معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute

Member of Qatar Foundation عضو في مؤسسة قطر

Reflections on Dimensionality Reduction for Visual Analytics: strengths, weaknesses and perspectives

Michael Aupetit

Qatar Computing Research Institute, HBKU, Doha, Qatar

Joint work with: L. Allano, I. Espagnon, J.-D. Fekete, N. Heulot, S. Lespinats, H. Meyer-Baese, G. Sannie

Mascot-Num –Visualization methods for uncertainty studies – Institut Henri Poincare, Paris, May 22, 2017

Outline

Outline

- What is Dimensionality Reduction

Outline

- What is Dimensionality Reduction
- Strengths

Outline

- What is Dimensionality Reduction
- Strengths
- Weaknesses

Outline

- What is Dimensionality Reduction
- Strengths
- Weaknesses
- Turning weaknesses into strengths

Outline

- What is Dimensionality Reduction
- Strengths
- Weaknesses
- Turning weaknesses into strengths
- Conclusion

Outline

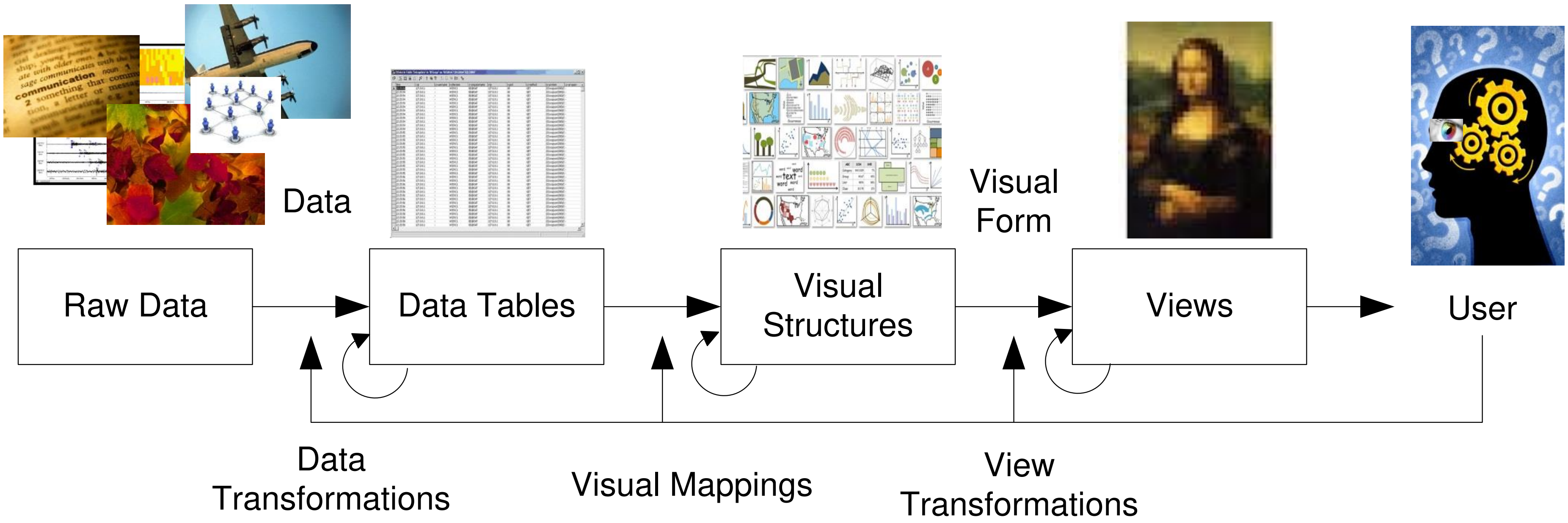
- What is Dimensionality Reduction
- Strengths
- Weaknesses
- Turning weaknesses into strengths
- Conclusion
- Perspectives

Outline

- **What is Dimensionality Reduction**
- Strengths
- Weaknesses
- Turning weaknesses into strengths
- Conclusion
- Perspectives

What is Dimensionality Reduction

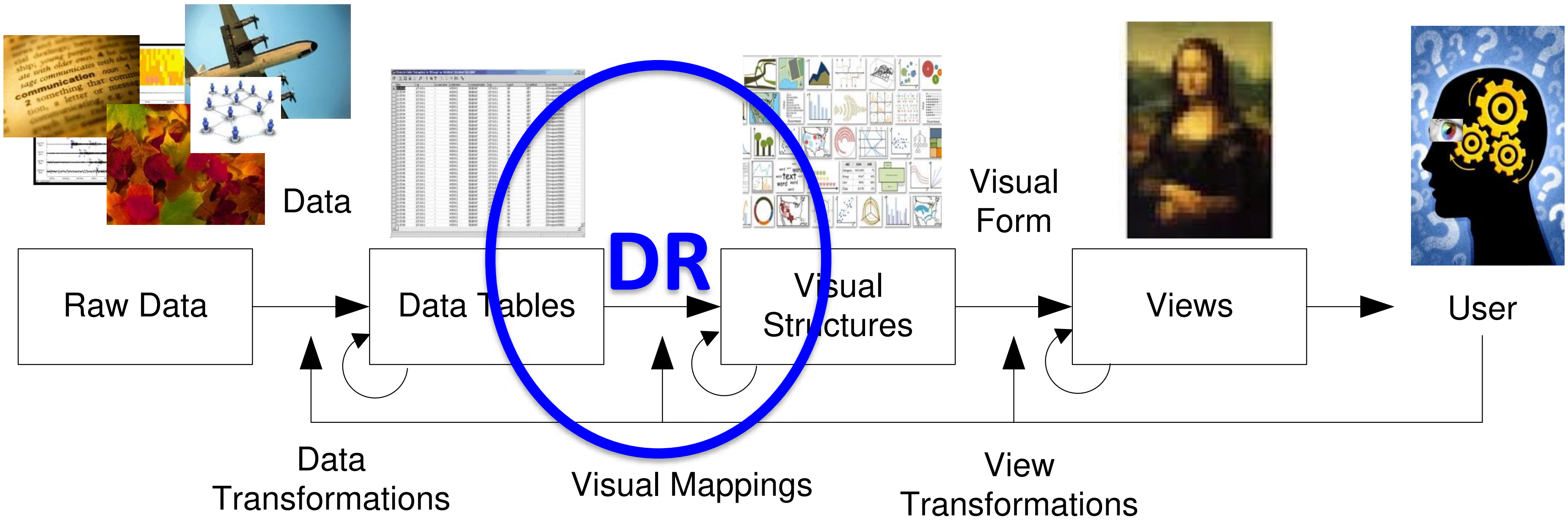
- DR in the visualization pipeline



S.K. Card, J. Mackinlay, B. Shneiderman,
Information Visualization Readings in Information Visualization: Using Vision to Think. Morgan Kaufman 1998.

What is Dimensionality Reduction

- DR in the visualization pipeline



S.K. Card, J. Mackinlay, B. Shneiderman,
Information Visualization Readings in Information Visualization: Using Vision to Think. Morgan Kaufman 1998.

What is Dimensionality Reduction

- Data table

Variable (attributes)

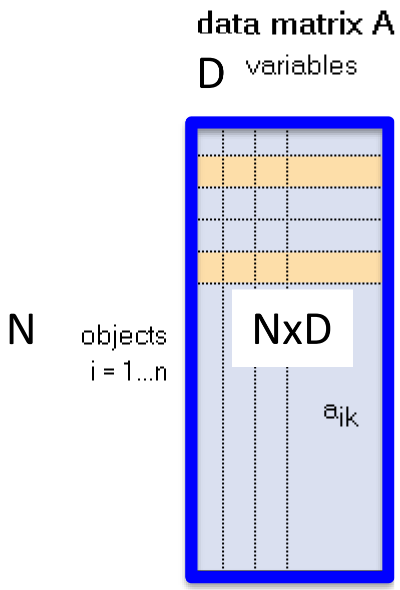
Name	Thread pitch (mm)	Minor diameter tolerance	Nominal diameter (mm)	Head shape	Price for 50 screws	Available at factory outlet?	Number in stock	Flat or Phillips head?
M4	0.7	4g	4	Pan	$10.08	Yes	276	Flat
M5	0.8	4g	5	Round	$13.89	Yes	183	Both
M6	1	5g	6	Button	$10.42	Yes	1043	Flat
M8	1.25	5g	8	Pan	$11.98	No	298	Phillips
M10	1.5	6g	10	Round	$16.74	Yes	488	Phillips
M12	1.75	7g	12	Pan	$18.26	No	998	Flat
M14	2	7g	14	Round	$21.19	No	235	Phillips
M16	2	8g	16	Button	$23.57	Yes	292	Both
M18	2.1	8g	18	Button	$25.87	No	664	Both
M20	2.4	8g	20	Pan	$29.09	Yes	486	Both
M24	2.55	9g	24	Round	$33.01	Yes	982	Phillips
M28	2.7	10g	28	Button	$35.66	No	1067	Phillips
M36	3.2	12g	36	Pan	$41.32	No	434	Both
M50	4.5	15g	50	Pan	$44.72	No	740	Flat

value

Instance (item)

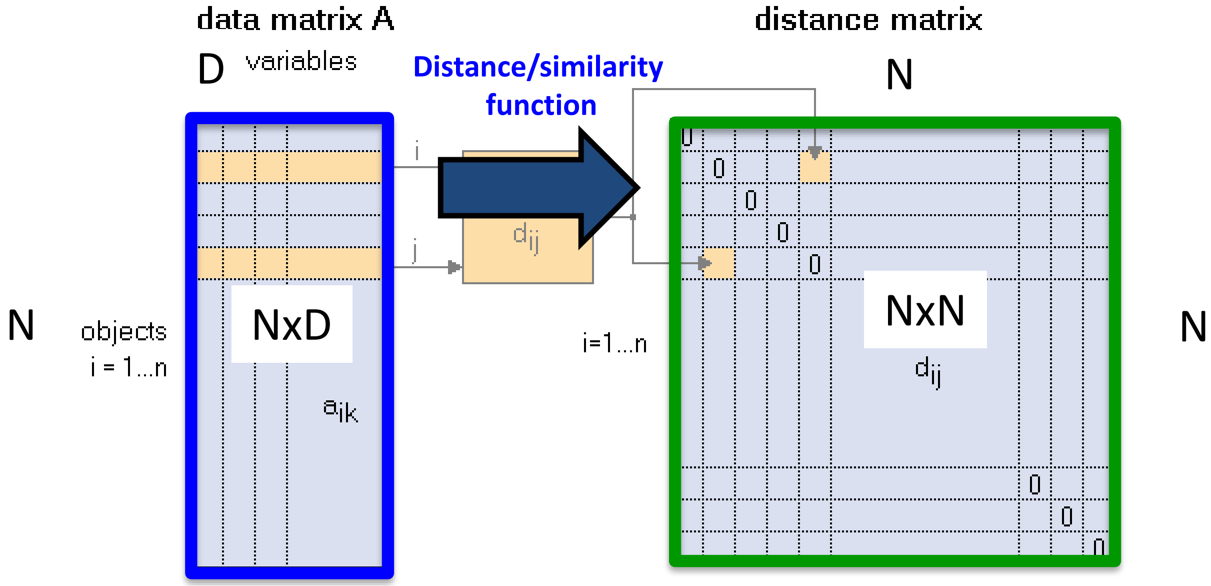
What is Dimensionality Reduction

- From **data table** to **similarity matrix**



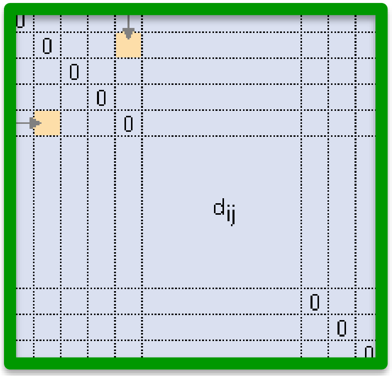
What is Dimensionality Reduction

- From **data table** to **similarity matrix**



What is Dimensionality Reduction

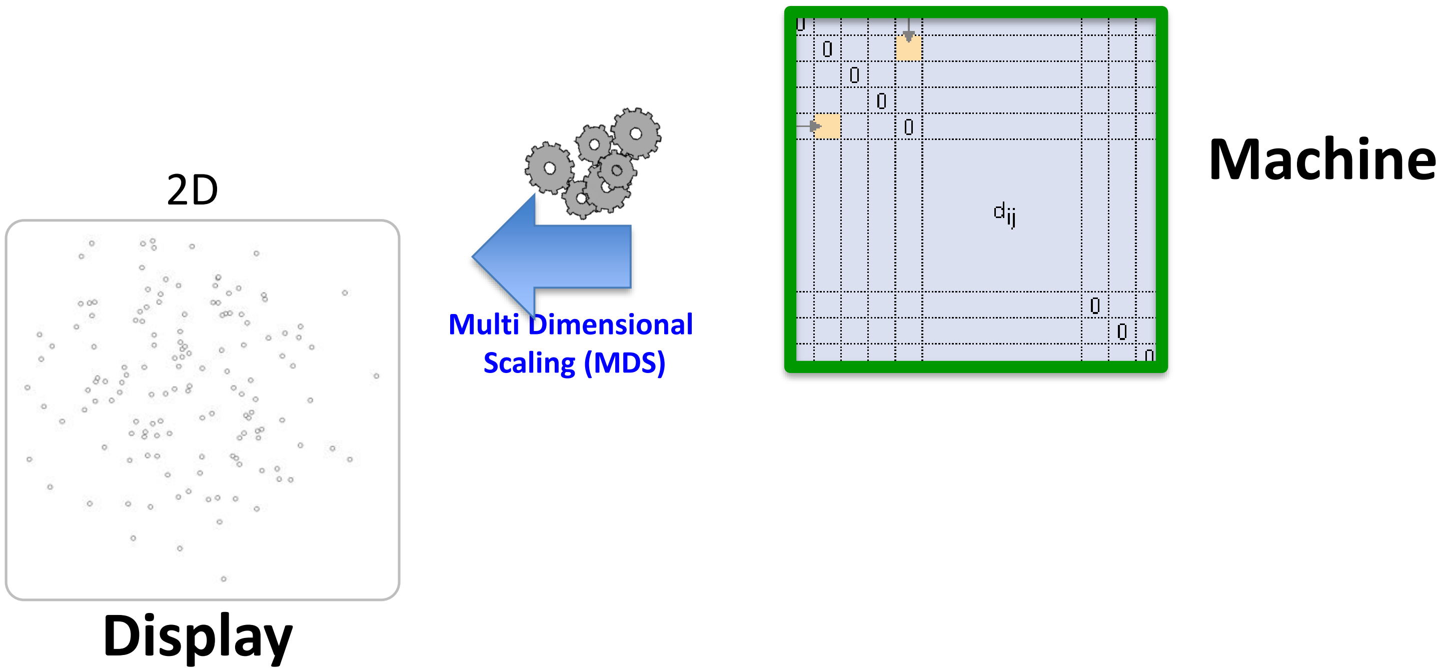
- From **similarity matrix** to **scatterplot**



Machine

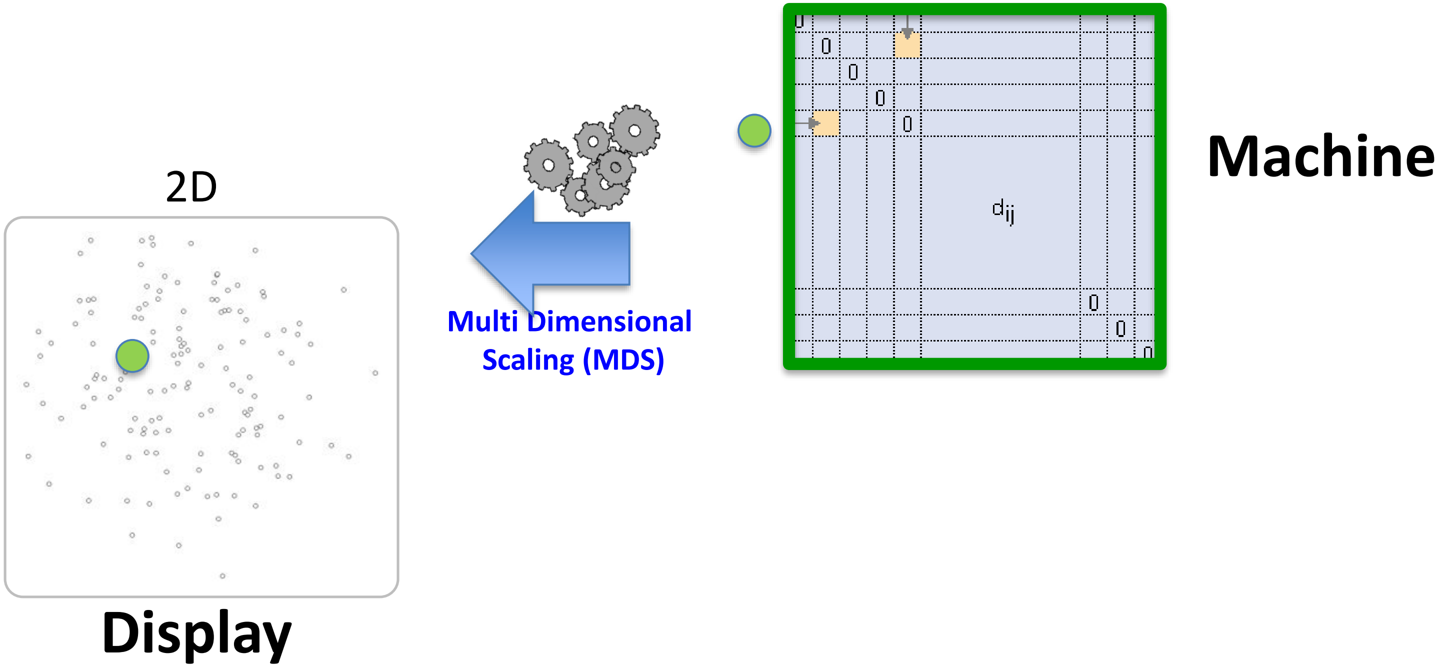
What is Dimensionality Reduction

- From **similarity matrix** to **scatterplot**



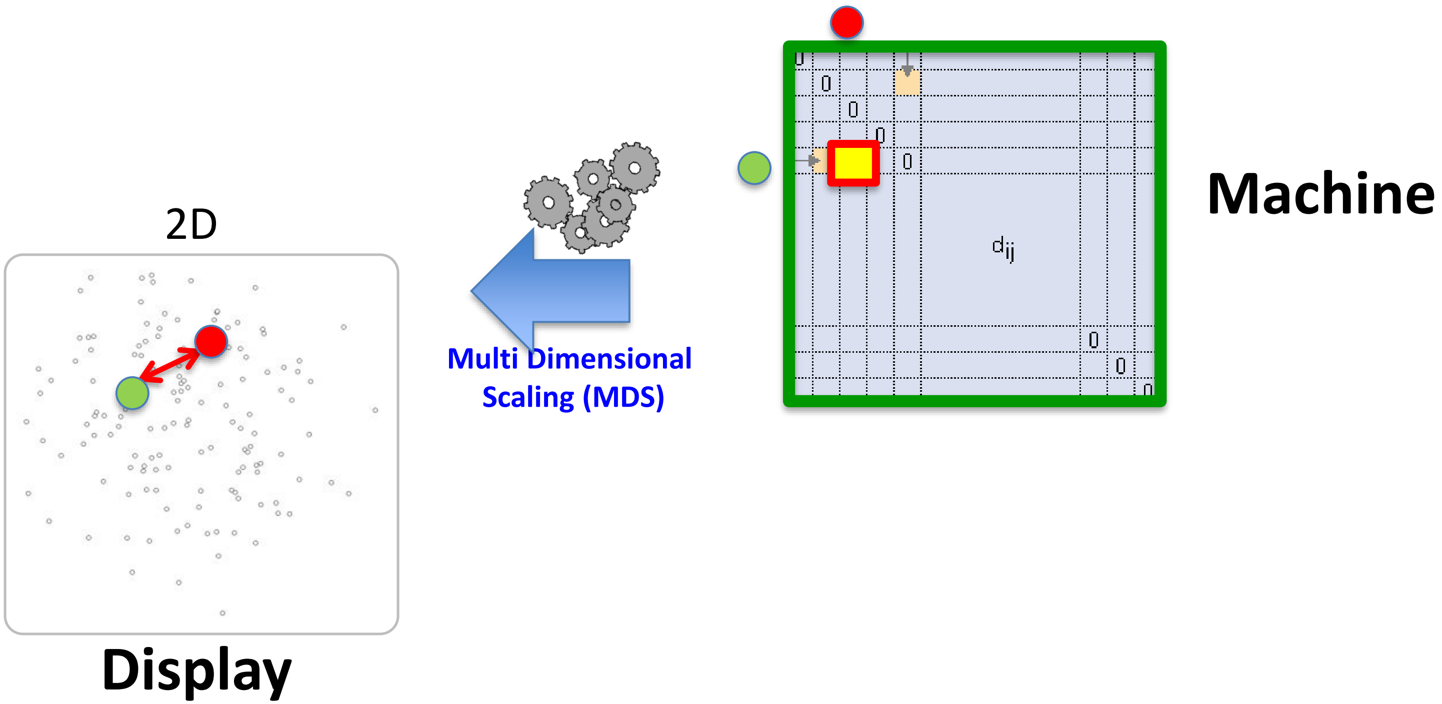
What is Dimensionality Reduction

- From **similarity matrix** to **scatterplot**



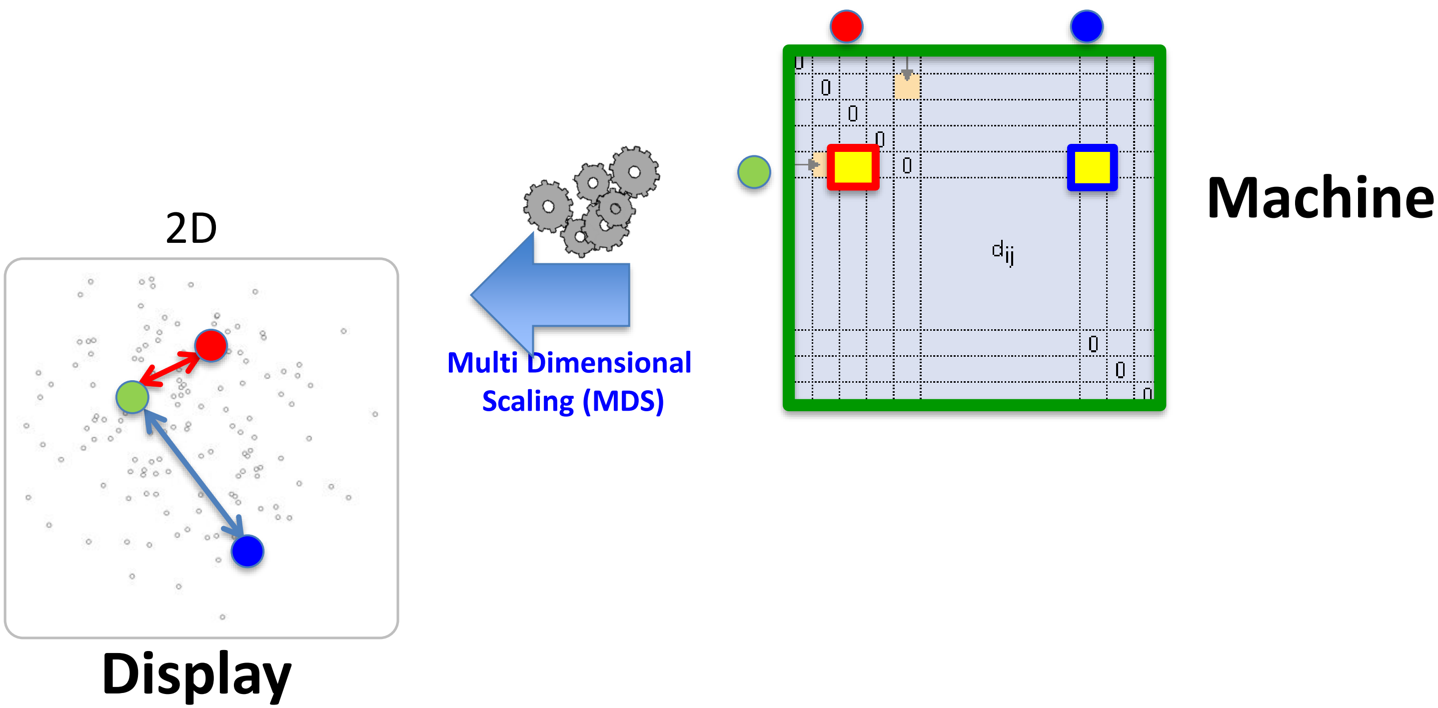
What is Dimensionality Reduction

- From **similarity matrix** to **scatterplot**



What is Dimensionality Reduction

- From **similarity matrix** to **scatterplot**

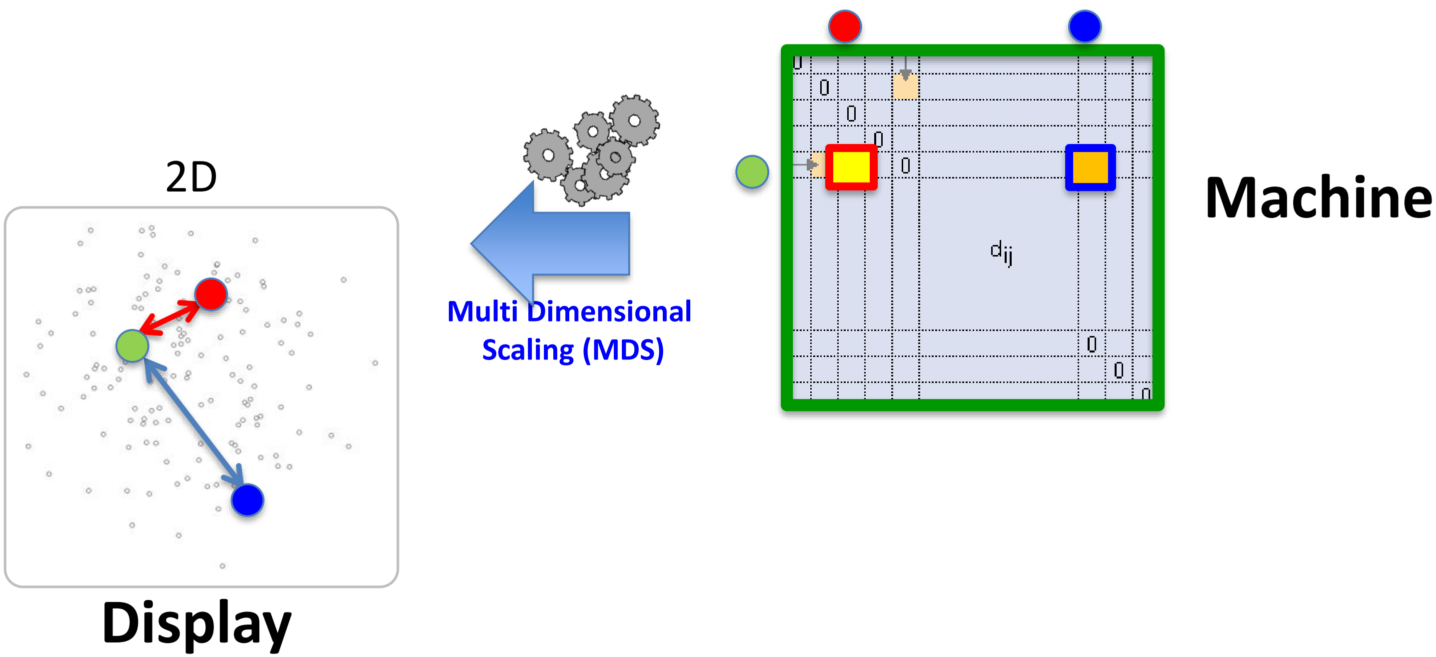


Outline

- What is Dimensionality Reduction
- **Strengths**
- Weaknesses
- Turning weaknesses into strengths
- Conclusion
- Perspectives

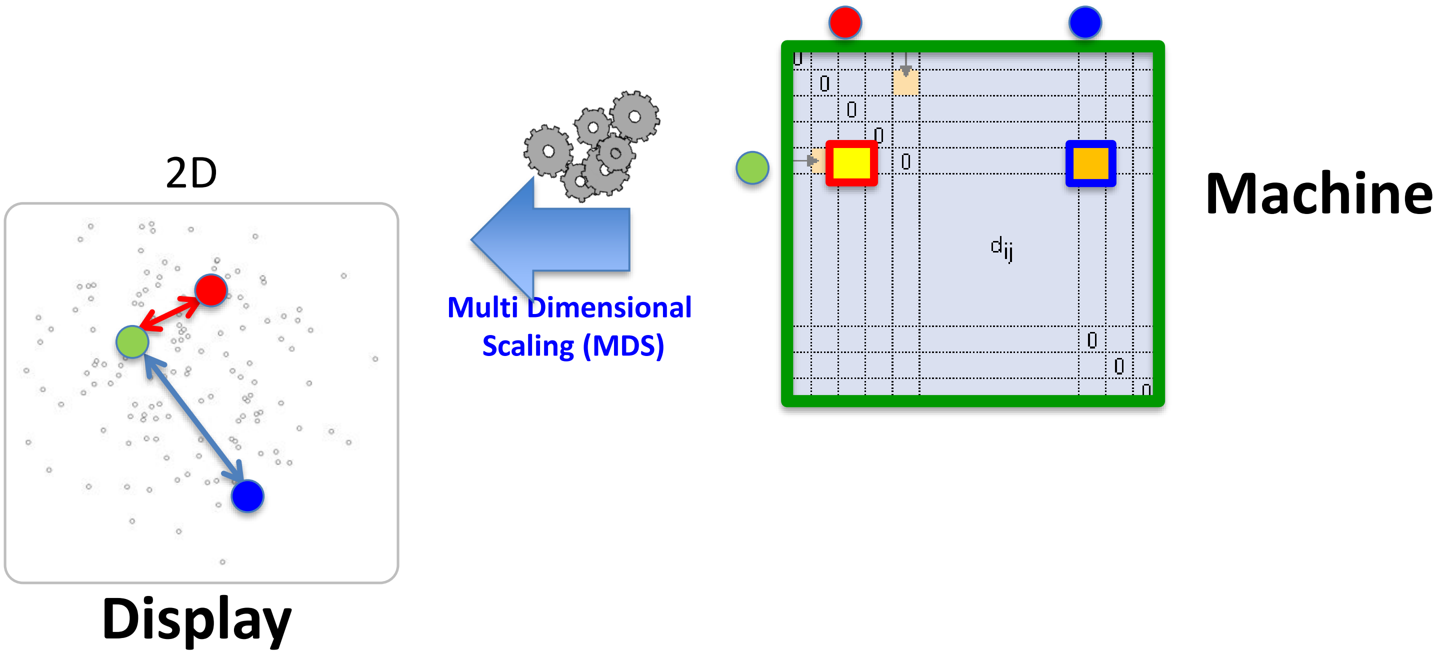
Strengths of Dimensionality Reduction

- Benefits from **Gestalt law of proximity**



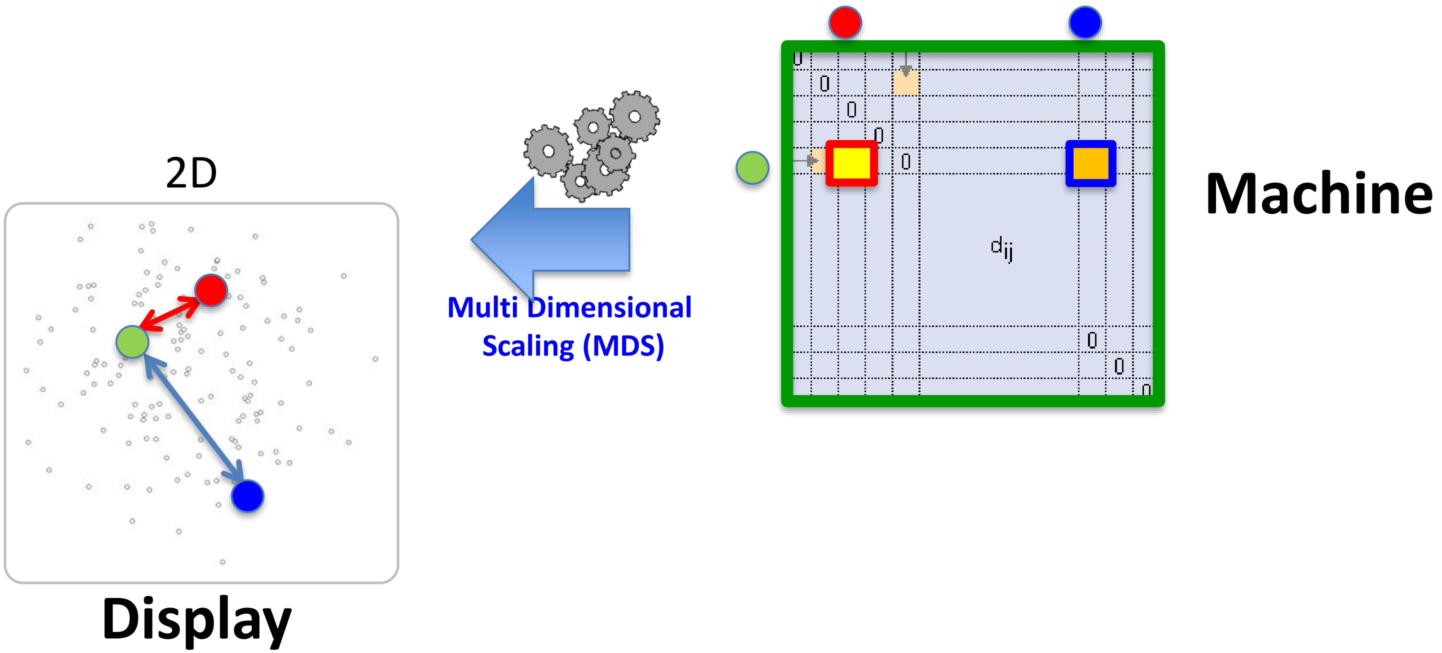
Strengths of Dimensionality Reduction

- Benefits from **Gestalt law of proximity**
  - spatial proximity => items **belong to same group**



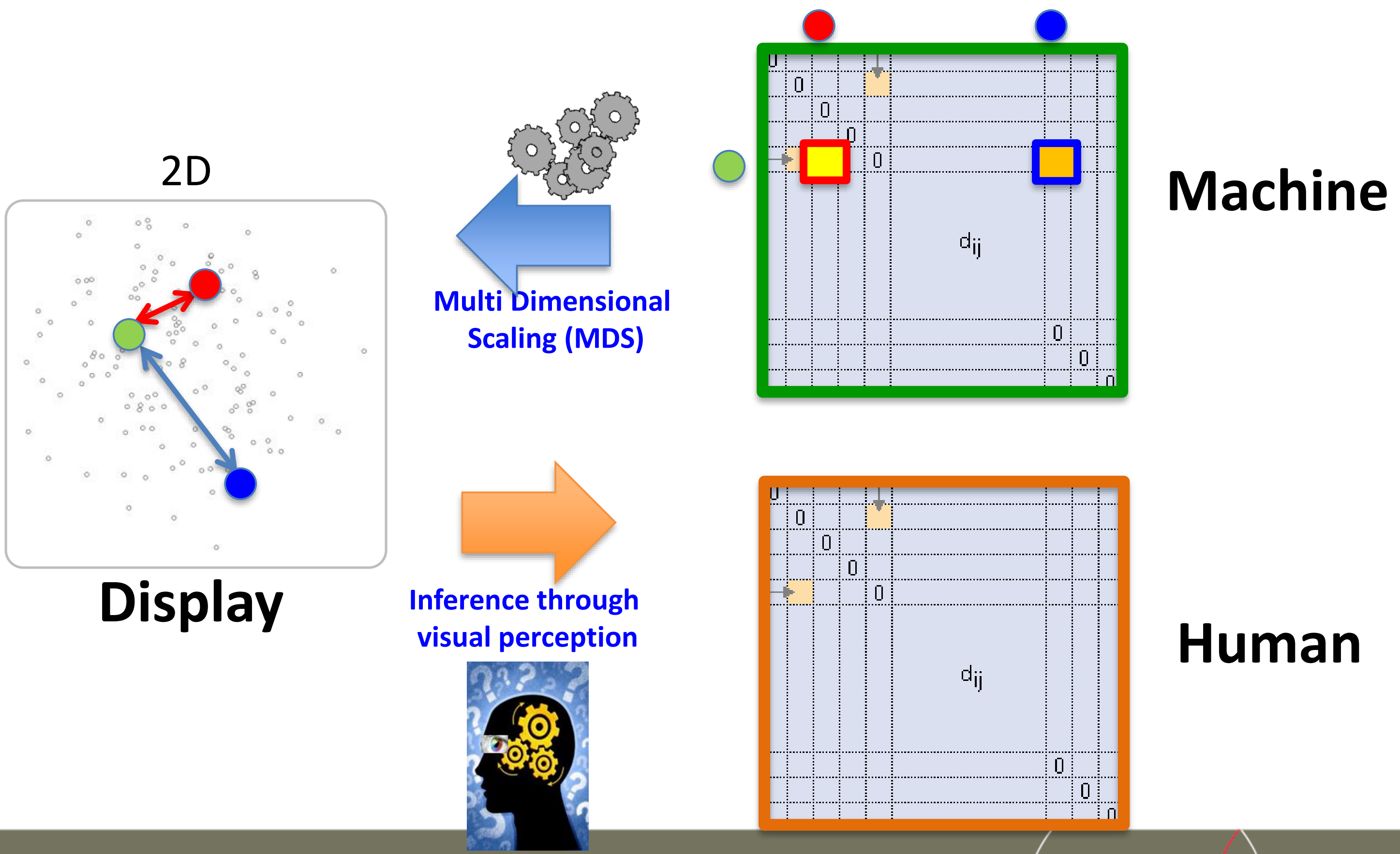
Strengths of Dimensionality Reduction

- Benefits from **Gestalt law of proximity**
  - spatial proximity => items **belong to same group**
  - pre-attentive = with **no cognitive effort**



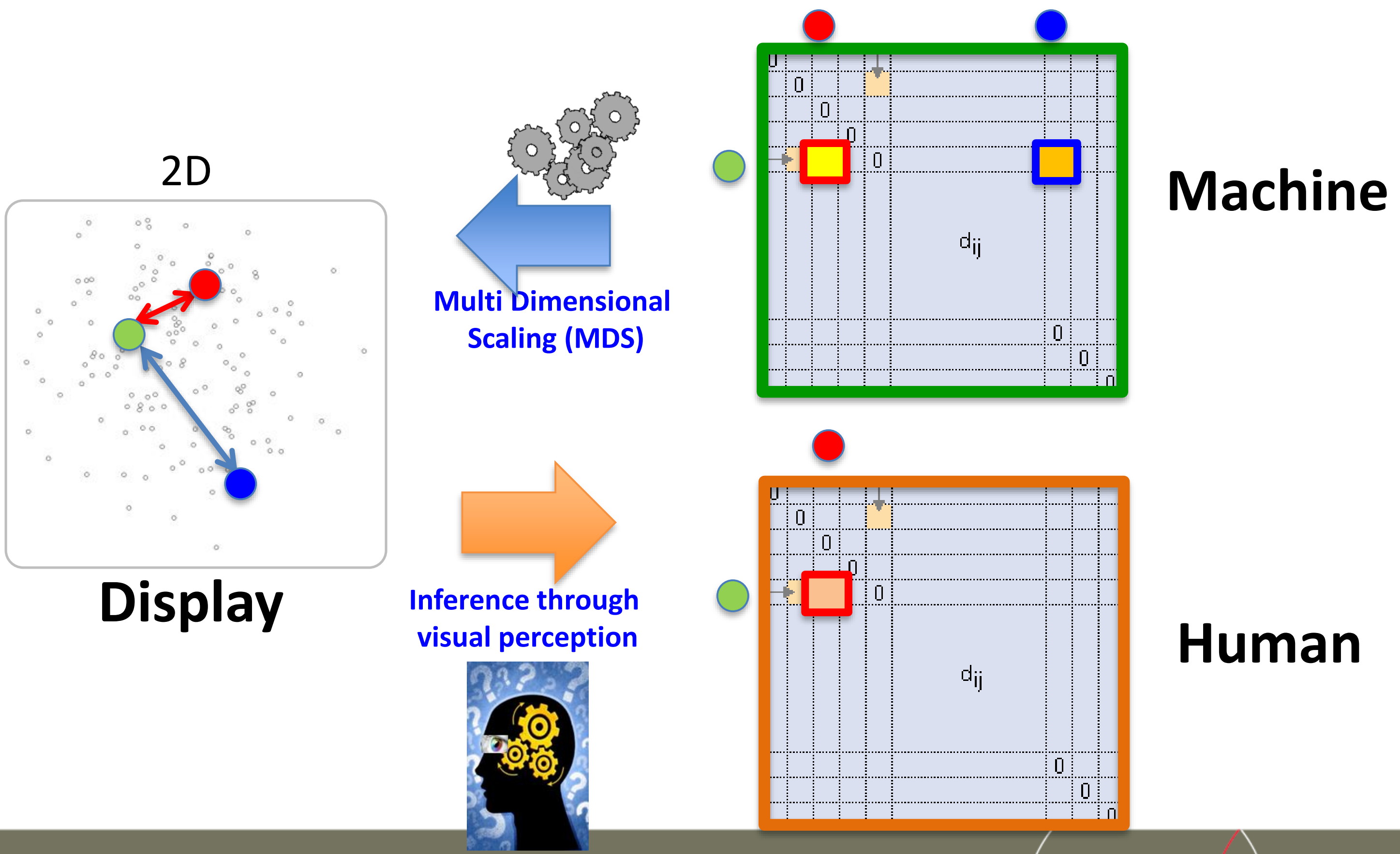
Strengths of Dimensionality Reduction

- Benefits from **Gestalt law of proximity**
  - spatial proximity => items **belong to same group**
  - pre-attentive = with **no cognitive effort**



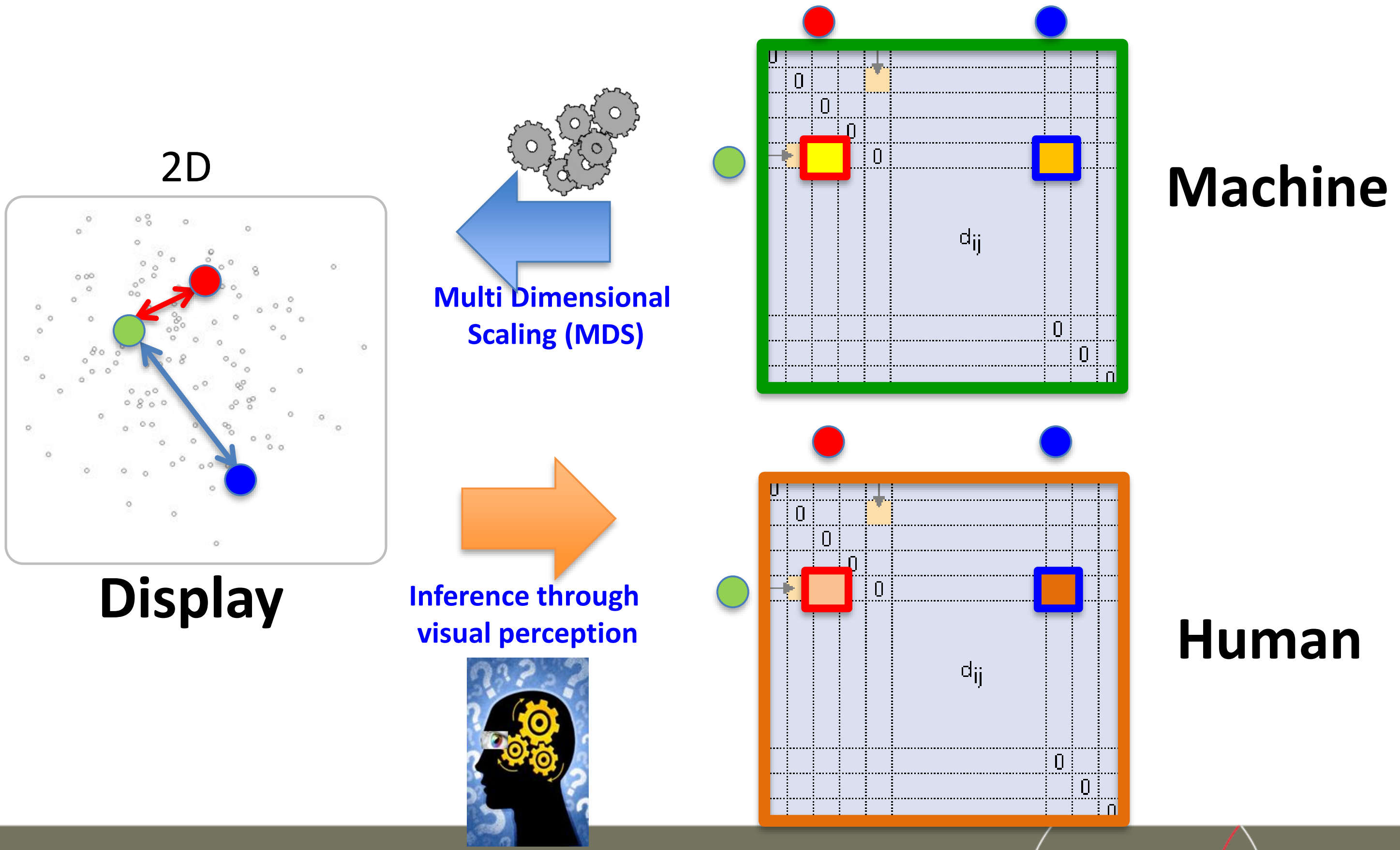
Strengths of Dimensionality Reduction

- Benefits from **Gestalt law of proximity**
  - spatial proximity => items **belong to same group**
  - pre-attentive = with **no cognitive effort**



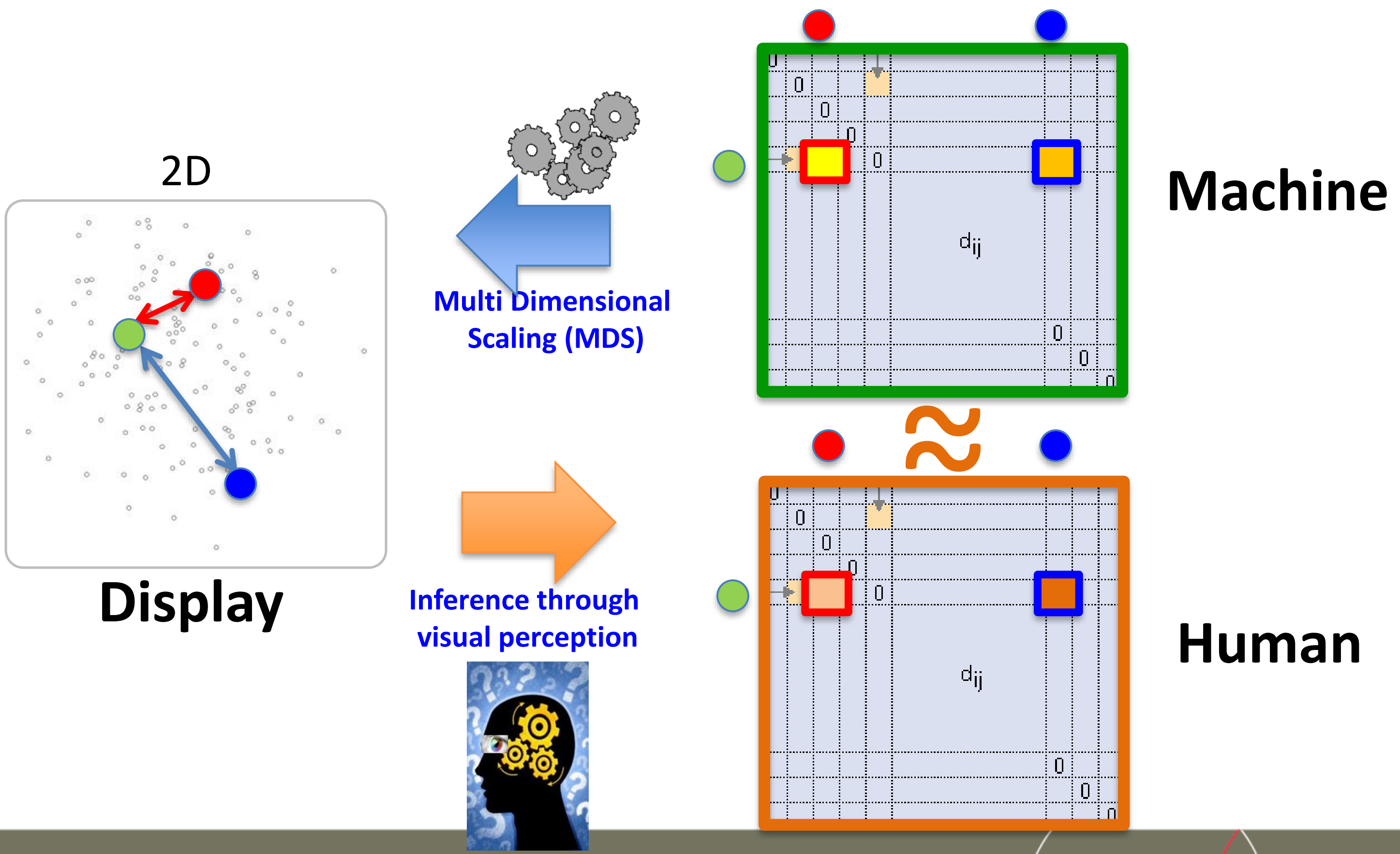
Strengths of Dimensionality Reduction

- Benefits from **Gestalt law of proximity**
  - spatial proximity => items **belong to same group**
  - pre-attentive = with **no cognitive effort**



Strengths of Dimensionality Reduction

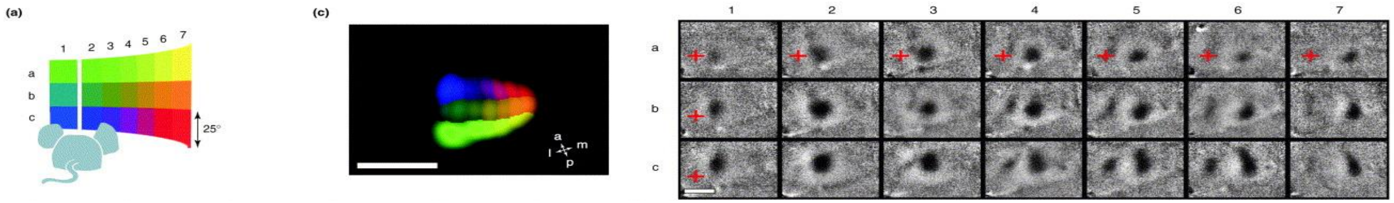
- Benefits from **Gestalt law of proximity**
  - spatial proximity => items **belong to same group**
  - pre-attentive = with **no cognitive effort**



Strengths of Dimensionality Reduction

Strengths of Dimensionality Reduction

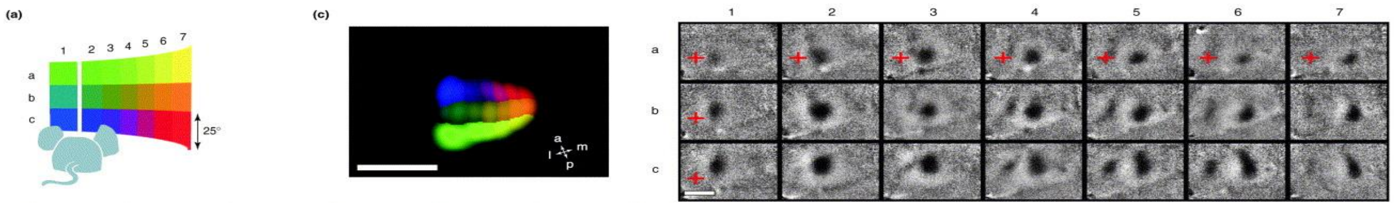
- Mammals' brains organize **similarities** to build **maps**



Retinotopic map of a mouse [Hübener 2003]

Strengths of Dimensionality Reduction

• Mammals' brains organize **similarities** to build **maps**

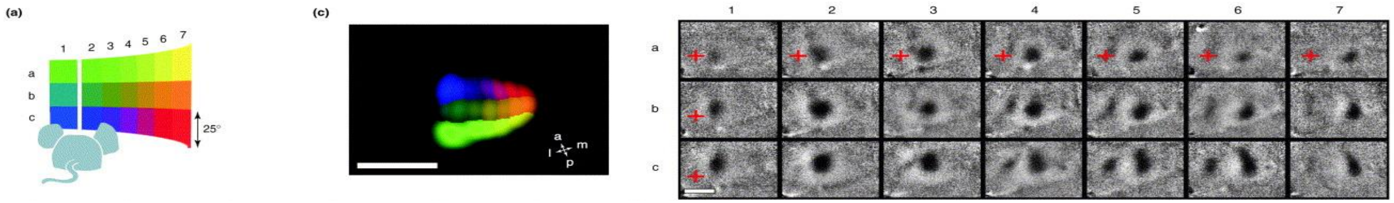


Retinotopic map of a mouse [Hübener 2003]

• This map-building process as been selected through evolution because it is of **survival importance**

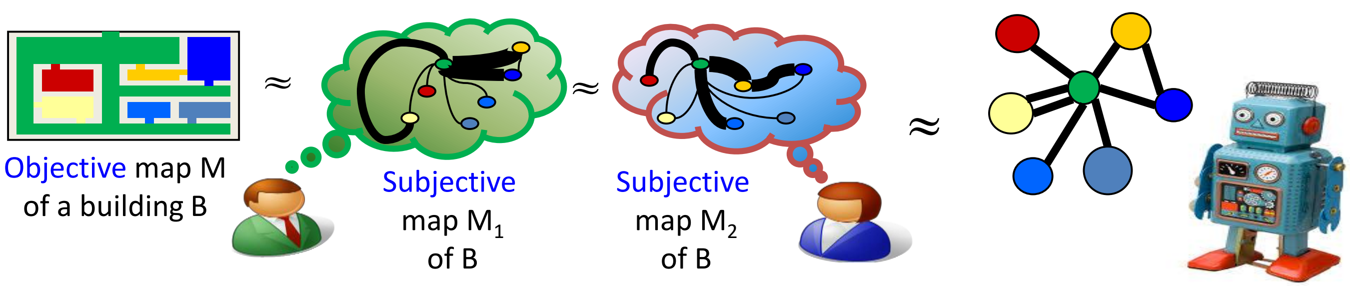
Strengths of Dimensionality Reduction

- Mammals' brains organize **similarities** to build **maps**



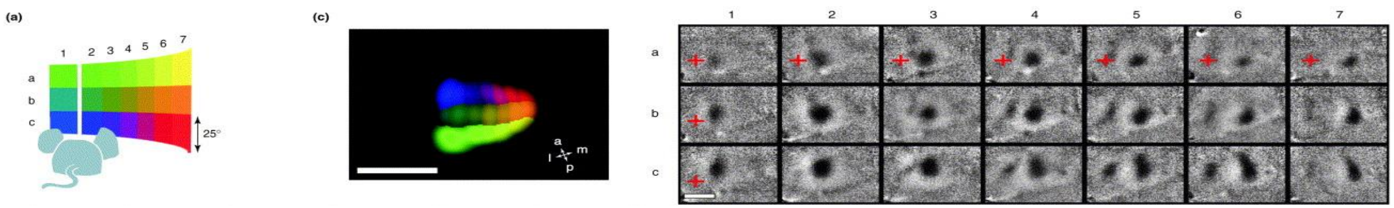
Retinotopic map of a mouse [Hübener 2003]

- This map-building process as been selected through evolution because it is of **survival importance**
  - **to navigate *physical* world…**



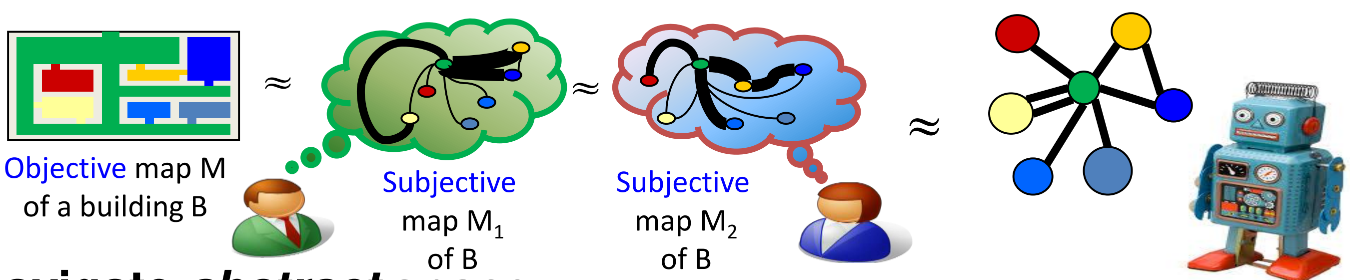
Strengths of Dimensionality Reduction

- Mammals' brains organize **similarities** to build **maps**



Retinotopic map of a mouse [Hübener 2003]

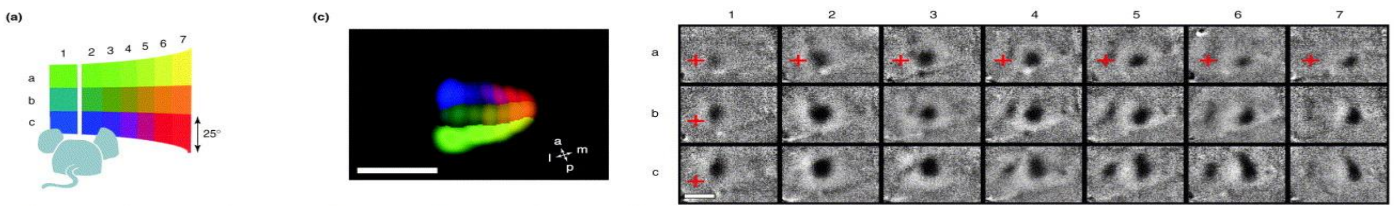
- This map-building process as been selected through evolution because it is of **survival importance**
  - **to navigate *physical* world…**



  - **and navigate *abstract* space…**

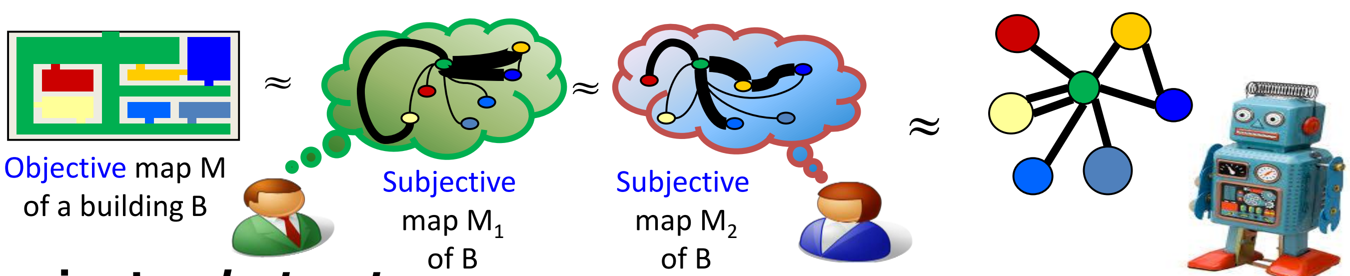
Strengths of Dimensionality Reduction

- Mammals' brains organize **similarities** to build **maps**



Retinotopic map of a mouse [Hübener 2003]

- This map-building process as been selected through evolution because it is of **survival importance**
  - **to navigate *physical* world…**

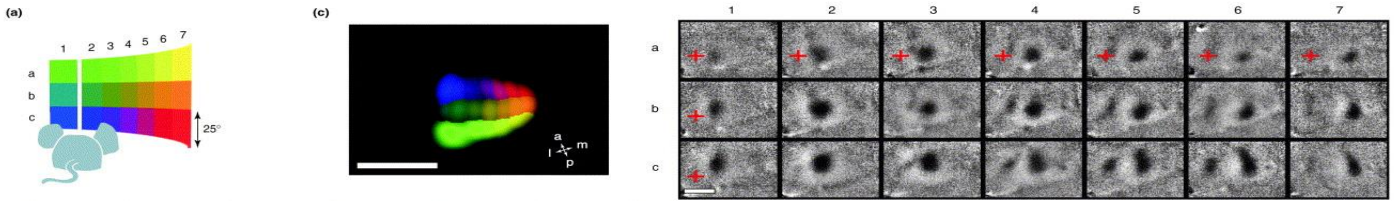


  - **and navigate *abstract* space…**

Te***l***er***a***e***dt***hs*i*a***a***e***a***lt***g***se***sl***sol***n***hs***p***a***ry***netimikwyet***r***sn***e***ny

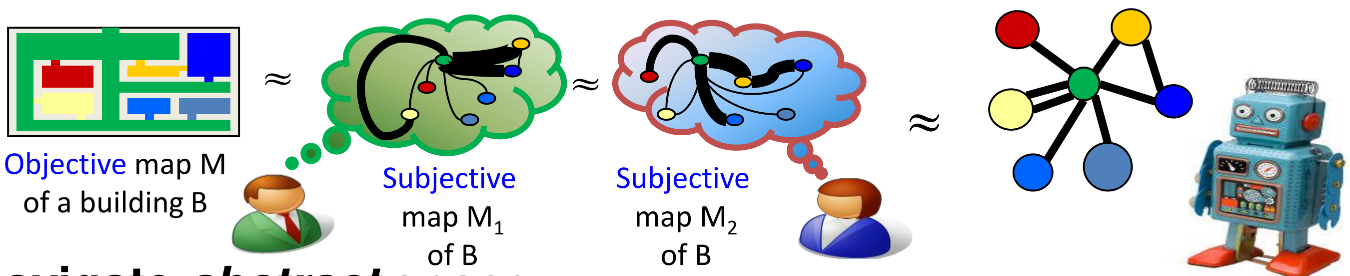
Strengths of Dimensionality Reduction

- Mammals' brains organize **similarities** to build **maps**



Retinotopic map of a mouse [Hübener 2003]

- This map-building process as been selected through evolution because it is of **survival importance**
  - **to navigate *physical* world…**



  - **and navigate *abstract* space…**

Te***l***er***a***e***dt***hs***i***a***a***e***a***lt***g***se***sl***sol***n***hs***p***a***ry***netimikwyet***r***sn***e***ny

These letters make sense only ***spatially arranged*** in this way

Strengths of Dimensionality Reduction

• **Correlation** between **spatial arrangement** and additional information is **crucial**

Strengths of Dimensionality Reduction

- **Correlation** between **spatial arrangement** and additional information is **crucial**

Strengths of Dimensionality Reduction

• **Correlation** between **spatial arrangement** and additional information is **crucial**

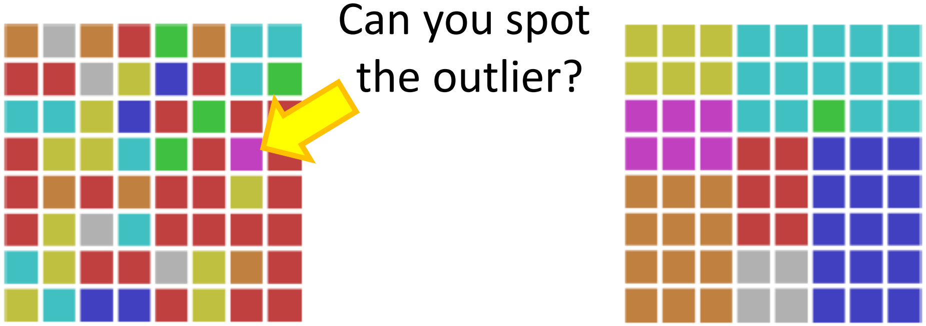
Can you spot the outlier?

Strengths of Dimensionality Reduction

- **Correlation** between **spatial arrangement** and additional information is **crucial**

Strengths of Dimensionality Reduction

- **Correlation** between **spatial arrangement** and additional information is **crucial**

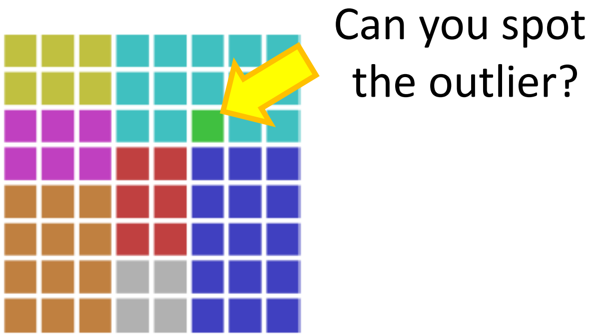


Strengths of Dimensionality Reduction

- **Correlation** between **spatial arrangement** and additional information is **crucial**

Strengths of Dimensionality Reduction

- **Correlation** between **spatial arrangement** and additional information is **crucial**



Strengths of Dimensionality Reduction

- **Correlation** between **spatial arrangement** and additional information is **crucial**

High cognitive burden
Proximity (*position*)
not correlated with
Similarity (*color/shape*)

Can you spot the outlier?

Can you spot the outlier?

Strengths of Dimensionality Reduction

- **Correlation** between **spatial arrangement** and additional information is **crucial**

***High* cognitive burden**
Proximity (*position*)
not correlated with
Similarity (*color/shape*)

Can you spot the outlier?

Can you spot the outlier?

***Low* cognitive burden**
Proximity
correlated with
Similarity

Strengths of Dimensionality Reduction

- **Correlation** between **spatial arrangement** and additional information is **crucial**

High cognitive burden
Proximity (*position*)
not correlated with
Similarity (*color/shape*)

Can you spot the outlier?

Can you spot the outlier?

Low cognitive burden
Proximity
correlated with
Similarity

[Haroz&Whitney, InfoVis2012]

Strengths of Dimensionality Reduction

- **Correlation** between **spatial arrangement** and additional information is **crucial**

High cognitive burden
Proximity (*position*)
not correlated with
Similarity (*color/shape*)

Can you spot the outlier?

Can you spot the outlier?

Low cognitive burden
Proximity
correlated with
Similarity

[Haroz&Whitney, InfoVis2012]

Strengths of Dimensionality Reduction

- **Correlation** between **spatial arrangement** and additional information is **crucial**

High cognitive burden
Proximity (*position*)
not correlated with
Similarity (*color/shape*)

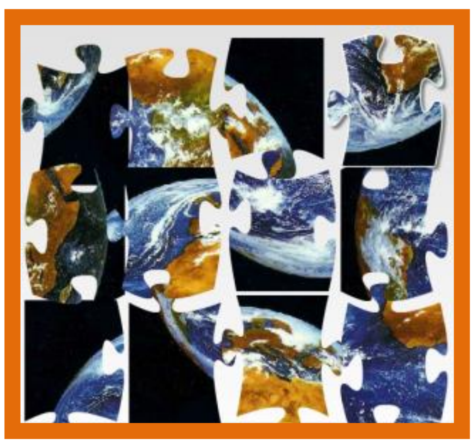
Can you spot the outlier?

Can you spot the outlier?

Low cognitive burden
Proximity
correlated with
Similarity

[Haroz&Whitney, InfoVis2012]

Error in *Position?*
or
error in *Color?*



Strengths of Dimensionality Reduction

- **Correlation** between **spatial arrangement** and additional information is **crucial**

High cognitive burden
Proximity (*position*)
not correlated with
Similarity (*color/shape*)

Can you spot the outlier?

Can you spot the outlier?

Low cognitive burden
Proximity
correlated with
Similarity

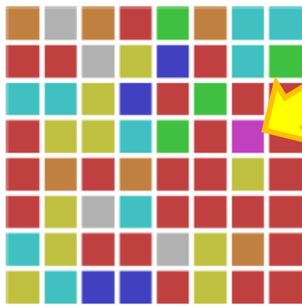
[Haroz&Whitney, InfoVis2012]

Error in *Position?*
or
error in *Color?*

Strengths of Dimensionality Reduction

- **Correlation** between **spatial arrangement** and additional information is **crucial**

High cognitive burden
Proximity (*position*)
not correlated with
Similarity (*color/shape*)



Can you spot the outlier?

Can you spot the outlier?

Low cognitive burden
Proximity
correlated with
Similarity

[Haroz&Whitney, InfoVis2012]

Error in *Position?*
or
error in *Color?*

Strengths of Dimensionality Reduction

- **Correlation** between **spatial arrangement** and additional information is **crucial**

High cognitive burden
Proximity (*position*)
not correlated with
Similarity (*color/shape*)

Can you spot the outlier?

Can you spot the outlier?

Low cognitive burden
Proximity
correlated with
Similarity

[Haroz&Whitney, InfoVis2012]

Error in *Position?*
or
error in *Color?*

Position and *Color* are **coherent with external information**

Strengths of Dimensionality Reduction

- **Correlation** between **spatial arrangement** and additional information is **crucial**

High cognitive burden
Proximity (*position*)
not correlated with
Similarity (*color/shape*)

Can you spot the outlier?

Can you spot the outlier?

Low cognitive burden
Proximity
correlated with
Similarity

[Haroz&Whitney, InfoVis2012]

Error in *Position?*
or
error in *Color?*

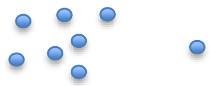
Position and *Color* are **coherent with external information**

Strengths of Dimensionality Reduction

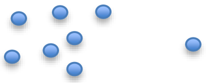
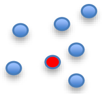
Strengths of Dimensionality Reduction

- **Many interesting (abstract) patterns** are **similarity-based**

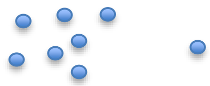
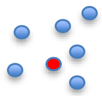
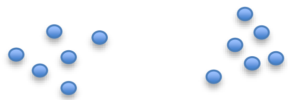
Strengths of Dimensionality Reduction

- **Many interesting (abstract) patterns** are **similarity-based**
  - Outliers 

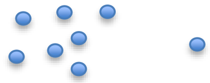
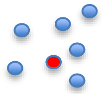
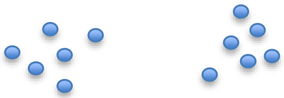
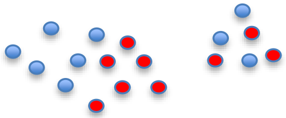
Strengths of Dimensionality Reduction

- **Many interesting (abstract) patterns** are **similarity-based**
  - Outliers 
  - Class outliers 

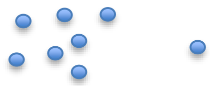
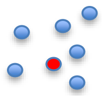
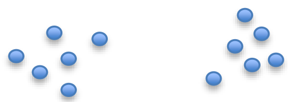
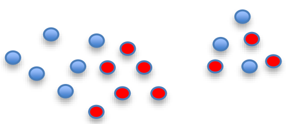
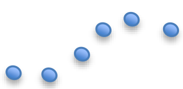
Strengths of Dimensionality Reduction

- **Many interesting (abstract) patterns** are **similarity-based**
  - Outliers 
  - Class outliers 
  - Clusters 

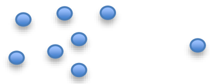
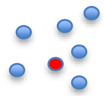
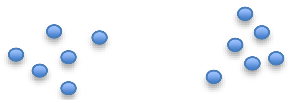
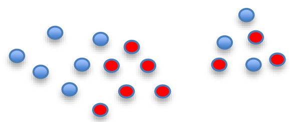
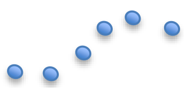
Strengths of Dimensionality Reduction

- **Many interesting (abstract) patterns** are **similarity-based**
  - Outliers 
  - Class outliers 
  - Clusters 
  - Class structure 

Strengths of Dimensionality Reduction

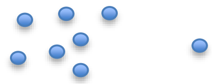
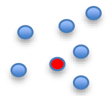
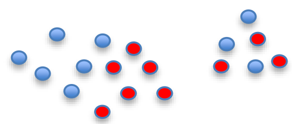
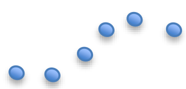
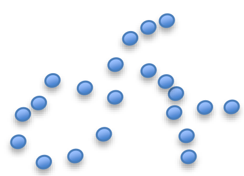
- **Many interesting (abstract) patterns** are **similarity-based**
  - Outliers 
  - Class outliers 
  - Clusters 
  - Class structure 
  - Path 

Strengths of Dimensionality Reduction

- **Many interesting (abstract) patterns** are **similarity-based**
  - Outliers 
  - Class outliers 
  - Clusters 
  - Class structure 
  - Path 
  - Shape

Strengths of Dimensionality Reduction

- **Many interesting (abstract) patterns** are **similarity-based**

  - Outliers 
  - Class outliers 
  - Clusters 
  - Class structure 
  - Path 
  - Shape 

- **DR** preserve similarities so is **likley to preserve these patterns**

Outline

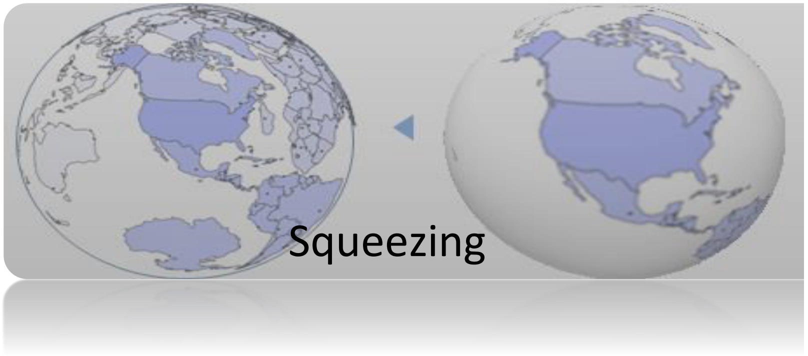
- What is Dimensionality Reduction
- Strengths
- **Weaknesses**
- Turning weaknesses into strengths
- Conclusion
- Perspectives

Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving

Weaknesses of Dimensionality Reduction

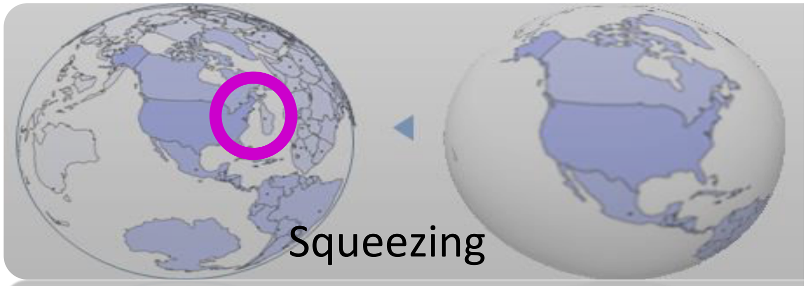
- Mapping **distortions** prevent perfect similarity preserving



Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving

False neighbors



Is Madagascar that
close to New-York?

Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving

False neighbors

Is Madagascar that
close to New-York?

Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving

Is Madagascar that close to New-York?

Do Alaska and Siberia are so far apart?

Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving

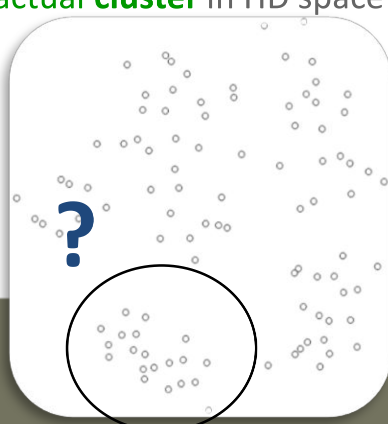
False neighbors

Missed neighbors

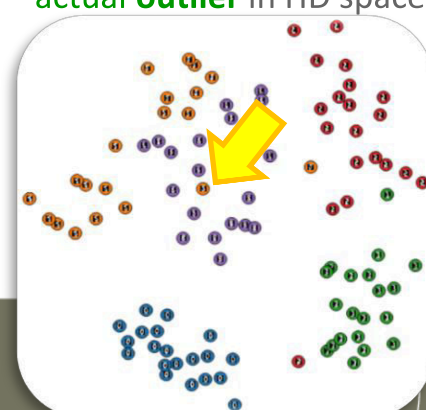
Is Madagascar that close to New-York?

Do Alaska and Siberia are so far apart?

DR artifact or actual **cluster** in HD space ?



DR artifact or actual **outlier** in HD space?



Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving

Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving

- **Inference** of HD patterns from 2D ones very likely to be **wrong**

Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving
- **Inference** of HD patterns from 2D ones very likely to be **wrong**
- Consequences for users

Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving
- **Inference** of HD patterns from 2D ones very likely to be **wrong**
- Consequences for users
  - **blind trust**

Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving

- **Inference** of HD patterns from 2D ones very likely to be **wrong**

- Consequences for users
  - **blind trust**
  - **no trust**

Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving

- **Inference** of HD patterns from 2D ones very likely to be **wrong**

- Consequences for users
  - **blind trust**
  - **no trust**
  - or use **external knowledge** for checking

Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving

- **Inference** of HD patterns from 2D ones very likely to be **wrong**

- Consequences for users
  - **blind trust**
  - **no trust**
  - or use **external knowledge** for checking

- Mapping distortions must be

Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving
- **Inference** of HD patterns from 2D ones very likely to be **wrong**
- Consequences for users
  - **blind trust**
  - **no trust**
  - or use **external knowledge** for checking
- Mapping distortions must be
  - **displayed**

Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving

- **Inference** of HD patterns from 2D ones very likely to be **wrong**

- Consequences for users
  - **blind trust**
  - **no trust**
  - or use **external knowledge** for checking

- Mapping distortions must be
  - **displayed**
  - **reduced**

Weaknesses of Dimensionality Reduction

- Mapping **distortions** prevent perfect similarity preserving

- **Inference** of HD patterns from 2D ones very likely to be **wrong**

- Consequences for users
  - **blind trust**
  - **no trust**
  - or use **external knowledge** for checking

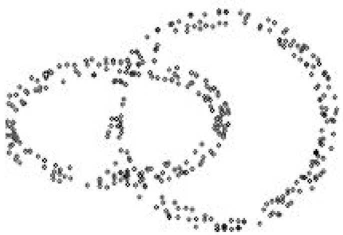
- Mapping distortions must be
  - **displayed**
  - **reduced**
  - or **overcome**

Weaknesses of Dimensionality Reduction

- Displaying **map accuracy** **as a number**

75%

85%

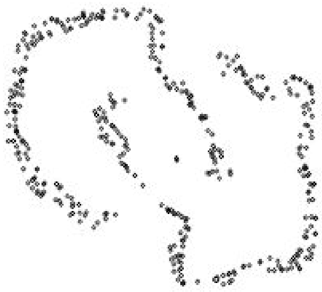


70%

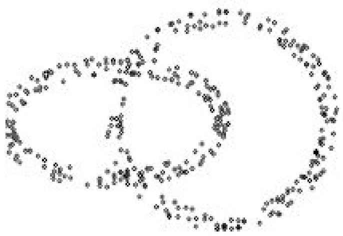
Weaknesses of Dimensionality Reduction

- Displaying **map accuracy** **as a number**

75%



85%



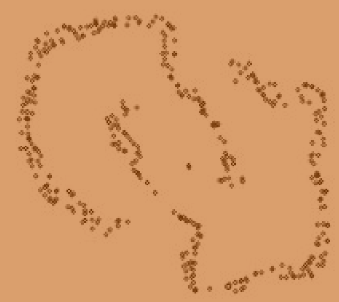
70%

- **Support** **selection** of the most accurate map

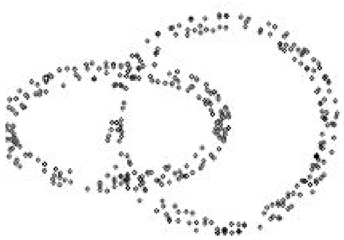
Weaknesses of Dimensionality Reduction

- Displaying **map accuracy as a number**

75%



85%

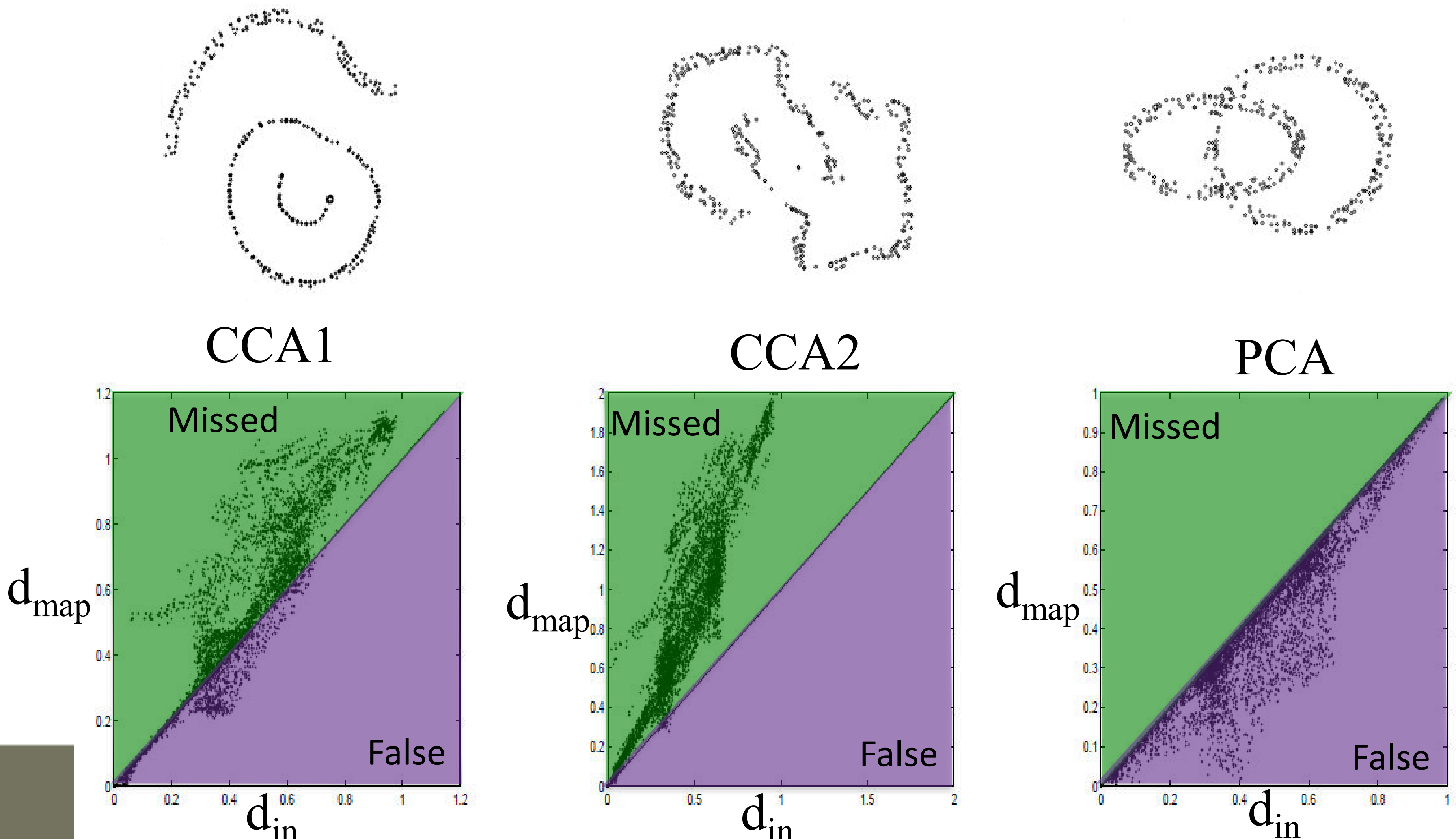


70%

- **Support selection** of the most accurate map
- **Do not support interpretation** even of the most accurate map
  - Are the 15% remaining distortions **spread** all over the map or **concentrated** somewhere???

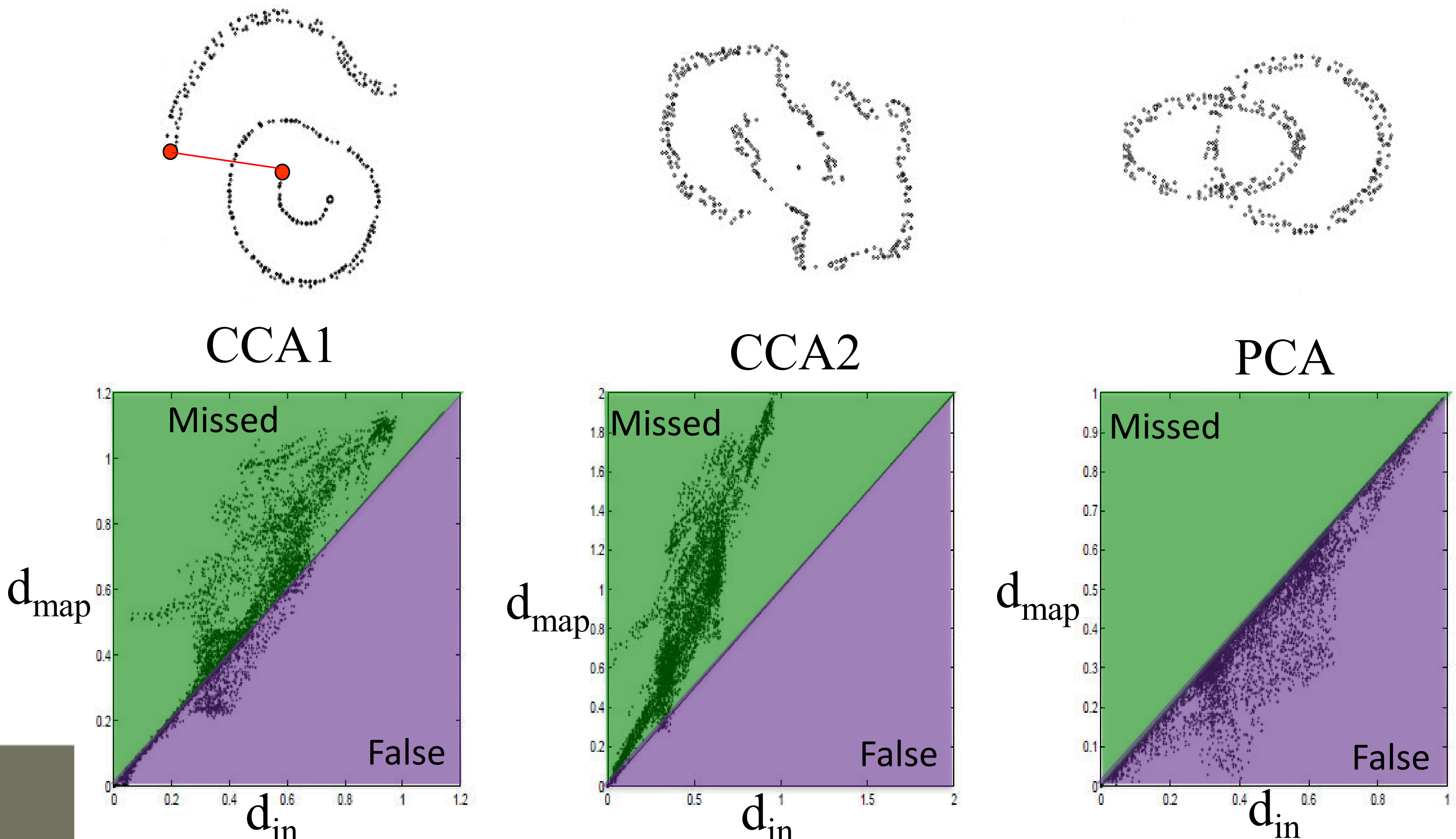
Weaknesses of Dimensionality Reduction

- Displaying **map accuracy** **in an additional view** (Shepard diagram)



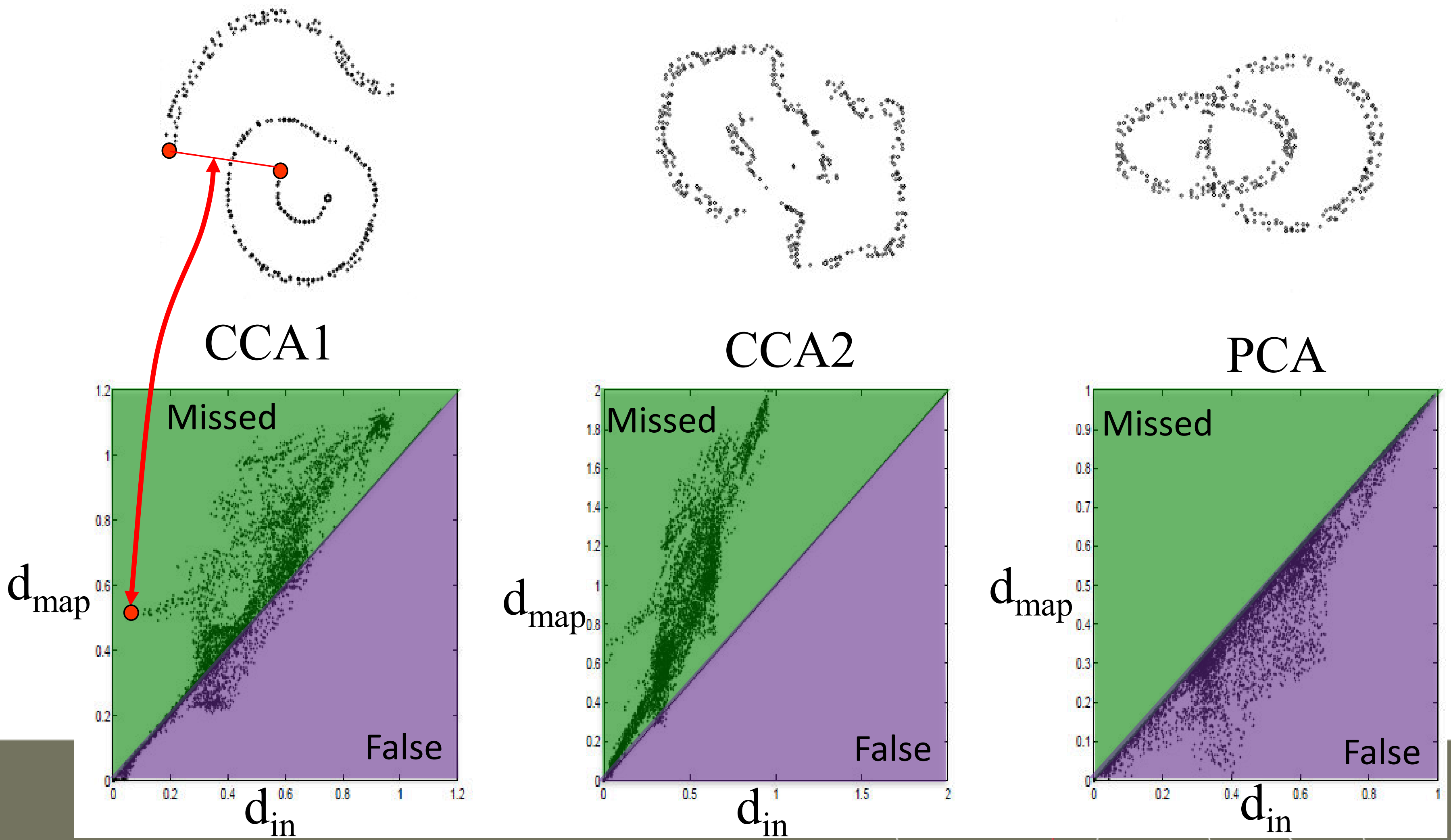
Weaknesses of Dimensionality Reduction

- Displaying **map accuracy in an additional view** (Shepard diagram)



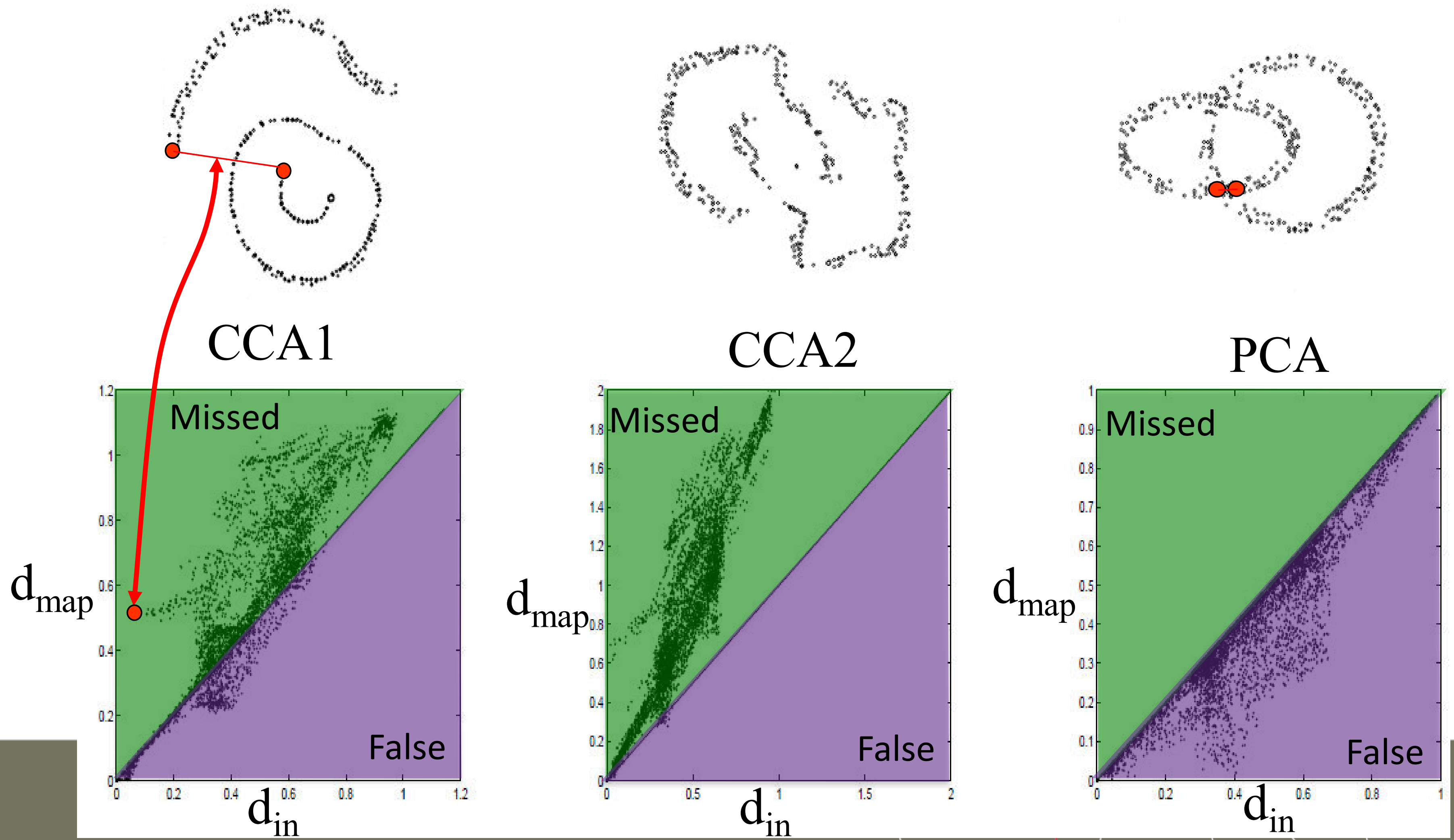
Weaknesses of Dimensionality Reduction

- Displaying **map accuracy** **in an additional view** (Shepard diagram)



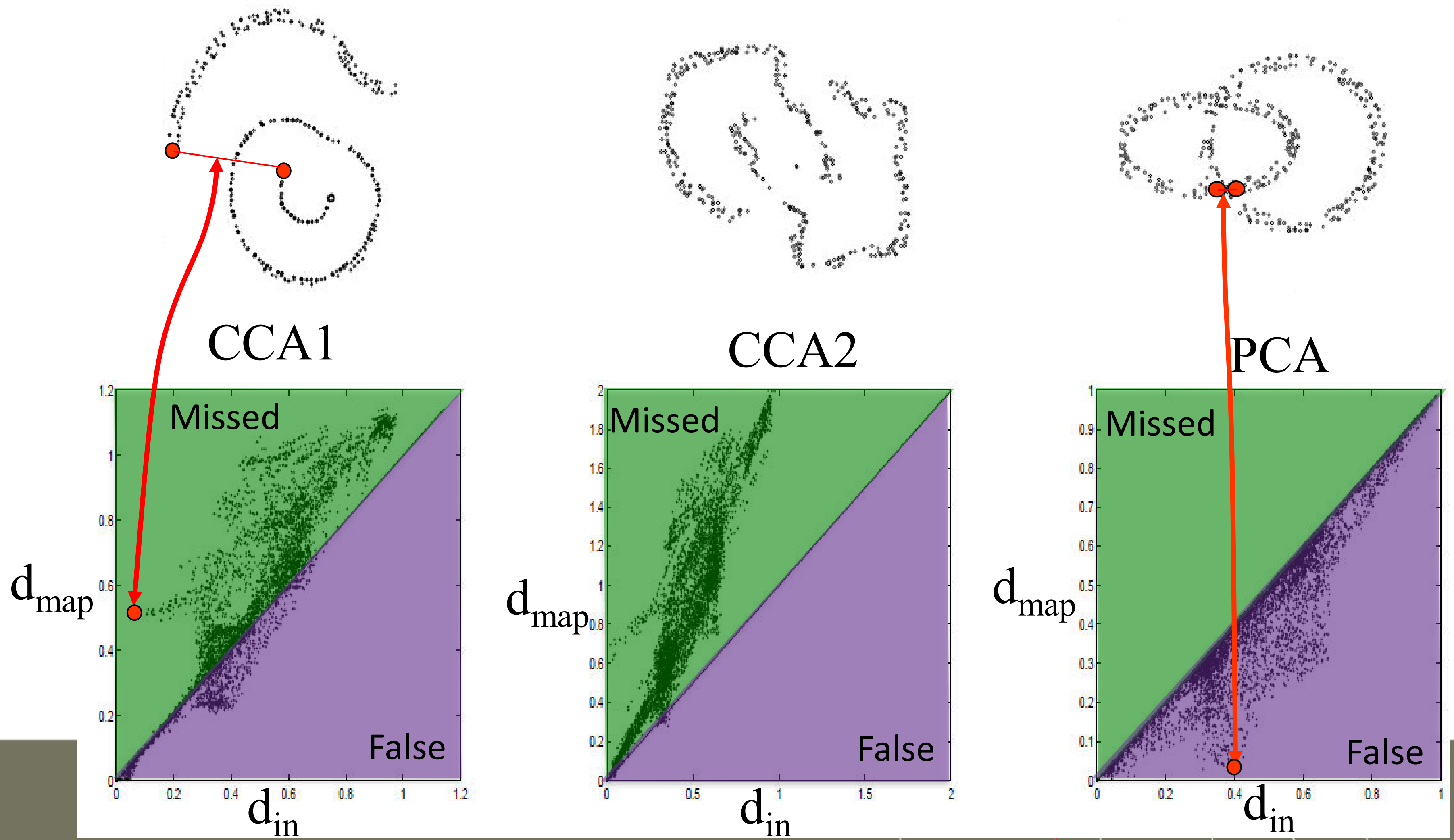
Weaknesses of Dimensionality Reduction

- Displaying **map accuracy in an additional view** (Shepard diagram)



Weaknesses of Dimensionality Reduction

- Displaying **map accuracy** **in an additional view** (Shepard diagram)



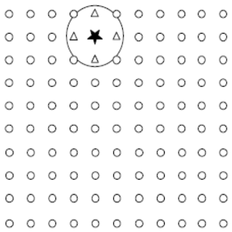
Outline

- What is Dimensionality Reduction
- Strengths
- Weaknesses
- **Turning weaknesses into strengths**
- Conclusion
- Perspectives

Turning weaknesses into strengths

- Use **distortions** within the map **to recover HD patterns**

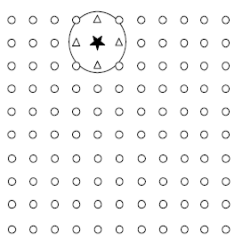
Original



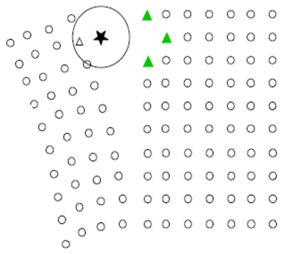
Turning weaknesses into strengths

- Use **distortions** within the map **to recover HD patterns**

Original

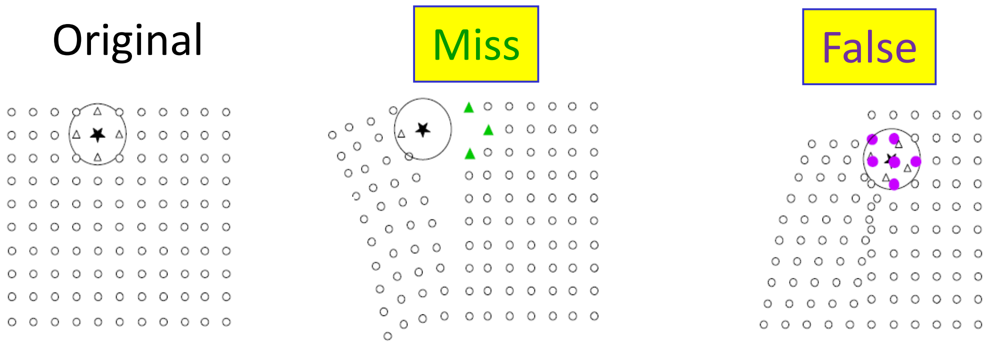


Miss



Turning weaknesses into strengths

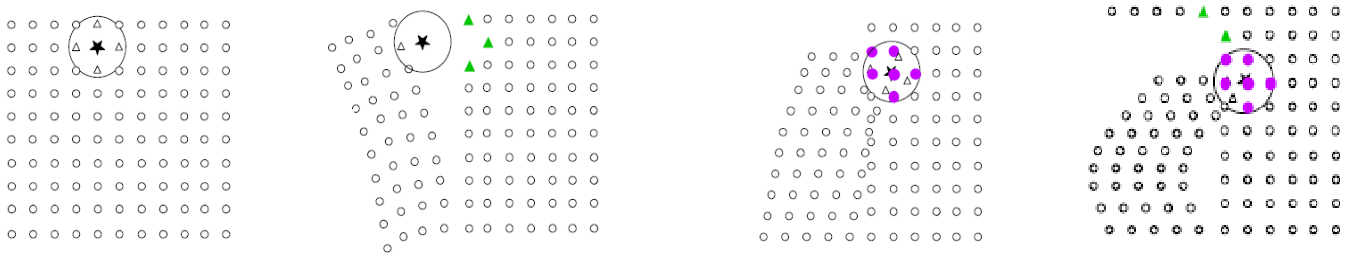
- Use **distortions** within the map **to recover HD patterns**



Turning weaknesses into strengths

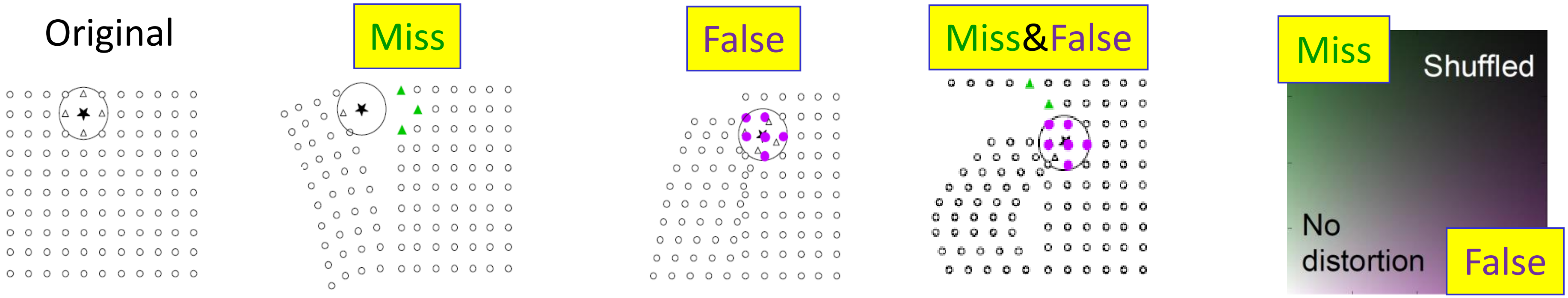
- Use **distortions** within the map **to recover HD patterns**

Original | Miss | False | Miss&False



Turning weaknesses into strengths

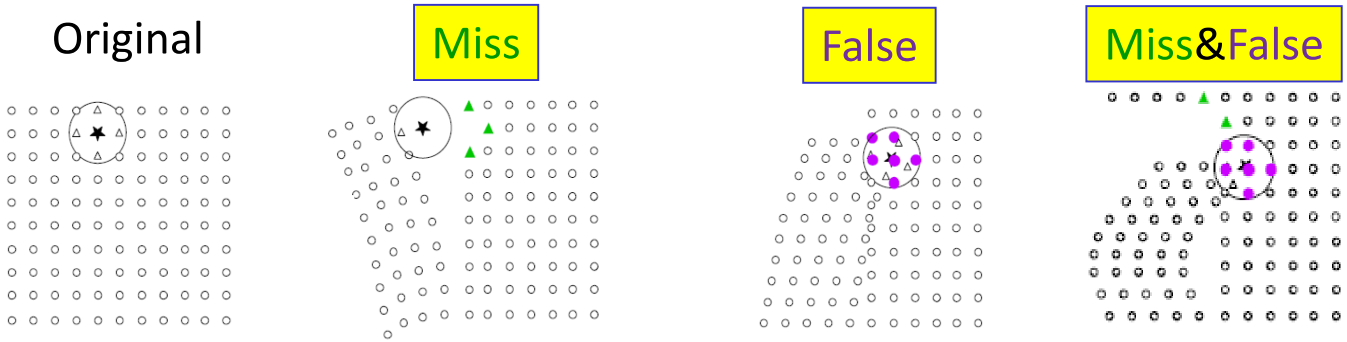
- Use **distortions** within the map **to recover HD patterns**



Perceptually uniform color map

Turning weaknesses into strengths

- Use **distortions** within the map **to recover HD patterns**

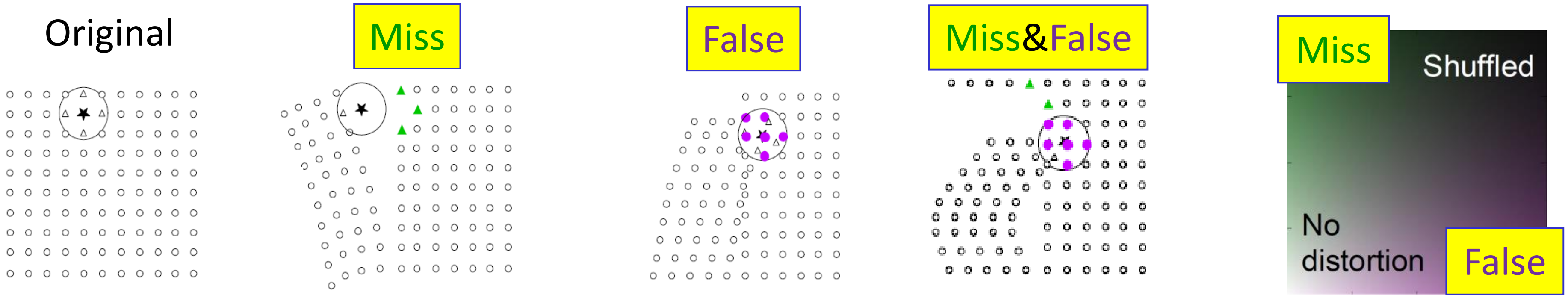


Perceptually uniform color map

Three **graphical inference rules** based on the colored map

Turning weaknesses into strengths

- Use **distortions** within the map **to recover HD patterns**



Perceptually uniform color map

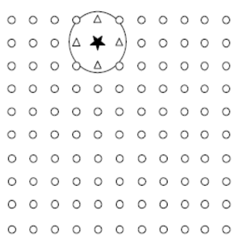
Three **graphical inference rules** based on the colored map

No distortion
Cluster exists in the multivariate space

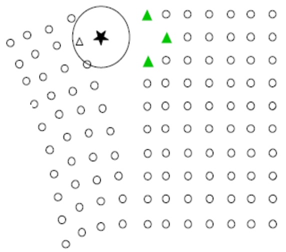
Turning weaknesses into strengths

- Use **distortions** within the map **to recover HD patterns**

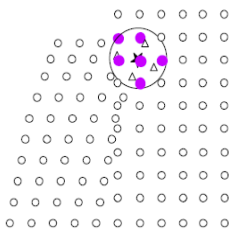
Original



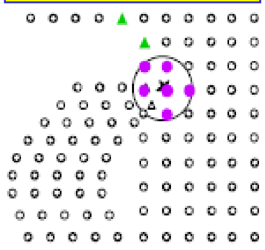
Miss

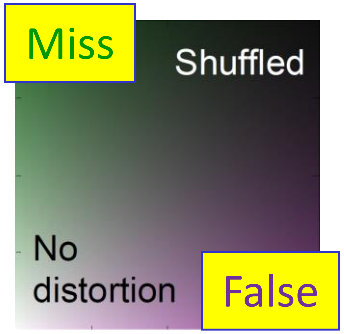


False



Miss&False



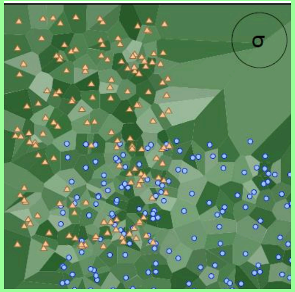


Perceptually uniform color map

Three **graphical inference rules** based on the colored map

No distortion

Cluster exists in the multivariate space

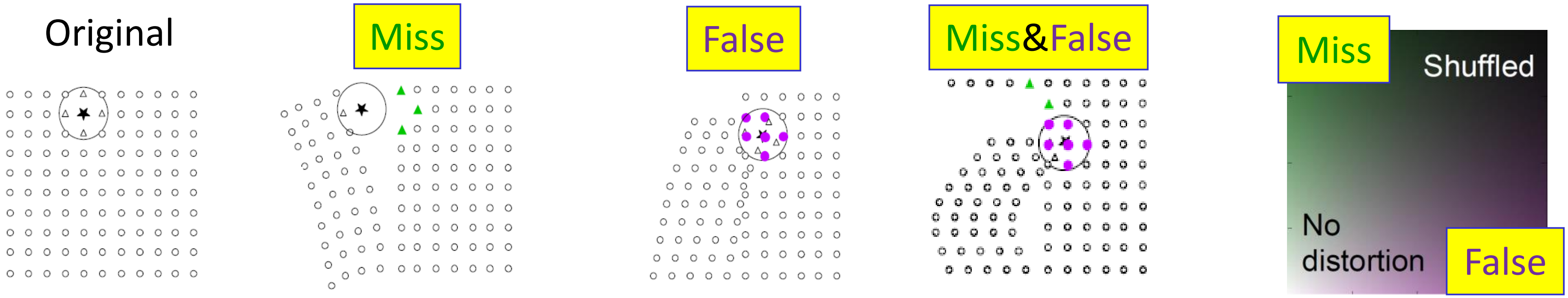


True overlap

Both classes overlap in the multivariate space

Turning weaknesses into strengths

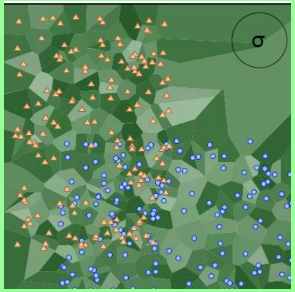
- Use **distortions** within the map **to recover HD patterns**



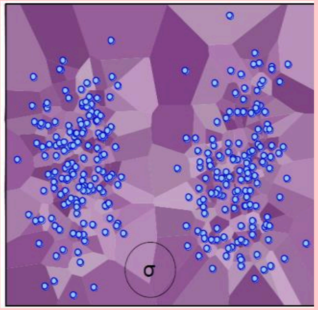
Perceptually uniform color map

Three **graphical inference rules** based on the colored map

No distortion
Cluster exists in the multivariate space



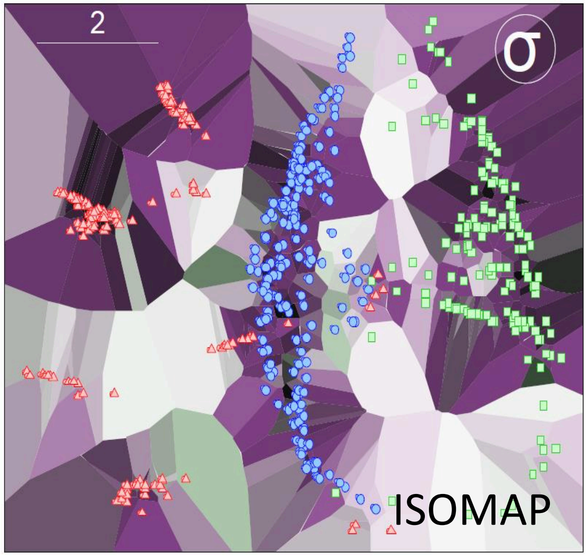
True overlap
Both classes overlap in the multivariate space



True separation
Both clusters are separated in the multivariate space

Turning weaknesses into strengths

- Use **distortions** within the map **to recover HD patterns**

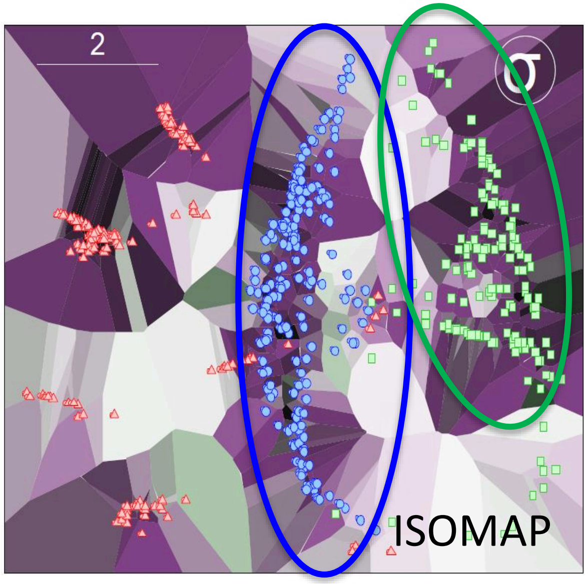


CheckViz

OilFlow data
(N=1000, D=12)

Turning weaknesses into strengths

- Use **distortions** within the map **to recover HD patterns**

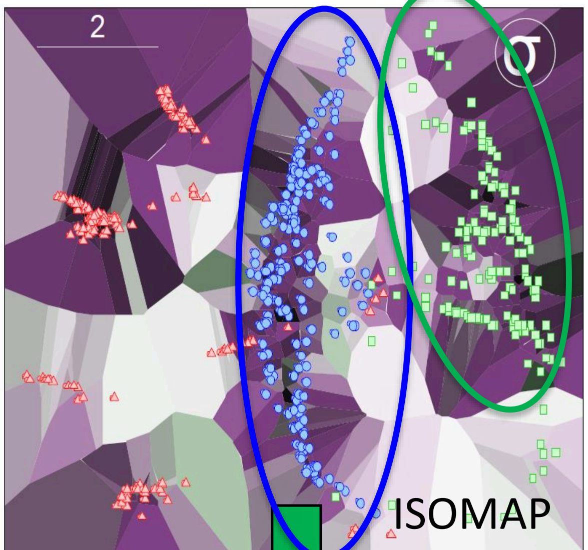


CheckViz

OilFlow data
(N=1000, D=12)

Turning weaknesses into strengths

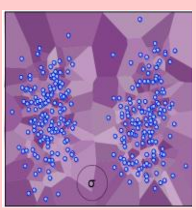
- Use **distortions** within the map **to recover HD patterns**



CheckViz

OilFlow data
(N=1000, D=12)

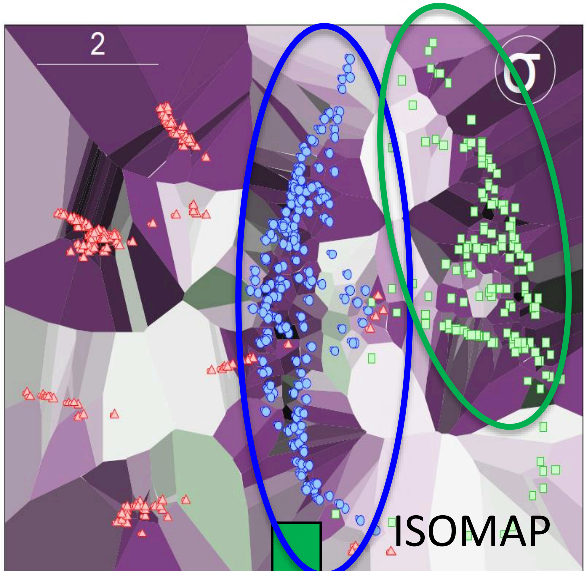
True separation



Blue and **Green** classes
actually separated in the 12D space

Turning weaknesses into strengths

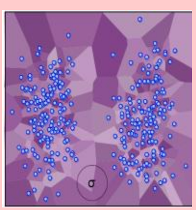
- Use **distortions** within the map **to recover HD patterns**



CheckViz

OilFlow data
(N=1000, D=12)

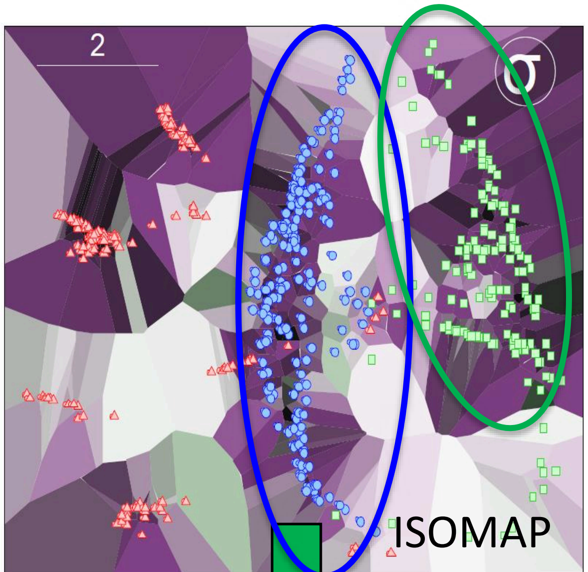
True separation



Blue and **Green** classes
actually separated in the 12D space

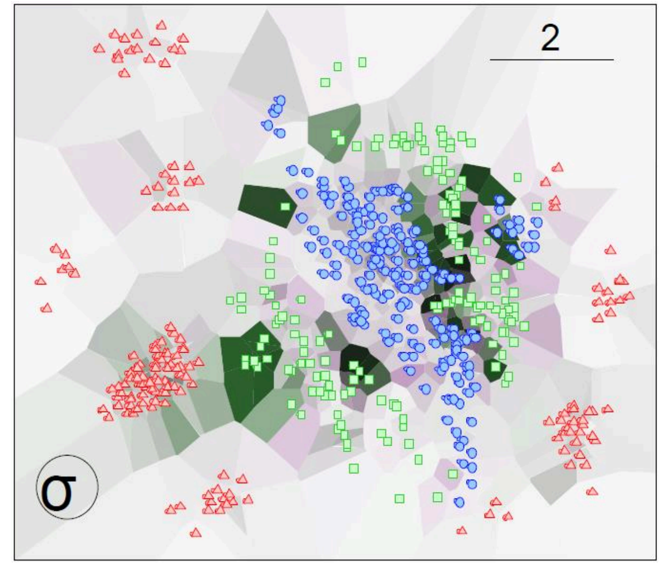
Turning weaknesses into strengths

- Use **distortions** within the map **to recover HD patterns**

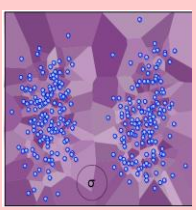


CheckViz

OilFlow data
(N=1000, D=12)



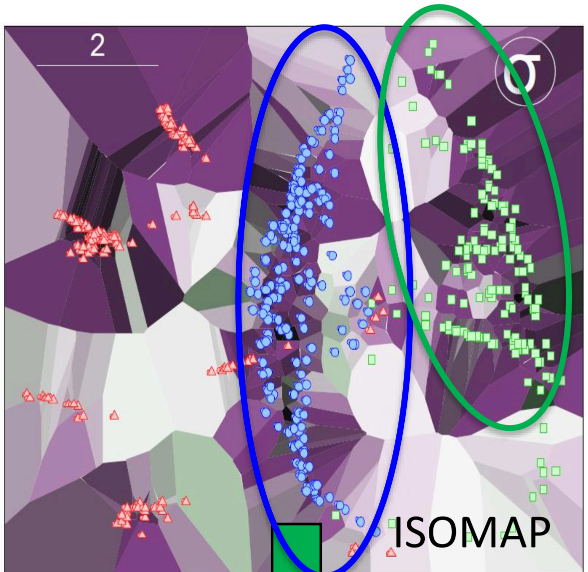
True separation



Blue and **Green** classes
actually separated in the 12D space

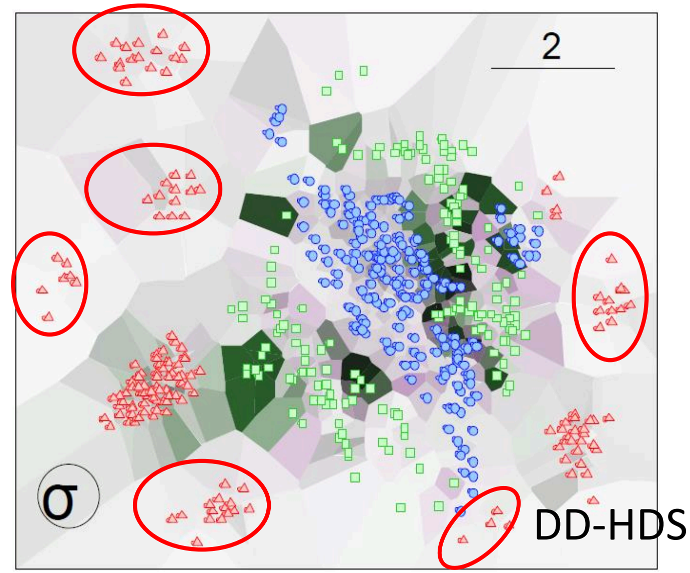
Turning weaknesses into strengths

- Use **distortions** within the map **to recover HD patterns**

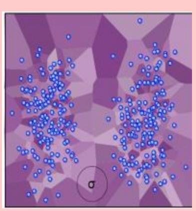


CheckViz

OilFlow data
(N=1000, D=12)



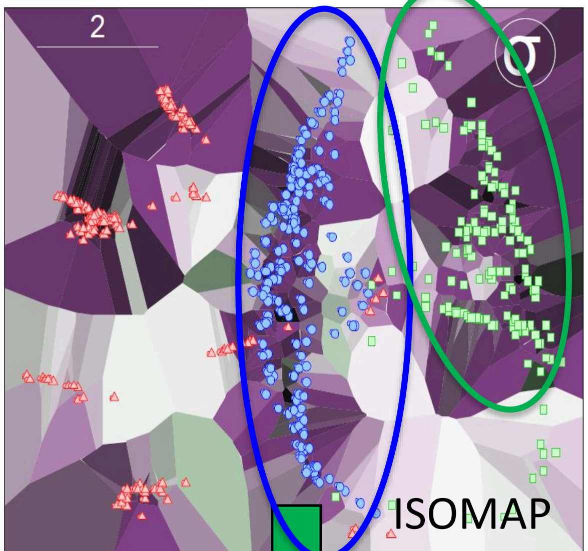
True separation



Blue and **Green** classes
actually separated in the 12D space

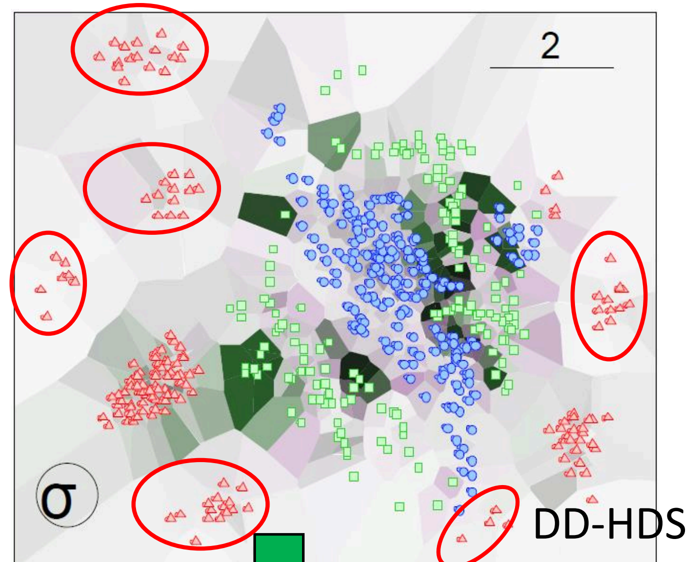
Turning weaknesses into strengths

- Use **distortions** within the map **to recover HD patterns**

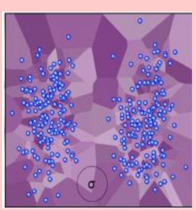


CheckViz

OilFlow data
(N=1000, D=12)



True separation



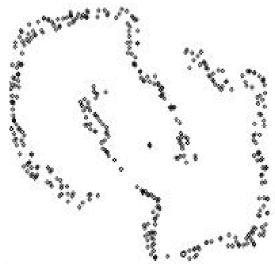
No distortion

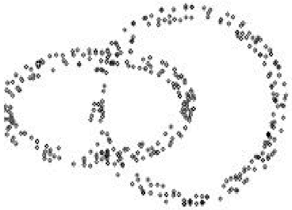
Blue and **Green** classes
actually separated in the 12D space

Red class is **actually disconnected**
in the 12D space

Turning weaknesses into strengths

- Displaying **original similarities** **within the map itself**





Turning weaknesses into strengths

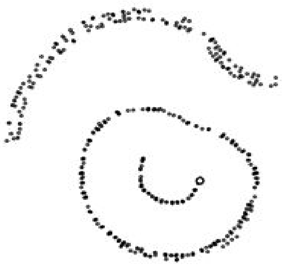
- Displaying **original similarities** **within the map itself**

Not **trustworthy**!!!

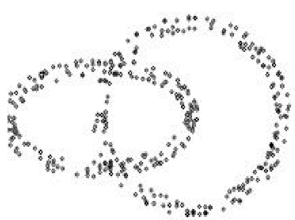
ProxiViz M. Aupetit, **Neurocomputing** 70(7-9):1304-1330, March 2007

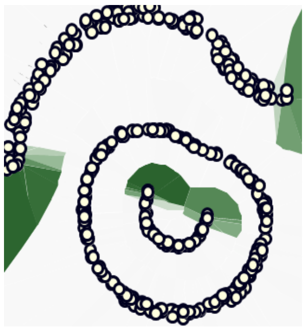
Turning weaknesses into strengths

- Displaying **original similarities within the map itself**



Not **trustworthy**!!!

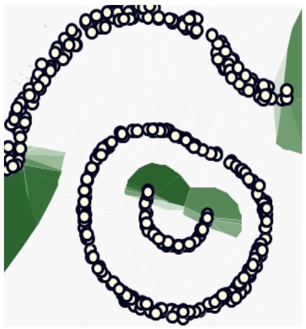




CheckViz

Turning weaknesses into strengths

- Displaying **original similarities** **within the map itself**



CheckViz

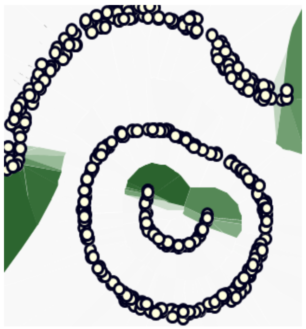
Not **enough**!!!

Not **trustworthy**!!!

Turning weaknesses into strengths

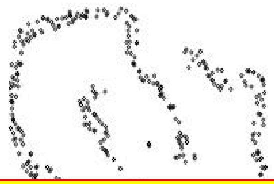
- Displaying **original similarities within the map itself**

ProxiViz

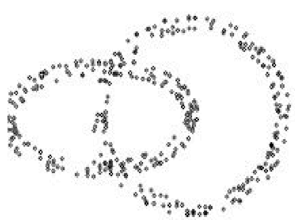


CheckViz

Not **enough**!!!



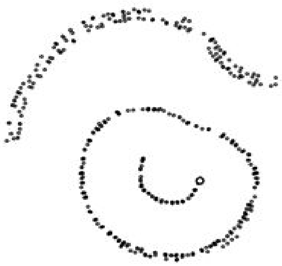
Not **trustworthy**!!!

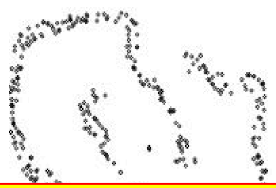


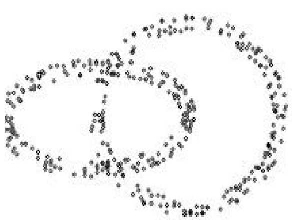
Turning weaknesses into strengths

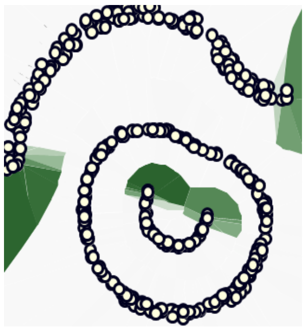
- Displaying **original similarities within the map itself**

ProxiViz









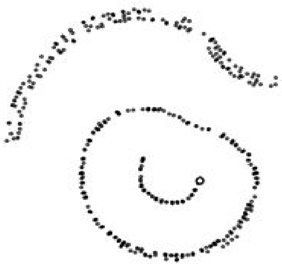
Not **trustworthy**!!!

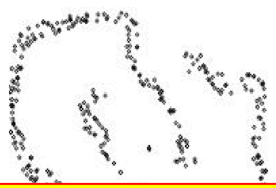
CheckViz

Not **enough**!!!

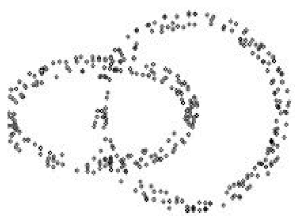
Turning weaknesses into strengths

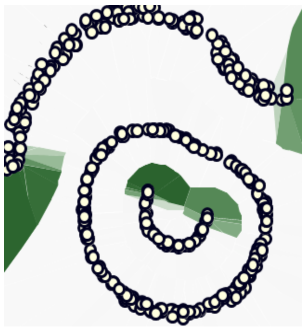
- Displaying **original similarities within the map itself**





Not **trustworthy**!!!

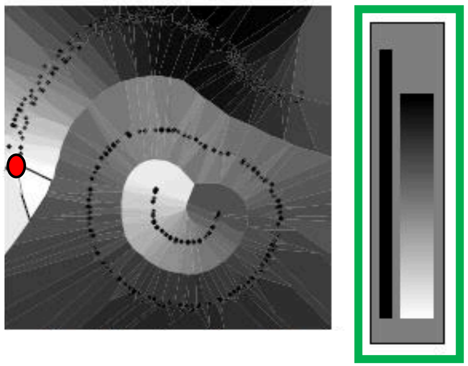




CheckViz

Not **enough**!!!

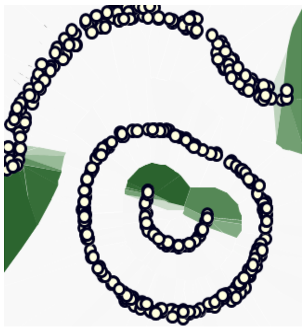
ProxiViz



Turning weaknesses into strengths

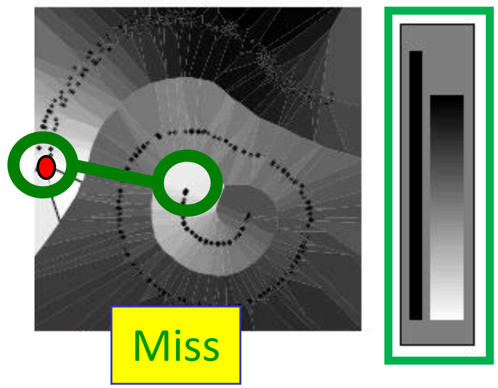
- Displaying **original similarities** **within the map itself**

ProxiViz

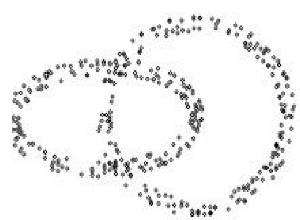


CheckViz

Not enough!!!



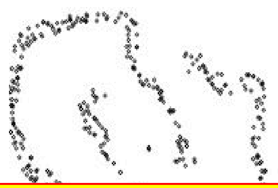
Not trustworthy!!!

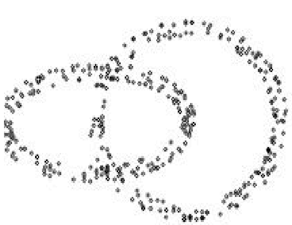


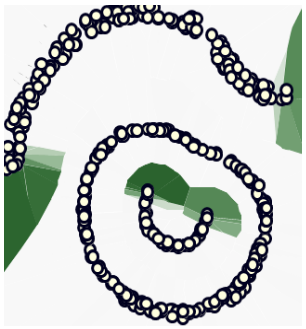
Turning weaknesses into strengths

- Displaying **original similarities within the map itself**

ProxiViz

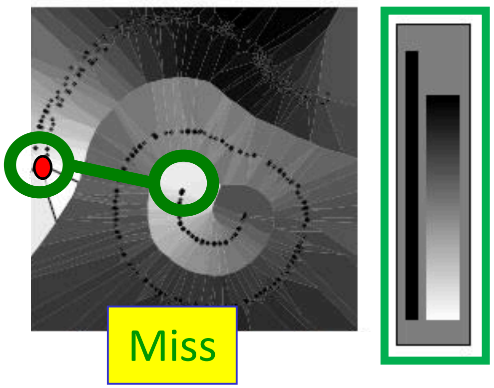






CheckViz

Not **enough**!!!

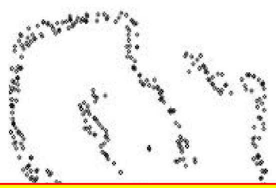


Not **trustworthy**!!!

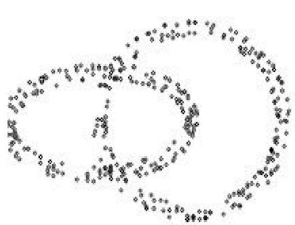
Turning weaknesses into strengths

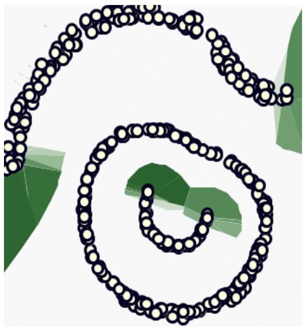
- Displaying **original similarities within the map itself**

ProxiViz



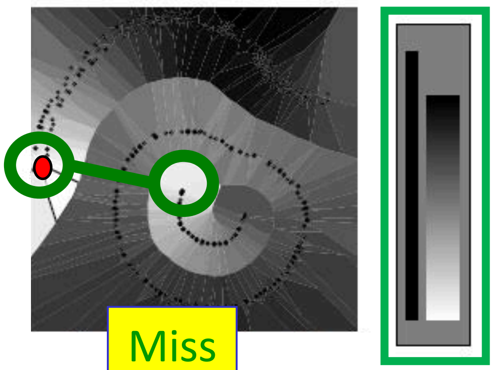
Not **trustworthy**!!!





CheckViz

Not **enough**!!!

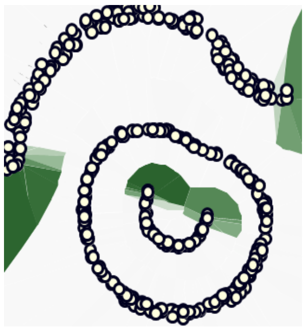


Miss

Turning weaknesses into strengths

- Displaying **original similarities within the map itself**

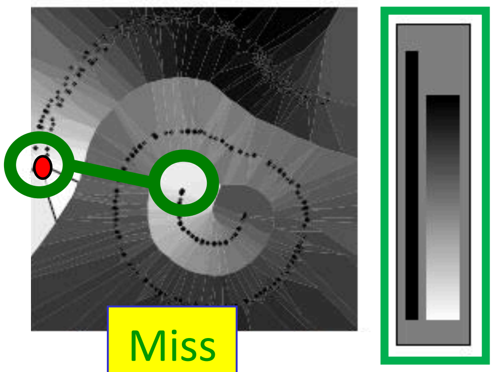
Not **trustworthy**!!!

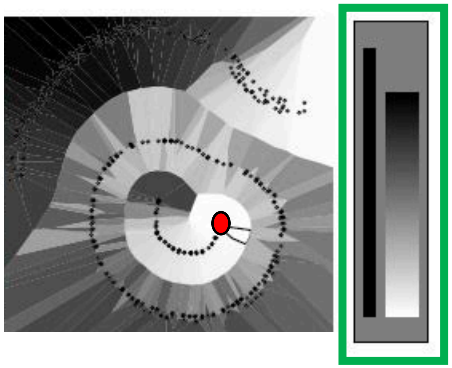


CheckViz

Not **enough**!!!

ProxiViz

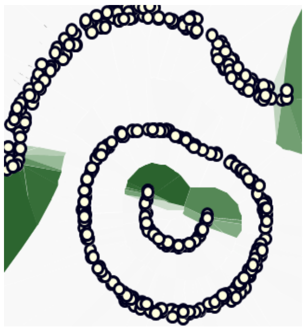




Turning weaknesses into strengths

- Displaying **original similarities within the map itself**

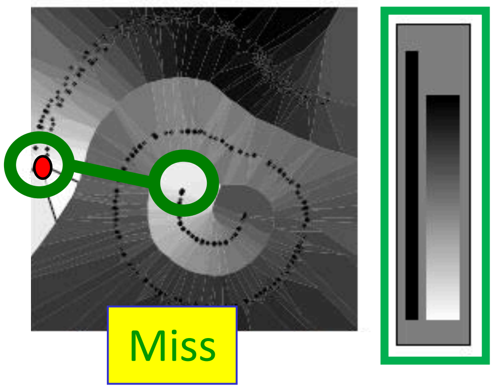
Not **trustworthy**!!!

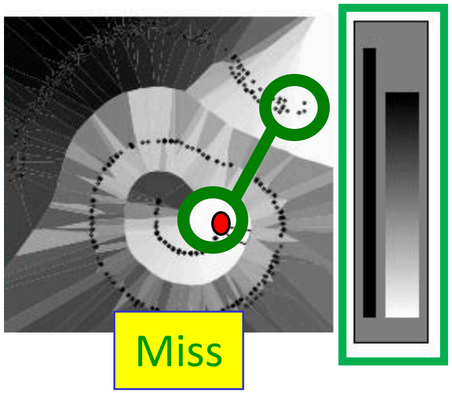


CheckViz

Not **enough**!!!

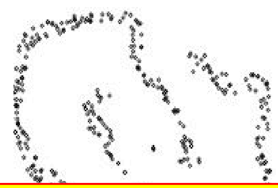
ProxiViz

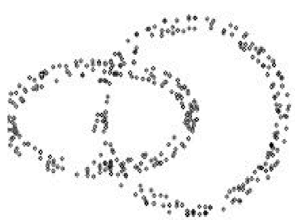




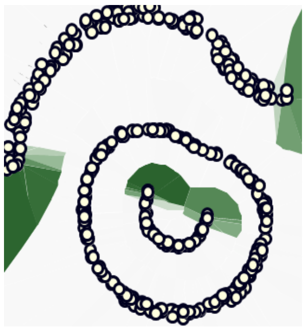
Turning weaknesses into strengths

- Displaying **original similarities** **within the map itself**





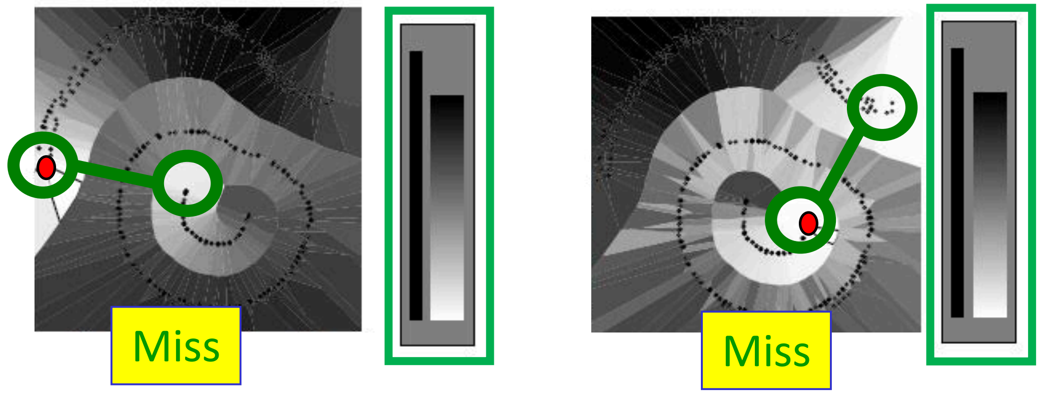
Not **trustworthy**!!!

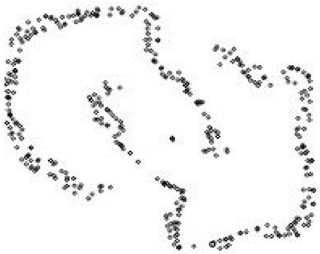


CheckViz

Not **enough**!!!

ProxiViz

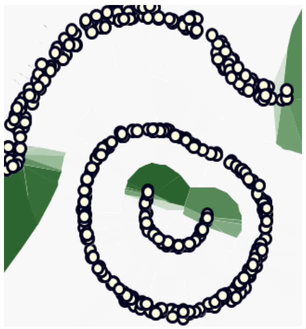




Turning weaknesses into strengths

- Displaying **original similarities** **within the map itself**

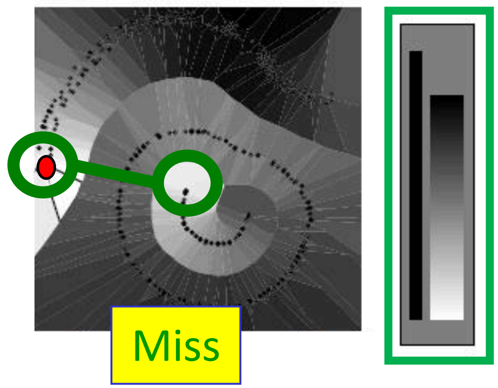
Not **trustworthy**!!!

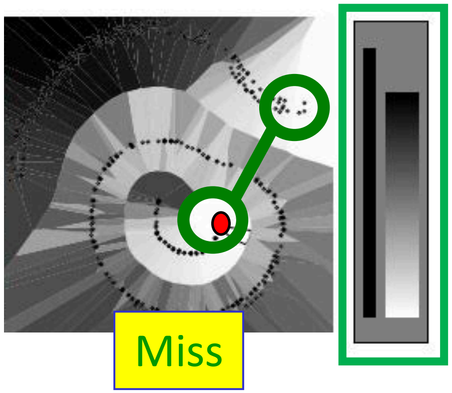


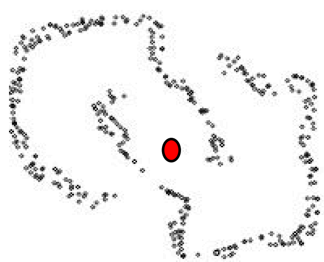
CheckViz

Not **enough**!!!

ProxiViz



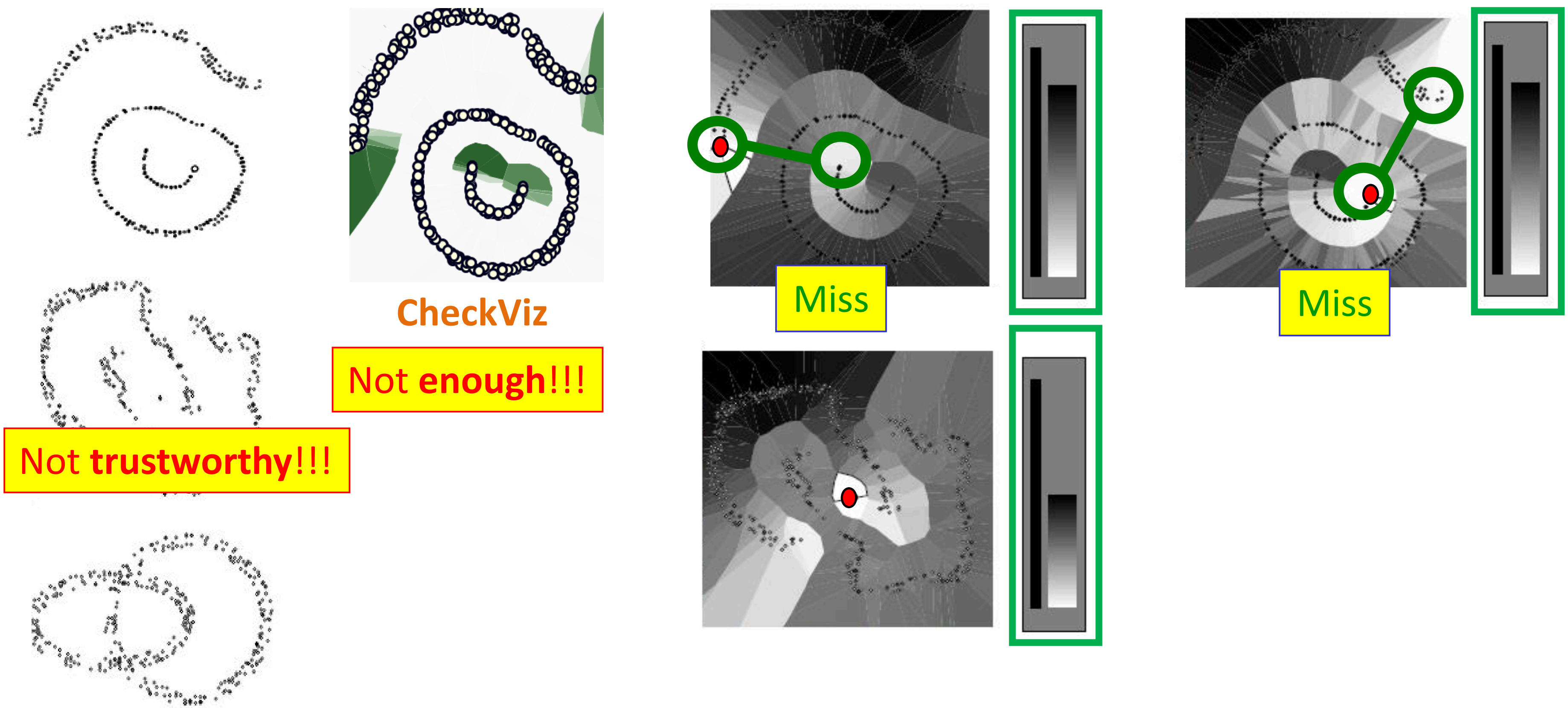




Turning weaknesses into strengths

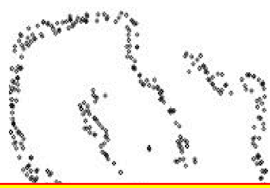
- Displaying **original similarities within the map itself**

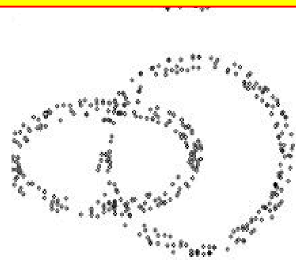
ProxiViz



Turning weaknesses into strengths

- Displaying **original similarities** **within the map itself**



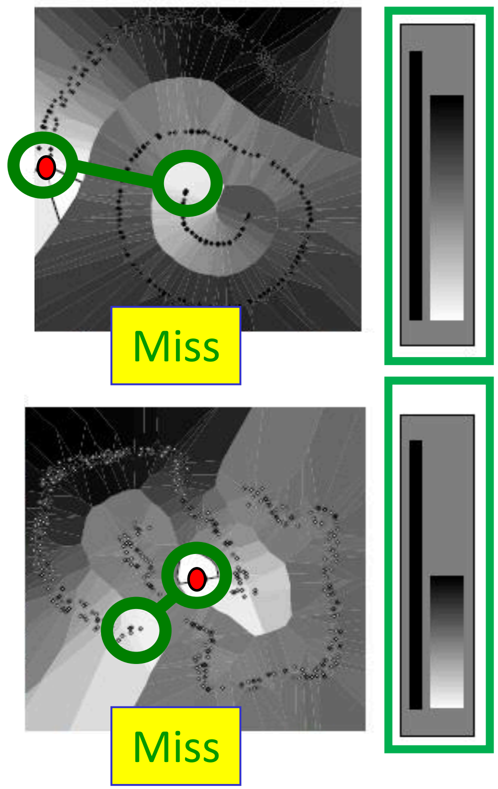


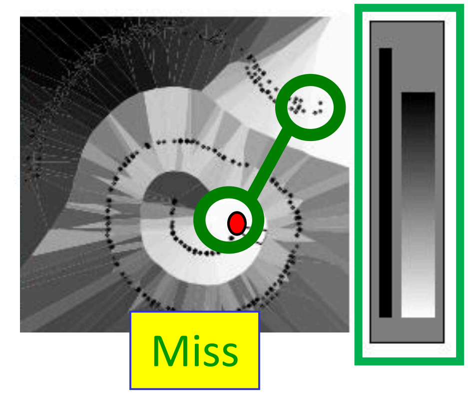
Not **trustworthy**!!!

CheckViz

Not **enough**!!!

ProxiViz

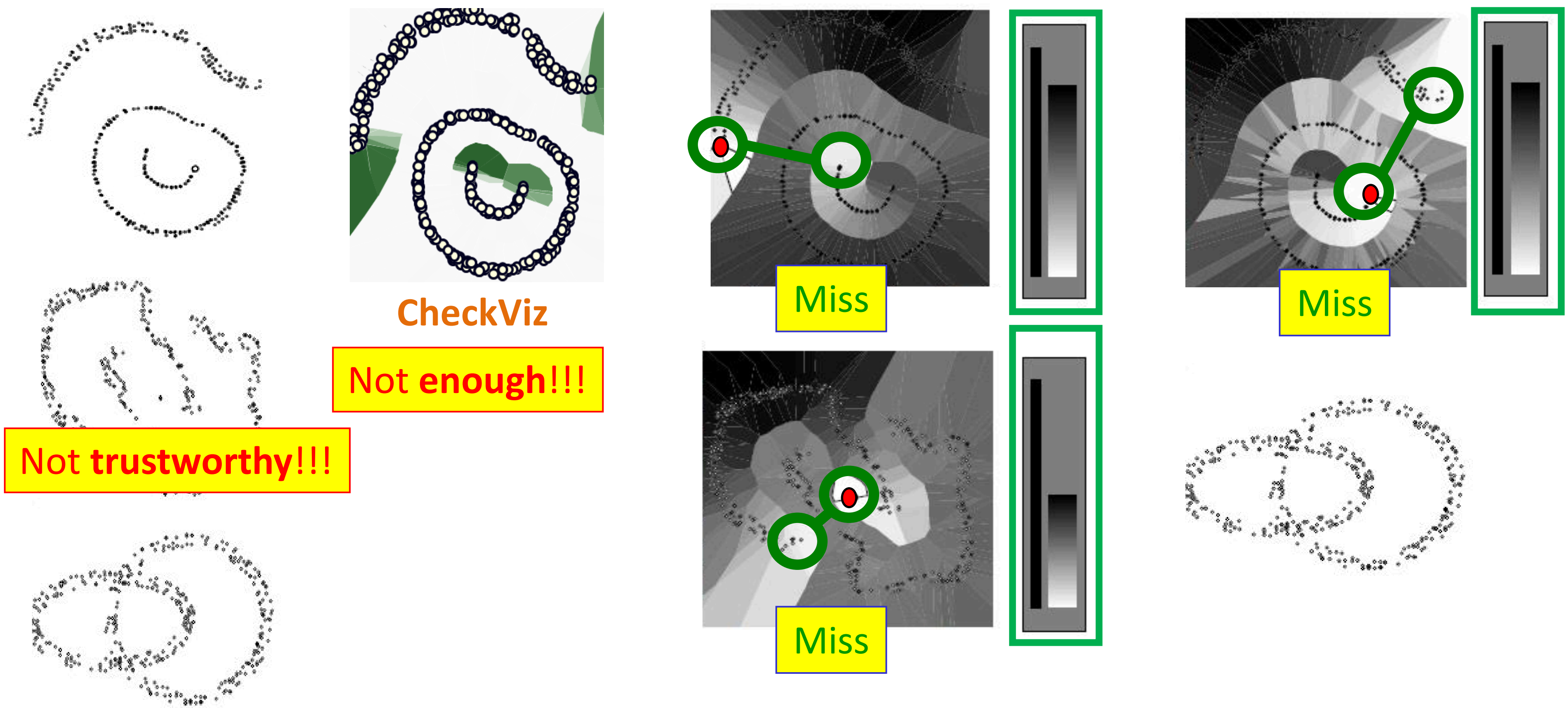




Turning weaknesses into strengths

- Displaying **original similarities within the map itself**

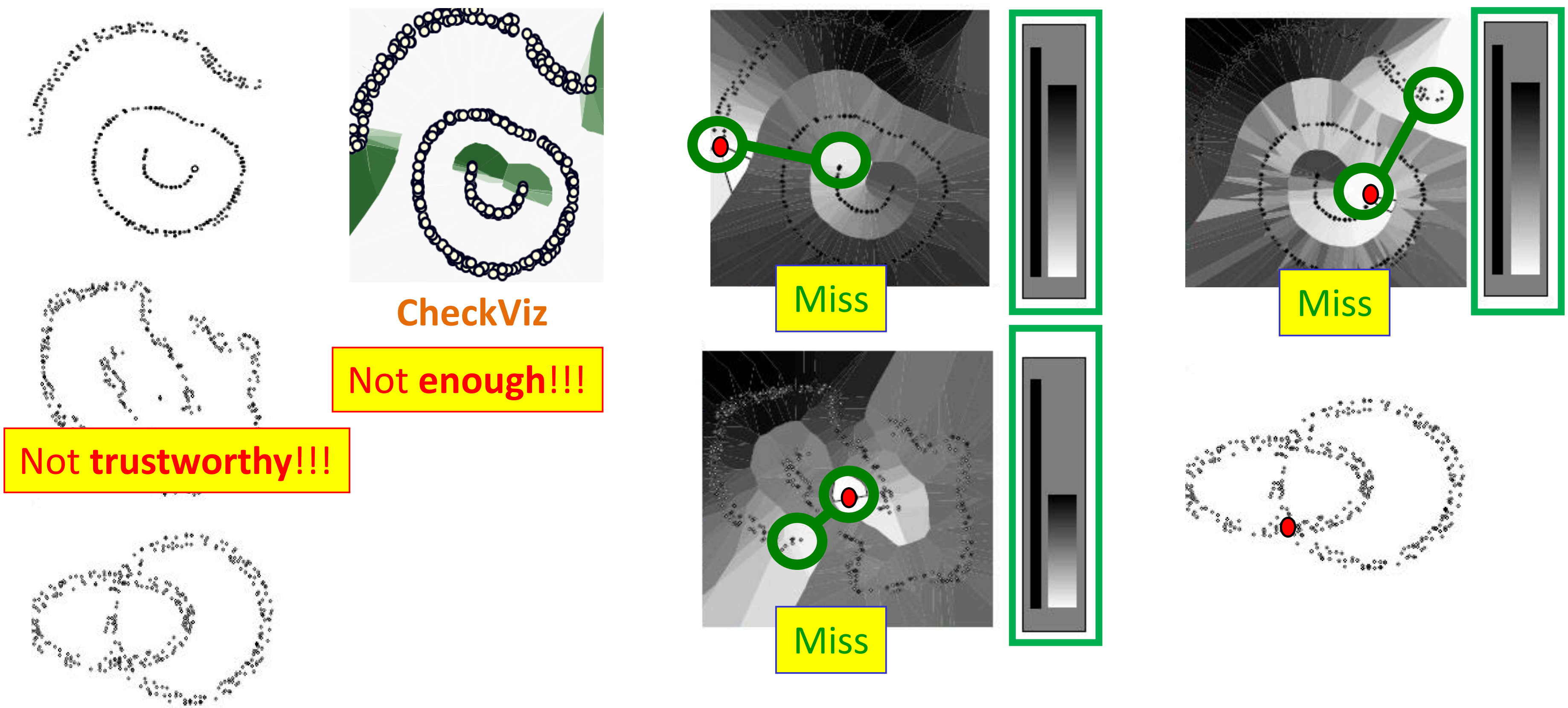
ProxiViz



Turning weaknesses into strengths

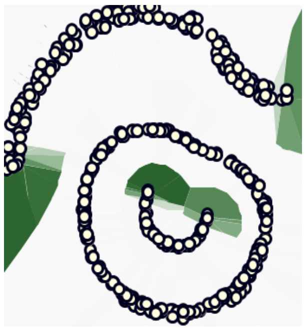
- Displaying **original similarities within the map itself**

ProxiViz



Turning weaknesses into strengths

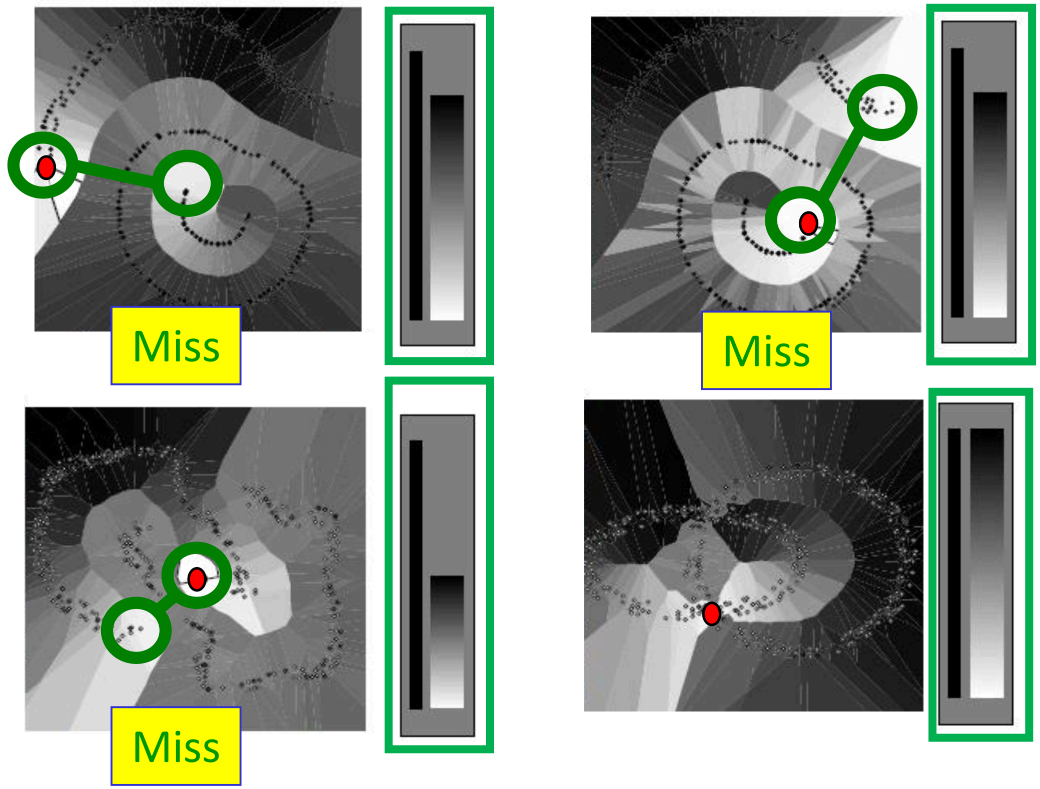
- Displaying **original similarities within the map itself**



CheckViz

Not **enough**!!!

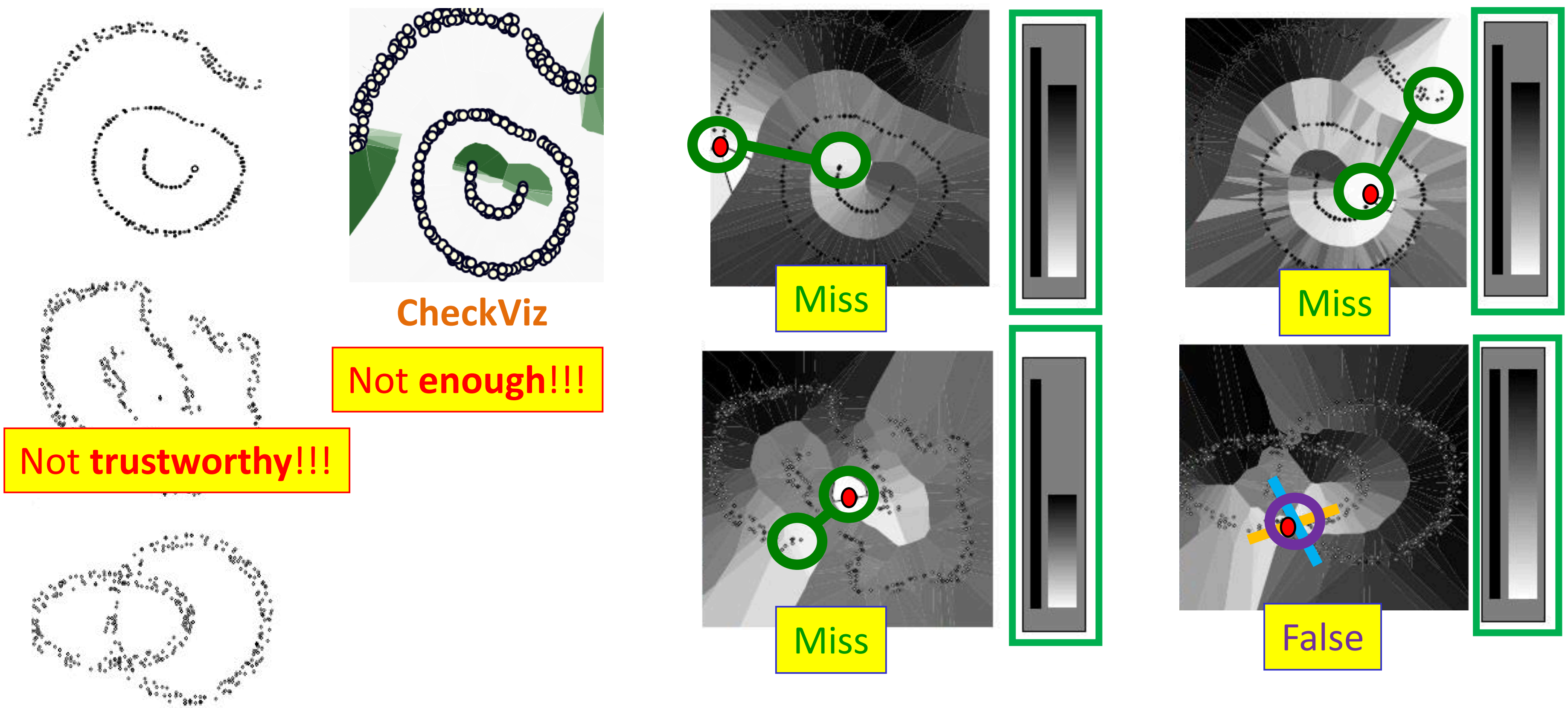
ProxiViz



Turning weaknesses into strengths

- Displaying **original similarities within the map itself**

ProxiViz



Turning weaknesses into strengths

- Displaying **original similarities** **within the map itself**

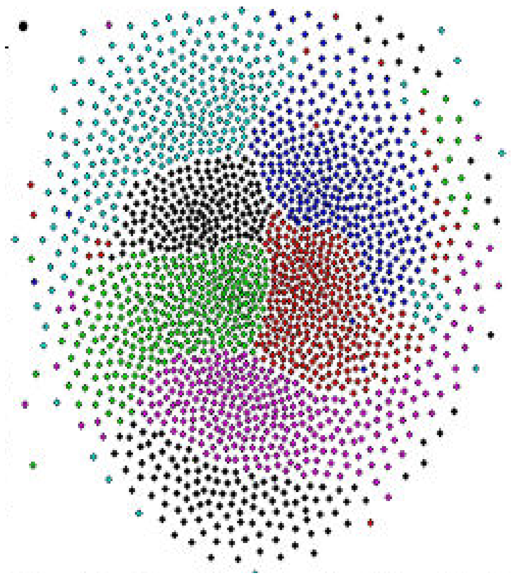
ISOLET data (**D=617**, N=1800, C=6)

Turning weaknesses into strengths

- Displaying **original similarities within the map itself**

ISOLET data (**D=617**, N=1800, C=6)

CCA projection

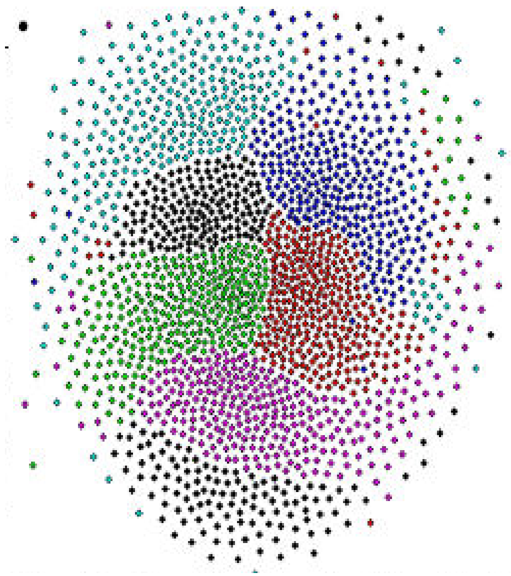


Turning weaknesses into strengths

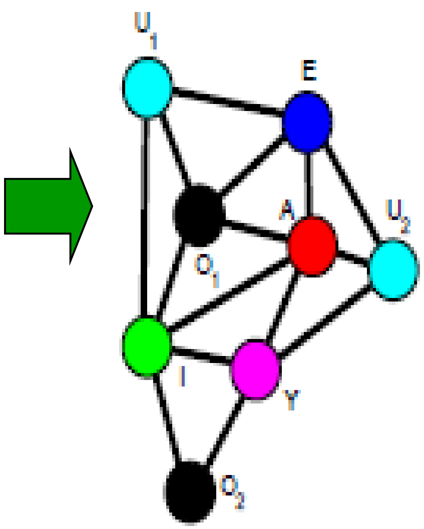
- Displaying **original similarities** **within the map itself**

ISOLET data (**D=617**, N=1800, C=6)

CCA projection



Naively inferred structure is **wrong**

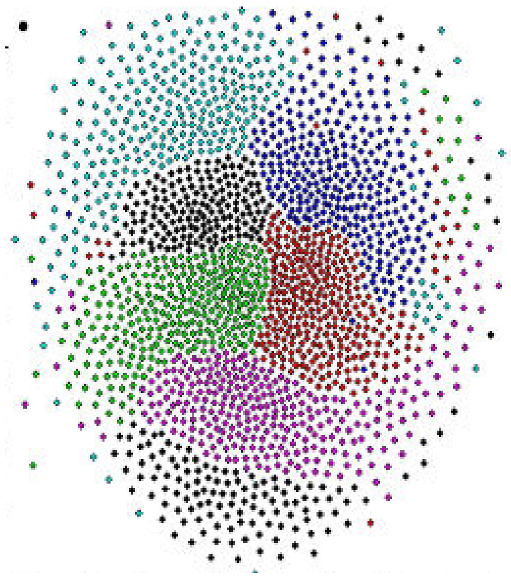


Turning weaknesses into strengths

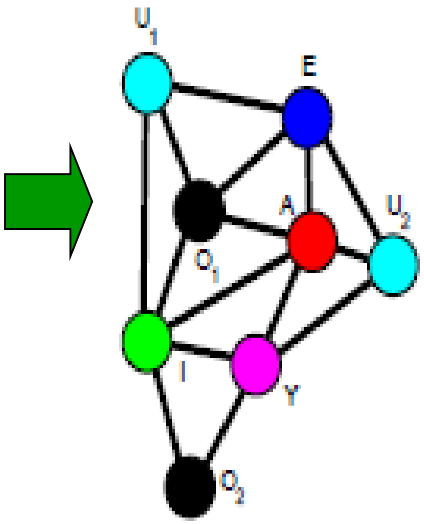
- Displaying **original similarities within the map itself**

ISOLET data (**D=617**, N=1800, C=6)

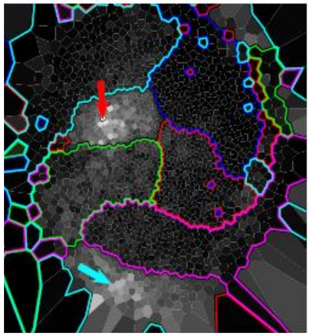
CCA projection



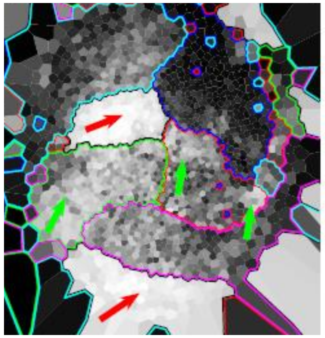
Naively inferred structure is **wrong**



ProxiViz



Within-class proximity



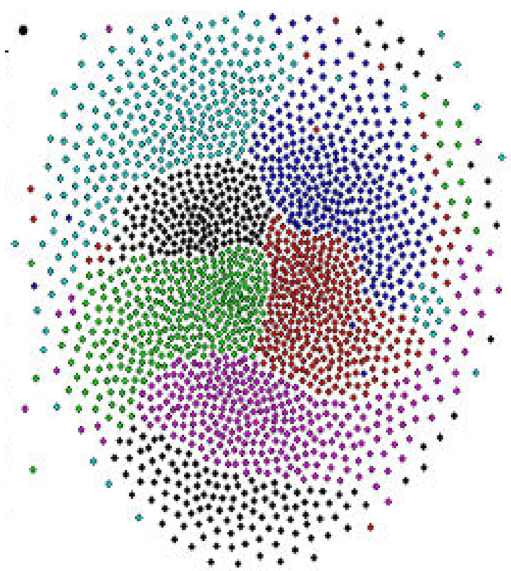
Between-class proximity

Turning weaknesses into strengths

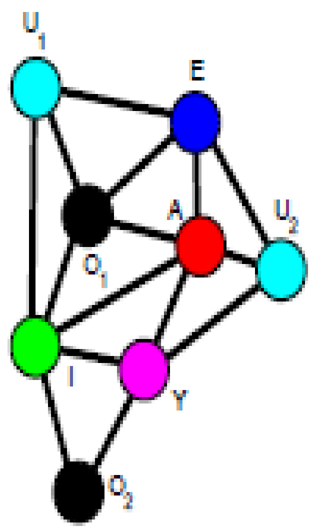
- Displaying **original similarities within the map itself**

ISOLET data (**D=617**, N=1800, C=6)

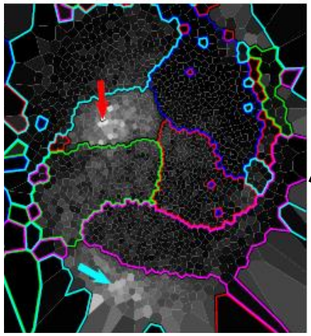
CCA projection



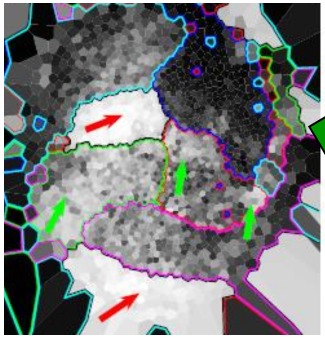
Naively inferred structure is **wrong**



ProxiViz

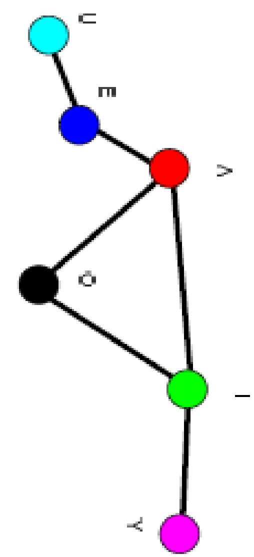


Within-class proximity



Between-class proximity

Likely original structure inferred with ProxiViz

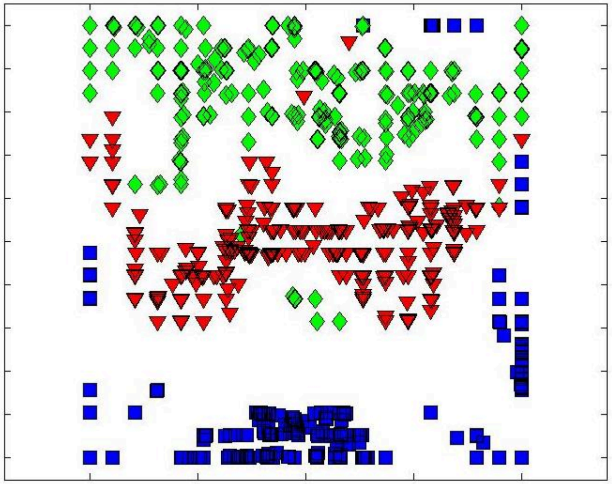


Turning weaknesses into strengths

- Displaying **original similarities within the map itself**

OIL FLOW data (**D=12**, N=1000, C=3)

GTM



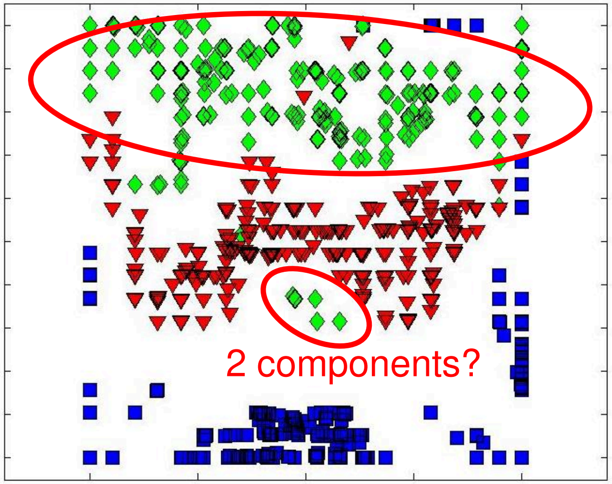
Turning weaknesses into strengths

- Displaying **original similarities** **within the map itself**

OIL FLOW data (**D=12**, N=1000, C=3)

Query

GTM



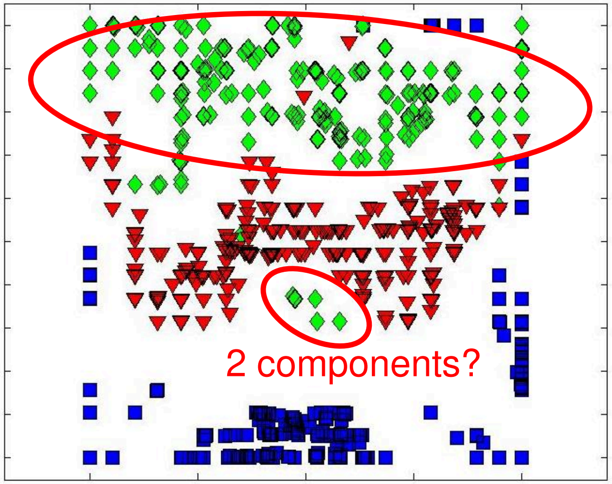
Turning weaknesses into strengths

- Displaying **original similarities within the map itself**

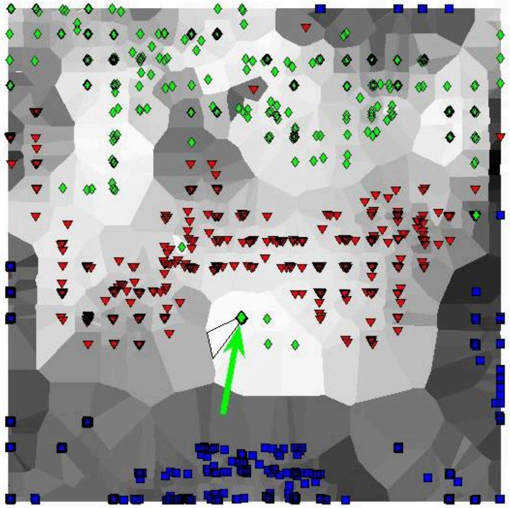
OIL FLOW data (**D=12**, N=1000, C=3)

Query

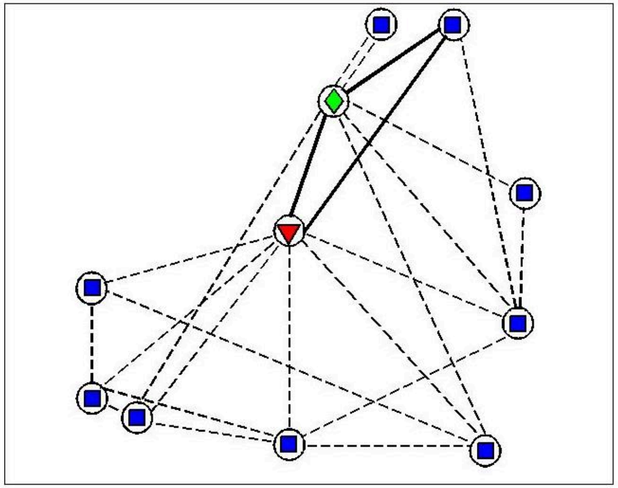
GTM



GTM + ProxiViz



Class graph from
the Supervised Generative Graph



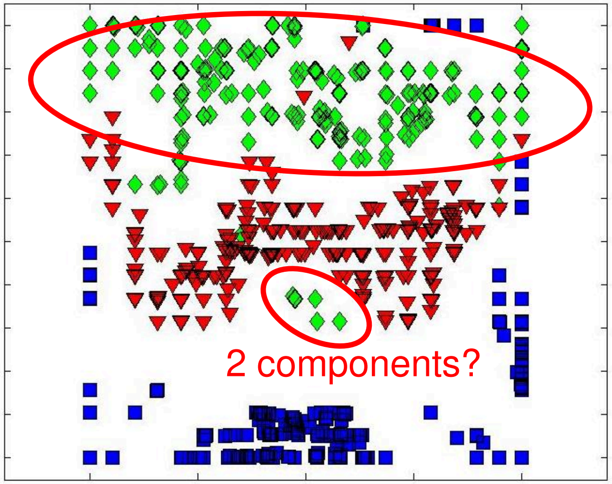
Turning weaknesses into strengths

- Displaying **original similarities within the map itself**

OIL FLOW data (**D=12**, N=1000, C=3)

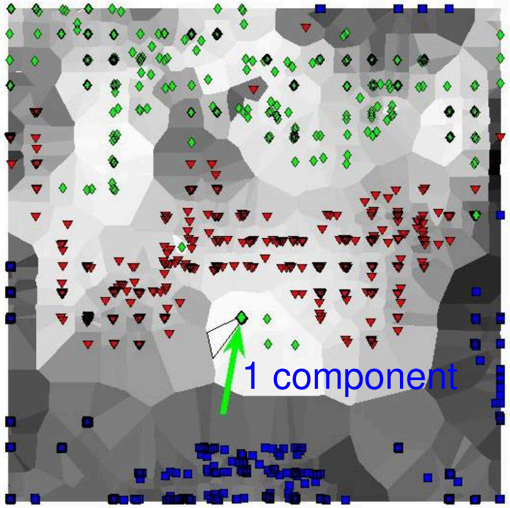
Query

GTM

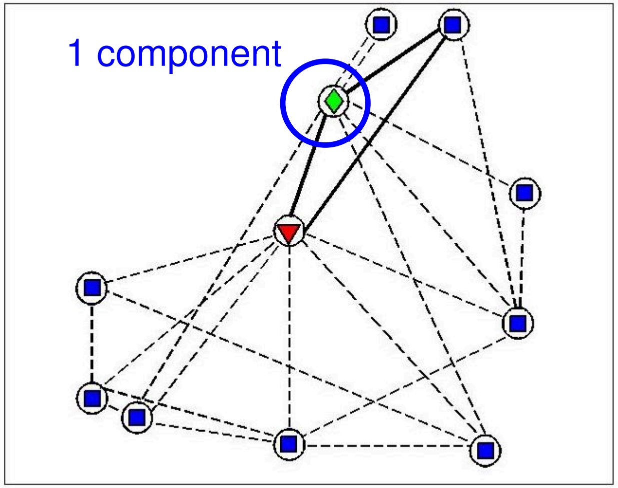


Answer

GTM + ProxiViz



Class graph from the Supervised Generative Graph

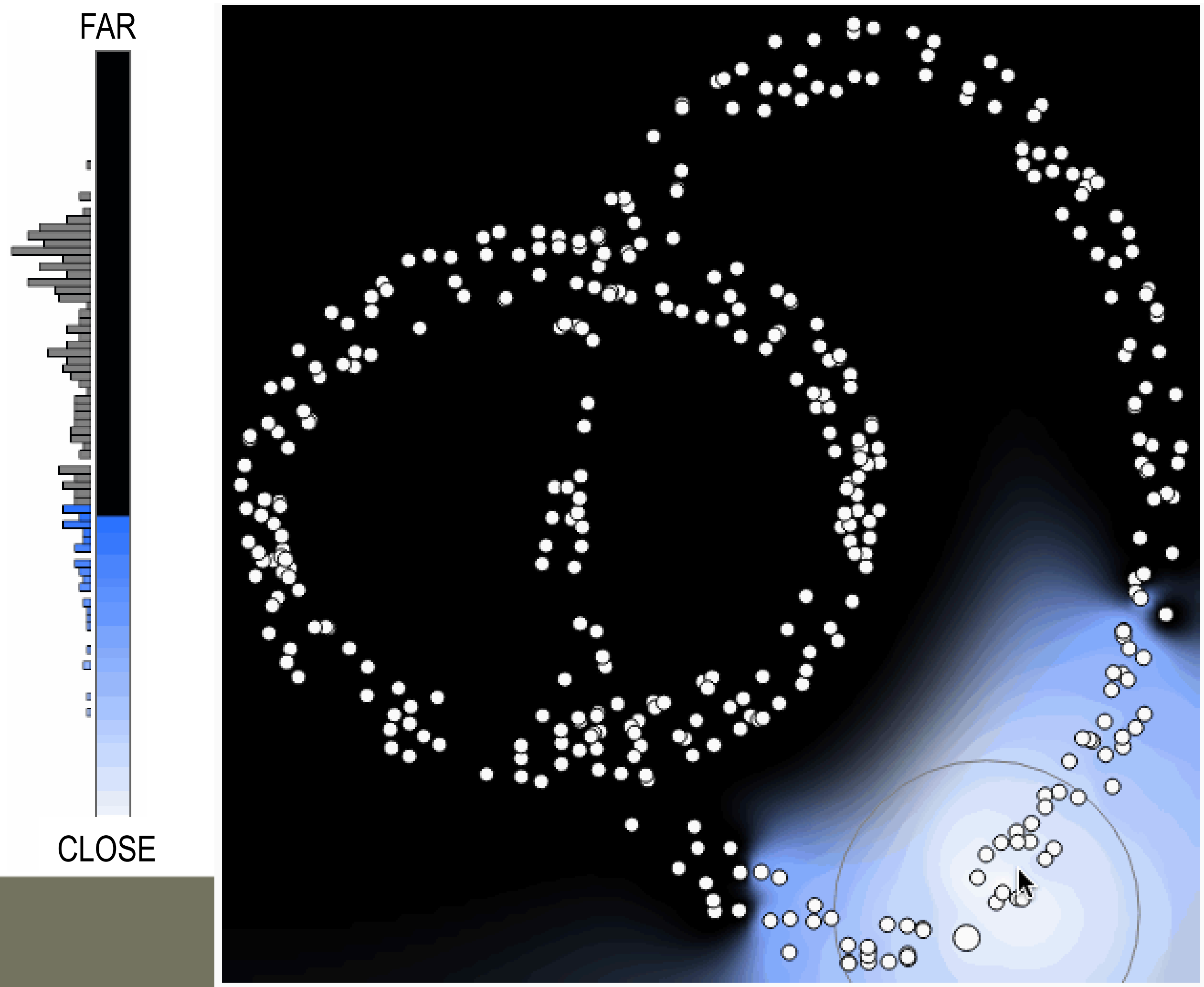


Turning weaknesses into strengths

- Displaying **original similarities** **within a lens** **to navigate**

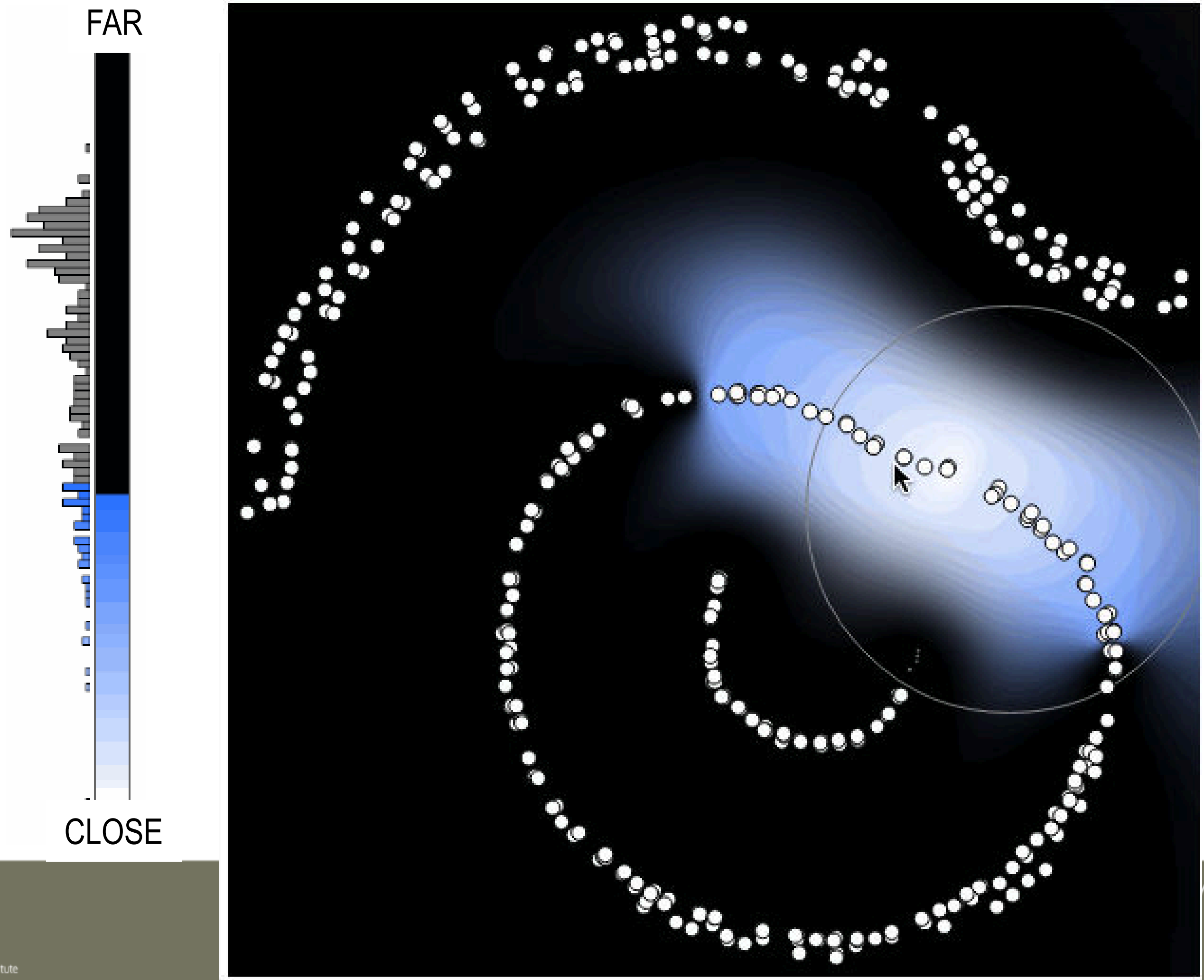
Turning weaknesses into strengths

- Displaying **original similarities** **within a lens** **to navigate**



Turning weaknesses into strengths

- Displaying **original similarities** **within a lens** **to navigate**



Turning weaknesses into strengths

- Displaying **original similarities** **within a lens** **to brush MD data**

Turning weaknesses into strengths

- Displaying **original similarities within a lens to brush HD data**

A multidimensional brush
for scatterplot data analytics

Michaël Aupetit – QCRI, Doha
Nicolas Heulot – IRT SystemX, Université Paris-Saclay
Jean-Daniel Fekete – INRIA AVIZ, Université Paris-Saclay

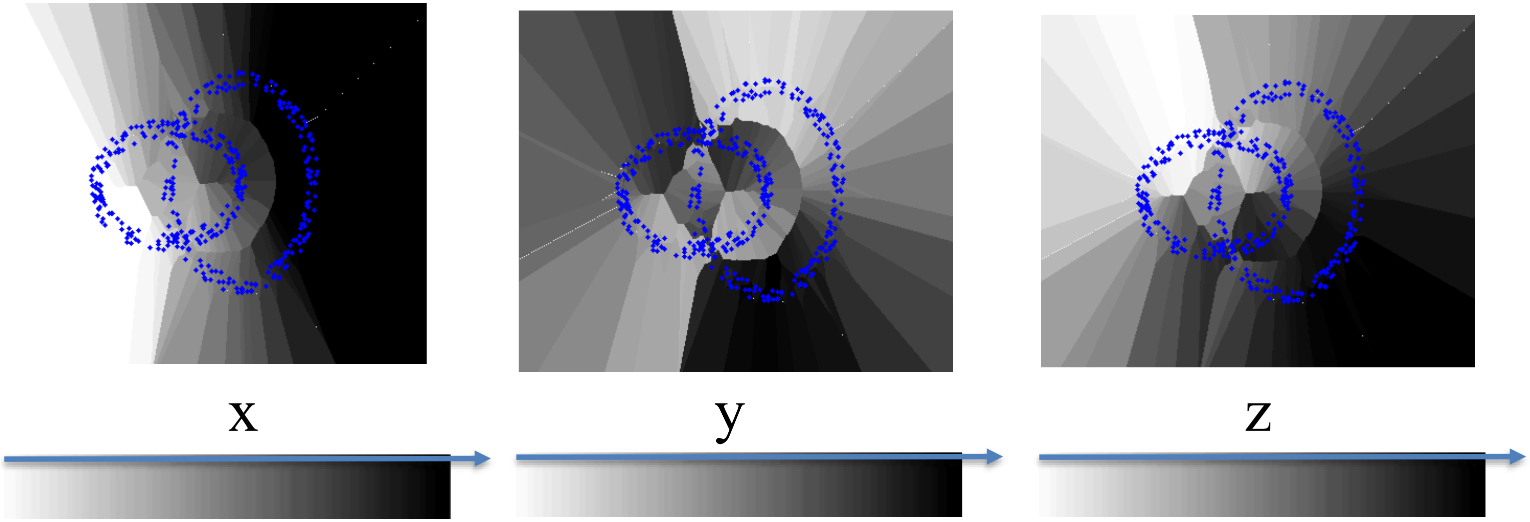
Contact: michael.aupetit@gmail.com

Component Plane Xiao, Wang, Teng, Zhang, Component plane presentation integrated self-organizing map for microarray data analysis, FEBS Letters, 538(1–3): 117-124, 2003

Turning weaknesses into strengths

- Displaying **original variable within the map itself**
- Can detect **correlations between variables**

PCA

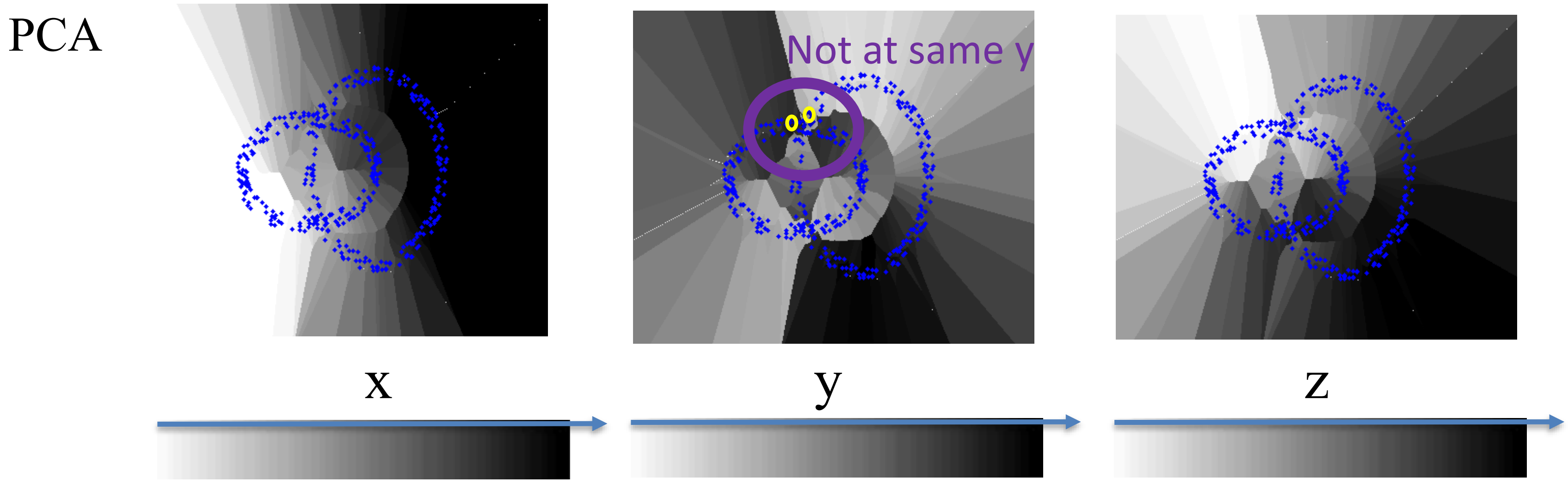


Component Plane

Xiao, Wang, Teng, Zhang, Component plane presentation integrated self-organizing map for microarray data analysis, FEBS Letters, 538(1–3): 117-124, 2003

Turning weaknesses into strengths

- Displaying **original variable within the map itself**
- Can detect **correlations between variables**

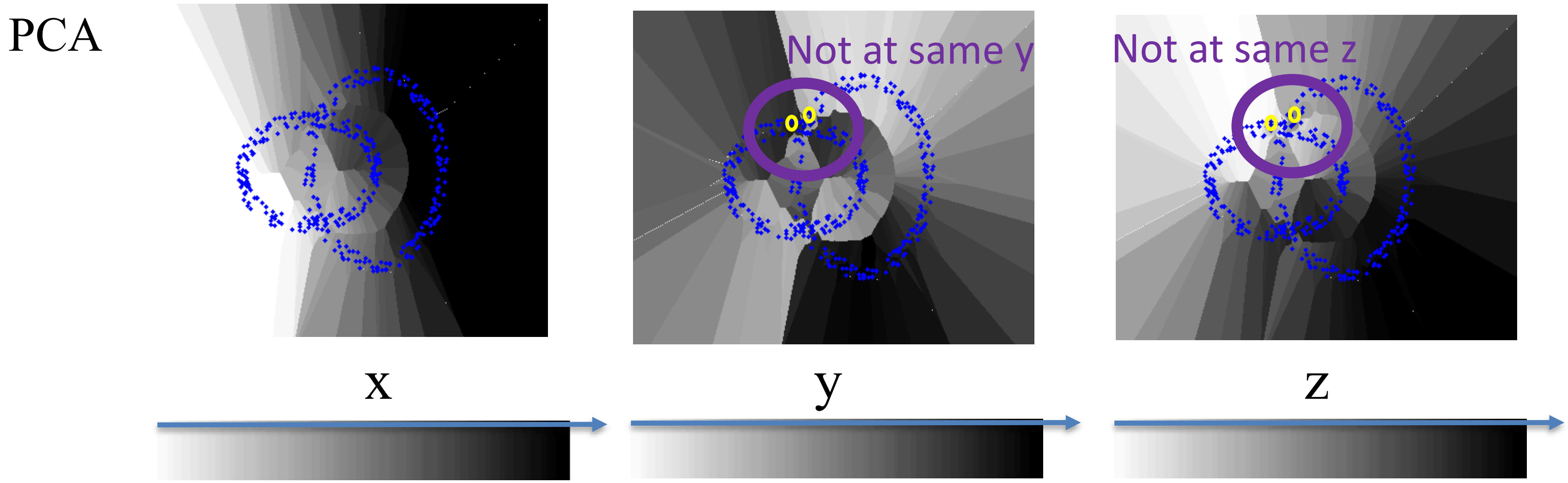


Component Plane

Xiao, Wang, Teng, Zhang, Component plane presentation integrated self-organizing map for microarray data analysis, FEBS Letters, 538(1–3): 117-124, 2003

Turning weaknesses into strengths

- Displaying **original variable within the map itself**
- Can detect **correlations between variables**



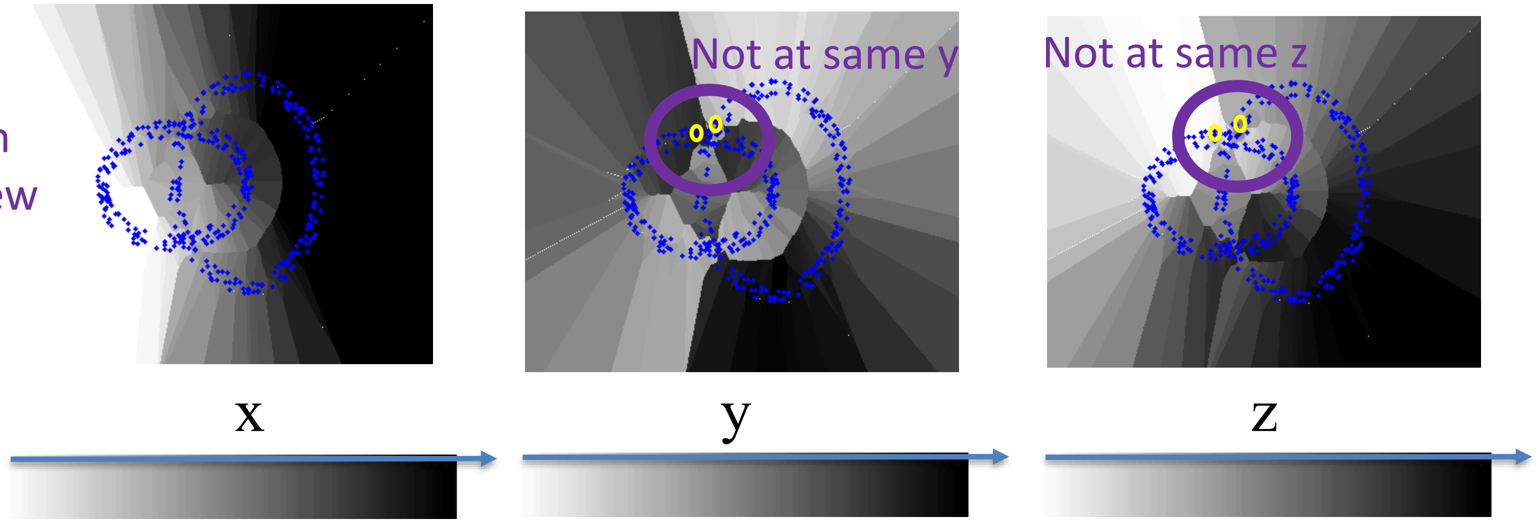
Component Plane Xiao, Wang, Teng, Zhang, Component plane presentation integrated self-organizing map for microarray data analysis, FEBS Letters, 538(1–3): 117-124, 2003

Turning weaknesses into strengths

- Displaying **original variable within the map itself**
- Can detect **correlations between variables**

PCA

Sure False even if in a single view (OR)



Component Plane Xiao, Wang, Teng, Zhang, Component plane presentation integrated self-organizing map for microarray data analysis, FEBS Letters, 538(1–3): 117-124, 2003

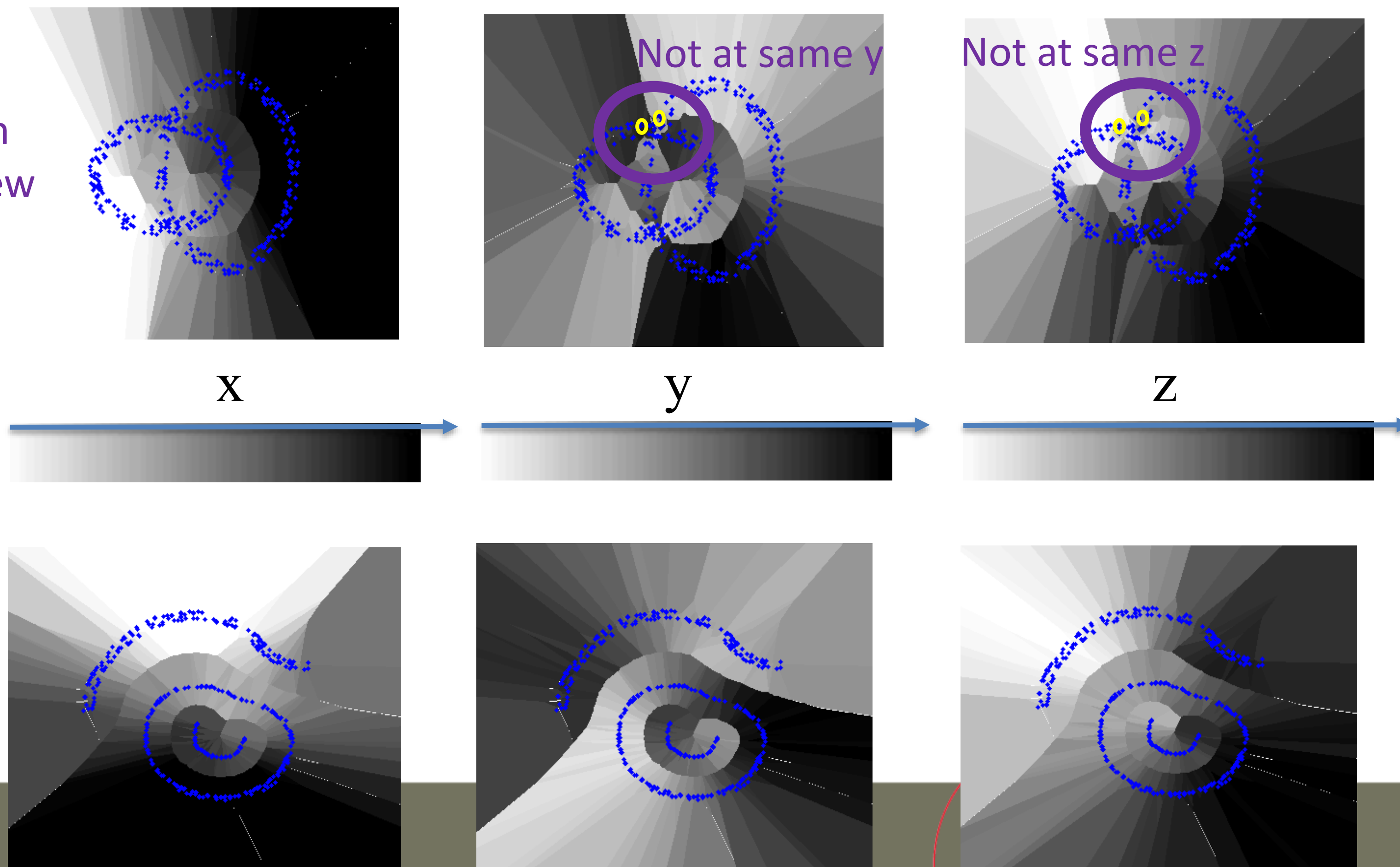
Turning weaknesses into strengths

- Displaying **original variable within the map itself**
- Can detect **correlations between variables**

PCA

Sure False even if in a single view (OR)

CCA



Component Plane

Xiao, Wang, Teng, Zhang, Component plane presentation integrated self-organizing map for microarray data analysis, FEBS Letters, 538(1–3): 117-124, 2003

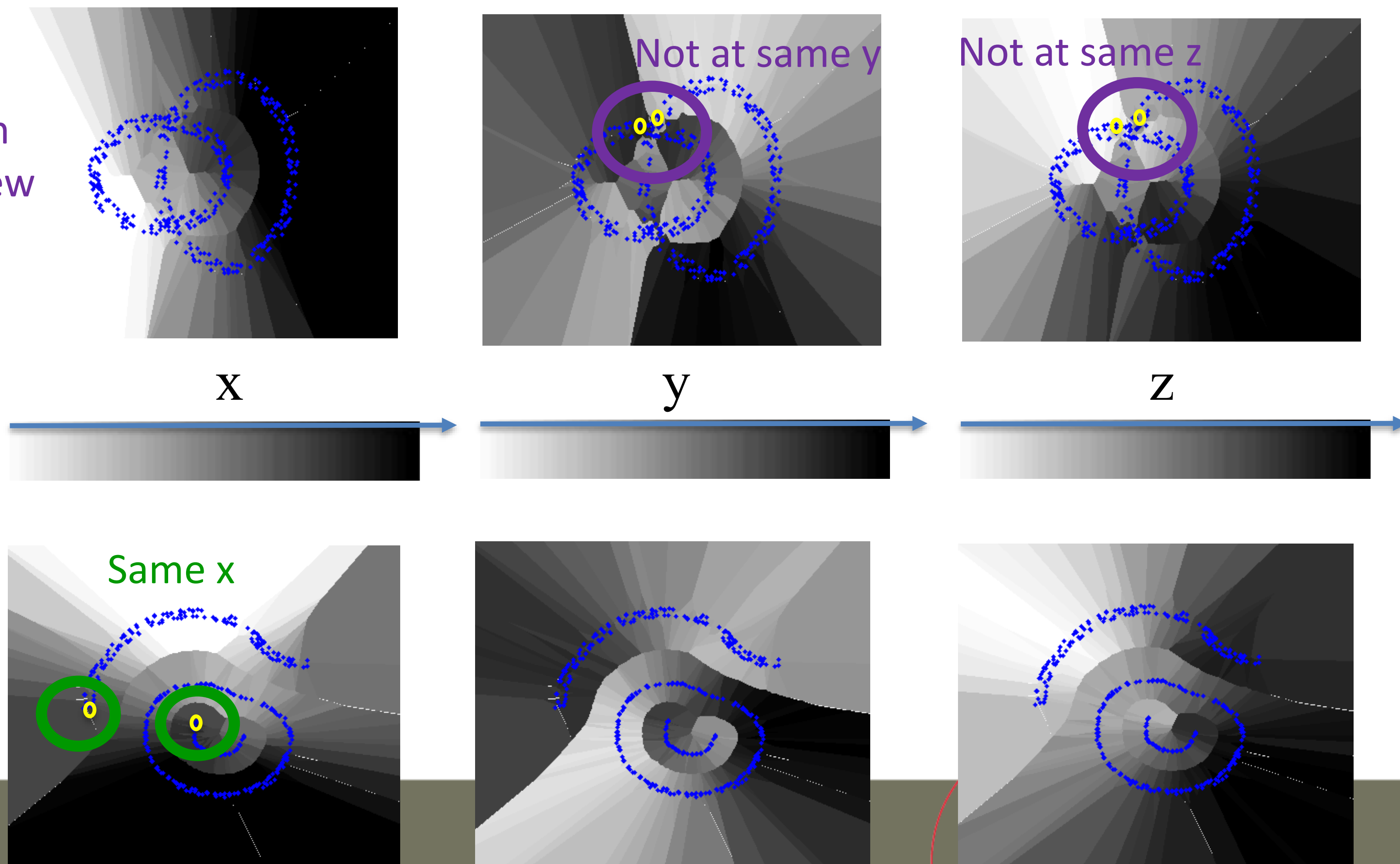
Turning weaknesses into strengths

- Displaying **original variable within the map itself**
- Can detect **correlations between variables**

PCA

Sure False even if in a single view (OR)

CCA



Component Plane

Xiao, Wang, Teng, Zhang, Component plane presentation integrated self-organizing map for microarray data analysis, FEBS Letters, 538(1–3): 117-124, 2003

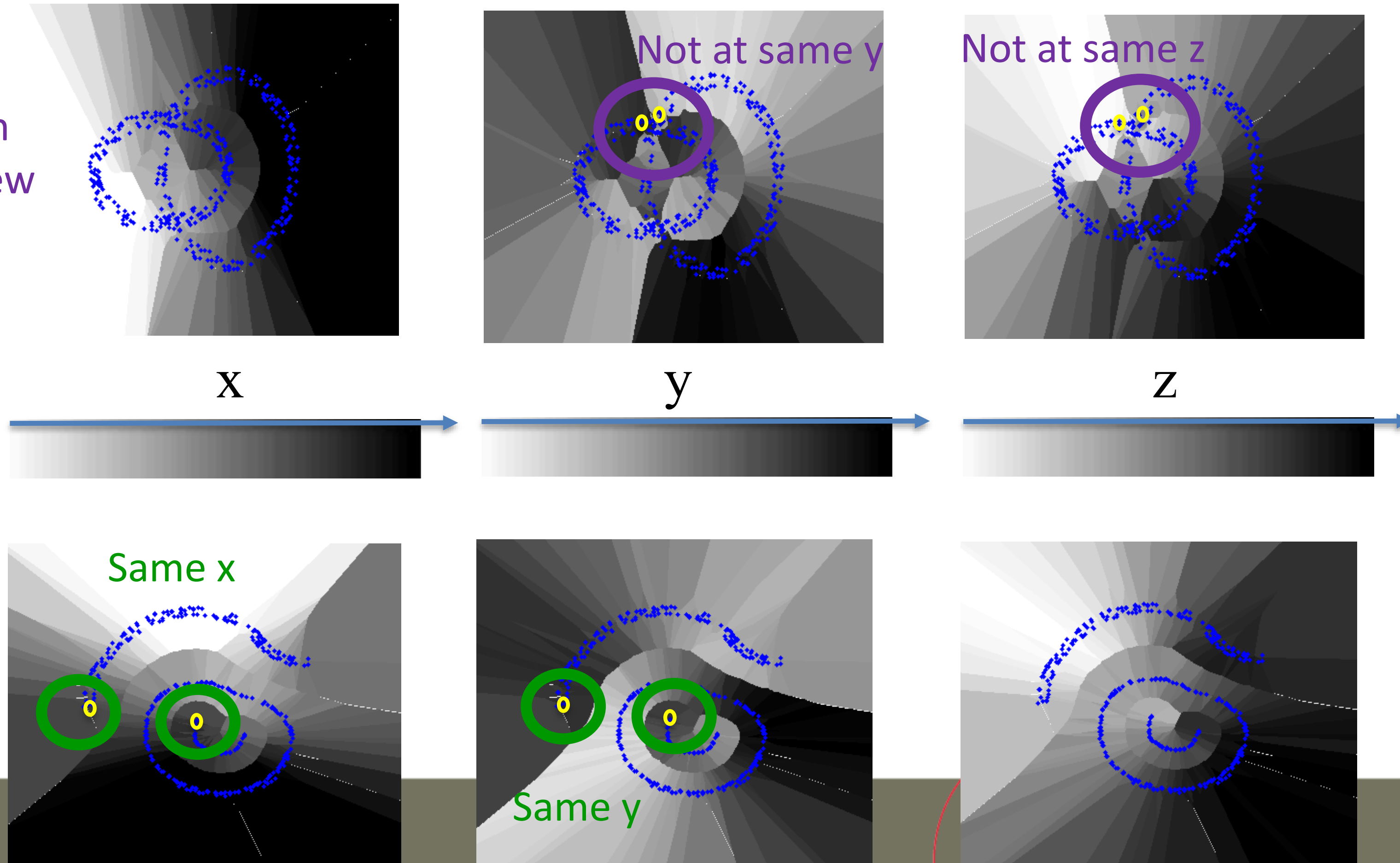
Turning weaknesses into strengths

- Displaying **original variable within the map itself**
- Can detect **correlations between variables**

PCA

Sure False even if in a single view (OR)

CCA



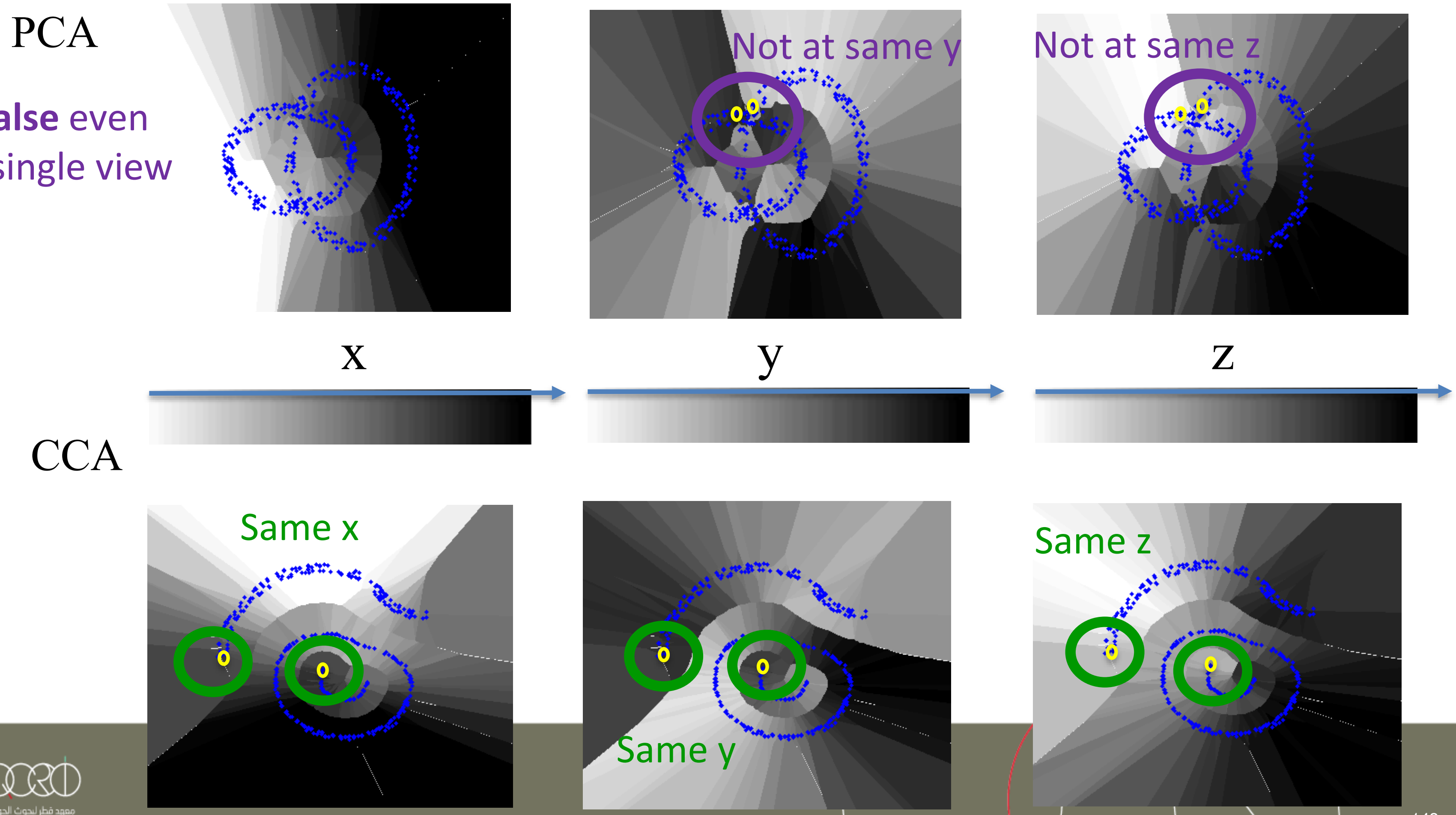
Component Plane

Xiao, Wang, Teng, Zhang, Component plane presentation integrated self-organizing map for microarray data analysis, FEBS Letters, 538(1–3): 117-124, 2003

Turning weaknesses into strengths

- Displaying **original variable within the map itself**
- Can detect **correlations between variables**

Sure False even if in a single view (OR)



Component Plane Xiao, Wang, Teng, Zhang, Component plane presentation integrated self-organizing map for microarray data analysis, FEBS Letters, 538(1–3): 117-124, 2003

Turning weaknesses into strengths

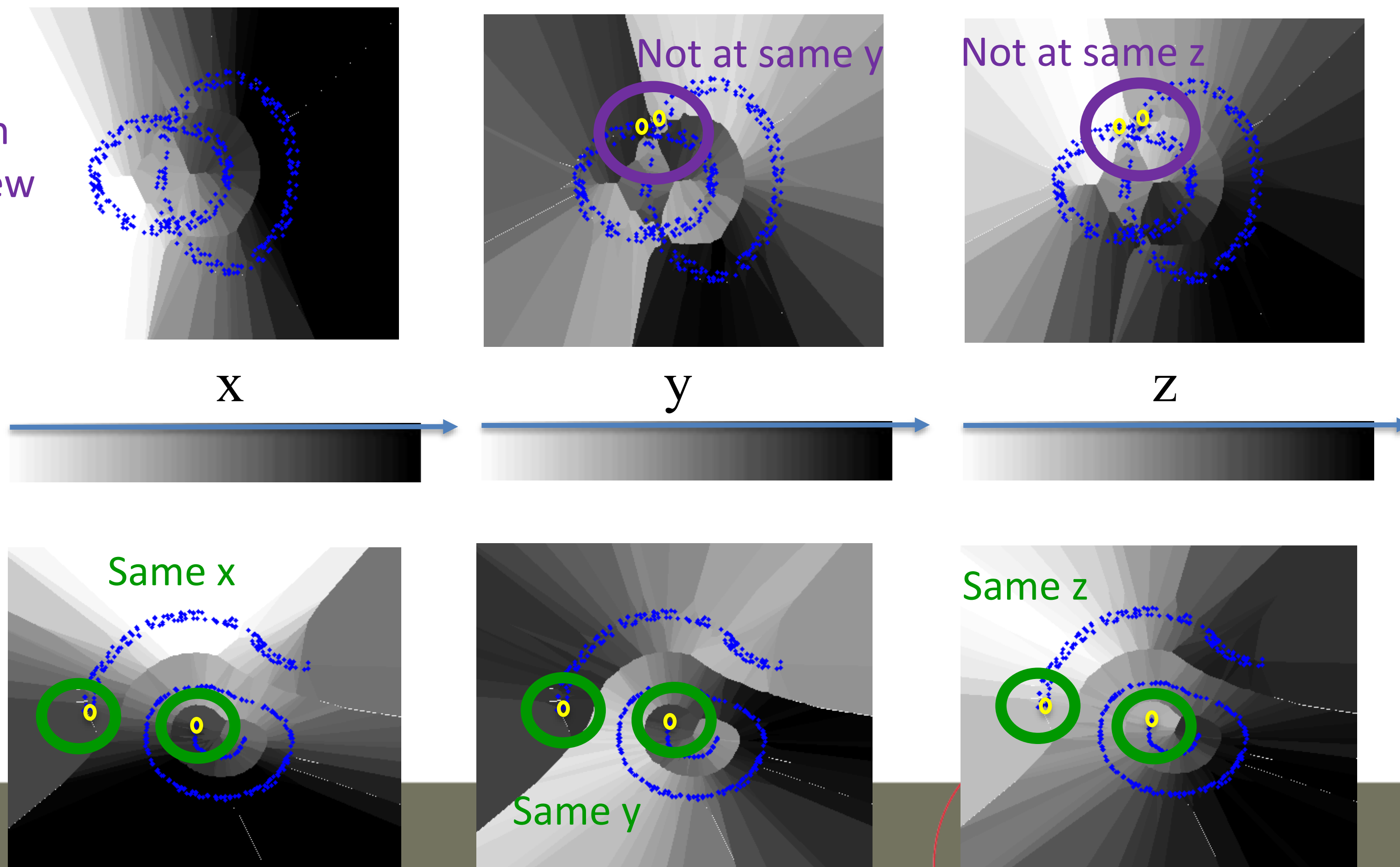
- Displaying **original variable within the map itself**
- Can detect **correlations between variables**

PCA

Sure False even
if in a single view
(OR)

CCA

Sure Miss
only if
in all 3 views
at once!
(AND)

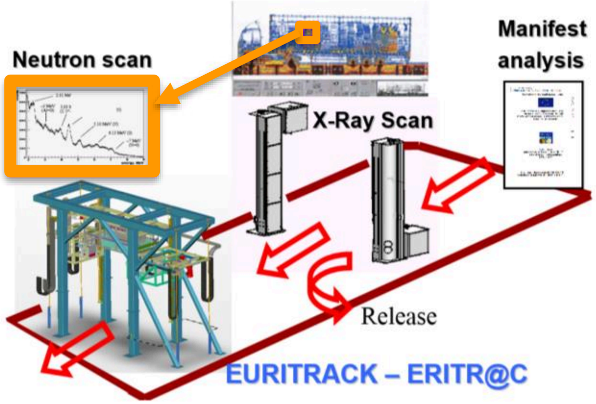


Turning weaknesses into strengths

- Displaying **out-of-sample instance** **within the map itself**

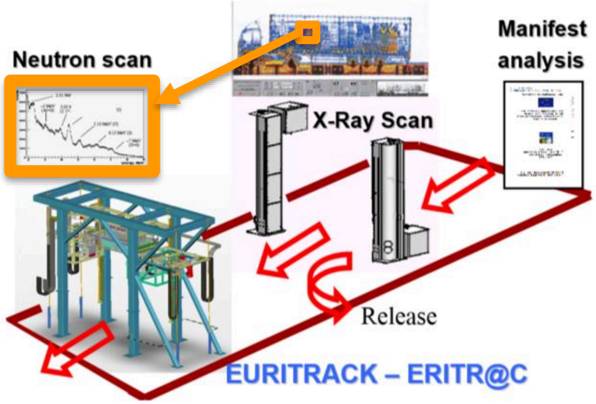
Turning weaknesses into strengths

- Displaying **out-of-sample instance** **within the map itself**



Turning weaknesses into strengths

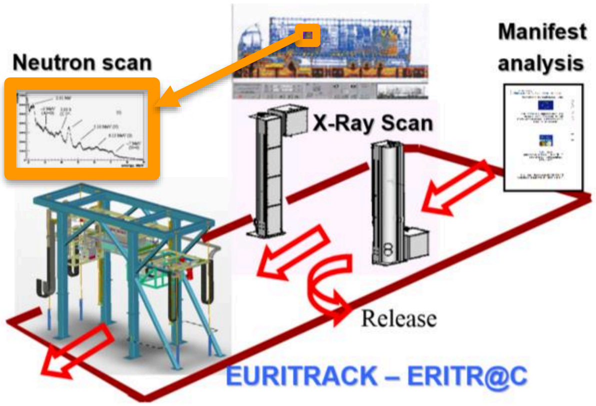
- Displaying **out-of-sample instance** **within the map itself**



40 reference materials

Turning weaknesses into strengths

- Displaying **out-of-sample instance** **within the map itself**



40 reference materials

Wood furniture

Textile

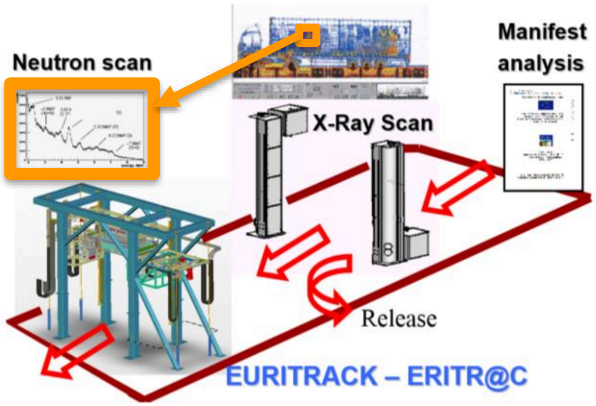
Metal box

Plastic boxes

…

Turning weaknesses into strengths

- Displaying **out-of-sample instance within the map itself**



40 reference materials

Wood furniture

Textile

Metal box

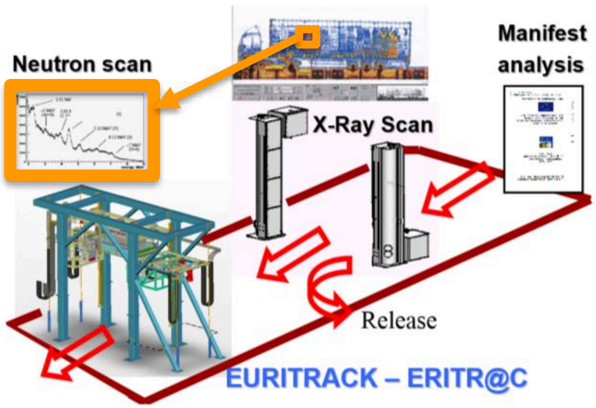
Plastic boxes

…

15D chemical signature

Turning weaknesses into strengths

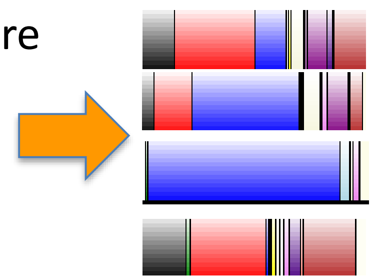
- Displaying **out-of-sample instance** **within the map itself**



40 reference materials

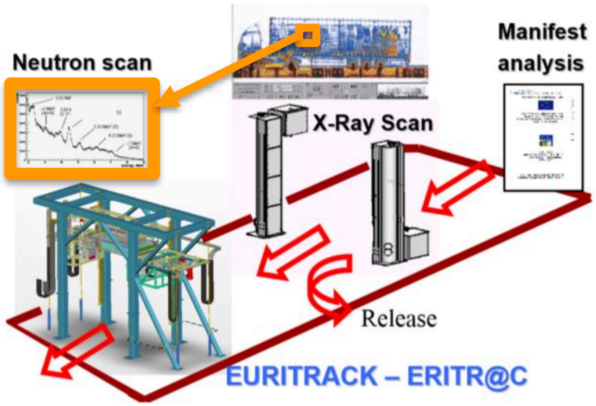
Wood furniture
Textile
Metal box
Plastic boxes
...

15D chemical signature



Turning weaknesses into strengths

- Displaying **out-of-sample instance** **within the map itself**



40 reference materials

Wood furniture

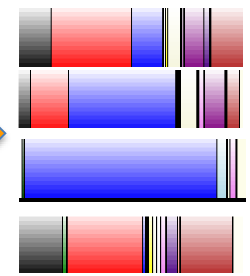
Textile

Metal box

Plastic boxes

…

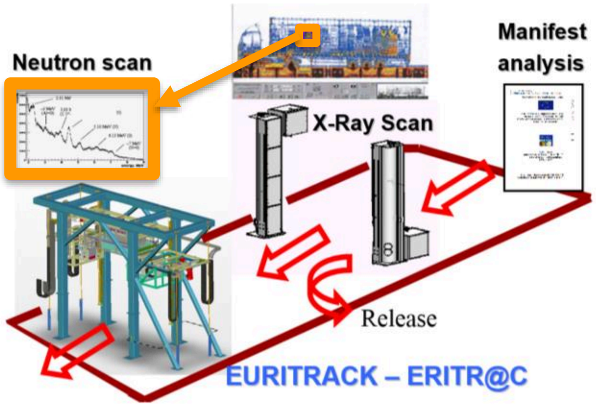
15D chemical signature



Reference Map

Turning weaknesses into strengths

- Displaying **out-of-sample instance** **within the map itself**



40 reference materials

Wood furniture

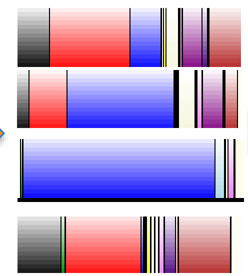
Textile

Metal box

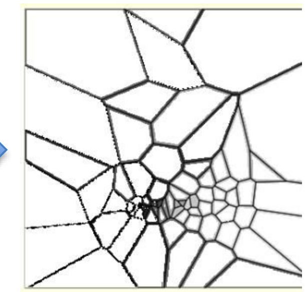
Plastic boxes

…

15D chemical signature

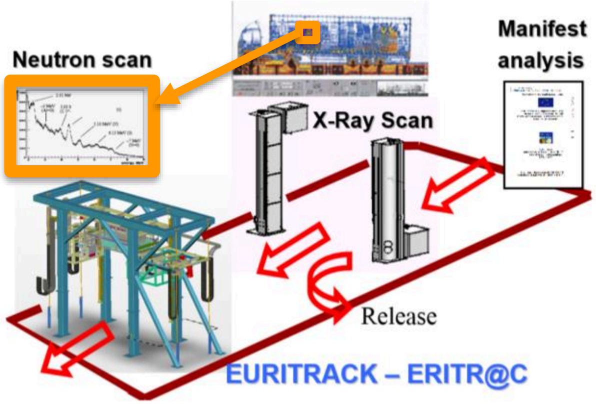


Reference Map



Turning weaknesses into strengths

- Displaying **out-of-sample instance** **within the map itself**



40 reference materials

Wood furniture

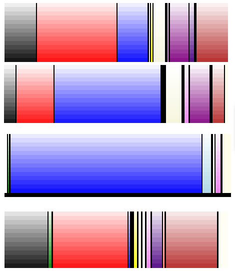
Textile

Metal box

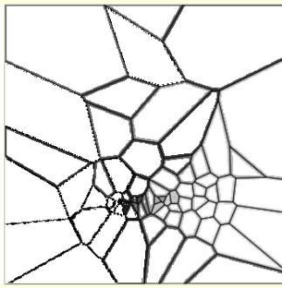
Plastic boxes

…

15D chemical signature



Reference Map

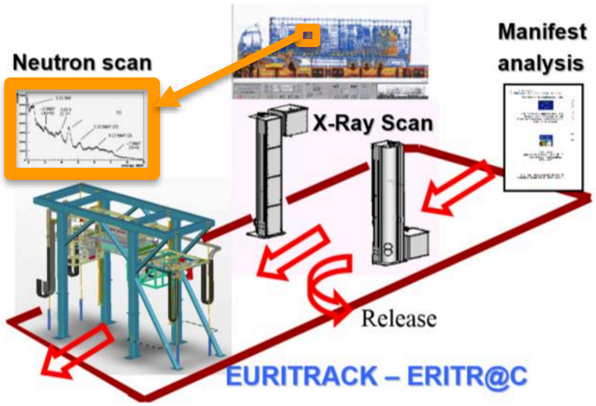


New content

Manifest: "paper"

Turning weaknesses into strengths

- Displaying **out-of-sample instance** **within the map itself**



40 reference materials

Wood furniture

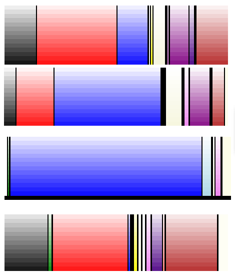
Textile

Metal box

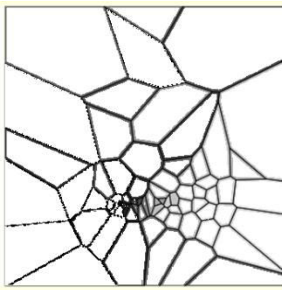
Plastic boxes

...

15D chemical signature



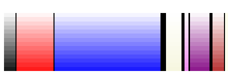
Reference Map



New content

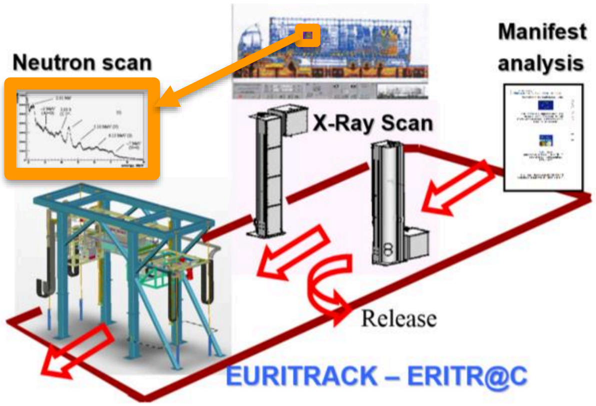
Manifest: "paper"

15D chemical signature



Turning weaknesses into strengths

- Displaying **out-of-sample instance within the map itself**



40 reference materials

Wood furniture

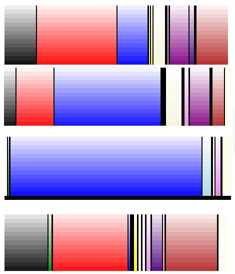
Textile

Metal box

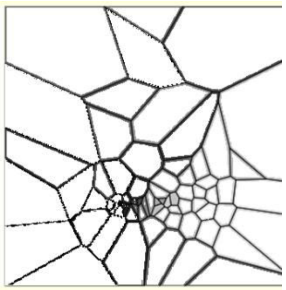
Plastic boxes

…

15D chemical signature



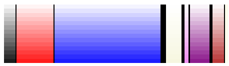
Reference Map



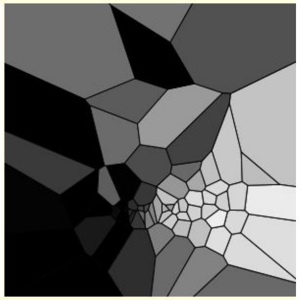
New content

Manifest: "paper"

15D chemical signature



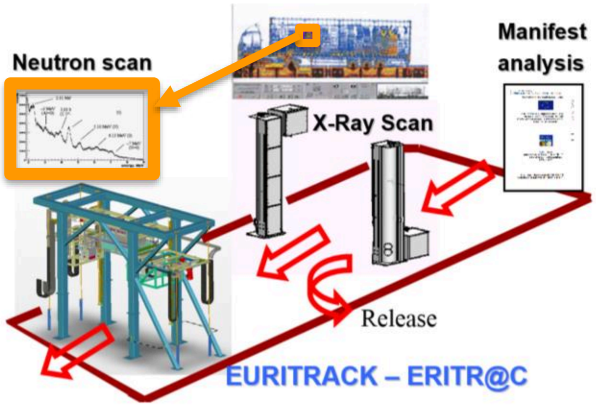
ClassiViz



Similarity to reference is color-coded

Turning weaknesses into strengths

- Displaying **out-of-sample instance** **within the map itself**



40 reference materials

Wood furniture

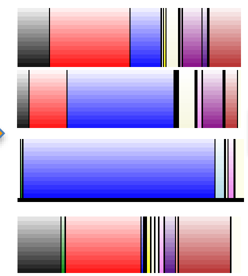
Textile

Metal box

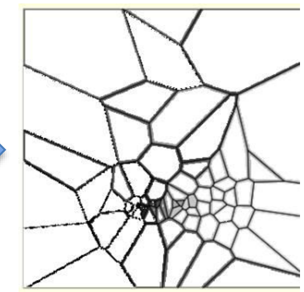
Plastic boxes

…

15D chemical signature



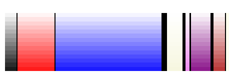
Reference Map



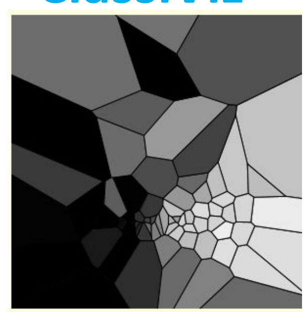
New content

Manifest: "paper"

15D chemical signature



ClassiViz

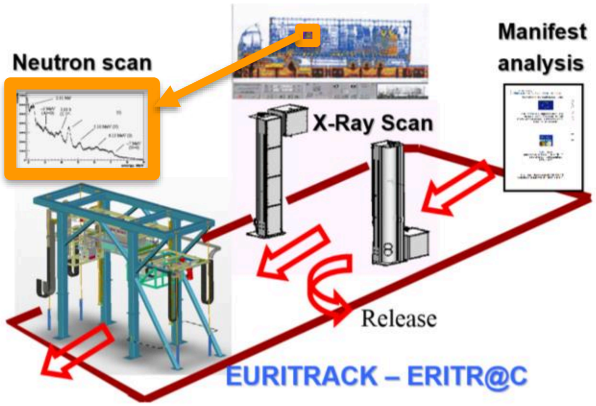


Similarity to reference is color-coded

*New content **is not added** to the map (where to put it???)*

Turning weaknesses into strengths

- Displaying **out-of-sample instance** **within the map itself**



40 reference materials

Wood furniture

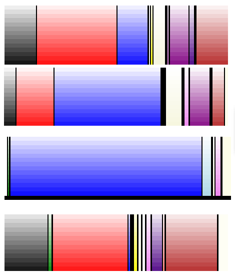
Textile

Metal box

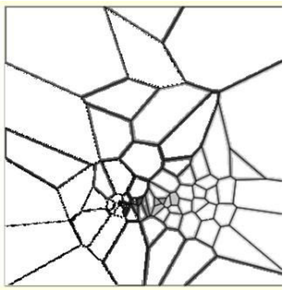
Plastic boxes

…

15D chemical signature



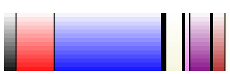
Reference Map



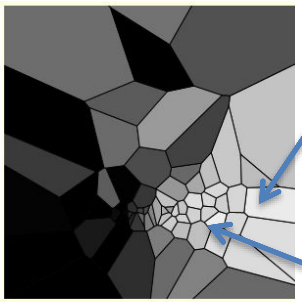
New content

Manifest: "paper"

15D chemical signature



ClassiViz



wood

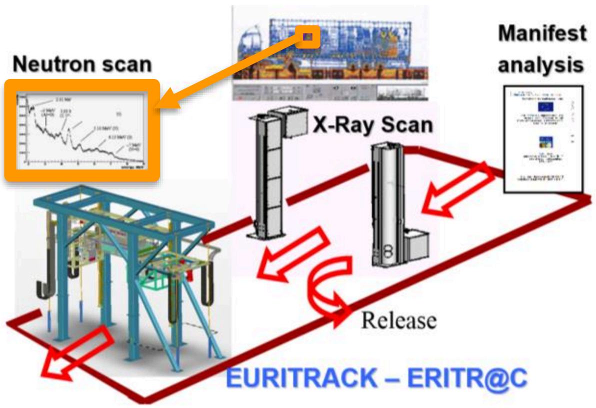
cardboard

Similarity to reference is color-coded

New content ***is not added*** *to the map (where to put it???)*

Turning weaknesses into strengths

- Displaying **out-of-sample instance** **within the map itself**



40 reference materials

Wood furniture

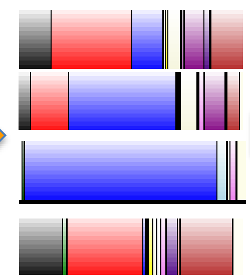
Textile

Metal box

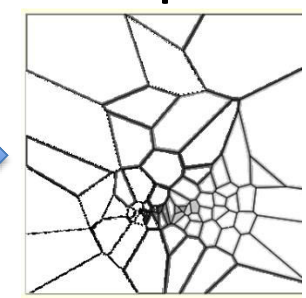
Plastic boxes

…

15D chemical signature



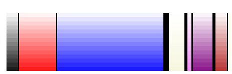
Reference Map



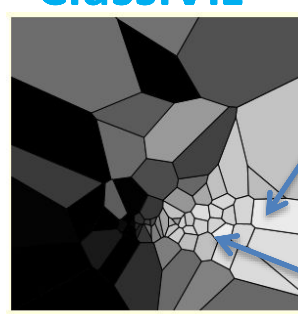
New content

Manifest: "paper"

15D chemical signature



ClassiViz



wood

cardboard

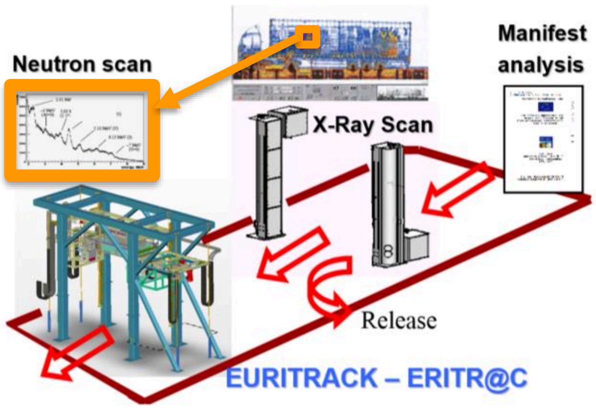
Chemical coherent with manifest

Similarity to reference is color-coded

New content ***is not added*** *to the map (where to put it???)*

Turning weaknesses into strengths

- Displaying **out-of-sample instance within the map itself**



40 reference materials

Wood furniture

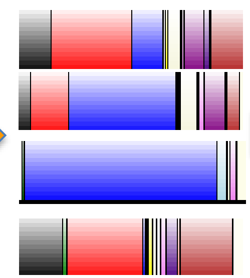
Textile

Metal box

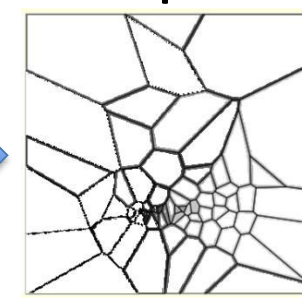
Plastic boxes

...

15D chemical signature



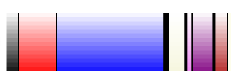
Reference Map



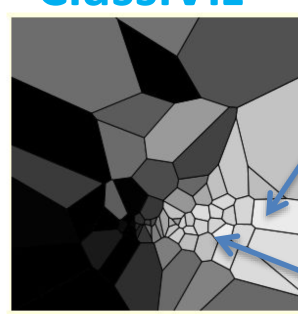
New content

Manifest: "paper"

15D chemical signature



ClassiViz



wood

cardboard

Similarity to reference is color-coded

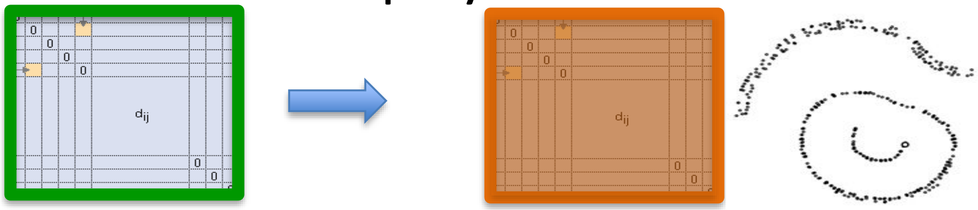
Chemical coherent with manifest

Custom does not open for deep checking

New content ***is not added*** *to the map (where to put it???)*

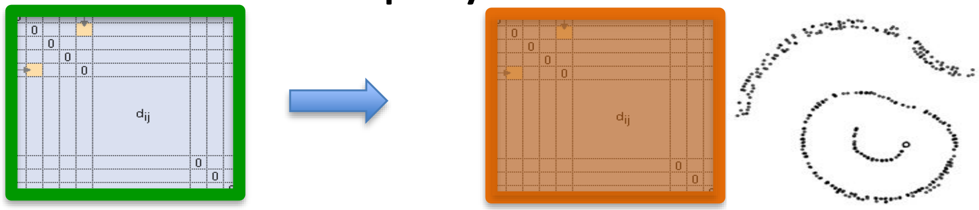
Summary

Display distortions



Summary

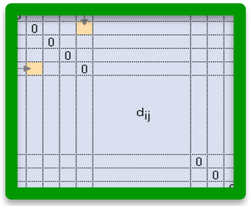
Display **distortions**

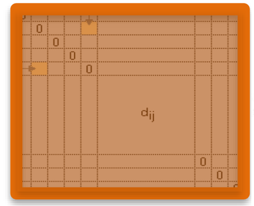


78%

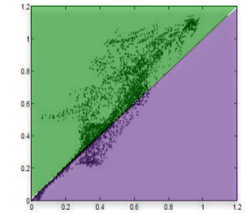
Summary

Display **distortions**



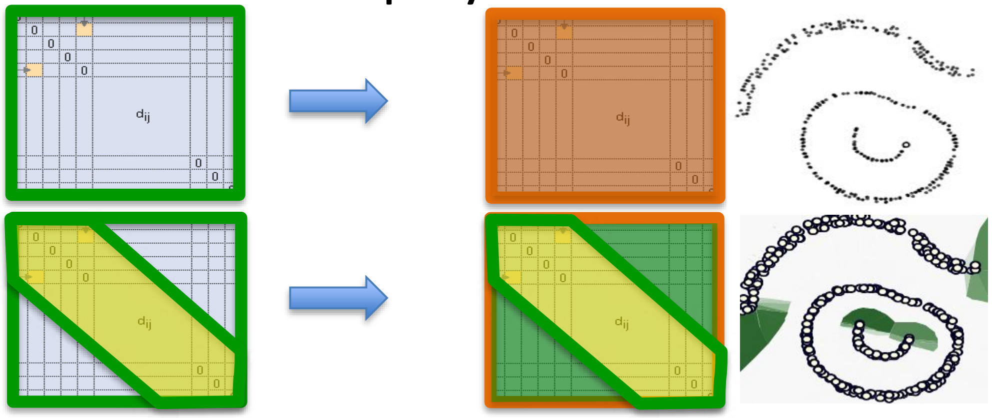


78%



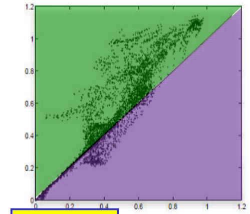
Summary

Display distortions



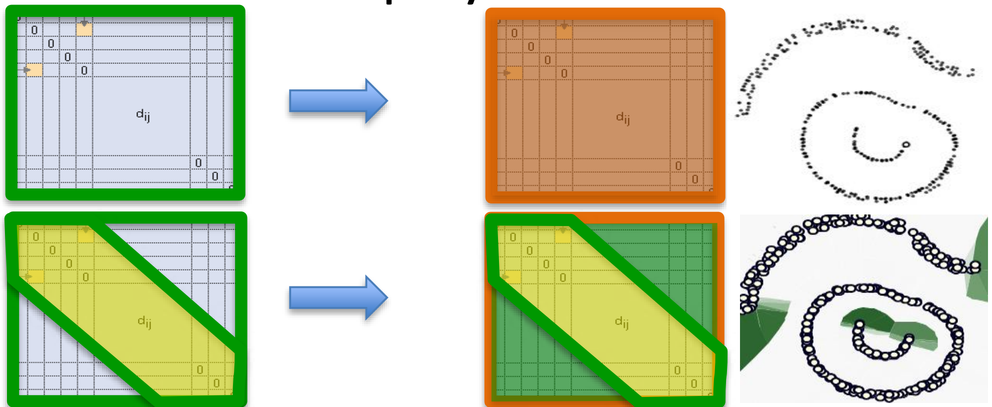
78%

CheckViz



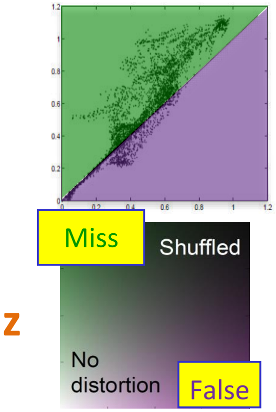
Summary

Display **distortions**



78%

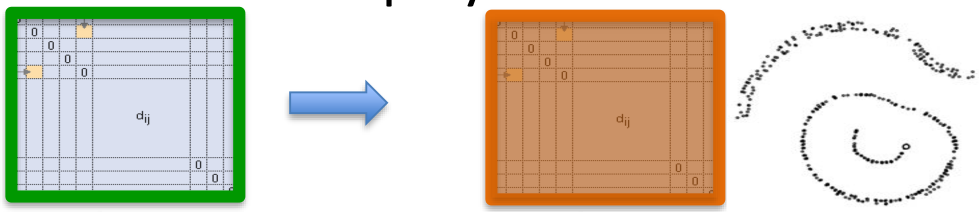
CheckViz



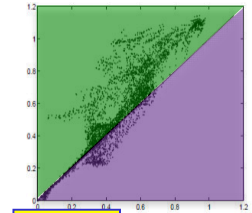
Display **original data**

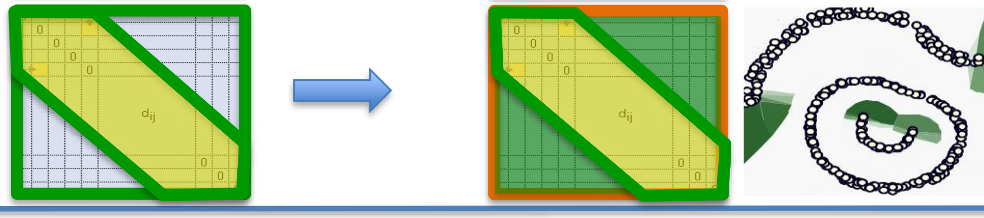
Summary

Display **distortions**

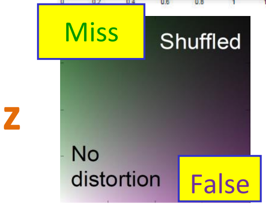


78%

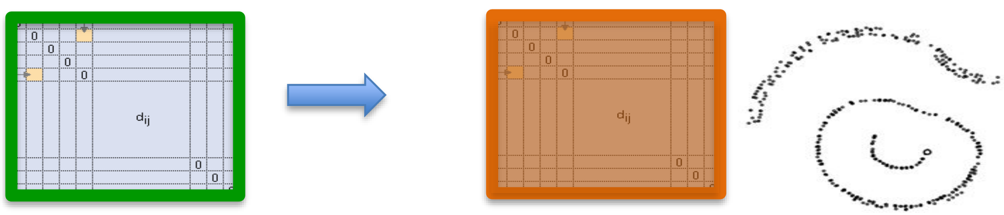




CheckViz

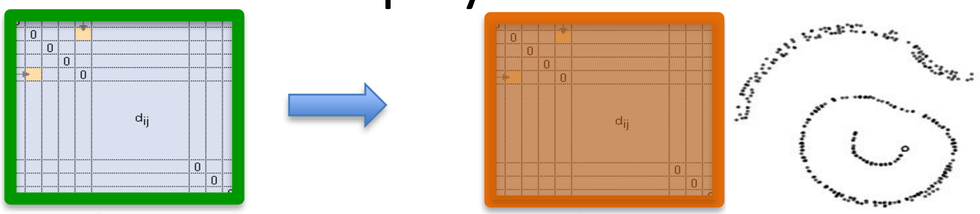


Display **original data**

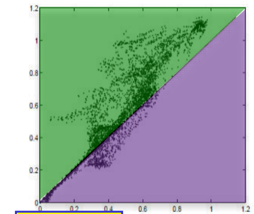


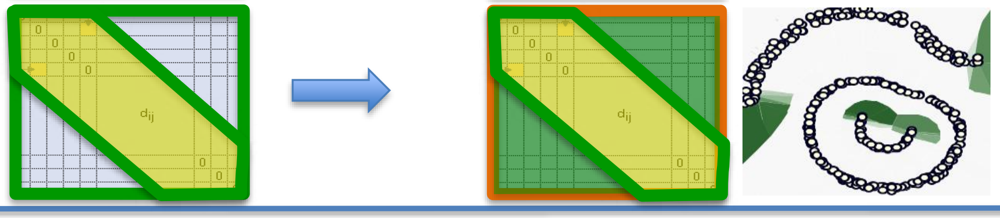
Summary

Display distortions



78%

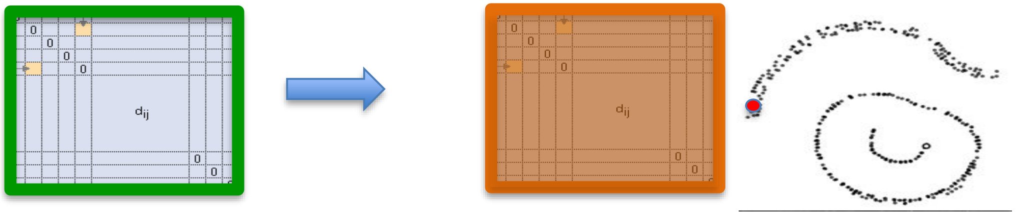




CheckViz

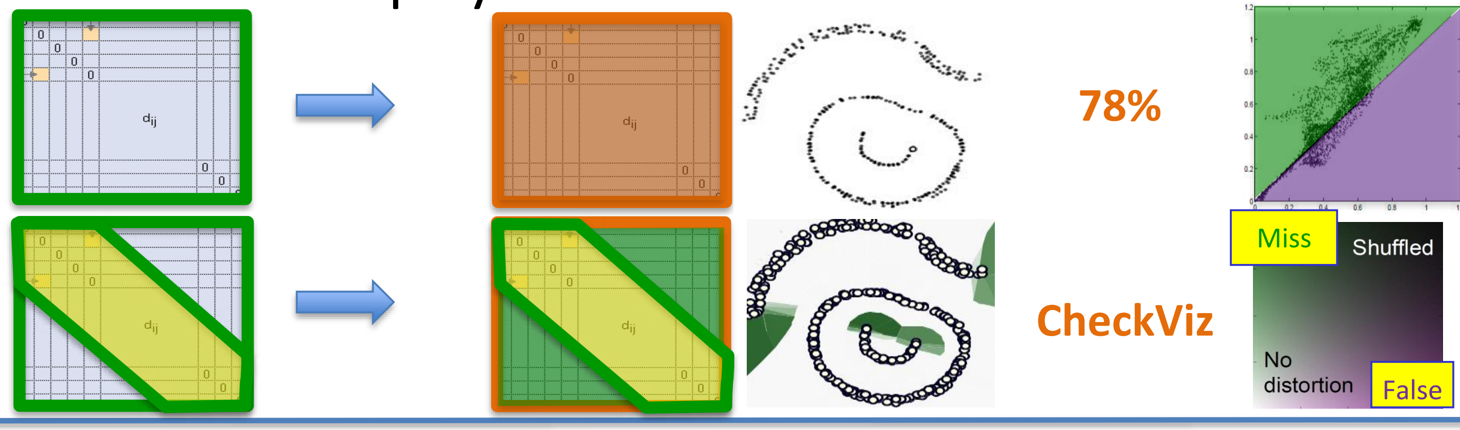
Display original data

Instance



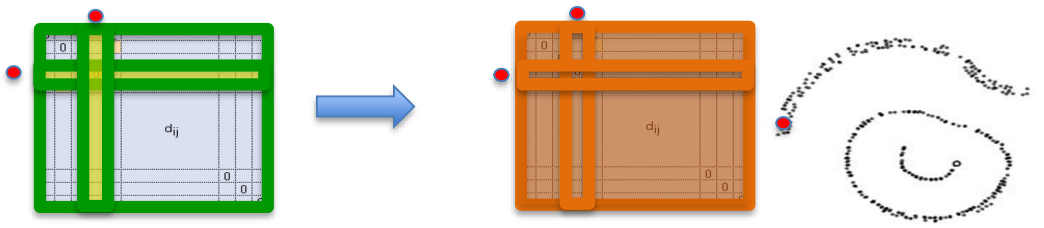
Summary

Display distortions



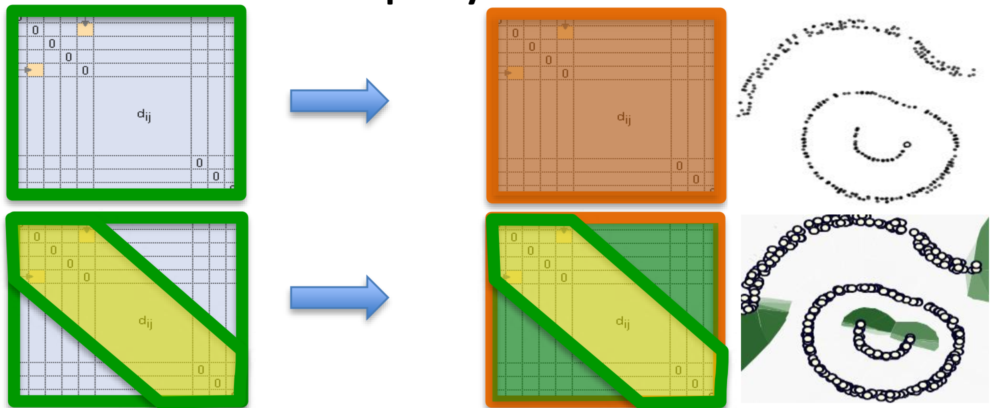
Display original data

Instance



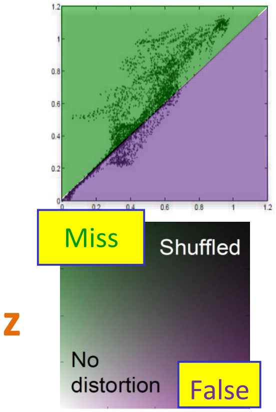
Summary

Display distortions



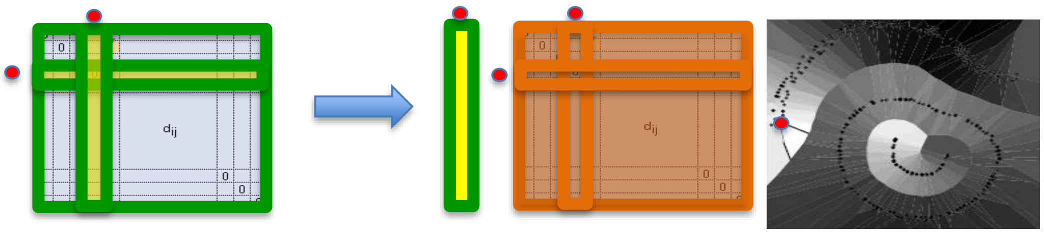
78%

CheckViz



Display original data

Instance

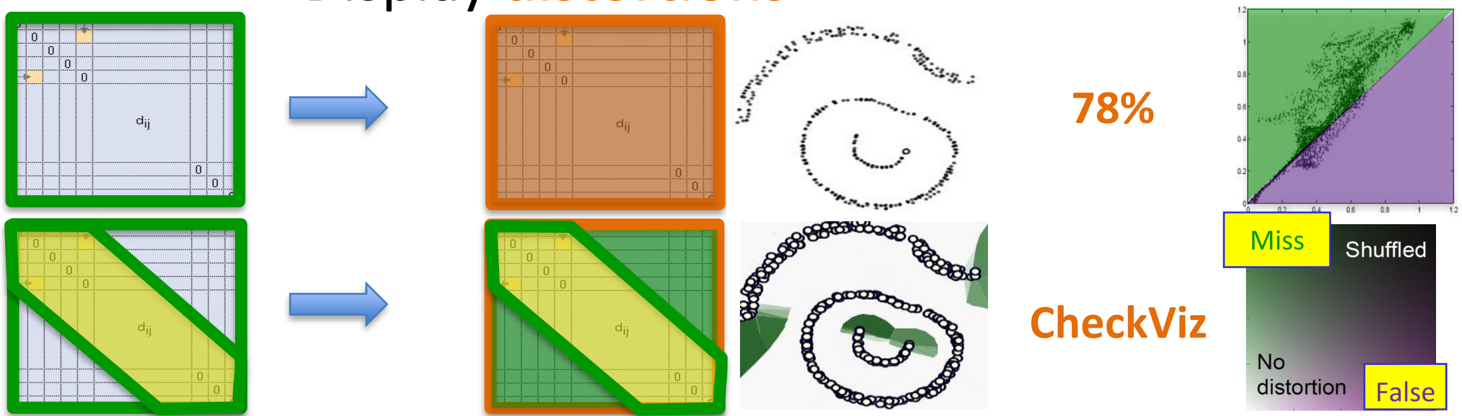


ProxiViz

ProxiLens

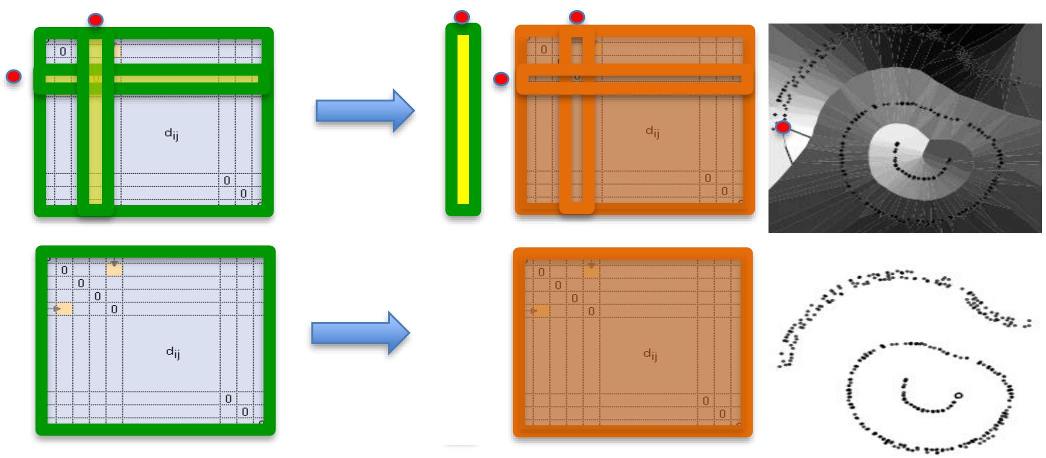
Summary

Display distortions



Display original data

Instance

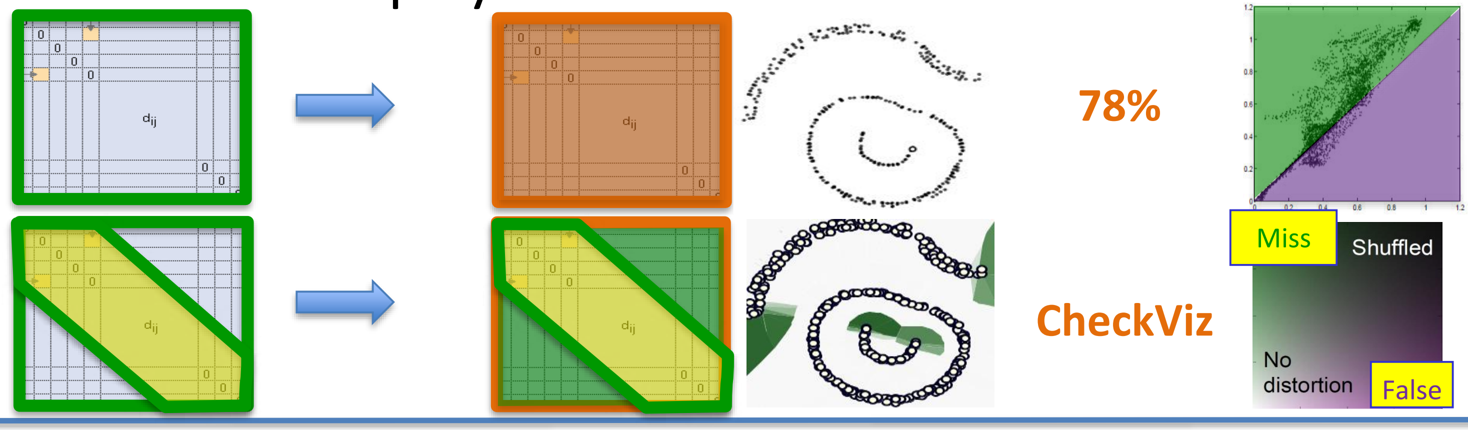


ProxiViz

ProxiLens

Summary

Display distortions

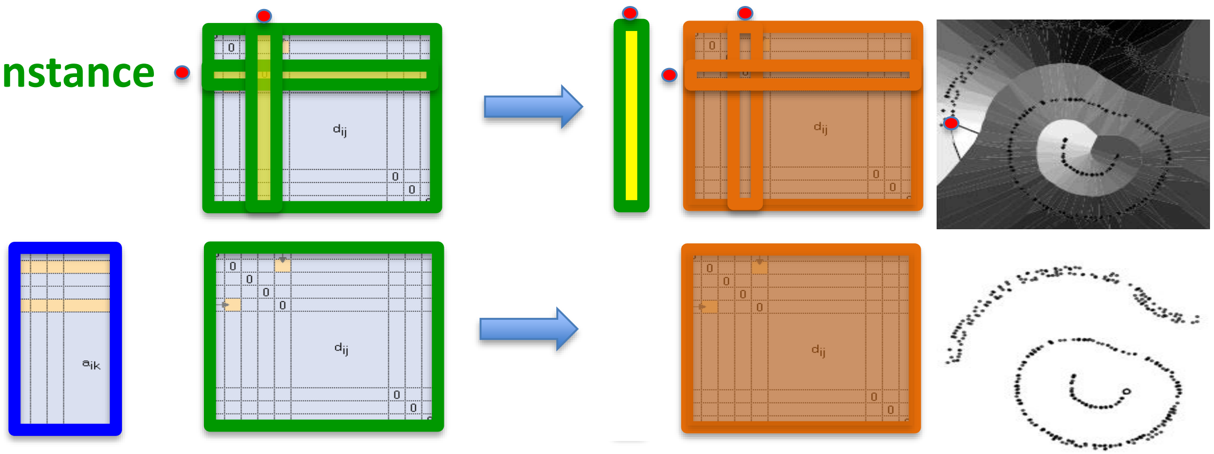


78%

CheckViz

Display original data

Instance

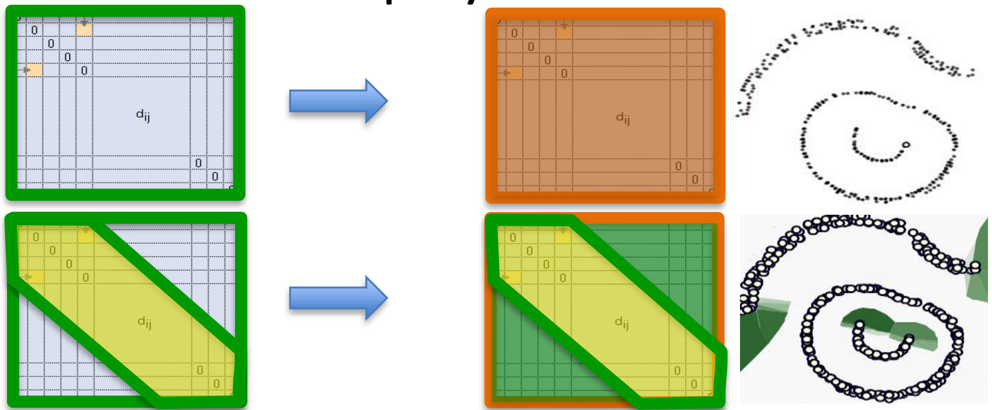


ProxiViz

ProxiLens

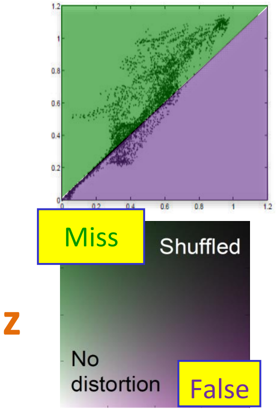
Summary

Display distortions



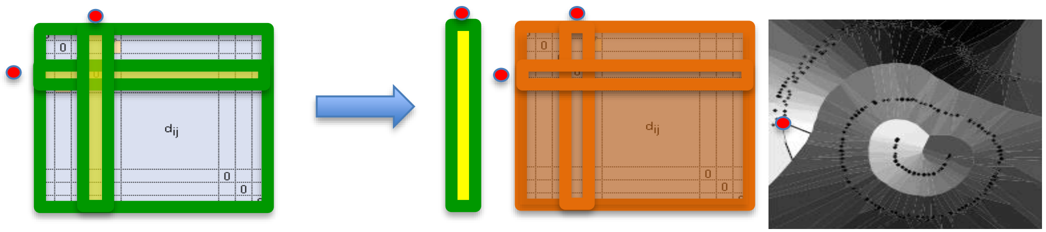
78%

CheckViz



Display original data

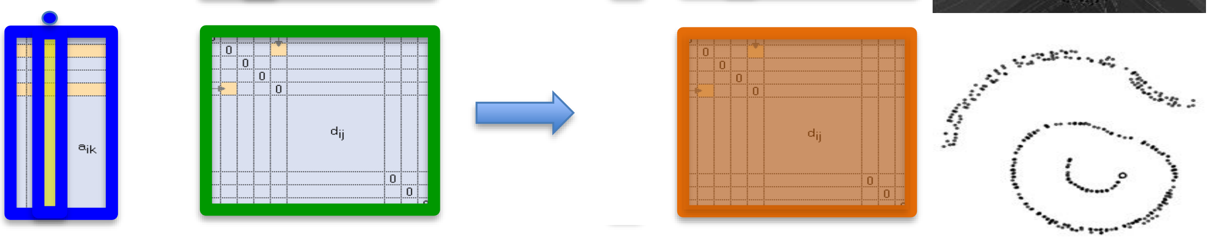
Instance



ProxiViz

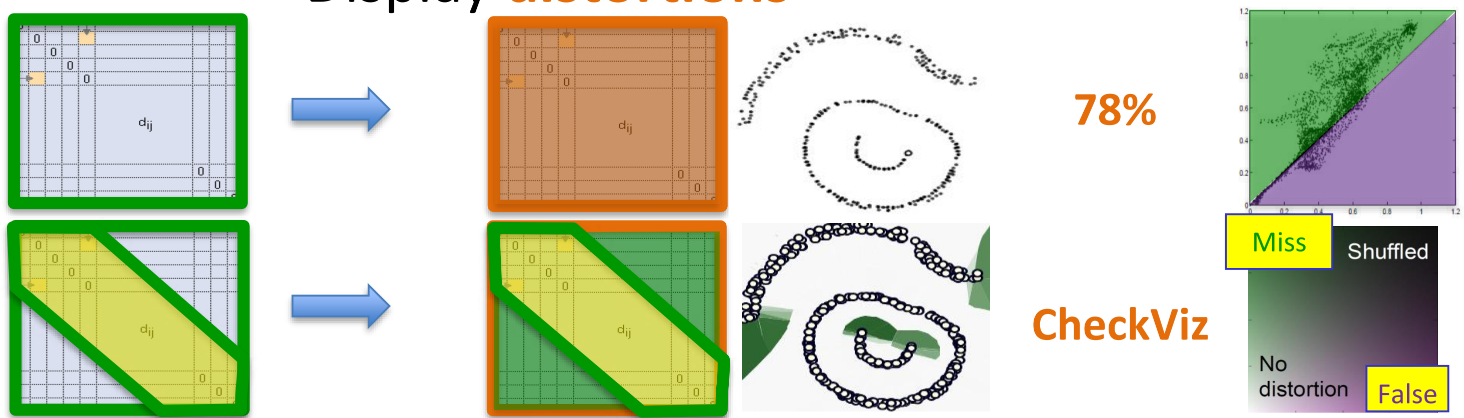
ProxiLens

Variable

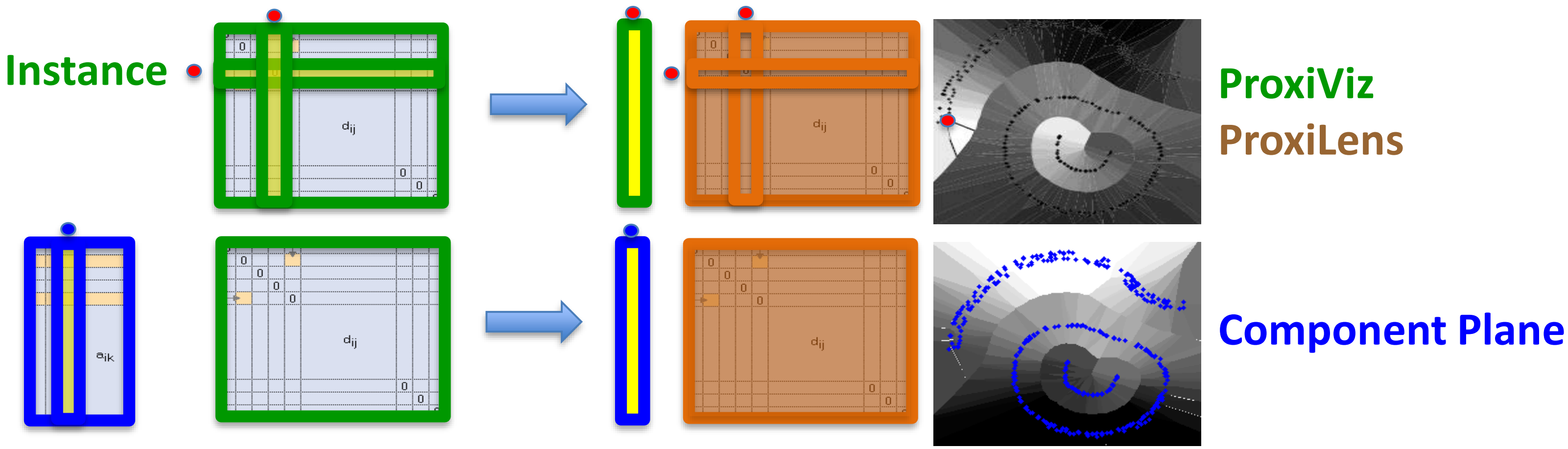


Summary

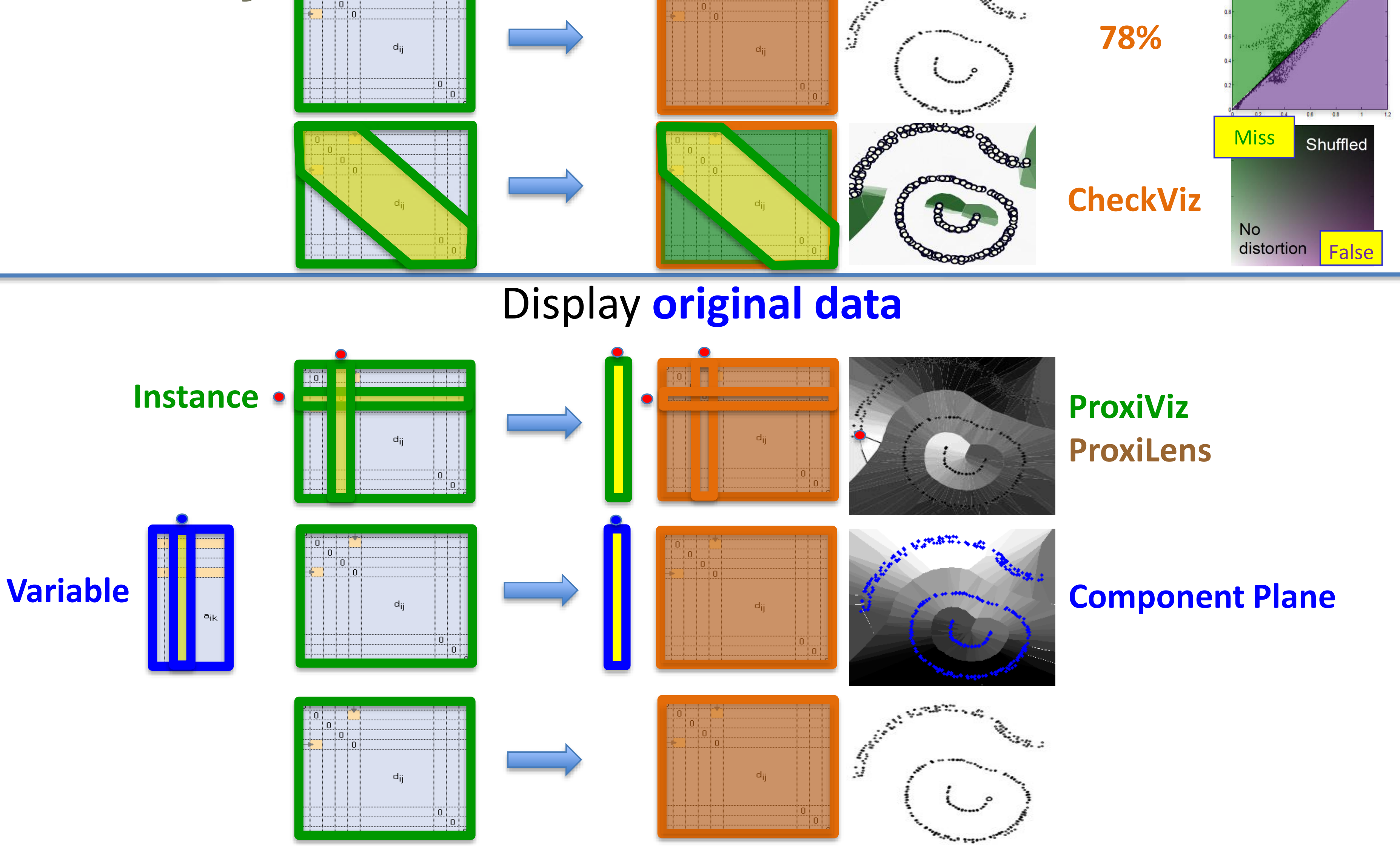
Display distortions



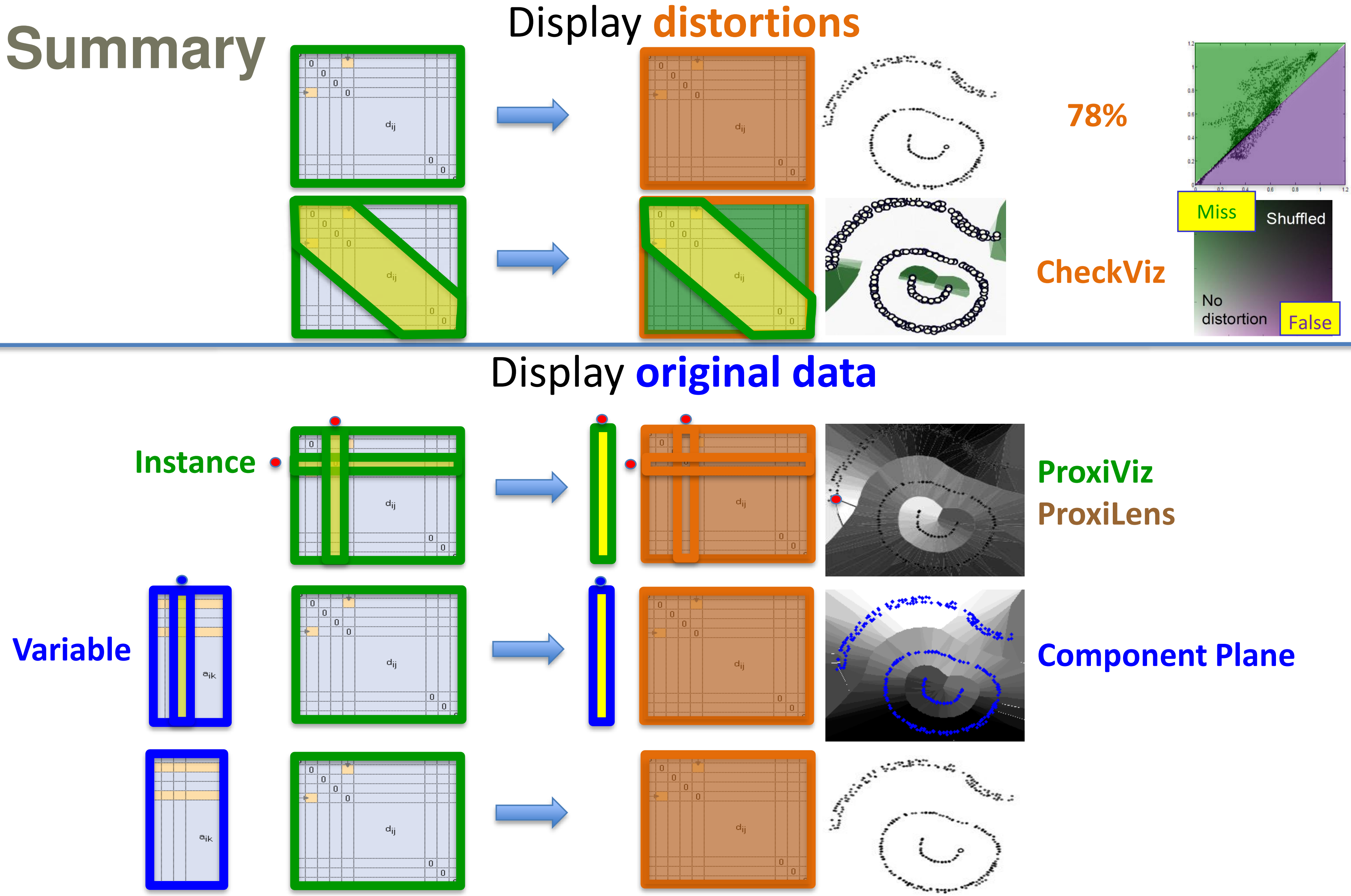
Display original data



Summary


Display distortions
0
0
0
0
dij
78%
Miss
Shuffled
No distortion
False
CheckViz
Display original data
Instance
ProxiViz
ProxiLens
Variable
aik
Component Plane

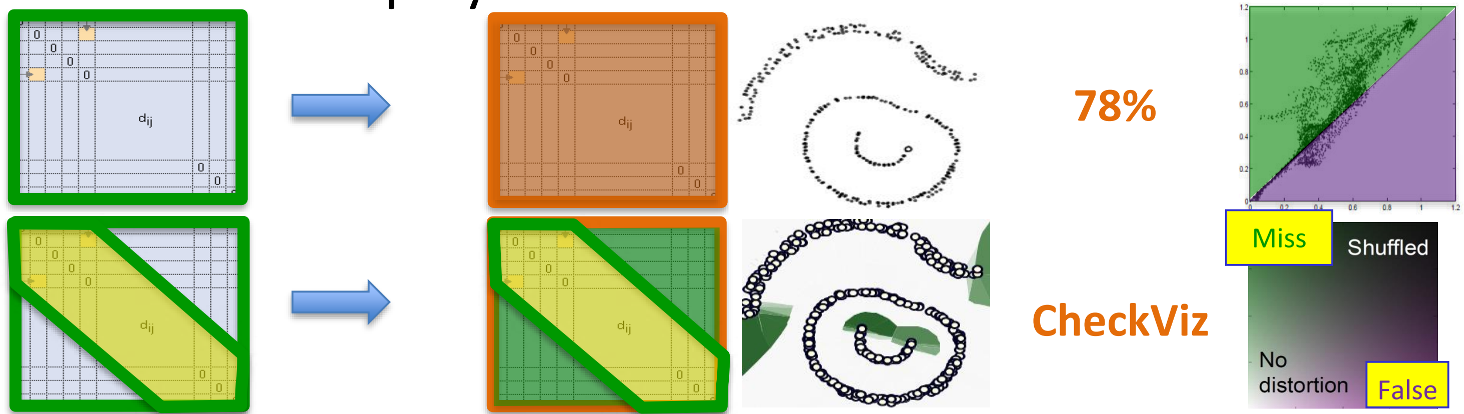
Summary



Display distortions
78%
Miss
Shuffled
No distortion
False
CheckViz
Display original data
Instance
ProxiViz
ProxiLens
Variable
Component Plane

Summary

Display distortions

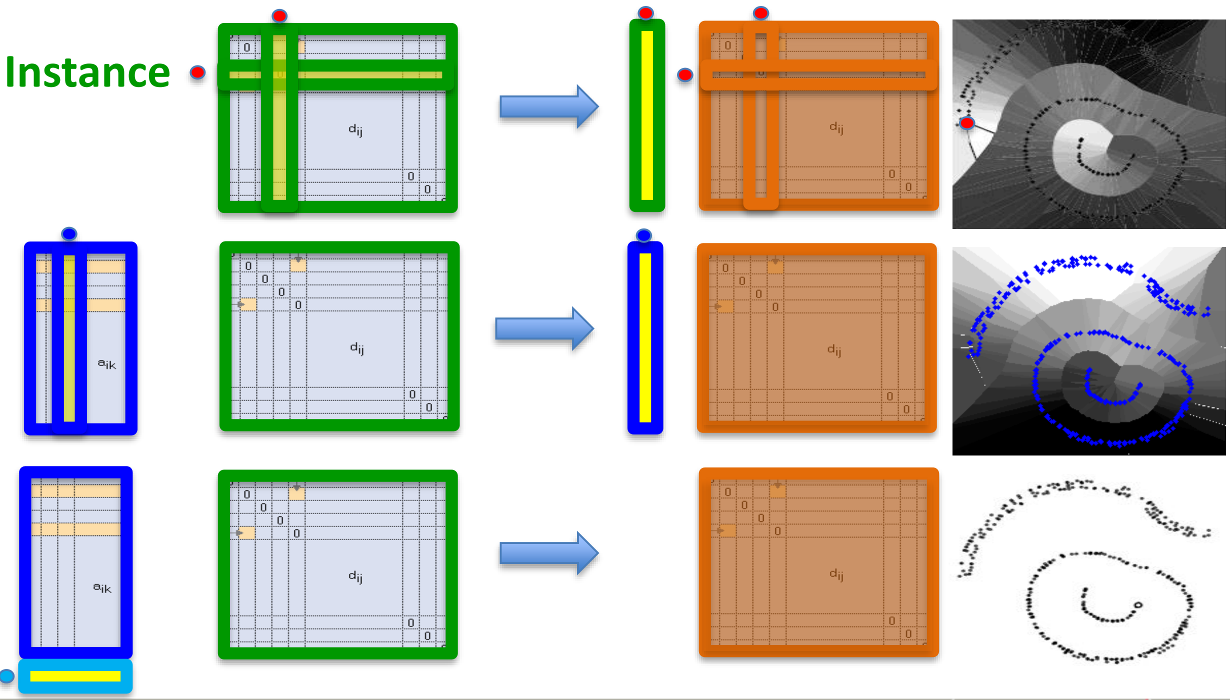


Display original data

Instance

Variable

Out-of-sample instance



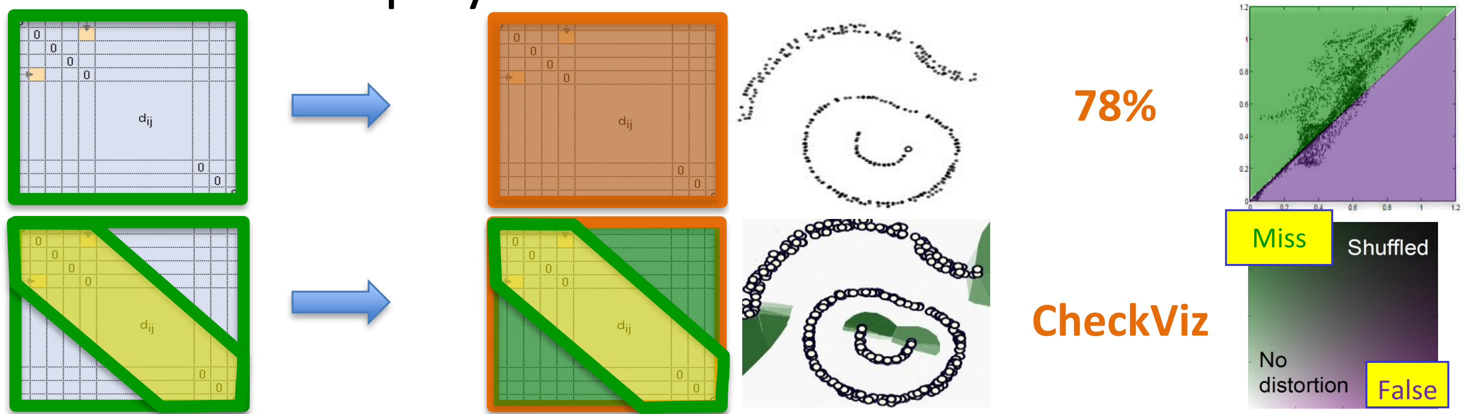
ProxiViz

ProxiLens

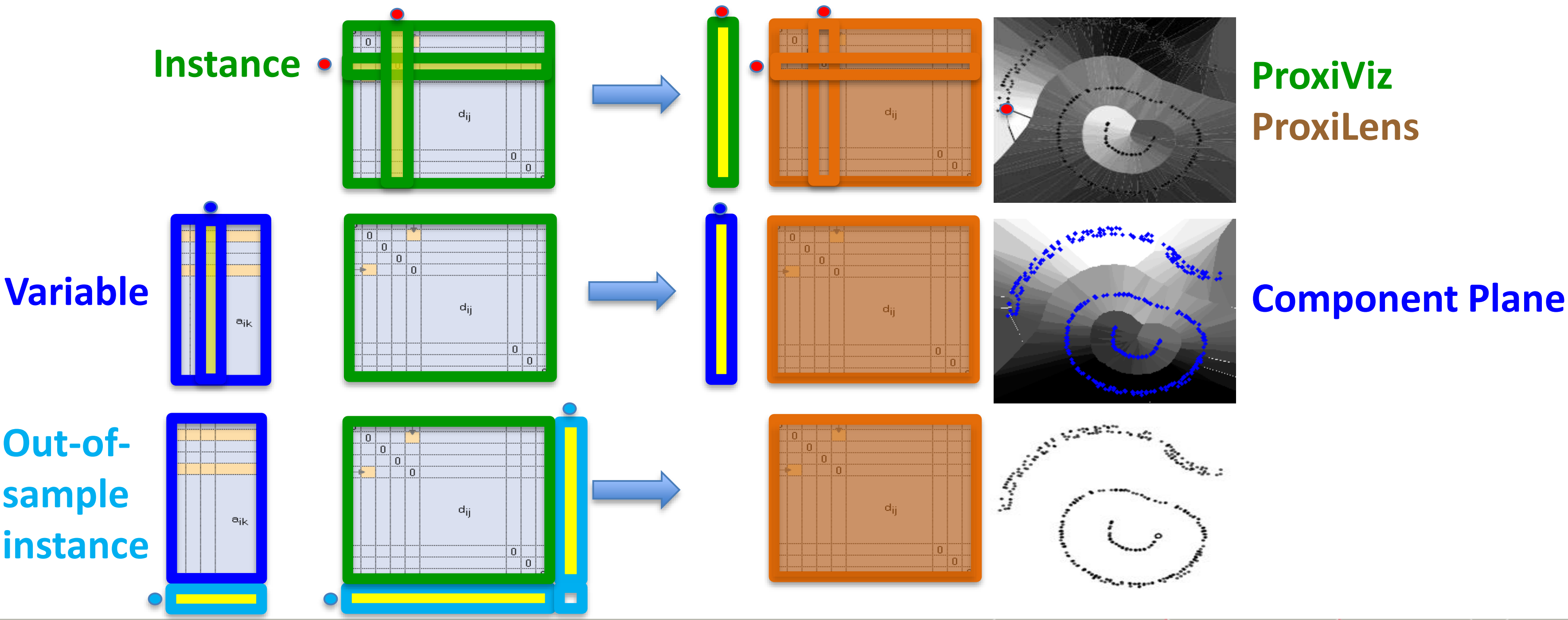
Component Plane

Summary

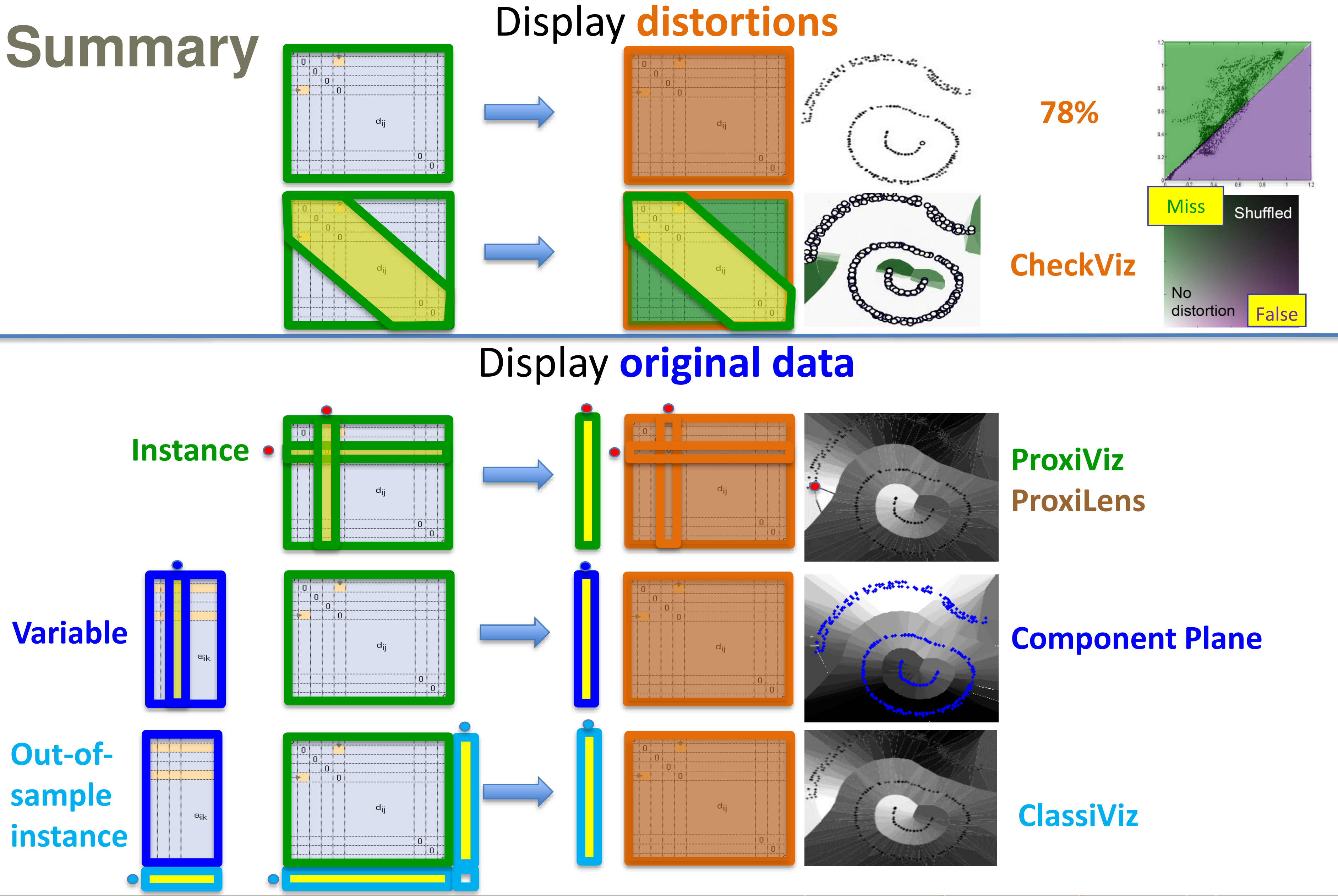
Display distortions



Display original data



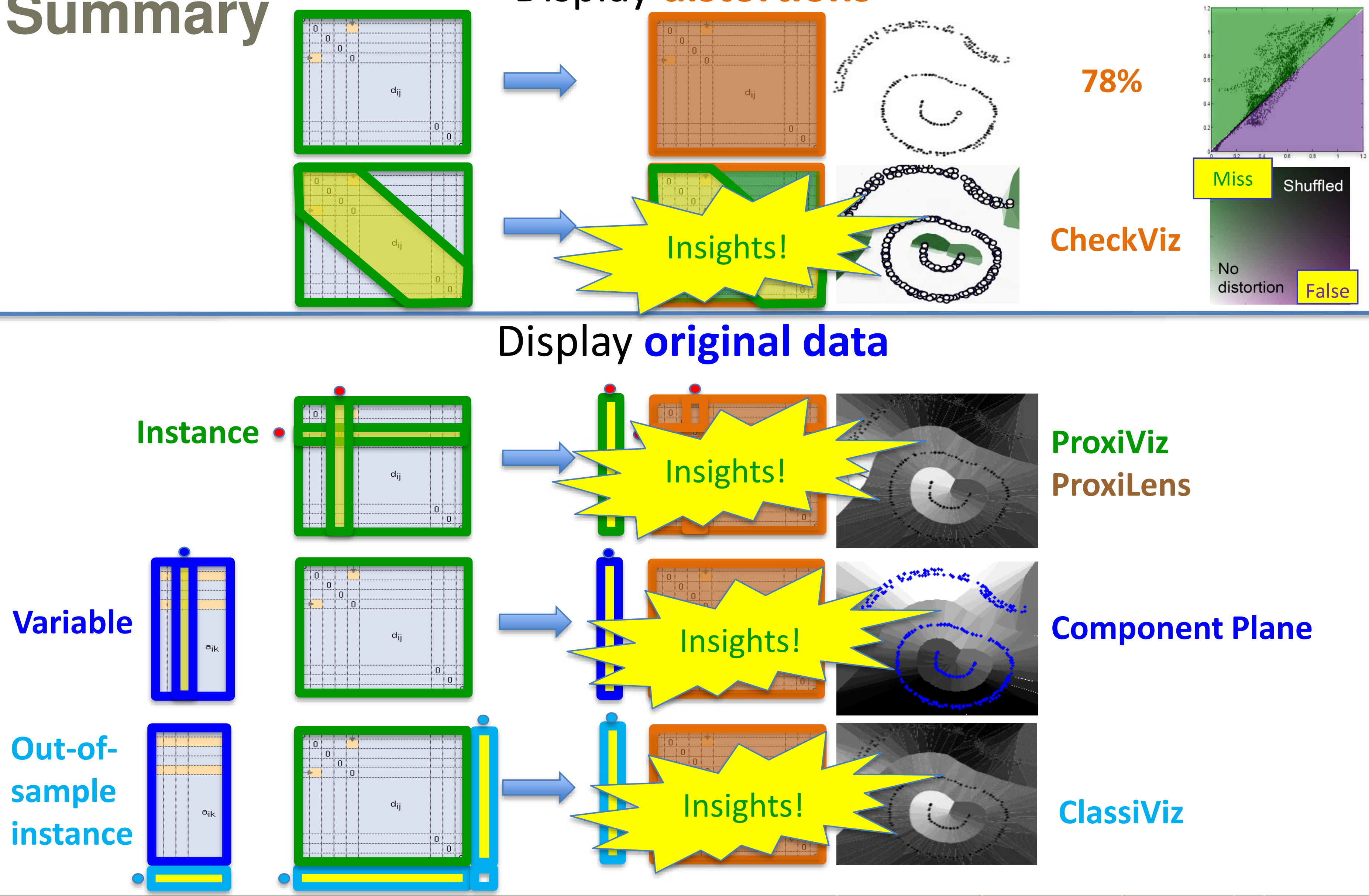
Summary



Display distortions
78%
Miss
Shuffled
CheckViz
No distortion
False
Display original data
Instance
ProxiViz
ProxiLens
Variable
Component Plane
Out-of-sample instance
ClassiViz

Summary

Display distortions

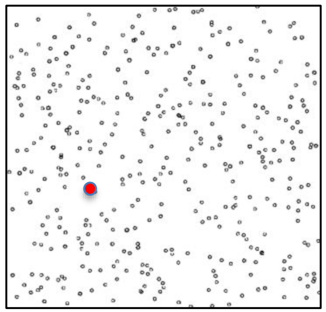


Analysis

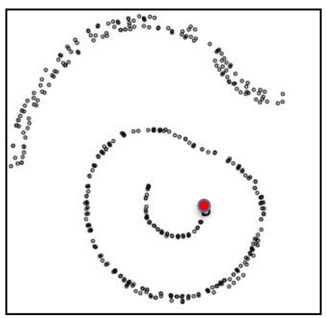
ProxiViz

Position

Random

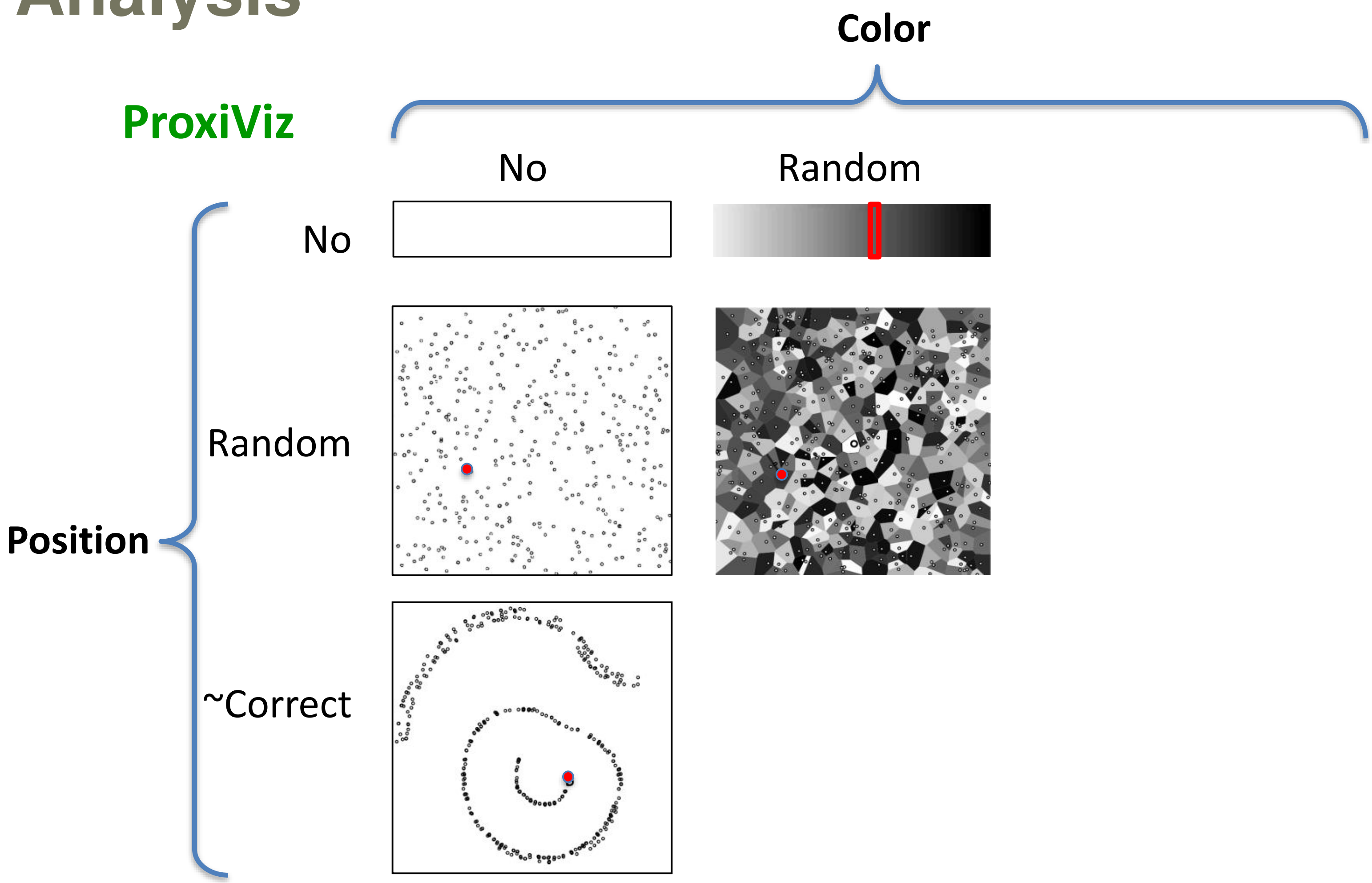


~Correct

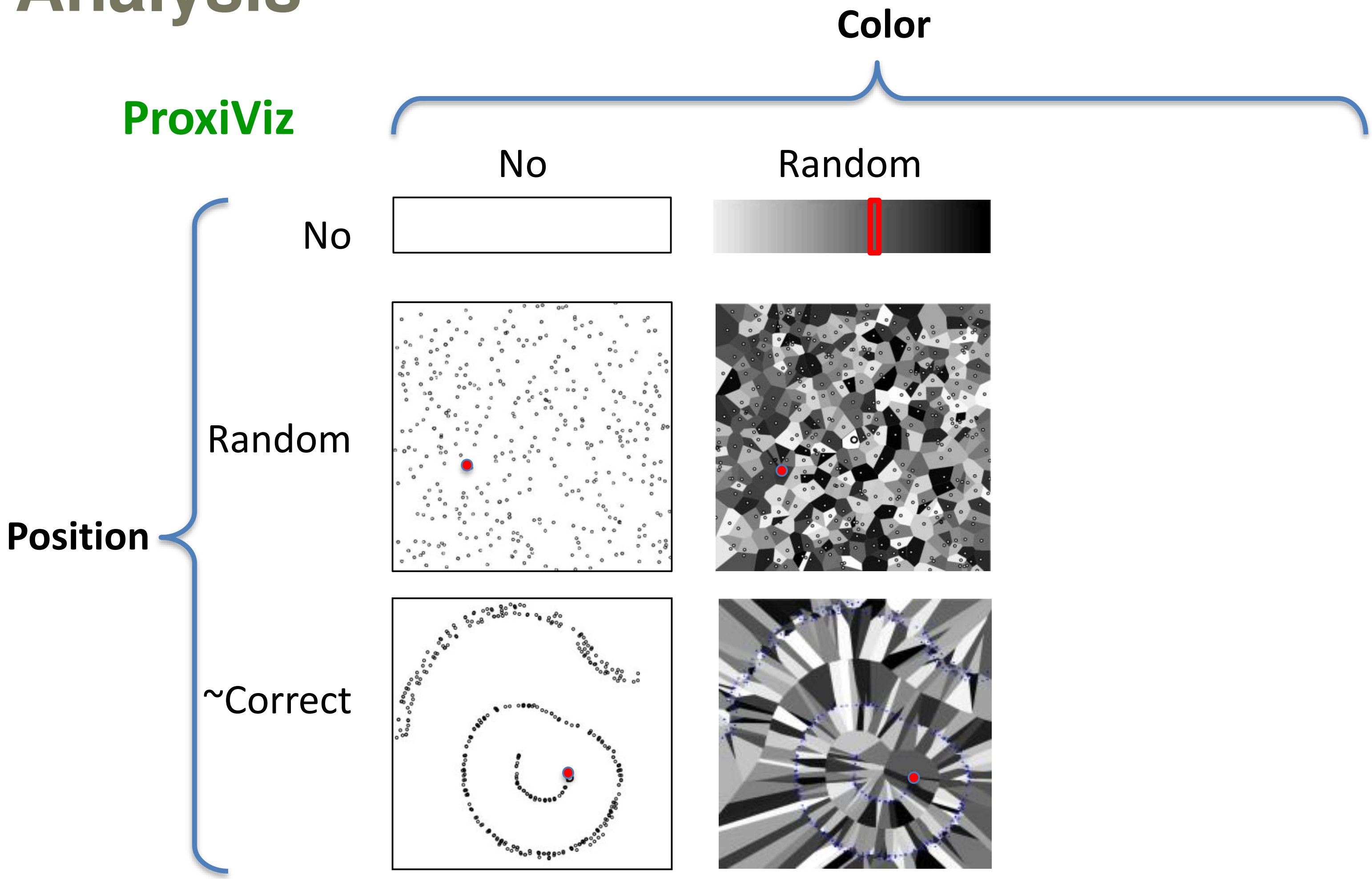


Analysis

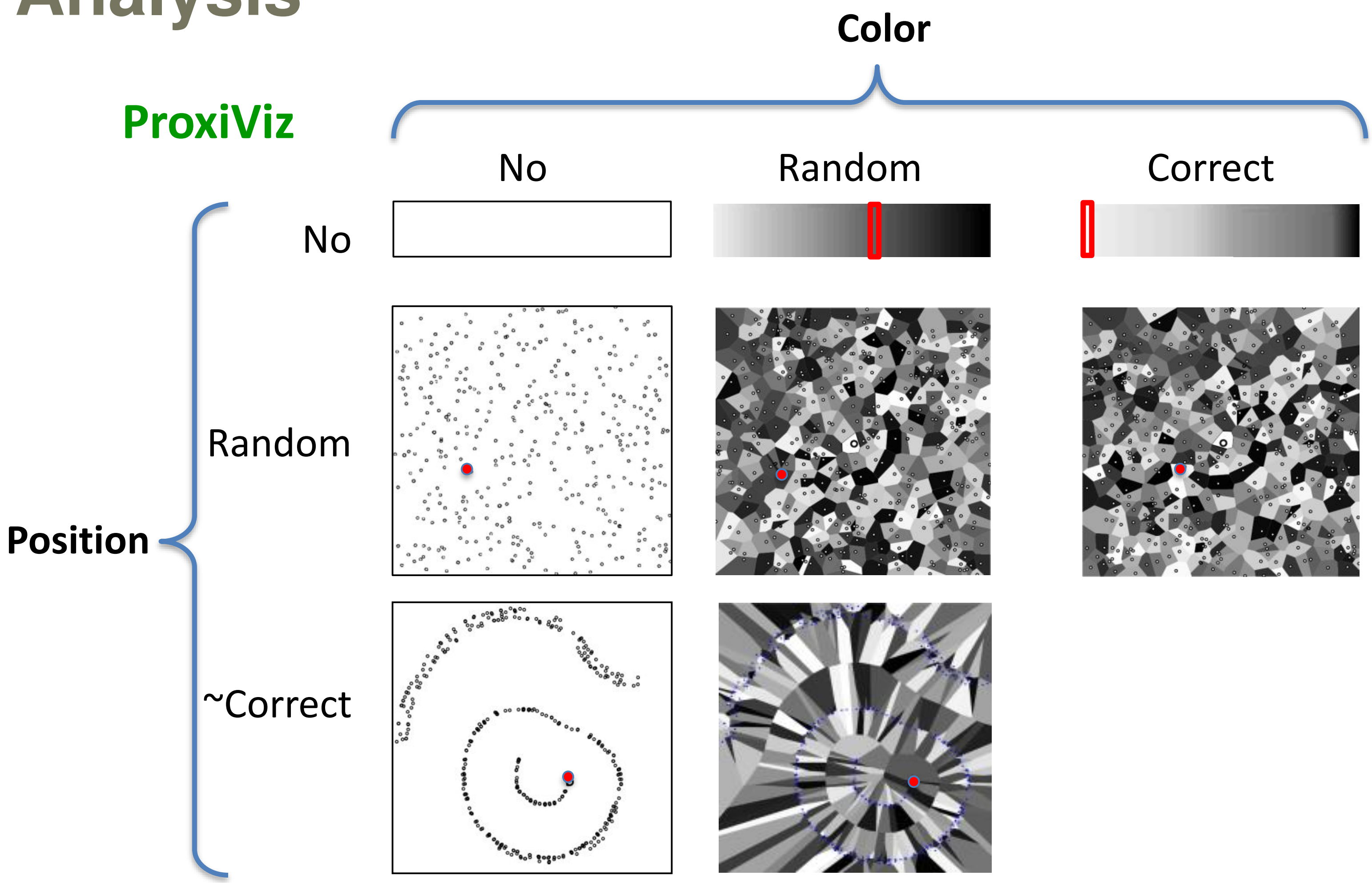
ProxiViz



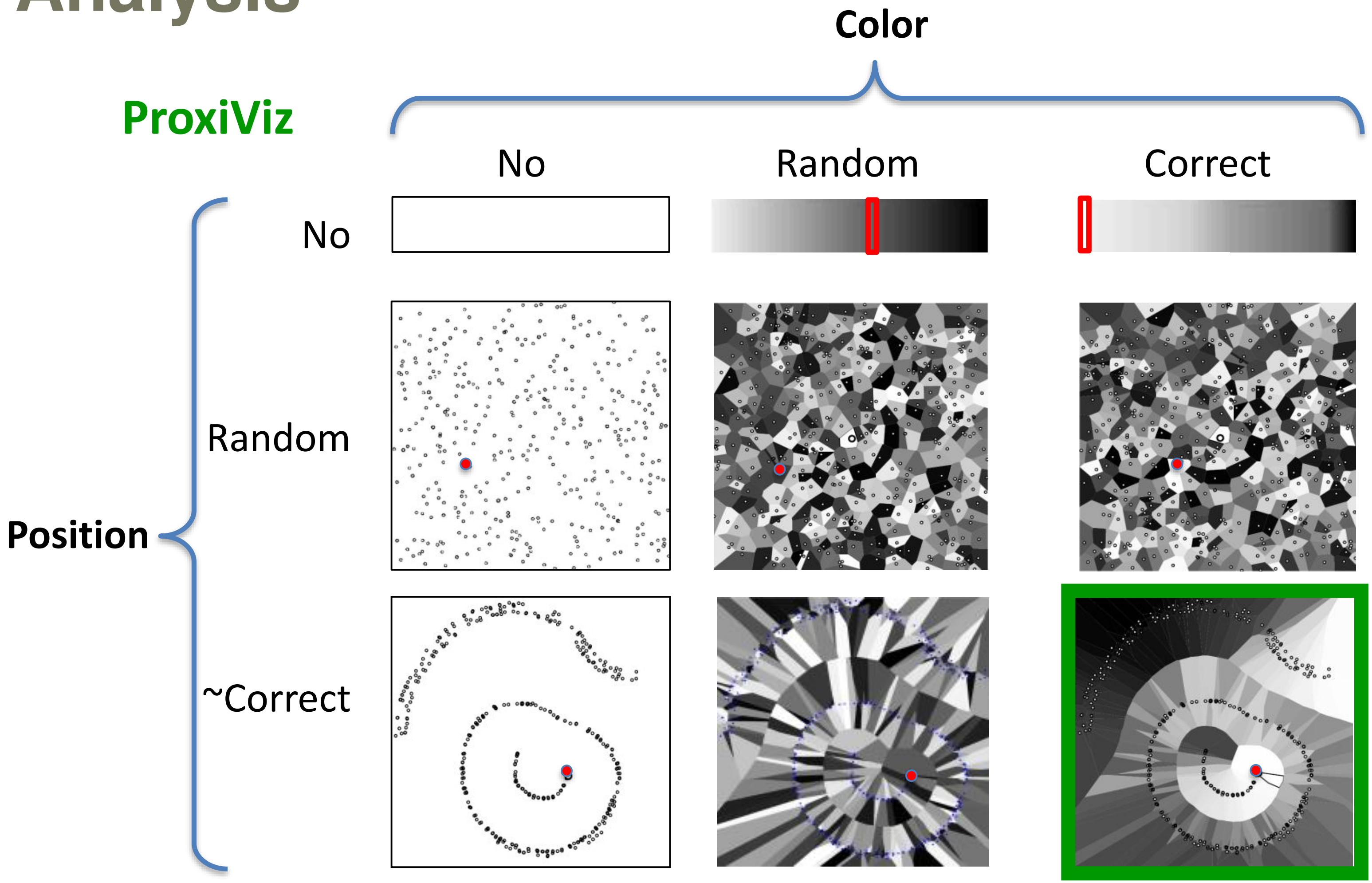
Analysis



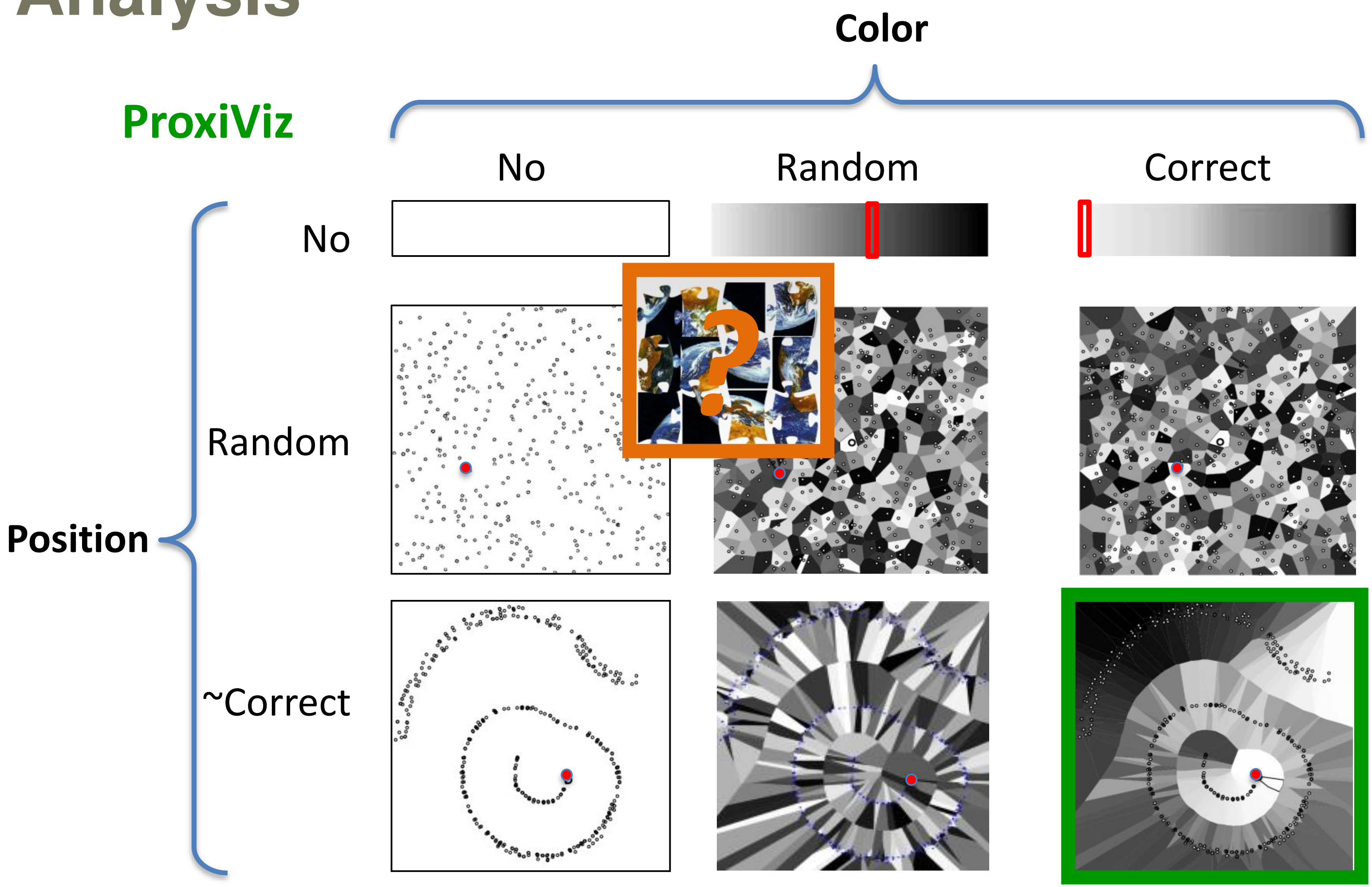
Analysis


ProxiViz
Color
No
Random
Correct
Position
No
Random
~Correct

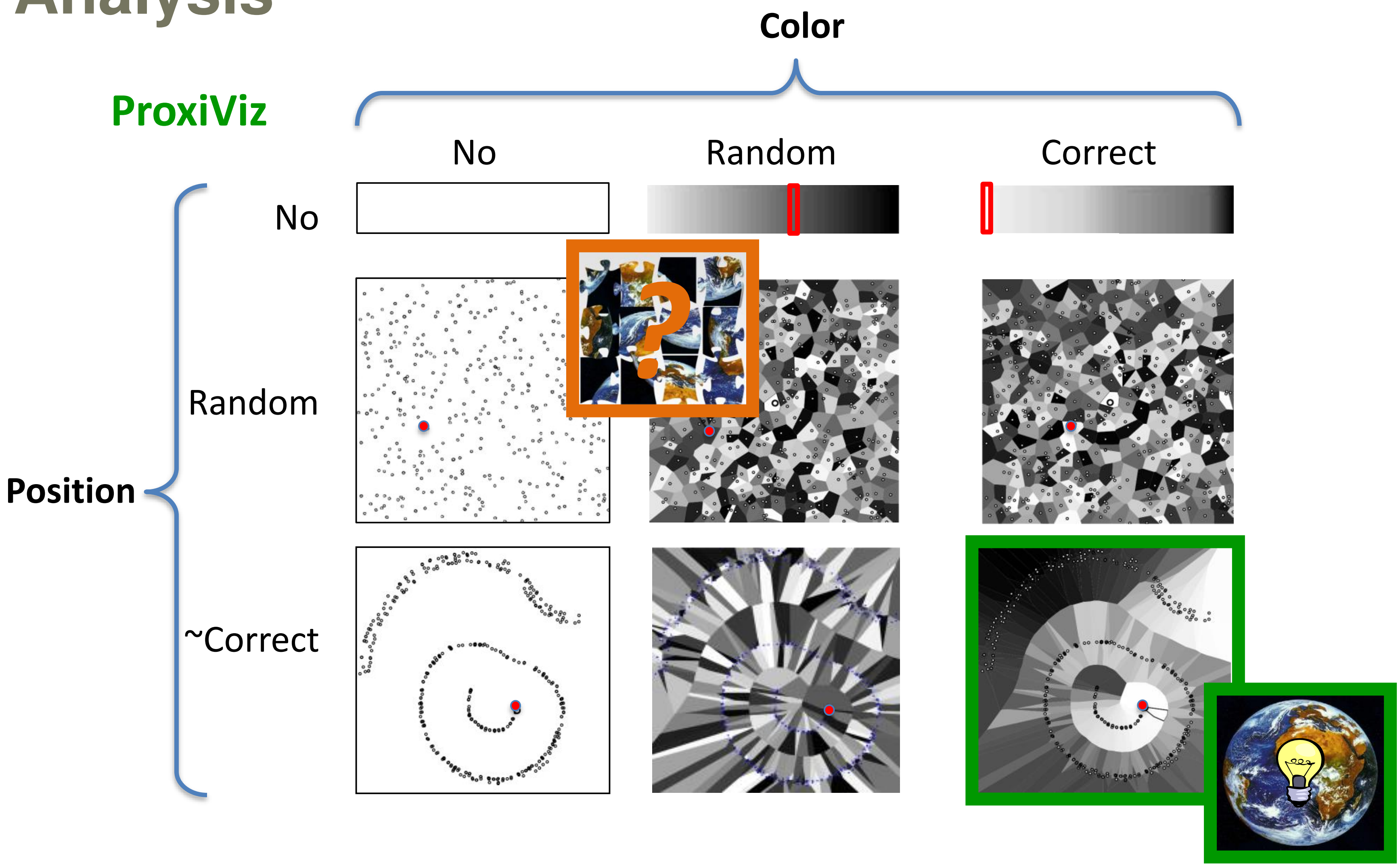
Analysis


ProxiViz
Color
No
Random
Correct
Position
No
Random
~Correct

Analysis

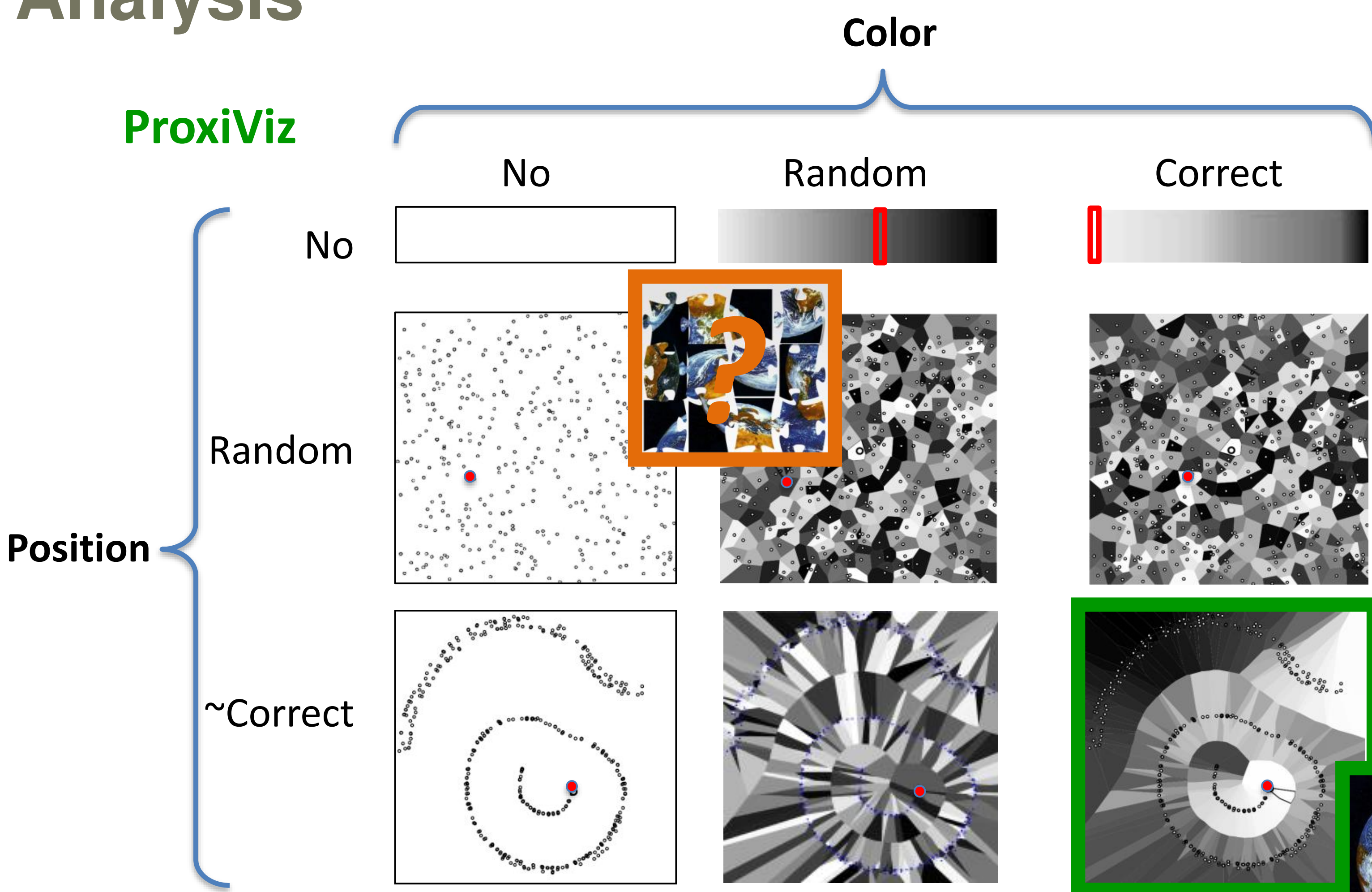

Color
ProxiViz
No
Random
Correct
Position
No
Random
~Correct

Analysis


ProxiViz
Color
No
Random
Correct
Position
No
Random
~Correct

Analysis

ProxiViz

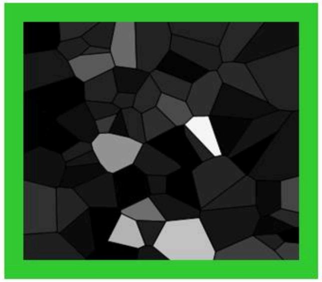


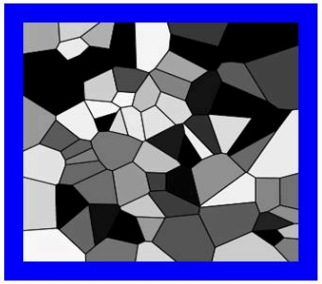
• **Correlation** between position and color is crucial to **detect patterns**

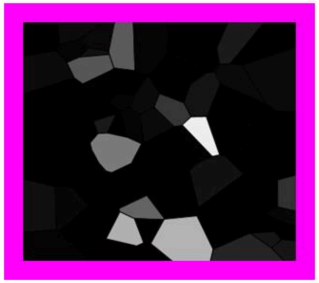
Analysis

ClassiViz

Random



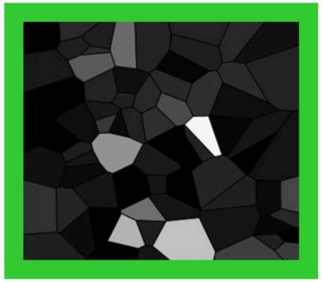
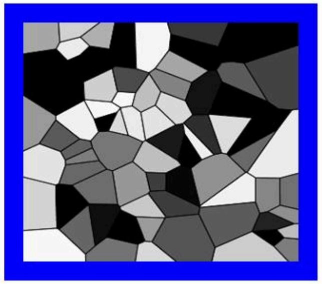
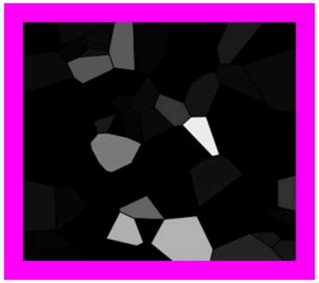




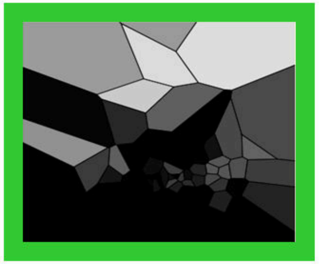
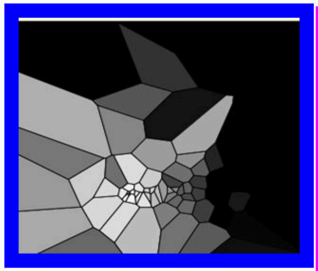
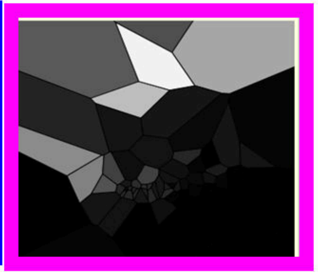
Analysis

ClassiViz

Random


MDS (DD-HDS)

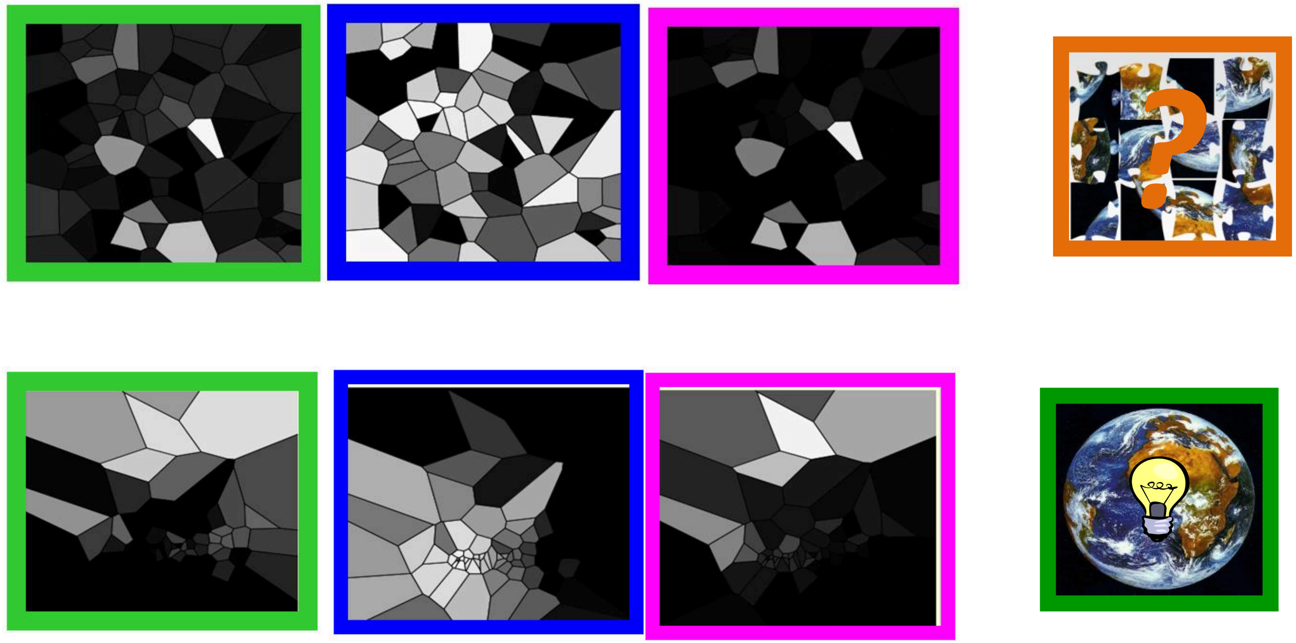





Analysis

ClassiViz

Random

MDS (DD-HDS)



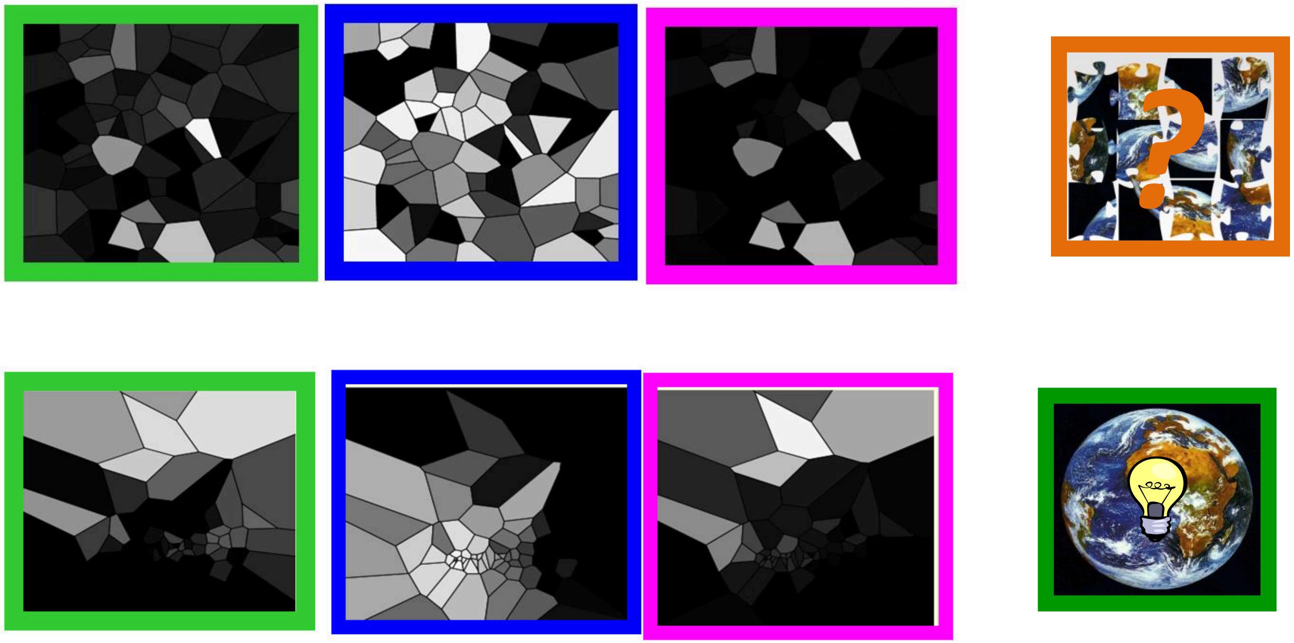
- **Correlation** between position and color is crucial

Analysis

ClassiViz

Random

MDS (DD-HDS)

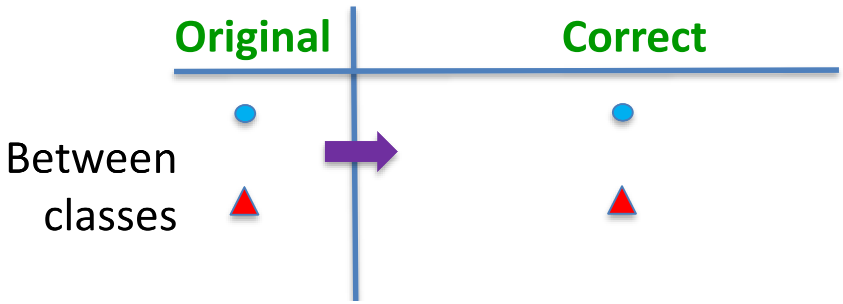


- **Correlation** between position and color is crucial to **lower cognitive load and ease decision**

Turning weaknesses into strengths

- Mapping **labeled data** **using distortions**

UCI Digits (D=64, N=5264, C=10)

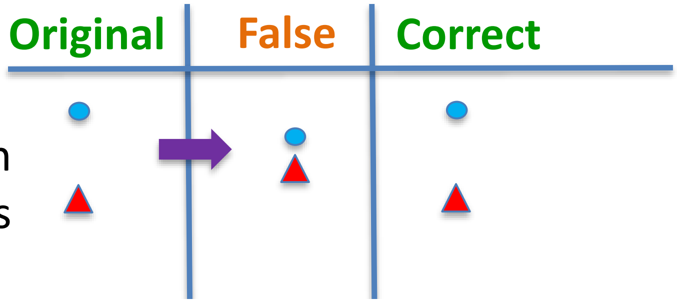


Turning weaknesses into strengths

- Mapping **labeled data using distortions**

UCI Digits (D=64, N=5264, C=10)

Between classes

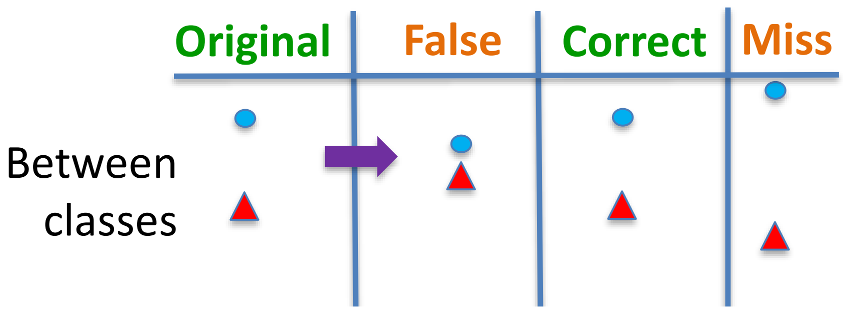


 Lespinats, Aupetit, Meyer-Baese. ClassiMap: A New Dimension Reduction Technique for Exploratory Data Analysis of Labeled Data. IJPRAI 29(6) (2015)

Turning weaknesses into strengths

- Mapping **labeled data** **using distortions**

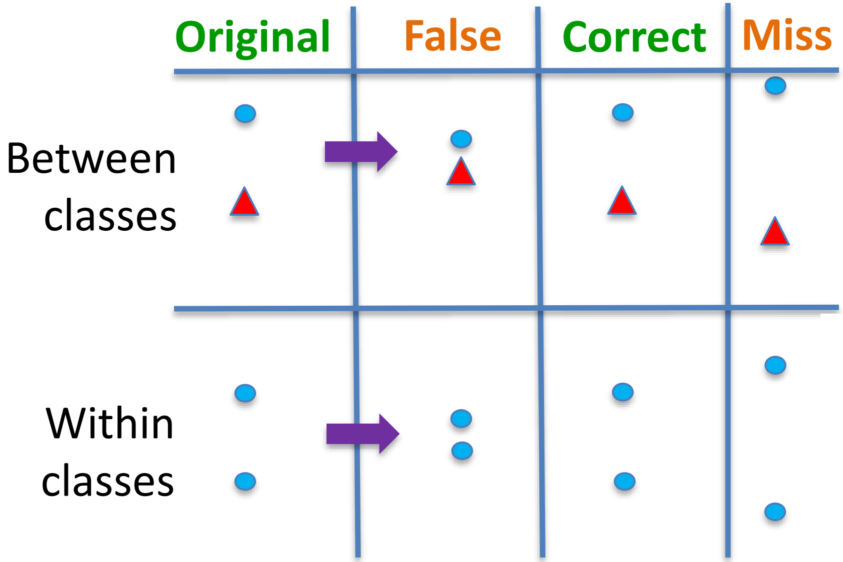
UCI Digits (D=64, N=5264, C=10)



Turning weaknesses into strengths

- Mapping **labeled data** **using distortions**

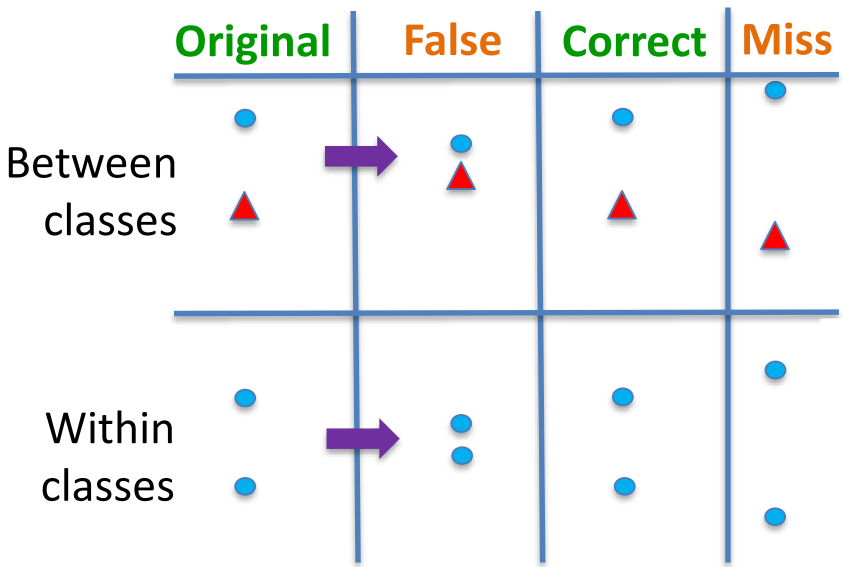
UCI Digits (D=64, N=5264, C=10)



Lespinats, Aupetit, Meyer-Baese. ClassiMap: A New Dimension Reduction Technique for Exploratory Data Analysis of Labeled Data. IJPRAI 29(6) (2015)

Turning weaknesses into strengths

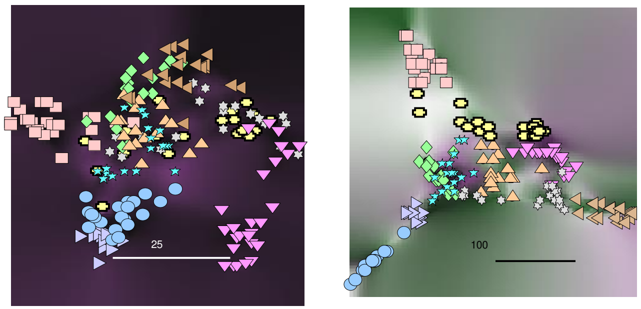
- Mapping **labeled data** **using distortions**



UCI Digits (D=64, N=5264, C=10)

Ignoring class labels

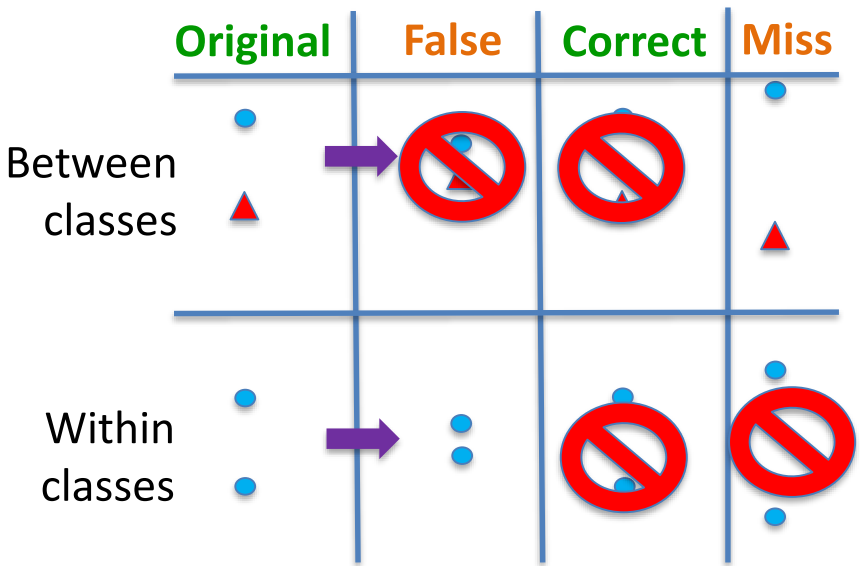
PCA **Isomap**



 Lespinats, Aupetit, Meyer-Baese. ClassiMap: A New Dimension Reduction Technique for Exploratory Data Analysis of Labeled Data. IJPRAI 29(6) (2015)

Turning weaknesses into strengths

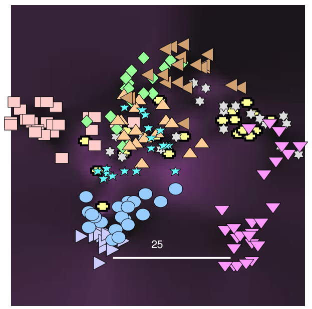
- Mapping **labeled data** **using distortions**



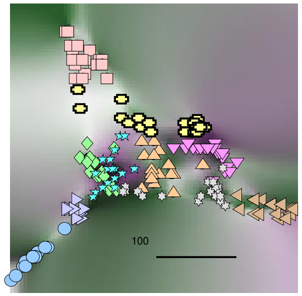
UCI Digits (D=64, N=5264, C=10)

Ignoring class labels

PCA

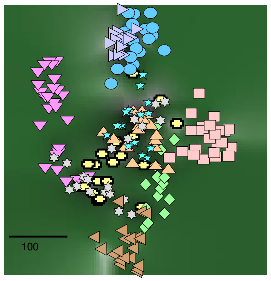


Isomap

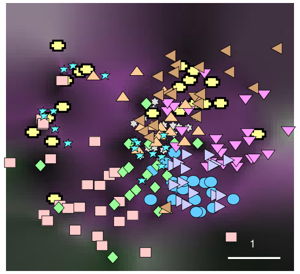


Forcing class separation

LDA



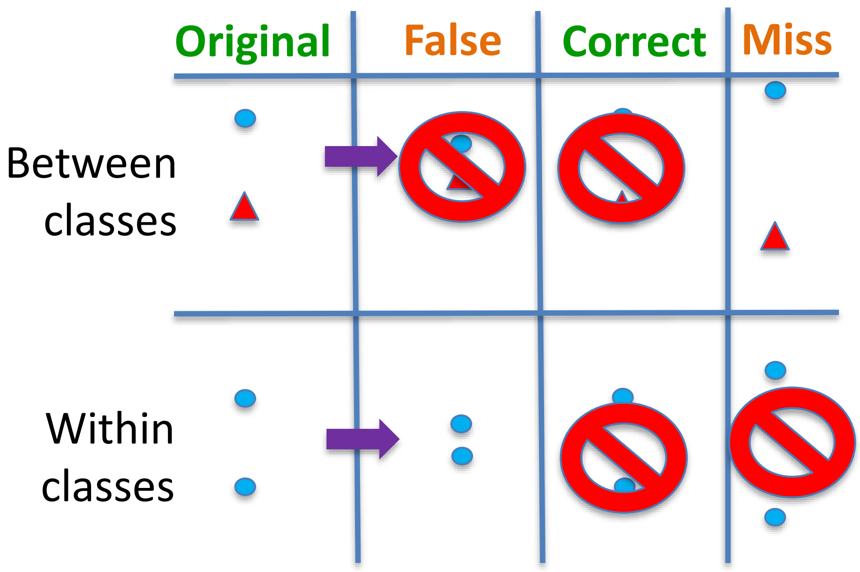
S-NeRV



ClassiMap Lespinats, Aupetit, Meyer-Baese. ClassiMap: A New Dimension Reduction Technique for Exploratory Data Analysis of Labeled Data. IJPRAI 29(6) (2015)

Turning weaknesses into strengths

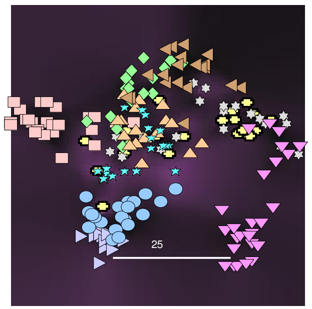
- Mapping **labeled data using distortions**



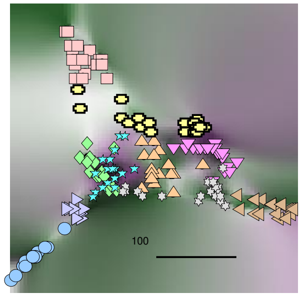
UCI Digits (D=64, N=5264, C=10)

Ignoring class labels

PCA

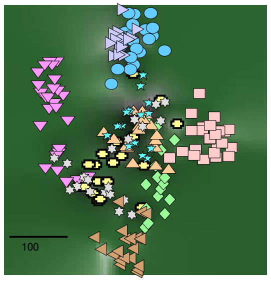


Isomap

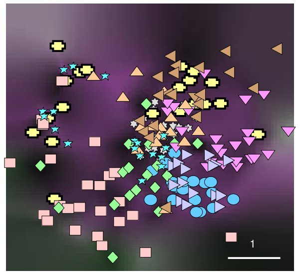


Forcing class separation

LDA

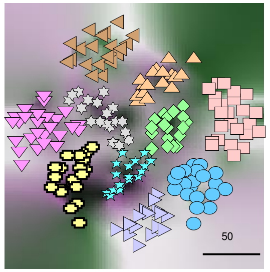


S-NeRV



Using distortions

ClassiMap



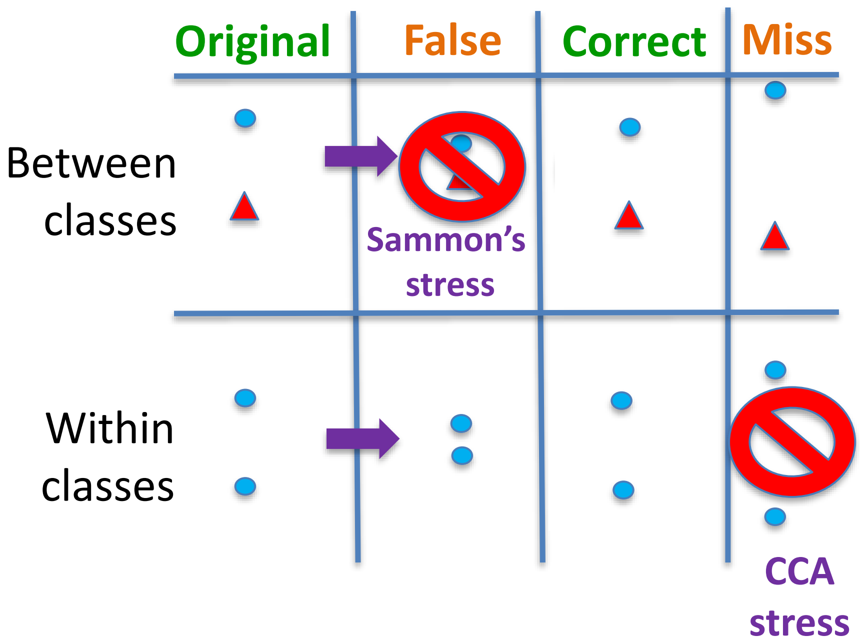
Lespinats, Aupetit, Meyer-Baese. ClassiMap: A New Dimension Reduction Technique for Exploratory Data Analysis of Labeled Data. IJPRAI 29(6) (2015)

Turning weaknesses into strengths

- Mapping **labeled data** **using distortions**

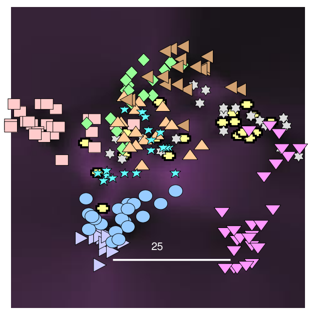
UCI Digits (D=64, N=5264, C=10)

ClassiMap

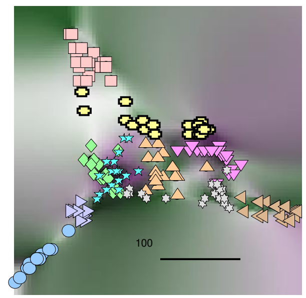


Ignoring class labels

PCA

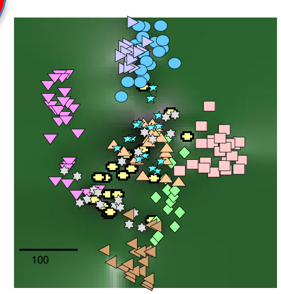


Isomap

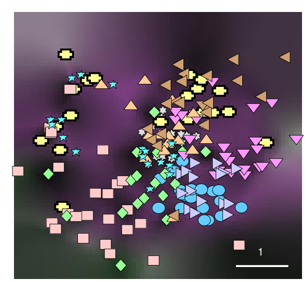


Forcing class separation

LDA

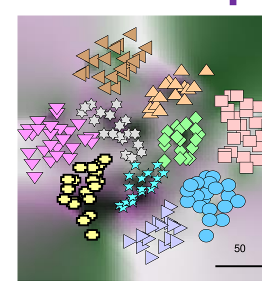


S-NeRV



Using distortions

ClassiMap



Outline

- What is Dimensionality Reduction
- Strengths
- Weaknesses
- Turning weaknesses into strengths
- **Conclusion**
- **Perspectives**

Conclusion

Conclusion

- Distortions cannot be avoided but

Conclusion

- Distortions cannot be avoided but
  - Can be used to **infer HD structures** (**CheckViz**)

Conclusion

- Distortions cannot be avoided but
  - Can be used to **infer HD structures** (CheckViz)
  - Can be exploited to get **better map of labeled data** (ClassiMap)

Conclusion

- Distortions cannot be avoided but
  - Can be used to **infer HD structures** (**CheckViz**)
  - Can be exploited to get **better map of labeled data** (**ClassiMap**)
  - Maps even incorrect but not random can be **used as a support**

Conclusion

- Distortions cannot be avoided but
  - Can be used to **infer HD structures** (**CheckViz**)
  - Can be exploited to get **better map of labeled data** (**ClassiMap**)
  - Maps even incorrect but not random can be **used as a support**
    - to **explore actual local similarities** (**ProxiVis**)

Conclusion

- Distortions cannot be avoided but
  - Can be used to **infer HD structures** (**CheckViz**)
  - Can be exploited to get **better map of labeled data** (**ClassiMap**)
  - Maps even incorrect but not random can be **used as a support**
    - to **explore actual local similarities** (**ProxiVis**)
    - to **navigate HD data** (**ProxiLens**)

Conclusion

- Distortions cannot be avoided but
  - Can be used to **infer HD structures** (**CheckViz**)
  - Can be exploited to get **better map of labeled data** (**ClassiMap**)
  - Maps even incorrect but not random can be **used as a support**
    - to **explore actual local similarities** (**ProxiVis**)
    - to **navigate HD data** (**ProxiLens**)
    - to **brush HD data** (**MD-Brush**)

Conclusion

- Distortions cannot be avoided but
    - Can be used to **infer HD structures** (**CheckViz**)
    - Can be exploited to get **better map of labeled data** (**ClassiMap**)
    - Maps even incorrect but not random can be **used as a support**
        - to **explore actual local similarities** (**ProxiVis**)
        - to **navigate HD data** (**ProxiLens**)
        - to **brush HD data** (**MD-Brush**)
        - to **classify out-of-sample** data (**ClassiViz**)

Conclusion

- Distortions cannot be avoided but
  - Can be used to **infer HD structures** (**CheckViz**)
  - Can be exploited to get **better map of labeled data** (**ClassiMap**)
  - Maps even incorrect but not random can be **used as a support**
    - to **explore actual local similarities** (**ProxiVis**)
    - to **navigate HD data** (**ProxiLens**)
    - to **brush HD data** (**MD-Brush**)
    - to **classify out-of-sample** data (**ClassiViz**)
    - to **detect correlation** between variables (**Component Planes**)

Conclusion

- Distortions cannot be avoided but
  - Can be used to **infer HD structures** (**CheckViz**)
  - Can be exploited to get **better map of labeled data** (**ClassiMap**)
  - Maps even incorrect but not random can be **used as a support**
    - to **explore actual local similarities** (**ProxiVis**)
    - to **navigate HD data** (**ProxiLens**)
    - to **brush HD data** (**MD-Brush**)
    - to **classify out-of-sample** data (**ClassiViz**)
    - to **detect correlation** between variables (**Component Planes**)
- Maps are **useless without additional information**

Conclusion

- Distortions cannot be avoided but
    - Can be used to **infer HD structures** (**CheckViz**)
    - Can be exploited to get **better map of labeled data** (**ClassiMap**)
    - Maps even incorrect but not random can be **used as a support**
        - to **explore actual local similarities** (**ProxiVis**)
        - to **navigate HD data** (**ProxiLens**)
        - to **brush HD data** (**MD-Brush**)
        - to **classify out-of-sample** data (**ClassiViz**)
        - to **detect correlation** between variables (**Component Planes**)

- Maps are **useless without additional information**
    - **Can generate false pattern**

Conclusion

- Distortions cannot be avoided but
  - Can be used to **infer HD structures** (**CheckViz**)
  - Can be exploited to get **better map of labeled data** (**ClassiMap**)
  - Maps even incorrect but not random can be **used as a support**
    - to **explore actual local similarities** (**ProxiVis**)
    - to **navigate HD data** (**ProxiLens**)
    - to **brush HD data** (**MD-Brush**)
    - to **classify out-of-sample** data (**ClassiViz**)
    - to **detect correlation** between variables (**Component Planes**)

- Maps are **useless without additional information**
  - **Can generate false pattern**
  - **Can miss actual pattern**

Conclusion

- Distortions cannot be avoided but
  - Can be used to **infer HD structures** (**CheckViz**)
  - Can be exploited to get **better map of labeled data** (**ClassiMap**)
  - Maps even incorrect but not random can be **used as a support**
    - to **explore actual local similarities** (**ProxiVis**)
    - to **navigate HD data** (**ProxiLens**)
    - to **brush HD data** (**MD-Brush**)
    - to **classify out-of-sample** data (**ClassiViz**)
    - to **detect correlation** between variables (**Component Planes**)
- Maps are **useless without additional information**
  - **Can generate false pattern**
  - **Can miss actual pattern**
- Maps are **not an end but a mean** to support user decision

Perspectives

Perspectives

- Formalization needed to go beyond similarity-based distortions

Perspectives

- Formalization needed to go beyond similarity-based distortions
  - Interesting patterns all based on similarities
    - Outliers
    - Class outliers
    - Clusters
    - Class structure...

Perspectives

- Formalization needed to go beyond similarity-based distortions
  - Interesting patterns all based on similarities
    - Outliers
    - Class outliers
    - Clusters
    - Class structure...
  - What is the impact of similarity distortions on specific pattern detection?

Perspectives

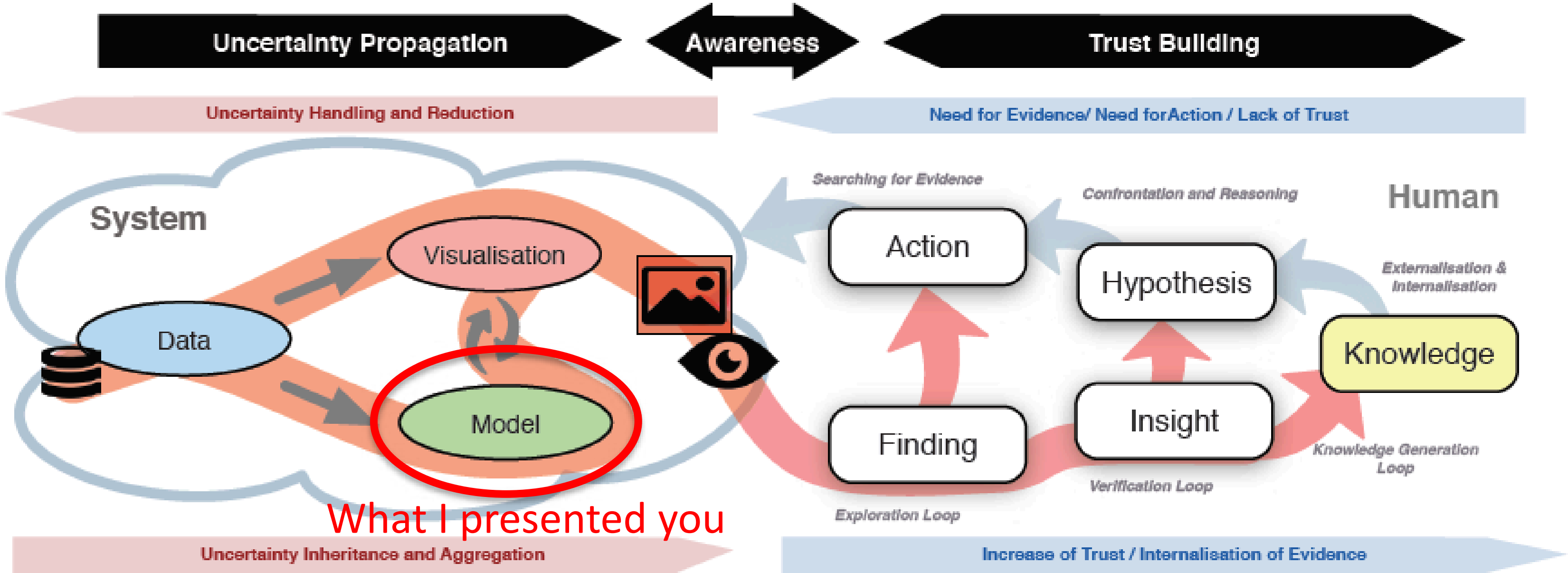
- Formalization needed to go beyond similarity-based distortions
  - Interesting patterns all based on similarities
    - Outliers
    - Class outliers
    - Clusters
    - Class structure...
  - What is the impact of similarity distortions on specific pattern detection?
  - Can we design DR techniques tuned to optimize preservation of specific patterns?

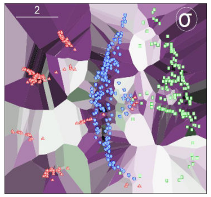
Worth reading

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 22, NO. 1, JANUARY 2016

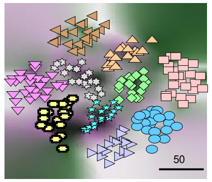
The Role of Uncertainty, Awareness, and Trust in Visual Analytics

Dominik Sacha, Hansi Senaratne, Bum Chul Kwon, *Member, IEEE*, Geoffrey Ellis, and Daniel A. Keim, *Member, IEEE*

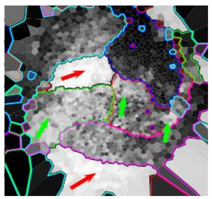




[CheckViz] Lespinats, Aupetit. ***CheckViz: Sanity Check and Topological Clues for Linear and Non-Linear Mappings***. Computer Graphics Forum 30(1): 113-125 (2011)

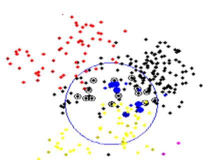


[ClassiMap] Lespinats, Aupetit, Meyer-Baese. ClassiMap: ***A New Dimension Reduction Technique for Exploratory Data Analysis of Labeled Data***. IJPRAI 29(6) (2015)

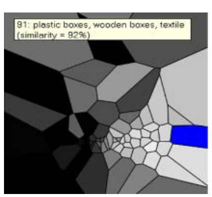


[ProxiViz] Aupetit. ***Visualizing distortions and recovering topology in continuous projection techniques***. Neurocomputing 70(7-9): 1304-1330 (2007)

[ProxiLens] Heulot, Aupetit, Fekete. ***ProxiLens: Interactive Exploration of High-Dimensional Data using Projections***
EuroVis 2013 Workshop on Visual analytics using Multidimensional Projections, Leipzig, Germany, June 2013



[HDBrush] Aupetit, Heulot, Fekete. ***A multidimensional brush for scatterplot data analytics*** Poster at IEEE VIS 2014 conference, Paris, France, November 2014



[ClassiViz] Aupetit, Allano, Espagnon, Sannie. ***Visual Analytics to Check Marine Containers in the Eritr@c Project*** EuroVAST (2010)

Qatar Computing Research Institute

Thank you !

Q&A

- How users perceive these maps, how long it takes them to learn how to use these techniques. For instance Euritrac ClassiViz -> qualitatively but not quantitatively evaluated
- Compare with Nonato connected missed with edge bundle, compare to color coding -> yes, all at one glance while proxiviz requires selecting a focus point, but proxiviz shows the actual data (1 row of the similarity matrix), not a summary (= no free lunch), what is best for users?
- Multi scale topological structure in HD used to enrich map -> yes, could be used to enrich and interact with data
- Dual problem between graph drawing drawing all edges, and approach here where we only draw missed neighbors -> interesting remark, would need more formalism to better understand relations between both approaches

maupetit@hbku.edu.qa
https://about.me/michaelaupetit