

Exemplo de gráficas

Esta primeira tarefa consiste en realizar un exemplo de gráfica apropiado para variables de diferentes tipos de datos: atributos, discretas, continuas e series de tempo.

Atributos

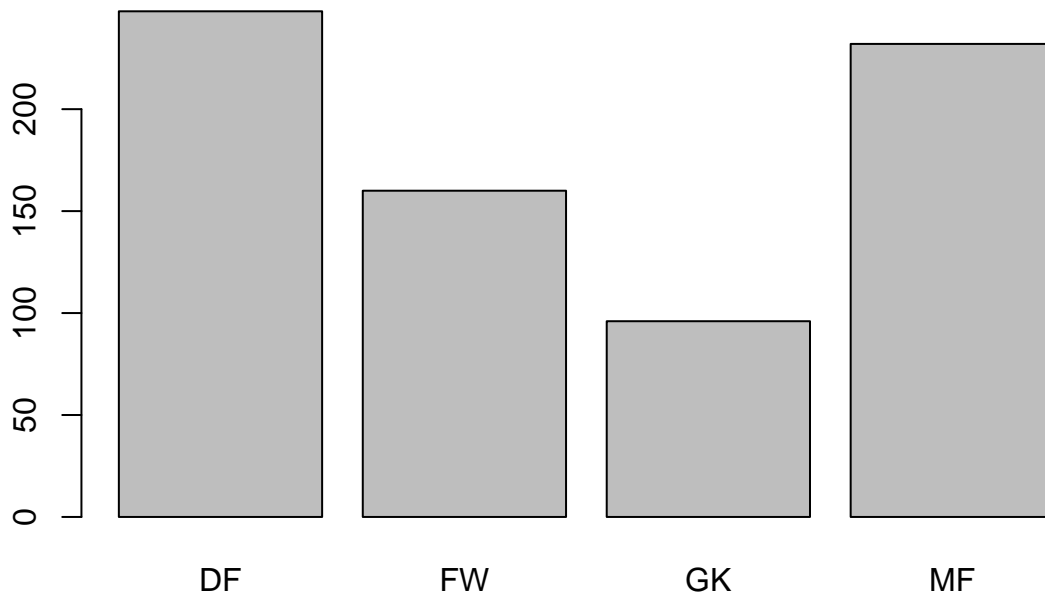
Son variables cualitativas, ou sexa, que os seus resultados non son valores numéricos. Tipos de gráficas apropiadas: similares ás variables discretas, como **digramas de barras**, **diagramas de puntos**, **diagramas de sector**, pero nos últimos tempos aparece algunha máis específica de atributos, como as **nubes de puntos**

```
##usase para estes exemplos o conxunto de datos Mundial 2018, que están nos "datos para a aula"  
datos=read.csv2("mundial2018.csv")  
##en particular, para os primeiros exemplos de atributos usase a columna "posicion"  
posicion=datos$Pos.
```

Diagramas de barras

Tamén se coñecen como *diagramas de rectángulos* cando se usan como atributos. Consisten en unha barra para cada categoría, que ten a frecuencia como altura

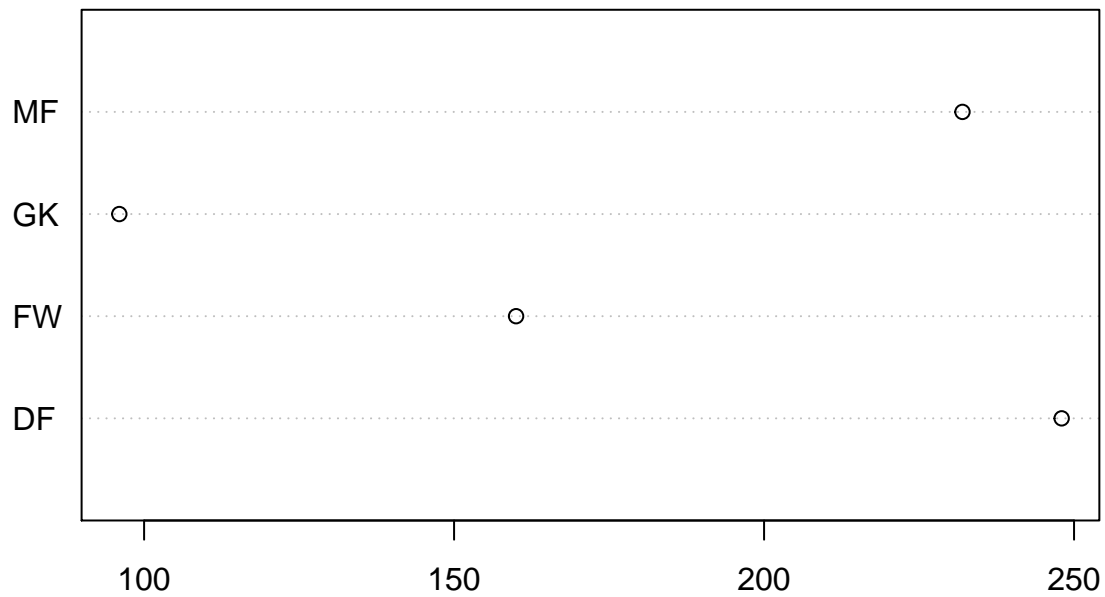
```
#diagrama de barras, comando barplot() aplicado a unha táboa de datos:  
barplot(table(posicion))
```



Diagramas de puntos

Neste caso é **un punto** o que marca a frecuencia de cada categoría:

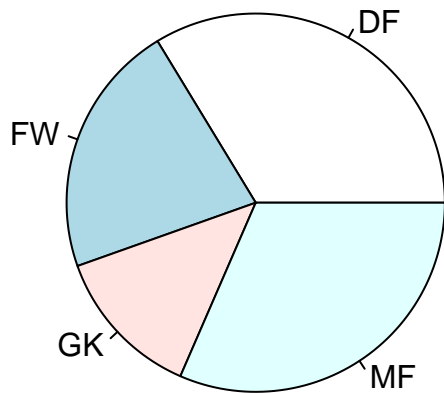
```
#diagrama de puntos, comando dotchart() aplicado a unha táboa de datos:  
dotchart(table(posicion))
```



Diagramas de sector

Ou *diagramas de tarta*. Consisten en separar un círculo en diferentes sectores, un por cada categoría, de área proporcional á súa frecuencia.

#diagrama de sectores, comando pie() aplicado a unha táboa de datos:
`pie(table(posicion))`



Nube de palabras

É unha gráfica relativamente moderna, que se ve con bastante frecuencia, xa que resulta vistosa. Do punto de vista estatístico permite unha visión global rápida das categorías, se ben, a visión é máis exacta (e menos vistosa) cun diagrama de barras

#cárgase o paquete (ten que estar instalado)
`library(wordcloud)`
Neste caso aplícase á táboa separada en dous vectores: valores e frecuencias
`auxiliar1=table(posicion)#facen a táboa`
`valores=dimnames(auxiliar1)[[1]]`
`frecuencias=as.vector(auxiliar1)`

```
#comando wordcloud; o parámetro "col" indica a lista de cores que debe usar
wordcloud(valores,frecuencias,col=rainbow(length(valores)))
```



```
#podese facer con poucas categorías diferentes, pero luce máis,
# e é máis útil, cando hai moitas categorías diferentes,
# coma na variable "country", que representa a
# liga na que xogaba cada xogador do mundial 2018
liga=datos$country #crear unha variable auxiliar para a liga
auxiliar1=table(datos$country)
valores=dimnames(auxiliar1)[[1]]
frecuencias=as.vector(auxiliar1)
wordcloud(valores,frecuencias,col=rainbow(length(valores)),min.freq = 1)
```



```
# ,min.freq = 1 quiere decir que poña todas as categorías que teñan frecuencia>=1
```

Variables discretas

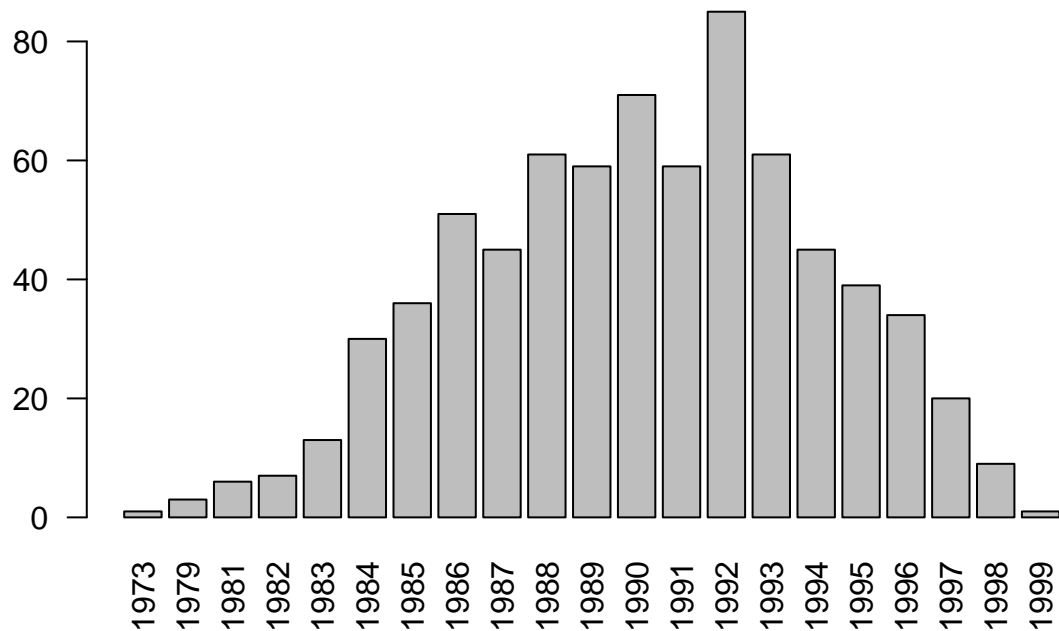
Son variables cuantitativas, ou sexa, que os seus resultados non valores numéricos. Tipos de gráficas apropiadas: similares aos atributos, como **digramas de barras**, **diagramas de puntos**, **diagramas de sector**.

```
##usase para estes exemplos o conxunto de datos Mundial 2018, que están nos "datos para a aula"
datos=read.csv2("mundial2018.csv")
##en particular, para os primeiros exemplos de atributos usase a columna "year", por que é
# a única discreta, aínda que ten demasiados valores diferentes, e case sería máis apropiada
#tratala como continua
ano=datos$year
```

Diagramas de barras

Tamén se coñecen como *diagramas de rectángulos* cando se usan como atributos. Consisten en unha barra para cada categoría, que ten a frecuencia como altura

```
#diagrama de barras, comando barplot() aplicado a unha táboa de datos:
barplot(table(ano),las=2)
```

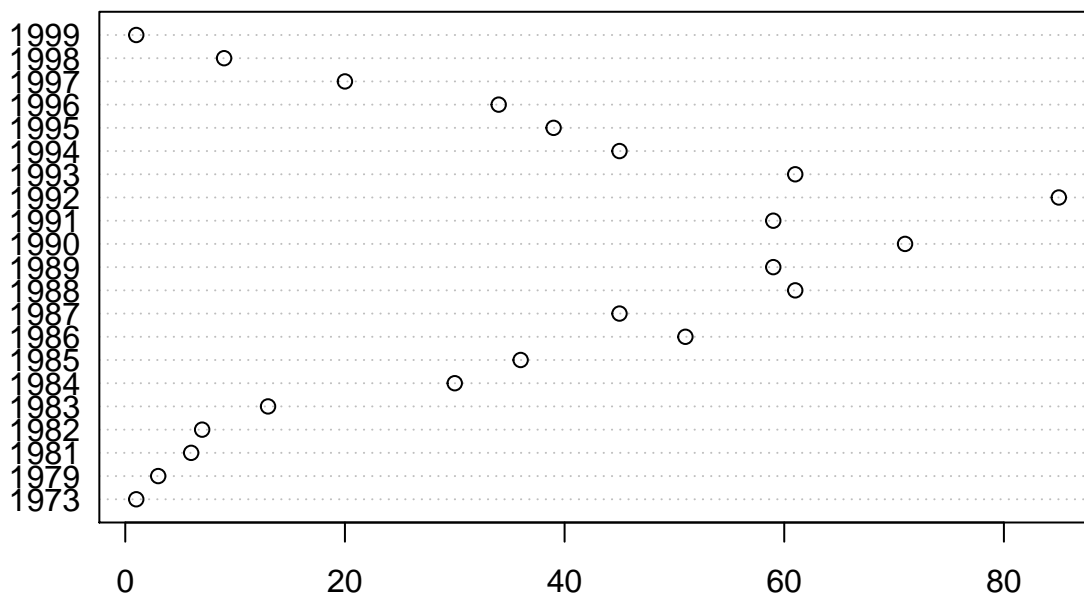


```
# ,las=2 indica que os textos no eixe X se coloquen en vertical
```

Diagramas de puntos

Neste caso é **un punto** o que marca a frecuencia de cada categoría:

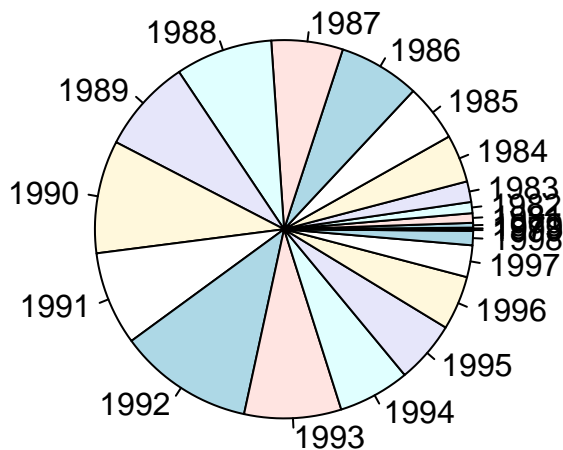
```
#diagrama de puntos, comando dotchart() aplicado a unha táboa de datos:
dotchart(table(ano))
```



Diagramas de sector

Ou *diagramas de tarta*. Consisten en separar un círculo en diferentes sectores, un por cada categoría, de área proporcional á súa frecuencia.

#diagrama de sectores, comando pie() aplicado a unha táboa de datos:
`pie(table(ano))`



Variables continuas

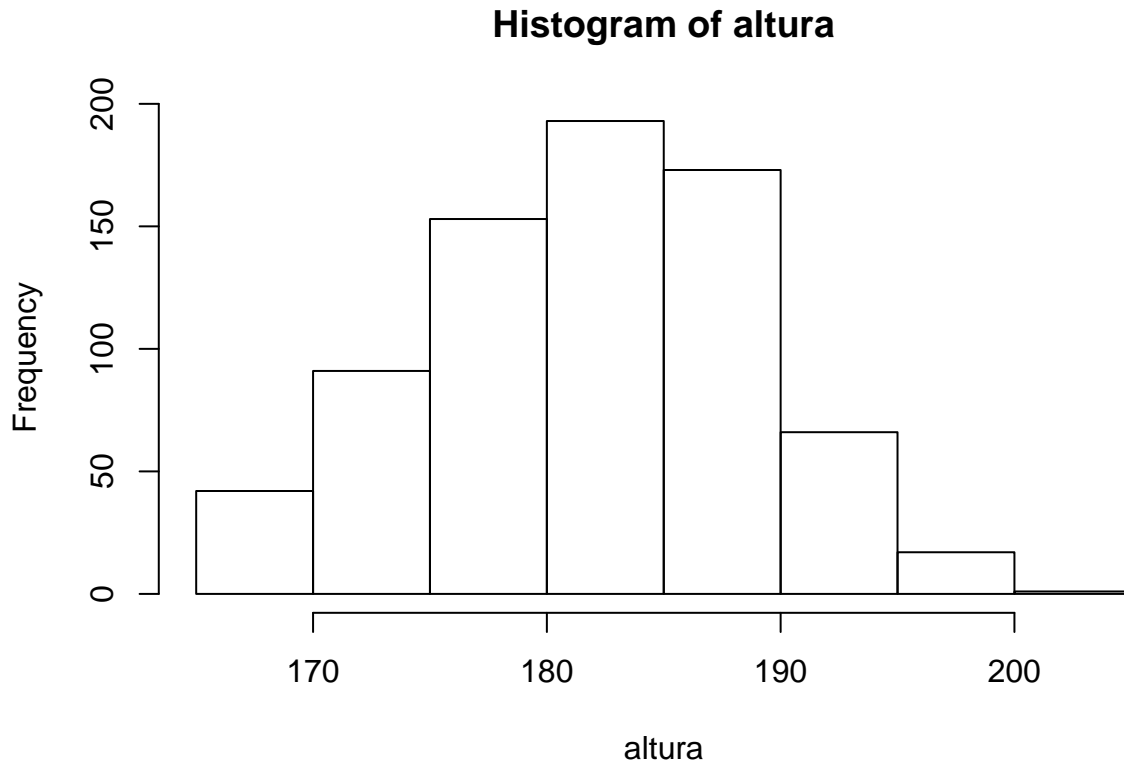
Son variables cuantitativas, ou sexa, que os seus resultados son valores numéricos. Tipos de gráficas apropiadas: **histogramas, densidades, diagramas de barras.**

```
##usase para estes exemplos o conxunto de datos Mundial 2018, que están nos "datos para a aula"  
datos=read.csv2("mundial2018.csv")  
##en particular, para os primeiros exemplos de atributos usase a columna "Height"  
altura=datos$Height
```

Histogramas

Son como diagramas de barras para variables continuas. Para realizalas os datos agrúpanse en intervalos, e a súa altura será a frecuencia ou a densidade do intervalo

```
#histograma, comando hist() aplicado a un un vector de datos:  
hist(altura)
```

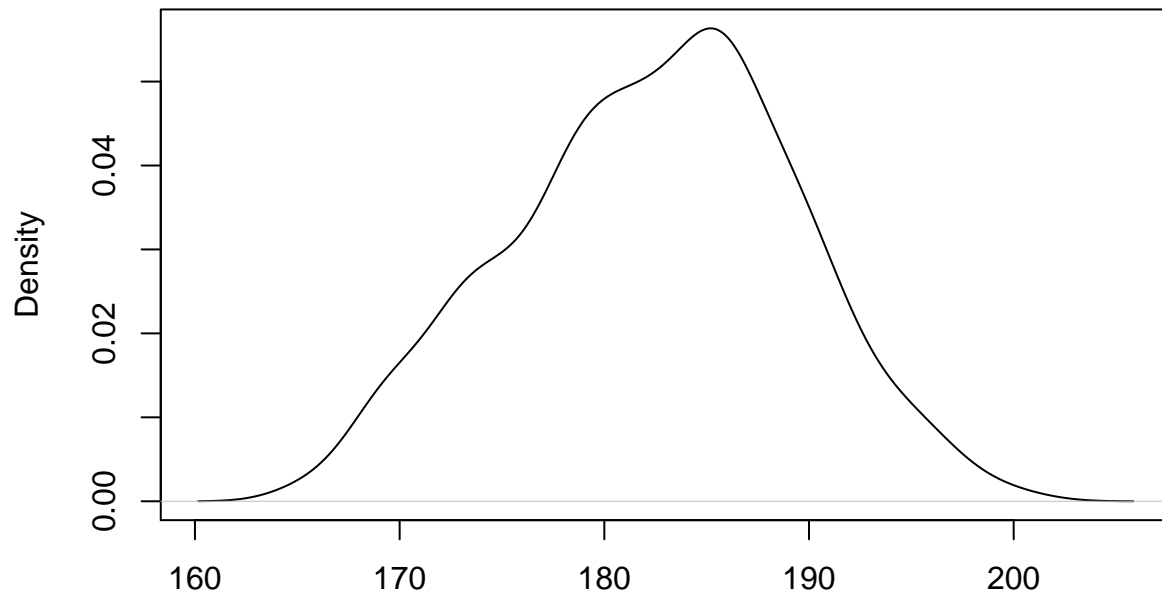


Densidade

É coma un histograma suavizado (a amplitude dos intervalos converxe a cero)

```
#hai que calcular a densidade, comando density(), e despois debuxala nunha gráfica  
# comando plot():  
plot(density(altura))
```

density.default(x = altura)

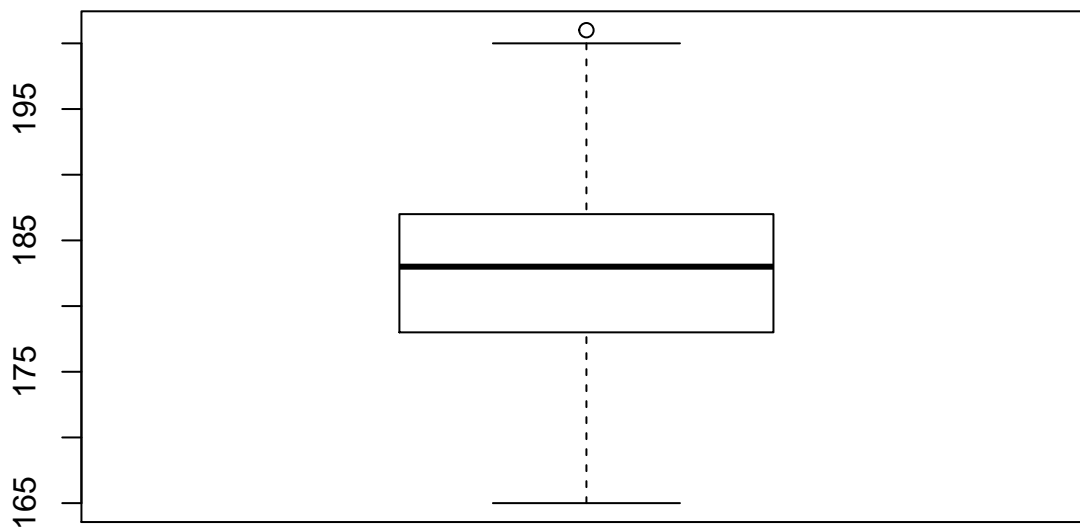


N = 736 Bandwidth = 1.614

Diagramas de caixa

Marcan os pontos onde estão o 1º quartil, mediana e terceiro quartil, ademais de indicar valores atípicos.

```
#diagrama de caixa, comando boxplot():  
boxplot(altura)
```



Serie de tempo

Soen ser variables numéricas (discretas ou continuas) nas que importa o período de tempo nas que foron producidas, polo que unha representación gráfica dunha serie de tempo debe ter en conta tanto os valores da serie como os periodos.

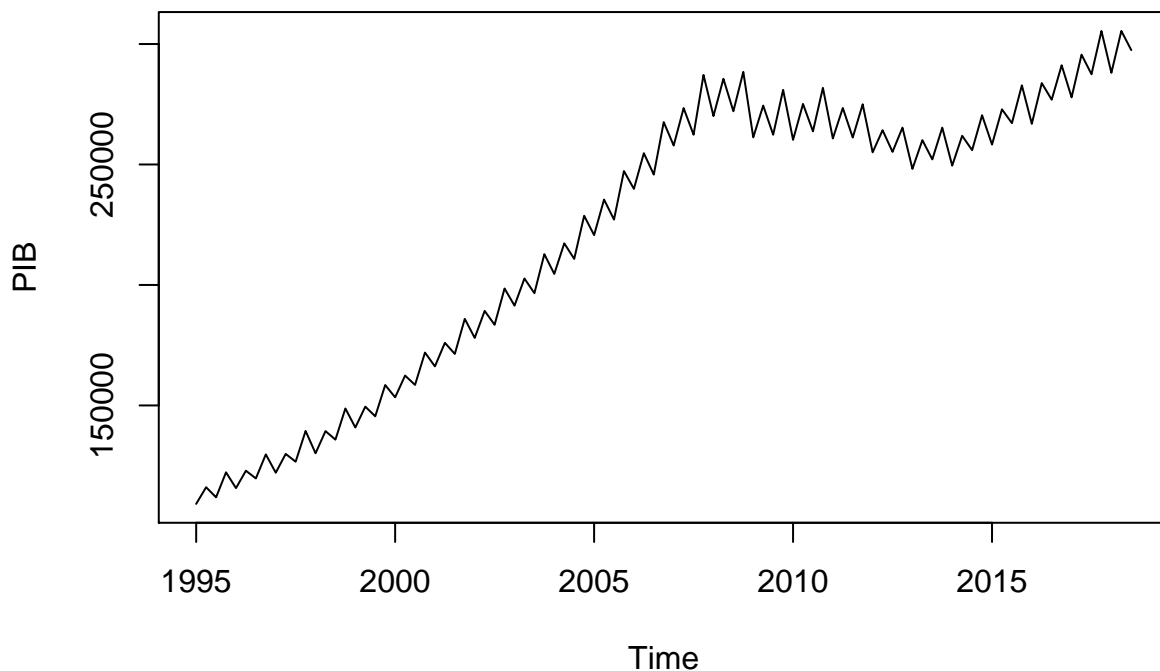
Por esa causa a gráfica máis habitual para este tipo de datos son os **diagramas de liñas**.

```
##usase para este exemplo unha das series que está no conxunto de "varias.series.csv",  
# colocada no cartafol de "datos para a aula"  
datos=read.csv2("varias.series.csv")  
##en particular, usase a columna "PIB.a.precios.de.mercado"  
PIB=datos$PIB.a.precios.de.mercado  
#Ademais é mellor darlle estrutura de serie, e así facilita a realización da gráfica.  
#dase estrutura de serie co comando ts(), indicándolle o periodo de inicio ("start"), e  
#o número de datos que ten cada ano ("frequency")  
PIB=ts(PIB,start = c(1995,1),frequency = 4)#comenza no 1º trimestre de 1995;  
# e 4 periodos (trimestres) cada ano
```

Diagramas de liñas

Faise habitualmente para datos ordenados. Consiste nunha liña que marca a evolución dos valores dos datos. No caso de series de tempo é a evolución dos datos con respecto ao tempo.

```
plot(PIB)
```



Variables bidimensionais

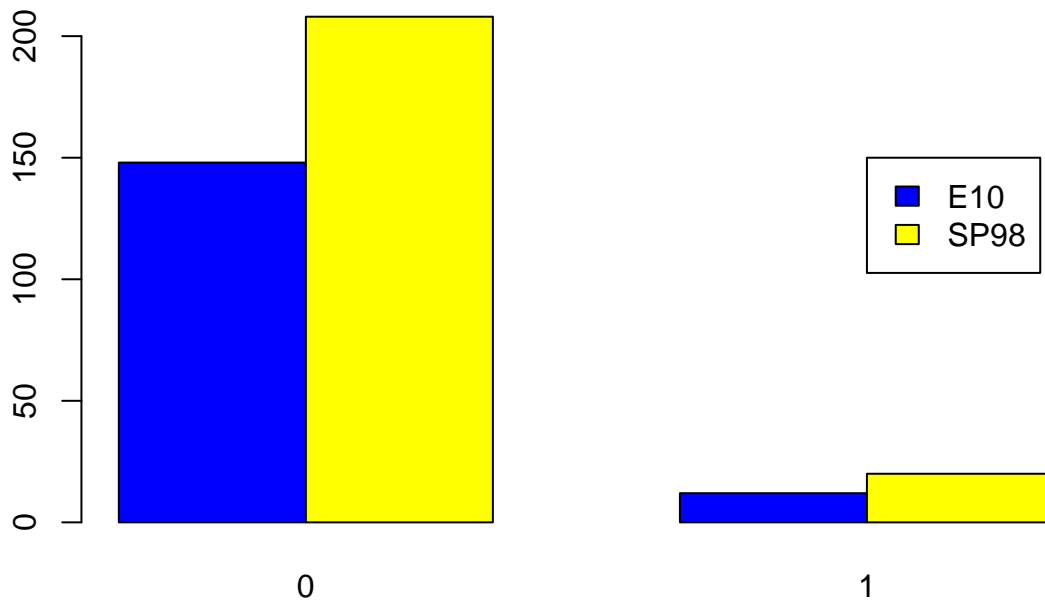
```
# Usaranse agora datos de "conductor.csv", as anotacións dun conductor sobre o consumo  
#de gasolina nas súas viaxes  
datos=read.csv2("conductor.csv")
```


Dous atributos (ou duas discretas, ou discreta e atributo)

Unha boa posibilidade son **diagramas de barras** bidimensionais.

```
# Usaranse neste exemplo as columnas "gas_type" e "sun" da colección de datos:
# Variables que indican se foi ou non un día soleado e o tipo de gasolina que se consumiu
gasolina=datos$gas_type
sol=datos$sun

#neste caso hai que usar barplot con táboa, pero de dobre entrada (2 variables)
barplot(table(gasolina,sol),col=c("blue","yellow"),beside = TRUE)
# e hai que engadir unha lenda para identificar as barras
legend(5,150,levels(gasolina),fill = c("blue","yellow"))
```

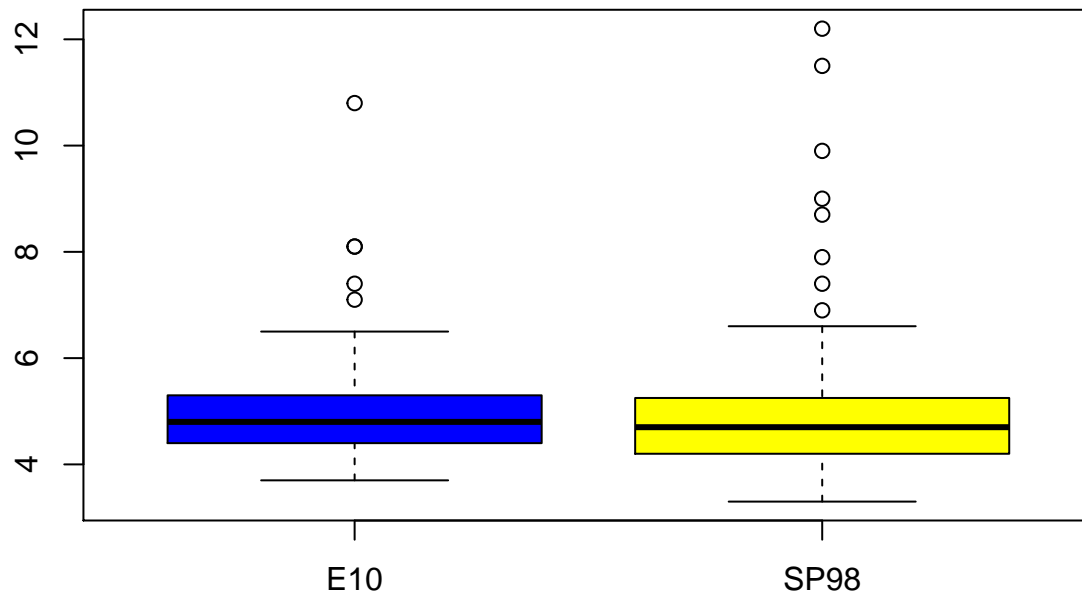


Un atributo (ou discreta) e unha continua

Unha boa posibilidade son **diagramas de caixa** da continua separado por categorías do atributo.

```
# Usaranse neste exemplo as columnas "gas_type" e "consume" da colección de datos:
# Variables que indican o consumo e o tipo de gasolina que se consumiu
consume=datos$consume
gasolina=datos$gas_type

#neste caso hai que usar boxplot da variable continua "en función" do atributo:
# boxplot(continua~gasolina)
boxplot(consume~gasolina,col=c("blue","yellow"))
```



Dúas con-

tinuas

Unha boa posibilidade é un **diagrama de dispersión** coas dúas continuas representadas nun eixe de coordenadas.

Usaranse neste exemplo as columnas "speed" e "consume" da colección de datos:

Variables que indican o consumo e a velocidade

`consume=datos$consume`

`velocidade=datos$speed`

#neste caso hai que usar boxplot da variable continua "en función" do atributo:

boxplot(continua~gasolina)

`plot(velocidade,consume)`

