

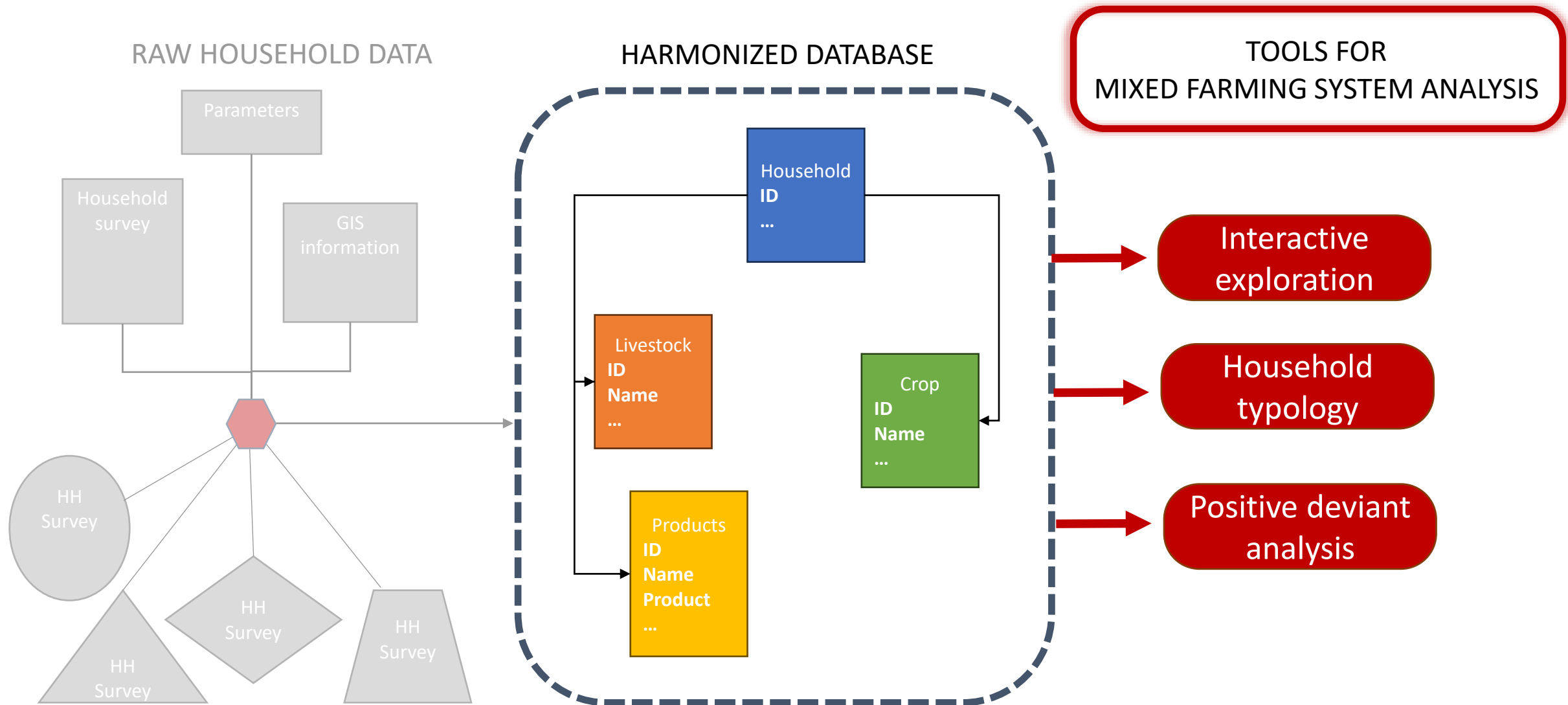
# Farm household Data analysis

Farmhousehold Workshop

2024

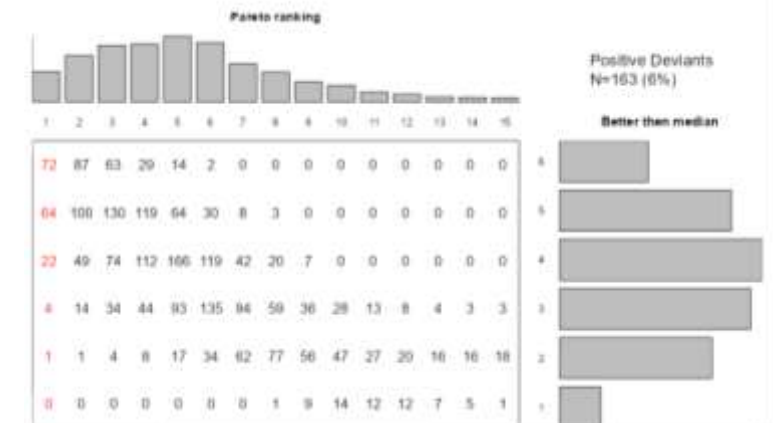
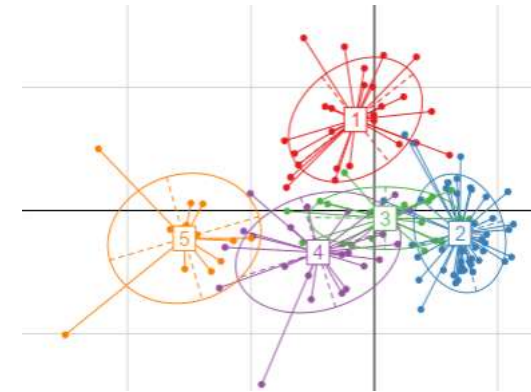
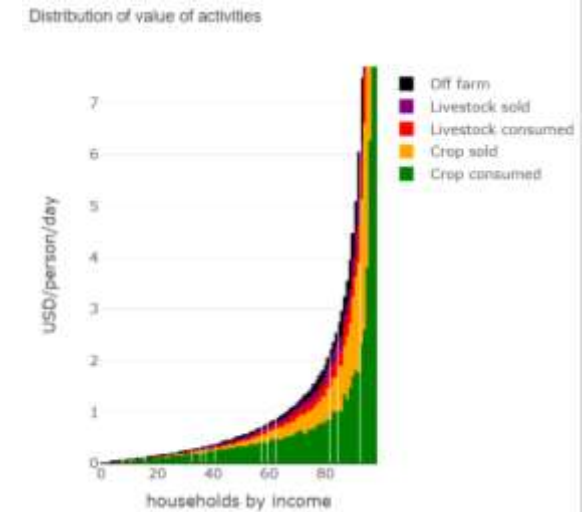
Romain Frelat

# Objectif



# Structure of the tutorial

1. Farm household data exploration
2. Household typology
3. Positive deviance analysis



# Requirement

One farmhousehold dataset as rds file

[https://github.com/rfrelat/farmhousehold/raw/main/inst/extdata/mini\\_rhomis.rds](https://github.com/rfrelat/farmhousehold/raw/main/inst/extdata/mini_rhomis.rds)

Open R-Studio

Make sure farmhousehold package is installed

```
> devtools::install_github("rfrelat/farmhousehold")
```

Ready?

# Farm household data exploration

# Quick overview

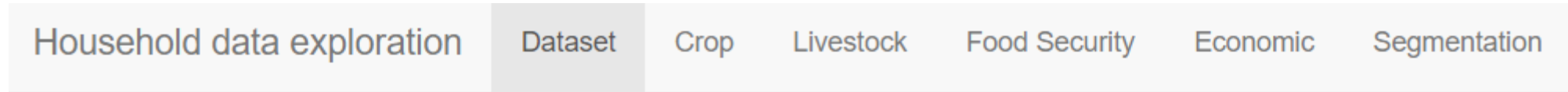


1. Load the data
2. See crop production
3. See livestock production
4. See food security status
5. See economic activities
6. Make groups
7. And go back and check the farm characteristics per group

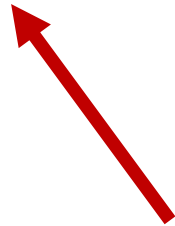
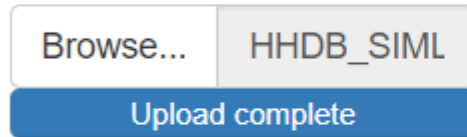
## Exploration steps

- A. Get to know your dataset (1 – 5)
- B. Make hypothesis and test them (6 – 7)

# Tips and tricks

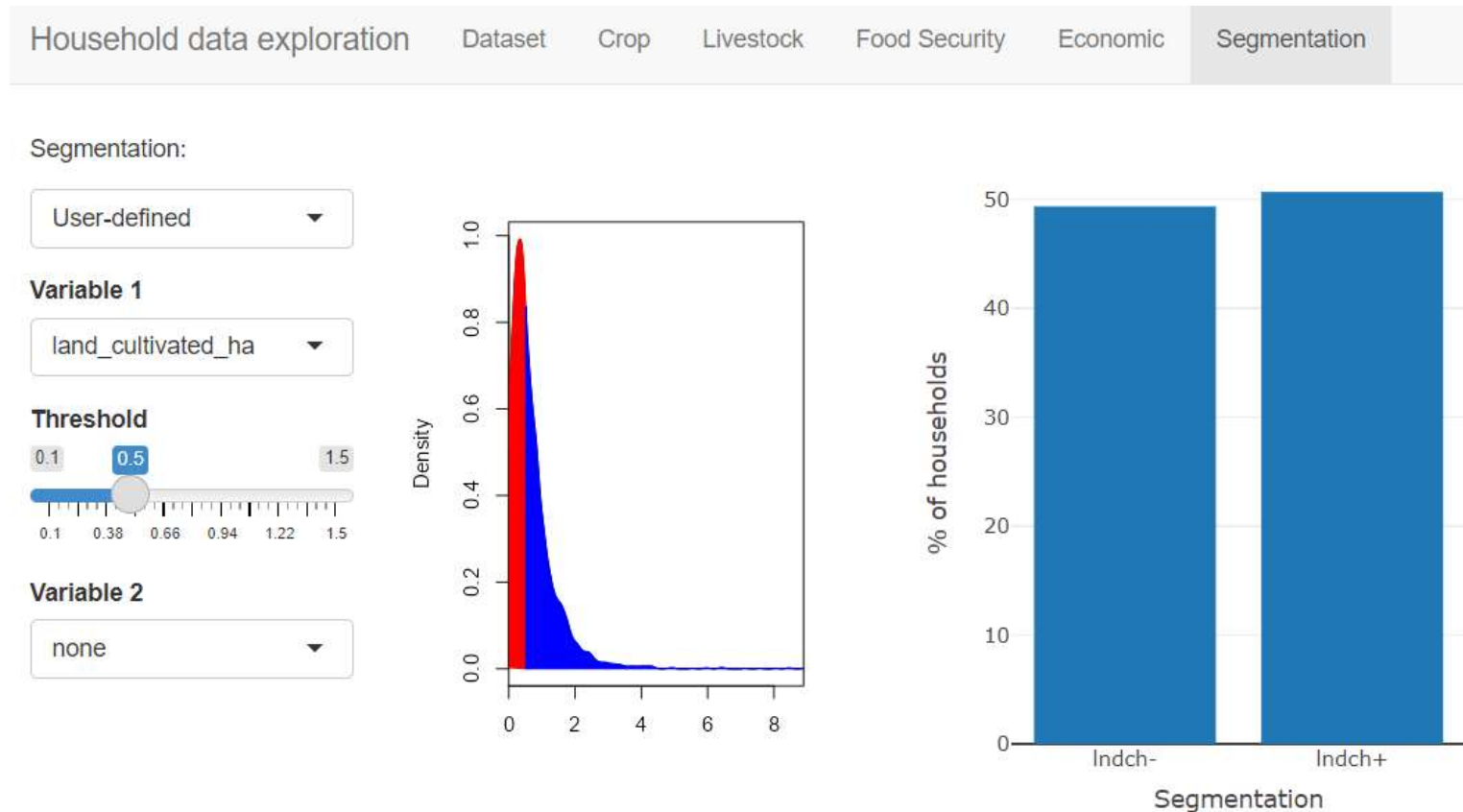


**Choose farmhousehold  
dataset**



Don't forget to click on Load RDS data, else nothing will happen.

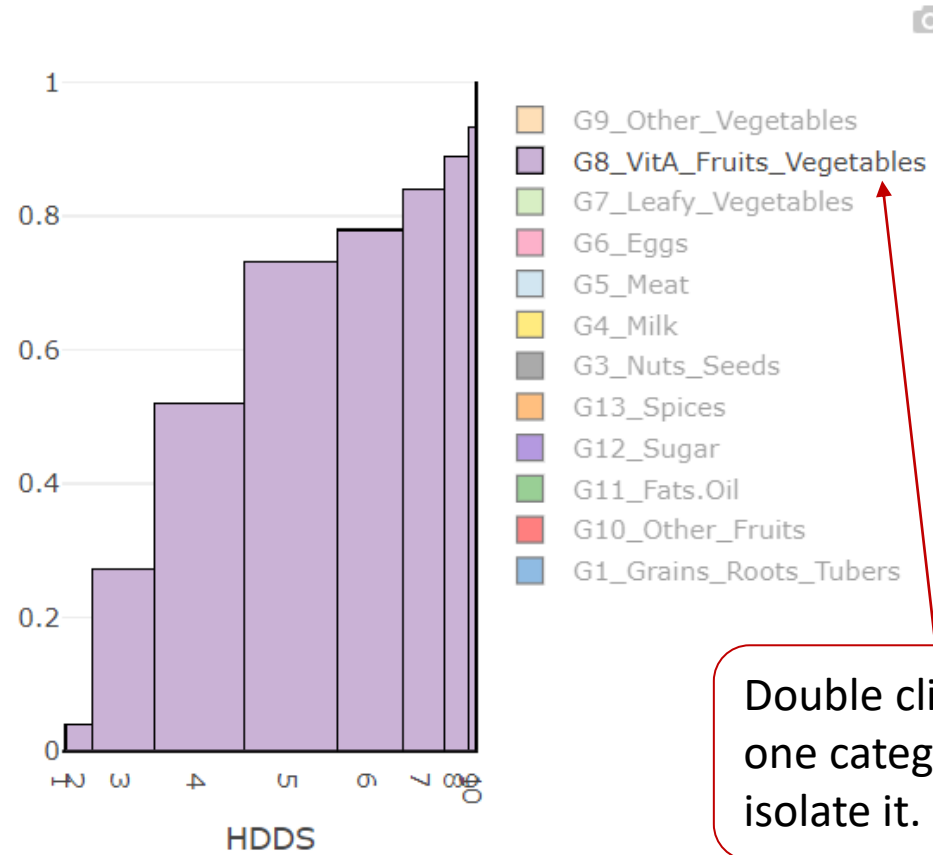
# Tips and tricks (2)



By default, the threshold selected in the segmentation panel correspond to the median. The label of the groups is an abbreviation of the variable name followed by (-) or (+).

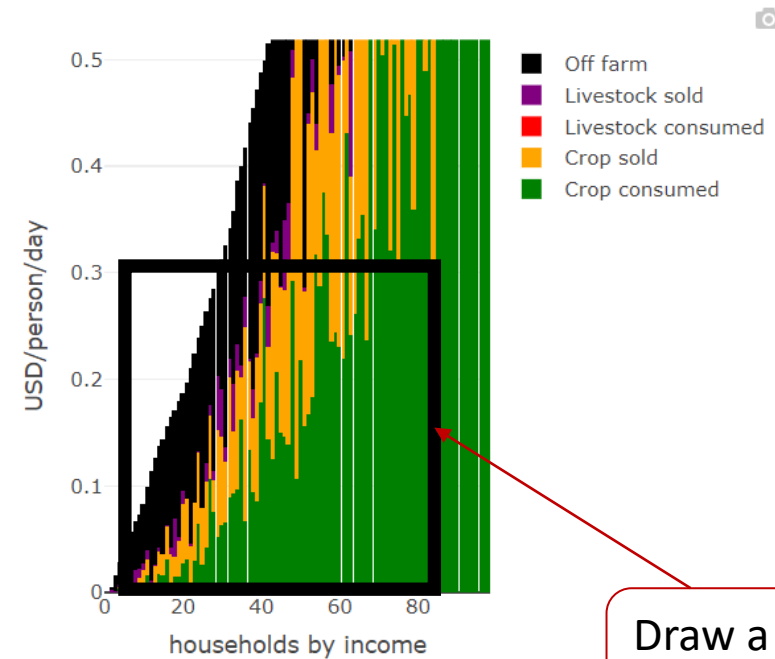


# Tips and tricks (3)



## Graphs are interactive!

Distribution of value of activities



# Your turn

```
> library(farmhousehold)

> runExplo()
```

Online version:

[https://startistic.shinyapps.io/farmhousehold\\_explo/](https://startistic.shinyapps.io/farmhousehold_explo/)

Questions that you can answer:

1. How many households are in the dataset?
2. What is the most important crop and livestock species?
3. Which food groups trigger changes between the diet diversity score of 5 and 6?
4. How many households do not report any income sources?
5. Are households with livestock more food secure?
6. Do crop species cultivated depend on the size of the land cultivated?
7. Are diverse farms more food secure? Or having higher number of sources of income?
8. Between land size and herd size, which is the most important driver of food security?
9. Do household size influence farm production or food security status?
10. What are the main differences among farm households in the different farming systems?

*Use your creativity and answer your own questions*

# Household typology

# Farm household typology

Capture the complexity and diversity of farming systems for:

- Targeting - which technology to which farmer?
- Scaling out – how a technology can be disseminated, to what extent?
- Selection – who are the representative farmers?

# Dataset step-by-step

## 1. Load data

Any csv file or farmhousehold rds file.

If large dataset ( $N > 10000$ ) you might want to subset your data

Don't forget to click on "Load data"

## 2. Select variables

Enough variables but not too correlated and normal-like distribution

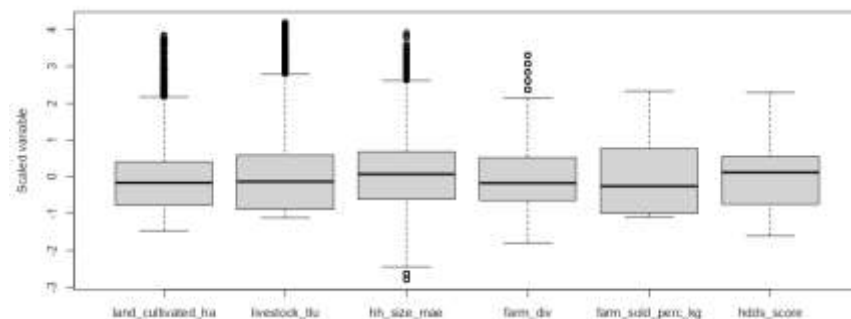
## 3. Outliers

Log transform right skewed variables and/or remove outliers

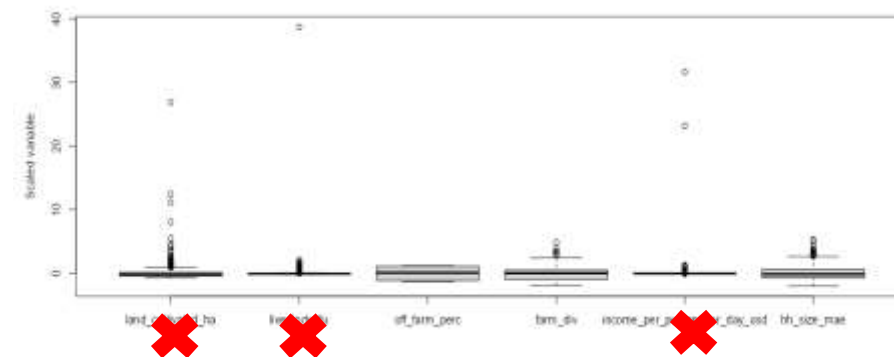
# Dataset

Box well spread around 0,  
no high outliers

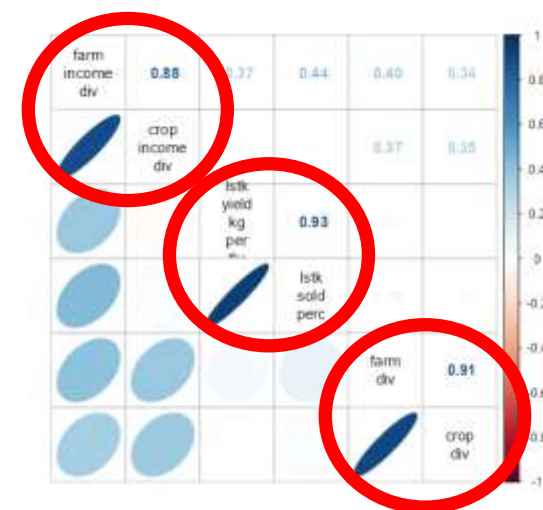
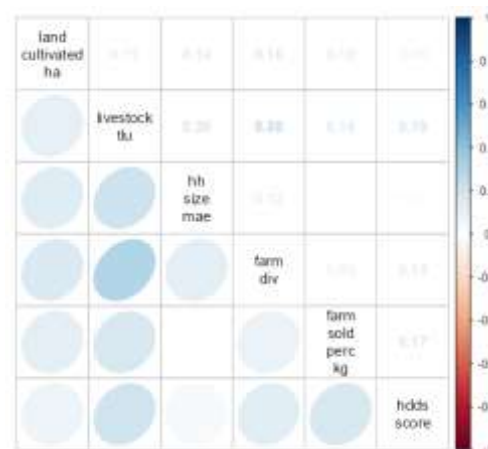
Ideal case



Worst case



No strong correlation  
 $r < 0,7$



# Analysis step-by-step

## 5. Multivariate analysis

Select the number of PCs that “best” summarize the data

## 6. Cluster analysis

Select either hierarchical clustering or ward’s criteria

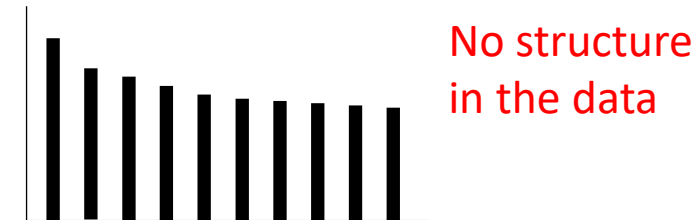
Select the number of cluster that is “best” based on the graph

The bending elbow rule  
(for PC and cluster selection)

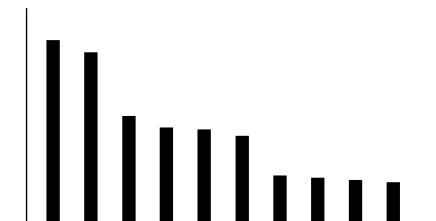
**Ideal case**



**Worst case**



**Normal case**

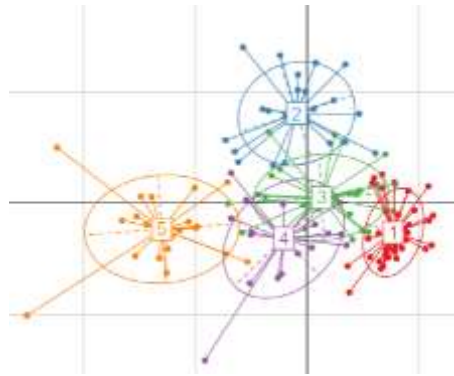


# Analysis

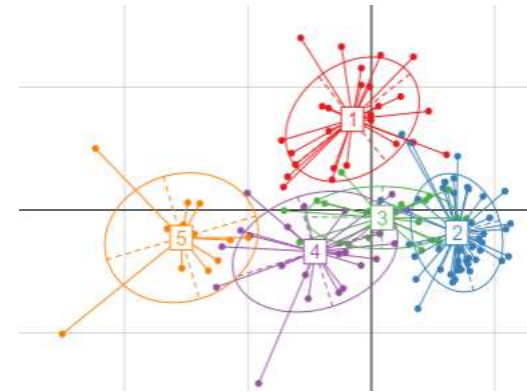
Ideally, a robust typology is:

- independent from the clustering method

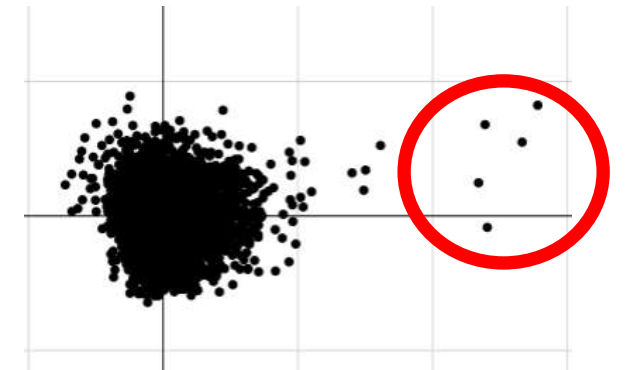
hierarchical



K-means



- independent from known geographic groups
- independent from a single variable or few outliers





# Your turn

```
> library(farmhousehold)  
  
> runCluster()
```

Online version:

[https://startistic.shinyapps.io/farmhousehold\\_cluster/](https://startistic.shinyapps.io/farmhousehold_cluster/)

Positive deviance analysis

# Problems it can solve

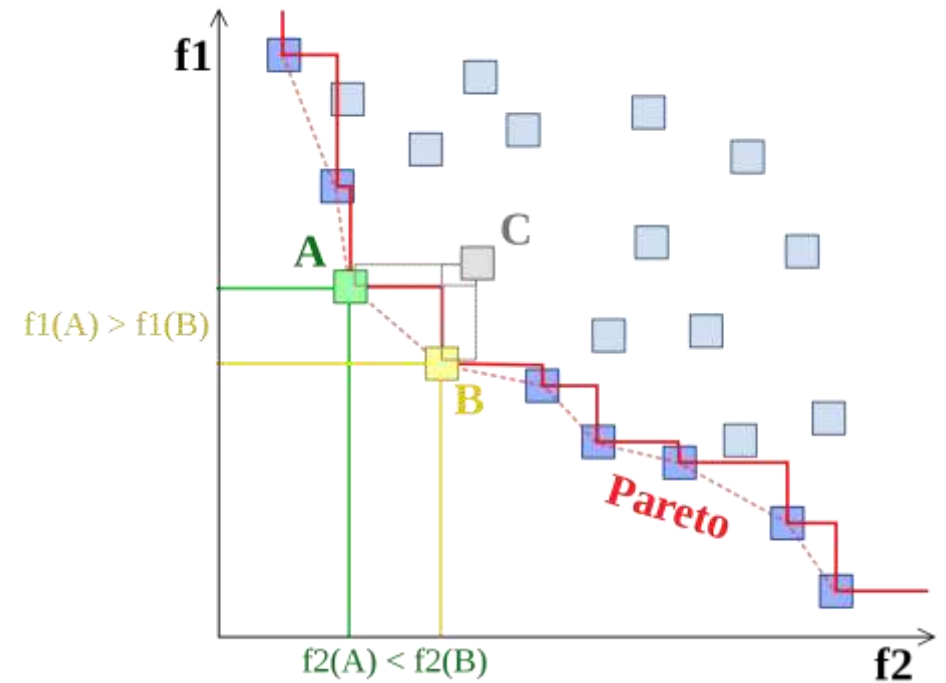
- “Identify farms with ‘outstanding’ **multi-functional performance**; and discuss the implications of the outstanding farms for the development of new systems of production” (Modernel *et al.* 2018)
- “From identifying exceptionally well-performing households to identifying **potentially superior practices**. [...] Provide suggestions on viable interventions, based on empirical, qualitative insights” (Steinke et al. 2019)
- “Support **redesign of farming systems**, with the aim to improve farm productive, economic and environmental performances” (Toorop et al. 2020)

# Pareto optimality

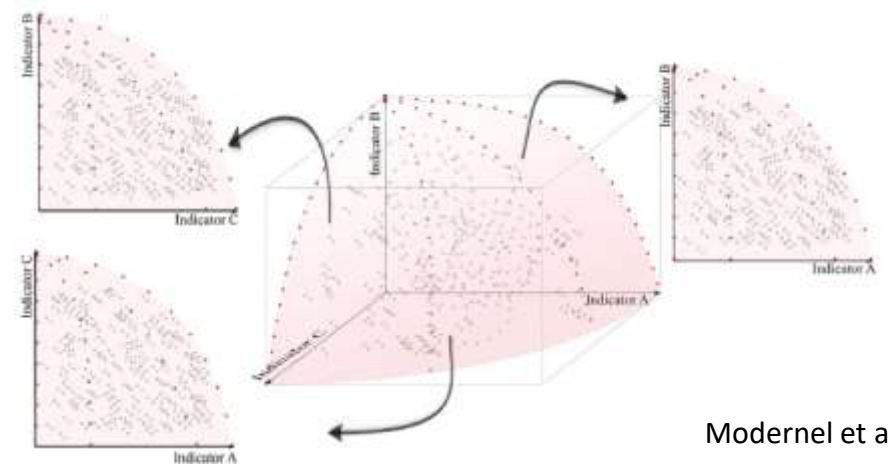
“A set of 'non-inferior' solutions in the objective space defining a boundary beyond which none of the objectives can be improved without sacrificing at least one of the other objectives”

Many different algorithms to solve this optimization problem.

In R, the function `eaf::pareto_rank()` use genetic algorithm: NSGA-II (Deb *et al.* 2002).



© Johann Dréo



Modernel et al. 2019

# Dataset step-by-step

## 1. Load data

Any csv file or farmhousehold rds file.

If large dataset ( $N > 10000$ ) you might want to subset your data

Don't forget to click on "Load data"

## 2. Select variables

Which one to be maximized? Which one to be minimized?

## 3. Outliers

Log transform right skewed variables and/or remove outliers

# Selection of Positive deviants

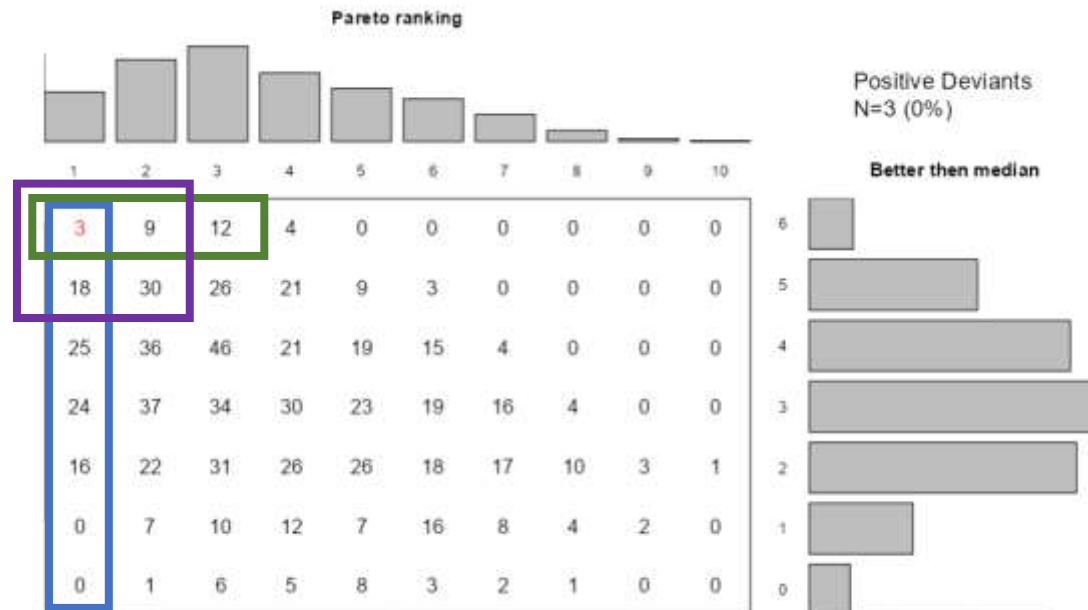
Four options:

A: Non-dominated households (Pareto rank 1)

B: Non-dominated households AND values better than the median

C: Pareto rank  $\leq 3$  AND values better than median

D: Pareto rank  $\leq 2$  AND almost all values better than median



Look at the distribution of the positive deviants and select the option that fit best your needs.

# Your turn

```
> library(farmhousehold)

> runPosdev()
```

Online version:

[https://startistic.shinyapps.io/farmhousehold\\_posdev/](https://startistic.shinyapps.io/farmhousehold_posdev/)