

# Importare tabelle da PDF

Per importare dati da un file pdf, usiamo il pacchetto [tabulizer](#)<sup>1)</sup>, che usa la libreria open source [tabula](#) :

## Installazione

```
install.packages("tabulizer")  
library(tabulizer)
```

Al momento, non è disponibile la versione per l'ultima versione di R. Per installare il pacchetto:

[inst\\_tabulizer.R](#)

```
if (!require("remotes")) {  
  install.packages("remotes")  
}  
# on 64-bit Windows  
remotes::install_github(c("ropensci/tabulizerjars",  
  "ropensci/tabulizer"),  
                        INSTALL_opts = "--no-multiarch")  
# elsewhere  
remotes::install_github(c("ropensci/tabulizerjars",  
  "ropensci/tabulizer"))
```

Vedi [repository](#) su [GitHub](#)

## Estrarre i dati

### Estrarre direttamente la tabella

```
mydata <- extract_tables(file.choose())
```

Si aprirà la finestra di dialogo, che consentirà di scegliere il file PDF. Naturalmente, è possibile indicare semplicemente il nome del file:

```
mydata <- extract_tables("myfile.pdf")
```

Possono essere estratte tabelle da più pagine. Il risultato sarà una [lista](#) di tabelle (formato carattere).

## Scegliere la pagina

Per indicare la pagina del file in cui è presente la tabella, usare l'argomento `pages`:

```
mydata <- extract_tables(file.choose(), pages = 3)
```

In questo caso, verrà estratta la tabella a pagina 3.

## Selezionare interattivamente la tabella

Se nella pagina sono incluse diverse tabelle o aree di testo e tabelle, possiamo usare la funzione `extract_areas()`:

```
mydata <- extract_areas(file.choose())
```

Si aprirà una interfaccia html ([Shiny](#)) che permetterà di selezionare la tabella o area di testo da importare.

## Trasformare i dati in dataframe o altro oggetto

Infine, trasformeremo i dati importati nel tipo di oggetto adatto, ad esempio un [dataframe](#):

```
mydf <- as.data.frame(mydata)
```

[Gestione dei dati](#), [Importazione](#), [PDF](#)

<sup>1)</sup>

Thomas J. Leeper (2018). `tabulizer`: Bindings for Tabula PDF Table Extractor Library.

From:  
<https://www.agnesevardanega.eu/wiki/> - Ricerca Sociale con R

Permanent link:  
[https://www.agnesevardanega.eu/wiki/r/gestione\\_dei\\_dati/importare\\_tabelle\\_da\\_pdf](https://www.agnesevardanega.eu/wiki/r/gestione_dei_dati/importare_tabelle_da_pdf)

Last update: 10/09/2025 06:33

