

Data Journalism Introduction

Andrea Marchetti

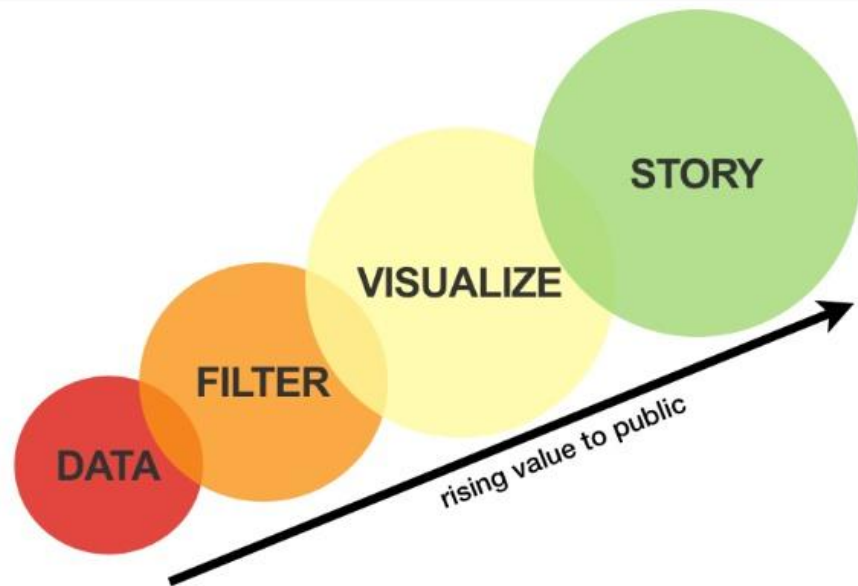
Infouma A.A. 2025/2026

Cos'è il Data Journalism

Il data journalism è una forma di giornalismo che utilizza **dati** come fonte primaria per scoprire, analizzare e raccontare **storie** di interesse pubblico

- I dati come fonte
- L'analisi come metodo
- La visualizzazione come linguaggio
- La narrazione come obiettivo

DATA-DRIVEN JOURNALISM = PROCESS



Giornalismo Vs Data Journalism

Giornalismo tradizionale

Fonti umane

Interviste

Narrazione lineare

Data Journalism

Dataset

Esplorazione/Analisi = Intervistare i dati

Narrazione (ipertestuale) con
visualizzazioni

Il data journalism è una evoluzione del giornalismo tradizionale



Incidenti stradali in Toscana anno 2024 - Fonte ISTAT

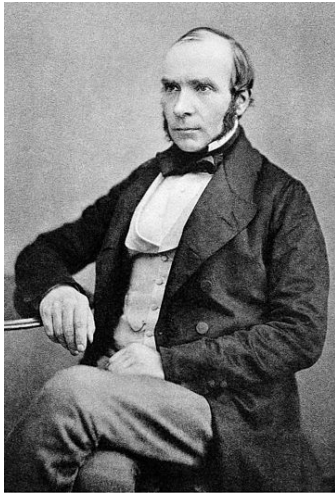
Incidenti stradali, morti, feriti e tasso di mortalità per provincia. Anno 2024 (valori assoluti, valori percentuali e tasso di mortalità per 100.000 abitanti)							
	Incidenti	Morti	Feriti	Morti: differenza 2024-2023	Morti: variazioni % 2024-2019 (b)	Morti: variazioni % 2024-2010	Tasso di mortalità 2024 (c)
Italia	173.364	3.030	233.853	-9	-4,5	-26,3	5,1
Toscana	15.174	188	19.465	-14	-10,0	-38,6	5,1
Arezzo	912	14	1.210	-9	-26,3	-51,7	4,2
Firenze	4.831	54	6.021	-7	22,7	-31,6	5,5
Grosseto	804	10	1.118	-6	-28,6	-67,7	4,6
Livorno	1.674	18	2.144	-2	-18,2	-45,5	5,5
Lucca	1.720	24	2.185	5	-20,0	-27,3	6,3
Massa Carrara	839	12	1.082	2	-14,3	-33,3	6,4
Pisa	1.706	22	2.343	-6	4,8	-33,3	5,3
Pistoia	931	11	1.195	1	-15,4	-26,7	3,8
Prato	998	10	1.152	6	-16,7	-16,7	3,8
Siena	759	13	1.015	2	-35,0	-43,5	5,0

Data (2024)	Ora	Latitudine	Longitudine	Indirizzo	Comune	Provincia	Regione	N. morti	N. feriti
15/01/2024	08:45	41.9028	12.4964	Via Appia Nuova 145	Roma	RM	Lazio	0	2
03/02/2024	22:10	45.4642	9.1900	Viale Monza 210	Milano	MI	Lombardia	1	3
18/03/2024	14:30	40.8518	14.2681	Corso Garibaldi 88	Napoli	NA	Campania	0	1
29/05/2024	11:20	45.0703	7.6869	Corso Francia 120	Torino	TO	Piemonte	0	2

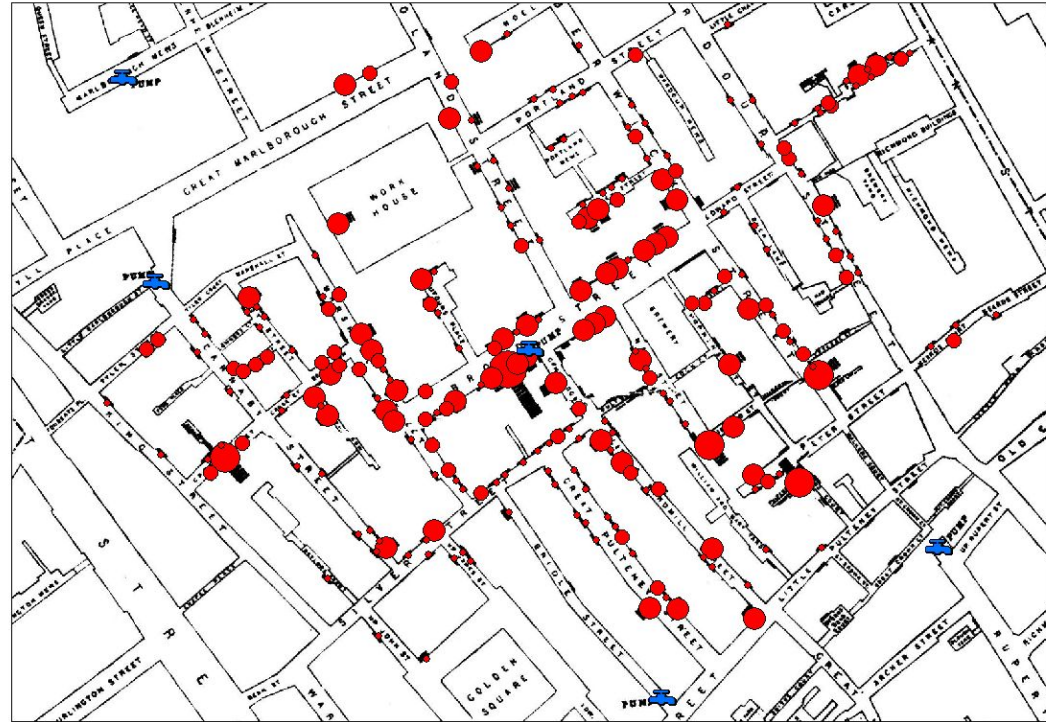
Origini del data journalism

Origini del Data Journalism

1854 – John Snow e la mappa del colera a Londra



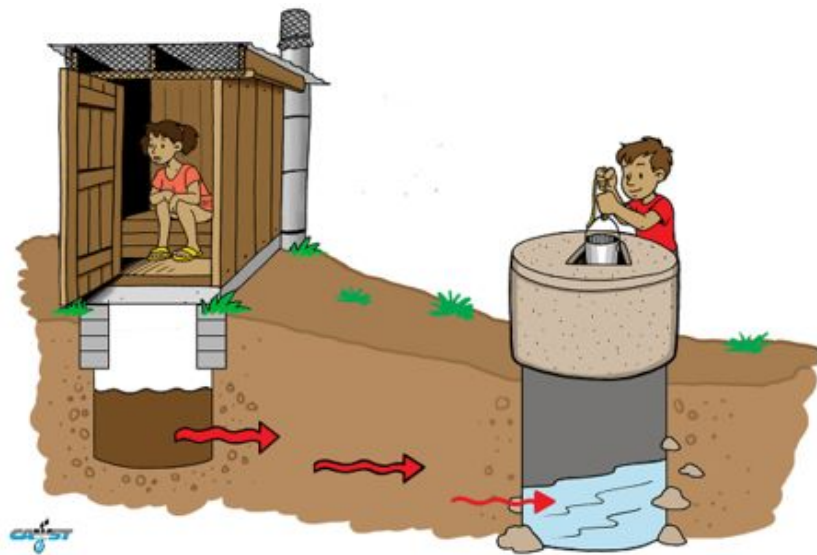
John Snow





Broadwick Street, la
pompa incriminata e il
pub

Il colera non si
trasmette per via aerea
ma per via idrica



Introduzione dell'informatica

1980 - arrivano i computer. Originariamente chiamato “computer-assisted reporting”

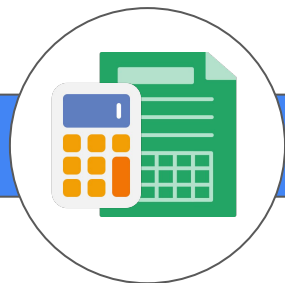
1990 - crescita sostanziale con la diffusione dei primi personal computer

2000 - open data, tools per data visualization, interactive maps, database

2010 - si afferma il termine data journalism con i primi casi come wikileaks

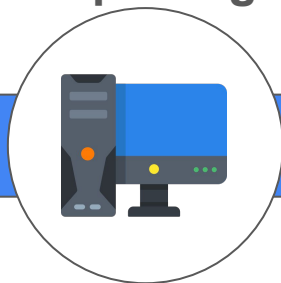
Evoluzione del data journalism

1960-1970
Precision Journalism



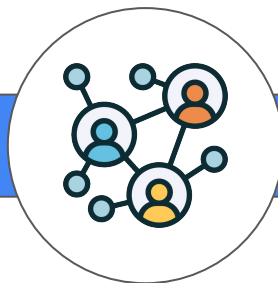
Philip Meyer
Sondaggi, statistica

1990-2000
Computer Assisted Reporting



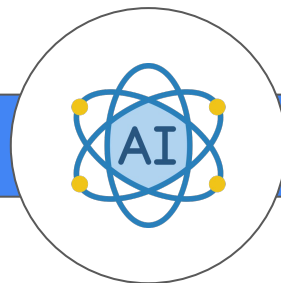
Database
Foglio di calcolo

2010-2018
Data Journalism



Open Data
Tools
Data Journalism Handbook

2020-Futuro
Journalism AI



LLM
Etica del dato

Esempi famosi

The Guardian – MPs' Expenses Scandal (2009)

Nel maggio 2009 furono resi pubblici centinaia di migliaia di documenti sulle spese dei Parlamentari britannici

Rivelarono un uso spesso improprio di fondi pubblici

The Guardian organizzò un'analisi collaborativa di questi documenti.



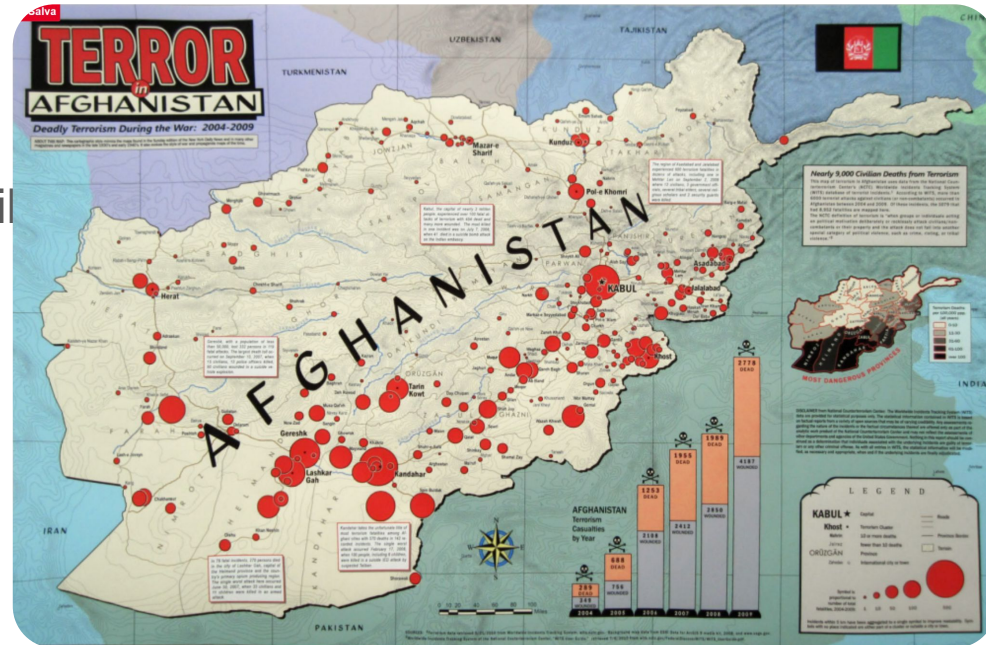
Wikileaks / Afghanistan War Logs (2010)

Divulgazione di una raccolta di registri militari interni top secret statunitensi a riguardo della guerra in Afghanistan, che sono stati pubblicati da WikiLeaks il 25 luglio 2010

Julian Assange giornalista australiano fondatore di wikileaks

<https://wardiary.wikileaks.org/>

Wiki (veloce) Leaks (fuga di notizie)

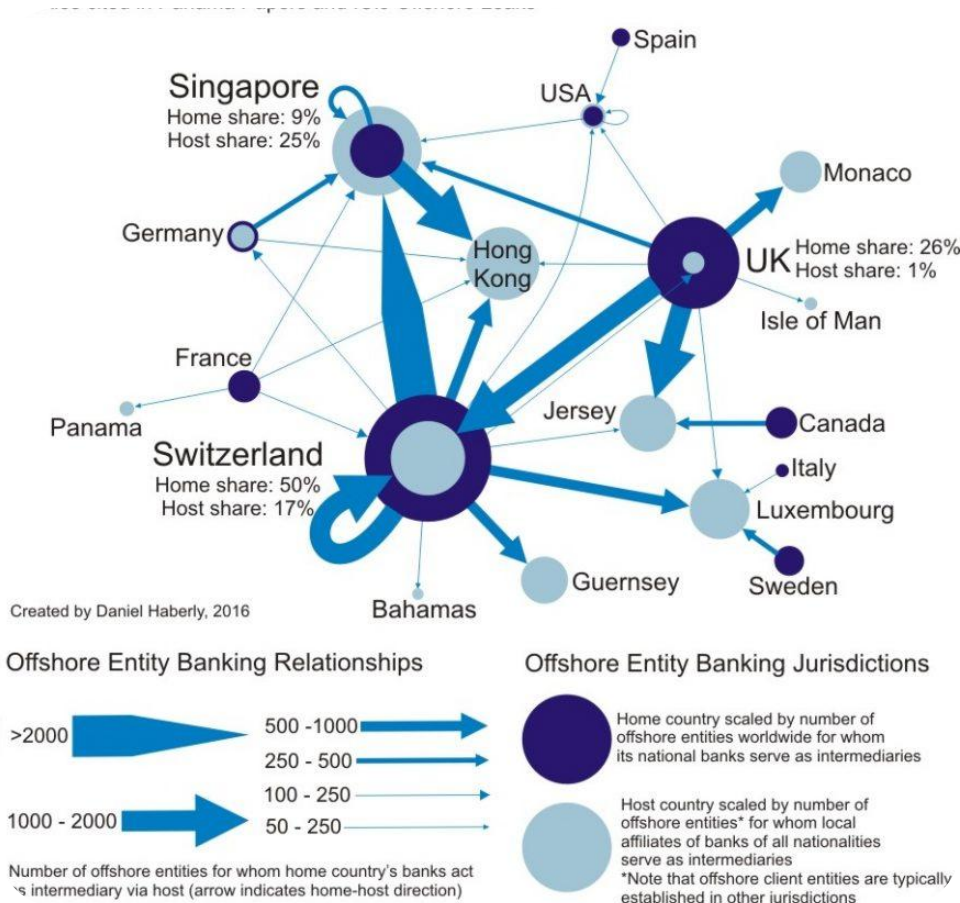


Panama Papers - ICIJ (2015)

11,5 milioni di documenti trapelati (leak) provenienti dallo **studio legale panamense Mossack Fonseca**, specializzato nella creazione di società offshore

International Consortium of Investigative Journalists, coordinò una collaborazione tra oltre 370 giornalisti di più di 100 testate in circa 80 paesi

L'inchiesta rivelò come politici, imprenditori, celebrità e criminali abbiano utilizzato società offshore per: nascondere patrimoni, evitare o eludere il pagamento delle tasse, occultare conflitti di interesse

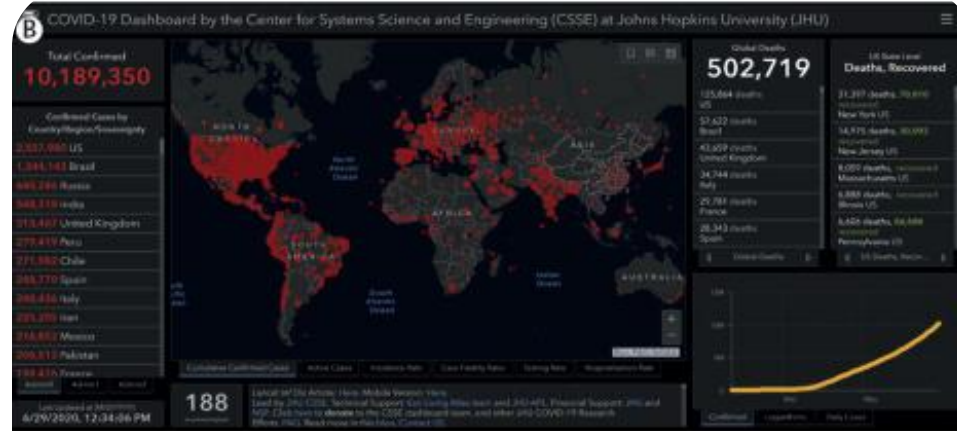


COVID-19 tracking (2020-2023)

Il progetto del New York Times si basava su un'enorme raccolta sistematica di dati epidemiologici ufficiali, aggiornati quotidianamente.

<https://www.nytimes.com/interactive/2021/us/covid-cases.html>

<https://coronavirus.jhu.edu/>



Epstein Files (2008-2024)

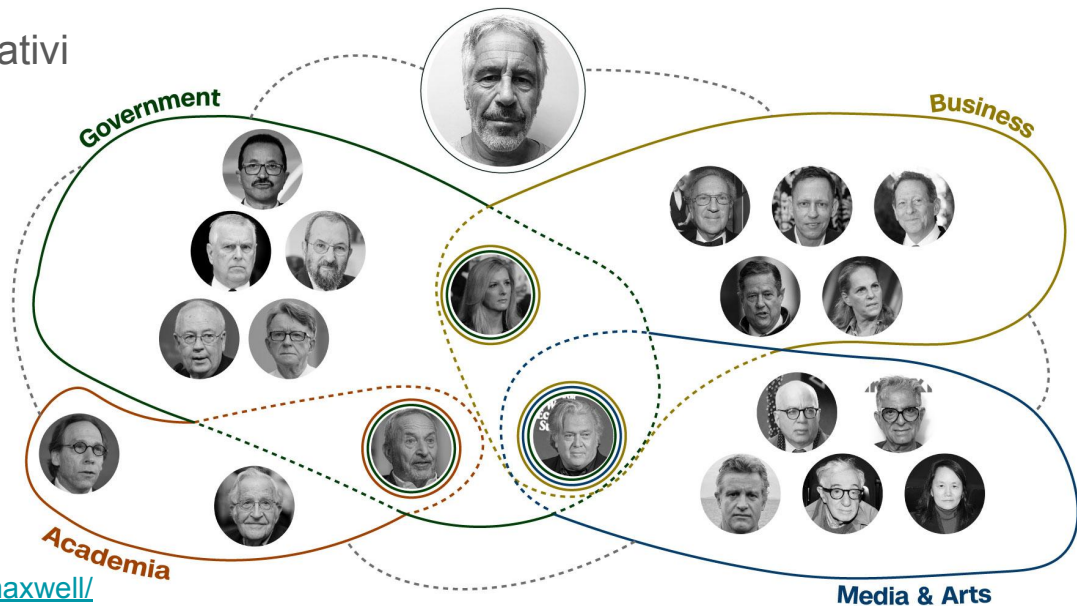
Insieme di documenti giudiziari e investigativi (PDF) pubblicati in momenti diversi:

- Atti giudiziari desecretati
- Deposizioni (testimonianze sotto giuramento)
- Documenti civili (cause per diffamazione, procedimenti federali)
- Flight logs ...

<https://www.theguardian.com/us-news/jeffrey-epstein>

<https://www.courtlistener.com/docket/4355835/giuffre-v-maxwell/>

<https://epsteinvisualizer.com/>



Openpolis

Punti di una ricerca di uno storico - Alessandro Barbero

All'inizio ci sono le **domande**: cosa vuoi sapere, quali sono le domande che ti poni

Poi ci sono i **fatti**: Stabilire i fatti, anche cosa sentono le persone è un fatto, ma ci sono fatti quantificabili e altri difficili da ricostruire. Per i fatti devi dichiarare dove sono le fonti.

Infine ci sono le **interpretazioni**: la sconfitta di Napoleone è un fatto, ma dire che questo ha segnato la nascita della potenza inglese, beh questa è una interpretazione, che può cambiare nel tempo.



Glossario

Agenda setting, framing, priming

Narrativa

Insight: fatto/pattern che risulta inaspettato o sorprendente o estremamente importante,

Enlight: fatto che cambia il nostro modo di percepire la realtà

Bias cognitivi

Correlazione $\neq$ Causalità

Agenda Setting, framing, priming

Agenda setting: Sono i media che stabiliscono quali sono i temi da trattare influenzando le priorità percepite.

Se per settimane i telegiornali aprono con notizie su: immigrazione, femminicidi, guerra

è probabile che l'opinione pubblica consideri questi temi tra i più urgenti del momento.



[Violenza di genere sui quotidiani nazionali italiani](#)

Agenda Setting, **framing**, priming

Framing: inquadramento di una notizia
influenza l'interpretazione

Esempio: immigrazione descritta come
“emergenza sicurezza” o come “crisi
umanitaria”



Agenda Setting, framing, **priming**

Priming: Il priming avviene quando l'attenzione ripetuta su un tema porta le persone a usarlo come metro di giudizio ad esempio dei politici

Esempio: se l'immigrazione è ripetutamente presentato come un problema di sicurezza i politici verranno giudicati su come intervengono per gestire il problema.



Narrativa

Il modo in cui i fatti vengono selezionati, organizzati e interpretati per costruire un racconto dotato di senso.

Narrativa 1:

L'IA rivoluzionerà il lavoro e la società

Narrativa 2:

L'IA minaccia occupazione e democrazia



Esempio di insight

Nelle carceri italiane la percentuale di donne detenute rappresenta il 4.4%

Fonte Ministero della Giustizia

31 gennaio 2026

https://www.giustizia.it/giustizia/it/mg_1_14_1.page?contentId=SST1489965

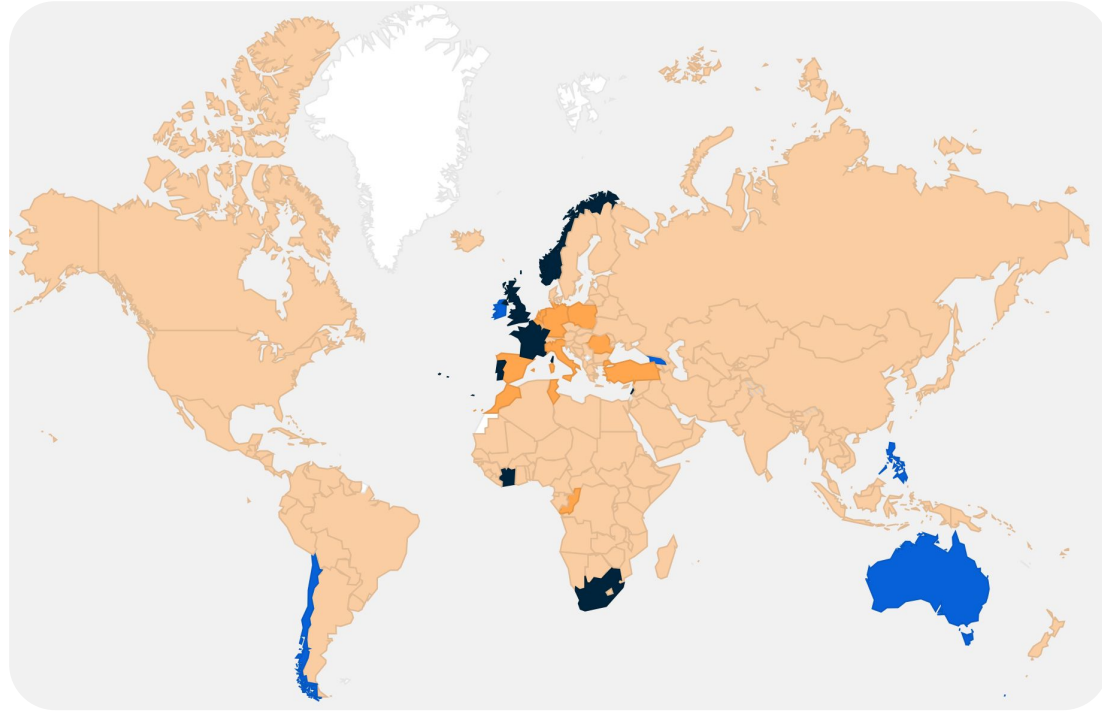
EMILIA ROMAGNA	10	2.984	3.976	169	2.037	150	59
FRIULI VENEZIA GIULIA	5	487	735	30	334	26	10
LAZIO	14	5.312	6.701	461	2.369	56	9
LIGURIA	6	1.111	1.394	71	757	36	16
LOMBARDIA	18	6.144	8.976	413	4.144	135	38
MARCHE	6	814	1.058	24	340	48	10
MOLISE	3	269	425	0	78	3	2
PIEMONTE	13	3.978	4.302	165	1.754	114	37
PUGLIA	11	2.945	4.536	204	495	157	10
SARDEGNA	10	2.583	2.460	53	779	57	4
SICILIA	23	6.439	7.064	242	1.055	170	8
TOSCANA	16	3.162	3.401	104	1.620	157	60
TRENTINO ALTO ADIGE	2	510	493	41	305	11	4
UMBRIA	4	1.340	1.700	78	595	19	2
VALLE D'AOSTA	1	181	152	0	87	1	1
VENETO	9	1.929	2.957	146	1.485	75	28
Totale	189	51.211	63.734	2.782	20.206	1.480	326

Detenuti presenti

Donne

Prison Insider

Piattaforma di riferimento per chi vuole sapere come funzionano le carceri nei vari Paesi, quali sono le condizioni di detenzione e quali problematiche si evidenziano, con dati, analisi e storie reali



<https://www.prison-insider.com/en/informer/pays>

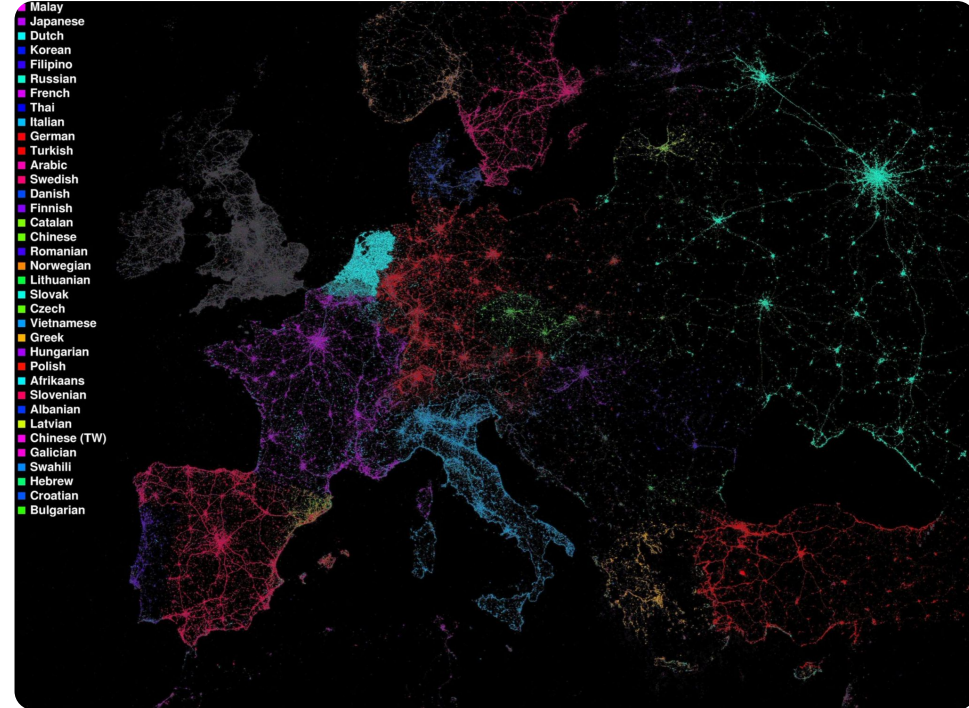
Esempio di insight

[Language communities of Twitter](#), Eric

Fisher 27 ottobre 2011

Tweet geolocalizzati e identificazione
del linguaggio

L'Insight è dato dalle eccezioni.



English
Portuguese
Indonesian
Spanish
Malay
Japanese
Dutch
Korean
Filipino
Russian
French
Thai
Italian
German
Turkish
Arabic
Swedish
Danish
Finnish
Catalan
Chinese
Romanian
Norwegian
Lithuanian
Slovak
Czech
Vietnamese
Greek
Hungarian
Polish
Afrikaans
Slovenian
Albanian
Latvian
Chinese (TW)
Galician
Swahili
Hebrew
Croatian
Bulgarian



L'Italia è un paese sicuro?

Enlight: come cambiare la propria percezione

L'Italia è un paese sicuro?
Consideriamo omicidi,
femminicidi, incidenti stradali
mortalità

DECRETO SICUREZZA

Più strumenti contro
terrorismo e criminalità
organizzata



Più tutele per forze
dell'ordine e personale
di sicurezza



Stretta sulla
cannabis light

Novità in
materia antimafia



Carcere e tutela dei
minori



Carcere e
tutela dei minori

Occupazioni abusive: pene
più severe e sgomberi
immediati



Lotta all'accattonaggio
e alla truffa

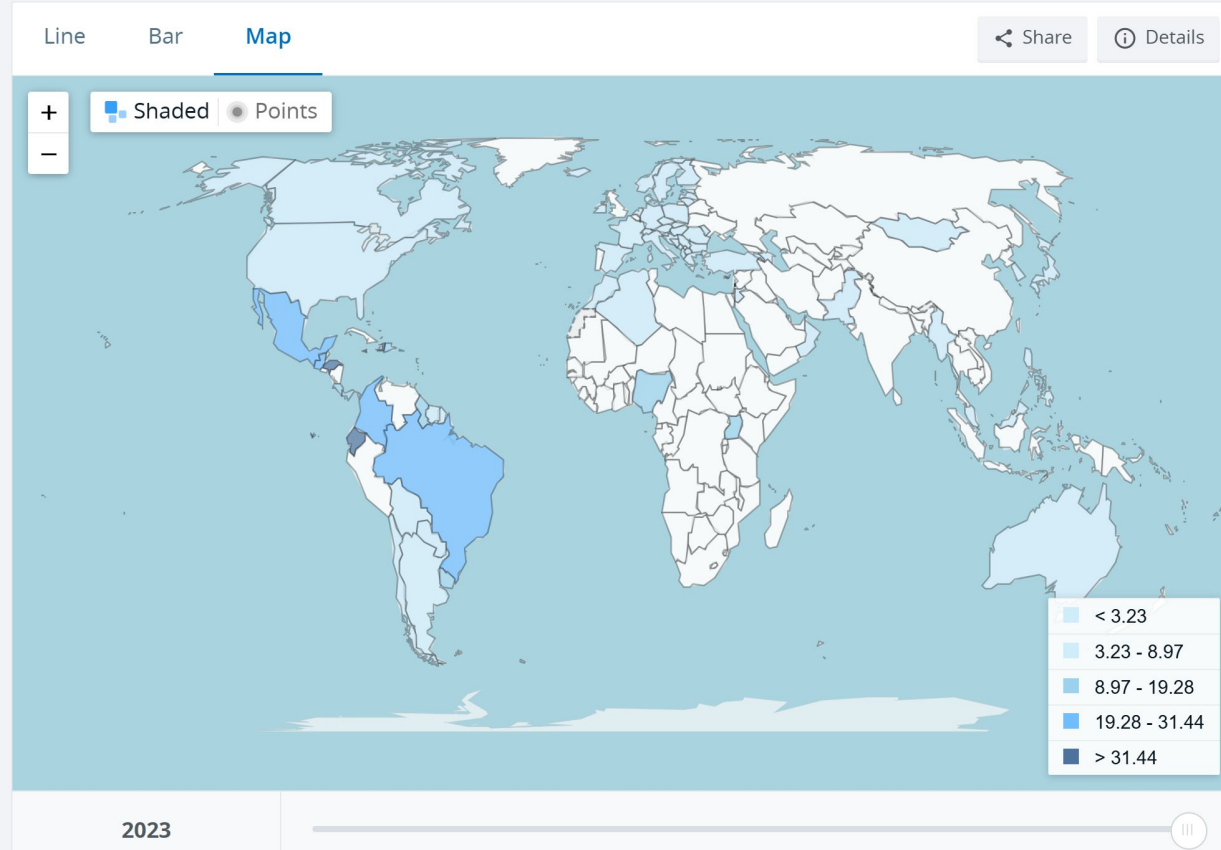


Misure per migranti e
centri di trattenimento

Intentional homicides (per 100,000 people)

UNODC Research - Data Portal – Intentional Homicide, UN Office on Drugs and Crime (UNODC)

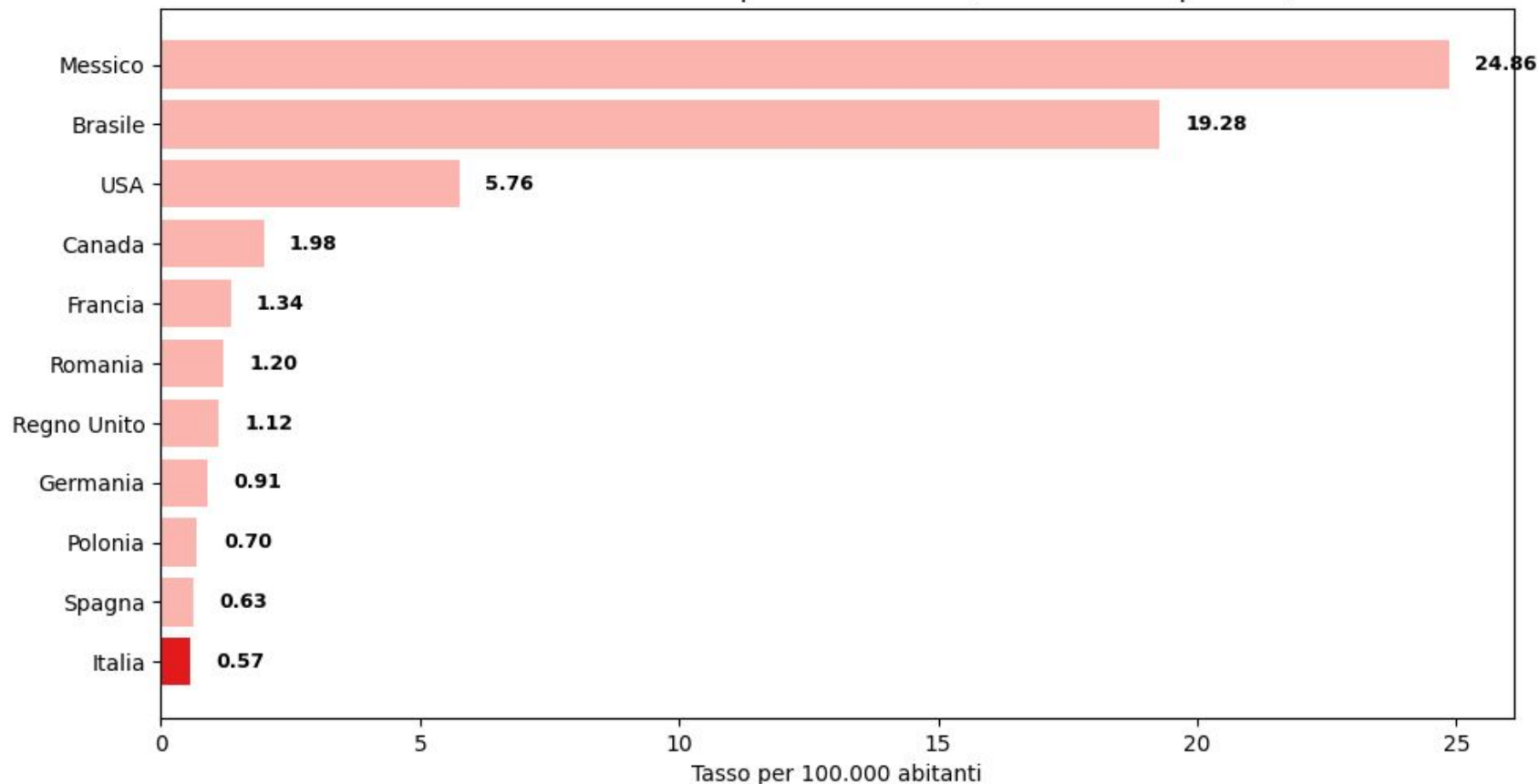
License : CC BY-4.0 [🔗](#)



Data World Bank

<https://data.worldbank.org/indicator/VC.IHR.PSRC.P5?end=2023&start=2023&view=map>

Omicidi intenzionali totali - paesi selezionati (ultimo anno disponibile)

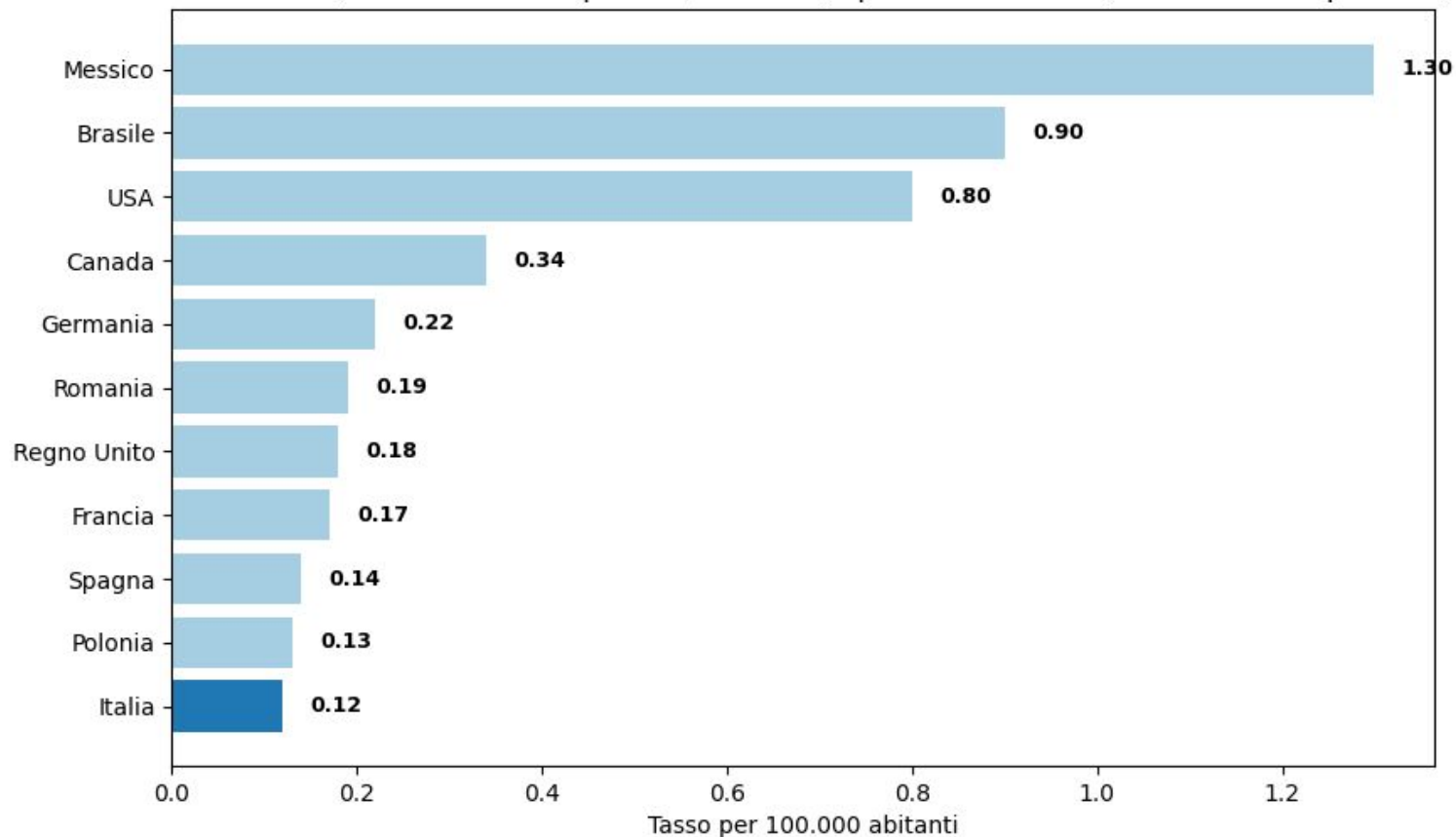


Fonte: UNODC - Global Study on Homicide / intentional homicide rate (per 100.000 abitanti).

Dataset UNODC: <https://dataunodc.un.org/dp-intentional-homicide-victims>

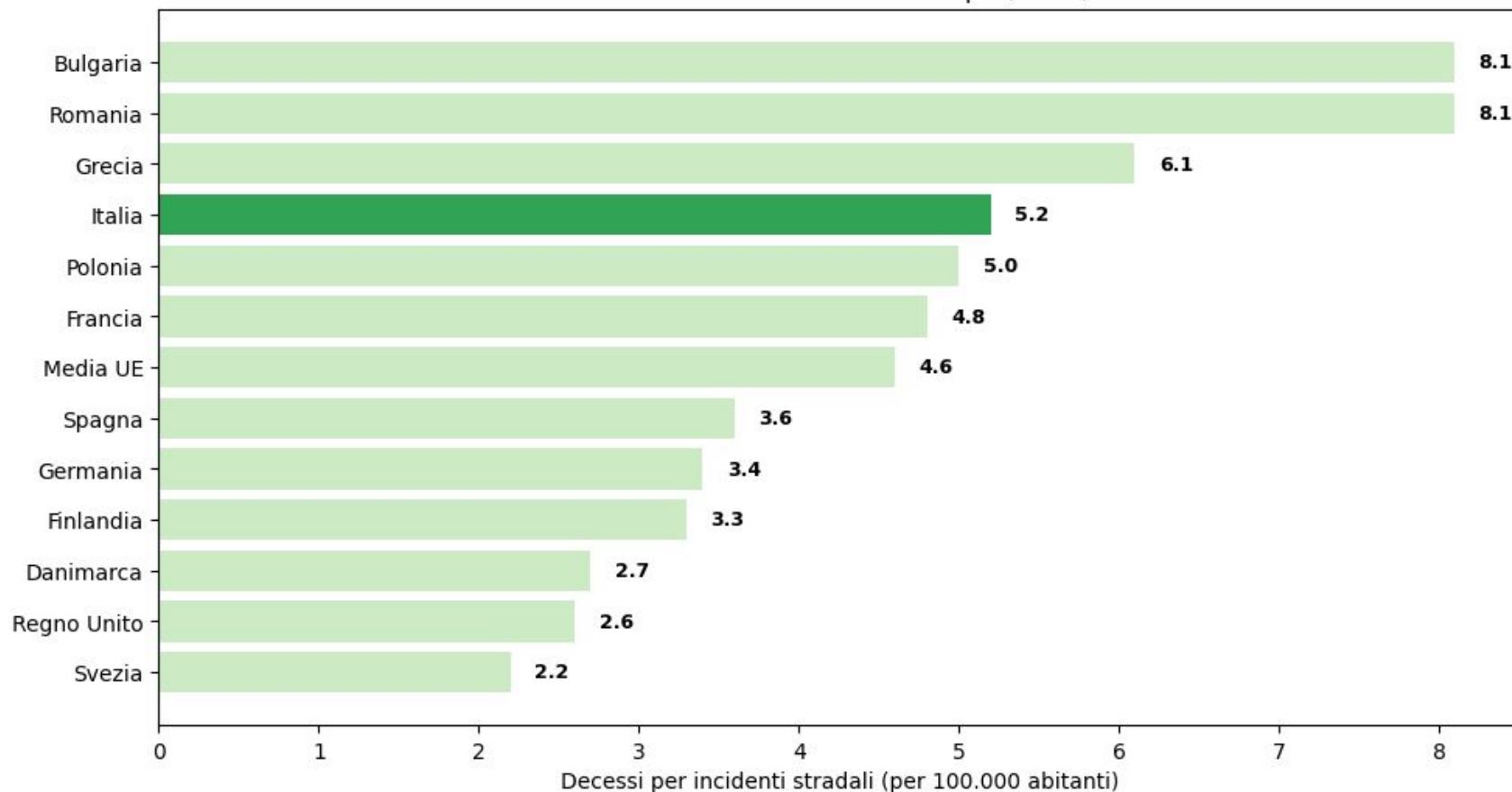
Global Study on Homicide: <https://www.unodc.org/unodc/en/data-and-analysis/global-study-on-homicide.html>

Femminicidi (donne uccise da partner/familiare) - paesi selezionati (ultimo anno disponibile)



Fonte: compilazione da UNODC (Global Study on Homicide / homicide country data) e dataset EIGE/Eurostat per dati specifici sui femminicidi in ambito familiare.
UNODC: <https://www.unodc.org/unodc/en/data-and-analysis/global-study-on-homicide.html>
EIGE / dataset crim_hom_vrel: https://dgs-p.eige.europa.eu/data/view?code=genvio_int_adm_oth_crim_hom_vrel

Tassi di mortalità da incidenti stradali in Europa (2023) - con Media UE



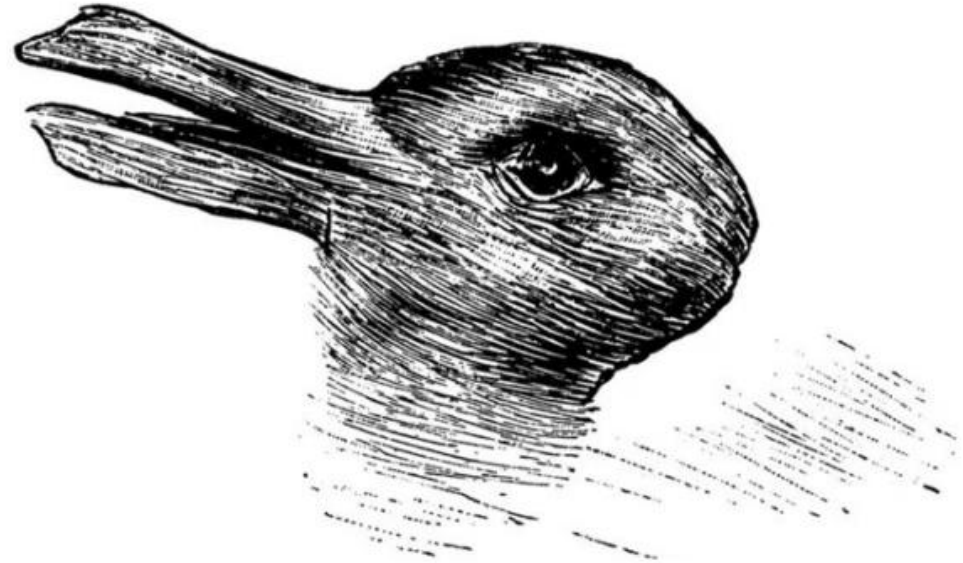
Fonte: Commissione Europea / Eurostat - Road safety statistics 2023 (dati CARE).

Eurostat: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Road_safety_statistics

Esempio di enlight: femminicidi, omicidi e incidenti stradali

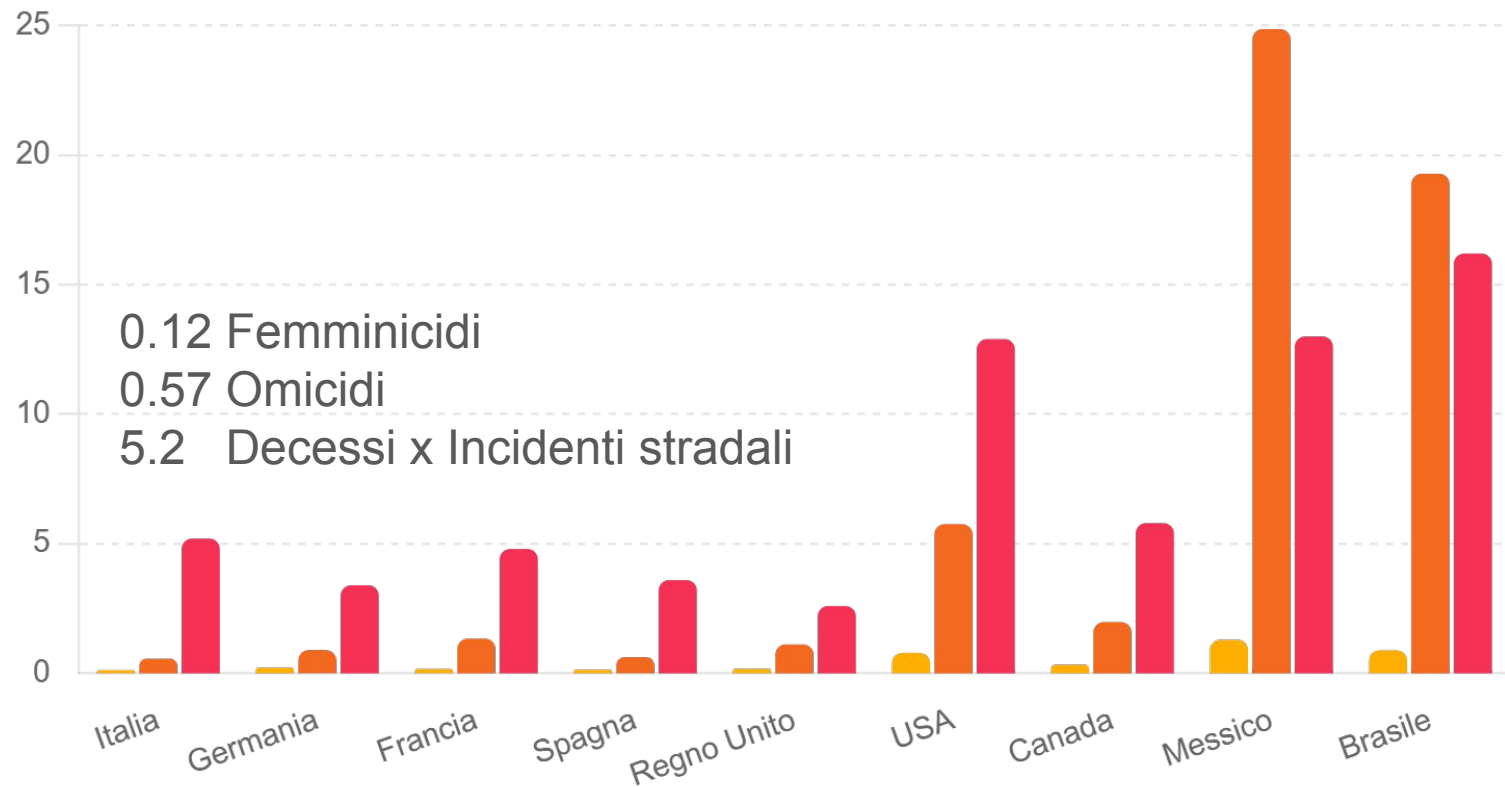
La percezione degli italiani è diversa, ma i dati parlano chiaro: l'Italia è diventata uno dei Paesi più sicuri d'Europa per **omicidi** e **femminicidi**, mentre resta in forte ritardo sulla **sicurezza stradale**

Federico Fubini, 9 febbraio 2026, Corriere.



Illusione Anatra-Coniglio - 1892

Femminicidi, Omicidi e Incidenti stradali



Confirmation Bias

È la tendenza a:

- Cercare informazioni che confermano ciò che già crediamo
- Dare più peso alle notizie coerenti con le nostre opinioni
- Ignorare o svalutare quelle che le contraddicono

Concetti collegati:

- **Filter bubble:** personalizzazione algoritmica che rafforza contenuti affini
- **Echo chamber:** ambienti informativi che riflettono le proprie idee



Correlazione $\neq$ Causalità

Correlazione significa che due variabili si muovono insieme.

Causalità significa che una variabile provoca l'altra.

Aumentano le vendite di gelato

Aumentano gli annegamenti

Sono correlate non sono causali

Il gelato non causa gli annegamenti.

La causa nascosta è la temperatura estiva.



News sul corso di Data Journalism

Strumenti per affrontare il corso

Pagina del corso su [Didawiki](http://didawiki.cli.di.unipi.it/doku.php/informaticaumanistica/dj/start)

<http://didawiki.cli.di.unipi.it/doku.php/informaticaumanistica/dj/start>

Prerequisiti informatici

Linguaggi: HTML, CSS, JAVASCRIPT, Python.

Formato dati: CSV, Json

Prerequisiti umanistici

Capacità di strutturare una narrazione

Capacità di strutturare una narrazione

Informazione pura:

“Il tasso di disoccupazione è salito al 9,2%.”

Narrativa giornalistica:

“Per Marco, 52 anni, la crisi non è un numero: è il terzo contratto non rinnovato in due anni. Intanto il tasso di disoccupazione sale al 9,2%.”

Nel secondo caso i dati sono inseriti in una cornice narrativa.

Struttura del corso

Parte teorico-pratica

- Workflow del Data Journalism
- Data Collection (web scraping)
- Data Cleaning
- Data Exploration (Tableau)
- Data Visualization
- Data Storytelling

Parte progettuale

Progetto di gruppo (1-3 persone)

- Domanda di ricerca
- Raccolta dati
- Esplorazione/Analisi dei dati
- Visualizzazione
- Racconto finale
- Pubblicazione

Esame

Progetto di data Journalism

- Lavoro di gruppo da 1 a 3 persone max
- Scelta del progetto concordata con i docenti
- Sottomettere il codice e i dati
- L'esame orale è una discussione del progetto seguito da alcune domande sui contenuti del corso

Cominciate subito a pensare ad un possibile progetto! Fin dalla 4° lezione inizieremo a discuterne

Elementi di valutazione

Quantità e qualità dei dati raccolti

Quantità e qualità delle interviste

Visualizzazione dei dati raccolti

Interesse della storia

Presentazione della storia

Cosa imparano

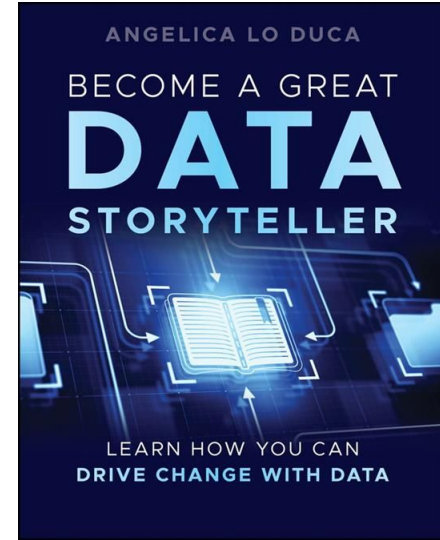
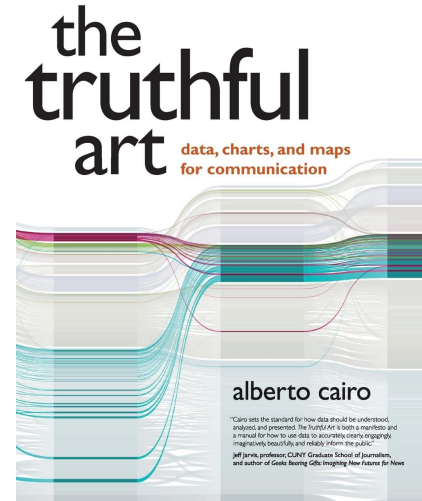
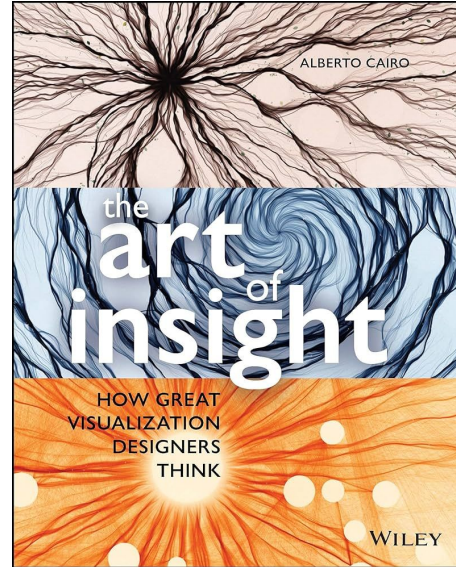
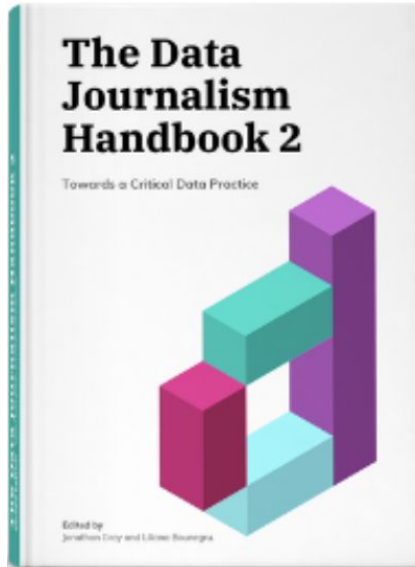
Competenze tecniche

- Scraping di dati dal web
- Pulizia dataset
- Uso di Tableau
- Costruzione di Visualizzazioni efficaci
- Tecniche di data storytelling

Competenze critiche

- Valutare qualità dei dati
- Riconoscere bias
- Evitare manipolazioni visive
- Trasformare numeri in storie

Riferimenti bibliografici



Bibliografia

Gray, Bounegru, Chambers – The Data Journalism Handbook (open access)

Alberto Cairo – The Functional Art

Alberto Cairo – How Charts Lie

Angelica Lo Duca - How to become a great storyteller

Sitografia

<https://openvisualizationacademy.org/>

Dataroom di Milena Gabanelli

Milena Gabanelli è una giornalista, famosa per Report (1997-2017) e dal 2017 cura la rubrica Dataroom sul Corriere, basata su inchieste di data journalism per spiegare fenomeni sociali e di attualità principalmente italiani.

