

# Sparse Generalized Singular Value Decomposition

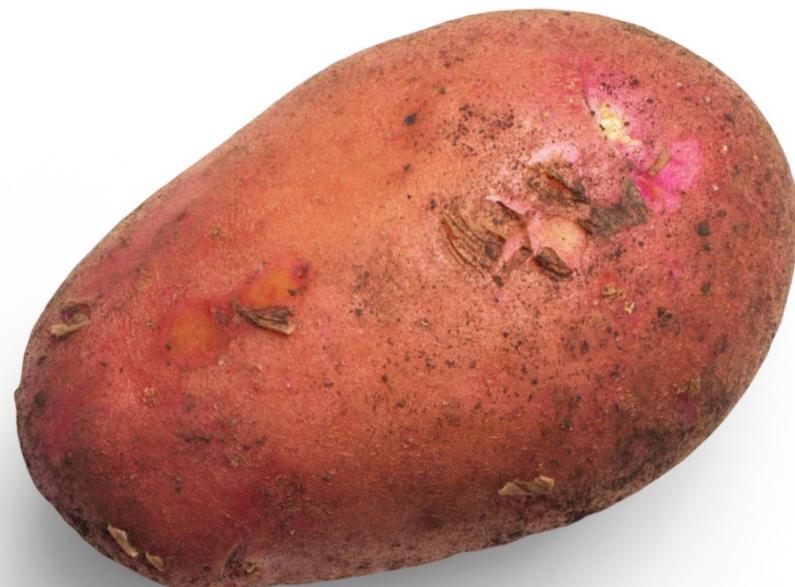
Vincent Guillemot, Vincent Le Goff, Ju-Chi Yu, & Hervé Abdi

---

# SOME MENTAL IMAGES

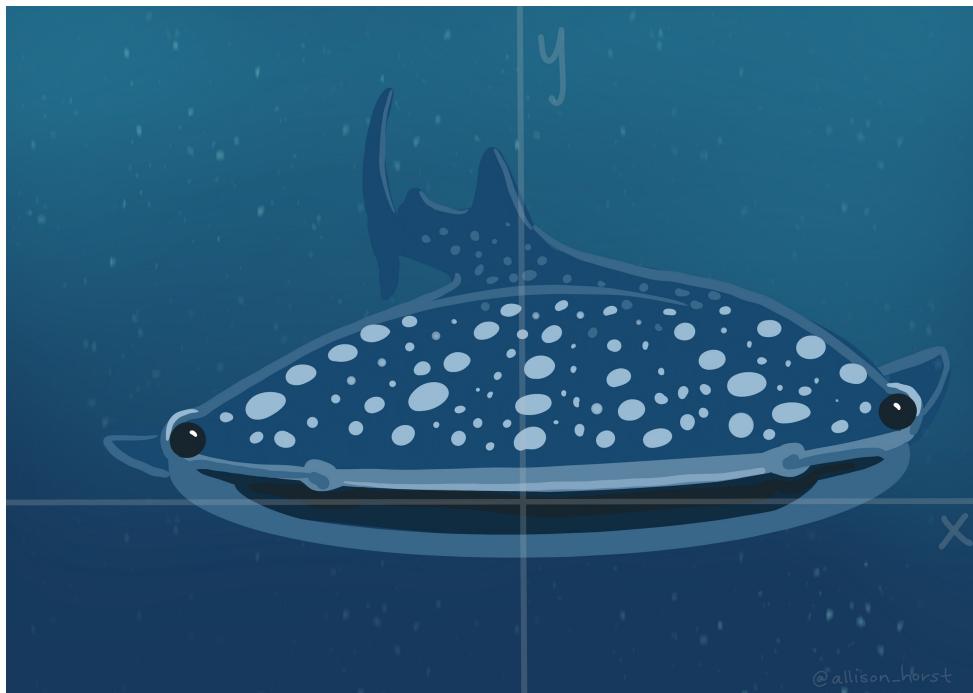


# Potato Chips Analysis



Cut the yummiest French fries

# Whale versus krill: this is you



Eat the most krill (put on your 3D glasses)

Artwork by @allison\_horst [https://twitter.com/allison\\_horst](https://twitter.com/allison_horst)

# Whale versus krill: this is your data



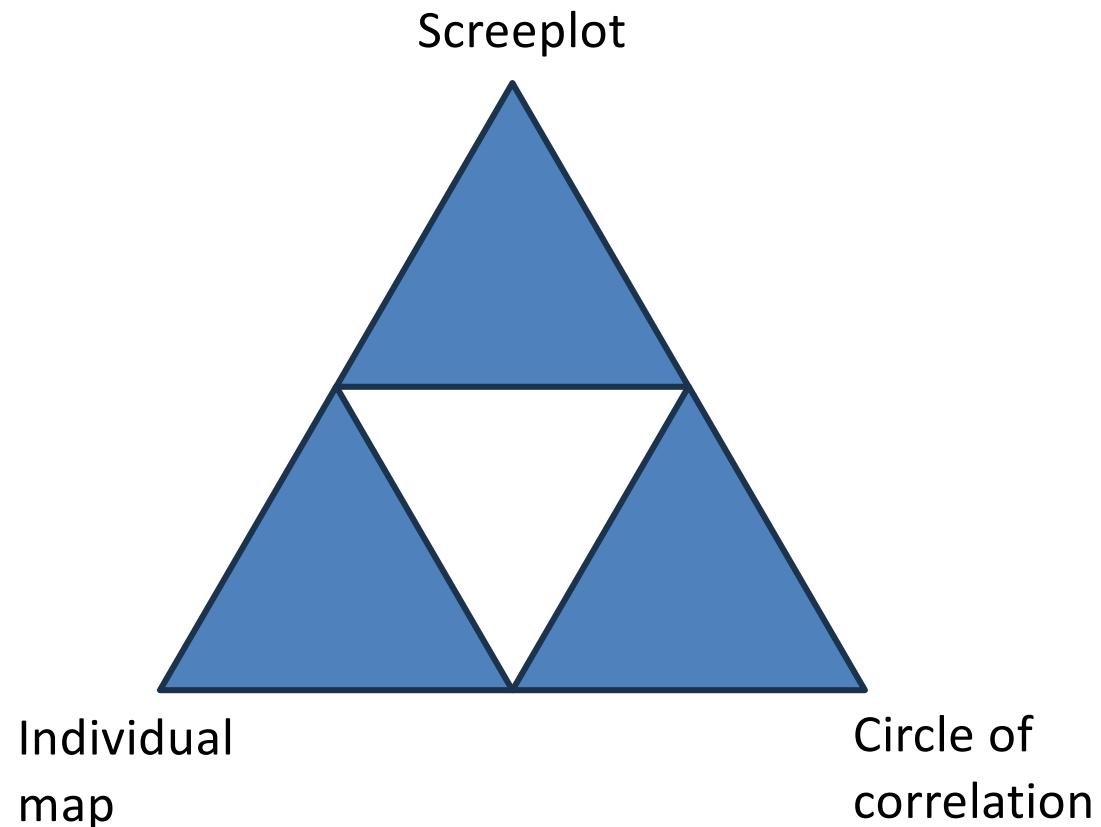
Eat the most krill (put on your 3D glasses)

Artwork by @allison\_horst [https://twitter.com/allison\\_horst](https://twitter.com/allison_horst)

# **THE TRI-FORCE OF PCA**



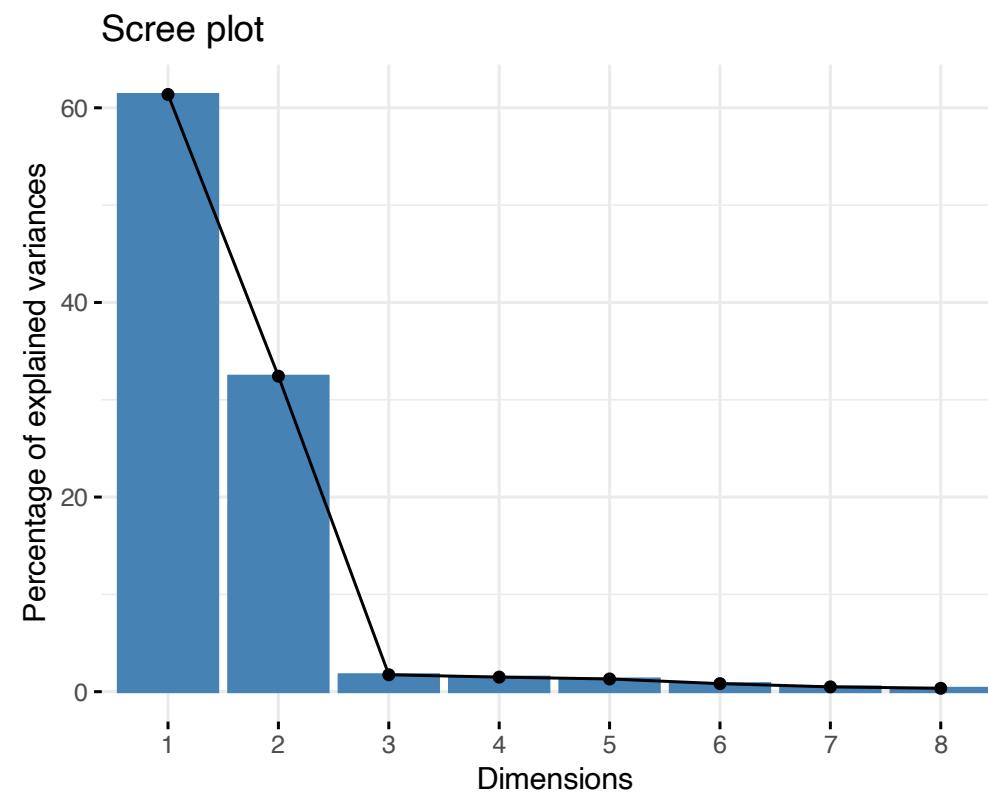
# The tri-force of PCA



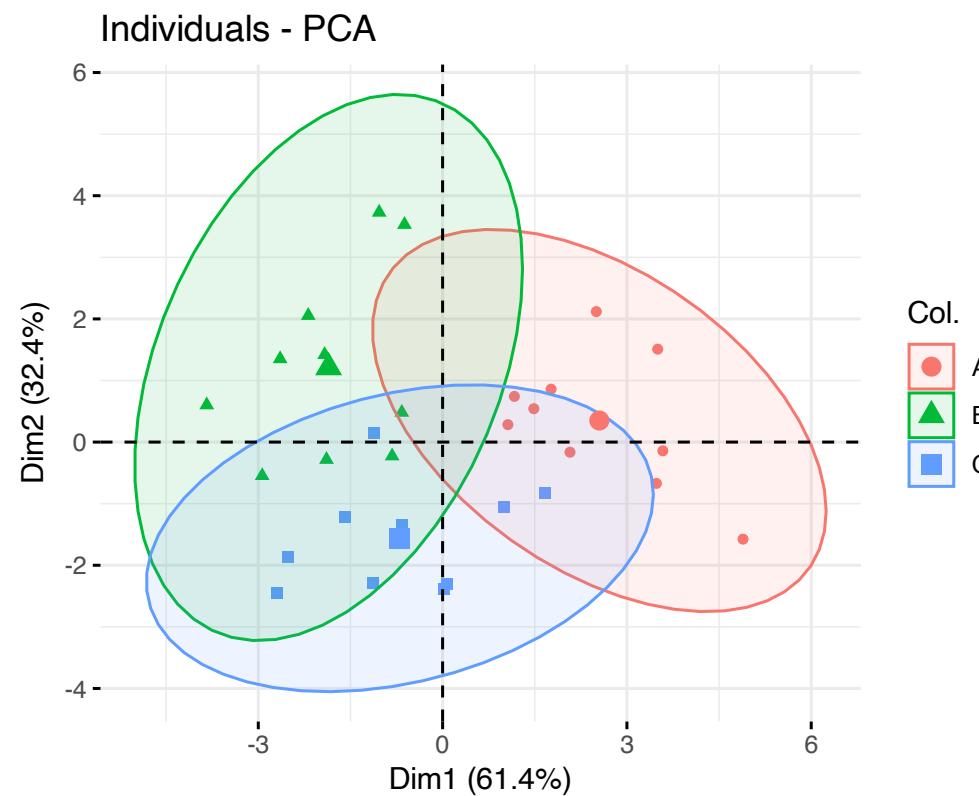
# Example data

```
library(readxl)
xtmp <- read_excel("simul.xlsx")
x <- as.matrix(xtmp[, -1])
rownames(x) <- xtmp$Ind
```

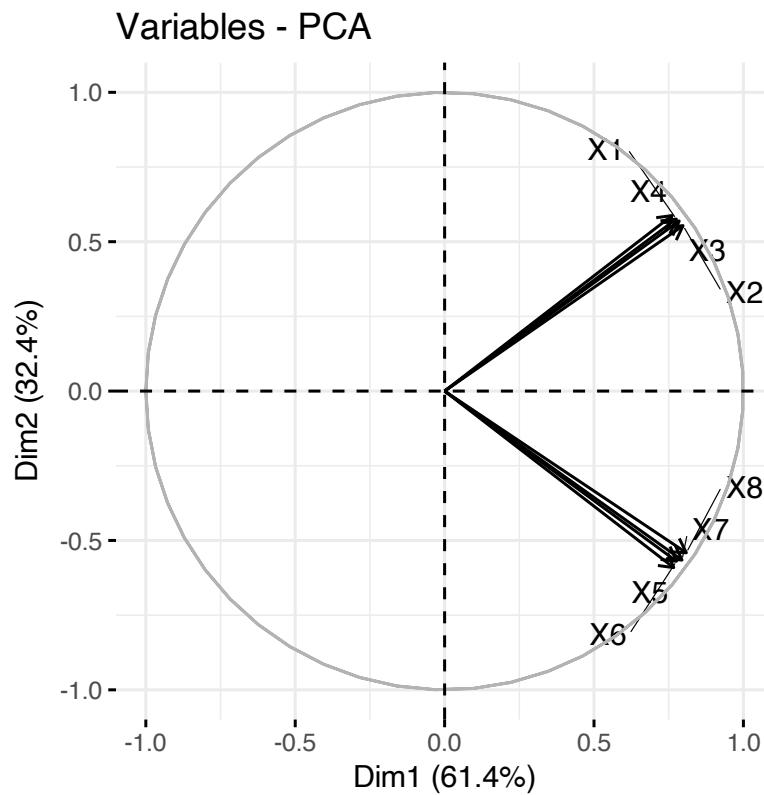
# Example screeplot



# Example individual map



# Example circle of correlation



# VOCABULARY



# French versus English

---

“Aaaaah, mais  
acépé en fait c'est  
la pisci-aïe !”

*(Anonymous student, after 6 hours of  
teaching PCA in French)*

English	French
PCA = principal component analysis	ACP = analyse en composantes principales
SVD = singular value decomposition	SVD = décomposition en valeurs singulières
EVD = eigenvalue decomposition	décomposition en éléments propres
ICA = independent component analysis	ICA = analyse en composantes indépendantes
MDS = multidimensional scaling	MDS = multidimensional scaling or analyse du triple

# R vocabulary

Base methods:

- `eigen` for eigenvalue decomposition, `svd` for singular value decomposition,
- `prcomp` and `princomp` for PCA,
- `biplot`

Nice packages:

- FactoMineR: PCA, MFA, CA, MCA and associates. In earlier versions, the graphs were “crude”...
- factoextra: “helper” package to make beautiful plots, and much more!
- ade4: more than “one block” type of analyses. Made by ecologists so ⇒ PCOA, coinertia analysis, STATIS, etc.
- ExPosition or TExPosition : made for psychometricians (they like PLS)

# And a few nice books and papers

- MOOC multivariate data analysis by François Husson:  
[https://husson.github.io/MOOC\\_AnaDo/index.html](https://husson.github.io/MOOC_AnaDo/index.html) (FR/EN)
- PCA paper(s) by Hervé Abdi:  
<https://personal.utdallas.edu/~herve/abdi-awPCA2010.pdf>

## A LITTLE BIT OF MATH

$x+h$

Tangent  
line

Lines

$$y = g(x)$$

$$\begin{aligned}f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\&= \lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} \\&= \lim_{h \rightarrow 0} \frac{x^2 + 2xh + h^2 - x^2}{h} \\&= \lim_{h \rightarrow 0} \frac{2xh + h^2}{h}\end{aligned}$$

$$\begin{aligned}&= \lim_{h \rightarrow 0} h(\sqrt{x+h} + \sqrt{x}) \\&= \lim_{h \rightarrow 0} \frac{1}{\sqrt{x+h} + \sqrt{x}} \\&= \frac{1}{2\sqrt{x}}\end{aligned}$$

$$\begin{aligned}f(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\f(a) &= \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}\end{aligned}$$

# Notations

(non-universal) Conventions: matrices and vectors are **bold**

- $I$  = number of observations,  $J$  = number of variables (only quantitative)
- $i$  for an individual observation, and  $j$  for a single variable
- $\mathbf{X}$  = data matrix, with  $I$  rows and  $J$  columns, sometimes already centered, and scaled, to make our life easy
- $\mathbf{x}_j$  = variable  $j$ , and  $j$ th column of  $\mathbf{X}$
- $\mathbf{v}$  a set of  $J$  coefficients
- $\mathbf{M}$ , masses for the individuals, and  $\mathbf{W}$ , weights for the variables

## A little detour: matrix multiplication

Take a pen and paper (or R), and do this multiplication:

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 2 & 3 & 4 \\ -1 & 0 & 0 & 1 \end{bmatrix}$$

[Cool video: 5 ways to see matrix multiplication](#)

# PCA

“Find a linear combination of the columns of the data that would capture the most information.”

In mathematical words, find

$$\mathbf{Xv} = v_1 \mathbf{x}_1 + \cdots + v_J \mathbf{x}_J$$

that maximizes... wait a minute! What are the dimensions?

- $\mathbf{X}$ :  $I$  rows and  $J$  columns,
- $\mathbf{v}$ :  $J$  rows and 1 column,
- $\mathbf{Xv}$ :  $I$  rows and 1 column.

# THE MATHEMATICAL TRANSLATION OF THE INTUITIONS BEHIND PCA

# Most popular intuition of PCA: how does it translate?

“PCA creates a linear combination of variables that maximizes variance.”

$$\arg \max_{\|\mathbf{v}\|_2^2=1} \text{var}(\mathbf{Xv})$$

- Why  $\|\mathbf{v}\|_2^2 = 1$ ?
- Dirty trick:  $\text{var}(\mathbf{Xv}) = \mathbf{v}^\top \mathbf{X}^\top \mathbf{Xv}$

# Least “well-known” intuition of PCA: how does it translate?

“PCA creates a linear combination of variables that maximizes correlation.”

$$\operatorname{argmax}_{\mathbf{v}} \sum_{j=1}^J \operatorname{cor}(\mathbf{Xv}, \mathbf{x}_j)^2$$

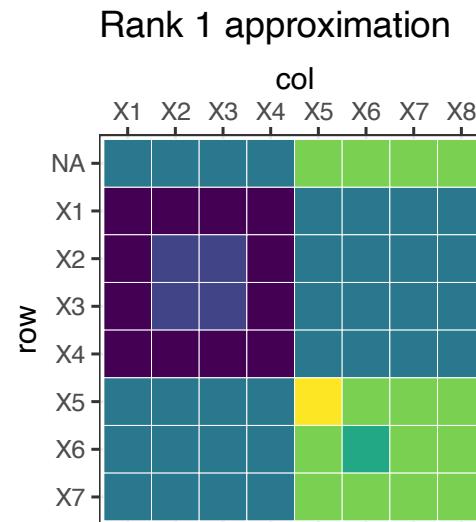
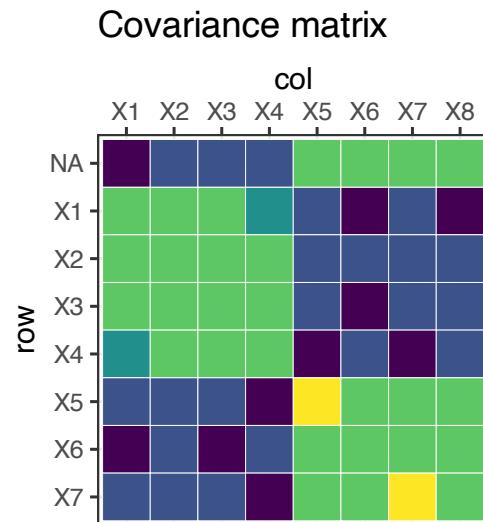
# Second least “well-known” intuition of PCA: how does it translate?

“PCA creates the best lower rank approximation of the covariance matrix.”

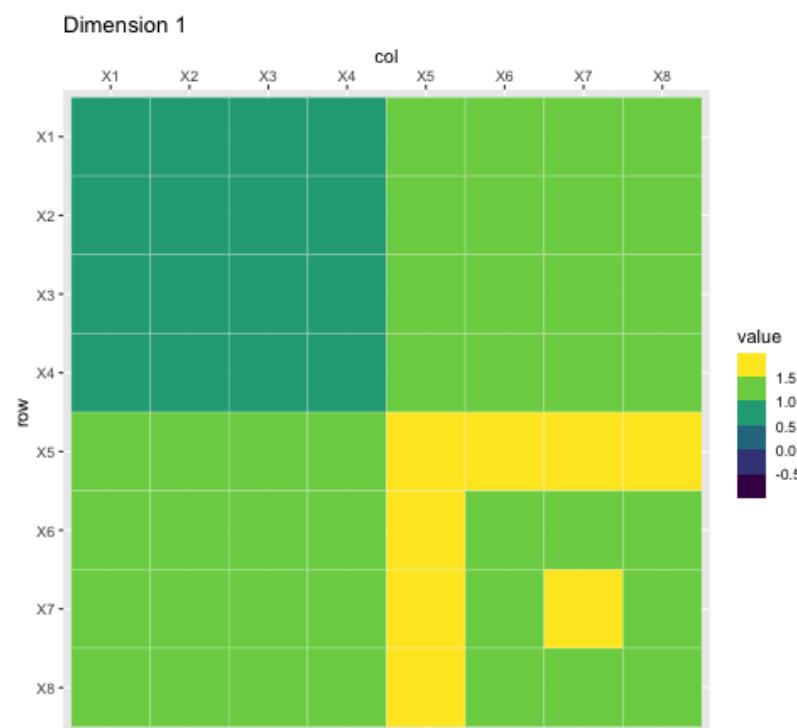
$$\arg \min_{\|v\|_2^2=1} \left\| \frac{1}{I} X^T X - \lambda v v^T \right\|_F^2$$

- $\frac{1}{I} X^T X$
- $\lambda$ : the [blank] of the covariance matrix
- $v$ : the [blank] of the covariance matrix

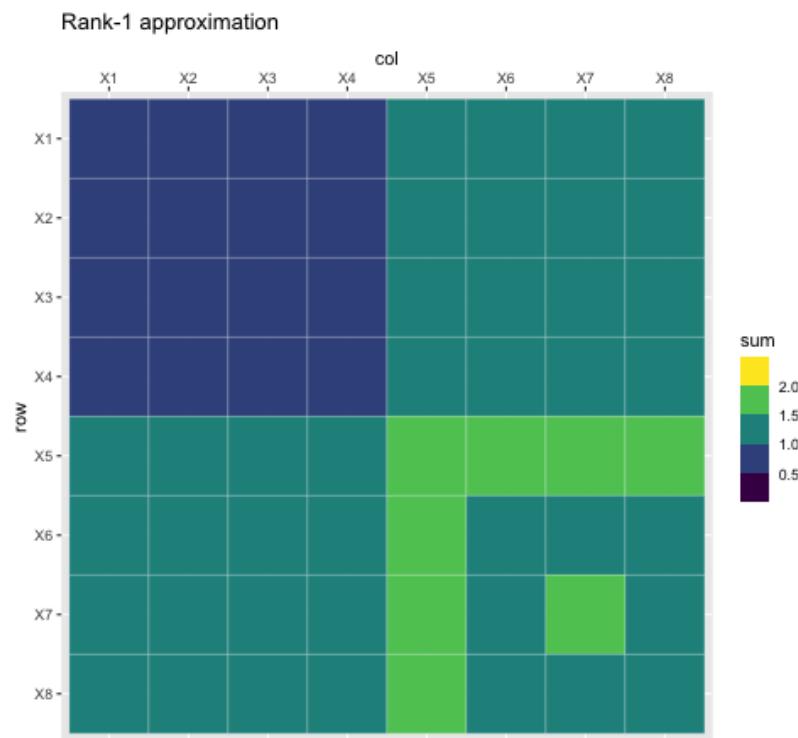
# A little image



# Rank-1 approximations



# Increasing rank approximations



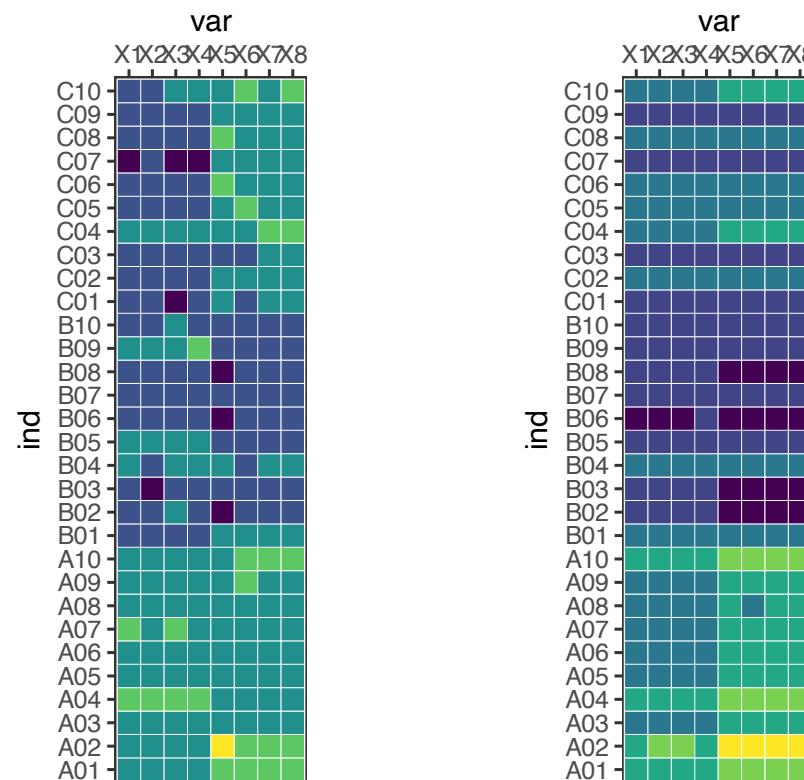
We can do the same kind of magic with the data itself

Singular value decomposition can be used to approximate a rectangular matrix with a lower ranked matrix of the same dimension.

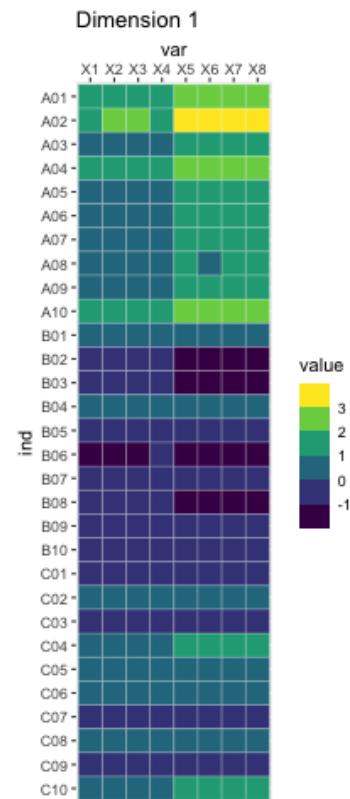
$$\arg \min_{\|u\|_2^2 = \|v\|_2^2 = 1} \|X - \delta uv^\top\|_F^2$$

- $\delta$ : singular value
- $u$ : left singular vector
- $v$ : right singular vector

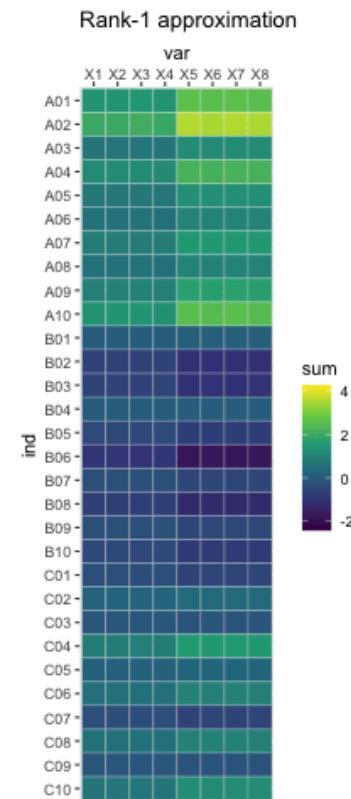
# Rank 1 approximation



# Rank-1 approximations



# Increasing rank approximations



# **CONSTRAINING THE (GENERALIZED) SVD**



# LASSO

LASSO is a (relatively) recent technique originally intended for regression problems:

$$\operatorname{argmin}_{\beta} \|y - X\beta\|_2^2 \text{ such that } \|\beta\|_1 \leq r$$

or the dual form

$$\operatorname{argmin}_{\beta} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

- the obtained weights are sparse (with zeros)
- the non-zeros coefficients correspond to important variables
- the result is biased
- selecting  $\lambda$  is done through cross-validation

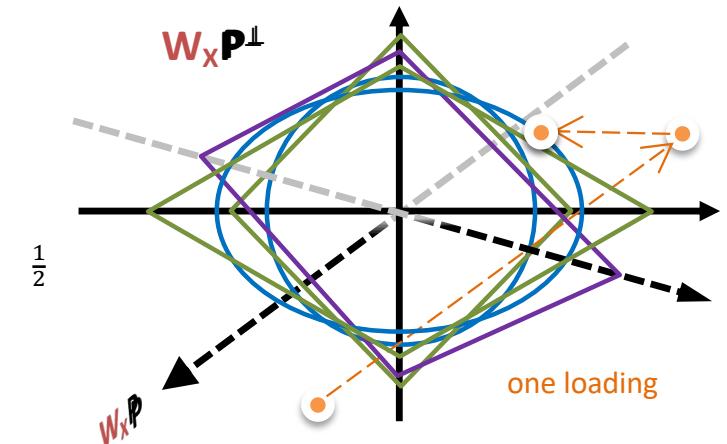
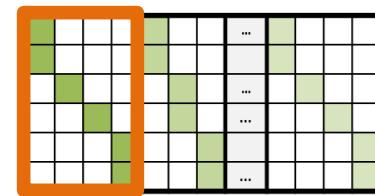
# Optimization problem

$$(\delta_\ell, \mathbf{p}_\ell, \mathbf{q}_\ell) = \underset{\beta}{\operatorname{argmin}} \|\mathbf{X} - \delta \mathbf{p}^\top \mathbf{q}\|_2^2$$

such that  $\begin{cases} \mathbf{p}^\top \mathbf{M} \mathbf{p} = \mathbf{q}^\top \mathbf{W} \mathbf{q} = 1 \\ \mathbf{p}^\top \mathbf{M} \mathbf{p}_{\ell'} = \mathbf{q}^\top \mathbf{W} \mathbf{q}_{\ell'} = 0, \forall \ell' < \ell \\ \|\mathbf{p}\|_1 \leq s_{\mathbf{p}, \ell} \text{ and } \|\mathbf{q}\|_1 \leq s_{\mathbf{q}, \ell} \end{cases}$

## Sparse GSVD (sGSVD) and sparse MCA (sMCA)

- Sparsify the GSVD: the sparse GSVD (sGSVD)
  - Intersection of 3 spaces
    - $L_1$ -ball
    - $L_2$ -ball with weights  $\mathbf{P}^T \mathbf{W}_x \mathbf{P} = \mathbf{I}$
    - Orthogonal space:
- Sparse MCA (sMCA)
  - Generalize  $L_1$ -constraint
  - Group constraint:
    - $L_{(1,2)}$ -norm: sum of  $L_2$ -norms
    - $L_2$ -norms:  $[(\text{Level 1 loading})^2 + (\text{Level 2 loading})^2 + \dots]$



# Optimization problem

$$(\delta_\ell, \mathbf{p}_\ell, \mathbf{q}_\ell) = \underset{\beta}{\operatorname{argmin}} \|\mathbf{X} - \delta \mathbf{p}^\top \mathbf{q}\|_2^2$$

such that  $\begin{cases} \mathbf{p}^\top \mathbf{M} \mathbf{p} = \mathbf{q}^\top \mathbf{W} \mathbf{q} = 1 \\ \mathbf{p}^\top \mathbf{M} \mathbf{p}_{\ell'} = \mathbf{q}^\top \mathbf{W} \mathbf{q}_{\ell'} = 0, \forall \ell' < \ell \\ \|\mathbf{p}\|_1 \leq s_{\mathbf{p}, \ell} \text{ and } \|\mathbf{q}\|_1 \leq s_{\mathbf{q}, \ell} \end{cases}$

# What are the parameters?

- $\mathbf{M}$  and  $\mathbf{W}$  = masses, weights... so metrics
- Number of dimensions
- $s_{\mathbf{p},\ell}$  and  $s_{\mathbf{q},\ell}$  = sparsity parameters, between 1 (strong sparsity), and  $\sqrt{\text{dimension}}$  (no sparsity)

# Lost & Found

- Transition formulas: from rows to columns and back: pseudo-version
- Supplementary projection: pseudo-inverse projector
- Asymmetric projection ( $\lambda = 1$ ): kept
- Distributional equivalence: kept
- Nested Solutions (i.e.,  $\mathbf{X}$  vs.  $\mathbf{X} - \mathbf{r}\mathbf{c}^T$ ) and ( $\lambda_1 = 1$ ): Lost

# A few fun papers

We belong to the fan club of: (Witten, Tibshirani, and Hastie 2009; Trendafilov 2014, Journee2010).

Our work on:

- Constrained Singular Value Decomposition (Guillemot et al. 2019)
- Sparse Correspondence Analysis (Abdi et al. 2024)
- Sparse Multiple Correspondence Analysis (Guillemot et al. 2020; Yu et al. 2024)

# References

- Abdi, Hervé, Vincent Guillemot, Ruiping Liu, Ndèye Niang, Gilbert Saporta, and Ju-chi Yu. 2024. “From Plain to Sparse Correspondence Analysis: a Generalized SVD Approach.” *Statistica Applicata - Italian Journal of Applied Statistics* 35 (3): 1–39. <https://doi.org/10.26398/IJAS.0035-014>.
- Guillemot, Vincent, Derek Beaton, Arnaud Gloaguen, Tommy Löfstedt, Brian Levine, Nicolas Raymond, Arthur Tenenhaus, and Hervé Abdi. 2019. “A constrained singular value decomposition method that integrates sparsity and orthogonality.” Edited by Shyamal D Peddada. *PLOS ONE* 14 (3): e0211463. <https://doi.org/10.1371/journal.pone.0211463>.
- Guillemot, Vincent, Julie Le Borgne, Arnaud Gloaguen, Arthur Tenenhaus, Gilbert Saporta, Sylvie Chollet, Derek Beaton, and Hervé Abdi. 2020. “Sparse Multiple Correspondence Analysis.” In *52èmes Journées de Statistique*, 830–35. 52èmes Journées de Statistiques de La Société Française de Statistique (SFdS). Nice, France: Société Française de Statistique (SFdS). <https://pasteur.hal.science/pasteur-03037346>.
- Trendafilov, Nickolay T. 2014. “From Simple Structure to Sparse Components: A Review.” *Computational Statistics* 29 (3-4): 431–54.
- Witten, Daniela M, Robert Tibshirani, and Trevor Hastie. 2009. “A Penalized Matrix Decomposition, with Applications to Sparse Principal Components and Canonical Correlation Analysis.” *Biostatistics* 10 (3): 515–34.
- Yu, Ju-Chi, Julie Le Borgne, Anjali Krishnan, Arnaud Gloaguen, Cheng-Ta Yang, Laura A. Rabin, Hervé Abdi, and Vincent Guillemot. 2024. “Sparse Factor Analysis for Categorical Data with the Group-Sparse Generalized Singular Value Decomposition.” *Computational Statistics and Data Analysis*.