Settlers and surnames: An atlas illustrating the origins of settlers in 19th century America.

Eff, Ellis Anthon

Middle Tennessee State University

15 April 2013

Settlers and surnames

An atlas illustrating the origins of settlers in 19th century America.

Ellis Anthon Eff

April 15, 2013

Contents

List of Figures

Preface

This is an atlas, where the message is largely visual and the average reader will not read more than the captions. The additional text serves primarily as documentation of the procedures used in creating the maps.

The atlas is divided into two main parts: the first shows migration patterns for the continental United States, between roughly 1800 and 1880; the second shows the ethnic composition of the easternmost states based upon surname distributions.

Data for these maps are drawn from the North American Population Project [4, 7, 13, 5]. Analysis is done using tools created by contributors to R [21, 2, 12, 28, 15, 27, 17, 14, 3, 18, 6]. The document is written with LyX [25].

1 Migration towards the west

This map book examines the composition of the US population between approximately 1820 and 1880. The data are drawn from the 1880 census of the United States, obtained from the North American Population Project [4, 7]. This dataset is a public-use microdata sample (PUMS), where each record is a surveyed person. It is an 100 percent sample, containing all of the approximately 50 million surveyed persons.

The dataset contains fields for ego's county of residence and birthplace of ego's mother and father. Birthplaces are recorded as states or foreign countries. In order to visualize migration patterns, these three fields (county of residence, mother's birthplace, father's birthplace) are extracted for all persons born in 1840 or earlier. The data are rearranged into two fields: current county of residence and parent's birthplace, each person appearing in two records, one for father and one for mother. A matrix is then constructed, with rows representing US counties and columns US states and foreign countries. Each cell m_{ij} in M gives the number of incidences in which a person residing in the row county has a parent from the column county.[1] Matrix M is then row-normalized (i.e., each row is divided by the row sum) to obtain matrix $\tilde{M}$, where each cell $\tilde{m}_{ij}$ gives the proportion of parents of persons in row county i that were born in column place j. A few sample rows and columns of $\tilde{M}$ are shown in Table1.

[1] Thus, a person in county i with both parents from the same state j would count as two instances in m_{ij}.

Table 1.1: Summary statistics for $\tilde{M}$

	Maryland	Ireland	Virginia	New York
60910	0.0065	0.1375	0.02	0.0541
240190	0.7923	0.0269	0.0322	0.0023
281630	0.0278	0.0161	0.256	0.0016
190870	0.0418	0.0847	0.1177	0.0648
80050	0.0118	0.1287	0.0507	0.1074
271570	0.0026	0.182	0.0083	0.1153
360490	1e-04	0.1459	5e-04	0.3489
280890	0.0269	0.0144	0.2309	0.0025

While the date of recorded residence in the row county is unambiguously 1880, the date at which a parent was born in the column location takes a much wider range. Figure 1.1 gives some insight into that range. The bottom chart shows the distribution of ages of persons 40 and older; note that years divisible by 10 are over-represented, indicating that many persons gave an approximate age. A 40 year old would have been born in 1840, an 80 year old in 1800. Extracting from the original 1880 census data all households with natural children

of the household head, the ages of the household head and spouse were calculated for the year of birth for each of these children.[2] The resulting empirical probabilty that a child at birth will have a parent of a given age is shown in the middle chart of Figure 1.1. Using this empirical probability to estimate the age of parents for all persons ages 40 and up leads to the top chart: a cdf for the estimated year of birth of parents in matrix $\tilde{M}$. One can see that all but a tiny fraction of parents were born before 1820, half were born before 1797, and about nine percent were born prior to 1776. Matrix $\tilde{M}$ thus provides insight into how the population of 1770-1820 rearranged itself so as to give the population of 1880. Note that the migration from source region to target region could have taken place as late as 1880.

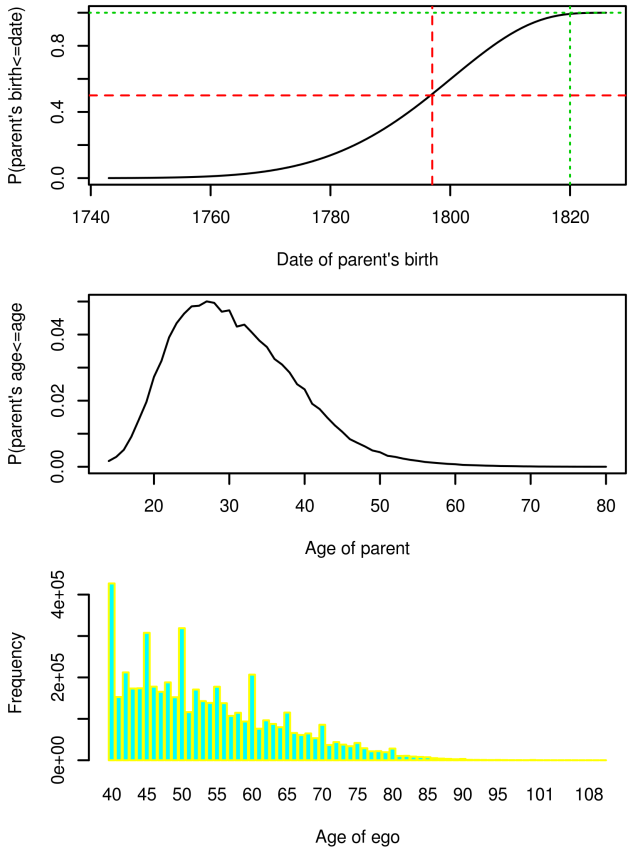

Figure 1.1: Bottom chart shows distribution of ages for persons 40 and older. Middle chart shows the empirical probability that a child at birth will have a parent of the age given in the x-axis. The top chart shows the empirical cdf for year of birth of parents in matrix $\tilde{M}$.

[2] Children were ages 0 to 13. The number of matches between parent and child was 31,846,218. Spouses may not have always been a natural parent.

The following pages present maps showing the 1880 distribution of population originating from specific states and foreign countries. Only the most important sources are considered. The first section covers migration from the northern states forbidding slavery; the second section looks at migration from the southern slave states; and the third section looks at the settlement patterns of immigrants from other countries.

1.1 Northern states

Slavery was forbidden in all of these states by 1820. The easternmost of these states were well settled by 1820 and served as major sources for westward migration during the 19th century. One striking feature of the westward migration was that it tended to be due west, fanning out only past the Mississippi River. Figure 1.2 shows the pattern of migration from Maine.

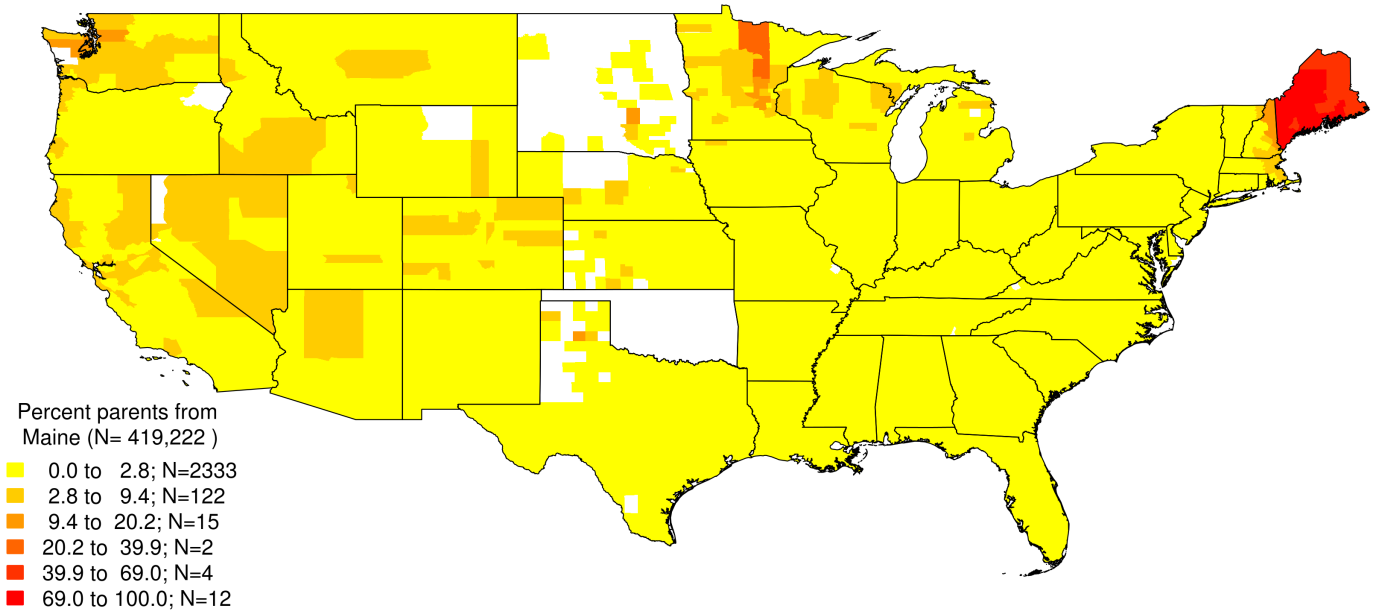

Figure 1.2: Maine, like other New England states, shows a marked tendency for migration to move due west, despite the intervening barrier of the Great Lakes. All migration maps on the following pages use the same intervals in the color scale.

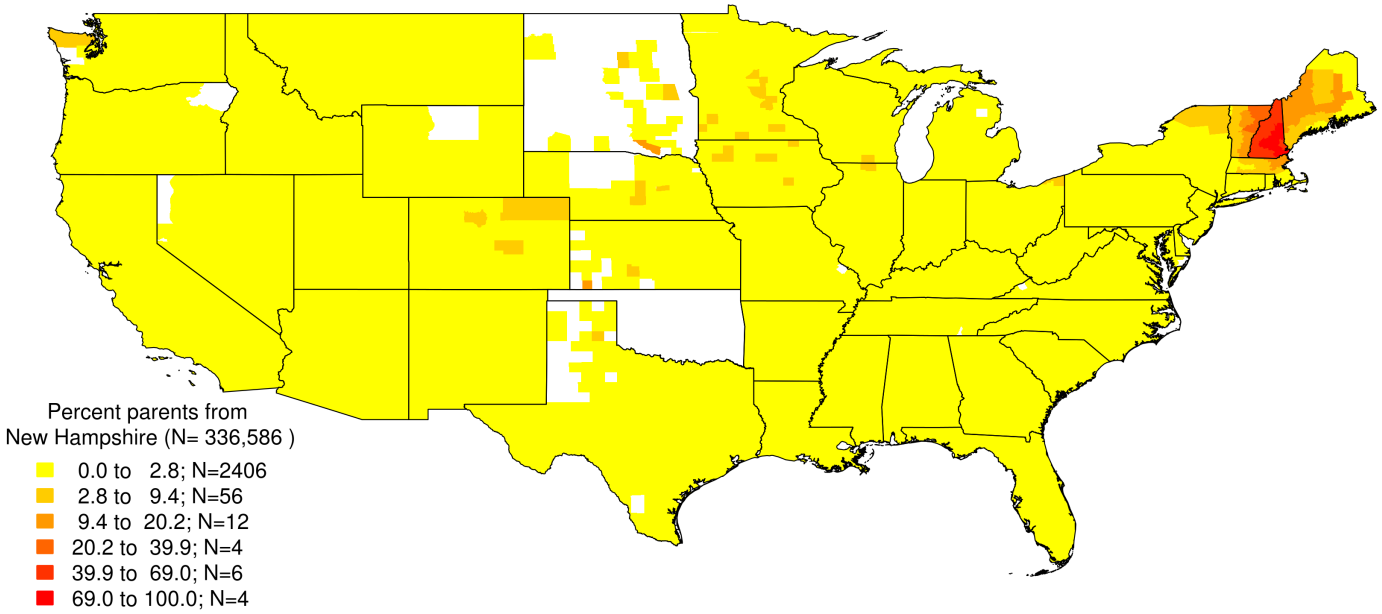

Figure 1.3: New Hampshire has a fairly pronounced migration shadow extending past its borders: eastward, into Maine, southward into eastern Massachusetts; and westward into Vermont and New York. There was negligible migration over a longer distance.

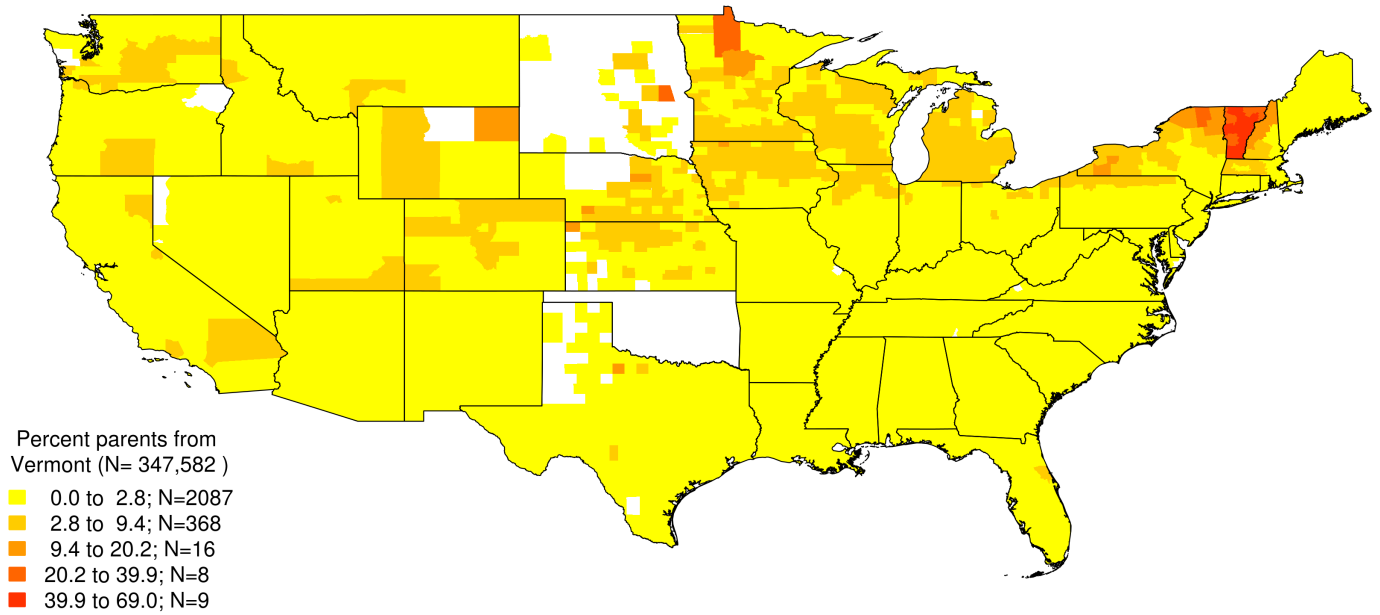

Figure 1.4: Vermont a exhibits the typical New England profile: westward through New York and into Michigan and beyond. Note how the migration shadow skims across northern Illinois, bypassing all of Missouri, but dipping down to cover Kansas. Like most Northern states, some migrants went to the Texas Panhandle, while avoiding the rest of Texas.

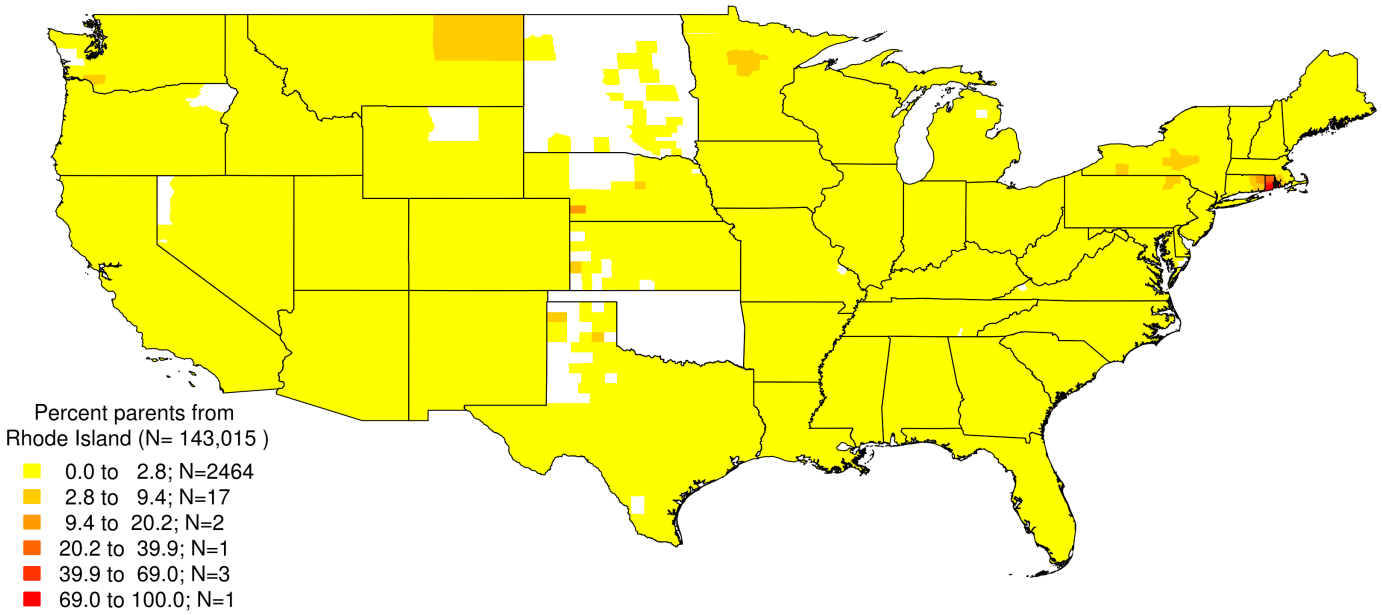

Figure 1.5: Rhode Island was too small to have much of a distinct migration shadow.

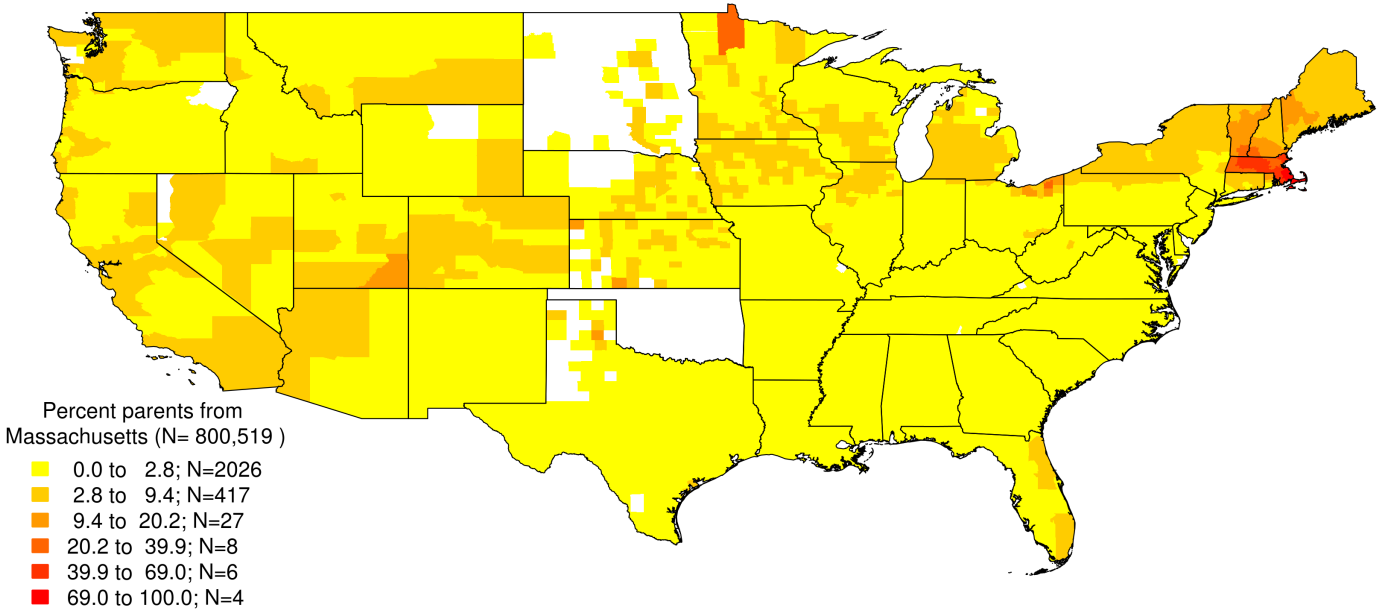

Figure 1.6: Massachusetts's migrants settled all through New England, and contributed to the population as far away as California. However, with the exception of Florida, they avoided the South, as did all other New England states.

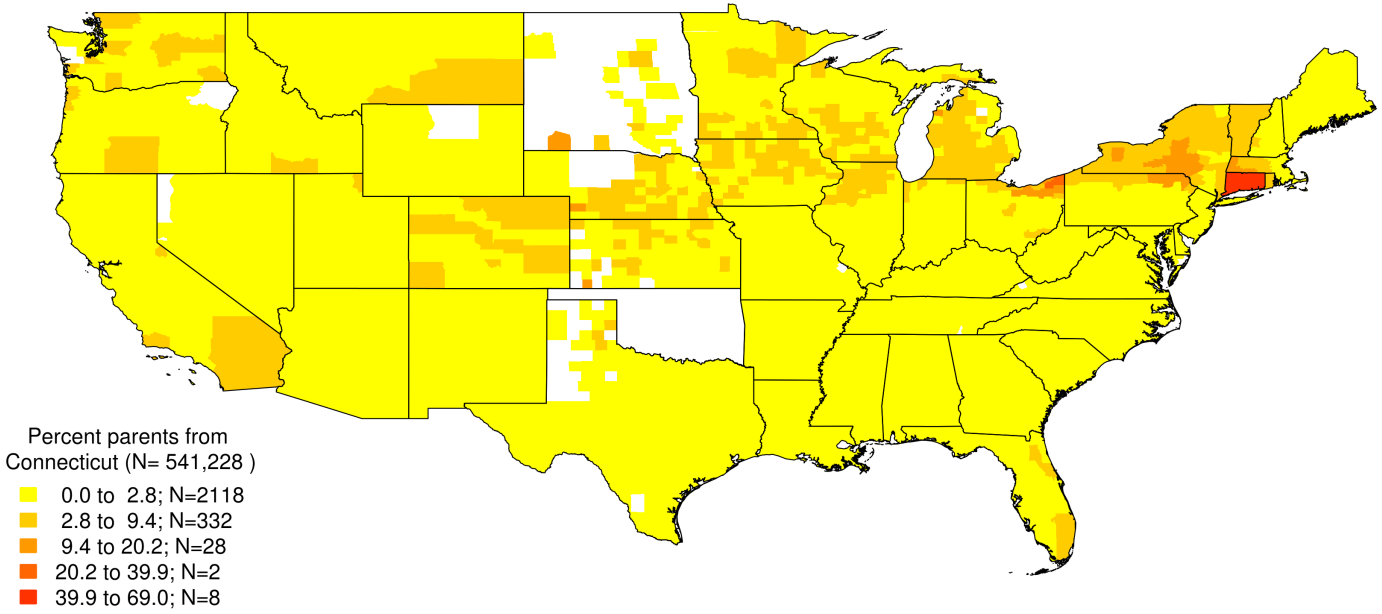

Figure 1.7: Connecticut's migrants did not have as much westward reach as those of Massachusetts, but were nevertheless of importance in New York and the Great Lakes states.

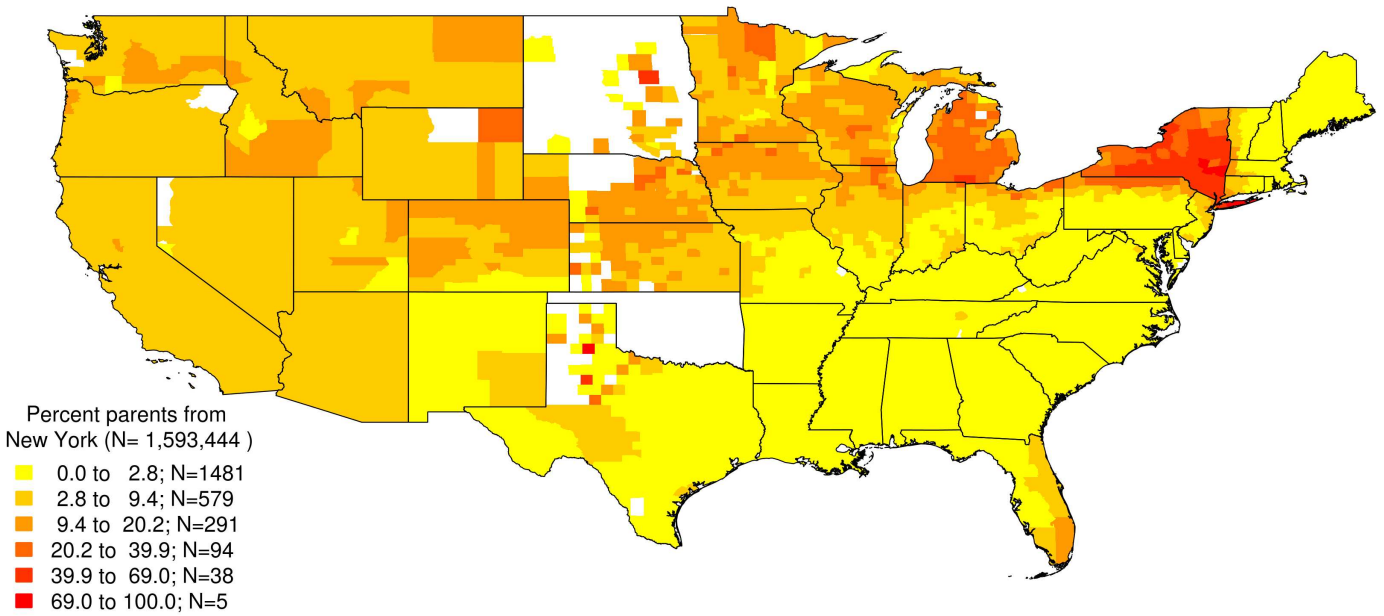

Figure 1.8: New York was primarily settled from New England and its migration shadow looks little different. Note how strongly the shadow hews to latitude until it reaches the Missouri River (the border between Iowa and Nebraska), when it fans out from the Dakotas to the Texas Panhandle. In the South, only the Texas Panhandle and Florida received significant migration from New York.

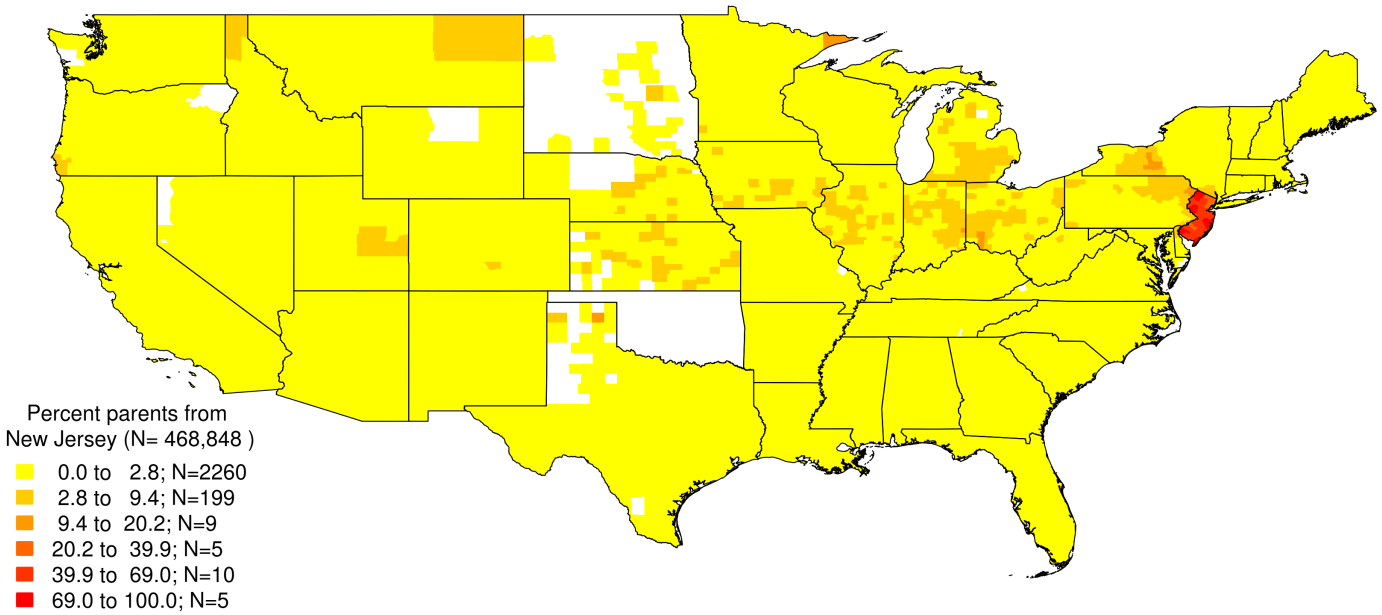

Figure 1.9: New Jersey shows a migration track further to the south than that of New York and New England. Migration shadows from the northern east coast all show a strong tendency to move due west, though New Jersey did send some migrants north into New York state. Note that New Jersey migrants are represented in Pennsylvania only along the northern border–a region also containing many migrants from New York.

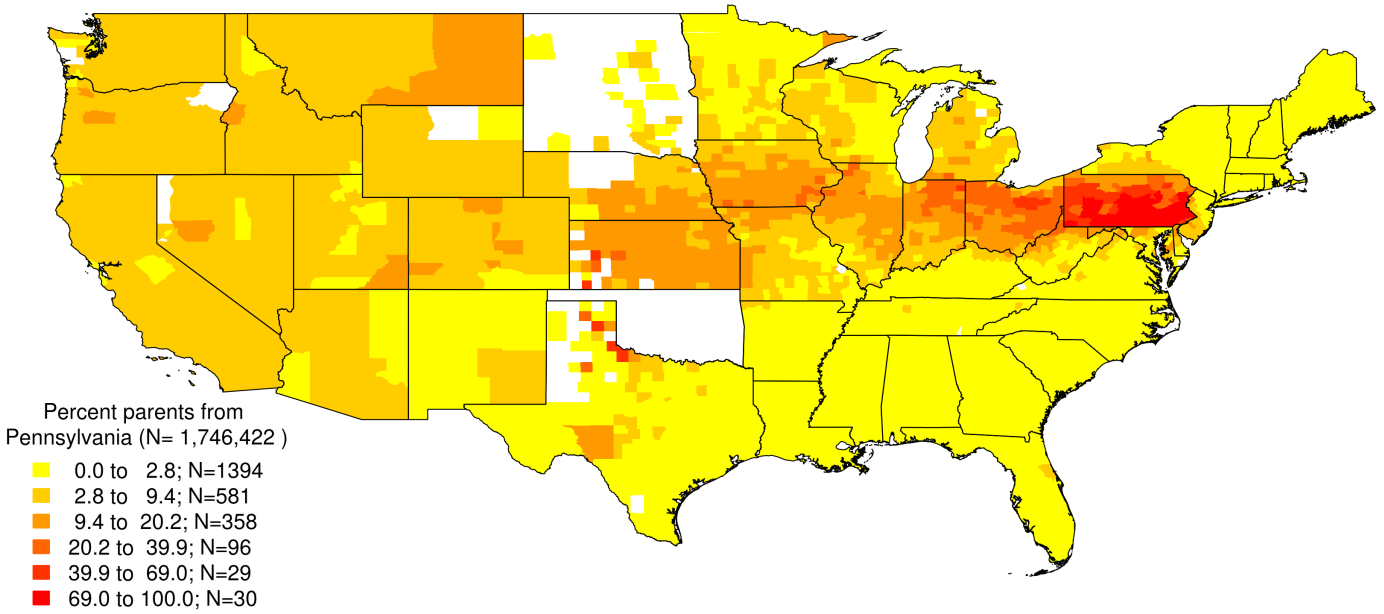

Figure 1.10: Pennsylvania is one of the major sources of westward moving settlers, who headed due west across Ohio, Indiana, and Illinois. Like the other northern states, it sent settlers to Kansas and the Texas panhandle, but not many to Missouri.

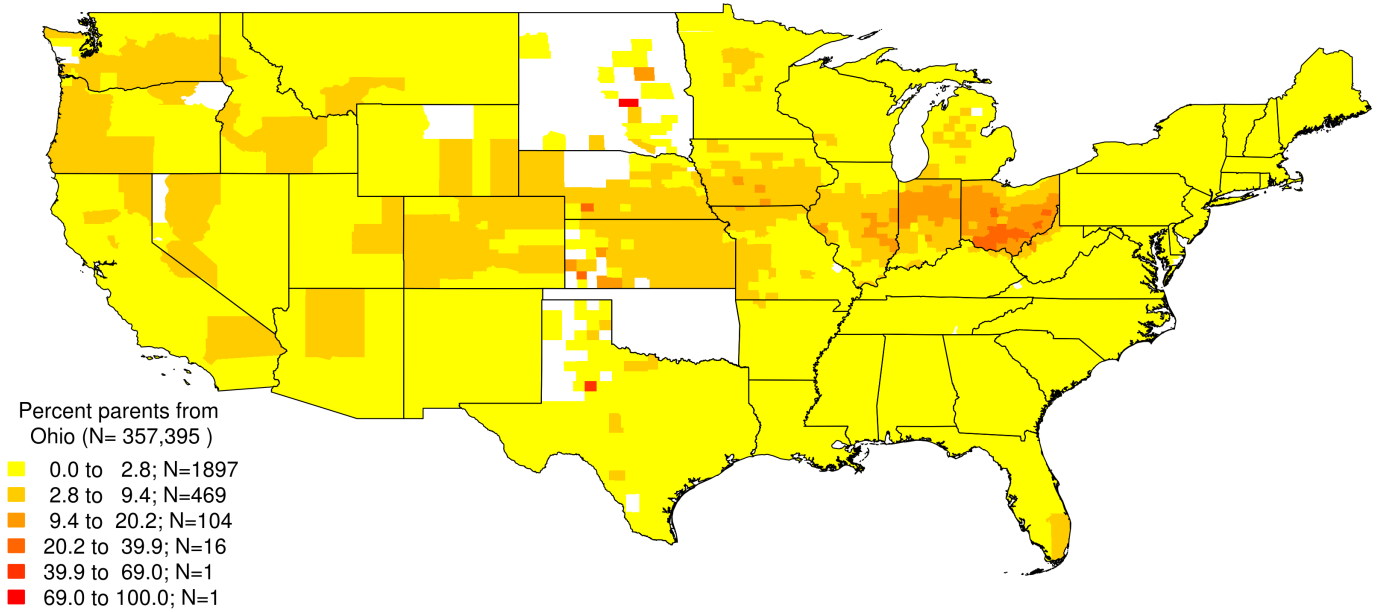

Figure 1.11: Ohio resembles Pennsylvania, following the line of latitude due west, largely avoiding Missouri but dipping south to Kansas.

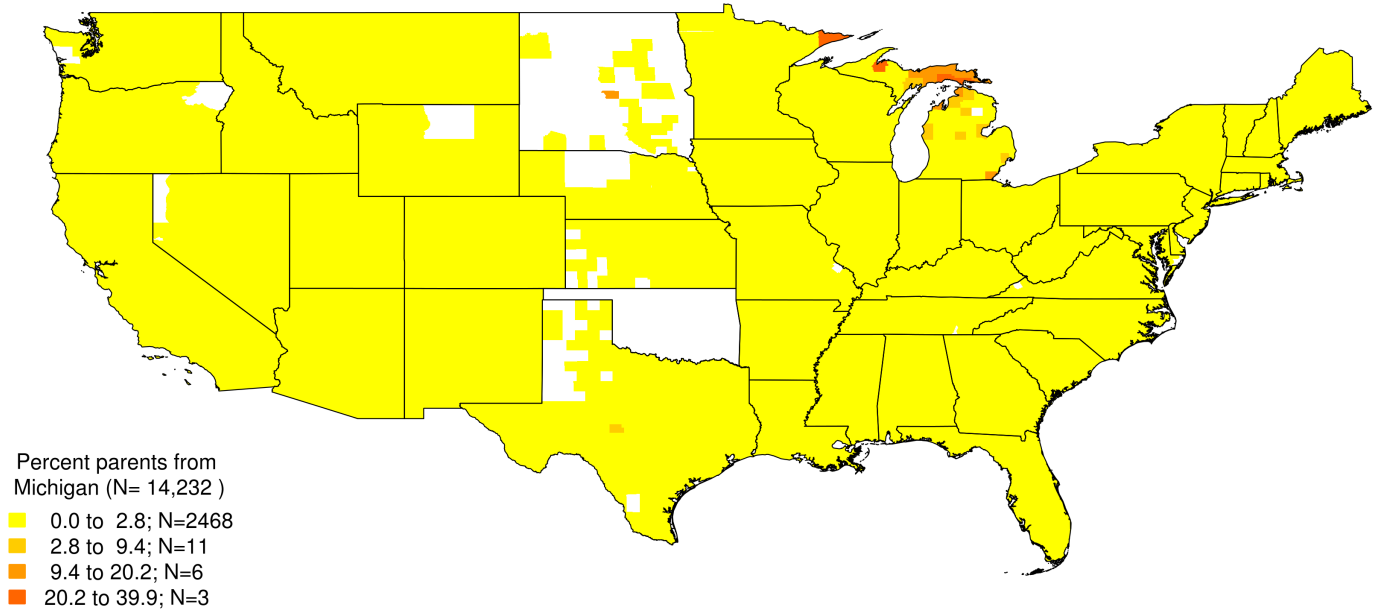

Figure 1.12: Michigan was sparsely settled around 1800 and most families established there tended to stay rather than migrate further.

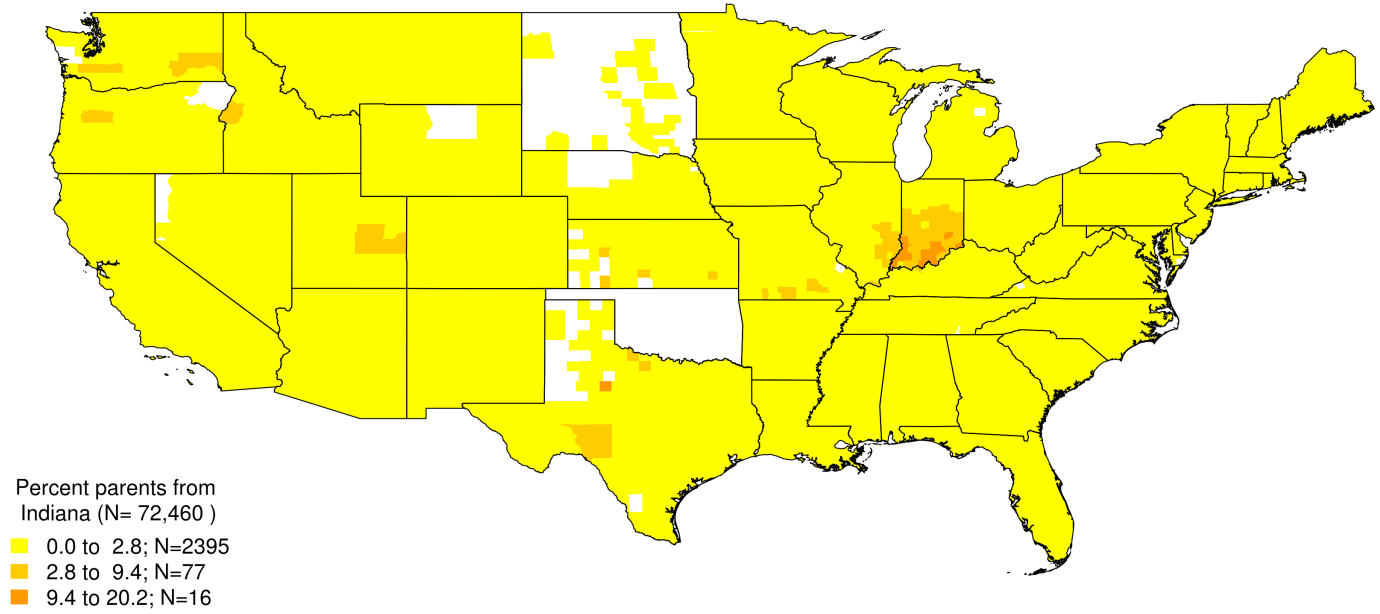

Figure 1.13: Indiana, like Michigan, was a destination state rather than a source: it sent few migrants westward, but those it sent did not avoid Missouri.

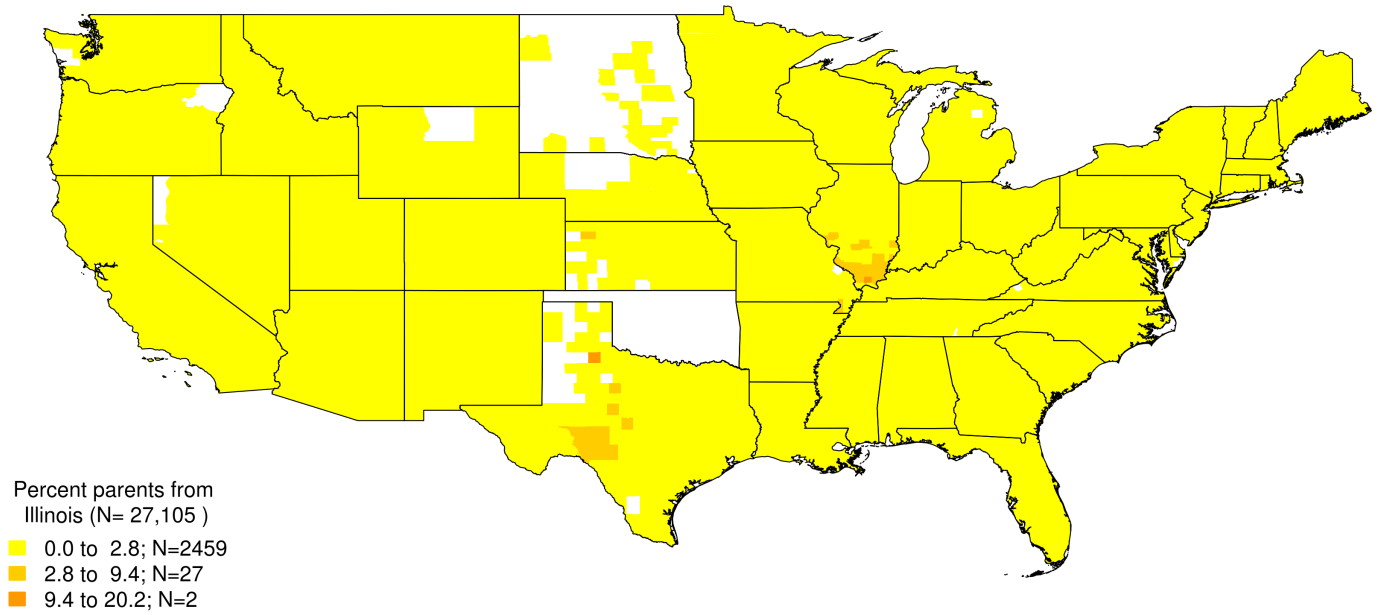

Figure 1.14: Illinois, like Indiana, sent few migrants westward. Even within the state, few had parents from Illinois.

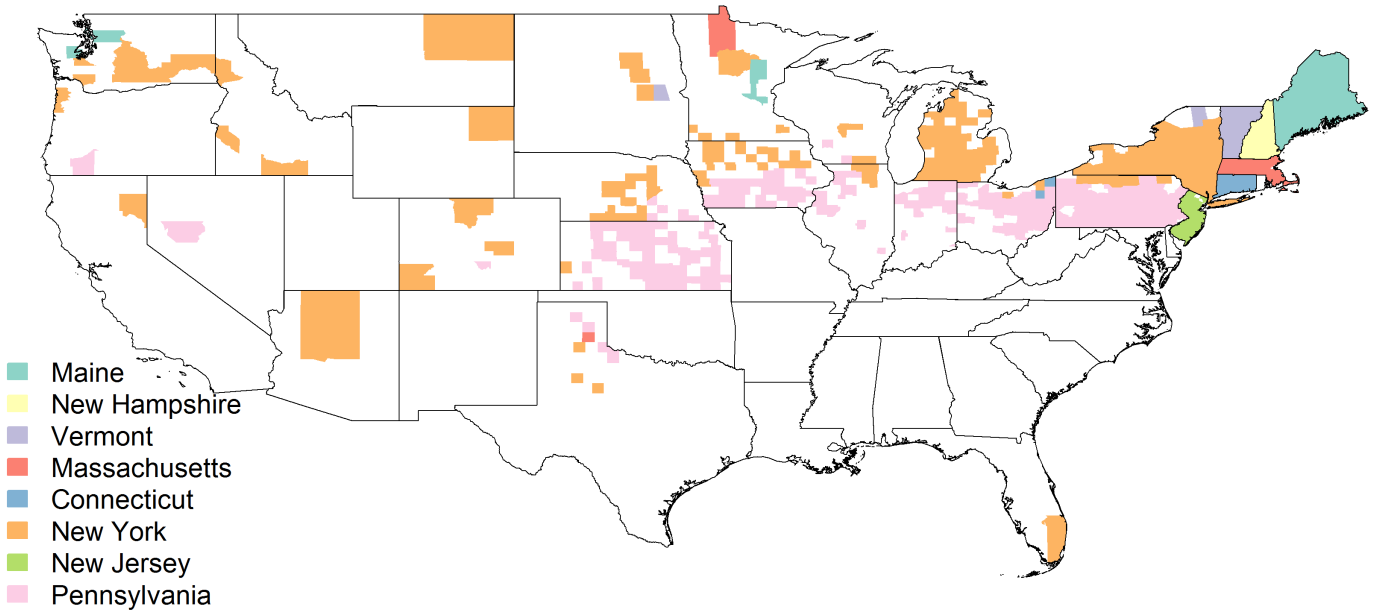

Figure 1.15: The state which makes the largest contribution to the population of each county is indicated. Only the principal Northern states are shown. New York and Pennsylvania were clearly dominant in the westward migration. Populations of the New England states were stable, showing no significant inmigration.

1.2 Southern states

Slavery was legal in all of these states up to the Civil War. Migration of African-Americans was often coerced, with slaves moving from the tobacco farming regions of Virginia and North Carolina to the cotton growing regions of the deep South and Mississippi Delta. As in the previous figures, each map shows the percent of a county's population coming from the specific southern source state. In cases where the whites coming from that specific state outnumber the blacks coming from that state, the map colors are shown in a yellow to red scale; in cases where the movement of blacks is predominant, the color scale runs from yellow to green.

Northern states sent few migrants to the South, but Kentucky and Virginia served as major sources for settlers to Ohio, Indiana, and Illinois.

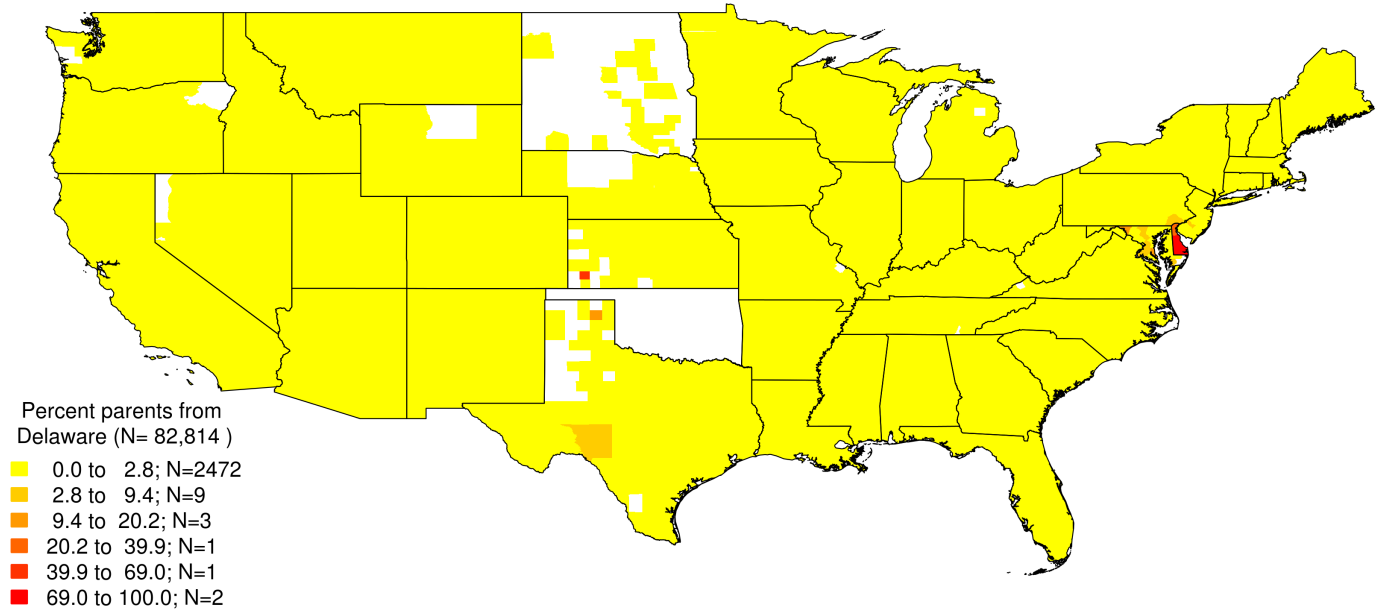

Figure 1.16: Delaware's westward migration was inconspicuous. Though slavery was legal there, it was rare, and there was virtually no movement of blacks to the cotton lands.

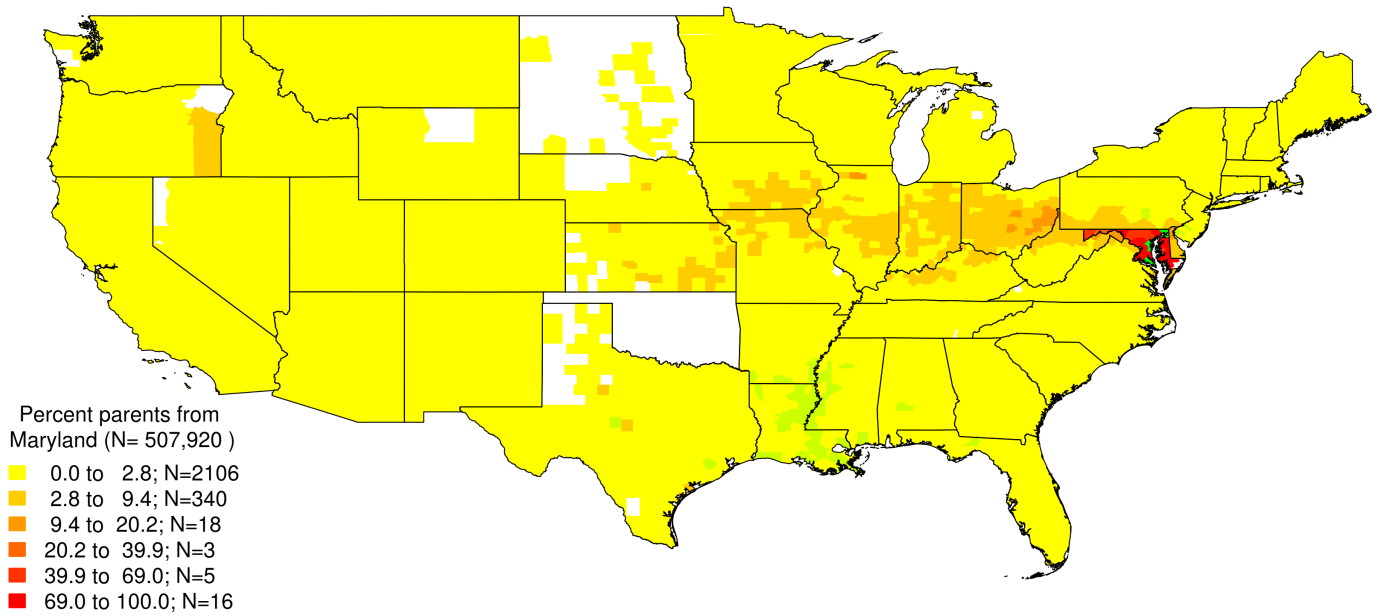

Figure 1.17: Maryland sends migrants due west, but it also sends migrants into slave states such as Kentucky and Missouri. The migration shadow across Missouri and into Kansas is a continuous carpet, unlike the northern pattern of skirting Missouri to reach Kansas. Counties where the largest proportion of Maryland settlers were white are shown in red scale; counties where the largest proportion were black are shown in green scale. Black migrants went to cotton lands along the Mississippi River.

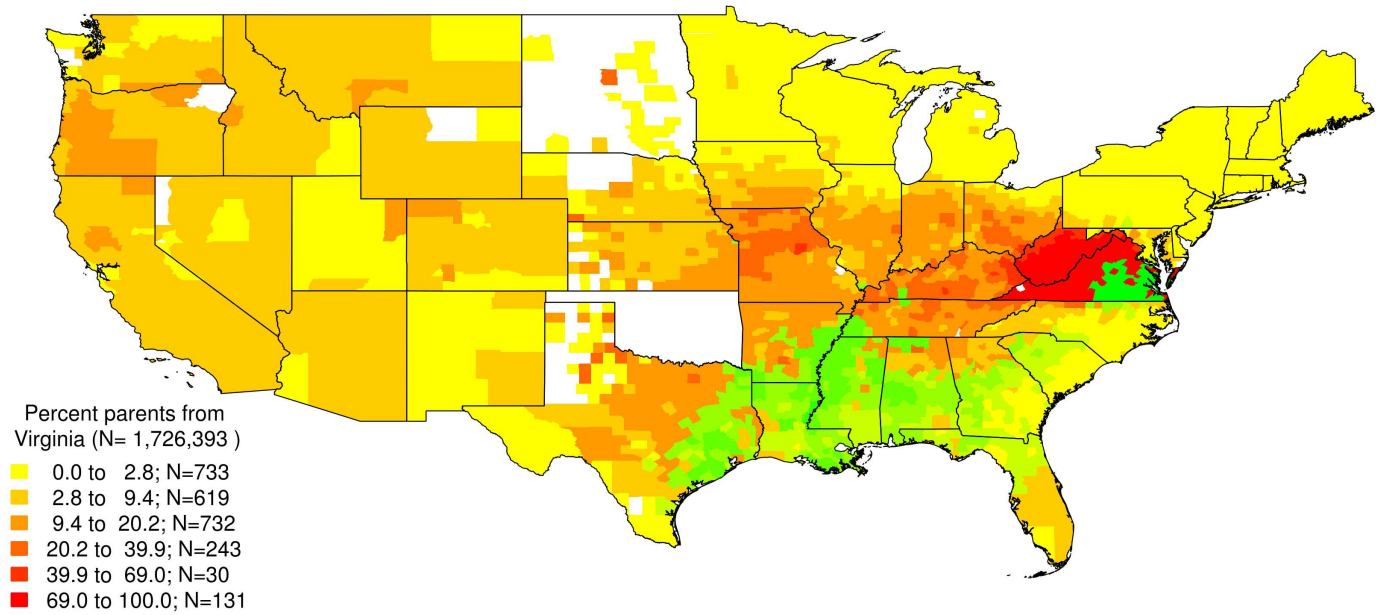

Figure 1.18: Virginia sent white migrants (in red scale) due west, who settled on both sides of the Ohio River. The state ranks with Pennsylvania as the biggest source of westward migrants, and like Pennsylvania sent settlers to Ohio, Indiana, and Illinois. Northern Missouri was a popular site of settlement. Black migrants (in green scale) were sent further south, to the cotton lands of the Mississippi Delta, east Texas, and southern Alabama and Georgia.

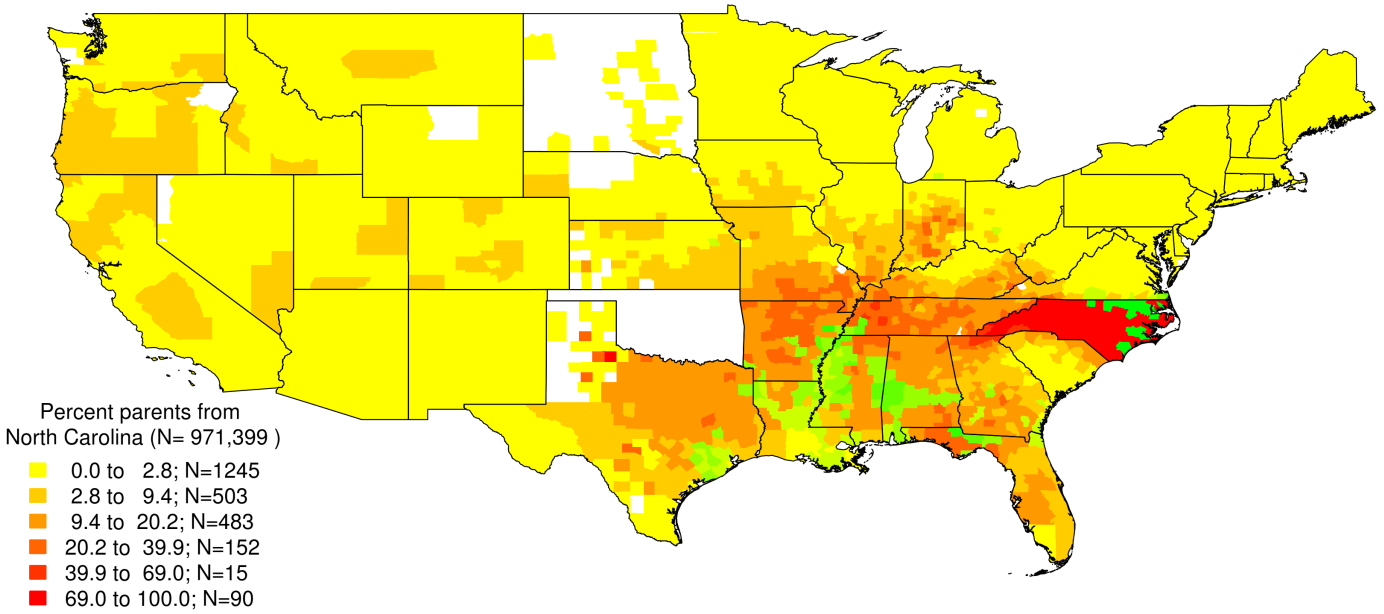

Figure 1.19: North Carolina was a major source of white settlers (in red scale) in Tennessee, Arkansas and Texas. A significant number were also found in southern Illinois and Indiana, and as far south as Florida. Most black migrants (in green scale) from North Carolina went to the cotton lands.

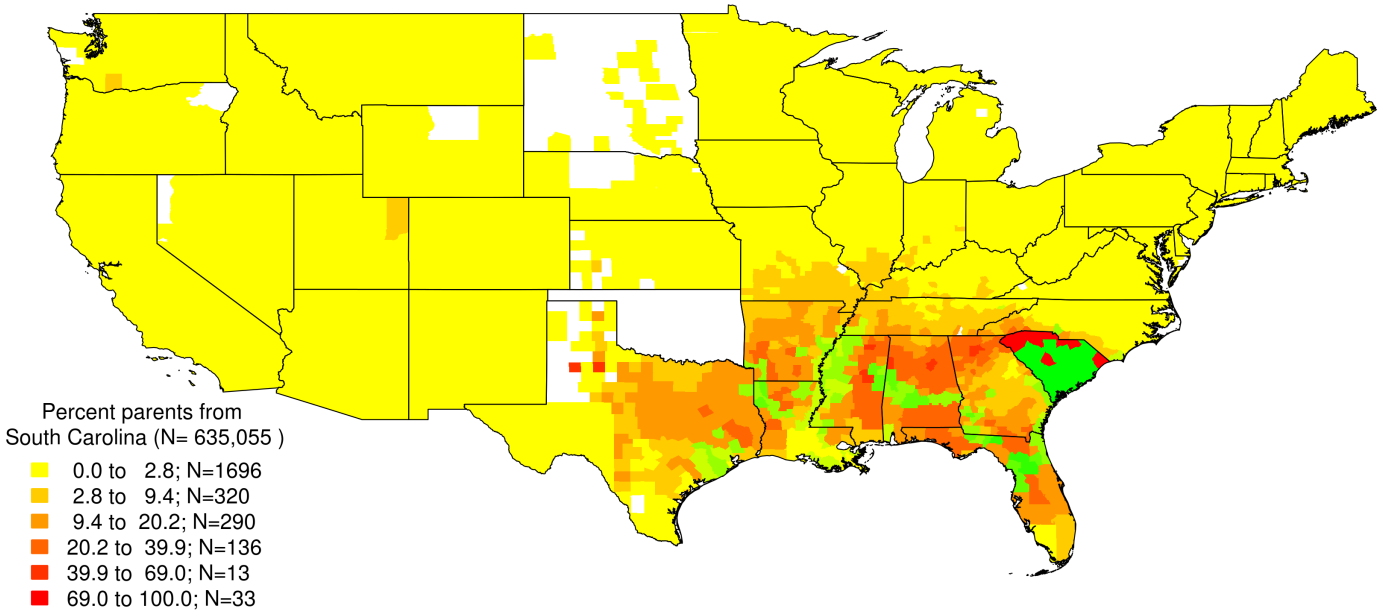

Figure 1.20: South Carolina shows a migration shadow one level to the south of North Carolina. Movement was to the west and south, with central Georgia bypassed. Many settled in Alabama, Mississippi, Arkansas, Texas, and even Florida.

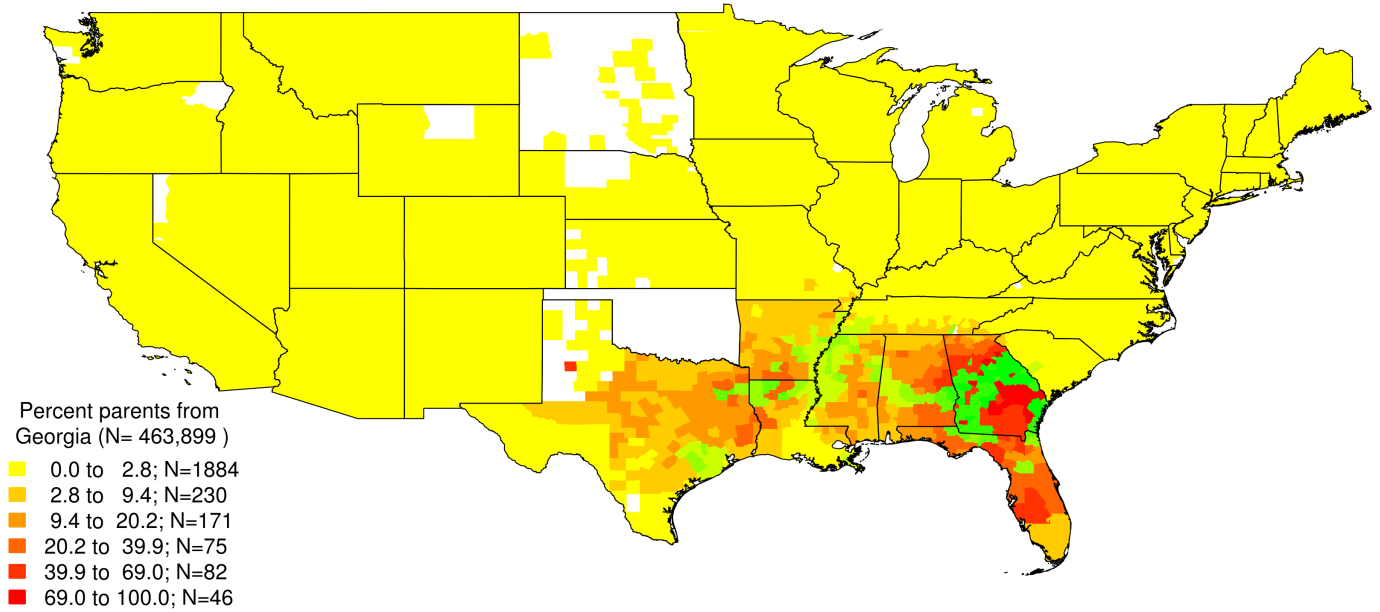

Figure 1.21: Georgia casts a significant migration shadow onto Florida and eastern Alabama, but all migration is to the west or south, a pattern holding true for all the deep South states.

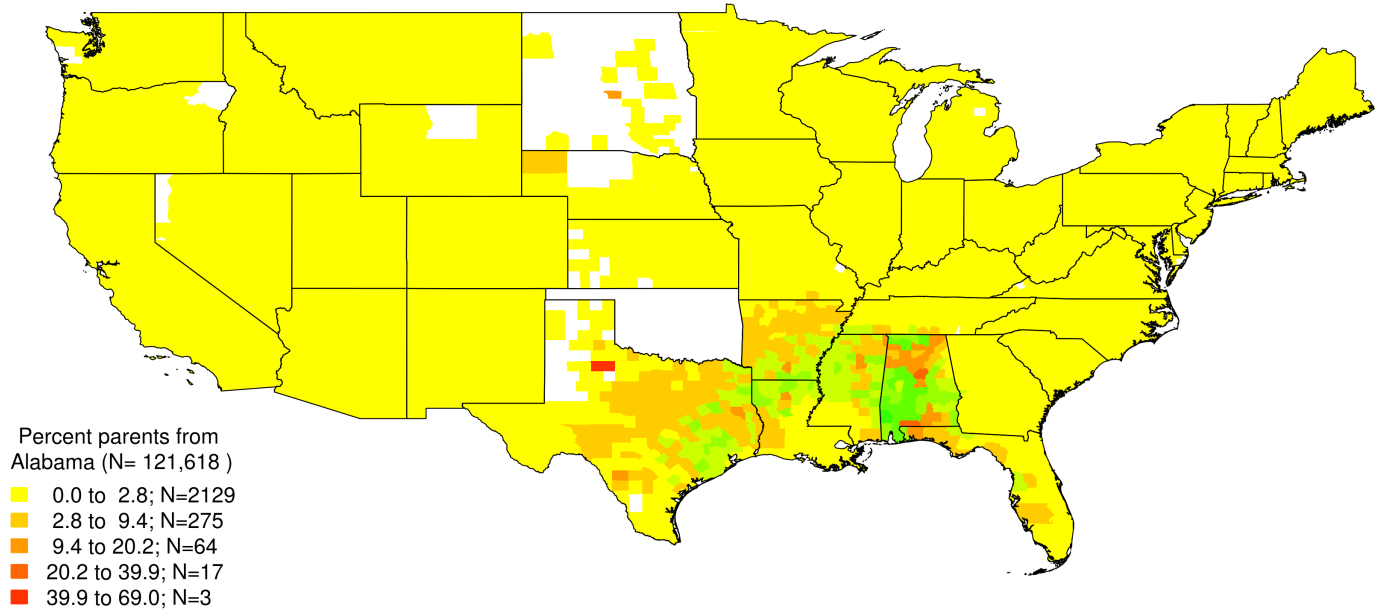

Figure 1.22: Alabama casts its migration shadow to the west and south. Like Indiana, it was primarily a destination state and did not send large numbers westward.

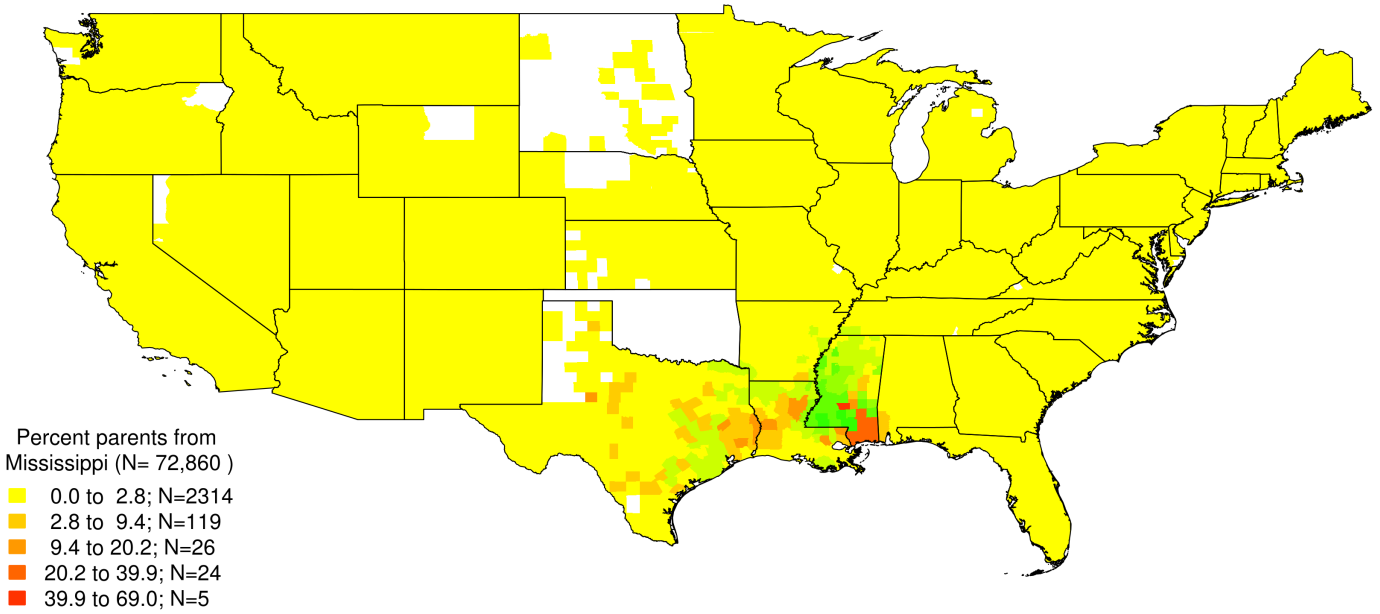

Figure 1.23: Mississippi follows the pattern of all the deep South states: a migration shadow only to the west, stretching as far as central Texas. Mississippi was more of a destination than a source, however, and it sent relatively few westward.

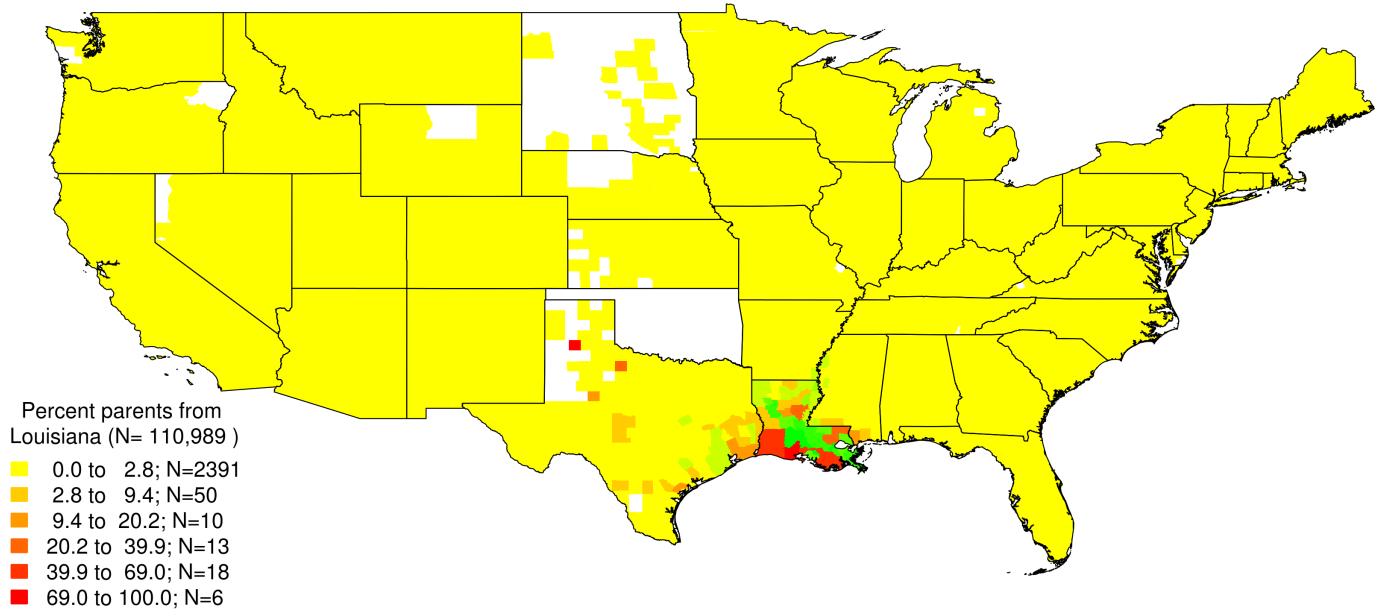

Figure 1.24: Louisiana sent relatively few migrants into Texas. The population of the southern portion remained fairly stable, attracting few migrants.

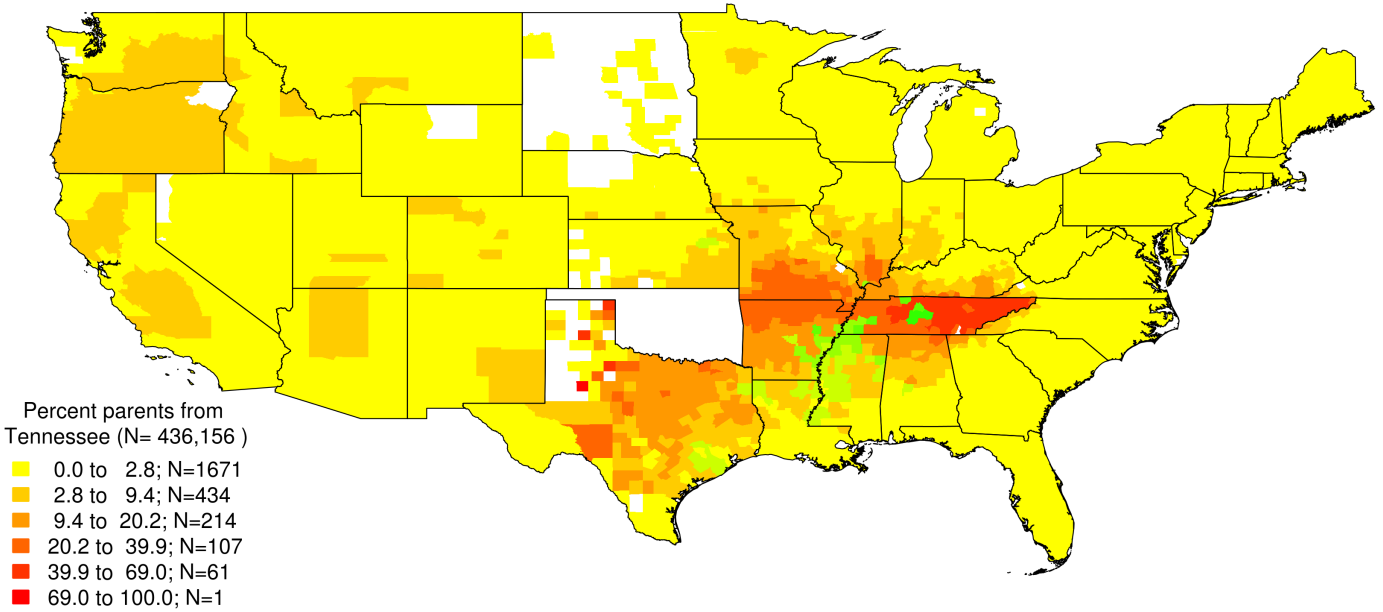

Figure 1.25: Tennessee cast a migration shadow across Arkansas and southern Missouri, and down to Texas. There was also some significant northward migration, into western Kentucky and southern Illinois.

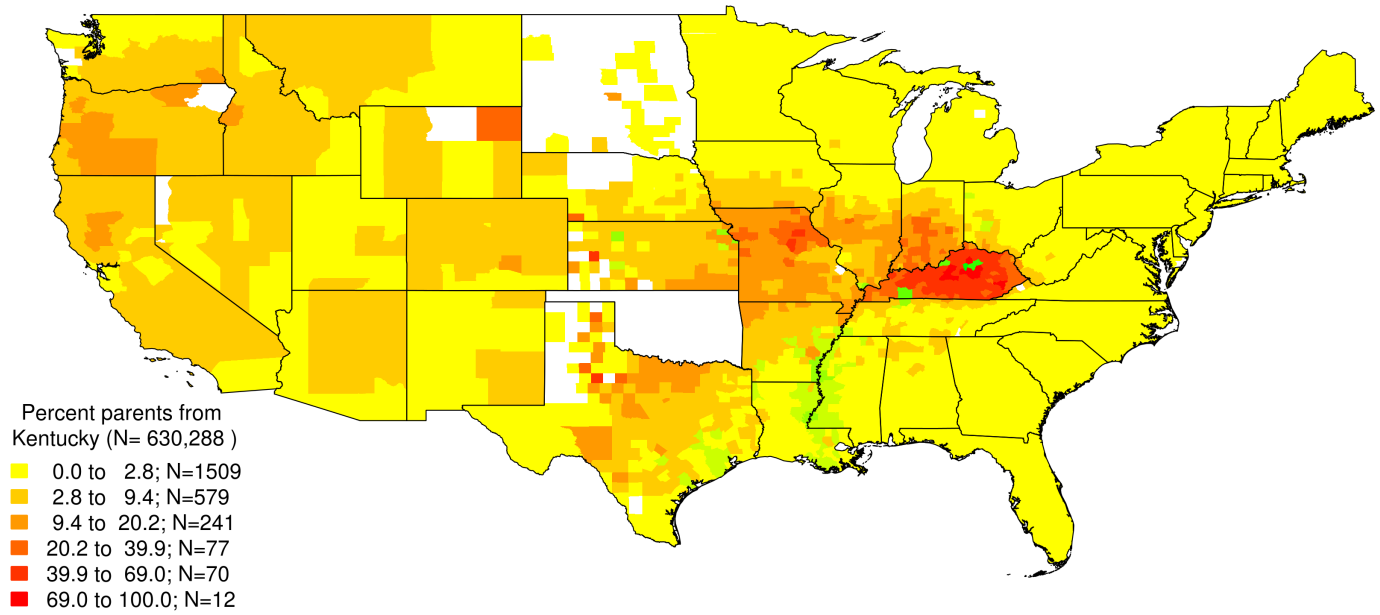

Figure 1.26: Kentucky sent migrants westward, to Missouri, as well as northward into Indiana and Illinois.

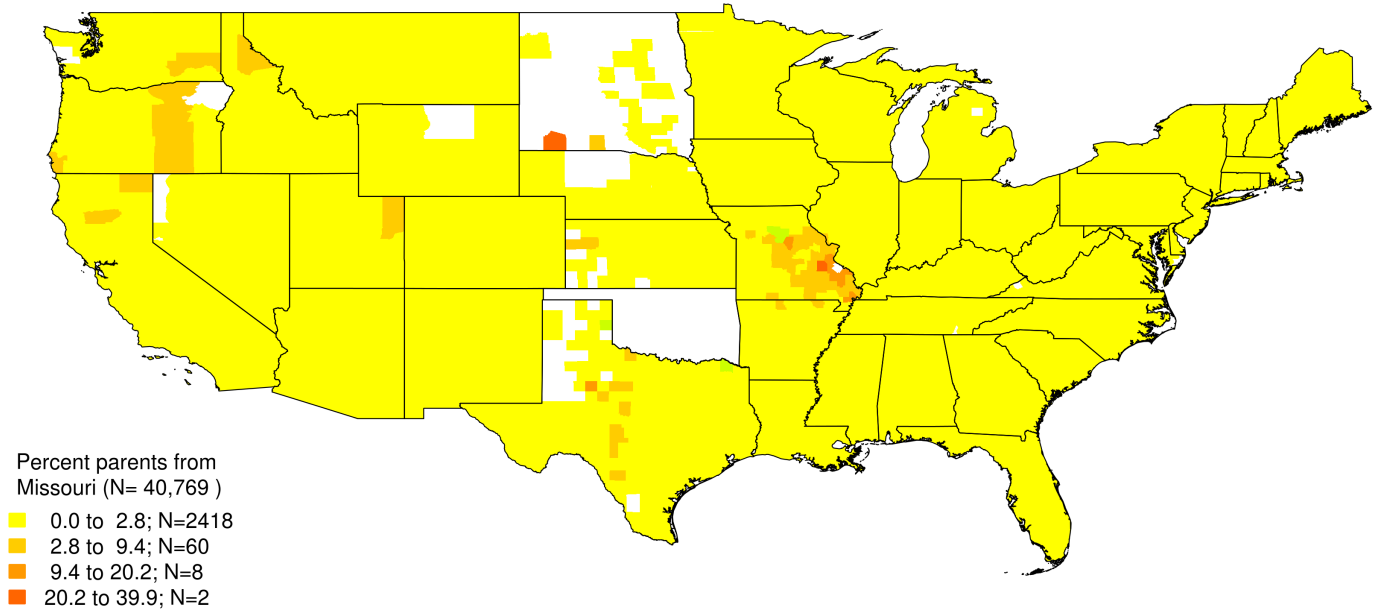

Figure 1.27: Missouri sent few migrants elsewhere, though the eastern part of the state contained a fair number of people whose parents were born there before 1820.

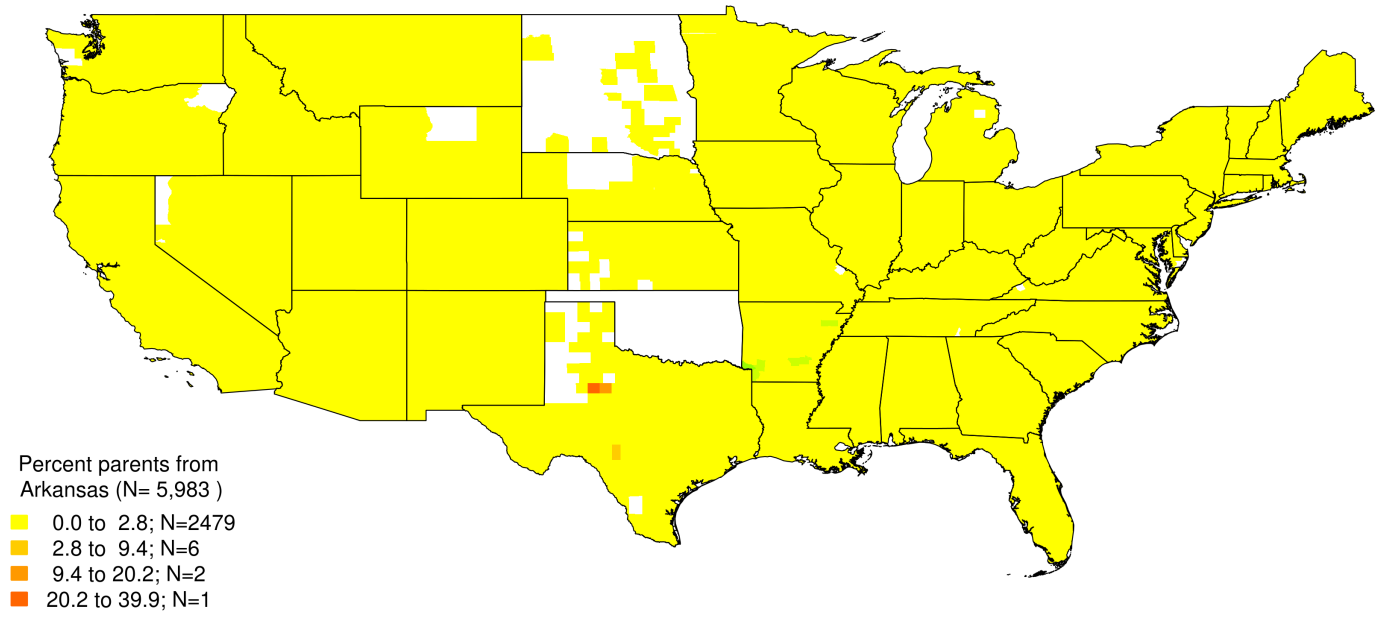

Figure 1.28: Arkansas, like Missouri and Texas, was a destination state and sent very few migrants elsewhere.

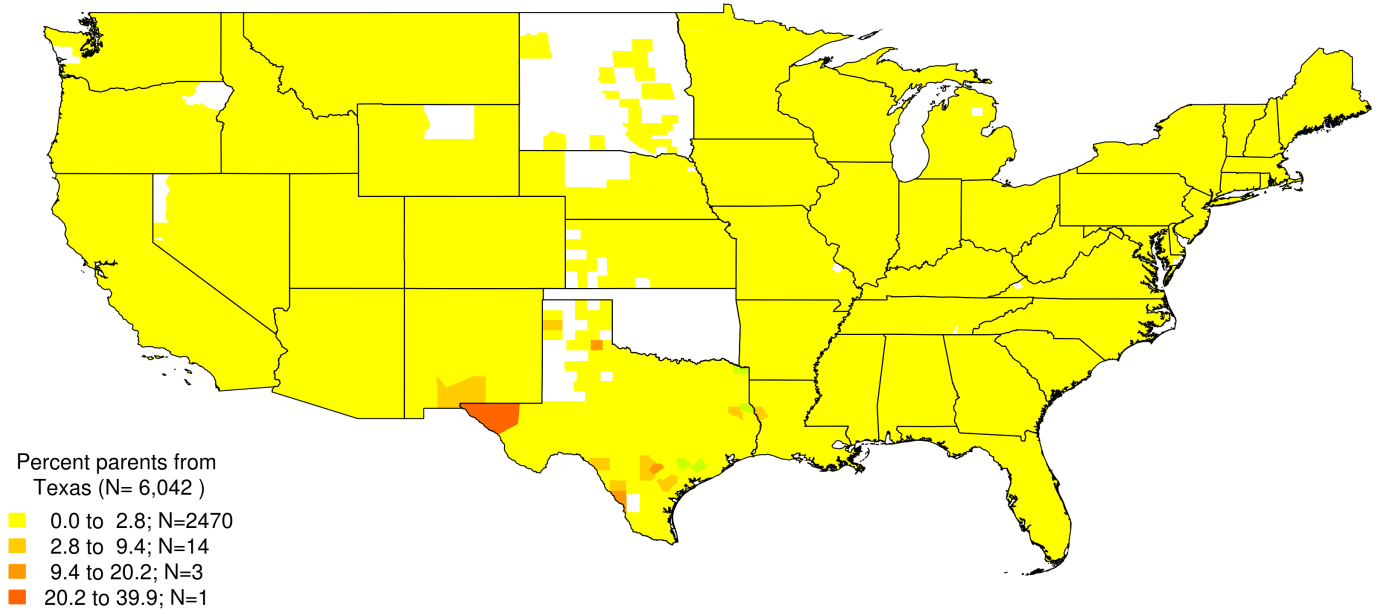

Figure 1.29: Texas had only a small population with parents born in the state before 1820, and this population is only significant in a few locations, most notably El Paso and San Antonio.

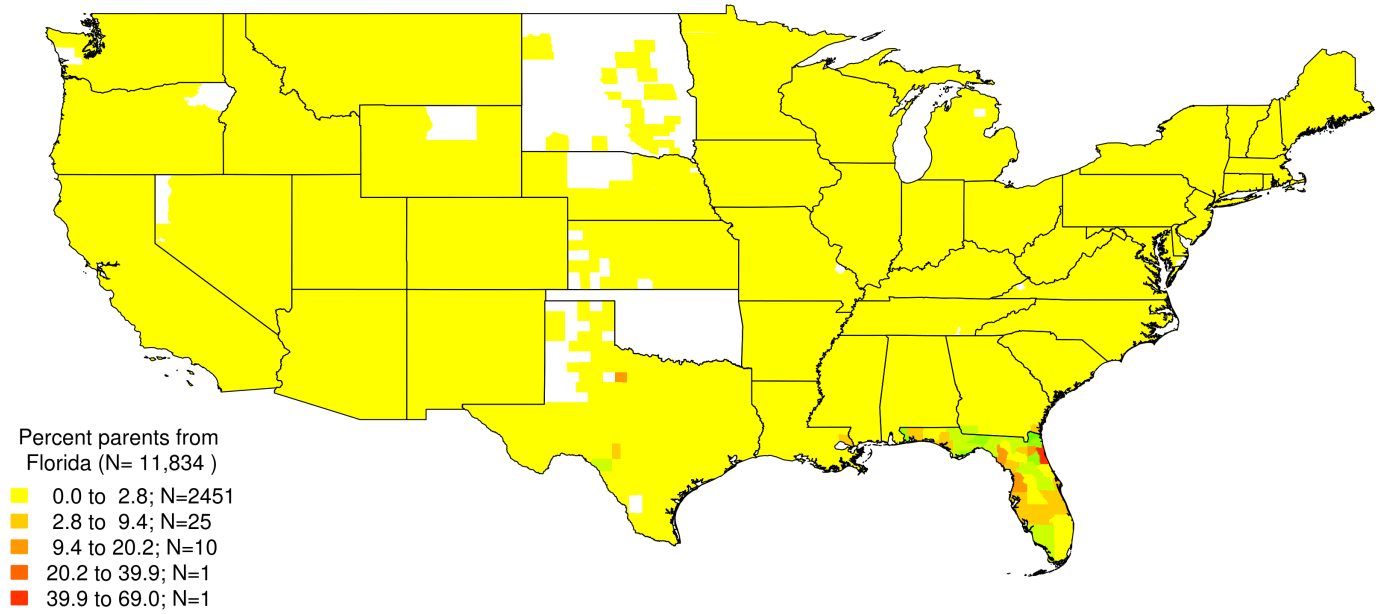

Figure 1.30: Florida sent very few migrants elsewhere. Like Texas, its long-established population may have been Latino, since Saint Augustine stands out as an area with a population whose parents were mostly born in the state before 1820.

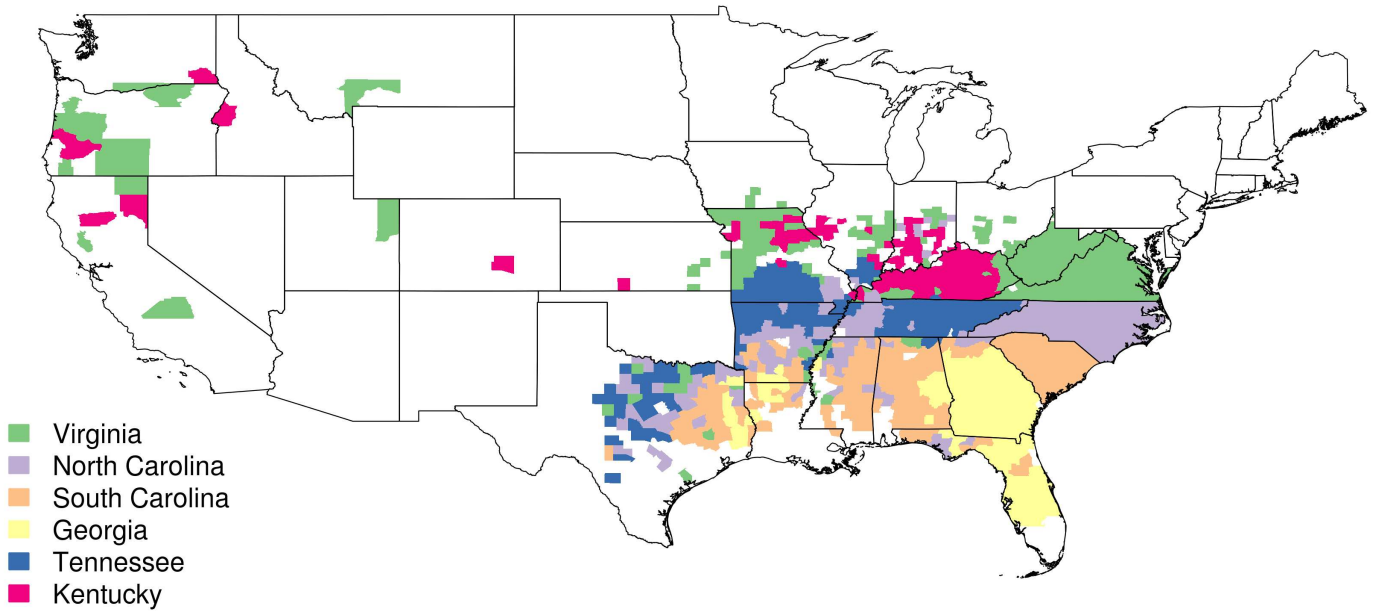

Figure 1.31: Counties where the principal Southern states are the top contributors to the white population. Tennessee and North Carolina played the lead role in settling the Ozarks and north central Texas. Northern Missouri was mostly settled from Virginia and Kentucky, a role the two states also played in southern Indiana and Illinois. South Carolina was the main source of settlers in Alabama, Mississippi, southern Arkansas, and east Texas. Georgia was the lead source in the settlement of Florida.

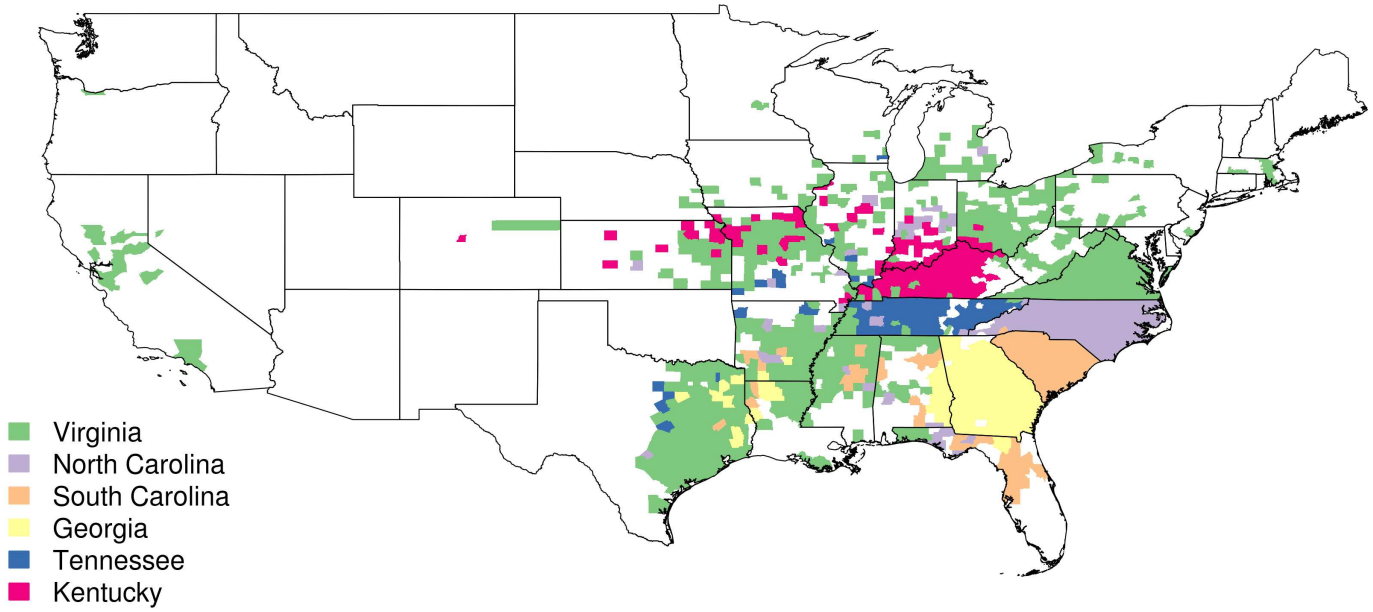

Figure 1.32: Counties where the principal Southern states are the top contributors to the black population. Virginia is by far the predominant source of black population, even in areas where white Virginians were not the main settlers. The Ozarks and north central Texas are blank since few blacks migrated to those areas.

1.3 Foreign countries

Foreign migration largely avoided the South, except for the coerced immigration of Africans. The largest white influxes were the Irish and Germans. The Chinese were concentrated in the West, as far east as the Rocky Mountains. Mexican and Canadian immigration clung near the borders. The Mormon settlements in Utah attracted British and Scandinavian converts. Germans and Scandinavians were most common in the western Great Lakes. Foreign migration played a large role in settling the West.

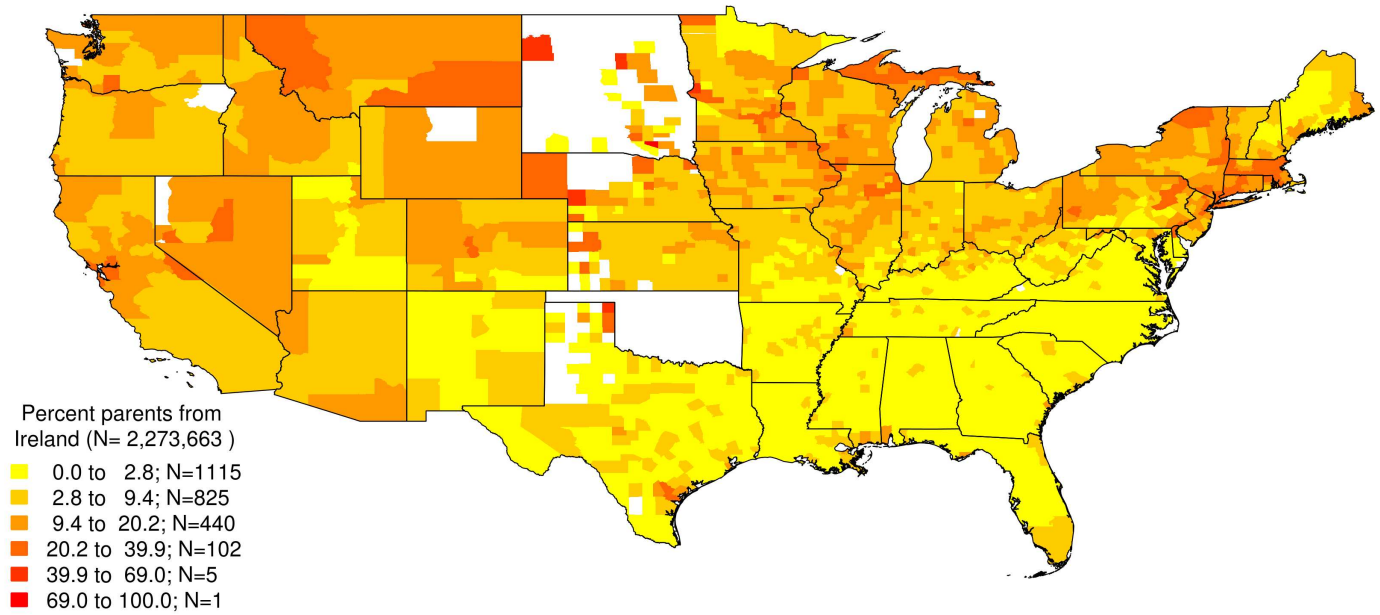

Figure 1.33: Ireland sent two kinds of immigrants: Protestants during the early to mid 18th century and Catholics beginning in 1840. The map shows persons whose parents were born in Ireland before 1820; most of these would be from the immigration of the 1840s and later. They are heavily represented in the West and near the northeast coast, and largely absent from the South.

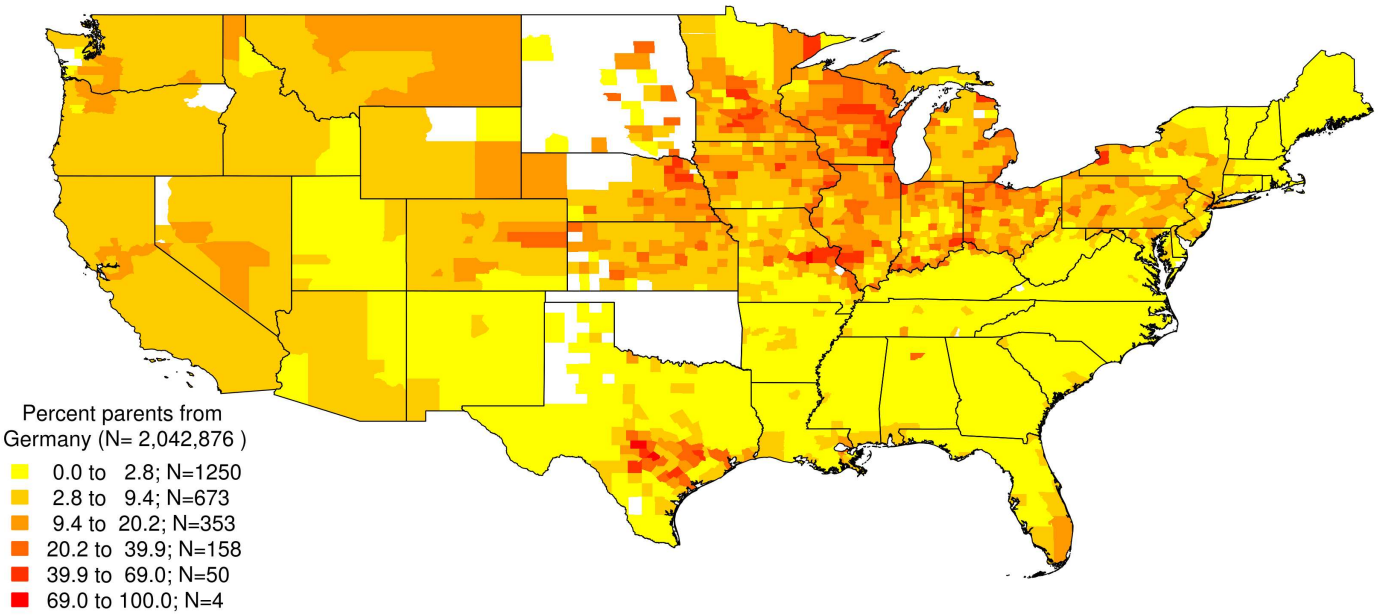

Figure 1.34: Germany sent many settlers during colonial times to Pennsylvania. The Germans shown on this map came during the 19th century. Unlike the Irish, they are found most heavily in the midWest. Like the Irish, they settled in the West and largely avoided the South, with the exception of the large German population that settled in central Texas. In Missouri, their settlements drove a wedge between the Tennesseans to the south and the Virginians and Kentuckians to the north.

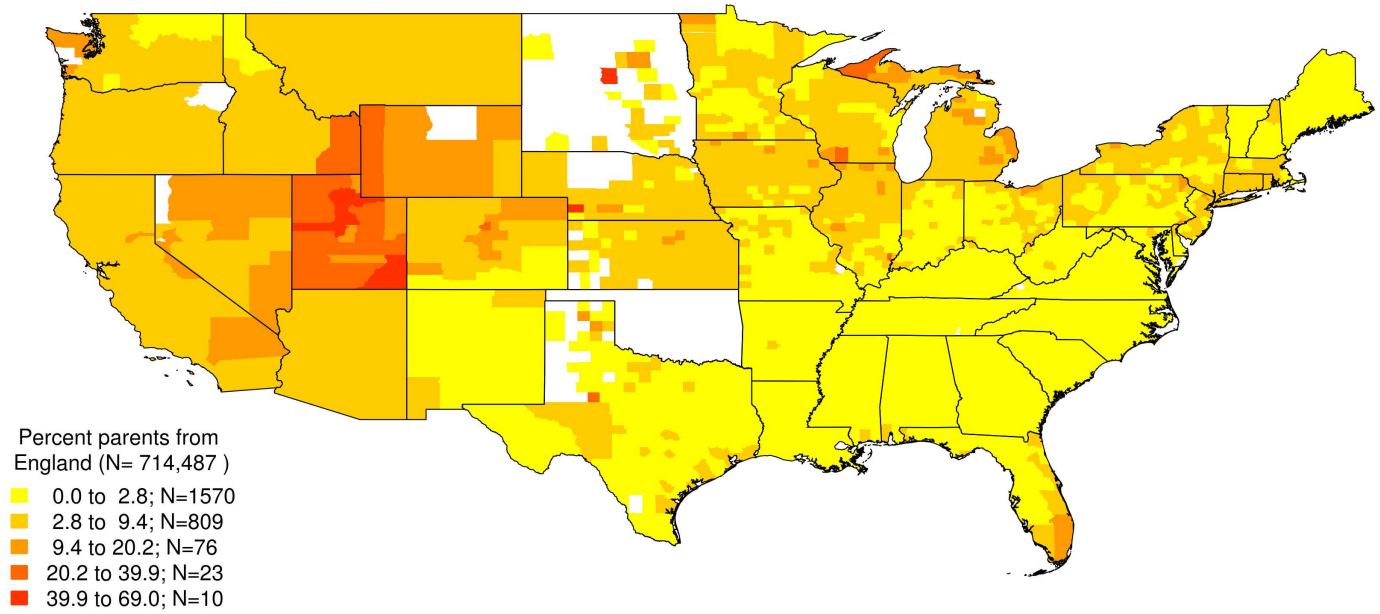

Figure 1.35: England was the main source of the first British settlers, but by the last half of the the 19^{th} century the numbers of immigrants were only a third of those of the Irish or Germans. They were especially predominant in Utah, since Mormon proselytization had been successful in Britain.

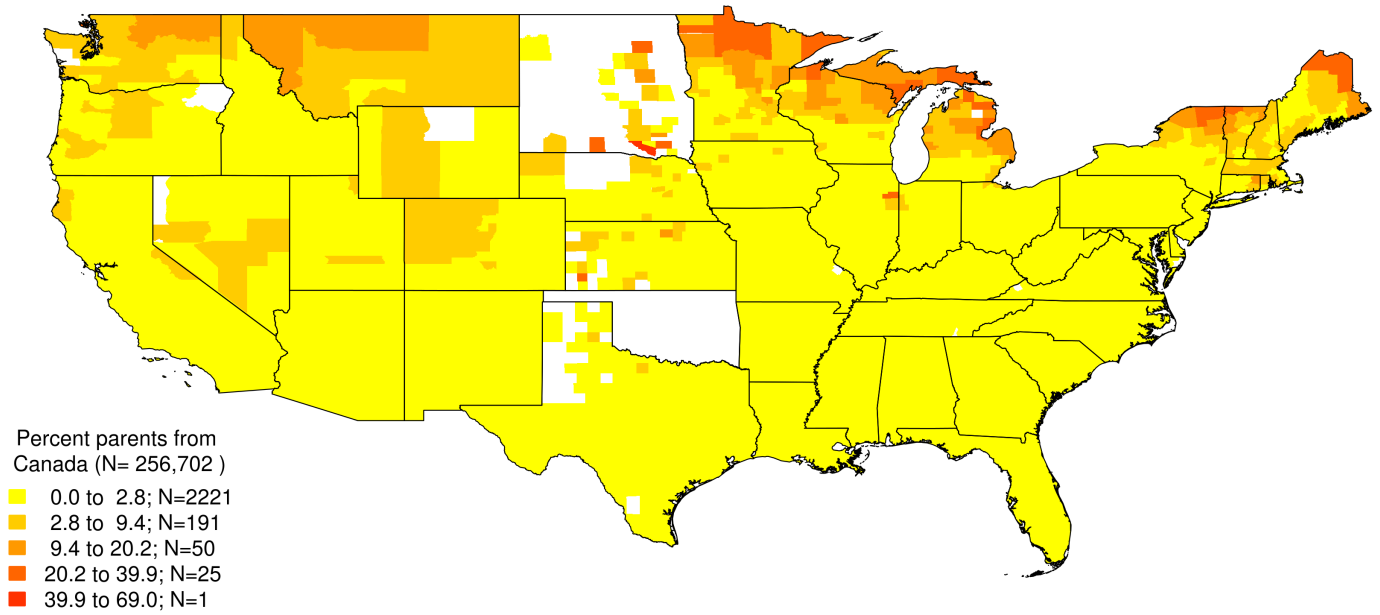

Figure 1.36: Canada sent migrants who tended to settle close to the Canadian border.

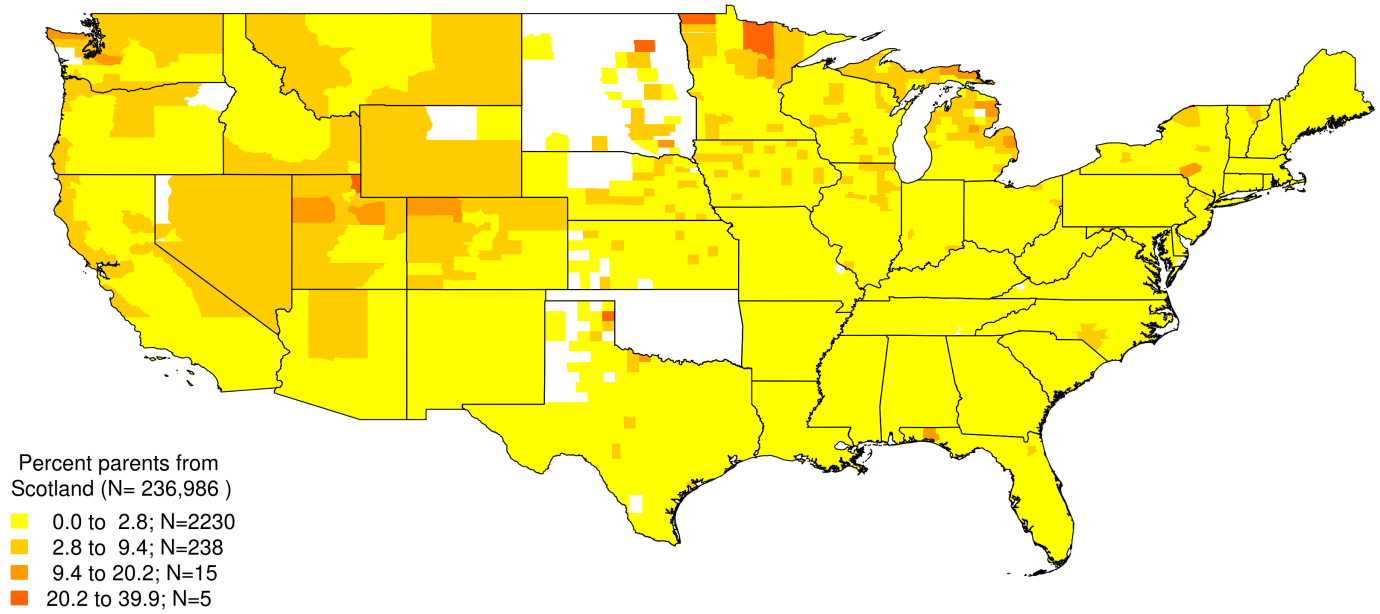

Figure 1.37: Scotland, like England, sent settlers to Utah, but in smaller numbers. Like other foreigners, Scots are heavily represented in the West, including as far east as Minnesota and Wisconsin. Note a small concentration near Scotland County, North Carolina, an area settled by Scots in the 18th century, an indication perhaps of homophily.

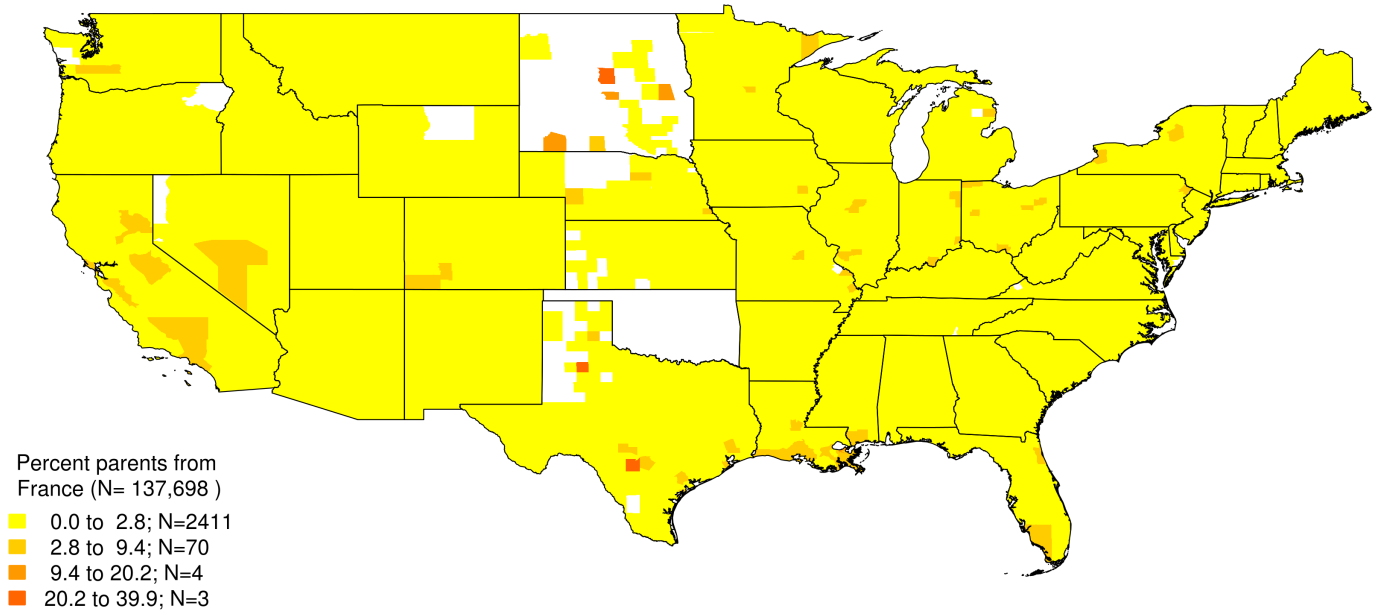

Figure 1.38: France sent a large number of migrants, but these mostly dispersed, leaving few areas of concentration. Nevertheless, there is a slight concentration of French migrants in southern Louisiana, another sign that migrants may prefer to settle among people similar to themselves.

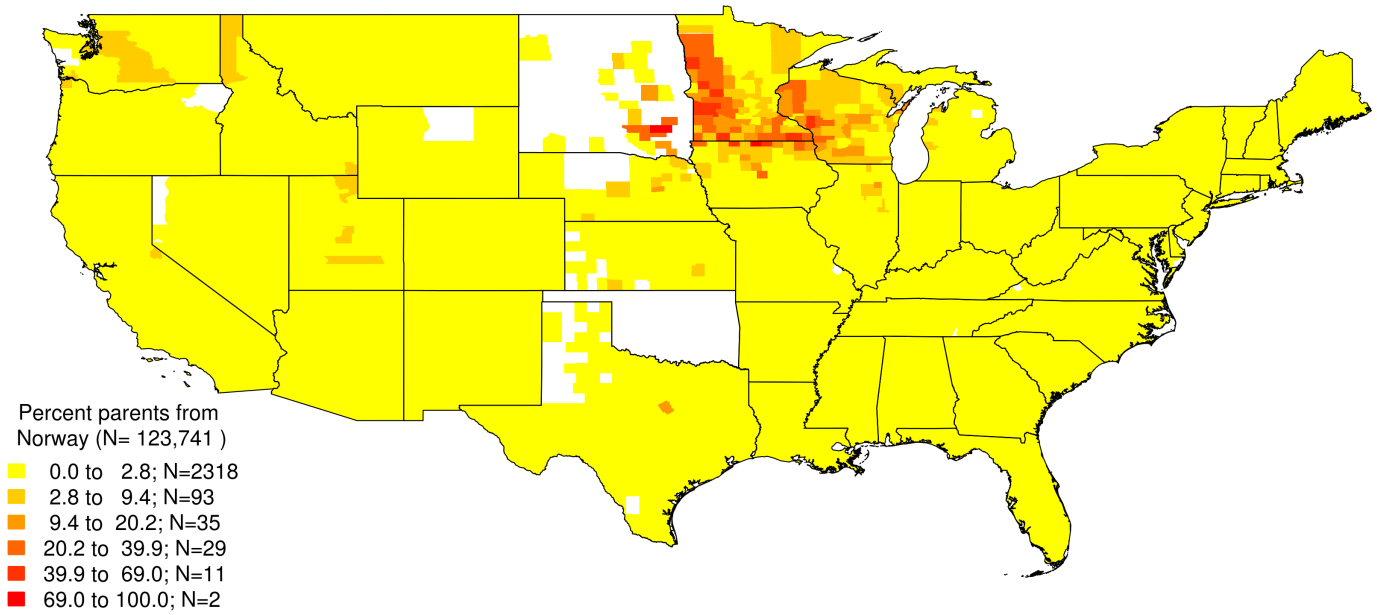

Figure 1.39: Norway has a highly concentrated presence in Minnesota and Wisconsin.

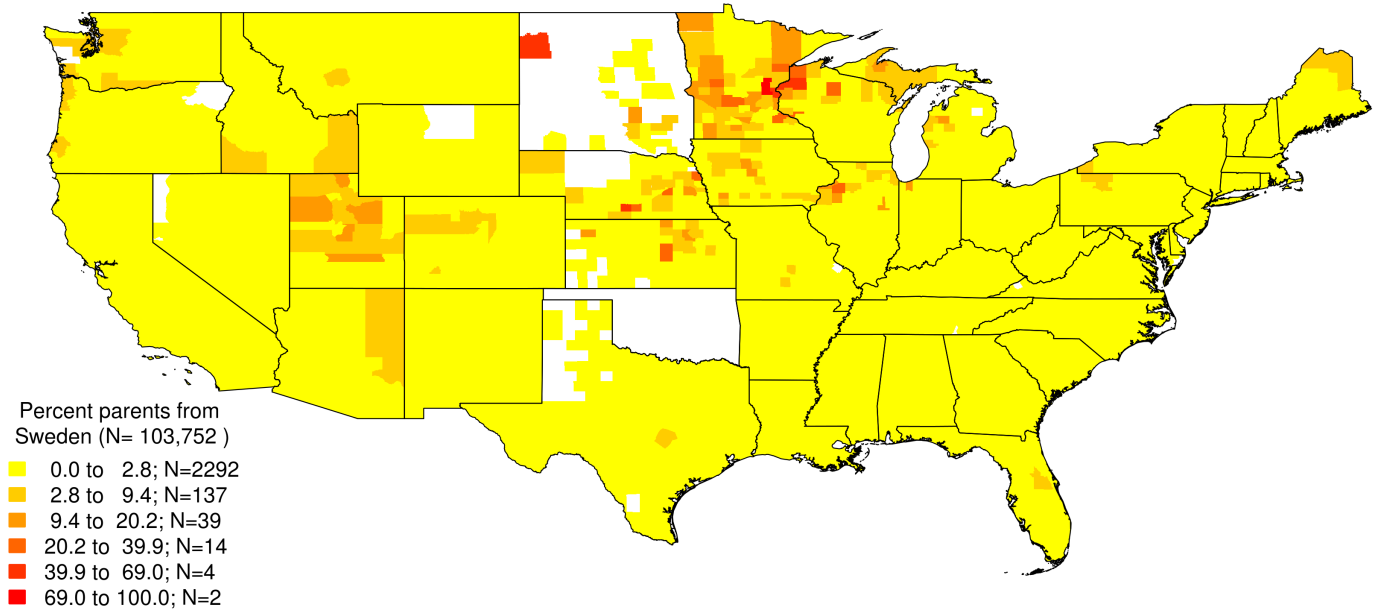

Figure 1.40: Sweden resembles Norway in that it has a strong presence in Minnesota, but it also has a strong presence in Utah, since Mormon proselytization was successful among Scandinavians.

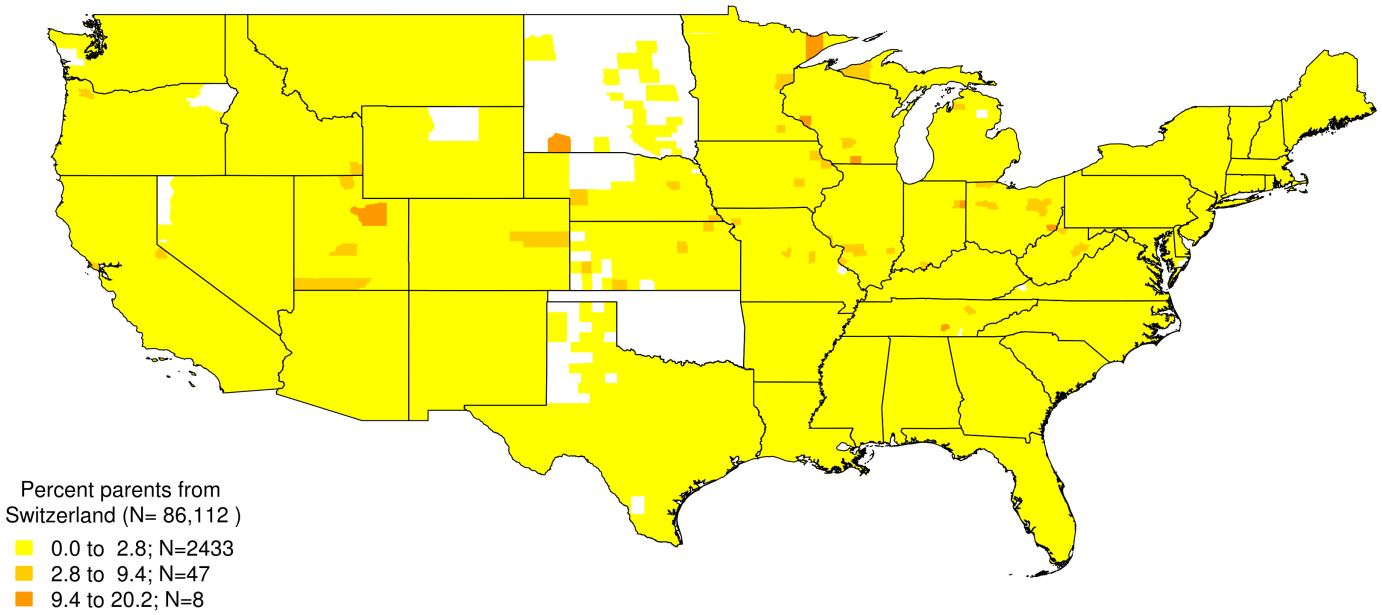

Figure 1.41: Switzerland sent migrants who dispersed, leaving no centers of concentration.

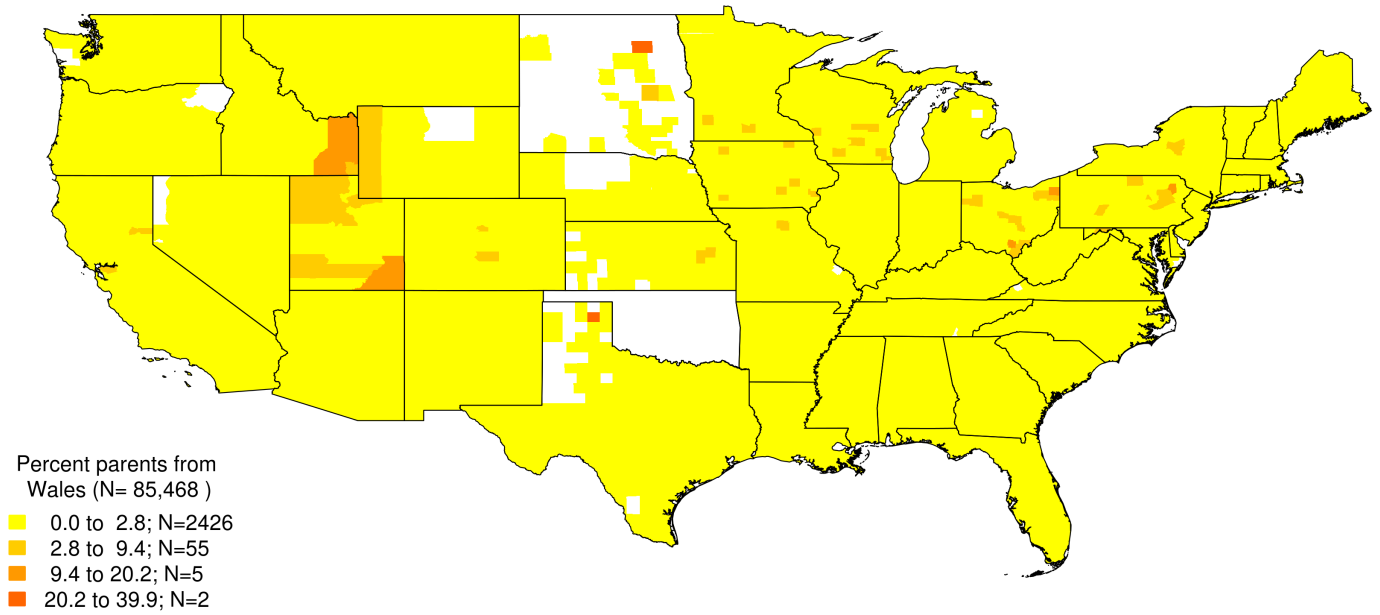

Figure 1.42: Wales resembles England and Scotland in that it sent migrants to Utah.

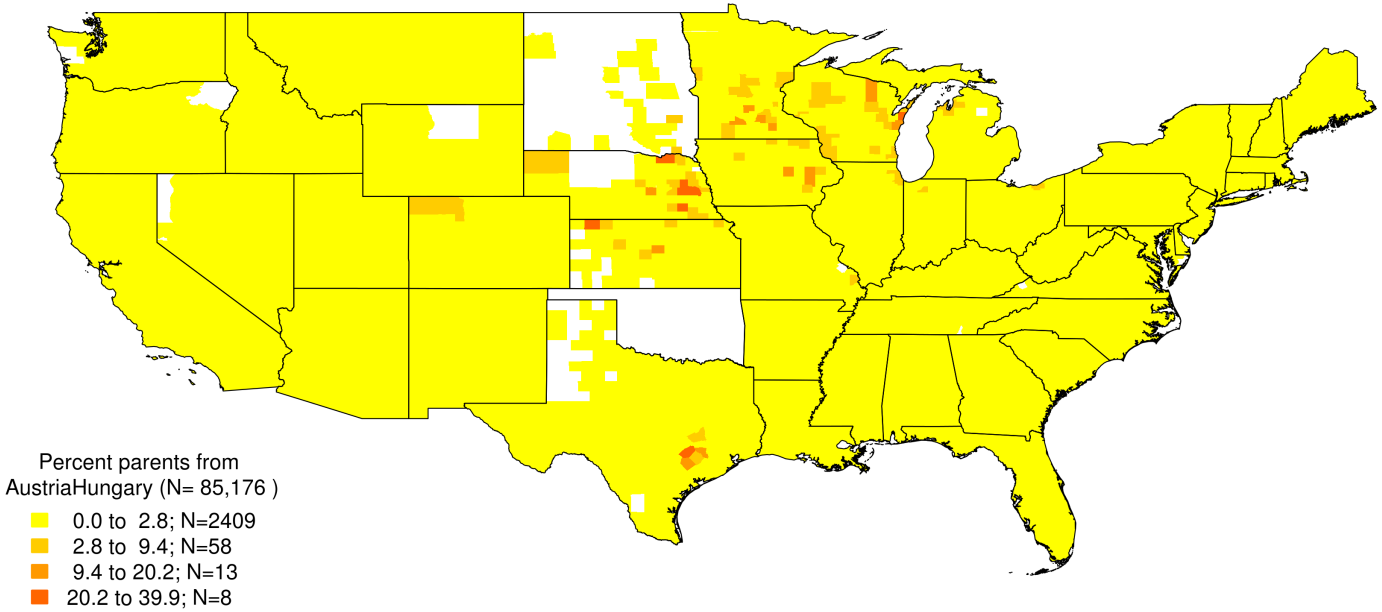

Figure 1.43: Austria-Hungary has a pattern something like the German: a strong presence in the western Great Lakes, and a strong presence in central Texas. Like Sweden, migrants also went to Nebraska, Kansas, and Iowa.

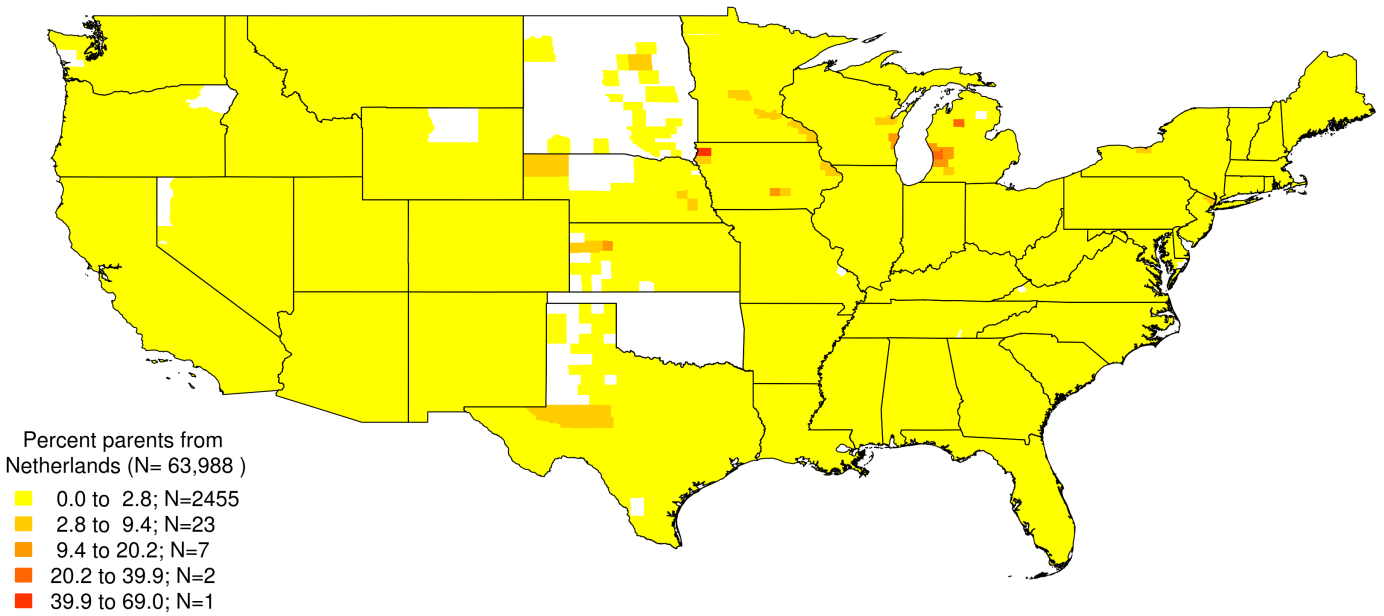

Figure 1.44: The Netherlands sent migrants in the 17th century, who occupied the Hudson Valley and parts of coastal New York and New Jersey. This map, however, shows those whose parents were born in the Netherlands before 1820. The biggest concentration is in western Michigan, a state largely settled by New Yorkers, many of whom would have had Dutch genetic or cultural heritage. Perhaps another example of homophily?

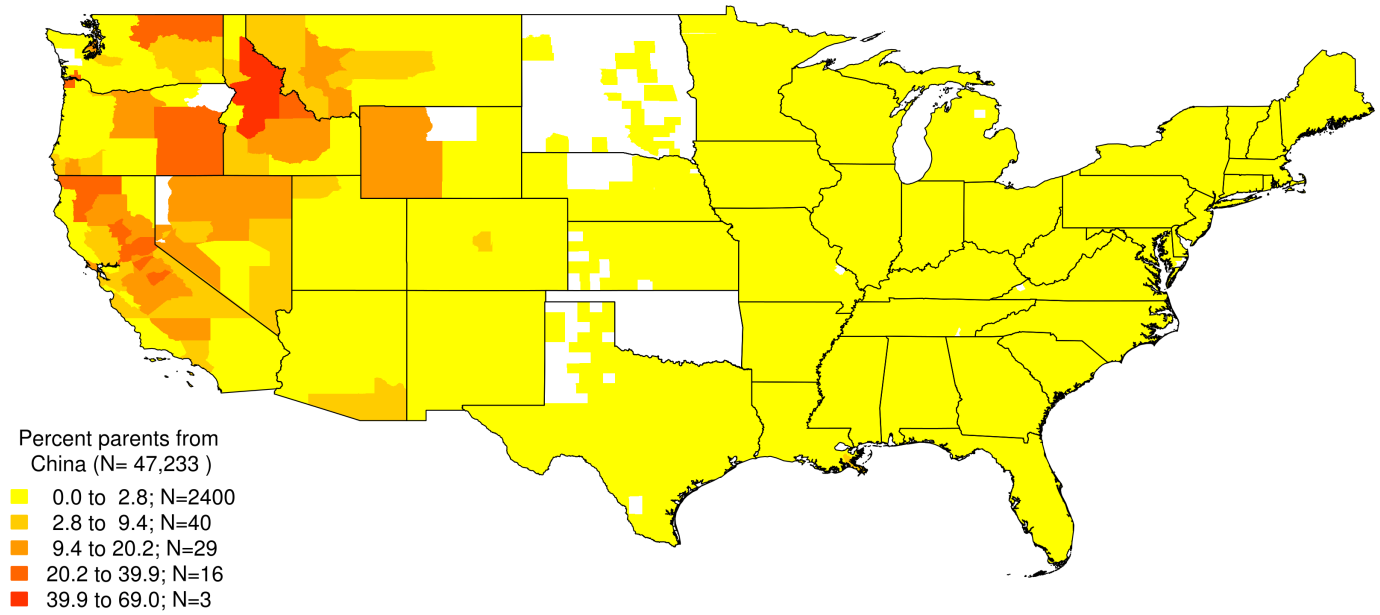

Figure 1.45: China's migrants made up a large share of the population in western counties, as far inland as Wyoming.

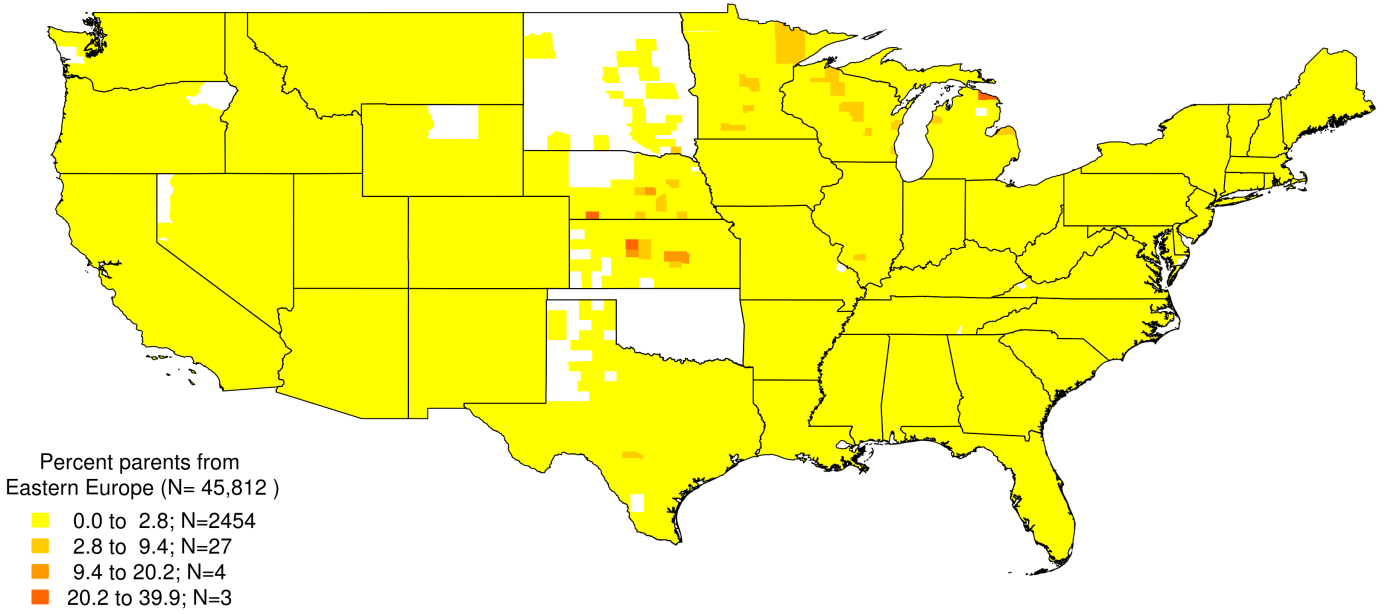

Figure 1.46: Eastern Europe shows a pattern of concentration not unlike that of Austria-Hungary: Minnesota and Wisconsin, Nebraska and Kansas.

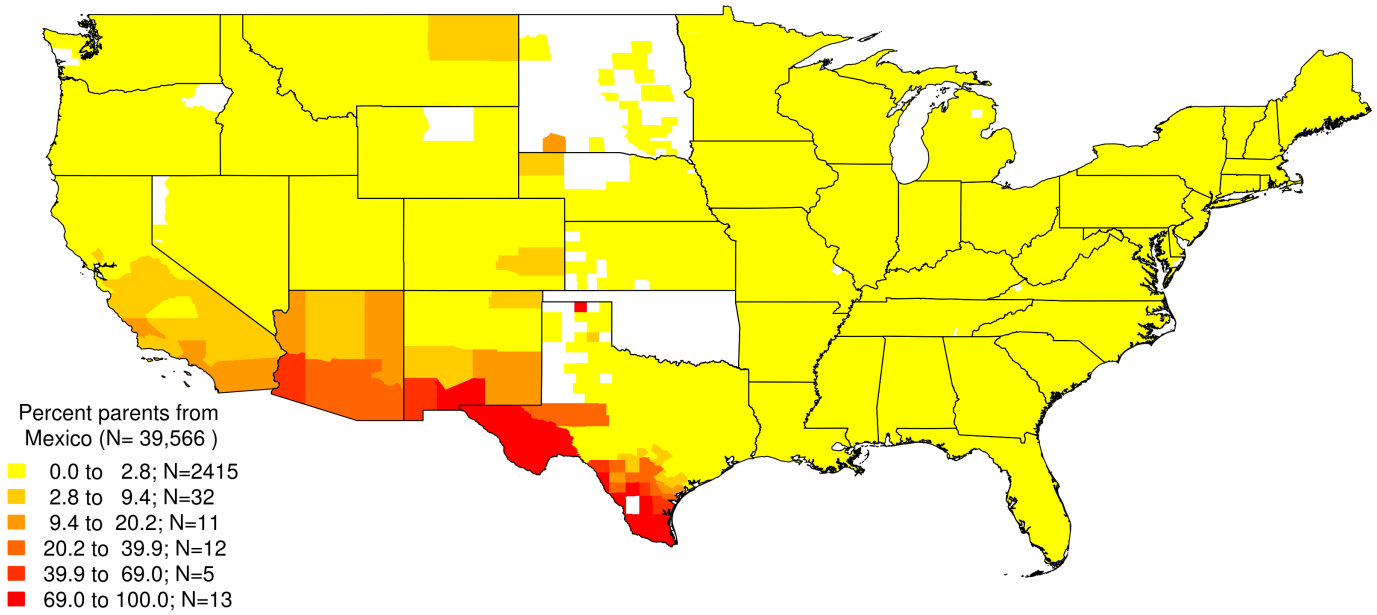

Figure 1.47: Mexico resembles Canada in that its migrants remain close to the border.

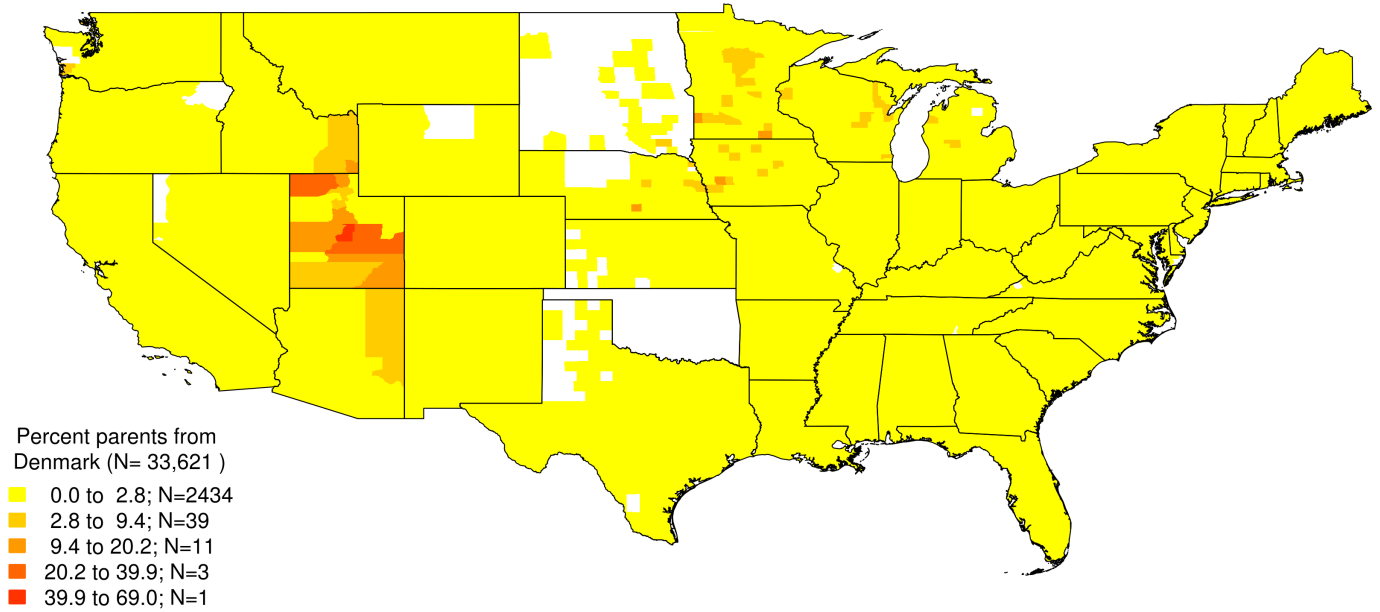

Figure 1.48: Denmark has a migration pattern similar to that of Sweden: Minnesota, Wisconsin, Iowa, Nebraska and Utah. It has the strongest presence in Utah of any Scandinavian country.

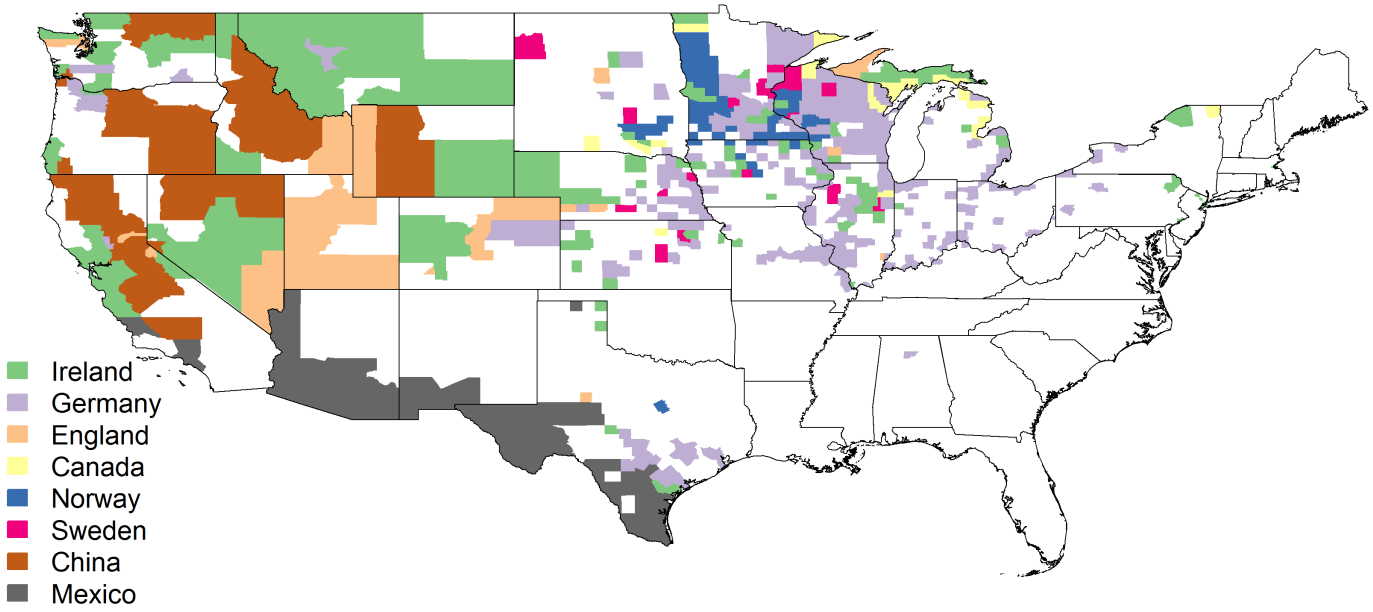

Figure 1.49: Counties where the top eight foreign locations provided the largest single share of population. Foreigners were especially prevalent in the West and in the western Great Lakes states. All groups avoided the South.

2 British origins of the US: evidence from surname frequencies

Crudely summarized, the overall story of the previous maps is that population moved west from the easternmost states, foreign immigration was largely confined to areas north of the Ohio River, and migration from Northern states seldom strayed south. Given that the easternmost source states played the key role in forming the populations and cultures of the western states, it is worthwhile to learn more about whence the population of the eastern states derived.

David Hackett Fischer [9] describes four principal migrations from Britain to colonial America, each with a different source area and each settling in different areas of the eastern United States. Fischer attributes differences in regional cultures in the US to the different regional British cultures brought by these four migration streams. In Fischer's scheme, New England was settled by Puritans from East Anglia, Virginia was settled by "Cavaliers and Servants" from southern England, Pennsylvania was settled by Quakers from the English North Midlands, and the western frontier was populated by Borderers from the border between Scotland and England, who arrived in large numbers in the mid 18th century. A poor people, with a tradition of skirmish warfare, they squatted on land on the western Pennsylvania frontier, then migrated along the Appalachian valleys toward the south. In Virginia, they mingled with poor farmers from Fischer's Cavalier and Servant culture, then migrated through the Cumberland Gap into Kentucky and Tennessee, whence they made the first inroads of European settlement both to the northwest and southwest [11].

Fischer argues that the first wave of migrants develop a culture to which subsequent migrants must adapt. A similar argument has been advanced by George Foster [10] for Latin America. Foster believes that the first Spanish immigrants to Latin America—predominantly from Seville—"crystallized" Latin American colonial culture, and that subsequent immigrants had little influence on Latin American regional cultures. The cultural geographer Wilbur Zelinsky [29] calls this the "doctrine of first effective settlement"—"the hypothesis that the first European or American white population that established the economic and social basis of an area had a decisive influence on later patterns"[11].

A small literature describes how practices associated with a particular ethnic group tend to be transmitted over the generations, even when the group's technology and environment change. A well-known example from ethnography is that of the American Plains Indians. With the introduction of the horse, a number of peoples moved onto the Plains and devel-

oped a culture based on nomadic buffalo hunting. While the Plains Indians all possessed a common core of cultural traits (such as the horse, the travois, and the teepee), they differed from each other in features such as political organization, and these differences can be attributed to cultural inertia—the persistence of traits from the time before they migrated onto the Plains [22, 19]. Likewise, Melanesians and Polynesians in the Western Pacific have very distinct political organizations, which are attributed to cultural inertia [22, 20, 23]. Cultural inertia is well documented even among European-Americans. For example, Sonya Salamon [24] describes how farming practices in the Midwest vary according to whether proprietors are of German, Irish, Yankee, or Swedish ancestry. Salamon's farmers are at least three or four generations removed from their European peasant ancestors, and yet the ethnic differences have persisted.

Surname analysis is one way to examine differential migration from British regions to US regions. Surnames are passed along male lines. Many surnames testify to their region of origin, either because they are in a particular language, or are toponyms, or are spelled in a distinct way. Even patronymics by their structure reveal regions of origins.[1] Figure 2.1 shows how the surname "Dow" is distributed in the British Isles and the eastern US. Like most surnames, it gives testimony to the specificity of origin and destination of migration flows.

The data used in this study are from the North American Population Project: the 1881 census of England, Wales, and Scotland [13]; and the 1880 census of the United States [7]. Both of these datasets are public-use microdata samples (PUMS), where each record is a surveyed person. Both are 100 percent samples, containing every surveyed person in that year. There are about 25 million persons in the British data and 50 million in the US data.

In order to visualize the pattern of surname distributions as far back in time as possible, all records for males aged 40 or older are extracted from the two data sets. In the British data, each male is assigned to his birth county, giving a picture of how surnames were distributed in the British Isles prior to 1841. In the US data, all males 40 or older are retained in the data set only if residing in a county located in the same state as their father had been born. This condition allows one to visualize the distribution of surnames as it was in 1820 or earlier. For most of the subsequent surname analyses, the set of US males is further restricted to include only white males.

In the original data, some records in the surname field are blank or contain non-alphabetic characters. After removing all non-alphabetic characters, all records with fewer than three characters are discarded. All surnames held by only one person are then discarded. These steps ensure that a missing value or spelling error is not counted as a distinct surname, and it makes it less likely that a stray foreigner's surname would be counted as British. Over 190,000 distinct surnames were left

The data for US surnames is arranged in a frequency table U, such that each row represents a surname and each column a US county; each cell in U gives the number of occurances of the row surname in the county column. A few sample rows and columns are shown in Table 2.1.

A similar matrix B is constructed for the British data, with each cell in B giving the number of occurances of the row sur-

[1] For example, a suffix -s is Welsh; a prefix Mac- is Irish or Scottish.

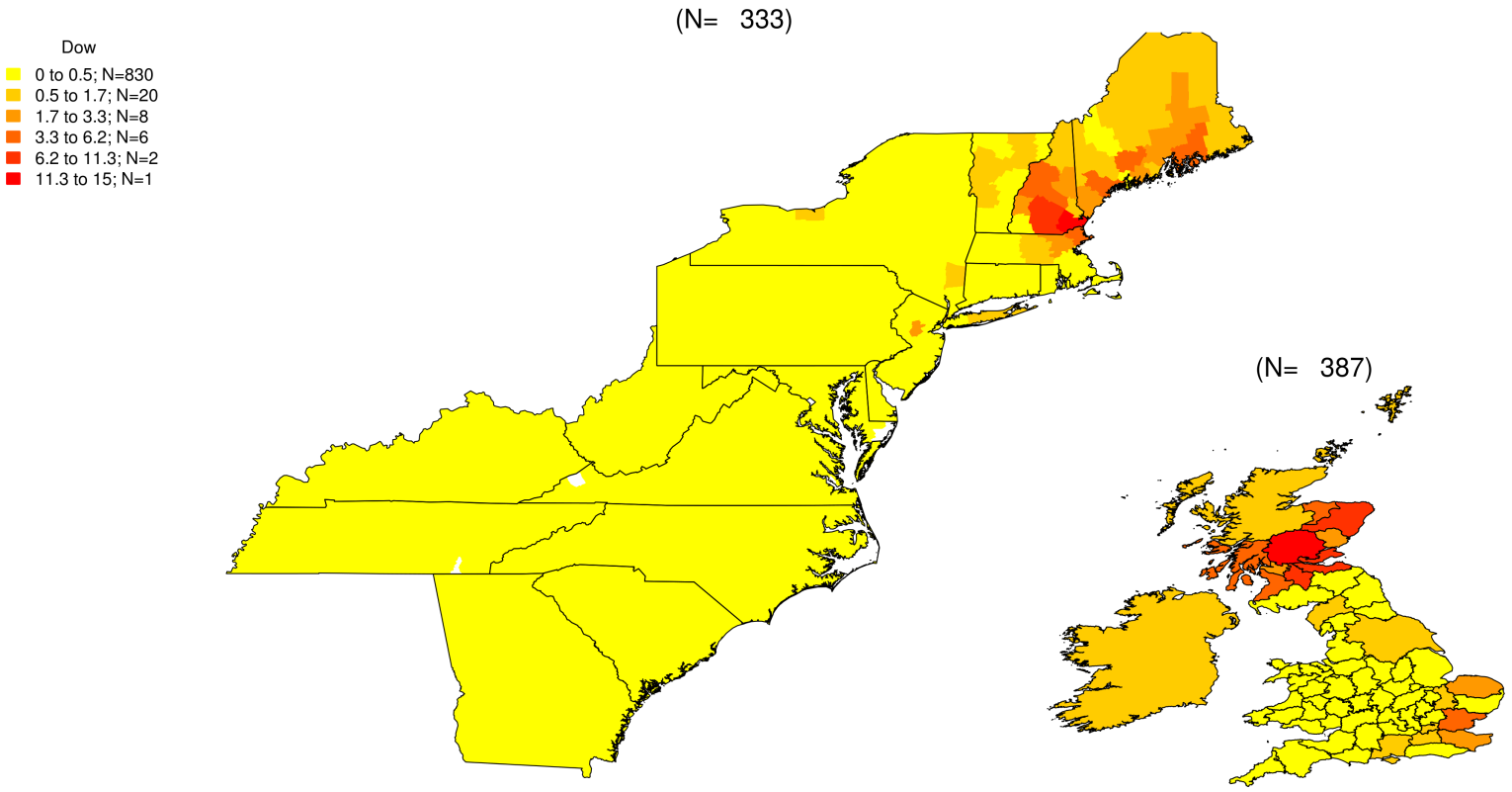

Figure 2.1: Example of a surname distribution in Britain 1881 and the eastern US 1880: the surname "Dow".

Table 2.1: Sample rows and columns from matrix U

	360610	360750	245100	360010
smith	266	54	117	81
marshall	16	2	10	4
locke	1	0	0	1
burke	5	1	7	1
sumner	1	1	1	1
petty	1	1	0	0
godwin	2	0	2	0
hume	4	1	0	0

name in the British county column. The British data cover England, Scotland, and Wales, with Ireland only present because a number of respondents listed Ireland as their place of birth. The British county boundaries are no longer used, and it proved convenient to aggregate the Scottish historic counties in order to use a GIS file of modern boundaries [1]. This gave a total of 52 counties in England and Wales, 13 counties in Scotland, and Ireland as an additional region.

These 66 counties seemed to provide too much regional detail in Britain, and additional aggregation seemed desirable. Adjacent counties tend to have similar surnames, which suggested that aggregation might be feasible. The surname data are analogous to data used by ecologists, in that each county represents a sampling location and each surname a species. A variety of methods exist in order to aggregate sampling locations into habitats, based on species abundance. Each of these methods produces a somewhat different result. I chose the quantitative Kulczynski distance measure, which is appropriate for count data and is considered to be effective at detecting gradients in ecological data.[8, 18] For the clustering algorithm I chose Ward's minimum variance method, using the square of the Kulczynski distance matrix[26, 18]. The resulting dendrogram is shown in Figure 2.2. The rectangular boxes identify the eight clusters selected for aggregation; these are also shown on the map. Figure 2.3 shows that cutting the dendogram into three clusters separates the map into Wales, England, and Scotland (with Ireland). As the number of clusters increases, the constituent counties remain contiguous. Overall, the aggregation appears very reasonable and well-behaved. Below is a list of the historic counties included in each of the eight categories:

1. North Wales: Anglesey; Caernarvonshire; Denbighshire; Flintshire; Merionethshire; Montgomeryshire; Shropshire

2. East England: Bedfordshire; Cambridgeshire; Essex; Hertfordshire; Huntingdonshire; Norfolk; Suffolk

3. South England: Berkshire; Buckinghamshire; Cornwall; Devonshire; Dorset; Gloucestershire; Hampshire; Kent; Middlesex; Oxford; Somerset; Surrey; Sussex; Wiltshire

4. South Wales: Brecknockshire; Cardiganshire; Carmarthenshire; Glamorganshire; Herefordshire; Monmouthshire; Pembrokeshire; Radnorshire

5. Mid England: Cheshire; Derbyshire; Lancashire; Leicestershire; Lincolnshire; Northamptonshire; Nottinghamshire; Rutland; Staffordshire; Warwick; Worcestershire; Yorkshire

6. North England: Cumberland; Durham; Northumberland; Westmorland

7. Scots-Irish: Aberdeenshire; Angus; Ayr; Banff; Berwickshire; Clackmannanshire; Dumfriesshire; Dunbartonshire; East Lothian; Fife; Kincardinshire; Kirkudbrightshire; Kinrossshire; Lanarkshire; Mid Lothian; Morayshire; Peeblesshire; Perthshire; Renfrewshire; Roxburghshire; Selkirk; Stirlingshire; Wigtown; West Lothian; Ireland

8. unknownBI: Alderney; Guernsey; Isle of Man; Jersey; England; Wales; Scotland; At Sea; Foreign; Unknown

9. Highlands: Argyll; Bute; Caithness; Invernessshire; Nairnshire; Orkney; Rossand Cromarty; Shetland; Sutherland; Adjacent Islands Herm; Adjacent Islands Little Sark

A similar hierarchical clustering procedure was performed on the matrix of US data U, using Ward's minimum variance on the square of the Kulczynski distance matrix. Five clusters seem indicated by the dendrogram, and these map out to five fairly contiguous regions.

Many US surnames are not in the set of British surnames. To find the national origin of these surnames I select from the full 1880 US Census dataset all white males 40 and over whose father was born outside the US.[2] Surnames with only one instance are dropped. Locations outside the US are aggregated into a small number of categories. Below is a list of geographical areas included in each of the categories.

1. unknownUS: Africa; Eastern Africa; Ethiopia; Madagascar; Mauritius; Mozambique; Congo; Northern Africa; Algeria; Egypt; Libya; Morocco; Sudan; Tunisia; Southern Africa; Botswana; Zimbabwe; Southern Africa, NS; Burkina Faso; Guinea; Liberia; Mali; Nigeria; Sierra Leone; French West Africa, NS; West Africa, NS; Africa, NS; Caribbean; West Indies; Other West Indies; Curacao; St Eustatius; Dutch Caribbean, NS; French St Maarten; Guadeloupe; Martinique; St Barthelemy; French Caribbean, NS; Caribbean, NS; West Indies, NS; Antigua Barbuda; Bahamas; Barbados; Cayman Isles; Dominica; Grenada; Haiti; Jamaica; St Kitts Nevis; St Lucia; St Vincent; Trinidad and Tobago; Turks and Caicos; British Honduras; French Guiana; British Guiana; Suriname; North America; US Outlying Areas and Territories; Eastern Asia; China; Hong Kong; Macau; Taiwan; Japan; Korea; South Korea; Mongolia; South Asia; Afghanistan; Bhutan; India; Iran; Pakistan; Ceylon; Indonesia; Malaysia; Burma; Philippines; Singapore; Thailand; Vietnam; South Eastern Asia, NS; Western Asia; Armenia; Cyprus; Iraq; Israel; West Bank; Jordan; Lebanon; Palestine; Yemen; Syria; Turkey; Asia, NS; Northern Europe; Europe, NS; Oceania; Fiji; New Caledonia; Papua New Guinea; Solomon Islands; Cook

[2] Males and father birthplaces are selected since surnames are transmitted from fathers to sons. The selection is limited to whites since blacks (the only sizable non-white racial group) mostly acquired surnames on emancipation, and many of these are simply the borrowed names of former slave owners. Only those 40 and older are selected in order to map the relationship between surname and national origin as it would have been in the early years of the 19th century.

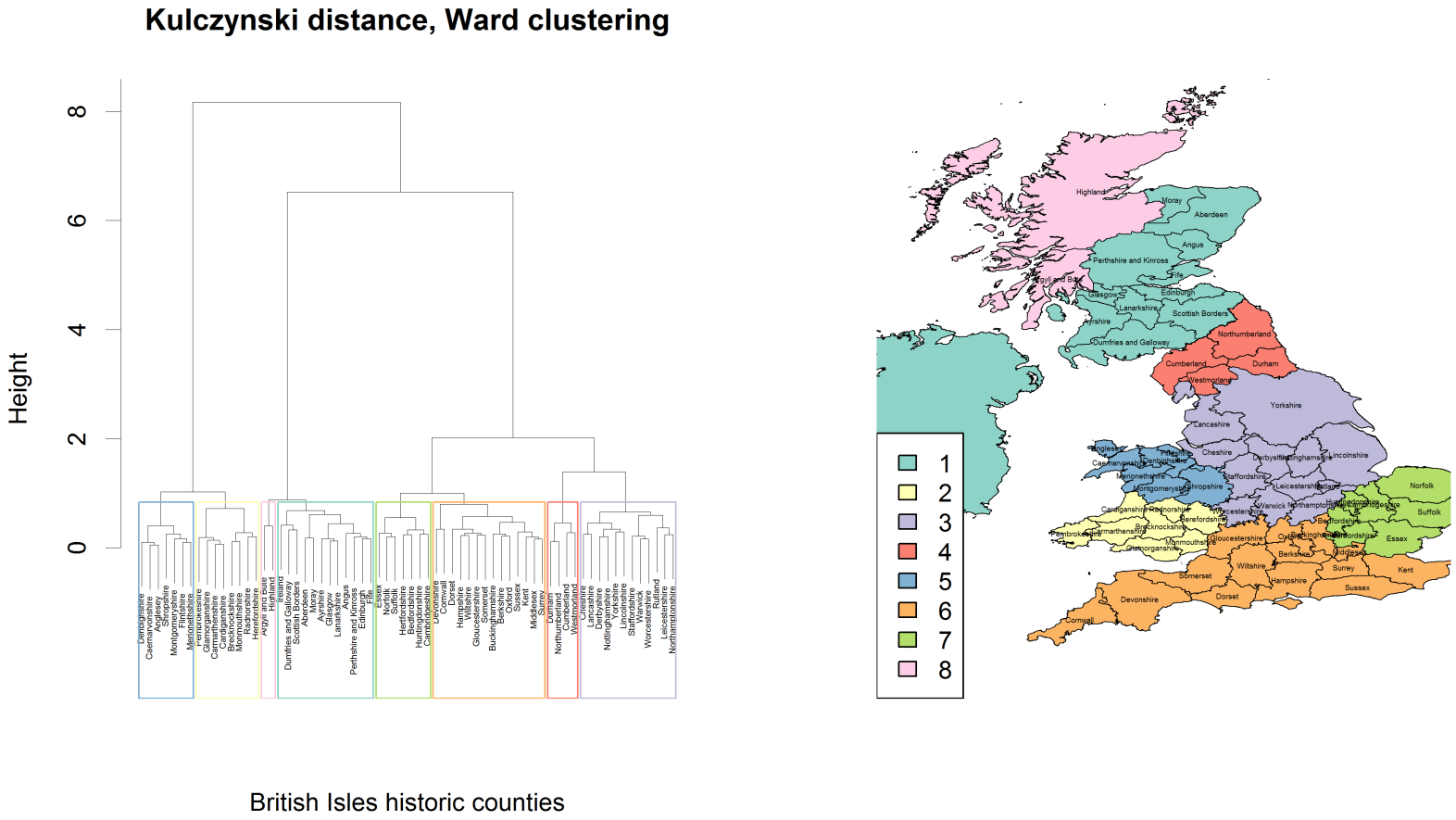

Figure 2.2: Hierarchical cluster analysis on surname frequency in British counties. The dendrogram is based upon a quantitative Kulczynski distance matrix and the Ward minimum variance clustering algorithm.

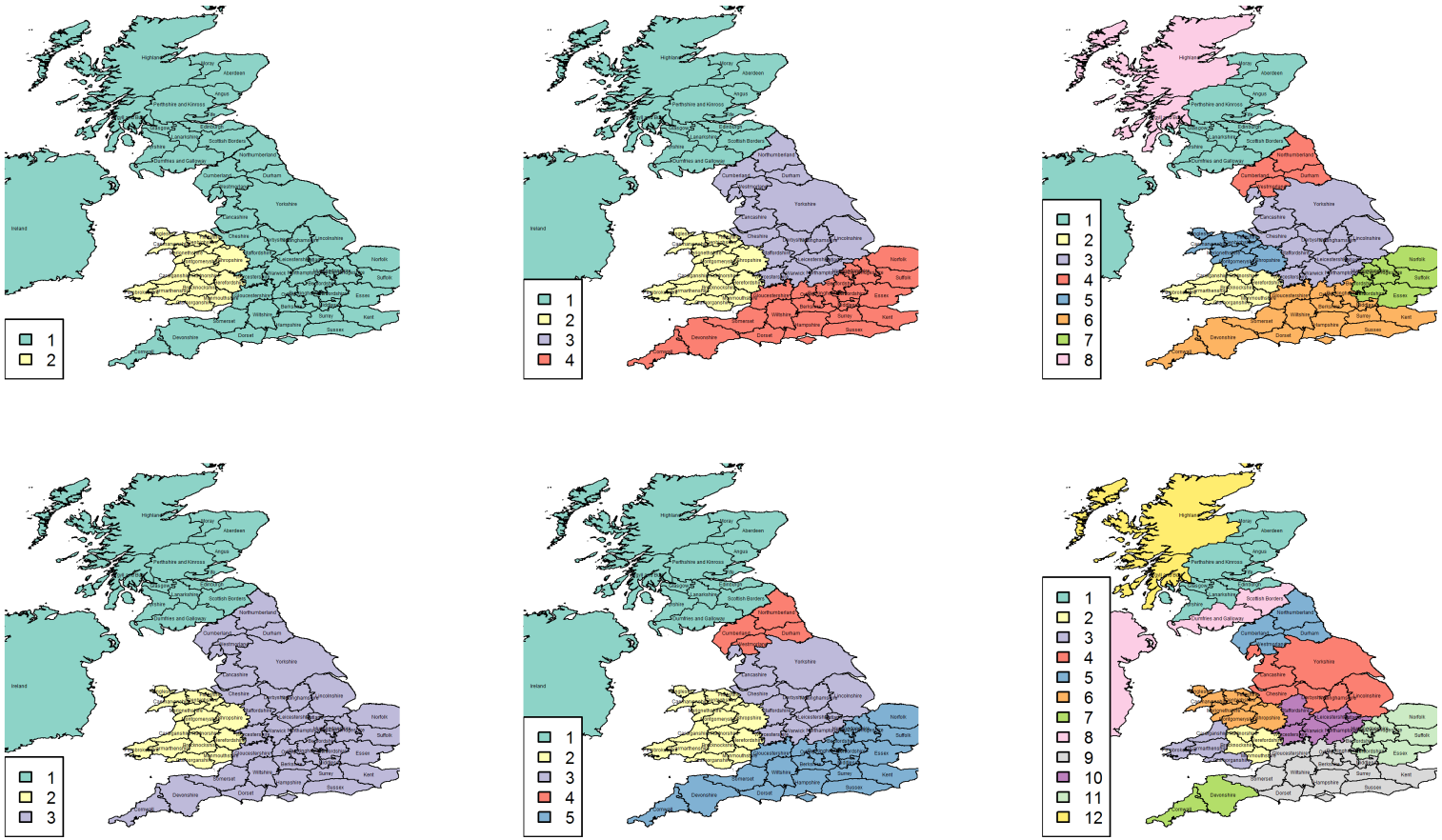

Figure 2.3: Cutting the dendrogram found in Figure 2.2 groups British counties into regions of similar surnames. Note that counties within clusters are always contiguous.

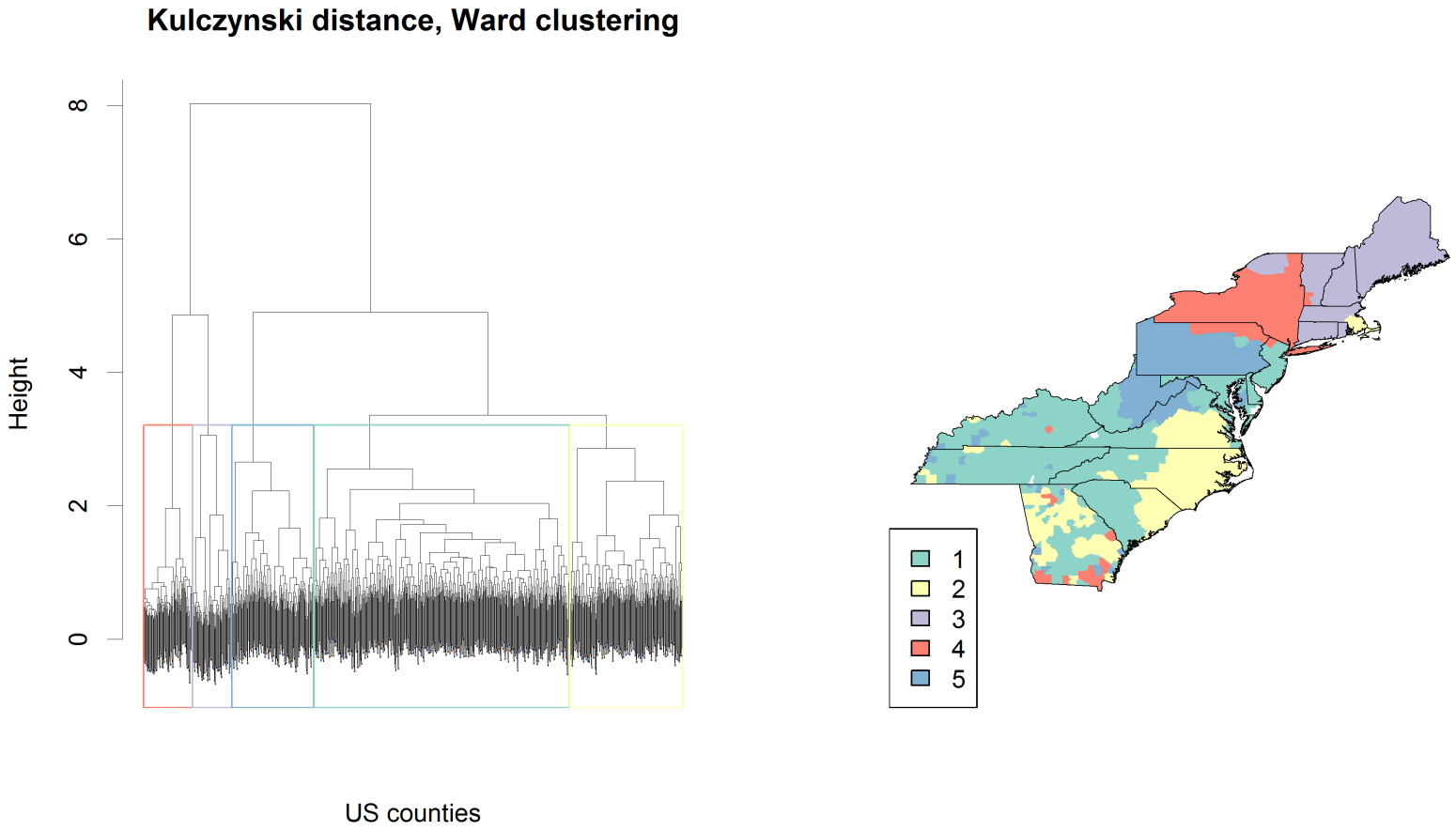

Figure 2.4: Hierarchical cluster analysis on surname frequency in US counties. The dendrogram is based upon a quantitative Kulczynski distance matrix and the Ward minimum variance clustering algorithm.

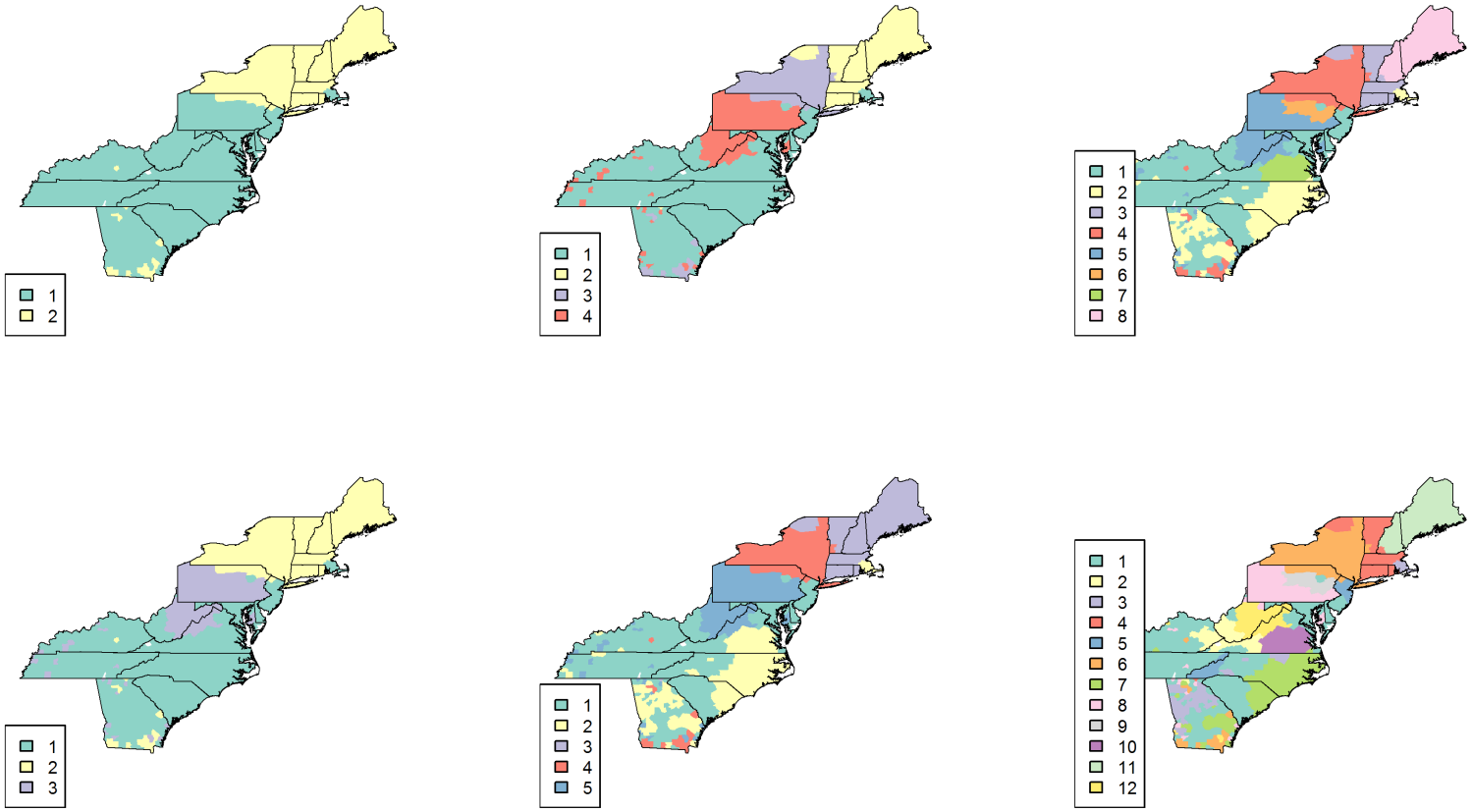

Figure 2.5: Cutting the dendrogram found in Figure 2.4 groups US counties into regions of similar surnames. Contiguity is not as pronounced as it is with the British counties in Figure 2.3, suggesting a settlement pattern of leapfrogging and infill.

Islands; French Polynesia; Oceania, NS; AtlanticIslands; Atsea; Unknown

2. otherBritish: Union of South Africa; St Helena and Ascension; Falkland Islands; Bermuda; United Kingdom; England; Channel Islands; Guernsey; Isle of Jersey; Isle of Man; Scotland; Wales; Australia and New Zealand, NS; Australia; New Zealand

3. IberoItalian: Cape Verde; Cuba; Dominican Republic; Central America; Costa Rica; El Salvador; Guatemala; Honduras; Mexico; Nicaragua; Panama; South America; Argentina; Bolivia; Brazil; Chile; Colombia; Ecuador; Paraguay; Peru; Uruguay; Venezuela; South America, NS; Gibraltar; Italy; Malta; Portugal; Azores; Madeira Islands; Spain; Southern Europe, NS

4. Canada: Canada; British Columbia; Manitoba; New Brunswick, Nova Scotia, and PEI; Ontario and Upper Canada; Saskatchewan; Quebec; Northwest; Newfoundland and Labrador

5. Scandinavia: Greenland; Denmark; Faroe Islands; Finland; Iceland; Norway; Svalbard and Jan Mayen Islands; Sweden

6. EasternEurope: Eastern Europe; Bulgaria; Czech Republic; Hungary; Poland; Romania; Russia; Slovakia; Ukraine; Estonia; Latvia; Lithuania; Albania; Bosnia; Croatia; Greece; Slovenia; Yugoslavia; Montenegro; Serbia; Austria

7. IrelandF: Ireland

8. France: Andorra; France; Alsace Lorraine; Lorraine; Monaco

9. Benelux: Belgium; Luxembourg; Netherlands

10. Germany: Alsace; Germany; Mecklenburg Schwerin; Hamburg; Bremen; Mecklenburg; Other Germany

11. Switzerland: Liechtenstein; Switzerland

From the surnames of these males and the consolidated birthplaces of their fathers, a matrix frequency table F is created, where each row i is a specific surname and each column j is a paternal birthplace. The matrices B and F are then merged in matrix G. The goal is to row-normalize G, so that each cell $\tilde{g}_{ij}$ in $\tilde{G}$ represents the percent of surname i with origin in region j. Merging is not straightforward, however, since the universes for B and F are different: the former is Britain and the latter is the US; differences in the values $\tilde{g}_{ij}$ will in large part be due to the differences in the sizes of the two universes. It is also true that immigration occuring in the 17^{th} or early 18^{th} century would not be represented in matrix F. As a result, there would be relatively few instances of Dutch and British surnames in F.

The following sequential manipulations of the matrix G help address these problems and to further reduce the number of categories.

1. For surnames where one or more of the British regions or foreign locations is present, the unknown categories (unknownBI and unknownUS) are set to zero.

2. If both "unknownBI" and "unknownUS" are non-zero, then "unknownUS" is set to zero. There will now be only

one unknown value for each surname, and it will exist only if there are zeros for all British regions and foreign locations.

3. For all surnames for which at least one other foreign location is non-zero, the value of "Canada" is divided up and added to these other foreign locations, each receiving an amount in proportion to the number of persons already present in that location. Values for "Canada" are then set to zero for these surnames.

4. "Canada", "IrelandF", and "unknownBI" are then added to "otherBritish", then set to zero. The three fields are removed, and "unknownUS" is relabeled "unknown".

5. Next, "otherBritish" is allocated over the eight British regions (if at least one of these is non-zero), in proportion to the number of persons already counted in each of these regions (an addition to the existing values, not a replacement). Values for "otherBritish" are then set to zero for these surnames. This step emphasizes the British regions in these specific cases where a surname is found both in Britian and in foreign locations, helping to correct the undercount for British surnames.

6. For the remaining values of "otherBritish", if any foreign locations are non-zero, the values of "otherBritish" are allocated to them, in proportion to the number of persons already counted in each of these locations (an addition to the existing values, not a replacement). Values for "otherBritish" are then set to zero for these surnames.

7. Surnames with non-zero values of "unknown" are assigned to "Benelux" if they begin with "van". This helps mitigate the under-representation of Dutch names.

8. Surnames with non-zero values of "unknown" are assigned to the British regions if they begin with "Mac". This is done by finding the proportion of surnames with the prefix "Mac" distributed across all British regions, and then allocating those in "unknown" to those regions in the same proportion. Values for "unknown" are then set to zero. The process is then repeated for the surname prefix "Mc".

9. Surnames with non-zero values of "otherBritish" are assigned to the British regions if they begin with "Mac". This is done by finding the proportion of surnames with the prefix "Mac" distributed across all British regions, and then allocating those in "otherBritish" to those regions in the same proportion. Values for "otherBritish" are then set to zero. The process is then repeated for the surname prefix "Mc".

10. Two different processes account for a surname's presence in both British regions and foreign locations. First, immigrants to the US may change their surnames to sound more British, as for example Mueller becomes Miller, so that Miller becomes a common surname among people of German descent. Second, a surname present in Britain may exist there only because of immigration from other European countries, especially France, the Netherlands, and Germany. In order to distinguish purely British surnames, I look at the relative frequencies of surnames in

> Britain and foreign regions. For all surnames where the sum of occurances among foreign locations is less than 80 percent of the sum of occurances in Britain, I assume a British surname and set the values of foreign locations to zero. For all surnames where the sum of foreign location occurances is more than 125 percent of the sum of British occurances, I assume a foreign surname and set the values of British occurances to zero. For all those ratios lying between 80 and 125 percent, I keep foreign and British occurances, considering that the surname may be one often adopted by immigrants. These ranges are admittedly arbitrary, but examination of the surnames selected in each of the three groups convinces me that these ranges are reasonable.

Table 2.2 presents the summary statistics for matrix G, giving the number of persons, the number of surnames, and the mean number of persons per surname for each of the British regions and foreign locations. One can see that the foreign locations are drawn from a different universe, with many fewer persons per surname. Overall, Germany has the largest number of surnames, followed by South England, which contains London and the coast facing the continent. Also given in the table is the number of persons holding the most common surname by place, and the actual most common surname. Note that "Miller" is the most common surname for four foreign locations, a clear indication of the degree to which immigrants may switch to British surnames. Note also that the name "Smith" is the most common surname in three English regions, illustrating the degree to which surnames overstep regional boundaries.

Table 2.2: Summary statistics for $\tilde{G}$

					Persons/Surnames			
	persons	surnames	unique	mean	max	max.name	max.unique	uniq.name
Scandinavia	44,014	7,110	2,945	6.2	3,113	Peterson	194	Halvorson
Germany	389,296	62,950	39,089	6.2	6,818	Miller	147	Behrens
France	27,496	11,047	3,135	2.5	279	Miller	70	Tatro
IberoItalian	15,195	4,320	2,246	3.5	464	Garcia	464	Garcia
EasternEurope	19,349	9,240	1,972	2.1	147	Cohen	22	Prochaska
Benelux	16,625	6,990	2,291	2.4	87	Miller	79	VanCleave
Switzerland	16,502	6,459	1,193	2.6	312	Miller	14	Luchsinger
unknown	126,514	25,289	25,289	5.0	192	Seavey	192	Seavey
ScotsIrish	737,991	25,701	8,107	28.7	7,412	Murphy	317	McMahan
NorthWales	122,723	8,495	376	14.4	17,446	Jones	14	Elkes
Highlands	74,314	5,179	313	14.3	3,939	McDonald	53	McAskill
EastEngland	339,965	23,347	3,384	14.6	6,956	Smith	61	Waldock
SouthEngland	1,028,635	47,261	14,017	21.8	14,067	Smith	276	Dunlap
SouthWales	125,305	8,383	389	14.9	10,920	Jones	24	Manny
MidEngland	1,130,315	39,453	9,644	28.6	20,152	Smith	191	Rostron
NorthEngland	157,159	13,332	1,168	11.8	2,376	Thompson	31	Mordue
otherBritish	49,858	22,402	16,249	2.2	121	Calhoun	73	Nicolle

The matrix G is then row-normalized by dividing each cell by the row sum, creating the matrix $\tilde{G}$. Each cell in $\tilde{G}$ can now be interpreted as the percent of instances of the row surname that occur in the column location. A few sample rows and columns are shown in Table 2.3.

Table 2.3: Sample rows and columns from matrix $\tilde{G}$

	ScotsIrish	MidEngland	SouthEngland
Smith	0.1269	0.3909	0.2726
Marshall	0.1851	0.4406	0.2256
Locke	0.0755	0.1396	0.6377
Burke	0.7963	0.0838	0.0931
Sumner	0.0019	0.7765	0.1713
Petty	0.0172	0.6046	0.3095
Godwin	0.0068	0.2146	0.7169
Hume	0.4846	0.0993	0.0875

The predicted number of persons in American county k descending from settlers from a location j is $x_{kj} = \sum_i u_{ik}\tilde{g}_{ij}$ or, in matrix format, $X = U'\tilde{G}$. The same method can be applied to the British surname data: $Y = B'\tilde{G}$. Row-normalizing Y gives $\tilde{Y}$ where each cell $\tilde{y}_{kj}$=$y_{kj}/\sum_j y_{kj}$ gives the probability that a surname selected in British county k comes from location j. The diagonal of $\tilde{Y}$ is interesting in that it is interpreted as the probability that a surname selected in British county k comes from British county k. This is less than one, since surnames are often shared across counties; it would equal one were each county endowed with a unique set of surnames, and the less frequent its surnames in outside counties, the higher it would be. The diagonal therefore provides some measure of the distinctiveness of the locations; Welsh counties are particularly low, reflecting the fact that a few surnames such as Jones are held by a large number of persons, both in Wales and in England.

Table 2.4 shows the overall composition of the US population within the geographic boundaries of the maps.

Table 2.4: Ethnic composition of U

	count	pct
Scandinavia	5,442	0.6
Germany	111,916	12.2
France	9,919	1.1
IberoItalian	1,344	0.1
EasternEurope	3,458	0.4
Benelux	8,645	0.9
Switzerland	6,102	0.7
unknown	27,638	3
ScotsIrish	95,989	10.4
NorthWales	19,611	2.1
Highlands	8,250	0.9
EastEngland	87,380	9.5
SouthEngland	254,772	27.7
SouthWales	21,341	2.3
MidEngland	203,853	22.1
NorthEngland	29,035	3.2
otherBritish	25,955	2.8

The ethnic composition of each of the five regions shown in

Figure 2.4 is given in Table 2.5.

Pennsylvania is the outlier, with its large non-British population. The Southern Coast and New England are quite similar in their composition, though there is some truth in the oft-heard assertion [16] that the South is more "Celtic" than the North: both Welsh and Scottish surnames are more common in the South. Nevertheless, the most salient difference between the North and the South is that the latter is much more British than the North.

The following maps show the degree to which US counties contain more than their share of each surname category in matrix X. The expected number of persons in US county k descending from settlers from location j is:

$$\hat{x}_{kj} = \left(\sum_k x_{kj} / \sum_k \sum_j x_{kj} \right) \sum_j x_{kj}$$

which is used to create a matrix $\dot{X}$ giving the percentage by which surname category j exceeds its expected level in county k:

$$\dot{x}_{kj} = 100\, max\left(0, \left(x_{kj} - \hat{x}_{kj}\right)\right) / \hat{x}_{kj}$$

The plotted values in the maps are drawn from $\dot{X}$. The histograms report values from $\tilde{X}$, which gives the percent of the population in each row county in each column surname category. Twelve randomly selected names are shown for each of the surname categories. For surnames stemming from the eight British regions, a map of the British Isles is shown in the upper right, with the relevant region given in black, and red scale indicating the degree to which other regions share surnames. For non-British surname categories, a scatterplot is given, with logged county population size on the abscissa and the percentage share of county population on the ordinate. Blue points represent counties south of the Mason-Dixon line (slave states) and red points counties north of the line. The green dotted line is the lowess smoother. The plot gives some indication whether the surname category is more represented in the North or South, and in urban or rural areas.

Table 2.5: Ethnic composition of the five US regions in Figure 2.4

	Pct. in surname category						Actual minus expected				
	Penn	Frntr	SCst	NwEng	NwYrk	expected	Penn	Frntr	SCst	NwEng	NwYrk
Scandinavia	0.6	0.6	0.4	0.6	0.8	0.6	0.0	0.0	-0.2	0.0	0.2
Germany	30.4	7.9	2.6	3.2	11.8	12.2	18.3	-4.2	-9.6	-9.0	-0.3
France	1.9	0.8	0.4	0.6	1.4	1.1	0.8	-0.3	-0.6	-0.5	0.3
IberoItalian	0.1	0.1	0.1	0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0
EasternEurope	0.8	0.3	0.1	0.1	0.4	0.4	0.5	-0.1	-0.2	-0.3	0.0
Benelux	0.8	1.0	0.2	0.2	2.4	0.9	-0.2	0.1	-0.8	-0.8	1.5
Switzerland	2.0	0.4	0.1	0.1	0.4	0.7	1.3	-0.3	-0.5	-0.6	-0.3
unknown	4.6	3.3	3.0	1.3	2.9	3.0	1.6	0.3	0.0	-1.7	-0.1
ScotsIrish	11.9	12.8	10.2	8.2	8.8	10.4	1.4	2.4	-0.2	-2.2	-1.6
NorthWales	1.6	2.6	3.0	2.1	1.9	2.1	-0.6	0.5	0.9	-0.1	-0.3
Highlands	1.0	1.0	1.1	0.7	0.9	0.9	0.1	0.1	0.2	-0.2	0.0
EastEngland	5.6	9.1	11.4	12.2	10.2	9.5	-3.9	-0.4	1.9	2.7	0.7
SouthEngland	17.1	27.5	33.0	35.6	27.7	27.7	-10.6	-0.2	5.3	7.9	0.1
SouthWales	1.8	2.8	3.2	2.1	2.1	2.3	-0.5	0.5	0.9	-0.2	-0.2
MidEngland	15.0	23.5	25.8	26.7	21.4	22.1	-7.2	1.4	3.7	4.5	-0.8
NorthEngland	2.8	3.7	3.4	3.1	2.7	3.2	-0.3	0.6	0.3	0.0	-0.4
otherBritish	2.1	2.5	1.9	3.1	4.1	2.8	-0.7	-0.3	-0.9	0.3	1.3

Notes: British regions shown in Figure 2.2

2.1 British regions

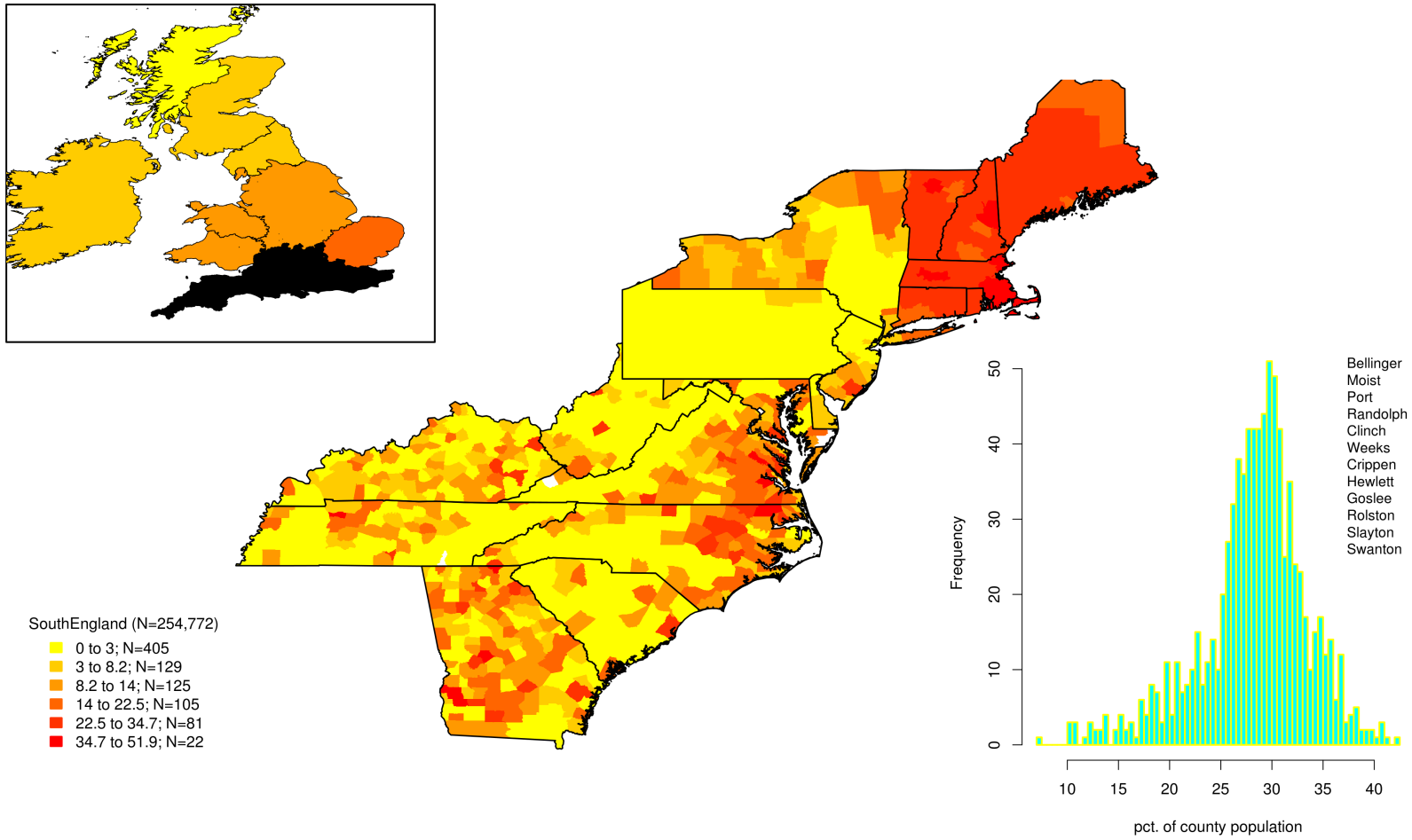

Figure 2.6: South England is the single biggest surname category. New England, followed by the Southern Coast, (areas defined in Figure 2.4) have the highest incidence of surnames from South England. Shown in black on the small map of the British Isles, the region includes London, Bristol, and other ports such as Plymouth, and seems the most likely recruiting ground for British colonists. It shares surnames most frequently with East England, but due to the cosmopolitan nature of its cities, it shares surnames with all British and foreign locations.

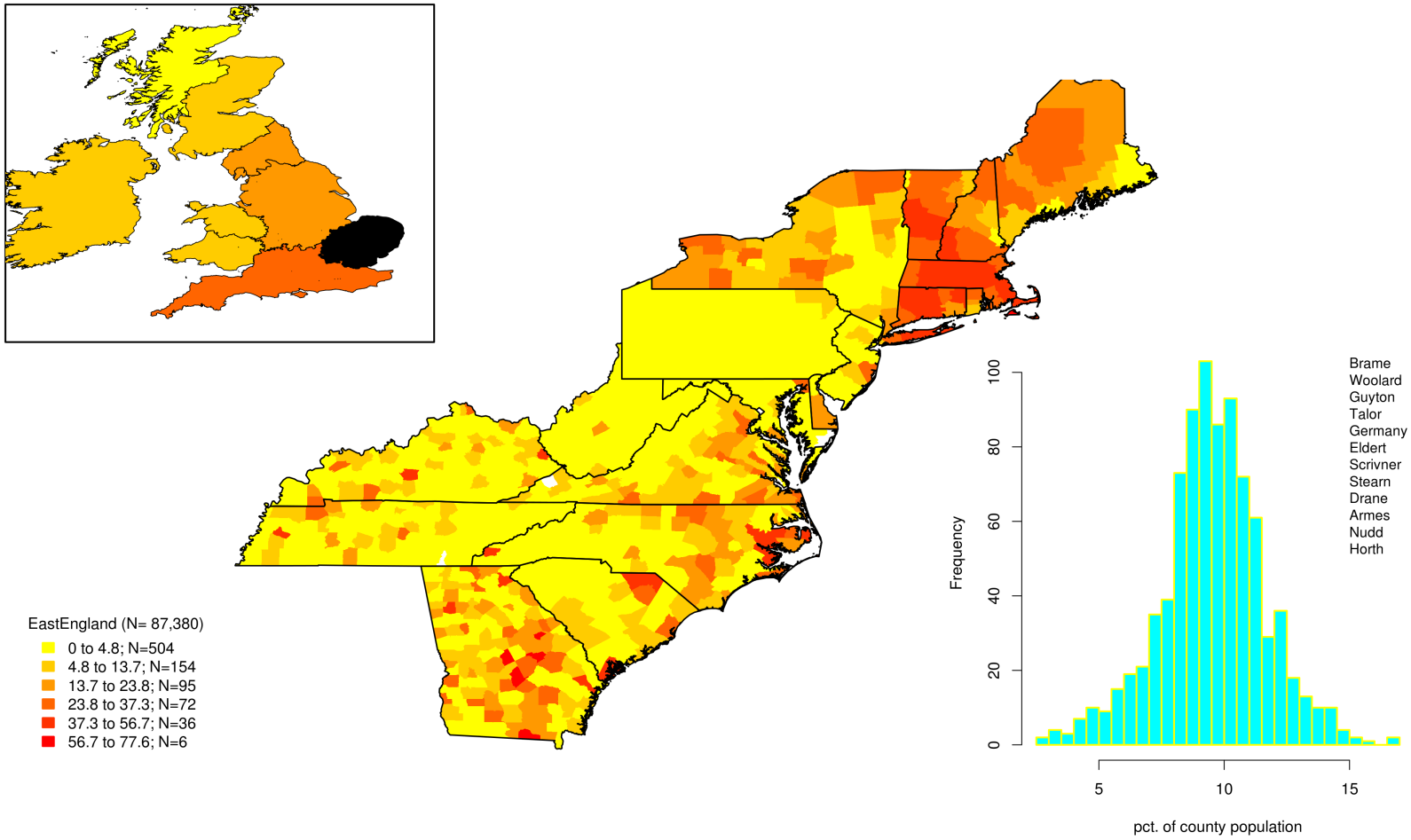

Figure 2.7: East England, or more specifically East Anglia, is considered to be the area of origin of the Puritans. And indeed East England surnames are more heavily represented in New England than they are elsewhere, but they are not as common in New England as surnames from South or Mid England. And in fact East England surnames are also common in New York and the South.

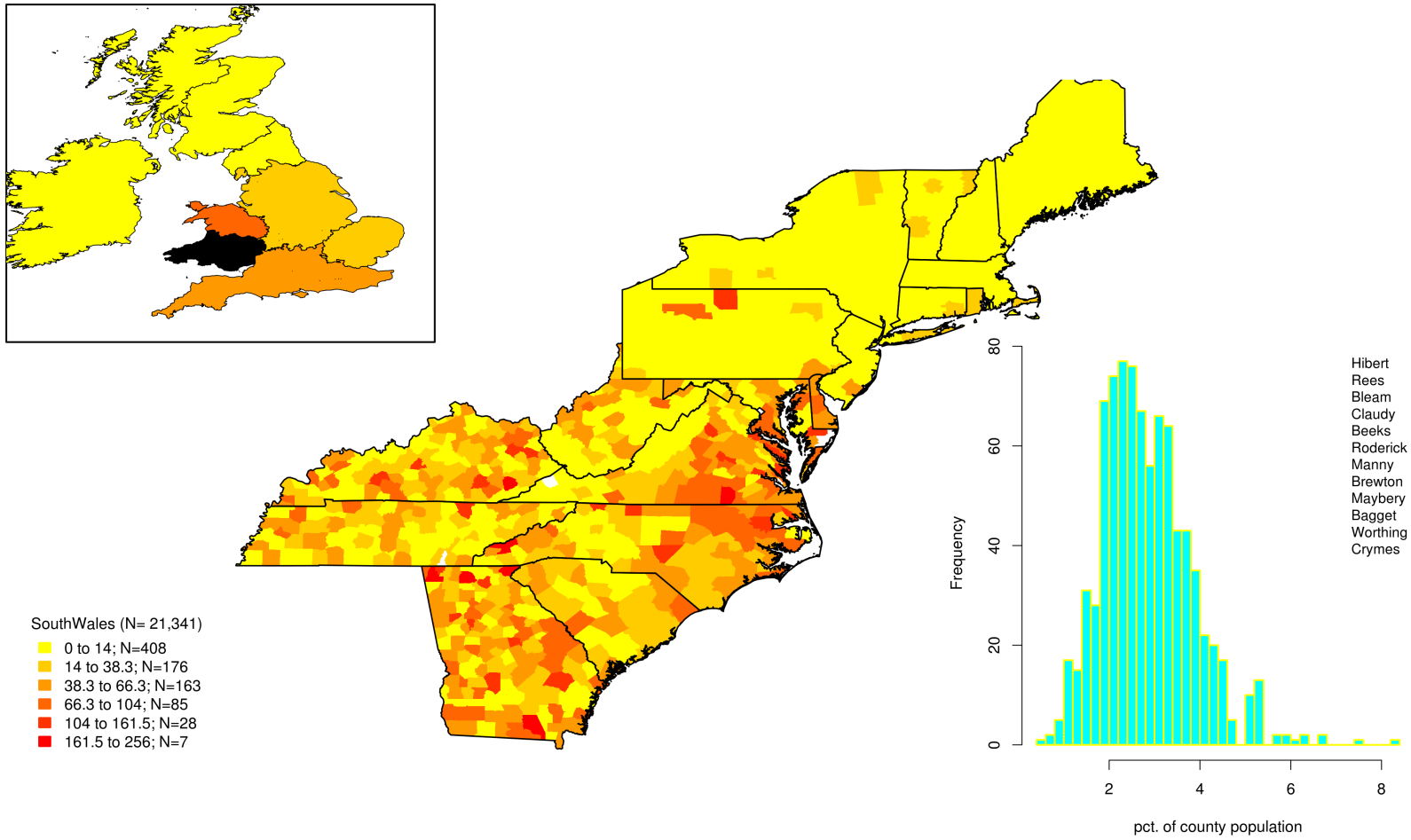

Figure 2.8: South Wales did not provide that many surnames, but those it did provide are noticably concentrated in the South, especially in the Southern Coast area. Its surnames are most often shared with North Wales, followed by South England. Located close to Bristol, it would have been a good recruiting ground for indentured servants destined for the Chesapeake. Puritanism was a largely English movement, making it unlikely that Welsh would have settled in the New England colonies.

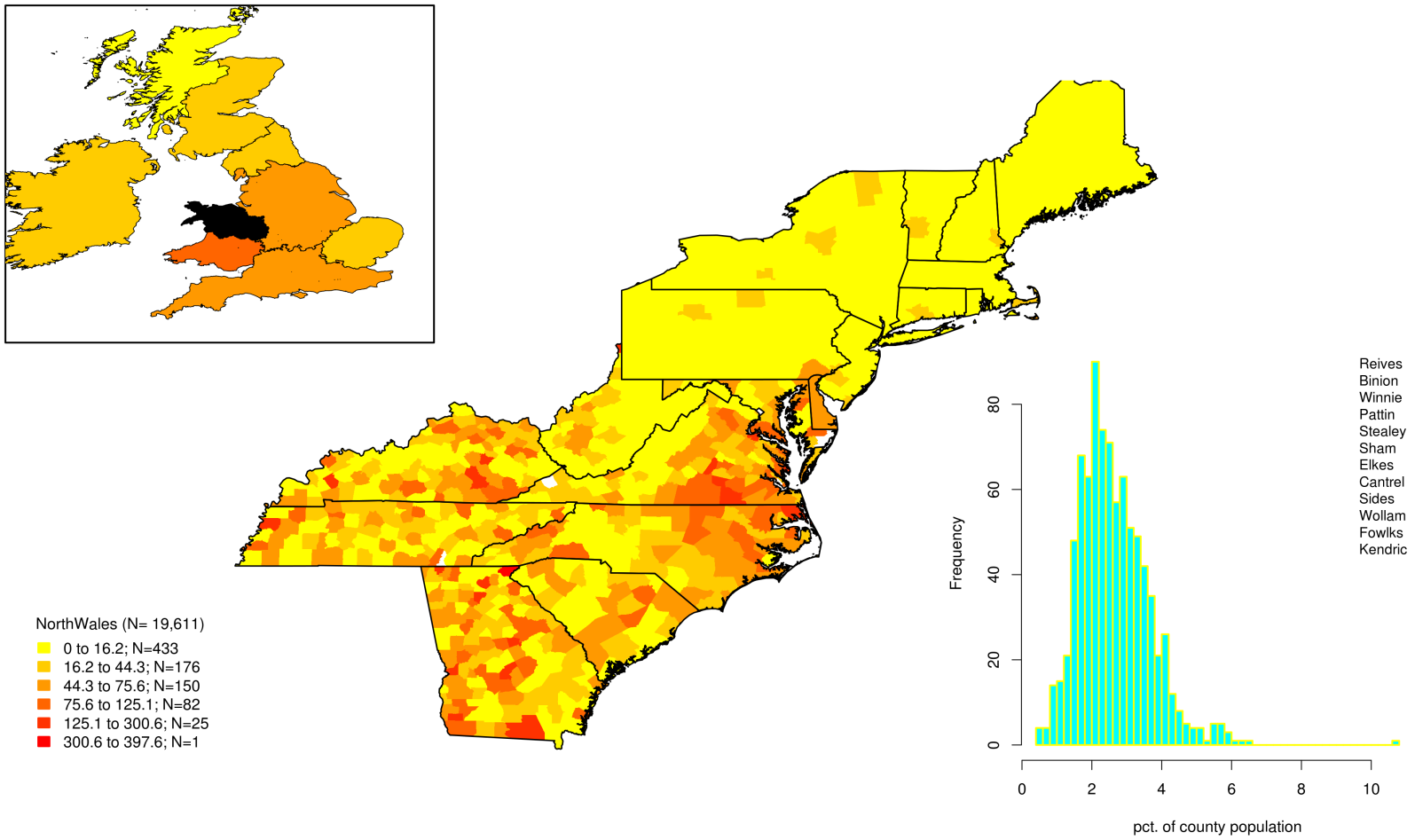

Figure 2.9: North Wales follows the Welsh pattern of a surname concentration in the South. Its surnames are most often shared with South Wales.

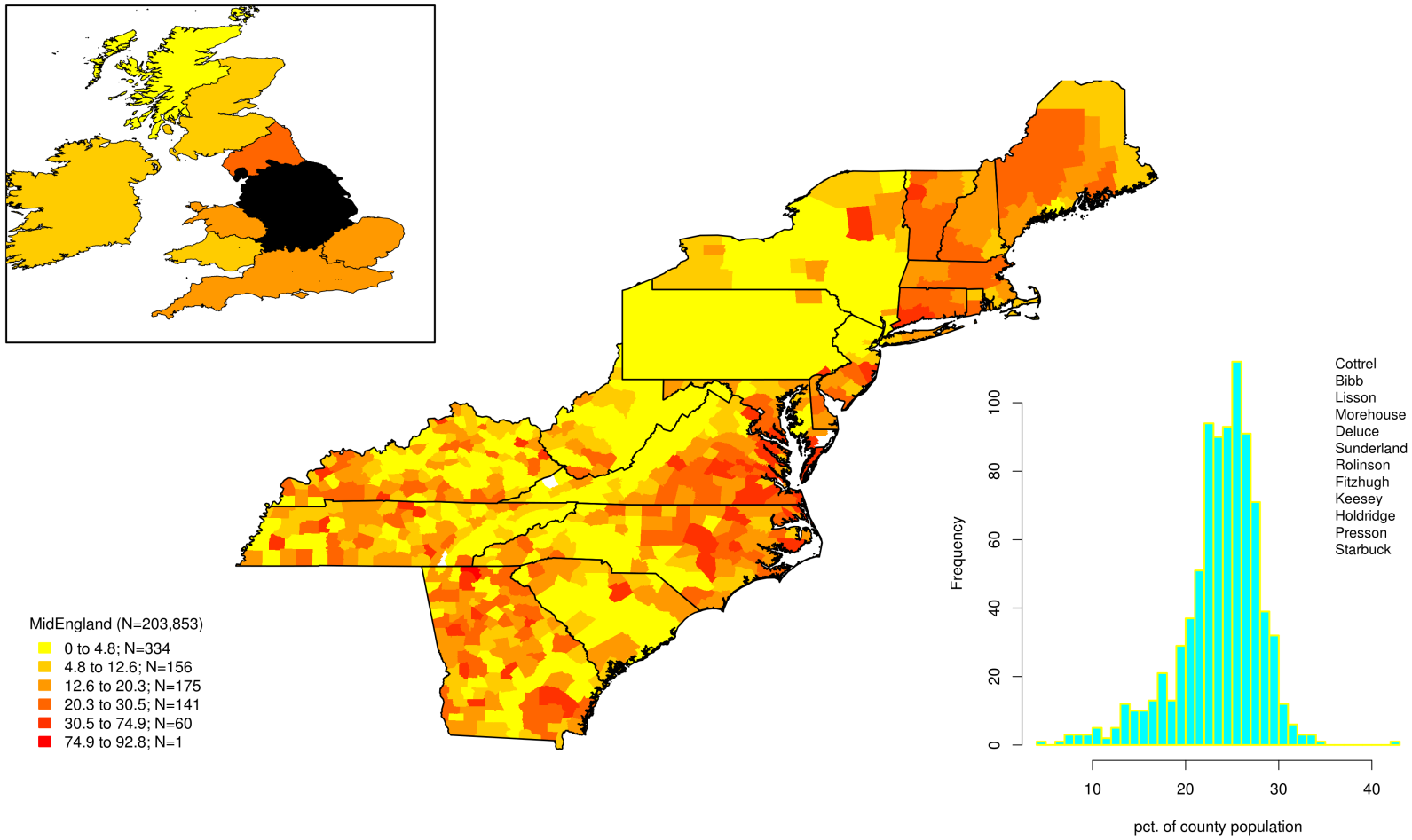

Figure 2.10: Mid England is the second largest surname category, after South England. The surnames are heavily represented both in New England and the Southern Coast. Mid England surnames are shared most frequently with North England.

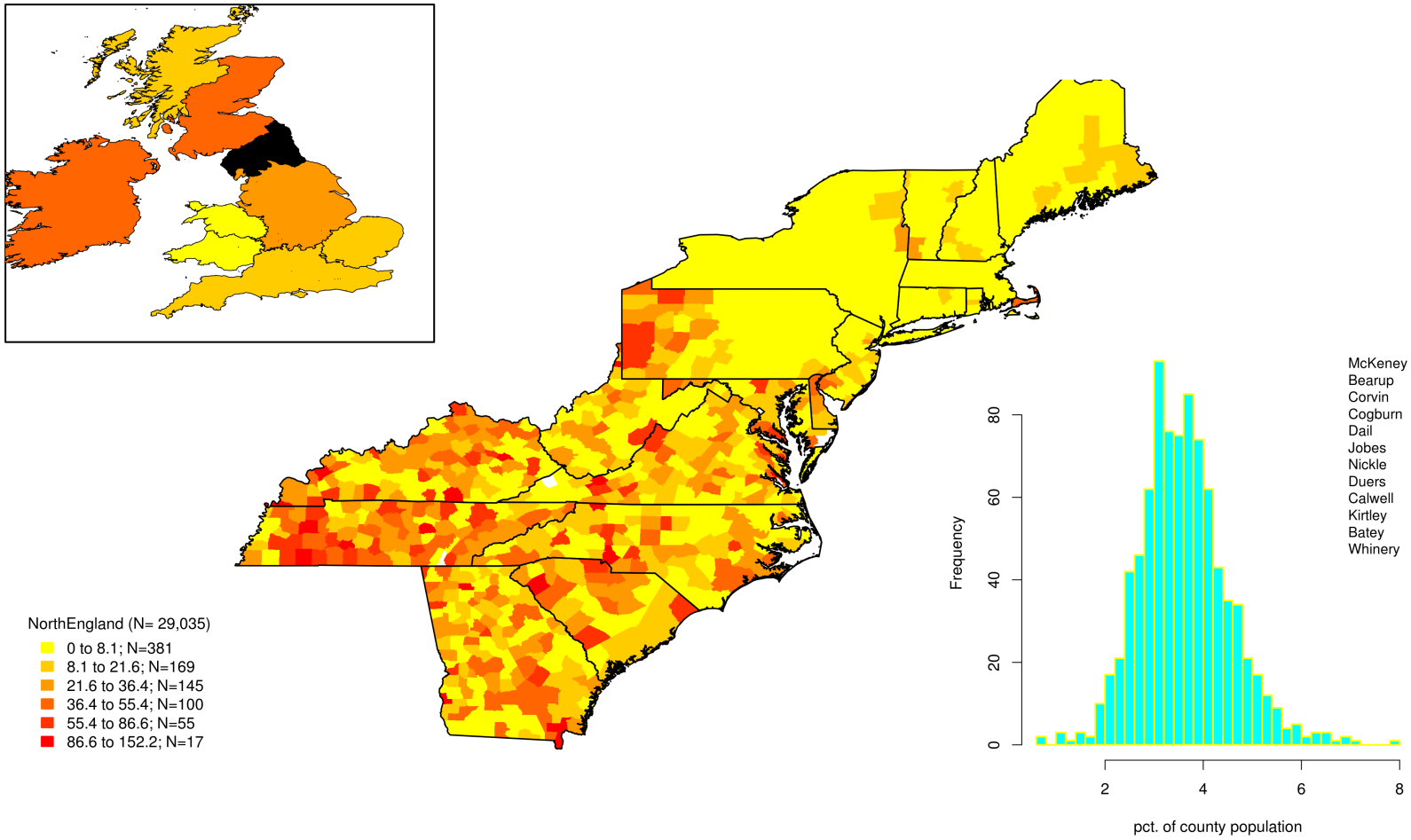

Figure 2.11: North England constitutes part of the region from which David Hackett Fischer's Borderers came. Their surnames are more prevalent in the South than in the North, and more in the Frontier region than on the Southern Coast. Nevertheless, they have a center of concentration in western Pennsylvania. North England surnames are shared most frequently with Scots-Irish, and then with Mid England.

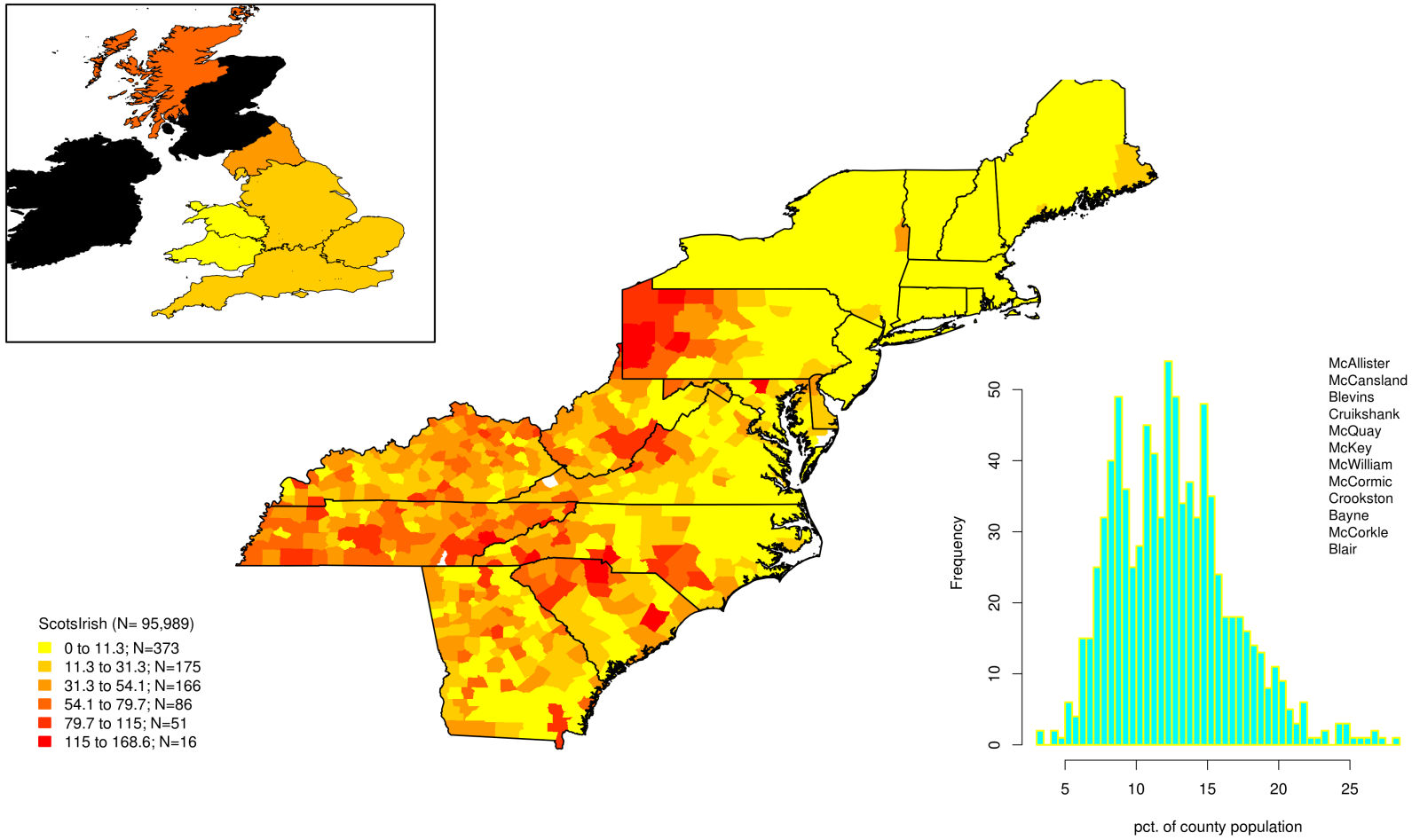

Figure 2.12: Scots-Irish (as I use the term here) includes Ireland and all but northern Scotland. Since the maps are based on a population of males at least 40 years old, whose fathers were born in the same state in which they (the sons) currently reside, the maps reflect the distribution of population before 1820, a time when settlers of Irish descent would have been overwhelmingly Borderers. The distribution of their surnames fits the conventional view of the Borderers: coming late, they settled far away from the coast, on the frontier. Like North England, Scots-Irish is more prevalent in the South than in the North and has its strongest concentrations in the west, including western Pennsylvania.

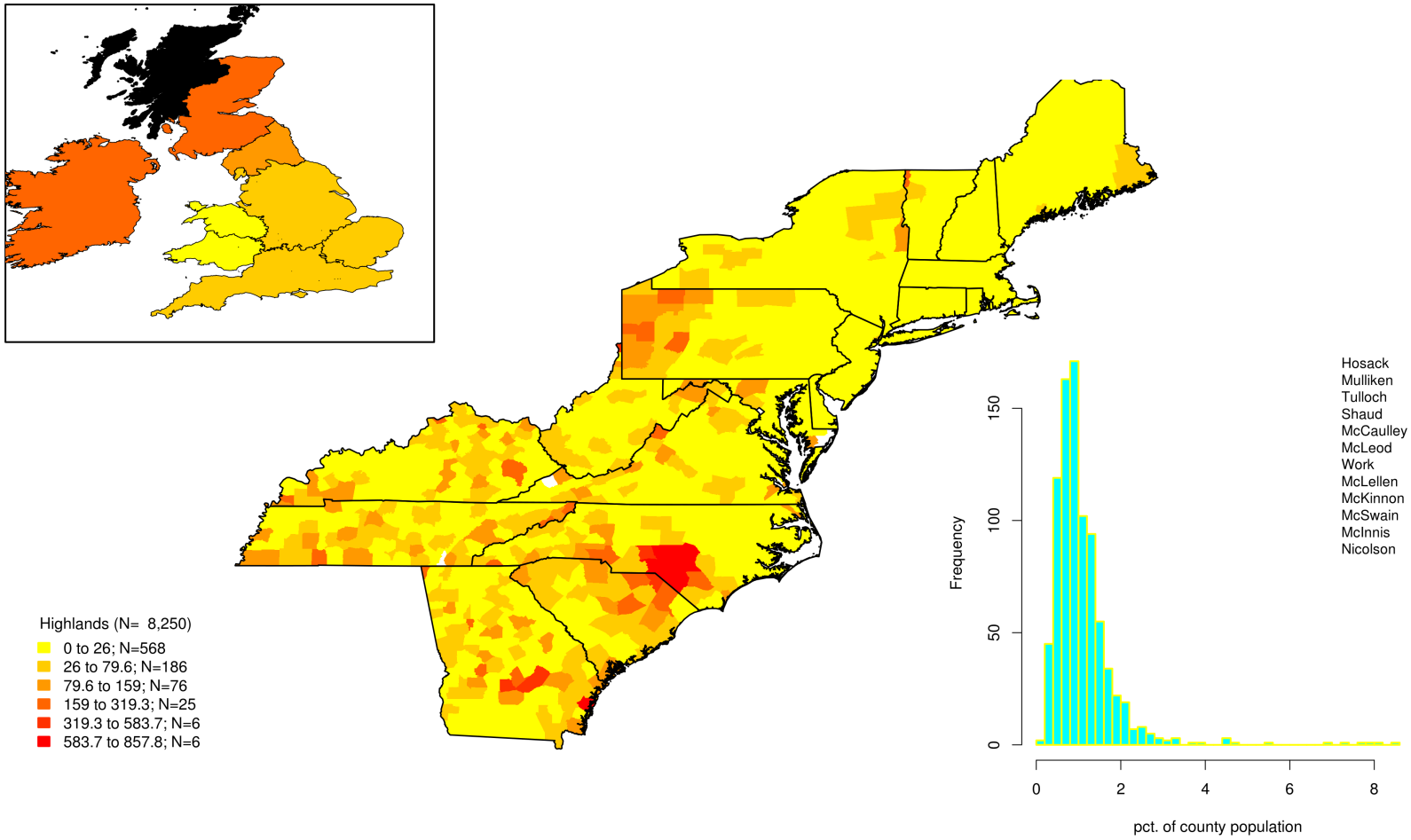

Figure 2.13: Scottish Highlands, which includes the northern islands, would have contained mostly Gaelic and Norn speakers in the 18th century, making these immigrants culturally distinct from the Scots-speaking Scots-Irish. Highlands surnames are shared most frequently with Scots-Irish. Though a much smaller immigrant group, they settled in a pattern similar to the Scots-Irish: on the frontier, with a concentration in western Pennsylvania. They also have a notable concentration around what is today Scotland County, North Carolina.

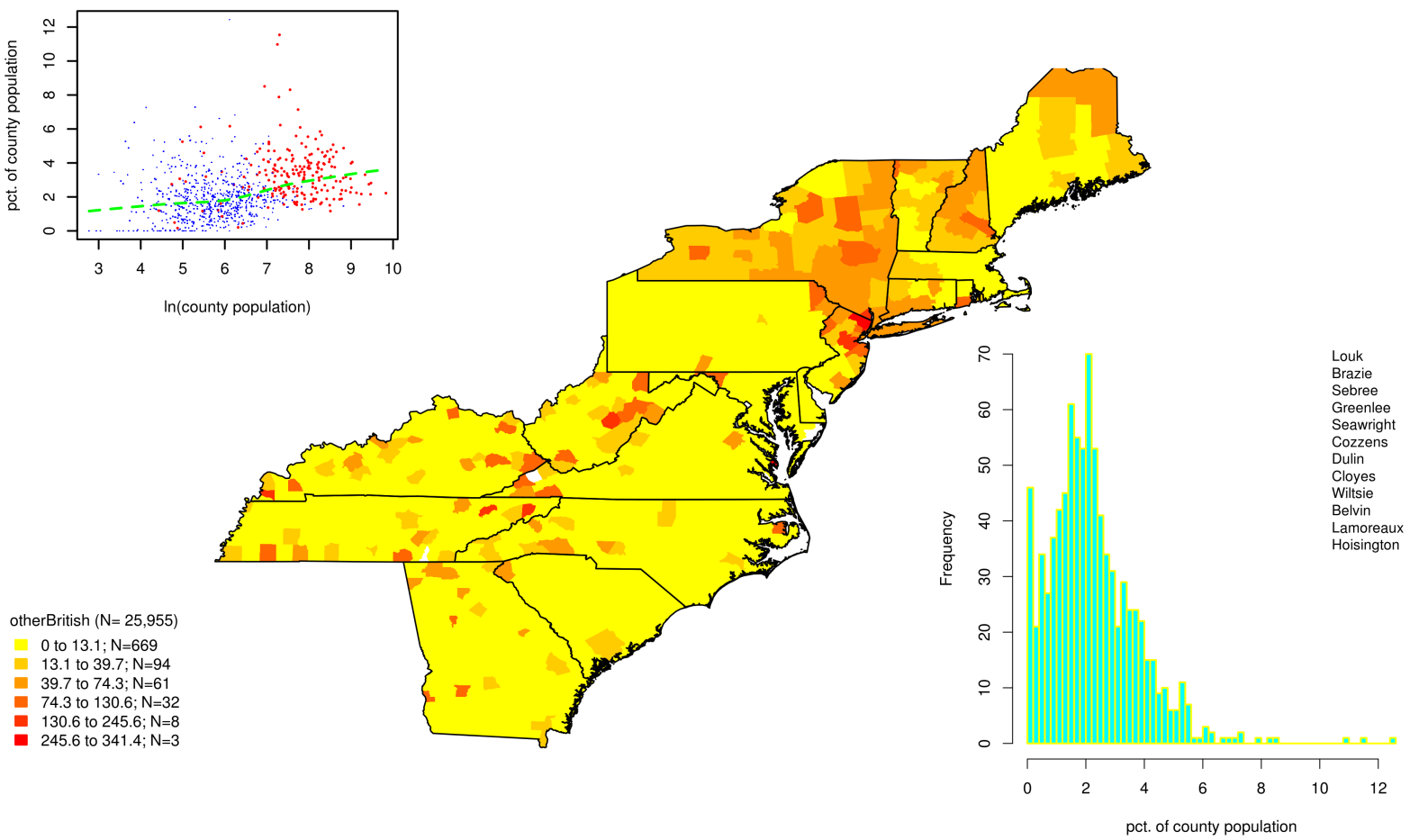

Figure 2.14: Other British Isles. These are surnames not regionally identifiable in the British census of 1881, but found in the United States, with at least one occurance of persons claiming to have a father born in the British Isles. Among the most likely reasons for their existence: they might be surnames newly invented by immigrants; they might be surnames that went through an orthographic change on only one side of the Atlantic; or they might be surnames of immigrants from another country who had a child while in transit to America through Britain. These surnames are much more frequent in the North (red points in the scatter plot) than in the South (blue points) and more frequent in counties with larger population (rising trend for the dashed green lowess smoother).

2.2 Foreign locations

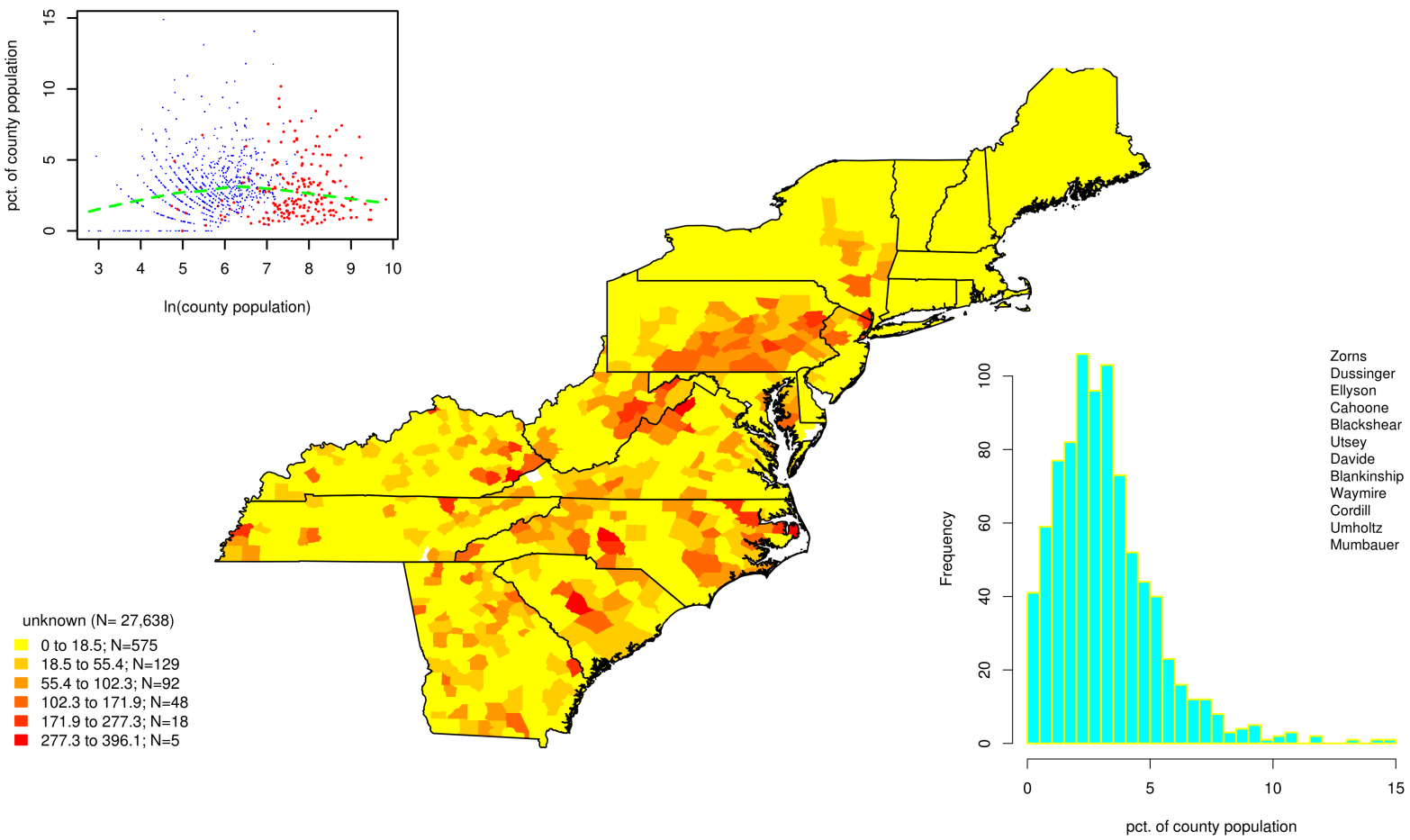

Figure 2.15: Unknown origin. These surnames, found in the US 1880 census, were not found in the British 1881 census, nor were there any instances of a person claiming a foreign location as father's birthplace. The highest incidence of these surnames is in areas of German and Dutch settlement in Pennsylvania and New York, and many of these surnames appear to be Germanic (see the randomly chosen examples shown on the map). That no person had a father born outside the US indicates that these names had been in the US at least two generations by 1840. The number of surnames of unknown origin was also high in the Southern Coast, where orthographic change may have led to new surnames among the oldest settlers.

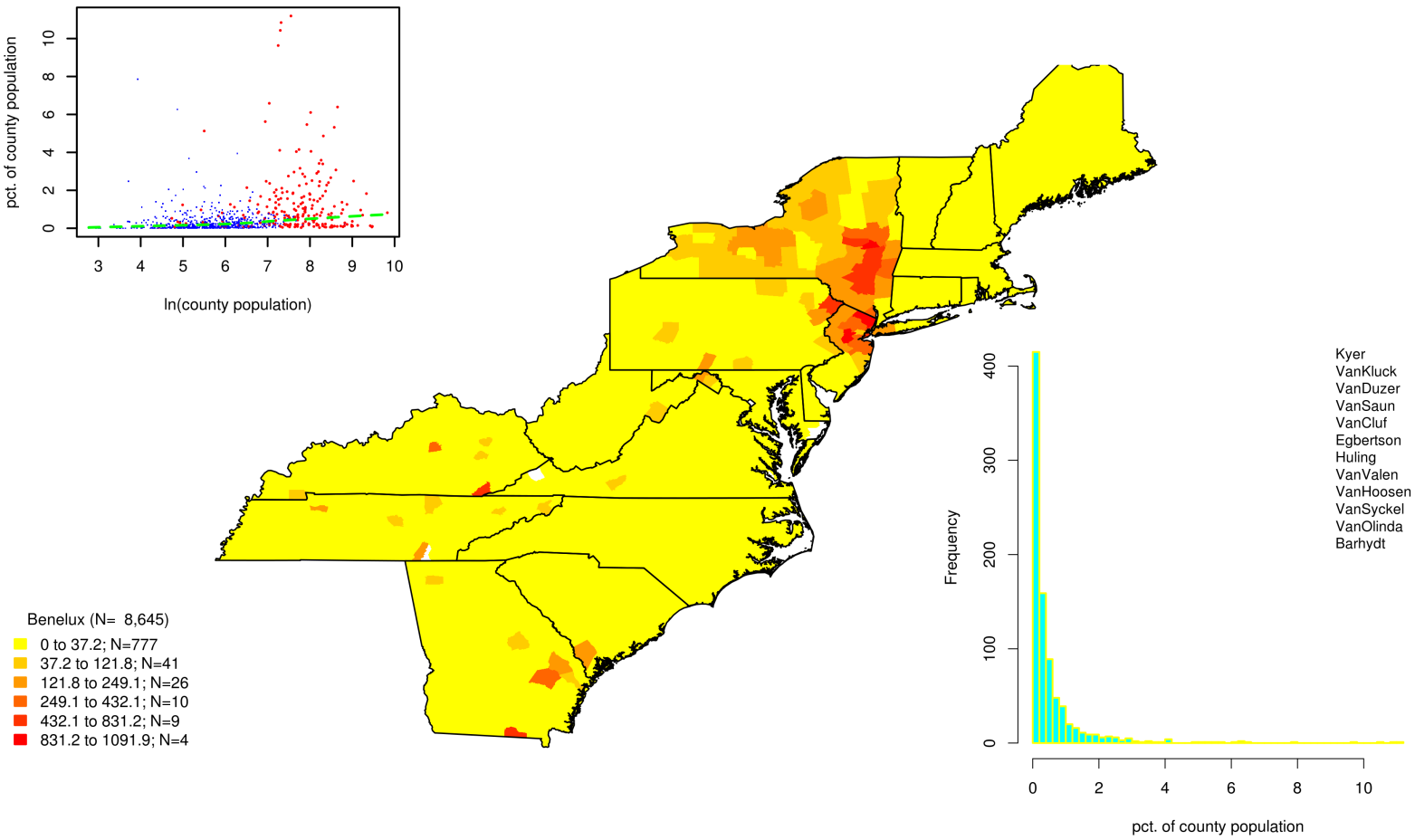

Figure 2.16: Netherlands, with Belgium and Luxembourg. The areas of early Dutch settlement are evident on the map: around New York City, and up the Hudson Valley.

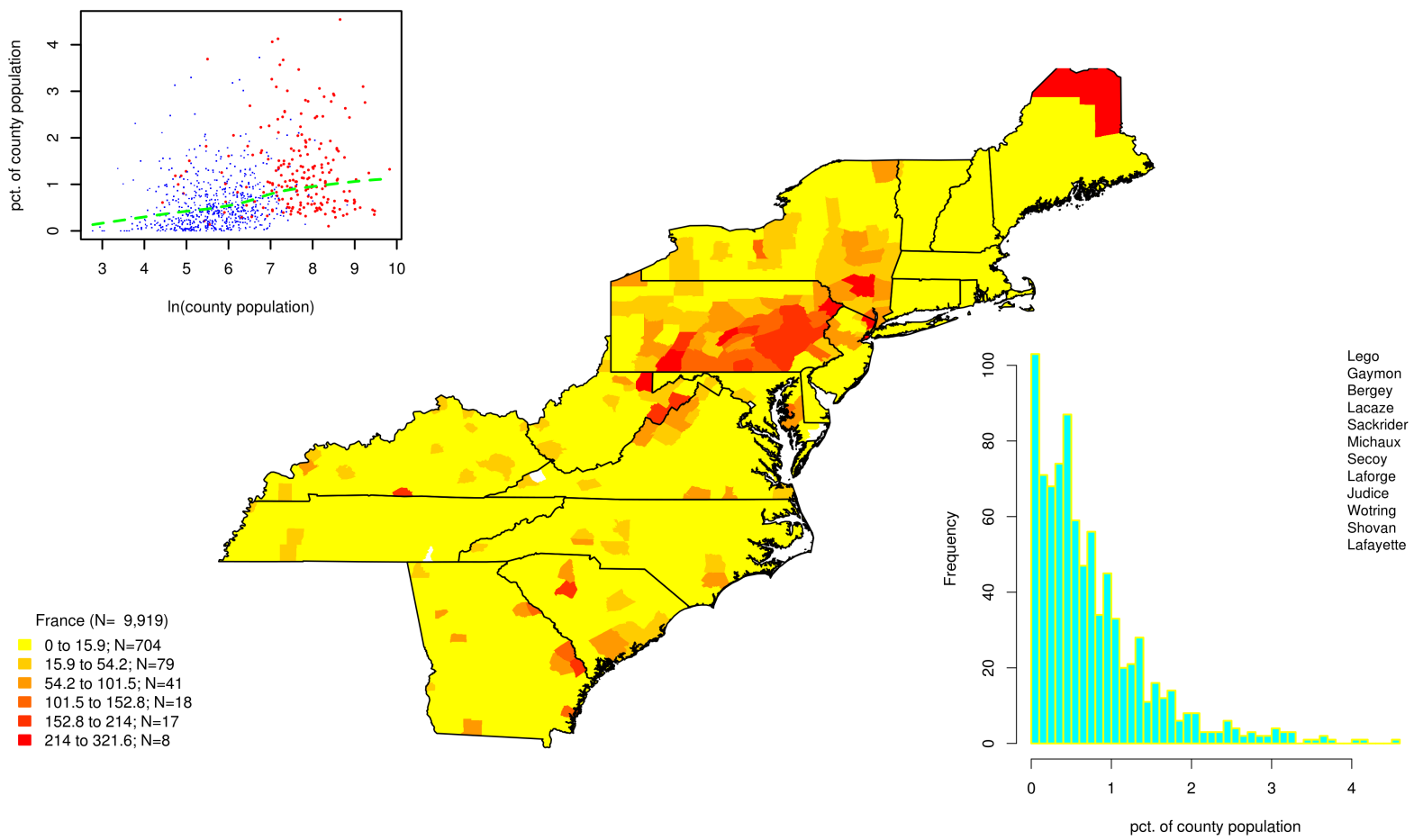

Figure 2.17: France. Most French surnames are associated with the areas of German and Dutch settlement, though there are other concentrations, near the border with Quebec and New Brunswick, and on the South Carolina coast. Dutch settlers included many Huguenots, as well as Walloons. These surnames are much more frequent in the North (red points in the scatter plot) than in the South (blue points) and more frequent in counties with larger population (rising trend for the dashed green lowess smoother).

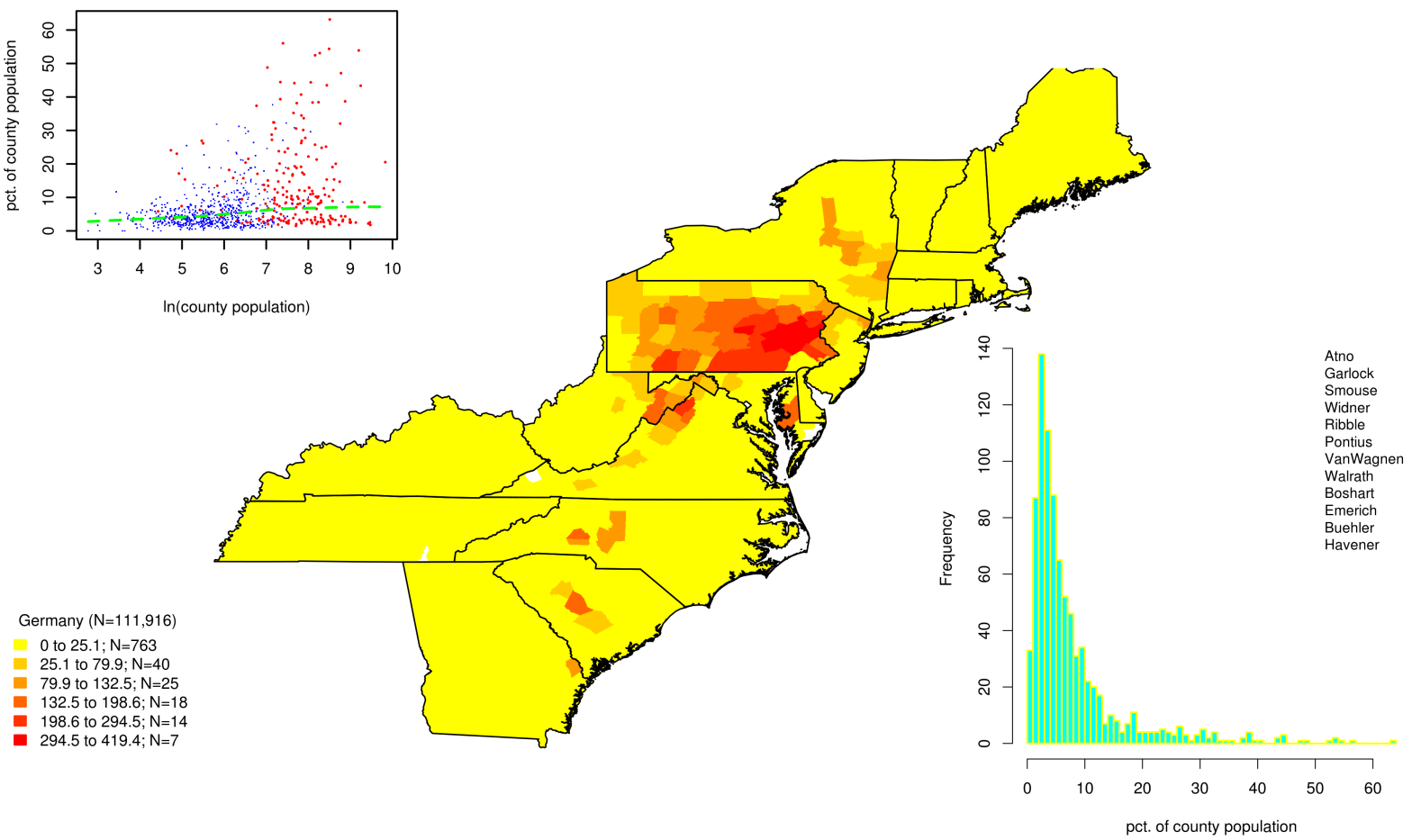

Figure 2.18: Germany. Among foreign locations, Germany has the largest number of persons represented on these maps. Pennsylvania and the Shenandoah Valley are the most conspicuous centers of concentration, while German surnames are conspicuously absent in New England and the Southern Coast. Many Germans sought refuge in the Netherlands during the Thirty Years War; this may account for the presence of German surnames in New York. Note that Germans in the South occupy a piedmont area: far from both the coast and the frontier.

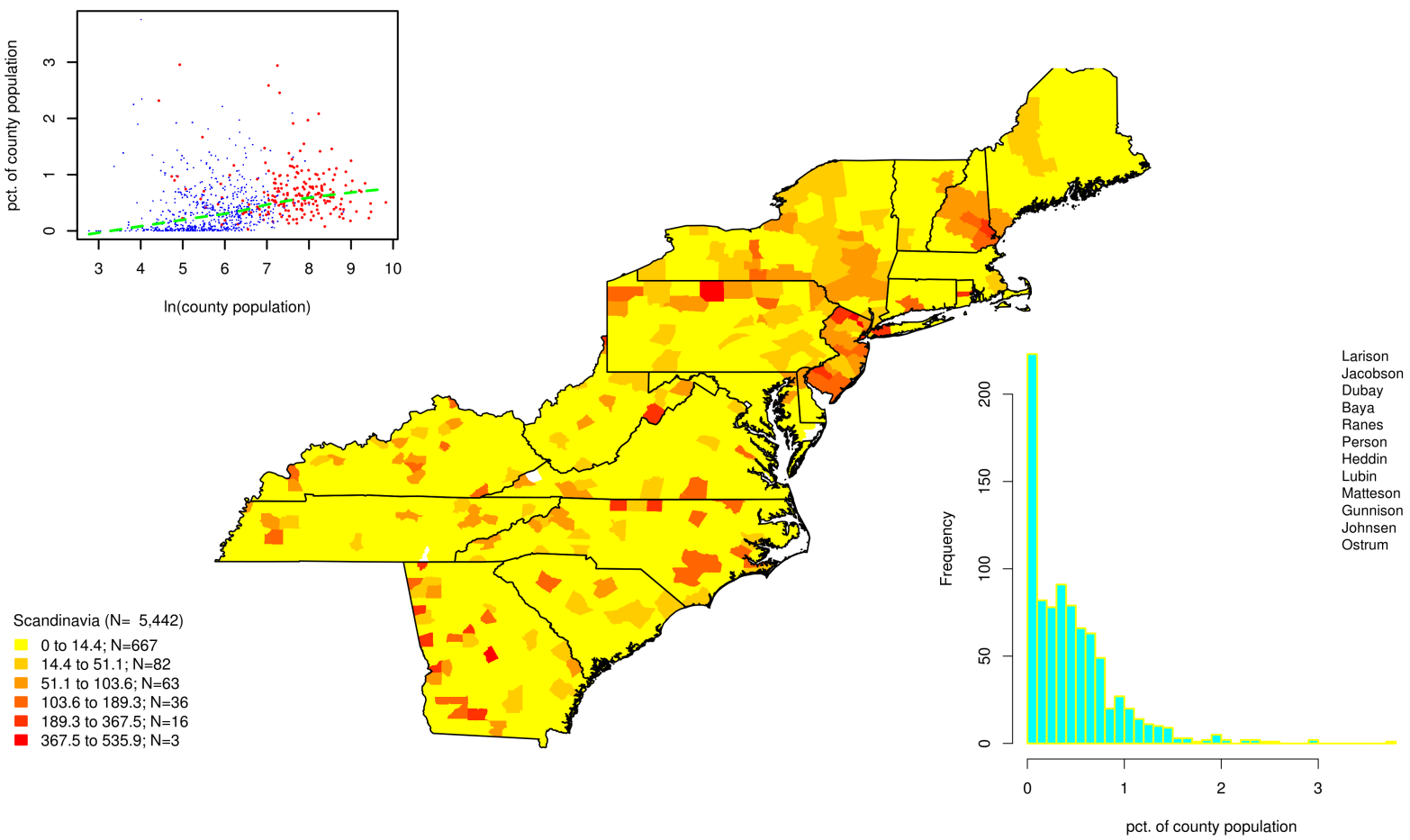

Figure 2.19: Scandinavia. The Swedish colony of New Sweden may account for the relatively high incidence of Scandinavian surnames in New Jersey. Dutch colonial enterprises often employed Norwegians and Danes, which may account for the relative density in areas of Dutch settlement. These surnames are more frequent in the North (red points in the scatter plot) than in the South (blue points) and more frequent in counties with larger population (rising trend for the dashed green lowess smoother). Nevertheless, these surnames are few in number and widely dispersed, indicating assimilated populations, unlike the late 19th century Scandinavian concentrations in the western Great Lakes.

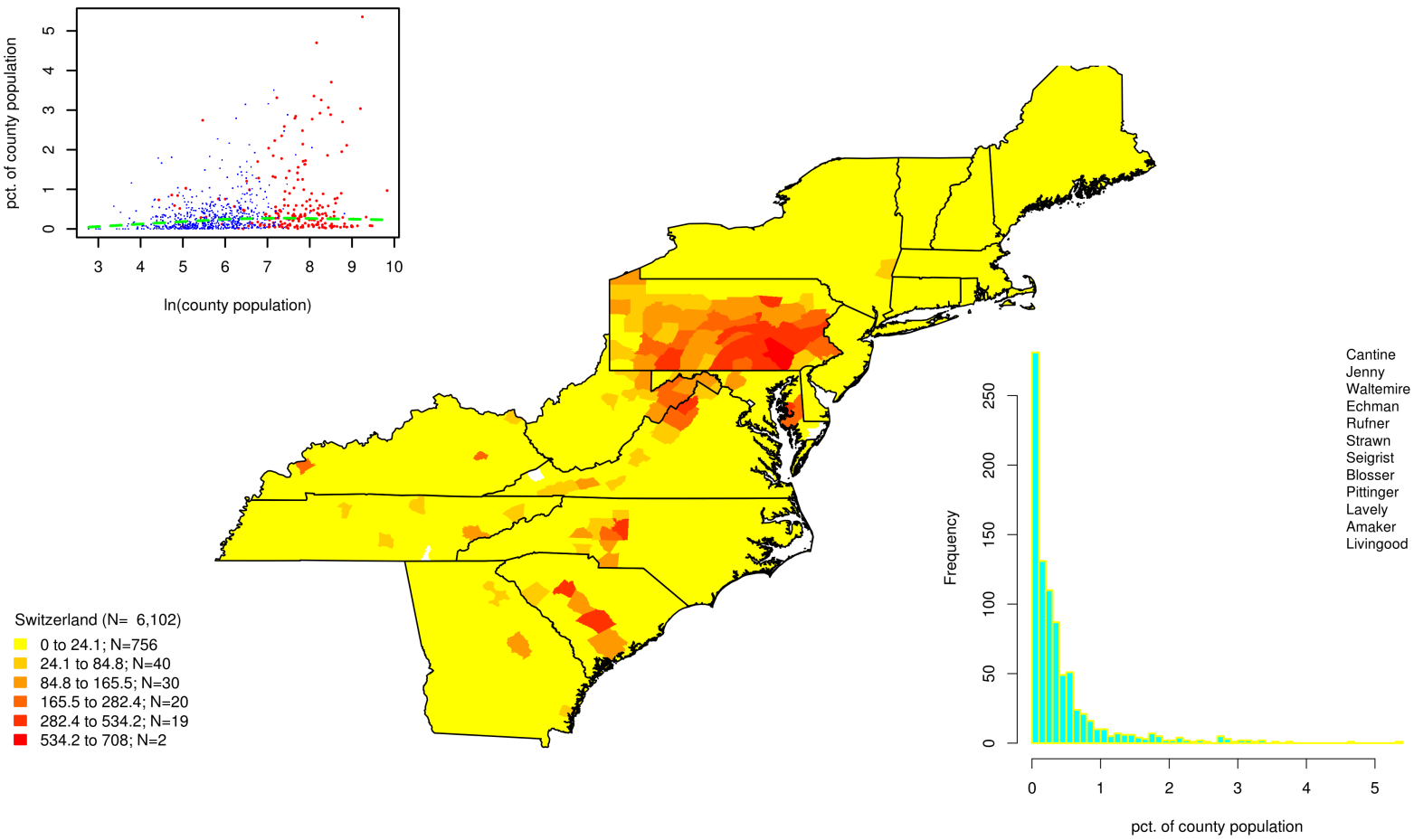

Figure 2.20: Switzerland. The distribution of Swiss surnames matches very closely the distribution of German surnames: Pennsylvania, then southward into the Shenandoah Valley. Most of the surnames appear in fact to be German, rather than French or Italian.

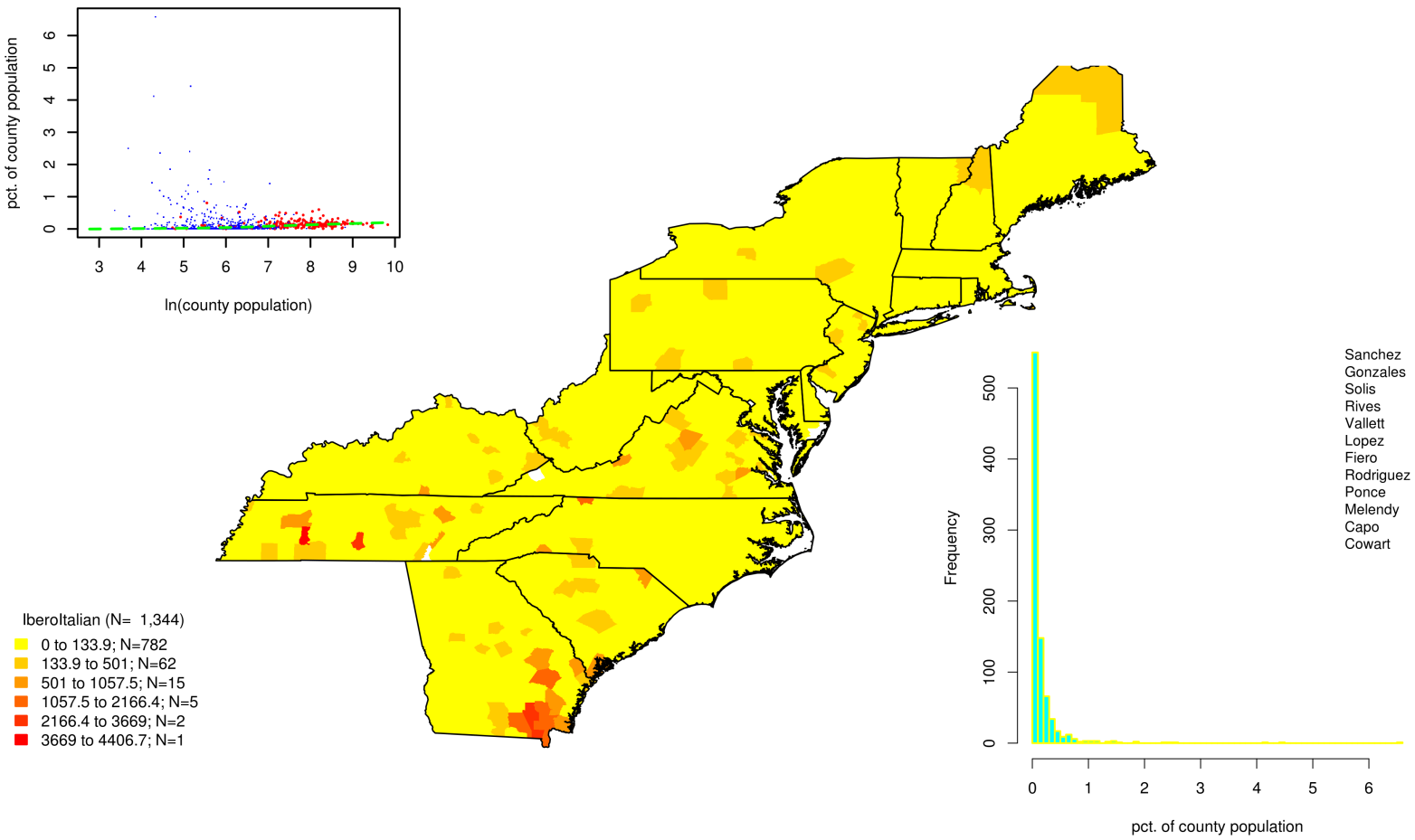

Figure 2.21: Ibero-Italian (Spain, Portugal, Italy). This surname category includes not only the Romance countries of south-western Europe, but also Latin America and Portugal's Atlantic islands. The only real concentration is in south Georgia, which would represent the influx of settlers from Spanish Florida.

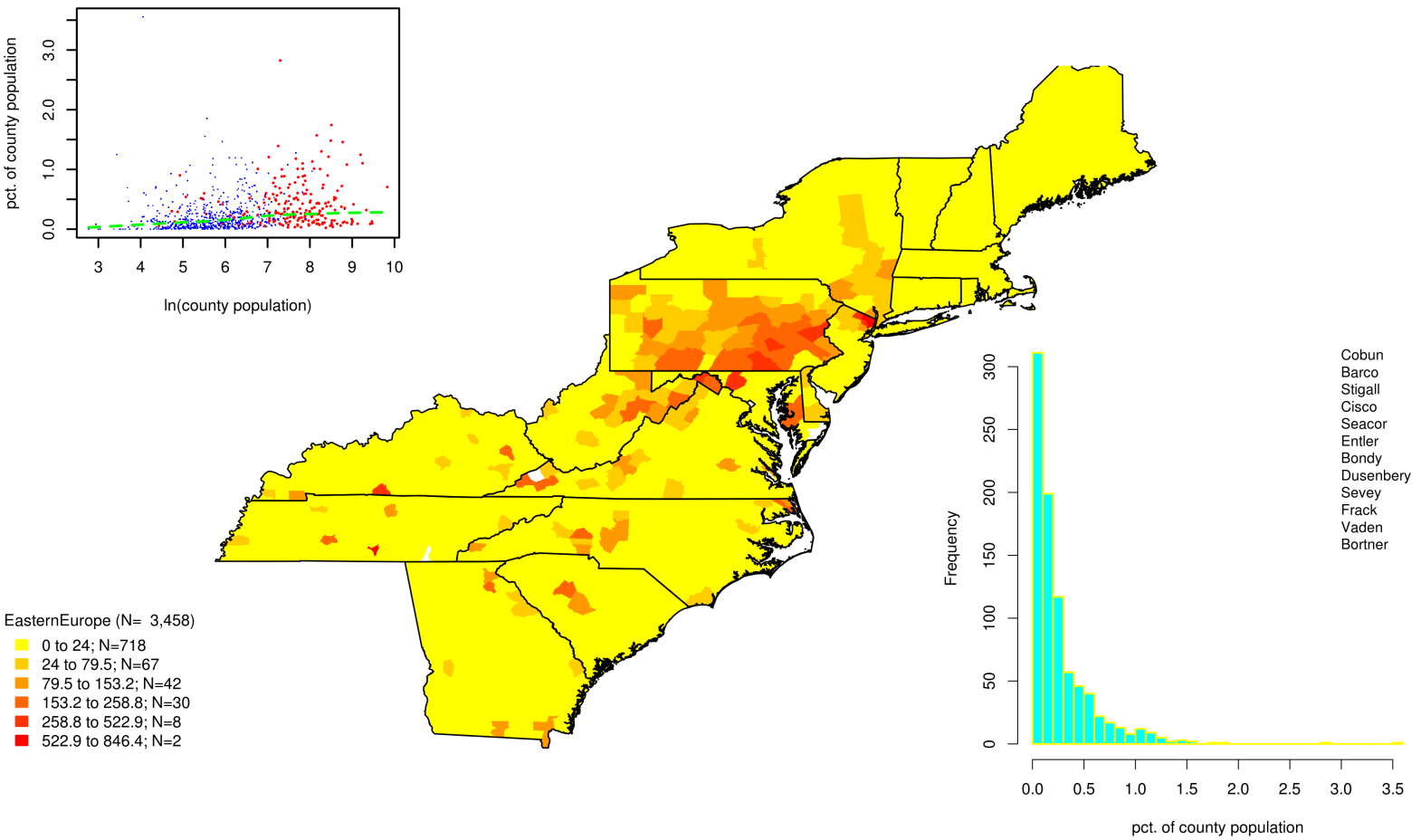

Figure 2.22: Eastern Europe. This surname category includes everything to the east and southeast of Germany, a vast area that has very low representation in early America. The distribution resembles that of Germany, and many of the sample surnames seem to be German.

3 Summary and conclusion

The first series of maps looked at migration from US states. For each person in the 1880 census, born in 1840 or earlier, the current county of residence was compared to the states in which the parents had been born. About half of all parents were estimated to be born before 1797 and virtually all before 1820, thus giving a view of how population for each state had dispersed in the period from approximately 1800 to 1880. The most striking pattern was the due westward flow of population, with only exceptional northward or southward movement, especially east of the Mississippi River. As migration moved further west, the flows tend to fan out and intermingle, and a strong southward flow into Texas becomes evident. Migration of blacks, however, was characterized by a transfer from the northern and eastern tobacco farming regions to the much more profitable cotton lands stretching from east Texas to Georgia, and up the Mississippi River as far as Kentucky.

Another striking pattern was the tendency of migrants from the North to avoid slave states, a tendency most noticeable in Missouri, which was skirted by Northern migrants, who then turned south to settle Nebraska and Kansas. The only exceptions were south Florida and the Texas Panhandle, both of which attracted Northern migrants. Immigrants from foreign countries in this period also avoided slave states, and again the major exception was in Texas, where Germans settled in large numbers in the area around Austin. In general, immigrants from foreign countries had a stronger presence in the West and the western Great Lakes.

The source states for the westward migration were largely on the eastern seaboard. The second series of maps took a closer look at the ethnic composition of those states. Largely settled from the British Isles, it has been hypothesized by David Hackett Fischer that these states differ in part because their original settlers derived from different parts of Britain. Surnames are collected from all males in these eastern states born in 1840 or before, who live in the same state in which their father was born. This is compared to the distribution of surnames in the British Isles (with imperfect coverage of Ireland) collected from all males born in 1840 or earlier, where the surname is assigned to their birth county. Yet a third set of surname assignments is created by identifying all US males, born in 1840 or earlier, whose father was born in a foreign location: this third set allows assignments to locations other than the British Isles. In this way, probable regions of origin are produced for the US easternmost counties.

The most striking pattern is that the mid-Atlantic states (and especially Pennsylvania) have by far the greatest non-British proportion of the population. New England is the most English area, and especially draws from South England

and East England. The South is also overwhelmingly English, but has more Welsh and Scottish than New England. Away from the coast, in areas which would have felt like the frontier in the late 18th century, Scottish and North English presence is stronger.

The areas predominantly settled from the mid-Atlantic states are the same areas that attracted foreign immigrants after 1840. The South and New England, areas with few non-British, were largely avoided by the post-1840 immigrants.

Bibliography

[1] Global administrative areas: Boundaries without limits. http://www.gadm.org/.

[2] Roger Bivand, with contributions by Hisaji Ono, and Richard Dunlap. *classInt: Choose univariate class intervals*, 2012. R package version 0.1-18.

[3] Roger S. Bivand, Edzer J. Pebesma, and Virgilio Gomez-Rubio. *Applied spatial data analysis with R*. Springer, NY, 2008.

[4] Minnesota Population Center. North atlantic population project: Complete count microdata. version 2.0. Machine-readable database, 2008.

[5] Minnesota Population Center. National historical geographic information system: Version 2.0. downloadable shapefile, 2011.

[6] David B. Dahl. *xtable: Export tables to LaTeX or HTML*, 2012. R package version 1.7-0.

[7] Steven Ruggles et al. United states 1880 integrated public use microdata series: Version 5.0. Machine-readable database, 2010.

[8] Daniel P. Faith, Peter R. Minchin, and Lee Belbin. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, 69(1-3):57–68, 1987.

[9] David H. Fischer. *Albion's Seed: Four British Folkways in America*. Oxford University Press, 1989.

[10] George McClelland Foster. *Culture and Conquest: America's Spanish Heritage*. Wenner-Gren Foundation for Anthropological Research, 1960.

[11] Raymond Gastil. *Cultural Regions of the United States*. University of Washington Press, January 1975.

[12] Frank E Harrell Jr and with contributions from many other users. *Hmisc: Harrell Miscellaneous*, 2012. R package version 3.9-3.

[13] M. Woollard K. Schurer. National sample from the 1881 census of great britain. Machine-readable database, 2003.

[14] Christian Kleiber and Achim Zeileis. *Applied Econometrics with R*. Springer-Verlag, New York, 2008. ISBN 978-0-387-77316-2.

[15] Nicholas J. Lewin-Koh, Roger Bivand, contributions by Edzer J. Pebesma, Eric Archer, Adrian Baddeley, Hans-Joerg Bibiko, Jonathan Callahan, German Carrillo,

Stephane Dray, David Forrest, Michael Friendly, Patrick Giraudoux, Duncan Golicher, Virgilio Gomez Rubio, Patrick Hausmann, Karl Ove Hufthammer, Thomas Jagger, Sebastian P. Luque, Don MacQueen, Andrew Niccolai, Tom Short, Greg Snow, Ben Stabler, and Rolf Turner. *maptools: Tools for reading and handling spatial objects*, 2012. R package version 0.8-16.

[16] Forrest McDonald and Grady McWhiney. The south from self-sufficiency to peonage: an interpretation. *American Historical Review*, 85(5):1095–1118, 1980.

[17] Erich Neuwirth. *RColorBrewer: ColorBrewer palettes*, 2011. R package version 1.0-5.

[18] Jari Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, and Helene Wagner. *vegan: Community Ecology Package*, 2012. R package version 2.0-4.

[19] Chad Oliver. *Ecology and cultural continuity as contributing factors in the social organization of the Plains Indians*, volume 48 of *University of California Publications in American Archaeology and Ethnology*. University of California Press, 1962.

[20] Benjamin S. Orlove. Ecological anthropology. *Annual review of anthropology*, 9:235–273, 1980.

[21] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

[22] Peter J. Richerson Robert Boyd. *Culture and the Evolutionary Process*. University of Chicago Press, June 1988.

[23] Marshall D. Sahlins. Poor man, rich man, big-man, chief: political types in melanesia and polynesia. *Comparative studies in society and history*, 5(03):285–303, 1963.

[24] Sonya Salamon. Ethnic origin as explanation for local land ownership patterns. *Research in rural sociology and development*, 1, 1984.

[25] The LyX Team. LyX 2.0.5.1- The Document Processor [Computer software and manual]. Internet: http://www.lyx.org, 2012.

[26] Joe H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.

[27] Hadley Wickham. The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1):1–29, 2011.

[28] Yihui Xie. *knitr: A general-purpose package for dynamic report generation in R*, 2012. R package version 0.8.

[29] Wilbur Zelinsky. *The cultural geography of the United States*. Prentice Hall, 1992.