Visualizing multi-dimensional spaces

Thomas Torsney-Weir

About me

About me

- Born:Allentown, PA

About me

- Born: Allentown, PA
- Undergrad: Georgetown

About me

- Born: Allentown, PA
- Undergrad: Georgetown
- Finance: NYC

About me

- Born: Allentown, PA
- Undergrad: Georgetown
- Finance: NYC
- Master’s: Simon Fraser

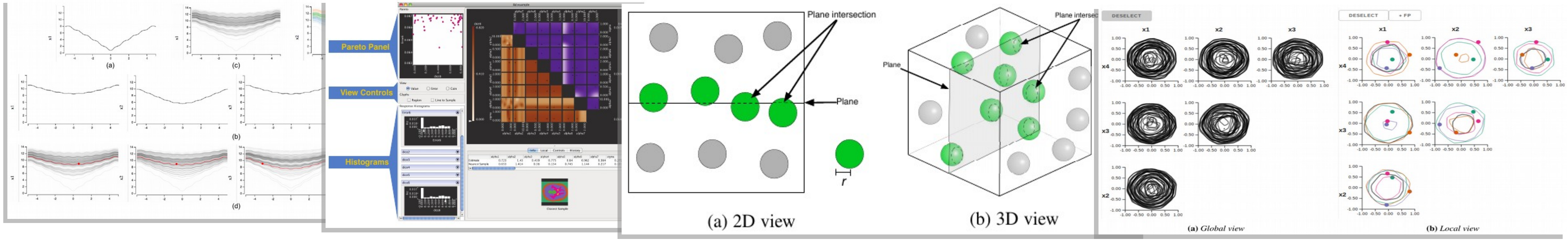
About me

- Born: Allentown, PA
- Undergrad: Georgetown
- Finance: NYC
- Master's: Simon Fraser
- PhD/Postdoc: University of Vienna

About me

- Born: Allentown, PA
- Undergrad: Georgetown
- Finance: NYC
- Master’s: Simon Fraser
- PhD/Postdoc: University of Vienna
- Now: Swansea University

My research

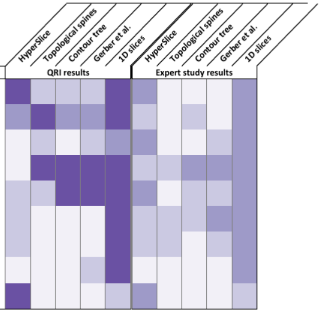


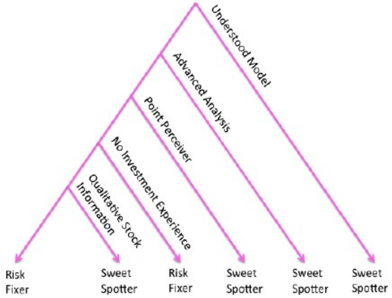
Multi-dimensional spaces

API design

Users/tasks

Task	Task description for discrete data items from [AES05]	Our adaption to continuous scalar functions
Retrieve value	"Given a set of specific cases, find attributes of those cases"	Given an x, what is the function value?
Filter	"Given some concrete conditions attribute values, find data cases satisfying those conditions."	For what parameter values is the function equal or over x?
Compute derived value	"Given a set of data cases, compute an aggregate numeric representation of those data cases"	Summary statistics: variance, mean, SA
Find extremum	"Find data cases possessing an extreme value of an attribute over its range within the data set"	Find local/global min/max
Determine range	"Given a set of data cases and an attribute of interest, find the span of values within the set"	What is the range of possible outputs?
Characterize distribution	"Given a set of data cases and a quantitative attribute of interest, characterize the distribution of that attribute's values over the set"	What types of shapes do the manifolds have
Find anomalies	"Identify any anomalies within a given set of data cases with respect to a given relationship or expectation, e.g. statistical outliers"	Do areas of the manifold have shapes unlike any others
Cluster	"Given a set of data cases, find clusters of similar attribute values"	Areas of the manifold have similar shapes
Correlate	"Given a set of data cases and two attributes, determine useful relationships between the values of those attributes"	1D vs 2D relationships





Multi-dimensional spaces

Outline

- Definition
- Applications
- Solutions
- Personal contributions

What is a multi-D space?

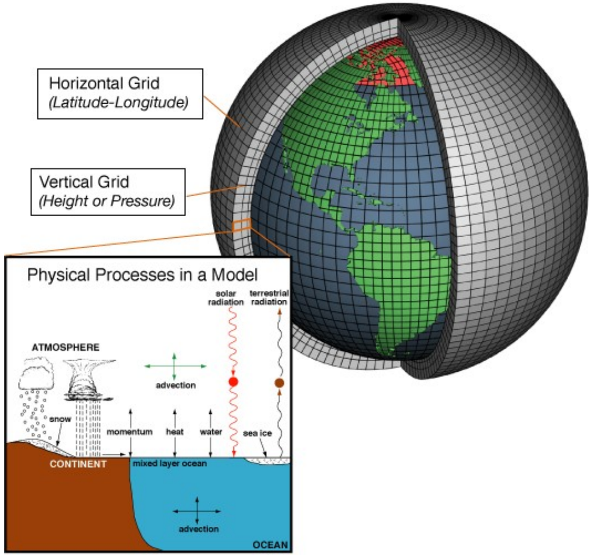
- 3-20 dimensions
- meaningful axes/dimensions
- continuous - derivatives

Application areas

- Simulations
- Regression models
- Optimization surfaces
- Multi-objective optimization

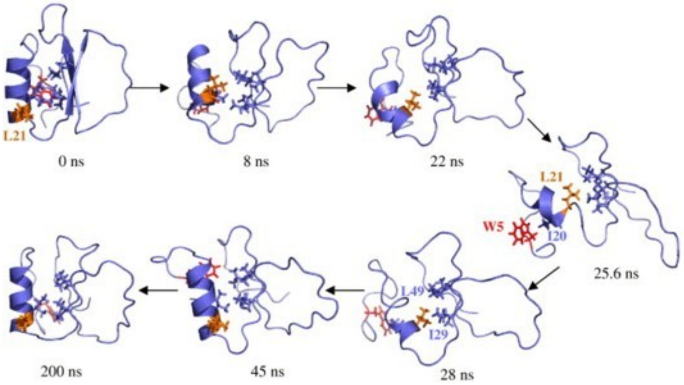
Simulations

Weather



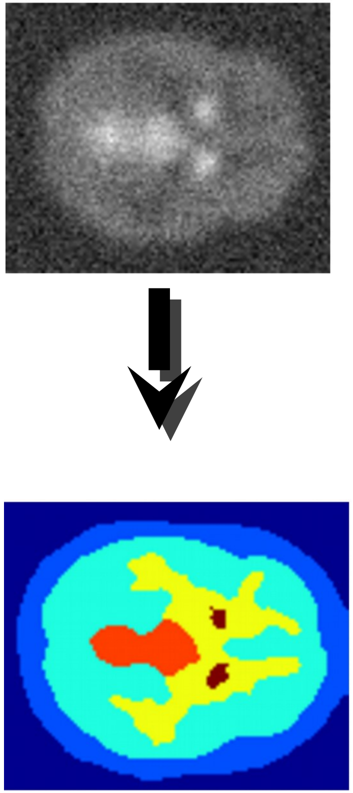
https://en.wikipedia.org/wiki/File:AtmosphericModelSchematic.png

Protein folding



http://depts.washington.edu/daglab/pom/07mar.jpg

Image segmentation



Regression models

$$f(x_1, x_2, \ldots, x_n) \to y$$

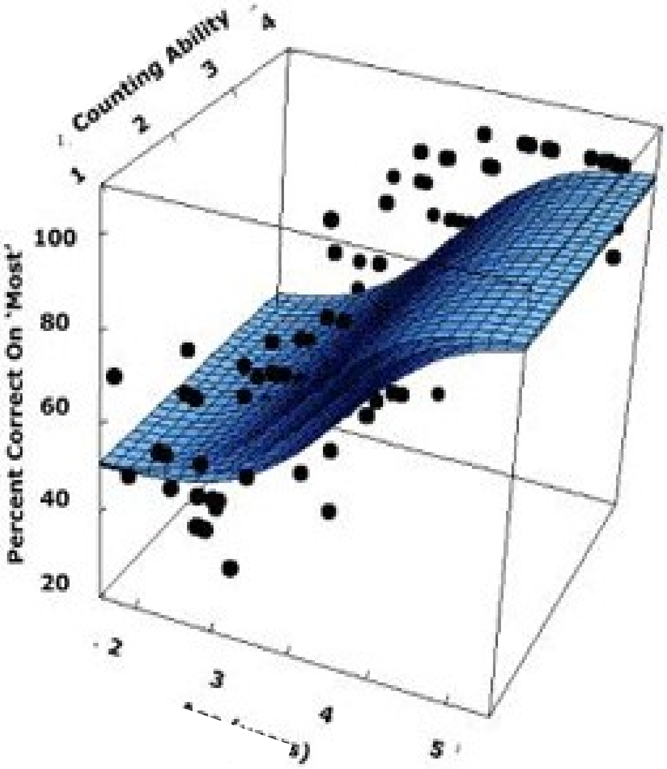
Crime rate

Rooms

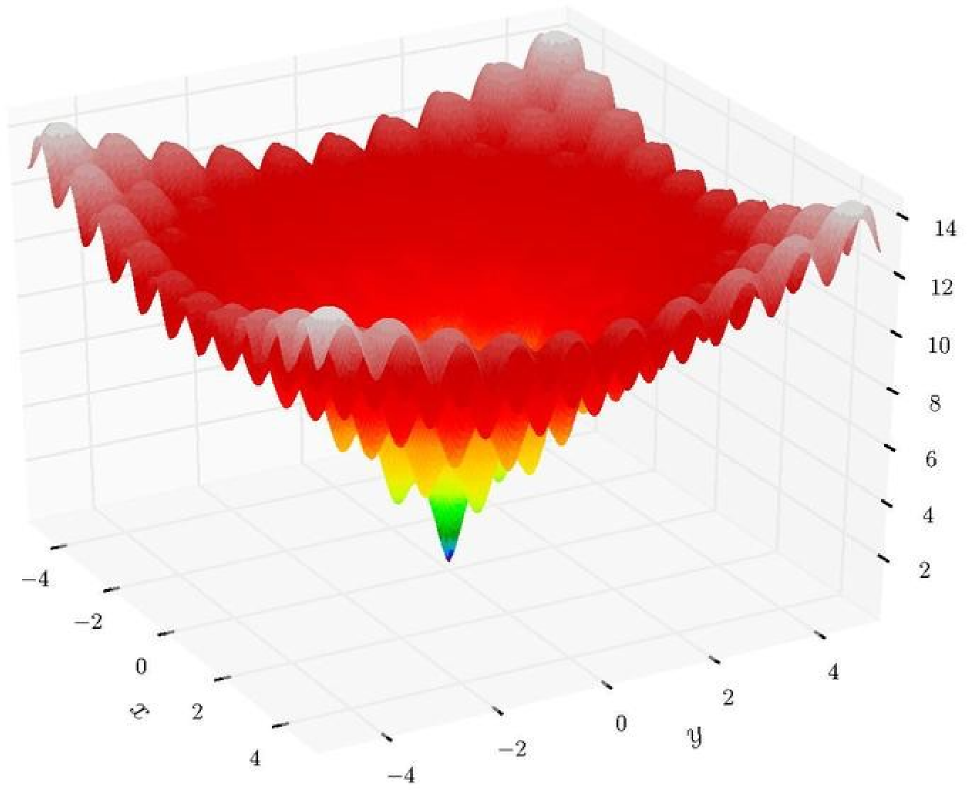
Schools

House price

Halberda et al, 2008

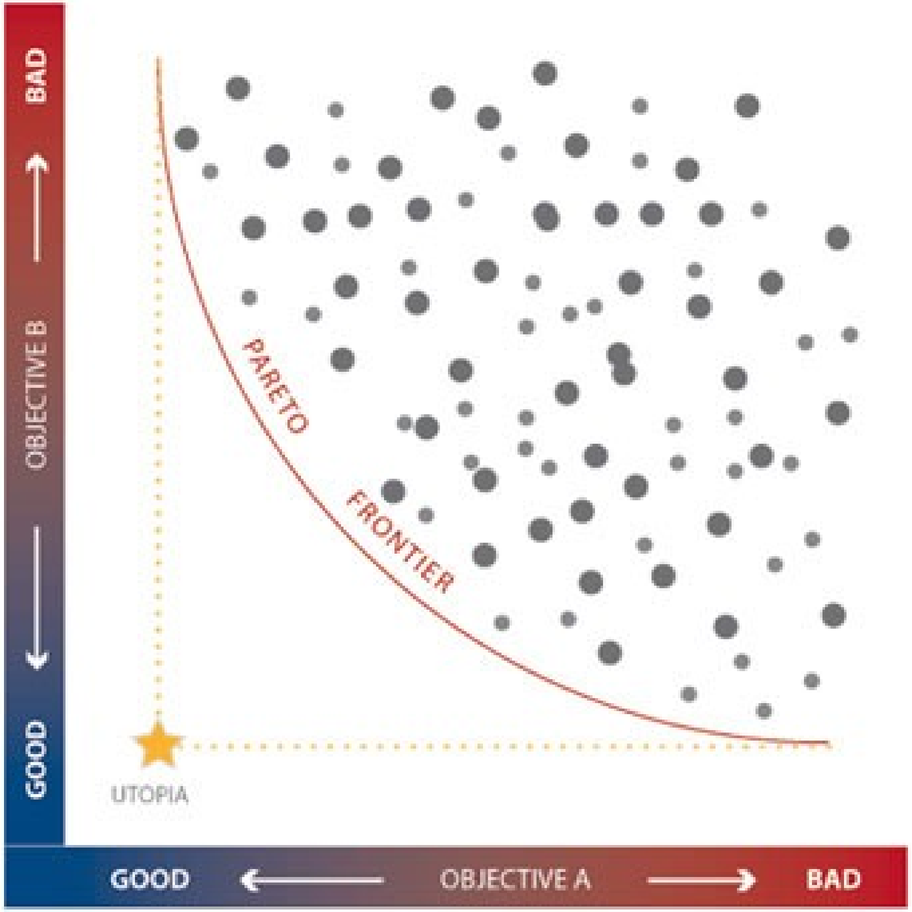


Optimization surfaces



https://en.wikipedia.org/wiki/File:Ackley%27s_function.pdf

Multi-objective optimization

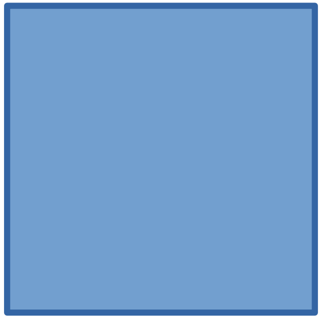


http://www.enginsoft.com/assets/img/tecnology/pido/MultiobjectiveOptimization_b.jpg

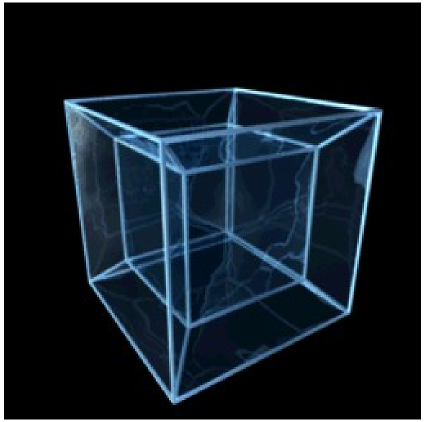
Challenges

Too many dimensions to show on screen

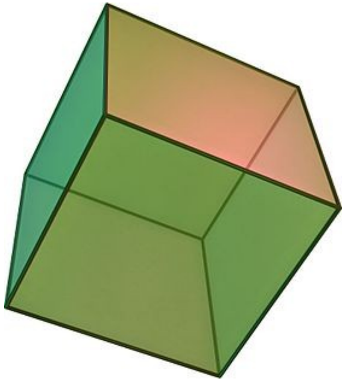
2D



4D



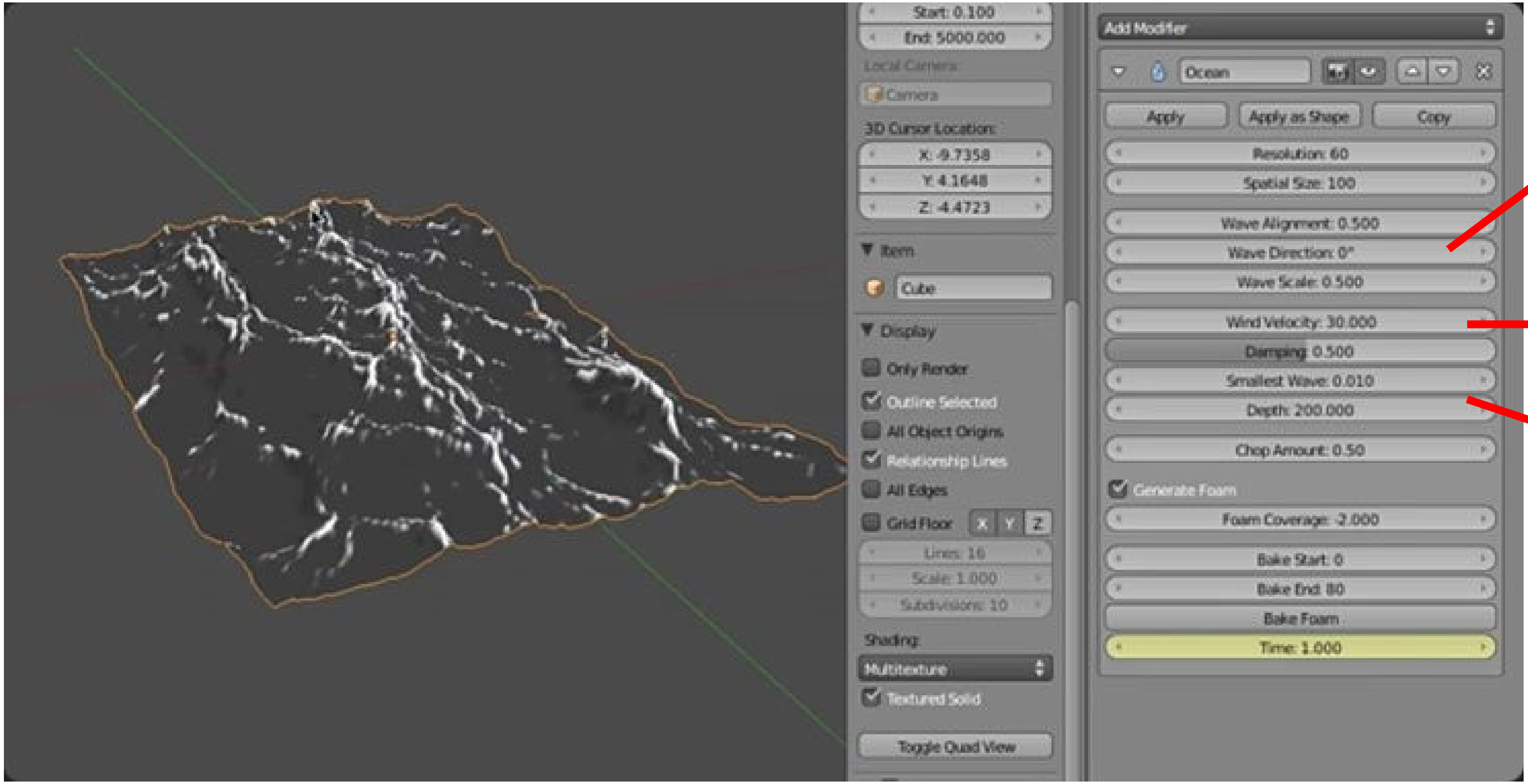
3D



>4D

Challenges

Dimensions are meaningful



Wave direction

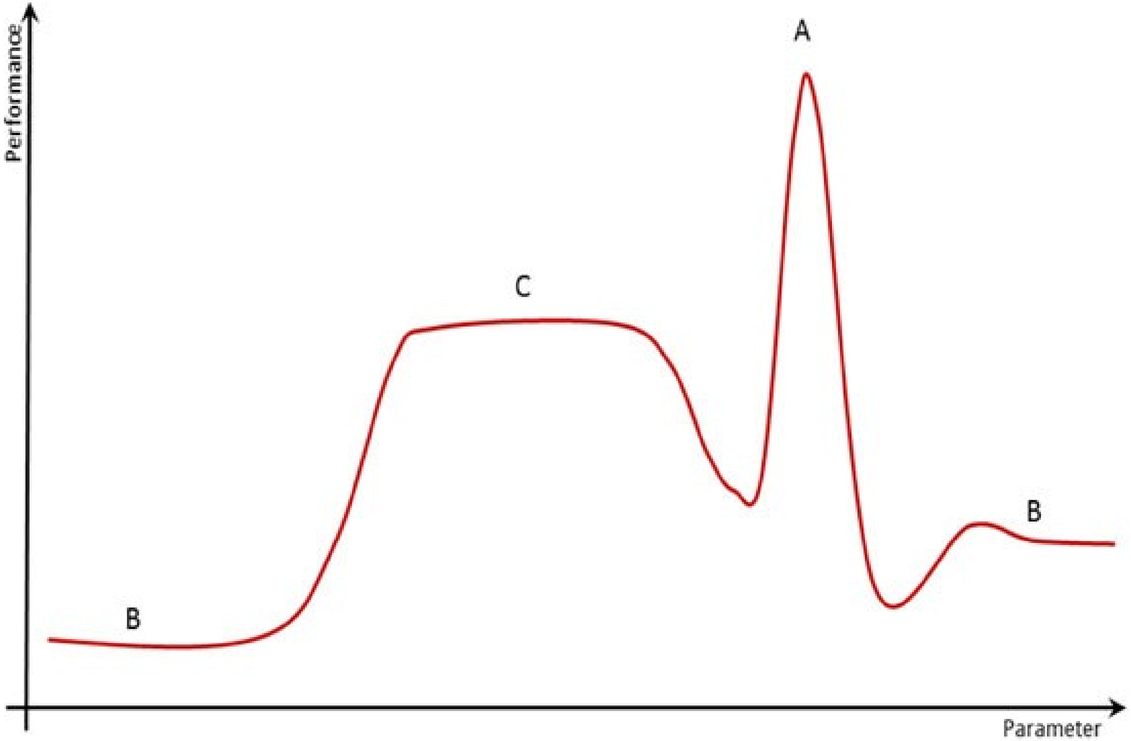
Wind velocity

Depth

http://eat3d.com/blog/metalliandy/blender-ocean-simulation-foam-accumulation

Challenges

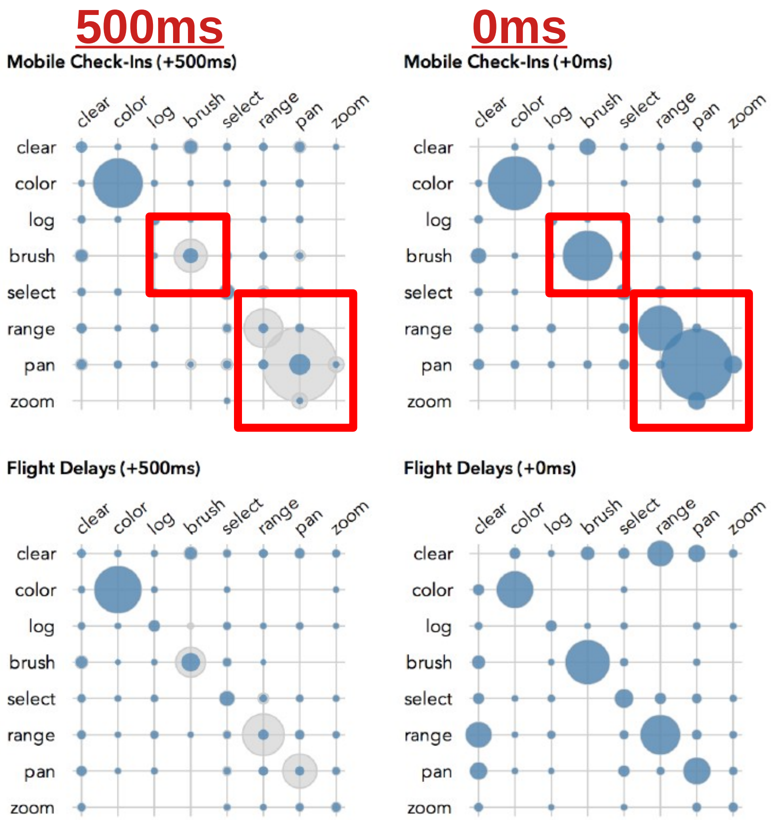
Showing changes in behavior



http://www.ansys-blog.com/wp-content/uploads/2012/02/thierryrobustdesignchart3.jpg

Challenges

Response times are important



- Delays decrease interaction
- Exploration decreases

Zhicheng Liu, Jeffrey Heer. The Effects of Interactive Latency on Exploratory Visual Analysis. IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis), 2014

Overview of solutions

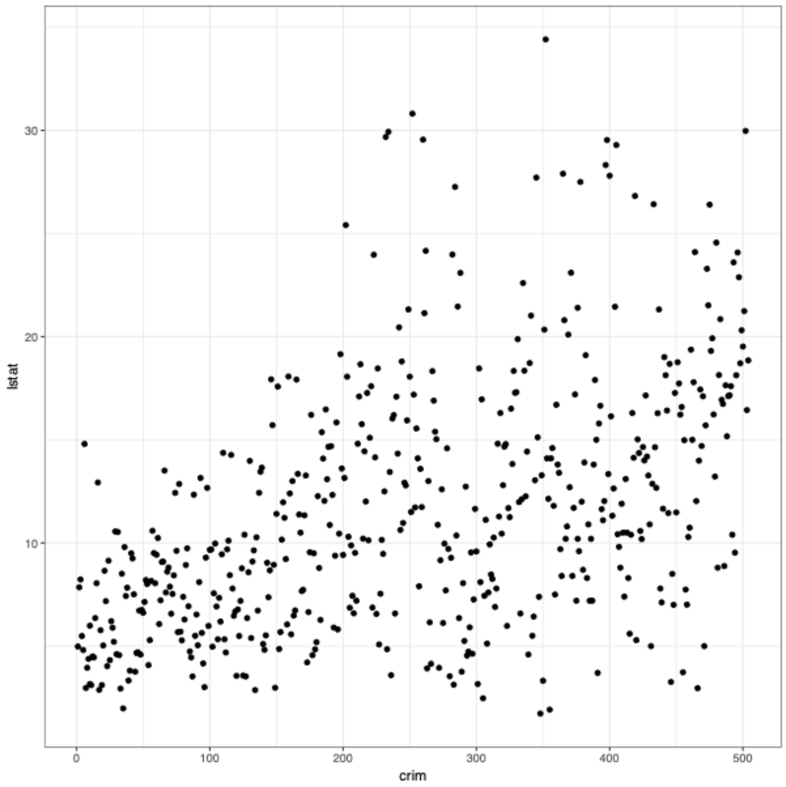
- Definition
- Applications
- Solutions
- Personal contributions

Overview

so what are approaches to solve this?

- discretization
- dimension reduction
- topology
- slicing

Discretization



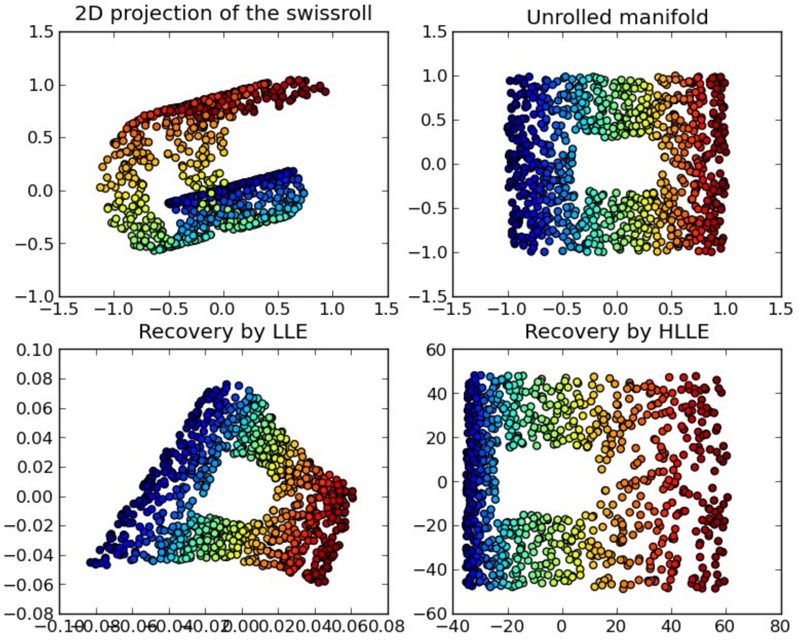
Pros:

- Many visualization techniques
- Less training

Cons:

- Connections between points lost

Dimension reduction



Pros:

- Reduce to 2D screen
- Can find patterns automatically

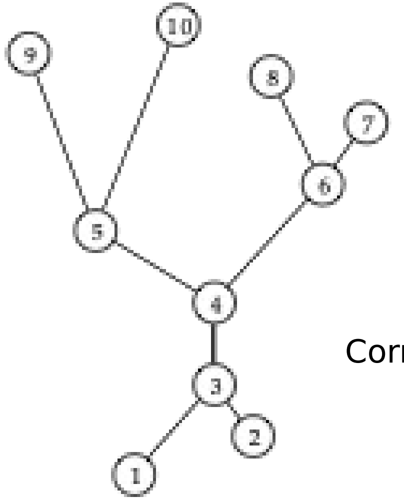
Cons:

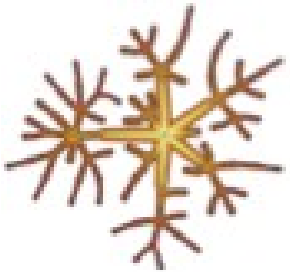
- Meaningless dimensions
- True dimensionality may be > 3

https://en.wikipedia.org/wiki/Nonlinear_dimensionality_reduction#/media/File:Lle_hlle_swissroll.png

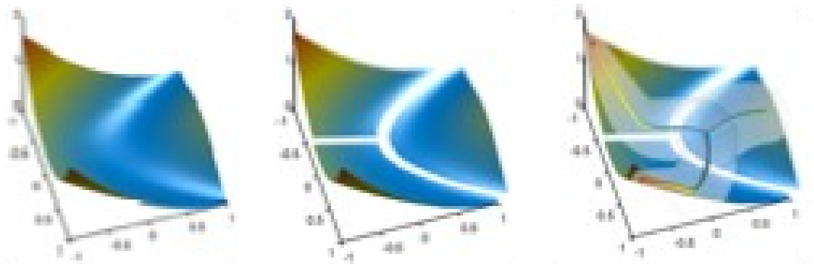
Topology

Carr, Snoeyink, and Axen 2003





Correa, Lindstrom, and Bremer 2011



Gerber et al. 2010

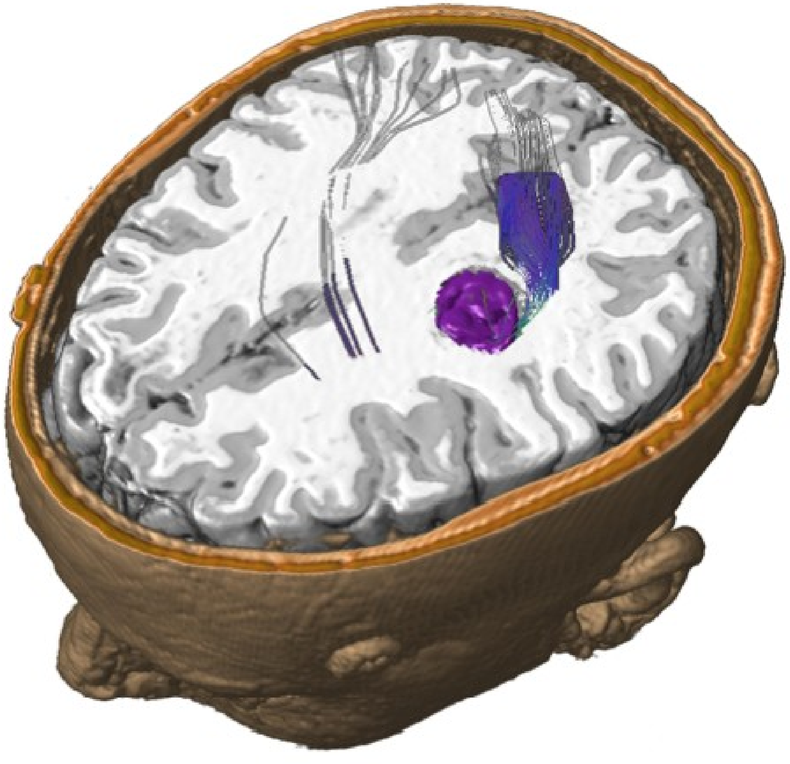
Pros:

- 2D representation of multi-D field
- Highlights key features

Cons:

- Difficult to understand
- Input space removed

Slicing



Pros:

- Reduces dimensionality
- Easy to understand metaphor

Cons:

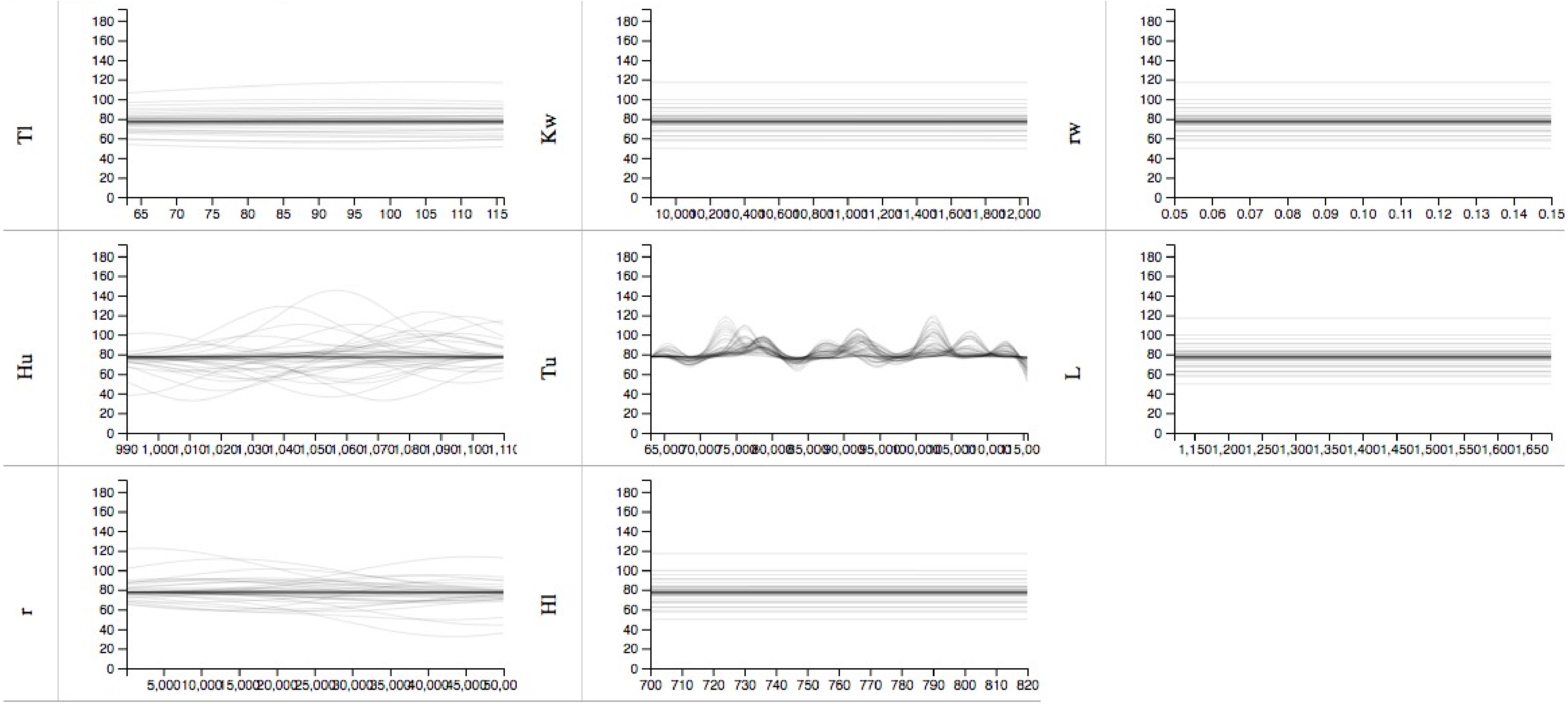
- Focus point selection important

My contributions

- Definition
- Applications
- Solutions
- Personal contributions

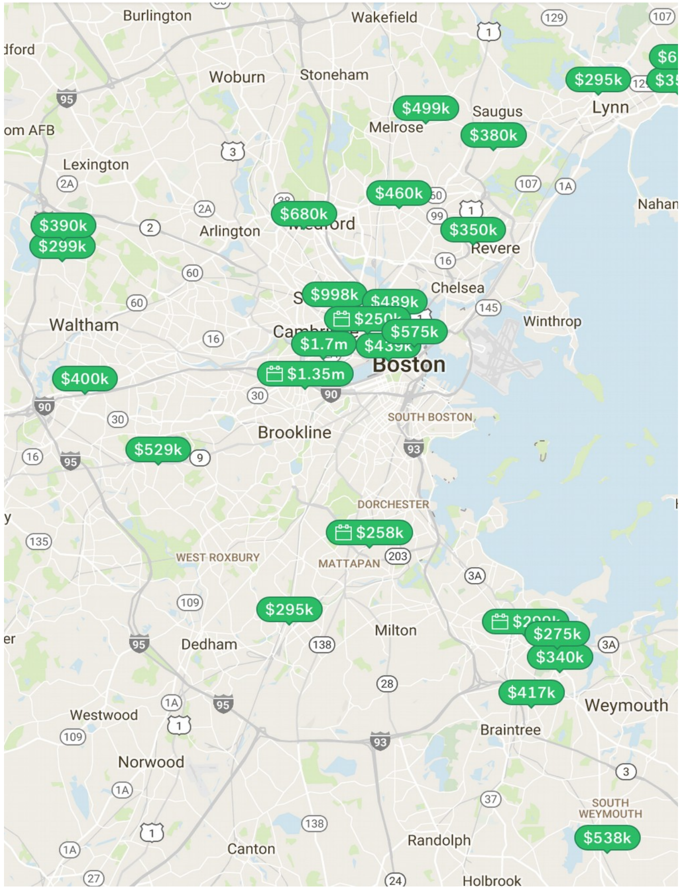
Sliceplorer

How do we look at multi-dimensional surfaces?

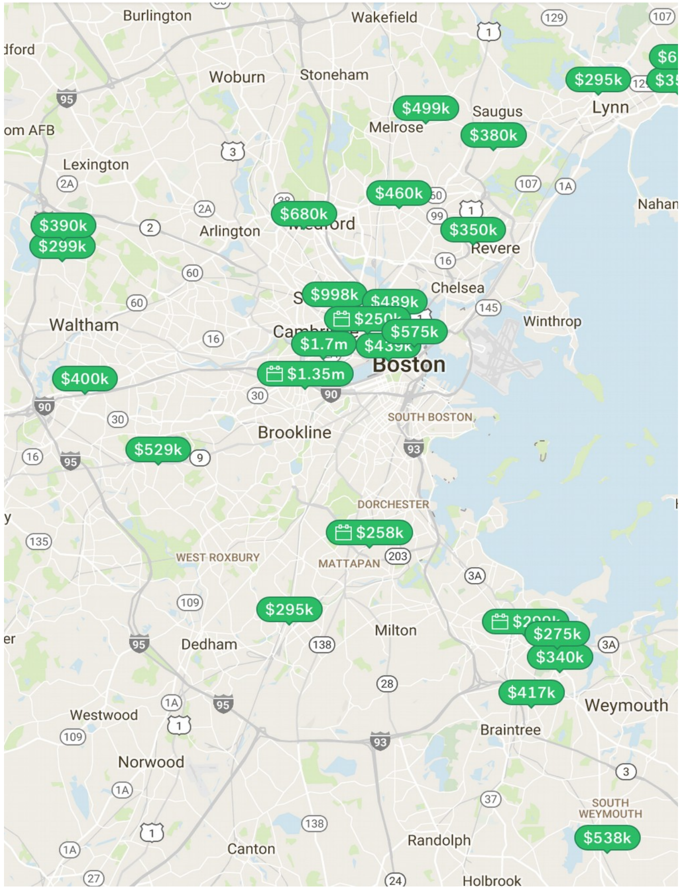


Torsney-Weir, T. , Sedlmair, M. and Möller, T. (2017), Sliceplorer: 1D slices for multi-dimensional continuous functions.
Computer Graphics Forum, 36: 167-177. doi:10.1111/cgf.13177

Housing prices in Boston



Housing prices in Boston



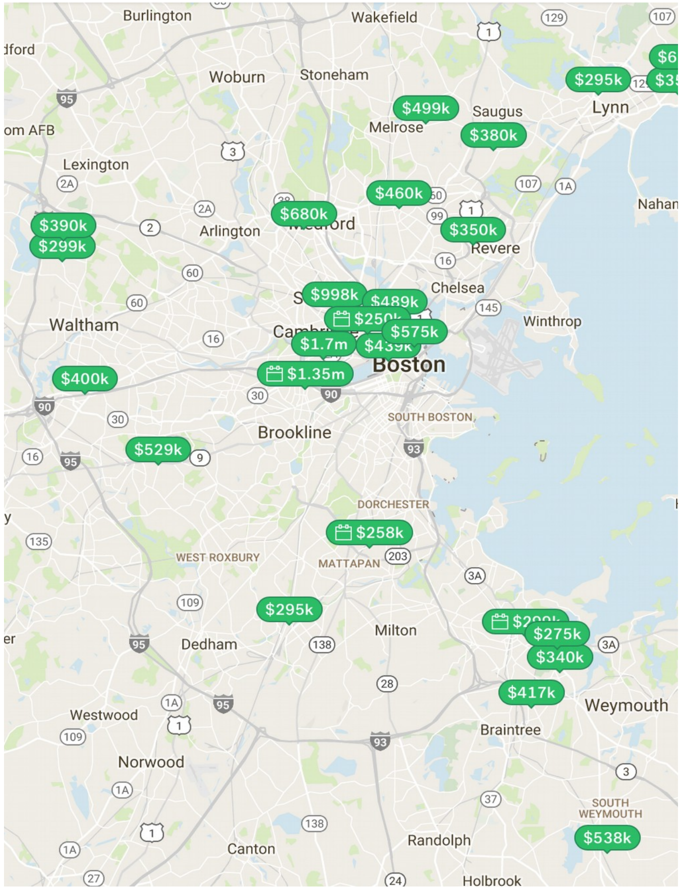
1) What's the most expensive house?

2) What factors contribute the most to changes in price?

3) How much does house price change with safety?

4) Is the relationship linear/logarithmic/etc?

Housing prices in Boston



1) What's the most expensive house?
2) What factors contribute the most to changes in price?
3) How much does house price change with safety?
4) Is the relationship linear/logarithmic/etc?

Continuous model

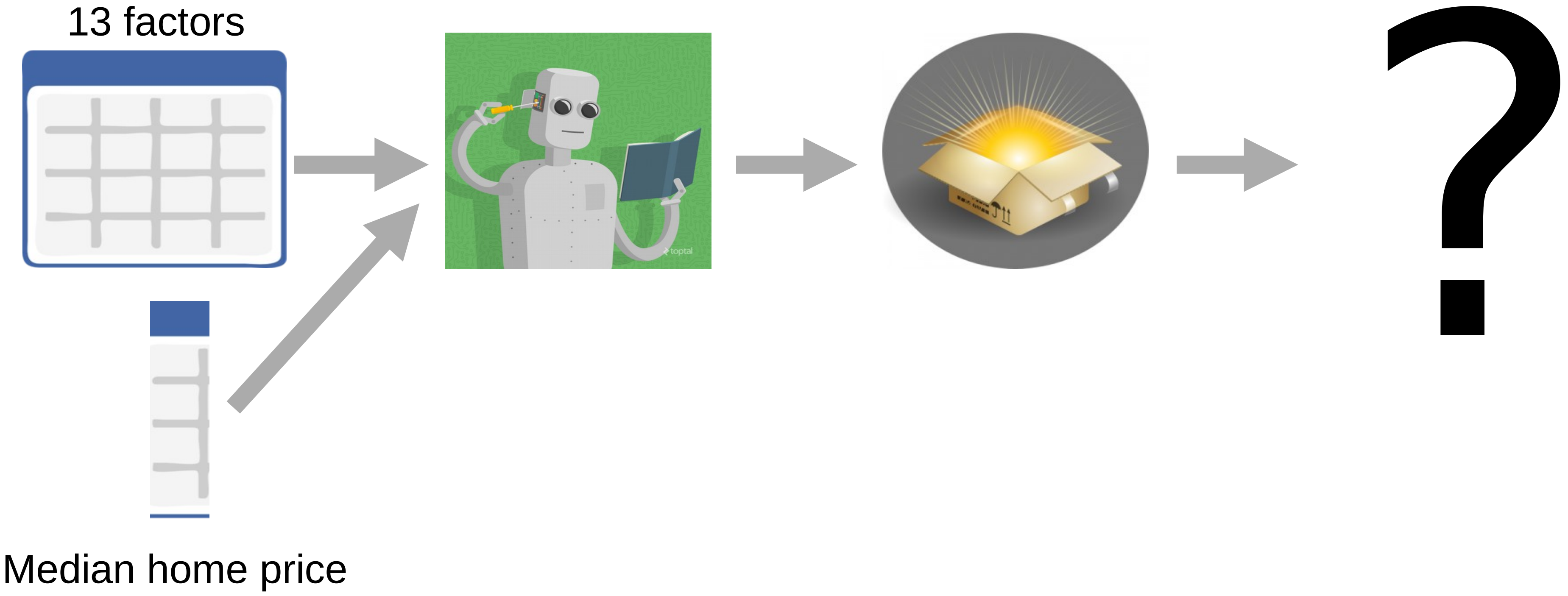
UCI housing dataset

13 factors → Median home price

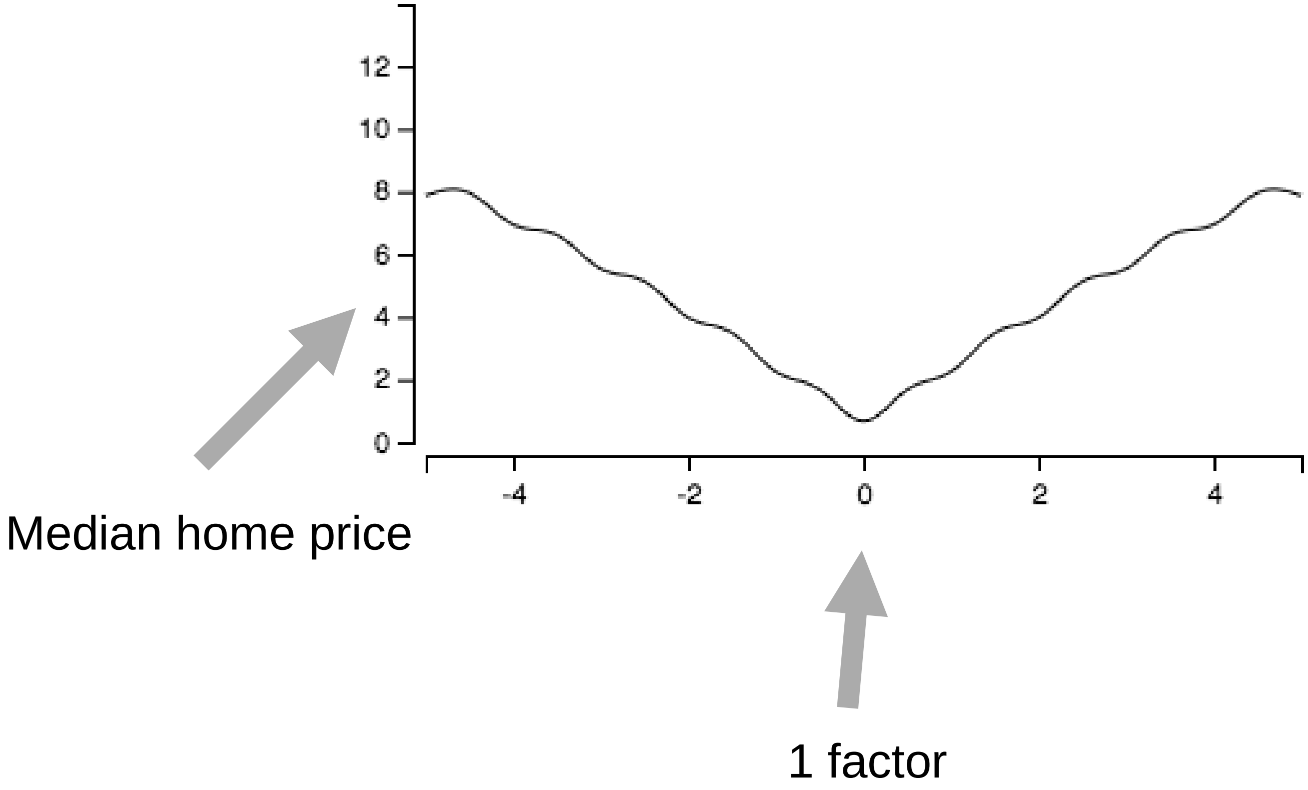
Key factors

CRIM	Crime rate
LSTAT	% lower income status
NOX	Nitric oxides concentration
RM	Average rooms per dwelling

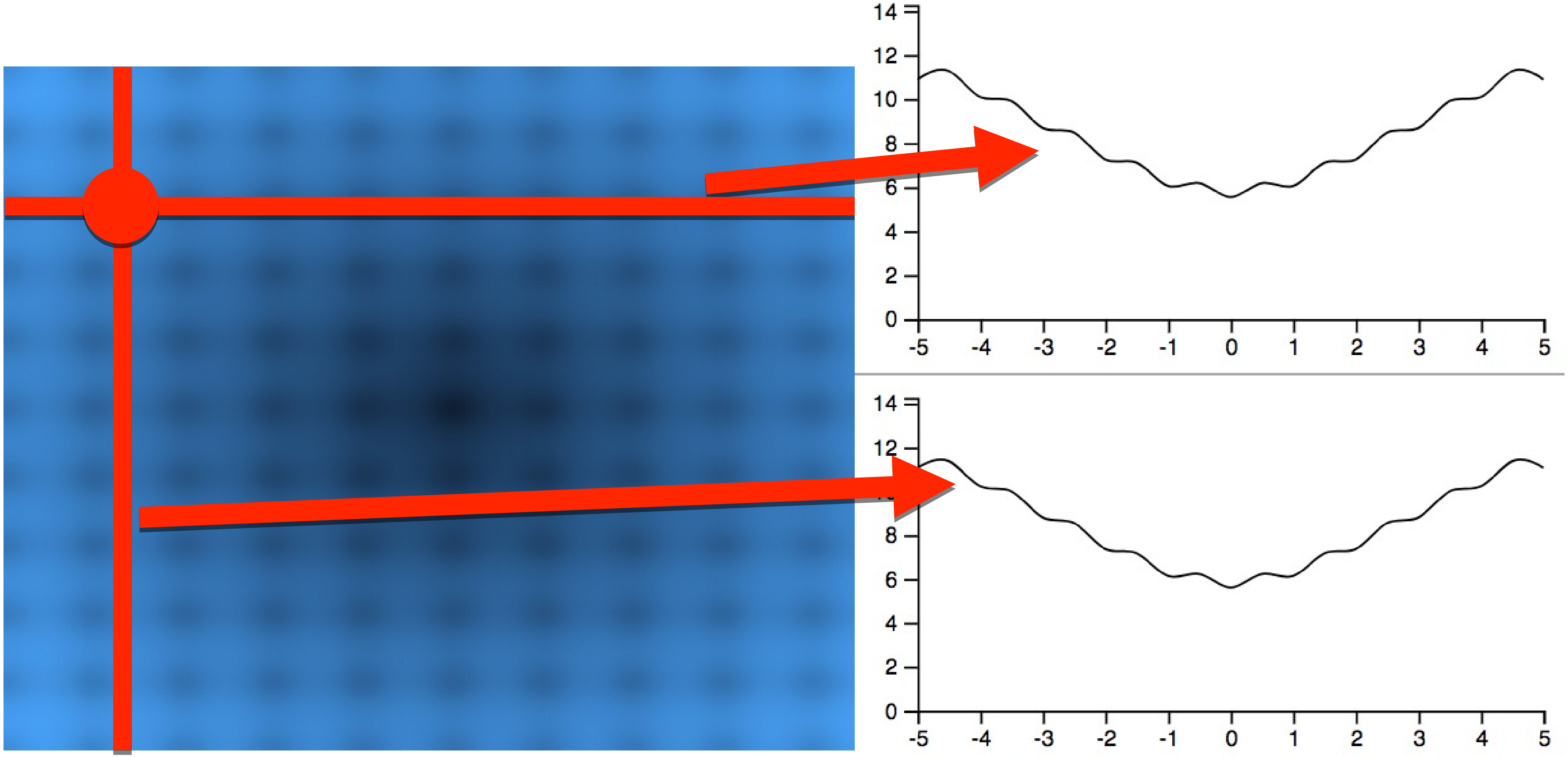
Building a model



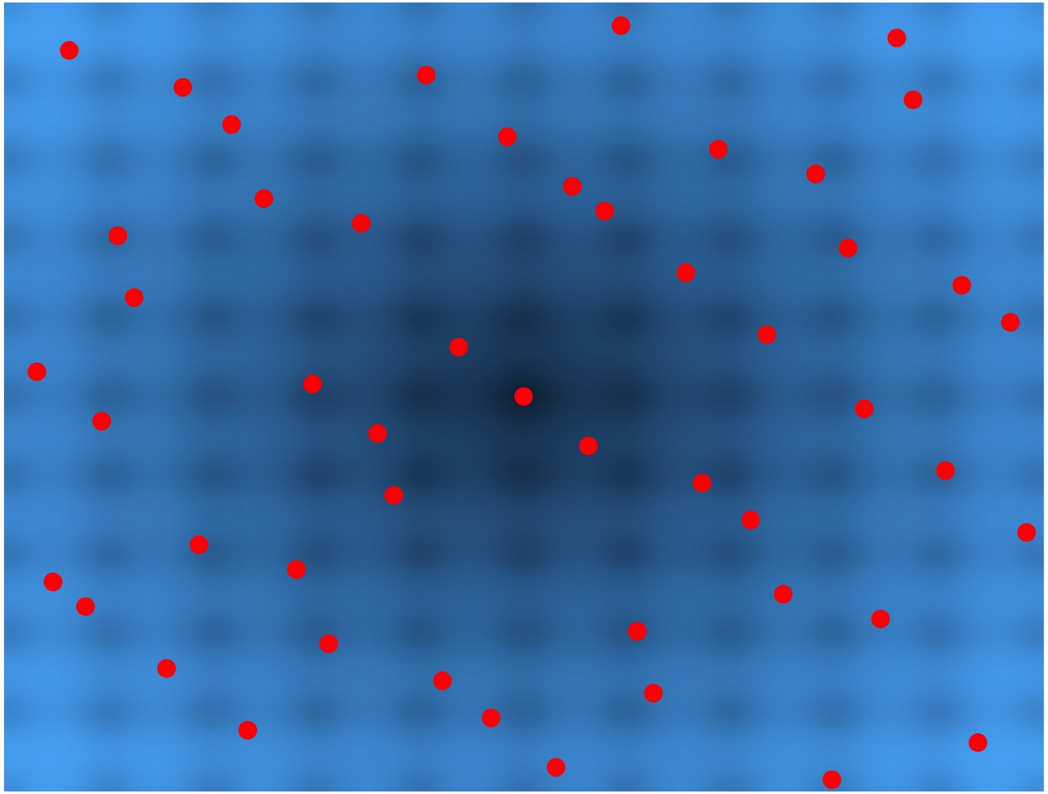
Visualization

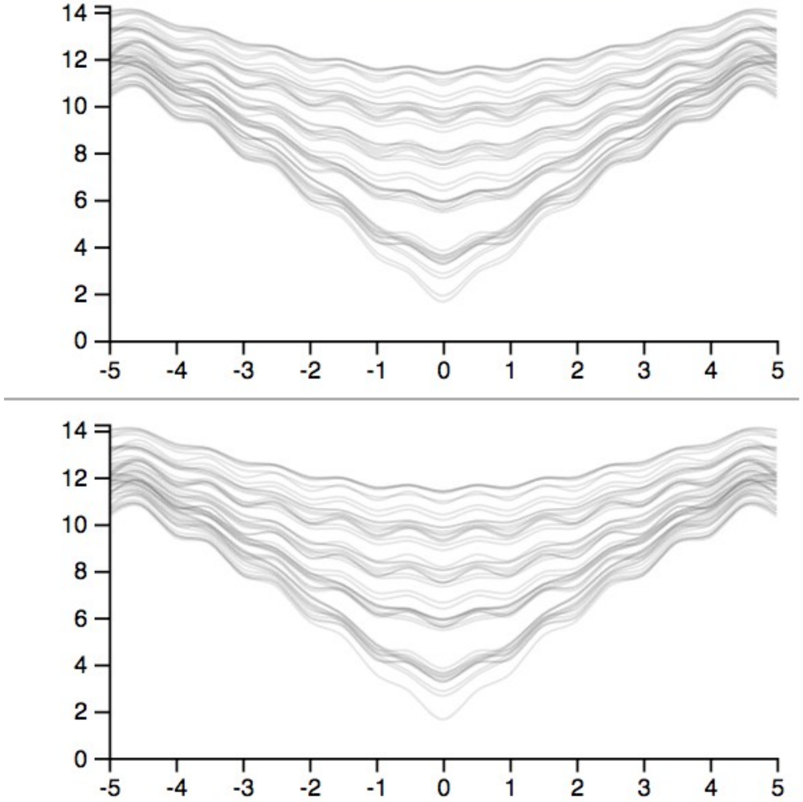


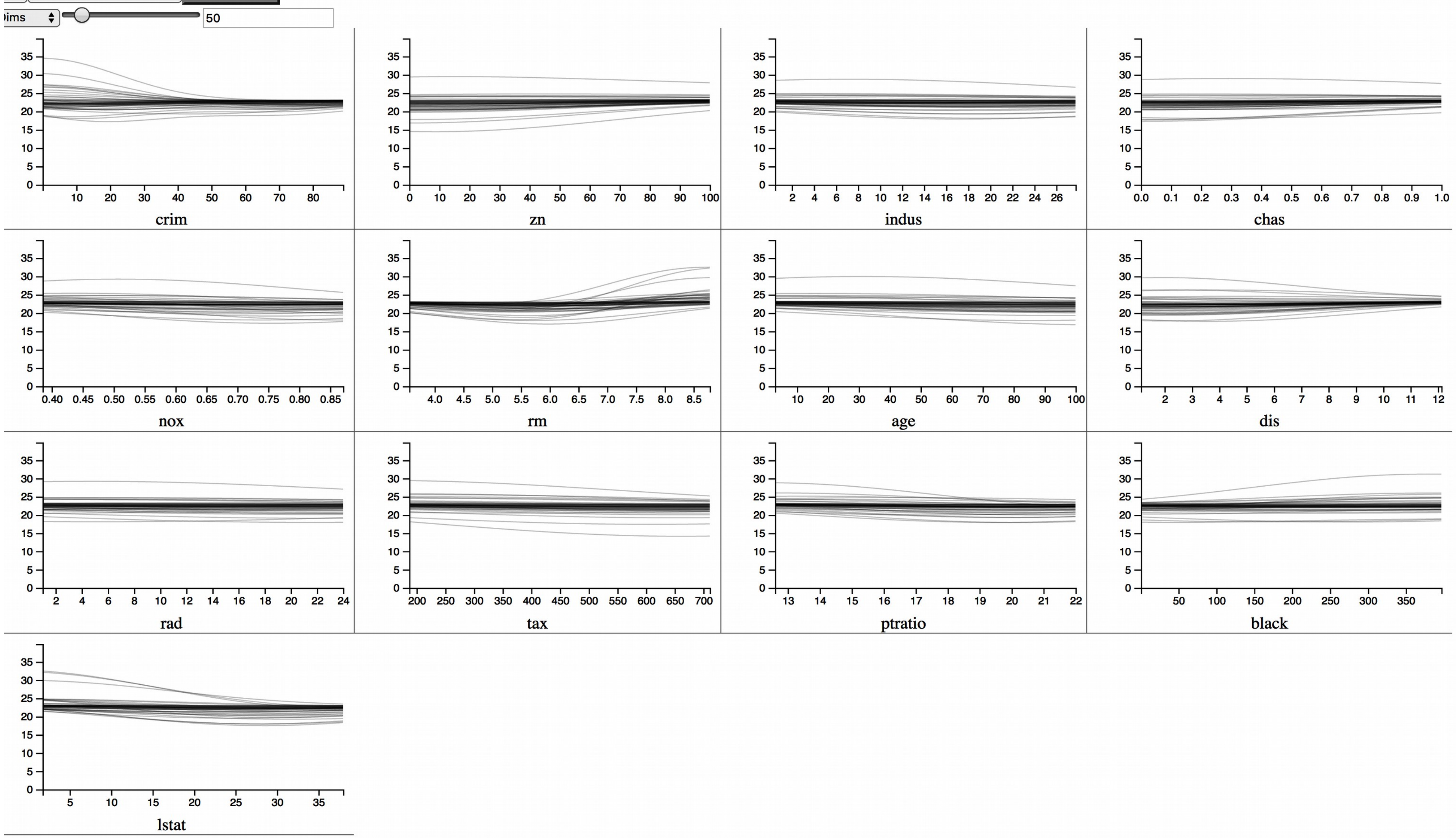
Slicing

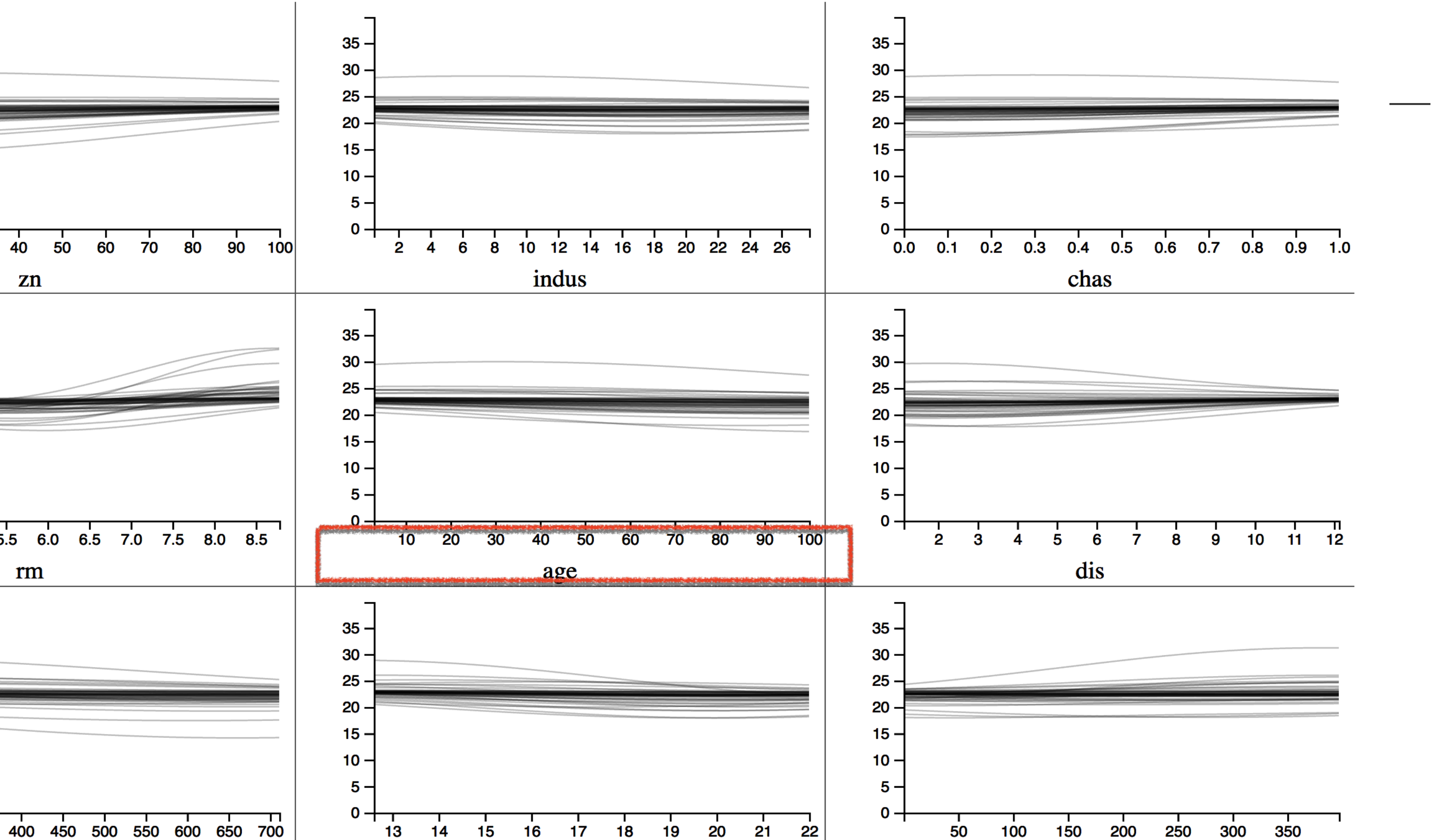


Slicing

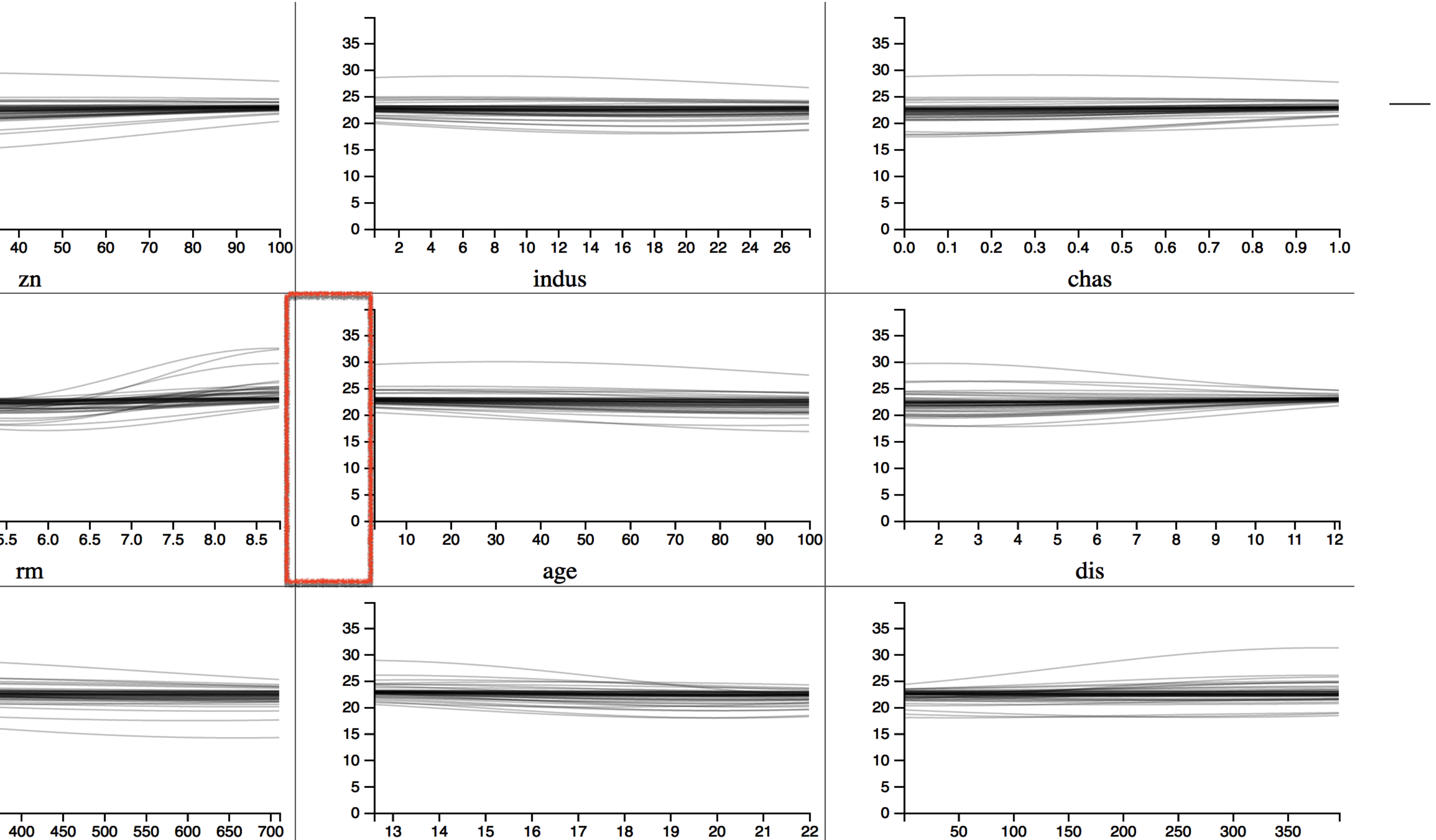


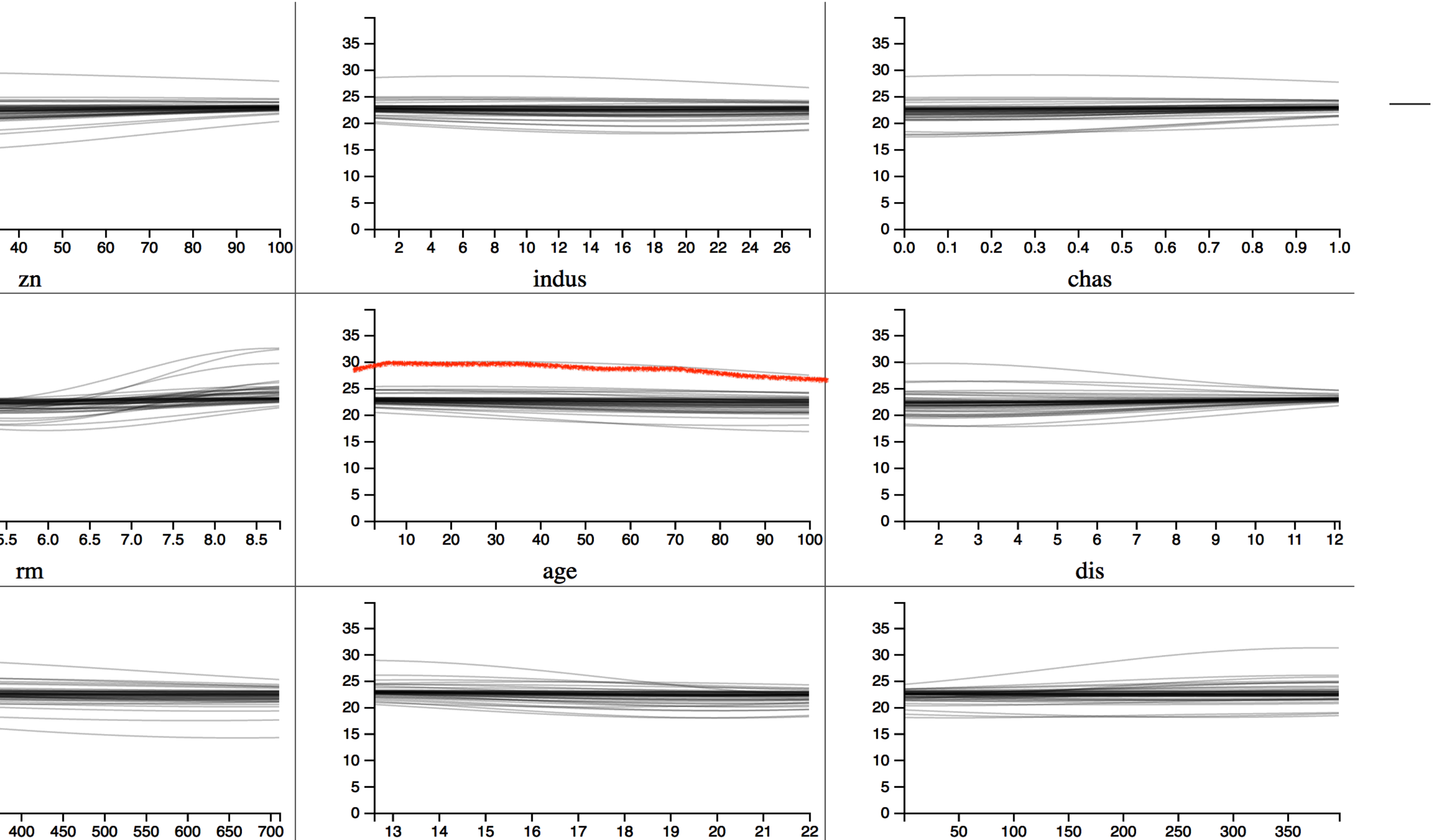



Dims
50
crim
zn
indus
chas
nox
rm
age
dis
rad
tax
ptratio
black
lstat

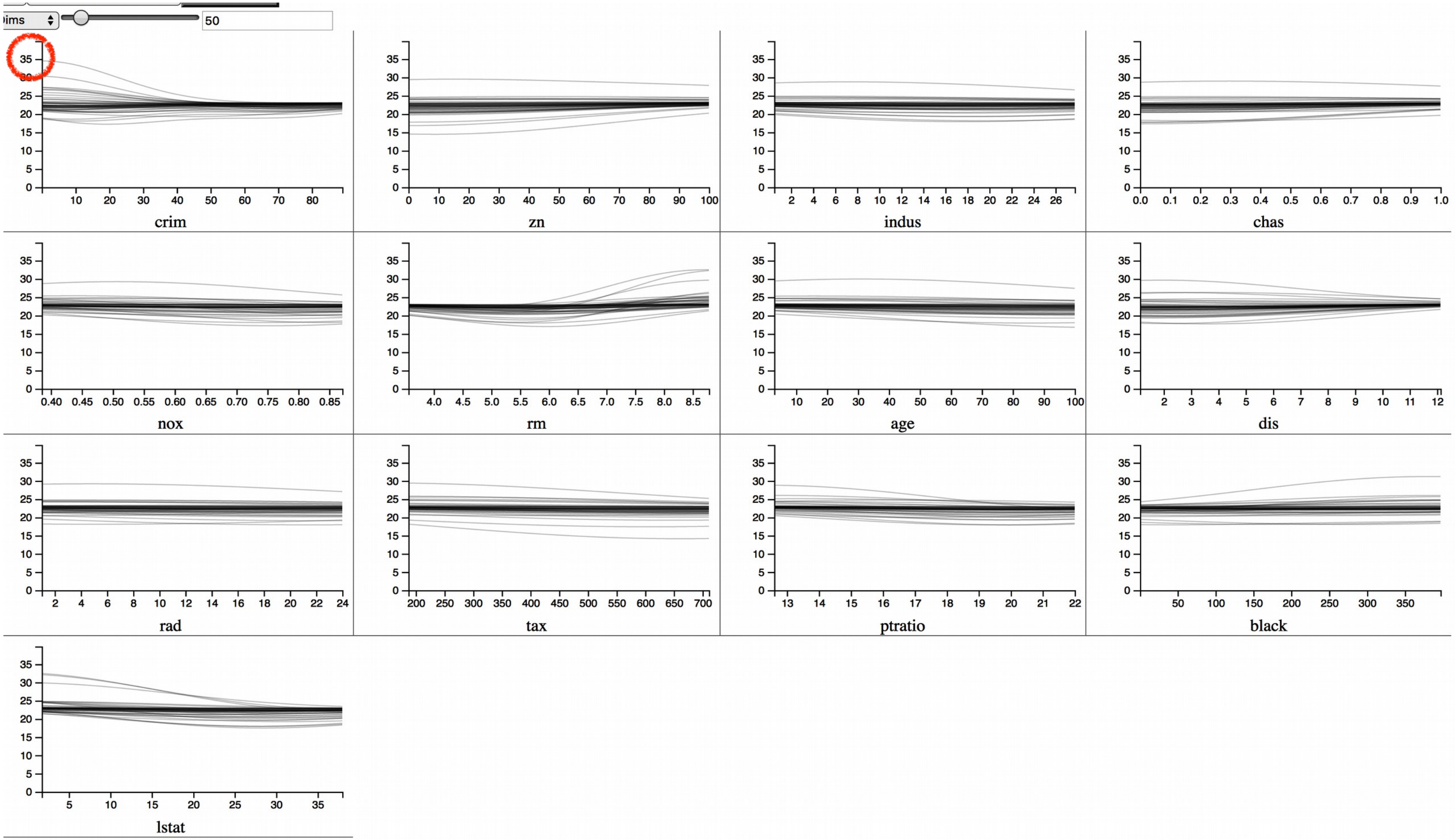


35
30
25
20
15
10
5
0
40 50 60 70 80 90 100
zn
2 4 6 8 10 12 14 16 18 20 22 24 26
indus
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
chas
5.5 6.0 6.5 7.0 7.5 8.0 8.5
rm
10 20 30 40 50 60 70 80 90 100
age
2 3 4 5 6 7 8 9 10 11 12
dis
400 450 500 550 600 650 700
13 14 15 16 17 18 19 20 21 22
50 100 150 200 250 300 350

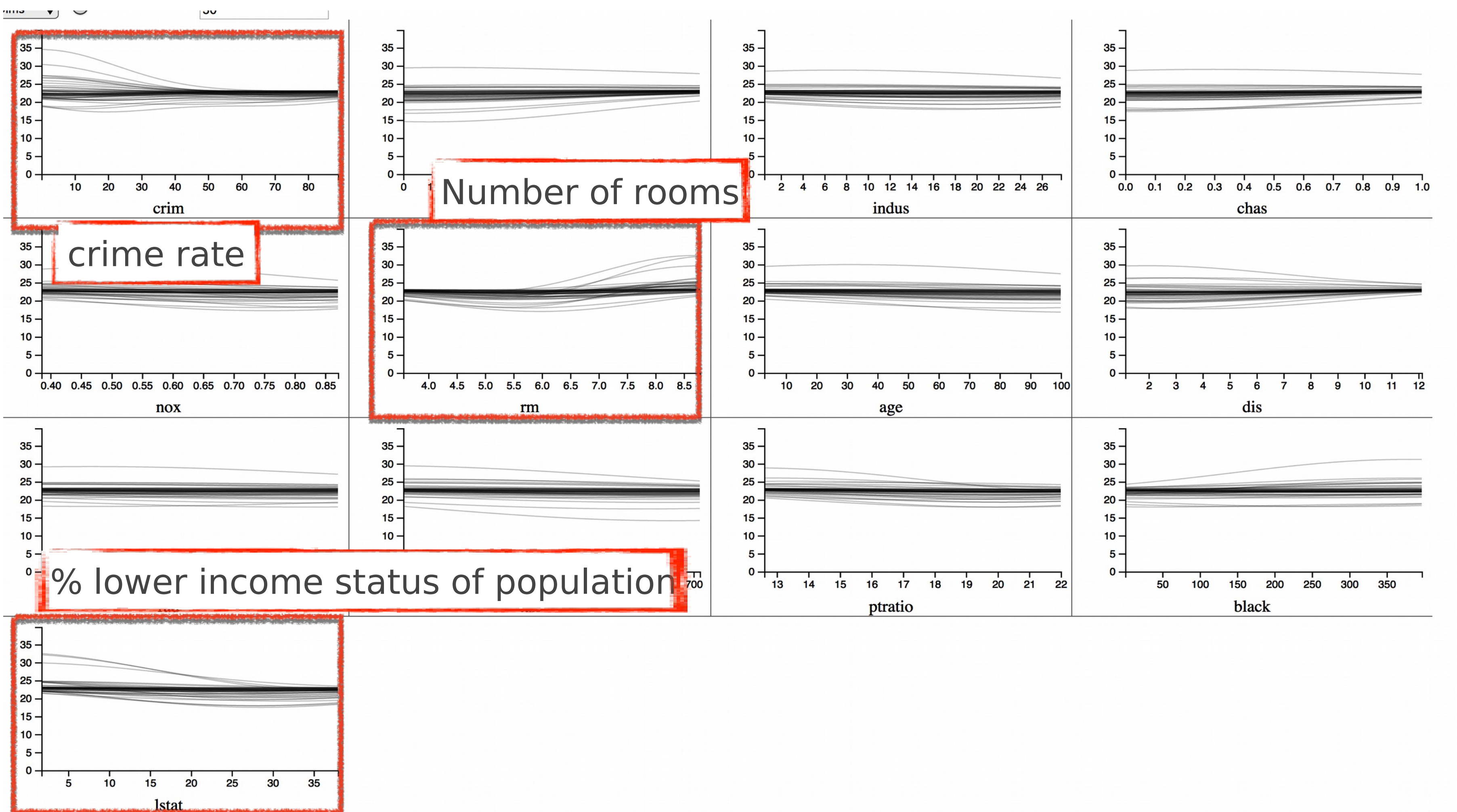

zn
indus
chas
rm
age
dis
35 30 25 20 15 10 5 0
40 50 60 70 80 90 100
2 4 6 8 10 12 14 16 18 20 22 24 26
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
5.5 6.0 6.5 7.0 7.5 8.0 8.5
10 20 30 40 50 60 70 80 90 100
2 3 4 5 6 7 8 9 10 11 12
400 450 500 550 600 650 700
13 14 15 16 17 18 19 20 21 22
50 100 150 200 250 300 350


40 50 60 70 80 90 100
zn
35 30 25 20 15 10 5 0
2 4 6 8 10 12 14 16 18 20 22 24 26
indus
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
chas
5.5 6.0 6.5 7.0 7.5 8.0 8.5
rm
10 20 30 40 50 60 70 80 90 100
age
2 3 4 5 6 7 8 9 10 11 12
dis
400 450 500 550 600 650 700
13 14 15 16 17 18 19 20 21 22
50 100 150 200 250 300 350

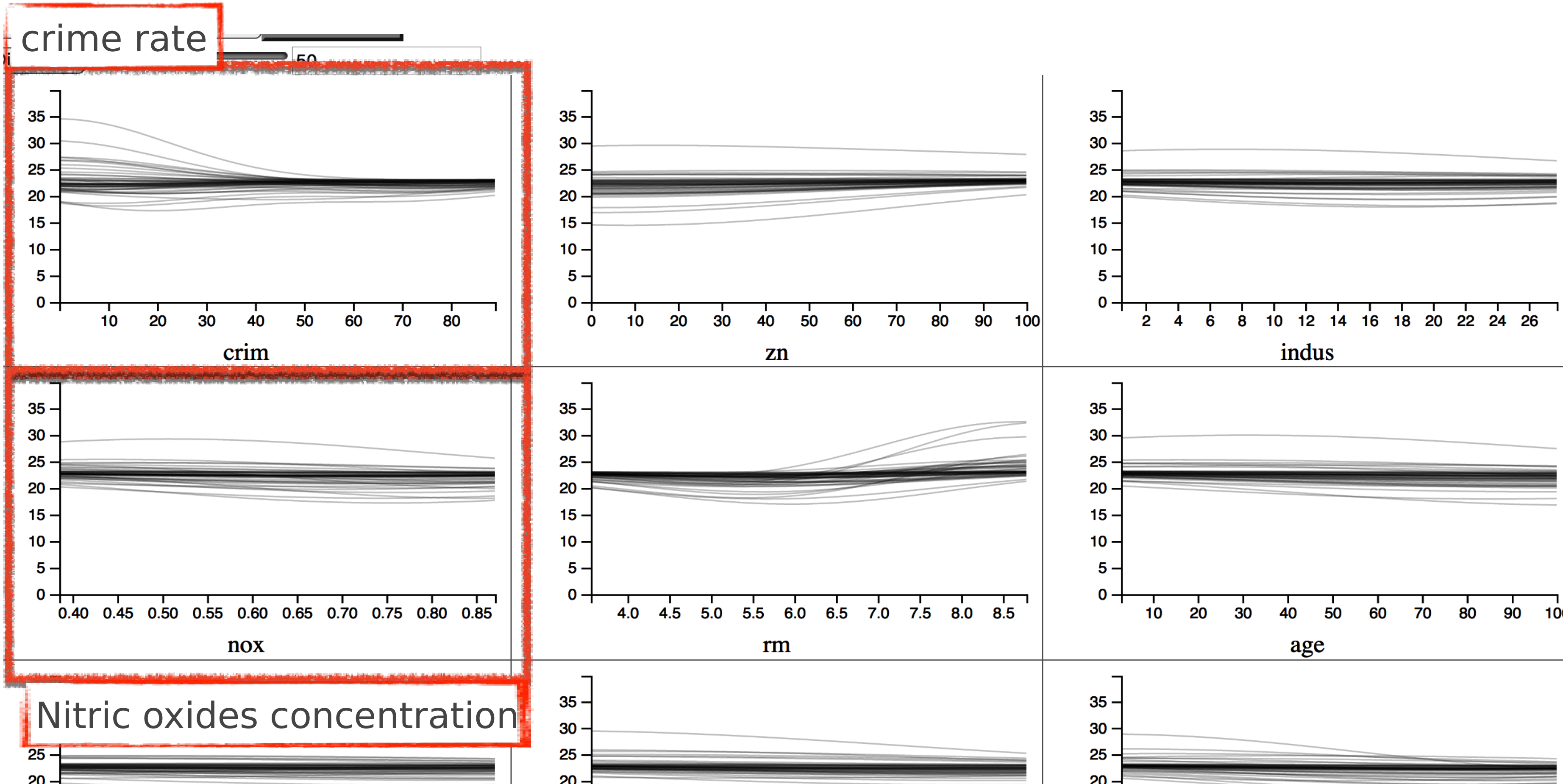
Q1: Most expensive house



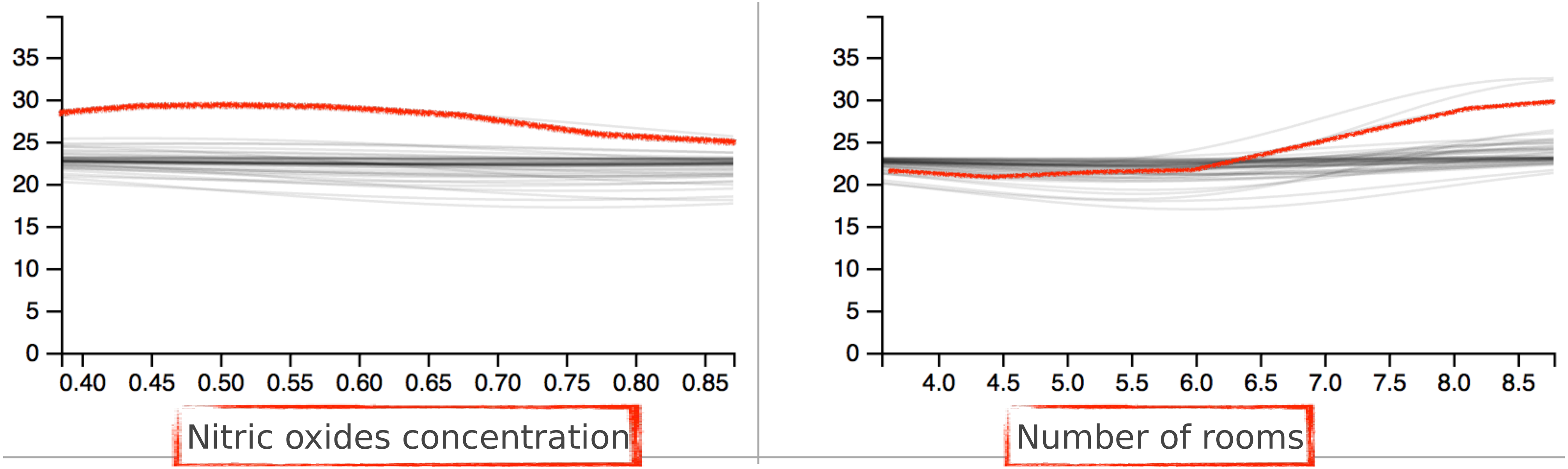
Q2:


Number of rooms
crime rate
% lower income status of population
crim
indus
chas
nox
rm
age
dis
ptratio
black
lstat

Q3: Influence of safety?

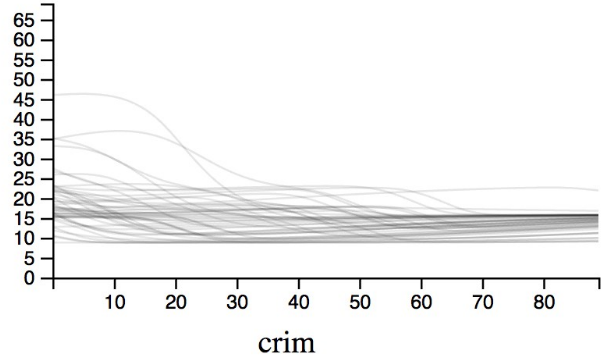


Q4: Type of relationship?

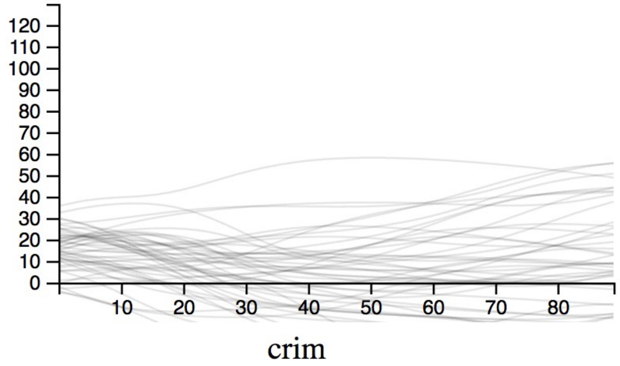


Model selection

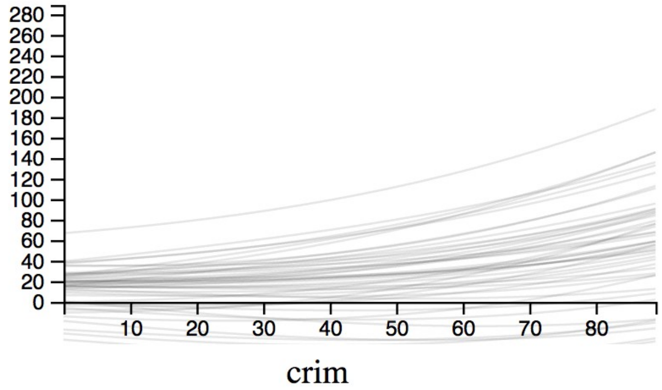
NN – 2 layer



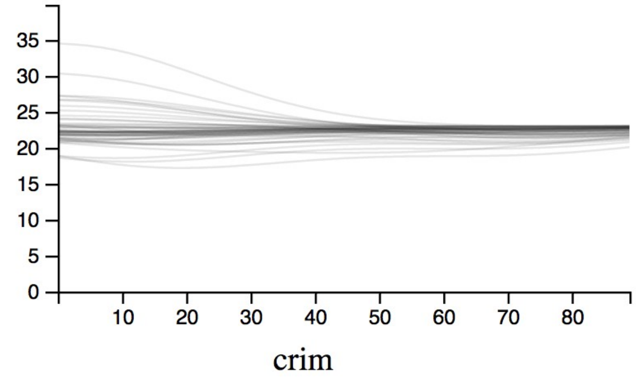
NN – single layer



SVM - polynomial



SVM - RBF

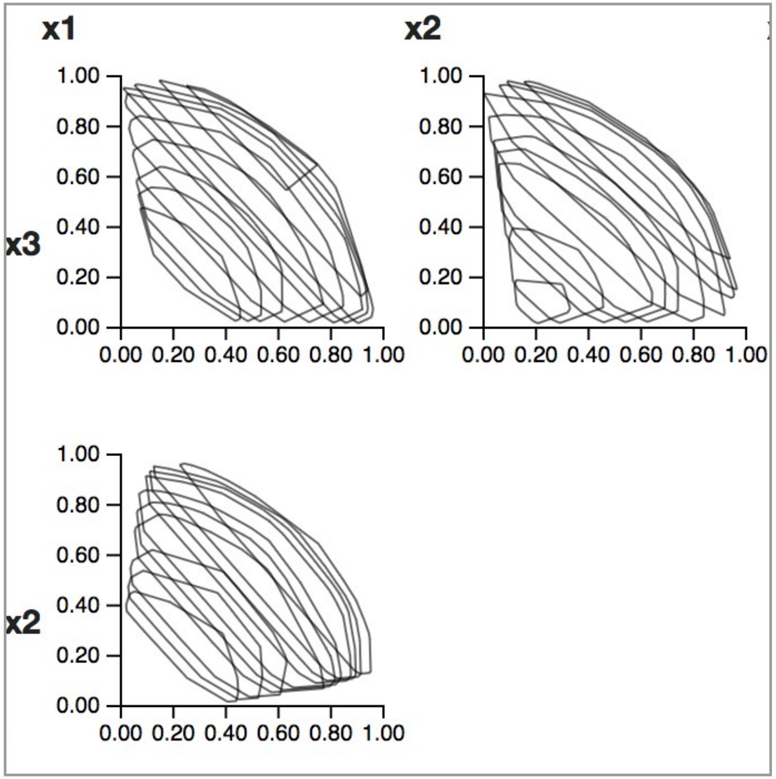


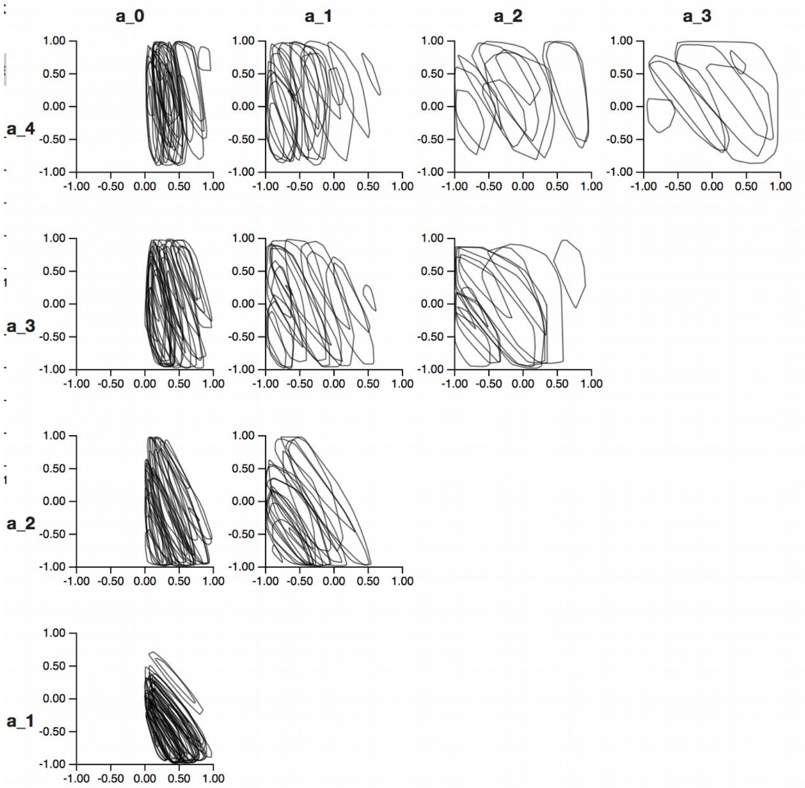
Sliceplorer

- Good “first pass” visualization
- Easy to implement
- Easy to understand

Hypersliceplorer

Relationships between parameters

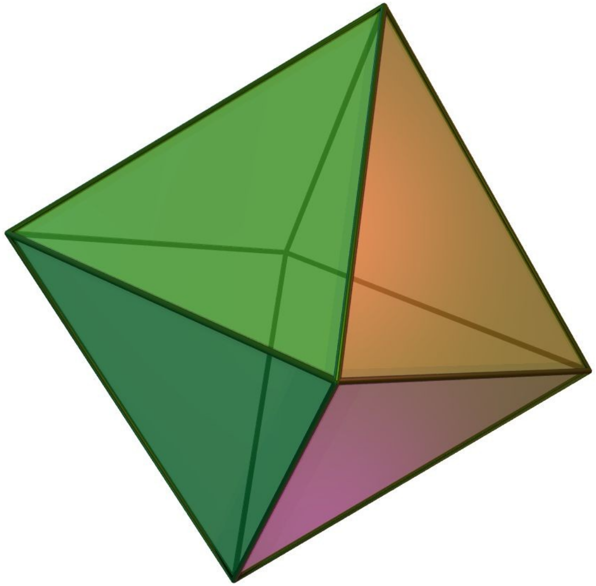




Torsney-Weir, T. , Möller, T. , Sedlmair, M. and Kirby, R. M. (2018), Hypersliceplorer: Interactive visualization of shapes in multiple dimensions. Computer Graphics Forum, 37: 229-240. doi:10.1111/cgf.13415

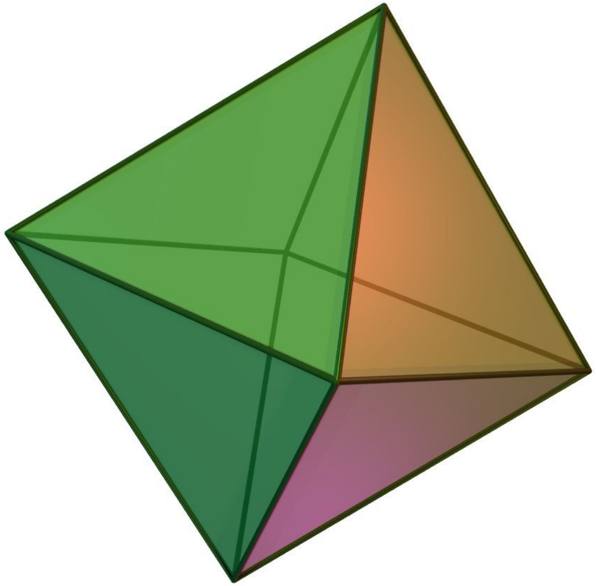
Orthoplex

3D

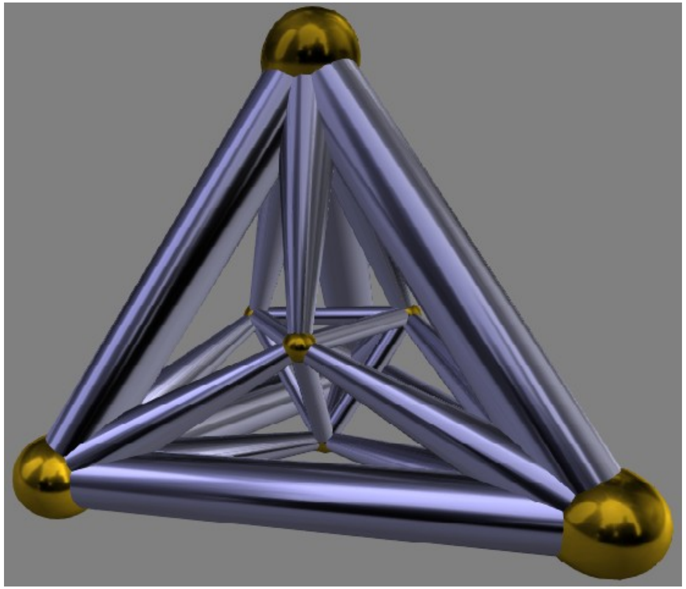


Orthoplex

3D



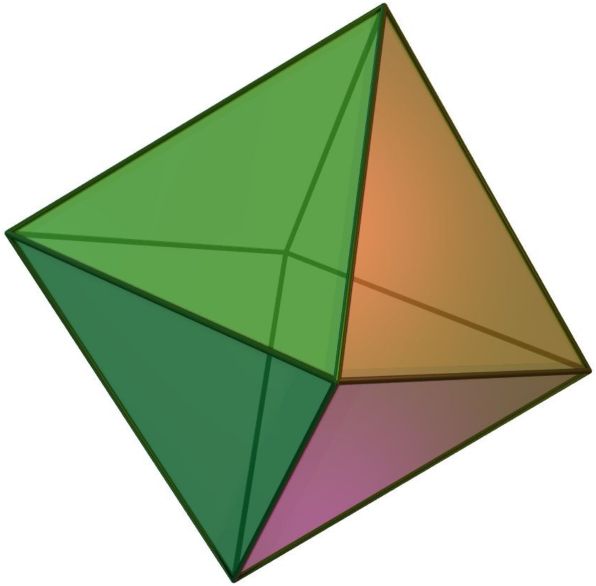
4D



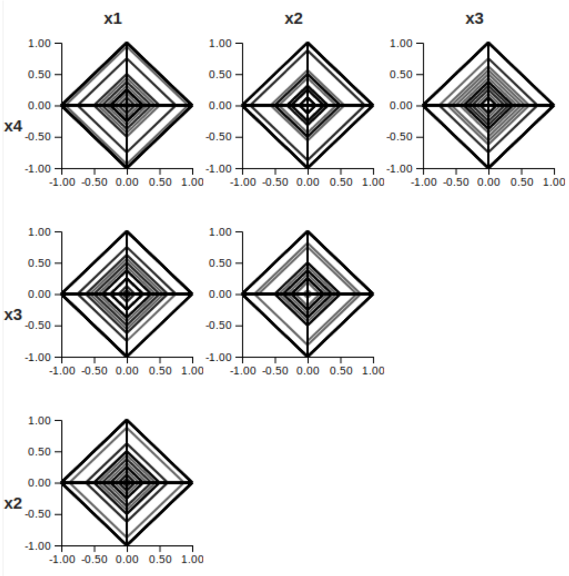
Sommerville, 1929

Orthoplex

3D

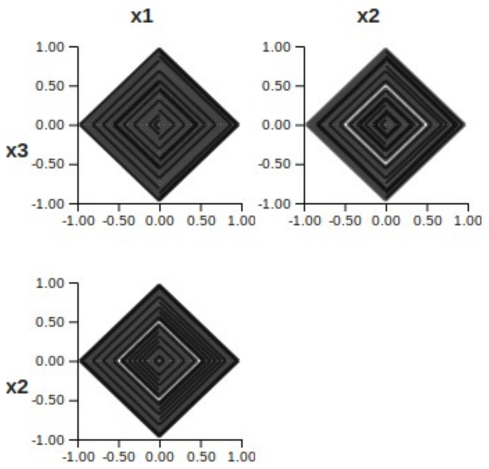


4D

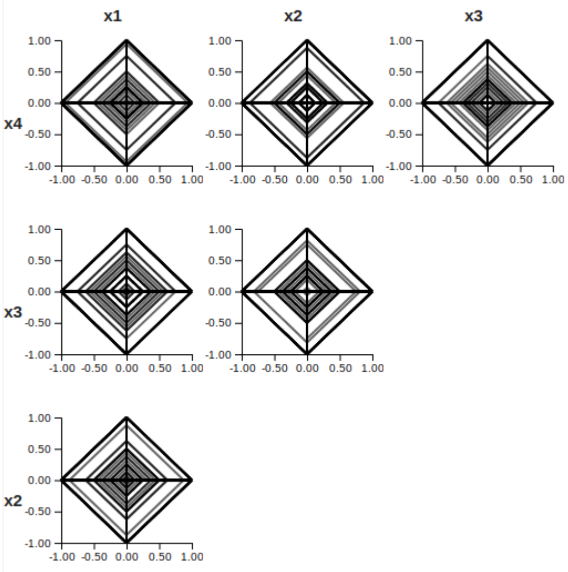


Orthoplex

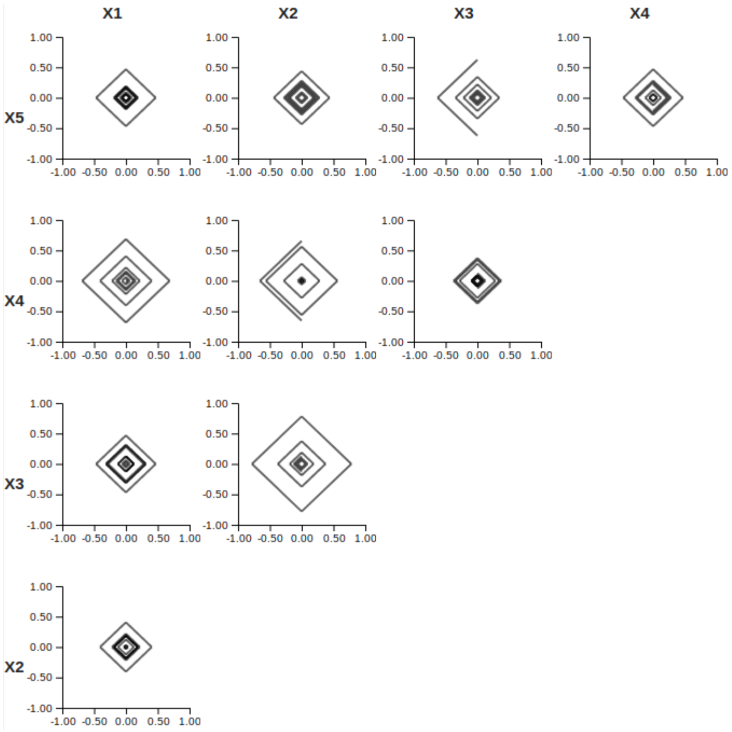
3D



4D

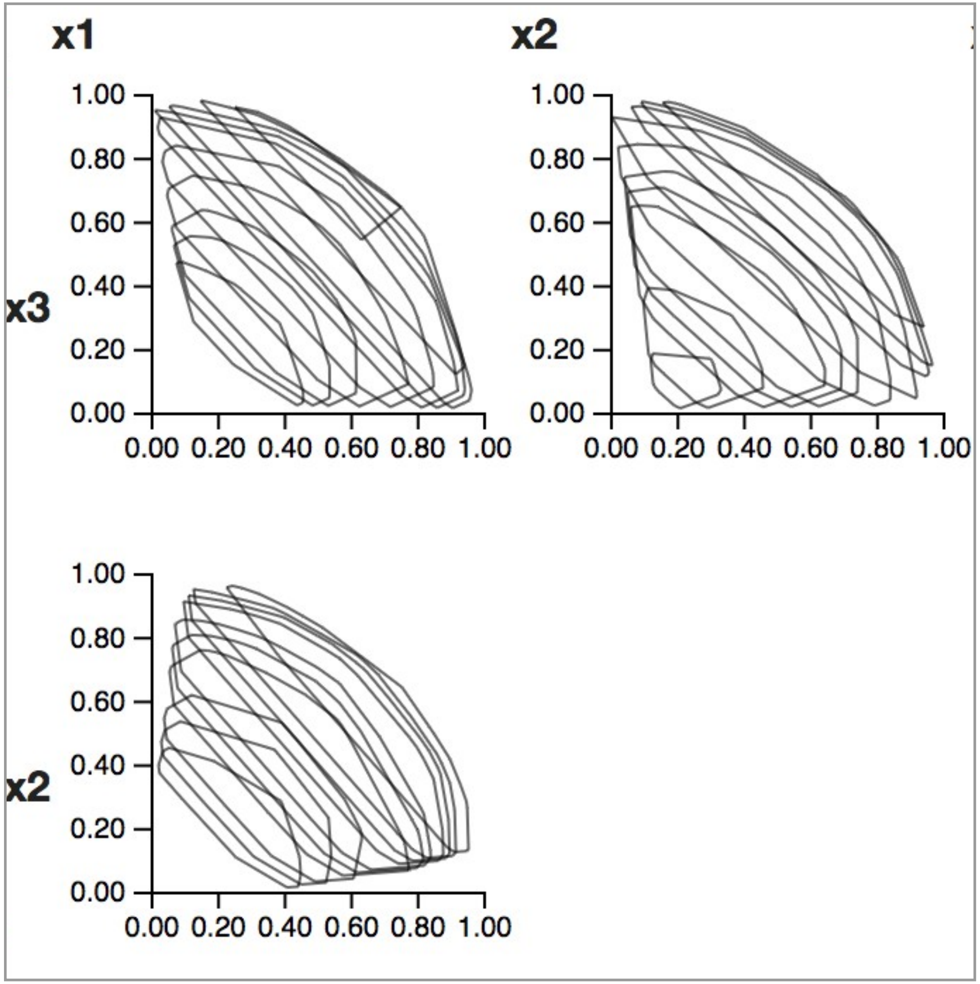


5D



What are we doing?

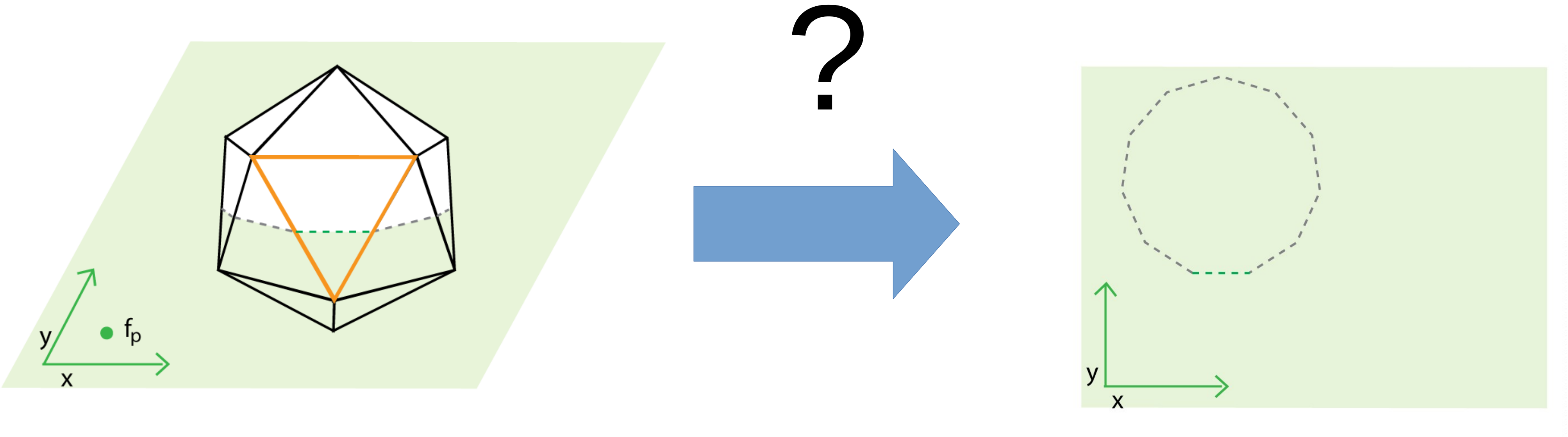
- Randomly sample focus points
- Projections of 2D slices
- Interactive viewer



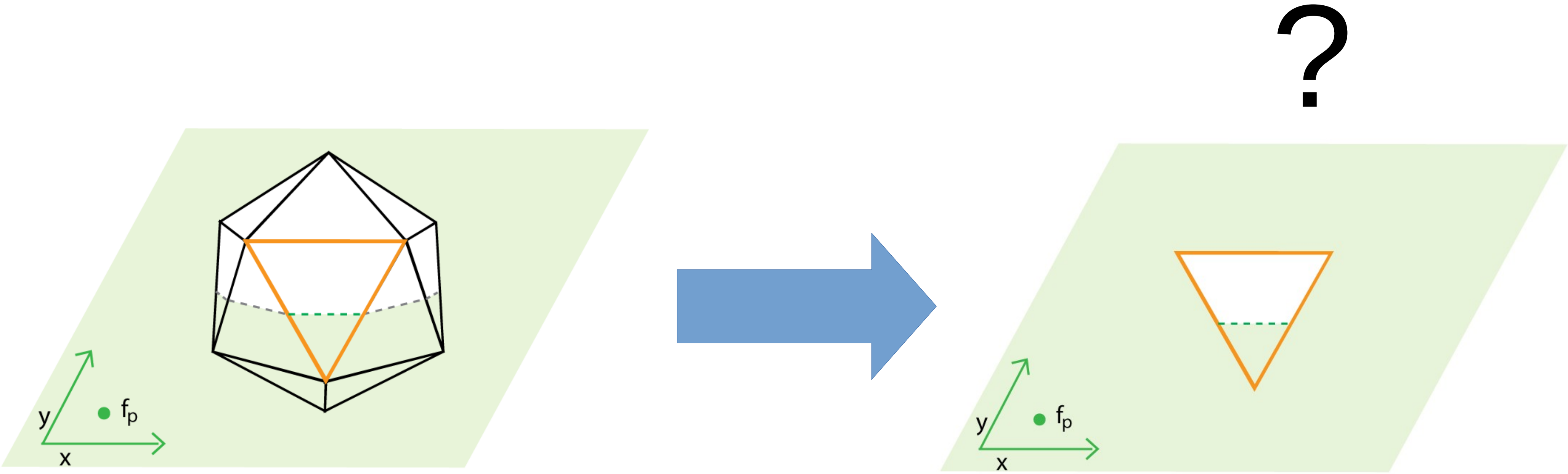
Algorithm

1) Input vertices
2) Compute the convex hull --- (d-1)-dimensional simplices (quickhull)
3) Generate m d-dimensional focus points (Sobol sequence)
4) For each 2D plane, focus point, and simplex compute the intersection between the 2D plane and the (d-1)-dimensional simplex
5) Draw each intersection line for each focus point in a SPLOM layout

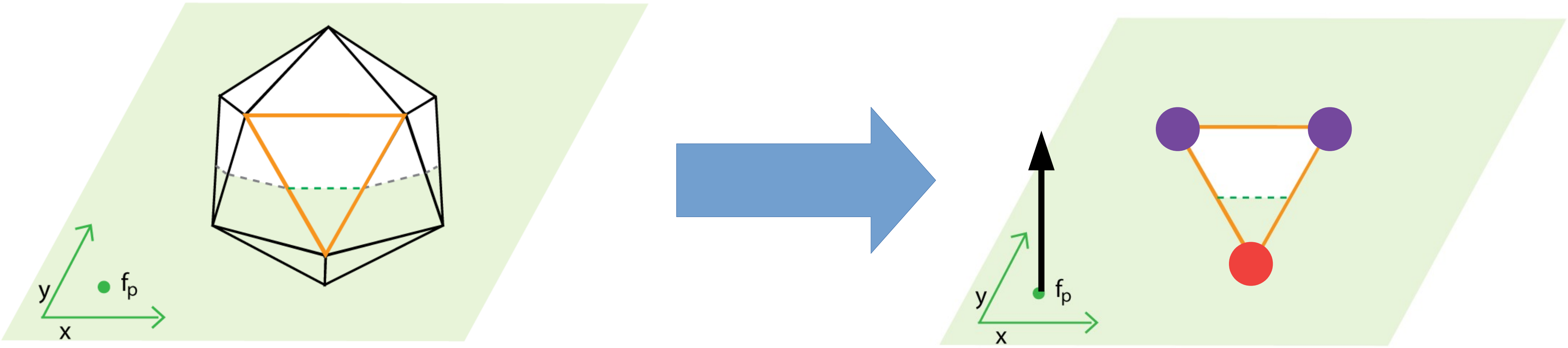
Polytopes



Polytopes

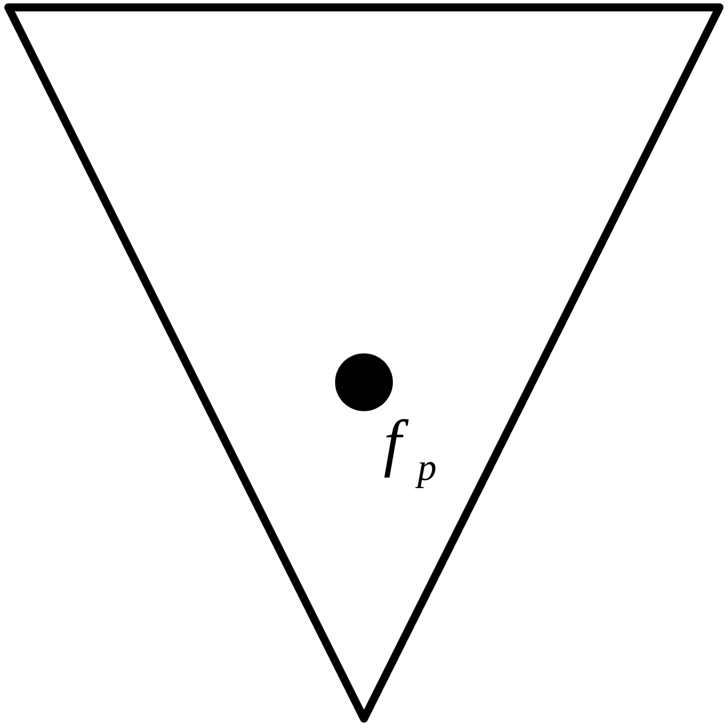


Polytopes



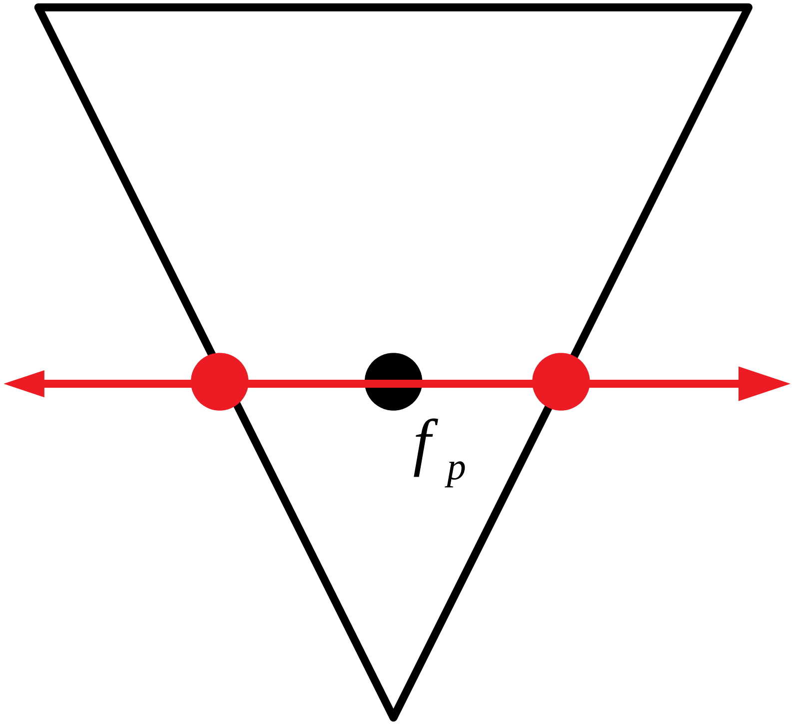
Hypersliceplorer

$$f_p = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_d \end{bmatrix}$$



Hypersliceplorer

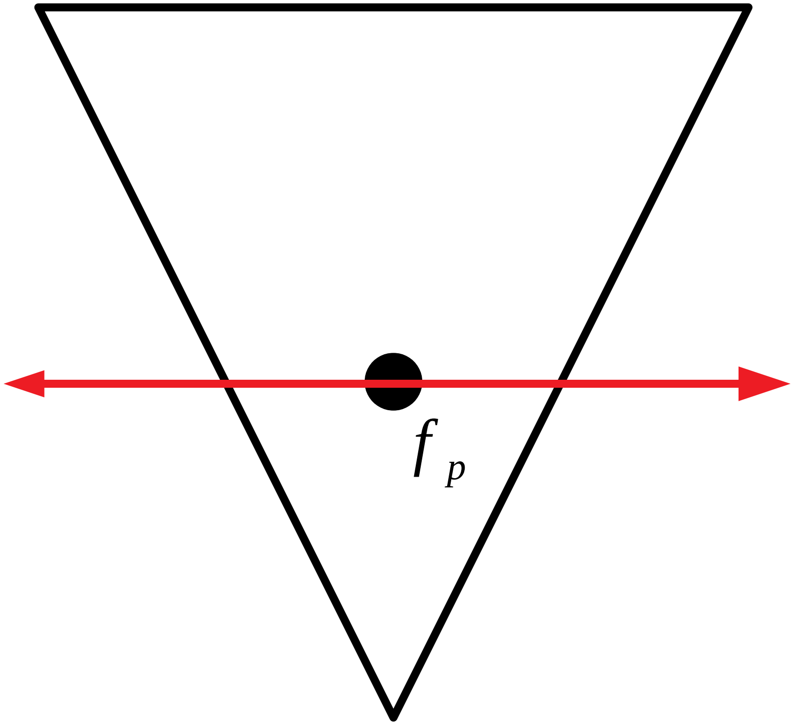
$$f_p = \begin{bmatrix} x \\ p_2 \\ p_3 \\ \vdots \\ p_d \end{bmatrix}$$



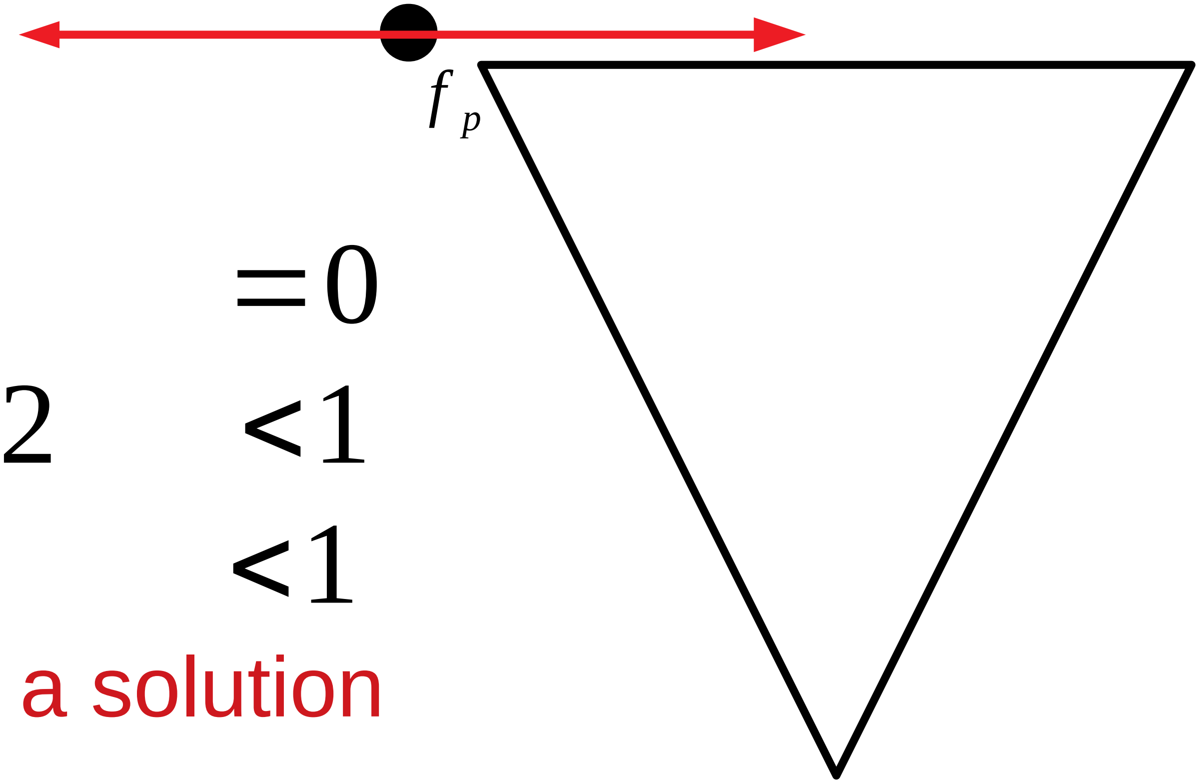
Hypersliceplorer

$$0.24x + 0.27 = 0$$
$$-0.43x + 0.12 < 1$$
$$0.78x - 0.2 < 1$$

There will be a solution



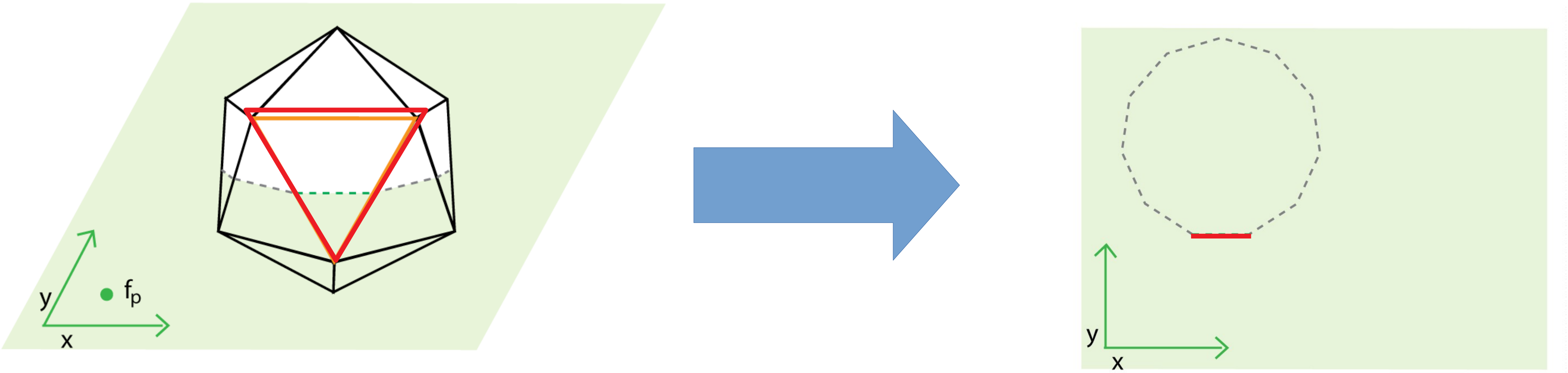
Hypersliceplorer



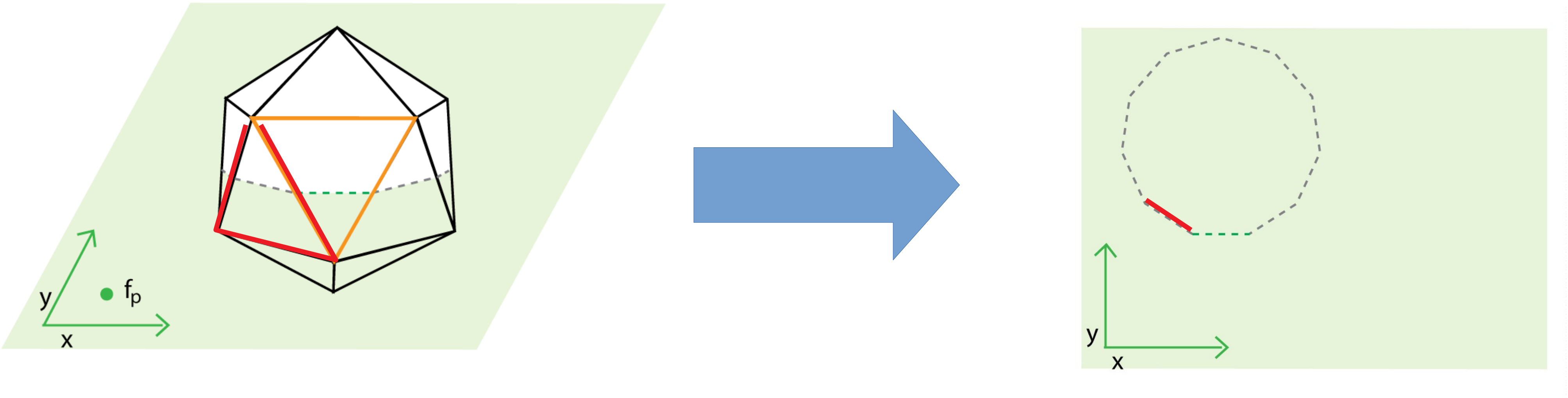
$$0.24x + 0.27 = 0$$
$$-0.43x + 0.12 < 1$$
$$0.78x - 0.2 < 1$$

There will not be a solution

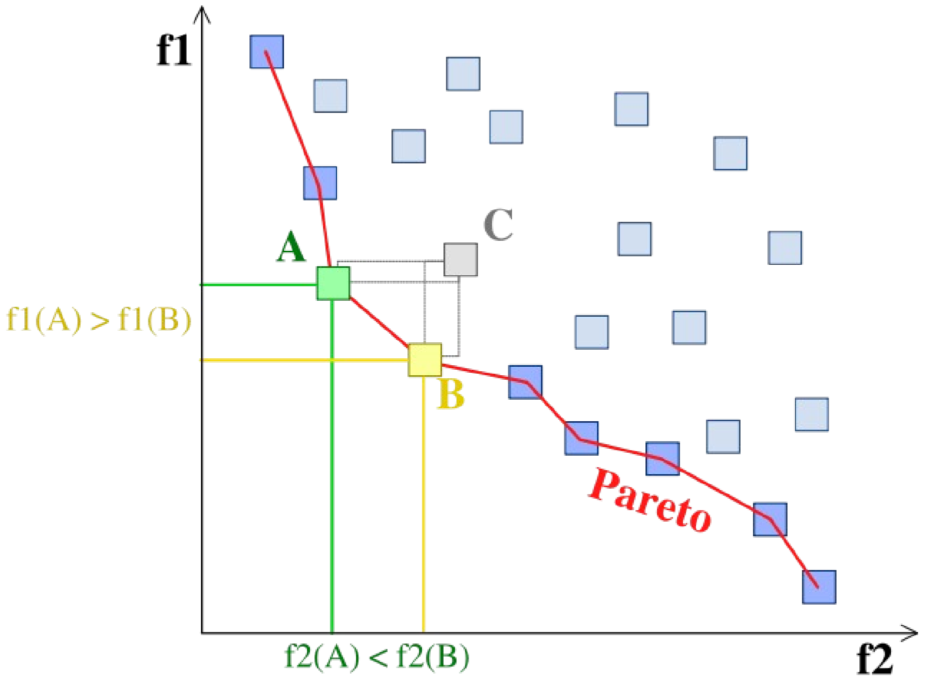
Polytopes



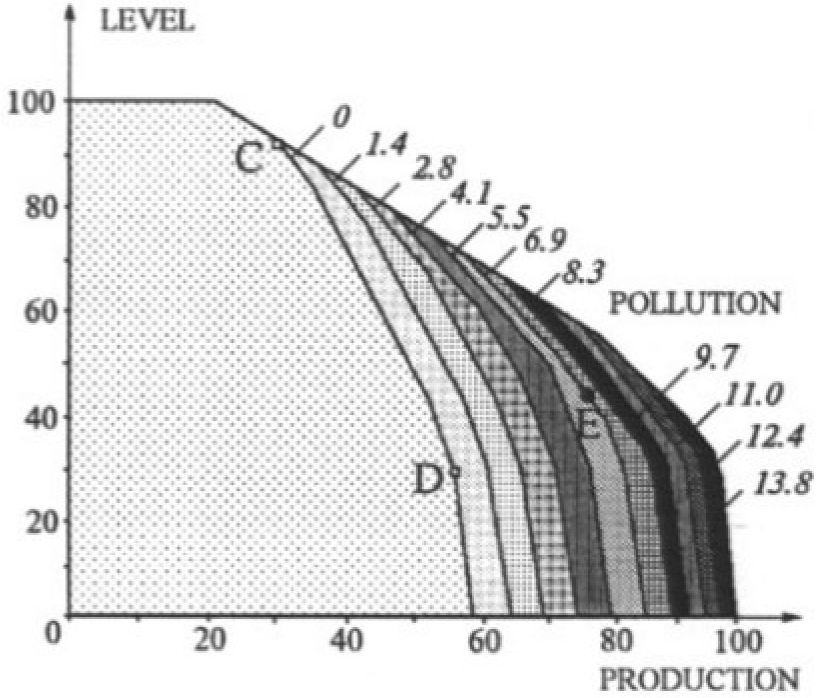
Polytopes



Pareto fronts



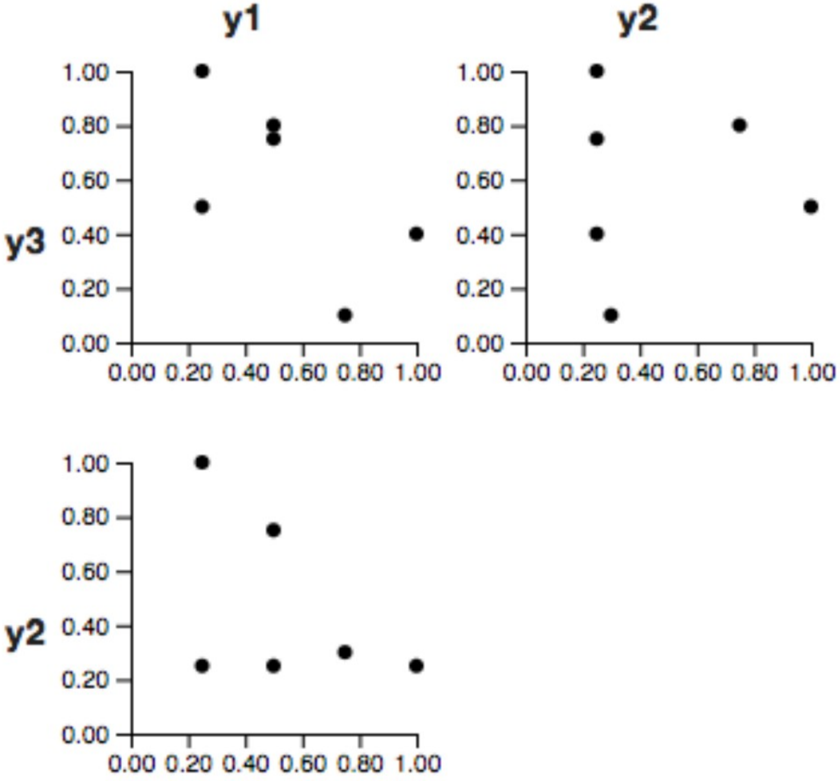
2 objectives: trade-off curve



3 objectives: Interactive decision maps [Lotov:2004]

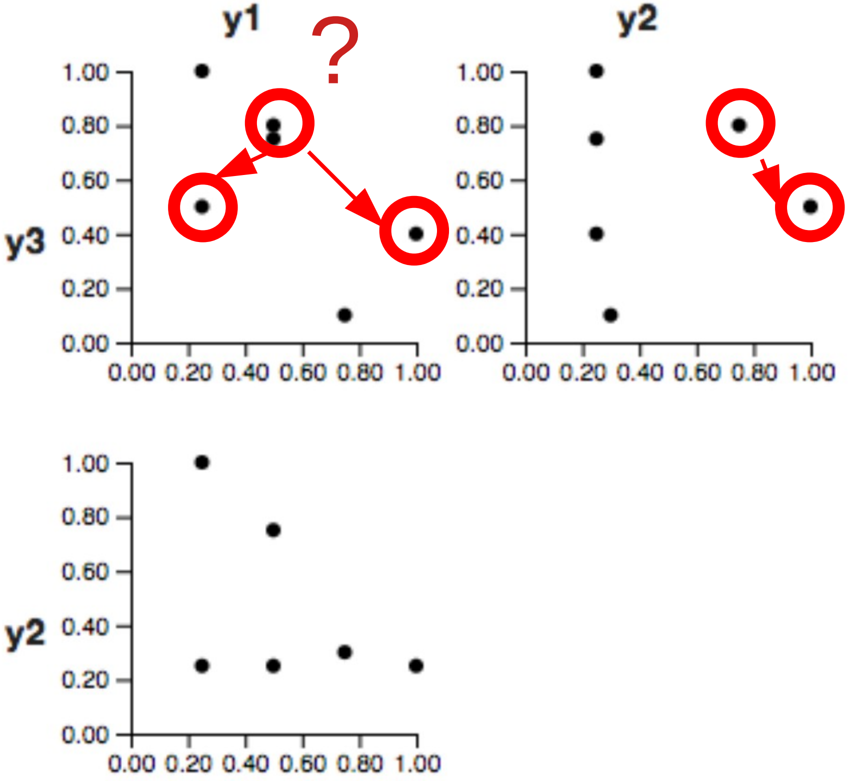
Typical method

4+ objectives: SPLOM

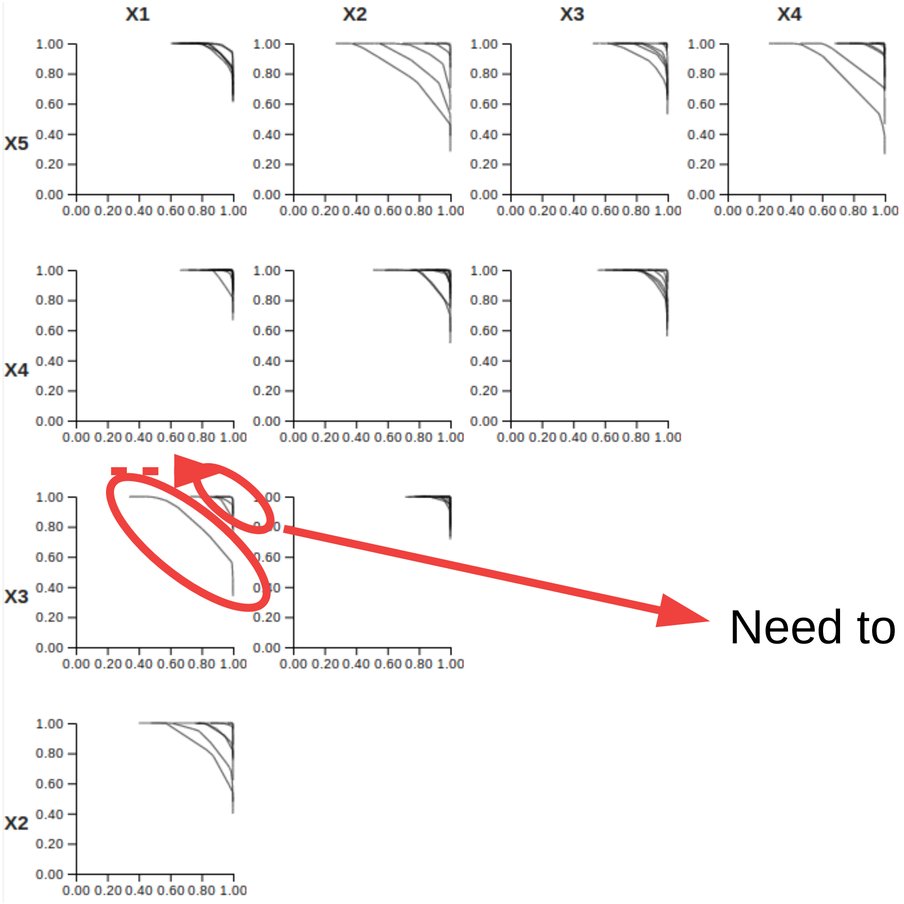


Typical method

4+ objectives: SPLOM



Pareto fronts



Need to lower other objective values

Hypersliceplorer

- Algorithm for 2D slices of polytopes
- Relationships between dimensions

Conclusion

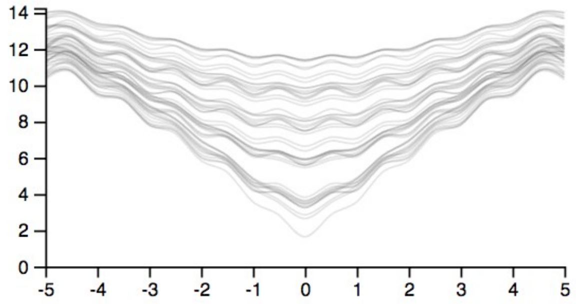
Future work

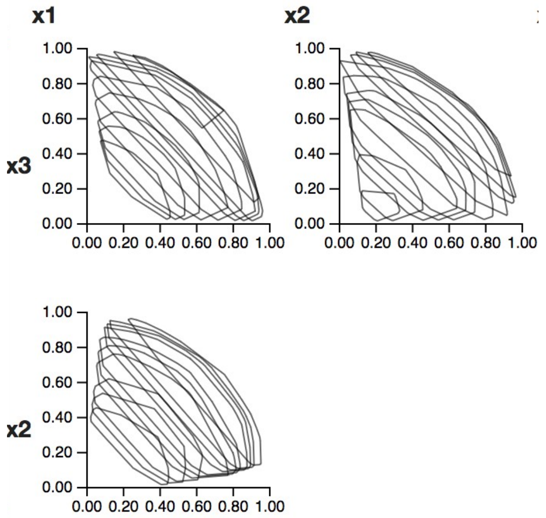
- Relatively un-explored area
  - Most work on discrete data visualization
- New visual encodings
- Refining tasks
- Improving algorithms

Task	Task description for discrete data items from [AES05]	Our adaption to continuous scalar functions	QRI results					Expert study results				
			HyperSlice	Topological spines	Contour tree	Gerber et al.	1D slices	HyperSlice	Topological spines	Contour tree	Gerber et al.	1D slices
Retrieve value	"Given a set of specific cases, find attributes of those cases"	Given an x, what is the function value?										
Filter	"Given some concrete conditions attribute values, find data cases satisfying those conditions."	For what parameter values is the function equal or over x?										
Compute derived value	"Given a set of data cases, compute an aggregate numeric representation of those data cases"	Summary statistics: variance, mean, SA										
Find extremum	"Find data cases possessing an extreme value of an attribute over its range within the data set"	Find local/global min/max										
Determine range	"Given a set of data cases and an attribute of interest, find the span of values within the set"	What is the range of possible outputs?										
Characterize distribution	"Given a set of data cases and a quantitative attribute of interest, characterize the distribution of that attribute's values over the set"	What types of shapes do the manifolds have										
Find anomalies	"Identify any anomalies within a given set of data cases with respect to a given relationship or expectation, e.g. statistical outliers"	Do areas of the manifold have shapes unlike any others										
Cluster	"Given a set of data cases, find clusters of similar attribute values"	Areas of the manifold have similar shapes										
Correlate	"Given a set of data cases and two attributes, determine useful relationships between the values of those attributes"	1D vs 2D relationships										

Conclusion

- 1D slices – sliceplorer
- 2D slices – hypersliceplorer
- Definition and challenges of multi-dimensional visualization



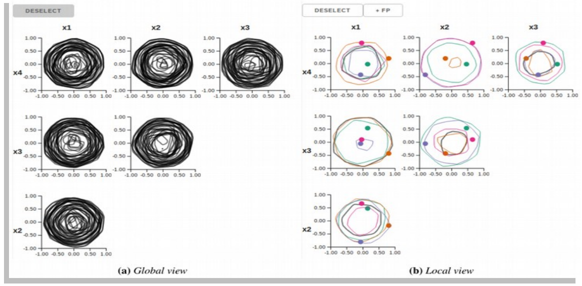


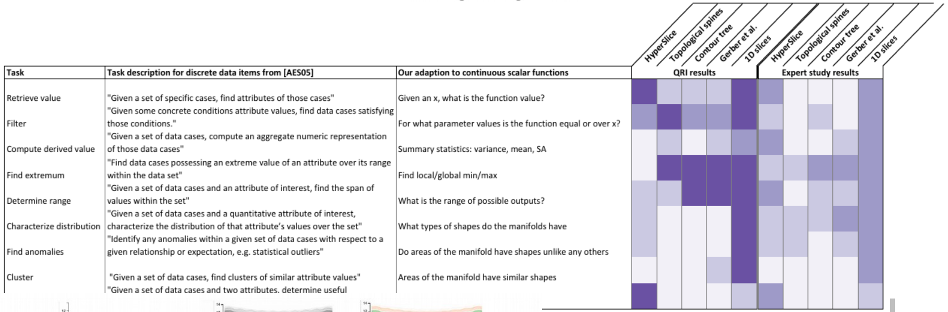
Thanks!

www.tomtorsneyweir.com

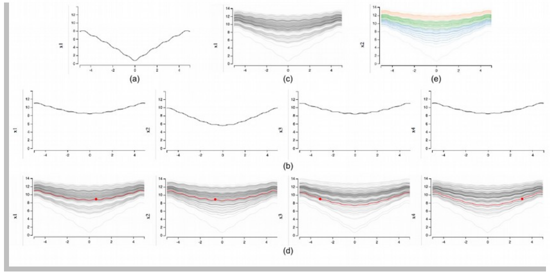
t.d.torsney-weir@swansea.ac.uk

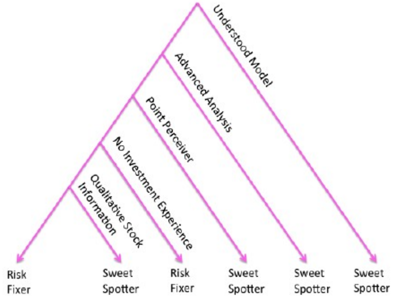
@gabysbrain

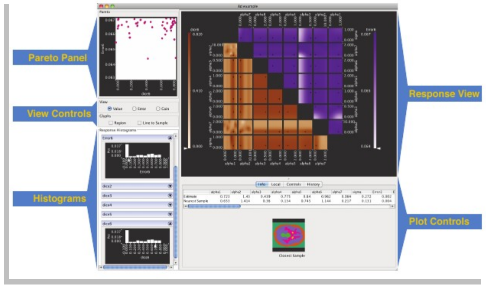


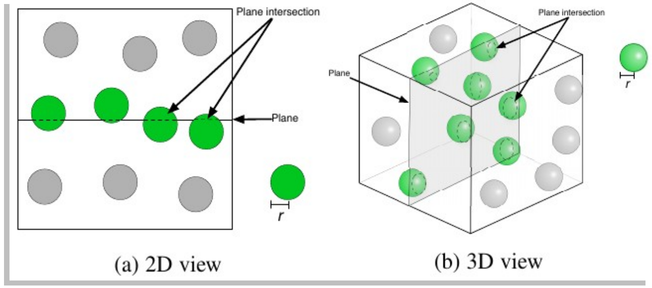


Task	Task description for discrete data items from [AES05]	Our adaption to continuous scalar functions
Retrieve value	"Given a set of specific cases, find attributes of those cases"	Given an x, what is the function value?
Filter	"Given some concrete conditions attribute values, find data cases satisfying those conditions."	For what parameter values is the function equal or over x?
Compute derived value	"Given a set of data cases, compute an aggregate numeric representation of those data cases"	Summary statistics: variance, mean, SA
Find extremum	"Find data cases possessing an extreme value of an attribute over its range within the data set"	Find local/global min/max
Determine range	"Given a set of data cases and an attribute of interest, find the span of values within the set"	What is the range of possible outputs?
Characterize distribution	"Given a set of data cases and a quantitative attribute of interest, characterize the distribution of that attribute's values over the set"	What types of shapes do the manifolds have
Find anomalies	"Identify any anomalies within a given set of data cases with respect to a given relationship or expectation, e.g. statistical outliers"	Do areas of the manifold have shapes unlike any others
Cluster	"Given a set of data cases, find clusters of similar attribute values"	Areas of the manifold have similar shapes
	"Given a set of data cases and two attributes, determine useful	





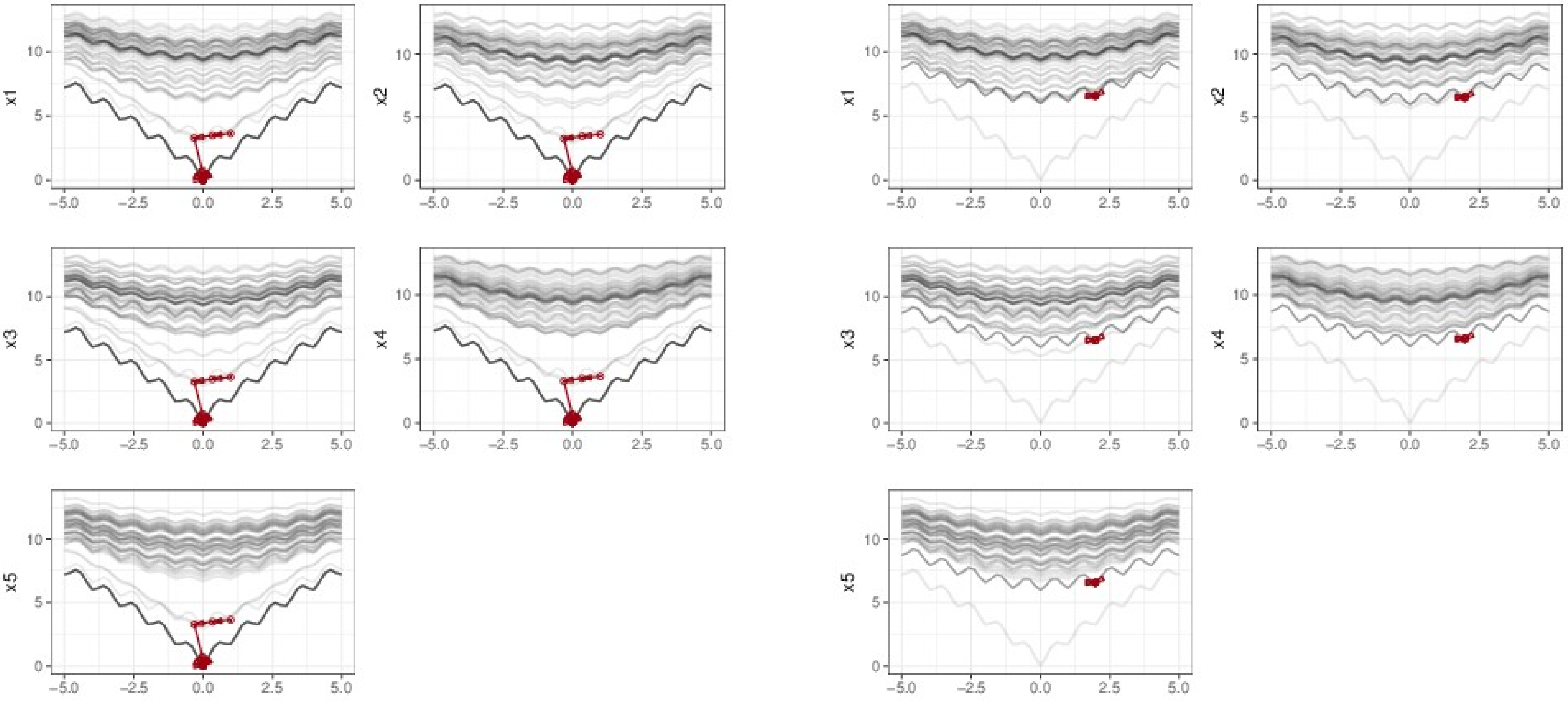




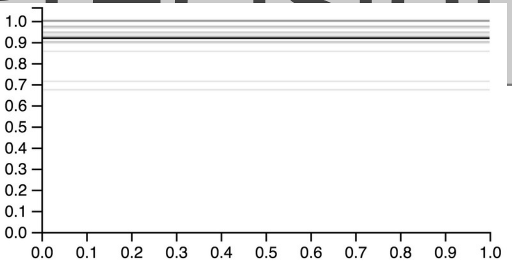
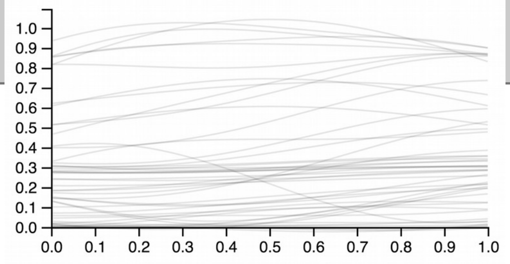
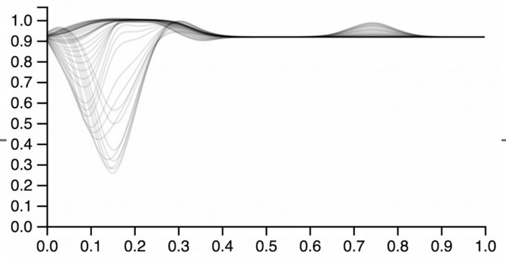
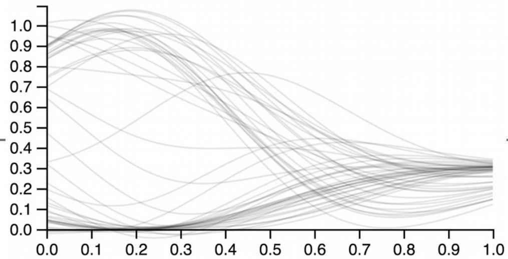
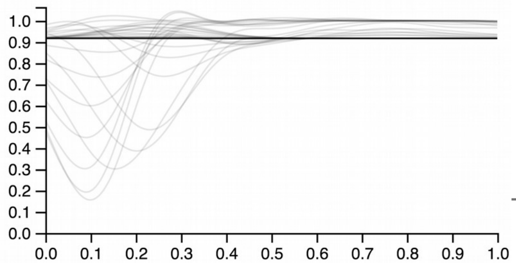
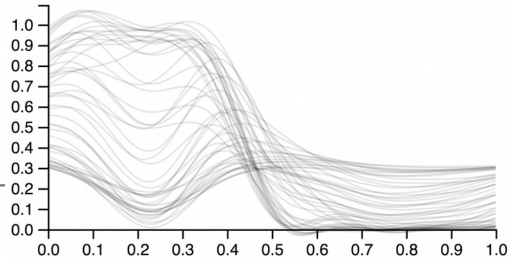
(a) 2D view (b) 3D view

Optimization functions

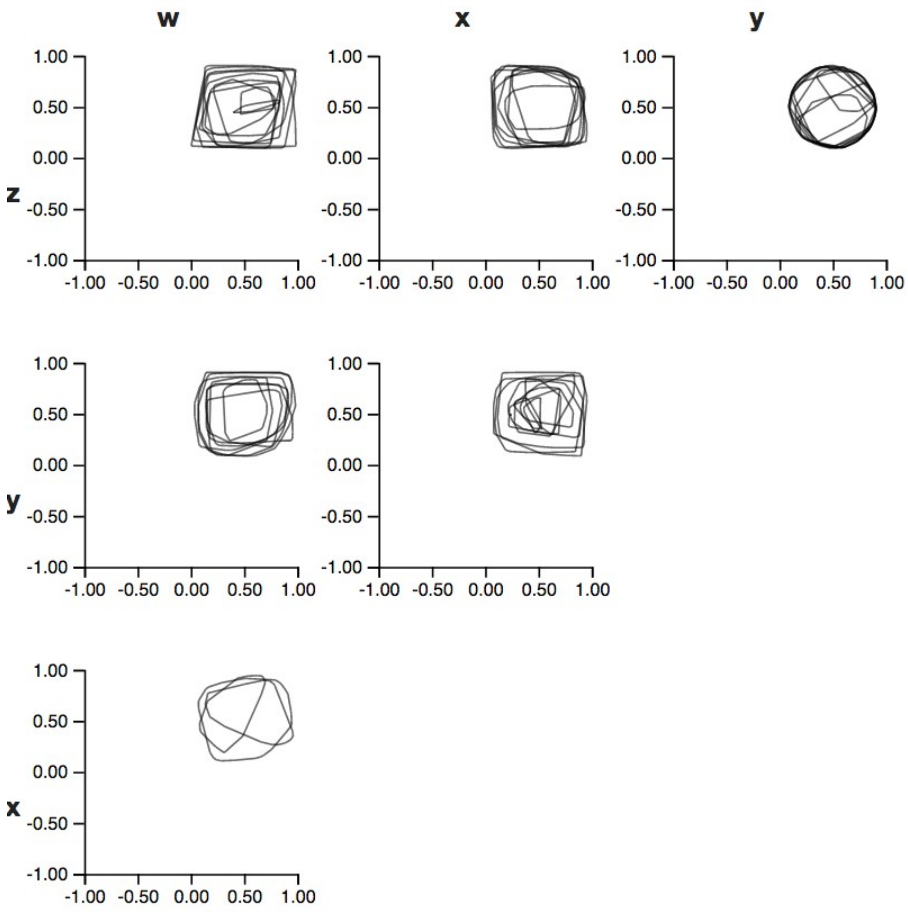
Torsney-Weir et al.: Sliceplorer

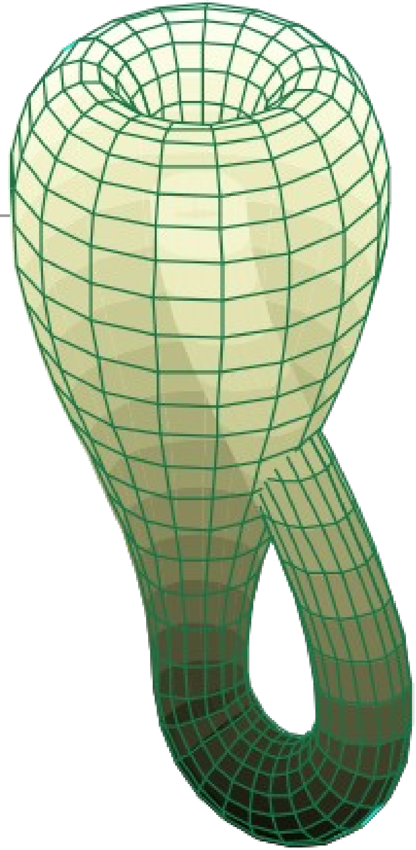


performance

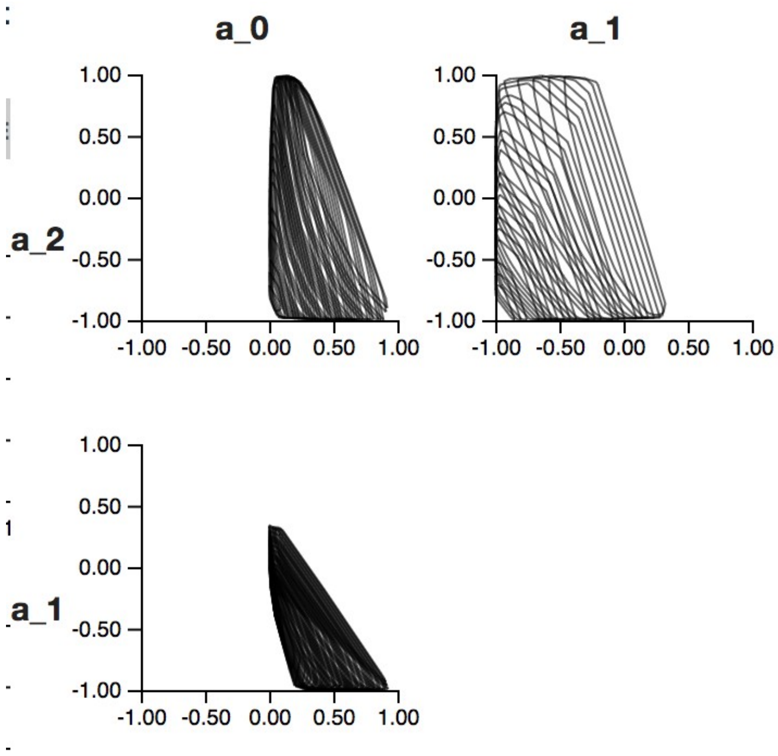
	Precision	Recall
p1		
p2		
p		

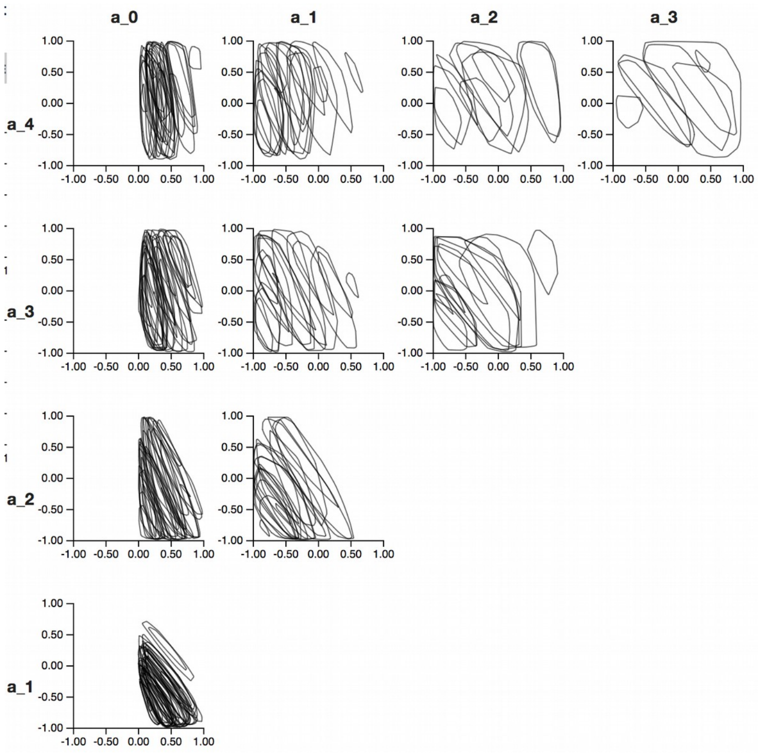
Klein bottle





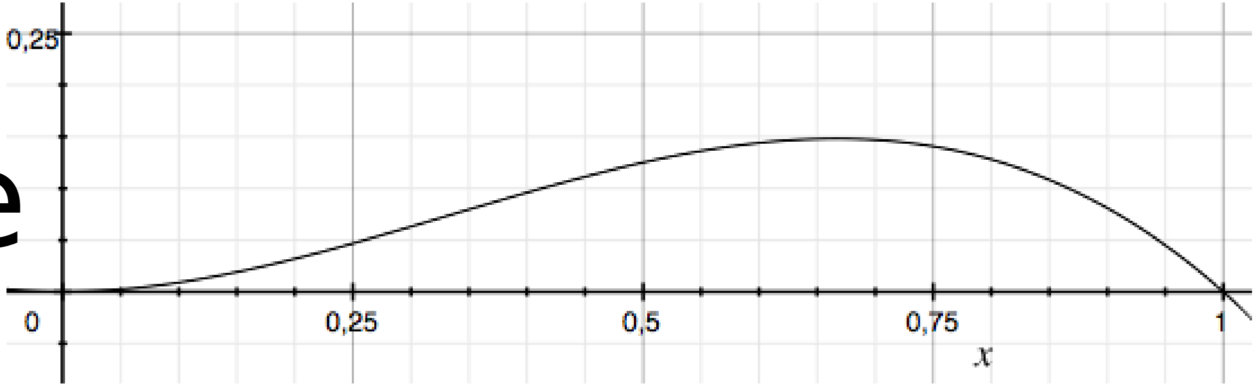
Comparing spaces



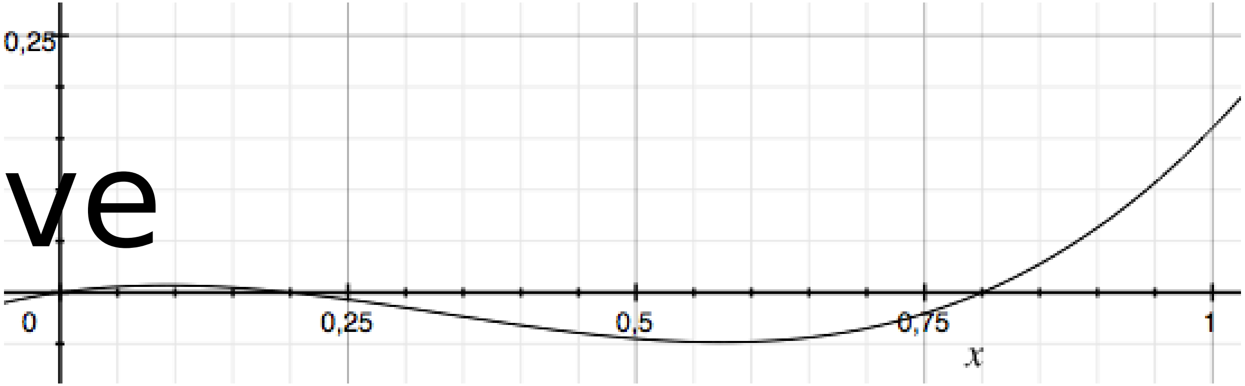


Positive polynomials

Positive



Not positive



Bernstein polynomials

$$b_{v,n} = \binom{n}{v} x^v (1-x)^{n-v}$$

$$B_n(x) = \sum_{v=0}^{n} \beta_v b_{v,n}(x)$$

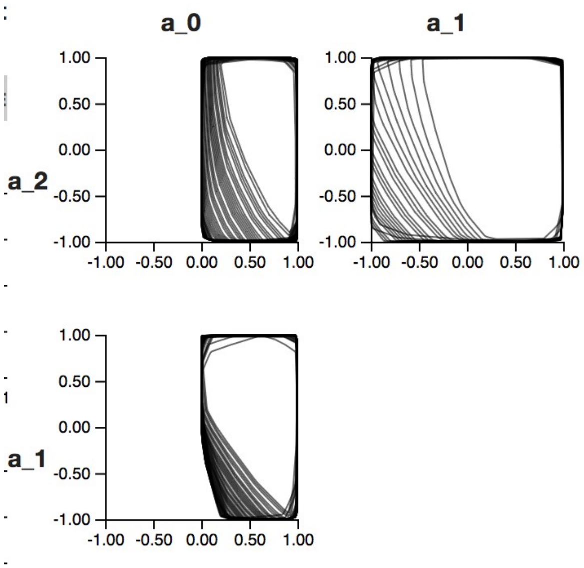
$$\beta_v \geq 0, v = 0, \ldots, n$$

Experiment

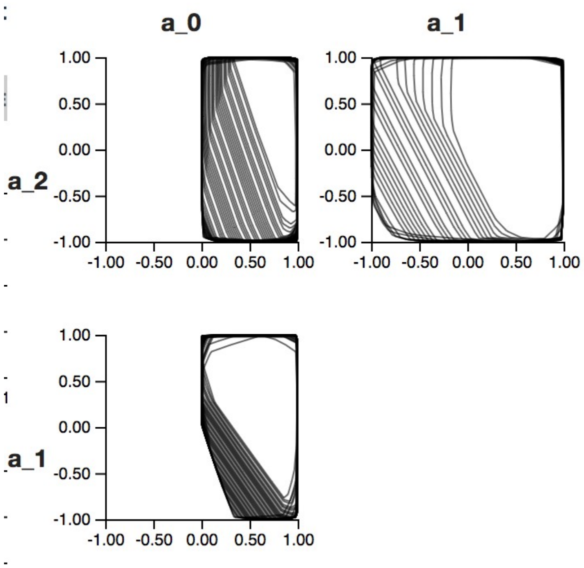
1. Pick a polynomial of degree d

2. Set one of the d+1 coefficients to 1

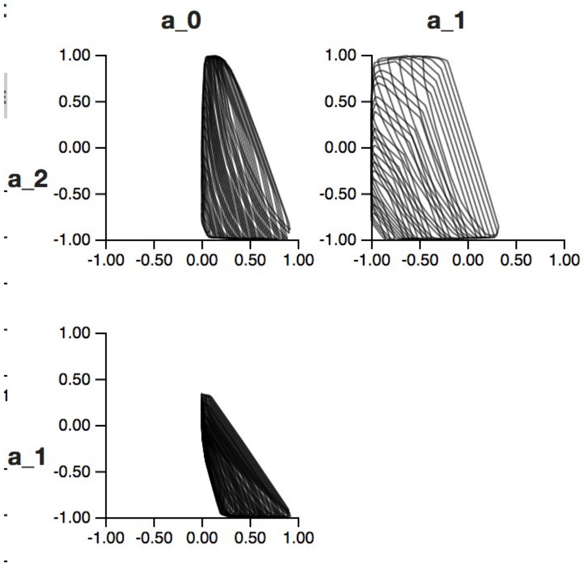
$$a_0 + a_1 x + a_2 x^2 + 1x^3$$



Positive

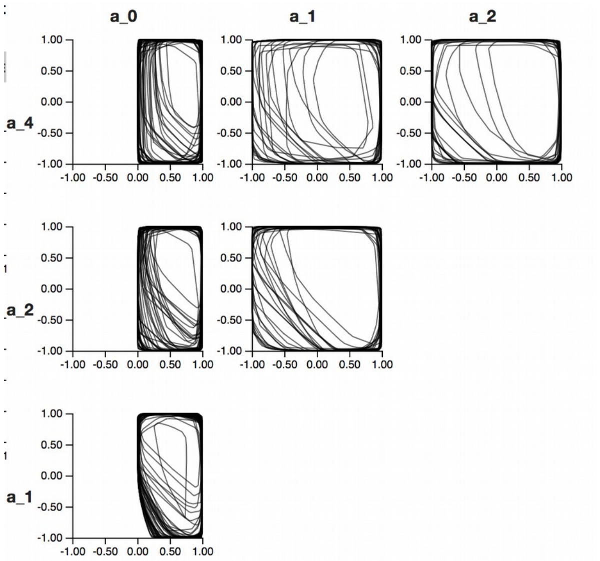


Bernstein

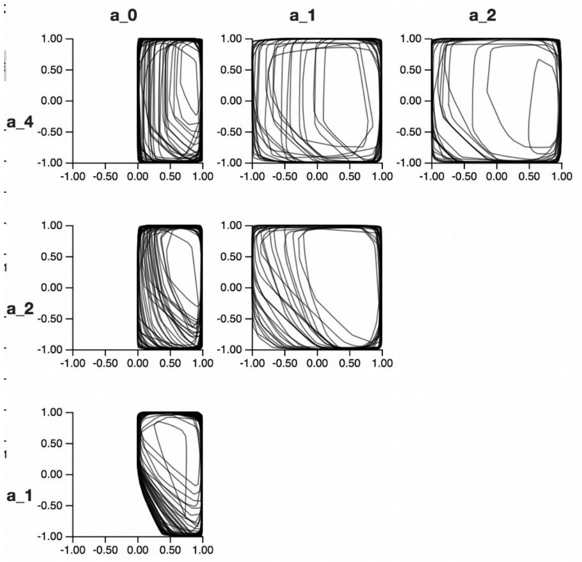


Difference

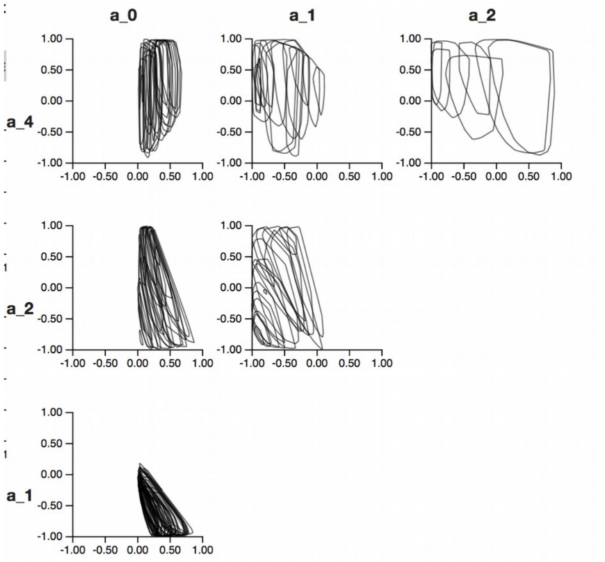
$$a_0 + a_1 x + a_2 x^2 + 1x^3 + a_4 x^4$$



Positive



Bernstein



Difference