

## 1. Explore and find stories in your data

- Here is the data, what is it trying to tell us; in particular, which question does it want us to ask? What seems to be going on?
- Use comparators:
  - Socio-economic groups (Farmers/pastoralists)
  - Geographical area (Province A/Province B)
  - Setting (Urban/rural, Conflict/no conflict, etc.)
  - Sectors (WASH/LFS, Shelter/health, etc.)
  - Affected groups (IDPS/Affected Residents, etc.)
  - Vulnerable groups (pregnant woman, elders, etc.)
  - Sex and age (Age intervals, Male/female)
  - Time period (Before/After)

## 2. Identify your message

Ask yourself:

- What **data** is important to show? Which dimensions and metrics need to be shown at the same time?
- What do I want to **emphasize** in the data? For example, do I want to compare different values, show relationships, or present changes over time? What story am I trying to tell?
- What **options** do I have for displaying this data? Which option is most **effective** at communicating my message? Which chart or visualization emphasizes what's important in the most direct, intuitive and readable way?

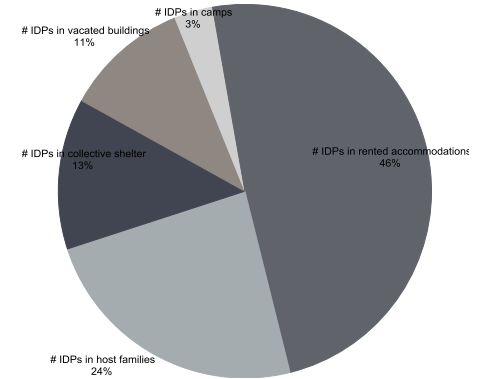
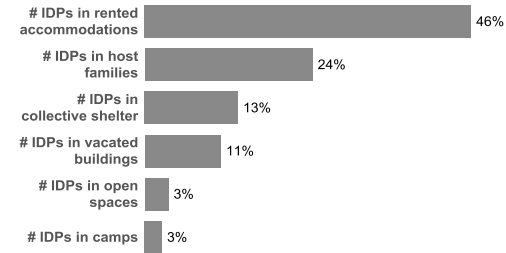
## 3. Prepare your data

Chose and organise your dimensions and metrics:

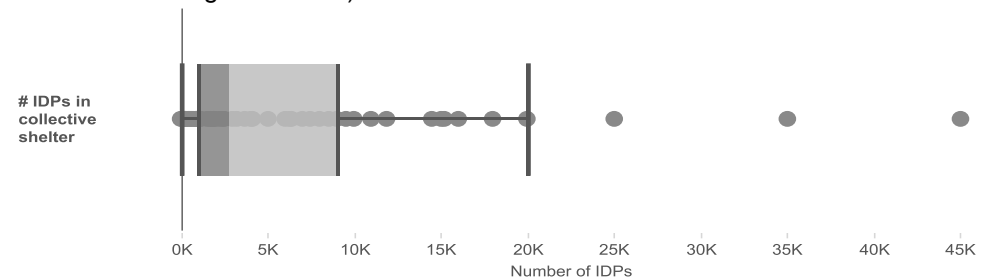
Governorate	# IDPs in camps	# IDPs in collective shelter	# IDPs in host families	# IDPs in open spaces	# IDPs in rented houses	# IDPs in vacated buildings
Al-Hasakeh	300	61,113	65,342	2,266	193,635	3,973
Aleppo	27,865	246,163	319,630	88,568	208,968	233,779
Ar-Raqqa	0	22,775	119,400	6,500	147,000	11,915
Deir-ez-Zor	300	81,659	289,869	4,670	70,659	70,268
Hama	0	35,700	74,000	3,250	59,550	60,800
Homs	0	26,000	38,267	17,056	51,529	29,165
Idleb	77,216	61,614	106,410	24,535	405,137	49,445
Lattakia	2,500	35,500	9,300	600	856,000	4,100

## 4. Compare within

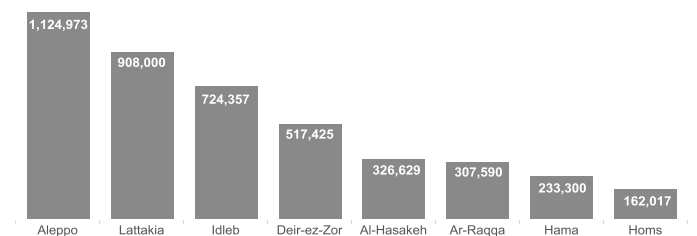
- **Composition or part to whole:** Measures of individual categorical subdivisions as ratios to the whole. Values represent parts of a whole, for example the percentage of IDPs per type of setting. Use when you want to show how various parts comprise the whole or how values and sizes compare to one another and the whole



- **Distribution:** Counts of something per categorical subdivisions (intervals) along a quantitative scale, from lowest to highest, for example, how IDPs values frequently appear (are distributed) for a particular setting. Use when you want to show how values relate to one another as a matter of proximity (i.e. their distribution through the entire range of values)



- **Ranking:** Categorical subdivisions of a measure ordered by size (either descending or ascending). Use when you want to show how values (associated with categorical items) are ranked according to size

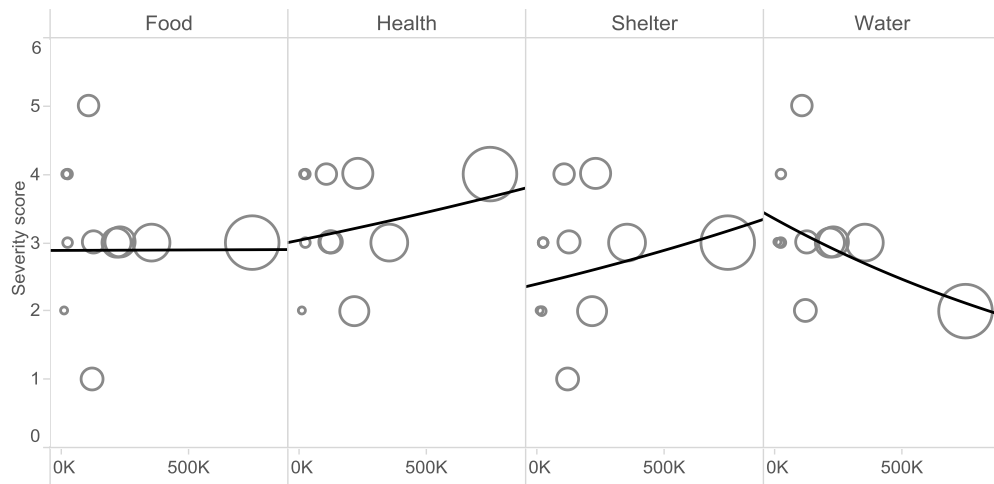


## 5. Compare between

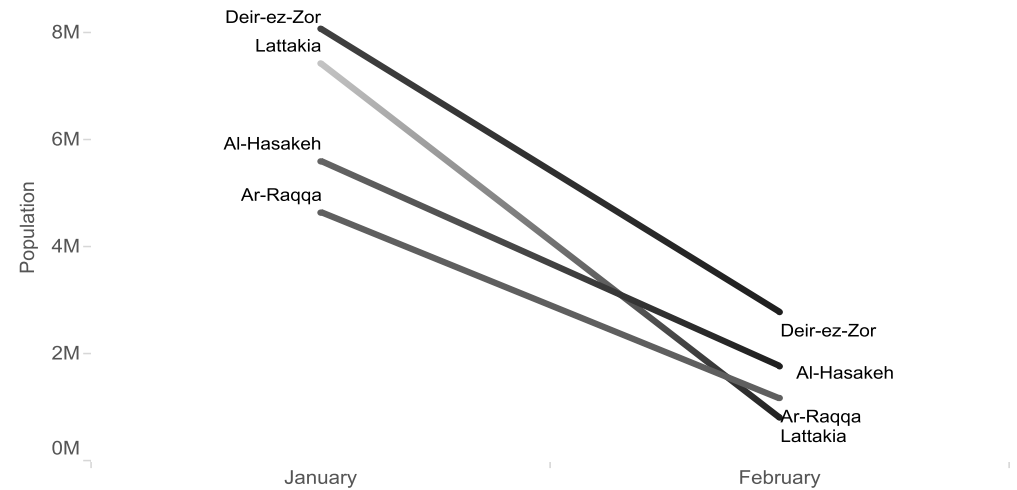
- **Nominal:** A simple comparison of the categorical subdivisions of one or more measures in no particular order, for example, number of IDPs per region.

Governorate	# IDPs in rented accommodations	# IDPs in host families	# IDPs in collective shelter	# IDPs in vacated buildings	# IDPs in open spaces	# IDPs in camps
Al-Hasakeh	193,635	65,342	61,113	3,973	2,266	300
Aleppo	208,968	319,630	246,163	233,779	88,568	27,865
Ar-Raqqa	147,000	119,400	22,775	11,915	6,500	0
Deir-ez-Zor	70,659	289,869	81,659	70,268	4,670	300
Hama	59,550	74,000	35,700	60,800	3,250	0
Homs	51,529	38,267	26,000	29,165	17,056	0
Idleb	405,137	106,410	61,614	49,445	24,535	77,216
Lattakia	856,000	9,300	35,500	4,100	600	2,500

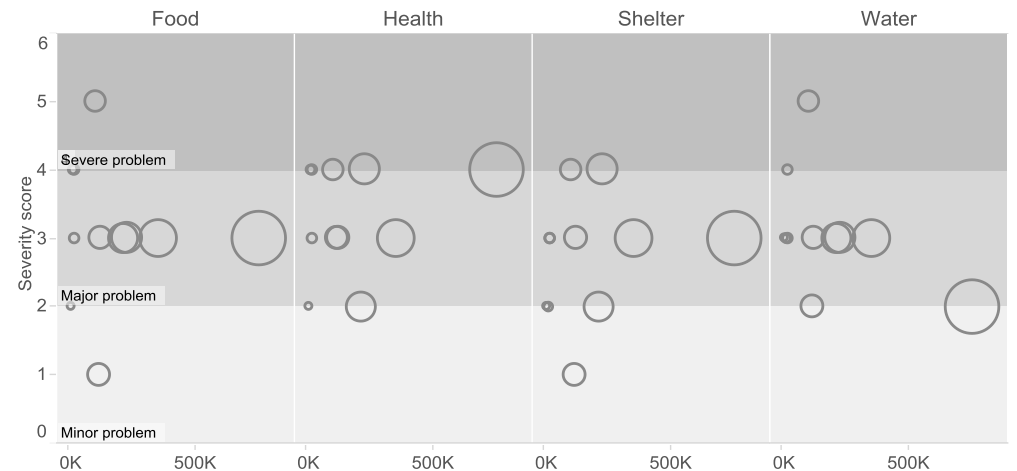
- **Relationships:** Comparisons of two paired sets of measures to determine if as one set goes up the other set goes either up or down in a corresponding manner, and if so, how strongly, for example, the number of people affected and the severity score per sector, to determine if there is a relationship between population density and humanitarian impact. Use when you want to establish (or show) relationship between 2 (or more) variables and how two sets of quantitative variables associated with a common set of entities behave in relation to one another.



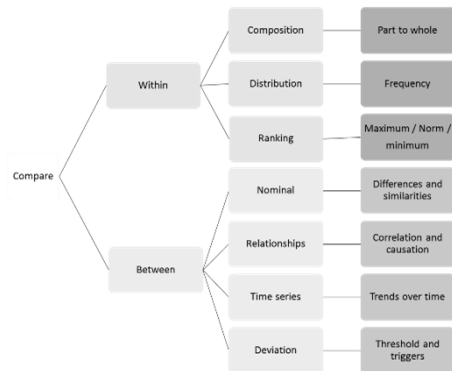
- **Time series:** Multiple instances of one or more measures taken at equidistant points in time. Values display how something changed through time (yearly, monthly, etc.). Use when you want to understand the trend over time of some variable(s) or how values change through time.



- **Deviation:** Categorical subdivisions of a measure compared to a reference measure, expressed as the differences between them (reference lines). The difference between two sets of values (e.g. targeted beneficiaries vs. reached beneficiaries). Use when you want to see which value deviate or differ from the norm or a reference point.



## 6. Go beyond comparisons



Comparison is at the core of the analytical process. It is used to identify similarities and differences within and between your data, forming the basis of pattern recognition.

Patterns are a repeating design or sequence and demonstrate recurring themes or categories, appearing in a predictable manner. They are regularities, variations or exceptions which stand out above the typical noise evident in nature or in raw data.

What makes some details in the material

analysts are studying more worthy of attention than others? Here are three principles for identifying which details in the material are more worthy of attention than others:

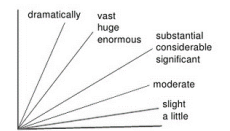
- **Look for patterns of repetition or resemblance:** Which details seem to repeat? What goes with what? Why? What does the detail mean? What else might it mean? How do the details fit together? What do they have in common? What does this pattern of details mean? What else might this same pattern of details mean? How else could it be explained? In virtually all subjects, repetition is a sign of emphasis. Once apparent similarities have been located, analysts can refine their thinking by pursuing significant distinctions among the similar things (looking at differences within the similarity or similarities despite the difference).
- **Look for binary oppositions:** What is opposed to what? Sometimes patterns of repetition are significant because they are part of a contrast around which the subject matter is structured. Detecting repetition will lead analysts to discover opposites which are central to locating issues and concerns.
- **Look for anomalies, outliers, and things that don't fit:** What details don't seem to fit? How might they be connected with other details to form a different pattern? An anomaly is literally something that cannot be named, a deviation from the normal order. Anomalies help us revise stereotypical assumptions, and noticing them often leads to new and better ideas.

Our perception of patterns in quantitative or qualitative data is fundamental to the sense-making process. They lead to insights that might never occur if the data were examined in any other way. For example, certain health conditions may cluster in particular geographical areas or people from a particular group may apply similar coping mechanisms. These patterns may not be specifically what was looked for or anticipated, but they may be important in themselves and deserve increased attention, or they may shed light on new areas of interest or specific elements of the data.

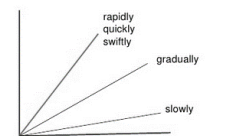
Meaningful patterns in quantitative data fall into three general categories:

- 1) **Large-scale patterns (a.k.a. trends).** These are patterns that reveal what is going on in general, that is, as a whole (i.e. cereal prices have trended downwards over the course of the year), such as repetitions, cycles, feedback systems, exponential growth, diminishing returns, etc. A trend is a general direction in which something is developing or changing and refers to the changes or movements in facts and figures over a period of time. The degree and the speed of change also need to be considered. Basic trends can be categorised as upward (↗) or downward (↘) movements, stability (→ no change or movement) and change in direction (↙ or ↘, ∩ or U). Recognising trends is often a matter of looking at the data at the appropriate level of scale. If looked at too closely, data is simply a series of peaks and troughs lacking any real sense of direction. However, when one zooms out and views a greater range of data, the overall shape of the data becomes clearer.

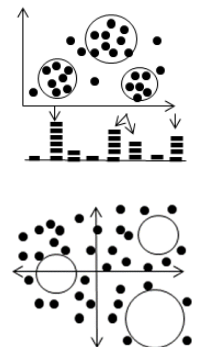
Describing the degree of change



Describing the speed of change



- 2) **Small-scale patterns:** These are patterns that reveal what is going on in specific subsets of data, such as clusters and gaps. Knowing how many concentrations are present is just as important as knowing where they fall. A cluster might represent something as simple as the distinguishing characteristics of different livelihood groups or geographical areas (urban vs. rural). On the other hand, gaps in the data represent the absence of any observable data, which can be just as informative as actual observations.



- 3) **Exceptions or outliers:** These are values that appear outside of what is normal, standard, expected, regular or acceptable. They can be described as data elements that deviate from other observations by so much that they arouse suspicion of being produced by a mechanism different than that which generated the other observations. For example, out of all visited areas, only one particular village showed a complete lack of food product in the local market. Every abnormal value can and ought to be explained. Three possible reasons can create outliers: Errors (caused by inaccurate data entry, measurement or bias), extraordinary events (storm, earthquake, etc.) or extraordinary entities (richest person in the village, etc.).

