Dataframe

Osemar Xavier

2 de outubro de 2018

Criando um Data Frame

```
d1 <-  data.frame(X = 1: 10, Y = c(51, 54, 61, 67, 68, 75, 77, 75, 80, 82 ))
d1
```

```
##      X  Y
## 1    1 51
## 2    2 54
## 3    3 61
## 4    4 67
## 5    5 68
## 6    6 75
## 7    7 77
## 8    8 75
## 9    9 80
## 10  10 82
```

names -> Nomes das colunas

```
names(d1)
```

```
## [1] "X" "Y"
```

Class - informa o tipo

```
class(d1)
```

```
## [1] "data.frame"
```

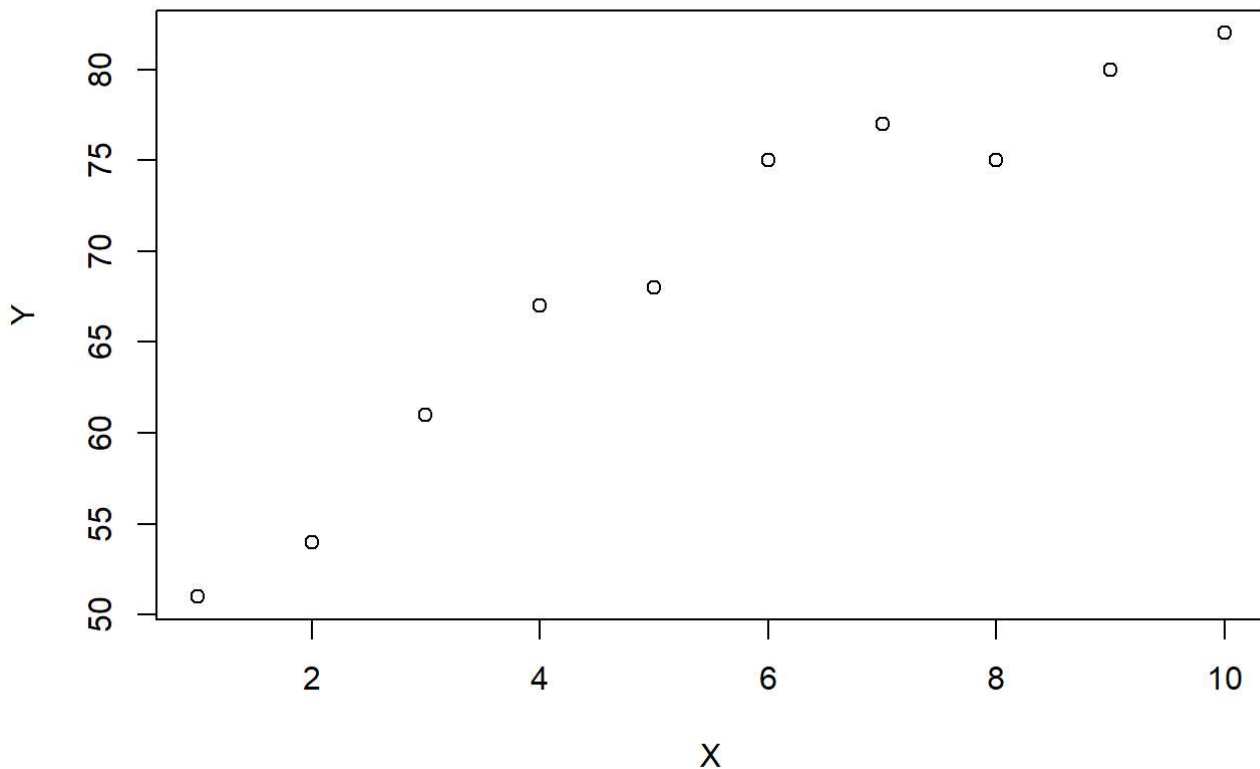## Acessando as colunas do dataframe

```
d1$X
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

```
d1$Y
```

```
##  [1] 51 54 61 67 68 75 77 75 80 82
```

Plotando os dados do dataframe

```
plot(d1)
```


plot(d1$X, d1$Y)

Carregando um arquivo de texto

```
ex1 <-  read.table('teste.txt')
ex1
```

```
##      V1 V2        V3 V4
## 1 Aula de Linguagem R.
```

carregando arquivo de texto com cabeçalho

```
ex2 <-  read.table('teste2.txt', head  = T)
ex2
```

```
##     nome           email sexo
## 1 maria maria@gmail.com     m
```

ler arquivo no formato csv

```
ex3 <-  read.table('teste3.csv', head = T, sep = ";", dec = ",")
ex3
```

```
##      nome              email sexo  salario
## 1   mari      maria@gmail.com   m   100.20
## 2 lucas      lucas@gmail.com    m   300.40
## 3 bruna bruna@gmail.com.com    f   345.67
```

Exibindo classes

```
class(ex3)
```

```
## [1] "data.frame"
```

```
class(ex3$salario)
```

```
## [1] "numeric"
```

```
ex3$salario
```

```
## [1] 100.20 300.40 345.67
```

Exibir os datasets

```
data()
```

Carrega pacotes de datasets no R

```
ls("package:datasets")
```

```
##     [1] "ability.cov"              "airmiles"
##     [3] "AirPassengers"            "airquality"
##     [5] "anscombe"                 "attenu"
##     [7] "attitude"                 "austres"
##     [9] "beaver1"                  "beaver2"
##    [11] "BJsales"                  "BJsales.lead"
##    [13] "BOD"                      "cars"
##    [15] "ChickWeight"              "chickwts"
##    [17] "co2"                      "CO2"
##    [19] "crimtab"                  "discoveries"
##    [21] "DNase"                    "esoph"
##    [23] "euro"                     "euro.cross"
##    [25] "eurodist"                 "EuStockMarkets"
##    [27] "faithful"                 "fdeaths"
##    [29] "Formaldehyde"             "freeny"
##    [31] "freeny.x"                 "freeny.y"
##    [33] "HairEyeColor"             "Harman23.cor"
##    [35] "Harman74.cor"            "Indometh"
##    [37] "infert"                   "InsectSprays"
##    [39] "iris"                     "iris3"
##    [41] "islands"                  "JohnsonJohnson"
##    [43] "LakeHuron"                "ldeaths"
##    [45] "lh"                       "LifeCycleSavings"
##    [47] "Loblolly"                 "longley"
##    [49] "lynx"                     "mdeaths"
##    [51] "morley"                   "mtcars"
##    [53] "nhtemp"                   "Nile"
##    [55] "nottem"                   "npk"
##    [57] "occupationalStatus"       "Orange"
##    [59] "OrchardSprays"            "PlantGrowth"
##    [61] "precip"                   "presidents"
##    [63] "pressure"                 "Puromycin"
##    [65] "quakes"                   "randu"
##    [67] "rivers"                   "rock"
##    [69] "Seatbelts"                "sleep"
##    [71] "stack.loss"               "stack.x"
##    [73] "stackloss"                "state.abb"
##    [75] "state.area"               "state.center"
##    [77] "state.division"           "state.name"
##    [79] "state.region"             "state.x77"
##    [81] "sunspot.month"            "sunspot.year"
##    [83] "sunspots"                 "swiss"
##    [85] "Theoph"                   "Titanic"
##    [87] "ToothGrowth"              "treering"
##    [89] "trees"                    "UCBAdmissions"
##    [91] "UKDriverDeaths"           "UKgas"
##    [93] "USAccDeaths"              "USArrests"
##    [95] "UScitiesD"                "USJudgeRatings"
##    [97] "USPersonalExpenditure"   "uspop"
##    [99] "VADeaths"                 "volcano"
##   [101] "warpbreaks"               "women"
##   [103] "WorldPhones"              "WWWusage"
```

View -> Permite visualização do arquivo ex3

```
View(ex3)
```

Datasets no R

```
data("mtcars")
mtcars
```

```
##                      mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4           21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag       21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710          22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive      21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout   18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant             18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D           24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230            22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280            19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C           17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128            32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic         30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla      33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona       21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E          21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

```
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
tail(mtcars)
```

```
##                   mpg cyl  disp  hp drat    wt qsec vs am gear carb
## Porsche 914-2    26.0   4 120.3  91 4.43 2.140 16.7  0  1    5    2
## Lotus Europa     30.4   4  95.1 113 3.77 1.513 16.9  1  1    5    2
## Ford Pantera L   15.8   8 351.0 264 4.22 3.170 14.5  0  1    5    4
## Ferrari Dino     19.7   6 145.0 175 3.62 2.770 15.5  0  1    5    6
## Maserati Bora    15.0   8 301.0 335 3.54 3.570 14.6  0  1    5    8
## Volvo 142E       21.4   4 121.0 109 4.11 2.780 18.6  1  1    4    2
```

```
fix(mtcars)
dim(mtcars)
```

```
## [1] 32 11
```

A forma mais direta de se obter um resumo estatístico das variáveis num 'data.frame' é através da função 'summary'

Com a Summary é possível apresentar estatísticas descritivas para as variáveis numéricas.

```
summary(mtcars)
```

```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##       am             gear            carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

Plotando um histograma do dataset mtcars

```
par(mfrow = c(2, 1))
hist(mtcars$mpg, main = "Distribution mpg - automatic transmission", xlab = "mpg")
```

Distribution mpg - automatic transmission