



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

Programación en R para Ciencia de Datos

Miguel Jorquera

DBDC-202010

Educación Profesional
Escuela de Ingeniería



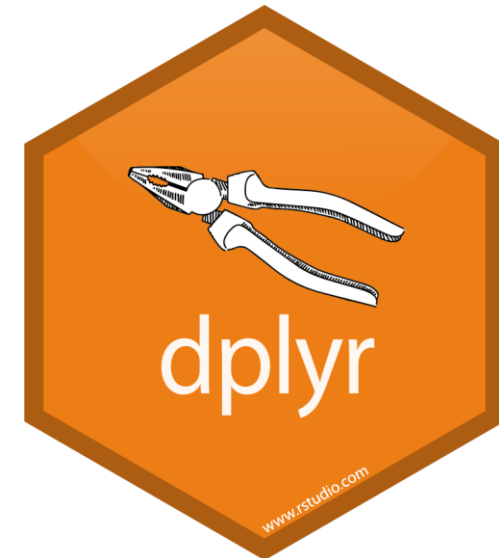
ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

RESUMEN

Manipulación de tablas

- Análisis exploratorio de datos
 - `dplyr`: Manipulación de tablas



Camino a la manipulación de tablas

- Si bien hay variedad de herramientas para llevar a cabo la fase exploratoria, nosotros nos centraremos en la utilización de dos packages principalmente
 - `dplyr` para consultas
 - Generación de información agregada.
 - Tablas de frecuencia.
 - Facilita el cálculo de estadísticos descriptivos en general



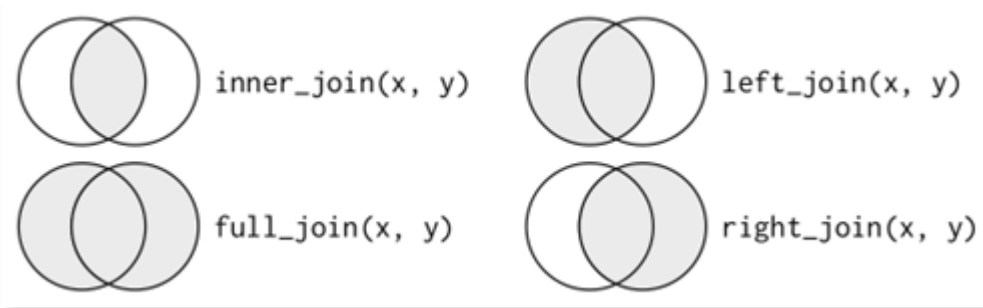
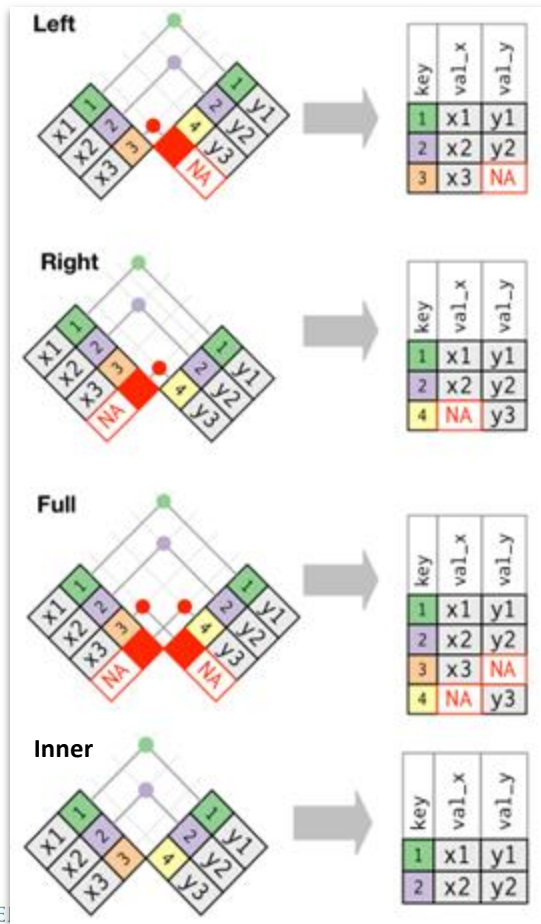
Camino a la manipulación de tablas

- `filter()` : Para filtrar los datos (por filas).
- `arrange()` : Para ordenar un dataset.
- `select()` : Para seleccionar y renombrar columnas.
- `mutate()` : Para crear (o modificar) columnas.
- `group_by()` : Para agrupar tablas
- `summarise()` : Para generar medidas agregadas.
- `sample_n()` y `sample_frac()` : Para generar muestras aleatorias.



Camino a la manipulación de tablas

- Datos relacionales y cruce entre tablas



dplyr	merge
inner_join(x, y)	merge(x, y)
left_join(x, y)	merge(x, y, all.x = TRUE)
right_join(x, y)	merge(x, y, all.y = TRUE)
full_join(x, y)	merge(x, y, all.x = TRUE, all.y = TRUE)

dplyr	SQL
inner_join(x, y, by = "z")	SELECT * FROM x INNER JOIN y USING (z)
left_join(x, y, by = "z")	SELECT * FROM x LEFT OUTER JOIN y USING (z)
right_join(x, y, by = "z")	SELECT * FROM x RIGHT OUTER JOIN y USING (z)
full_join(x, y, by = "z")	SELECT * FROM x FULL OUTER JOIN y USING (z)





ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

TEMAS PARA HOY



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

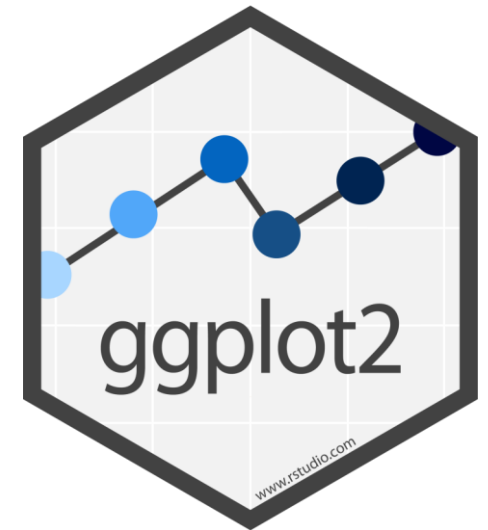
INTRODUCCIÓN A GGPLOT

ggplot2

Ggplot2 es un package basado en la gramática de gráficos, creada por Wilkinson (2015). Básicamente la gramática de gráficos nos dice que un gráfico estadístico corresponde a un mapeo entre la data hacia atributos estéticos (como el color, forma o tamaño) de objetos geométricos (como puntos, líneas o barras).

Referencias útiles

- <https://ggplot2.tidyverse.org/reference/index.html>
- <https://r4ds.had.co.nz/data-visualisation.html>
- <https://www.rdocumentation.org/packages/ggplot2/versions/3.3.0>



<https://ggplot2.tidyverse.org/>



Otras referencias

highcharter

- Official package website: <http://jkunst.com/highcharter>
- Replicating Highcharts Demos: <https://cran.rstudio.com/web/packages/highcharter/vignettes/replicating-highcharts-demos.html>
- CRAN site: <https://cran.r-project.org/web/packages/highcharter/>.
- Shiny demo code: <https://github.com/jbkunst/shiny-apps/tree/master/highcharter>.
- Referencia oficial (No R): <http://highcharts.com>

leaflet

- <https://rstudio.github.io/leaflet/>
- <https://www.rdocumentation.org/packages/leaflet/versions/2.0.3>
- <https://github.com/rstudio/leaflet>
- Referencia oficial (No R): <https://leafletjs.com/reference-1.6.0.html>

Shiny

- <https://shiny.rstudio.com/>
- <https://www.shinyapps.io/>



Vamos!

