DOCUMENT RESUME

ED 135 004 CS 203 217

AUTHOR Hirshberg, Jeffrey
TITLE Instant Mapping of American Regional Vocabulary.
PUB DATE 76
NOTE 44p.; Paper presented at the Annual Meeting of the National Council of Teachers of English (66th, Chicago, November 25-27, 1976)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *American English; Computational Linguistics; *Computer Graphics; *Dictionaries; Geographic Regions; *Maps; *Regional Dialects; Social Dialects; *Speech Habits; Word Frequency

ABSTRACT

When it is published in four or five years, the "Dictionary of American Regional English" (DARE) will be the official dictionary of the American Dialect Society. This dictionary will contain information concerning vocabulary, pronunciation, and grammatical forms which are restricted regionally or socially in American speech. One distinctive feature of DARE is its innovative use of the computer to graphically display geographical and social speech variation. Computer data were obtained through responses to questionnaires--consisting of 1847 questions covering aspects of daily life--administered in 1,002 communities in the United States between 1965 and 1970. The computer data are available to the DARE editors in two forms: maps, distorted to reflect population size and settlement histories, and data-summary slips. With its graphic linguistic patterns, DARE is expected to complement the "Linguistic Atlas." (Sixteen figures illustrate the text.) (JM)

INSTANT MAPPING OF AMERICAN REGIONAL VOCABULARY

Jeffrey Hirshberg
Dictionary of American Regional English
University of Wisconsin-Madison

The Dictionary of American Regional English--which we in its home offices affectionately call by its acronym, "DARE" -- will be the official dictionary of the American Dialect Society when it is published by Harvard's Belknap Press within the next four or five years. The information it will contain will concern vocabulary, pronunciation, and grammatical forms which are restricted either regionally or socially in American speech. Such information will have been the result of analyses of two kinds of data. The first includes private collections of word lists donated to the dictionary as well as such written sources as the several publications of the American Dialect Society and data gleaned from some 200 regional American writings and 115 early American diaries.

Our present concern will be with the second kind of data being employed by DARE. This consists of the information derived from the responses to questionairs administered in 1002 communities across the 50 States between 1965 and 1970. The DARE Questionair consists of 1847 questions, designed to cover aspects of everyone's daily life -- from terms one uses to describe the weather to the kinds of foods he eats and how he spends his occupational and leisure time. In contrast to the Linguistic Atlases, which selectively choose to specify distributions of words and expressions known to vary geographically and socially, the dictionary hopes to

provide a more exhaustive account of American dialect variation, though it should be understood that a total description of all dialectally significant elements is an ideal not humanly attainable.

Nevertheless, the questionair's 1847 questions were answered, all or in part, by 2752 informants -- all chosen because they had lived their entire lives in the area in which we found them. Besides lexical usage -- in phonetic transcription where appropriate -- the DARE fieldworkers recorded the community-type, age, education, race, and sex of every informant. All of this information has now been placed on computer tapes, along with the nearly 2 1/2 million different responses to the many questions of the questionair. By means of what is known as an "interactive processing program" -- "interactive" means essentially that one can carry on a conversation with the computer -- we are now eliciting data from the computer and preparing our entries. What we would like to do this morning is to demonstrate some of the varieties of information the computer is helping us to determine and to understand.

It should be useful before we proceed to examples to take a look at the two forms in which DARE's editors see the computer data. The first of these are maps which have been distorted to reflect population size and settlement histories -- the criteria which governed selection of sites for DARE's fieldwork (figure 1). Though distorted, the map is quite easy to get used to. Notice that New York, Pennsylvania, Illinois, and California loom large -- as their populations would lead us to suspect. There is exactly one space on the map for each of the 1002 communities in which

interviews for DARE took place, and there are no extra spaces. Thus, the 1002 communities are spread evenly on the map. The maps for our examples will also be distorted, but they have been redesigned by cartographers at the University of Wisconsin for ease of presentation.

The second way in which the editors get their computer information is on a data summary slip. Figure 2 represents an actual slip of the type prepared for every word in the dictionary. The word or expression will appear under "Response," the appropriate question number will appear to its immediate right, and the number of informants giving the word at the far right of the top line.

We'll look at the rest of the slip later, but let's recognize that we must at all times consider the question to which a particular response is given. This should be clear: a word may well mean one thing in one geographical area or to one socially distinct group of people while denoting quite another thing to a different group. In fact this is the first thing which the computer-generated maps force us to recognize. An example is bayou, which showed up as a response to two questions. Our question number 61 reads, "What do you call a small stream of water not big enough to be a river?" The most frequent responses were brook and branch. Brook is predominantly a northern term; and over 60% of the 230 informants who used it hailed from New England. Over 92% of the informants who said branch, on the other hand, were from the South and Central states, where, I am told, bourbon and branch water is a favorite means of assuaging a thirst. The interesting fact is that very few informants who gave branch were from Louisiana. The reason

is clear at a glance at figure 3. The most frequent expression for a small stream in Louisiana is bayou, reflective, of course, of the French settlement of the area. The use of bayou in this sense extends to only a few states neighboring Louisiana. But this is not the only meaning of bayou. Question 71 asks the name for "a stretch of still water going off to the side of a river or a lake." Figure 4 shows the geographical distribution of bayou when used in this sense -- the sense which is most often captured by backwater, lagoon, or slough. Two facts are revealed by this map. First, though most heavily concentrated in the South -- look at Texas and Tennessee in particular -- bayou meaning "slough" is current as far north as Wisconsin and Michigan -- which had its share of French settlers. The word is rare north of North Carolina in the east, and west of the Rockies. But notice also -- and this is our second point -- that bayou denoting lagoon or slough does not occur in Louisiana. And any self-respecting Louisianan will tell you he knows the difference between a bayou and a lagoon.[2]

The comparison of maps in this way has proved an important editorial technique -- one might venture to call it an editorial principle -- for DARE. In precisely this way, for instance, the geographical distribution of a term like armful (figure 5), which appears to be widespread -- and therefore not properly to be included in the dictionary -- takes on significance when contrasted with the more regionally specific armload (figure 6) to denote "the amount of wood a person can carry in both arms." Notice the triangle-like pattern extending with its base at the Pacific Coast towards southern New Jersey for armload (figure 6).

Then notice that this same triangle is the area of least density of distribution for armful (figure 5). In his Word Geography of the Eastern United States Hans Kurath notes armful to be the general Northern expression and to be replacing the old Midland form armload.[2] Our data coincides with that of the Linguistic Atlas of New England from which Professor Kurath drew his conclusion, but it also shows the currency of armload throughout the Midland, Central, and Western states, and the extension of armful into the South since the field-work for the atlas was completed in the '30's. It certainly seems to be as prevalent in most areas of the South as the genuine Southern expression turn or turn of wood (figure 7). The computer program allows us to combine expressions whenever we so desire. Here turn and turn of wood have been mapped together to demonstrate their clear southern orientation. Again, the comparison of maps -- such as we have just made for armload, armful, and turn -- is a significant aspect of DARE editorial procedures, and maps will be included with the entries in the dictionary wherever they are deemed particularly revealing.

Such a map will probably be included for baby buggy (figure 8). Though widespread, the geographical distribution of baby buggy seems to stop short of the east coast around the Appalachian Mountains. East of the Appalachians, and in fact often overlapping the area of baby buggy between the Appalachians and the Mississippi River, the word most often used to describe the kind of vehicle for a baby or small child which allows him to lie down, is baby carriage (figure 9). But notice here the curious gap in the use of baby carriage in south-western New Jersey and Southern Pennsylvania. There, a most frequent expression is baby coach (figure 10).

Our computer program also allows us to contrast as many as three expressions on a given map, such as we have done in figure 11. We have mapped here three expressions used to designate that which hangs below the roof to carry off rain-water. Notice the overlap of the two northern terms in Michigan: ● represents both eave troth and eaves troth, a predominantly northeastern expression, while O , which is mostly concentrated in the northwest and upper midwest, designates eaves trough. For a third contrast, guttering is represented by Δ ; it is most frequent in the upper South and lower midwest.

So far we've been looking at examples of strictly geographical distributions. But DARE's interactive program allows us to determine social as well as regional patterns. And these social patterns may be either mapped or printed in tables. Look again at figure two , the data summary slip, which indicates the subdivisions of each of the five social variables for which we have stored data. There are five community types, ranging from Urban to Rural, three age types, and so on. In tabular form we may generate a chart of any one variable, say race, or two variables, say race against age, or three variables, which will result in a series of charts. If we want to know how many old women for a given response come from the largest community size, we would plot a frequency count of age against sex against community size. Most often, however, we are simply concerned with the breakdown of one variable. The numbers printed beneath the subdivisions are the percentages represented by each subdivision within DARE's entire informant population. 66% of our informants, for example,

were old -- for our purposes, 60 years and older. The computer operator will indicate in the spaces below these numbers the exact population breakdown -- either in percentages or number-count -- for each response. Now, let's say that 67% of the informants giving a particular response are old. Do we conclude that the phrase is old fashioned? Not at all; for what is crucial here is not the actual percentage, but the actual percentage relative to that of the entire DARE population. And a 1% variation is too negligible to be considered statistically significant. On the other hand, 80% of the 195 informants who told us that an addition to a house is called an ell were over 60 years old. The difference between this and the expected 66% is substantial, and only 2% of the informants were young: we therefore have reason to believe that ell is relatively old fashioned. A stronger statement may be made about the use of tidies to designate "knitted or crocheted pieces placed on the back and arms of a chair for decoration and cleanliness." Only 1 of the 128 informants using the word was young, and a full 88% were old; tidies seems to be passing out of use. So, for that matter, is antimacassar, another response to the same question, for which 76% of 190 informants were old. What is of particular interest here is that antimacassar seems to be passing out of use in some areas, while still used by the middle-aged and young in others. This is seen from figure 12, an age map of the informants saying antimacassar. Notice, for instance, that almost every informant using the word in New England is old, but that the situation is quite different in Wisconsin. A New Englander on the DARE staff says she knew the word as a book word -- an

observation that corresponds with another of our findings, that 55% of the informants using antimacassar had at least two years of college.

We can do precisely this kind of mapping for all five social variables, and, in fact, for every subdivision as well. We can command the computer, for example, to specify the race of each informant -- as we have just seen it specify the age for informants using antimacassar, or we can command it to map just the black or white or Indian or oriental informants. Thirty informants told us that the time of day just before the sun comes up is called before-day. Figure 13 displays the geographic distribution of the phrase: though current in some parts of upper New York State, and though one informant was from New Hampshire, the expression is predominantly Southern. After observing from his tables that 11 of the 30 informants were black, the computer operator commanded the computer to print a map of just the black informants using before-day. The result is figure 14: all black informants are Southern. That 11 of 30 -- or 37% in contrast to an expected 7% -- of the total number of informants were black is also significant. Our impression that before-day is a Southern black expression is reinforced by several facts about the informant population. The word is current among all age groups of black informants, while 17 of 19 white informants were old: this substantiates the observation of one informant from South Carolina who told us that before-day is said by "old-fashioned people and Negroes." Secondly, the expression is said by all educational levels amongst the black group, better than half of whom had attended college;

while 14 of the 19 white informants had never attended college. Finally, the phrase is known in Jamaican English and is cited from an 1877 written source in the Cassidy-LePage Dictionary of Jamaican English.[3] Before-day is a black speech form which has spread among rural and relatively uneducated southern white speakers.

A matter about which there has been a growing body of literature since the 1930's is the use of anymore in a positive, affirmative sense. Our question 24 asked informants to fill in the final word of the following sentence: "People used to walk a lot, but everybody drives a car _____." Now, nowadays, and these days were frequent responses, but 110 informants said anymore: "People used to walk a lot, but everybody drives a car anymore." Figure 15 reflects the geographical distribution of this usage, which, though not actually restricted, is certainly more frequent in some areas than in others. As teachers of English you'll probably be gratified to know that the phrase is most common among informants with less than a high school education, but the phrase does seem to be genuinely regional in Kentucky, Indiana, Minnesota, Iowa, and Oklahoma. In Indiana and North Dakota, better than half of the informants using the word had attended college. This suggests, in turn, that the fact that 31% of all the informants using anymore in this affirmative sense attended college is not necessarily reflective of the educational plight of the country. I'll leave you to ponder that for yourselves.

One last comment. The dictionary will not be and does not pretend to be a substitute for the Linguistic Atlases which are

its editors' constant companions: its informants, though considered representative of their areas, are too wide-spread to allow for the specificity of dialect boundaries so meticulously detailed in LANE and LAUM. This should go without saying. Nevertheless, what the "instant mapping" of American Regional vocabulary does allow us to do -- and the reason for which DARE should properly complement the atlases -- is to obtain in graphic form an overall pattern which the atlases cannot yet provide. One of the interesting aspects of this picture, for a final example, is its demonstration of the encroachment of Southern expressions in the central states, forming a quite consistent pattern of isoglosses moving northwest from the Central-Atlantic or South-Atlantic Coast, cutting through the upper central states and most often peaking in central Illinois, before turning sharply southwest through Missouri and the middle of Texas. This is the pattern reflected in figure 16. And similar patterns have been found for the North, the Northeast, the Central Atlantic, and the Western States.

What I hope to have suggested is that one of the features which makes DARE an exciting lexicographical project is its innovative use of the computer to graphically display geographical and social variation in American Speech. Students of dialect in the past have had to spend hours tabulating and mapping the same kind of information, and often not even attempting the more detailed or thickly-filled maps, thus missing completely what DARE, with its interactive mapping capacity, can readily obtain in a few seconds -- anymore.

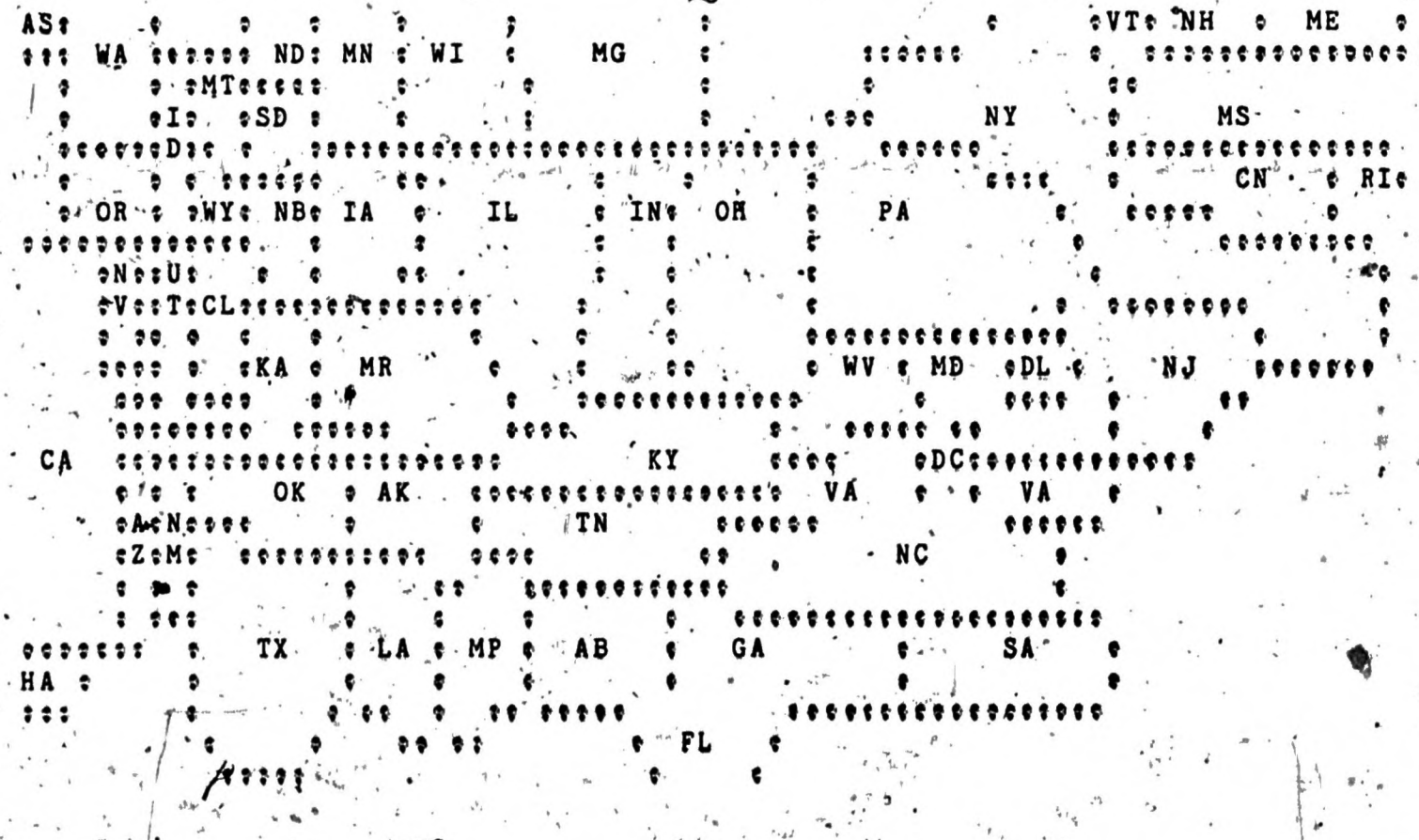
Footnotes

[1]DARE's data for bayou is confirmed by toponymic study. See Robert C. West, "The Term 'Bayou' in the United States: A Study in the Geography of Place Names," abstracted in Annals of the Association of American Geographers, 43(1953), pp. 197-198.

[2]Hans Kurath, A Word Geography of the Eastern United States (Ann Arbor, 1949), p. 57.

[3]F. G. Cassidy and R. B. LePage, Dictionary of Jamaican English (Cambridge, 1967), p. 36, s.v. "Before-day."

Figure 1



13

14

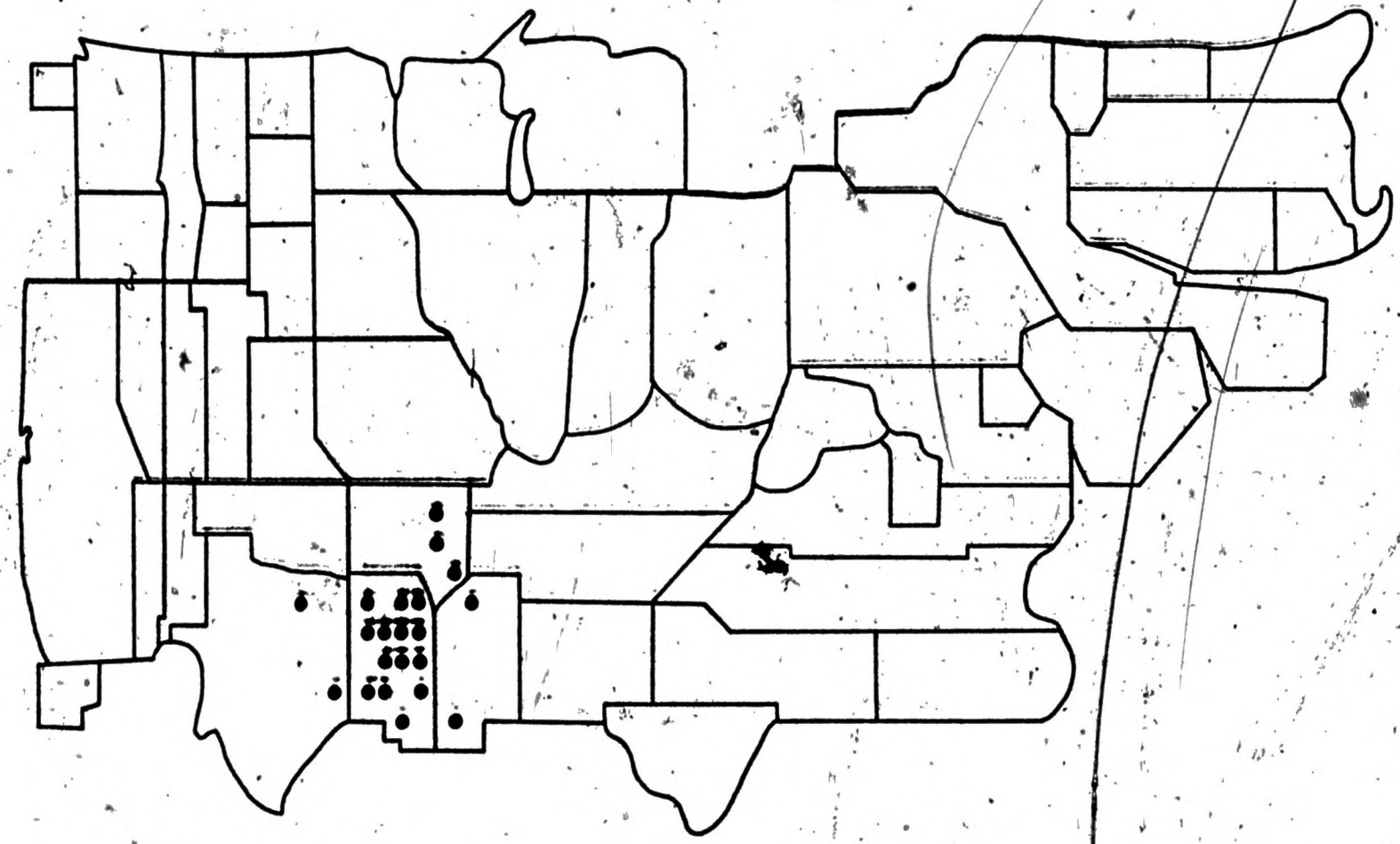
Figure 2

Response: Question: No. of informants:

Geograph: Map? Comments:

Variables:	Comm.					Age			Educ.					Race				Sex	
	1 U	2	3	4	5 R	Y	M	O	X	L	G	H	C	W	N	I	O	F	M
DARE %	6	5	20	38	31	9	24	66	1	3	24	41	31	92	7	.25	.3	50	50
Actual %																			
Frequency:																			
Conclusions:																			

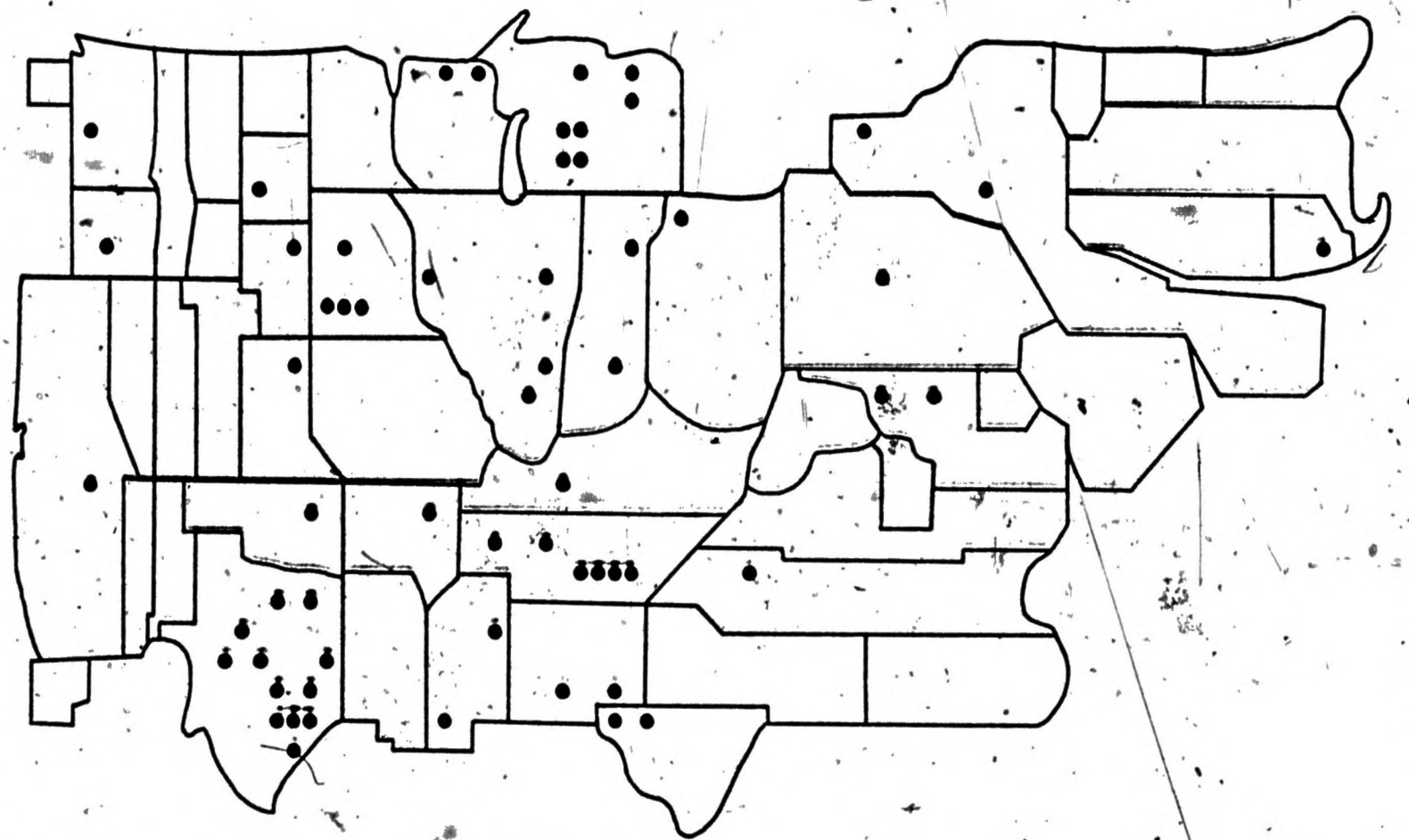
Figure 3



QUESTION 61, "BAYOU"

21 Informants

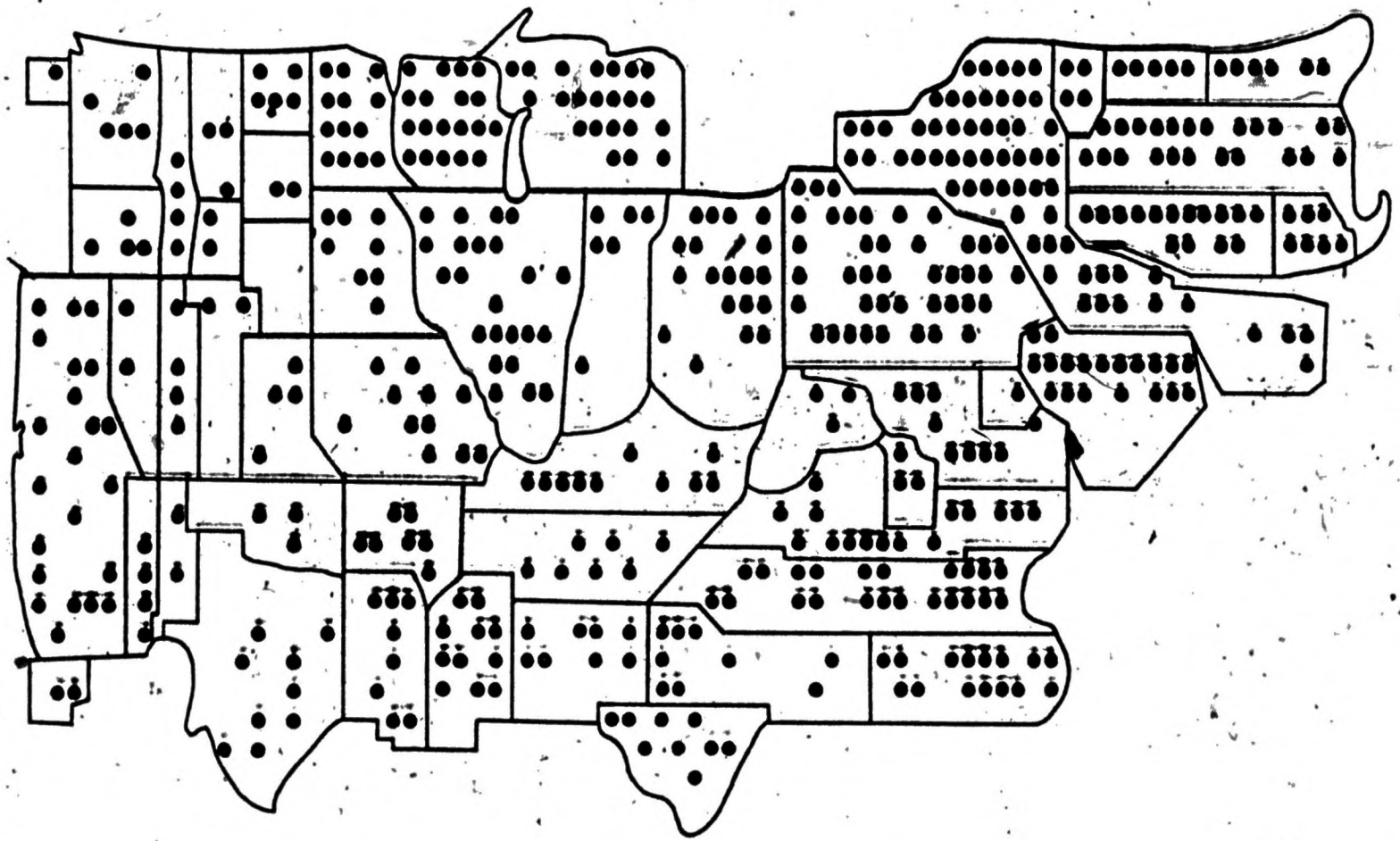
Figure 4



QUESTION 71, "BAYOU"

60 Informants

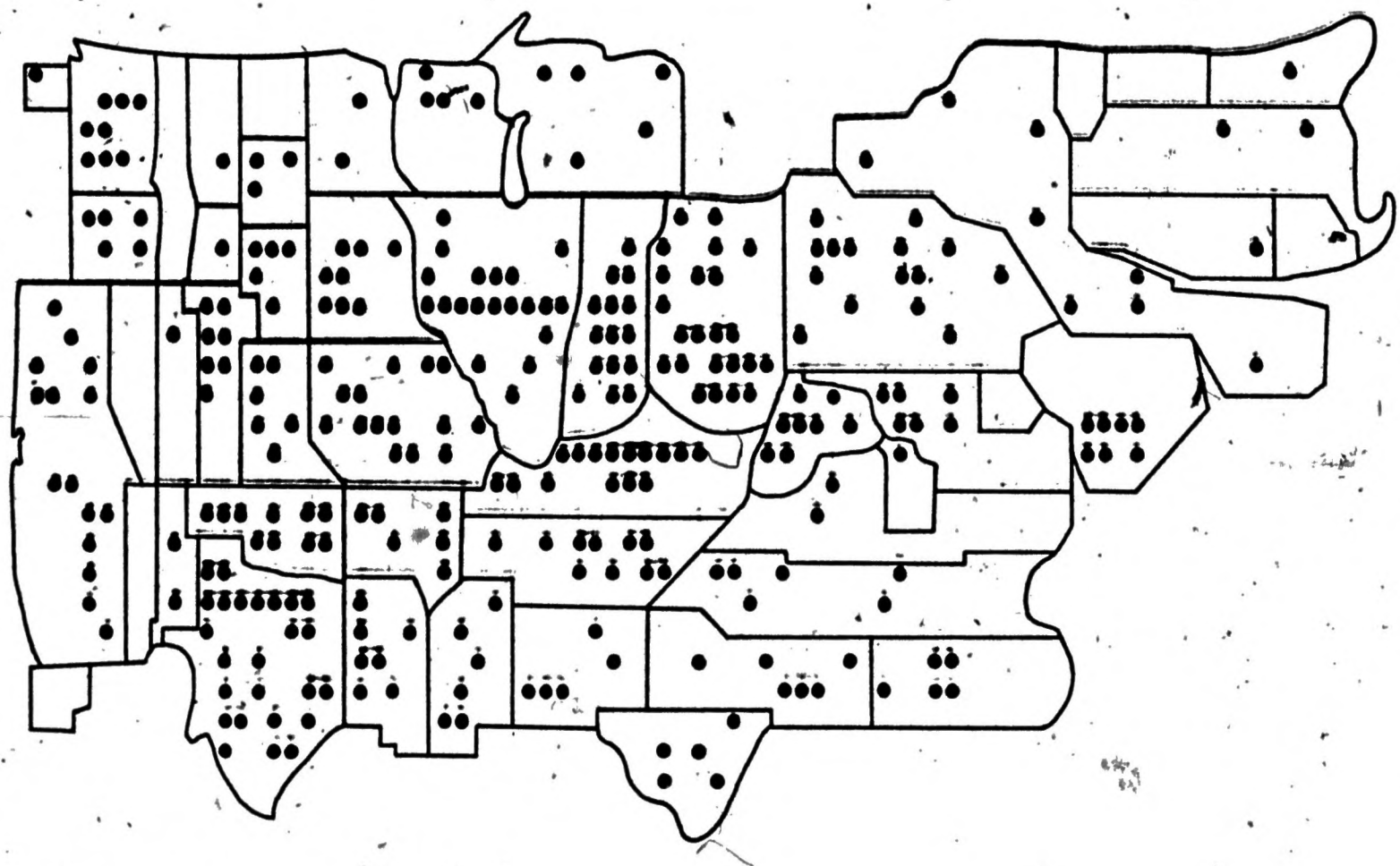
Figure 5



QUESTION 452, "ARM-FUL"

520 Informants

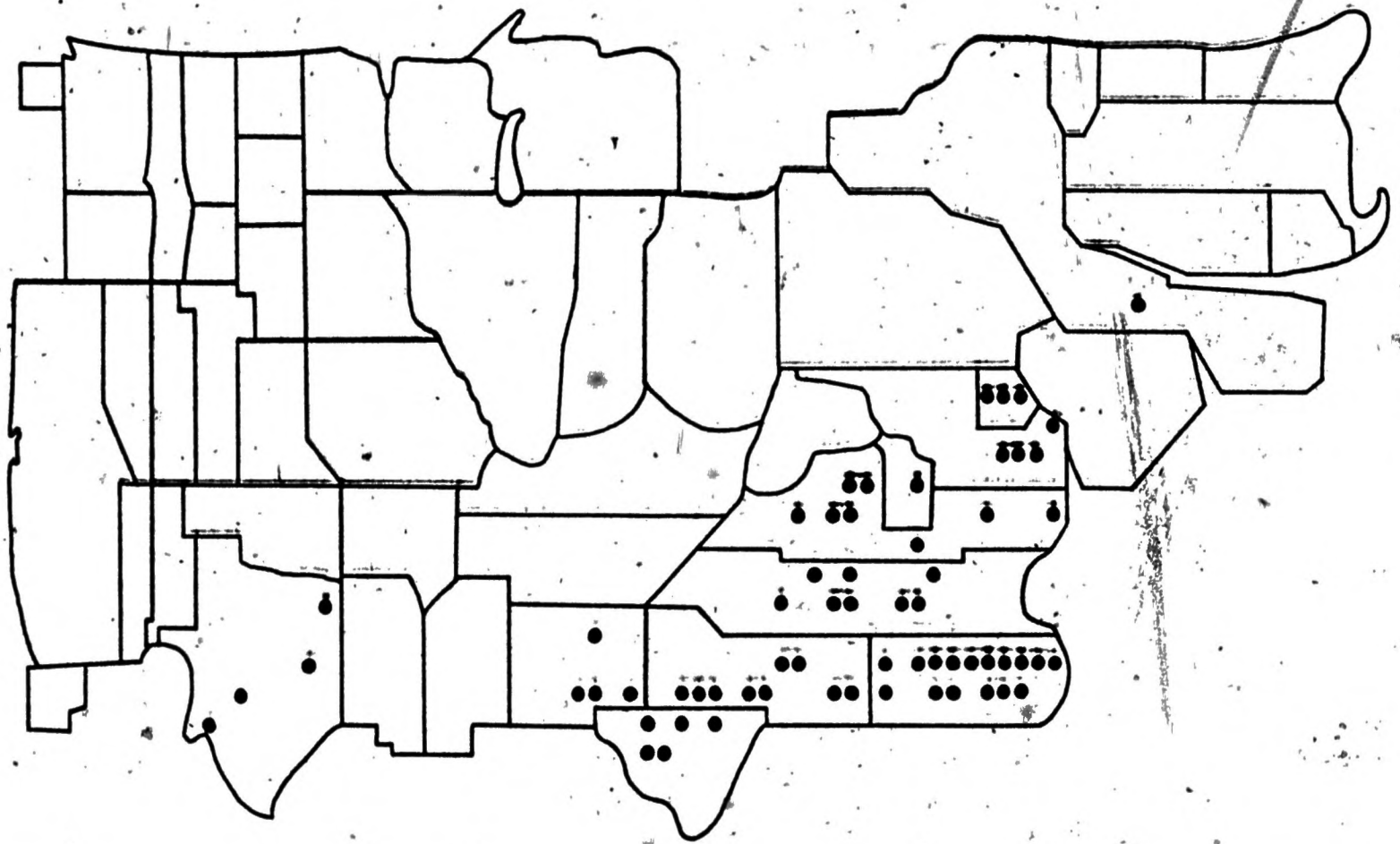
Figure 6



QUESTION 452, "ARMLOAD"

304 Informants

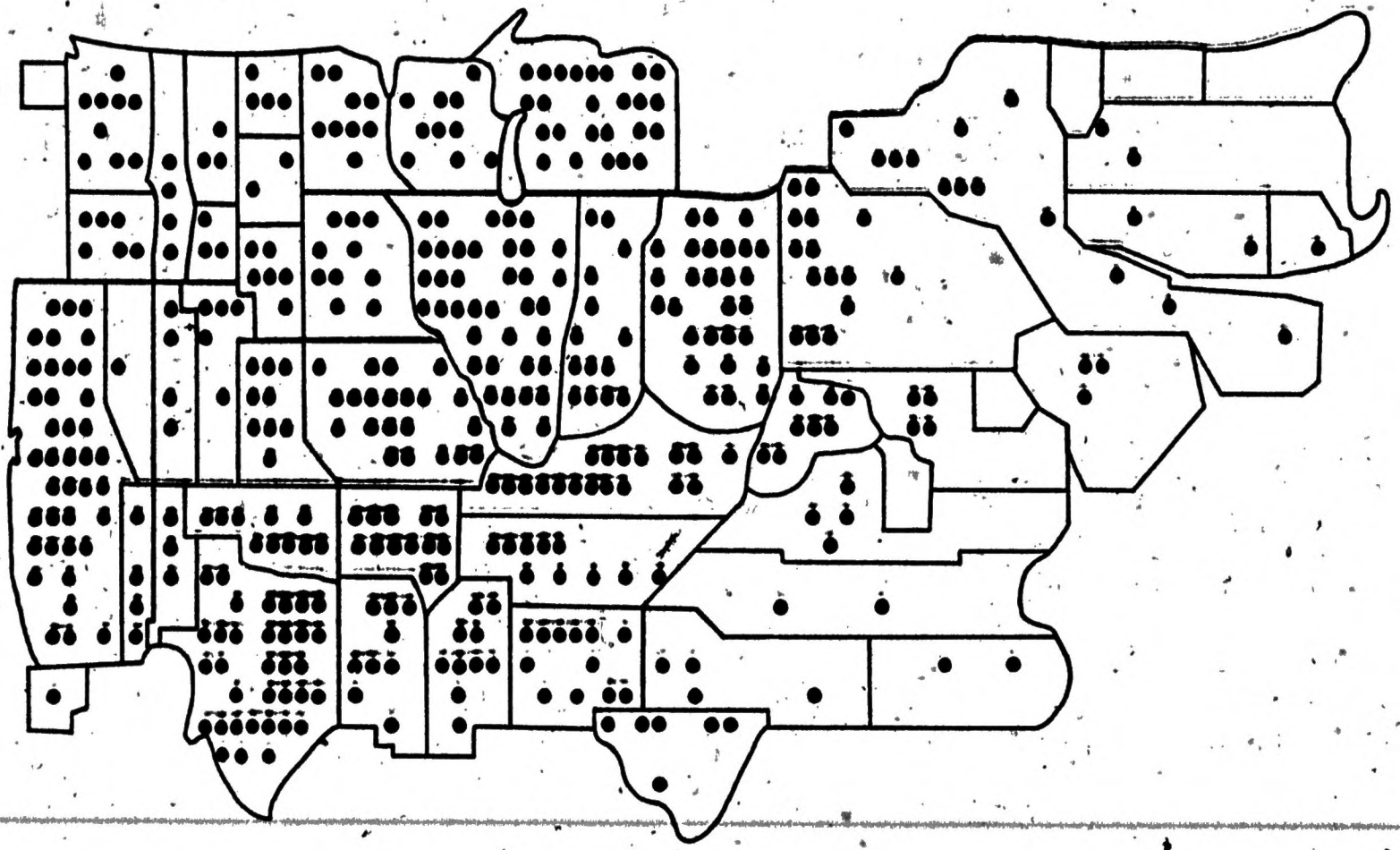
Figure 7



QUESTION 452, "TURN" 60 Informants

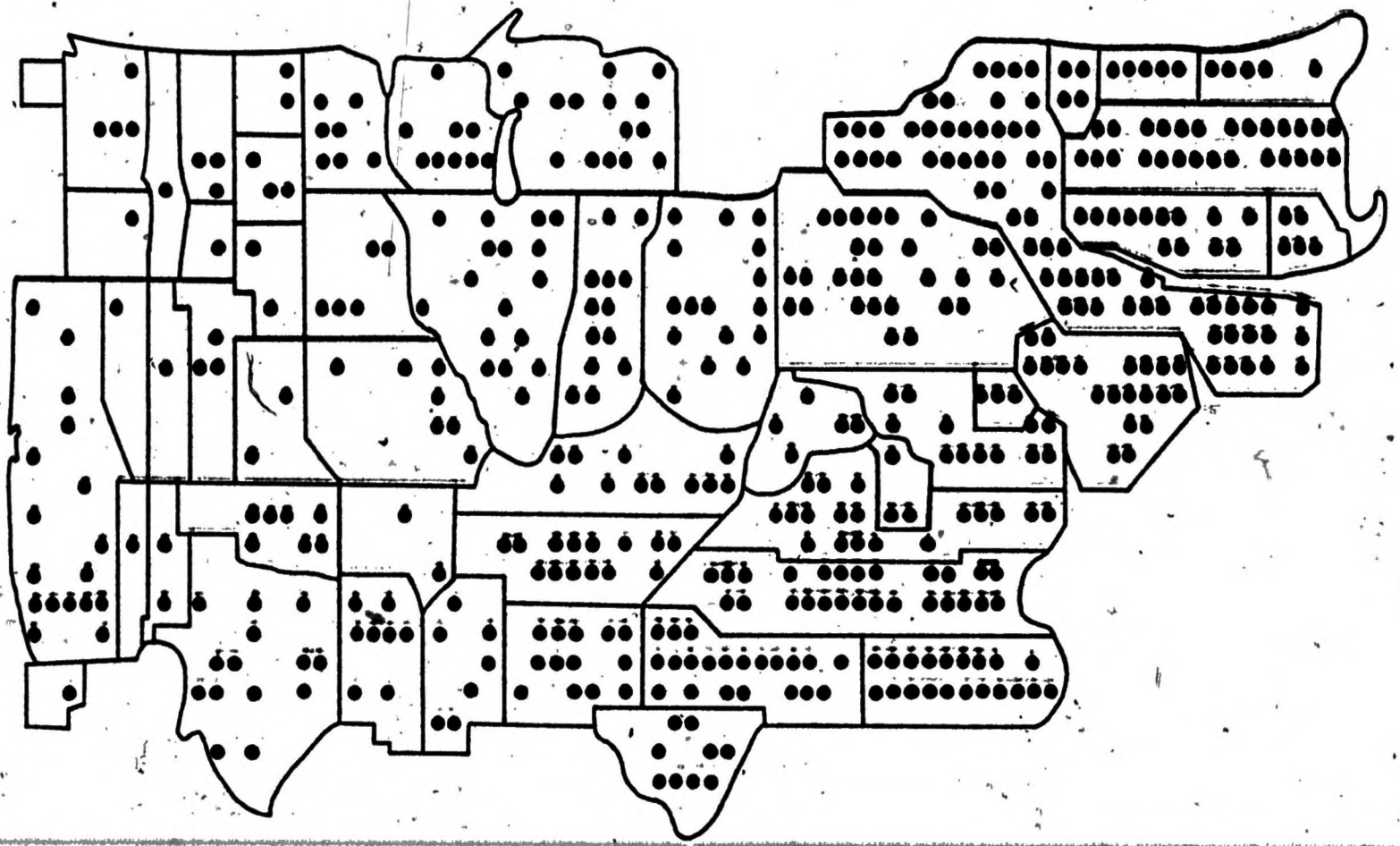
"TURN OF WOOD" 3 Informants

Figure 8



QUESTION 529, "BABY BUGGY"

Figure 9

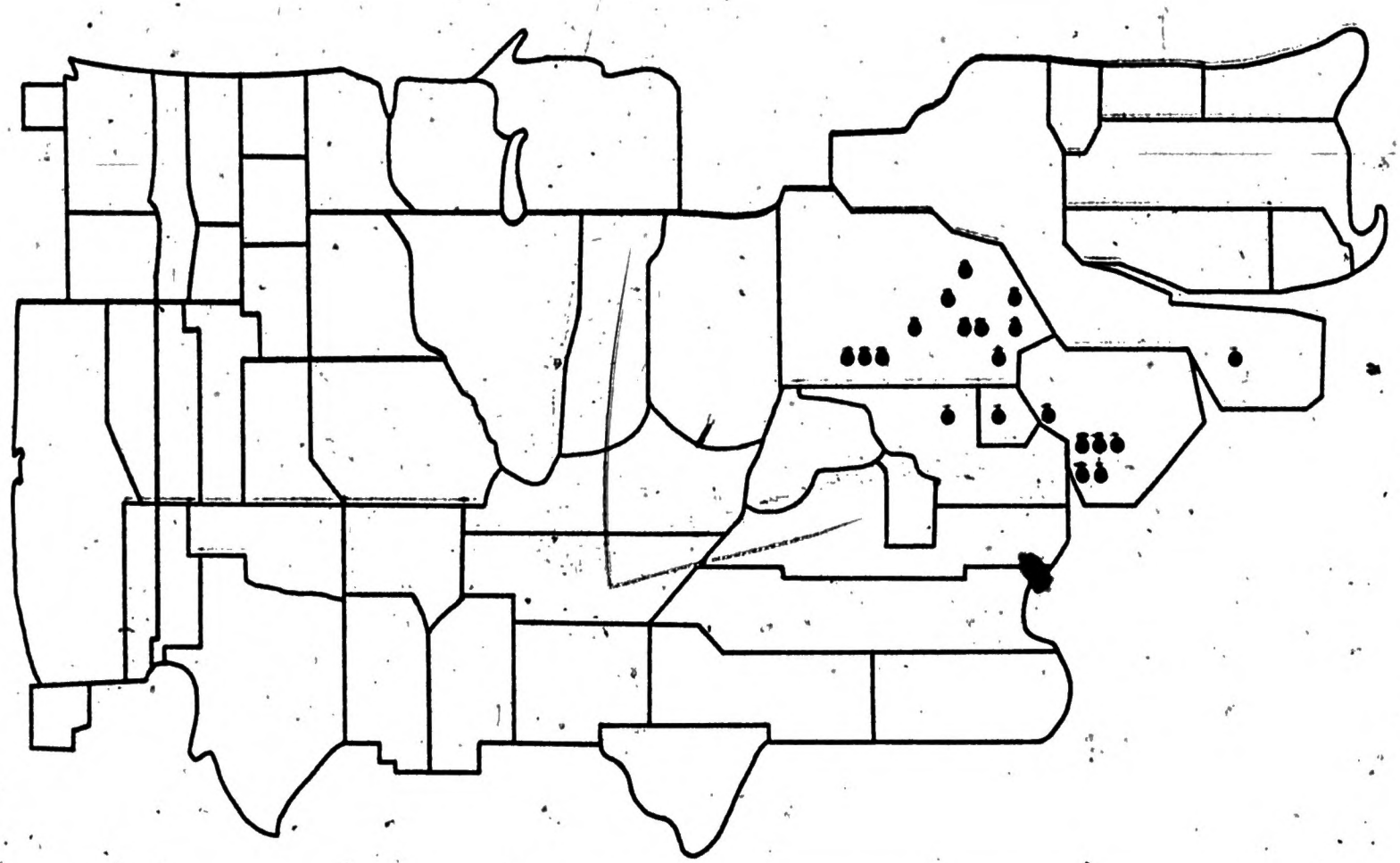


29

30

QUESTION 529, "BABY CARRIAGE"

Figure 10

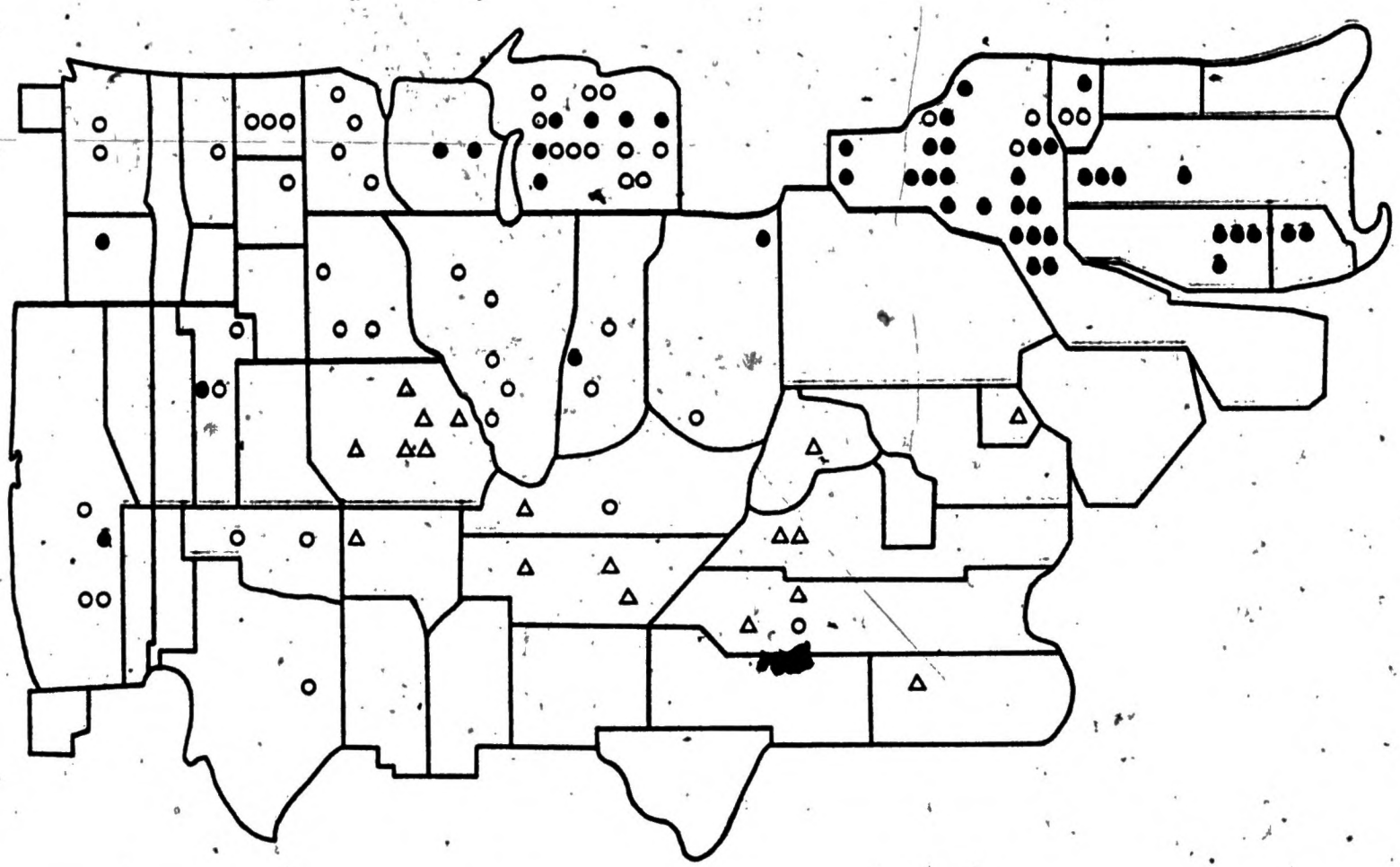


31

32

QUESTION 529, "BABY COACH"

Figure 11

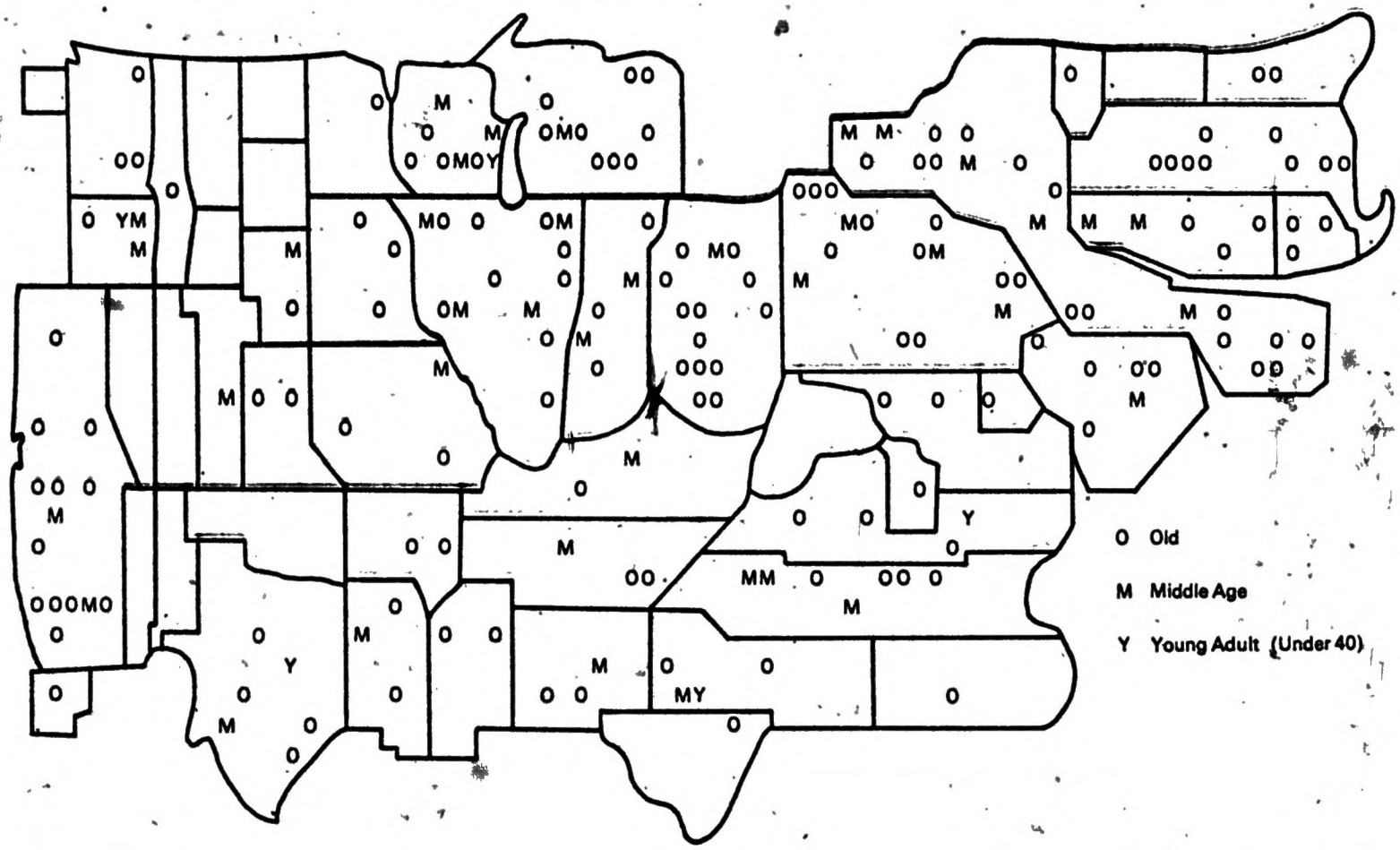


QUESTION 110, "GUTTERING" 19 Informants △

"EAVE TROUGH" 48 Informants ○

"EAVE TROTH"]
"EAVES TROTH"] 46 Informants ●

Figure 12



35

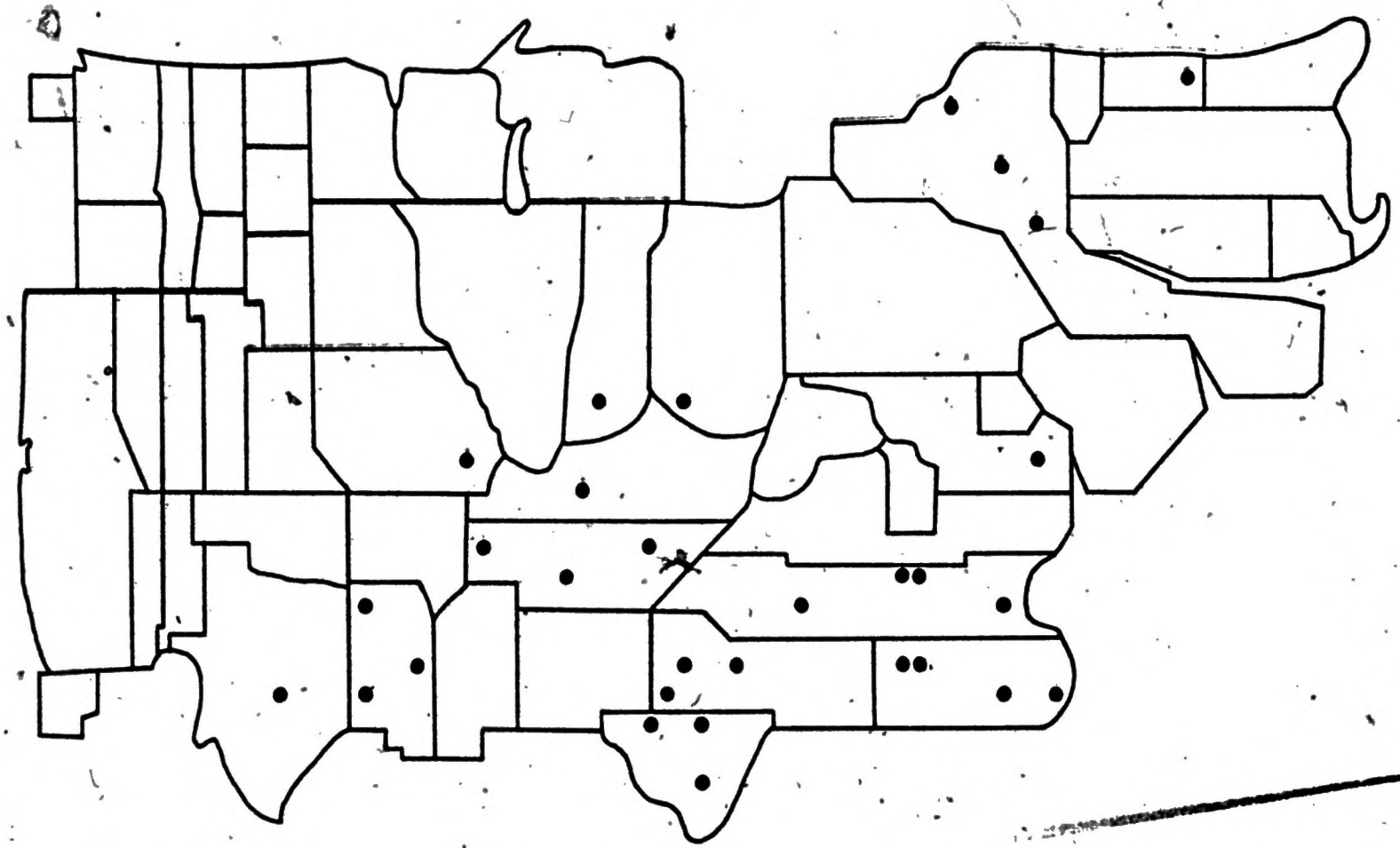
36

QUESTION 129, "ANTIMACASSARS"

190 Informants

AGE MAP

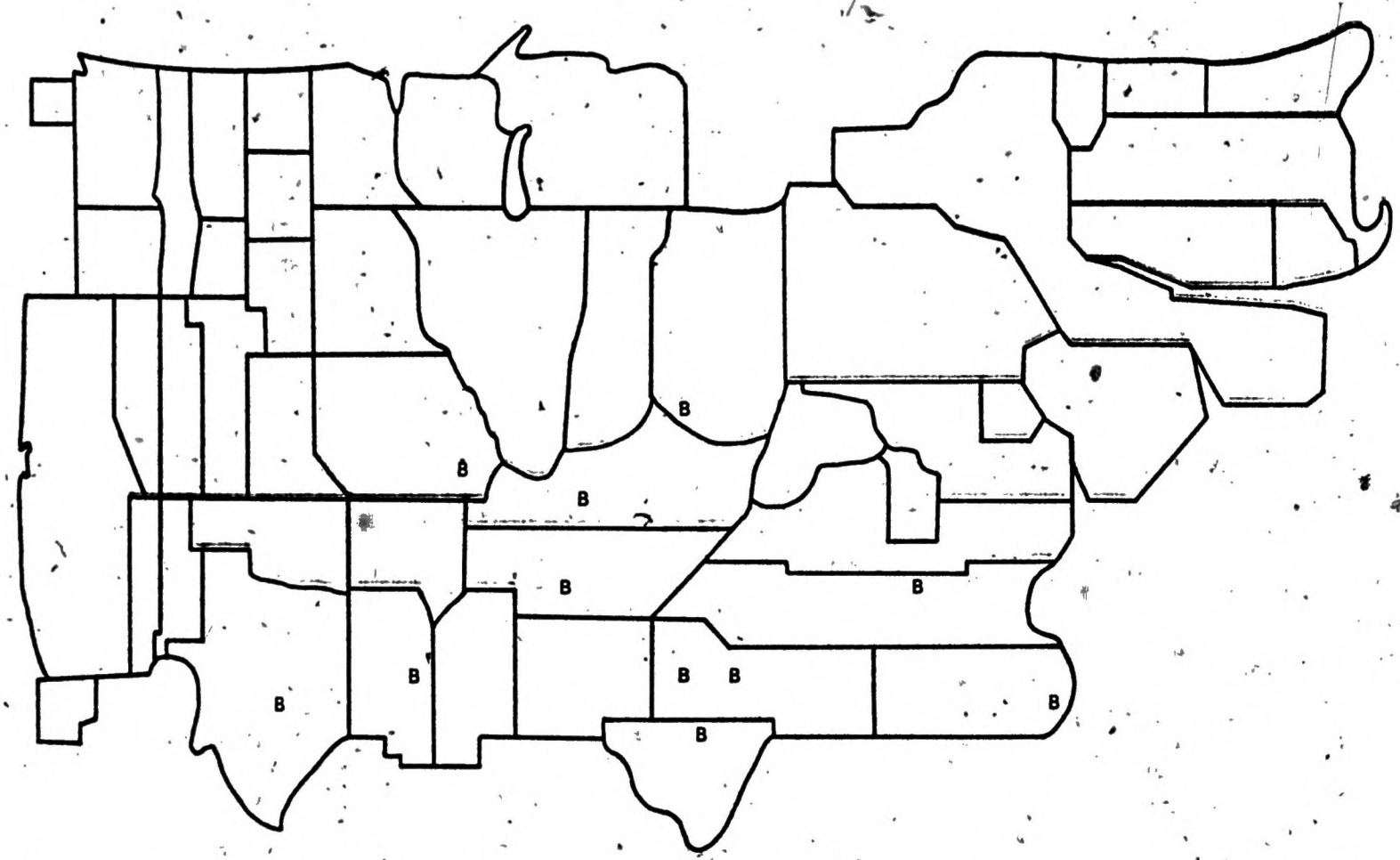
Figure 13



QUESTION 1, "BEFORE DAY"

30 Informants

Figure 14



39

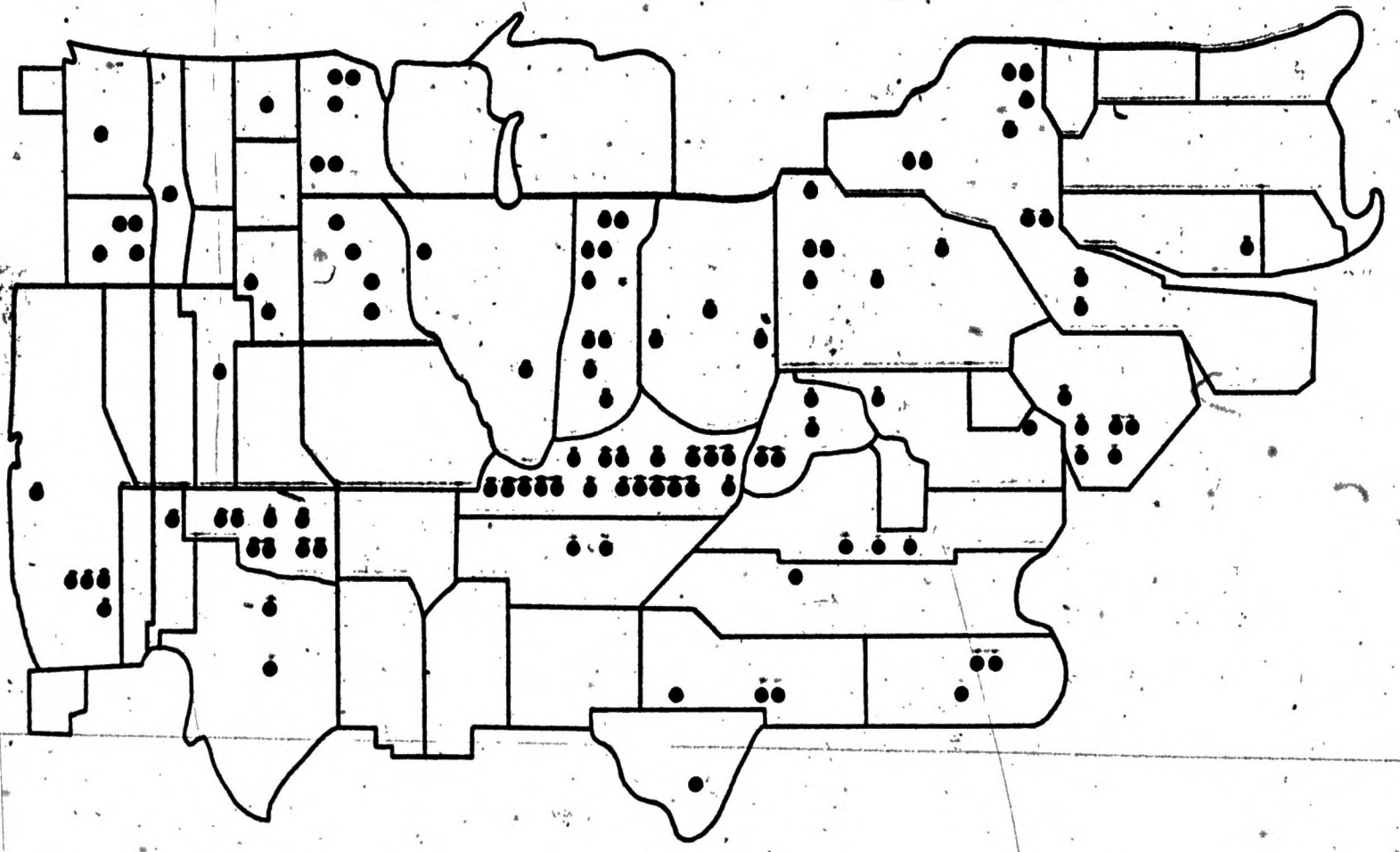
40

QUESTION 1, "BEFORE-DAY"

30 Informants

RACE MAP,
BLACK SPEAKERS ONLY

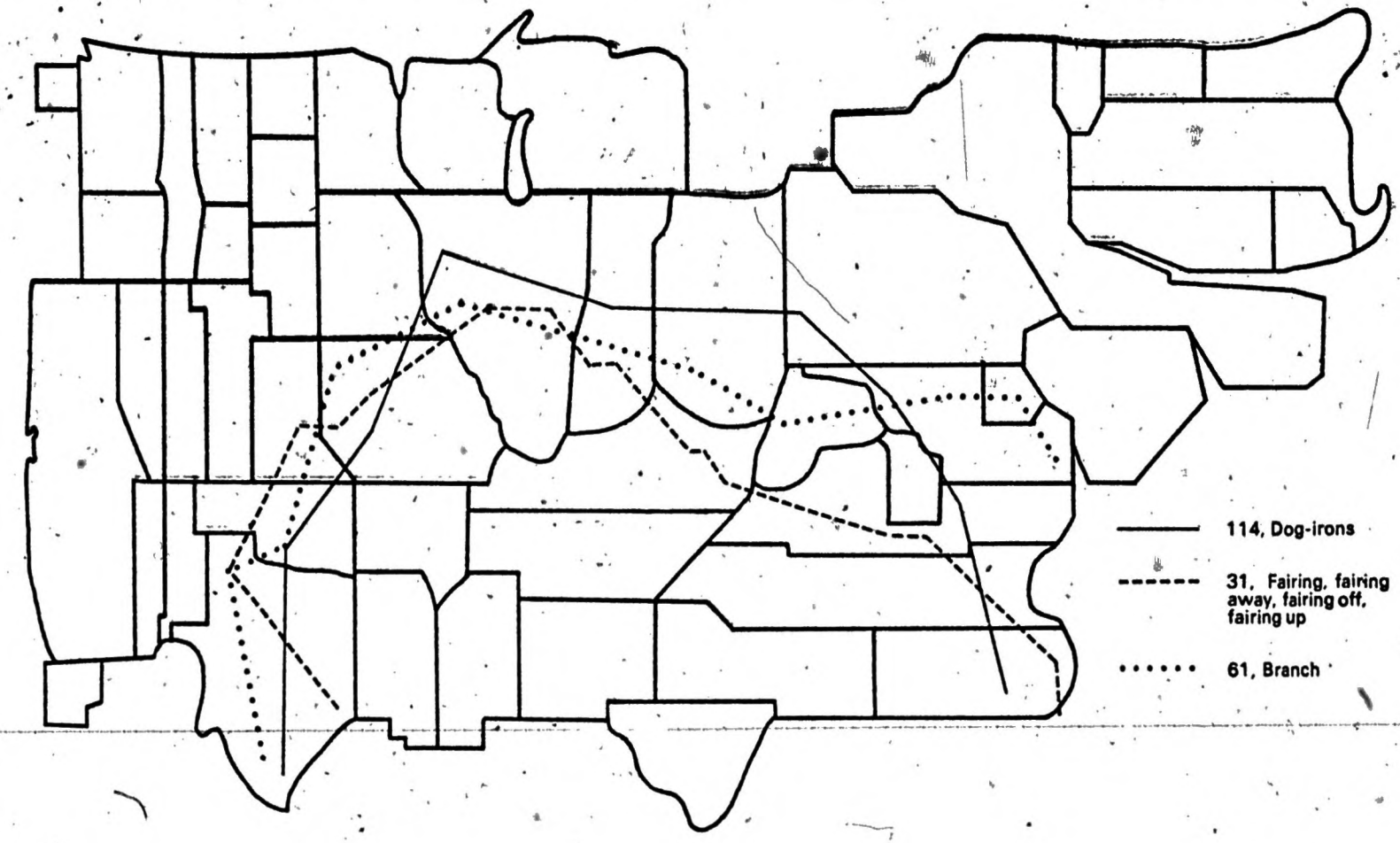
Figure 15



QUESTION 24, "ANYMORE"

110 Informants

Figure 16


114, Dog-irons
31, Fairing, fairing away, fairing off, fairing up
61, Branch