



Curso de introdução ao software R - Modulo 1

Darlene Duarte

Renata Veroneze

Dúvida de ontem

Trocar o nome das colunas

```
>d <- data.frame(alpha=1:3, beta=4:6, gamma=7:9)
```

```
>names(d)[names(d)=="beta"] <- "two"
```

```
>names(d)[3] <- "three"
```

```
>names(d) <- sub("^alpha$", "one", names(d))
```

Usando o pacote

```
>library(plyr)
```

```
>rename(d, c("beta"="two", "gamma"="three"))
```

Gerando distribuições

No R podemos gerar dados usando uma distribuição desejada

`>?distributions` #abre ajuda do R sobre distribuições



Gerando distribuições

```
>d<-rnorm(n=100,mean=5,sd=2)#gera valores  
de uma dist. normal
```

```
>d<-rnorm(100,5,2)#gera valores de uma dist.  
normal
```

```
>plot(density(d))#constrói um gráfico de  
densidade com os valores gerados
```



Análise exploratória de dados

Média

No R a média pode ser calculada através do comando *mean()*.

```
> x<-c(3,9,10,12)
```

```
> mean(x)
```

Mediana

```
> y<-c(3,9,10,12,21)
```

```
> median(y)
```



E se houverem valores perdidos?

```
>x <- c(12,7,3,4.2,18,2,54,-21,8,-5,NA)  
#criando vetor com valor perdido
```

```
>mean(x)
```

```
>mean(x,na.rm = TRUE)
```

```
>median(x,na.rm=TRUE)
```



Moda

Duas formas *table()* ou *subset()*

```
> z<-c(2,5,8,7,5,6,9,5)
```

```
> table(z)
```

```
> subset(table(z),table(z)==max(table(z)))
```

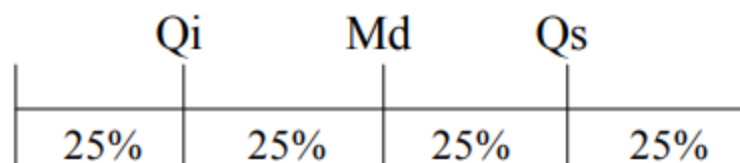


Quantis

Tal como a mediana, é uma medida que se calcula a partir da amostra ordenada. E representa as partes em que a série de dados é dividida.

```
> b<-rnorm(50,10)
```

```
> quantile(b)
```



Desafio



Calcule a média da coluna resposta do arquivo “Exemplo01.csv” usando a função *mean()* e usando o R como calculadora (escreva seu próprio código). Calcule a amplitude e a mediana.



Respostas

```
>r<-ex03$resposta
```

```
#ou
```

```
>r<-ex03[,3]
```

```
>mean(r)
```

```
>med<-sum(r)/length(r)
```

```
>range(r)
```

```
>median(r)
```

```
>summary(r)
```



Por tratamento...

```
>m<-by(ex03$resposta, ex03$trat, mean)
```

```
>by(ex03$resposta, ex03$trat, median)
```

```
>by(ex03$resposta, ex03$trat, summary)
```

```
>mean<-do.call("rbind", as.list(m))
```



Variância e desvio padrão

No R usamos as funções *var()* e *sd()* para calcular essas medidas.

```
> p<-c(20,23,23,28,33,37,37,37,40,44)
```

```
> var(p)
```

```
> sd(p)
```



Variância e desvio padrão

O coeficiente de correlação mede o grau de associação entre as variáveis dependente e independente (x e y).

```
> x <- rnorm(20, mean=2, sd=0.5)
> y <- rnorm(20, 1, 3)
> cor(x, y)
```



Análise exploratória de dados

Função	Descrição
length(x)	Tamanho
max(x)	Máximo
min(x)	Mínimo
range(x)	Amplitude
sum(x)	Soma dos elementos
prod(x)	Produto dos elementos
mean(x)	Média
median(x)	Mediana
var(x)	Variância
sd(x)	Desvio padrão
summary(x)	“Sumário”



Função `tapply`

Tal função permite aplicar operações semelhantes em subgrupos de dados determinados por fatores.

```
>tapply(ex03$resposta,  
ex03$tratamento, mean)
```

```
>tapply(ex03$resposta, ex03[, 2],  
var)
```

```
>tapply(ex03$resposta, ex03[, 2],  
range)
```


Desafio



Usando o arquivo `exercicio3.5_2.txt` , calcule a média da biomassa das folhas e a variância da biomassa dos troncos. Calcule a correlação entre a biomassa dos troncos e a nova variável obtida no exercício anterior(soma das biomassas das folhas e troncos). Qual é a altura (ht) máxima? Qual é a classe que mais ocorre?



Resposta

```
> mean(dados.esaligna$folha)
[1] 7.231944
> var(dados.esaligna$tronco)
[1] 4882.308
> cor(dados.esaligna$tronco,T_F)
[1] 0.9916532
```



Resposta

```
> max(dados.esaligna$ht)
[1] 25.52
```

```
> table(dados.esaligna$classe)
```

a	b	c	d
10	10	10	6



Desafio

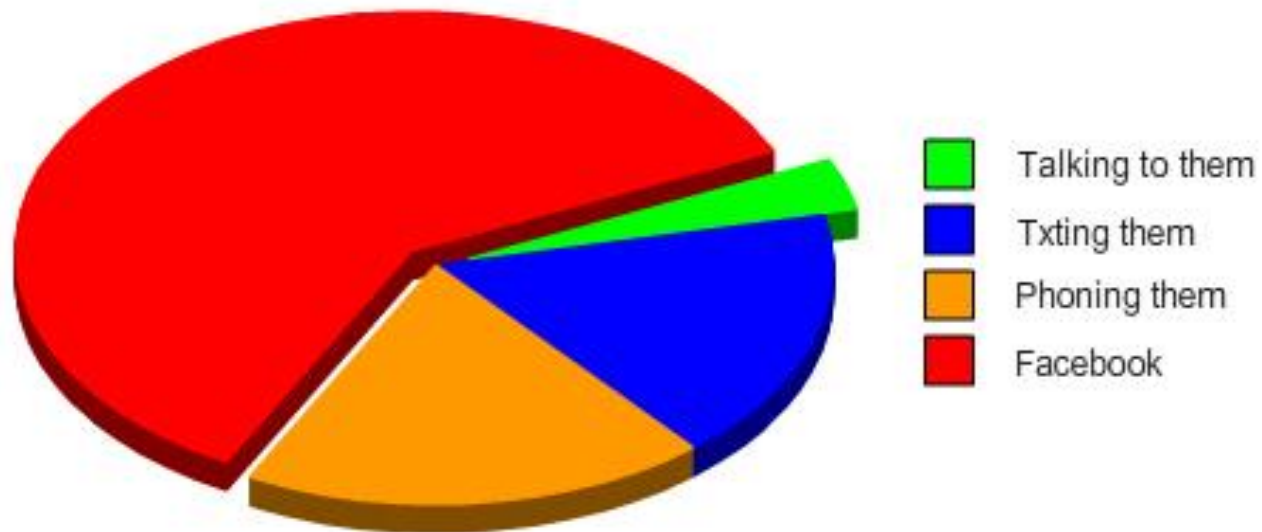


1. Leia o arquivo dados_suinos.csv
2. Faça um summary das variáveis PESOINC, CRT, CMD, GPT, GPD, ETT, ETA e CA
3. Usando a função tapply faça um summary de cada variável por tratamento



Gráficos

How I Communicate With People In The Next Room



Gráficos

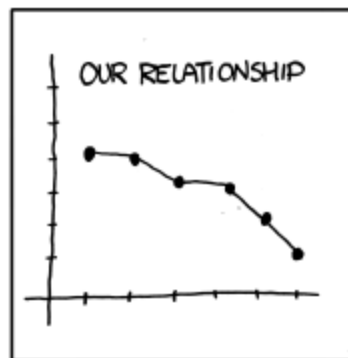
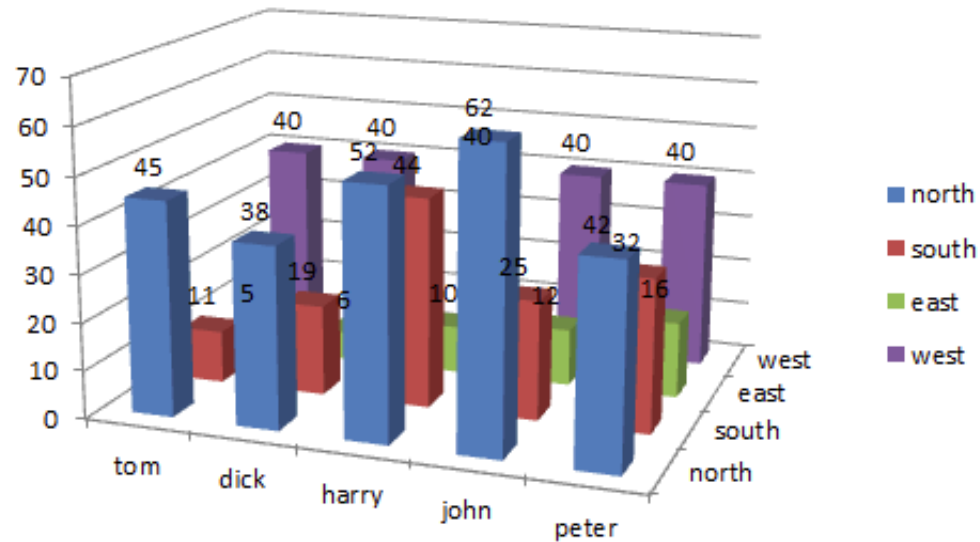
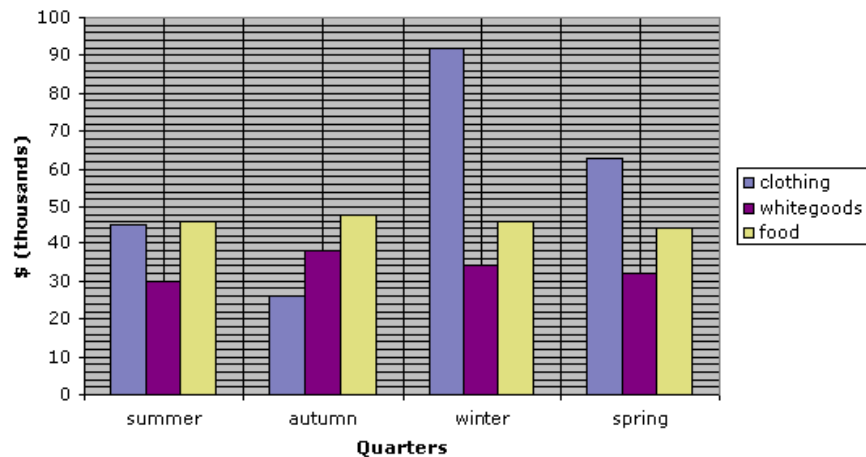


- Representação esquemática dos dados
- Visualmente melhor que as tabelas
- Valoriza o seu trabalho



Gráficos

Company sales in 1998



Gráficos

Gráficos de dispersão

Gráficos de barra

Box-plot

Histogramas

QQ-plot



<http://rgraphgallery.blogspot.com.br/>



Gráficos

- A função *plot()*.

```
> x<-c(10:20)
```

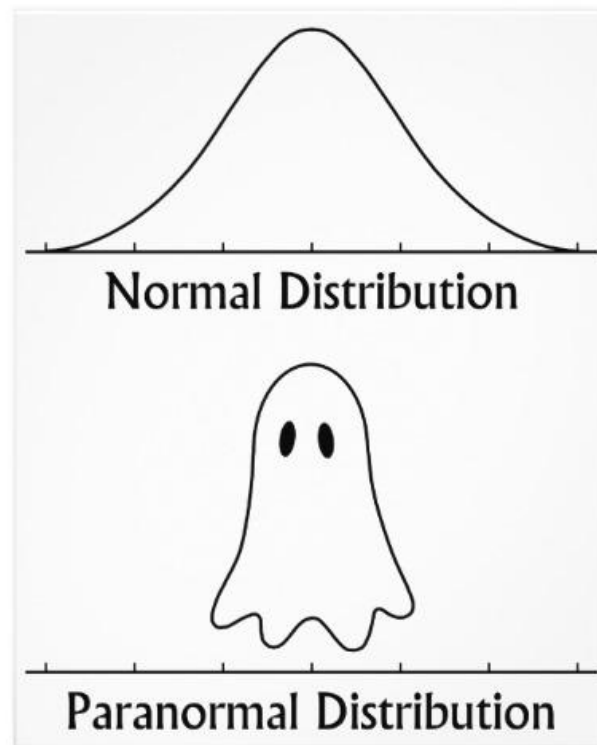
```
> y<-seq(30,50,2)
```

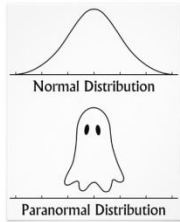
```
> plot(x,y)
```



QQ-plot

- Comparar distribuição dos dados com a distribuição teórica (qqnorm e qqline).





QQ-plot

- Comparar distribuições dos dados com a distribuição teórica (qqnorm e qqline).

```
>set.seed(183)
```

```
>x <- rnorm(80, mean=50, sd=5) # Normally distributed  
  numbers
```

```
>z <- runif(80) # Uniformly distributed numbers
```

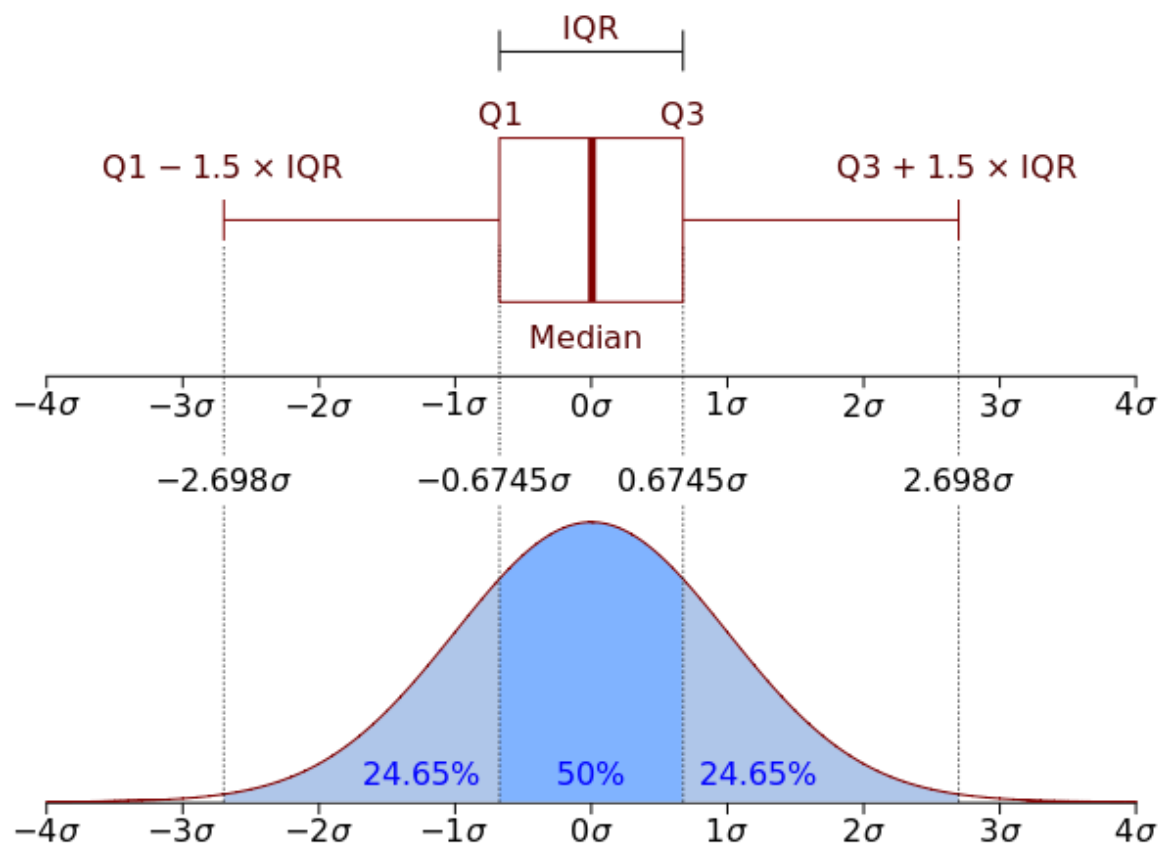
```
>qqnorm(x)
```

```
> qqline(x)
```



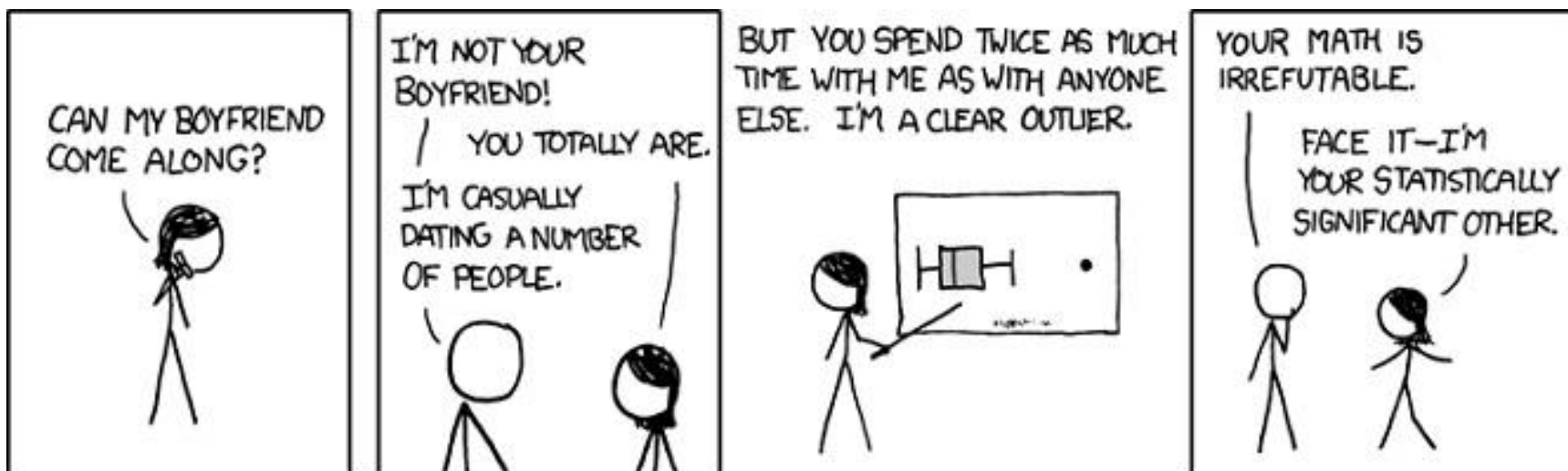
Boxplot

- Permite avaliar a simetria dos dados e a sua dispersão



Boxplot

- Permite avaliar a simetria dos dados e a sua dispersão



Boxplot

- Permite avaliar a simetria dos dados e a sua dispersão

```
riqueza <- c(15,18,22,24,25,30,31,34,37,39,41,45)
```

```
area <- c(2,4.5,6,10,30,34,50,56,60,77.5,80,85)
```

```
area.cate <- rep(c("pequeno", "grande"), each=6)
```

```
plot(riqueza~area)
```

```
plot(area,riqueza) # o mesmo que o anterior
```

```
boxplot(riqueza~area.cate)
```



Desafio



1. Leia o arquivo de dados `exercicio3.5_2.txt`
2. Faça um plot utilizando as variáveis `dap` e `ht`
3. Faça um boxplot utilizando as variáveis `dap` e `classe`



Resposta



```
a<-read.table("exercicio3.5_2.txt",h=T, sep=",")
```

```
head(a)
```

```
plot(a$dap,a$ht)
```

```
boxplot(a$dap~a$classe)
```



Barplot

```
a<-read.table("exercicio3.5_2.txt",h=T, sep=",")  
counts <- table(a$classe)  
counts  
barplot(counts)
```



Gráfico de pizza

```
slices <- c(10, 12, 4, 16, 8)
```

```
lbls <- c("US", "UK", "Australia", "Germany",  
  "France")
```

```
pie(slices, labels = lbls, main="Pie Chart of  
  Countries")
```



Desafio



1. Utilize a função `tapply` para calcular a média de `dap` para cada classe do `exercicio3.5_2.txt`
2. Faça um `barplot` utilizando as médias geradas com a função `tapply`



Resposta



```
m_classe<-tapply(a$dap, a$classe, mean)  
barplot(m_classe)
```

