

resultados rtesis

oscar lazo

2024-01-31

```
#leer Librerias
```

```
library(readxl)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(readr)
library(psych)
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'
## The following objects are masked from 'package:psych':
##
##   %+%, alpha
```

```
library(stargazer)
```

```
##
## Please cite as:
## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
library(lmtest)
```

```
## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
```

```
library(gridExtra)
```

```

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##      combine
library(PerformanceAnalytics)

## Loading required package: xts

##
## ##### Warning from 'xts' package #####
## #
## # The dplyr lag() function breaks how base R's lag() function is supposed to #
## # work, which breaks lag(my_xts). Calls to lag(my_xts) that you type or #
## # source() into this session won't work correctly. #
## #
## # Use stats::lag() to make sure you're not using dplyr::lag(), or you can add #
## # conflictRules('dplyr', exclude = 'lag') to your .Rprofile to stop #
## # dplyr from breaking base R's lag() function. #
## #
## # Code in packages is not affected. It's protected by R's namespace mechanism #
## # Set `options(xts.warn_dplyr_breaks_lag = FALSE)` to suppress this warning. #
## #
## #####
##
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':
##
##      first, last

##
## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':
##
##      legend
library(car)

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:psych':
##
##      logit

## The following object is masked from 'package:dplyr':
##
##      recode
library(corrplot)

## corrplot 0.92 loaded

```

```
library(ellipse)
```

```
##
## Attaching package: 'ellipse'
## The following object is masked from 'package:car':
##
##     ellipse
## The following object is masked from 'package:graphics':
##
##     pairs
```

```
library(knitr)
```

```
#cargar datos
```

```
data <- read.csv("data.csv", sep = ";", header = TRUE, stringsAsFactors = FALSE)
attach(data)
```

```
# Mostrar las primeras filas del dataframe para verificar que se haya cargado correctamente
head(data)
```

```
##      N.BASES      Nombre.la.base total.de.observaciones
## 1      1 natallyFile_130_C78700xV600      47220000
## 2      2 natallyFILE_128_C62600xV300      18780000
## 3      3 natallyFile_137_C67500xV800      54000000
## 4      4 natallyFile_144_C4700xV300      1410000
## 5      5 natallyFile_149_C10900xV300      3270000
## 6      6 natallyFile_159_C80700xV3      242100
##      filas.observaciones. variables total.de.valores.en.la.matriz.de...similaridad
## 1      78700      600      360000
## 2      62600      300      90000
## 3      67500      800      640000
## 4      4700      300      90000
## 5      10900      300      90000
## 6      80700      3      9
##      Rchic.total.similaridad Pchic.total.similaridad cant..valores.no.similares
## 1      179700      179101      599
## 2      44850      44551      299
## 3      319600      318801      799
## 4      44850      44551      299
## 5      44850      44551      299
## 6      3      3      0
##      porcentaje.de.valores.similares.entre.2.matrices porcentaje.de.no.similaridad
## 1      0.997      0.0030
## 2      0.993      0.0070
## 3      0.998      0.0020
## 4      0.993      0.0067
## 5      0.993      0.0067
## 6      1.000      0.0000
##      decimales.6pchic..y.7.rchic nodospchic nodosrchic porcentaje.nodos.iguales
## 1      7      38      38      1.00000
## 2      7      34      35      0.97143
## 3      7      35      36      0.97222
## 4      7      36      37      0.97297
## 5      7      36      36      1.00000
```

```
## 6              7              1              1              1.00000
## porcentaje.nodos.no.iguales dendogramas.r.chic dendrograma.pchic
## 1              0.00000              1              0
## 2              0.02857              1              0
## 3              0.02778              1              0
## 4              0.02703              1              0
## 5              0.00000              1              0
## 6              0.00000              1              1
## Rchic.total.similaridad.6. Pchic.total.similaridad.6.
## 1              179700              179700
## 2              44850              44850
## 3              319600              319600
## 4              44850              44850
## 5              44850              44850
## 6              3              3
## cant..valores.no.similares.1 porcentaje.de.valores.similares.6.
## 1              0              1
## 2              0              1
## 3              0              1
## 4              0              1
## 5              0              1
## 6              0              1
## porcentaje.de.no.similaridad.1
## 1              0
## 2              0
## 3              0
## 4              0
## 5              0
## 6              0
```

```
str(data)
```

```
## 'data.frame':  383 obs. of  23 variables:
## $ N.BASES : int  1 2 3 4 5 6 7 8 9 10 ...
## $ Nombre.la.base : chr  "nattallyFile_130_C78700xV600" "nattallyFILE
## $ total.de.observaciones : int  47220000 18780000 54000000 1410000 3270000
## $ filas.observaciones. : int  78700 62600 67500 4700 10900 80700 26300 3
## $ variables : int  600 300 800 300 300 3 500 900 300 38 ...
## $ total.de.valores.en.la.matriz.de...similaridad : int  360000 90000 640000 90000 90000 9 250000 8
## $ Rchic.total.similaridad : int  179700 44850 319600 44850 44850 3 124750 4
## $ Pchic.total.similaridad : int  179101 44551 318801 44551 44551 3 124251 4
## $ cant..valores.no.similares : int  599 299 799 299 299 0 499 899 299 0 ...
## $ porcentaje.de.valores.similares.entre.2.matrices: num  0.997 0.993 0.998 0.993 0.993 1 0.996 0.99
## $ porcentaje.de.no.similaridad : num  0.003 0.007 0.002 0.0067 0.0067 0 0.004 0.
## $ decimales.6pchic..y.7.rchic : int  7 7 7 7 7 7 7 7 7 7 ...
## $ nodospchic : int  38 34 35 36 36 1 36 37 37 10 ...
## $ nodosrchic : int  38 35 36 37 36 1 36 38 38 10 ...
## $ porcentaje.nodos.iguales : num  1 0.971 0.972 0.973 1 ...
## $ porcentaje.nodos.no.iguales : num  0 0.0286 0.0278 0.027 0 ...
## $ dendogramas.r.chic : int  1 1 1 1 1 1 1 1 1 1 ...
## $ dendrograma.pchic : int  0 0 0 0 0 1 0 0 0 1 ...
## $ Rchic.total.similaridad.6. : num  179700 44850 319600 44850 44850 ...
## $ Pchic.total.similaridad.6. : num  179700 44850 319600 44850 44850 ...
## $ cant..valores.no.similares.1 : int  0 0 0 0 0 0 0 0 0 0 ...
## $ porcentaje.de.valores.similares.6. : int  1 1 1 1 1 1 1 1 1 1 ...
```

```
## $ porcentaje.de.no.similaridad.1 : int 0 0 0 0 0 0 0 0 0 0 ...
total.de.observaciones<-as.integer(total.de.observaciones)
filas.observaciones.<-as.integer(filas.observaciones.)
Rchic.total.similaridad<-as.integer(Rchic.total.similaridad)
Pchic.total.similaridad<-as.integer(Pchic.total.similaridad)
Rchic.total.similaridad.6.<-as.integer(Rchic.total.similaridad.6.)

## Warning: NAs introduced by coercion to integer range
Pchic.total.similaridad.6.<-as.integer(Pchic.total.similaridad.6.)

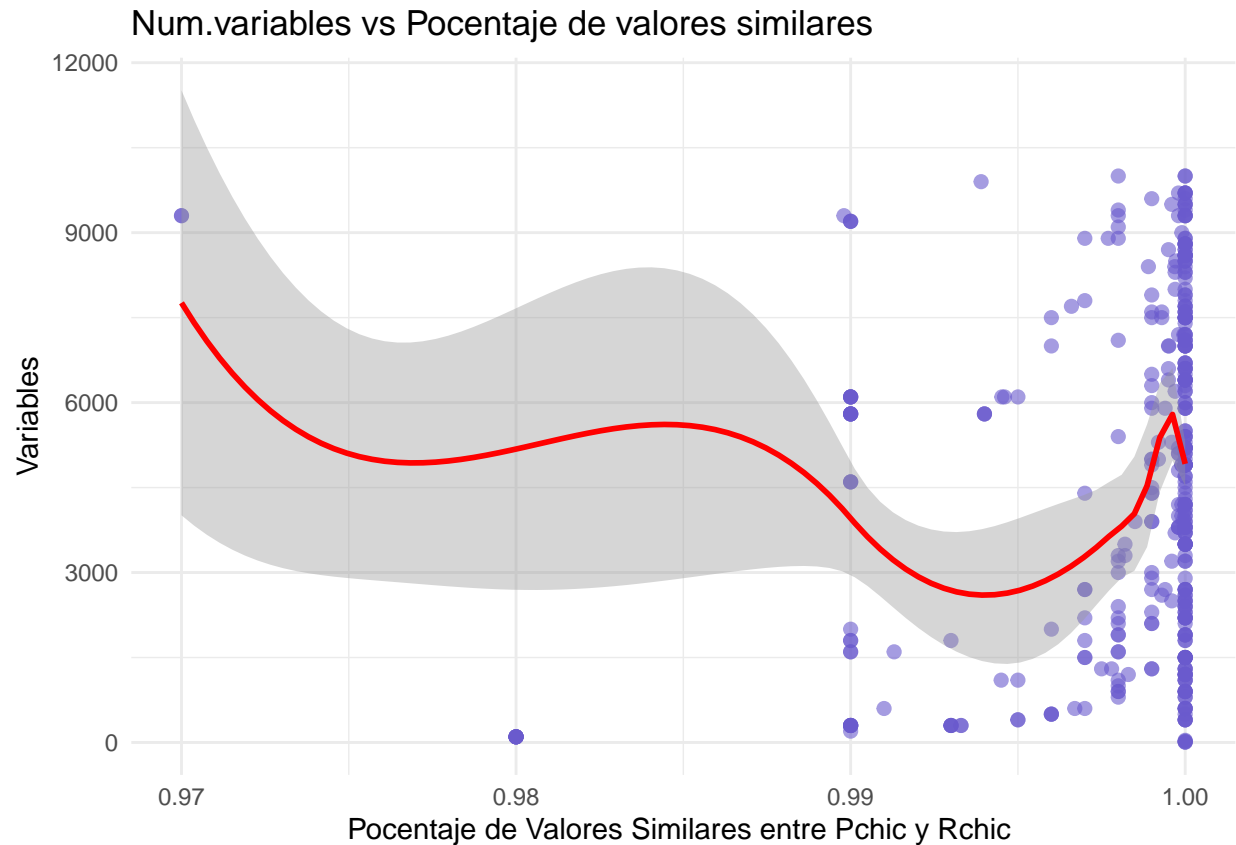
## Warning: NAs introduced by coercion to integer range
data<-data.frame(N.BASES,total.de.observaciones,filas.observaciones.,variables,
  total.de.valores.en.la.matriz.de...similaridad,Rchic.total.similaridad,Pchic.total.similaridad,
  cant..valores.no.similares,porcentaje.de.valores.similares.entre.2.matrices,
  porcentaje.de.no.similaridad,nodospchic,nodosrchic,porcentaje.nodos.iguales,
  Rchic.total.similaridad.6.,Pchic.total.similaridad.6.,porcentaje.de.valores.similares.6.)
```

describe

```
library(ggplot2)

ggplot(data = data, aes(x = porcentaje.de.valores.similares.entre.2.matrices, y = variables)) +
  geom_point(color = 'slateblue', size = 2, alpha = 0.6) +
  geom_smooth(color = 'red') +
  xlab('Pocentaje de Valores Similares entre Pchic y Rchic ') +
  ylab('Variables') +
  ggtitle('Num.variables vs Pocentaje de valores similares ') +
  theme_minimal()

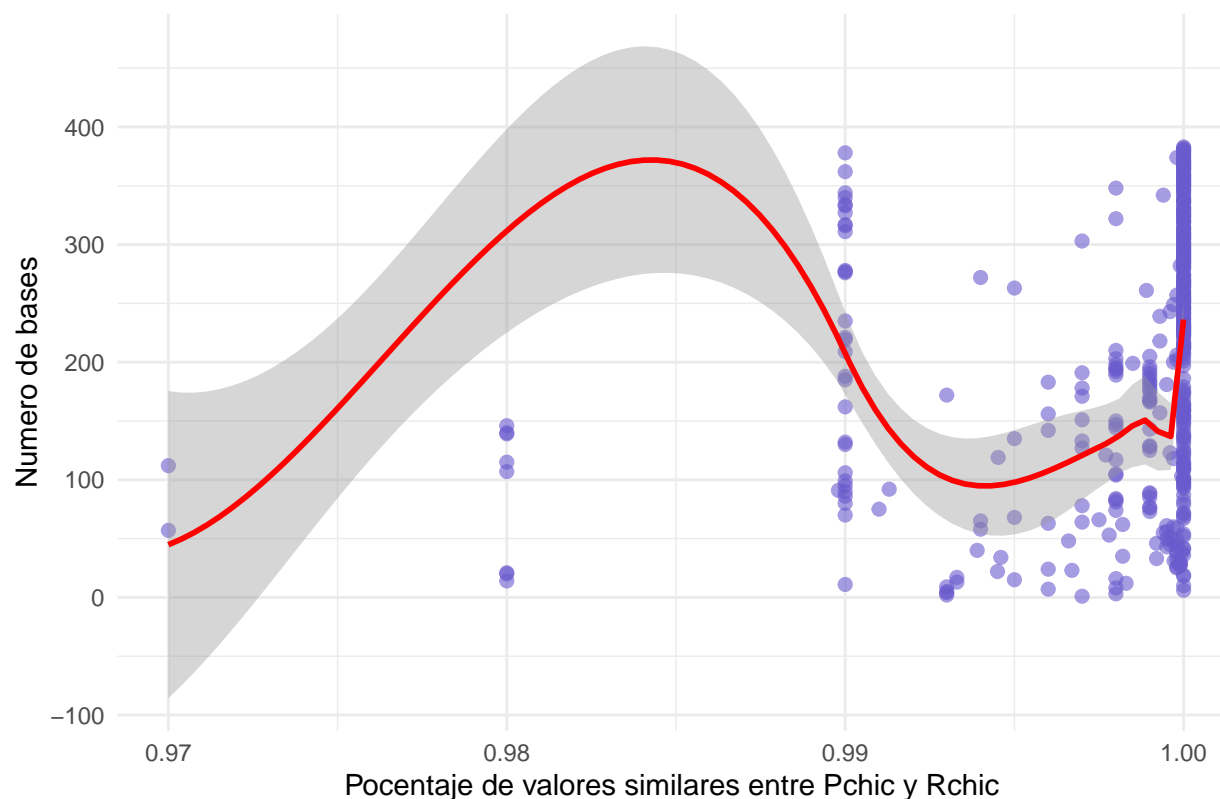
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



```
ggplot(data = data, aes(x = porcentaje.de.valores.similares.entre.2.matrices, y = N.BASES)) +
  geom_point(color = 'slateblue', size = 2, alpha = 0.6) +
  geom_smooth(color = 'red') +
  xlab('Pocentaje de valores similares entre Pchic y Rchic ') +
  ylab('Numero de bases') +
  ggtitle('Num.bases vs Pocentaje de valores similares ') +
  theme_minimal()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

Num.bases vs Pocentaje de valores similares



```
summary_table <- describe(data)
```

```
# Display the resultant table
print(summary_table)
```

	vars	n	mean
## N.BASES	1	383	192.00
## total.de.observaciones	2	383	198751776.50
## filas.observaciones.	3	383	49696.14
## variables	4	383	4582.99
## total.de.valores.en.la.matriz.de...similaridad	5	383	29700580.34
## Rchic.total.similaridad	6	383	14847998.67
## Pchic.total.similaridad	7	383	14816570.85
## cant..valores.no.similares	8	383	31427.83
## porcentaje.de.valores.similares.entre.2.matrices	9	383	1.00
## porcentaje.de.no.similaridad	10	383	0.00
## nodospchic	11	383	37.47
## nodosrchic	12	383	39.43
## porcentaje.nodos.iguales	13	383	0.95
## Rchic.total.similaridad.6.	14	382	14693171.45
## Pchic.total.similaridad.6.	15	382	14693171.45
## porcentaje.de.valores.similares.6.	16	383	1.00
##		sd	median
## N.BASES		110.71	192.00
## total.de.observaciones	184671616.88		144420000.00
## filas.observaciones.	57433.32		43200.00

## variables	2952.88	4600.00
## total.de.valores.en.la.matriz.de...similaridad	29049143.28	21160000.00
## Rchic.total.similaridad	14523147.61	10577700.00
## Pchic.total.similaridad	14484626.50	10465799.00
## cant..valores.no.similares	110338.97	7121.00
## porcentaje.de.valores.similares.entre.2.matrices	0.00	1.00
## porcentaje.de.no.similaridad	0.00	0.00
## nodospchic	3.60	38.00
## nodosrchic	3.80	40.00
## porcentaje.nodos.iguales	0.01	0.95
## Rchic.total.similaridad.6.	14439706.39	10350225.00
## Pchic.total.similaridad.6.	14439706.39	10350225.00
## porcentaje.de.valores.similares.6.	0.00	1.00
##	trimmed	mad
## N.BASES	192.00	142.33
## total.de.observaciones	177585374.59	167326236.00
## filas.observaciones.	45254.14	32320.68
## variables	4525.73	3854.76
## total.de.valores.en.la.matriz.de...similaridad	26169837.13	28866222.00
## Rchic.total.similaridad	13082655.70	14430664.71
## Pchic.total.similaridad	13065322.61	14266570.54
## cant..valores.no.similares	12187.75	8435.99
## porcentaje.de.valores.similares.entre.2.matrices	1.00	0.00
## porcentaje.de.no.similaridad	0.00	0.00
## nodospchic	38.00	0.00
## nodosrchic	40.00	0.00
## porcentaje.nodos.iguales	0.95	0.00
## Rchic.total.similaridad.6.	12920466.34	14093410.27
## Pchic.total.similaridad.6.	12920466.34	14093410.27
## porcentaje.de.valores.similares.6.	1.00	0.00
##	min	max
## N.BASES	1.00e+00	3.830000e+02
## total.de.observaciones	1.52e+05	1.577800e+09
## filas.observaciones.	2.00e+02	8.190000e+05
## variables	3.00e+00	1.000000e+04
## total.de.valores.en.la.matriz.de...similaridad	9.00e+00	1.000000e+08
## Rchic.total.similaridad	3.00e+00	4.999500e+07
## Pchic.total.similaridad	3.00e+00	4.998845e+07
## cant..valores.no.similares	0.00e+00	1.250351e+06
## porcentaje.de.valores.similares.entre.2.matrices	9.70e-01	1.000000e+00
## porcentaje.de.no.similaridad	0.00e+00	3.000000e-02
## nodospchic	1.00e+00	3.900000e+01
## nodosrchic	1.00e+00	4.100000e+01
## porcentaje.nodos.iguales	8.00e-01	1.000000e+00
## Rchic.total.similaridad.6.	3.00e+00	4.999500e+07
## Pchic.total.similaridad.6.	3.00e+00	4.999500e+07
## porcentaje.de.valores.similares.6.	1.00e+00	1.000000e+00
##	range	skew kurtosis
## N.BASES	3.820000e+02	0.00 -1.21
## total.de.observaciones	1.577648e+09	1.98 9.02
## filas.observaciones.	8.188000e+05	8.84 104.55
## variables	9.997000e+03	0.11 -1.22
## total.de.valores.en.la.matriz.de...similaridad	9.999999e+07	0.79 -0.61
## Rchic.total.similaridad	4.999500e+07	0.79 -0.61


```
## Pchic.total.similaridad          4.998845e+07  0.79   -0.61
## cant..valores.no.similares      1.250351e+06  8.39   81.52
## porcentaje.de.valores.similares.entre.2.matrices 3.000000e-02 -3.08   11.40
## porcentaje.de.no.similaridad     3.000000e-02  2.86    9.40
## nodospchic                      3.800000e+01 -8.78   79.03
## nodosrchic                      4.000000e+01 -8.68   77.55
## porcentaje.nodos.iguales         2.000000e-01 -3.84   65.00
## Rchic.total.similaridad.6.       4.999500e+07  0.80   -0.60
## Pchic.total.similaridad.6.       4.999500e+07  0.80   -0.60
## porcentaje.de.valores.similares.6. 0.000000e+00  NaN    NaN
##                                se
## N.BASES                        5.66
## total.de.observaciones         9436278.82
## filas.observaciones            2934.71
## variables                       150.89
## total.de.valores.en.la.matriz.de...similaridad 1484341.88
## Rchic.total.similaridad          742098.18
## Pchic.total.similaridad          740129.84
## cant..valores.no.similares       5638.06
## porcentaje.de.valores.similares.entre.2.matrices 0.00
## porcentaje.de.no.similaridad      0.00
## nodospchic                      0.18
## nodosrchic                      0.19
## porcentaje.nodos.iguales          0.00
## Rchic.total.similaridad.6.       738799.65
## Pchic.total.similaridad.6.       738799.65
## porcentaje.de.valores.similares.6. 0.00
```

histgrama

```
A<-ggplot(data,aes(x=Rchic.total.similaridad))+
geom_histogram(bins=10,aes(y = ..density..),fill="gray", color = "black", size=0.1)+
geom_density(col="blue")+
stat_function(fun=dnorm,color="red",args = list( mean(Rchic.total.similaridad),sd(Rchic.total.similaridad),
labs(title=" Distribucion de valores de similaridad de Rchic ", y="Densidad",x="rchicsimilardiad")
```

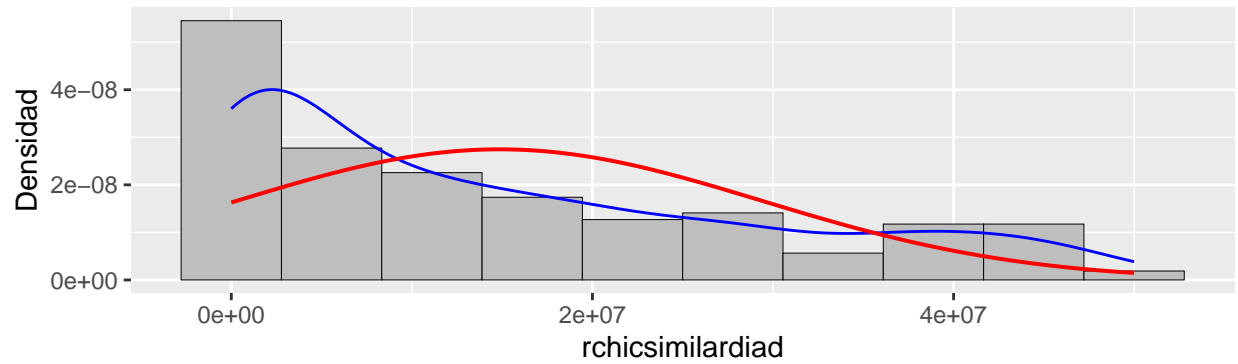
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
B<-ggplot(data,aes(x=Pchic.total.similaridad))+
geom_histogram(bins=10,aes(y = ..density..),fill="gray", color = "black", size=0.1)+
geom_density(col="blue")+
stat_function(fun=dnorm,color="red",args = list( mean(Pchic.total.similaridad),sd(Pchic.total.similaridad),
labs(title="Distribucion de valores de similaridad de Pchic", y="Densidad",x="Pchicsimilardiad")
```

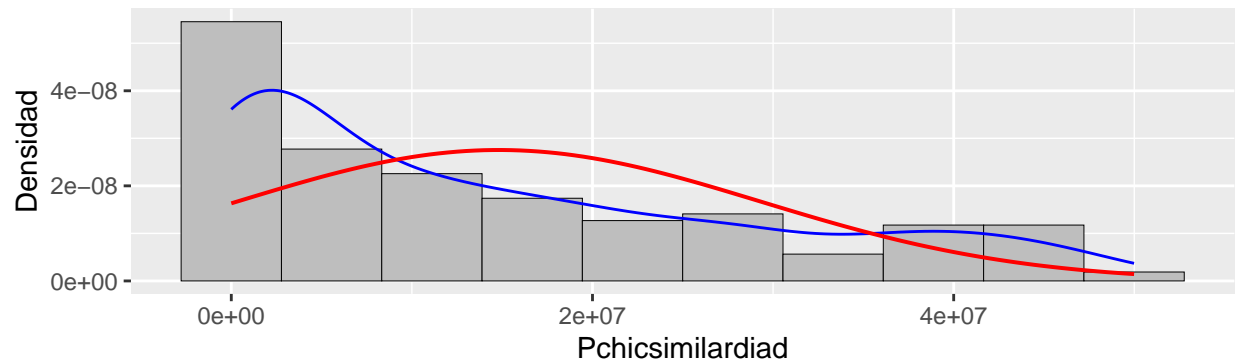
```
grid.arrange(A,B)
```

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

Distribucion de valores de similaridad de Rchic



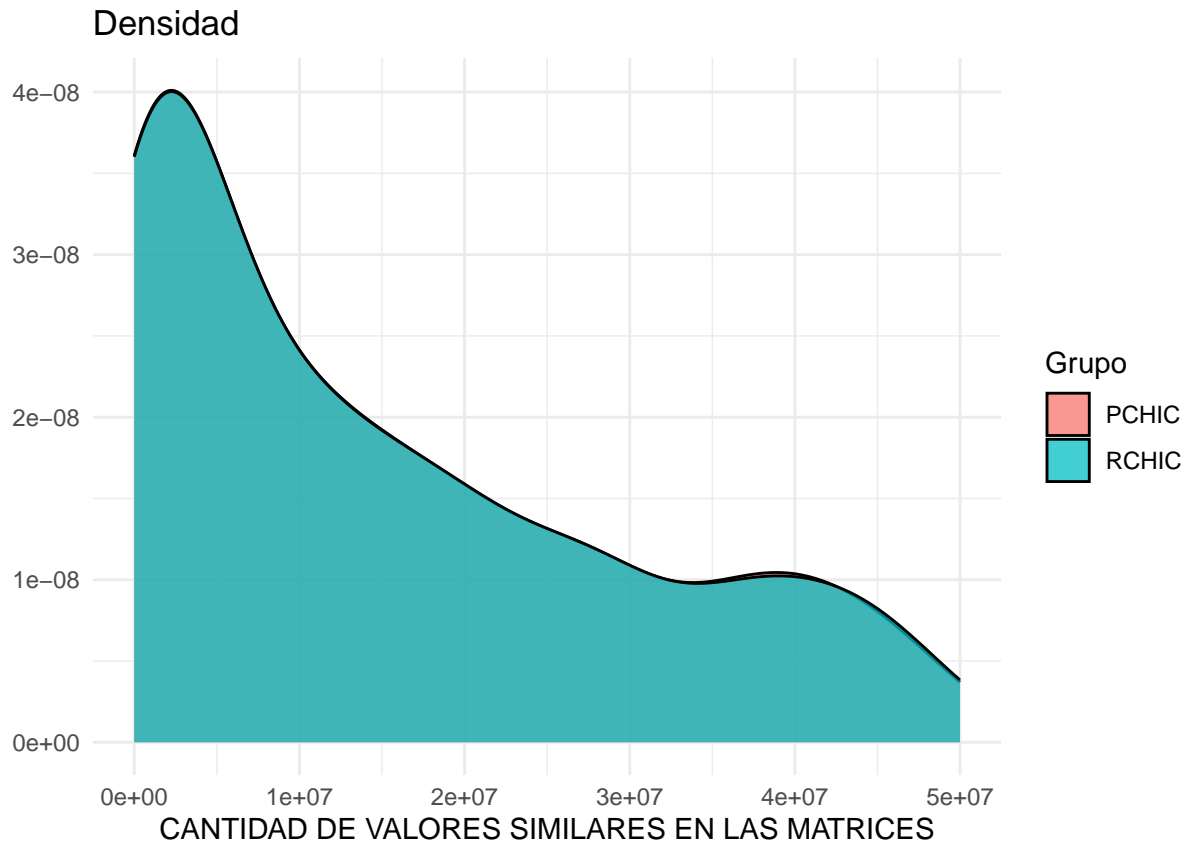
Distribucion de valores de similaridad de Pchic



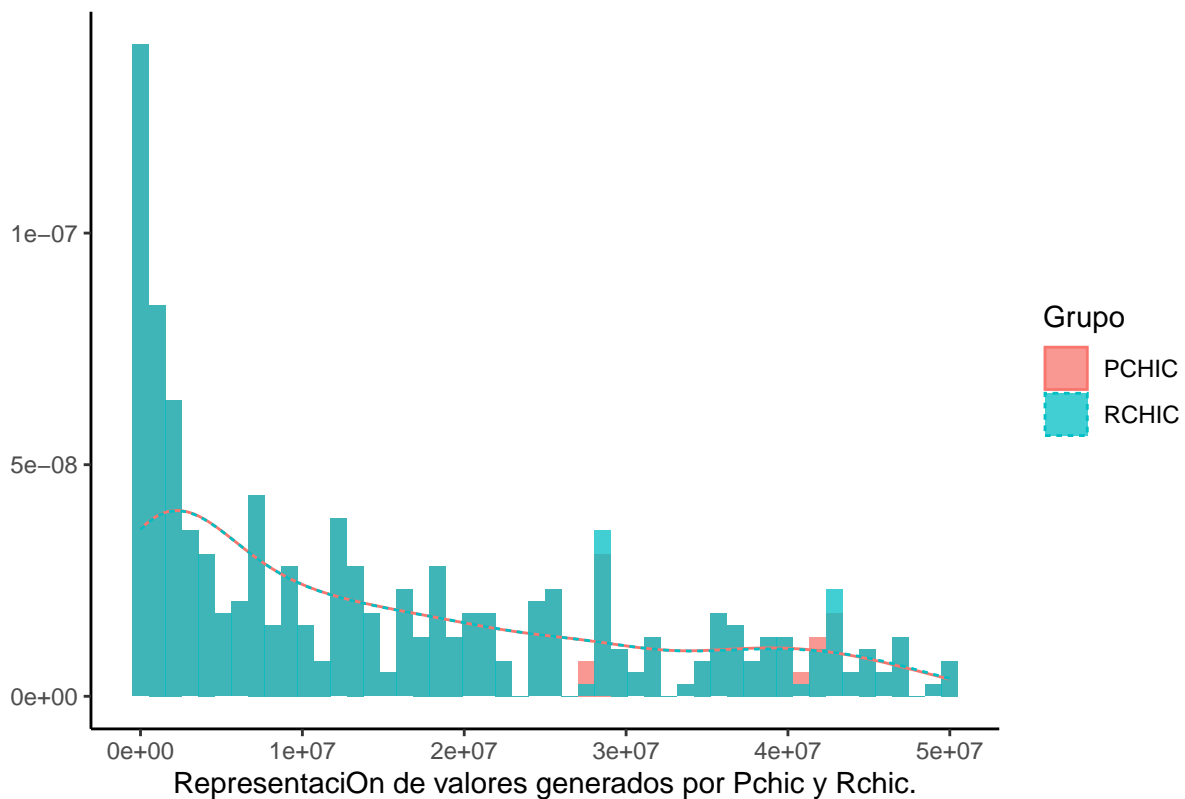
representacion conjunta

```
x1 <- Pchic.total.similaridad
x2 <- Rchic.total.similaridad
x <- c(x1, x2)
Grupo <- c(rep("PCHIC", length(Pchic.total.similaridad)), rep("RCHIC", length(Rchic.total.similaridad)))
df <- data.frame(x, group = Grupo)

ggplot(df)+
  aes(x=x, fill=Grupo)+
  geom_density(alpha=0.75)+
  labs(title="Densidad", x="CANTIDAD DE VALORES SIMILARES EN LAS MATRICES", y="")+
  guides(colour="none")+
  theme_minimal()
```



```
ggplot(df)+
  aes(x=x,fill=Grupo)+
  geom_histogram(aes(y=..density..),bins=50,alpha=0.75,position="identity")+
  geom_density(aes(linetype=Grupo,colour=Grupo),alpha=0)+
  labs(title="",x="RepresentaciOn de valores generados por Pchic y Rchic.",y="")+
  theme_classic()
```



prueba de hipotesis

$H_0: u \leq 0.9$

$H_1: u > 0.9$

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v forcats 1.0.0      v stringr 1.5.0
## v lubridate 1.9.3    v tibble 3.2.1
## v purrr 1.0.2       v tidyr 1.3.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x ggplot2::%+%( )      masks psych::%+%( )
## x ggplot2::alpha( )    masks psych::alpha( )
## x gridExtra::combine( ) masks dplyr::combine( )
## x dplyr::filter( )     masks stats::filter( )
## x xts::first( )        masks dplyr::first( )
## x dplyr::lag( )        masks stats::lag( )
## x xts::last( )         masks dplyr::last( )
## x car::recode( )       masks dplyr::recode( )
## x purrr::some( )       masks car::some( )
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
t.test(x=porcentaje.de.valores.similares.entre.2.matrices , alternative = "greater", mu=0.9 , conf.level=0.95)
```

```
##
```

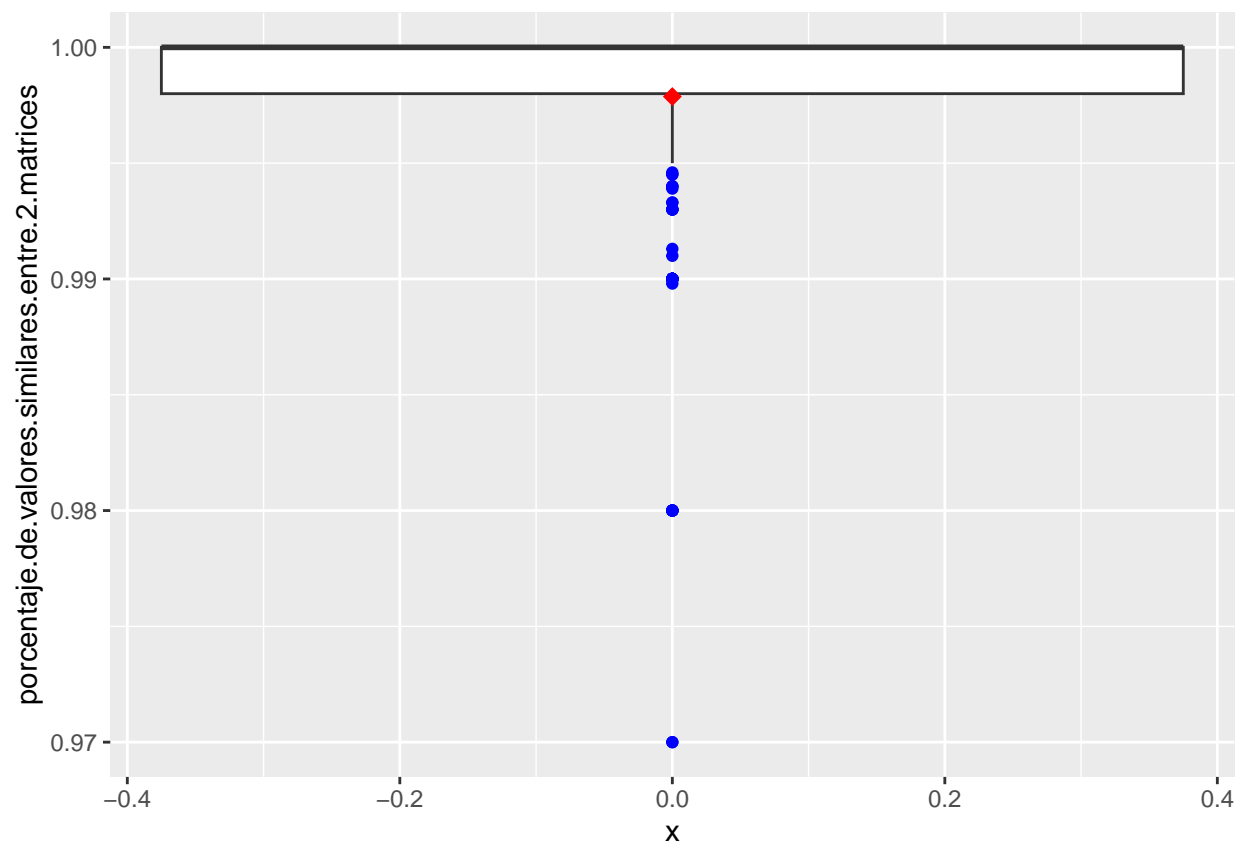
```
## One Sample t-test
##
## data: porcentaje.de.valores.similares.entre.2.matrices
## t = 430.69, df = 382, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 0.9
## 95 percent confidence interval:
## 0.9974996      Inf
## sample estimates:
## mean of x
## 0.9978743
```

```
library(tidyverse)
t.test(x=porcentaje.nodos.iguales, alternative = "greater", mu=0.9 , conf.level = 0.95)
```

```
##
## One Sample t-test
##
## data: porcentaje.nodos.iguales
## t = 81.382, df = 382, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 0.9
## 95 percent confidence interval:
## 0.9494437      Inf
## sample estimates:
## mean of x
## 0.9504661
```

```
ggplot(data, aes(x=0, y=porcentaje.de.valores.similares.entre.2.matrices))+ geom_boxplot(outlier.colour=
stat_summary(fun.y=mean, geom="point", shape=18,size=3, color="red")
```

```
## Warning: The `fun.y` argument of `stat_summary()` is deprecated as of ggplot2 3.3.0.
## i Please use the `fun` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



```
resumen_psych <- describe(data)
```

```
kable(print(resumen_psych))
```

	vars	n	mean
## N.BASES	1	383	192.00
## total.de.observaciones	2	383	198751776.50
## filas.observaciones.	3	383	49696.14
## variables	4	383	4582.99
## total.de.valores.en.la.matriz.de...similaridad	5	383	29700580.34
## Rchic.total.similaridad	6	383	14847998.67
## Pchic.total.similaridad	7	383	14816570.85
## cant..valores.no.similares	8	383	31427.83
## porcentaje.de.valores.similares.entre.2.matrices	9	383	1.00
## porcentaje.de.no.similaridad	10	383	0.00
## nodospchic	11	383	37.47
## nodosrchic	12	383	39.43
## porcentaje.nodos.iguales	13	383	0.95
## Rchic.total.similaridad.6.	14	382	14693171.45
## Pchic.total.similaridad.6.	15	382	14693171.45
## porcentaje.de.valores.similares.6.	16	383	1.00
##		sd	median
## N.BASES		110.71	192.00
## total.de.observaciones	184671616.88	144420000.00	
## filas.observaciones.	57433.32	43200.00	
## variables	2952.88	4600.00	

## total.de.valores.en.la.matriz.de...similaridad	29049143.28	21160000.00
## Rchic.total.similaridad	14523147.61	10577700.00
## Pchic.total.similaridad	14484626.50	10465799.00
## cant..valores.no.similares	110338.97	7121.00
## porcentaje.de.valores.similares.entre.2.matrices	0.00	1.00
## porcentaje.de.no.similaridad	0.00	0.00
## nodospchic	3.60	38.00
## nodosrchic	3.80	40.00
## porcentaje.nodos.iguales	0.01	0.95
## Rchic.total.similaridad.6.	14439706.39	10350225.00
## Pchic.total.similaridad.6.	14439706.39	10350225.00
## porcentaje.de.valores.similares.6.	0.00	1.00
##	trimmed	mad
## N.BASES	192.00	142.33
## total.de.observaciones	177585374.59	167326236.00
## filas.observaciones.	45254.14	32320.68
## variables	4525.73	3854.76
## total.de.valores.en.la.matriz.de...similaridad	26169837.13	28866222.00
## Rchic.total.similaridad	13082655.70	14430664.71
## Pchic.total.similaridad	13065322.61	14266570.54
## cant..valores.no.similares	12187.75	8435.99
## porcentaje.de.valores.similares.entre.2.matrices	1.00	0.00
## porcentaje.de.no.similaridad	0.00	0.00
## nodospchic	38.00	0.00
## nodosrchic	40.00	0.00
## porcentaje.nodos.iguales	0.95	0.00
## Rchic.total.similaridad.6.	12920466.34	14093410.27
## Pchic.total.similaridad.6.	12920466.34	14093410.27
## porcentaje.de.valores.similares.6.	1.00	0.00
##	min	max
## N.BASES	1.00e+00	3.830000e+02
## total.de.observaciones	1.52e+05	1.577800e+09
## filas.observaciones.	2.00e+02	8.190000e+05
## variables	3.00e+00	1.000000e+04
## total.de.valores.en.la.matriz.de...similaridad	9.00e+00	1.000000e+08
## Rchic.total.similaridad	3.00e+00	4.999500e+07
## Pchic.total.similaridad	3.00e+00	4.998845e+07
## cant..valores.no.similares	0.00e+00	1.250351e+06
## porcentaje.de.valores.similares.entre.2.matrices	9.70e-01	1.000000e+00
## porcentaje.de.no.similaridad	0.00e+00	3.000000e-02
## nodospchic	1.00e+00	3.900000e+01
## nodosrchic	1.00e+00	4.100000e+01
## porcentaje.nodos.iguales	8.00e-01	1.000000e+00
## Rchic.total.similaridad.6.	3.00e+00	4.999500e+07
## Pchic.total.similaridad.6.	3.00e+00	4.999500e+07
## porcentaje.de.valores.similares.6.	1.00e+00	1.000000e+00
##	range	skew kurtosis
## N.BASES	3.820000e+02	0.00 -1.21
## total.de.observaciones	1.577648e+09	1.98 9.02
## filas.observaciones.	8.188000e+05	8.84 104.55
## variables	9.997000e+03	0.11 -1.22
## total.de.valores.en.la.matriz.de...similaridad	9.999999e+07	0.79 -0.61
## Rchic.total.similaridad	4.999500e+07	0.79 -0.61
## Pchic.total.similaridad	4.998845e+07	0.79 -0.61

```

## cant..valores.no.similares          1.250351e+06  8.39    81.52
## porcentaje.de.valores.similares.entre.2.matrices 3.000000e-02 -3.08    11.40
## porcentaje.de.no.similaridad         3.000000e-02  2.86     9.40
## nodospchic                          3.800000e+01 -8.78    79.03
## nodosrchic                          4.000000e+01 -8.68    77.55
## porcentaje.nodos.iguales             2.000000e-01 -3.84    65.00
## Rchic.total.similaridad.6.           4.999500e+07  0.80    -0.60
## Pchic.total.similaridad.6.           4.999500e+07  0.80    -0.60
## porcentaje.de.valores.similares.6.   0.000000e+00  NaN     NaN
##                                     se
## N.BASES                             5.66
## total.de.observaciones               9436278.82
## filas.observaciones.                 2934.71
## variables                            150.89
## total.de.valores.en.la.matriz.de...similaridad 1484341.88
## Rchic.total.similaridad               742098.18
## Pchic.total.similaridad               740129.84
## cant..valores.no.similares           5638.06
## porcentaje.de.valores.similares.entre.2.matrices 0.00
## porcentaje.de.no.similaridad          0.00
## nodospchic                           0.18
## nodosrchic                           0.19
## porcentaje.nodos.iguales              0.00
## Rchic.total.similaridad.6.            738799.65
## Pchic.total.similaridad.6.            738799.65
## porcentaje.de.valores.similares.6.    0.00

```

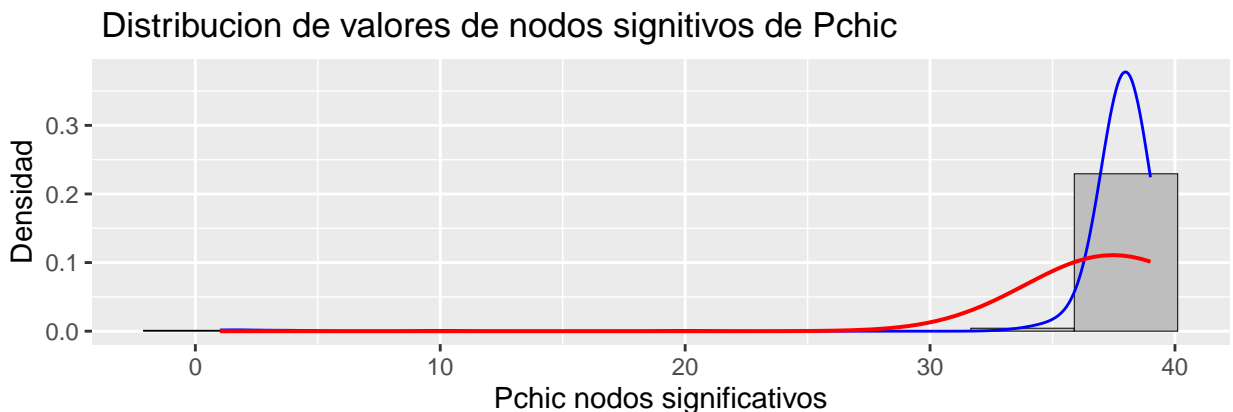
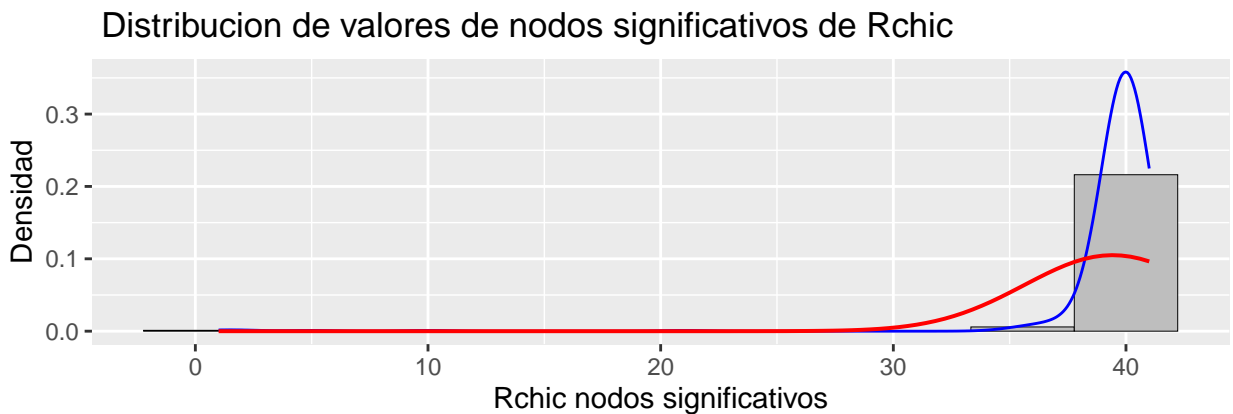
		vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
N.BASES	1	383	192.00	110.71	192.00	192.00	142.33	1.00e+00	3.000000e+00	3.000000e+00	0.00	-1.21	5.66	
total.de.observaciones	2	383	198751776.40	1614482000.00	770853167392623520	15778001570648989.02	9436278.82							
filas.observaciones.	3	383	49696.14	7433.32	3200.00	5254.14	2320.68	2.00e+01	8190000e+00	8190000e+00	0.00	14.55	2934.71	
variables	4	383	4582.99	2952.88	4600.00	4525.73	3854.76	3.00e+00	1000000e+00	997000e+00	1.03	-1.22	150.89	
total.de.valores.en.la.matriz	5	383	295005809.49	1431280000.00	1698378866229.00	1000000e+00	998999e+00	7.97	-0.61	1484341.88				
Rchic.total.similaridad	6	383	148479984523141057770030826554730663.70	1099500e+00	997500e+00	7.97	-0.61	742098.18						
Pchic.total.similaridad	7	383	148165704854626050579930653224266570.00	1099845e+00	99845e+00	7.97	-0.61	740129.84						
cant..valores.no.similares	8	383	31427.83	110338.97	21.00	12187.78	8435.99	0.00e+00	1250351e+06	1250351e+06	8.39	81.52	5638.06	
porcentaje.de.valores.similares.entre.2.matrices	9	383	0.00	0.00	1.00	1.00	0.00	9.70e-01	1.000000e+00	1.000000e+00	-3.08	11.40	0.00	
porcentaje.de.no.similaridad	10	383	0.00	0.00	0.00	0.00	0.00	0.00e+00	3.000000e+00	3.000000e+00	2.86	9.40	0.00	
nodospchic	11	383	37.47	3.60	38.00	38.00	0.00	1.00e+01	3.800000e+01	3.800000e+01	-8.78	79.03	0.18	
nodosrchic	12	383	39.43	3.80	40.00	40.00	0.00	1.00e+01	4.000000e+01	4.000000e+01	-8.68	77.55	0.19	
porcentaje.nodos.iguales	13	383	0.95	0.01	0.95	0.95	0.00	8.00e-01	1.000000e+00	1.000000e+00	-3.84	65.00	0.00	
Rchic.total.similaridad	14	382	146931714459706035022529204664993410.00	1099500e+00	997500e+00	8.07	-0.60	738799.65						

Pchic.total.similaridad	15.382	14693171445397061350225129204663393418.07	e+09	99500e+99500e+807	-	738799.65				
porcentaje.de.valores.similares	16.383	6.1.00	0.00	1.00	1.00	0.00	1.00e+10	000000e+000000e+NaN	NaN	0.00

```
C<-ggplot(data,aes(x=nodosrhcic))+
geom_histogram(bins=10,aes(y = ..density..),fill="gray", color = "black", size=0.1)+
geom_density(col="blue")+
stat_function(fun=dnorm,color="red",args = list( mean(nodosrhcic),sd(nodosrhcic)),size=0.7)+
labs(title=" Distribucion de valores de nodos significativos de Rchic ", y="Densidad",x="Rchic nodos si

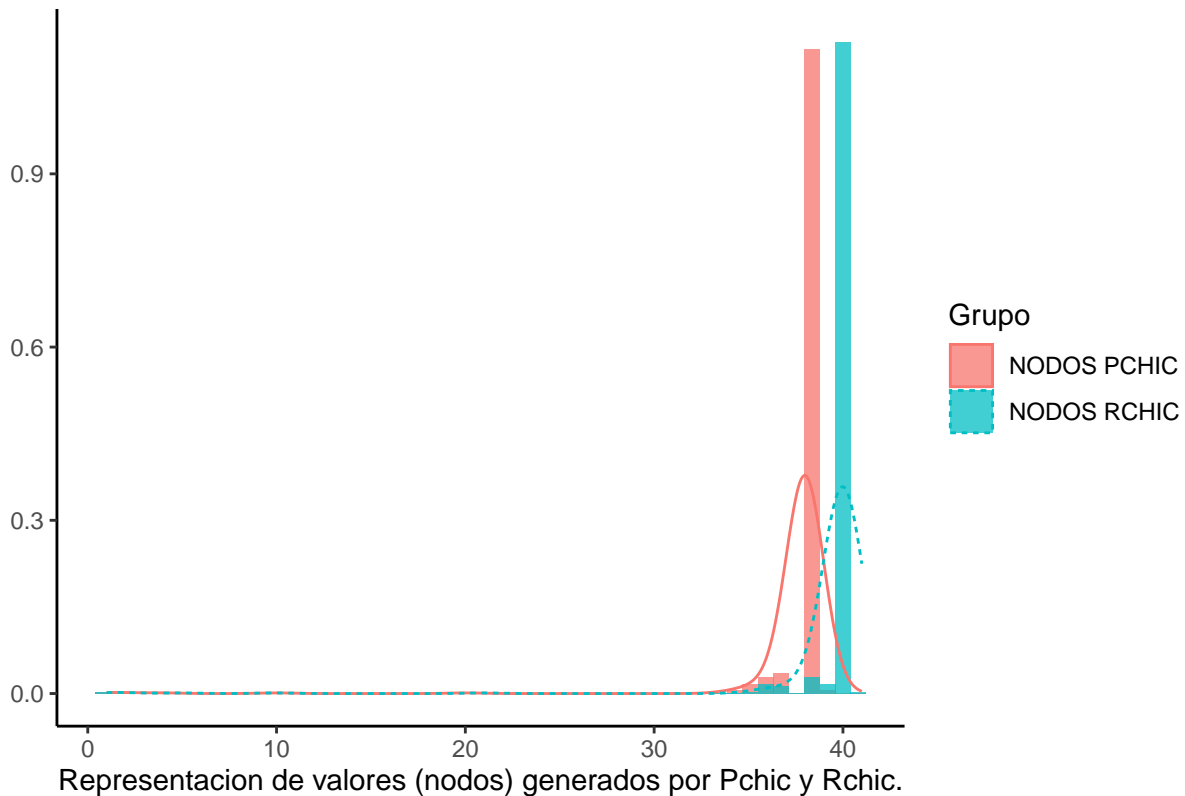
D<-ggplot(data,aes(x=nodospchic))+
geom_histogram(bins=10,aes(y = ..density..),fill="gray", color = "black", size=0.1)+
geom_density(col="blue")+
stat_function(fun=dnorm,color="red",args = list( mean(nodospchic),sd(nodospchic)),size=0.7)+
labs(title=" Distribucion de valores de nodos signitivos de Pchic ", y="Densidad",x="Pchic nodos signif

grid.arrange(C,D)
```



```
x1 <- nodospchic
x2 <- nodosrchic
x <- c(x1, x2)
Grupo <- c(rep("NODOS PCHIC", length(nodospchic)), rep("NODOS RCHIC", length(nodosrchic)))
df <- data.frame(x, group = Grupo)
```

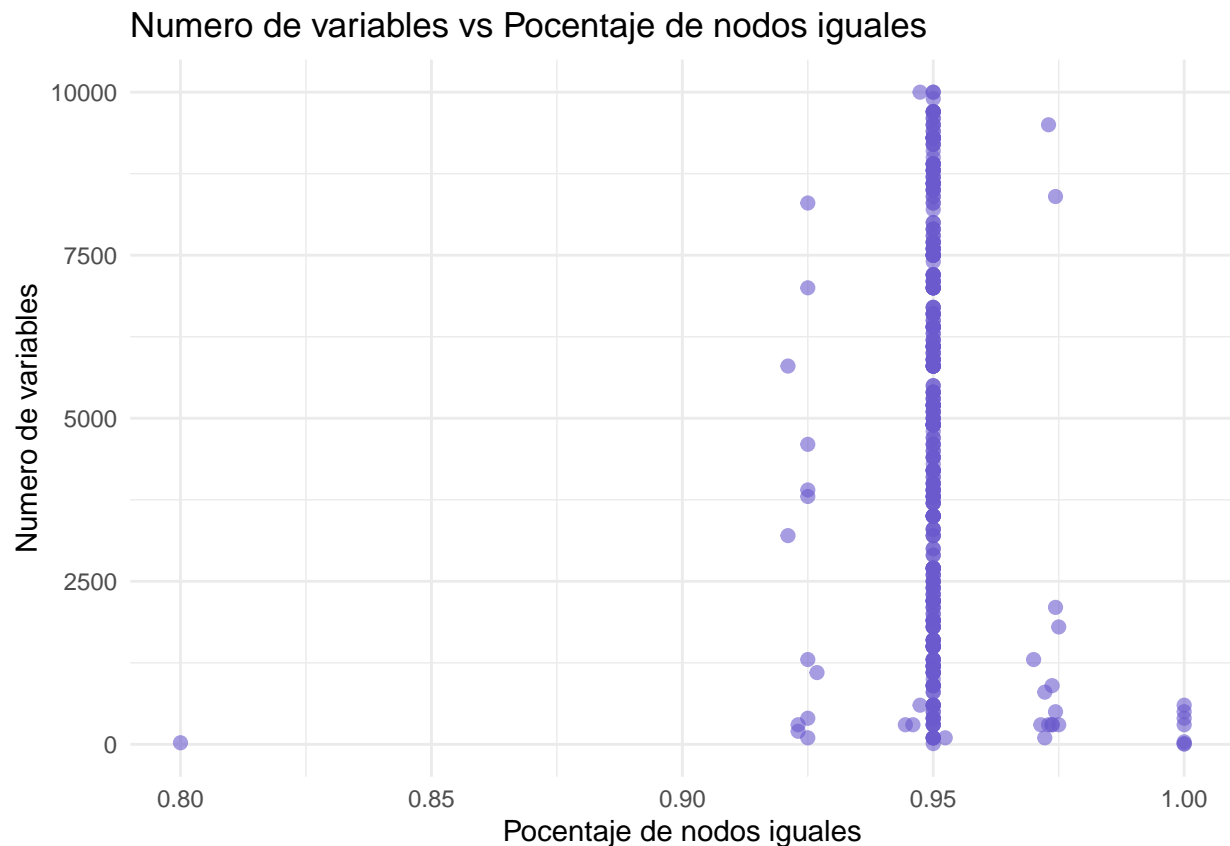
```
ggplot(df)+
  aes(x=x,fill=Grupo)+
  geom_histogram(aes(y=..density..),bins=50,alpha=0.75,position="identity")+
  geom_density(aes(linetype=Grupo,colour=Grupo),alpha=0)+
  labs(title="",x="Representacion de valores (nodos) generados por Pchic y Rchic.",y="")+
  theme_classic()
```



```
ggplot(data = data, aes(x = porcentaje.nodos.iguales, y = variables)) +
  geom_point(color = 'slateblue', size = 2, alpha = 0.6) +
  geom_smooth(color = 'red') +
  xlab('Pocentaje de nodos iguales') +
  ylab('Numero de variables') +
  ggtitle('Numero de variables vs Pocentaje de nodos iguales') +
  theme_minimal()
```

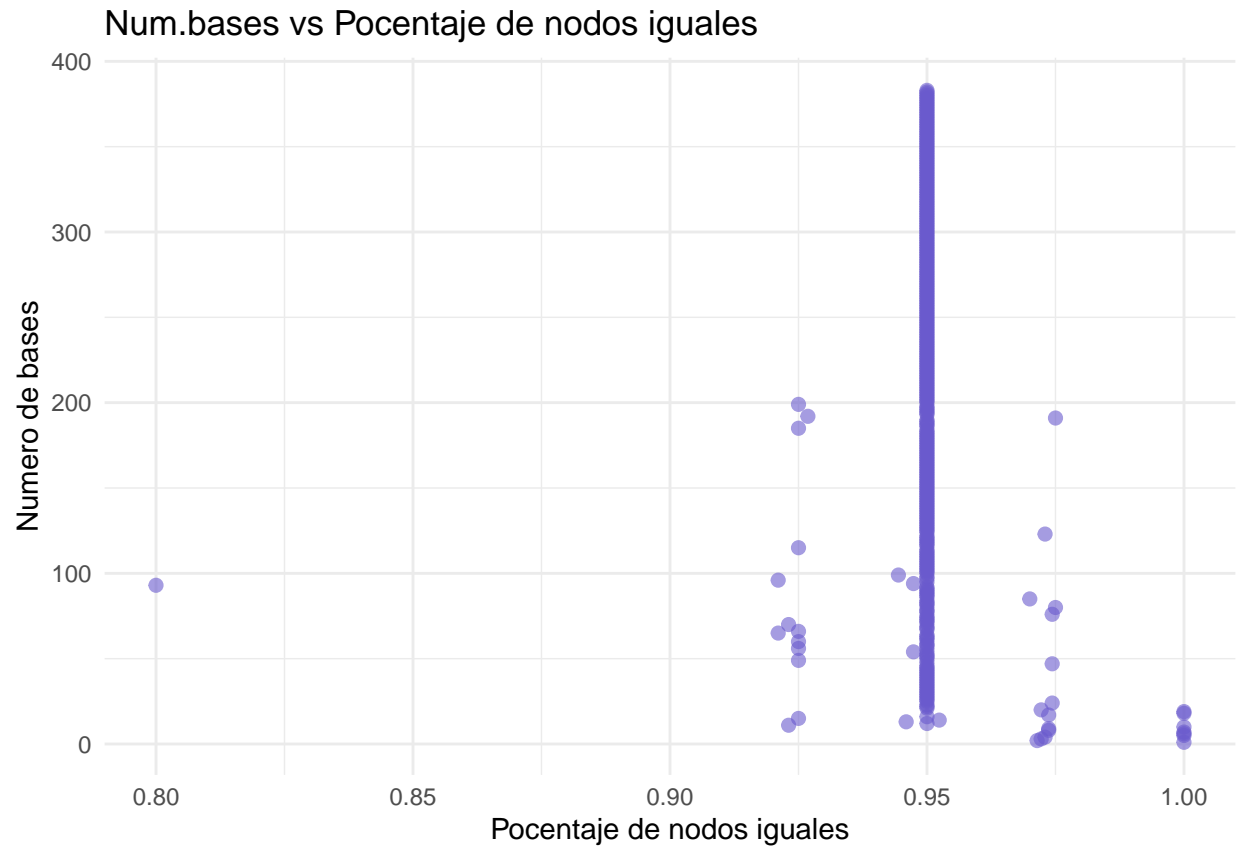
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : zero-width neighborhood. make span bigger
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : pseudoinverse used at 0.94737
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : neighborhood radius 0.00263
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : reciprocal condition number 5.1288e-17
```

```
## Warning: Computation failed in `stat_smooth()`
## Caused by error in `predLoess()`:
## ! NA/NaN/Inf en llamada a una función externa (arg 5)
```



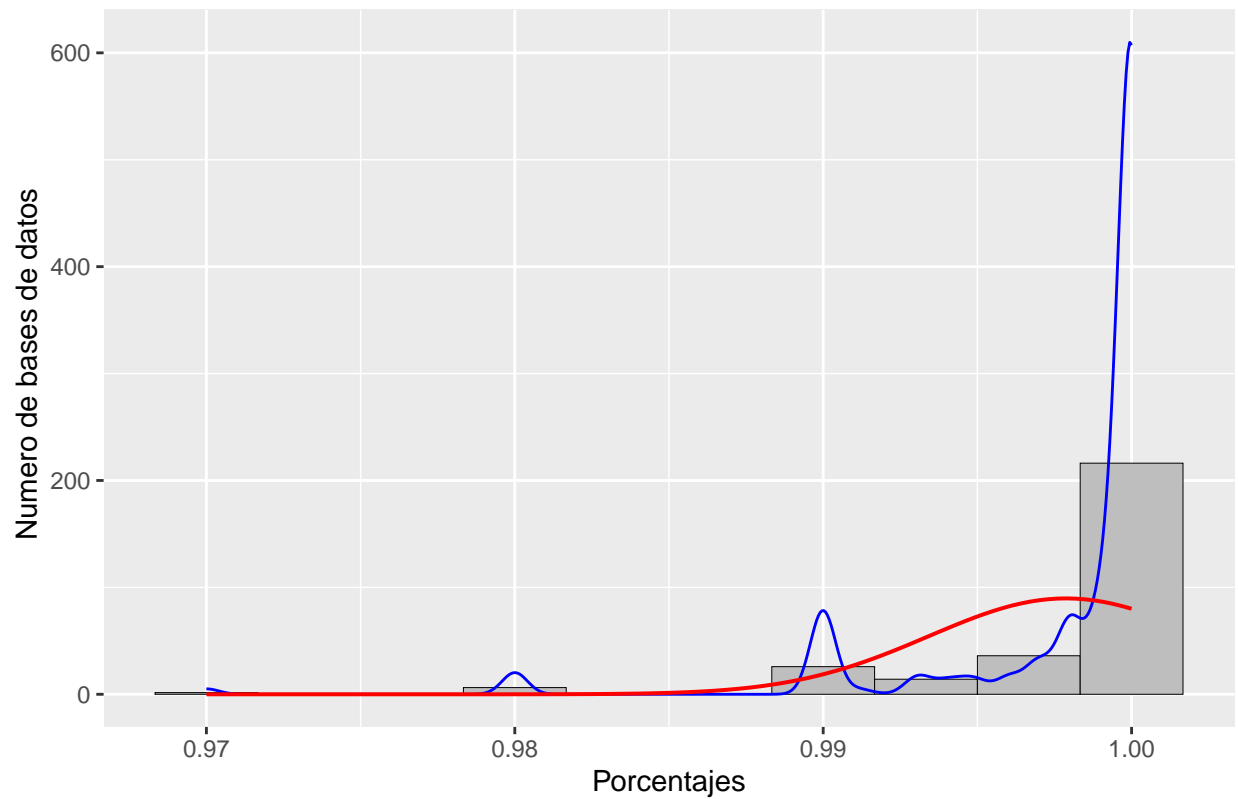
```
ggplot(data = data, aes(x = porcentaje.nodos.iguales, y = N.BASES)) +
  geom_point(color = 'slateblue', size = 2, alpha = 0.6) +
  geom_smooth(color = 'red') +
  xlab('Pocentaje de nodos iguales') +
  ylab('Numero de bases') +
  ggtitle('Num.bases vs Pocentaje de nodos iguales') +
  theme_minimal()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : zero-width neighborhood. make span bigger
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : pseudoinverse used at 0.94737
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : neighborhood radius 0.00263
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : reciprocal condition number 5.1288e-17
## Warning: Computation failed in `stat_smooth()`
## Caused by error in `predLoess()`:
## ! NA/NaN/Inf en llamada a una función externa (arg 5)
```



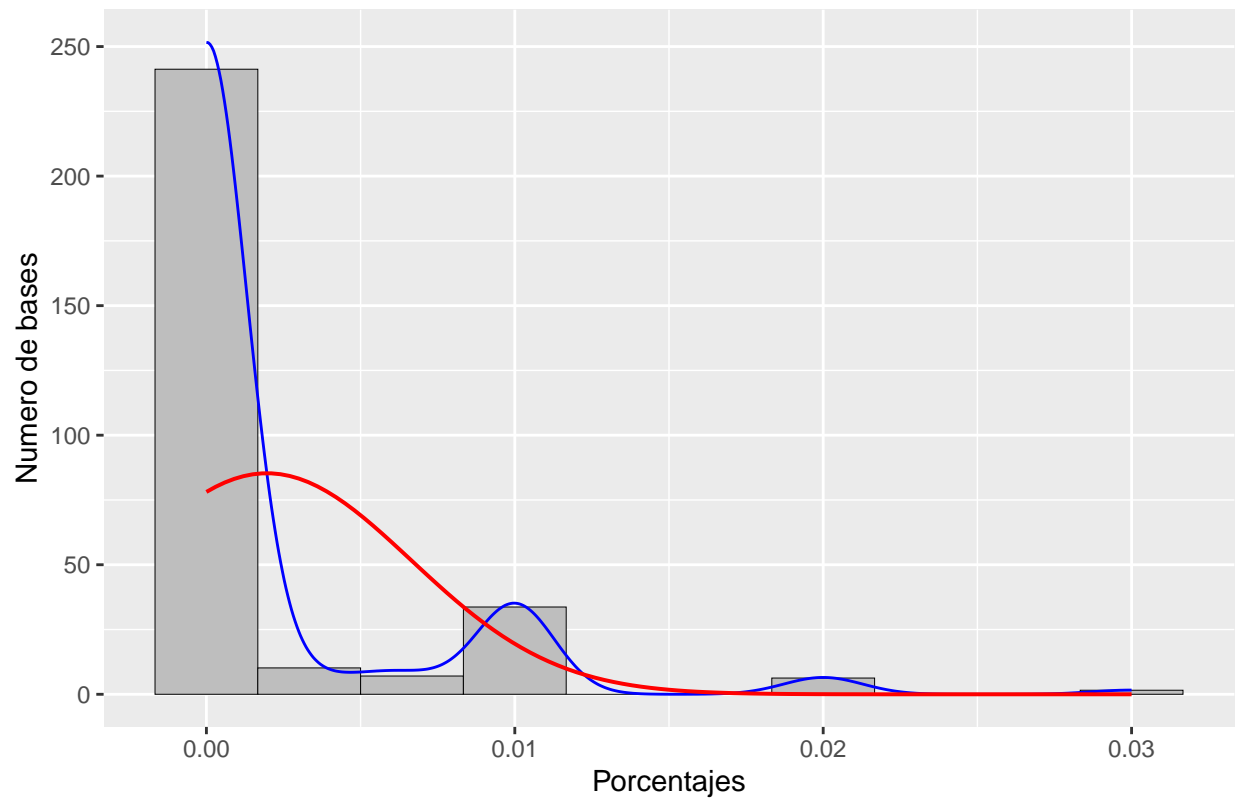
```
ggplot(data,aes(x=porcentaje.de.valores.similares.entre.2.matrices))+
geom_histogram(bins=10,aes(y = ..density..),fill="gray", color = "black", size=0.1)+
geom_density(col="blue")+
stat_function(fun=dnorm,color="red",args = list( mean(porcentaje.de.valores.similares.entre.2.matrices)
labs(title="Distribucion de los porcentajes de valores similares entre Pchic y Rchic", y=" Numero de ba
```

Distribucion de los porcentajes de valores similares entre Pchic y Rchic



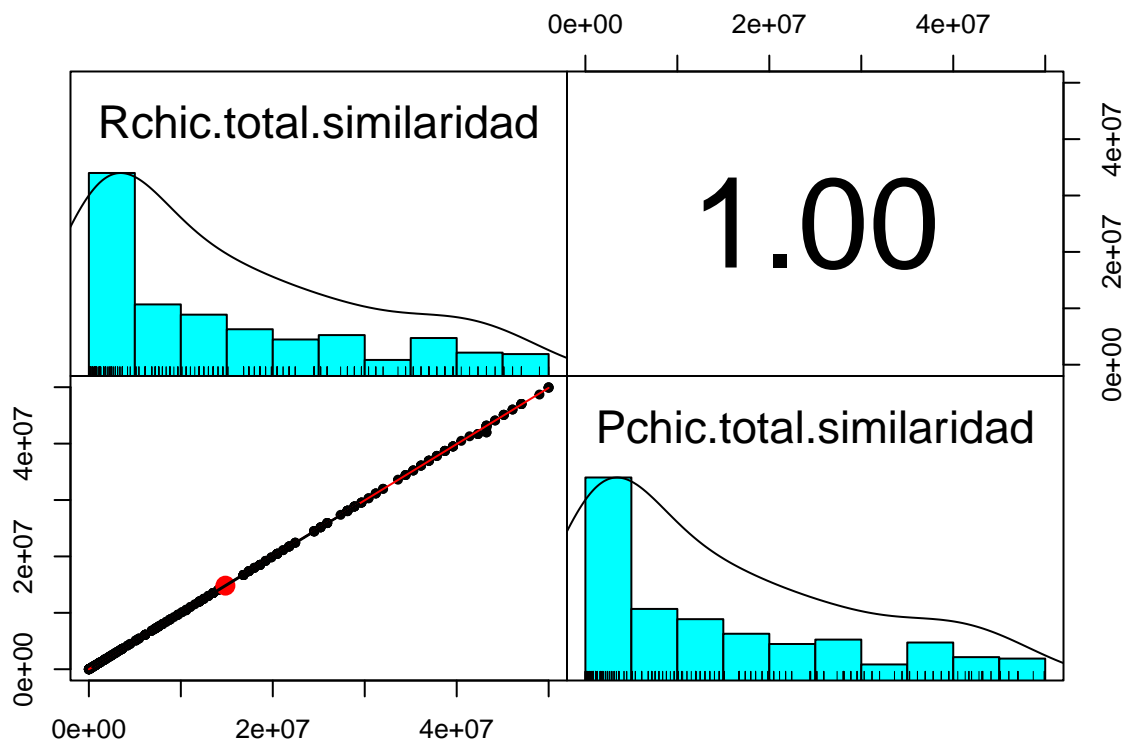
```
ggplot(data,aes(x=porcentaje.de.no.similaridad))+
geom_histogram(bins=10,aes(y = ..density..),fill="gray", color = "black", size=0.1)+
geom_density(col="blue")+
stat_function(fun=dnorm,color="red",args = list( mean(porcentaje.de.no.similaridad),sd(porcentaje.de.no
labs(title="Distribucion de  porcentajes de valores no similares entre Pchic y Rchic", y="Numero de bas
```

Distribucion de porcentajes de valores no similares entre Pchic y Rchic



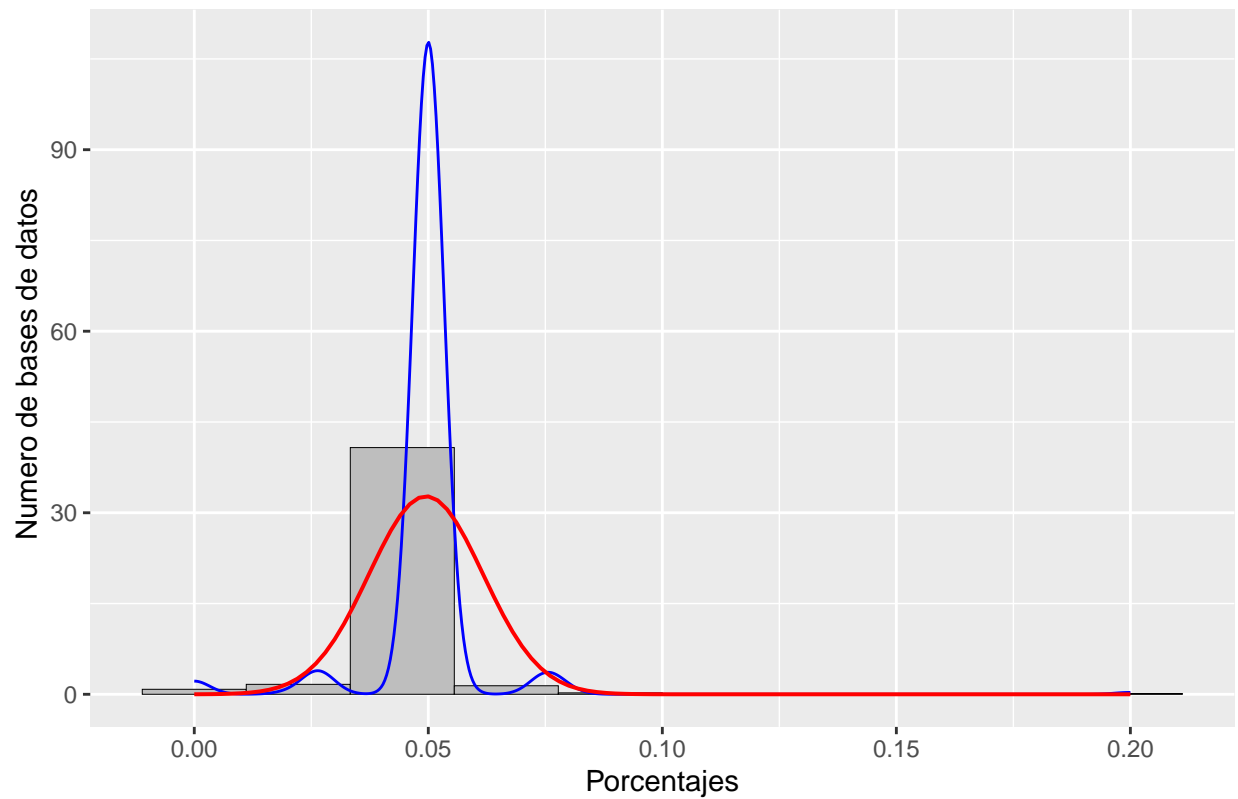
```
#dendogrograma
```

```
pairs.panels(data[6:7],gap=0)
```



```
ggplot(data,aes(x=porcentaje.nodos.no.iguales))+
geom_histogram(bins=10,aes(y = ..density..),fill="gray", color = "black", size=0.1)+
geom_density(col="blue")+
stat_function(fun=dnorm,color="red",args = list( mean(porcentaje.nodos.no.iguales),sd(porcentaje.nodos.no.iguales)))
labs(title="Distribucion de los porcentajes de nodos no significativos entre Pchic y Rchic", y="Numero de nodos no significativos")
```

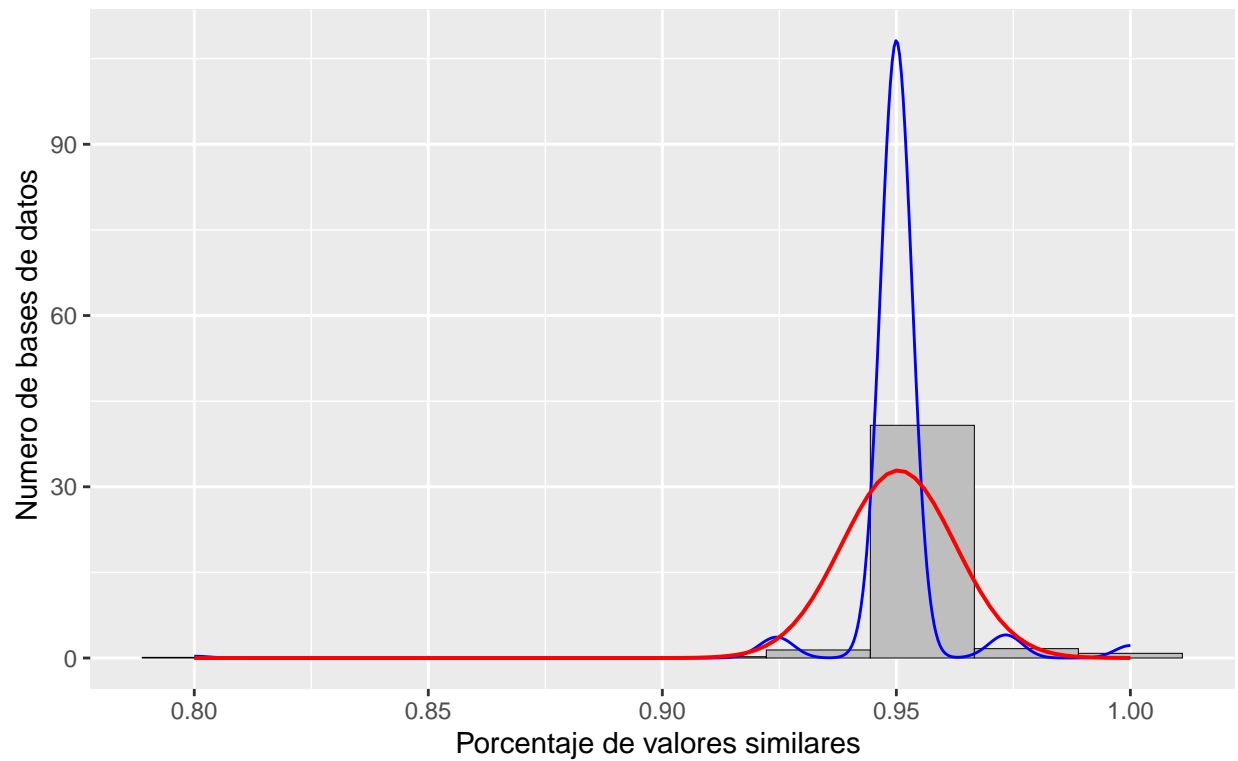
Distribucion de los porcentajes de nodos no significativos entre Pchic y Rch



DISTRIBUCION

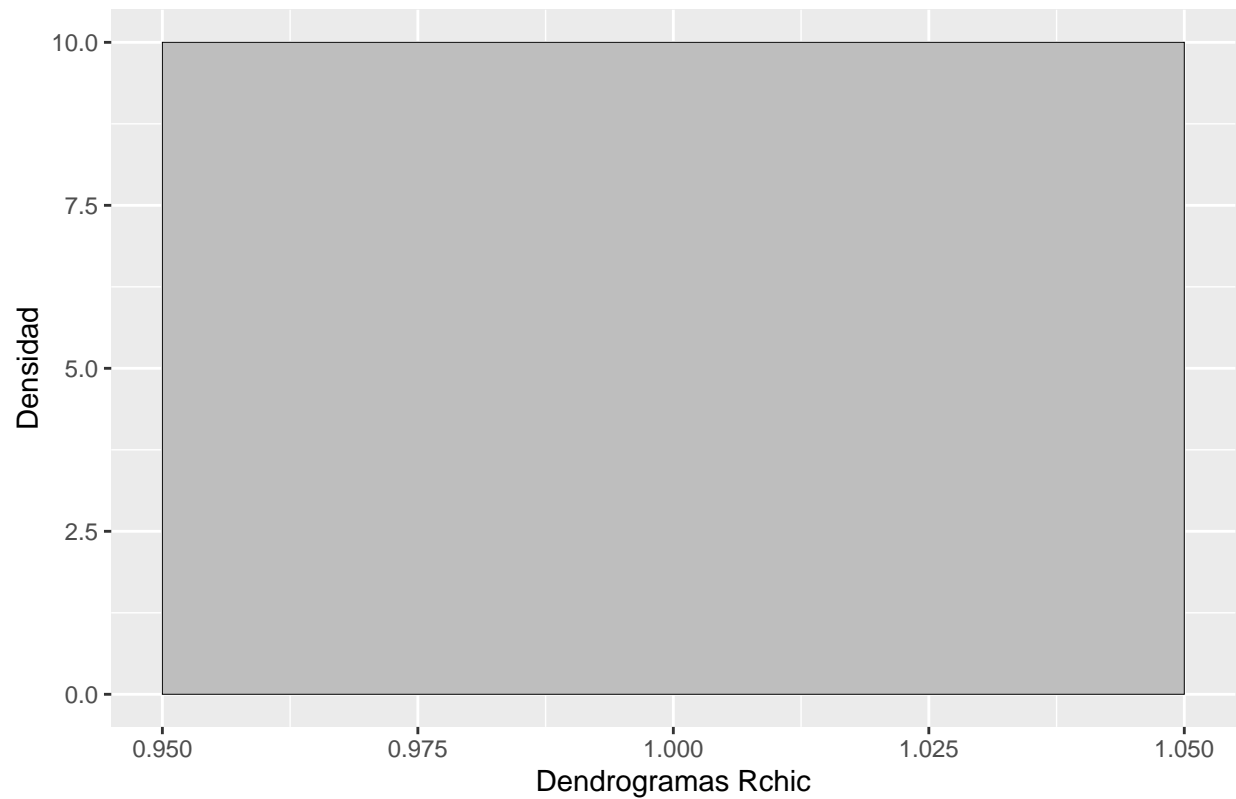
```
ggplot(data,aes(x=porcentaje.nodos.iguales))+
geom_histogram(bins=10,aes(y = ..density..),fill="gray", color = "black", size=0.1)+
geom_density(col="blue")+
stat_function(fun=dnorm,color="red",args = list( mean(porcentaje.nodos.iguales),sd(porcentaje.nodos.iguales)))
labs(title="Distribucion porcentajes de valores  nodos significativos
", y=" Numero de bases de datos",x="Porcentaje de valores similares ")
```


Distribucion porcentajes de valores nodos significativos



```
ggplot(data,aes(x=dendogramas.r.chic))+  
geom_histogram(bins=10,aes(y = ..density..),fill="gray", color = "black", size=0.1)+  
geom_density(col="blue")+  
stat_function(fun=dnorm,color="red",args = list( mean(dendogramas.r.chic),sd(dendogramas.r.chic)),size=0.1)+  
labs(title=" Histograma de dendogramas de Rchic ", y="Densidad",x="Dendogramas Rchic")
```

Histograma de dendrogramas de Rchic



```
ggplot(data,aes(x=dendrograma.pchic))+
geom_histogram(bins=10,aes(y = ..density..),fill="gray", color = "black", size=0.1)+
geom_density(col="blue")+
stat_function(fun=dnorm,color="red",args = list( mean(dendrograma.pchic),sd(dendrograma.pchic)),size=0.1)+
labs(title=" Histograma de dendrogramas de Rchic ", y="Densidad",x="Dendrogramas pchic")
```

Histograma de dendrogramas de Rchic

