



# 設計空間特徴抽出へのDecisionSiteの適用

- Multidisciplinary Design Exploration (MDE) Lecture Series 2 -

酒井浩行 – Manager, Sr. Application Consultant

June 8<sup>th</sup>, 2007 @ 東北大学





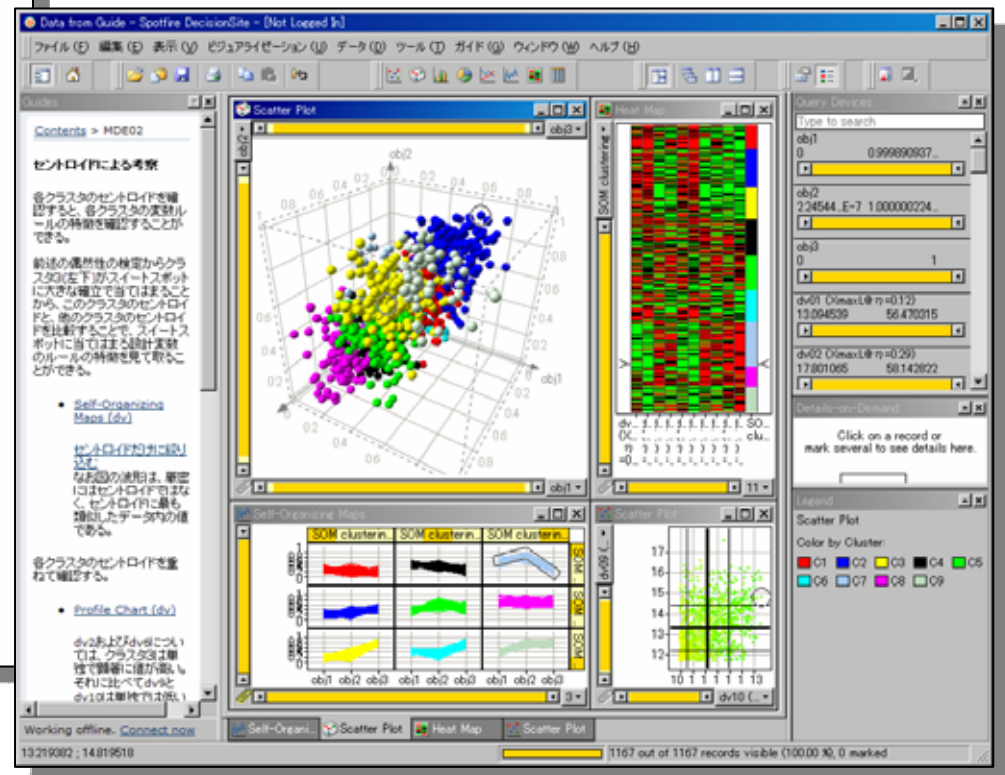
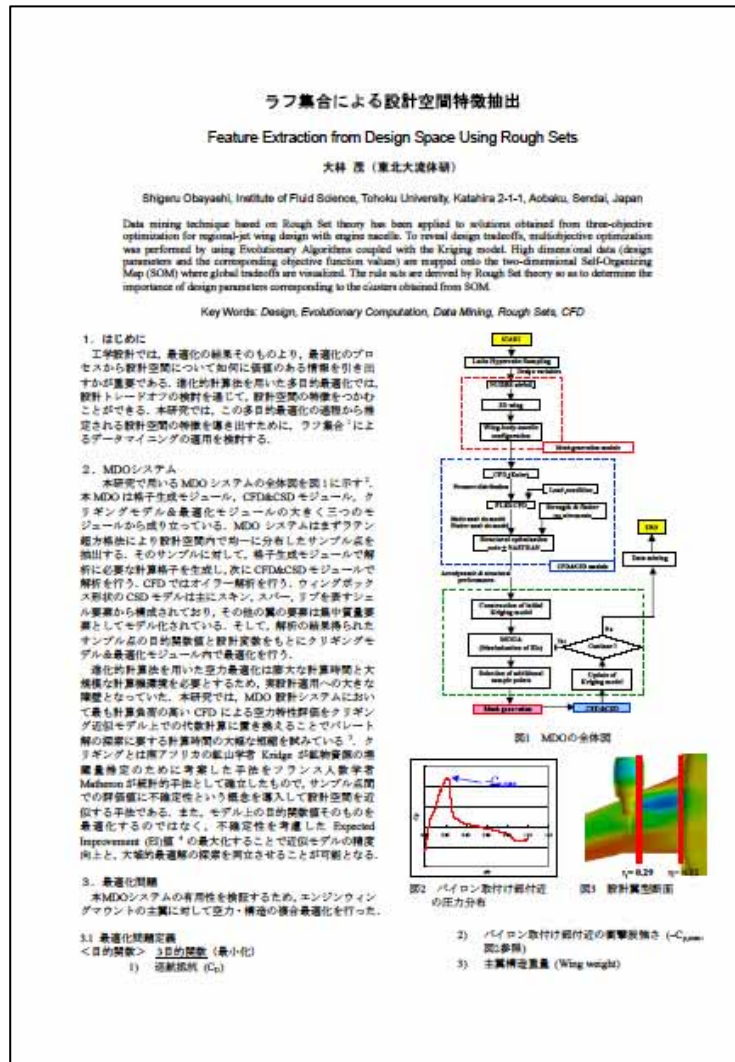
# はじめに

- 東北大学大林茂教授による「ラフ集合による設計空間特徴抽出」の手法に、Spotfire DecisionSite(以下DS)の適用を試みる。
- SOMはじめ、変数のルールを可視化するためにプロフィールチャートを適用する。
- DSに備わるSOMを適用する。
- 注意点
  - SOMの適用結果が、上記論文の結果と一部異なります。これは適用アルゴリズム、または/かつSOM実行時のパラメータ等の違いと考えられますが、詳細の調査が必要です。
  - 今回は、設計変数のSOMの結果がSOMineの結果と異なることが予想されたため、DS単独での解析を試み、一定の結論に達しました。設計変数のSOMineによるSOM番号があれば、同じ手法を検定できます。



# 論文「ラフ集合による設計空間特徴抽出」

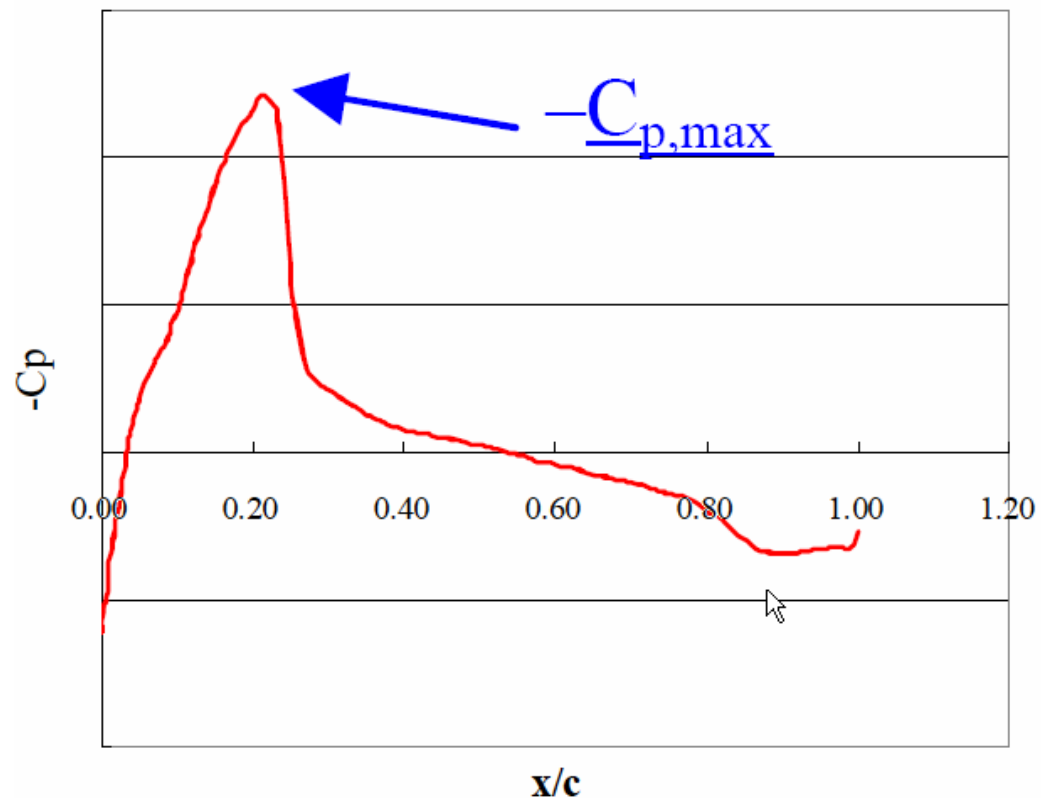
- 設計空間の特徴を導き出す
  - 自己組織化マップ (SOM)
  - 分散分析 (ANOVA)
  - ラフ集合 (データの離散化)
  - モデル化 (Rosetta)





# 目的関数

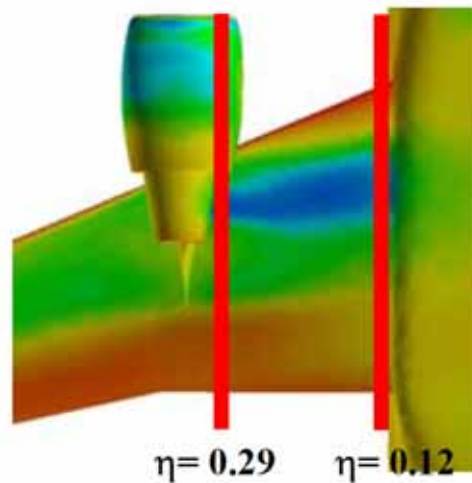
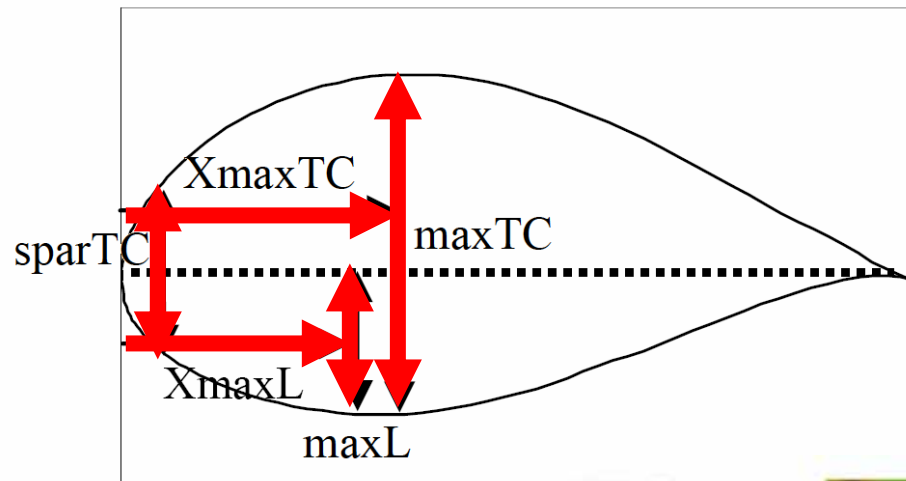
- 以下の目的関数の最小化を目指す
  - 巡航抵抗 (CD)
  - パイロン取付け部付近の衝撃波強さ ( $-C_{p,max}$ )
  - 主翼構造重量 (Wing weight)





# 設計変数

- 翼断面の設計値



- $X_{maxL}$ 
  - 翼下面最大厚み位置の前縁からの距離
- $maxL$ 
  - 翼下面最大厚み
- $X_{maxTC}$ 
  - 翼型最大厚み位置の前縁からの距離
- $maxTC$ 
  - 翼型最大厚み
- $sparTC$ 
  - 前桁位置における翼の厚み



# 方針

- 以下三段階の考察を試みる
  - 目的関数によるSOMとデータの離散化からの考察
  - 決定属性を満たす理想空間とANOVAの適用による考察
  - 設計変数によるSOMと偶然性の検定の適用による考察
    - セントロイドによる考察

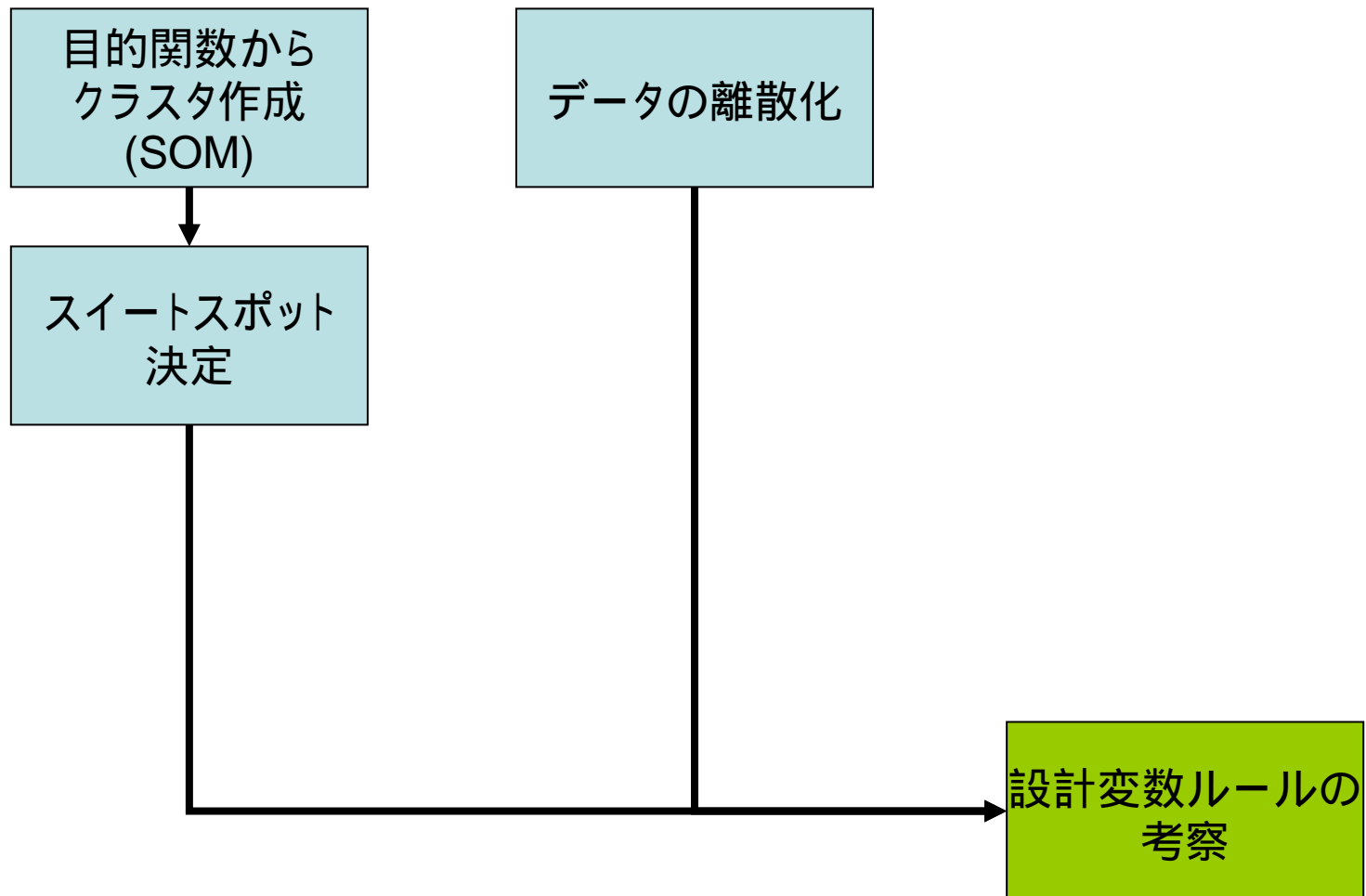




## 目的関数のSOMとデータの離散化

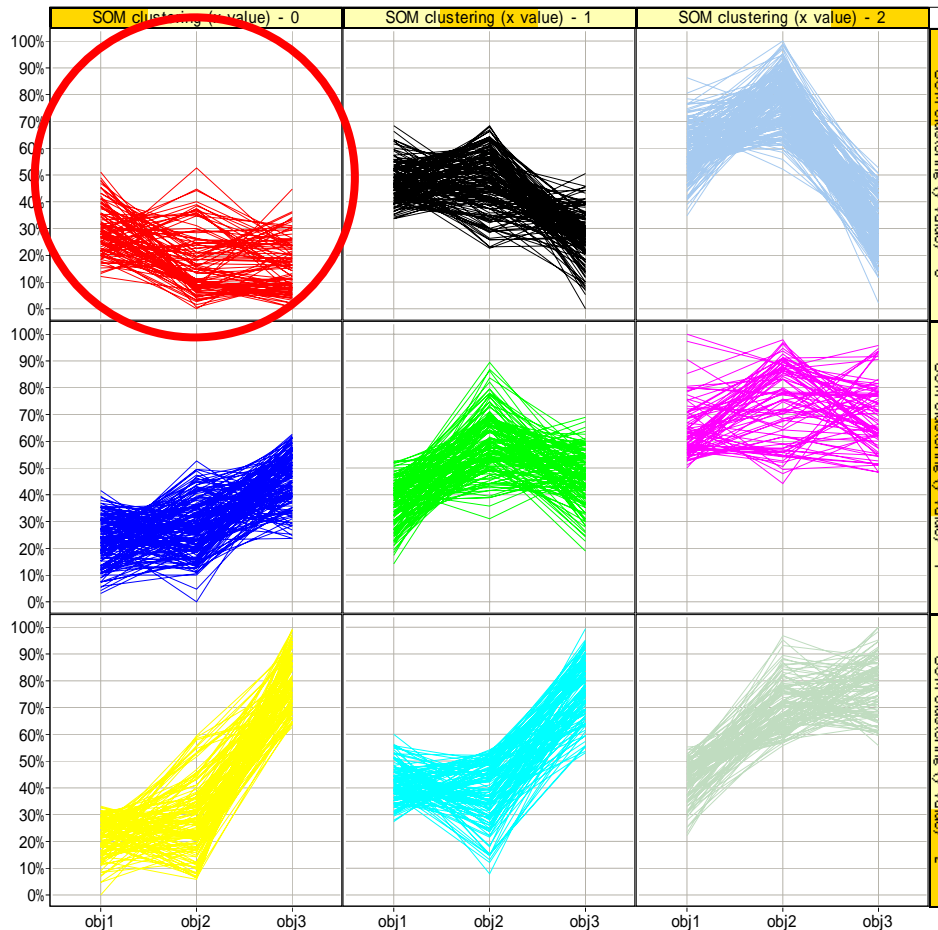


# 解析フロー





# 目的関数からSOMを作成



目的関数に対してSOMを実行。グリッドサイズは同様の3x3とした。  
DecisionSiteのSOMではプロファイルチャートを用いるが、目的関数ごとに色を変更することなく、左上のクラス(1)がスイートスポットであることがわかる。

なお縦軸は各目的関数の値を正規化して表示している。

Color by SOM clustering:

1 2 3 4 5 6 7 8 9

All columns use individual scales.

Generated by Self-Organizing Maps

Grid size (width x height): 3 x 3

Neighborhood function: Bubble  
Radius (begin x end): 2.5 x 0

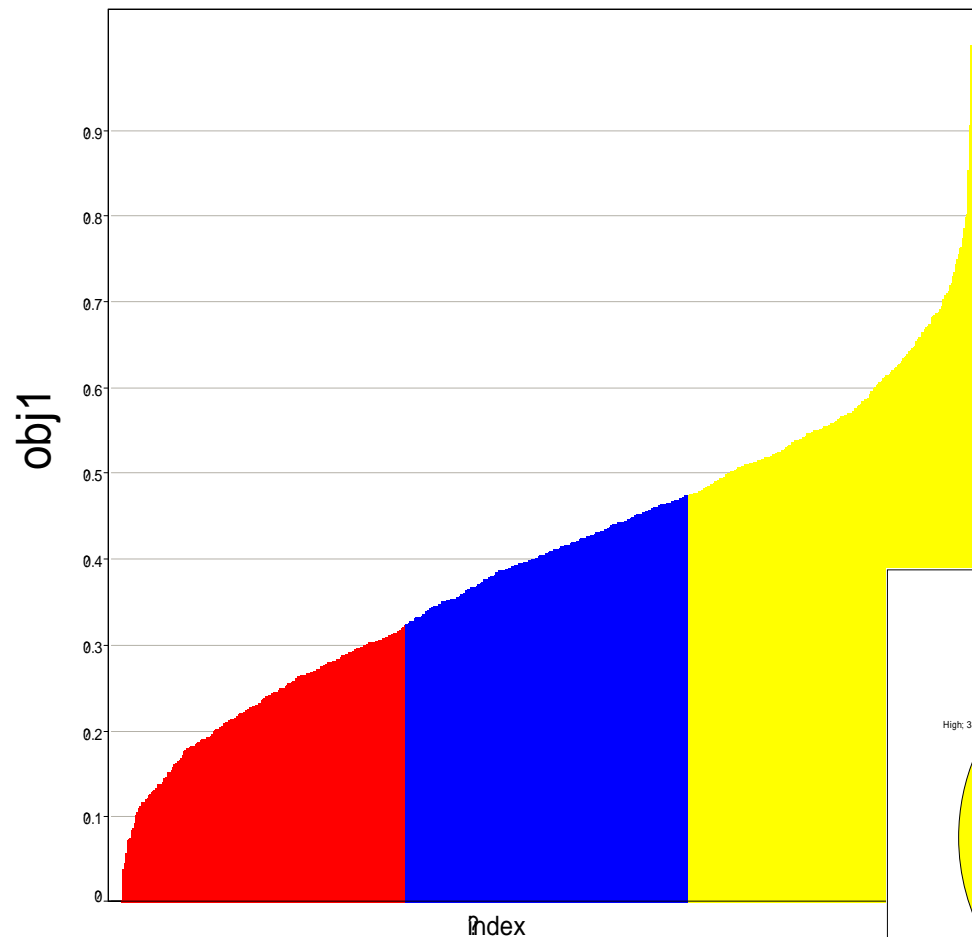
Learning function: Linear  
Initial rate: 0.05

Number of training steps: 12500

Output parameters:  
Mapping Precision: 0.1693  
Topology Preservation: 1.714e-3



# データの離散化



設計変数の離散化を実行し、ヒストグラムにて離散化後の分布を確認する。

なお離散化のルールは、各変数を三分割し、各区間でのデータ数が同じになるように分割幅を自動的に定めた。

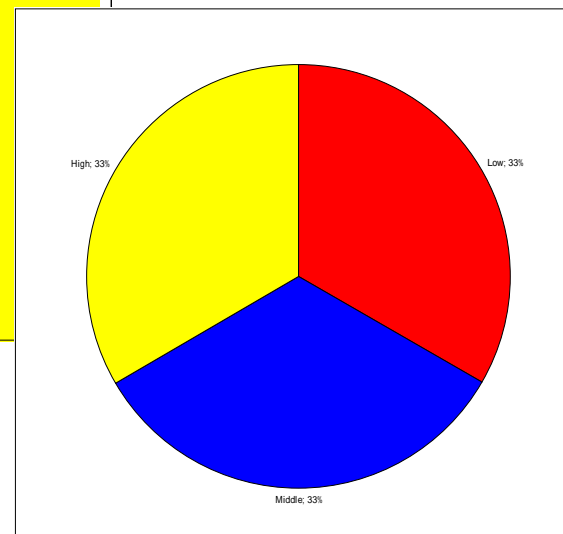
The height of a bar represents the sum of obj1.

Color by obj1 (ラフ集合):

Low Middle High

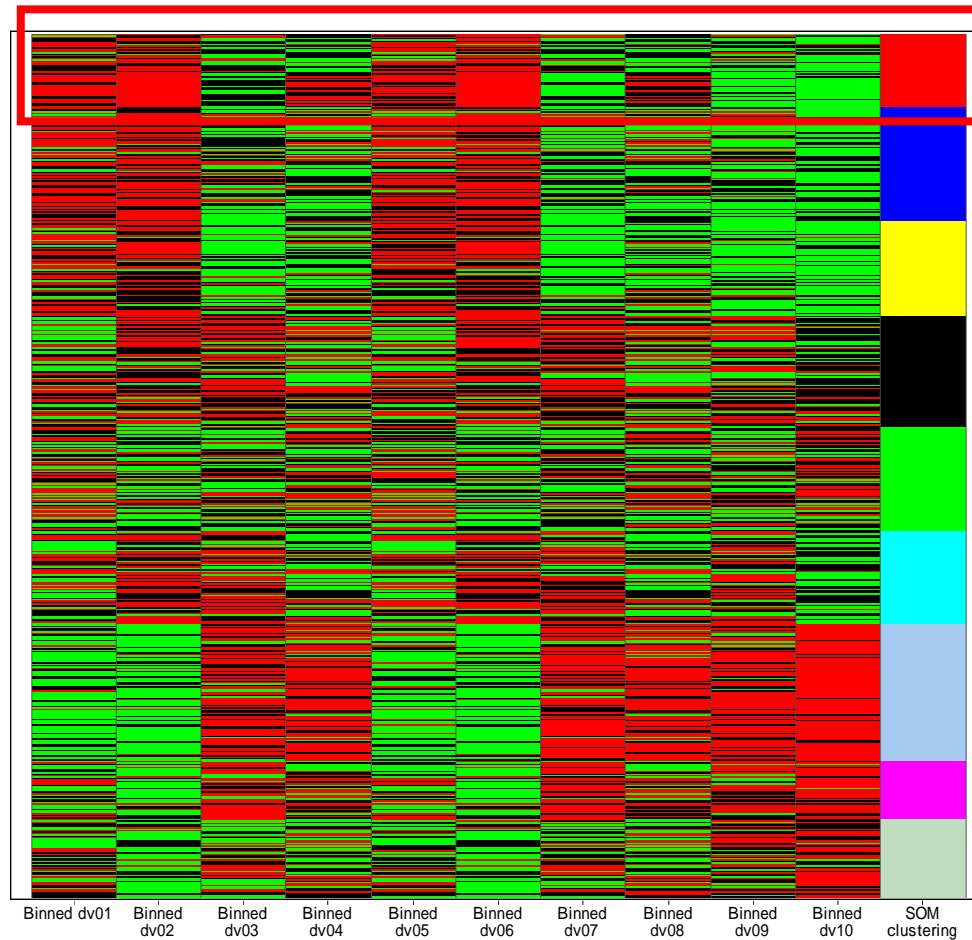
The labels show the height of each bar.

The x-axis values are sorted.





# データの離散化



設計変数の離散化を実行し、Heat Mapにてクラスタごとの分布を確認する。スイートスポットの設計変数の分布が読める。

なお離散化のルールは同様に、各変数を三分割し、各区間でのデータ数が同じになるように分割幅を自動的に定めた。

Low Middle High

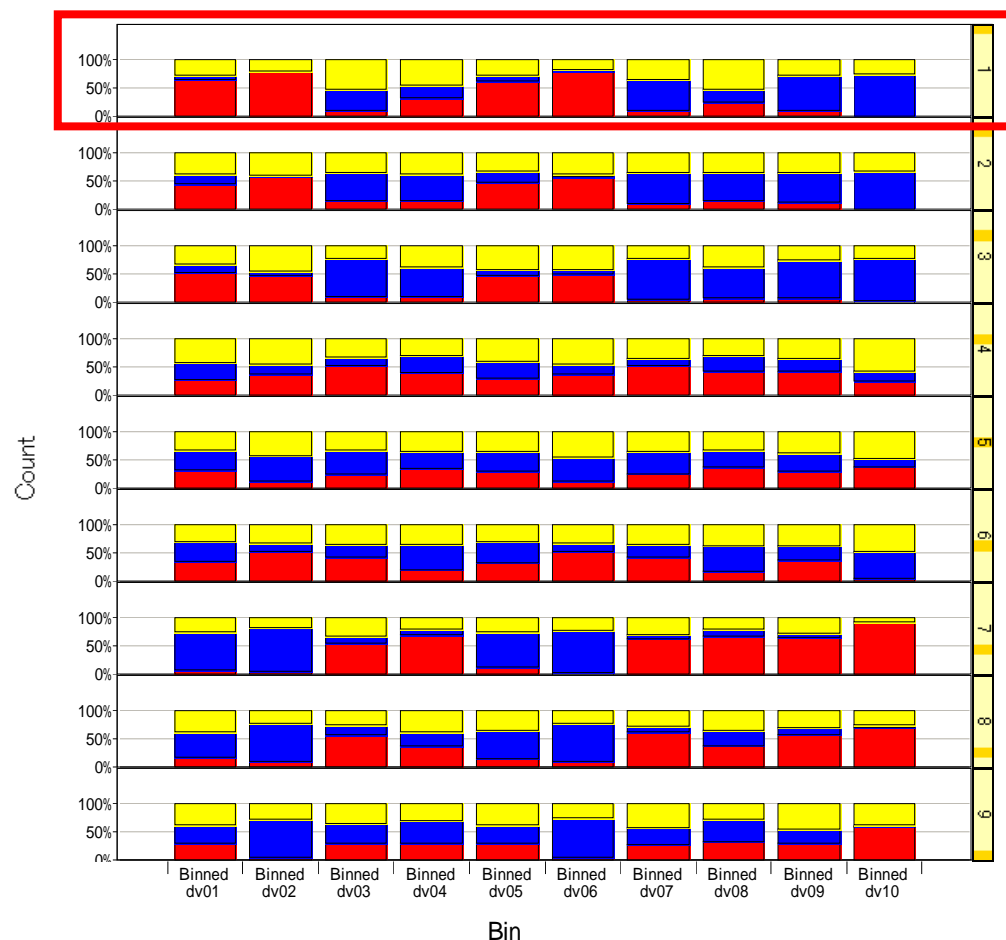
SOM clustering  
Colored by SOM

1 2 3 4 5 6 7 8 9

Order by SOM clustering



# 可視化による設計変数ルールの考察 (1)



左図の第一行目がスイートスポットである。縦軸は各離散値の割合を示し、横軸に各設計変数を並べた。

スイートスポットで特徴的なのは、  
・dv2およびdv6の約80%がHigh(大)  
・dv9およびdv10の約70%がLow(小)  
となっており、他のクラスタに比較すると特にdv2およびdv6にその傾向が顕著である。

The height of a bar represents the number of records.

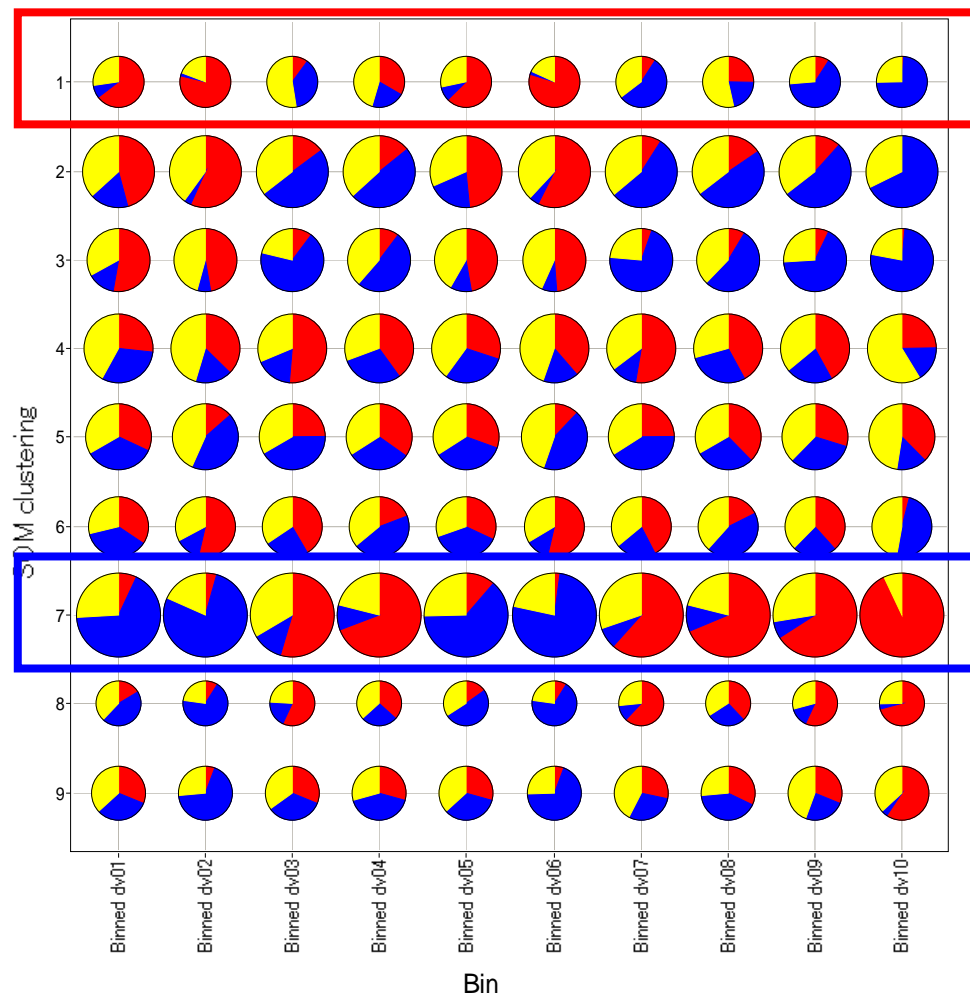
Color by Group:

High Low Middle

The labels show the height of each bar.



# 可視化による設計変数ルールの考察 (2)



左図の第一行目がスイートスポットである。色は離散値の割合を示すし、円のサイズはレコードの数を表す。

スイートスポットで特徴的なのは、  
・dv2およびdv6の約80%がHigh(大)  
・dv9およびdv10の約70%がLow(小)  
となっており、他のクラスタに比較すると特にdv2およびdv6にその傾向が顕著である。

こちらのチャートでは、クラスタ7の特徴も読める。

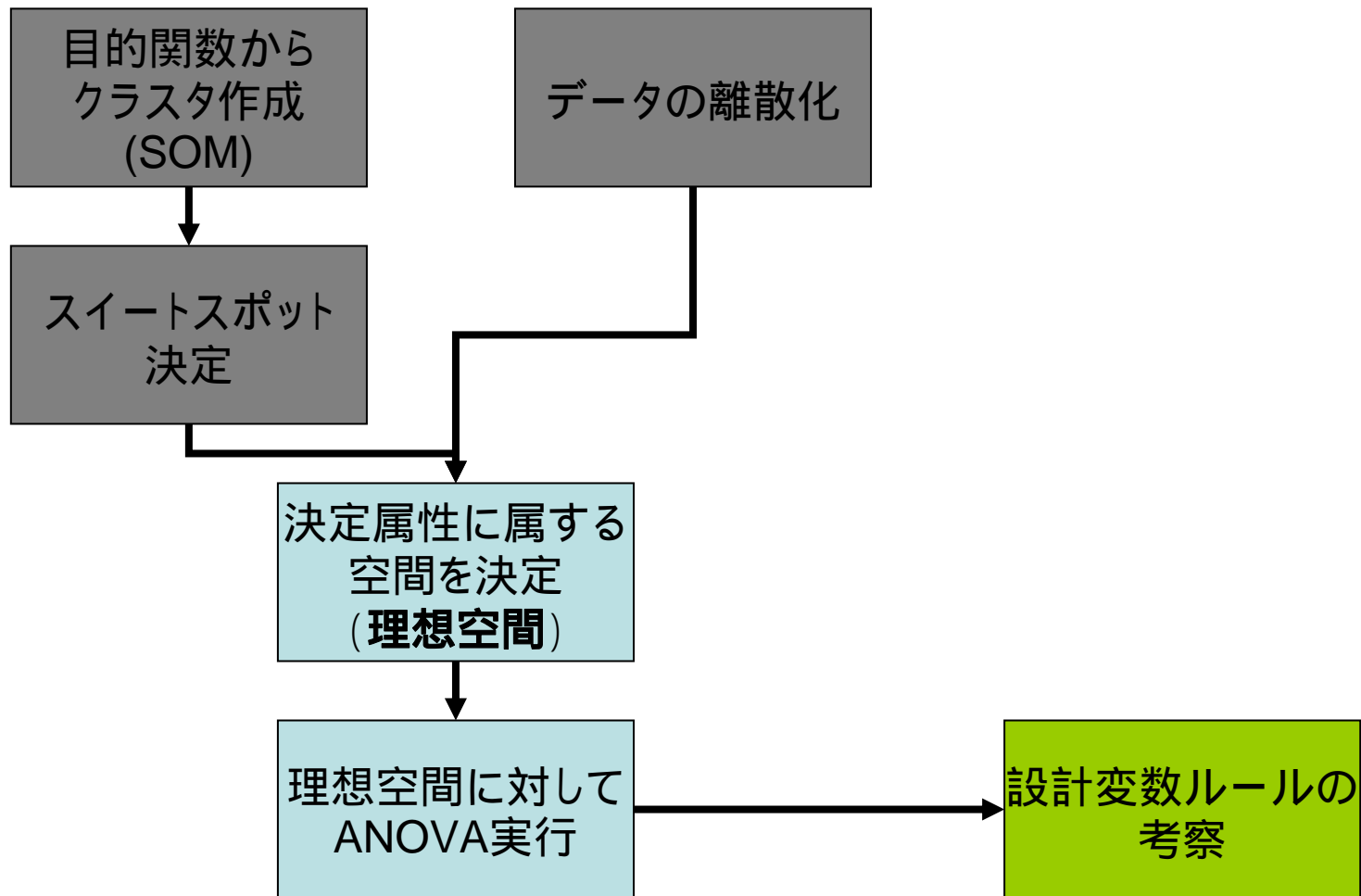


# 決定属性とANOVAの適用



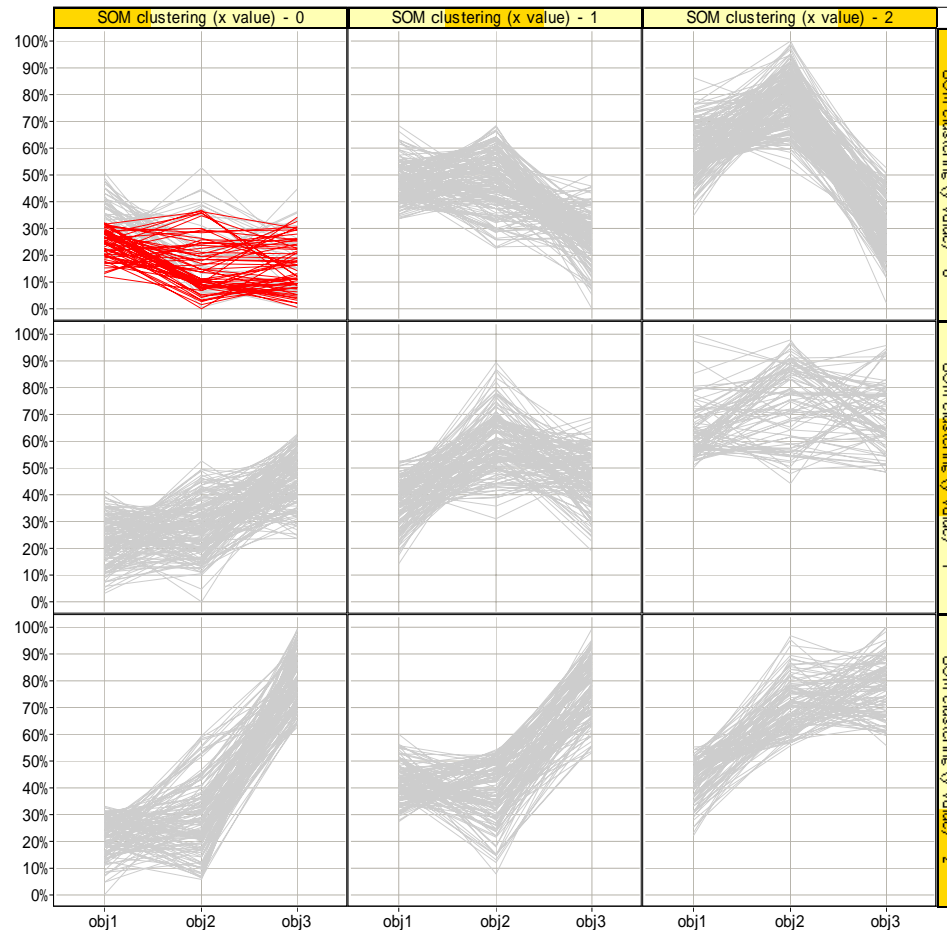


# 解析フロー





# 決定属性に属する理想空間を決定

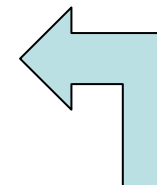


決定属性を同様に  
・各目的関数が「Low(小)」に属する  
・スイートスポットに属する  
とし、この空間を**理想空間**とする。

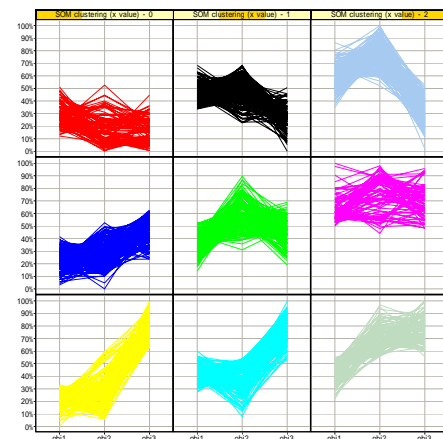
Color by SOM clustering:

1 2 3 4 5 6 7 8 9

All columns use individual scales.



決定属性に従うデータだけを抽出する。





# 理想空間に対してANOVA実行 (1)

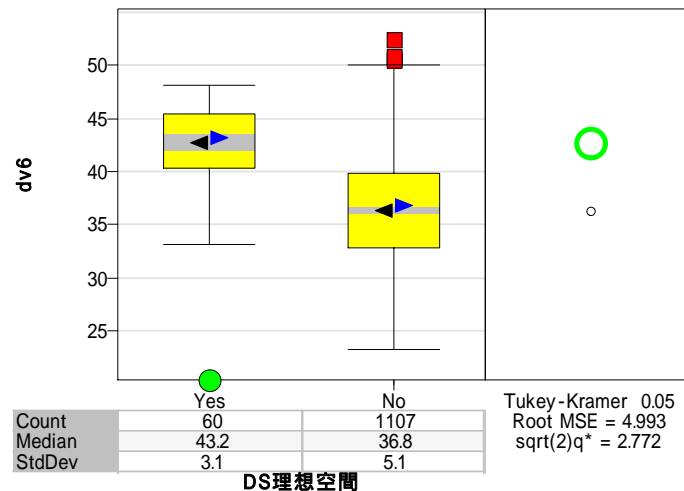
Y (numerical)	X (categorical)	p-value
dv6	DS理想空間	6.1e-021
dv2	DS理想空間	9.917e-018
dv10	DS理想空間	8.701e-016
dv9	DS理想空間	9.678e-015
dv5	DS理想空間	9.414e-010
dv1	DS理想空間	2.142e-009
dv7	DS理想空間	3.211e-008
dv3	DS理想空間	6.032e-005
dv4	DS理想空間	0.3271
dv8	DS理想空間	0.8008

理想空間に対して、各設計変数をANOVAにて検定した。

dv6を筆頭にdv3までが5%以下のP値を示したが、ここでは大きく桁の変わるdv9までを有意差のある設計変数と仮定する。

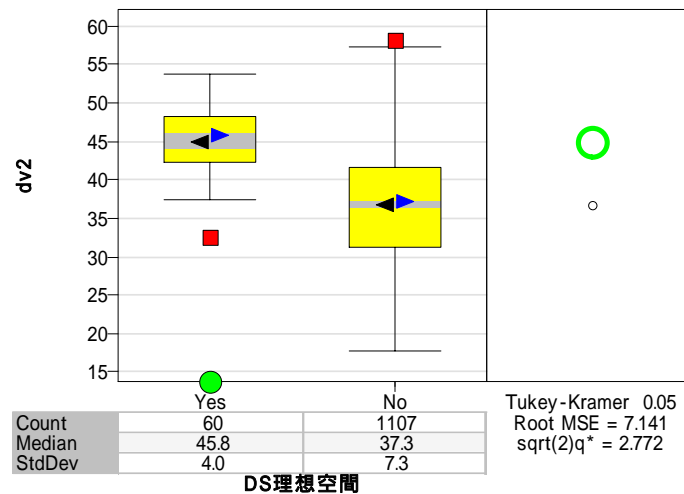


# 理想空間に対してANOVA実行 (2)



上位一位、二位のdv6とdv2をBox Plotで比較する。横軸は理想空間とそれ以外の空間である。

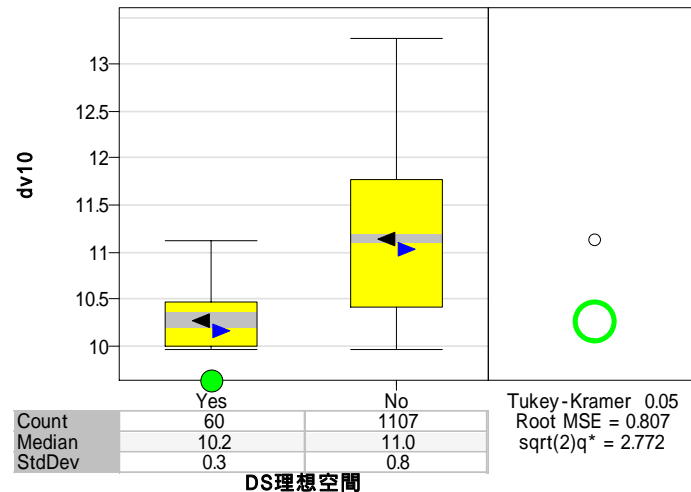
Comparison Circlesを見てもわかるとおり、理想空間とそれ以外の値で、大きな有意差があることが読み取れる。またdv6およびdv2を大きくとればいいことも推測できる。



Y (numerical)	X (categorical)	p-value
dv6	DS理想空間	6.1e-021
dv2	DS理想空間	9.917e-018
dv10	DS理想空間	8.701e-016
dv9	DS理想空間	9.678e-015
dv5	DS理想空間	9.414e-010
dv1	DS理想空間	2.142e-009
dv7	DS理想空間	3.211e-008
dv3	DS理想空間	6.032e-005
dv4	DS理想空間	0.3271
dv8	DS理想空間	0.8008



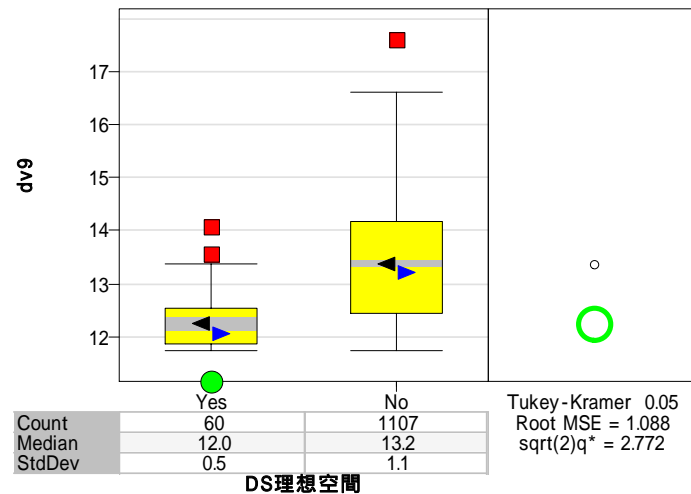
# 理想空間に対してANOVA実行 (3)



上位三位、四位のdv10とdv9をBox Plotで比較する。横軸は理想空間とそれ以外の空間である。

こちらComparison Circlesを見てもわかるとおり、理想空間とそれ以外の値で、大きな有意差があることが読み取れる。

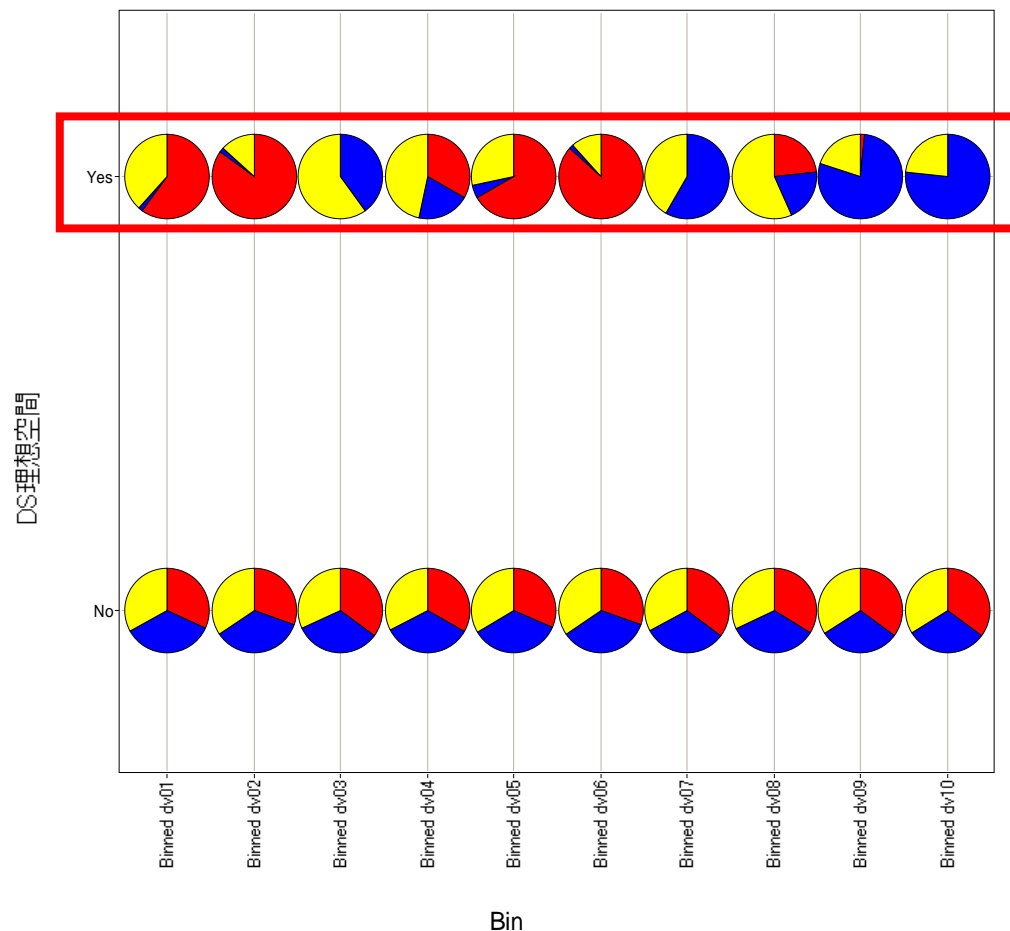
またdv10およびdv9を小さくとればいいことも推測できる。



Y (numerical)	X (categorical)	p-value
dv6	DS理想空間	6.1e-021
dv2	DS理想空間	9.917e-018
dv10	DS理想空間	8.701e-016
dv9	DS理想空間	9.678e-015
dv5	DS理想空間	3.414e-010
dv1	DS理想空間	2.142e-009
dv7	DS理想空間	3.211e-008
dv3	DS理想空間	6.032e-005
dv4	DS理想空間	0.3271
dv8	DS理想空間	0.8008



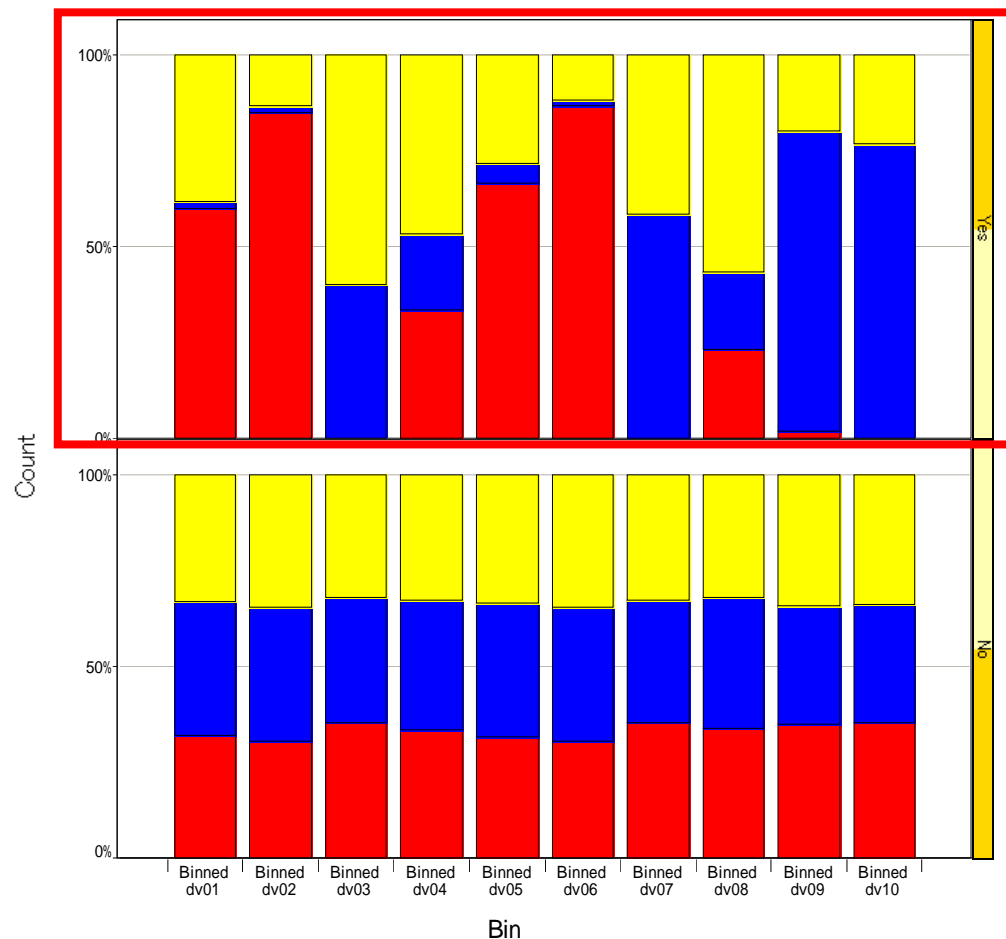
# 可視化による設計変数ルールの考察 (1)



Pie Chartで示してみる。やはり理想空間では、dv2、dv6、dv9、dv10の分布に前述の特徴が見られる。



## 可視化による設計変数ルールの考察 (2)



Bar Chartで示してみる。同じように理想空間では、dv2、dv6、dv9、dv10の分布に前述の特徴が見られる


The height of a bar represents the number of records.

Color by Group:

High Low Middle

The labels show the height of each bar.



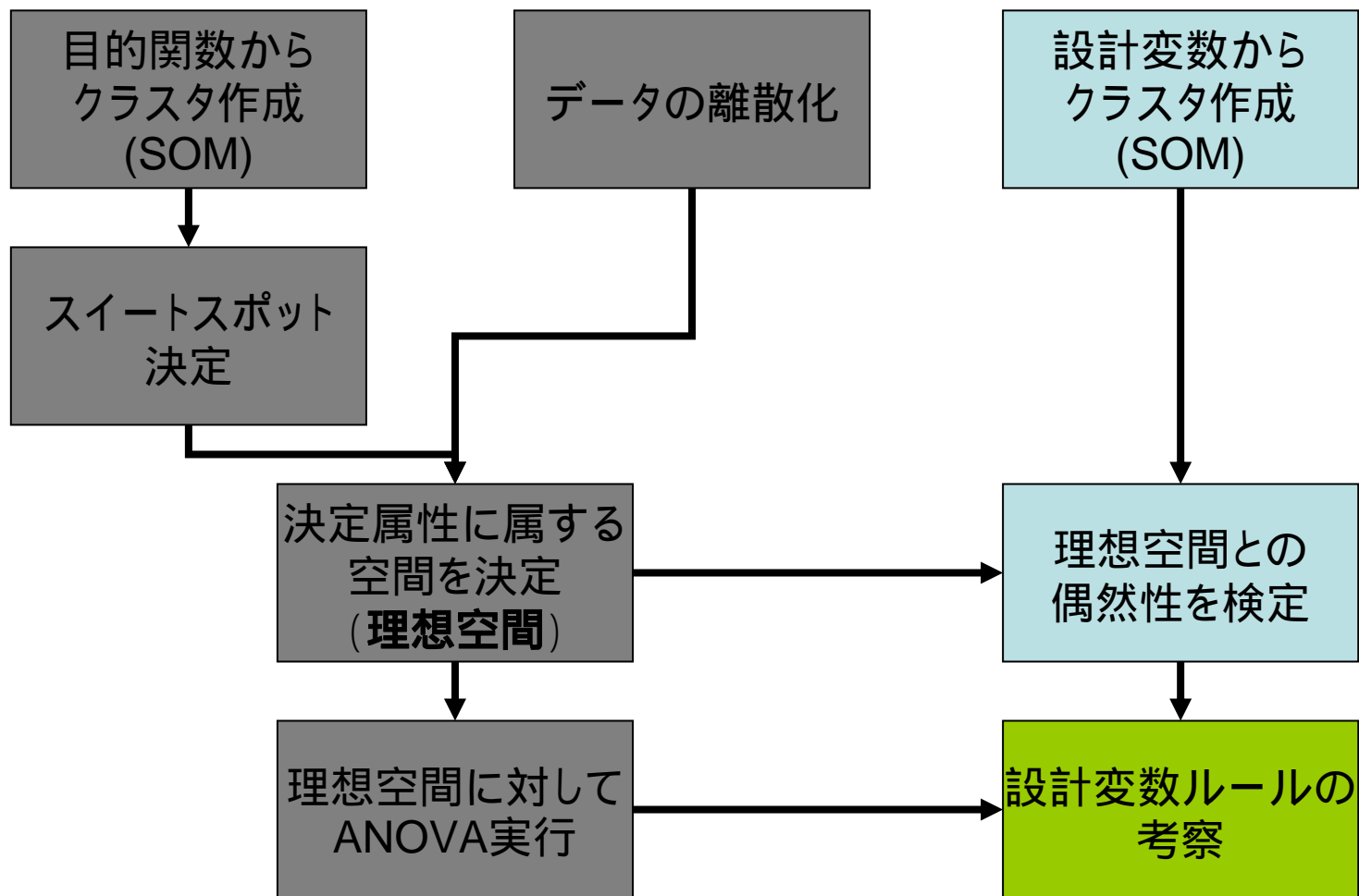


## 設計変数のSOMと偶然性の検定の適用



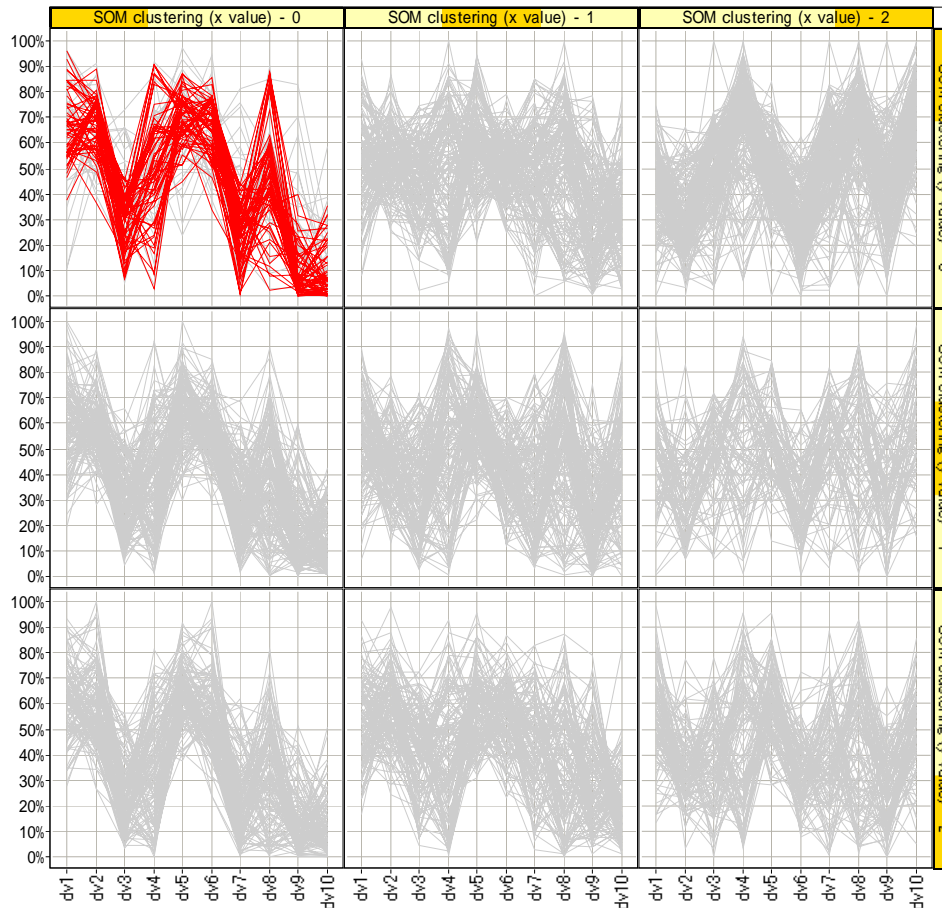


# 解析フロー





# 決定属性を満たす設計変数



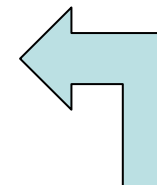
視点を変えて目的関数のSOMクラスタにしたがって、目的関数を構成する設計変数を分割してみる。

するとより設計変数のルールが明確に見える。

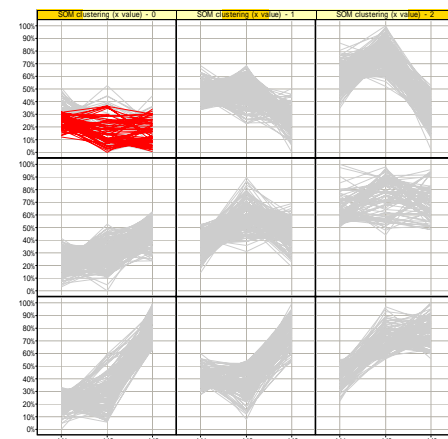
Color by SOM clustering:

1 2 3 4 5 6 7 8 9

All columns use individual scales.

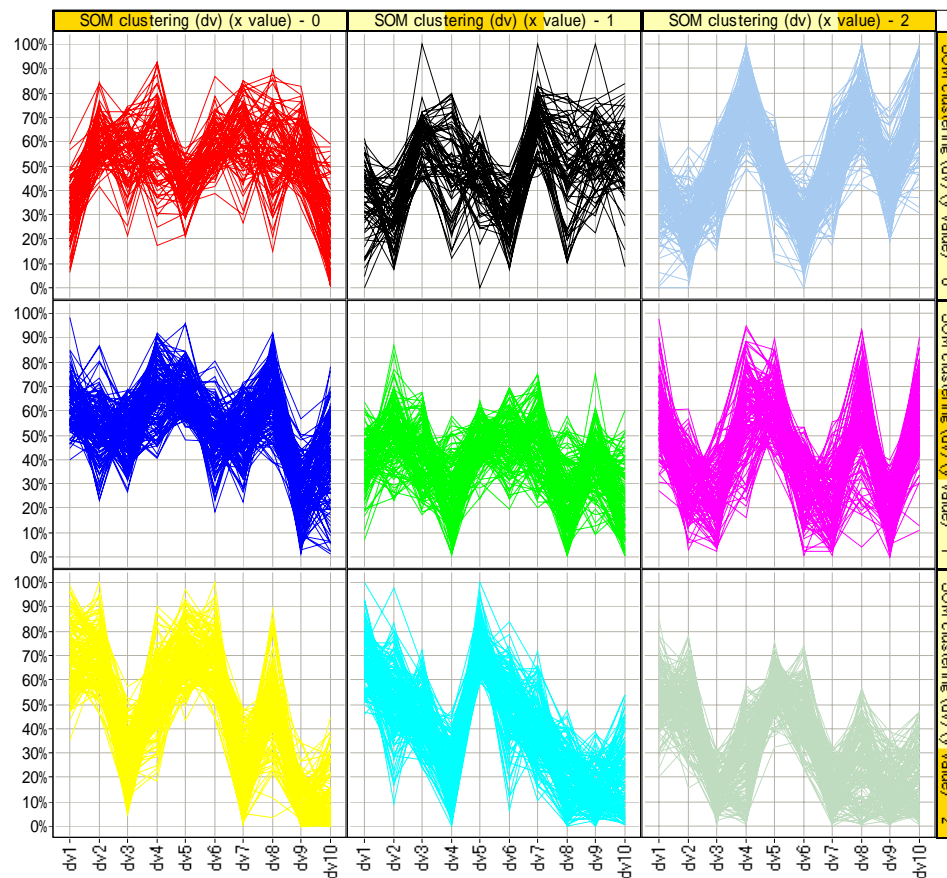


目的関数を構成する設計変数を、目的関数のSOMクラスタに従って分割してみる。





# 設計変数からSOMを作成



ここでは設計変数の特徴をより明確にするため、目的関数を無視し設計変数に対してSOMを実行してみる。

すると各設計変数の組み合わせの特徴を掴んで、クラスタリングされた。

Color by SOM clustering (dv):

1 2 3 4 5 6 7 8 9

All columns use individual scales.

Generated by Self-Organizing Maps

Grid size (width x height): 3 x 3

Neighborhood function: Bubble  
Radius (begin x end): 2.5 x 0

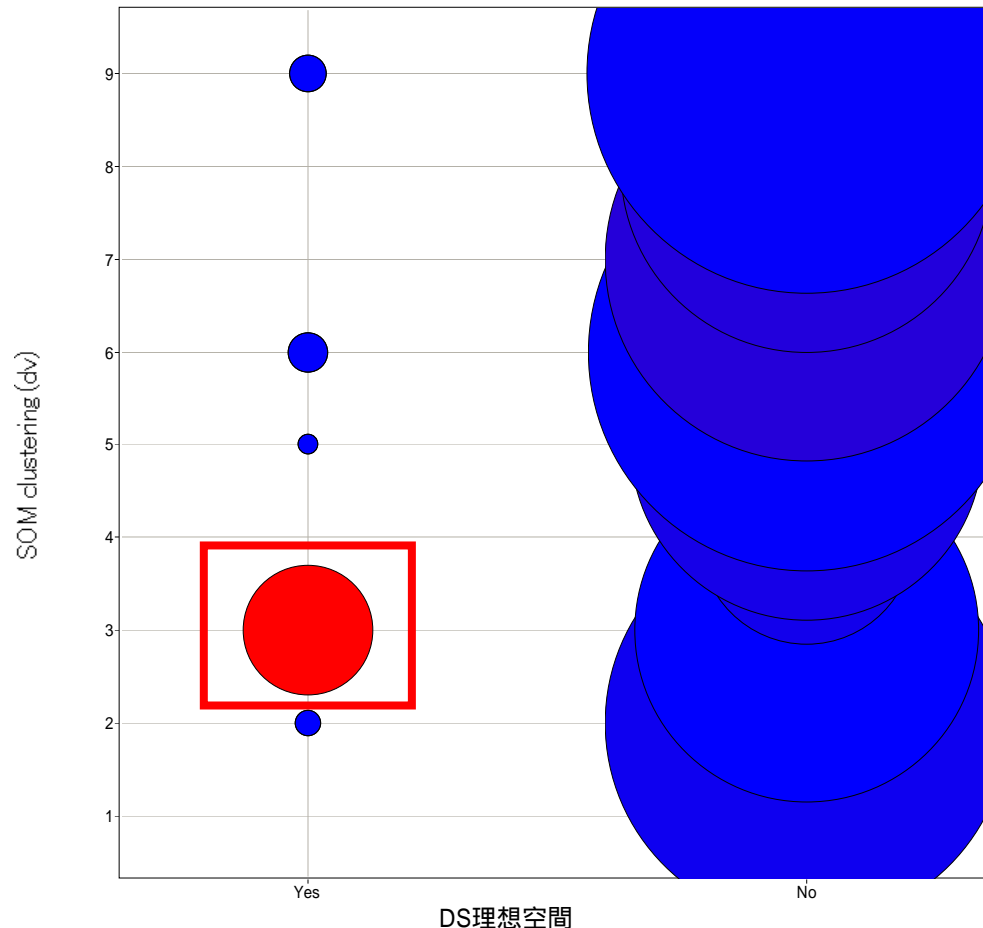
Learning function: Linear  
Initial rate: 0.05

Number of training steps: 12500

Output parameters:  
MappingPrecision: 0.3128  
TopologyPreservation: 0.1063



# 偶然性の検定



そこで設計変数のSOMクラスタに対し、先の理想空間への「偶然性の検定」を適用してみる。

ここではクラスタ3が理想空間となる偶然性が高いことがわかる。

Color by Coincidence DSTarget vs SOMdv:  
5.899761930758E-24



Size by records count



Labels show sector value; sector percentage

between categories for two columns.

Calculation settings:

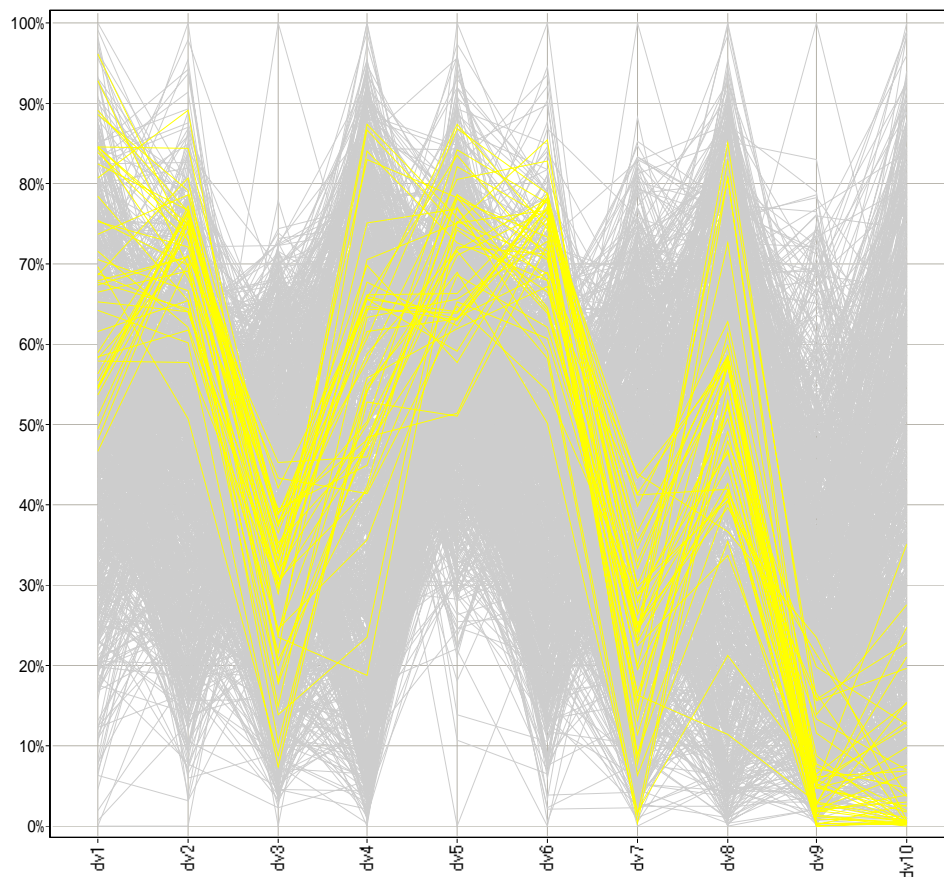
-----  
Total no of records: 1167 (All records)

First category column: DSTarget

Second category column: SOM clustering (dv)

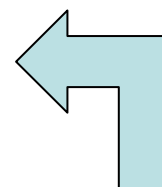


# 決定属性を満たす設計変数の考察 (1)

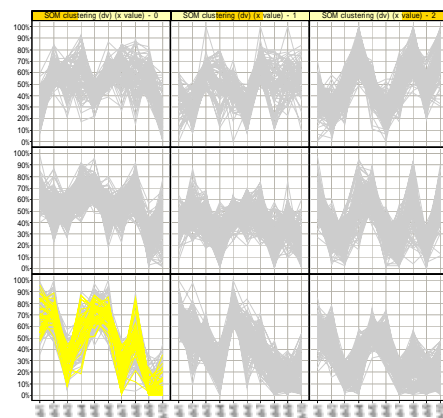


ここでも今までと同じように、dv2、dv6、dv9、およびdv10に特徴が見られる。

しかしそれぞれの設計変数単独で目的関数への影響があるとはいえず、これら設計変数の設定値の特徴が組み合わさって、ひとつの変数設定を特徴付けていることも考えられる。

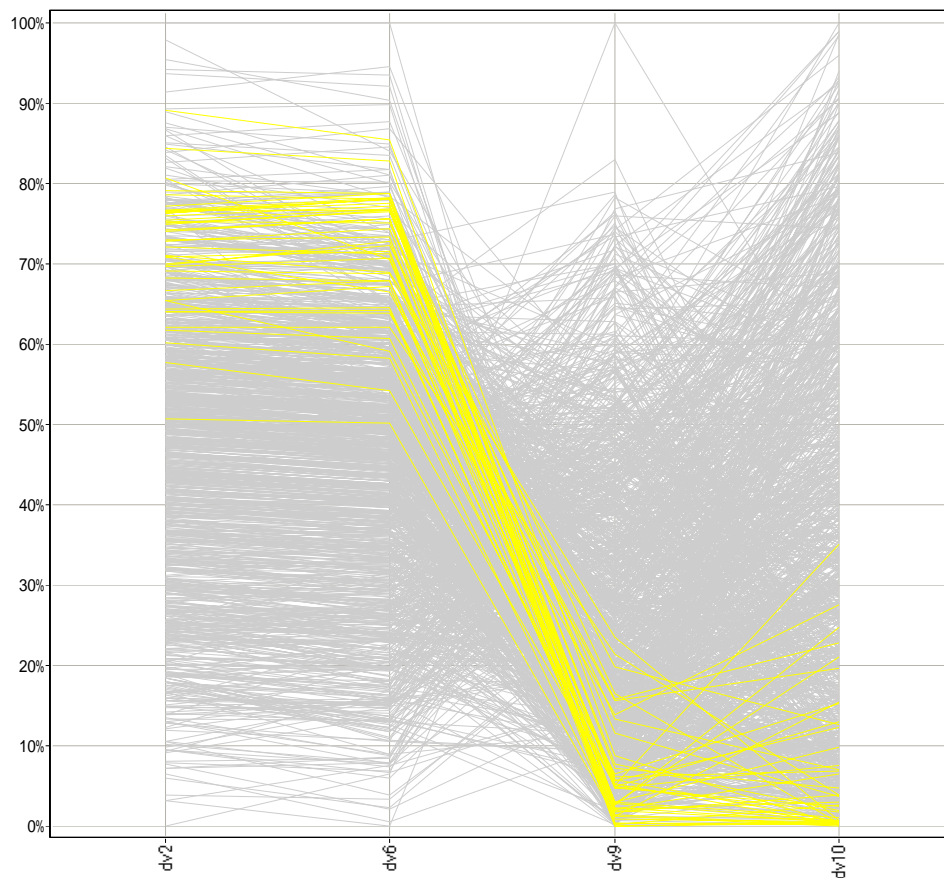


クラスタ3のうち、理想空間に属するものだけを抽出する。



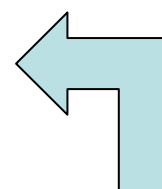


# 決定属性を満たす設計変数の考察 (2)

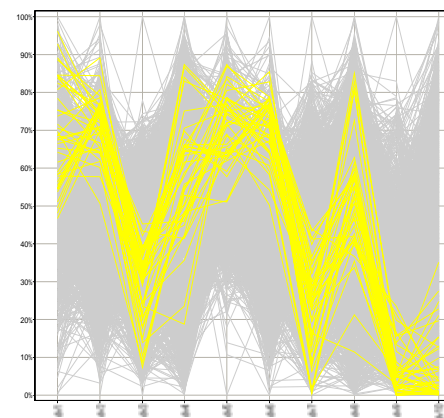


最後に先のANOVAで目的関数への貢献度が上位にランクされた四設計変数だけを可視化してみる。

このチャートから、それぞれの設計変数の特徴がさらによく見てとれる。

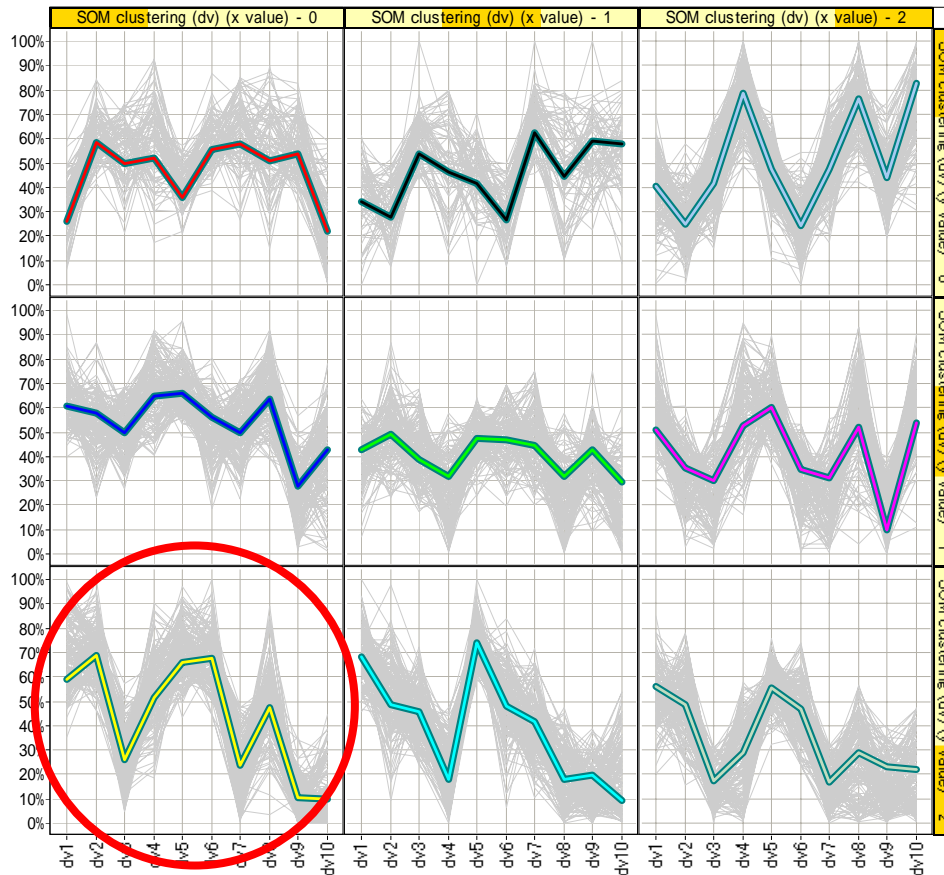


設計変数のうちANOVAで上位にランクされた設計変数に絞り込む。





# セントロイド考察 (1)



なお各クラスターのセントロイドを確認すると、各クラスターの変数ルールの特徴を確認することができる。

前述の偶然性の検定からクラスター3(左下)がスイートスポットに大きな確立で当てはまることから、このクラスターのセントロイドと、他のクラスターのセントロイドを比較することで、スイートスポットに当てはまる設計変数のルールの特徴を見て取ることができる。

なお図の波形は、厳密にはセントロイドではなく、セントロイドに最も類似したデータ内の値である。

Color by SOM clustering (dv):

1 2 3 4 5 6 7 8 9

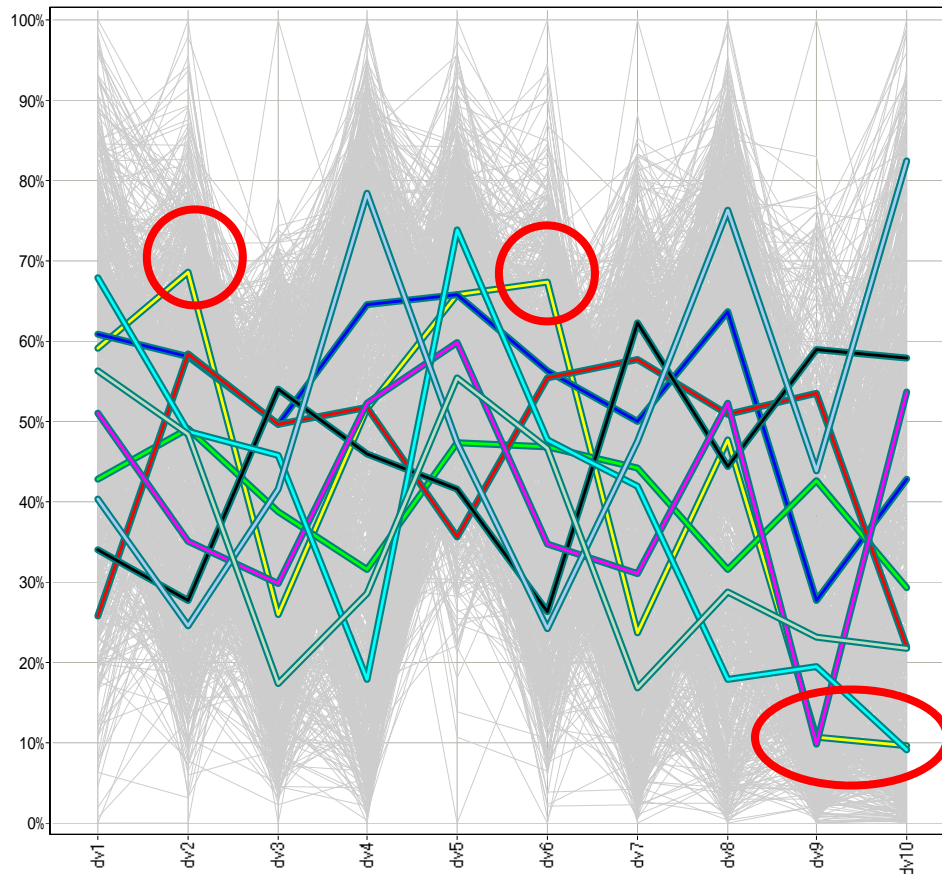
Line width by Centroid Thickness

All columns use individual scales.

The labels show obj1.



## セントロイド考察 (2)



各クラスターのセントロイドを重ねて確認する。黄色がクラスター3である。

dv2およびdv6については、クラスター3は単独で顕著に値が高い。それに比べてdv9とdv10は単独では低いクラスターが他にある(クラスター6とクラスター8)ことから、dv9とdv10は同時に低いことがスイートスポットを構成する設計変数の特徴といえる。

Color by SOM clustering (dv):

1 2 3 4 5 6 7 8 9

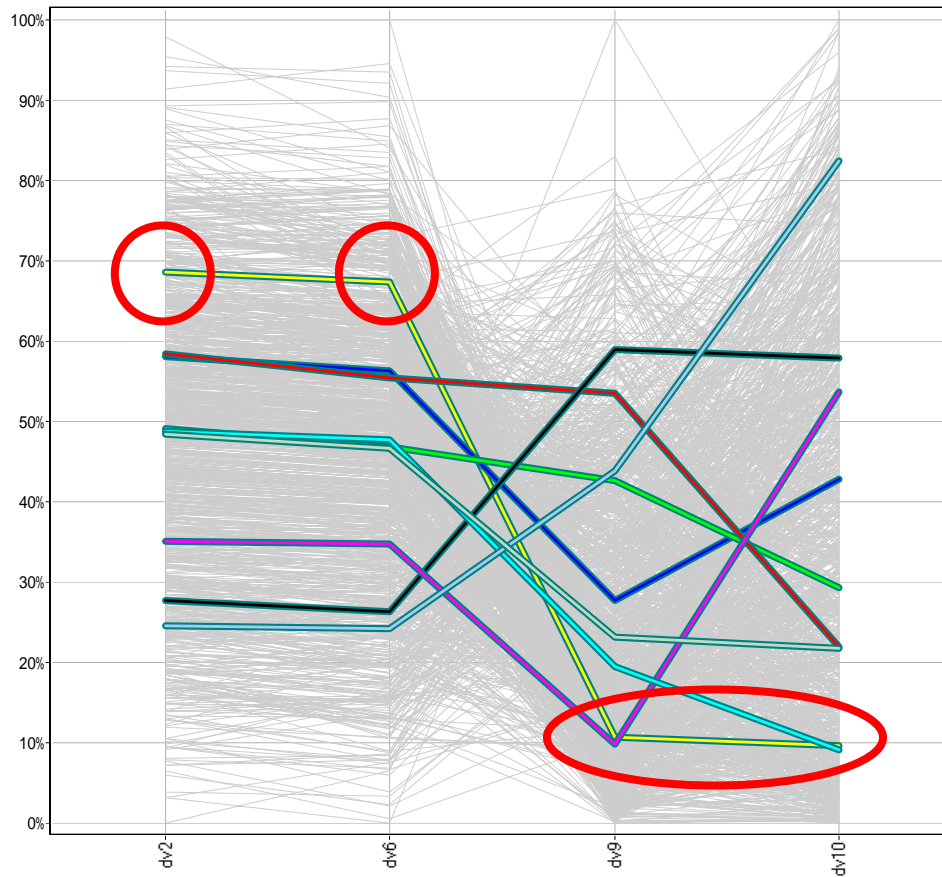
Line width by Centroid Thickness

All columns use individual scales.

The labels show SOM clustering (dv).



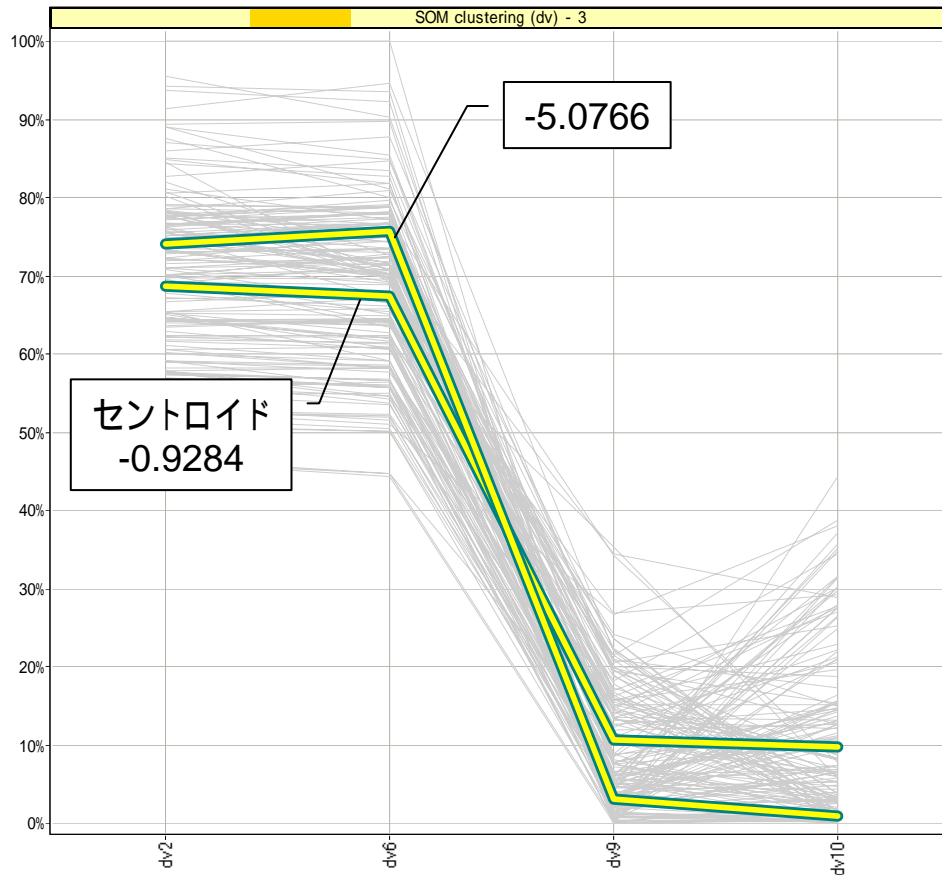
# セントロイド考察 (3)



さらにANOVAの上位四位までの設計変数だけを可視化する。前述の考察をさらに明確に示している。



# セントロイド考察 (4)



さらに目的関数がすべて小さいほうが  
いいということから、目的関数(正規化  
済)の合計値が最も小さくなる(-5.0766)  
場合の設計変数を重ねてみた。

やはりその場合の設計変数もクラスタ3  
に分類されていた。dv2、dv6およびdv9、  
dv10の傾向がさらに顕著に表れている。  
特にdv9およびdv10はクラスタ内で同  
時に最小値に近い値を示していること  
から、さらにこの二つの設計変数は小さ  
いほど良いということが伺える。

Color by SOM clustering (dv):

1 2 3 4 5 6 7 8 9

Line width by Calculated Column

All columns use individual scales.

The labels show Sum (obj1, obj2, obj3) (normalizaed).



## 結論

- SOMやANOVA単体では見えない設計変数のルールを、SOMやANOVAに加え、データの離散化や偶然性の検定など、複数の手法を組み合わせることで、比較的容易かつ明確に捉えることができた。
- すべての解析フローを、DSのみの使用で作業することができた。



# Beyond this analysis

- パレート解のうち、最適候補解とManufacturability
  - 最適解から少しくらいずれても、目的変数に大きな変動がないか？
    - 理想目的変数の設計変数を元にProfile Search
    - 目的変数に戻って暴れてないかどうか確認
- モデル外パラメータ (目的変数)
  - コスト、環境負荷、長期信頼性、市場評価などなど
- モデル検証
  - 実験結果とシミュレーション(モデル)結果の比較
  - 他のシミュレータとの比較
- 階層パレート
  - ここの最適クラスタを集めて、それらでまたパレート解を求める



## 参考文献





## 参考文献 (1)

- 大林茂「ラフ集合による設計空間特徴抽出」
  - 科研費シンポ原稿(2007年1月)
- 大林茂「ラフ集合による多目的最適化データマイニング」
  - 科研費シンポ原稿改訂版(2006年1月)



## 参考文献 (2)

- 自己組織化マップ (SOM)
  - Mirkin, B. 著、『Mathematical Classification and Clustering, Nonconvex Optimization and Its Applications Volume 11』、Pardalos, P. / Horst, R. 編集、The Netherlands、Kluwer Academic Publishers 刊、1996 年。
  - MacQueen, J. 著、『Some methods for classification and analysis of multivariate observations』、In Le Cam, L. M. / Neyman, J. 編集、『Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Volume I: Statistics』(p281 ~ 297)、Berkeley and Los Angeles, CA.、University of California Press 刊、1967 年。
- 偶然性の検定
  - Tavazoie, S.、Hughes, J.D.、Campbell, M.J.、Cho, R.J.、Church, G.M. 著、『Systematic determination of genetic network architecture』、Nature Genetics、22 (3)、pp 281-285、1999 年
- Comparison Circles
  - Hsu, J.C. 著、『Multiple Comparisons: Theory and Methods』、London、Chapman & Hall 刊、1996年。
  - Sall, J. 著、『"Graphical Comparison of Means" Statistical Computing and Statistical Graphics Newsletter, 3』、1992年、p.27-32。

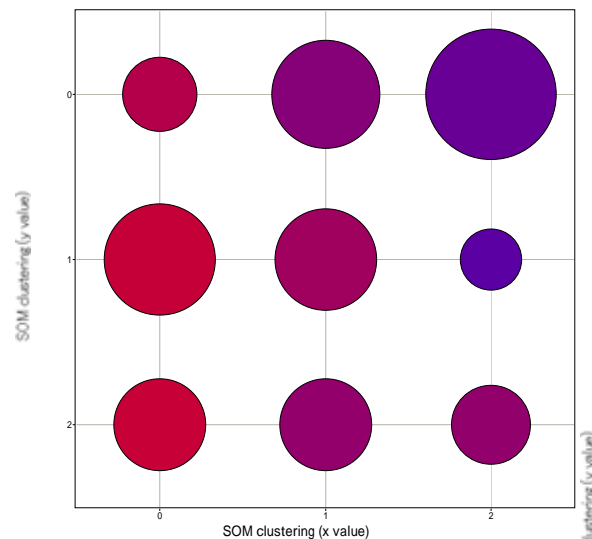


補足資料

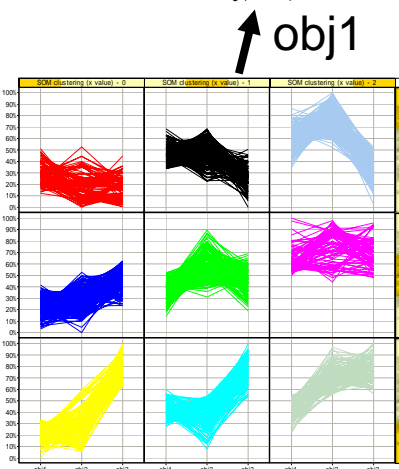




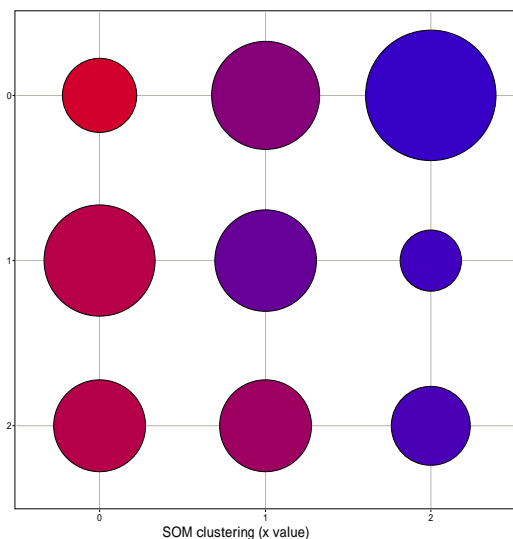
# SOMクラスタの色づけ (1)



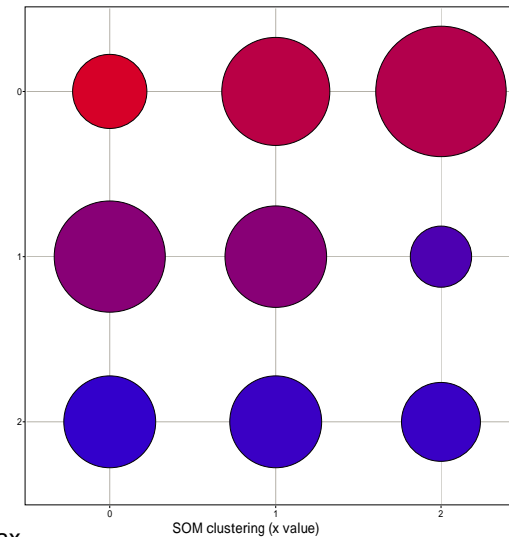
バブルチャートを使ってSOMの領域に色をつけることができる。  
ここでは各目的関数のクラスタごとの平均値(正規化済)を色付けしている。



obj2

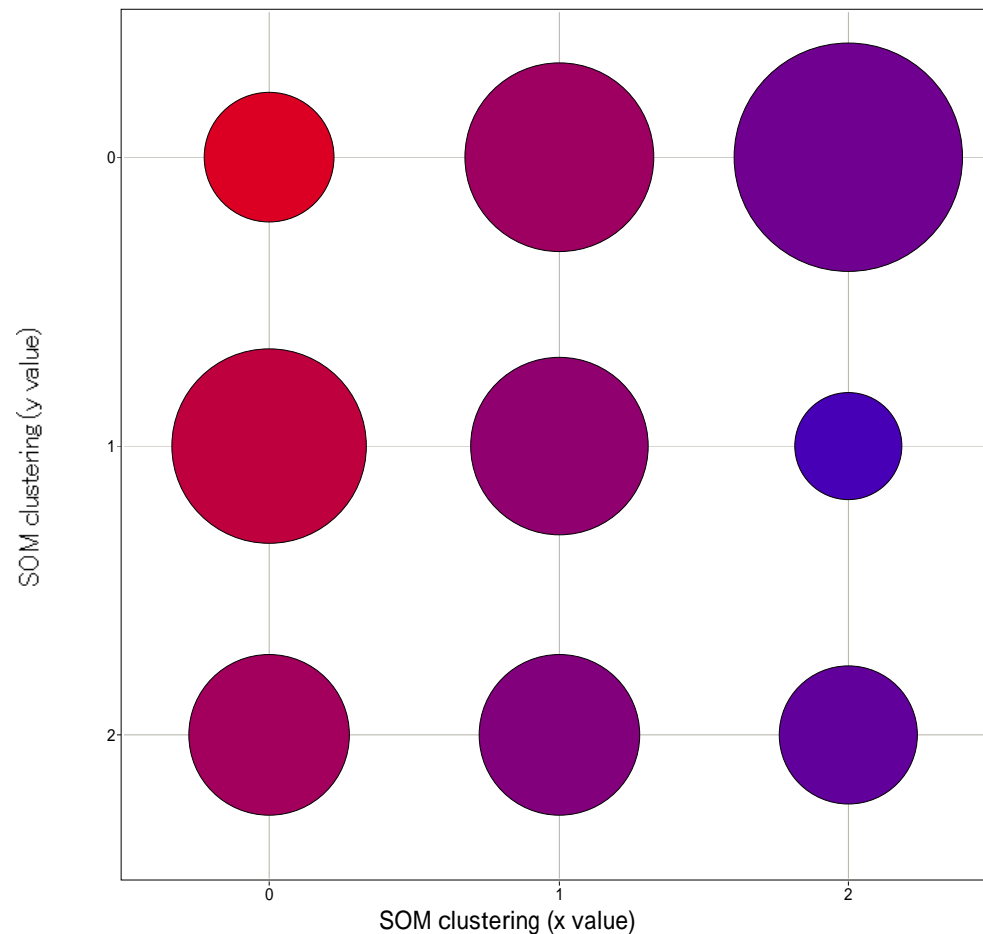


obj3





# SOMクラスタの色づけ (2)



目的関数はそれぞれ低ければ低いほどいいので、三目的関数の各クラスター内の平均値(正規化済)を算出し、それで色付けした。  
ここでも左上のクラスター(1)がスイートスポットであることがわかる。

