta médica.
- **R2.** Se tiene una relación fuerte (positiva) entre la **satisfacción** del paciente con la atención médica y la **facilidad** que tiene para obtener una cita.



# Práctica 3

## Fuente de datos

Estudios clínicos a partir de imágenes digitalizadas de pacientes con cáncer de mama de Wisconsin (WDBC, Wisconsin Diagnostic Breast Cancer)

Variable	Descripción	Tipo
Radius	Media de las distancias del centro y puntos del perímetro	Continuo
Texture	Desviación estándar de la escala de grises	Continuo
Perimeter	Valor del perímetro del cáncer de mama	Continuo
Area	Valor del área del cáncer de mama	Continuo
Smoothness	Variación de la longitud del radio	Continuo
Compactness	$\text{Perímetro}^2 / \text{Área} - 1$	Continuo
Concavity	Caída o gravedad de las curvas de nivel	Continuo
Concave points	Número de sectores de contorno cóncavo	Continuo
Symmetry	Simetría de la imagen	Continuo
Fractal dimension	"Aproximación de frontera" - 1	Continuo

**Fuente:** [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))

## Paso 1: Acceso a los datos

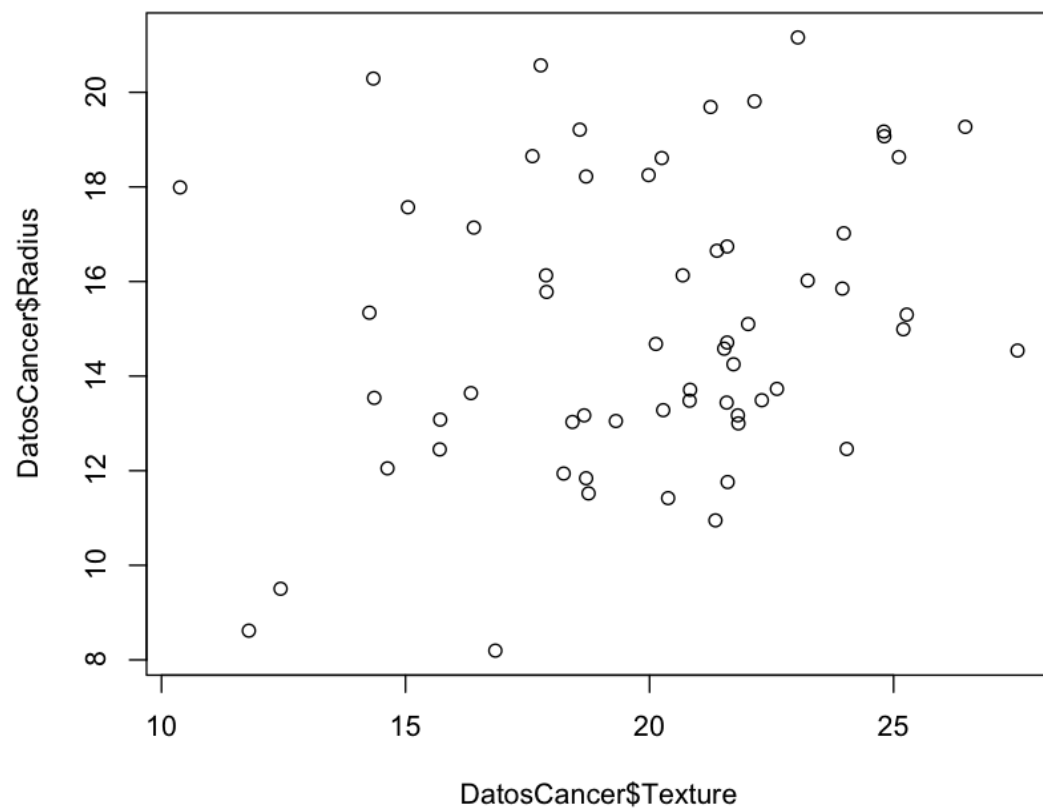
```
DatosCancer <- read.table("/Users/guille/Documents/1 FI-UNAM/1 Cursos/2021-1/1 IA2021-1/2 CasosPracticos/2  
Correlaciones/1Cancer.txt", header=T, sep="\t")
```

DatosCancer

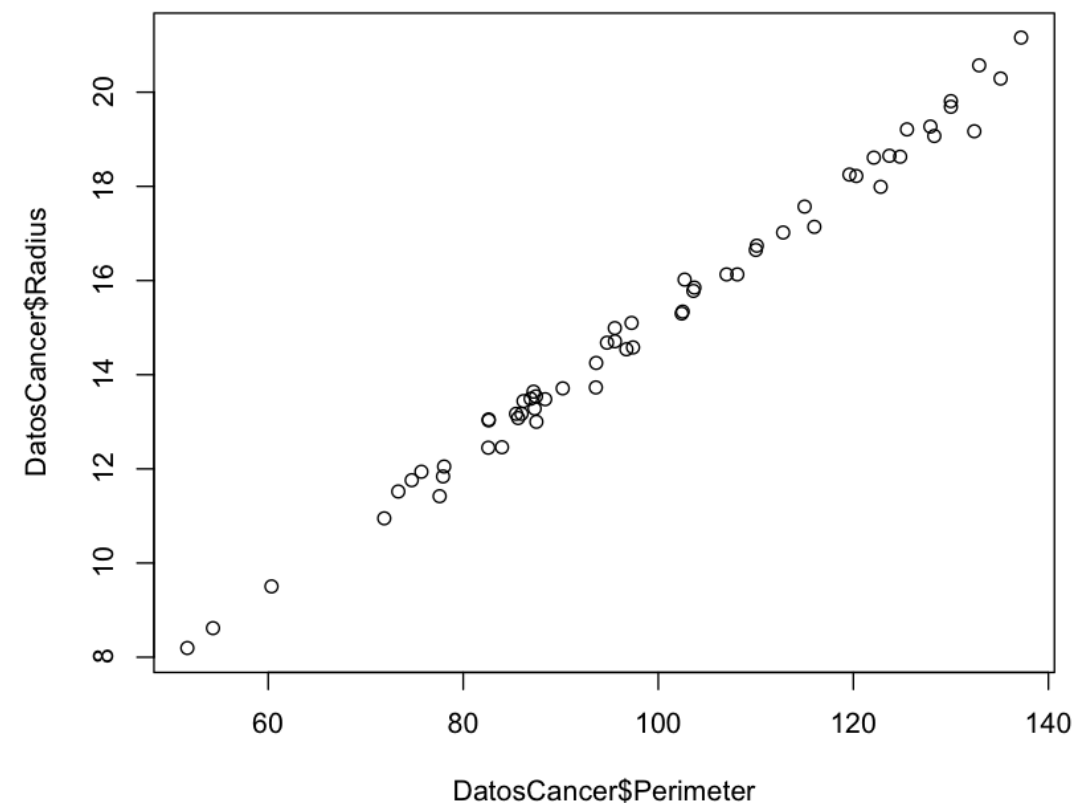
	Radius	Texture	Perimeter	Area	Smoothness	Compactness	Concavity	ConcavePoints	Symmetry	FractalDimension
1	17.990	10.38	122.80	1001.0	0.11840	0.27760	0.300100	0.147100	0.2419	0.07871
2	20.570	17.77	132.90	1326.0	0.08474	0.07864	0.086900	0.070170	0.1812	0.05667
3	19.690	21.25	130.00	1203.0	0.10960	0.15990	0.197400	0.127900	0.2069	0.05999
4	11.420	20.38	77.58	386.1	0.14250	0.28390	0.241400	0.105200	0.2597	0.09744
5	20.290	14.34	135.10	1297.0	0.10030	0.13280	0.198000	0.104300	0.1809	0.05883
6	12.450	15.70	82.57	477.1	0.12780	0.17000	0.157800	0.080890	0.2087	0.07613
7	18.250	19.98	119.60	1040.0	0.09463	0.10900	0.112700	0.074000	0.1794	0.05742
8	13.710	20.83	90.20	577.9	0.11890	0.16450	0.093660	0.059850	0.2196	0.07451
9	13.000	21.82	87.50	519.8	0.12730	0.19320	0.185900	0.093530	0.2350	0.07389
10	12.460	24.04	83.97	475.9	0.11860	0.23960	0.227300	0.085430	0.2030	0.08243
11	16.020	23.24	102.70	797.8	0.08206	0.06669	0.032990	0.033230	0.1528	0.05697
12	15.780	17.89	103.60	781.0	0.09710	0.12920	0.099540	0.066060	0.1842	0.06082

## Paso 2: Evaluación visual

`plot(DatosCancer$Radius~DatosCancer$Texture)`

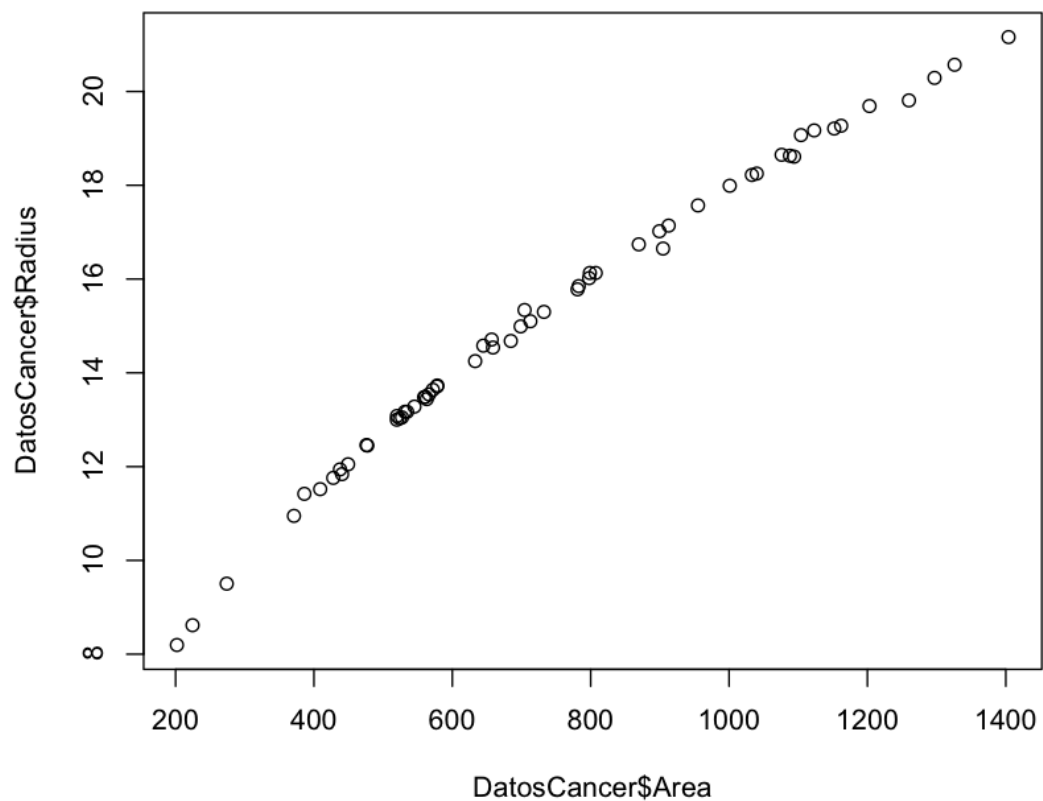


`plot(DatosCancer$Radius~DatosCancer$Perimeter)`

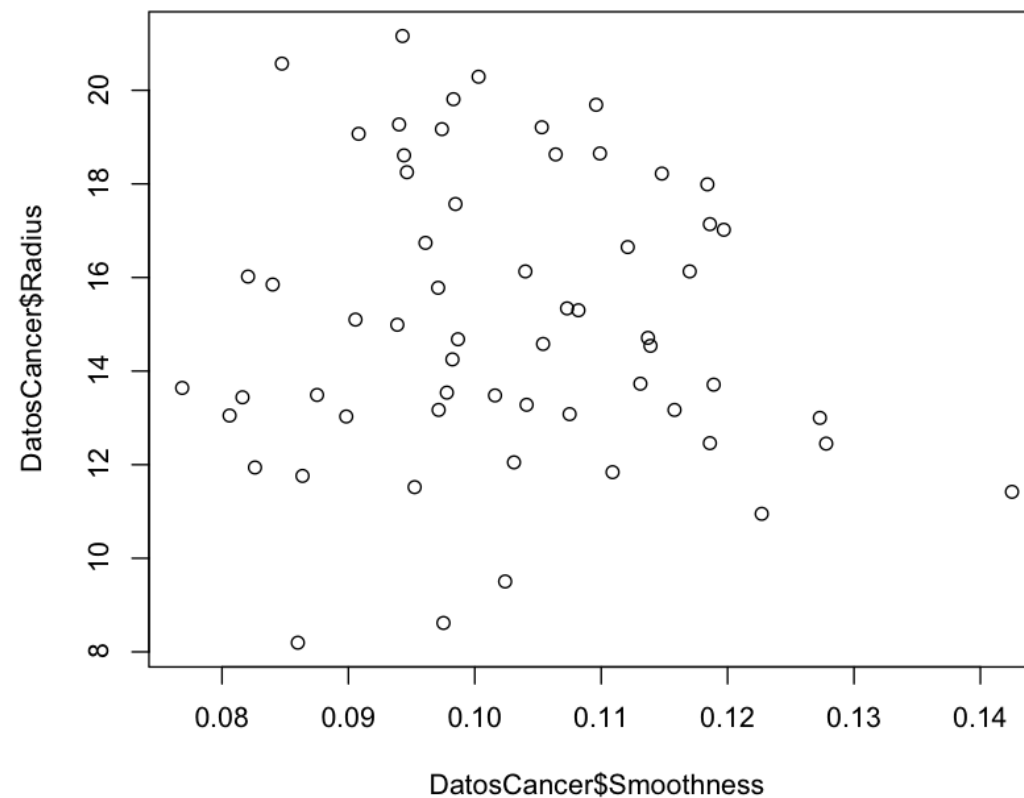


## Paso 2: Evaluación visual

```
plot(DatosCancer$Radius~DatosCancer$Area)
```

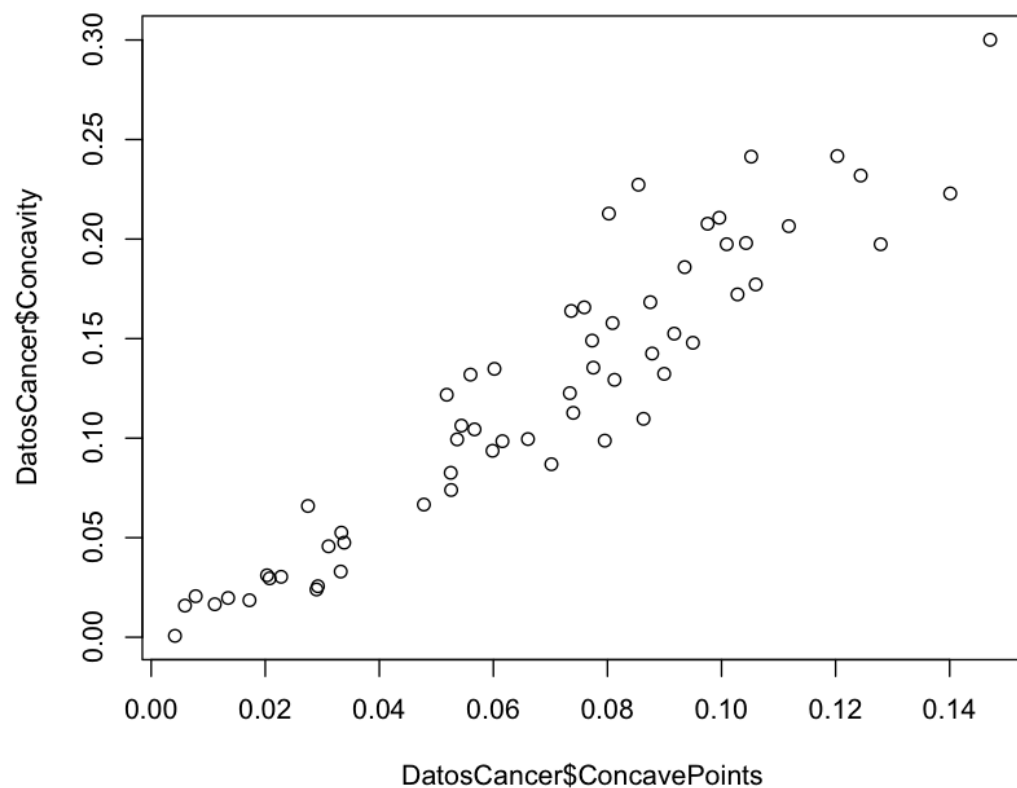


```
plot(DatosCancer$Radius~DatosCancer$Smoothness)
```

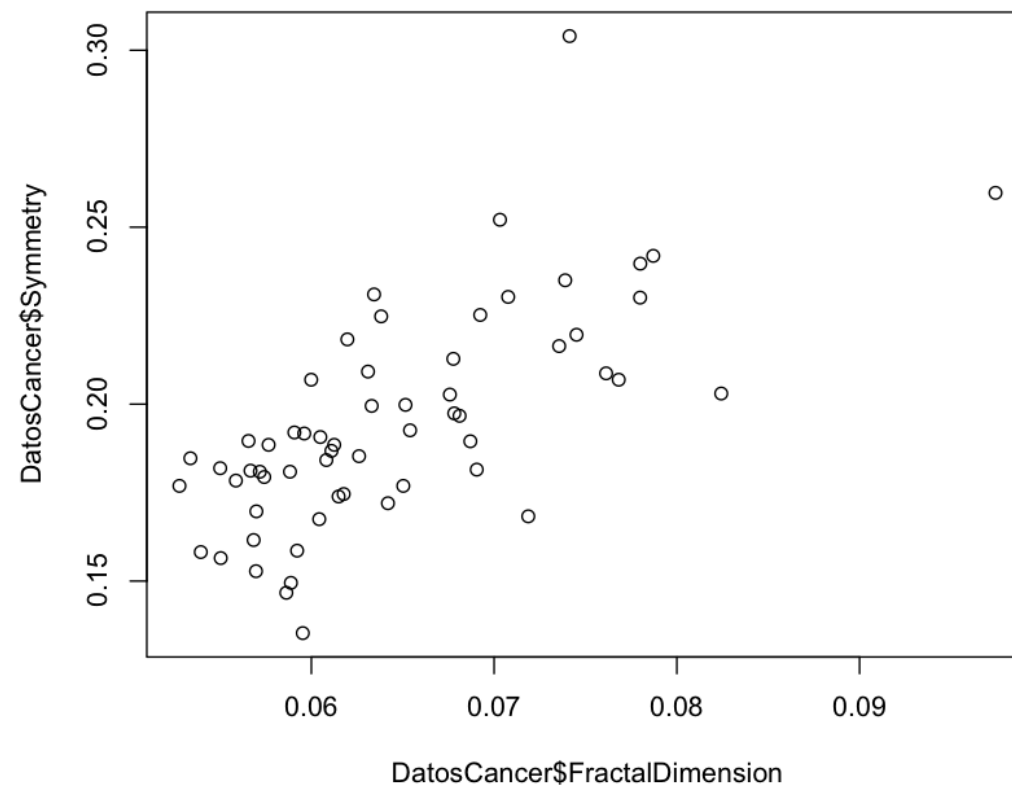


## Paso 2: Evaluación visual

```
plot(DatosCancer$Concavity~DatosCancer$ConcavePoints)
```



```
plot(DatosCancer$Symmetry~DatosCancer$FractalDimension)
```



## Paso 3: Obtención de correlaciones

```
MCorrelaciones <- cor(DatosCancer)
```

MCorrelaciones

	Radius	Texture	Perimeter	Area	Smoothness	Compactness	Concavity	ConcavePoints	Symmetry	FractalDimension
Radius	1.00000000	0.24434333	0.9959636052	0.99220293	-0.0614966710	0.2935017	0.5082676	0.6679423	0.11492876	-0.3263750
Texture	0.24434333	1.00000000	0.2497315308	0.21296532	-0.0382080720	0.1356041	0.1663960	0.1327581	0.01340155	-0.1096147
Perimeter	0.99596361	0.24973153	1.0000000000	0.98552330	-0.0008416856	0.3728748	0.5748909	0.7204580	0.18326569	-0.2534726
Area	0.99220293	0.21296532	0.9855232974	1.00000000	-0.0810373416	0.2567310	0.4843345	0.6464840	0.08306744	-0.3391931
Smoothness	-0.06149667	-0.03820807	-0.0008416856	-0.08103734	1.0000000000	0.7214197	0.6729152	0.6162395	0.67953080	0.7605514
Compactness	0.29350174	0.13560407	0.3728747717	0.25673095	0.7214197469	1.00000000	0.9082936	0.8193251	0.81140166	0.7376877
Concavity	0.50826756	0.16639599	0.5748908650	0.48433453	0.6729152399	0.9082936	1.00000000	0.9398877	0.71730711	0.5111389
ConcavePoints	0.66794234	0.13275808	0.7204579967	0.64648400	0.6162395155	0.8193251	0.9398877	1.00000000	0.67547335	0.3596574
Symmetry	0.11492876	0.01340155	0.1832656888	0.08306744	0.6795307970	0.8114017	0.7173071	0.6754734	1.00000000	0.6896536
FractalDimension	-0.32637502	-0.10961472	-0.2534725532	-0.33919311	0.7605514200	0.7376877	0.5111389	0.3596574	0.68965360	1.00000000



## Paso 3: Obtención de correlaciones

```
MCorrelaciones <- cor(DatosCancer)
```

MCorrelaciones

	Radius	Texture	Perimeter	Area	Smoothness	Compactness	Concavity	ConcavePoints	Symmetry	FractalDimension
Radius	1.00000000	0.24434333	0.9959636052	0.99220293	-0.0614966710	0.2935017	0.5082676	0.6679423	0.11492876	-0.3263750
Texture	0.24434333	1.00000000	0.2497315308	0.21296532	-0.0382080720	0.1356041	0.1663960	0.1327581	0.01340155	-0.1096147
Perimeter	0.99596361	0.24973153	1.0000000000	0.98552330	-0.0008416856	0.3728748	0.5748909	0.7204580	0.18326569	-0.2534726
Area	0.99220293	0.21296532	0.9855232974	1.00000000	-0.0810373416	0.2567310	0.4843345	0.6464840	0.08306744	-0.3391931
Smoothness	-0.06149667	-0.03820807	-0.0008416856	-0.08103734	1.0000000000	0.7214197	0.6729152	0.6162395	0.67953080	0.7605514
Compactness	0.29350174	0.13560407	0.3728747717	0.25673095	0.7214197469	1.00000000	0.9082936	0.8193251	0.81140166	0.7376877
Concavity	0.50826756	0.16639599	0.5748908650	0.48433453	0.6729152399	0.9082936	1.00000000	0.9398877	0.71730711	0.5111389
ConcavePoints	0.66794234	0.13275808	0.7204579967	0.64648400	0.6162395155	0.8193251	0.9398877	1.00000000	0.67547335	0.3596574
Symmetry	0.11492876	0.01340155	0.1832656888	0.08306744	0.6795307970	0.8114017	0.7173071	0.6754734	1.00000000	0.6896536
FractalDimension	-0.32637502	-0.10961472	-0.2534725532	-0.33919311	0.7605514200	0.7376877	0.5111389	0.3596574	0.68965360	1.00000000

## Paso 3: Obtención de correlaciones

```
MCorrelaciones <- round(cor(DatosCancer), 3)
```

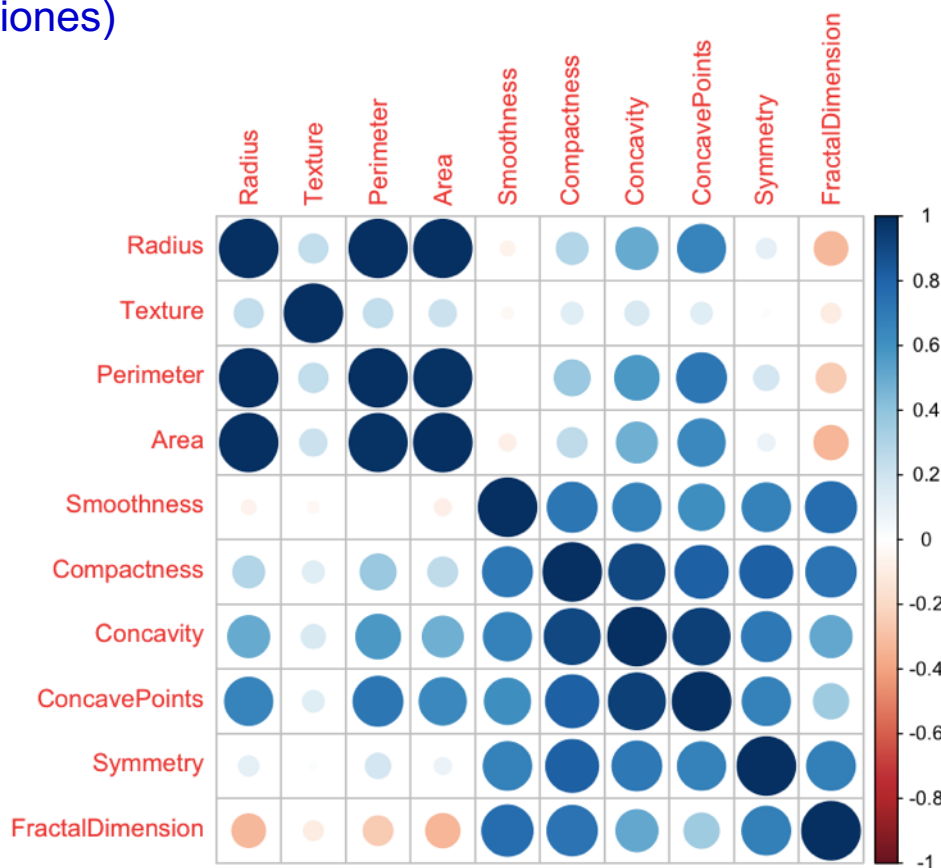
MCorrelaciones

	Radius	Texture	Perimeter	Area	Smoothness	Compactness	Concavity	ConcavePoints	Symmetry	FractalDimension
Radius	1.000	0.244	0.996	0.992	-0.061	0.294	0.508	0.668	0.115	-0.326
Texture	0.244	1.000	0.250	0.213	-0.038	0.136	0.166	0.133	0.013	-0.110
Perimeter	0.996	0.250	1.000	0.986	-0.001	0.373	0.575	0.720	0.183	-0.253
Area	0.992	0.213	0.986	1.000	-0.081	0.257	0.484	0.646	0.083	-0.339
Smoothness	-0.061	-0.038	-0.001	-0.081	1.000	0.721	0.673	0.616	0.680	0.761
Compactness	0.294	0.136	0.373	0.257	0.721	1.000	0.908	0.819	0.811	0.738
Concavity	0.508	0.166	0.575	0.484	0.673	0.908	1.000	0.940	0.717	0.511
ConcavePoints	0.668	0.133	0.720	0.646	0.616	0.819	0.940	1.000	0.675	0.360
Symmetry	0.115	0.013	0.183	0.083	0.680	0.811	0.717	0.675	1.000	0.690
FractalDimension	-0.326	-0.110	-0.253	-0.339	0.761	0.738	0.511	0.360	0.690	1.000

## Paso 4: Visualización de correlaciones

```
library(corrplot)
```

```
corrplot(MCorrelaciones)
```

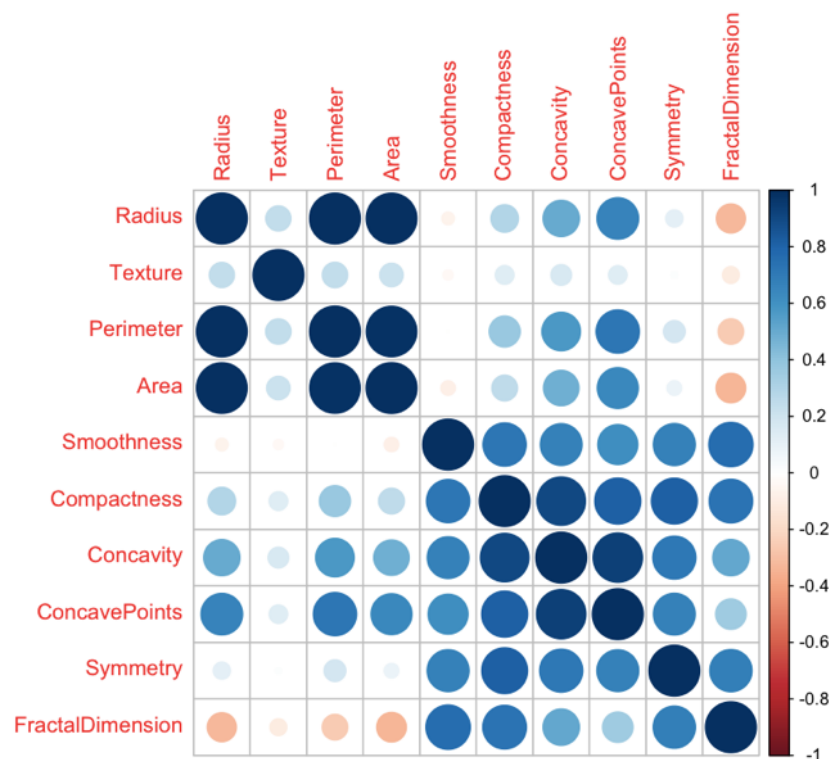


- La función **corrplot()** es útil para trazar gráficas de una matriz de correlaciones.
- Las correlaciones positivas se muestran en azul y las negativas en rojo.
- La intensidad del color y el tamaño del círculo son proporcionales a los valores de los coeficientes de correlación.

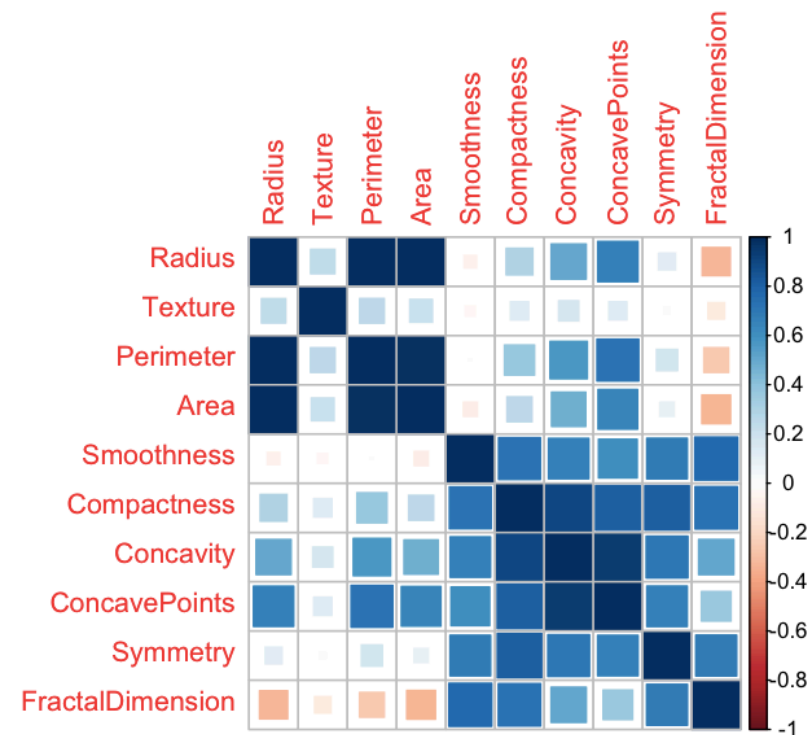
## Paso 4: Visualización de correlaciones

Siete métodos diferentes de **visualización**: "circle", "square", "ellipse", "number", "shade", "color", "pie"

`corrplot(MCorrelaciones, method="circle")`



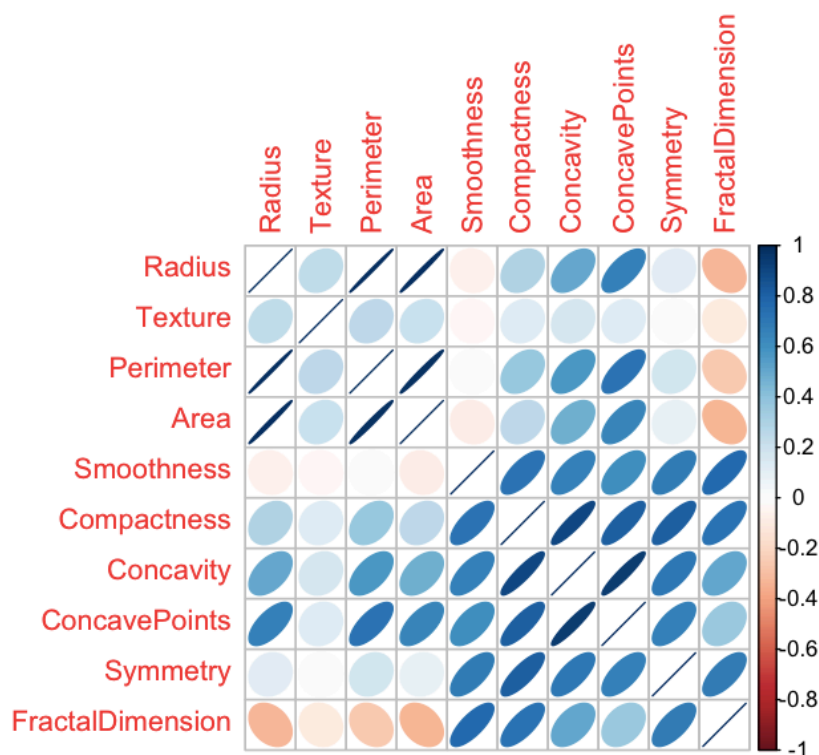
`corrplot(MCorrelaciones, method="square")`



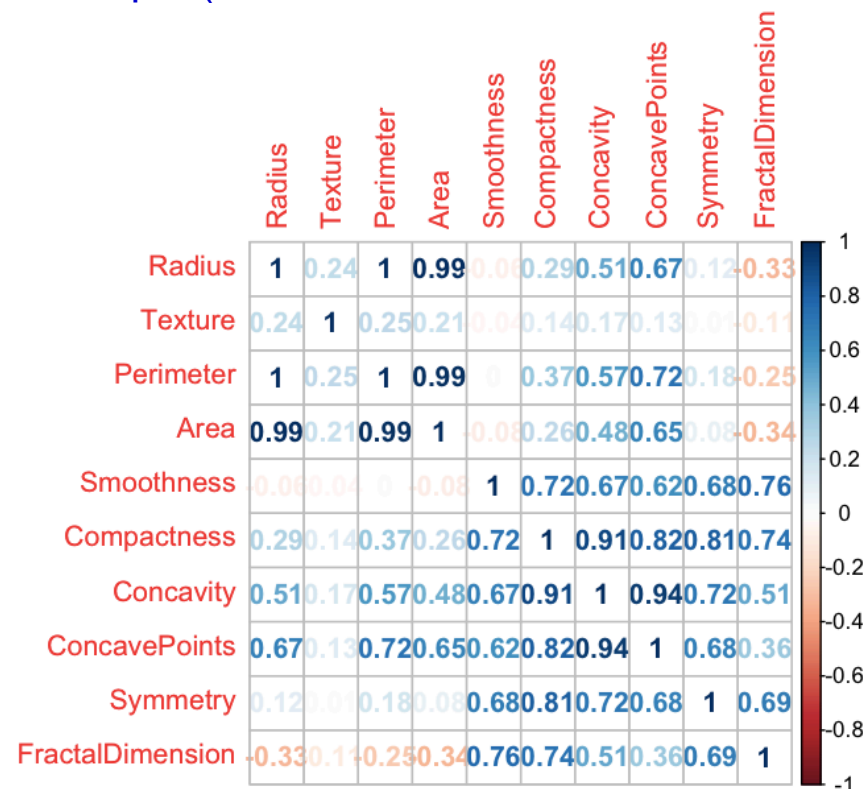
## Paso 4: Visualización de correlaciones

Siete diferentes métodos de **visualization**: "circle", "square", "ellipse", "number", "shade", "color", "pie"

`corrplot(MCorrelaciones, method="ellipse")`



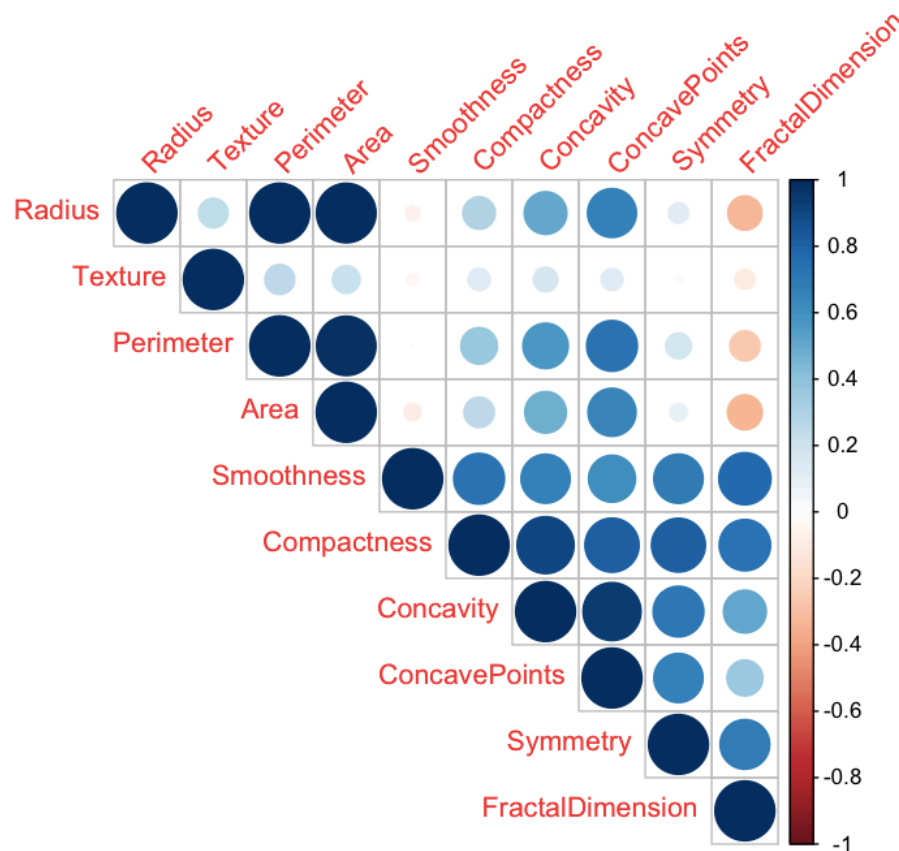
`corrplot(MCorrelaciones, method="number")`



## Paso 4: Visualización de correlaciones

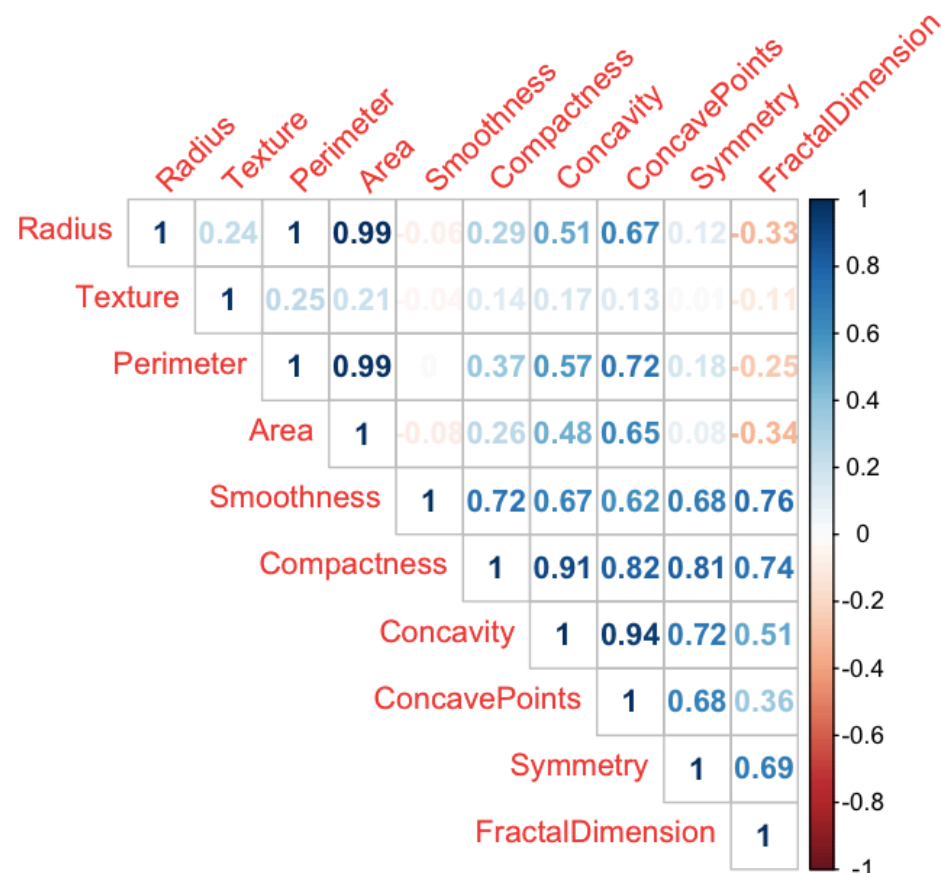
```
corrplot(MCorrelaciones, type="upper", tl.srt=45)
```

text label string rotation



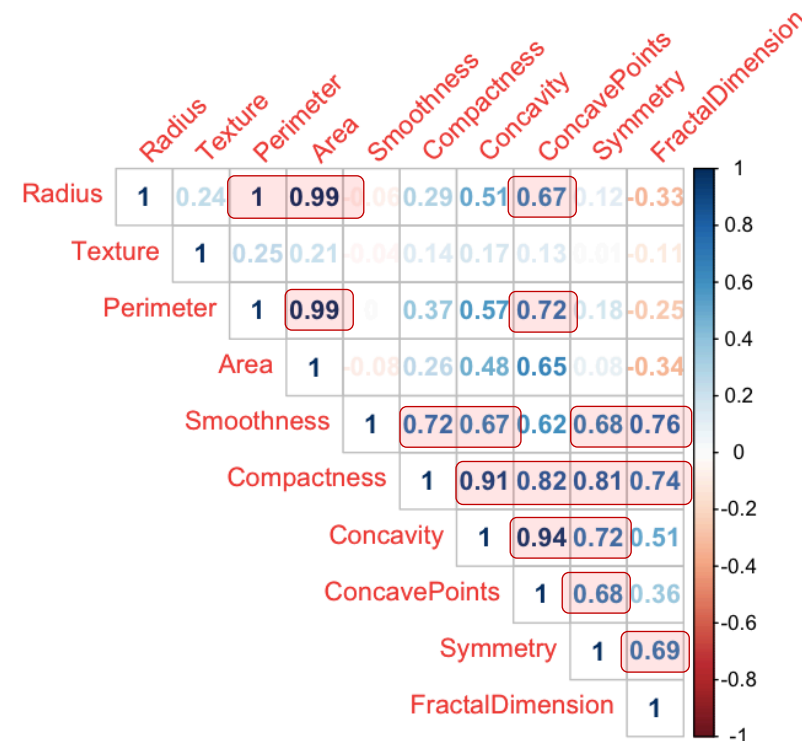
## Paso 4: Visualización de correlaciones

```
corrplot(MCorrelaciones, method="number", type="upper", tl.srt=45)
```





## Paso 5: Elección de variables



	1	2	3	4	5	6	7	8	9	10
	Radius	Texture	Perimeter	Area	Smoothness	Compactness	Concavity	ConcavePoints	Symmetry	FractalDimension
1	17.990	10.38	122.80	1001.0	0.11840	0.27760	0.300100	0.147100	0.2419	0.07871
2	20.570	17.77	132.90	1326.0	0.08474	0.07864	0.086900	0.070170	0.1812	0.05667
3	19.690	21.25	130.00	1203.0	0.10960	0.15990	0.197400	0.127900	0.2069	0.05999
4	11.420	20.38	77.58	386.1	0.14250	0.28390	0.241400	0.105200	0.2597	0.09744
5	20.290	14.34	135.10	1297.0	0.10030	0.13280	0.198000	0.104300	0.1809	0.05883
6	12.450	15.70	82.57	477.1	0.12780	0.17000	0.157800	0.080890	0.2087	0.07613
7	18.250	19.98	119.60	1040.0	0.09463	0.10900	0.112700	0.074000	0.1794	0.05742
8	13.710	20.83	90.20	577.9	0.11890	0.16450	0.093660	0.059850	0.2196	0.07451
9	13.000	21.82	87.50	519.8	0.12730	0.19320	0.185900	0.093530	0.2350	0.07389
10	12.460	24.04	83.97	475.9	0.11860	0.23960	0.227300	0.085430	0.2030	0.08243
11	16.020	23.24	102.70	797.8	0.08206	0.06669	0.032990	0.033230	0.1528	0.05697
12	15.780	17.89	103.60	781.0	0.09710	0.12920	0.099540	0.066060	0.1842	0.06082