

Trabajo Evaluacion

Miguel Santiago Gómez Suárez - AX751708

Nicolás Leornado Maldonado Garzón - AX840282

2024-01-15

Trabajo Final

En este trabajo se realizará el estudio de:

- Análisis previo de los datos.
- Análisis de componentes principales.
- Análisis Factorial.
- Análisis de conglomerados.

Todo esto se hará sobre el conjunto de datos de prueba [Wine Quality](#)

Análisis previo de los datos

Estadísticos

Cargamos nuestros datos y observamos nuestros 5 primero y los 5 últimos datos de nuestro dataSet.

Nota: Es importante tener en cuenta que es necesario mover el directorio de trabajo a la ubicación del proyecto

```
datos_vino <- read.csv("../Datos/Brutos/winequality-white.csv", sep = ";")
```

```
head(datos_vino)
```

```
## fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
## 1          7.0           0.27         0.36          20.7        0.045
## 2          6.3           0.30         0.34           1.6        0.049
## 3          8.1           0.28         0.40           6.9        0.050
## 4          7.2           0.23         0.32           8.5        0.058
## 5          7.2           0.23         0.32           8.5        0.058
## 6          8.1           0.28         0.40           6.9        0.050
## free.sulfur.dioxide total.sulfur.dioxide density    pH sulphates
alcohol
## 1          45          170  1.0010 3.00         0.45
8.8
## 2          14          132  0.9940 3.30         0.49
9.5
```

```
## 3          30          97  0.9951 3.26      0.44
10.1
## 4          47         186  0.9956 3.19      0.40
9.9
## 5          47         186  0.9956 3.19      0.40
9.9
## 6          30          97  0.9951 3.26      0.44
10.1
##    quality
## 1         6
## 2         6
## 3         6
## 4         6
## 5         6
## 6         6
```

```
tail(datos_vino)
```

```
##      fixed.acidity volatile.acidity citric.acid residual.sugar
chlorides
## 4893          6.5          0.23          0.38          1.3
0.032
## 4894          6.2          0.21          0.29          1.6
0.039
## 4895          6.6          0.32          0.36          8.0
0.047
## 4896          6.5          0.24          0.19          1.2
0.041
## 4897          5.5          0.29          0.30          1.1
0.022
## 4898          6.0          0.21          0.38          0.8
0.020
##      free.sulfur.dioxide total.sulfur.dioxide density    pH sulphates
alcohol
## 4893          29          112 0.99298 3.29      0.54
9.7
## 4894          24          92 0.99114 3.27      0.50
11.2
## 4895          57         168 0.99490 3.15      0.46
9.6
## 4896          30         111 0.99254 2.99      0.46
9.4
## 4897          20         110 0.98869 3.34      0.38
12.8
## 4898          22          98 0.98941 3.26      0.32
11.8
##      quality
## 4893         5
## 4894         6
## 4895         5
```

```
## 4896      6
## 4897      7
## 4898      6
```

Ahora observamos un resumen de nuestros datos:

```
summary(datos_vino)
```

```
## fixed.acidity    volatile.acidity    citric.acid    residual.sugar
## Min.   : 3.800    Min.   :0.0800    Min.   :0.0000    Min.   : 0.600
## 1st Qu.: 6.300    1st Qu.:0.2100    1st Qu.:0.2700    1st Qu.: 1.700
## Median : 6.800    Median :0.2600    Median :0.3200    Median : 5.200
## Mean   : 6.855    Mean   :0.2782    Mean   :0.3342    Mean   : 6.391
## 3rd Qu.: 7.300    3rd Qu.:0.3200    3rd Qu.:0.3900    3rd Qu.: 9.900
## Max.   :14.200    Max.   :1.1000    Max.   :1.6600    Max.   :65.800
## chlorides        free.sulfur.dioxide    total.sulfur.dioxide    density
## Min.   :0.00900    Min.   : 2.00      Min.   : 9.0      Min.
:0.9871
## 1st Qu.:0.03600    1st Qu.: 23.00      1st Qu.:108.0      1st
Qu.:0.9917
## Median :0.04300    Median : 34.00      Median :134.0      Median
:0.9937
## Mean   :0.04577    Mean   : 35.31      Mean   :138.4      Mean
:0.9940
## 3rd Qu.:0.05000    3rd Qu.: 46.00      3rd Qu.:167.0      3rd
Qu.:0.9961
## Max.   :0.34600    Max.   :289.00      Max.   :440.0      Max.
:1.0390
##          pH          sulphates          alcohol          quality
## Min.   :2.720    Min.   :0.2200    Min.   : 8.00    Min.   :3.000
## 1st Qu.:3.090    1st Qu.:0.4100    1st Qu.: 9.50    1st Qu.:5.000
## Median :3.180    Median :0.4700    Median :10.40    Median :6.000
## Mean   :3.188    Mean   :0.4898    Mean   :10.51    Mean   :5.878
## 3rd Qu.:3.280    3rd Qu.:0.5500    3rd Qu.:11.40    3rd Qu.:6.000
## Max.   :3.820    Max.   :1.0800    Max.   :14.20    Max.   :9.000
```

- La acidez fija (`fixed.acidity`) varía entre 3.8 y 14.2, con una mediana de 6.8. Esto indica que la mitad de los vinos tienen una acidez fija de 6.8 o menos.
- La acidez volátil (`volatile.acidity`) tiene un rango similar, pero su mediana es menor (0.26), lo que indica que la mayoría de los vinos tienen una acidez volátil bastante baja.
- El azúcar residual (`residual.sugar`) tiene un rango muy amplio, desde 0.6 hasta 65.8. Su mediana es 5.2, lo que indica que la mitad de los vinos tienen un azúcar residual de 5.2 o menos.
- Los cloruros (`chlorides`) varían entre 0.009 y 0.346, con una mediana de 0.043. Esto indica que la mayoría de los vinos tienen un nivel de cloruros bastante bajo.

- El dióxido de azufre libre (`free.sulfur.dioxide`) y el dióxido de azufre total (`total.sulfur.dioxide`) tienen rangos bastante amplios, lo que indica una gran variabilidad en estos valores entre los vinos.
- La densidad (`density`) de los vinos varía poco, con una mediana de 0.9937.
- El pH varía entre 2.72 y 3.82, con una mediana de 3.18. Esto indica que la mayoría de los vinos tienen un pH ligeramente ácido.
- Los sulfatos (`sulphates`) varían entre 0.22 y 1.08, con una mediana de 0.47.
- El alcohol (`alcohol`) varía entre 8 y 14.2, con una mediana de 10.4. Esto indica que la mitad de los vinos tienen un contenido de alcohol de 10.4% o menos.
- La calidad (`quality`) de los vinos varía entre 3 y 9, con una mediana de 6. Esto indica que la mayoría de los vinos tienen una calidad media.

Coefficiente de asimetría

Para el calculo del coeficiente de asimetría usamos la función proporcionada en las practicas. Esta está contenida en nuestros scripts, así que los cargamos.

Luego, realizamos el calculo del coeficiente.

```
apply(datos_vino, 2, asim)
```

```
##      fixed.acidity  volatile.acidity      citric.acid
##      0.6473548      1.5760137      1.2811353
##      residual.sugar      chlorides  free.sulfur.dioxide
##      1.0764341      5.0202543      1.4058834
## total.sulfur.dioxide      density      pH
##      0.3904706      0.9771742      0.4575022
##      sulphates      alcohol      quality
##      0.9765952      0.4870435      0.1557010
```

De lo anterior podemos decir que:

- `fixed.acidity`, `volatile.acidity`, `citric.acid`, `residual.sugar`, `chlorides`, `free.sulfur.dioxide`, `density`, `pH`, `sulphates` y `alcohol` tienen asimetría positiva. Esto significa que estas variables tienen una cola a la derecha, es decir, un número de valores extremadamente altos.
- En particular, `chlorides` tiene un coeficiente de asimetría muy alto (5.02), lo que indica una fuerte asimetría positiva. Esto sugiere que hay algunos vinos con un nivel de cloruros mucho más alto que la mayoría.
- `total.sulfur.dioxide` tiene una asimetría positiva más baja (0.39), lo que indica que su distribución es más simétrica que las otras variables.

- quality tiene un coeficiente de asimetría cercano a 0 (0.15), lo que indica que su distribución es bastante simétrica. Esto sugiere que hay un equilibrio entre los vinos de calidad inferior y superior.

Estadísticos por calidades de vinos

```
# Crea una copia de datos_vino sin la columna quality
datos_vino_sin_quality <- datos_vino[, !(names(datos_vino) %in%
"quality")]

# Aplica aggregate() a cada columna de datos_vino_sin_quality
resultados <- lapply(datos_vino_sin_quality, function(x) {
  aggregate(x = x,
    by = list(datos_vino$quality),
    FUN = function(x) return(c(media = mean(x), varianza =
var(x), asim = asim(x))))
})
```

resultados

```
## $fixed.acidity
##   Group.1    x.media x.varianza    x.asim
## 1      3  7.6000000  2.9747368  0.4737407
## 2      4  7.1294479  1.1696213  0.7621756
## 3      5  6.9339739  0.7029693  0.5955481
## 4      6  6.8376706  0.7013090  0.7482633
## 5      7  6.7347159  0.5710375  0.1400228
## 6      8  6.6571429  0.6713136 -0.5018612
## 7      9  7.4200000  0.9670000  0.8448491
##
## $volatile.acidity
##   Group.1    x.media x.varianza    x.asim
## 1      3  0.333250000  0.019832303  0.881019984
## 2      4  0.381226994  0.030089535  1.375039840
## 3      5  0.302010981  0.010013261  1.426030891
## 4      6  0.260564149  0.007769026  1.531596857
## 5      7  0.262767045  0.008300384  0.808666124
## 6      8  0.277400000  0.011670356  0.974598322
## 7      9  0.298000000  0.003320000  0.214034185
##
## $citric.acid
##   Group.1    x.media    x.varianza    x.asim
## 1      3  0.336000000  0.006635789 -0.028282108
## 2      4  0.304233129  0.026849254  0.463520975
## 3      5  0.337652711  0.019828552  0.701458900
## 4      6  0.338025478  0.014238339  1.995302881
## 5      7  0.325625000  0.006269916  0.750045879
## 6      8  0.326514286  0.007299849  0.536131973
## 7      9  0.386000000  0.006730000  0.140320839
##
```

```

## $residual.sugar
##   Group.1      x.media x.varianza      x.asim
## 1      3  6.3925000 28.2740197  0.5672563
## 2      4  4.6282209 17.3356493  1.0903758
## 3      5  7.3349691 28.3835274  0.4863829
## 4      6  6.4416060 26.6940135  1.4254225
## 5      7  5.1864773 18.4727066  1.1296764
## 6      8  5.6714286 18.1765066  0.8309328
## 7      9  4.1200000 14.1320000  0.9015932
##
## $chlorides
##   Group.1      x.media      x.varianza      x.asim
## 1      3  0.0543000000  0.0021592737  3.3572656502
## 2      4  0.0500981595  0.0006701878  5.5398987152
## 3      5  0.0515463281  0.0007020173  4.7466725073
## 4      6  0.0452174704  0.0004183205  4.6055416837
## 5      7  0.0381909091  0.0001144345  1.9888533862
## 6      8  0.0383142857  0.0001732857  2.9107643515
## 7      9  0.0274000000  0.0000553000 -0.2479179961
##
## $free.sulfur.dioxide
##   Group.1      x.media      x.varianza      x.asim
## 1      3  53.3250000 4819.2440789  2.0594029
## 2      4  23.3588957  415.8071272  2.3195984
## 3      5  36.4320522  329.2769733  0.4765345
## 4      6  35.6505914  247.6115859  0.6796834
## 5      7  34.1255682  175.4230657  0.8033747
## 6      8  36.7200000  262.5590805  1.4395381
## 7      9  33.4000000  180.3000000  0.9814352
##
## $total.sulfur.dioxide
##   Group.1      x.media      x.varianza      x.asim
## 1      3  1.706000e+02  1.161186e+04  8.107155e-01
## 2      4  1.252791e+02  2.782960e+03  2.064177e-01
## 3      5  1.509046e+02  1.943592e+03 -3.170024e-02
## 4      6  1.370473e+02  1.704552e+03  3.559145e-01
## 5      7  1.251148e+02  1.072103e+03  5.003653e-01
## 6      8  1.261657e+02  1.089418e+03  5.369317e-01
## 7      9  1.160000e+02  3.930000e+02 -4.392805e-01
##
## $density
##   Group.1      x.media      x.varianza      x.asim
## 1      3  9.948840e-01  8.012288e-06  3.056671e-01
## 2      4  9.942767e-01  6.063201e-06  1.819554e-01
## 3      5  9.952626e-01  6.475670e-06  6.141087e-02
## 4      6  9.939613e-01  9.141611e-06  1.850747e+00
## 5      7  9.924524e-01  7.659998e-06  9.774719e-01
## 6      8  9.922359e-01  7.771407e-06  1.116485e+00
## 7      9  9.914600e-01  9.724250e-06  1.036894e+00
##

```

```
## $pH
##   Group.1      x.media  x.varianza      x.asim
## 1      3  3.187500000  0.044030263  0.081240164
## 2      4  3.182883436  0.026794721  0.652699610
## 3      5  3.168833219  0.019851522  0.651182115
## 4      6  3.188598726  0.022799856  0.422825860
## 5      7  3.213897727  0.025091161  0.269147971
## 6      8  3.218685714  0.023191941  0.088280675
## 7      9  3.308000000  0.006870000 -0.001433029
##
## $sulphates
##   Group.1      x.media x.varianza      x.asim
## 1      3  0.47450000  0.01436289  0.46240808
## 2      4  0.47613497  0.01391769  0.68835350
## 3      5  0.48220316  0.00964425  0.84465236
## 4      6  0.49110555  0.01284980  0.97910851
## 5      7  0.50310227  0.01695066  0.95078147
## 6      8  0.48622857  0.02163052  0.95612331
## 7      9  0.46600000  0.00858000  0.42784736
##
## $alcohol
##   Group.1      x.media  x.varianza      x.asim
## 1      3 10.34500000  1.49839474  0.01547826
## 2      4 10.15245399  1.00644456  0.69746876
## 3      5  9.80884008  0.71751960  1.07892774
## 4      6 10.57537155  1.31739018  0.39767478
## 5      7 11.36793561  1.55385152 -0.30465601
## 6      8 11.63600000  1.63875402 -0.89139282
## 7      9 12.18000000  1.02700000 -0.97667486
```

De lo anterior podemos ver que:

1. La acidez fija (`fixed.acidity`) tiende a aumentar con la calidad del vino. Los vinos de calidad 9 tienen la mayor media y varianza.
2. La acidez volátil (`volatile.acidity`) tiende a disminuir a medida que aumenta la calidad del vino. Los vinos de calidad 3 tienen la mayor media y varianza.
3. El ácido cítrico (`citric.acid`) parece aumentar ligeramente con la calidad del vino. Los vinos de calidad 9 tienen la mayor media.
4. El azúcar residual (`residual.sugar`) no muestra una tendencia clara con respecto a la calidad del vino.
5. Los cloruros (`chlorides`) tienden a disminuir a medida que aumenta la calidad del vino.
6. El dióxido de azufre libre (`free.sulfur.dioxide`) y total (`total.sulfur.dioxide`) no muestran una tendencia clara con respecto a la calidad del vino.

7. La densidad (density) tiende a disminuir a medida que aumenta la calidad del vino.
8. El pH no muestra una tendencia clara con respecto a la calidad del vino.
9. Los sulfatos (sulphates) parecen aumentar ligeramente con la calidad del vino.
10. El alcohol (alcohol) tiende a aumentar con la calidad del vino. Los vinos de calidad 9 tienen la mayor media y varianza.

Matriz de Covarianza

Usando la función definida en las practicas obtenemos la matriz de covarianza

```
mcof(datos_vino)
```

```
##               fixed.acidity volatile.acidity   citric.acid
## fixed.acidity      0.7119681971    -1.930176e-03  2.952648e-02
## volatile.acidity   -0.0019301764     1.015747e-02 -1.822905e-03
## citric.acid        0.0295264821    -1.822905e-03  1.464280e-02
## residual.sugar     0.3809440223     3.285862e-02  5.781712e-02
## chlorides          0.0004255386     1.552458e-04  3.023221e-04
## free.sulfur.dioxide -0.7087739060    -1.662665e-01  1.935902e-01
## total.sulfur.dioxide 3.2653465871     3.822758e-01  6.228615e-01
## density            0.0006695405     8.172264e-06  5.410275e-05
## pH                 -0.0542537470    -4.856539e-04 -2.991734e-03
## sulphates          -0.0016506552    -4.109063e-04  8.607072e-04
## alcohol            -0.1255071925     8.398008e-03 -1.127594e-02
## quality            -0.0849299662    -1.737889e-02 -9.868271e-04
##
##               residual.sugar   chlorides free.sulfur.dioxide
## fixed.acidity      0.380944022  4.255386e-04             -0.708773906
## volatile.acidity    0.032858624  1.552458e-04             -0.166266506
## citric.acid         0.057817120  3.023221e-04              0.193590244
## residual.sugar     25.720517863  9.825496e-03             25.795310325
## chlorides          0.009825496  4.772363e-04              0.037666806
## free.sulfur.dioxide 25.795310325  3.766681e-02             289.183666769
## total.sulfur.dioxide 86.513636310  1.846498e-01             444.775064918
## density            0.012724567  1.680410e-05              0.014962476
## pH                 -0.148653305 -2.983039e-04             -0.001586231
## sulphates          -0.015431592  4.178834e-05              0.114914468
## alcohol            -2.812166069 -9.682258e-03             -5.233439970
## quality            -0.438226605 -4.061276e-03              0.122853162
##
##               total.sulfur.dioxide      density      pH
## fixed.acidity      3.26534659  6.695405e-04 -5.425375e-02
## volatile.acidity    0.38227584  8.172264e-06 -4.856539e-04
## citric.acid         0.62286152  5.410275e-05 -2.991734e-03
## residual.sugar     86.51363631  1.272457e-02 -1.486533e-01
## chlorides          0.18464978  1.680410e-05 -2.983039e-04
## free.sulfur.dioxide 444.77506492  1.496248e-02 -1.586231e-03
## total.sulfur.dioxide 1805.71675147  6.733827e-02  1.489118e-02
```



```
## density          0.06733827  8.943698e-06 -4.225998e-05
## pH               0.01489118 -4.225998e-05  2.279653e-02
## sulphates        0.65251133  2.542228e-05  2.686974e-03
## alcohol          -23.47181150 -2.870844e-03  2.256045e-02
## quality          -6.57540361 -8.133613e-04  1.329389e-02
##               sulphates      alcohol      quality
## fixed.acidity    -1.650655e-03 -0.125507193 -0.0849299662
## volatile.acidity -4.109063e-04  0.008398008 -0.0173788916
## citric.acid      8.607072e-04 -0.011275936 -0.0009868271
## residual.sugar   -1.543159e-02 -2.812166069 -0.4382266051
## chlorides        4.178834e-05 -0.009682258 -0.0040612763
## free.sulfur.dioxide 1.149145e-01 -5.233439970  0.1228531625
## total.sulfur.dioxide 6.525113e-01 -23.471811503 -6.5754036072
## density          2.542228e-05 -0.002870844 -0.0008133613
## pH               2.686974e-03  0.022560445  0.0132938853
## sulphates        1.302205e-02 -0.002447856  0.0054243430
## alcohol          -2.447856e-03  1.514117789  0.4746294463
## quality          5.424343e-03  0.474629446  0.7841955475
```

De lo anterior podemos ver que:

1. `fixed.acidity` y `volatile.acidity`: La covarianza es negativa, lo que indica que tienden a moverse en direcciones opuestas. Cuando la acidez fija aumenta, la acidez volátil tiende a disminuir, y viceversa.
2. `fixed.acidity` y `citric.acid`: La covarianza es positiva, lo que indica que tienden a moverse juntas. Cuando la acidez fija aumenta, el ácido cítrico también tiende a aumentar.
3. `fixed.acidity` y `residual.sugar`: La covarianza es positiva, lo que indica que tienden a moverse juntas. Cuando la acidez fija aumenta, el azúcar residual también tiende a aumentar.
4. `volatile.acidity` y `residual.sugar`: La covarianza es positiva, lo que indica que tienden a moverse juntas. Cuando la acidez volátil aumenta, el azúcar residual también tiende a aumentar.
5. `citric.acid` y `residual.sugar`: La covarianza es positiva, lo que indica que tienden a moverse juntas. Cuando el ácido cítrico aumenta, el azúcar residual también tiende a aumentar.

Matriz de correlaciones

```
cor(datos_vino)
##               fixed.acidity volatile.acidity citric.acid
residual.sugar      1.00000000      -0.02269729  0.289180698
## fixed.acidity      0.08902070      1.00000000      -0.02269729
## volatile.acidity    -0.02269729      0.06428606      1.00000000
citric.acid          -0.149471811
```

## citric.acid	0.28918070	-0.14947181	1.000000000	
0.09421162				
## residual.sugar	0.08902070	0.06428606	0.094211624	
1.00000000				
## chlorides	0.02308564	0.07051157	0.114364448	
0.08868454				
## free.sulfur.dioxide	-0.04939586	-0.09701194	0.094077221	
0.29909835				
## total.sulfur.dioxide	0.09106976	0.08926050	0.121130798	
0.40143931				
## density	0.26533101	0.02711385	0.149502571	
0.83896645				
## pH	-0.42585829	-0.03191537	-0.163748211	-
0.19413345				
## sulphates	-0.01714299	-0.03572815	0.062330940	-
0.02666437				
## alcohol	-0.12088112	0.06771794	-0.075728730	-
0.45063122				
## quality	-0.11366283	-0.19472297	-0.009209091	-
0.09757683				
##	chlorides	free.sulfur.dioxide		
total.sulfur.dioxide				
## fixed.acidity	0.02308564	-0.0493958591		
0.091069756				
## volatile.acidity	0.07051157	-0.0970119393		
0.089260504				
## citric.acid	0.11436445	0.0940772210		
0.121130798				
## residual.sugar	0.08868454	0.2990983537		
0.401439311				
## chlorides	1.00000000	0.1013923521		
0.198910300				
## free.sulfur.dioxide	0.10139235	1.0000000000		
0.615500965				
## total.sulfur.dioxide	0.19891030	0.6155009650		
1.000000000				
## density	0.25721132	0.2942104109		
0.529881324				
## pH	-0.09043946	-0.0006177961		
0.002320972				
## sulphates	0.01676288	0.0592172458		
0.134562367				
## alcohol	-0.36018871	-0.2501039415		-
0.448892102				
## quality	-0.20993441	0.0081580671		-
0.174737218				
##	density	pH	sulphates	alcohol
## fixed.acidity	0.26533101	-0.4258582910	-0.01714299	-0.12088112
## volatile.acidity	0.02711385	-0.0319153683	-0.03572815	0.06771794
## citric.acid	0.14950257	-0.1637482114	0.06233094	-0.07572873

## residual.sugar	0.83896645	-0.1941334540	-0.02666437	-0.45063122
## chlorides	0.25721132	-0.0904394560	0.01676288	-0.36018871
## free.sulfur.dioxide	0.29421041	-0.0006177961	0.05921725	-0.25010394
## total.sulfur.dioxide	0.52988132	0.0023209718	0.13456237	-0.44889210
## density	1.00000000	-0.0935914935	0.07449315	-0.78013762
## pH	-0.09359149	1.0000000000	0.15595150	0.12143210
## sulphates	0.07449315	0.1559514973	1.00000000	-0.01743277
## alcohol	-0.78013762	0.1214320987	-0.01743277	1.00000000
## quality	-0.30712331	0.0994272457	0.05367788	0.43557472
##	quality			
## fixed.acidity	-0.113662831			
## volatile.acidity	-0.194722969			
## citric.acid	-0.009209091			
## residual.sugar	-0.097576829			
## chlorides	-0.209934411			
## free.sulfur.dioxide	0.008158067			
## total.sulfur.dioxide	-0.174737218			
## density	-0.307123313			
## pH	0.099427246			
## sulphates	0.053677877			
## alcohol	0.435574715			
## quality	1.000000000			

De lo anterior podemos decir que:

1. fixed.acidity y density: Tienen una correlación positiva de 0.27, lo que indica una relación moderada. A medida que la acidez fija aumenta, la densidad también tiende a aumentar.
2. volatile.acidity y quality: Tienen una correlación negativa de -0.19, lo que indica una relación débil. A medida que la acidez volátil aumenta, la calidad tiende a disminuir.
3. residual.sugar y density: Tienen una correlación positiva muy fuerte de 0.84, lo que indica que a medida que el azúcar residual aumenta, la densidad también tiende a aumentar.
4. chlorides y quality: Tienen una correlación negativa de -0.21, lo que indica una relación débil. A medida que los cloruros aumentan, la calidad tiende a disminuir.
5. free.sulfur.dioxide y total.sulfur.dioxide: Tienen una correlación positiva fuerte de 0.62, lo que indica que a medida que el dióxido de azufre libre aumenta, el dióxido de azufre total también tiende a aumentar.
6. alcohol y density: Tienen una correlación negativa muy fuerte de -0.78, lo que indica que a medida que el alcohol aumenta, la densidad tiende a disminuir.

7. alcohol y quality: Tienen una correlación positiva de 0.44, lo que indica una relación moderada. A medida que el alcohol aumenta, la calidad también tiende a aumentar.

Visualización de nuestros datos

Histograma de nuestros datos

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.2

# Obtén los nombres de las columnas de tu conjunto de datos
columnas <- names(datos_vino)

datos_vino$quality <- factor(datos_vino$quality)

# Excluye la columna 'quality' ya que es la que usaremos para el relleno
columnas <- columnas[columnas != "quality"]

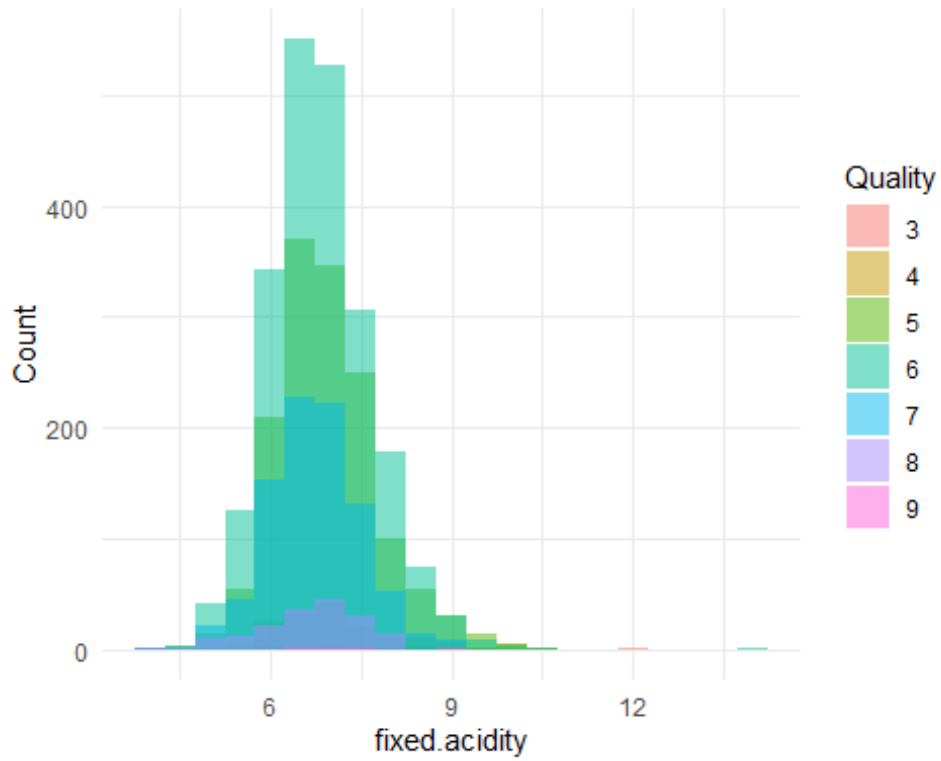
# Crea un histograma para cada columna
for (columna in columnas) {
  p <- ggplot(datos_vino, aes_string(x = columna, fill = "quality")) +
    geom_histogram(binwidth = .5, position = "identity", alpha = 0.5) +
    labs(x = columna, y = "Count", fill = "Quality") +
    theme_minimal()

  # Imprime el gráfico en la consola
  print(p)

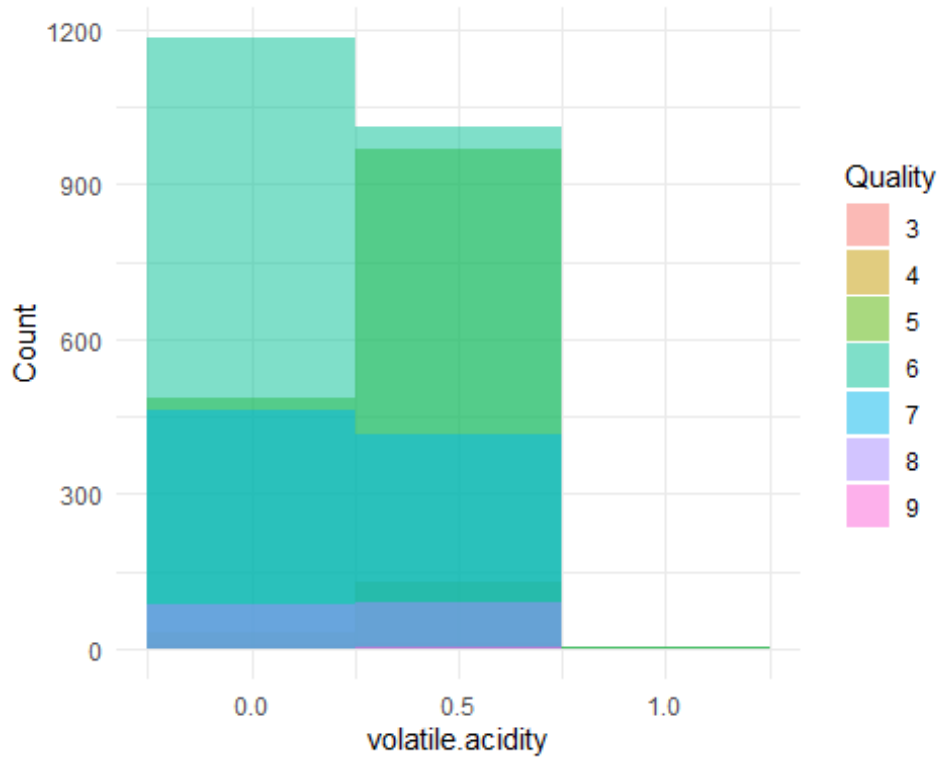
  # Guarda el gráfico en un archivo
  ggsave(paste0("../Figuras/histograma_", columna, ".png"), plot = p)
}

## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with `aes()`.
## i See also `vignette("ggplot2-in-packages")` for more information.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning
was
## generated.

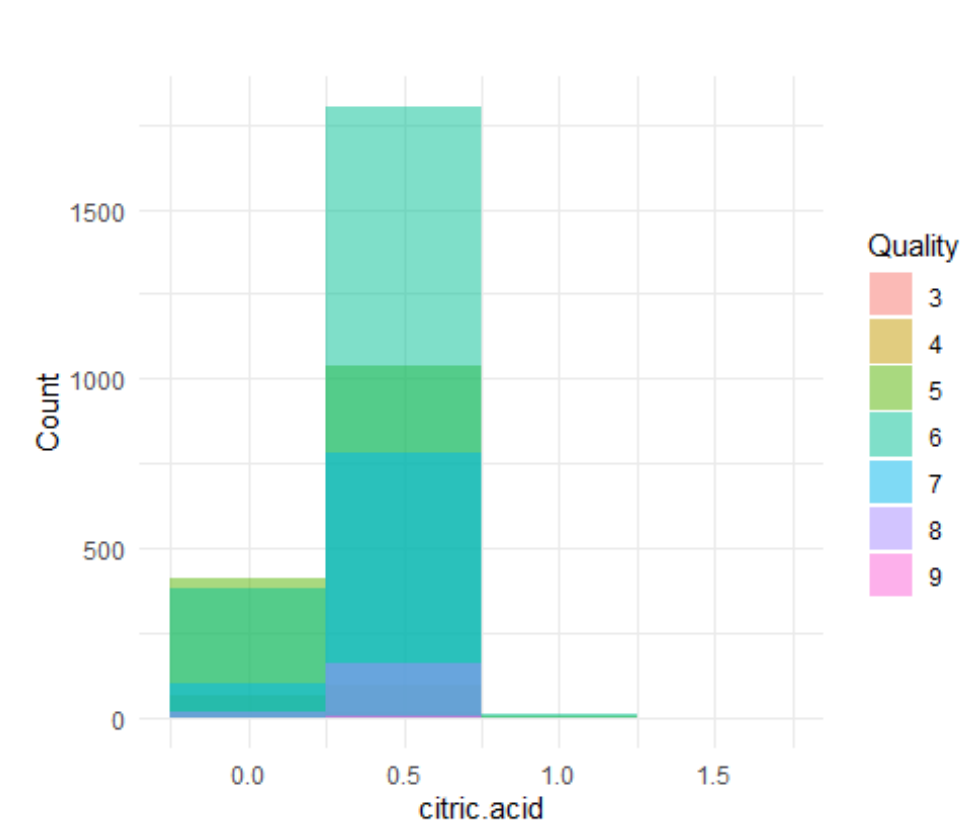
## Saving 5 x 4 in image
```



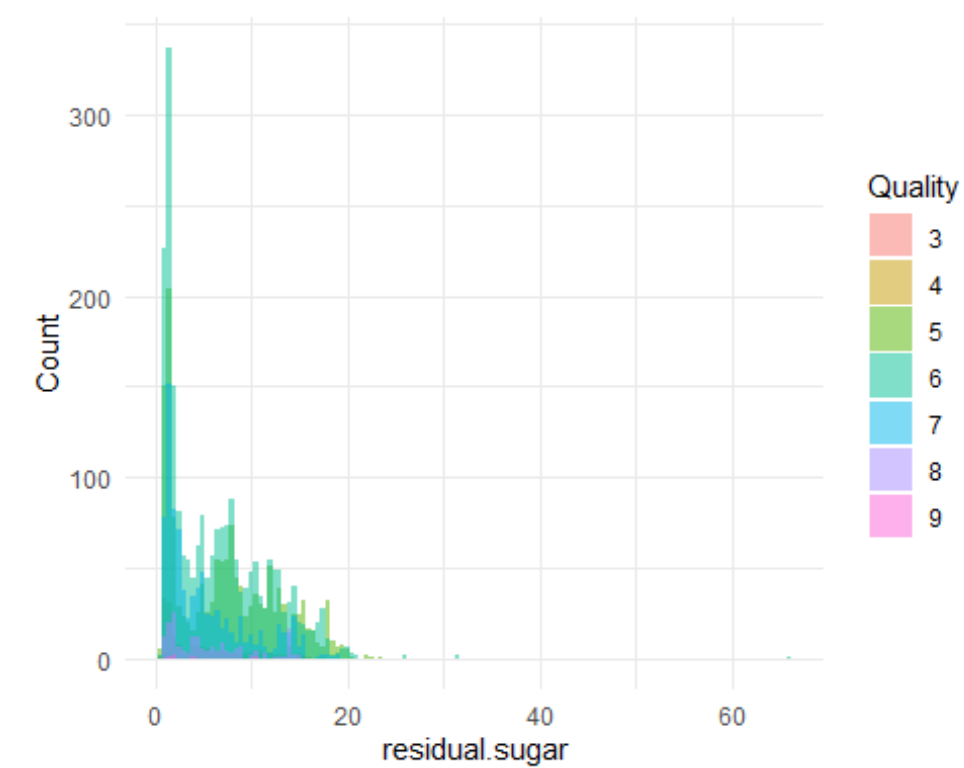
```
## Saving 5 x 4 in image
```



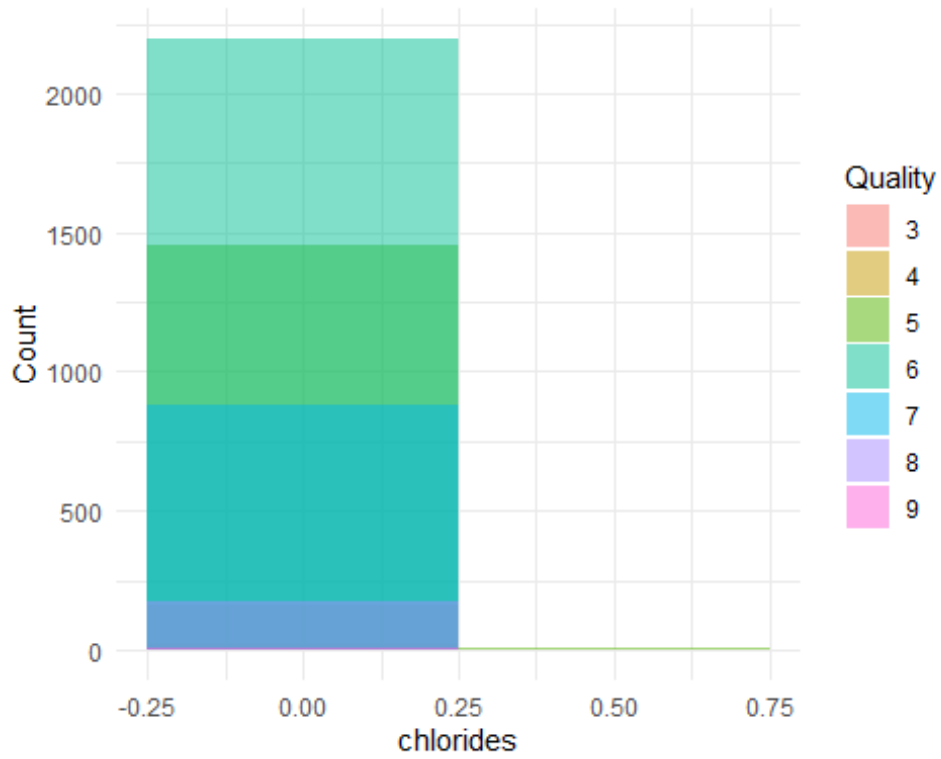
```
## Saving 5 x 4 in image
```



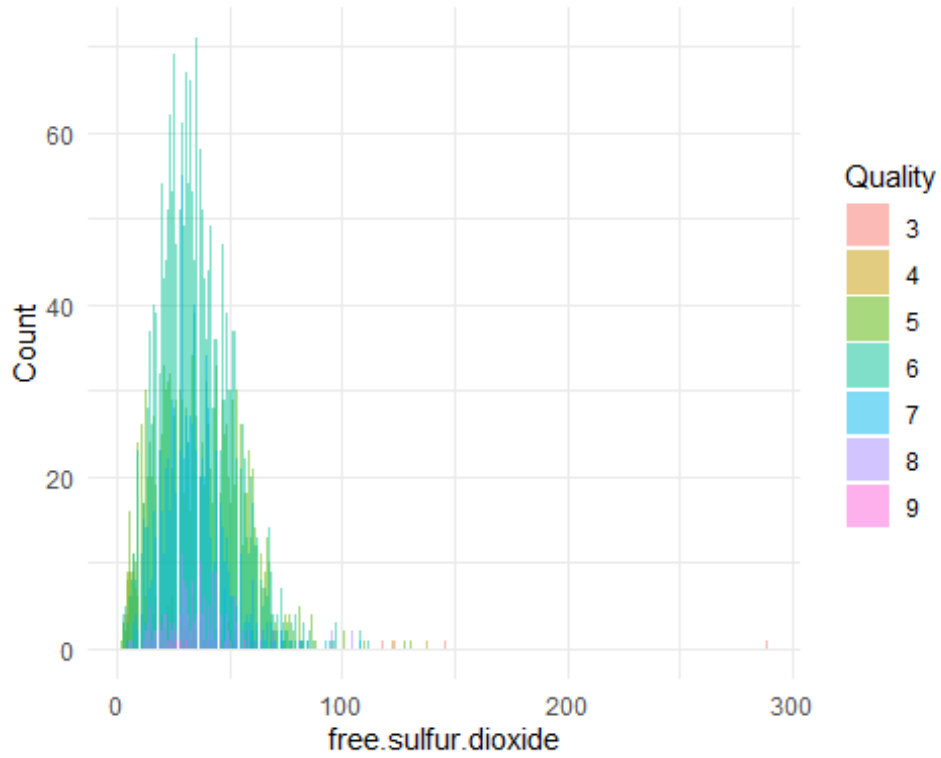
```
## Saving 5 x 4 in image
```



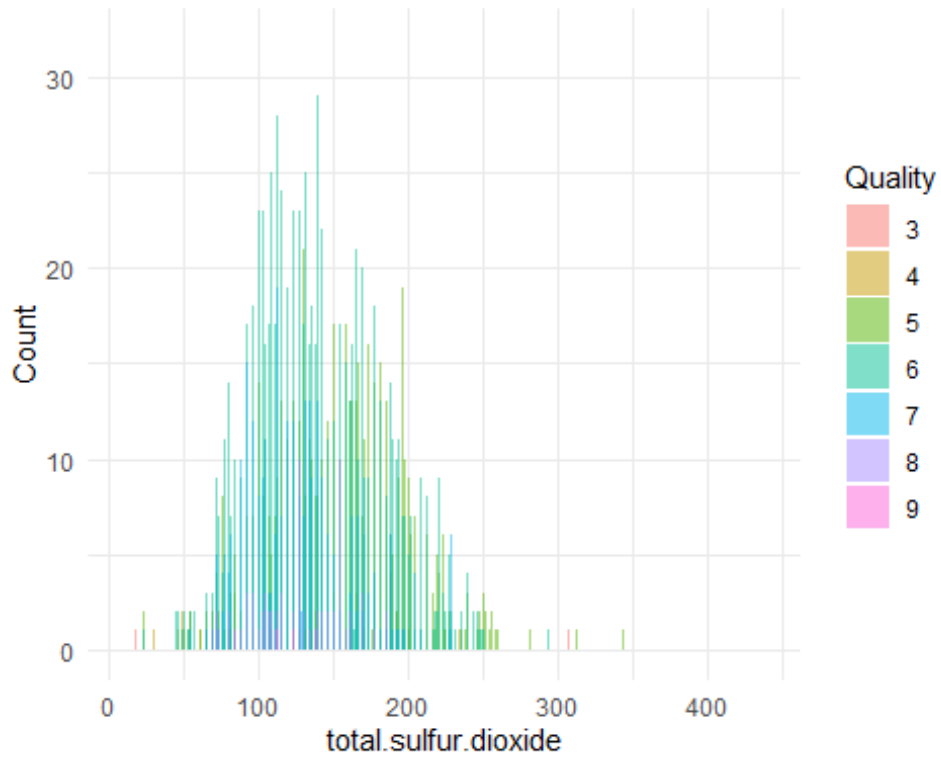
```
## Saving 5 x 4 in image
```



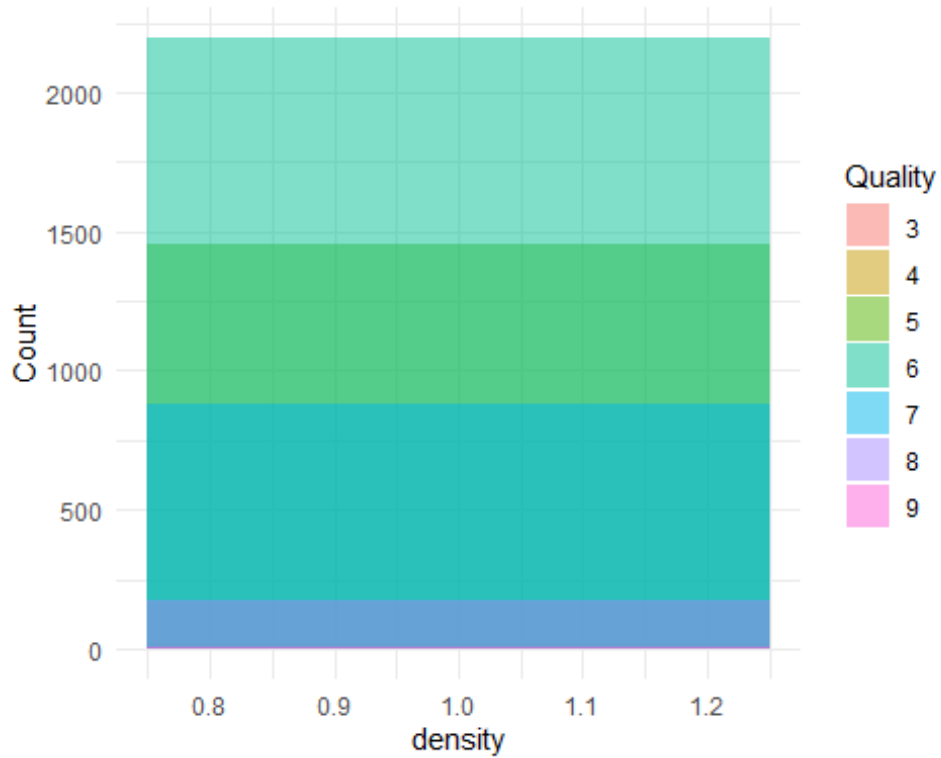
```
## Saving 5 x 4 in image
```



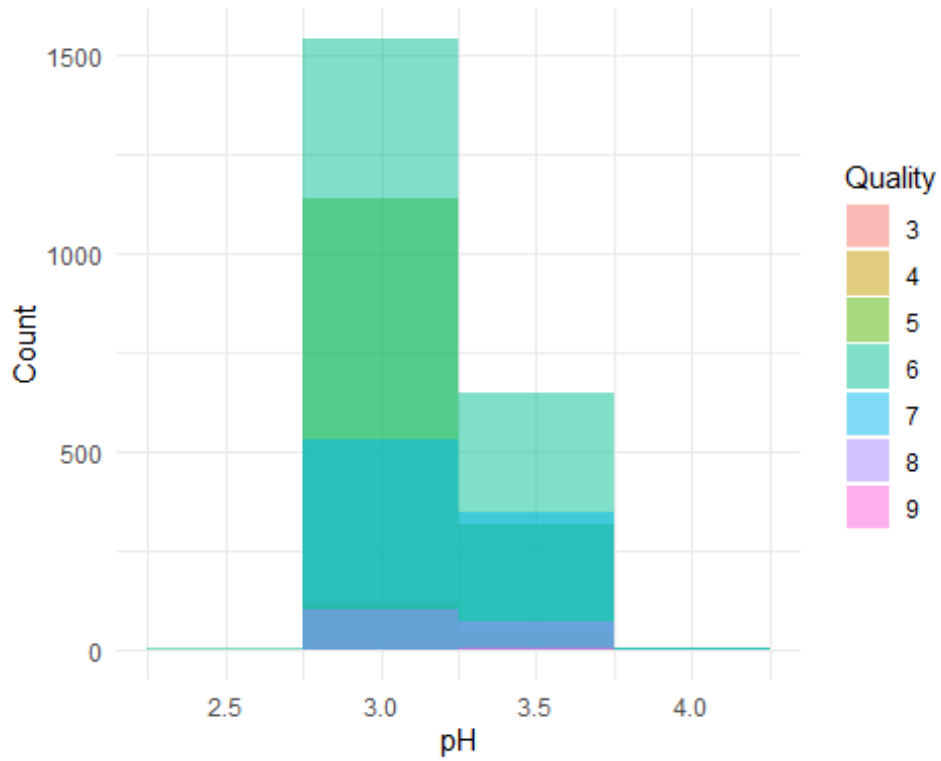
```
## Saving 5 x 4 in image
```



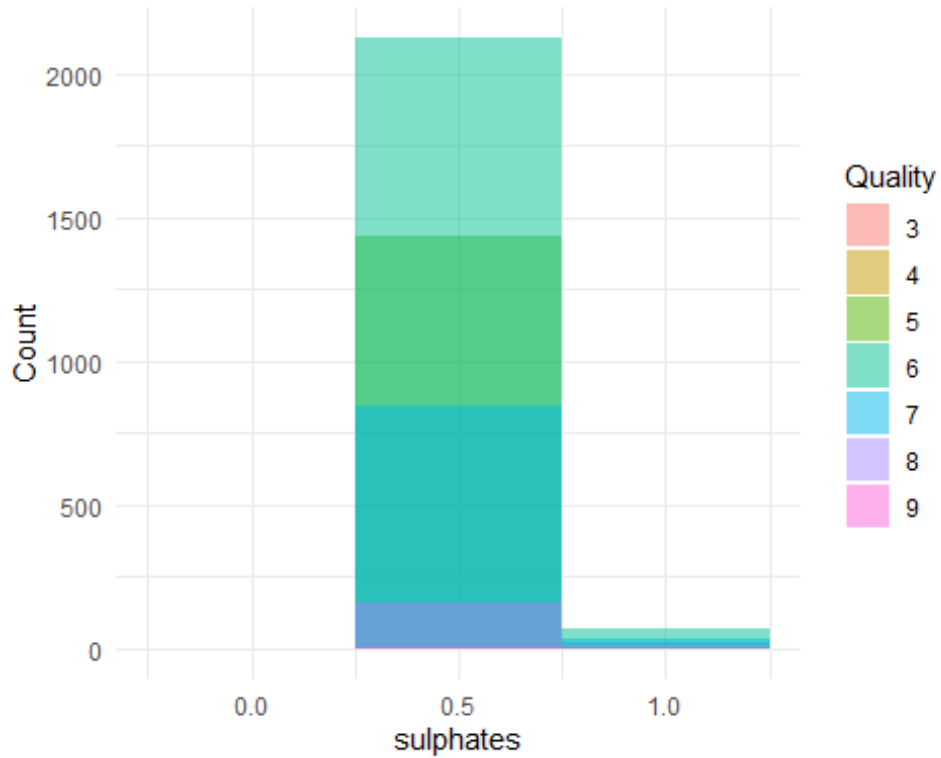
```
## Saving 5 x 4 in image
```



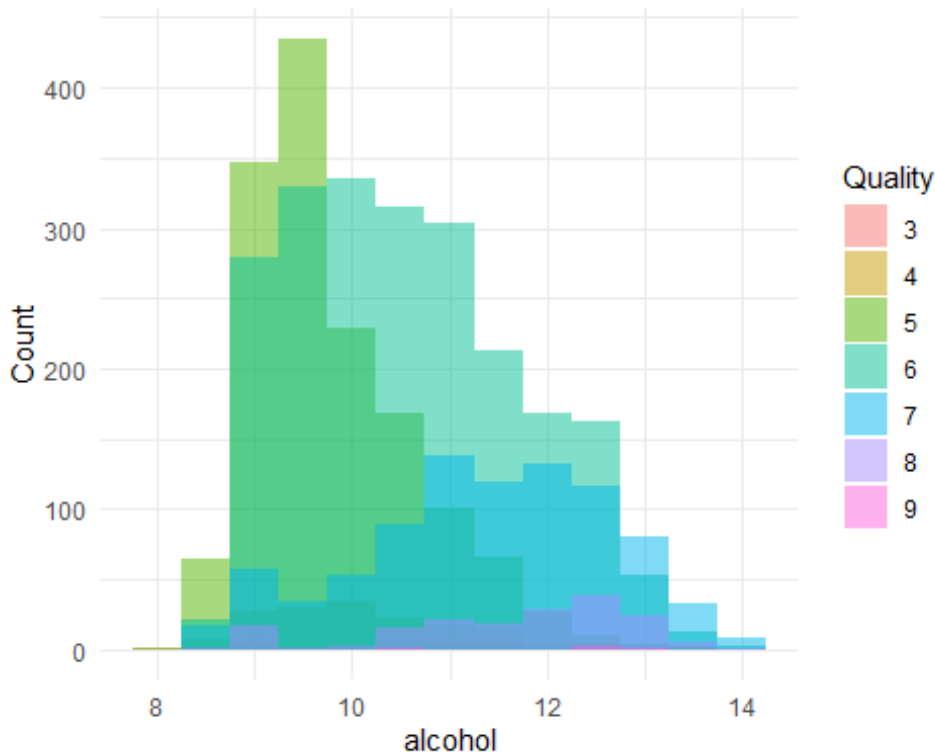
```
## Saving 5 x 4 in image
```

```
## Saving 5 x 4 in image
```



```
## Saving 5 x 4 in image
```



En el anterior apartado generamos cada una de las graficas de los histogramas de cada una de las variables de nuestro dataSet y las dividimos coloreadas por el tipo de calidad de los vinos.

Algo que se puede resaltar es que, la cantidad de alcohol en un vino tiene relación con la calificación en su calidad.

Diagrama de bigotes

```
library(reshape2)

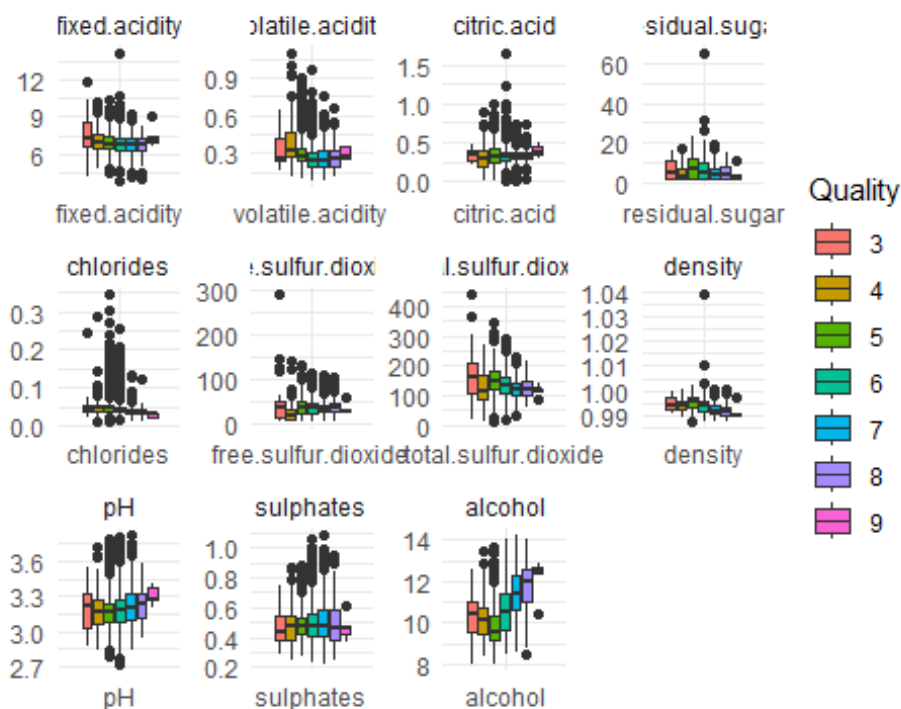
## Warning: package 'reshape2' was built under R version 4.3.2

# Reorganiza Los datos en formato Largo
datos_largos <- melt(datos_vino)

## Using quality as id variables

# Crea el gráfico
p <- ggplot(datos_largos, aes(x = variable, y = value, fill =
factor(quality))) +
  geom_boxplot() +
  facet_wrap(~ variable, scales = "free") +
  theme_minimal() +
  labs(x = "", y = "", fill = "Quality")

# Imprime el gráfico
print(p)
```



```
# Guarda el gráfico en un archivo
ggsave("../Figuras/diagramas_bigotes_por_calidad.png", plot = p)

## Saving 5 x 4 in image
```

De los diagramas de bigotes anteriores podemos observar que, todas las variables poseen datos atípicos. Y que estos datos tienen en cada una de las calidades una cantidad amplia de las mismas.

Grafico de dispersión

```
library(GGally)

## Warning: package 'GGally' was built under R version 4.3.2

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2

# Crea el diagrama de dispersión
p <- ggpairs(datos_vino)

# Imprime el gráfico
print(p)

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


Del diagrama de dispersión podemos ver el resumen de todas las anteriores graficas realizadas. Por otro lado, podemos ver cómo la forma de todas las distribuciones de nuestros datos vienen siendo parecidas a una forma gaussiana.

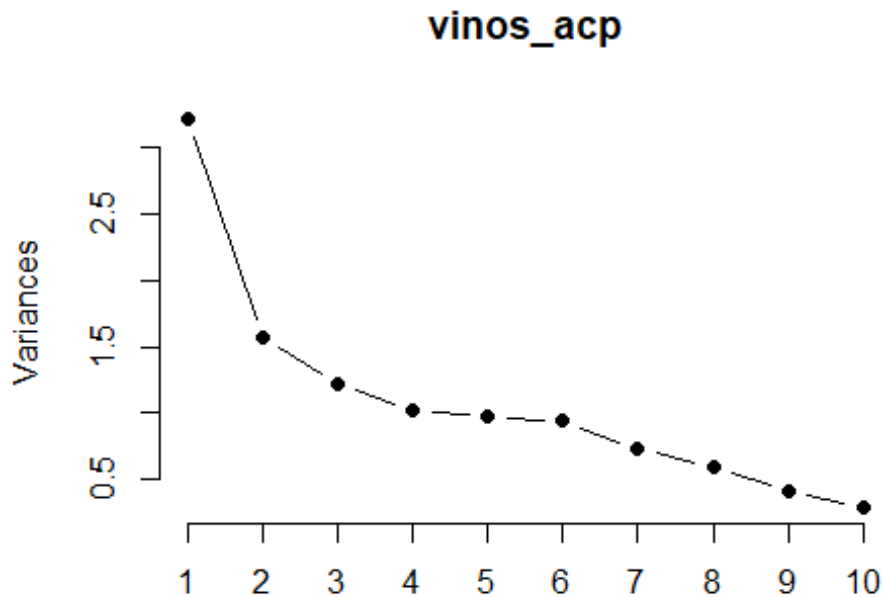
Análisis de componentes principales

Para realizar el análisis de las componentes principales de nuestro dataSet, haremos uso de la función que trae R ya integrada.

```
vinos_acp <- prcomp(datos_vino_sin_quality, scale. = TRUE)
summary(vinos_acp)

## Importance of components:
##              PC1      PC2      PC3      PC4      PC5      PC6
PC7
## Standard deviation    1.7951 1.2551 1.1053 1.00922 0.98658 0.96889
0.85241
## Proportion of Variance 0.2929 0.1432 0.1111 0.09259 0.08848 0.08534
0.06605
## Cumulative Proportion 0.2929 0.4361 0.5472 0.63979 0.72827 0.81361
0.87967
##              PC8      PC9      PC10     PC11
## Standard deviation    0.77418 0.64354 0.53804 0.14370
## Proportion of Variance 0.05449 0.03765 0.02632 0.00188
## Cumulative Proportion 0.93416 0.97181 0.99812 1.00000

plot(vinos_acp, type = "l", pch = 19)
```



```
png("../Figuras/diagrama_metodo_codo.png")
plot(vinos_acp, type = "l", pch = 19)
dev.off()
```

```
## png
## 2
```

De lo anterior podemos decir que:

- La primera componente principal (PC1) explica el 29.29% de la varianza en los datos. Esto indica que PC1 captura una gran parte de la información en el conjunto de datos.
- La segunda componente principal (PC2) explica un 14.32% adicional de la varianza, llevando el total acumulado a 43.61%.
- La tercera componente principal (PC3) explica un 11.11% adicional de la varianza, llevando el total acumulado a 54.72%.
- Las componentes principales restantes (PC4 a PC11) explican el resto de la varianza, con cada una explicando una cantidad decreciente.
- En total, las primeras tres componentes principales (PC1, PC2 y PC3) explican más del 50% de la varianza en los datos. Esto sugiere que estas tres componentes principales pueden ser suficientes para capturar la mayoría de la información en el conjunto de datos.
- La última componente principal (PC11) explica solo el 0.188% de la varianza, lo que sugiere que esta componente principal puede no ser muy útil para entender la variación en los datos.
- En general, estos resultados sugieren que se podría reducir la dimensionalidad del conjunto de datos a tres o quizás incluso dos dimensiones sin perder demasiada información.

Ahora miremos cómo cada variable contribuye a la componente:

```
vinos_acp$rotation[, 1:3]
```

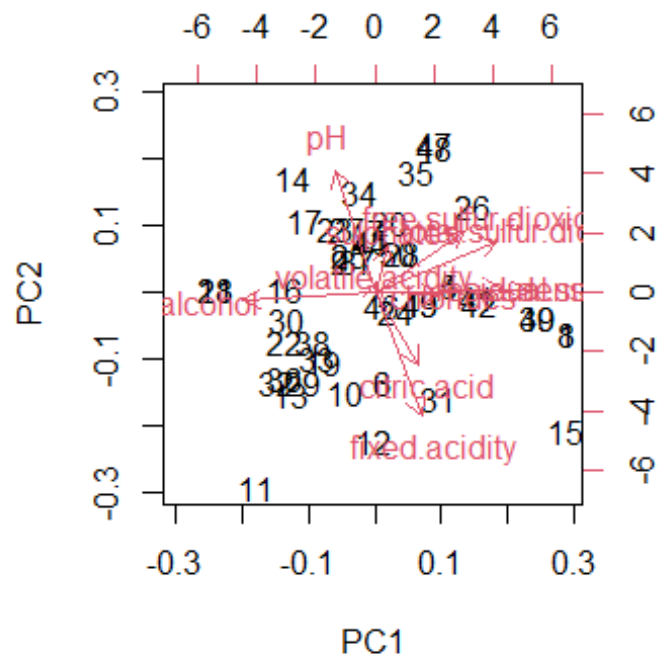
```
##                PC1                PC2                PC3
## fixed.acidity    0.157218451 -0.587558208  0.1213683
## volatile.acidity 0.005089494  0.051728054 -0.5909715
## citric.acid      0.144049843 -0.345294562  0.5043969
## residual.sugar   0.427408368  0.008749392 -0.2143199
## chlorides        0.212011065 -0.008800308 -0.1023674
## free.sulfur.dioxide 0.300334387  0.290355136  0.2794101
## total.sulfur.dioxide 0.406652203  0.244032391  0.1243753
## density          0.511523597  0.006296796 -0.1292029
## pH               -0.128831885  0.581344397  0.1266715
## sulphates        0.043379327  0.222695370  0.4332440
## alcohol          -0.437237835 -0.035568666  0.1059032
```

De lo anterior podemos decir que:

- La primera componente principal (PC1) está fuertemente influenciada por `density`, `total.sulfur.dioxide`, `residual.sugar`, y `free.sulfur.dioxide`, con coeficientes de 0.511523597, 0.406652203, 0.427408368, y 0.300334387 respectivamente. Esto indica que estos factores son los más importantes en la variación de los datos a lo largo de PC1.
- La segunda componente principal (PC2) está principalmente influenciada por `fixed.acidity` y `pH`, con coeficientes de -0.587558208 y 0.581344397 respectivamente. Esto indica que estos factores son los más importantes en la variación de los datos a lo largo de PC2.
- La tercera componente principal (PC3) está fuertemente influenciada por `volatile.acidity`, `citric.acid`, y `sulphates`, con coeficientes de -0.5909715, 0.5043969, y 0.4332440 respectivamente. Esto indica que estos factores son los más importantes en la variación de los datos a lo largo de PC3.
- En general, estos resultados sugieren que `density`, `total.sulfur.dioxide`, `residual.sugar`, `free.sulfur.dioxide`, `fixed.acidity`, `pH`, `volatile.acidity`, `citric.acid`, y `sulphates` son los factores más importantes en la variación de los datos de vino.

Tenemos tambien, el siguiente grafico que nos resume la relación entre las componentes.

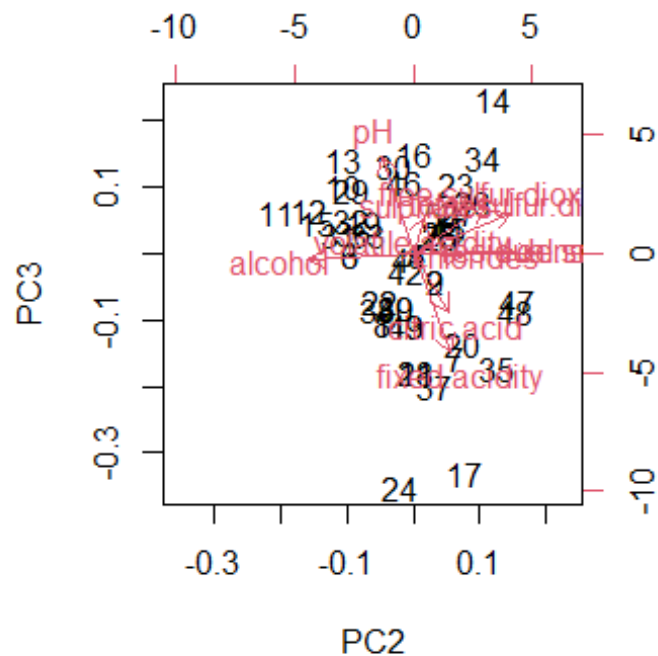
```
# Genera un biplot con solo las primeras 5 variables
# Subconjunto vinos_acp$x para incluir solo las primeras 50 observaciones
vinos_acp_sub <- vinos_acp
vinos_acp_sub$x <- vinos_acp$x[1:50,1:2]
biplot(vinos_acp_sub)
```



```
# Genera un biplot con las primeras 50 observaciones
png("../Figuras/diagrama_biplot1.png", width = 2000, height = 2000, res =
300)
biplot(vinos_acp_sub)
dev.off()

## png
## 2

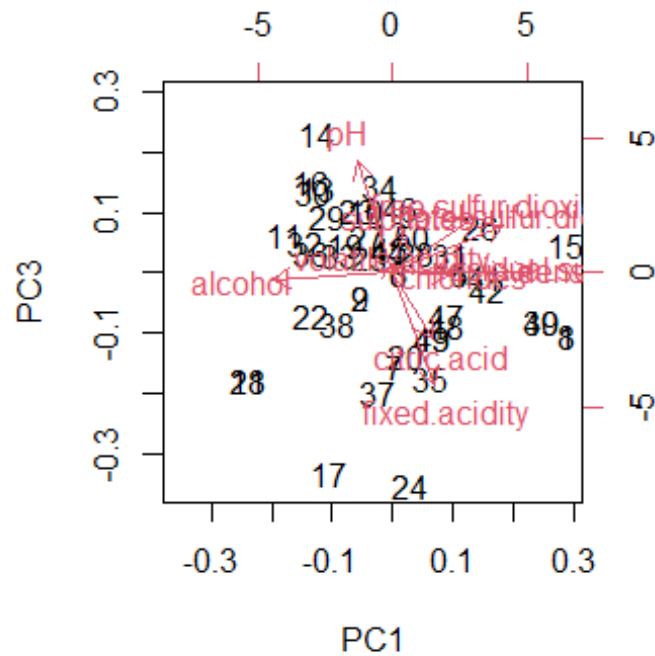
# Genera un biplot con solo las primeras 5 variables
# Subconjunta vinos_acp$x para incluir solo las primeras 50 observaciones
vinos_acp_sub <- vinos_acp
vinos_acp_sub$x <- vinos_acp$x[1:50,2:3]
biplot(vinos_acp_sub)
```

```
# Genera un biplot con las primeras 50 observaciones
png("../Figuras/diagrama_biplot2.png", width = 2000, height = 2000, res =
300)
biplot(vinos_acp_sub)
dev.off()

## png
## 2

# Genera un biplot con solo las primeras 5 variables
# Subconjunta vinos_acp$x para incluir solo las primeras 50 observaciones
vinos_acp_sub <- vinos_acp
vinos_acp_sub$x <- vinos_acp$x[1:50, c(1,3)]
biplot(vinos_acp_sub)
```



```
# Genera un biplot con las primeras 50 observaciones
png("../Figuras/diagrama_biplot3.png", width = 2000, height = 2000, res =
300)
biplot(vinos_acp_sub)
dev.off()

## png
## 2
```

De lo anterior podemos decir que, la primera componente es una composición entre el alcohol, los chloridres, la azucar residual y el total de dioxido de sulfuro.

Por otro lado, la segunda componente principalmente del PH y de la acides ajustada.

Aunque para la relación entre la componente 1 y la componente 2 el Alcohol tiene una relevancia alta para la componente 1. Al contrastar con la componente 3 nos damos cuenta que, para dicha componente el aporte que el alcohol hace a esta, contrastando contra la componente 2, es mucho menor, ya que no se muestra un vector diagonal en el alcohol en dicho resultado y vemos cómo casi el vector de alcohol es paralelo a la componente 2.

Análisis Factorial

En este caso igualmente usaremos la funcion integrada de R para realizar el analisis factorial usando el metodo de las componentes principales.

Aplicamos el analisis factorial, eliminando el factor que no queremos incluir en el analisis. Definimos como 3 el numero de factores, tomando el resultado del analisis de componentes previo y escogemos la rotación varimax, para simplificar la interpretación de los factores.

```
library(psych)

## Warning: package 'psych' was built under R version 4.3.2

##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha

datos_vinos_af <- principal(datos_vino[, -12], nfactors = 3, rotate =
"varimax")

datos_vinos_af$values

## [1] 3.22225389 1.57523993 1.22167134 1.01852235 0.97333458 0.93874151
## [7] 0.72659802 0.59935848 0.41414367 0.28948714 0.02064909

datos_vinos_af$loadings

##
## Loadings:
##
##          RC1    RC2    RC3
## fixed.acidity    0.106  0.791
## volatile.acidity  0.126 -0.132 -0.631
## citric.acid             0.537  0.521
## residual.sugar    0.782  0.129 -0.129
## chlorides          0.384
## free.sulfur.dioxide 0.544 -0.204  0.426
## total.sulfur.dioxide 0.745 -0.125  0.275
## density           0.913  0.175
## pH                -0.102 -0.743  0.209
## sulphates             -0.202  0.519
## alcohol            -0.787 -0.113
##
##          RC1    RC2    RC3
## SS loadings  3.107 1.648 1.264
## Proportion Var 0.282 0.150 0.115
## Cumulative Var 0.282 0.432 0.547

datos_vinos_af$communality

##          fixed.acidity    volatile.acidity    citric.acid
##          0.6414537          0.4309639          0.5654892
##          residual.sugar    chlorides    free.sulfur.dioxide
##          0.6448702          0.1577601          0.5188279
```

## total.sulfur.dioxide	density	pH
## 0.6455580	0.8635796	0.6054545
## sulphates	alcohol	
## 0.3134930	0.6317151	

Vemos que el primer factor explica 3.22 de la varianza, el segundo factor explica 1.57 y el tercer factor explica 1.22. Los factores restantes explican menos de 1 de la varianza, lo que sugiere que los primeros tres factores son los más importantes, confirmando nuestro anterior análisis.

En cuanto a las cargas factoriales, la variable fixed.acidity tiene una alta correlación con el segundo factor (0.791), mientras que residual.sugar tiene una alta correlación con el primer factor (0.782). Las variables que tienen cargas factoriales altas en el mismo factor se pueden considerar como agrupadas o relacionadas, por ende podríamos decir que residual.sugar esta relacionada a total.sulfur.dioxide y a density, al tener cargas altas sobre el primer factor.

Finalmente para las comunales, que representan la cantidad de varianza de cada variable que es explicada por los factores, tenemos que la variable density tiene una comunalidad de 0.86, lo que significa que el 86% de su varianza es explicada por los tres factores.

```
datos_vinos_af$complexity
```

## fixed.acidity	volatile.acidity	citric.acid
## 1.050195	1.170285	2.038962
## residual.sugar	chlorides	free.sulfur.dioxide
## 1.110084	1.141881	2.201571
## total.sulfur.dioxide	density	pH
## 1.329758	1.074104	1.197339
## sulphates	alcohol	
## 1.320860	1.041533	

```
indice_Hoffman <- function(x) (sum(x^2)^2)/sum(x^4)
comp <- apply(datos_vinos_af$loadings, 1, indice_Hoffman)
comp
```

## fixed.acidity	volatile.acidity	citric.acid
## 1.050195	1.170285	2.038962
## residual.sugar	chlorides	free.sulfur.dioxide
## 1.110084	1.141881	2.201571
## total.sulfur.dioxide	density	pH
## 1.329758	1.074104	1.197339
## sulphates	alcohol	
## 1.320860	1.041533	

```
mean(comp)
```

```
## [1] 1.334234
```

Ahora obtenemos los índices de complejidad de Hoffman de cada variable, de los cuales podemos apreciar que la variable free.sulfur.dioxide tiene el índice de Hoffman más alto (2.20), lo que indica que es la variable más importante en la formación de los clusters en tus datos. La siguiente variable más importante es citric.acid con un índice de 2.03.

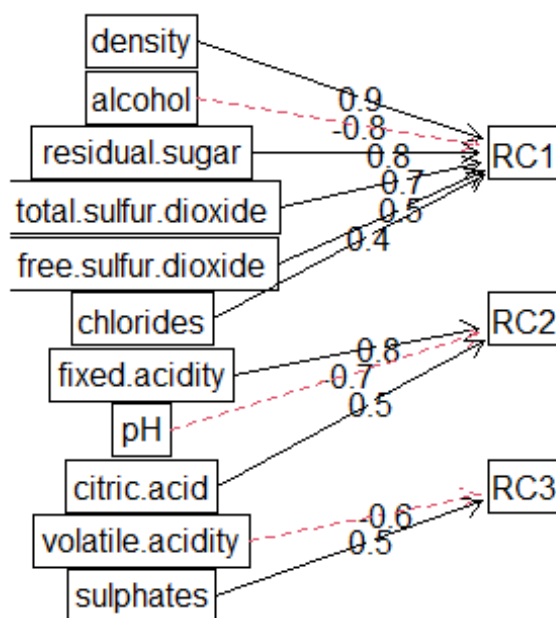
Las variables fixed.acidity, residual.sugar, density y alcohol tienen índices de Hoffman cercanos a 1, lo que indica que son menos importantes para la formación de los clusters y que son influenciadas por menos factores subyacentes.

También tenemos que nuestra media de los índices de complejidad es 1.334234, indicando que las variables generalmente están asociadas de forma moderada con varios factores.

Procedemos a realizar un diagrama para localizar la mayor carga de cada variable.

```
p <- fa.diagram(datos_vinos_af)
```

Components Analysis



```
print(p)

## NULL

png("../Figuras/diagrama_analisis_factorial.png")
fa.diagram(datos_vinos_af)
dev.off()

## png
## 2
```

De este diagrama podemos observar la representación grafica de las cargas en el espacio de los factores, donde vemos que la mayor carga de las variables densidad, alcohol, total.sulfur.dioxide, residual.sugar, free.sulfur.dioxide y chlorides esta sobre el RC1, donde vemos que es bastante alta(teniendo en cuenta el valor absoluto), para las primeras 4 mientras que las ultimas dos, tienen una carga menor. Por otra parte tenemos las variables fixed.acidity, pH y citric.acid, que tienen su mayor carga sobre el RC2 y finalmente encontramos sulphates y volatile.acidity que tienen carga sobre el RC3. No se encontraron variables que no estuvieran correlacionadas con ninguno de los factores identificados en el análisis factorial.

Análisis de conglomerados

Finalmente realizaremos un analisis de conglomerados, esto lo realizaremos usando el metodo kmeans de clustering, que viene integrado en R. Marcamos nstart igual a 1000, para que el proceso se inicialice 1000 veces diferentes y pueda seleccionar la mejor opción.

```
valores_unicos <- unique(datos_vino$quality)
print(valores_unicos)
```

```
## [1] 6 5 7 8 4 3 9
## Levels: 3 4 5 6 7 8 9
```

Procedemos a estandarizar los datos para aplicar el algoritmo de clustering.

```
datos_vino_estand <- scale(datos_vino[, -12])
datos_vino_km <- kmeans(datos_vino_estand[, -12], 7, nstart = 1e3)
```

```
## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations
```

```
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 244900)
```

```
## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations
```

```
datos_vino_km
```

```
## K-means clustering with 7 clusters of sizes 971, 728, 972, 817, 415,
895, 100
```

```

##
## Cluster means:
##   fixed.acidity volatile.acidity citric.acid residual.sugar
chlorides
## 1  -0.46881948      -0.5346002  -0.1369238   -0.5833974 -
0.10167434
## 2  -0.03725229      -0.3644059  -0.2484769    1.3435630
0.16250693
## 3   0.98330546      -0.3139700   0.3221581   -0.5463049 -
0.26670136
## 4   0.28158850       0.1549744   0.6978185    0.9014914
0.09434201
## 5  -0.15972597       1.5719662  -1.0262081   -0.2153006
0.21964186
## 6  -0.69159433       0.3126466  -0.2657250   -0.5430820 -
0.53652938
## 7  -0.18224133       0.3076444   0.9428911   -0.4173680
5.51619461
##   free.sulfur.dioxide total.sulfur.dioxide      density          pH
sulphates
## 1      0.004108624           0.01652227 -0.28273565  0.936854527
0.6479079
## 2      0.152910444           0.19903755  1.15972621 -0.292809709 -
0.3510744
## 3     -0.532090728          -0.52961887 -0.44315733 -0.691853830 -
0.2553059
## 4      1.171428021           1.30040749  1.02009251 -0.269470056
0.2986364
## 5     -0.574153736          -0.11457207  0.13282731  0.003340681 -
0.2233099
## 6     -0.390518660          -0.75344314 -1.15784538  0.285480073 -
0.2846310
## 7      0.326152189           0.13293176  0.08737267 -0.607723676 -
0.2194672
##           alcohol
## 1  0.01482373
## 2 -0.81704781
## 3  0.36370380
## 4 -0.84663661
## 5 -0.56713638
## 6  1.37607307
## 7 -0.77624824
##
## Clustering vector:
##   [1] 2 1 3 2 2 3 5 2 1 3 3 3 3 1 4 1 5 6 3 5 6 6 1 5 1 4 1 1 3 6 2 3
3 1 2 3 5
##  [38] 3 2 2 7 7 5 1 1 1 5 5 5 1 1 1 1 6 7 1 4 2 5 1 2 2 5 5 3 5 6 1 1
2 4 4 1 3
##  [75] 1 2 6 6 2 5 1 2 5 4 4 4 4 4 4 4 4 4 6 6 3 4 4 3 3 4 4 2 2 2 4 2
2 4 4 1 4

```

[112] 4 4 5 5 5 6 3 4 4 5 2 2 2 1 1 5 1 1 1 6 1 4 2 4 2 3 4 3 3 6 4 3
3 3 3 1 5
[149] 6 3 3 4 3 3 3 4 4 6 6 6 6 2 5 4 4 4 4 6 2 3 3 3 3 6 2 3 3 2 5 1
1 5 2 4 4
[186] 4 4 6 6 4 4 2 3 1 7 7 7 4 4 4 4 4 4 1 1 2 2 3 5 1 6 1 2 1 6 2 3
2 2 3 3 5
[223] 6 1 5 2 1 4 1 2 5 4 4 4 4 4 4 4 3 2 2 3 3 4 4 1 6 6 6 5 1 2 4 6
1 1 6 6 6
[260] 6 1 2 3 4 1 2 2 5 5 5 2 5 2 5 4 4 4 3 3 6 6 6 4 4 2 4 4 4 4 4 4
1 4 1 5 1
[297] 4 4 4 3 5 3 5 1 5 5 5 5 5 1 1 6 1 1 1 7 3 3 1 1 1 3 3 2 3 4 4 1
4 3 6 3 6
[334] 2 6 6 6 2 1 1 2 3 2 1 1 1 6 4 2 4 1 3 5 3 4 2 2 1 3 1 2 5 1 2 1
1 1 5 3 6
[371] 5 3 5 3 6 6 1 3 1 1 3 2 2 3 6 6 1 4 1 4 4 6 5 3 4 1 3 1 2 6 1 4
5 4 1 1 6
[408] 5 1 3 2 4 3 3 2 4 3 4 3 1 6 4 4 4 1 4 4 4 3 4 4 5 4 5 6 6 1 1 4
5 1 1 6 4
[445] 2 6 6 5 5 1 5 1 5 1 1 5 6 2 1 4 2 2 2 5 4 2 2 5 4 4 4 1 6 1 1 4
3 3 5 1 5
[482] 1 2 5 7 1 1 6 1 4 6 6 2 1 1 6 4 6 1 4 4 4 1 1 2 2 5 1 6 1 1 3 3
6 3 1 1 3
[519] 1 3 3 3 3 6 6 7 3 6 6 1 1 7 4 4 4 4 4 4 1 1 1 4 1 1 2 1 1 3 1 4
4 3 3 4 3
[556] 3 4 5 2 3 1 5 4 1 5 2 3 1 1 1 4 6 4 6 3 3 3 6 6 6 4 1 2 2 1 2 2
1 1 1 2 1
[593] 5 5 5 1 5 3 1 3 7 3 5 5 3 3 2 2 2 1 6 3 4 4 4 4 3 1 1 3 4 4 1 3
5 2 5 4 4
[630] 1 2 2 1 2 1 1 1 2 1 1 1 2 4 2 2 2 3 2 1 2 2 2 2 2 1 3 3 5 1 4 3
1 5 3 4 5
[667] 1 4 1 4 4 6 1 1 4 2 2 6 6 6 1 3 2 7 3 4 1 7 4 1 1 5 5 5 4 2 1 4
4 4 4 6 5
[704] 6 6 3 4 1 2 5 2 3 3 4 1 1 4 2 1 1 4 1 6 6 3 5 1 3 1 7 2 4 3 2 5
1 2 5 3 1
[741] 6 1 1 2 5 3 4 3 4 3 3 4 4 4 7 6 2 4 1 1 4 4 4 5 1 3 4 3 2 2 1 7
7 2 3 7 3
[778] 2 4 4 6 4 4 4 2 2 2 2 5 2 2 1 5 1 5 2 4 4 1 1 4 4 4 2 3 4 4 2 4
4 4 5 5 2
[815] 4 5 3 5 5 5 6 5 1 1 3 1 1 6 3 3 1 3 6 6 1 6 3 3 3 3 3 2 2 1 1 1
2 3 2 6 1
[852] 4 1 4 1 1 4 4 4 7 3 4 1 6 1 6 1 4 4 1 4 4 6 3 3 1 6 7 7 1 1 3 3
3 4 6 1 1
[889] 1 1 1 6 1 6 6 1 4 3 3 4 1 2 4 4 2 3 3 6 2 4 5 5 5 5 3 5 1 3 3 3
1 4 4 4 1
[926] 5 5 4 2 4 3 4 4 4 4 4 6 4 4 2 4 4 5 6 2 1 4 3 5 3 5 2 1 3 2 5 3
1 1 2 2 3
[963] 2 1 2 3 5 2 1 6 6 6 2 1 1 5 1 4 1 7 7 1 6 6 1 5 3 3 1 3 3 3 4 1
3 3 1 2 2
[1000] 3 3 4 4 1 1 1 1 5 1 3 3 1 3 4 2 5 4 3 3 1 3 3 1 2 1 1 1 5 1 4 1
4 4 4 7 3

[1037] 1 5 1 3 5 3 5 4 3 3 1 3 3 6 2 7 3 3 3 3 1 5 1 1 3 4 1 1 1 2 4 3
3 1 3 2 3
[1074] 3 5 6 4 3 4 1 1 2 2 2 1 4 1 3 2 3 4 2 1 3 2 6 2 1 2 6 2 6 1 3 4
1 1 6 6 3
[1111] 6 3 2 3 5 3 5 4 5 5 2 5 6 3 3 4 6 6 1 3 1 1 2 3 5 2 1 1 3 3 3 3
3 3 3 4 3
[1148] 3 3 4 4 4 5 6 1 1 2 2 7 2 1 2 1 7 3 1 5 3 3 1 1 5 3 1 4 4 2 1 3
3 5 3 1 1
[1185] 6 4 3 6 3 3 3 1 2 2 4 4 1 2 2 3 1 3 2 6 6 3 4 2 1 3 3 2 3 6 3 4
3 7 3 3 1
[1222] 6 3 4 6 6 6 3 3 3 4 3 3 5 1 3 6 3 6 3 1 1 3 6 5 4 3 3 3 4 1 1 1
6 7 1 1 4
[1259] 4 4 4 1 1 4 1 1 3 4 2 4 4 3 7 3 4 1 4 2 1 1 3 3 4 3 3 3 3 3 6 6
3 4 1 5 5
[1296] 4 5 6 6 3 3 3 4 4 4 3 6 3 3 3 5 1 3 3 4 4 3 1 1 4 3 6 1 1 1 3 4
1 6 3 3 2
[1333] 4 1 3 1 3 4 4 5 3 5 5 2 6 6 6 3 3 3 6 6 1 4 1 3 4 4 6 1 1 1 6 5
3 1 3 5 4
[1370] 7 3 3 4 4 6 6 3 1 1 5 6 6 5 3 1 1 6 6 3 3 1 6 6 1 6 6 3 2 2 3 2
4 5 1 3 5
[1407] 3 3 2 3 6 3 1 3 1 5 1 4 3 3 3 6 1 4 3 3 3 3 3 1 5 1 6 3 3 2 4 3
4 4 4 3 3
[1444] 3 4 3 6 3 3 3 4 3 4 4 4 4 6 3 3 3 3 1 3 3 3 1 3 3 3 3 4 4 3 3 3
1 5 5 3 3
[1481] 1 1 1 3 5 3 1 4 4 4 1 3 4 3 3 3 5 6 3 6 3 3 5 6 3 3 4 4 4 3 3 3
3 3 3 1 3
[1518] 4 3 3 3 5 1 4 2 4 3 4 3 2 4 3 4 3 4 3 3 3 1 3 4 5 3 1 3 3 3 3 6
3 3 7 6 1
[1555] 4 1 4 3 5 3 3 3 6 3 3 3 2 3 4 4 4 4 1 3 3 1 1 5 4 4 2 3 3 4 4 4
3 4 4 3 3
[1592] 1 3 3 3 4 5 4 7 7 3 4 6 6 5 6 3 1 4 4 3 3 3 3 3 1 4 1 4 1 3 3 5
1 3 1 4 3
[1629] 3 3 1 3 1 3 4 3 5 3 7 3 5 1 2 4 2 4 1 3 1 1 3 4 3 2 5 3 1 4 4 4
4 4 3 2 4
[1666] 1 3 3 3 3 2 6 7 7 4 4 6 3 1 5 4 1 2 2 2 2 1 2 4 1 5 2 1 2 1 3 4
1 1 1 1 4
[1703] 5 2 2 1 1 2 5 2 3 1 3 5 7 1 1 3 3 1 1 1 4 1 1 2 6 4 7 1 3 4 5 3
2 3 5 2 3
[1740] 3 5 6 2 4 7 4 3 4 3 1 2 1 4 1 1 2 4 4 3 4 1 3 1 2 4 4 4 1 5 6 5
1 5 5 3 4
[1777] 2 1 1 1 3 1 3 5 1 3 3 6 3 3 3 5 3 4 5 5 6 6 2 1 4 3 2 1 4 2 5 4
2 4 5 5 6
[1814] 3 3 3 1 5 6 6 3 3 3 1 2 1 2 1 1 5 5 5 5 5 1 7 7 1 4 7 1 1 4 1 6
5 4 5 4 1
[1851] 1 6 1 3 1 4 5 3 3 4 1 4 1 4 1 7 2 2 3 3 1 1 2 2 3 2 3 3 4 2 4 2
4 4 1 2 5
[1888] 2 2 1 4 4 2 2 4 4 4 4 1 6 2 3 2 3 4 4 1 3 3 2 4 3 3 3 6 3 4 3 4
4 4 2 1 3
[1925] 3 7 7 3 1 4 4 4 5 4 1 4 4 1 3 2 4 2 4 4 4 3 2 4 1 2 2 5 3 1 2 4
3 2 4 3 3

[1962] 3 3 4 4 2 3 4 1 1 3 1 2 2 2 4 1 2 3 4 4 4 4 4 4 3 4 4 6 5 4 1
3 4 4 4 4
[1999] 4 2 5 5 1 3 1 4 4 1 1 3 3 1 1 2 3 3 6 3 3 2 1 1 4 1 7 7 7 2 3 4
3 3 6 2 3
[2036] 3 1 3 1 3 4 1 1 3 5 3 2 2 2 1 3 2 2 3 3 3 2 1 4 5 4 6 3 1 1 3 3
3 6 6 5 1
[2073] 1 1 1 1 1 1 1 5 1 5 2 3 1 1 3 3 3 1 2 2 5 2 2 2 5 2 4 1 2 6 4 5
1 5 4 2 4
[2110] 4 2 2 2 1 4 2 6 3 6 3 2 6 2 2 6 2 5 4 5 1 6 2 2 3 1 1 3 1 6 2 2
2 4 1 2 2
[2147] 1 2 2 6 6 2 3 3 5 2 3 3 1 6 1 6 3 3 2 5 5 4 4 4 4 4 5 4 4 4 5 5
4 3 2 3 2
[2184] 5 5 2 7 5 2 2 6 2 4 2 2 5 6 3 4 4 2 1 4 3 1 4 3 1 3 6 1 1 6 1 3
2 1 3 1 6
[2221] 3 2 5 2 5 2 4 4 2 2 2 2 3 5 1 5 2 2 1 5 2 2 4 1 2 3 2 1 3 2 4 1
2 2 1 1 3
[2258] 7 4 7 4 5 3 2 1 2 3 1 1 4 1 2 1 5 1 2 4 1 4 7 1 1 4 4 4 4 7 7 4
6 1 3 5 6
[2295] 6 5 1 4 6 6 6 2 2 5 5 6 1 2 3 3 2 6 3 6 6 4 3 4 4 6 3 5 5 1 4 5
2 5 2 3 4
[2332] 3 4 6 4 2 4 4 1 1 1 1 6 1 1 4 4 3 3 7 4 4 5 1 1 6 1 3 4 7 2 3 1
2 1 5 4 4
[2369] 4 1 1 4 6 3 1 6 1 2 4 1 1 1 3 6 3 4 1 1 3 6 1 1 1 1 4 1 4 6 3 1
3 3 4 3 3
[2406] 2 6 1 5 1 1 4 4 1 5 3 1 5 5 4 3 4 5 1 5 1 1 4 1 4 4 2 4 4 4 5 5
3 5 1 7 1
[2443] 2 1 2 1 1 1 1 1 5 5 1 1 4 4 3 3 4 2 2 3 3 1 5 4 4 1 1 5 3 1 3 2
6 5 7 6 4
[2480] 2 1 4 4 4 2 4 2 3 3 2 1 2 2 2 2 1 3 2 2 1 2 4 3 3 4 4 4 6 2 4 4
4 1 3 6 3
[2517] 1 4 3 4 2 1 7 3 2 7 1 1 1 5 1 5 5 2 3 3 2 5 5 1 4 3 3 2 6 1 5 5
4 4 5 4 2
[2554] 1 1 1 4 1 1 1 3 2 3 1 5 4 6 1 4 1 1 6 1 5 1 4 4 4 6 7 2 2 2 2 2
2 2 1 2 5
[2591] 2 1 1 2 6 5 1 1 1 4 1 4 1 1 1 6 1 4 4 6 4 4 1 2 6 2 2 2 6 5 2 2
6 2 6 1 1
[2628] 1 1 5 6 6 4 2 4 3 2 4 3 3 3 2 1 5 7 1 1 1 2 7 4 5 6 2 4 1 4 1 4
6 2 1 3 3
[2665] 6 5 6 1 5 4 6 6 2 3 3 3 3 6 5 2 3 2 3 5 1 3 3 4 2 2 5 6 3 1 6 6
2 3 3 5 6
[2702] 2 2 2 7 7 4 2 4 4 4 1 4 4 4 4 4 2 6 3 3 4 6 1 3 2 1 3 4 6 2 5 5
6 2 4 3 2
[2739] 5 6 2 5 1 3 3 6 3 2 1 2 1 1 6 6 6 4 4 1 1 1 1 1 3 3 2 6 3 2 3 3
4 1 6 3 3
[2776] 3 3 3 3 6 6 2 1 1 1 2 3 2 4 4 1 4 4 6 4 3 6 4 6 6 4 4 6 6 3 1 4
4 4 1 3 3
[2813] 1 1 6 6 3 6 2 1 7 4 2 3 4 1 4 1 1 2 2 5 6 5 1 1 4 2 3 6 6 6 1 1
3 3 3 6 5
[2850] 7 2 3 3 1 6 6 1 6 6 1 1 6 6 6 1 1 1 6 6 6 6 3 1 1 1 3 2 4 6 6 3
3 2 6 6 6

[2887] 5 6 3 3 6 3 4 1 6 5 4 4 3 2 1 6 5 6 1 5 4 6 6 6 5 5 6 3 3 3 6 6
6 2 3 1 6
[2924] 3 2 4 1 3 3 2 1 1 4 3 6 3 6 4 2 2 3 3 4 2 3 6 6 6 4 2 5 1 5 5 6
3 3 6 1 6
[2961] 2 1 2 6 1 6 2 2 3 3 3 3 1 2 3 3 6 2 2 6 6 3 1 6 6 2 2 3 3 1 6 6
2 3 3 6 6
[2998] 3 6 3 2 2 6 3 6 6 1 6 6 4 1 6 6 2 2 3 2 6 3 3 3 1 5 2 2 1 3 2 3
5 5 2 1 4
[3035] 4 6 1 4 3 4 2 4 4 7 2 1 6 3 3 7 4 7 4 3 6 6 6 1 4 3 4 4 6 4 4 2
4 3 6 6 6
[3072] 6 4 6 1 6 5 4 3 1 3 4 6 6 1 3 6 2 6 6 1 4 3 3 3 3 1 5 6 6 3 3 5
3 3 3 3 3
[3109] 4 1 3 6 3 2 1 1 1 6 6 6 4 3 6 6 2 2 3 3 1 6 6 4 1 6 1 4 1 6 3 2
4 6 3 1 2
[3146] 2 2 4 2 4 6 2 4 6 6 6 2 6 6 1 6 2 6 6 3 6 6 6 6 1 3 3 6 6 6 2 3
5 3 3 6 3
[3183] 1 6 6 1 3 6 3 3 6 3 3 1 6 3 3 3 3 6 6 3 6 3 2 6 6 1 1 2 5 3 6 3
6 3 3 5 7
[3220] 6 3 6 3 3 6 6 1 4 4 1 4 3 6 6 1 6 6 1 3 6 1 6 1 6 6 3 3 3 6 1 3
4 6 6 3 2
[3257] 2 2 2 2 2 6 2 6 1 1 3 1 6 4 3 6 3 6 1 5 6 1 4 3 3 6 1 7 3 1 1 1
7 4 3 6 4
[3294] 3 1 4 4 4 6 6 3 6 6 6 6 6 6 4 5 6 6 1 6 3 4 6 6 6 1 1 2 6 6 3 3
6 1 1 6 3
[3331] 4 2 6 1 6 4 4 4 2 3 6 3 6 4 4 4 4 1 2 1 3 6 6 6 1 6 2 6 6 6 3 1
3 6 6 3 6
[3368] 6 1 3 6 6 1 6 1 2 1 5 6 4 6 3 6 1 1 6 6 4 3 6 6 6 6 3 4 4 2 1 4
6 5 3 5 3
[3405] 6 1 6 4 4 1 3 1 2 4 3 6 4 6 4 6 2 6 6 4 6 4 4 1 6 4 4 4 3 6 3 6
1 5 4 5 1
[3442] 3 6 6 4 1 2 2 6 1 3 3 3 6 6 3 5 1 6 2 2 4 6 6 4 6 6 4 6 2 4 2 1
6 1 4 6 3
[3479] 4 6 6 1 6 6 6 3 3 1 3 3 3 3 6 6 1 6 1 4 1 6 6 1 4 6 1 1 1 6 4 4
4 6 1 3 6
[3516] 6 6 6 3 6 4 1 4 4 6 6 3 2 5 6 4 4 3 2 4 4 3 7 6 6 3 3 2 1 1 4 2
4 6 1 3 2
[3553] 2 2 3 1 1 3 2 3 6 5 3 5 6 3 3 6 3 3 6 5 3 3 3 3 2 3 3 3 3 6 6 3
3 6 1 2 2
[3590] 2 1 1 1 1 3 3 6 2 1 4 2 1 1 6 6 4 5 1 2 2 4 1 1 2 5 3 4 6 2 2 4
2 3 2 6 4
[3627] 6 7 7 7 3 3 2 3 3 6 6 3 6 3 1 1 1 1 6 1 6 6 3 6 3 2 1 2 1 6 3 2
1 6 6 1 5
[3664] 6 6 6 2 6 3 6 2 6 6 6 5 1 6 6 2 3 4 6 1 4 6 4 3 2 1 6 6 3 3 6 2
1 1 4 6 3
[3701] 5 2 2 2 2 2 3 6 4 5 6 2 2 2 3 1 3 1 2 4 6 5 5 1 1 6 6 1 6 3 2 2
3 6 2 7 6
[3738] 7 6 2 2 2 2 2 2 2 3 4 2 1 2 1 1 1 6 3 1 1 1 3 3 6 3 6 6 4 4 4 2
4 3 2 2 6
[3775] 4 6 6 6 1 6 4 3 3 4 2 6 4 4 4 4 4 4 2 4 1 2 2 5 3 1 3 3 3 3 3 3
5 3 3 3 2

[3812] 3 3 1 2 6 5 6 6 6 6 2 1 4 3 6 6 6 6 6 5 5 2 5 5 3 3 5 5 6 4 3
6 6 2 6 2

[3849] 7 3 6 6 6 1 1 1 6 1 6 4 4 4 4 4 6 6 3 4 4 4 2 4 4 4 4 1 2 3 4 5
2 2 3 6 2

[3886] 6 6 6 2 2 6 3 1 5 6 6 1 4 3 1 5 6 7 6 6 6 6 6 6 6 6 7 6 3 6 6 6
4 6 6 3 4

[3923] 6 3 3 3 2 6 2 2 6 6 1 3 3 4 1 7 3 1 5 1 3 1 1 1 6 6 1 5 1 2 3 3
2 3 1 3 1

[3960] 1 4 4 6 6 4 5 1 4 4 3 2 3 7 4 4 1 6 6 1 1 2 1 1 1 6 2 2 4 1 2 4
3 3 1 1 3

[3997] 1 6 1 1 1 3 2 1 6 1 6 3 2 1 3 6 1 4 4 5 5 6 4 5 5 3 5 4 1 6 4 1
6 6 3 3 2

[4034] 3 2 1 5 5 3 5 1 2 2 2 2 2 2 2 2 3 2 3 2 2 2 3 5 3 2 3 5 5 3 1 6
6 5 5 5 2

[4071] 1 3 4 5 5 2 5 1 1 6 2 2 1 1 2 3 3 3 3 5 2 1 5 5 6 5 3 1 1 5 2 2
2 6 6 3 1

[4108] 2 6 1 5 6 1 1 3 3 2 2 6 6 2 2 2 6 6 1 1 4 4 6 3 4 4 6 6 1 6 4 6
1 2 2 1 2

[4145] 2 4 2 1 2 6 5 3 2 1 4 1 4 2 2 2 2 2 2 3 3 6 6 6 6 6 2 6 1 7 4 1
1 1 6 4 2

[4182] 3 4 6 6 4 3 6 6 5 6 3 3 3 3 6 6 1 4 3 3 6 6 1 2 6 2 1 6 6 1 4 5
7 4 4 4 5

[4219] 3 4 6 6 5 6 6 1 4 1 2 3 3 6 6 1 1 2 2 3 2 1 2 1 1 6 2 2 6 7 6 1
2 1 2 6 2

[4256] 5 5 6 6 6 6 6 6 5 6 6 6 6 6 4 4 4 2 2 2 2 6 4 6 2 2 2 2 6 6 6 6
3 3 2 2 4

[4293] 3 4 3 4 3 3 6 7 7 5 4 6 1 1 2 1 6 6 2 6 6 6 6 6 5 6 6 6 4 2 3 3
6 2 4 2 5

[4330] 2 2 4 2 2 2 2 2 2 2 1 2 4 1 1 7 3 7 3 4 4 6 6 3 3 1 6 1 1 2 4 3
1 6 2 2 2

[4367] 6 2 2 2 5 1 3 6 5 1 5 1 5 5 1 6 1 1 2 2 1 1 2 6 2 4 2 2 2 2 4 4
4 2 4 1 6

[4404] 1 2 2 6 3 2 6 4 6 3 6 3 2 4 1 2 2 2 2 1 1 2 1 1 2 6 1 4 6 6 5 1
1 6 1 1 5

[4441] 2 6 2 5 6 6 6 6 3 2 2 2 5 6 2 2 2 2 6 1 2 1 6 6 2 2 1 1 1 1 1 6
3 7 2 6 6

[4478] 4 4 5 2 4 6 1 5 6 3 1 6 6 6 6 6 2 3 4 3 7 3 6 3 5 3 6 5 1 1 6 6
5 6 6 6 6

[4515] 5 3 6 3 3 2 4 5 3 2 2 2 2 3 5 4 4 4 6 2 6 4 4 3 3 3 6 6 3 3 6 6
6 6 4 3 3

[4552] 3 6 6 6 5 1 4 1 1 6 6 1 1 3 1 3 1 6 1 6 6 6 6 6 6 6 2 1 2 2 1 1
4 6 1 6 3

[4589] 1 6 1 4 6 5 6 5 6 5 5 6 3 3 2 6 5 6 3 2 6 5 6 6 5 1 5 6 3 6 6 5
6 3 6 6 6

[4626] 5 4 6 6 6 2 1 4 1 1 5 5 6 5 4 6 3 6 1 6 6 1 6 5 5 5 6 4 6 4 4 4
4 6 3 6 3

[4663] 3 1 1 3 4 1 1 6 4 1 1 6 1 1 2 6 4 1 5 6 6 6 1 2 5 2 2 2 2 2 1 2
2 6 1 6 7

[4700] 4 4 5 5 1 4 5 6 2 1 6 6 2 5 6 6 3 6 1 1 3 6 1 2 4 1 2 6 2 2 3 3
6 6 1 6 6

```

## [4737] 6 6 6 6 1 1 1 1 1 4 4 1 2 2 1 4 6 1 2 3 6 6 3 2 2 2 6 6 6 6 6 1
4 4 4 4 4
## [4774] 6 3 1 1 6 2 5 5 5 2 2 4 3 6 6 4 6 6 2 6 7 7 1 1 6 2 2 6 1 6 6 6
1 2 3 3 1
## [4811] 1 1 6 7 6 5 2 6 6 6 7 3 3 6 2 1 1 1 2 6 6 5 6 1 6 1 5 6 1 6 6 1
6 6 6 7 6
## [4848] 6 3 4 3 4 1 6 6 2 2 3 4 5 5 6 6 6 5 6 6 6 1 2 6 6 4 6 1 3 6 5 5
4 4 2 6 6
## [4885] 4 4 1 6 3 2 6 1 1 6 4 3 6 6
##
## Within cluster sum of squares by cluster:
## [1] 5886.856 3990.819 5737.880 5804.424 3154.632 4761.188 1210.372
## (between_SS / total_SS = 43.3 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
"tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"

datos_vino_km$tot.withinss

## [1] 30546.17

datos_vino_km$withinss

## [1] 5886.856 3990.819 5737.880 5804.424 3154.632 4761.188 1210.372

sum(datos_vino_km$withinss)

## [1] 30546.17

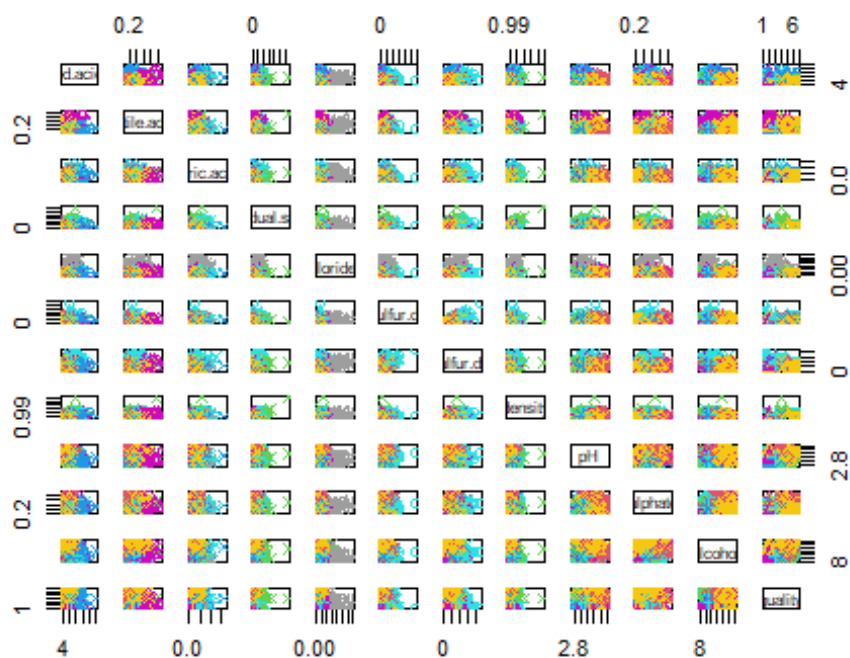
```

Cruzamos estos resultados con nuestra variable Factor(quality) para comparar los resultados.

```

plot(datos_vino,
     col = datos_vino_km$cluster + 1,
     pch = as.numeric(as.factor(datos_vino$quality)))

```



```
png("../Figuras/diagrama_kmeans.png", width = 10000, height = 10000, res
= 300)
plot(datos_vino,
     col = datos_vino_km$cluster + 1,
     pch = as.numeric(as.factor(datos_vino$quality)))
dev.off()

## png
## 2

table(datos_vino_km$cluster, datos_vino$quality)

##
##      3    4    5    6    7    8    9
## 1    2   19  190  502  220   38    0
## 2    3    9  278  341   82   15    0
## 3    4   44  272  459  163   29    1
## 4    5   19  373  361   48   11    0
## 5    2   54  240  101   15    3    0
## 6    3   15   58  388  350   77    4
## 7    1    3   46   46    2    2    0
```

De esta matriz podemos observar que:

1. Los clusters 2, 5, 6 y 7 tienen la mayor cantidad de vinos, especialmente de calidad 5 y 6. Esto podría indicar que estos clusters representan vinos de calidad media.

2. El cluster 4, por otro lado, tiene la menor cantidad de vinos en general y parece tener una mayor concentración de vinos de calidad 5 y 6, pero en menor cantidad que los otros clusters.
3. Los vinos de calidad 3 y 9 son muy raros en todos los clusters, lo que indica que estos vinos de calidad extremadamente baja o alta son menos comunes en el conjunto de datos.
4. Los clusters 1, 3 y 7 parecen tener una distribución más equilibrada de calidades de vino, con una ligera predominancia de vinos de calidad 5 o 6.

En conclusión, los clusters generados por el algoritmo k-means parecen agrupar los vinos principalmente en torno a la calidad media (5 y 6), con algunos clusters que tienen una mayor concentración de estas calidades que otros. Los vinos de calidad extremadamente alta o baja son menos comunes y están distribuidos de manera bastante uniforme entre los clusters. Como podemos ver, observamos que los clusters creados por el algoritmo, no se corresponden con los reales, puesto a que los vinos de cada clase se encuentran demasiado dispersos entre todos los clusters creados.