Plotting intersections

To the Editor: Recently, Lex and Gehlenborg[1] published an interesting column about set visualization full of valuable advice for scholars seeking an appropriate representation of intersecting data sets. However, we think that two slightly misleading statements deserve further explanation: "There are 2^n possible intersections for n sets" and "Euler diagrams represent intersecting sets as overlapping shapes."

Strictly speaking, an intersection of two (or more) sets is the set that contains elements belonging to both (or at least two) sets. Thus, the number of possible intersections given n sets is $< 2^n$. In fact, 2^n is the number of all possible intersections, plus the number of subsets formed by elements belonging to only one set (that is, disjoint elements), plus the set formed by elements which belong to neither of the above subsets.

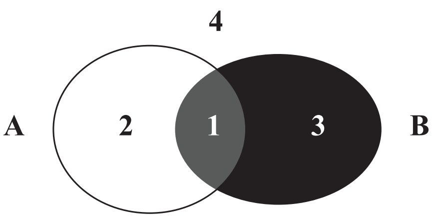

Figure 1 | Euler diagram displaying the intersection of two genes, A and B.

Consider an Euler diagram (**Fig. 1**) that illustrates all possible situations arising from mutations in two genes, A and B. Area 1 is properly defined as the intersection (mutations occurred in both A and B genes—co-mutated genes). Area 2 and 3 correspond, respectively, to cases in which only A or only B was mutated. Finally, region 4, outside the circles, represents the case in which no mutation occurred in either A or B. Thus four possible situations are represented by four corresponding regions of the plane ($2^n = 2^2$), including two populated by genes that are not co-mutated ($n = 2$) and one that defines the intersection ($2^n - n - 1$). Of course, the above considerations hold for less trivial analyses. For example, five genes generate 32 ($2^n = 2^5$) possible cases—five corresponding to mutations in one gene and 26 corresponding to intersections (two or more genes co-mutated); this describes **Figure 1b** of Lex and Gehlenborg's column[1].

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Giovanni Lentini[1] & Solomon Habtemariam[2]

[1]Dipartimento di Farmacia–Scienze del Farmaco, Università degli Studi di Bari 'Aldo Moro', Bari, Italy. [2]Pharmacognosy Research Laboratories, Medway School of Science, University of Greenwich, Kent, UK.
e-mail: giovanni.lentini@uniba.it

1. Lex, A. & Gehlenborg, N. *Nat. Methods* **11**, 779 (2014).

Lex and Gehlenborg reply: We thank Lentini and Habtemariam[1] for their thoughtful comments on our article regarding the formula for the number of possible intersections of n sets.

The formula used in our original article[2] implicitly includes intersections of each of the n sets with the universe (set of all objects one wants to consider and that may or may not be included in the n sets) as well as the set of elements in the universe that are not included in any of the n sets. The latter set indeed does not conform to the usual definition of an intersection, but for the sake of brevity we chose not to introduce an additional term to refer to both this set and the intersections. Further confusion may have arisen from the fact that the shape representing the universe is often not explicitly drawn in Venn and Euler diagrams, as in our article[2] or in the comment by Lentini and Habtemariam. Figure 1 shows an example of a Venn diagram with two sets and an explicit representation of the universe.

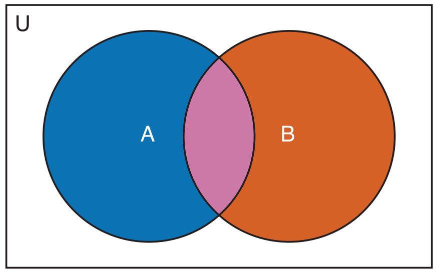

Figure 1 | A Venn diagram showing the intersections of sets A and B with each other and the implicit universe U, which is shown explicitly in this illustration.

Finally, we want to emphasize that the exponential growth of the number of possible intersections of n sets is the primary visualization challenge in this context and a key motivation for the development of interactive techniques such as UpSet[3].

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests

Alexander Lex[1] & Nils Gehlenborg[2]

[1]School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA. [2]Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, USA.
e-mail: nils@hms.harvard.edu

1. Lentini, G. & Habtemariam, S. Plotting intersections. *Nat. Methods* **12**, 281 (2015).
2. Lex, A. & Gehlenborg, N. *Nat. Methods* **11**, 779 (2014).
3. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992 (2014).