Component Selection for the Metro Visualisation of the Self-Organising Map

Robert Neumayer, Rudolf Mayer, and Andreas Rauber
{**neumayer,mayer,rauber**}**@ifs.tuwien.ac.at**

Vienna University of Technology
Institute for Software Technology and Interactive Systems
Favoritenstraße 9-11, 1040, Vienna, Austria

presented at WSOM'07
September 4, 2007

som visualisations

plain clusterings overwhelmingly difficult to understand

visualisation of

- cluster quality
- instances
- attributes/components across clustering

component planes

- + visualisation of all components
- - one illustration per component

the metro visualisation

component planes $\longrightarrow$ component lines

metro line metaphor for data visualisation

- skewed distances
- layer of abstraction / simplified structure
- easier to understand, interpret and memorise
- overlayed onto any som visualisation

unsupervised feature selection with respect to a certain som

aggregation of component lines

london metro maps

real-world map

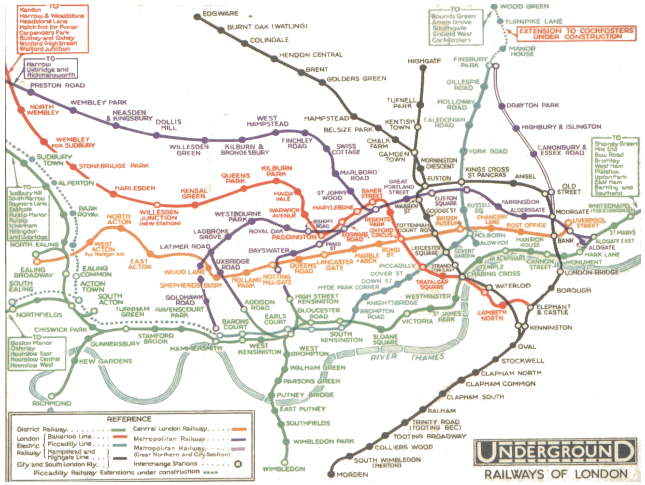


Figure: correct 1932 metro map

skewed distances

Figure: classic 1933 metro map

component planes

model vector visualisation

partitions of projections of single variables

number of plots equals number of components

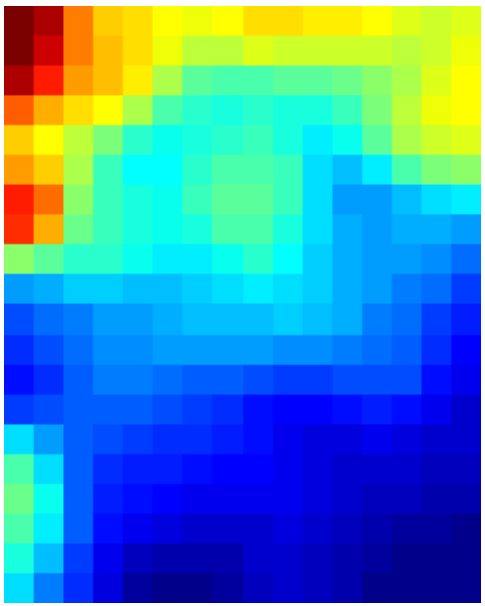


Figure: component plane visualisation for a single variable

one single, continuous component plane

projection of single variable

discretisation

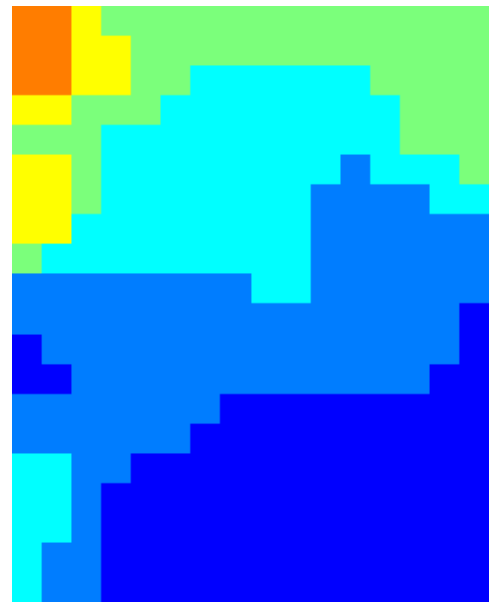


Figure: after discretisation step for a single component

binned component plane

computation of centres of gravity

interconnecting lines

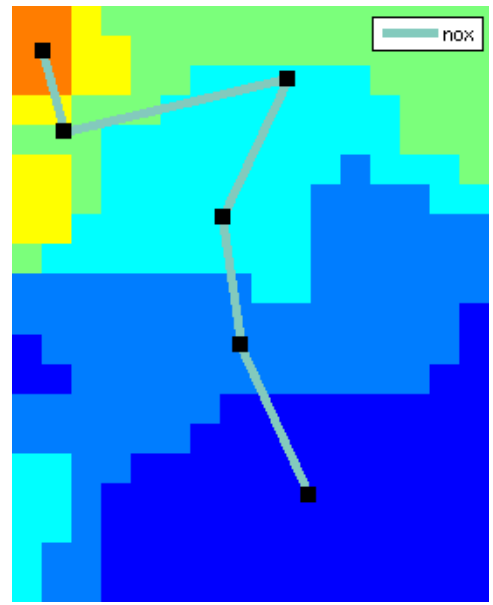


Figure: metro visualisation based on centres of gravity for a single component

distances between component lines

distance measure necessary

minimum pairwise distances

computed for both directions

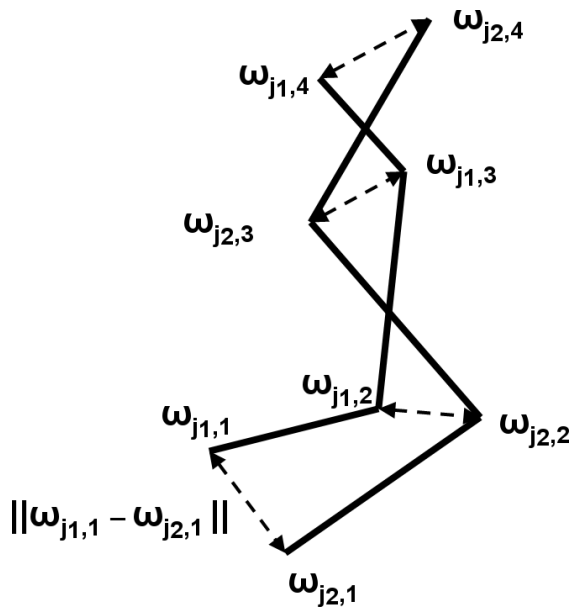


Figure: distance function for component lines

visual enhancements

snapping of lines onto som grid

heuristic algorithm

leads to aligned metro lines

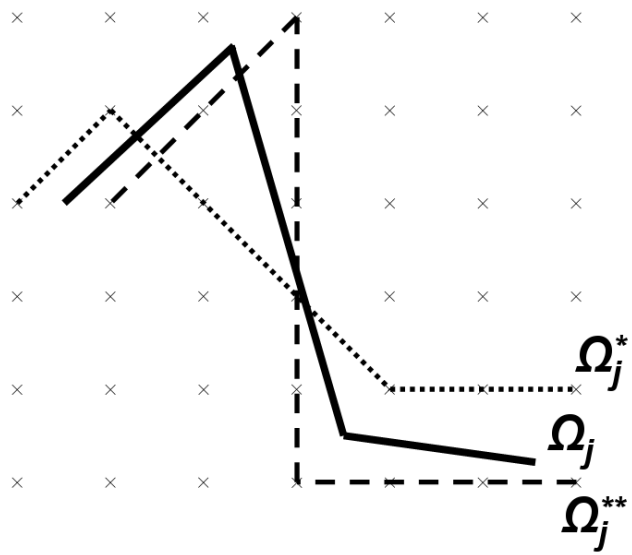


Figure: snapping of metro lines

component selection 1/2

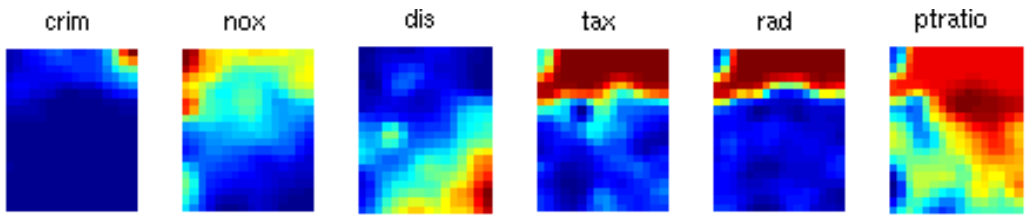
visualisation not always feasible

selection of feasible components

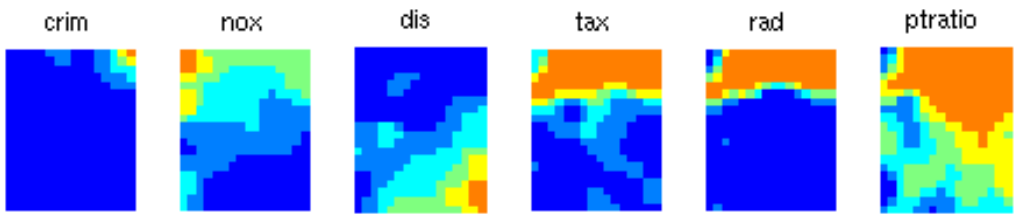
for a given som clustering

visualisation of scattered components makes less sense

component selection 2/2



(a) plain component planes



(b) binned component planes

Figure: not all components are equally feasible for metro visualisation

.750	.546	.500	.214	.210	.177

Table: component region ratios

component aggregation

plotting of all components might overload the illustration

selection of most feasible components

- ward's clustering on component lines
- based on line distance function

resultant illustration is less crowded

the boston housing data set

describes housing in the boston area

506 instances

14 components

$20 \times 16 = 320$ units

discretisation performed for six bins

boston housing discretisation step

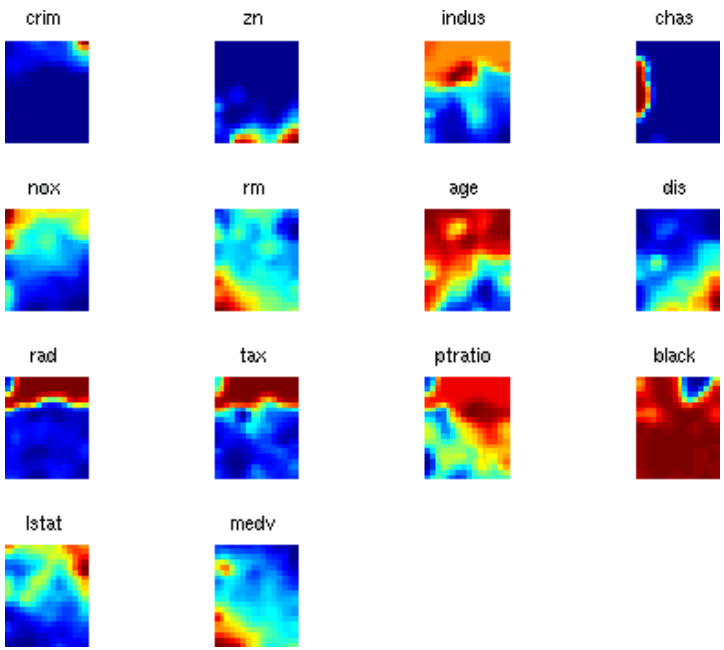


Figure: component planes visualisation of all variables

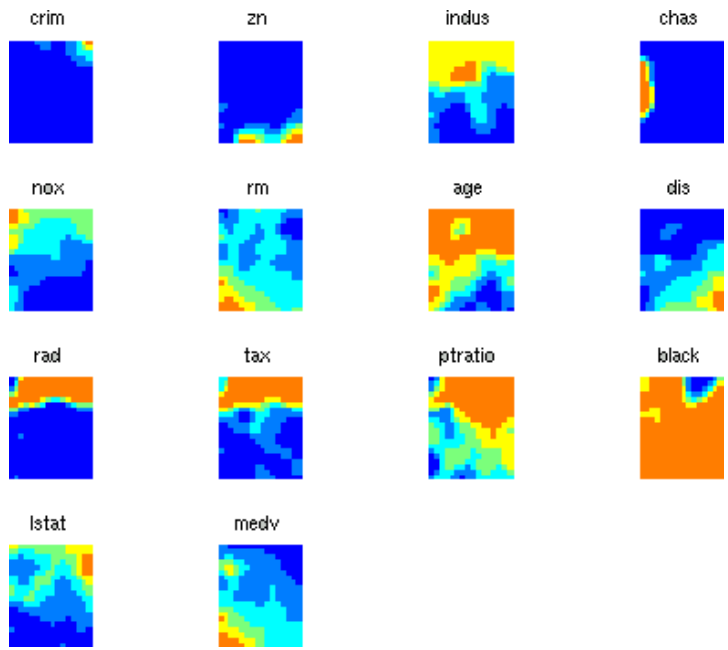


Figure: all variables after discretisation

binned component lines

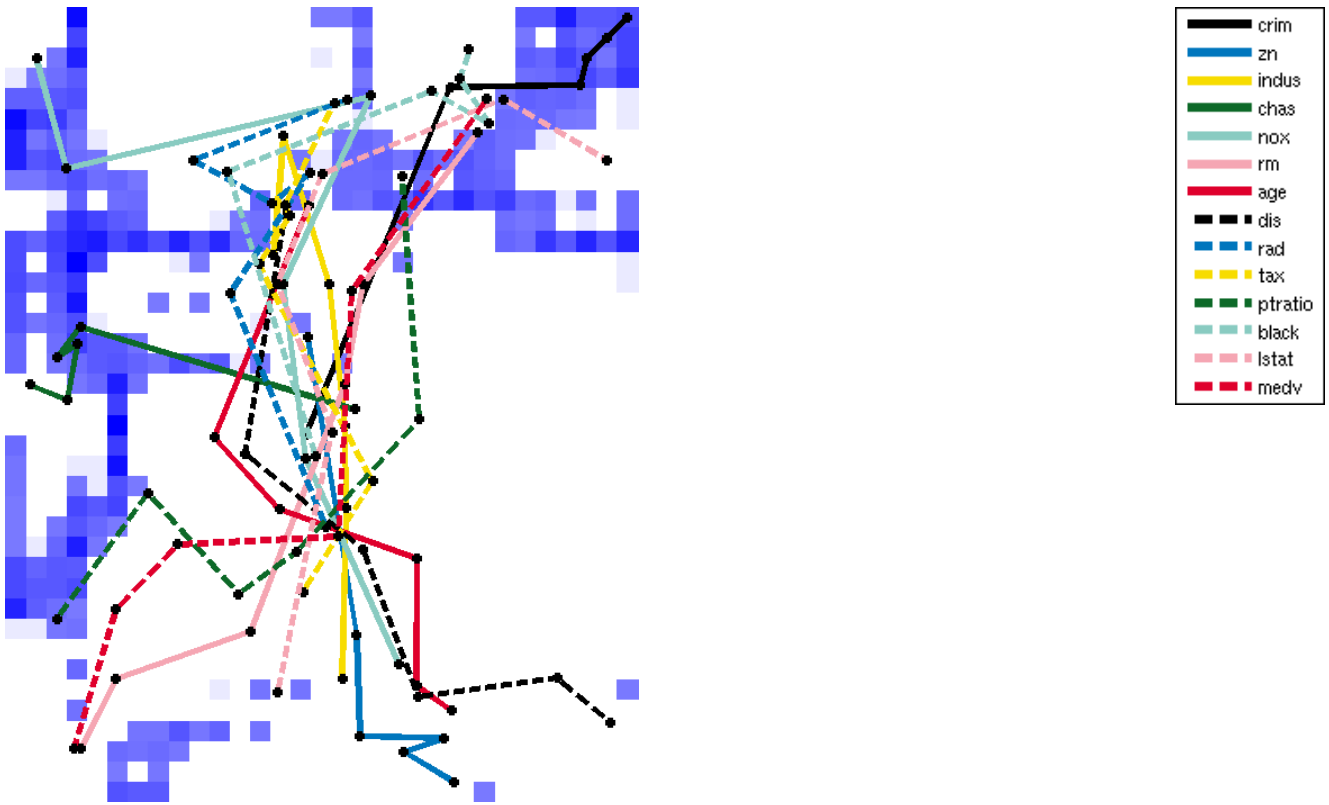


Figure: discretisation is performed for all components

snapped component lines

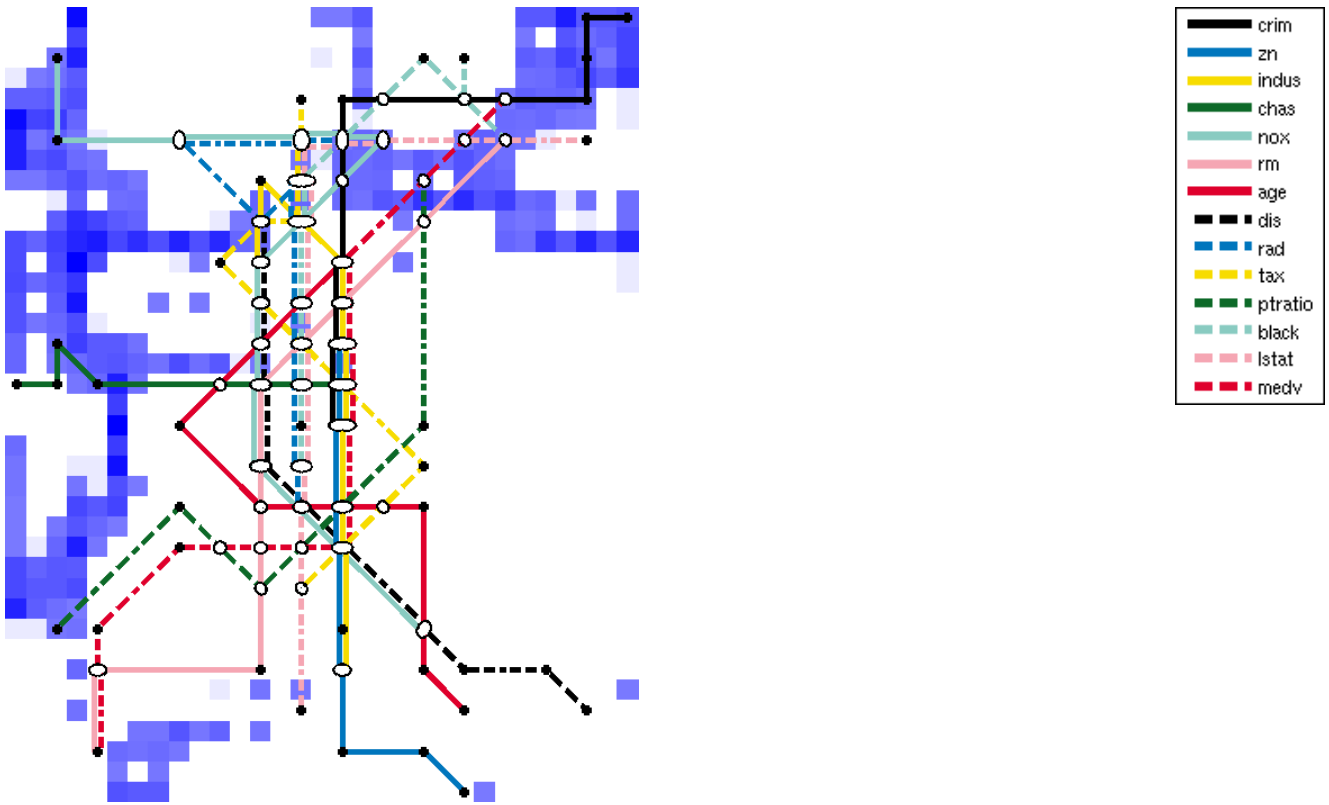


Figure: component lines are snapped onto the som grid

aggregated component lines

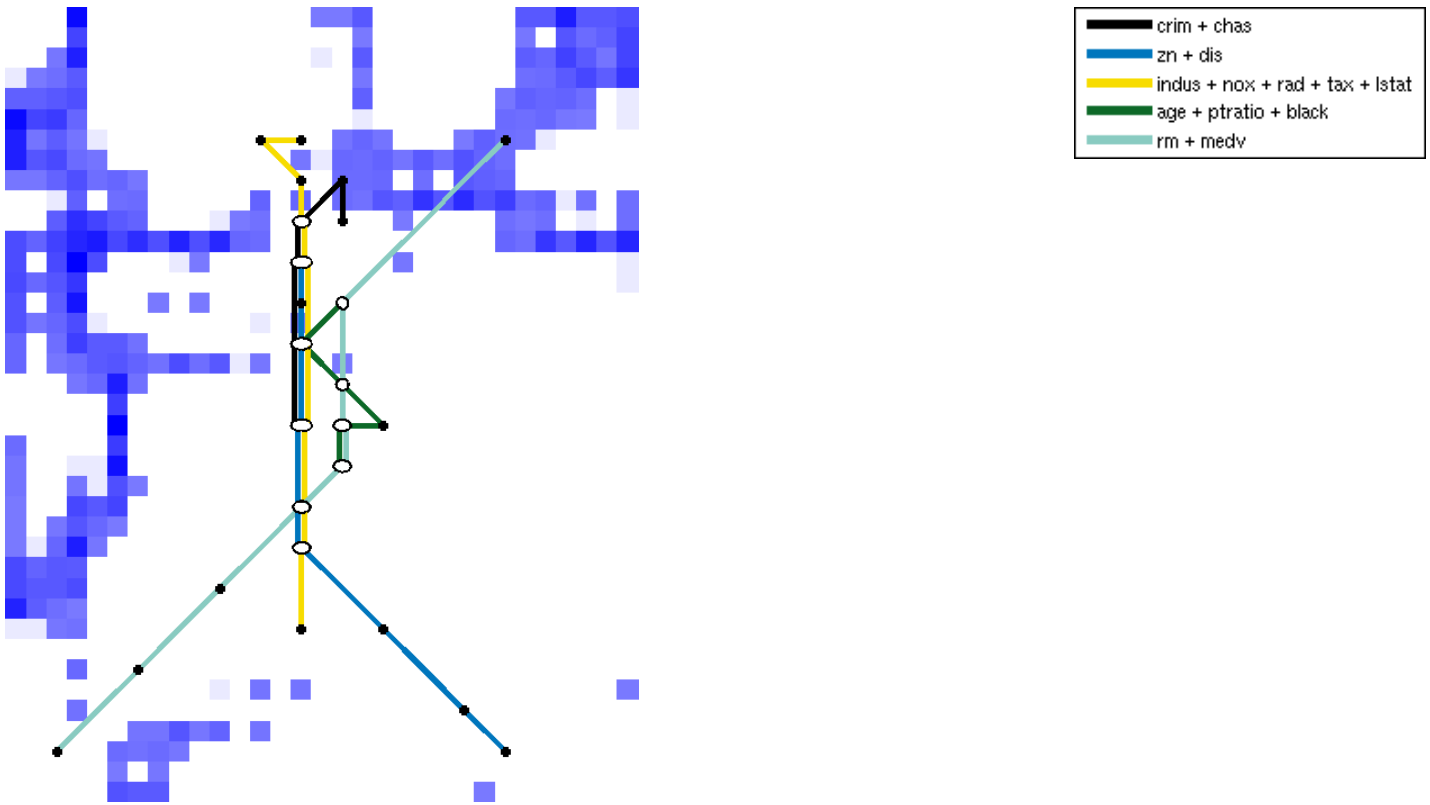


Figure: aggregation of similar lines according to line distance

selected components

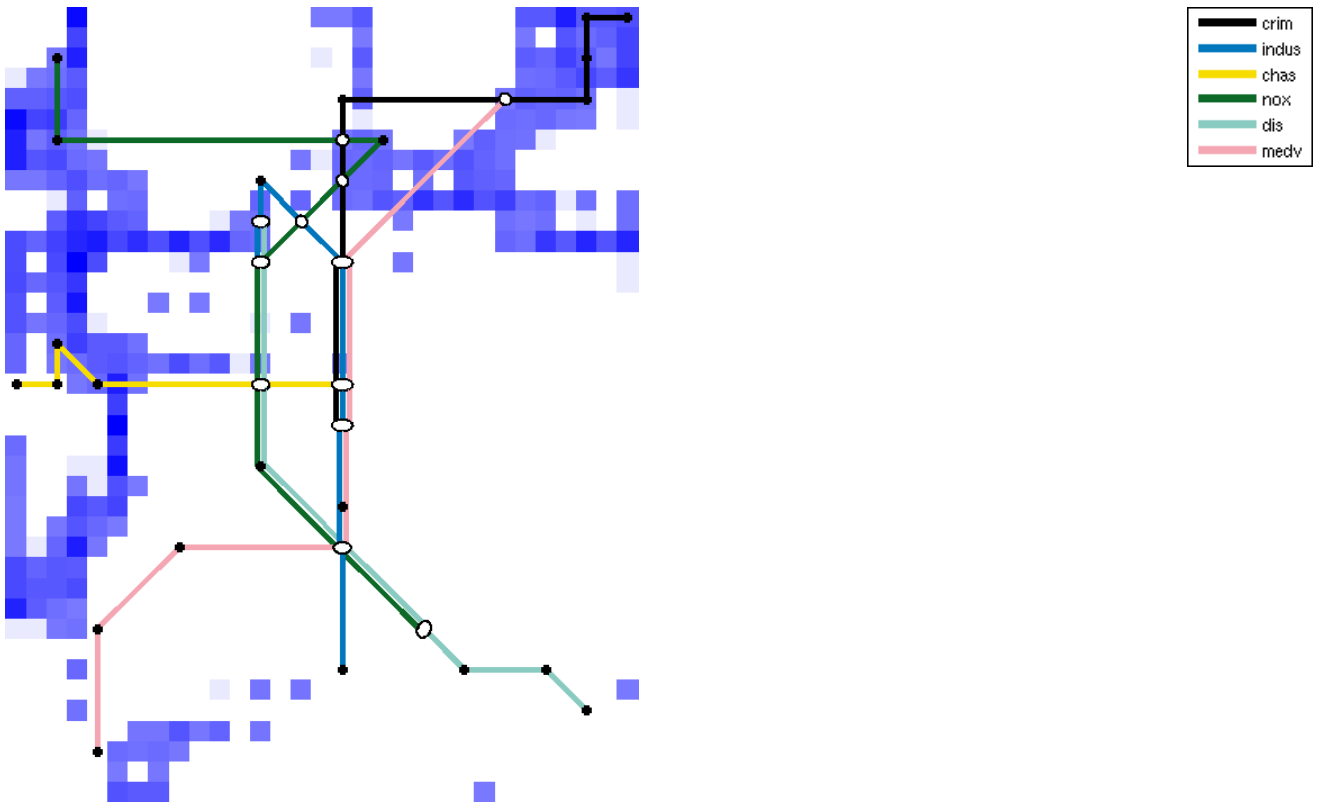


Figure: only components selected by scattering measures are visualised

selected and aggregated components

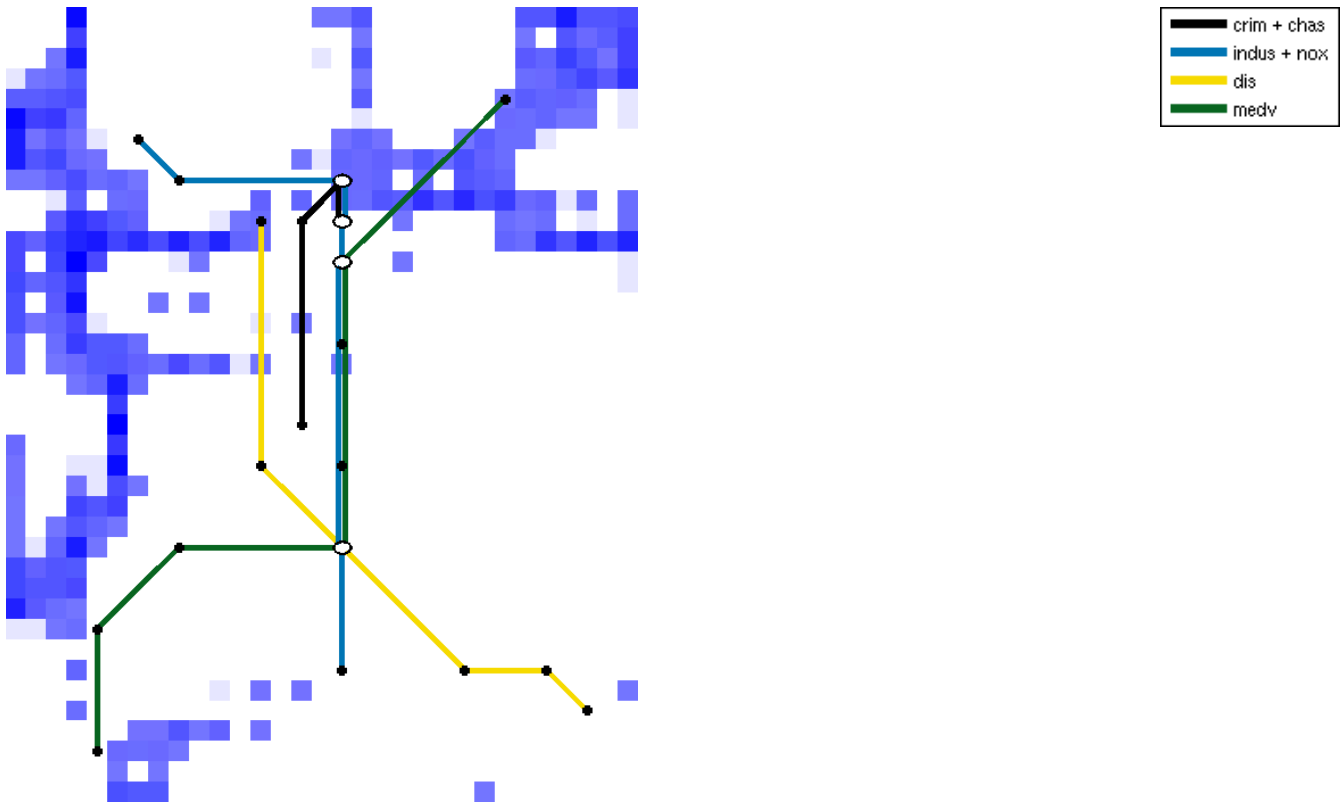


Figure: only selected components are aggregated and subsequently visualised

. . . we've come a long way

from 14 component planes plots

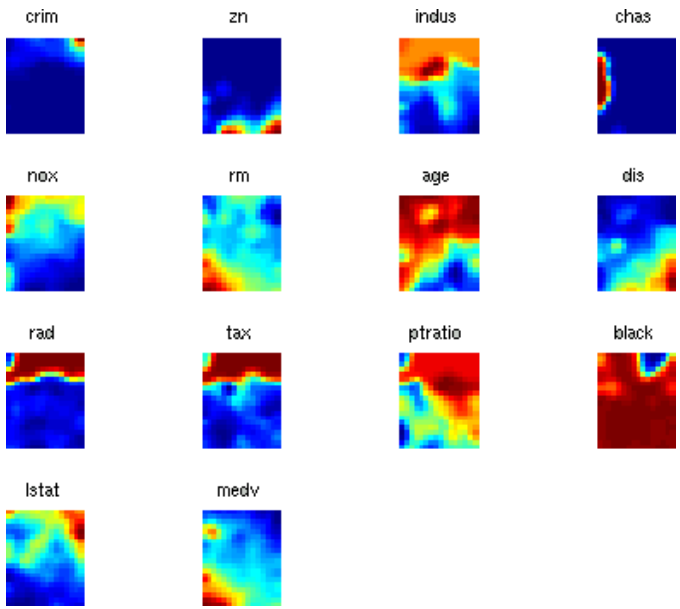


Figure: component planes visualisation of all variables

. . . to a single, slim visualisation

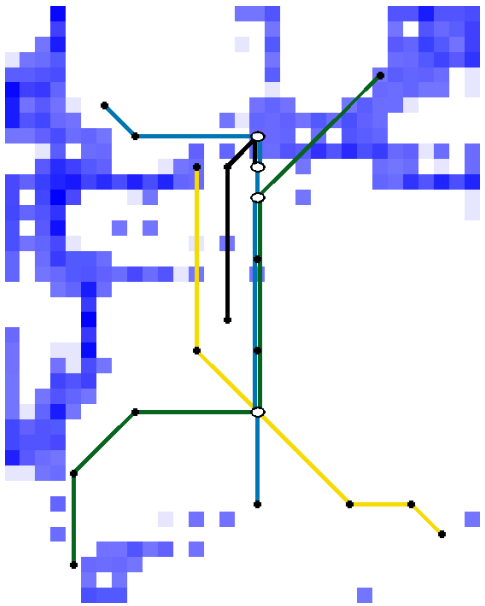


Figure: only selected and aggregated components are visualised

recap

plotting of component planes in one single illustration

visualisation of correlations between components

aggregation of highly correlated components

overlaying existing som visualisations

things to do and see

line distance functions

weighting criteria for snappings

heuristics for setting parameter values

intersections as som quality criteria

- intersections independent from clusterings and initialisation?

more infos

http://www.ifs.tuwien.ac.at/~neumayer

http://www.ifs.tuwien.ac.at/dm