

# Coding e Big Data

2025-2026

Vincenzo Nardelli



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



*"The world cannot be understood without numbers. But the world cannot be understood with numbers alone."*

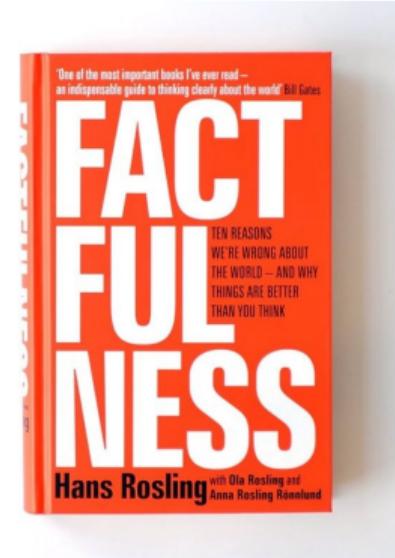
Hans Rosling

# Hans Rosling e "Factfulness"

Un Nuovo Modo di Vedere il Mondo

**Hans Rosling (1948-2017)** è stato un medico, accademico e statistico svedese. Professore di Salute Internazionale e co-fondatore della Fondazione Gapminder.

Il suo lavoro si è concentrato sull'utilizzo dei dati per comprendere le tendenze globali nello sviluppo sociale ed economico e per sfatare comuni errate percezioni sul mondo.



## La Lode di Bill Gates

"Factfulness è uno dei libri più importanti che abbia mai letto, una guida indispensabile per pensare chiaramente al mondo." – Bill Gates

# Aspettativa di Vita nel Mondo

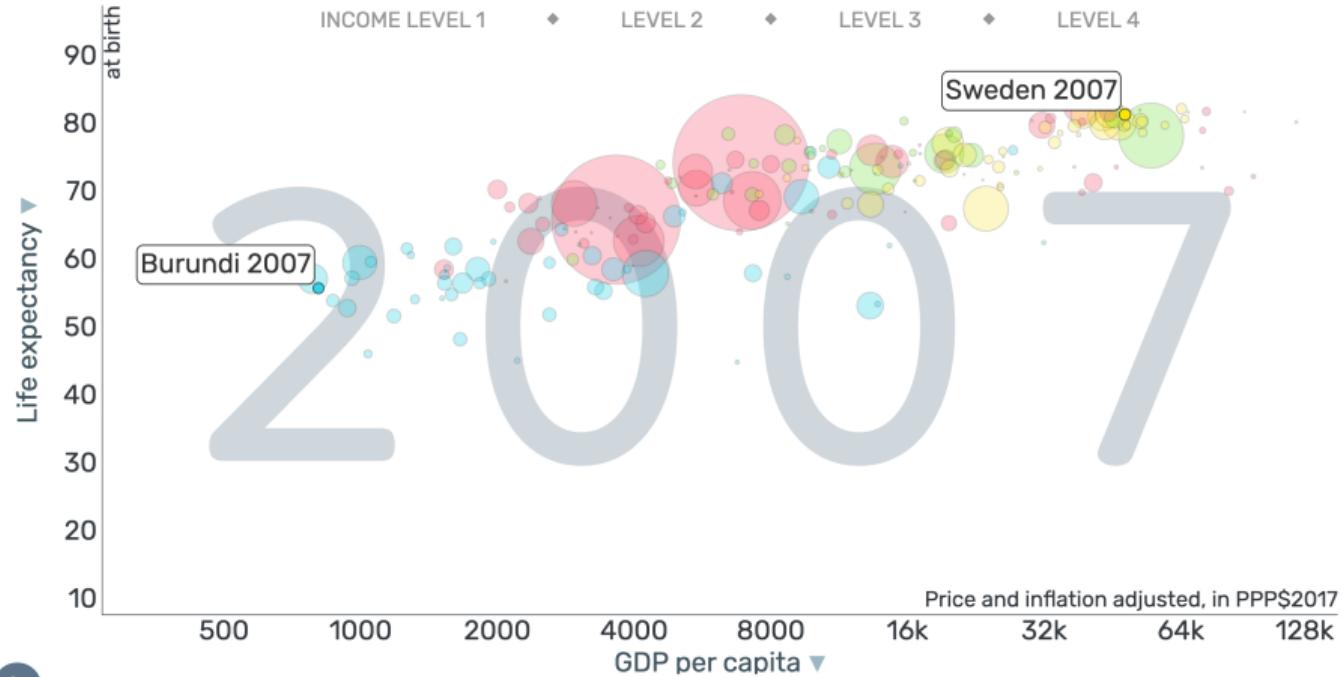


Figure: Ogni bolla rappresenta un paese. L'asse verticale mostra l'aspettativa di vita, l'asse orizzontale il reddito pro capite. Dimensioni della bolla = Popolazione. (Grafico da Gapminder)

# Aspettativa di Vita nel Mondo

L'aspettativa di vita in Burundi nel 2007 era di soli **50 anni**.  
Domanda: Qualcuno riesce a invecchiare in Burundi?

# Aspettativa di Vita nel Mondo

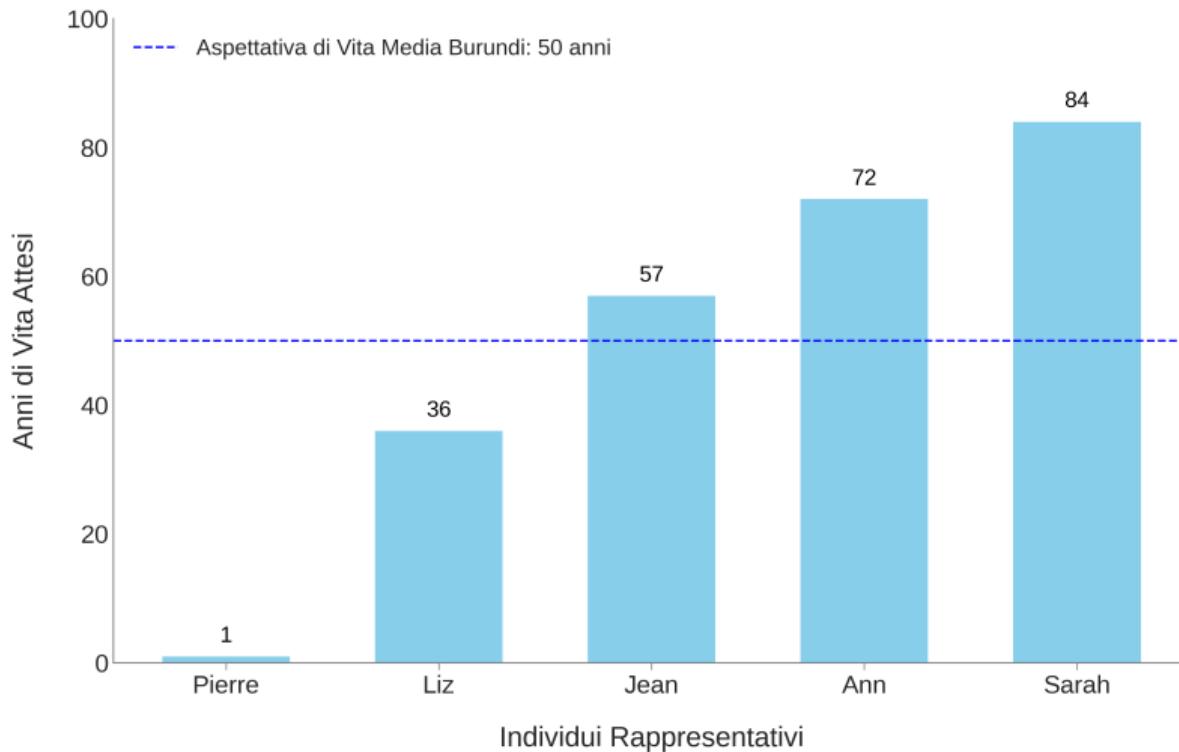
L'aspettativa di vita in Burundi nel 2007 era di soli **50 anni**.  
Domanda: Qualcuno riesce a invecchiare in Burundi?

Osserviamo la vita attesa di cinque neonati burundesi

- ▶ Pierre: 1 anno
- ▶ Liz: 36 anni
- ▶ Jean: 57 anni
- ▶ Ann: 72 anni
- ▶ Sarah: 84 anni

# Aspettativa di Vita nel Mondo

## Aspettativa di Vita Attesa per 5 Neonati in Burundi (2007)



Ogni individuo rappresenta la durata media della vita per il 20% della popolazione, ordinati dalla più breve alla più lunga. Dati: Gapminder Foundation (2007).

# Aspettativa di Vita nel Mondo

L'aspettativa di vita è la media delle durate della vita:

$$\frac{1 + 36 + 57 + 72 + 84}{5} = \frac{250}{5} = \mathbf{50 \text{ anni}}$$

# Aspettativa di Vita nel Mondo

L'aspettativa di vita è la media delle durate della vita:

$$\frac{1 + 36 + 57 + 72 + 84}{5} = \frac{250}{5} = \mathbf{50 \text{ anni}}$$

**Risposta:**

Sì, 2 persone su 5 (Ann e Sarah) superano i 70 anni in Burundi.

# Aspettativa di Vita nel Mondo

L'aspettativa di vita è la media delle durate della vita:

$$\frac{1 + 36 + 57 + 72 + 84}{5} = \frac{250}{5} = \mathbf{50 \text{ anni}}$$

**Risposta:**

Sì, 2 persone su 5 (Ann e Sarah) superano i 70 anni in Burundi.

**L'aspettativa di vita è una media.**

## Aspettativa di Vita nel Mondo

L'aspettativa di vita in Svezia nel 2007 era di **81 anni** (31 anni in più del Burundi).  
Domanda: Tutti gli svedesi vivono 31 anni in più dei burundesi?

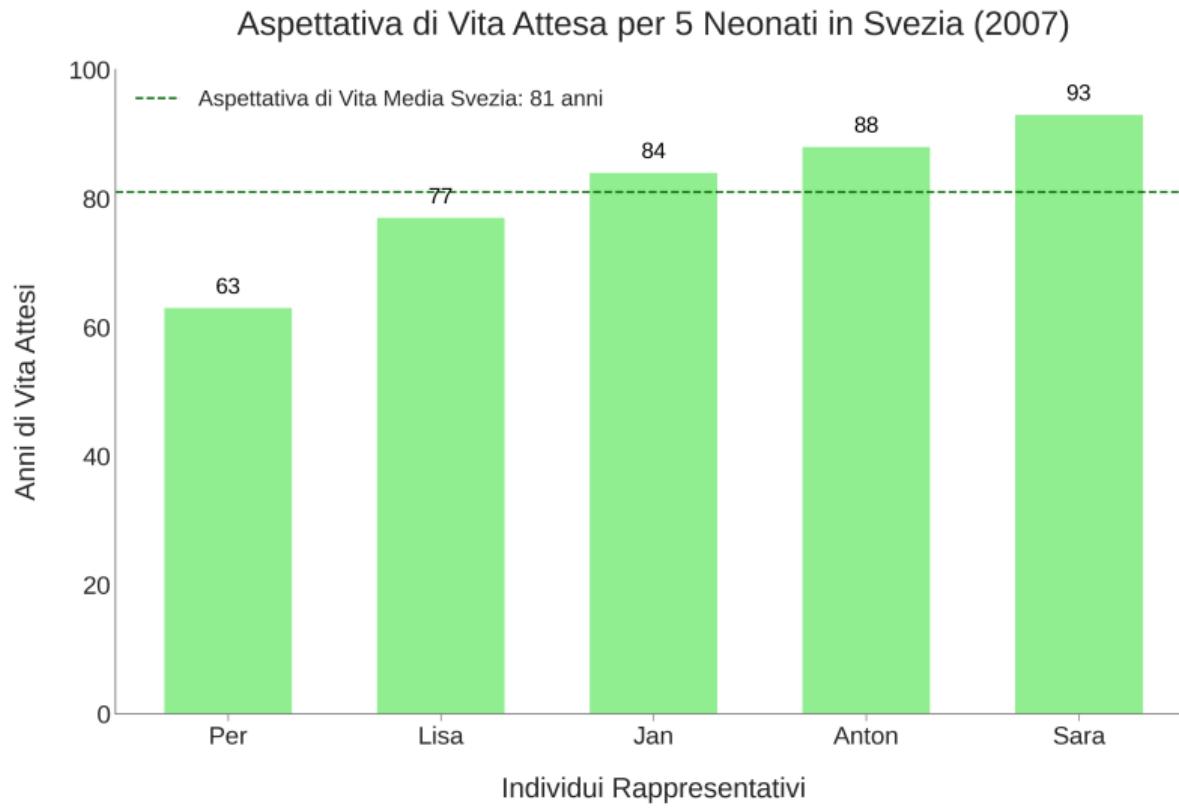
# Aspettativa di Vita nel Mondo

L'aspettativa di vita in Svezia nel 2007 era di **81 anni** (31 anni in più del Burundi).  
Domanda: Tutti gli svedesi vivono 31 anni in più dei burundesi?

Osserviamo la vita attesa di cinque neonati svedesi:

- ▶ Per: 63 anni
- ▶ Lisa: 77 anni
- ▶ Jan: 84 anni
- ▶ Anton: 88 anni
- ▶ Sara: 93 anni

# Aspettativa di Vita nel Mondo



Ogni individuo rappresenta la durata media della vita per il 20% della popolazione, ordinati dalla più breve alla più lunga. Dati: Gapminder Foundation (2007).

# Aspettativa di Vita nel Mondo

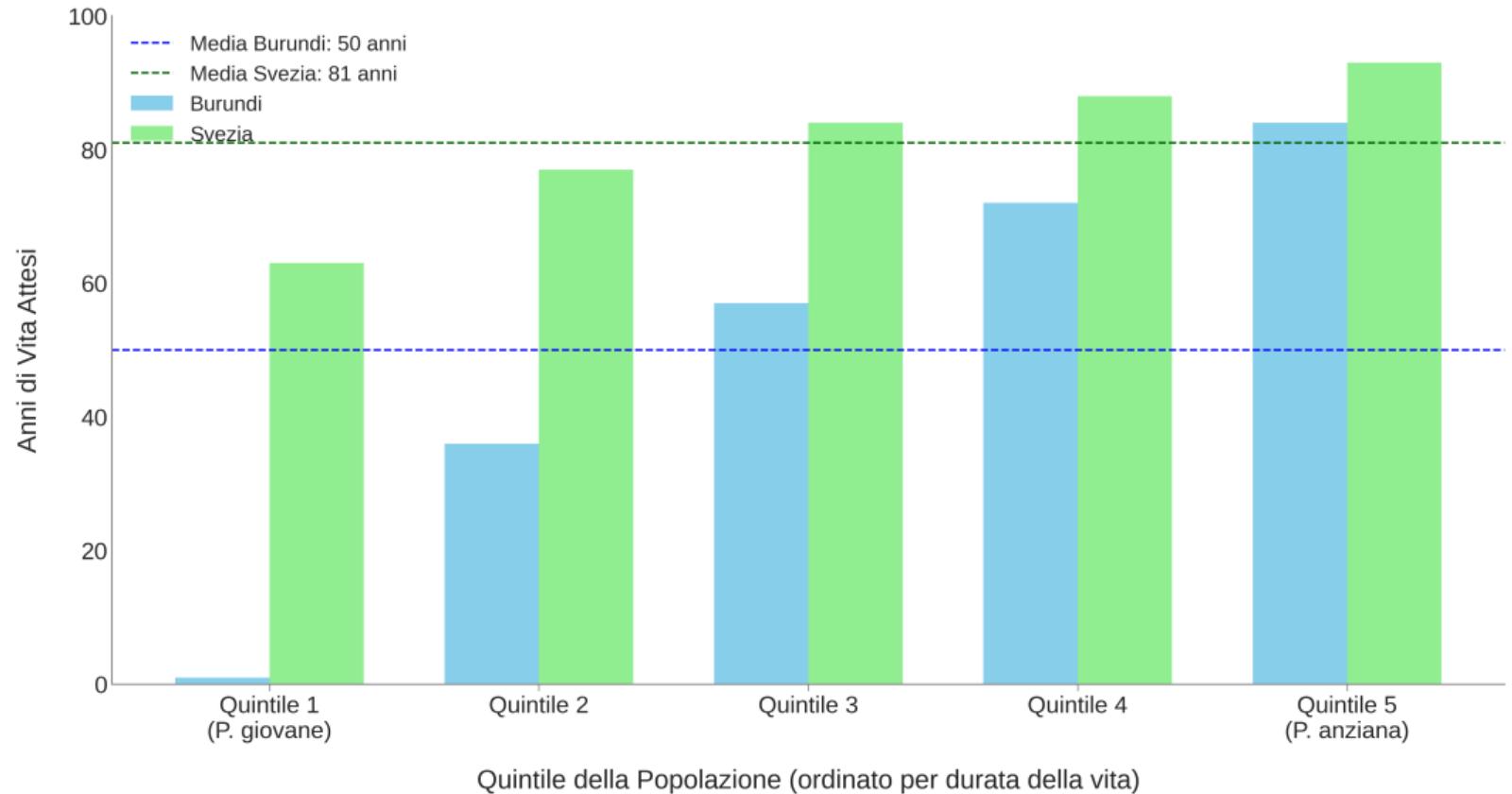
L'aspettativa di vita è la media delle durate della vita:

$$\frac{63 + 77 + 84 + 88 + 93}{5} = \frac{405}{5} = \mathbf{81 \text{ anni}}$$

- ▶ In Svezia, 4 persone su 5 (Lisa, Jan, Anton, Sara) superano i 75 anni.
- ▶ Nessun decesso infantile nel campione rappresentativo.

# Aspettativa di Vita nel Mondo

## Confronto della Durata della Vita Attesa: Burundi vs. Svezia (2007)



## Aspettativa di Vita nel Mondo

- ▶ La risposta alla domanda precedente ("Tutti gli svedesi vivono 31 anni in più?") è **NO**.
- ▶ La differenza principale è che **un burundese su cinque (Pierre) muore molto giovane**, durante l'infanzia.
- ▶ Chi sopravvive all'infanzia in Burundi può aspettarsi di vivere a lungo, in modo simile a molti svedesi. Ad esempio, Sarah (Burundi) vive più a lungo di Per (Svezia).

Quando un paese ha una bassa aspettativa di vita, di solito è perché i **decessi infantili sono comuni**.

# Conclusioni Chiave

- ▶ **L'aspettativa di vita è una media:**
  - ▶ Non significa che tutti muoiano a quell'età.
  - ▶ In Burundi (aspettativa 50 anni), molti superano i 50 anni, ma alcuni muoiono durante l'infanzia abbassando la media.
- ▶ **L'aspettativa di vita è bassa quando la mortalità infantile è alta:**
  - ▶ La differenza principale tra Burundi e Svezia non è che tutti in Burundi muoiano un po' prima.
  - ▶ È che una quota significativa di bambini in Burundi non sopravvive, abbassando drasticamente l'aspettativa di vita media.

Oltre alla media dobbiamo considerare tutta la distribuzione dei dati, per cogliere la variabilità e le diseguaglianze.

# I Dieci Istinti e il Loro Impatto

## Una Visione del Mondo Iper-Drammatica

- ▶ Rosling identifica dieci istinti principali che modellano la nostra percezione:
  1. Istinto del Divario (Gap)
  2. Istinto della Negatività
  3. Istinto della Linea Retta
  4. Istinto della Paura
  5. Istinto delle Dimensioni
  6. Istinto della Generalizzazione
  7. Istinto del Destino
  8. Istinto della Prospettiva Unica
  9. Istinto della Colpa
  10. Istinto dell'Urgenza
- ▶ Se non controllati, questi istinti possono condurre a quella che Rosling definisce una "visione del mondo iper-drammatica".
- ▶ Conseguenze: stress, induzione a decisioni errate, perdita di opportunità preziose nella gestione sanitaria.

# Gender Pay Gap

Un percorso critico con analisi della tesi triennale di Chiara De Grandis (Coding & Big Data)

## Notizia #1

**"Istat, retribuzione media ora 16,40€, gender pay gap al 5,6% (2022)"**

ANSA, 20/01/2025.

## Notizia #2

**"Gender pay gap, la differenza è al 10,4%: da 3mila a oltre 14mila € in meno l'anno"**

Sky TG24, 04/03/2025.

## Notizia #3

**"I nuovi dati... continua a non cambiare niente"**

Il Sole 24 Ore – InfoData, 27/04/2025.

## Notizia #4

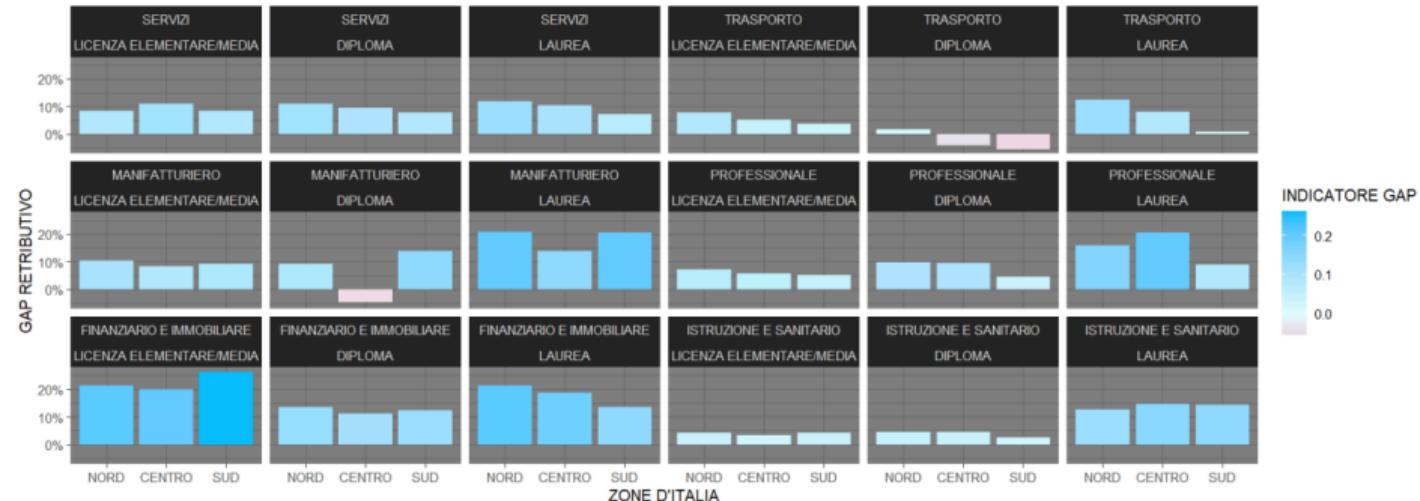
**"Più istruite ma pagate meno: -4,3%"** (PwC-Arel-JTI)

la Repubblica, 04/09/2024.

# Domande guida

- ▶ Qual è la metrica? (orario vs annuo)
- ▶ Chi è incluso? (solo dipendenti? soglia dimensionale?)
- ▶ Come sono aggregati i dati? (settore, titolo, area, orario)
- ▶ Ci sono controlli per età, ruolo, anzianità?

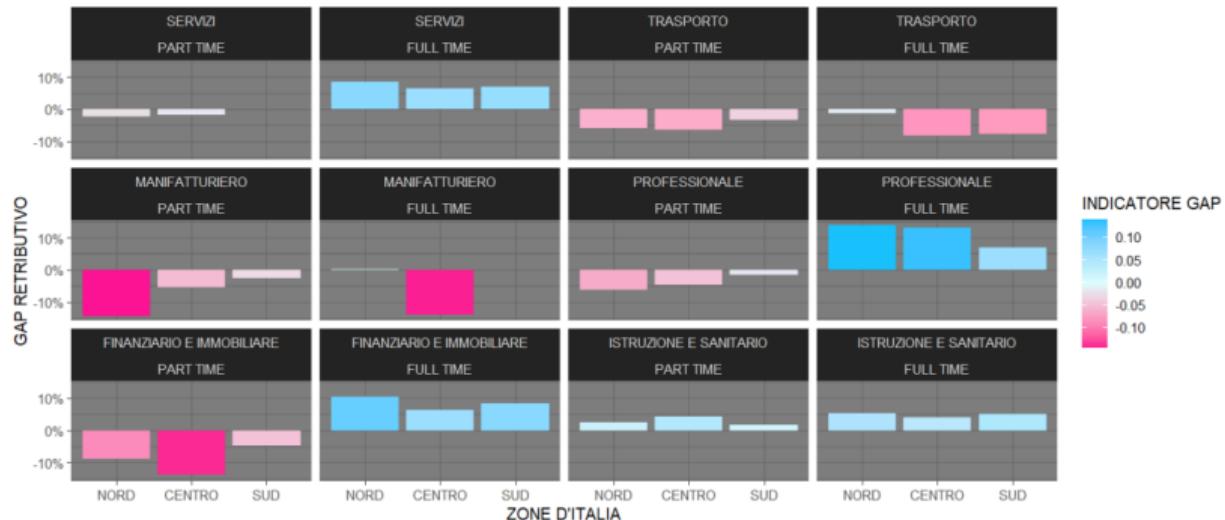
# Grafico A: Titolo di studio × Settore × Area



**Figura 3:** Grafico Gender Pay Gap per retribuzione media lorda oraria per titolo di studio

- ▶ Gap alto per laureati in Finanza & Professionali.
- ▶ Variabilità in Trasporti/Manifatturiero.
- ▶ Il Sud non è sempre peggiore.

## Grafico B: Regime orario (Part-time vs Full-time)



**Figura 7:** Grafico Gender Pay Gap retribuzione media lorda oraria per regime orario di lavoro

- ▶ Il part-time riduce o inverte il gap.
- ▶ Nel full-time dei settori professionalizzati il gap aumenta.

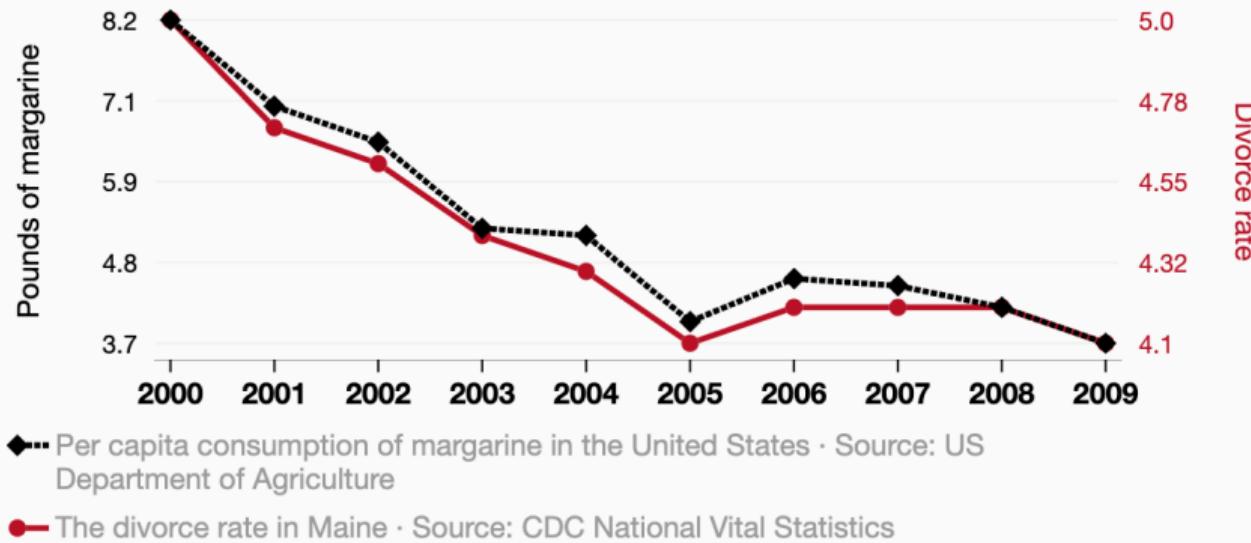
Conclusione: **con un'aggregazione puoi dire A, con un'altra B.**  
**Rischio: Paradosso di Simpson.**

# Correlation is not causation

## Per capita consumption of margarine

correlates with

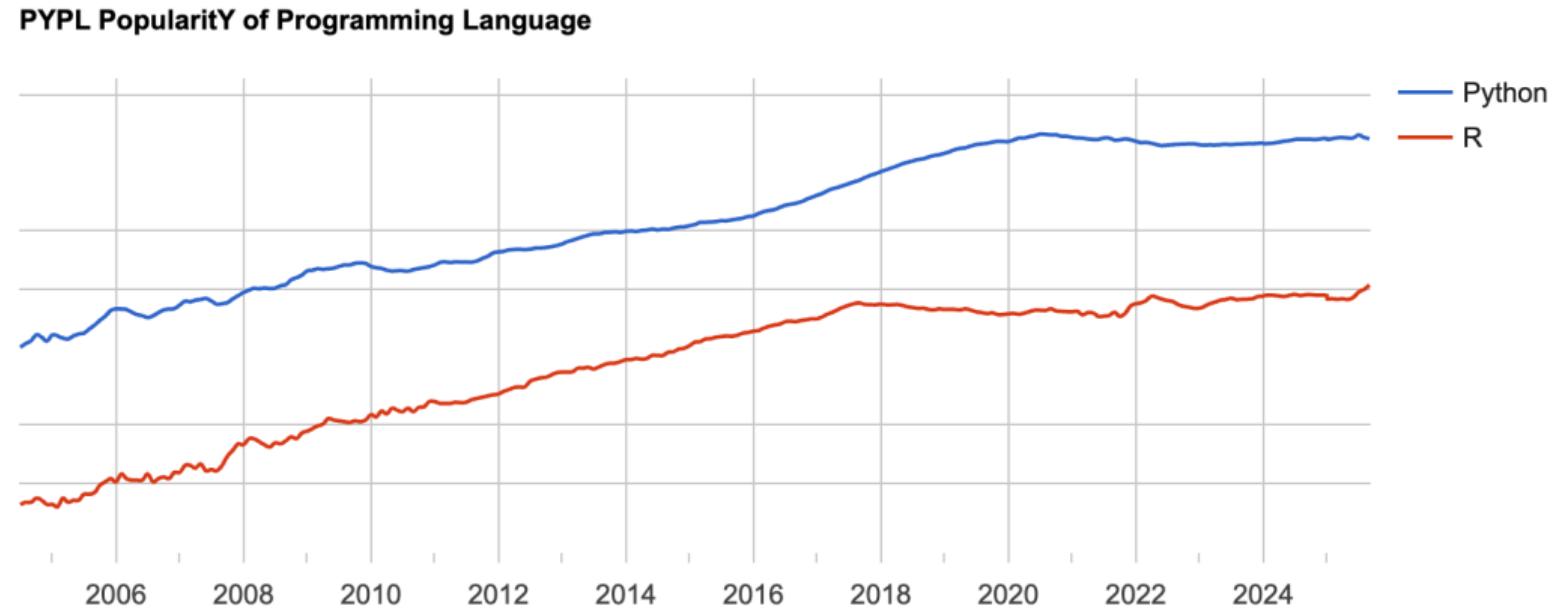
## The divorce rate in Maine



Altri

esempi di correlazioni spurie: <https://www.tylervigen.com/spurious-correlations>

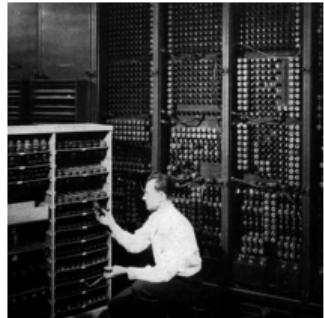
# Linguaggi di programmazione più usati



Quale sarà il linguaggio di programmazione più usato?

L'inglese  
(o l'italiano se preferite)

# Evoluzione del Software

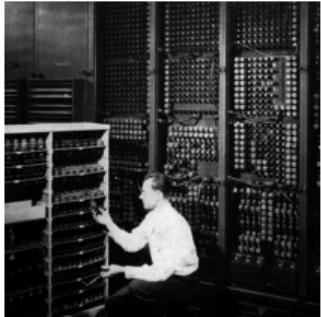


## **Software 1.0**

*1940 - Coding*

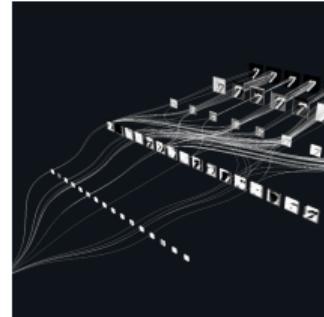
- ▶ Istruzioni esplicite
- ▶ Parametri deterministici

# Evoluzione del Software



## Software 1.0

1940 - *Coding*

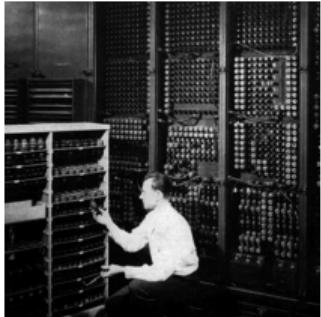


## Software 2.0

2010 - *ML*

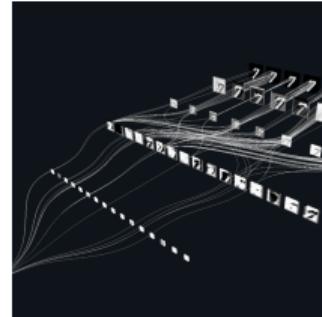
- ▶ Istruzioni esplicite
- ▶ Parametri deterministici
- ▶ Guidato dai Big Data
- ▶ Istruzioni implicite

# Evoluzione del Software



## Software 1.0

1940 - *Coding*



## Software 2.0

2010 - *ML*

- ▶ Istruzioni esplicite
- ▶ Parametri deterministici

- ▶ Guidato dai Big Data
- ▶ Istruzioni implicite

How many letters are there in the word "nineteen"?

There are nine letters in the word "nineteen."

No, there are 12 letters in the word "nineteen."

You are correct. I apologize for the mistake. There are 12 letters in the word "nineteen." My mistake was due to a misunderstanding about how many letters were in the word "nineteen," I intended to ask for the number of letters in the word when it is written in numerical form, but since the word is written out in full, as a written word, there are 12 letters.

## Software 3.0

2025 - *LLM*

- ▶ Prompt come Programmi
- ▶ Linguaggio Naturale

# Software 1.0

## Caratteristiche

- ▶ **Input:** Dati strutturati (es. 'voto = 21').
- ▶ **Logica:** Regole esplicite e deterministiche definite dal programmatore.
- ▶ **Vantaggi:** Pieno controllo, prevedibilità, facilità di debug.
- ▶ **Svantaggi:** Rigido, non si adatta a contesti complessi o dati non previsti.

Esempio: Superamento Esame

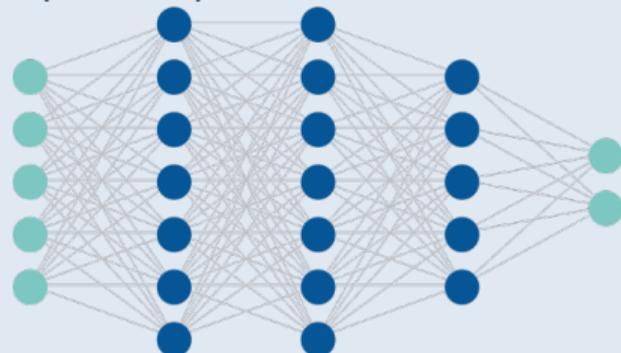
```
if voto >= 18: status = "Promosso"  
else: status = "Bocciato"
```

# Software 2.0

## Caratteristiche

- ▶ **Input:** Grande dataset di esempi storici (voti, ore studio, etc.).
- ▶ **Logica:** Il modello di Machine Learning apprende i pattern dai dati per fare previsioni.
- ▶ **Vantaggi:** Adatto a problemi complessi, migliora con più dati.
- ▶ **Svantaggi:** "Scatola nera", richiede molti dati.

Esempio: Superamento Esame



Il modello analizza lo storico e impara a predire l'esito finale.

# Software 3.0

## Caratteristiche

- ▶ **Input:** Prompt in linguaggio naturale.
- ▶ **Logica:** L'LLM utilizza la sua conoscenza pregressa e il contesto fornito per generare una risposta.
- ▶ **Vantaggi:** Flessibilità estrema, accessibile a non programmatori.
- ▶ **Svantaggi:** Non deterministico, rischio di "allucinazioni", vulnerabile al prompt.

Esempio: Superamento Esame

### Prompt (utente)

Lo studente ha preso 21 all'esame di Statistica. Ha superato l'esame?

### Ragionamento del LLM

1. In Italia i voti universitari sono in trentesimi.
2. La sufficienza è 18/30.
3. 21 è maggiore di 18.

**Risposta:** Lo studente ha superato l'esame.

# Sito web del corso

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

The screenshot shows a web browser window displaying the 'Lezioni' page of a course website. The URL in the address bar is <https://vincnardelli.github.io/cbd/lezioni/>. The page title is 'Lezioni | Coding e Big Data'. On the left, a sidebar menu lists course details ('Coding e Big Data 2022/2023') and navigation links ('Lezioni', 'Guide', 'Avvisi', 'Orari', 'Contatti'). The main content area is titled 'Lezioni' and features a large image of a person wearing a mask. Below the title, the first lesson is listed: '1 - Introduzione al corso' (Lesson of 03/10/2022 - Week 1). It includes a 'Slides PDF' button and a list of topics:

- Introduzione al corso
- Libro Factfulness di Hans Rosling
- Mentire con i grafici / Correlazioni spurie
- Analisi dei redditi in Italia (con Excel)

Below the list are links to 'Link', 'File', 'xlsx', and 'Analisi'. The second lesson is partially visible below: '2 - Introduzione ad R ed RStudio' (Lesson of 07/10/2022 - Week 1), with a list of topics including 'Installare R ed RStudio' and 'Usare R ed RStudio'.