

ProyectoFinal

2024-06-20

Conjunto de datos

Conjunto seleccionado

El conjunto de datos seleccionado es sobre cáncer de mama, este incluye información completa sobre pacientes diagnosticadas con cáncer de mama, que abarca datos demográficos, atributos del tumor (tamaño, tipo, ubicación) y resultados clínicos.

Por qué se seleccionó este conjunto de datos

La razón por la que se seleccionó este conjunto de datos es su importancia crucial en el campo de la medicina y la salud pública. Permite el desarrollo y la mejora de modelos predictivos que pueden ayudar a diagnosticar el cáncer de mama de manera más precisa y temprana, lo que puede llevar a tratamientos más oportunos y mejorar las tasas de supervivencia. Además, este conjunto de datos es particularmente interesante debido a su alto número de atributos, que permiten un análisis detallado.

Preguntas de investigación

- ¿Cuáles son las características más influyentes para determinar si un tumor es maligno o benigno?
- ¿Es posible desarrollar un modelo predictivo para el diagnóstico del cáncer de mama utilizando este conjunto de datos?
- ¿Existen patrones o relaciones significativas entre las diferentes características del tumor y su diagnóstico?

```
library(readr)
dataproyectofinal <- read_csv("C:/Users/pablo/Desktop/Analisis/analisisExploratorio/dataproyectofinal.csv",
  col_types = cols(id = col_integer(),
    diagnosis = col_factor(levels = c("M",
      "B"))))

## New names:
## * ' ' -> '...33'

## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

Limpieza y preparación de datos

Primero realizaremos una revisión de los datos para tener más información al respecto

```
#Primeros valores
```

```
head(dataprojectofinal)
```

```
## # A tibble: 6 x 33
##       id diagnosis radius_mean texture_mean perimeter_mean area_mean
##   <int> <fct>         <dbl>         <dbl>         <dbl>         <dbl>
## 1  842302 M             18.0           10.4           123.          1001
## 2  842517 M             20.6           17.8           133.          1326
## 3 84300903 M             19.7           21.2           130           1203
## 4 84348301 M             11.4           20.4            77.6          386.
## 5 84358402 M             20.3           14.3           135.          1297
## 6  843786 M             12.4           15.7            82.6          477.
## # i 27 more variables: smoothness_mean <dbl>, compactness_mean <dbl>,
## #   concavity_mean <dbl>, 'concave points_mean' <dbl>, symmetry_mean <dbl>,
## #   fractal_dimension_mean <dbl>, radius_se <dbl>, texture_se <dbl>,
## #   perimeter_se <dbl>, area_se <dbl>, smoothness_se <dbl>,
## #   compactness_se <dbl>, concavity_se <dbl>, 'concave points_se' <dbl>,
## #   symmetry_se <dbl>, fractal_dimension_se <dbl>, radius_worst <dbl>,
## #   texture_worst <dbl>, perimeter_worst <dbl>, area_worst <dbl>, ...
```

```
#resumen de los datos
```

```
summary(dataprojectofinal)
```

```
##           id           diagnosis radius_mean texture_mean
## Min.      : 8670      M:212      Min.   : 6.981  Min.   : 9.71
## 1st Qu.: 869222      B:356      1st Qu.:11.707 1st Qu.:16.17
## Median : 906157                      Median :13.375 Median :18.84
## Mean   : 30425140                      Mean  :14.139 Mean  :19.28
## 3rd Qu.: 8825022                      3rd Qu.:15.797 3rd Qu.:21.79
## Max.    :911320502                      Max.   :28.110 Max.   :39.28
## perimeter_mean area_mean smoothness_mean compactness_mean
## Min.   : 43.79  Min.   : 143.5  Min.   :0.06251  Min.   :0.01938
## 1st Qu.: 75.20  1st Qu.: 420.3  1st Qu.:0.08640  1st Qu.:0.06517
## Median : 86.29  Median : 551.4  Median :0.09589  Median :0.09312
## Mean   : 92.05  Mean   : 655.7  Mean   :0.09644  Mean   :0.10445
## 3rd Qu.:104.15  3rd Qu.: 784.1  3rd Qu.:0.10533  3rd Qu.:0.13043
## Max.    :188.50 Max.    :2501.0  Max.    :0.16340  Max.    :0.34540
## concavity_mean concave points_mean symmetry_mean fractal_dimension_mean
## Min.   :0.00000  Min.   :0.00000  Min.   :0.1060  Min.   :0.04996
## 1st Qu.:0.02958  1st Qu.:0.02035  1st Qu.:0.1620  1st Qu.:0.05770
## Median :0.06155  Median :0.03360  Median :0.1792  Median :0.06155
## Mean   :0.08896  Mean   :0.04901  Mean   :0.1812  Mean   :0.06280
## 3rd Qu.:0.13100  3rd Qu.:0.07401  3rd Qu.:0.1957  3rd Qu.:0.06613
## Max.    :0.42680  Max.    :0.20120  Max.    :0.3040  Max.    :0.09744
## radius_se texture_se perimeter_se area_se
## Min.   :0.1115  Min.   :0.3602  Min.   : 0.757  Min.   : 6.802
## 1st Qu.:0.2324  1st Qu.:0.8331  1st Qu.: 1.605  1st Qu.:17.850
## Median :0.3240  Median :1.1080  Median : 2.285  Median :24.565
## Mean   :0.4052  Mean   :1.2165  Mean   : 2.867  Mean   :40.374
```

```
## 3rd Qu.:0.4798 3rd Qu.:1.4743 3rd Qu.: 3.360 3rd Qu.: 45.237
## Max. :2.8730 Max. :4.8850 Max. :21.980 Max. :542.200
## smoothness_se compactness_se concavity_se concave points_se
## Min. :0.001713 Min. :0.002252 Min. :0.00000 Min. :0.000000
## 1st Qu.:0.005166 1st Qu.:0.013133 1st Qu.:0.01510 1st Qu.:0.007663
## Median :0.006374 Median :0.020460 Median :0.02592 Median :0.010950
## Mean :0.007041 Mean :0.025515 Mean :0.03195 Mean :0.011817
## 3rd Qu.:0.008151 3rd Qu.:0.032455 3rd Qu.:0.04212 3rd Qu.:0.014730
## Max. :0.031130 Max. :0.135400 Max. :0.39600 Max. :0.052790
## symmetry_se fractal_dimension_se radius_worst texture_worst
## Min. :0.007882 Min. :0.0008948 Min. : 7.93 Min. :12.02
## 1st Qu.:0.015128 1st Qu.:0.0022445 1st Qu.:13.03 1st Qu.:21.07
## Median :0.018725 Median :0.0031955 Median :14.97 Median :25.41
## Mean :0.020531 Mean :0.0037967 Mean :16.28 Mean :25.67
## 3rd Qu.:0.023398 3rd Qu.:0.0045585 3rd Qu.:18.80 3rd Qu.:29.68
## Max. :0.078950 Max. :0.0298400 Max. :36.04 Max. :49.54
## perimeter_worst area_worst smoothness_worst compactness_worst
## Min. : 50.41 Min. : 185.2 Min. :0.07117 Min. :0.02729
## 1st Qu.: 84.15 1st Qu.: 515.7 1st Qu.:0.11660 1st Qu.:0.14758
## Median : 97.67 Median : 686.5 Median :0.13135 Median :0.21300
## Mean :107.35 Mean : 881.7 Mean :0.13244 Mean :0.25460
## 3rd Qu.:125.53 3rd Qu.:1085.0 3rd Qu.:0.14602 3rd Qu.:0.33930
## Max. :251.20 Max. :4254.0 Max. :0.22260 Max. :1.05800
## concavity_worst concave points_worst symmetry_worst fractal_dimension_worst
## Min. :0.0000 Min. :0.00000 Min. :0.1565 Min. :0.05504
## 1st Qu.:0.1159 1st Qu.:0.06497 1st Qu.:0.2504 1st Qu.:0.07147
## Median :0.2275 Median :0.10002 Median :0.2821 Median :0.08005
## Mean :0.2727 Mean :0.11481 Mean :0.2901 Mean :0.08397
## 3rd Qu.:0.3835 3rd Qu.:0.16168 3rd Qu.:0.3180 3rd Qu.:0.09208
## Max. :1.2520 Max. :0.29100 Max. :0.6638 Max. :0.20750
## ...33
## Mode:logical
## NA's:568
##
##
##
##
```

Datos faltantes

De lo anterior podemos observar que solamente existen valores faltantes en toda la columna de nombre ...33, por lo tanto, esta será eliminada del conjunto de datos.

```
#Eliminar la columna ...33
dataproyectofinal <- dataproyectofinal %>% select(-...33)
head(dataproyectofinal)
```

```
## # A tibble: 6 x 32
##       id diagnosis radius_mean texture_mean perimeter_mean area_mean
##   <int> <fct>          <dbl>         <dbl>         <dbl>         <dbl>
## 1  842302 M             18.0           10.4          123.          1001
## 2  842517 M             20.6           17.8          133.          1326
## 3 84300903 M             19.7           21.2          130           1203
```

```
## 4 84348301 M          11.4          20.4          77.6          386.
## 5 84358402 M          20.3          14.3          135.          1297
## 6 843786 M           12.4          15.7           82.6          477.
## # i 26 more variables: smoothness_mean <dbl>, compactness_mean <dbl>,
## #   concavity_mean <dbl>, 'concave points_mean' <dbl>, symmetry_mean <dbl>,
## #   fractal_dimension_mean <dbl>, radius_se <dbl>, texture_se <dbl>,
## #   perimeter_se <dbl>, area_se <dbl>, smoothness_se <dbl>,
## #   compactness_se <dbl>, concavity_se <dbl>, 'concave points_se' <dbl>,
## #   symmetry_se <dbl>, fractal_dimension_se <dbl>, radius_worst <dbl>,
## #   texture_worst <dbl>, perimeter_worst <dbl>, area_worst <dbl>, ...
```

Limpieza de duplicados

Para ver si todos los datos son únicos, se realizará una revisión por IDs.

```
# Revisar valores duplicados en las IDs
```

```
dataproyectofinal %>%
  group_by(id) %>%
  summarise(count = n()) %>%
  filter(count > 1)
```

```
## # A tibble: 0 x 2
## # i 2 variables: id <int>, count <int>
```

Luego de aplicado el proceso, podemos observar que no existen IDs repetidas, por lo tanto, borraremos la columna “id”, ya que no servirá para el análisis.

```
#borrar la columna Id
```

```
dataproyectofinal <- dataproyectofinal %>% select(-id)
head(dataproyectofinal)
```

```
## # A tibble: 6 x 31
##   diagnosis radius_mean texture_mean perimeter_mean area_mean smoothness_mean
##   <fct>         <dbl>         <dbl>         <dbl>         <dbl>         <dbl>
## 1 M           18.0          10.4          123.          1001          0.118
## 2 M           20.6          17.8          133.          1326          0.0847
## 3 M           19.7          21.2          130           1203          0.110
## 4 M           11.4          20.4          77.6           386.          0.142
## 5 M           20.3          14.3          135.          1297          0.100
## 6 M           12.4          15.7           82.6          477.          0.128
## # i 25 more variables: compactness_mean <dbl>, concavity_mean <dbl>,
## #   'concave points_mean' <dbl>, symmetry_mean <dbl>,
## #   fractal_dimension_mean <dbl>, radius_se <dbl>, texture_se <dbl>,
## #   perimeter_se <dbl>, area_se <dbl>, smoothness_se <dbl>,
## #   compactness_se <dbl>, concavity_se <dbl>, 'concave points_se' <dbl>,
## #   symmetry_se <dbl>, fractal_dimension_se <dbl>, radius_worst <dbl>,
## #   texture_worst <dbl>, perimeter_worst <dbl>, area_worst <dbl>, ...
```

Datos atípicos

No se eliminarán los valores atípicos del conjunto de datos, dado que representan variaciones significativas que podrían ser fundamentales para el análisis. Estos valores pueden indicar comportamientos o casos extremos

que, aunque no comunes, son parte integral del fenómeno bajo estudio. Eliminar estos valores podría llevar a una interpretación errónea de los datos, omitiendo potencialmente patrones importantes.

Análisis exploratorio de datos

Estadísticas descriptivas

Primero realizaremos un análisis descriptivo de los datos para resumir las características principales del conjunto de datos.

```
psych_stats <- psych::describe(dataprojectofinal)
psych_stats
```

##	vars	n	mean	sd	median	trimmed	mad	min
## diagnosis*	1	568	1.63	0.48	2.00	1.66	0.00	1.00
## radius_mean	2	568	14.14	3.52	13.38	13.83	2.82	6.98
## texture_mean	3	568	19.28	4.30	18.84	19.03	4.13	9.71
## perimeter_mean	4	568	92.05	24.25	86.29	89.79	18.87	43.79
## area_mean	5	568	655.72	351.66	551.40	606.75	227.13	143.50
## smoothness_mean	6	568	0.10	0.01	0.10	0.10	0.01	0.06
## compactness_mean	7	568	0.10	0.05	0.09	0.10	0.05	0.02
## concavity_mean	8	568	0.09	0.08	0.06	0.08	0.06	0.00
## concave points_mean	9	568	0.05	0.04	0.03	0.04	0.03	0.00
## symmetry_mean	10	568	0.18	0.03	0.18	0.18	0.03	0.11
## fractal_dimension_mean	11	568	0.06	0.01	0.06	0.06	0.01	0.05
## radius_se	12	568	0.41	0.28	0.32	0.36	0.16	0.11
## texture_se	13	568	1.22	0.55	1.11	1.16	0.47	0.36
## perimeter_se	14	568	2.87	2.02	2.29	2.51	1.14	0.76
## area_se	15	568	40.37	45.52	24.57	31.72	13.65	6.80
## smoothness_se	16	568	0.01	0.00	0.01	0.01	0.00	0.00
## compactness_se	17	568	0.03	0.02	0.02	0.02	0.01	0.00
## concavity_se	18	568	0.03	0.03	0.03	0.03	0.02	0.00
## concave points_se	19	568	0.01	0.01	0.01	0.01	0.01	0.00
## symmetry_se	20	568	0.02	0.01	0.02	0.02	0.01	0.01
## fractal_dimension_se	21	568	0.00	0.00	0.00	0.00	0.00	0.00
## radius_worst	22	568	16.28	4.83	14.97	15.74	3.62	7.93
## texture_worst	23	568	25.67	6.15	25.41	25.38	6.40	12.02
## perimeter_worst	24	568	107.35	33.57	97.66	103.49	24.95	50.41
## area_worst	25	568	881.66	569.28	686.55	788.91	319.35	185.20
## smoothness_worst	26	568	0.13	0.02	0.13	0.13	0.02	0.07
## compactness_worst	27	568	0.25	0.16	0.21	0.23	0.13	0.03
## concavity_worst	28	568	0.27	0.21	0.23	0.25	0.19	0.00
## concave points_worst	29	568	0.11	0.07	0.10	0.11	0.07	0.00
## symmetry_worst	30	568	0.29	0.06	0.28	0.28	0.05	0.16
## fractal_dimension_worst	31	568	0.08	0.02	0.08	0.08	0.01	0.06
##		max	range	skew	kurtosis	se		
## diagnosis*		2.00	1.00	-0.52	-1.73	0.02		
## radius_mean		28.11	21.13	0.95	0.82	0.15		
## texture_mean		39.28	29.57	0.65	0.74	0.18		
## perimeter_mean		188.50	144.71	0.99	0.95	1.02		
## area_mean		2501.00	2357.50	1.64	3.60	14.76		
## smoothness_mean		0.16	0.10	0.50	0.77	0.00		

## compactness_mean	0.35	0.33	1.18	1.61	0.00
## concavity_mean	0.43	0.43	1.39	1.95	0.00
## concave points_mean	0.20	0.20	1.17	1.03	0.00
## symmetry_mean	0.30	0.20	0.72	1.25	0.00
## fractal_dimension_mean	0.10	0.05	1.29	2.94	0.00
## radius_se	2.87	2.76	3.07	17.41	0.01
## texture_se	4.88	4.52	1.64	5.26	0.02
## perimeter_se	21.98	21.22	3.42	21.07	0.08
## area_se	542.20	535.40	5.41	48.52	1.91
## smoothness_se	0.03	0.03	2.30	10.30	0.00
## compactness_se	0.14	0.13	1.89	5.03	0.00
## concavity_se	0.40	0.40	5.09	48.31	0.00
## concave points_se	0.05	0.05	1.45	5.09	0.00
## symmetry_se	0.08	0.07	2.19	7.79	0.00
## fractal_dimension_se	0.03	0.03	3.90	25.89	0.00
## radius_worst	36.04	28.11	1.10	0.91	0.20
## texture_worst	49.54	37.52	0.50	0.20	0.26
## perimeter_worst	251.20	200.79	1.12	1.04	1.41
## area_worst	4254.00	4068.80	1.85	4.32	23.89
## smoothness_worst	0.22	0.15	0.42	0.50	0.00
## compactness_worst	1.06	1.03	1.47	2.99	0.01
## concavity_worst	1.25	1.25	1.15	1.58	0.01
## concave points_worst	0.29	0.29	0.49	-0.55	0.00
## symmetry_worst	0.66	0.51	1.42	4.36	0.00
## fractal_dimension_worst	0.21	0.15	1.65	5.15	0.00

Visualizaciones

A continuación, se realizarán gráficos de pares, lo cual ayudará a identificar qué variables tienen fuertes correlaciones lineales o patrones discernibles entre sí. En este caso se realizará para alguno de los atributos. Además, para obtener una vista clara y comparativa de las características estos serán normalizados.

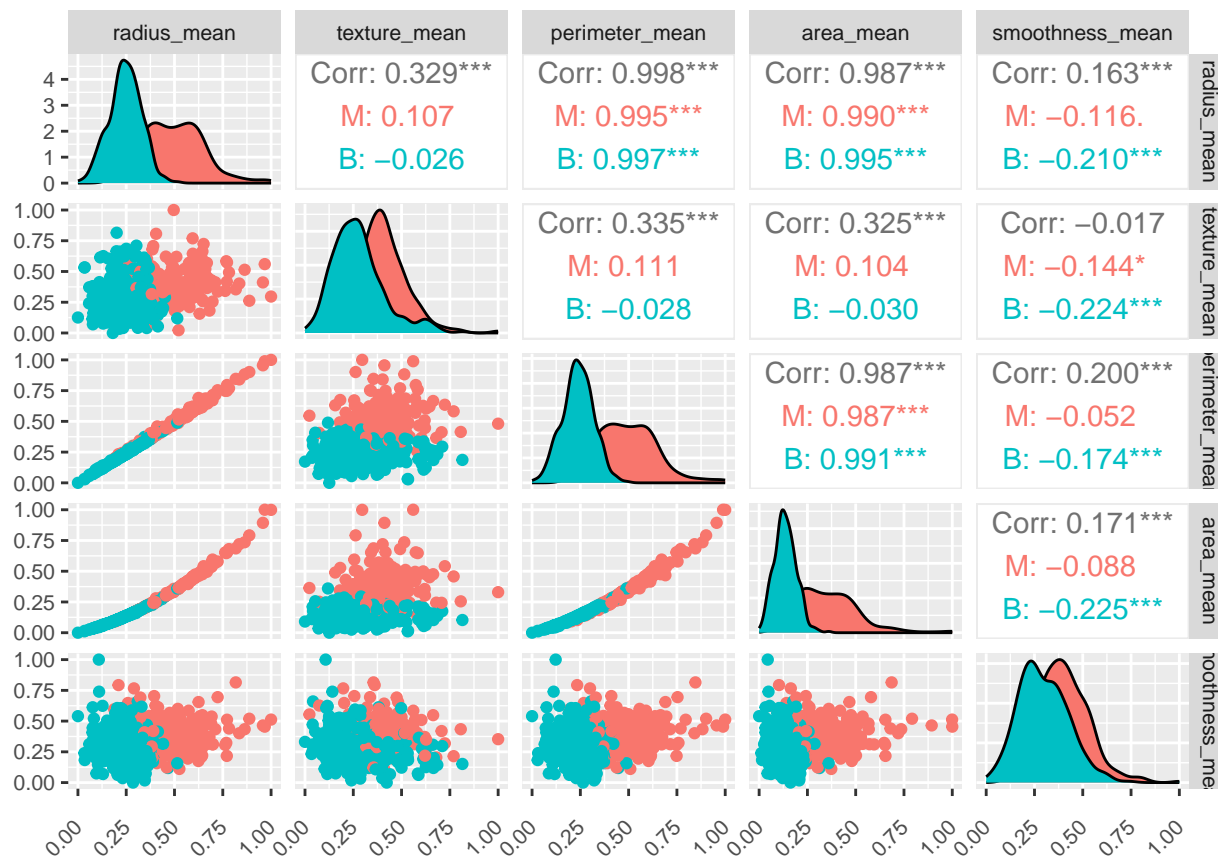
```
normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))
}
data_normalized <- as.data.frame(lapply(dataprojectofinal[, 2:ncol(dataprojectofinal)], normalize))

# Agregar de nuevo la columna diagnosis
data_normalized$diagnosis <- dataprojectofinal$diagnosis

#seleccionar alguno de los atributos
p <- ggpairs(data_normalized,
             columns = c("radius_mean", "texture_mean", "perimeter_mean", "area_mean", "smoothness_mean",
                        "compactness_mean", "concavity_mean", "concave points_mean", "symmetry_mean", "fractal_dimension_mean",
                        "radius_worst", "texture_worst", "perimeter_worst", "area_worst", "smoothness_worst",
                        "compactness_worst", "concavity_worst", "concave points_worst", "symmetry_worst", "fractal_dimension_worst"),
             ggplot2::aes(color = diagnosis))

# Ajustar los ángulos de las etiquetas de los ejes
p <- p + theme(text = element_text(size = 10), # Ajusta el tamaño del texto general
               axis.text.x = element_text(angle = 45, vjust = 0.5, hjust = 1))

print(p)
```

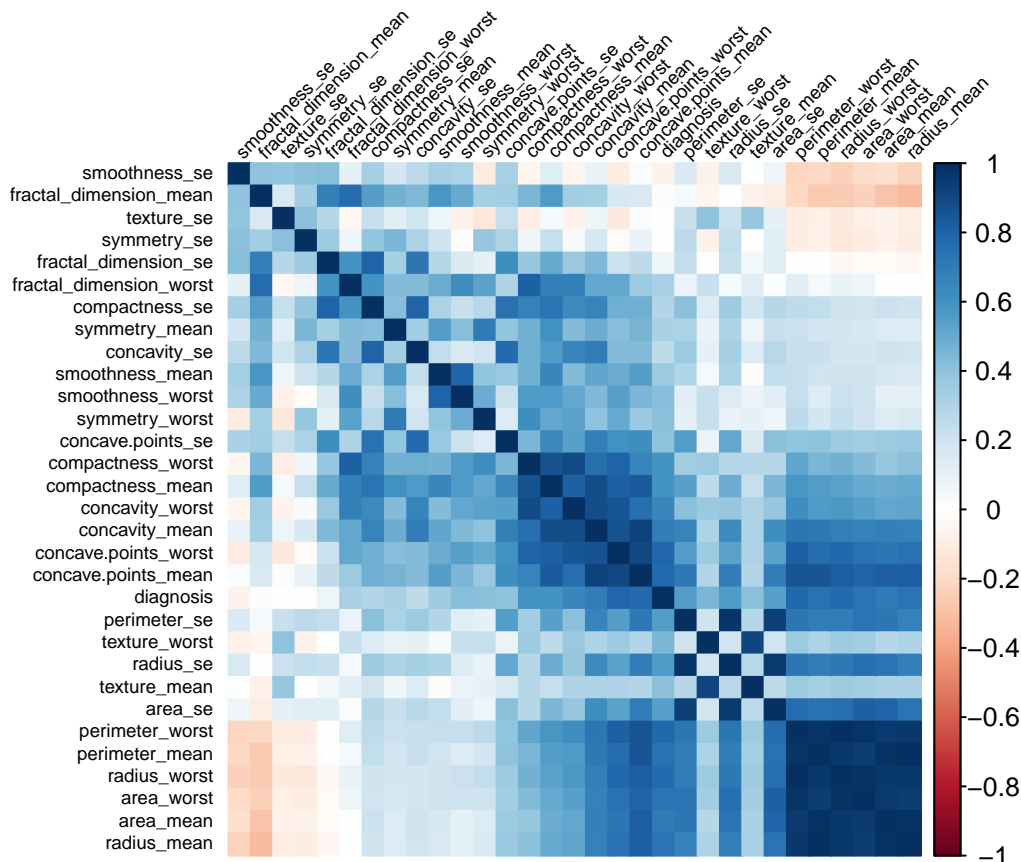


Realizaremos una matriz de correlación, para observar si existen relaciones significativas entre las diferentes características del conjunto de datos. Además, aplicaremos una transformación a diagnosis.

```
# Codificar 'diagnosis' como numérica
data_normalized$diagnosis <- as.numeric(factor(data_normalized$diagnosis, levels = c("B", "M")))

# Calcular la matriz de correlación
cor_matrix <- cor(data_normalized, use = "complete.obs")

corrplot(cor_matrix, method = "color",
          order = "AOE",
          tl.col = "black",
          tl.srt = 45,
          tl.cex = 0.6,
          )
```



Discusión de hallazgos

En la visualización de pares se puede observar que variables como `radius_mean`, `perimeter_mean` y `area_mean` muestran correlaciones muy altas entre sí. Esta fuerte correlación es lógica, dado que todas son medidas de tamaño y dimensión del tumor. De la matriz podemos observar que hay bloques de alta correlación que confirman lo observado en el gráfico de pares. Variables relacionadas con las dimensiones físicas del tumor están altamente correlacionadas entre sí.

Transformaciones de datos

Aun que anteriormente ya se realizaron transformaciones, estaban destinadas a mejorar las visualizaciones. Por lo tanto, se decidió estandarizar los datos para asegurar que todas las características contribuyan equitativamente a futuros modelos que se vean beneficiados. En caso de necesitar los datos normalizados, tendremos los utilizados anteriormente.

```
data_scaled <- as.data.frame(scale(dataproyectofinal[ , 2:ncol(dataproyectofinal)]))

# Agregar de nuevo la columna diagnosis
data_scaled$diagnosis <- dataproyectofinal$diagnosis

# Codificación de la variable categórica
data_scaled$diagnosis <- ifelse(data_scaled$diagnosis == 'M', 1, 0)
head(data_scaled)
```


##	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	
## 1	1.0951132	-2.0702630	1.2681916	0.9818460	1.5737125	
## 2	1.8286958	-0.3513251	1.6846895	1.9060324	-0.8381537	
## 3	1.5784816	0.4581342	1.5651010	1.5562634	0.9431593	
## 4	-0.7729634	0.2557694	-0.5965646	-0.7667148	3.3005686	
## 5	1.7490822	-1.1491542	1.7754118	1.8235666	0.2767791	
## 6	-0.4800990	-0.8328138	-0.3907899	-0.5079426	2.2472580	
##	compactness_mean	concavity_mean	concave	points_mean	symmetry_mean	
## 1	3.2795417	2.6491483		2.5293505	2.2134838912	
## 2	-0.4888074	-0.0257915		0.5457278	-0.0000513617	
## 3	1.0502760	1.3606102		2.0342828	0.9371456070	
## 4	3.3988652	1.9126616		1.4489684	2.8625930758	
## 5	0.5369957	1.3681382		1.4257621	-0.0109914041	
## 6	1.2415724	0.8637640		0.8221406	1.0027858616	
##	fractal_dimension_mean	radius_se	texture_se	perimeter_se	area_se	
## 1	2.2514169	2.4852417	-0.5636703	2.8278325	2.4828576	
## 2	-0.8683511	0.4982551	-0.8741419	0.2625928	0.7404184	
## 3	-0.3984042	1.2263965	-0.7781385	0.8491733	1.1786643	
## 4	4.9026535	0.3256772	-0.1095557	0.2858188	-0.2887454	
## 5	-0.5626025	1.2681899	-0.7882822	1.2707009	1.1876709	
## 6	1.8862172	-0.2547462	-0.5910222	-0.3210227	-0.2896241	
##	smoothness_se	compactness_se	concavity_se	concave	points_se	symmetry_se
## 1	-0.2135390	1.31406725	0.7216048		0.65842022	1.1486323
## 2	-0.6042006	-0.69458047	-0.4422989		0.25717181	-0.8031113
## 3	-0.2963965	0.81246390	0.2110510		1.42355382	0.2380605
## 4	0.6885767	2.74067634	0.8170230		1.11327672	4.7280380
## 5	1.4805483	-0.05053964	0.8259685		1.14251749	-0.3593133
## 6	0.1561587	0.44324362	0.1580408		-0.07259915	0.1352735
##	fractal_dimension_se	radius_worst	texture_worst	perimeter_worst	area_worst	
## 1	0.90493154	1.8841959	-1.3562568		2.3011925	1.9978631
## 2	-0.09995439	1.8034341	-0.3673997		1.5326797	1.8871966
## 3	0.29240806	1.5093785	-0.0226008		1.3450196	1.4533136
## 4	2.04350015	-0.2839470	0.1351610		-0.2524743	-0.5515066
## 5	0.49784134	1.2960846	-1.4635999		1.3360834	1.2179277
## 6	0.48537939	-0.1679813	-0.3121017		-0.1175377	-0.2460319
##	smoothness_worst	compactness_worst	concavity_worst	concave	points_worst	
## 1	1.3060984	2.6132966	2.1066853		2.2951312	
## 2	-0.3793743	-0.4323646	-0.1490096		1.0850177	
## 3	0.5248115	1.0802926	0.8524556		1.9537390	
## 4	3.3953822	3.8894224	1.9867781		2.1747294	
## 5	0.2175639	-0.3153706	0.6107226		0.7268607	
## 6	2.0478819	1.7186734	1.2606199		0.9036531	
##	symmetry_worst	fractal_dimension_worst	diagnosis			
## 1	2.7457049	1.9332477	1			
## 2	-0.2435458	0.2795140	1			
## 3	1.1501459	0.1998160	1			
## 4	6.0353340	4.9274577	1			
## 5	-0.8669119	-0.3979191	1			
## 6	1.7509029	2.2376499	1			

Interpretación y conclusiones

Interpretación de resultados

La visualización de pair plot muestra agrupaciones distintas para tumores benignos y malignos en función de múltiples características. Se puede observar que `radius_mean` y `texture_mean` muestran diferencias significativas entre tumores benignos y malignos. Los tumores malignos tienden a tener valores más altos en estas características.

Respuesta a las preguntas de investigación

- **¿Cuáles son las características más influyentes para determinar si un tumor es maligno o benigno?**

Las características más influyentes a simple vista son `radius_mean`, `perimeter_mean` y `area_mean`, ya que presentan las diferencias más significativas entre tumores benignos y malignos y están altamente correlacionadas entre sí.

- **¿Es posible desarrollar un modelo predictivo para el diagnóstico del cáncer de mama utilizando este conjunto de datos?**

Sí, es posible. La alta correlación y la clara separación observada en las visualizaciones sugieren que un modelo predictivo puede diferenciar entre tumores benignos y malignos.

- **¿Existen patrones o relaciones significativas entre las diferentes características del tumor y su diagnóstico?**

Sí, existen patrones significativos. Las características relacionadas con el tamaño del tumor (`radius_mean`, `perimeter_mean`, `area_mean`) están altamente correlacionadas y muestran una separación clara entre tumores benignos y malignos. Estos patrones se repiten en múltiples visualizaciones.