

```
%Fecha de actualizacion 14/abril/2023
clc
clear all
close all
```

Portada

```
figure(1);
% tamaño de la figura
% set(gcf, 'Position', [100, 100, 800, 600]);
tamano_letra=10;
% Cargar y mostrar la primera imagen
img1 = imread('UNAMLogo.png'); % Reemplaza 'Imagen1.jpg' con el nombre de tu primera imagen
subplot(4, 4, 1); % Divide la figura en 1 fila y 2 columnas, y selecciona la primera posición
imshow(img1);

% Cargar y mostrar la segunda imagen
img2 = imread('FIlogo.jpg'); % Reemplaza 'Imagen2.jpg' con el nombre de tu segunda imagen
subplot(4, 4, 4); % Selecciona la segunda posición
imshow(img2);

% Agregar texto a un subplot específico (por ejemplo, el subplot 8)
subplot(4, 4, [2,3]);
text(0.5, 0.9, 'UNAM', 'FontSize', 14,'HorizontalAlignment', 'center','Color','b');
text(0.5, 0.5, 'Facultad de ingeniería', 'FontSize', tamano_letra, 'FontWeight', 'bold', 'HorizontalAlignment', 'center');
text(0.5, 0.0, 'TSISB', 'FontSize', tamano_letra,'HorizontalAlignment', 'center');
axis off
subplot(4, 4, [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]);
text(0.5, 0.9, 'Alumna: Flores Morín María Alejandra', 'FontSize', tamano_letra,'HorizontalAlignment', 'center');
text(0.5, 0.8, 'No. Cuenta: 315165805', 'FontSize', tamano_letra,'HorizontalAlignment', 'center');
text(0.5, 0.6, 'Profesor: Dr. en C. Luis Antonio Aguilar Pérez ', 'FontSize', tamano_letra,'HorizontalAlignment', 'center');
text(0.5, 0.4, 'Semestre 2024-1', 'FontSize', tamano_letra,'HorizontalAlignment', 'center');
text(0.5, 0.3, 'tareaEjercicio1 Ejercicios de DATASTORE en MATLAB', 'FontSize', tamano_letra,'HorizontalAlignment', 'center');
text(0.5, 0.1, 'Fecha de entrega: 04 de octubre del 2023', 'FontSize', tamano_letra,'HorizontalAlignment', 'center');
axis off
```



UNAM
Facultad de ingeniería
TSISB



Alumna: Flores Morín María Alejandra

No. Cuenta: 315165805

Profesor: Dr. en C. Luis Antonio Aguilar Pérez

Semestre 2024-1

tareaEjercicio1 Ejercicios de DATASTORE en MATLAB

Fecha de entrega: 04 de octubre del 2023

Análisis exploratorio de una base de datos

```
%%%%
%En este ejercicio solo estaremos trabajando con un archivo
%contenido dentro del datastore
%%%%

%%
%Creacion de directorio de trabajo
%rootFolder = 'D:\TSISB_IA';
rootFolder = 'C:\Users\puma_\Documents\TSISB_IA';
workingFolder = 'practica_3';
tempFolder = 'temp';
savePath = fullfile(rootFolder,workingFolder);
saveTempPath = fullfile(rootFolder,workingFolder,tempFolder);

prefix = ['\' 'heart'];
sufix = '.csv';
```

```

newName = [prefix, sufix];

%Despues de correr esta celda una vez, crea un bloque de comentarios a partir de esta linea

if ~exist(savePath,'dir')
    [status, message, ~] = mkdir(savePath);
    if status == 0
        disp(message)
    end
end

if ~exist(saveTempPath,'dir')
    [status, message, ~] = mkdir(saveTempPath);
    if status == 0
        disp(message)
    end
end

%%%
%Organizacion y copia de archivos
[fileName, pathFileName] = uigetfile('C:\','*.txt');

if isequal(fileName,0)
    disp('Se canceló la busqueda de archivos');
else
    disp(['El usuario seleccionó el archivo ', fullfile(pathFileName,fileName)]);
    [status, message, ~] = copyfile([pathFileName,fileName],[saveTempPath,newName]);
    prefix = ['heart'];
    sufix = '.csv';
    newName = [prefix, sufix];
    if status == 0
        disp(message)
    else
        disp(['El cual se movio a la dirección ', fullfile(saveTempPath)]);
    end
end
end

```

El usuario seleccionó el archivo C:\Users\puma_\Documents\TSISB_IA\practica_3\heart.csv
The process cannot access the file because it is being used by another process.

%}

Análisis exploratorio de datos

```
%Se determina la direccion de los archivos que integran el "Datastore"
archivoCSV = "heart.csv"
```

archivoCSV = "heart.csv"

```
%Visualización rápida de las primeras lineas del archivo de texto
%Como solo es un archivo se puede utilizar el comando dbtype
dbtype("heart.csv",'1:5')
```

```
1 age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,oldpeak,slope,ca,thal,target
2 52,1,0,125,212,0,1,168,0,1,2,2,3,0
3 53,1,0,140,203,1,0,155,1,3,1,0,0,3,0
4 70,1,0,145,174,0,1,125,1,2,6,0,0,3,0
5 61,1,0,148,203,0,1,161,0,0,2,1,3,0
```

```
%Visualizacion general de las opciones de importacion de los archivos
%dentro del datastore
opts = detectImportOptions("heart.csv")
```

opts =
DelimitedTextImportOptions with properties:

Format Properties:
Delimiter: {','}
Whitespace: '\b\t '
LineEnding: {'\n' '\r' '\r\n'}
CommentStyle: {}
ConsecutiveDelimitersRule: 'split'
LeadingDelimitersRule: 'keep'
TrailingDelimitersRule: 'ignore'
EmptyLineRule: 'skip'
Encoding: 'UTF-8'

Replacement Properties:

```
%Visualizacion de los nombres y tipo de variables de las columnas
disp([opts.VariableNames' opts.VariableTypes'])
```

```

{'age' } {'double'}
{'sex' } {'double'}
{'cp' } {'double'}
{'trestbps'} {'double'}
{'chol' } {'double'}
{'fbs' } {'double'}
{'restecg' } {'double'}
{'thalach' } {'double'}
{'exang' } {'double'}
{'oldpeak' } {'double'}
{'slope' } {'double'}
{'ca' } {'double'}
{'thal' } {'double'}
{'target' } {'double'}

```

```

%Desde la version 2019, los nombres de las variables pueden incluir
%cualquier tipo de simbolos, ademas de no necesariamente comenzar solo
%con letras, por lo que matlab requiere el Flag "preserve" para considerar
%esta opcion
opts.VariableNamingRule='preserve';

```

Tabla de resumen de los datos

```

%Asignacion a una variable particular
values=opts.VariableNames;
%Modificacion como si fuera un indexado de valores clasico
newValues={'Edad','Sexo','Tipo Dolor','Presión arterial en reposo [mmHg]','Colesterol sérico [mg/dl]','Azúcar en sangre (ayunas) [mg/dl]',...
opts.VariableNames = newValues;
Heart_1 = readtable(archivoCSV,opts);
head(Heart_1)

```

Edad	Sexo	Tipo Dolor	Presión arterial en reposo [mmHg]	Colesterol sérico [mg/dl]	Azúcar en sangre
52	1	0	125	212	€
53	1	0	140	203	1
70	1	0	145	174	€
61	1	0	148	203	€
62	0	0	138	294	1
58	0	0	100	248	€
58	1	0	114	318	€
55	1	0	160	289	€

```

%Cuantas mujeres y hombres hay
Genero = categorical(Heart_1.Sexo);
grpCatGenero = categories(Genero)

```

```

grpCatGenero = 2x1 cell
'0'
'1'

```

```

nuevosGeneros = {'Mujer',...
    'Hombre'};
nuevoGenero = renamecats(Genero,nuevosGeneros);
summary(nuevoGenero)

```

Mujer	312
Hombre	713

```

%Cuantos sanos y enfermos hay
Estado = categorical(Heart_1.("Enfermedad cardiaca"));
grpCatGenero = categories(Estado);
nuevosEstados = {'No hay enfermedad cardiaca',...
    'Enfermedad cardiaca'};
nuevoEstado = renamecats(Estado,nuevosEstados);
summary(nuevoEstado)

```

No hay enfermedad cardiaca	499
Enfermedad cardiaca	526

Tabla de frecuencias

Principales signos para diagnosticar Enfermedad Cardiaca.

```

Heart_1 = readtable(archivoCSV,opts);
%Eliminacion de caracteristicas de la tabla
Heart_2 = removevars(Heart_1,{'Tipo Dolor','Resultados ECG (0, 1, 2)'}, 'Pendiente segmento ST del má ejercicio', 'No de vasos principales < 0.01');
head(Heart_2)

```

Edad	Sexo	Presión arterial en reposo [mmHg]	Colesterol sérico [mg/dl]	Azúcar en sangre (ayunas) [mg/dl]
—	—	—	—	—

52	1	125	212	0
53	1	140	203	1
70	1	145	174	0
61	1	148	203	0
62	0	138	294	1
58	0	100	248	0
58	1	114	318	0
55	1	160	289	0

```
Edad = categorical(Heart_2.Edad);
summary(Edad)
```

29	4
34	6
35	15
37	6
38	12
39	14
40	11
41	32
42	26
43	26
44	36
45	25
46	23
47	18
48	23

Tabla personas Enfermedad Cardiaca

```
%Filtro para personas con Enfermedad Cardiaca
filtro = (Heart_2."Enfermedad cardiaca") == 1;
Heart_2_Enfermos = Heart_2(filtro, :)
```

Heart_2_Enfermos = 526x9 table

	Edad	Sexo	Presión arterial en reposo [mmHg]	Colesterol sérico [mg/dl]	Azúcar en sangre (ayunas) [mg/dl]	Glucosa en orina [mg/dl]	Presión arterial media [mmHg]	Coeficiente de masa corporal [kg/m²]
1	58	0	100	248				
2	71	0	112		149			
3	34	0	118		210			
4	34	0	118		210			
5	51	0	140		308			
6	50	0	120		244			
7	58	1	140		211			
8	67	0	106		223			
9	45	1	104		208			
10	63	0	135		252			
11	42	0	120		209			
12	44	1	130		233			
13	50	0	120		244			
14	50	1	129		196			
15	51	1	125		213			
16	59	1	138		271			
17	64	1	128		263			
18	65	0	160		360			
19	54	1	120		258			
20	55	0	132		342			
21	42	1	140		226			
22	41	1	135		203			
23	66	0	146		278			
24	58	0	150		283			
25	38	1	138		175			
26	56	1	120		193			
27	48	1	130		245			
28	29	1	130		204			
29	66	0	146		278			
30	59	1	150		212			
31	29	1	130		204			
32	53	1	130		197			
33	37	0	120		215			

	Edad	Sexo	Presión arterial en reposo [mmHg]	Colesterol sérico [mg/dl]	Azúcar en sangre (ayunas) [mg/dl]
34	47	1	138	257	
35	48	1	124	255	
36	52	1	134	201	
37	52	1	134	201	
38	38	1	138	175	
39	37	0	120	215	
40	44	1	120	220	
41	58	1	140	211	
42	54	0	108	267	
43	62	0	124	209	
44	62	1	128	208	
45	45	0	138	236	
46	57	0	128	303	
47	76	0	140	197	
48	43	0	122	213	
49	57	1	150	126	
50	54	1	108	309	
51	47	1	138	257	
52	52	1	118	186	
53	53	1	142	226	
54	29	1	130	204	
55	42	1	120	295	
56	54	1	108	309	
57	65	0	140	417	
58	60	0	150	240	
59	66	1	120	302	
60	53	1	130	197	
61	52	1	138	223	
62	57	1	140	192	
63	60	0	150	240	
64	51	0	130	256	
65	41	1	135	203	
66	50	1	129	196	
67	54	1	108	309	
68	55	0	132	342	
69	64	0	180	325	
70	47	1	138	257	
71	41	1	110	235	
72	34	1	118	182	
73	47	1	112	204	
74	51	0	120	295	
75	52	1	152	298	
76	39	1	140	321	
77	54	1	120	258	
78	54	1	108	309	
79	40	1	140	199	
80	54	1	120	258	
81	67	0	115	564	
82	41	1	120	157	
83	51	1	100	222	
84	57	0	120	354	
85	43	1	130	315	
86	45	0	112	160	
87	43	1	150	247	
88	56	1	120	240	
89	39	0	94	199	
90	64	1	110	211	
91	42	1	130	180	
92	45	1	128	308	
93	41	1	112	250	

	Edad	Sexo	Presión arterial en reposo [mmHg]	Colesterol sérico [mg/dl]	Azúcar en sangre (ayunas) [mg/d
94	56	1	130	221	
95	67	0	115	564	
96	69	1	160	234	
97	59	1	150	212	
98	45	1	115	260	
99	60	0	102	318	
100	62	0	124	209	
:					

frecuencias

```
%Cuantas personas enfermas hay TOTAL
```

```
PersonasEnfermasTotal=sum(Heart_2_Enfermos.( "Enfermedad cardíaca"))
```

```
PersonasEnfermasTotal = 526
```

```
%Principales Signos Enfermedad Cardiaca PA
```

```
filtro = (Heart_2_Enfermos.( "Presión arterial en reposo [mmHg]" ) >= 140);
```

```
Heart_2_Enfermos_PA = Heart_2_Enfermos(filtro, :);
```

```
PersonasEnfermasPA=nnz(Heart_2_Enfermos_PA.( "Presión arterial en reposo [mmHg]" ))
```

```
PersonasEnfermasPA = 137
```

```
FrecuenciaPA=(PersonasEnfermasPA/PersonasEnfermasTotal*100);
```

```
disp("Frecuencia: "+FrecuenciaPA+"%")
```

```
Frecuencia: 26.0456%
```

```
%Principales Signos Enfermedad Cardiaca Colesterol Alto:
```

```
filtro = (Heart_2_Enfermos.( "Colesterol sérico [mg/dl]" ) >= 200);
```

```
Heart_2_Enfermos_Colesterol = Heart_2_Enfermos(filtro, :);
```

```
PersonasEnfermasColesterol=nnz(Heart_2_Enfermos_Colesterol.( "Colesterol sérico [mg/dl]" ))
```

```
PersonasEnfermasColesterol = 427
```

```
FrecuenciaColesterol=(PersonasEnfermasColesterol/PersonasEnfermasTotal*100);
```

```
disp("Frecuencia: "+FrecuenciaColesterol+"%")
```

```
Frecuencia: 81.1787%
```

```
%Principales Signos Enfermedad Cardiaca Diabetes :D 1= Presentaron niveles
```

```
%mayor a 120
```

```
filtro = (Heart_2_Enfermos.( "Azúcar en sangre (ayunas) [mg/dl]" ) == 1);
```

```
Heart_2_Enfermos_Diabetes = Heart_2_Enfermos(filtro, :);
```

```
PersonasEnfermasDiabetes=nnz(Heart_2_Enfermos_Diabetes.( "Azúcar en sangre (ayunas) [mg/dl]" ))
```

```
PersonasEnfermasDiabetes = 71
```

```
FrecuenciaDiabetes=(PersonasEnfermasDiabetes/PersonasEnfermasTotal*100);
```

```
disp("Frecuencia: "+FrecuenciaDiabetes+"%")
```

```
Frecuencia: 13.4981%
```

```
%Principales Signos Enfermedad Cardiaca Frecuencia Cardiaca
```

```
filtro = (Heart_2_Enfermos.( "Máx Fr cardiaca[latidos/min]" ) >= 100);
```

```
Heart_2_Enfermos_FrMax = Heart_2_Enfermos(filtro, :);
```

```
PersonasEnfermasConFrMax=nnz(Heart_2_Enfermos_FrMax.( "Máx Fr cardiaca[latidos/min]" ))
```

```
PersonasEnfermasConFrMax = 523
```

```
FrecuenciaFrMax=(PersonasEnfermasConFrMax/PersonasEnfermasTotal*100);
```

```
disp("Frecuencia: "+FrecuenciaFrMax+"%")
```

```
Frecuencia: 99.4297%
```

```
%Quienes tienen todos los síntomas anteriores más significativos
```

```
filtro = (Heart_2_Enfermos.( "Colesterol sérico [mg/dl]" ) >= 200) & (Heart_2_Enfermos.( "Presión arterial en reposo [mmHg]" ) >= 140) & (Heart_2_Enfermos.( "Azúcar en sangre (ayunas) [mg/dl]" ) == 1) & (Heart_2_Enfermos.( "Máx Fr cardiaca[latidos/min]" ) >= 100)
```

```
Heart_2_Enfermos_Col_PA_FrMax = Heart_2_Enfermos(filtro, :);
```

```
PersonasEnfermasVarios=nnz(Heart_2_Enfermos_Col_PA_FrMax.( "Edad" ))
```

```
PersonasEnfermasVarios = 113
```

```
FrecuenciaVariosSintomas=(PersonasEnfermasVarios/PersonasEnfermasTotal*100);
```

```
disp("Frecuencia: "+FrecuenciaVariosSintomas+"%")
```

```
Frecuencia: 21.4829%
```

```
%Agrupar por Edad y Sexo
```

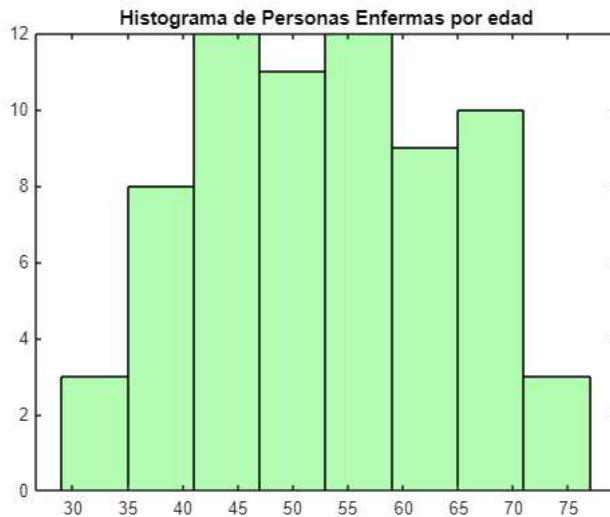
```
Heart_3_Enfermos = groupsummary(Heart_2_Enfermos, [ "Edad", "Sexo"], "mean")
```

Heart_3_Enfermos = 68x10 table

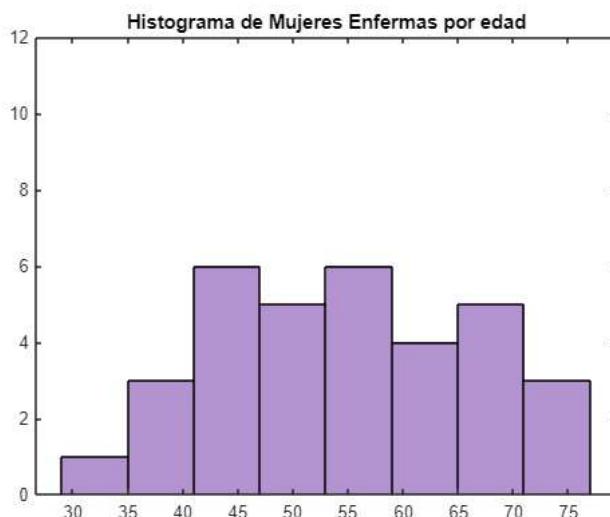
	Edad	Sexo	GroupCount	mean_Presión arterial en reposo [mmHg]	mean_Colesterol sérico [mg/dl]	
1	29	1	4	130	204	
2	34	0	3	118	210	
3	34	1	3	118	182	
4	35	0	4	138	183	
5	35	1	4	122	192	
6	37	0	3	120	215	
7	37	1	3	130	250	
8	38	1	8	138	175	
9	39	0	7	119.1429	211	
10	39	1	3	140	321	
11	40	1	3	140	199	
12	41	0	12	118.2500	244	
13	41	1	17	122.1176	208.0588	
14	42	0	6	111	237	
15	42	1	16	131.5000	233.4375	
16	43	0	3	122	213	
17	43	1	12	126.2500	269	
18	44	0	6	113	191.5000	
19	44	1	20	126	231.7000	
20	45	0	10	127	212.4000	
21	45	1	9	115.6667	258.6667	
22	46	0	9	128.3333	208	
23	46	1	3	101	197	
24	47	1	10	127	239.5000	
25	48	0	3	130	275	
26	48	1	9	125.3333	240.6667	
27	49	0	6	132	270	
28	49	1	3	130	266	
29	50	0	9	116.6667	239	
30	50	1	3	129	196	
31	51	0	9	130	286.3333	
32	51	1	20	113.8000	223.9000	
33	52	0	3	136	196	
34	52	1	25	133.1200	231.8400	
35	53	0	9	132	238	
36	53	1	9	134	223	
37	54	0	15	129	254.8000	
38	54	1	17	126.8824	264.7059	
39	55	0	7	133.2857	302.5714	
40	55	1	3	130	262	
41	56	0	3	140	294	
42	56	1	14	122.8571	223.6429	
43	57	0	6	124	328.5000	
44	57	1	15	136.4000	178.8000	
45	58	0	13	125.3846	261.6154	
46	58	1	11	125	222.1818	
47	59	1	15	148.2000	241.6000	
48	60	0	10	126.6000	244.8000	
49	61	1	3	150	243	
50	62	0	7	130.8571	288.2857	
51	62	1	6	129	219.5000	
52	63	0	6	137.5000	223.5000	
53	63	1	3	145	233	
54	64	0	9	150	313.6667	
55	64	1	10	139.4000	233	
56	65	0	9	151.6667	348.6667	
57	65	1	3	120	177	
58	66	0	7	147.7143	255.7143	
59	66	1	7	142.8571	259.7143	

	Edad	Sexo	GroupCount	mean_Presión arterial en reposo [mmHg]	mean_Colesterol sérico [mg/dl]	
60	67	0	9	124.3333	354.6667	
61	68	0	3	120	211	
62	68	1	3	118	277	
63	69	0	3	140	239	
64	69	1	3	160	234	
65	70	1	3	156	245	
66	71	0	11	124.3636	232.9091	
67	74	0	3	120	269	
68	76	0	3	140	197	

```
figure(1)
subplot(1,1,1)
Img3=histogram(Heart_3_Enfermos.Edad,min(Heart_2.Edad):6:max(Heart_2.Edad));
Img3.FaceColor = [0.5 0.99 0.5];
title('Histograma de Personas Enfermas por edad');
```



```
close all
%Filtro para MUJERES con Enfermedad Cardiaca
filtro = (Heart_3_Enfermos.(“Sexo”) == 0);
Heart_3_Enfermos_Mujeres = Heart_3_Enfermos(filtro, :);
%Histogramas
figure(5)
subplot(1,1,1)
Img5=histogram(Heart_3_Enfermos_Mujeres.Edad,min(Heart_2.Edad):6:max(Heart_2.Edad));
Img5.FaceColor = [0.5 0.3 0.7];
title('Histograma de Mujeres Enfermas por edad');
ylim([0, 12]); % Límites del eje y
```



```

close all
%Filtro para HOMBRES con Enfermedad Cardiaca
filtro = (Heart_3_Enfermos.(“Sexo”) == 1);
Heart_3_Enfermos_Hombres = Heart_3_Enfermos(filtro, :);
figure(6)
subplot(1,1,1)
Img6=histogram(Heart_3_Enfermos_Hombres.Edad,min(Heart_2.Edad):6:max(Heart_2.Edad));
Img6.FaceColor = [0 0 0.5];
title(‘Histograma de Hombres Enfermos por edad’);
ylim([0, 12]); % Límites del eje y

```

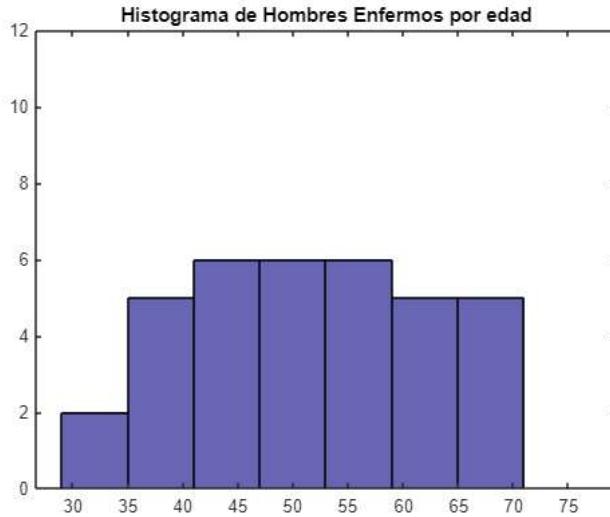


Tabla personas "Sanas"

```

%Filtro para personas 'Sanas'
filtro = (Heart_2.(“Enfermedad cardíaca”) == 0);
Heart_2_Sanas = Heart_2(filtro, :)

```

Heart_2_Sanas = 499x9 table

	Edad	Sexo	Presión arterial en reposo [mmHg]	Colesterol sérico [mg/dl]	Azúcar en sangre (ayunas) [mg/dl]
1	52	1	125	212	
2	53	1	140	203	
3	70	1	145	174	
4	61	1	148	203	
5	62	0	138	294	
6	58	1	114	318	
7	55	1	160	289	
8	46	1	120	249	
9	54	1	122	286	
10	43	0	132	341	
11	51	1	140	298	
12	52	1	128	204	
13	54	1	124	266	
14	60	1	140	185	
15	61	0	145	307	
16	58	0	136	319	
17	56	1	130	256	
18	55	0	180	327	
19	44	1	120	169	
20	57	1	130	131	
21	70	1	160	269	
22	46	1	150	231	
23	57	1	128	229	
24	61	0	130	330	
25	46	1	120	249	
26	66	0	178	228	

	Edad	Sexo	Presión arterial en reposo [mmHg]	Colesterol sérico [mg/dl]	Azúcar en sangre (ayunas) [mg/dl]
27	60	1	117	230	
28	57	0	140	241	
29	49	1	120	188	
30	55	1	140	217	
31	55	1	140	217	
32	67	1	152	212	
33	57	1	154	232	
34	67	1	100	299	
35	59	1	170	288	
36	42	1	136	315	
37	62	0	160	164	
38	59	1	170	326	
39	61	1	140	207	
40	56	1	125	249	
41	59	1	140	177	
42	48	1	130	256	
43	63	1	140	187	
44	50	1	140	233	
45	49	1	118	149	
46	46	1	150	231	
47	59	0	174	249	
48	62	0	140	268	
49	68	1	144	193	
50	63	1	140	187	
51	44	1	120	169	
52	53	1	123	282	
53	65	1	110	248	
54	47	1	110	275	
55	51	1	140	299	
56	62	1	120	281	
57	40	1	152	223	
58	54	1	110	206	
59	44	1	110	197	
60	48	1	130	256	
61	57	1	110	335	
62	59	1	126	218	
63	61	0	145	307	
64	63	1	130	254	
65	43	1	120	177	
66	44	1	120	169	
67	60	1	145	282	
68	61	1	120	260	
69	58	0	170	225	
70	57	1	152	274	
71	63	0	124	197	
72	61	1	134	234	
73	40	1	110	167	
74	41	1	110	172	
75	58	1	114	318	
76	54	1	192	283	
77	58	1	125	300	
78	63	1	130	330	
79	77	1	125	304	
80	77	1	125	304	
81	48	1	124	274	
82	56	1	125	249	
83	59	1	170	326	
84	56	1	132	184	
85	56	1	130	283	
86	54	1	110	239	

	Edad	Sexo	Presión arterial en reposo [mmHg]	Colesterol sérico [mg/dl]	Azúcar en sangre (ayunas) [mg/d]
87	56	0	200	288	
88	56	1	130	283	
89	64	1	120	246	
90	44	1	110	197	
91	56	0	134	409	
92	63	1	140	187	
93	60	1	140	293	
94	57	1	165	289	
95	40	1	110	167	
96	56	1	125	249	
97	63	1	130	254	
98	64	1	125	309	
99	67	1	160	286	
100	58	1	100	234	

%Agrupar por Edad y Sexo Per

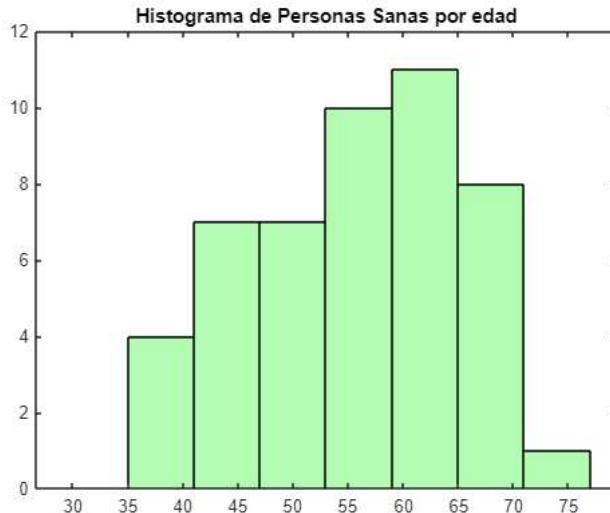
```
Heart_3_Sanas = groupsummary(Heart_2_Sanas, ["Edad", "Sexo"], "mean")
```

```
Heart_3_Sanas = 48x10 table
```

	Edad	Sexo	GroupCount	mean_Presión arterial en reposo [mmHg]	mean_Colesterol sérico [mg/dl]	
1	35	1	7	122.5714	234	
2	38	1	4	120	231	
3	39	1	4	118	219	
4	40	1	8	131	195	
5	41	1	3	110	172	
6	42	1	4	136	315	
7	43	0	4	132	341	
8	43	1	7	125.1429	207	
9	44	1	10	114.6000	213.7000	
10	45	1	6	126	286.5000	
11	46	1	11	135.4545	266.6364	
12	47	1	8	109	259	
13	48	1	11	121.0909	251.0909	
14	49	1	8	119	168.5000	
15	50	1	9	144.6667	225.3333	
16	51	0	3	130	305	
17	51	1	7	140	298.5714	
18	52	1	15	122.9333	223.2667	
19	53	1	8	131.5000	242.5000	
20	54	1	21	127.4286	244.5238	
21	55	0	8	154	266	
22	55	1	12	144	286.3333	
23	56	0	7	171.7143	339.8571	
24	56	1	15	129.0667	246.9333	
25	57	0	6	135	238.5000	
26	57	1	30	139	256.7667	
27	58	0	8	153	272	
28	58	1	36	125.0556	256.5833	
29	59	0	3	174	249	
30	59	1	28	146.1429	235.3571	
31	60	0	3	150	258	
32	60	1	24	132	243.7500	
33	61	0	8	137.5000	318.5000	
34	61	1	20	136	214	
35	62	0	17	142.5882	248.8235	
36	62	1	7	120	275	
37	63	0	11	127.6364	299.5455	
38	63	1	12	133.3333	257	
39	64	1	15	131.6667	279.7333	
40	65	0	4	150	225	

	Edad	Sexo	GroupCount	mean_Presión arterial en reposo [mmHg]	mean_Colesterol sérico [mg/dl]	
41	65	1	11	129.2727	262.5455	
42	66	0	4	178	228	
43	66	1	7	132.5714	226.5714	
44	67	1	22	131.0455	250.6818	
45	68	1	6	162	233.5000	
46	69	1	3	140	254	
47	70	1	11	143.6364	253.7273	
48	77	1	3	125	304	

```
figure(2)
subplot(1,1,1)
Img4=histogram(Heart_3_Sanas.Edad,min(Heart_2.Edad):6:max(Heart_2.Edad));
Img4.FaceColor = [0.5 0.99 0.5];
title('Histograma de Personas Sanas por edad');
```

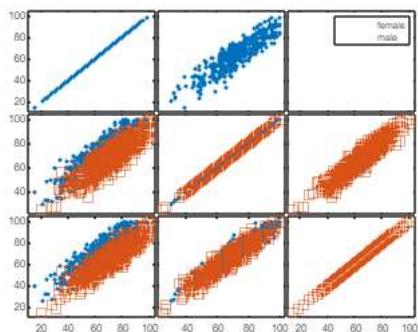


Visualizaciones exploratorias de datos

```
fig01=figure;
hold on
h1 = gplotmatrix(xVar,yVar,valGenero,colors,'.s',10)

h1 =
    3x3x2 Line array.

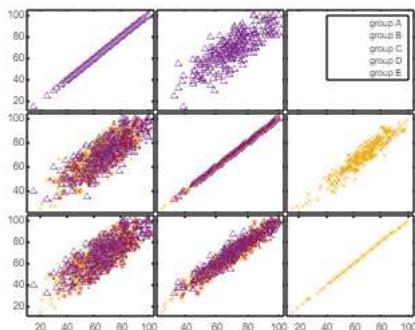
hold off
```



```
fig02 = figure;
hold on
h2 = gplotmatrix(xVar,yVar,valRaza,colors,'.*x^',5)

h2 =
    3x3x5 Line array.

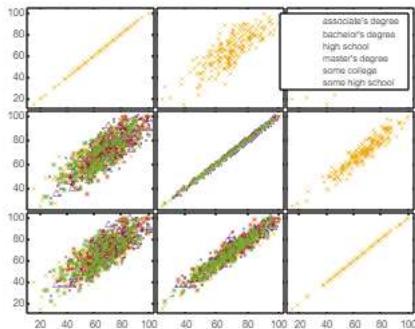
hold off
```



```
fig03 = figure;
hold on
h3 = gplotmatrix(xVar,yVar,valPadres,colors,'.*x^',5)
```

```
h3 =
3x3x6 Line array.

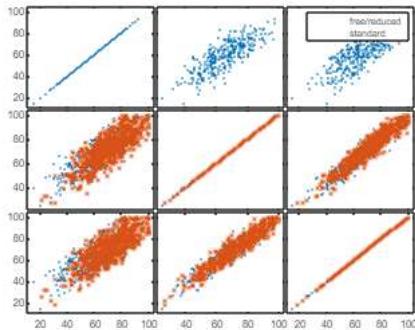
hold off
```



```
fig04 = figure;
hold on
h4 = gplotmatrix(xVar,yVar,valDesayuno,colors,'.*x^',5)
```

```
h4 =
3x3x2 Line array.

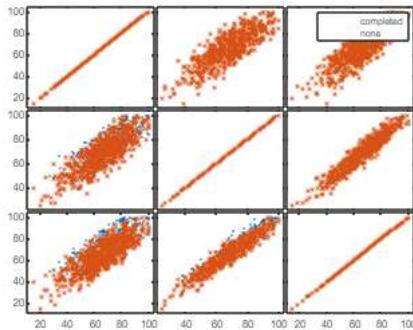
hold off
```



```
fig05 = figure;
hold on
h5 = gplotmatrix(xVar,yVar,valPrecurso,colors,'.*x^',5)
```

```
h5 =
3x3x2 Line array.

hold off
```



```
%Variables categoricas
%valGenero 2cat
%valRaza 5cat
%valPadres 5cat
%valDesayuno 2cat
%valPrecurso 2cat
%%
%Etiquetas de variables
%labelsGenero
%labelsRaza
%labelsPadres
%labelsDesayuno
%labelsPrecurso
%%
%Variables numéricas
%math
%reading
%writing

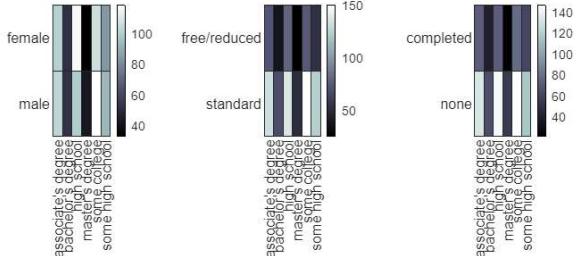
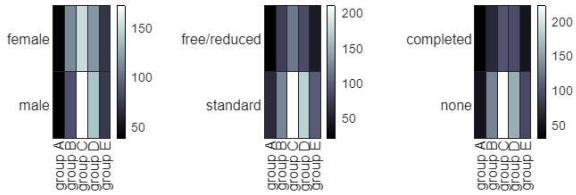
cts1 = crosstab(valGenero,valRaza);
cts2 = crosstab(valDesayuno,valRaza);
cts3 = crosstab(valPrecurso,valRaza);
cts4 = crosstab(valGenero,valPadres);
cts5 = crosstab(valDesayuno,valPadres);
cts6 = crosstab(valPrecurso,valPadres);

fig06=figure;
subplot(2,3,1)
h1 = heatmap(labelsRaza,labelsGenero,cts1,'Colormap',bone);

subplot(2,3,2)
h2 = heatmap(labelsRaza,labelsDesayuno,cts2,'Colormap',bone);
%sortx(h2,'male')

subplot(2,3,3)
h3 = heatmap(labelsRaza,labelsPrecurso,cts3,'Colormap',bone);
%sortx(h3,'male')

subplot(2,3,4)
h4 = heatmap(labelsPadres,labelsGenero,cts4,'Colormap',bone);
%sortx(h4,'male')
subplot(2,3,5)
h5 = heatmap(labelsPadres,labelsDesayuno,cts5,'Colormap',bone);
%sortx(h5,'male')
subplot(2,3,6)
h6 = heatmap(labelsPadres,labelsPrecurso,cts6,'Colormap',bone);
```



```

fig07 = figure;
subplot(2,3,1)
hold on
    x = valRaza;
    y = valGenero;
    zM = math;
    zR = reading;
    zW = writing;
    swarmchart3(x,y,zM,50,'red','o')
    swarmchart3(x,y,zR,100,'green','pentagram')
    swarmchart3(x,y,zR,50,'blue','x')
    grid on
    legend('Math','Reading','Writing')
    view([60 30])
hold off

subplot(2,3,2)
hold on
    x = valRaza;
    y = valDesayuno;
    zM = math;
    zR = reading;
    zW = writing;
    swarmchart3(x,y,zM,50,'red','o')
    swarmchart3(x,y,zR,100,'green','pentagram')
    swarmchart3(x,y,zR,50,'blue','x')
    grid on
    legend('Math','Reading','Writing')
    view([60 30])
hold off

subplot(2,3,3)
hold on
    x = valRaza;
    y = valPrecurso;
    zM = math;
    zR = reading;
    zW = writing;
    swarmchart3(x,y,zM,50,'red','o')
    swarmchart3(x,y,zR,100,'green','pentagram')
    swarmchart3(x,y,zR,50,'blue','x')
    grid on
    legend('Math','Reading','Writing')
    view([60 30])
hold off

subplot(2,3,4)
hold on
    x = valPadres;
    y = valGenero;
    zM = math;
    zR = reading;
    zW = writing;
    swarmchart3(x,y,zM,50,'red','o')
    swarmchart3(x,y,zR,100,'green','pentagram')
    swarmchart3(x,y,zR,50,'blue','x')
    grid on
    legend('Math','Reading','Writing')

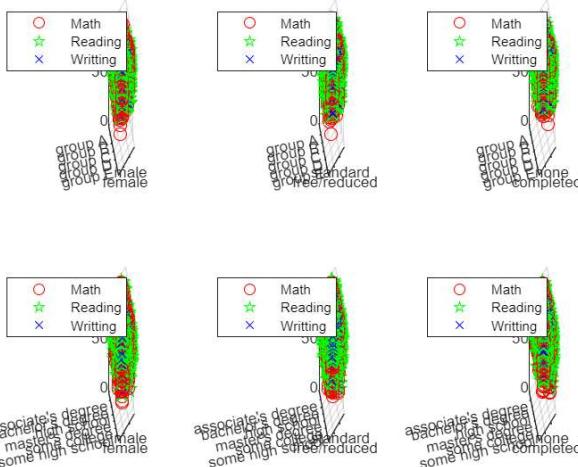
```

```

view([60 30])
hold off

subplot(2,3,5)
hold on
x = valPadres;
y = valDesayuno;
zM = math;
zR = reading;
zW = writing;
swarmchart3(x,y,zM,50,'red','o')
swarmchart3(x,y,zR,100,'green','pentagram')
swarmchart3(x,y,zR,50,'blue','x')
grid on
legend('Math','Reading','Writing')
view([60 30])
hold off
subplot(2,3,6)
hold on
x = valPadres;
y = valPrecurso;
zM = math;
zR = reading;
zW = writing;
swarmchart3(x,y,zM,50,'red','o')
swarmchart3(x,y,zR,100,'green','pentagram')
swarmchart3(x,y,zR,50,'blue','x')
grid on
legend('Math','Reading','Writing')
view([60 30])
hold off

```



```

%Tablas
%dsStudents_1
%Variables categoricas
%valGenero 2cat
%valRaza 5cat
%valPadres 5cat
%valDesayuno 2cat
%valPrecurso 2cat
%%
%Etiquetas de variables
%labelsGenero
%labelsRaza
%labelsPadres
%labelsDesayuno
%labelsPrecurso
%%
%Variables numericas
%math
%reading
%writing

mathAE    = [dsStudents_1.mathScore(valRaza == 'group B'); dsStudents_1.mathScore(valRaza == 'group C');dsStudents_1.mathScore(valRaza == 'group D');dsStudents_1.mathScore(valRaza == 'group E')];
readingAE = [dsStudents_1.readingScore(valRaza == 'group B');dsStudents_1.readingScore(valRaza == 'group C');dsStudents_1.readingScore(valRaza == 'group D');dsStudents_1.readingScore(valRaza == 'group E')];
writingAE = [dsStudents_1.writingScore(valRaza == 'group B');dsStudents_1.writingScore(valRaza == 'group C');dsStudents_1.writingScore(valRaza == 'group D');dsStudents_1.writingScore(valRaza == 'group E')];

```

```

generoAE = [dsStudents_1.gender(valRaza == 'group B'); dsStudents_1.gender(valRaza == 'group C'); dsStudents_1.gender(valRaza == 'group D');
pruebaAE = [dsStudents_1.testPreparation(valRaza == 'group B'); dsStudents_1.testPreparation(valRaza == 'group C'); dsStudents_1.testPreparation(valRaza == 'group D');
desayunoAE = [dsStudents_1.lunch(valRaza == 'group B'); dsStudents_1.lunch(valRaza == 'group C'); dsStudents_1.lunch(valRaza == 'group D');
tableDataVF = table(generoAE, pruebaAE, desayunoAE, mathAE, readingAE, writingAE)

```

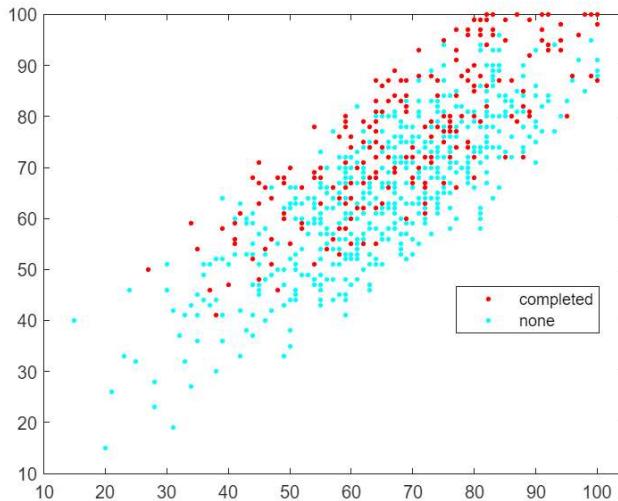
tableDataVF = 778x6 table

	generoAE	pruebaAE	desayunoAE	mathAE	readingAE	writingAE
1	male	none	free/reduced	70	70	63
2	female	none	standard	46	61	58
3	male	none	standard	74	69	69
4	male	none	standard	53	50	49
5	male	none	free/reduced	76	74	76
6	female	none	standard	70	71	71
7	male	none	standard	74	83	72
8	female	completed	standard	80	81	85
9	female	completed	free/reduced	69	78	75
10	male	none	standard	86	82	72
11	female	none	standard	52	56	54
12	female	completed	standard	69	72	72
13	female	none	free/reduced	61	74	71
14	female	none	free/reduced	35	34	36
:						

```

fig09=figure;
gscatter(tableDataVF.mathAE,tableDataVF.writingAE,tableDataVF.pruebaAE)

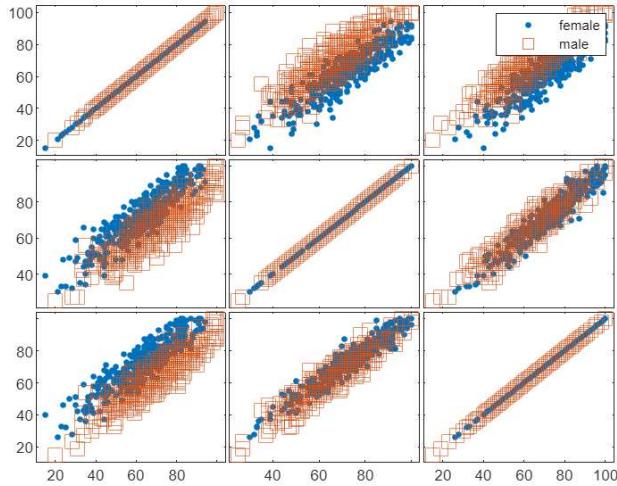
```



```

valGenero = tableDataVF.generoAE;
xVar = [mathAE, readingAE, writingAE];
yVar = [mathAE, readingAE, writingAE];
xnames = {'math'; 'reading'; 'writing'};
fig09=figure;
gplotmatrix(xVar,yVar,valGenero,colors,'.s',10)

```



```
XTotal=[tableDataVF.mathAE, tableDataVF.readingAE];
tam = size(XTotal,1);
PD = 0.3;
index = cvpartition(tam,'holdout',PD)
```

```
index =
Hold-out cross validation partition
NumObservations: 778
NumTestSets: 1
TrainSize: 545
TestSize: 233
```

```
pTrain = XTotal(index.training, :);
pTest = XTotal(index.test, :);
size(pTrain)
```

```
ans = 1x2
 545     2
```

```
size(pTest)
```

```
ans = 1x2
 233     2
```

```
X=pTrain;
```

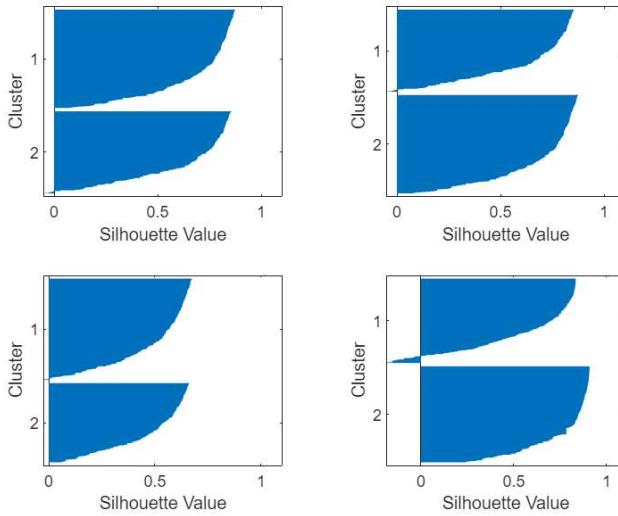
```
%X=[tableDataVF.mathAE, tableDataVF.readingAE];
%cat = [tableDataVF.generoAE];
%
num = 2;
[grp,C] = kmeans(X,2);
%eva = evalclusters(X,"kmeans","silhouette",...
%    "KList",2,"Distance","cosine")
%num = eva.OptimalK
[grp1,C1] = kmeans(X,num,"Distance","sqeuclidean");
[grp2,C2] = kmeans(X,num,"Distance","cityblock");
[grp3,C3] = kmeans(X,num,"Distance","cosine");
%[grp4,C4] = kmeans(X,2,"Distance","correlation");
%[grp5,C5] = kmeans(X,2,"Distance","hamming");
```

```
fig10a=figure;
tiledlayout(2,2);
nexttile
    axis square
    title('Automatico')
    silhouette(X,grp)
nexttile
    axis square
    title('sqeuclidean')
    silhouette(X,grp1,"sqEuclidean")
nexttile
    axis square
    title('cityblock')
    silhouette(X,grp2,"cityblock")
nexttile
```

```

axis square
title('cosine')
silhouette(X,grp3, "cosine")

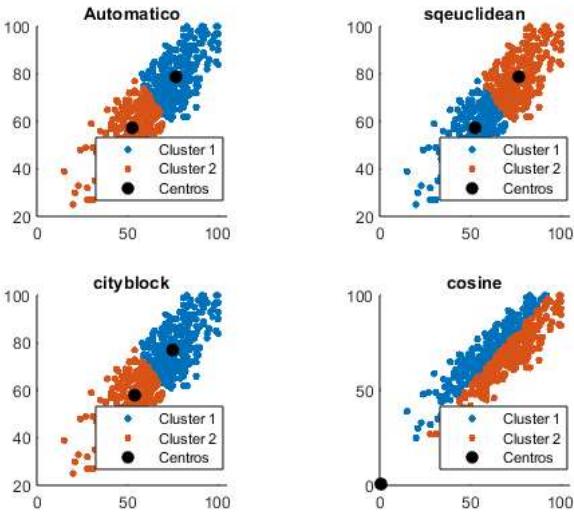
```



```

fig10b=figure;
tiledlayout(2,2);
nexttile
hold on
axis square
title('Automatico')
gscatter(X(:,1),X(:,2),grp,colors)
plot(C(:,1),C(:,2),'.k','MarkerSize',25)
legend('Cluster 1','Cluster 2','Centros')
hold off
nexttile
hold on
axis square
title('sqeuclidean')
gscatter(X(:,1),X(:,2),grp1,colors)
plot(C1(:,1),C1(:,2),'.k','MarkerSize',25)
legend('Cluster 1','Cluster 2','Centros')
hold off
nexttile
hold on
axis square
title('cityblock')
gscatter(X(:,1),X(:,2),grp2,colors)
plot(C2(:,1),C2(:,2),'.k','MarkerSize',25)
legend('Cluster 1','Cluster 2','Centros')
hold off
nexttile
hold on
axis square
title('cosine')
gscatter(X(:,1),X(:,2),grp3,colors)
plot(C3(:,1),C3(:,2),'.k','MarkerSize',25)
legend('Cluster 1','Cluster 2','Centros')
hold off

```



```

mdl = fitgmdist(X,2,...
    'RegularizationValue',0.1,...
    'CovarianceType','full',...
    'SharedCovariance',true,...
    'MaxIter',1000);
mdl1 = fitgmdist(X,2,...
    'RegularizationValue',0.1,...
    'CovarianceType','diagonal',...
    'SharedCovariance',true,...
    'MaxIter',1000);
mdl2 = fitgmdist(X,2,...
    'RegularizationValue',0.1,...
    'CovarianceType','full',...
    'SharedCovariance',false,...
    'MaxIter',1000);
mdl3 = fitgmdist(X,2,...
    'RegularizationValue',0.1,...
    'CovarianceType','diagonal',...
    'SharedCovariance',false,...
    'MaxIter',1000);

%mdl1 = fitgmdist(X,2,'RegularizationValue',0.1,"CovarianceType","diagonal","Replicates",5);
%threshold = [0.49 0.51];
%P = posterior(mdl,X);
%n = size(X,1);
%[~,order] = sort(P(:,1));
%figure
%plot(1:n,P(order,1),'r-',1:n,P(order,2),'b-')
%legend({'Cluster 1', 'Cluster 2'})
%idx = cluster(mdl,X);
%idxBoth = find(P(:,1)>=threshold(1) & P(:,1)<=threshold(2));
%numInBoth = numel(idxBoth)
%figure
%gscatter(X(:,1),X(:,2),idx,'rb','+o',5)
%hold on
%plot(X(idxBoth,1),X(idxBoth,2),'ko','MarkerSize',10)

grp = cluster(mdl,X);
grp1 = cluster(mdl1,X);
grp2 = cluster(mdl2,X);
grp3 = cluster(mdl3,X);

fig11a=figure;
tiledlayout(2,2);
nexttile
    axis square
    silhouette(X,grp)

nexttile
    axis square
    silhouette(X,grp1)

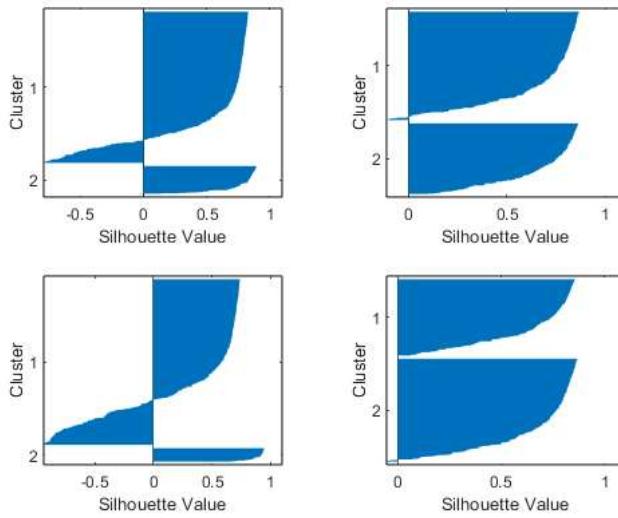
nexttile
    axis square
    silhouette(X,grp2)

```

```

nexttile
axis square
silhouette(X,grp3)

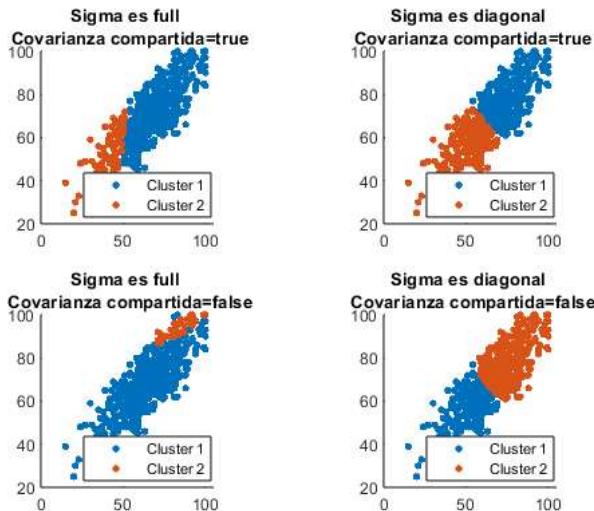
```



```

fig11b=figure;
tiledlayout(2,2);
nexttile
hold on
    axis square
    title(sprintf('Sigma es full \n Covarianza compartida=true'))
    gscatter(X(:,1),X(:,2),grp,colors)
    legend('Cluster 1','Cluster 2')
hold off
nexttile
hold on
    axis square
    title(sprintf('Sigma es diagonal \n Covarianza compartida=true'))
    gscatter(X(:,1),X(:,2),grp1,colors)
    legend('Cluster 1','Cluster 2')
hold off
nexttile
hold on
    axis square
    title(sprintf('Sigma es full \n Covarianza compartida=false'))
    gscatter(X(:,1),X(:,2),grp2,colors)
    legend('Cluster 1','Cluster 2')
hold off
nexttile
hold on
    axis square
    title(sprintf('Sigma es diagonal \n Covarianza compartida=false'))
    gscatter(X(:,1),X(:,2),grp3,colors)
    legend('Cluster 1','Cluster 2')
hold off

```



```
M='cosine';
Z1 = linkage(X);
Z2 = linkage(X, 'average',M);
Z3 = linkage(X, 'centroid',M);
```

Warning: centroid linkage specified with non-Euclidean distance metric.
Warning: Non-monotonic cluster tree -- the centroid linkage is probably not appropriate.

```
Z4 = linkage(X, 'complete',M);
Z5 = linkage(X, 'median',M);
```

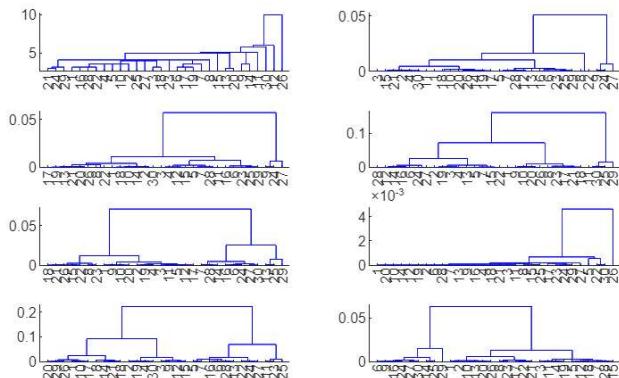
Warning: median linkage specified with non-Euclidean distance metric.
Warning: Non-monotonic cluster tree -- the median linkage is probably not appropriate.

```
Z6 = linkage(X,'single',M);
Z7 = linkage(X, 'ward',M);
```

Warning: ward's linkage specified with non-Euclidean distance metric.

```
Z8 = linkage(X, 'weighted',M);
```

```
fig12=figure;
subplot(5,2,1)
    dendrogram(Z1)
subplot(5,2,2)
    dendrogram(Z2)
subplot(5,2,3)
    dendrogram(Z3)
subplot(5,2,4)
    dendrogram(Z4)
subplot(5,2,5)
    dendrogram(Z5)
subplot(5,2,6)
    dendrogram(Z6)
subplot(5,2,7)
    dendrogram(Z7)
subplot(5,2,8)
    dendrogram(Z8)
```



```
grp1 = cluster(Z1,'maxclust',2);
D1 = pdist(X);
c1 = cophenet(Z1,D1)
```

c1 = 0.5752

```
grp2 = cluster(Z2,'maxclust',2);
D2 = pdist(X);
c2 = cophenet(Z2,D2)
```

c2 = 0.2305

```
grp3 = cluster(Z3,'maxclust',2);
D3 = pdist(X);
c3 = cophenet(Z3,D3)
```

c3 = 0.2138

```
grp4 = cluster(Z4,'maxclust',2);
D4 = pdist(X);
c4 = cophenet(Z4,D4)
```

c4 = 0.1583

```
grp5 = cluster(Z5,'maxclust',2);
D5 = pdist(X);
c5 = cophenet(Z5,D5)
```

c5 = 0.2129

```
grp6 = cluster(Z6,'maxclust',2);
D6 = pdist(X);
c6 = cophenet(Z6,D6)
```

c6 = 0.1645

```
grp7 = cluster(Z7,'maxclust',2);
D7 = pdist(X);
c7 = cophenet(Z7,D7)
```

c7 = 0.1830

```
grp8 = cluster(Z8,'maxclust',2);
D8 = pdist(X);
c8 = cophenet(Z8,D8)
```

c8 = 0.2170

```
fig13=figure;
subplot(5,2,1)
```

```
gscatter(X(:,1),X(:,2),grp1,colors)

subplot(5,2,2)
gscatter(X(:,1),X(:,2),grp2,colors)

subplot(5,2,3)
gscatter(X(:,1),X(:,2),grp3,colors)

subplot(5,2,4)
gscatter(X(:,1),X(:,2),grp4,colors)

subplot(5,2,5)
gscatter(X(:,1),X(:,2),grp5,colors)

subplot(5,2,6)
gscatter(X(:,1),X(:,2),grp6,colors)

subplot(5,2,7)
gscatter(X(:,1),X(:,2),grp7,colors)

subplot(5,2,8)
gscatter(X(:,1),X(:,2),grp8,colors)
```

