Geographical Text Analysis:
Analysing digital texts using GIS

Ian Gregory,

Lancaster University

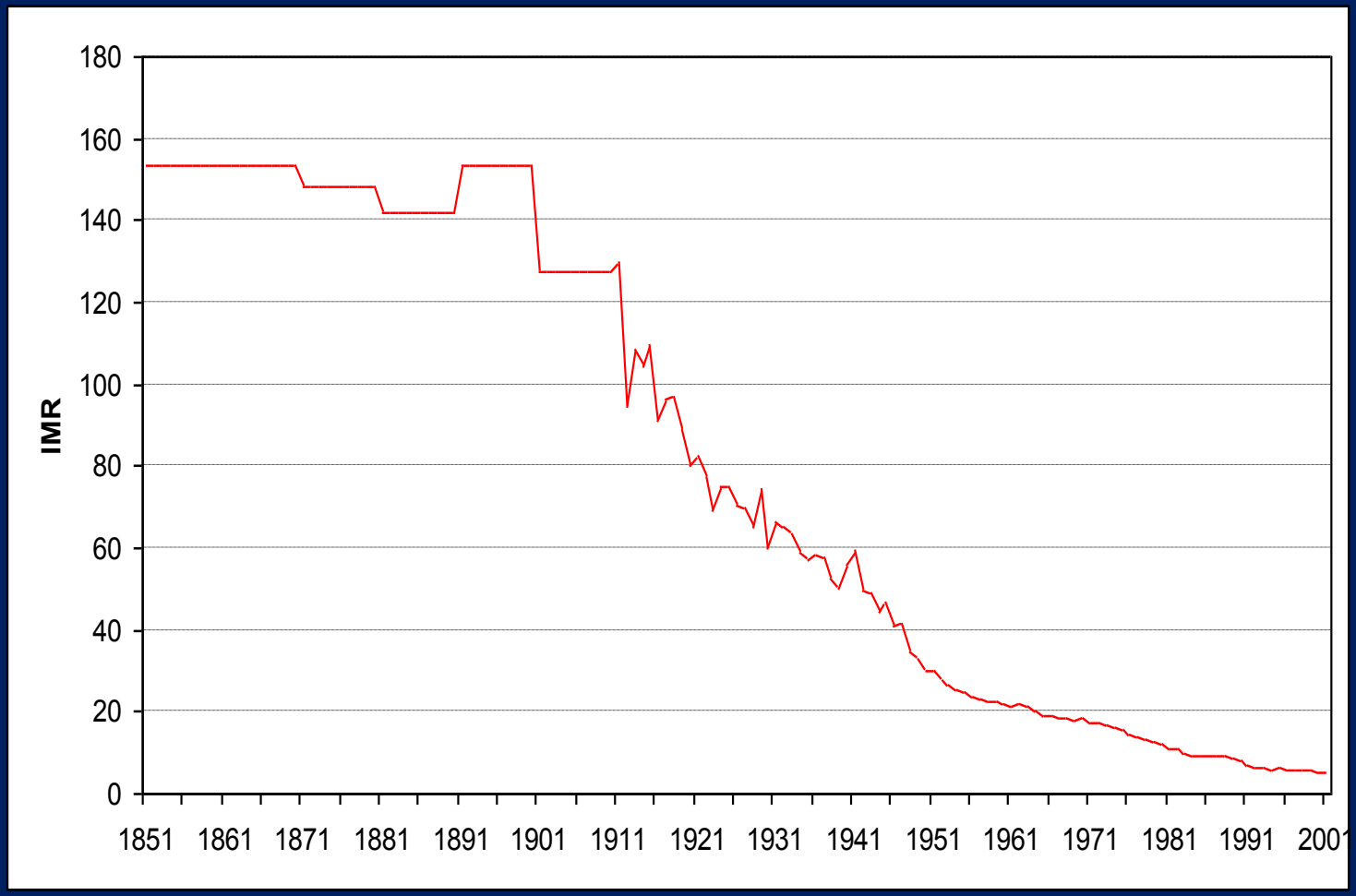
Acknowledgements:

Alistair Baron, David Cooper, Chris Donaldson, Andrew Hardie, Patricia Murrieta-Flores, Paul Rayson and C.J. Rupp (Lancaster)

Claire Grover (Edinburgh) – providing access to the geo-reference Histpop data

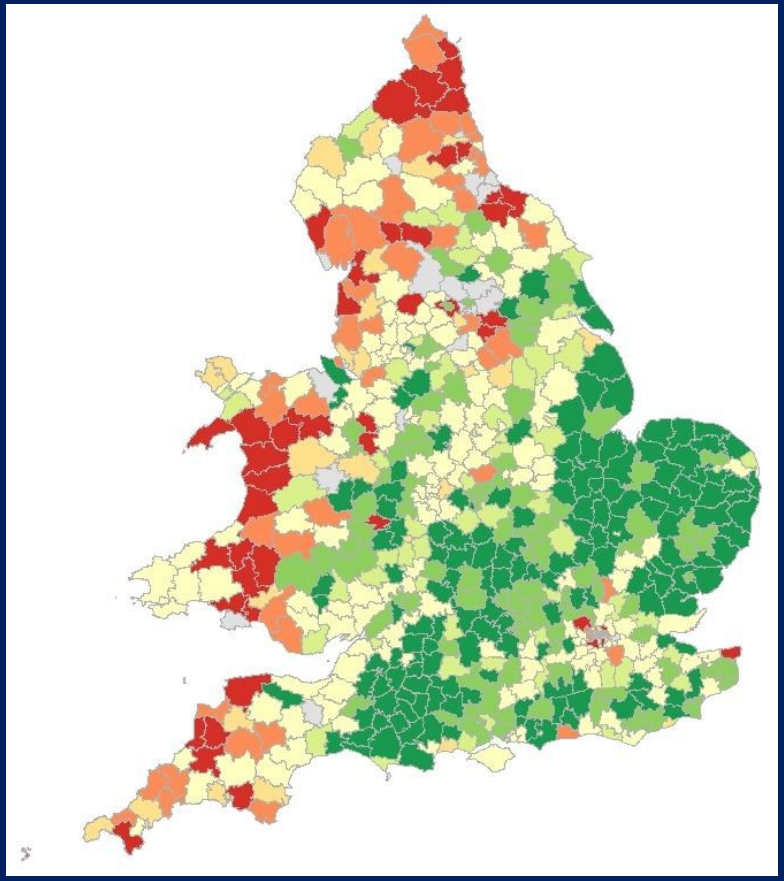
Richard Deswarte – help with the HistPop data

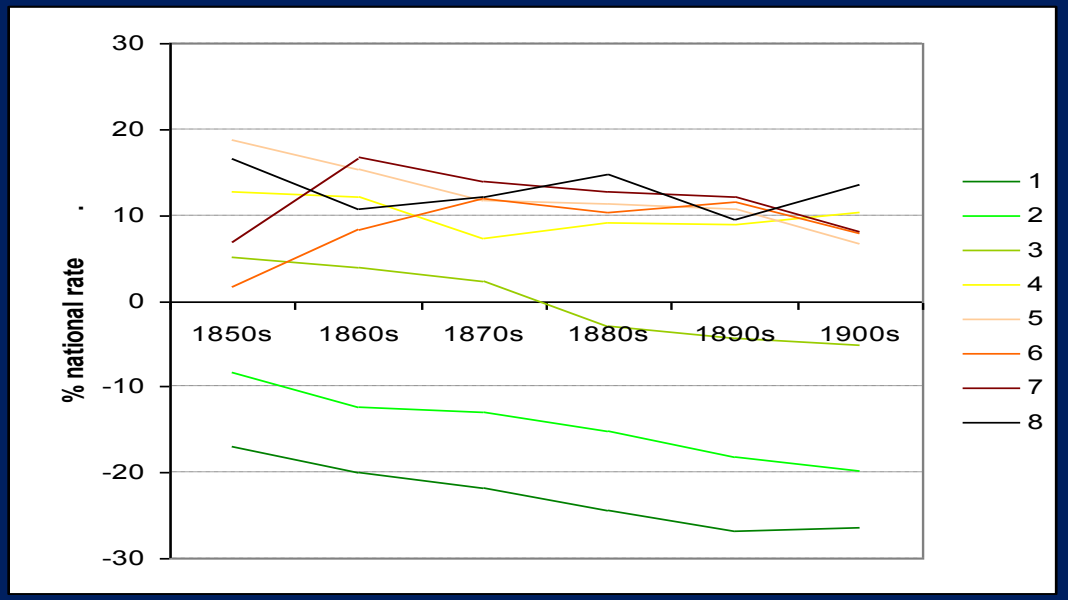
Change in Infant Mortality in England & Wales, 1851-2001



Traditional HGIS:

Infant mortality decline in England & Wales, 1851-1911





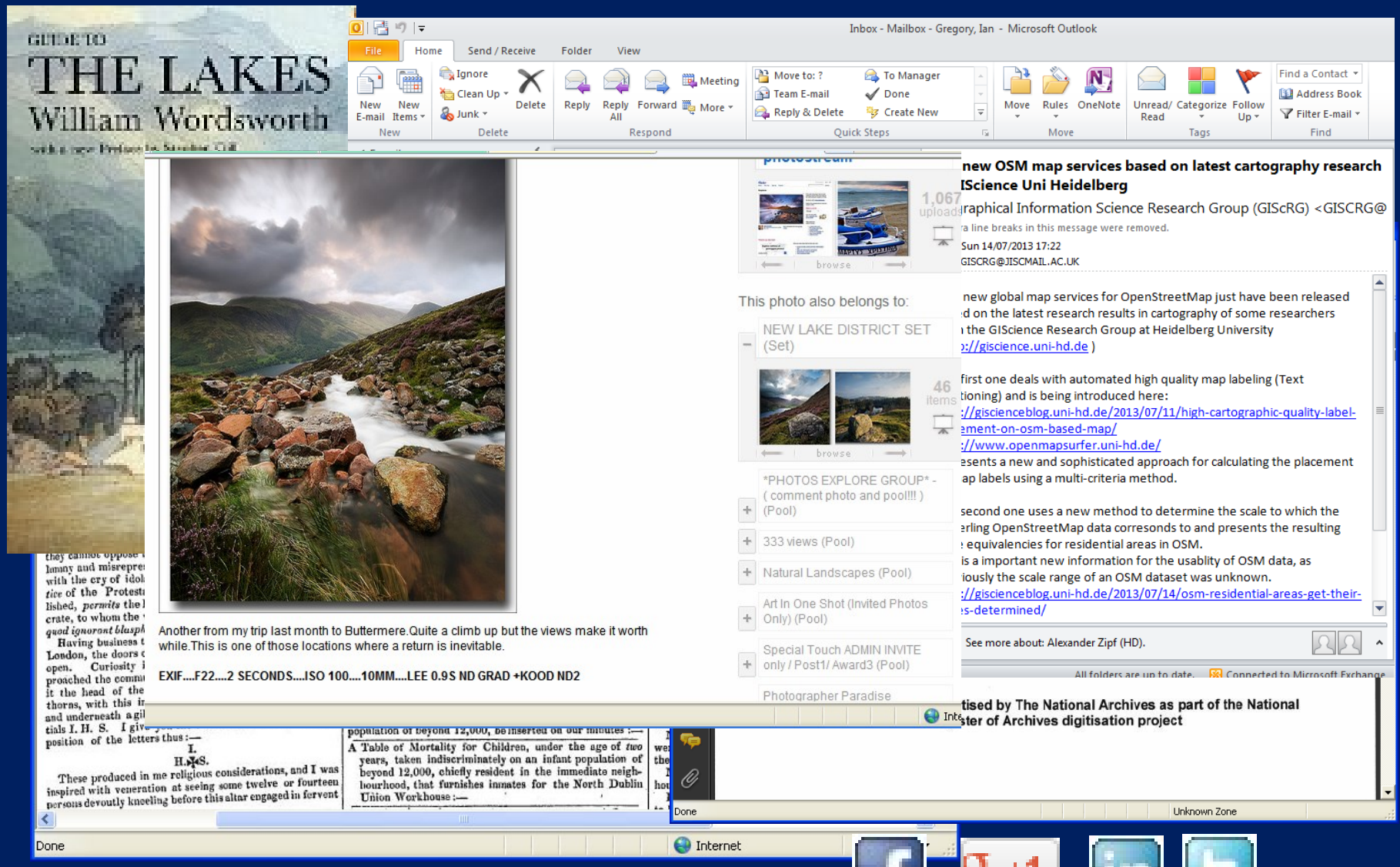
Source: Gregory (2008)
Annals of the Assoc. of American Geographers

Data volumes

- 2013: Estimated that 90% in existence were created in the past two years
- Amount of data will double every two years between now and 2020
  - Risen, J. and Lichtblau, E. "How the U.S. uses technology to mine more data more quickly." *New York Times.* 8th June 2013.

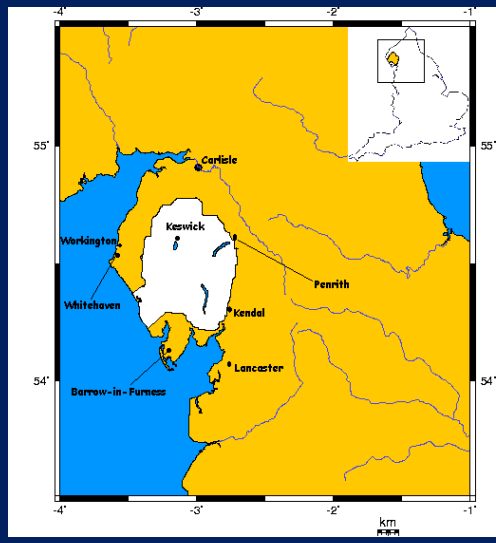
REG_DIST	REG_CNTY	TOT_POP	TOT_MALE	M_0	M_1	M_2	M_3	M_4	M_0_4	M_5_9	M
MILTON	KENT	23270	11966	380	351	326	384	336	1777	1463	
MITFORD	NORFOLK	27367	13590	371	347	371	358	351	1798	1627	
MONMOUTH	MONMOUTHSI	30340	15377	389	379	418	456	400	2042	2034	
MORPETH	NORTHUMBER	36077	18444	585	491	510	549	531	2666	2464	
MUTFORD	SUFFOLK	39478	19057	616	579	621	553	592	2961	2488	
NANTWICH	CHESHIRE	61566	31062	887	847	885	863	810	4292	4011	
NARBERTH	PEMBROKESHI	19541	9149	258	235	270	243	279	1285	1196	
NEATH	GLAMORGANS	52077	26219	792	751	735	763	765	3806	3508	
NEW FOREST	HAMPSHIRE	13221	6609	161	160	160	162	174	817	885	
NEWARK	NOTTINGHAM	30616	15114	419	376	414	397	408	2014	1833	
NEWBURY	BERKSHIRE	21327	10208	276	274	250	237	273	1310	1251	
NEWCASTLE IN EMLYN	CARDIGANSHI	19014	8643	232	230	228	229	231	1150	1152	
NEWCASTLE UNDER LY	STAFFORDSHIF	34661	17724	556	489	500	497	556	2598	2350	
NEWCASTLE UPON TYN	NORTHUMBER	150252	73572	2317	2143	2023	2108	2032	10623	9189	
NEWENT	GLOUCESTERSH	11030	5496	133	125	137	134	151	680	678	
NEWMARKET	CAMBRIDGESH	28247	14185	363	317	368	353	383	1784	1724	
NEWPORT	SHROPSHIRE	15352	7707	206	178	197	211	200	992	949	
NEWPORT	MONMOUTHSI	70542	35197	1073	924	1019	1026	964	5006	4545	
NEWPORT PAGNELL	BUCKINGHAM	24583	12030	372	319	351	339	346	1727	1537	
NEWTON ABBOT	DEVONSHIRE	74996	32934	972	850	926	856	889	4493	3964	
NEWTOWN	MONTGOMERY	25439	12732	338	311	370	333	361	1713	1574	
NORTH AYLESFORD	KENT	27437	14683	440	363	413	375	397	1988	1723	
NORTH WITCHFORD	CAMBRIDGESH	15464	7713	201	207	180	207	215	1010	978	
NORTHALLERTON	NORTH RIDING	11884	5989	174	135	175	151	163	798	736	
NORTHAMPTON	NORTHAMPTO	64244	31520	1010	843	875	849	871	4448	3959	
NORTHLEACH	GLOUCESTERSH	9884	5131	121	123	129	142	131	646	636	

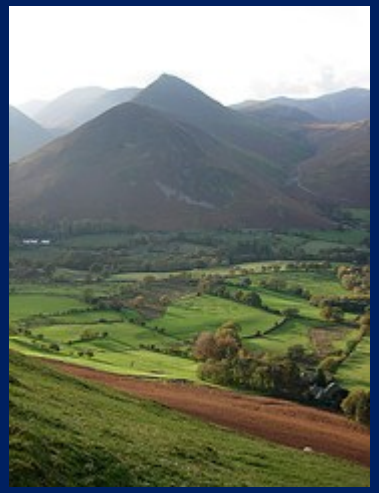
Modern ‘data’


GUIDE TO
THE LAKES
William Wordsworth
Inbox - Mailbox - Gregory, Ian - Microsoft Outlook
File
Home
Send / Receive
Folder
View
New E-mail
New Items
Ignore
Clean Up
Junk
Delete
Reply
Reply All
Forward
Meeting
More
Move to: ?
Team E-mail
Reply & Delete
To Manager
Done
Create New
Quick Steps
Move
Rules
OneNote
Unread/ Read
Categorize
Follow Up
Find a Contact
Address Book
Filter E-mail
New
Delete
Respond
Move
Tags
Find
Another from my trip last month to Buttermere.Quite a climb up but the views make it worth while.This is one of those locations where a return is inevitable.
EXIF....F22....2 SECONDS....ISO 100....10MM....LEE 0.9S ND GRAD +KOOD ND2
This photo also belongs to:
NEW LAKE DISTRICT SET (Set)
PHOTOS EXPLORE GROUP - (comment photo and pool!!!) (Pool)
333 views (Pool)
Natural Landscapes (Pool)
Art In One Shot (Invited Photos Only) (Pool)
Special Touch ADMIN INVITE only / Post1/ Award3 (Pool)
Photographer Paradise
new OSM map services based on latest cartography research
IScience Uni Heidelberg
raphical Information Science Research Group (GIScRG) <GISCRG@
Sun 14/07/2013 17:22
GISCRG@JISCMAIL.AC.UK
new global map services for OpenStreetMap just have been released
on the latest research results in cartography of some researchers
the GIScience Research Group at Heidelberg University
://giscience.uni-hd.de)
first one deals with automated high quality map labeling (Text
tioning) and is being introduced here:
://giscienceblog.uni-hd.de/2013/07/11/high-cartographic-quality-label-
ement-on-osm-based-map/
://www.openmapsurfer.uni-hd.de/
esents a new and sophisticated approach for calculating the placement
ap labels using a multi-criteria method.
second one uses a new method to determine the scale to which the
erling OpenStreetMap data corresonds to and presents the resulting
equivalencies for residential areas in OSM.
is a important new information for the usablity of OSM data, as
iously the scale range of an OSM dataset was unknown.
://giscienceblog.uni-hd.de/2013/07/14/osm-residential-areas-get-their-
s-determined/
See more about: Alexander Zipf (HD).
tised by The National Archives as part of the National
ter of Archives digitisation project
A Table of Mortality for Children, under the age of two years, taken indiscriminately on an infant population of beyond 12,000, chiefly resident in the immediate neighbourhood, that furnishes inmates for the North Dublin Union Workhouse :—
Done
Internet
Unknown Zone

Literary Mapping of the Lakes

- British Academy funded pilot project with David Cooper and Sally Bushell
- Two tours of the Lake District
  - Thomas Gray, 1769 (9,000 words)
    - Proto-Picturesque
  - ST Coleridge, 1802 (10,000 words)
    - Romantic
- Aims:
  - Can we create a GIS of text?
  - What can it offer to literary research?
- Method:
  - Texts typed up by hand
  - Places tagged manually
  - Conversion
  - Analysis

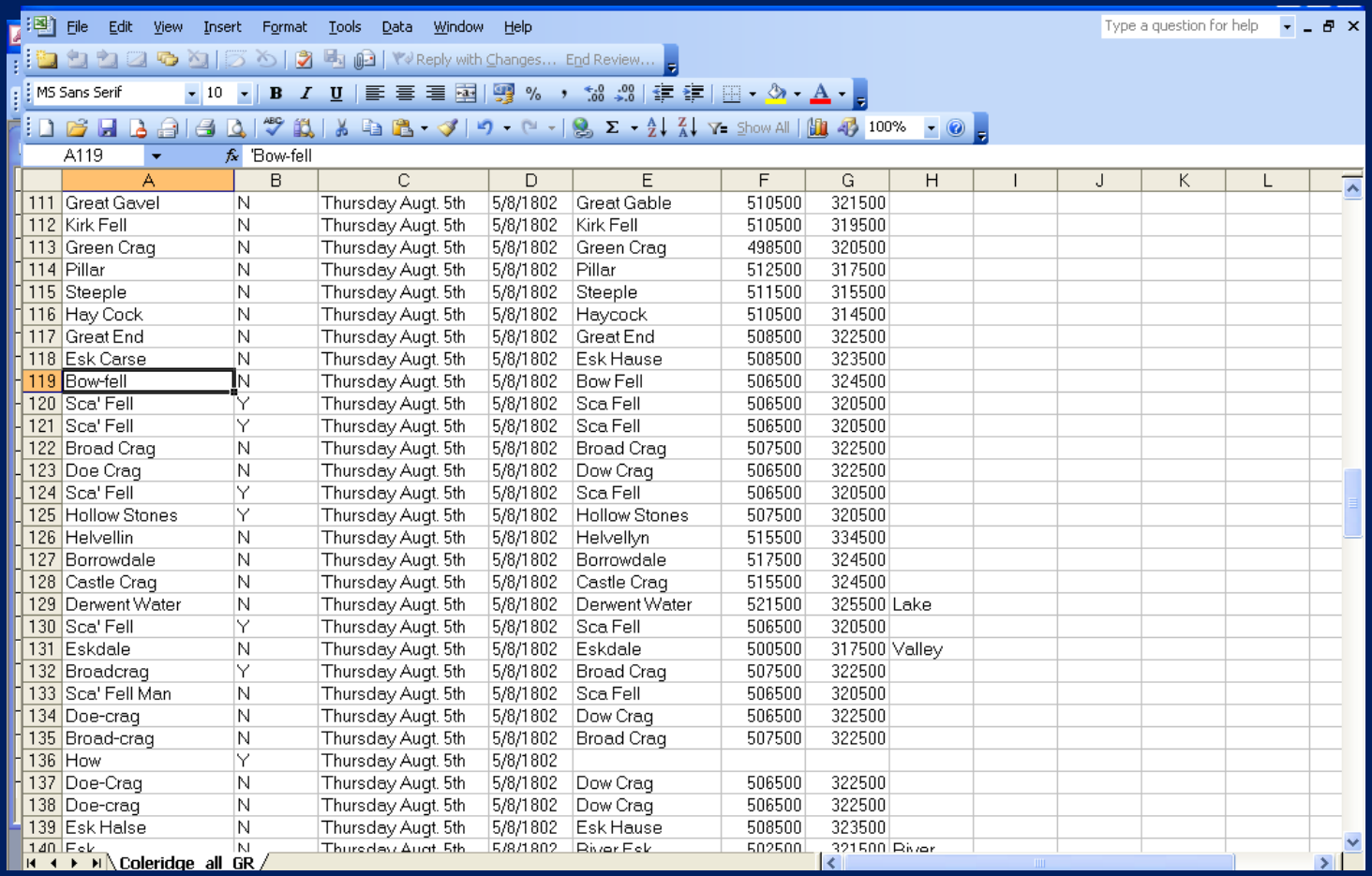




Place names coded in XML

<p in_text="Y">On Sunday Augt. 1st - half after 12 I had a Shirt, cravat, 2 pair of Stockings, a little paper & half a dozen Pens, a German Book (Voss's Poems) & a little Tea & Sugar, with my Night Cap, packed up in my natty green oil-skin, neatly squared, and put into my <format format_type="I">net</format> Knapsack / and the Knap-sack on my back & the Besom stick in my hand, which for want of a better, and in spite of <person>Mrs C.</person> & <person>Mary</person>, who both raised their voices against it, especially as I left the Besom scattered on the Kitchen Floor, off I sallied - over the Bridge<my_comment><pl_name visited="Y">Greta Bridge, Keswick</pl_name></my_comment>, thro' the Hop-Field, thro' the <pl_name visited="Y">Prospect Bridge</pl_name> at <pl_name visited="Y">Portinscale</pl_name>, so on by the tall Birch that grows out of the center of the huge Oak, along into <pl_name visited="Y">Newlands</pl_name>-- <pl_name visited="Y">Newlands</pl_name>is indeed a lovely Place-the houses...

Convert to a GIS

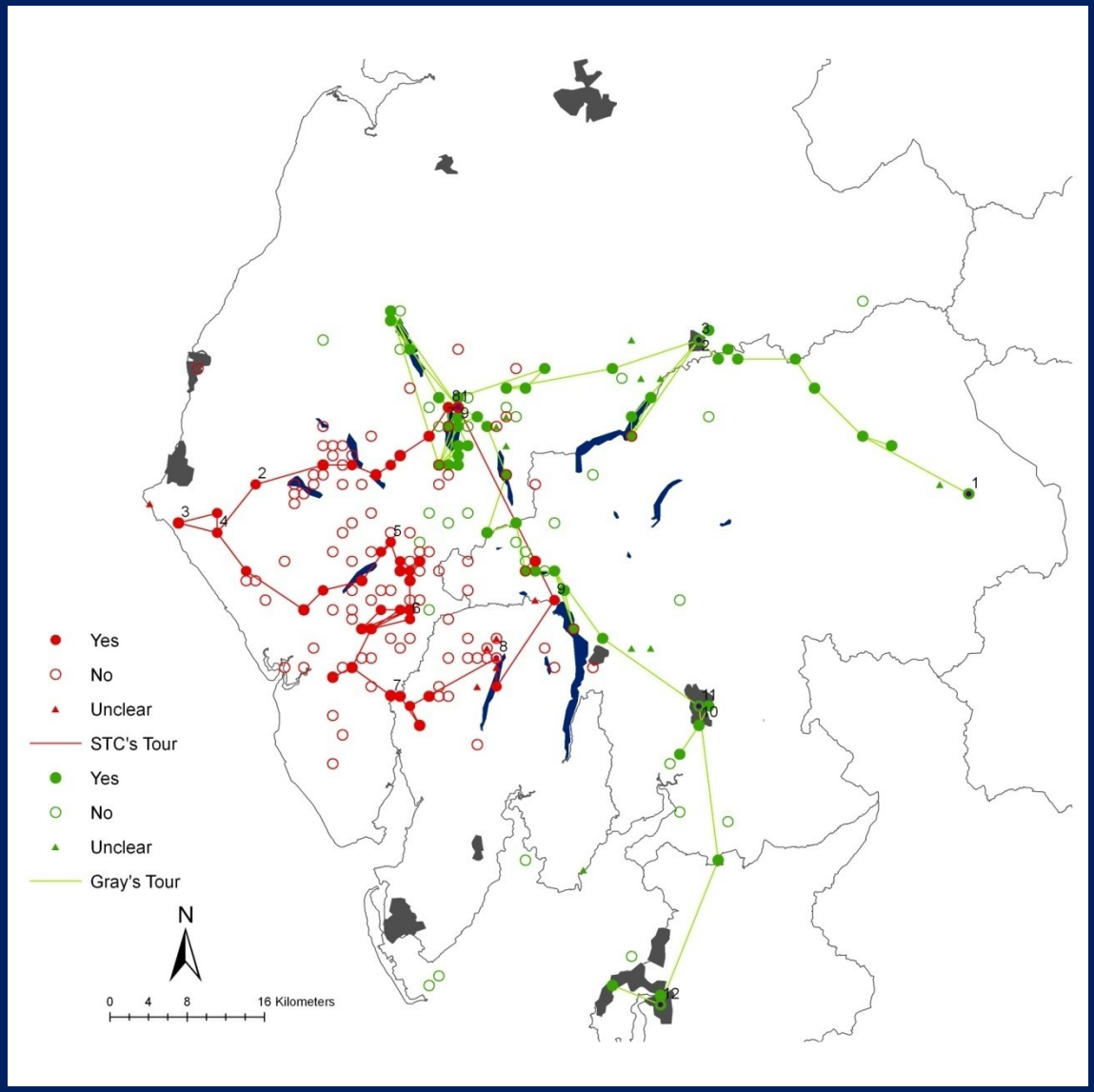


	A	B	C	D	E	F	G	H
111	Great Gavel	N	Thursday Augt. 5th	5/8/1802	Great Gable	510500	321500	
112	Kirk Fell	N	Thursday Augt. 5th	5/8/1802	Kirk Fell	510500	319500	
113	Green Crag	N	Thursday Augt. 5th	5/8/1802	Green Crag	498500	320500	
114	Pillar	N	Thursday Augt. 5th	5/8/1802	Pillar	512500	317500	
115	Steeple	N	Thursday Augt. 5th	5/8/1802	Steeple	511500	315500	
116	Hay Cock	N	Thursday Augt. 5th	5/8/1802	Haycock	510500	314500	
117	Great End	N	Thursday Augt. 5th	5/8/1802	Great End	508500	322500	
118	Esk Carse	N	Thursday Augt. 5th	5/8/1802	Esk Hause	508500	323500	
119	Bow-fell	N	Thursday Augt. 5th	5/8/1802	Bow Fell	506500	324500	
120	Sca' Fell	Y	Thursday Augt. 5th	5/8/1802	Sca Fell	506500	320500	
121	Sca' Fell	Y	Thursday Augt. 5th	5/8/1802	Sca Fell	506500	320500	
122	Broad Crag	N	Thursday Augt. 5th	5/8/1802	Broad Crag	507500	322500	
123	Doe Crag	N	Thursday Augt. 5th	5/8/1802	Dow Crag	506500	322500	
124	Sca' Fell	Y	Thursday Augt. 5th	5/8/1802	Sca Fell	506500	320500	
125	Hollow Stones	Y	Thursday Augt. 5th	5/8/1802	Hollow Stones	507500	320500	
126	Helvellin	N	Thursday Augt. 5th	5/8/1802	Helvellyn	515500	334500	
127	Borrowdale	N	Thursday Augt. 5th	5/8/1802	Borrowdale	517500	324500	
128	Castle Crag	N	Thursday Augt. 5th	5/8/1802	Castle Crag	515500	324500	
129	Derwent Water	N	Thursday Augt. 5th	5/8/1802	Derwent Water	521500	325500	Lake
130	Sca' Fell	Y	Thursday Augt. 5th	5/8/1802	Sca Fell	506500	320500	
131	Eskdale	N	Thursday Augt. 5th	5/8/1802	Eskdale	500500	317500	Valley
132	Broadcrag	Y	Thursday Augt. 5th	5/8/1802	Broad Crag	507500	322500	
133	Sca' Fell Man	N	Thursday Augt. 5th	5/8/1802	Sca Fell	506500	320500	
134	Doe-crag	N	Thursday Augt. 5th	5/8/1802	Dow Crag	506500	322500	
135	Broad-crag	N	Thursday Augt. 5th	5/8/1802	Broad Crag	507500	322500	
136	How	Y	Thursday Augt. 5th	5/8/1802				
137	Doe-Crag	N	Thursday Augt. 5th	5/8/1802	Dow Crag	506500	322500	
138	Doe-crag	N	Thursday Augt. 5th	5/8/1802	Dow Crag	506500	322500	
139	Esk Halse	N	Thursday Augt. 5th	5/8/1802	Esk Hause	508500	323500	
140	Esk	N	Thursday Augt. 5th	5/8/1802	River Esk	502500	321500	River

OS 1:50,000 gazetteer – all places on 1:50,000 maps

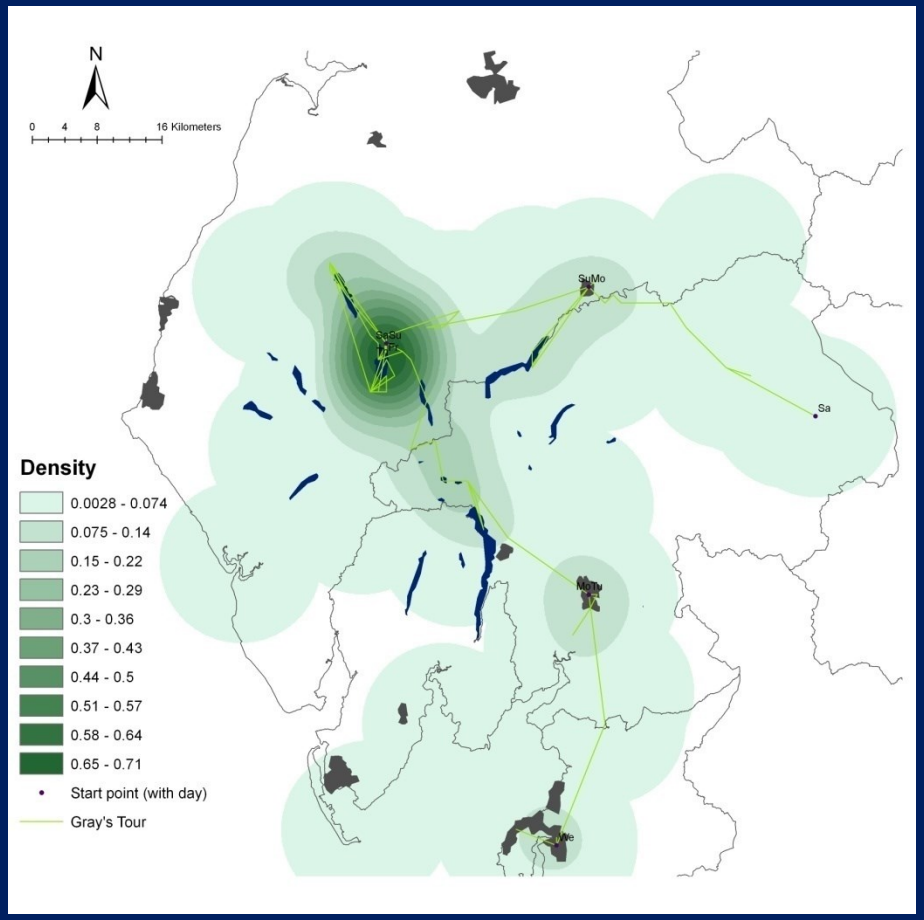
- Accuracy
- Spelling problems
- Disambiguation

Coleridge & Gray in a GIS

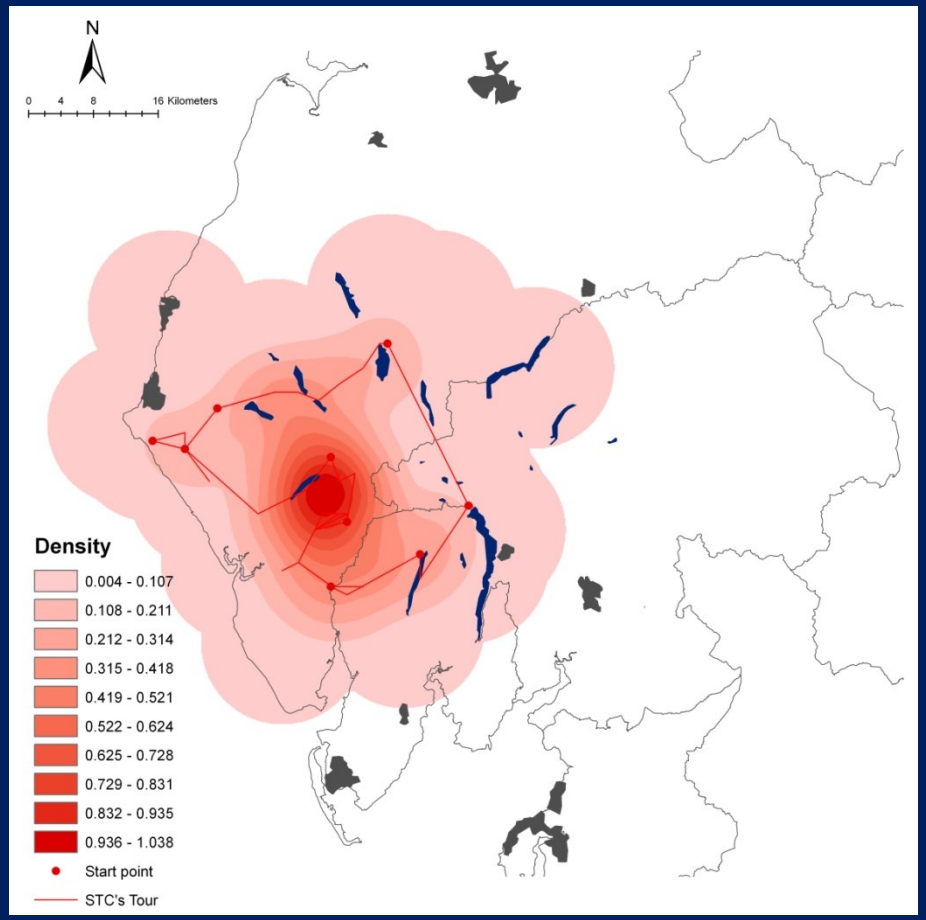


Smoothed surfaces of places

Gray

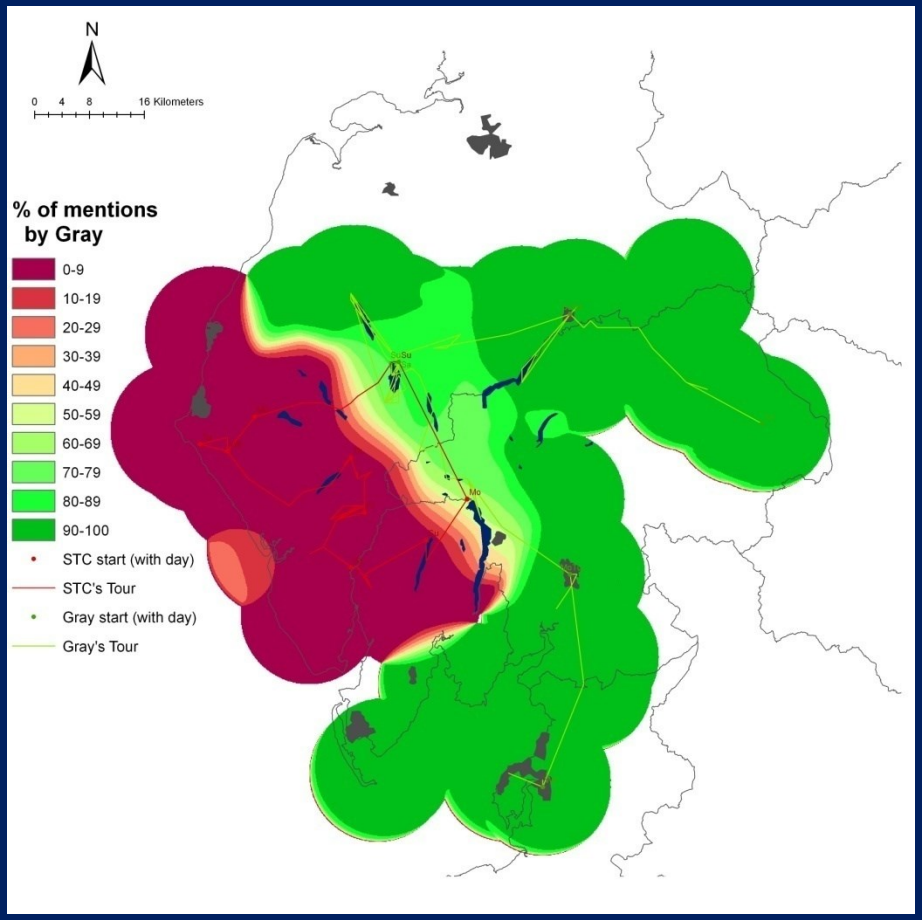


Coleridge



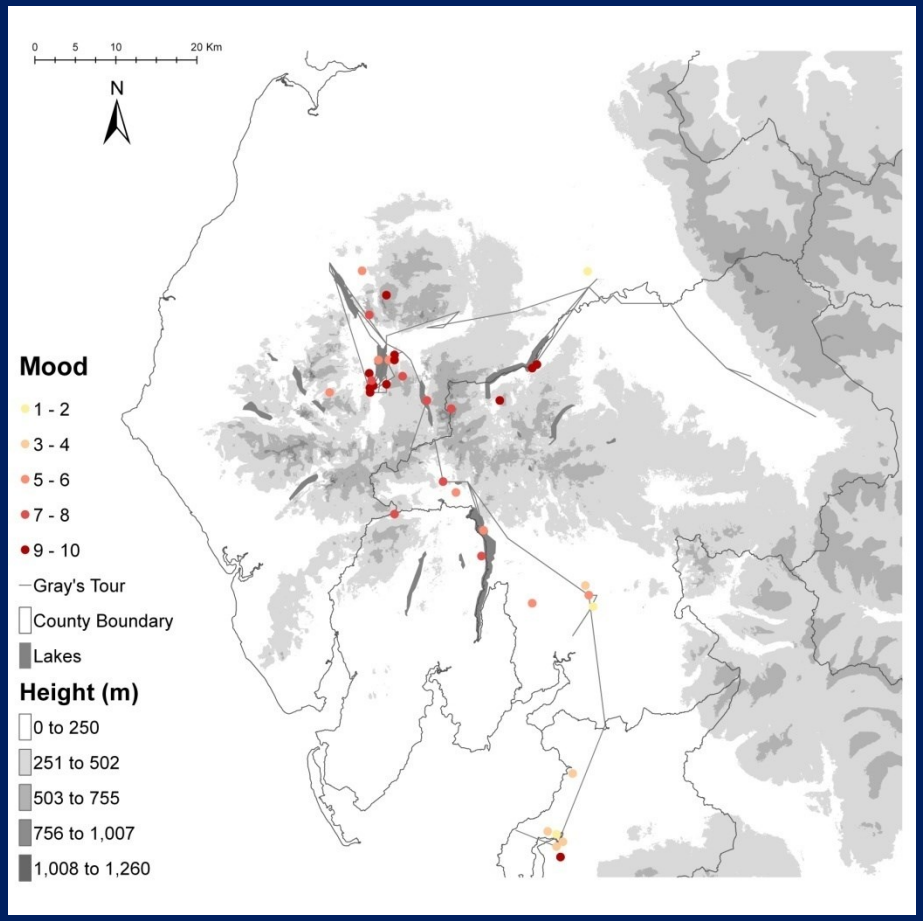
Comparing Coleridge and Gray

All mentions

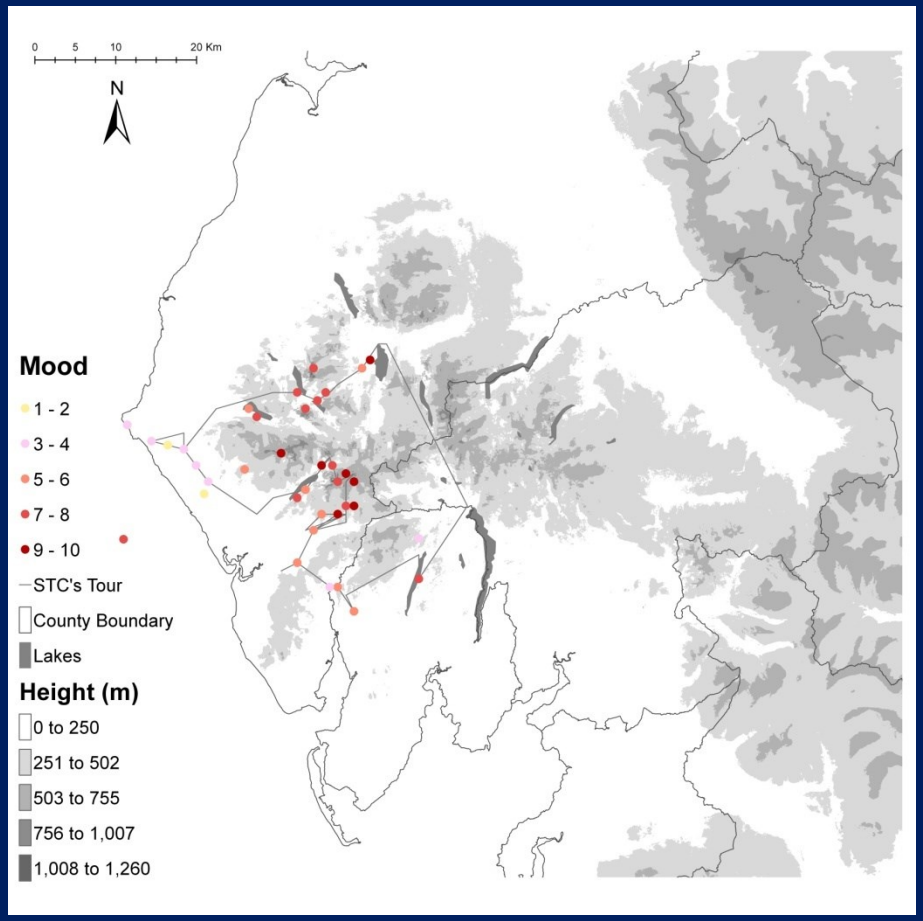


Green: Only in Gray
Yellow: Evenly in both
Red: Only in Coleridge

Mapping Emotional Response

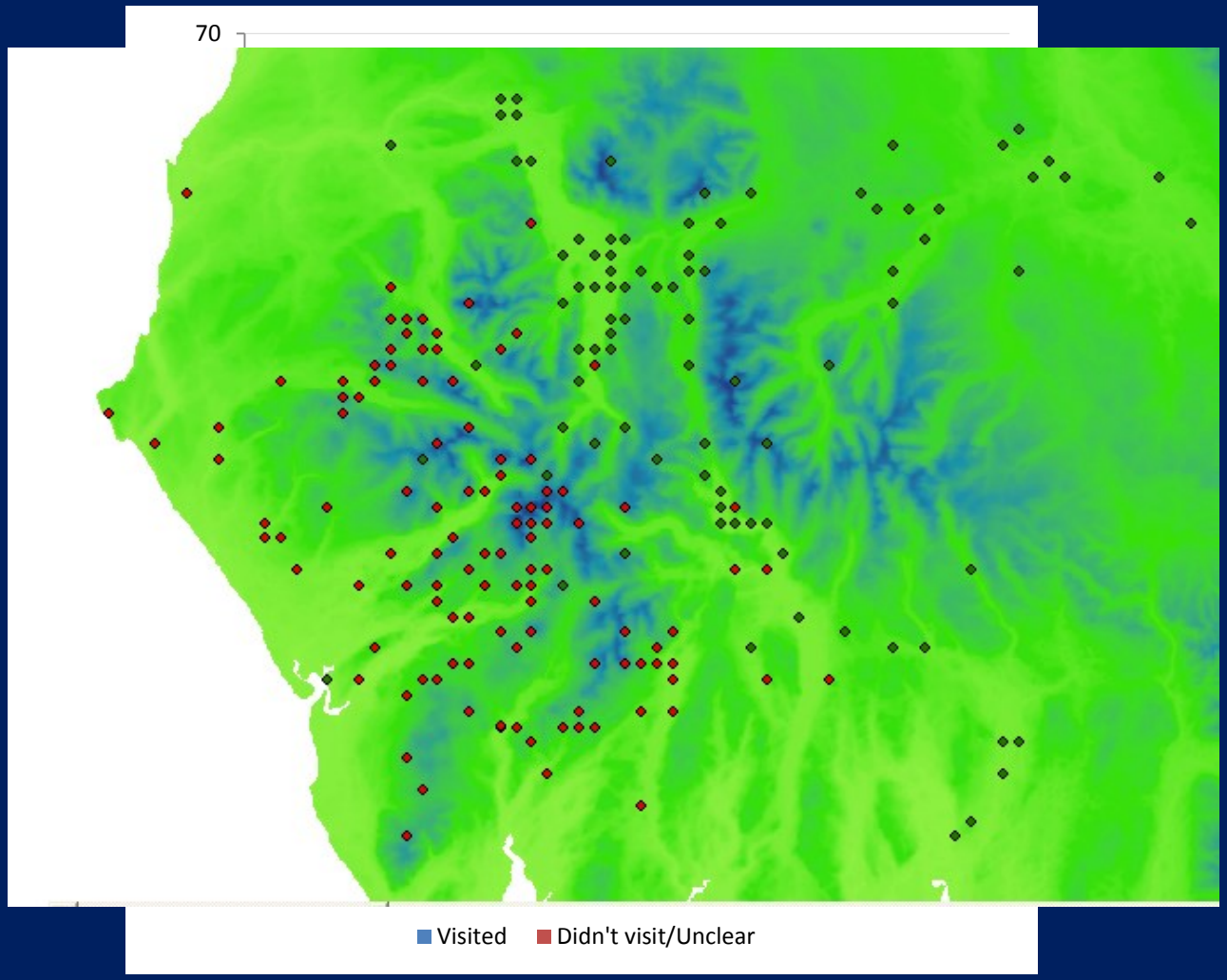


Gray



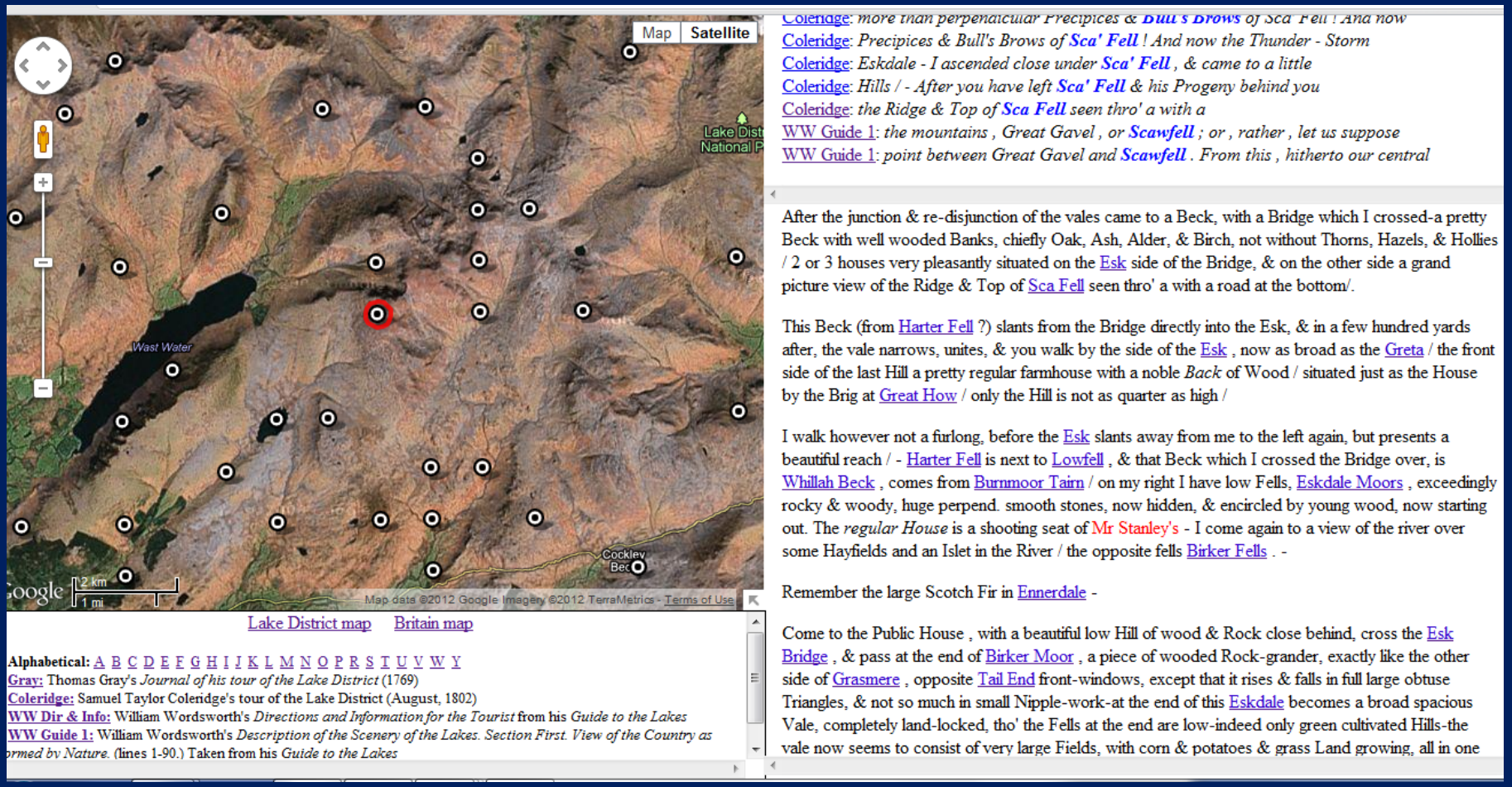
Coleridge

Physical Characteristics of Tours



Altitude of mentions

Close Reading with Internet Mapping



http://www.lancs.ac.uk/mappingthelakes
http://www.lancs.ac.uk/mappingthelakes/v2

The Histpop Collection

- Covers the printed reports published in the Census and the Registrar General's Annual Reports, 1801-1937
  - Nearly 13,000,000 words
- Georeferenced by C. Grover et al (University of Edinburgh)
  - Recall: 81%
  - Precision: 82%
  - Correct with locality 75% (Tobin et al 2010)
- Just concerned with the Registrar General's Reports, 1851-1911 and places in England & Wales
  - England & Wales: 2,000,000 words
- http://www.histpop.org

Geographical Text Analysis

- Combination of Computational Linguistics and GIS allows:
  - 1. Exploratory spatial approach:
    - Ask *where* is this corpus talking about?
    - Identify place-names in areas that the corpus concentrates on.
    - Find out what it is saying about these places
  - 2. Thematic approach:
    - Find out which places are associated with our theme
    - Find out what else it is saying in relation to this theme
    - Find out what other themes are associated with these places
  - 3. Textual/statistical comparisons
    - Explore biases in instances against background patterns
    - Explore how discourse varies within this

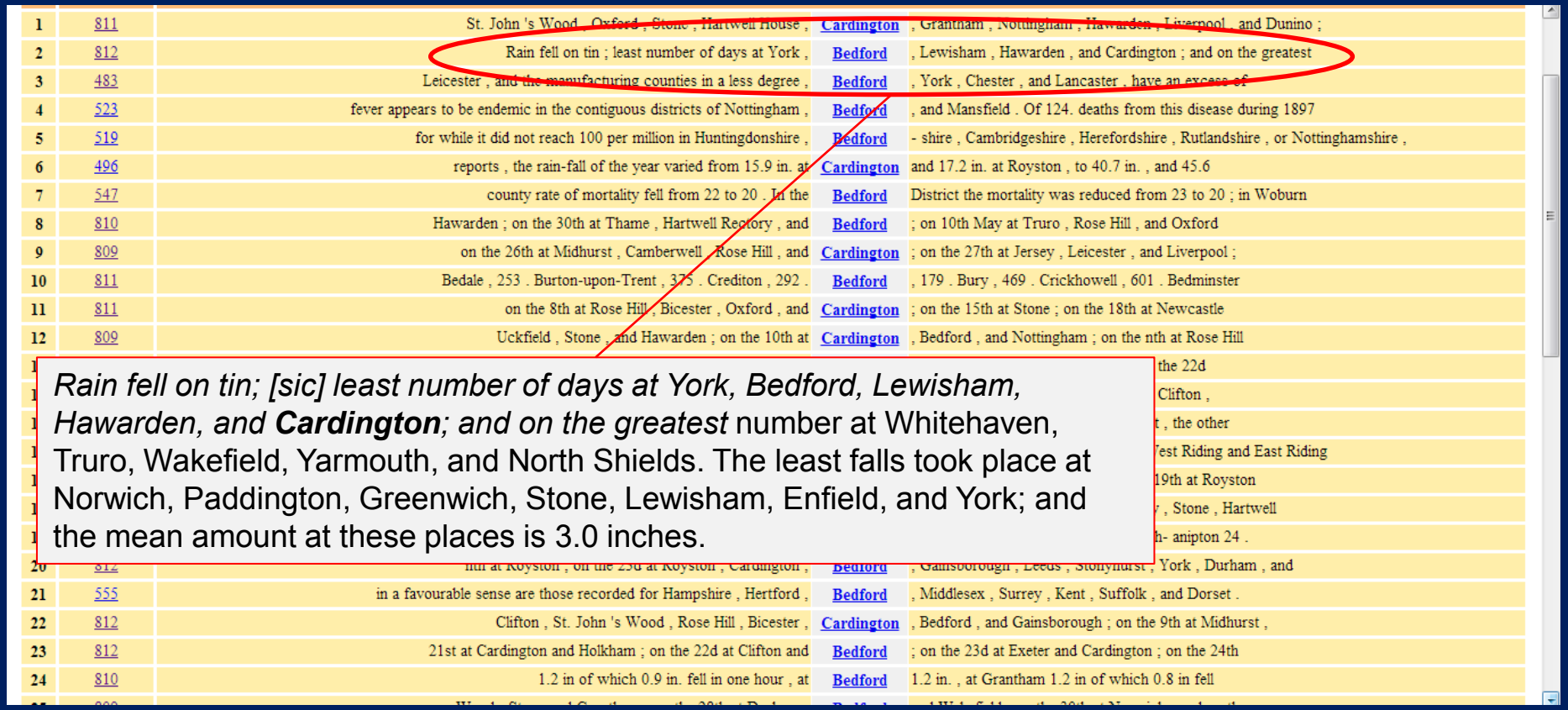
Place-names instances, 1851-1910

17,667 instances classed as “populated places”

1. Tyne & Wear (564)
2. Durham (567)
3. Liverpool (533)
4. Manchester (639)
5. Leeds (226)
6. Nottingham (308)
7. West Bromwich (50)
8. Bedford (299)
9. Bristol (123)
10. London (3,274)

What are they saying?

- ‘Bedford’ cluster consists of 3 different place-names (Bedford, Cardington, Kempston)
- Concordance using Key Word in Context in random order



1	811	St. John 's Wood , Oxford , Stone , Hartwell House ,	Cardington	, Grantham , Nottingham , Hawarden , Liverpool , and Dunino ;
2	812	Rain fell on tin ; least number of days at York ,	Bedford	, Lewisham , Hawarden , and Cardington ; and on the greatest
3	483	Leicester , and the manufacturing counties in a less degree ,	Bedford	, York , Chester , and Lancaster , have an excess of
4	523	fever appears to be endemic in the contiguous districts of Nottingham ,	Bedford	, and Mansfield . Of 124. deaths from this disease during 1897
5	519	for while it did not reach 100 per million in Huntingdonshire ,	Bedford	- shire , Cambridgeshire , Herefordshire , Rutlandshire , or Nottinghamshire ,
6	496	reports , the rain-fall of the year varied from 15.9 in. at	Cardington	and 17.2 in. at Royston , to 40.7 in. , and 45.6
7	547	county rate of mortality fell from 22 to 20 . In the	Bedford	District the mortality was reduced from 23 to 20 ; in Woburn
8	810	Hawarden ; on the 30th at Thame , Hartwell Rectory , and	Bedford	; on 10th May at Truro , Rose Hill , and Oxford
9	809	on the 26th at Midhurst , Camberwell , Rose Hill , and	Cardington	; on the 27th at Jersey , Leicester , and Liverpool ;
10	811	Bedale , 253 . Burton-upon-Trent , 375 . Crediton , 292 .	Bedford	, 179 . Bury , 469 . Crickhowell , 601 . Bedminster
11	811	on the 8th at Rose Hill , Bicester , Oxford , and	Cardington	; on the 15th at Stone ; on the 18th at Newcastle
12	809	Uckfield , Stone , and Hawarden ; on the 10th at	Cardington	, Bedford , and Nottingham ; on the nth at Rose Hill
				the 22d
				Clifton ,
				t , the other
				Vest Riding and East Riding
				19th at Royston
				, Stone , Hartwell
				h- anipton 24 .
20	812	nth at Royston , on the 23d at Royston , Cardington ,	Bedford	, Gainsborough , Leeds , Stonyhurst , York , Durham , and
21	555	in a favourable sense are those recorded for Hampshire , Hertford ,	Bedford	, Middlesex , Surrey , Kent , Suffolk , and Dorset .
22	812	Clifton , St. John 's Wood , Rose Hill , Bicester ,	Cardington	, Bedford , and Gainsborough ; on the 9th at Midhurst ,
23	812	21st at Cardington and Holkham ; on the 22d at Clifton and	Bedford	; on the 23d at Exeter and Cardington ; on the 24th
24	810	1.2 in of which 0.9 in. fell in one hour , at	Bedford	1.2 in. , at Grantham 1.2 in of which 0.8 in fell

Rain fell on tin; [sic] least number of days at York, Bedford, Lewisham, Hawarden, and ***Cardington****; and on the greatest* number at Whitehaven, Truro, Wakefield, Yarmouth, and North Shields. The least falls took place at Norwich, Paddington, Greenwich, Stone, Lewisham, Enfield, and York; and the mean amount at these places is 3.0 inches.

Collocation approaches to the 'Bedford' cluster

- Statistics based on:
  - Number of times each word occurs near the search place-name(s)
  - Compare to expected based on frequency of the word in the corpus as a whole
  - Summarise using stats such as z-scores, log-likelihood, or mutual information
  - In this case z-scores on a minimum of three collocates

- Lists:
  - 71 of the top 100 z-scores are placenames,
  - 15 of the remainder are numeric
  - *, ; ( ) sub-district* and *sub-districts* are also there
- Others include:
  - feet (z=12.42, 78th) – height at which barometric readings are taken
  - Observatory (10.64, 90th),
  - inches (10.20, 92nd) – used in barometric readings
  - least (9.23, 103rd) – usually in terms of temperature, rainfall, etc.
  - amount (8.77, 107th) – again in terms of readings
  - falls (7.24, 119th) – falls of rain/snow

- The observatory/weather station at Cardington means that the place is named far often than might be expected in lists and in relation to weather readings

Collocation approaches to 'London'

- 115 unique place-names
- Lists:
  - 71 of the top 100 z-scores are to place-names and comma is 16th (z=37.34)
- Other high z-scores are related to water quality/supply:
  - water (z=44.81, 6th)
  - company (48.16 , 5th) & companies (36.01 , 20th)
  - supplied (36.40, 18th) & supply (29.45, 37th)
  - cholera (30.77, 33rd)
  - waterworks (27.92, 43rd) & reservoirs (18.41, 100th)
  - sewage (14.49, 143rd).
- Other unusually frequent words:
  - hospital (27.10, 48th)
  - health (21.38, 77th)
  - bankers (21.94, 73rd)
  - parliament (20.60, 83rd)
  - asylum (17.67, 105th)
  - streets (17.47, 106th)
  - Saturday (16.22, 119th)

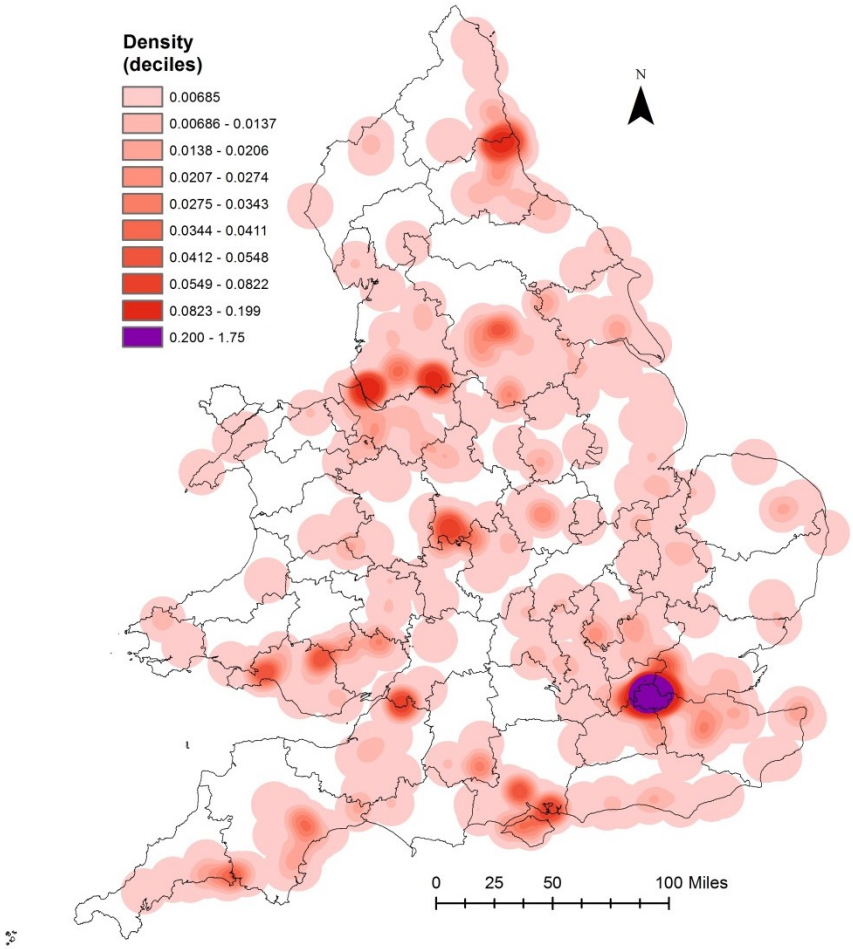
Other clusters

- Liverpool & Manchester
  - 37 place-names
  - Lists:
    - 76/100 placenames. 17 numbers; plus , ; *township*, *sub-district* – all associated with lists as well.
  - Remainder of top 100 only have small nos :
    - Bayliss (Dr. Bayliss – 4), depot (emigration depot – 3) and adverting (3)
  - Beyond this:
    - diarrhoea (112th , z=13.85), towns (123rd 13.12), cholera (184th 9.36), declining (188th 9,23), prevalent (207th 8.69), lung (as in 'disease', 234th 7.85), fatally (235th 7.79).
  - Much more descriptive of diseases - no discourse on water supply.
    - Water has a z-score of -3.171, supply z=-1.099, supplied z=-0.489, sewage z=-0.248, company and companies n<3
    - "Diarrhoea is reported as having been prevalent and fatal in Manchester , **Salford** , Liverpool , Sheffield , York , Yarmouth , Norwich , and many parts both of town and country ." (24th Annual Report, 1861)

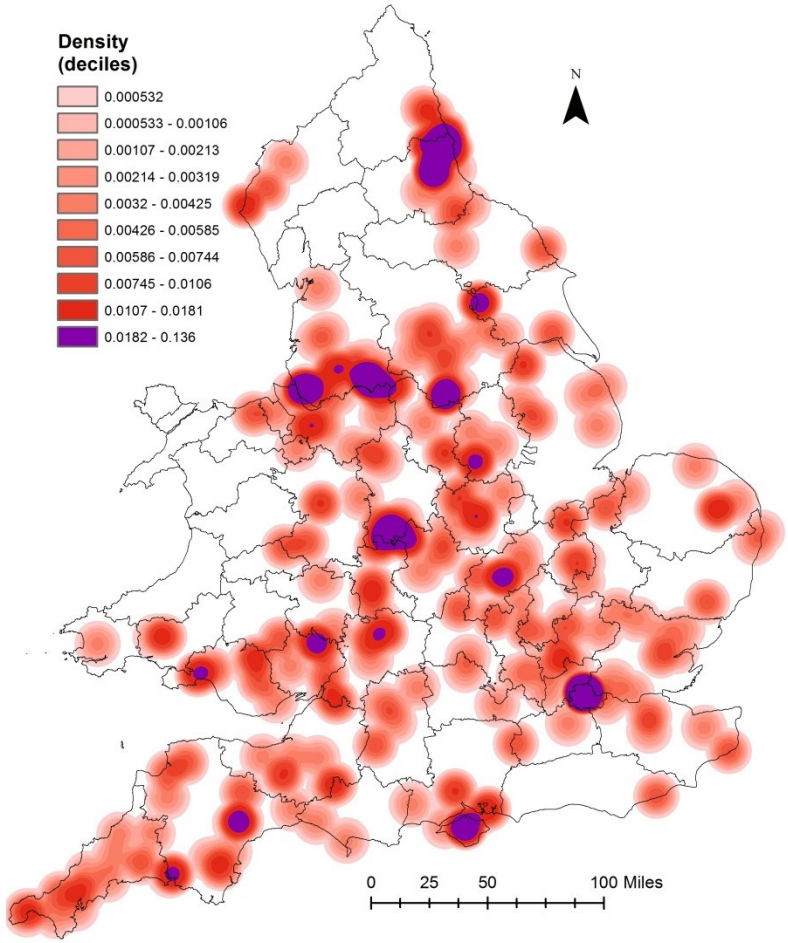
Mapping collocations

Place-names that occur in the same sentences as:

Cholera, diarrhoea, dysentery

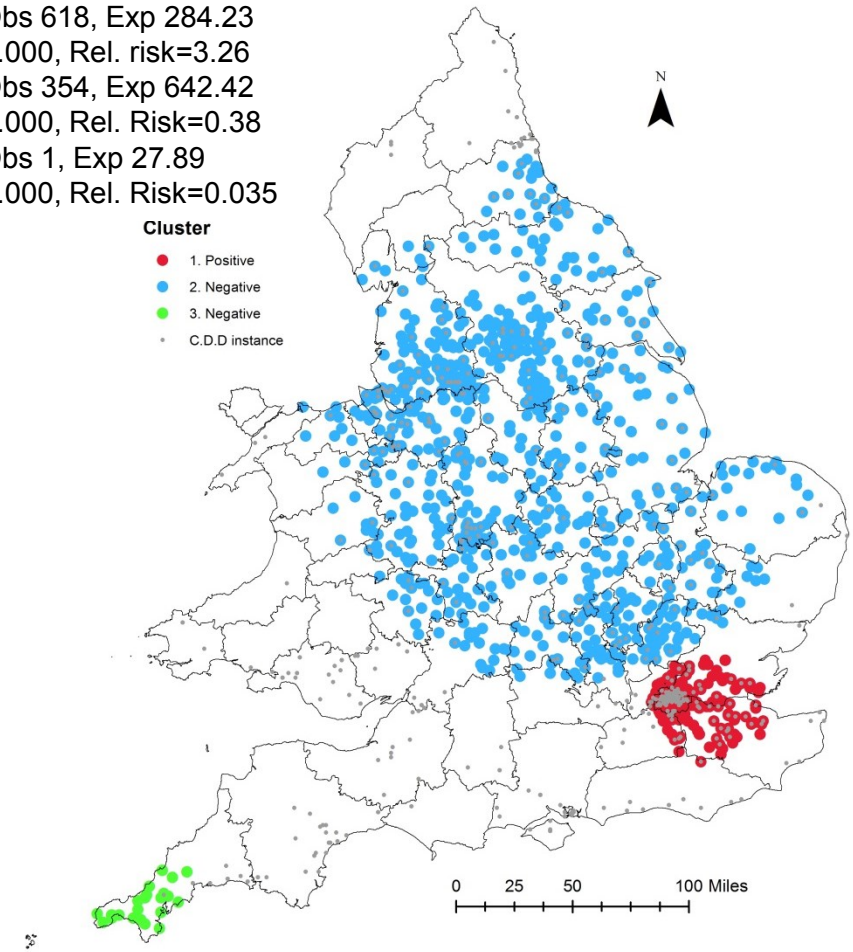


Small-pox



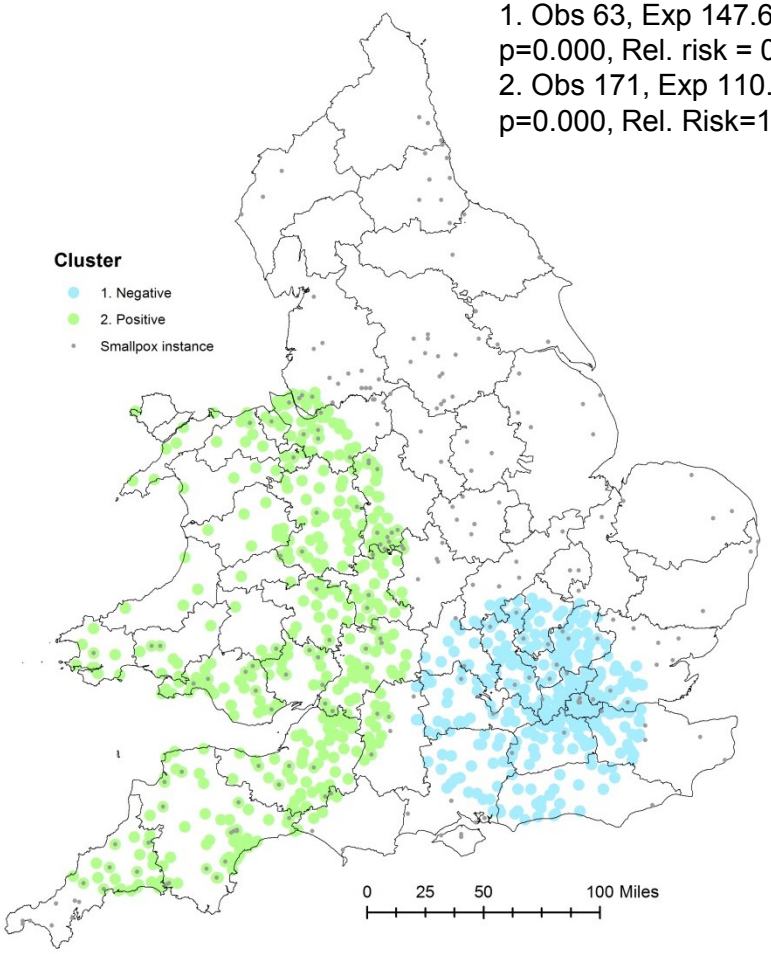
Kulldorf Scan Statistic

1. Obs 618, Exp 284.23
p=0.000, Rel. risk=3.26
2. Obs 354, Exp 642.42
p=0.000, Rel. Risk=0.38
3. Obs 1, Exp 27.89
p=0.000, Rel. Risk=0.035



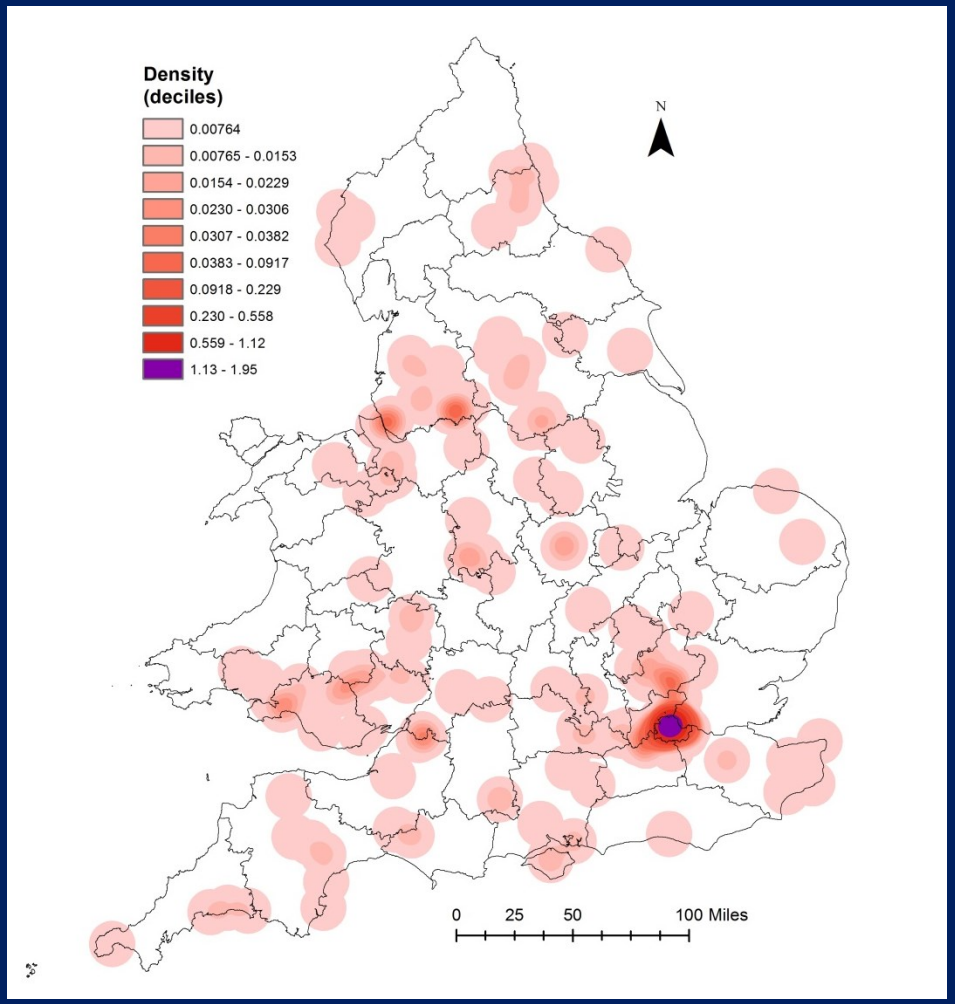
Cholera, diarrhoea, dysentery

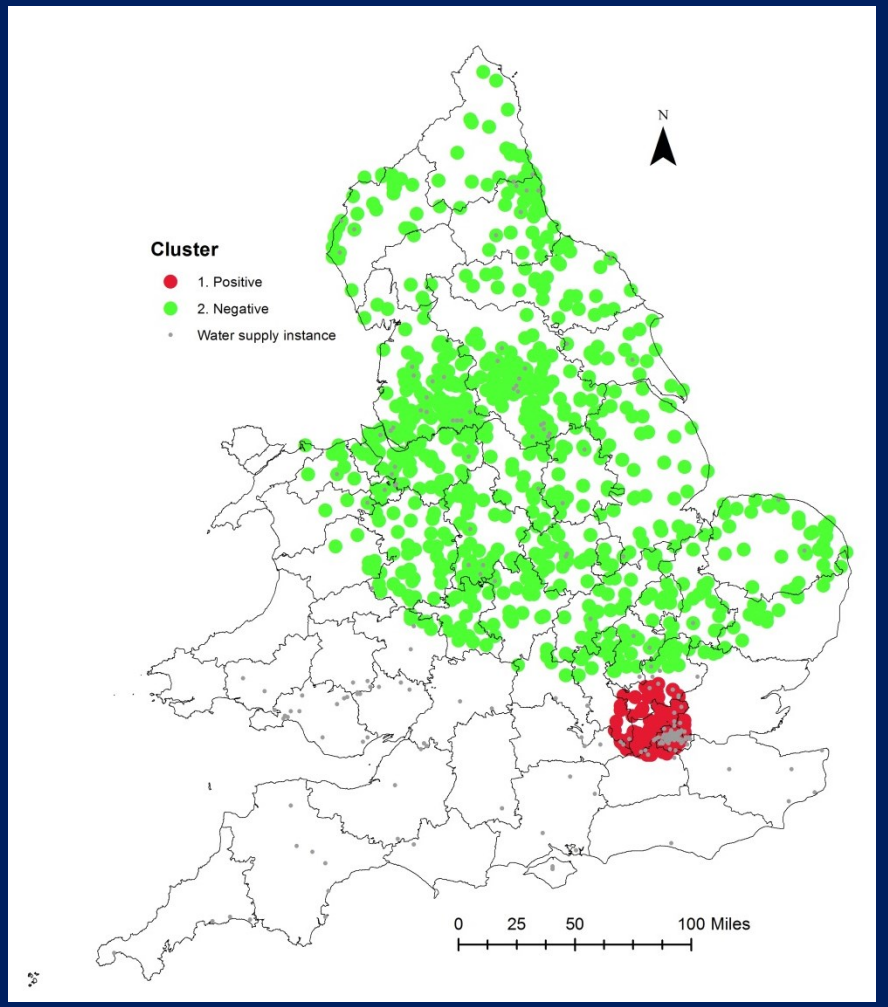
1. Obs 63, Exp 147.64
p=0.000, Rel. risk = 0.34
2. Obs 171, Exp 110.94
p=0.000, Rel. Risk=1.84



Small-pox

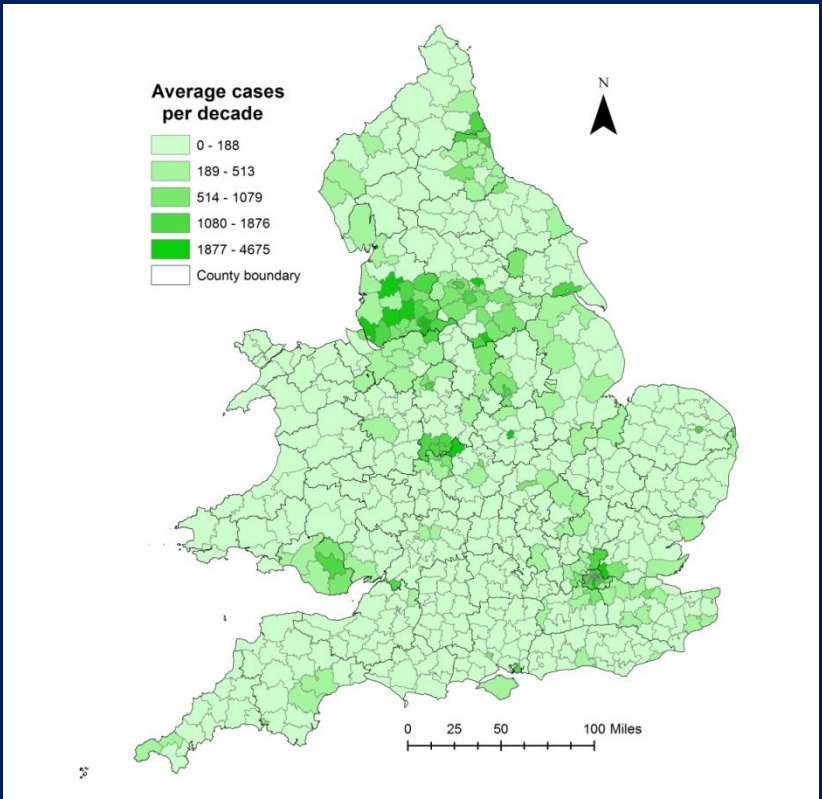
Water Supply





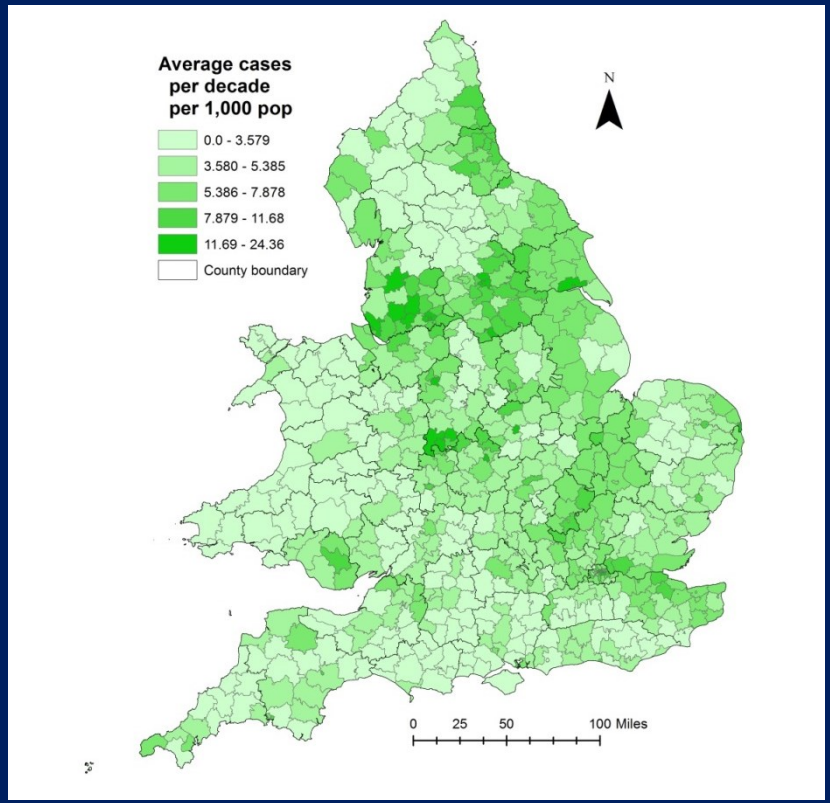
Supply, supplied, company, companies, sewage, reservoirs and waterworks – 860 collocations with place-names

Cholera, diarrhoea, dysentery 1850s-1900s



Deaths

Highest 10: Liverpool, West Derby, Birmingham, Sheffield, Chorlton, Salford, Leeds, Aston, West Ham, Leicester (Highest London: *Southwark*,14th)



Death rate

Highest 10: Liverpool, Preston, Manchester, Salford, Leicester, Birmingham, Prestwich, *St. Olave Bermondsey*, *Whitechapel*, and *St. George in the East*

Variations in cities

Place	Reg Districts	Instances	Deaths (per decade)	Death Rate (per 1000)	Instance Rate (per 1,000 deaths)
London	LCC area	512	33,636	9.09	15.22
London	Cluster	551	33,788	8.77	16.31
Liverpool	L'pool, W. Derby, Toxteth Pk	35	10,242	14.69	2.44
Manchester	Mcr, Chorlton, Prestwich, Salford	25	10,567	13.95	2.37
Birmingham	B'ham, Aston	8	6,387	13.75	1.25
Newcastle	N'cstle on Tyne	14	1,876	10.49	7.46
Gateshead	Gateshead	24	1,300	10.34	18.46
Southampton	Soton	13	621	7.28	20.93
Portsmouth	Portsmouth	14	1,131	7.22	12.38
Bristol	Bristol	13	1,563	5.53	8.31
Leeds	Leeds	7	2,867	14.19	2.44
Sheffield	Sheffield	6	3,162	14.80	1.90
Preston	Preston	2	2,370	17.50	0.84
Bolton	Bolton	2	2,443	11.81	0.82
Blackburn	Blackburn	0	1,875	10.19	0
Burnley	Burnley	1	1,321	9.08	0.76
Chester	Chester	7	260	4.61	30.77(26.90)
England & Wales		**1,278**	**206,552**	**7.53**	**8.31**

Collocation analysis

- Words that collocate with sentences that contain place-names and cholera/diarrhoea/dysentery
- London:
  - Water related: Water (n=154, z=32.755); water-fields (11, 31.34); supplied (59, 26.49); companies (34, 23.75); waterfields (7, z=22.39); elevation (21, 22.31); company (46, 22.00); supplying (9, 14.51); elevations (6, 12.66); waters (17, 10.88); supply (32, 10.86); impure (11, 10.60); ditches (4, 8.90); pipes (6, 8.14); matter [organic or cholera] (15, 7.04); waterworks (5, 6.723); etc
  - Research related: Map (n=8, z=14.98); extract [from report] (5, 12.16); circular [a circular] (4, 10.09); exhibiting (4, 9.44); Professor (6, 7.54); diagrams (3, 6.98); Dr. (17, 5.73); report (32, 5.67); return [a return] (10, 3.81)
  - Descriptive: Epidemic (n=56; z=18.47); outbreak (19, 15.70); infected (5, 6.56); epidemics (5, 3.58)
- Liverpool, Manchester, Birmingham:
  - Descriptive: Prevailed, epidemic, occurred, deaths, prevalent, mortality
  - Eg. “In Liverpool more children of this age died from lung diseases than from diarrhoea or diseases of the brain, and the high rate of mortality there was mainly due to those three causes…” 37th Annual Report, 1874
  - Eg. “If English towns are selected for comparison it will be seen that the borough of Liverpool was the most unhealthy in 1866; for by a malignant fever in winter and cholera in summer, the mortality of the year was raised to 4.185, while that of Manchester was 3.195”

Collocation analysis

- Gateshead:
  - Certified (n=4, z=8.99); hours (3, 8.71); aged (4, 8.58)
  - *Report on the cholera epidemic of 1866 in England: Supplement to the twenty-ninth annual report of the Registrar General* (1868) "First Cases of Epidemic Cholera in Newcastle-upon- Tyne and Gateshead in 1853"
    - "1st September, Bill Quay, Heworth, Gateshead Elizabeth Handy, aged 40 years, wife of coke-burner, cholera, 12 hours"
- Southampton, Portsmouth, Bristol:
  - Cholerafield (n=4, z=17.01); Dr. (4, 6.59); Health [board of] (4, 4.64); July (3, 3.99); Medical [journal, officer, board] (3, 3.04); report (3, 2.40)
- Leeds, Sheffield:
  - Prevalent (n=3, z-5.90); epidemic (3, 4.00), occurred (3.33)
- Blackburn, Bolton, Preston, Burnley:
  - Nothing of interest
- Chester:
  - August (n=3, z=8.44); water (3, 5.00)

Conclusions

- Registrar General's reports:
  - Shows:
    - Strong discourse on cholera/diarrhoea/dysentery and its relationship to the water supply in London
    - Relative lack of interest in these diseases elsewhere
    - Very little discourse on causes of these diseases and measures to reduce them outside London
    - Biases are not justified by deaths. Northern industrial towns particularly ignored.

Conclusions

- Three approaches
  - Exploratory spatial
  - Thematic
  - Textual/Statistical comparisons
- Implications:
  - Geo-parsed corpora are noisy but the implications are relatively easy to understand
    - Geo-parsing will get better
  - Ability to use statistical summaries, macro-readings and close reading together to understand an issue
    - Analyse abstractions of millions (plus) of words of text
      - Complete accountability
    - Decide which parts to close read
    - Ability to use texts to help explain quantitative results

Where next?

- Further research
  - Corpus of Lake District writing to 1900
    - Over 80 texts
    - 1,000,000 words
  - More historical sources
    - British Library Nineteenth Century Newspapers Corpus
    - 2 million pages of newspapers
  - Bring quantitative sources in as well

- http://www.lancs.ac.uk/spatialhum