

# A Novel Use of Correspondence and Cluster Analysis to Study and Visualize High-Dimensional Occupations Data

Leigh Shaw TAYLOR

Murat GUVENC

Tobias LUNDE

- Introduction \*\*\*\*
- Definition and advantages of correspondence analysis (CA)
- Methodological drawbacks of correspondence analysis
- *Complementary use of correspondence and cluster Analyses: a summary* Lebart's procedure
- Adapting Lebart's procedure to disaggregated qualitative data
- Interpreting the results of the user interface, legend tables and templates for abbreviated legends
- An applied example: mapping occupational structure of women in 1851 in England and Wales at the parish level
- Conclusions

## Appendices

- Appendix 1 Maps Series for Men's Occupations in 1851 and in England and Wales
- Appendix 2 Maps Series for Men's and Women's Occupations in 1851 and in England and Wales
- Appendix 3 Legend Tables for the Women's Occupations in 1851 and in England and Wales
- Appendix 4 Legend Tables for the Men's's Occupations in 1851 and in England and Wales
- Appendix 5 Legend Tables for the Women's and Men's Occupations in 1851 and in England and Wales:

Correspondence Analysis (CA) is defined as :

‘An exploratory statistical technique for describing, summarizing and visualizing data matrices, typically consisting of frequencies or count data’

A method of *geometric data analysis (GDA)*, where data is interpreted as a cloud of points in a multidimensional, Euclidian space, with each point representing individual observations.

It provides a systematic method based on mathematical structures that allows for an investigation of the magnitude and substantive nature of associations between observations and attributes, such as

- *extracting underlying, “hidden” relationships and dimensions of data tables*
- *Visualising the the relationships and relative position of observations and attributes within the cloud of data points, typically using diagrams called correspondence maps.*

Unlike the more common multivariate statistical approaches, it is *not*

- limited to numerical data or estimates
- based on the notion of a sample from a wider population
- reliant on *a priori* assumptions about the data, allowing analysis to remain “impervious to the expectations and prejudices of researchers” (Jean-Paul Benzécri, 1973)

While GDA is often presented in opposition to multivariate techniques, they can be used in conjunction with each other

Can be derived in different, mathematically equivalent ways (see)

It is conceptually analogous to Principle Component Analysis, but for count and frequency data instead of continuous data

A detailed discussion is beyond the scope of this presentation. See for example Karl M. Van Meter et al. Correspondence Analysis : A History and French Sociological Perspective, Correspondence Analysis in the Social Sciences pp. 128-138; Blasius and Greenacre, 2014, *Vizualisation and Verbalization of Data*; Le Roux and Rouanet, 2004, Geometric Data Analysis.

## The method of CA

- The starting point of CA is a cross-tabulation of data, where each row represents an observation and each column a variable
- The objective of CA is to project the cross-table onto a lower-dimensional space while retaining the maximum possible variation in the original data
- In this lower-dimensional space, observations and variables are represented by principal coordinates, which describe how far from the centre of the cloud of points each observation and variable is
- The first principal coordinates captures the most variation in the original data, the second coordinate the second most, and so on
- The projection is based on relative distances from the centre of the cloud of points, so that rows or columns with relatively small or large totals are not disproportionally weighted in the analysis. Without this step, an analysis of the below table would not place much emphasis on people above 75 years, as relatively few of those surveyed were that age
- These relative positions of rows and columns are often referred to as *profiles*

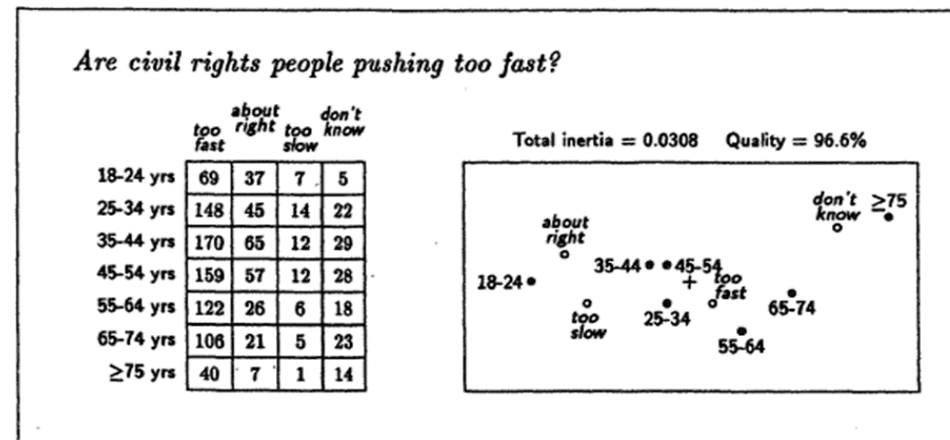


FIGURE 1.1 Example 1: A 7 × 4 table of frequencies from a social attitudes survey together with the two-dimensional symmetric CA map and the amount of inertia accounted for by the map.

## The CA as a tool for pattern recognition

In many applications, the first two principle coordinates captures the vast majority of variation and can be plotted on a so-called correspondence map (in the below case, 96.6% of the variation in the table is captured by the diagram), where each row/observation and column/variable is represented by a point

The position of row and column points on this map depend upon the relative “shape” of row and column profiles.

- Points that are close together resemble each other, either similar observations (e.g. people that are similar to each other) or variables (e.g. average air humidity and average rainfall in a region)
- Points further away from the centre represent outliers, and those close to it represent “average cases” or variables that do not vary systematically by observation
- In the below example, young people are more likely to think that civil rights people are pushing “about right” or “too slow”, while people above 75 tend to not know what to think.
- Middle aged people are not distinguished in any particular direction in terms of civil rights views

In other words, CA establishes the **correspondence** between the **shapes** of profiles and their **position** on the projection plane

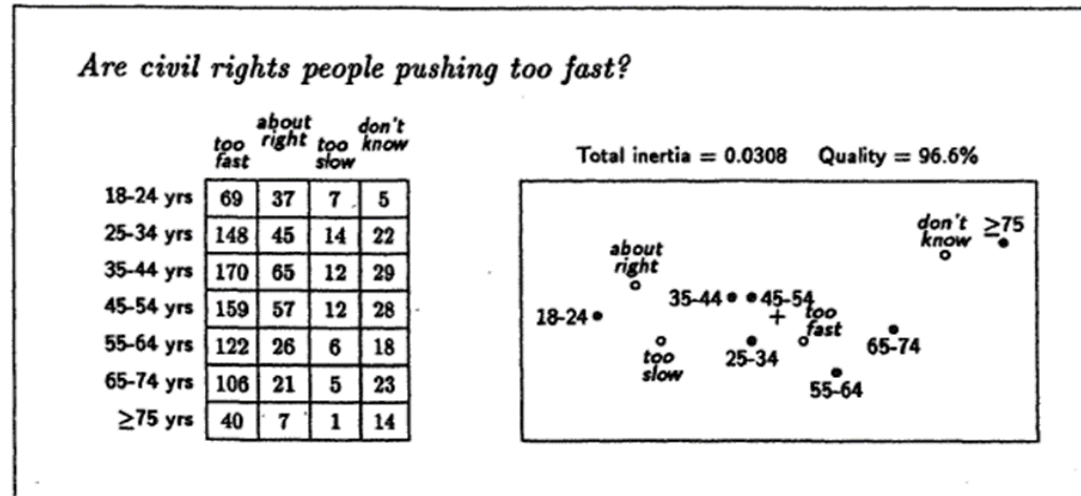


FIGURE 1.1 Example 1: A  $7 \times 4$  table of frequencies from a social attitudes survey together with the two-dimensional symmetric CA map and the amount of inertia accounted for by the map.

*How is the UN doing in solving the problems it has had to face?*

	very good	good	poor	very poor	don't know
W. Germany	2	46	21	6	26
Britain	7	39	28	9	17
France	2	45	22	3	28
Japan	1	11	43	5	41
USA	5	46	27	13	10

Total inertia = 0.1463    Quality = 96.9%

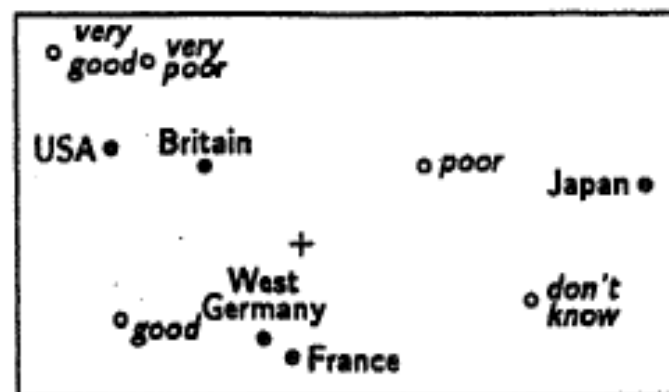


FIGURE 1.3 Example 3: A 5 x 5 table of percentages (rows sum to 100%) from a series of surveys in five different countries, together with the two-dimensional symmetric CA map and the amount of inertia accounted for by the map.

## Examples of research questions CA allows for

- Systematic descriptions and visualisations of large, complex sets of count and categorical data
- Shedding light upon latent associations or connections in data tables that are too large for humans to interpret or visualise, given the limits of our cognition
- identifying the typical and distinctive observations and variables
- developing new concepts and models (such as «*Cultural Capital*» or «*Political Capital*» coined by Bourdieu)
- formulating original research questions and guide model selection in more traditional, multivariate statistical techniques
- designing relevant case studies
- monitoring and visualizing social and economical change in time

## Limitations of CA applied to *large* data sets

While CA has proven a useful tool to study data tables in many fields, they have typically been of limited size.

We would like to study the occupational structure of England and Wales in 1851 at the parish level, using micro-level census data. Counting the number of people working in each occupational category at PST level 3 in each parish, we get a table of more than 3 million cells.

This creates two problems for a correspondence analysis of the table:

- The results and correspondence map is much too complex to be easily read – there are approximately 13,000 points to plot on the diagram (the total number of rows and columns) While some of these could be excluded, that would also exclude potentially valuable information, and the exclusion criteria would be determined *ad hoc*
- CA is highly sensitive to outliers. «remote profile point(s) can notably influence, for example, the first principal axis, and therefore all the subsequent dimensions, since these dimensions are related to the first axis through the constraints of orthogonality of axes.” (\*) This is particularly true for large tables, where a small number of outliers can obfuscate much of the variation between the other variables

This do not affect the representation of *distinctive* observations, but hinders the representation of nuances amongst *typical* cases

The problem is much more important if the dataset comprises – as it is the case with our data – a mixture of distinctive and diversified cases and/attributes. As a consequence of these two drawbacks CA is not a particularly useful tool to analyze our case of highly disaggregated data at high levels of spatial resolution.

(\*) Lebart L. 1994 Complementary Use of Correspondence Analysis and Cluster Analysis

## Lebart's Procedure for Complementary use of Correspondence and Cluster Analysis

The reliability of Correspondence Analysis and the legibility of Correspondence Maps depend upon:

- the level resolution (as far as cases/observations are concerned)
- the level of disaggregation (as far as attributes/variables are concerned)

It follows that CA is useful in analysing aggregated data at low levels of resolution, and particularly inappropriate for large disaggregated data sets at high resolutions. Conventional CA is therefore of limited value where disaggregated data sets are analysed

A novel procedure proposed by L. Lebart (\*) solves these issues by

- reducing the sensitivity of CA to outliers, while retaining the ability to distinguish between typical observations
- enhancing the interpretability of results by producing results that are readable by humans and possible to map

(\*) Lebart L., 1994, Complementary Use of Correspondence Analysis and Cluster Analysis.

## Outline of Lebart's Procedure

1. Perform CA on the large data matrix
2. Perform cluster analysis on the lower-dimensional projection produced by the CA, assigning each observation to a cluster
  - Any type of cluster analyses may be used depending on the application
3. Aggregate the observations by adding together the observations in the same cluster (column sum by cluster membership)
4. Perform a CA on this new “table of clusters”, where each row represents one or more of the original observations
  - Because CA is performed using relative distances from the cloud centre, adding together observations does not mean these affect the CA disproportionately
5. Perform a cluster analysis on this new projection, assigning each cluster to a cluster of clusters
6. Repeat until a desired number of clusters is reached
  - For example, our original data consists of 12,600 observations. The first step produces around 5,000 clusters of observations, the second 2,000 clusters of clusters, the third 800 clusters of clusters of clusters, and so on

## Lebart's Procedure for Complementary use of Correspondence and Cluster Analysis

Referring to the «exact relationship between the **principal axes of CA**, and the **nodes of a cluster analysis**»

Lebart emphasizes the «Complementarity between Correspondence and Cluster analysis»

- 'The two clusters which are **merged last** by the cluster analysis determine the **strongest contrast in CA**, loading on opposite sides of the first principal axis»
- As opposed to Correspondence analysis, "classification algorithms and particularly agglomerative algorithms, are locally robust in the sense that the lower parts of the produced dendrograms **are largely independent of outliers.**"
- It is much easier to describe sets of clusters than a continuous space
- For example, in our case, we can now describe the large occupational table by referring to clusters of parishes instead of 12,600 individual parishes. Each of these clusters may be characterised by their distinctive occupational characteristics, for example one cluster of parishes may have a large number of textile workers and a small number of farmers relative to parishes outside of this cluster

Greenacre, M and Blasius, J, 'Preface', in *Correspondence Analysis in the Social Sciences*, M. Greenacre, and J. Blasius eds Academic Press p. xii.

Lebart, L., 'Complementary Use of Correspondence Analysis and Cluster Analysis' in M. Greenacre, and J. Blasius eds pp. *Correspondence Analysis in the Social Sciences* p.163.

## The role of the researcher in Lebart's procedure

- The number of principal dimensions the cluster analyses are performed on depends on the researcher, and can for example be done with a criteria that selects the number of dimensions necessary to capture 95% of the variation in the original table
- The number of clusters at each iteration of the is chosen by the researcher
  - While this can be done algorithmically, selecting the optimal number of clusters is a difficult problem mathematically
  - We have chosen to use a combination of the “elbow method”, and choosing a number of clusters that retain at least 95% of the variation in the pre-clustered data.
- The final number of clusters is decided by the researcher depending on the application, but at each step the cluster assignments can be saved, allowing for a hierarchical display of cluster memberships.

## Applying Lebart's Procedure to study the Occupational Structure of England and Wales in 1851

The exploratory analysis addresses the following four research questions

1. What kind of organization or *affinities* can be detected among the occupation profiles of parishes in England and Wales in 1851?
2. Is it possible to build a meaningful **typology of parishes** in England and Wales with respect to their occupation profiles?
3. What kind of organization or affinities can be detected among occupations with respect to ways in which they are deployed across England and Wales in 1851?
4. Is it possible to build a meaningful **typology of occupations** with respect to their deployment pattern accross England and Wales? This typology can summarize and shed light upon distinctive features of the economic division of labour in England and Wales in mid- nineteen century.

## Applying Lebart's Procedure to study the Occupational Structure of England and Wales in 1851

The original data set comprises some 12600 parishes and approximately 290 occupation categories at PST level 3. It has been transformed into a reduced table of 12 clusters of parishes and 290 occupations.

Clusters and indicators derived from 1851 Women's Occupations data are produced in Appendix 3. This second parish cluster (Cluster N0:11) would be useful in illustrating the above presented indicators. (Cluster N0:11) comprises

- (206) parishes (1.6 % of all parishes) and
- accounts for 7,3% of Women's Employment 1851 (422231 operatives)

For simplicity, only over-represented Secondary occupations are taken up. (see next slide)

Notice that Occupations are ranked in the descending order of their signed-chi Indices. The signed chi indices gives a numerical answer to how over *or* underrepresented each occupation is in a given cluster of parishes,

- Cells with **distinctively high** Signed-Chi Indices are shaded **blue**, indicating a relatively **high** presence of that occupation in the cluster of parishes
- Cells with **moderately high** Signed-Chi Indices are **yellow**, indicating a relatively **low** presence of that occupation in the cluster of parishes
- Cells with Signed-Chi Indices just above zero are shaded in light red.

In the example on the next slide, Cotton Manufacturing, Textiles (PST code 220040) and in Secondary Occupations (PST code 220000) have **distinctively high** Signed-Chi Indices and accounts for 76,5 % of England and Wales's Women employment in Cotton Manufacturing and 35,9 % in Textiles

	Women's Secondary Occupations : Legend Category number 11																												
	220040 S. TXTs, Cotton man	220000 S. TXTs	200000 SECONDARY.	210000 S. Clothing	230010 S. Ind. using leather, bone etc., Leatherind	235990 S. Furnishing, Furnishing, other	255070 S. Chemical, soap, adhesives, man, Match man	265040 S. Machines & tools, Machine operation	265010 S. Machines & tools, Machine making	225050 S. Wood ind, Cork man	240010 S. Paper ind, Paper making (vellum &parchment)	265031 S. Machines & tools, Engineering	255990 S. Chemicals,Chemical man, other	280060 S. B&C, Plasterer, painter, decorator	265030 S. Machines & tools, m&operation, Engineering	225080 S. Wood ind, Woodworking	255020 S. Chemical, man, Colour industries	241000 S. Printing	258030 S. Fuelind, Electricity generation and supply	280030 S. B&C, Masonry	201060 S. Food ind, Minor Food ind	240020 S. Paper ind, Paper pr	258990 S. Fuelind, Fuelind, other	201040 S. Food ind, Meat, fish, poultry pr	270020 S. RT Vehicles, Cart building	266000 S. Electrical goods	258010 S. Fuelind, Coal gas manufacturing	276020 S. Stone & mineral processing Slate	
S_CHI	804	313	158	31	26	25	23	23	19	10	8	7	7	6	5	5	5	4	4	4	3	3	3	2	2	2	2	1	
COL%	76,5	35,5	40,9	22,8	18,2	41,7	34,4	52,8	17,1	19,6	9,5	28,4	14,4	9,9	12,6	11,1	24,6	8,7	17,0	11,6	12,7	9,5	16,7	8,5	19,2	9,1	10,5	16,7	
ROW %	17,7	7,5	1,6	0,2	0,2	0,0	0,0	0,0	0,1	0,0	0,2	0,0	0,0	0,1	0,0	0,0	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,1	0,0	0,0	0,0	0,0	28,0
OVERALL%	1,7	1,5	0,3	0,1	0,1	0,0	0,0	0,0	0,1	0,0	0,2	0,0	0,0	0,1	0,0	0,0	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,1	0,0	0,0	0,0	0,0	4,2
CONT_R	10,5	4,9	5,6	3,1	2,5	5,7	4,8	7,3	2,4	2,7	1,3	3,9	2,0	1,4	1,7	1,5	3,4	1,2	2,4	1,6	1,8	1,3	2,3	1,2	2,7	1,3	1,5	2,3	6,7
In Numbers	31541	6587	633	718	169	169	84	464	84	844	42	84	338	84	169	3800	549	42	84	42	169	42	253	1816	84	42	169	127	49232

Source: Derived from CAMPOP 1851 data set on Women's Occupations (at PST3 level and by Parishes)

## The second set comprises the following occupations

- Clothing Industries
- Leather industries
- Furnishing
- Match Manufacturing
- Machine Operation

with **moderately high** S\_CHI indices (shaded yellow in slide No.19)

Entries along COL% suggest relatively high shares from total employment (i.e. That of England and Wales) (\*)

- 22,8 % of Women in Clothing Industries (718 Operatives)
- 18,2 % of Women in Leather industries (169 Operatives)
- 41,7 % of Women in Furnishing (169 Operatives)
- 34,4 % of Women in Match Manufacturing ( 84 Operatives)
- 52,8 % of Women in Machine Operation (464 Operatives)

Yet they and account for minor shares (\*\*) in the local employment profile.

Location quotients (Listed in in **CONT\_R**) are relatively high

(\*) Recall that this cluster accounts for 7.3 % of total Women's employment in England and Wales

(\*\*) although higher than national shares listed in **OVERALL %**

The same holds true for the following occupations with relatively low concentrations and low shares in the local employment profile.

1.	17,1 % of Women in Machine making	(84)
2.	19,6 % of Women in Cork Manufacturing	(844)
3.	9,5 % of Women in Paper (incl Vellum & Parchment)	(42)
4.	28,4 % of Women in Machines and Tools Engineering	(84)
5.	14,4 % of Women in Chemicals	(338)
6.	9,9 % of Women Plasterers & Decorators	(84)
7.	12,6 % of Women in Machine Operarating Engineering	(169)
8.	11,1 % of Women in Woodworking	(3800)
9.	24,1 % of Women in Chemical Man. Color ind.	(549)
10.	8,7 % of Women in Printing	(42)
11.	17,7 % of Women in Fuel industries	(84)
12.	11,6 % of Women in Building & Const. Masonry	(42)
13.	12,7 % of Women in Minor Food ind.	(169)
14.	9,5 % of Women in Paper	(42)
15.	16,7 % of Women in Other Fuel ind.	(253)
16.	8,5 % of Women in Meat Fish, Poultry Products	(1816)
17.	19,2 % of Women in Cart Building	(84)
18.	9,1 % of Women in Electrical Goods	(42)
19.	10,5 % of Women in Coal gas Manufacturing	(169)
20.	16,7 % of Women in Stone Processing: Slate	(127)

Notice that:

- the employment in secondary occupations is 2.3 times **higher than** that of England and Wales.
- The employment share of over-represented occupations (28.0 %) is approximately 7 times higher their share in England and Wales (4.2 %) (see the last column in slide 22)

Notice also that **within cluster differences in labour specialization** are accurately detected:

- Total employment in Cotton Manufacturing and TEXTILES is **17 times** higher than the total employment in the six occupations with **intermediate levels of specialization**
- Total employment in Cotton Manufacturing and TEXTILES with **highest level of concentration** is **4,3** times higher than the total employment in 20 occupations showing **low levels of specialization**.
- Finally total employment in Cotton Manufacturing and TEXTILES is **3,4** times higher than the total employment in 26 occupations with **intermediate or low levels of specialization**

As expected, the employment shares of remaining sectors are significantly different than those observed in England and Wales (see below)

In fact:

- The employment share of Services (14.8 %) is significantly **less than** that of England and Wales (19,9 %),
- The employment share of Agriculture (1,39 %) is significantly **less than** that of England and Wales (3,17 %),
- The employment share of Mining & Quarrying (0,08 %) is significantly **less than** that of England and Wales (0,16%),
- The employment share of Tertiary Dealers (0,03 %) is significantly **less than** that of England and Wales (0,06 %),

And that the same holds true for those:

Coded under '**Unspecified Occupations' Without employment', 'Aged, Sick, Poor'**

which account for significantly **lower** shares (44,7%) as opposed to 57,4 % in England and Wales

#### Employment profile in Cluster 11 of Women's Occupations Map and Average Employment of in England and Wales in 1851

Sectors	% in Cluster No. 11*	% in England and Wales
Agriculture	1,39	3,17
Mining and Quarrying	0,08	0,16
Secondary Occ.	35,1	15,2
Tertiary Dealers	0,3	0,6
Tertiary Sellers	1,1	0,9
Tertiary Services & Professions	14,8	19,9
Tertiary Transport & Comm.	0,17	0,20
Without-Unspecified Occupations	44,7	57,4

Source Derived from CAMPOP 1851 data set on Women's Occupations (at PST3 level and by Parishes)

This set on indicators facilitate interpretation yet for practical reasons they can hardly be used as map legends.

Yet they can be re-calculated for the following seven general categories (see slide 26)

- **Primary Occupations**
  - o Agriculture,
  - o Mining and Quarrying,
- **Secondary Occupations**
  - o Secondary occupations with high concentration indices
- **Tertiary Occupations**
  - O Dealers,
  - O Sellers,
  - O Services and Professions,
  - O Transport and Communication,
- **Unspecified or Without Occupation**

**TEMPLATE for A B R I D G E D LEGENDS**

	Primary Occupations		Secondary Occ. S	Tertiary Occupations				Without - Unspecified Occ's
CLU No.	Agriculture	Mining & Q.	Secondary	Dealers	Sellers	Services & Prof.	Transport & Comn.	W_Occ & .. ASP
1								
2								
3								

Employment profiles are denoted through Location Quotients and with respect to the following conventions

1. If the sector's  $LQ < 1$  and no occupation has a  $LQ > 1$ . The column for this sector is left blank.  
The heading for this sector is written in small characters.
2. If the sector's  $LQ < 1$  but if it comprises occupations with  $LQ's > 1$   
The LQ 's first three occupations with highest SCI's are listed  
The heading for this sector is written in **bold** characters.
3. If the sector's  $LQ > 1$  The LQ 's first three occupations with highest SCI's are listed  
The heading for this sector is written in **bold and oversized** characters.

Abbreviated Legend for the first plate of Women’s Occupations Map (see Slide No. 25 below).

11	Agriculture 0.4	Mining & Q. 0.5	<b>Secondary 2.3</b>	<b>Dealers 1.4</b>	<b>Sellers 1.2</b>	Services & P 0.7	Tr&Com 0.8	W_Occ & .. ASP 0.8
	Estate Work 1.5	MQ 13.8	TXTs, cotton 10.5	Metal, zinc 4.8	Itinerant 2.0	CAs, Regul.&measur. 9.2	Road Tr.-Anim. Pow. 2.1	Unspessific. 1.4
		Quarrying 1.2	TXTs 5.0	Food 3.0	Small traders 1.8	Army 3.1	Road Tr.-motorised 1.7	
		Coal 1.1	Secondary 5.6	TXTs, cotton 3.4	Iron & steel 1.7	DACC, Alcoholic drinks 1.9	Inland navig. 1.3	

- 1. Occupations in Agriculture, Mining and Quarrying, Services and Professions, Transport and Communications are significantly low. Yet, concentrations in Estate Work , Mining &Quarrying, Coal Mining, Regulation and Measurement, Army, Alchoolic Drinks in Animal Power and Motorized, Road Transoirt, Inland Navigation, and in Unspecified Occupations.
- 1. Secondary Occupations,  
Cotton Manufacturer, Operatives in various trades of Textiles or Secondary are **over-represented**
- 3. Tertiary Dealers  
in Food, Cotton and in Metals: Zinc are **over-represented**
- 4 Tertiary Sellers, Itinerant sellers, Small Traders, Iron and Steel Products are slightly **over-represented**

Referring Legent tables produced in Appendix 4 the Parish clusters can be assigned to three general groups and labelled as follows

The first group comprises six clusters specialized in various braches of Textiles

- **Cotton Manuf. Textiles** Plate 2 in 1851 Women's occupations Map (Cluster No 11) 7.3 % of Women's Employment
  - **Wool and Worsted** Plate 3 in 1851 Women's occupations Map (Cluster No 10) 4,4 % of Women's Employment
  - **Worsted and Wool** Plate 4 in 1851 Women's occupations Map (Cluster No 12) 1,1 % of Women's Employment
  - **Silk and Cotton** Plate 5 in 1851 Women's occupations Map (Cluster No (8) 3,2 % of Women's Employment
  - **Laces Agr. Gardens etc** Plate 6 in 1851 Women's occupations Map (Cluster No (6) 2,4 % of Women's Employment
  - **Fibre, Rush, Straw** Plate 7 in 1851 Women's occupations Map (Cluster No (4) 7,1 % of Women's Employment
- 

in 1851 these six clusters accounted for 25,4 % of Women's Employment

Parish clusters with high levels of Specialization in Tertiary Occupations

- **Linen, Pottery, Iron & Steel** Plate 8 in 1851 Women's occupations Map (Cluster No (9) 4,6 % of Women's Employment
  - **Footware, Clothing** Plate 9 in 1851 Women's occupations Map (Cluster No (2) 7,8 % of Women's Employment
  - **Clothing, Kitchen,**
  - **Laundry Services** Plate 10 in 1851 Women's occupations Map (Cluster No (1) 26,8 % of Women's Employment
- 

in 1851 these three clusters accounted for 39,2 % of Women's Employment

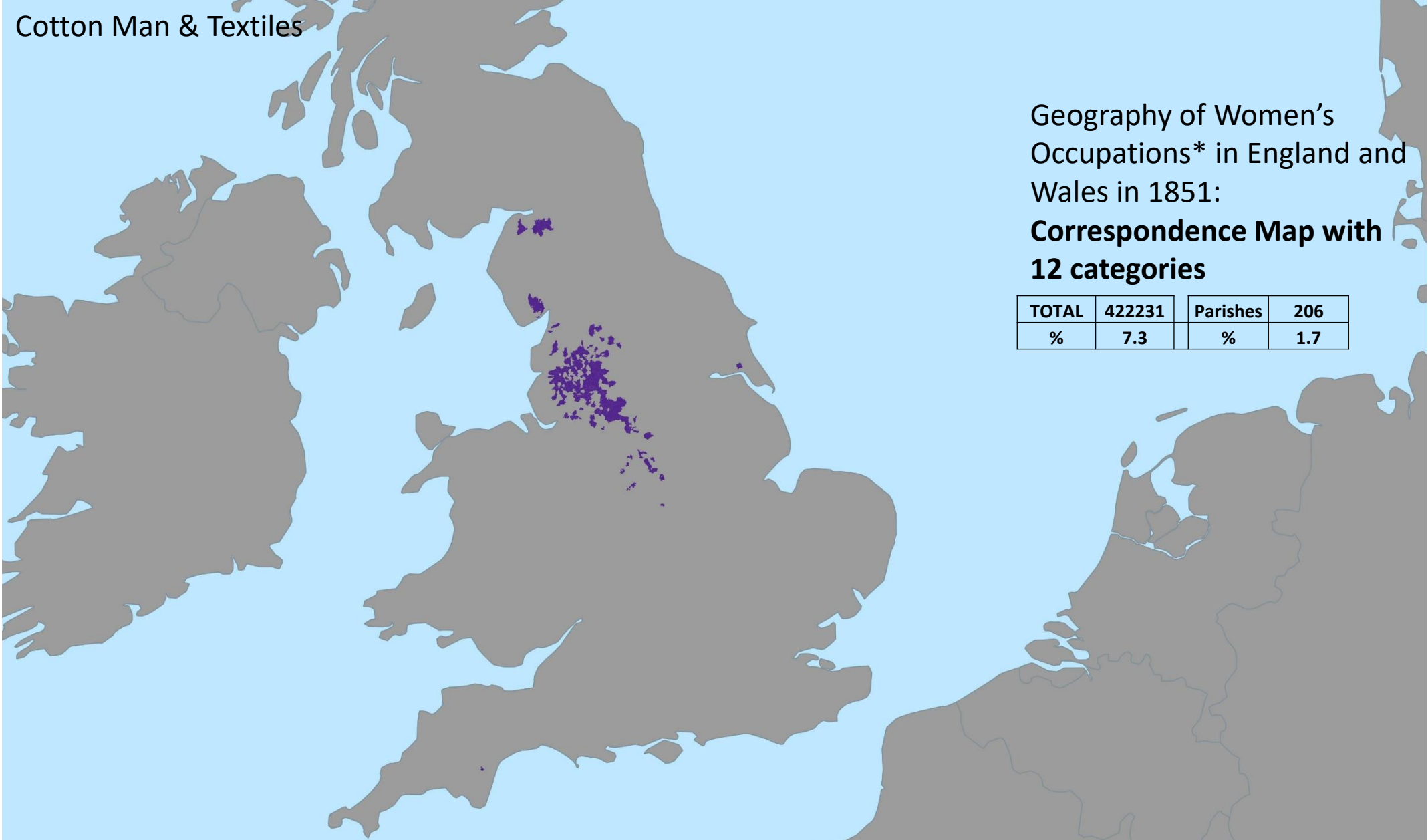
Parish clusters specialized in Agriculture industries services compatible with agriculture

- **Farming, Fishing Tertiary** Plate 11 in 1851 Women's occupations Map (Cluster No (3) 8,3 % of Women's Employment
  - **Agriculture, Fishing Tertiary** Plate 12 in 1851 Women's occupations Map (Cluster No (7) 16,4 % of Women's Employment
  - **Farming Ind & Services & P** Plate 13 in 1851 Women's occupations Map (Cluster No (5) 10,6 % of Women's Employment
- 

in 1851 these three clusters accounted for 35,3 % of Women's Employment

**The first group comprises six clusters specialized in various braches of Textiles (see Slides 30-35)**

- **Cotton Manuf. Textiles**      Plate 2 in 1851 Women's occupations Map (Cluster No 11)      7.3 % of Women's Employment
- **Wool and Worsted**      Plate 3 in 1851 Women's occupations Map (Cluster No 10)      4,4 % of Women's Employment
- **Worsted and Wool**      Plate 4 in 1851 Women's occupations Map (Cluster No 12)      1,1 % of Women's Employment
- **Silk and Cotton**      Plate 5 in 1851 Women's occupations Map (Cluster No (8)      3,2 % of Women's Employment
- **Laces Agr. Gardens etc**      Plate 6 in 1851 Women's occupations Map (Cluster No (6)      2,4 % of Women's Employment
- **Fibre, Rush, Straw**      Plate 7 in 1851 Women's occupations Map (Cluster No (4)      7,1 % of Women's Employment



Geography of Women’s Occupations\* in England and Wales in 1851:

**Correspondence Map with 12 categories**

TOTAL	422231	Parishes	206
%	7.3	%	1.7

11	Agriculture 0.4	Mining & Q. 0.5	<b>Secondary 2.3</b>	<b>Dealers 1.4</b>	<b>Sellers 1.2</b>	Services & P 0.7	Tr&Com 0.8	W_Occ & .. ASP 0.8
	Estate Work 1.5	MQ 13.8	TXTs, cotton 10.5	Metal, zinc 4.8	Itinerant 2.0	CAs, Regul.&measur. 9.2	Road Tr.-Anim. Pow. 2.1	Unspessific. 1.4
		Quarrying 1.2	TXTs 5.0	Food 3.0	Small traders 1.8	Army 3.1	Road Tr.-motorised 1.7	
		Coal 1.1	Secondary 5.6	TXTs, cotton 3.4	Iron & steel 1.7	DACC, Alcoholic drinks 1.9	Inland navig. 1.3	

Geography of Women’s  
Occupations\* in England and  
Wales in 1851:  
**Correspondence Map with  
12 categories**

TOTAL	258848	Parishes	219
%	4.4	%	1.8

10	Agriculture 0.8	Mining & Q. 0.6	Secondary 2.1	Dealers 1.6	Sellers 1.1	Services & P 0.7	Tr&Com 0.7	W_Occ & .. ASP 0.8
	Forestry 1.2	Coal 1.3	TXTs, wool 16.8	Clothing 7.6	Small traders 6.5	CAs, Agents 11.2	Road Tr.-Anim. Pow. 1.2	
		Quarrying 1.2	TXTs 7.1	Metal, zinc 6.6	Food 1.2	Army, 1.5	Inland navig. 1.5	
		Lead 0.6	TXTs, worsted 7.9	Food, bread, confect. 1.7		DACC, Alcoholic drinks 1.3	Inland navig.goods 0.9	

Geography of Women’s Occupations\* in England and Wales in 1851:  
**Correspondence Map with 12 categories**

TOTAL	63633	Parishes	27
%	1.1	%	0.2

12	Agriculture 0.4	Mining & Q. 0.6	Secondary 2.7	Dealers 1.3	Sellers 0.8	Services & P 0.6	Tr&Com 0.5	W_Occ & .. ASP 0.8
		Coal 1.0	TXTs, worsted 50.8	TXTs, wool 11.0	Itirenant 1.7	Entert. theather 3.4	Road Tr.-Anim. Pow. 1.5	Uncert. age, sickness 0.6
			TXTs 3.9	Food, bread, confect. 1.9	Iron & steel 1.0	Mis. sanitary 4.5	Communications 2.9	Students 0.5
			TXTs, wool 4.2	Machines, tools 22.8		Entertainment 1.4	Inland navig. 1.2	Uncert-others 0.7

Geography of Women’s Occupations\* in England and Wales in 1851:

Correspondence Map with 12 categories

TOTAL	184494	Parishes	109
%	3.2	%	0.9

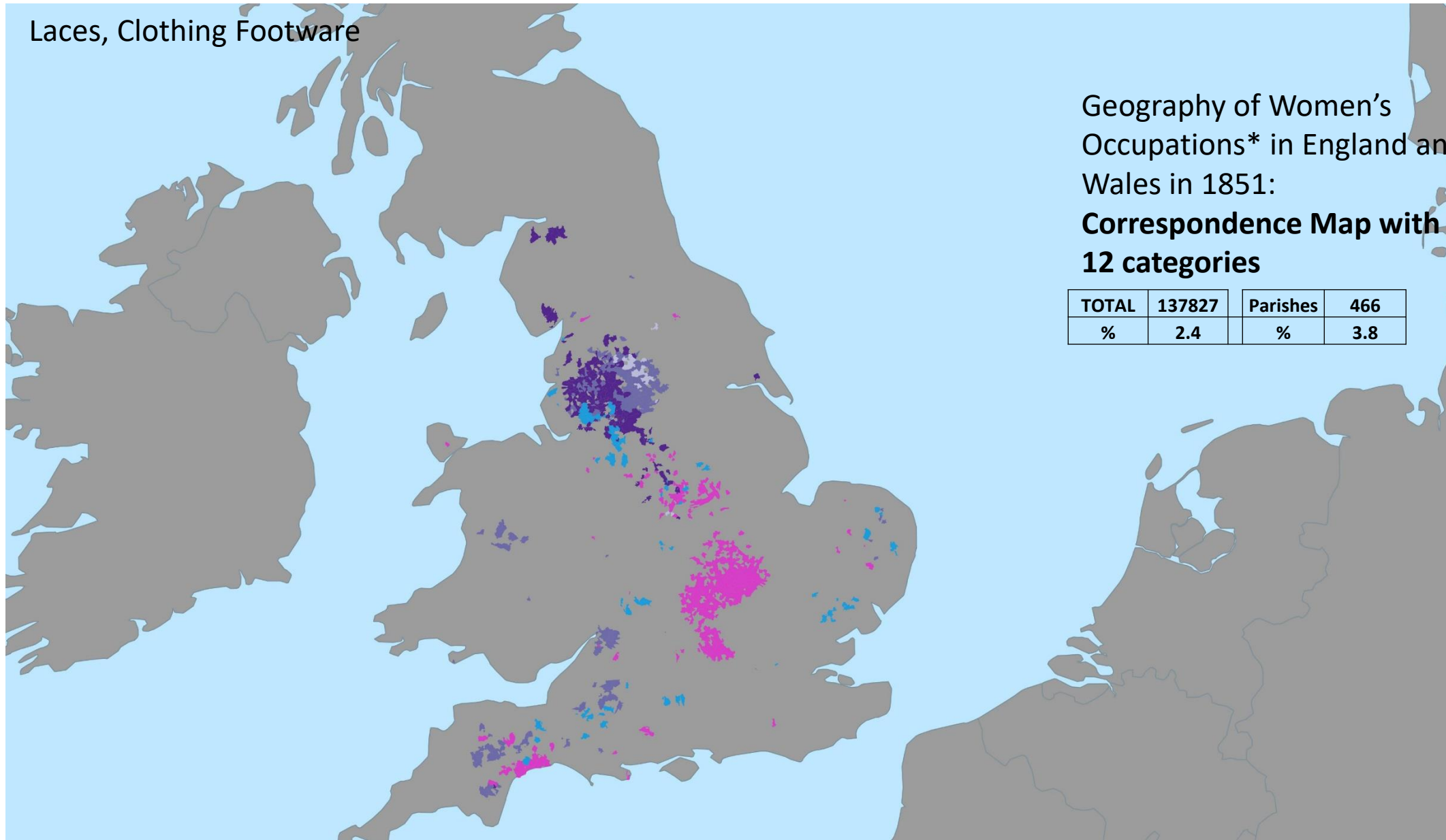
8	Agriculture 0.5	Mining & Q. 0.2	Secondary 2.3	Dealers 1.2	Sellers 1.2	Services & P 0.8	Tr&Com 1.0	W_Occ & .. ASP 0.8
	Fishing 1.9		TXTs, silk 20.7	Metal, zinc 3.9	Ter. 1.5	Mis. funeral 3.5	Communications 1.5	Unspessific 1.3
			TXTs, cotton 2.0	Food 2.5	Initerant 1.4	Mis. laundry 1.0	TP 1.6	
			TXTs 1.7	TXTs, silk 5.2	Chemical 2.8	DACC, Alcoholic drinks 1.2	Road Tr.-motorised 2.1	

Laces, Clothing Footware

Geography of Women's  
Occupations\* in England and  
Wales in 1851:

**Correspondence Map with  
12 categories**

TOTAL	137827	Parishes	466
%	2.4	%	3.8



6	<b>Agriculture 1.0</b>	Mining & Q. 0.1	<b>Secondary 1.7</b>	<b>Dealers 1.1</b>	Sellers 0.8	Services & P 0.8	Tr&Com 0.9	W_Occ & .. ASP 0.9
	Agr. 1.5		TXTs, lace 28.3	TXTs, lace 7.9	Small traders 1.7	Education 1.1	Road Tr.-Anim. Pow. 1.2	Unspessific 1.2
	Agr.Gardens 1.1		Clothing 1.6	Wood 2.5	Food 1.2	DACmn 1.15	Rail tr. 1.2	Uncert. age, sickness 1.2
	Agr.Animal 1.0		Footwear 1.1	Food, bread, confect. 1.1	Furnishing 4.2	CA commercial 1.3	Communication 1.2	Students 0.9

Fibre, Straw, Rush, Clothing

Geography of Women’s Occupations\* in England and Wales in 1851:

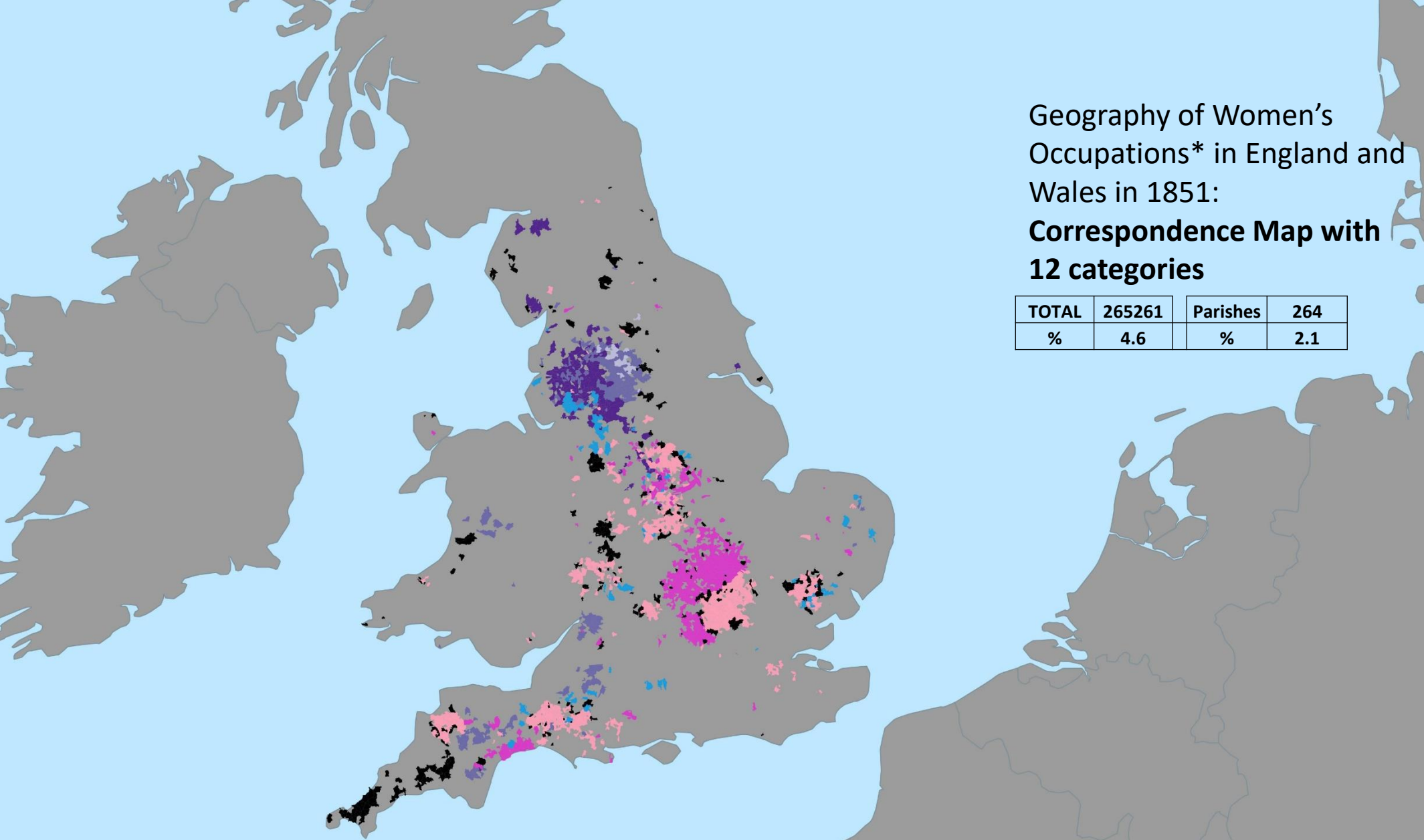
Correspondence Map with 12 categories

TOTAL	415005	Parishes	517
%	7.1	%	4.2

4	Agriculture 0.6	Mining & Q. 0.2	Secondary 2.0	Dealers 1.2	Sellers 0.9	Services & P 0.8	Tr&Com 1.3	W_Occ & .. ASP 0.8
			Fibre, straw, rush prod. 12.4	Clothing 11.1	Clothing 3.1	Storage 8.7	Rail tr. 3.8	
			Clothing 5.5	TXTs, lace 6.0	Ter. 1.0	Army 1.7	Communications 3.9	
			TXTs 13.0	TXTs, cotton 4.4	Small traders 1.1	Local gov. 1.3	Road Tr.-motorised 2.2	

**Parish clusters with high levels of Specialization in Tertiary Occupations**

- **Linen, Pottery, Iron & Steel** Plate 8 in 1851 Women’s occupations Map (Cluster No (9) 4,6 % of Women’s Employment
- **Footware, Clothing** Plate 9 in 1851 Women’s occupations Map (Cluster No (2) 7,8 % of Women’s Employment
- **Clothing, Kitchen,**
- **Laundry Services** Plate 10 in 1851 Women’s occupations Map (Cluster No (1) 26,8 % of Women’s Employment

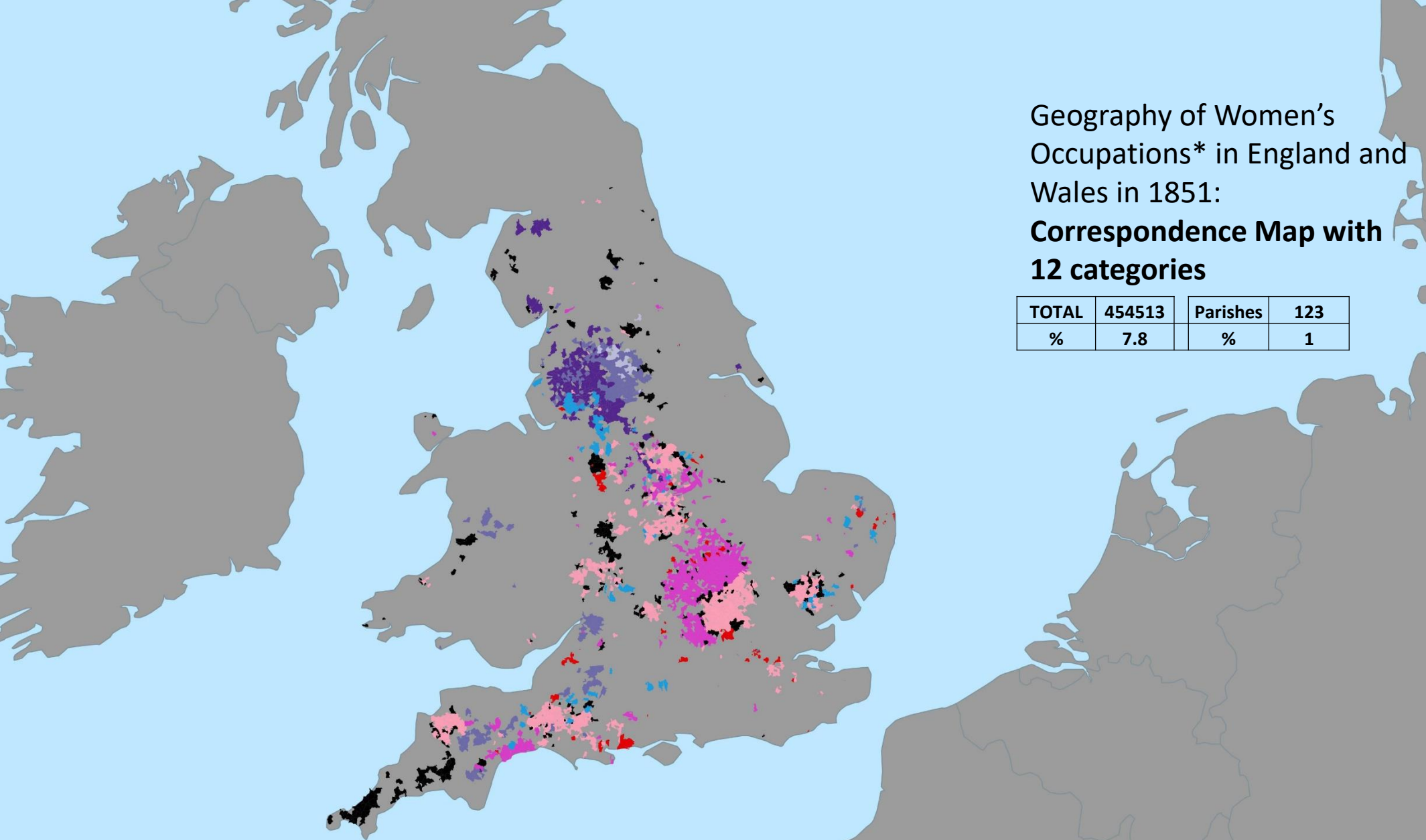


Geography of Women’s Occupations\* in England and Wales in 1851:

Correspondence Map with 12 categories

TOTAL	265261	Parishes	264
%	4.6	%	2.1

9	Agriculture 0.6	<b>Mining &amp; Q. 11.2</b>	<b>Secondary 1.2</b>	<b>Dealers 1.0</b>	<b>Sellers 1.0</b>	Services & P 0.7	<b>Tr&amp;Com 1.6</b>	<b>W_Occ &amp; .. ASP 1.1</b>
		Quarrying 15.6	Iron, steel pr. 13.1	TXTs 1.5	Small traders 1.3	Artistic prof. 11.1	Rail tr. 5.5	Not working 1.1
		Copper 20.8	Pottery 16.6	Metal 3.8	Food 1.1	Storage 1.9	Inland navig. 2.1	Uncert. age, sickness 0.9
		Tin 19.0	TXTs, linen 15.6	Pottery 1.9	Food, fruit,vegetable 1.1	Engineering 4.2	Com. Post Of. 1.2	Unstated 0.9

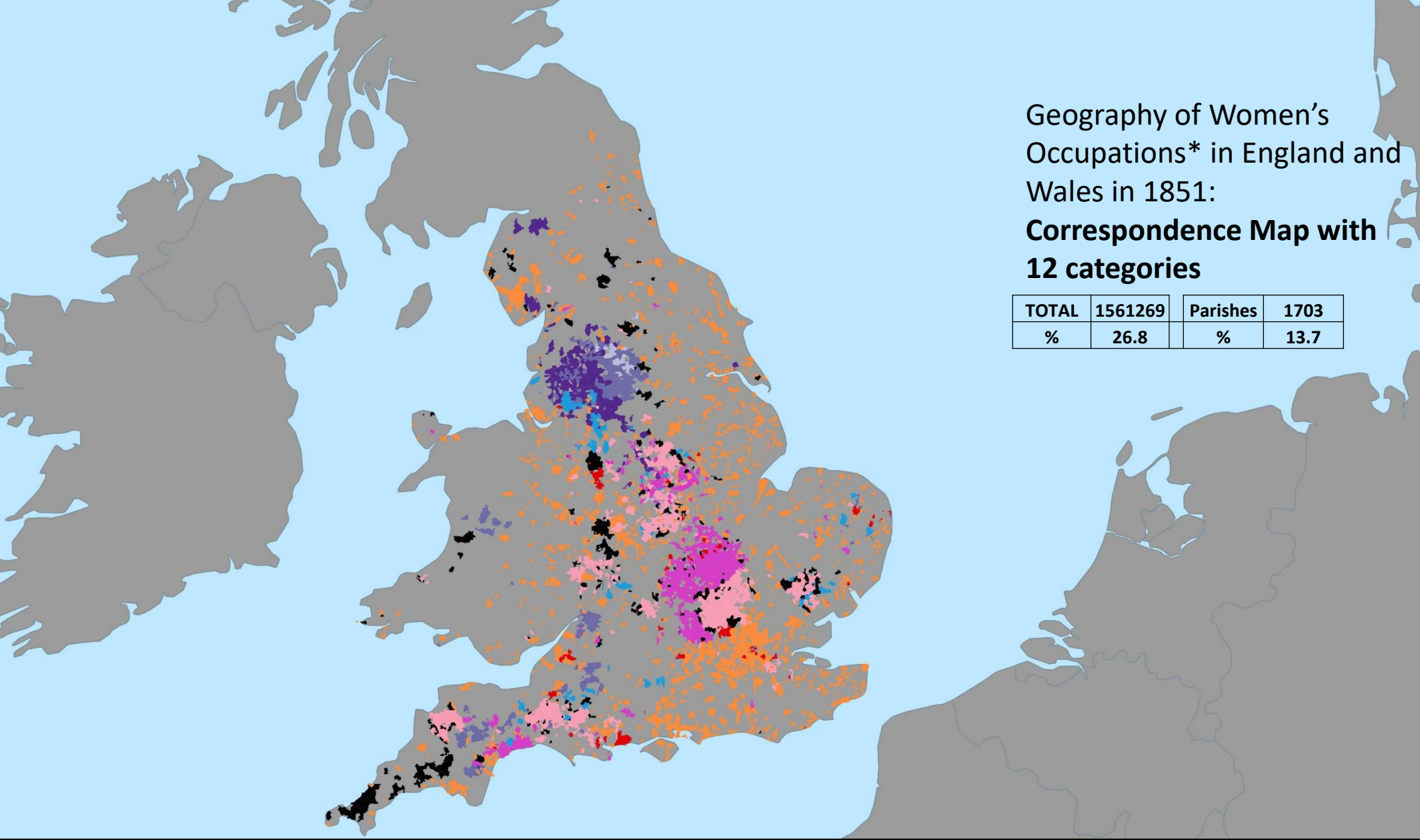


Geography of Women’s Occupations\* in England and Wales in 1851:

**Correspondence Map with 12 categories**

TOTAL	454513	Parishes	123
%	7.8	%	1

2	Agriculture 0.2	Mining & Q. 0.1	<b>Secondary 1.3</b>	<b>Dealers 1.1</b>	<b>Sellers 1.7</b>	<b>Services &amp; P 1.0</b>	<b>Tr&amp;Com 1.0</b>	<b>W_Occ &amp; .. ASP 1.0</b>
	Forestry 4.7	Gold 1.4	Footwear 4.6	Minor pr 2.4	Food, fruit,vegetable 2.9	Mis. laundry 1.5	Sea Tr. 1.7	Not working 1.1
	Agr.Gardens 1.8		Clothing 1.9	Wood 2.0	Initerant 2.3	Domestic 1.6	Inland navig. 2.9	Uncert. age, sickness 1.1
	Agr.Animal 1.2		Printing 5.6	Leather 4.1	Ter. 1.8	Financial, broking 3.6	Communication 1.4	Students 0.9



Geography of Women’s Occupations\* in England and Wales in 1851:

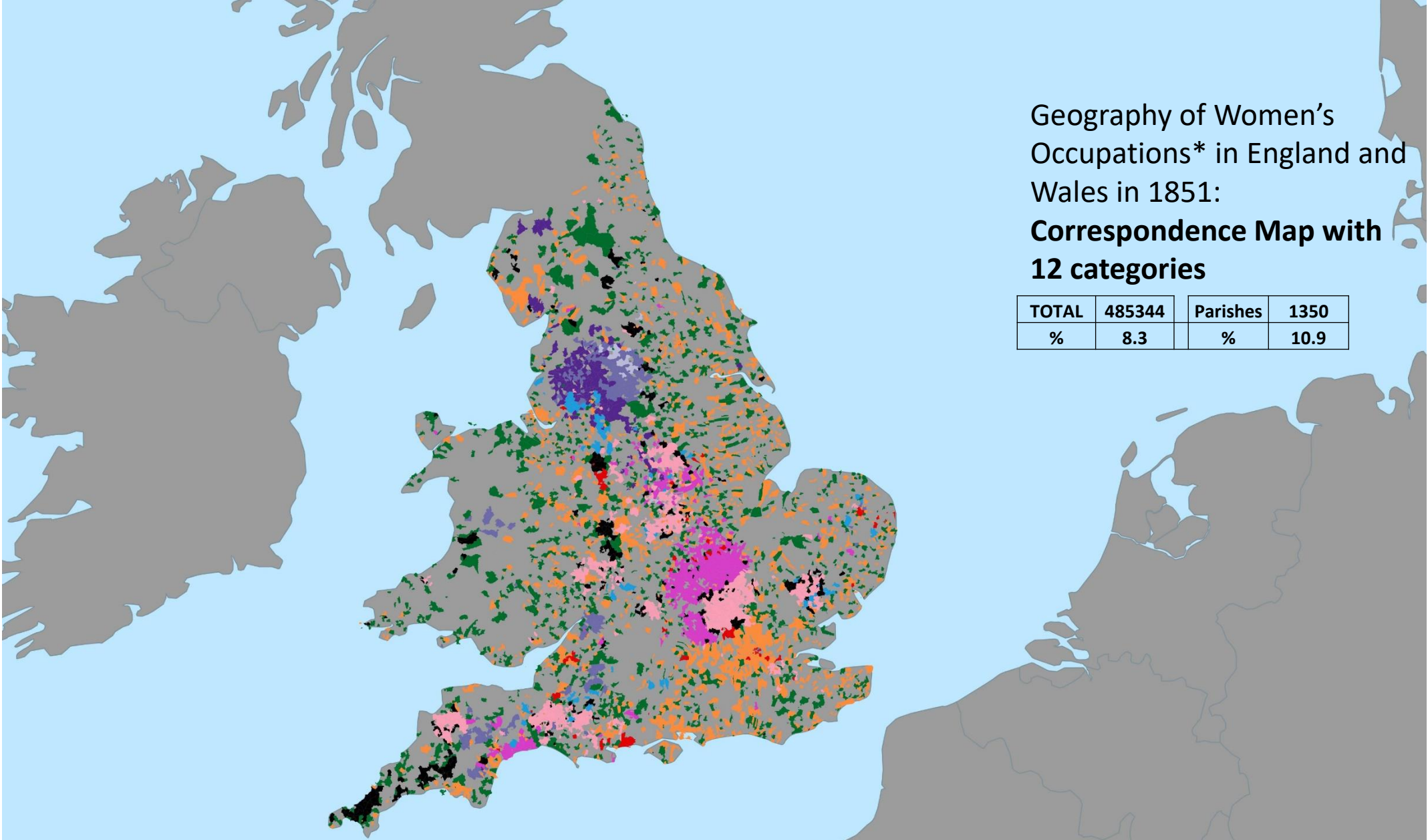
Correspondence Map with 12 categories

TOTAL	1561269	Parishes	1703
%	26.8	%	13.7

1	Agriculture 0.3	Mining & Q. 0.1	Secondary 0.6	Dealers 1.0	Sellers 1.1	Services & P 1.5	Tr&Com 1.0	W_Occ & .. ASP 1.0
	Agr. Gardens 1.6		Clothing 1.2	TXTs 1.3	Food, fruit,vegetable 1.5	Domestic, house 1.35	Sea Tr. 1.4	Students 1.4
	Agr. Animal 1.2		RT Vehicles 2.0	Minor pr 1.3	Food, meat-poultry 1.6	Misc. Laundry 1.7	Road Tr.-Anim. Pow. 1.7	Not Working 1.0
			B&C, Carpentry 1.3	Food, meat, poultry 1.6	Tobacco 1.8	Doms, Kitchen Stf. 2.1	Communications 1.2	Age, Sickness 1.1

**Parish clusters with diversified occupation structures (urban fringes agriculture sector)**

- Farming, Fishing tertiary  
light Industries Plate 11 in 1851 Women's occupations Map (Cluster No (3)
- Agriculture, Fishing tertiary  
light Industries Plate 11 in 1851 Women's occupations Map (Cluster No (2)
- Farming Rural Industries  
and Services Plate 12 in 1851 Women's occupations Map (Cluster No (5)



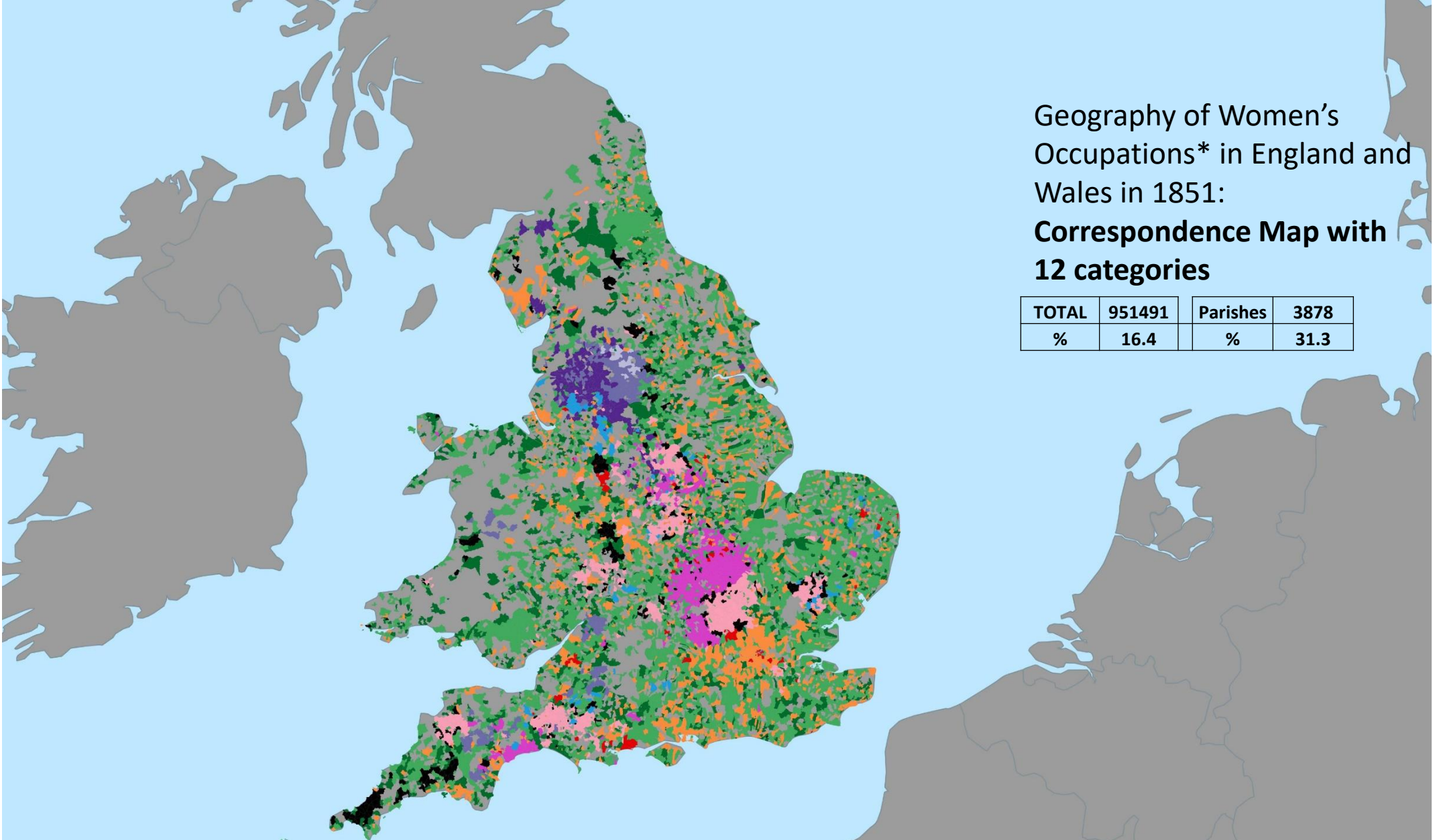
Geography of Women’s Occupations\* in England and Wales in 1851:  
**Correspondence Map with 12 categories**

TOTAL	485344	Parishes	1350
%	8.3	%	10.9

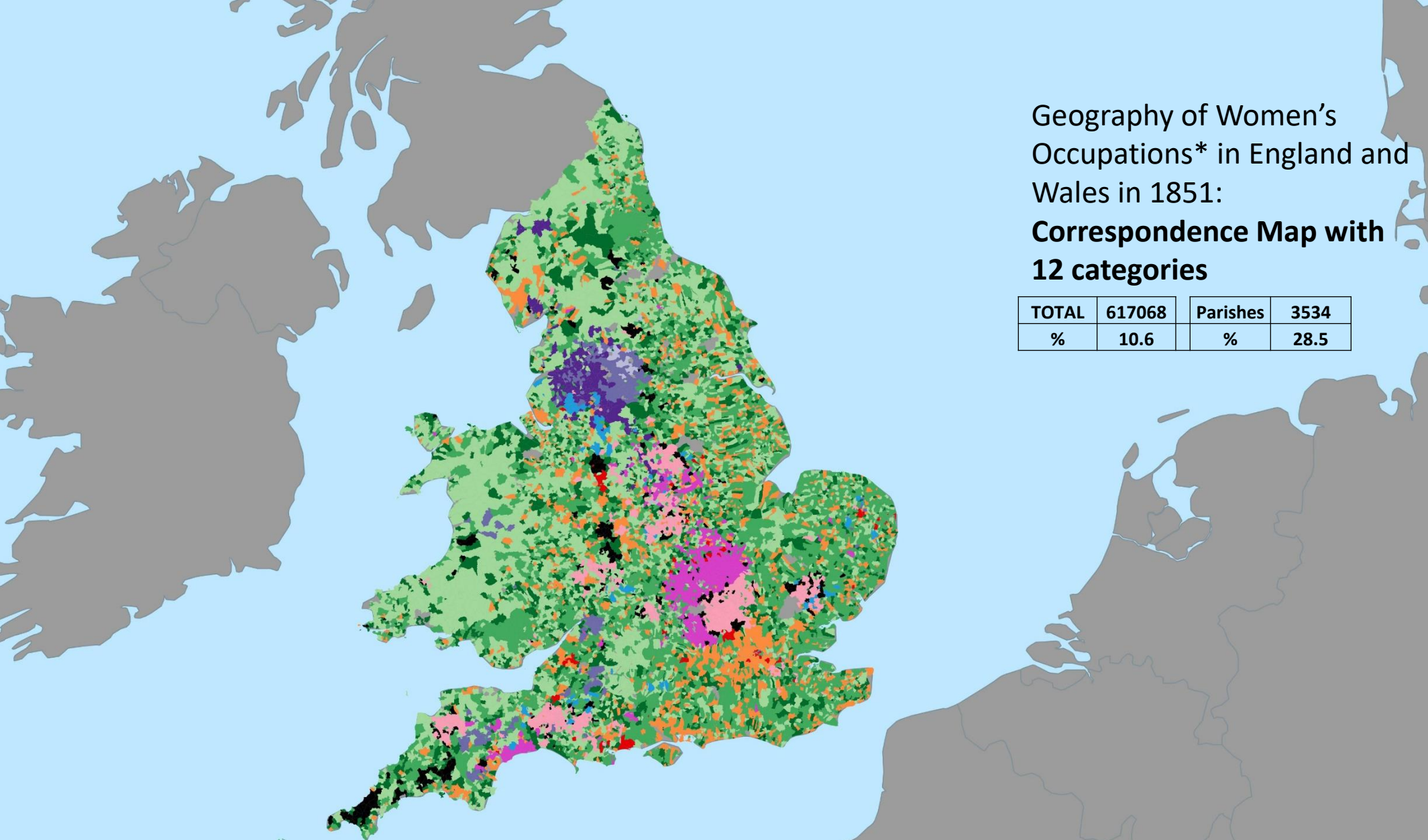
3	<b>Agriculture 1.4</b>	Mining & Q. 0.7	Secondary 0.5	<b>Dealers 1.1</b>	<b>Sellers 1.1</b>	<b>Services &amp; P 1.1</b>	<b>Tr&amp;Com 1.1</b>	<b>W_Occ &amp; .. ASP 1.1</b>
	Agr.Farming 1.4	Lead 1.5	Fibres, rope, cord prod. 5.4	Fibrous veget. 7.3	Food 1.6	Domestic, house 1.1	Com. Post Of. 1.6	Unstated 2.1
	Fishing 4.3	Quarrying 1.1	Chemical 3.7	Food, bread, confect. 1.4	Food, poultry 2.3	DACCmn 1.4	Road Tr.-Anim. Pow. 1.3	Uncertain stat. 1.9
	Agr. 1.7	Gold 1.3	Chem.-natural oil ref. 4.6	Food, cereal 1.7	Iron & metal pr. 1.2	Prof.Religion 1.6	Inland navig. 1.5	Unspecific 1.4

Geography of Women's  
Occupations\* in England and  
Wales in 1851:  
Correspondence Map with  
12 categories

TOTAL	951491	Parishes	3878
%	16.4	%	31.3



7	<b>Agriculture 1.1</b>	<b>Mining &amp; Q. 1.6</b>	<b>Secondary 0.3</b>	<b>Dealers 0.6</b>	<b>Sellers 0.7</b>	<b>Services &amp; P 0.8</b>	<b>Tr&amp;Com 0.9</b>	<b>W_Occ &amp; .. ASP 1.3</b>
	Agr.Farming 1.1	Coal 3.3	Glass prod. 3.4	NFM and metal 2.8		Financial 1.3	TP 1.3	Not working 1.4
	Agr. 1.3	Iron 5.4	Iron pr. ind. 2.6	Stone 1.5		DACCmn 1.0	Com. Post Of. 1.1	Unspecific 1.1
		Quarrying 1.5	Iron pr. 4.5	Building materials 1.11			Inland navig. 1.2	Uncert. age, sickness 1.1



Geography of Women’s Occupations\* in England and Wales in 1851:  
**Correspondence Map with 12 categories**

TOTAL	617068	Parishes	3534
%	10.6	%	28.5

5	<b>Agriculture 4.1</b>	Mining & Q. 0.4	Secondary 0.4	Dealers 0.7	Sellers 0.7	<b>Services &amp; P 1.0</b>	Tr&Com 0.9	<b>W_Occ &amp; .. ASP 1.0</b>
	Agr.Farming 4.3	Lead 1.5	Food, milling 2.0	Food, cereal 1.4	Food 1.2	Domestic, house 1.1	Com. Post Of. 1.5	Unstated 2.6
	Agr. 2.6	Quarrying 1.4	Clothing 1.6	Wood 1.1		Prof.Religion 1.4	Road Tr.-Anim. Pow. 1.4	Uncertain stat. 2.3
	Agr.Animal 1.3	MQ 9.4	NFM, zinc 3.3	Live animals 1.5		Construction 1.5	Inland navg. 1.7	Unspecific 1.3

Dear Leigh

**I have identified the following methodological advantages.**

**If you see wish we can add few points on empirical findingsrmat.**

Conclusions

We start to see that the model is successful in;

- detecting *affinities* amongst thousands of parish occupation profiles,
- generating useful *typologies* and in defining economic regions
- elucidating the nationwide *deployment* of occupations through reliable quantitative indices
- Thanks to sharp contrasts in occupation profiles, parish clusters are easy to interpret and label
- The parish clusters can be used as legends and generate legible maps
  - Map series for *Men's Occupations and in 1851* (See Appendix 1)
  - Map series for *Men's and Women's Occupations (Combined)* in 1851 (See Appendix 2)drawn with the same approach,can be interpreted similarly

Complementary Use of Correspondence and Cluster Analysis” constitutes an efficient tool for pattern recognition in disaggregated data sets. Thanks to the mathematical symmetry between Correspondence and ClusterAnalyses

- negative impacts of outliers are kept under control,
- The information loss is minimized and measurable
- Efficient Parish Clusters are obtained (i.e. with minimum within group variances and maximum between group variances)
- The latter constitute sound references for case studies, and efficient legends for mapping and visualization
- The accuracy of the data reduction exercise can be enhanced either by:
  - Increasing the number of categories  
(See Appendix 1 Men’s Occupation Map with 4 categories (Slides 4-7) and  
Men’s Occupation Map with 19 categories (Slides 10-28)
  - Or by unpacking previously aggregated clusters
- (See Appendix 1 Re-clustering parishes specialized in tertiary occupations (Slides 31-40)