MULTIDIMENSIONAL DATA

SOURCES AND TYPES OF *N*D DATA

MEASUREMENTS, SIMULATIONS
Physical, geographical, chemical...

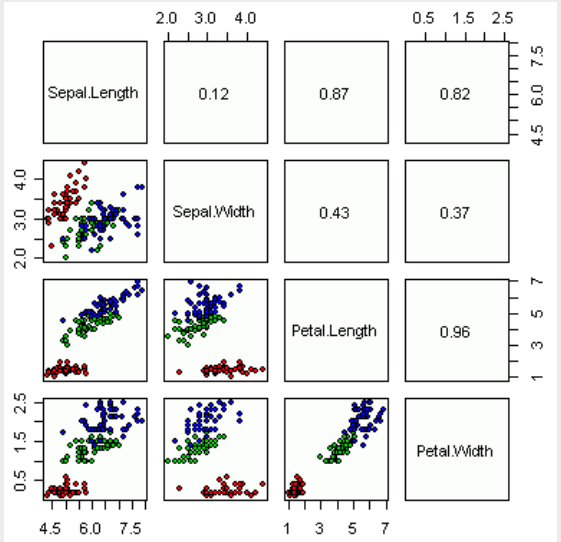
CENSUSES, QUESTIONNAIRES
Sociology, marketing, medicine...

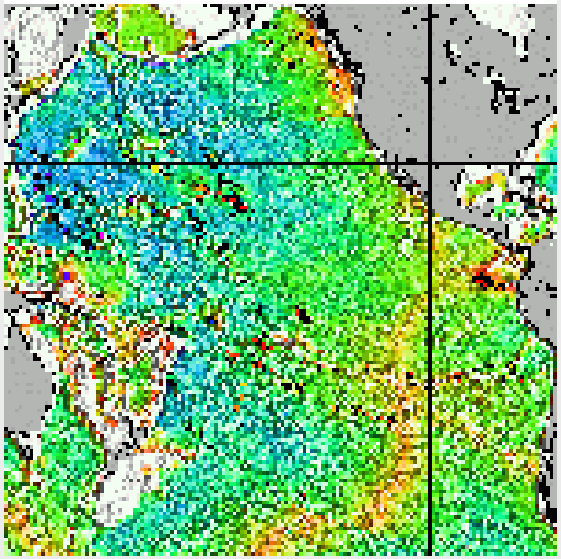
Our focus:
UNSTRUCTURED, MULTIDIMENSIONAL,
POSSIBLY HOMOGENEOUS,
USUALLY NUMERICAL

REAL WORLD *N*D DATA:

IRIS, CARS (*N* = 4)
CENSUS INCOME (*N* = 42)
SKYSERVER (*N* = 361)

TALL TABLES VS
WIDE TABLES DUALITY
attributes ↔ observations





PURPOSE OF *N*D VISUALIZATION

RELATIONS

Clusters

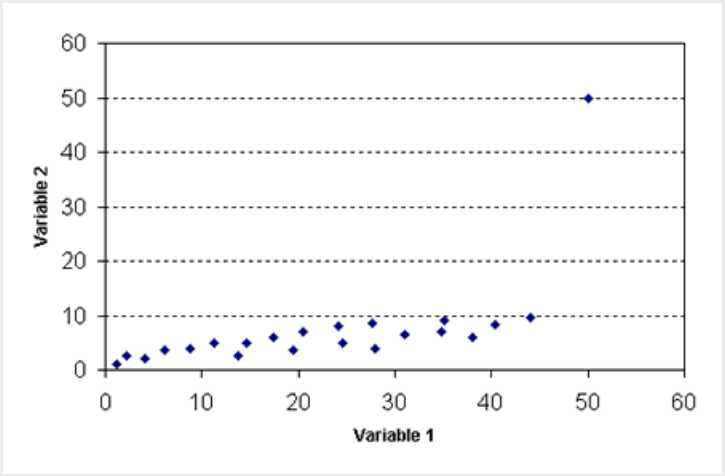
Trends

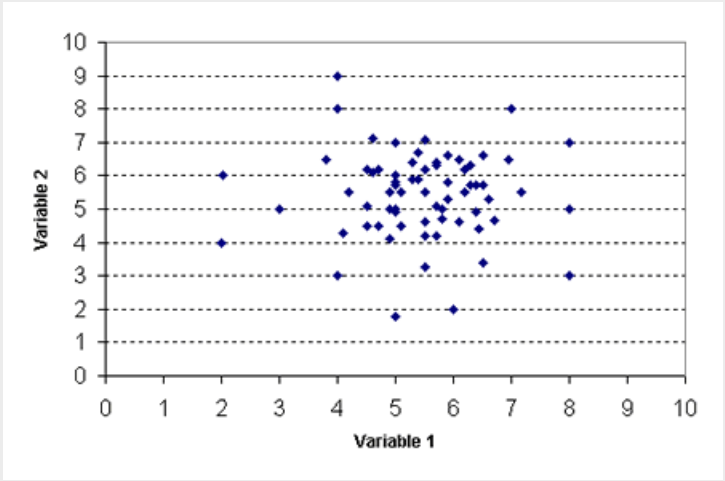
OUTLIERS

SEMANTICS OF DIMENSIONS

Correlated dimensions

Redundant dimensions





DEFINITION OF THE PROBLEM

*N*D → 2D

run, rerun, camcol, field, obj, mode, nChild, type, catID, flags, rowc, rowcErr, colc, colcErr, rowv, rowvErr, colv, colvErr, rowc_u, rowc_g, rowc_r, rowc_i, rowc_z, rowcErr_u, rowcErr_g, rowcErr_r, rowcErr_i, rowcErr_z, colc_u, colc_g, colc_r, colc_i, colc_z, colcErr_u, colcErr_g, colcErr_r, colcErr_i, colcErr_z, sky_u, sky_g, sky_r, sky_i, sky_z, skyErr_u, skyErr_g, skyErr_r, skyErr_i, skyErr_z, psfMag_u, psfMag_g, psfMag_r, psfMag_i, psfMag_z, psfMagErr_u, psfMagErr_g, psfMagErr_r, psfMagErr_i, psfMagErr_z, fiberMag_u, fiberMag_g, fiberMag_r, fiberMag_i, fiberMag_z, fiberMagErr_u, fiberMagErr_g, fiberMagErr_r , fiberMagErr_i, fiberMagErr_z, petroMag_u, petroMag_g, petroMag_r, petroMag_i, petroMag_z, petroMagErr_u, petroMagErr_g, petroMagErr_r, petroMagErr_i, petroMagErr_z, petroRad_u, petroRad_g, petroRad_r, petroRad_i, petroRad_z, petroRadErr_u, petroRadErr_g, petroRadErr_r, petroRadErr_i, petroRadErr_z, petroR50_u, petroR50_g, petroR50_r, petroR50_i, petroR50_z, petroR50Err_u, petroR50Err_g, petroR50Err_r, petroR50Err_i, petroR50Err_z, petroR90_u, petroR90_g, petroR90_r, petroR90_i, petroR90_z, petroR90Err_u, petroR90Err_g, petroR90Err_r, petroR90Err_i, petroR90Err_z, q_u, q_g, q_r, q_i, q_z, qErr_u, qErr_g, qErr_r, qErr_i, qErr_z , u_u, u_g, u_r, u_i, u_z, uErr_u, uErr_g, uErr_r, uErr_i, uErr_z , isoRowc_u, isoRowc_g, isoRowc_r, isoRowc_i, isoRowc_z, isoRowcErr_u, isoRowcErr_g, isoRowcErr_r, isoRowcErr_i, isoRowcErr_z, isoRowcGrad_u, isoRowcGrad_g, isoRowcGrad_r, isoRowcGrad_i, isoRowcGrad_z , isoColc_u, isoColc_g, isoColc_r, isoColc_i, isoColc_z, isoColcGrad_u, isoColcGrad_g, isoColcGrad_r, isoColcGrad_i, isoColcGrad_z , isoA_u, isoA_g, isoA_r, isoA_i, isoA_z, isoAErr_u, isoAErr_g, isoAErr_r, isoAErr_i, isoAErr_z, isoB_u, isoB_g, isoB_r, isoB_i, isoB_z, isoBErr_u, isoBErr_g, isoBErr_r, isoBErr_i, isoBErr_z, isoAGrad_u, isoAGrad_g, isoAGrad_r, isoAGrad_i , isoAGrad_z, isoBGrad_u, isoBGrad_g, isoBGrad_r, isoBGrad_i, isoBGrad_z, isoPhi_u, isoPhi_g, isoPhi_r, isoPhi_i, isoPhi_z , isoPhiErr_u, isoPhiErr_g, isoPhiErr_r, isoPhiErr_i, isoPhiErr_z, isoPhiGrad_u, isoPhiGrad_g, isoPhiGrad_r, isoPhiGrad_i, isoPhiGrad_z, deVRad_u, deVRad_g, deVRad_r, deVRad_i, deVRad_z, deVRadErr_u, deVRadErr_g, deVRadErr_r, deVRadErr_i, deVRadErr_z, deVAB_u, deVAB_g, deVAB_r, deVAB_i, deVAB_z, deVABErr_u, deVABErr_g, deVABErr_r, deVABErr_i, deVABErr_z , deVPhi_u, deVPhi_g, deVPhi_r, deVPhi_i, deVPhi_z, deVPhiErr_u, deVPhiErr_g, deVPhiErr_r, deVPhiErr_i , deVPhiErr_z, deVMag_u, deVMag_g, deVMag_r, deVMag_i, deVMag_z , deVMagErr_u, deVMagErr_g, deVMagErr_r, deVMagErr_i, deVMagErr_z, expRad_u, expRad_g, expRad_r, expRad_i, expRad_z, expRadErr_u, expRadErr_g, expRadErr_r, expRadErr_i, expRadErr_z , expAB_u, expAB_g, expAB_r, expAB_i, expAB_z, expABErr_u, expABErr_g, expABErr_r, expABErr_i, expABErr_z, expPhi_u, expPhi_g, expPhi_r, expPhi_i, expPhi_z, expPhiErr_u, expPhiErr_g, expPhiErr_r, expPhiErr_i, expPhiErr_z, expMag_u, expMag_g, expMag_r, expMag_i, expMag_z, expMagErr_u, expMagErr_g, expMagErr_r, expMagErr_i, expMagErr_z, modelMag_u, modelMag_g, modelMag_r, modelMag_i, modelMag_z, modelMagErr_u, modelMagErr_g, modelMagErr_r, modelMagErr_i, modelMagErr_z, texture_u, texture_g, texture_r, texture_i, texture_z, lStar_u, lStar_g, lStar_r, lStar_i, lStar_z, lExp_u, lExp_g, lExp_r, lExp_i, lExp_z, lDeV_u, lDeV_g, lDeV_r, lDeV_i, lDeV_z, fracPSF_u, fracPSF_g, fracPSF_r, fracPSF_i, fracPSF_z, flags_u, flags_g, flags_r, flags_i, flags_z, type_u, type_g, type_r, type_i, type_z, status, ra, dec, cx, cy, cz, offsetRa_u, offsetRa_g, offsetRa_r, offsetRa_i, offsetRa_z, offsetDec_u, offsetDec_g, offsetDec_r, offsetDec_i, offsetDec_z, primTarget, secTarget, reddening_u, reddening_g, reddening_r, reddening_i, reddening_z, priority, rho, htmID, fieldID, u, g, r, i, z, Err_u, Err_g, Err_r, Err_i, Err_z

→ x,y

SOLUTIONS

- ADD MORE DIMENSIONS TO DISPLAY

Color, Transparency
Motion

- LOWER THE NUMBER OF DIMENSIONS

Restrict the original dimensions (subsetting)
Create new dimensions (reduction)

- PROJECT FROM DIFFERENT VIEWPOINT-S
- USE ADVANCED PROJECTION METHODS
- LINK DIFFERENT VIEWPORTS

QUICK OVERVIEW

SUBSETTING

EMBEDDING

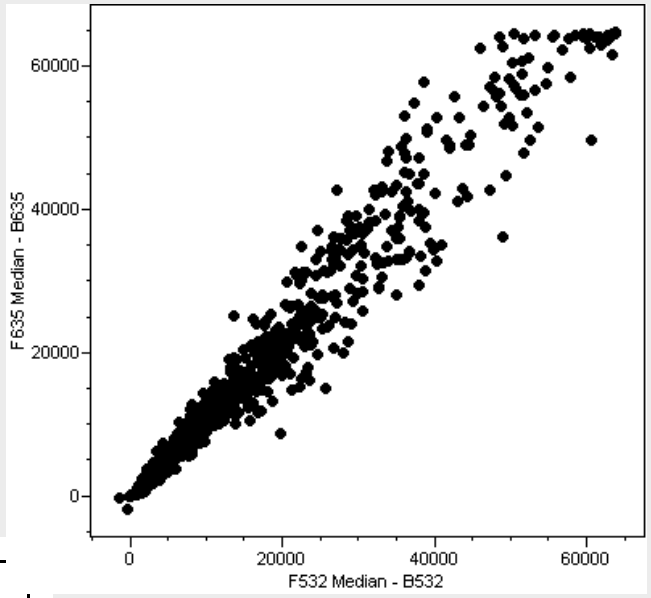
REDUCTION

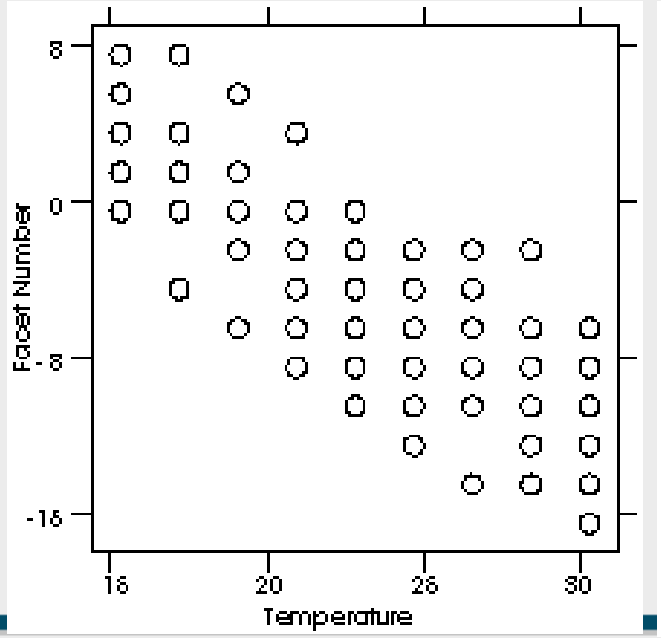
AXIS RECONFIGURATION

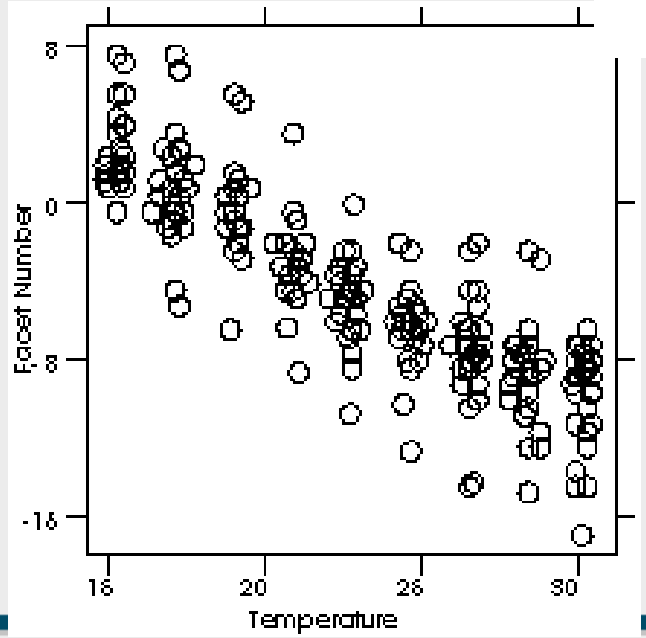
DIMENSIONAL SUBSETTING

SUBSETTING - SCATTERPLOT

- ⊕ SIMPLE AND INTUITIVE
- ⊖ DISCARDS (*N*-2) DIMENSIONS
- ⊖ OVERPLOTTING





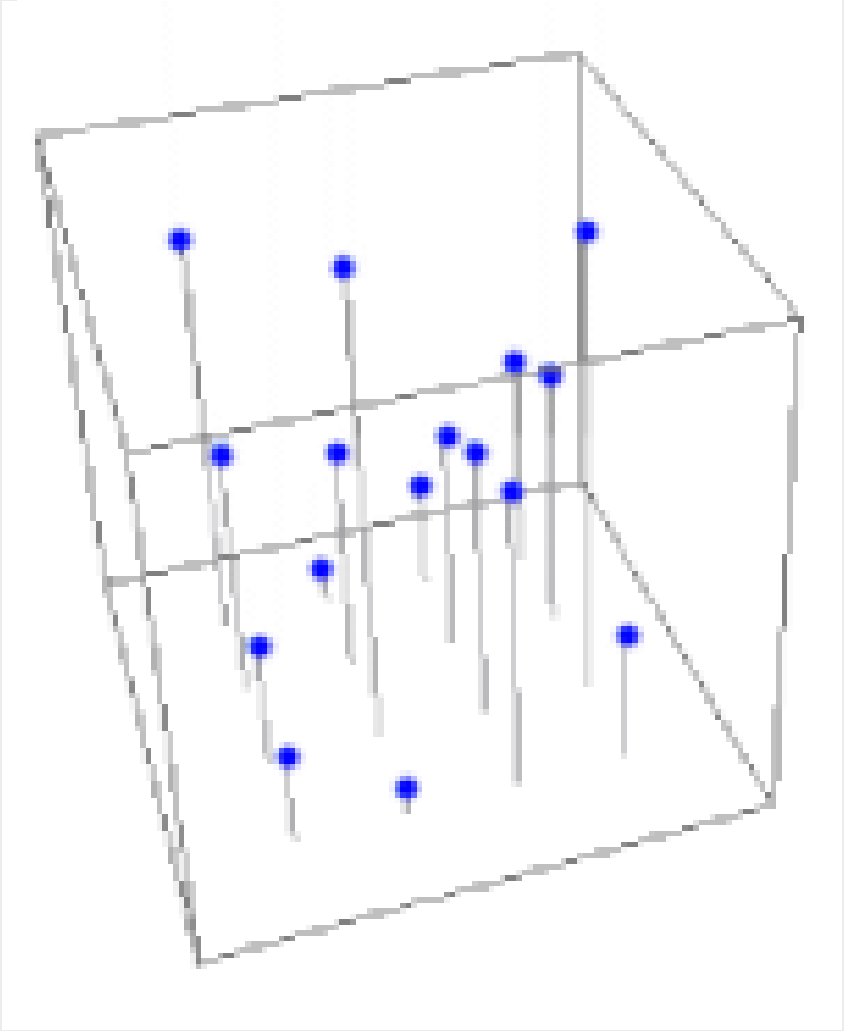


SUBSETTING - 3D SCATTERPLOT

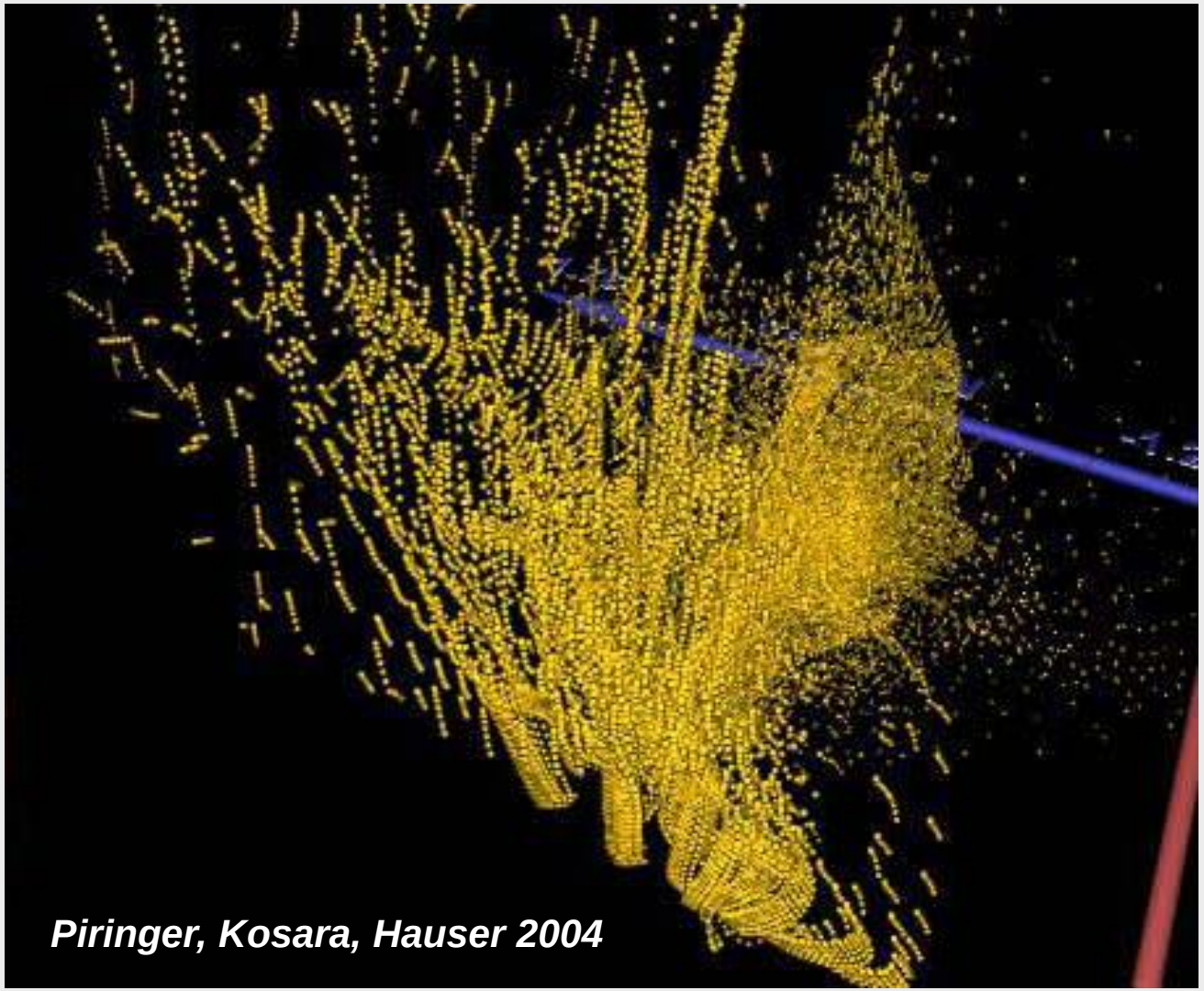
- ⊕ 1 MORE DIMENSION
- ⊖ 3D IN 2D

WAYS TO IMPROVE:

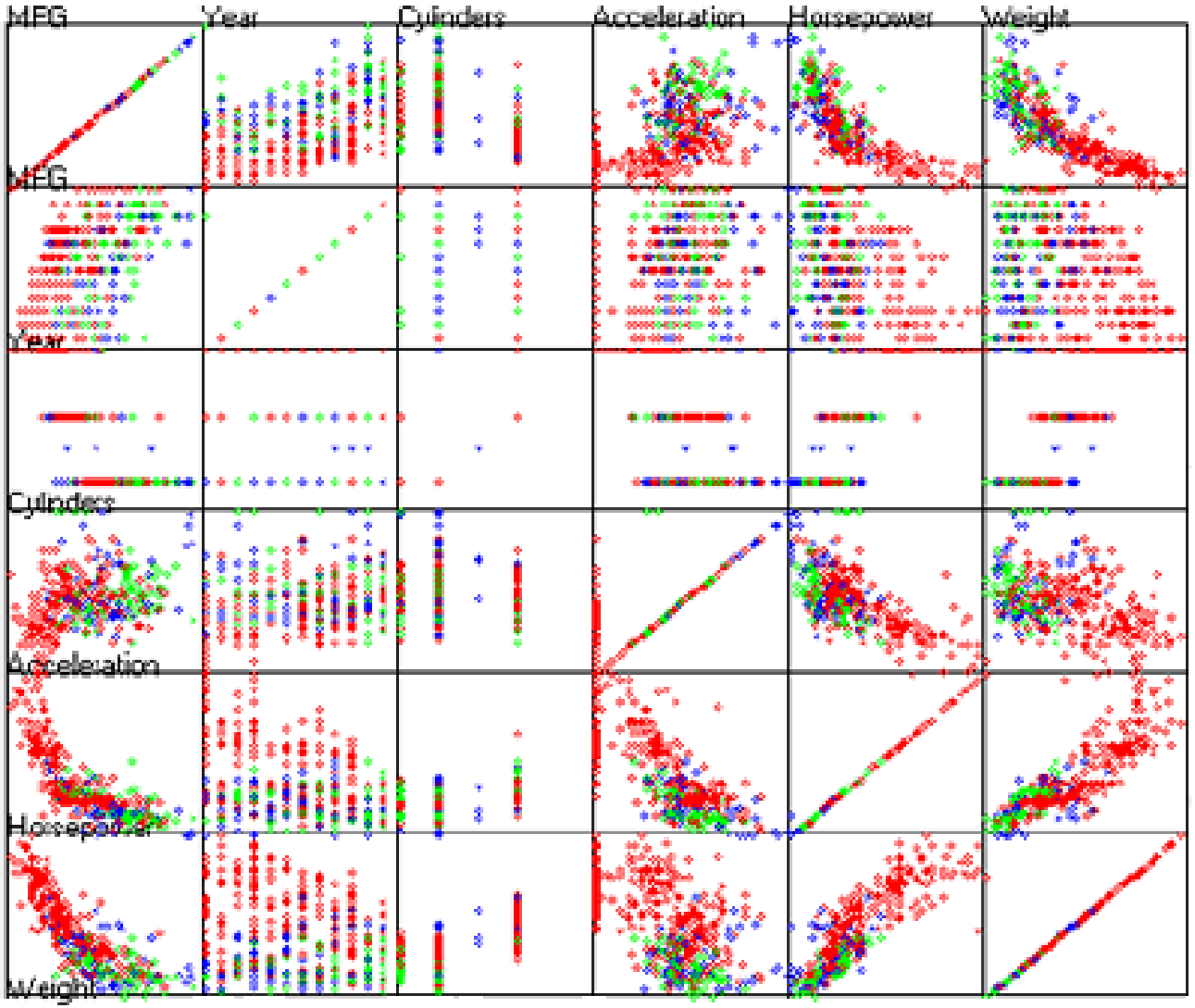
- Perspective
- Height lines
- Halos
- Blur



SUBSETTING - 3D SCATTERPLOT



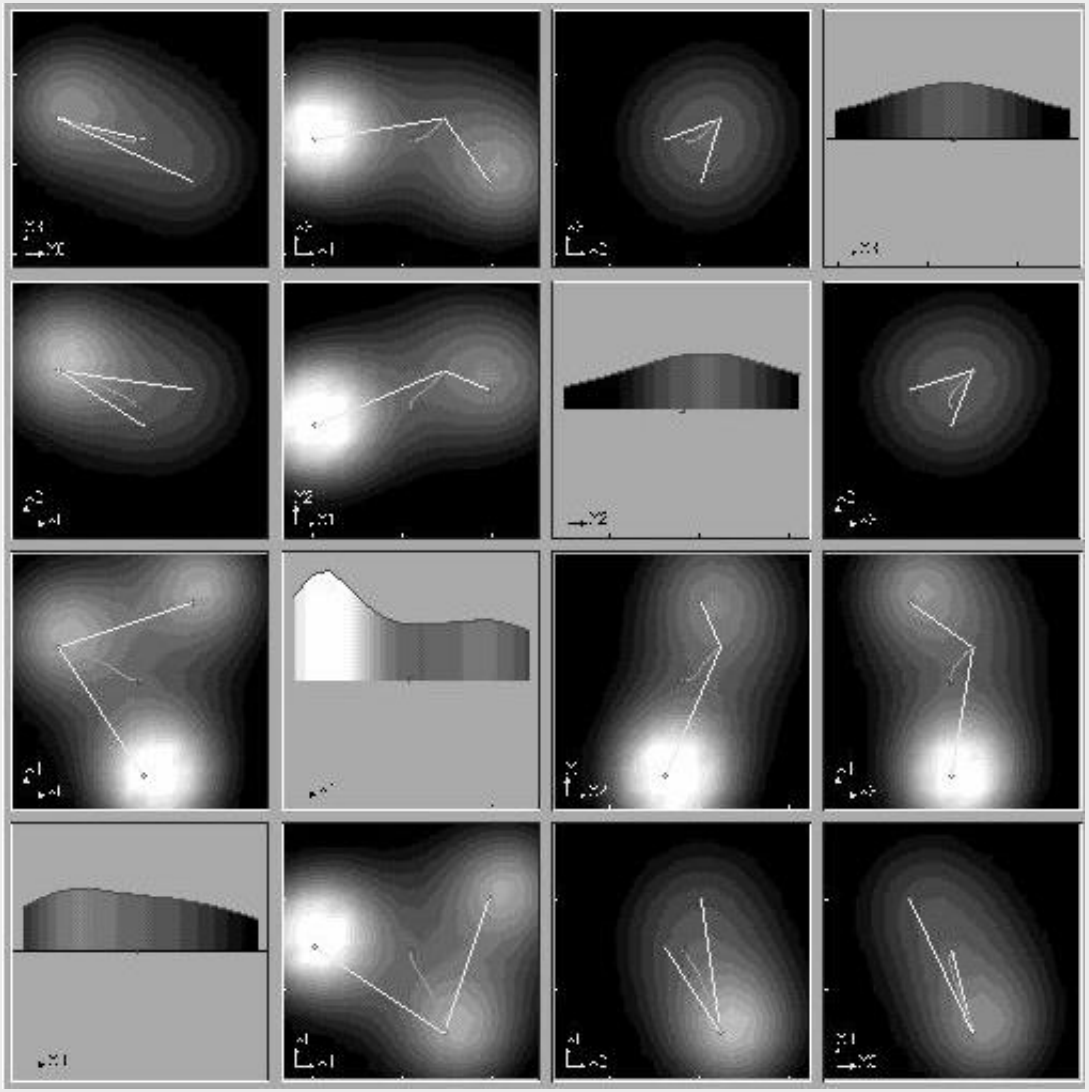
SCATTERPLOT MATRIX



HYPERSLICE

SCALAR
FUNCTION
IN
*N*D SPACE

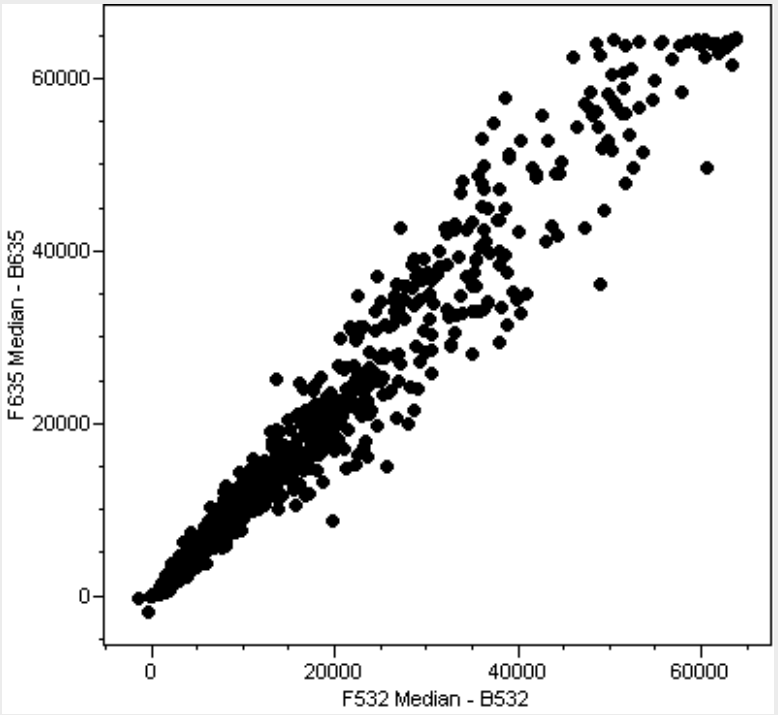
2D SLICES
1D HISTOGRAMS

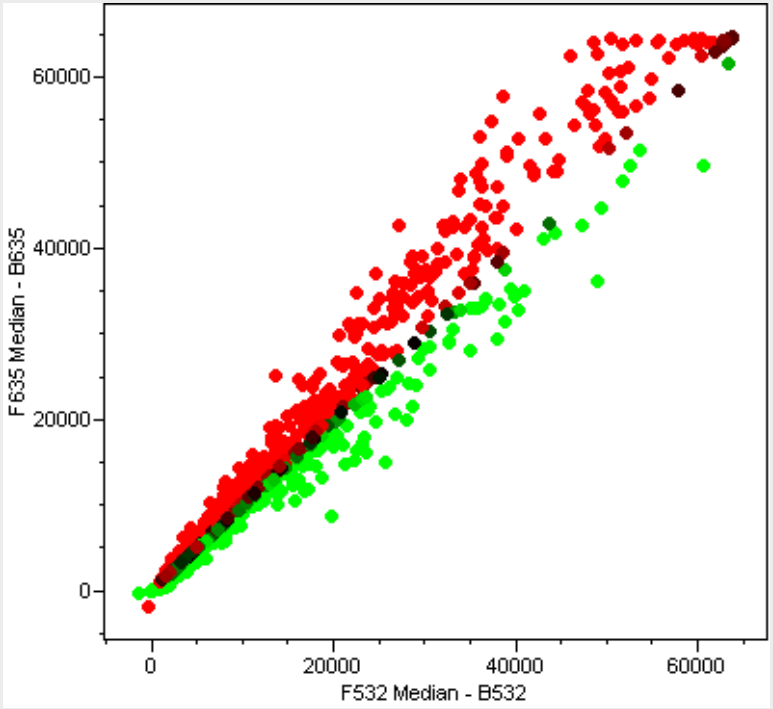


DIMENSIONAL EMBEDDING

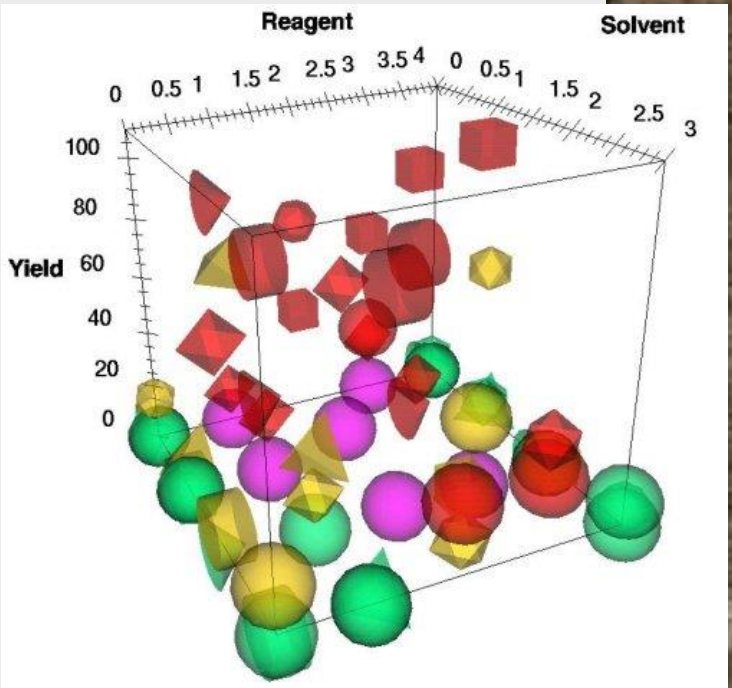
DIMENSIONAL EMBEDDING

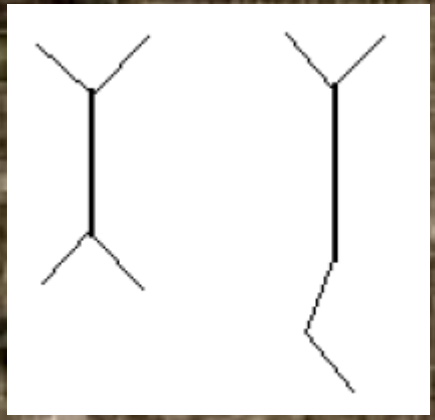
NATURALLY EMBEDDED DIMENSIONS:
Color, motion





DIMENSIONAL EMBEDDING – GLYPHS

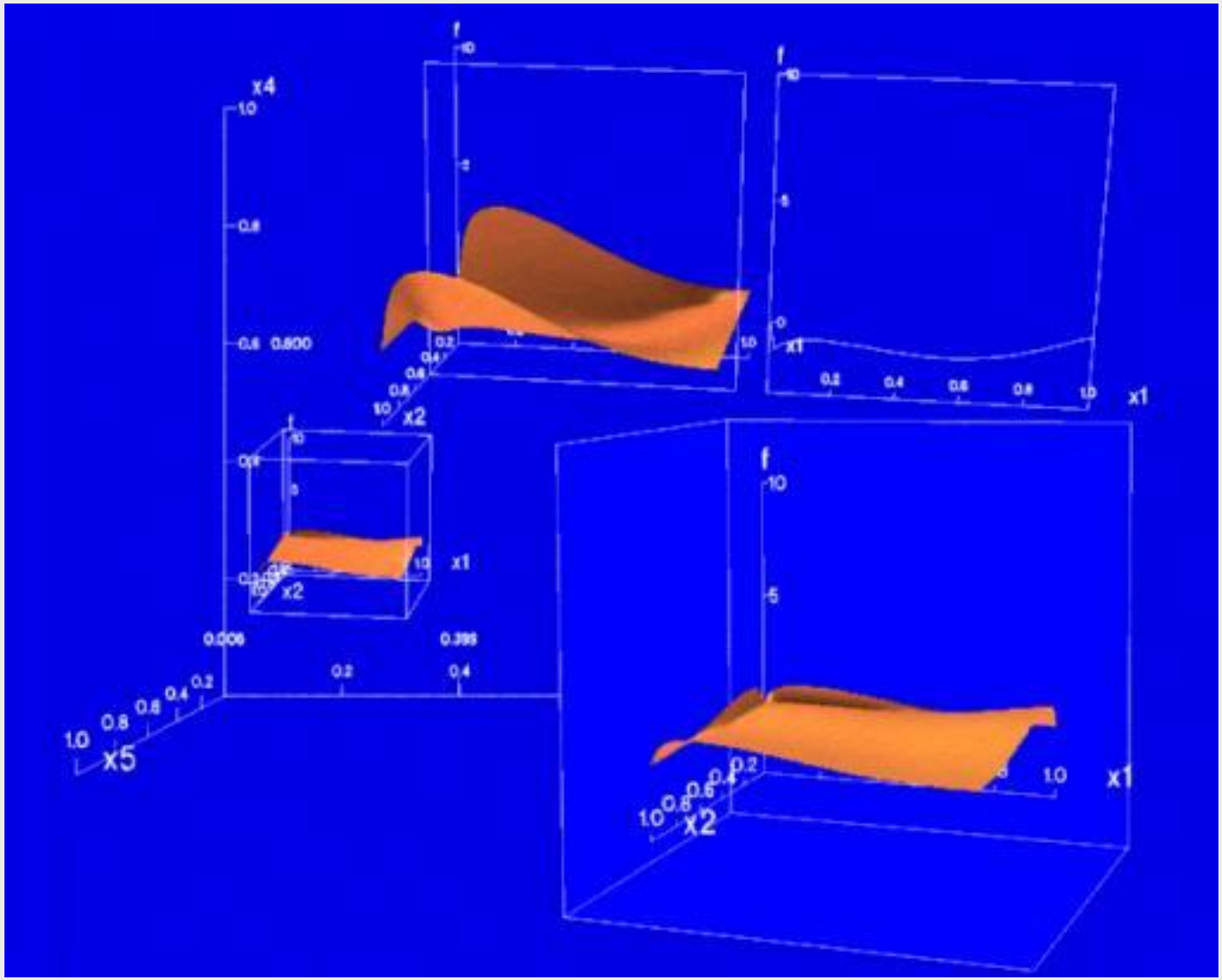




DIMENSIONAL EMBEDDING

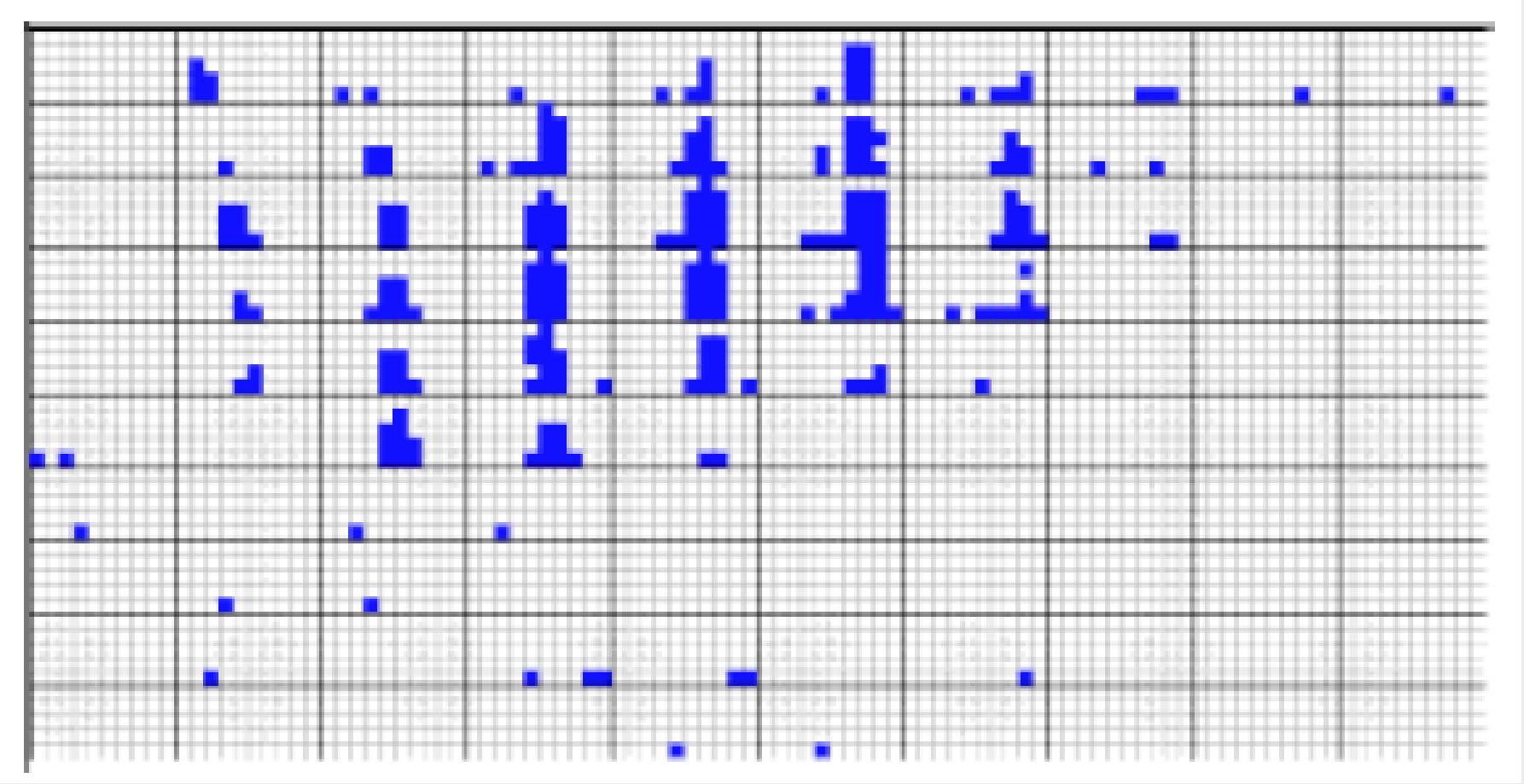
WORLDS
WITHIN
WORLDS

THE GLYPH
IS A WHOLE
NEW
GRAPH



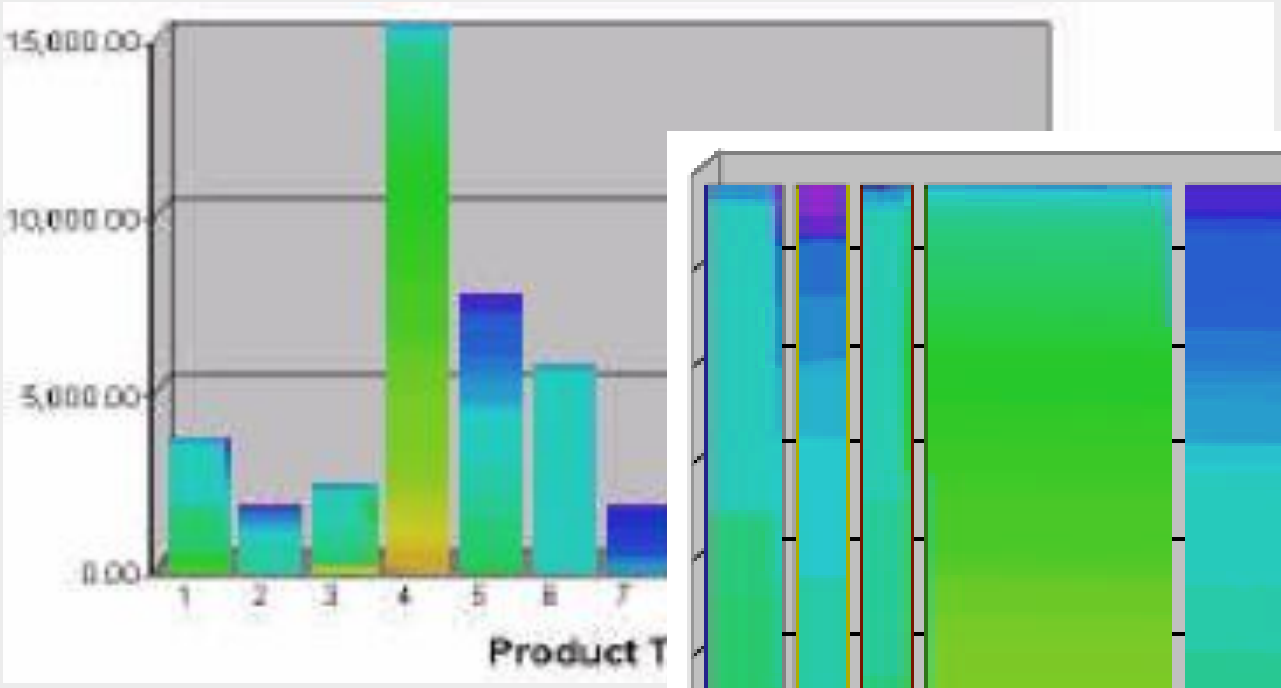
DIMENSIONAL STACKING

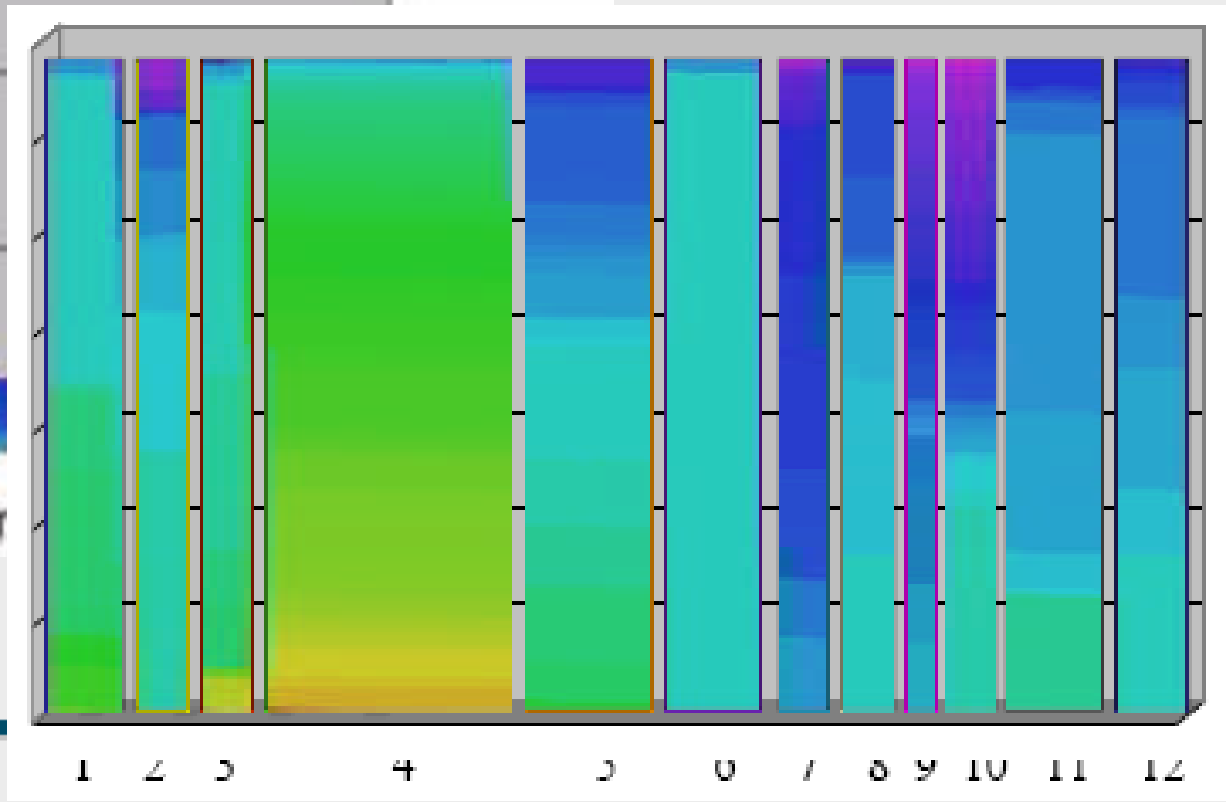
DISCRETIZE THE OUTER 2D SPACE
EMBED INNER 2D SPACES



PIXEL-ORIENTED TECHNIQUES

DISCRETIZATION @ 1 PIXEL
COLOR USED TO EMBED +1 DIMENSION





DIMENSION REDUCTION

DIMENSION REDUCTION

$N\text{D} \rightarrow P\text{D}, P < N$

P_i DON'T NEED TO BE MEMBERS OF $\{ N_i \}$

THINGS TO PRESERVE:

Relations - groups, trends, outliers

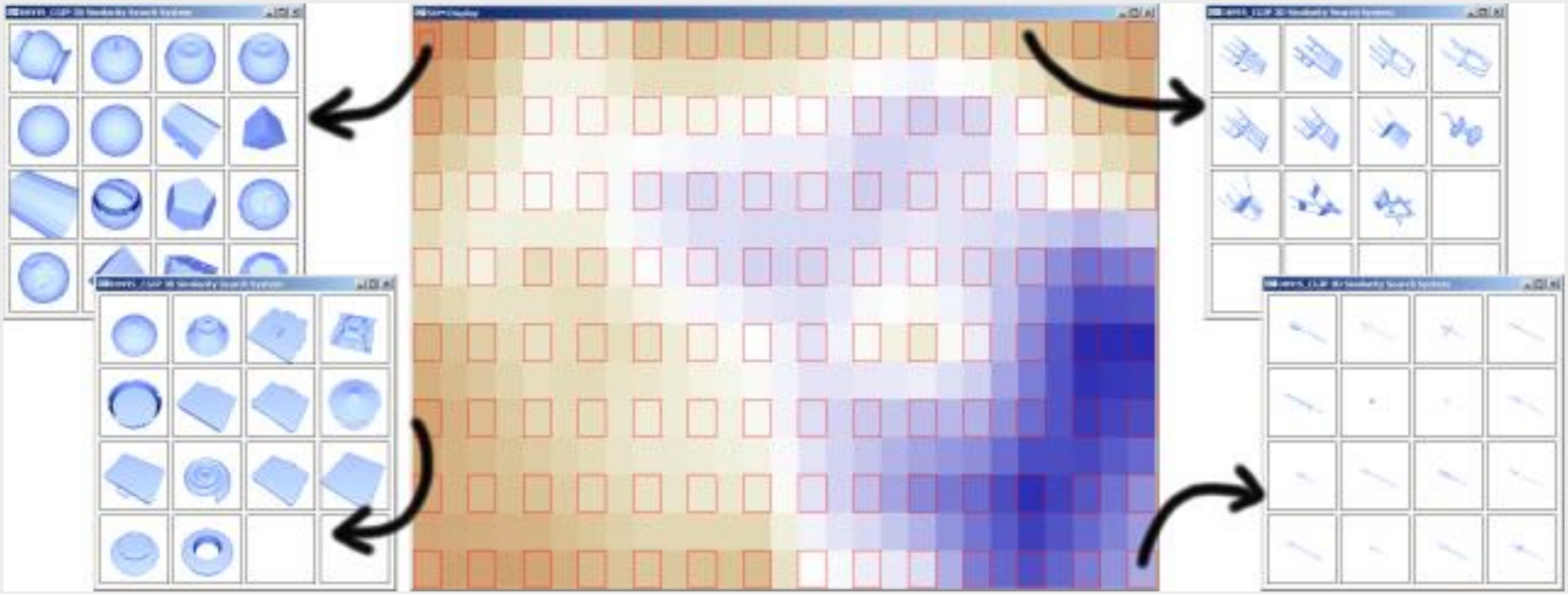
Topology

Relative measures

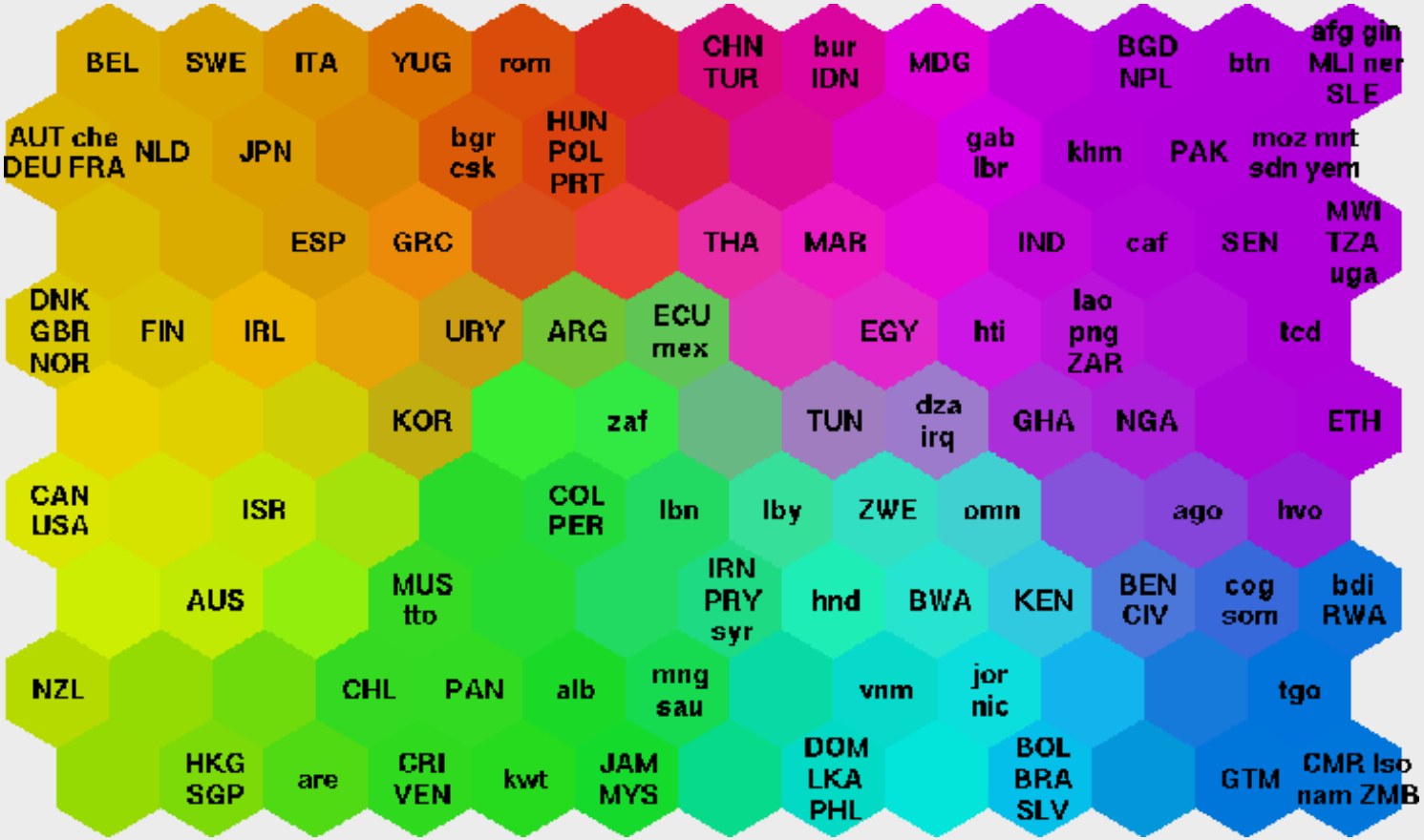
SELF ORGANIZING MAPS

NEURAL NETWORK, UNSUPERVISED LEARNING

SMALL SET OF NEURONS
(prototype vectors)



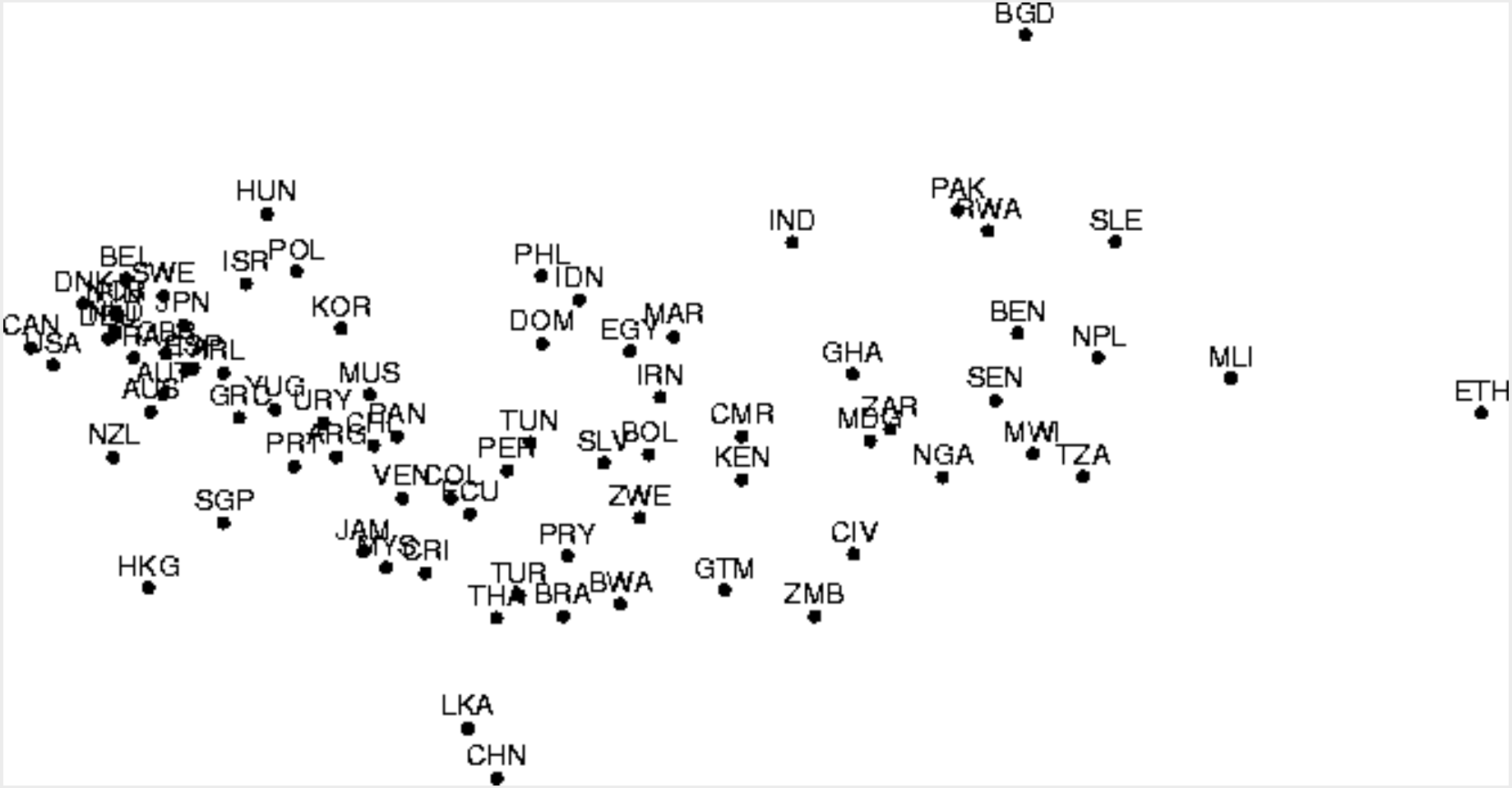
SOM – WORLD POVERTY MAP



39 DIMENSIONS: HEALTH, EDUCATION, NUTRITION, STANDARD OF LIVING,...

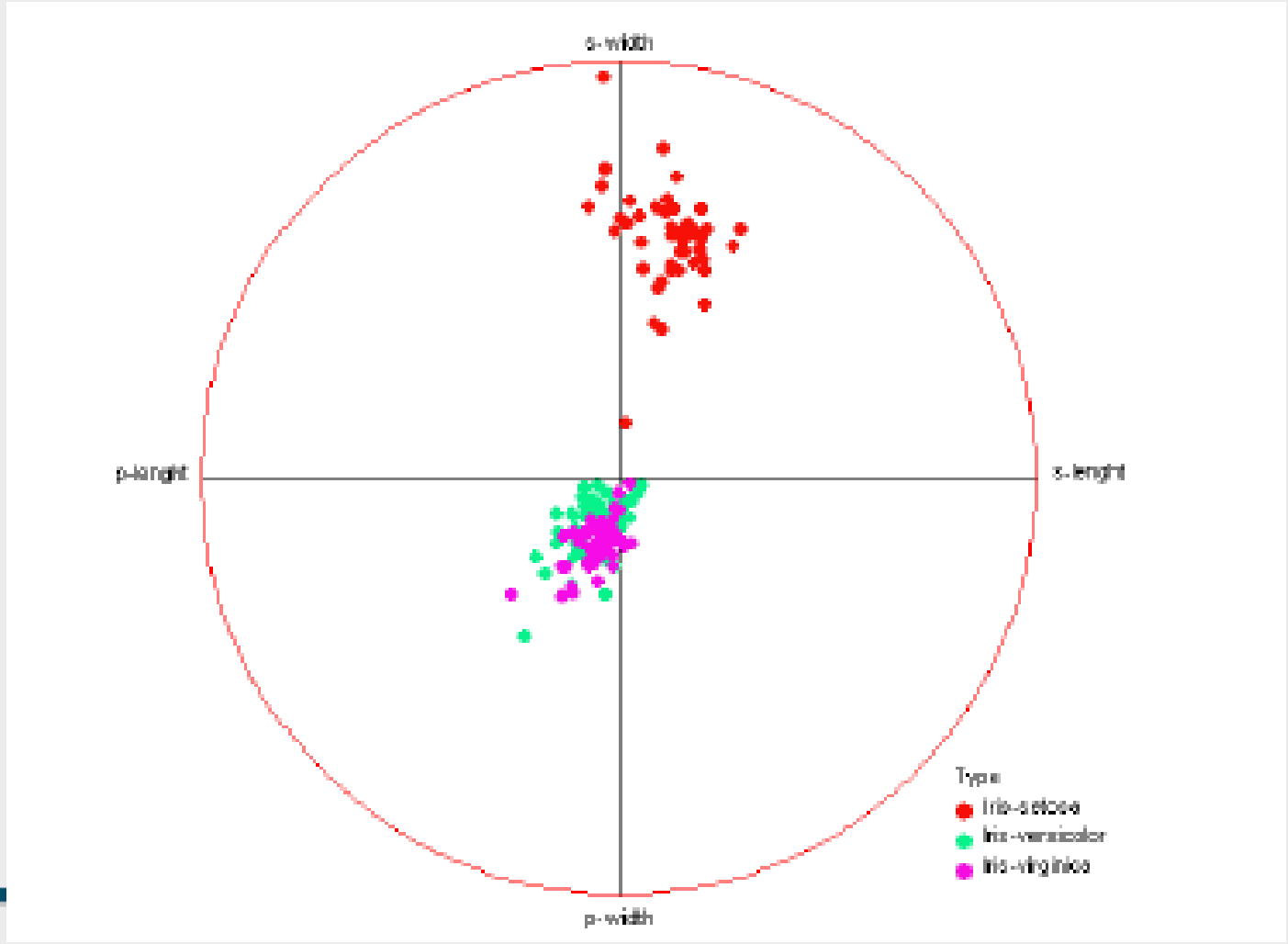
MULTIDIMENSIONAL SCALING

SPRING MODEL (FORCE-DIRECTED LAYOUT)



RADVIZ

FORCE-DIRECTED, ANCHORED SPRINGS



PITFALLS OF DIMENSION REDUCTION

NEW COORDINATE SYSTEM HAS NO SEMANTICS

Self-organizing maps
Multidimensional scaling,
Force-directed layouts
Vector quantization

DISTANCES IN *N*D MEANINGLESS FOR BIG *N*

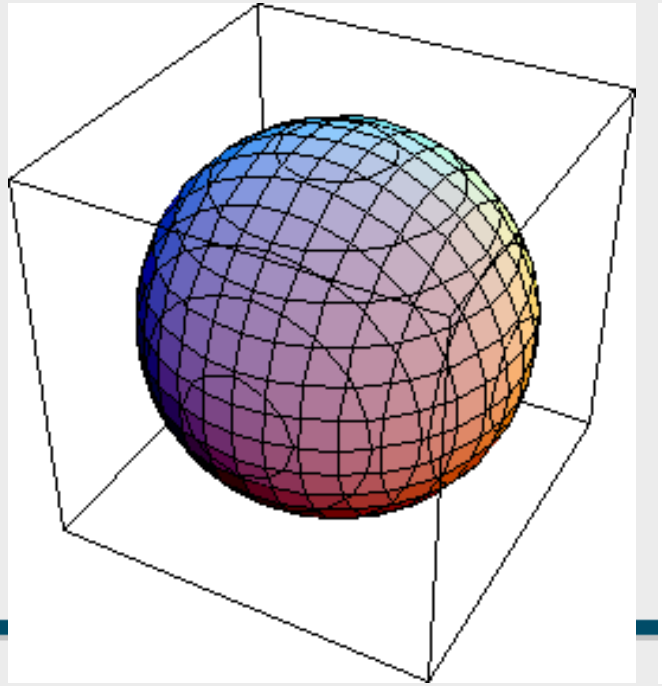
Curse of dimensionality
Help: e.g. *k* nearest neighbors

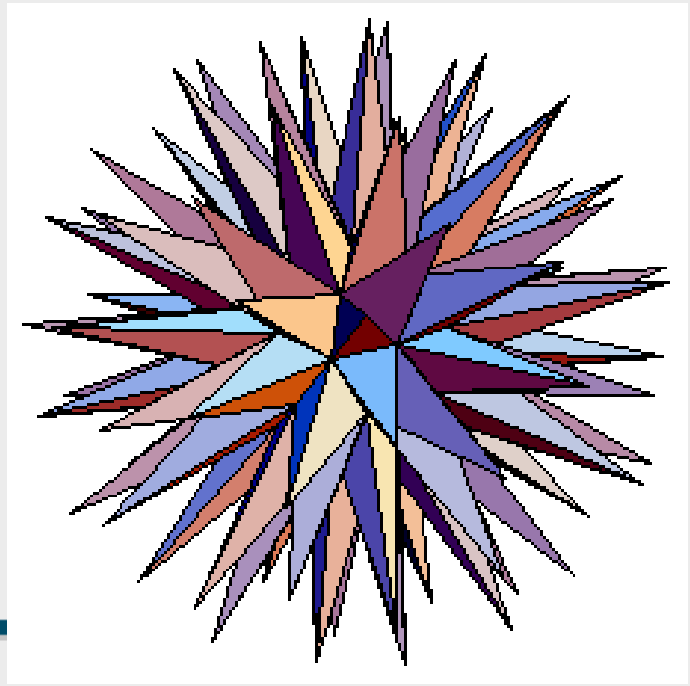
CURSE OF DIMENSIONALITY

UNIT SPHERE INSIDE UNIT BOX

2D sphere = 78%, 3D sphere = 52%,
4D sphere = 31%, 10D sphere = 0.2%

IN *N*D, EVERYTHING IS FAR AWAY...





DISTANCE MEASURES IN *N*D

EUCLID DISTANCE
Intuitive

MANHATTAN DISTANCE
Fast computation

MAHALANOBIS DISTANCE
Uses dimension covariance

K-NEAREST NEIGHBORS
Works for large *n*

SUMMARY

GENERAL PROBLEM: *N*D → 2D

SUBSETTING
EMBEDDING
REDUCTION

Next lesson:
AXIS RECONFIGURATION