sessionInfo()

R version 4.2.2 (2022-10-31 ucrt)

Platform: x86\_64-w64-mingw32/x64 (64-bit)

Running under: Windows 10 x64 (build 19044)

Matrix products: default

locale:

[1] LC\_COLLATE=Spanish\_Argentina.utf8 LC\_CTYPE=Spanish\_Argentina.utf8

[3] LC\_MONETARY=Spanish\_Argentina.utf8 LC\_NUMERIC=C

[5] LC\_TIME=Spanish\_Argentina.utf8

attached base packages:

[1] stats graphics grDevices utils datasets methods base

loaded via a namespace (and not attached):

[1] Rcpp\_1.0.9 compiler\_4.2.2 later\_1.3.0

[4] urlchecker\_1.0.1 prettyunits\_1.1.1 profvis\_0.3.7

[7] remotes\_2.4.2 tools\_4.2.2 digest\_0.6.30

[10] pkgbuild\_1.3.1 pkgload\_1.3.2 memoise\_2.0.1

[13] lifecycle\_1.0.3 rlang\_1.0.6 shiny\_1.7.3

[16] cli\_3.4.1 rstudioapi\_0.14 fastmap\_1.1.0

[19] withr\_2.5.0 stringr\_1.4.1 fs\_1.5.2

[22] htmlwidgets\_1.5.4 devtools\_2.4.5 rprojroot\_2.0.3

[25] glue\_1.6.2 R6\_2.5.1 processx\_3.8.0

[28] sessioninfo\_1.2.2 callr\_3.7.3 purrr\_0.3.5

[31] magrittr\_2.0.3 ps\_1.7.2 promises\_1.2.0.1

[34] ellipsis\_0.3.2 htmltools\_0.5.3 usethis\_2.1.6

[37] mime\_0.12 xtable\_1.8-4 httpuv\_1.6.6

[40] stringi\_1.7.8 miniUI\_0.1.1.1 cachem\_1.0.6

[43] crayon\_1.5.2

> ?merge

> help(package=ratter)

> help(package=dplyr)

> args(sample)

function (x, size, replace = FALSE, prob = NULL)

NULL

> getwd()

[1] "C:/Users/LENOVO/Documents"

> ls()

[1] "application\_train" "applicationTest" "bureau"

> rm(list=ls())

> ls()

character(0)

> **install.packages("knitr")**

Installing package into ‘C:/Users/LENOVO/AppData/Local/R/win-library/4.2’

(as ‘lib’ is unspecified)

There is a binary version available but the source version is

later:

binary source needs\_compilation

knitr 1.40 1.41 FALSE

installing the source package ‘knitr’

trying URL 'https://cran.rstudio.com/src/contrib/knitr\_1.41.tar.gz'

Content type 'application/x-gzip' length 890610 bytes (869 KB)

downloaded 869 KB

\* installing \*source\* package 'knitr' ...

\*\* package 'knitr' successfully unpacked and MD5 sums checked

\*\* using staged installation

\*\* R

\*\* demo

\*\* inst

\*\* byte-compile and prepare package for lazy loading

\*\* help

\*\*\* installing help indices

\*\* building package indices

\*\* installing vignettes

\*\* testing if installed package can be loaded from temporary location

\*\* testing if installed package can be loaded from final location

\*\* testing if installed package keeps a record of temporary installation path

\* DONE (knitr)

The downloaded source packages are in

‘C:\Users\LENOVO\AppData\Local\Temp\Rtmp0qHRu1\downloaded\_packages’

**> library(knitr)**

> x <-1

> print(x)

[1] 1

> x

[1] 1

>

> "hola"

[1] "hola"

> 1:20

[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

> is.na(1)

[1] FALSE

> c(1:20,NA)

[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 NA

> c

function (...) .Primitive("c")

> is.na(c)

[1] FALSE

Warning message:

In is.na(c) : is.na() applied to non-(list or vector) of type 'builtin'

> c(1,2,3,na)

Error: object 'na' not found

> c(1,2,3,NA)

[1] 1 2 3 NA

> c

function (...) .Primitive("c")

> is.na(c)

[1] FALSE

Warning message:

In is.na(c) : is.na() applied to non-(list or vector) of type 'builtin'

> missing <- is.na(c)

Warning message:

In is.na(c) : is.na() applied to non-(list or vector) of type 'builtin'

> c[!missing]

Error in c[!missing] : object of type 'builtin' is not subsettable

> x <- c(1,2,3,NA)

> missing <- is.na(x)

> x[!missing]

[1] 1 2 3

> missingEnx <- complete.cases(x)

> x[missingEnx]

[1] 1 2 3

> **library(gplots)**

Attaching package: ‘gplots’

The following object is masked from ‘package:stats’:

lowess

> miMatriz <- matrix(c(1,2,3,4), byrow=T, nrow=2)

> textplot(miMatriz)

> miDataFrame <- data.frame(edades=vEdades, nombres=vNombres)

Error in data.frame(edades = vEdades, nombres = vNombres) :

object 'vEdades' not found

> miDataFrame <- data.frame(edades= vEdades, nombres= vNombres)

Error in data.frame(edades = vEdades, nombres = vNombres) :

object 'vEdades' not found

> library(data.table)

data.table 1.14.4 using 4 threads (see ?getDTthreads). Latest news: r-datatable.com

> miDt1 <- data.table( nombre=c("Juan","Maria","Pablo","Ana"), edades=c(25,30,22,29))

> miDt2 <- data.table( nombre=c("Juan","Maria","Pedro","Ana", "Juanjo"), alturas=c(1.65,1.70,1.68,1.85,1.77))

> setkey(miDt1, nombre)

> setkey(miDt2, nombre)

> merge(miDt1,miDt2)

nombre edades alturas

1: Ana 29 1.85

2: Juan 25 1.65

3: Maria 30 1.70

> alturas <- c("alto", "bajo", "medio", "alto", "alto")

> fAlturas <- as.factor(alturas)

> fAlturas

[1] alto bajo medio alto alto

Levels: alto bajo medio

> relevel(fAlturas, ref="bajo") # "bajo"" será el nivel de referencia en los modelos

[1] alto bajo medio alto alto

Levels: bajo alto medio

> x <- factor(c("alto", "bajo", "medio", "alto", "alto"), levels=c("bajo", "medio", "alto"))

> x

[1] alto bajo medio alto alto

Levels: bajo medio alto

> x <- 0:5

> class(x)

[1] "integer"

> as.numeric(x)

[1] 0 1 2 3 4 5

> as.logical(x)

[1] FALSE TRUE TRUE TRUE TRUE TRUE

> as.character(x)

[1] "0" "1" "2" "3" "4" "5"

> as.complex(x)

[1] 0+0i 1+0i 2+0i 3+0i 4+0i 5+0i

> m <- matrix(nrow = 2, ncol = 3)

> m

[,1] [,2] [,3]

[1,] NA NA NA

[2,] NA NA NA

> dim(m)

[1] 2 3

> attributes(m)

$dim

[1] 2 3

> dim

function (x) .Primitive("dim")

> x <- 1:20

> y <- 10:14

> cbind(x,y)

x y

[1,] 1 10

[2,] 2 11

[3,] 3 12

[4,] 4 13

[5,] 5 14

[6,] 6 10

[7,] 7 11

[8,] 8 12

[9,] 9 13

[10,] 10 14

[11,] 11 10

[12,] 12 11

[13,] 13 12

[14,] 14 13

[15,] 15 14

[16,] 16 10

[17,] 17 11

[18,] 18 12

[19,] 19 13

[20,] 20 14

> rbind(x,y)

[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]

x 1 2 3 4 5 6 7 8 9 10 11 12 13

y 10 11 12 13 14 10 11 12 13 14 10 11 12

[,14] [,15] [,16] [,17] [,18] [,19] [,20]

x 14 15 16 17 18 19 20

y 13 14 10 11 12 13 14

> x <- 1:3

> names(x)

NULL

> names(x) <- c("nombre", "edad", "altura")

> x

nombre edad altura

1 2 3

> names(x)

[1] "nombre" "edad" "altura"

> l <- list(nombre = "Pepe", edad = 35, altura = 1.85)

> 1

[1] 1

> $edad [1] 35

Error: unexpected '$' in "$"

> m <- matrix(1:4, nrow = 2, ncol = 2)

> dimnames(m) <- list(c("a", "b"), c("c", "d"))

> m

c d

a 1 3

b 2 4

> valturas = c(180.5, 175.0, 192.4, 180.8)

> vnombres = c("Pepe", "Jaime", "David", "Tomas")

> miDataFrame = data.frame(heights = valturas, nombres = vnombres)

> valturas[1]

[1] 180.5

> valturas[3]

[1] 192.4

> valturas[c(1, 2, 4)]

[1] 180.5 175.0 180.8

> miDataFrame[1, 1:2]

heights nombres

1 180.5 Pepe

> miDataFrame$nombres

[1] "Pepe" "Jaime" "David" "Tomas"

> x <- list(enteros = 1:4, num = 0.6, nombre = "Pepe")

> name <- "enteros"

> x[[name]]

[1] 1 2 3 4

> x$numbers ## El elemento "numbers" no existe

NULL

> x$num

[1] 0.6

> x <- c("a", "b", "c", "c", "d", "a")

> mayoresA <- x > "a"

> mayoresA

[1] FALSE TRUE TRUE TRUE TRUE FALSE

> x[mayoresA]

[1] "b" "c" "c" "d"

> x <- matrix(1:6, 2, 3)

> x[1, 2]

[1] 3

> x[2, 1]

[1] 2

> x[1, ]

[1] 1 3 5

> x[, 2]

[1] 3 4

> x <- list(enteros = 1:4, floats = 0.6)

> x[1]

$enteros

[1] 1 2 3 4

> x[2]

$floats

[1] 0.6

> x[[1]]

[1] 1 2 3 4

> x$floats

[1] 0.6

> x[["floats"]]

[1] 0.6

> x <- list(a = list(10, 12, 14), b = c(3.14, 2.81))

> x[[c(1, 3)]]

[1] 14

> x[[c(2, 2)]]

[1] 2.81

> x[[1]][[3]]

[1] 14

> x[[c(2, 1)]]

[1] 3.14

> alturas = rnorm(10,mean=165,sd=3)

> alturas

[1] 160.2162 165.6149 163.9647 165.7578 161.1180 162.1225 168.2573

[8] 166.2113 166.7595 170.4457

> lanzaMonedas = rbinom(10,size=10,prob=0.5)

> lanzaMonedas

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

> rnorm(10,mean=165,sd=3)

[1] 160.8694 161.5474 162.8825 161.8378 163.0628 164.4439 161.3963

[8] 171.1109 165.3233 164.7477

> rnorm(10,mean=165,sd=3)

[1] 166.4869 165.1122 164.6037 169.4304 164.3489 161.1492 166.1570

[8] 163.9455 163.4346 161.7956

> set.seed(10000)

> rnorm(10,mean=165,sd=3)

[1] 164.5612 163.7027 164.2982 167.0530 165.2161 161.5100 166.1805

[8] 161.8225 166.4687 170.8105

> rnorm(10,mean=165,sd=3)

[1] 166.9099 163.1052 164.0768 163.4775 164.5692 159.0439 166.5027

[8] 165.5233 164.0010 162.9210

> set.seed(10000)

> rnorm(10,mean=165,sd=3)

[1] 164.5612 163.7027 164.2982 167.0530 165.2161 161.5100 166.1805

[8] 161.8225 166.4687 170.8105

> edad <- rnorm(10,mean=6.5,sd=3)

> edad

[1] 8.4099390 4.6051920 5.5768219 4.9775433 6.0692379 0.5438885 8.0027309

[8] 7.0232654 5.5010006 4.4209824

> sample(edad,size=6,replace=TRUE)

[1] 8.409939 4.977543 4.605192 7.023265 7.023265 5.576822

> probs <- c(0,0.3,0.1,0.3,0,0,0.2,0.1,0,0)

> sum(probs)

[1] 1

> sample(edad,size=6,replace=TRUE, prob=probs)

[1] 7.023265 5.576822 4.977543 5.576822 4.605192 8.002731

> sample(alturas,size=6,replace=FALSE)

[1] 160.2162 168.2573 163.9647 165.6149 162.1225 170.4457

> set.seed(1234)

> train\_row <- sample(1:nrow(dataset), nrow(dataset)\*0.7)

Error in nrow(dataset) : object 'dataset' not found

> search()

[1] ".GlobalEnv" "package:data.table" "package:gplots"

[4] "package:knitr" "tools:rstudio" "package:stats"

[7] "package:graphics" "package:grDevices" "package:utils"

[10] "package:datasets" "package:methods" "Autoloads"

[13] "package:base"

> train\_row <- sample(1:nrow(datasets), nrow(datasets)\*0.7)

Error in nrow(datasets) : object 'datasets' not found

> setwd("~/")

> install.packages("downloader")

Installing package into ‘C:/Users/LENOVO/AppData/Local/R/win-library/4.2’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/downloader\_0.4.zip'

Content type 'application/zip' length 25049 bytes (24 KB)

downloaded 24 KB

package ‘downloader’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\LENOVO\AppData\Local\Temp\Rtmp0qHRu1\downloaded\_packages

> library(downloader)

> download(fileUrl,"./datos/student.zip", mode ="wb")

Error in is.factor(x) : object 'fileUrl' not found

> download("../datos/student.zip", mode ="wb")

Error in download.file(url, ...) :

argument "destfile" is missing, with no default

> studentMat <- read.table("C:/Users/LENOVO/Downloads/student-mat.csv",

+ row.names=NULL, sep=";", header=TRUE)

> kable(head(studentMat[,1:5]))

|school |sex | age|address |famsize |

|:------|:---|---:|:-------|:-------|

|GP |F | 18|U |GT3 |

|GP |F | 17|U |GT3 |

|GP |F | 15|U |LE3 |

|GP |F | 15|U |GT3 |

|GP |F | 16|U |GT3 |

|GP |M | 16|U |LE3 |

> con <- file("C:/Users/LENOVO/Downloads/student-por.csv","r")

> studentPor <- read.csv2(con)

> close(con)

> head(studentPor[,1:5])

school sex age address famsize

1 GP F 18 U GT3

2 GP F 17 U GT3

3 GP F 15 U LE3

4 GP F 15 U GT3

5 GP F 16 U GT3

6 GP M 16 U LE3

**> library(downloader)**

> studentMat <- read.table("C:/Users/LENOVO/Downloads/student-mat.csv",

+ row.names=NULL, sep=";", header=TRUE)

> kable(head(studentMat[,1:5]))

Error in kable(head(studentMat[, 1:5])) : could not find function "kable"

> install.packages("knitr")

Installing package into ‘C:/Users/LENOVO/AppData/Local/R/win-library/4.2’

(as ‘lib’ is unspecified)

There is a binary version available but the source version is later:

binary source needs\_compilation

knitr 1.40 1.41 FALSE

installing the source package ‘knitr’

trying URL 'https://cran.rstudio.com/src/contrib/knitr\_1.41.tar.gz'

Content type 'application/x-gzip' length 890610 bytes (869 KB)

downloaded 869 KB

\* installing \*source\* package 'knitr' ...

\*\* package 'knitr' successfully unpacked and MD5 sums checked

\*\* using staged installation

\*\* R

\*\* demo

\*\* inst

\*\* byte-compile and prepare package for lazy loading

\*\* help

\*\*\* installing help indices

\*\* building package indices

\*\* installing vignettes

\*\* testing if installed package can be loaded from temporary location

\*\* testing if installed package can be loaded from final location

\*\* testing if installed package keeps a record of temporary installation path

\* DONE (knitr)

The downloaded source packages are in

‘C:\Users\LENOVO\AppData\Local\Temp\RtmpO0r6NB\downloaded\_packages’

**> library(knitr)**

**> library(gplots)**

Attaching package: ‘gplots’

The following object is masked from ‘package:stats’:

lowess

> studentMat <- read.table("C:/Users/LENOVO/Downloads/student-mat.csv",

+ row.names=NULL, sep=";", header=TRUE)

> kable(head(studentMat[,1:5]))

|school |sex | age|address |famsize |

|:------|:---|---:|:-------|:-------|

|GP |F | 18|U |GT3 |

|GP |F | 17|U |GT3 |

|GP |F | 15|U |LE3 |

|GP |F | 15|U |GT3 |

|GP |F | 16|U |GT3 |

|GP |M | 16|U |LE3 |

> con <- file("C:/Users/LENOVO/Downloads/student-por.csv","r")

> studentPor <- read.csv2(con)

> close(con)

> head(studentPor[,1:5])

school sex age address famsize

1 GP F 18 U GT3

2 GP F 17 U GT3

3 GP F 15 U LE3

4 GP F 15 U GT3

5 GP F 16 U GT3

6 GP M 16 U LE3

> library(RJSONIO)

Error in library(RJSONIO) : there is no package called ‘RJSONIO’

> tail(studentPor)

school sex age address famsize Pstatus Medu Fedu Mjob Fjob reason

644 MS F 18 R GT3 T 4 4 teacher at\_home reputation

645 MS F 19 R GT3 T 2 3 services other course

646 MS F 18 U LE3 T 3 1 teacher services course

647 MS F 18 U GT3 T 1 1 other other course

648 MS M 17 U LE3 T 3 1 services services course

649 MS M 18 R LE3 T 3 2 services other course

guardian traveltime studytime failures schoolsup famsup paid activities nursery

644 mother 3 1 0 no yes no yes yes

645 mother 1 3 1 no no no yes no

646 mother 1 2 0 no yes no no yes

647 mother 2 2 0 no no no yes yes

648 mother 2 1 0 no no no no no

649 mother 3 1 0 no no no no no

higher internet romantic famrel freetime goout Dalc Walc health absences G1 G2

644 yes yes yes 4 4 3 2 2 5 4 7 9

645 yes yes no 5 4 2 1 2 5 4 10 11

646 yes yes no 4 3 4 1 1 1 4 15 15

647 yes no no 1 1 1 1 1 5 6 11 12

648 yes yes no 2 4 5 3 4 2 6 10 10

649 yes yes no 4 4 1 3 4 5 4 10 11

G3

644 10

645 10

646 16

647 9

648 10

649 11

> summary(studentPor)

school sex age address

Length:649 Length:649 Min. :15.00 Length:649

Class :character Class :character 1st Qu.:16.00 Class :character

Mode :character Mode :character Median :17.00 Mode :character

Mean :16.74

3rd Qu.:18.00

Max. :22.00

famsize Pstatus Medu Fedu

Length:649 Length:649 Min. :0.000 Min. :0.000

Class :character Class :character 1st Qu.:2.000 1st Qu.:1.000

Mode :character Mode :character Median :2.000 Median :2.000

Mean :2.515 Mean :2.307

3rd Qu.:4.000 3rd Qu.:3.000

Max. :4.000 Max. :4.000

Mjob Fjob reason guardian

Length:649 Length:649 Length:649 Length:649

Class :character Class :character Class :character Class :character

Mode :character Mode :character Mode :character Mode :character

traveltime studytime failures schoolsup

Min. :1.000 Min. :1.000 Min. :0.0000 Length:649

1st Qu.:1.000 1st Qu.:1.000 1st Qu.:0.0000 Class :character

Median :1.000 Median :2.000 Median :0.0000 Mode :character

Mean :1.569 Mean :1.931 Mean :0.2219

3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:0.0000

Max. :4.000 Max. :4.000 Max. :3.0000

famsup paid activities nursery

Length:649 Length:649 Length:649 Length:649

Class :character Class :character Class :character Class :character

Mode :character Mode :character Mode :character Mode :character

higher internet romantic famrel

Length:649 Length:649 Length:649 Min. :1.000

Class :character Class :character Class :character 1st Qu.:4.000

Mode :character Mode :character Mode :character Median :4.000

Mean :3.931

3rd Qu.:5.000

Max. :5.000

freetime goout Dalc Walc health

Min. :1.00 Min. :1.000 Min. :1.000 Min. :1.00 Min. :1.000

1st Qu.:3.00 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:1.00 1st Qu.:2.000

Median :3.00 Median :3.000 Median :1.000 Median :2.00 Median :4.000

Mean :3.18 Mean :3.185 Mean :1.502 Mean :2.28 Mean :3.536

3rd Qu.:4.00 3rd Qu.:4.000 3rd Qu.:2.000 3rd Qu.:3.00 3rd Qu.:5.000

Max. :5.00 Max. :5.000 Max. :5.000 Max. :5.00 Max. :5.000

absences G1 G2 G3

Min. : 0.000 Min. : 0.0 Min. : 0.00 Min. : 0.00

1st Qu.: 0.000 1st Qu.:10.0 1st Qu.:10.00 1st Qu.:10.00

Median : 2.000 Median :11.0 Median :11.00 Median :12.00

Mean : 3.659 Mean :11.4 Mean :11.57 Mean :11.91

3rd Qu.: 6.000 3rd Qu.:13.0 3rd Qu.:13.00 3rd Qu.:14.00

Max. :32.000 Max. :19.0 Max. :19.00 Max. :19.00

> any(studentMat$G3 == 20)

[1] TRUE

> all(studentMat$G3 > 0)

[1] FALSE

**Graficos**

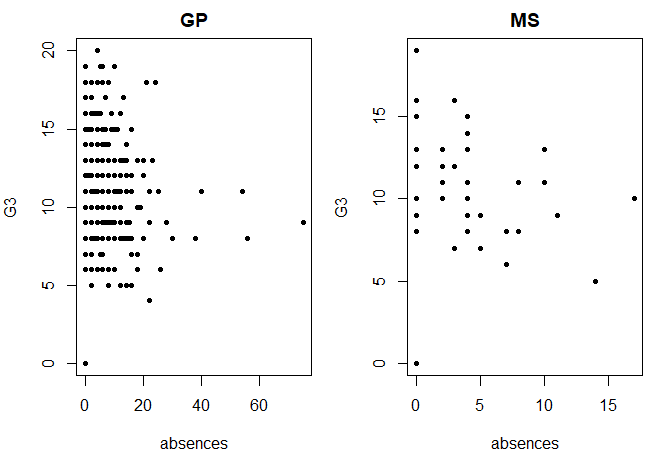
> par(mfrow = c(1, 2), mar = c(5, 4, 2, 1))

> with(subset(studentMat, school == "GP"),

+ plot(absences, G3, main = "GP", pch=20))

> with(subset(studentMat, school == "MS"),

+ plot(absences, G3, main = "MS", pch=20))



**Con lattice**

**library(lattice)**

**> library(gridExtra)**

> plot1 <- xyplot(G3 ~ absences,

+ data = studentMat[studentMat$school=="GP",],

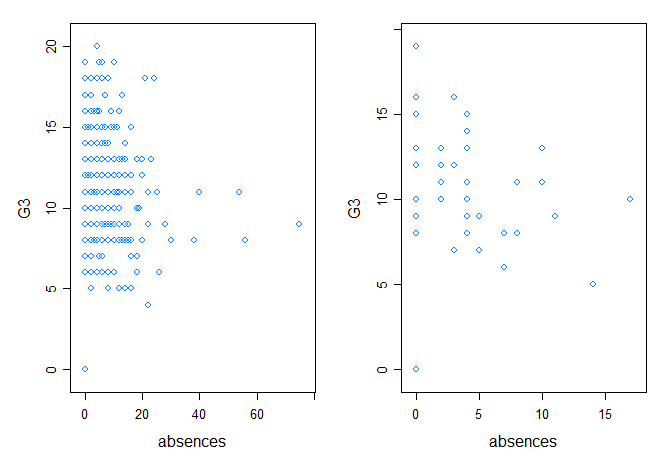
+ scales=list(relation="free"))

> plot2 <- xyplot(G3 ~ absences,

+ data = studentMat[studentMat$school=="MS",],

+ scales=list(relation="free"))

> grid.arrange(plot1, plot2, nrow=1)



El mismo ejemplo con **qplot** de **{ggplot2}**

**library(ggplot2)**

**> miGgplot1 <- qplot(absences, G3,**

**+ data = studentMat[studentMat$school=="GP",])**

**Warning message:**

**`qplot()` was deprecated in ggplot2 3.4.0.**

**This warning is displayed once every 8 hours.**

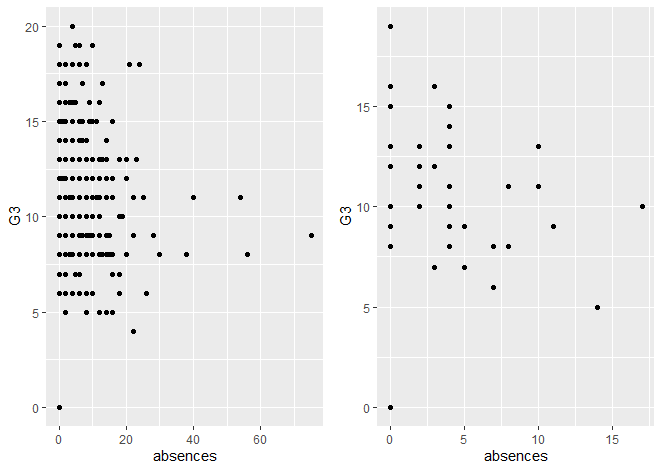
**Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was**

**generated.**

**> miGgplot2 <- qplot(absences, G3,**

**+ data = studentMat[studentMat$school=="MS",])**

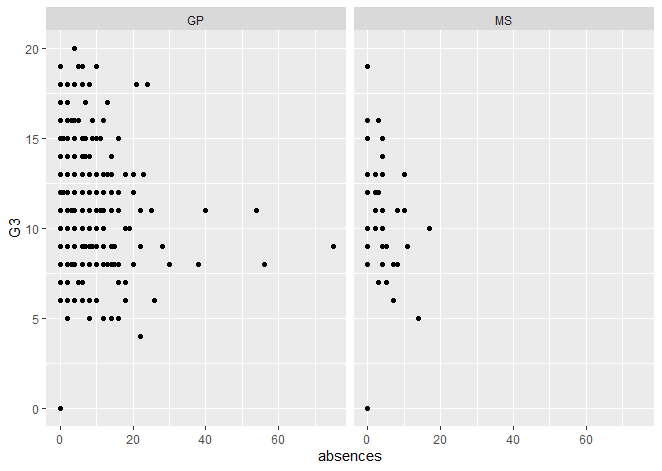
**> grid.arrange(miGgplot1, miGgplot2, nrow=1)**



**library(ggplot2)**

**> g1 <- ggplot(studentMat, aes(absences, G3))**

**> g1 + geom\_point() + facet\_grid(. ~ school)**



**> library(xtable)**

**> TableInputData <- xtable(summary(studentPor))**

**> print(TableInputData, type = "latex")**

**% latex table generated in R 4.2.2 by xtable 1.8-4 package**

**% Thu Nov 24 14:15:29 2022**

**\begin{table}[ht]**

**\centering**

**\begin{tabular}{rlllllllllllllllllllllllllllllllll}**

**\hline**

**& school & sex & age & address & famsize & Pstatus & Medu & Fedu & Mjob & Fjob & reason & guardian & traveltime & studytime & failures & schoolsup & famsup & paid & activities & nursery & higher & internet & romantic & famrel & freetime & goout & Dalc & Walc & health & absences & G1 & G2 & G3 \\**

**\hline**

**X & Length:649 & Length:649 & Min. :15.00 & Length:649 & Length:649 & Length:649 & Min. :0.000 & Min. :0.000 & Length:649 & Length:649 & Length:649 & Length:649 & Min. :1.000 & Min. :1.000 & Min. :0.0000 & Length:649 & Length:649 & Length:649 & Length:649 & Length:649 & Length:649 & Length:649 & Length:649 & Min. :1.000 & Min. :1.00 & Min. :1.000 & Min. :1.000 & Min. :1.00 & Min. :1.000 & Min. : 0.000 & Min. : 0.0 & Min. : 0.00 & Min. : 0.00 \\**

**X.1 & Class :character & Class :character & 1st Qu.:16.00 & Class :character & Class :character & Class :character & 1st Qu.:2.000 & 1st Qu.:1.000 & Class :character & Class :character & Class :character & Class :character & 1st Qu.:1.000 & 1st Qu.:1.000 & 1st Qu.:0.0000 & Class :character & Class :character & Class :character & Class :character & Class :character & Class :character & Class :character & Class :character & 1st Qu.:4.000 & 1st Qu.:3.00 & 1st Qu.:2.000 & 1st Qu.:1.000 & 1st Qu.:1.00 & 1st Qu.:2.000 & 1st Qu.: 0.000 & 1st Qu.:10.0 & 1st Qu.:10.00 & 1st Qu.:10.00 \\**

**X.2 & Mode :character & Mode :character & Median :17.00 & Mode :character & Mode :character & Mode :character & Median :2.000 & Median :2.000 & Mode :character & Mode :character & Mode :character & Mode :character & Median :1.000 & Median :2.000 & Median :0.0000 & Mode :character & Mode :character & Mode :character & Mode :character & Mode :character & Mode :character & Mode :character & Mode :character & Median :4.000 & Median :3.00 & Median :3.000 & Median :1.000 & Median :2.00 & Median :4.000 & Median : 2.000 & Median :11.0 & Median :11.00 & Median :12.00 \\**

**X.3 & & & Mean :16.74 & & & & Mean :2.515 & Mean :2.307 & & & & & Mean :1.569 & Mean :1.931 & Mean :0.2219 & & & & & & & & & Mean :3.931 & Mean :3.18 & Mean :3.185 & Mean :1.502 & Mean :2.28 & Mean :3.536 & Mean : 3.659 & Mean :11.4 & Mean :11.57 & Mean :11.91 \\**

**X.4 & & & 3rd Qu.:18.00 & & & & 3rd Qu.:4.000 & 3rd Qu.:3.000 & & & & & 3rd Qu.:2.000 & 3rd Qu.:2.000 & 3rd Qu.:0.0000 & & & & & & & & & 3rd Qu.:5.000 & 3rd Qu.:4.00 & 3rd Qu.:4.000 & 3rd Qu.:2.000 & 3rd Qu.:3.00 & 3rd Qu.:5.000 & 3rd Qu.: 6.000 & 3rd Qu.:13.0 & 3rd Qu.:13.00 & 3rd Qu.:14.00 \\**

**X.5 & & & Max. :22.00 & & & & Max. :4.000 & Max. :4.000 & & & & & Max. :4.000 & Max. :4.000 & Max. :3.0000 & & & & & & & & & Max. :5.000 & Max. :5.00 & Max. :5.000 & Max. :5.000 & Max. :5.00 & Max. :5.000 & Max. :32.000 & Max. :19.0 & Max. :19.00 & Max. :19.00 \\**

**\hline**

**\end{tabular}**

**\end{table}**

**> kable(summary(studentPor))**

**| | school | sex | age | address | famsize | Pstatus | Medu | Fedu | Mjob | Fjob | reason | guardian | traveltime | studytime | failures | schoolsup | famsup | paid | activities | nursery | higher | internet | romantic | famrel | freetime | goout | Dalc | Walc | health | absences | G1 | G2 | G3 |**

**|:--|:----------------|:----------------|:-------------|:----------------|:----------------|:----------------|:-------------|:-------------|:----------------|:----------------|:----------------|:----------------|:-------------|:-------------|:--------------|:----------------|:----------------|:----------------|:----------------|:----------------|:----------------|:----------------|:----------------|:-------------|:------------|:-------------|:-------------|:------------|:-------------|:--------------|:------------|:-------------|:-------------|**

**| |Length:649 |Length:649 |Min. :15.00 |Length:649 |Length:649 |Length:649 |Min. :0.000 |Min. :0.000 |Length:649 |Length:649 |Length:649 |Length:649 |Min. :1.000 |Min. :1.000 |Min. :0.0000 |Length:649 |Length:649 |Length:649 |Length:649 |Length:649 |Length:649 |Length:649 |Length:649 |Min. :1.000 |Min. :1.00 |Min. :1.000 |Min. :1.000 |Min. :1.00 |Min. :1.000 |Min. : 0.000 |Min. : 0.0 |Min. : 0.00 |Min. : 0.00 |**

**| |Class :character |Class :character |1st Qu.:16.00 |Class :character |Class :character |Class :character |1st Qu.:2.000 |1st Qu.:1.000 |Class :character |Class :character |Class :character |Class :character |1st Qu.:1.000 |1st Qu.:1.000 |1st Qu.:0.0000 |Class :character |Class :character |Class :character |Class :character |Class :character |Class :character |Class :character |Class :character |1st Qu.:4.000 |1st Qu.:3.00 |1st Qu.:2.000 |1st Qu.:1.000 |1st Qu.:1.00 |1st Qu.:2.000 |1st Qu.: 0.000 |1st Qu.:10.0 |1st Qu.:10.00 |1st Qu.:10.00 |**

**| |Mode :character |Mode :character |Median :17.00 |Mode :character |Mode :character |Mode :character |Median :2.000 |Median :2.000 |Mode :character |Mode :character |Mode :character |Mode :character |Median :1.000 |Median :2.000 |Median :0.0000 |Mode :character |Mode :character |Mode :character |Mode :character |Mode :character |Mode :character |Mode :character |Mode :character |Median :4.000 |Median :3.00 |Median :3.000 |Median :1.000 |Median :2.00 |Median :4.000 |Median : 2.000 |Median :11.0 |Median :11.00 |Median :12.00 |**

**| |NA |NA |Mean :16.74 |NA |NA |NA |Mean :2.515 |Mean :2.307 |NA |NA |NA |NA |Mean :1.569 |Mean :1.931 |Mean :0.2219 |NA |NA |NA |NA |NA |NA |NA |NA |Mean :3.931 |Mean :3.18 |Mean :3.185 |Mean :1.502 |Mean :2.28 |Mean :3.536 |Mean : 3.659 |Mean :11.4 |Mean :11.57 |Mean :11.91 |**

**| |NA |NA |3rd Qu.:18.00 |NA |NA |NA |3rd Qu.:4.000 |3rd Qu.:3.000 |NA |NA |NA |NA |3rd Qu.:2.000 |3rd Qu.:2.000 |3rd Qu.:0.0000 |NA |NA |NA |NA |NA |NA |NA |NA |3rd Qu.:5.000 |3rd Qu.:4.00 |3rd Qu.:4.000 |3rd Qu.:2.000 |3rd Qu.:3.00 |3rd Qu.:5.000 |3rd Qu.: 6.000 |3rd Qu.:13.0 |3rd Qu.:13.00 |3rd Qu.:14.00 |**

**| |NA |NA |Max. :22.00 |NA |NA |NA |Max. :4.000 |Max. :4.000 |NA |NA |NA |NA |Max. :4.000 |Max. :4.000 |Max. :3.0000 |NA |NA |NA |NA |NA |NA |NA |NA |Max. :5.000 |Max. :5.00 |Max. :5.000 |Max. :5.000 |Max. :5.00 |Max. :5.000 |Max. :32.000 |Max. :19.0 |Max. :19.00 |Max. :19.00 |**

**> names(studentMat)**

**[1] "school" "sex" "age" "address" "famsize"**

**[6] "Pstatus" "Medu" "Fedu" "Mjob" "Fjob"**

**[11] "reason" "guardian" "traveltime" "studytime" "failures"**

**[16] "schoolsup" "famsup" "paid" "activities" "nursery"**

**[21] "higher" "internet" "romantic" "famrel" "freetime"**

**[26] "goout" "Dalc" "Walc" "health" "absences"**

**[31] "G1" "G2" "G3"**

**> table(studentMat$internet)**

**no yes**

**66 329**

**> any(studentMat$G3 == 20)**

**[1] TRUE**

**> all(studentMat$G3 > 0)**

**[1] FALSE**

**> studentMatPor <- merge(studentMat,studentPor,**

**+ by=c("school","sex","age",**

**+ "address","famsize",**

**+ "Pstatus","Medu",**

**+ "Fedu","Mjob",**

**+ "Fjob","reason",**

**+ "nursery","internet"),**

**+ all=FALSE,**

**+ suffixes=c("mat","por"))**

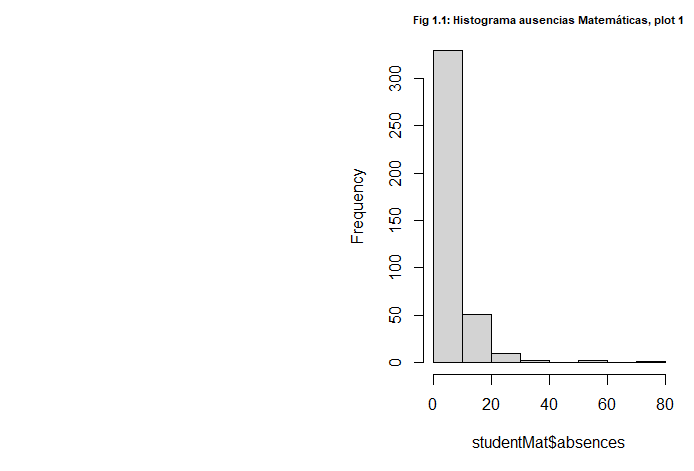
**> dim(studentMatPor)**

**[1] 382 53**

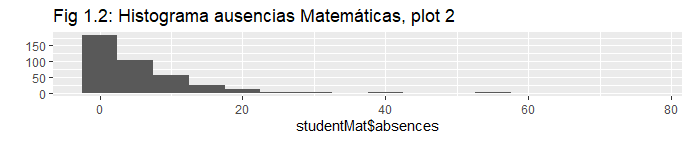
**Histograma para ver la distribución de las ausencias de los alumnos de matemáticos. (3 graficos)**

**> library(ggplot2)**

**> hist(studentMat$absences, main = "Fig 1.1: Histograma ausencias Matemáticas, plot 1", cex.main=0.7)**



**qplot(studentMat$absences, binwidth = 5, main = "Fig 1.2: Histograma ausencias Matemáticas, plot 2")**

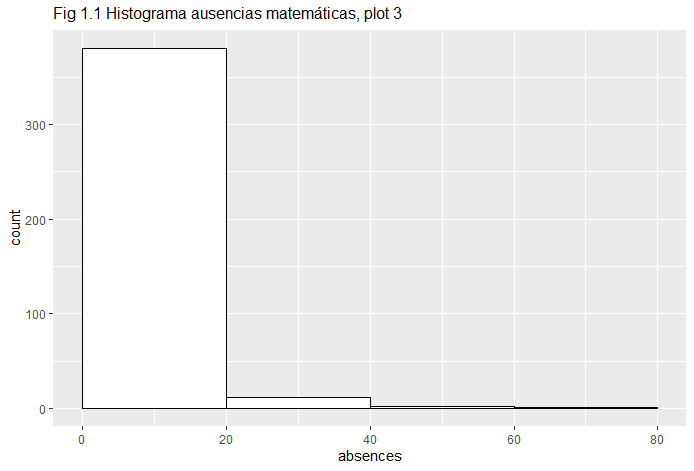


**ggplot(studentMat, aes(x=absences)) +**

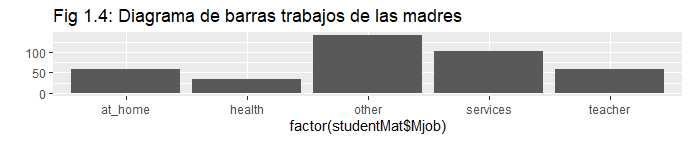
**+ geom\_histogram(binwidth = 20, fill = "white", colour = "black", origin = 0) +**

**+ ggtitle ("Fig 1.1 Histograma ausencias matemáticas, plot 3") +**

**+ theme(plot.title = element\_text(vjust = +1.5, size = 12))**



qplot(factor(studentMat$Mjob), main = "Fig 1.4: Diagrama de barras trabajos de las madres")



# Clasificamos en binario la variable target: Suponemos que "fail" es una nota # final de 0 a 9. Y pass >9.

# Para la variable tipo factor Mjob, podemos ver su distribución en relación al # target, con diagramas de barras.

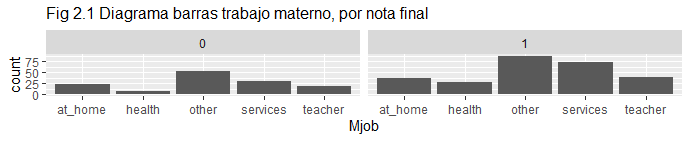
studentMat$pass <- ifelse(studentMat$G3>9, 1, 0)

> ggplot(studentMat, aes(Mjob)) + geom\_bar() +

+ facet\_wrap(~ pass) +

+ ggtitle ("Fig 2.1 Diagrama barras trabajo materno, por nota final") +

+ theme(plot.title = element\_text(vjust = +1.5, size = 12))



*# Para las variables continuas, podemos ver su distribución en relación al target, con histogramas.*

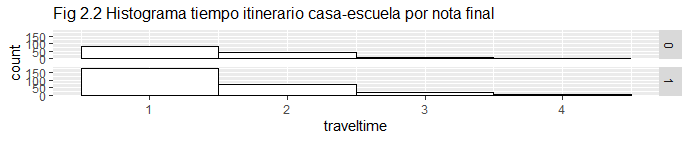
ggplot(studentMat, aes(x = traveltime)) + geom\_histogram(binwidth = 1, fill =

+ "white", colour = "black") +

+ facet\_grid(pass ~ .) +

+ ggtitle ("Fig 2.2 Histograma tiempo itinerario casa-escuela por nota final")+

+ theme(plot.title=element\_text(vjust = +1.5, size = 12))



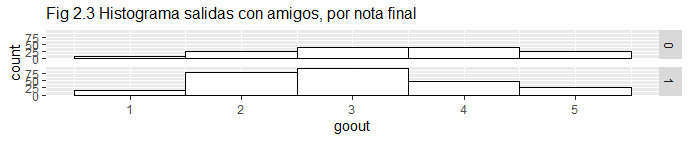
ggplot(studentMat, aes(x = goout)) + geom\_histogram(binwidth = 1, fill =

+ "white", colour = "black") +

+ facet\_grid(pass ~ .) +

+ ggtitle ("Fig 2.3 Histograma salidas con amigos, por nota final") +

+ theme(plot.title=element\_text(vjust = +1.5, size = 12))



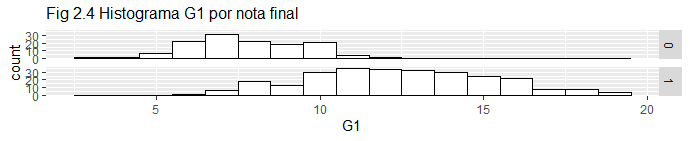
ggplot(studentMat, aes(x = G1)) + geom\_histogram(binwidth = 1, fill = "white",

+ colour = "black") +

+ facet\_grid(pass ~ .) +

+ ggtitle ("Fig 2.4 Histograma G1 por nota final") +

+ theme(plot.title=element\_text(vjust = +1.5, size = 12))



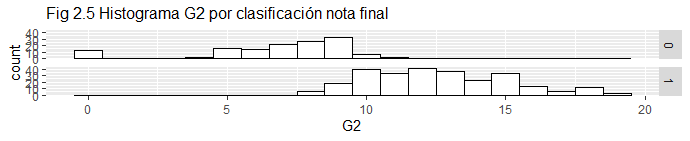
ggplot(studentMat, aes(x = G2)) + geom\_histogram(binwidth = 1, fill = "white",

+ colour = "black") +

+ facet\_grid(pass ~ .) +

+ ggtitle ("Fig 2.5 Histograma G2 por clasificación nota final") +

+ theme(plot.title=element\_text(vjust = +1.5, size = 12))



También podemos ver la relación de las variables continuas con el target, utilizando box plot

library(grid)

> library(gridExtra)

> plot1 <- ggplot(studentMat, aes(factor(pass), traveltime, fill=factor(pass))) +

+ geom\_boxplot() +

+ scale\_colour\_discrete(name = "Type") +

+ scale\_fill\_discrete(name="Type", breaks=c("0", "1"),

+ labels=c("fail", "pass")) +

+ scale\_x\_discrete(breaks=c("0", "1"), labels=c("fail", "pass")) +

+ xlab("") +

+ ggtitle ("Fig 3.1. Tiempo casa-escuela, por nota final") +

+ theme(plot.title=element\_text(vjust = +2.5, size = 7),

+ axis.text.x=element\_blank(), axis.title.x=element\_blank())

> plot2 <- ggplot(studentMat, aes(factor(pass), G1, fill=factor(pass))) +

+ geom\_boxplot() +

+ scale\_colour\_discrete(name = "Type") +

+ scale\_fill\_discrete(name="Type", breaks=c("0", "1"),

+ labels=c("fail", "pass")) +

+ scale\_x\_discrete(breaks=c("0", "1"), labels=c("fail", "pass")) +

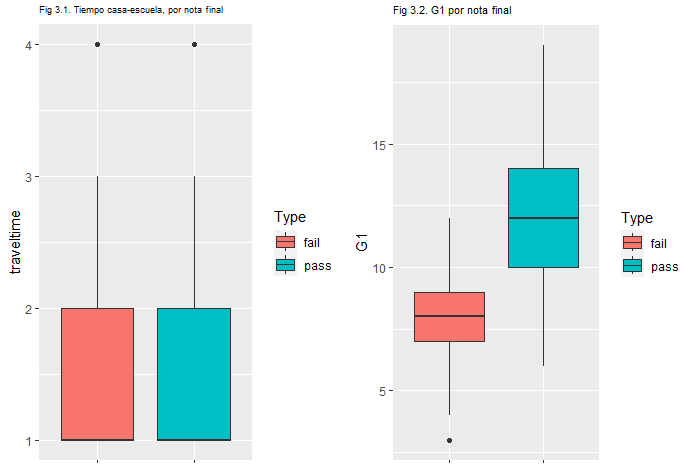
+ xlab("") +

+ ggtitle ("Fig 3.2. G1 por nota final") +

+ theme(plot.title=element\_text(vjust = +2.5, size = 8),

+ axis.text.x=element\_blank(), axis.title.x=element\_blank())

> grid.arrange(plot1, plot2, nrow=1, ncol=2)



Matriz de correlación

Para ver lo correladas que están unas variables con otras, y también con el target, podemos calcular y dibujar una matriz de correlación.

# Primero creamos variables dummies para las variables de tipo factor que queramos

# incluir en la matriz de correlación. Por ejemplo:

studentMat$GP <- **ifelse**(studentMat$school == "GP", 1, 0) studentMat$MS <- **ifelse**(studentMat$school == "MS", 1, 0)

# Por simplificación del ejemplo, sólo consideramos algunas variables para construir # la matriz de correlación

#plot.new()

# Generamos una paleta de colores más claros

col <- **colorRampPalette**(**c**("#BB4444", "#EE9988", "#FFFFFF", "#77AADD", "#4477AA"))

# Dibujamos la matriz de correlación con cuadrados de colores y etiquetas negras

studentMat$GP <- ifelse(studentMat$school == "GP", 1, 0)

> studentMat$MS <- ifelse(studentMat$school == "MS", 1, 0)

> matCor <- cor(studentMat[,

+ c("GP","MS","absences","goout","Dalc","Walc","traveltime","G1","G2","G3","pass")])

> library(corrplot)

> matCor[is.na(matCor)] <- 0

> plot.new()

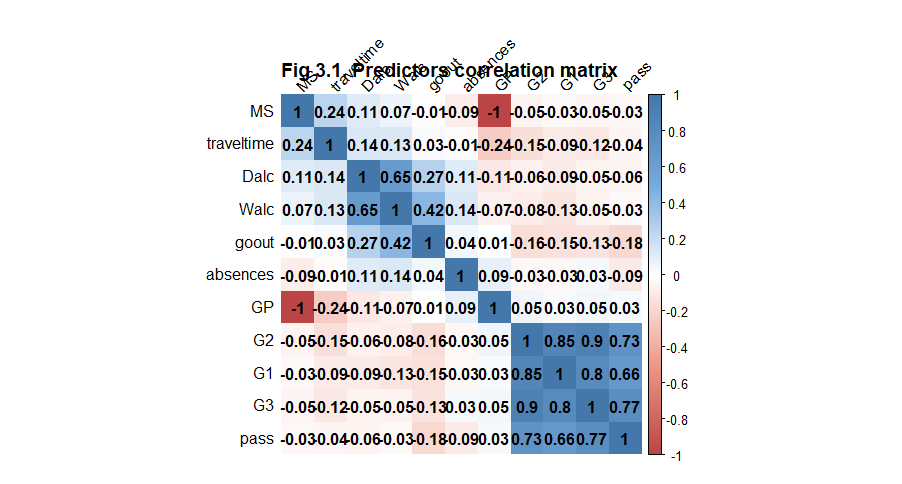
> col <- colorRampPalette(c("#BB4444", "#EE9988", "#FFFFFF", "#77AADD", "#4477AA"))

> corrplot(matCor, method = "shade", shade.col = NA, tl.col = "black",

+ tl.srt = 45, col = col(200), addCoef.col="black", order="AOE",

+ mar = c(1,0,2,0), line=-2, is.corr=FALSE,

+ main = "Fig 3.1. Predictors correlation matrix")



Conexion Base de datos Postgres (https://cran.r-project.org/web/packages/RPostgres/RPostgres.pdf)

## CONEXIÓN ENTRE POSTGRESL Y R

# 1.0 Instalación y carga de paquetes

install.packages("DBI", dependencies = TRUE)

install.packages("RPostgres", dependencies = TRUE)

install.packages("sf", dependencies = TRUE)

library(DBI)

library(RPostgres)

library(sf)

# 2.0 Parámetros de conexión a PostgreSQL

dvr <- RPostgres::Postgres()

db <- 'diana' ##Nombre de la BBDD

host\_db <- 'localhost'

db\_port <- '5432'

db\_user <- 'postgres' ##Tu usuario

db\_password <- 'postgres' ##Tu contraseña

# 3.0 Conexión

con <- dbConnect(dvr, dbname = db, host=host\_db, port=db\_port,

user=db\_user, password=db\_password)

# 4.0 Listado de tablas de la Base de Datos

dbListTables(con)

# 5.0 Lectura de una tabla

roads <- st\_read(con, layer = "roads")

print(roads)

# 6.0 Query --> Seleccion de todos los elementos de la tabla

res <- dbSendQuery(con, "SELECT road\_id, road\_name FROM roads where road\_name LIKE \'Jeff Rd\'")

dbFetch(res)

dbClearResult(res)

# 7.0 Query --> Borrado de un elemento de la tabla

res <- dbSendQuery(con, "DELETE FROM roads WHERE road\_id = 6")

dbClearResult(res)

# 8.0 Cerrar conexión

dbDisconnect(con)