Introducción a R

Oscar Perpiñán Lamigueiro http://oscarperpinan.github.io

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software fibre

,

2...1...

\- (- - - - - T

¿Qué es R?

R está muy bien documentado R es un proyecto colaborativo

Objetos en R

Indexado

Bucles

Datos con R

¿Qué es R?

¿Qué es R?

Es un lenguaje de programación principalmente orientado al análisis estadístico y visualización de información cuantitativa y cualitativa y publicado como software libre con licencia GNU-GPL.

http://www.R-project.org

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

¿Qué es R?

R está muy bier documentado

R es un proyecto colaborativo

Objetos en F

Indexad

Bucles

. .

Patos con R

Para instalar R

- Windows: http: //cran.es.r-project.org/bin/windows/base/
- ► Mac: http://cran.es.r-project.org/bin/macosx/
- ► Linux: http://cran.es.r-project.org/bin/linux/

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

¿Qué es R?

R está muy bien documentado

R es un proyecto colaborativo

Objetos en F

naexaa

Bucles

Datos con P

Interfaces para R

- En mi opinión, la mejor interfaz para R es ESS con Emacs.
- Para los que prefieren una interfaz gráfica es recomendable RStudio:
 - ► Instalador:

http://www.rstudio.com/ide/download/desktop

► Introducción:

http://www.rstudio.com/ide/docs/using/source

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

¿Qué es R?

R está muy bier documentado

R es un proyect colaborativo

Objetos en I

Bucles

Datos con R

¿Qué es R?

R está muy bien documentado

R es un proyecto colaborativo

Objetos en R

Indexado

Bucles

Datos con R

Fuentes de Información

- Manuales Oficiales
 - ► Introduction to R
 - ► R Data Import/Export
 - ► R Installation and Administration
 - Writing R Extensions
 - ► R language definition
 - ► R Internals
- ▶ Manuales externos

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

¿Qué es R?

R está muy bien documentado

R es un proyect colaborativo

Objetos en l

пислии

ucles

Sakaa aan

/isualización d

Fuentes de Información

- Listas de correo (sin olvidar respetar estos consejos)
 - ► Generales: R-announce, R-help, R-devel
 - ► Special Interest Group (SIG) mailing lists
- ► R-bloggers
- ▶ stackoverflow

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

¿Qué es R?

R está muy bien documentado

K es un proy colaborativo

Objetos en 1

пислас

Bucles

. . .

isualización de

¿Qué es R? R está muy bien documentado R es un proyecto colaborativo

Objetos en R

Indexado

Bucles

Datos con R

R es un proyecto colaborativo

- Una de las grandes riquezas de R es la cantidad de paquetes (más de 6000 actualmente) que amplían sus funcionalidades.
- ► La lista completa está en http://cran.es.r-project.org/web/packages/.
- ► Las CRAN Task Views agrupan por temáticas: http://cran.r-project.org/web/views/

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

¿Qué es R? R está muy h

R es un proyecto

Objetos en R

ucies

atos con R

CRAN Task Views

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

Introducción a R

R es software libre

Qué es R?

R es un provecto

colaborativo

. . .

D 1

oucles

Datos con P

Alos Coll IC

Visualización de datos

Big Data

- ► High-Performance and Parallel Computing with R
- Machine Learning & Statistical Learning
- ► Web Technologies and Services
- **...**

Más de 6000 paquetes disponibles

► Algunos vienen instalados y se cargan al empezar:

sessionInfo()

R version 3.1.1 (2014-07-10)

```
Platform: i586-pc-linux-gnu (32-bit)
locale:
[1] LC CTYPE=es ES.utf8
                              L.C NUMERIC=C
[3] LC TIME=es ES.utf8
                            LC COLLATE=es ES.utf8
[5] LC MONETARY-es ES.utf8 LC MESSAGES-es ES.utf8
[7] LC PAPER=es ES.utf8
                              LC_NAME=C
[9] LC ADDRESS=C
                              LC TELEPHONE=C
[11] LC MEASUREMENT-es ES.utf8 LC IDENTIFICATION-C
attached base packages:
[1] stats
             graphics grDevices utils
                                           datasets methods
                                                               hase
other attached packages:
[1] RColorBrewer 1.0-5 lattice 0.20-27
loaded via a namespace (and not attached):
[1] compiler_3.1.1 grid_3.1.1 tools_3.1.1
```

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

¿Qué es R?

R es un proyecto colaborativo

Objetos en I

ucies

Datos con R

visualización de latos

Más de 6000 paquetes disponibles

Otros vienen instalados pero hay que cargarlos:

```
library(lattice)
packageVersion('lattice')
```

[1] '0.20.27'

Package: lattice Version: 0.20-27 Date: 2014/02/27

Repository: CRAN

packageDescription('lattice')

```
Priority: recommended
Title: Lattice Graphics
Author: Deepayan Sarkar <deepayan.sarkar@r-project.org>
Maintainer: Deepayan Sarkar <deepayan.sarkar@r-project.org>
Description: Lattice is a powerful and elegant high-level data
        visualization system, with an emphasis on multivariate data,
        that is sufficient for typical graphics needs, and is also
        flexible enough to handle most nonstandard requirements. See
        ?Lattice for an introduction.
Depends: R (>= 2.15.1)
Suggests: KernSmooth, MASS
Imports: grid, grDevices, graphics, stats, utils
Enhances: chron
LazyLoad: yes
LazvData: ves
License: GPL (>= 2)
URL: http://lattice.r-forge.r-project.org/
Packaged: 2014-02-27 06:16:44 UTC; deepayan
NeedsCompilation: ves
```

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

R está muy bien locumentado

R es un proyecto

Obietos en I

пиехаис

Bucles

Sakaa aan

...

atos

Más de 6000 paquetes disponibles

Otros hay que instalarlos y después cargarlos:

```
install.packages('data.table')
library('data.table')
packageDescription('data.table')
```

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

¿Qué es R? R está muy b

R es un proyecto colaborativo

Objetos en R

пислии

ucies

Datos con R

Objetos en R

- Existen varios objetos en R:
 - Vectores
 - Listas
 - Funciones
 - **.**..
- ▶ A partir de estos objetos se definen varias clases:
 - ▶ matrix
 - ▶ data.frame
 - factor
 - ▶ Date, POSIXct
 - **.**..

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

Objetos en R

Matrices

Lietze

Data.frame

. .

пиехацо

Datos con F

Objetos en R

Vectores

Matrices

Listas

Data.frame

Funciones

Indexado

Bucles

Datos con R

Primeros pasos

x <- 1:5

X

[1] 1 2 3 4 5

length(x)

[1] 5

class(x)

[1] "integer"

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

Objetos en R

Vectores

Matrices Listas

Data.frame

ndovado

Datos con R

Generar vectores con seq

```
x1 <- seq(1, 100, by=2)
x1
```

[1] 1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 [26] 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99

```
seq(1, 100, length=10)
```

[1] 1 12 23 34 45 56 67 78 89 100

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

x es sortware in

Objetos en R

Vectores

Matrices

Listas

Data.iraine

Indexa

Bucles

Bucles

Datos con R

Unir vectores con c

```
x <- c(1, 2, 3)
x
```

[1] 1 2 3

```
x <- seq(1, 100, length=10)
y <- seq(2, 100, length=50)
z <- c(x, y)
```

Z

```
[1]
                       45
                                                                                   18
[20]
              24
                  26
                       28
                                                 40
                                                                  48
                                                                       50
                                                                               54
                                                                                   56
[39]
          60
              62
                       66
                              70
[58]
          98 100
```

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

Objetos en R

Vectores

Matrices

Data.frame

пислии

Bucles

atos con R

Operaciones sencillas con vectores

[1] 2 3 4 5 6

x^2

[1] 1 4 9 16 25

[1] 2 4 6 8 10 7 9 11 13 15

[1] 1 4 9 16 25 6 14 24 36 50

$$x^2 + y^3$$

[1] 2 12 36 80 150 217 347 521 745 1025

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

Objetos en R

Vectores

Matrices Listas

Data.frame Funciones

ndexado

Bucles

Datos con R

Visualización de

Objetos en R

Vectores

Matrices

Listas

Data.frame

Funciones

Indexado

Bucles

Datos con R

М

```
[1,1] [,2] [,3] [,4]
[1,] 1 4 7 10
[2,] 2 5 8 11
[3,] 3 6 9 12
```

class(M)

[1] "matrix"

dim(M)

[1] 3 4

summary(M)

```
V1
                  V2
                              V3
                                           V4
Min
     :1 0
            Min
                   :4.0
                         Min
                               · 7 0
                                      Min
                                            :10 0
1st Qu.:1.5
            1st Qu.:4.5
                        1st Qu.:7.5
                                      1st Qu.:10.5
Median :2 0
            Median :5 0
                        Median :8 0
                                      Median :11 0
Mean
      :2.0
            Mean
                 :5.0
                        Mean
                               :8.0
                                     Mean
                                            :11.0
3rd Qu.: 2.5
            3rd Qu.:5.5
                         3rd Qu.:8.5
                                      3rd Qu.:11.5
Max.
      :3.0
            Max.
                   :6.0
                         Max.
                                            :12.0
                                      Max.
```

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

Objetos en R

Matrices

Matrices

Data fran

Funciones

naexaa

Bucles

Datos con R

```
x y z
[1,] 1 1 1
[2,] 2 2 2
[3,] 3 3 3
[4,] 4 4 4
[5,] 5 5 5
[6,] 6 6 6
```

[7,] [8,] [9.]

М

[10,] 10 10 10

M <- rbind(x, y, z)
M</pre>

```
[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] 
x 1 2 3 4 5 6 7 8 9 10 
y 1 2 3 4 5 6 7 8 9 10 
z 1 2 3 4 5 6 7 8 9 10
```

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

K es software libre

Objetos en R

Matrices

Matrices

Listas

Funciones

ndexad

lucles

Datos con R

/isualización de latos

Objetos en R

Vectores

Matrices

Listas

Data.frame

Funciones

Indexado

Bucles

Datos con R

Para crear una lista usamos la función list

class(lista)

[1] "list"

length(lista)

[1] 3

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

Objetos en

Matrices

Lietae

Data.frame

unciones

ucles

Datos con R

Visualización d

Objetos en R

Vectores

Matrices

Listas

Data.frame

Funciones

Indexado

Bucles

Datos con R

Para crear un data frame...

```
1 1 - 0.7545423 0

2 2 - 1.2616725 0

3 3 0.7946587 0

4 4 - 0.9536974 0

5 5 1.7672158 0

6 1 1.3973362 0

7 2 - 0.1051969 0

8 3 1.0237416 0

9 4 - 0.1157114 0

10 5 - 0.3946580 0
```

length(df)

[1] 3

dim(df)

[1] 10 3

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

Objetos en R

Vectores Matrices

Data.frame

uncione

Bucles

Datos con R

Objetos en R

Vectores

Matrices

Listas

Data.frame

Funciones

Indexado

Bucles

Datos con R

Para definir una función usamos la función function

```
myFun <- function(x, y) x + y
myFun</pre>
```

function(x, y) x + y

class(myFun)

[1] "function"

myFun(3, 4)

[1] 7

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

bjetos en R

Matrices

Listas

Punciones

Bucles

Datos con R

Visualización de

Podemos construir a partir de funciones

```
foo <- function(x, ...){
  mx <- mean(x, ...)
  medx <- median(x, ...)
  sdx <- sd(x, ...)
  c(mx, medx, sdx)
}</pre>
```

O en forma resumida:

```
foo <- function(x, ...){c(mean(x, ...), median(x, ...), sd(x, ...)}
```

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

Objetos en R

Vectores Matrices

Listas

Punciones

runciones

Bucles

Datos con R

Jatos con K

Y ahora usamos la función con vectores

foo(1:10)

[1] 5.50000 5.50000 3.02765

foo(rnorm(1e5))

[1] 1.931042e-03 -3.182508e-05 1.001589e+00

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

ix es software nore

Objetos en R

Matrices

Listas

ata.fram

Funciones

111010710101

Bucles

Datos con R

Visualización de

Objetos en R

Indexado

Vectores

Matrices

Listas

Data Frame

Bucles

Datos con R

Indexado numérico

[1] 1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 [26] 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99

x[1:5]

[1] 1 3 5 7 9

x[10:5]

[1] 19 17 15 13 11 9

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

--,----

ndexado

Vectores

Matrices

istas lata Framo

ucles

Datos con R

Indexado con condiciones lógicas

$$x == 37$$

- [1] FALSE FALSE
- [25] FALSE FALSE
- [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
- [49] FALSE FALSE

$$x[x == 37]$$

[1] 37

$$x[x != 9]$$

- 7 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 [26] 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99

- [1] 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69
- [26] 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

Vectores

Indexado con condiciones múltiples

$$z \leftarrow seq(-10, 10, by = .5)$$

z[z < -5 | z > 5]

cond <-
$$(z >= 0 \& z <= 5)$$

cond

- [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
- [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE TRUE TRUE
- [25] TRUE TRUE TRUE TRUE TRUE TRUE TRUE FALSE FALSE FALSE FALSE
- [37] FALSE FALSE FALSE FALSE

z[cond]

[1] 0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

--,----

ndexado

Vectores

Matrices Listas

Listas Data Fran

Ruclas

atos con R

Objetos en F

Indexado

Vectores

Matrices

Listas

Data Frame

Bucles

Datos con R

Indexado de matrices

M[1:2,]

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

    x
    1
    2
    3
    4
    5
    6
    7
    8
    9
    10

    y
    1
    2
    3
    4
    5
    6
    7
    8
    9
    10
```

M[1:2, 2:3]

```
[,1] [,2]
x 2 3
y 2 3
```

M[1, c(1, 4)]

[1] 1 4

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

Objetos en K

ndexado

Vectores Matrices

Matrices

Data Frame

ucles

Datos con R

Indexado de matrices

M[-1,]

```
[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
y 1 2 3 4 5 6 7 8 9 10
z 1 2 3 4 5 6 7 8 9 10
```

M[-c(1, 2),]

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

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

Objetos en R

Indexado

Matrices

stas ata Framo

Bucles

Datos con R

/isualización de latos

Objetos en R

Indexado

Vectores

Matrices

Listas

Data Frame

Bucles

Datos con R

Podemos acceder a los elementos...

Por su nombre

lista\$a

[1] 1 3 5

o por su índice

lista[1]

\$a [1] 1 3 5

lista[[1]]

[1] 1 3 5

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

Objetos en r

Indexado

Vectores

Listas

ata Frame

ıcles

Datos con R

/isualización de latos

Objetos en R

Indexado

Vectores

Matrices

Listas

Data Frame

Bucles

Datos con R

Podemos acceder a los elementos

Por su nombre (como una lista)

```
df$x
```

```
[1] 1 2 3 4 5 1 2 3 4 5
```

▶ Por su índice (como una matriz)

```
df[1,]
```

```
x yz
11-0.75454230
```

```
df[,1]
```

```
[1] 1 2 3 4 5 1 2 3 4 5
```

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

Objetos en r

Indexado

Vectores Matrices

Listas Data Frame

1

ucies

atos con F

Jatos con K

Objetos en R

Indexado

Bucles

Matrices

Listas / data.frame

Datos con R

La función apply

```
apply(M, 1, sum)
```

x y z 55 55 55

rowSums(M)

x y z 55 55 55

apply(M, 2, mean)

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

colMeans(M)

[1] 1 2 3 4 5 6 7 8 9 1

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

Objetos eri

ndexado

ucles

Matrices

Listas / data.frame

Visualización de

Objetos en R

Indexado

Bucles

Matrices

Listas / data.frame

Datos con R

lapply y sapply

```
$y
[1] 30
```

[1] 55

\$x

\$z [1] 4.45118

sapply(lista, sum)

```
x y z
55.00000 30.00000 4.45118
```

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

--,----

Indexado

Bucles

Listas / data.frame

atos con R

Objetos en R

Indexado

Bucles

Datos con R
Lectura de datos
Indexado
Datos agregados
Regresión lineal

setwd, getwd, dir

```
getwd()
old <- setwd("~/github/r-intro-eoi")
dir()</pre>
```

dir(pattern='.R')

[1] "intro.R"

dir('data')

[1] "aranjuez.csv"

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

,

Lectura de datos

Indexado

Datos agregados Regresión lineal

/isualización de latos dats <- read.table('data/aranjuez.csv')</pre>

head(dats)

V1

2 2004-01-01 3 2004-01-02

4 2004-01-03 5 2004-01-04 6 2004-01-05

3

Lectura de datos

```
1 ,"TempAvg","TempMax","TempMin","HumidAvg","HumidMax","WindAvg","WindMax","Rain","Radiation","ETwización de
```

```
,4.044,10.71,-1.969,88.3,95.9,0.746,3.528,0,5.49,0.535268783569336
```

```
.5.777.11.52.1.247.83.3.98.5.1.078.6.88.0.6.537.0.771049916744232
```

Lectura de datos con read.table

```
dats <- read.table('data/aranjuez.csv', sep=',',
    header=TRUE)
head(dats)</pre>
```

```
X TempAvg TempMax TempMin HumidAvg HumidMax WindAvg WindMax Rain
               4 044
                             -1 969
1 2004-01-01
                       10 71
                                         88 3
                                                  95 9
                                                         0 746
                                                                 3 528
2 2004-01-02
               5.777
                       11.52
                             1.247
                                         83.3
                                                  98.5
                                                         1.078
                                                                 6.880
                                                                          0
3 2004-01-03
              5.850
                       13.32
                             0.377
                                         75.0
                                                  94.4
                                                         0.979
                                                                 6.576
             4.408
4 2004-01-04
                      15.59
                             -2.576
                                         82.0
                                                  97.0
                                                         0.633
                                                                 3.704
5 2004-01-05
             3.081 14.58 -2.974
                                         83.2
                                                  97.0
                                                         0.389
                                                                 2.244
6 2004-01-06
               2.304
                       11.83 -3.379
                                         84.5
                                                  96.5
                                                         0.436
                                                                 2.136
  Radiation
                   ET
      5.490 0.5352688
     6.537 0.7710499
     8 810 0 8361229
      9 790 0 6861381
    10.300 0.5152422
      9 940 0 4886631
```

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

Objetos eri

Indexado

ducle

atos con R

Lectura de datos

Datos agregados

Regresión lineal

Lectura de datos con read. table

```
aranjuez <- read.csv('data/aranjuez.csv')
head(aranjuez)</pre>
```

```
X TempAvg TempMax TempMin HumidAvg HumidMax WindAvg WindMax Rain
              4.044
                       10.71
                             -1.969
                                         88.3
                                                  95.9
                                                         0.746
                                                                 3.528
1 2004-01-01
              5 777
2 2004-01-02
                       11 52
                             1 247
                                         83 3
                                                  98 5
                                                         1 078
                                                                 6 880
3 2004-01-03
             5.850
                       13 32
                             0.377
                                         75.0
                                                  94 4
                                                         0 979
                                                                 6.576
4 2004-01-04
             4.408
                      15.59 -2.576
                                         82.0
                                                  97.0
                                                         0.633
                                                                 3.704
             3 081
5 2004-01-05
                       14 58 -2 974
                                         83 2
                                                  97 0
                                                         0.389
                                                                 2 244
6 2004-01-06
               2 304
                       11.83 -3.379
                                         84 5
                                                  96 5
                                                         0 436
                                                                 2 136
 Radiation
     5.490 0.5352688
     6 537 0 771 04 99
     8.810 0.8361229
    9.790 0.6861381
    10 300 0 5152422
     9 940 0 4886631
```

class(aranjuez)

[1] "data.frame"

names(aranjuez)

```
[1] "X" "TempAvg" "TempMax" "TempMin" "HumidAvg" "HumidMax" [7] "WindAvg" "WindMax" "Rain" "Radiation" "ET"
```

4 日) 4 周) 4 三) 4 三)

= 900 €

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

--,-----

Indexado

Bucles

Patos con R

Lectura de datos

Indexado

Datos agregados Regresión lineal

Objetos en R

Indexado

Bucles

Datos con R

Lectura de datos

Indexado

Datos agregados

Regresión linea



```
X TempAvg TempMax TempMin HumidAvg HumidMax WindAvg WindMax Rain
              4 044
                                                                3.528
1 2004-01-01
                      10 71
                             -1.969
                                        88 3
                                                 95.9
                                                         0.746
2 2004-01-02
             5.777
                      11.52
                             1.247
                                        83.3
                                                 98.5
                                                        1.078
                                                                6.880
3 2004-01-03
             5.850
                      13.32
                             0.377
                                        75.0
                                                 94.4
                                                        0.979
                                                                6.576
                                                                          0
4 2004-01-04
             4.408
                      15.59 -2.576
                                        82.0
                                                 97 0
                                                        0 633
                                                                3.704
                                                                          0
5 2004-01-05
             3.081
                      14.58 -2.974
                                        83.2
                                                 97.0
                                                        0.389
                                                                2.244
 Radiation
     5 490 0 5352688
     6.537 0.7710499
    8.810 0.8361229
     9 790 0 6861381
    10 300 0 5152422
```

► Filas y Columnas

aranjuez[10:14, 1:5]

```
X TempAvg TempMax TempMin HumidAvg
10 2004-01-10
                10 85
                        16 59
                                 5 676
                                           84 9
               7.59
                       9.23
                                4.806
                                          95.4
11 2004-01-11
               7.41
                        10.24
                                5.200
                                          93.1
12 2004-01-12
13 2004-01-13
                 8.35
                        11 38
                                4 137
                                          91 3
14 2004-01-14
                 8.74
                        13.32
                                 2.857
                                           86.9
```

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libr

Objetos en

Indexado

Bucles

atos con P

Lectura de datos

atos agregados

Regresión lineal

/isualización d latos

Indexado con []

Condición basada en los datos

```
idx <- with(aranjuez, Radiation > 20 & TempAvg < 10)
head(aranjuez[idx, ])</pre>
```

```
X TempAvg TempMax TempMin HumidAvg HumidMax WindAvg WindMax Rain
   2004-03-22
                 9.78
                       16.12
                               4.340
                                        51.65
                                                 87.9
                                                        1.526
                                                                7.660
                8 50
                              -0.290
                                                                6 027
   2004-03-23
                       15 52
                                       50 10
                                                 83 3
                                                        1 533
                              1.584
                                              76.6
                                                               5.939
   2004-03-25
               7.47
                      14.58
                                       49.66
                                                       1.138
100 2004-04-09
               8.83
                      15.52
                              2.056
                                       47.50
                                              70.8
                                                        1.547
                                                               6.125
                              -0 155
                                       54 45
101 2004-04-10
                7 04
                      13 85
                                                85.8
                                                        1 448
                                                                6 958
102 2004-04-11
                7.50
                       15.19
                              -1.699
                                       54.98
                                                 91.0 1.126
                                                               7 590
   Radiation
       21.92 3.075785
82
83
       20.62 2.881419
       22.44 2.849603
100
    25.45 3.566452
101
       21.07 2.943239
102
       20.99 2.905479
```

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software fibre

Objetos en K

Bucles

itos con R

Indevado

Datos agregado:

Regresión lineal

⁷isualización de latos

```
Radiation TempAvg TempMax TempMin
82
         21.92
                 9.780
                         16.12
                                  4.340
83
         20.62
                 8 500
                         15.52
                                 -0.290
                                 1.584
85
         22.44
                 7.470
                         14.58
100
         25.45
                 8.830
                         15.52
                                 2.056
101
         21.07
                 7.040
                        13.85
                                 -0.155
                                 -1 699
102
         20.99
                 7.500
                        15.19
104
         25.76
                 9.420
                        17.47
                                 0.115
461
         24.29
                 7.460
                        14.66
                                 -0.081
462
         25.25
                 7.930
                         17.35
                                 -1 686
463
         24.56
                 9.800
                         19.08
                                 -1.484
1146
         20.08
                 7.170
                         18.20
                                 -3.746
         20 90
                 4 378
1157
                         12 03
                                 -6 353
         21 87
                 7 920
                                 -2 941
1159
                         18 54
1160
         20.35
                 7.830
                         16.49
                                 -2.807
1521
         21.54
                 8.100
                         19.29
                                 -4.075
2244
         20.49
                 6 121
                         15.15
                                 -0.940
2245
         21.02
                 5.989
                        16.94
                                 -3.208
2246
         20.22
                 9.020
                         19.74
                                 -2.068
2261
         23 00
                 9 500
                         14 96
                                  3 662
2262
         20.40
                 9.910
                         14.70
                                 4.668
2263
         24.09
                 9.440
                         16.89
                                 0.794
                 9.680
2265
         23.64
                         16.35
                                  2.938
2295
         22 46
                 8 730
                         13 84
                                  1.740
```

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libr

Objetos er

Indexado

Bucles

tos con R

Indexado

Datos agregados Regresión lineal

/isualización d latos

Objetos en R

Indexado

Bucles

Datos con R

Lectura de datos Indexado

Datos agregados

Regresión lineal

aggregate

rainy Radiation 1 FALSE 19.63325 2 TRUE 10.26028 Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

K es software fibre

Indexado

Bucles

atos con R

Indexado

Datos agregados

Regresión lineal

/isualización de latos

Variable categórica con cut

(16.3,23.5] TRUE 15.802522

TRUE 22.545862

10 (23.5.30.7]

```
aranjuez$tempClass <- cut(aranjuez$TempAvg, 5)</pre>
aggregate(Radiation ~ tempClass, data = aranjuez,
          FUN = mean)
    tempClass Radiation
1 (-5.34.1.89] 8.805389
2 (1.89.9.09] 9.014178
3 (9.09,16.3] 14.554177
4 (16.3.23.5] 21.912414
5 (23.5.30.7] 26.192742
aggregate(Radiation ~ tempClass + rainy,
          data = aranjuez, FUN = mean)
    tempClass rainy Radiation
1 (-5.34.1.89] FALSE 9.869134
 (1.89,9.09] FALSE 10.718837
 (9.09.16.3] FALSE 17.238283
 (16.3.23.5] FALSE 23.238145
 (23.5,30.7] FALSE 26.392665
 (-5.34,1.89] TRUE
                  6.822955
7 (1.89.9.09] TRUE
                 7 063932
8 (9.09,16.3] TRUE 11.091063
```

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

--,------

Indexado

Bucles

atos con R

dexado

Datos agregados

Regresión line

Fechas

as.Date

```
aranjuez$date <- as.Date(aranjuez[,1],</pre>
                 format='%Y-%m-%d')
```

Funciones para extraer mes y año

```
Year <- function(x)as.numeric(format(x, "%Y"))</pre>
Month <- function(x)as.numeric(format(x, "%m"))</pre>
```

Añado columnas en data, frame

```
aranjuez$month <- Month(aranjuez$date)</pre>
aranjuez$year <- Year(aranjuez$date)</pre>
aranjuez$quarter <- quarters(aranjuez$date)</pre>
```

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

Datos agregados

ndexado

Datos agregados

Ziemalización :

```
year Radiation
1 2004 16.39449
2 2005 17.34966
3 2006 16.00713
4 2007 16.40107
5 2008 16.19843
6 2009 17.44067
7 2010 16.66956
8 2011 17.47265
```

```
month year Radiation
       1 2004 7.773161
       2 2004 10.329241
       3 2004 13 290032
      4 2004 20.314500
       5 2004 20.925000
       6 2004 28.436667
       7 2004 26.897097
       8 2004 22.999032
       9 2004 19.156733
      10 2004 11 427774
11
      11 2004 8.668667
12
      12 2004 6.479419
13
       1 2005
              9.845774
14
       2 2005 11.310240
15
       3 2005 16.141000
```

Agregamos varias variables con fechas

```
month year Radiation
                          TempAvg
              7.773161
                         6.096258
       1 2004
       2 2004 10.329241
                         6 458621
3
       3 2004 13.290032
                        8.761290
       4 2004 20.314500 11.834333
       5 2004 20.925000 15.095161
       6 2004 28.436667 24.276000
       7 2004 26.897097 25.036774
       8 2004 22.999032 23.569355
       9 2004 19.156733 20.823667
10
      10 2004 11.427774 14.745161
11
      11 2004
              8.668667 6.803933
      12 2004
              6.479419
                         5.082161
13
       1 2005
               9.845774 1.767387
14
       2 2005 11.310240 3.685400
       3 2005 16.141000
15
                        9 976600
16
       4 2005 21.693333 13.966667
17
       5 2005 23.938387 19.226452
18
       6 2005 26.779333 25.191333
19
       7 2005 27.256774 25.983226
20
       8 2005 23.423871 24.433871
21
       9 2005 19.575333 19.292000
      10 2005 12.052581 14.588710
23
              8.021000 7.184733
      11 2005
24
      12 2005
              7.165516
                         3.991032
25
       1 2006
               7.082871
                         3 984581
26
       2 2006 10.224857
                         4.737500
27
       3 2006 15.199323 10.308065
28
       4 2006 20 135000 14 314333
29
       5 2006 23.439815 18.887407
~ ~
       C 000C OF 040000 00 C00400
```

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

Objetos en

Indexado

Bucles

tos con R

Indexado Datos agregados

legresión lineal

Objetos en R

Indexado

Bucles

Datos con R

Lectura de datos Indexado Datos agregados

Regresión lineal

1m: Regresión simple

```
radTemp <- lm(Radiation ~ TempAvg, data = aranjuez)
summary(radTemp)</pre>
```

```
Call:
lm(formula = Radiation ~ TempAvg, data = aranjuez)
Residuals:
    Min
            10 Median 30
                                     Max
-20 1437 -3 6000 0 2903 3 7191 15 8617
Coefficients:
          Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.07258
                      0.20585 24.64 <2e-16 ***
TempAvg 0.80932 0.01244 65.08 <2e-16 ***
Signif. codes: 0 '***, 0.001 '**, 0.01 '*, 0.05 '., 0.1 ', 1
Residual standard error: 5.431 on 2883 degrees of freedom
 (13 observations deleted due to missingness)
Multiple R-squared: 0.595, Adjusted R-squared: 0.5948
F-statistic: 4235 on 1 and 2883 DF. p-value: < 2.2e-16
```

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

Objetos en

Indexado

Bucles

atos con R

Indexado

Datos agregados

Regresión lineal

1m: Regresión multiple

summary(radTempHumid)

Call:

```
Residuals:
    Min
             10 Median 30 Max
-22.1493 -2.7493 0.0142 2.7647 13.9198
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept) 34.5033445 0.8610398 40.072 < 2e-16 ***
              0 0527658 0 0410350 1 286 0 199
TempAvg
HumidAvg -0.3588349 0.0115441 -31.084 < 2e-16 ***
TempAvg:HumidAvg 0.0047893 0.0006621 7.233 6.03e-13 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
Residual standard error: 4.236 on 2881 degrees of freedom
  (13 observations deleted due to missingness)
Multiple R-squared: 0.7537, Adjusted R-squared: 0.7535
F-statistic: 2939 on 3 and 2881 DF, p-value: < 2.2e-16
```

lm(formula = Radiation ~ TempAvg * HumidAvg. data = aranjuez)

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

Objetos err.

Indexado

Bucles

atos con R

ndexado

Datos agregados

Regresión lineal

1m: Regresión multiple

```
radTempRain <- lm(Radiation ~ TempAvg * rainy,
                     data = aranjuez)
summary(radTempRain)
Call:
lm(formula = Radiation ~ TempAvg * rainy, data = aranjuez)
Residuals:
            10 Median 30
    Min
                                  Max
-17.0597 -3.3220 0.0441 3.3454 15.0028
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)
               7.72388 0.24232 31.874 < 2e-16 ***
               0 73334 0 01328 55 204 < 2e-16 ***
TempAvg
rainvTRUE -3.78620 0.41024 -9.229 < 2e-16 ***
TempAvg:rainyTRUE -0.12163 0.03077 -3.953 7.89e-05 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
Residual standard error: 4.929 on 2881 degrees of freedom
  (13 observations deleted due to missingness)
Multiple R-squared: 0.6665, Adjusted R-squared: 0.6662
F-statistic: 1920 on 3 and 2881 DF, p-value: < 2.2e-16
```

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

Objetoo eri

Indexado

Bucles

atos con R

Indexado

Datos agregados

Regresión lineal

Objetos en R

Indexado

Bucles

Datos con R

Visualización de datos Catálogo Aiustes

Lattice

► Documentación: Código y Figuras del libro

library(lattice)

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software fibre

,

пиехацо

Bucle

atos con R

Visualización de

Catálogo

xyplot

xyplot(Radiation ~ TempAvg, data=aranjuez)

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libr

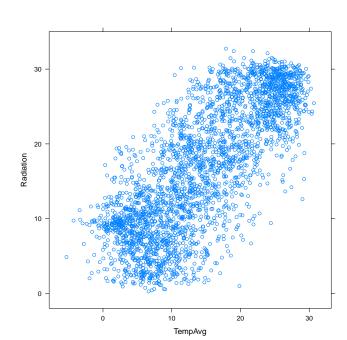
--,------

lucles

atos con R

Visualización de

Catálogo



Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

--,-----

ucies

atos con R

Visualización de datos

Catálogo Ajustes

Paneles

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

K es software fibre

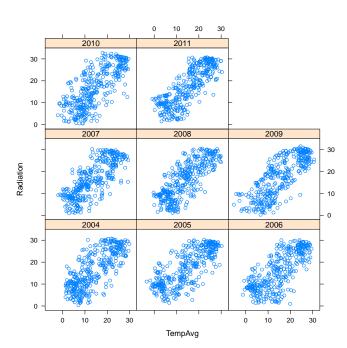
пислаи

ucles

atos con R

Visualización de datos

Catálogo Aiustes



Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

Objetos en

ndexado

ucles

atos con R

Visualización de datos

Catálogo Aiustes

Grupos

 Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

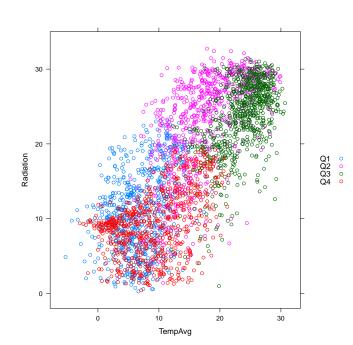
,

lucles

atos con R

Visualización de

Catálogo Aiustes



Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

Objetos er

шелаці

icles

tos con R

datos

Matriz de gráficos de dispersión

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

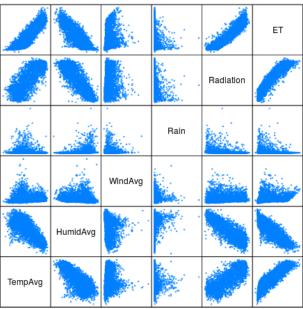
R es software libre

пиехацо

ucles

Datos con R

Visualización de



Scatter Plot Matrix

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

пислацо

ucies

Datos con R

Visualización de datos

Matriz de gráficos de dispersión

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

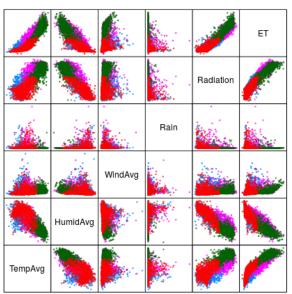
bjetos en K

ndexado

Bucles

atos con R

Visualización de



Scatter Plot Matrix

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

,

ndexado

cles

atos con R

Visualización de

Catálogo

Q1 Q2 Q3 Q4

Box-and-Whiskers

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

Objetos ert i

ucles

atos con R

Visualización de datos



Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

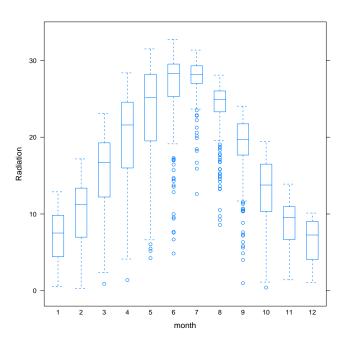
K es software libre

'

Bucles

Datos con R

Visualización de datos



Histogramas

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

Objetos CII

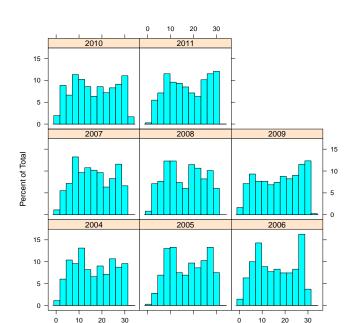
naexado

ucles

histogram(~ Radiation|factor(year), data=aranjuez)

Dutoo com re

datos



Radiation

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

--,----

ucles

atos con R

Visualización de datos

Gráficos de densidad

Introducción a R

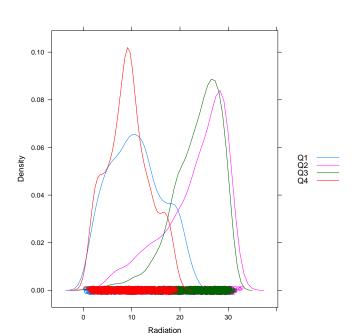
Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

Bucles

atos con R

Visualización de datos



Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

,

пасласо

oucles

atos con R

Visualización de datos

R es software libre

Objetos en R

Indexado

Bucles

Datos con R

Visualización de datos Catálogo Ajustes

Añadimos regresión lineal

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

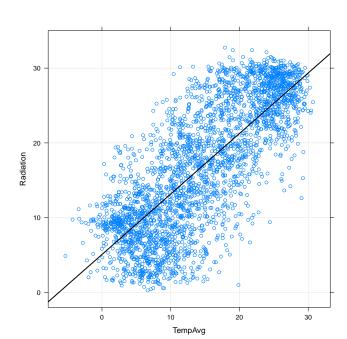
objetos en K

Hackaa

Bucles

atos con R

Visualización de datos



Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

--,------

oucles

atos con R

Visualización de datos

Catálogo

Ajustes

Añadimos ajuste local

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

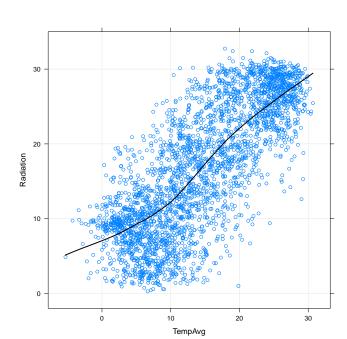
objetos en

ndexado

Bucles

atos con R

Visualización de datos



Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

Objetos Ci

шелацо

oucles

atos con R

Visualización de

Catálogo

Ajustes

Colores y tamaños

```
xyplot(Radiation ~ TempAvg,
    type=c('p', 'r'),
    cex=2, col='blue',
    alpha=.5, pch=19,
    lwd=3, col.line='black',
    data=aranjuez)
```

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

K es software fibre

Objetos err.

HICCARC

Bucles

Datos con R

Visualización de



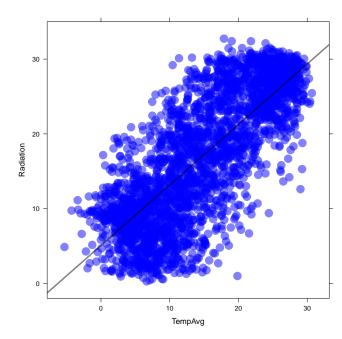
Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

ucies

atos con R

visualizacion de datos



Colores con grupos: par.settings y simpleTheme

▶ Primero definimos el tema con simpleTheme

Introducción a R

Oscar Perpiñán
Lamigueiro
http://
oscarperpinan.
github.io

R es software libre

Objetos err.

iuexauo

ucles

atos con R

Visualización de

datos

Aiustes

Colores con grupos: par.settings y simpleTheme

Aplicamos el resultado en par.settings

```
xyplot(Radiation ~ TempAvg,
    groups=quarter,
    par.settings=myTheme,
    auto.key=list(space='right'),
    data=aranjuez)
```

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

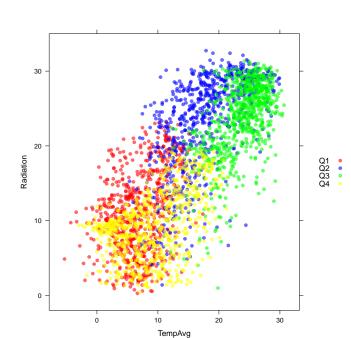
objetos err.

ndexado

Bucles

tos con R

Visualización de datos



Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

objetos en .

ucles

ntos con R

Visualización de datos

Catálogo

Ajustes

Colores: brewer.pal

```
library(RColorBrewer)
myPal <- brewer.pal(n = 4, 'Dark2')</pre>
myTheme <- simpleTheme(col = myPal,
                   pch=19, alpha=.6)
```

ColorBrewer: http://colorbrewer2.org/

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

Aiustes

Asignamos paleta con par.settings

```
xyplot(Radiation ~ TempAvg,
    groups=quarter,
    par.settings=myTheme,
    auto.key=list(space='right'),
    data=aranjuez)
```

Introducción a R

Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

Objetos err.

Indexado

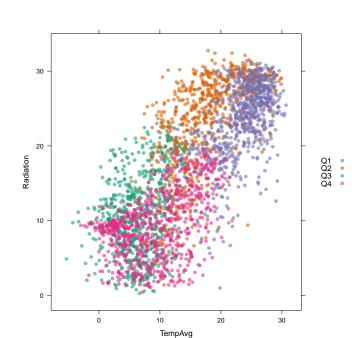
Bucles

atos con R

Visualización de datos

Catálogo

Aiustes



Oscar Perpiñán Lamigueiro http:// oscarperpinan. github.io

R es software libre

Objetoo err 1

ucies

itos con R

datos