the Sampling Lens: making sense of saturated visualisations

Geoffrey Ellis, Enrico Bertini & Alan Dix

information visualisations frequently have to deal with large amounts of data and this often leads to saturated areas in the display with considerable overplotting. This poster introduces the Sampling Lens, a novel tool that utilises random sampling to reduce the clutter within a moveable region, thus allowing the user to uncover any potentially interesting patterns and trends in the data while still being able to view the sample in context. We demonstrate the versatility of the tool by adding sampling lenses to scatter and parallel coordinate visualisations.

1 the need for density reduction

with all visualisation techniques dealing with substantial amount of data, apart from space-filling approaches, there is the possibility that portions of the display will be saturated - data points or lines are overplotted or the points are clustered as to be indistinct and in many cases, patterns will be hidden.

2 density reduction techniques

general density reduction

filtering
distortion
aggregation
clustering
zooming

avoiding overplotting

space-filling
point dithering
spatial distortion
expansion lens

3 the case for random sampling

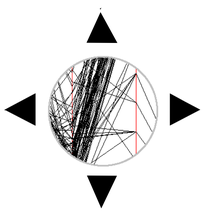
randomness ...

- is a compromise
- makes things possible
- makes other things faster
- is used widely in computer science
- results may not be perfect but are 'good enough'

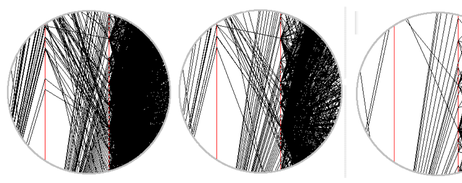
traditional algorithms are
- optimisation and
- machine learning:
- digital signal proces
- telecommunications
- network protocols:
- parallel computing:
- cryptography:

4 the Sampling Lens

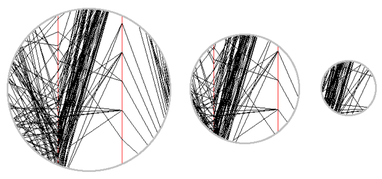
To ensure display continuity, points that are removed as the sample rate increases, reappear in reverse order when rate decreases



drag around



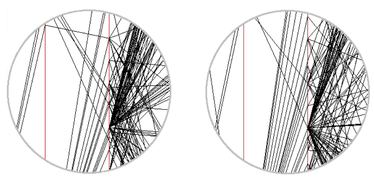
change the sampling rate within the lens between 1 and 100%

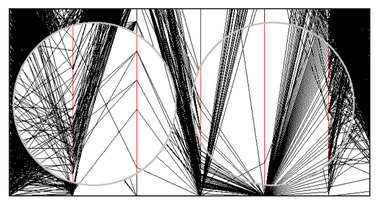


change the size of the lens

Reality Check

the user can click on this button to view a completely new sample within the lens, thus 'real' patterns will persist whilst sampling induced artefacts will disappear

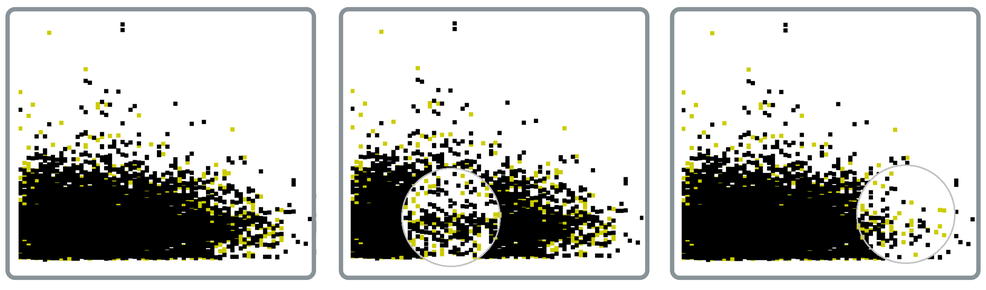




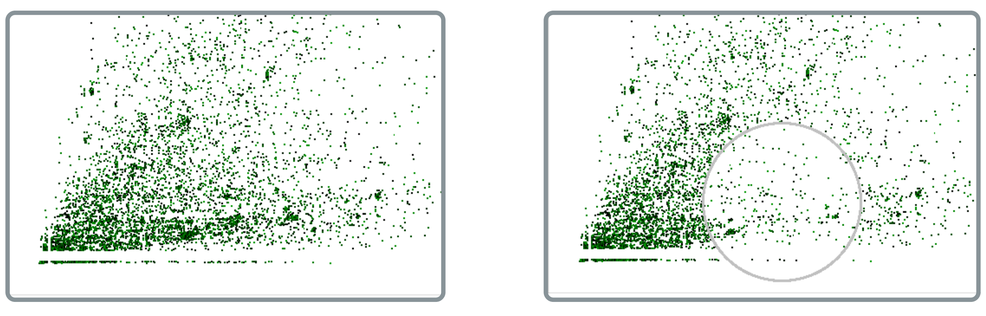
auto-sampling

the system dynamically adjusts the sampling rate by applying 'greater sampling' in saturated areas and 'light sampling' in sparse areas so low density patterns are not removed; hence maintaining a constant proportion of overlapping items

5 scatterplots

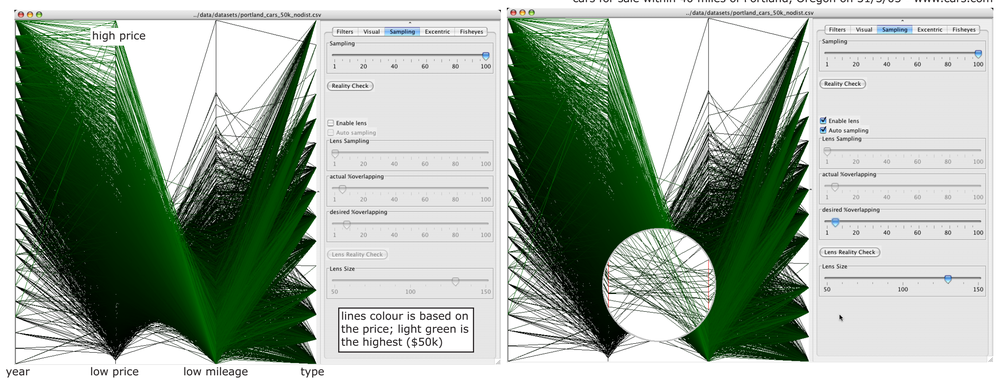


the graph has considerable overplotting and to make matters worse, the data happens to be sorted so that the darker points obscures the lighter points. In these circumstances, the lens not only enables patterns to be found but also gives a good indication of the proportion of light and dark points



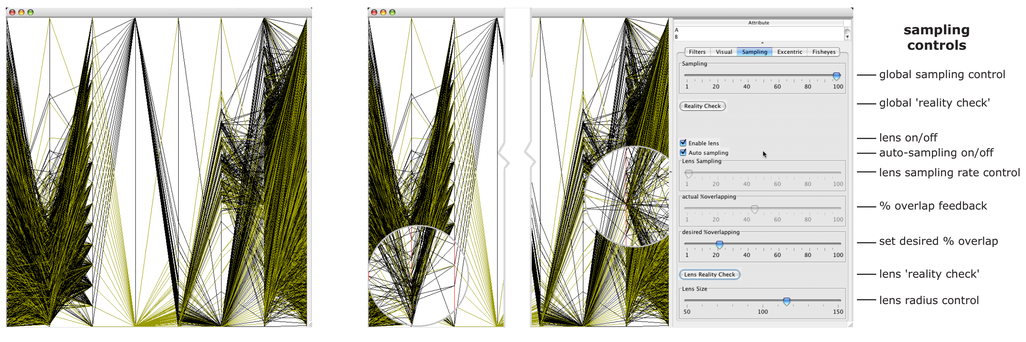
this visualisation of mail parcel data from the German post-office plotted according to weight and volume. Without the lens, it is not obvious which clusters are dense and hence more significant. Applying the lens over the area of interest reveals clusters representing common combination of goods and boxes

6 parallel coordinate plots



without the lens, it appears that low mileage cars are on sale at high prices

with the lens, it appears that some fairly low mileage cars are on sale at reasonably low prices. The lens sampling rate is 1%



patterns are not obvious

the lens exposes interesting trends

http://www.hcibook.com/alan/papers/chi2005-samplinglens/