Aspectos básicos de los data frames

Ricardo

23/10/2020

Operaciones con Data frames básico

Veresmoe ejemplso de opraciones en data frames

El Iris data set de R. Fisher

Si buscamos en goolge iris data set

```
str(iris)
```

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 ...
```

```
str(iris3)
```

```
##  num [1:50, 1:4, 1:3] 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  - attr(*, "dimnames")=List of 3
##   ..$ : NULL
##   ..$ : chr [1:4] "Sepal L." "Sepal W." "Petal L." "Petal W."
##   ..$ : chr [1:3] "Setosa" "Versicolor" "Virginica"
```

```
class(iris)
```

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

```
str(iris)
```

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 ...
```

```r
class(iris$Sepal.Length)
```

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

```r
apply(iris,MARGIN=2,FUN=mean)
```

```
## Warning in mean.default(newX[, i], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(newX[, i], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(newX[, i], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(newX[, i], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(newX[, i], ...): argument is not numeric or logical:
## returning NA

## Sepal.Length  Sepal.Width Petal.Length  Petal.Width      Species
##           NA           NA           NA           NA           NA
```

```r
estadisticos=function(x) {c(Media=mean(x,na.rm=TRUE),
                            Desv_tipica=sd(x,na.rm=TRUE),
                            Min=min(x,bar.rm=TRUE)
)
  }
apply(iris[,1:4],MARGIN=2,FUN=mean)
```

```
## Sepal.Length  Sepal.Width Petal.Length  Petal.Width
##     5.843333     3.057333     3.758000     1.199333
```

```r
apply(iris[,-5],MARGIN=2,FUN=mean)
```

```
## Sepal.Length  Sepal.Width Petal.Length  Petal.Width
##     5.843333     3.057333     3.758000     1.199333
```

```r
apply(iris[,-5],MARGIN=2,FUN=estadisticos)
```

```
##             Sepal.Length Sepal.Width Petal.Length Petal.Width
## Media          5.8433333   3.0573333     3.758000   1.1993333
## Desv_tipica    0.8280661   0.4358663     1.765298   0.7622377
## Min            1.0000000   1.0000000     1.000000   0.1000000
```

```r
# help(aggregate)
aggregate(iris$Sepal.Length,by=list(iris$Species),FUN=mean)
```

```
##      Group.1     x
## 1     setosa 5.006
## 2 versicolor 5.936
## 3  virginica 6.588
```

```
lapply(iris,class)
```

```
## $Sepal.Length
## [1] "numeric"
##
## $Sepal.Width
## [1] "numeric"
##
## $Petal.Length
## [1] "numeric"
##
## $Petal.Width
## [1] "numeric"
##
## $Species
## [1] "factor"
```

```
dim(iris)
```

```
## [1] 150   5
```

```
dim(iris)[1]
```

```
## [1] 150
```

```
rownames(iris)
```

```
##   [1] "1"   "2"   "3"   "4"   "5"   "6"   "7"   "8"   "9"   "10"  "11"  "12"
##  [13] "13"  "14"  "15"  "16"  "17"  "18"  "19"  "20"  "21"  "22"  "23"  "24"
##  [25] "25"  "26"  "27"  "28"  "29"  "30"  "31"  "32"  "33"  "34"  "35"  "36"
##  [37] "37"  "38"  "39"  "40"  "41"  "42"  "43"  "44"  "45"  "46"  "47"  "48"
##  [49] "49"  "50"  "51"  "52"  "53"  "54"  "55"  "56"  "57"  "58"  "59"  "60"
##  [61] "61"  "62"  "63"  "64"  "65"  "66"  "67"  "68"  "69"  "70"  "71"  "72"
##  [73] "73"  "74"  "75"  "76"  "77"  "78"  "79"  "80"  "81"  "82"  "83"  "84"
##  [85] "85"  "86"  "87"  "88"  "89"  "90"  "91"  "92"  "93"  "94"  "95"  "96"
##  [97] "97"  "98"  "99"  "100" "101" "102" "103" "104" "105" "106" "107" "108"
## [109] "109" "110" "111" "112" "113" "114" "115" "116" "117" "118" "119" "120"
## [121] "121" "122" "123" "124" "125" "126" "127" "128" "129" "130" "131" "132"
## [133] "133" "134" "135" "136" "137" "138" "139" "140" "141" "142" "143" "144"
## [145] "145" "146" "147" "148" "149" "150"
```

```
rownames(iris)<- paste0("Flor ",rownames(iris))
rownames(iris)
```

```
##   [1] "Flor 1"   "Flor 2"   "Flor 3"   "Flor 4"   "Flor 5"   "Flor 6"
```

```
##   [7] "Flor 7"   "Flor 8"   "Flor 9"   "Flor 10"  "Flor 11"  "Flor 12"
##  [13] "Flor 13"  "Flor 14"  "Flor 15"  "Flor 16"  "Flor 17"  "Flor 18"
##  [19] "Flor 19"  "Flor 20"  "Flor 21"  "Flor 22"  "Flor 23"  "Flor 24"
##  [25] "Flor 25"  "Flor 26"  "Flor 27"  "Flor 28"  "Flor 29"  "Flor 30"
##  [31] "Flor 31"  "Flor 32"  "Flor 33"  "Flor 34"  "Flor 35"  "Flor 36"
##  [37] "Flor 37"  "Flor 38"  "Flor 39"  "Flor 40"  "Flor 41"  "Flor 42"
##  [43] "Flor 43"  "Flor 44"  "Flor 45"  "Flor 46"  "Flor 47"  "Flor 48"
##  [49] "Flor 49"  "Flor 50"  "Flor 51"  "Flor 52"  "Flor 53"  "Flor 54"
##  [55] "Flor 55"  "Flor 56"  "Flor 57"  "Flor 58"  "Flor 59"  "Flor 60"
##  [61] "Flor 61"  "Flor 62"  "Flor 63"  "Flor 64"  "Flor 65"  "Flor 66"
##  [67] "Flor 67"  "Flor 68"  "Flor 69"  "Flor 70"  "Flor 71"  "Flor 72"
##  [73] "Flor 73"  "Flor 74"  "Flor 75"  "Flor 76"  "Flor 77"  "Flor 78"
##  [79] "Flor 79"  "Flor 80"  "Flor 81"  "Flor 82"  "Flor 83"  "Flor 84"
##  [85] "Flor 85"  "Flor 86"  "Flor 87"  "Flor 88"  "Flor 89"  "Flor 90"
##  [91] "Flor 91"  "Flor 92"  "Flor 93"  "Flor 94"  "Flor 95"  "Flor 96"
##  [97] "Flor 97"  "Flor 98"  "Flor 99"  "Flor 100" "Flor 101" "Flor 102"
## [103] "Flor 103" "Flor 104" "Flor 105" "Flor 106" "Flor 107" "Flor 108"
## [109] "Flor 109" "Flor 110" "Flor 111" "Flor 112" "Flor 113" "Flor 114"
## [115] "Flor 115" "Flor 116" "Flor 117" "Flor 118" "Flor 119" "Flor 120"
## [121] "Flor 121" "Flor 122" "Flor 123" "Flor 124" "Flor 125" "Flor 126"
## [127] "Flor 127" "Flor 128" "Flor 129" "Flor 130" "Flor 131" "Flor 132"
## [133] "Flor 133" "Flor 134" "Flor 135" "Flor 136" "Flor 137" "Flor 138"
## [139] "Flor 139" "Flor 140" "Flor 141" "Flor 142" "Flor 143" "Flor 144"
## [145] "Flor 145" "Flor 146" "Flor 147" "Flor 148" "Flor 149" "Flor 150"
```

```
head(iris)
```

```
##        Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## Flor 1          5.1         3.5          1.4         0.2  setosa
## Flor 2          4.9         3.0          1.4         0.2  setosa
## Flor 3          4.7         3.2          1.3         0.2  setosa
## Flor 4          4.6         3.1          1.5         0.2  setosa
## Flor 5          5.0         3.6          1.4         0.2  setosa
## Flor 6          5.4         3.9          1.7         0.4  setosa
```

```
colnames(iris)
```

```
## [1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
```

```
names(iris)
```

```
## [1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
```

```
table(iris$Species)
```

```
##
##     setosa versicolor  virginica
##         50         50         50
```

```r
prop.table(table(iris$Species))
```

```
##
##     setosa versicolor  virginica
##  0.3333333  0.3333333  0.3333333
```

```r
A=matrix(rnorm(20),byrow=TRUE,ncol=4)
class(A)
```

```
## [1] "matrix" "array"
```

```r
dim(A)
```

```
## [1] 5 4
```

```r
dimnames(A)<-list(letters[1:5],letters[1:4])
dimnames(A)
```

```
## [[1]]
## [1] "a" "b" "c" "d" "e"
##
## [[2]]
## [1] "a" "b" "c" "d"
```

```r
x=1:10
x%%2==0
```

```
##  [1] FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE
```

```r
lapply(x, FUN=function(y) y%%2==0)
```

```
## [[1]]
## [1] FALSE
##
## [[2]]
## [1] TRUE
##
## [[3]]
## [1] FALSE
##
## [[4]]
## [1] TRUE
##
## [[5]]
## [1] FALSE
##
## [[6]]
## [1] TRUE
##
## [[7]]
```

```
## [1] FALSE
##
## [[8]]
## [1] TRUE
##
## [[9]]
## [1] FALSE
##
## [[10]]
## [1] TRUE
```

```
lapply(x, FUN=function(y) y%%2==0)
```

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

```
unlist(lapply(x, FUN=function(y) y%%2==0))
```

```
##  [1] FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE
```