The many dimensions of data

Author: **Inês Ventura**

Do you feel safe walking around after dark? Does the size of the city affect how you feel? How do these feelings compare between men and women? For data analysts, these questions come with unwieldy amounts of data. Luckily, Dr Gianmarco Alberti from the Department of Criminology (Faculty of Social Wellbeing, University of Malta) has authored a free software that visually portrays data patterns in a practical way.

So how does the software work? Going back to our safety in the dark example, the data is plugged into the software. The programme then explores how the feeling of safety relates to the size of the city. In this example, we'll split the variable 'feeling safe after dark' by gender and see if it's influenced by the number of people living in the city ('town size', represented at the top of the first image). The table below is small yet highly complex (see figure 1), finding any obvious pattern of association between categories is hardly an easy task.

Plugging this data into Alberti's software, named 'CAinterprTools', provides a body of facilities that allow users to get the most of those results (see figure 2). With this image, it's easier to understand that the bigger the town (right side along the horizontal axis), the less safe interviewees felt. Both male (identified with a M) and female (identified with a F) have the same perception of their safety in a smaller town (points in the left-hand side of the chart). However, as the size of the town increases, females start to feel less secure than men (points to the right-hand side of the chart). It is clear that this approach can provide insights into data structure and help reveal hidden patterns.

The beauty of Alberti's program is that it increases data analysis efficiency, not only to researchers (e.g criminologists, political scientists, or biologists) but also to citizens and several entities that work with data (e.g. banks, customers, or companies). The fact that this software was developed under the free R statistical programming language means it is a free resource which can allow for an easier way to interpret data — especially for those who aren't math oriented! Let's hope this leads to safer streets for all!

The programme can be found on https://cran.r-project.org/package=CAinterprTools.

FIG. 1

FEEL SAFE AFTER DARK	TOWN SIZE -10,000	10,000-50,000	50,000-100,000	100,000-500,000	500,000-1,000,000	1,000,000+	GRAND TOTAL
FEMALE	14628	11792	6139	21334	6609	11426	71928
bit unsafe	2930	3152	1680	5898	2130	3594	19384
very unsafe	968	1171	640	3573	1553	2747	10652
fairly safe	6237	4988	2559	8093	2073	3501	27451
very safe	4277	2376	1206	3634	799	1485	13777
unknown	216	105	54	136	54	99	664
MALE	13041	10755	5099	19044	5480	9742	63161
bit unsafe	1030	1336	817	3677	1367	2504	10731
very unsafe	193	288	165	1529	552	850	3577
fairly safe	4889	4497	2186	8383	2234	3857	26046
very safe	6880	4589	1913	5414	1309	2501	22606
unknown	49	45	18	41	18	30	201
GRAND TOTAL	**27669**	**22547**	**11238**	**40378**	**12089**	**21168**	**135089**

Figure 1: presents how safe male and female participants feel after dark (column 1) compared to the population size of the city (row 1)

The data used in the table was taken from the International Crime Victim Survey - https://wp.unil.ch/icvs/

FIG. 2

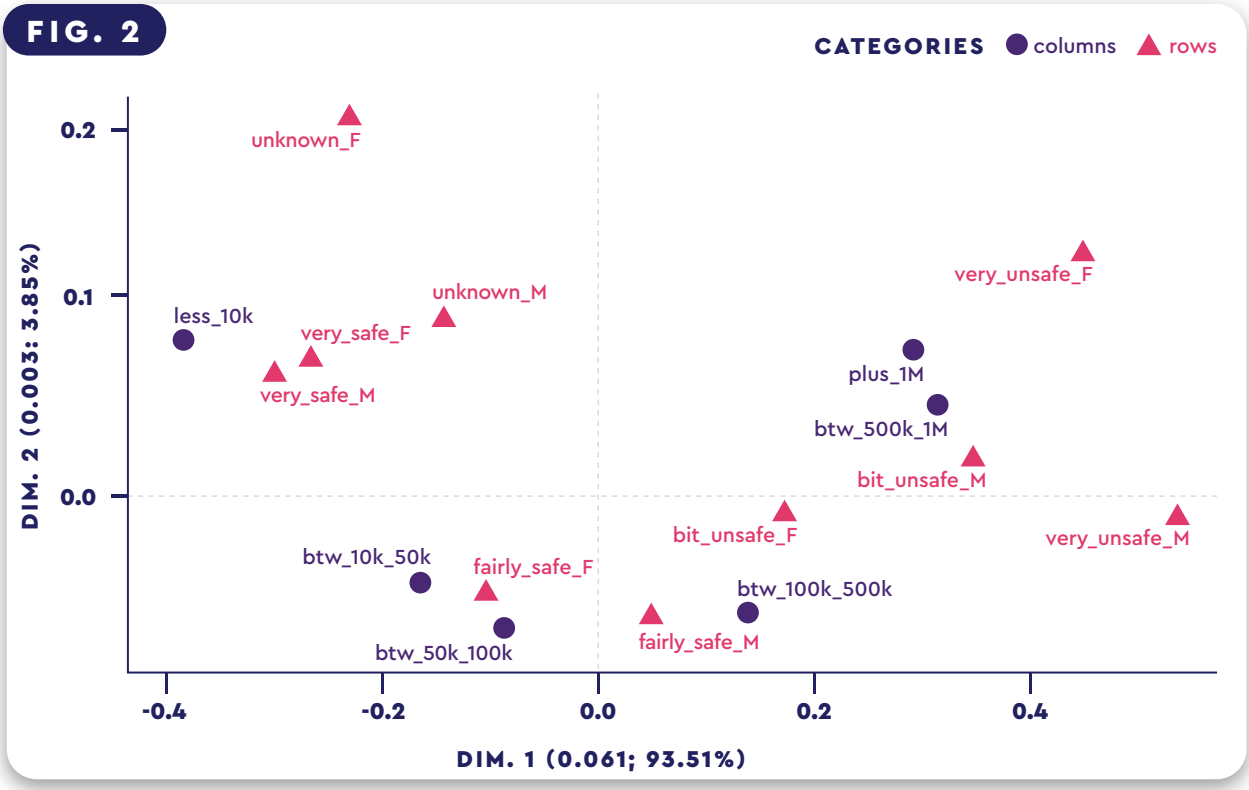

Figure 2: shows a visual representation of the data presented in Table 1. The size of the city is marked with black points, while how safe participants felt is symbolised by the red triangles.

This graph has been stylised for editorial purposes.