

# Tableau: Estrazione di insight

Angelica Lo Duca  
angelica.loduca@iit.cnr.it

## Due tecniche per analizzare i dati

- Approccio più giornalistico
  - Formulo delle domande e cerco la risposta nei dati.
- Approccio basato sulla “data science”
  - Parto dai dati e cerco di capire cosa nascondono.

## Due tecniche per analizzare i dati

- Approccio più giornalistico
  - Formulo delle domande e cerco la risposta nei dati.
- Approccio basato sulla “data science”
  - Parto dai dati e cerco di capire cosa nascondono.

Che cos'è un insight?

Un insight è una scoperta all'interno dei dati.

E' qualcosa di interessante, come ad esempio un cambiamento nel tempo, una anomalia o altro.

Come estrarre un insight?

## Tipologie Analisi

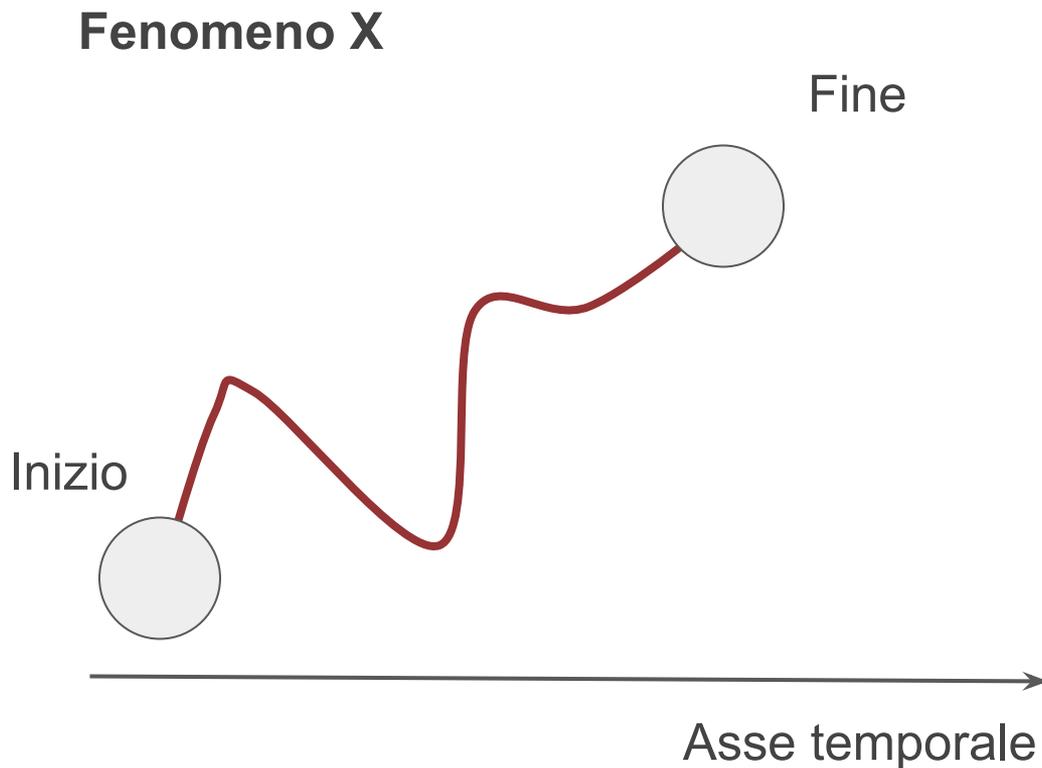
- Analisi temporale
- Zoom
- Confronto tra più categorie
- Analisi spaziale
- Valori anomali

Faremo degli esempi dal  
Global Terrorism Database

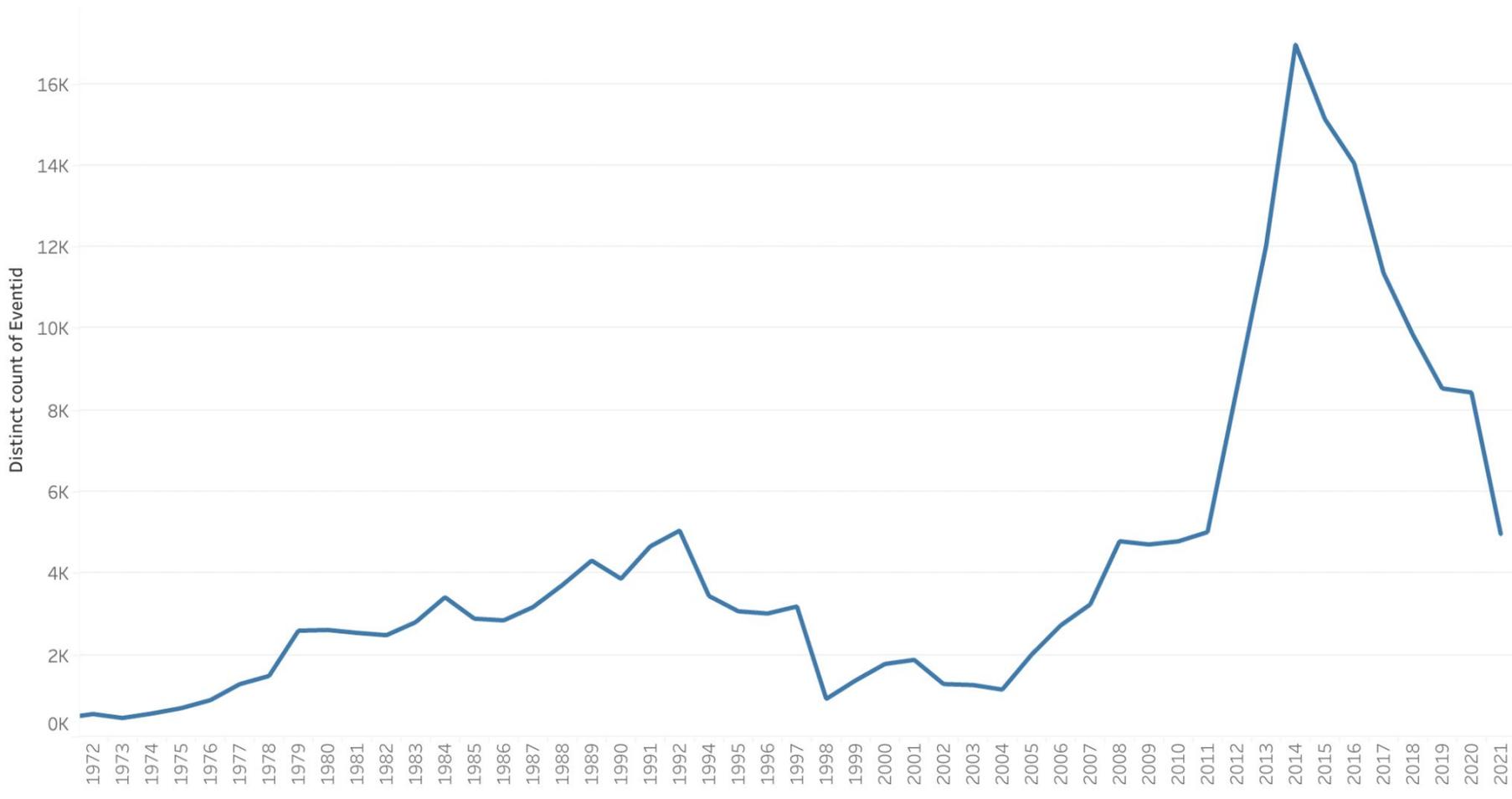
# Analisi Temporale

# Analisi Temporale: Narrare i cambiamenti nel tempo

I dati sono utilizzati per analizzare l'evoluzione nel tempo di un fenomeno.



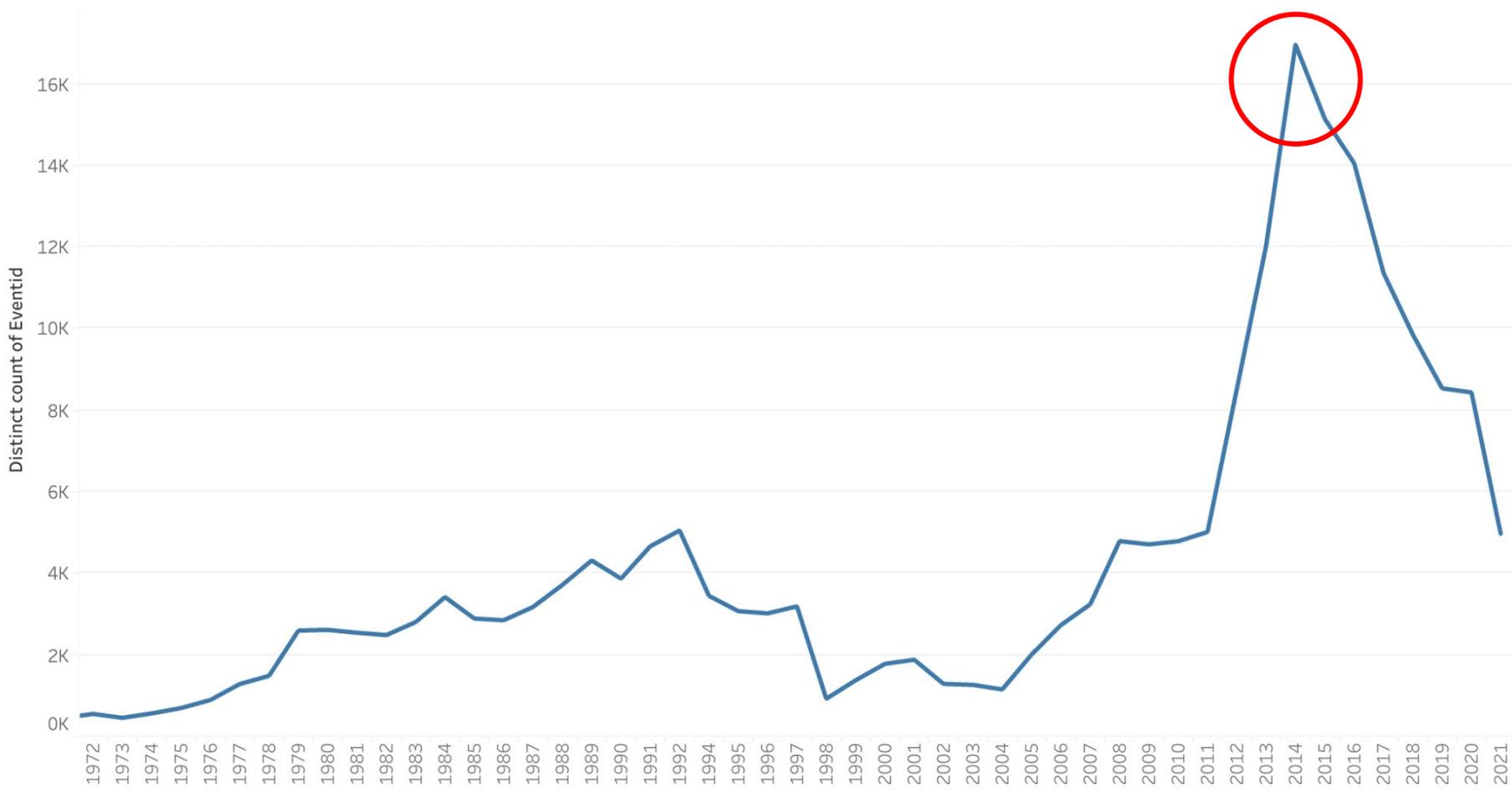
# Number of Events by Date



## Prima strategia:

cerco un picco nei dati e cerco di capire cosa accade nei pressi del picco (prima, durante e dopo).

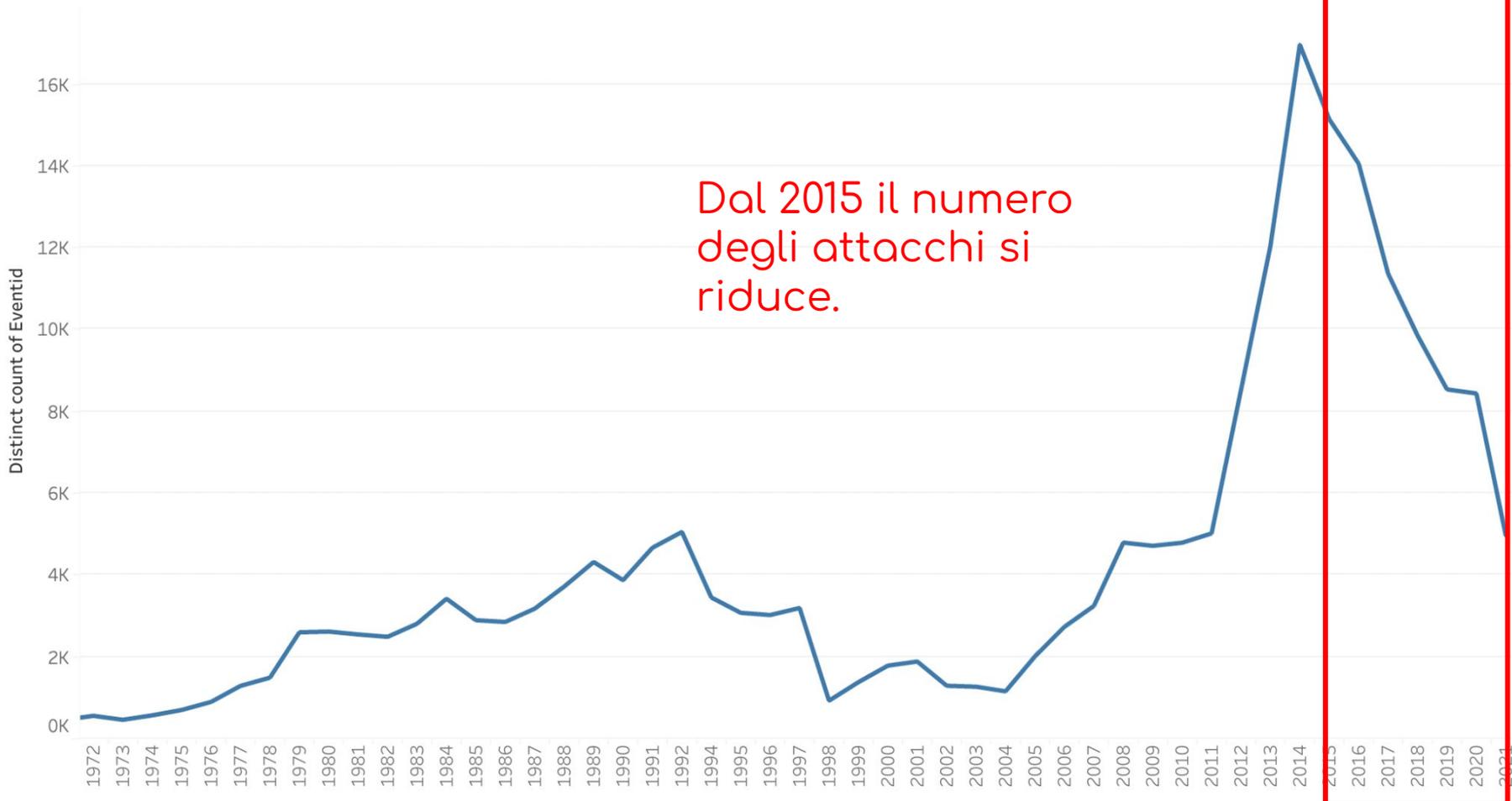
# Number of Events by Date



# Number of Events by Date



# Number of Events by Date



Cerco cosa è successo negli anni in  
corrispondenza del picco.

2011 - L'inizio risale al 2001 con la  
caduta delle Torri Gemelle  
2015 - Perdita di territorio dell'ISIS

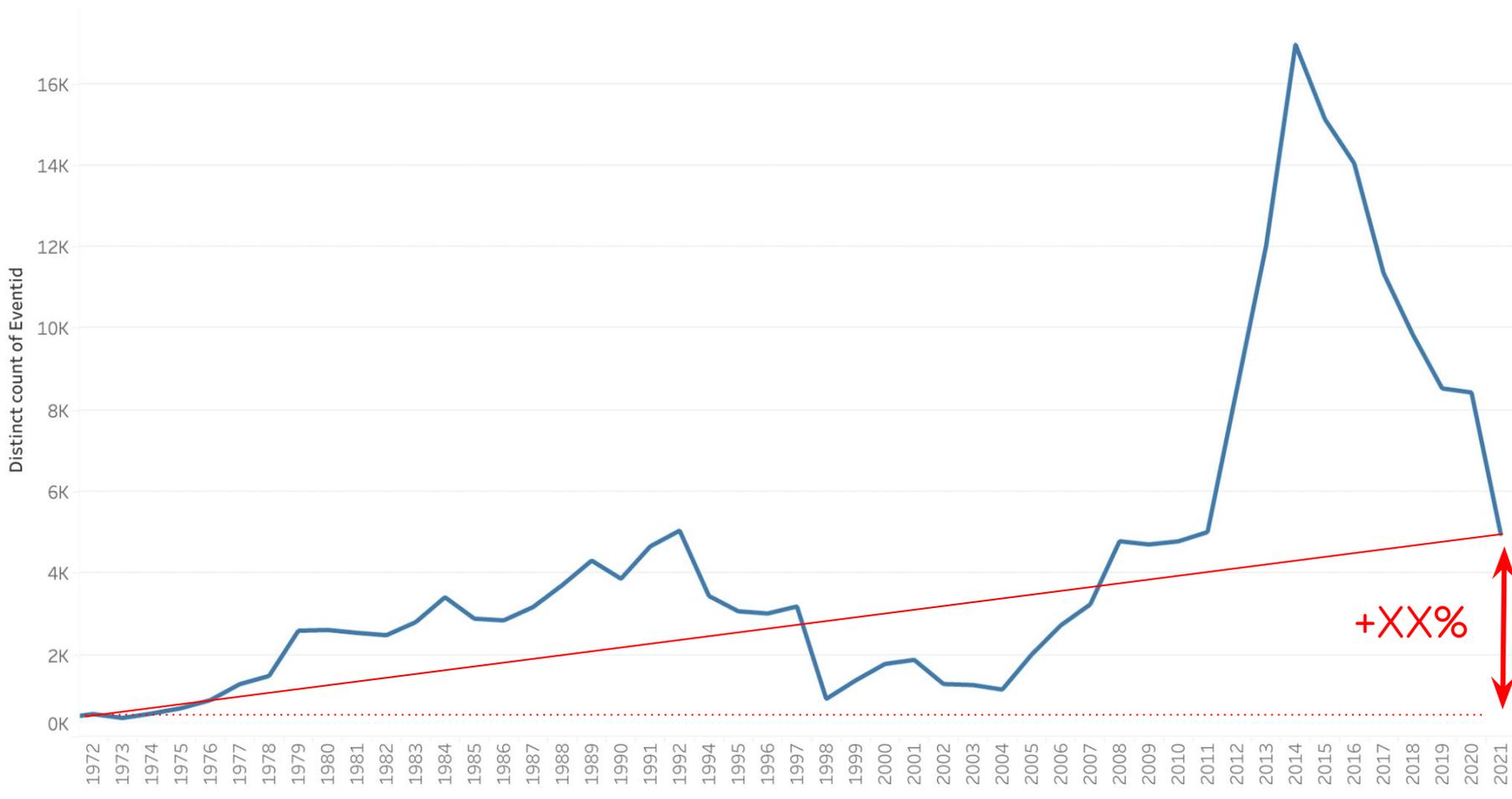
Approfondisco questi eventi,  
cercando altri dati o altre  
sorgenti.

## Seconda strategia:

Confronto gli anni agli estremi e calcolo l'incremento percentuale.

Incr. perc. = (valore anno finale - valore anno iniziale) / valore anno iniziale \* 100

# Number of Events by Date



Approfondisco l'argomento per capire  
cosa ha determinato il cambiamento  
negli anni (altre fonti, altri dati).

## Terza strategia:

### Confronto con dei valori soglia

- Confronto con un valore base
- Uso di una soglia

# Altri esempi

## Esempio 1 – Salute:

Tra il 2000 e il 2020 si osserva una diminuzione del tasso di mortalità per malattie cardiovascolari nei Paesi dell'UE. Il calo coincide con l'introduzione di nuove linee guida sanitarie e l'aumento della prevenzione primaria.

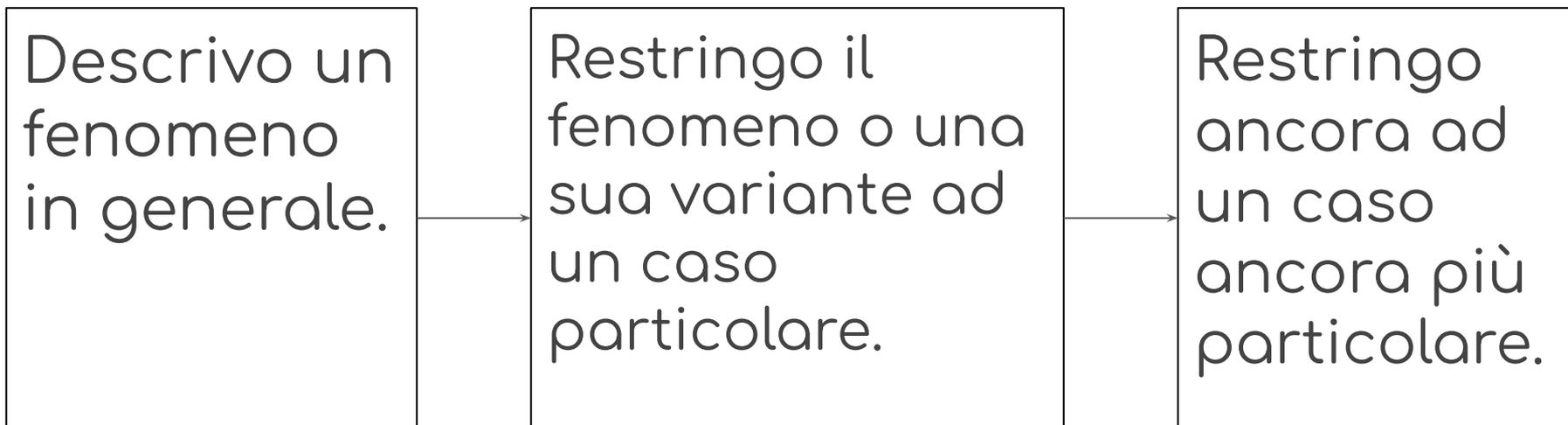
## Esempio 2 – Economia:

Durante la crisi economica del 2008, il tasso di disoccupazione giovanile in Italia è salito vertiginosamente. Confrontando il 2007 e il 2013, si nota un incremento del 60%, suggerendo una forte correlazione con la recessione globale.

Zoom

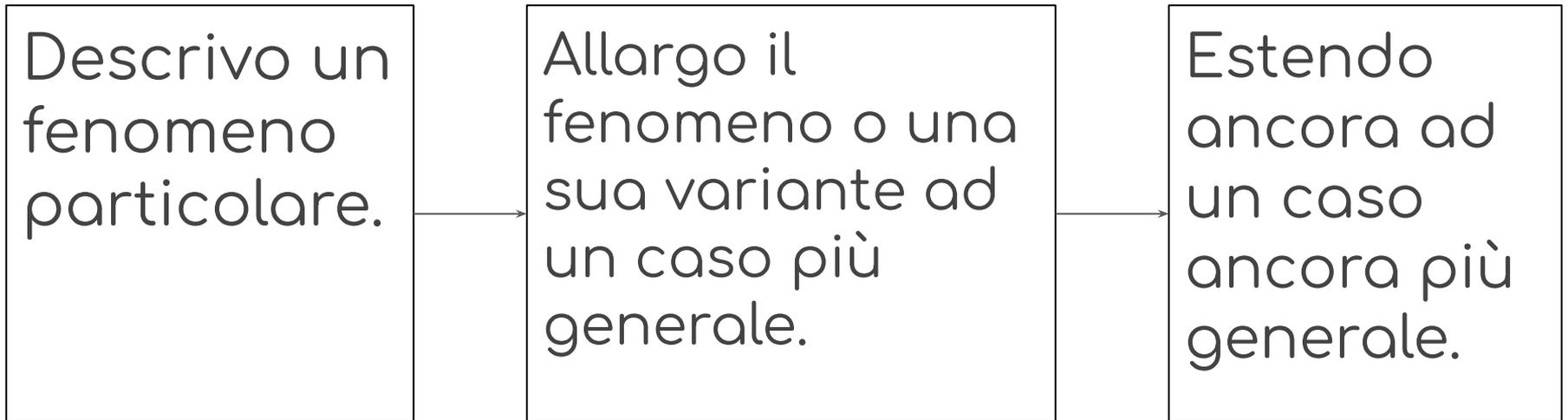
## Zoom In: Dalla panoramica al dettaglio

Si parte da una situazione generale e si va verso un caso particolare.



## Zoom Out: Dal dettaglio alla panoramica

Si analizza prima un particolare dell'intera informazione disponibile e dopo si espande la sua vista ad una prospettiva più generale.





## Prima Strategia:

Mi concentro su una regione o un Paese specifico e poi passo ad un livello più ampio (ad esempio continenti).

# Numero di morti in seguito ad attacchi terroristici



Descrivo la  
situazione in Italia.

## Numero di morti in seguito ad attacchi terroristici



Estendo la mia  
descrizione a tutta  
l'Europa: mostro  
quindi dati  
aggregati a livello di  
continenti.

Posso estendere la mia descrizione a tutto il mondo.

In questo caso, mostro solo un numero senza mappa, che corrisponde al numero totale di morti.

Vale anche il viceversa (zoom out).  
Parto dal mondo, poi passo ad un  
continente, ed infine, ad un Paese.

Posso usare dati diversi (ma simili) per livelli diversi di zoom.

Ad esempio, prendo i dati italiani da ISTAT e quelli europei da Eurostat.

## Seconda Strategia:

Mi concentro su una regione o un Paese specifico, ne calcolo una metrica specifica (ad esempio il valor medio) e lo confronto con la stessa metrica a livello più ampio.

Ad esempio confronto il numero medio di attacchi terroristici in Italia per anno con il valor medio in Europa e nel mondo.

# Altri esempi

Esempio 1 – Ambiente (Zoom In):

Parto dall'analisi delle emissioni globali di CO<sub>2</sub>, restringo l'osservazione all'Europa, poi alla Germania. Scopro che la Germania ha ridotto del 25% le sue emissioni tra il 2000 e il 2020, soprattutto grazie alla transizione energetica ("Energiewende").

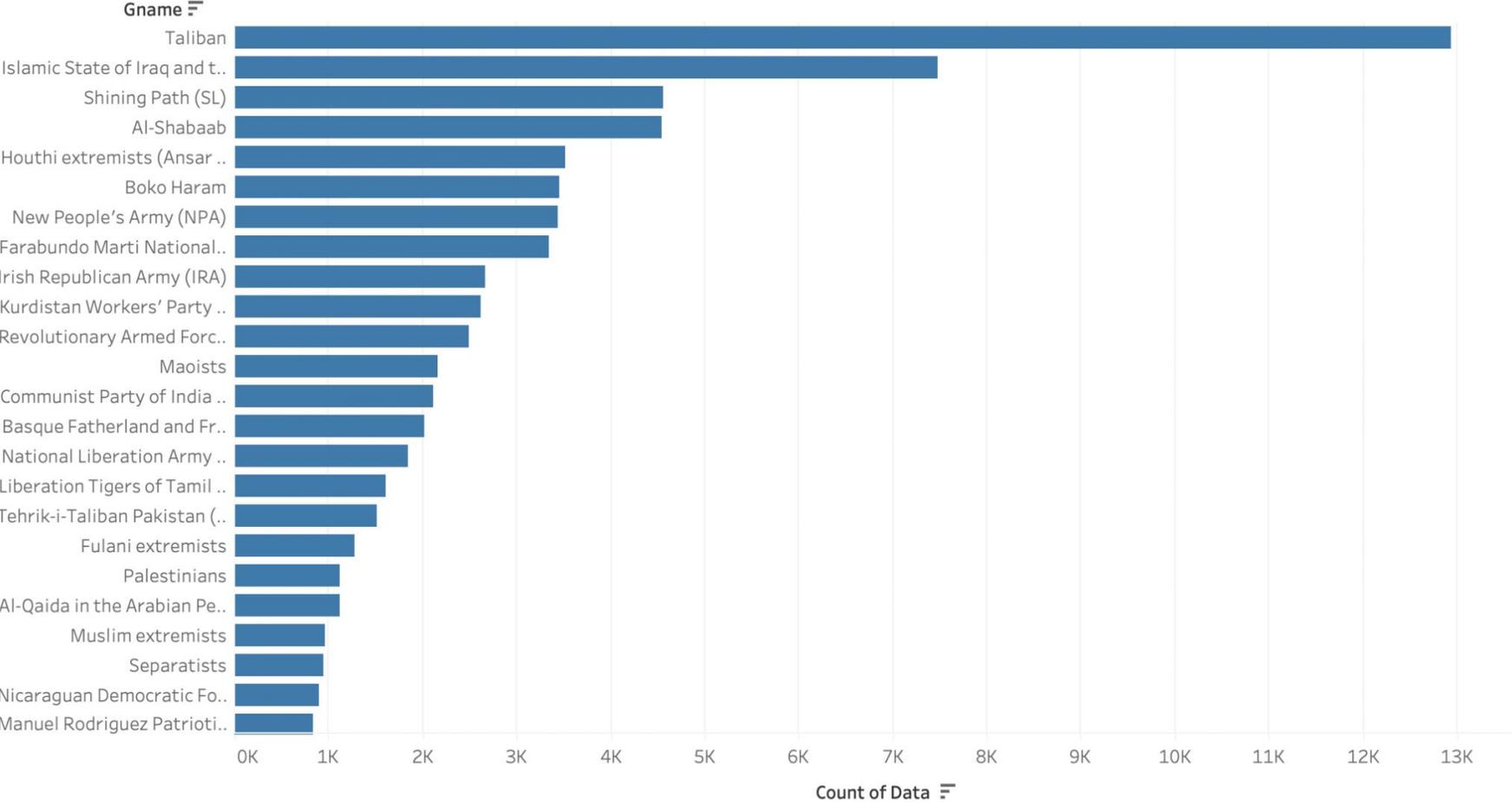
Esempio 2 – Istruzione (Zoom Out):

Esamino il tasso di abbandono scolastico in Sicilia, poi lo confronto con il dato del Sud Italia e infine con l'intera nazione. Questo mostra che la Sicilia ha uno dei valori più alti, contribuendo significativamente alla media nazionale.

Confronto tra più categorie

Confronto due o più categorie

# NAttackxGroup

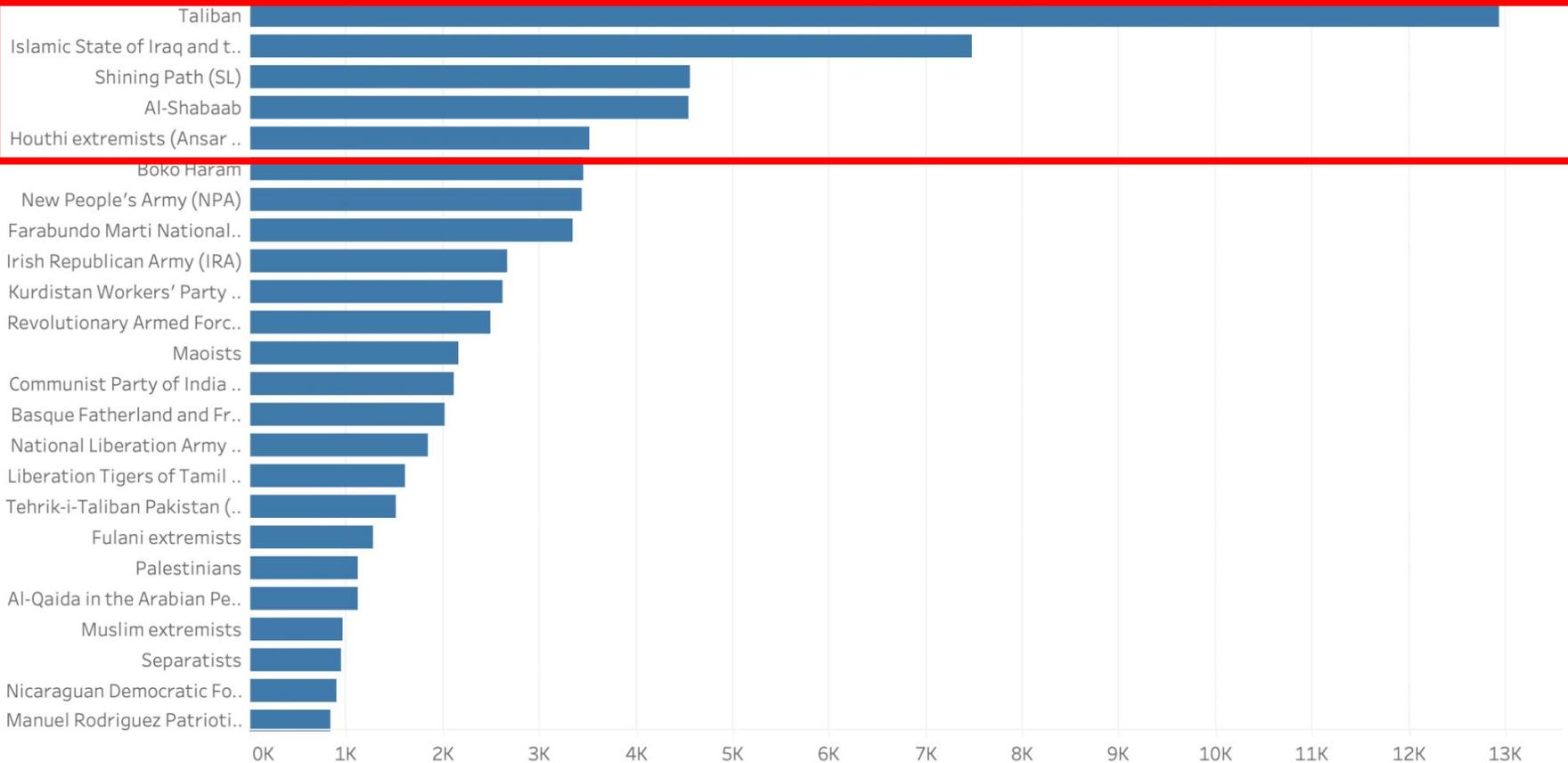


## Prima Strategia:

Mi concentro sulle prime N categorie e cerco informazioni su ciascuna di esse.

# NAttackxGroup

Gname 



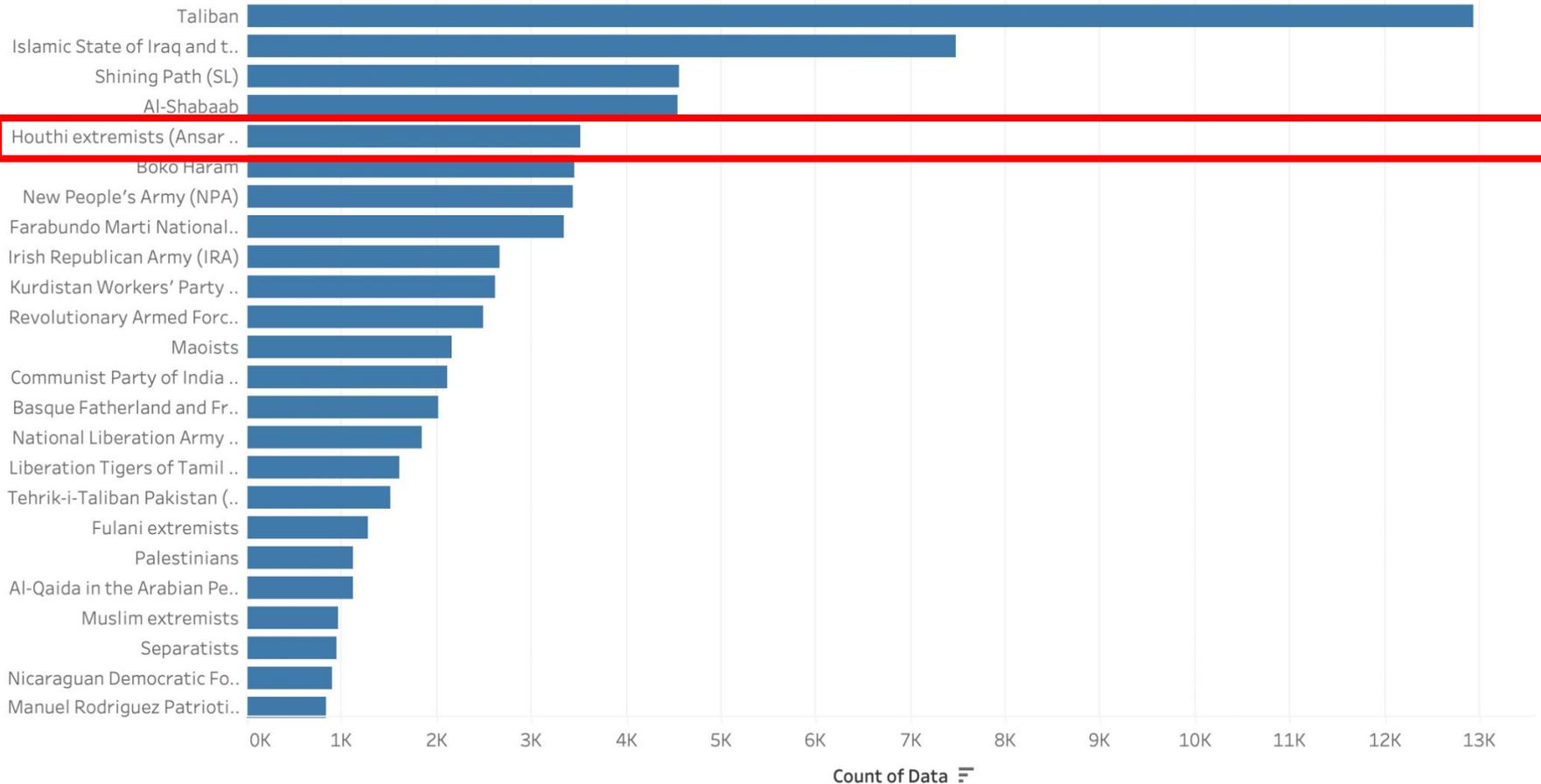
Count of Data 

## Seconda Strategia:

Mi concentro su una categoria e cerco di capire perché occupa quella posizione.

# NAttackxGroup

Gname



## Terza Strategia:

Mi concentro su due categorie e confronto la loro evoluzione nel tempo.

# Posso anche cercare di capire se esistono delle correlazioni

## Cause and Effect

Changes in one factor affect changes in another factor

## Common Cause

An external factor causes changes in both factors

## Accidental Relationship

There is a relationship between the two factors even though there is no cause-and-effect relationship

## Presumed Relationship

A correlation does not appear to be accidental, but there is no cause-and-effect relationship or common cause relationship

# Altri esempi

## Esempio 1 – Economia:

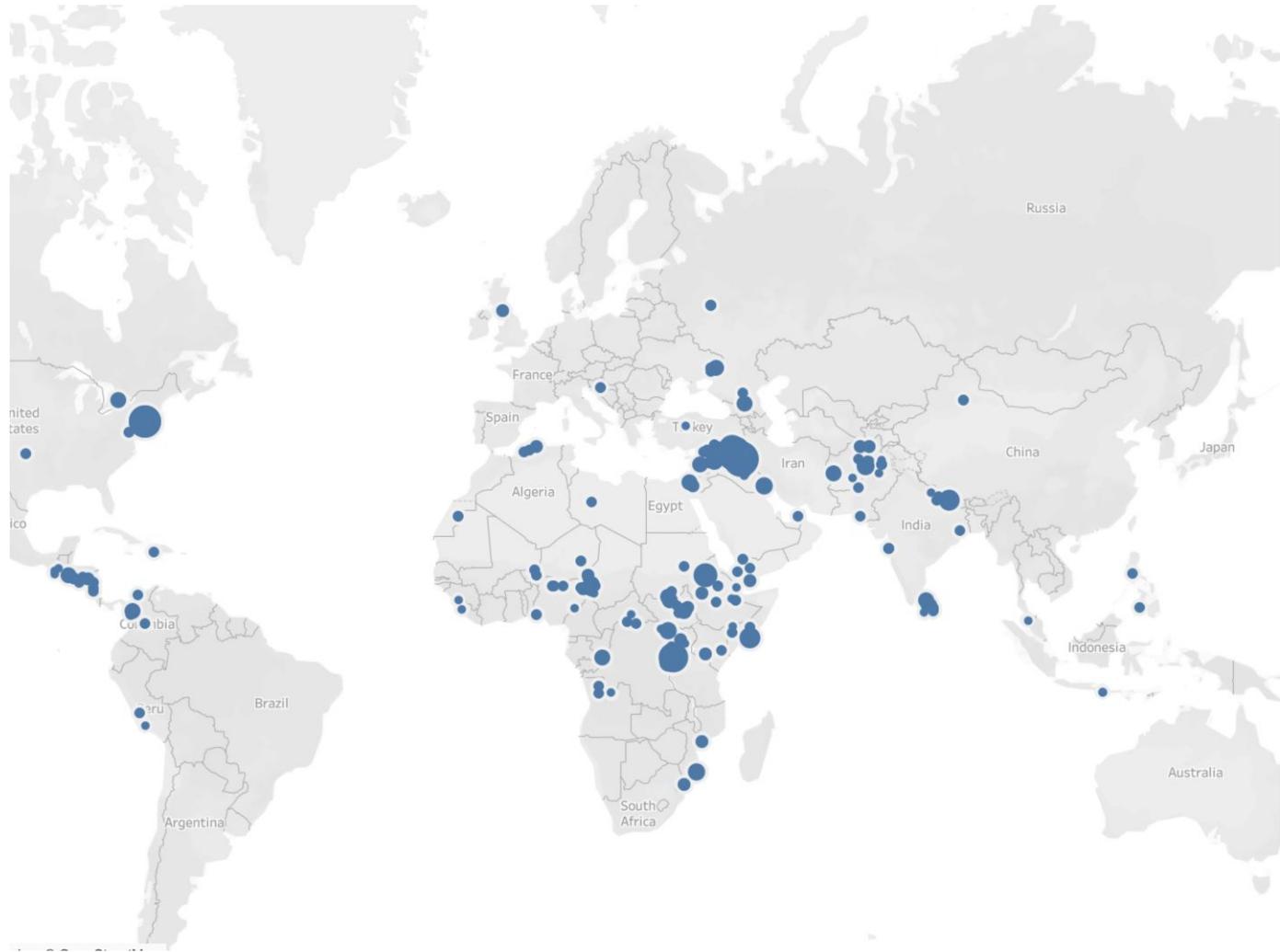
Confronto l'inflazione in Germania, Spagna e Italia tra il 2015 e il 2022. La Spagna mostra oscillazioni maggiori, forse a causa della forte dipendenza dal turismo e dell'impatto della pandemia su questo settore.

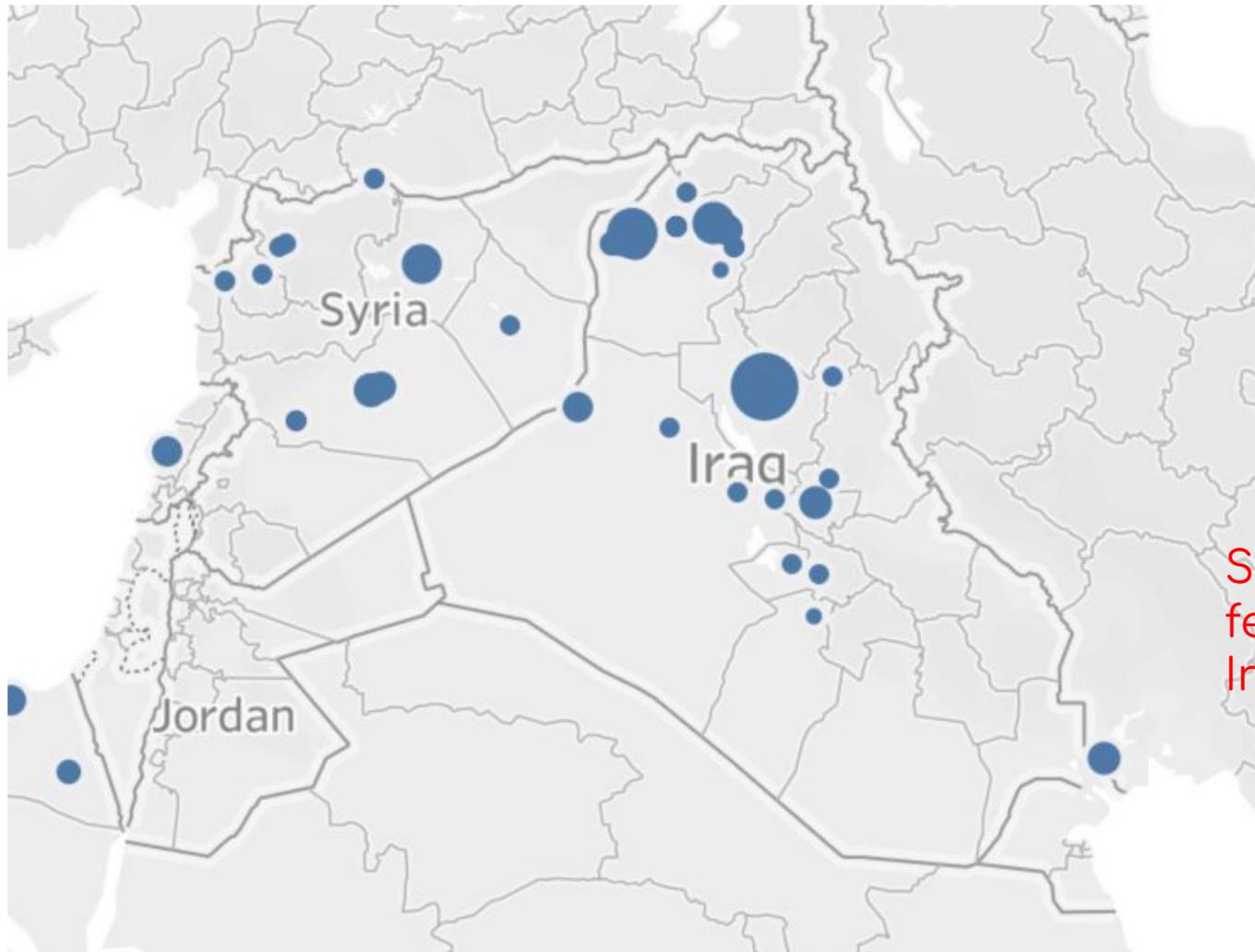
## Esempio 2 – Salute:

Metto a confronto le cause principali di morte in Francia: malattie cardiovascolari, tumori e diabete. I tumori risultano essere la causa dominante dal 2017, superando le malattie cardiovascolari per la prima volta.

# Analisi spaziale

Confronto la stessa categoria o lo stesso fenomeno in due o più luoghi diversi della stessa dimensione (ad esempio in due città, in due nazioni, ecc).





Spiego come varia il fenomeno in Syria e Iraq.

Posso anche usare un grafico a barre o colonne per confrontare i luoghi diversi.

In questo caso, si ricade nell'analisi che fa il confronto tra categorie.

# Altri esempi

## Esempio 1 – Salute:

Confronto i tassi di vaccinazione antinfluenzale tra Lombardia e Sicilia. La Lombardia mostra una copertura superiore del 30%, probabilmente grazie a una maggiore organizzazione delle campagne vaccinali e accessibilità ai servizi sanitari.

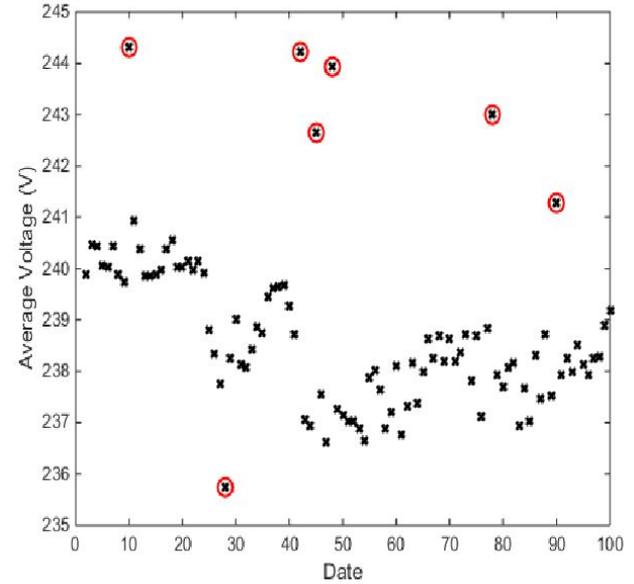
## Esempio 2 – Economia:

Analizzo il reddito medio disponibile in Piemonte e Campania. Il Piemonte mostra valori più alti del 40%, suggerendo disuguaglianze territoriali nel livello di benessere economico.

Valori anomali

# Points Anomalies\*

Una point anomaly è quando un singolo dato si distingue dal pattern, range o norma previsti. In altre parole, il punto dati è inaspettato.

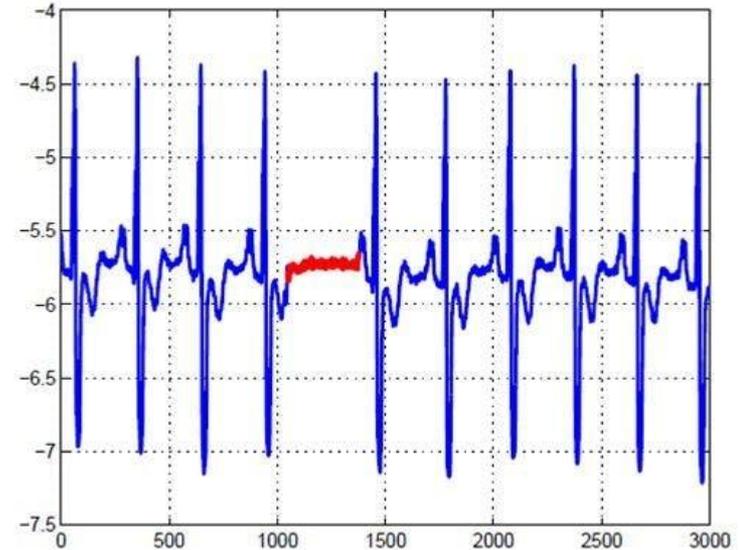


\*3 Types of Anomalies in Anomaly Detection

<https://hackernoon.com/3-types-of-anomalies-in-anomaly-detection>

# Collective Anomalies\*

Si verifica un'anomalia collettiva quando singoli punti dati esaminati isolatamente sembrano normali. Quando si esamina un gruppo di questi punti dati, tuttavia, diventano chiari modelli, comportamenti o risultati inaspettati.

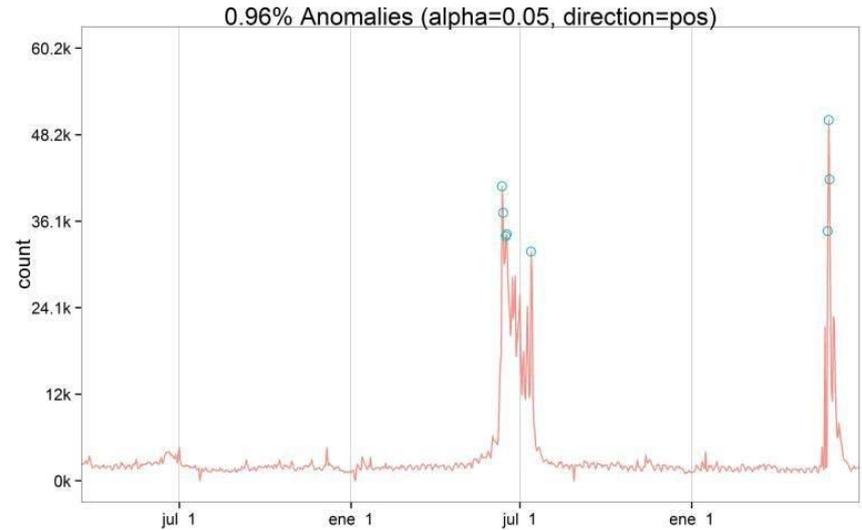


\*3 Types of Anomalies in Anomaly Detection

<https://hackernoon.com/3-types-of-anomalies-in-anomaly-detection>

# Contextual Anomalies\*

Risultati inaspettati che derivano da quella che sembra un'attività normale.



\*3 Types of Anomalies in Anomaly Detection

<https://hackernoon.com/3-types-of-anomalies-in-anomaly-detection>

# Altri esempi

Esempio 1 – Istruzione (Point Anomaly):

Nel 2020, una scuola in provincia di Bolzano registra un tasso di promozione del 100% durante la pandemia. Un'anomalia rispetto alla media regionale del 92%, probabilmente dovuta all'adozione anticipata della didattica a distanza.

Esempio 2 – Economia (Contextual Anomaly):

Il tasso di disoccupazione cala nel 2020 in Italia, un dato anomalo visto il lockdown. Contestualizzando, si scopre che molti lavoratori non erano considerati "disoccupati" poiché impossibilitati a cercare lavoro attivamente.