

# Multivariate Data Analysis (MVDA) & Visualization

Mpho Mafata

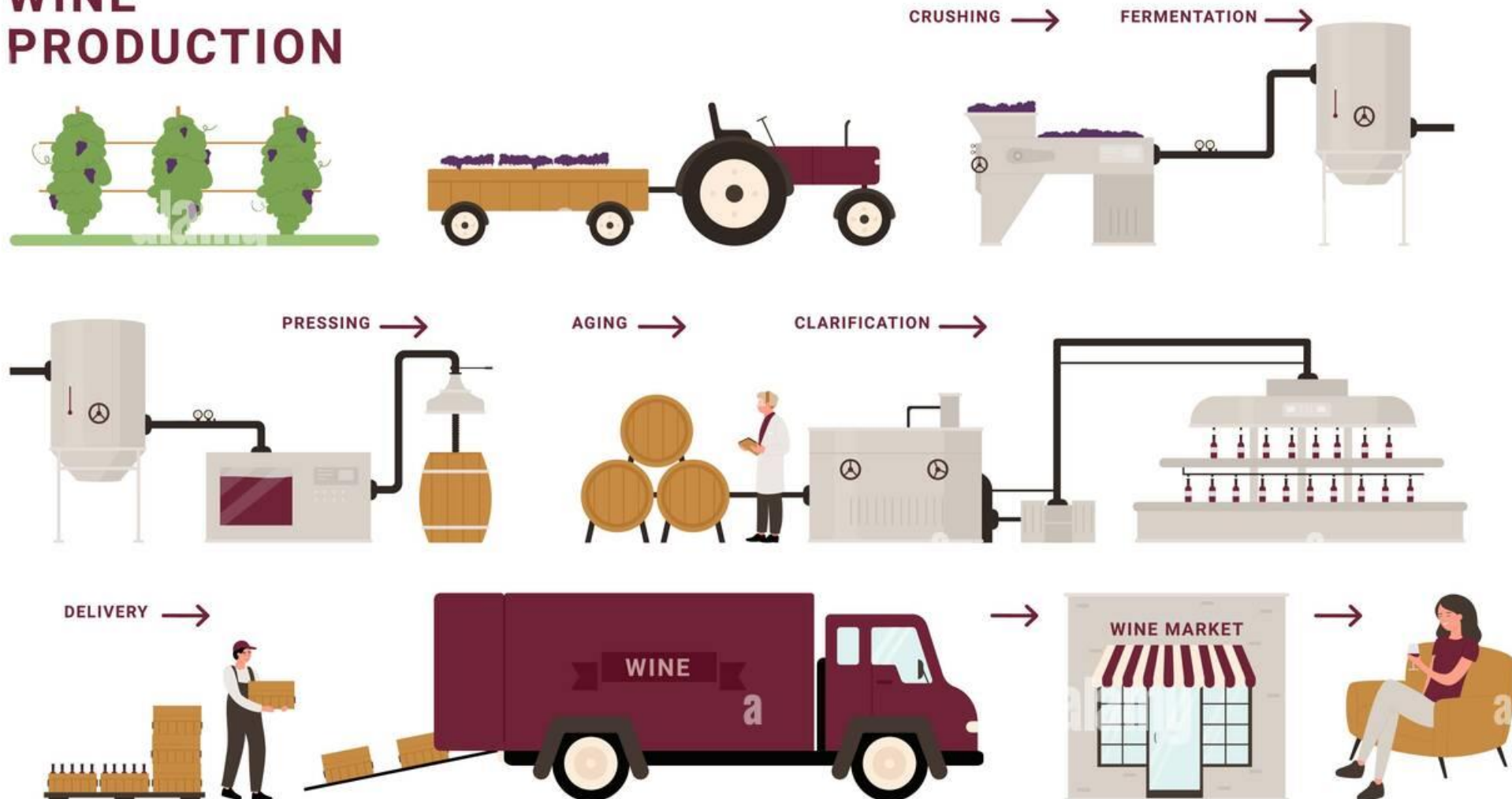
10 October 2024





# Data integration and fusion

## WINE PRODUCTION



# Data wrangling & Caveats in MVDA

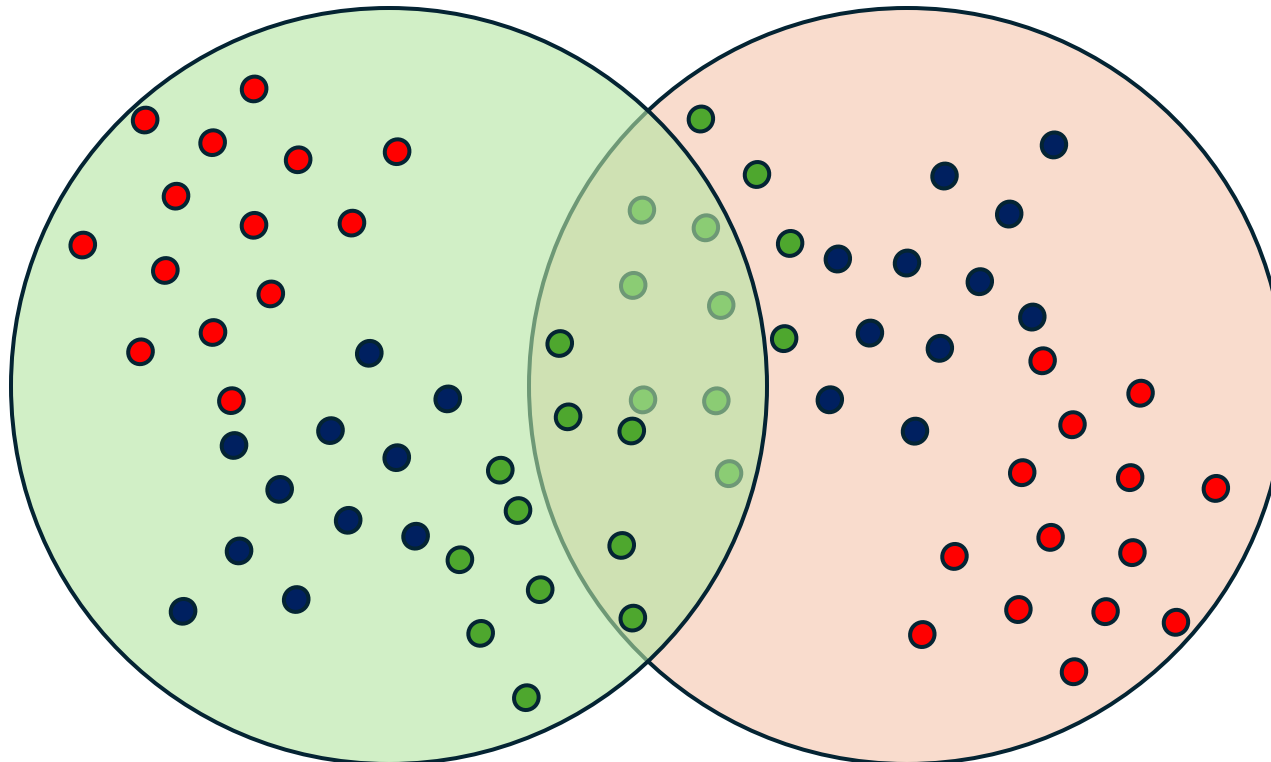
## Grouping observations

The variance across observational groups of **samples** is greater than between each sample

## Grouping variables

The variance across observational groups of **measurements** is greater than between each measurements

## Combined effects



## Multiblock Approaches

### Purpose of exploratory

- Reduce number of dimensions
- \* Used prior to confirmatory analyses

### Purpose of confirmatory

- Prediction
- \* Calibration, validation, and testing
- \* Large sample variation and variability

### Examples

- Variations of PCA: sum-PCA, m-PCA, h-PCA, etc.
- Factor analysis: PARAFac, PARAdise, and variations, MFA, ComDim, etc.
- Predictive analysis: PLS variations OPLS, OPLS-DA, P-ComDim, LDA, etc.

### Practical optimization criteria

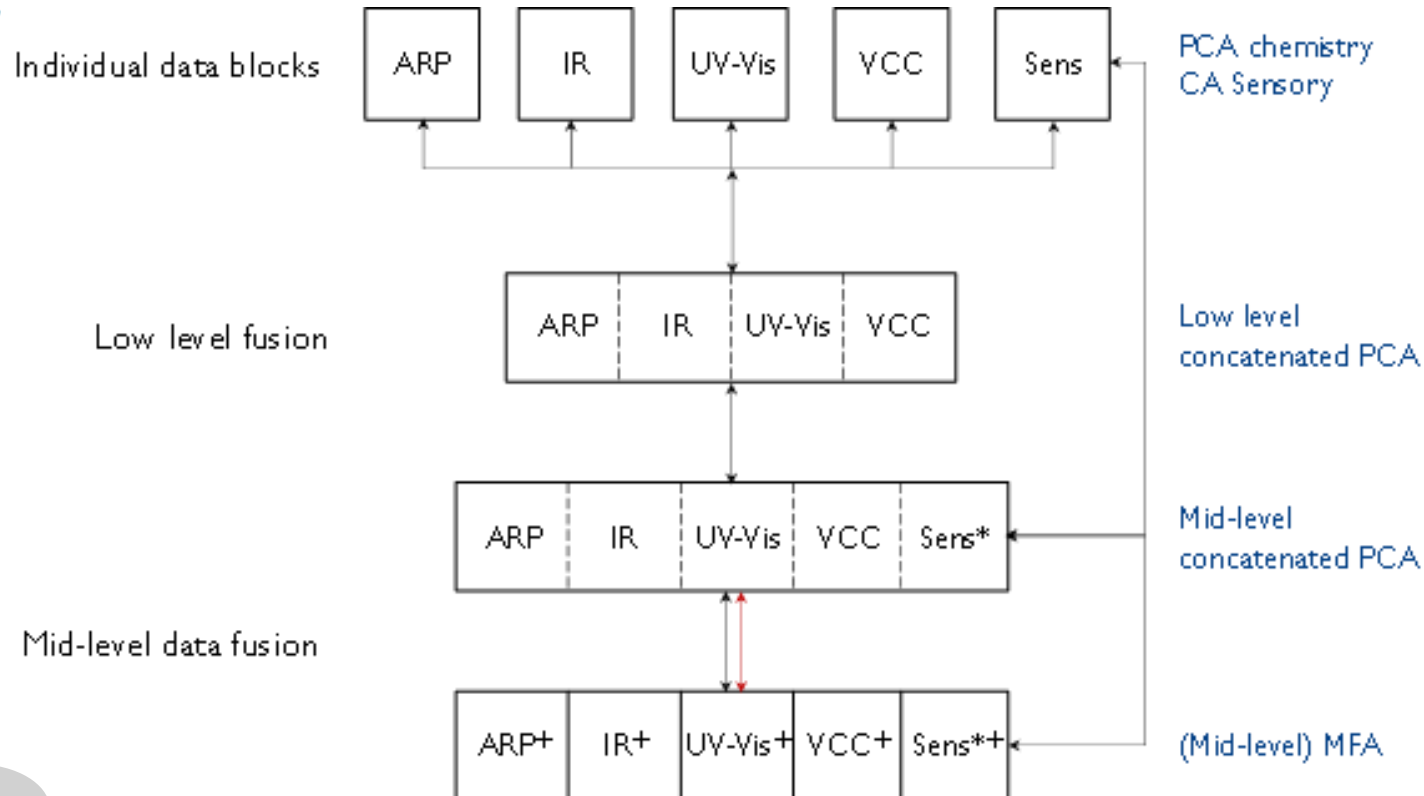
- Factors/dimensions with Eigenvalue less than 1
- Cumulative variation of 70%
- Dimension/point of first inflection in eigenvalue decay/ scree plot
- Optimize particular criterion ex. indices such as coefficients of fit (covariance, correlation or regression)

# Data integration and fusion

The chemistry:

\* Matrix effect

Love affair:  
sensory  
scientists



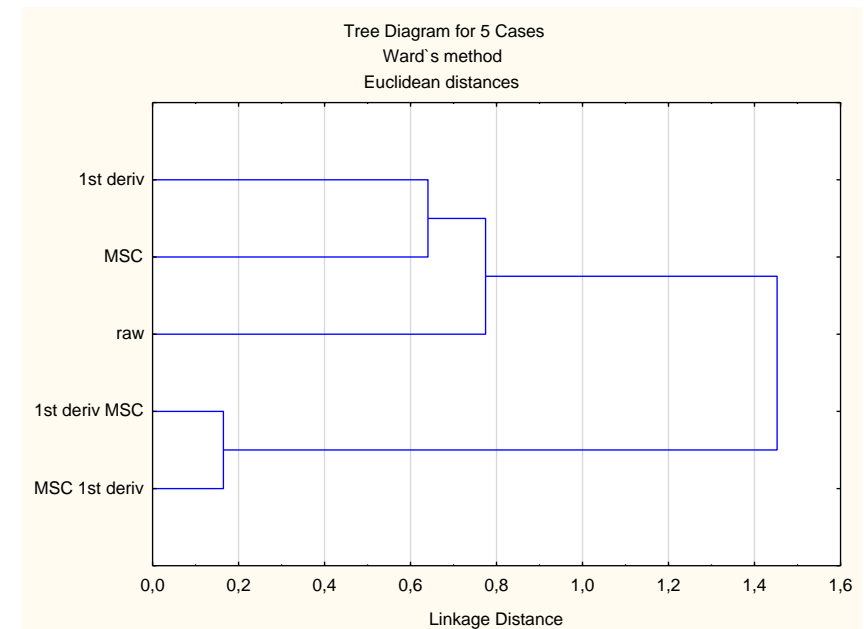
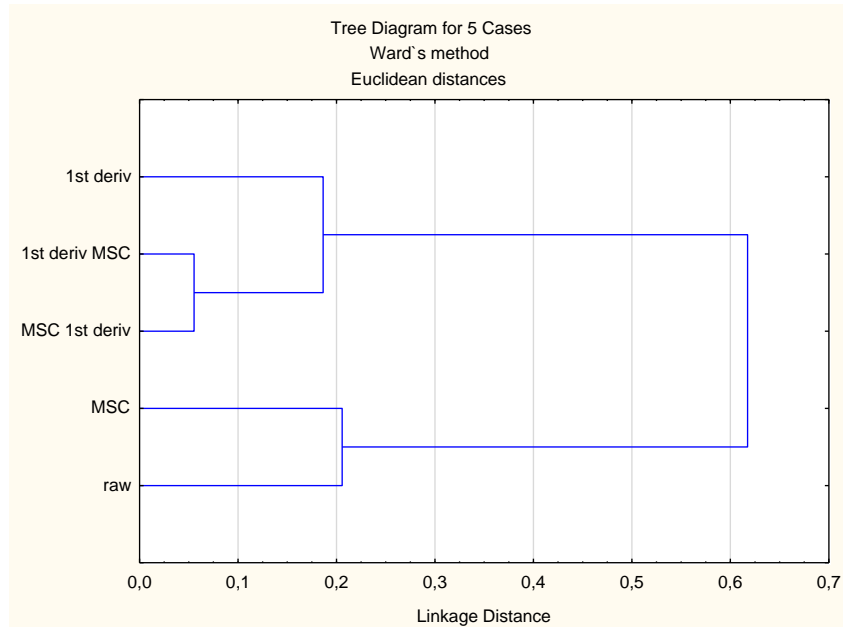
Winemaking as an  
industrial process  
PAT

Think of wine  
as a complex  
chemical  
solution



# Data integration and fusion

AVN	CHENIN BLANC							SAUVIGNON BLANC						
		1st deriv	1st deriv MSC	MSC	MSC 1st deriv	raw	MFA	1st deriv	1st deriv MSC	MSC	MSC 1st deriv	raw	MFA	
	1st deriv	1	0.99	0.81	0.98	0.83	0.98	1	0.97	0.89	0.97	0.84	0.99	
	1st deriv MSC	0.99	1	0.82	0.99	0.78	0.97	0.97	1	0.90	0.99	0.74	0.98	
	MSC	0.81	0.82	1	0.87	0.88	0.92	0.89	0.90	1	0.92	0.83	0.95	
	MSC 1st deriv	0.98	0.99	0.87	1	0.82	0.98	0.97	0.99	0.92	1	0.74	0.98	
	raw	0.83	0.78	0.88	0.82	1	0.90	0.84	0.74	0.83	0.74	1	0.85	
	MFA	0.98	0.97	0.92	0.98	0.90	1	0.99	0.98	0.95	0.98	0.85	1	



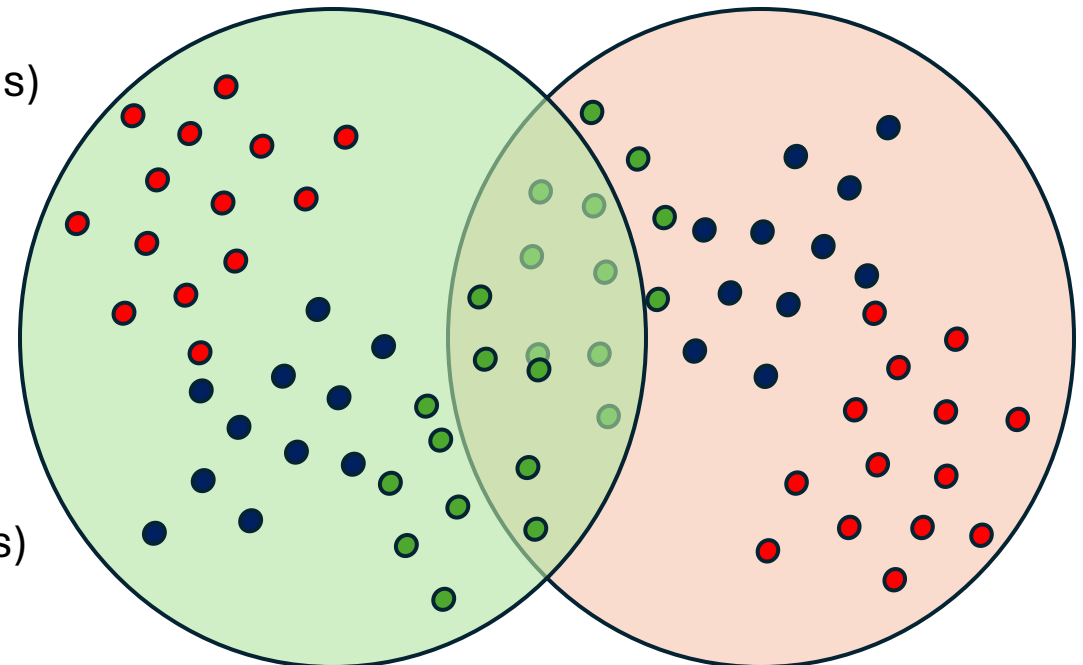
## MFA – Multiple Factor Analysis

This specific method is useful in many fields where variables are structured into groups, for example:

- Genomic: protein variables, DNA variables
- Sensory analysis: sensorial and physico-chemical variables
- Questionnaires: student health (addicted consumptions variables, psychological conditions variables, sleep, identification variables...)
- Comparison of coding (continuous variables, categorical variables)

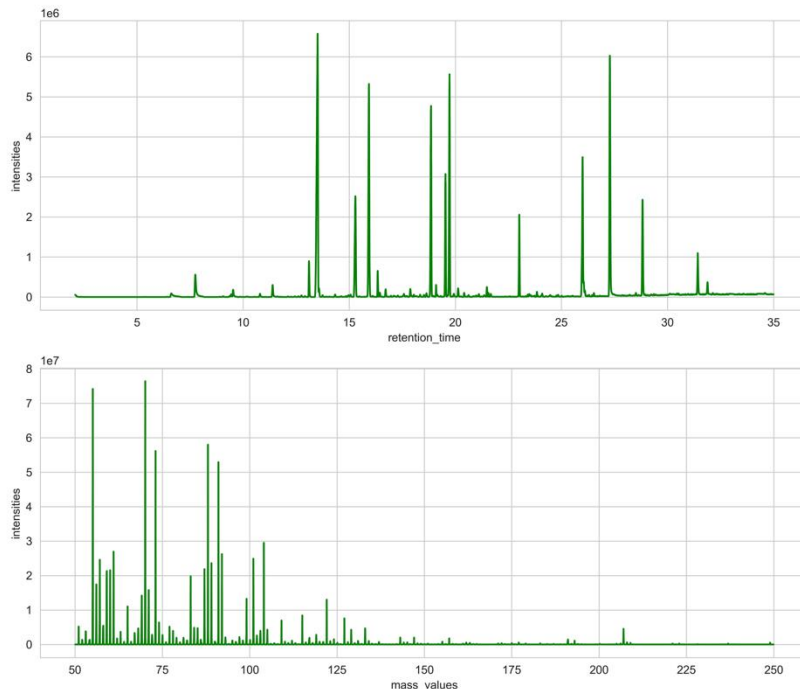
Taking into account the structure of the data allows to:

- Balance the influence of each group of variables
- Study the links between the sets of variables
- Give the classical graphs but also specific ones:
- Partial representation (individuals seen by one group of variables)
- Groups of variables



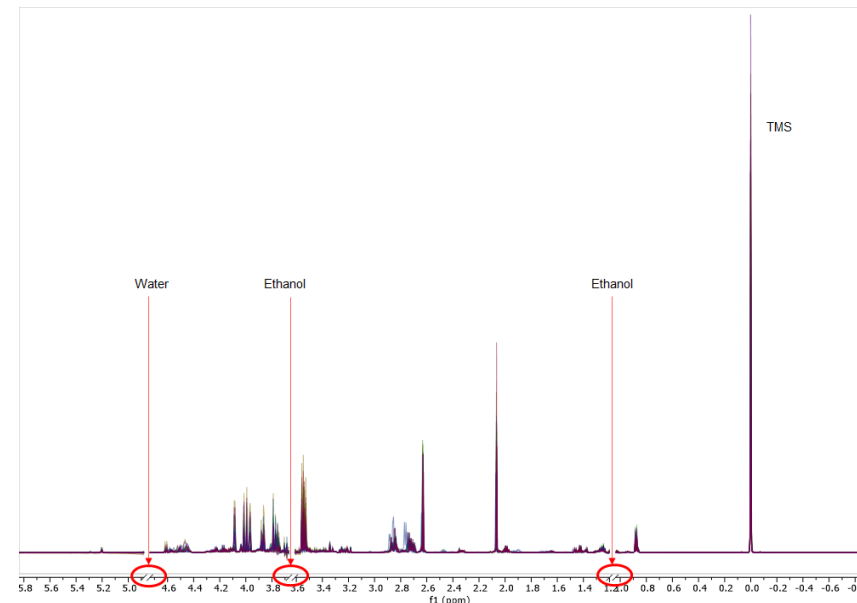
# Our dataset

1. Sensory data from sorting (2 forms)
2. NMR
3. HRMS (2 modes)
4. IR
5. Physicochemical parameters



High Resolution Mass Spectrometry (HRMS)

## Sensory data from sorting



Nuclear Magnetic Resonance (NMR) spectrum



## Explore each block separately

1. Sensory data from sorting
  1. Non-verbose data - MDS
  2. Verbose data – Heatmap
2. Oenological/physicochemical parameters - PCA

## Explore data fusion (MFA)

1. FactomineR
2. Factoextra

Tutorial videos from FactoMineR creators

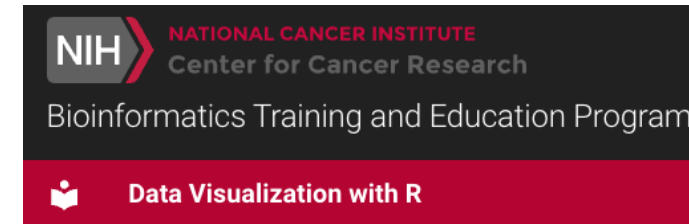
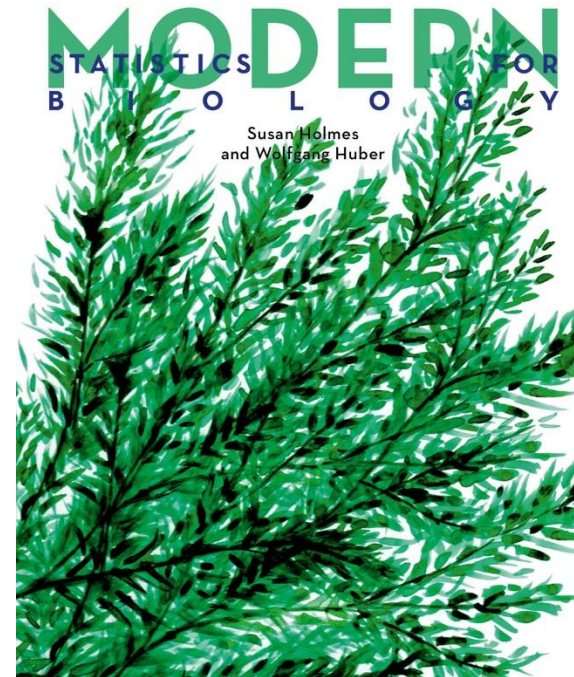
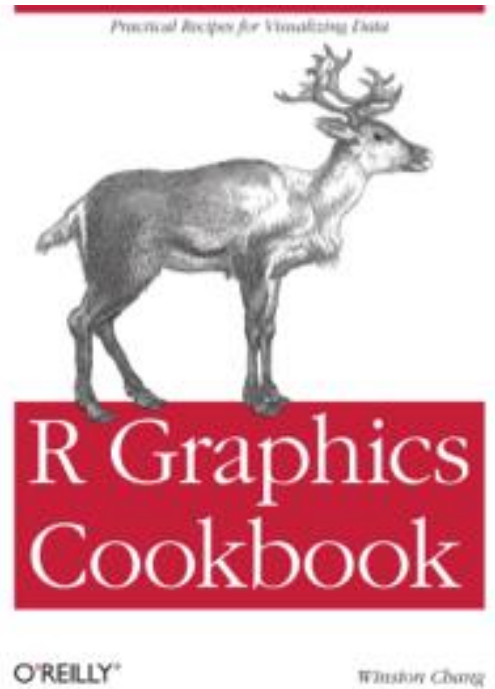
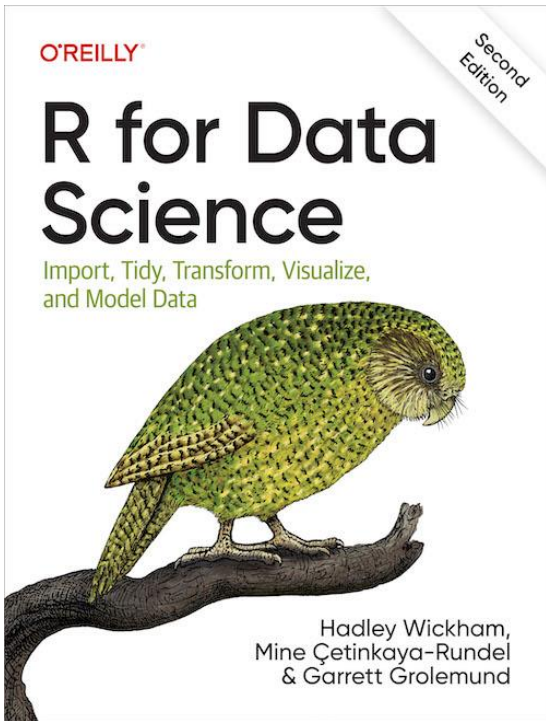
- ☐ Multiple Factor Analysis (MFA)

<https://www.youtube.com/watch?v=pks8m2ka7Pk>

- ☐ Textbook and material

<http://factominer.free.fr/bookV2/index.html>

[https://www.youtube.com/watch?v=MOl0Aw1TTFE&list=PLnZgp6epRBbRX8TEp1HlFGqfMf\\_AxYEj7](https://www.youtube.com/watch?v=MOl0Aw1TTFE&list=PLnZgp6epRBbRX8TEp1HlFGqfMf_AxYEj7)





# Thank you!



<https://github.com/mpho-mafata>



<https://www.linkedin.com/in/mafatampho>



<https://orcid.org/0000-0002-6468-7193>

## A chemometric approach to investigating South African wine behaviour using chemical and sensory markers

by  
**Mpho Mafata**

Dissertation presented for the degree of  
**Doctor of Philosophy (Agricultural Sciences)**



at  
**Stellenbosch University**  
Department of viticulture and Oenology, Faculty of AgriSciences

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.

*Supervisor:* Dr. Astrid Buica  
*Co-supervisors:* Dr. Jeanne Brand and Prof. Andrei V. Medvedovici

March 2021

<https://doi.org/10.13140/RG.2.2.17899.00804>