

Métodos Cuantitativos

II

Clase 8

Sebastián Muñoz Tapia
Mail: semunoz@uahurtado.cl

Tabla de contenidos

01

Repaso

02

Importación/Exportación
Bases de datos

03

tydiverse

04

Ejercicio Casa

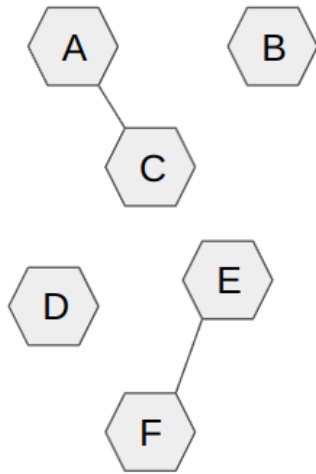


01

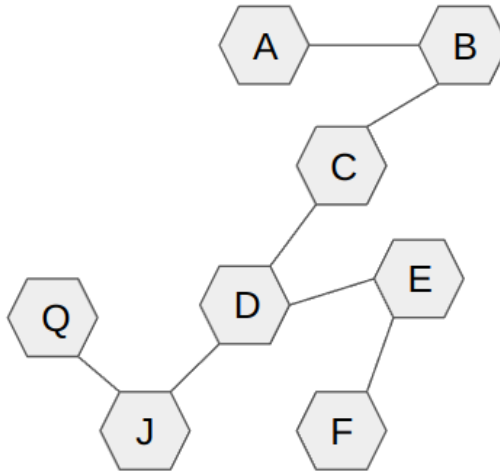
Repaso



De persona novata a experta

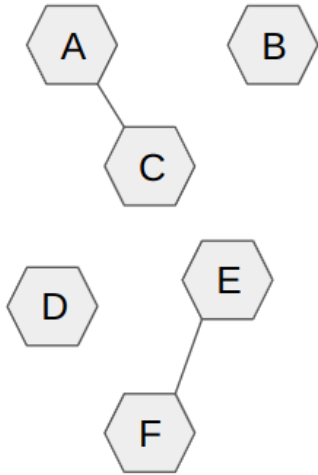


Novato/a

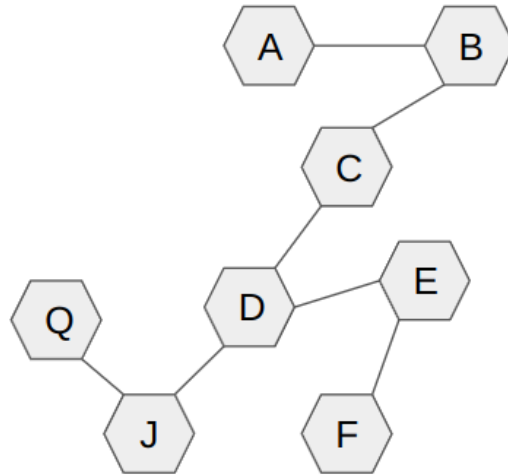


Competente

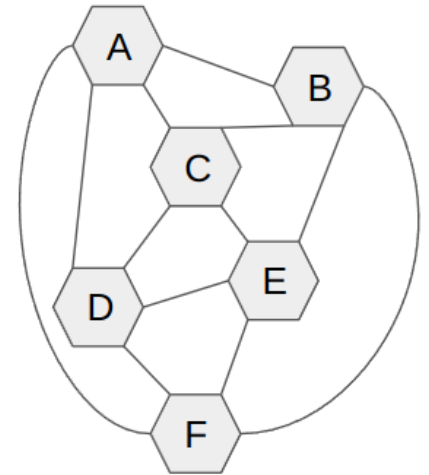
De persona novata a experta



Novato/a



Competente

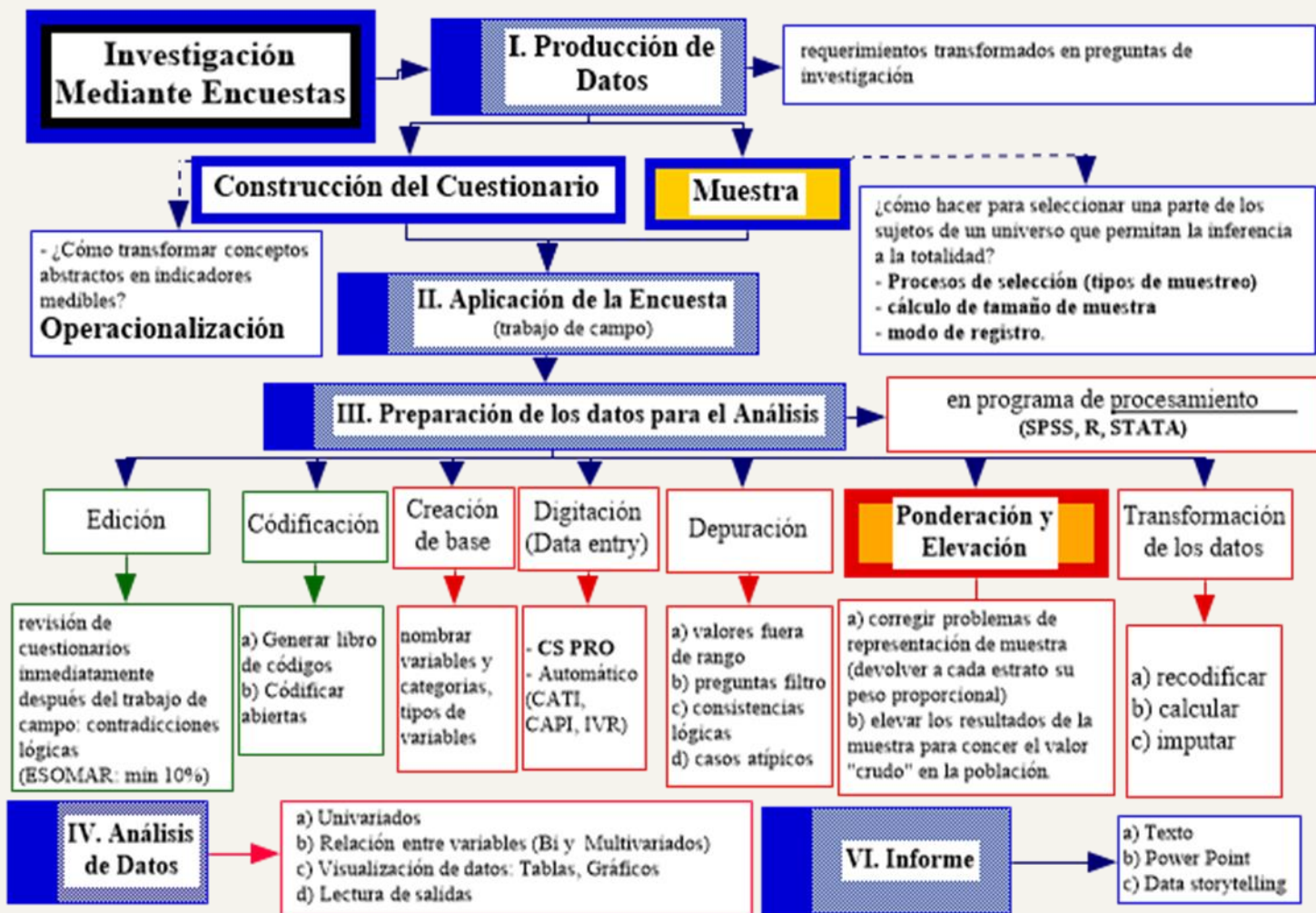


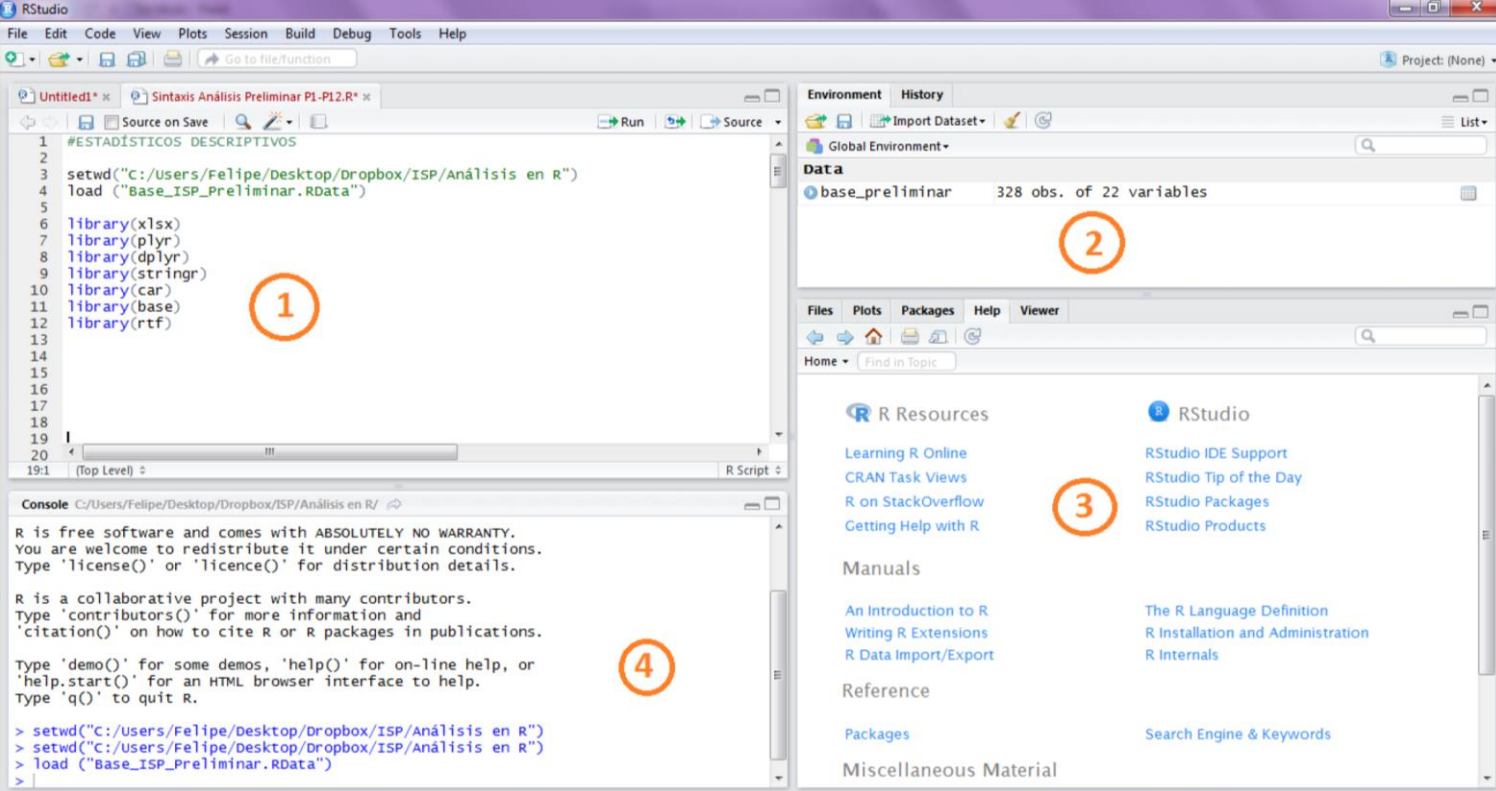
Experta/o

```
var scrollHeight =  
element.clientHeight + 0.02 * width;  
window.scroll(0, scrollHeight);  
}
```



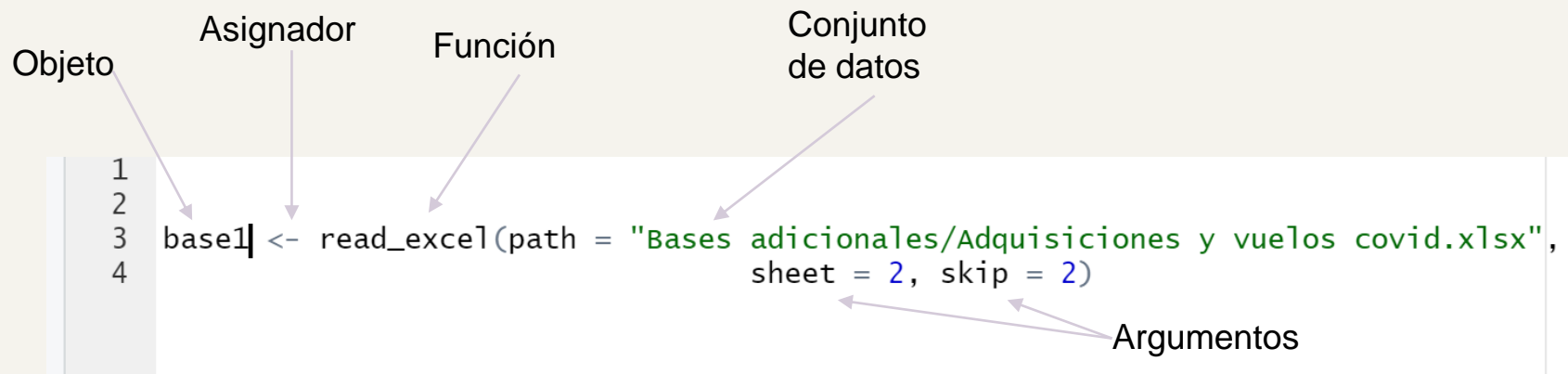






- Ventana (1): Editor de Sintaxis
- Ventana (2): “Entorno de Trabajo”
- Ventana (3): Sub-pestañas; Files (Historial de archivos), plots (Gráficos), packages (Paquetes), help (ayuda)
- Ventana (4): Visualizador de resultados.

Lógica y estructura de un comando



- A objeto le asigno el resultado de una acción/proceso (función) ...
- que se ejecutó sobre un conjunto de datos...
- considerando una serie de configuraciones (argumentos: especificaciones de la función)
- En este caso: leer una base de Excel en la pestaña dos desde la segunda fila; y asignarlo a la base1

Tipos de objetos → Vector

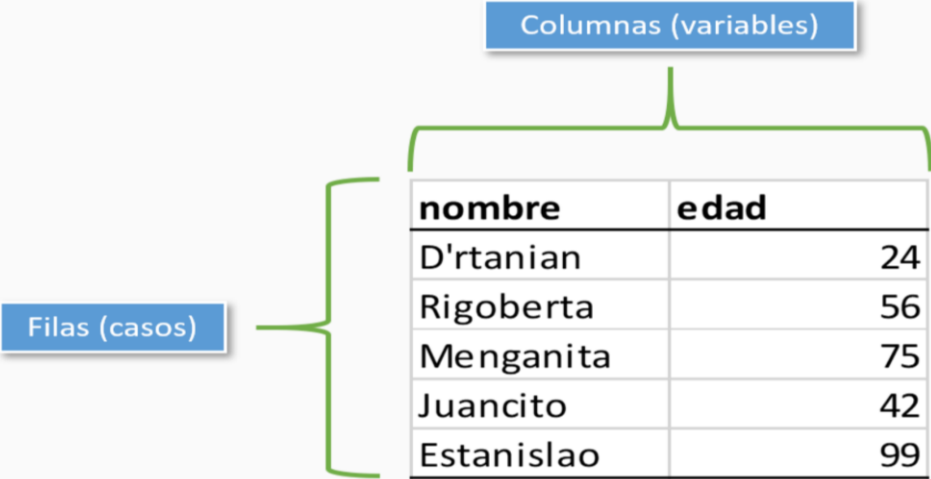
- Almacenamiento contiguo (filas o columnas), tienen una dimensión
- Puede almacenar información numérica, caracteres, valores lógicos (TRUE or FALSE)
- Ejemplo:
 - (1, 2, 4, 78, 42, 3, 65)
 - (“casa”, “auto”, “bus”, “bicicleta”)
 - (TRUE, FALSE, FALSE, TRUE)

Tipos de objetos → Factor

- Se trata de un vector que puede almacenar dos capas de información (tiene dos dimensiones): números y letras.
- Es útil para variables nominales (como género) u ordinales (muy satisfecho, satisfecho, poco satisfecho, muy insatisfecho)
- `genero <- (1, 2, 2, 2, 1, 1)`
- `factor(genero, labels = c("masculino", "femenino"))`

Tipos de objetos → Bases de datos (Data Frame)

- El elemento data.frame es lo que conocemos como una base de datos: Filas (casos) y columnas (variables) relacionadas entre sí:



The diagram illustrates the structure of a data frame. A blue box labeled 'Columnas (variables)' is positioned above the table, with a green bracket spanning the width of the table. Another blue box labeled 'Filas (casos)' is positioned to the left of the table, with a green bracket spanning the height of the table. The table itself has two columns: 'nombre' and 'edad'. The rows contain the following data:

nombre	edad
D'rtanian	24
Rigoberta	56
Menganita	75
Juancito	42
Estanislao	99
Fuente: un excel	

Paquetes

- Podemos descargarlos e instalarlos con el siguiente comando:
 - `install.packages("nombre_del_paquete")`
- Sólo es necesario instalar el paquete una vez por computadora. Se descargarán de la pagina de CRAN los archivos correspondientes, ubicándose en el directorio donde hayamos instalado el programa
- Una vez instalado el paquete, cada vez que abramos una nueva sesión de R y querramos utilizar el mismo debemos cargarlo al ambiente de trabajo mediante la siguiente función:
 - `library(nombre_del_paquete)`

- `install.packages("openxlsx")`
 - Para abrir archivos excel
- `install.packages("tidyverse")`
 - Conjunto de paquetes para limpiar, ordenar, recodificar y graficar datos
 - con comillas
- abrir librerías (cada inicio de sesión)
 - `library(openxlsx)`
 - `library(tidyverse)`

Los proyectos

- ¿Cómo garantizar que la ruta para importar bases de datos al ambiente de R y para exportar resultados pueda funcionar si muevo mis carpetas o las comparto?

```
read.csv("C:\\Users\\Guido\\Investigación\\Bases_COvid\\Base Covid19.csv")  
dir.create("C:\\Users\\Guido\\Investigación\\Resultados")  
setwd("C:\\Users\\Guido\\Cursos\\Diplomatura UNSAM\\")
```



Importación

- A la hora de importar una base de datos nos podemos llegar a enfrentar a distintos tipos de archivos. En R contamos con distintos paquetes y funciones según el tipo de extensión del archivo:

Tipo de archivo	Paquete	Extension	Funciones
Texto Plano	readr	.csv	read_csv()
Texto Plano	readr	.txt	read_txt()
Texto Plano	readr	.tsv	read_tsv()
Extension de R	RBase	.RDS	readRDS()
Extension de R	RBase	.RDATA	open()
Otros Softwares	haven	.dta	read_dta()
Otros Softwares	haven	.sav	read_spss()
Excel	openxlsx	.xlsx	read.xlsx()
Excel	readxl	.xls	read_excel()

Comma
Separated
Values

STATA

Importación de resultados

Por lo general, cada paquete que presenta funciones para importar bases de dato, tiene como complemento una función para exportar (guardar en el disco de nuestra PC) un objeto con la misma extensión.

Ejemplos:

- el paquete **openxlsx** tiene una función denominada **write.xlsx()** que nos permite exportar un dataframe creando un archivo **.xlsx**

- En RBase la función **saveRDS()** nos permite exportar archivos de extensión **.RDS** (son menos pesados para trabajarlos luego desde R)

En general estas funciones tienen un primer parametro para especificar el objeto a exportar, y un segundo para especificar la ruta y el nombre de archivo a crear ()

```
write.xlsx(x = objeto_resultados,file = "Resultados/cuadro1.xlsx")
```

```
saveRDS(object = objeto_resultados,file = "Resultados/base_nueva.RDS")
```

Lenguaje R

```
variable1 <- c(1:3)
```

```
variable2 <- c("juan", "pedro", "ale")
```

```
base <- data.frame(variable1, variable2)
```



Otras funciones:

class(nuevo_objeto): observar tipo de objeto

sum (): sumatoria de valores al interior de objeto

NA [sin comillas]: valores perdidos

ls (): lista los objetos del entorno

rm (): borra objetos específicos del entorno

rm (list =ls ()): borra todos los objetos del entorno

Una las funciones con su explicación para explorar data.frame

- **dim(base)**
- **summary(base**
- **names(base)**
- **unique(base\$variable)**
- **table(base\$variable)**
- **base_sample <- sample_n(tbl = base_a_muestrar, size = cantidad_de_casos)**
- frecuencias de variable en particular
- nombre de columnas
- conjunto resumen de estadísticos
- categorías de variable en particular
- muestra aleatoria simple de n casos
- cantidad de filas y columnas)

Para que sirve el \$ cuando estoy con una base de datos

- `base$variable`
- `class(base$variable)`
- `table(base$variable)`
- `mean(base$variable)`
- Observar el tipo de variable específica en un `data.frame` (si es “numeric”, “character”, “factor”)
- Entrega valor de columna/variable entera
- Realizar una media en variable específica del `data.frame`
- Realizar una tabla de una variable específica del `data.frame`

magrittr

Sin EL PIPE:

```
table(base_covid$sexo)
--
F          M      NR
93963 86742 1975
```

Con EL PIPE

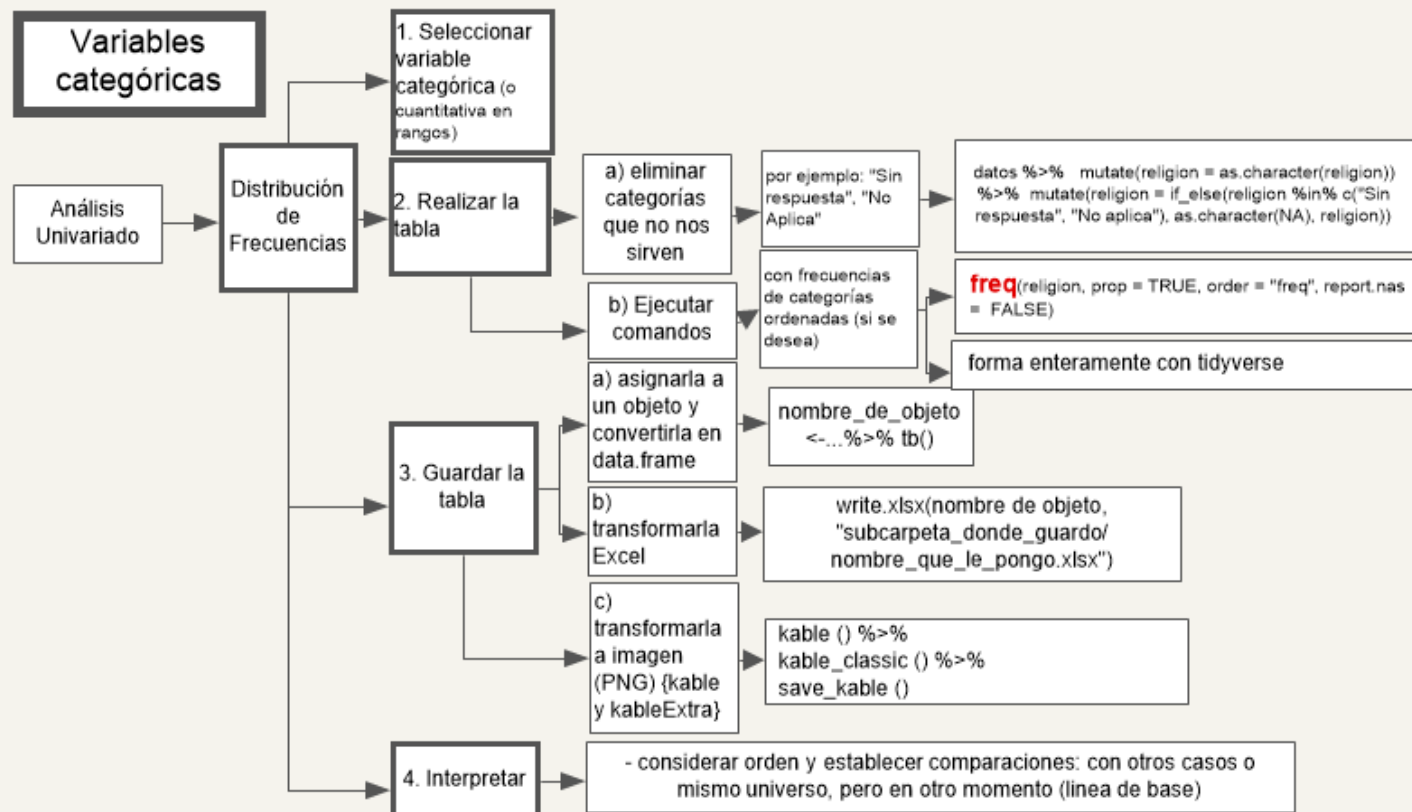
```
base_covid$sexo %>%
table()
--
.
F M NR
93963 86742 1975
```

dplyr

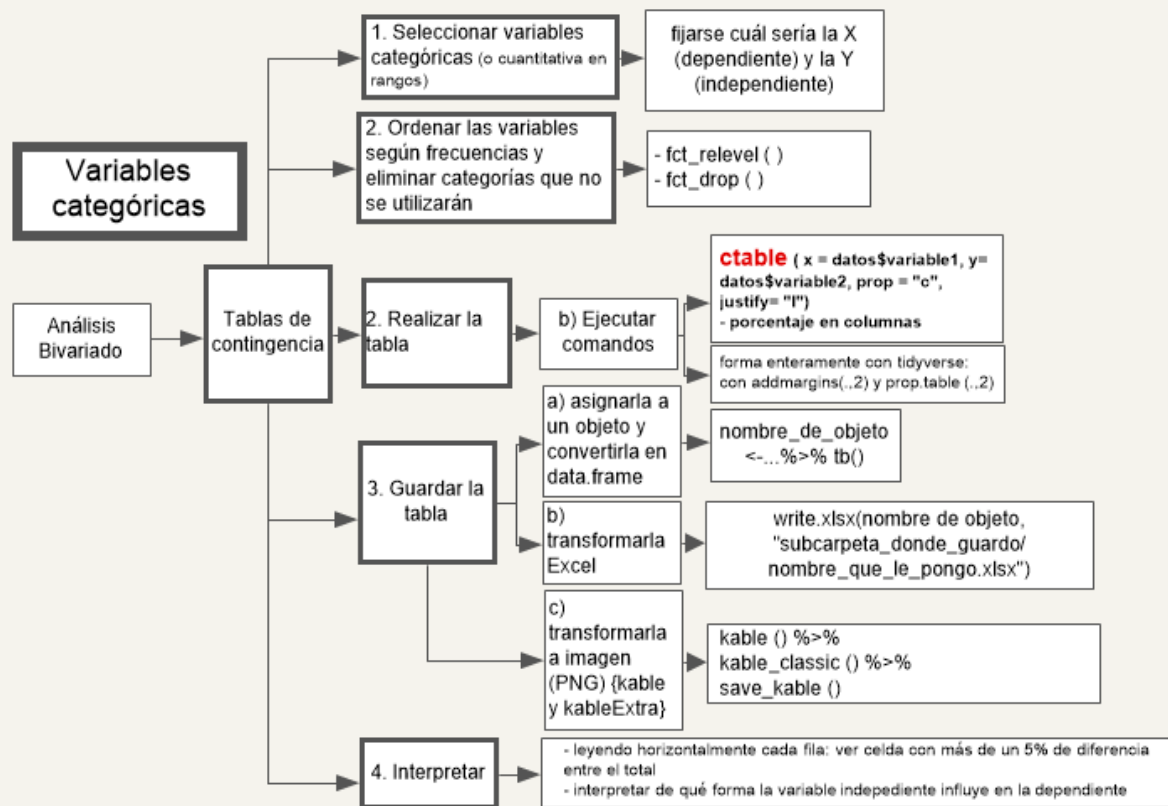
Función	Acción
<code>select()</code>	<i>selecciona o descarta variables</i>
<code>filter()</code>	<i>selecciona filas</i>
<code>mutate()</code>	<i>crea / edita variables</i>
<code>rename()</code>	<i>renombra variables</i>
<code>group_by()</code>	<i>segmenta en funcion de una variable</i>
<code>summarize()</code>	<i>genera una tabla de resúmen</i>

Análisis de datos cualitativos

Distribución de frecuencias

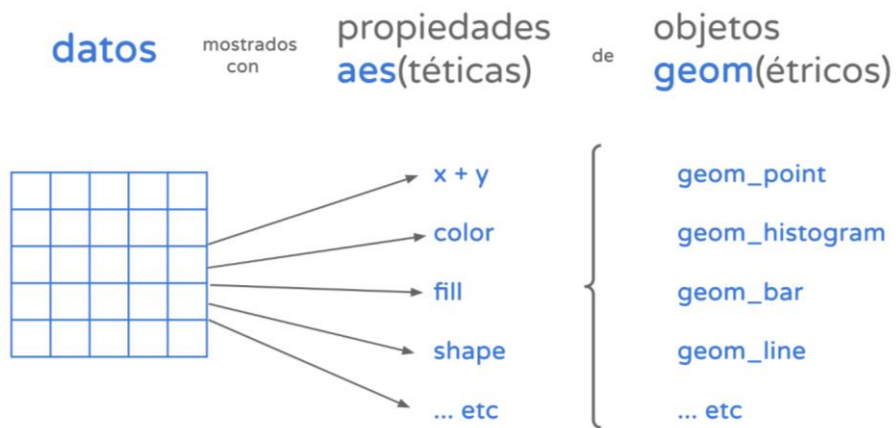


Tablas de contingencia

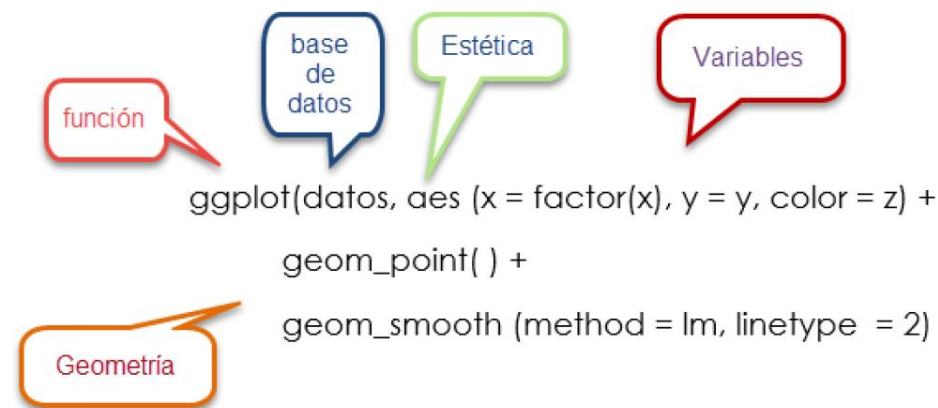




ggplot2



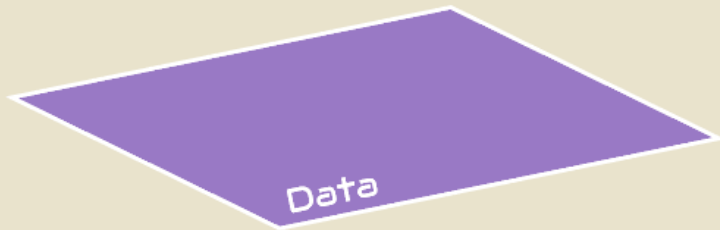
- ¿Cómo funciona esto en la práctica?
- El caso de uso más simple de ggplot consiste en:
 - una llamada a la función `ggplot()`, pasándole un dataset y una “asignación de atributos estéticos” (aesthetic mapping en inglés) usando `aes()`
 - al menos una capa “geom”, que define el recurso gráfico que mostrará los datos; por ejemplo `geom_line()` para dibujar líneas o `geom_point()` para dibujar puntos.



Capa	Descripción
Datos	Conjunto de información que se representará de manera gráfica. En nuestro caso se trata de una o más variables, o una base de datos.
Estética	Escala en la cual se posicionará la información de interés. Refiere al posicionamiento de la información a representar sobre los diferentes ejes y dimensiones del gráfico resultante. Hablamos del posicionamiento de variables en los ejes X e Y como también de la posibilidad de indicar variables que pueden ser posicionadas como color de relleno dentro de los diferentes ejes, como una función de transparencia, etc.
Geometría	Formas, elementos visuales, que se emplearán para representar visualmente la información ya consignada en los <i>datos</i> y ubicada en las diferentes posiciones del gráfico mencionadas en la <i>estética</i> . ⁴ Cada especificación de geometría permite visualizar diferentes características de la(s) variable(s) y su distribución.

Grammar of Graphics

xy, 3902, 29, 9,
4756, x, 72, 633,
647, 617, 827, 3,
1, 21, 45, tyu, 6,
987, 457, 283, 8,
4, 5, 671, 34, 67,
x, 981, hu, 89, 5



siguiente paso: ¡manos a la obra!

```
32: #----- Visualising your boxplot -----
33:
34: # Before plotting if not installed install.packages("ggplot2")
35: # Then activate ggplot2 package
36: library(ggplot2)
37:
38: # Create new variable for plot only x and y axis. ('data' and 'aesthetics' layer)
39: plot <- ggplot(data=new_data, aes(x=Genre, y=Gross..US))
40:
41: # Create new variable with geometries layer
42: q <- plot + geom_jitter(aes(fill=Studio, size=Budget..mill.),
43:                        shape = 21, # this will shape a border around data points.
44:                        colour = "black") + # with the border color of black.
45:   geom_boxplot(alpha=0.7, outlier.color = NA) # places the boxplot on the data points
46:                                           # and removes boxplot layer outliers.
47: |
48: # Change axis and title if needed.
49: q <- q +
50:   xlab("Genre") +
51:   ylab("Gross $ US") +
52:   ggtitle("Domestic Gross $ by Genre")
53:
54: # Make your plot visually attractive and readable with the 'theme' function. (Theme layer)
55: q = theme(axis.title = element_text(colour = "blue", size = 14),
56:           axis.text = element_text(size = 12),
57:           legend.title = element_text(size = 12),
58:           legend.text = element_text(size = 10),
59:           plot.title = element_text(size = 14, hjust = 0.5), # 'hjust' will center your text.
60:           panel.background = element_rect(fill = "#B3CC99"))
61:
```