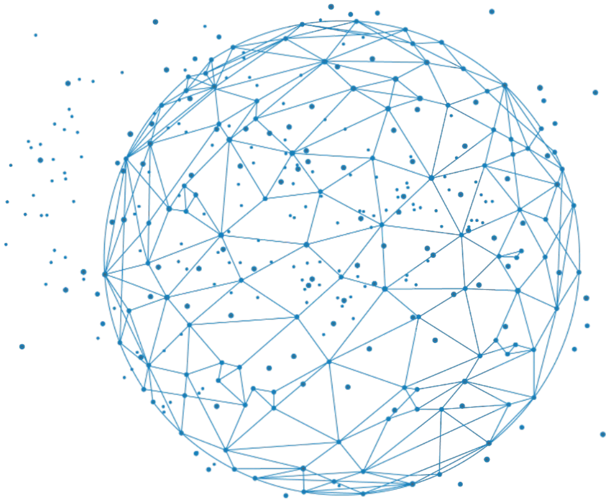
Second edition 2024 in Fréjus



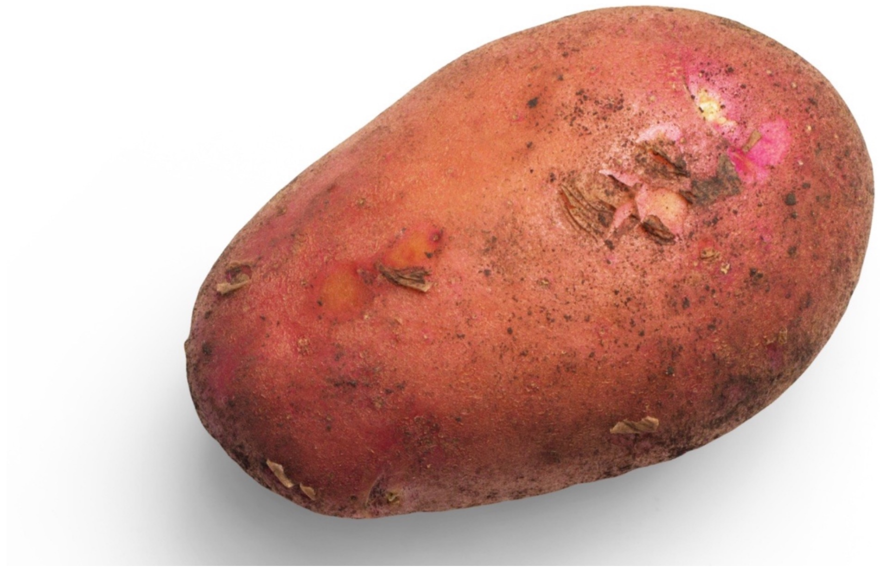
Principal Component Analysis and Singular Value Decomposition

Vincent Guillemot

DOI version final

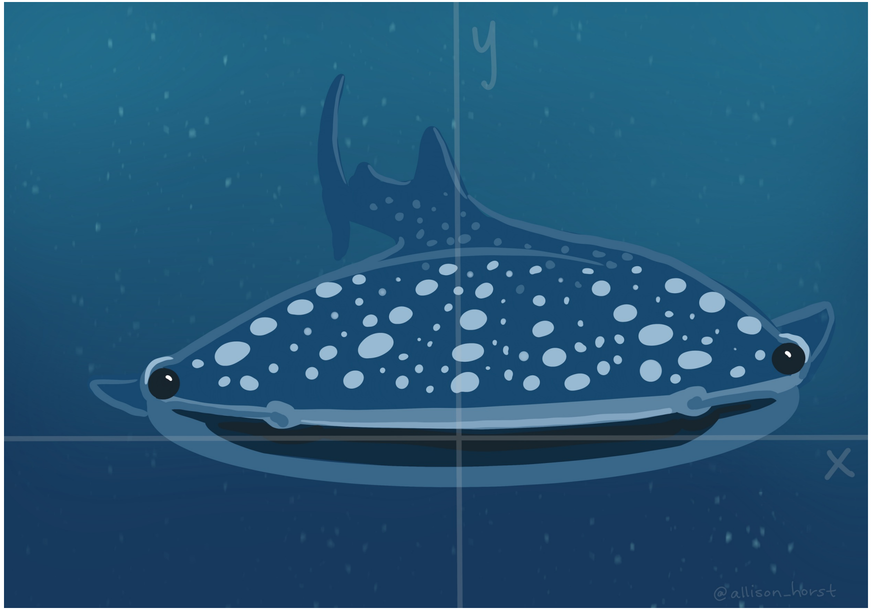
Some mental images

Potato Chips Analysis



Cut the yummiest French fries

Whale versus krill: this is you (credit: Allison Horst)



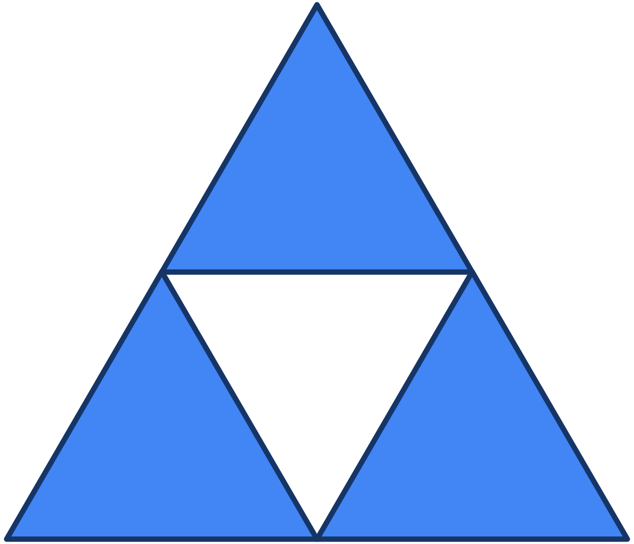
Eat the most krill (put on your 3D glasses)

Whale versus krill: this is your data (credit: Allison Horst)

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

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

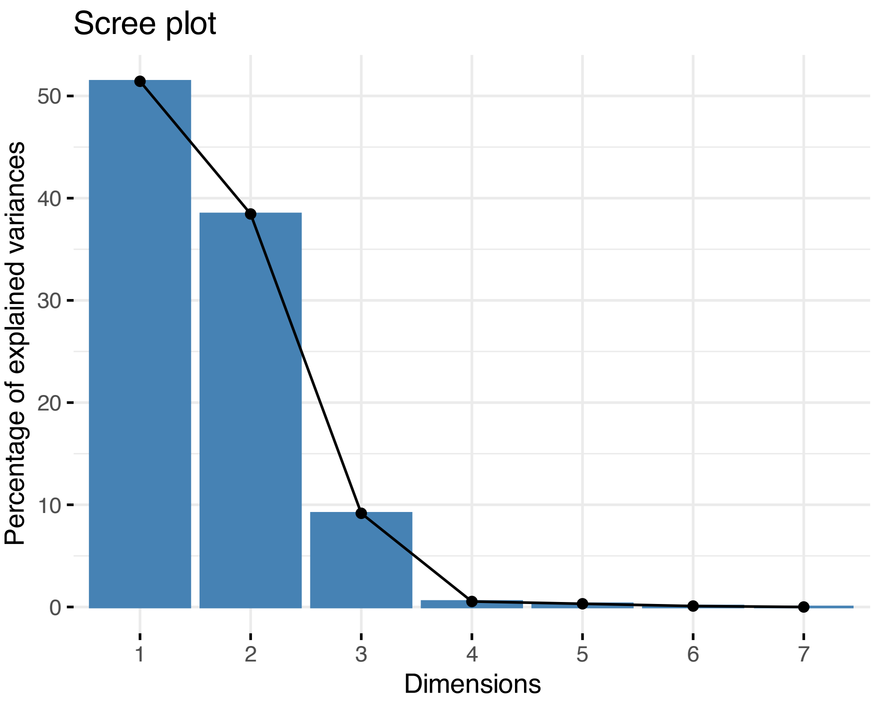
The tri-force of PCA



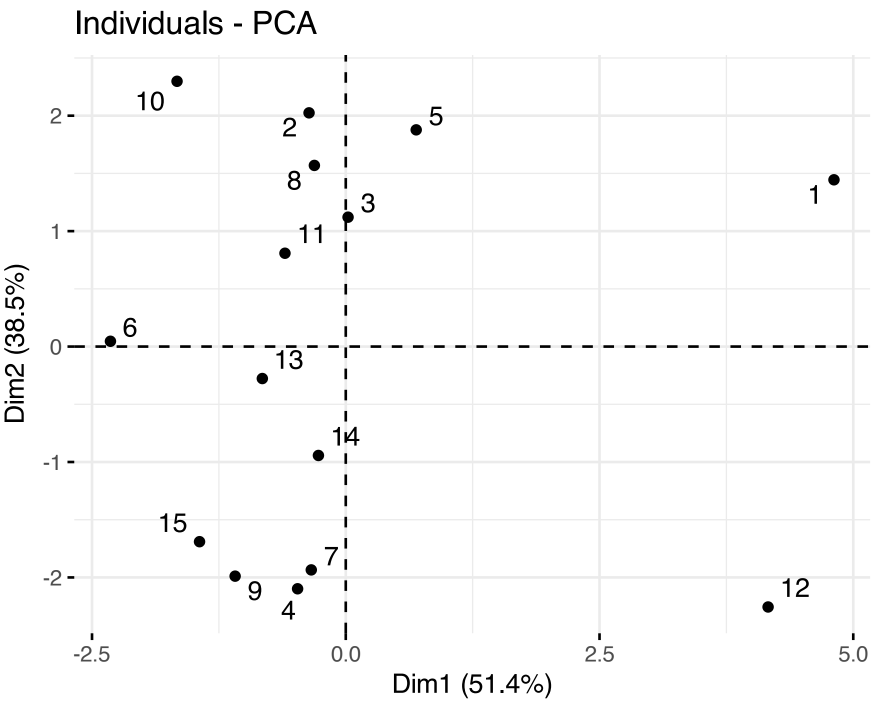
Example data

```
n <- 15
dat.ex <- tibble(
  X1 = rnorm(n),
  X2 = rnorm(n),
  X3 = -X1,
  X4 = 2 * X2  + 0.25 * rnorm(n),
  X5 = X1 + X2 + 0.25 * rnorm(n),
  X6 = X1 - X2 + 0.25 * rnorm(n),
  X7 = rnorm(n)
)
```

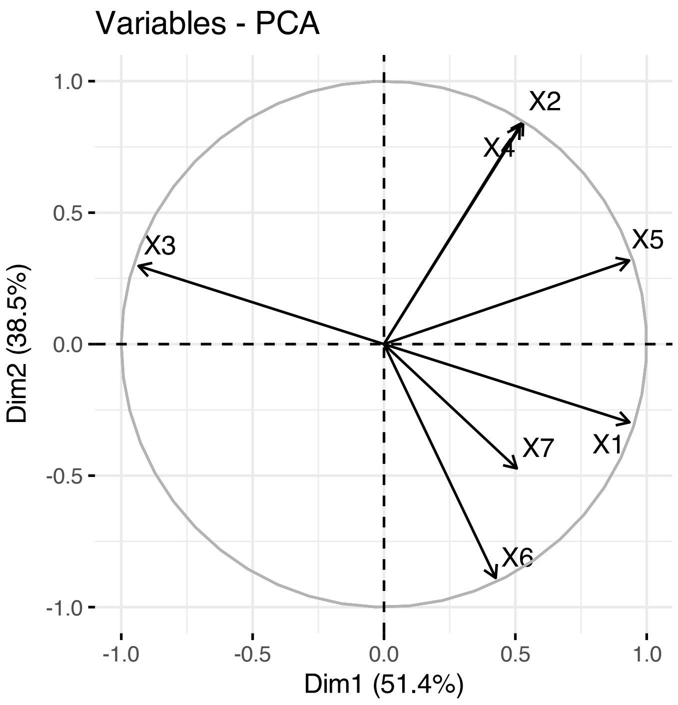
Example screeplot



Example individual map



Example circle of correlation



Vocabulary

French versus English

"Aaaaah, mais acépé en fait c'est la PCA !"

(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

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`: made for psychometricians (they like PLS)

And a few nice books and papers

MOOC analyse de données de François Husson :
https://husson.github.io/MOOC_AnaDo/index.html
(also in English)
PCA paper(s) by Hervé Abdi:
https://personal.utdallas.edu/~herve/abdi-awPCA2010.pdf

(more?)

A little bit of Math

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

Notations

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

- n = number of observations, p = number of variables (only quantitative)
- i for an individual observation, and j for a single variable
- $\mathbf{X}$ = data matrix, with n rows and p columns, sometimes already centered, and scaled, to make our life easy
- $\mathbf{X}_j$ = variable j, and jth column of $\mathbf{X}$
- $\mathbf{w}$ a set of weights

A little detour: matrix multiplication

Take a pen and paper, 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

$$\mathbf{Xw} = w_1\mathbf{X}_1 + \cdots + w_p\mathbf{X}_p$$

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

- $\mathbf{X}$: n rows and p columns,
- $\mathbf{w}$: p rows and 1 columns,
- $\mathbf{Xw}$: n rows and 1 column.

The mathematical translation of the intuitions behind PCA

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

ELIXIR FRANCE

Most popular intuition of PCA: how does it translate?

"PCA creates a linear combination of variables that maximizes variance."

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

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

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

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

$$\underset{\mathbf{w}}{\operatorname{argmax}} \sum_{j=1}^{p} \operatorname{cor}\left(\mathbf{X}\mathbf{w}, \mathbf{X}_j\right)^2$$

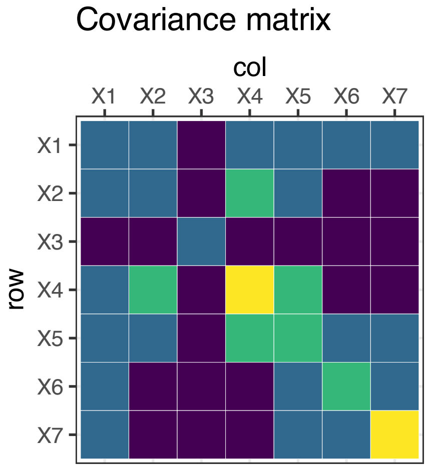
Second least “well-known” intuition of PCA: how does it translate?

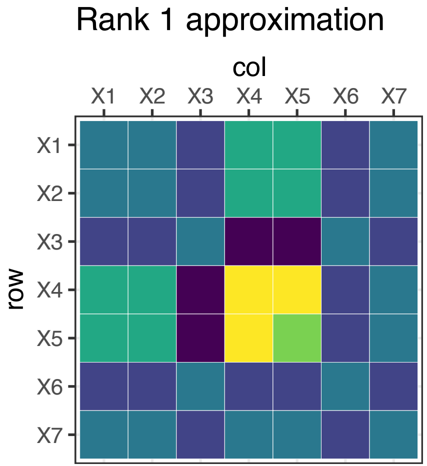
“PCA creates the best lower rank approximation of the covariance matrix.”

$$\arg\min_{\|w\|_2^2=1} \left\| \frac{1}{n}\mathbf{X}^\top\mathbf{X} - \lambda\mathbf{w}\mathbf{w}^\top \right\|_F^2$$

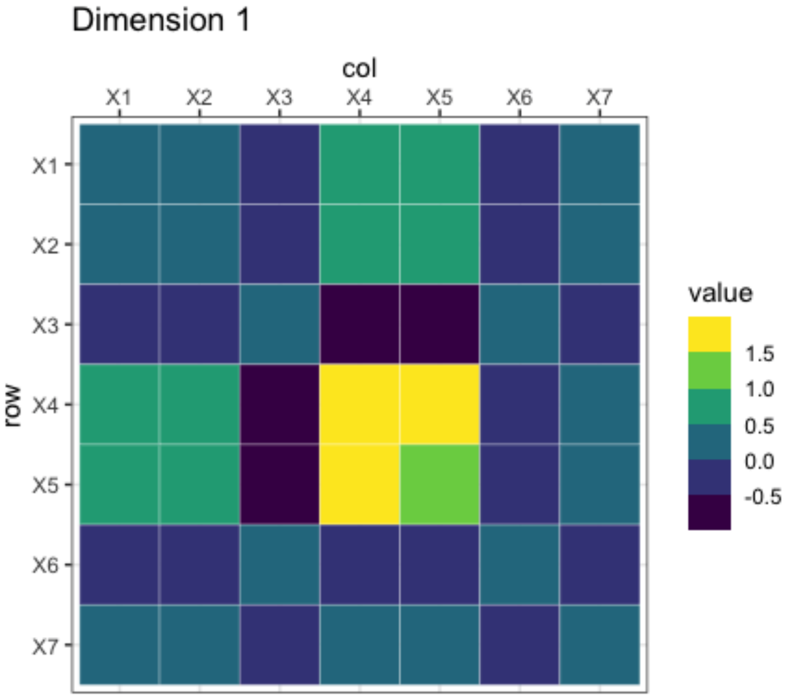
- $\frac{1}{n}\mathbf{X}^\top\mathbf{X}$
- λ: the [blank] of the covariance matrix
- $\mathbf{w}$: the [blank] of the covariance matrix

A little image

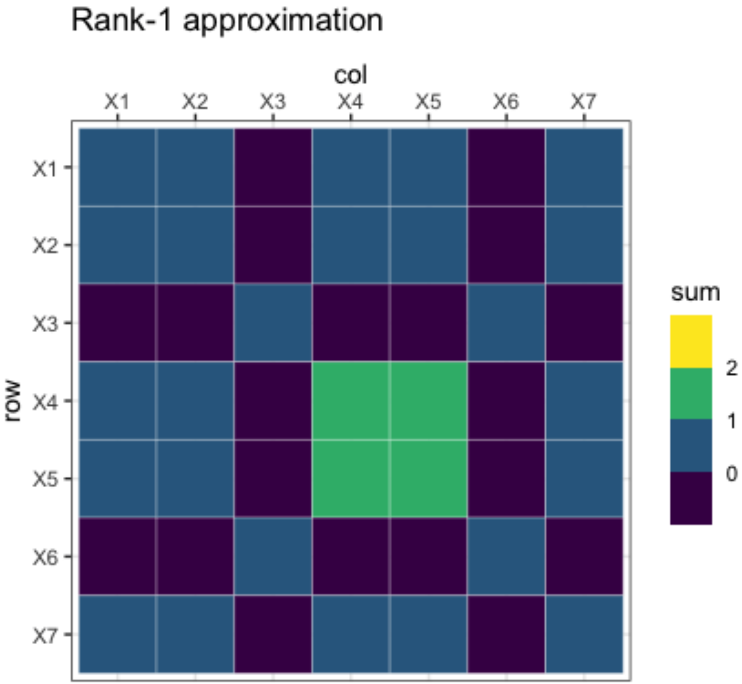




Rank-1 approximations



Increasing rank approximations



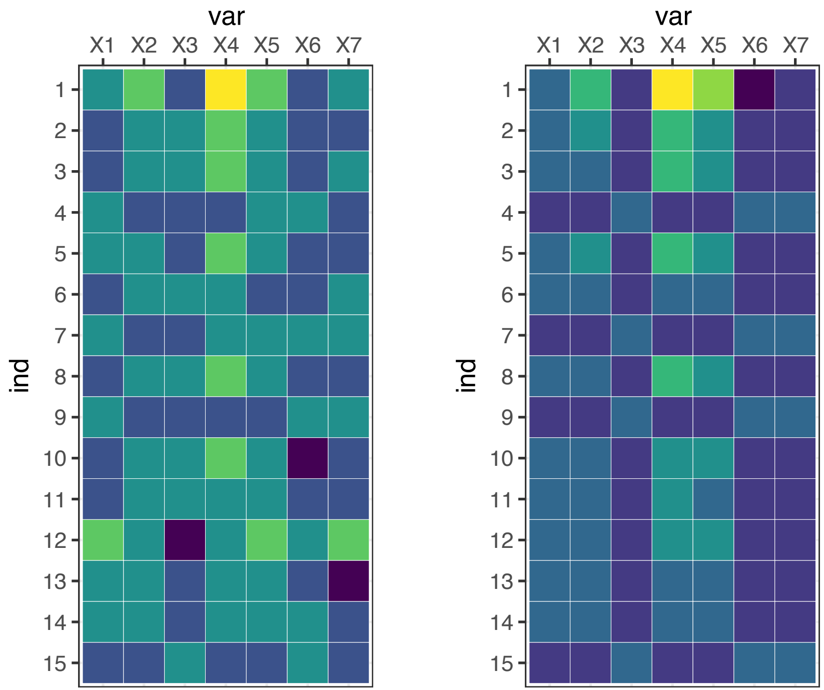
We can do the same kind of magic with the data itself

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

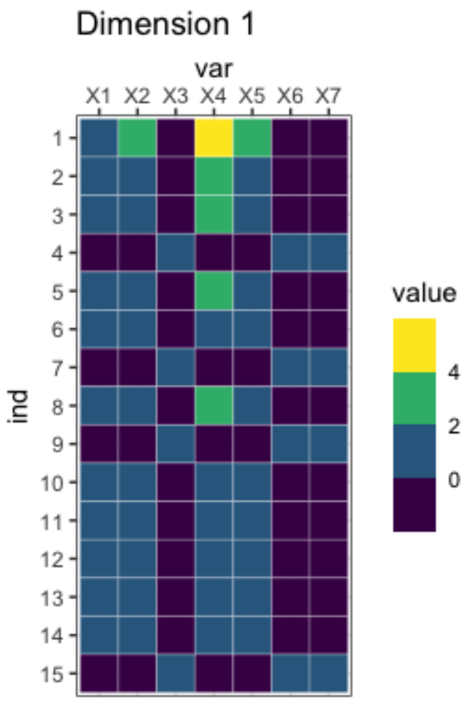
$$\arg \min_{\|\mathbf{u}\|_2^2 = \|\mathbf{w}\|_2^2 = 1} \|\mathbf{X} - \delta \mathbf{u} \mathbf{w}^\top\|_F^2$$

- δ: singular value
- $\mathbf{u}$: left singular vector
- $\mathbf{w}$: right singular vector

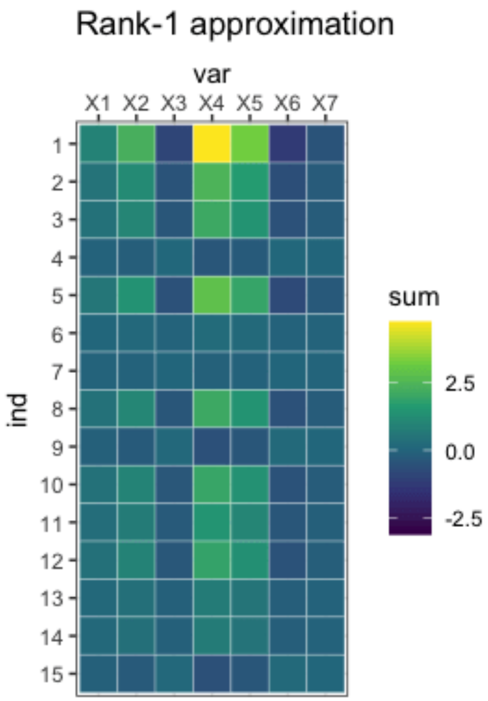
Rank 1 approximation



Rank-1 approximations



Increasing rank approximations



Un peu de code

```
library(dplyr)
library(FactoMineR)
library(factoextra)
library(ggplot2)


n <- 15
dat.ex <- tibble(
  X1 = rnorm(n),
  X2 = rnorm(n),
  X3 = -X1,
  X4 = 2 * X2  + 0.25 * rnorm(n),
  X5 = X1 + X2 + 0.25 * rnorm(n),
  X6 = X1 - X2 + 0.25 * rnorm(n),
  X7 = rnorm(n)
)
```

```
res.pca.ex <- PCA(dat.ex, scale.unit = TRUE, graph = FALSE)
fviz_screeplot(res.pca.ex)
fviz_pca_ind(res.pca.ex, repel = TRUE)
fviz_pca_var(res.pca.ex, repel = TRUE)
```

INSTITUT FRANÇAIS DE BIOINFORMATIQUE