

Introducción práctica a R y visualización de datos con ggplot2

M. en C. Luis Mario Hernández Soto



“No reinventes la rueda”

- En otras palabras, no pierdas mucho tiempo creando lo que alguien más ya hizo.
- Práctica de desarrollar soluciones o algoritmos desde cero, en lugar de utilizar soluciones existentes y bien establecidas.



“No reinventes la rueda”.

Pros:

- Entendimiento completo del problema.
- Personalización a necesidades específicas.
- Aprendizaje profundo.
- Libertad Creativa.

Contras:

- Tiempo y recursos.
- Calidad y mantenimiento dando soluciones menos eficientes.
- Reinención de problemas ya resueltos por otros.
- Omisión de aprender de soluciones existentes y mejores prácticas.

Paquetes en R

- Las expresiones (o funciones) son conjuntos de instrucciones que se utilizan para que R realice operaciones sobre objetos.
- R tiene un core básico. Con funciones como las operaciones aritméticas.
- Comunidad expande esas capacidades para hacer tareas mucho más complejas.

Paquetes

- Los usuarios han construido paquetes.
 - Diversas funciones:
 - Tareas generales (gráficas, instalación de herramientas).
 - Tareas específicas (Estadísticas particulares, implementación de modelos matemáticos).
- Los paquetes de R son un conjunto de funciones y bases de datos desarrollados por la comunidad.



Paquetes

Files

Plots

Packages

Help

Viewer

+

Install

↺

Update

🔍

🌐

	Name	Description	Version		
User Library					
<input type="checkbox"/>	AnnotationDbi	Manipulation of SQLite-based annotations in Bioconductor	1.56.2	🌐	🔍
<input type="checkbox"/>	AnnotationFilter	Facilities for Filtering Bioconductor Annotation Resources	1.18.0	🌐	🔍
<input type="checkbox"/>	ape	Analyses of Phylogenetics and Evolution	5.6-2	🌐	🔍
<input type="checkbox"/>	aplot	Decorate a 'ggplot' with Associated Information	0.1.4	🌐	🔍
<input type="checkbox"/>	askpass	Safe Password Entry for R, Git, and SSH	1.1	🌐	🔍
<input type="checkbox"/>	assertthat	Easy Pre and Post Assertions	0.2.1	🌐	🔍
<input type="checkbox"/>	babelgene	Gene Orthologs for Model Organisms in a Tidy Data Format	22.3	🌐	🔍
<input type="checkbox"/>	backports	Reimplementations of Functions Introduced Since R-3.0.0	1.4.1	🌐	🔍
<input type="checkbox"/>	base64enc	Tools for base64 encoding	0.1-3	🌐	🔍
<input type="checkbox"/>	BH	Boost C++ Header Files	1.78.0-0	🌐	🔍
<input type="checkbox"/>	Biobase	Biobase: Base functions for Bioconductor	2.54.0	🌐	🔍
<input type="checkbox"/>	BiocFileCache	Manage Files Across Sessions	2.2.1	🌐	🔍
<input type="checkbox"/>	BiocGenerics	S4 generic functions used in Bioconductor	0.40.0	🌐	🔍
<input type="checkbox"/>	BiocIO	Standard Input and Output for Bioconductor Packages	1.4.0	🌐	🔍
<input type="checkbox"/>	BiocManager	Access the Bioconductor Project Package Repository	1.30.17	🌐	🔍
<input type="checkbox"/>	BiocParallel	Bioconductor facilities for parallel evaluation	1.28.3	🌐	🔍
<input type="checkbox"/>	BiocVersion	Set the appropriate version of Bioconductor packages	3.14.0	🌐	🔍
<input type="checkbox"/>	biomaRt	Interface to BioMart databases (i.e. Ensembl)	2.50.3	🌐	🔍
<input type="checkbox"/>	Biostrings	Efficient manipulation of biological strings	2.62.0	🌐	🔍
<input type="checkbox"/>	biovizBase	Basic graphic utilities for visualization of genomic data.	1.42.0	🌐	🔍
<input type="checkbox"/>	bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.0.4	🌐	🔍

Instalación de paquetes

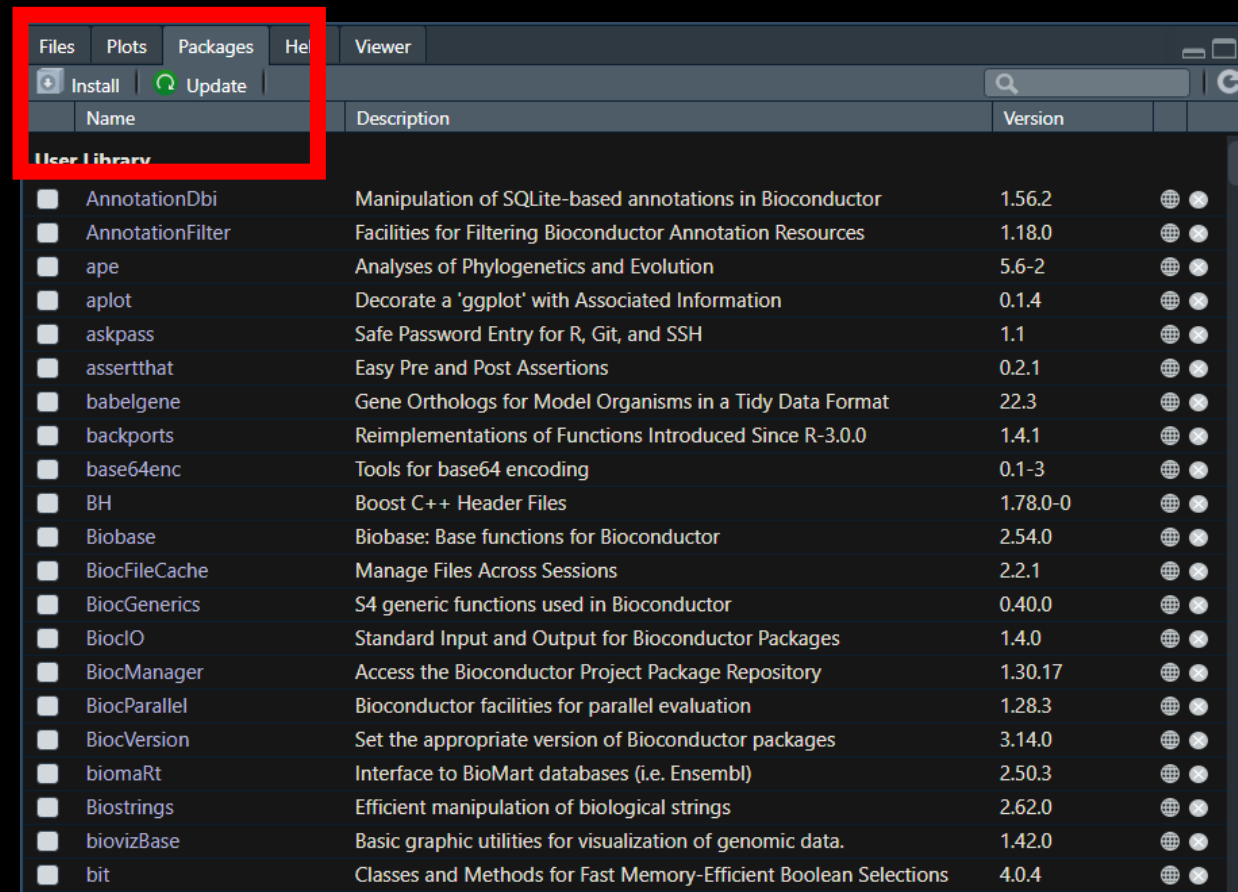
- Varias formas.
- Usar la función `install.packages()` .
 - El argumento es el nombre del paquete y se expresa entre comillas; es decir,
 - > `install.packages("nombre del paquete")`



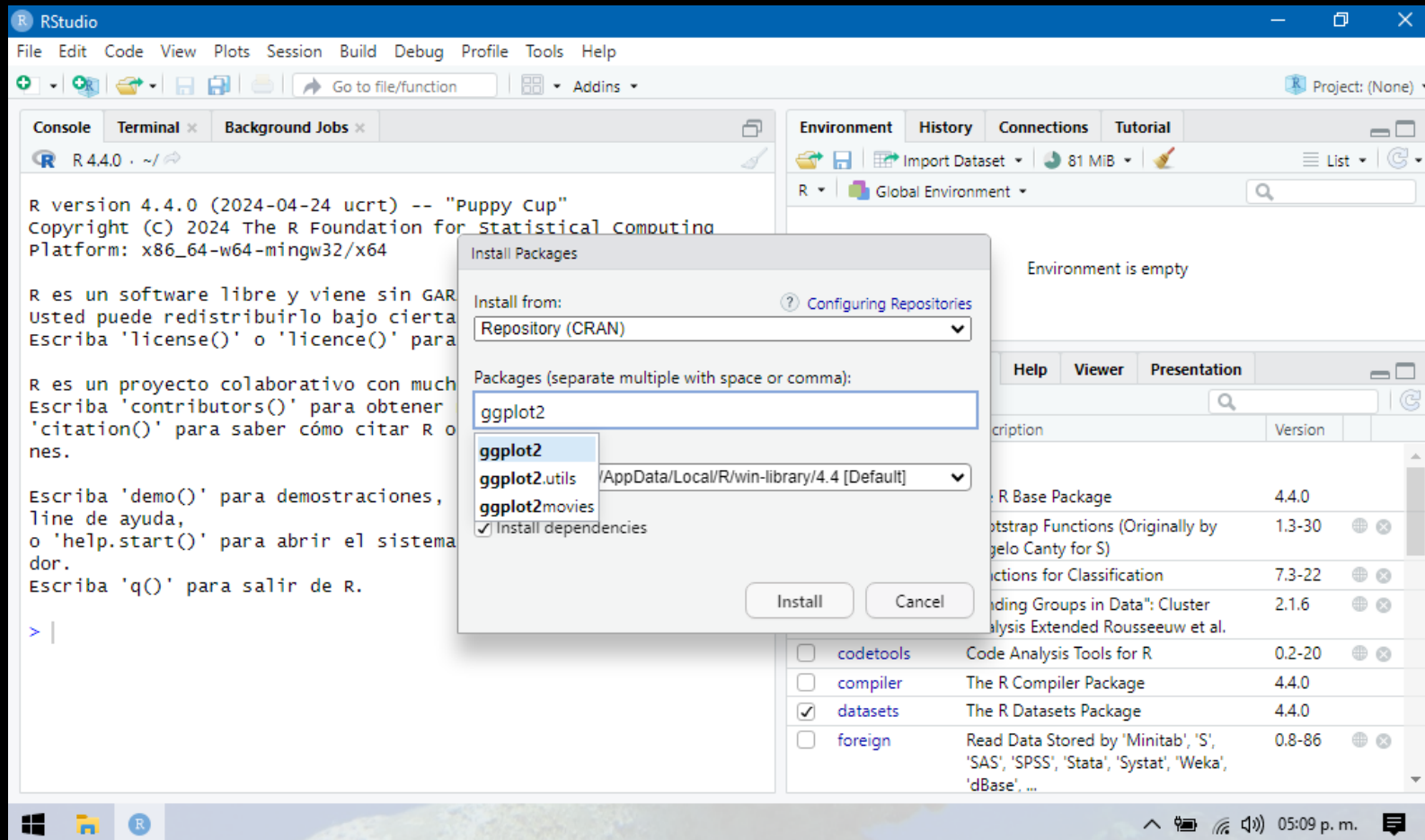
> `install.packages("ggplot2")`

Instalación de paquetes

- Alternativamente se pueden instalar desde el multipanel o desde la pestaña Tools, Install packages.

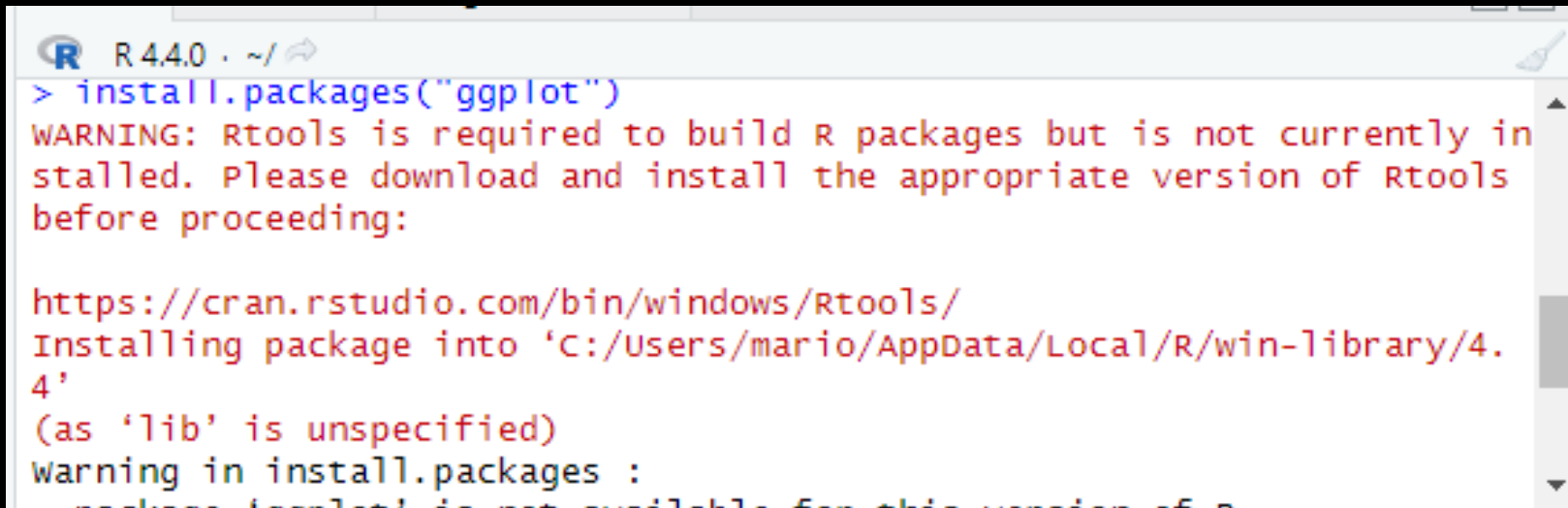


- Adicionalmente ofrece instalar **dependencias**.



Instalación de paquetes

- Esto último es particularmente útil si vemos errores como este:

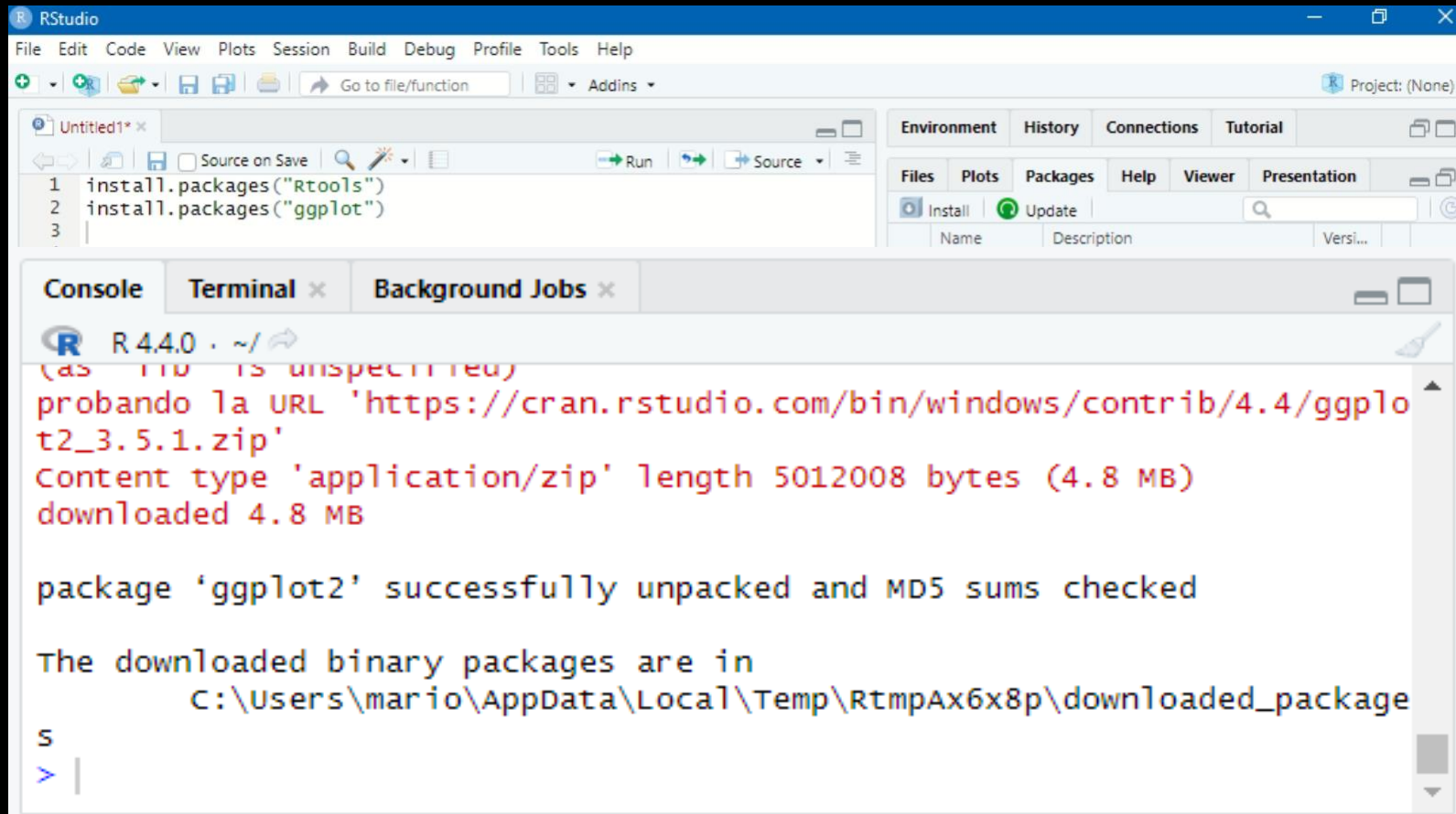


```
R 4.4.0 . ~/
> install.packages("ggplot2")
WARNING: Rtools is required to build R packages but is not currently in
stalled. Please download and install the appropriate version of Rtools
before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/mario/AppData/Local/R/win-library/4.
4'
(as 'lib' is unspecified)
warning in install.packages :
package 'ggplot2' is not available for this version of R
```

Instalación de paquetes

- Si la instalación es correcta primero mostrará el proceso y después volverá el prompt.



The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains the code `install.packages("Rtools")` and `install.packages("ggplot2")`.
- Console:** Displays the output of the installation process for `ggplot2`. The text is as follows:
`R 4.4.0 ~/
(as rtd is unspecified)
probando la URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/ggplot2_3.5.1.zip'
Content type 'application/zip' length 5012008 bytes (4.8 MB)
downloaded 4.8 MB

package 'ggplot2' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\mario\AppData\Local\Temp\RtmpAx6x8p\downloaded_packages
S
> |`
- Environment/History/Connections/Tutorial:** Tabs are visible at the top right.
- Files/Plots/Packages/Help/Viewer/Presentation:** Tabs are visible below the top right.

Library

- Funciones y material (datos de ejemplo) de un paquete de R se almacenan en una **library**.
- Ruta en computadora que aloja los paquetes deseados.
- Se cargan con:

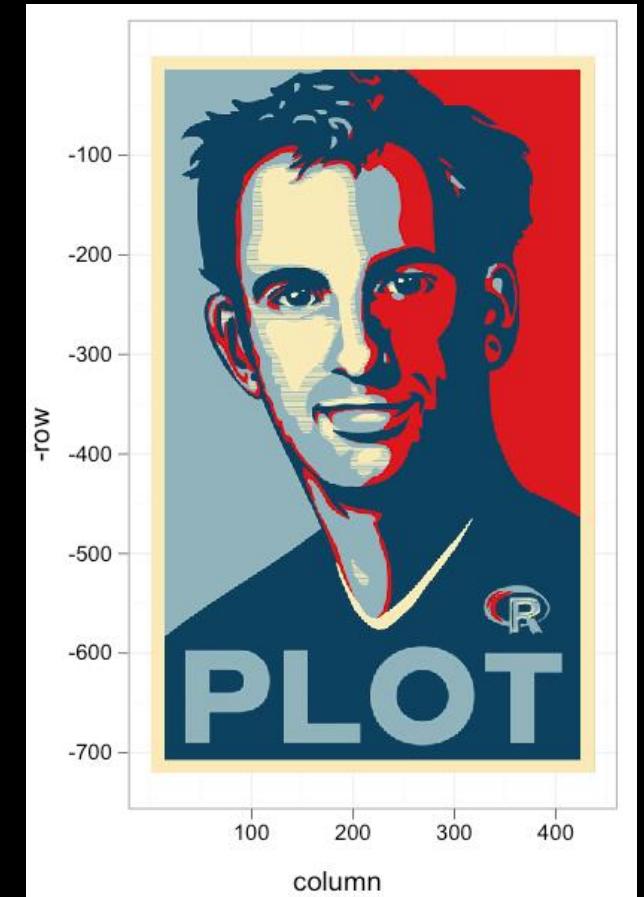
```
> library("nombre de la library")
```



```
> library("ggplot2")
```

¿Qué es ggplot2?

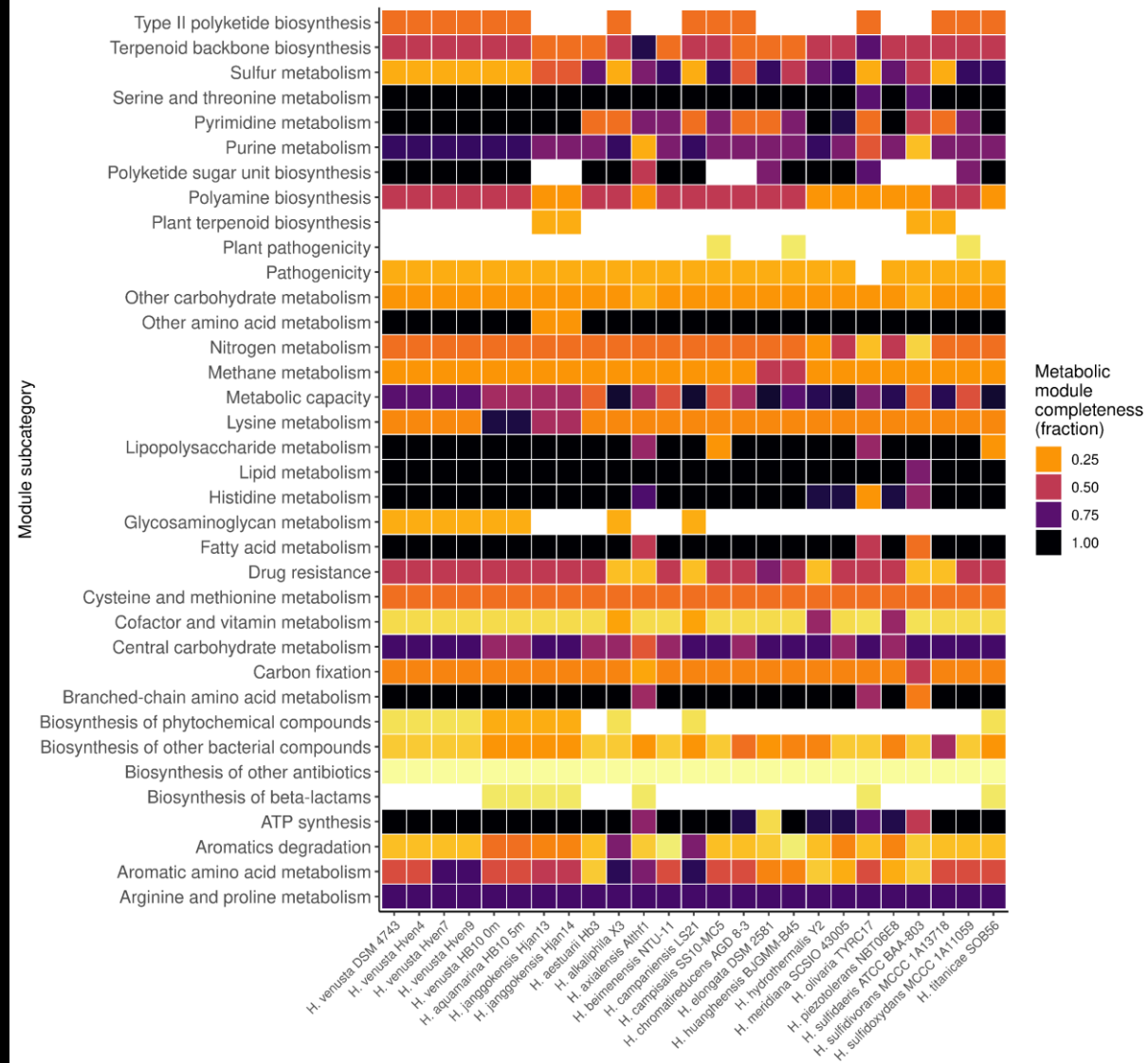
- Paquete R para generar visualizaciones de datos.
- Creado por **Hadley Wickham**.
- Gran cantidad de usuarios:
 - Su versatilidad.
 - Una interfaz clara y coherente.
 - Resultados visuales atractivos.
- Grammar of Graphics Ploting (ggplot).
 - Por la forma estructurada de describir visualizaciones de datos.
- Genera resultados visualmente atractivos y listos para publicación.



¿Qué es ggplot2?

- La gramática de los gráficos representa y abstrae ideas/objetos gráficos.
- Piense en "verbo", "sustantivo" y "adjetivo" para gráficos.
- Permite una "teoría" de gráficos sobre la cual construir nuevos gráficos y objetos gráficos.
- Acorta la distancia de la mente a la página

All strains



Geometrías

- Las geometrías son tipos de gráficos.
 - `geom_point()`, `geom_jitter()`
 - Puntos
 - `geom_line()`
 - Líneas
 - `geom_boxplot()`
 - Caja y bigotes
 - `geom_col()`, `geom_bar()`
 - Barras
 - `geom_histogram()`
 - Histogramas
 - `geom_density()`, `geom_violín()`
 - Dispersión

Estéticas

- Las estéticas (Aesthetics) son los parámetros que pueden ser ajustados en una geometría dada
 - Color
 - Colour
 - Fill
 - Opacidad
 - Alpha
 - Forma
 - Tamaño

Estructura básica



Estructura básica

- Se construye:

`ggplot(data=mpg, aes(x=cty, y=hwy)) +`
`geom_point()+`
`ggtitle("Ejemplo")+`
`xlab("Leyenda")`

Diagram annotations:

- `data.frame a graficar` points to `data=mpg`
- `Estética` points to `aes(x=cty, y=hwy)`
- `Geometría` points to `geom_point()`
- `Otras capas` points to the group containing `ggtitle("Ejemplo")` and `xlab("Leyenda")`

Aesthetic mappings

← y ← x ← color

hwy	disp	cyl	class
17	5.0	8	suv
20	2.7	4	pickup
17	4.0	6	suv
25	2.8	6	compact
27	3.1	6	compact
30	2.0	4	compact
25	2.8	6	compact
23	2.8	6	compact
26	3.0	6	midsize
17	5.4	8	pickup
28	2.5	5	subcompact
29	3.5	6	midsize
26	2.4	4	midsize
29	2.0	4	midsize
15	5.4	8	pickup
29	1.8	4	compact
18	5.7	8	suv
12	4.7	8	pickup
26	2.8	6	compact
24	3.3	6	minivan

Data

Geom

A graficar

Ejemplo1. Uso de cubrebocas

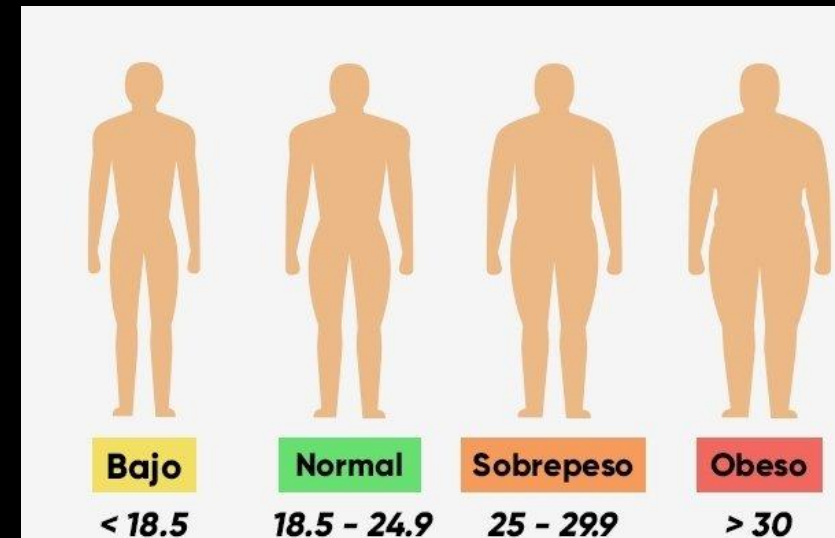
- Porque la pandemia deo muchos datos.
 - Descargue el archivo
usodecubrebocasCORRECTO.csv

Ejemplo 2. IMC

- Método de evaluación fácil y económico para la categoría de peso: bajo peso, peso saludable, sobrepeso, y obesidad*.

$$IMC = \frac{Peso\ (kg)}{Altura\ (m)^2}$$

- Descargue el archivo IMC.csv



Ejemplo 3. Extracción de ADN

- El ADN (o DNA) se puede extraer de muestras biológicas.
- Es importante:
 - Cantidad (mg/ μ L)
 - Calidad (A_{260}/A_{280} y $A_{260}/230$).
- Descargue el archivo DNA.csv

Algunos trucos extra

- Se pueden cambiar los fondos (temas)
 - `theme_gray()`, `theme_bw()`, `theme_linedraw()`, `theme_light()`,
`theme_minimal()`, `theme_classic()`, `theme_void()`
- Las etiquetas (Leyendas)
 - `theme()`, `labs()`

Los más importante...

- ¡El artículo! ¡La tesis!
- ¿Cómo se guarda todo esto?

```
ggsave("mi_grafica.png", plot=plot1)
```

Gracias por su atención

lhernandezs1212@egresado.ipn.mx

<https://www.researchgate.net/profile/Luis-Mario-Hernandez-Soto>