



FAKULTÄT FÜR
INFORMATIK



Visualization Group

SCIENTIFIC PROJECT REPORT

Enhancement of GPLOM for use with Epidemiological Data

Author:
Matthias Graf

Supervisors:
Paul Klemm
Prof. Bernhard Preim

February 22, 2014

Abstract

This report proposes an enhancement for filtering and highlighting data in generalised plot matrices (GPLOM). Its goal is to improve data analysts ability to explore multivariate datasets composed of both categorical and quantitative information. A prototype illustrates the technique for the epidemiological Study of Health in Pomerania (SHIP).

Contents

1	Introduction	3
1.1	Problem Description	3
1.2	The SHIP Dataset	3
1.3	Outline	3
2	Related Work	4
2.1	Parallel Sets	5
2.2	Scatterplot Matrix	5
2.3	Generalised Pairs Plot	5
2.4	Generalised Plot Matrix	6
3	Enhanced GPLOM Prototype	8
3.1	Interface Overview	8
3.2	Filtering & Highlighting	10
4	Discussion	11
4.1	Dataset Related Issues	11
4.2	Axis Labels	12
4.3	Matrix Diagonal Histograms	12
4.4	Variable Order	12
4.5	Heatmap	12
4.6	Histogram	13
4.7	Scatterplot	13
4.8	Filtering & Highlighting	13
5	Conclusion and Future Work	14

1 Introduction

The Study of Health in Pomerania (SHIP) is a longitudinal epidemiological study. The initial cohort consisted of 4308 study participants. One of its goals is to "investigate associations and interactions among common risk factors, subclinical disorders and clinical diseases" [Völzke et al., 2011]. In order to achieve this goal, a visual exploration of the dataset can be a first step towards understanding the structures and patterns inherent in the data. This can be a basis for the formulation of hypotheses later used in multivariate analysis. One of the research tools for the visual analysis of the SHIP is the Cargo framework [Klemm et al., 2014]. It integrates classical information visualisation techniques, a table, scatterplots and parallel coordinates into a linked visual analytics tool. The heterogeneous nature of the data, which is comprised of a wide range of sociodemographic and medical variables, poses several challenges.

1.1 Problem Description

The primary problem that this scientific project set out to tackle is the unsuitability of scatterplots and parallel coordinates for displaying categorical variables (measures). Those plots translate categories to a continuous numerical scale, leading to overlapping datapoints (overplotting) which obscures important properties of the underlying dataset. For example, frequency of a certain category compared to another is hidden. Furthermore, in the case of nominal variables (e.g. sex), an arbitrary order and spacing is imposed, which may confuse analysts into perceiving non-existing relations.

1.2 The SHIP Dataset

Overplotting is a major issue because a large proportion of the SHIP variables are categorical. The survey uses standardised questionnaires which commonly predefine a set of possible answers. The sample SHIP-2 dataset (n=2333) used in this project contains 29 quantitative (metric) variables and 95 categorical variables. Of those, 62 are dichotom (binary), 27 ordinal and 6 nominal. Dichotomous variables often stem from Yes-No-Questions and ordinal variables often stem from binning of quantitative information. In addition, error codes exist to indicate, for example, that a certain question was not answered. Ordinal and nominal variables have the following cardinality histogram: 6x 8, 17x 9, 4x 10, 4x 11, 1x 13, 1x 14.

1.3 Outline

The first part of this project focused on finding a suitable information visualisation technique for the given task and dataset. It is required to enable a visually guided exploratory data analysis that allows epidemiologists to find structures and patterns of interest. Specifically, it has to support the visual display of relations and comparisons between both categorical and quantitative variables. In order to achieve this, overplotting has to be avoided. The visualisation technique has to scale well to many variables, provide

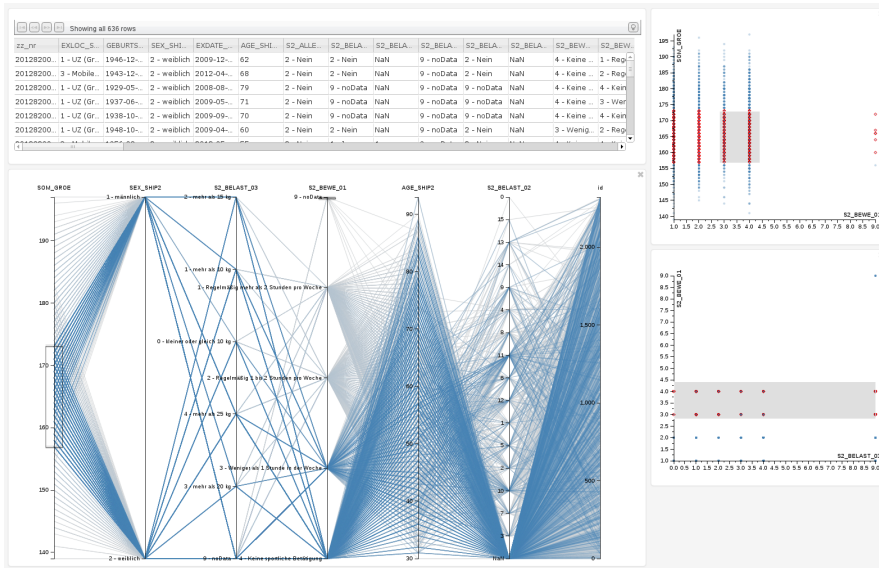


Figure 1: Cargo user interface with sample selection and applied brushing. Overlapping datapoints and lines conceal frequencies and patterns.

an overview first and details on demand. The main focus lies on preliminary exploration, not on providing tools for sound statistical hypothesis testing.

The related work section discusses existing techniques: parallel plots, parallel sets, scatterplot matrix, generalised pairs plot and generalised plot matrix. Of those, the generalised plot matrix (GPLOM) was found to be most promising, but had only limited filtering and highlighting capabilities. Those are important for revealing details in the dataset by letting the analyst formulate specific search directions. The original GPLOM allows only one category of one variable to be highlighted at a time, and only single categories to be filtered successively. The second part of this project focused on developing improved filtering and highlighting features for GPLOMs which work on all variables, both categorical and quantitative, on multiple categories and multiple variables at a time. A prototype was developed to demonstrate the proposed techniques. A discussion and outlook conclude this report.

2 Related Work

The Cargo framework is the primary reference for this project [Klemm et al., 2014]. It is a research prototype developed for the testing of visual analytics techniques for the SHIP dataset. Its webbased frontend combines a table, scatterplot and parallel coordinates into a coordinated view. Figure 1 shows its user interface with a particular selection of variables for the plots. Brushing and linking among all views is possible. The figure shows a rectangle being brushed in the scatterplot for SOM_GROE (body height) and

S2_BEWE_01 (amount of workout done in winter). The colour blue indicates the brushed area in the parallel coordinates. Both variables are essentially quantitative (metric) but appear as categorical in the dataset due to binning. As a result, many datapoints overlap one another in the plots, which renders them unsuitable for identifying patterns because it is not possible to see how many dots (in the scatterplot) or lines (in the parallel coordinates) actually overlap. Categories are simply coded as continuous numerical values. In the case of nominal variables, which do not have an inherent order (e.g. job), this translation imposes a meaningless order and spacing between the categories.

Heinrich et al. summarise three ways of dealing with overlapping in parallel coordinates [Heinrich and Weiskopf, 2012]: colouring lines, plotting lines as curves and mapping categories onto intervals (see also [Teoh and Ma, 2003]). However none of them alleviates the basic problem: that categories are interpreted as numerical.

2.1 Parallel Sets

Parallel Sets are an adaption of parallel coordinates for categorical variables [Bendix et al., 2005]. The idea of mapping onto intervals on the axis remains, but not lines but streams are drawn. This clearly improves the visibility of frequency of categories, which is indicated by the breadth of the stream. However, parallel sets do not support quantitative variables. How can the benefits of both techniques be combined to be applicable to both types of variables? This may be a question worth investigating in the future. But other drawbacks remain. Parallel Sets do not scale well with the number of variables and their cardinality. The higher those, the thinner the streams get. The order of variables and categories also highly affects the visual display, because the first variable determines the sink of the streams and their colour. Also, variables can only be directly compared if they are situated side-by-side. An unfortunate order may therefore conceal patterns.

2.2 Scatterplot Matrix

The pairs plot or scatterplot matrix (SPLOM) is another traditional information visualisation technique for visualising relationships in multivariate datasets [Hartigan, 1975]. It combines scatterplots for all variable pairings into a matrix grid. This provides a good overview of the dataset, because all datapoints are visible at once, in all bivariate combinations. Scatterplots are easy to interpret. Another advantage of the SPLOM is that the order of the variables does not alter the visual display much, because the individual plots remain unchanged.

2.3 Generalised Pairs Plot

The generalised pairs plot is a recent improvement of the SPLOM to account for the problem of overplotting that occurs with categorical variables [Emerson et al., 2013]. Instead of scatterplots, it uses more appropriate plots depending on the variable type combination. For example, mosaic plots can be used if two categorical variables are plotted, or side-by-side box plots if a categorical meets a quantitative variable. This is shown in figure 2. If the same plots are used on both sides of the matrix diagonal,

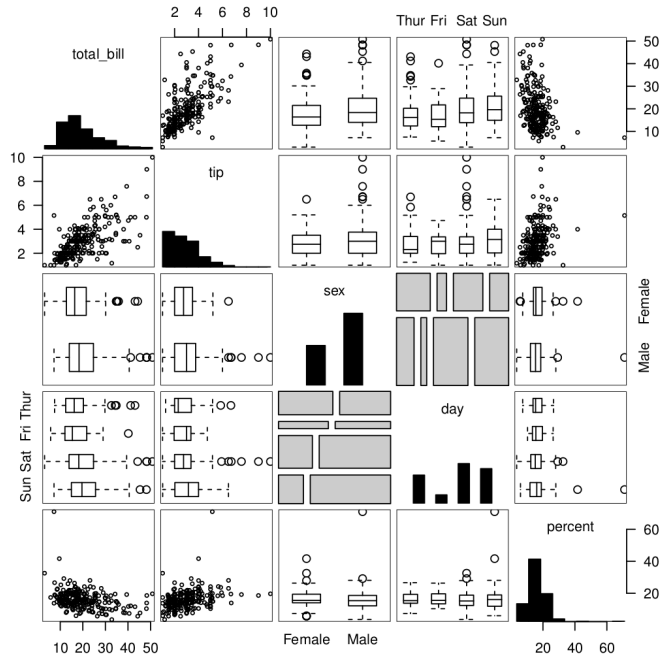


Figure 2: Generalised Pairs Plot example

one is a mirror of the other. This can be considered a drawback, because almost half of the space is devoted to displaying redundant information. Another problem is that the visual clarity of the mosaic plot and side-by-side box plot does not scale well with the cardinality of categorical variables. Furthermore, no filtering or highlighting is supported.

2.4 Generalised Plot Matrix

The generalised plot matrix (GPLOM) proposes solutions for those drawbacks [Im et al., 2013]. It only uses one half of the matrix and sorts the variables so that all plot types are grouped (see figure 3). This eases comparisons between plots of the same type, due to the improved visual proximity and alignment. Instead of side-by-side box plots it uses histograms. Depending on the aggregation function, the histograms may show the minimum, average, maximum, count or sum of the quantitative variable for each category. Compared to side-by-side box plots, histograms are easier to read with higher cardinalities of categorical variables [Im et al., 2013, see Figure 7]. However, they also contain less information. Mosaic plots are replaced by heatmaps, which also scale better to higher cardinalities. This comes with a drawback: as values are now coded by colour brightness, not by space, comparing values is harder. The combined use of histograms and heatmaps has the significant advantage of aligning all categories vertically and horizontally. Each bar and compartment can therefore be traced and compared across each axis. The GPLOM also proposes techniques for interactivity: brushing, linking, filtering, tooltips

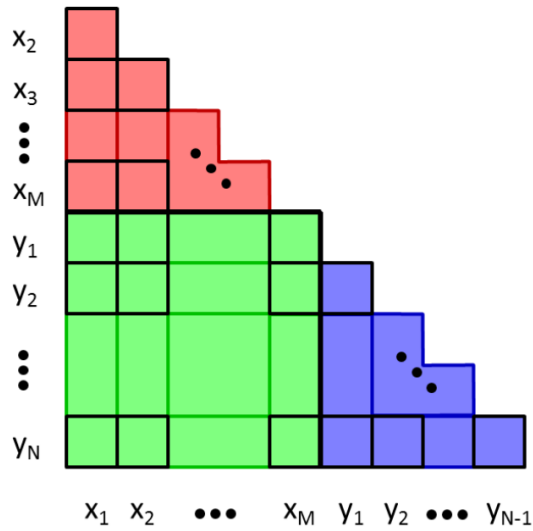


Figure 3: GPLOM structure - red: heatmaps, green: histograms, blue: scatterplots [Im et al., 2013, see Figure 6]

and textual search.

Of all techniques considered, I chose the GPLOM to be the most potent for an exploratory data analysis of the SHIP sample dataset. It provides an overview of all bivariate relations, is designed to work well with both categorical and quantitative variables, scales well with category cardinality, allows comparisons to be made between different plots and supports filtering and highlighting.

Yet, much potential for improvement remains. Due to the high number of variables in the sample dataset, one crucial aspect of the visualisation technique is the ability to narrow the perspective on the data. The only way GPLOMs support this is by filtering distinct categories via double clicking the corresponding histogram bars. Filtering works by excluding all the other categories of the same variable from the dataset. This operation automatically hides the variable, adds it to the filter list and updates all plots. Multiple variables can successively be filtered in this manner. A datapoint has to meet all filter criteria to be shown, which means the resulting set is the intersection of all filters. Each filter criterion is limited to including exactly one category. Neither exclusion, nor inclusion of multiple categories is possible. Quantitative variables can not be filtered at all. This is a major restriction.

Associative highlighting (linking) is performed by single-clicking a histogram bar. All datapoints that belong to this particular category are highlighted by colour in all scatterplots and histograms. It thus allows the analyst to see the effect of the filter before it is applied by showing a comparison between the current selection and the post-filter selection. This is crucial because it shows how the filtered subset is distributed and how this distribution compares to the rest of the data. As with filtering, highlighting is

restricted to only one category of one categorical variable at a time. Selections can not be made on quantitative variables and they can not be narrowed or broadened any further.

Variables can also not be hidden without applying a filter. This means that analysts are not able to pick only certain variables of interest. Focusing without filtering is therefore impossible. All things considered, the analysts ability to explore the dataset in detail is limited, which may hamper him from finding certain patterns and structures.

It is important to note that the GPLOM is not tailored towards epidemiological studies. It is assumed that the core requirements of an exploratory data analysis can be generalised. This may not be the case. To the best of my knowledge, no such differences have yet been discussed in the literature.

3 Enhanced GPLOM Prototype

The prototype developed as part of this project is an alternative GPLOM implementation. It is webbased and builds on the D3 framework ¹. This prototype was build primarily for the sake of demonstrating the enhanced highlighting and filtering techniques proposed in this report. It does not contain all features of the GPLOM. For example, bendy highlights, textual search and the infobox (with kernel density estimation) are not implemented. In contrast to the original implementation, this prototype is free and open source software ². The sample SHIP dataset could not be included because it is deemed confidential. The interactive features of the prototype are illustrated in a screencast accompanying the report.

3.1 Interface Overview

Figure 4 shows an annotated screenshot of the user interface. The GPLOM is a lower triangular matrix filled with 3 plot types: heatmaps, histograms and scatterplots. Each row and column is labeled with the variable name it belongs to. In the upper right corner is a variable selection menu. On its left is the list of filter criteria currently defined. Hovering the mouse over various elements shows tooltips and labels.

Heatmap A heatmap is drawn for two categorical variables. It contains compartments for each combination of categories and colours them according to the number of datapoints found in each. The darker the colours the higher the frequencies. Hovering the mouse over a compartment shows a tooltip revealing what categories it belongs to and how many datapoints it contains. The colour mapping (lightness) is scaled for each heatmap and selection, which means that colours are not comparable between different heatmaps or before and after a filter is applied.

Histogram Histograms are drawn when a categorical variable is plotted against a quantitative variable. The original GPLOM implementation lets the analyst select which

¹<http://d3js.org/>

²<https://github.com/rbyte/F-GPLOM>

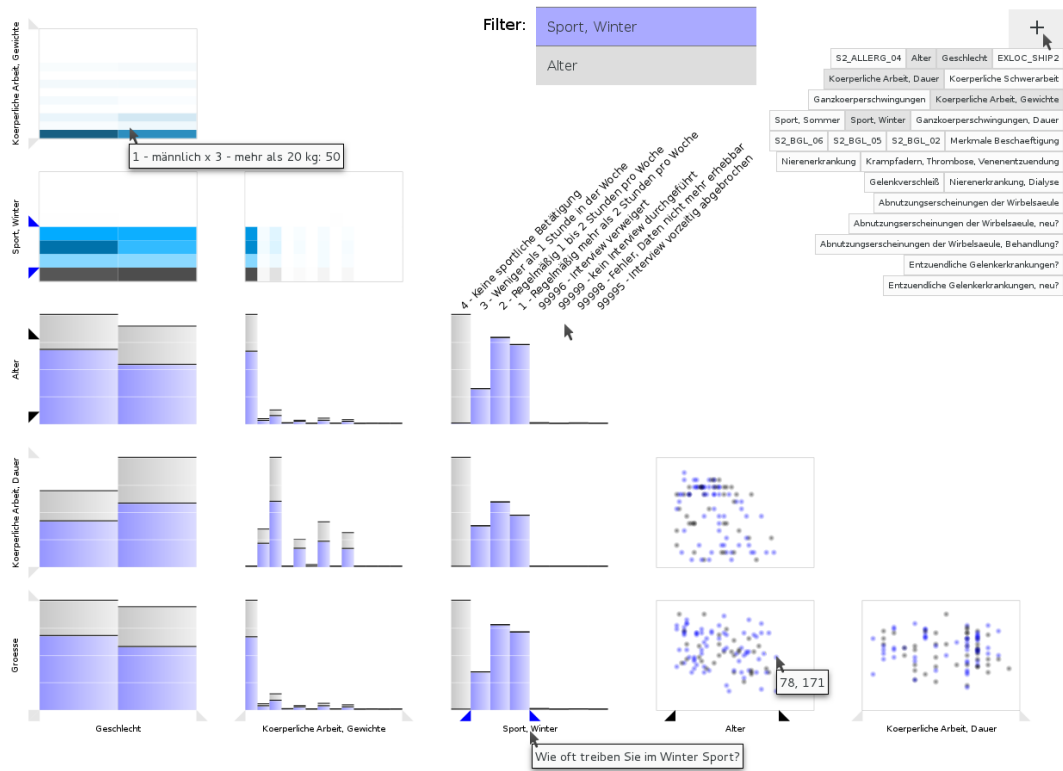


Figure 4: User interface of the prototype with sample variable selection, filtering and highlighting applied and tooltips shown for datapoints in scatterplots, details on variables, categories and heatmap compartments

aggregation function should be used (min, max, average, sum, count). The prototype currently only supports sum aggregation. The bars height is scaled between zero and the histograms maximum. Comparing bars between histograms is therefore not possible. Hovering over a histogram reveals the categories each bar belongs to.

Scatterplot Scatterplots are used for two quantitative variables. Their bounding boxes are scaled to each variables global minimum and maximum. This allows for comparisons to be made between scatterplots that are on the same row or column. Each point is a black circle with a radial transparency gradient. If the semi-transparent points overlap, the resulting point is darker, which makes it possible for the analyst to identify overlap and its magnitude. This is a proposed improvement of scatterplots.

3.2 Filtering & Highlighting

It follows a description of proposals for improvements of the filtering and highlighting capabilities of the GPLOM.

Show & Hide Variables The first is the variable selection menu. It allows analysts to show and hide variables. This is important because the GPLOMs visual complexity (number of plots) increases exponentially with the number of variables displayed.

Range Sliders The second major improvement is the introduction of filtering for quantitative variables with range sliders. The sliders allow the definition of an upper and lower bound for each variable. Everything outside that range is excluded. The sliders are also used for categorical variables and enable the selection of multiple categories for filtering and highlighting. The sliders exist for each row and column and are located next to the variable labels. Since all except two variables are listed on both axis, a pair of sliders on one axis often has a corresponding pair of siblings on the other axis. Those are tied and change synchronously. GPLOMs so called bendy highlights visualise this relationship between rows and columns.

Associative Highlighting Dragging sliders interactively highlights the selected range in all plots (brushing and linking). Figure 4 shows the variable "Sport, Winter" being highlighted. Sliders are grey in their default position, a range that included everything. They turn blue if they are dragged. Points in scatterplots are coloured blue if they belong to the defined range. The same is true for histograms and heatmaps. Histograms are overlaid by blue bars and in heatmaps, highlighted proportions are coloured blue, with higher saturations for higher proportions. Highlighted values are not visible in the heatmaps of the original GPLOM implementation. Utilising saturation to indicate proportions is a proposal for improving the highlighting technique.

Two Filter Stages Highlighted variables are added to the filter criteria list, with a blue background colour. This is a deviation from the original implementation. The list now

features two stages: highlighted filters and applied filters on variables. Highlighted filters can be applied by clicking a blue list item. The background turns grey and the sliders black to indicate this. In the example in Figure 4, "Alter" has an applied filter. All points outside its filter range are excluded entirely, while "Sport, Winter" is only highlighted in blue, but still included in all computations. This distinction between highlighted and applied filter criterion has two advantages.

First, multiple criteria can be defined for highlighted selections. This allows analysts to define more fine grained comparisons between the current selection (grey) and a specified subset (blue). The subset can now contain criteria on quantitative variables and multiple categories from categorical variables. This was not previously possible.

Second, applied filters can easily be highlighted (undone) by clicking the grey list items. Filtered variables are also not hidden automatically. This allows applied filters to be altered by adjusting the black sliders. In addition, context is preserved. Variables can always be included or excluded with the variable selection menu, but this does not affect the filter list, which keeps a record of the steps taken so far in narrowing down the dataset.

4 Discussion

The prototype still has many issues and problems. First of all, it does not implement all GPLOM features. For example, bendy highlights could make the relations between columns and rows clearer. A textual search could reduce time needed for a visual search and letting the user change the histogram aggregation function enables important changes of perspective on the data.

If all variables of the SHIP dataset are viewed at once, the GPLOM is too complex to read. Although the prototype lets the analyst select only a subset of all variables to be displayed, this comes with the drawback that no complete overview is given.

4.1 Dataset Related Issues

Error Codes The dataset contains error codes. They are used to include various meta information, for example, that a question was not answered. In categorical variables, those error codes are simply coded as additional categories. In quantitative variables, fields with error codes are ignored. Both approaches may be inappropriate. Suggestions would be to code errors separately, ignore error fields by default or group error codes into one category only.

Binning Binning of quantitative variables is another problem. Age, given in years, is a good example. Those variables are coded as ordinal categorical, but this does not reflect their quantitative nature. How classes (bins) are defined has an important effect on the perception of distribution. If individual bin ranges have different sizes, for example low income 0-3000, mid income 3000-5000, high income 5000-13000, but are visualised in histogram bars of the same breadth, data distortion occurs. Also, coarse binning impedes

fine grained analysis and worsens the effect of overlapping in the data. Since binning reduces information contained in the dataset, it should be an option for the visualisation technique or analyst to choose, not be preimposed by the survey, because it can not be undone.

Dichotomous Variables Another property of the sample dataset are its many dichotomous variables (alternatives), mainly yes-no questions. Currently, the prototype codes them as nominal categorical variables with cardinality two. A possible improvement may be to use mosaic plots or violin plots instead.

4.2 Axis Labels

Another important drawback is the missing axis labeling for each plot. Category names, ranges and numbers are only retrievable by checking labels and tooltips. How can this information be integrated in the display without crowding it? Important key values, like domain boundaries for scatterplots, or the baseline and maximum of histograms should be included in future implementations.

4.3 Matrix Diagonal Histograms

Compared to the generalised pairs plot, the GPLOM does not include a matrix diagonal for displaying a one-variable histogram (see figure 2). Its inclusion may improve the visualisation. However, it is important to include visual clues to indicate that those histograms on the diagonal show something completely different than the ones that plot a categorical against a quantitative variable.

4.4 Variable Order

The order of variables and of nominal categories is another issue worth considering further. Currently, the prototype orders them arbitrarily. It may however be useful to order variables by similarity, so that proximity reflects similarity. This requires a measure for similarity, such as correlation. Johansson et al. has even proposed including correlation coefficients into each plot [Johansson and Johansson, 2010].

4.5 Heatmap

In heatmaps, saturation is used to decode the proportion of highlighted items in a certain bucket or compartment. This is a very dense coding, since lightness is already used for overall compartment frequency. But the main drawback is the difficulty to visually quantify differences in colour. Another problem is that blue saturation may have an impact on the perceived lightness of a colour, even though the HSL colour model is used.

In addition, the local scaling makes different heatmaps incomparable. However, a per row or global scaling may hide (local) details. If the two variables have a low cardinality, mosaic plots may be a better choice. But this would degrade the advantages of the heatmaps alignment and render bendy highlights useless. Another suggestion for

improving heatmaps is to allow highlighting and filtering through hovering, clicking and dragging over compartments and areas.

4.6 Histogram

Although histograms are generally easy to read, to understand which aggregation function is used is difficult and unintuitive, because it is ambiguous. The sum function, which the prototyp uses, may not be suitable for certain variables. For example, plotting age against sex, the average or count function probably are a better choice.

The data density of the histogram is directly linked to the variable cardinality. This visual data density could be equalised by giving each histogram bar, not variable, a fixed amount of space. This change would also affect the display of the heatmap then, which would have a more even spacing. The problem with this approach may be that variables with a huge number of categories take up too much space. Limits on the space each variable takes could solve this.

Due to the local scaling of each histogram's bar height, multiple histograms are not directly comparable. This may be desirable on a per row basis.

A general problem of histograms is that they hide the distribution of the datapoints for each category. This can only be alleviated by using a different plot, for example side-by-side kernel density distributions. Providing the analyst with the option to switch between multiple types of plots is worth considering.

4.7 Scatterplot

The performance of the interface degrades with the number of points plotted in each scatterplot. To counter this effect, each scatterplot only shows a random selection of 100 points. This has the drawback that whenever the scatterplot is redrawn, the random selection changes and flickering occurs. Future implementation have to improve the scatterplot performance.

4.8 Filtering & Highlighting

Effects between variables can be complex. If those effects only occur under certain preconditions, filtering and highlighting may help reveal the relevant subsets, if the defining criteria for such a subset are sufficient. This report proposed enhanced means of defining criteria through range sliders for each variable. The combination of different criteria allow fine grained subsets to be defined. The interactive display of those subsets on all bivariate plots allows the analyst to reveal multivariate relationships and explore the dataset. However, there are still limits. For example, selecting an area in a scatterplot is limited to a rectangle, no individual points can be selected and unions or exclusions are also not supported. All filtering is currently based on one dimensional range inclusion and set intersection. More diverse options require enhanced user interface utilities.

The authors of the GPLoM mentioned that associative highlighting lacked clear affordances and was rarely used. The sliders provide those affordances. However, that

items in the filter list can be clicked is still not obvious. This has to be improved in the future.

Another improvement of the associative highlighting could be to allow the analysts to set colours for each filter criterion and group criteria. Different groups could then be highlighted in different colours and intersections as a mixture of the two colours.

The current range implementation imposes a binary in-out decision on the dataset. Smooth brushing [Hauser et al., 2002] may improve the brushing and highlighting in scatterplots by colouring datapoints with gradual increases in blue saturation, as the selected point moves closer to the selection center.

A major problem occurs when a filter range is defined through dragging the sliders of a quantitative variable on the y-axis. Because the sliders are aligned with a row and the histograms blue overlay bars are interactively responding to the sliders movements, a confusing cause-effect connection is visible. This is due to the fact that the histogram bars hide the distribution of the datapoints inside each category and that the slider domain and the histogram height domain are not the same, in fact have little to do with one another.

5 Conclusion and Future Work

To solve the problem of overplotting, I reviewed several information visualisation tools and proposed the GPLOM to complement the Cargo framework. It provides a broader overview of the dataset, a more diverse set of perspectives and, with the new filtering and highlighting techniques, a finer grained exploration facility.

A prototypal implementation has been developed as part of this project. An important next step is the evaluation of the prototype. Considerations about the suitability of the prototype have to center around user needs and tasks. Those where framed in the context of a large scale epidemiological study that requires analysts to explore the gathered data in an exploratory, visually guided fashion. Whether the proposed interface is suitable is therefore a question left open for the epidemiologists to answer. A study could provide insights into whether the tool is preferred and how suitable it is for the task. In particular, do the proposed additions actually make it easier to find patterns and understand the dataset? Various problems and drawbacks have been discussed and suggestions have been made for future research.

References

- F. Bendix, R. Kosara, and H. Hauser. Parallel sets: visual analysis of categorical data. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 133–140, 2005. doi: 10.1109/INFVIS.2005.1532139.
- John W Emerson, Walton A Green, Barret Schloerke, Jason Crowley, Dianne Cook, Heike Hofmann, and Hadley Wickham. The generalized pairs plot. *Journal of Computational and Graphical Statistics*, 22(1):79–91, 2013.

- John A Hartigan. Printer graphics for clustering. *Journal of Statistical Computation and Simulation*, 4(3):187–213, 1975.
- Helwig Hauser, Florian Ledermann, and Helmut Doleisch. Angular brushing of extended parallel coordinates. In *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*, pages 127–130. IEEE, 2002.
- Julian Heinrich and Daniel Weiskopf. State of the art of parallel coordinates. In *Eurographics 2013-State of the Art Reports*, pages 95–116. The Eurographics Association, 2012.
- J.-F. Im, M.J. McGuffin, and R. Leung. Gplom: The generalized plot matrix for visualizing multidimensional multivariate data. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2606–2614, 2013. ISSN 1077-2626. doi: 10.1109/TVCG.2013.160.
- Sara Johansson and Jimmy Johansson. Visual analysis of mixed data sets using interactive quantification. *ACM SIGKDD Explorations Newsletter*, 11(2):29–38, 2010.
- Paul Klemm, Lisa Frauenstein, David Perlich, Katrin Hegenscheid, Henry Völzke, and Bernhard Preim. Clustering Socio-demographic and Medical Attribute Data in Cohort Studies. In *Bildverarbeitung für die Medizin (BVM)*, page in print, 2014.
- Soon Tee Teoh and Kwan-Liu Ma. Paintingclass: interactive construction, visualization and exploration of decision trees. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 667–672. ACM, 2003.
- Henry Völzke, Dietrich Alte, Carsten Oliver Schmidt, Dörte Radke, Roberto Lorbeer, Nele Friedrich, Nicole Aumann, Katharina Lau, Michael Piontek, Gabriele Born, et al. Cohort profile: the study of health in pomerania. *International journal of epidemiology*, 40(2):294–307, 2011.