UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY

100
1918 · 2018

forward together · saam vorentoe · masiye phambili

Automated sensory attribute consolidation:

Efficient integration with data analysis applied to craft gins and beers

13 September 2023

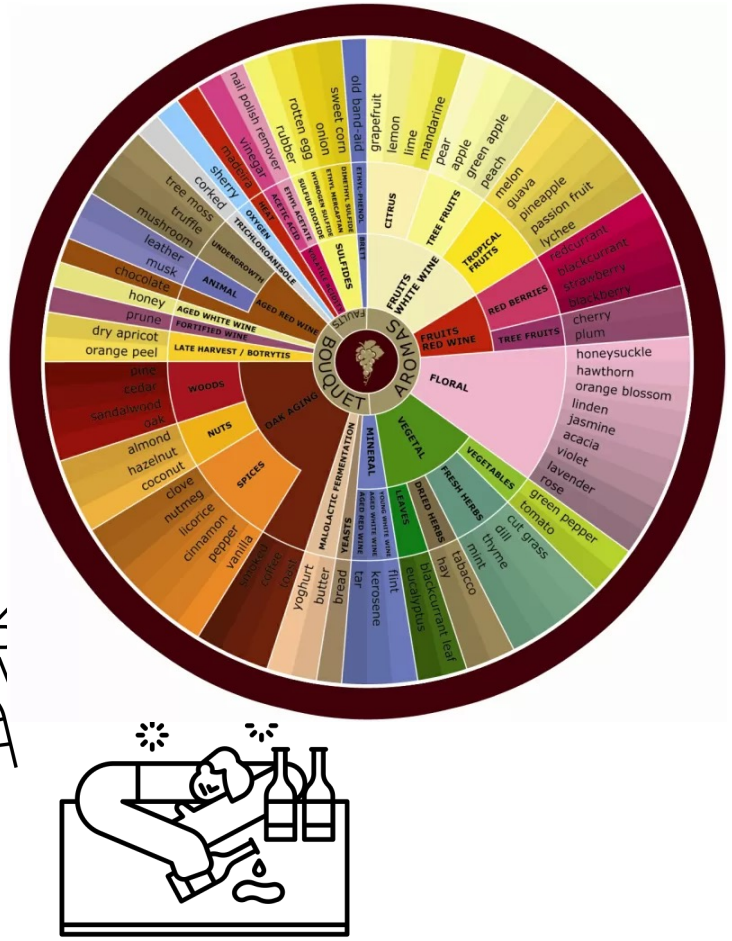
Sasev 2023

Contents

Background and problems identified

How do you define the taste of a gin or a beer?

- Map what sensory professionals experience
- This has been extensively researched, both in academia and industry
- Industry best practice, requires training of panel members
  - Cost and time constraints
  - Sensory fatigue
  - Human physiology constraints
- These factors severely limit the size of any comparison



Our proposal

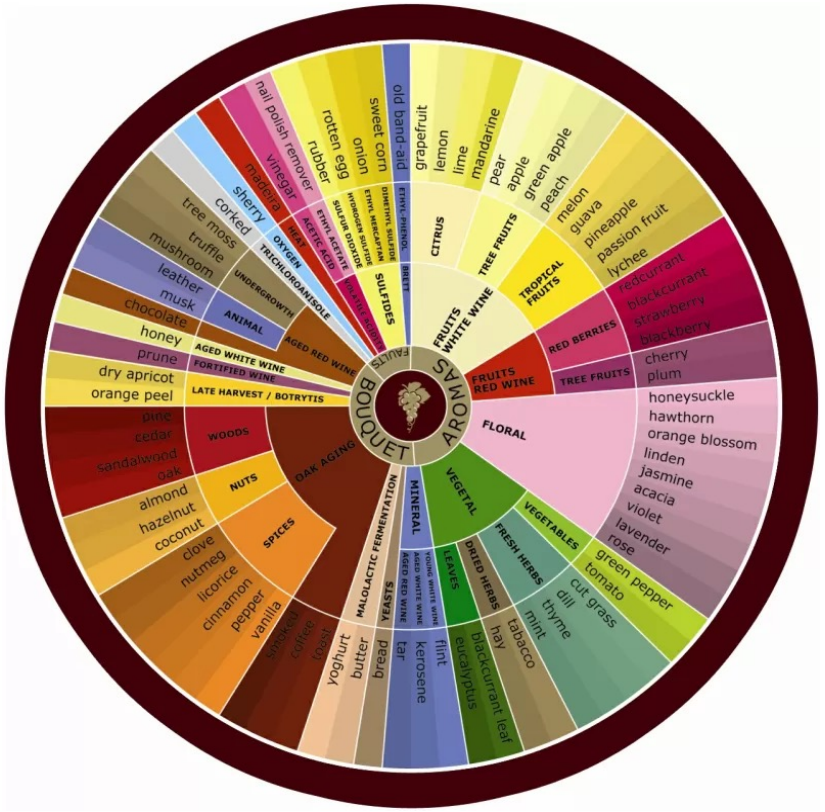
Bigger net with bigger holes

- Instead of benchmarking products together, we got sensory descriptions from a wider array of professional sources
- Creating a sensible sensory map
- Investigating and testing automation and expansion of this concept

Current study

Used a four-stage approach

1. Gathered data manually (Automation starting to become a necessity) [Scrapy, Beautiful Soup, Power Automate]
2. Extracted data either manually or by using NLP to tokenize the text, and POS analysis to extract sensory terms (2-tupple and 3-tupple) [Spacy] [Textacy]
   outputs => a list of descriptive terms for each product
3. Consolidation, including: [Pandas]
   - Manually by a panel of sensory experts
   - Automated procedure truncating the list based on citation frequency and mapping sensory terms to a standardized tiered lexicon
   - Pruning
4. Final output = descriptor map {Absence/Presence table}



Output analysis

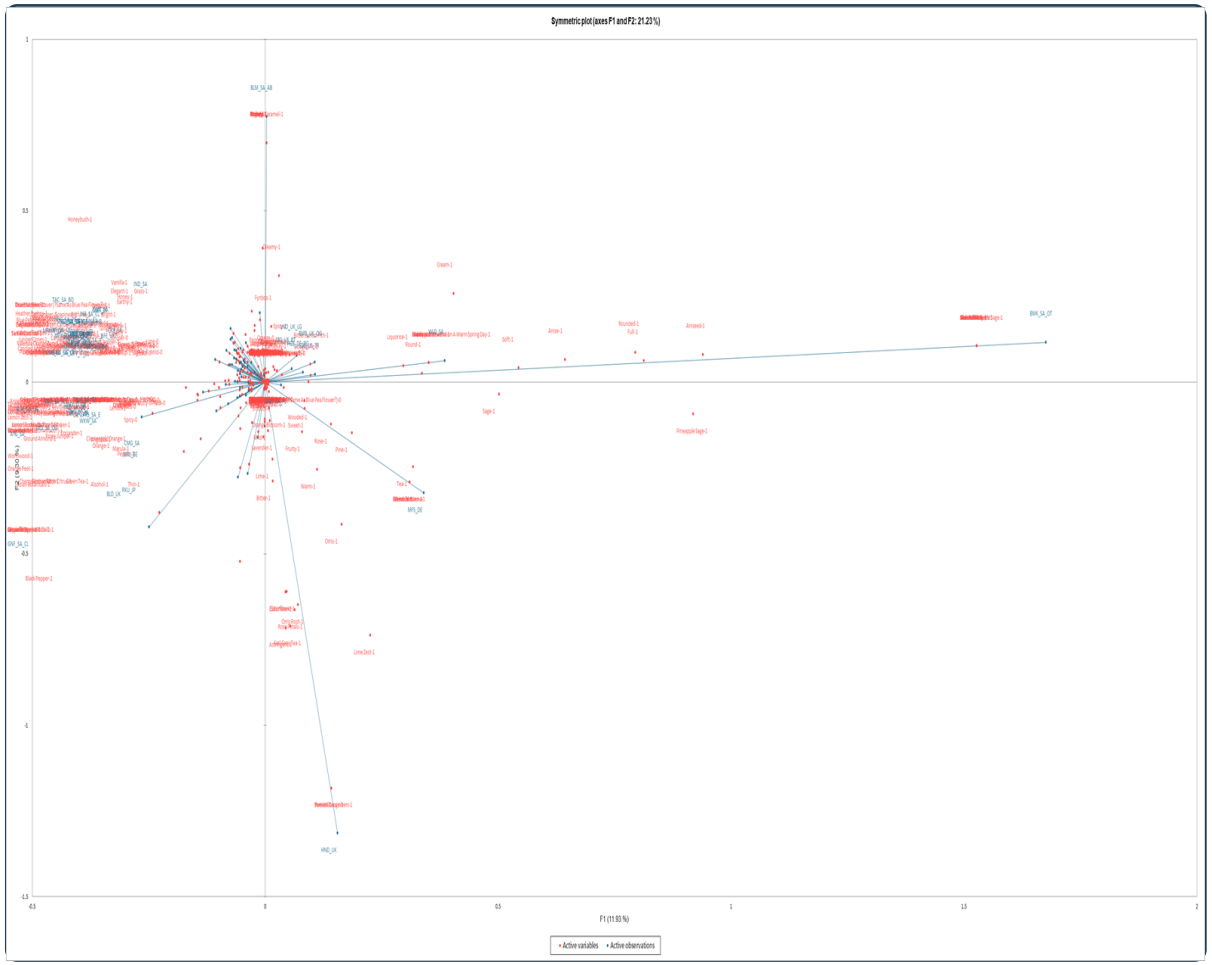
Maps, what are they good for?

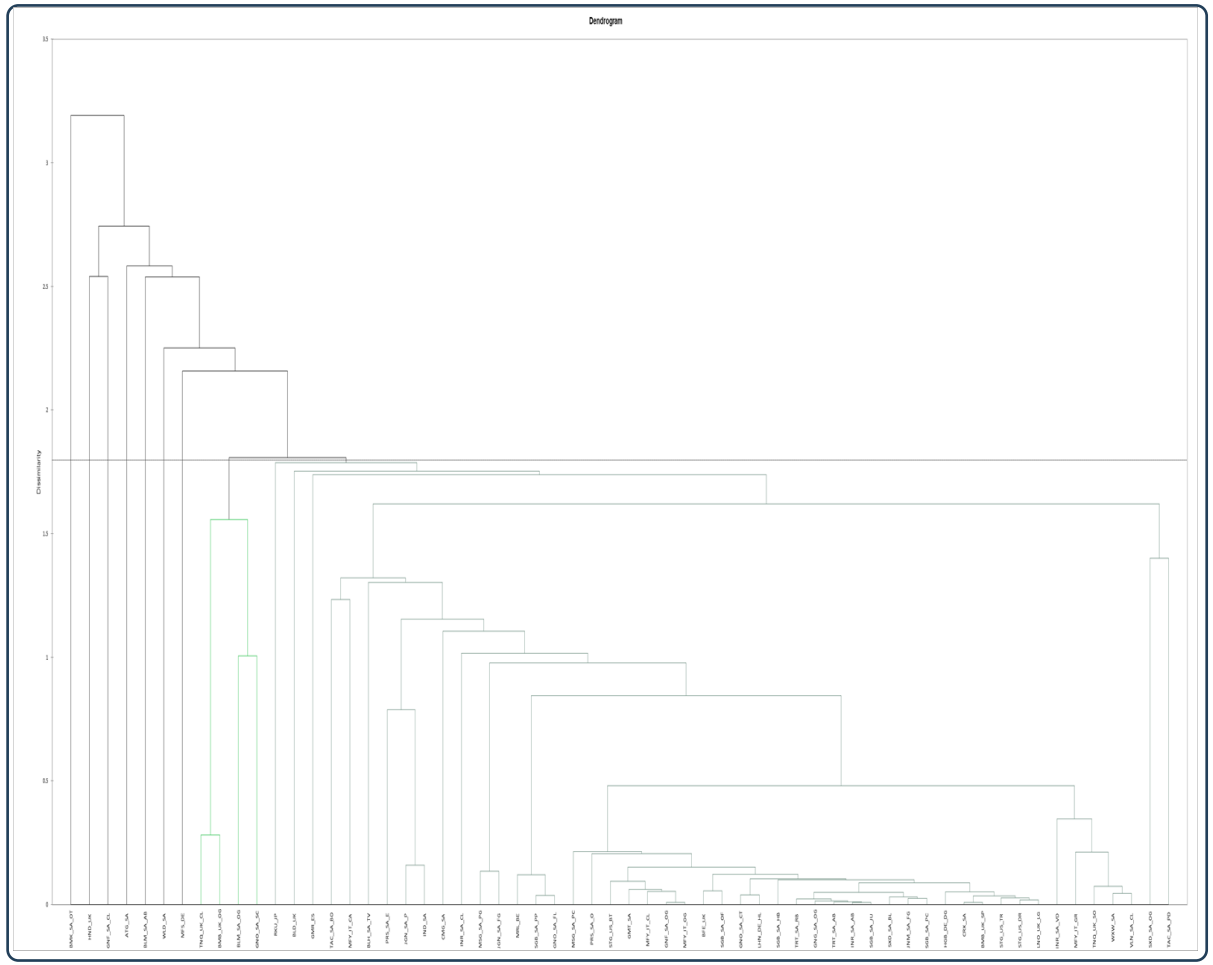
- To show the representative value of each map, and to verify that distinctiveness and inter relationships between products where preserved

How was this verified?

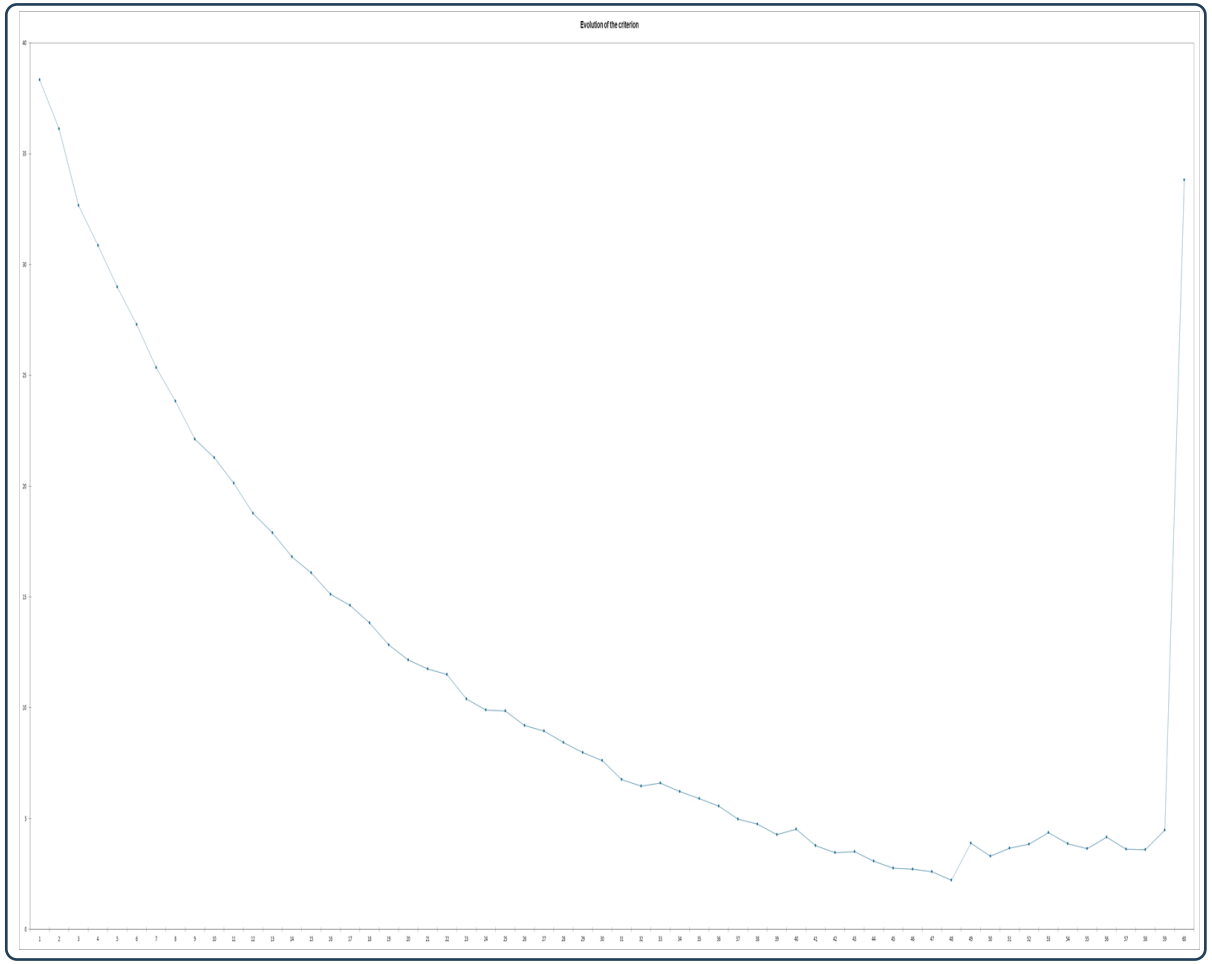
- Spaces investigated with HAC, MCA, fKMeans, and Concept Lattices

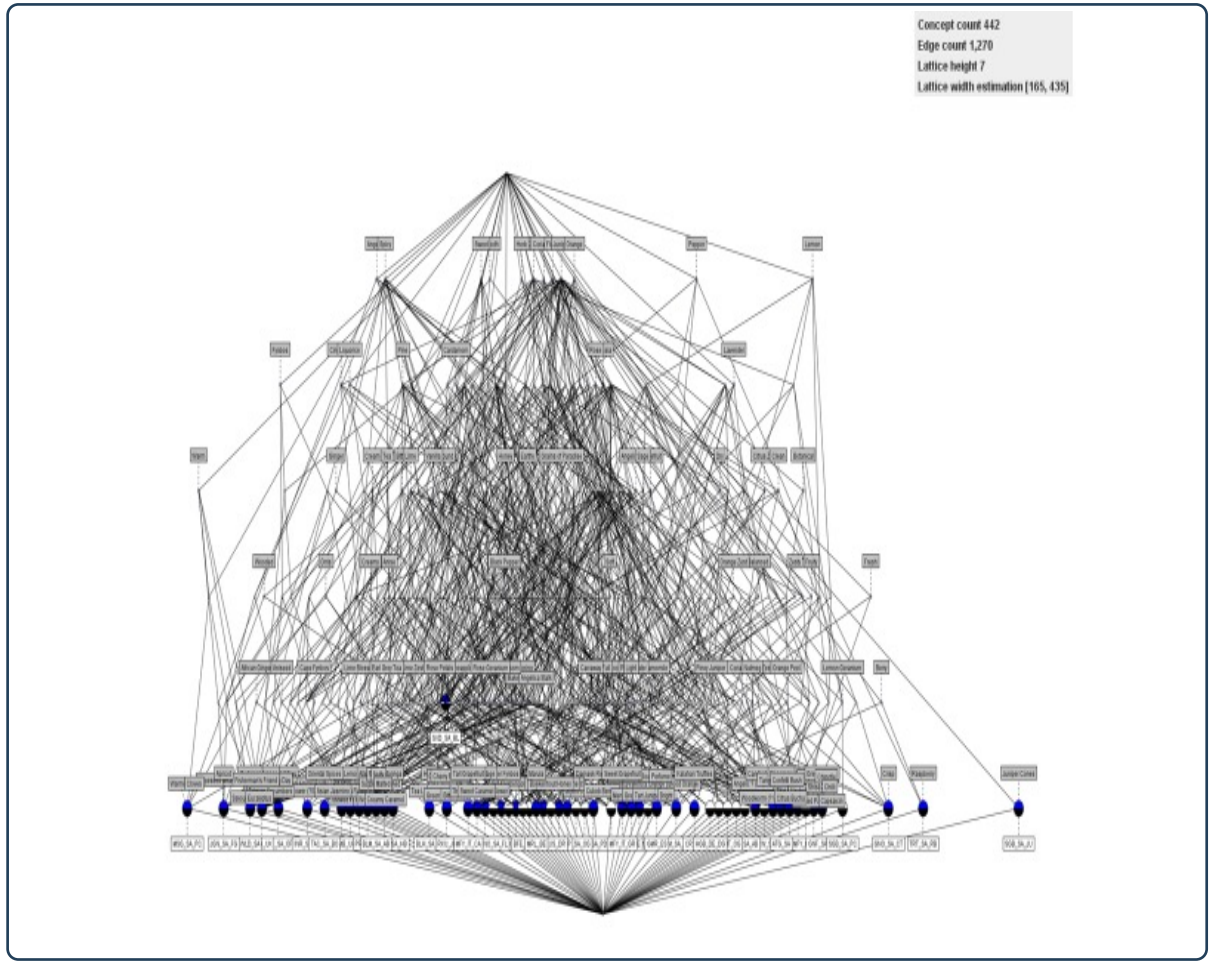
Application: Craft gin results from raw data analysis



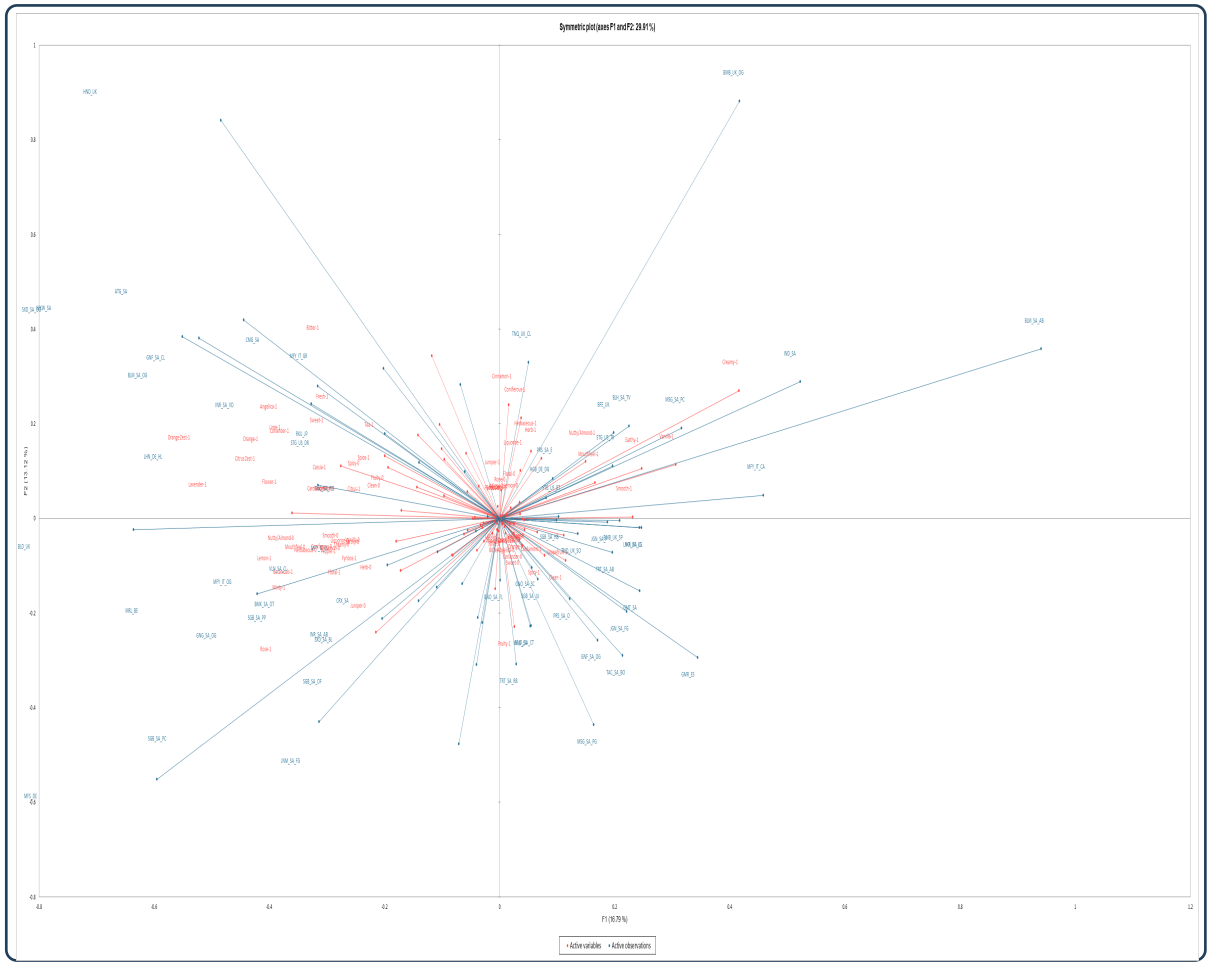


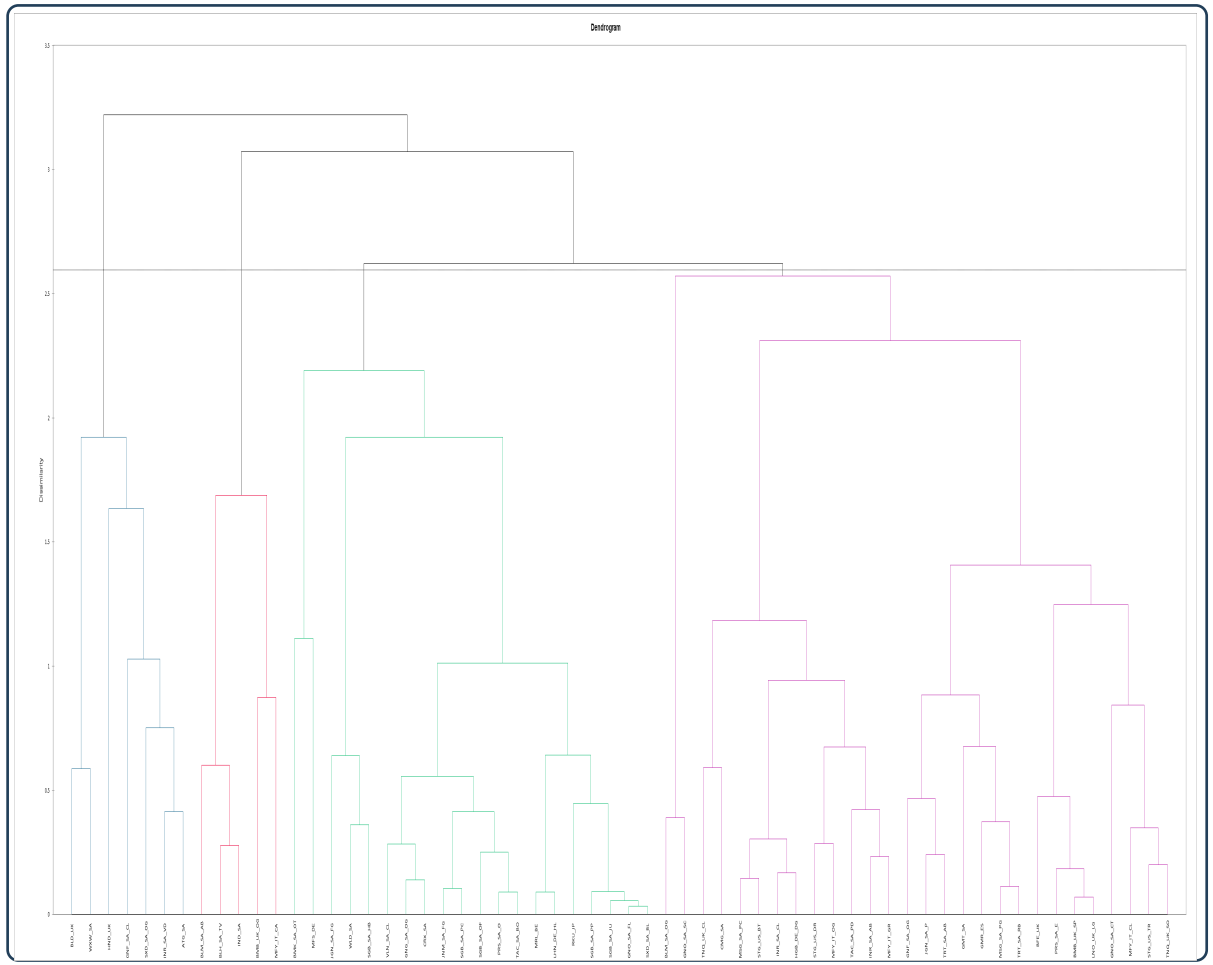
Application: Craft gin results from raw data analysis



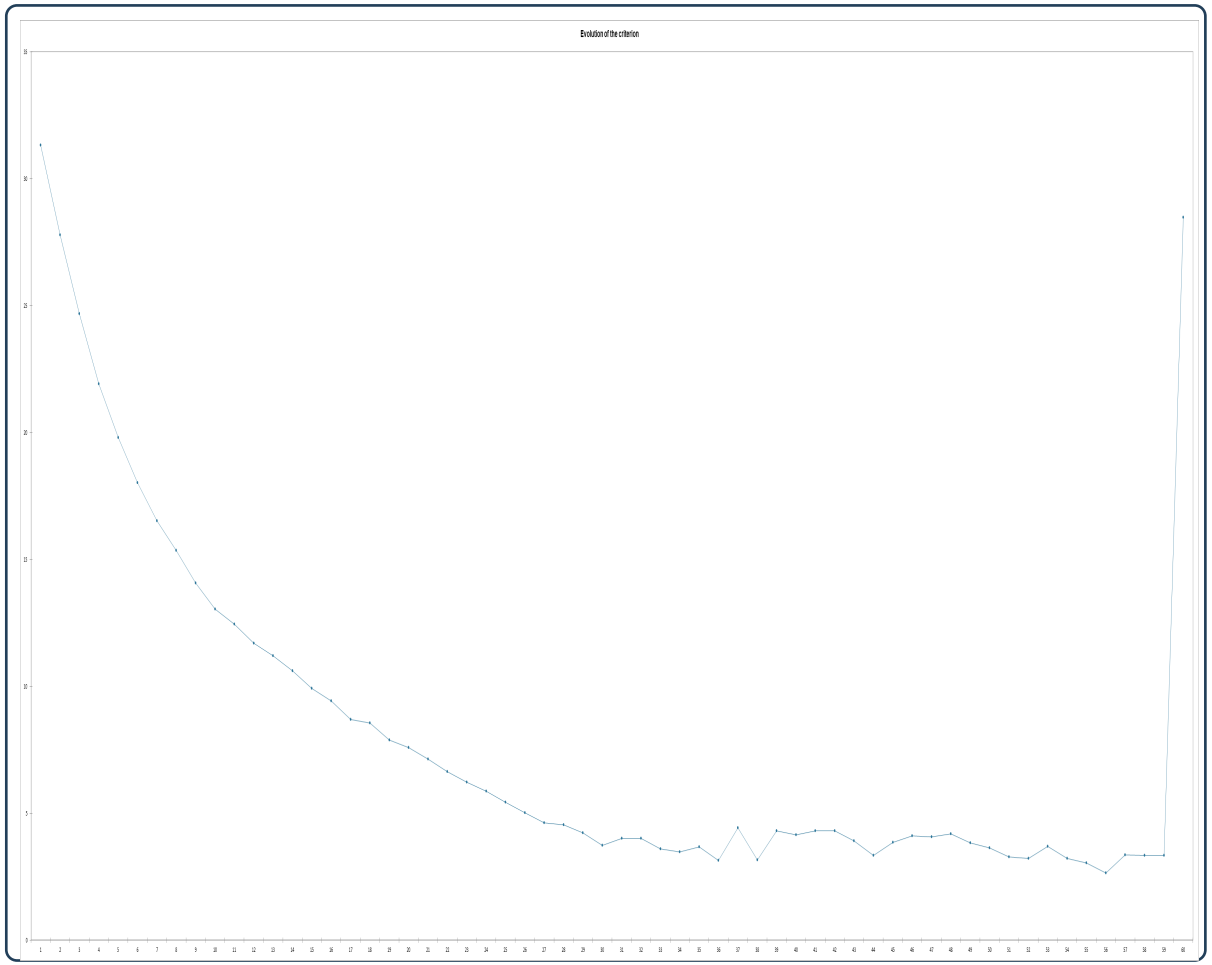


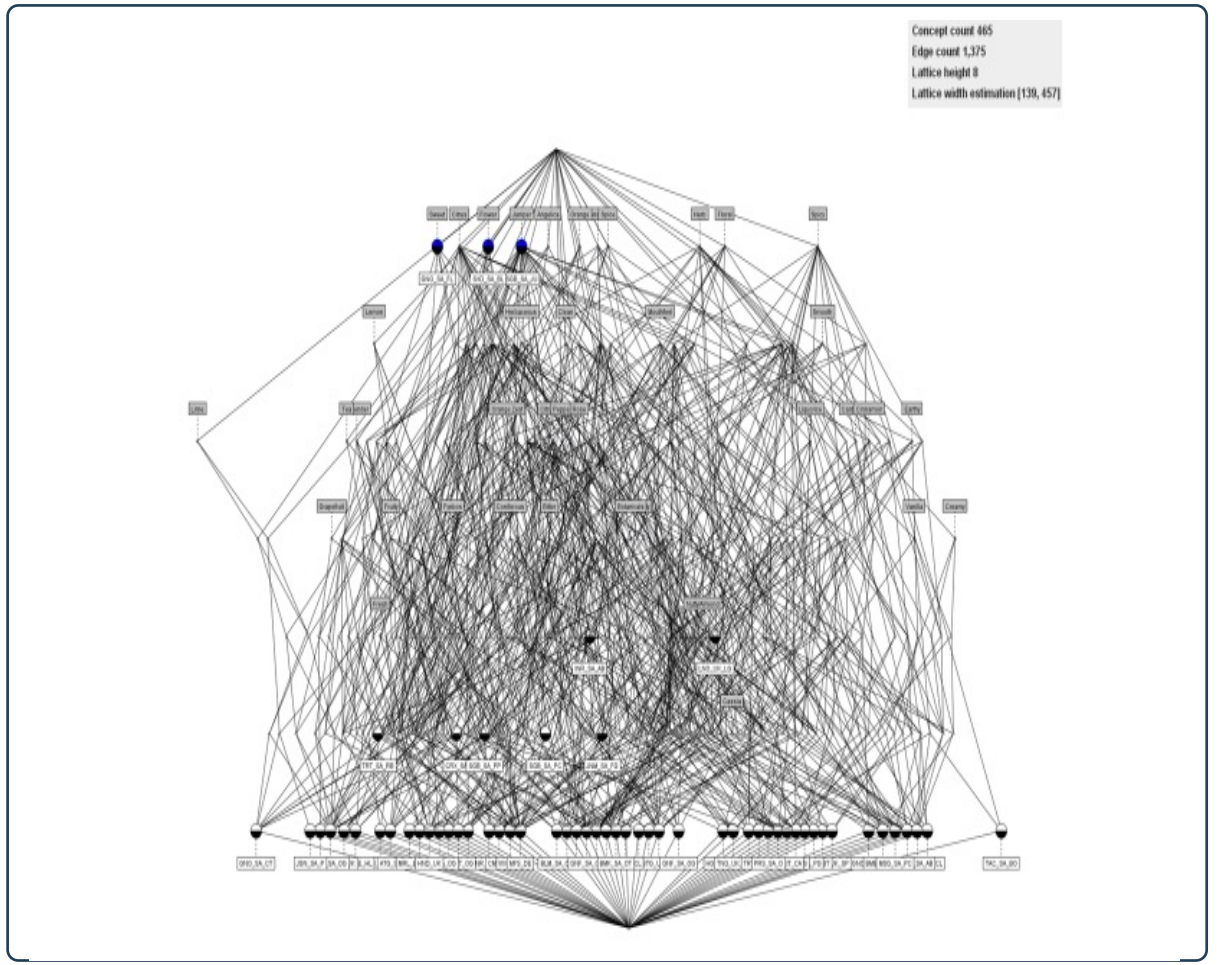
Application: Craft gin results from consolidated data analysis





Application: Craft gin results from consolidated data analysis





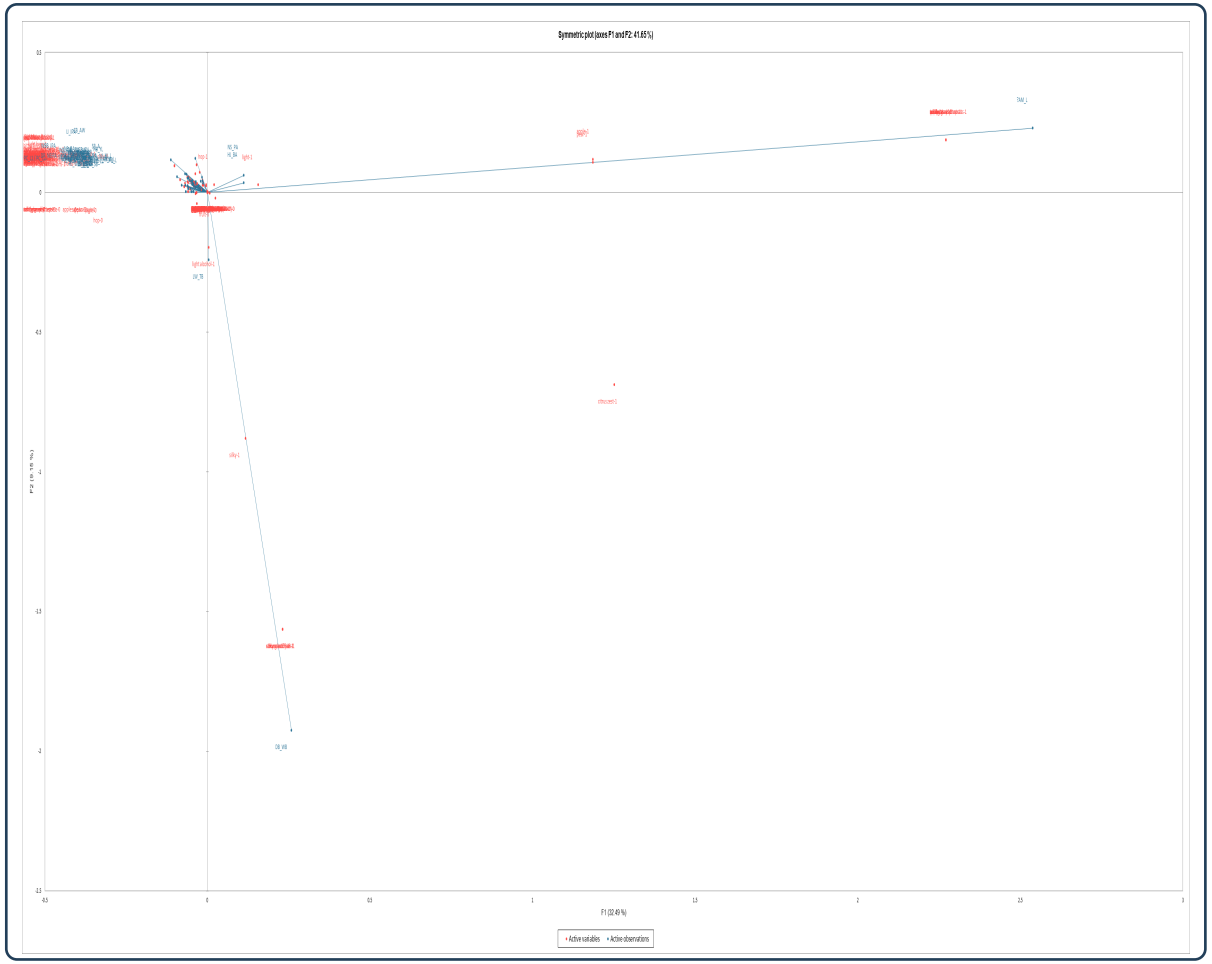
Application (Gin model comparison)

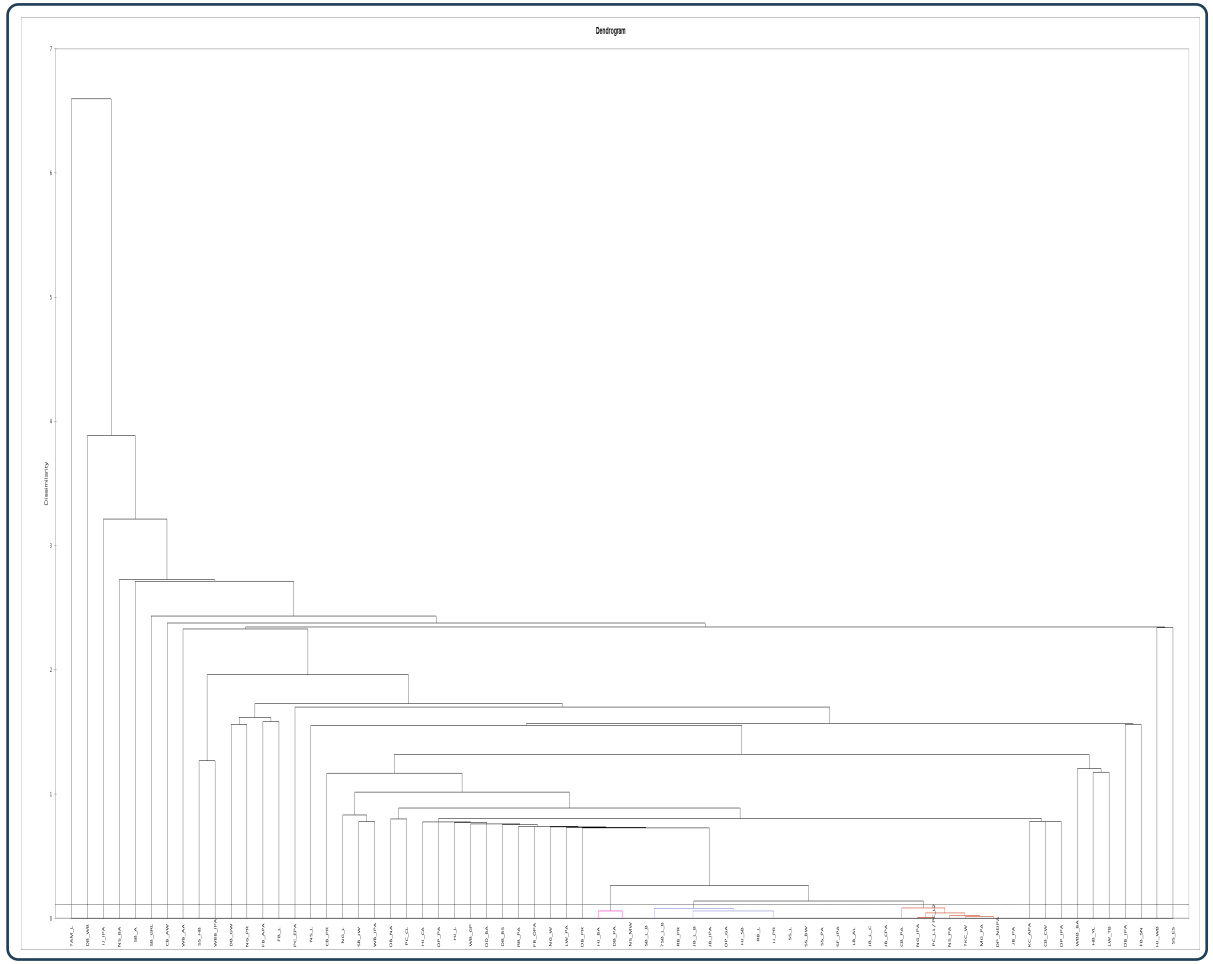
Comparison Metric?

To verify that our choices lead to comparable results with use RV coefficients to compare the different choices.

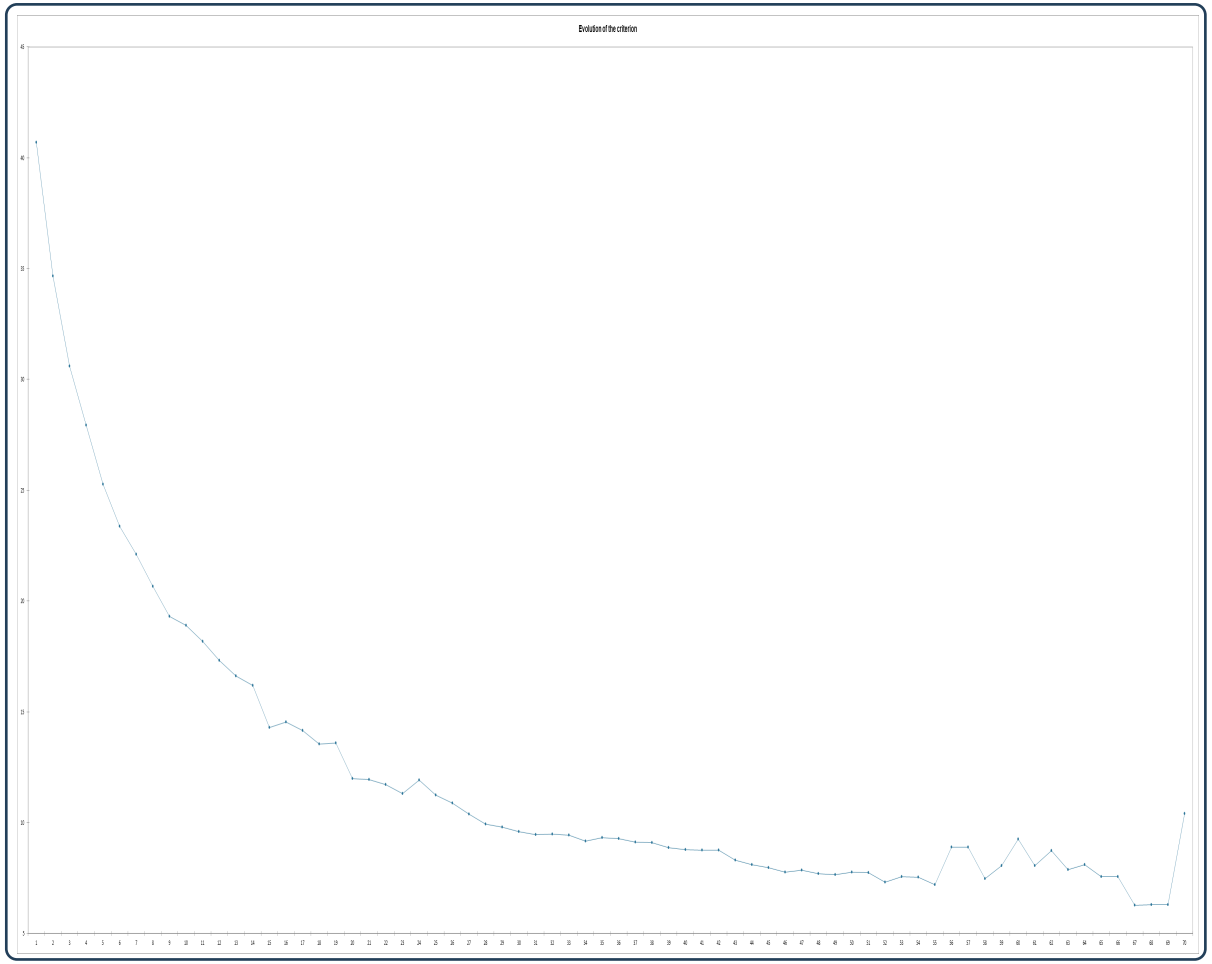
Datasets	Coverage	Dimensions	RV coefficients	Adjusted RV coefficients	p-value	Correlation
HRAW vs ARAW	100%	All	0.697	0.431	0.000	Low
HRAW vs HCON	100%	All	0.760	0.625	0.000	Moderate
ARAW vs AMAP	100%	All	0.786	0.656	0.000	Moderate
ARAW vs ACON	100%	All	0.822	0.727	0.000	High
HCON vs ACON	100%	All	0.446	0.246	0.000	Negligible

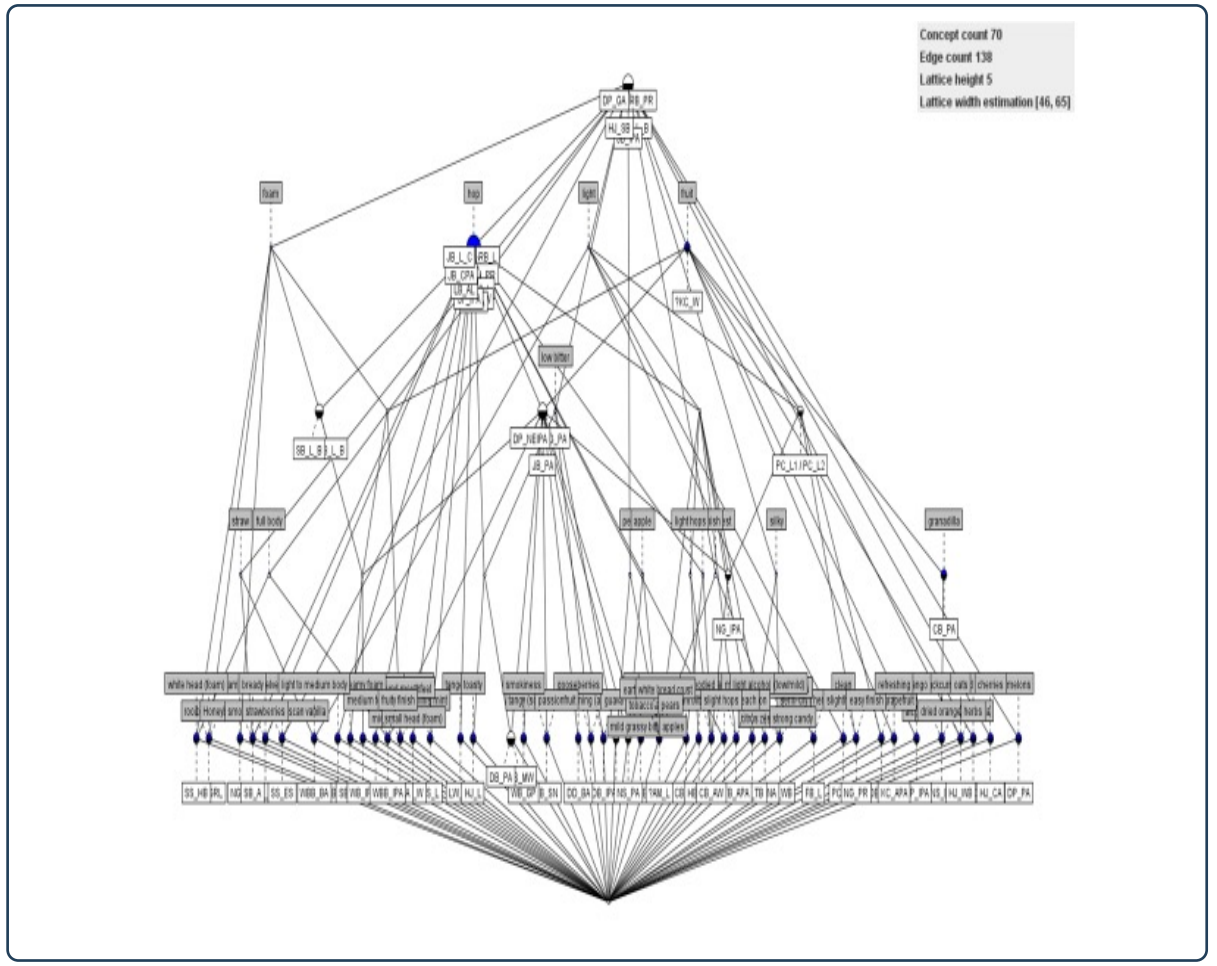
Application: Craft beer results from raw data analysis



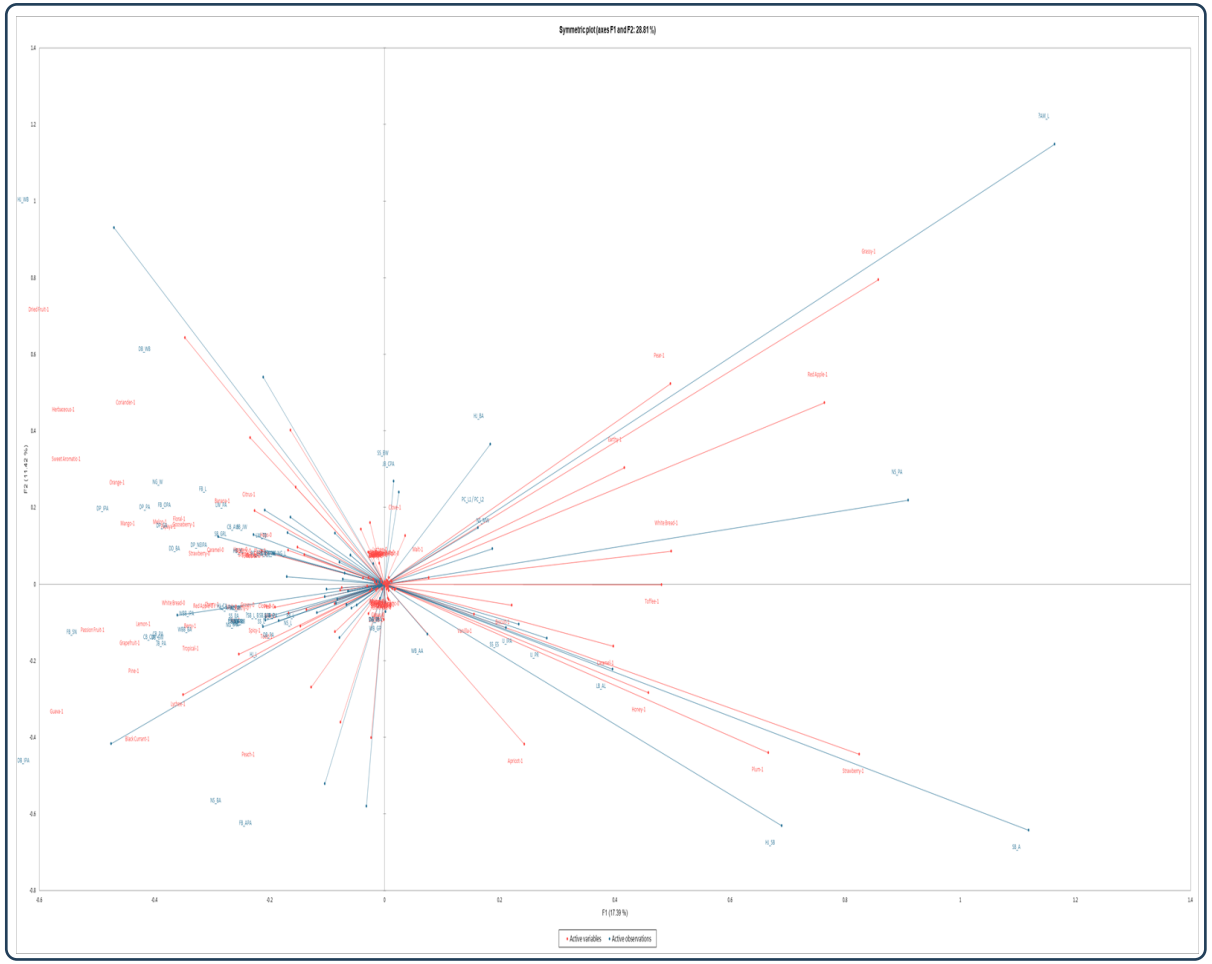


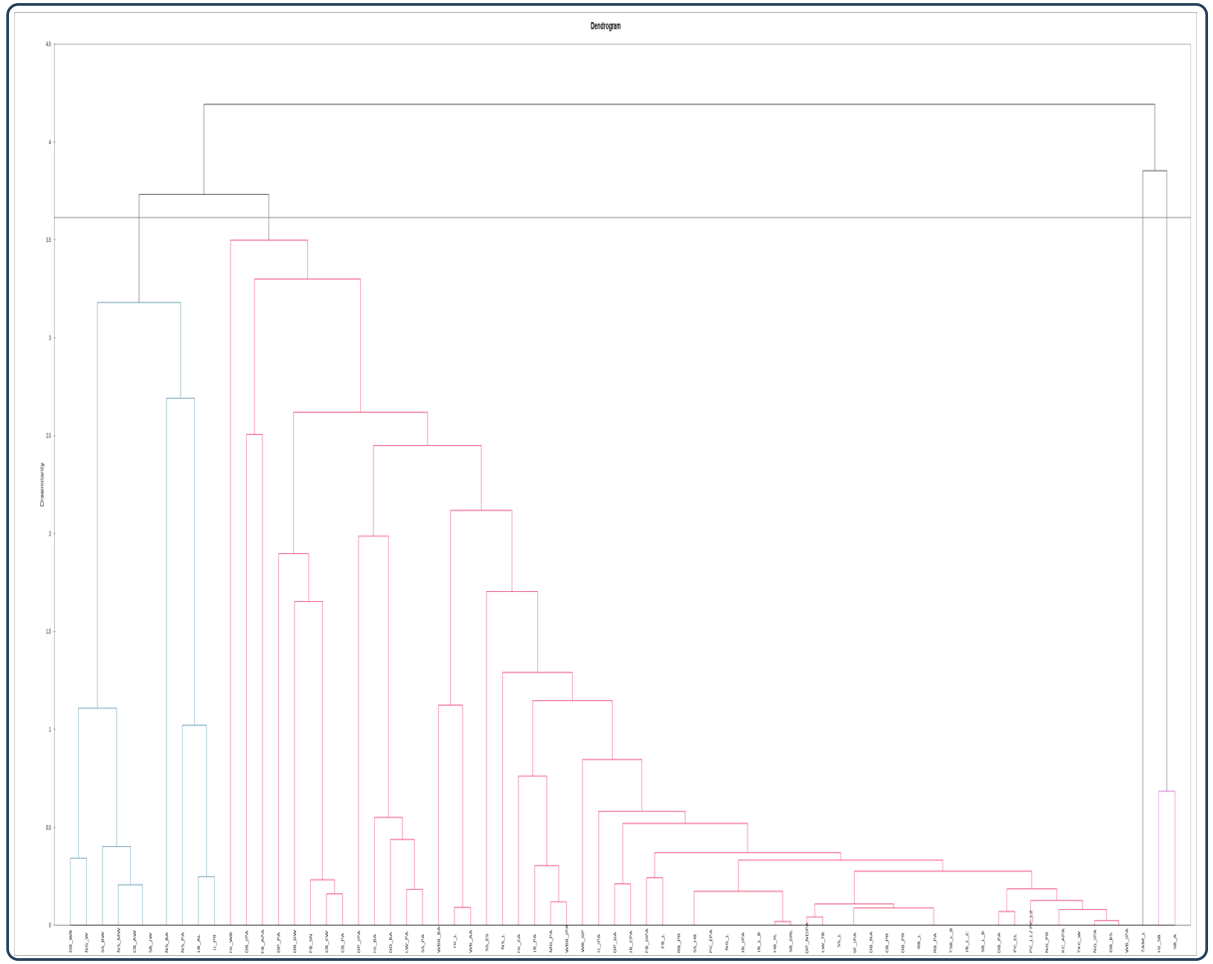
Application: Craft beer results from raw data analysis



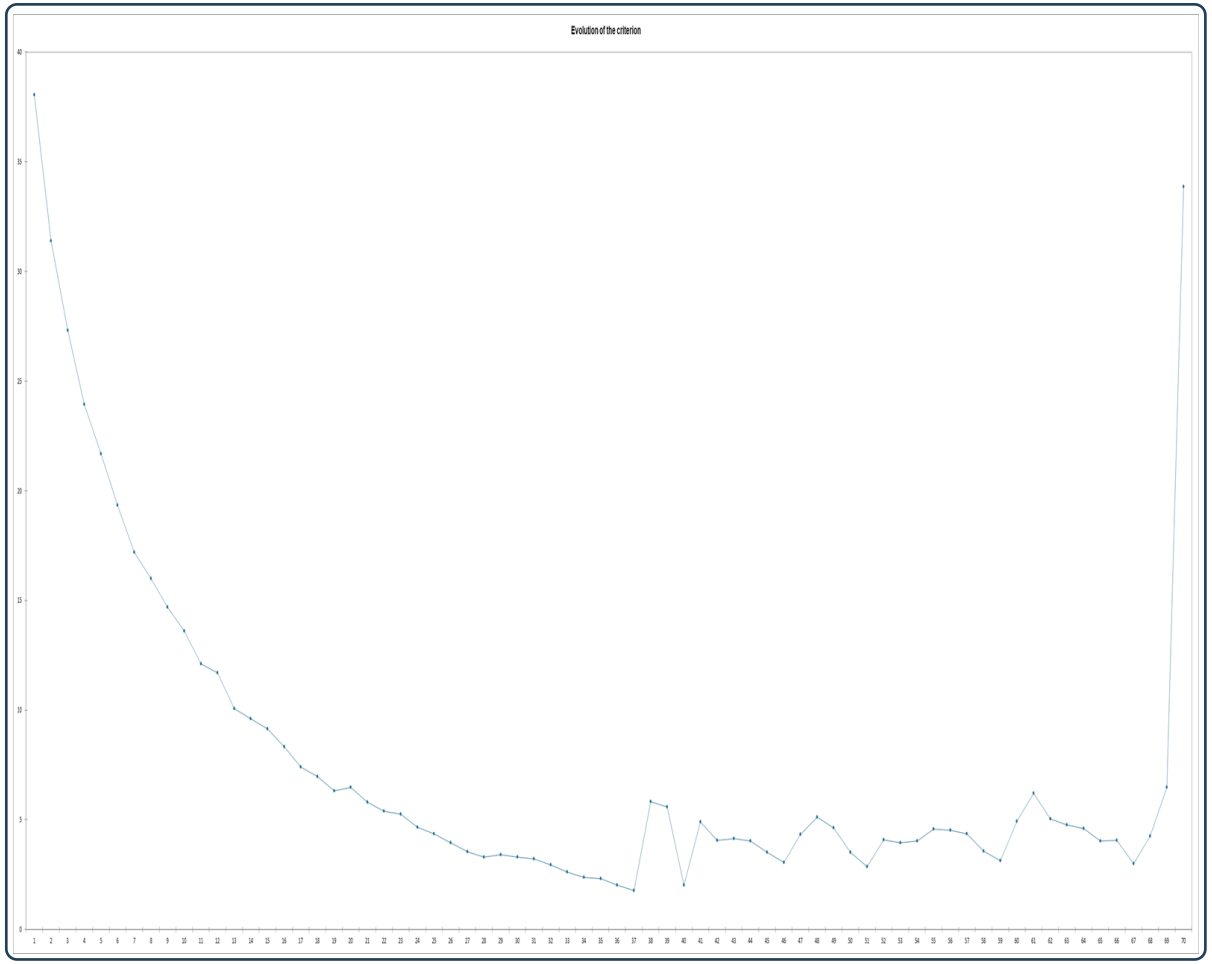


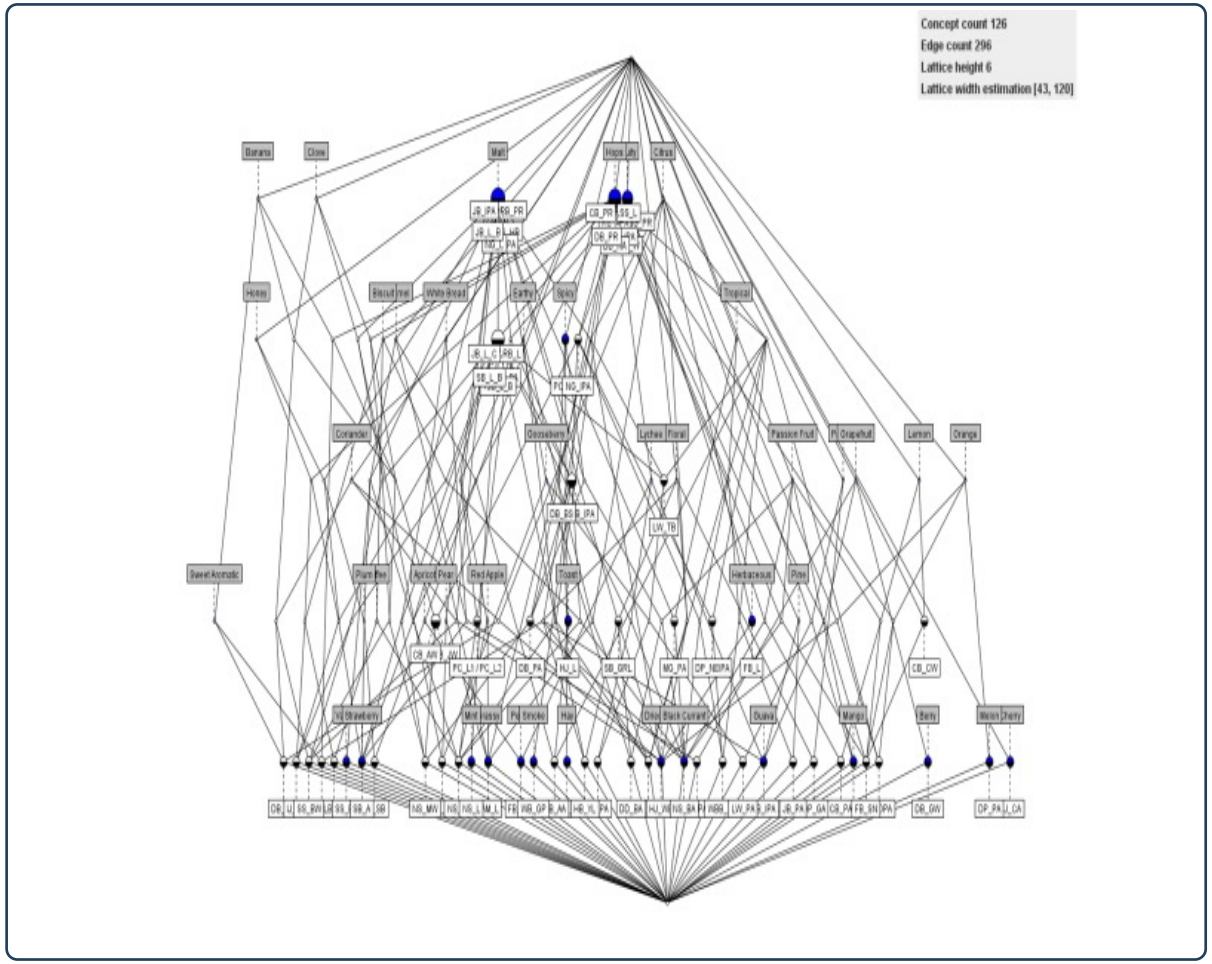
Application: Craft beer results from consolidated data analysis





Application: Craft beer results from consolidated data analysis





Application (Beer model comparison)

Comparison Metric?

To verify that our choices lead to comparable results with use RV coefficients to compare the different choices.

Datasets	Coverage	Dimensions	RV coefficients	Adjusted RV coefficients	p-value	Correlation
HRAW vs ARAW	100%	All	0.411	0.229	0.000	Negligible
HRAW vs HCON	100%	All	0.527	0.373	0.000	Low
ARAW vs AMAP	100%	All	0.538	0.426	0.000	Low
ARAW vs ACON	100%	All	0.569	0.466	0.000	Low
HCON vs ACON	100%	All	0.841	0.801	0.000	High

Application (gin and beer)

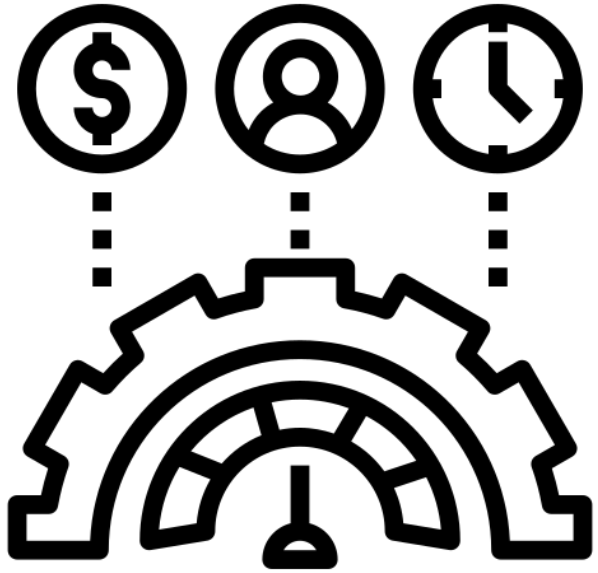
Does it work?

Yes (for a given value of work)

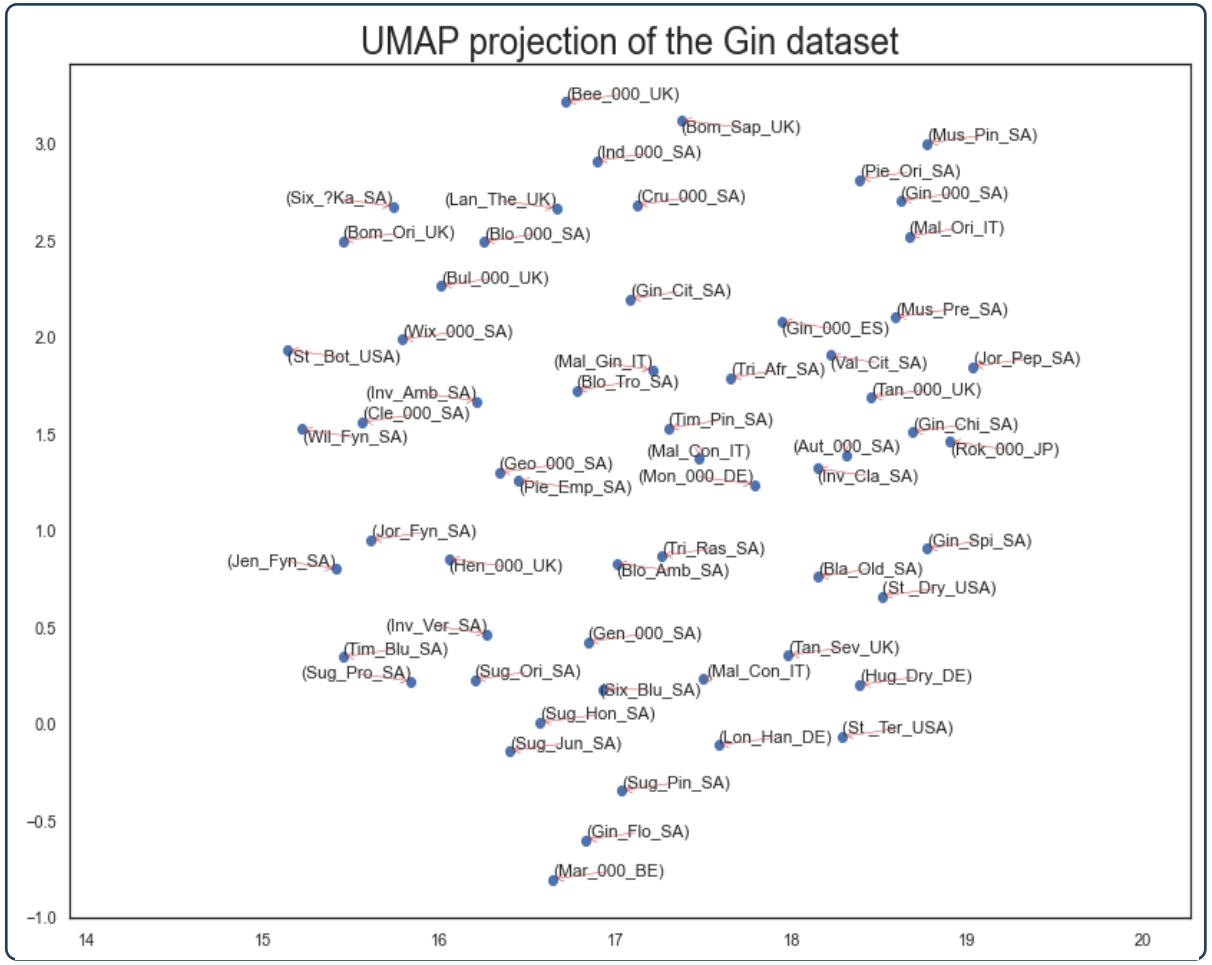
But there are shortcomings…

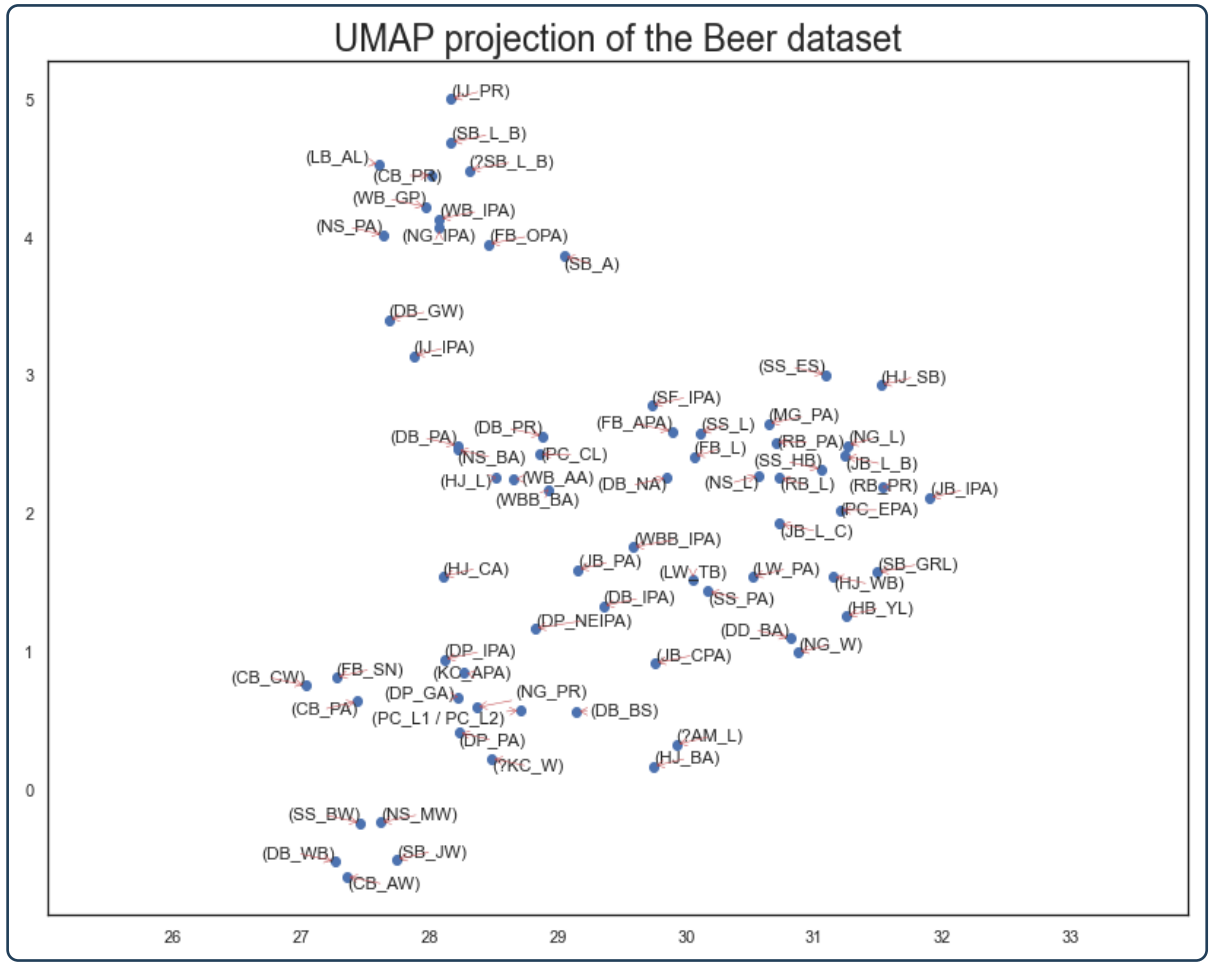
The application of this approach to small product sets:

- Evaluation of various module choices
- Optimization of
    - time
    - effort
    - cost



Application (gin and beer): Better visualization?





Future

How can this be used?

- Once established, easy and cheap access to information
- Benefits for product placement and marketing = increased revenue
- Adds value to the development of future products where the desired properties can be assessed
- = better wine, beer, and gin (Chocolate, soft drinks, etc.)

Future

What is the next product for investigation?

The optimized tool can be deployed in much larger projects like investigating the sensory space of Chenin Blanc wines, specifically looking at old vine and new vine comparisons (Requires rewriting of module for description sources) [SQL, Scrapy, Beautiful Soup, Power Automate]

The older the vine, the better the wine?

Thank you very much for your attention. Questions are most welcome.

Author/Presenter Details

- Name: Markus Kruger
- WhatsApp/Mobile Number: +27 82 969 8255
- Email address: KrugerMarkusG@Gmail.com

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

forward together
sonke siya phambili
saam vorentoe

Acknowledgements:

I wish to thank my supervisors: Dr. J Brand, Prof. B Watson, and Dr. A Buica as well as my collaborator Dr. M Mafata.

Financial support from Department of Information Science, the School for Data Science and Computational Thinking, and the South African Grape and Wine Research Institute, Department of Viticulture and Oenology, all at Stellenbosch University, South Africa has made this project possible.