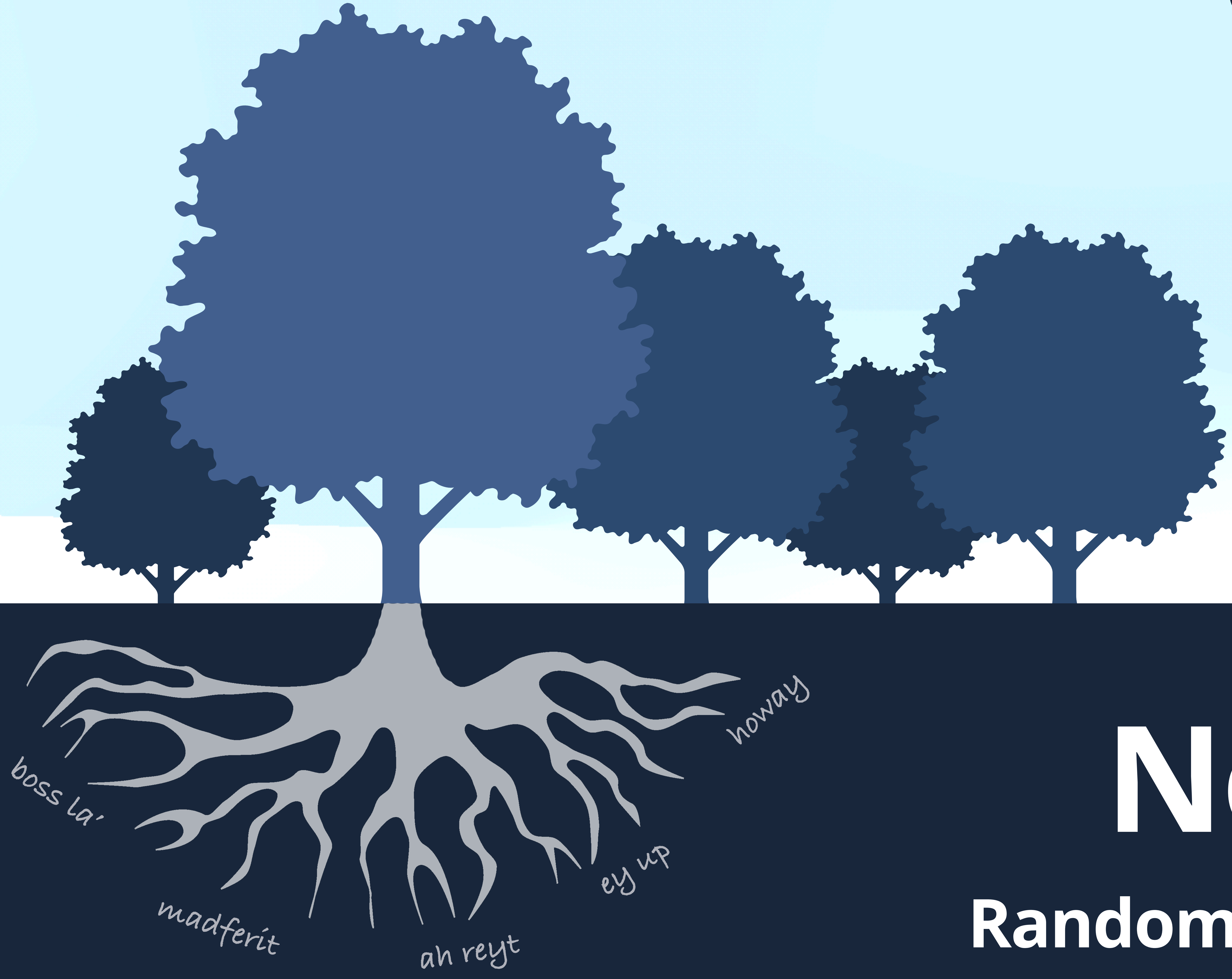
UKLVC15, Lancaster University

George Bailey · *University of York*



Northern roots

Random forests and northern English dialect levelling revisited

Dialect levelling

- “the eradication of socially or locally marked variants […] in conditions of social or geographical mobility and resultant dialect contact” (Milroy 2002: 7)

- **Multiple sources of evidence:**

  - Reduction of local forms in studies of specific dialects, e.g. the FACE and GOAT vowels in Tyneside English (Watt 2002)

  - Loss of regional diversity in more spatially-widespread dialectological studies (e.g. Britain, Blaxter and Leemann 2021; MacKenzie, Bailey and Turton 2022)

  - Perceptual evidence from dialect recognition tasks (e.g. Kerswill & Williams 2002)

  - ‘Machine learning’ dialect classification (e.g. Strycharczuk et al. 2020)

Random forests and *General Northern English* (GNE)

(Strycharczuk et al. 2020)

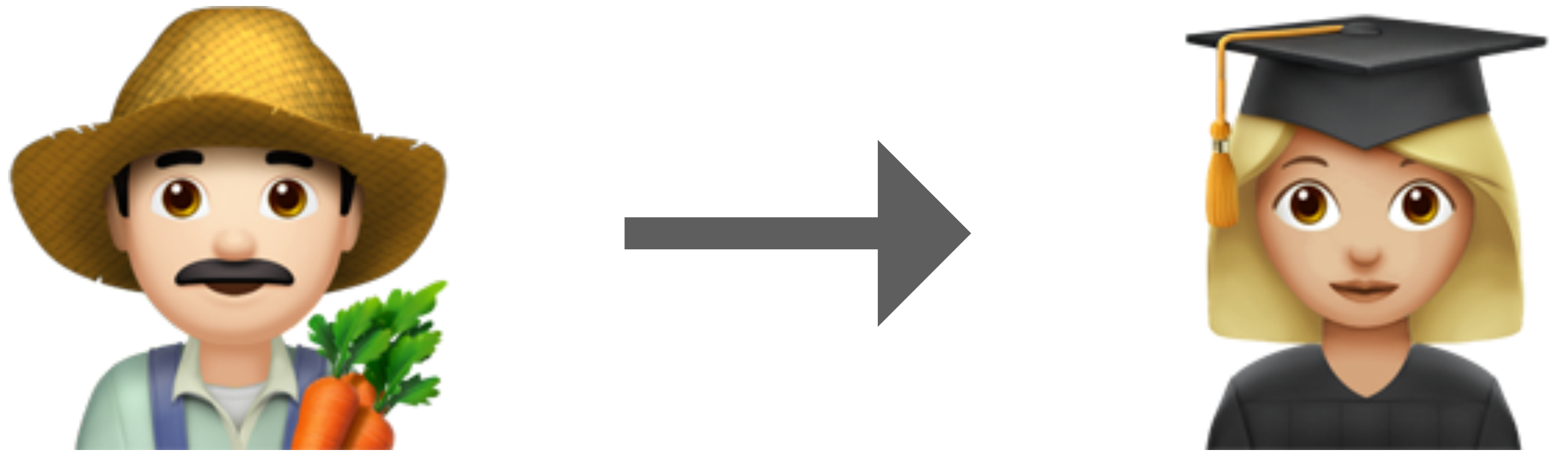
Random forests: machine-learning classification technique to generate predictions based on the output of multiple decision trees

- Used by Strycharczuk et al. (2020) in a novel computational approach to identifying dialect levelling in the North of England
  - use statistical models to quantify the level of **mutual confusability** between the dialects of **Manchester**, **Liverpool**, **Leeds**, **Sheffield** and **Newcastle**
  - if the models struggle to accurately classify speakers into their respective dialect groups → dialect levelling has taken place

Random forests and *General Northern English* (GNE)

(Strycharczuk et al. 2020)

- They train models based on vowel systems: F1 and F2 measurements for 23 vowel categories in English
- Recordings taken from the *English Dialects App* corpus: read passage from 105 speakers
  - “a typical speaker in our sample is an urban white woman in her 30s with a university degree”



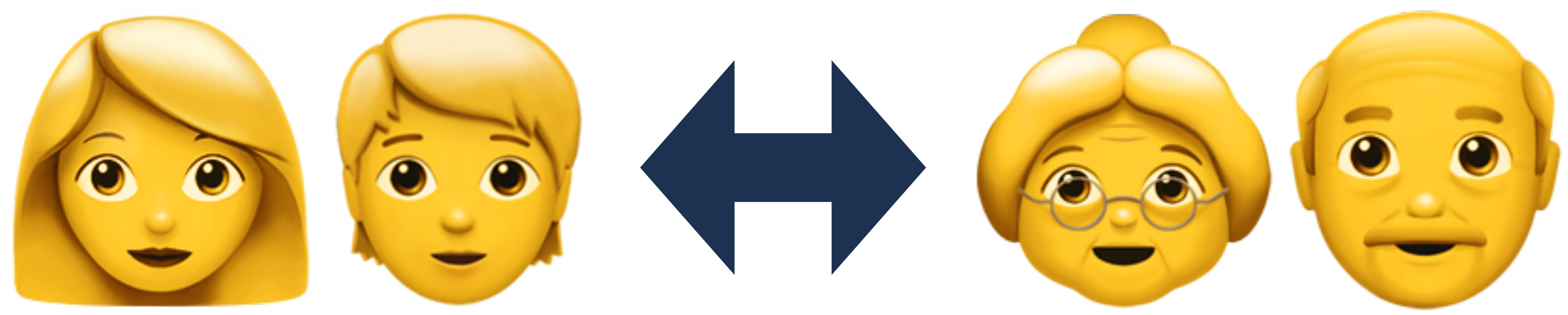
Random forests and *General Northern English* (GNE)

(Strycharczuk et al. 2020)

- Results reveal higher confusability rates between **Manchester**~**Leeds**, and between **Leeds**~**Sheffield** → **dialect levelling** to a *General Northern English*
  - "a pan-regional standard accent associated with middle-class speakers"
  - speakers who demonstrate broadly northern features (e.g. absence of FOOT–STRUT split and BATH–TRAP split) but lacking more locally-specific features

This study

Adopting the same computational approach using random forests, *but…*



…modelling older and younger speakers separately to investigate levelling *diachronically*.

Are younger speakers more difficult to classify?

…modelling dialects more holistically using survey data covering **phonological**, **lexical**, and **morphosyntactic** features

Methodology

Data: Sample

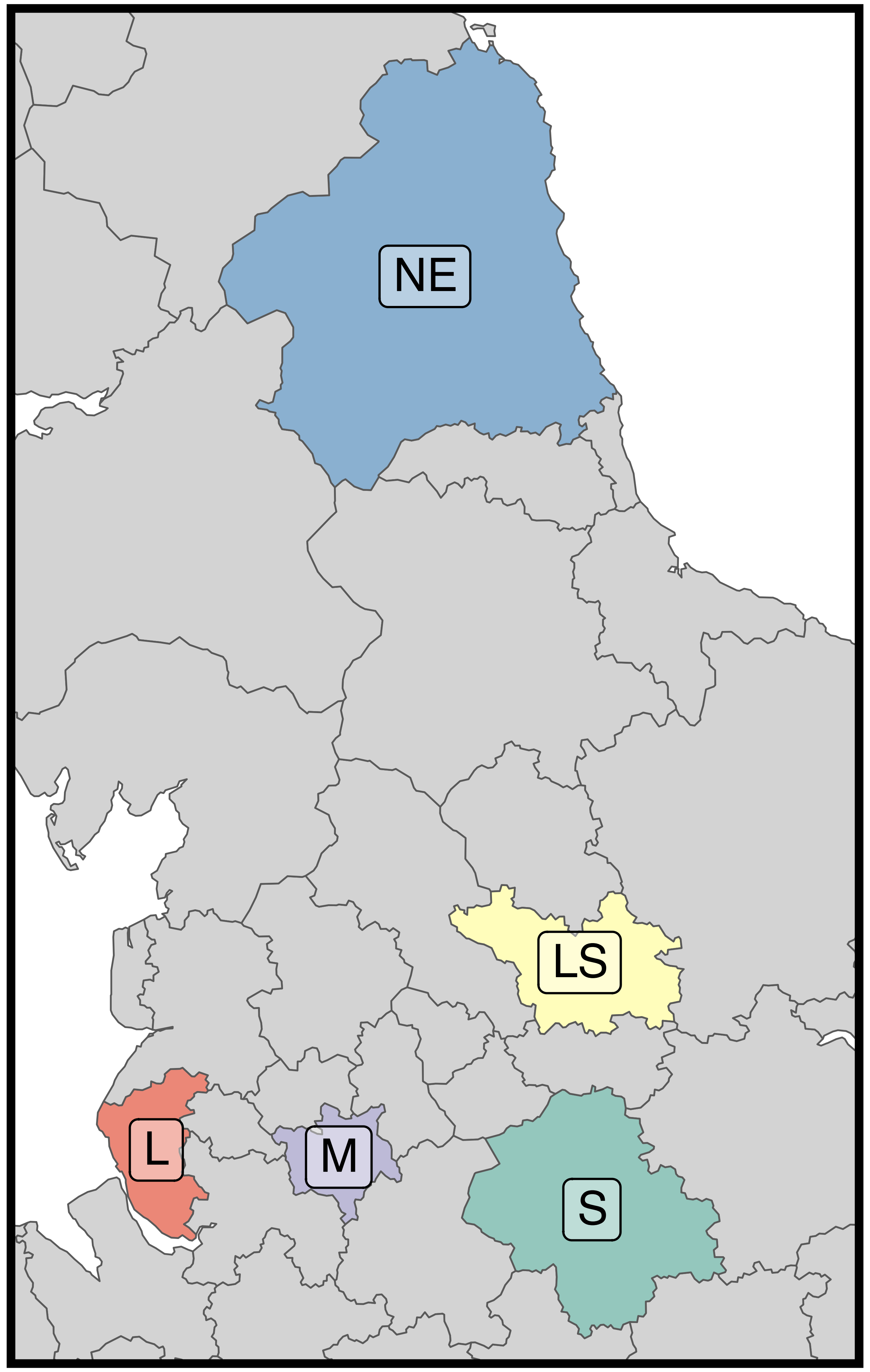
www.ourdialects.uk

- *Our Dialects* survey: **over 20,000 responses** geolocated by the postcode district they lived longest between ages 4–13 (see MacKenzie et al. 2022)

- ~4,000 speakers from the 5 northern cities of interest:

Leeds	Liverpool	Manchester	Newcastle	Sheffield
LS	**L**	**M**	**NE**	**S**
(N = 470)	(N = 441)	(N = 1065)	(N = 1352)	(N = 616)

- 'Younger' group (N=2499): born 1981–2010, mean = **1995**
- 'Older' group (N=1445): born 1924–1980, mean = **1961**



Data: Mobility

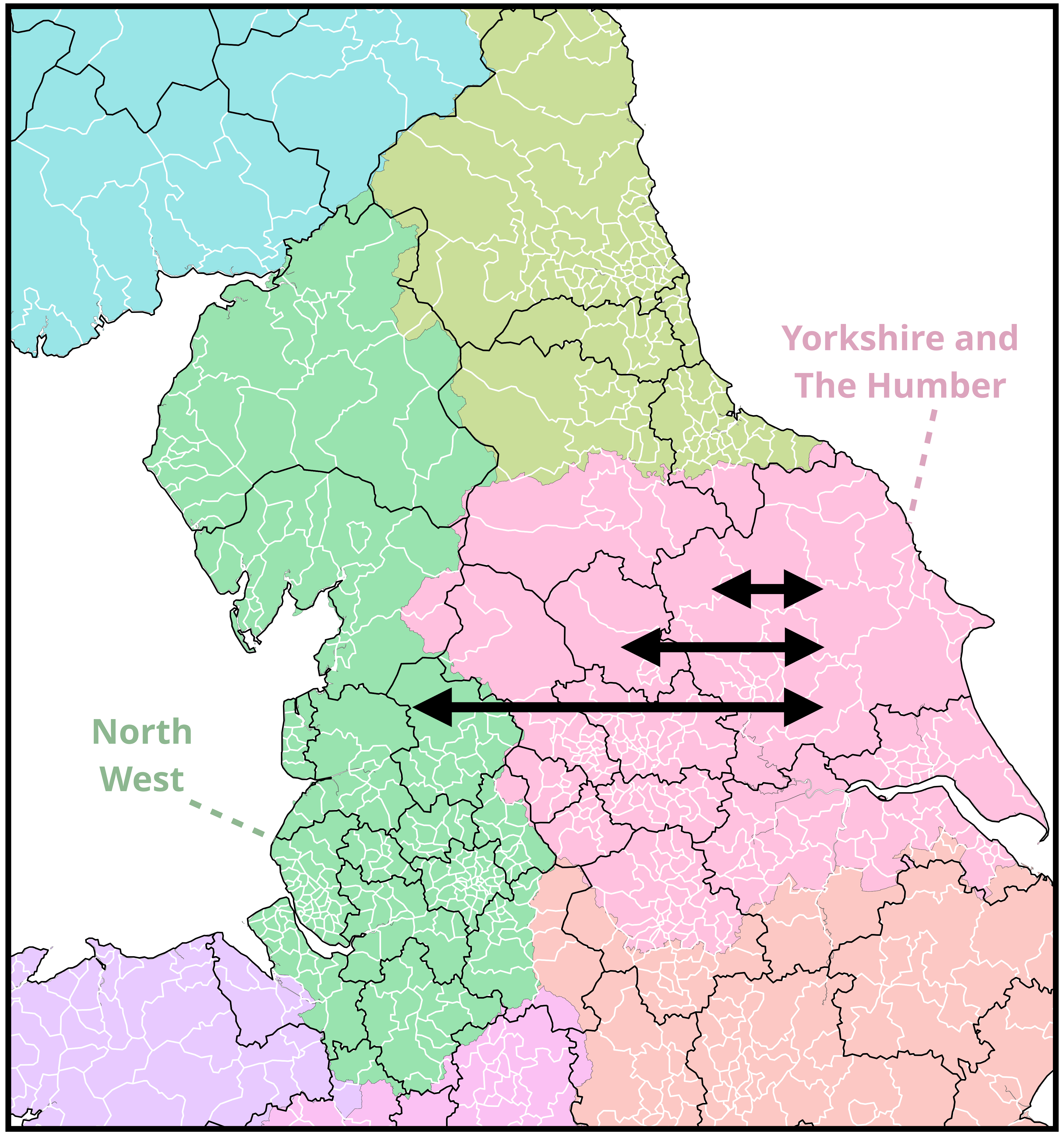
- Respondents were also asked for a full list of *everywhere* they lived during childhood and early adolescence

- Most were non-mobile (93.4%), but there are enough responses from mobile individuals to consider this as a factor in the analysis:

  - 78 moved **between postcode districts** *(within the same postcode area)*

  - 49 moved **between postcode areas** *(within the same region)*

  - 96 moved **between regions** *(within England)*



Data: Survey questions

The survey includes 35 questions covering three *types* of dialect features:

lexical

e.g. what word do you use to refer to the ***evening meal***?

morphosyntactic

e.g. could you use the phrase ‘***we was watching a film***’?

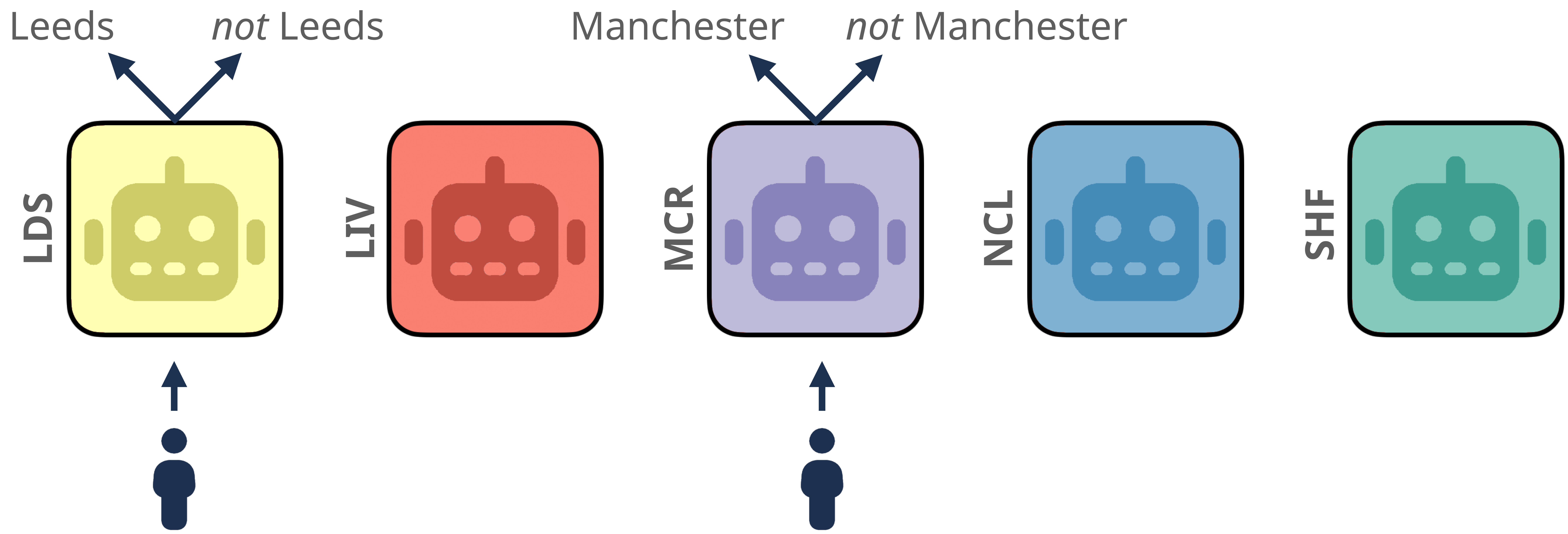
phonological

e.g. do the words ***book*** and ***spook*** rhyme for you?

e.g. do the words ***thin*** and ***fin*** sound the same or different to you?

Analysis

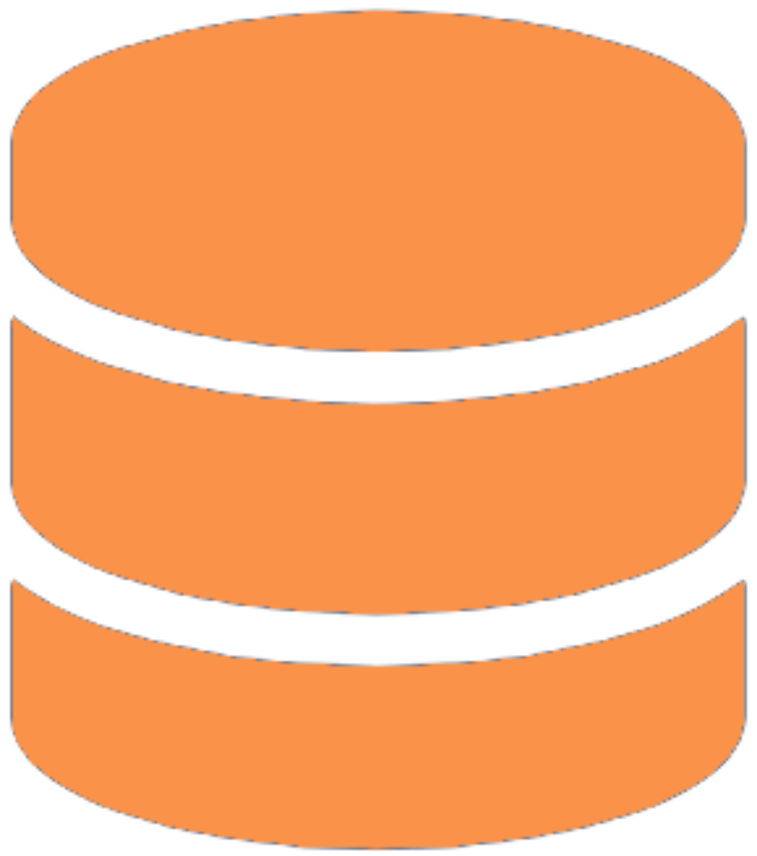
- Turn the dependent variable into a binary (location vs *not* location) and fit separate random forest models to predict membership of each location



Analysis

training data

used to train
classification models

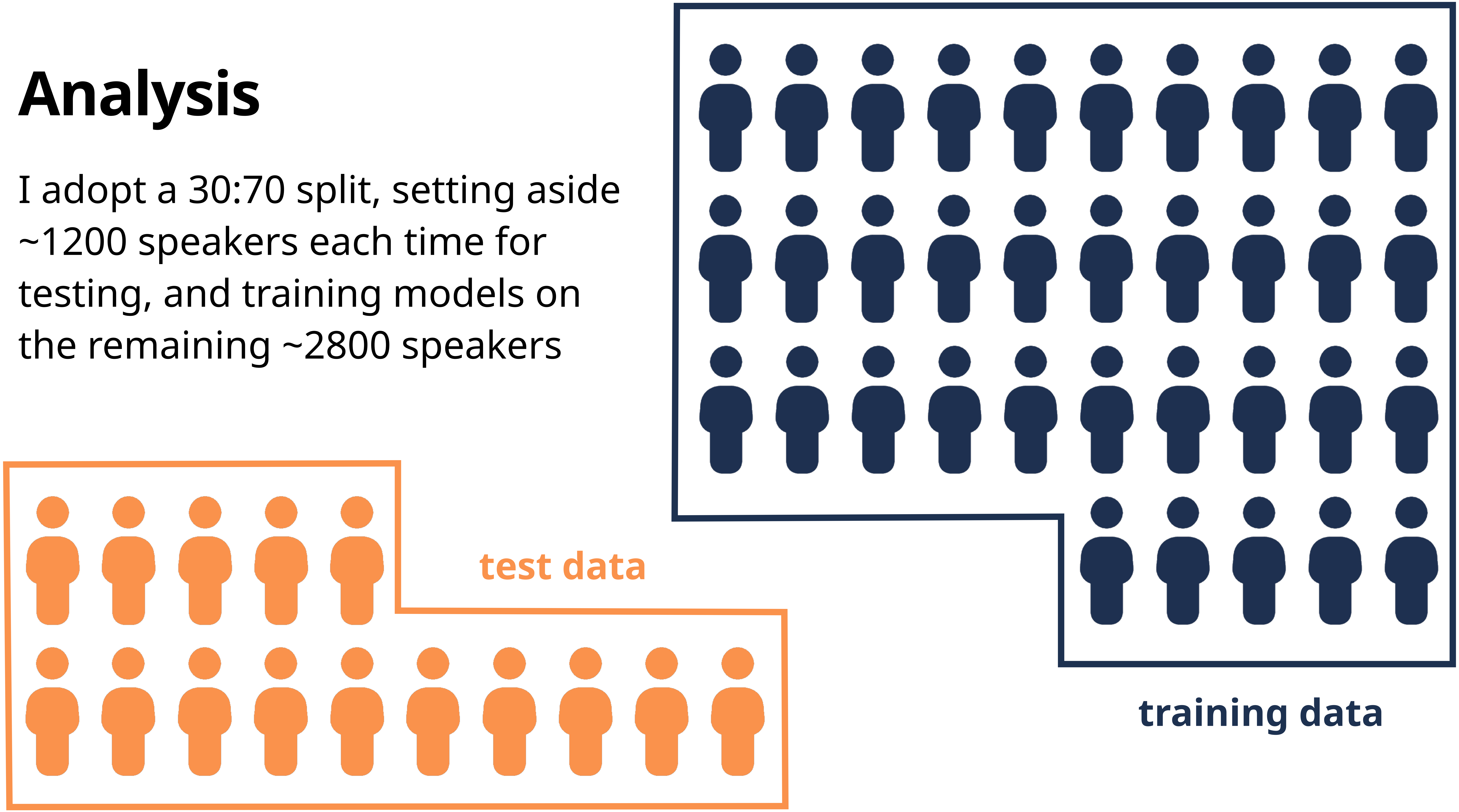


testing data

those models are then
tested on unseen data

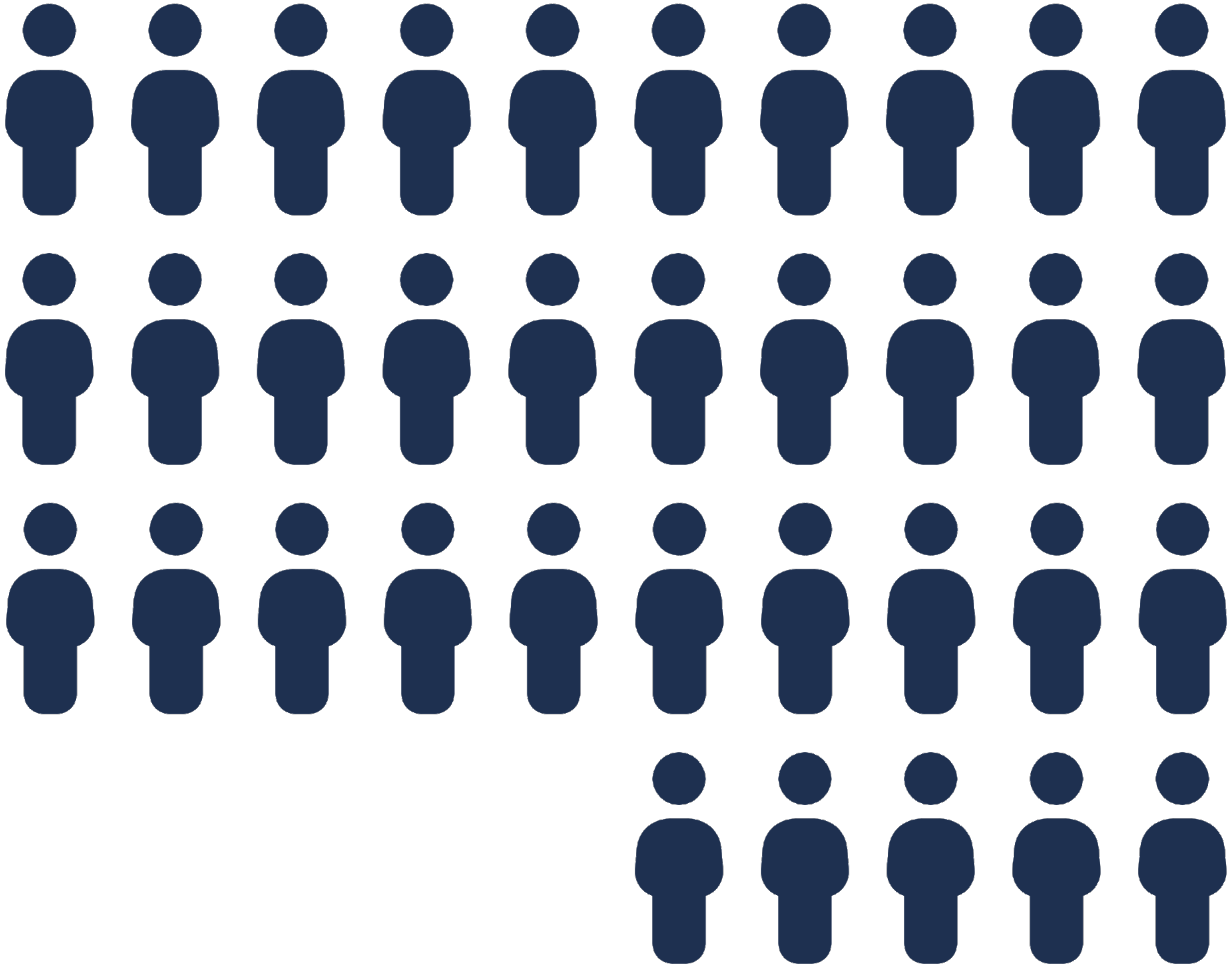
Analysis

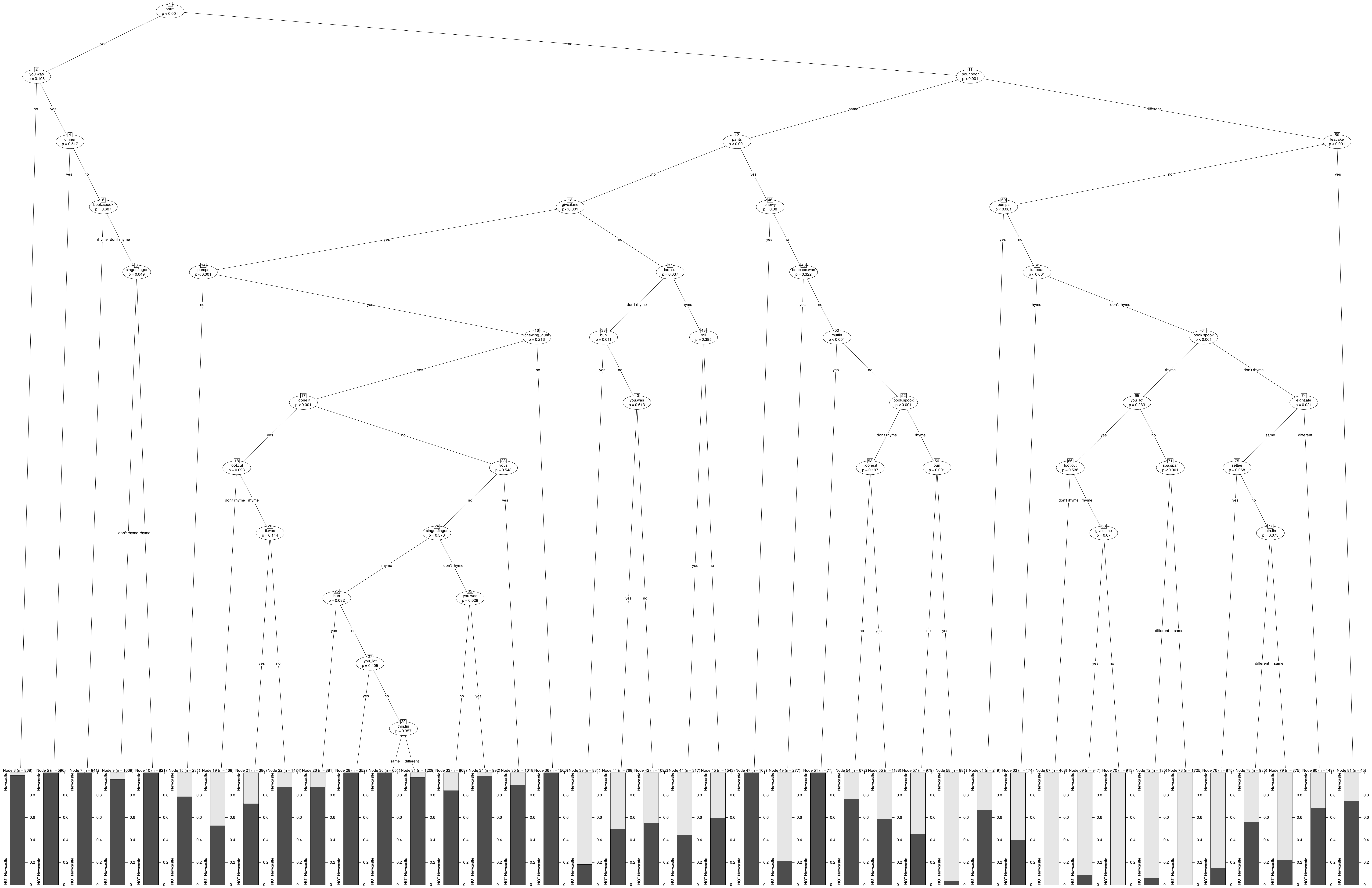
I adopt a 30:70 split, setting aside ~1200 speakers each time for testing, and training models on the remaining ~2800 speakers



Analysis

- Fit a random forest that learns from this training data
- Each forest contains 500 classification *trees*



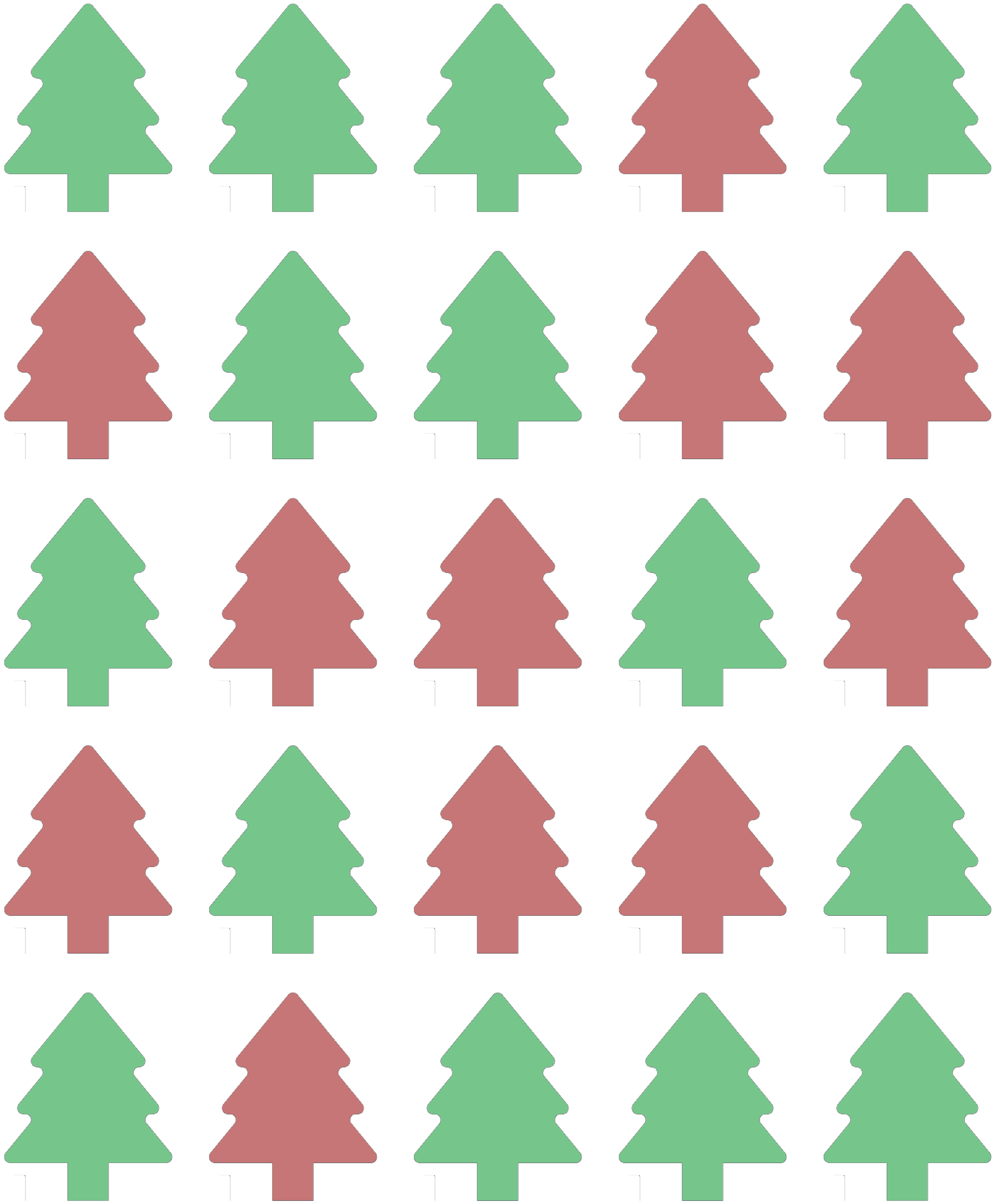

1 barm p < 0.001
2 you.was p = 0.108
4 dinner p = 0.517
6 book.spook p = 0.607
8 singer.finger p = 0.049
11 pour.poor p < 0.001
12 pants p < 0.001
13 give.it.me p < 0.001
14 pumps p < 0.001
16 chewing_gum p = 0.213
17 I.done.it p < 0.001
18 foot.cut p = 0.093
20 it.was p = 0.144
23 yous p = 0.543
24 singer.finger p = 0.573
25 bun p = 0.082
27 you_lot p = 0.405
29 thin.fin p = 0.357
32 you.was p = 0.029
37 foot.cut p = 0.037
38 bun p = 0.011
40 you.was p = 0.613
43 roll p = 0.385
46 chewy p = 0.08
48 beaches.was p = 0.322
50 muffin p < 0.001
52 book.spook p < 0.001
53 I.done.it p = 0.197
56 bun p = 0.001
59 teacake p < 0.001
60 pumps p < 0.001
62 fur.bear p < 0.001
64 book.spook p < 0.001
65 you_lot p = 0.233
66 foot.cut p = 0.536
68 give.it.me p = 0.07
71 spa.spar p < 0.001
74 eight.ate p = 0.021
75 settee p = 0.068
77 thin.fin p = 0.075
Newcastle
NOT Newcastle

Analysis

- Each tree in a random forest generates a prediction
- The random forest settles upon one single outcome based on the majority ‘vote’
- Here: I also analyse tree ‘agreement’ as a gradient measure of confidence

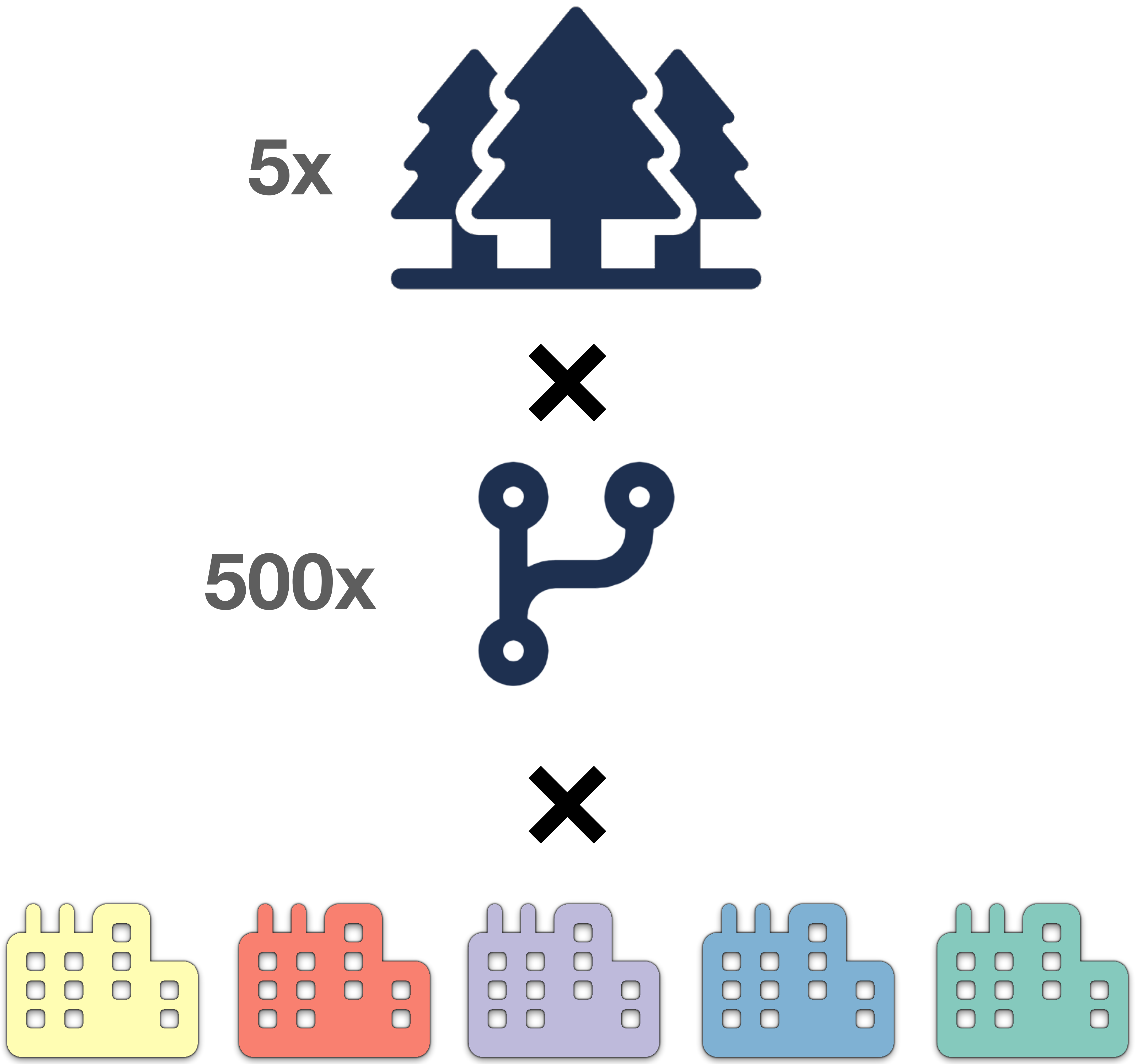
yes, from Liverpool

(56% agreement)



Analysis

- This entire process is repeated with:
  - different random samples of predictors (dialect features)
  - different random samples of the speaker population for the training/testing allocation
- This is called **bootstrap aggregation** (bagging)



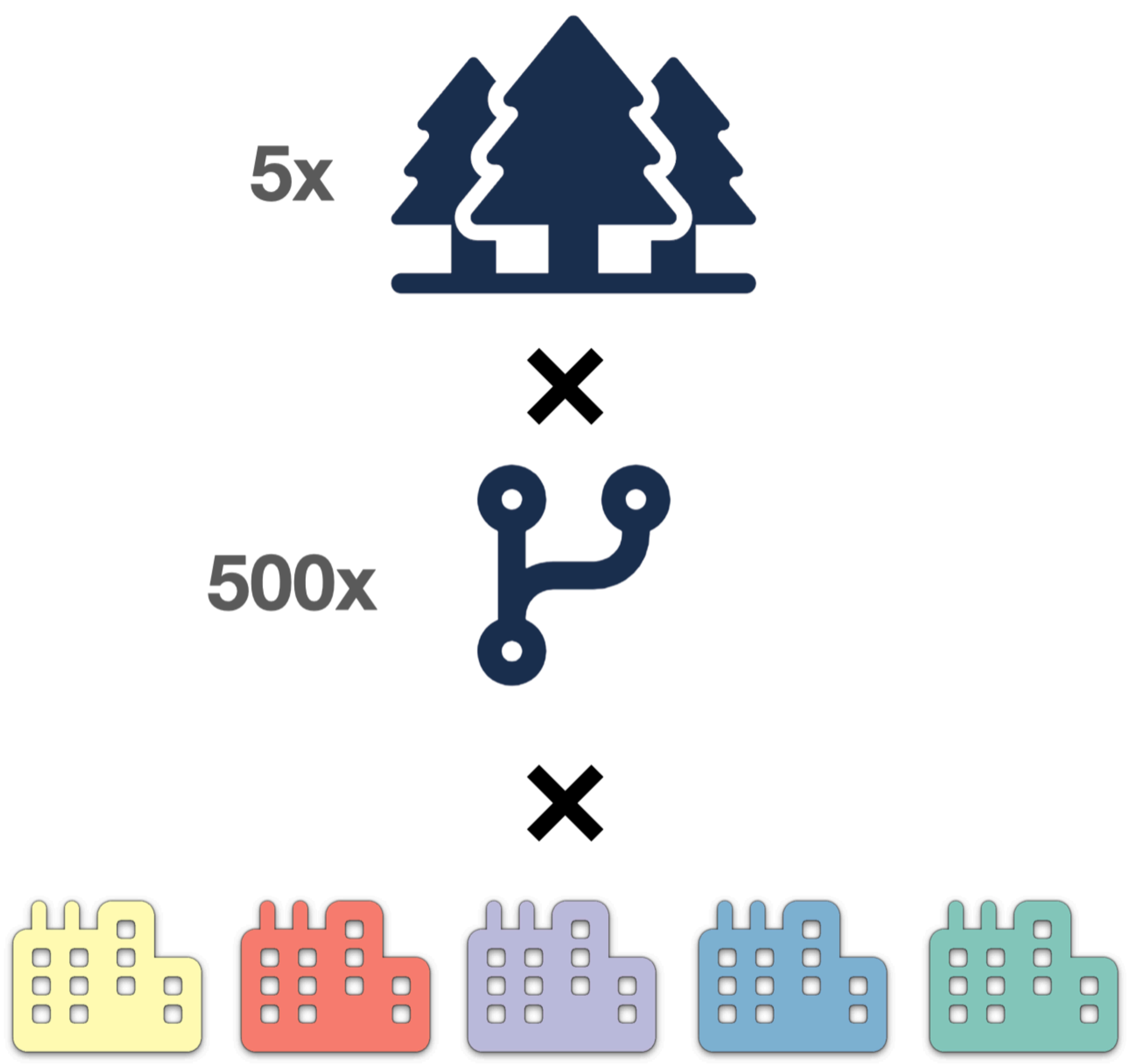
2500 classification trees per city

Analysis

- Then *that entire process* is repeated, but on:
  - only younger speakers
  - only older speakers

Analysis

- This entire process is repeated with:
  - different random samples of predictors (dialect features)
  - different random samples of the speaker population for the training/testing allocation
- This is called **bootstrap aggregation** (bagging)

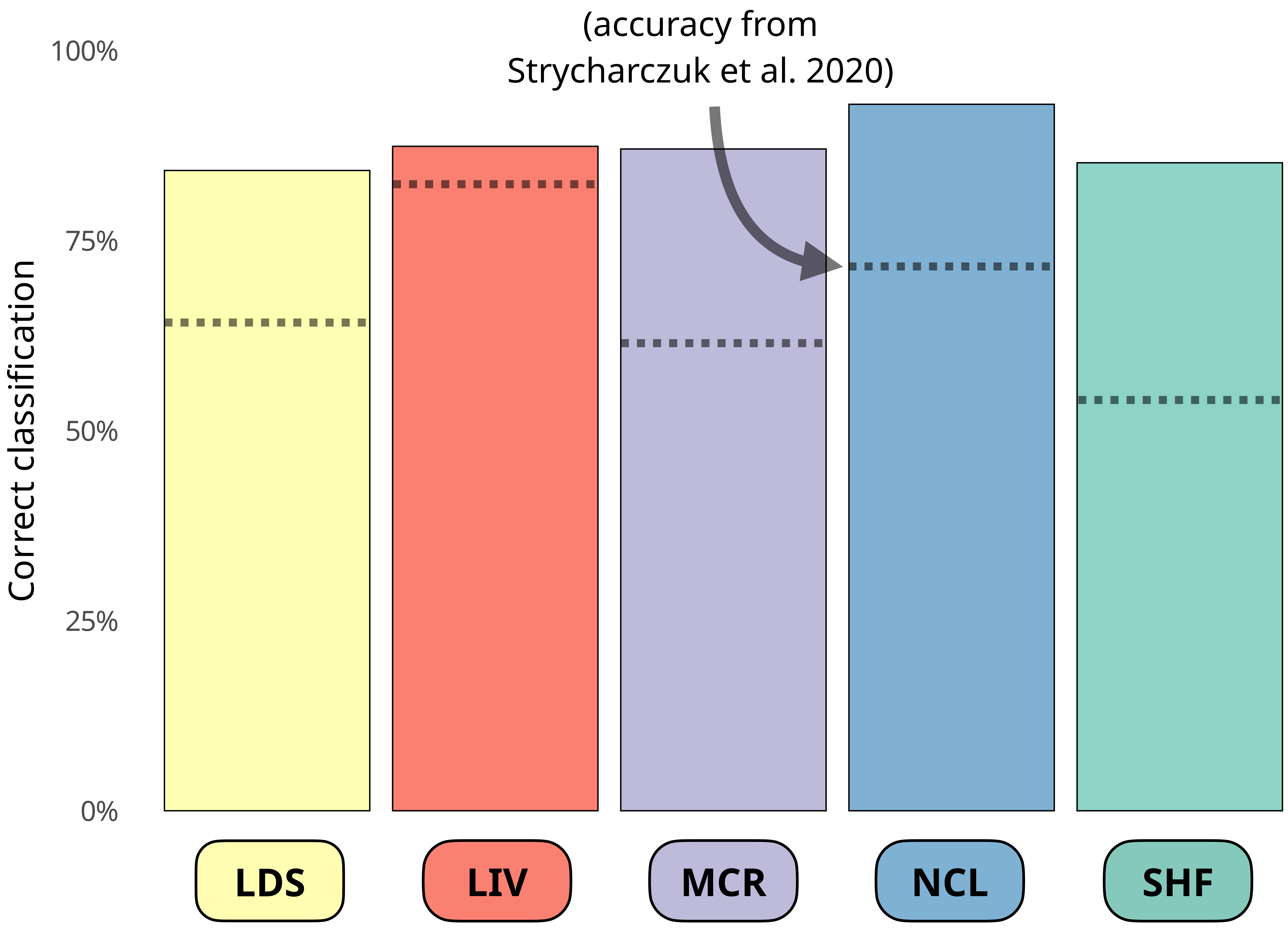


2500 classification trees per city

- Resulting in **3 sets of 5 random forests**:
  - the ‘overall’ set (for a general analysis like Strycharczuk et al)
  - the ‘young’ + ‘old’ sets (to investigate apparent-time change)

Results

Accuracy by city



Newcastle model is most accurate (93%), followed by Liverpool and Manchester (87%)

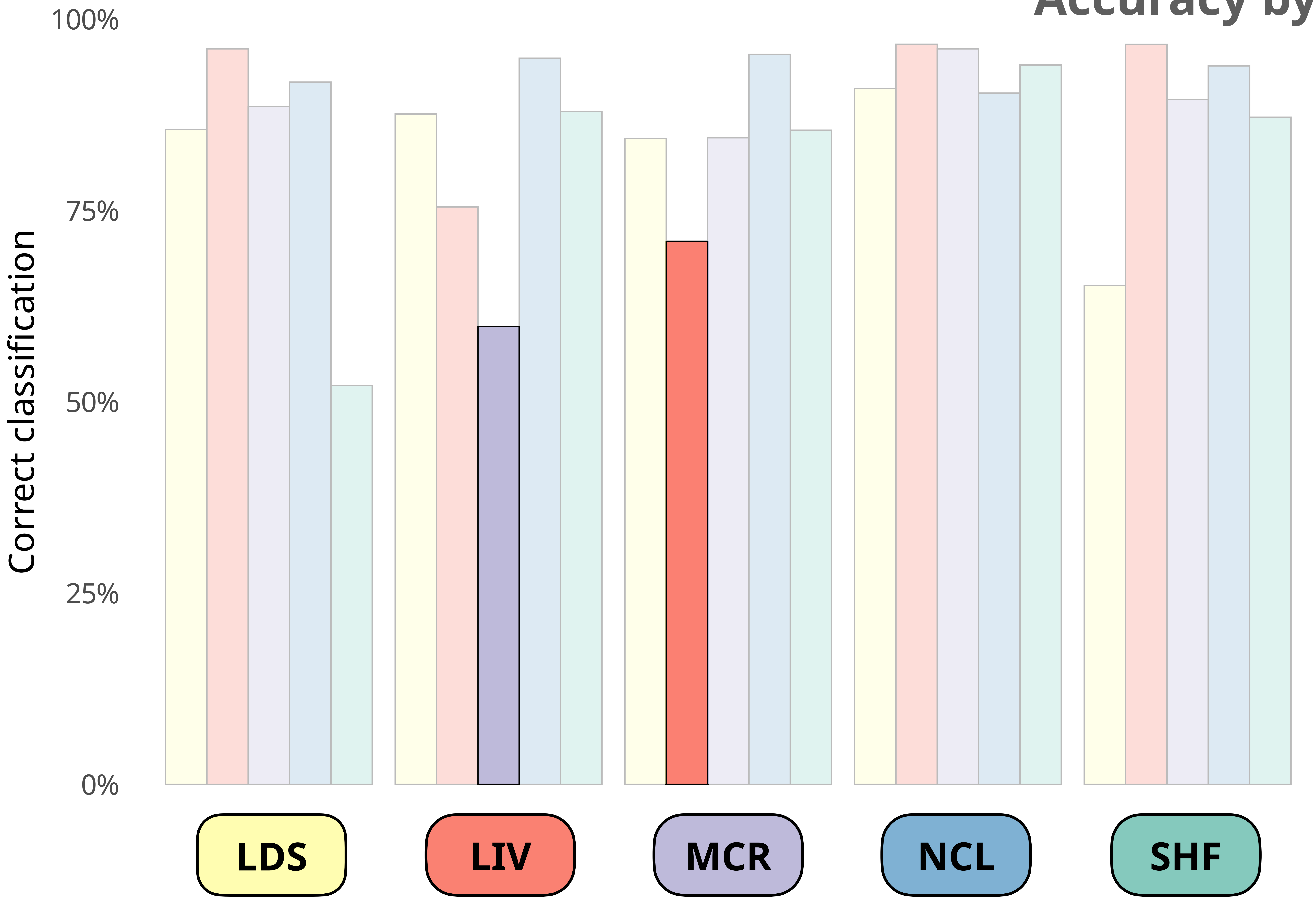
Sheffield (85%) and Leeds (84%) are least accurate

But very high rates across the board!

Accuracy by speaker dialect
(split by model)
100%
75%
50%
25%
0%
Correct classification
LDS
LIV
MCR
NCL
SHF
Most errors:
Sheffield speakers incorrectly classified as being from Leeds
Leeds speakers incorrectly classified as being from Sheffield

Accuracy by speaker dialect

(split by model)



Most errors:

The same also applies to **Manchester** and **Liverpool**

i.e. Liverpudlians getting mistaken as Mancunian and (to a lesser extent) vice versa

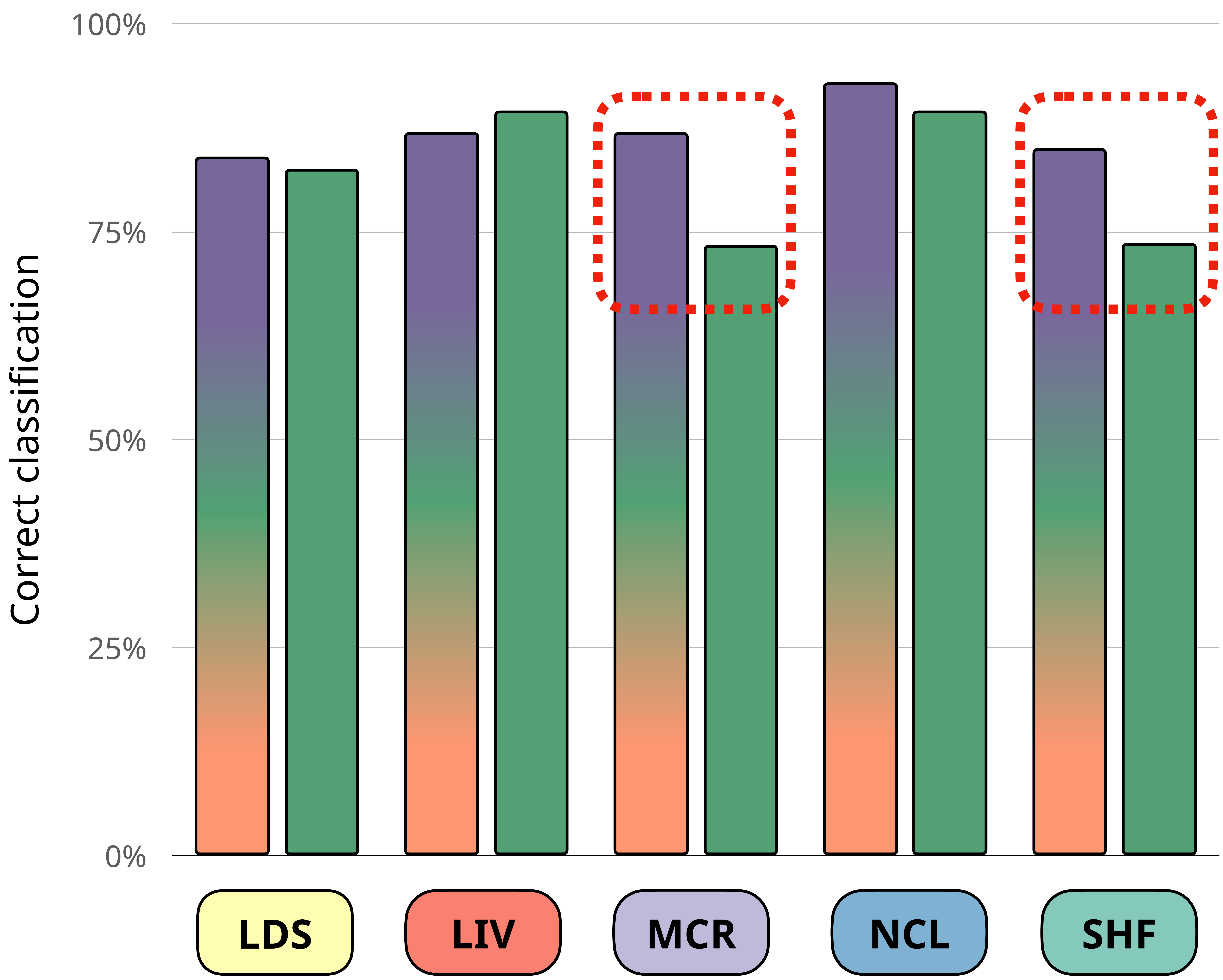
Branching out #1

Phonological-only forests

These random forests were trained on a combination of **lexical**, **phonological**, and **grammatical** dialect features

What if we train models *only* on **phonological** features (more closely mirroring the models of Strycharczuk et al. 2020)?

Combined models vs phonological models



Overall, accuracy rates don't change that much

Biggest drops are for **Manchester** (87% → 73%) and **Sheffield** (85% → 74%)

Accuracy by city/speaker

phon. features only

Branching out #2

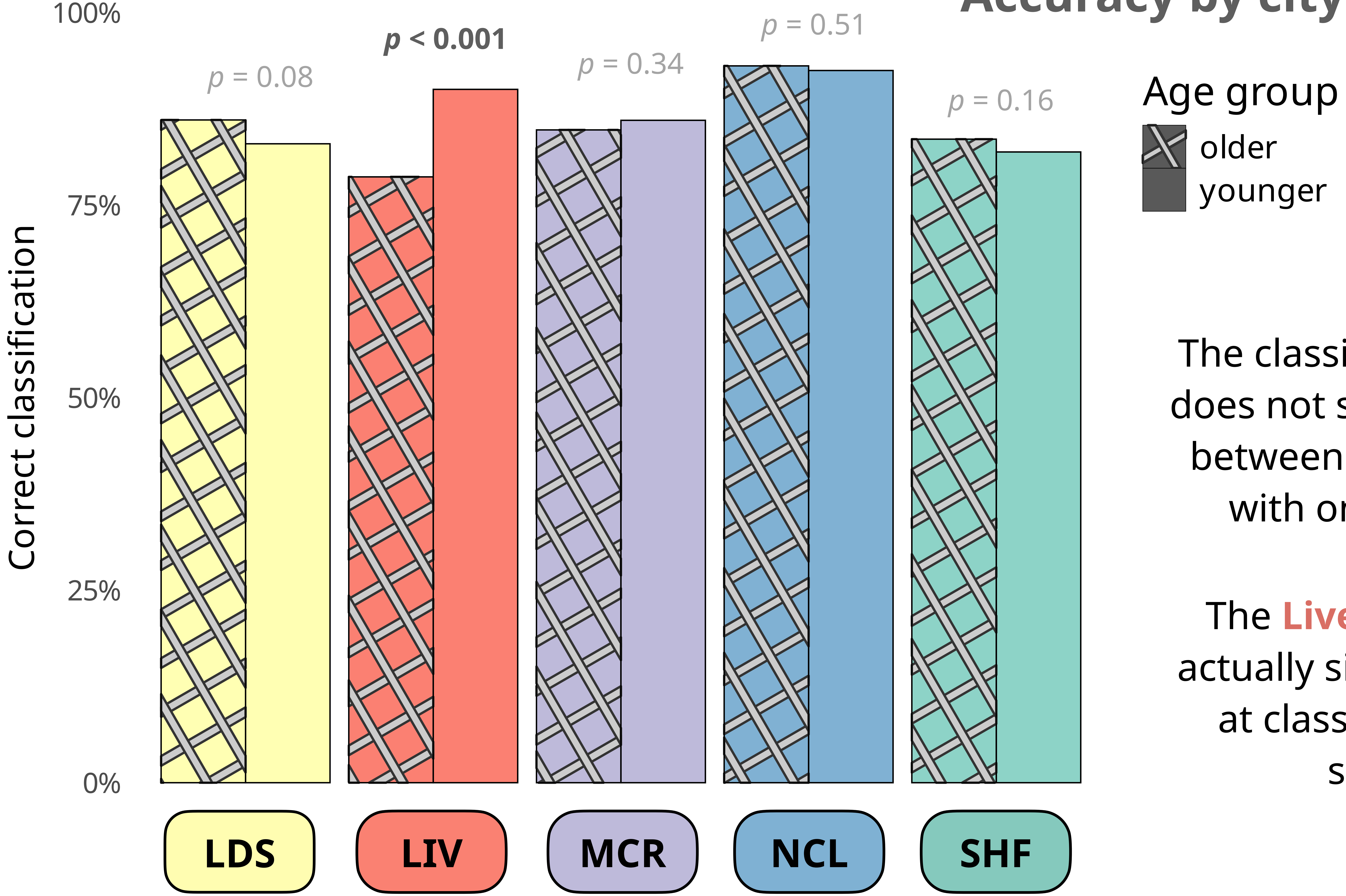
Apparent-time analysis

These models have all been randomly sampling from the *whole* population of respondents

What happens if we train (and test) models specifically on *younger* vs *older* speakers?

Hypothesis:
Younger speakers are more difficult to classify, due to levelling

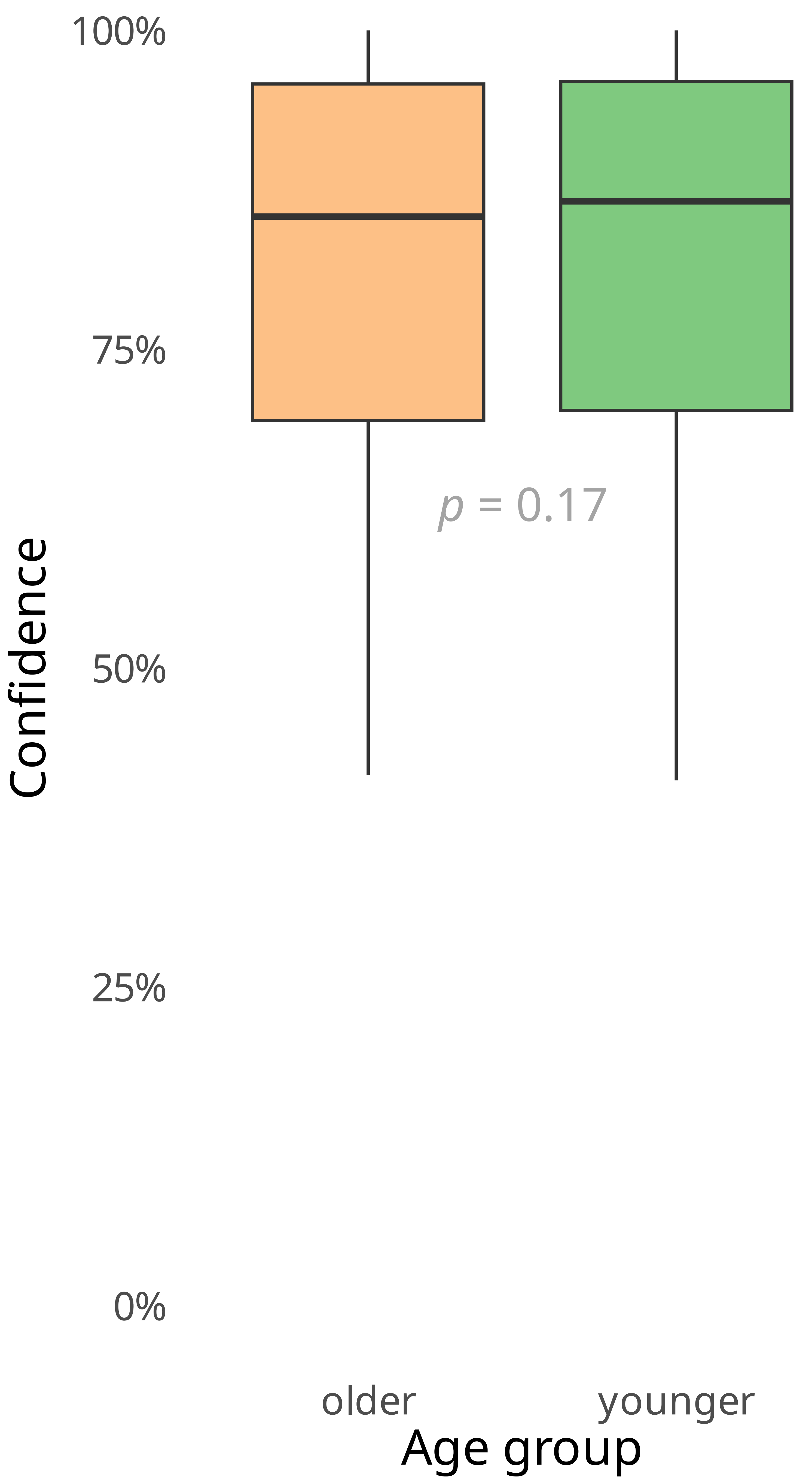
Accuracy by city and age group



The classification accuracy does not significantly differ between the age groups, with one exception...

The **Liverpool** model is actually significantly *better* at classifying younger speakers!

Confidence by age group



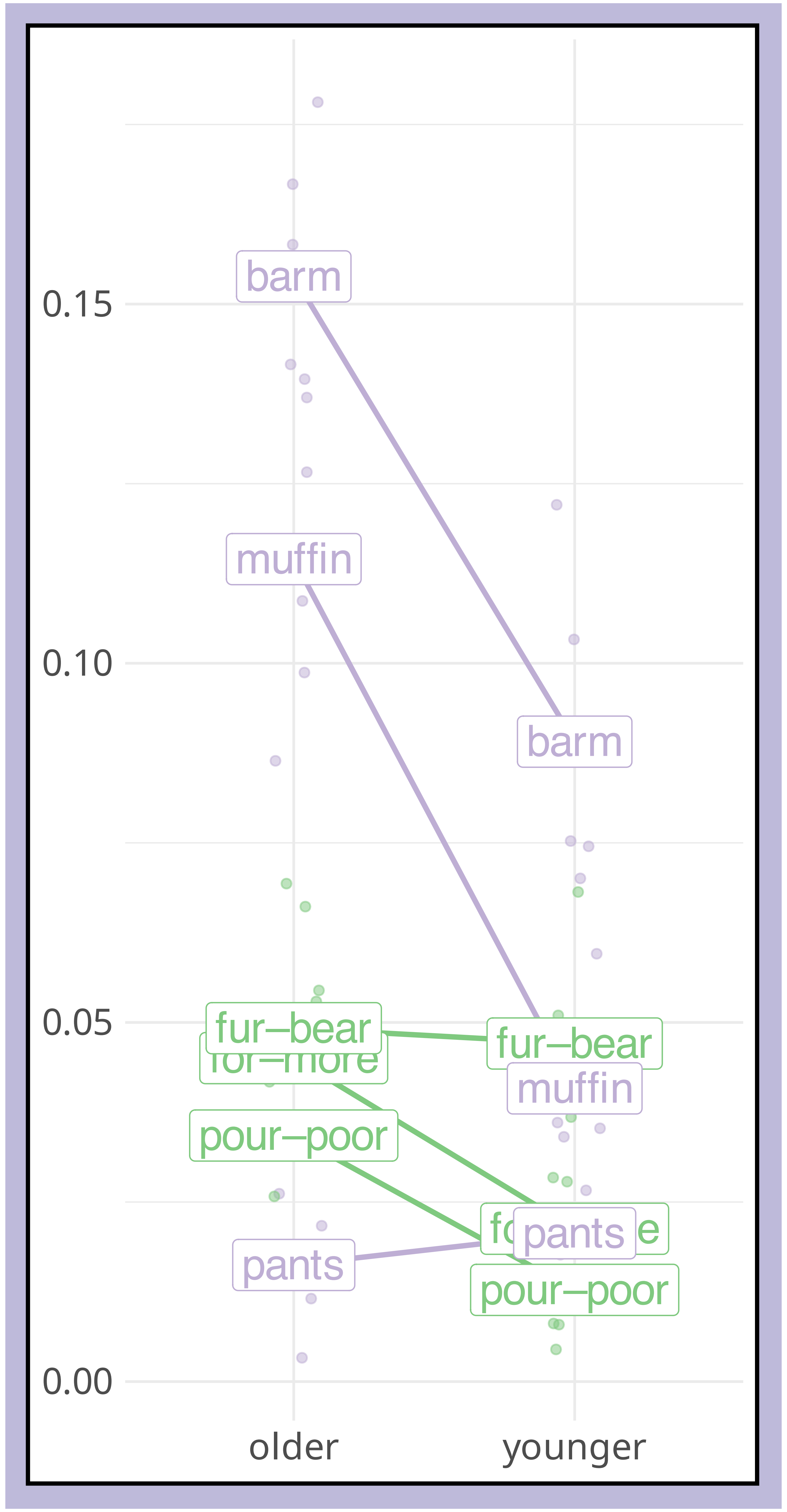
The models may be (somewhat) equally accurate in their overall classifications...

...but is there lower consensus (i.e. fewer correct classifications from the individual trees of a forest) for younger speakers?

Also no.

Conditional variable importance

- *Conditional variable importance* (Strobl et al. 2008) measures the **relative influence** of each dialect feature in a random forest
    - i.e. how useful the presence (or absence) of a particular feature is in classifying a speaker as being from that location (or not)
- **Do *these* show any differences in apparent time?**



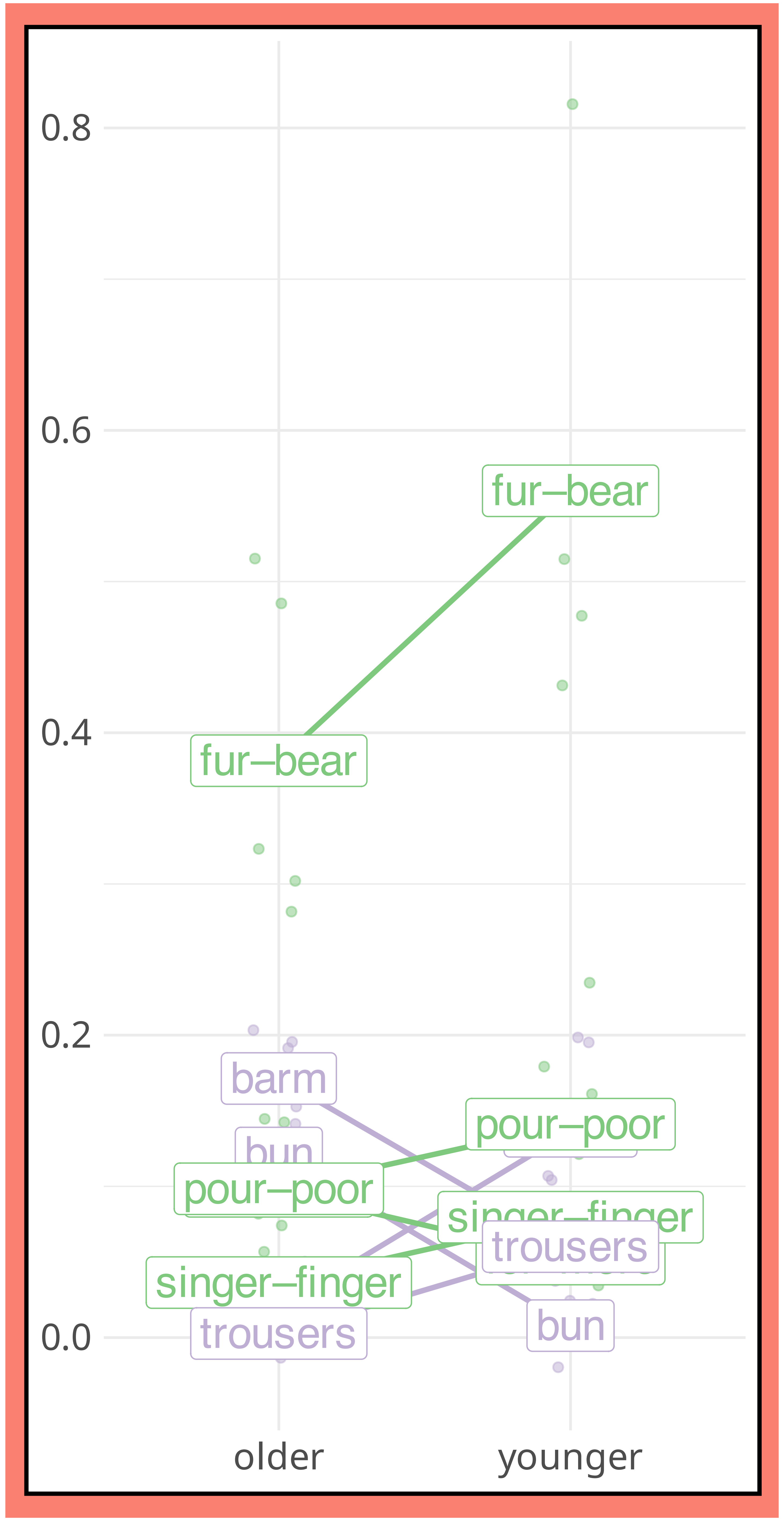
MCR

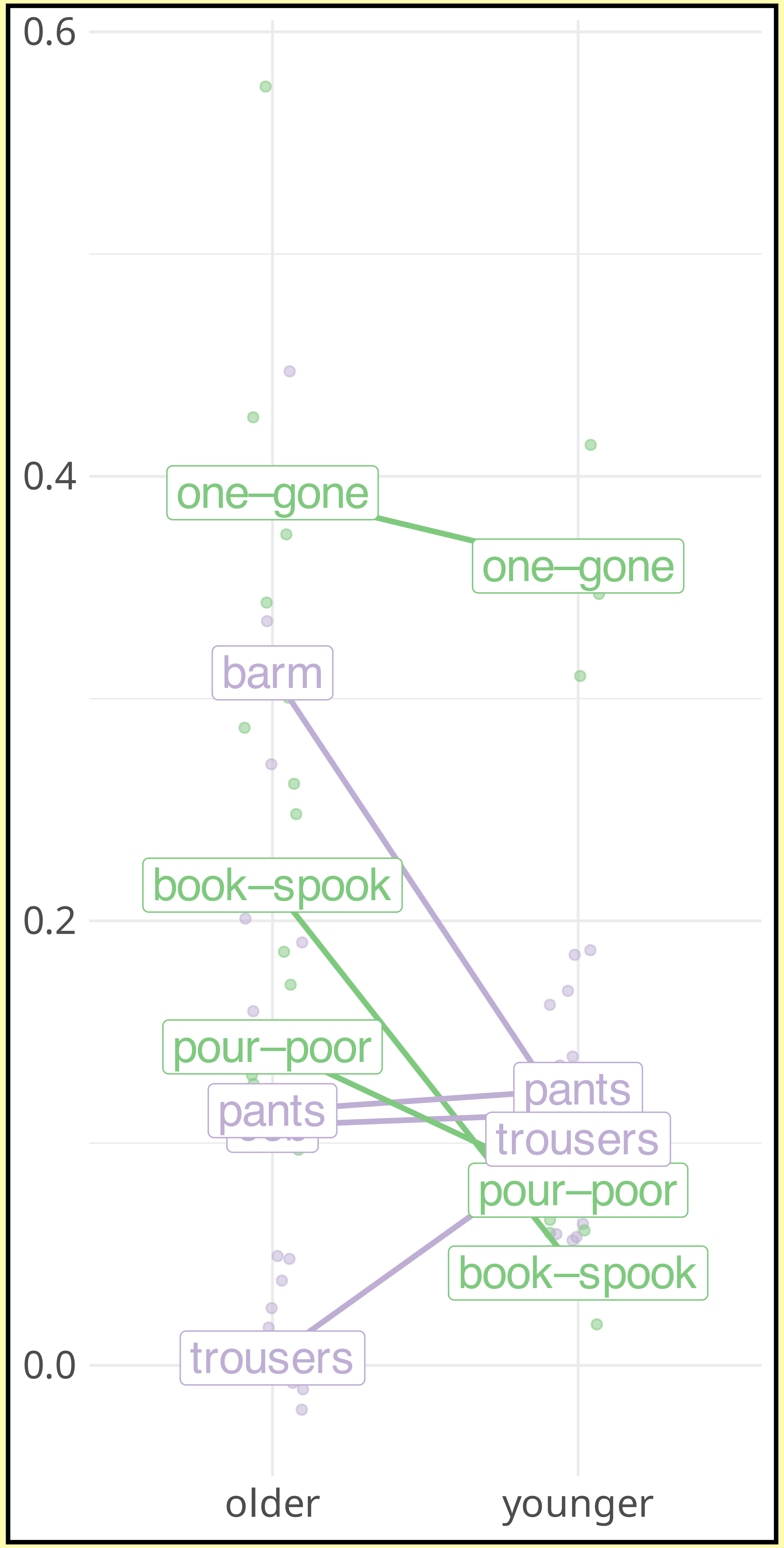
lexical items like *barm* and *muffin* have much lower importance

NORTH–FORCE distinction and FORCE–CURE merger have also weakened

NURSE–SQUARE merger has become increasingly important in classifying Liverpool English

LIV





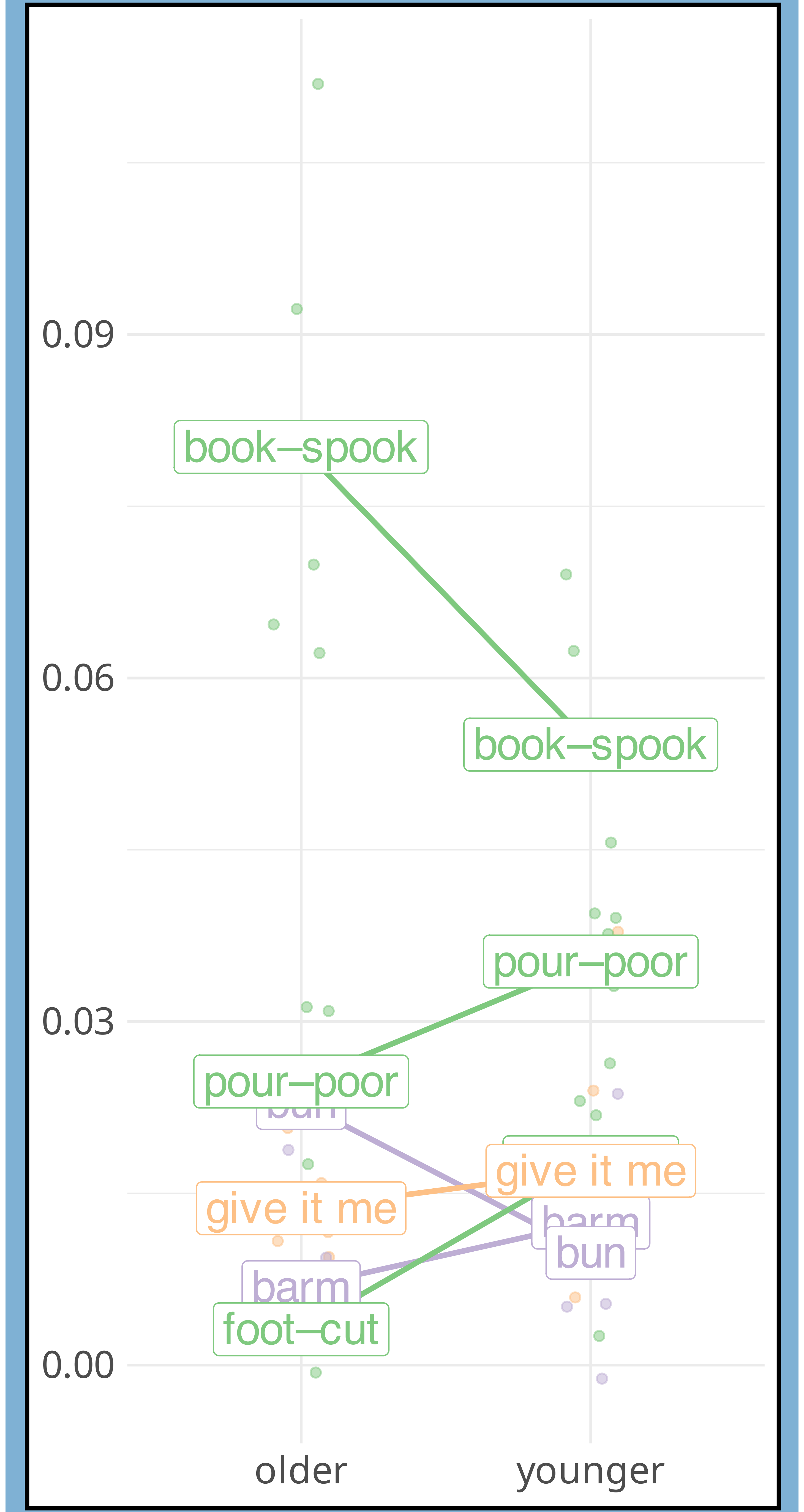
LDS

one–gone distinction is fairly stable → now by far the most important feature for classifying young Leeds speakers

long [uː] in *book* is less useful for classifying young speakers (levelling!)

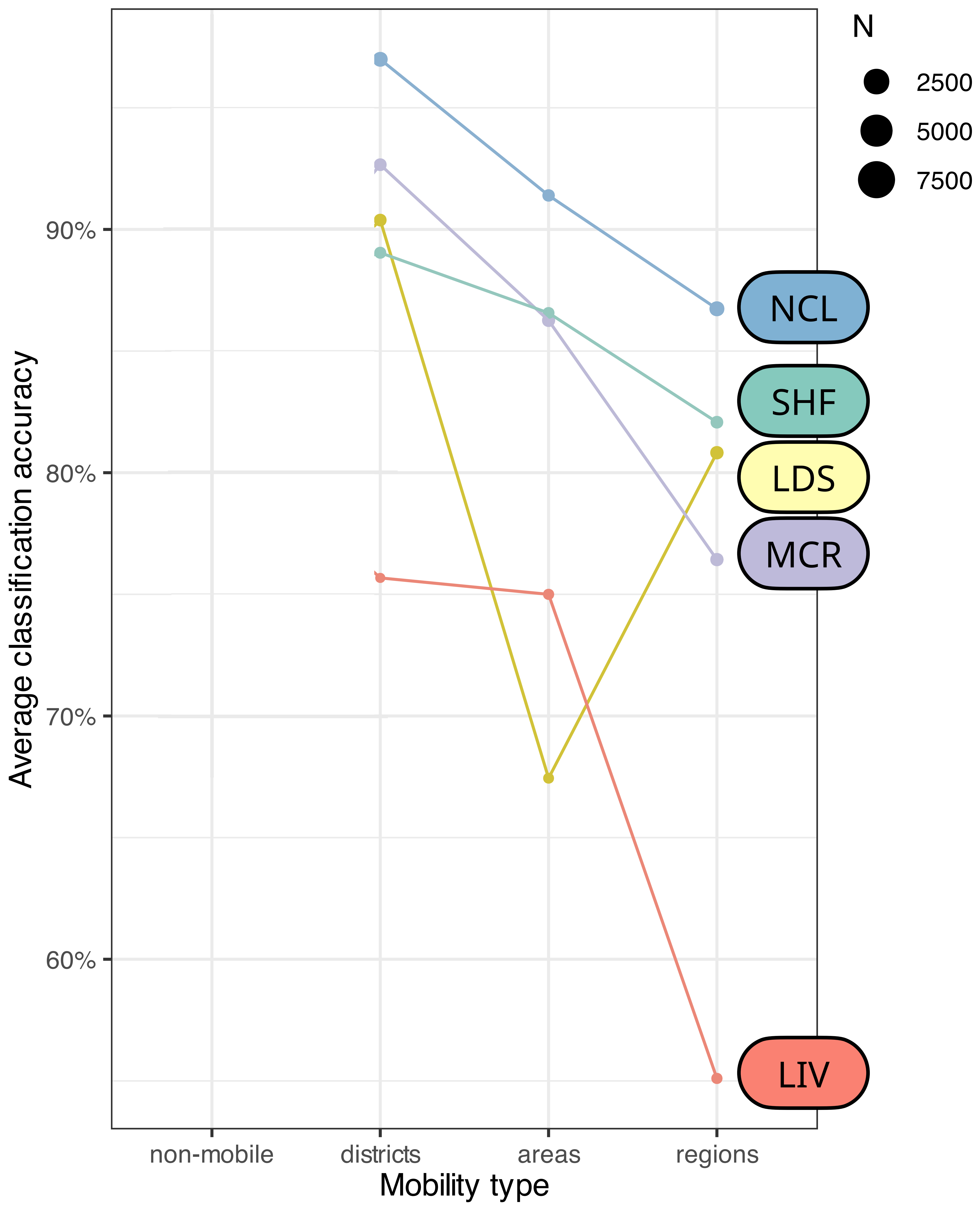
FORCE–CURE distinction has become increasingly important at its expense

NEW



Childhood mobility

- General tendency for classification accuracy to **decrease** as extent of mobility **increases**
- Biggest decrease is for **Liverpool** speakers when mobility is between *regions*
- Surprisingly, non-mobile speakers generally harder to classify than those who moved *within* the limits of a postcode area



Discussion

Discussion

Overall results

- Overall, classification accuracy is much higher than that reported by Strycharczuk et al. (2020)
    - clean, binary self-report data vs messy acoustic formant data?
    - considering different *dimensions* of dialectal variability (i.e. lexical, morphosyntactic and consonantal features, not just vowels)?
    - speakers shifting away from their regional accents due to formality of read passage in the data they use?
- Despite higher overall accuracy, the results are similar in terms of the **hierarchy of dialects** and the specific **confusability patterns**

Discussion

Dialect levelling

- Strycharczuk et al. (2020) conclude that the lower classification success for certain dialects suggests **levelling** has taken place
- But this presupposes that the random forest models would, at some earlier point in time, have had *higher* classification accuracy
- This isn't supported by the apparent-time analysis here:
    - no consistent increase in accuracy for models trained (and tested) exclusively on older speakers

Discussion

Does this mean dialect levelling *hasn't* taken place? no!

- **Possible explanations:**
  - Looking at too narrow a time window: the results here don't mean that levelling *didn't* take place, but rather that it likely **slowed down around the 1950s/60s onwards**
  - Survey data: great for tracing systematic phonological changes (i.e. mergers and splits), not so great for levelling that manifests in **smaller-scale, gradient phonetic shifts**
- Variable importance scores indicate that some features *are* becoming less useful in dialect classification (i.e. because of levelling), but **not to the point where speakers are becoming indistinguishable**
- Geographically mobile speakers *are* more difficult to classify, and **mobility→levelling**


"It's grim up north"
Thanks! Any questions?