ANALYSE MULTI-TABLEAUX:

LA FAMILLE STATIS

Hervé Abdi

The University of Texas at Dallas

Paris CNAM November 2011

Part two: La Famille STATIS

CNAM: 7 NOVEMBRE 2011

La Famille STATIS

Biblio. Allez voir à www.utdallas.edu/~herve

A87: STATIS & DISTATIS

A71, A59: DISTATIS

C71: Rv Coefficient

C40: Multiple Factor Analysis

C33: STATIS

CNAM: 7 NOVEMBRE 2011

The STATIS Family

START WITH *K* TABLES

The STATIS Family

START WITH *K* TABLES

- Center and Normalize
  (or not) columns (almost always)

BEFORE STATIS

- Normalize (or not) rows (rarely)

  exception: CA

  → sum of $x = 1$

  Escofier/Volle/Rao/Hellinger (etc.)

  → sum of $x^2 = 1$

BEFORE STATIS

- What about the tables?

BEFORE STATIS

Normalizing Tables: Why?

- Divide all elements of $\mathbf{X}_k$ by J_k

HOW TO NORMALIZE TABLE *K*.

- Divide all elements of $\mathbf{X}_k$ by J_k or better by $J_k^{½}$

- Plain multi-block ➔ Tucker 1 & consensus PCA

HOW TO NORMALIZE TABLE *K*.

- Divide all elements of $\mathbf{X}_k$ by $(sum\ \mathbf{X}_k)^{½}$

- Plain multi-block → SUM-PCA

How to normalize Table *k*.

- Divide all elements of $\mathbf{X}_k$ by $(sum\ \mathbf{X}_k\mathbf{X}_k^T)^{½}$

- Plain multi-block → R_V-PCA

HOW TO NORMALIZE TABLE *K*.

- Divide all elements of $\mathbf{X}_k$ by first singular value

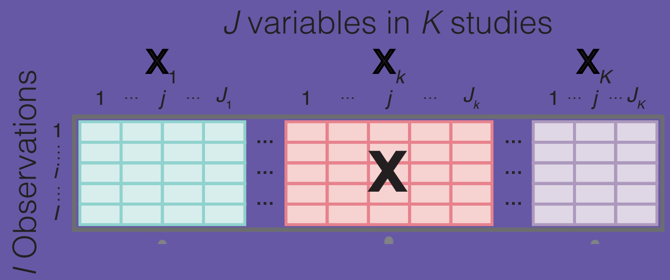
- Plain multi-block → Multiple Factor Analysis

How to normalize Table *k*.

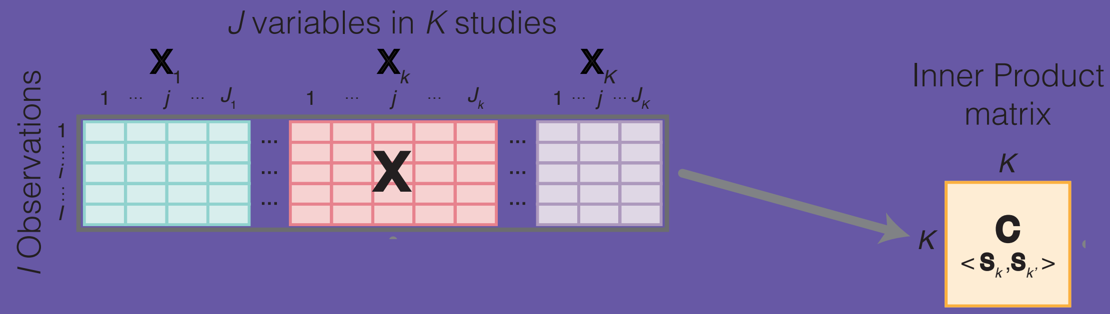
Table normalization: Most important step!

- STATIS: 1,2, 3 …

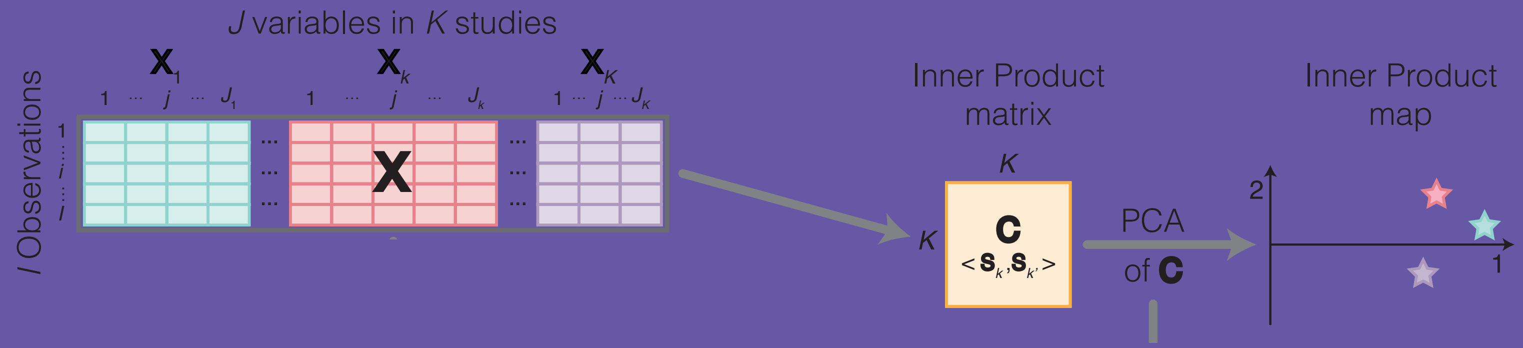
WHEN THE TABLES ARE NORMALIZED: WHAT TO DO?



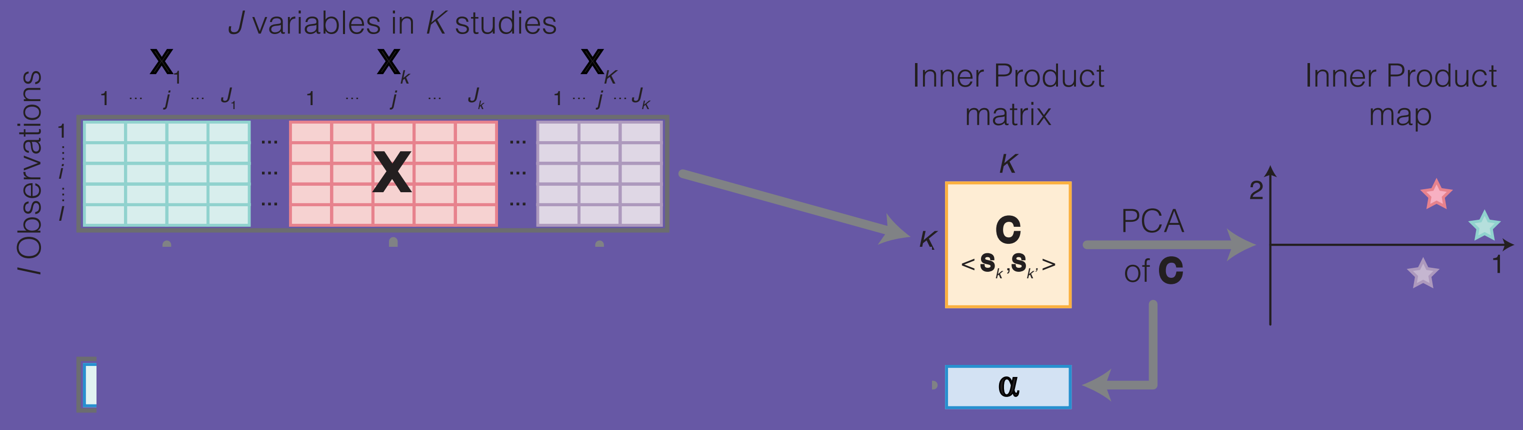
START WITH A MULTI TABLE MATRIX



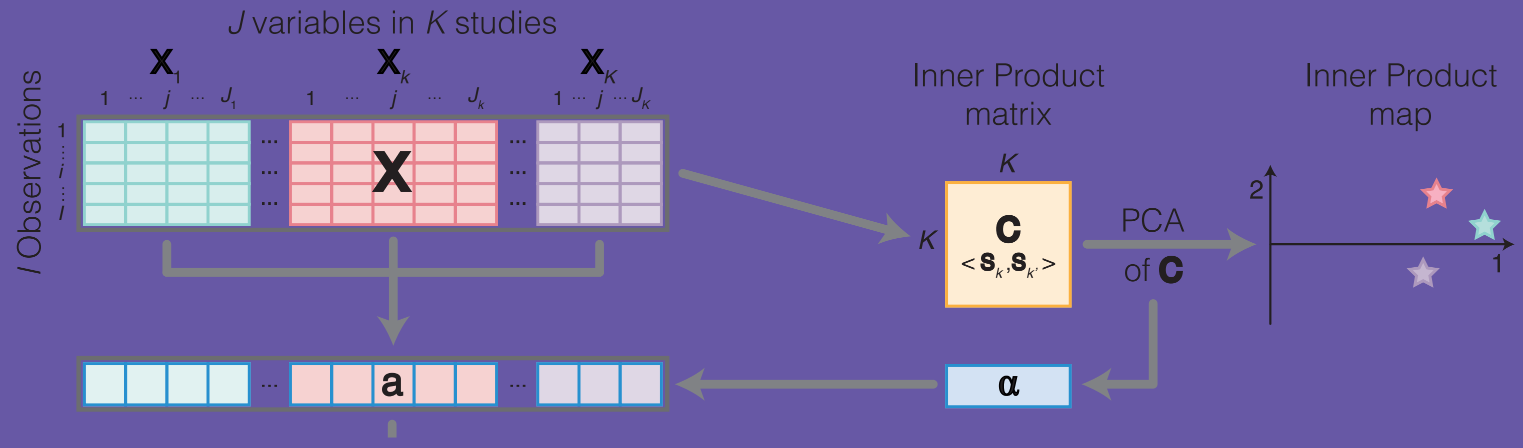
Compute the between table similarity



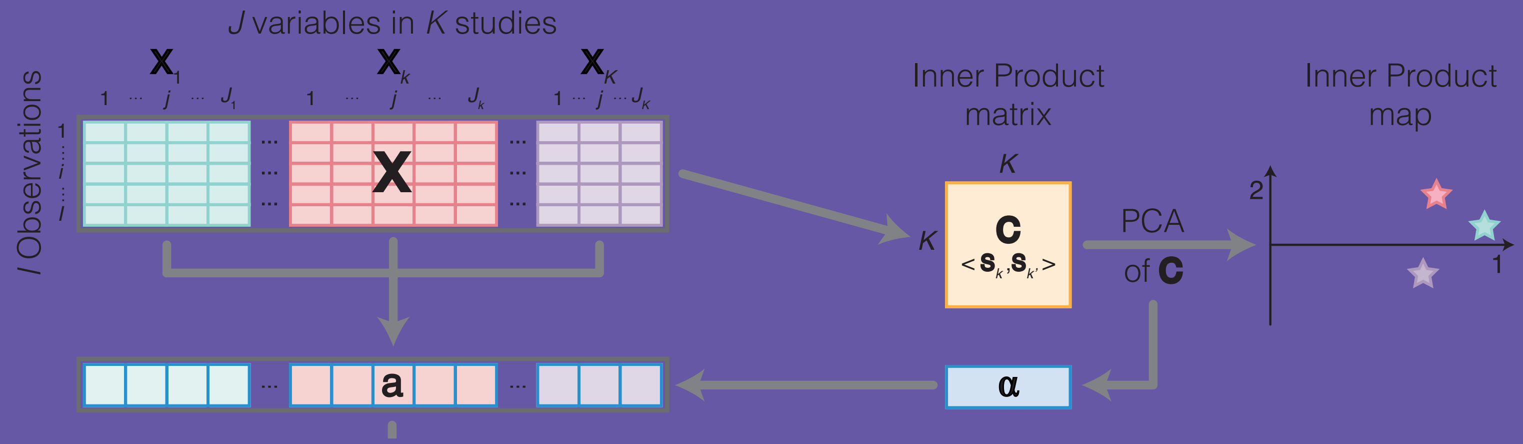
Get a PCA of the between table similarity



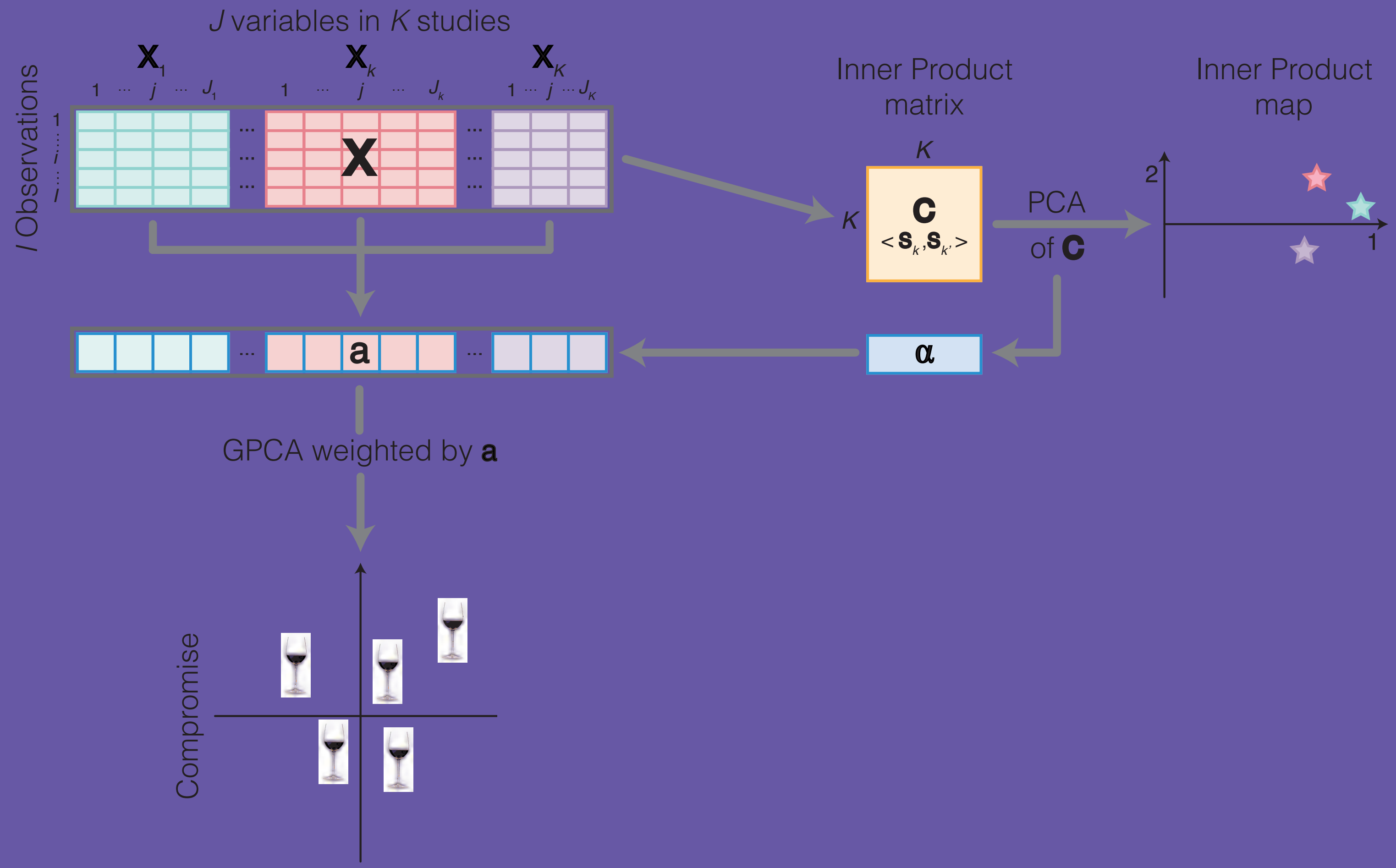
PCA OF C GIVES OPTIMAL ALPHA WEIGHTS



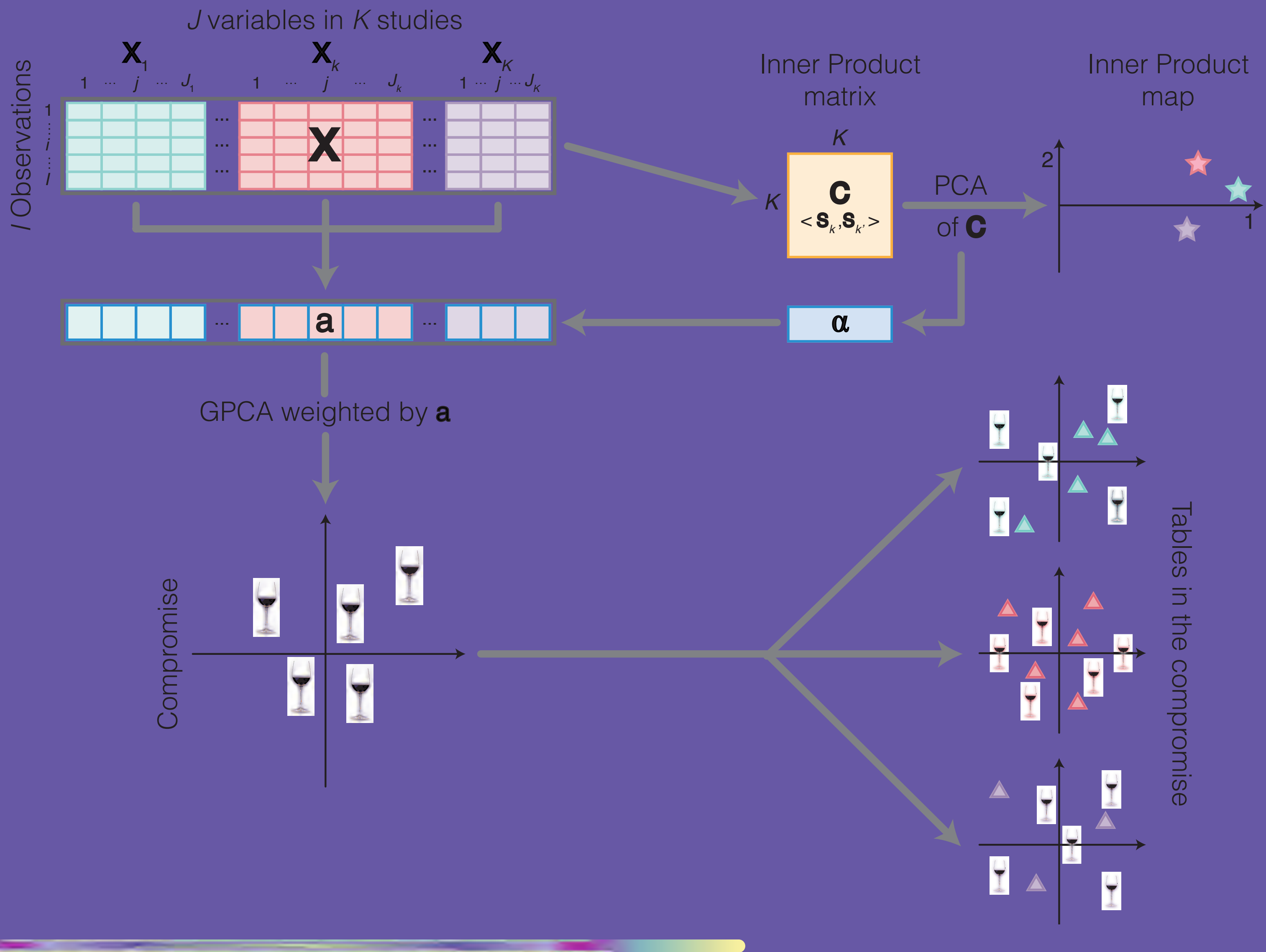
ALPHA WEIGHTS ARE USED FOR GPCA OF X



START WITH A MULTI TABLE MATRIX



GPCA OF X → FACTOR SCORES F (COMPROMISE)



PROJECT THE X_k ON COMPROMISE → F_k

An Example

- Sauvignon Blanc Wines
- From New-Zealand, France, and Canada
- Chemical/Physical measurements
- Specific scales + Four commons scales:
  cat-pee, passion, green pepper, mineral

EXAMPLE: 10 PARTICIPANTS TASTE 3*4 =12 WINES

- 1. Between Table Structure
- 2. Derive Optimal Weights
- 3. Compute Compromise from Weights
- 4. Eigen-decompose Compromise
- 5. Project Original Tables (factor scores)
- 6. … *Is the Earth round? …*

REMEMBER: THE STEPS OF STATIS

- 1. Between Table Structure

THE STEPS OF STATIS

	Assessor 1						Assessor 2						Assessor 3						Assessor 4					Assessor 5					
	V1	V2	V3	V4	V5	V6	V1	V2	V3	V4	V7	V8	V1	V2	V3	V4	V9	V10	V1	V2	V3	V4	V8	V1	V2	V3	V4	V11	V12
NZ_1	8	6	7	4	1	6	8	6	8	3	7	5	8	6	8	3	7	2	9	5	8	2	6	9	6	9	3	8	2
NZ_2	7	5	8	1	2	8	6	5	6	3	7	7	8	7	7	2	8	2	8	7	7	3	5	7	7	7	1	9	2
NZ_3	6	5	6	5	3	4	6	6	6	5	8	7	8	7	7	6	9	1	8	8	9	2	7	7	7	7	1	7	2
NZ_4	9	6	8	4	3	5	8	6	8	4	6	6	8	2	8	3	9	3	8	8	9	4	7	8	9	7	5	6	1
FR_1	2	2	2	8	7	3	2	3	1	7	4	3	3	4	3	6	4	6	4	2	2	4	3	4	4	4	2	4	4
FR_2	3	4	4	9	6	1	4	3	4	9	3	5	4	3	4	8	3	9	3	2	2	6	2	4	5	5	6	1	5
FR_3	5	3	5	4	8	3	3	3	2	7	4	4	5	4	5	2	3	6	4	4	4	6	4	6	5	7	2	3	1
FR_4	5	2	4	8	7	4	4	3	5	5	3	3	6	3	7	7	1	7	5	2	2	9	4	6	6	5	8	4	5
CA_1	8	6	8	4	4	7	8	6	9	5	5	6	8	5	9	1	5	2	7	5	6	3	2	8	6	8	2	5	4
CA_2	4	6	2	5	3	4	5	5	5	6	5	8	5	5	4	6	5	1	5	6	6	4	4	6	6	6	4	6	3
CA_3	8	4	8	1	3	3	8	4	8	3	7	7	8	3	7	3	5	4	7	3	6	1	6	7	4	8	4	5	1
CA_4	5	3	6	4	4	2	5	3	7	4	8	5	5	4	4	5	4	3	5	2	2	6	6	5	5	5	5	6	1

	Assessor 6					Assessor 7				Assessor 8						Assessor 9					Assessor 10			
	V1	V2	V3	V4	V13	V1	V2	V3	V4	V1	V2	V3	V4	V14	V5	V1	V2	V3	V4	V15	V1	V2	V3	V4
NZ_1	8	5	6	2	9	8	5	8	4	7	6	7	4	9	2	8	6	9	1	7	8	6	7	5
NZ_2	6	6	6	2	4	7	6	8	4	6	5	6	2	7	2	8	7	9	1	6	7	5	7	3
NZ_3	7	7	7	2	7	6	7	6	3	6	6	6	4	9	2	7	7	8	4	7	7	6	6	2
NZ_4	8	7	8	2	8	7	8	6	1	8	7	8	2	8	2	8	9	9	3	9	8	7	7	4
FR_1	3	2	2	7	2	4	2	3	6	3	3	4	4	4	4	3	4	4	5	4	2	3	1	7
FR_2	3	3	3	3	4	4	4	4	4	4	4	4	7	3	6	5	5	5	7	2	3	3	3	9
FR_3	4	2	3	3	3	4	3	4	4	5	3	5	3	3	5	5	5	5	6	3	4	2	5	8
FR_4	5	3	5	9	3	5	3	5	7	6	4	6	3	2	4	5	5	6	5	3	3	4	2	8
CA_1	7	7	7	1	4	8	4	9	4	8	6	5	4	5	4	8	7	8	4	7	8	6	7	4
$CA_2$4	6	2	4	6	4	7	5	2	5	7	5	4	6	1	5	6	4	5	6	5	6	4	4	
CA_3	7	4	8	2	3	8	5	7	3	7	4	8	2	6	2	8	4	7	4	5	7	4	8	5
CA_4	4	5	3	3	7	4	3	5	2	5	4	6	2	4	3	5	4	5	3	4	5	4	6	6

The Data: 10 Assessors by 3*4 = 12 wines

	Chemical Properties			
	Titratable Acidity	pH	Alcohol	Residual Sugar
NZ_1	5.60	3.38	14.00	3.00
NZ_2	5.30	3.53	13.50	3.60
NZ_3	6.20	3.27	14.00	3.00
NZ_4	8.50	3.19	13.50	3.90
F_1	5.00	3.60	12.50	1.50
F_2	5.88	3.00	12.50	2.00
F_3	4.50	3.33	13.00	0.80
F_4	5.60	3.40	12.00	2.10
CA_1	7.60	3.30	13.00	2.80
CA_2	5.70	3.43	13.50	2.10
CA_3	6.20	3.30	12.50	2.50
CA_4	6.90	2.20	13.00	2.00

Supplementary table: Chemistry

$$\mathbf{Y}_{[1]} = \begin{bmatrix} 8 & 6 & 7 & 4 & 1 & 6 \\ 7 & 5 & 8 & 1 & 2 & 8 \\ 6 & 5 & 6 & 5 & 3 & 4 \\ 9 & 6 & 8 & 4 & 3 & 5 \\ 2 & 2 & 2 & 8 & 7 & 3 \\ 3 & 4 & 4 & 9 & 6 & 1 \\ 5 & 3 & 5 & 4 & 8 & 3 \\ 5 & 2 & 4 & 8 & 7 & 4 \\ 8 & 6 & 8 & 4 & 4 & 7 \\ 4 & 6 & 2 & 5 & 3 & 4 \\ 8 & 4 & 8 & 1 & 3 & 3 \\ 5 & 3 & 6 & 4 & 4 & 2 \end{bmatrix}.$$

A matrix: Assessor 1

$$
\mathbf{X}_{[1]} = \begin{bmatrix}
0.12 & 0.13 & 0.07 & -0.04 & -0.18 & 0.11 \\
0.07 & 0.05 & 0.13 & -0.18 & -0.12 & 0.23 \\
0.01 & 0.05 & 0.02 & 0.01 & -0.07 & -0.01 \\
0.18 & 0.13 & 0.13 & -0.04 & -0.07 & 0.05 \\
-0.21 & -0.18 & -0.20 & 0.16 & 0.15 & -0.07 \\
-0.16 & -0.03 & -0.09 & 0.21 & 0.10 & -0.19 \\
-0.05 & -0.11 & -0.04 & -0.04 & 0.21 & -0.07 \\
-0.05 & -0.18 & -0.09 & 0.16 & 0.15 & -0.01 \\
0.12 & 0.13 & 0.13 & -0.04 & -0.01 & 0.17 \\
-0.10 & 0.13 & -0.20 & 0.01 & -0.07 & -0.01 \\
0.12 & -0.03 & 0.13 & -0.18 & -0.07 & -0.07 \\
-0.05 & -0.11 & 0.02 & -0.04 & -0.01 & -0.13
\end{bmatrix}
$$

Pre-Processed (center, Sum of $x^2 = 1$)

$$\mathbf{S}_{[1]} = \mathbf{X}_{[1]}\mathbf{X}_{[1]}^{T}$$

$$= \begin{bmatrix}
0.08 & 0.08 & 0.02 & 0.07 & -0.11 & -0.08 & -0.07 & -0.07 & 0.06 & 0.00 & 0.03 & -0.03 \\
0.08 & 0.13 & 0.01 & 0.06 & -0.11 & -0.12 & -0.05 & -0.07 & 0.08 & -0.02 & 0.05 & -0.03 \\
0.02 & 0.01 & 0.01 & 0.01 & -0.02 & -0.01 & -0.02 & -0.02 & 0.01 & 0.01 & 0.01 & -0.00 \\
0.07 & 0.06 & 0.01 & 0.07 & -0.11 & -0.07 & -0.04 & -0.06 & 0.07 & -0.02 & 0.04 & -0.02 \\
-0.11 & -0.11 & -0.02 & -0.11 & 0.17 & 0.12 & 0.07 & 0.11 & -0.10 & 0.03 & -0.08 & 0.03 \\
-0.08 & -0.12 & -0.01 & -0.07 & 0.12 & 0.12 & 0.04 & 0.07 & -0.08 & 0.03 & -0.06 & 0.02 \\
-0.07 & -0.05 & -0.02 & -0.04 & 0.07 & 0.04 & 0.06 & 0.05 & -0.04 & -0.02 & -0.01 & 0.02 \\
-0.07 & -0.07 & -0.02 & -0.06 & 0.11 & 0.07 & 0.05 & 0.09 & -0.05 & -0.01 & -0.05 & 0.01 \\
0.06 & 0.08 & 0.01 & 0.07 & -0.10 & -0.08 & -0.04 & -0.05 & 0.08 & -0.02 & 0.02 & -0.04 \\
0.00 & -0.02 & 0.01 & -0.02 & 0.03 & 0.03 & -0.02 & -0.01 & -0.02 & 0.07 & -0.04 & -0.01 \\
0.03 & 0.05 & 0.01 & 0.04 & -0.08 & -0.06 & -0.01 & -0.05 & 0.02 & -0.04 & 0.07 & 0.02 \\
-0.03 & -0.03 & -0.00 & -0.02 & 0.03 & 0.02 & 0.02 & 0.01 & -0.04 & -0.01 & 0.02 & 0.03
\end{bmatrix}$$

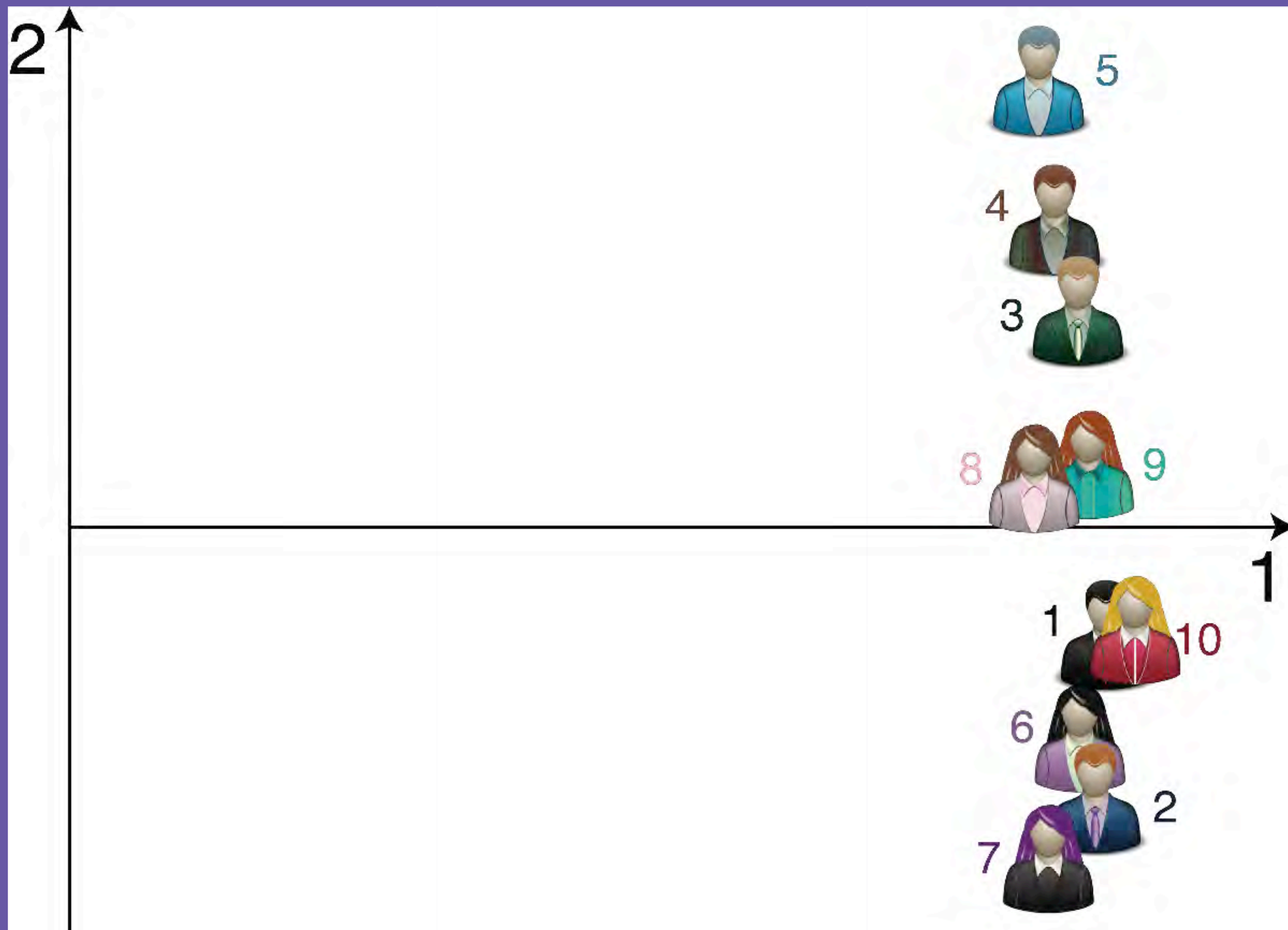
A Cross-Product Matrix: $X_1X_1^T$

$$\mathbf{C} = \begin{bmatrix}
0.51 & 0.44 & 0.40 & 0.38 & 0.34 & 0.40 & 0.41 & 0.35 & 0.48 & 0.52 \\
0.44 & 0.52 & 0.36 & 0.41 & 0.30 & 0.42 & 0.40 & 0.39 & 0.43 & 0.54 \\
0.40 & 0.36 & 0.42 & 0.39 & 0.33 & 0.35 & 0.33 & 0.31 & 0.42 & 0.46 \\
0.38 & 0.41 & 0.39 & 0.56 & 0.37 & 0.40 & 0.37 & 0.39 & 0.46 & 0.51 \\
0.34 & 0.30 & 0.33 & 0.37 & 0.35 & 0.30 & 0.27 & 0.29 & 0.37 & 0.39 \\
0.40 & 0.42 & 0.35 & 0.40 & 0.30 & 0.48 & 0.39 & 0.34 & 0.45 & 0.51 \\
0.41 & 0.40 & 0.33 & 0.37 & 0.27 & 0.39 & 0.46 & 0.32 & 0.41 & 0.46 \\
0.35 & 0.39 & 0.31 & 0.39 & 0.29 & 0.34 & 0.32 & 0.43 & 0.38 & 0.44 \\
0.48 & 0.43 & 0.42 & 0.46 & 0.37 & 0.45 & 0.41 & 0.38 & 0.60 & 0.54 \\
0.52 & 0.54 & 0.46 & 0.51 & 0.39 & 0.51 & 0.46 & 0.44 & 0.54 & 0.68
\end{bmatrix}.$$

Cosine (or R_V) Matrix

$$\mathbf{C} = \mathbf{U}\Theta\mathbf{U}^{T} = \begin{bmatrix} 0.325 & -0.212 \\ 0.326 & -0.359 \\ 0.289 & 0.212 \\ 0.326 & 0.561 \\ 0.253 & 0.443 \\ 0.311 & -0.237 \\ 0.294 & -0.391 \\ 0.279 & 0.124 \\ 0.349 & 0.137 \\ 0.389 & -0.165 \end{bmatrix} \times \mathrm{diag}\begin{Bmatrix} 4.135 \\ 0.224 \end{Bmatrix} \times \begin{bmatrix} 0.325 & -0.212 \\ 0.326 & -0.359 \\ 0.289 & 0.212 \\ 0.326 & 0.561 \\ 0.253 & 0.443 \\ 0.311 & -0.237 \\ 0.294 & -0.391 \\ 0.279 & 0.124 \\ 0.349 & 0.137 \\ 0.389 & -0.165 \end{bmatrix}^{T}$$

Eigen decomposition of C



Eigen of c: The assessors' map

$$G = U\Theta^{\frac{1}{2}} = \begin{bmatrix} 0.662 & -0.100 \\ 0.662 & -0.170 \\ 0.588 & 0.100 \\ 0.662 & 0.265 \\ 0.515 & 0.209 \\ 0.633 & -0.112 \\ 0.598 & -0.185 \\ 0.567 & 0.059 \\ 0.710 & 0.065 \\ 0.791 & -0.078 \end{bmatrix}$$

Factor scores from C

- 1. Between Table Structure
- 2. Derive Optimal Weights

THE STEPS OF STATIS

Rescale factor scores dimension 1 to sum of 1

$$\alpha = \mathbf{u}_1 \times (\mathbf{u}_1 \mathbf{1})^{-1} = \begin{bmatrix} 0.325 \\ 0.326 \\ 0.289 \\ 0.326 \\ 0.253 \\ 0.311 \\ 0.294 \\ 0.279 \\ 0.349 \\ 0.389 \end{bmatrix} \times 3.141^{-1} = \begin{bmatrix} 0.104 \\ 0.104 \\ 0.092 \\ 0.104 \\ 0.081 \\ 0.099 \\ 0.094 \\ 0.089 \\ 0.111 \\ 0.124 \end{bmatrix}$$

Weights (equal weights = .10)

- From **α** get diagonal matrix **A**

WEIGHTS (EQUAL WEIGHTS = .10)

- 1. Between Table Structure
- 2. Derive Optimal Weights
- 3. Compute Compromise from α weights

THE STEPS OF STATIS

- Get diagonal matrix of masses for rows **M**

- **M** = $^{1}/_{I}\,\mathbf{I}$ (equal masses)

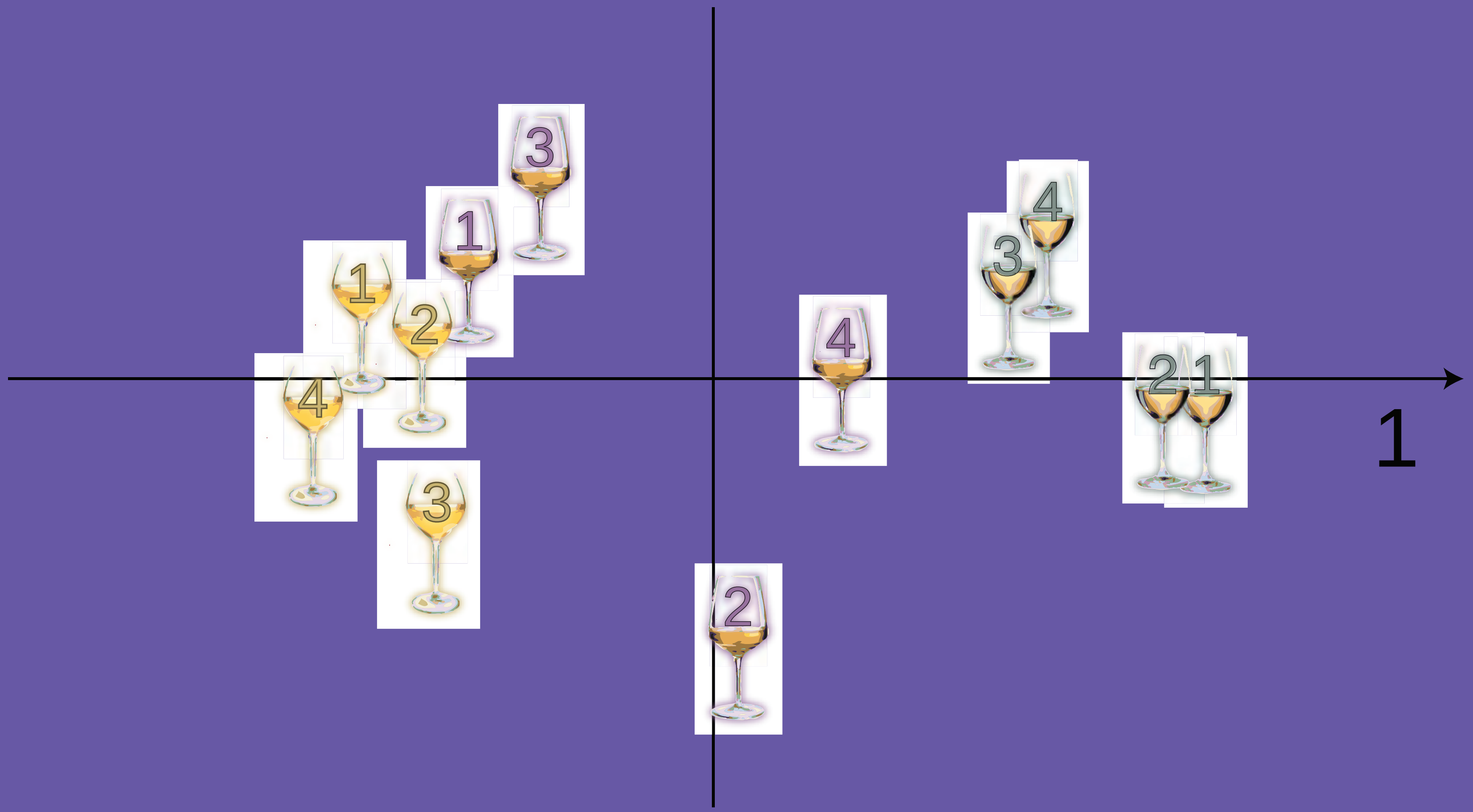
MASSES ARE FOR THE ROWS (EQUAL MASSES = .08)

- $\mathbf{X} = \mathbf{P\Delta Q}^T$ with $\mathbf{Q}^T\mathbf{AQ} = \mathbf{P}^T\mathbf{MP} = \mathbf{I}$

GET COMPROMISE. 1 GENERALIZED SVD OF X

- $\mathbf{X} = \mathbf{P \Delta Q}^T$ with $\mathbf{Q}^T\mathbf{AQ} = \mathbf{P}^T\mathbf{MP} = \mathbf{I}$
- $\mathbf{F} = \mathbf{P\Delta} = \mathbf{XAQ}$

GET COMPROMISE. 2 FACTOR SCORES



Compromise: Plot of factor scores

- 1. Between Table Structure
- 2. Derive Optimal Weights
- 3. Compute Compromise from Weights
- 4. Eigen-decompose Compromise
- 5. Project Original Tables (factor scores)

THE STEPS OF STATIS

- $\mathbf{X} = \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^{T}$ with $\mathbf{Q}^{T}\mathbf{A}\mathbf{Q} = \mathbf{P}^{T}\mathbf{M}\mathbf{P} = \mathbf{I}$
- $\mathbf{F} = \mathbf{P}\boldsymbol{\Delta} = \mathbf{X}\mathbf{A}\mathbf{Q}$

- $\mathbf{F}_{k} = \mathbf{X}_{k}\mathbf{Q}_{k}$

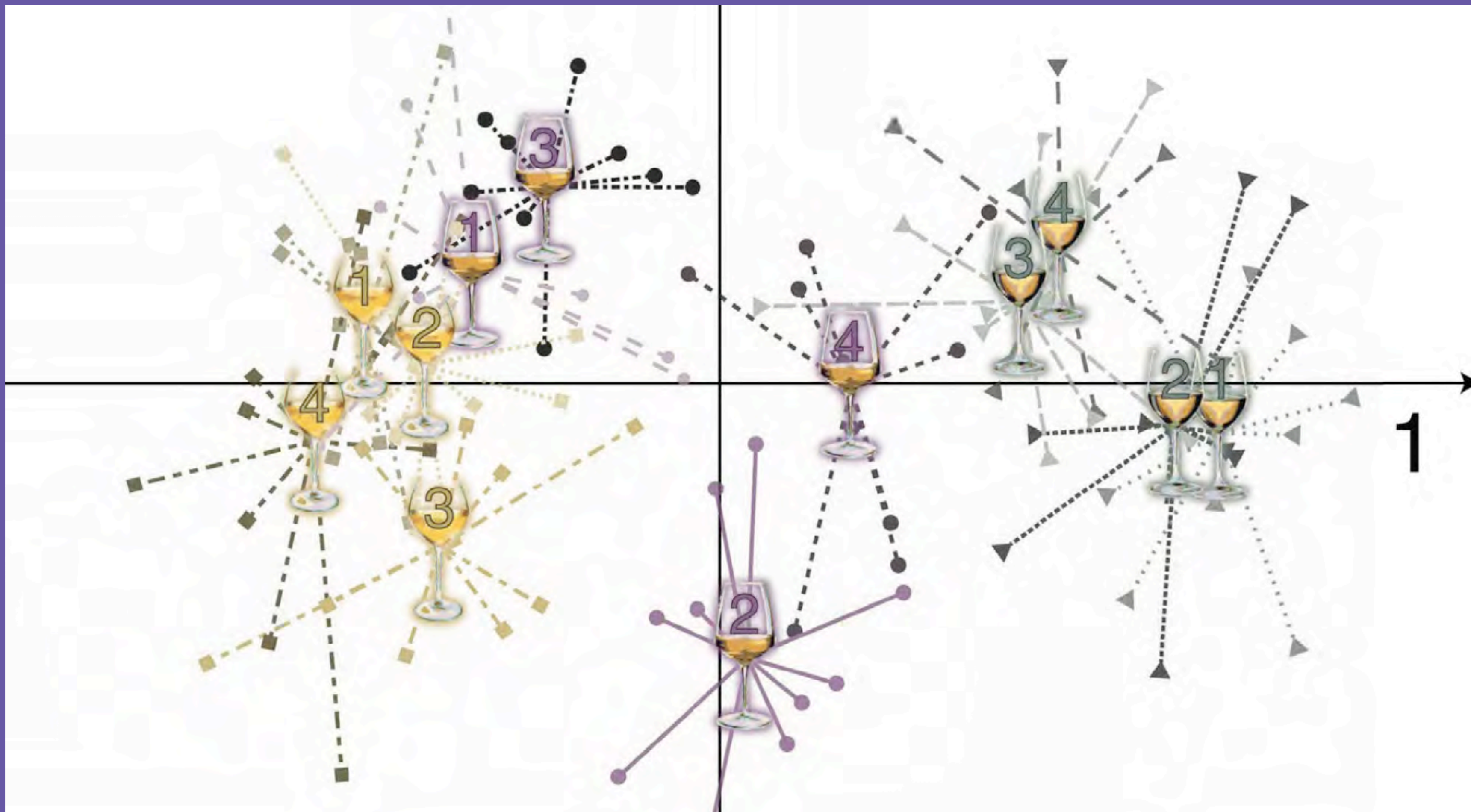
PARTIAL FACTOR SCORES

- $\mathbf{X} = \mathbf{P\Delta Q}^T$ with $\mathbf{Q}^T\mathbf{AQ} = \mathbf{P}^T\mathbf{MP} = \mathbf{I}$
- $\mathbf{F} = \mathbf{P\Delta} = \mathbf{XAQ}$

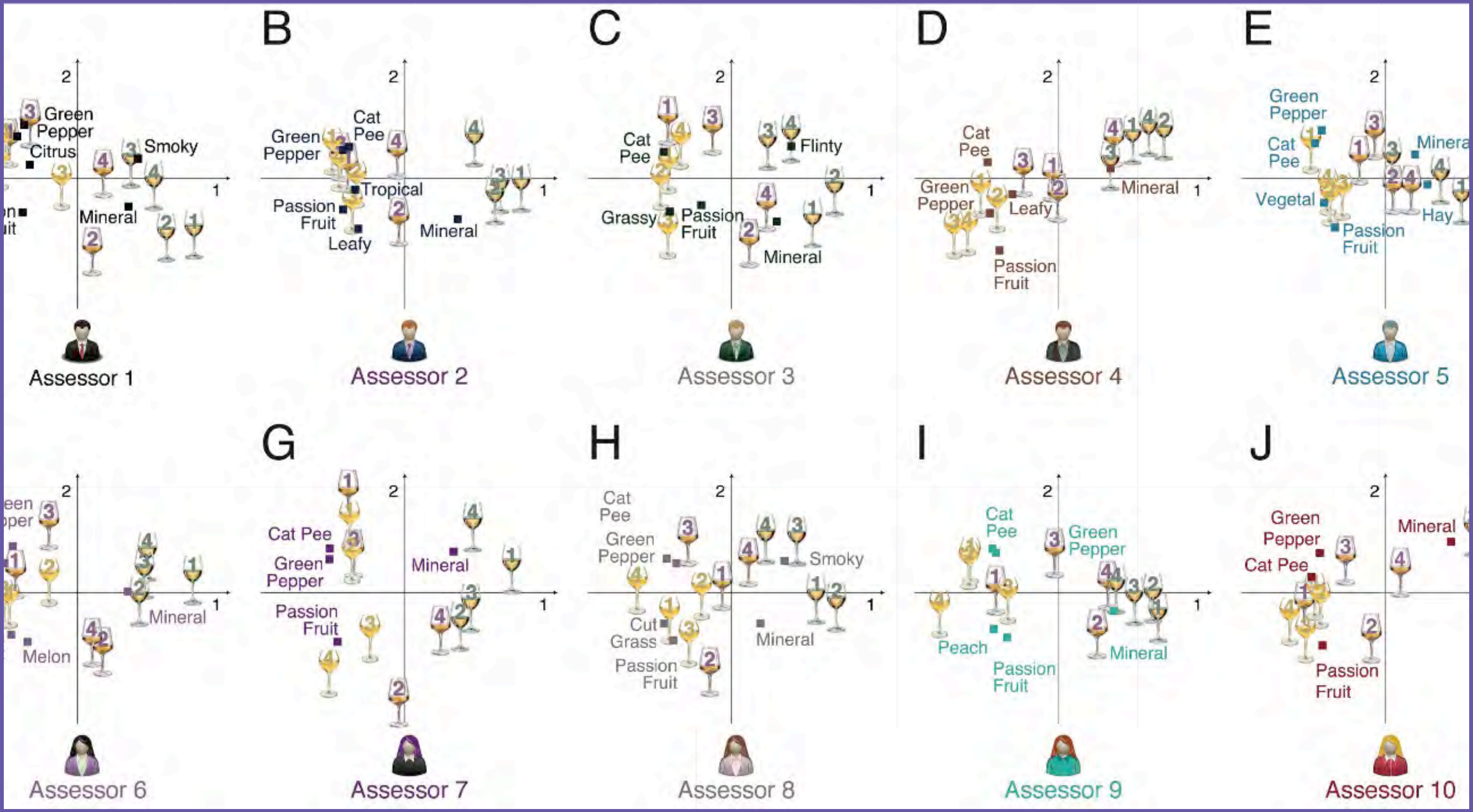
- $\mathbf{F}_k = \mathbf{X}_k\mathbf{Q}_k$

- $\mathbf{F} = \Sigma\, \alpha_k \mathbf{F}_k = \Sigma\, \alpha_k \mathbf{X}_k\mathbf{Q}_k$

PARTIAL FACTOR SCORES: BARYCENTRIC PROPERTY

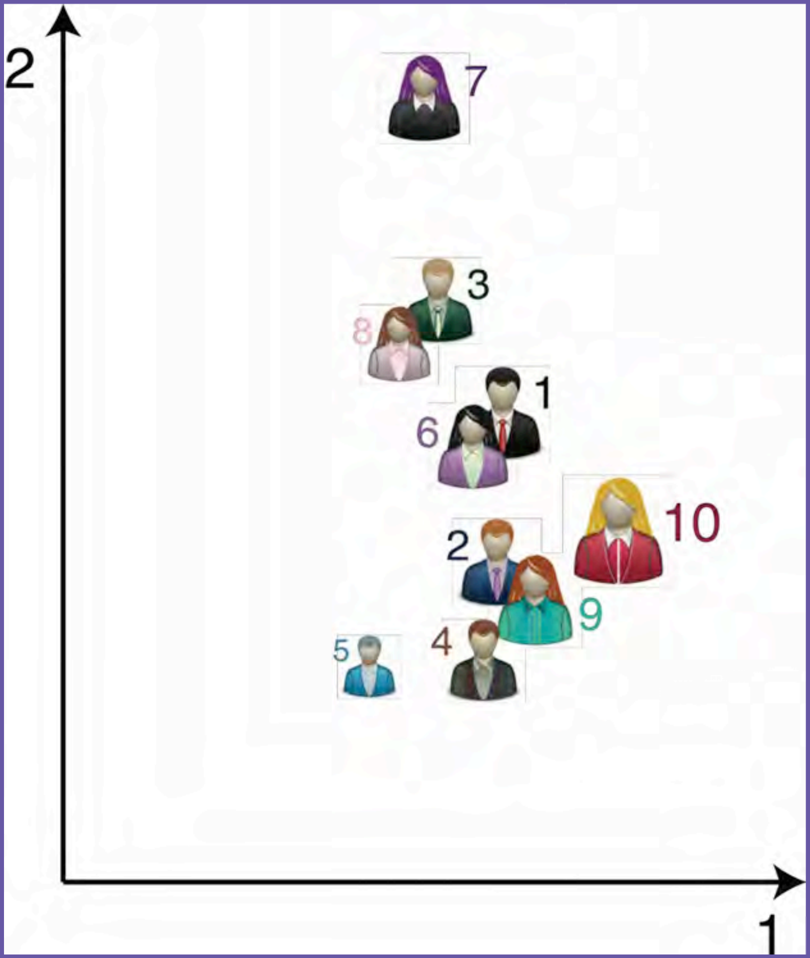


Compromise with "Tables"



THE TABLES AS "BIPLOTS"

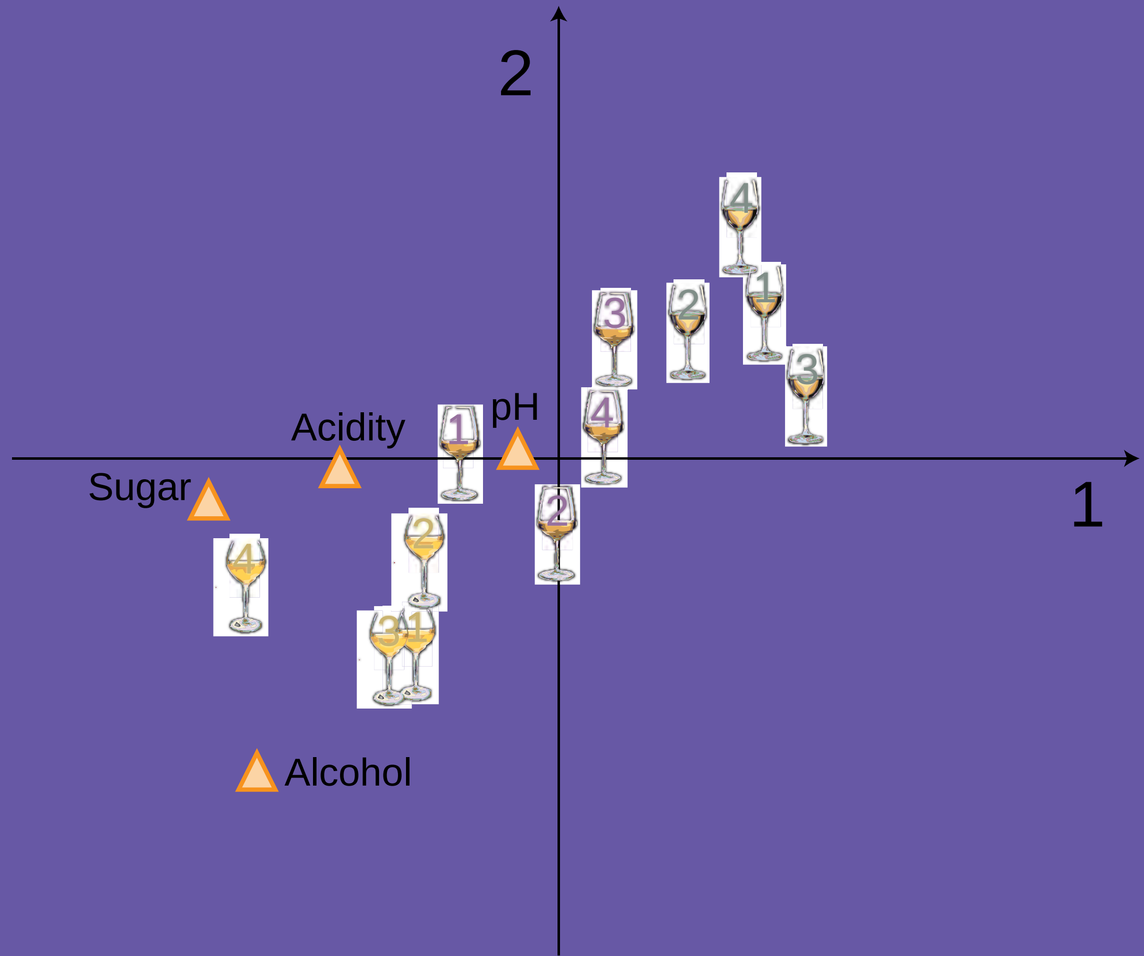
WHAT ARE THE IMPORTANT TABLES?



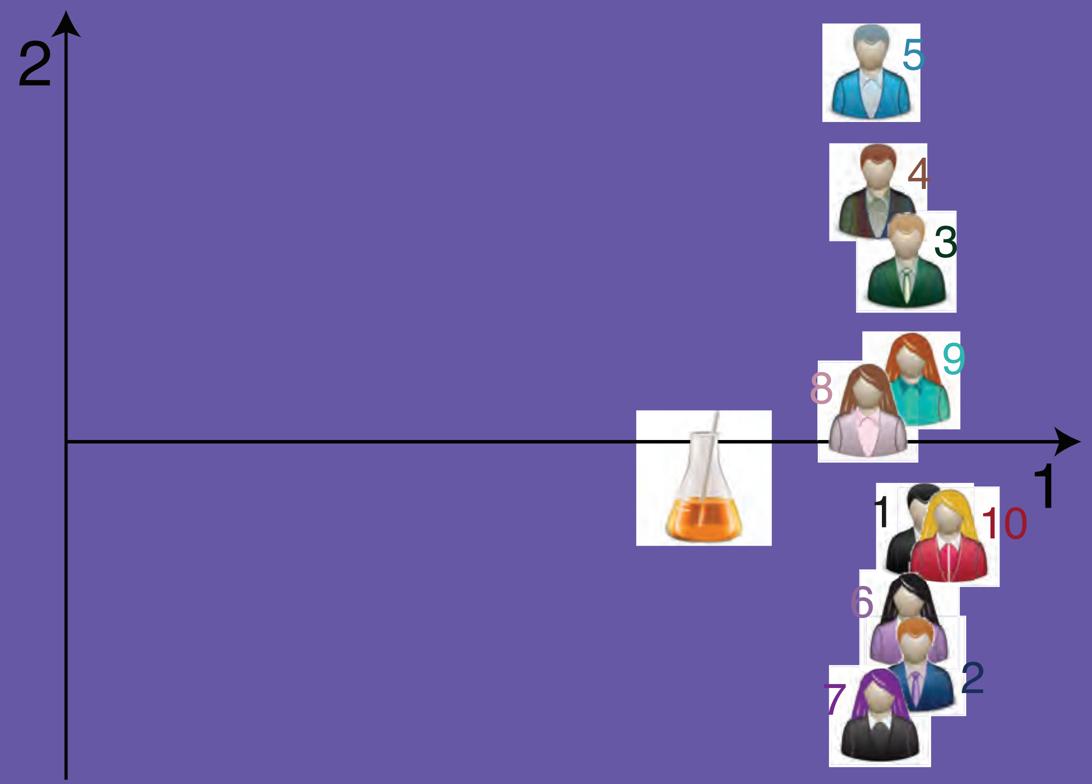
Contributions to Inertia

PARTIAL INERTIA

How to help: Projecting new tables



PHYSICO AS SUP (FACTOR SCORES + LOADINGS)



PHYSICO. RV AS SUP

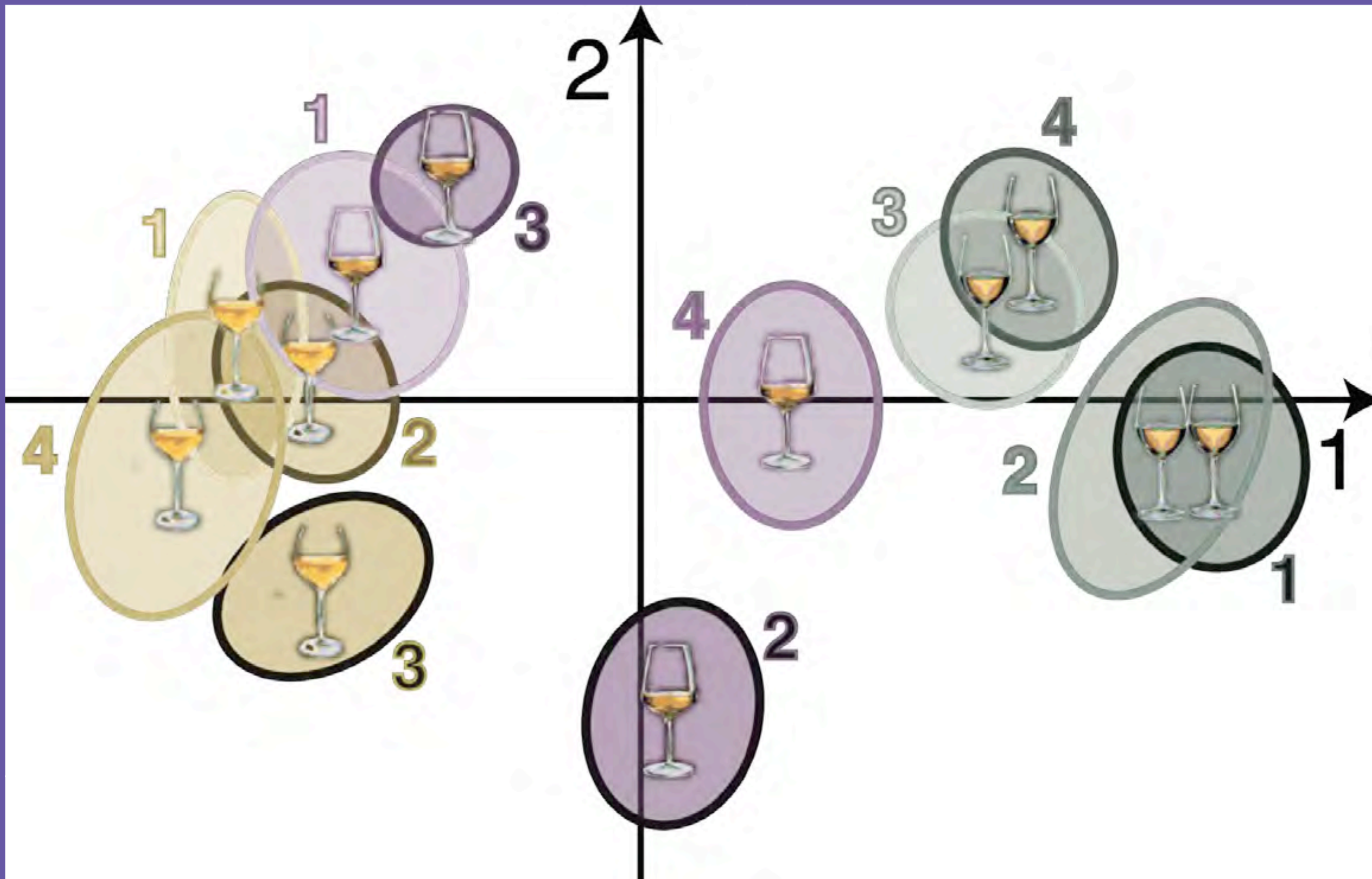
- 1. Between Table Structure
- 2. Derive Optimal Weights
- 3. Compute Compromise from Weights
- 4. Eigen-decompose Compromise
- 5. Project Original Tables (factor scores)
- 6… ? *Is the Earth ....*

THE STEPS OF STATIS

Is the earth round $p < .05$?

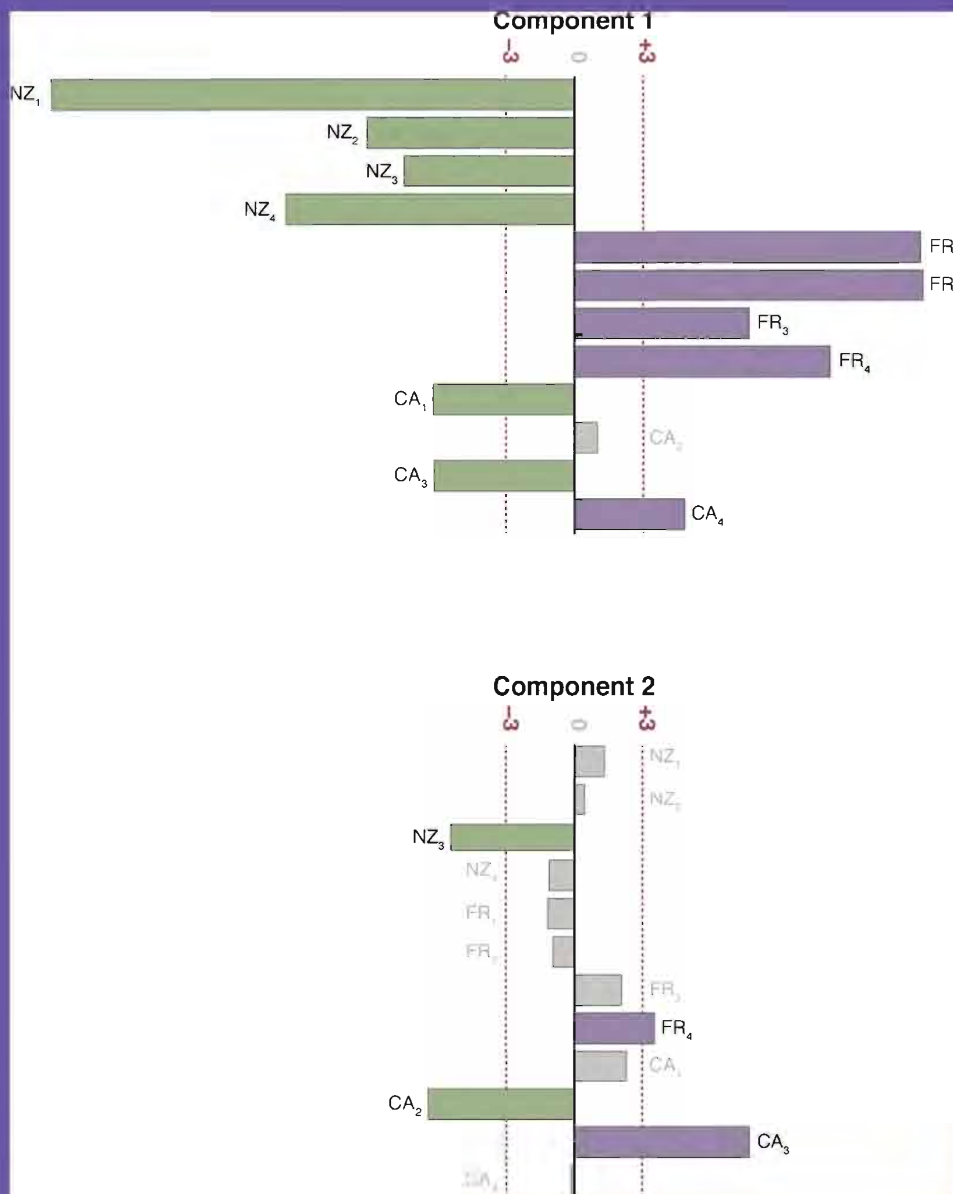
- Bootstrap again: The assessors are random

Compute confidence intervals



BOOTSTRAP 95% CI

Bootstrap Ratios (what is that?)



Component 1
−3
0
+3
NZ1
NZ2
NZ3
NZ4
FR1
FR2
FR3
FR4
CA1
CA2
CA3
CA4
Component 2
−3
0
+3
NZ1
NZ2
NZ3
NZ4
FR1
FR2
FR3
FR4
CA1
CA2
CA3
CA4

Extensions of STATIS

Partial Triadic Analysis

- Same variables all over:

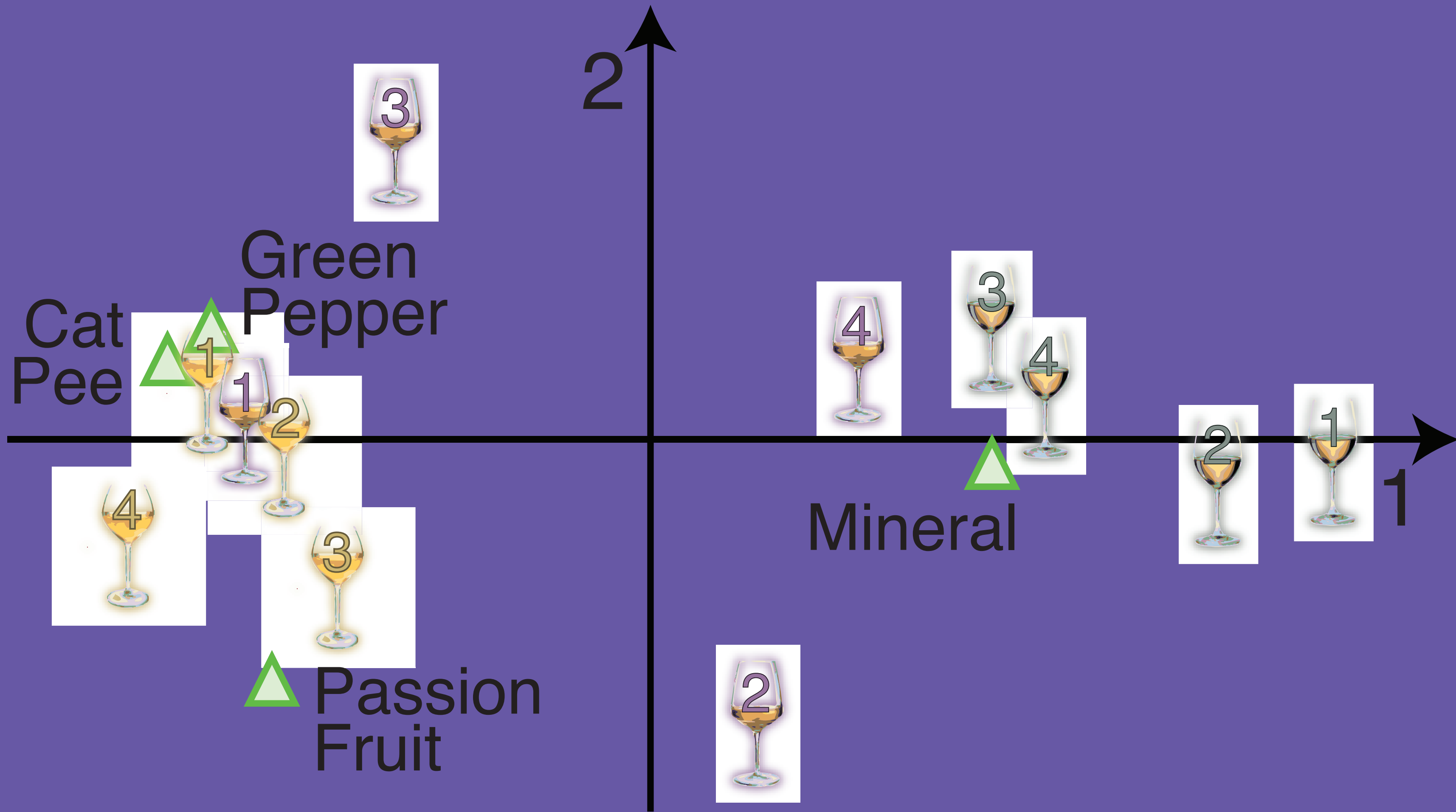
PARTIAL TRIADIC ANALYSIS

- Same variables all over:
  use $\mathbf{X}_k$ in lieu of $\mathbf{S}_k$

- Same variables all over:
  use $\mathbf{X}_k$ in lieu of $\mathbf{S}_k$

  Possible problem: negative cosine

PARTIAL TRIADIC ANALYSIS



PTA Factor Scores

DISTATIS

- *K* (squared Euclidean) distance matrices $\mathbf{D}_k$

DISTATIS: START WITH DISTANCE MATRICES

I . D . Double Centering > I . S

Distance

With a formula:

$s_{i,j} = d_{i,j} - (d_{i,+} - d_{+,+}) - (d_{+,j} - d_{+,+})$

With matrices:

$\mathbf{S} = -.5\Xi\mathbf{D}\Xi^{T}$ with $\Xi = \mathbf{I} - \mathbf{1}\mathbf{m}^{T}$ and $\mathbf{m}^{T}\mathbf{1} = 1$

TRANSFORMS THE DISTANCES INTO COVARIANCE

And back to standard STATIS

- Here 3 groups:

  France, Canada, New Zealand

N GROUPS: CANONICAL STATIS: CANOSTATIS.

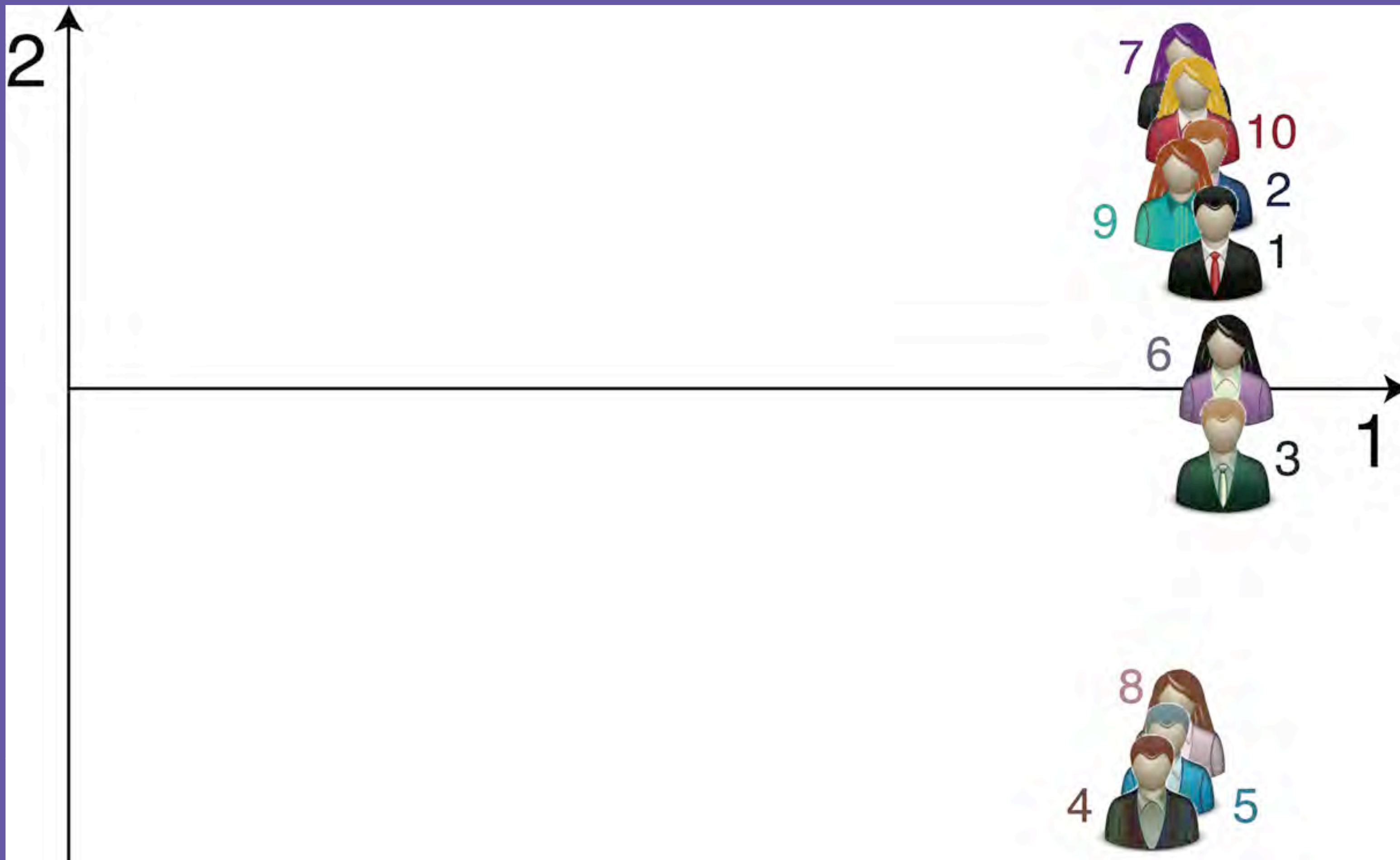
- Compute Mahalanobis distance per Table

N GROUPS

- Compute Mahalanobis distance per Table

- And back to DISTATIS

Wine Example



CANOSTATIS: THE ASSESSORS

Canonical Statis: 3 Groups

CANOSTATIS WITH CONFIDENCE

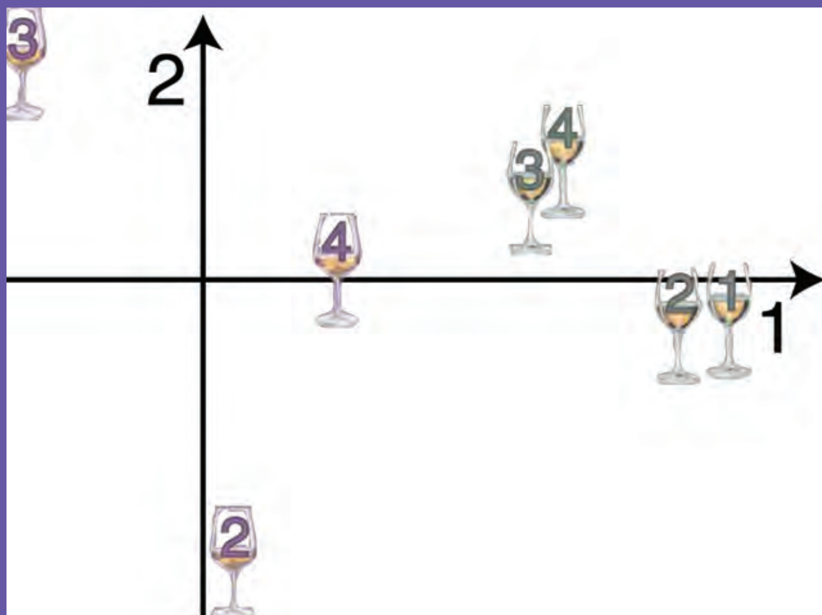
One more table: (K+1) statis

- Use $\mathbf{S}_k^* = \mathbf{HX}_k$ instead of $\mathbf{S}_k$

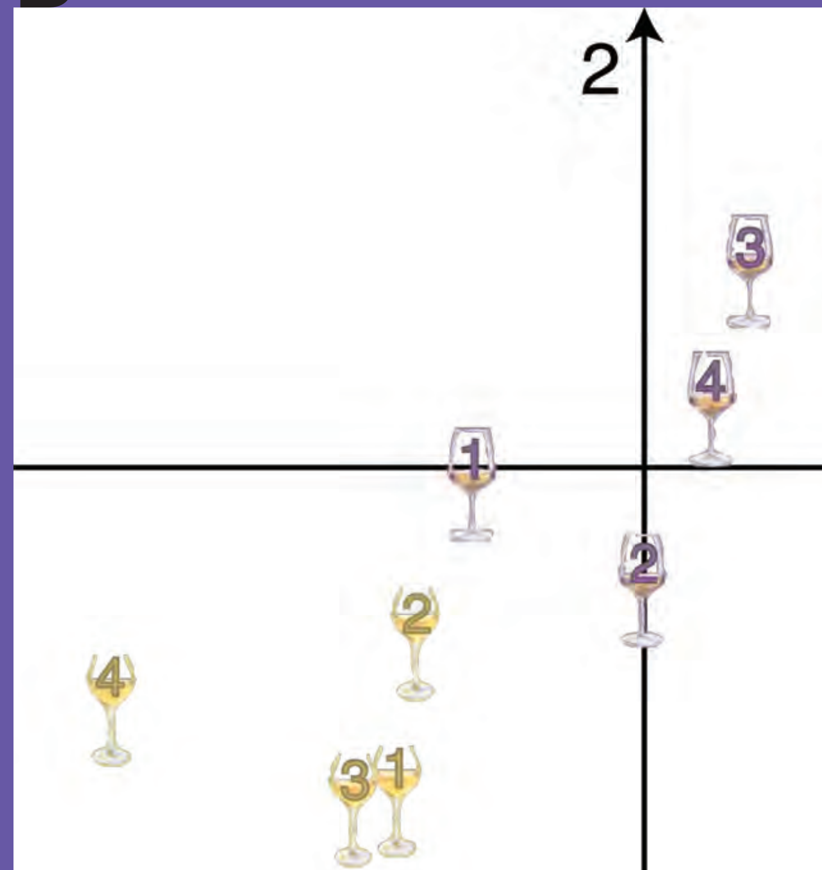
ONE MORE TABLE: (K+1) STATIS

- Use $\mathbf{S}_k^* = \mathbf{HX}_k$ instead of $\mathbf{S}_k$
- and back to STATIS

ONE MORE TABLE: (K+1) STATIS



B



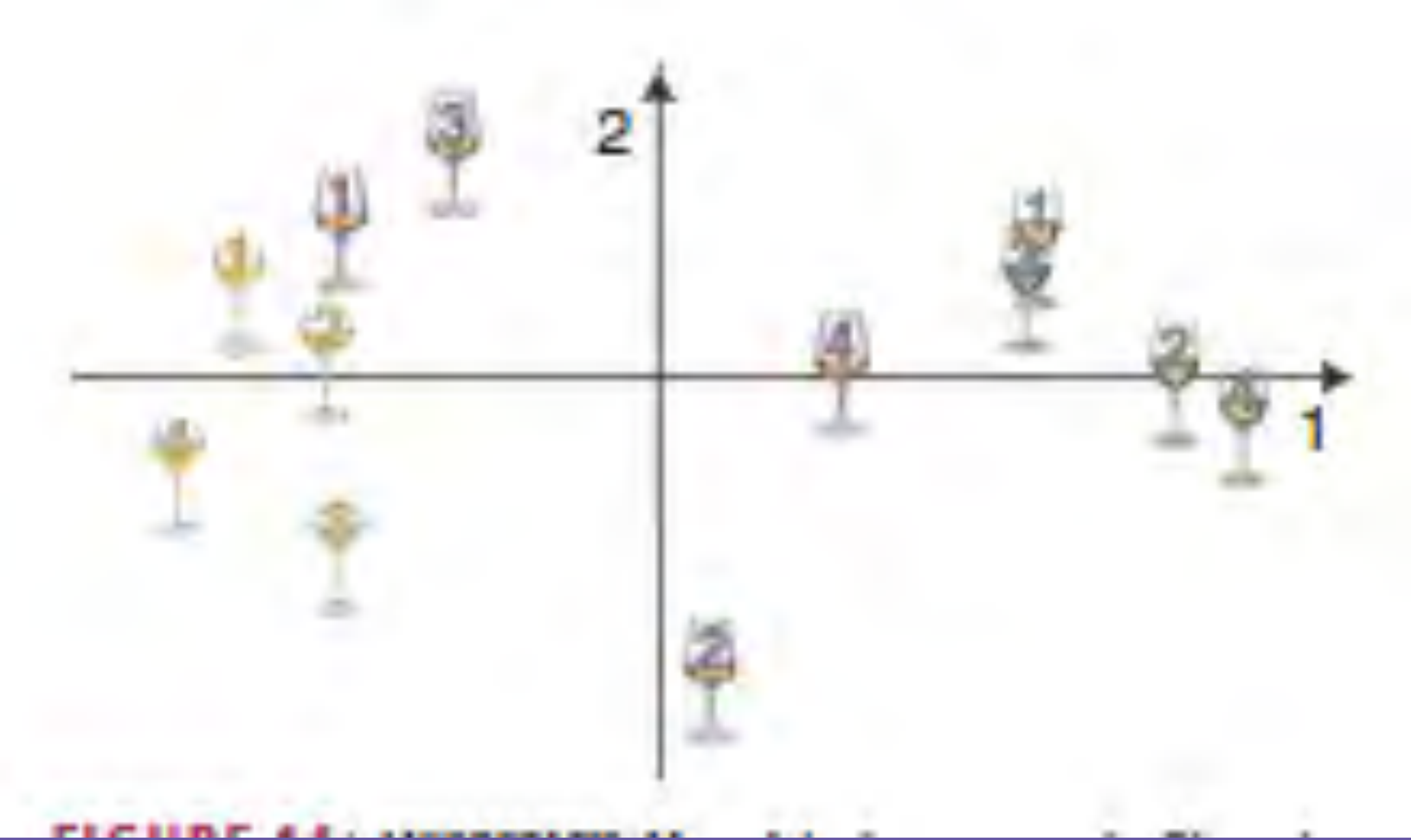
(*K*+1) STATIS

1) COMPROMISE & 2) PHYSICO

ANISO-STATIS

- One weight per column

ANISO-STATIS



ANISOSTATIS

Double statis: do-statis or do-act

- Two sets of matrices

DOUBLE STATIS: DO-STATIS OR DO-ACT

- Generalized canonical correlation
- Multiple factor analysis & SUM-PCA
- INDSCAL

RELATED TECHNIQUES