

Introduzione al linguaggio di programmazione open source R

...

Sommario

- R software
- Programmazione e R (recap)
- RStudio IDE (Integrated Development Environment)
- Connessione al DB
- Lettura e scrittura
- Il pacchetto Tidyverse

Benefici

- Si basa sull'approccio Free-licensing
- Si integra in modo facile
- Big-data compliant
- Il tools più conveniente in termine di costi
- Una varietà infinita di soluzioni statistiche
- Customizzabile
- It brings the state of the art research to the world in real time
- Utilizzato dai maggiori data providers
- Utilizzato dalle maggiori università al mondo

R - linguaggio di programmazione

- Free
- Copy and re-use
- Accedere ai sorgenti e modificarli
- Ricercatori forniscono funzioni tramite pacchetti
- Grafica eccellente
- Eccellente per simulazione, programmazione, computer intensive analysis
- Si interfaccia ai database
- Community molto attiva

Svantaggi

- Curva di apprendimento inizialmente lenta
- Non ha una GUI
- Facile sbagliare
- L'uso di dataset grandi è limitato dalla RAM

Architettura

- 32 e 64 bit
 - driver ODBC
 - Java
 - Installazione di R
- Multiplatforma
 - Windows
 - MacOS
 - Linux

Oggetti

- Vettori
 - numerici: `num=c(1,2,45)`
 - carattere: `car=c("one","two","three")`
 - logici: `num>3` oppure `c(FALSE,FALSE,TRUE)`
 - misti: `mixed=c("foo",1,TRUE)` ma diventano carattere
 - vettori di vettori: `c(num,car)`
- Matrici
 - `matrix()`, `cbind()`, `rbind()`
- Liste (collezioni di oggetti)
- Fattori (classificazioni discrete per l'aggregazione)
 - `factor()`
- Data Frames
 - `data.frame()`
 - Matrice per colonne fattori o vettori della stessa lunghezza ma di tipi diversi
 - Lista di vettori di qualsiasi tipo o di fattori tutti della stessa lunghezza

Valutazione espressioni

- Simple evaluation

- costanti `1`
- assegnazione `y<-1`
- richiamo funzioni `mean(1:5)`
- operatori `1+1` oppure `y <- 2 * (a + log(x))`

- Control Structure

- if
- for
- while

```
for ( name in vector )  
  statement1
```

```
while ( statement1 ) statement2
```

```
if ( statement1 ) {  
  statement2  
} else if ( statement3 ) {  
  statement4  
} else if ( statement5 ) {  
  statement6  
} else  
  statement8
```

- Indexing

```
x[i]  
x[i, j]  
x[[i]]  
x[[i, j]]  
x$a  
x$a"
```

- Scope of variabile

- Environment
- Global environment
-

Mappa i nome dei valori alle variabili
Livello workspace

Assegnazione

Tre modi di assegnare un valore ad una variabile

- `=`
 - valutata nell'ambiente di valutazione della funzione
- `<-`
 - valutata nel local environment
- `<<-`
 - utilizzata per assegnare nell'parent environment

- Assegnazioni multiple: `y <- yy <- 0`

```
1 | median(x = 1:10)
1 | ## [1] 5.5
1 | exists("x")
1 | ## [1] FALSE
1 | median(x <- 1:10)
1 | ## [1] 5.5
1 | exists("x")
1 | ## [1] TRUE
```

Buona programmazione

- Commentare in R #
- Indentare
- Utilizzare le funzioni - raccogliere azioni ripetitive
- Documentare
- Versioning

Environment

- Composta
 - Frame: contenitore di oggetti
 - Enclosed Environment: Il proprietario dell'ambiente
 - Associare nomi ad oggetti
- Perché environment
 - richiamare oggetti nell'environment corretto
 - evitare il masking (utilizzo della funzione corretta

L'Environment è come una lista con alcune eccezioni, in un environment:

- Ciascun nome è unico
- I nomi non sono ordinati
- ha un parent

Workspace==R_GlobalEnv

Valutazione espressioni - Environment

- Scope of variabile

- Frame
 - Insieme di oggetti
- Environment
 - Una porzione di memoria che contiene un frame
 - Mappa i nome dei valori alle variabili
 - Ciascun environment può accedere ad uno e un solo enviroment chiamato parent environment
- Ogni **sessione R ha un environment associato** chiamato Global environment
- R usato interattivamente usa una Frame nel GlobalEnv come container degli oggetti
- Ogni pacchetto ha uno e un solo environment
- Ogni funzione ha un environment

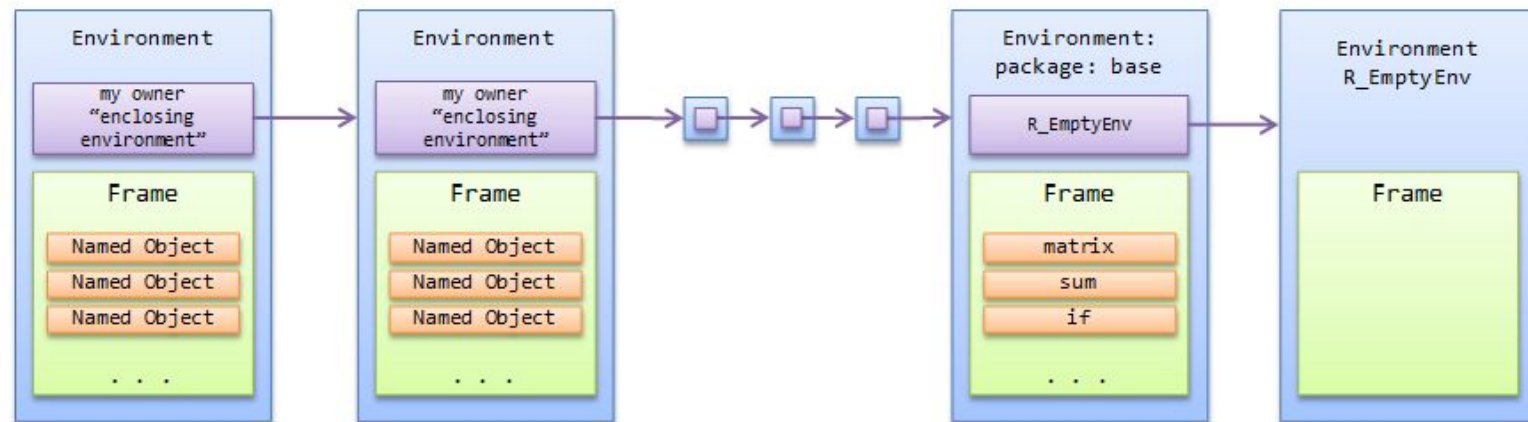
- Perchè environment

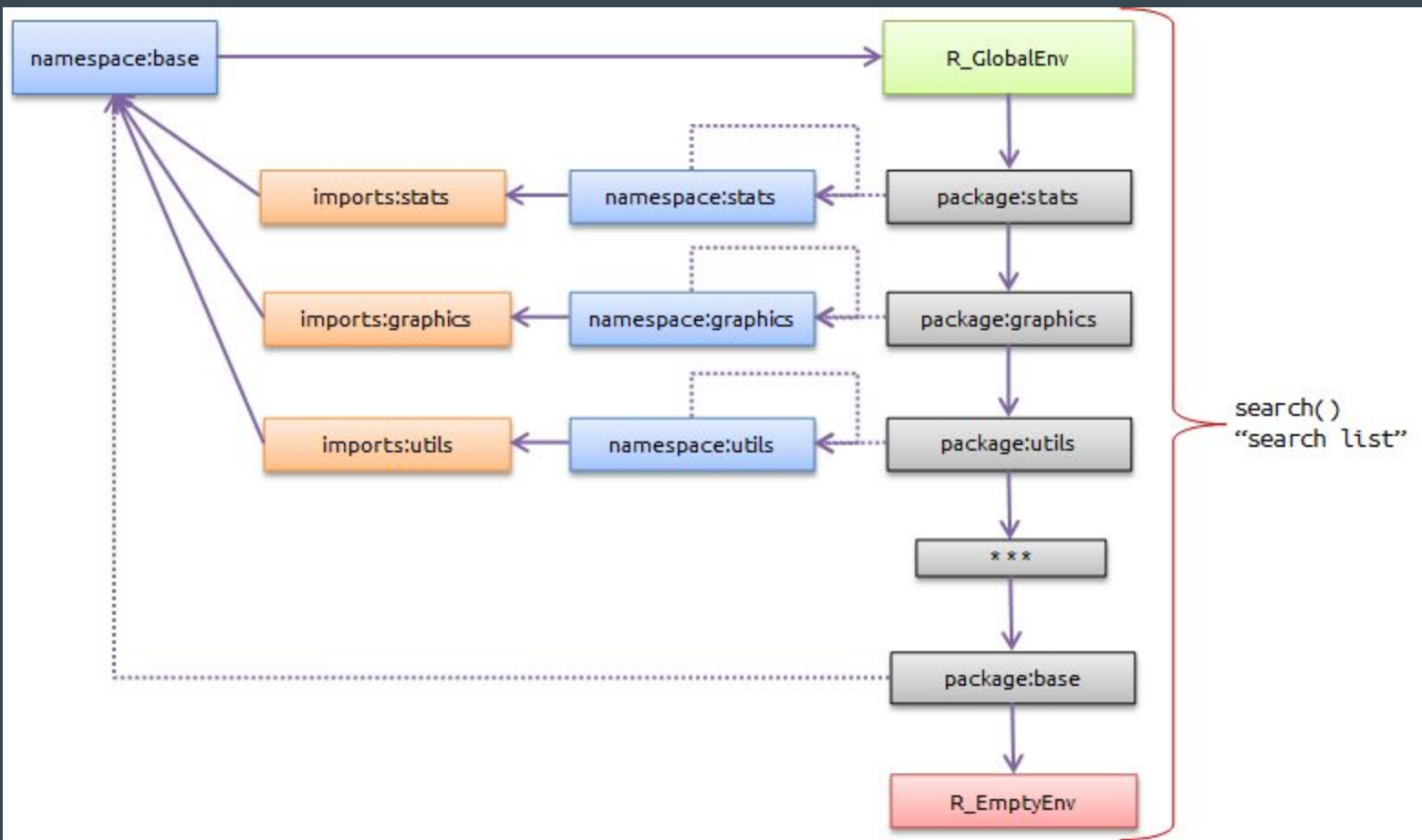
- richiamare oggetti nell'environment corretto
- evitare il masking

L'Environment è come una lista con alcune eccezioni, in un environment:

- Ciascun nome è unico
- I nomi non sono ordinati
- ha un parent

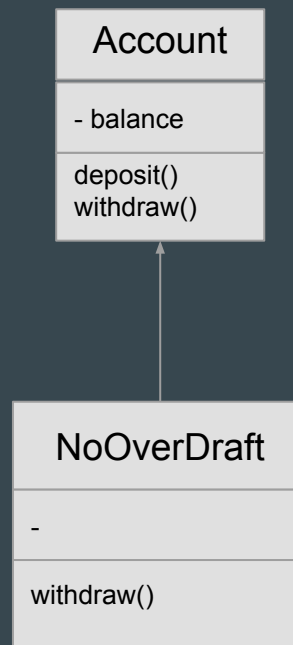
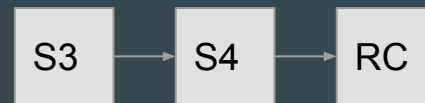
Workspace==R_GlobalEnv





Object oriented programming

- Classe
 - Definizione astratta del mondo reale
- Metodo
 - Azioni sulla classe
- Oggetto
 - Implementazione della classe

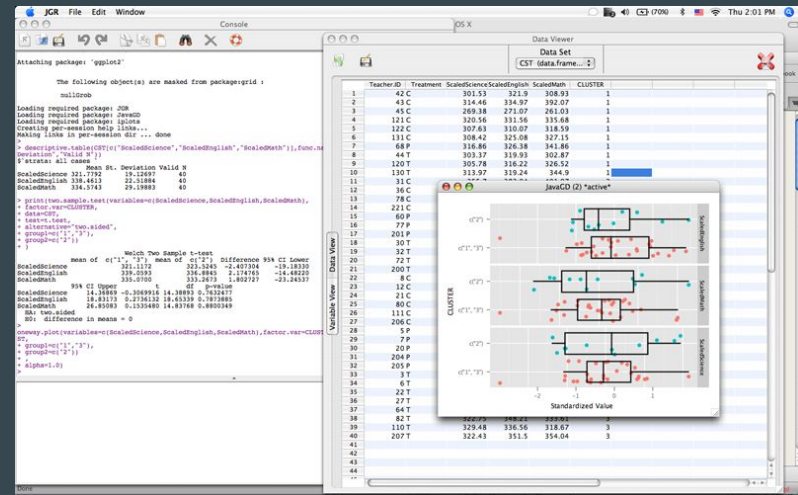
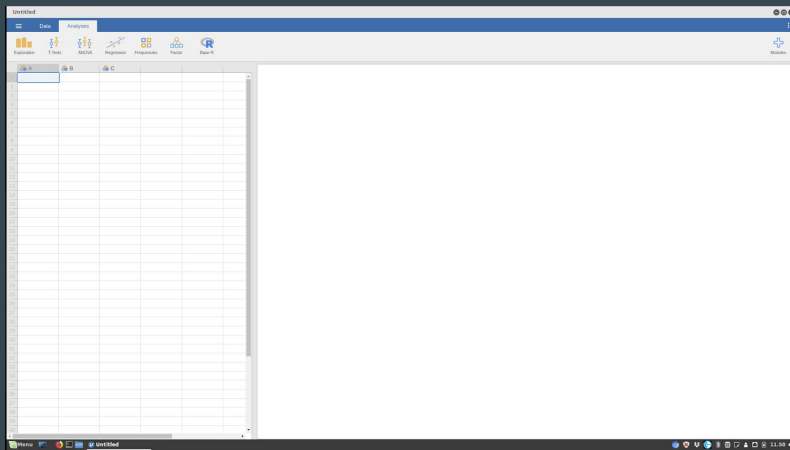


Rstudio

Rstudio

- Rstudio
 - Un IDE non una GUI
- GUI
 - R-commander
 - Deducer (JGR)
 - Red-r
 - Jamovi

• • •



Jamovi

IDE



Cosa significa

Integrated Development Environment

Lavoro più fluido

- Integra l'help e la documentazione di R
- E' possibile gestire diversi ambiente di lavoro (manage working directories utilizzando i Progetti)
- Si visualizza il workspace e i dati nello stesso ambiente

Sviluppato per R

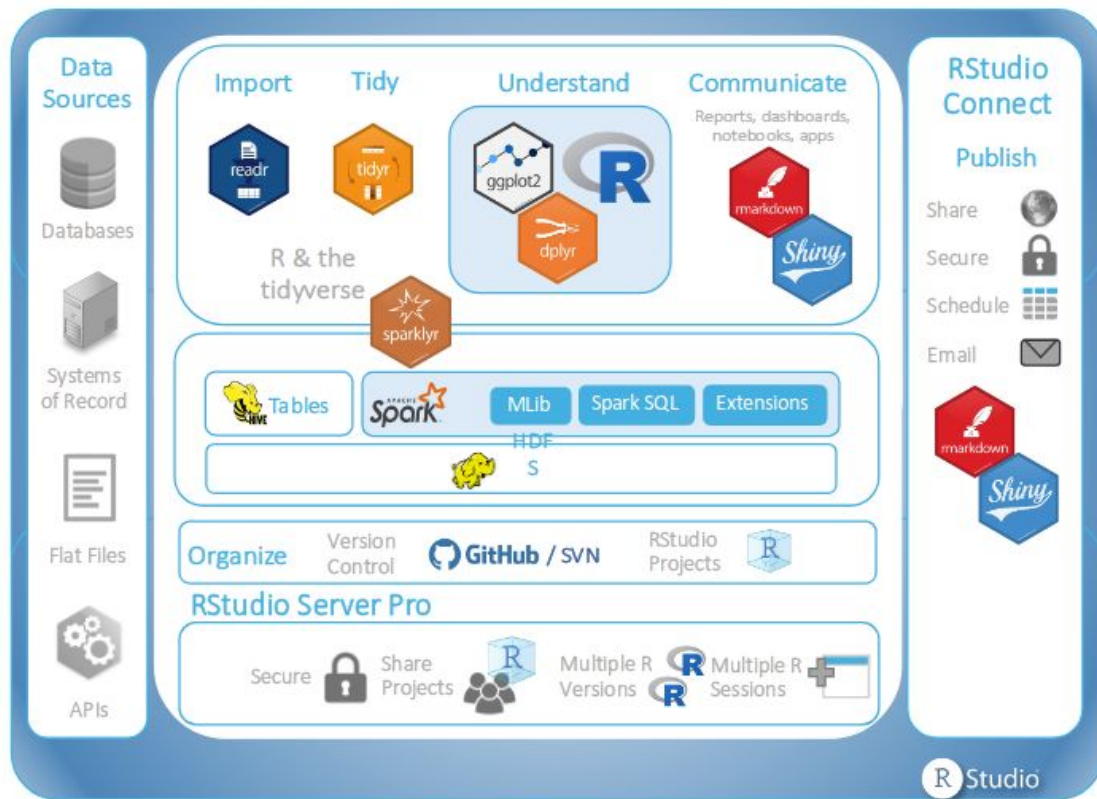
- Syntax highlighting, code completion, and smart indentation
- Execute R code directly from the source editor
- Quickly jump to function definitions

Sviluppo

- Interactive debugger to diagnose and fix errors quickly
- Extensive package development tools
- Authoring with Sweave and R Markdown

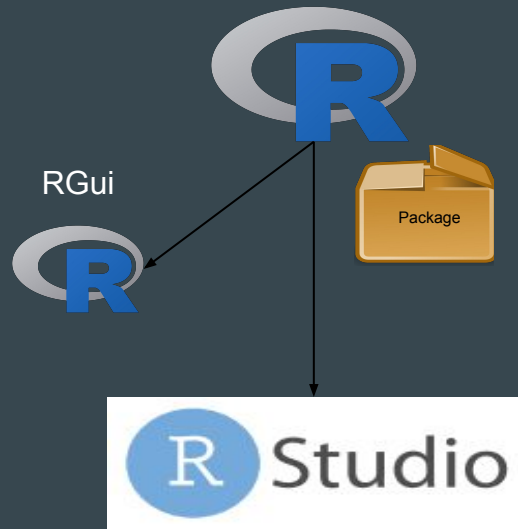
RStudio Toolchain Expands R into Industry

Enables
Enterprise
adoption
of R



- Free professional licenses for teaching
- Steeply discounted professional licenses for academic research

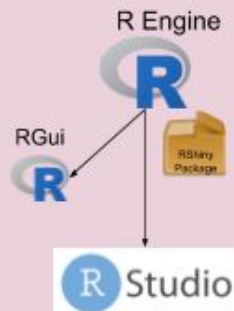
R Engine



Ambienti

Sviluppo

Desktop



Server Test



Preproduzione

Server Test

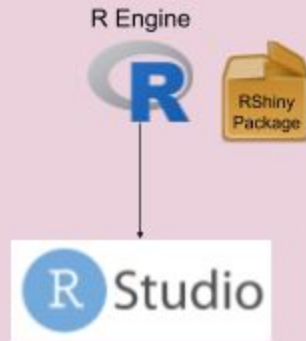


Produzione

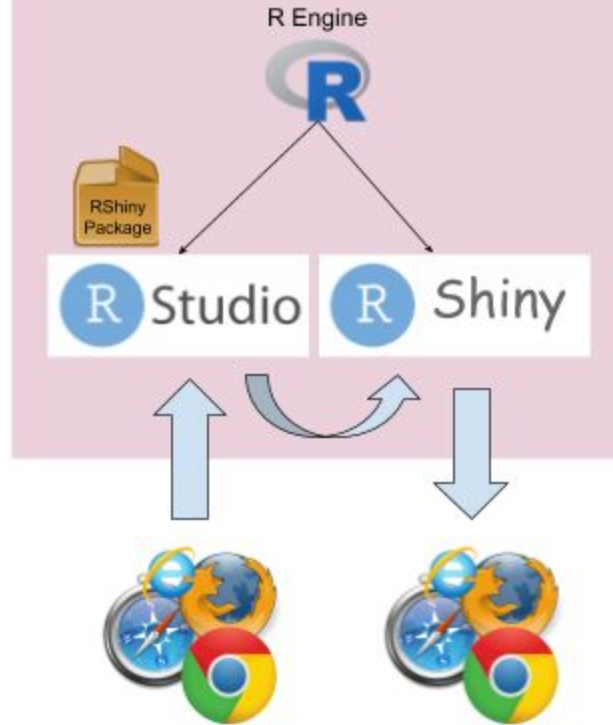
Server Produzione



Desktop



Server



Lavoro più fluido

- Layout
- Progetti
- File
- Installa pacchetti (per esempio jsonlite)
- Help
- Environment

Sviluppo

- Debug
- Profiling
- History
- Connection

Sviluppato per R

- Aiuti nello scrivere codice
- Editor
- Refactoring
- Pacchetti
 - Librerie di sistema
 - Librerie utente

Pubblicazione

- Markdown
- Presentazioni
- Html
- Sweave

Esercitazione

- Scaricare il progetto dal seguente indirizzo
 - <https://github.com/omargiorgetti/corsoR>
- Creazione del progetto CorsoR dallo zip
- Chiudere il progetto
- Creazione di un progetto Normale da directory
- Chiudere e riaprire il progetto CorsoR
- Installazione di un pacchetto
 - Installare il pacchetto tidyverse
- Spostare file tra le cartelle
 - Creare una cartella imieiesercizi
 - spostare i file

Tidyverse

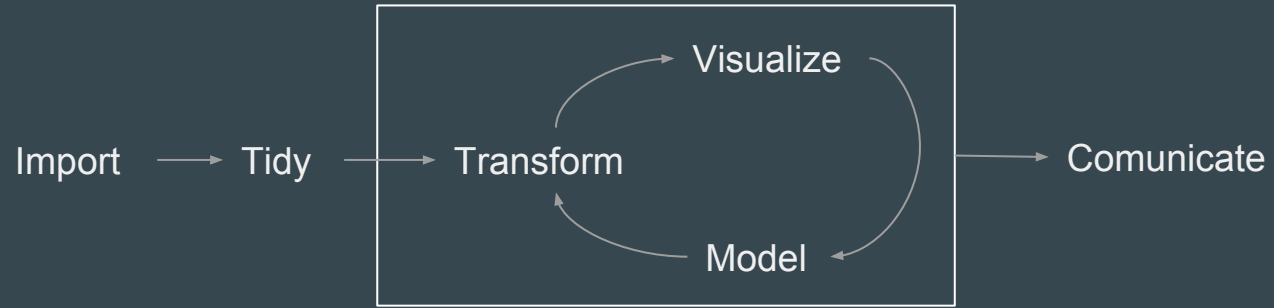


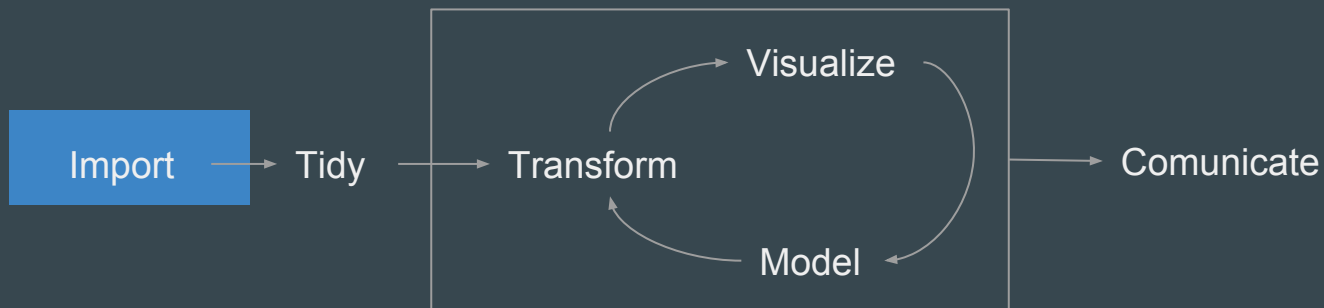
R packages for data science

The tidyverse is an opinionated **collection of R packages** designed for data science. All packages share an underlying philosophy and common APIs.

Install the complete tidyverse with:

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





Wrangle

Importare dati in R e disporli nella forma adeguata per visualizzazione e modellazione

Import

- File
 - Excel
 - OpenDocument
 - Testo
 - etc...
- Database
- Web

Esercizio 1

- Leggere i file csv nella cartella esercizi
- Contare il numero di record che devono essere sempre 3483
- provare ad usare sia le funzioni di utils che di tidyverse

Connessione ai db

Tipologie di connessione

- ODBC
 - Open DataBase Connectivity
- JDBC
 - Java DataBase Connectivity

Tipo	Pacchetto R
ODBC	RODBC
Jdbc	RJDBC
Driver db2	ibmdbR

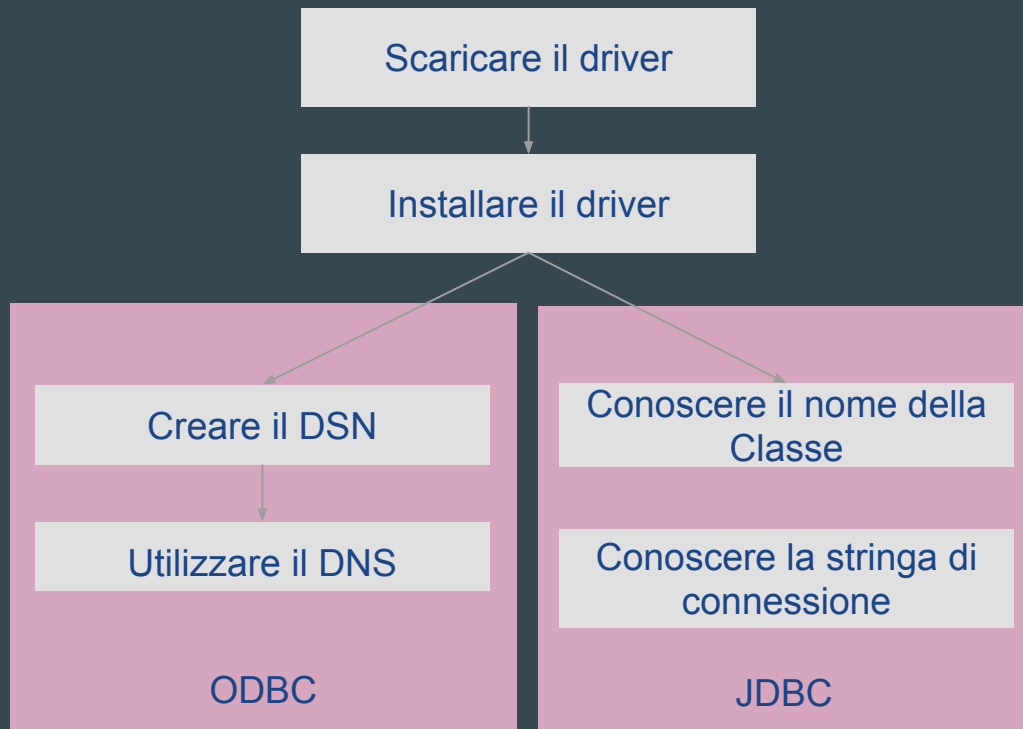
Connessione a db

- Driver

- Connettore
- Software per la connessione
- Scelta in funzione del DBMS utilizzato e alla tipologia

- Conoscere

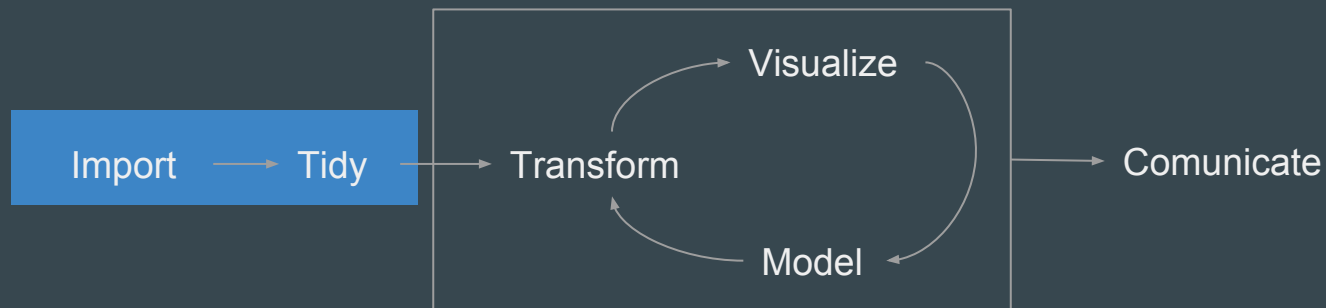
- Host
- Porta
- Utente
- Password



DSN
(Database Source Name)

Accesso al db

- In-database function
 - RDBL package -> generico
 - R services -> connessione a MS sql server
 - dbplyr



Wrangle

Importare dati in R e disporli nella forma adeguata per visualizzazione e modellazione

Import

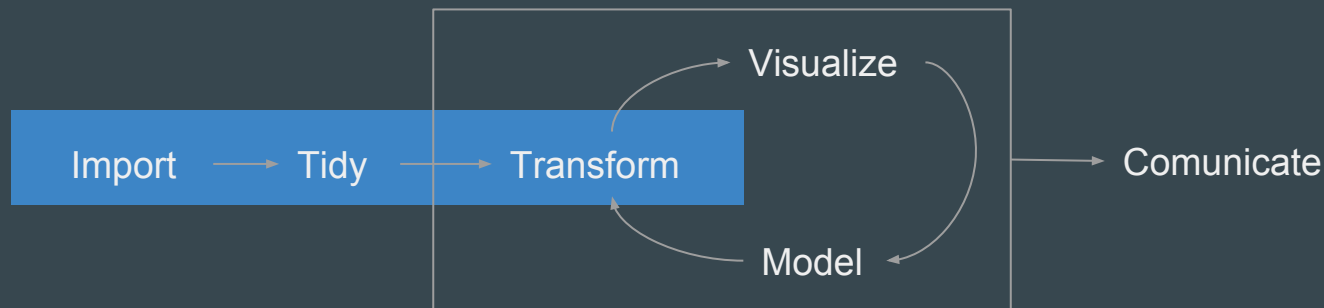
- File
 - Excel
 - OpenDocument
 - Testo
 - etc...
- Database
- Web

Tidy

- Gathering
- Spreading
- Separate
- Unite

Esercizio 2

- convertire la table4a come la table1 (per la misura cases)
- convertire la table2 come la table1



Wrangle

Importare dati in R e disporli nella forma adeguata per visualizzazione e modellazione

Import

- File
 - Excel
 - OpenDocument
 - Testo
 - etc...
- Database
- Web

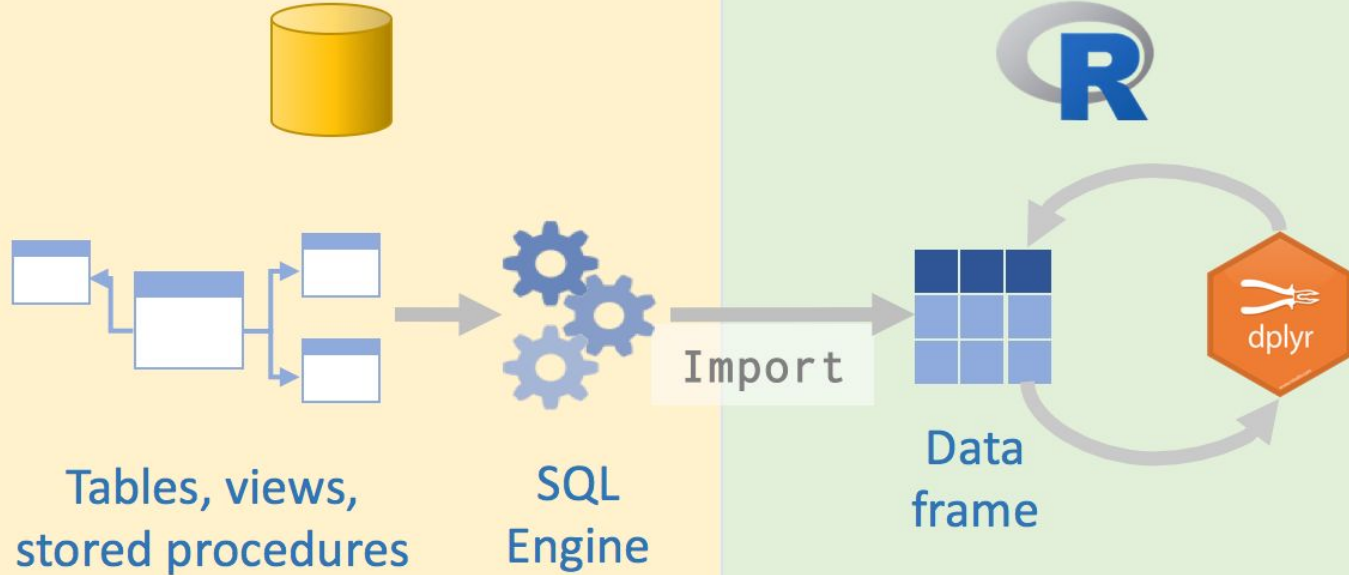
Tidy

- Gathering
- Spreading
- Separate
- Unite

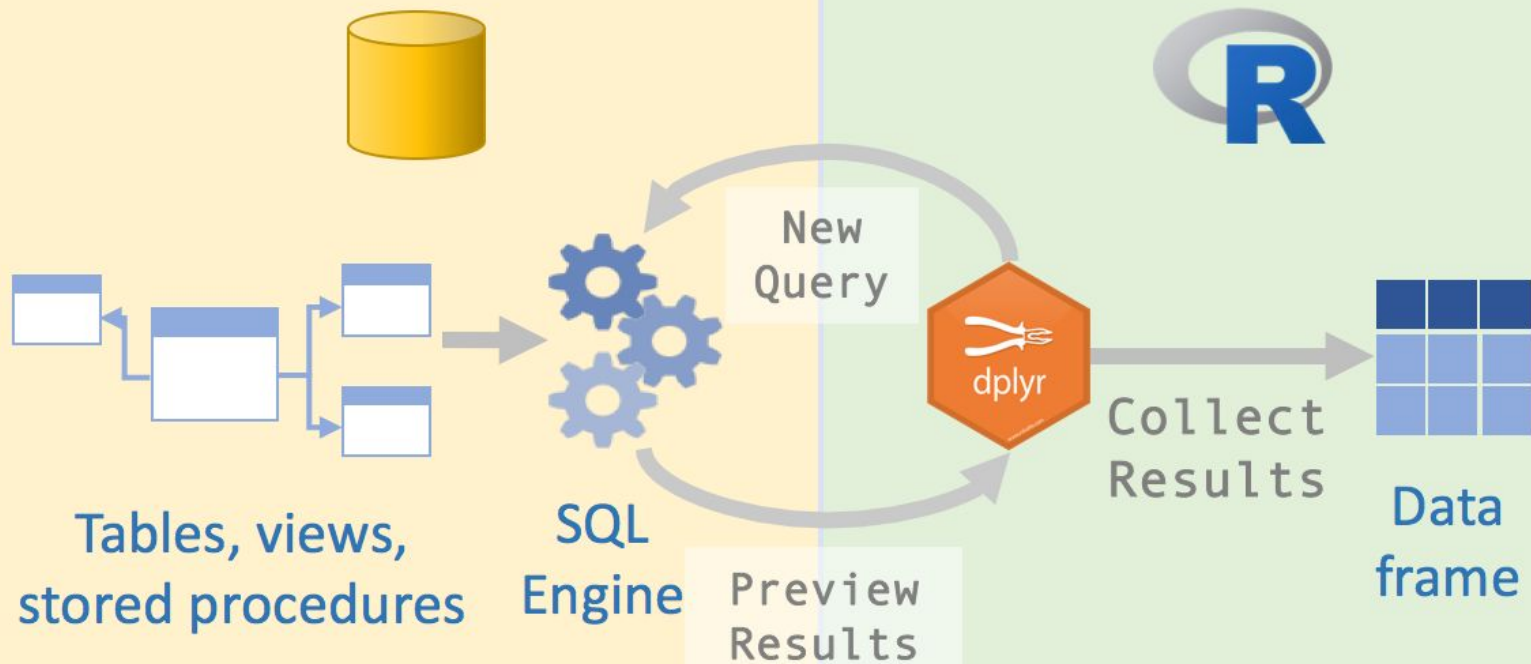
Transform

- dplyr
- forcats
- stringr
- lubridate

Reading the data all at once

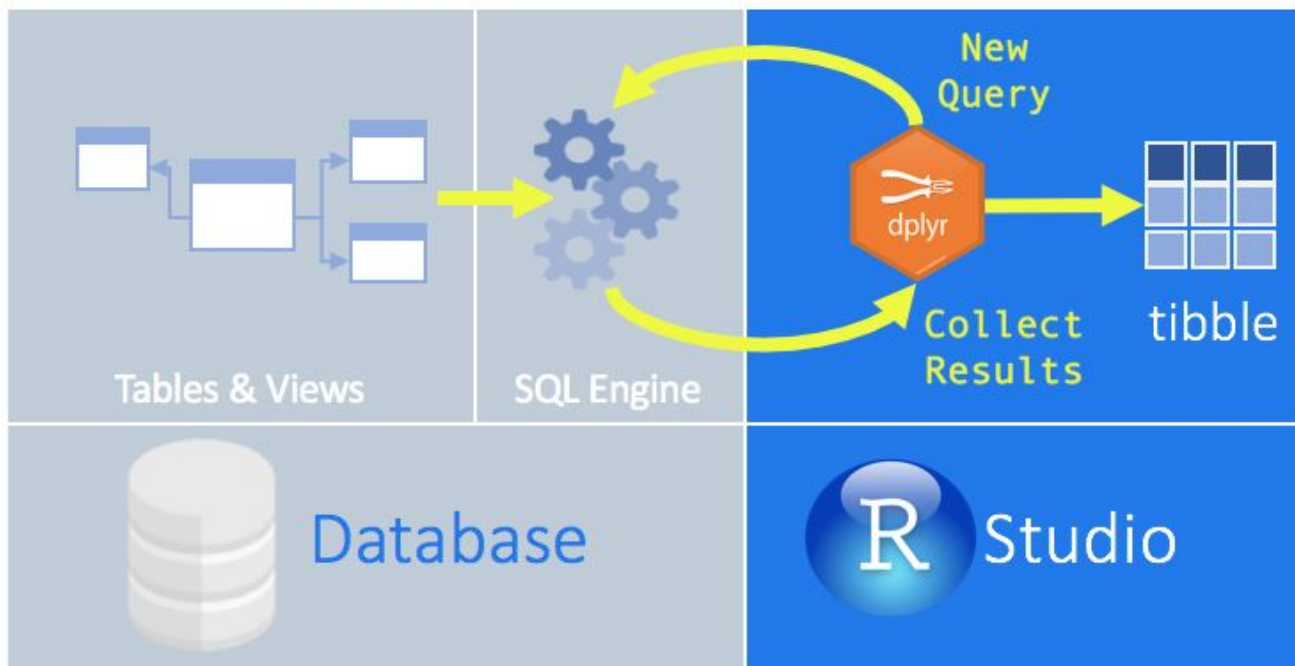


Using dplyr to interact with the database



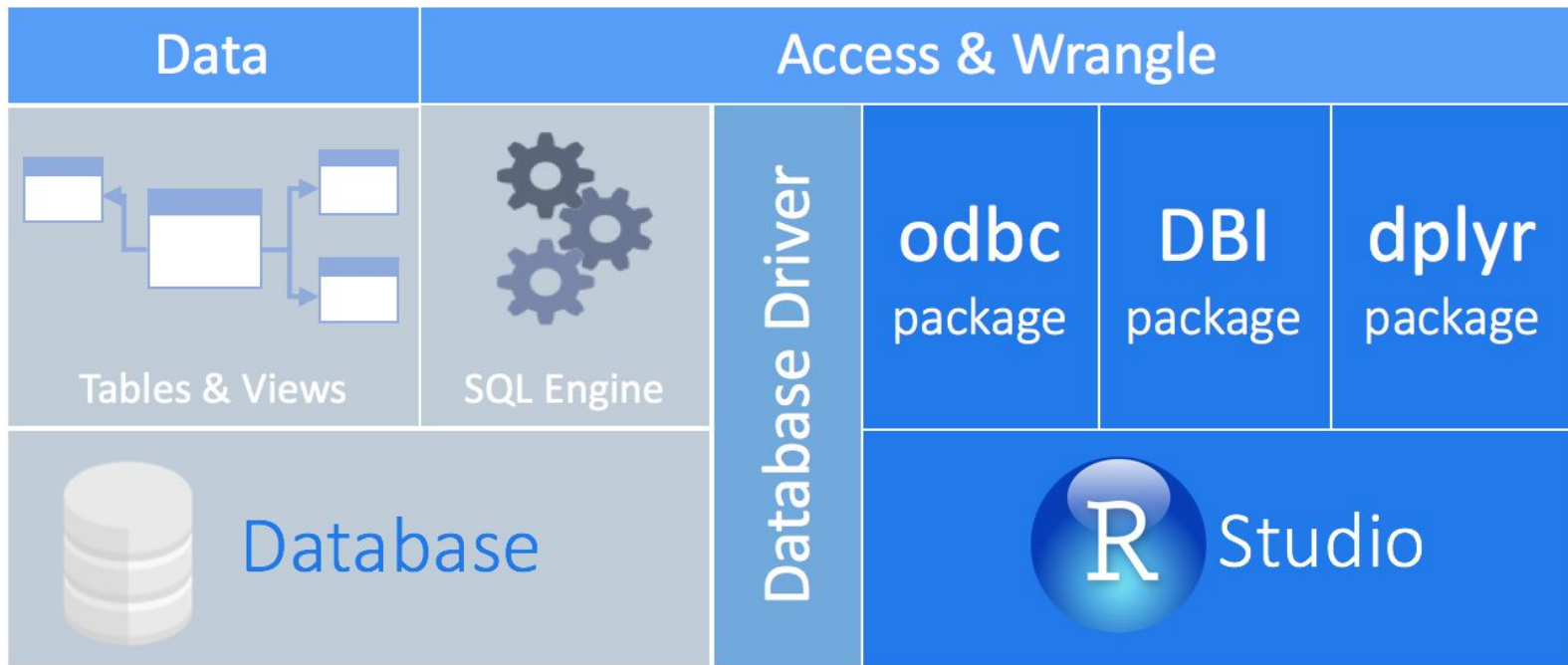
dplyr

Use dplyr to interact with the database



dplyr

Commercial Databases



dplyr

Grammatica per la manipolazione dei dati

Mette a disposizione i seguenti verbi:

- `mutate()` : aggiunge nuove variabili come funzioni o variabili stesse
- `select()` : seleziona le variabili in base al nome
- `filter()` : seleziona i valori in base al valore
- `summarise()` aggrega i valori rispetto ad una variabile
- `arrange()` cambia l'ordine delle variabili

combinare con `group_by()`

magrite %>%

Semplificazione della lettura

function(data) diventa data %>% function()

function(data1,data2) diventa data1 %>% function(data2)

Inventory

film	
film_id SMALLINT	
title VARCHAR(255)	
description TEXT	
release_year YEAR	
language_id TINYINT	
original_language_id TINYINT	
rental_duration TINYINT	
rental_rate DECIMAL(4,2)	
length SMALLINT	
replacement_cost DECIMAL(5,2)	
rating ENUM(...)	
special_features SET(...)	
last_update TIMESTAMP	
Indexes	
Triggers	

film_category	
film_id SMALLINT	
category_id TINYINT	
last_update TIMESTAMP	
Indexes	

category	
category_id TINYINT	
name VARCHAR(25)	
last_update TIMESTAMP	
Indexes	

language	
language_id TINYINT	
name CHAR(20)	
last_update TIMESTAMP	
Indexes	

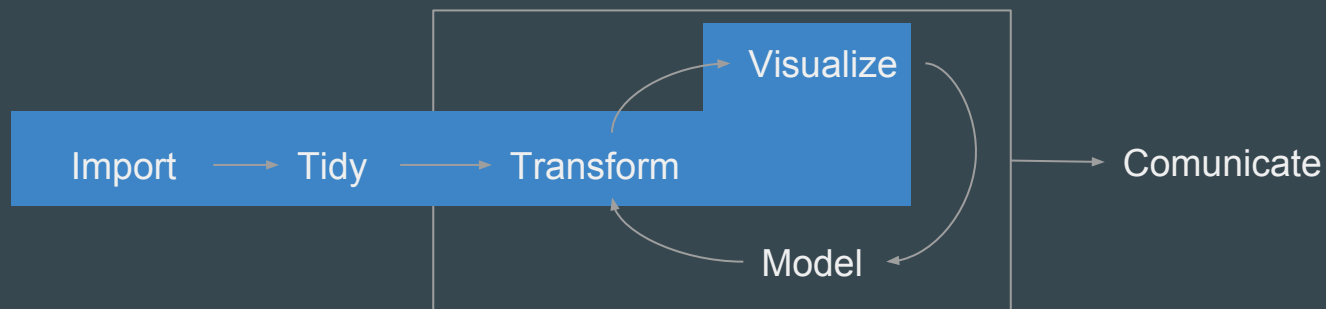
actor	
actor_id SMALLINT	
first_name VARCHAR(45)	
last_name VARCHAR(45)	
last_update TIMESTAMP	
Indexes	

film_actor	
actor_id SMALLINT	
film_id SMALLINT	
last_update TIMESTAMP	
Indexes	

inventory	
inventory_id MEDIUMINT	
film_id SMALLINT	
store_id TINYINT	
last_update TIMESTAMP	
Indexes	

film_text	
film_id SMALLINT	
title VARCHAR(255)	
description TEXT	
Indexes	

Movie database



Import

- File
 - Excel
 - OpenDocument
 - Testo
 - etc...
- Database
- Web

Tidy

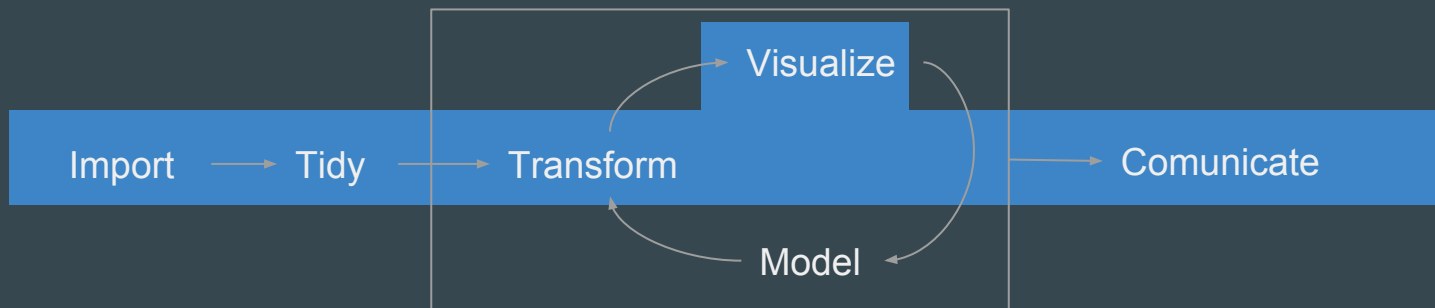
- Gathering
- Spreading
- Separate
- Unite

Transform

- dplyr

Visualize

- ggplot2



Wrangle

Importare dati in R e disporli nella forma adeguata per visualizzazione e modellazione

Import

- File
 - Excel
 - OpenDocument
 - Testo
 - etc...
- Database
- Web

Tidy

- Gathering
- Spreading
- Separate
- Unite

Transform

- dplyr

Visualize

- ggplot2

Communicate

- Rmarkdown
- Rhtml

varie

- R notebook
 - Jupyter notebook
- Personalizzazione applicazioni Shiny
 - www directory
 - tags
 - includeCSS
 - `tags$style(HTML())`
 - includeScript
- Sql injection
 - `DBI::sqlInterpolate`
- Riferimenti
 - R Advance (<http://adv-r.had.co.nz/>)