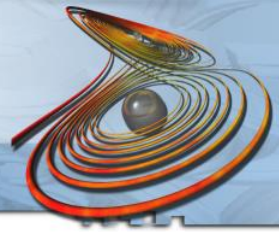


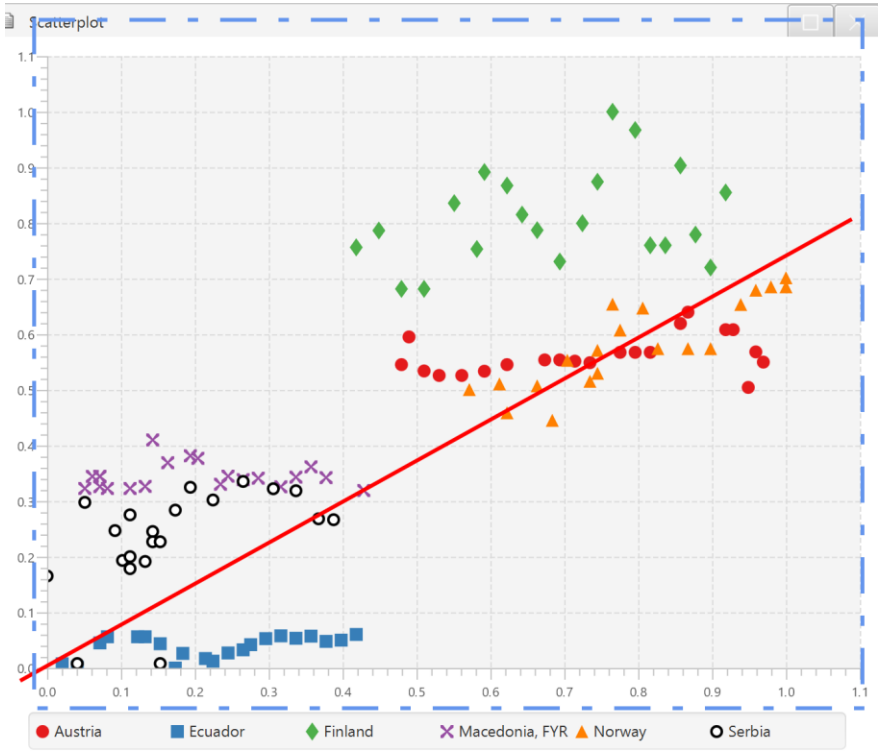
Interactive Visual Exploration of Local Patterns in Large Scatterplot Spaces

Mohammad Chegini*, Lin Shao*, Robert Gregor*,
Dirk Joachim Lehmann§, Keith Andrews*, Tobias Schreck*

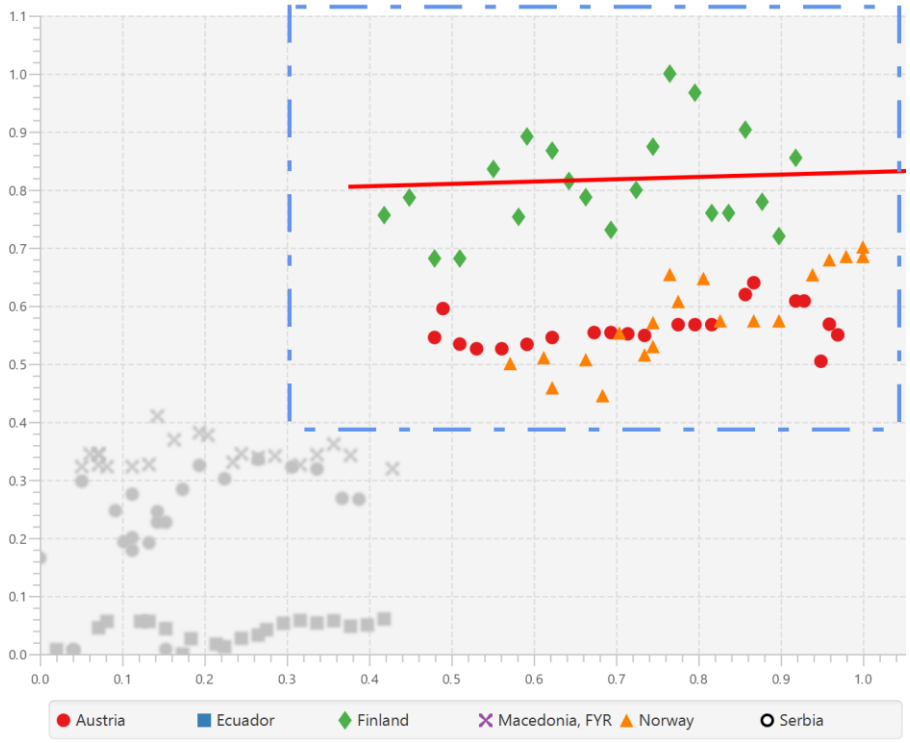
* Graz University of Technology, Austria
§ University of Magdeburg, Germany

June 5, 2018

What is a pattern?



A global pattern.



Local pattern (Also a query).

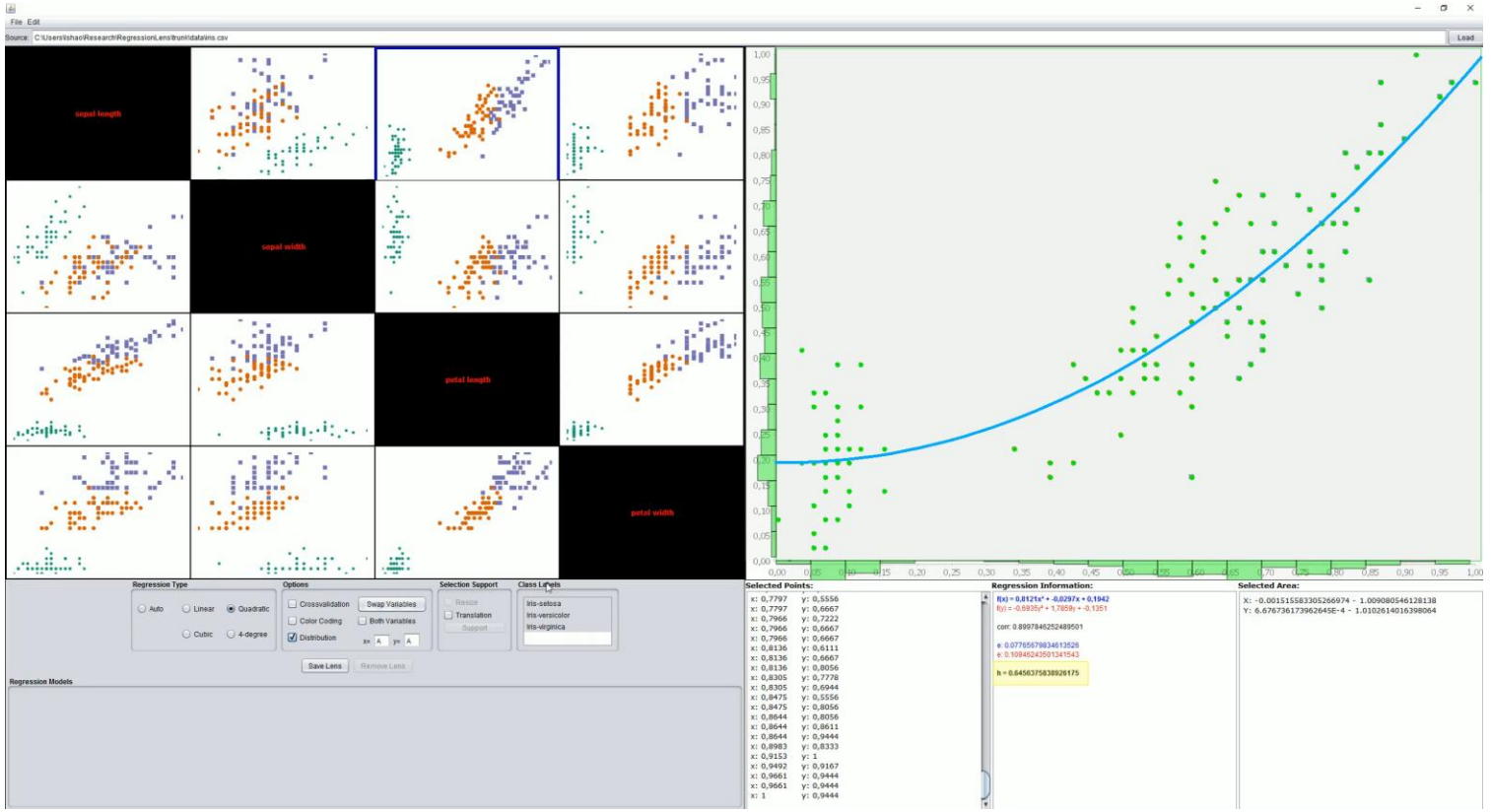
Motivation

- Hidden information in local patterns.
- Finding redundant attributes.
- Finding local relationship in data.

Overview of the System and Dataset

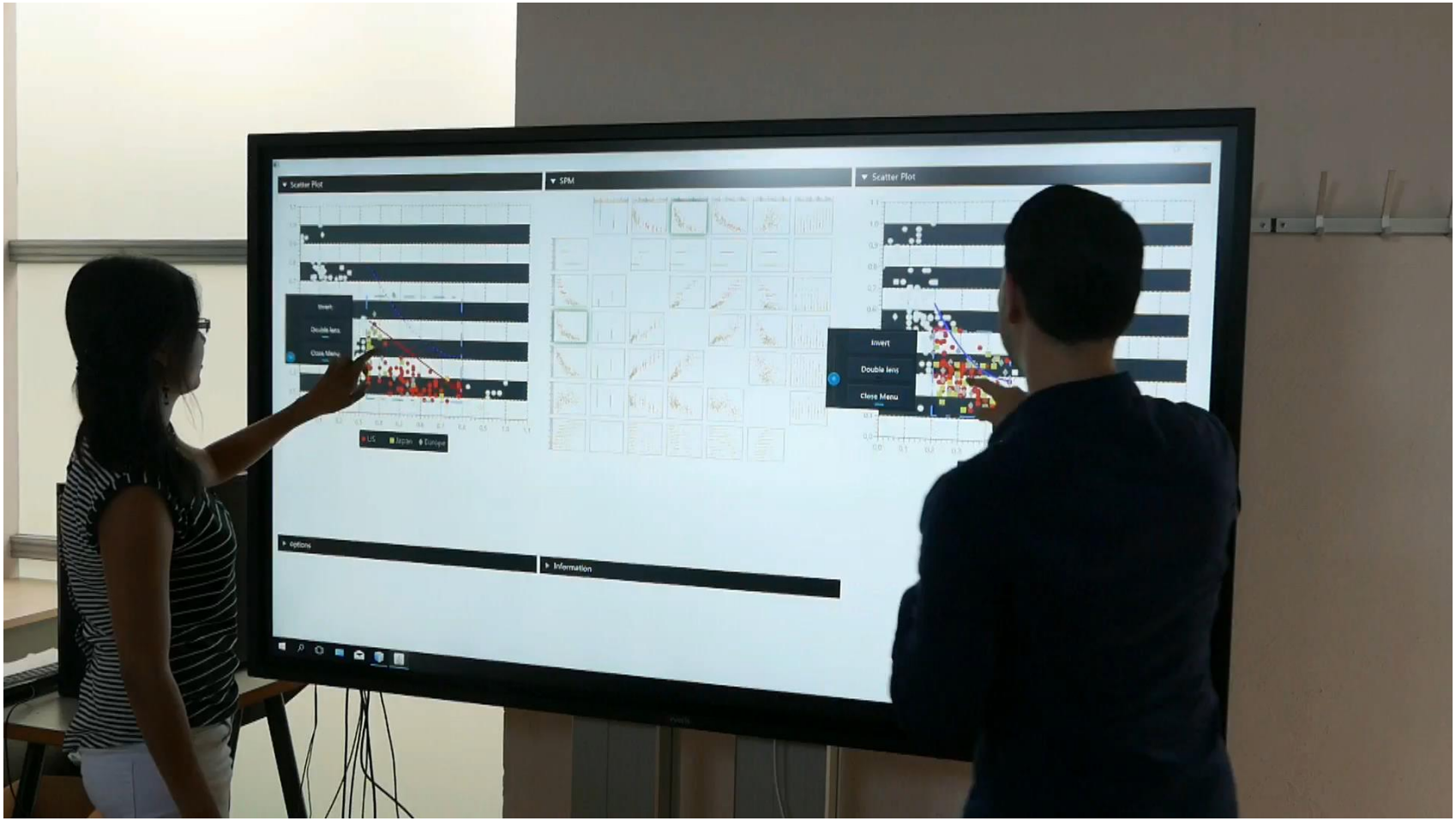
Renewable energy	GDP	Unemployment	class	year
25.13	3.55	3.11	AT	1991
24.47	2.41	3.31	AT	1992
26.13	0.97	3.47	AT	1993
26.35	0.29	4.09	AT	1994

Regression Lens [1]



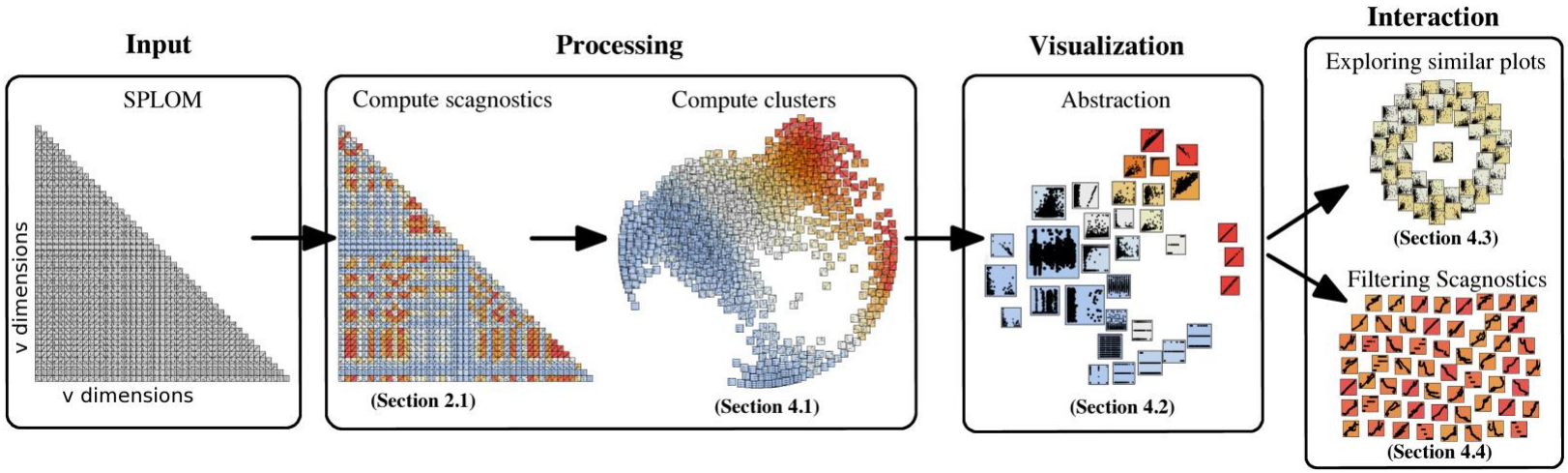
[1] Shao, Lin, et al. "Interactive Regression Lens for Exploring Scatter Plots." *Computer Graphics Forum*. Vol. 36. No. 3. 2017.

Collaborative Scatterplot Analysis [2]



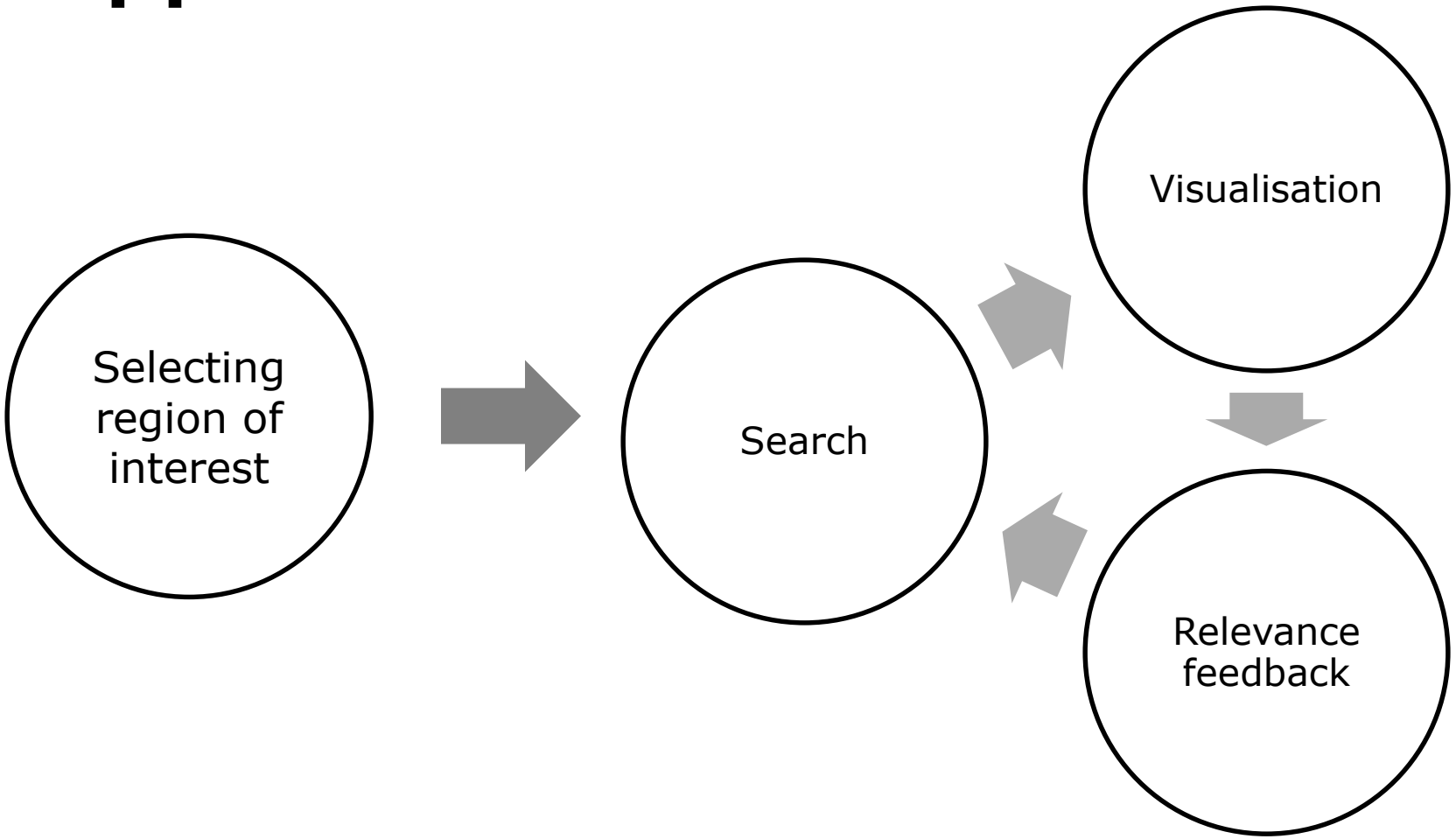
[2] Mohammad Chegini, et al. "Interaction Concepts for Collaborative Visual Analysis of Scatterplots on Large Vertically-Mounted High-Resolution Multi-Touch Displays." Forum Media Technology & All Around Audio Symposium. 2017.

ScagExplorer [3]



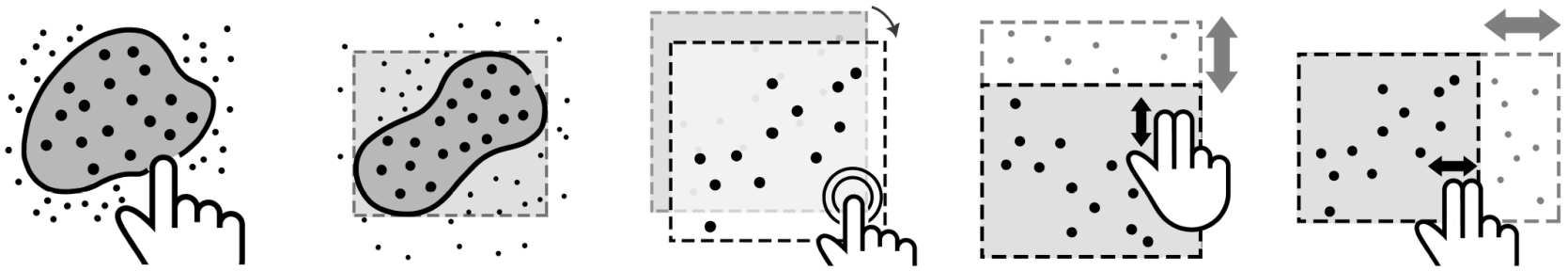
[3] Dang, Tuan Nhon, and Leland Wilkinson. "Scagexplorer: Exploring scatterplots by their scagnostics." Visualization Symposium (PacificVis), 2014 IEEE Pacific. IEEE, 2014.

Approach


Selecting region of interest
Search
Visualisation
Relevance feedback

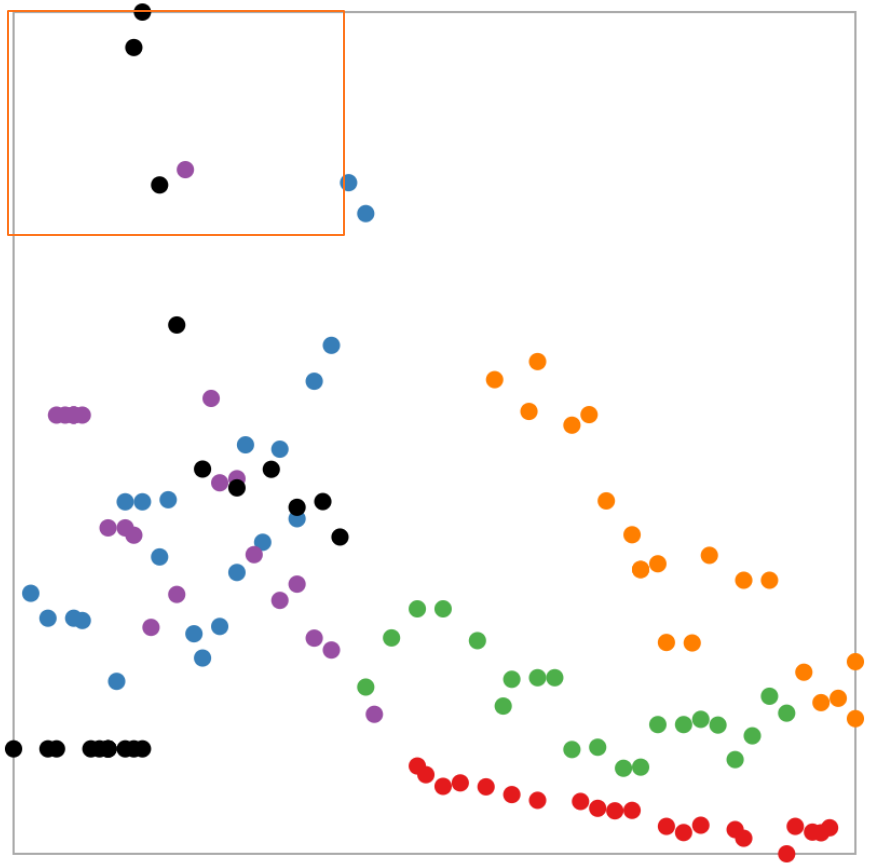
Initializing the Query (Step 1)

- Selecting a scatterplot from SPLOM.
- The user selects a region of interest inside a scatterplot as a query.

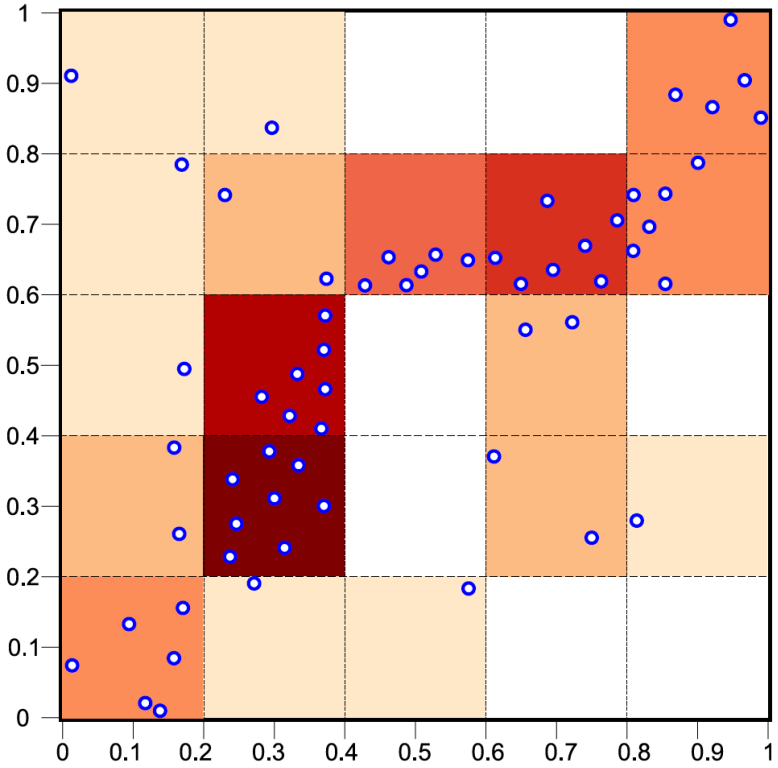


Sliding Window Approach (Step 2)

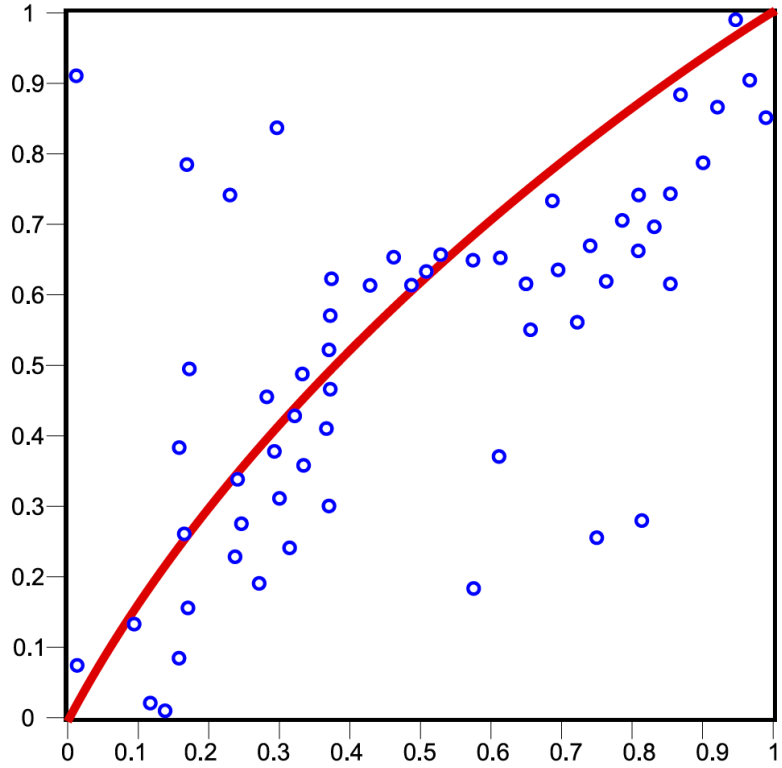
- Step size.
- Scale size.



Shape and Model-based Descriptors (Step 2)

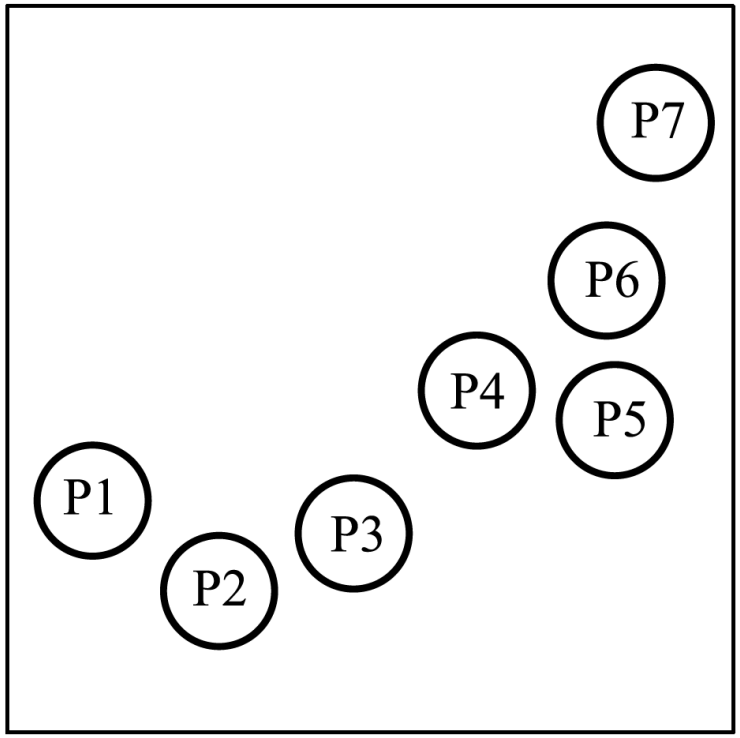


Shape-based descriptor.

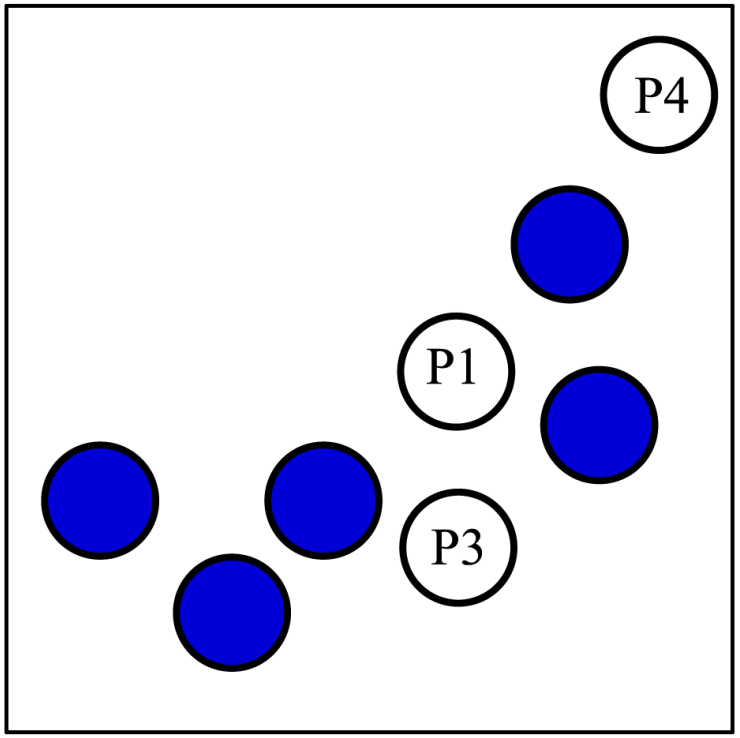


Model-based descriptor.

Purity Scores (Step 2)



User defined query pattern.

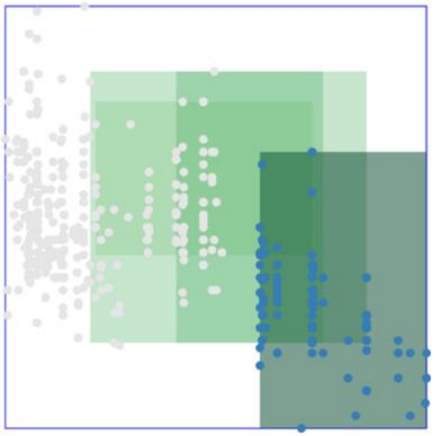


A matching pattern.

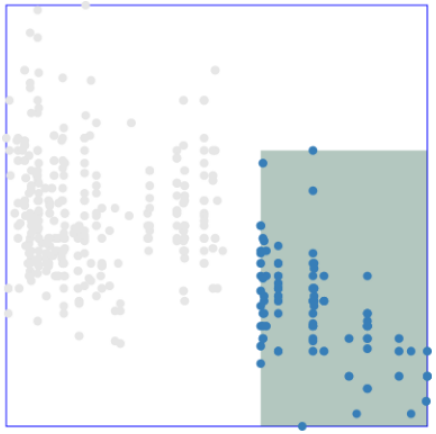
Ranking Algorithm (Step 2)

- Create a ranking of all patterns in the SPLOM, according to similarity with the query pattern.
- The ranking depends on the chosen weights for the descriptors.

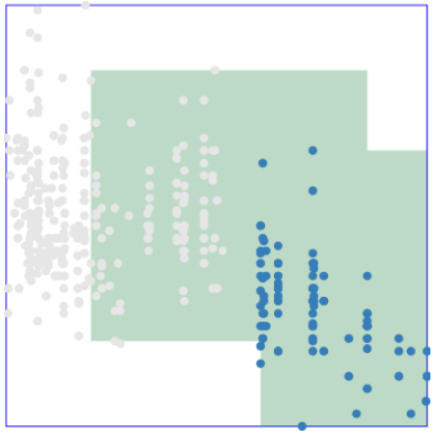
Visualisation of Patterns (Step 3)



Aggregation



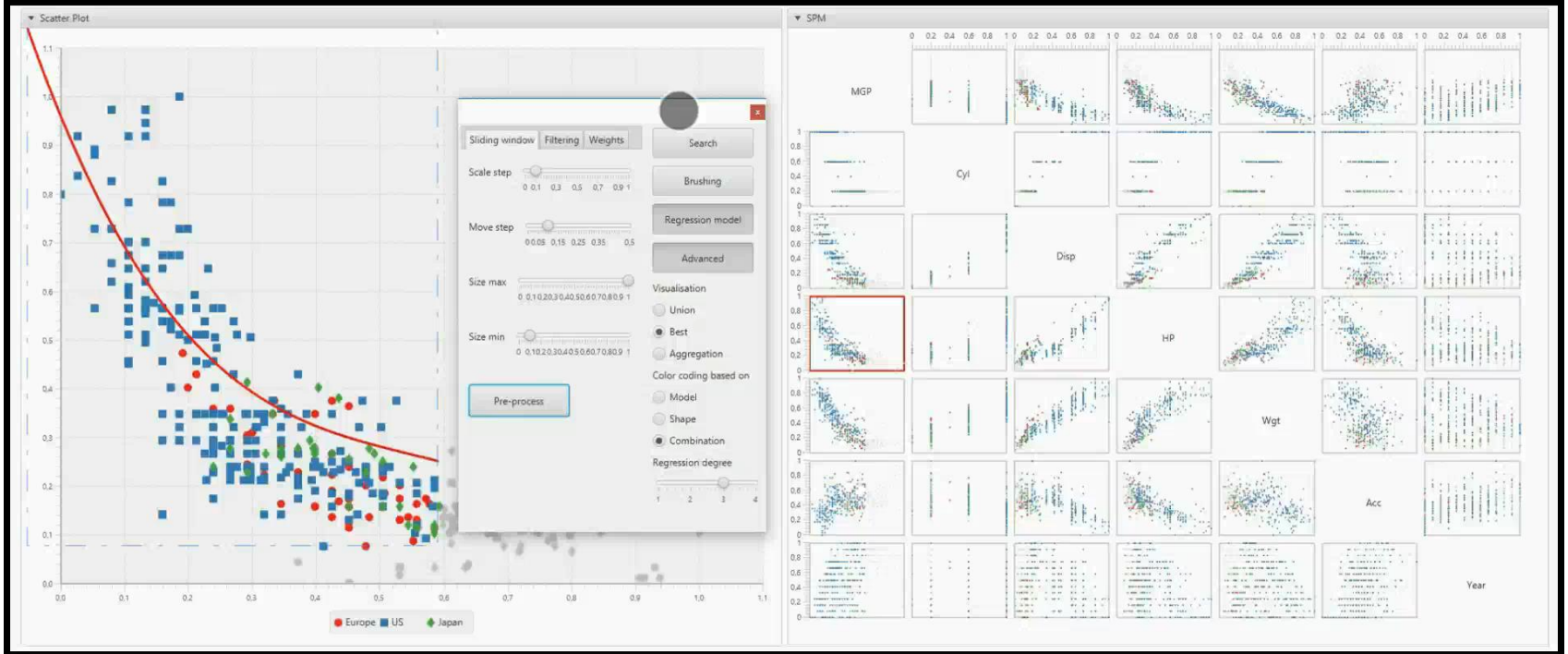
Best-Matches



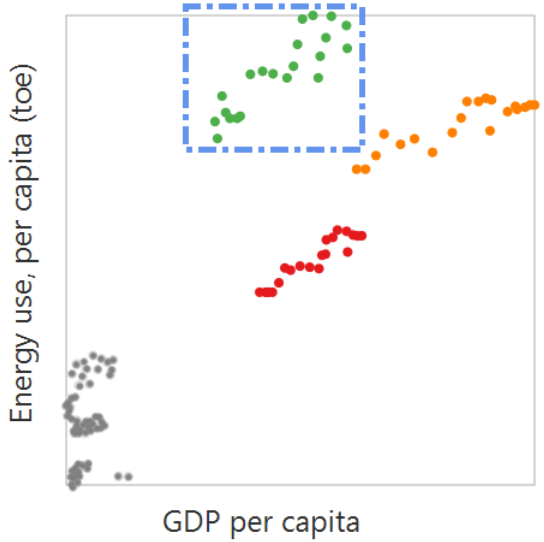
Union

Relevance Feedback (Step 4)

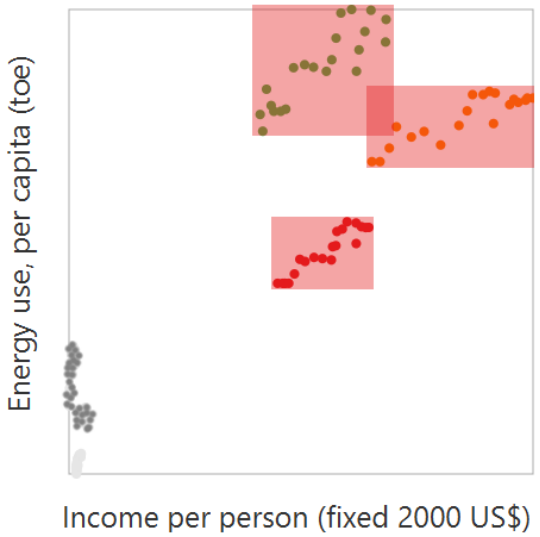
- The user can select related local patterns from the highlighted matches in the SPLOM.
- Based on the patterns marked as relevant by the user, the system adapts the parameters (weights, thresholds, step and scale sizes).
- The system searches again with the new parameters.


Scatter Plot
Sliding window
Filtering
Weights
Search
Brushing
Regression model
Advanced
Scale step
Move step
Size max
Size min
Visualisation
Union
Best
Aggregation
Color coding based on
Model
Shape
Combination
Regression degree
Pre-process
Europe
US
Japan
SPM
MGP
Cyl
Disp
HP
Wgt
Acc
Year

Use Case 1: Finding Similar Patterns

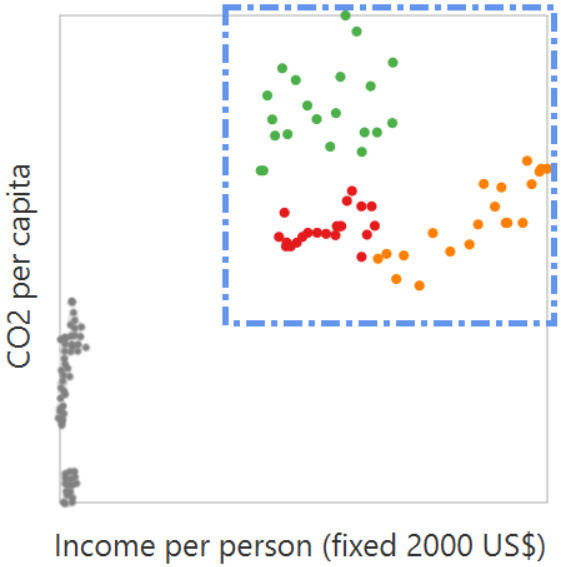


Original query pattern.

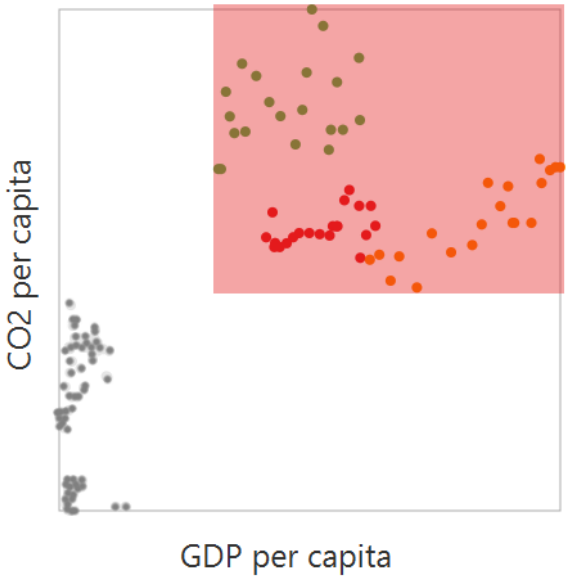


Similar patterns to the query.

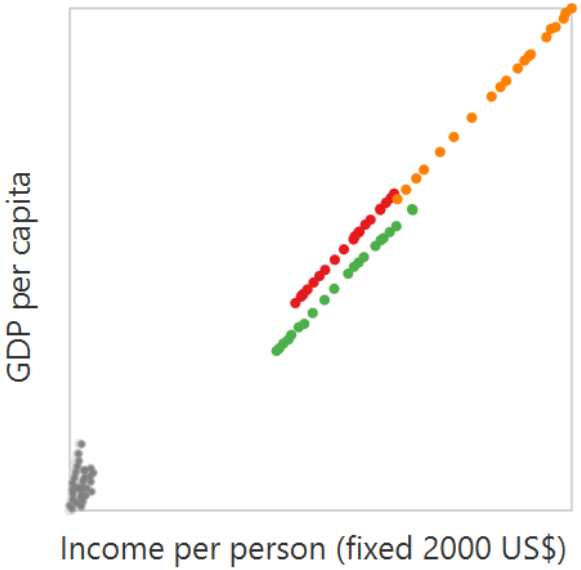
Use Case 2: Finding Positive Correlations



Original query pattern

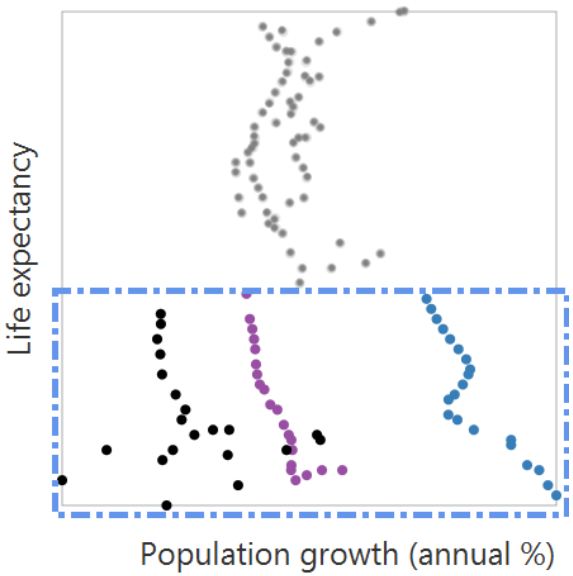


A similar pattern

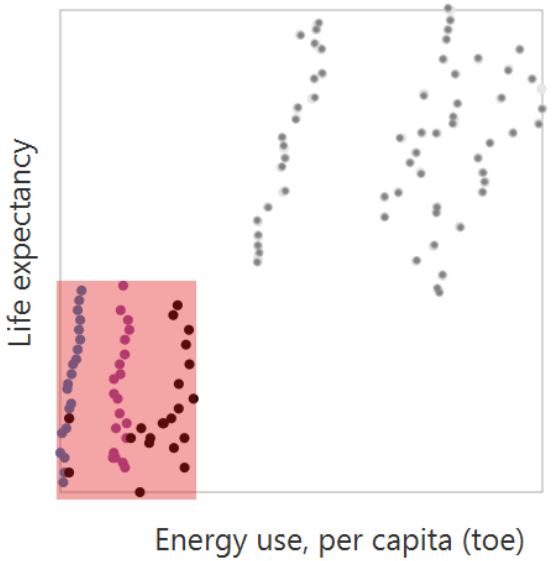


A local positive correlation

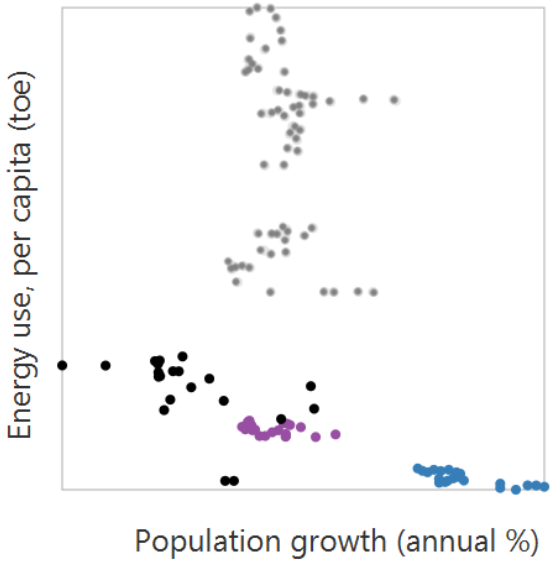
Use Case 3: Finding Negative Correlations



Original query pattern



A similar pattern

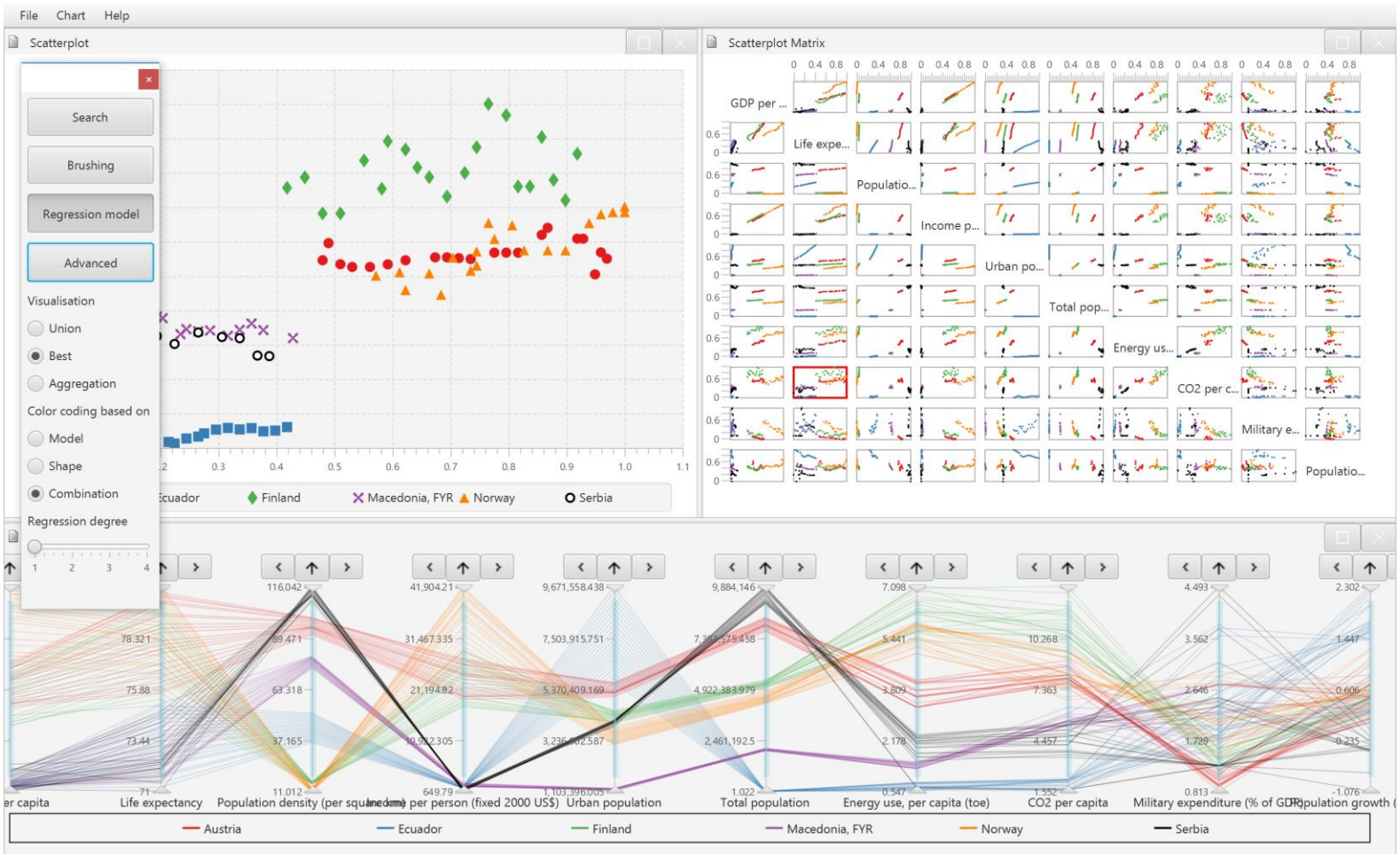


A local negative correlation

Future work

- Including more than two dimensions.
- Using production dataset.
- Adding other modalities for interaction.
- Conducting user studies.
- Considering other descriptors.

Questions?


File
Chart
Help
Scatterplot
Search
Brushing
Regression model
Advanced
Visualisation
Union
Best
Aggregation
Color coding based on
Model
Shape
Combination
Regression degree
Finland
Macedonia, FYR
Norway
Serbia
Scatterplot Matrix
GDP per ...
Life expe...
Populatio...
Income p...
Urban po...
Total pop...
Energy us...
CO2 per c...
Military e...
Populatio...
Life expectancy
Population density (per square km)
Urban population
Total population
Energy use, per capita (toe)
CO2 per capita
Austria
Ecuador
Finland
Macedonia, FYR
Norway
Serbia

Extra Slide: Ranking Formula

$$s(q,p) = \begin{cases} 0 & \text{if } P_{precision} < P_{pmin} \text{ or } P_{recall} < P_{rmin} \\ w_m(1-d_m) + s(1-w_m)(1-d_s), & \text{otherwise} \end{cases}$$

$$\underset{C_k}{\operatorname{argmin}} \left(\sum_{i=1}^{n} ranking_{C_k}(q, u_i) \right)$$

Extra Slide: Scalability

- Step and scale size for sliding-window (385).
- Number of dimensions (10 -> 34650).
- Number of descriptors (34650 x 3136 x t).