

Manipulación de datos con dplyr

J. Eduardo Gamboa U.

2025-10-10

Contents

1	Introducción	2
2	Paquete dplyr	3
2.1	Aplicación	4
2.2	Función select	4
2.3	Función pull	6
2.4	Función mutate	7
2.5	Función rename	10
2.6	Función relocate	11
2.7	Función slice	13
2.8	Función filter	15
2.9	Función arrange	18
2.10	Función summarise	21
2.11	Función group_by	22
2.12	Función inner_join	24
2.13	Función right_join	25
2.14	Función left_join	27
2.15	Función full_join	29
3	Funciones de tidyr	30
3.1	Función pivot_longer	30
3.2	Función pivot_wider	32
4	Resumen de funciones para preparación de datos	33
5	Ejercicios	33

1 Introducción

“Garbage in, garbage out”

Antes de analizar o modelar, debemos preparar los datos. Estas tareas de preprocesamiento son esenciales para convertir un conjunto de datos crudo en una fuente confiable de información. Entre ellas destacan:

- Limpieza: detección y corrección de errores, valores faltantes o inconsistencias.
- Transformación: modificación de estructuras y formatos para facilitar el análisis.
- Integración: combinación de diferentes fuentes de datos de forma coherente.

En esta sesión utilizaremos los paquetes `dplyr` y `magrittr`. Este último introduce el operador pipe `%>%` (atajo: Ctrl + Shift + M en RStudio), aunque también es posible emplear el pipe nativo de R (`|>`).

El operador pipe permite que dispongamos los elementos de una función de manera opuesta, permitiendo la concatenación. Por ejemplo:

```
library(dplyr)
x = c(2,4,5,3)
sum(x)
```

```
## [1] 14
```

```
x %>% sum
```

```
## [1] 14
```

```
x |> sum()
```

```
## [1] 14
```

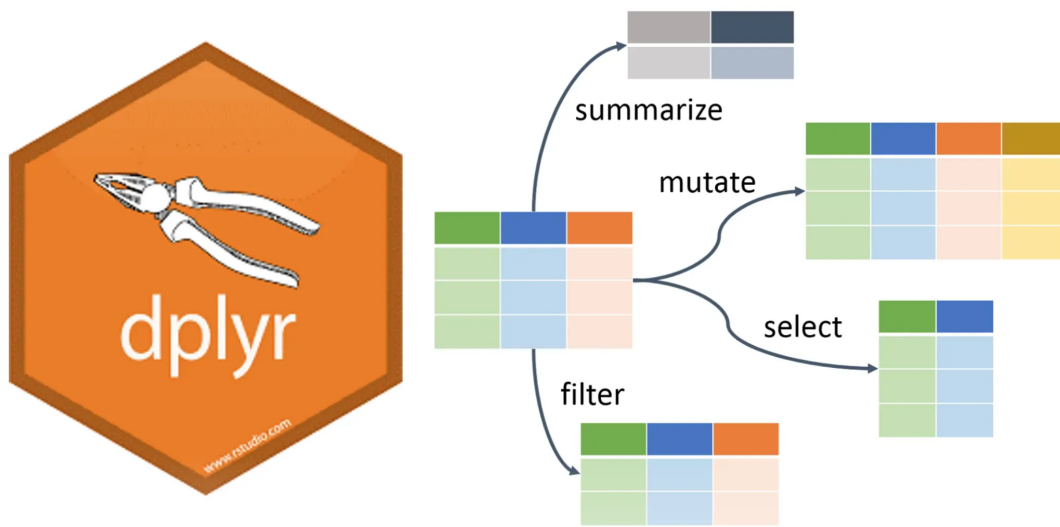
En general `f(x)` es expresado como `x %>% f`.

Tiene más utilidad cuando se usa una serie de funciones anidadas, por ejemplo: `h(g(f(x)))` es equivalente a `x %>% f %>% g %>% h` o `x %>% f() %>% g() %>% h()`

A partir de la versión 4.1.0, existe un nuevo pipe en R: `|>`

Es decir `h(g(f(x)))` es equivalente a `x |> f() |> g() |> h()`

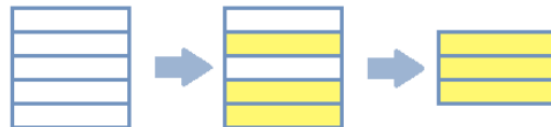
2 Package dplyr



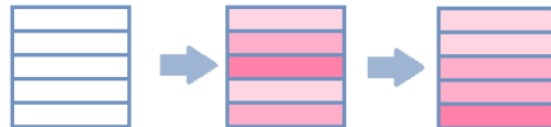
select



filter



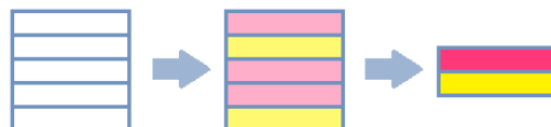
arrange



mutate



summarise



2.1 Aplicación

El archivo `Alumnos Matriculados 2025-II-UNALM.csv` contiene el listado de Alumnos matriculados del año 2025 - 2do Semestre. Cada fila corresponde a un curso, ofrecido por un departamento académico, el cual pertenece a alguna de las facultades que conforman la UNALM.

```
datos <- read.csv2('Alumnos Matriculados 2025-II-UNALM.csv')
datos |> head(4)
```

```
##   FACULTAD DEPARTAMENTO PERIODO          CURSO
## 1 Agronomia  Entomologia 2025-II      Apicultura General
## 2 Agronomia  Entomologia 2025-II      Control Biologico
## 3 Agronomia  Entomologia 2025-II  Crianza y Evaluacion de Insectos
## 4 Agronomia  Entomologia 2025-II      Entomologia Agricola
##   NRO_MATRICULADOS
## 1                  30
## 2                  34
## 3                  24
## 4                 196
```

2.2 Función select

¿Cómo se usa?

```
df |> select(posiciones, nombres o reglas de las variables a ser seleccionadas)
```

```
df |> select(-posiciones, nombres o reglas de las variables a ser eliminadas)
```

La(s) columna(s) seleccionadas se almacenan en un data frame.

2.2.1 Ejemplos

En todos los casos se mostrarán sólo las 4 primeras filas:

1. Seleccionar la variable CURSO

```
datos |> select(CURSO)|> head(4)
```

```
##           CURSO
## 1  Apicultura General
## 2   Control Biologico
## 3 Crianza y Evaluacion de Insectos
## 4   Entomologia Agricola
```

2. Seleccionar la tercera variable del data frame

```
datos |> select(3)|> head(4)
```

```
##   PERIODO
## 1 2025-II
## 2 2025-II
## 3 2025-II
## 4 2025-II
```

3. Seleccionar las variables FACULTAD y DEPARTAMENTO

```
datos |> select(FACULTAD, DEPARTAMENTO) |> head(4)
```

```
##   FACULTAD DEPARTAMENTO
## 1 Agronomia Entomologia
## 2 Agronomia Entomologia
## 3 Agronomia Entomologia
## 4 Agronomia Entomologia
```

4. Seleccionar todas las variables excepto DEPARTAMENTO

```
datos |> select(-DEPARTAMENTO) |> head(4)
```

```
##   FACULTAD PERIODO          CURSO NRO_MATRICULADOS
## 1 Agronomia 2025-II      Apicultura General         30
## 2 Agronomia 2025-II      Control Biologico          34
## 3 Agronomia 2025-II Crianza y Evaluacion de Insectos  24
## 4 Agronomia 2025-II      Entomologia Agricola       196
```

5. Seleccionar todas las variables excepto la primera y la segunda

```
datos |> select(-1, -3) |> head(4)
```

```
##   DEPARTAMENTO          CURSO NRO_MATRICULADOS
## 1 Entomologia      Apicultura General         30
## 2 Entomologia      Control Biologico          34
## 3 Entomologia Crianza y Evaluacion de Insectos  24
## 4 Entomologia      Entomologia Agricola       196
```

6. Seleccionar las columnas que contengan la silaba DO

```
datos |> select(contains("DO")) |> head(4)
```

```
##   PERIODO NRO_MATRICULADOS
## 1 2025-II         30
## 2 2025-II         34
## 3 2025-II         24
## 4 2025-II        196
```

7. Seleccionar las columnas de tipo character

```
datos |> select(where(is.character)) |> head(4)
```

```
##   FACULTAD DEPARTAMENTO PERIODO          CURSO
## 1 Agronomia Entomologia 2025-II      Apicultura General
## 2 Agronomia Entomologia 2025-II      Control Biologico
## 3 Agronomia Entomologia 2025-II Crianza y Evaluacion de Insectos
## 4 Agronomia Entomologia 2025-II      Entomologia Agricola
```

8. Seleccionar las columnas de tipo numeric o que inicien con F

```
datos |> select(where(is.numeric), starts_with('F')) |> head(4)
```

```
##   NRO_MATRICULADOS  FACULTAD
## 1                30 Agronomia
## 2                34 Agronomia
## 3                24 Agronomia
## 4               196 Agronomia
```

9. Seleccionar las columnas que no terminen con “O”

```
datos |> select(-ends_with('O')) |> head(4)
```

```
##   FACULTAD NRO_MATRICULADOS
## 1 Agronomia                30
## 2 Agronomia                34
## 3 Agronomia                24
## 4 Agronomia               196
```

2.3 Función pull

¿Cómo se usa?

```
df |> pull(posición o nombre de la variable a ser seleccionada)
```

La columna seleccionada se almacena en un vector.

2.3.1 Ejemplos

En todos los casos se mostrarán sólo los 4 primeras registros:

1. Seleccionar la columna FACULTAD usando pull.

```
datos |> pull(FACULTAD) |> head(4)
```

```
## [1] "Agronomia" "Agronomia" "Agronomia" "Agronomia"
```

2. Seleccionar la primera columna usando pull.

```
datos |> pull(1) |> head(4)
```

```
## [1] "Agronomia" "Agronomia" "Agronomia" "Agronomia"
```

2.4 Función mutate

¿Cómo se usa?

```
df |> mutate(nueva_variable = regla para crear la nueva variable)
```

```
df |> mutate(variable_existente = regla para modificar la variable)
```

2.4.1 Ejemplos

En todos los casos se mostrarán sólo las 6 primeras filas:

1. Incrementar en 10% la cantidad de matriculados en cada curso, de modo que esta nueva cantidad sea entera, y se almacene en una nueva variable de nombre MATRI_NUEVO.

```
datos |>
  mutate(MATRI_NUEVO = round(1.10*NRO_MATRICULADOS,0)) |> head()
```

	FACULTAD	DEPARTAMENTO	PERIODO	CURSO
## 1	Agronomia	Entomologia	2025-II	Apicultura General
## 2	Agronomia	Entomologia	2025-II	Control Biologico
## 3	Agronomia	Entomologia	2025-II	Crianza y Evaluacion de Insectos
## 4	Agronomia	Entomologia	2025-II	Entomologia Agricola
## 5	Agronomia	Entomologia	2025-II	Entomologia General
## 6	Agronomia	Entomologia	2025-II	Principios de Control de Plagas
	NRO_MATRICULADOS	MATRI_NUEVO		
## 1	30	33		
## 2	34	37		
## 3	24	26		
## 4	196	216		
## 5	162	178		
## 6	126	139		

2. Incrementar en 10% la cantidad de matriculados en cada curso, de modo que esta nueva cantidad sea entera, y se almacene en la misma variable.

```
datos |>
  mutate(NRO_MATRICULADOS = round(1.10*NRO_MATRICULADOS,0)) |> head(5)
```

	FACULTAD	DEPARTAMENTO	PERIODO	CURSO
## 1	Agronomia	Entomologia	2025-II	Apicultura General
## 2	Agronomia	Entomologia	2025-II	Control Biologico
## 3	Agronomia	Entomologia	2025-II	Crianza y Evaluacion de Insectos
## 4	Agronomia	Entomologia	2025-II	Entomologia Agricola
## 5	Agronomia	Entomologia	2025-II	Entomologia General
	NRO_MATRICULADOS			
## 1	33			
## 2	37			
## 3	26			
## 4	216			
## 5	178			

- Incrementar en 10% la cantidad de matriculados en cada curso, de modo que esta nueva cantidad sea entera, y se almacene en una variable de nombre MATRI_NUEVO y solo se muestre dicha variable.

```
datos |>
  transmute(MATRI_NUEVO = round(1.10*NRO_MATRICULADOS,0)) |>
  head()
```

```
##   MATRI_NUEVO
## 1           33
## 2           37
## 3           26
## 4          216
## 5           178
## 6           139
```

- Crear una nueva variable que se llame TAMAÑO, de modo que si el curso se dicta para menos de 40 alumnos tome el valor PEQUEÑO, de lo contrario, GRANDE

```
datos |>
  mutate(TAMAÑO = ifelse(NRO_MATRICULADOS < 40, "PEQUEÑO", "GRANDE")) |>
  head(5)
```

```
##   FACULTAD DEPARTAMENTO PERIODO CURSO
## 1 Agronomia Entomologia 2025-II Apicultura General
## 2 Agronomia Entomologia 2025-II Control Biologico
## 3 Agronomia Entomologia 2025-II Crianza y Evaluacion de Insectos
## 4 Agronomia Entomologia 2025-II Entomologia Agricola
## 5 Agronomia Entomologia 2025-II Entomologia General
##   NRO_MATRICULADOS TAMAÑO
## 1                30 PEQUEÑO
## 2                34 PEQUEÑO
## 3                24 PEQUEÑO
## 4               196 GRANDE
## 5               162 GRANDE
```


5. Crear una nueva variable que se llame TAMAÑO, de modo que si el curso se dicta para menos de 30 alumnos tome el valor PEQUEÑO, si son de 30 a 60 alumnos es MEDIANO, y más de 60 alumnos es GRANDE.

```
datos |>
  mutate(TAMAÑO = case_when(NRO_MATRICULADOS < 30 ~ "PEQUEÑO",
                             NRO_MATRICULADOS >= 30 & NRO_MATRICULADOS <=60 ~ "MEDIANO",
                             NRO_MATRICULADOS > 60 ~ "GRANDE")) |>
  head()
```

```
##   FACULTAD DEPARTAMENTO PERIODO          CURSO
## 1 Agronomia  Entomologia 2025-II      Apicultura General
## 2 Agronomia  Entomologia 2025-II      Control Biologico
## 3 Agronomia  Entomologia 2025-II  Crianza y Evaluacion de Insectos
## 4 Agronomia  Entomologia 2025-II      Entomologia Agricola
## 5 Agronomia  Entomologia 2025-II      Entomologia General
## 6 Agronomia  Entomologia 2025-II  Principios de Control de Plagas
##   NRO_MATRICULADOS  TAMAÑO
## 1                 30 MEDIANO
## 2                 34 MEDIANO
## 3                 24 PEQUEÑO
## 4                196  GRANDE
## 5                162  GRANDE
## 6                126  GRANDE
```

6. Crear una variable que se llame NUMEROS, de modo que si el curso es ofrecido por el departamento de Matemática o el de Estadística e Informática, toma el valor de 1, de lo contrario 0.

```
datos |>
  mutate(NUMEROS = ifelse(DEPARTAMENTO %in% c("Matematica", "Estadistica e Informatica"), 1, 0))
  head()
```

```
##   FACULTAD DEPARTAMENTO PERIODO          CURSO
## 1 Agronomia  Entomologia 2025-II      Apicultura General
## 2 Agronomia  Entomologia 2025-II      Control Biologico
## 3 Agronomia  Entomologia 2025-II  Crianza y Evaluacion de Insectos
## 4 Agronomia  Entomologia 2025-II      Entomologia Agricola
## 5 Agronomia  Entomologia 2025-II      Entomologia General
## 6 Agronomia  Entomologia 2025-II  Principios de Control de Plagas
##   NRO_MATRICULADOS NUMEROS
## 1                 30      0
## 2                 34      0
## 3                 24      0
## 4                196      0
## 5                162      0
## 6                126      0
```

2.5 Función rename

¿Cómo se usa?

```
df |> rename(nuevo_nombre = posición de la columna)
```

```
df |> rename(nuevo_nombre = antiguo_nombre)
```

2.5.1 Ejemplos

En todos los casos se mostrarán sólo las 3 primeras filas:

1. Renombrar la primera columna como Facultad.

```
datos |> rename(FACULTAD = 1) |> head(3)
```

```
##   FACULTAD DEPARTAMENTO PERIODO          CURSO
## 1 Agronomia  Entomologia 2025-II      Apicultura General
## 2 Agronomia  Entomologia 2025-II      Control Biologico
## 3 Agronomia  Entomologia 2025-II Crianza y Evaluacion de Insectos
##   NRO_MATRICULADOS
## 1                 30
## 2                 34
## 3                 24
```

2. Renombrar la columna PERIODO como Ciclo

```
datos |> rename(Ciclo = PERIODO) |> head(3)
```

```
##   FACULTAD DEPARTAMENTO  Ciclo          CURSO
## 1 Agronomia  Entomologia 2025-II      Apicultura General
## 2 Agronomia  Entomologia 2025-II      Control Biologico
## 3 Agronomia  Entomologia 2025-II Crianza y Evaluacion de Insectos
##   NRO_MATRICULADOS
## 1                 30
## 2                 34
## 3                 24
```

3. Renombrar la primera columna como Facultad y la segunda como Depto

```
datos |> rename(Facultad = 1, Depto = 2) |> head(3)
```

```
##   Facultad  Depto PERIODO          CURSO
## 1 Agronomia Entomologia 2025-II      Apicultura General
## 2 Agronomia Entomologia 2025-II      Control Biologico
## 3 Agronomia Entomologia 2025-II Crianza y Evaluacion de Insectos
##   NRO_MATRICULADOS
## 1                 30
## 2                 34
## 3                 24
```

4. Pasar a minúsculas los nombres de todas las variables

```
datos |> rename_with(tolower) |> head(3)
```

```
##   facultad departamento periodo                curso
## 1 Agronomia  Entomologia 2025-II          Apicultura General
## 2 Agronomia  Entomologia 2025-II          Control Biologico
## 3 Agronomia  Entomologia 2025-II Crianza y Evaluacion de Insectos
##   nro_matriculados
## 1                 30
## 2                 34
## 3                 24
```

5. Pasar a minúsculas los nombres de todas las variables, excepto la primera letra, que debe ir en mayúscula

```
datos |> rename_with(~ tools::toTitleCase(tolower(.x))) |> head(3)
```

```
##   Facultad Departamento Periodo                Curso
## 1 Agronomia  Entomologia 2025-II          Apicultura General
## 2 Agronomia  Entomologia 2025-II          Control Biologico
## 3 Agronomia  Entomologia 2025-II Crianza y Evaluacion de Insectos
##   Nro_matriculados
## 1                 30
## 2                 34
## 3                 24
```

2.6 Función relocate

¿Cómo se usa?

```
df |> relocate(.before = nombre de columna)
```

```
df |> relocate(.after = nombre de columna)
```

2.6.1 Ejercicios

En todos los casos se mostrarán sólo las 3 primeras filas:

1. Mover la columna PERIODO a la primera posición

```
datos |> relocate(PERIODO) |> head(3)
```

```
##   PERIODO  FACULTAD DEPARTAMENTO                CURSO
## 1 2025-II Agronomia  Entomologia          Apicultura General
## 2 2025-II Agronomia  Entomologia          Control Biologico
## 3 2025-II Agronomia  Entomologia Crianza y Evaluacion de Insectos
##   NRO_MATRICULADOS
## 1                 30
## 2                 34
## 3                 24
```

2. Mover la columna PERIODO a la última posición

```
datos |> relocate(PERIODO, .after = last_col()) |> head(3)
```

```
##   FACULTAD DEPARTAMENTO                                CURSO NRO_MATRICULADOS
## 1 Agronomia Entomologia                        Apicultura General           30
## 2 Agronomia Entomologia                        Control Biologico           34
## 3 Agronomia Entomologia Crianza y Evaluacion de Insectos           24
##   PERIODO
## 1 2025-II
## 2 2025-II
## 3 2025-II
```

3. Mover la columna PERIODO antes de NRO_MATRICULADOS

```
datos |> relocate(PERIODO, .before = NRO_MATRICULADOS) |> head(3)
```

```
##   FACULTAD DEPARTAMENTO                                CURSO PERIODO
## 1 Agronomia Entomologia                        Apicultura General 2025-II
## 2 Agronomia Entomologia                        Control Biologico 2025-II
## 3 Agronomia Entomologia Crianza y Evaluacion de Insectos 2025-II
##   NRO_MATRICULADOS
## 1                30
## 2                34
## 3                24
```

4. Mover la columna PERIODO después de DEPARTAMENTO

```
datos |> relocate(PERIODO, .after = DEPARTAMENTO) |> head(3)
```

```
##   FACULTAD DEPARTAMENTO PERIODO                                CURSO
## 1 Agronomia Entomologia 2025-II                        Apicultura General
## 2 Agronomia Entomologia 2025-II                        Control Biologico
## 3 Agronomia Entomologia 2025-II Crianza y Evaluacion de Insectos
##   NRO_MATRICULADOS
## 1                30
## 2                34
## 3                24
```

5. Mover las columnas FACULTAD y DEPARTAMENTO después de CURSO

```
datos |> relocate(FACULTAD, DEPARTAMENTO, .after = CURSO) |> head(3)
```

```
##   PERIODO                                CURSO FACULTAD DEPARTAMENTO
## 1 2025-II                        Apicultura General Agronomia Entomologia
## 2 2025-II                        Control Biologico Agronomia Entomologia
## 3 2025-II Crianza y Evaluacion de Insectos Agronomia Entomologia
##   NRO_MATRICULADOS
## 1                30
## 2                34
## 3                24
```

2.7 Función slice

¿Cómo se usa?

```
df |> slice(fila o vector de filas a ser seleccionadas)
```

```
df |> slice(-(fila o vector de filas a ser seleccionadas))
```

2.7.1 Ejemplos

1. Seleccionar la primera fila.

```
datos |> slice(1)
```

```
##      FACULTAD DEPARTAMENTO PERIODO          CURSO NRO_MATRICULADOS
## 1 Agronomia  Entomologia 2025-II Apicultura General             30
```

2. Seleccionar las filas 3, 7, 48 y 120.

```
datos |> slice(c(3,7,48, 120))
```

```
##      FACULTAD      DEPARTAMENTO PERIODO          CURSO
## 1 Agronomia      Entomologia 2025-II Crianza y Evaluacion de Insectos
## 2 Agronomia      Fitopatologia 2025-II      Fitopatologia Agricola
## 3 Agronomia      Horticultura 2025-II      Manejo de Viveros
## 4 Ciencias Fisica y Meteorologia 2025-II      Meteorologia Dinamica
##      NRO_MATRICULADOS
## 1              24
## 2             330
## 3              23
## 4              16
```

3. Seleccionar todas las filas excepto la 8 en adelante.

```
datos |> slice(-(8:669))
```

```
##      FACULTAD DEPARTAMENTO PERIODO          CURSO
## 1 Agronomia  Entomologia 2025-II      Apicultura General
## 2 Agronomia  Entomologia 2025-II      Control Biologico
## 3 Agronomia  Entomologia 2025-II Crianza y Evaluacion de Insectos
## 4 Agronomia  Entomologia 2025-II      Entomologia Agricola
## 5 Agronomia  Entomologia 2025-II      Entomologia General
## 6 Agronomia  Entomologia 2025-II Principios de Control de Plagas
## 7 Agronomia Fitopatologia 2025-II      Fitopatologia Agricola
##      NRO_MATRICULADOS
## 1              30
## 2              34
## 3              24
## 4             196
## 5             162
## 6             126
## 7             330
```

4. Seleccionar las 5 primeras filas.

```
datos |> slice_head(n = 5)
```

```
##      FACULTAD DEPARTAMENTO PERIODO      CURSO
## 1 Agronomia  Entomologia 2025-II      Apicultura General
## 2 Agronomia  Entomologia 2025-II      Control Biologico
## 3 Agronomia  Entomologia 2025-II Crianza y Evaluacion de Insectos
## 4 Agronomia  Entomologia 2025-II      Entomologia Agricola
## 5 Agronomia  Entomologia 2025-II      Entomologia General
##      NRO_MATRICULADOS
## 1                30
## 2                34
## 3                24
## 4               196
## 5               162
```

5. Seleccionar las 3 últimas filas

```
datos |> slice_tail(n = 3)
```

```
##      FACULTAD      DEPARTAMENTO PERIODO
## 1 Zootecnia Produccion Animal 2025-II
## 2 Zootecnia Produccion Animal 2025-II
## 3 Zootecnia Produccion Animal 2025-II
##
##                                     CURSO NRO_MATRICULADOS
## 1      Ultrasonografia y Laparoscopia Reproductiva      10
## 2 Utilizacion de Pastizales en la Produccion Animal      64
## 3                                     Zootecnia General      175
```

6. Selecciona una muestra aleatoria de 7 filas.

```
datos |> slice_sample(n = 7)
```

```
##      FACULTAD      DEPARTAMENTO PERIODO
## 1 Economia y Planificacion Economia y Planificacion 2025-II
## 2 Economia y Planificacion Economia y Planificacion 2025-II
## 3      Ingenieria Agricola      Recursos Hidricos 2025-II
## 4 Economia y Planificacion      Gestion Empresarial 2025-II
## 5 Economia y Planificacion Economia y Planificacion 2025-II
## 6      Ciencias      Fisica y Meteorologia 2025-II
## 7      Ingenieria Agricola      Recursos Hidricos 2025-II
##
##                                     CURSO NRO_MATRICULADOS
## 1      Finanzas Corporativas      21
## 2      Macroeconomia II      43
## 3      Ingenieria de Riegos I      49
## 4 Planeamiento y Direccion Estrategica      47
## 5      Macroeconomia Dinamica      33
## 6      Meteorologia General      244
## 7 Hidrologia y Drenaje Urbano y Vial      6
```

7. Seleccionar los 4 registros con mayor cantidad de matriculados

```
datos |> slice_max(NRO_MATRICULADOS, n = 4)
```

```
##          FACULTAD          DEPARTAMENTO PERIODO
## 1      Ciencias          Matematica 2025-II
## 2      Ciencias          Quimica 2025-II
## 3      Ciencias          Biologia 2025-II
## 4 Economia y Planificacion Economia y Planificacion 2025-II
##          CURSO NRO_MATRICULADOS
## 1 Analisis Matematico I          1039
## 2      Quimica General          934
## 3      Ecologia General          892
## 4      Economia General          728
```

2.8 Función filter

¿Cómo se usa?

```
df |> filter(reglas de filtro)
```

Operadores lógicos utilizados con `filter`:

- `<`: menor que
- `>`: mayor que
- `==`: igual (un valor)
- `\texttt{%in%}`: pertenencia a un conjunto
- `&`: y
- `|`: o

2.8.1 Ejemplos

En todos los casos se mostrarán sólo las 6 primeras filas:

1. Seleccionar los registros de cursos de la facultad de Ciencias

```
datos |> filter(FACULTAD == "Ciencias")|> head()
```

```
##  FACULTAD DEPARTAMENTO PERIODO          CURSO NRO_MATRICULADOS
## 1 Ciencias      Biologia 2025-II      Biodiversidad          24
## 2 Ciencias      Biologia 2025-II      Biologia Celular        33
## 3 Ciencias      Biologia 2025-II      Biologia Experimental I    10
## 4 Ciencias      Biologia 2025-II      Biologia Experimental II   13
## 5 Ciencias      Biologia 2025-II      Biologia Experimental III    4
## 6 Ciencias      Biologia 2025-II      Biologia General         562
```

2. Seleccionar los registros de cursos de la facultad de Ciencias y que su cantidad de matriculados sea mayor a 200

```
datos |> filter(FACULTAD == "Ciencias" & NRO_MATRICULADOS > 200) |> head()
```

##	FACULTAD	DEPARTAMENTO	PERIODO	CURSO	NRO_MATRICULADOS
## 1	Ciencias	Biologia	2025-II	Biologia General	562
## 2	Ciencias	Biologia	2025-II	Botanica General	203
## 3	Ciencias	Biologia	2025-II	Ecologia General	892
## 4	Ciencias	Biologia	2025-II	Fisiologia Vegetal	336
## 5	Ciencias Fisica y Meteorologia	2025-II	Fisica General	553	
## 6	Ciencias Fisica y Meteorologia	2025-II	Meteorologia General	244	

3. Seleccionar los registros de cursos de la facultad de Ciencias o que su cantidad de matriculados sea mayor a 200

```
datos |> filter(FACULTAD == "Ciencias" | NRO_MATRICULADOS > 200) |> head()
```

##	FACULTAD	DEPARTAMENTO	PERIODO	CURSO	NRO_MATRICULADOS
## 1	Agronomia	Fitopatologia	2025-II		
## 2	Agronomia	Fitopatologia	2025-II		
## 3	Agronomia	Fitotecnica	2025-II		
## 4	Agronomia	Fitotecnica	2025-II		
## 5	Agronomia	Fitotecnica	2025-II		
## 6	Agronomia	Fitotecnica	2025-II		
##					
## 1				Fitopatologia Agricola	330
## 2				Fitopatologia General	244
## 3				Agroecologia	202
## 4				Cultivos II (Papa, camote, leguminosas)	252
## 5				Cultivos IV (Cultivos tropicales, cultivos forrajeros)	219
## 6				Introduccion a la Agronomia	211

4. Seleccionar los registros de cursos de la facultad de Ciencias o de Agronomia

```
datos |> filter(FACULTAD %in% c("Ciencias", "Agronomia")) |> head()
```

##	FACULTAD	DEPARTAMENTO	PERIODO	CURSO
## 1	Agronomia	Entomologia	2025-II	Apicultura General
## 2	Agronomia	Entomologia	2025-II	Control Biologico
## 3	Agronomia	Entomologia	2025-II	Crianza y Evaluacion de Insectos
## 4	Agronomia	Entomologia	2025-II	Entomologia Agricola
## 5	Agronomia	Entomologia	2025-II	Entomologia General
## 6	Agronomia	Entomologia	2025-II	Principios de Control de Plagas
##				
##				NRO_MATRICULADOS
## 1				30
## 2				34
## 3				24
## 4				196
## 5				162
## 6				126

5. Seleccionar los registros de cursos de la facultad de Ciencias o de Agronomía, y que tengan más de 299 matriculados

```
datos |> filter((FACULTAD %in% c("Ciencias", "Agronomía")) & NRO_MATRICULADOS > 299) |> head(6)
```

	FACULTAD	DEPARTAMENTO	PERIODO	CURSO	NRO_MATRICULADOS
## 1	Agronomía	Fitopatología	2025-II	Fitopatología Agrícola	330
## 2	Agronomía	Horticultura	2025-II	Olericultura General	339
## 3	Agronomía	Suelos	2025-II	Edafología	343
## 4	Ciencias	Biología	2025-II	Biología General	562
## 5	Ciencias	Biología	2025-II	Ecología General	892
## 6	Ciencias	Biología	2025-II	Fisiología Vegetal	336

6. Seleccionar los registros de cursos que no sean de la facultad de Pesquería

```
datos |> filter(FACULTAD != "Pesquería") |> head(6)
```

	FACULTAD	DEPARTAMENTO	PERIODO	CURSO	NRO_MATRICULADOS
## 1	Agronomía	Entomología	2025-II	Apicultura General	30
## 2	Agronomía	Entomología	2025-II	Control Biológico	34
## 3	Agronomía	Entomología	2025-II	Crianza y Evaluación de Insectos	24
## 4	Agronomía	Entomología	2025-II	Entomología Agrícola	196
## 5	Agronomía	Entomología	2025-II	Entomología General	162
## 6	Agronomía	Entomología	2025-II	Principios de Control de Plagas	126

7. Seleccionar los registros de cursos de la facultad de Economía y Planificación que tengan entre 20 y 40 matriculados, inclusive

```
datos |> filter(FACULTAD == "Economía y Planificación" & between(NRO_MATRICULADOS, 20, 40)) |> head(4)
```

	FACULTAD	DEPARTAMENTO	PERIODO	CURSO	NRO_MATRICULADOS
## 1	Economía y Planificación	Ciencias Humanas y Educación	2025-II		
## 2	Economía y Planificación	Ciencias Humanas y Educación	2025-II		
## 3	Economía y Planificación	Ciencias Humanas y Educación	2025-II		
## 4	Economía y Planificación	Economía y Planificación	2025-II		
## 1				ética	25
## 2				Proyecto Universitario Personalizado	26
## 3				Redacción Técnica	22
## 4				Desarrollo y Política Agraria	36

2.9 Función arrange

¿Cómo se usa?

```
df |> arrange(variable) # orden alfabético A Z, o de menor a mayor
```

```
df |> arrange(desc(variable)) # orden alfabético Z A, o de mayor a menor
```

```
df |> arrange(variable numérica) # orden de mayor a menor
```

2.9.1 Ejemplos

En todos los casos se mostrarán sólo las primeras filas:

1. Ordenar el conjunto de datos según el número de matriculados, de manera ascendente

```
datos |> arrange(NRO_MATRICULADOS) |> head(4)
```

```
##          FACULTAD          DEPARTAMENTO PERIODO
## 1          Ciencias          Biologia 2025-II
## 2          Ciencias Fisica y Meteorologia 2025-II
## 3          Ciencias  Ingenieria Ambiental 2025-II
## 4 Economia y Planificacion  Gestion Empresarial 2025-II
##                                CURSO NRO_MATRICULADOS
## 1          Ecologia de Montanas                1
## 2                                Seminario I        1
## 3 Seminario en Ingenieria Ambiental            1
## 4          Administracion por Objetivos          1
```

2. Ordenar el conjunto de datos según el número de matriculados, de manera descendente

```
datos |> arrange(-NRO_MATRICULADOS) |> head(3)
```

```
##  FACULTAD DEPARTAMENTO PERIODO          CURSO NRO_MATRICULADOS
## 1 Ciencias  Matematica 2025-II Analisis Matematico I          1039
## 2 Ciencias   Quimica 2025-II   Quimica General              934
## 3 Ciencias   Biologia 2025-II   Ecologia General             892
```

```
datos |> arrange(desc(NRO_MATRICULADOS)) |> head(3)
```

```
##  FACULTAD DEPARTAMENTO PERIODO          CURSO NRO_MATRICULADOS
## 1 Ciencias  Matematica 2025-II Analisis Matematico I          1039
## 2 Ciencias   Quimica 2025-II   Quimica General              934
## 3 Ciencias   Biologia 2025-II   Ecologia General             892
```

3. Ordenar el conjunto de datos según el departamento académico, alfabéticamente de la A a la Z.

```
datos |> arrange(DEPARTAMENTO) |> head()
```

```
##                                FACULTAD      DEPARTAMENTO PERIODO
## 1 Oficina de Bienestar Estudiantil Actividades Culturales 2025-II
## 2 Oficina de Bienestar Estudiantil Actividades Culturales 2025-II
## 3 Oficina de Bienestar Estudiantil Actividades Culturales 2025-II
## 4 Oficina de Bienestar Estudiantil Actividades Culturales 2025-II
## 5 Oficina de Bienestar Estudiantil Actividades Culturales 2025-II
## 6 Oficina de Bienestar Estudiantil Actividades Culturales 2025-II
##                                CURSO NRO_MATRICULADOS
## 1 Artes Plasticas y Creatividad                40
## 2                                Banda Unalm          75
## 3                                Capoeira             43
## 4                                Danza de Costa       21
## 5                                Danzas Folcloricas   117
## 6                                Dibujo y Pintura     41
```

4. Ordenar el conjunto de datos según el departamento académico, alfabéticamente de la Z a la A.

```
datos |> arrange(desc(DEPARTAMENTO)) |> head()
```

```
##                                FACULTAD DEPARTAMENTO PERIODO
## 1 Industrias Alimentarias      T.A.P.A. 2025-II
## 2 Industrias Alimentarias      T.A.P.A. 2025-II
## 3 Industrias Alimentarias      T.A.P.A. 2025-II
## 4 Industrias Alimentarias      T.A.P.A. 2025-II
## 5 Industrias Alimentarias      T.A.P.A. 2025-II
## 6 Industrias Alimentarias      T.A.P.A. 2025-II
##                                CURSO NRO_MATRICULADOS
## 1                                Fermentaciones Industriales 70
## 2                                Industrias Carnicas          10
## 3                                Industrias Lacteas           11
## 4 Introduccion a la Industria Alimentaria                    54
## 5 Seminario en Industrias Alimentarias I                     37
## 6 Seminario en Industrias Alimentarias II                     19
```

5. Filtrar los registros de cursos con 10 a 20 matriculados y luego ordenar el conjunto de datos según la cantidad de matriculados (descendentemente) y la facultad (de la Z a la A). Mostrar los 18 primeros registros.

```
datos |>
  filter(between(NRO_MATRICULADOS,10,20)) |>
  arrange(-NRO_MATRICULADOS, desc(FACULTAD)) |>
  head(18)
```

##	FACULTAD	DEPARTAMENTO	PERIODO
## 1	Zootecnia	Nutricion	2025-II
## 2	Zootecnia	Produccion Animal	2025-II
## 3	Pesqueria	Acuicultura e Industrias Pesqueras	2025-II
## 4	Pesqueria	Manejo Pesquero y Medio Ambiente	2025-II
## 5	Oficina de Bienestar Estudiantil	Educacion Fisica y Deportes	2025-II
## 6	Ingenieria Agricola	Mecanizacion y Energia	2025-II
## 7	Industrias Alimentarias	I.A.P.A.	2025-II
## 8	Industrias Alimentarias	T.A.P.A.	2025-II
## 9	Industrias Alimentarias	T.A.P.A.	2025-II
## 10	Economia y Planificacion	Economia y Planificacion	2025-II
## 11	Ciencias	Biologia	2025-II
## 12	Ciencias	Fisica y Meteorologia	2025-II
## 13	Ciencias	Matematica	2025-II
## 14	Ciencias	Quimica	2025-II
## 15	Pesqueria	Acuicultura e Industrias Pesqueras	2025-II
## 16	Pesqueria	Manejo Pesquero y Medio Ambiente	2025-II
## 17	Ingenieria Agricola	Mecanizacion y Energia	2025-II
## 18	Ingenieria Agricola	Recursos Hidricos	2025-II
##		CURSO	NRO_MATRICULADOS
## 1		Farmacologia	20
## 2		Gestion de la Produccion de Pavos, Patos y Codornices	20
## 3		Procesamiento de Aceite y Harina de Pescado	20
## 4		Evaluacion de Recursos Hidrobiologicos	20
## 5		Judo	20
## 6		Ingenieria y Responsabilidad Social	20
## 7		Fundamentos de Control y Automatizacion	20
## 8		Tecnologia de Aceites y Grasas	20
## 9		Tecnologia de Alcoholes y Derivados	20
## 10		Economia de la Informacion	20
## 11		Comunicacion Cientifica	20
## 12		Fisica Aplicada	20
## 13		Calculo Diferencial	20
## 14		Cinetica Quimica y Enzimatica	20
## 15		Procesamiento de Enlatado de Recursos Hidrobiologicos	19
## 16		Materiales, Metodos y Artes de Pesca	19
## 17		Mantenimiento de Maquinaria	19
## 18		Estructuras Hidraulicas II	19

2.10 Función summarise

¿Cómo se usa?

```
df |> summarise(Resultado = funcion_de_resumen(variable))
```

2.10.1 Ejemplos

1. ¿Cuál es la cantidad promedio de matriculados por curso?

```
datos |> summarise(Media = mean(NRO_MATRICULADOS))
```

```
##      Media
## 1 57.76532
```

2. ¿Cuál es el percentil 10 y la mediana del numero de matriculados por curso?

```
datos |> summarise(P10      = quantile(NRO_MATRICULADOS, 0.10),
                  Mediana = median(NRO_MATRICULADOS))
```

```
##    P10 Mediana
## 1    9      30
```

3. ¿Cuál es el departamento académico que más cursos ofrece?

```
datos |> summarise(Moda = modeest::mfv(DEPARTAMENTO))
```

```
##      Moda
## 1 Gestion Empresarial
```

4. Obtener la lista de cursos ofrecidos por el departamento de Estadística e Informática.

```
datos |>
  filter(DEPARTAMENTO == "Estadística e Informática") |>
  summarise(lista_cursos = paste(CURSO, collapse = ", "))
```

```
##
## 1 Algoritmia, Analisis de Regresion, Analisis de Series de Tiempo, Analisis de Sobrevivencia
```

5. ¿Cuántos cursos ofrece la UNALM en el ciclo 2025-2?

```
datos |>
  summarise(total_cursos = n_distinct(CURSO))
```

```
##    total_cursos
## 1             658
```

2.11 Función group_by

¿Cómo se usa?

```
df |> group_by(Variable de agrupacion |> ...
```

1. ¿Cuál es la cantidad promedio de matriculados por curso en cada facultad?

```
datos |>
  group_by(FACULTAD) |>
  summarise(Media = mean(NRO_MATRICULADOS))
```

```
## # A tibble: 10 x 2
##   FACULTAD           Media
##   <chr>           <dbl>
## 1 Agronomia         87.8
## 2 Ciencias          79.1
## 3 Ciencias Forestales 37.1
## 4 Economia y Planificacion 64.6
## 5 Estudios Generales 426
## 6 Industrias Alimentarias 55
## 7 Ingenieria Agricola 34.7
## 8 Oficina de Bienestar Estudiantil 40.9
## 9 Pesqueria        22.3
## 10 Zootecnia        38.5
```

2. ¿Cuál es el percentil 10, la mediana y desviación estándar del numero de matriculados en cada uno de los cursos ofrecidos por la facultad de Economia y Planificacion?

```
datos |>
  filter(FACULTAD == "Economia y Planificacion") |>
  group_by(DEPARTAMENTO) |>
  summarise(P10 = quantile(NRO_MATRICULADOS, 0.10),
            Mediana = median(NRO_MATRICULADOS),
            DesvEst = sd(NRO_MATRICULADOS))
```

```
## # A tibble: 5 x 4
##   DEPARTAMENTO           P10 Mediana DesvEst
##   <chr>           <dbl>   <dbl>   <dbl>
## 1 Ciencias Humanas y Educacion 23.8     46    308.
## 2 Economia y Planificacion    12.5     28    104.
## 3 Estadistica e Informatica   15.6     30     61.8
## 4 Gestion Empresarial        17.8     40     43.6
## 5 Sociologia Rural           26      56    278.
```

3. ¿Cuál es el departamento académico que más cursos ofrece en cada facultad?

```
datos |>
  group_by(FACULTAD) |>
  summarise(Moda = modeest::mfv(DEPARTAMENTO))
```

```
## # A tibble: 10 x 2
##   FACULTAD                               Moda
##   <chr>                                <chr>
## 1 Agronomia                           Fitotecnia
## 2 Ciencias                           Fisica y Meteorologia
## 3 Ciencias Forestales                 Manejo Forestal
## 4 Economia y Planificacion            Gestion Empresarial
## 5 Estudios Generales                 Estudios Generales
## 6 Industrias Alimentarias             I.A.P.A.
## 7 Ingenieria Agricola                 Recursos Hidricos
## 8 Oficina de Bienestar Estudiantil    Educacion Fisica y Deportes
## 9 Pesqueria                          Acuicultura e Industrias Pesqueras
## 10 Zootecnia                          Produccion Animal
```

4. Obtener la lista de cursos ofrecidos por cada departamento académico.

```
datos |>
  group_by(DEPARTAMENTO) |>
  summarise(lista_cursos = paste(CURSO, collapse = ", "))
```

```
## # A tibble: 30 x 2
##   DEPARTAMENTO                               lista_cursos
##   <chr>                                <chr>
## 1 Actividades Culturales                 Artes Plasticas y Creatividad, Banda Unal~
## 2 Acuicultura e Industrias Pesqueras    Acuicultura I, Acuicultura II, Conservaci~
## 3 Biologia                             Biodiversidad, Biologia Celular, Biologia~
## 4 Ciencias Humanas y Educacion          etica, etica y Ciudadania, Historia Econo~
## 5 Economia y Planificacion              Desarrollo y Politica Agraria, Econometri~
## 6 Educacion Fisica y Deportes           Aerobicos, Aikido, Ajedrez, Basquet, Beis~
## 7 Entomologia                          Apicultura General, Control Biologico, Cr~
## 8 Estadistica e Informatica             Algoritmia, Analisis de Regresion, Analis~
## 9 Estudios Generales                   Trabajo de investigacion
## 10 Fisica y Meteorologia                 Agrometeorologia, Cambio Climatico, Adapt~
## # i 20 more rows
```

2.12 Función inner_join

Retorna las filas de x que tienen valores coincidentes en y, considerando todas las columnas de x e y
¿Cómo se usa?

```
x |> inner_join(y)
```

2.12.1 Ejemplo

Leer el archivo `Matriculados Alumnos Año 2025.csv`, que contiene los mismos datos que el archivo que hemos venido trabajando, pero del ciclo 2025-1. Combinar los registros de ambos data frames usando `inner_join`.

```
datos2 <- read.csv2('Matriculados Alumnos Año 2025.csv')

datos |>
  inner_join(datos2, by = c("FACULTAD" = "Facultad",
                           "DEPARTAMENTO" = "Departamento",
                           "CURSO" = "curso")) -> datos_2025

datos_2025 |> head(4)
```

```
##   FACULTAD DEPARTAMENTO PERIODO CURSO
## 1 Agronomia Entomologia 2025-II Apicultura General
## 2 Agronomia Entomologia 2025-II Control Biologico
## 3 Agronomia Entomologia 2025-II Crianza y Evaluacion de Insectos
## 4 Agronomia Entomologia 2025-II Entomologia Agricola
##   NRO_MATRICULADOS Periodo cantidad_matriculados
## 1                30 2025-I                28
## 2                34 2025-I                 4
## 3                24 2025-I                 4
## 4               196 2025-I               109
```

Dándole forma:

```
datos_2025 |>
  relocate(CURSO, .before = PERIODO) |>
  rename(MAT_25II = 5,
         MAT_25I = 7) |> head(4)
```

```
##   FACULTAD DEPARTAMENTO CURSO PERIODO MAT_25II
## 1 Agronomia Entomologia Apicultura General 2025-II 30
## 2 Agronomia Entomologia Control Biologico 2025-II 34
## 3 Agronomia Entomologia Crianza y Evaluacion de Insectos 2025-II 24
## 4 Agronomia Entomologia Entomologia Agricola 2025-II 196
##   Periodo MAT_25I
## 1 2025-I 28
## 2 2025-I 4
## 3 2025-I 4
## 4 2025-I 109
```


2.13 Función right_join

Retorna las filas de y, además de todas las columnas de x e y. Completa con NA cuando es necesario.

¿Cómo se usa?

```
x |> right_join(y)
```

2.13.1 Ejemplo

Combinar los registros de ambos data frames usando right_join.

```
datos |>
  right_join(datos2, by = c("FACULTAD" = "Facultad",
                           "DEPARTAMENTO" = "Departamento",
                           "CURSO" = "curso")) -> datos_2025

datos_2025 |> head(5)
```

```
##   FACULTAD DEPARTAMENTO PERIODO CURSOR
## 1 Agronomia Entomologia 2025-II Apicultura General
## 2 Agronomia Entomologia 2025-II Control Biologico
## 3 Agronomia Entomologia 2025-II Crianza y Evaluacion de Insectos
## 4 Agronomia Entomologia 2025-II Entomologia Agricola
## 5 Agronomia Entomologia 2025-II Entomologia General
##   NRO_MATRICULADOS Periodo cantidad_matriculados
## 1                30 2025-I                    28
## 2                34 2025-I                     4
## 3                24 2025-I                     4
## 4               196 2025-I                   109
## 5               162 2025-I                   131
```

Dándole forma:

```
datos_2025 |>
  relocate(CURSO, .before = PERIODO) |>
  rename(MAT_25II = 5,
         MAT_25I = 7) -> datos_2025

datos_2025 |> head(5)
```

```
##   FACULTAD DEPARTAMENTO CURSOR PERIODO MAT_25II
## 1 Agronomia Entomologia Apicultura General 2025-II 30
## 2 Agronomia Entomologia Control Biologico 2025-II 34
## 3 Agronomia Entomologia Crianza y Evaluacion de Insectos 2025-II 24
## 4 Agronomia Entomologia Entomologia Agricola 2025-II 196
## 5 Agronomia Entomologia Entomologia General 2025-II 162
##   Periodo MAT_25I
## 1 2025-I 28
## 2 2025-I 4
## 3 2025-I 4
## 4 2025-I 109
## 5 2025-I 131
```

¿Qué cursos se dictaron solo en el ciclo 2025-I?

```
datos_2025 |>
  filter(is.na(MAT_25II)) |>
  head(13)
```

##	FACULTAD	DEPARTAMENTO		
## 1	Agronomia	Horticultura		
## 2	Agronomia	Horticultura		
## 3	Agronomia	Horticultura		
## 4	Agronomia	Suelos		
## 5	Agronomia	Suelos		
## 6	Agronomia	Suelos		
## 7	Ciencias	Biologia		
## 8	Ciencias	Biologia		
## 9	Ciencias	Biologia		
## 10	Ciencias	Biologia		
## 11	Ciencias	Biologia		
## 12	Ciencias Fisica y Meteorologia			
## 13	Ciencias Fisica y Meteorologia			
##			CURSO PERIODO MAT_25II	
## 1		Certificaciones Agricolas	<NA>	NA
## 2		Floricultura I	<NA>	NA
## 3		Floricultura II	<NA>	NA
## 4		Analisis de Suelos y Plantas	<NA>	NA
## 5		Cartografia del Suelo	<NA>	NA
## 6		Suelos Forestales	<NA>	NA
## 7	Diversidad de Briofitas, Pteridofitas y Gimnospermas		<NA>	NA
## 8		Diversidad de Invertebrados	<NA>	NA
## 9		Ecofisiologia	<NA>	NA
## 10		Ecologia de Ecosistemas	<NA>	NA
## 11		Genetica Vegetal	<NA>	NA
## 12		Biofisica	<NA>	NA
## 13		Circuitos Electricos	<NA>	NA
##	Periodo MAT_25I			
## 1	2025-I	29		
## 2	2025-I	7		
## 3	2025-I	19		
## 4	2025-I	13		
## 5	2025-I	7		
## 6	2025-I	8		
## 7	2025-I	24		
## 8	2025-I	38		
## 9	2025-I	15		
## 10	2025-I	12		
## 11	2025-I	31		
## 12	2025-I	17		
## 13	2025-I	5		

2.14 Función left_join

Retorna las filas de x, además de todas las columnas de x e y. Completa con NA cuando es necesario.

¿Cómo se usa?

```
x |> left_join(y)
```

2.14.1 Ejemplo

Combinar los registros de ambos data frames usando left_join.

```
datos |>
  left_join(datos2, by = c("FACULTAD" = "Facultad",
                          "DEPARTAMENTO" = "Departamento",
                          "CURSO" = "curso")) -> datos_2025
datos_2025 |> head(5)
```

```
##   FACULTAD DEPARTAMENTO PERIODO CURSOS
## 1 Agronomia Entomologia 2025-II Apicultura General
## 2 Agronomia Entomologia 2025-II Control Biologico
## 3 Agronomia Entomologia 2025-II Crianza y Evaluacion de Insectos
## 4 Agronomia Entomologia 2025-II Entomologia Agricola
## 5 Agronomia Entomologia 2025-II Entomologia General
##   NRO_MATRICULADOS Periodo cantidad_matriculados
## 1                30 2025-I                    28
## 2                34 2025-I                     4
## 3                24 2025-I                     4
## 4               196 2025-I                   109
## 5               162 2025-I                   131
```

Dándole forma:

```
datos_2025 |>
  relocate(CURSO, .before = PERIODO) |>
  rename(MAT_25II = 5,
         MAT_25I = 7) -> datos_2025
datos_2025 |> head(5)
```

```
##   FACULTAD DEPARTAMENTO CURSOS PERIODO MAT_25II
## 1 Agronomia Entomologia Apicultura General 2025-II 30
## 2 Agronomia Entomologia Control Biologico 2025-II 34
## 3 Agronomia Entomologia Crianza y Evaluacion de Insectos 2025-II 24
## 4 Agronomia Entomologia Entomologia Agricola 2025-II 196
## 5 Agronomia Entomologia Entomologia General 2025-II 162
##   Periodo MAT_25I
## 1 2025-I 28
## 2 2025-I 4
## 3 2025-I 4
## 4 2025-I 109
## 5 2025-I 131
```

¿Qué cursos se dictaron solo en el ciclo 2025-II?

```
datos_2025 |>
  filter(is.na(MAT_25I)) |>
  head(13)
```

##	FACULTAD	DEPARTAMENTO		CURSO	PERIODO
## 1	Ciencias	Biologia			
## 2	Ciencias	Biologia			
## 3	Ciencias	Biologia			
## 4	Ciencias	Biologia			
## 5	Ciencias	Biologia			
## 6	Ciencias	Biologia			
## 7	Ciencias	Biologia			
## 8	Ciencias	Biologia			
## 9	Ciencias	Fisica y Meteorologia			
## 10	Ciencias	Fisica y Meteorologia			
## 11	Ciencias	Fisica y Meteorologia			
## 12	Ciencias	Fisica y Meteorologia			
## 13	Ciencias	Fisica y Meteorologia			
##					
## 1				Biodiversidad	2025-II
## 2				Comunicacion Cientifica	2025-II
## 3				Diversidad de Angiospermas	2025-II
## 4				Diversidad de Cordados	2025-II
## 5				Ecologia de Poblaciones	2025-II
## 6				Ecologia del Individuo	2025-II
## 7				Genetica de Poblaciones	2025-II
## 8				Herramientas para la Descripcion de los Ecosistemas	2025-II
## 9				Introduccion a la Inteligencia Artificial	2025-II
## 10				Meteorologia Tropical	2025-II
## 11				Observaciones Agrometeorologicas	2025-II
## 12				Sistema de Informacion Geografica para la Gestion de Riesgos	2025-II
## 13				Tecnicas de Pronostico del Tiempo	2025-II
##					
##	MAT_25II	Periodo	MAT_25I		
## 1	24	<NA>	NA		
## 2	20	<NA>	NA		
## 3	36	<NA>	NA		
## 4	41	<NA>	NA		
## 5	23	<NA>	NA		
## 6	31	<NA>	NA		
## 7	48	<NA>	NA		
## 8	18	<NA>	NA		
## 9	54	<NA>	NA		
## 10	18	<NA>	NA		
## 11	2	<NA>	NA		
## 12	8	<NA>	NA		
## 13	7	<NA>	NA		

2.15 Función full_join

Retorna todas las filas de x e y, hayan o no coincidencias. Completa con NA cuando es necesario.

¿Cómo se usa?

```
x |> full_join(y)
```

2.15.1 Ejemplo

Combinar los registros de ambos data frames usando full_join.

```
datos |>
  full_join(datos2, by = c("FACULTAD" = "Facultad",
                          "DEPARTAMENTO" = "Departamento",
                          "CURSO" = "curso")) -> datos_2025
datos_2025 |> head(5)
```

```
##   FACULTAD DEPARTAMENTO PERIODO CURSOS
## 1 Agronomia Entomologia 2025-II Apicultura General
## 2 Agronomia Entomologia 2025-II Control Biologico
## 3 Agronomia Entomologia 2025-II Crianza y Evaluacion de Insectos
## 4 Agronomia Entomologia 2025-II Entomologia Agricola
## 5 Agronomia Entomologia 2025-II Entomologia General
##   NRO_MATRICULADOS Periodo cantidad_matriculados
## 1                30 2025-I                    28
## 2                34 2025-I                     4
## 3                24 2025-I                     4
## 4               196 2025-I                   109
## 5               162 2025-I                   131
```

Dándole forma:

```
datos_2025 |>
  relocate(CURSO, .before = PERIODO) |>
  rename(MAT_25II = 5,
         MAT_25I = 7) -> datos_2025
datos_2025 |> head(5)
```

```
##   FACULTAD DEPARTAMENTO CURSOS PERIODO MAT_25II
## 1 Agronomia Entomologia Apicultura General 2025-II 30
## 2 Agronomia Entomologia Control Biologico 2025-II 34
## 3 Agronomia Entomologia Crianza y Evaluacion de Insectos 2025-II 24
## 4 Agronomia Entomologia Entomologia Agricola 2025-II 196
## 5 Agronomia Entomologia Entomologia General 2025-II 162
##   Periodo MAT_25I
## 1 2025-I 28
## 2 2025-I 4
## 3 2025-I 4
## 4 2025-I 109
## 5 2025-I 131
```

¿Qué cursos se dictaron solo en uno de los ciclos en la facultad de Industrias Alimentarias?

```
datos_2025 |>
  filter(FACULTAD == "Industrias Alimentarias" & (is.na(MAT_25I) | is.na(MAT_25II)))
```

```
##              FACULTAD DEPARTAMENTO              CURSO
## 1 Industrias Alimentarias      T.A.P.A. Tecnología de Alkoholes y Derivados
## 2 Industrias Alimentarias      T.A.P.A.  Tecnología de Frutas y Hortalizas
## 3 Industrias Alimentarias      T.A.P.A.              Enologia
## PERIODO MAT_25II Periodo MAT_25I
## 1 2025-II      20    <NA>      NA
## 2 2025-II      48    <NA>      NA
## 3    <NA>      NA 2025-I      26
```

3 Funciones de tidyr

Pivoteo se refiere al proceso de transformar la estructura de un conjunto de datos, reorganizando cómo se distribuyen las filas y columnas.

Hay dos grandes movimientos:

- `pivot_longer`: Convierte columnas en filas (hace el dataset “más largo”)
- `pivot_wider`: Convierte filas en columnas (hace el dataset “más ancho”)

Se necesita cargar el paquete `tidyr`. Más información sobre este paquete aquí.

3.1 Función `pivot_longer`

¿Cómo se usa `pivot_longer`?

```
df |> pivot_longer(cols = Col1, names_to = Col2, values_to = Col3)
```

Col1 = Columnas que se transformarán en filas

Col2 = Columna que almacenará los nombres originales de las columnas

Col3 = Columna que almacenará los valores

3.1.1 Ejemplo

```
datos_2025 |> filter(FACULTAD == "Industrias Alimentarias") |> head(9)
```

```
##          FACULTAD DEPARTAMENTO
## 1 Industrias Alimentarias I.A.P.A.
## 2 Industrias Alimentarias I.A.P.A.
## 3 Industrias Alimentarias I.A.P.A.
## 4 Industrias Alimentarias I.A.P.A.
## 5 Industrias Alimentarias I.A.P.A.
## 6 Industrias Alimentarias I.A.P.A.
## 7 Industrias Alimentarias I.A.P.A.
## 8 Industrias Alimentarias I.A.P.A.
## 9 Industrias Alimentarias I.A.P.A.
##          CURSO PERIODO MAT_25II Periodo MAT_25I
## 1          Administracion 2025-II      49 2025-I      49
## 2          Agronegocios 2025-II      64 2025-I      26
## 3 Alimentacion y Nutricion Humana 2025-II      50 2025-I      47
## 4 Alimentos Funcionales y Nutraceuticos 2025-II       7 2025-I      19
## 5 Analisis de Alimentos 2025-II     123 2025-I      43
## 6 Contabilidad y Finanzas 2025-II      50 2025-I      49
## 7 Control de Calidad e Inocuidad de Alimentos 2025-II      99 2025-I      33
## 8 Dise\xfflo en Ingenieria 2025-II      50 2025-I      18
## 9 Envases y Embalajes de Alimentos 2025-II      32 2025-I      25
```

```
library(tidyr)
datos_2025 |>
  select(-PERIODO, -Periodo) |>
  filter(FACULTAD == "Industrias Alimentarias") |>
  pivot_longer(cols = c("MAT_25I", "MAT_25II"),
               names_to = "Ciclo",
               values_to = "Alumnos") |>
  mutate(Ciclo = ifelse(Ciclo == "MAT_25I", "2025-I", "2025-II")) -> CURSOS_IA
CURSOS_IA |> head(9)
```

```
## # A tibble: 9 x 5
##   FACULTAD DEPARTAMENTO CURSO          Ciclo Alumnos
##   <chr>      <chr>      <chr>      <chr>    <int>
## 1 Industrias Alimentarias I.A.P.A. Administracion 2025~      49
## 2 Industrias Alimentarias I.A.P.A. Administracion 2025~      49
## 3 Industrias Alimentarias I.A.P.A. Agronegocios 2025~      26
## 4 Industrias Alimentarias I.A.P.A. Agronegocios 2025~      64
## 5 Industrias Alimentarias I.A.P.A. Alimentacion y Nutricion H~ 2025~      47
## 6 Industrias Alimentarias I.A.P.A. Alimentacion y Nutricion H~ 2025~      50
## 7 Industrias Alimentarias I.A.P.A. Alimentos Funcionales y Nu~ 2025~      19
## 8 Industrias Alimentarias I.A.P.A. Alimentos Funcionales y Nu~ 2025~       7
## 9 Industrias Alimentarias I.A.P.A. Analisis de Alimentos 2025~      43
```

3.2 Función pivot_wider

¿Cómo se usa pivot_wider?

```
df |> pivot_longer(names_from = Col1, values_from = Col2)
```

Col1 = Columna que contiene las categorías que se convertirán en nuevos nombres de columnas

Col2 = Columna que contiene los valores que se distribuirán en las nuevas columnas

pivot_wider() realiza el proceso inverso a pivot_longer(), es decir, convierte filas en columnas.

3.2.1 Ejemplo

```
CURSOS_IA |>
  filter(DEPARTAMENTO == "T.A.P.A.") |>
  pivot_wider(names_from = Ciclo, values_from = Alumnos)
```

```
## # A tibble: 17 x 5
##   FACULTAD          DEPARTAMENTO CURSO          `2025-I` `2025-II`
##   <chr>            <chr>      <chr>          <int>    <int>
## 1 Industrias Alimentarias T.A.P.A. Fermentaciones Indus~    34      70
## 2 Industrias Alimentarias T.A.P.A. Industrias Carnicas    11      10
## 3 Industrias Alimentarias T.A.P.A. Industrias Lacteas    12      11
## 4 Industrias Alimentarias T.A.P.A. Introduccion a la In~    65      54
## 5 Industrias Alimentarias T.A.P.A. Seminario en Industr~    23      37
## 6 Industrias Alimentarias T.A.P.A. Seminario en Industr~    33      19
## 7 Industrias Alimentarias T.A.P.A. Taller de Emprendimi~     8      17
## 8 Industrias Alimentarias T.A.P.A. Tecnologia de Aceite~    17      20
## 9 Industrias Alimentarias T.A.P.A. Tecnologia de Alcoho~    NA      20
## 10 Industrias Alimentarias T.A.P.A. Tecnologia de Alimen~    29      66
## 11 Industrias Alimentarias T.A.P.A. Tecnologia de Azucar~    32      25
## 12 Industrias Alimentarias T.A.P.A. Tecnologia de Frutas~    NA      48
## 13 Industrias Alimentarias T.A.P.A. Tecnologia de Legumi~    15       8
## 14 Industrias Alimentarias T.A.P.A. Tecnologia Poscosecha    39      80
## 15 Industrias Alimentarias T.A.P.A. Tecnologias de Conse~    42      76
## 16 Industrias Alimentarias T.A.P.A. Tecnologias de Extra~    36      74
## 17 Industrias Alimentarias T.A.P.A. Enologia                26      NA
```


4 Resumen de funciones para preparación de datos

- **select**: permite seleccionar o retirar una o más columnas de un data frame.
- **pull**: permite seleccionar una columna de un data frame.
- **mutate**: permite crear una nueva columna de un data frame o modificar una existente.
- **rename**: permite renombrar una o más columnas de un data frame.
- **relocate**: cambia el orden de las columnas.
- **slice**: selecciona filas específicas.
- **filter**: selecciona filas según una o más reglas.
- **arrange**: ordena las filas de un data frame según los valores de una o más columnas.
- **summarise**: resume variables (por ejemplo, media, desviación estándar, conteo).
- **group_by**: agrupa datos por una o más variables para análisis agregados.
- **inner_join(x,y)** retorna las filas de x que tienen valores coincidentes en y, considerando todas las columnas de x e y
- **right_join(x,y)** retorna las filas de y, además de todas las columnas de x e y. Completa con NA cuando es necesario.
- **left_join(x,y)** retorna las filas de x, además de todas las columnas de x e y. Completa con NA cuando es necesario.
- **full_join(x,y)** retorna todas las filas de x e y, hayan o no coincidencias. Completa con NA cuando es necesario.
- **pivot_longer**: Convierte columnas en filas (hace el dataset “más largo”)
- **pivot_wider**: Convierte filas en columnas (hace el dataset “más ancho”)

5 Ejercicios

1. ¿Cuál es la cantidad total de matrículas por facultad?
2. ¿Cuál es el porcentaje de cursos “grandes” (≥ 200 matriculados) por facultad?
3. ¿Cuál es el rango (máx – mín) de matriculados por departamento académico?
4. ¿Cuál es la cantidad de total de matrículas por departamento académico en el año 2025?