

# Of digestive biscuits, old leather, and smoked ham

Making sense of the language of Scotch whisky

Praveen Gowtham

# Scotch



Sugars from dried malted barley fermented, distilled, and aged for years in a numerous possible array of casks. Must be made in Scotland.

# A complex and long process



# Complex Taste. Colorful lexicon.

- "The finish dries and becomes pleasantly rubbery, as a touch of truffle oil emerges."
- "Seaweed-led, with a hint of vanilla ice cream and more than a whiff of notes from the first aid box (TCP, plasters etc)."



# Making sense of the language

- Language of Scotch tastes/smells: metaphor, simile
- Across a set of Scotch tasting notes:
  - Construct lexicon used to describe whisky tastes/smells.
  - Are there groups of descriptors sharing common themes?  
Salty, fruity, bittersweet, etc.
  - Can scotches be described in a more robust way using these groups?

# Getting expert tasting notes



Master of Malt (MoM): online store + expert tasting notes from The Chaps at MoM.

Scrapy spider: extract tasting notes + bottling details on  $\approx 8,600$  Scotches

# Getting expert tasting notes

Glenfarclas 25 Year Old (70cl, 43%)

★★★★★ (73 Reviews)



Want Glenfarclas 25 Year Old for Christmas?

Ask #WhiskySanta

Wish

What is this?

Awards for Glenfarclas 25 Year Old



Gold  
Distillers' Single Malts 21 years old and over - 2019  
International Spirits Challenge



Silver Outstanding  
Cask Strength Malts - Cask Strength - 2019

## Glenfarclas 25 Year Old Bottling Note

Here we have an absolute classic Speyside from a long-time family-owned distillery. This 25 year old single malt from Glenfarclas spent its days in 100% Oloroso sherry casks. It's complex and delicately sherried, a whisky that's sure not to disappoint. It won gold at the 2019 International Spirits Challenge and makes a great after dinner treat. Awesome.

Consider this similar product...

Speyside 26yo TBWC

\$119.08



Sorry about this...

We can't actually ship Glenfarclas 25 Year Old to you in United States at the moment. We're working on a solution.



Shop now

More Alternatives...



\$106.59

21 Year Old



\$54.35

Glenfarclas 105



\$37.21

10 Year Old



\$55.54

15 Year Old



\$136.15

105 Special Edtn

Add to a Wishlist

## Tasting Note by The Chaps at Master of Malt

**Nose:** A touch of menthol. Classic Sherry notes, less pronounced than the 15 year old perhaps, but the complexity is greater.

**Palate:** Incredibly clean central palate, but amazingly tangy. Really zingy top notes. Sherry and creamy barley, hints of gingerbread and nutty chocolate.

**Finish:** Very, very long and complex finish, oak-rich, hints of smoke, cocoa, buttery...

**Overall:** One of our all-time favourite Speyside malts.



# Extracted Information

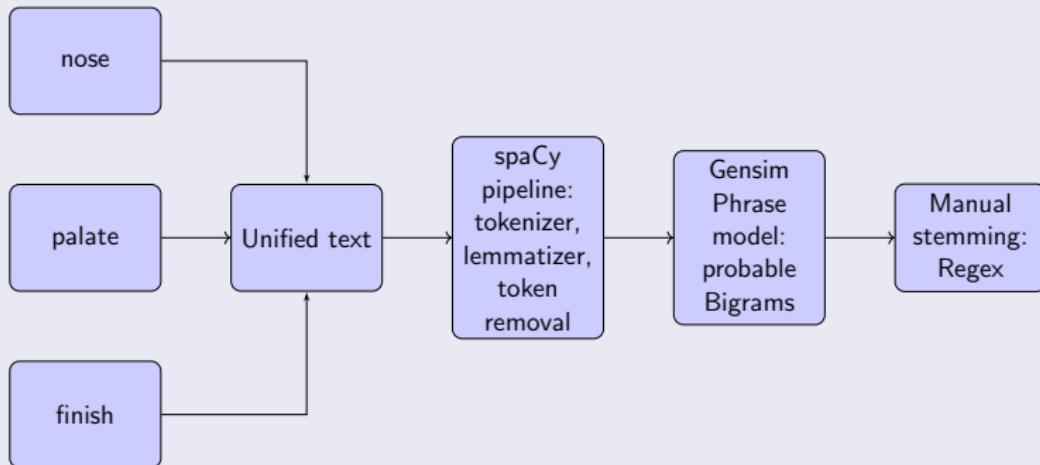
name	The name of the scotch.
distillery	Distillery that produced the scotch.
bottler	Some scotches are aged/bottled by 3rd parties.
region	Whisky making region that the distillery belongs to.
age	Maturation time (in years).
ABV	Alcohol by volume.
price	MoM is also a store. Price in USD.
description	General notes about the Scotch.
nose	Free text description of the scent of the Scotch.
palate	Free text description of the taste of the whisky.
finish	Free text description of aftertastes.
chill filter	Is the whisky chill filtered?
cask strength	Is the whisky bottled at cask strength?
maturity	Cask(s) used to age the whisky.

Table: Scotch attributes scraped from MoM.

# Text Preprocessing/Normalization



## Text Processing Flow



## Dictionary and corpus generation

- Gensim dictionary from tokens
    - 4145 unique descriptors
  - Filter tokens:
    - Lower bound: 0.5% of reviews
    - Upper bound 50% of reviews
    - Reduced dictionary: 477 descriptors
  - Bag-of-words (BoW) representation
    - Gensim corpus
    - Sparse review/descriptor-frequency matrix

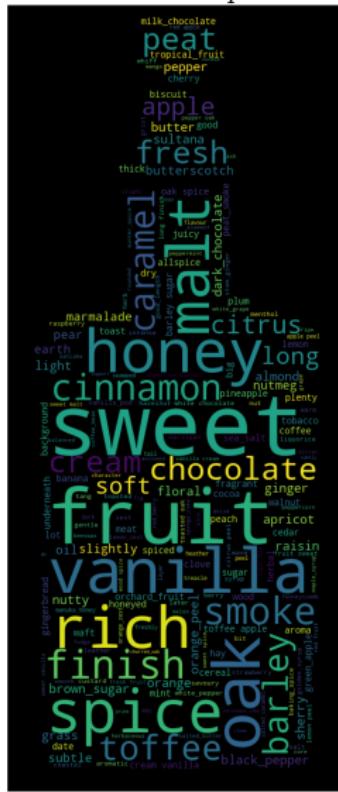
## Scotch descriptors



## Descriptors

- Common descriptors: sweet, vanilla, fruit, spice, oak, malt, etc.
  - Groupings among descriptors: taste types?
  - Regional differences

## Scotch descriptors



# Scotch producing regions

Scotches from different regions  
are said to have certain  
character. Let's see.



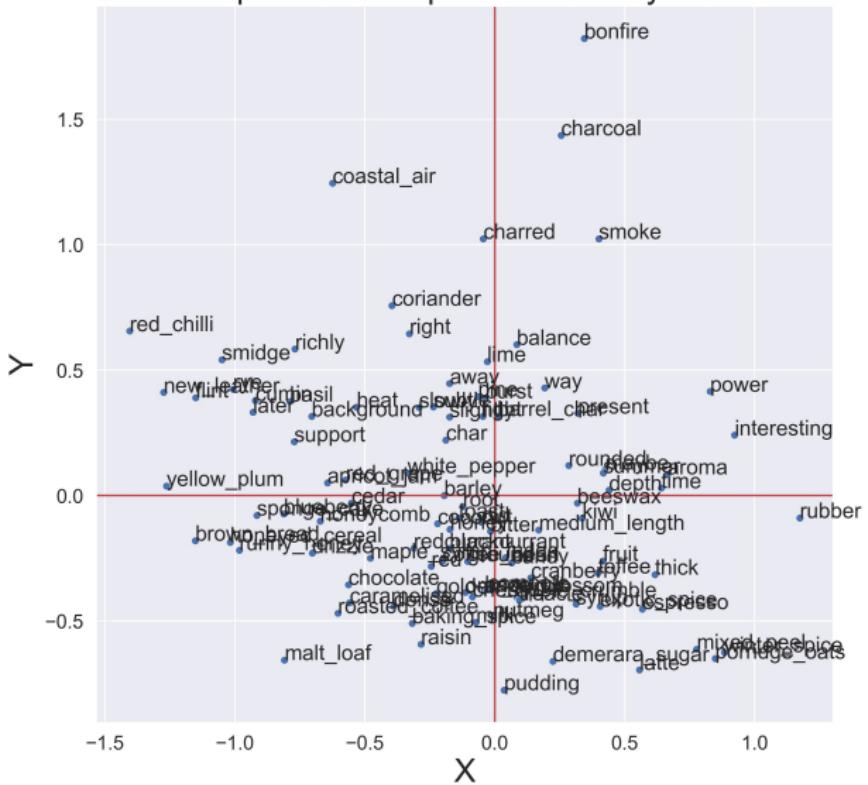
# Correspondence Analysis (CA)

Visualize descriptors and their groupings in a low-dimensional space.

- CA embeds descriptors in low-D space: similar co-occurrence of descriptors across reviews are related.
- Document/term-frequency correlations: embed tasting notes / scotches in same space as descriptors.
- *Technical:*
  - Double-center document/term-frequency matrix.
  - Take residual (co-occurrences beyond what is expected on average), SVD.
  - Project into 2D (first two singular values). Symmetrize bases for descriptors and scotches for 2D embedding.

## Correspondence Analysis: Descriptors

Scotch Descriptors: Correspondence Analysis Visualization



# Correspondence Analysis: Scotches by region



Scotches likely exhibit multiple flavor types. Flavor breakdown?

# Topic Modeling

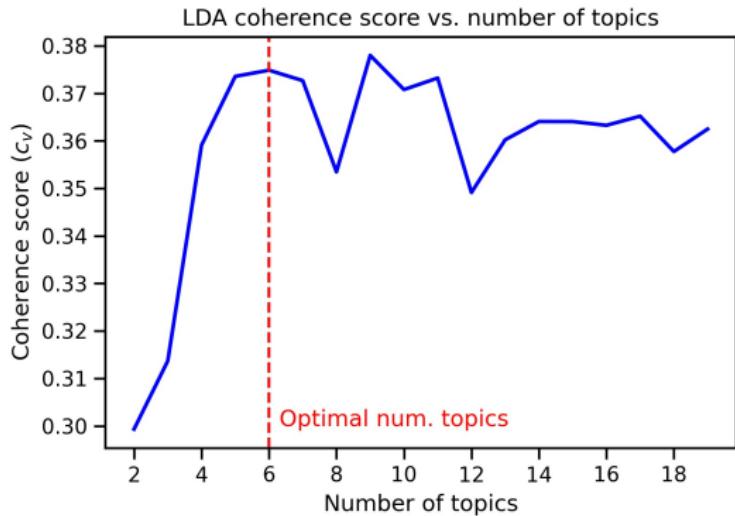
Want to construct groups of descriptors and break each tasting note down into a combination of these groups.

- Latent Dirichlet Allocation (LDA)
- Assume  $K$  topics (taste groups)
- Two-level probabilistic inference:
  - Distribution of topics on each document (tasting note)
  - Distribution of tokens in each topic
- Nice thing about LDA: Human interpretable topics.

# Tuning the number of topics

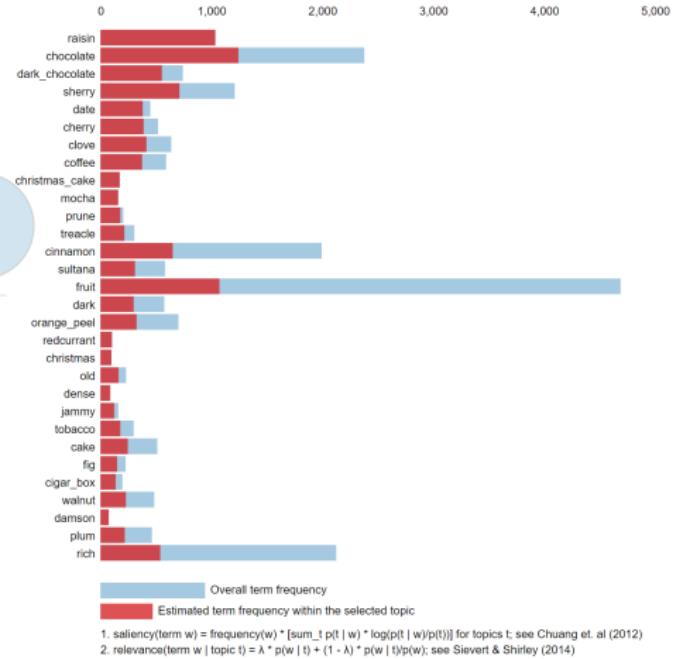
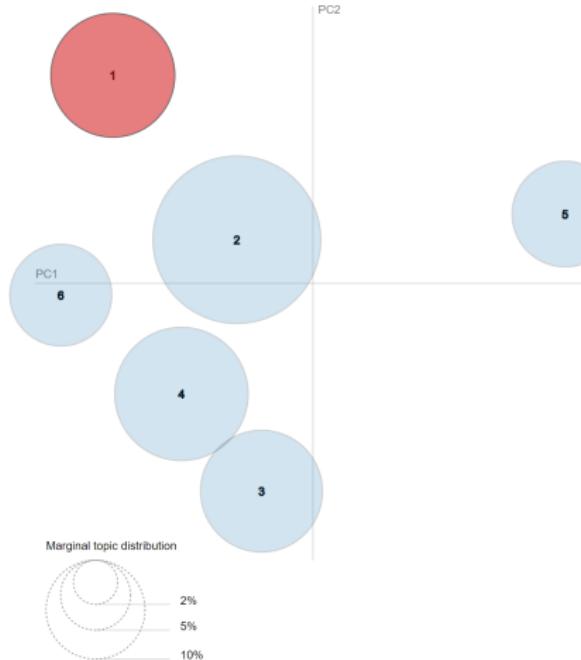
Number of topics as model hyperparameter.

- Tuning for human interpretability.
- Quantitative measures?
- $c_v$  score: agg. mutual info between descriptors in topic



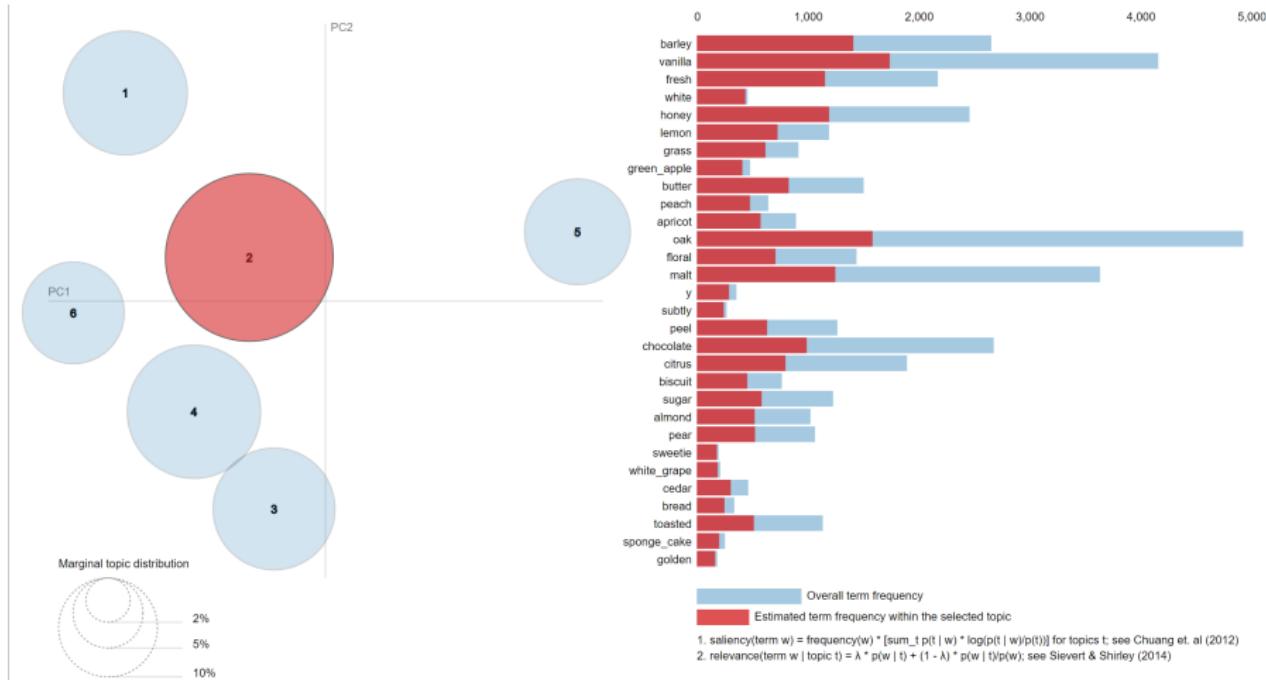
Tuning from 2-20 topics: 6 looks good.

# Visualizing the topic model

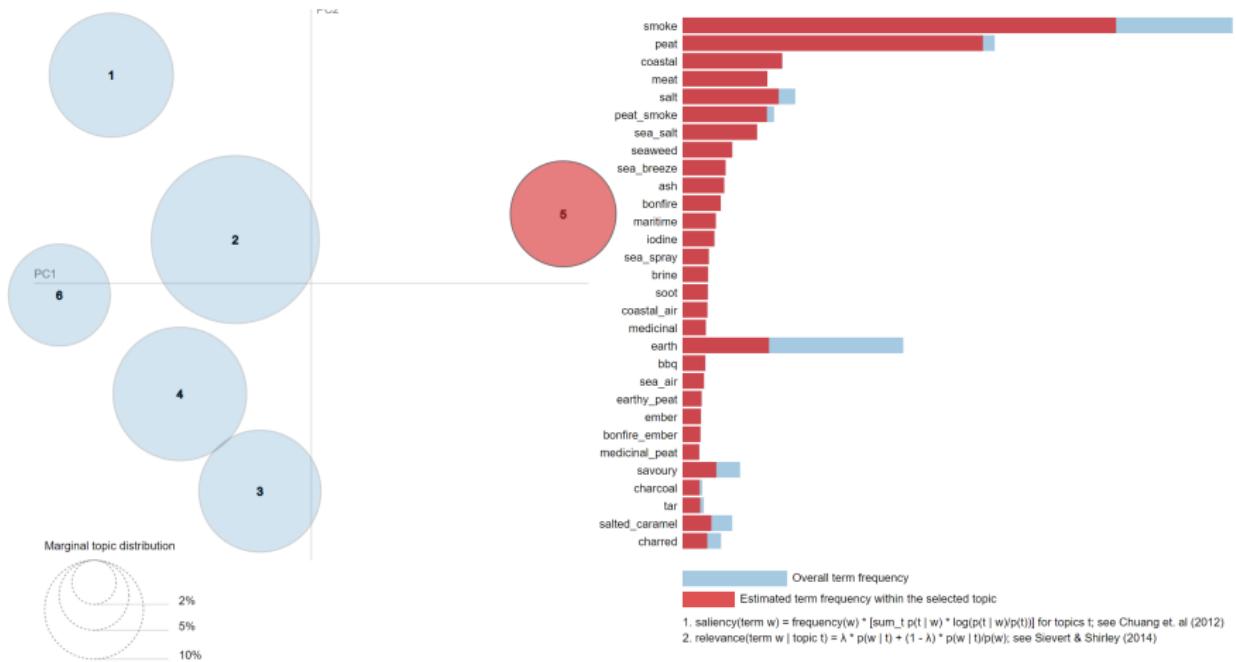


1. saliency(term w) = frequency(w) \* [sum\_t p(t | w) \* log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)  
2. relevance(term w | topic t) =  $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$ ; see Sievert & Shirley (2014)

# Visualizing the topic model



# Visualizing the topic model



# Gist of the topics

Topics	Description
1	Aromatic spice, candied fruit, molasses, Christmas. Dark, sweet flavors.
2	Citrus and floral with cream, malty notes.
3	Herbal, tannin, citrus, wood spice. Drier notes.
4	Wood, smoke, and pepper/heat. But also combined with cream and custard. Richness.
5	Smoke, peat, salt, meat. All the savoury notes.
6	Berries, rhubarb, dark chocolate, jam, fruits, tartness but sweet.



Figure: Topic 1: Christmas cake



Figure: Topic 5: Bacon.

# Taste profile comparison: Islay Scotches

- **Ardbeg 10:** sweet, vanilla, lemon, lime, ardbeg, smoke, love, ridge, vanilla, mountain, peat, citrus, fruit, cloud, sea\_spray, long, glorious, sea, caramel, beach\_bonfire, smoke
- **Laphroaig 10:** 'seaweed', 'vanilla', 'ice\_cream', 'tcp', 'plaster', 'oak', 'spice', 'cardamom', 'black\_pepper', 'chilli', 'big', 'muscular', 'peat', 'spice', 'liquorice', 'big', 'dose', 'salt', 'slightly', 'sweet', 'beauty', 'classic', 'iodine', 'plaster', 'cool\_wood', 'smoke', 'big', 'savoury', 'tarry', 'iodine'

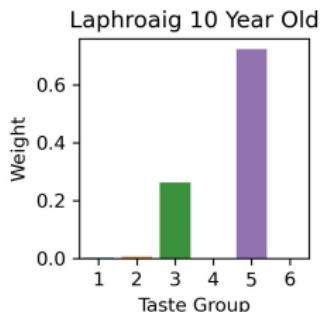
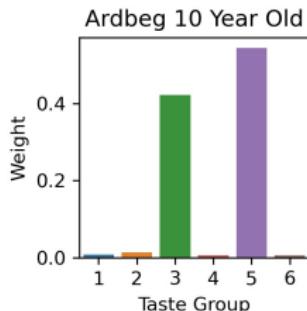
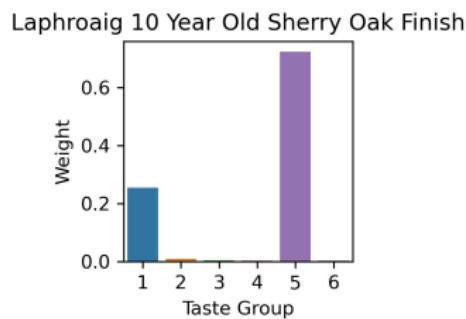
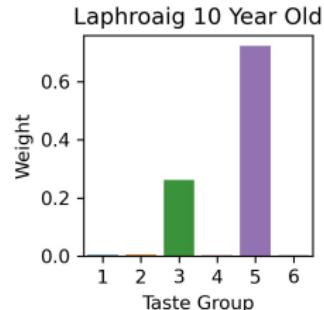


Figure: Taste group 3 = Herbal, tannin, citrus, wood spice.  
Drier notes. Taste group 5 = Peaty, salty, meaty notes.

# Taste profile comparison: Effect of Sherry Finish

- **Laphroaig 10:** 'seaweed', 'vanilla', 'ice\_cream', 'tcp', 'plaster', 'oak', 'spice', 'cardamom', 'black\_pepper', 'chilli', 'big', 'muscular', 'peat', 'spice', 'liquorice', 'big', 'dose', 'salt', 'slightly', 'sweet', 'beauty', 'classic', 'iodine', 'plaster', 'cool\_wood', 'smoke', 'big', 'savoury', 'tarry', 'iodine'

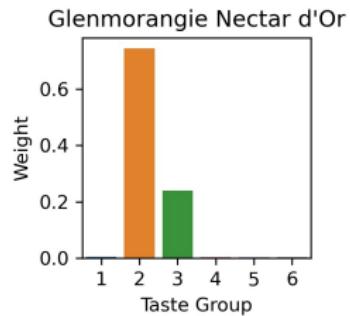
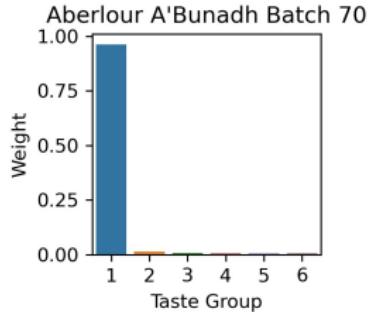
- **Laphroaig 10 Sherry Finish:**  
'roasted', 'cedar', 'peat\_smoke', 'iodine', 'away', 'dark\_chocolate', 'honey', 'vanilla\_pod', 'meat', 'maple\_syrup', 'bbq', 'lemon', 'charred\_oak', 'smidge', 'coffee', 'balanced', 'finish', 'sherry', 'sweet', 'smouldering', 'peat'



**Figure:** Taste group 3 = Herbal, tannin, citrus, wood spice. Drier notes. Taste group 5 = Peaty, salty, meaty notes. Taste group 1 = Nuts, molasses, candied berries, aromatic spice, and dark chocolate. Dark, sweet flavors

# Taste profile comparison: On the sweeter side

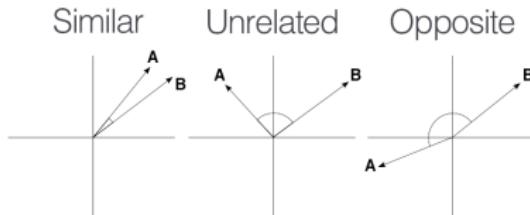
- **Aberlour A'Bunadh:**  
'gingerbread', 'orange\_zest',  
'black\_pepper', 'thick',  
'sherry', 'vanilla\_pod',  
'christmas\_cake', 'cherry',  
'bakewells', 'hefty',  
'helping', 'cinnamon',  
'cedar', 'cinnamon', 'raisin'
- **Glenmorangie Nectar d'Or:**  
'silk', 'syrup', 'syrup',  
'sponge', 'vanilla', 'custard',  
'butter', 'baklava', 'toasted',  
'brown\_sugar', 'sweet',  
'citrus', 'lemon\_curd',  
'vanilla', 'shortbread', 'oak',  
'spice', 'gingerbread', 'fruit',  
'drizzle', 'runny\_honey',  
'dry', 'oak', 'lemon',  
'shortbread',  
'gingersnap\_biscuit'



**Figure:** Taste group 1 = Nuts, molasses, candied berries, aromatic spice, and dark chocolate. Dark, sweet flavors.  
Taste group 2 = Citrus and floral with cream, malty notes.  
Taste group 3 = Herbal, tannin, citrus, wood spice. Drier notes.

# Building Recommendation Engine

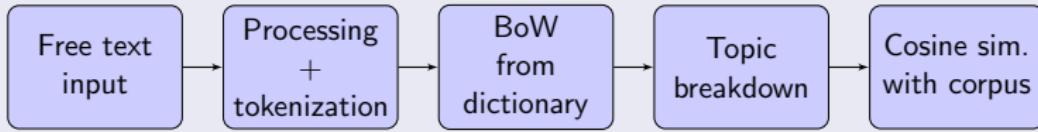
- Scotch review as topic vector
- Cosine similarity and ranking



**In-Corpus:** rank similarity of scotches to input scotch (all in corpus). Gensim index similarities.

**Out-of-Corpus:** Free text input of tastes. Get scotch recommendations.

## Out-of-corpus recommender flow



# In-corpus recommendations

**Aberlour A'Bunadh:** 'gingerbread', 'orange\_zest', 'black\_pepper',  
'thick', 'sherry', 'vanilla\_pod', 'christmas\_cake', 'cherry',  
'bakewells', 'hefty', 'helping', 'cinnamon', 'cedar', 'cinnamon',  
'raisin'

Edradour 10 Year Old

'brown\_sugar', 'treacle', 'clove',  
'cinnamon', 'load', 'raisin',  
'walnut\_loaf', 'damson\_jam',  
'richly', 'chocolate', 'rum', 'like',  
'spice', 'chocolate', 'date'



# In-corpus recommendations

**Aberlour A'Bunadh:** 'gingerbread', 'orange\_zest', 'black\_pepper',  
'thick', 'sherry', 'vanilla\_pod', 'christmas\_cake', 'cherry',  
'bakewells', 'hefty', 'helping', 'cinnamon', 'cedar', 'cinnamon',  
'raisin'

Edradour 10 Year Old

'soft', 'red', 'fruit', 'brown\_sugar',  
'clove', 'chocolate', 'brazil\_nut',  
'dry', 'orange\_peel', 'red\_wine',  
'pecan', 'raspberry\_jam',  
'crystallised\_ginger', 'cacao',  
'juicy', 'blackcurrant'



# Out-of-corpus recommendation

**Text input:** "light, citrus, lemon, grapefruit, dry, crisp, sea salt, salted caramel"

Longrow Peated



## Recommendation: *Tasting*

Notes: 'light', 'sweet', 'green',  
'grape', 'rhubarb', 'great',  
'contrast', 'big', 'blast', 'smoke',  
'salt', 'leather', 'maritime',  
'combination', 'harbourmaster',  
'jacket', 'salt'.

# Conclusion/Future Directions

- Dictionary for Scotch tasting notes
- Different meaningful sectors of taste descriptors
- LDA Topic model captures general taste profiles of Scotches
- Room for improvement:
  - Topics (vector) orthogonal but topics actually have some overlap (Topics 1 and 6, 3 and 4)
  - When is a taste descriptor 'big' ? Or when is it just a 'hint of' a given descriptor? Degree.
  - LDA + word embedding model for context?
- Get me some of that Longrow Peated and have a whee dram by the fireplace this evening!