

# INDICACIONES

UTI-Labos - Unidad de Tecnologías de la Información Administración de Laboratorios Informáticos Facultad de Ciencias Exactas y Naturales - UBA

Cuando abran la ventana de terminal:

1. Escriban: `conda init`
2. Luego cierren la ventana y vuelvan a abrirla.
3. Ahí escriben: `conda activate mejoramiento_genetico`
4. Luego → RStudio

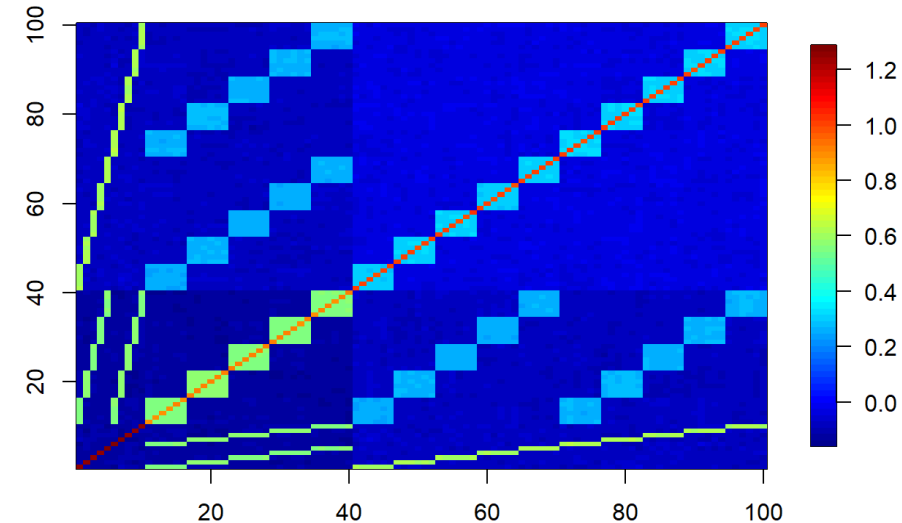
# simulMGF

**simGeno**: simula genotipos en **matrices de SNP** como valores aleatorios desde una distribución uniforme, para organismos diploides (codificados por 0, 1 y 2). (VanRaden et al., 2009; Sikorska et al., 2013)

**simPheno**: simula un **fenotipo desde una matriz de SNP** con loci de caracteres cuantitativos (**QTLs**) con efectos muestreados desde una distribución Normal.

**simulFS** y **simulHS**: simula los **genotipos de progenies de hermanos completos** y de **medios hermanos** desde el genotipo de los parentales.

**simulN** y **simulU**: simulan una **matriz de SNP** y **caracteres controlados por determinado número de QTLs** y sus **efectos muestreados desde una distribución Normal** o una **distribución Uniforme** respectivamente.



# simulMGF

**simGeno**(Nind, Nmarkers)

→

**simPheno**(x, Nqtl, Esigma, Pmean, Perror)

→

$$y = Pmean + \sum QTN \times Meffects + \varepsilon$$

**simulFS**(x, y, Nprogeny)

→

una matriz de dimensiones (nrow(x)\*Nprogeny) x ncol(x)

**simulHS**(x, Nprogeny)

→

una matriz de dimensiones (nrow(x)\*Nprogeny) x ncol(x)

**simulN**(Nind, Nmarkers, Nqtl, Esigma, Pmean, Perror)

→

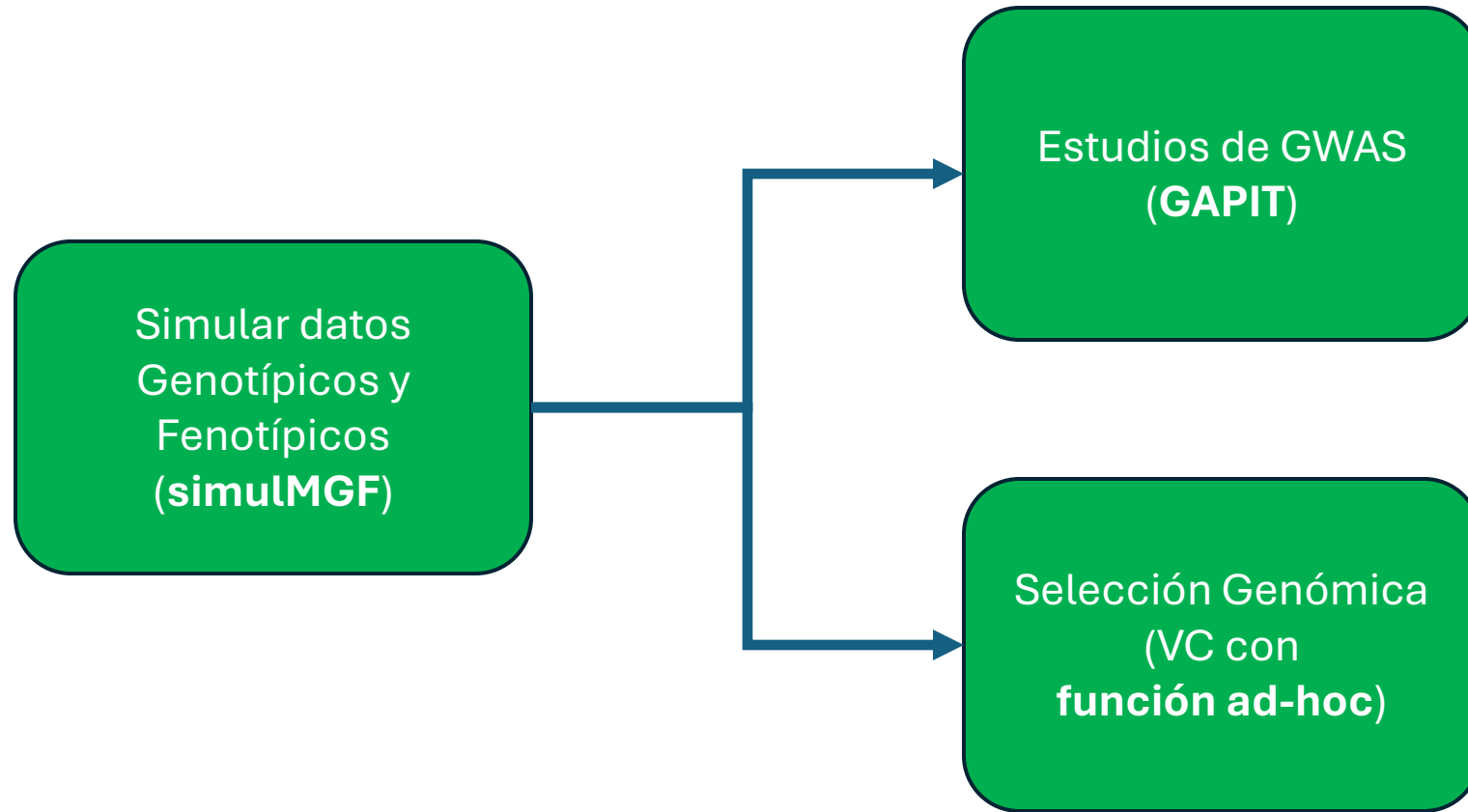
un objeto de clase “list” conteniendo una matriz de SNP, el caracter, los marcadores asociados y sus efectos.

**simulU**(Nind, Nmarkers, Nqtl, Pmean, Perror)

→

un objeto de clase “list” conteniendo una matriz de SNP, el caracter, los marcadores asociados y sus efectos.

TPs de “Mapeo por Asociación” y “Selección Genómica” de la materia de grado  
(UNSAM) y curso de posgrado (UBA)  
“Mejoramiento Genético y Genómico Vegetal”



# Simulación de datos

```
install.packages("simulMGF")
```

```
library(simulMGF)
```

```
set.seed(1234)
```

```
simulN(Nind = 1000, Nmarkers = 10000, Nqtl = 50,  
Esigma = .5, Pmean = 25, Perror = .25)
```

Código disponible en: <https://github.com/mngar/CURSO>

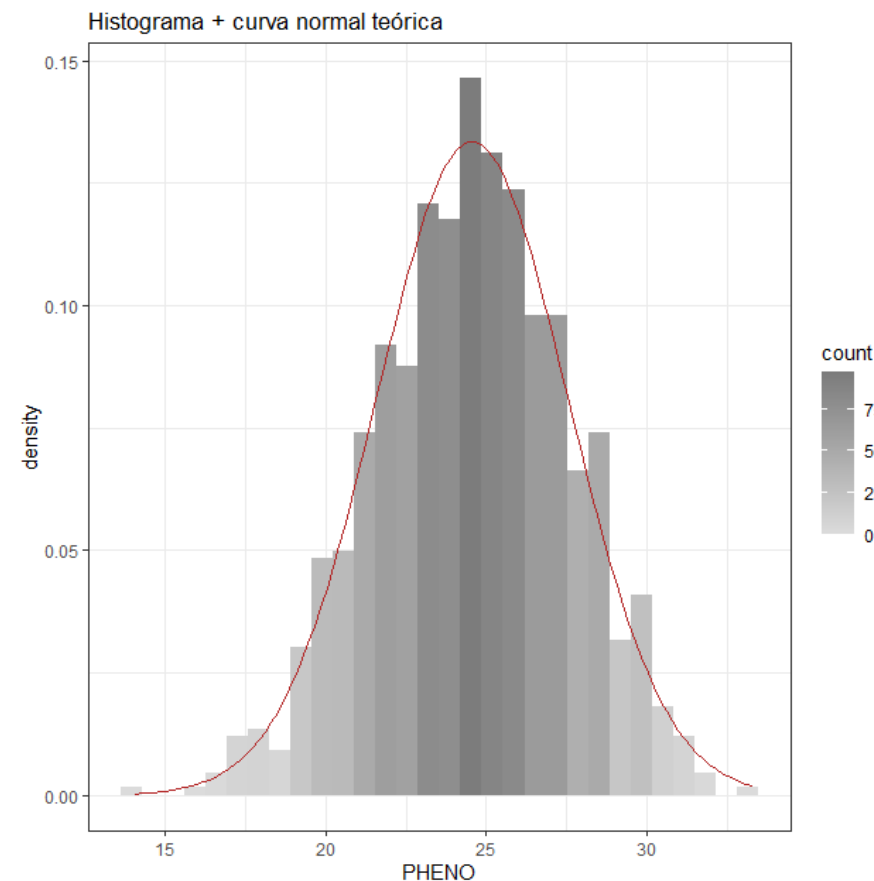
# OBJETO “nsimout”

`str(nsimout)`

List of 4

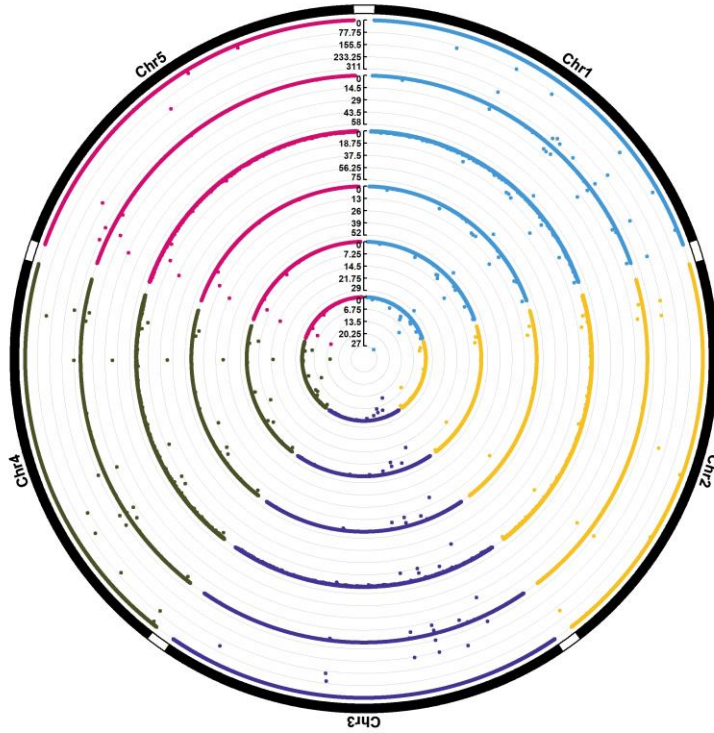
\$ geno : matriz de genotipos  
\$ pheno : vector de fenotipos  
\$ QTN : vector de marcadores asociados  
\$ Meffects : vector de efectos

$$y = P_{\text{mean}} + \sum \text{QTN} \times \text{Meffects} + \varepsilon$$



Código disponible en: <https://github.com/mngar/CURSO>

# GWAS



**Manhattan plot** con marcadores asociados para el carácter simulado para cada metodología (de adentro hacia afuera: GLM, MLM, SUPER, MLMM, FarmCPU y Blink).

**Poder estadístico** estimado para cada metodología de asociación ordenadas de acuerdo a lo esperado (mayor poder estadístico arriba).

Método	SNP_asociados	SNP_realmente_asociados	Falso_positivos	PODER
blink	49	48	1	0.96
FarmCPU	31	28	3	0.56
MLMM	32	32	0	0.64
SUPER	30	28	2	0.56
MLM	26	24	2	0.48
GLM	30	26	4	0.52

$$\text{PODER} = (\text{SNP\_realmente\_asociados} / \# \text{QTL})$$

Tibbs Cortes, L., Zhang, Z., & Yu, J. (2021). Status and prospects of genome-wide association studies in plants. The plant genome, 14(1), e20077.

Wang, J., Zhang, Z. (2021). GAPIT version 3: boosting power and accuracy for genomic association and prediction. Genomics, proteomics & bioinformatics, 19(4), 629-640.

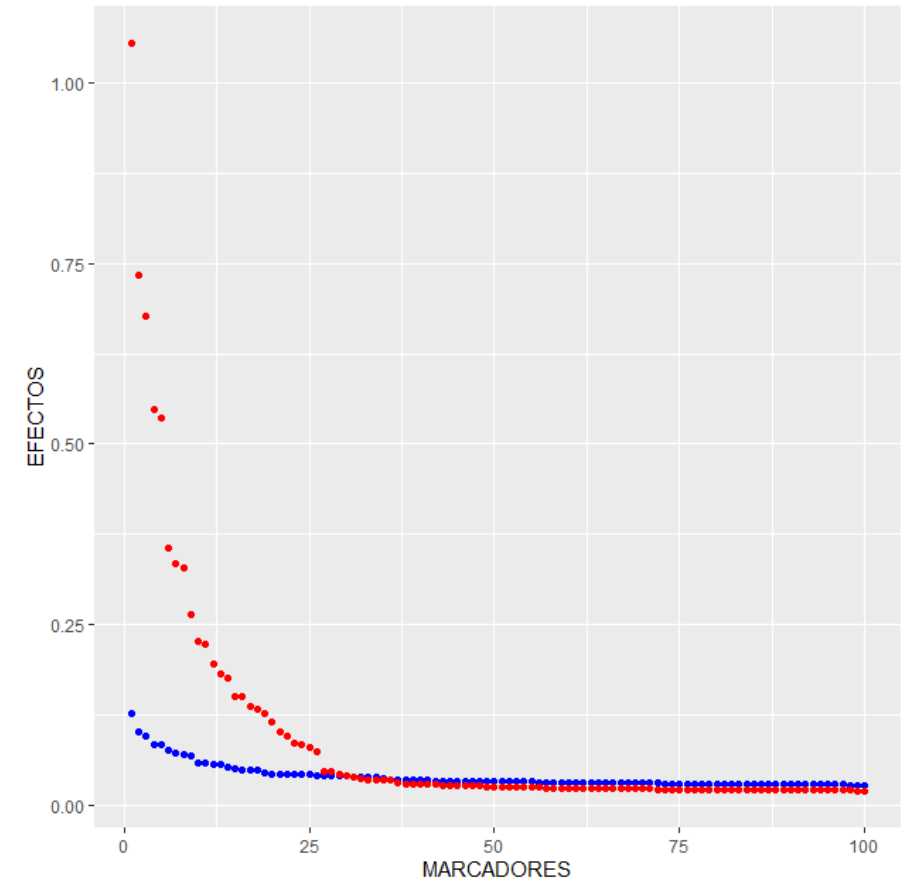
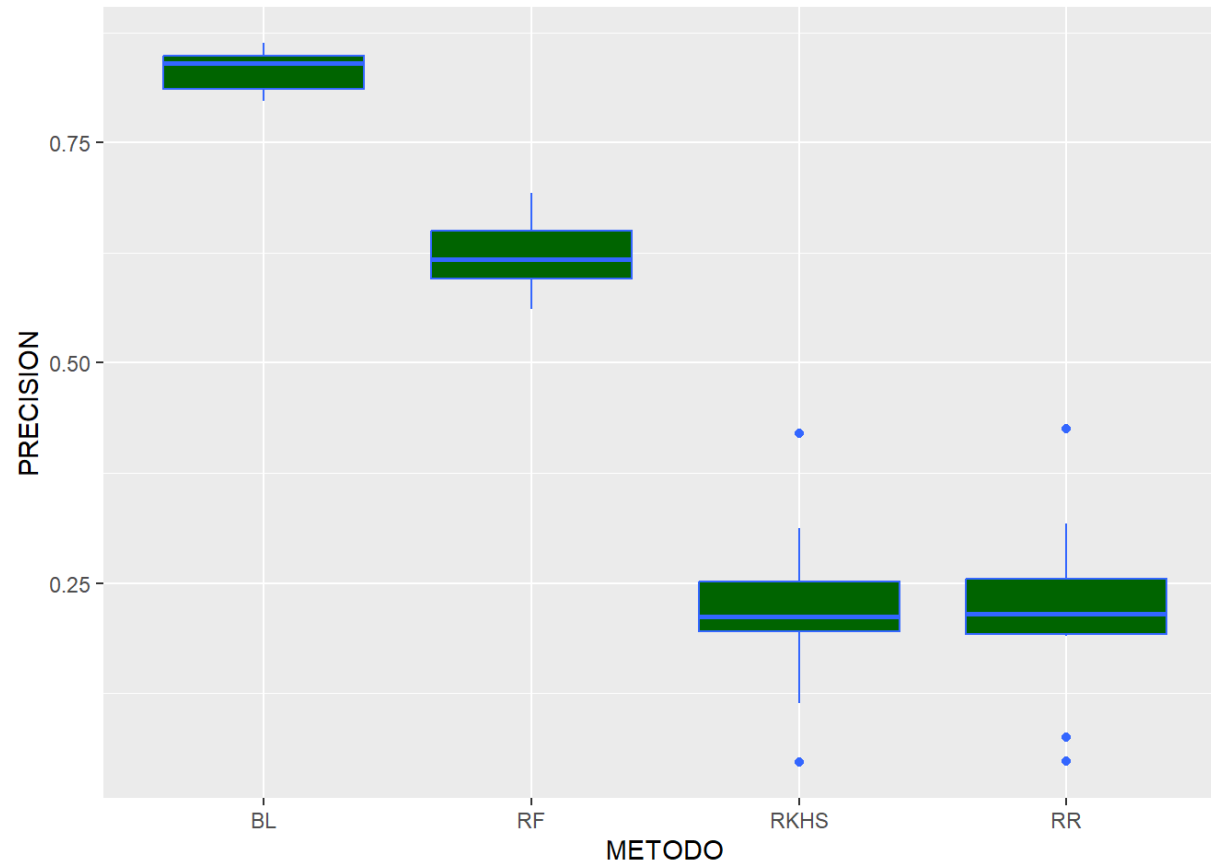
# SG

**RR:** Ridge Regression (regresión lineal)

**BL:** Bayesian LASSO (regresión lineal bayesiana)

**RF:** Random Forest (semiparamétrica)

**RKHS:** Reproducing Kernel Hilbert Spaces (no paramétrica)



**BL**

**RR**

- Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. The plant genome, 4(3).
- de Los Campos, et al., (2013). Genome-enabled prediction using the BLR (Bayesian Linear Regression) R-package. Genome-wide association studies and genomic prediction, 299-320.
- Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.
- Pérez, P., & de Los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. Genetics, 198(2), 483-495.



# Ventajas usar un “Proyecto R”

- Todos los archivos que generas están en el mismo lugar.
- Si quieres compartir tu proyecto, puedes compartir la carpeta completa y otros pueden reproducir rápidamente tu investigación o ayudar a solucionar problemas.
- Puedes usar GitHub para realizar copias de seguridad y el llamado "control de versiones", que te permite hacer un seguimiento de los cambios que has realizado en tu código a lo largo del tiempo.

**<https://github.com/mngar/ceroinfinito>**



New File



New Project...

Open File...



Open File in New Column...

Recent Files



Open Project...

Open Project in New Session...

Recent Projects



Import Dataset



Save



Save As...

Save All



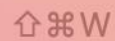
Publish...

Print...

Close



Close All



Close All Except Current



Close Project

Quit Session...



Console

Terminal x

 R 4.1.0 · /Volumes/

&gt; |

## New Project Wizard

## Create Project

**New Directory**

Start a project in a brand new working directory

**Existing Directory**

Associate a project with an existing working directory

**Version Control**

Checkout a project from a version control repository



Cancel

☐ arm

Data Analysis Using Regression 1.11-

Console

Terminal x

R 4.1.0 · /Volumes/

&gt; |

## New Project Wizard

## Create Project

**New Directory**

Start a project in a brand new working directory

**Existing Directory**

Associate a project with an existing working directory

**Version Control**

Checkout a project from a version control repository



Cancel

arm

Data Analysis Using Regression 1.11-

Go to file/function

Addins

R-For-Non-Programmers — GitHub

Console Terminal x

R 4.1.0 · /Volumes/

>

### New Project Wizard

## Project Type

Back

- New Project >
- R Package >
- Shiny Web Application >
- R Package using Rcpp >
- R Package using RcppArmadillo >
- R Package using RcppEigen >
- R Package using RcppParallel >

Cancel

arm Data Analysis Using Regression 1.11-



Console

Terminal x

 R 4.1.0 · /Volumes/

&gt; |

New Project Wizard

Back

## Create New Project



Directory name:

Create project as subdirectory of:

~/Downloads

Browse...

- ☐ Create a git repository
- ☐ Use renv with this project

☐ Open in new session

Create Project

Cancel

Tutorial

 List

Search

Search

Vers...

al 1.4-5

ata 1.8.0

oft 1.3

logical 2.0.8

tables

☐ arm

Data Analysis Using Regression 1.11-

Console

Terminal x

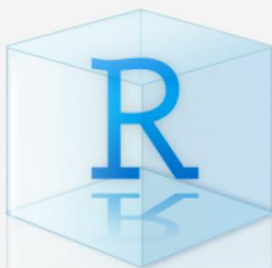
R 4.1.0 · /Volumes/

&gt;

New Project Wizard

Back

## Create New Project



Directory name:

tidy\_tuesday\_2021\_08\_03

Create project as subdirectory of:

~/Downloads

Browse...

☐ Create a git repository☐ Use renv with this project☒ Open in new session

Create Project

Cancel

Tutorial

List



Vers...

al 1.4-5

ata 1.8.0

oft 1.3

logical 2.0.8

tables

arm

Data Analysis Using Regression 1.11-

Console

Terminal ×

Jobs ×

 R 4.1.0 · ~/Downloads/tidy\_tuesday\_2021\_08\_03/ 

Y.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.

&gt;

Environment

History


Connections

Tutorial



Import Dataset ▾

402 MiB ▾

List ▾ R ▾ |  Global Environment ▾

Environment is empty

Files

Plots

Packages

Help

Viewer



New Folder



Delete



Rename



More ▾



Home &gt; Downloads &gt; tidy\_tuesday\_2021\_08\_03




▲ Name

Size

Modified

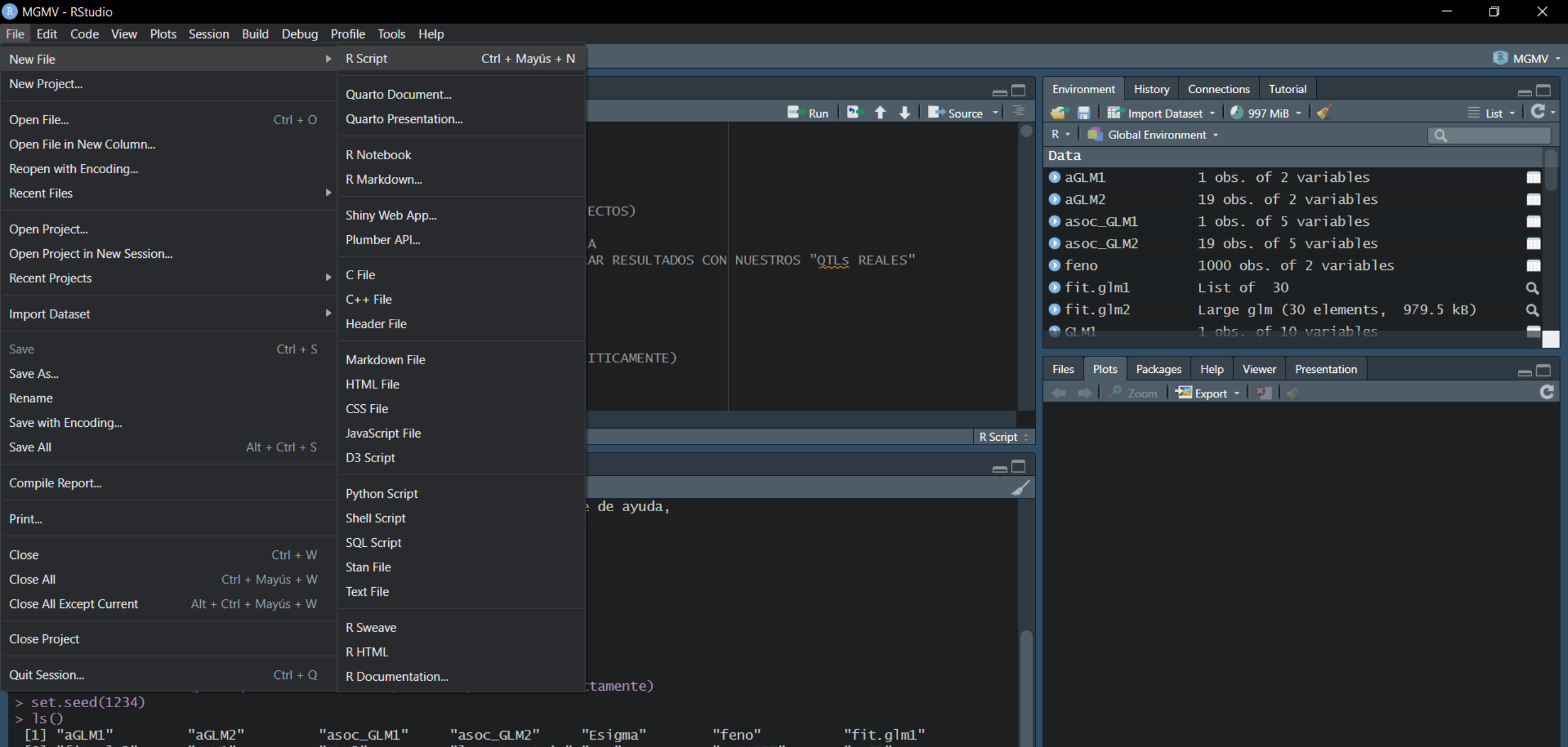


..



 tidy\_tuesday\_2021\_08\_03.Rproj 205 B

Jul 30, 2021







<https://github.com/mngar/ceroinfinito>




**ceroinfinito**
Public
Pin
Unwatch 1
Fork 0
Star 0

main 1 Branch Tags Go to file t Add file <> Code

 <b>mngar</b>	Clase en 0+inf	0b532a8 · 7 hours ago
 GWAS.R	Clase en 0+inf	7 hours ago
 LICENSE	Initial commit	7 hours ago
 README.md	Initial commit	7 hours ago
 SG.R	Clase en 0+inf	7 hours ago

 README

 MIT license




---

# ceroinfinito

---

Clase MGyGV\_2 de noviembre 2024

About 

Clase MGyGV\_2 de noviembre 2024

-  [Readme](#)
-  [MIT license](#)
-  [Activity](#)
-  [0 stars](#)
-  [1 watching](#)
-  [0 forks](#)

## Releases

No releases published

[Create a new release](#)

## Packages

No packages published

[Publish your first package](#)

## Languages

ceroinfinito/GWAS.R at main · mngar / ceroinfinito

github.com/mngar/ceroinfinito/blob/main/GWAS.R

mngar / ceroinfinito

Type / to search

Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

main

ceroinfinito / GWAS.R

Go to file

mngar

Clase en 0+inf

0b532a8 · 7 hours ago

History

Code

Blame

313 lines (233 loc) · 9.72 KB

Code 55% faster with GitHub Copilot

Raw

```
1 #clase MGyGV
2 #GWAS
3
4 # 1. INSTALAR Y CARGAR LOS PAQUETES NECESARIOS
5 # 2. SETEAR UNA SEMILLA ASI TODOS TIENEN LOS MISMOS DATOS
6 # 3. SIMULAR MATRIZ DE DATOS GENOMICOS Y FENOTIPICOS (QTLs y EFECTOS)
7 # 4. PONERLE NOMBRE A LOS INDIVIDUOS Y MARCADORES
8 # 5. RANKEAR MARCADORES ASOCIADOS Y EFECTOS: EXPORTAR COMO TABLA
9 # 6. REALIZAR ANALISIS GWAS CON DISTINTAS METODOLOGIAS Y COMPARAR RESULTADOS CON NUESTROS "QTLs REALES"
10
11 # instalacion de paquetes
12 #PAQUETE DE SIMULACION DE DATOS
13 install.packages("simulMGF")
14
15 #PAQUETES PARA CHECKEAR NORMALIDAD DEL FENOTIPO (GRAFICA Y ANALITICAMENTE)
16 install.packages("ggplot2")
17 install.packages("nortest")
18
19 #PAQUETE DE ANALISIS DE ASOCIACION
20 install.packages("remotes")
21 remotes::install_github("jiabowang/GAPIT3")
22
```

Buscar

6:15

02/11/2024

ceroinfinito/SG.R at main · mngar

github.com/mngar/ceroinfinito/blob/main/SG.R

mngar / ceroinfinito

Type / to search

Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

main

ceroinfinito / SG.R

Go to file

mngar

Clase en 0+inf

0b532a8 · 7 hours ago

History

Code

Blame

94 lines (67 loc) · 2.54 KB

Code 55% faster with GitHub Copilot

Raw

```
1 #clase MGyGV
2 #GWAS
3
4 # 1. INSTALAR Y CARGAR LOS PAQUETES NECESARIOS
5 # 2. SETEAR UNA SEMILLA ASI TODOS TIENEN LOS MISMOS DATOS
6 # 3. SIMULAR MATRIZ DE DATOS GENOMICOS Y FENOTIPICOS (QTLs y EFECTOS)
7 # 4. PONERLE NOMBRE A LOS INDIVIDUOS Y MARCADORES
8 # 5. RANKEAR MARCADORES ASOCIADOS Y EFECTOS: EXPORTAR COMO TABLA
9 # 6. REALIZAR ANALISIS GWAS CON DISTINTAS METODOLOGIAS Y COMPARAR RESULTADOS CON NUESTROS "QTLs REALES"
10
11 # instalacion de paquetes
12 #PAQUETE PARA REALIZAR SELECCION GENOMICA MEDIANTE RRBLUP
13 install.packages("rrBLUP")
14
15
16 #CARGA DE PAQUETES
17 library(simuIMGF)
18 library(rrBLUP)
19 source("https://raw.githubusercontent.com/mngar/forest/main/rrblup.R")
20
21 #setear la semilla (VAMOS A SIMULAR LOS MISMOS DATOS QUE UTILIZAMOS EN LA PRACTICA DE GWAS)
22 set.seed(1234)
23
```

Buscar

6:16

02/11/2024

- GAPIT.Association.Manhattan\_Geno.GLM.PHENO.pdf
- GAPIT.Association.GWAS\_Results.GLM.PHENO.csv
- GAPIT.Association.Manhattan\_Geno.GLM.PHENO.csv
- GAPIT.Association.Prediction\_results.GLM.PHENO.csv
- GAPIT.Genotype.Kin\_Zhang.csv / pdf