



Universidad del  
**Rosario**

Facultad de  
Economía

Laboratorio de  
Finanzas

Escuela de Ingeniería,  
Ciencia y Tecnología

# TALLER DE **DATA SCIENCE EN R**



Sueño**SER**

Potencializa tu perfil  
académico y profesional.

6:00 p.m a 8:00 p.m

**10** al **12** de octubre

Trasciende,  
construye  
y Lidera

Rosarios  
con  
*Propósito*

# Contenido

- Introducción
- Algunos conceptos preliminares
- Operaciones básicas de transformaciones
- Tubos de datos (pipes)



# Introducción

# Sobre los datos a analizar

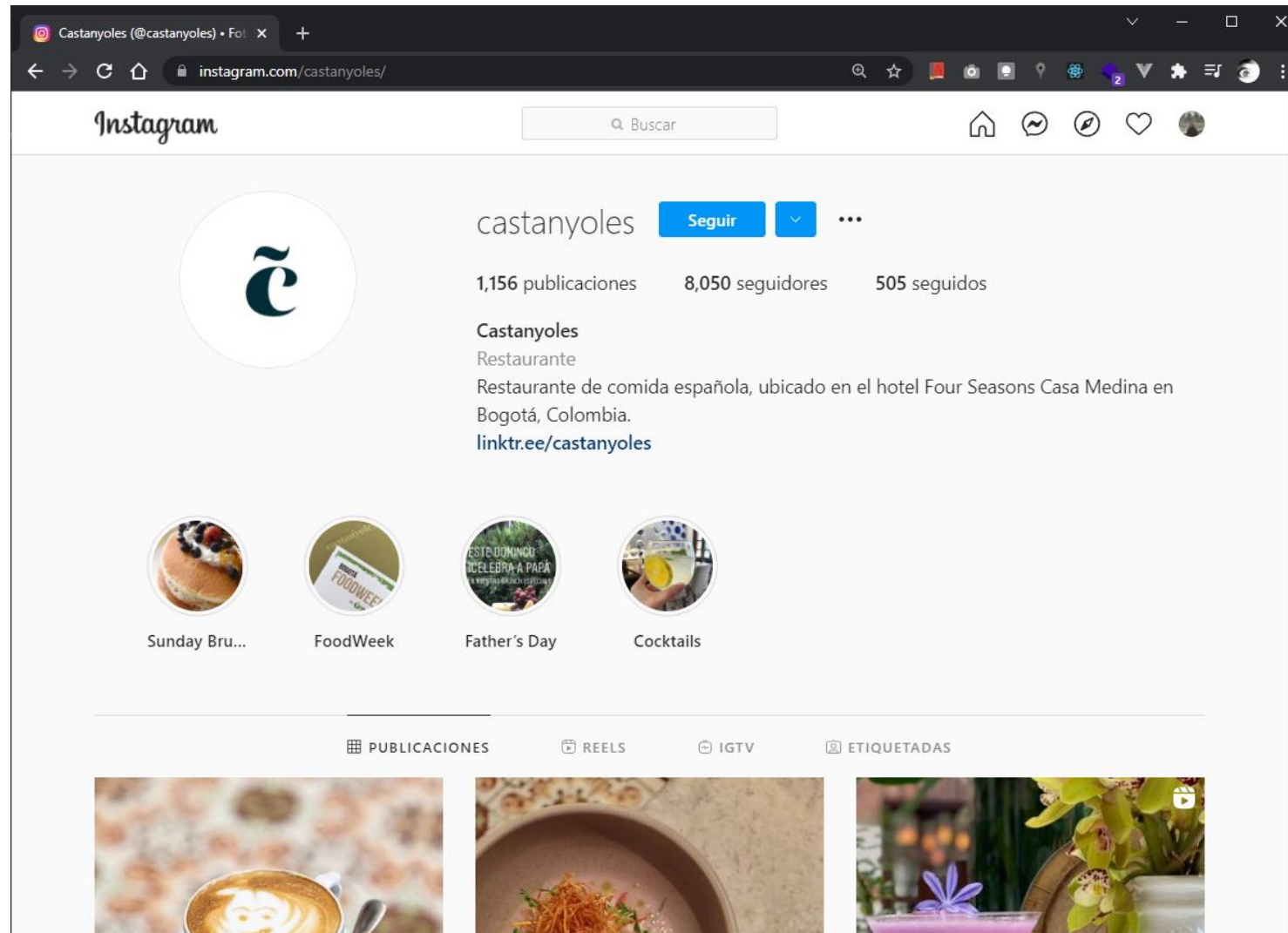
Es raro que vengan exactamente como los necesitamos

Algunas operaciones que se hacen con frecuencia:

- Cambiarle el nombre a las variables
- Reordenar los datos
- Crear datos a partir de los existentes
- Hacer resumen de los datos

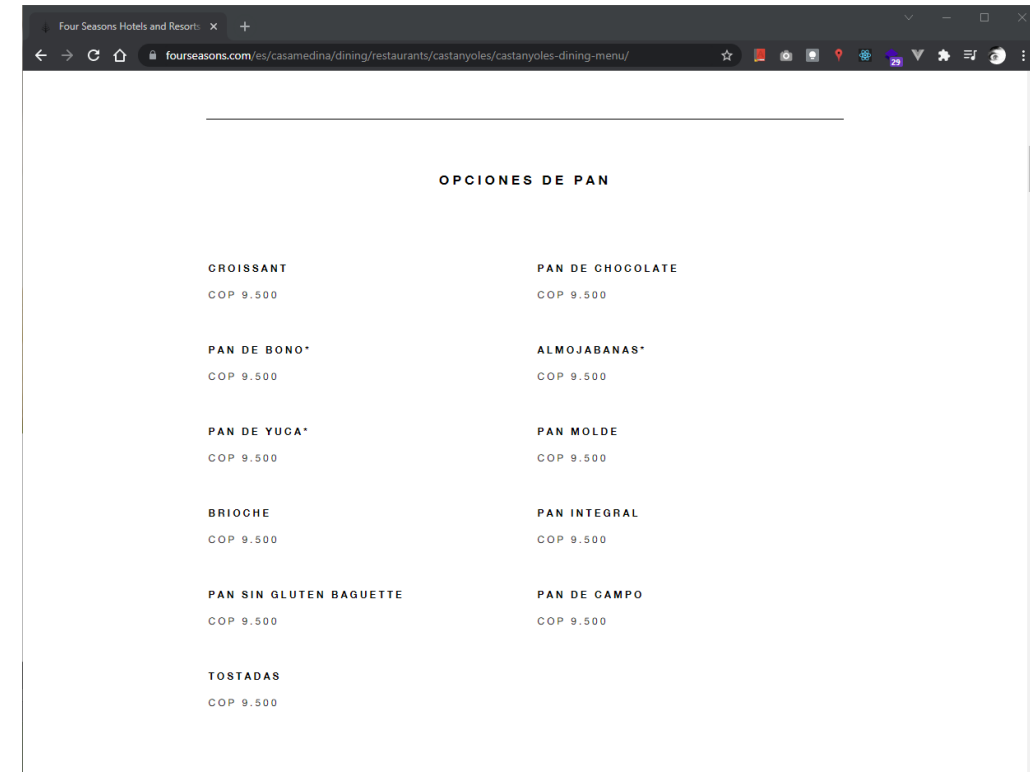
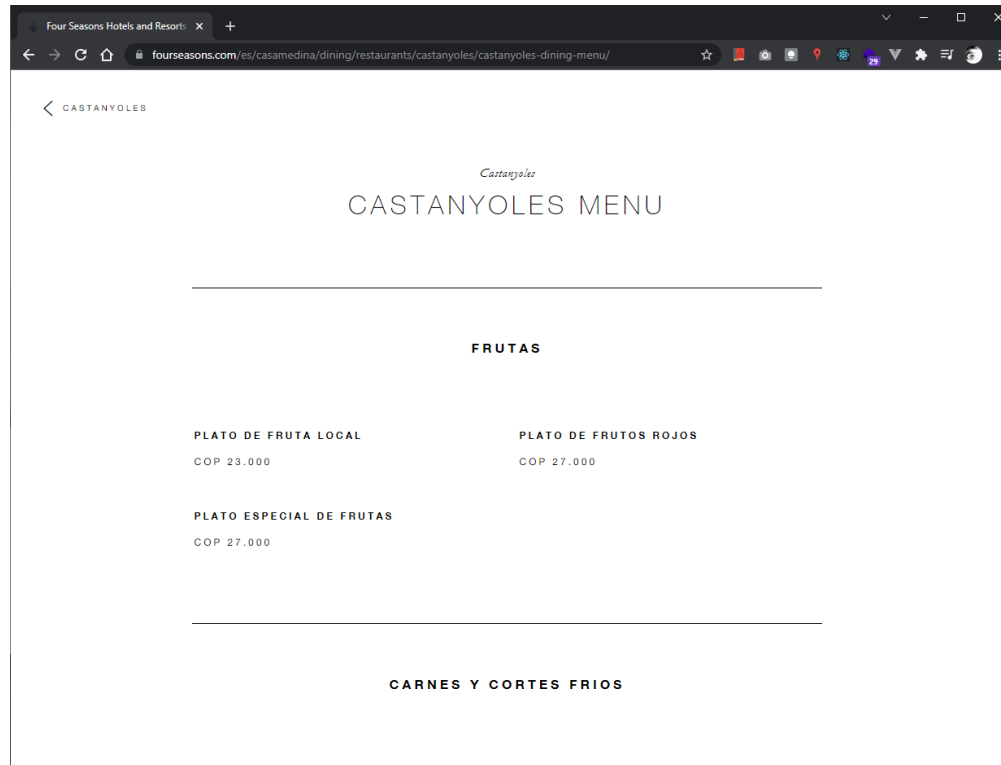
# El restaurante

## Introducción



# El menú

## Introducción



<https://www.fourseasons.com/es/casamedina/dining/restaurants/castanyoles/castanyoles-dining-menu/>

# Una noche en el restaurante

## Introducción



Los datos que se usan:

- Menú

Nombre del producto, categoría y precio

- Reservas

Nombre de la persona que reserva, datos de contacto

- Consumo en cada reserva

Nombre de la persona que reserva, nombre de cada producto, cantidad



# Algunas preguntas

■ ¿Cuántas especialidades tiene el restaurante?

ESPECIALIDADES DE CASTAÑOLES

HUEVOS BENEDICTINOS - TOCINO COP \$ 45.000	HUEVOS BENEDICTINOS - SALMÓN COP \$ 57.000
CALDO DE COSTILLAS COP \$ 23.000	CALENTAO COLOMBIANO COP \$ 38.000
DOS HUEVOS AL GUSTO COP \$ 38.000	PANCAKES COP \$36.000
WAFFLES COP \$38.000	TOSTADAS FRANCESAS COP \$36.000

■ ¿Cuánto cuesta en promedio una bebida?

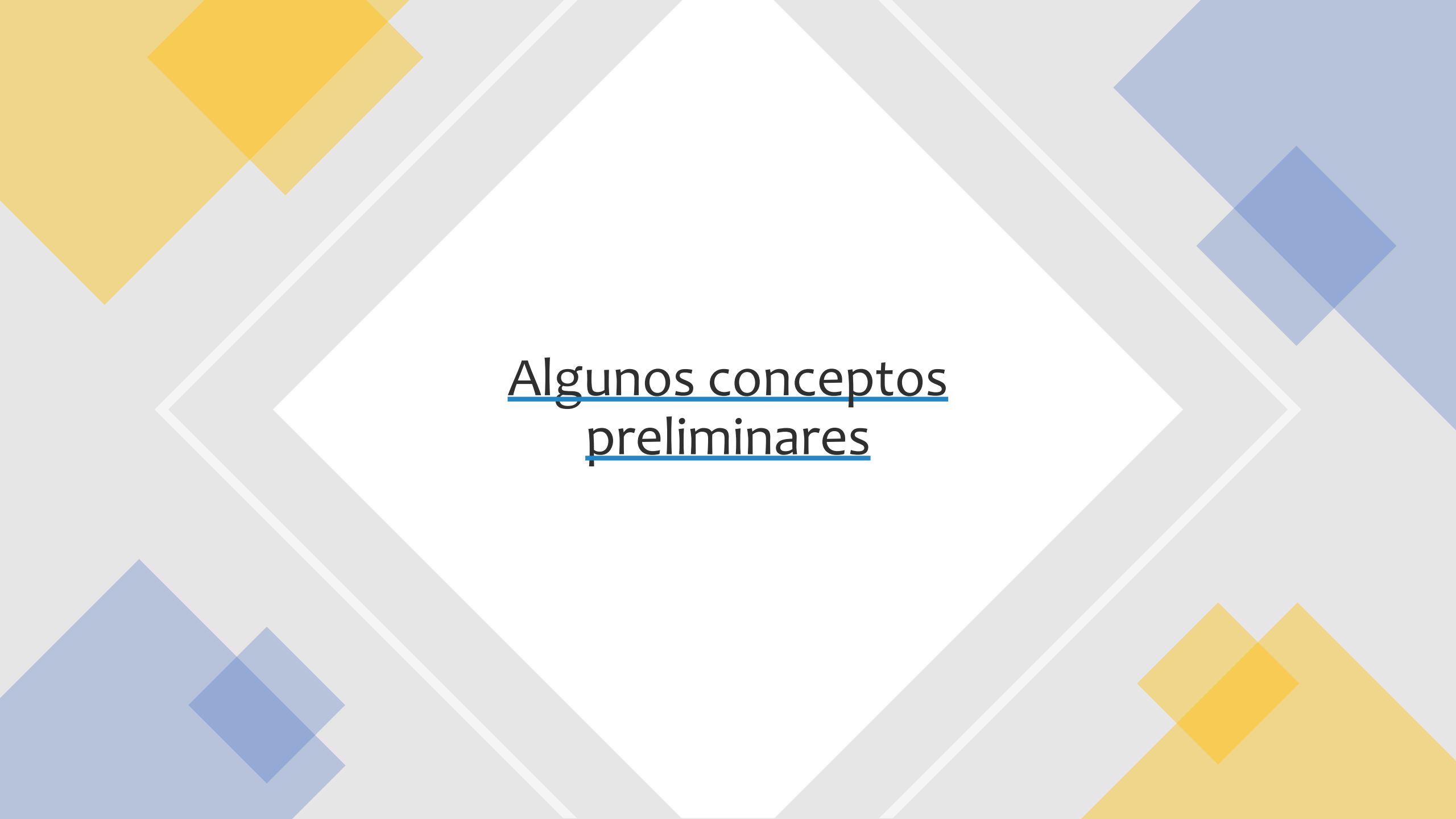
BEBIDAS

AROMÁTICA COP 7.000	TINTO COP 7.000
CAFÉ AMERICANO COP 7.000	ESPRESSO COP 7.000
SENCILLO COP \$7.500	ESPRESSO DOBLE COP 10.500
CAPPUCCINO COP 7.500	MACHIATTO COP 7.500
CAFÉ LATTE COP 8.500	CHOCOLATE COP 8.500
MILO CALIENTE COP 8.500	MILO FRIO COP 9.500
MOCCA COP 7.500	TÉ COP 7.500
ICE COFFEE COP 10.000	



# Algunas preguntas

- ¿Cuánto se recaudó anoche en total en el restaurante?
- ¿Cuánto pagó cada una de las reservas?
- ¿Cuántos diferentes platos ordenó cada reserva?
- ¿Cuál reserva pagó más y cuál pagó menos?
- ...



# Algunos conceptos preliminares

# Programas y paquetes a usar

## ■ Librería tidyverse

```
# install.packages("tidyverse")  
library(tidyverse)
```

```
> library(tidyverse)  
-- Attaching packages ----- tidyverse 1.3.1 --  
v ggplot2 3.3.3      v purrr   0.3.4  
v tibble  3.1.1      v dplyr   1.0.5  
v tidyr   1.1.3      v stringr 1.4.0  
v readr   1.4.0      v forcats 0.5.1  
-- Conflicts ----- tidyverse_conflicts() --  
x dplyr::filter() masks stats::filter()  
x dplyr::lag()    masks stats::lag()
```

## ■ Un paquete para carga de datos

```
library(readxl)
```

# La estructura dataframe

## Algunos conceptos preliminares

```
x <- c(25,0,75,200)
y <- c(11,0,53,300)
forma <- c("circulo", "circulo", "cuadrado", "cuadrado")

datos <- data.frame(x,y,forma)

View(datos)
```

x	y	forma
25	11	circulo
0	0	circulo
75	53	cuadrado
200	300	cuadrado

# Almacenamiento en Excel

## Algunos conceptos preliminares

### BEBIDAS

#### AROMÁTICA

COP 7.000

#### CAFÉ AMERICANO

COP 7.000

#### SENCILLO

COP \$7.500

#### CAPPUCCINO

COP 7.500

#### CAFÉ LATTE

COP 8.500

#### TINTO

COP 7.000

#### ESPRESSO

COP 7.000

#### ESPRESSO DOBLE

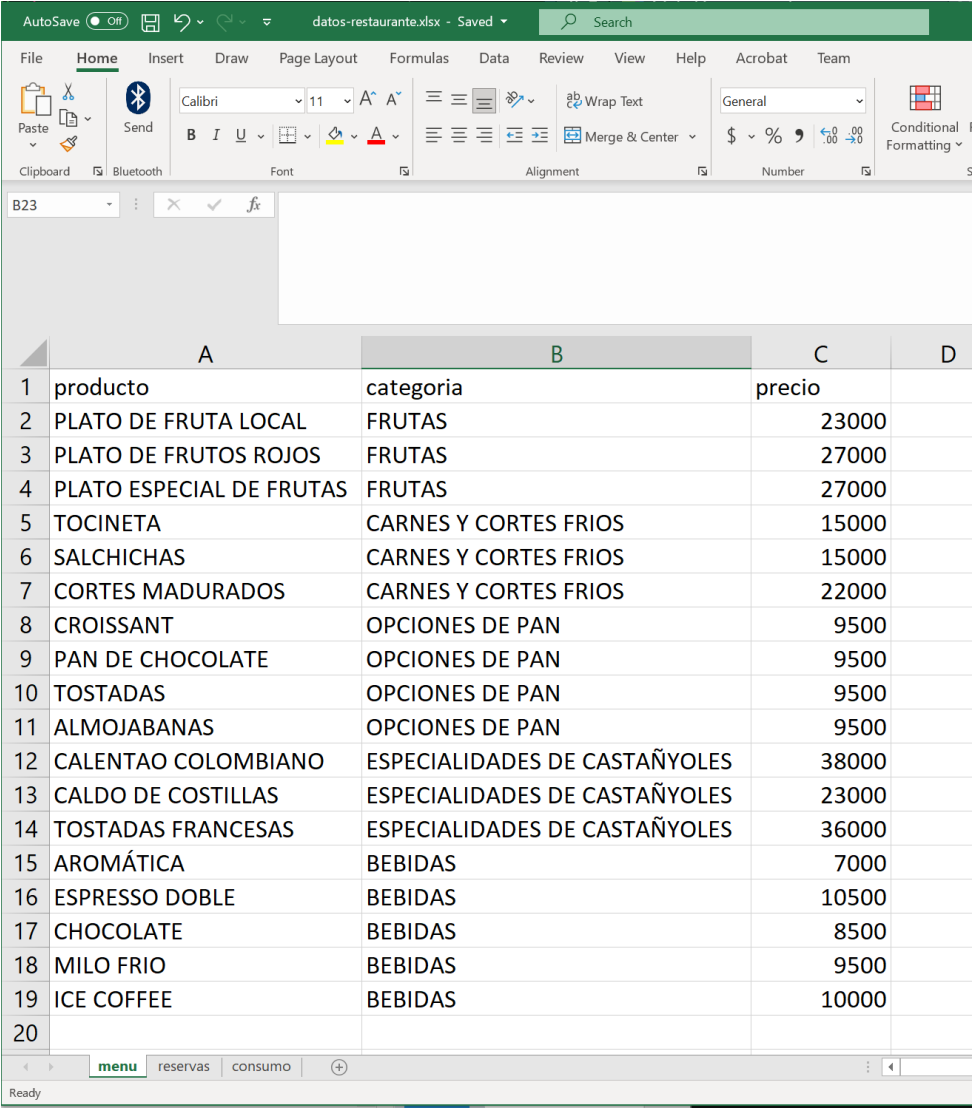
COP 10.500

#### MACHIATTO

COP 7.500

#### CHOCOLATE

COP 8.500



	A	B	C	D
1	producto	categoria	precio	
2	PLATO DE FRUTA LOCAL	FRUTAS	23000	
3	PLATO DE FRUTOS ROJOS	FRUTAS	27000	
4	PLATO ESPECIAL DE FRUTAS	FRUTAS	27000	
5	TOCINETA	CARNES Y CORTES FRIOS	15000	
6	SALCHICHAS	CARNES Y CORTES FRIOS	15000	
7	CORTES MADURADOS	CARNES Y CORTES FRIOS	22000	
8	CROISSANT	OPCIONES DE PAN	9500	
9	PAN DE CHOCOLATE	OPCIONES DE PAN	9500	
10	TOSTADAS	OPCIONES DE PAN	9500	
11	ALMOJABANAS	OPCIONES DE PAN	9500	
12	CALENTAO COLOMBIANO	ESPECIALIDADES DE CASTAÑOLES	38000	
13	CALDO DE COSTILLAS	ESPECIALIDADES DE CASTAÑOLES	23000	
14	TOSTADAS FRANCESAS	ESPECIALIDADES DE CASTAÑOLES	36000	
15	AROMÁTICA	BEBIDAS	7000	
16	ESPRESSO DOBLE	BEBIDAS	10500	
17	CHOCOLATE	BEBIDAS	8500	
18	MILO FRIO	BEBIDAS	9500	
19	ICE COFFEE	BEBIDAS	10000	
20				

# Carga de archivo de Excel en R

Algunos conceptos preliminares

## ■ Datos

<https://docs.google.com/spreadsheets/d/1oYPw8v3eDo5oCuPgjPpk7oxhMCqoPsKj/edit?usp=sharing&oid=116142505152764245433&rtpof=true&sd=true>

```
library(tidyverse)
library(readxl)

menu <- read_xlsx("datos-restaurante.xlsx", sheet = "menu")
View(menu)
```

## ■ Los datos cargados quedan en formato dataframe

# Algunos tipos de datos

- `int`: números enteros
- `dbl`: Números reales
- `chr`: Vectores de caracteres
- `dtm`: Fechas y horas
- `lgl`: Lógicas (tipos de datos que son CIERTO o FALSO)
- ...





# Operaciones básicas de transformaciones

# Desafíos de manipulación de datos

- Elegir observaciones por sus valores.
- Reordenar las filas.
- Elegir las variables por sus nombres.
- Crear nuevas variables con funciones de variables existentes.
- Contraer muchos valores en un solo resumen.

## Funciones principales de dplyr

- Elegir observaciones por sus valores (`filter()`)
- Reordenar las filas (`arrange()`)
- Elegir las variables por sus nombres (`select()`)
- Crear nuevas variables con funciones de variables existentes (`mutate()`)
- Contraer muchos valores en un solo resumen (`summarize()`)

# Estructura

- El primer parámetro es un dataframe
- Los siguientes parámetros dicen qué hacer con el dataframe
- Siempre retornan un dataframe
- Nunca modifican el dataframe de entrada

# dataframe de ensayo

## Operaciones básicas de transformaciones

```
df <- data.frame(  
  color = c("blue", "black", "blue", "blue",  
            "black"),  
  value = 1:5  
)
```



df

color	value
blue	1
black	2
blue	3
blue	4
black	5

# filter()

Operaciones básicas de transformaciones

df

color	value
blue	1
black	2
blue	3
blue	4
black	5

→

color	value
blue	1
blue	3
blue	4

```
filter(df, color == "blue")
```

- ACTIVIDAD: Filtrar el menú para ver únicamente las bebidas

`filter()`

Operaciones básicas de transformaciones

df

color	value
blue	1
black	2
blue	3
blue	4
black	5

→

color	value
blue	1
blue	3
blue	4

```
filter(df, value %in% c(1, 4))
```

- ACTIVIDAD: Filtrar el menú para ver únicamente las BEBIDAS y las FRUTAS



`filter()`

Operaciones básicas de transformaciones

df

color	value
blue	1
black	2
blue	3
blue	4
black	5

→

color	value
blue	1
blue	3
blue	4

```
filter(df, value %in% c(1, 4))
```

- ACTIVIDAD: Filtrar el menú para ver únicamente las BEBIDAS y las FRUTAS

# Operadores de comparación

Operaciones básicas de transformaciones

Operador relacional en R	Descripción
>	Mayor que
<	Menor que
>=	Mayor o igual que
<=	Menor o igual que
==	Igual a
!=	Distinto a



```
3 > 5 # TRUE
3 < 5 # FALSE
3 >= 5 # FALSE
3 <= 5 # TRUE
3 == 5 # FALSE
3 != 5 # TRUE
```

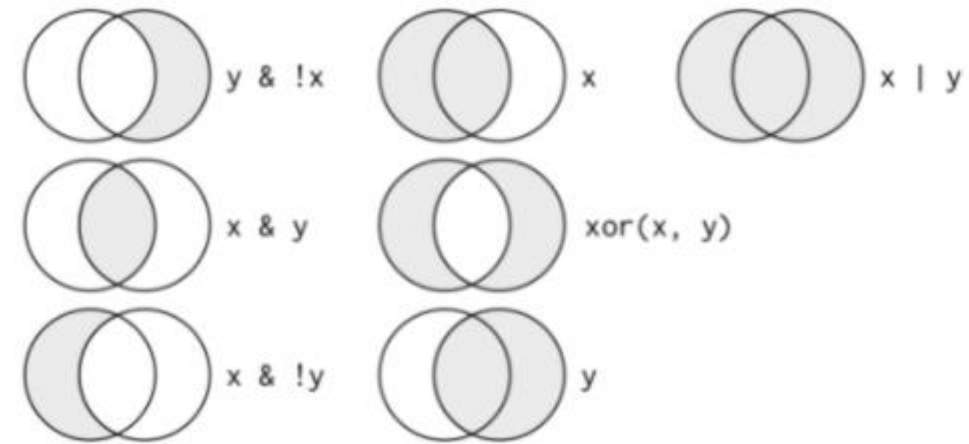
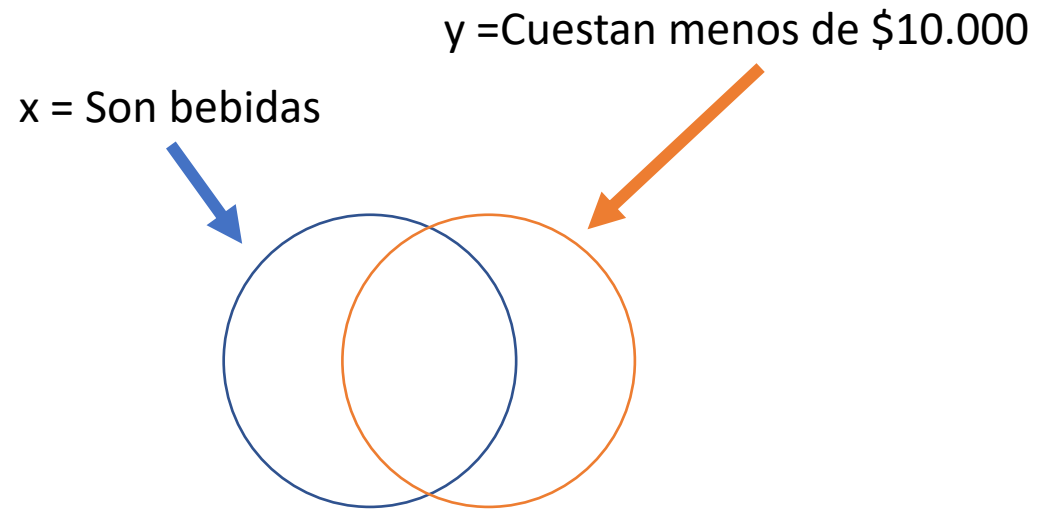
# Operadores lógicos

## Operaciones básicas de transformaciones

Operador	Comparación	Ejemplo	Resultado
<code>x   y</code>	x Ó y es verdadero	<code>TRUE   FALSE</code>	<code>TRUE</code>
<code>x &amp; y</code>	x Y y son verdaderos	<code>TRUE &amp; FALSE</code>	<code>FALSE</code>
<code>!x</code>	x no es verdadero (negación)	<code>!TRUE</code>	<code>FALSE</code>

# Condiciones complejas

## Operaciones básicas de transformaciones



- ACTIVIDAD: Filtrar el menú para ver únicamente las BEBIDAS o las opciones del menú que cuestan menos de \$10.000

## arrange()

df

color	value
4	1
1	2
5	3
3	4
2	5



color	value
1	2
2	5
3	4
4	1
5	3

```
arrange(df, color)
```

## Operaciones básicas de transformaciones

color	value
5	3
4	1
3	4
2	5
1	2

```
arrange(df, desc(color))
```

■ ACTIVIDAD: Ordenar el menú por precio

`select()`

Operaciones básicas de transformaciones

df

color	value
blue	1
black	2
blue	3
blue	4
black	5



color
blue
black
blue
blue
black

`select(df, color)`

value
1
2
3
4
5

`select(df, value)`

- ACTIVIDAD: Descarte la columna de categoría de los productos

```
mutate()
```

Operaciones básicas de transformaciones

df

color	value
blue	1
black	2
blue	3
blue	4
black	5

→

color	value	double
blue	1	2
black	2	4
blue	3	6
blue	4	8
black	5	10

```
mutate(df, double = 2 * value)
```

- ACTIVIDAD: El restaurante hace una promo. Todos los productos tienen **60%** de descuento. Calcule los precios.



`mutate()`

Operaciones básicas de transformaciones

df

color	value
blue	1
black	2
blue	3
blue	4
black	5

→

color	value	double	quadruple
blue	1	2	4
black	2	4	8
blue	3	6	12
blue	4	8	16
black	5	10	20

```
mutate(df, double = 2 * value,  
        quadruple = 2 * double)
```

- ACTIVIDAD: El restaurante hace una promo. Todos los productos tienen **60%** de descuento. Calcule los precios sin IVA en una columna y en otra con el IVA (**19%**)

```
summarise()
```

## Operaciones básicas de transformaciones

```
consumo <- read_xlsx("datos-restaurante.xlsx", sheet = "consumo")  
View(consumo)
```

numero_reserva	nombre	telefono	correo
1	Miguel	3143141414	<a href="mailto:miguel@correo.com">miguel@correo.com</a>
2	Andrés	3133133131	<a href="mailto:andres@correo.com">andres@correo.com</a>
3	Familia Martha	3152151515	<a href="mailto:martha@correo.com">martha@correo.com</a>

numero_reserva	producto	cantidad
	1 PLATO DE FRUTOS ROJOS	2
	1 SALCHICHAS	1
	1 PAN DE CHOCOLATE	1
	2 CALENTAO COLOMBIANO	1
	2 TOSTADAS	1
	3 CROISSANT	5
	3 AROMÁTICA	1
	3 ESPRESSO DOBLE	1
	3 CHOCOLATE	1
	3 MILO FRIO	1
	3 ICE COFFEE	1
	3 CALENTAO COLOMBIANO	1
	3 TOSTADAS FRANCESAS	1

```
summarise()
```

Operaciones básicas de transformaciones

df

color	value
blue	1
black	2
blue	3
blue	4
black	5

→

total
15

```
summarise(df, total = sum(value))
```

- ACTIVIDAD: Calcule el total de productos vendidos en la noche

`summarise()`

Operaciones básicas de transformaciones

df

color	value
blue	1
black	2
blue	3
blue	4
black	5



color	total
blue	8
black	7

```
by_color <- group_by(df, color)
summarise(by_color, total = sum(value))
```

- ACTIVIDAD: Calcule el total de productos vendidos en la noche POR RESERVA

# Algunas funciones de resumen

## Operaciones básicas de transformaciones

```
by_color <- group_by(df, color)
summarise(by_color, total = sum(value))
```

df

color	value
blue	1
black	2
blue	3
blue	4
black	5

→

color	total
blue	8
black	7

min(x), median(x), max(x),  
quantile(x, p)

n(), n\_distinct(x), sum(x), mean(x)

sum(x > 10), mean(x > 10)

sd(x), var(x), IQR(x), mad(x)

# Valores nulos (NA)

- ¿Mary tiene la misma edad de Jaime?
- ¿La edad y el sexo de Jaime es el mismo?

Name	Age	Sex
John	35	M
Mary	NA	F
Jaime	NA	NA

```
NA + 5
10 * NA
10 < NA
10 == NA
NA == NA
is.na(NA)

mean(c(10, 20, NA))
mean(c(10, 20, NA), na.rm = TRUE)
```

# Tubos de datos (pipes)



# Operador pipe (%>%)

Se lee como “después o luego”

```
menu %>%  
  filter(categoria != "FRUTAS") %>%  
  select(-producto) %>%  
  group_by(categoria) %>%  
  summarize(num_productos= n(),  
            prom = mean(precio))
```

1. Toma el menú,
2. ... después lo filtras para sacar las frutas,
3. ... después seleccionas todos menos la columna producto,
4. ... después lo agrupas por categoría,
5. ... después cuentas y sacas promedios

# Gracias por tu asistencia y participación 😊

## Contacto

✉ [miguela.orjuela@urosario.edu.co](mailto:miguela.orjuela@urosario.edu.co)

🌐 <https://www.linkedin.com/in/miguel-orjuela/>

🐙 <https://github.com/maorjuela73>

## Links de interés

- <https://bookdown.org/rdpeng/rprogdatascience/>
- <https://www.rstudio.com/resources/cheatsheets/>