



13TH PANGBORN SENSORY SCIENCE SYMPOSIUM
PANGBORN 2019



28 July – 1 August 2019 | Edinburgh International Convention Centre, Edinburgh, UK

www.pangbornsymposium.com

Fast and automated sensory analysis:

*Using natural language processing
for descriptive lexicon development*

Leah Hamilton and Jacob Lahne
Virginia Polytechnic Institute and State University

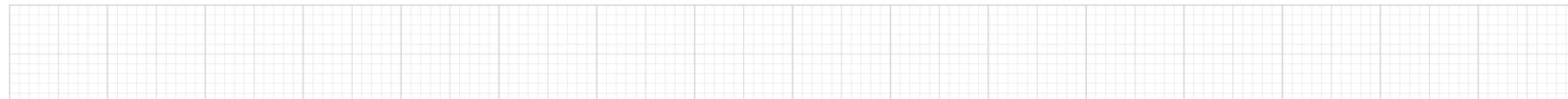
Tweet #Pangborn19

Organised by:



┌ Natural Language Processing (NLP)

- NLP- Computer science methods for automatically processing large amounts of human-generated text
- Sensory science deals with humans, and thus language
- A **case study**: transforming whisky reviews into a lexicon
- Think about applications in your own work

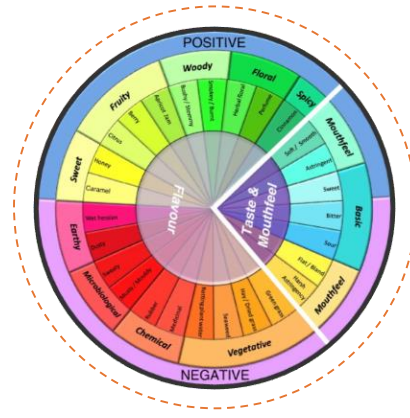


Traditional Lexicon Development



Noble et al., 1984

Modified 3 years after publication



Koch et al., 2012

69 Rooibos samples
22 hours of training
Still only a “first pass”



Drake et al., 2001

240 Cheeses
Narrowed to 70
3-Day workshop
75 hours of training

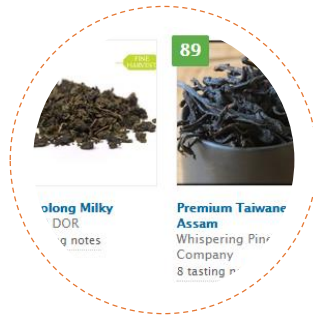
What if we use existing product descriptions?



WhiskyAdvocate

> 4200 reviews

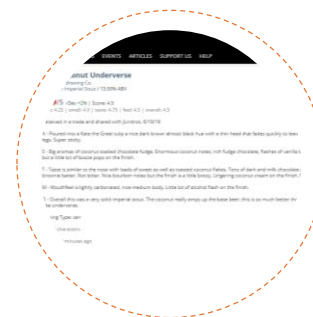
- 2296 Scotches
- 912 Americans
- 227 Canadians
- 211 Irishes



Whiskycast

> 2300 reviews

- 1282 Scotches
- 538 Americans
- 150 Irishes
- 114 Canadians

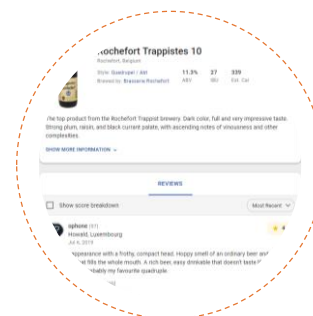


BeerAdvocate

> 1.5 million reviews

Steepster

> 350,000 reviews



RateBeer

> 3 million reviews

Previous Work:

Ickes, Lee, and Cadwallader, 2017

Valente, 2016

Objective

Transform over 6500 existing product reviews into a whisky lexicon or flavor wheel

BERRY BROS. & RUDD NORTH BRITISH 50 YEARS OLD

Country: Scotland
Region:
Type: Single Grain
Bottler: Berry Bros. & Rudd
ABV: 58.9%
Score: 95 points



This single grain Scotch whisky comes from the initial release in the Exceptional Casks range from Berry Bros. & Rudd, and comes from two ex-Sherry butts filled at the old North British distillery on December 18, 1962 and bottled on the same day 50 years later.

The nose is very smooth and creamy with a soft touch of ginger along with vanilla, banana, honey, coconut, and a gentle oakly note. The taste is spicy, tart, and peppery with white pepper, lemon zest, and touches of honey, vanilla, and berries that come out as the spices fade with a dry mouthfeel. The finish is equally dry and tart with hints of white pepper and honey.

An outstanding example of just how good single grains can be! (October, 2014)

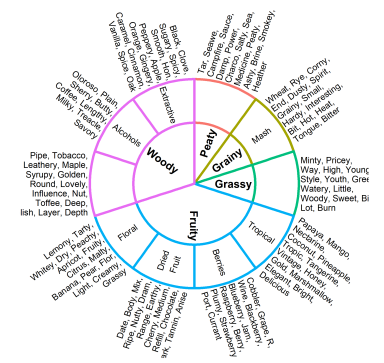
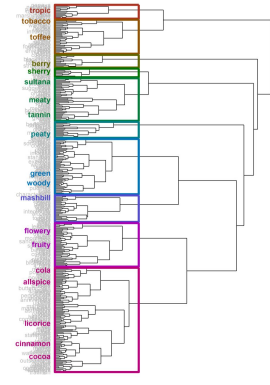
```

<div class="post-header">
  <div class="inner">
    <span class="meta-category"></span>
    <h1 class="post-title">
      Berry Bros. 46030: Rudd North British 50 Years Old
    </h1>
  </div>
</div>
<!--END .post-header -->
<div class="featured-image">
</div>
<!--BEGIN .featured-image -->
</div>
<!--END .featured-image -->
</div>
<!--BEGIN .post-content -->
<div class="post-content">
  <strong>Country</strong> Scotland</div>
<strong>Region</strong> </div>
<strong>Type</strong> Single Grain</div>
<strong>Bottler</strong> Berry Bros. & Rudd</div>
<strong>ABV</strong> 58.9%</div>
<strong>Score</strong> 95 points</div>
<p><a href="http://whiskybase.com/wp-content/uploads/2014/10/BBB-NorthBritish19">The nose is very smooth and creamy with a soft touch of ginger along with vani
</p>
<p><a href="http://whiskybase.com/wp-content/uploads/2014/10/BBB-NorthBritish19">An outstanding example of just how good single grains can be! (October, 2
  
```

AW	Bottle	Country	Region	Score	Title	Type
90	50 50.0%	Whitlaking Farm	USA	90 points	Whitlaking Farm The Best Whisky in Spirit of Heaven Ever Made	Single Grain
91	50 50.0%	Berry Bros. & Rudd	Scotland	95 points	Berry Bros. & Rudd North British 50 Years Old	Single Grain
92	90 90%	Douglas Laing & Co.	Scotland	90 points	Old Highland Cask Four Grain	Single Grain

reviews

anise	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
apple	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bean	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
brandy	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
burn	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
caramel	2	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
cigar	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
close	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
dry	2	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0





Focus Groups

Free Text Surveys

Manual Text Coding

Descriptive Analysis Consensus Training

NLP Tools

- Collecting Text
 - Web Scrapers
- Text Cleaning
 - Language Models
 - Lemmatizers/Stemmers
- Grouping Related Words
 - Principal Component Clustering

Web Scrapers



PROBLEM:

Collecting descriptions for a large, diverse sample of products

SOLUTION:

Web scraping of product reviews

BERRY BROS. & RUDD NORTH BRITISH 50 YEARS OLD

Country: Scotland
Region:
Type: Single Grain
Bottler: Berry Bros. & Rudd
ABV: 58.9%
Score: 95 points



This single grain Scotch whisky comes from the initial release in the Exceptional Casks range from Berry Bros. & Rudd, and comes from two ex-Sherry butts filled at the old North British distillery on December 18, 1962 and bottled on the same day 50 years later.

The nose is very smooth and creamy with a soft touch of ginger along with vanilla, banana, honey, coconut, and a gentle oaky note. The taste is spicy, tart, and peppery with white pepper, lemon zest, and touches of honey, vanilla, and berries that come out as the spices fade with a dry mouthfeel. The finish is equally dry and tart with hints of white pepper and honey.

An outstanding example of just how good single grains can be! (October, 2014)

```

<div class="post-header">
  <div class="inner">
    <span class="meta-category"></span>
    <h1 class="post-title">
      Berry Bros. &#038; Rudd North British 50 Years Old
    </h1>
  </div>
<!--END .post-header -->
</div>

<!--BEGIN .featured-image -->
<div class="featured-image">
</div>
<!--END .featured-image -->

<!--BEGIN .post-content -->
<div class="post-content">
  <p><strong>Country:</strong> Scotland<br />
<strong>Region:</strong> <br />
<strong>Type:</strong> Single Grain<br />
<strong>Bottler:</strong> Berry Bros. & Rudd<br />
<strong>ABV:</strong> 58.9%<br />
<strong>Score: 95 points</strong></p>
<p><img src="http://whiskycaast.com/wp-content/uploads/2014/10/BBR-NorthBritish19
<p>The nose is very smooth and creamy with a soft touch of ginger along with van
<p>An outstanding example of just how good single grains can be! <em>(October, 2

```

	ABV	Bottler	Country	Region	Review	Score	Title	Type	
90	88	57.90%	WhistlePig Farm	USA	The 2018 release of WhistlePig's Boss Hog is named for Mauve, one of the WhistlePig Farm pigs that was the soulmate of Mortimer, who was honored with the 2014 edition of Boss Hog. Mauve loved apples, so Master Distiller Dave Pickerell finished this rye whiskey in Calvados casks. The nose has notes of baked apples, nutmeg, muted rye spices, vanilla, and a touch of rising bread dough. The taste is mouth coating, vibrant, and explodes with a great complexity. Spices, apples, and dried fruits start off separately, then come together to form an apple pie sweetness that is subtle and elegant. The finish is amazingly long with gently fading spices and hints of apple pie, nutmeg, and honey. Wow! (September, 2018)	95 points	WhistlePig Rye The Boss Hog V: Spirit of Mauve Rye Whis		
91	89	58.90%	Berry Bros. & Rudd	Scotland	This single grain Scotch whisky comes from the initial release in the Exceptional Casks range from Berry Bros. & Rudd, and comes from two ex-Sherry butts filled at the old North British distillery on December 18, 1962 and bottled on the same day 50 years later. The nose is very smooth and creamy with a soft touch of ginger along with vanilla, banana, honey, coconut, and a gentle oaky note. The taste is spicy, tart, and peppery with white pepper, lemon zest, and touches of honey, vanilla, and berries that come out as the spices fade with a dry mouthfeel. The finish is equally dry and tart with hints of white pepper and honey. An outstanding example of just how good single grains can be! (October, 2014)	95 points	Berry Bros. & Rudd North British 50 Years Old	Single Gra	
92	90	50%	Douglas Laing & Co.	Scotland	Islay	This is a traditional Old Malt Cask bottling of Port Ellen distilled in 1982 and bottled in September of 2001 at the label's usual 50% ABV. The nose has a touch of driftwood smoke and a subtle peatiness, along with a slight hint of citrus tartness, butterscotch, honey, soft oak, and a sulphur-like flintiness. The taste is thick and spicy with a great black pepperiness, along with butterscotch, honey, lemon zest, grilled pineapple, and a touch of sulphur. The finish is long with a gentle peaty smokiness balanced nicely by honey and butterscotch. Amazing! (November, 2017) (Edin's note: Thanks to Davin de Kergommeaux and Frank Scott for sharing this bottle during a Port Ellen vertical tasting at the 2017 New Brunswick Spirits Festival.	95 points	Old Malt Cask Port Ellen 1982	Single Ma

Tools Used:
BeautifulSoup (Python),
rvest (R)



Pros

Fast (>1 product/second)
Don't need product samples
Large sample size

Cons

Messy Data
Lots of data = lots of analysis time
No universal tool (website-specific code)

Text Cleaning



PROBLEM:

Non-descriptive words (e.g. good, the, pricey) in product descriptions

Desired Descriptors

The taste is spicy, tart, and peppery with white pepper, lemon zest, and touches of honey, vanilla, and berries that come out as the spices fade with a dry mouthfeel.



PROBLEM:

Non-descriptive words (e.g. good, the, pricey) in product descriptions

SOLUTION:

Language Model- Part of Speech Filtering

Parts of Speech (Manual)

The **taste** **is** **spicy** **tart**, and **peppery** with **white** **pepper**, **lemon** **zest** and **touches** of **honey**, **vanilla**, and **berries** that **come** out as the **spices** **fade** with a **dry** **mouthfeel**.

Nouns

Verbs

Adjectives



PROBLEM:

Non-descriptive words (e.g. good, the, pricey) in product descriptions

SOLUTION:

Language Model- Part of Speech Filtering

Parts of Speech (Automated)

The **taste** **is** **spicy** **tart**, and **peppery** with **white** **pepper**, **lemon** **zest** and **touches** of **honey**, **vanilla**, and **berries** that **come** out as the **spices** **fade** with a **dry** **mouthfeel**.

Nouns

Verbs

Adjectives

The word "CLOVE" is written in a large, bold, red font. The letters have a jagged, dripping appearance, with red liquid-like drips hanging from the bottom of each letter. The background is white.

15% of occurrences of the word
“clove” are tagged as verbs

Most NLP tools were trained on
novels, journal articles, tweets,
NOT descriptive text



Pros

Takes context into account
Flexible

Cons

No tagger for sensory relevance (yet)
No language model is 100% accurate



PROBLEM:

Multiple forms of the same word

SOLUTION:

Lemmatizing & Stemming

FRUIT -S
-IER
-Y
-INESS



CLOVE

Lemmatizers are more state of the art, but they're not perfect

Some lemmatizers turn "clove" into "cleave", something no stemmer would do!

┌

Pros

Cons

Stemmers

Easy to write and customize
Small list of endings for adjectives

Creates artifacts (e.g. turning “red” to “r”)
Can’t handle irregular forms

Lemmatizers

Maintains readability
Better handling of irregular forms
Utilizes part of speech tags

Doesn’t combine adjectives and nouns
May be limited by dictionary
More computationally intensive

Clustering

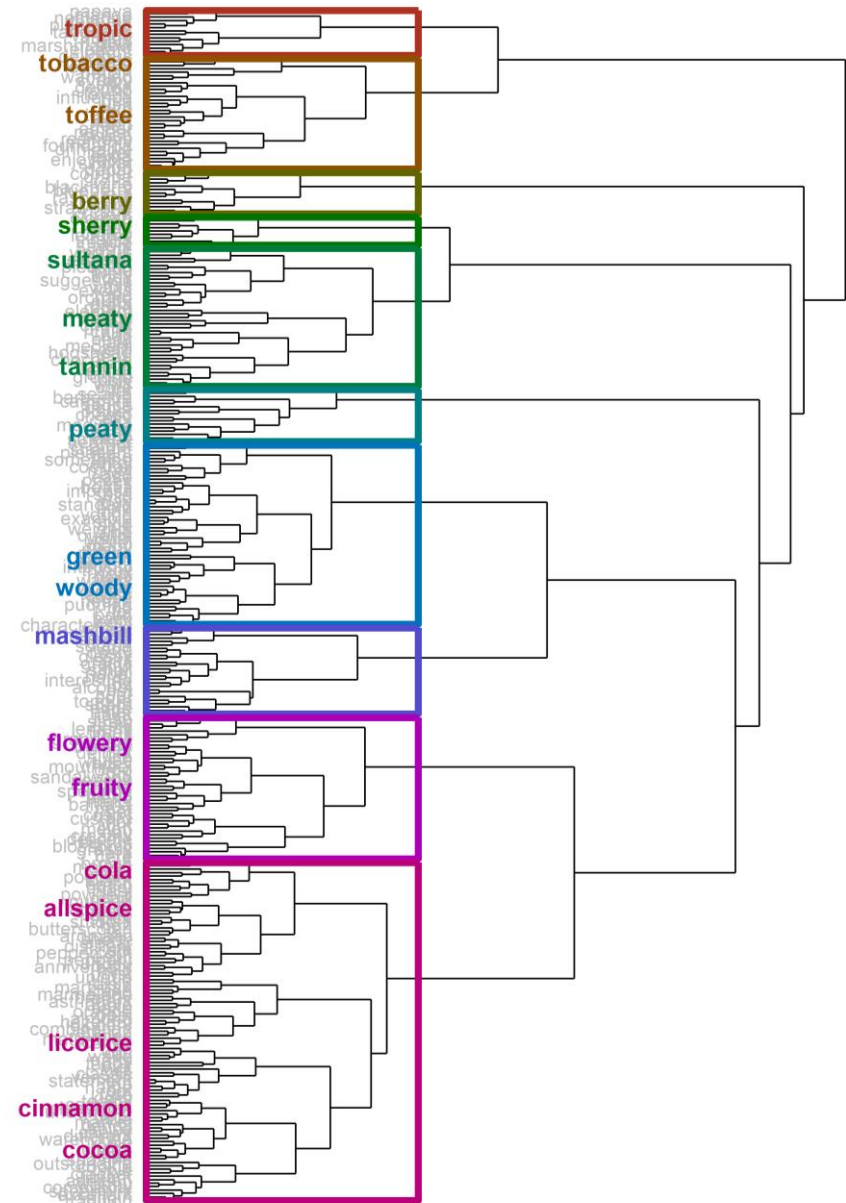


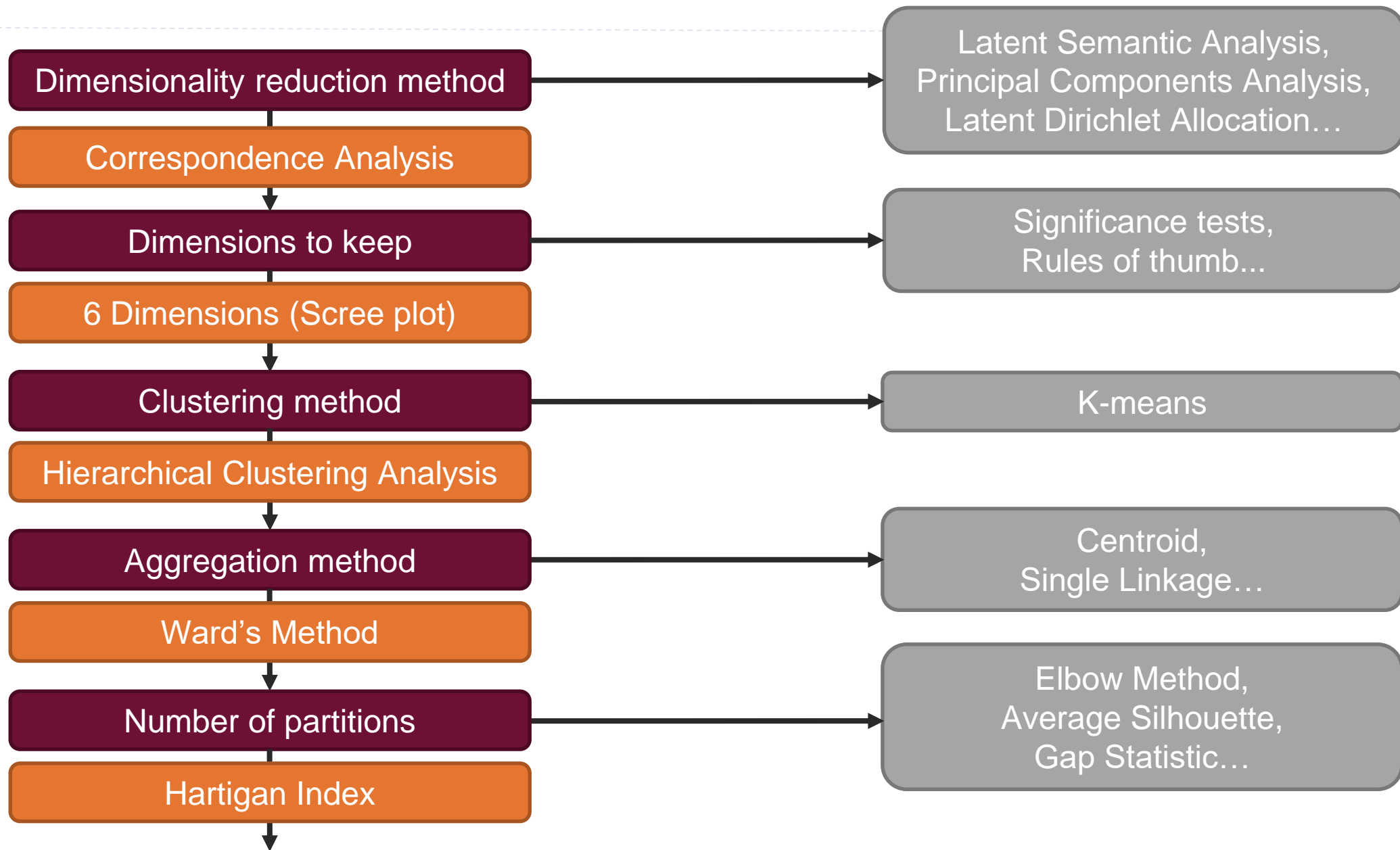
PROBLEM:

Grouping a large ($>10^3$) list of words into categories, like in a flavor wheel

SOLUTION:

Singular Value Decomposition + Clustering





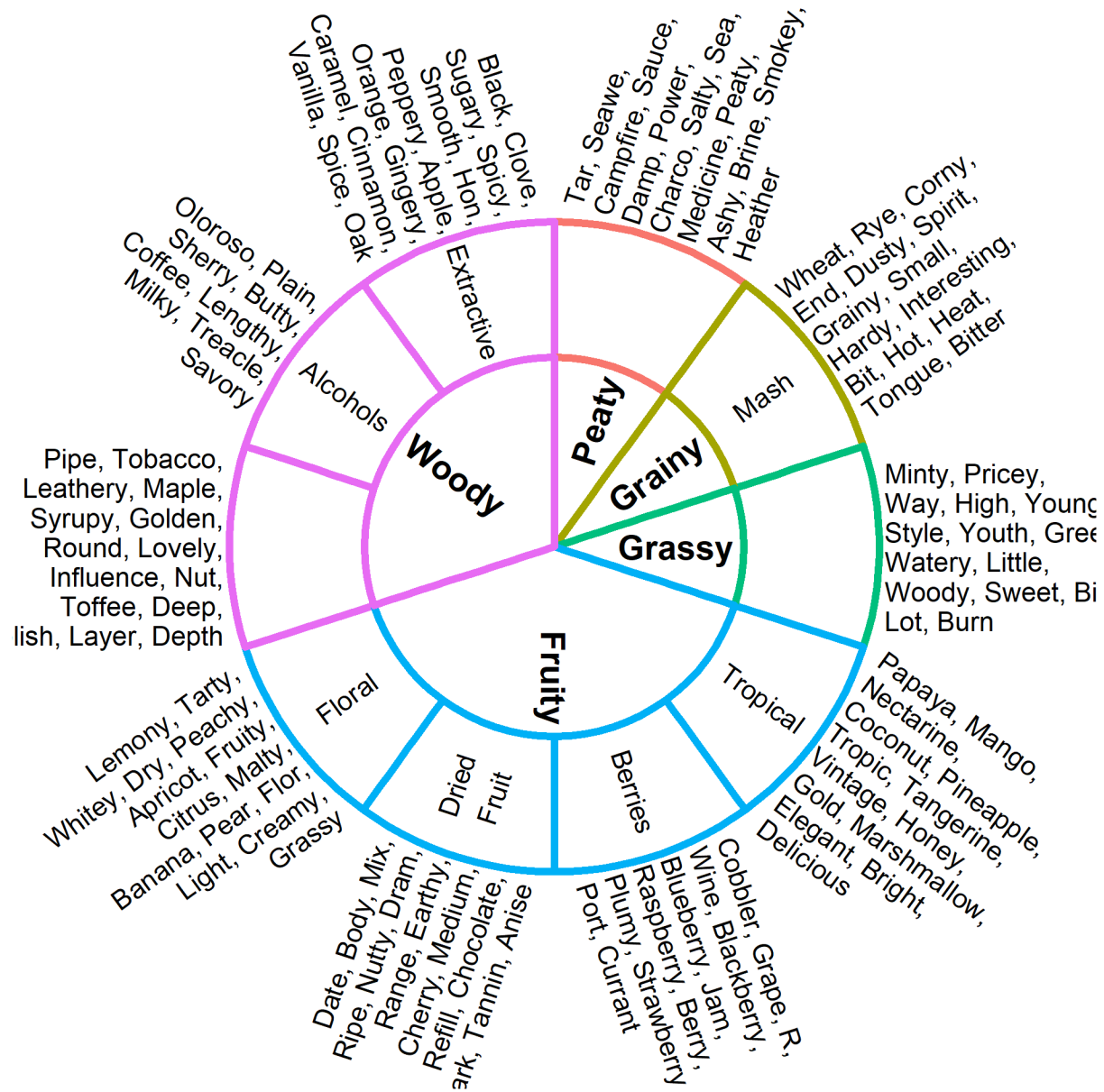
┌

Pros

- Can group very large numbers of words
- Visualization options
- Can find broad synonym clusters

Cons

- Many arbitrary decisions
- Highly dependent on arbitrary decisions
- Commonly co-occurring \neq synonymous

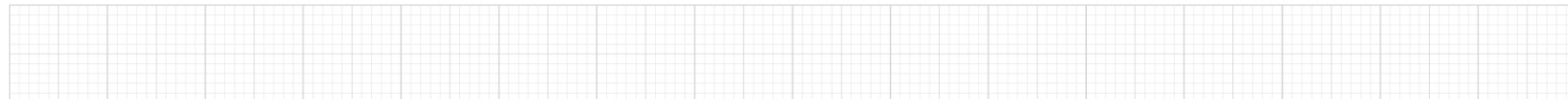


Tools

	Potential Tools
Web Scraping	BeautifulSoup (Python) rvest (R)
Language Modeling	spaCy (Python) Stanford CoreNLP (Java) cleanNLP (R)
Lemmatizing	textstem , koRpus , cleanNLP , lexicon (R) spaCy (Python) TreeTagger
Clustering	Base R , cluster , flashClust (R) scikit-learn , SciPy (Python)
Network Analysis	IGraph (R/Python/Mathematica/C++) Gephi

References

- Drake, M. A., McIngvale, S. C., Gerard, P. D., Cadwallader, K. R., & Civille, G. V. (2001). Development of a Descriptive Language for Cheddar Cheese. *Journal of Food Science*, 66(9), 1422-1427.
- Ickes, C. M., Lee, S. Y., & Cadwallader, K. R. (2017). Novel Creation of a Rum Flavor Lexicon Through the Use of Web-Based Material. *Journal of Food Science*, 82(5), 1216-1223. <https://doi.org/10.1111/1750-3841.13707>
- Koch, I. S., Muller, M., Joubert, E., van der Rijst, M., & Næs, T. (2012). Sensory characterization of rooibos tea and the development of a rooibos sensory wheel and lexicon. *Food Research International*, 46(1), 217-228. <https://doi.org/10.1016/j.foodres.2011.11.028>
- Noble, A., Arnold, R., Buechsenstein, J., Leach, E., Schmidt, J., & Stern, P. (1987). Modification of a standardized system of wine aroma terminology. *American Journal of Enology and Viticulture*, 38(2), 143-146.
- Valente, C. C. (2016). *Understanding South African Chenin Blanc wine by using data mining techniques applied to published sensory data*. Stellenbosch University.



┌

Questions?

LEAH HAMILTON (HLEAH@VT.EDU)

JACOB LAHNE (JLAHNE@VT.EDU)

