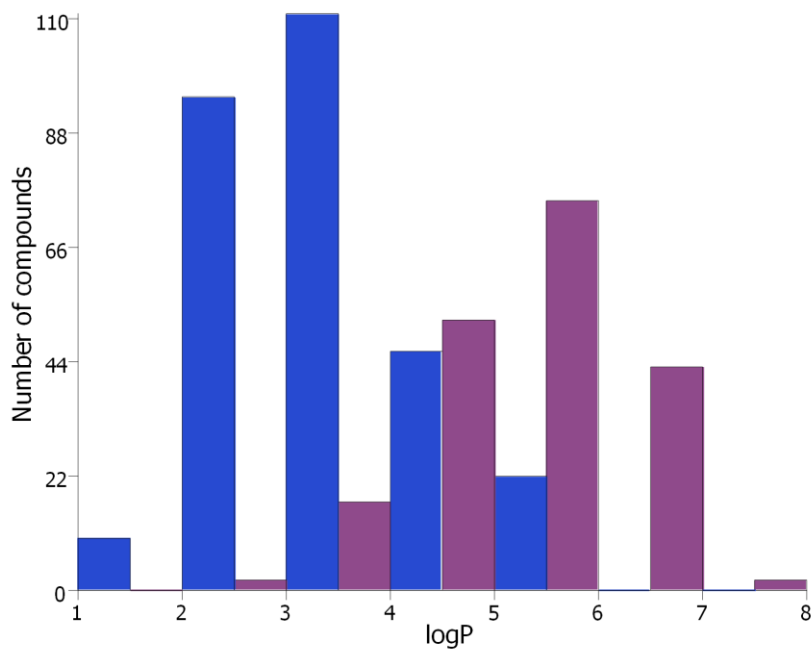


Supporting Information

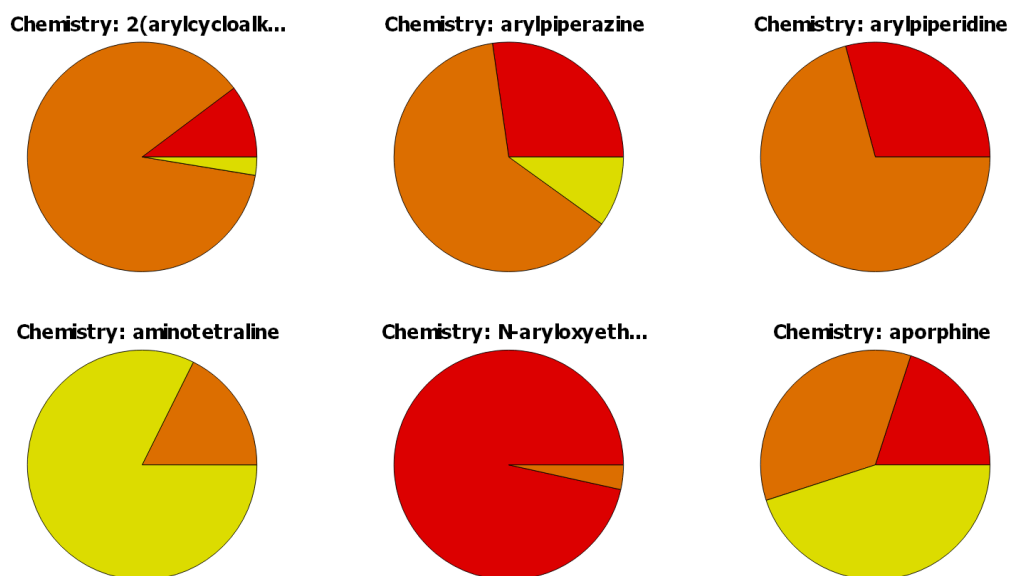
Further examples of common data visualisations

Histogram



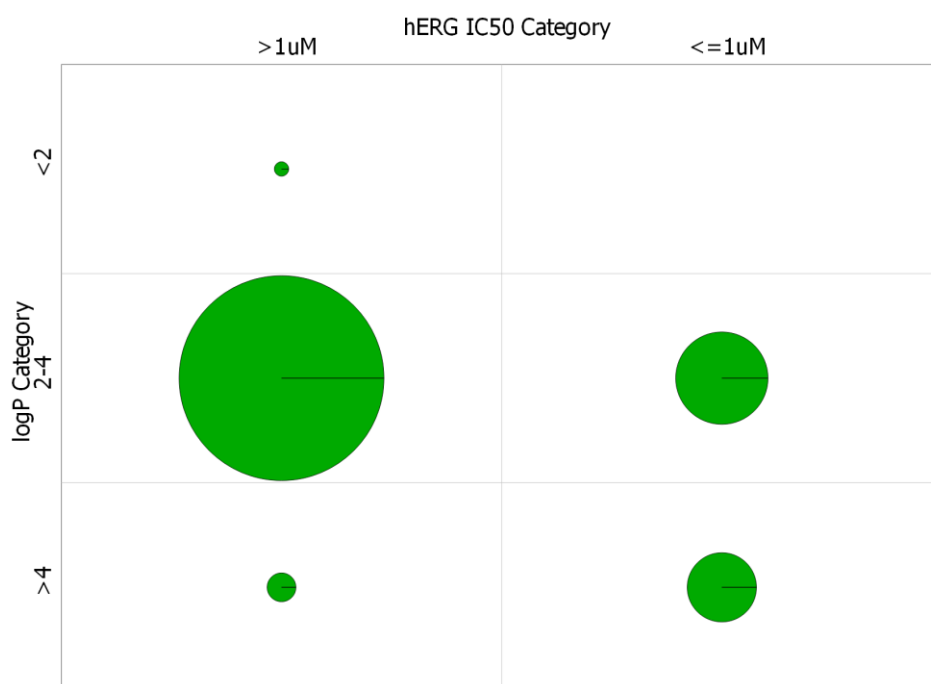
A histogram is a useful approach to view the distribution of a property for multiple compounds. The example above shows distributions of logP for two libraries, represented by the blue and purple bars.

Pie Chart



Pie charts provide another approach to summarising distributions of a property value, binned into appropriate ranges. The angle of a 'wedge' is proportional to the number of compounds with the property value in the indicated range. In the example above, distributions of solubility are compared for six chemical series. Red represents a solubility of less than 100 μM, orange between 100 μM and 1 mM, and yellow greater than 1 mM.

Categorical Distributions



For pairs of properties in which the values have been binned, the number of compounds with each combination of property bins can be displayed to visualise probability distributions. This can help to spot qualitative correlations, although care should be taken to avoid over interpretation of correlations between binned variables [1]. The example above compares the distribution of hERG IC50 with logP, from which it can be seen that, for this data set, no compounds with $\log P < 2$ have a hERG IC50 $\leq 1\mu\text{M}$, whereas the majority of compounds with $\log P > 4$ have hERG IC50 $\leq 1\mu\text{M}$.

Bibliography

- 1 Kenny PW and Montanari CA. Inflation of correlation in the pursuit of drug-likeness. J. Comput.-Aided Mol. Des. 2013;27(1):1-13.