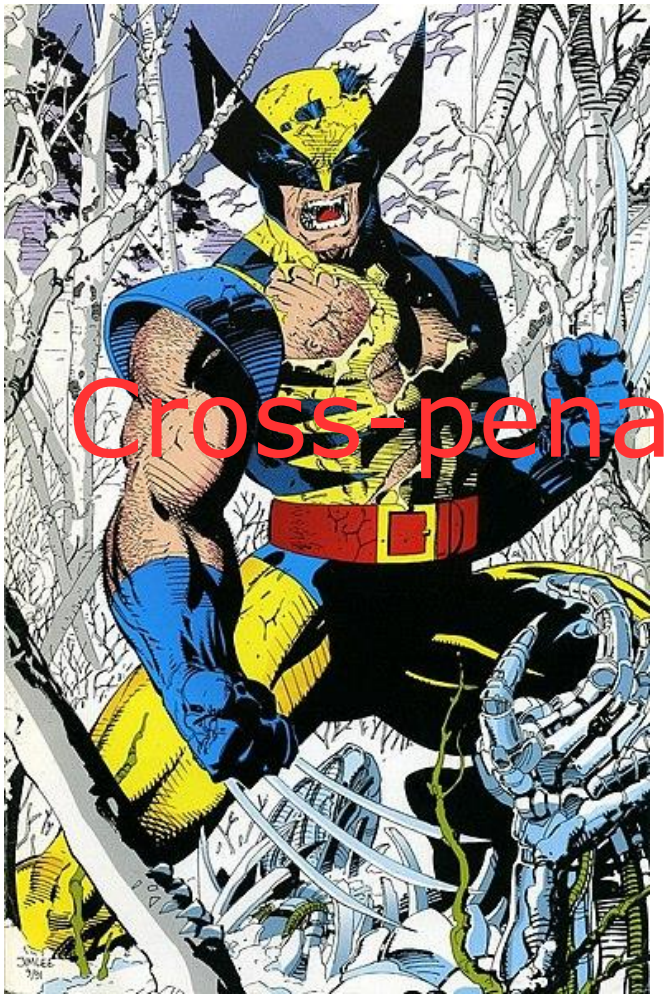


X-CAN

Cross-penalized Component ANalysis

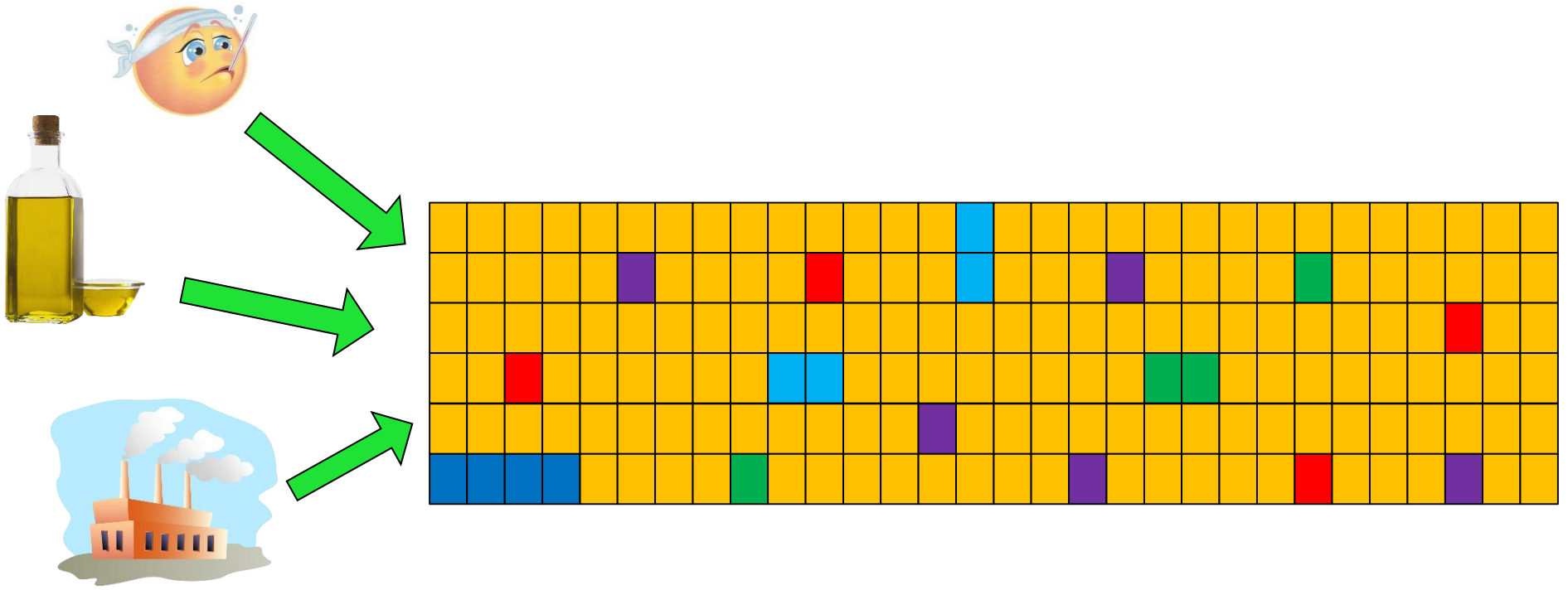
SSC'16, June 2019

José Camacho, Ph.D. (josecamacho@ugr.es)

http://wpd.ugr.es/~josecamacho/

Joint work with E. Acar, M. Rasmussen and R. Bro

Goal: Understand complex data



University of Granada
José Camacho, Ph.D.

The Animal Data

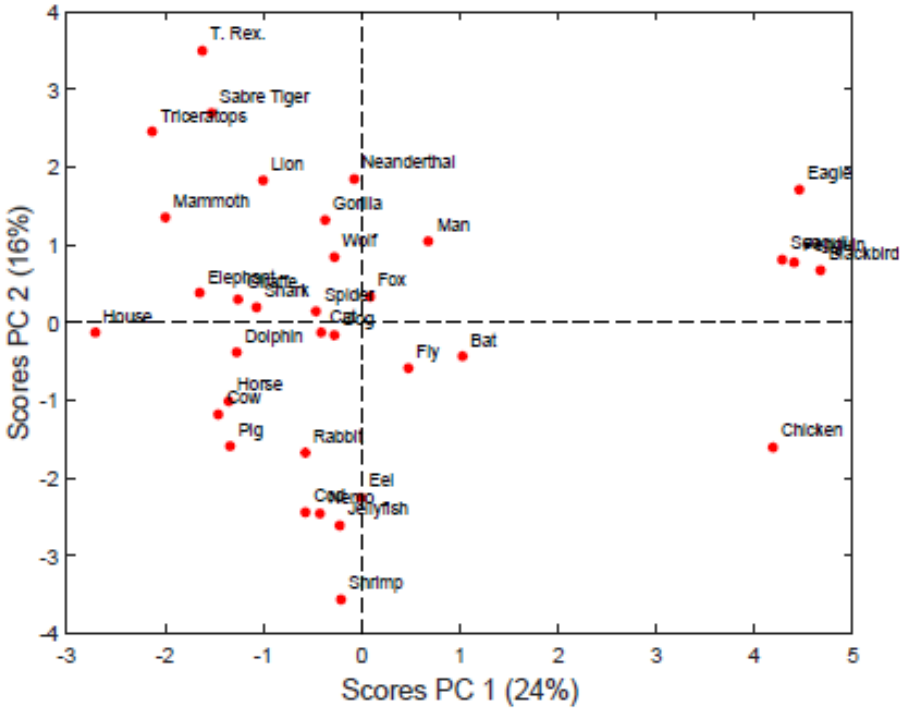
- ✓ Animal Data (Bro et al., J. Chemometrics 2012; 26: 256–263)

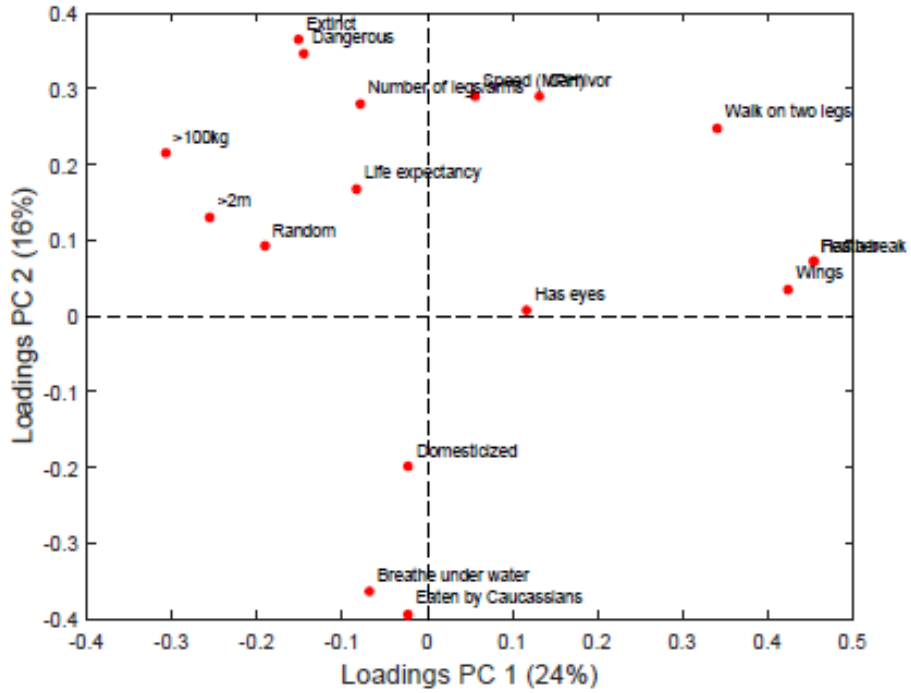
Table I. Animal data set used to illustrate coclustering

	Has eyes	Number of legs/ arms	Carnivore	Feather	Wings	Domesticized	Eaten by Caucasians	>100 kg	>2 m	Breathe under water	Extinct	Dangerous	Life expectancy	Random	Has a beak	Walk on two legs	Speed (MPH)
Giraffe	1	4	0	0	0	0	0	1	1	0	0	0	30	1	0	0	32
Cow	1	4	0	0	0	1	1	1	1	0	0	0	15	3	0	0	30
Lion	1	4	1	0	0	0	0	1	0	0	0	1	15	6	0	0	50
Gorilla	1	4	0	0	0	0	0	1	0	0	0	1	30	2	0	1	25
Fly	1	6	0	0	1	0	0	0	0	0	0	0	0,1	7	0	0	5
Spider	1	8	1	0	0	0	0	0	0	0	0	0	1	8	0	0	1
Shark	1	0	1	0	0	0	0	1	0	1	0	1	50	4	0	0	30
House	0	0	0	0	0	0	0	1	1	0	0	0	100	9	0	0	0
Horse	1	4	0	0	0	1	1	1	1	0	0	0	15	2	0	0	40
Elephant	1	4	0	0	0	0	0	1	1	0	0	0	35	6	0	0	25
Mammoth	1	4	0	0	0	0	0	1	1	0	1	0	35	5	0	0	25
Sabre Tiger	1	4	1	0	0	0	0	1	0	0	1	1	15	7	0	0	40
Pig	1	4	0	0	0	1	1	1	0	0	0	0	25	8	0	0	11
Cod	1	0	1	0	0	0	1	0	0	1	0	0	40	9	0	0	2
Eel	1	0	1	0	0	0	1	0	0	1	0	0	55	1	0	0	20
Jellyfish	1	0	0	0	0	0	0	0	0	1	0	0	0,7	3	0	0	1
Dolphin	1	0	1	0	0	0	0	1	1	1	0	0	30	5	0	0	35
Nemo	1	0	0	0	0	0	0	0	0	1	0	0	1	6	0	0	4
Shrimp	1	0	0	0	0	0	1	0	0	1	0	0	1	2	0	0	0,5
Dog	1	4	1	0	0	1	0	0	0	0	0	0	13	8	0	0	35
Cat	1	4	1	0	0	1	0	0	0	0	0	0	25	9	0	0	30
Fox	1	4	1	0	0	0	0	0	0	0	0	0	14	4	0	0	42
Wolf	1	4	1	0	0	0	0	0	0	0	0	1	18	3	0	0	25
Rabbit	1	4	0	0	0	1	1	0	0	0	0	0	9	8	0	0	35
Chicken	1	2	0	1	1	1	1	0	0	0	0	0	15	1	1	1	9
Eagle	1	2	1	1	1	0	0	0	0	0	0	0	55	3	1	1	60
Seagull	1	2	1	1	1	0	0	0	0	0	0	0	10	6	1	1	25
Blackbird	1	2	1	1	1	0	0	0	0	0	0	0	18	0	1	1	25
Bat	1	2	1	0	1	0	0	0	0	0	0	0	24	4	0	0	8
T. Rex.	1	4	1	0	0	0	0	1	1	0	1	1	40	9	0	1	25
Neanderthal	1	4	1	0	0	0	0	0	0	0	1	0	50	8	0	1	18
Triceratops	1	4	1	0	0	0	0	1	1	0	1	1	30	5	0	0	10
Man	1	4	1	0	0	0	0	0	0	0	0	0	80	2	0	1	28
Penguin	1	2	1	1	1	0	0	0	0	0	0	0	15	4	1	1	25

PCA

✓ Animal Data (PC1 24%, PC2 16%)





Rotation & SPCA

- ✓ The Problem:

"PCA is difficult to interpret because the resulting PCs are linear combinations of all the variables and observations."

- ✓ Solution: Simple structure
  - ✓ Rotation
  - ✓ SPCA

XCAN

✓ Cross-product Penalized Component Analysis

$$\mathbf{X} = \hat{\mathbf{U}}^X \hat{\mathbf{S}}^X (\hat{\mathbf{P}}^X)^T + \mathbf{E}$$

$$\{\hat{\mathbf{P}}^X, \hat{\mathbf{S}}^X, \hat{\mathbf{U}}^X\} = \arg\min_{\mathbf{P},\mathbf{S},\mathbf{U}} \|\mathbf{X} - \mathbf{USP}^T\|_F^2 + \lambda_r F_r + \lambda_c F_c \ s.t. \ \|\mathbf{u}_h\|_2^2 \leq 1, \ \|\mathbf{p}_h\|_2^2 \leq 1$$

INPUTS

$$F_r = \sum_{h=1}^{H} \|(\mathbf{u}_h \mathbf{u}_h^T) \oslash \mathbf{XXt}\|_F^2,$$

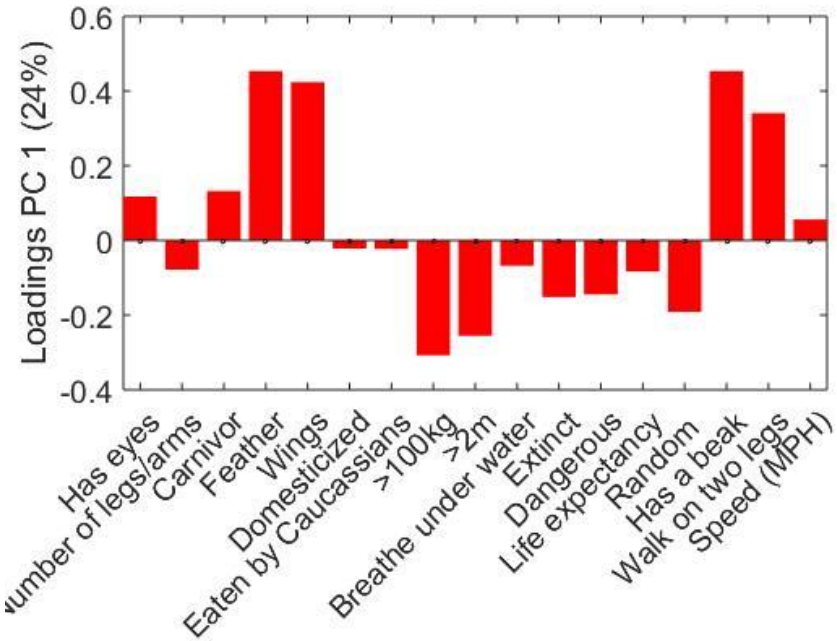
$$F_c = \sum_{h=1}^{H} \|(\mathbf{p}_h \mathbf{p}_h^T) \oslash \mathbf{XtX}\|_F^2,$$

INPUTS

XCAN

✓ Cross-product Penalized Component Analysis

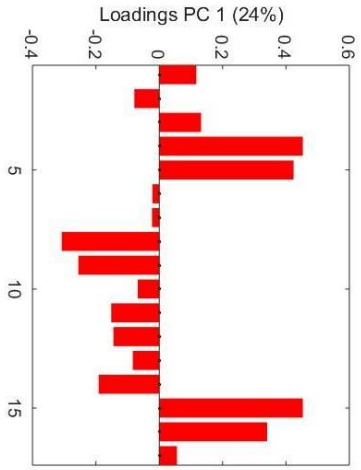
$$F_c = \sum_{h=1}^{H} \|(\mathbf{p}_h \mathbf{p}_h^T) \oslash \mathbf{XtX}\|_F^2,$$

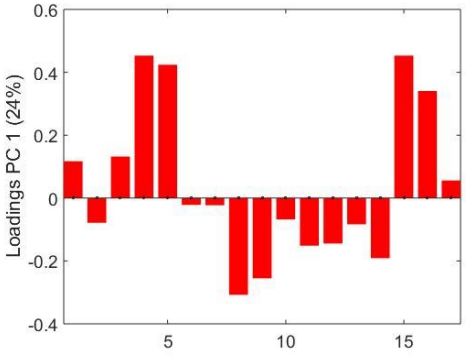


University of Granada
José Camacho, Ph.D.

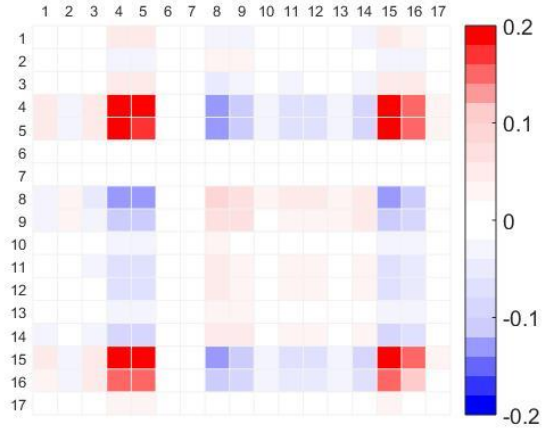
✓ Cross-product Penalized Component Analysis

$$F_c = \sum_{h=1}^{H} \|(\mathbf{p}_h \mathbf{p}_h^T) \oslash \mathbf{XtX}\|_F^2,$$





=

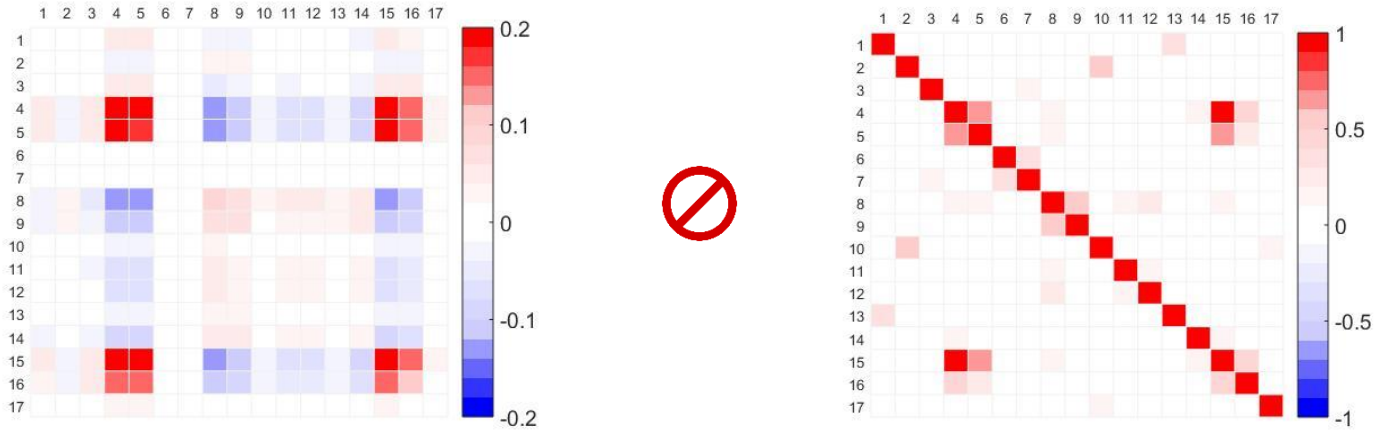


University of Granada
José Camacho, Ph.D.

XCAN

✓ Cross-product Penalized Component Analysis

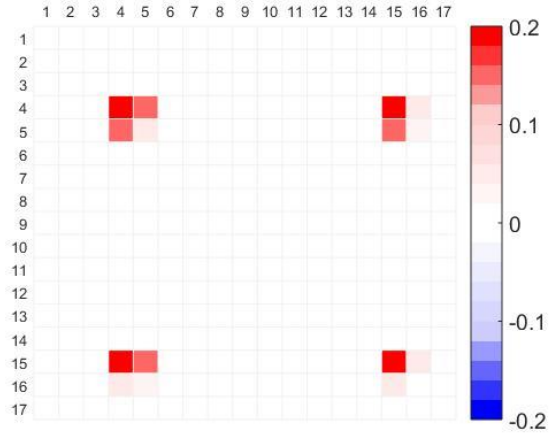
$$F_c = \sum_{h=1}^{H} \|(\mathbf{p}_h \mathbf{p}_h^T) \oslash \mathbf{XtX}\|_F^2,$$

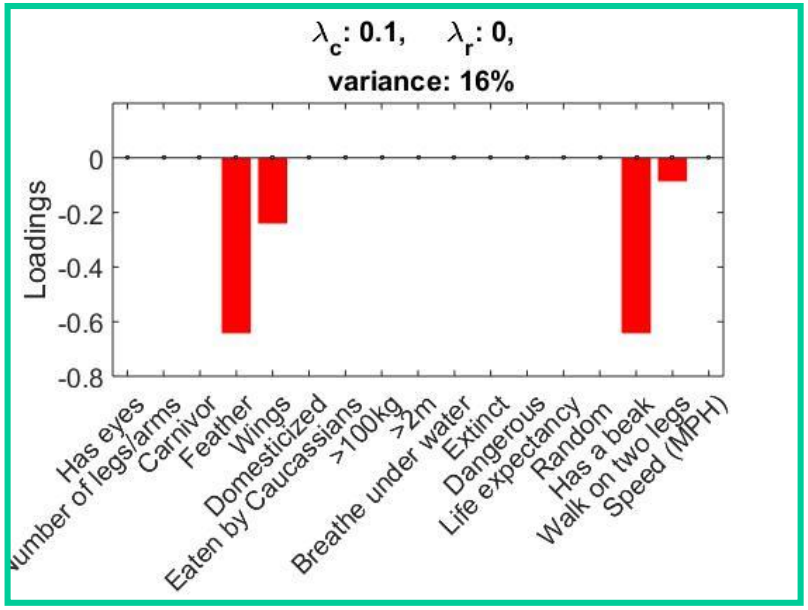


University of Granada
José Camacho, Ph.D.

✓ Cross-product Penalized Component Analysis

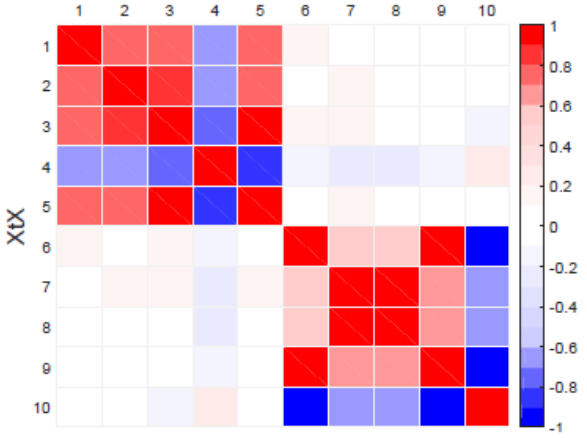
$$F_c = \sum_{h=1}^{H} \|(\mathbf{p}_h \mathbf{p}_h^T) \oslash \mathbf{XtX}\|_F^2,$$

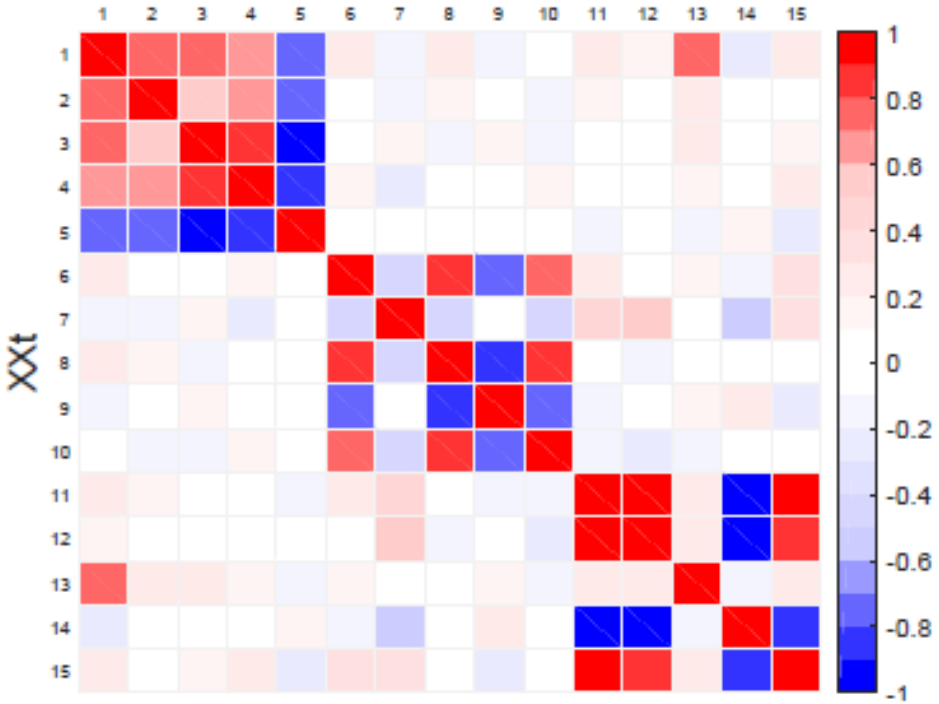


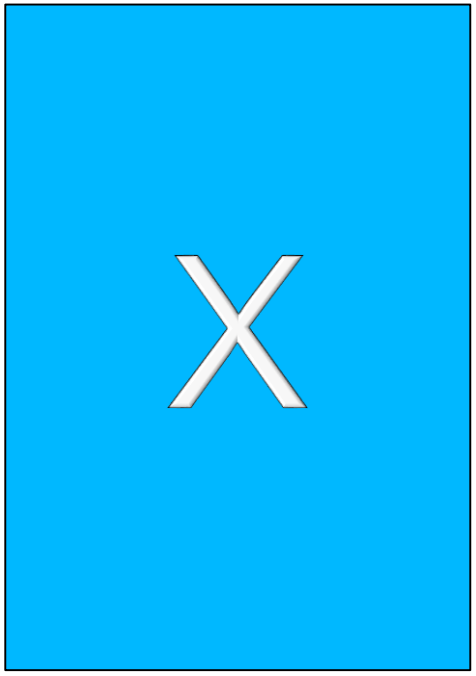


Simulated Example

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \\ \mathbf{0} & 2\mathbf{X}_3 \end{bmatrix} + 0.15 \cdot \mathbb{N}(\mathbf{0}, \mathbf{I}),$$



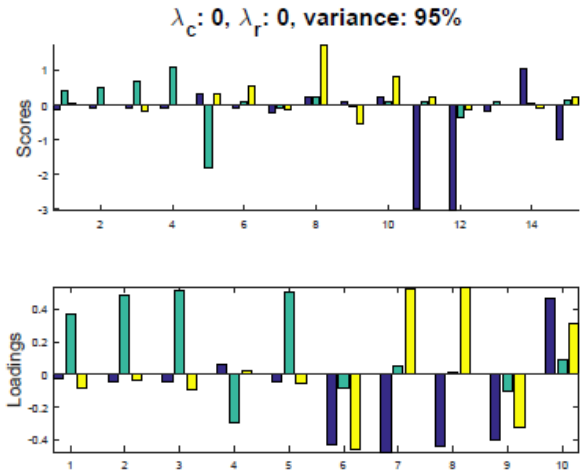




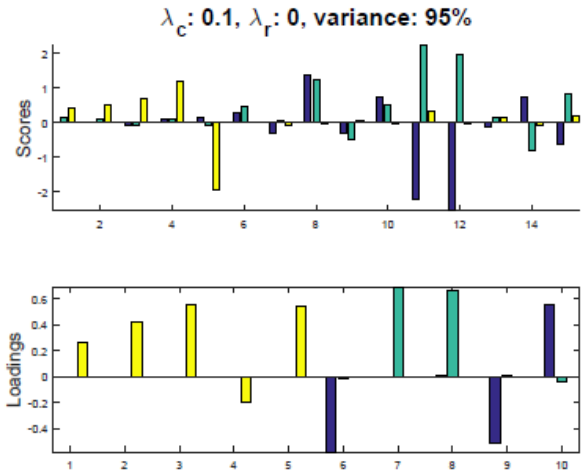
University of Granada
José Camacho, Ph.D.

Simulated Example

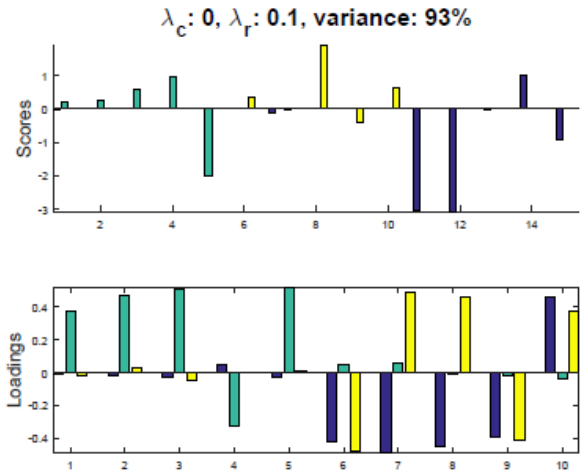
$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \\ \mathbf{0} & 2\mathbf{X}_3 \end{bmatrix} + 0.15 \cdot \mathbb{N}(\mathbf{0}, \mathbf{I}),$$



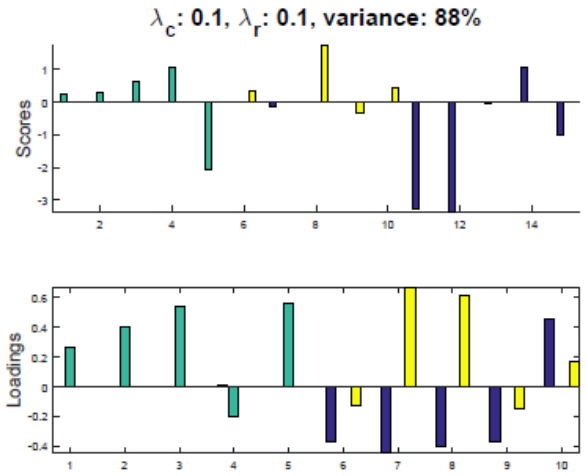
(a) PCA



(b) sparse loadings



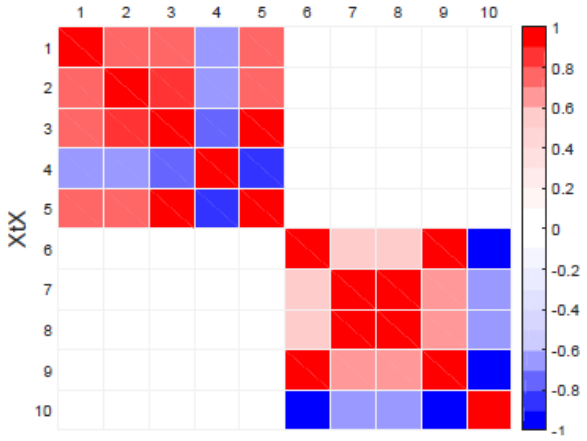
(c) sparse scores

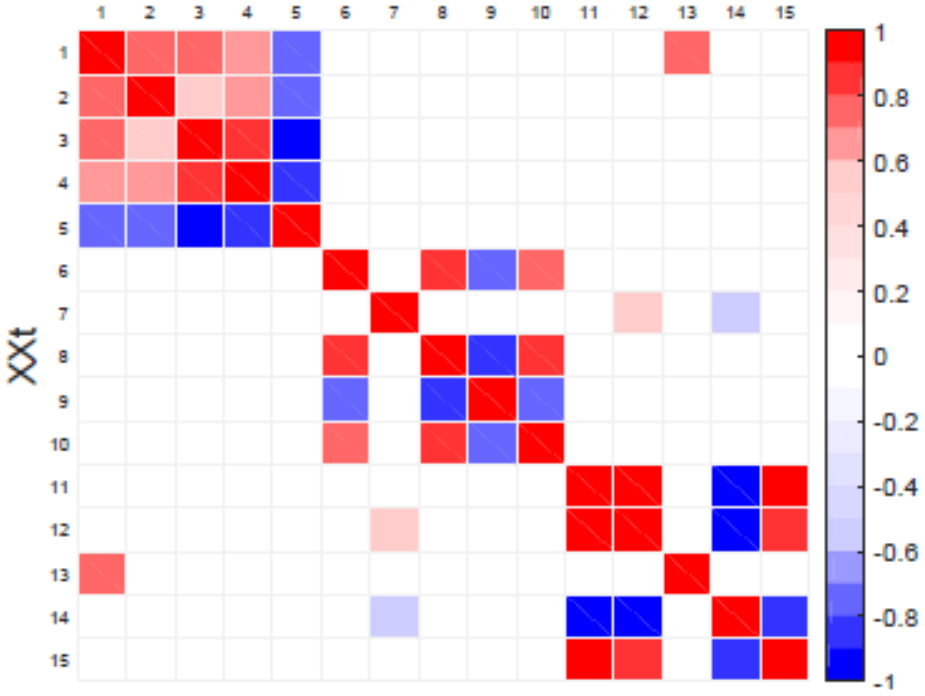


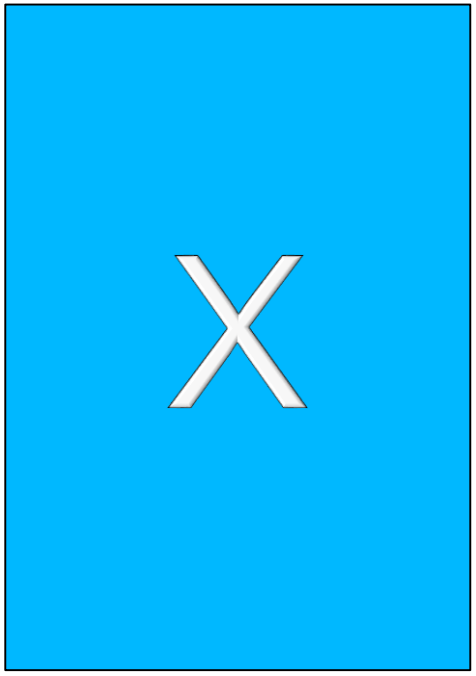
(d) sparse loadings and scores

Simulated Example

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \\ \mathbf{0} & 2\mathbf{X}_3 \end{bmatrix} + 0.15 \cdot \mathbb{N}(\mathbf{0}, \mathbf{I}),$$



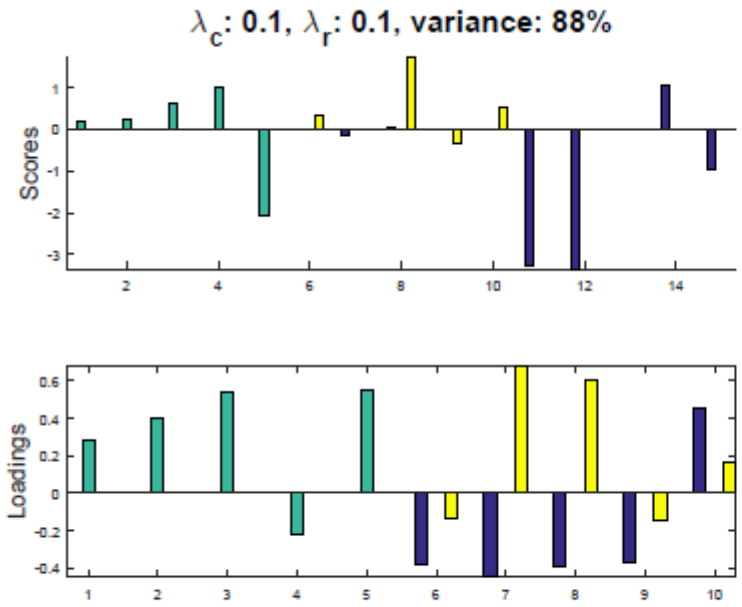




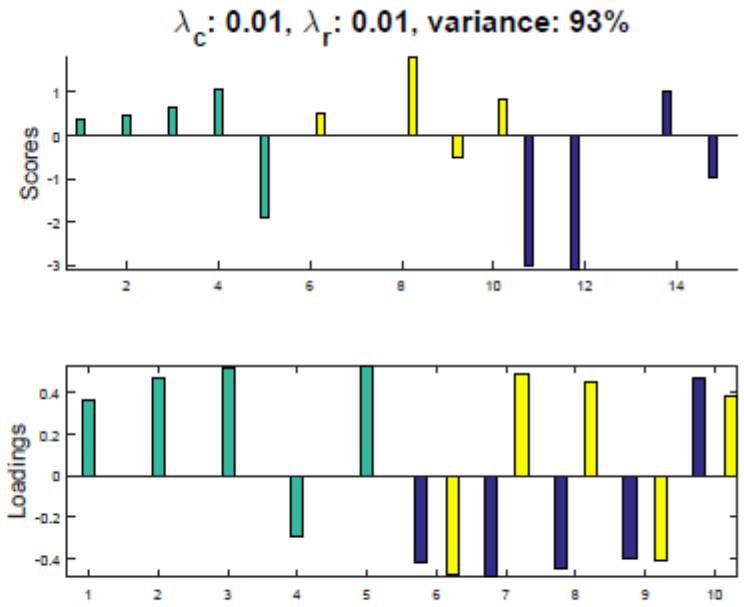
University of Granada
José Camacho, Ph.D.

Simulated Example

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \\ \mathbf{0} & 2\mathbf{X}_3 \end{bmatrix} + 0.15 \cdot \mathbb{N}(\mathbf{0}, \mathbf{I}),$$

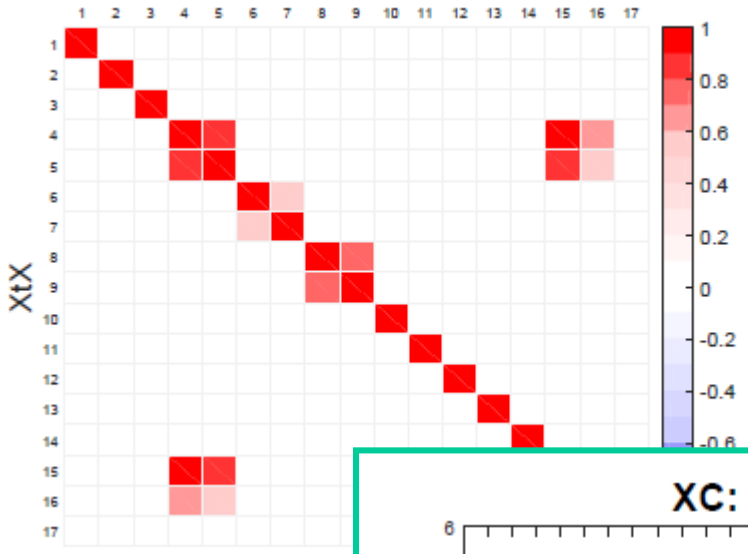


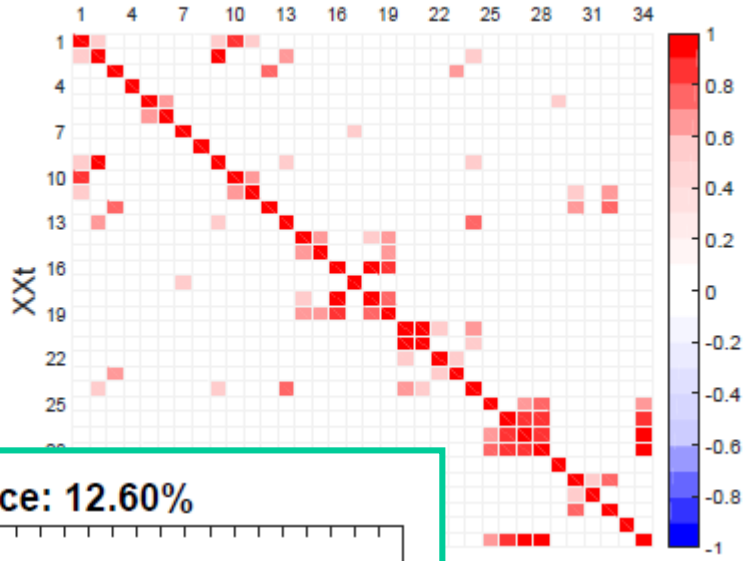
(a) Non-thresholded

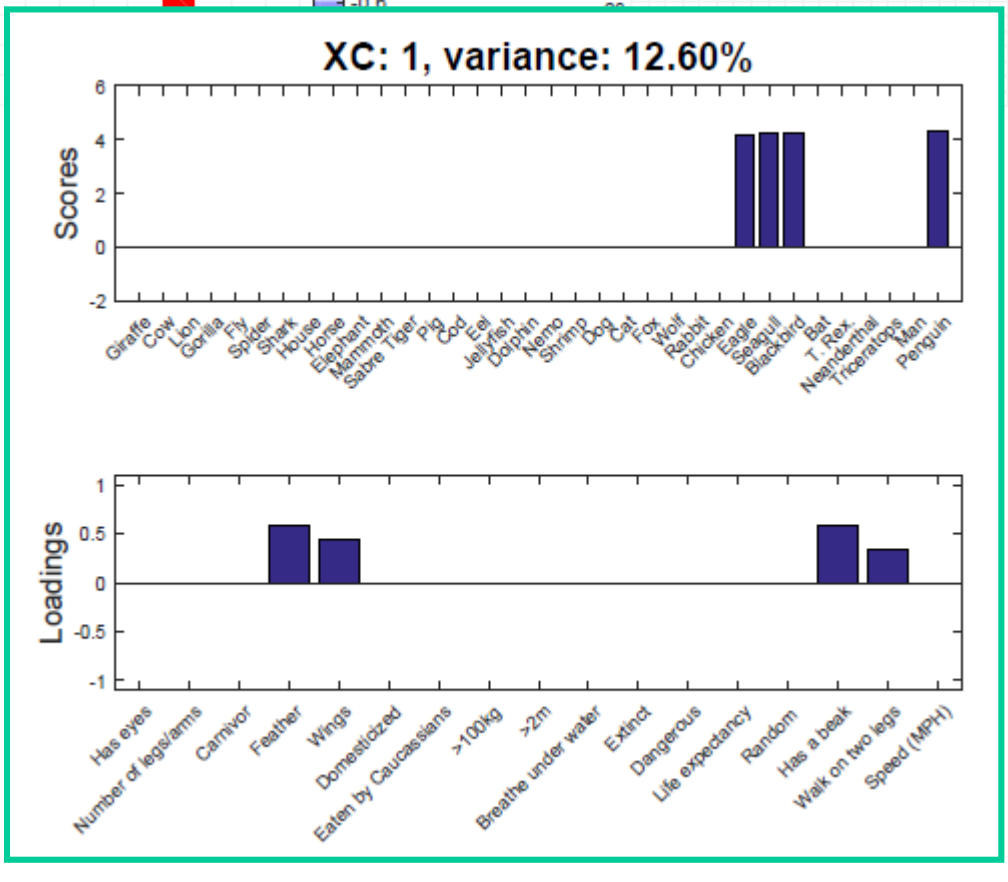


(b) Thresholded

Animal Data

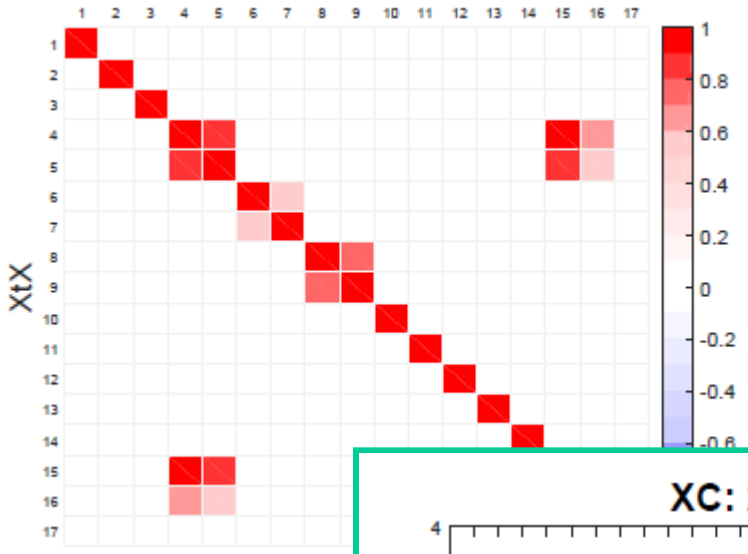


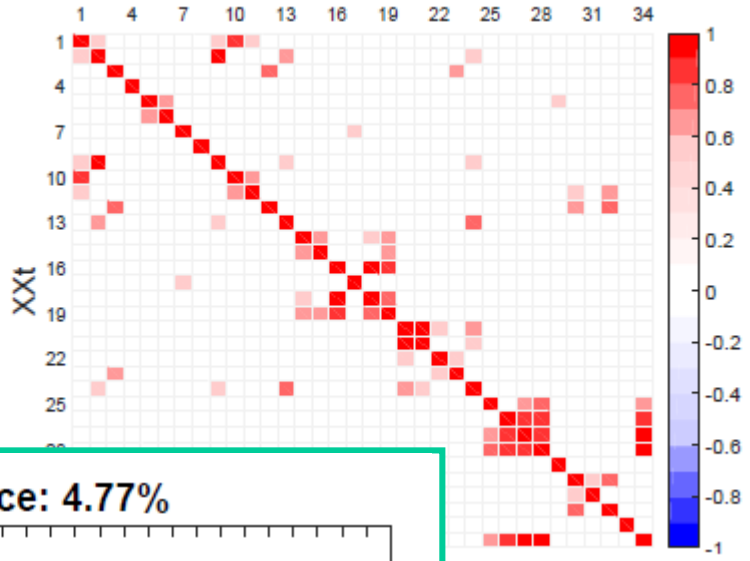


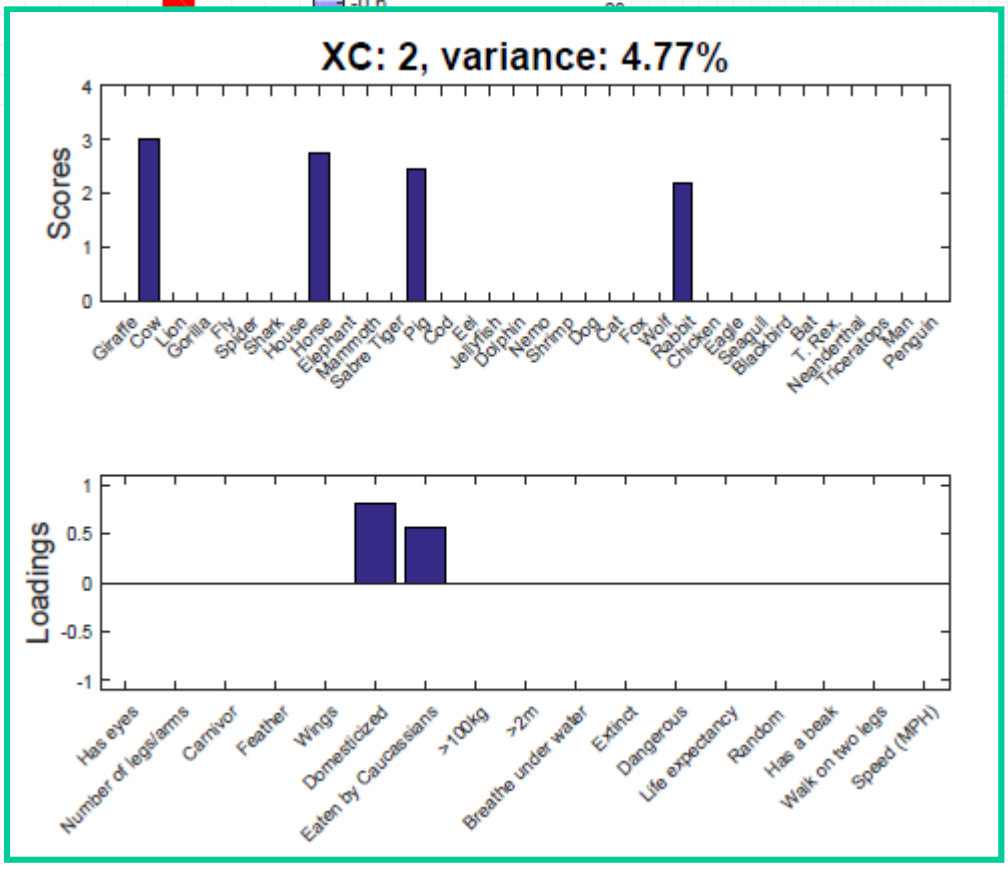


University of Granada
José Camacho, Ph.D.

Animal Data

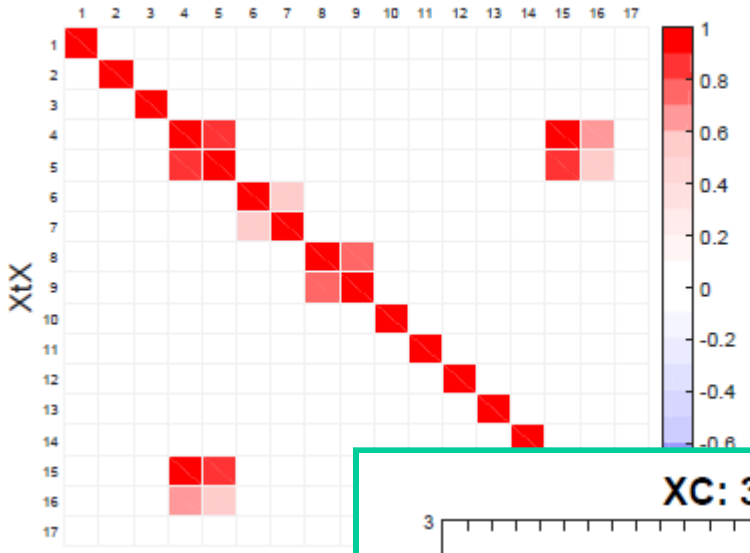


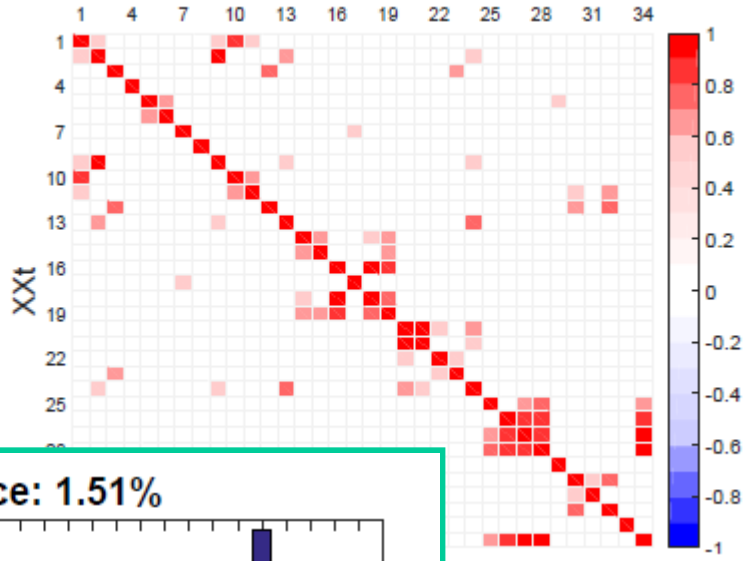


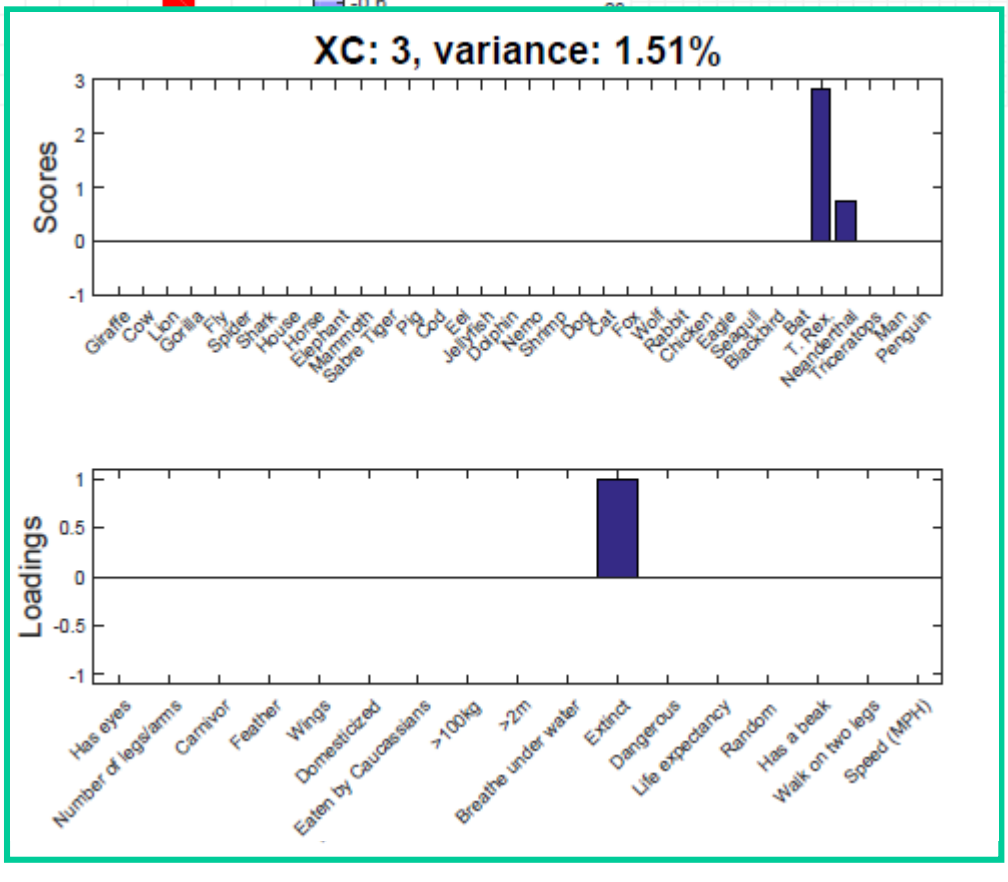


University of Granada
José Camacho, Ph.D.

Animal Data


XtX


XXt


XC: 3, variance: 1.51%
Scores
Giraffe
Cow
Lion
Gorilla
Fly
Spider
Shark
House
Horse
Elephant
Mammoth
Sabre Tiger
Pig
Cod
Eel
Jellyfish
Dolphin
Nemo
Shrimp
Dog
Cat
Fox
Wolf
Rabbit
Chicken
Eagle
Seagull
Blackbird
Bat
T. Rex
Neanderthal
Triceratops
Man
Penguin
Loadings
Has eyes
Number of legs/arms
Carnivor
Feather
Wings
Domesticized
Eaten by Caucassians
>100kg
>2m
Breathe under water
Extinct
Dangerous
Life expectancy
Random
Has a beak
Walk on two legs
Speed (MPH)

Conclusions

- ✓ XCAN:
  - ✓ variance maximization + structural penalties
  - ✓ One or two modes

- ✓ Future work
  - ✓ Omics data
  - ✓ Data fusion
  - ✓ Gray models

X-CAN

Cross-penalized Component ANalysis

SSC'16, June 2019

José Camacho, Ph.D. (josecamacho@ugr.es)

http://wpd.ugr.es/~josecamacho/

Joint work with E. Acar, M. Rasmussen and R. Bro