

# Coding e Big Data

2025-2026

Vincenzo Nardelli



[vincenzo.nardelli@unicatt.it](mailto:vincenzo.nardelli@unicatt.it)

# Tidyverse



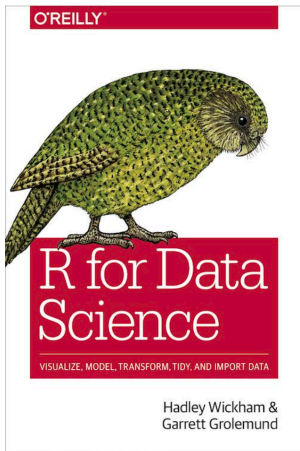
## R packages for data science

The tidyverse is an opinionated **collection of R packages** designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

Install the complete tidyverse with:

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

# Libro Suggestito



Versione italiana: <https://it.r4ds.hadley.nz/>

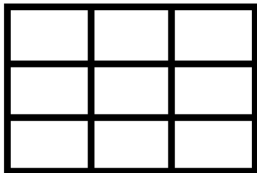
Versione inglese: <https://r4ds.hadley.nz/>



# Cosa è dplyr?

dplyr è una **grammatica** della manipolazione dei dati, che fornisce un **insieme coerente di verbi** che ti aiutano a risolvere le sfide più comuni di manipolazione dei dati

**d** = data, data.frame



**plyr** = pinze (pliers)



# Dplyr - funzioni base

## **select()**



Subset variabili  
(colonne)

## **filter()**



Subset osservazioni  
(righe)

## **mutate()**



Crea nuove variabili

## **summarise()**



Riassumi

# Dplyr - group by



**group\_by()**  
Raggruppa per

+

**mutate()**  
Crea nuove variabili



**summarise()**  
Riassumi



LAB

## Analisi dati Ecommerce

