How to Extract an Insight from Data

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

An insight is a discovery within data. It is something interesting, such as a change over time, an anomaly, or something else.

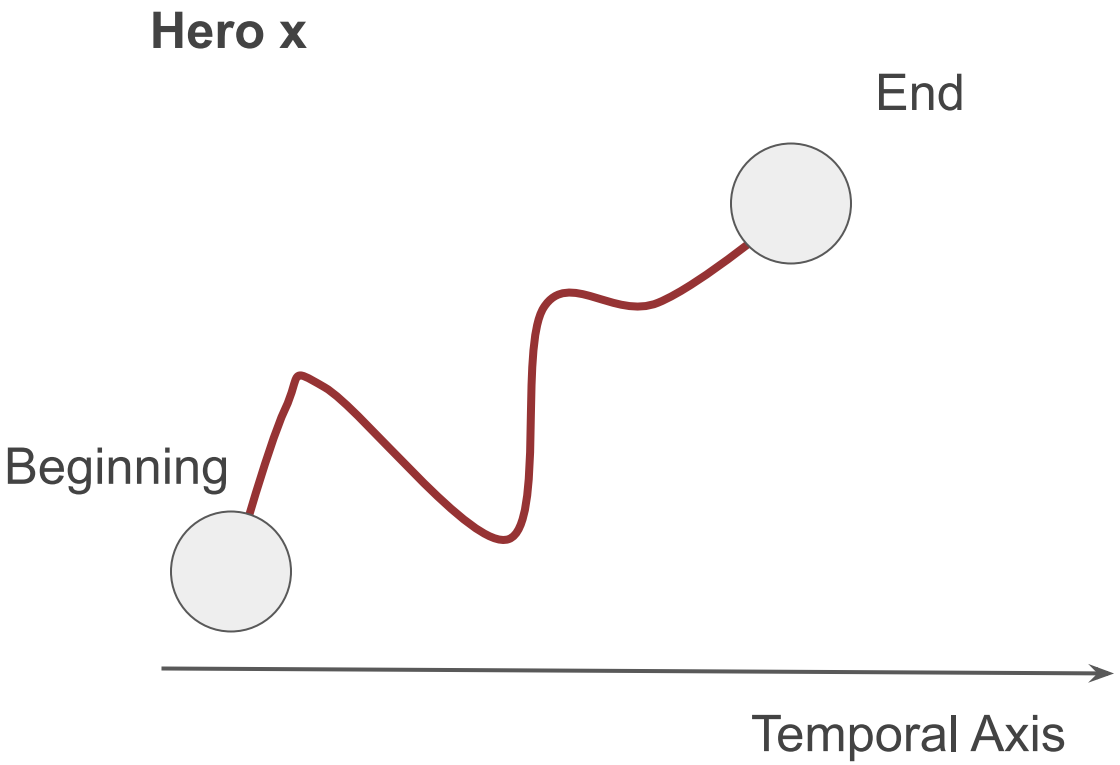
How to extract an insight

- Temporal Analysis
- Zoom
- Multi-Category Comparison
- Spatial Analysis
- Outliers

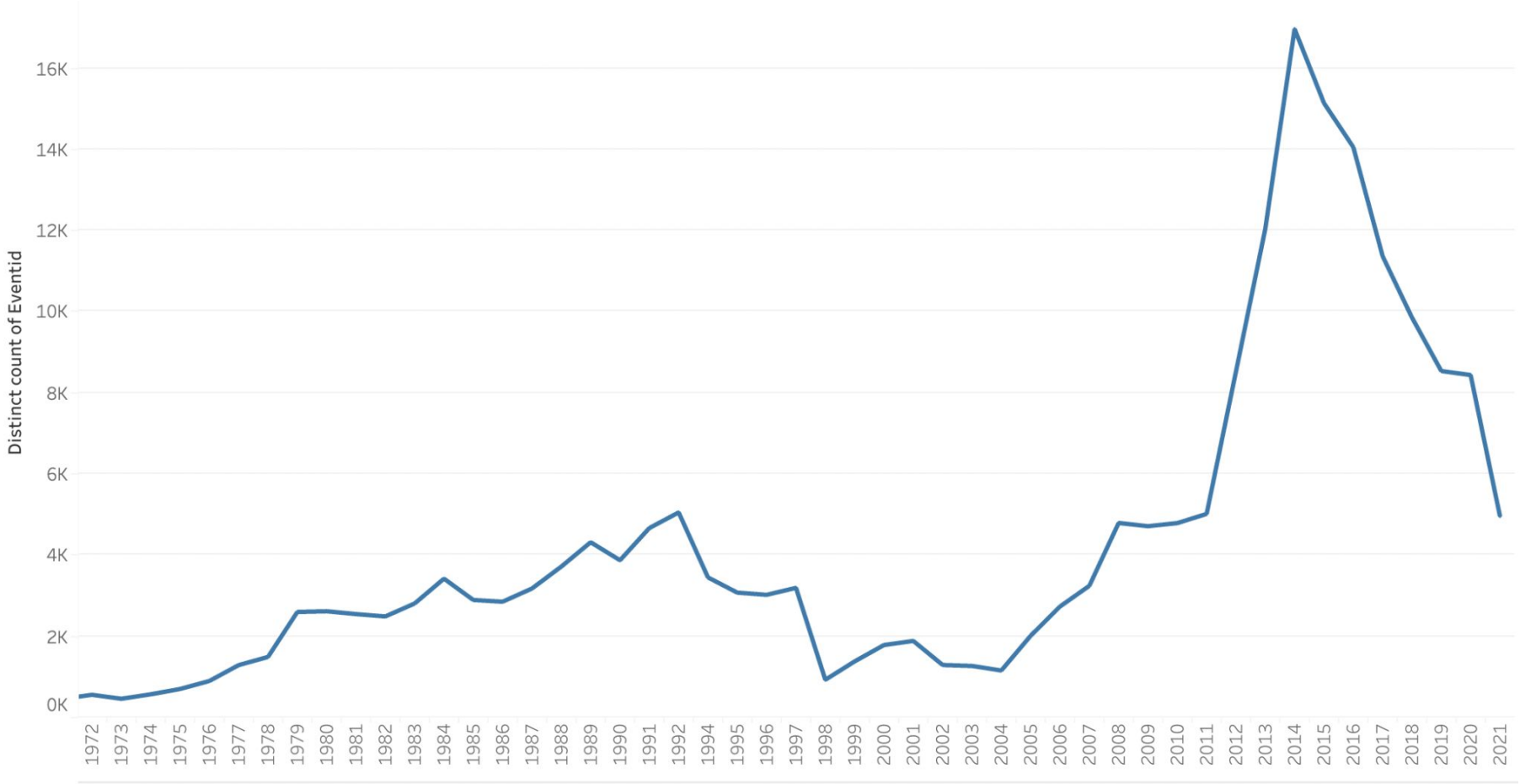
Temporal Analysis

Temporal Analysis: Narrating changes over time

The data is used to analyze the evolution of a phenomenon over time.



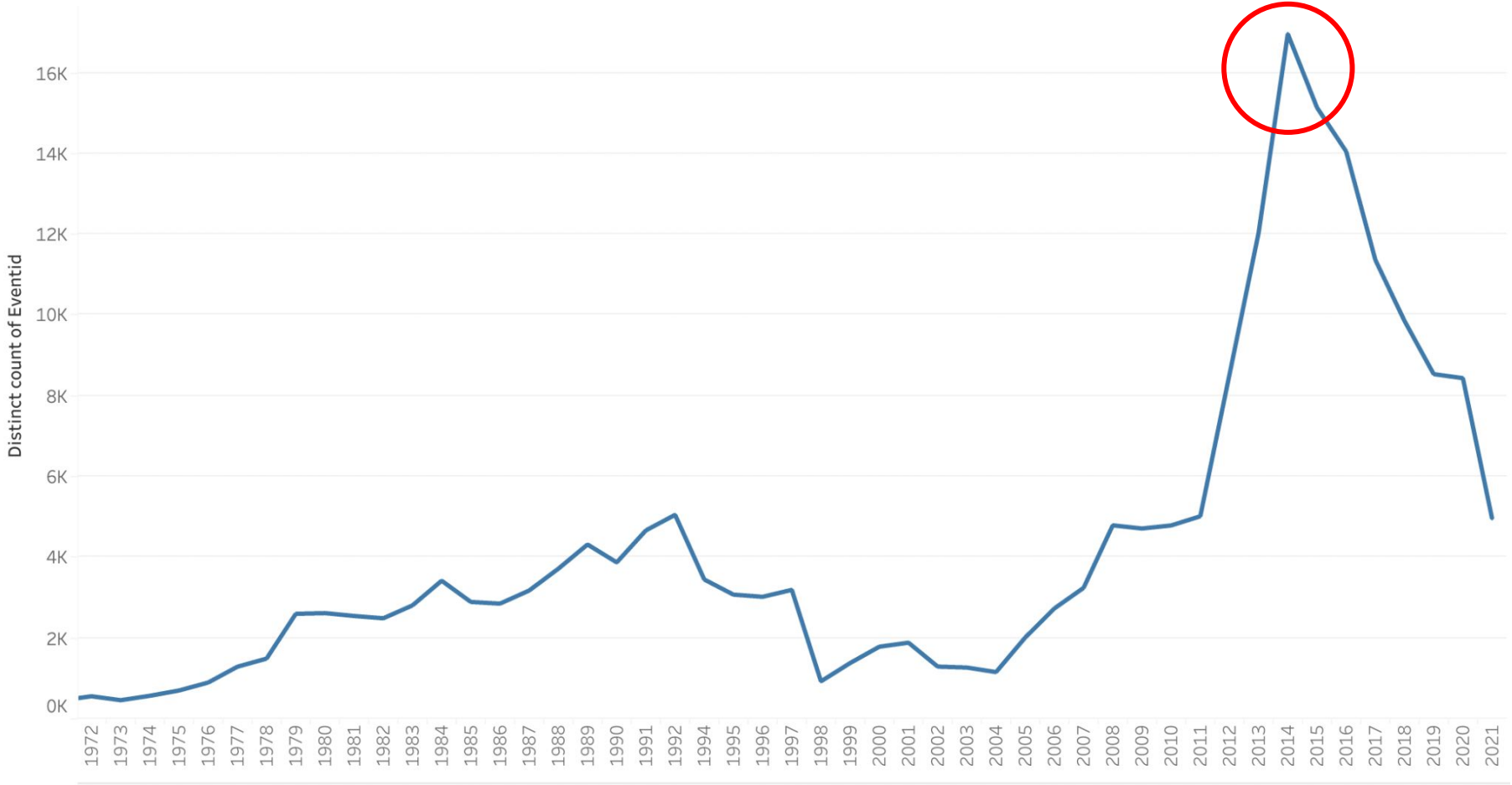
Number Global Terrorist Attacks by Data



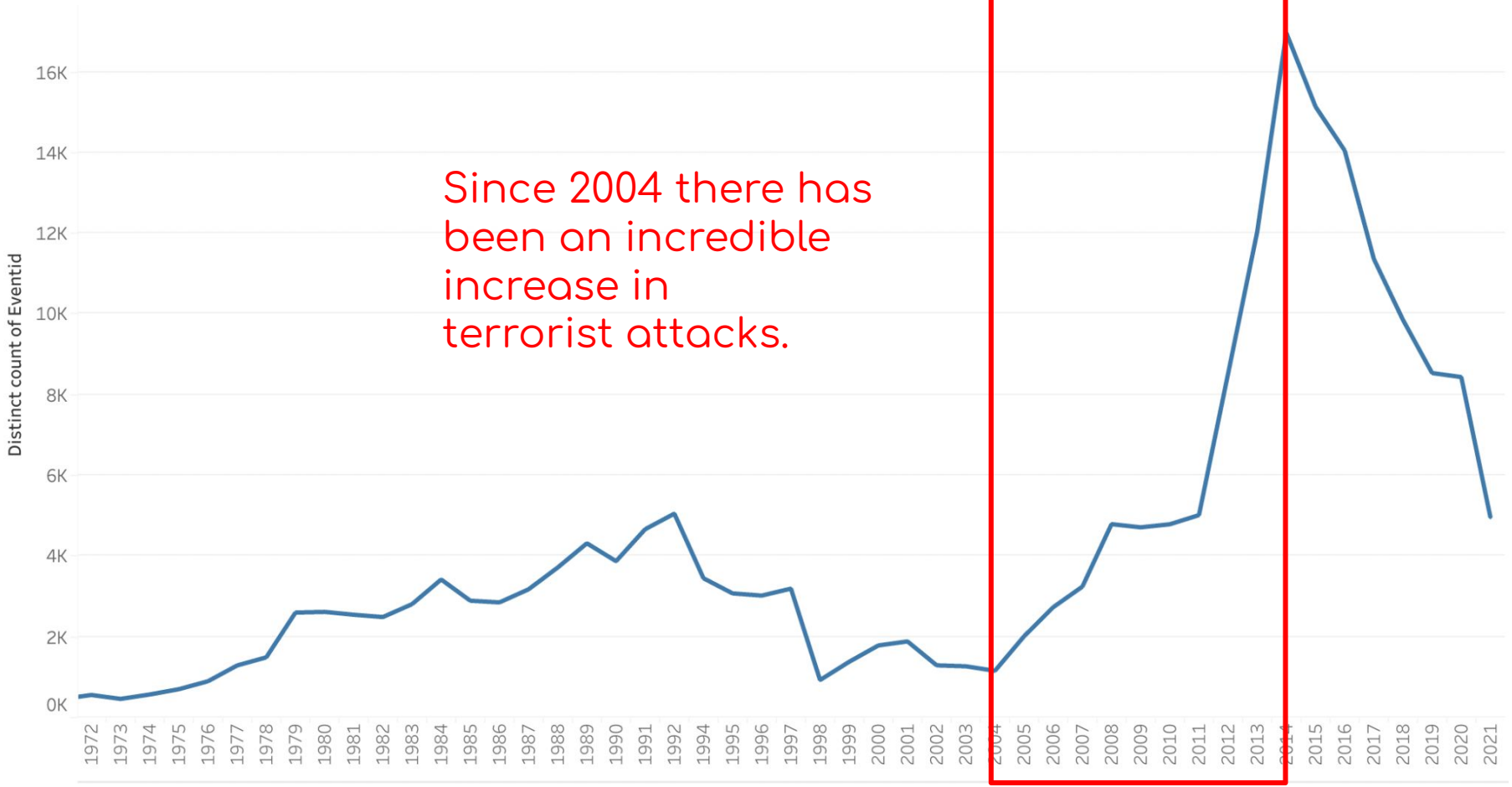
First strategy:

I search for a peak in the data and try to understand what happens near the peak (before, during, and after).

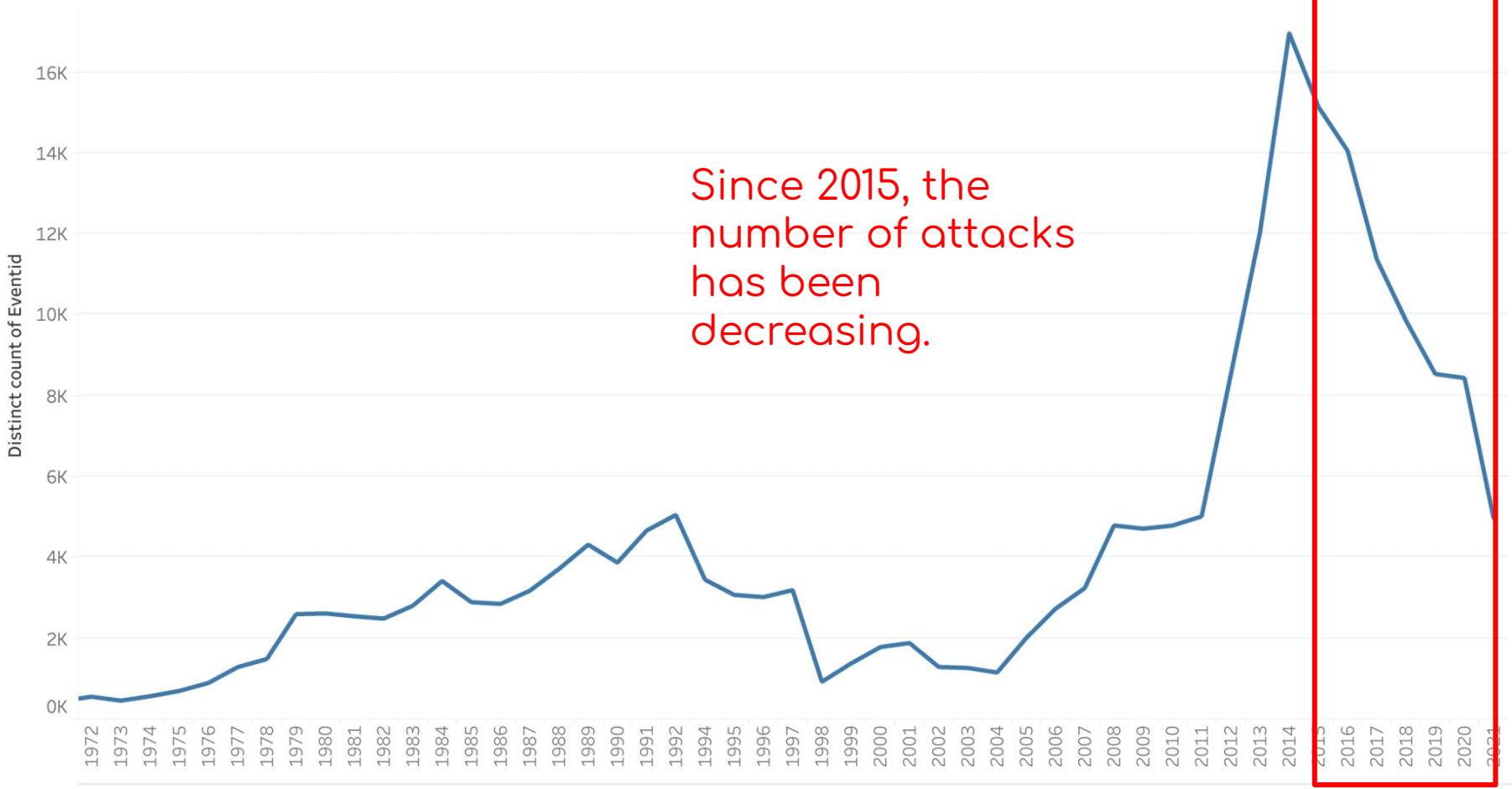
Number Global Terrorist Attacks by Data



Number Global Terrorist Attacks by Data



Number Global Terroristic Attacks by Data



Search for what happened in the years around the peak.

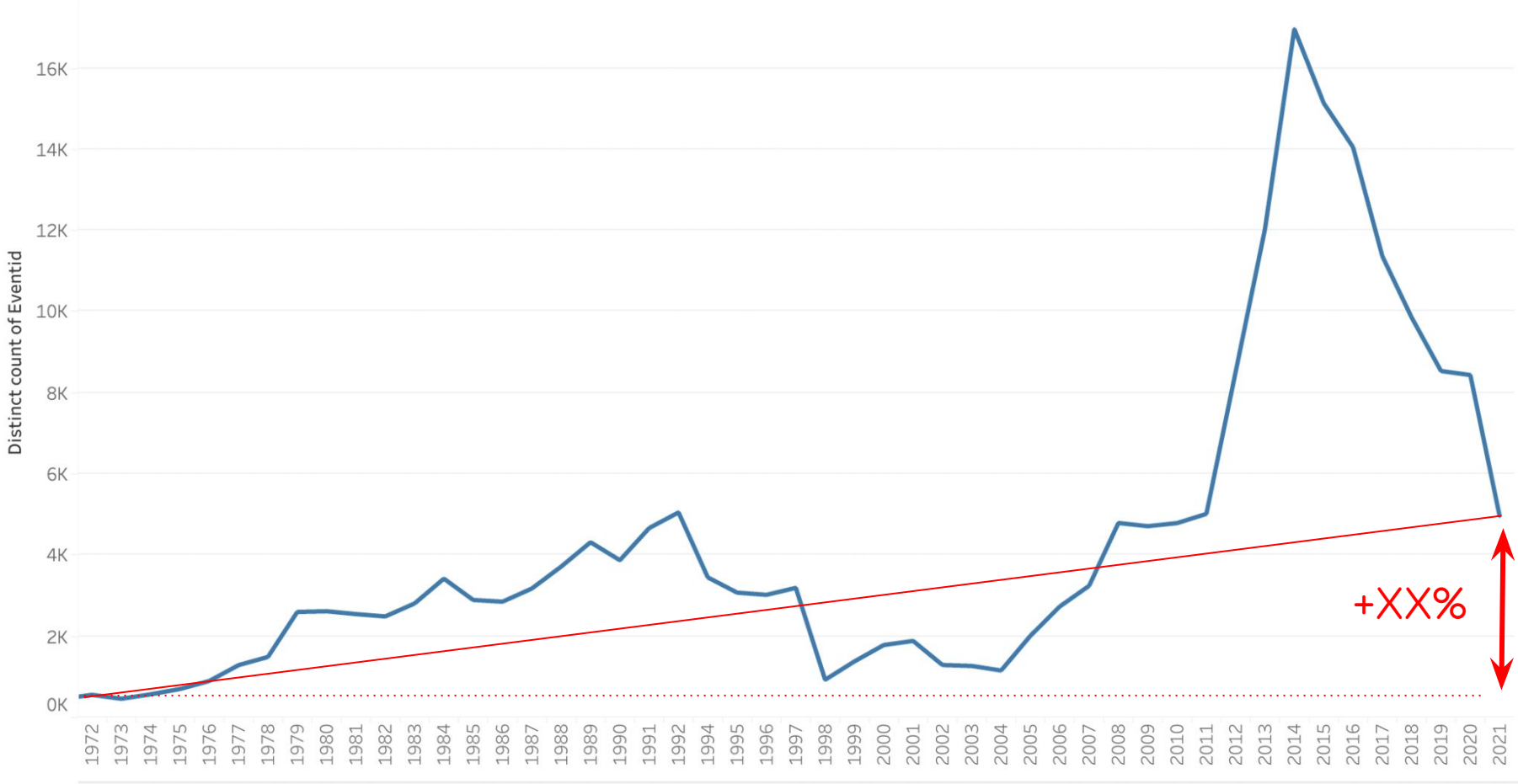
2003 - After the US invasion of Iraq in 2003, the Iraqi state collapsed, creating a power vacuum.
2015 - ISIS Territory Loss

Delve deeper into these events, looking for other data or other sources.

Second strategy:
Compare the years at the extremes and calculate the percentage increase.

Incr. perc. = (end year value - start year value) / start year value * 100

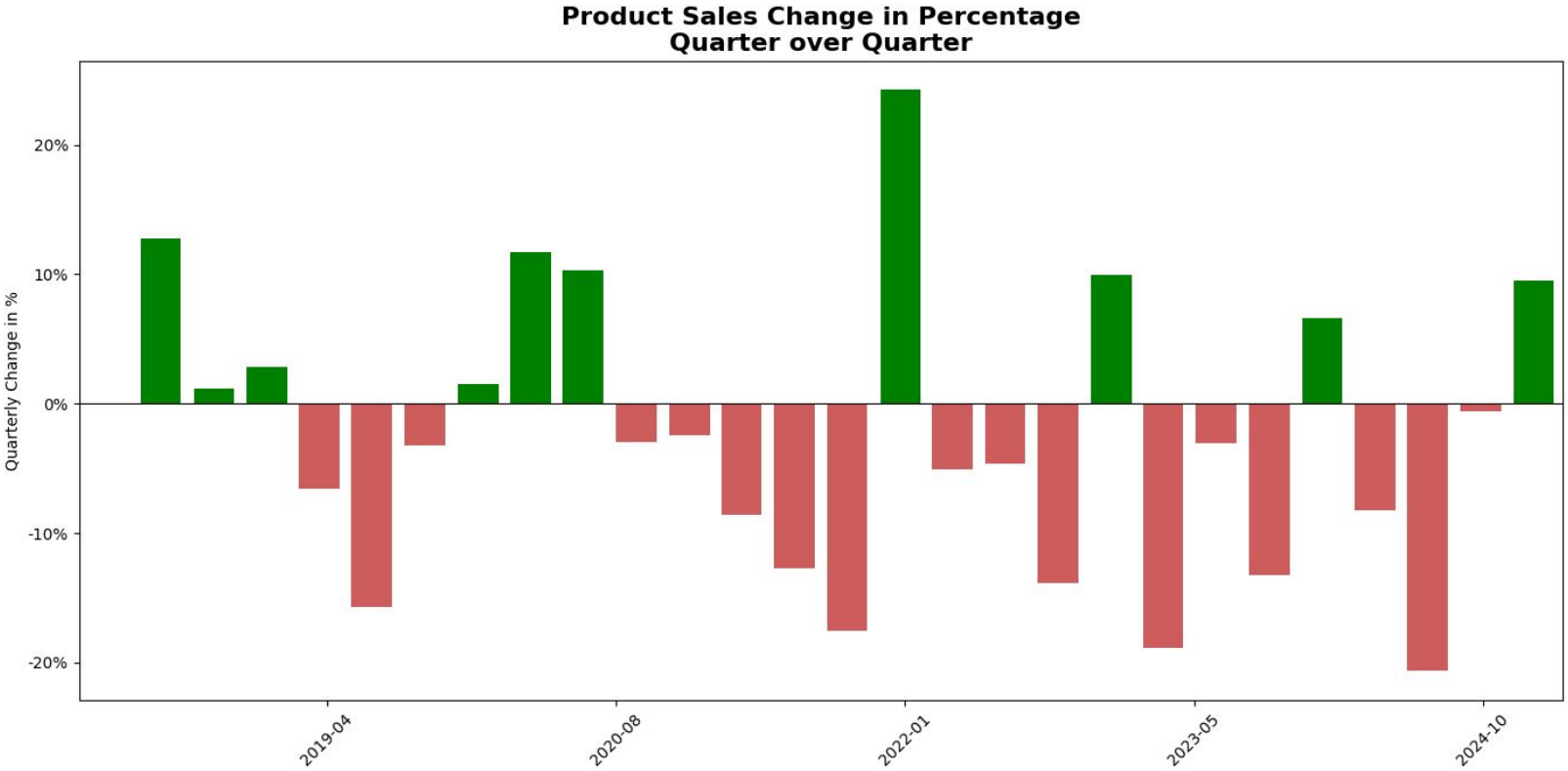
Number Global Terrorist Attacks by Data



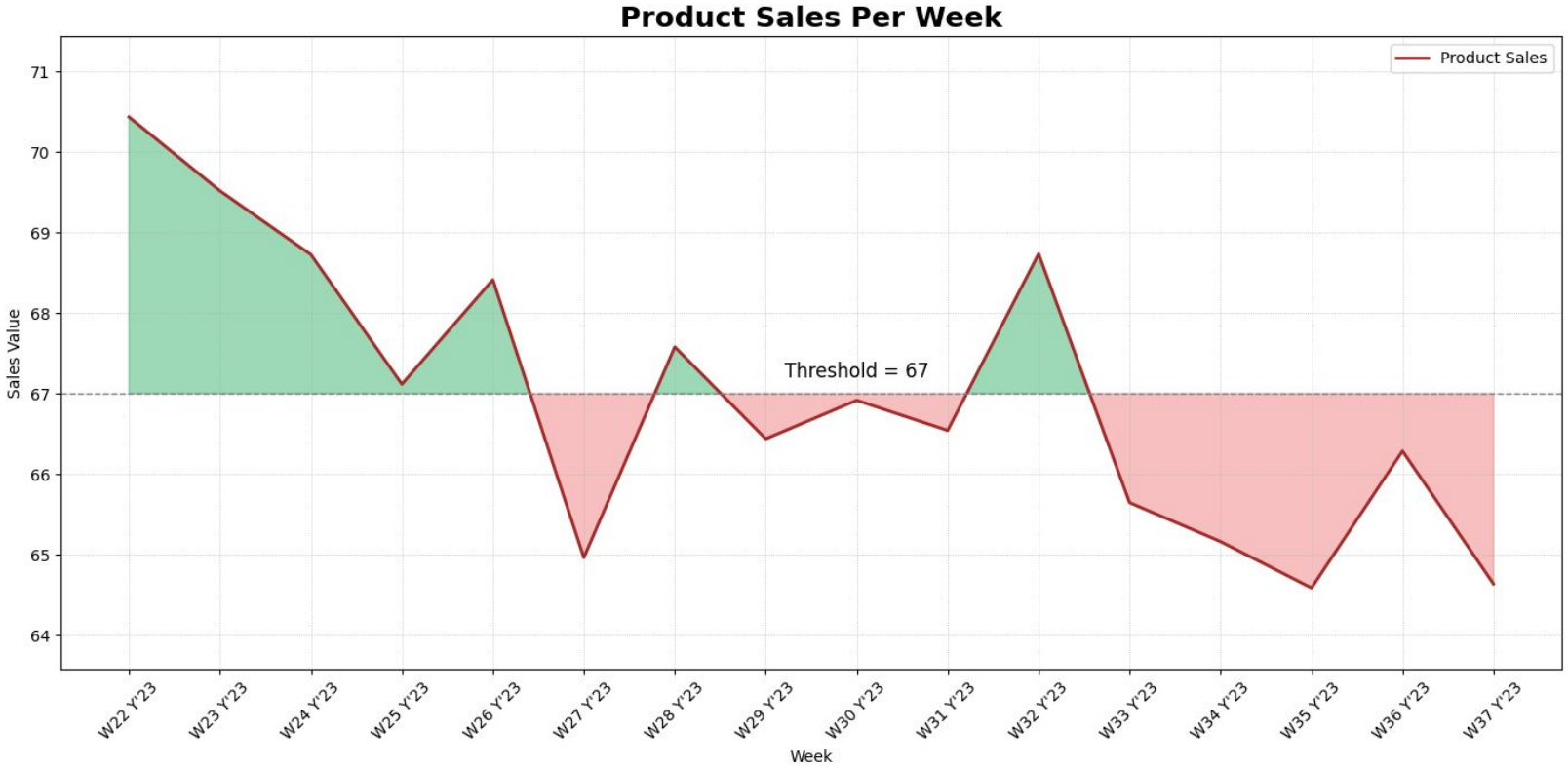
Delve deeper into the topic to understand what has determined the change over the years (other sources, other data).

Third strategy:
Comparison with other values

Compare with a baseline



Compare with a threshold



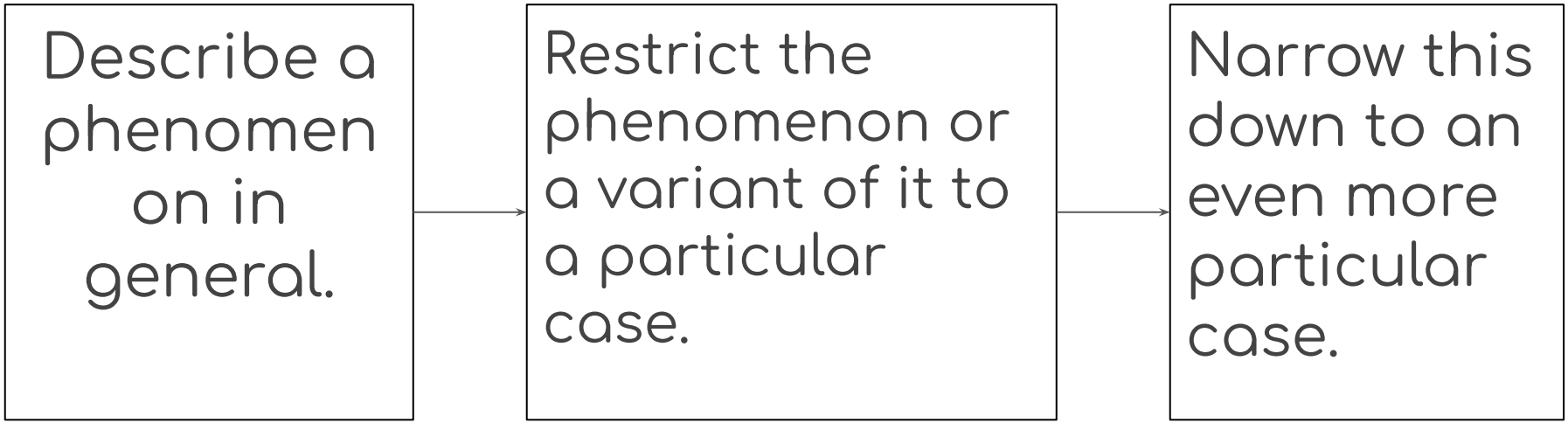
The objective is to understand if something happens in your data.

Remember:
No problem means no story.

Zoom

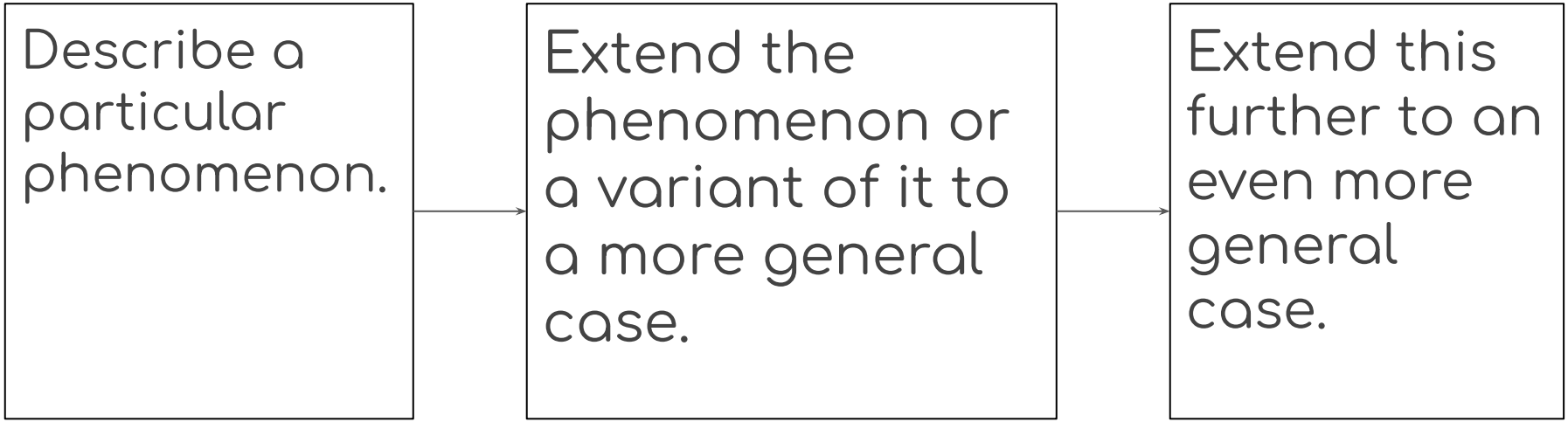
Zoom In: From overview to detail

Start from a general situation and moves towards a particular case.

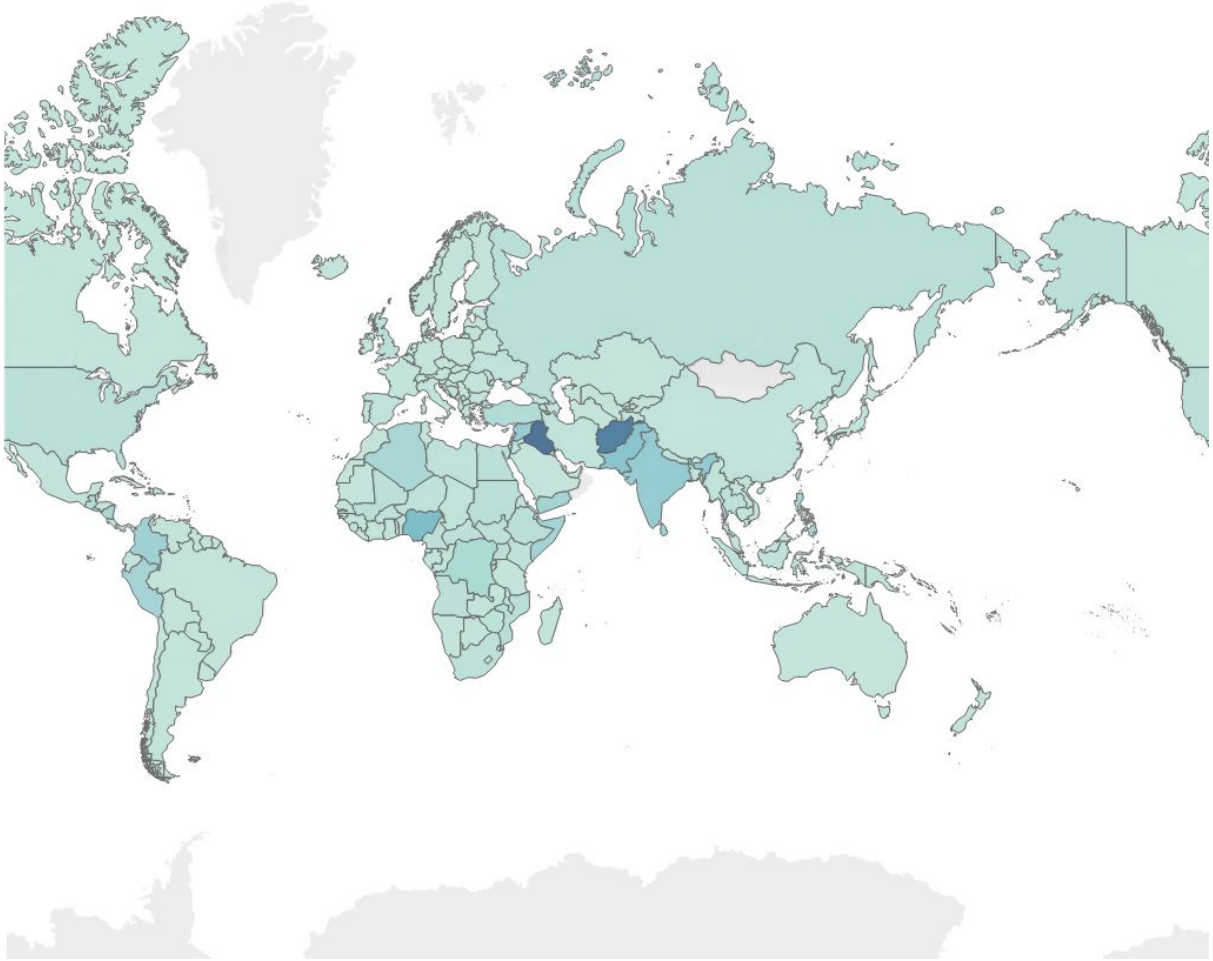


Zoom Out: From detail to overview

First, a particular part of the available information is analyzed and then the view is expanded to a more general perspective.



Number of deaths caused by terrorist attacks



First Strategy:

Focus on a specific region or country and then move to a broader level (e.g. continents).

Number of deaths caused by terrorist attacks

Describe the situation in Italy

Number of deaths caused by terrorist attacks

Extend my description to the whole of Europe: Show aggregated data at the continental level.

I can extend my description to the whole world.
In this case, I only show a number without a map, which corresponds to the total number of deaths.

The vice versa also applies (zoom out). I start from the world, then move on to a continent, and finally, to a country.

I can use different (but similar) data for different zoom levels.
For example, I take the Italian data from ISTAT and the European ones from Eurostat.

Second Strategy:

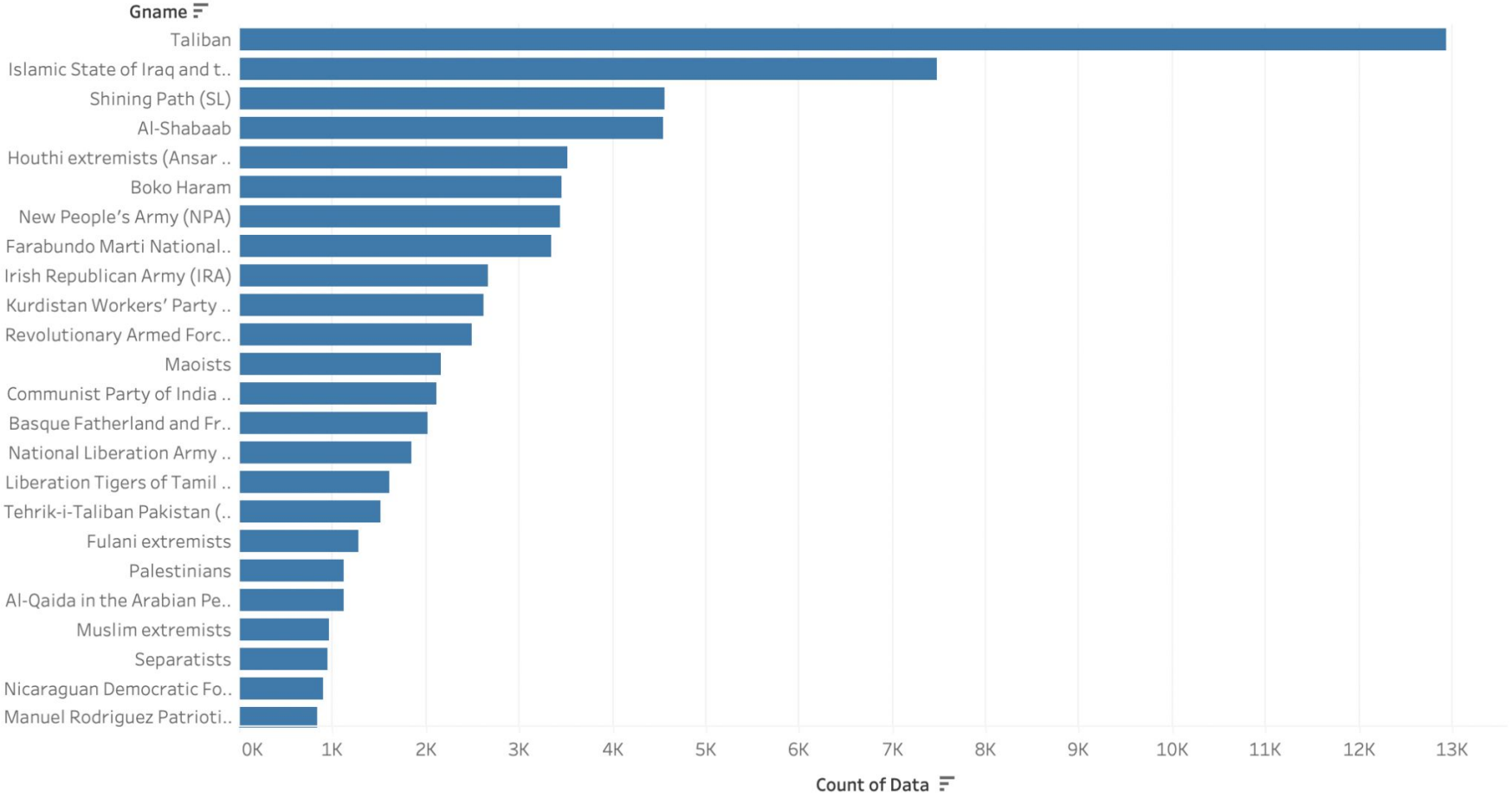
I focus on a specific region or country, calculate a specific metric for it (e.g. average value) and compare it to the same metric at a broader level.

For example, I compare the average number of terrorist attacks in Italy per year with the average value in Europe and in the world.

Comparison between multiple categories

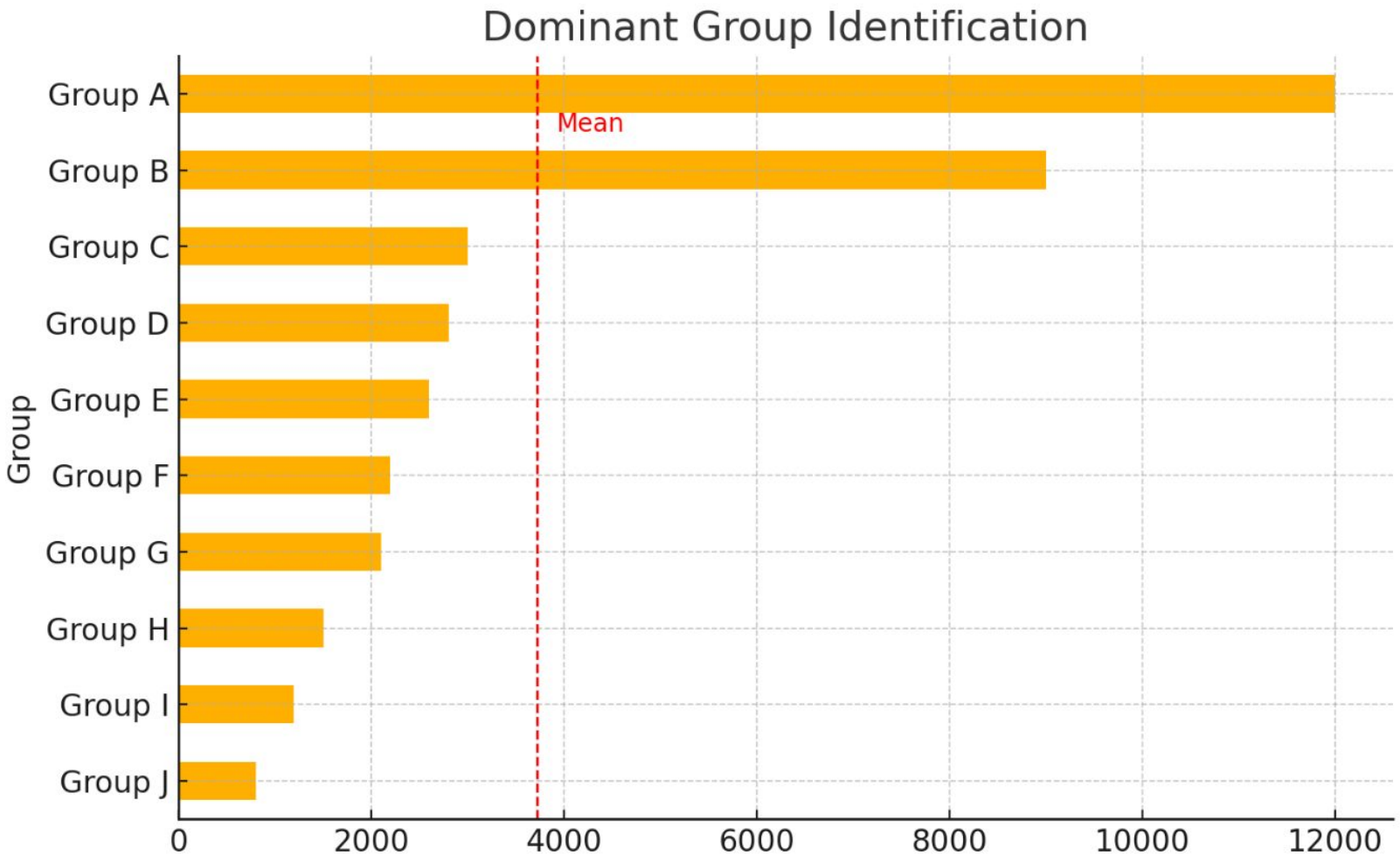
I have a static comparison among categories in a specific time and space.

NAttackxGroup



First Strategy:
Dominant Group Identification

Look for categories with significantly higher values than others.


Dominant Group Identification
Group A
Group B
Group C
Group D
Group E
Group F
Group G
Group H
Group I
Group J
Group
Mean
0
2000
4000
6000
8000
10000
12000

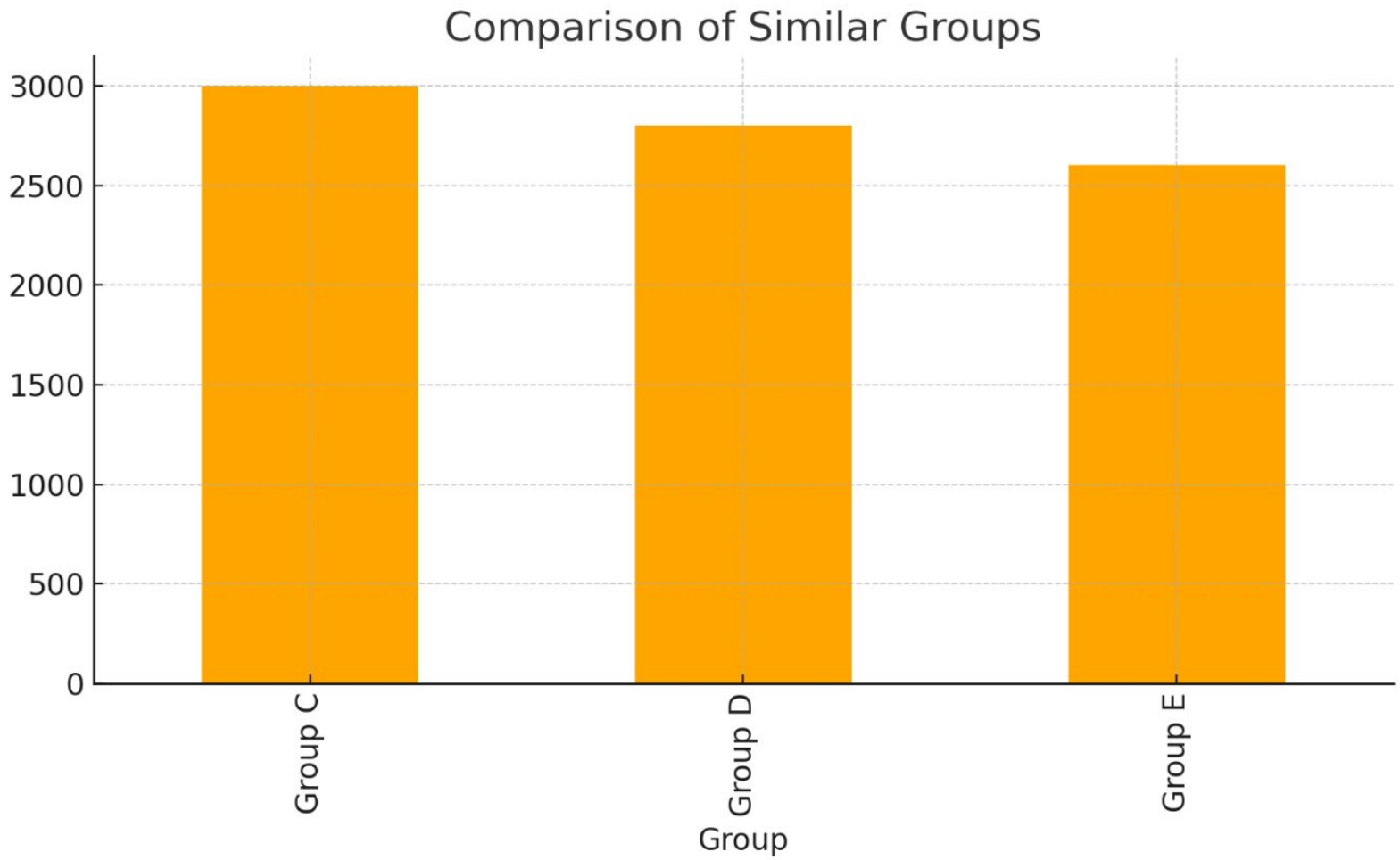
Different possible problems:

- Dominant groups should not be dominant but from the data, they are dominant (e.g., I discover that group A is incredibly the best group)

- Even if one group is truly dominant in numbers, it is essential to ask questions about context, evolution over time, impact, and distribution.

Second Strategy:
Comparison only of Similar Categories

Compare groups with similar backgrounds, goals, or regions to highlight differences in activity.


Comparison of Similar Groups
3000
2500
2000
1500
1000
500
0
Group C
Group D
Group E
Group

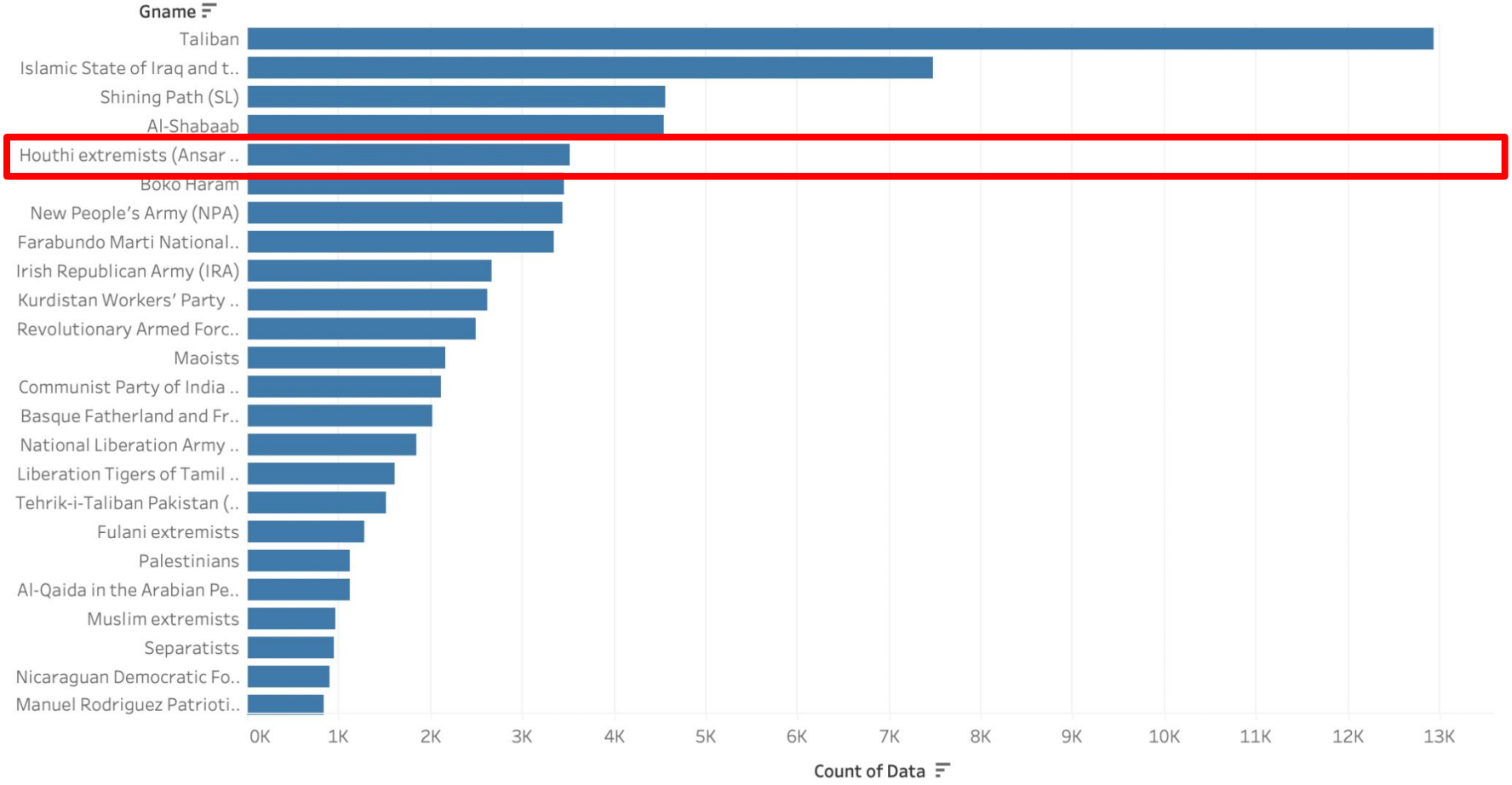
Possible problems:

- Similar groups are very different

Third Strategy:

Focus on a category and try to understand if that position is anomalous or correct

NAttackxGroup



Fourth Strategy:

Focus on two categories and compare their evolution over time.

I can also try to understand if there are any correlations

Cause and Effect
Changes in one factor affect changes in another factor

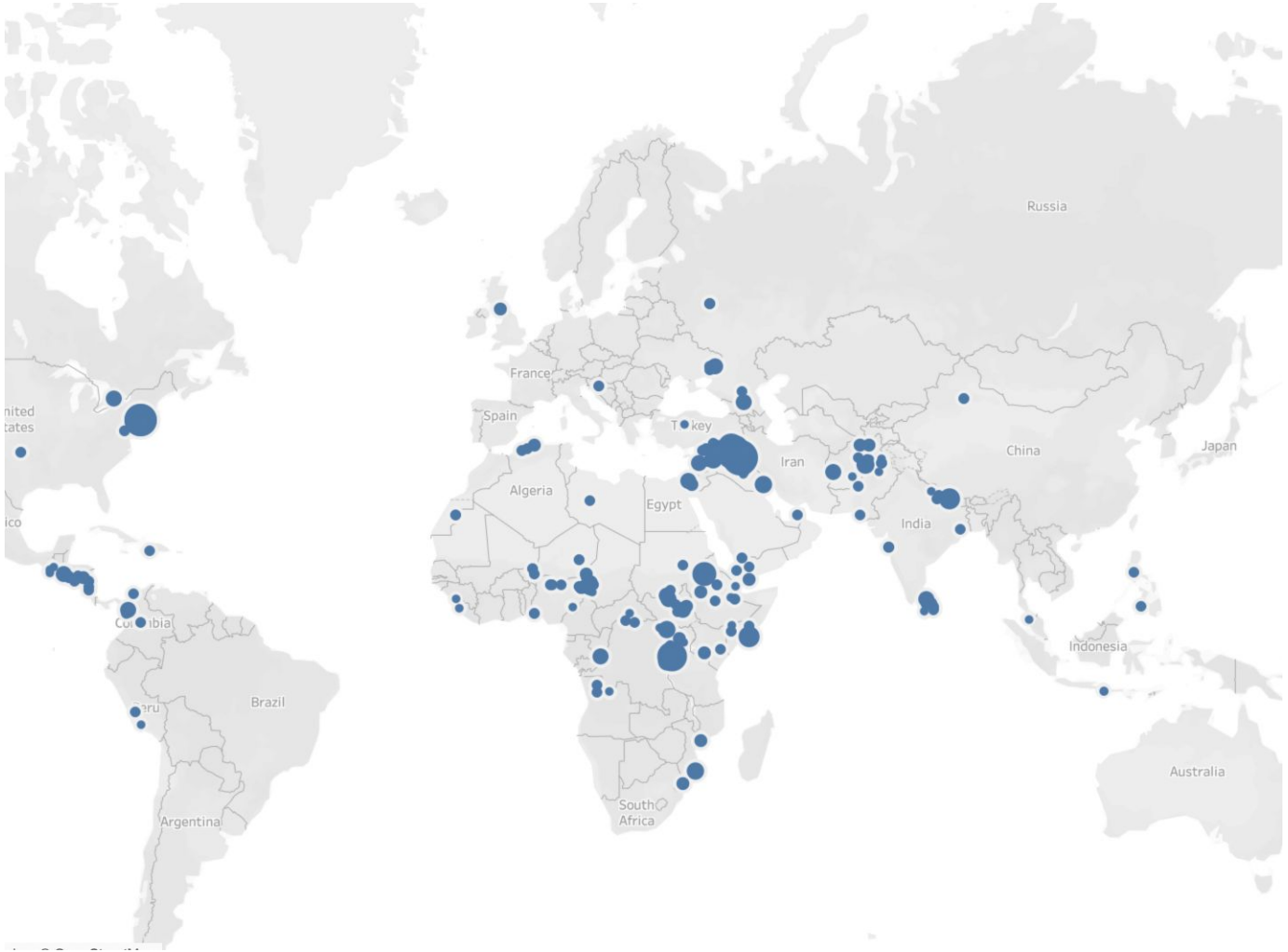
Common Cause
An external factor causes changes in both factors

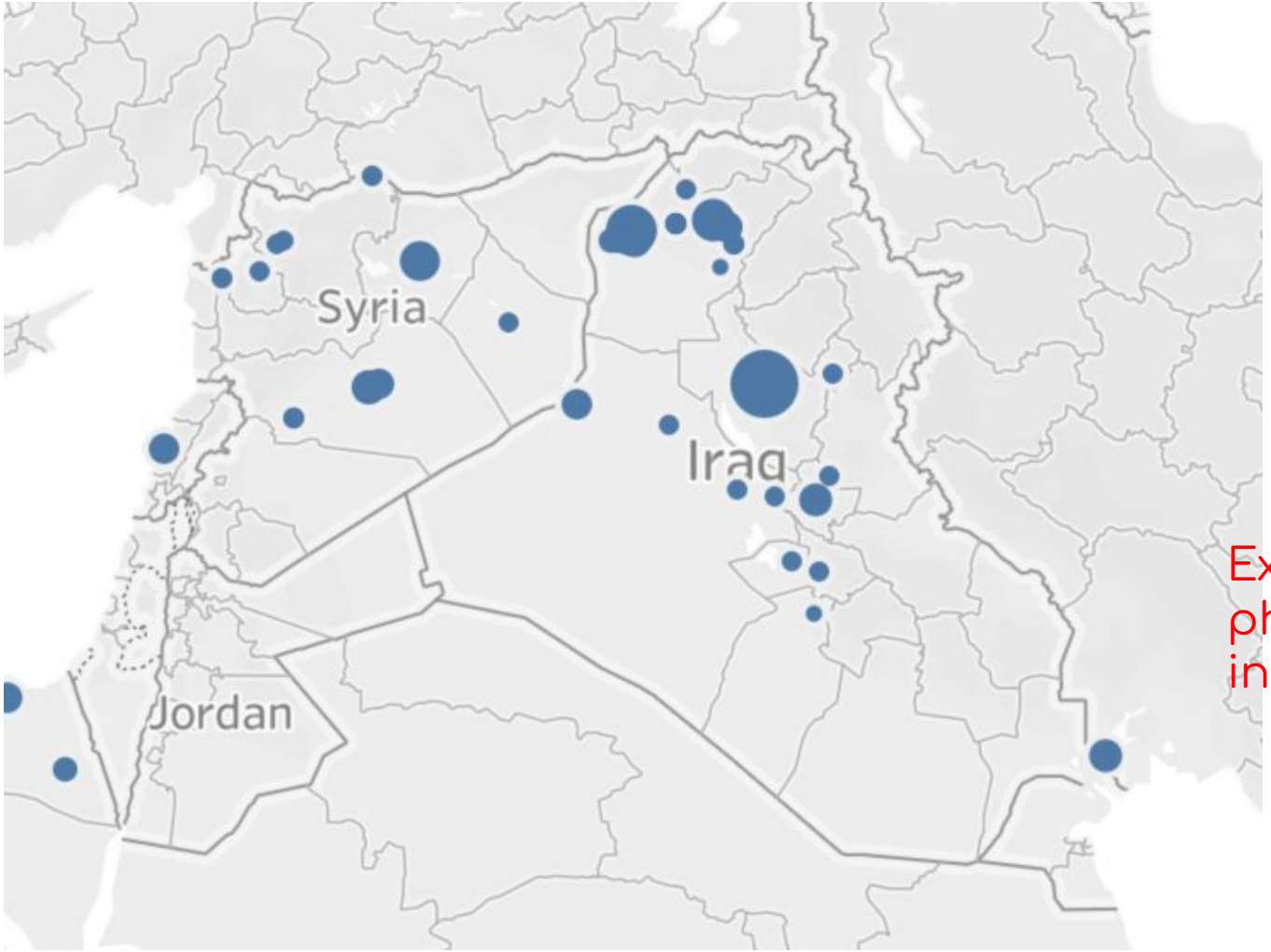
Accidental Relationship
There is a relationship between the two factors even though there is no cause-and-effect relationship

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

Spatial Analysis

Compare the same category or phenomenon in two or more different places of the same size (e.g. in two cities, in two countries, etc.).


Russia
France
Spain
Turkey
Iran
China
Japan
Algeria
Egypt
India
Indonesia
Brazil
Argentina
South Africa
Australia



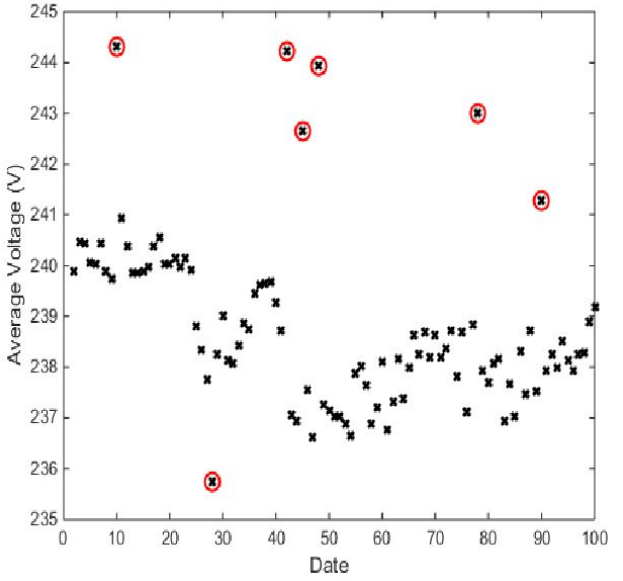
Explain how the phenomenon varies in Syria and Iraq.

I can also use a bar or column chart to compare different locations.
In this case, we fall back into the analysis that makes the comparison between categories.

Outliers

Points Anomalies*

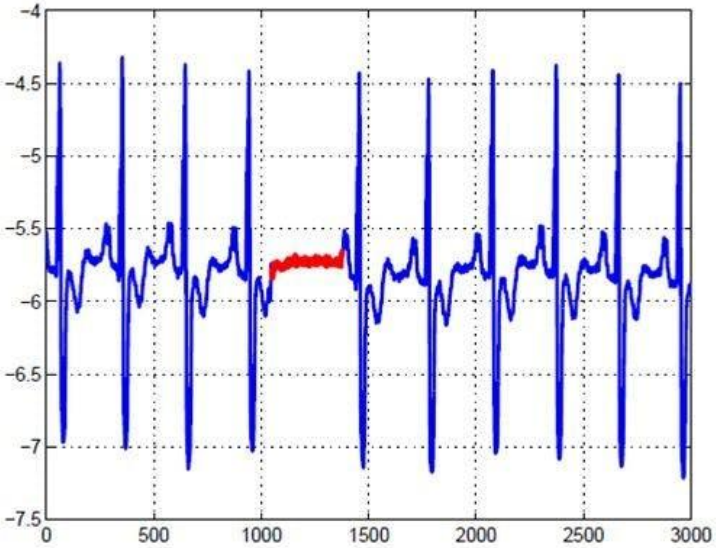
A point anomaly is when a single piece of data deviates from the expected pattern, range, or norm. In other words, the data point is unexpected.



*3 Types of Anomalies in Anomaly Detection
https://hackernoon.com/3-types-of-anomalies-in-anomaly-detection

Collective Anomalies*

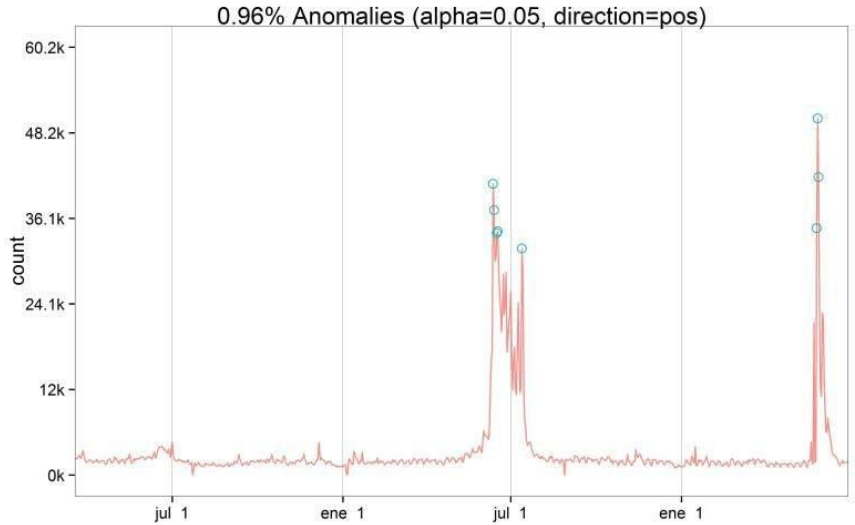
A collective anomaly occurs when individual data points examined in isolation appear normal. When a group of these data points is examined, however, unexpected patterns, behaviors, or results become apparent.



*3 Types of Anomalies in Anomaly Detection
https://hackernoon.com/3-types-of-anomalies-in-anomaly-detection

Contextual Anomalies*

Unexpected results that come from what seems like normal activity.



*3 Types of Anomalies in Anomaly Detection
https://hackernoon.com/3-types-of-anomalies-in-anomaly-detection