

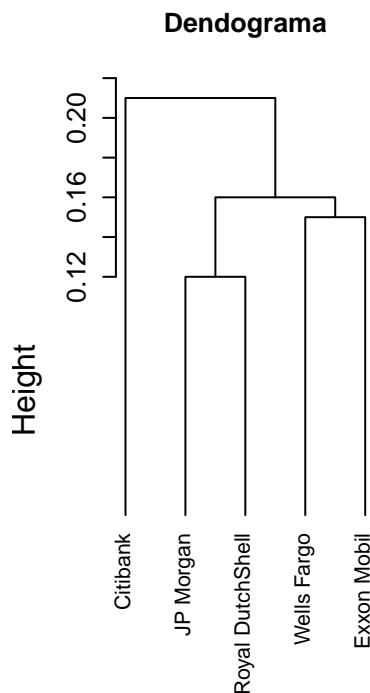
Prova 3 Multivariada

Davi Wentrück Feijó

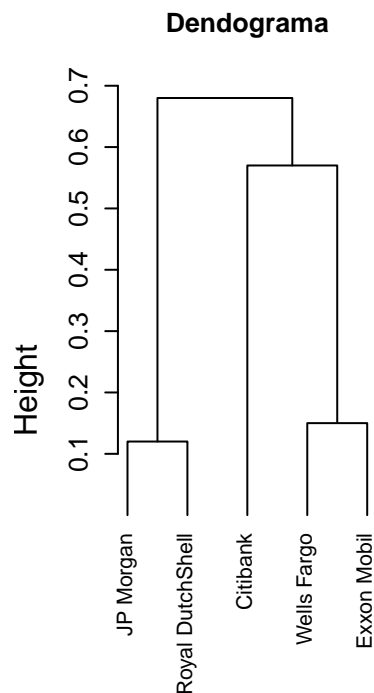
2023-07-18

79. Johnson e Wichern - Exercício 12.7

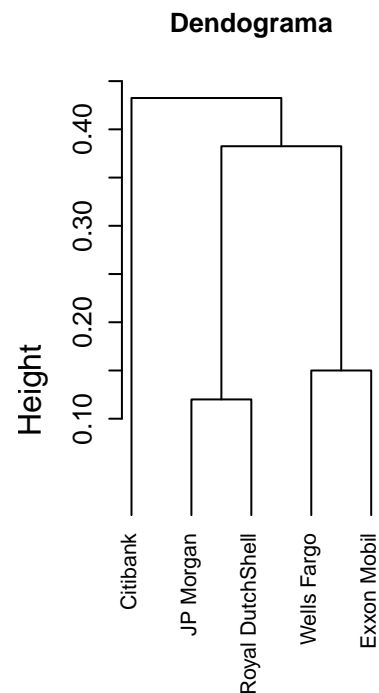
```
##                JP Morgan Citibank Wells Fargo Royal DutchShell
## Citibank                0.63
## Wells Fargo            0.51      0.57
## Royal DutchShell       0.12      0.32      0.18
## Exxon Mobil            0.16      0.21      0.15      0.68
```



Single
hclust (*, "single")



Complete
hclust (*, "complete")

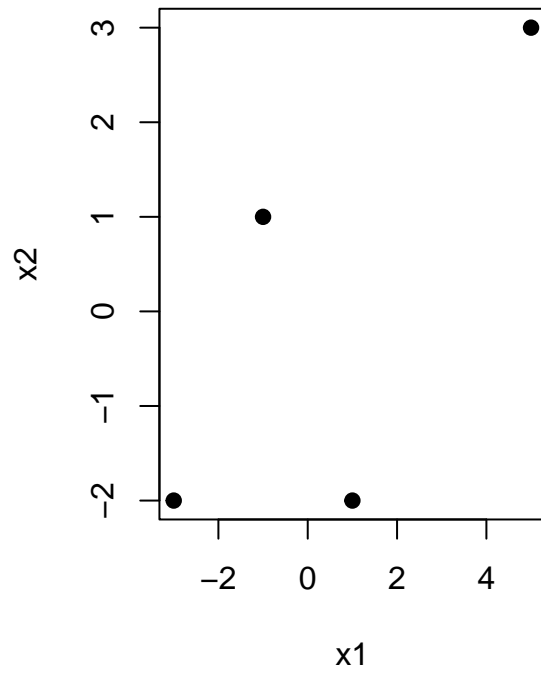


Average
hclust (*, "average")

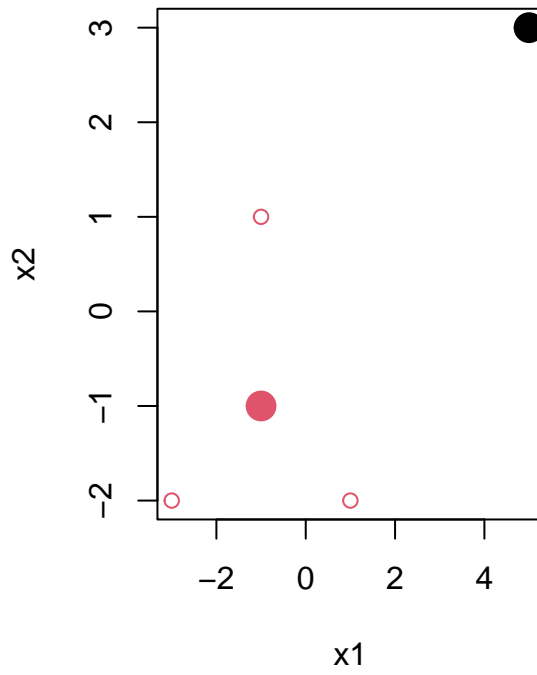
Podemos observar que os agrupamentos por ligação simples e pela média tem um comportamento parecido, se diferenciando na distância para o corte. Já o por ligação completa agrupa os elementos de forma diferente das do demais.

81. Johnson e Wichern - Exercício 12.12

Grafico de dispersao



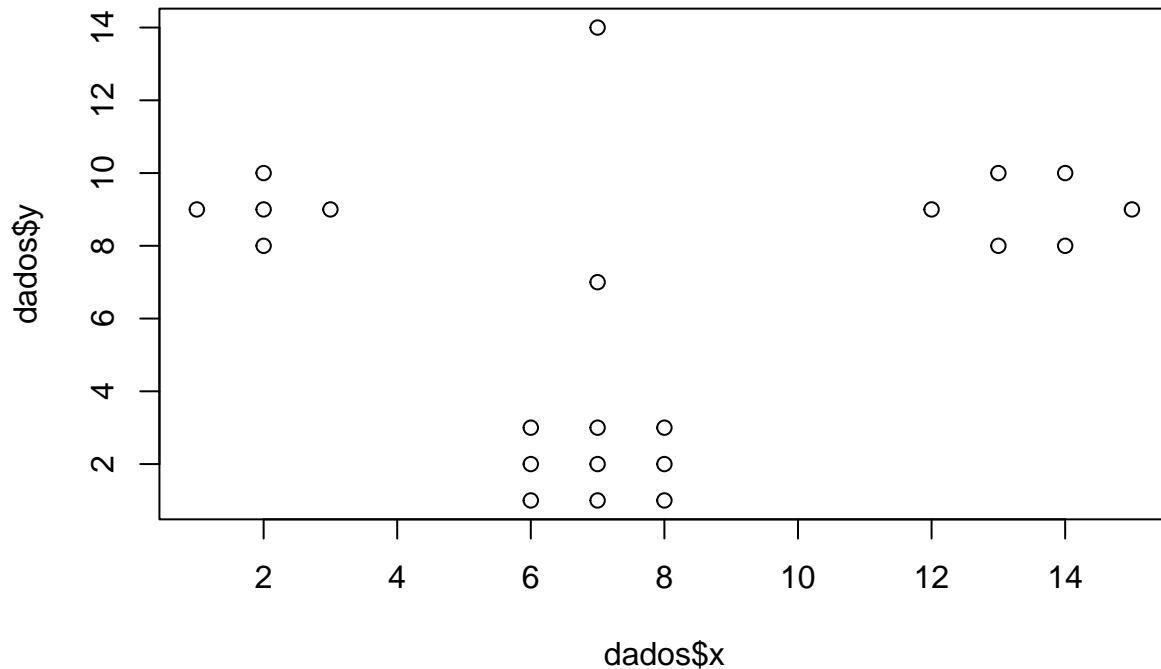
K-means



83. Considere o seguinte conjunto de 22 pontos (x, y).

##	ponto	x	y
## 1	1	1	9
## 2	2	2	10
## 3	3	2	9
## 4	4	2	8
## 5	5	3	9
## 6	6	7	14
## 7	7	12	9
## 8	8	13	10
## 9	9	13	8
## 10	10	14	10
## 11	11	14	8
## 12	12	15	9
## 13	13	7	7
## 14	14	6	3
## 15	15	7	3
## 16	16	8	3
## 17	17	6	2
## 18	18	7	2
## 19	19	8	2
## 20	20	6	1
## 21	21	7	1
## 22	22	8	1

(a) Plote o gráfico de dispersão para o conjunto de pontos. Quantos e quais grupos você indicaria através do gráfico.



Ao observar pelo gráfico é indicado claramente 3 grandes grupos e mais um ou dois grupos de uma única observação que seriam aqueles no meio.

(b) Construa a matriz de distâncias D , calculando as distâncias Euclidiana, de Manhattan e Mahalanobis. Compare os resultados. Explique a razão de possíveis diferenças nos resultados para as três distâncias.

Distancia Euclidiana

##	1	2	3	4	5	6	7
## 2	1.414214						
## 3	1.000000	1.000000					
## 4	1.414214	2.000000	1.000000				
## 5	2.000000	1.414214	1.000000	1.414214			
## 6	7.810250	6.403124	7.071068	7.810250	6.403124		
## 7	11.000000	10.049876	10.000000	10.049876	9.000000	7.071068	
## 8	12.041595	11.000000	11.045361	11.180340	10.049876	7.211103	1.414214
## 9	12.041595	11.180340	11.045361	11.000000	10.049876	8.485281	1.414214
## 10	13.038405	12.000000	12.041595	12.165525	11.045361	8.062258	2.236068
## 11	13.038405	12.165525	12.041595	12.000000	11.045361	9.219544	2.236068
## 12	14.000000	13.038405	13.000000	13.038405	12.000000	9.433981	3.000000
## 13	6.324555	5.830952	5.385165	5.099020	4.472136	7.000000	5.385165
## 14	7.810250	8.062258	7.211103	6.403124	6.708204	11.045361	8.485281

## 15	8.485281	8.602325	7.810250	7.071068	7.211103	11.000000	7.810250
## 16	9.219544	9.219544	8.485281	7.810250	7.810250	11.045361	7.211103
## 17	8.602325	8.944272	8.062258	7.211103	7.615773	12.041595	9.219544
## 18	9.219544	9.433981	8.602325	7.810250	8.062258	12.000000	8.602325
## 19	9.899495	10.000000	9.219544	8.485281	8.602325	12.041595	8.062258
## 20	9.433981	9.848858	8.944272	8.062258	8.544004	13.038405	10.000000
## 21	10.000000	10.295630	9.433981	8.602325	8.944272	13.000000	9.433981
## 22	10.630146	10.816654	10.000000	9.219544	9.433981	13.038405	8.944272
##	8	9	10	11	12	13	14
## 2							
## 3							
## 4							
## 5							
## 6							
## 7							
## 8							
## 9	2.000000						
## 10	1.000000	2.236068					
## 11	2.236068	1.000000	2.000000				
## 12	2.236068	2.236068	1.414214	1.414214			
## 13	6.708204	6.082763	7.615773	7.071068	8.246211		
## 14	9.899495	8.602325	10.630146	9.433981	10.816654	4.123106	
## 15	9.219544	7.810250	9.899495	8.602325	10.000000	4.000000	1.000000
## 16	8.602325	7.071068	9.219544	7.810250	9.219544	4.123106	2.000000
## 17	10.630146	9.219544	11.313708	10.000000	11.401754	5.099020	1.000000
## 18	10.000000	8.485281	10.630146	9.219544	10.630146	5.000000	1.414214
## 19	9.433981	7.810250	10.000000	8.485281	9.899495	5.099020	2.236068
## 20	11.401754	9.899495	12.041595	10.630146	12.041595	6.082763	2.000000
## 21	10.816654	9.219544	11.401754	9.899495	11.313708	6.000000	2.236068
## 22	10.295630	8.602325	10.816654	9.219544	10.630146	6.082763	2.828427
##	15	16	17	18	19	20	21
## 2							
## 3							
## 4							
## 5							
## 6							
## 7							
## 8							
## 9							
## 10							
## 11							
## 12							
## 13							
## 14							
## 15							
## 16	1.000000						
## 17	1.414214	2.236068					
## 18	1.000000	1.414214	1.000000				
## 19	1.414214	1.000000	2.000000	1.000000			
## 20	2.236068	2.828427	1.000000	1.414214	2.236068		
## 21	2.000000	2.236068	1.414214	1.000000	1.414214	1.000000	
## 22	2.236068	2.000000	2.236068	1.414214	1.000000	2.000000	1.000000

Distancia de Manhattan

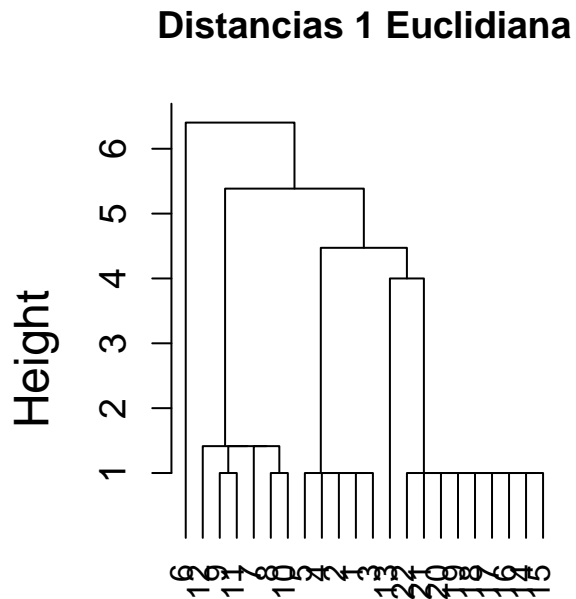
```

##      1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21
## 2    2
## 3    1  1
## 4    2  2  1
## 5    2  2  1  2
## 6   11  9 10 11  9
## 7   11 11 10 11  9 10
## 8   13 11 12 13 11 10  2
## 9   13 13 12 11 11 12  2  2
## 10  14 12 13 14 12 11  3  1  3
## 11  14 14 13 12 12 13  3  3  1  2
## 12  14 14 13 14 12 13  3  3  3  2  2
## 13   8  8  7  6  6  7  7  9  7 10  8 10
## 14  11 11 10  9  9 12 12 14 12 15 13 15  5
## 15  12 12 11 10 10 11 11 13 11 14 12 14  4  1
## 16  13 13 12 11 11 12 10 12 10 13 11 13  5  2  1
## 17  12 12 11 10 10 13 13 15 13 16 14 16  6  1  2  3
## 18  13 13 12 11 11 12 12 14 12 15 13 15  5  2  1  2  1
## 19  14 14 13 12 12 13 11 13 11 14 12 14  6  3  2  1  2  1
## 20  13 13 12 11 11 14 14 16 14 17 15 17  7  2  3  4  1  2  3
## 21  14 14 13 12 12 13 13 15 13 16 14 16  6  3  2  3  2  1  2  1
## 22  15 15 14 13 13 14 12 14 12 15 13 15  7  4  3  2  3  2  1  2  1

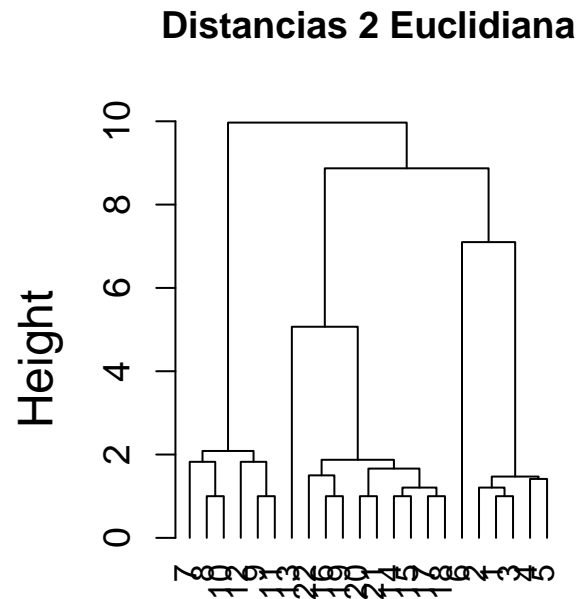
```

O que pode causar diferenca na forma de agrupamento por essas distancias é que uma usa a distancia euclidiana que tem uma formula diferente da de manhattan que usa valores absolutos, ou seja, é esperado agrupamentos diferentes.

(c) Utilize os algoritmos de ligação simples e média, e obtenha os respectivos dendogramas. Em cada caso indique sua escolha para o número de grupos e liste os elementos de cada grupo. Os dendogramas obtidos são únicos? Justifique.

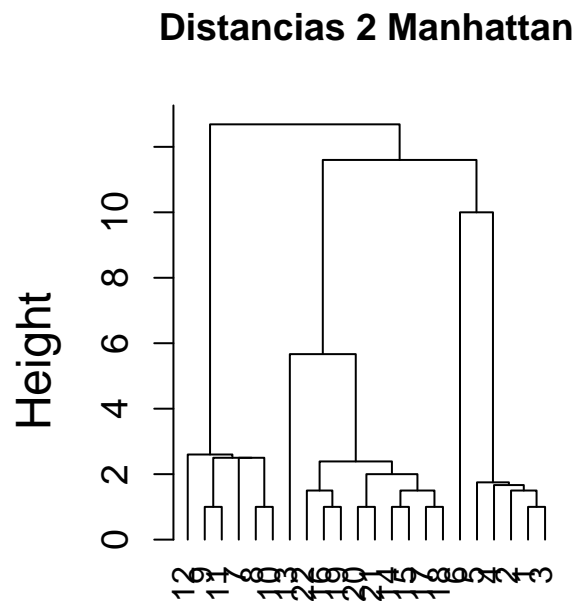
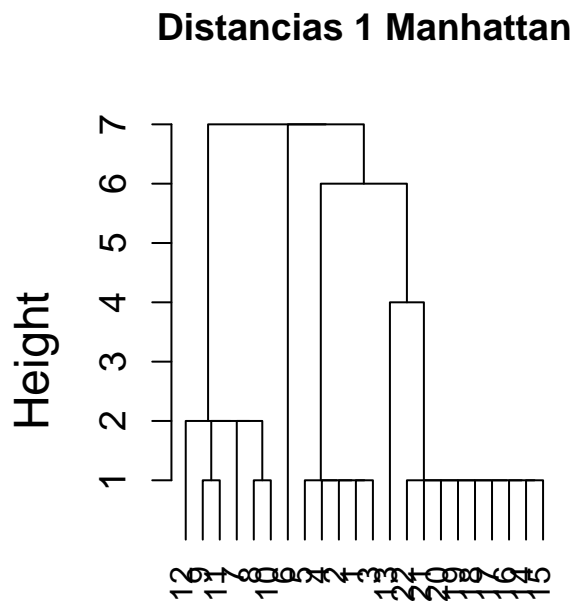


Single
hclust (*, "single")



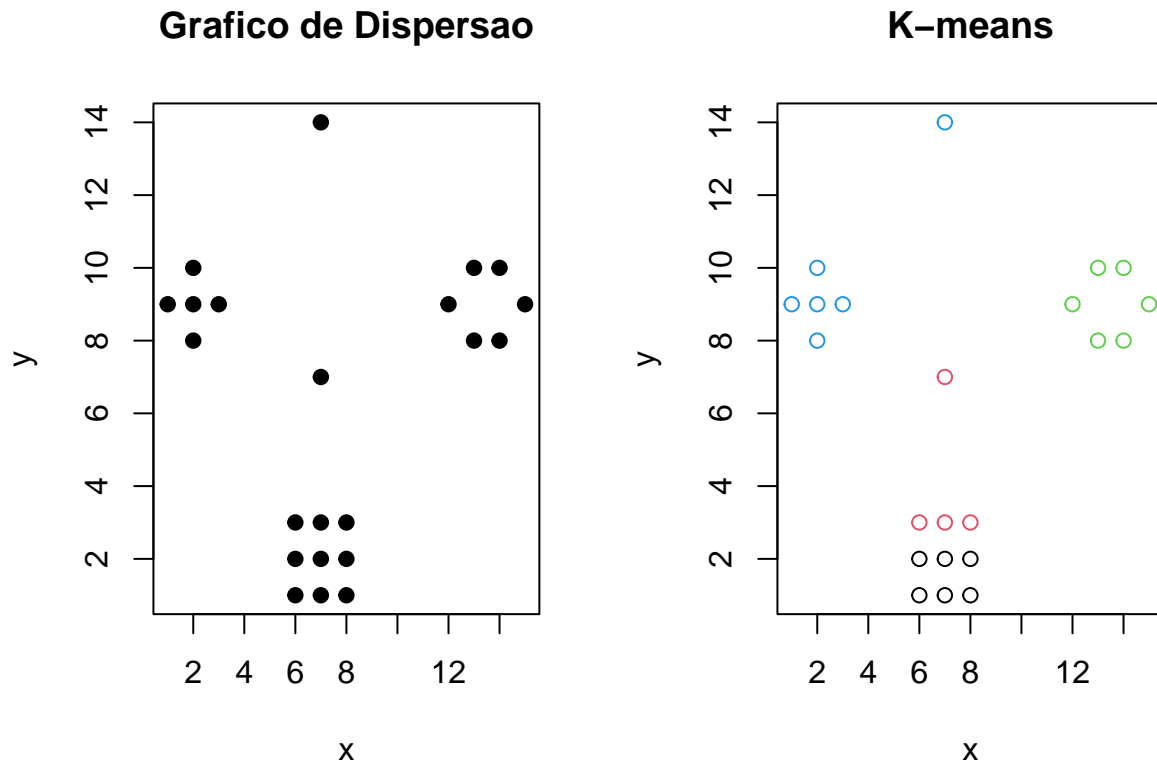
Average
hclust (*, "average")

Assim como discutido antes, as diferentes medidas de distancia trazem consigo agrupamentos diferentes.



O mesmo pode ser observado aqui

(d) Descreva e aplique o algoritmo de k-means para agrupar estes dados. Apresente os resultados e discuta.

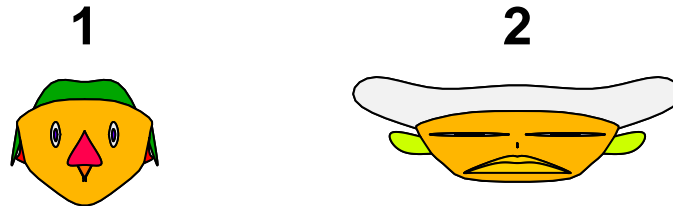


Podemos perceber que o k-means obteve um resultado muito bom, porem nao perfeito. Identificando 4 grupos porem nao contendo os elementos esperados de cada um. no caso o ponto do meio seria mais correto estar agrupado sozinho ou com o outro valor extremo.

84. Seis variáveis são medidas de 100 notas genuínas e 100 notas falsificadas (Flury and Riedwyl, 1988.)

##	Status	Length	Left	Right	Bottom	Top	Diagonal
## 1	0	214.8	131.0	131.1	9.0	9.7	141.0
## 2	0	214.6	129.7	129.7	8.1	9.5	141.7
## 3	0	214.8	129.7	129.7	8.7	9.6	142.2
## 4	0	214.8	129.7	129.6	7.5	10.4	142.0
## 5	0	215.0	129.6	129.7	10.4	7.7	141.8
## 6	0	215.7	130.8	130.5	9.0	10.1	141.4

(a) Calcule as médias de cada variável em cada população (notas verdadeiras e falsificadas). Obtenha também a matriz de variância-covariâncias Σ em cada população. Obtenha um face plot de cada população utilizando os valores médios. Compare os resultados numéricos com os gráficos. Discuta. Utilize o face plot para comparar os valores médios das variáveis para notas verdadeiras e notas falsas. Discuta os resultados.



```
## effect of variables:
## modified item      Var
## "height of face   " "Length"
## "width of face    " "Left"
## "structure of face" "Right"
## "height of mouth  " "Bottom"
## "width of mouth   " "Top"
## "smiling          " "Diagonal"
## "height of eyes   " "Length"
## "width of eyes    " "Left"
## "height of hair   " "Right"
## "width of hair    " "Bottom"
## "style of hair    " "Top"
## "height of nose   " "Diagonal"
## "width of nose    " "Length"
## "width of ear     " "Left"
## "height of ear    " "Right"
```

Medidas resumo das notas verdadeiras

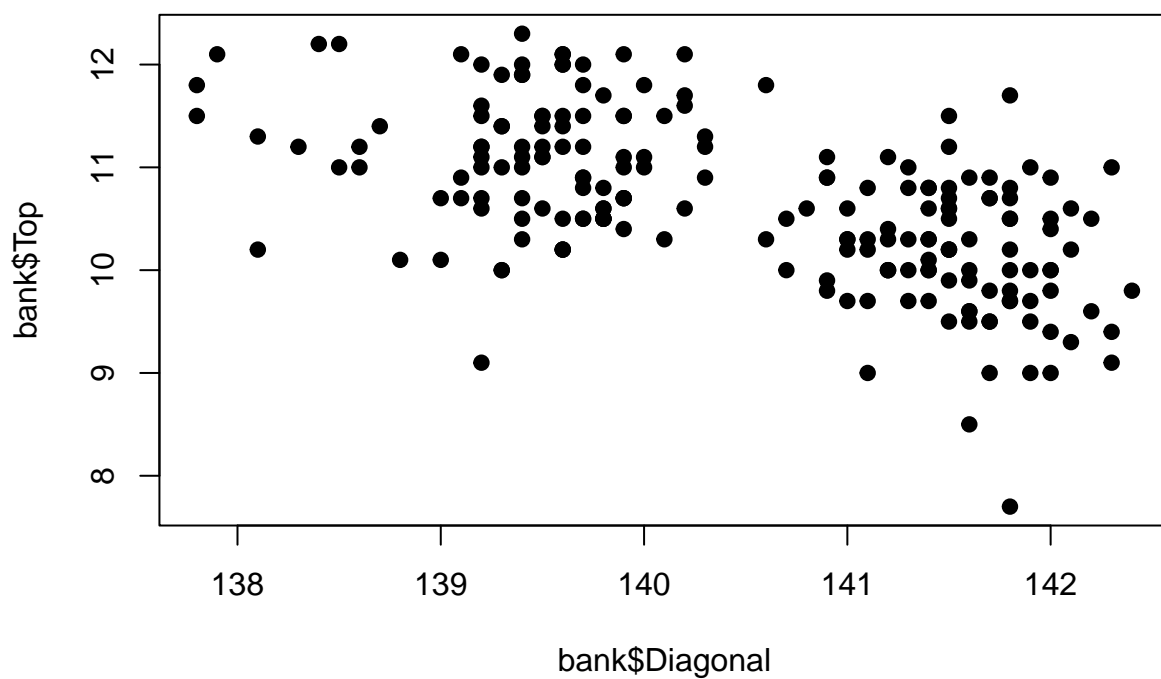
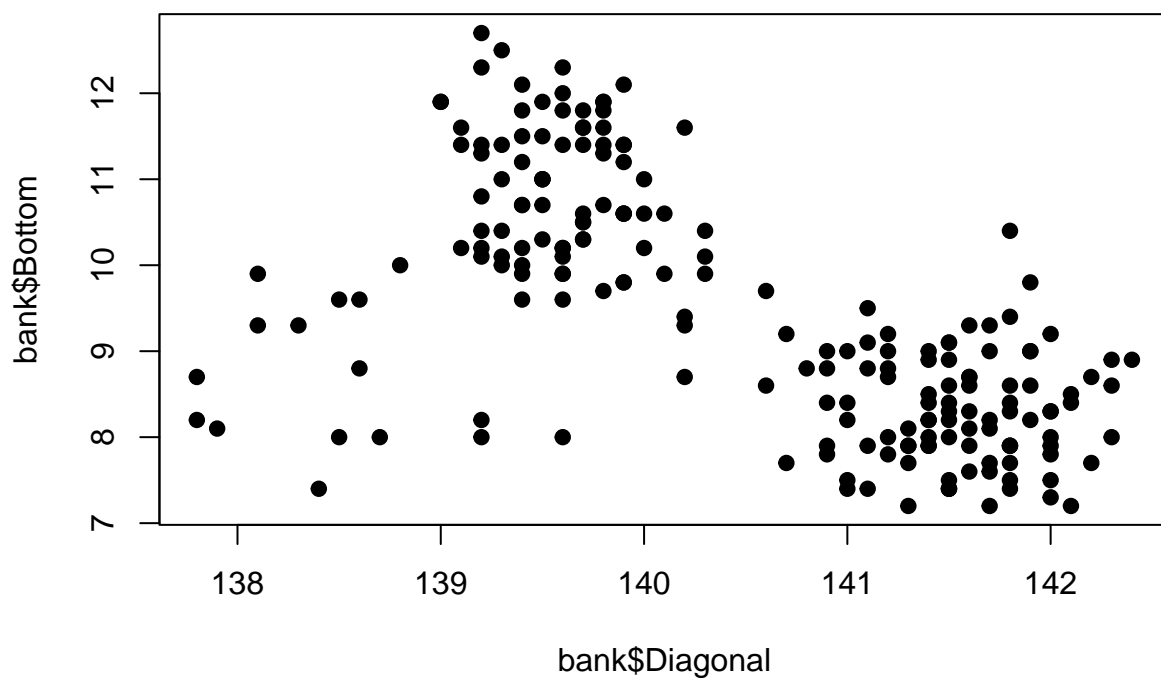
```
## Status      Length      Left      Right      Bottom
## 0:100  Min.    :213.8  Min.    :129.0  Min.    :129.0  Min.    : 7.200
## 1: 0    1st Qu.:214.7  1st Qu.:129.7  1st Qu.:129.4  1st Qu.: 7.900
##          Median :215.0  Median :129.9  Median :129.7  Median : 8.250
##          Mean   :215.0  Mean   :129.9  Mean   :129.7  Mean   : 8.305
##          3rd Qu.:215.2  3rd Qu.:130.2  3rd Qu.:130.0  3rd Qu.: 8.800
##          Max.   :215.9  Max.   :131.0  Max.   :131.1  Max.   :10.400
##          Top      Diagonal
## Min.    : 7.700  Min.    :139.6
## 1st Qu.: 9.775  1st Qu.:141.2
## Median :10.200  Median :141.5
## Mean   :10.168  Mean   :141.5
## 3rd Qu.:10.600  3rd Qu.:141.8
## Max.   :11.700  Max.   :142.4
```

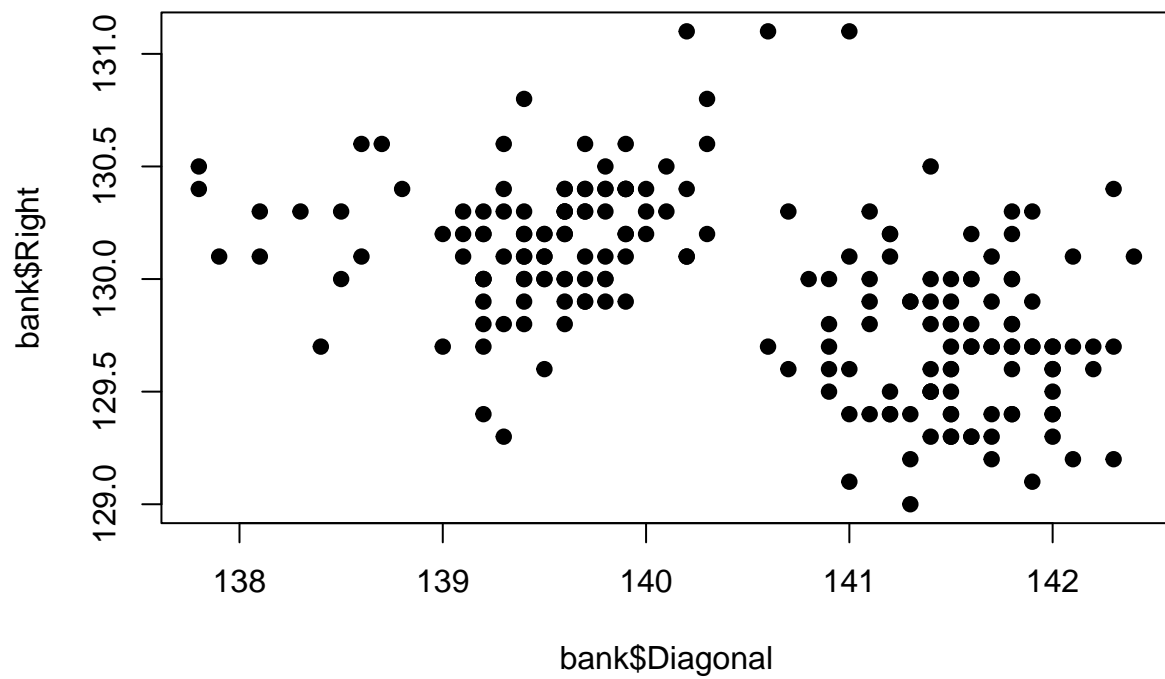
Medidas resumo das notas falsas

```
## Status      Length      Left      Right      Bottom
## 0: 0    Min.    :213.9  Min.    :129.6  Min.    :129.3  Min.    : 7.40
## 1:100   1st Qu.:214.6  1st Qu.:130.1  1st Qu.:130.0  1st Qu.: 9.90
##          Median :214.8  Median :130.3  Median :130.2  Median :10.60
##          Mean   :214.8  Mean   :130.3  Mean   :130.2  Mean   :10.53
##          3rd Qu.:215.0  3rd Qu.:130.5  3rd Qu.:130.4  3rd Qu.:11.40
##          Max.   :216.3  Max.   :130.8  Max.   :131.1  Max.   :12.70
##          Top      Diagonal
## Min.    : 9.10  Min.    :137.8
## 1st Qu.:10.68  1st Qu.:139.2
## Median :11.10  Median :139.5
## Mean   :11.13  Mean   :139.4
## 3rd Qu.:11.53  3rd Qu.:139.8
## Max.   :12.30  Max.   :140.6
```

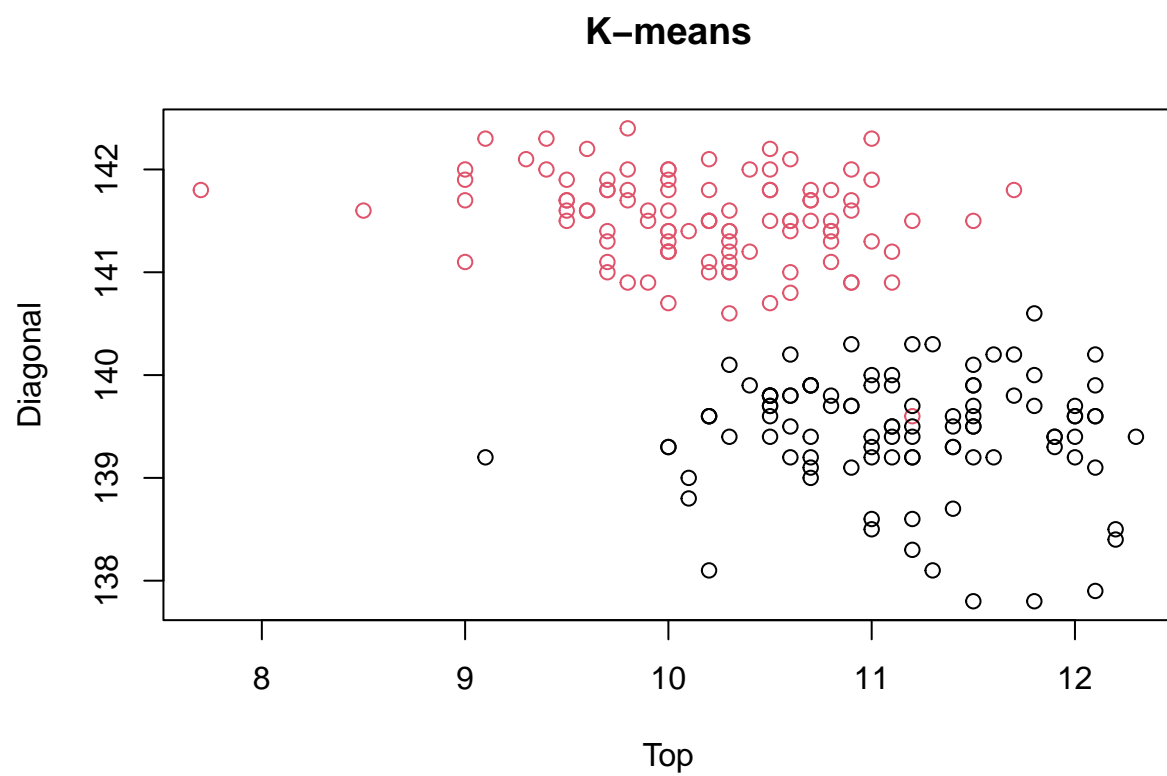
Podemos perceber pelo face plot que os 2 grupos tem características bem diferentes em especial nas variáveis, Bottom, Top e Diagonal, que no faceplot são responsáveis pelo cabelo do personagem. Levando em consideração as medidas resumo podemos ver que tem uma diferença grande entre os grupos na escala de uma nota de dinheiro

(b) Utilize um método de agrupamento para verificar se é possível distinguir as notas falsas das verdadeiras, imaginando que você não soubesse se 100 últimas notas eram verdadeiras.



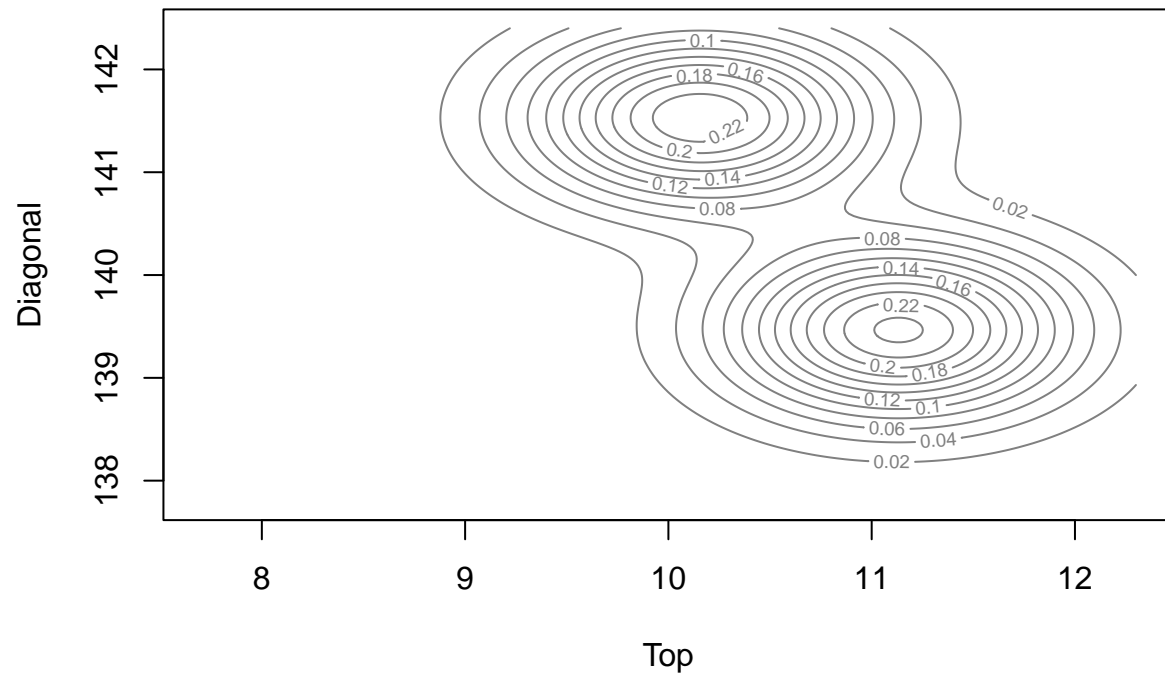


Podemos notar pelos graficos seleccionados que a diagonal da nota parece agrupar bem as observacoes em relacao as outras variaveis.



Podemos perceber que esta bem dividido com apenas uma observacao ficando agrupada no cluster errado

(c) Compare os resultados do agrupamento anterior com o agrupamento por mistura de normais



O modelo utilizado foi o EII para agrupar os dados

(d) Utilize o Índice Ajustado de Rand para comparar seu agrupamento das 200 notas com a verdadeira classificação em genuínas e falsas. Avalie a qualidade do agrupamento.

Comparando os resultados do k-mean com a verdadeira classificação

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

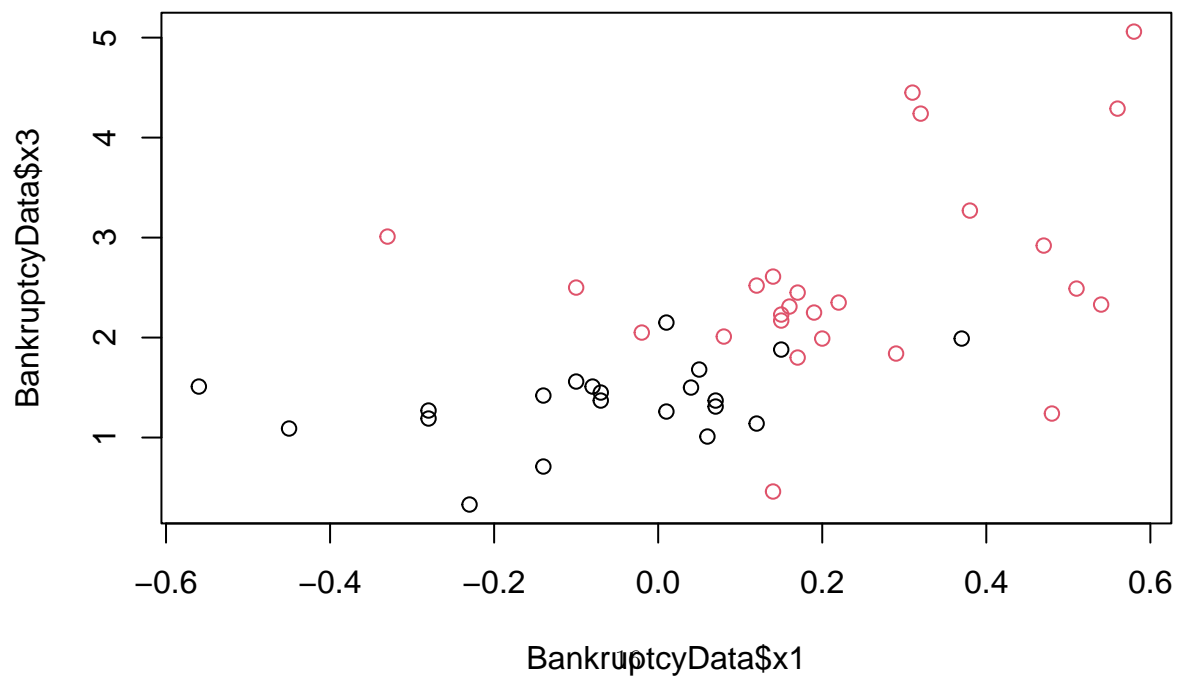
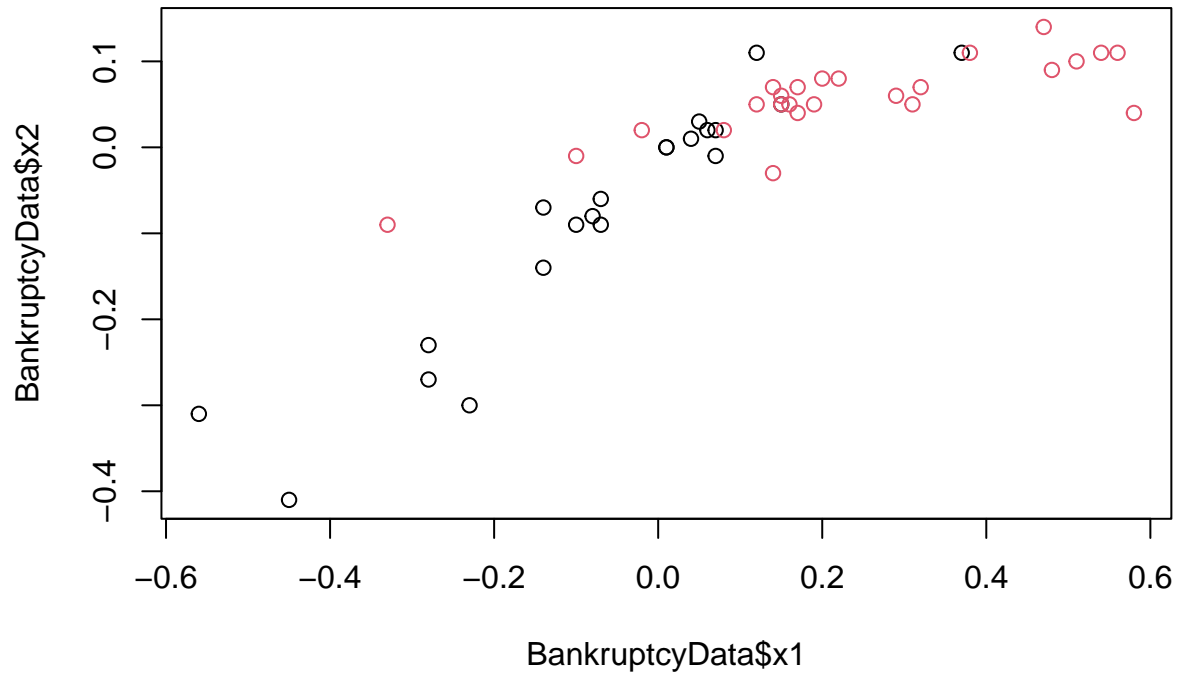
Comparando os resultados da mistura de normais com a verdadeira classificação

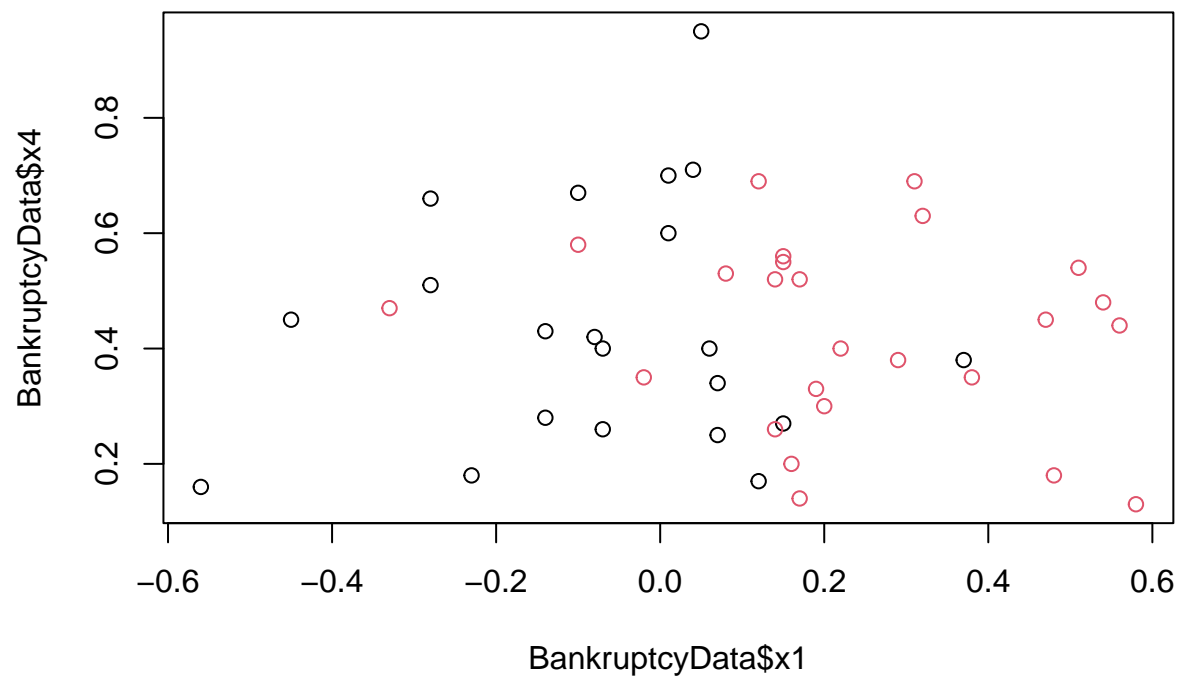
```
## [1] 0.9799995
```

Podemos perceber que o K-means agrupou todos muito bem mostrando um desempenho melhor que a mistura de normais, que ainda sim deu um bom resultado

89. Johnson e Wichern - Exercício 11.24. - Dados: T11-4-BankruptcyData.dat.

a)





b)

```
## Vetor de medias dos ativos: -0.06904762 -0.08142857
```

```
## Vetor de medias dos falidos: 0.2352 0.0556
```

```
## Covariancia dos ativos
```

```
##      x1      x2
## x1 0.04413 0.02848
## x2 0.02848 0.02100
```

```
## Covariancia dos falidos
```

```
##      x1      x2
## x1 0.04705 0.00851
## x2 0.00851 0.00238
```

Temos matrizes de covariancia diferente entre as populacoes.

c)

```
## Matriz de confusao
```

```
##  
##      0  1  
##  0 13  8  
##  1  1 24
```

```
## Acerto de classificacao por grupo 0.6190476 0.96
```

```
## Porporcao total de acerto 0.8043478
```

Com um calssificardo quadratico obtivemos um bom desempenho com 80% de acerto

d)

```
## erro aparente (APER) deste conjunto foi: 0.1956522
```

```
## estimacão da taxa de erro aparente foi: 0.2173913
```

Nosso erro aparente foi de 20% como esperado pelo resultado anterior

e)

```
## Matriz de confusao
```

```
##  
##      0  1  
##  0  9 12  
##  1  0 25
```

```
## Acerto de classificacao por grupo 0.4285714 1
```

```
## Porporcao total de acerto 0.7391304
```

```
## erro aparente (APER) deste conjunto foi: 0.2608696
```

```
## estimacão da taxa de erro aparente foi: 0.2608696
```

Quando mudamos as priores nos conseguimos acertar todos do grupo 1 porem erramos mais da metade do grupo 2

f)

```
## `summarise()` has grouped output by 'x5'. You can override using the `.groups`  
## argument.
```

```
## # A tibble: 4 x 3  
## # Groups:   x5 [2]  
##   x5     pop[,1] freq  
##   <fct>    <dbl> <int>  
## 1 0             0    15  
## 2 0             1     6  
## 3 1             0     3  
## 4 1             1    22
```

```
## erro aparente (APER) deste conjunto foi: 0.1956522
```

Usando o classificador de fisher que assume matrizes de covariancia iguais entre as duas populacoes, o que nao é o caso, obteve umm desempenho melhor que o da funcao quadratica apesar de nao ser indicado para situacoes como essa.

g)

Analise com X_1 e X_3

```
## Vetor de medias dos ativos: -0.06904762 1.366667
```

```
## Vetor de medias dos falidos: 0.2352 2.5936
```

```
## Covariancia dos ativos
```

```
##           x1      x3  
## x1 0.04413 0.03449  
## x3 0.03449 0.16430
```

```
## Covariancia dos falidos
```

```
##           x1      x3  
## x1 0.04705 0.07493  
## x3 0.07493 1.04677
```

Novamente matriz de covariancia diferentes

Usando a priori 0.5|0.5

```
## Matriz de confusao
```

```
##  
##      0  1  
## 0 19  2  
## 1  3 22
```

```
## Acerto de classificacao por grupo 0.9047619 0.88
```

```
## Porporcao total de acerto 0.8913043
```

```
## erro aparente (APER) deste conjunto foi: 0.1086957
```

```
## estimacao da taxa de erro aparente foi: 0.1304348
```

Usando a o X_3 no lugar do X_2 ajudou na melhora do classificador, errando apenas 10% neste conjunto de dados

Usando a priori 0.05|0.95

```
## Matriz de confusao
```

```
##  
##      0  1  
##  0  4 17  
##  1  0 25
```

```
## Acerto de classificacao por grupo 0.1904762 1
```

```
## Porporcao total de acerto 0.6304348
```

```
## erro aparente (APER) deste conjunto foi: 0.3695652
```

```
## estimacao da taxa de erro aparente foi: 0.3913043
```

Ao mudar a priori temos um desempenho pior do que quando usamos X_2 , ele continua acertando todos do grupo 1 porem errando mais do grupo 2.

Analise com X_1 e X_4

```
## Vetor de medias dos ativos: -0.06904762 0.437619
```

```
## Vetor de medias dos falidos: 0.2352 0.4268
```

```
## Covariancia dos ativos
```

```
##      x1      x4  
## x1 0.04413 0.00415  
## x4 0.00415 0.04458
```

```
## Covariancia dos falidos
```

```
##      x1      x4  
## x1 0.04705 -0.00670  
## x4 -0.00670 0.02638
```

As matrizes estao mais parecidas em alguns lugares porem ainda sao diferentes

```
## Matriz de confusao
```

```
##
##      0  1
##  0 17  4
##  1  4 21
```

```
## Acerto de classificacao por grupo 0.8095238 0.84
```

```
## Porporcao total de acerto 0.826087
```

```
## erro aparente (APER) deste conjunto foi: 0.173913
```

```
## estimacao da taxa de erro aparente foi: 0.2173913
```

Usando X_4 obtivemos um bom desempenho, errando somente 17% neste conjunto de dados, o que é um bom resultado

```
## Matriz de confusao
```

```
##
##      0  1
##  0  3 18
##  1  0 25
```

```
## Acerto de classificacao por grupo 0.1428571 1
```

```
## Porporcao total de acerto 0.6086957
```

```
## erro aparente (APER) deste conjunto foi: 0.3913043
```

```
## estimacao da taxa de erro aparente foi: 0.4565217
```

Assim como nos outros resultados, temos um desempenho geral ruim, porem acertando tudo do grupo 1

h)

```
## Vetor de medias dos ativos: -0.06904762 0.437619
```

```
## Vetor de medias dos falidos: 0.2352 0.4268
```

```
## Covariancia dos ativos
```

```
##      x1      x2      x3      x4
## x1 0.04413 0.02848 0.03449 0.00415
## x2 0.02848 0.02100 0.02602 0.00344
## x3 0.03449 0.02602 0.16430 0.03278
## x4 0.00415 0.00344 0.03278 0.04458
```

```
## Covariancia dos falidos
```

```
##          x1          x2          x3          x4
## x1  0.04705 0.00851 0.07493 -0.00670
## x2  0.00851 0.00238 0.00858  0.00019
## x3  0.07493 0.00858 1.04677  0.03263
## x4 -0.00670 0.00019 0.03263  0.02638
```

Novamente matrizes diferentes!

Usando a priori 0.5|0.5

```
## Matriz de confusao
```

```
##
##      0  1
##  0 19  2
##  1  1 24
```

```
## Acerto de classificacao por grupo 0.9047619 0.96
```

```
## Porporcao total de acerto 0.9347826
```

```
## erro aparente (APER) deste conjunto foi: 0.06521739
```

```
## estimacao da taxa de erro aparente foi: 0.1086957
```

Usando todas as variaveis disponiveis, temos um classificador ainda melhor errando somente 6% neste conjunto de dados

Usando a priori 0.05|0.95

```
## Matriz de confusao
```

```
##
##      0  1
##  0 12  9
##  1  0 25
```

```
## Acerto de classificacao por grupo 0.5714286 1
```

```
## Porporcao total de acerto 0.8043478
```

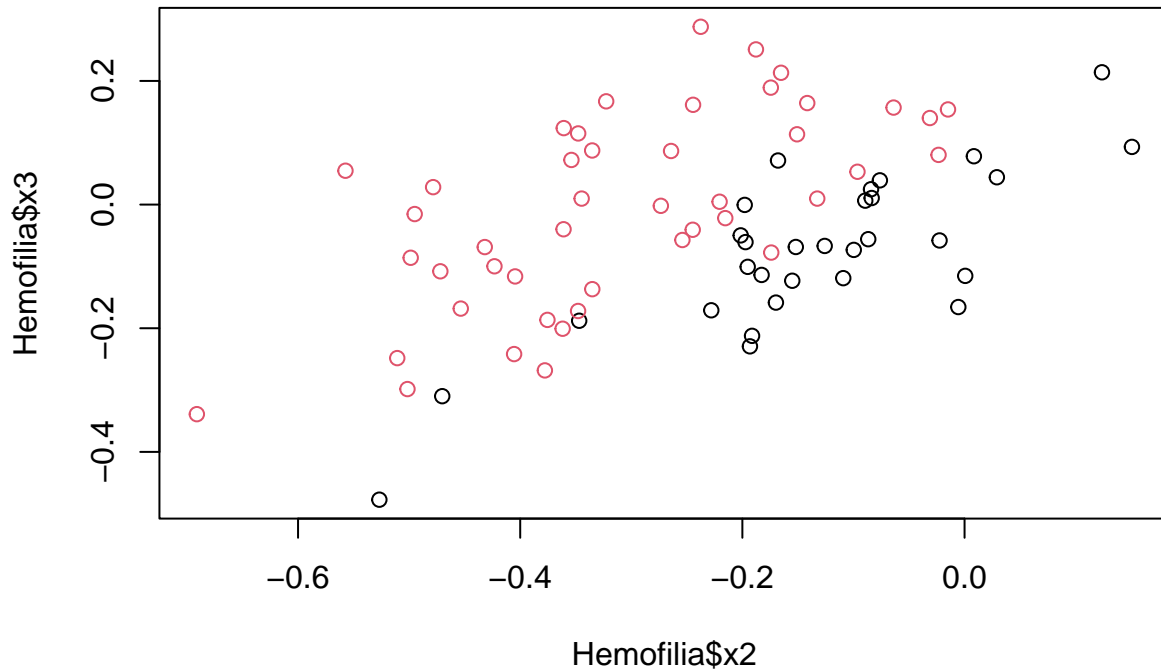
```
## erro aparente (APER) deste conjunto foi: 0.1956522
```

```
## estimacao da taxa de erro aparente foi: 0.2391304
```

Usando todas as variaveis disponiveis, temos um classificador ainda melhor e mesmo com a priori tendendo forte para um grupo, conseguimos minimizar os erros em relacao aos classificadores anteriores

90. Johnson e Wichern - Exercício 11.32. - Dados: T11-8-Hemofilia.dat.

a)



b)

```
## Vetor de medias do grupo 1: -0.13487 -0.07785667
```

```
## Vetor de medias dos grupo 2: -0.3079467 -0.005991111
```

```
## Covariancia do grupo 1
```

```
##           x2           x3
## x2 0.02089726 0.01551495
## x3 0.01551495 0.01792005
```

```
## Covariancia do grupo 2
```

```
##           x2           x3
## x2 0.02378380 0.01537617
## x3 0.01537617 0.02403505
```

Validacao Holdout

Para essa passo o banco de dados foi semaprado em 2 cada um com metade das observacoes e foi rodado um classificador linear no banco de treinamento e depois foi aplicado no de teste.

```
## Matriz de confusao

##
##      FALSE TRUE
##  1      14     1
##  2       8    15

## Acerto de classificacao por grupo 0.9333333 0.6521739

## Porporcao total de acerto 0.7631579

## erro aparente (APER) deste conjunto foi: 0.1466667

## estimacao da taxa de erro aparente foi: 0.16
```

Obtemos um classificador bom, com apenas 14% de taxa de erro!

c)

```
##
## -- Column specification -----
## cols(
##   X1 = col_double(),
##   X2 = col_double()
## )

## $class
## [1] 1 1 1 1 1 1 1 1 1 1
## Levels: 1 2
##
## $posterior
##           1           2
## 1  0.9975913 0.002408672
## 2  0.9509347 0.049065315
## 3  0.9693462 0.030653836
## 4  0.9490333 0.050966732
## 5  0.9750316 0.024968358
## 6  0.9622121 0.037787867
## 7  0.9653879 0.034612064
## 8  0.9592781 0.040721901
## 9  0.7211708 0.278829171
## 10 0.7273436 0.272656411
##
## $x
##           LD1
## 1  -3.0316295
## 2  -1.4912443
## 3  -1.7375324
## 4  -1.4711104
## 5  -1.8436782
## 6  -1.6285590
```



```
## 7 -1.6743790
## 8 -1.5894042
## 9 -0.4780549
## 10 -0.4936047
```

Todas as observacoes foram para o grupo 1 de acordo com nosso classificador

d)

Nessa questao a unica coisa que foi mudada era as probabilidades a priori, o resto ficou igual!

```
## Matriz de confusao
```

```
##
##      FALSE TRUE
##  1      14     1
##  2       8    15
```

```
## Acerto de classificacao por grupo 0.9333333 0.6521739
```

```
## Porporcao total de acerto 0.7631579
```

```
## erro aparente (APER) deste conjunto foi: 0.2368421
```

Obtemos um modelo com uma taxa de erro de 23%

```
## $class
## [1] 1 1 1 1 1 1 1 1 1 1
## Levels: 1 2
##
## $posterior
##           1           2
## 1  0.9991958 0.0008041821
## 2  0.9830918 0.0169081749
## 3  0.9895689 0.0104311143
## 4  0.9824135 0.0175864602
## 5  0.9915363 0.0084636687
## 6  0.9870785 0.0129214723
## 7  0.9881901 0.0118098639
## 8  0.9860472 0.0139527554
## 9  0.8858353 0.1141646872
## 10 0.8889242 0.1110758283
##
## $x
##           LD1
## 1 -2.534679891
## 2 -0.994294738
## 3 -1.240582786
## 4 -0.974160772
## 5 -1.346728650
## 6 -1.131609395
```

```
## 7 -1.177429369
## 8 -1.092454572
## 9 0.018894691
## 10 0.003344922
```

Novamente todas as novas observacoes vao para o mesmo grupo 1

91. Considere os dados sobre falsificação de notas do Exercício 84. Realize as análises Linear e Quadrática de discriminantes. É possível utilizar mistura de normais (função `mclust`) como análise de discriminantes? Se sim, faça o estudo e compare as estimativas APER para as três situações. Resuma os resultados e conclusões.

Usando um discriminante linear

```
## Matriz de confusao

##
##      0  1
##  0  99  1
##  1   0 100

## Acerto de classificacao por grupo 0.99 1

## Porporcao total de acerto 0.995

## erro aparente (APER) deste conjunto foi: 0.005
```

Nosso discriminante obteve um erro de apenas 0.005 ou seja errou somente uma classificacao

```
## Matriz de confusao

##
##      0  1
##  0  99  1
##  1   0 100

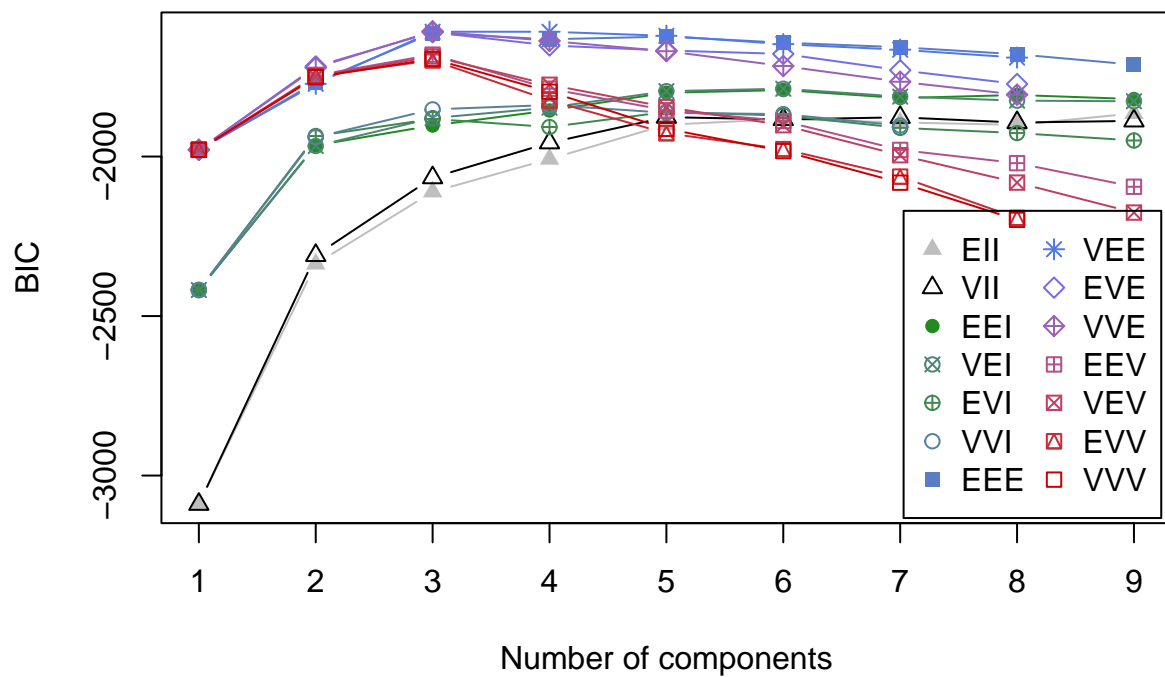
## Acerto de classificacao por grupo 0.99 1

## Porporcao total de acerto 0.995

## erro aparente (APER) deste conjunto foi: 0.005
```

Usando o `mclust` podemos rodar discriminantes por mistura de normais, ao usar a funcao `mclustBIC()` pode gerar um gráfico com diferentes discriminantes de misturas de normais e a quantidade de grupos que ele indica ter.

```
## Best BIC values:
##           VVE,3      VEE,4      VEE,3
## BIC      -1607.574 -1608.767736 -1608.793746
## BIC diff      0.000   -1.194096   -1.220106
```



```
## -----
## Gaussian finite mixture model fitted by EM algorithm
## -----
##
## Mclust VVE (ellipsoidal, equal orientation) model with 3 components:
##
## log-likelihood  n df      BIC      ICL
##      -663.3814 200 53 -1607.574 -1607.71
##
## Clustering table:
##  1  2  3
## 18 98 84

##
## Class      1  2  3
## Genuína    2 98  0
## Falsificada 16  0 84
```

```
cat("Matriz de confusao")
```

```
## Matriz de confusao
```

```
M
```

```
##  
##      1  2  3  
##    0  2 98  0  
##    1 16  0 84
```

```
cat("Index de Rand",RAND)
```

```
## Index de Rand 0.8418856
```

```
cat("erro aparente (APER) deste conjunto foi:",APER)
```

```
## erro aparente (APER) deste conjunto foi: 0.09
```