

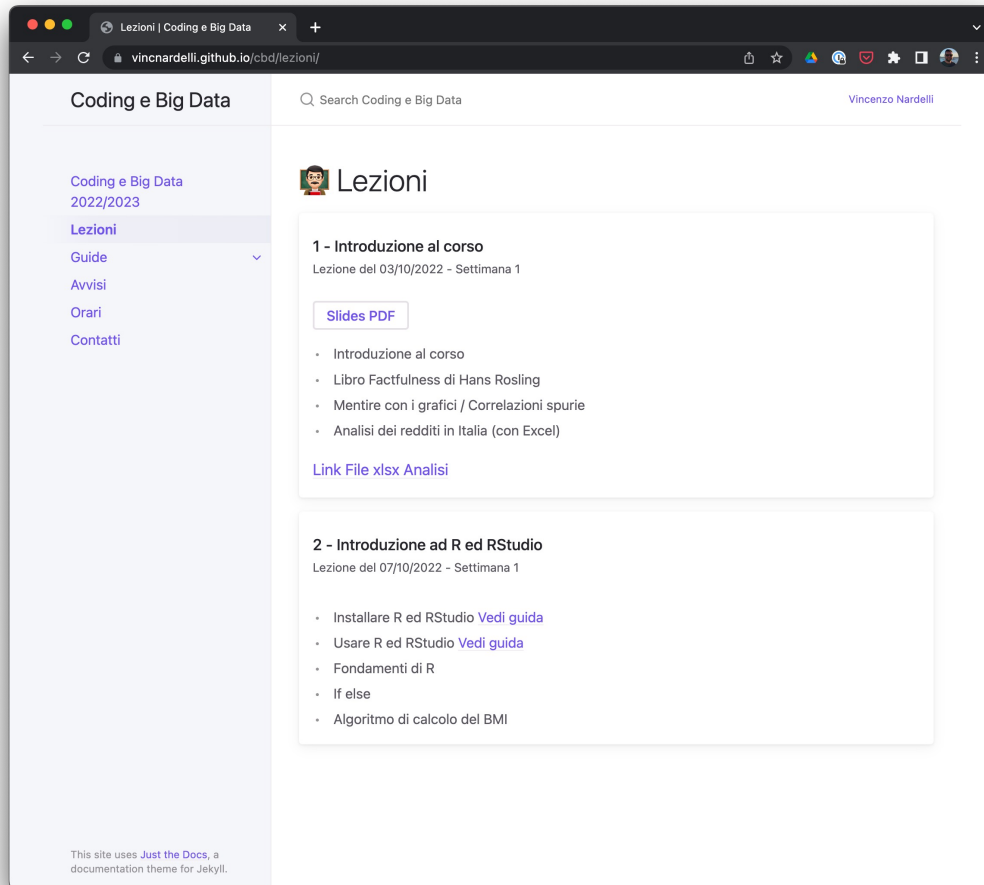
# Coding e Big Data

2022-2023



Vincenzo Nardelli  
vincenzo.nardelli@unicatt.it

# Sito web del corso



<https://vincnardelli.github.io/cbd/>

# Analisi di big data e programmazione

- Efficienza
  - Le interfacce punta e clicca non sono efficienti in termini di tempo
  - Automatizzare significa velocizzare le operazioni
- Riproducibilità
  - Crescente necessità di fornire dati, materiali ed analisi insieme ad i risultati
  - Assicura la possibilità di controllare i risultati e le procedure
  - Rende possibile effettuare analisi in produzione



# Linguaggi di programmazione più usati

Worldwide, Oct 2022 compared to a year ago:

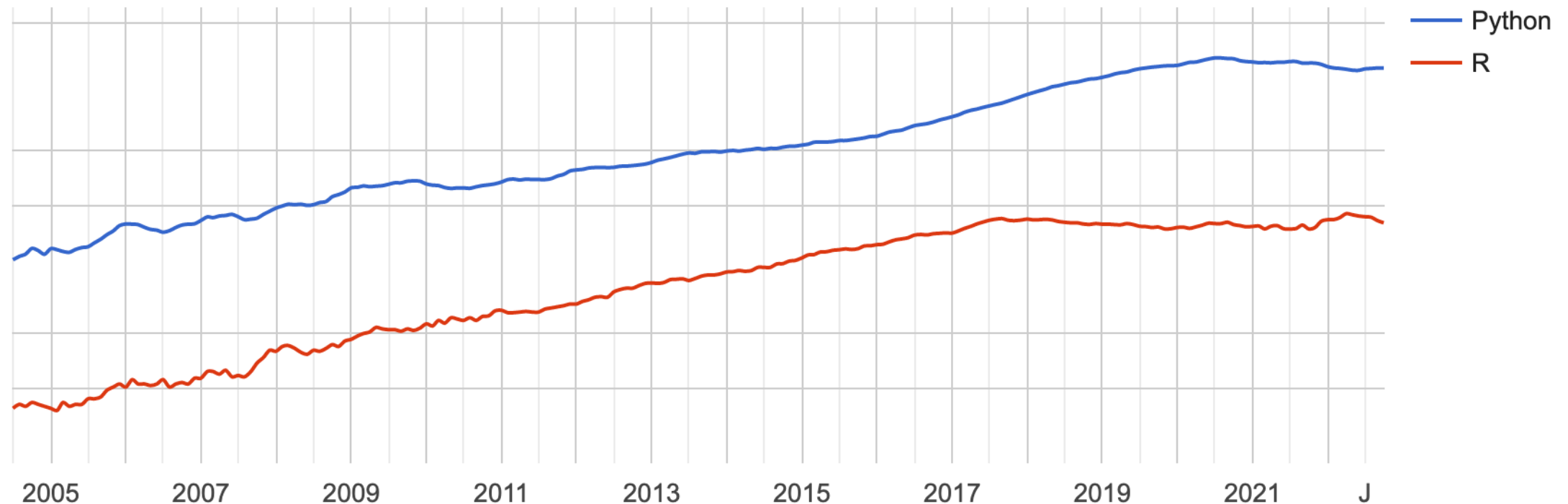
Rank	Change	Language	Share	Trend
1		Python	28.3 %	-1.8 %
2		Java	17.2 %	-0.9 %
3		JavaScript	9.69 %	+0.4 %
4		C#	7.2 %	-0.2 %
5		C/C++	6.45 %	-0.5 %
6		PHP	5.39 %	-0.9 %
7		R	4.03 %	+0.3 %
8	↑↑↑↑	TypeScript	2.71 %	+1.1 %
9	↓	Objective-C	2.16 %	+0.2 %
10	↑↑	Go	2.08 %	+0.5 %
11	↓	Swift	2.05 %	+0.4 %
12	↓↓↓	Kotlin	1.81 %	+0.1 %
13	↑↑↑↑↑	Rust	1.57 %	+0.8 %
14	↓	Matlab	1.52 %	+0.1 %
15		Ruby	1.13 %	+0.1 %

Fonte

<https://pypl.github.io/PYPL.html>

# Linguaggi di programmazione più usati

PYPL PopularitY of Programming Language



# R & Python

- Open source
  - Gratuito e liberamente utilizzabile
- Strumenti avanzati
  - Pacchetti e librerie per ogni tipo di analisi
- Documentazione e comunità
  - Nessun supporto cliente a pagamento ma comunità!



# R vs Python



- Data analytics, statistica
- Usato da statistici e dalla ricerca
- 12000 package on CRAN
- Semplice comunicazione (visualization, reporting and dashboard)



- Deployment and production
- Usato da programmatori e sviluppatori
- Integrazione con diversi sistemi operativi
- Algoritmi complessi e struttura ad oggetti

# R e RStudio

Linguaggio



Motore

IDE

(integrated development environment)



Cruscotto



# Python

Linguaggio



Motore

IDE

(integrated development environment)



Cruscotto

# Curva di apprendimento

## EXCEL

La maggior parte delle persone probabilmente ha già appreso almeno alcune nozioni di base in Microsoft Excel. Questo è un vantaggio sostanziale dell'utilizzo di Excel: la curva di **apprendimento iniziale è piuttosto minima** e la maggior parte delle analisi può essere eseguita puntando e facendo clic sul pannello superiore. Una volta che un utente ha importato i propri dati nel programma, non è eccessivamente difficile creare grafici e diagrammi di base.

## R

R è un linguaggio di programmazione, tuttavia, il che significa che la curva di **apprendimento iniziale è più ripida**. Ci vorranno almeno alcune settimane per familiarizzare con l'interfaccia e padroneggiare le varie funzioni. Fortunatamente, l'uso di R può diventare rapidamente una seconda natura con la pratica.

# Perchè programmare invece di usare Excel?

- R aiuta a leggere qualsiasi tipo di dato disponibile.
- L'automazione è molto più semplice rispetto a Excel.
- Supporta set di dati più grandi
- Offre anche un calcolo più veloce
- Le capacità di manipolazione dei dati sono potenti rispetto a Excel.
- Organizzazione del prodotto più semplice
- R può anche aiutare a rilevare qualsiasi tipo di errore (riproducibilità)
- È gratuito e open source. Quindi non è necessario pagare come Excel.
- R fornisce funzionalità statistiche avanzate.

# Installazione e configurazione di R e RStudio

- [https://vincnardelli.github.io/cbd/guide/installazione\\_r\\_rstudio](https://vincnardelli.github.io/cbd/guide/installazione_r_rstudio)
- [https://vincnardelli.github.io/cbd/guide/usare\\_r\\_rstudio](https://vincnardelli.github.io/cbd/guide/usare_r_rstudio)