Tarefa1

Lucas G Nachtigall

2022-09-14

```
## Warning: package 'gridExtra' was built under R version 4.2.1
```

```
## Warning: package 'GGally' was built under R version 4.2.1
```

1.Carregue a base de dados e mostre a estrutura do dataset (str()). Comente sobre o número de amostras, de variáveis (e seu tipo). O arquivo do dataset não pode ser modificado de forma alguma. A leitura deverá ser realizada de tal maneira qualquer característica dos dados.

```
rm(list=ls())
#setwd("./Downloads")
dados = read.csv2("Dry_Bean_Dataset.csv",header=T)
str(dados)
```

```
## 'data.frame':    13611 obs. of  17 variables:
##  $ Area           : int  28395 28734 29380 30008 30140 30279 30477 30519 30685 30834 ...
##  $ Perimeter      : num  610 638 624 646 620 ...
##  $ MajorAxisLength: num  208 201 213 211 202 ...
##  $ MinorAxisLength: num  174 183 176 183 190 ...
##  $ AspectRation   : num  1.2 1.1 1.21 1.15 1.06 ...
##  $ Eccentricity   : num  0.55 0.412 0.563 0.499 0.334 ...
##  $ ConvexArea     : int  28715 29172 29690 30724 30417 30600 30970 30847 31044 31120 ...
##  $ EquivDiameter  : num  190 191 193 195 196 ...
##  $ Extent         : num  0.764 0.784 0.778 0.783 0.773 ...
##  $ Solidity       : num  0.989 0.985 0.99 0.977 0.991 ...
##  $ roundness      : num  0.958 0.887 0.948 0.904 0.985 ...
##  $ Compactness    : num  0.913 0.954 0.909 0.928 0.971 ...
##  $ ShapeFactor1   : num  0.00733 0.00698 0.00724 0.00702 0.0067 ...
##  $ ShapeFactor2   : num  0.00315 0.00356 0.00305 0.00321 0.00366 ...
##  $ ShapeFactor3   : num  0.834 0.91 0.826 0.862 0.942 ...
##  $ ShapeFactor4   : num  0.999 0.998 0.999 0.994 0.999 ...
##  $ Class          : chr  "SEKER" "SEKER" "SEKER" "SEKER" ...
```

O dataset possui 17 colunas: 16 colunas numéricas e 1 categórica. A variável categórica é o rótulo da classe.

2.Altere a variável do tipo do feijão (Class) para um factor. Utilizando um comando mostre como estimar o número de classes existentes.

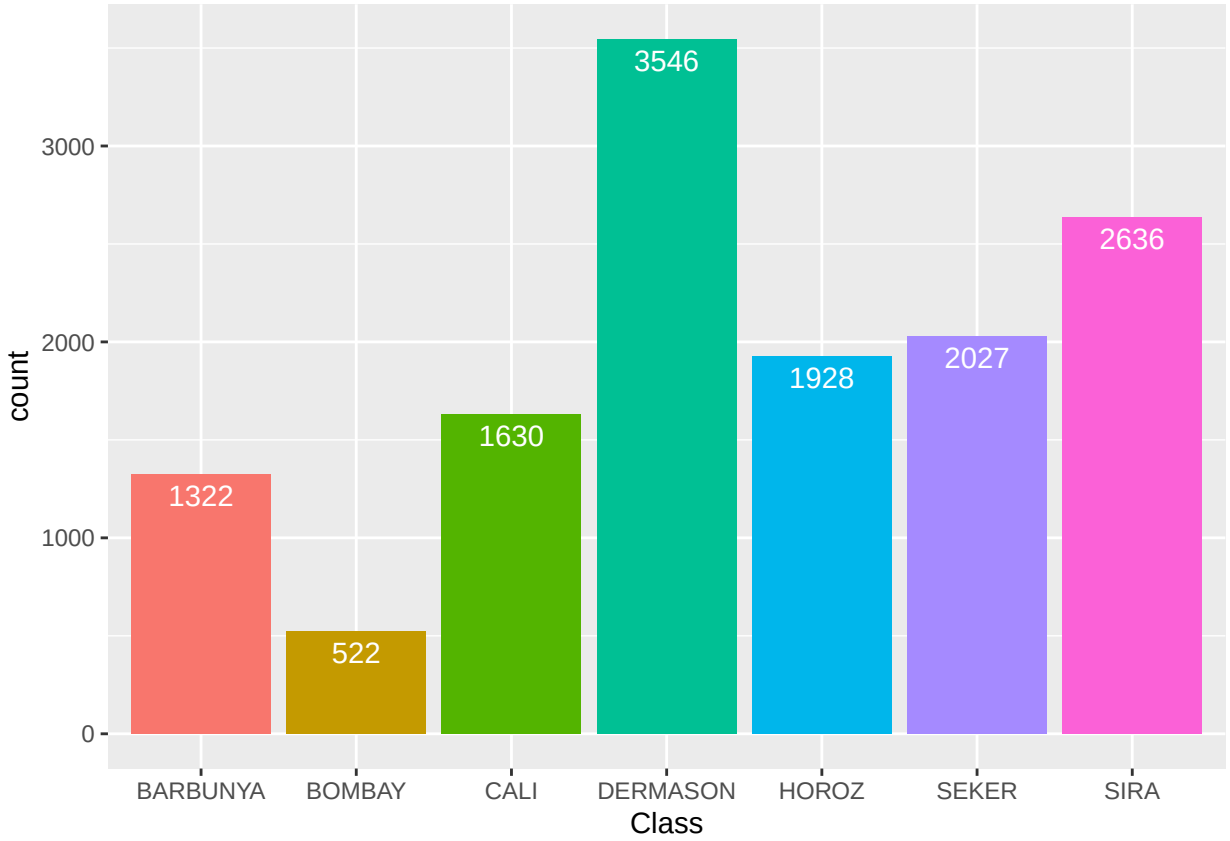
```
dados$Class=factor(dados$Class)
str(dados$Class)
```

```
##  Factor w/ 7 levels "BARBUNYA","BOMBAY",..: 6 6 6 6 6 6 6 6 6 6 ...
```

O database possui 7 classes.

3.Quantas amostras existem por classe? Use um gráfico de barras para ilustrar as quantidades.

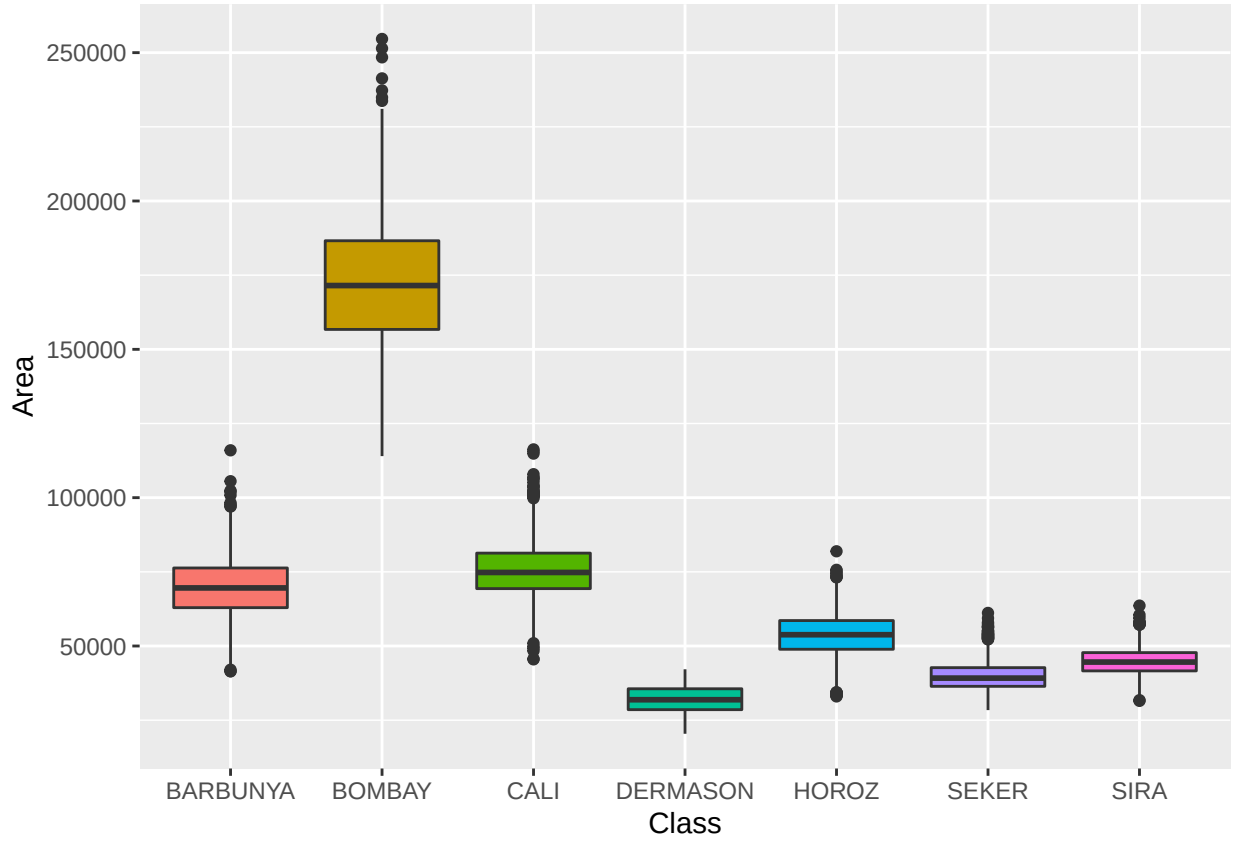
```
ggplot(dados,aes(Class,fill=Class)) +
  geom_bar()+ geom_text(aes(label = ..count..), stat = "count", vjust = 1.5, colour = "white") + theme(
```

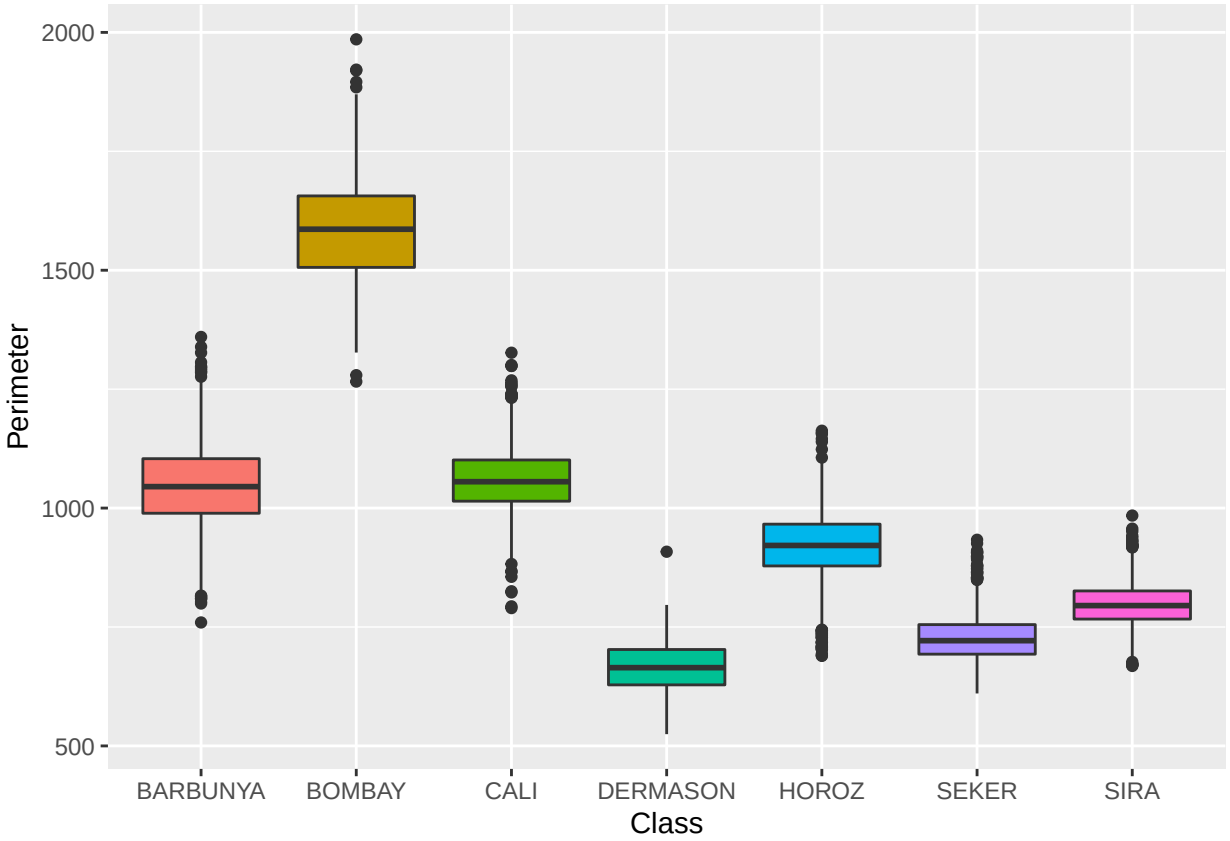


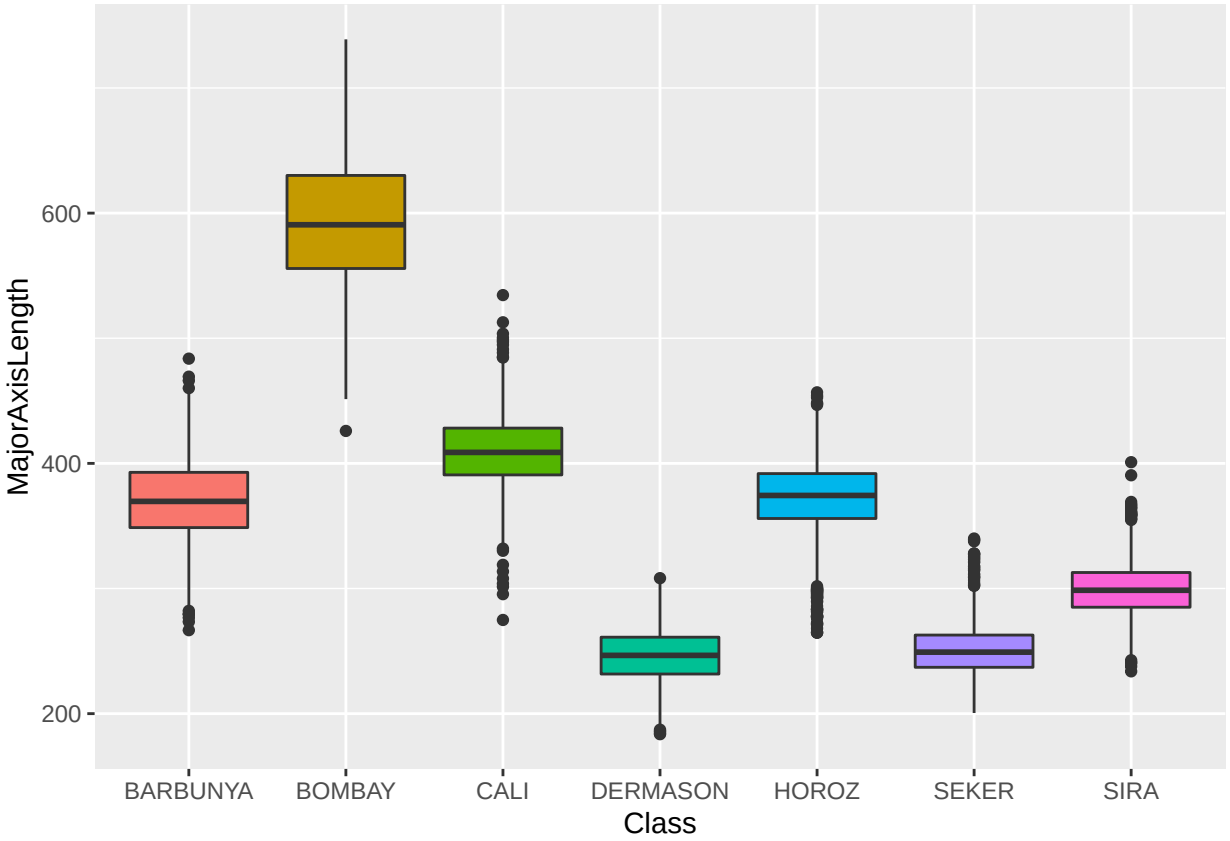
Das 7 classes (Barbunya, Bombay, Cali, Dermason, Horoz, Seker e Sira) existentes, sendo que as classes possuem o número de amostras divergentes. Isto é, o dataset é desbalanceado.

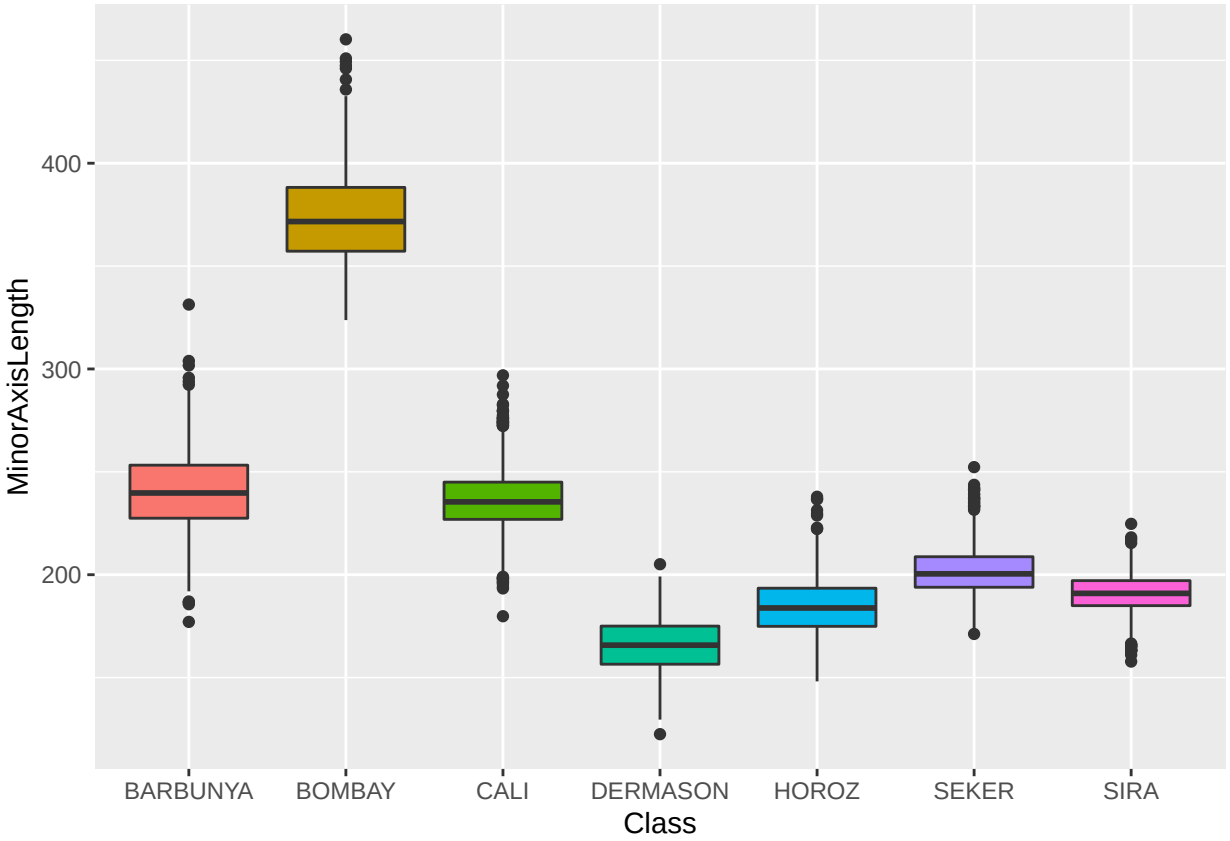
4.Alguma variável apresenta outliers? Tem alguma que não apresenta? Se as amostras com outliers fossem removidas, reduziria em quanto o número de amostras? Alguma classe sofreria uma redução maior do que a outra? Crie um boxplot por variável (boxplot()) para auxiliar na explicação. Não remova os outliers para as próximas etapas!

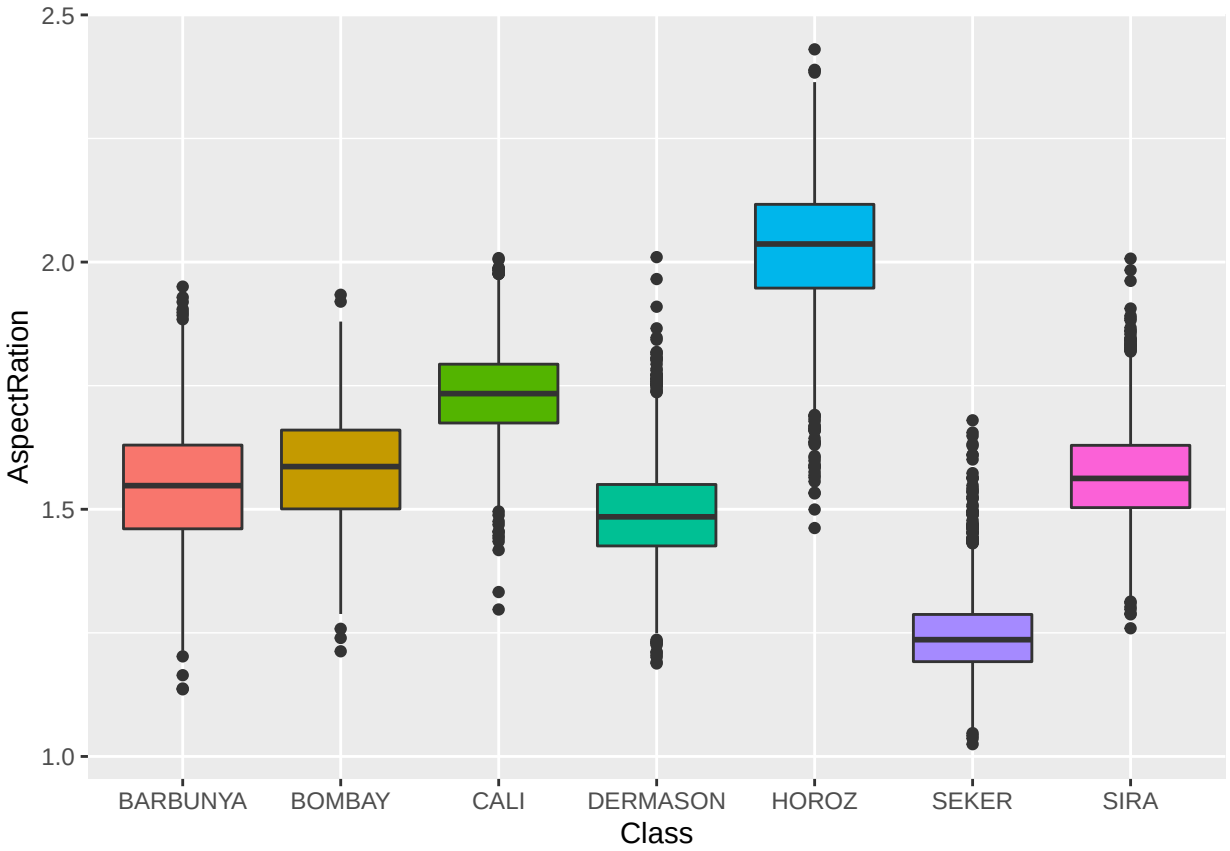
```
p = list()
for(i in 1:16){
  p[[i]] = ggplot(dados, aes_string(x="Class", y=names(dados)[i],fill="Class")) + geom_boxplot() + theme
  do.call(grid.arrange,c(p[i],ncol=1))
}
```

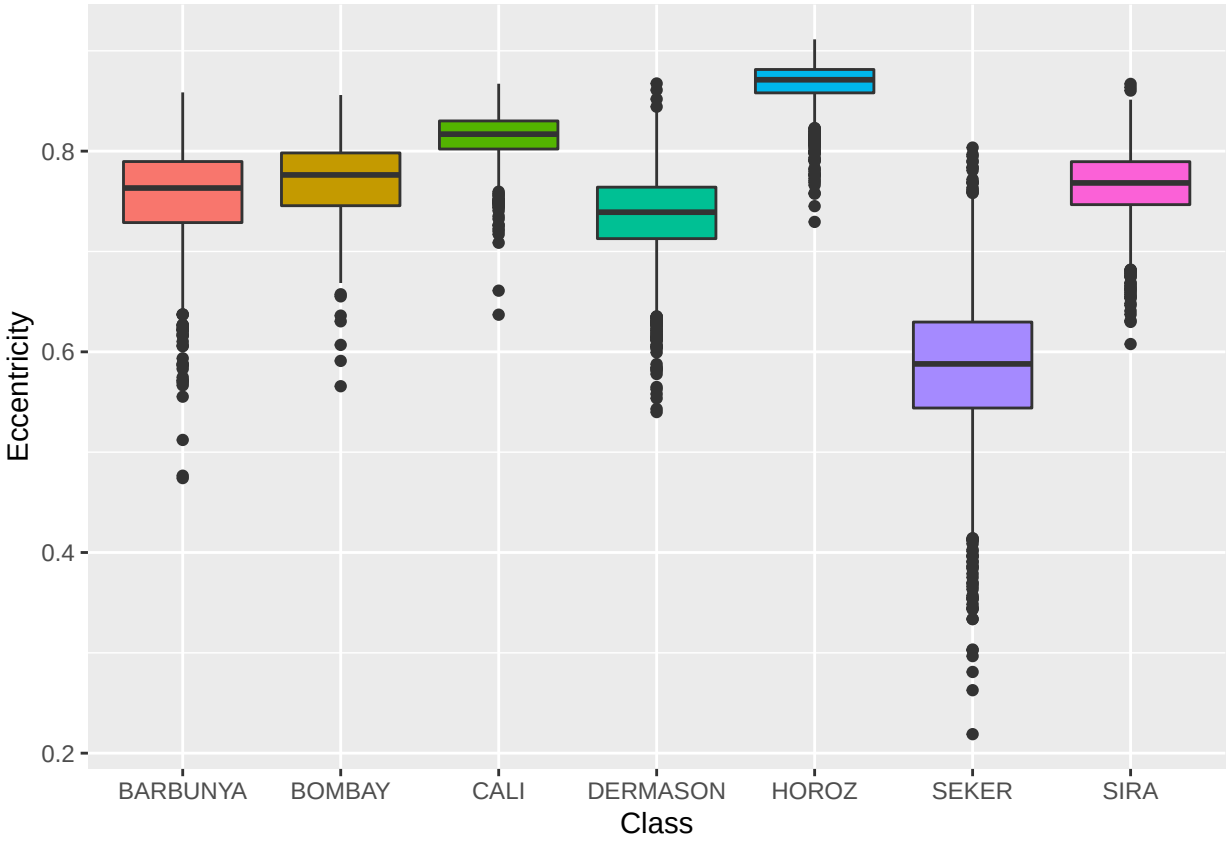


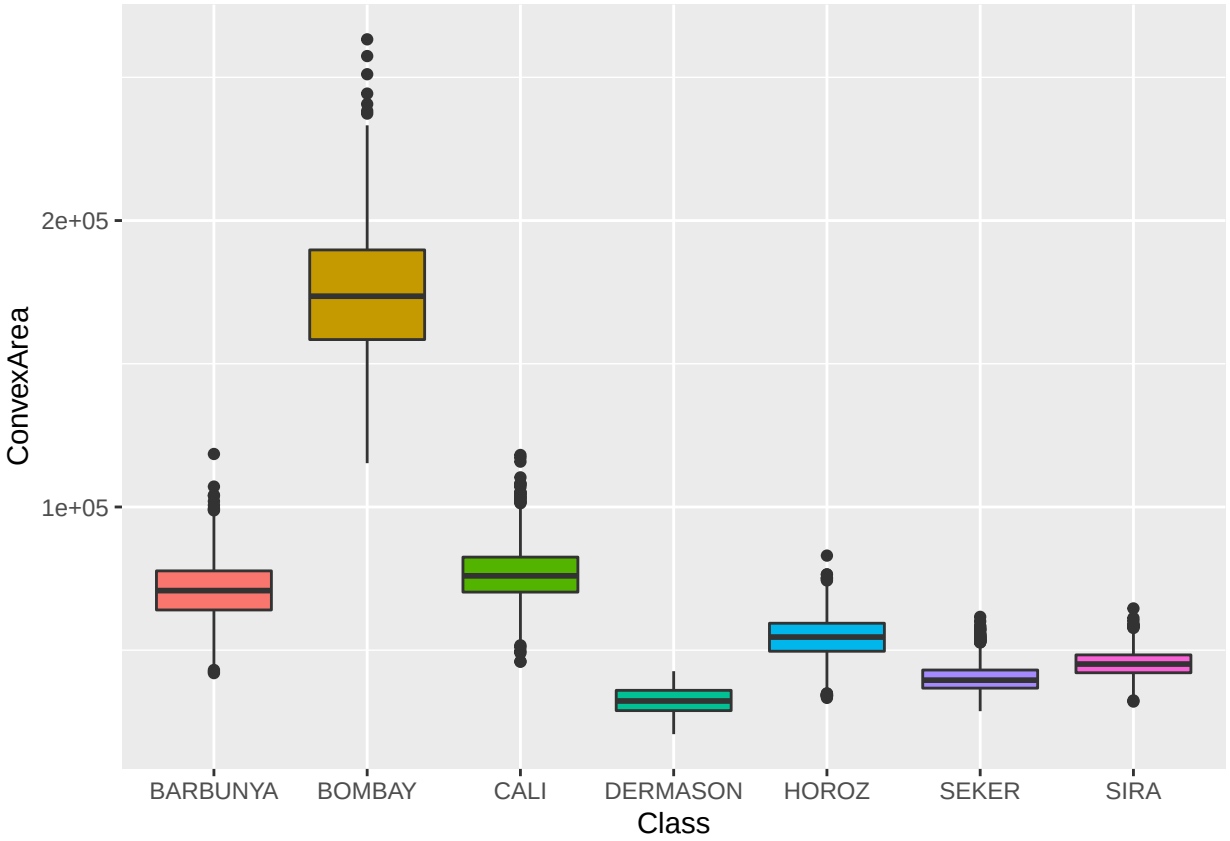

2000
1500
1000
500
Perimeter
BARBUNYA
BOMBAY
CALI
DERMASON
HOROZ
SEKER
SIRA
Class

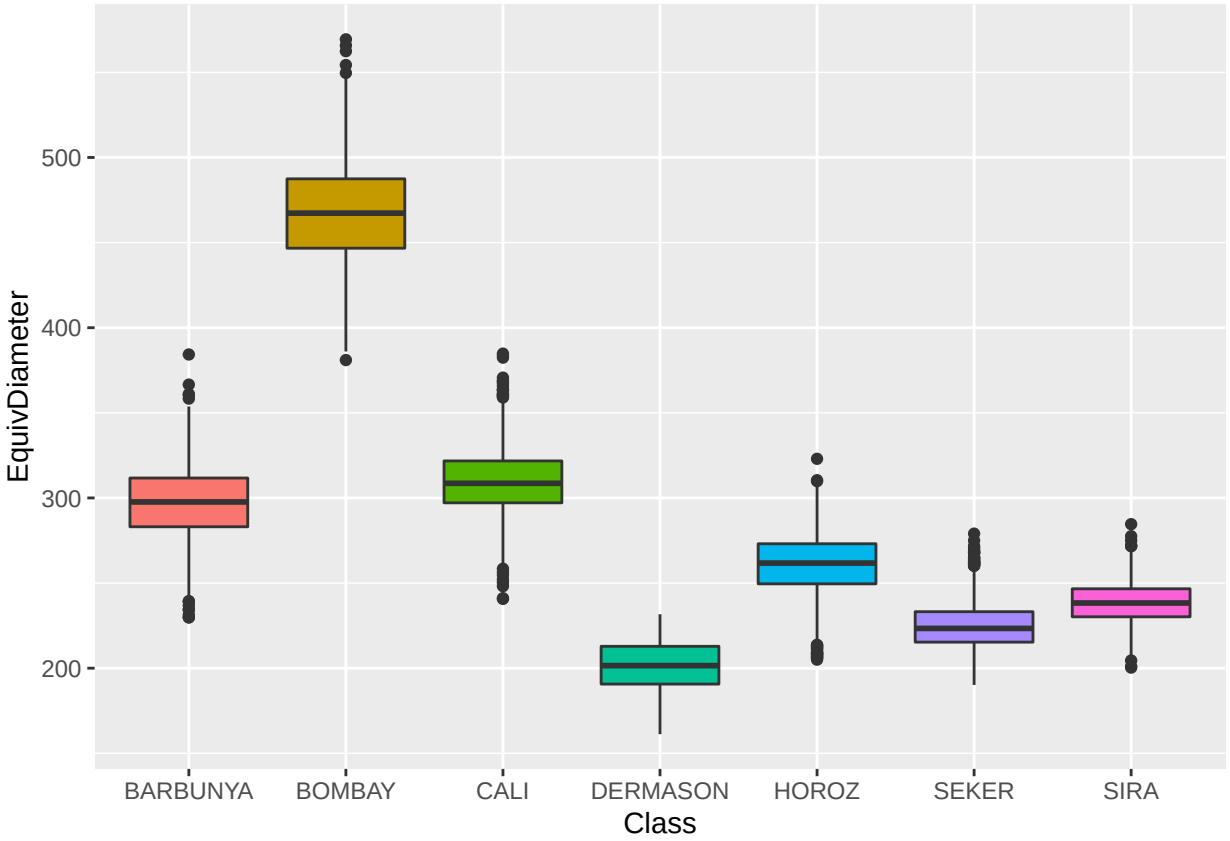

MajorAxisLength
600
400
200
BARBUNYA
BOMBAY
CALI
DERMASON
HOROZ
SEKER
SIRA
Class

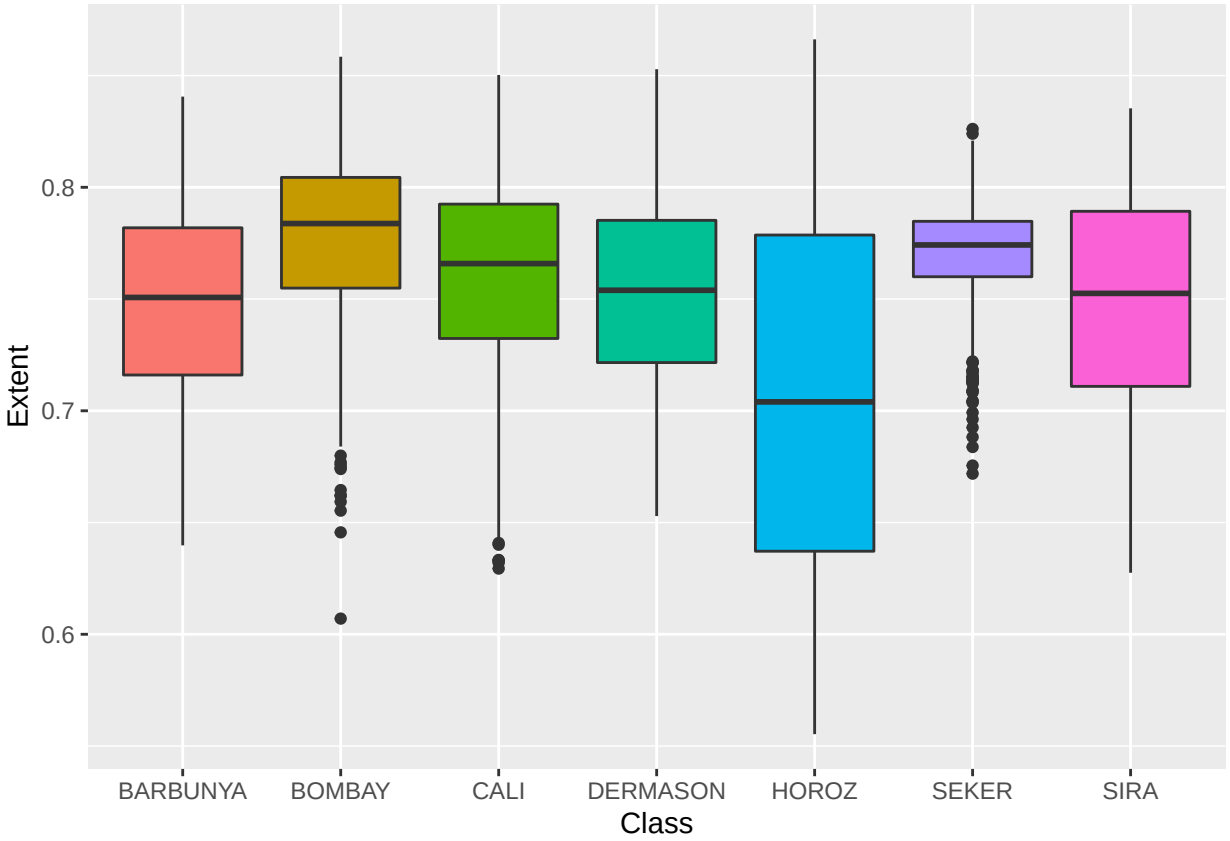

MinorAxisLength
400
300
200
BARBUNYA
BOMBAY
CALI
DERMASON
HOROZ
SEKER
SIRA
Class

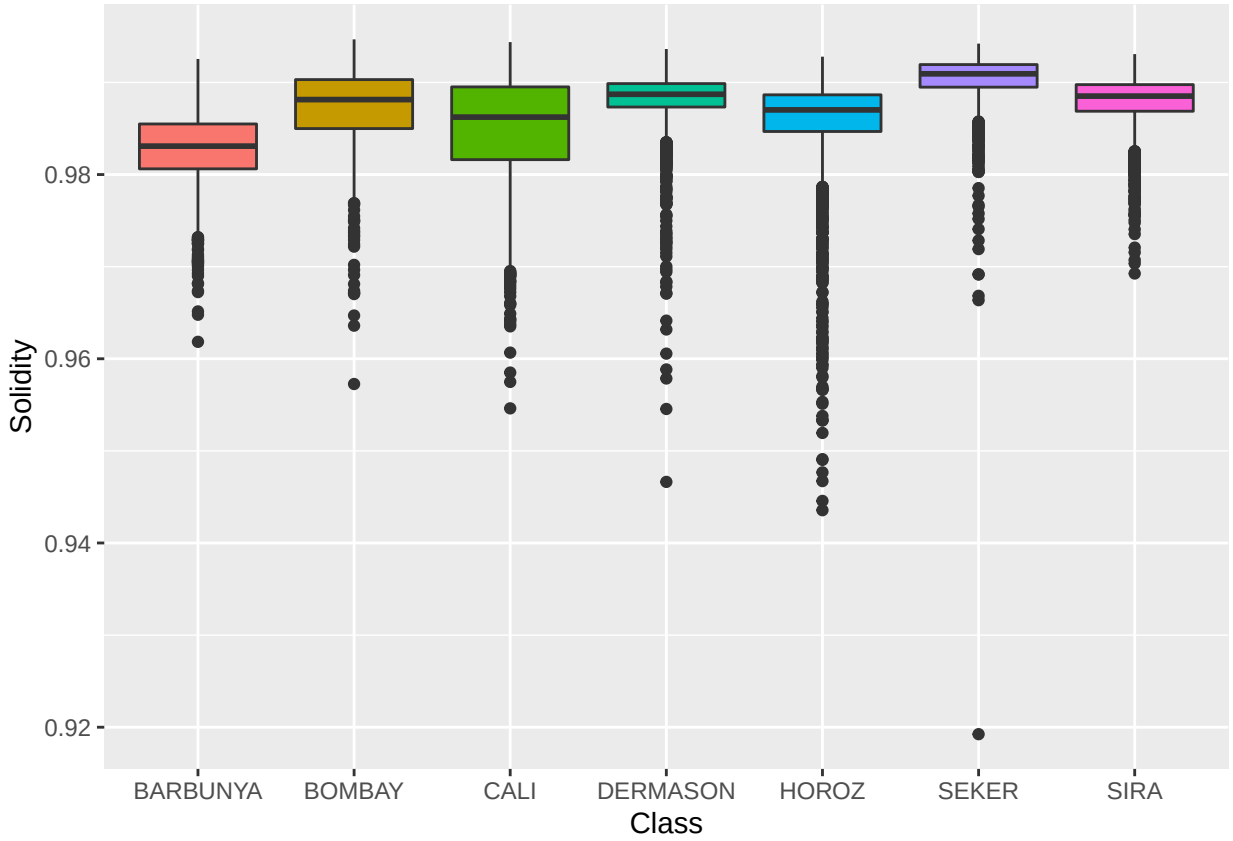

2.5
2.0
1.5
1.0
AspectRation
BARBUNYA
BOMBAY
CALI
DERMASON
HOROZ
SEKER
SIRA
Class

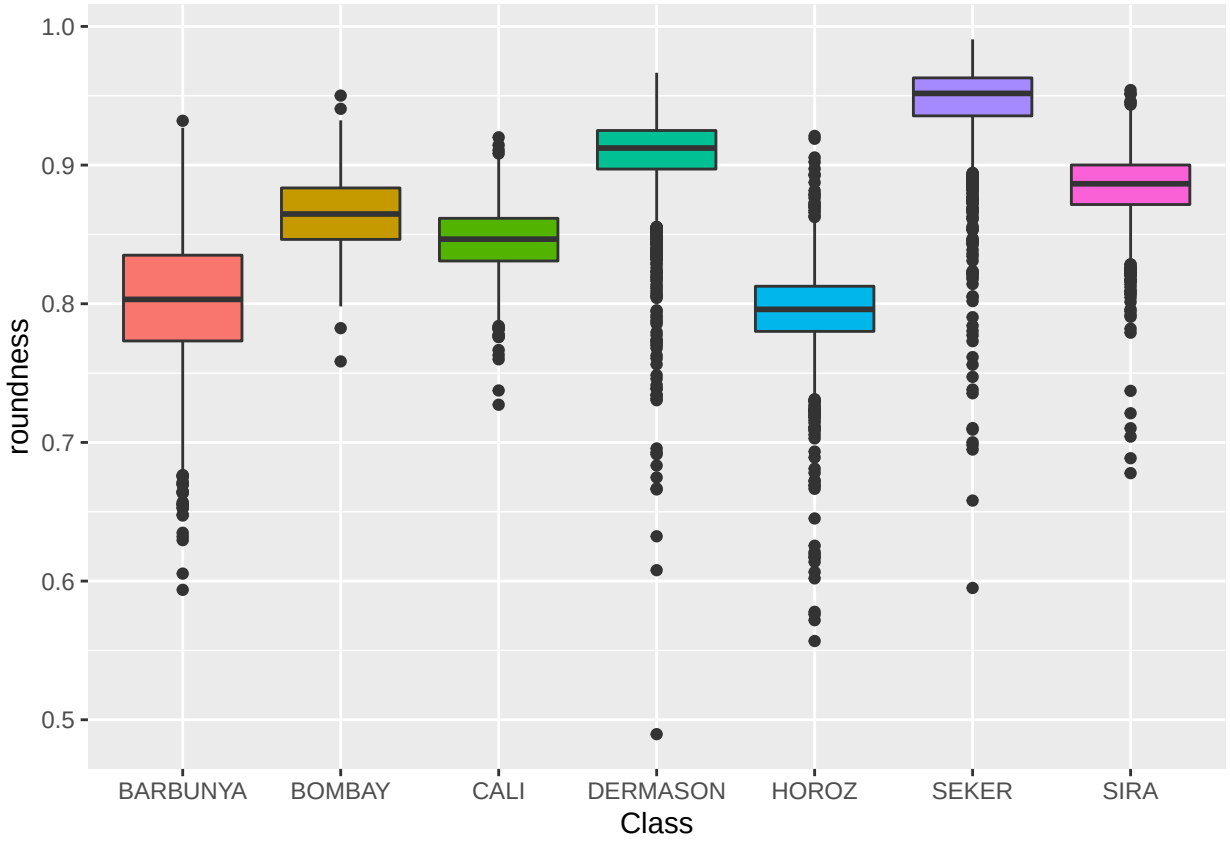

Eccentricity
0.8
0.6
0.4
0.2
BARBUNYA
BOMBAY
CALI
DERMASON
HOROZ
SEKER
SIRA
Class

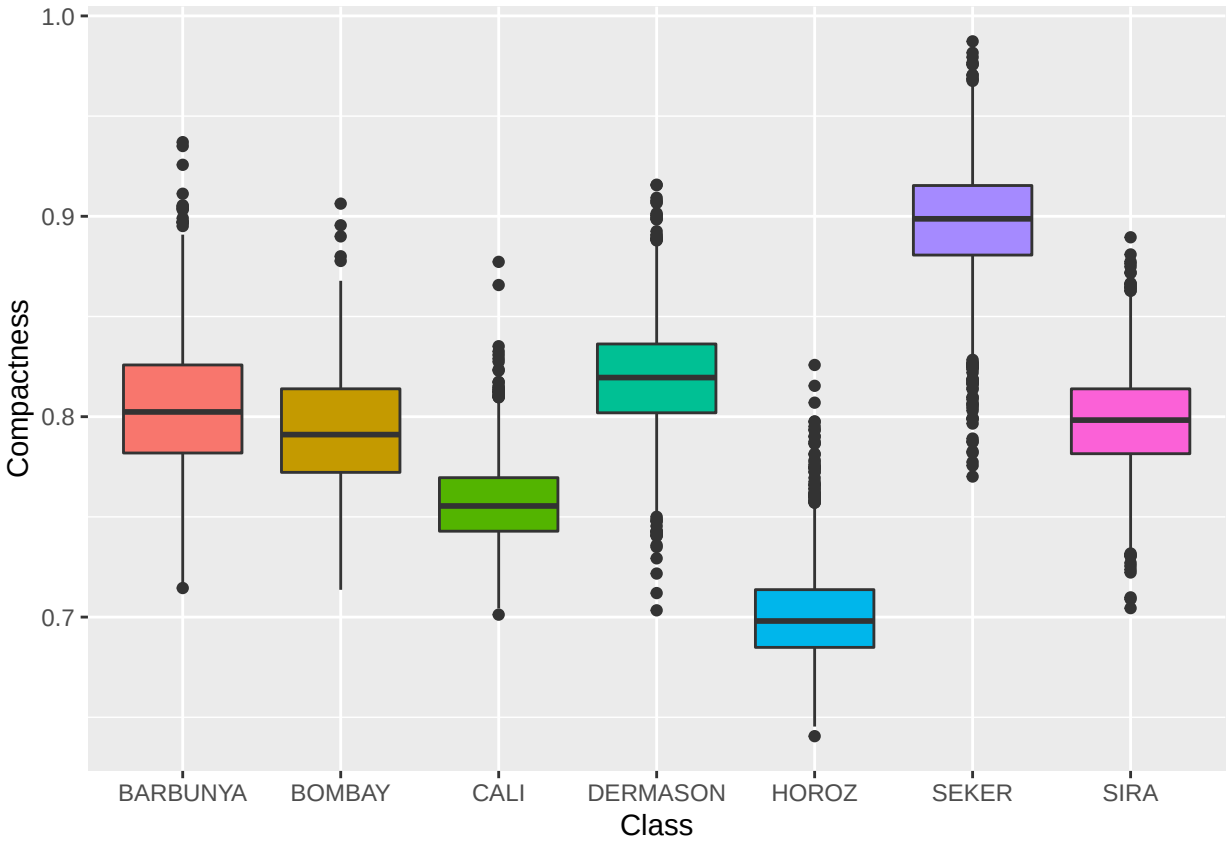

ConvexArea
2e+05
1e+05
BARBUNYA
BOMBAY
CALI
DERMASON
HOROZ
SEKER
SIRA
Class

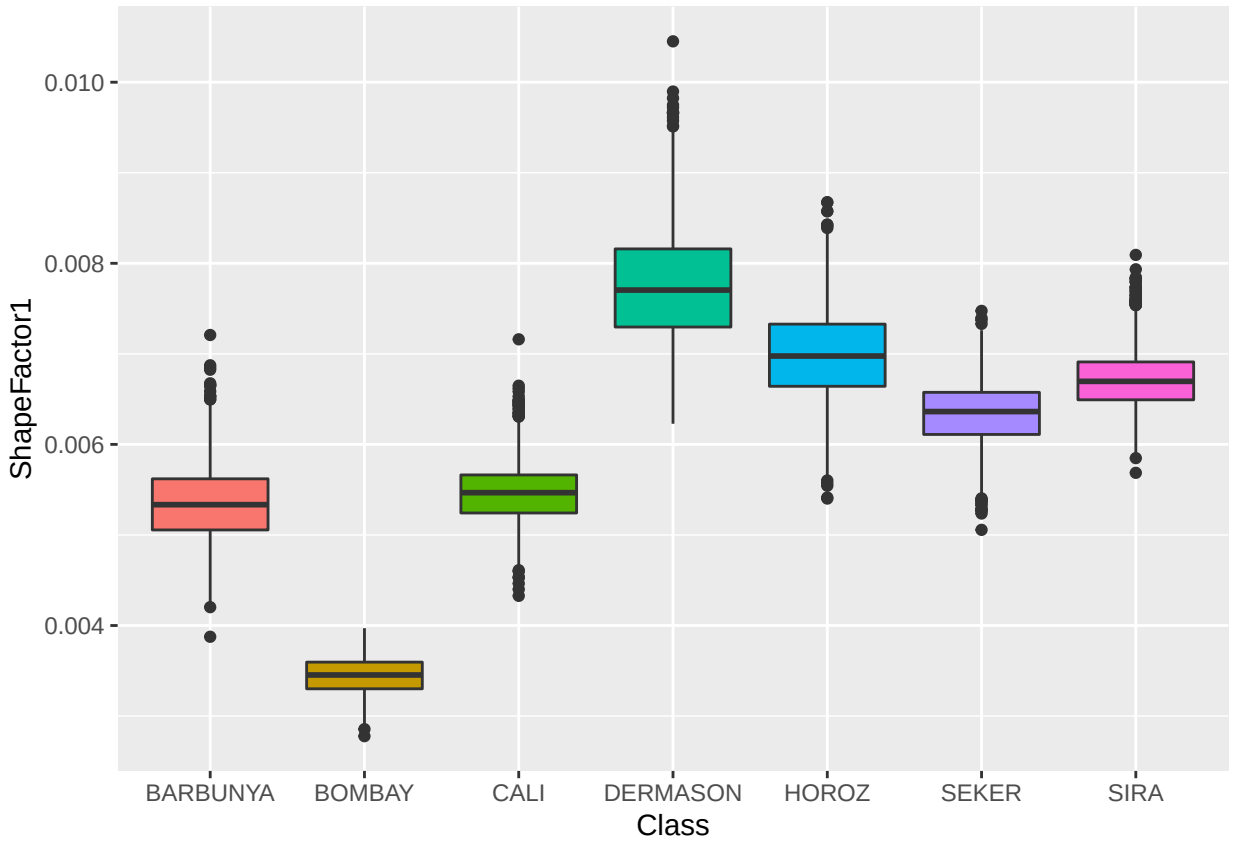

EquivDiameter
500
400
300
200
BARBUNYA
BOMBAY
CALI
DERMASON
HOROZ
SEKER
SIRA
Class

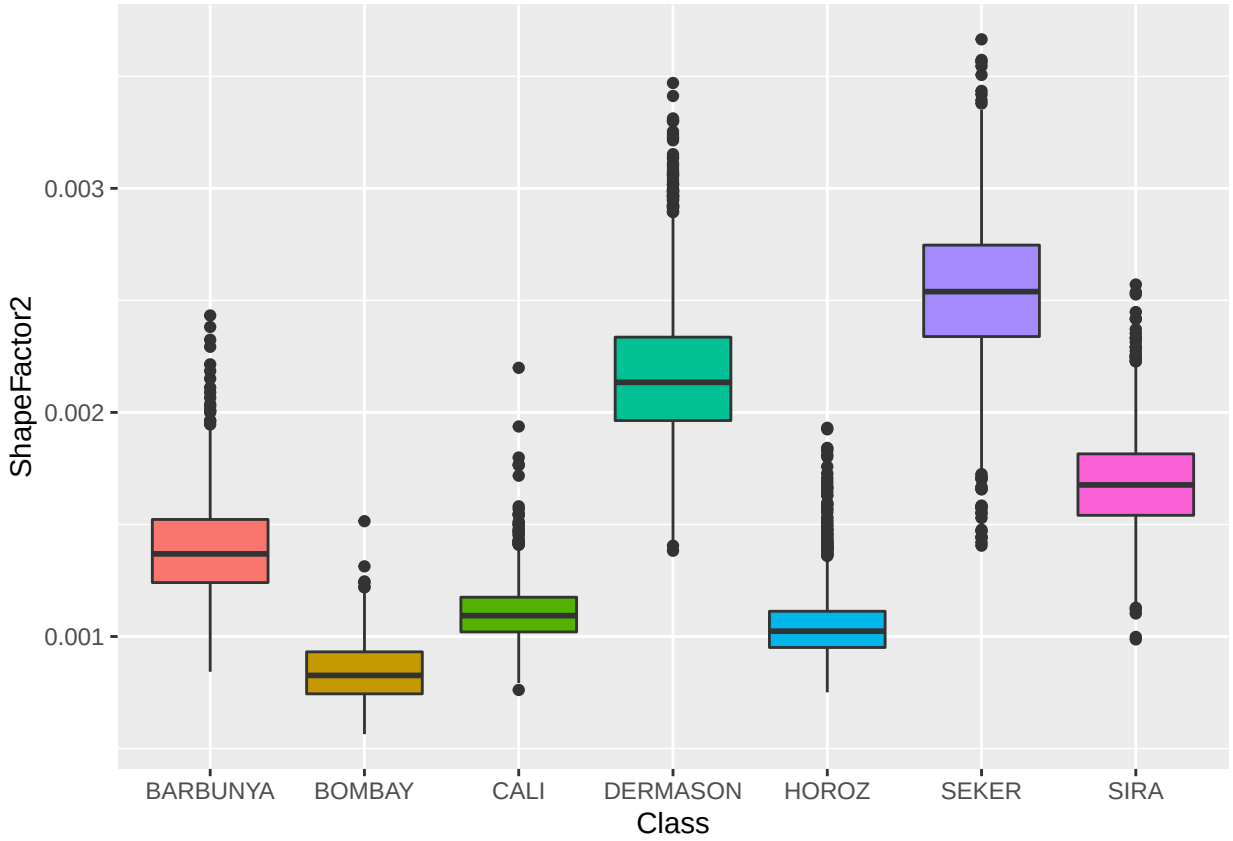

Extent
0.8
0.7
0.6
BARBUNYA
BOMBAY
CALI
DERMASON
HOROZ
SEKER
SIRA
Class

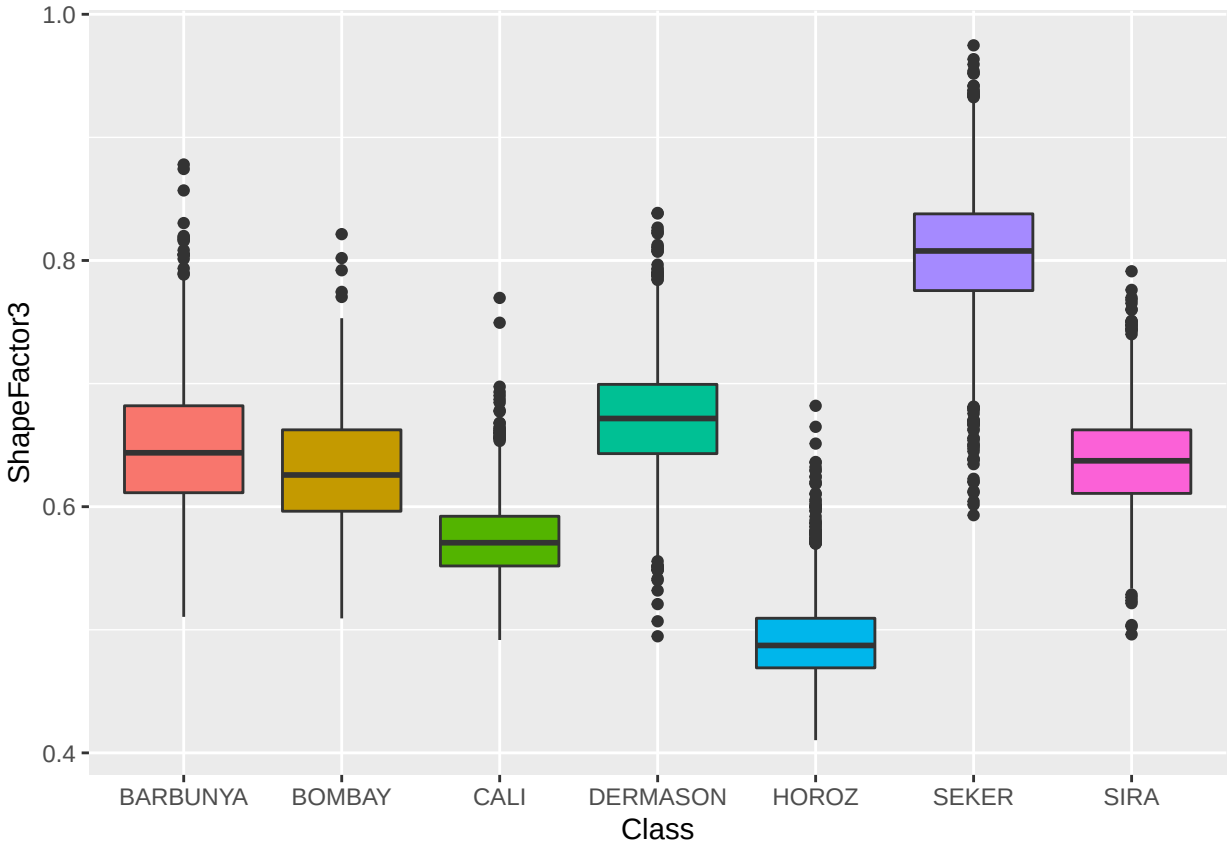

Solidity
0.98
0.96
0.94
0.92
BARBUNYA
BOMBAY
CALI
DERMASON
HOROZ
SEKER
SIRA
Class

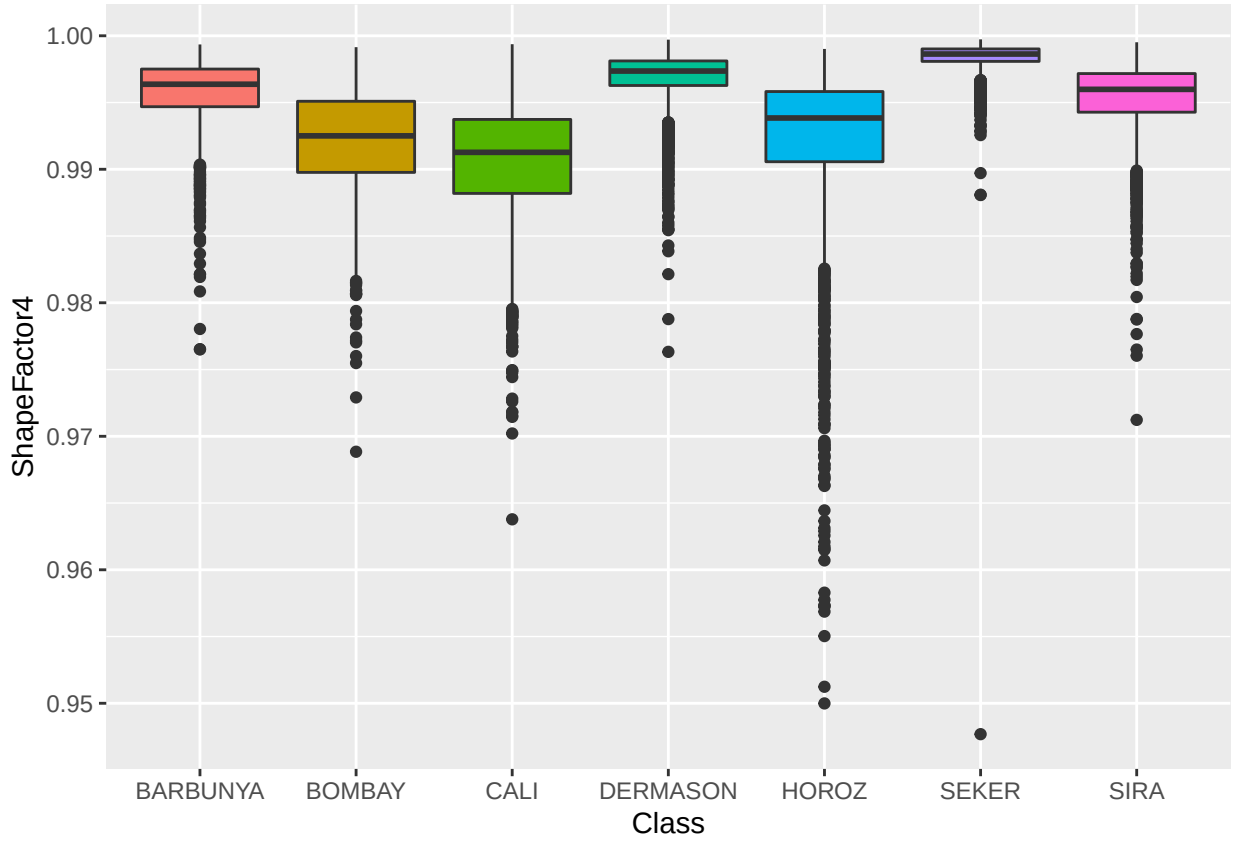

1.0
0.9
0.8
0.7
0.6
0.5
roundness
BARBUNYA
BOMBAY
CALI
DERMASON
HOROZ
SEKER
SIRA
Class


1.0
0.9
0.8
0.7
Compactness
BARBUNYA
BOMBAY
CALI
DERMASON
HOROZ
SEKER
SIRA
Class


0.010
0.008
0.006
0.004
ShapeFactor1
BARBUNYA
BOMBAY
CALI
DERMASON
HOROZ
SEKER
SIRA
Class


ShapeFactor2
0.003
0.002
0.001
BARBUNYA
BOMBAY
CALI
DERMASON
HOROZ
SEKER
SIRA
Class


1.0
0.8
0.6
0.4
ShapeFactor3
BARBUNYA
BOMBAY
CALI
DERMASON
HOROZ
SEKER
SIRA
Class



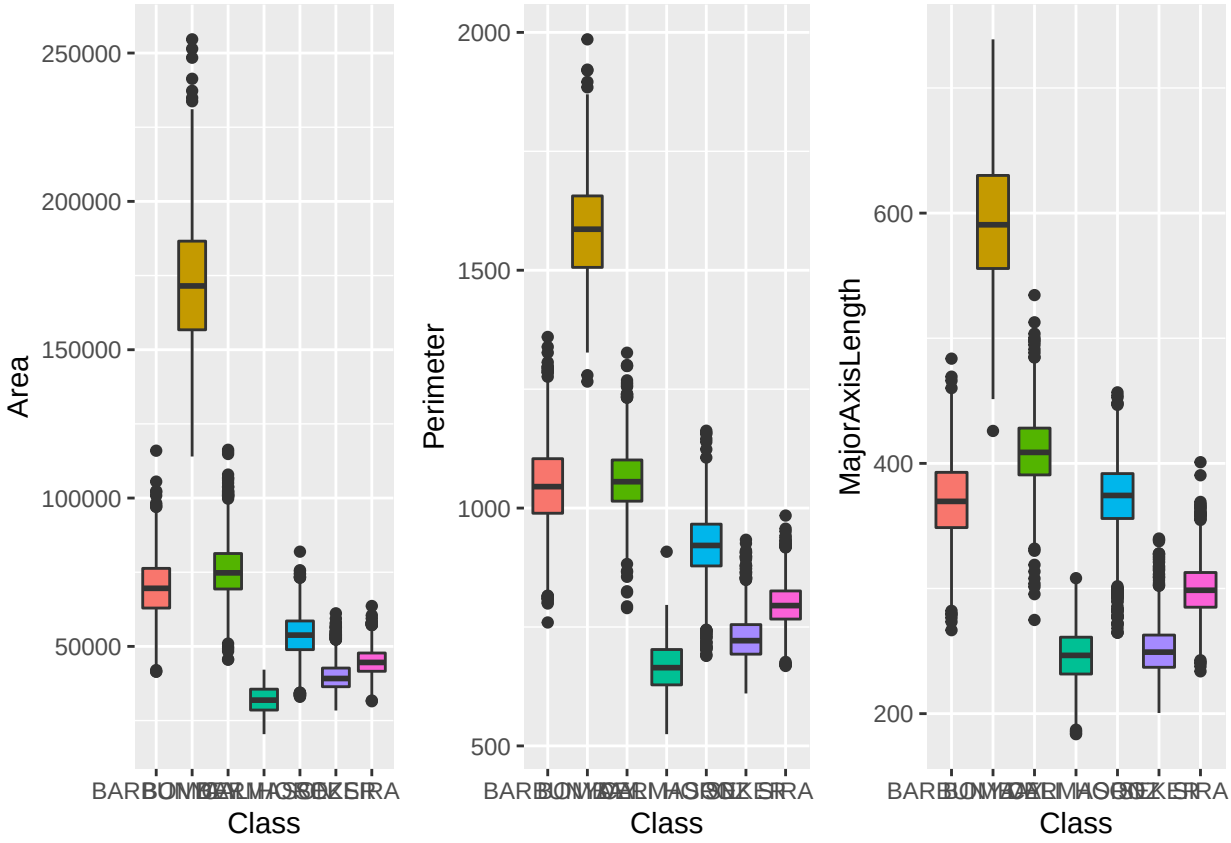

```
colnames(dados)
```

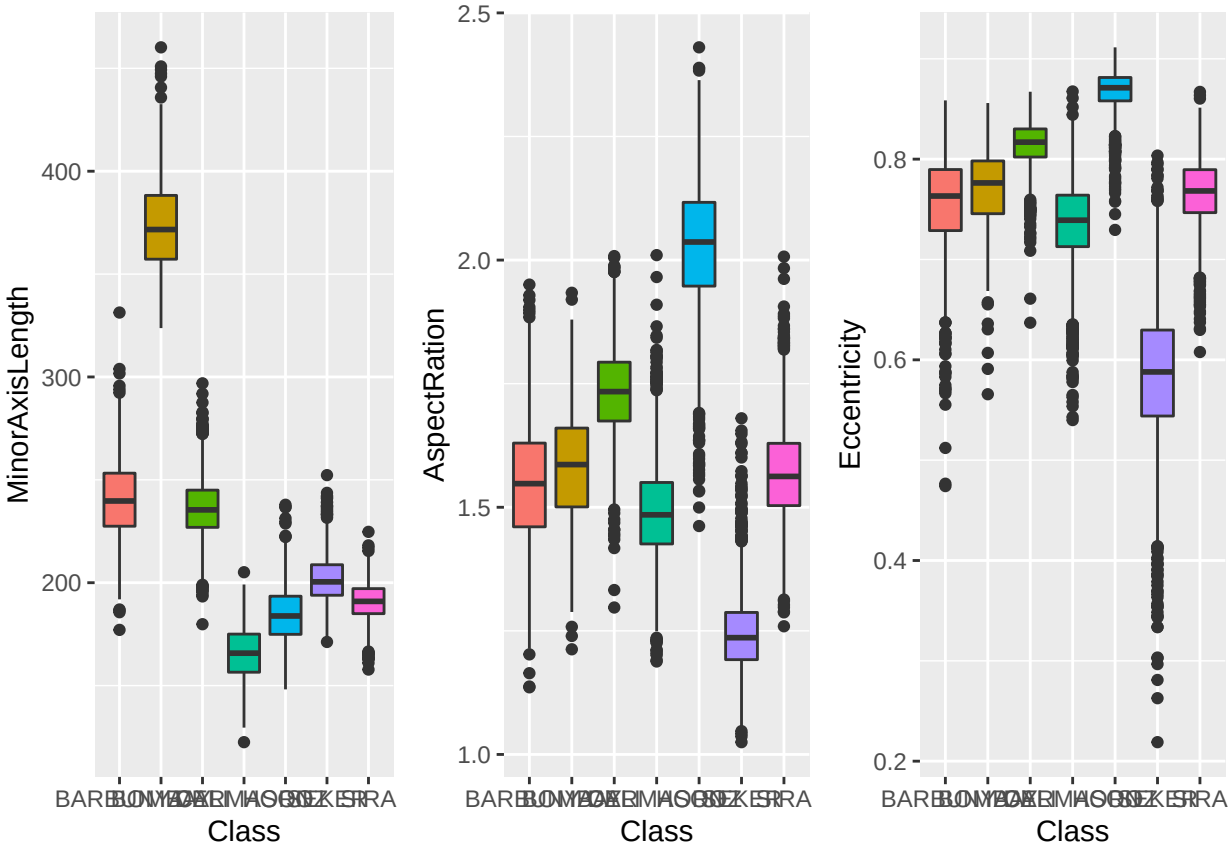
```
##  [1] "Area"             "Perimeter"        "MajorAxisLength" "MinorAxisLength"
##  [5] "AspectRation"     "Eccentricity"     "ConvexArea"      "EquivDiameter"
##  [9] "Extent"           "Solidity"         "roundness"       "Compactness"
## [13] "ShapeFactor1"     "ShapeFactor2"     "ShapeFactor3"    "ShapeFactor4"
## [17] "Class"
```

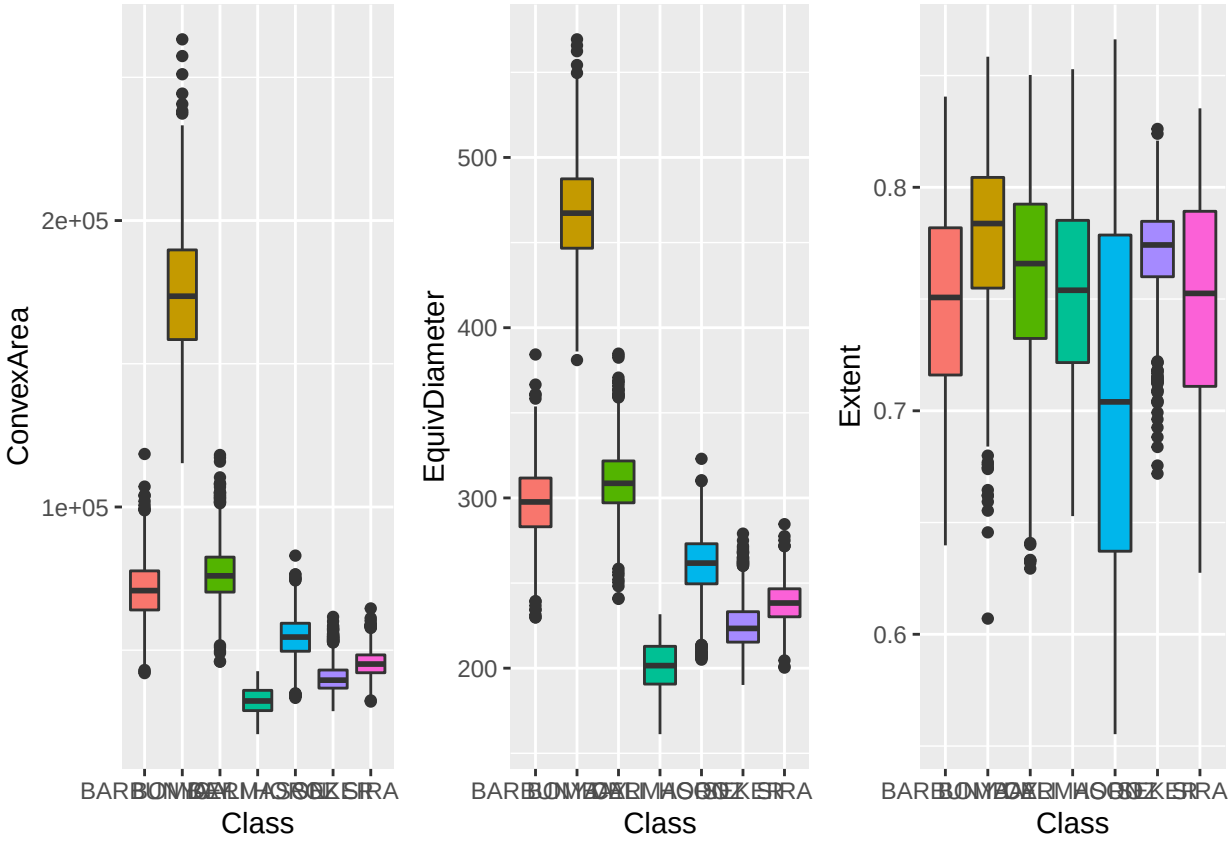
A Classe "Dermason" não apresenta outliers nas variálveis Area e EquivDiameter. Na Variável Extent as classes Barbunya, Dermason, Horoz e Sira não apresentam outliers. A classe Bombay pode ser facilmente discriminada em relação às demais classes, nas variáveis "Area", "Perimeter", "MinorAxisLength", "MajorAxisLength", "ConvexArea" e "EquiveDiameter".

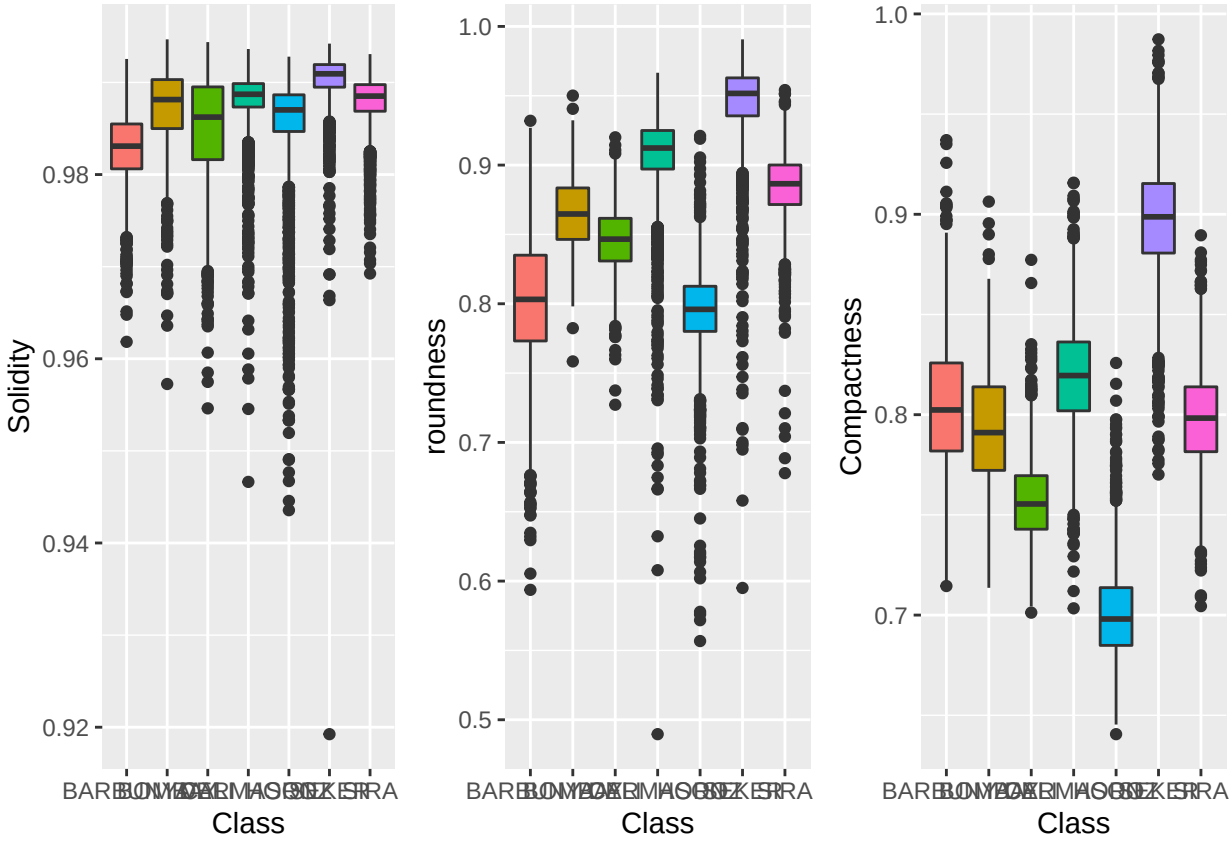
5.Utilizando um gráfico boxplot por variável x classe (organize em 3 colunas), diga qual é a variável que teria maior poder de discriminação? Justifique a sua escolha.

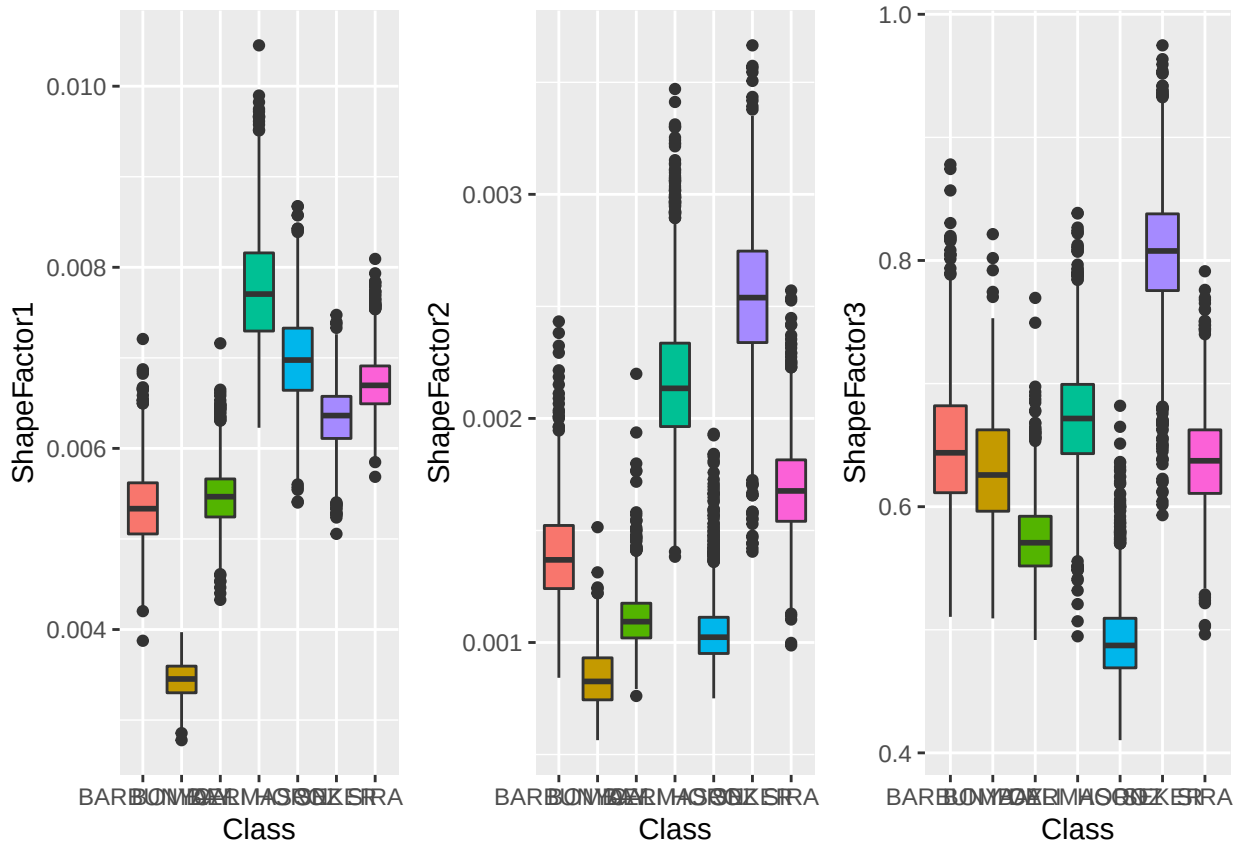
```
p = list()
for(i in 1:16){
  p[[i]] = ggplot(dados, aes_string(x="Class", y=names(dados)[i],fill="Class")) + geom_boxplot() + theme
  if((i==3) || (i==6) || (i==9) || (i==12) || (i==15)){
    do.call(grid.arrange,c(p[(i-2):i],ncol=3))
  }
}
```


Area
250000
200000
150000
100000
50000
Class
Perimeter
2000
1500
1000
500
Class
MajorAxisLength
600
400
200
Class


MinorAxisLength
400
300
200
Class
AspectRation
2.5
2.0
1.5
1.0
Class
Eccentricity
0.8
0.6
0.4
0.2
Class


ConvexArea
2e+05
1e+05
Class
EquivDiameter
500
400
300
200
Class
Extent
0.8
0.7
0.6
Class

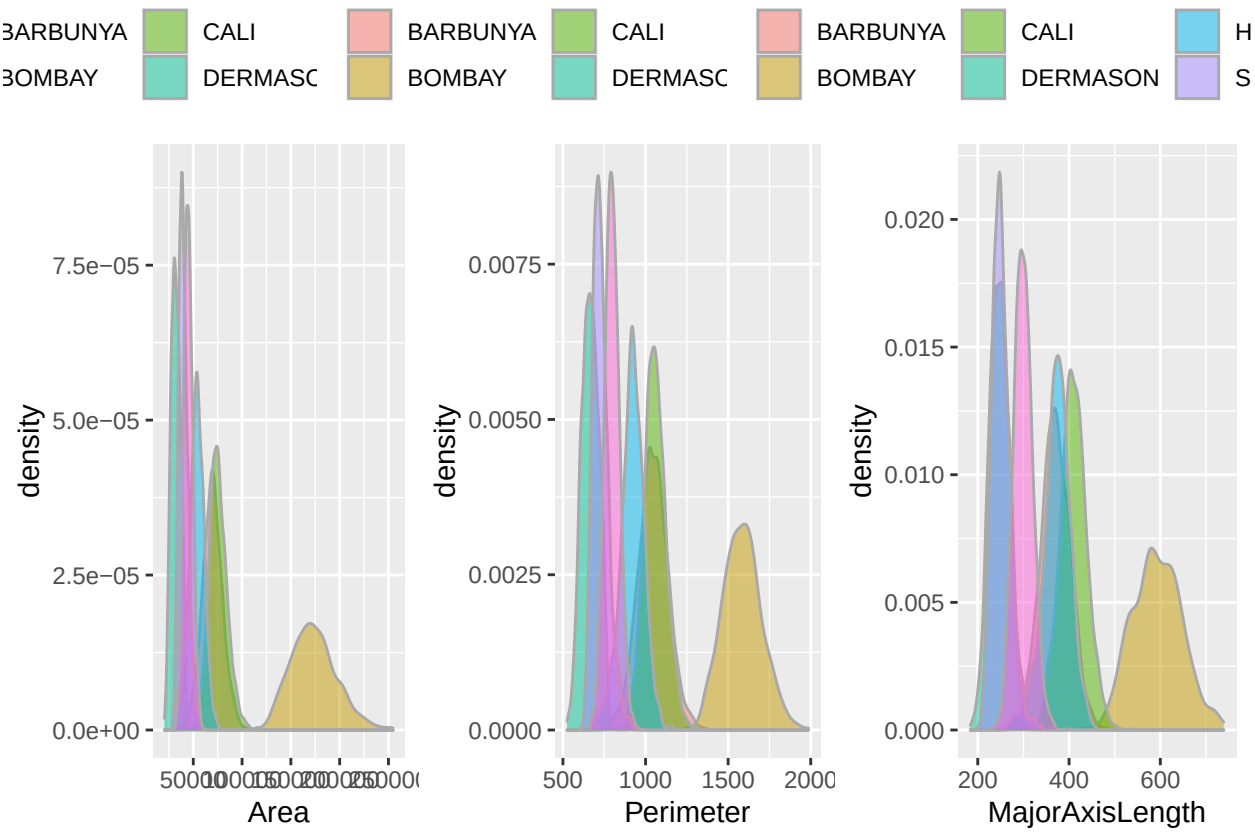

Solidity
0.98
0.96
0.94
0.92
Class
roundness
1.0
0.9
0.8
0.7
0.6
0.5
Class
Compactness
1.0
0.9
0.8
0.7
Class

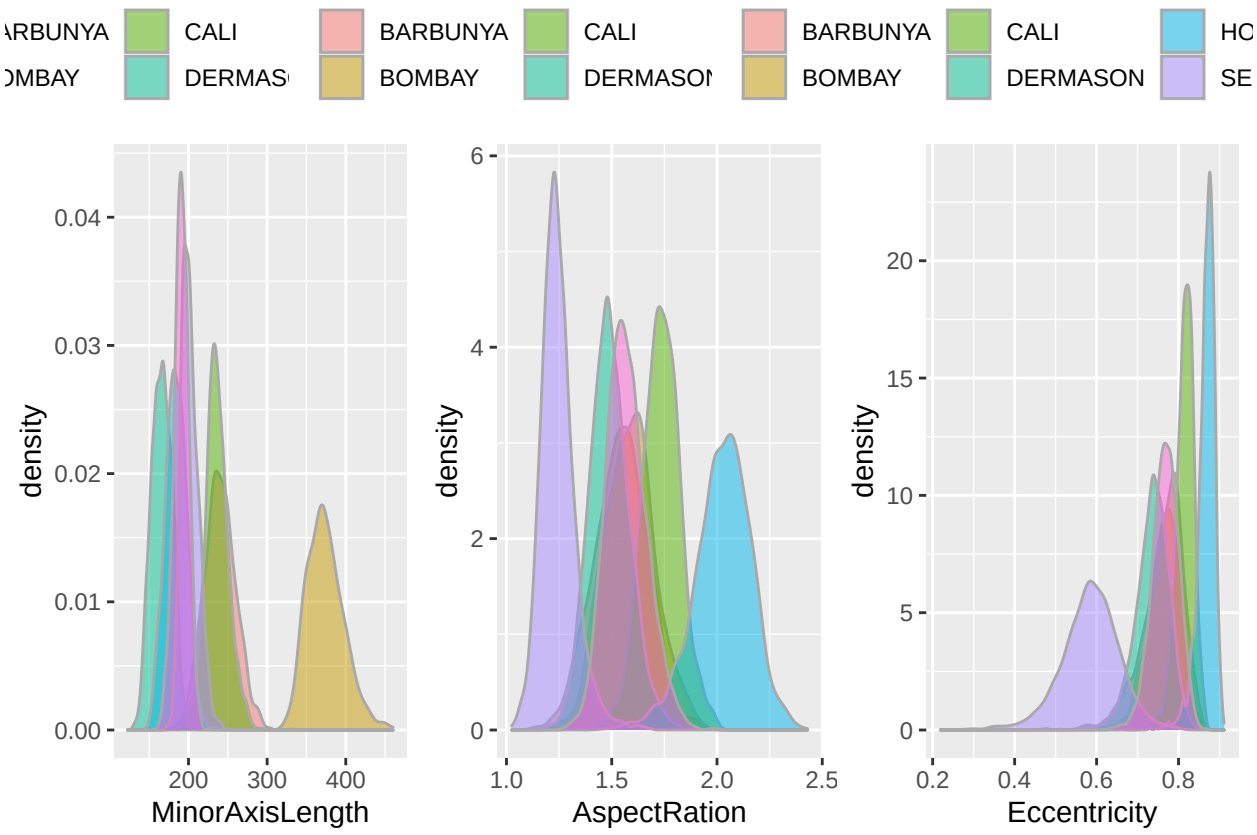


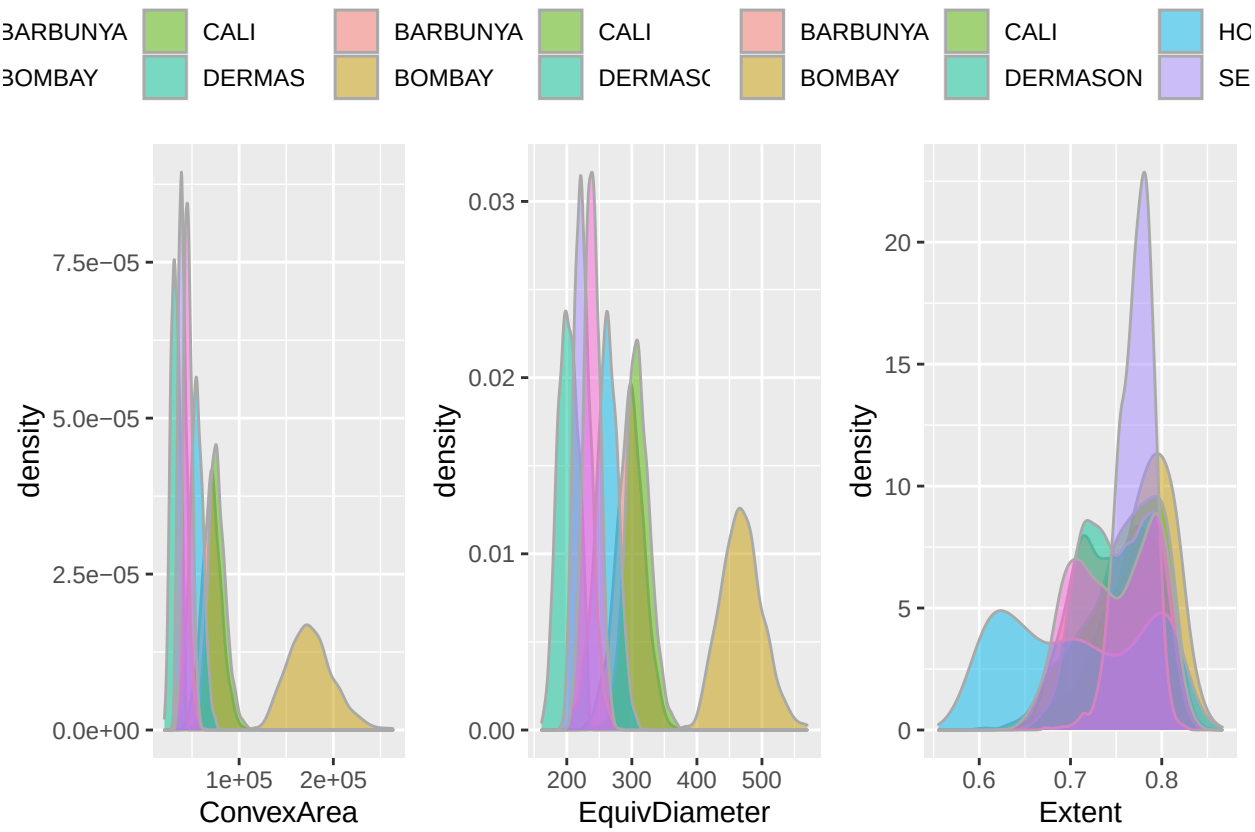
Algumas das variáves que poderia ter mais poder de discriminação seria: "Eccentricity", "ConvexArea", "ShapedFactor1" e "ShapedFactor2". Os principais pontos pelo qual estas variáveis se destacam são seus pontos mínimos e máximos, serem proximos e com poucos, ou nenhum, outlier, como também os pontos são diferentes entre as classes.

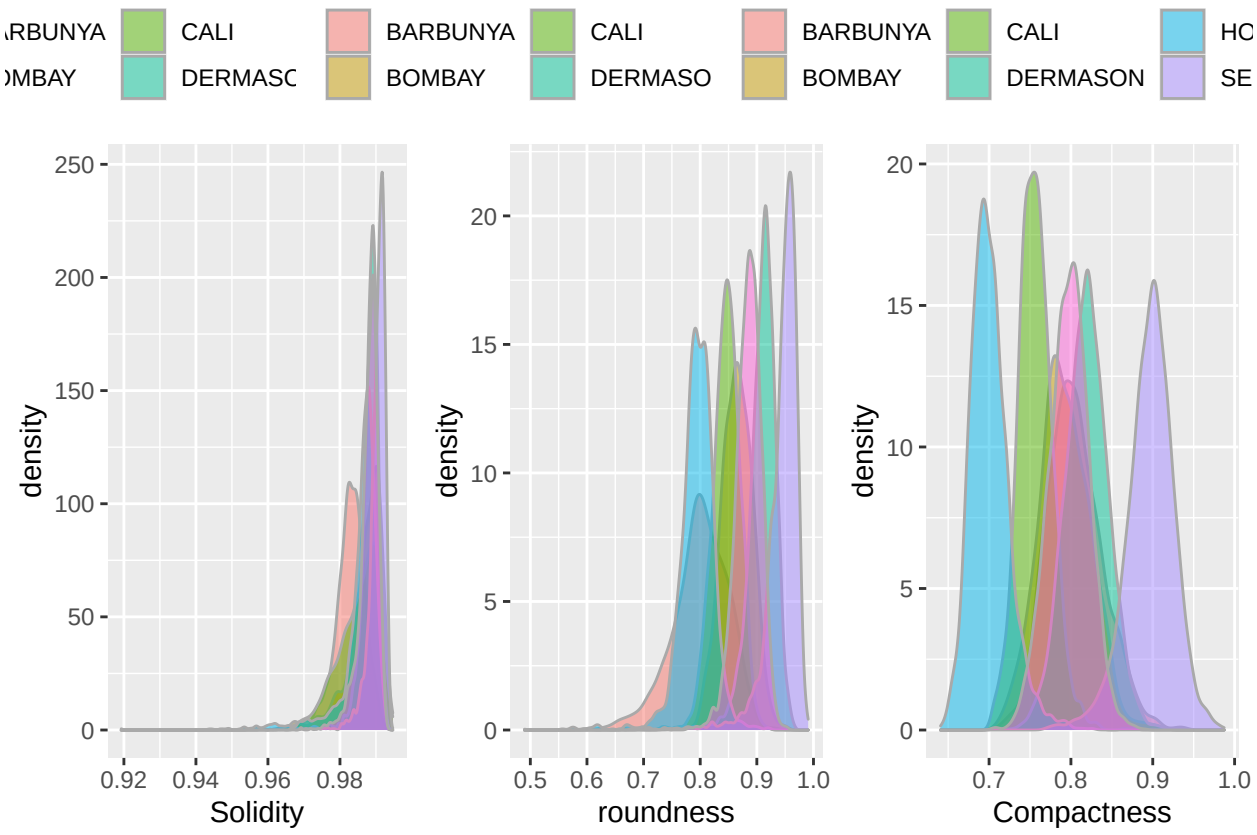
6.Utilizando gráficos de densidade por variável (organize em 3 colunas), é possível fazer alguma afirmação sobre a discriminabilidade de alguma classe? Pode utilizar os boxplots gerados na etapa anterior para auxiliar nas conclusões.

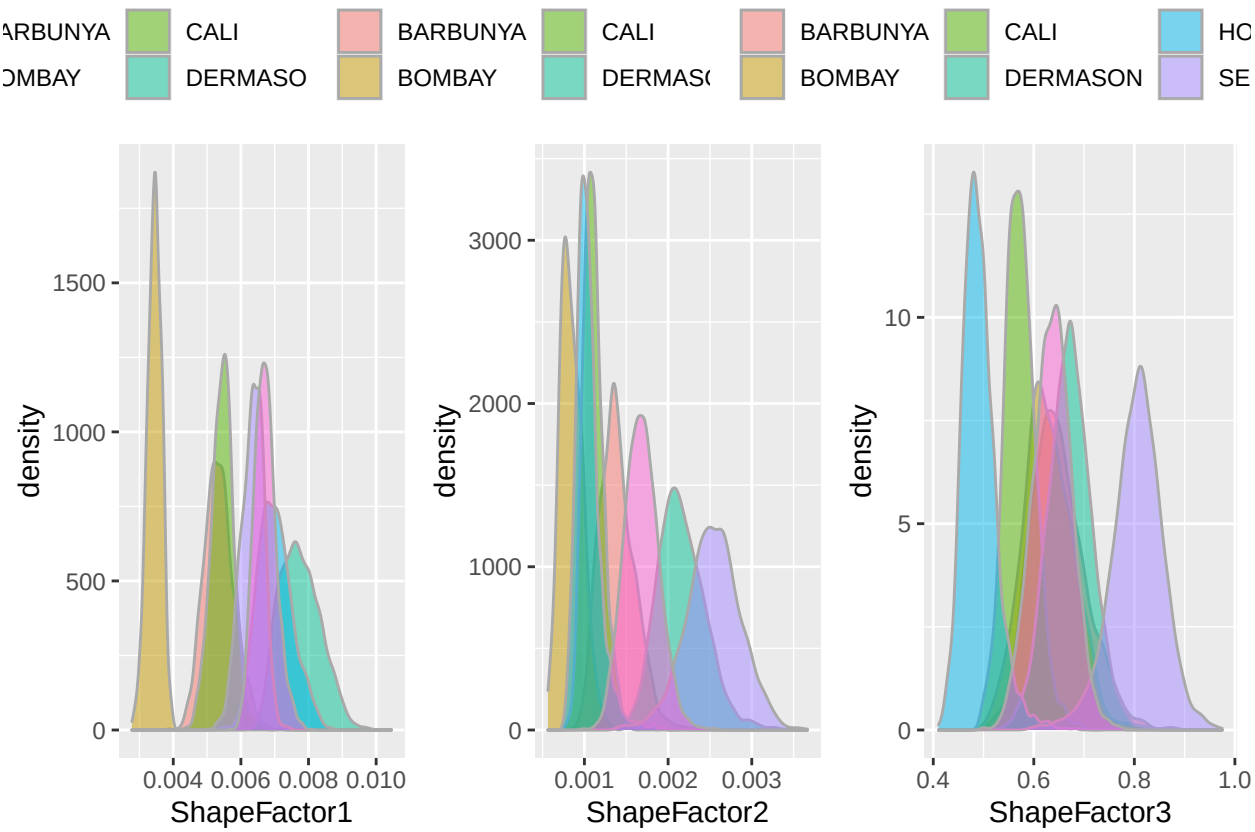
```
p = list()
for(i in 1:16){
   p[[i]] = ggplot(dados, aes_string(x=names(dados)[i],fill="Class")) +
            geom_density(alpha=0.5,color="darkgray") +
            theme(legend.position="top", legend.title = element_blank())
   if((i==3) || (i==6) || (i==9) || (i==12) || (i==15)){
    do.call(grid.arrange,c(p[(i-2):i],ncol=3))
  }
}
```


BARBUNYA
BOMBAY
CALI
DERMASC
BARBUNYA
BOMBAY
CALI
DERMASC
BARBUNYA
BOMBAY
CALI
DERMASON
H
S
density
7.5e−05
5.0e−05
2.5e−05
0.0e+00
Area
density
0.0075
0.0050
0.0025
0.0000
500
1000
1500
2000
Perimeter
density
0.020
0.015
0.010
0.005
0.000
200
400
600
MajorAxisLength


ARBUNYA
OMBAY
CALI
DERMAS
BARBUNYA
BOMBAY
CALI
DERMASON
BARBUNYA
BOMBAY
CALI
DERMASON
HO
SE
density
0.04
0.03
0.02
0.01
0.00
200
300
400
MinorAxisLength
density
6
4
2
0
1.0
1.5
2.0
2.5
AspectRation
density
20
15
10
5
0
0.2
0.4
0.6
0.8
Eccentricity


BARBUNYA
BOMBAY
CALI
DERMAS
BARBUNYA
BOMBAY
CALI
DERMASC
BARBUNYA
BOMBAY
CALI
DERMASON
HC
SE
density
7.5e−05
5.0e−05
2.5e−05
0.0e+00
1e+05
2e+05
ConvexArea
density
0.03
0.02
0.01
0.00
200
300
400
500
EquivDiameter
density
20
15
10
5
0
0.6
0.7
0.8
Extent


ARBUNYA
OMBAY
CALI
DERMASC
BARBUNYA
BOMBAY
CALI
DERMASO
BARBUNYA
BOMBAY
CALI
DERMASON
HC
SE
density
Solidity
roundness
Compactness

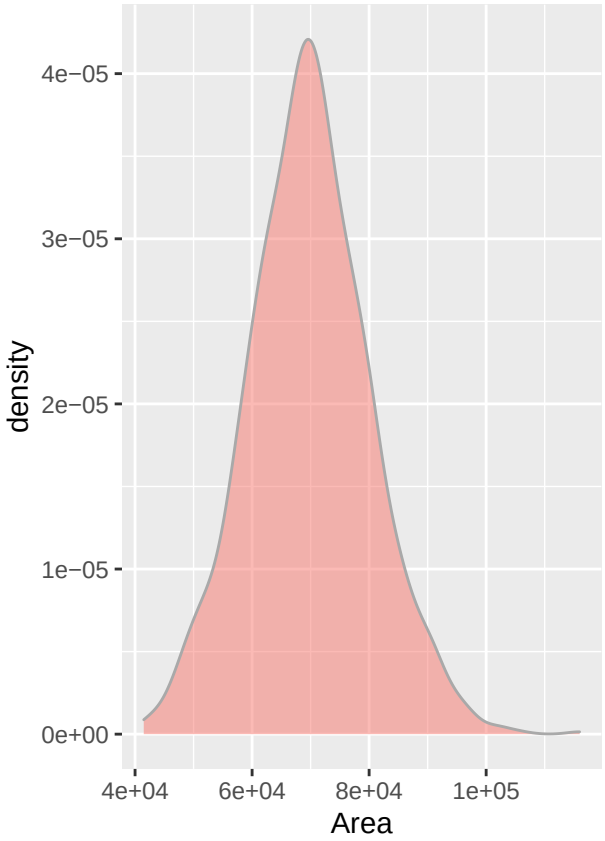


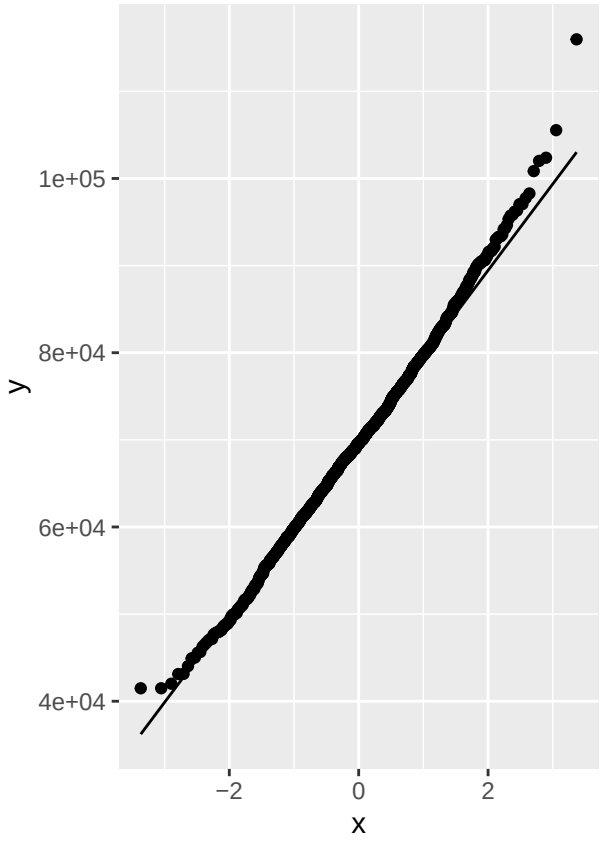
As Classes Bombay e Seker possuem grande discrepancia em algumas Variáveis.

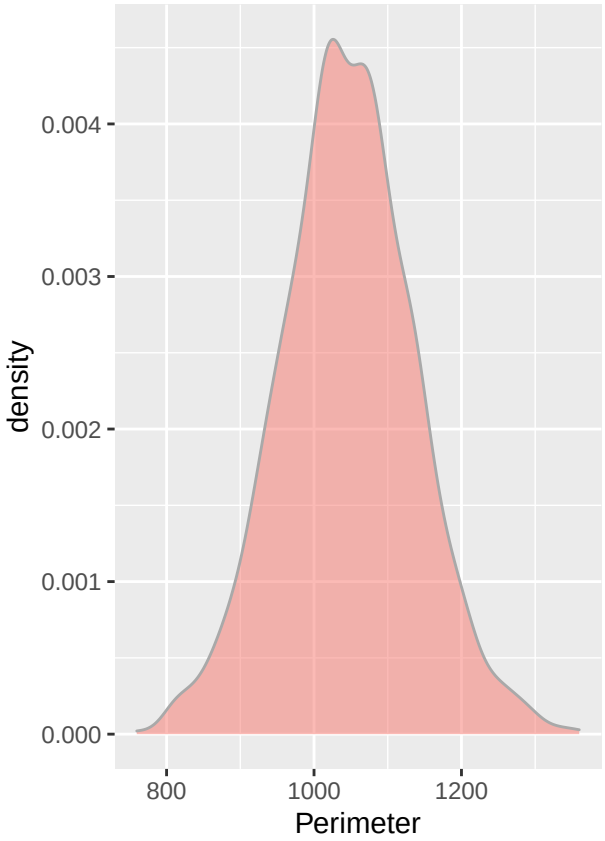
7.Algumas variável (por classe) possui uma distribuição normal (curva do sino)? É possível verificar numericamente se é verdade?

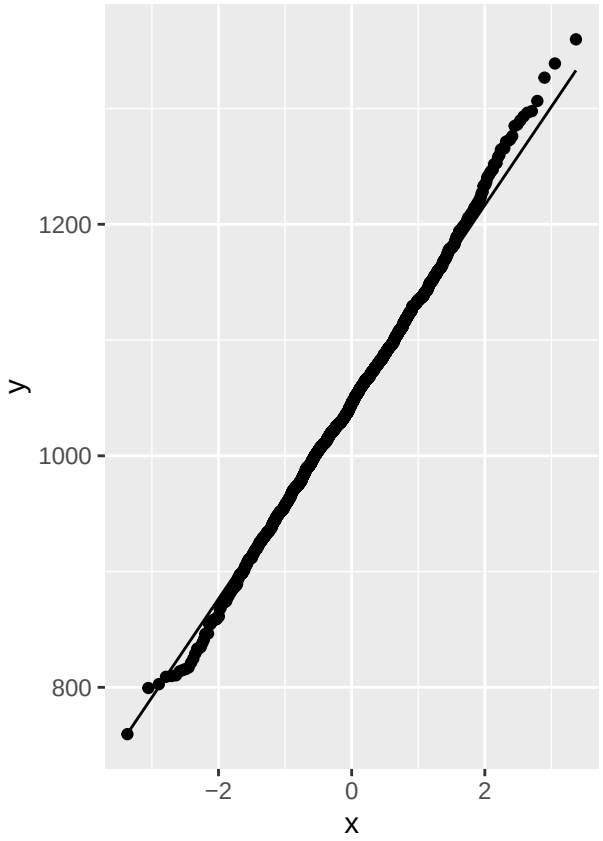
```
leves = list("BARBUNYA", "BOMBAY", "CALI", "DERMASON", "HOROZ", "SEKER", "SIRA")
p = list()
for(j in 1:7){
  print(leves[j])
  for(i in 1:16){
    p[[(i-1)*2+1]] = ggplot(dados[dados$Class==leves[j],], aes_string(x=names(dados)[i],fill="Class")) +
    p[[i*2]] = ggplot(dados[dados$Class==leves[j],], aes_string(sample=names(dados)[i]),color="Class") +
    if((i%%2)==0){
      do.call(grid.arrange,c(p[(i-1):i],ncol=2))
    }
  }
}
```

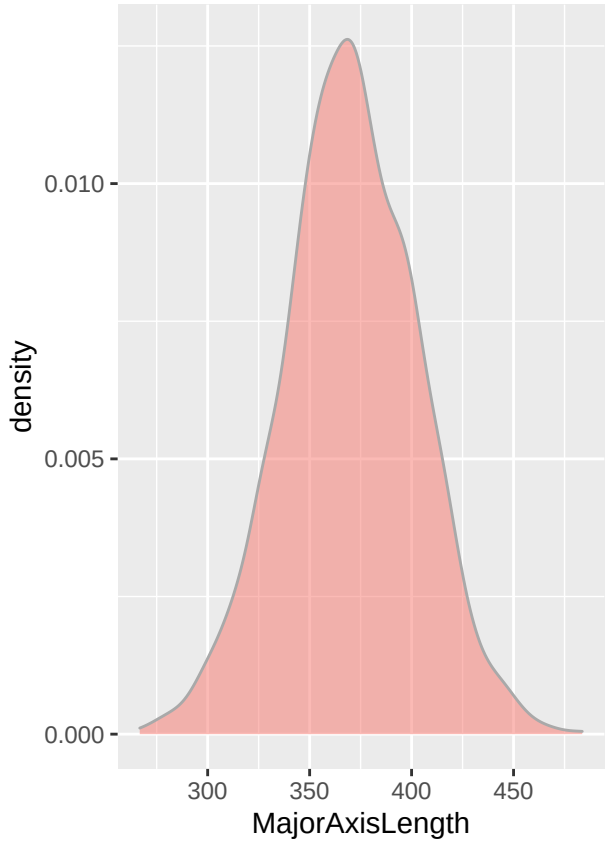
```
## [[1]]
## [1] "BARBUNYA"
```

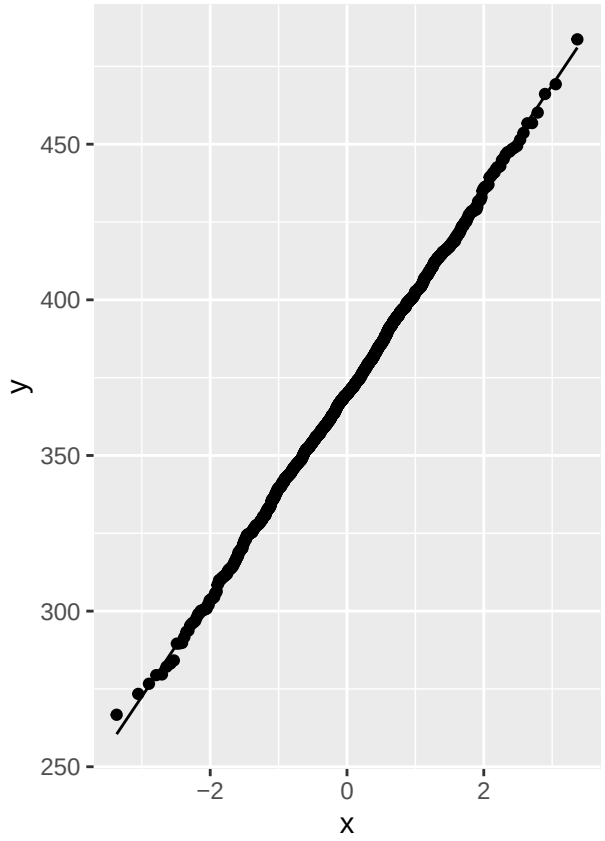

4e−05
3e−05
2e−05
1e−05
0e+00
density
4e+04
6e+04
8e+04
1e+05
Area

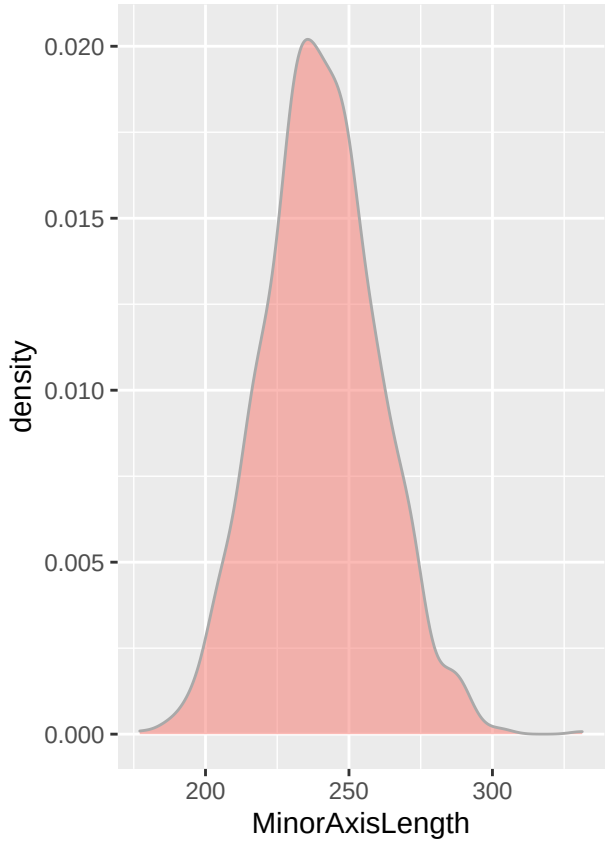

1e+05
8e+04
6e+04
4e+04
y
−2
0
2
x

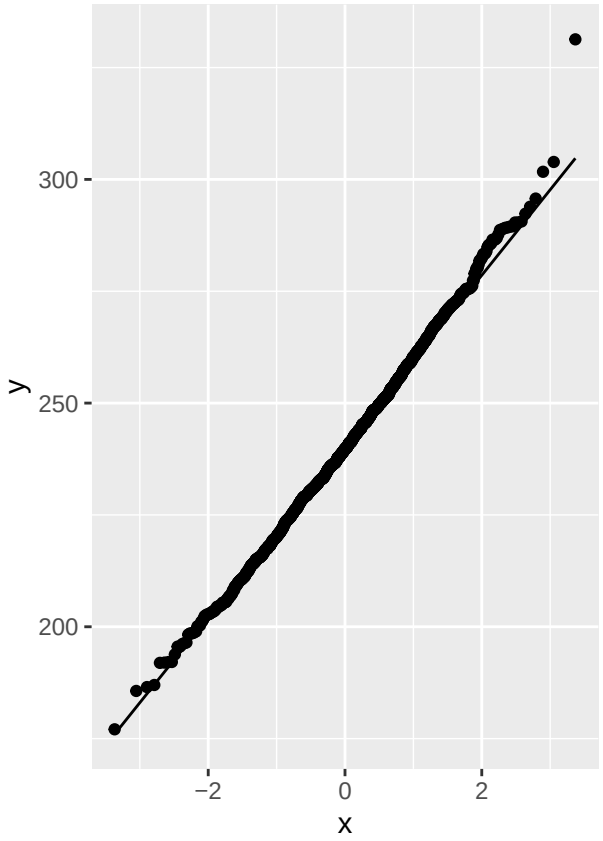

0.004
0.003
0.002
0.001
0.000
density
800
1000
1200
Perimeter

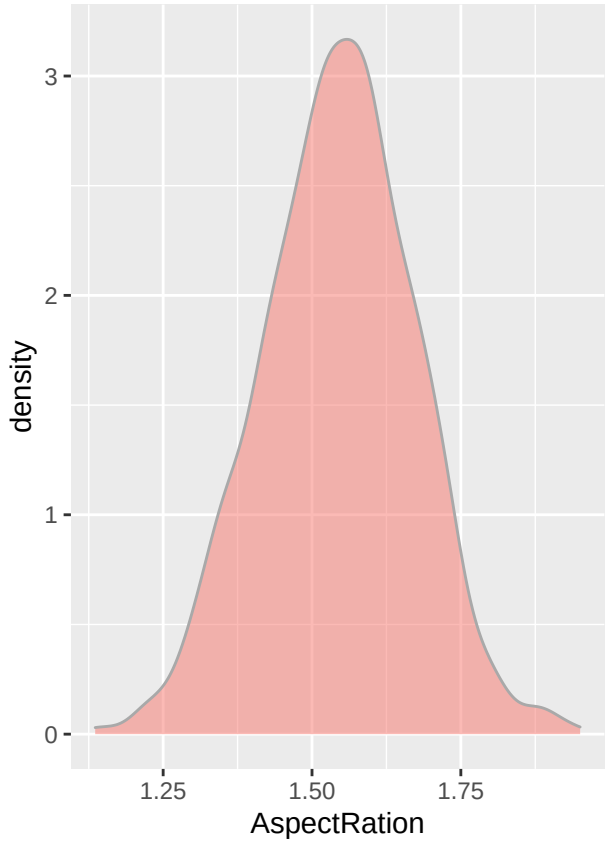

1200
1000
800
y
−2
0
2
x

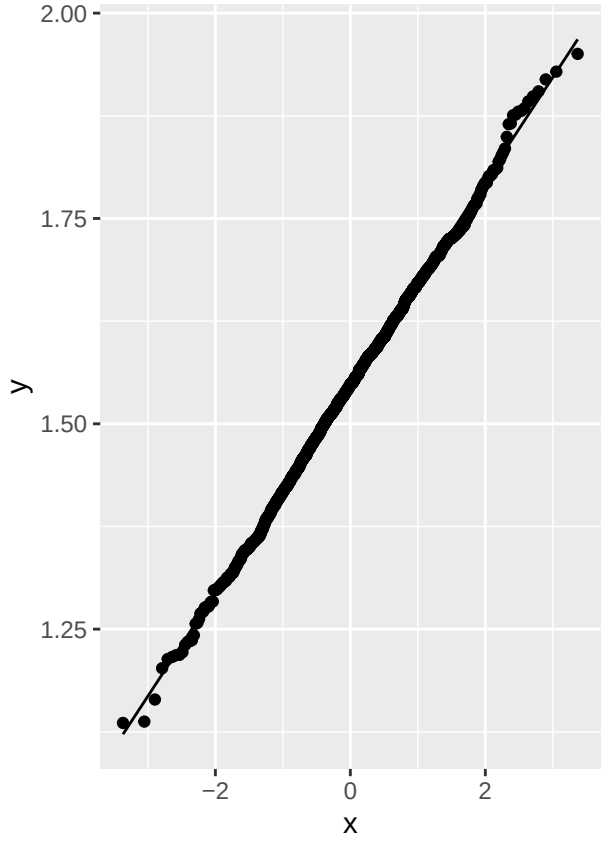

density
0.010
0.005
0.000
300
350
400
450
MajorAxisLength

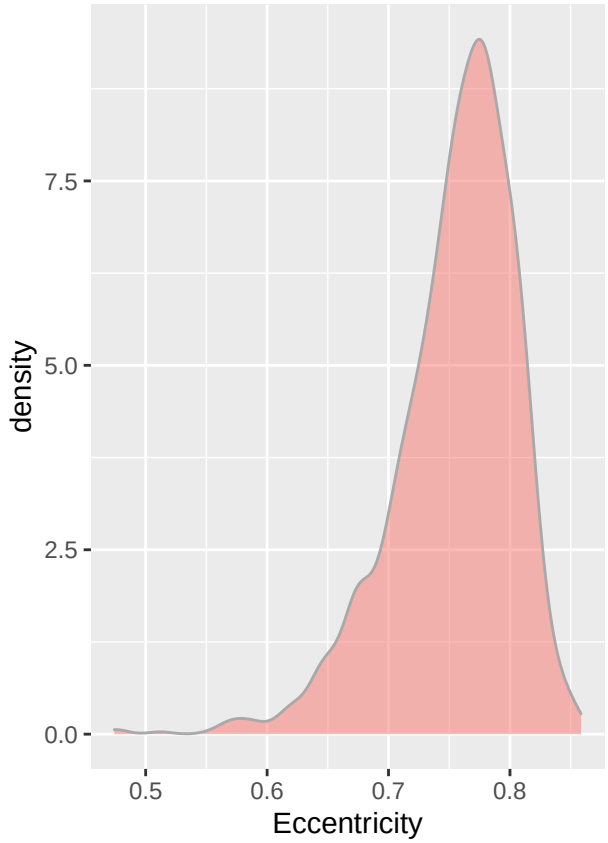

y
450
400
350
300
250
−2
0
2
x

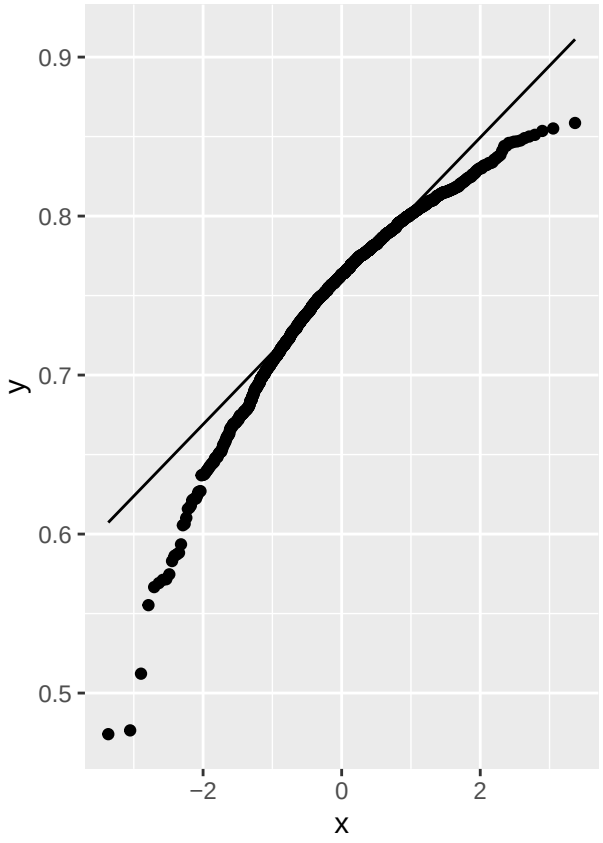

0.020
0.015
0.010
0.005
0.000
density
200
250
300
MinorAxisLength

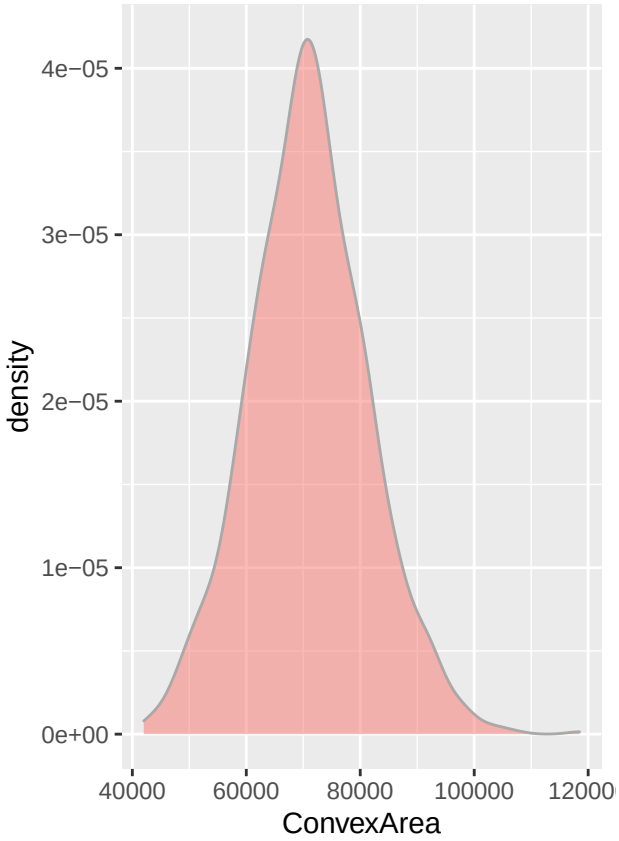

300
250
200
y
−2
0
2
x

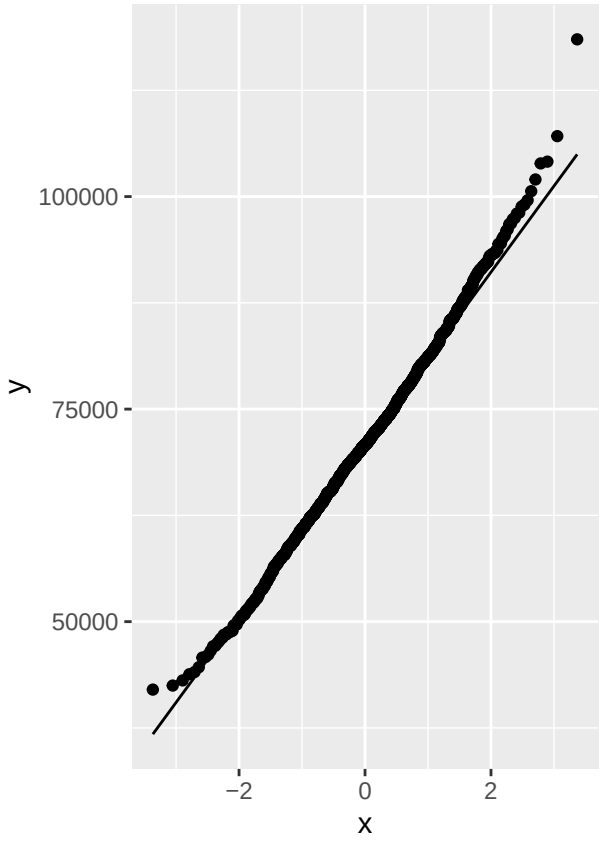

density
3
2
1
0
1.25
1.50
1.75
AspectRation

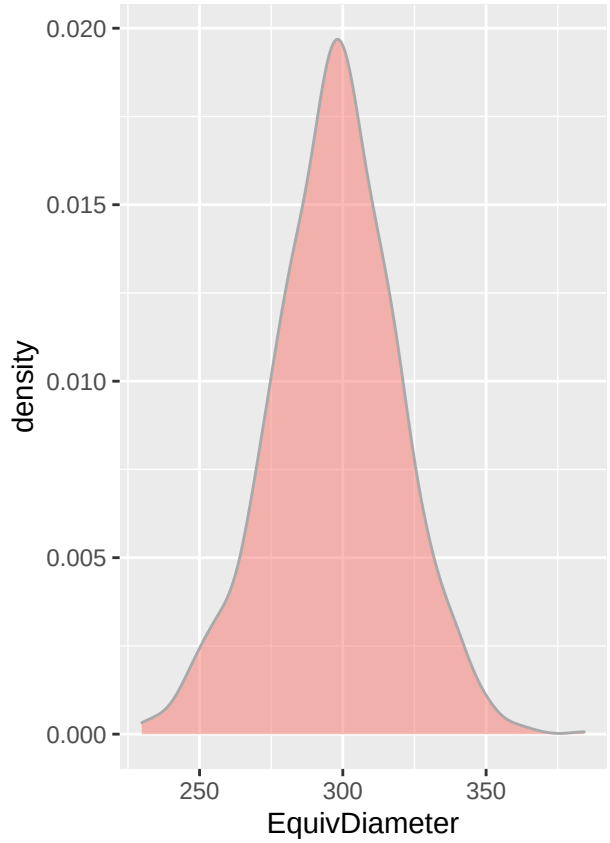

y
2.00
1.75
1.50
1.25
−2
0
2
x

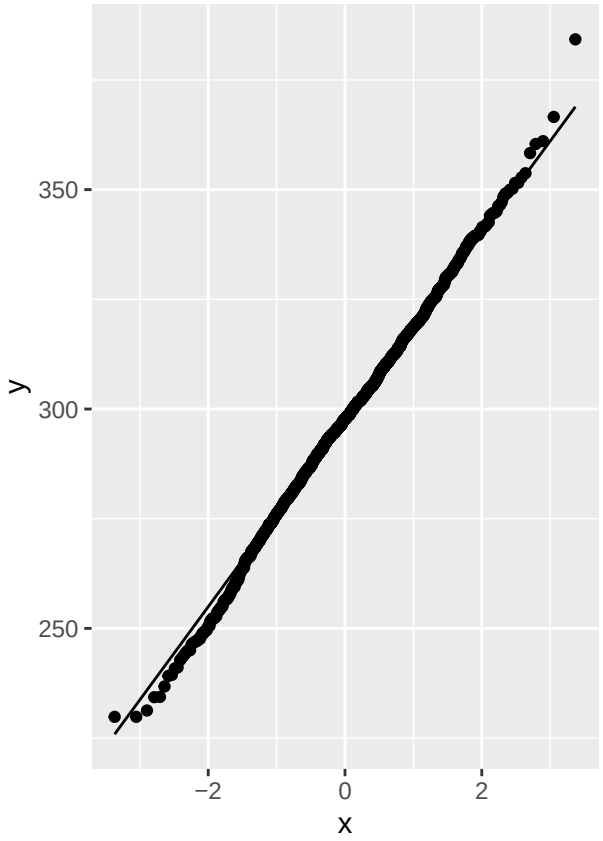

density
7.5
5.0
2.5
0.0
0.5
0.6
0.7
0.8
Eccentricity


y
0.9
0.8
0.7
0.6
0.5
−2
0
2
x

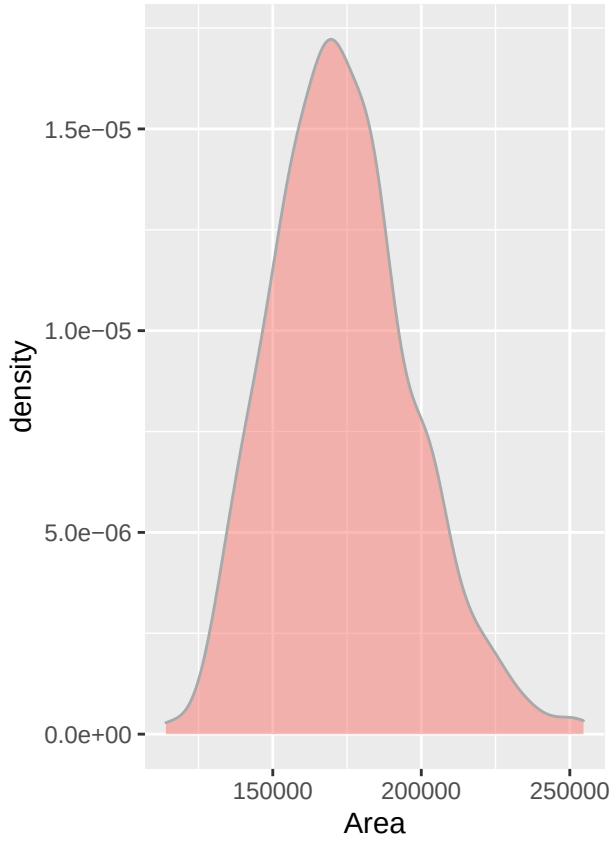

4e−05
3e−05
2e−05
1e−05
0e+00
density
40000
60000
80000
100000
12000
ConvexArea

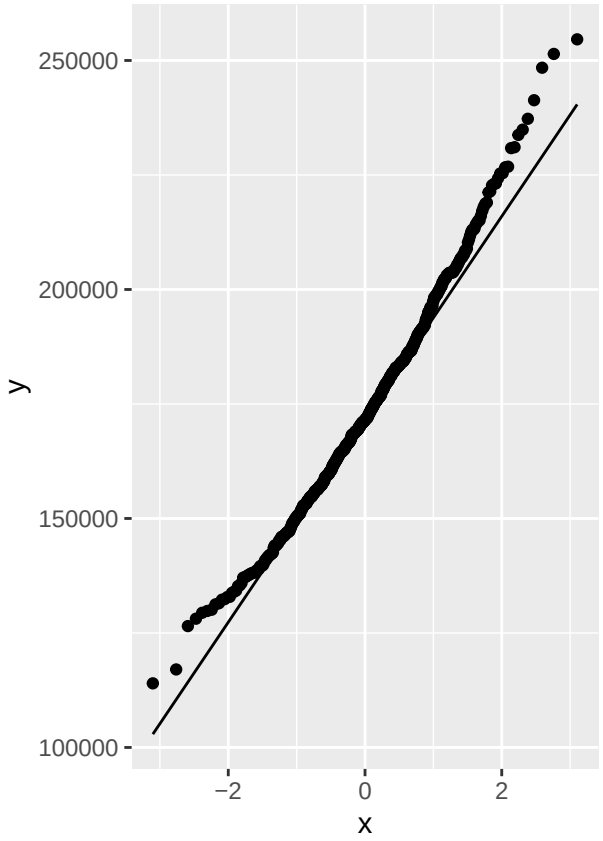

100000
75000
50000
y
−2
0
2
x

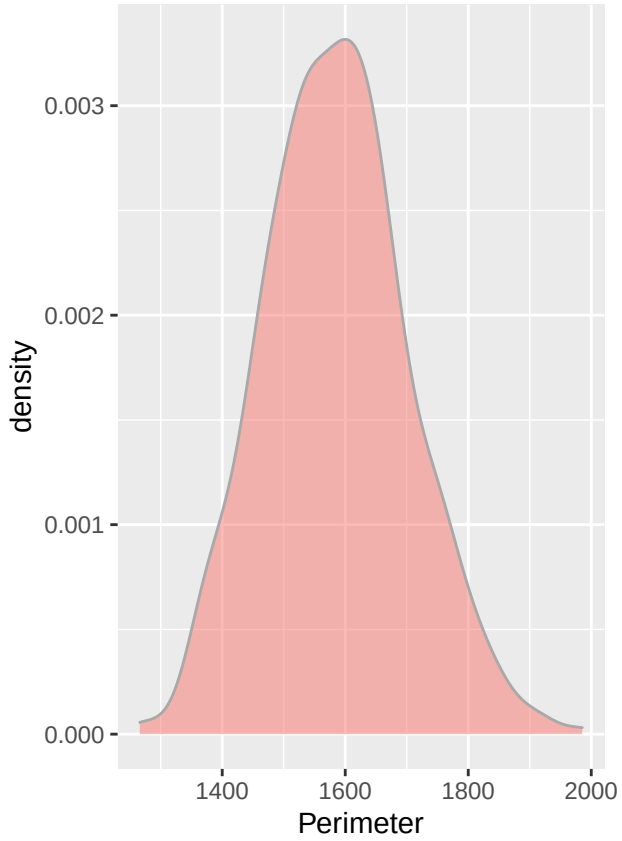

0.020
0.015
0.010
0.005
0.000
density
250
300
350
EquivDiameter

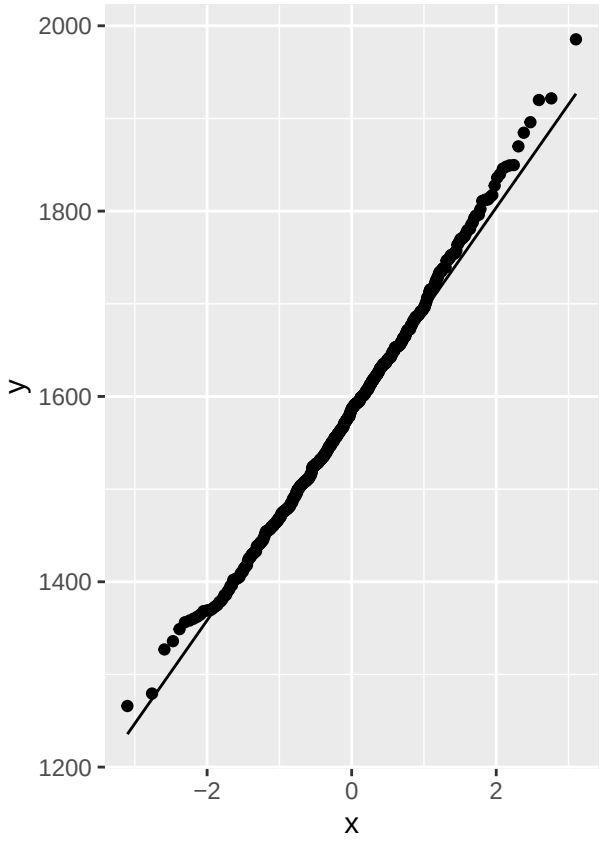

350
300
250
y
−2
0
2
x

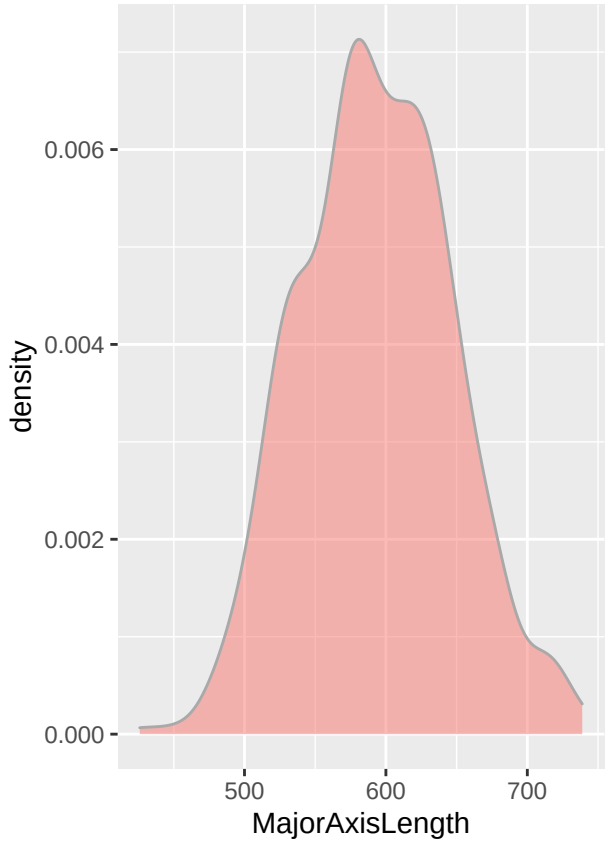
```
## [[1]]
## [1] "BOMBAY"
```

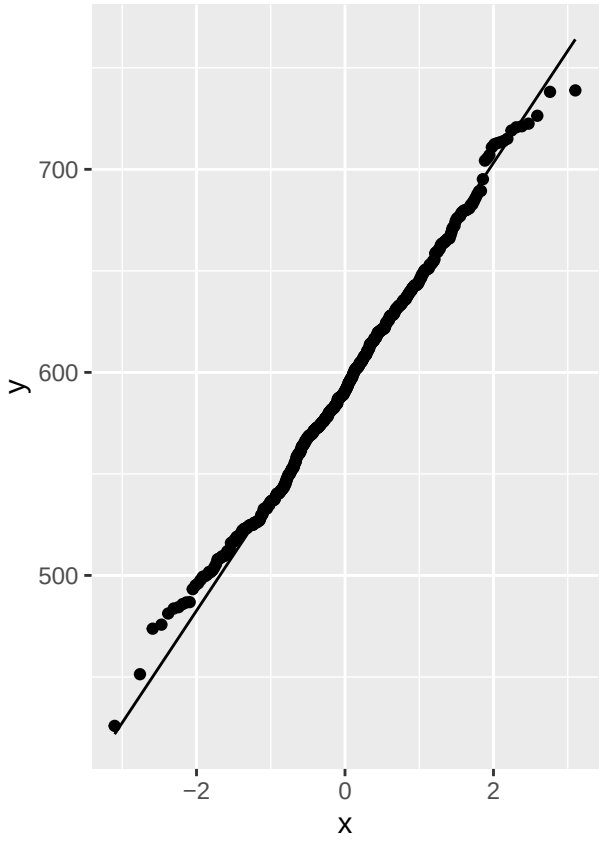

density
1.5e−05
1.0e−05
5.0e−06
0.0e+00
150000
200000
250000
Area

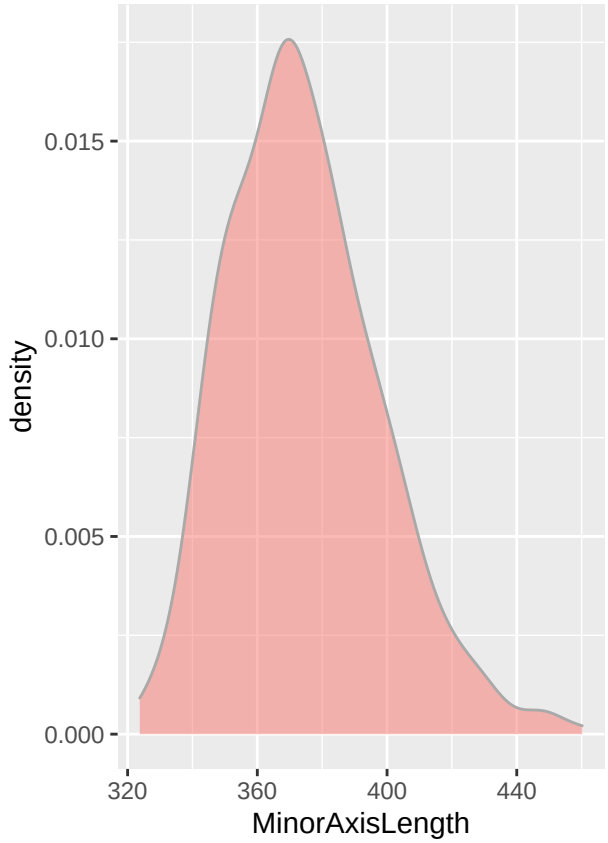

y
250000
200000
150000
100000
−2
0
2
x

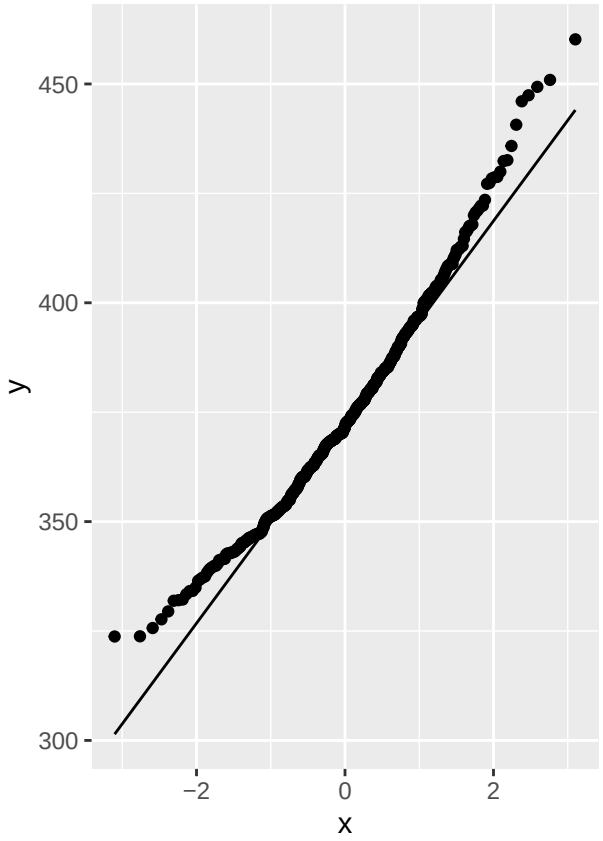

0.003
0.002
0.001
0.000
density
1400
1600
1800
2000
Perimeter

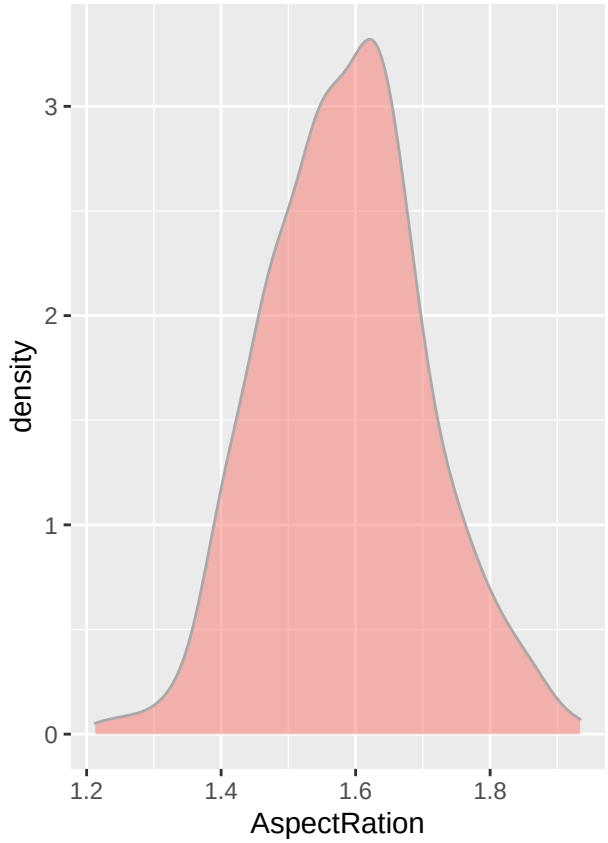

2000
1800
1600
1400
1200
y
−2
0
2
x

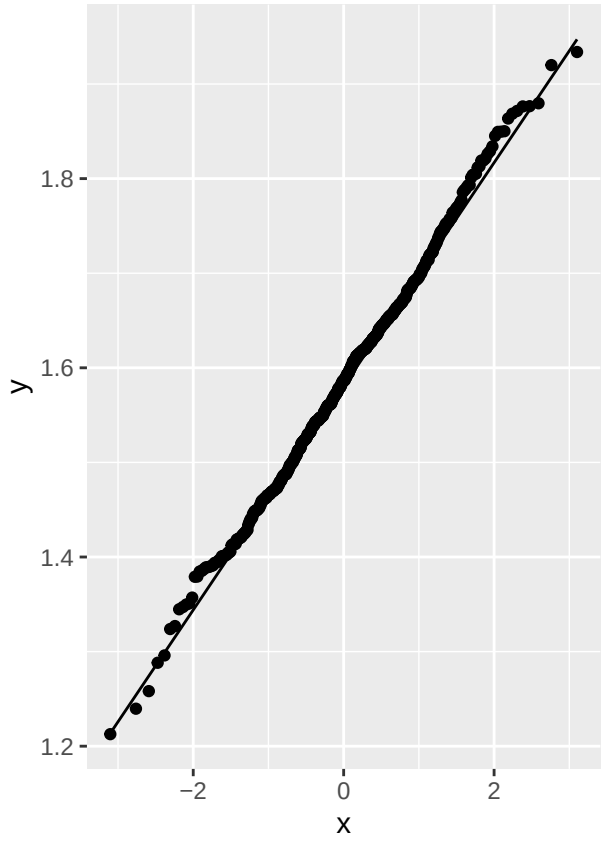

0.006
0.004
0.002
0.000
density
500
600
700
MajorAxisLength

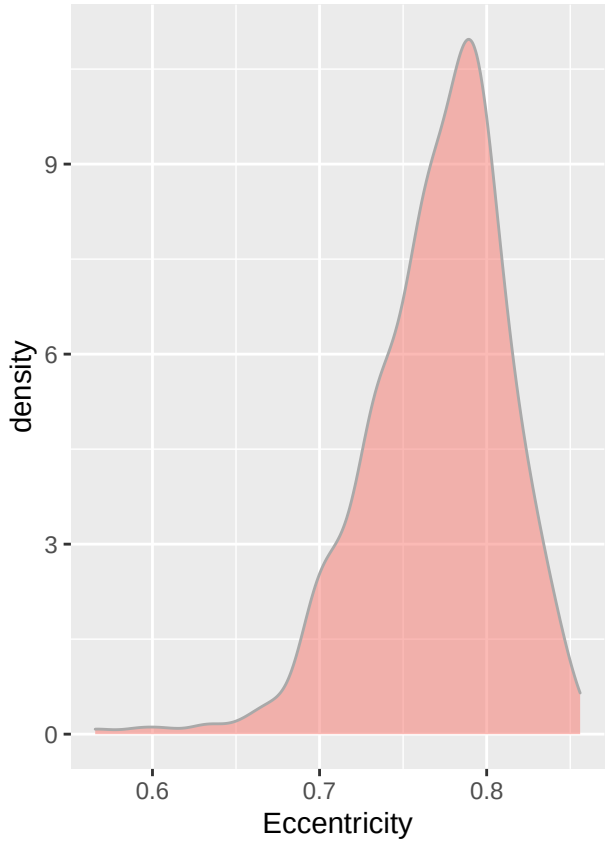

700
600
500
y
−2
0
2
x

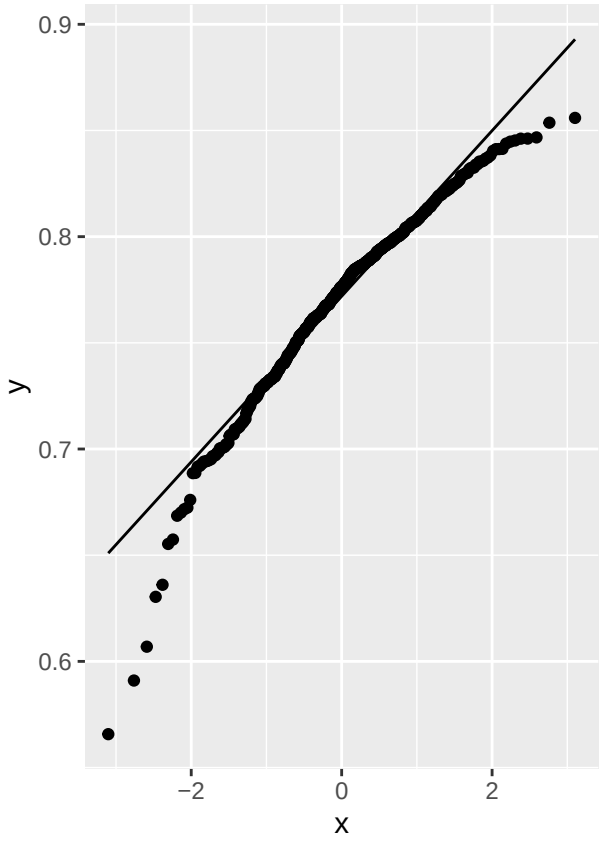

density
0.015
0.010
0.005
0.000
320
360
400
440
MinorAxisLength

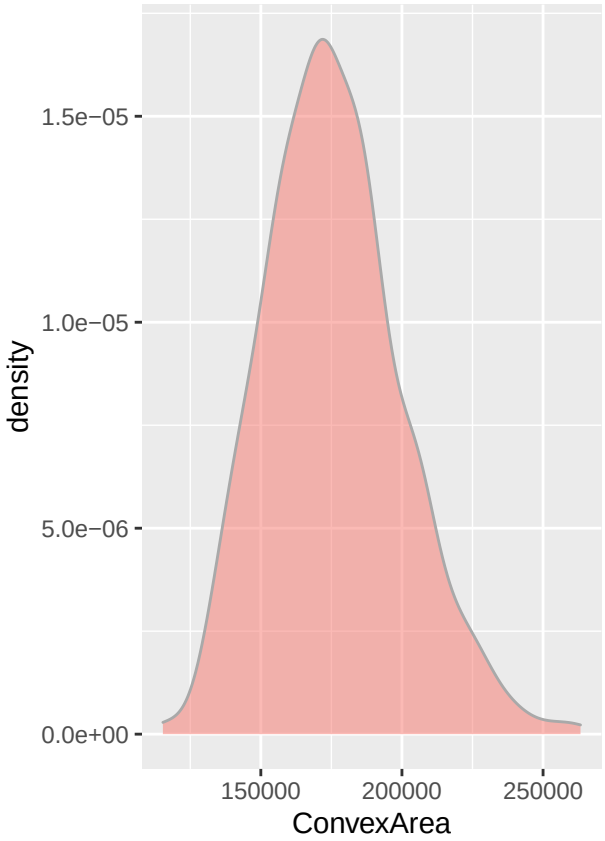

y
450
400
350
300
−2
0
2
x

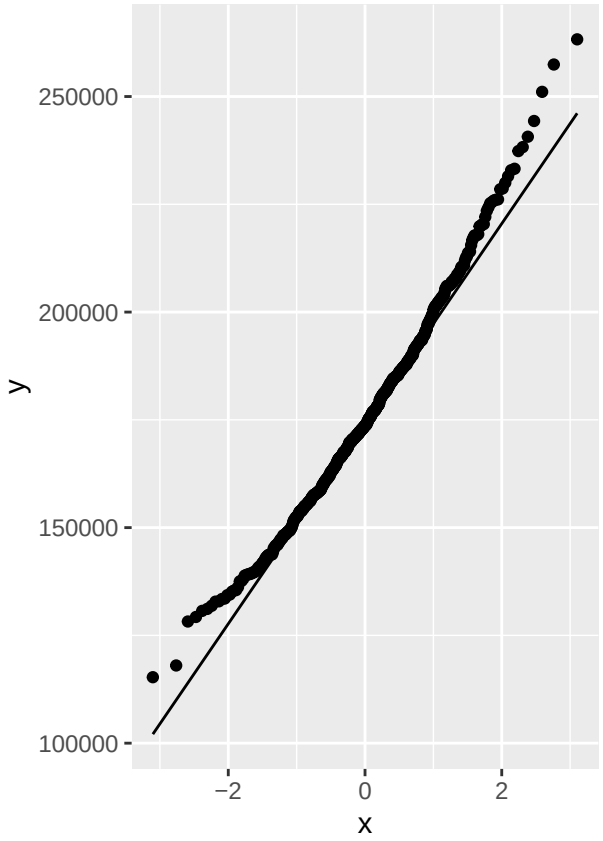

density
3
2
1
0
1.2
1.4
1.6
1.8
AspectRation

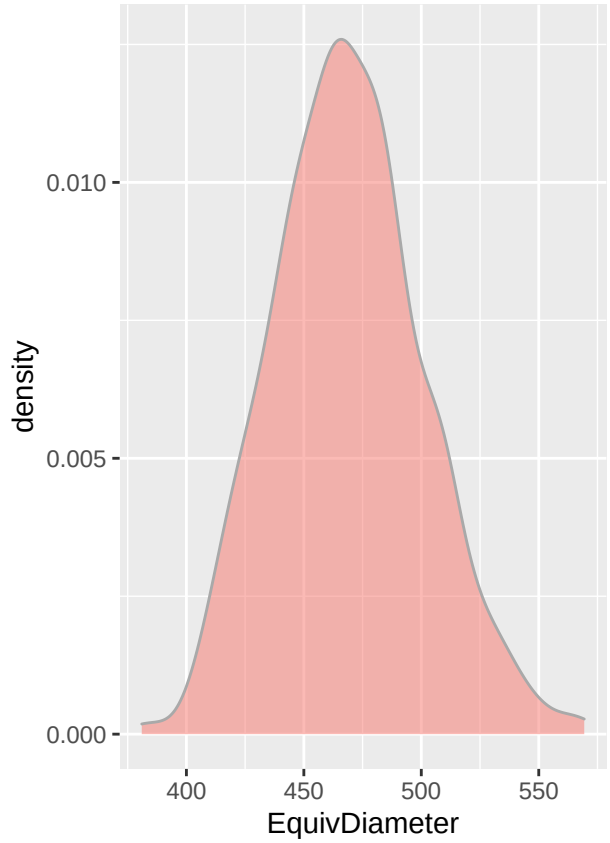

y
1.8
1.6
1.4
1.2
−2
0
2
x

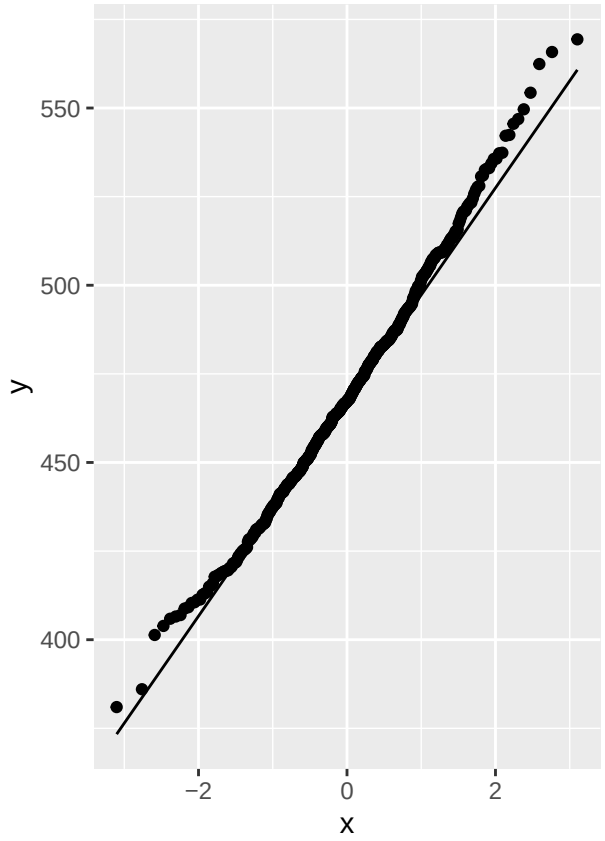

density
9
6
3
0
0.6
0.7
0.8
Eccentricity


y
0.9
0.8
0.7
0.6
−2
0
2
x

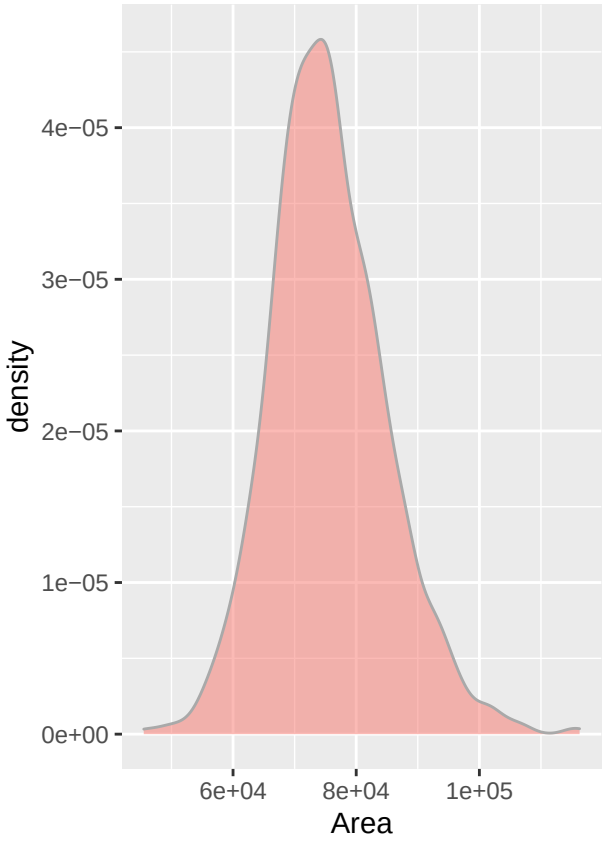

1.5e−05
1.0e−05
5.0e−06
0.0e+00
density
150000
200000
250000
ConvexArea

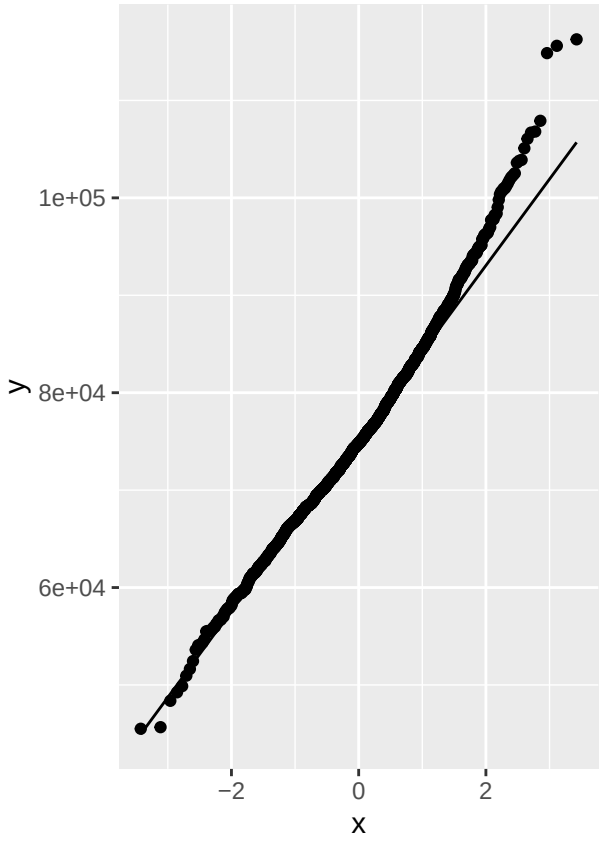

250000
200000
150000
100000
y
−2
0
2
x

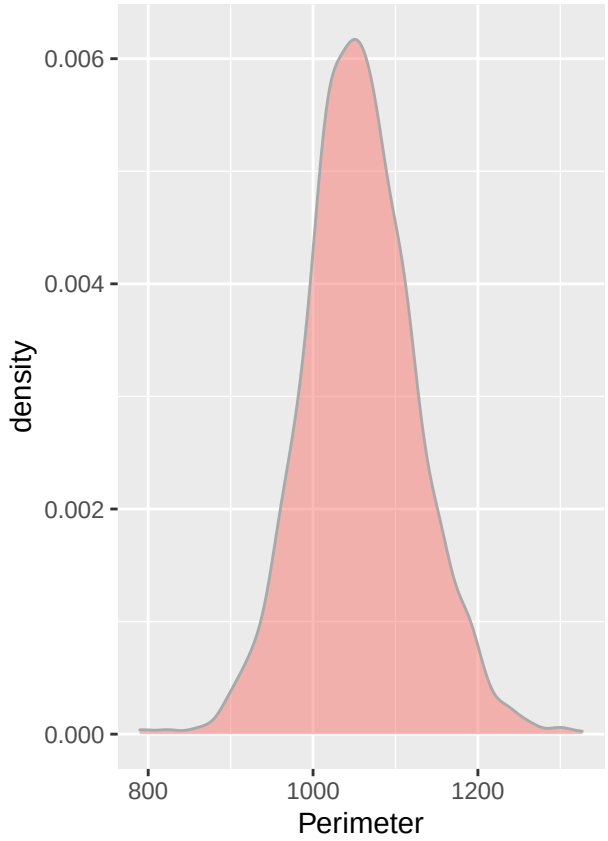

density
0.010
0.005
0.000
400
450
500
550
EquivDiameter

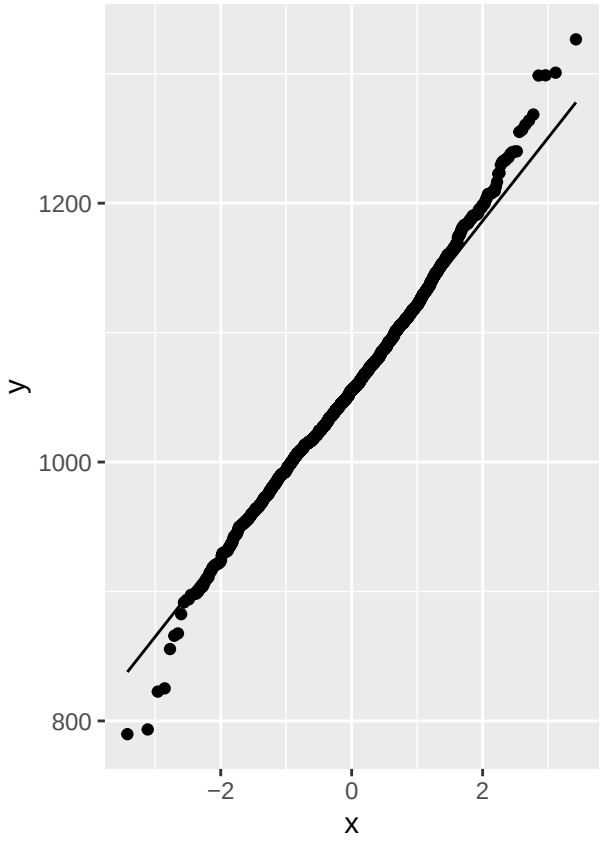

y
550
500
450
400
−2
0
2
x

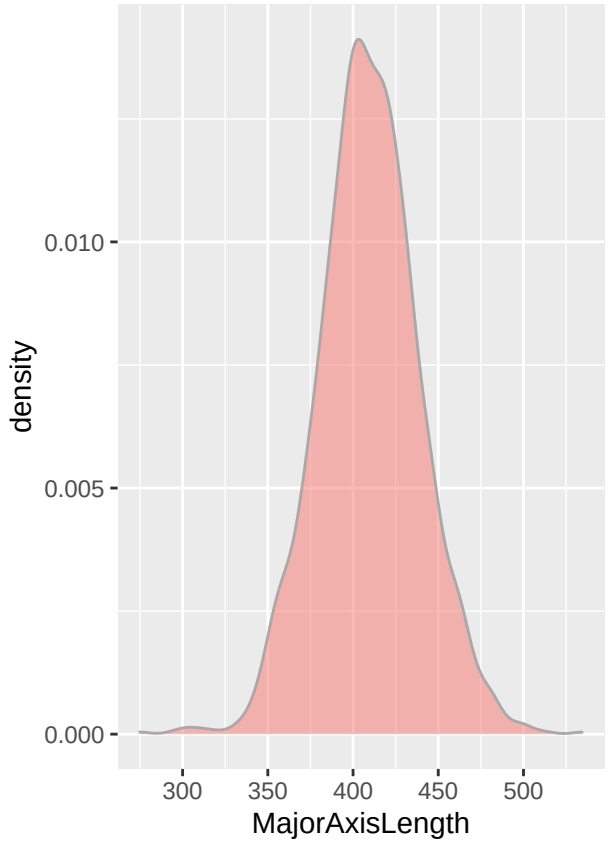
```
## [[1]]
## [1] "CALI"
```

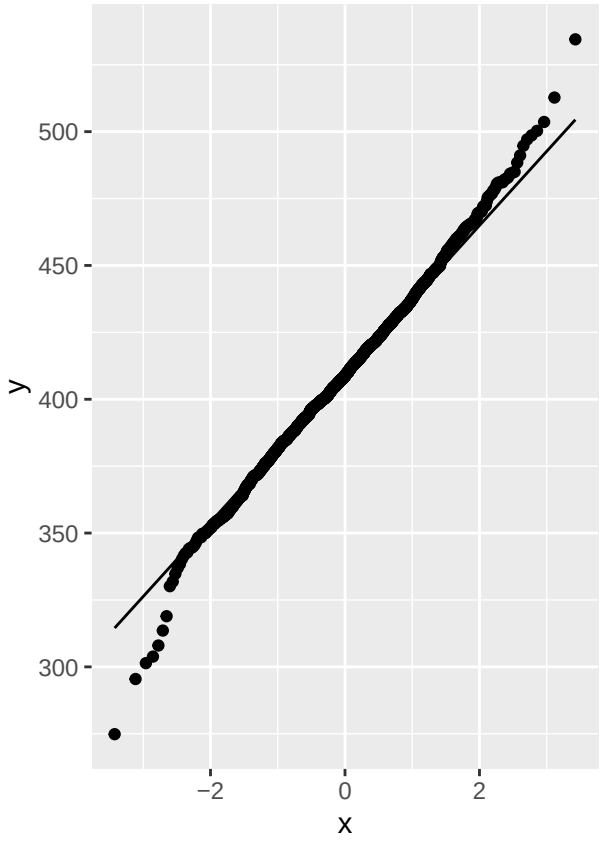

density
4e−05
3e−05
2e−05
1e−05
0e+00
6e+04
8e+04
1e+05
Area

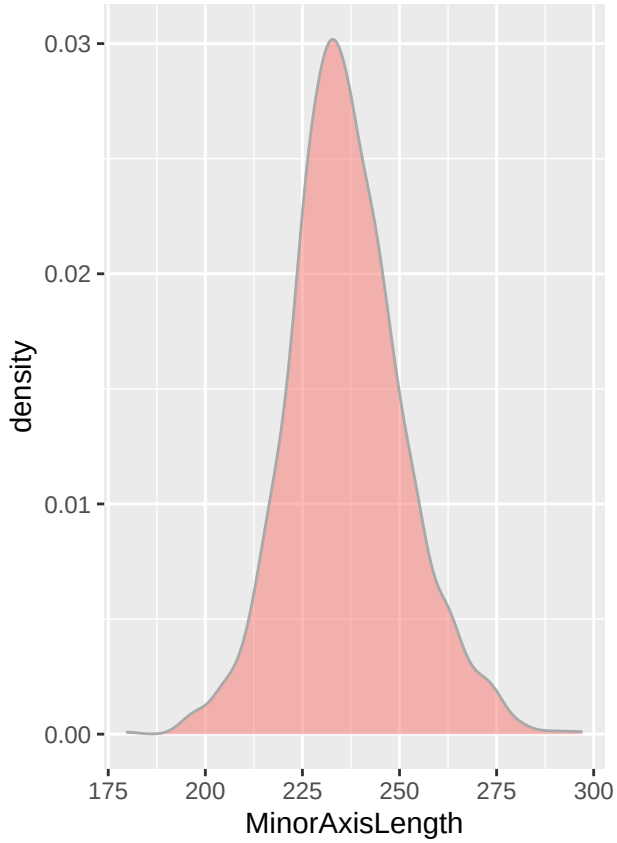

y
1e+05
8e+04
6e+04
−2
0
2
x

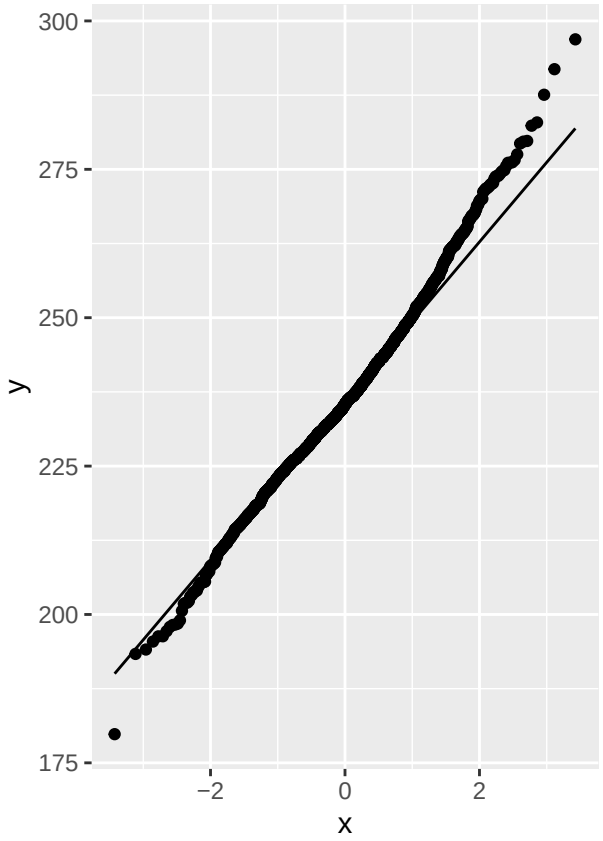

0.006
0.004
0.002
0.000
density
800
1000
1200
Perimeter

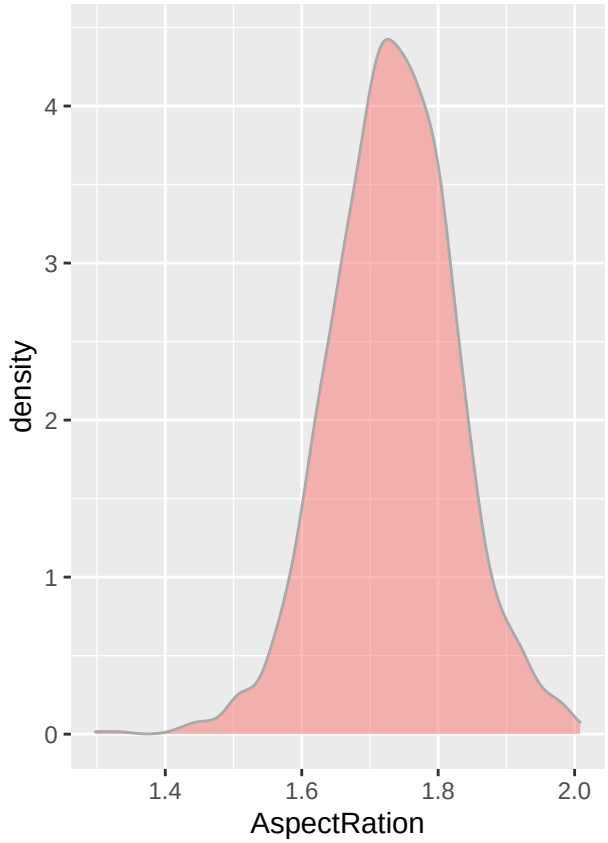

1200
1000
800
y
−2
0
2
x

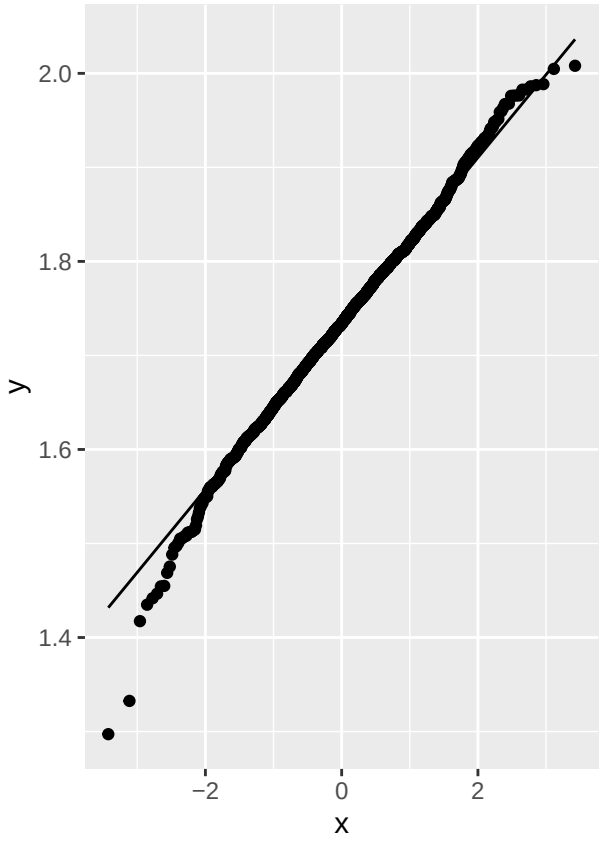

density
0.010
0.005
0.000
300
350
400
450
500
MajorAxisLength

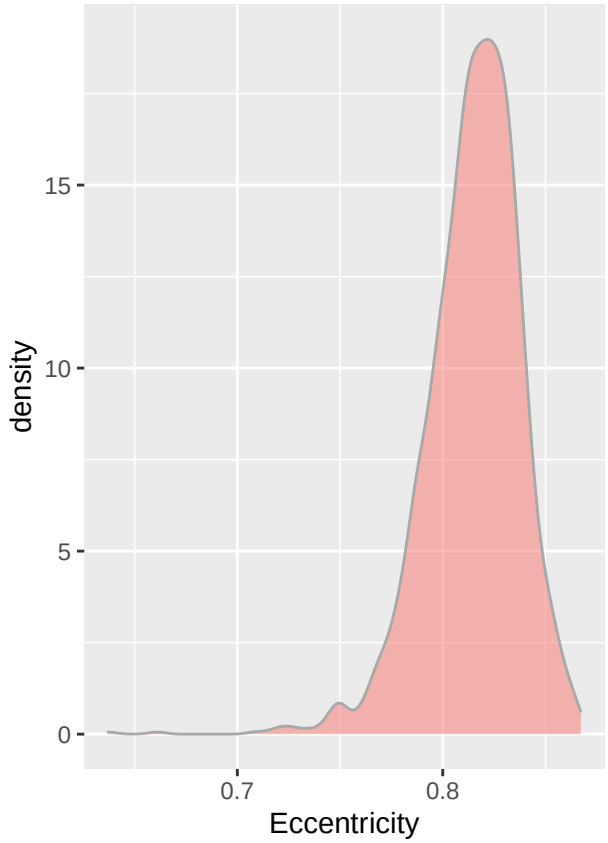

y
500
450
400
350
300
−2
0
2
x

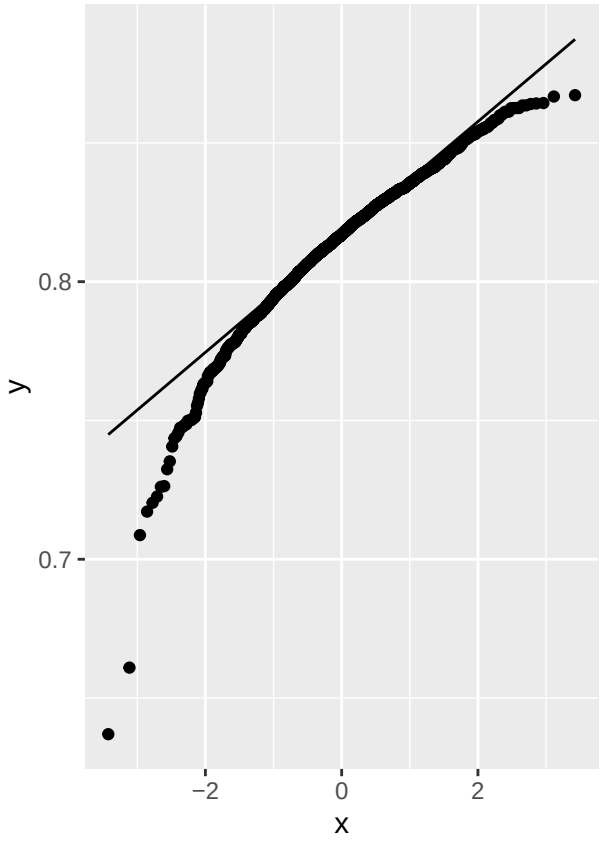

0.03
0.02
0.01
0.00
density
175
200
225
250
275
300
MinorAxisLength

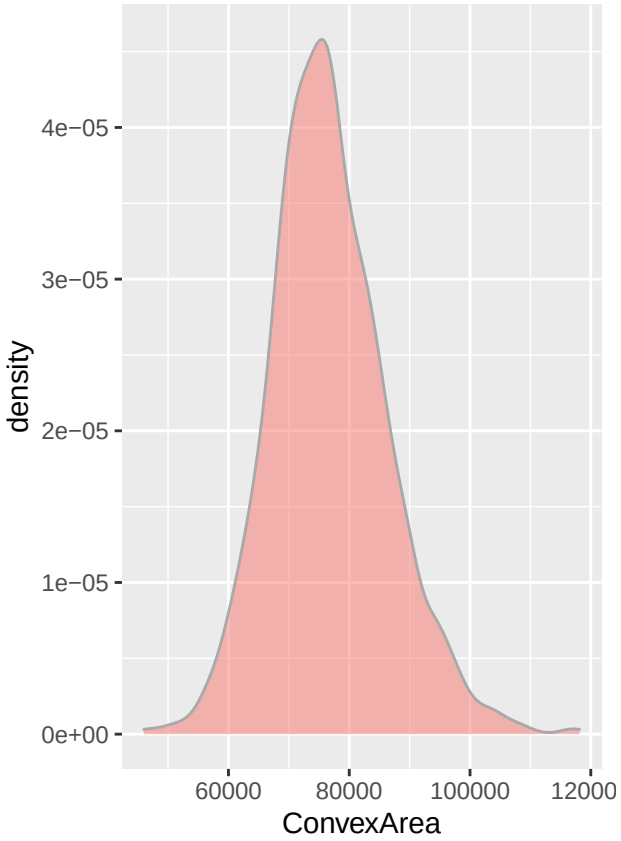

300
275
250
225
200
175
y
−2
0
2
x

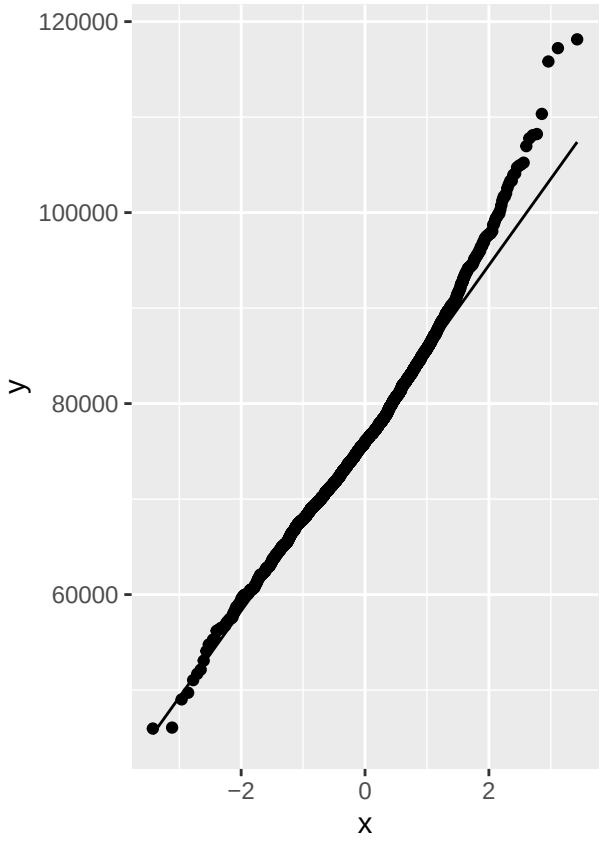

density
AspectRation
0
1
2
3
4
1.4
1.6
1.8
2.0

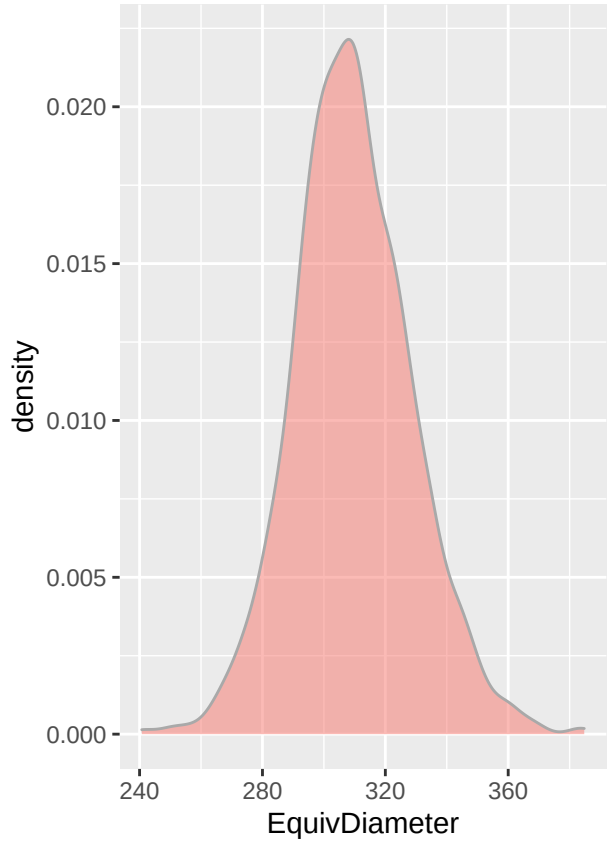

y
x
1.4
1.6
1.8
2.0
−2
0
2

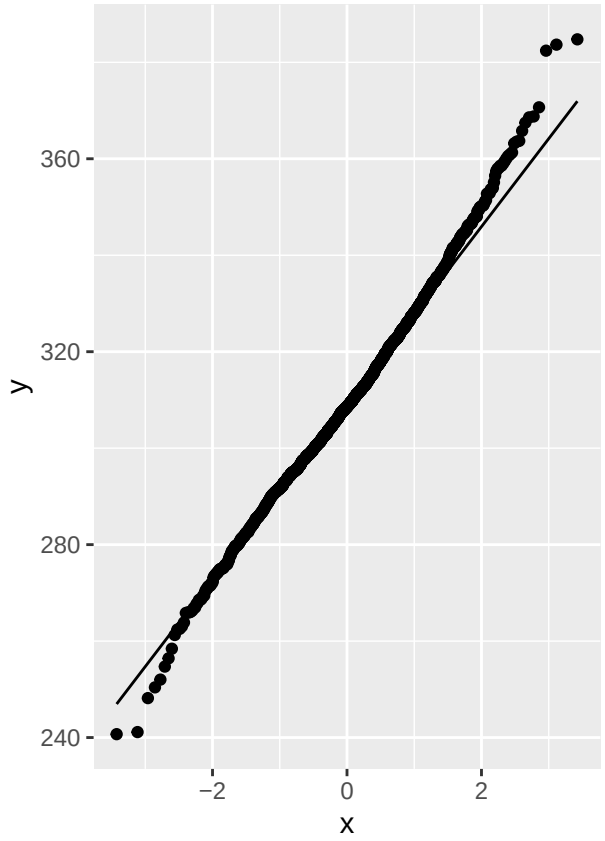

density
15
10
5
0
0.7
0.8
Eccentricity


y
0.8
0.7
−2
0
2
x

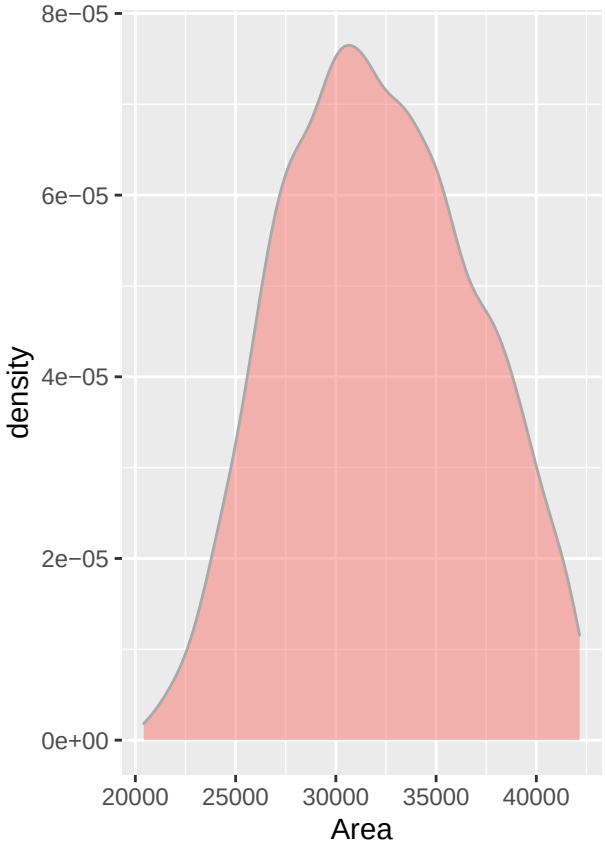

density
4e−05
3e−05
2e−05
1e−05
0e+00
60000
80000
100000
12000
ConvexArea

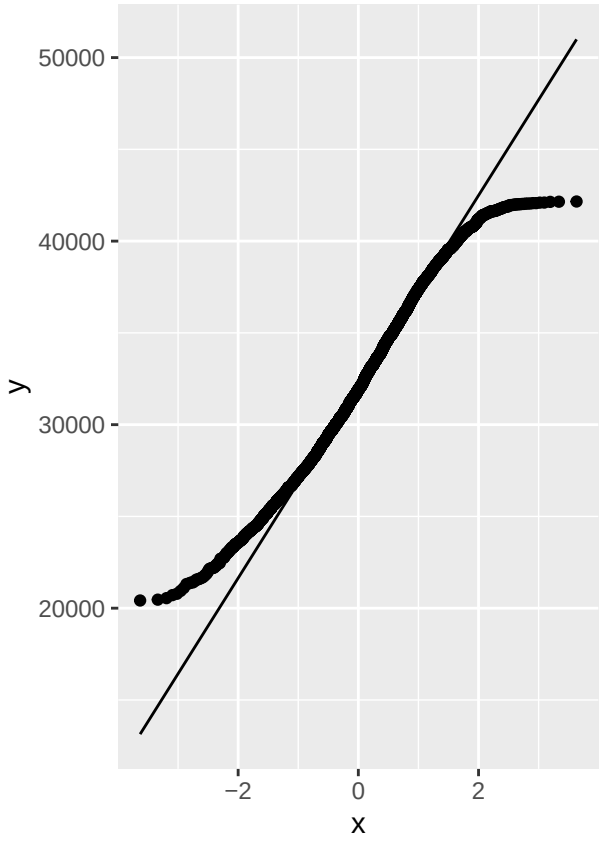

y
120000
100000
80000
60000
−2
0
2
x

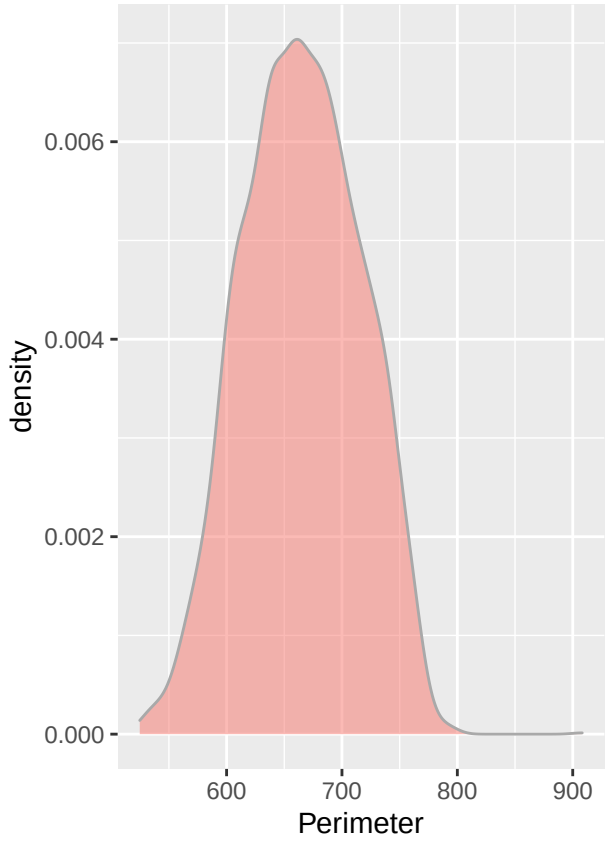

0.020
0.015
0.010
0.005
0.000
density
240
280
320
360
EquivDiameter

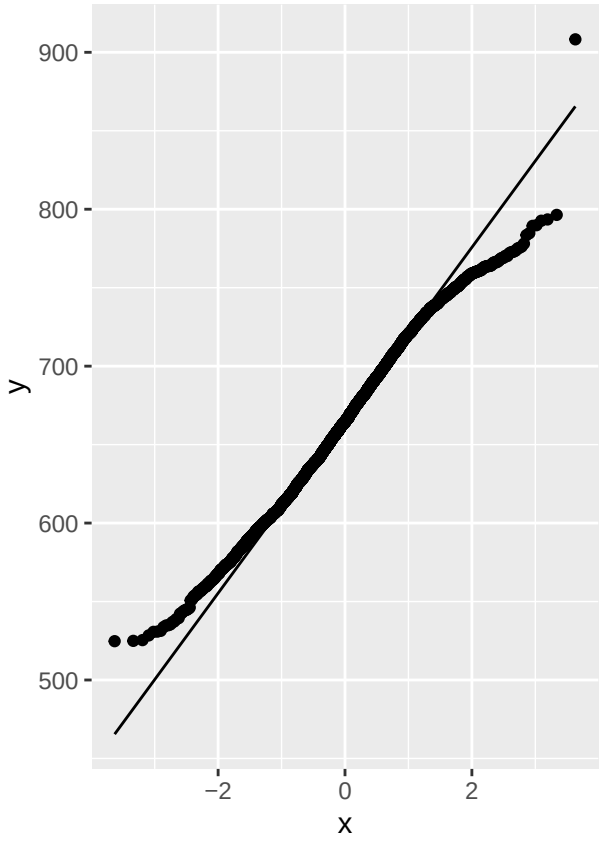

360
320
280
240
y
−2
0
2
x

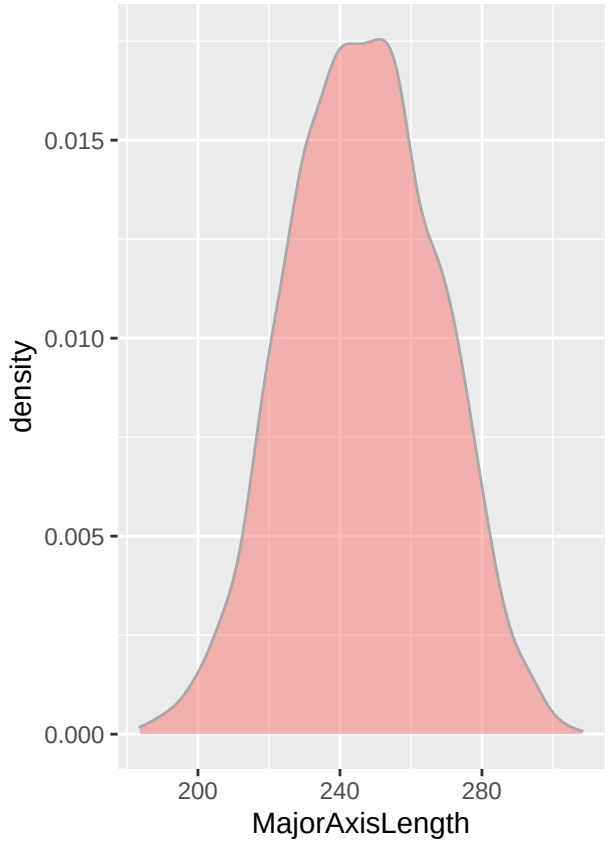
```
## [[1]]
## [1] "DERMASON"
```

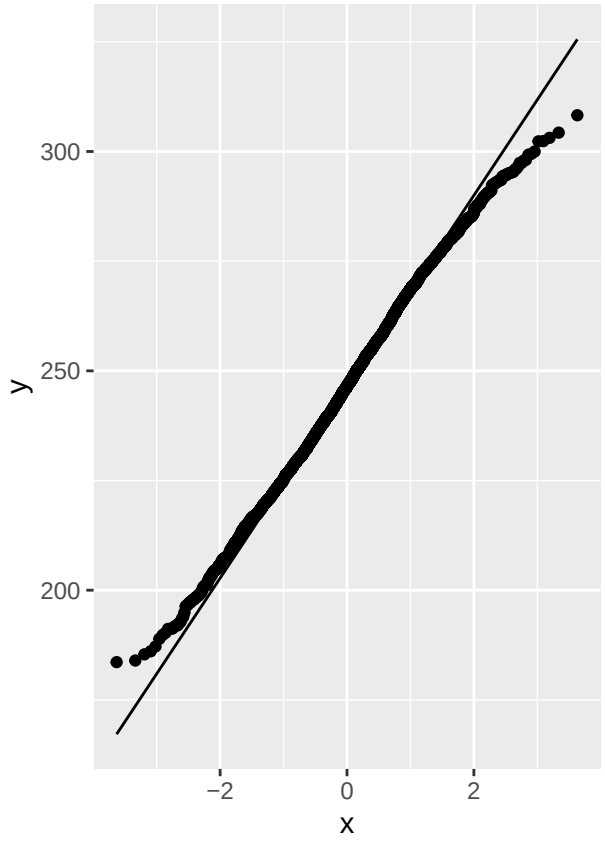

8e−05
6e−05
4e−05
2e−05
0e+00
density
20000
25000
30000
35000
40000
Area

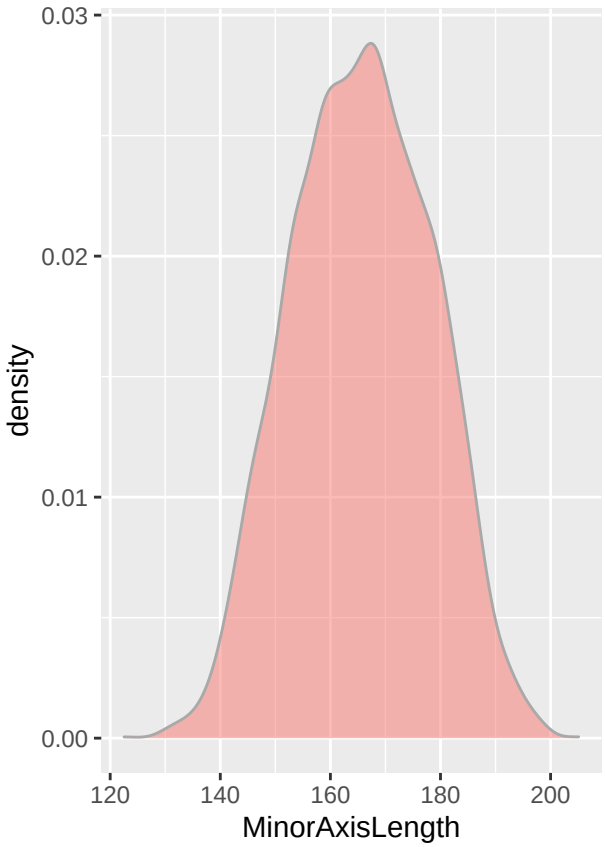

50000
40000
30000
20000
y
−2
0
2
x

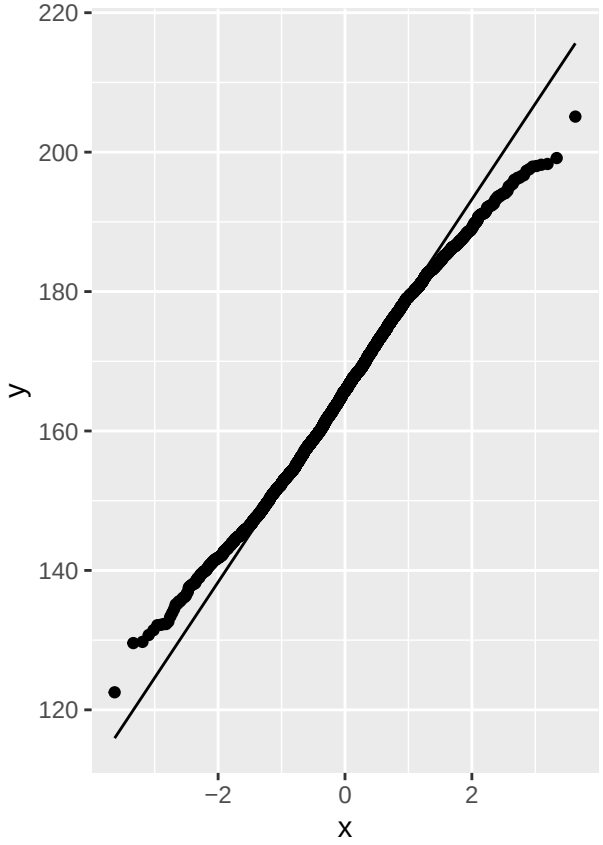

density
0.006
0.004
0.002
0.000
600
700
800
900
Perimeter

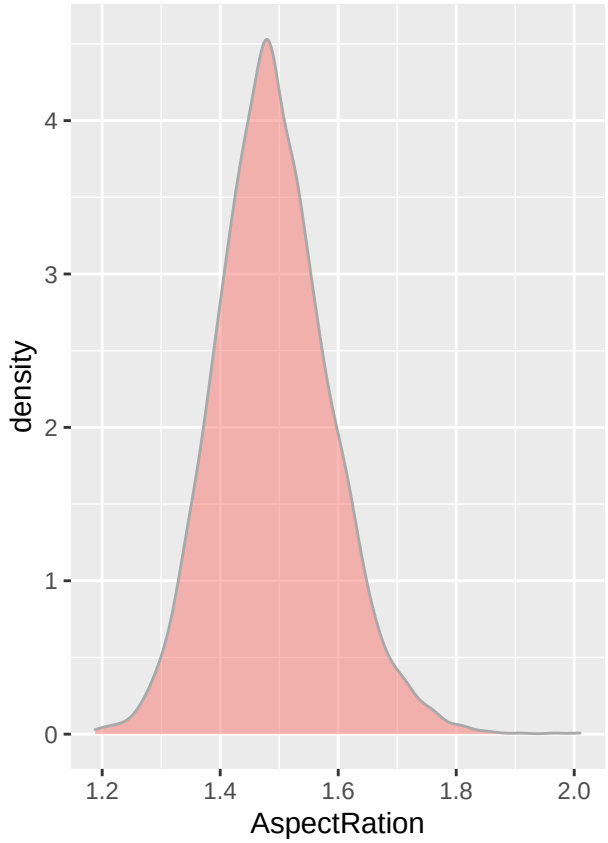

900
800
700
600
500
y
−2
0
2
x

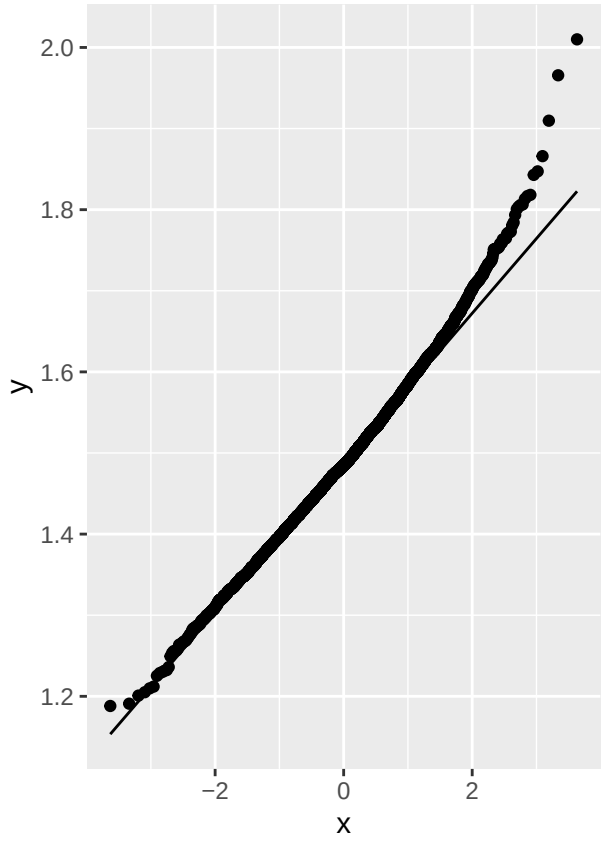

0.015
0.010
0.005
0.000
density
200
240
280
MajorAxisLength

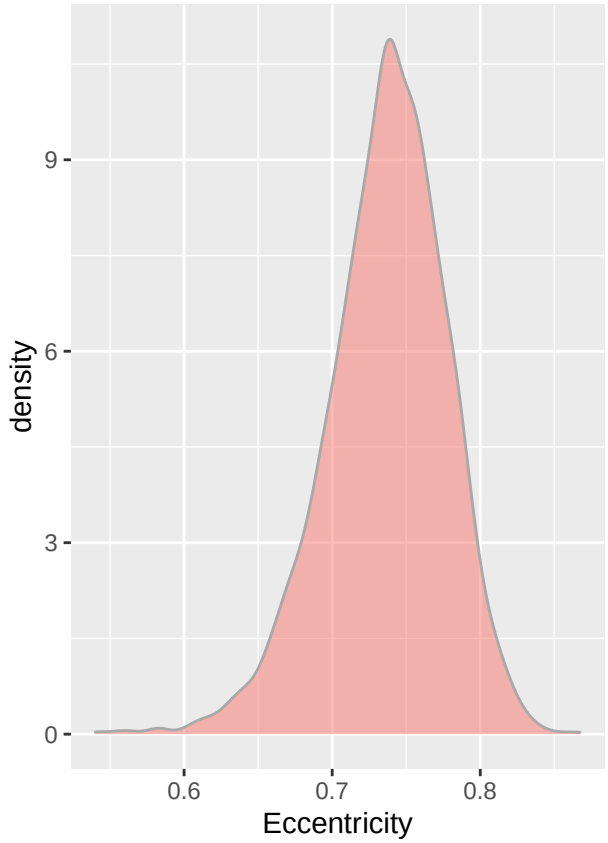

300
250
200
y
−2
0
2
x

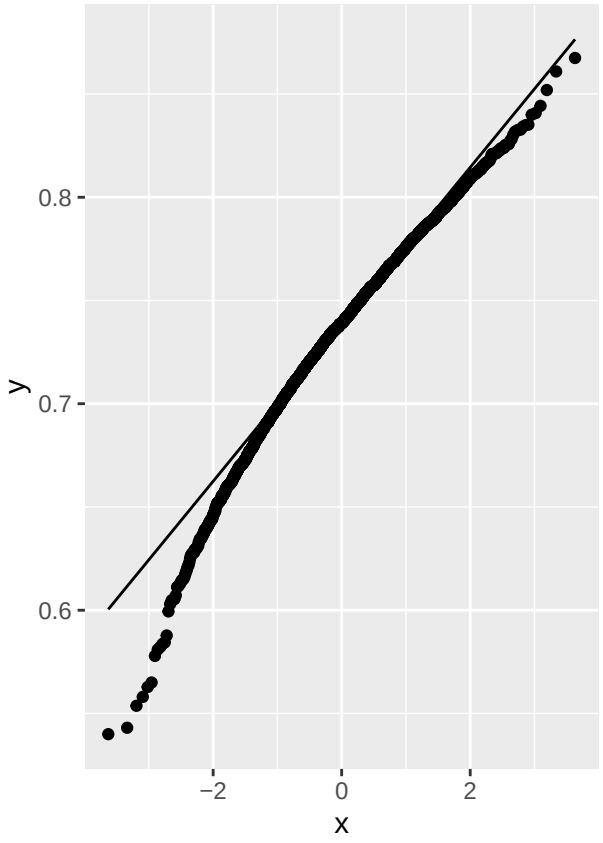

0.03
0.02
0.01
0.00
density
120
140
160
180
200
MinorAxisLength

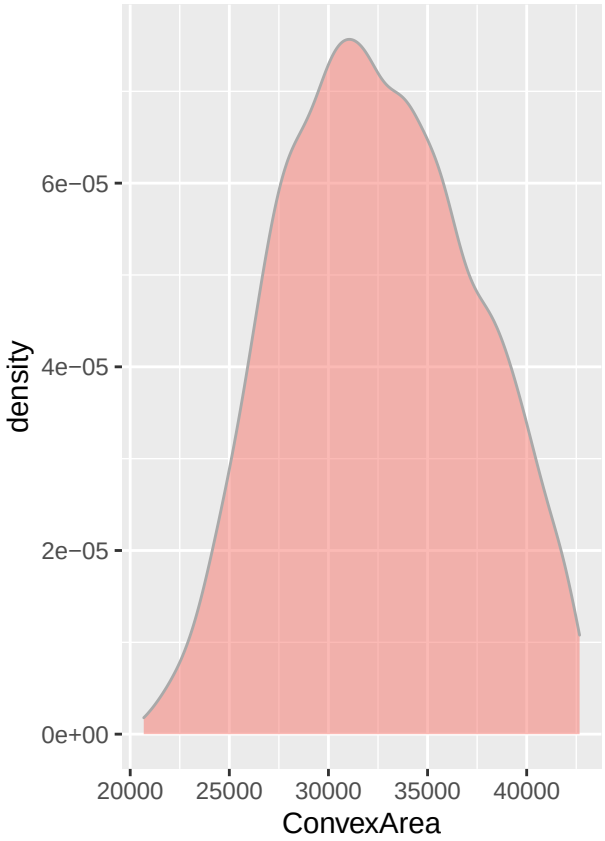

220
200
180
160
140
120
y
−2
0
2
x

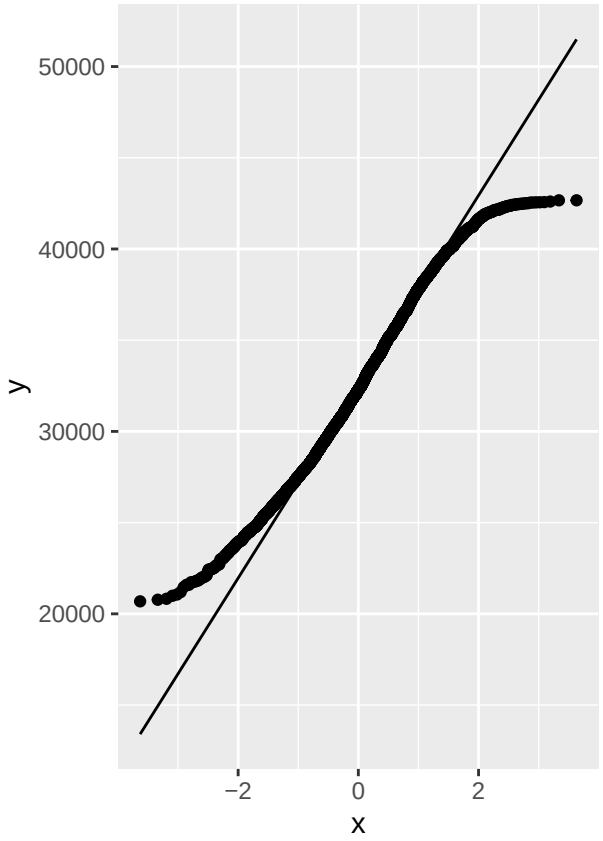

density
AspectRation
0
1
2
3
4
1.2
1.4
1.6
1.8
2.0

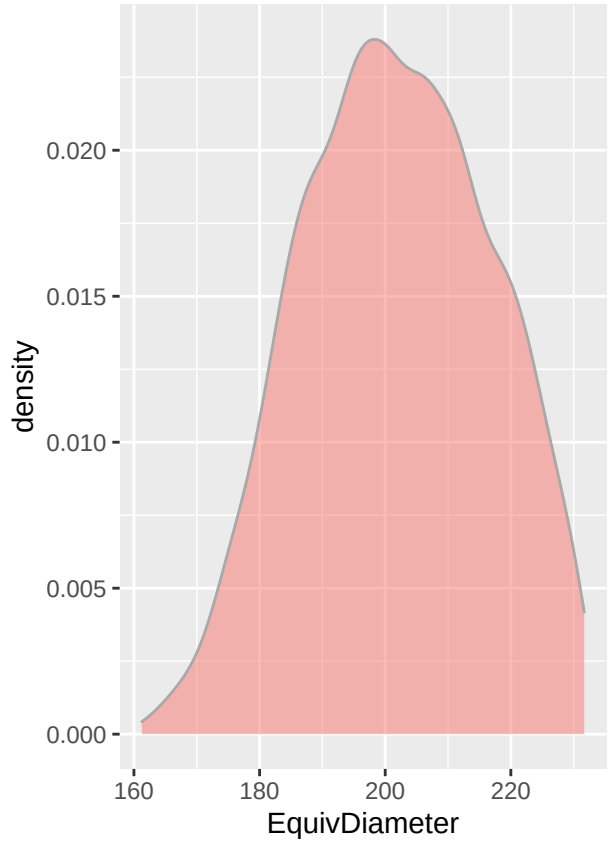

y
x
1.2
1.4
1.6
1.8
2.0
−2
0
2

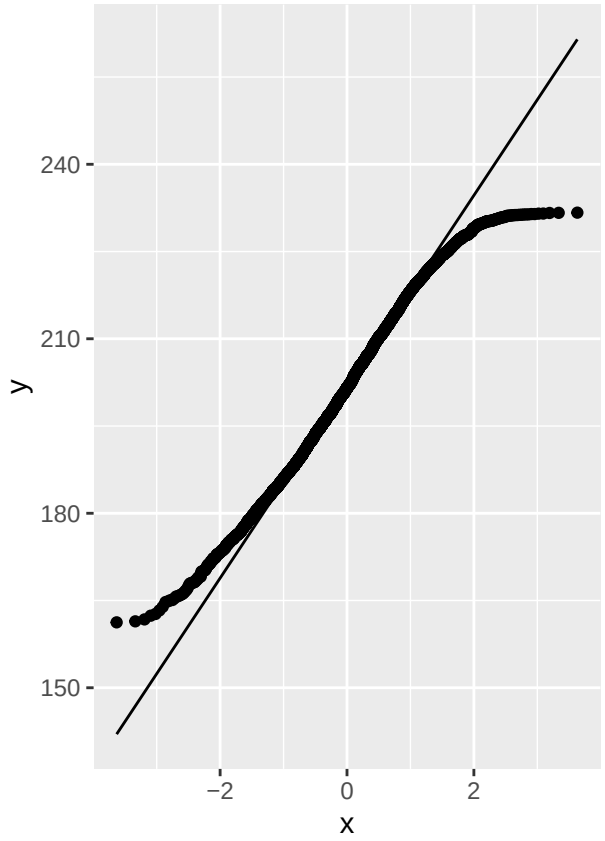

density
9
6
3
0
0.6
0.7
0.8
Eccentricity


y
0.8
0.7
0.6
−2
0
2
x

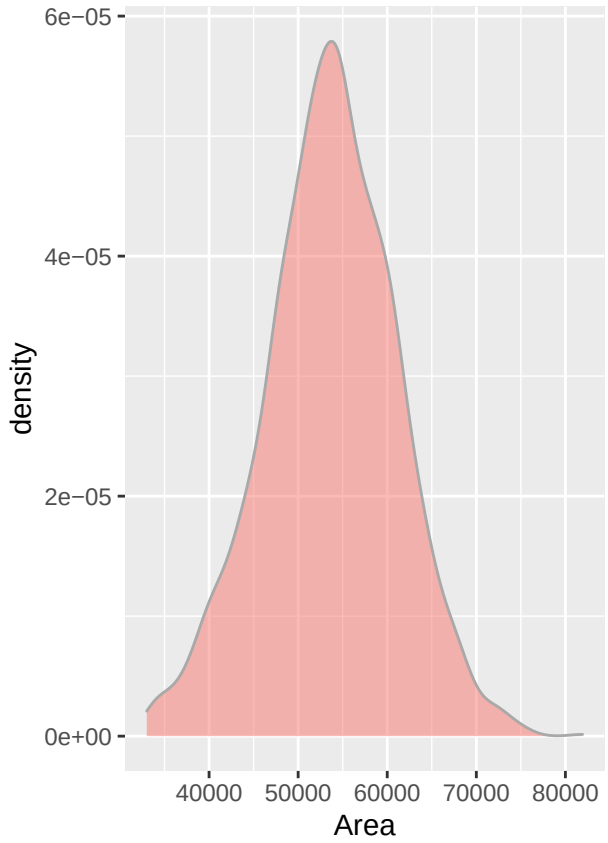

density
6e−05
4e−05
2e−05
0e+00
20000
25000
30000
35000
40000
ConvexArea

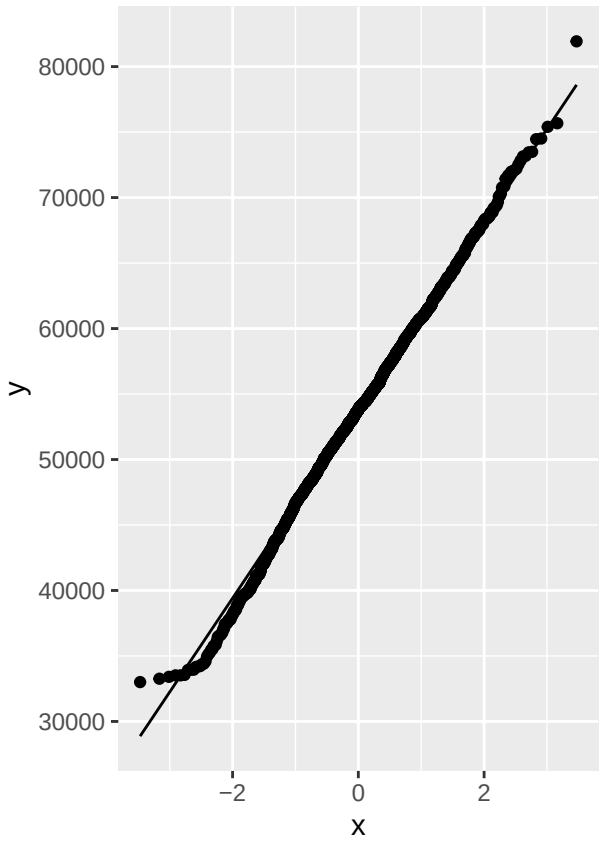

y
50000
40000
30000
20000
−2
0
2
x

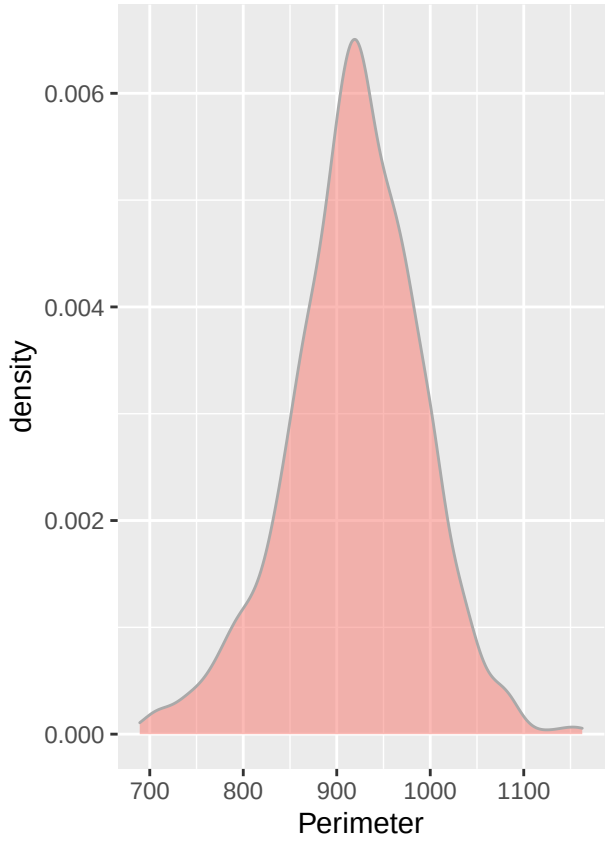

0.020
0.015
0.010
0.005
0.000
density
160
180
200
220
EquivDiameter

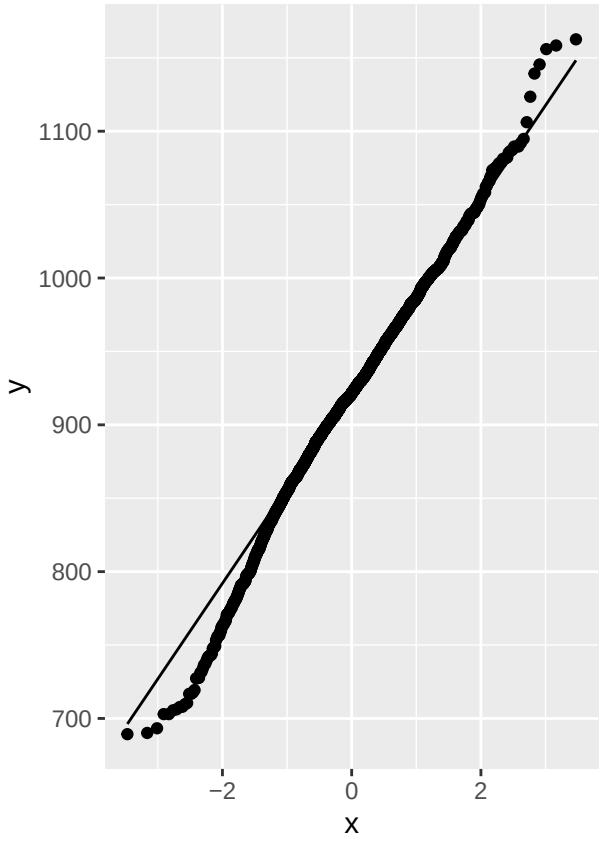

240
210
180
150
y
−2
0
2
x

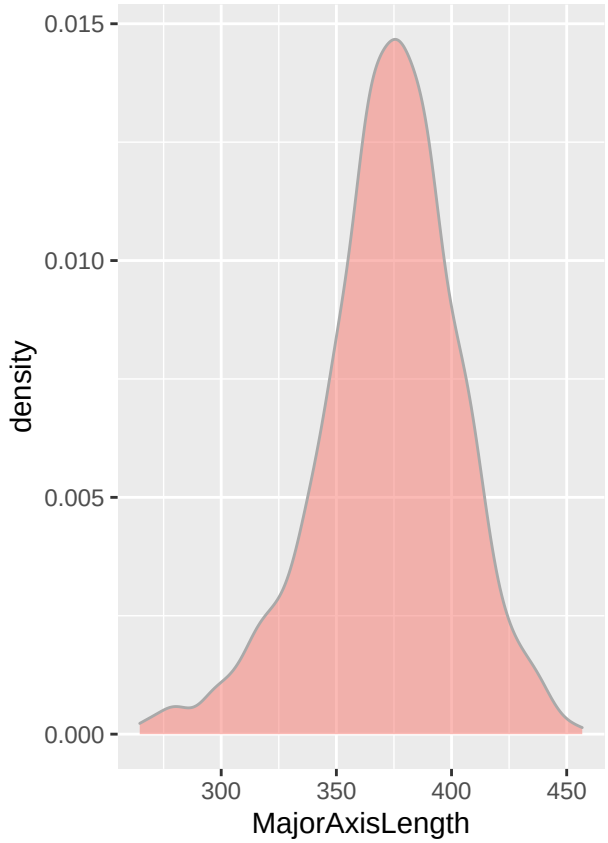
```
## [[1]]
## [1] "HOROZ"
```

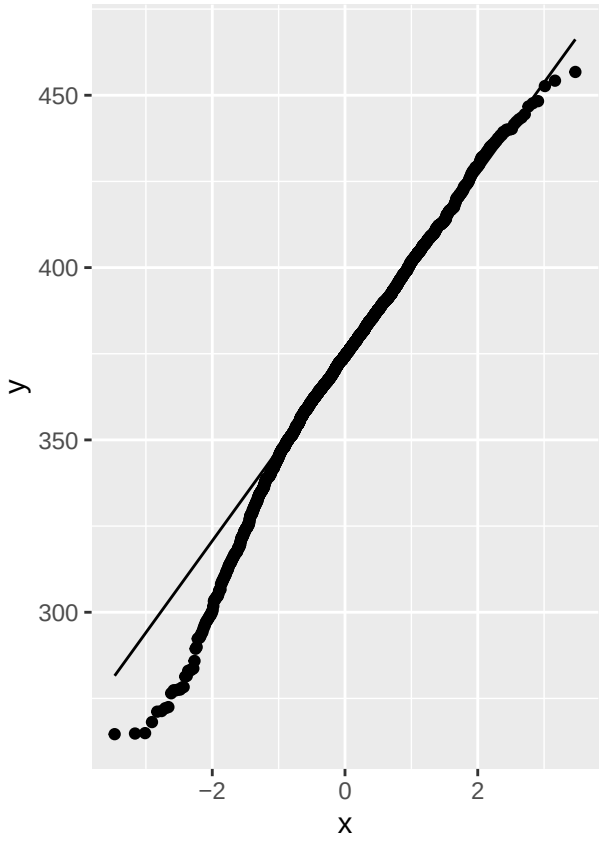

6e−05
4e−05
2e−05
0e+00
density
40000
50000
60000
70000
80000
Area

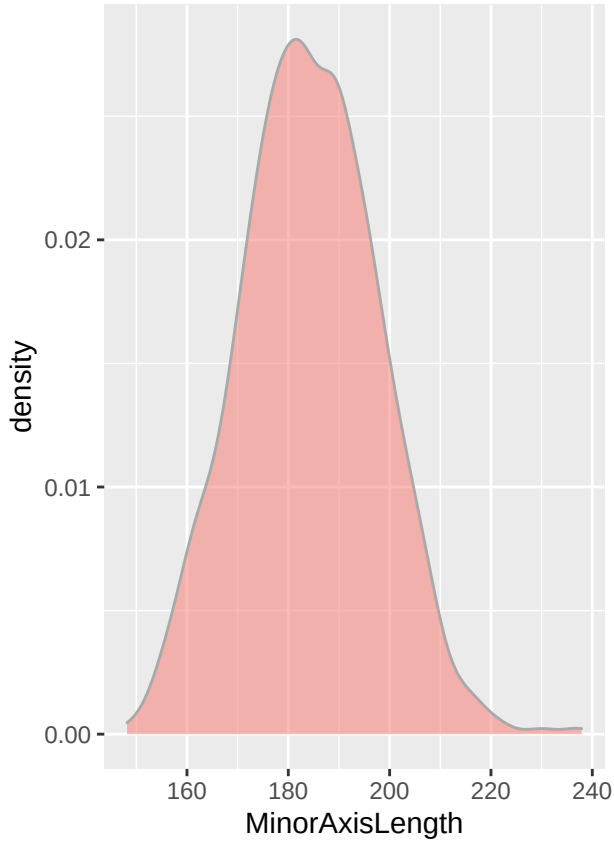

80000
70000
60000
50000
40000
30000
y
−2
0
2
x

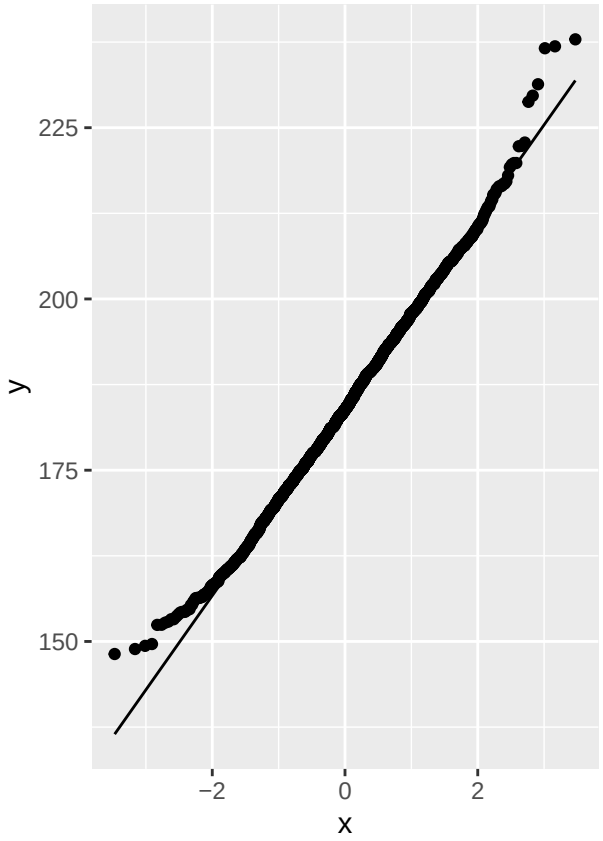

0.006
0.004
0.002
0.000
density
700
800
900
1000
1100
Perimeter

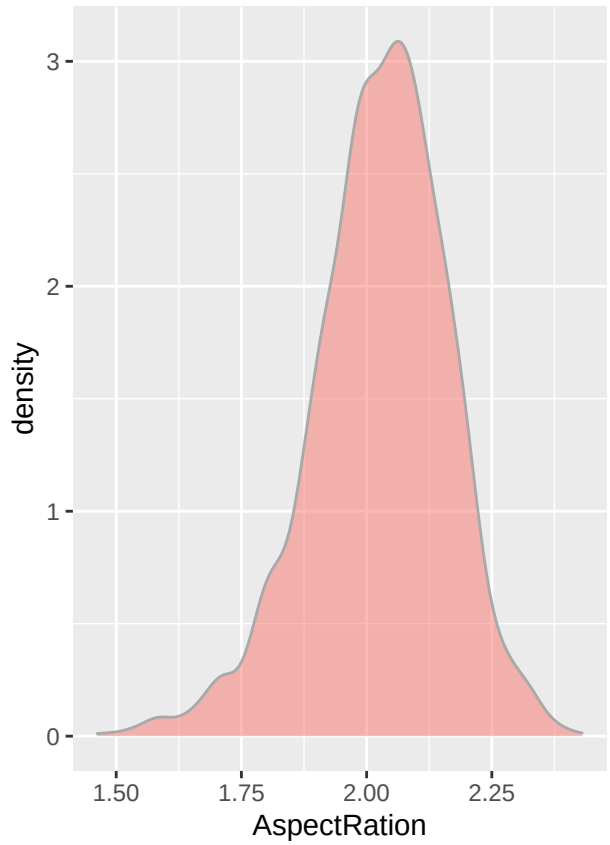

1100
1000
900
800
700
y
−2
0
2
x

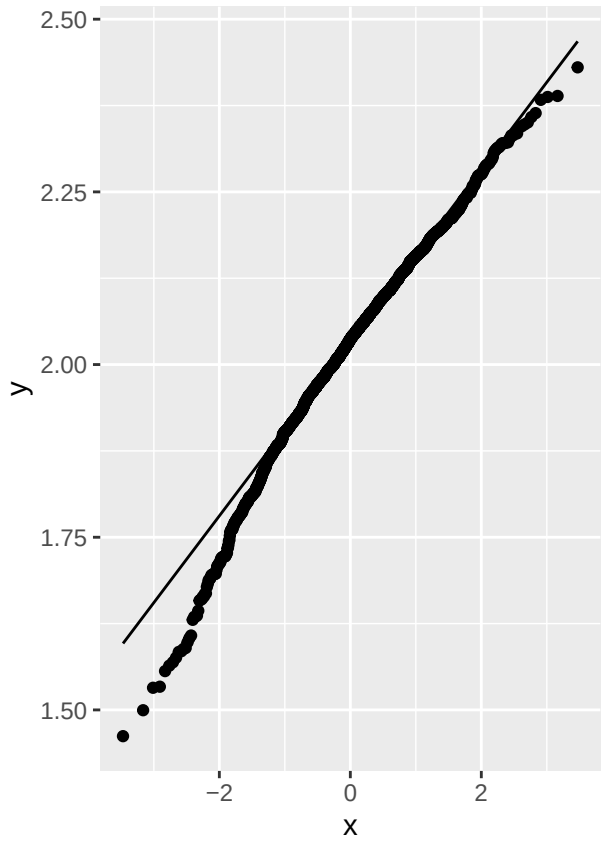

0.015
0.010
0.005
0.000
density
300
350
400
450
MajorAxisLength

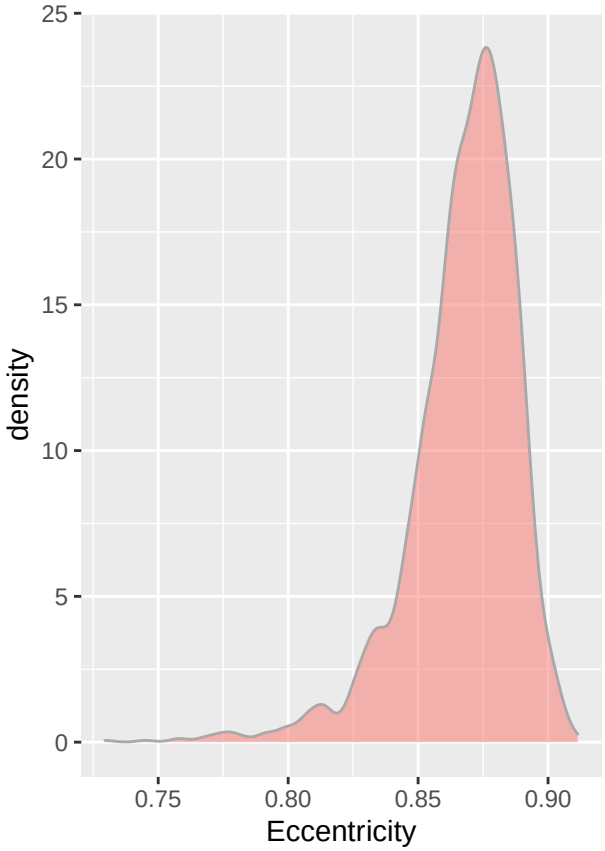

450
400
350
300
y
−2
0
2
x

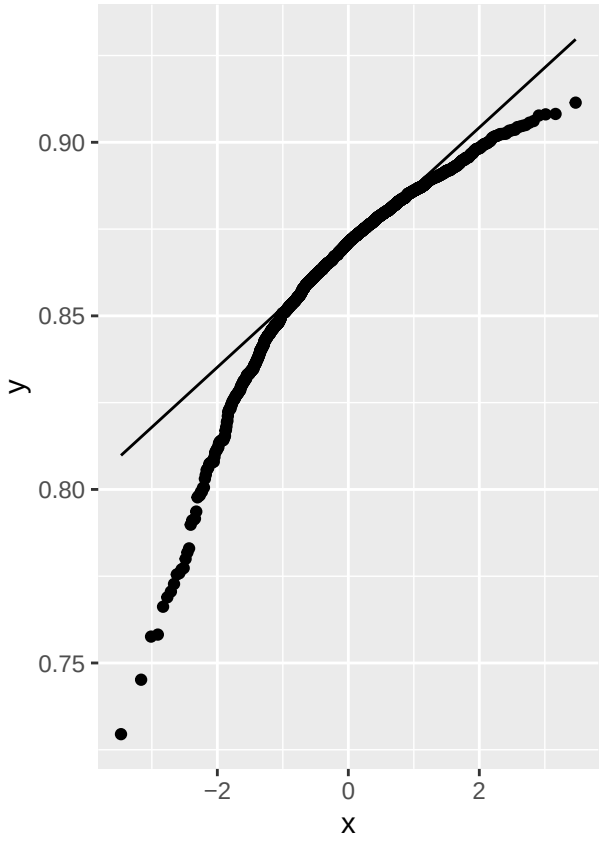

density
0.02
0.01
0.00
160
180
200
220
240
MinorAxisLength

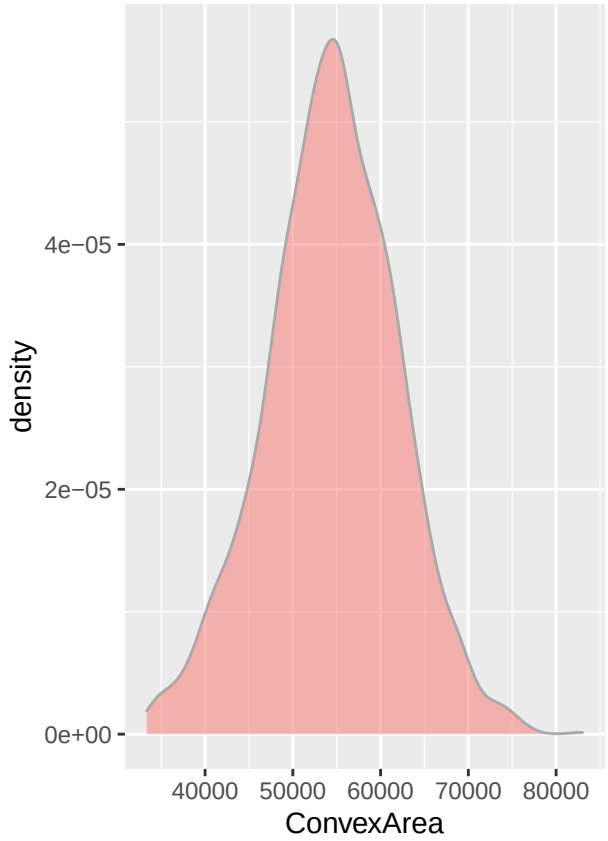

y
225
200
175
150
−2
0
2
x

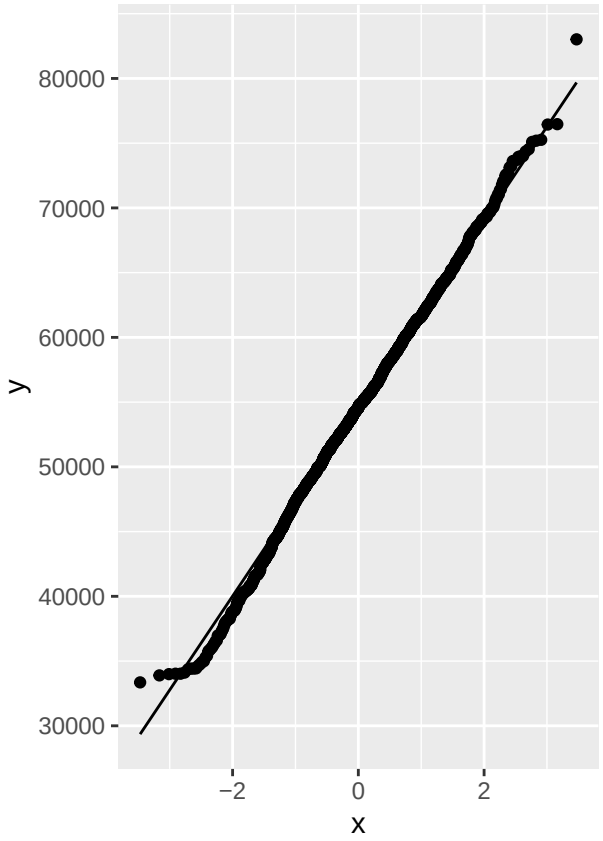

density
AspectRation
0
1
2
3
1.50
1.75
2.00
2.25

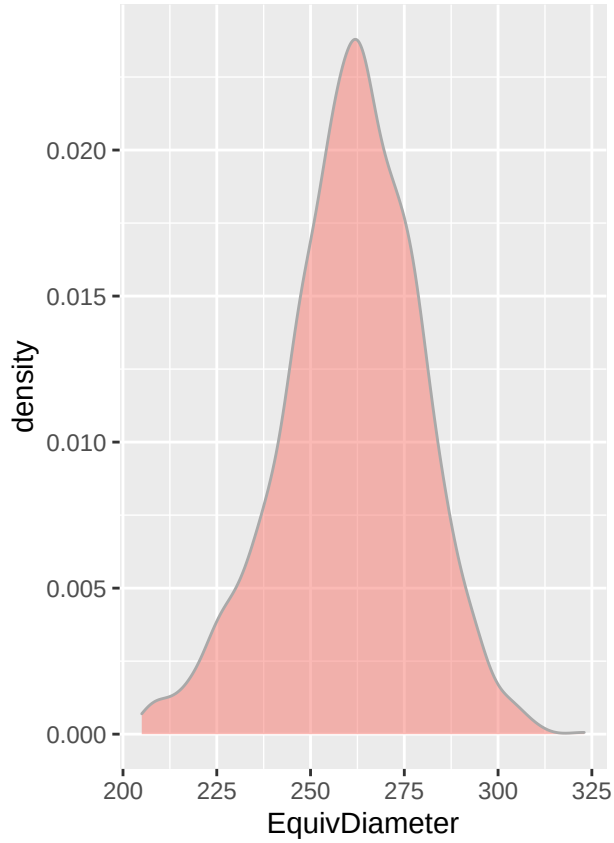

y
x
1.50
1.75
2.00
2.25
2.50
−2
0
2

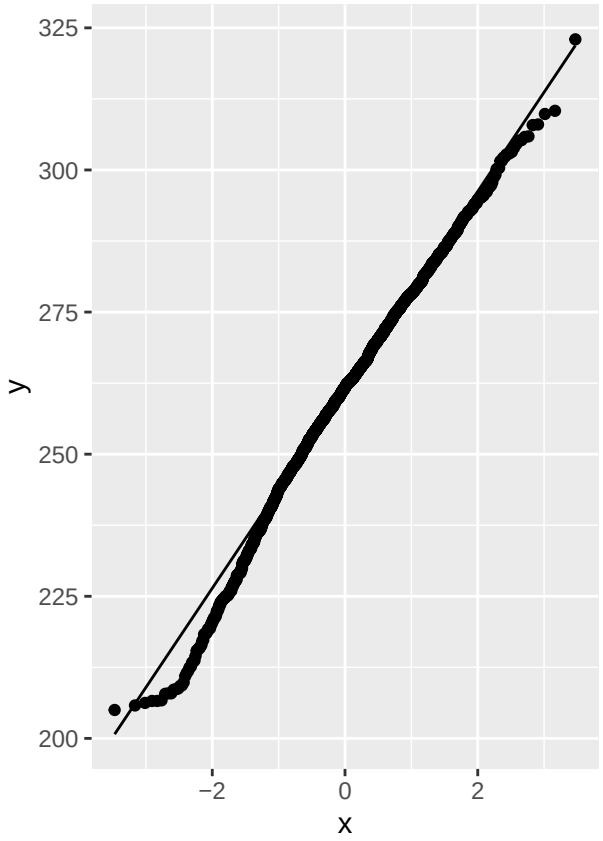

25
20
15
10
5
0
density
0.75
0.80
0.85
0.90
Eccentricity


0.90
0.85
0.80
0.75
y
−2
0
2
x

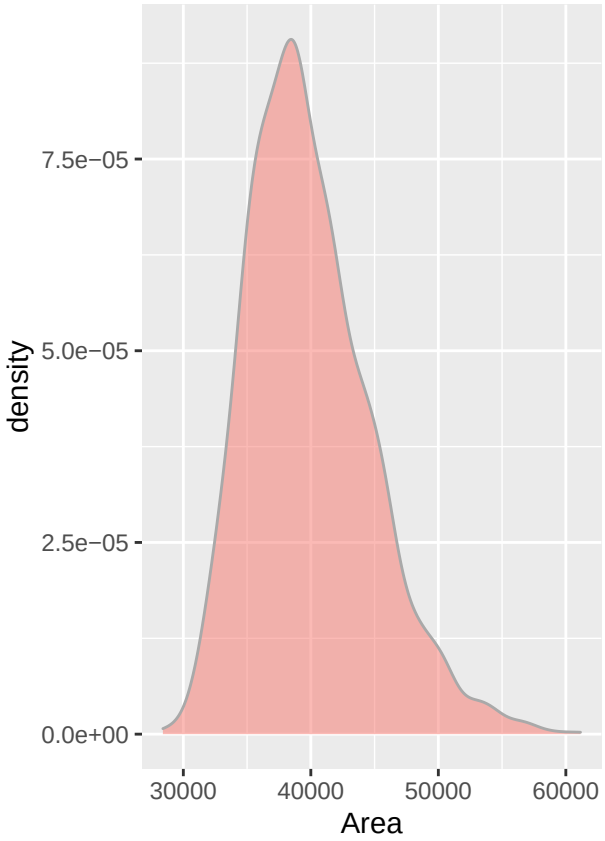

density
4e−05
2e−05
0e+00
40000
50000
60000
70000
80000
ConvexArea

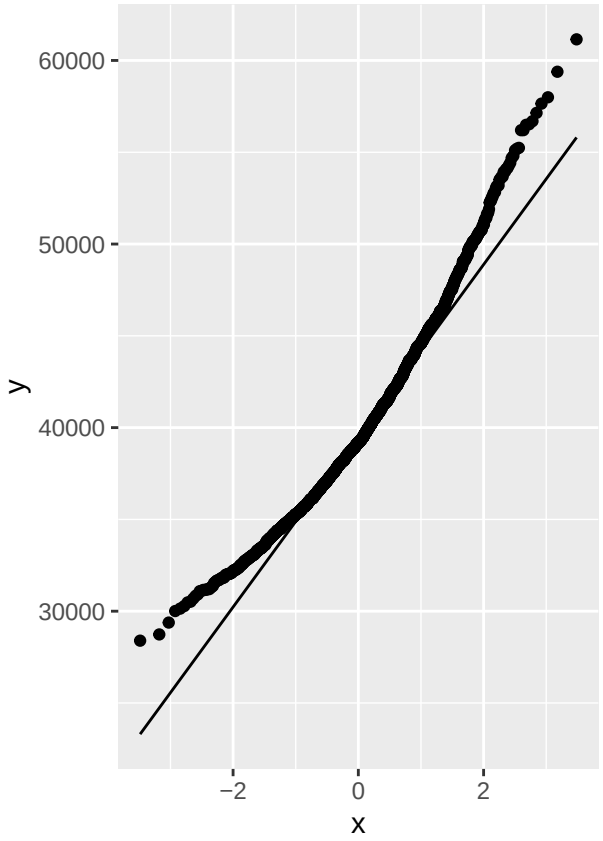

y
80000
70000
60000
50000
40000
30000
−2
0
2
x

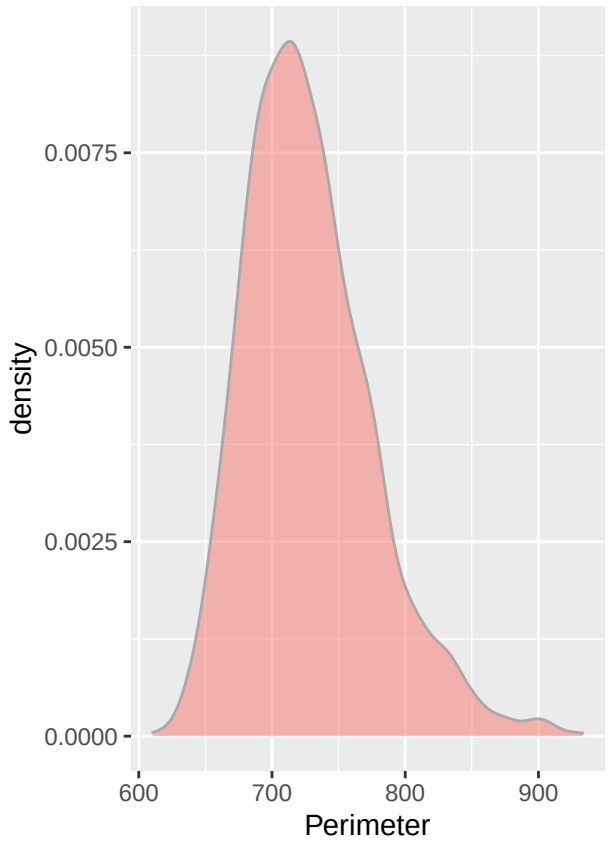

density
0.020
0.015
0.010
0.005
0.000
200
225
250
275
300
325
EquivDiameter

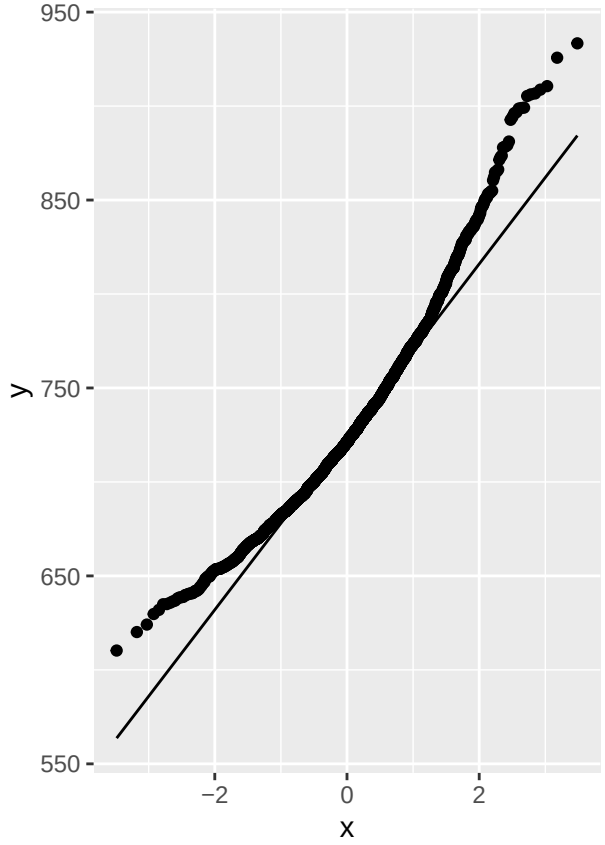

y
325
300
275
250
225
200
−2
0
2
x

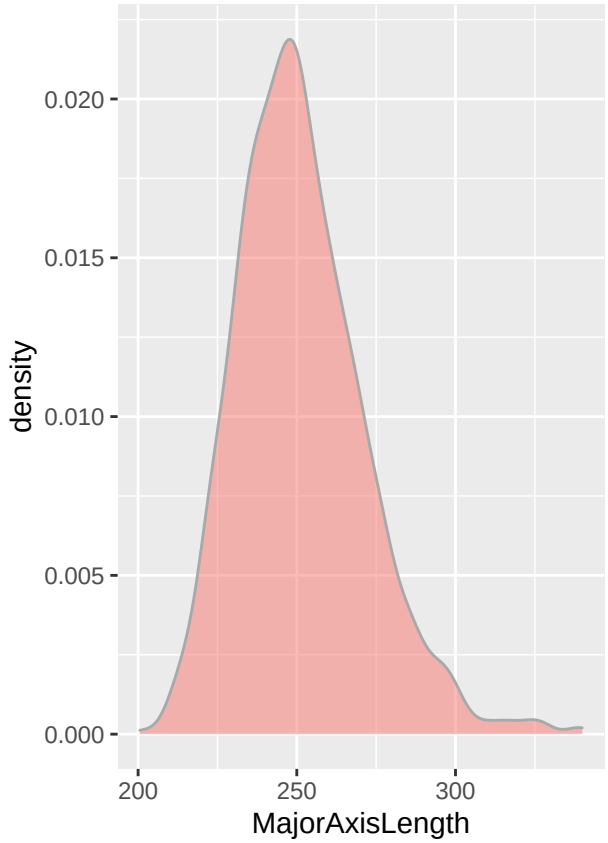
```
## [[1]]
## [1] "SEKER"
```

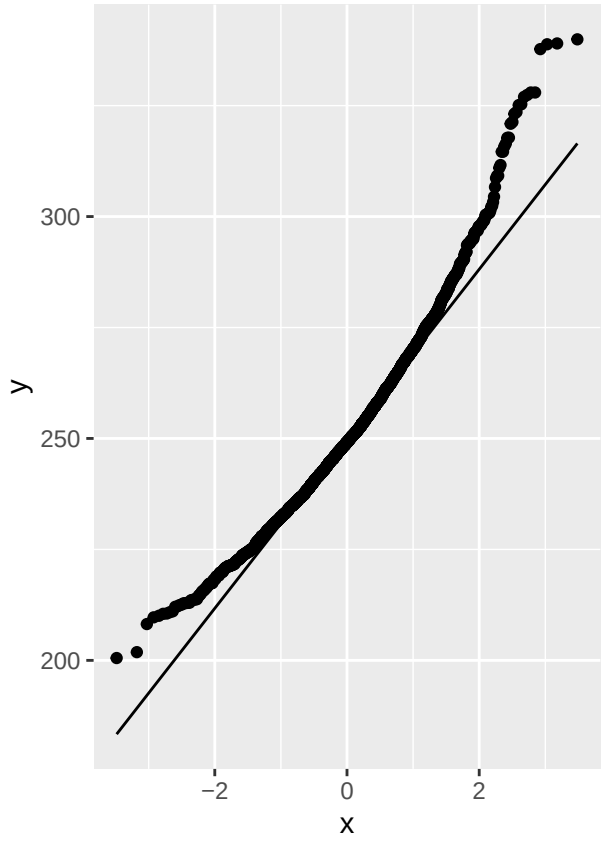

density
7.5e−05
5.0e−05
2.5e−05
0.0e+00
30000
40000
50000
60000
Area

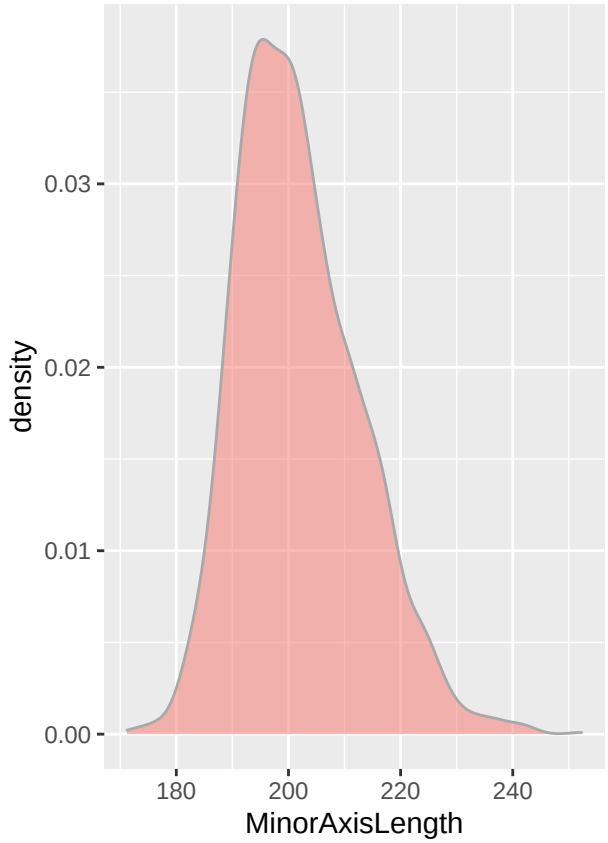

y
60000
50000
40000
30000
−2
0
2
x

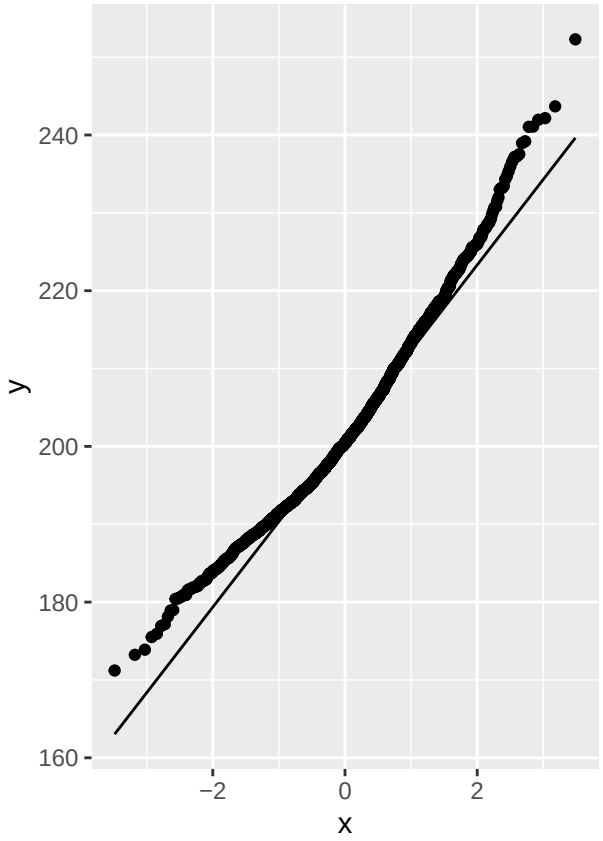

density
0.0075
0.0050
0.0025
0.0000
600
700
800
900
Perimeter

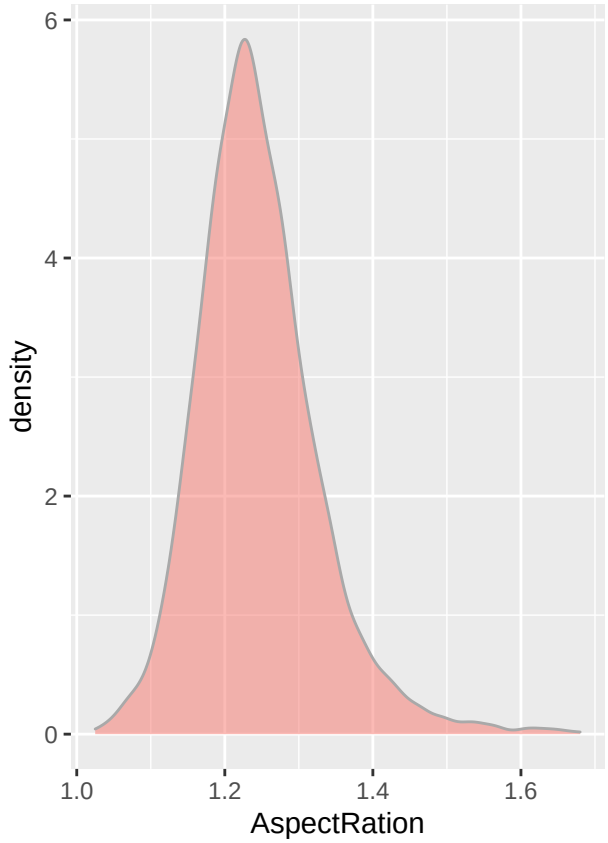

y
950
850
750
650
550
−2
0
2
x

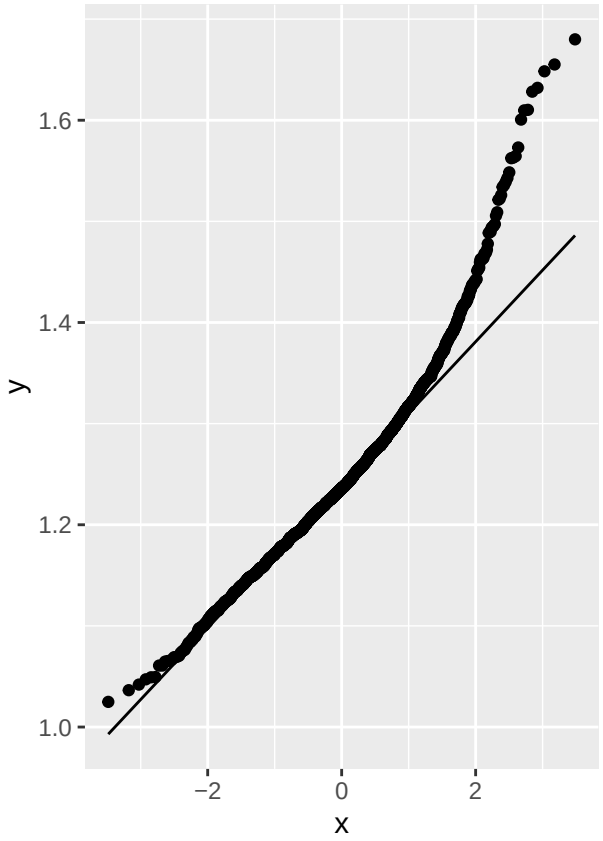

0.020
0.015
0.010
0.005
0.000
density
200
250
300
MajorAxisLength

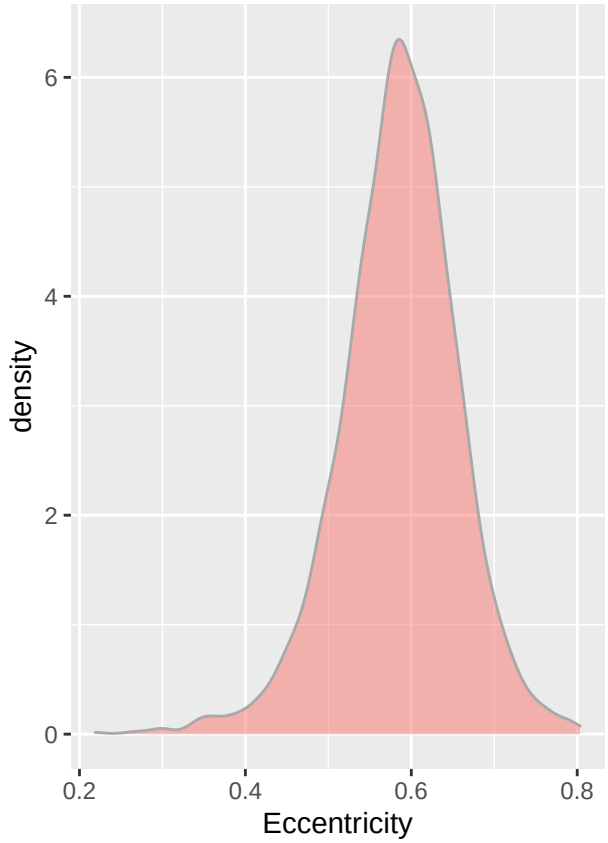

300
250
200
y
−2
0
2
x

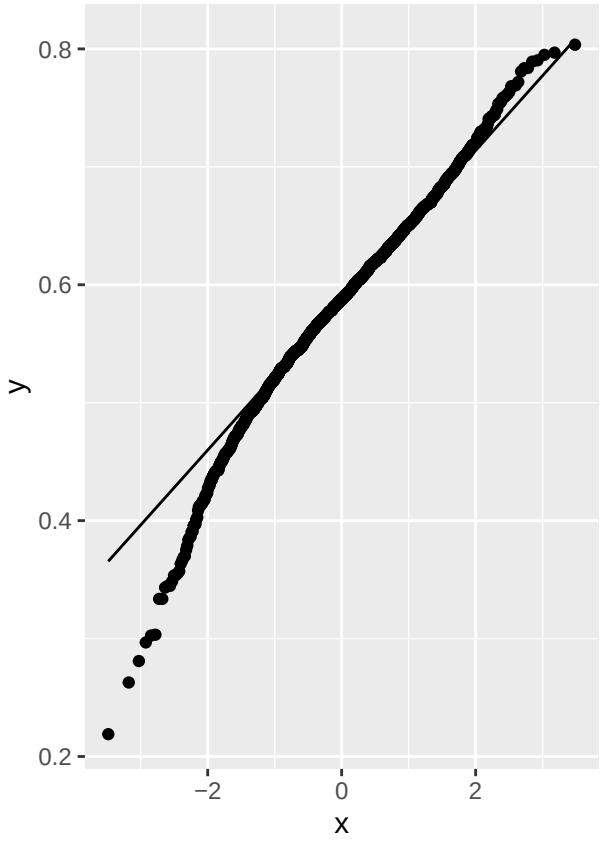

density
0.03
0.02
0.01
0.00
180
200
220
240
MinorAxisLength

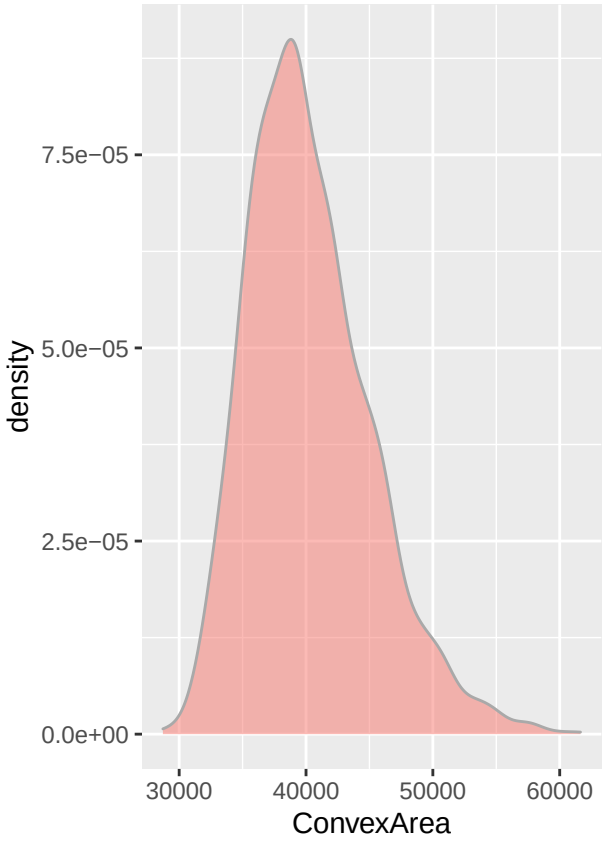

y
240
220
200
180
160
−2
0
2
x

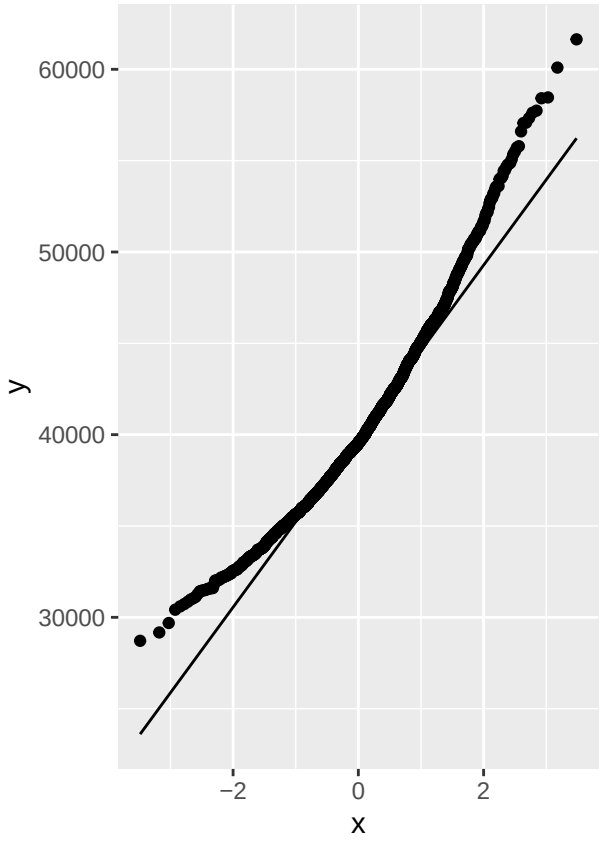

6
4
2
0
density
1.0
1.2
1.4
1.6
AspectRation

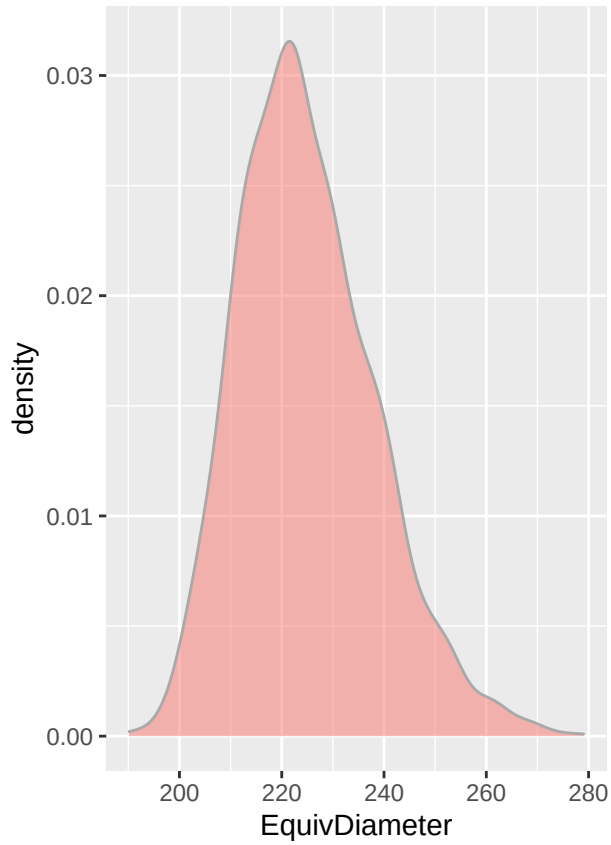

1.6
1.4
1.2
1.0
y
−2
0
2
x

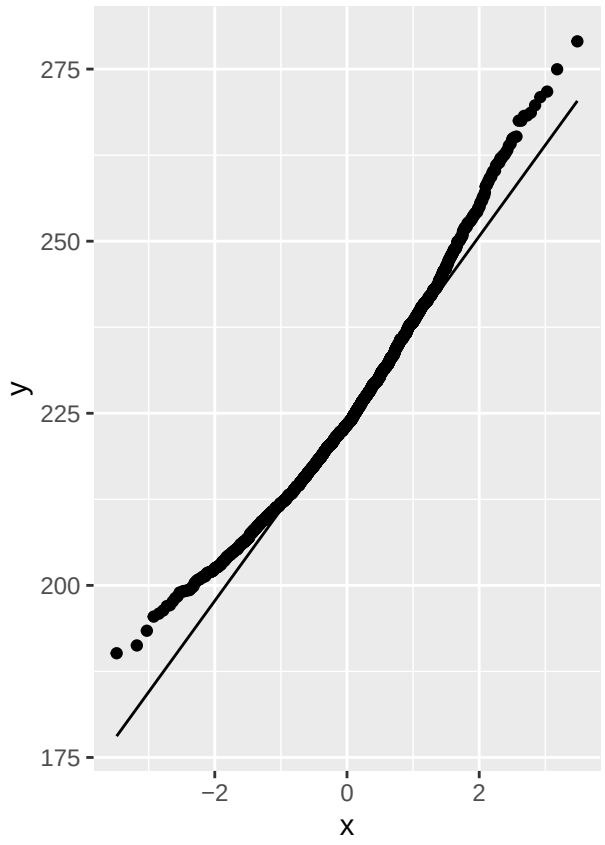

density
6
4
2
0
0.2
0.4
0.6
0.8
Eccentricity


y
0.8
0.6
0.4
0.2
−2
0
2
x

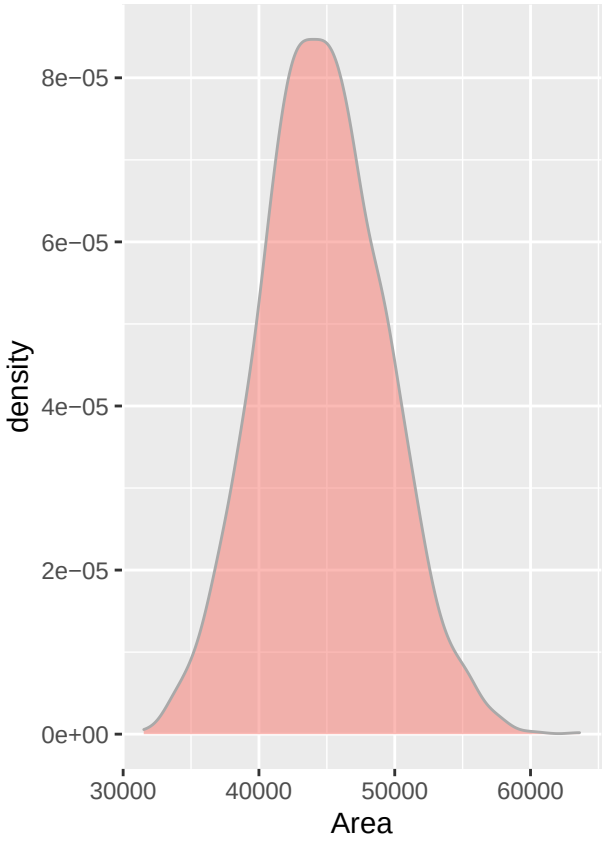

7.5e−05
5.0e−05
2.5e−05
0.0e+00
density
30000
40000
50000
60000
ConvexArea

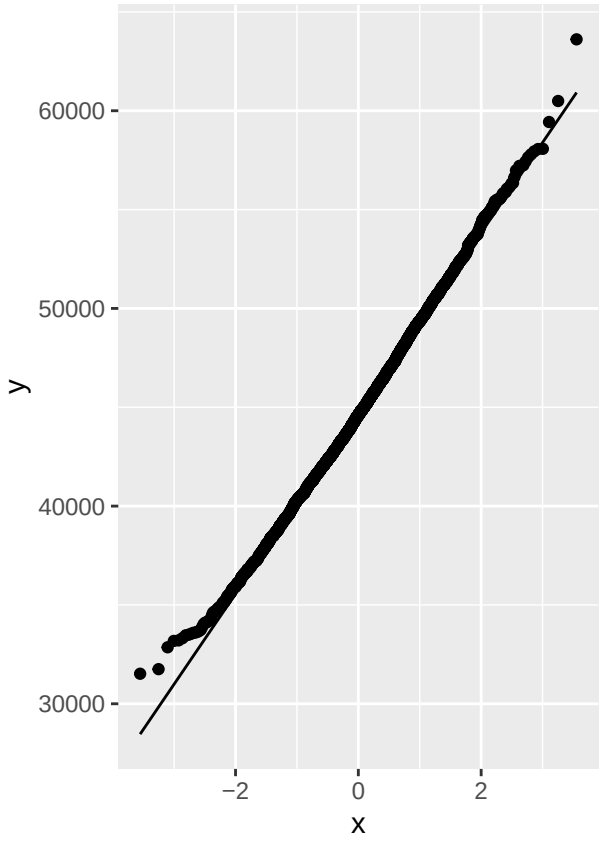

60000
50000
40000
30000
y
−2
0
2
x

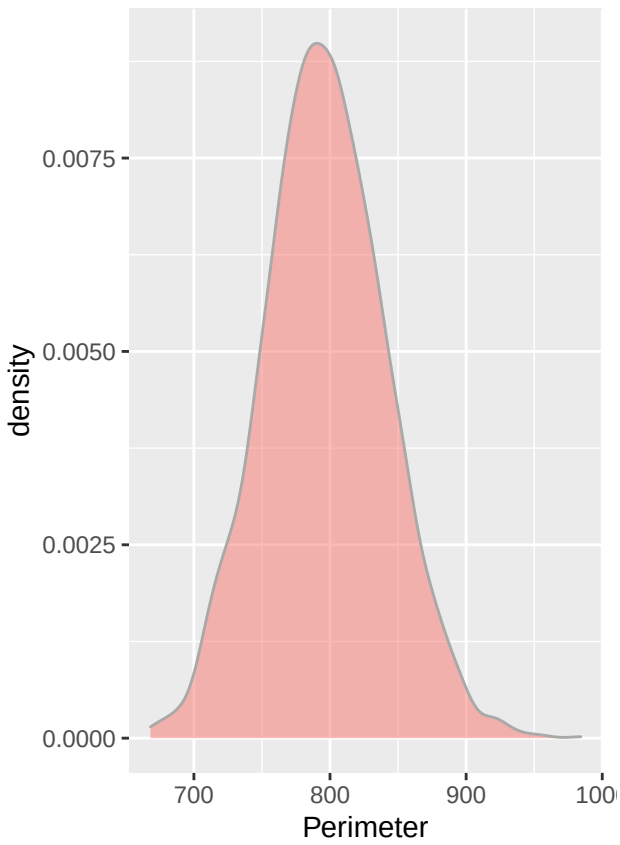

0.03
0.02
0.01
0.00
density
200
220
240
260
280
EquivDiameter

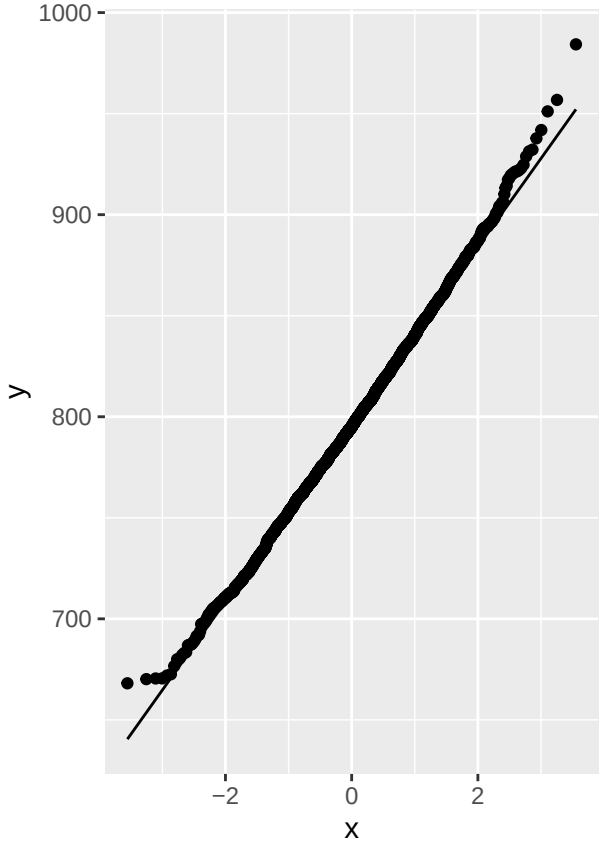

275
250
225
200
175
y
−2
0
2
x

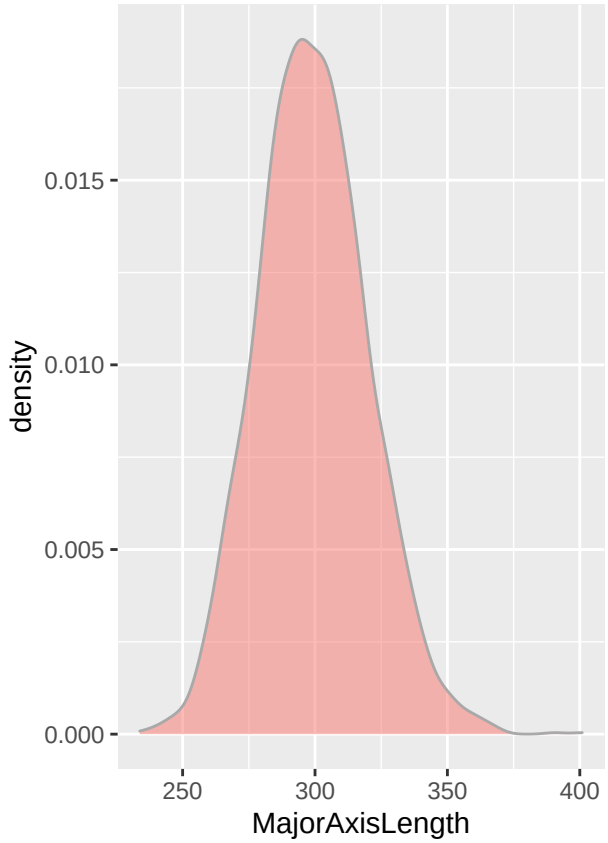
```
## [[1]]
## [1] "SIRA"
```

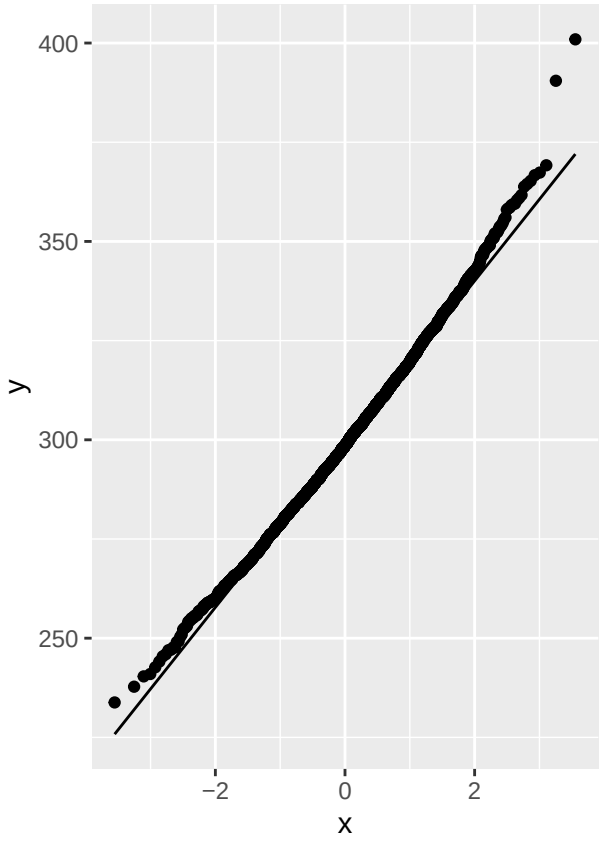

8e−05
6e−05
4e−05
2e−05
0e+00
density
30000
40000
50000
60000
Area

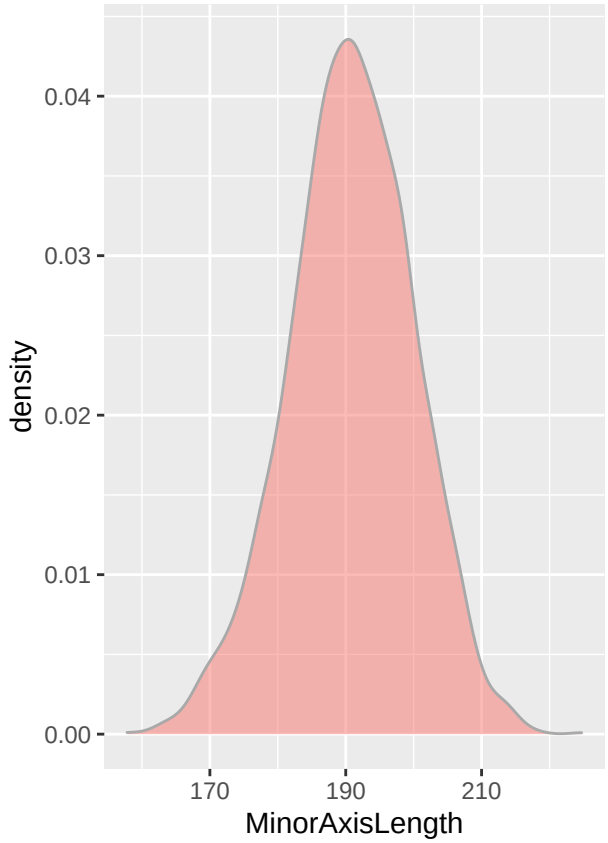

60000
50000
40000
30000
y
−2
0
2
x

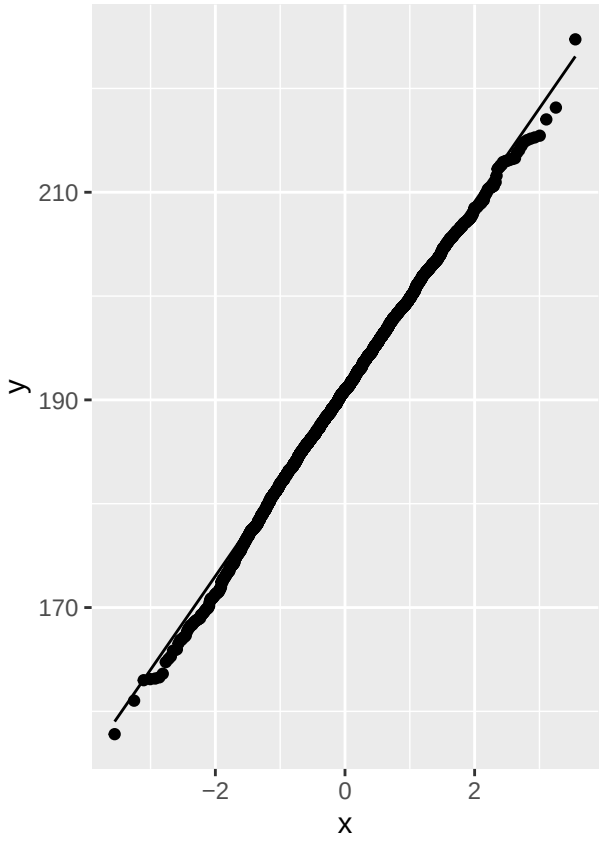

density
0.0075
0.0050
0.0025
0.0000
700
800
900
Perimeter

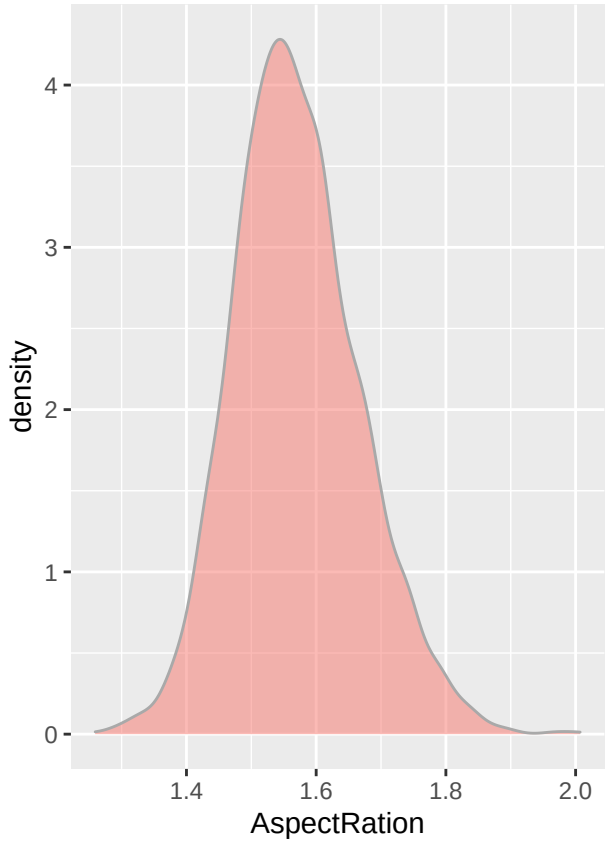

1000
900
800
700
y
−2
0
2
x

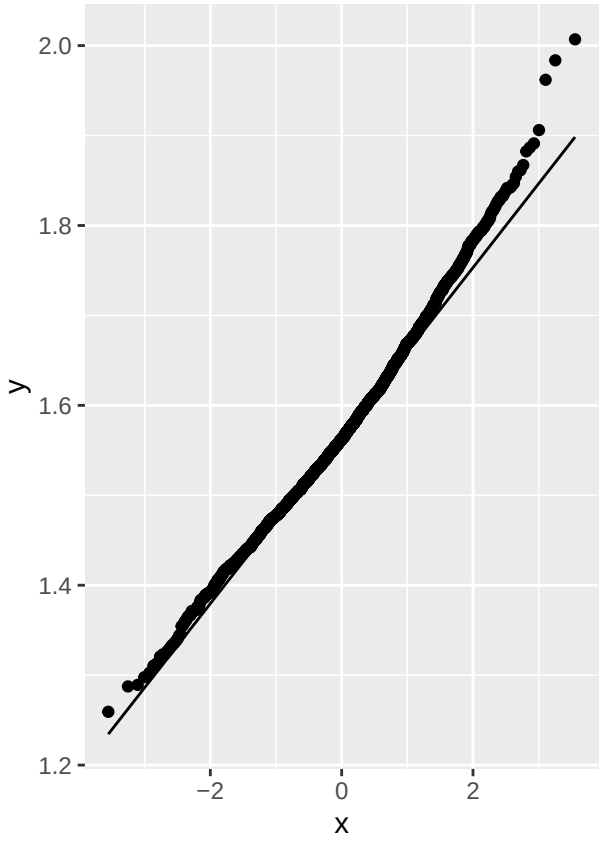

density
0.015
0.010
0.005
0.000
250
300
350
400
MajorAxisLength

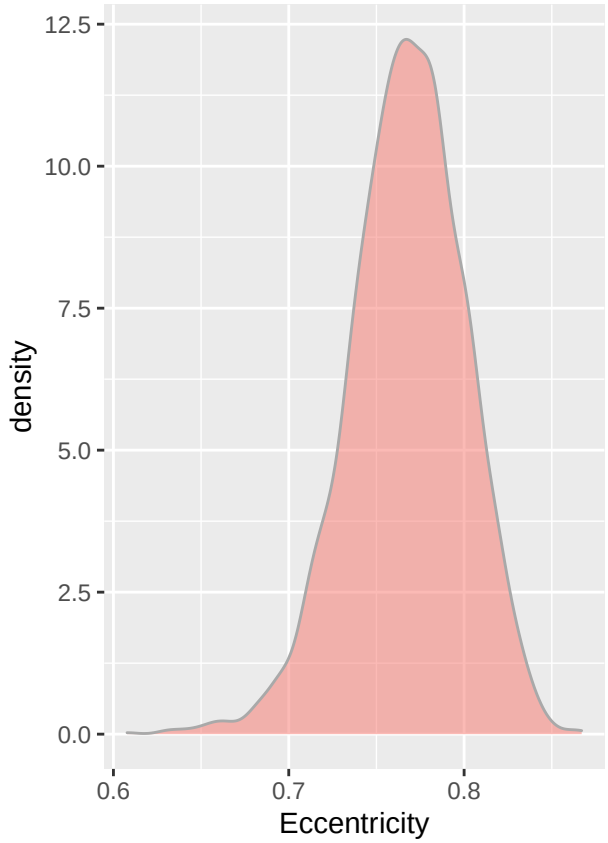

y
400
350
300
250
−2
0
2
x

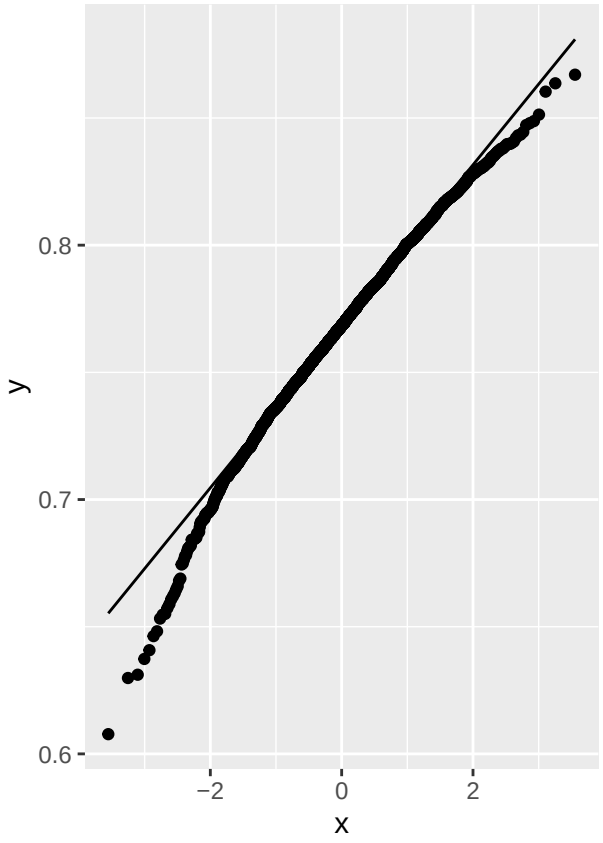

0.04
0.03
0.02
0.01
0.00
density
170
190
210
MinorAxisLength

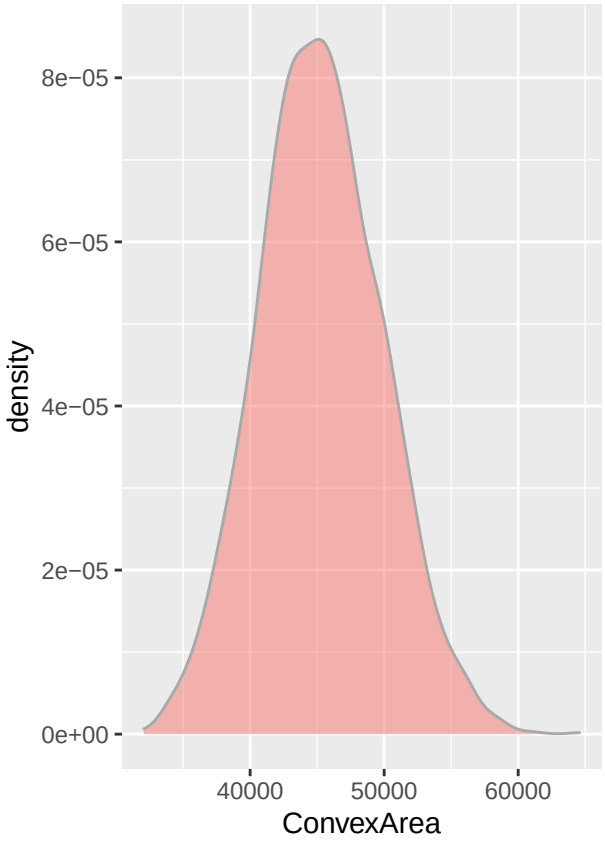

210
190
170
y
−2
0
2
x

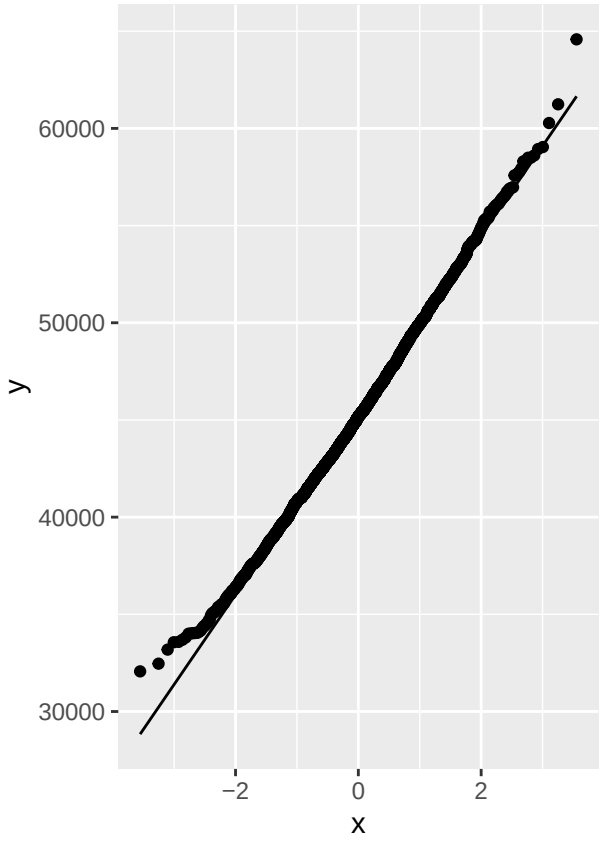

density
AspectRation
0
1
2
3
4
1.4
1.6
1.8
2.0

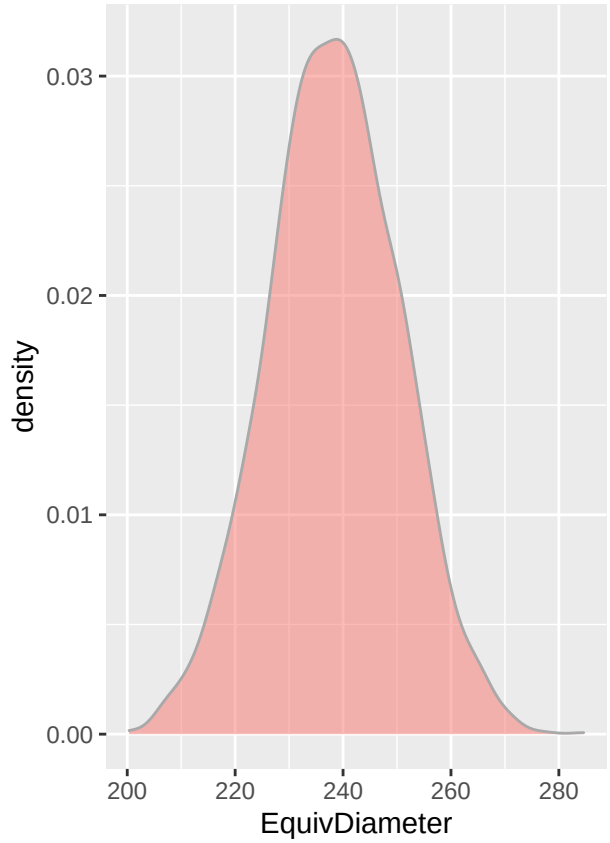

y
x
1.2
1.4
1.6
1.8
2.0
−2
0
2

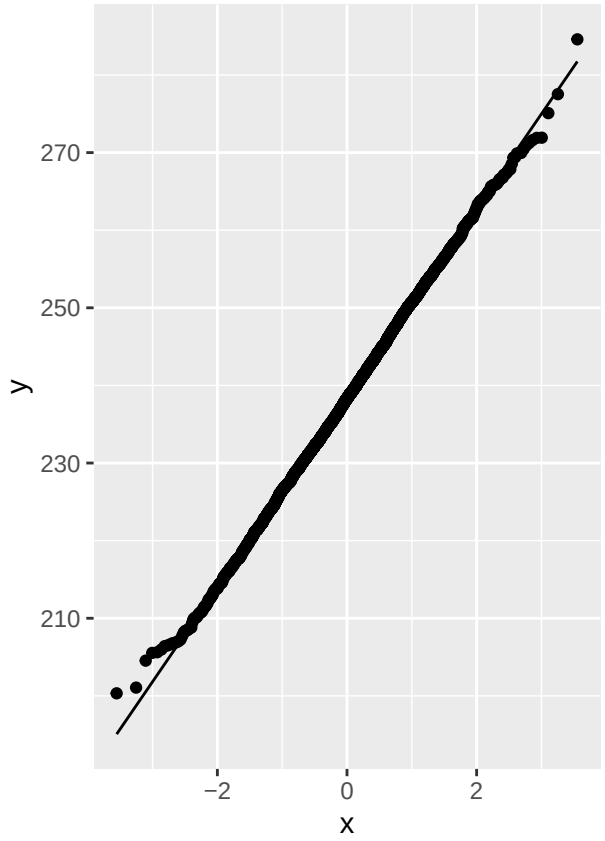

12.5
10.0
7.5
5.0
2.5
0.0
density
0.6
0.7
0.8
Eccentricity


0.8
0.7
0.6
y
−2
0
2
x

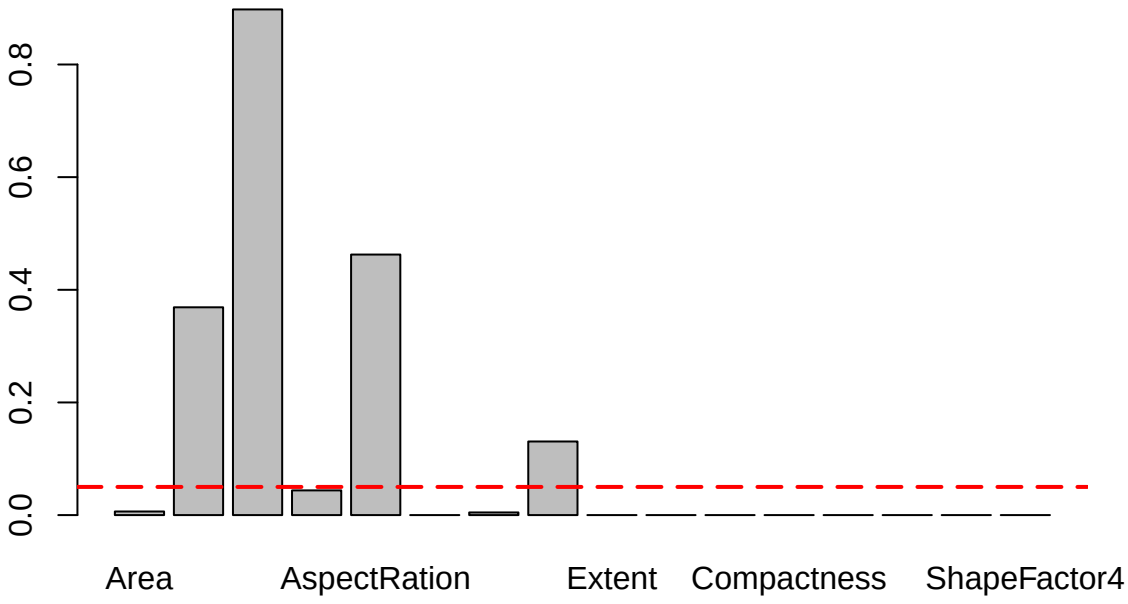

8e−05
6e−05
4e−05
2e−05
0e+00
density
40000
50000
60000
ConvexArea

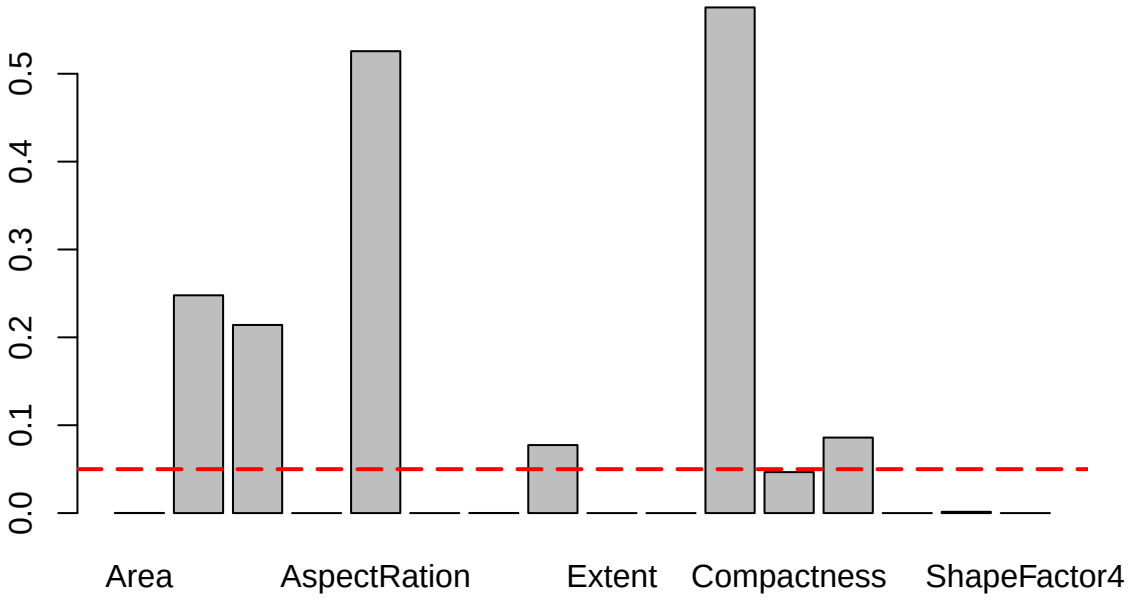

60000
50000
40000
30000
y
−2
0
2
x

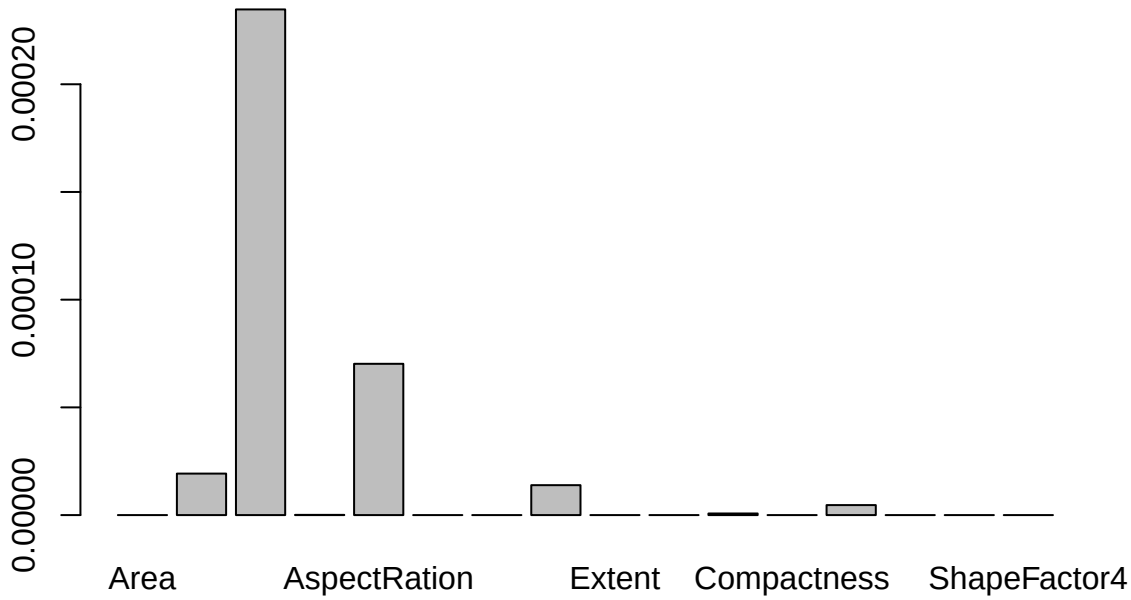


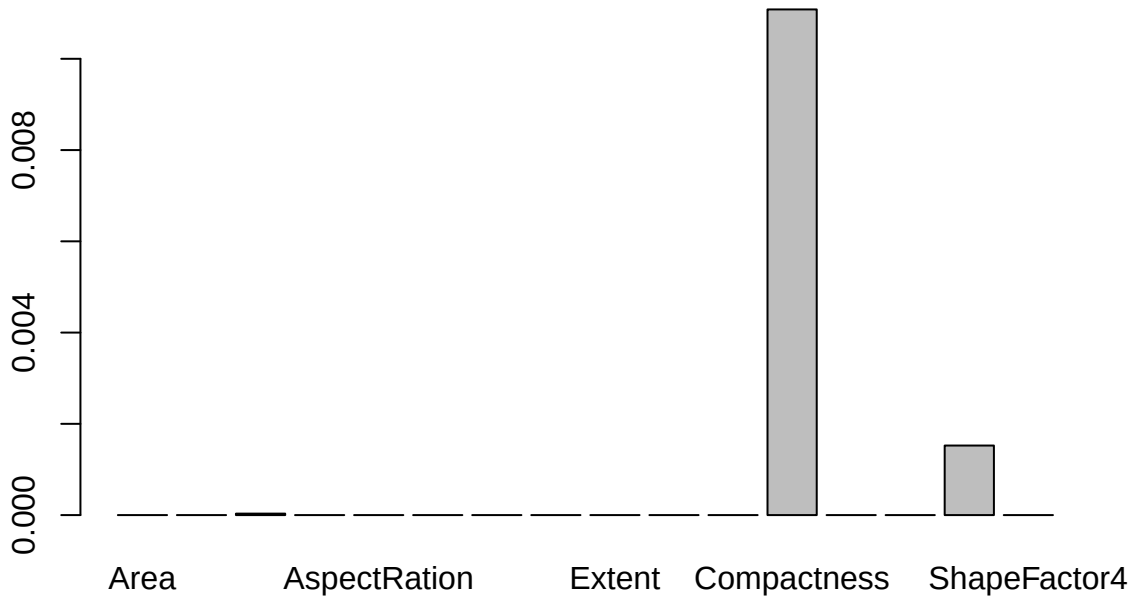


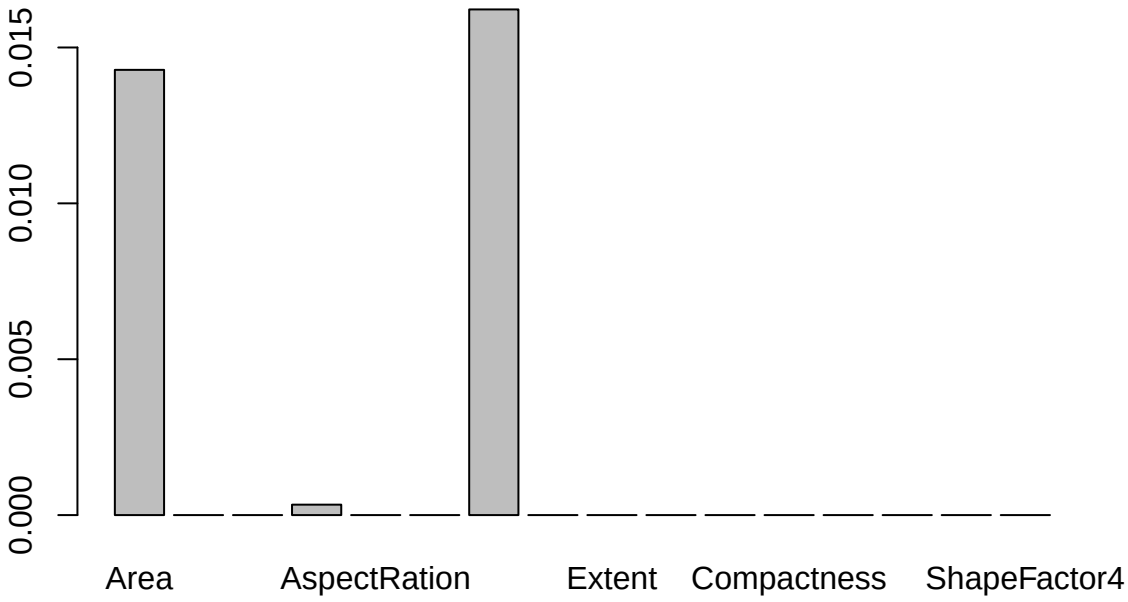

```
leves = list("BARBUNYA", "BOMBAY", "CALI", "DERMASON", "HOROZ", "SEKER", "SIRA")
pvalue = c()
for(j in 1:7){
  for(i in 1:16){
    pvalue[i] = shapiro.test(dados[dados$Class==leves[j],i])$p.value
  }
  barplot(pvalue,names.arg = names(dados)[1:16])
  abline(h=0.05, col = "Red", lty = 5, lwd = 2)
}
```

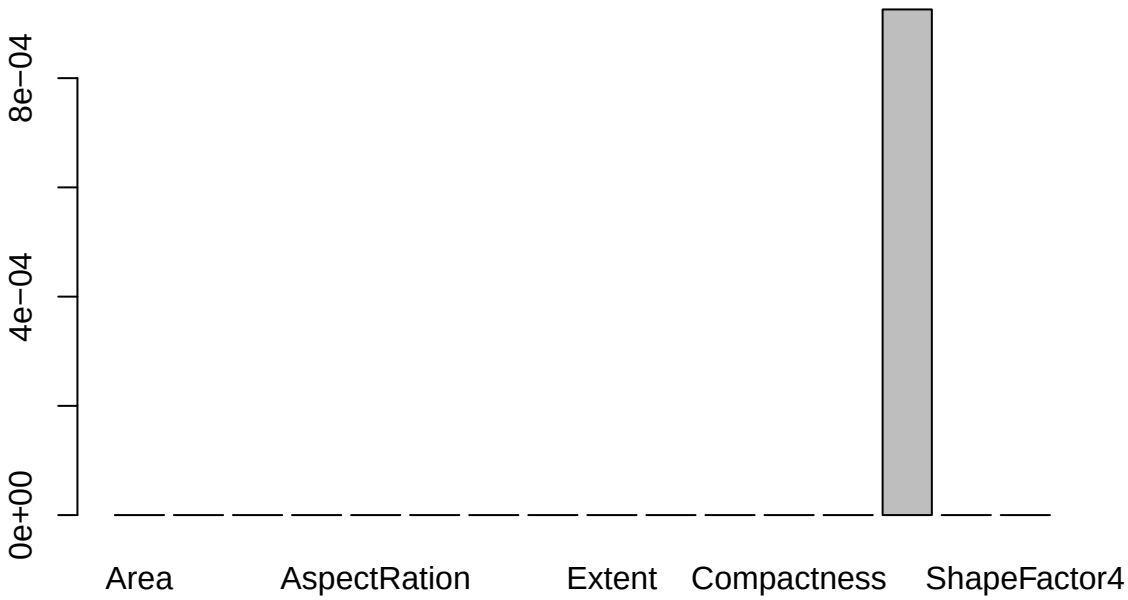

0.0
0.2
0.4
0.6
0.8
Area
AspectRation
Extent
Compactness
ShapeFactor4

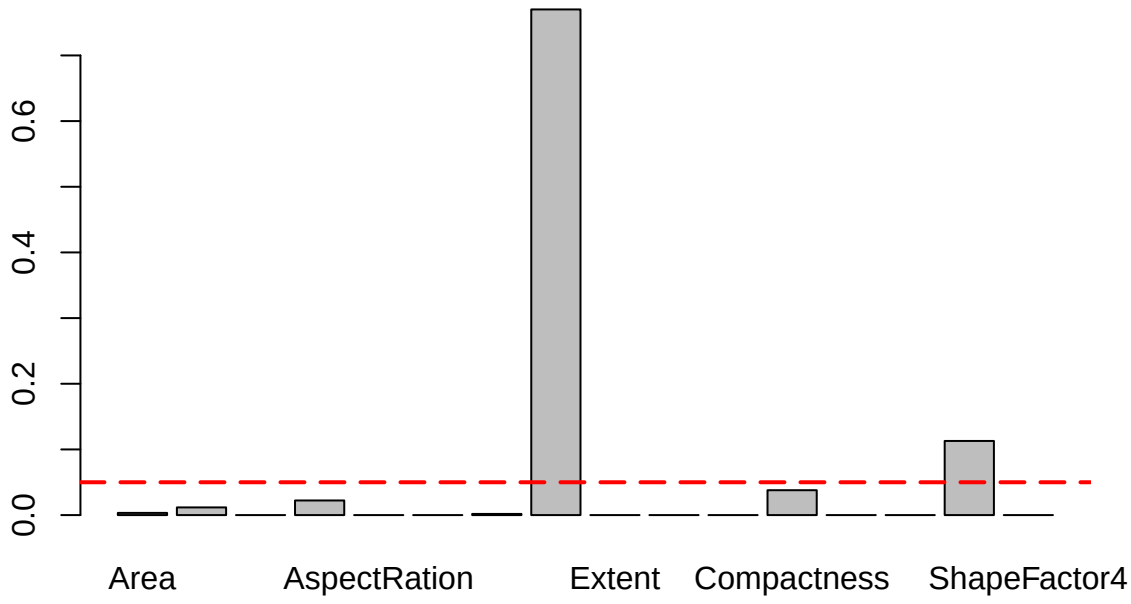

0.0
0.1
0.2
0.3
0.4
0.5
Area
AspectRation
Extent
Compactness
ShapeFactor4


0.00000
0.00010
0.00020
Area
AspectRation
Extent
Compactness
ShapeFactor4


0.000
0.004
0.008
Area
AspectRation
Extent
Compactness
ShapeFactor4


0.000
0.005
0.010
0.015
Area
AspectRation
Extent
Compactness
ShapeFactor4

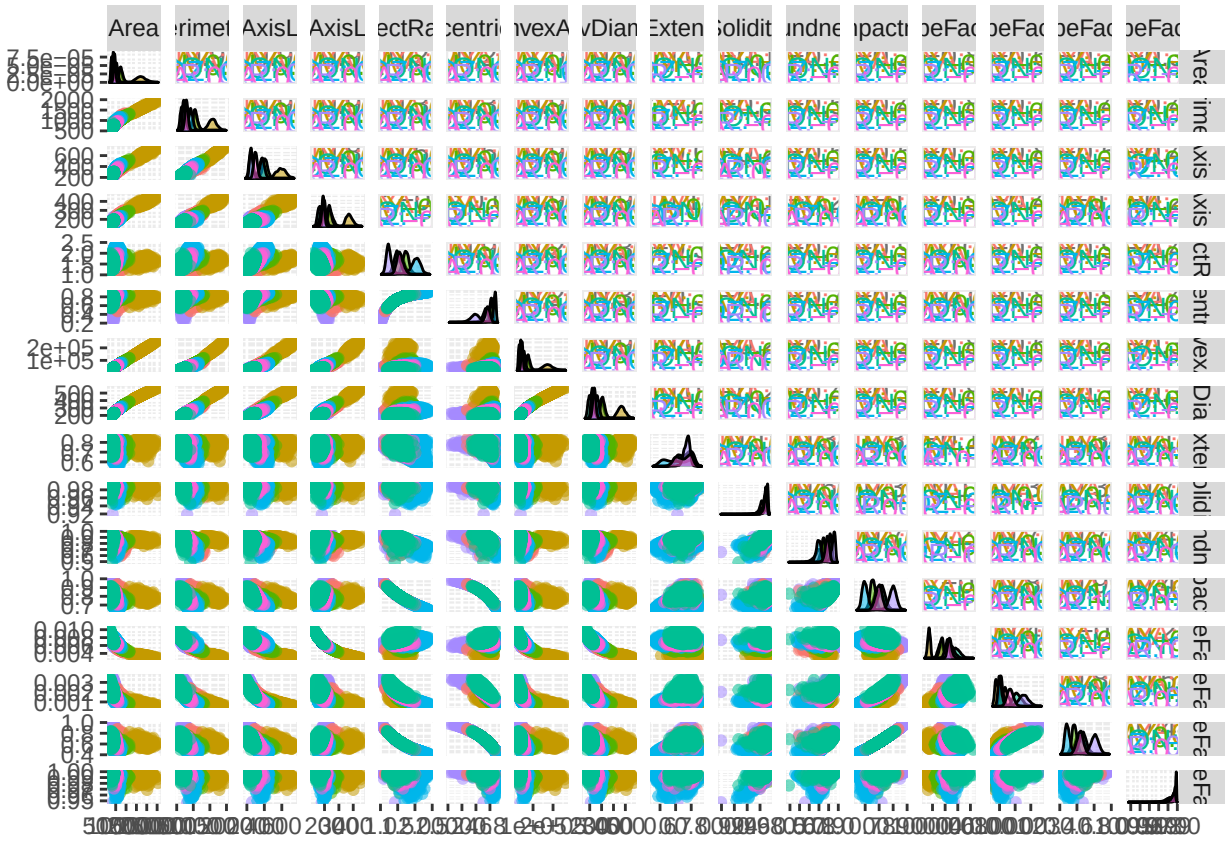

0e+00
4e−04
8e−04
Area
AspectRation
Extent
Compactness
ShapeFactor4



Em nenhuma das classes possui uma distribuição normal, em todas as classes há uma variável que possui uma discrêpancia dos dados.

8.Utilizando um gráfico de dispersão entre pares de variáveis, diga se existe alguma associação entre variáveis que permite uma maior discriminação?

```
ggpairs(dados[,1:16], aes(colour = dados$Class, alpha = 0.4))
```



Há discriminação de algumas classes em determinadas variáveis com maiores diferenças que em outras mesmo desconsiderando os outliers.

9.Após ter analisado estas informações, quais considerações você faz sobre este conjunto de dados (ou tarefa)?

Com relação à classificação das sementes de feijão, as características de dimensão e forma das variedades de feijão não possuem características discriminatórias externas, o que torna esse processo de classificação complexo.