BIE5782

Unidade 5:

ANÁLISE EXPLORATÓRIA

Análise Exploratória de Dados (AED)

- ✓ O que é AED?
- ✓ Quem utiliza AED?
- ✓ Objetivos da AED:
  - ➢ controle de qualidade dos dados
  - ➢ detectar padrões
  - ➢ orientar os testes de hipóteses
  - ➢ ajudar a interpretar os testes de hipóteses
- ✓ Apresentação gráfica é muito útil para AED

`summary()`, `str()`, `head()`, `tail()`

A Primeira Olhada

DEMONSTRAÇÃO NO R

is.na()

Teste Lógico para Valores Perdidos

```
> a
 [1]  1  2  3  4  5 NA  6  7  8  9 10 NA
> is.na(a)
 [1] FALSE FALSE FALSE FALSE FALSE  TRUE
 [7] FALSE FALSE FALSE FALSE FALSE  TRUE
> a[is.na(a)==F]
 [1]  1  2  3  4  5  6  7  8  9 10
> a[is.na(a)==T] <- 0
> a
 [1]  1  2  3  4  5  0  6  7  8  9 10  0
```

mean(), median()

Medidas de Tendência Central

```
> mean( c(0,1,2,3,4,5))
[1] 2.5
> median( c(0,1,2,3,4,5))
[1] 2.5
> mean( c(0,1,2,3,4,100))
[1] 18.33333
> median( c(0,1,2,3,4,100))
[1] 2.5
```

`mean(trim=)`, `mean()`, `median()`, `quantile()`

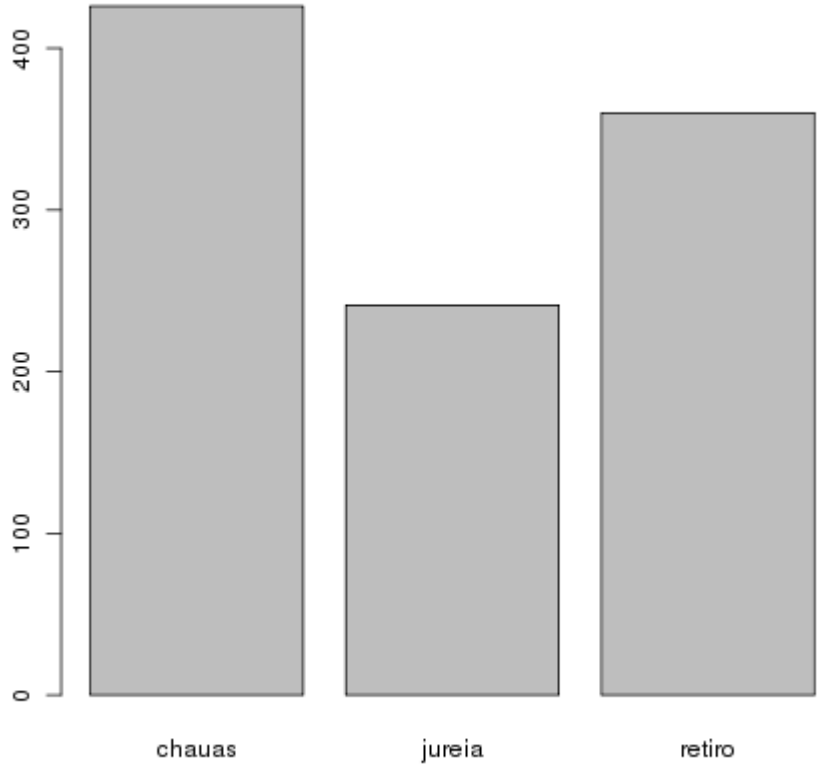
Média (normal e truncada) mediana, quantis: o pacote básico.

DEMONSTRAÇÃO NO R

table(), barplot()

Contagens de Fatores

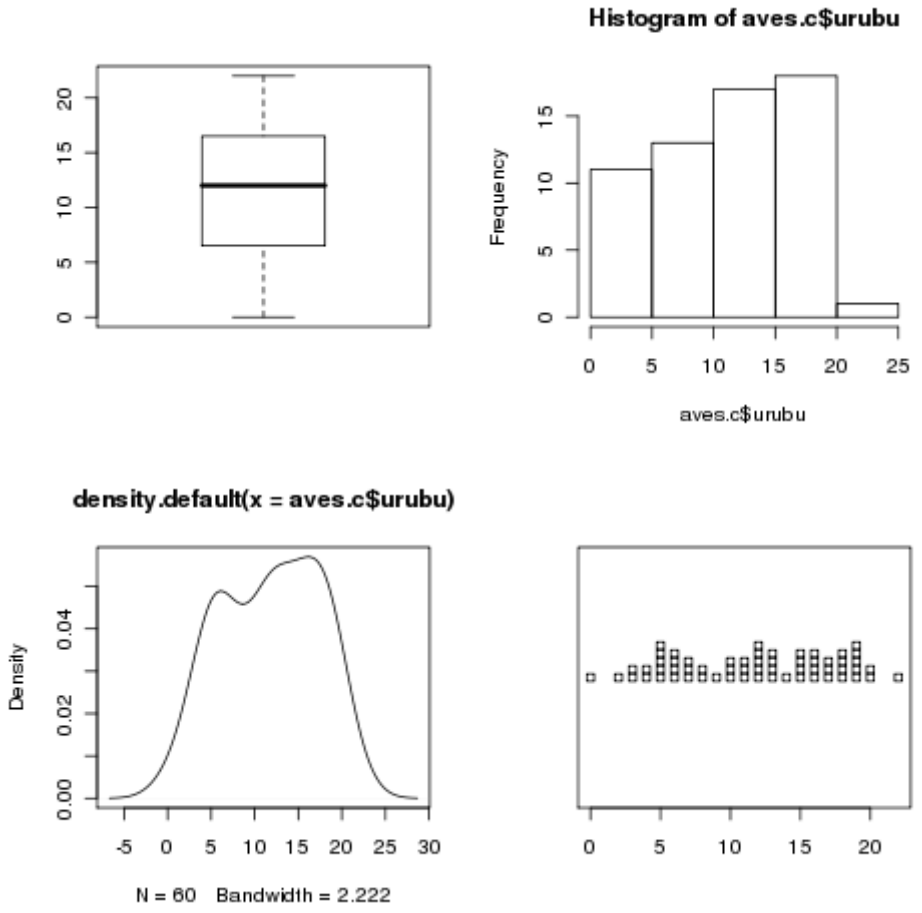
```
> table(caixeta$local)

chauas jureia retiro
   426    241    360
```

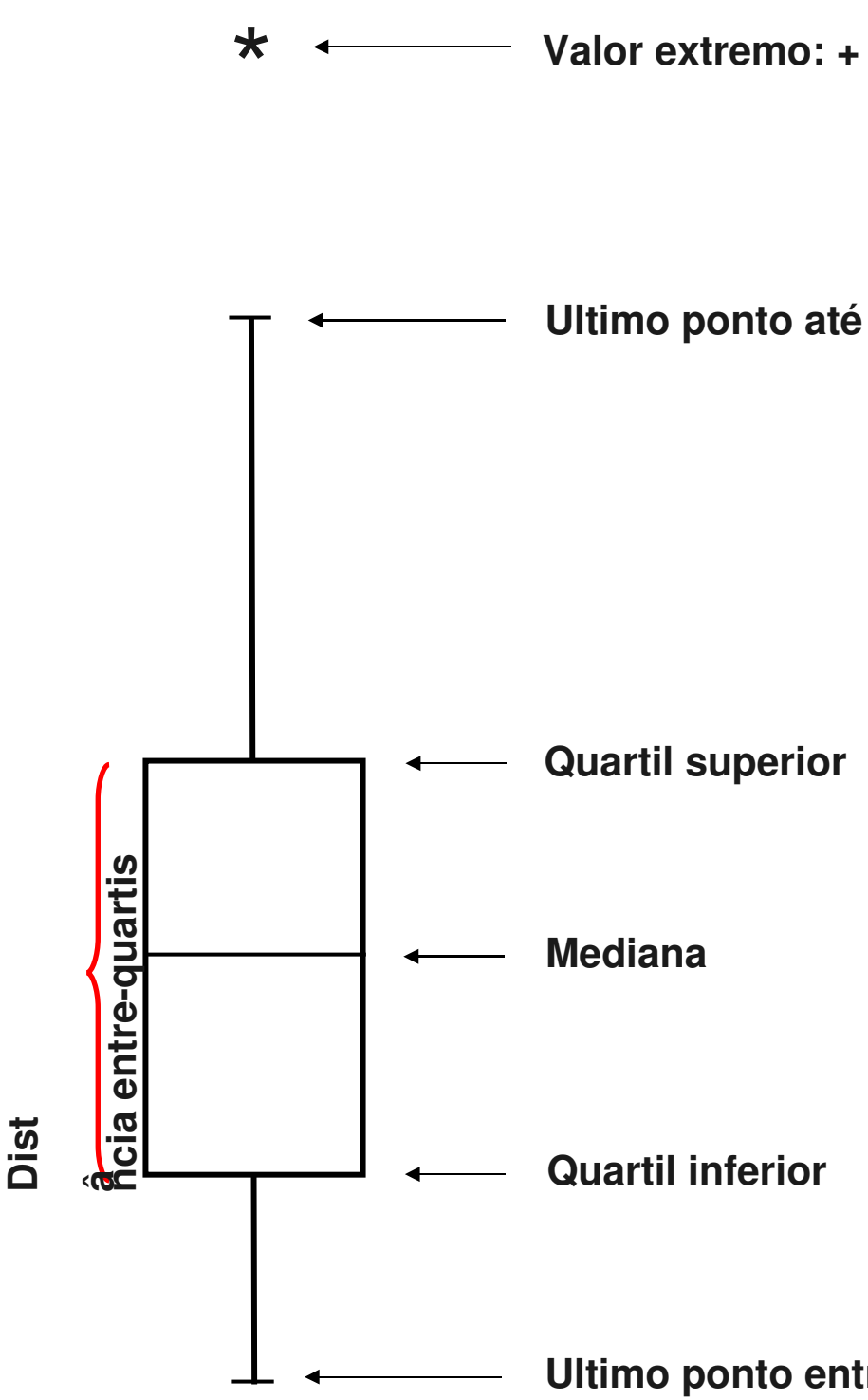



```
> barplot(table(caixeta$local))
```

`boxplot()`, `hist()` , `density()`, `stripchart()`

Gráficos Univariados Básicos

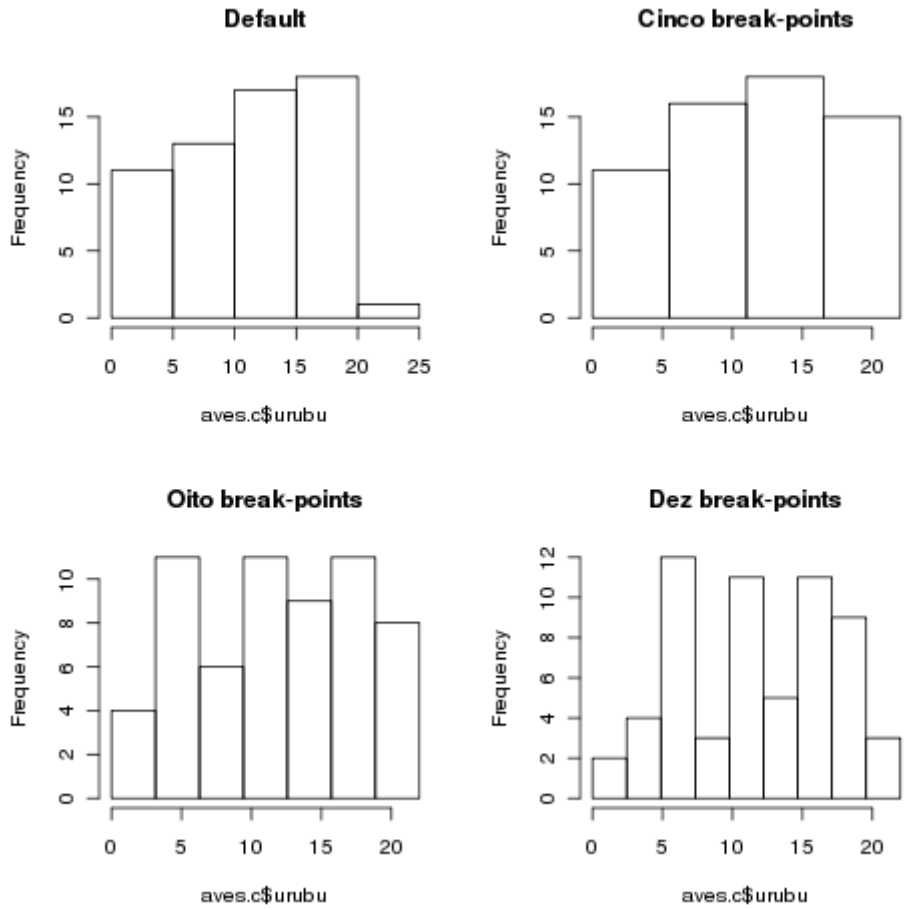




Box-and-whisker plot

ou

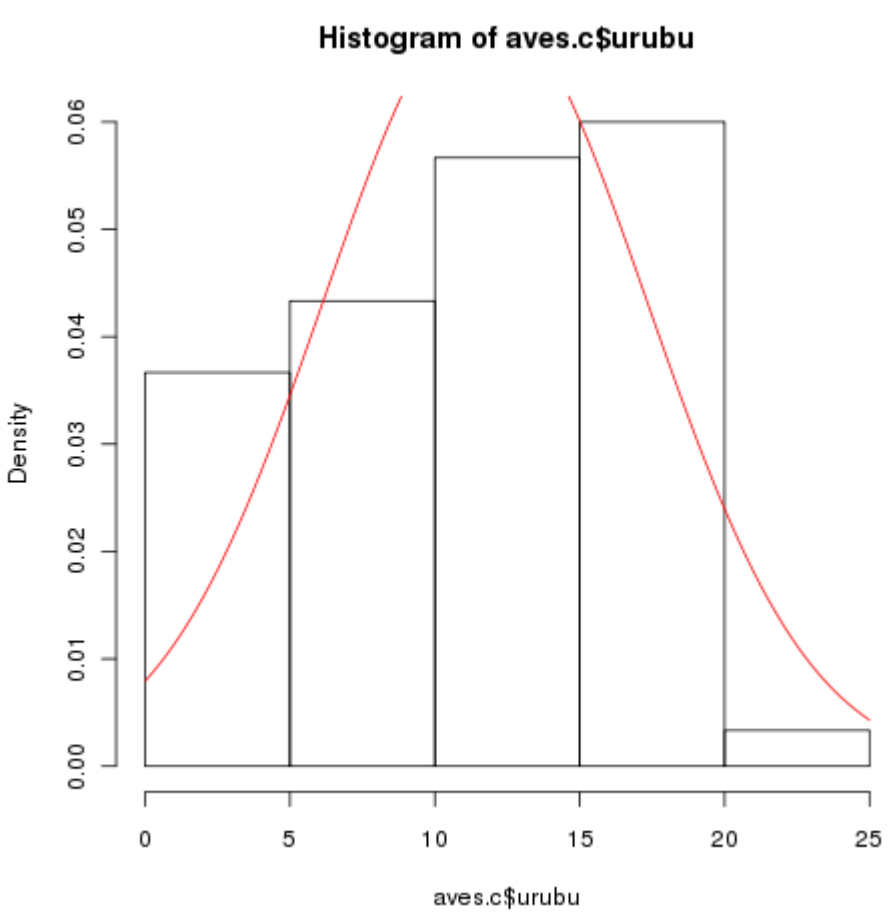
Diagrama de caixa e bigode de gato

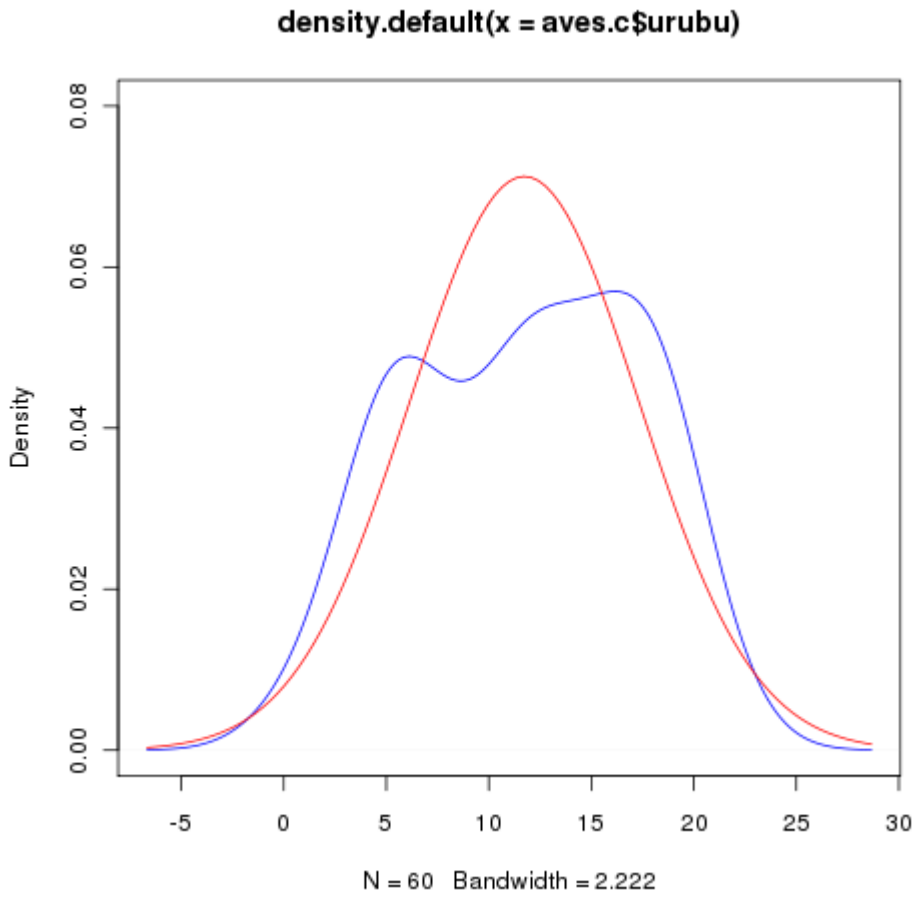



```
hist(aves.c$urubu)
hist(aves.c$urubu,breaks=seq(0,max(aves.c$urubu),length=5))
hist(aves.c$urubu,breaks=seq(0,max(aves.c$urubu),length=8))
hist(aves.c$urubu,breaks=seq(0,max(aves.c$urubu),length=10)
```

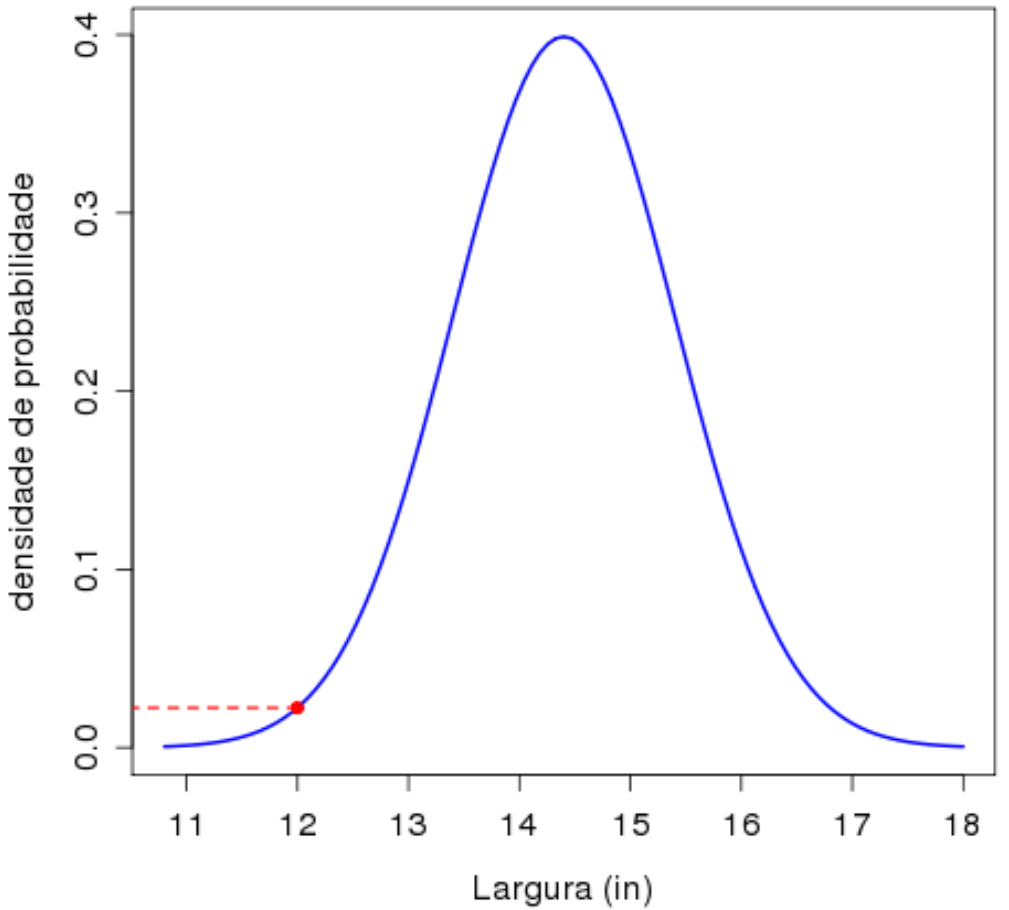
`hist(prob=T)` , `density()` , `curve()`

Curvas Empíricas e Teóricas de Densidade





Relembrando ...




```
> dnorm(x=12,mean=14.4,sd=1)
[1] 0.02239453
```

Vai para o R!

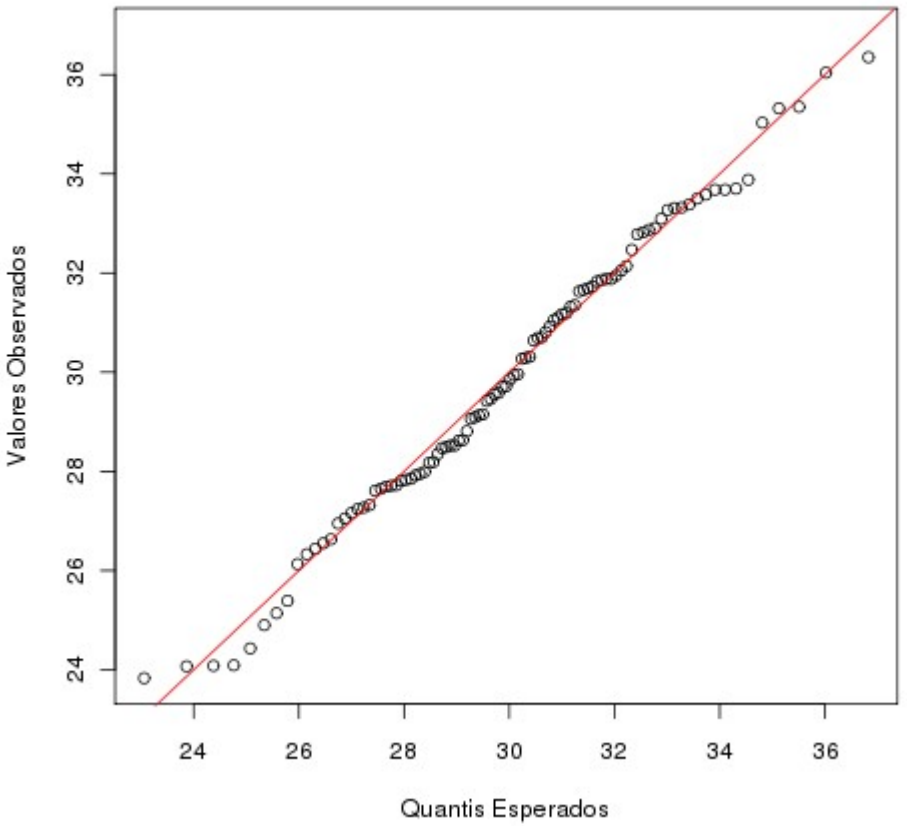
qqnorm() , qqline()

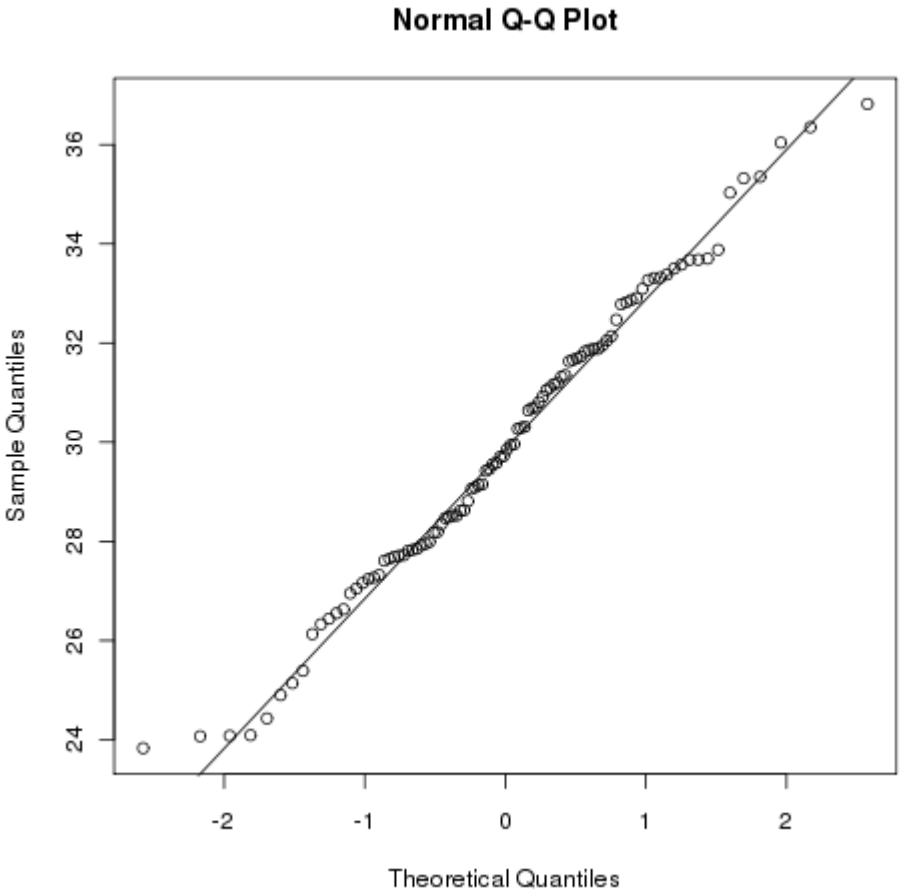
O melhor teste de normalidade

	x	percentil	q.norm
1	23.83	0.01	23.05859
2	24.07	0.02	23.86540
3	24.08	0.03	24.37730
4	24.09	0.04	24.76238
5	24.43	0.05	25.07561
...			
95	35.03	0.95	34.81219
96	35.32	0.96	35.12542
97	35.35	0.97	35.51050
98	36.04	0.98	36.02240
99	36.35	0.99	36.82921
100	36.82	1.00	Inf

`qqnorm()` , `qqline()`

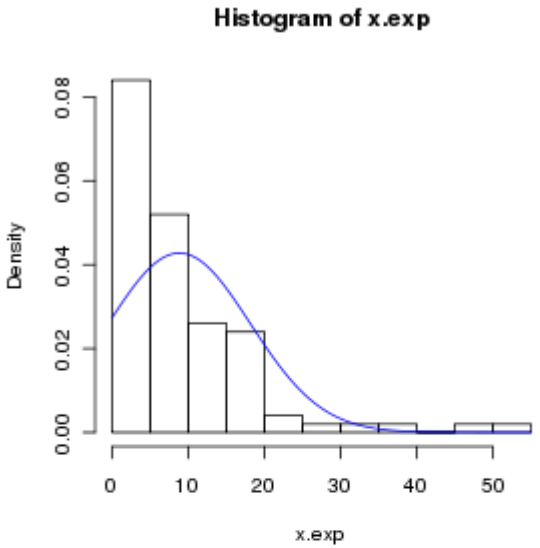
O melhor teste de normalidade

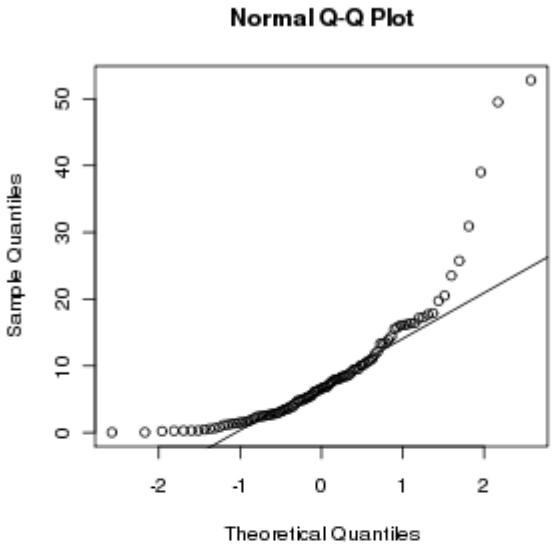


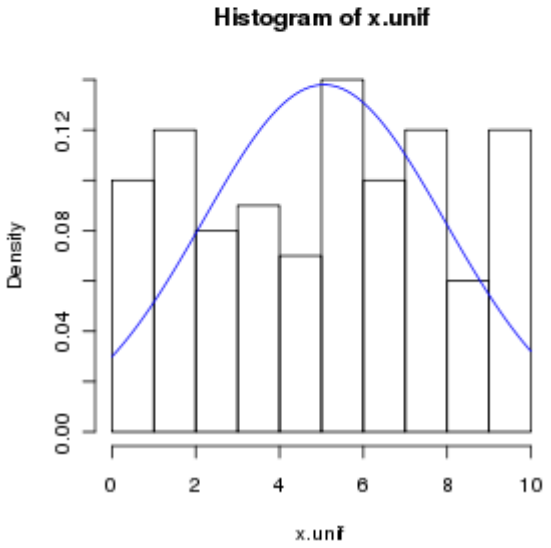


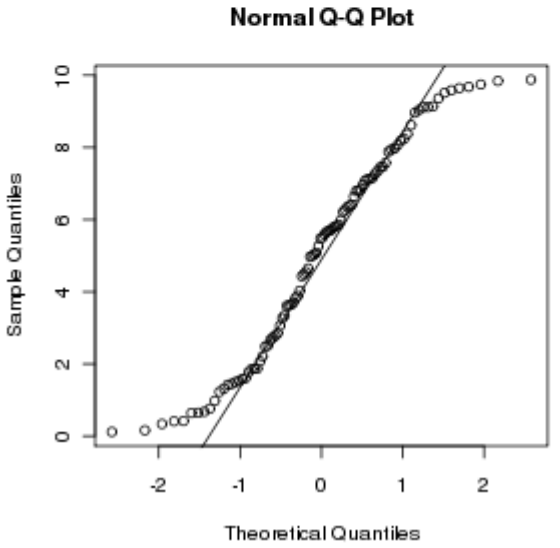
`qqnorm()` , `qqline()`

O melhor teste de normalidade









DUAS OU MAIS VARIÁVEIS

table()

Tabelas de Contingência

```
> table(caixeta$especie,caixeta$local)

                          chauas jureia retiro
  Alchornea triplinervia       0       3     12
  Andira fraxinifolia          0       4      0
  bombacaceae                  0       1      0
  Cabralea canjerana           0       4      0
  Callophyllum brasiliensis    7       0      0
  Calophyllum brasiliensis     0       4      0
  Cecropia sp                  0       0      1
  Coussapoa macrocarpa         0       3      0
  Coussapoa micropoda          2       0      7
  Cryptocaria moschata         0       2      0
  Cyathea sp                   0       0      2
```

xtabs()

Tabulação de Frequências

```
> head(Titanic.df)
  Class    Sex   Age Survived Freq
1   1st   Male Child       No    0
2   2nd   Male Child       No    0
3   3rd   Male Child       No   35
4  Crew   Male Child       No    0
5   1st Female Child       No    0
6   2nd Female Child       No    0
> xtabs(Freq~Sex+Survived, data=Titanic.df)

        Survived
Sex        No  Yes
  Male   1364  367
  Female  126  344
```

aggregate()

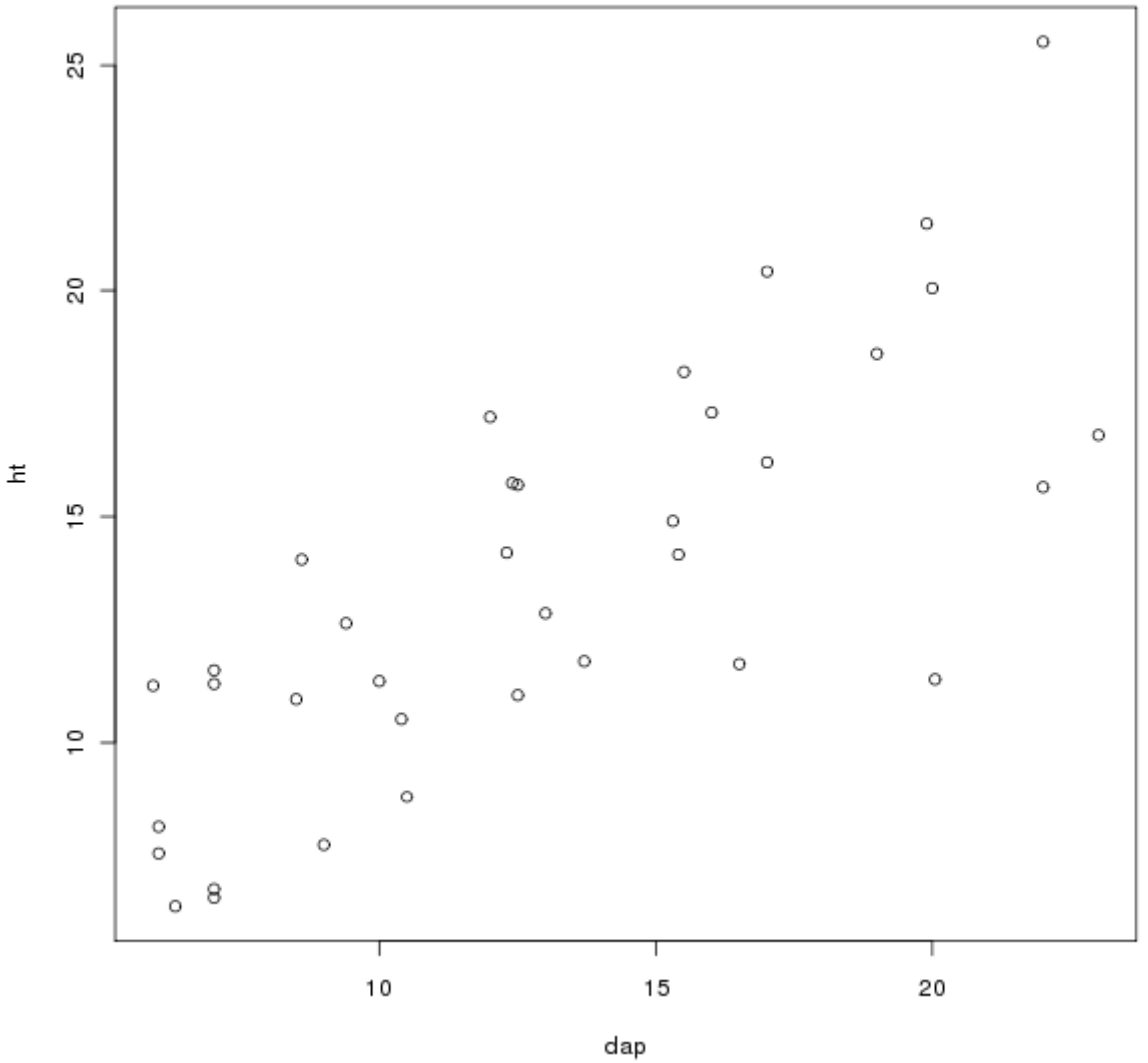
"Tabelas Dinâmicas"

```
> names(caixeta)
[1] "local"   "parcela" "arvore"  "fuste"   "cap"
[5] "h"       "especie" "ab"

> caixeta.alt <- aggregate(caixeta$h,
+ by=list(local=caixeta$local,
+ especie=caixeta$especie),FUN=max)

> head(caixeta.alt)
   local                      especie   x
1 jureia    Alchornea triplinervia 140
2 retiro    Alchornea triplinervia 100
3 jureia       Andira fraxinifolia  90
4 jureia                bombacaceae 150
5 jureia         Cabralea canjerana 150
6 chauas Callophyllum brasiliensis 200
```

Vai para o R!

plot(y~x)

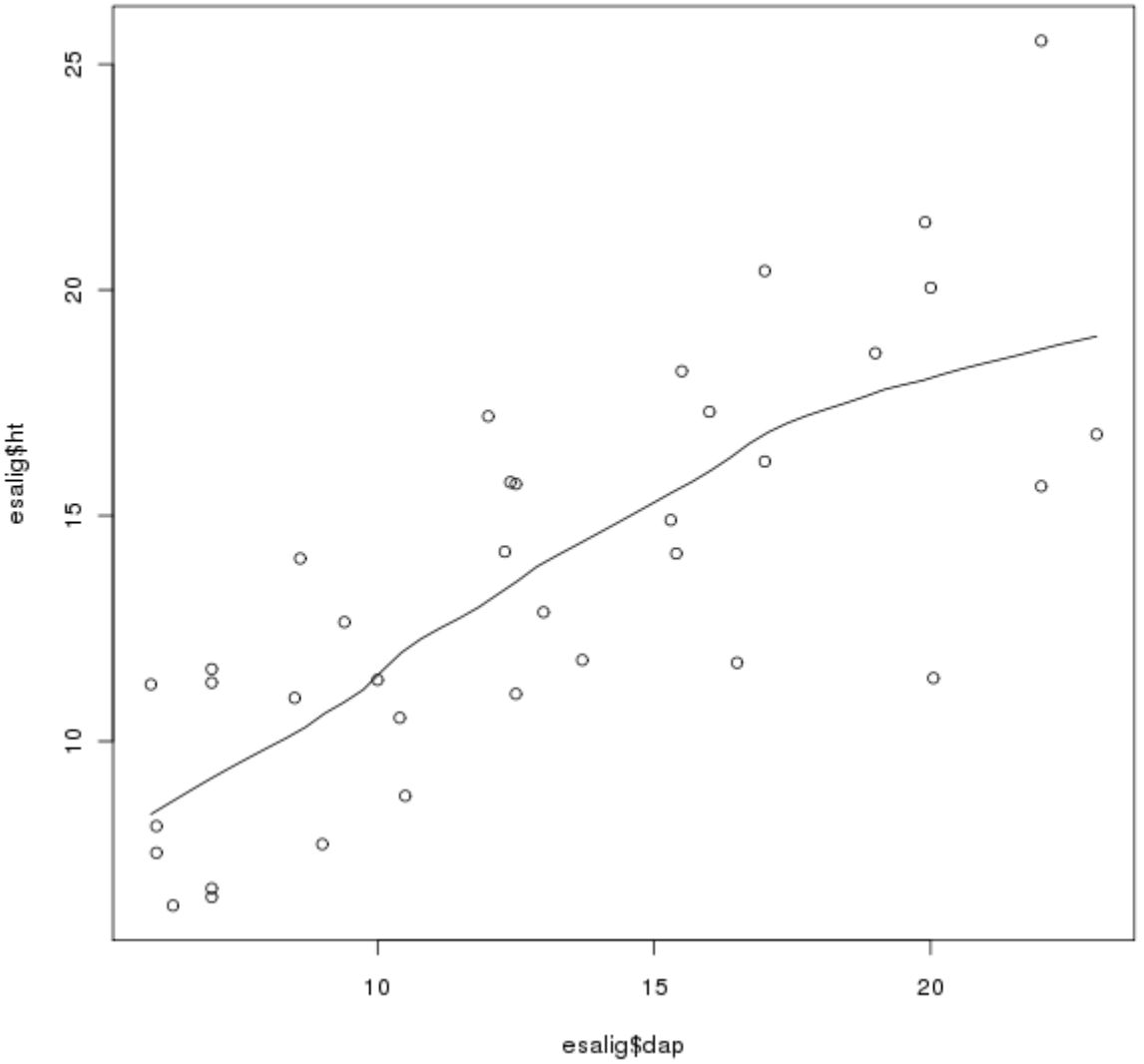
Espalhagramas




```
> plot(ht~dap, data=esalig)
```

scatter.smooth(y~x)

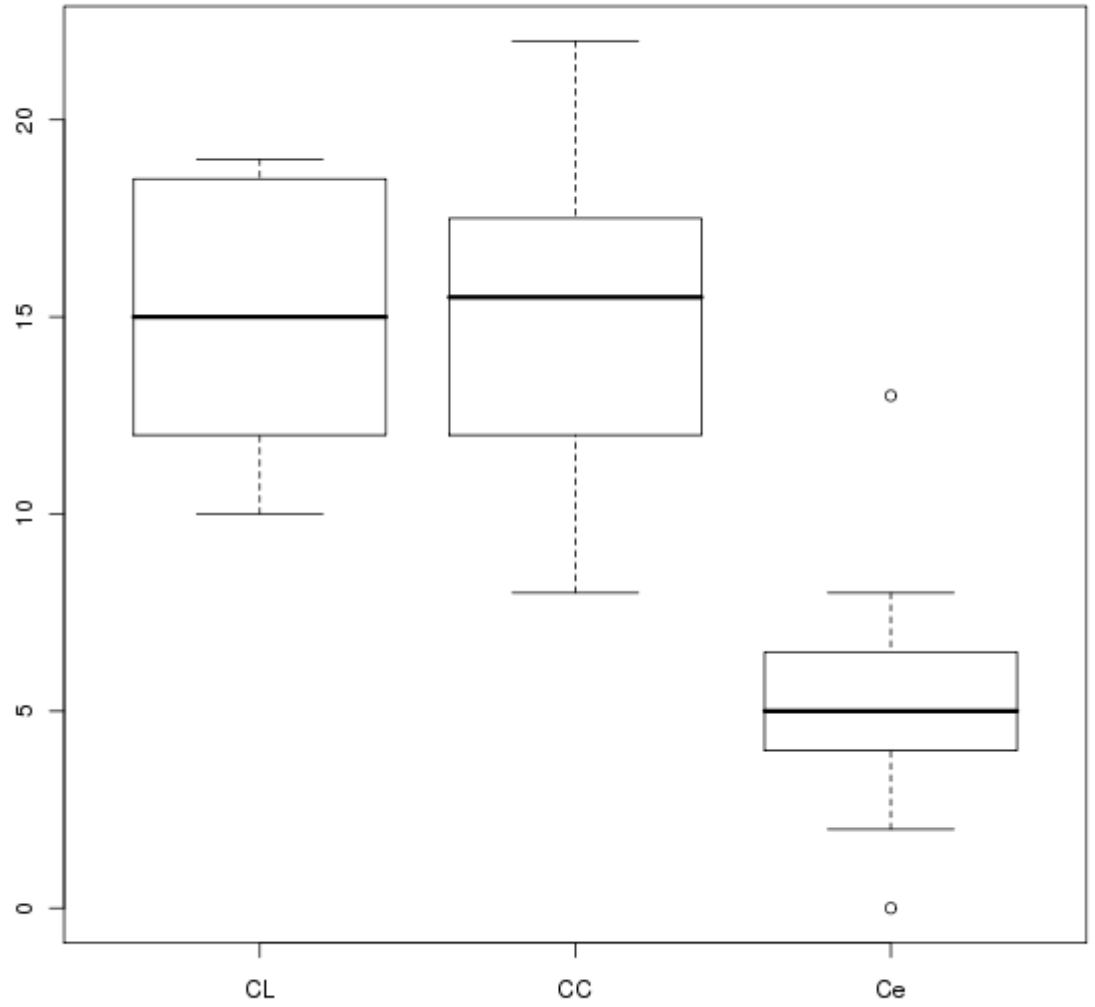
Espalhagramas com Linha de Tendência




```
> scatter.smooth(esalig$ht~esalig$dap, span=1/2)
```

boxplot(y~x)

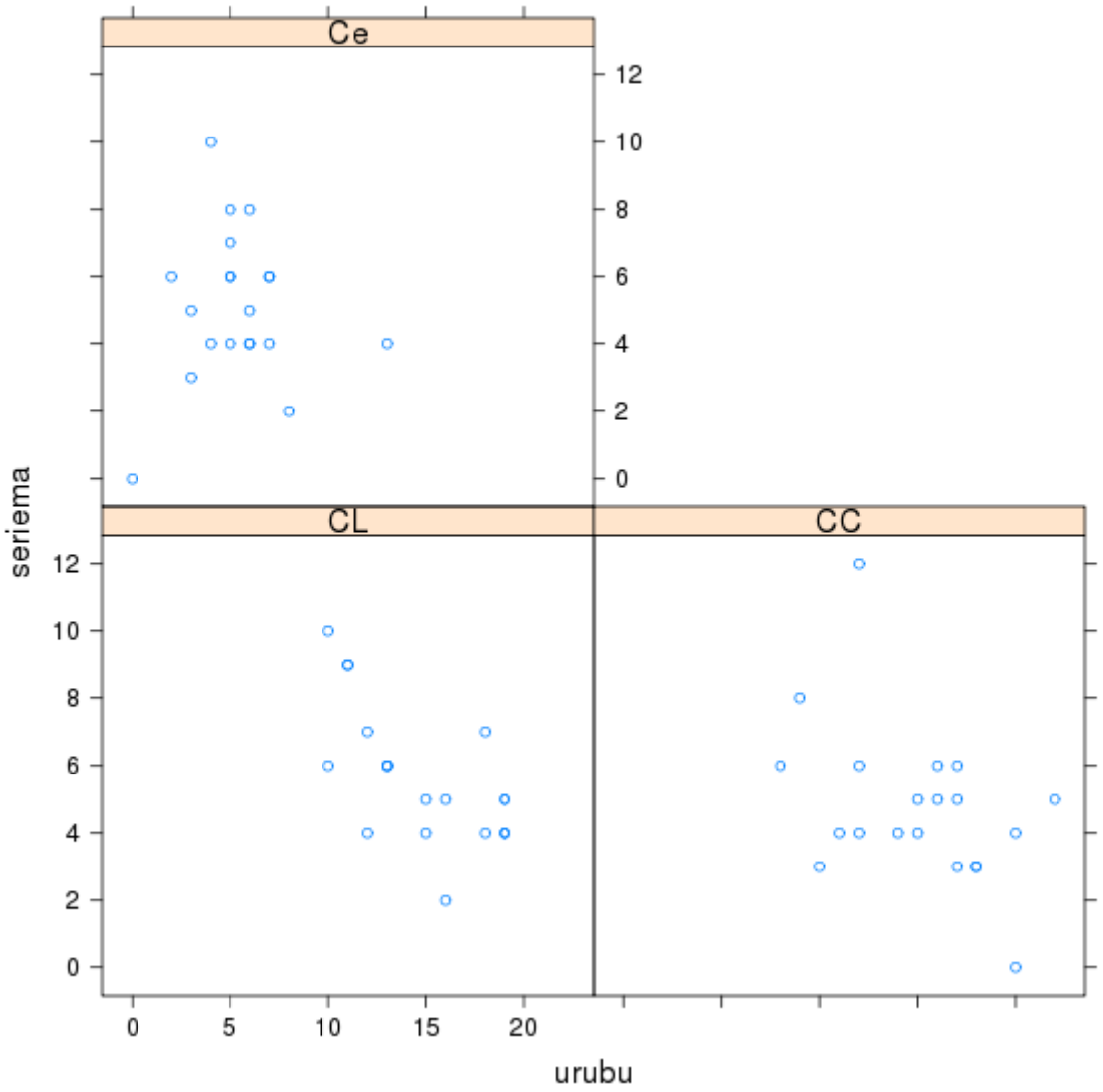
Boxplot por Classes




```
> boxplot(urubu~fisionomia, data=aves.c)
```

xyplot(y~x|z)

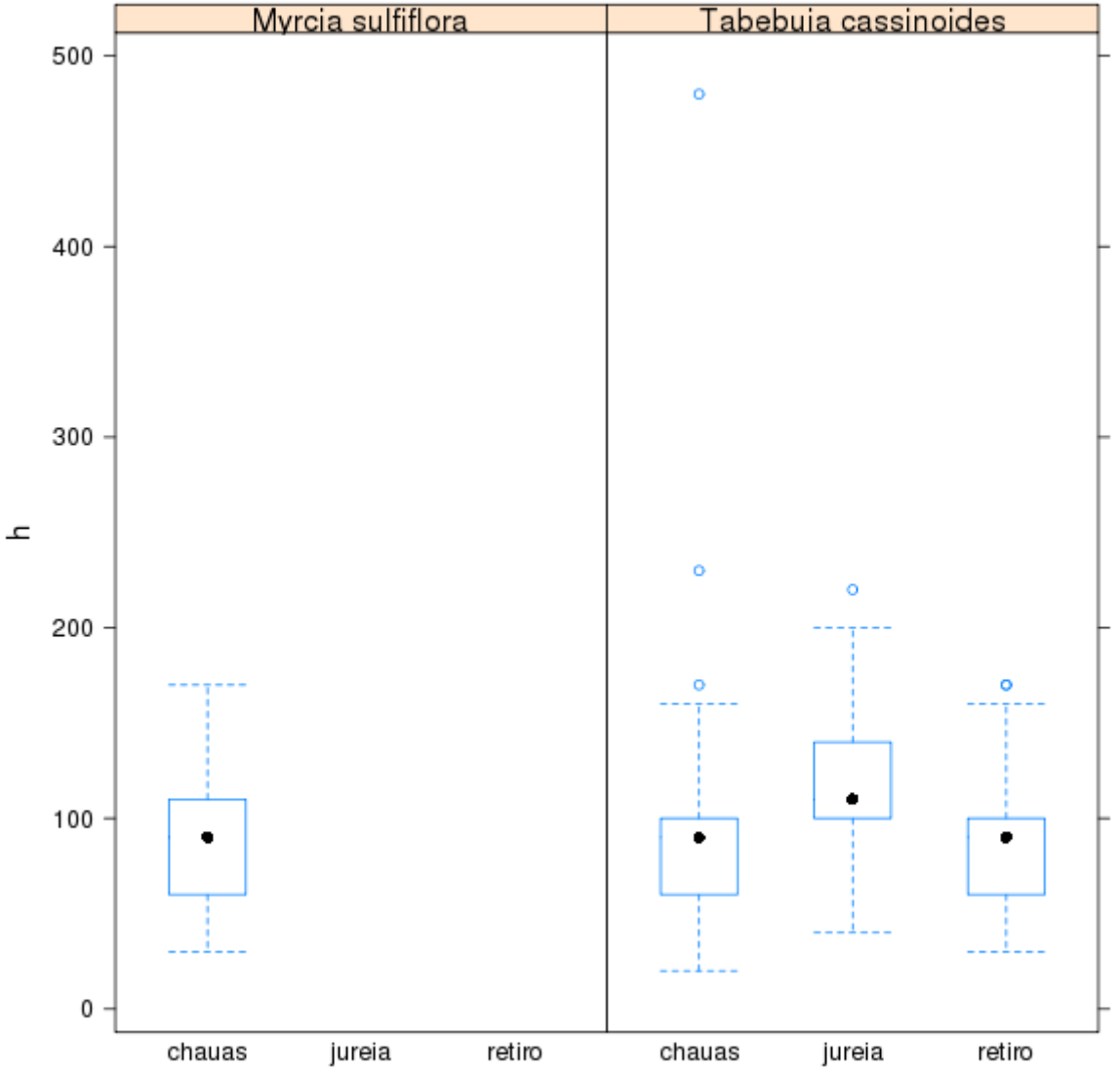
Pacote lattice: gráficos condicionados




```
> xyplot(seriema~urubu|fisionomia, data= aves.c)
```

bwplot(y~x|z)

Box-plot no lattice




```
>  bwplot(h~local|especie, data=caixeta.abund)
```

Sugestão de leitura

Ellison, A. M. 1993. Exploratory data analysis and graphic display. In: Scheiner, S. M. (ed.), *Design and analysis of ecological experiments*. Chapman & Hall, pp. 14-45.

McGill, R., Tukey, J. W. and Larsen, W. A. 1978. Variations of Box Plots. *Am. Statist*. 32: 12-16.

FIM DA UNIDADE 5

Para a tarde:

Tutorial da Unidade 5

http://ecologia.ib.usp.br/bie5782/doku.php?id=bie5782:02_tutoriais:start

Lista 4 de Exercícios:

http://ecologia.ib.usp.br/bie5782/doku.php id=bie5782:01_curso2009:exercicios4