

# Diachrony and dialect variation in British English prepositional *wh*-questions: Evidence from social-media data

David Willis  
University of Oxford

REEDS 2023, Amsterdam  
29 June 2023



# The Tweetolectology project



- ESRC project 'Investigating the diffusion of morphosyntactic innovations using social media', 1 September 2017–31 July 2021
- David Willis (PI, Oxford), Adrian Leemann (CI, Bern), Tamsin Blaxter (CI, Oxford), Deepthi Gopal (RA, Uppsala), Jonathan Mackenzie (IT)
- investigates geospatial patterns of grammatical variation in British/Irish English, Welsh and Norwegian/mainland Scandinavian

# Using Twitter for language variation/change

- Large-scale social-media data is increasingly of interest to linguists (Eisenstein et al. 2010, 2014; Russ 2012; Doyle 2012, 2014; Bamman et al. 2014; Gonçalves & Sánchez 2014; Jones 2015; Huang et al. 2016; Grieve et al. 2016, 2018).
- Much existing work shares the property of dealing with frequent (often lexical) patterns appearing in very large languages (American English, World Spanish), and the granularity of geographic information is often coarse.
- Can this work be integrated into theoretical and more traditional variationist work in syntax?
- large speech-like corpora avoid some of the problems of elicitation in dialect(-syntax) surveys and questionnaires

# Using Twitter for language variation/change

- British and Irish English Twitter corpus: selection of tweets from two sources:
  - seed corpus geotagged within a bounding box around Britain and Ireland during a period in 2017–18
  - users mentioned by accounts gathered this way
- initial corpus of 546.4m tweets, 6.1m unique user accounts
- of these, 2.0m accounts were successfully assigned a location within Britain, Ireland, Isle of Man or the Channel Islands using keyword localization used to map users to best guess based on: user location field, bio, and tweet text searched for Ordnance Survey place names, informal names, and names of regions, weighted by number of occurrences, context and distance between places mentioned

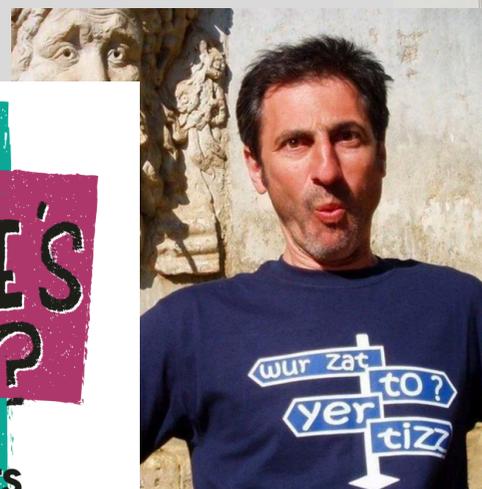
# Using Twitter for language variation/change



- 294k users were classified for age at different levels of granularity: precise, fine (decades) or broad (<30, 30–60, >60)
- interactive atlas will be published online

# Where ... to ...?

- *Where's that to?* (*Wur zat to?*) (split)
- *Where to's that?* (contiguous)
- very little existing research, but known to be a feature of Bristol (where it is a stereotype) and Cardiff (where there is awareness, but it is less central to identity)
- search is for *where is/ 's*; comparator is a set of other nonstandard forms: *Where's that at?* *Where's that about?*



## Where ... to ...?

He walked into one of the bedrooms, and there found the prisoner standing at a dressing table searching a drawer. Witness said “Hollon!” and prisoner turned round and said “Where’s the governor to?” Witness replied “Just outside,” and then prisoner rushed past him and ran up the road. (British Library Newspapers, Temple Cloud Police Court: Tuesday, Bristol Mercury, 28 May 1898, p. 7) [Temple Cloud = village 16km south of Bristol]

Interviewee: and where'd you live down town

Interviewer: oh I live out nearer Weston super Mare now

Interviewee: oh down further

Interviewer : yeah yeah

Interviewee: yeah where to

Interviewer: er Winscombe

Interviewee: yeah we used to go there



# Where ... to ...?

S0493: fingers in lots of different financial pies (.) what 's she doing ?

S0668: er admin things I think

S0493: *where to is that ?*

S0668: in Cardiff (.) --

ANONnameM's doing his erm chartered exams at the moment for (.) to become a chartered accountant (BNC2014, 31yo teacher from Cardiff)

- found also in Newfoundland English (Clarke 2010):

*Where did you see the woman to?* (Irish Avalon (Ferryland)) (Clarke 2010: 95)

- although canonical examples are typically with present-tense *be*, appears to be found with other verbs and tenses

# Where ... to ...?

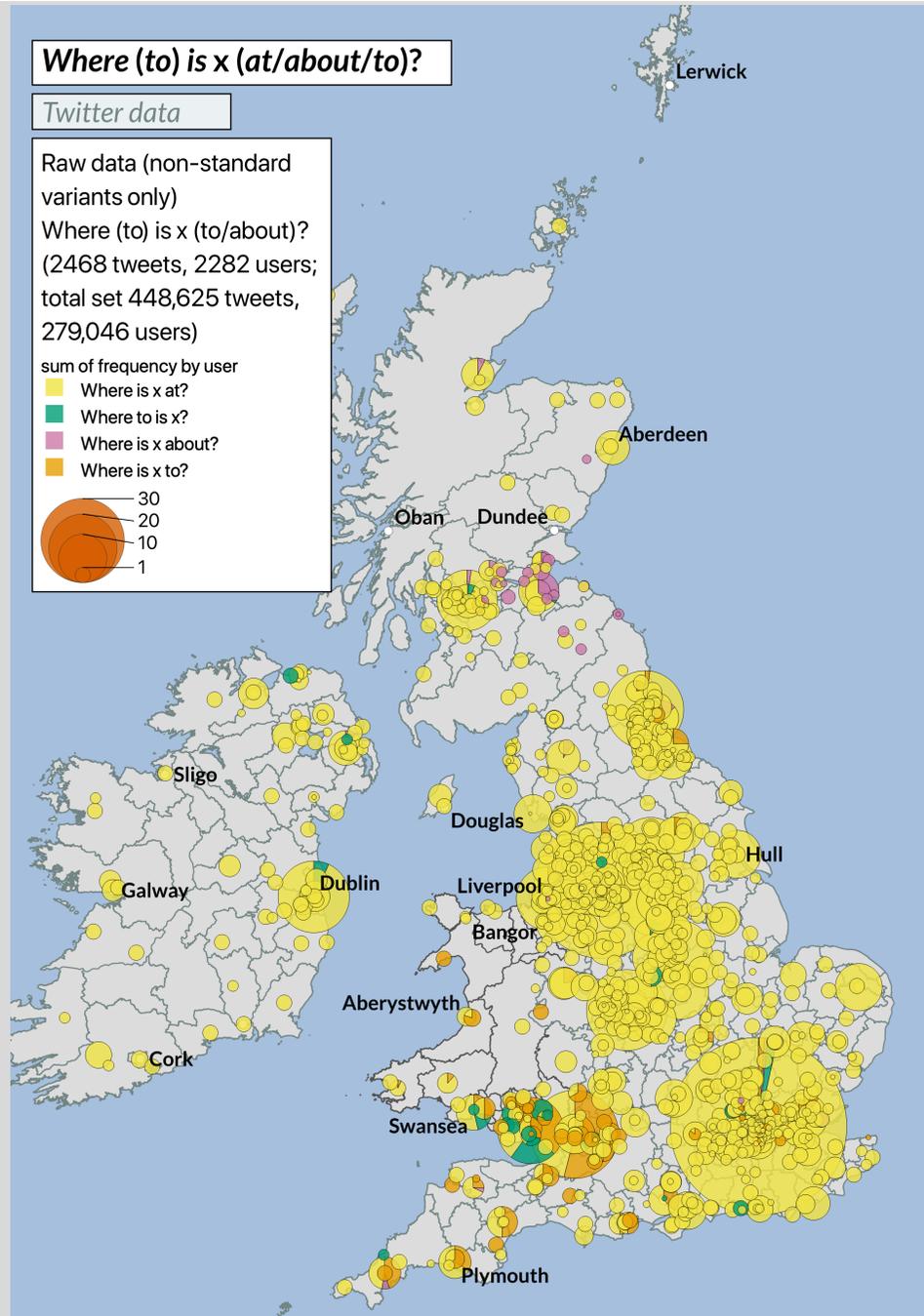
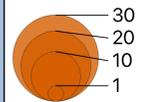
## Where (to) is x (at/about/to)?

### Twitter data

Raw data (non-standard variants only)  
Where (to) is x (to/about)?  
(2468 tweets, 2282 users;  
total set 448,625 tweets,  
279,046 users)

sum of frequency by user

- Where is x at?
- Where to is x?
- Where is x about?
- Where is x to?



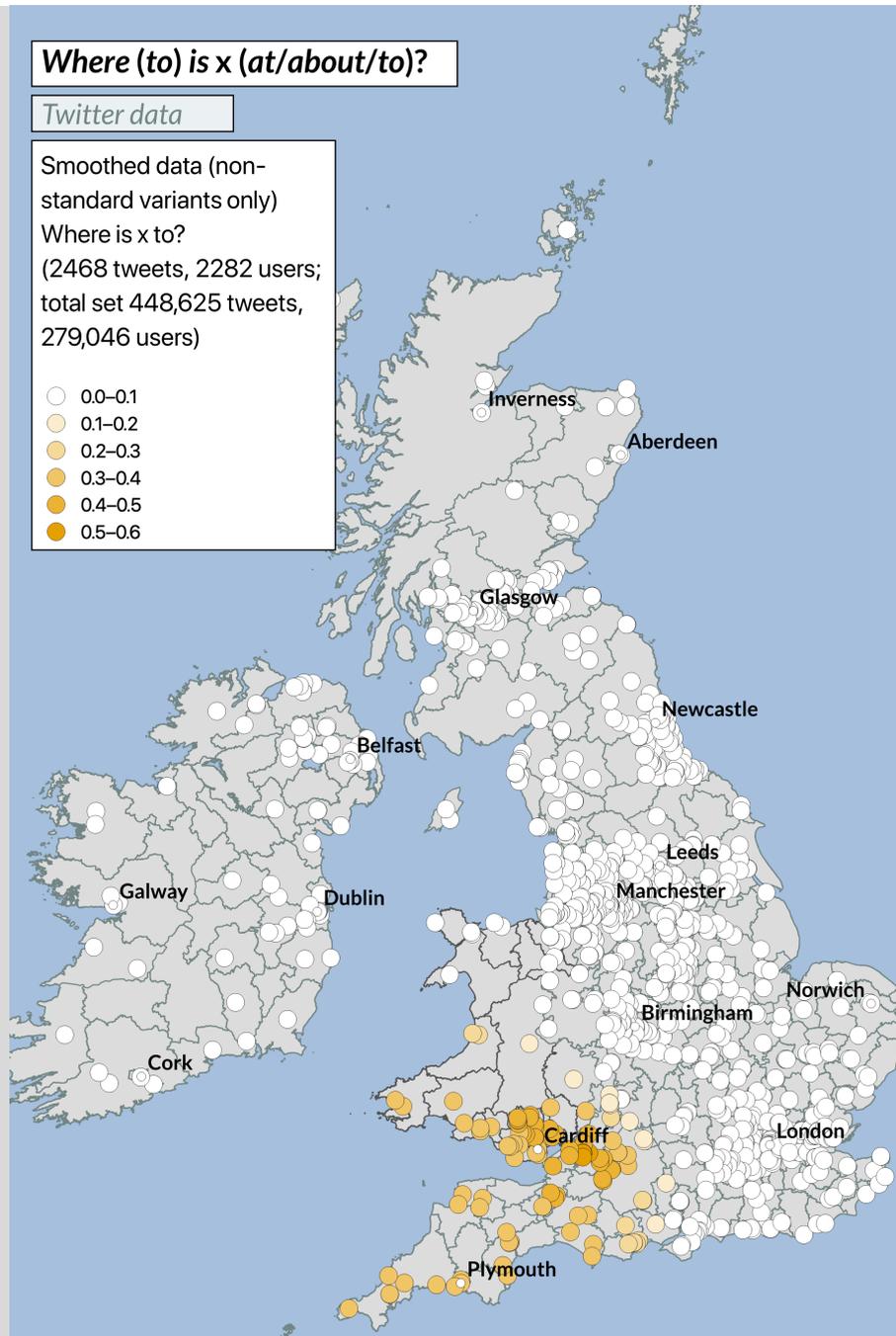
# Where ... to ...?

## Where (to) is x (at/about/to)?

### Twitter data

Smoothed data (non-standard variants only)  
Where is x to?  
(2468 tweets, 2282 users;  
total set 448,625 tweets,  
279,046 users)

- 0.0-0.1
- 0.1-0.2
- 0.2-0.3
- 0.3-0.4
- 0.4-0.5
- 0.5-0.6

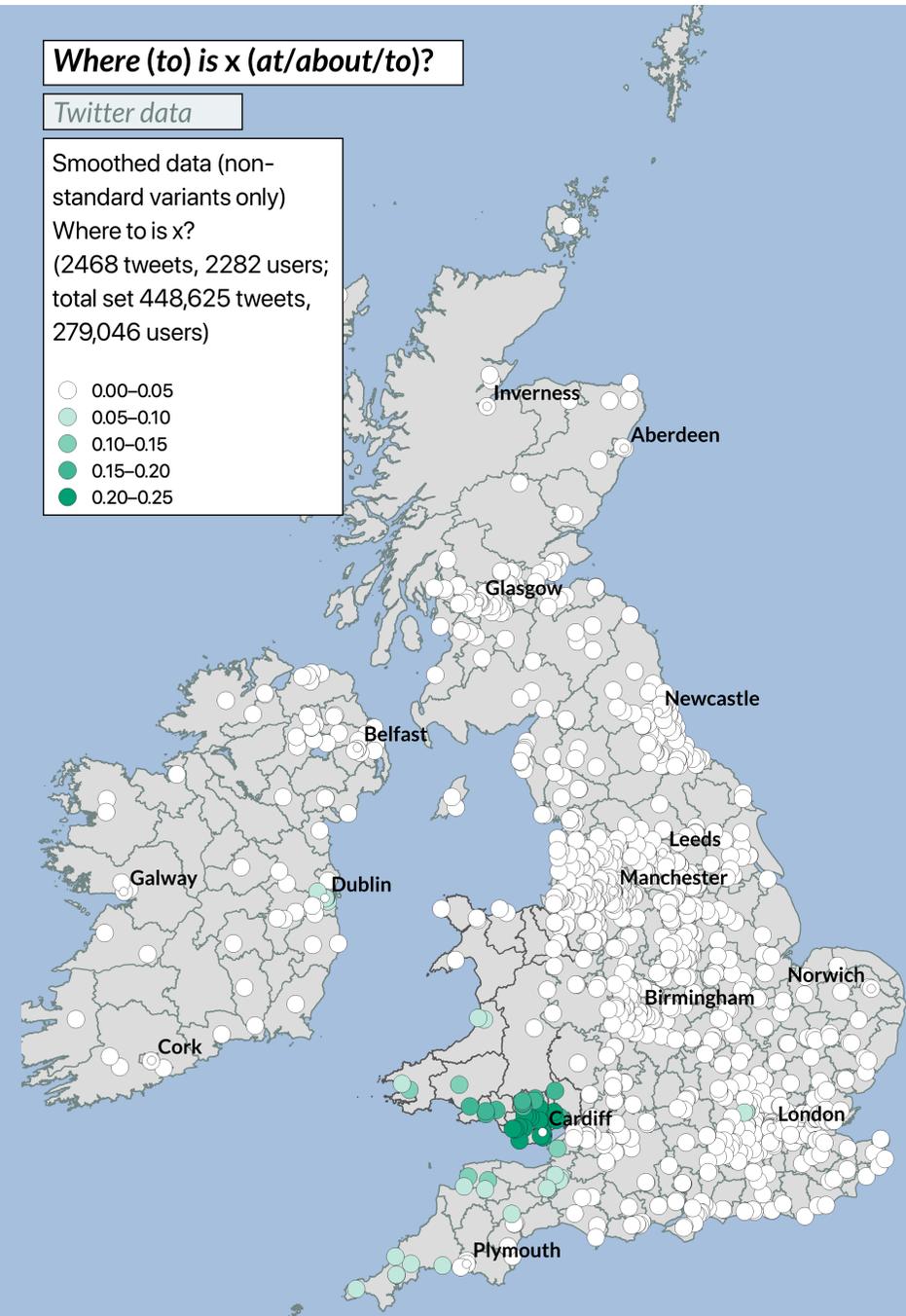


## Where (to) is x (at/about/to)?

### Twitter data

Smoothed data (non-standard variants only)  
Where to is x?  
(2468 tweets, 2282 users;  
total set 448,625 tweets,  
279,046 users)

- 0.00-0.05
- 0.05-0.10
- 0.10-0.15
- 0.15-0.20
- 0.20-0.25



## *Where ... to ...?*

- range of *where to is* x? is a subset of range of *where is* x *to*?
- even in its heartland, rates of *where to is* x are lower than rates of *where is* x *to*?
- the distribution implies the former is a development of the latter, presumably via reanalysis in ellipsis contexts: *Where is* ~~x~~ *to*? => *Where to is* ~~x~~?
- neither form is found in OED or EDD => recent but prior to end of settlement of Newfoundland (Late Modern English)?
- realignment of locative expressions is common crosslinguistically (Nikitina 2017):
  - Source (ablative) > Place (locative) e.g. Ancient Gk *ópisthe(n)* ‘behind’ < ‘from behind’
  - Goal (allative) > Place (locative) e.g. English *to the left*, German *zu Linken*

# Where ... to ...?

- requires obsolescence of *whither*, with *where* (*to*) filling the gap (Stolz, Levkovych & Urdze 2017: 217–218):

	Place	Goal	Source
	locative	allative	ablative
EModE	<i>where</i>	<i>whither</i>	<i>whence</i>
	<i>where</i>	<i>where (to)</i>	<i>from whence</i>
	<i>where</i>	<i>where (to)</i>	<i>where (from)</i>

- Bristol–Cardiff has gone one stage further, partially levelling Place ≠ Goal in favour of Goal (contrast earlier (?) levelling in favour of Place):

Br/Cdf	<i>where (to)</i>	<i>where (to)</i>	<i>where (from)</i>
--------	-------------------	-------------------	---------------------

- formally, the dialect differences here are to be represented as lexical differences

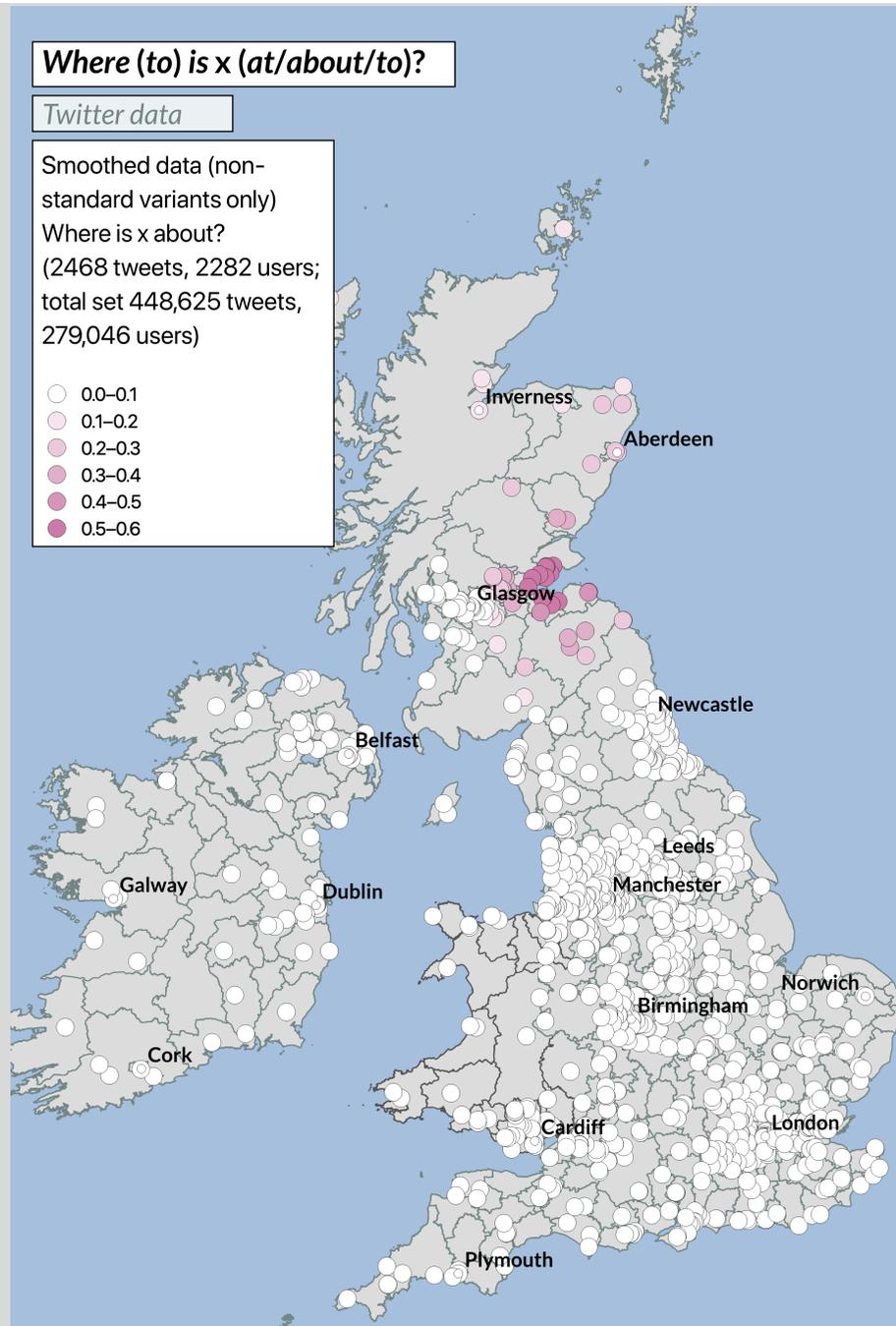
# Where ... about?

## Where (to) is x (at/about/to)?

### Twitter data

Smoothed data (non-standard variants only)  
Where is x about?  
(2468 tweets, 2282 users;  
total set 448,625 tweets,  
279,046 users)

- 0.0-0.1
- 0.1-0.2
- 0.2-0.3
- 0.3-0.4
- 0.4-0.5
- 0.5-0.6



# Where ... about?

- unlike *where ... to*, *where ... about* is historically compositional in meaning and old in its contiguous form:
  - Quar abute a-bide yee nu? (Cursor Mundi (Vesp.), line 15429)  
(composed after 1325)
- some of the early examples are Scottish
- *whereabout* > *whereabouts* is an analogical extension of the adverbial genitive *-s* (original in e.g. OE *tō-gēanes* 'against', *tō-middes* 'amidst') (cf. OE *ongēan* > *ongēans* > *against*; ME *alway* > *always*; *forward* > *forwards* etc.; cf. German *nachts*)
- this just needs the innovation of the split form by a reanalysis of ellipsis as before (cf. *where ... to*, but in reverse, i.e. the reanalysis is bidirectional)
- in all other aspects, this is a conservative relic feature

# Why for?

- posited by Brookes et al. (2017) as a feature of Multicultural London English for reason and purpose *wh-questions*, where other varieties use *why*, *what...for* or (for reason) *how come*
- associated with ‘pragmatically charged environments’?? (Brookes et al. 2017: 6)
- however, they found variation was not conditioned by sociodemographic factors, including residence within London
- contiguous *forwhy* goes back to Old English (preposition *for* + neuter instrumental of *hwā/hwæt* ‘who, what’), with contiguous *why for* in Middle English (exx. from Brookes et al. 2017: 40):

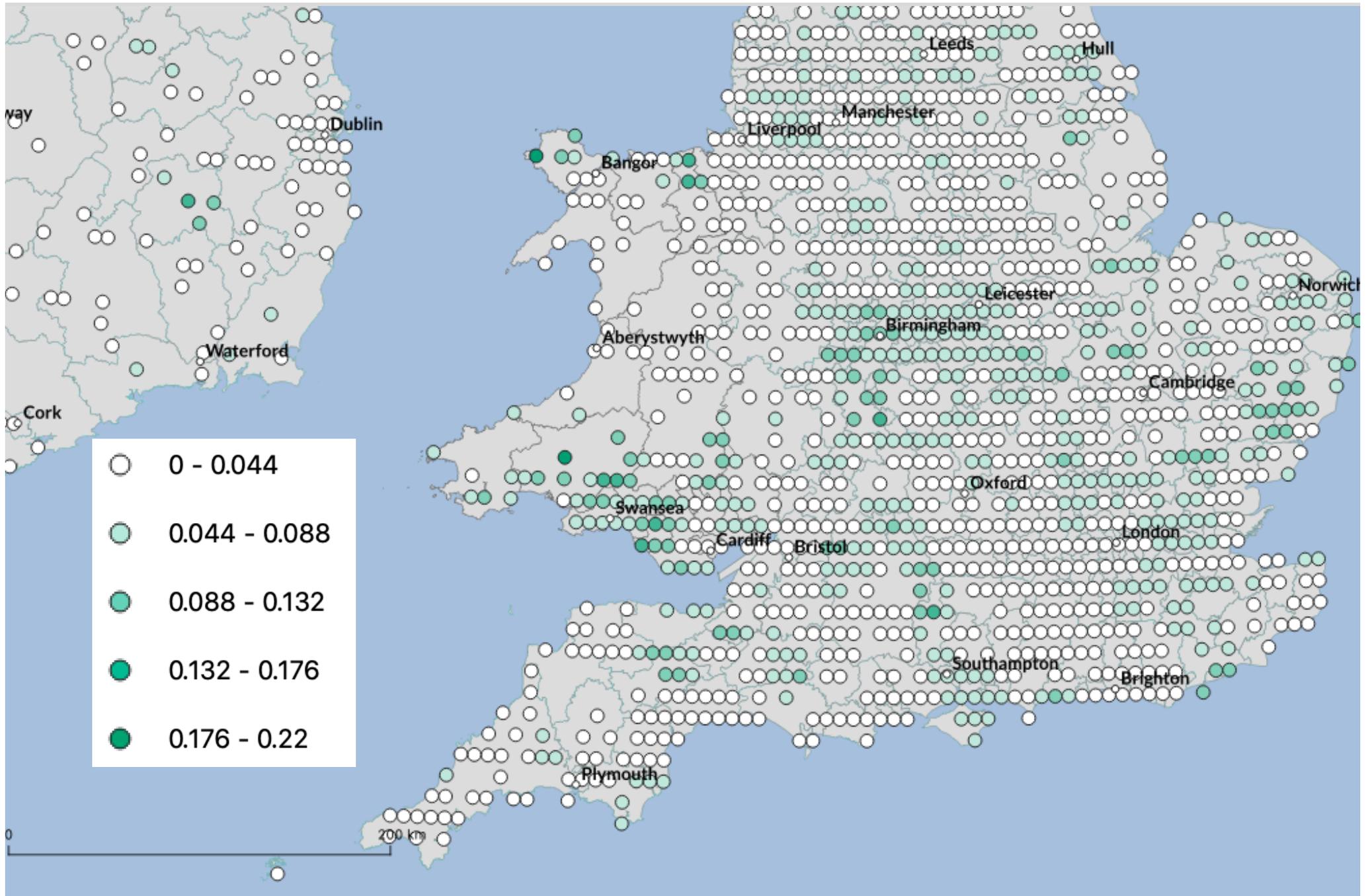
Nay, *why d’you kneel to me for?* I a’n’t your God-father. (Peter Anthony Motteux, *The Novelty*, 1697)

# *Why for?*

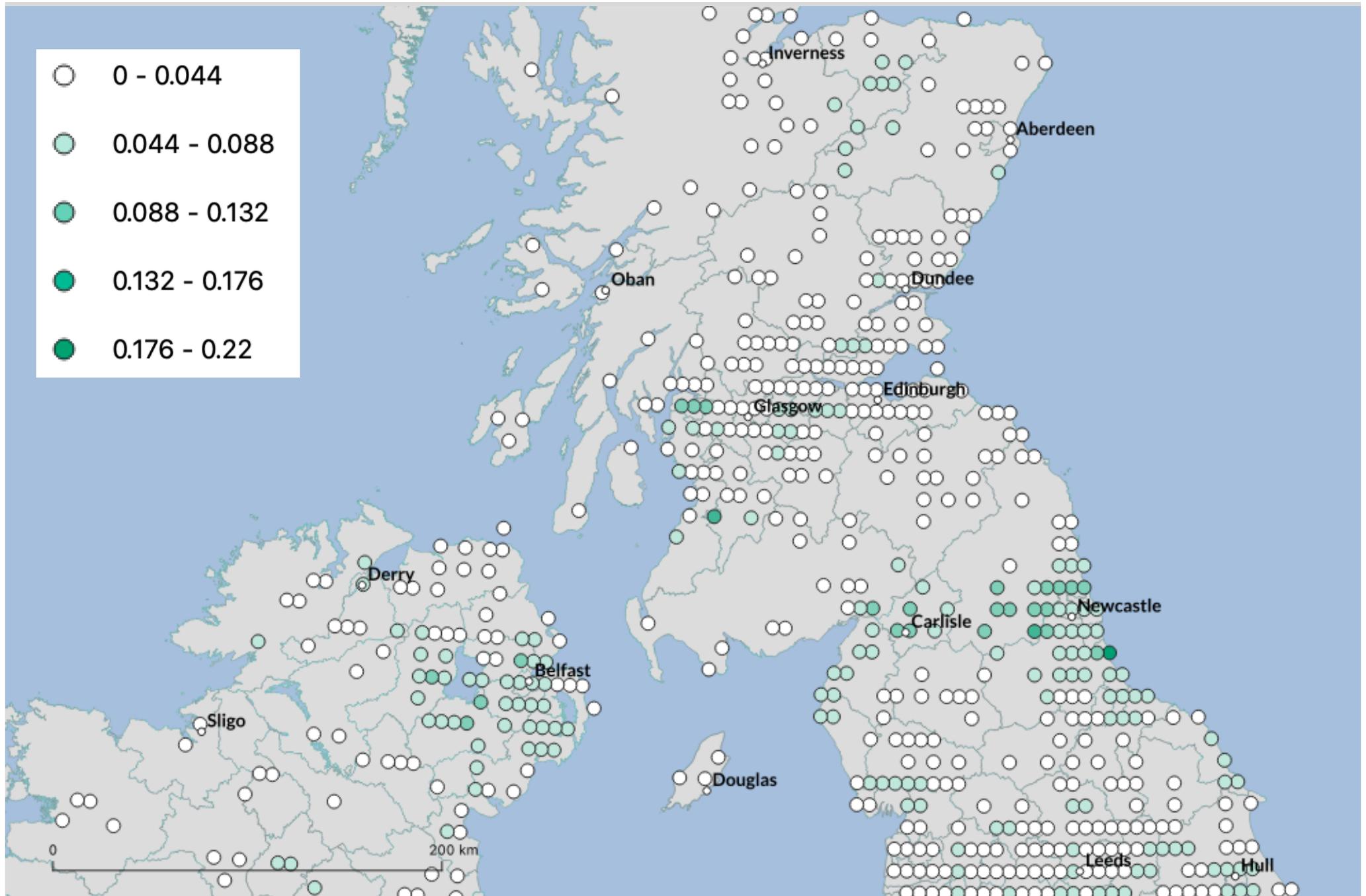
When I came to, I looked round, and he was gone. With that I walked towards where he was going, and I see him. With that my nephew see me. He said, "What is the matter with your jaw." McKen was three or four yards away. I said to him, "**Why did you want to run away so quick for?**" (Old Bailey Corpus 1910)

- found in US and other varieties of English

# Why for?



# Why for?



# Why for?

- no clear/strong geospatial distribution
- possible association with conservative urban varieties (Birmingham/West Midlands, Leicester, Newcastle/Tyneside, Belfast)
- little presence in rural areas in England and Wales
- not much found in Scotland or in Ireland outside the Ulster Scots area
- not associated with London, consequently...
- no obvious association with Multicultural London English
- data consistent with this as a relic feature going back to OE/ME (reformed as *what for* once the instrumental case was lost), perhaps reinvigorated or less recessive in some urban varieties
- for this reason, the contiguous variant is probably historically primary (cf. *whereabouts*)

# Conclusion

- social-media data can enrich our interpretations of variation and change in English, providing valuable additional sources over large numbers of speakers across wide areas
- plausible historical narratives emerge from large-scale geospatial distributions that would be difficult to establish without extensive surveying
- formally, these narratives require a bi-directional reanalysis involving ellipsis

# References

- Brookes, James, David Hall, Jenny Cheshire & David Adger. 2017. Causal interrogative variation in multicultural and traditional varieties of London English. *Queen Mary's Occasional Papers Advancing Linguistics* 36. 1–45.
- Campbell, Rowan Hope. 2021. *Directions of change in Cardiff English: Levelling, standardisation, or drift?* Cardiff: Cardiff University PhD dissertation.
- Clarke, Sandra. 2010. *Newfoundland and Labrador English*. Edinburgh: Edinburgh University Press.
- Nikitina, Tatiana. 2017. Ablative and allative marking of static locations. In Silvia Luraghi, Tatiana Nikitina & Chiara Zanchi (eds.), *Space in diachrony*, 67–94. Amsterdam: John Benjamins.
- Stolz, Thomas, Nataliya Levkovych & Aina Urdze. 2017. Typology and dynamics (with special focus on the development from Latin to Romance). In Silvia Luraghi, Tatiana Nikitina & Chiara Zanchi (eds.), *Space in diachrony*, 207–240. Amsterdam: John Benjamins.