Research Memorandum No. 36
For Administrative Use

RELATIVE STRATIFICATION EFFICIENCY OF SELECTED AREA DELINEATIONS, NORTHERN GREAT PLAINS

by

Glen V. Vergeront
Agricultural Economist

UNITED STATES DEPARTMENT OF AGRICULTURE
Bureau of Agricultural Economics
Division of Land Economics

May 1946

Lincoln, Nebraska

RELATIVE STRATIFICATION EFFICIENCY OF SELECTED AREA DELINEATIONS, NORTHERN GREAT PLAINS

CONTENTS

SUMMARY

1. To the extent that an area delineation is based upon the natural and land use differences of the region, the more strata that are used the more efficient will be the delineation on an over-all basis. When counties are used as sample units, the number of strata should be limited only by the number of counties it is possible to include in the study as sample units.

2. Greater uniformity of counties within areas can be attained by giving careful consideration to the natural and land use differences of the Great Plains region in planning a stratification. States, which are arbitrary areas not closely related to natural and use differences, are poor bases for stratification.

3. Items do not vary with location within the region in the same manner. There are wide variations in the effectiveness of area stratifications for different items. When studying specific items, stratifications should be designed with special reference to the particular item being sampled. For studies of the general survey type in which many items are estimated, a general-purpose stratification such as the livelihood-areas delineation will probably be most adequate.

4. In the study here reported the livelihood areas gave the best over-all results for the 19 items tested. This delineation requires at least 39 counties as sampling units and therefore may not be useful for all studies. When less than 39 could be used, but 20 or more counties could be, the type-of-farming or the land-use areas would offer the highest relative stratification efficiency of the delineations tested on the over-all basis. The production areas and the tenure-problem areas offer better bases for further subdivision than do the States.

5. If a contiguous-area delineation is used as a stratification for county sampling in the Great Plains, delineations of areas that are compact will have higher stratification efficiency than will delineations that contain long, narrow areas. The low stratification efficiency of the river-basin delineation as compared with the livelihood areas, type-of-farming areas, and the land-use areas makes this point clear.

6. When an area delineation is used as a basis for the description and comparison of the different parts of the region, it is as important that the delineation contain a minimum of variation between counties within areas as when the delineation is used as the basis for stratification in sampling.

INTRODUCTION

Area delineation is often used both as a tool for analysis and as a basis for sampling in economic research. This is true whether the land area being studied is a Nation, a region, a State, or a county. A large mass of land such as the Great Plains has great variety in physical characteristics, in the use of land, and in the institutional arrangements for the control of land. Owing to this great variety it is generally neither convenient nor fruitful to describe and analyze the characteristics of the region in total. The division of the region into smaller areas offers a means of exerting some control over this wide variation so that description can be more incisive and so that divergent trends within the region can be more readily discovered. When the data for studying regional totals are not available, area delineation serves as a basis for making sampling more effective.

Area delineations are effective for sampling to the degree that they group the population units into areas that are more homogeneous than the population as a whole. If a delineation does not group the population units into groups that are substantially more similar than the population as a whole, the delineation is ineffective. This criterion of usefulness is equally applicable to area delineations to be used for analysis of the region as a whole. The more homogeneous the areas of a delineation are, the more marked will be the differences between areas and the greater will be the reliance that can be placed on the conclusions made for individual

areas of the delineation. Therefore, the relative efficiency of delineations for stratification should be as useful in evaluating sets of areas for descriptive analysis as for sampling. 1/

Many area delineations have been used in economic research in the Great Plains. Little attempt has been made to measure the effectiveness of these delineations in creating homogeneous areas, or in other words, their usefulness. The purpose of this study is to examine a limited number of delineations made in the past with particular reference to their usefulness for county sampling in land economic research and to make suggestions for improving future delineations.

AREA DELINEATIONS TESTED

Of the many area delineations used in economic research in the Great Plains region, eight were selected to test. Delineations based upon a single criterion such as size of farm, or percentage of land in crops, etc., were not selected, because it was desired to test delineations that were believed to be more suitable for general research needs. The eight sets of areas tested, together with a brief description and the source of the delineations, are as follows:

1. Production areas.- This five-area delineation divides the region into a corn area, a spring-wheat area, a winter-wheat area, a wheat-range area, and a mountain area. The delineation was used in the analysis carried out for the Production Goals work, 1942. It was prepared by committees of State and Federal workers, with headquarters at the regional office of the Bureau of Agricultural Economics at Lincoln, Nebraska.

2. Tenure-problem areas.- A five-area delineation with subdivision based upon the similarity of tenure problems within the areas. This delineation was prepared by the Bureau of Agricultural Economics at the regional office at Lincoln.

3. States.- The seven States of the Great Plains region are included in this delineation. States are often used as stratifications for administrative reasons.

4. Type-of-farming areas.- A 20-area delineation based upon the types of farming areas, prepared by the Bureau of Agricultural Economics at Washington, D. C.

1/ In this report the terms delineation and stratification as well as areas and strata are used interchangeably. Stratification and strata are merely the applicable statistical terms for the more familiar terms, delineation and areas. Counties are the population units used in this study.

5. Land-use areas.- A 20-area delineation self-weighting for numbers of farmers with subdivision based mainly upon the wheat-risk factor, percentage of land in crops, and average size of farm. The areas were delineated by the Bureau of Agricultural Economics at the regional office at Lincoln.

6. Land-resource areas.- A 29-area delineation based upon soil differences and conservation needs. The areas are used as a basis for planning and research by the Soil Conservation Service.

7. River basins.- A 31-area delineation based upon the river-basin map prepared by the Bureau of Agricultural Economics from material furnished by the U. S. Geological Survey.

8. Livelihood areas.- A 39-area delineation based upon the natural resources and the economic use of the land. It was published by the National Resources Planning Board, 1943.

All except one of the maps tested have a background of research use in the region. Two of the sets of areas used--livelihood areas and the land resource areas--have been proposed at various times as being suitable for all economic research carried out in this region. With the increased interest in the use of river basins as a basis for administration and planning, various persons have proposed that areas based on river basins are appropriate for research uses. Of the areas tested, type-of-farming areas are probably the best known for they have often been used in agricultural economic research.

County data from the census and other sources were used for testing the various delineations. In some cases the basic maps were slightly revised in order that each county would fall wholly into a single area. In a few cases an area containing only one or two counties was combined with an adjoining area. Every effort was made not to change the basic design of the delineations.

Seven counties each of which contained less than 100 farmers in the 1940 Census were combined with adjoining counties. In addition, four counties--Denver, Colo.; Wyandotte, Kans.; Silver Bow, Mont., and Douglas, Neb.--were excluded because of the extreme influence of large cities upon the counties. Owing to the combinations and exclusions, 451 population units were used for the tests, whereas there are 462 counties in the region.

METHOD USED

The criterion used to measure the efficiency of a delineation is the degree to which counties similar for a particular item have been grouped into areas. To test the delineations, county values for 19 items of

interest to land economists were selected as the basis of the tests. Sixteen of the items are variables which it might be desired to estimate by sampling, whereas three--rainfall, length of growing season, and wheat risk--are items in which land economists are primarily interested for use as control items in stratification.

Analysis of variance is the statistical device which was used to measure the variation among counties within the areas of a delineation and to measure the total variation among counties of the region. 2/ These measurements were made for each item and not for the 19 items as a group. For comparison, the ratio between the variation among all counties of the region (total variance) and the variation among counties within areas (unexplained variance) was calculated. The ratio was then multiplied by 100 to put the ratio in percentage form. The results are shown in table 1. A ratio may be thought of as meaning the number of counties which would have to be sampled at random from the region as a whole to be comparable in efficiency to 100 counties selected from the particular stratification being tested for the item. A high ratio indicates a gain in sampling accuracy by using the delineation for the particular item concerned while a low ratio indicates little or no gain by use of the delineation.

RELATIVE STRATIFICATION EFFICIENCIES OF SELECTED DELINEATIONS

In evaluating the results shown in table 1, it is necessary to consider both the delineations and the items tested. The average relative efficiencies for the 19 items tested are a convenient tool for comparing the over-all stratification efficiencies of the eight delineations tested. It is not meant to imply that the weighting inherent in this selection of items is that which is desirable for the needs of all research in the region. On the other hand, it is not likely that the addition of other items would substantially alter the relative positions of the delineations as based upon the results for the 19 items.

The livelihood-areas delineation is the best delineation based upon results for the 19 items tested. This delineation gives the greatest uniformity of counties within areas for all items tested except wheat risk and percentage of cropland in wheat. One limiting factor in the use of the livelihood-areas delineation is that at least 39 counties would have to be sampled in order to obtain the full benefit of the stratification. It is important to notice that for the livelihood-areas stratification there are wide variations in the effectiveness of stratification for the various items tested. This point is discussed further on page 7.

2/ Variance is another name for squared standard deviation. Analysis of variance is a method developed by R. A. Fisher for comparing a number of different series of data in the same operation. For description of method see Snedecor, G. W., Statistical Methods; Pearson and Bennett, Statistical Methods, Croxton and Cowden, Applied General Statistics; etc.

The type-of-farming delineation and the land-use delineation are of about comparable efficiency on the basis of the items tested. Either of these stratifications are superior to all except the livelihood areas. The type-of-farming delineation is more efficient than the land-use delineation for such items as length of growing season, percentage of tenancy, amount of publicly-owned land per county, and the percentage of irrigation farmers. The land-use delineation is better for such items as wheat risk, proportion of total cropland in wheat, migration 1930-40, and level of living. The land-use delineation is particularly poor for publicly-owned land because in order to make the stratification self-weighting for farmers, a ranching area of Wyoming which contains a high percentage of publicly owned land had to be included with the Sand Hills ranching area of Nebraska where practically all of the land is privately owned. This illustrates a situation in which there is a wide variation among counties within an area. Such areas make a delineation poor for sampling and for description.

The land-resources delineation is almost as efficient as the type-of-farming areas or land-use delineation. In practice it cannot be considered as useful since both the type-of-farming and the land-use delineations offer more possibilities for further subdivision. The land-resource delineation could be greatly improved if two or three large, narrow areas were split into 4 or 6 smaller areas. Such large, narrow areas tend to include counties of wide diversity for many of the items tested, particularly for the items regarding climate and for items which are closely related to climate. This delineation is more effective than are the type-of-farming areas or the land-use areas for the four tenure items included in the tests.

The river-basins delineation is inferior to the delineations discussed above. This holds true because areas delineated on the basis of river basins in the Great Plains region tend to be long, narrow areas which include counties that have widely different item values. For instance, the James River Basin as delineated includes 22 counties stretching from the northern border of Nebraska into north central North Dakota in a long, narrow band never more than 3 counties wide. The Platte River Basin stretches practically the entire length of Nebraska in a narrow band. The Republican River Basin includes dryland wheat counties of eastern Colorado as well as Corn Belt counties of north-central Kansas. This delineation is relatively so poor that its suitability as a basis for research or planning in any field except possibly for problems connected directly with the supply of water may be doubted.

If a sample of 20 or more counties is contemplated for use in the Great Plains area, three delineations--production areas, tenure problem areas, and States--do not have a high enough intensity of subdivision to compare favorably with the delineations discussed above. Both the production areas and the tenure-problem areas are superior to States.

As States are arbitrary political units, they do not give adequate consideration to the natural features and the use differences that cause variation among counties and are thus not efficient for stratification purposes. If States are used as a basis of delineation, relatively larger samples should be taken in order to compensate for the inefficiency of the stratification. It is possible that in sampling for certain items that are directly affected by State laws and institutions, special consideration should be given to States in the sample design.

The production-area delineation is decidedly better than the tenure-areas delineation for two items--length of growing season and the percentage corn of total cropland. The tenure areas which emphasized an east-west division of the Great Plains are superior for these items: wheat risk, percentage of land in crops, percentage irrigation farmers are of total farmers, and value of land per acre. Both the production areas and the tenure areas offer better basic designs for further subdivision than do the States.

DIFFERENCES IN DISTRIBUTION OF ITEMS WITHIN REGION

The various items tested are not distributed over the region in the same way (table 1). The efficiency of a stratification based on areas is directly related to the geographic distribution of the items to be sampled. The distribution of items should be a first consideration in delineating areas. These data are not adequate for indicating how all items that might be of interest are distributed, but they do indicate some of the patterns that will be met.

For certain items the size of a county datum is closely related to the location of county within the region. Location and factors related to location are important in determining the size of county values. These items may be referred to as having a high correlation with location. Making contiguous area delineations is an effective method for grouping counties of similar values in areas for such items. The size of county datum is not closely associated with location within the region, for other items and contiguous-area delineations are not highly efficient as a basis of stratification.

To show some of the differences in the distribution of variation between counties, a number of comparisons of the ratios in table 1 are pertinent. Items that are highly correlated with location appear in this table as having high stratification efficiencies. Items that are not highly correlated with location have low stratification efficiencies. By making comparisons of the ratios for a single item between the various delineations, further differences are indicated. Four patterns of the distribution of items are discussed below.

Table 1.- Relative efficiency of various area delineations on the basis of selected items, Northern Great Plains

(The efficiency of samples drawn at random without stratification represents 100)

Selected items	Area delineations 1/							
	Production areas	Tenure problem areas	States	Type of farming areas	Land use areas	Land resource areas	River basins	Livelihood areas
Rainfall 2/	198	203	195	576	555	561	550	762
Length of growing season 2/	265	194	291	445	302	328	398	489
Wheat risk 3/	150	221	115	246	478	291	266	433
Percent of tenancy 4/	212	198	200	295	224	333	282	413
Percent of owner-operated land 4/	131	136	141	141	145	168	128	183
Percent of publicly-owned land 4/	337	332	219	319	272	470	305	480
Mortgage-debt ratio, part owners 4/	153	136	169	196	193	226	200	256
Size of farms 4/	148	154	136	175	179	168	157	199
Value of land per acre 4/	172	199	134	314	329	363	294	410
Percent of land in crops 5/	305	361	193	613	608	524	433	721
Percent wheat of total cropland 4/	135	133	179	281	341	277	256	333
Percent corn of total cropland 4/	270	143	166	307	407	355	424	451
Percent irrigation farmers of total farmers 4/	307	334	230	415	266	210	225	516
Changes in cropland, 1930-40 4/	123	119	119	146	149	128	152	177
Value of products per farm worker 4/	116	119	135	122	139	143	149	156
Farm population per square mile 4/	235	234	143	395	379	350	290	473
Migration, 1930-40 6/	152	164	114	149	194	190	163	211
Level of living, rural farm 6/	107	119	104	149	221	136	144	366
Average value of farm homes 4/	133	137	124	163	175	178	167	189
Average relative efficiency	192	191	164	287	292	284	262	380
Number of areas	5	5	7	20	20	29	31	39

1/ For source of delineation, see body of report.

Source of items:

2/ USDA Yearbook, 1941 - Climate and Man.

3/ Adjusting Wheat Acreage in Northern Great Plains to Wartime Demand, Ralph E. Ward, BAE, 1944.

4/ U. S. Census, 1940.

5/ From records of the Agricultural Adjustment Administration, compiled by BAE, Lincoln, Nebr.

6/ Rural Level of Living Indexes for Counties of the United States, 1940, Margaret J. Hagood, BAE, 1943.

First, there is a type of geographical distribution of county values in which a few major strata are effective in creating homogeneous groups of counties, and substantial gains in stratification efficiency are achieved by further subdivision within the large strata. Percentage of total land area in crops shown in the table is illustrative of this type of distribution. Percentage of land area in crops is an item for which a county value is largely determined by the location of the county within the region. The patterning of county values changes gradually from east to west and from north to south. Thus, the delineation of contiguous groups of counties makes areas within which the counties are relatively uniform for percentage of land in crops. For such an item the total variation between counties may be great but this variation is highly related to location within the region. Area stratification is highly effective for such an item. Unfortunately all items are not of this type so that contiguous-area stratification is not equally effective for most other items.

A second type of geographical distribution of county values is represented by that in which a few major areas give substantial stratification gains, but further subdivision within these major areas adds relatively small gains. This type of distribution is represented by the percentage of publicly-owned land and the percentage that irrigation farmers are of total farmers. Such an item is so concentrated that most of the counties with high item values lie in one part of the region while most of the counties with low values are in another part. Because the counties in each major division tend to be very similar over the whole division or because the variation between counties does not tend to be highly related to location within the division, further subdivision does not greatly increase homogeneity. If only a single item of this type is being sampled, it is likely that a sample design giving special attention to the major areas containing high variation between counties would be superior to general-purpose delineations. If such a special stratification were used it is probable that some basis of stratification other than by contiguous areas would be more efficient.

A third type of distribution of variation among counties is represented in table 1 by the level-of-living index. For such an item, stratification efficiency is low when only a few major geographical strata are used, but it is relatively high when a large number of areas are used. This happens because the variation among counties is almost as great within major areas as it is within the region as a whole. Nevertheless, the items are correlated with location so that further area subdivision tends to make fairly homogeneous groups of counties. Thus, for the levels-of-living index, 100 counties taken from the production-area delineation would be equal to 107 taken at random, a negligible stratification gain; 100 counties taken from the livelihood stratification would be equal to 366 taken at random, making the use of the livelihood stratification for this item very valuable.

In the fourth type of distribution of variation among counties, only relatively small gains from area stratification are achieved within the limits tested in this report. These items have a relatively low correlation with location. Even though county data differ widely over the region as a whole, contiguous-area groupings of counties are not homogeneous for these items. The delineated areas tend to have almost as much variation between counties as is present in the region as a whole. Value of products per farm worker in 1940 is typical of such an item.

The four patterns of variation among county data discussed above make clear that general-purpose stratifications cannot be expected to be equally efficient for a large number of items. For projects in which a single item is being studied, it will generally prove most efficient to use a stratification more closely related to the particular item than a general-purpose stratification can be; however, for the studies in which many items with widely different stratification efficiencies are sampled, a general-purpose stratification will probably be the best choice.

CONSIDERATIONS IN SELECTION OF A DELINEATION

Certain general conclusions may be drawn regarding the use of area delineation as a basis for stratification. On the basis of the items tested it is apparent that the more strata used the greater is the efficiency of the delineation. As the number of areas per delineation are increased, higher relative stratification efficiencies are obtained. For studies in which counties are to be used as sampling units, the number of strata should be limited only by the number of counties it is possible to study. When counties are used as sampling units it is particularly important to make the greatest possible gain from the use of complete stratification for the sampling error will tend to be high, because of the limited number of sampling units.

This conclusion has a major qualification. Merely increasing the number of areas of a delineation without regard to the physical and land use characteristics of the region will not necessarily be effective. Thus, the river-basin delineation which has more areas than the type-of-farming delineation is decidedly inferior in stratification efficiency. This is because river basins do not account so adequately for the geographical distribution of the characteristics tested. The more closely a delineation can be related to the items to be sampled or described, the more effective the delineation will be. The closest attention must be given to the geographical distribution of items that are to be sampled when contiguous area delineations are to be used as the basis for stratification. Only by so doing can the increased stratification efficiency due to the addition of strata be gained.

Contiguous-area delineations are not effective for all items in the Great Plains region. Certain items, such as the value of product per farm worker and change of acres in cropland, 1930-40, have a relatively low relationship to location within the region, and this makes contiguous-area stratification ineffective. For other items, high altitude and

intensive irrigation in the western mountainous third of the region cause great heterogeneity among counties. For some purposes it would be well not to use contiguous areas as a basis for delineation in the western third of the region. For the region as a whole, however, it seems likely that real advantages in homogeneity of counties within areas will be achieved by keeping the delineated areas contiguous and as compact as possible. Attention should be given to other factors in delineation, but compactness of areas should be a first consideration.

Wheat risk appears from the ratios in table 1 to be a highly important item to use in stratifying for studies measuring change in this region. The evidence that bears upon this is not complete, and further research could profitably be applied in investigating this point. The stratification used in the 20-area land-use delineation was largely based upon the wheat-risk factor for counties of the region. (Note the high ratio under land-use delineation for the wheat-risk item). The land-use delineation showed a clear-cut superiority over the type-of-farm delineation in creating homogeneous areas for migration, 1930 to 1940, and a smaller superiority for the change in percentage of cropland, 1930 to 1940. This is a hint that risk is an item of value in identifying those areas where the more dynamic changes take place in this region. As the measurement of change is often a major reason for carrying out research studies, areas of great change should be given special consideration in any stratification used.

Stress should be placed upon the fact that the variation between county data is related to location within the region in widely different ways for the various items. This is again stated to lend emphasis to the limitations of general-purpose stratification based upon contiguous-area delineation as tested in this study. When it is proposed to estimate the regional value for a single item, higher stratification efficiency will probably be given by a stratification designed especially for such an item. In many cases it is possible that such special designs will not be contiguous-area delineations. Nevertheless, the fact that all items were correlated with location to some extent should make the general-purpose area delineations highly useful for general surveys in which many items are being estimated.

NORTHERN GREAT PLAINS REGION

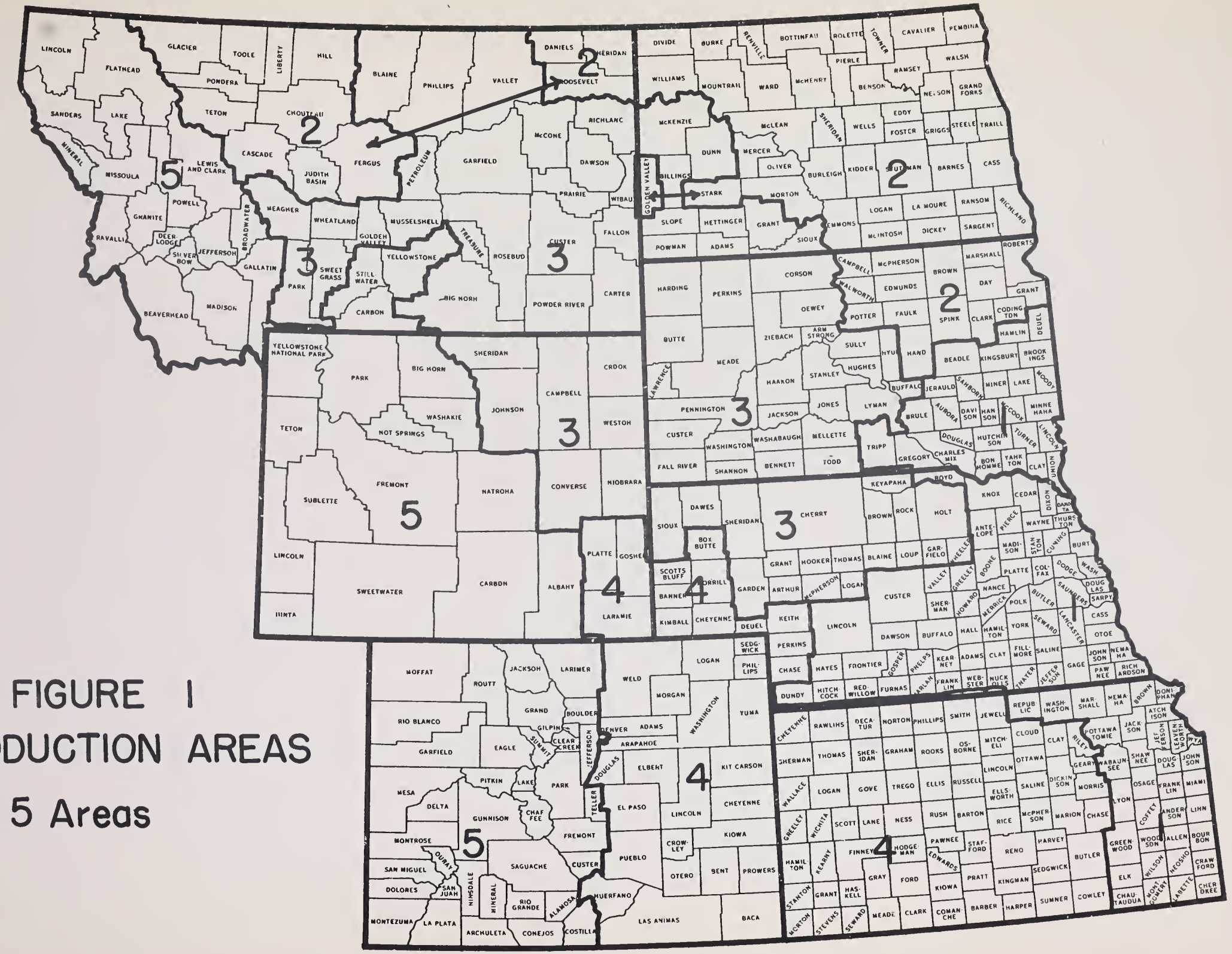

FIGURE 1

PRODUCTION AREAS

5 Areas

U.S. DEPARTMENT OF AGRICULTURE

BUREAU OF AGRICULTURAL ECONOMICS

NORTHERN GREAT PLAINS REGION

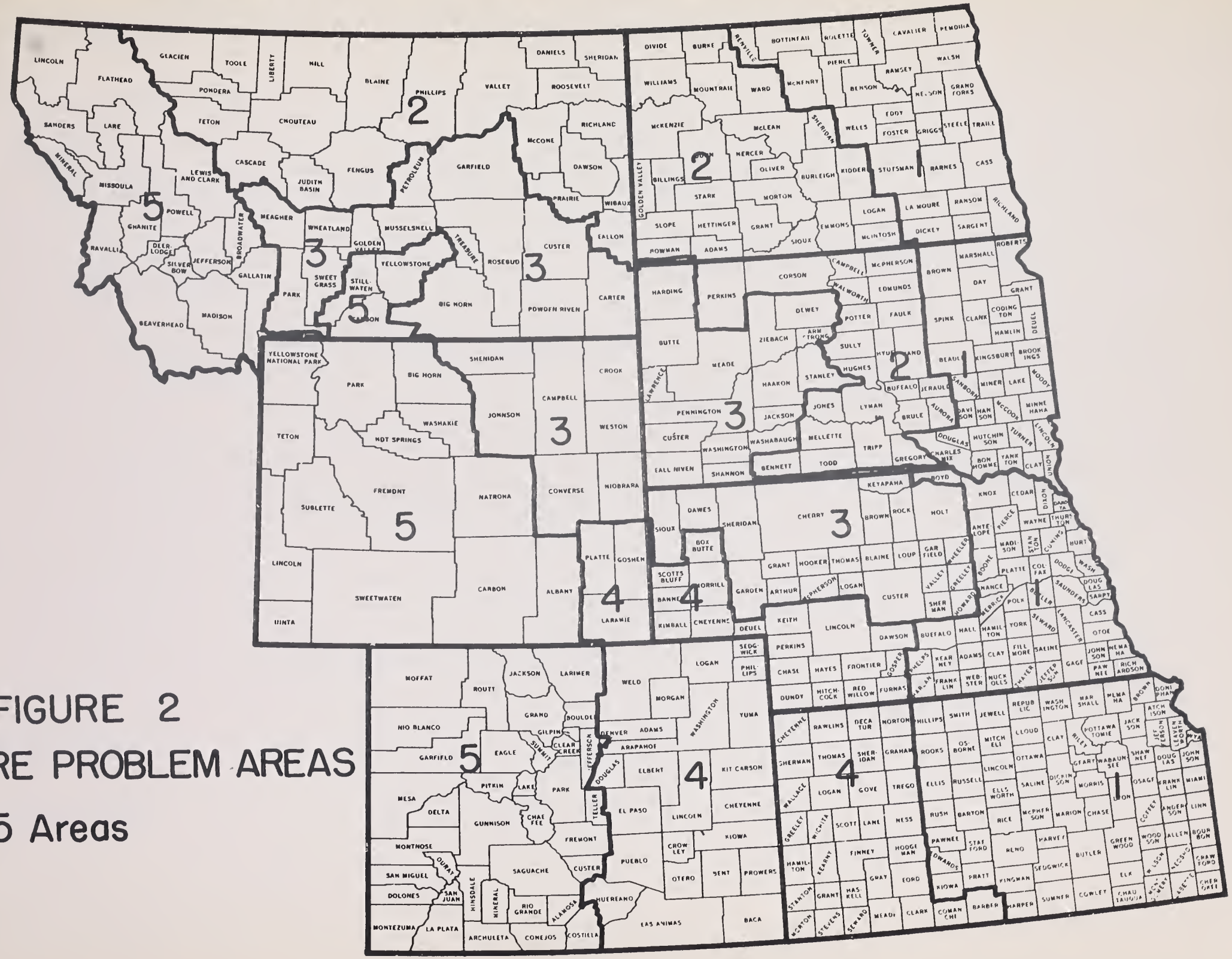

FIGURE 2
TENURE PROBLEM AREAS
5 Areas

U.S. DEPARTMENT OF AGRICULTURE

BUREAU OF AGRICULTURAL ECONOMICS

NORTHERN GREAT PLAINS REGION

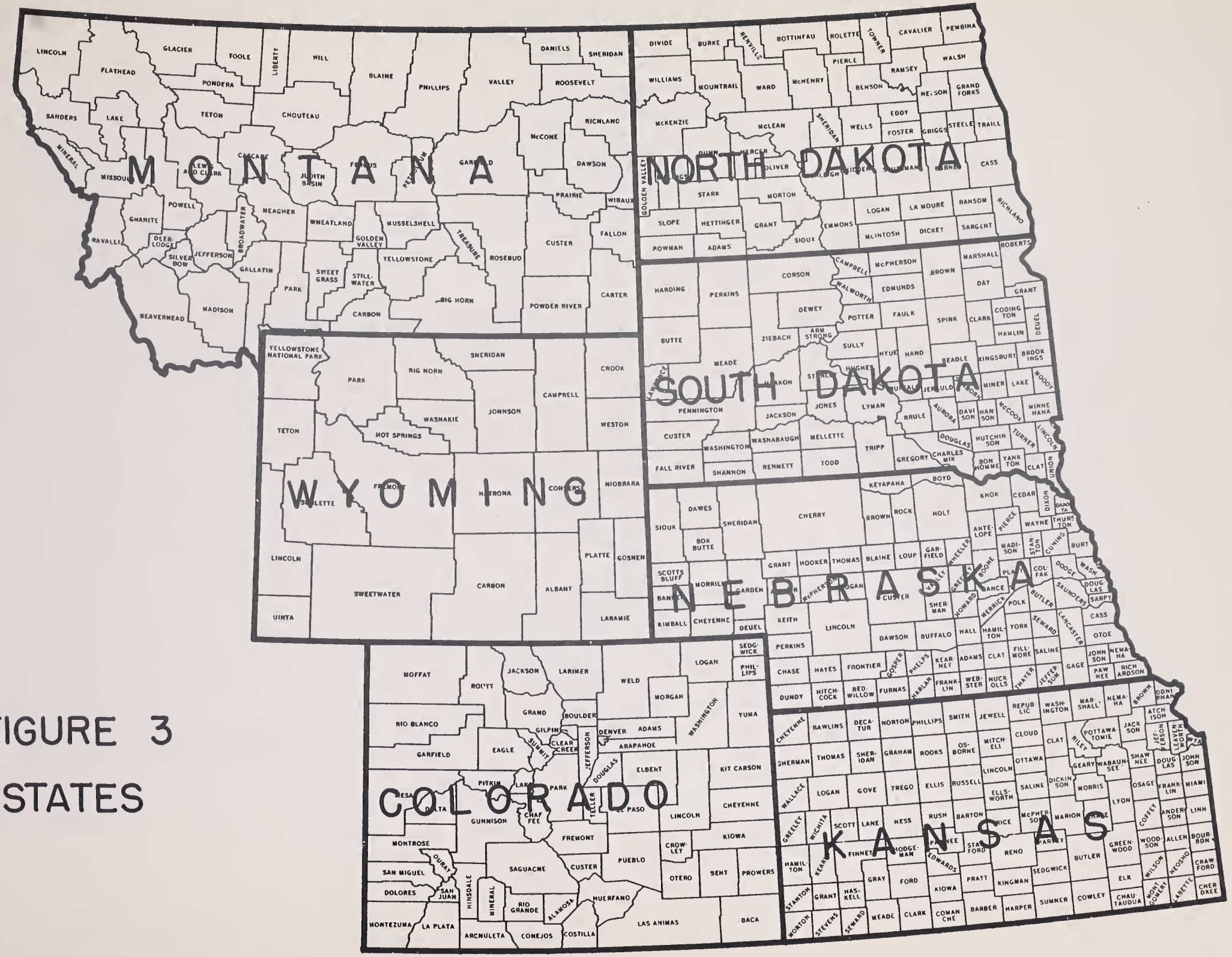

FIGURE 3

STATES

U.S. DEPARTMENT OF AGRICULTURE

BUREAU OF AGRICULTURAL ECONOMICS

NORTHERN GREAT PLAINS REGION

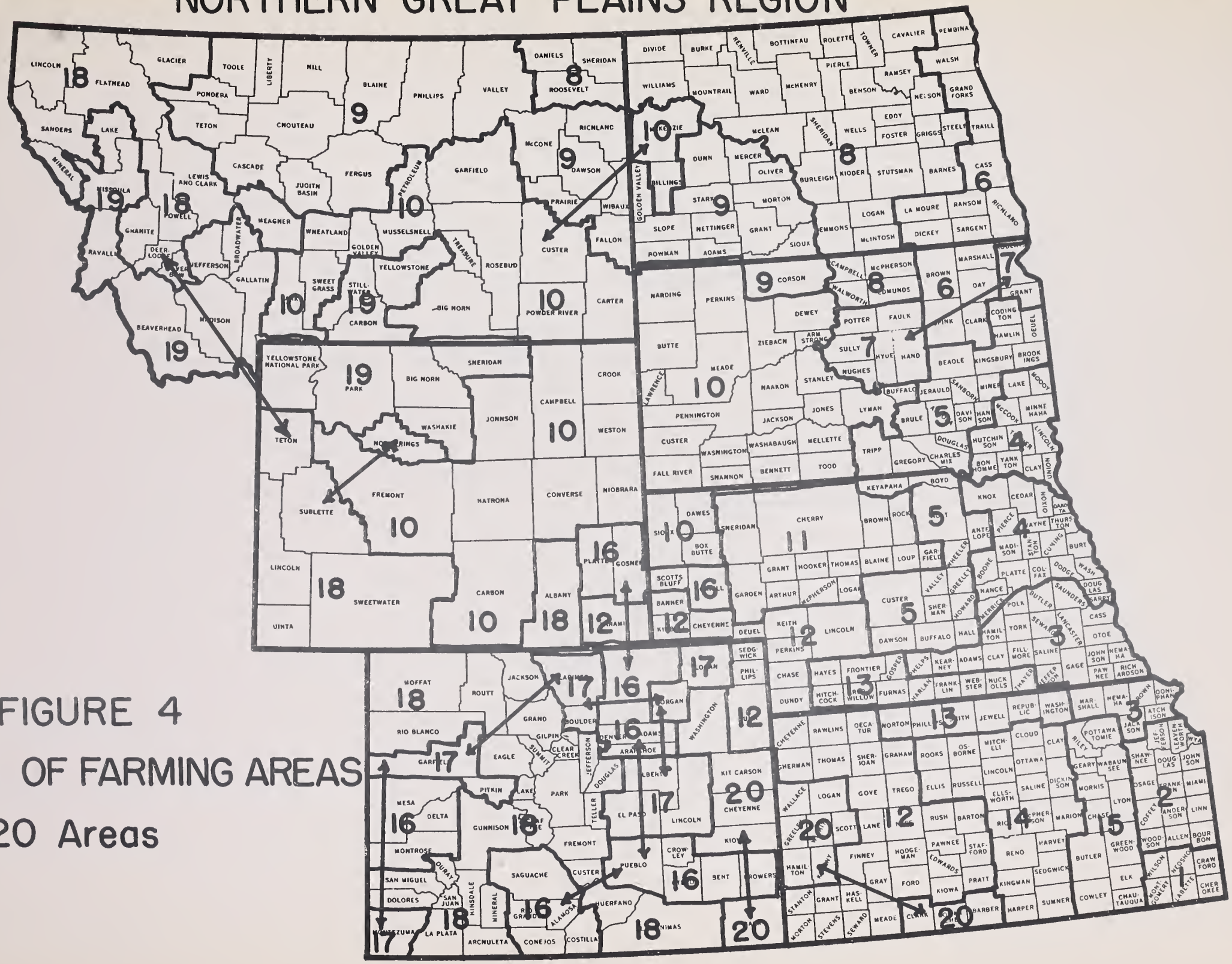

FIGURE 4

TYPE OF FARMING AREAS

20 Areas

U.S. DEPARTMENT OF AGRICULTURE

BUREAU OF AGRICULTURAL ECONOMICS

NORTHERN GREAT PLAINS REGION

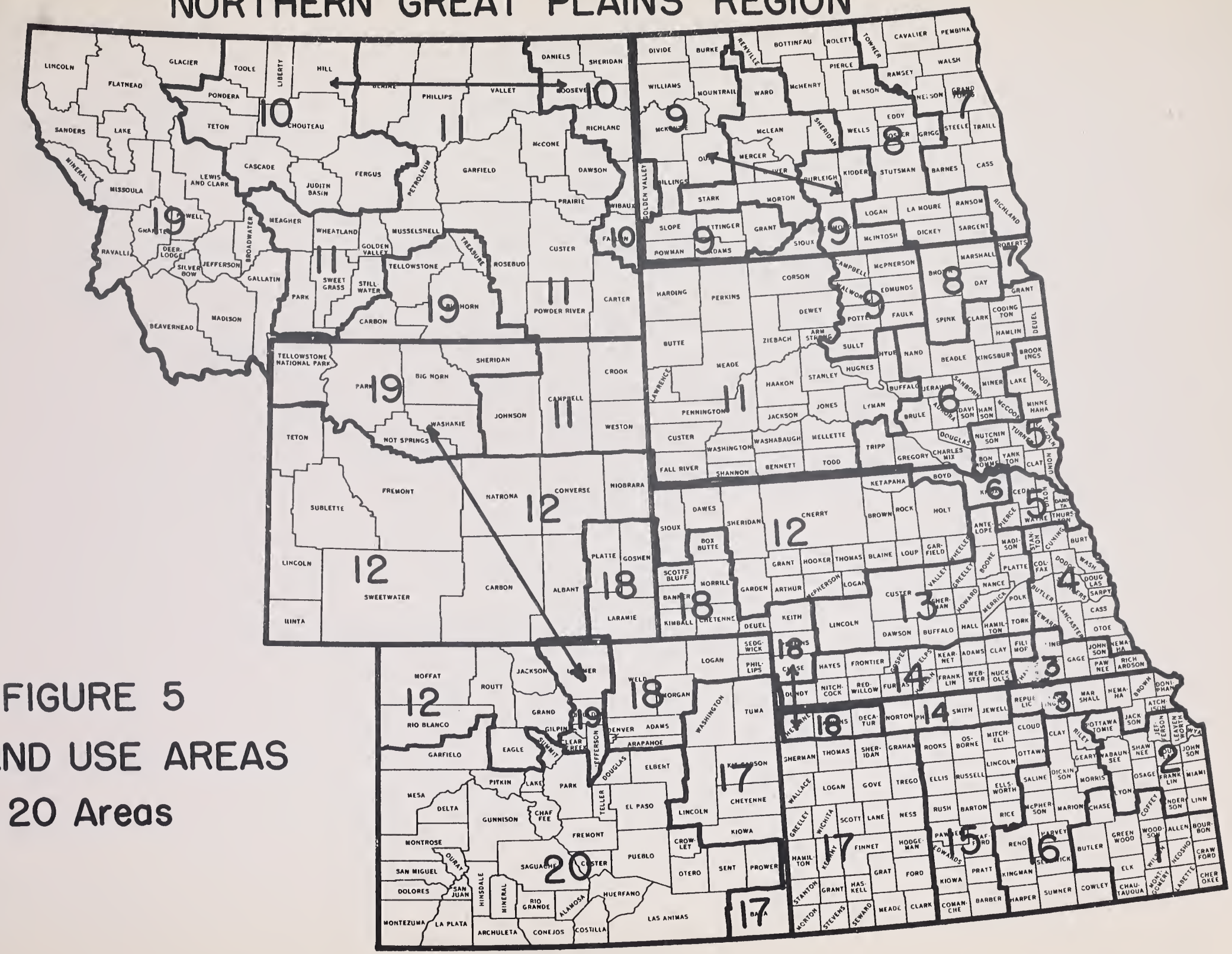

FIGURE 5
LAND USE AREAS
20 Areas

U.S. DEPARTMENT OF AGRICULTURE

BUREAU OF AGRICULTURAL ECONOMICS

NORTHERN GREAT PLAINS REGION

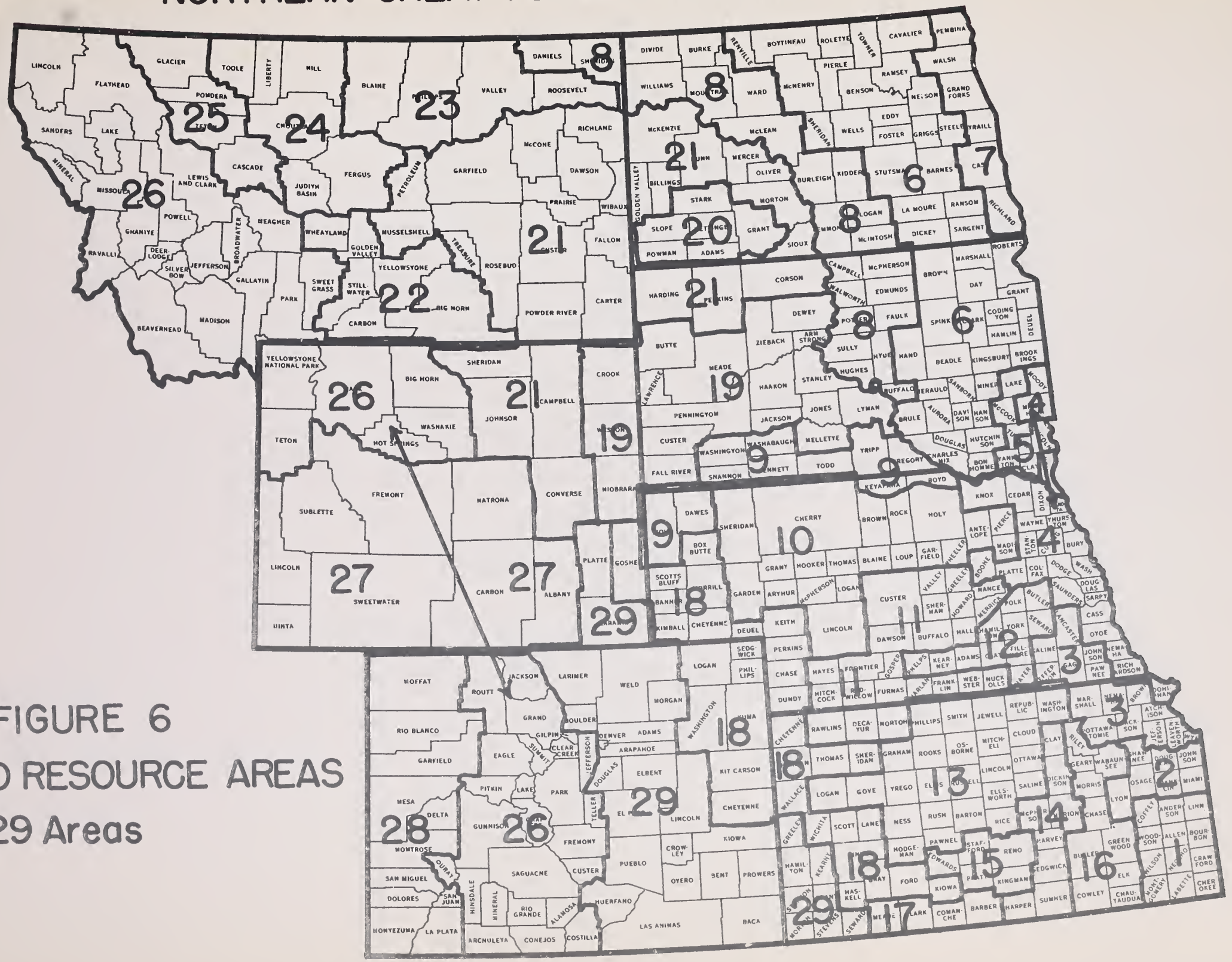

FIGURE 6
LAND RESOURCE AREAS
29 Areas

U.S. DEPARTMENT OF AGRICULTURE

BUREAU OF AGRICULTURAL ECONOMICS

NORTHERN GREAT PLAINS REGION

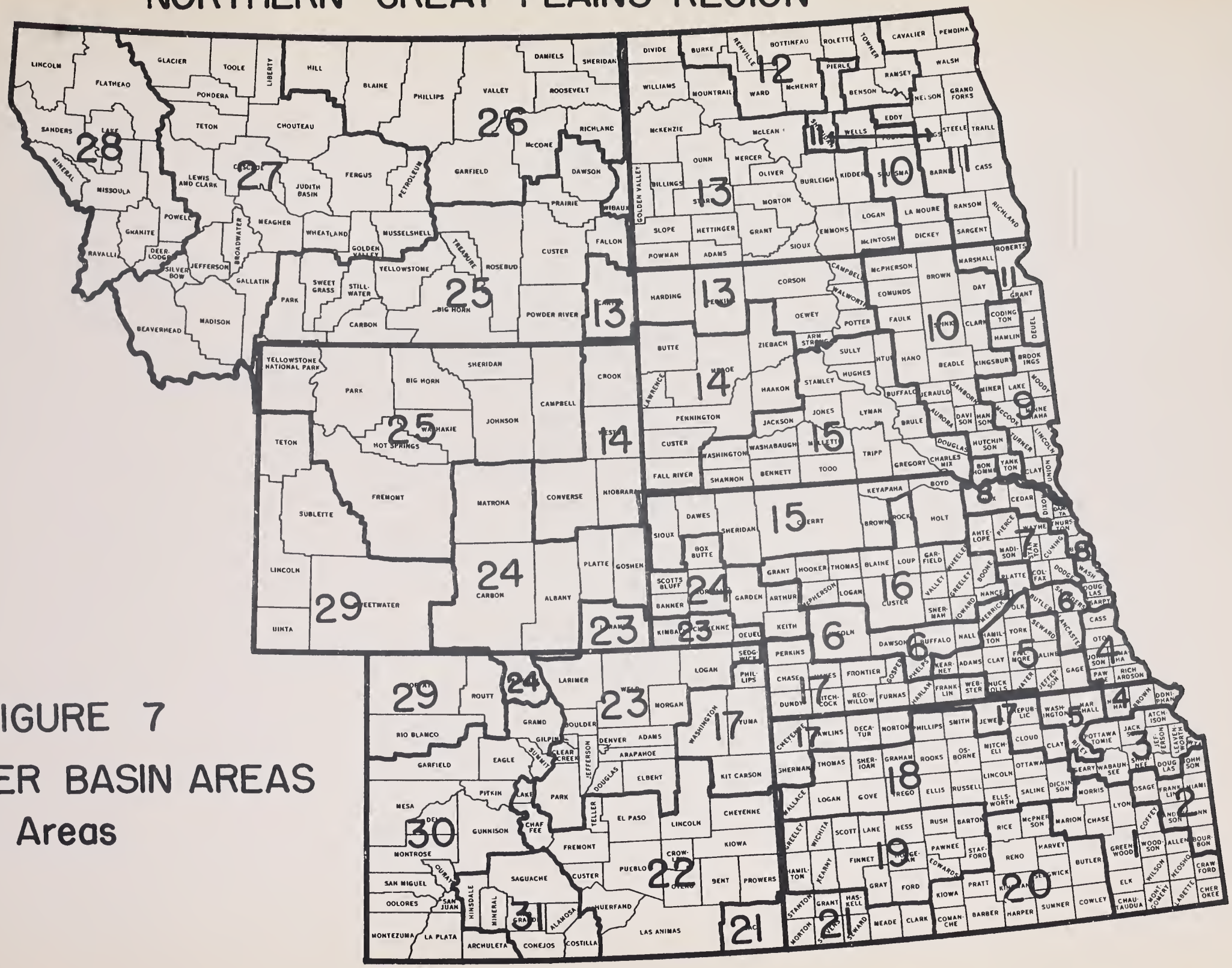

FIGURE 7

RIVER BASIN AREAS

31 Areas

U.S. DEPARTMENT OF AGRICULTURE

BUREAU OF AGRICULTURAL ECONOMICS

NORTHERN GREAT PLAINS REGION

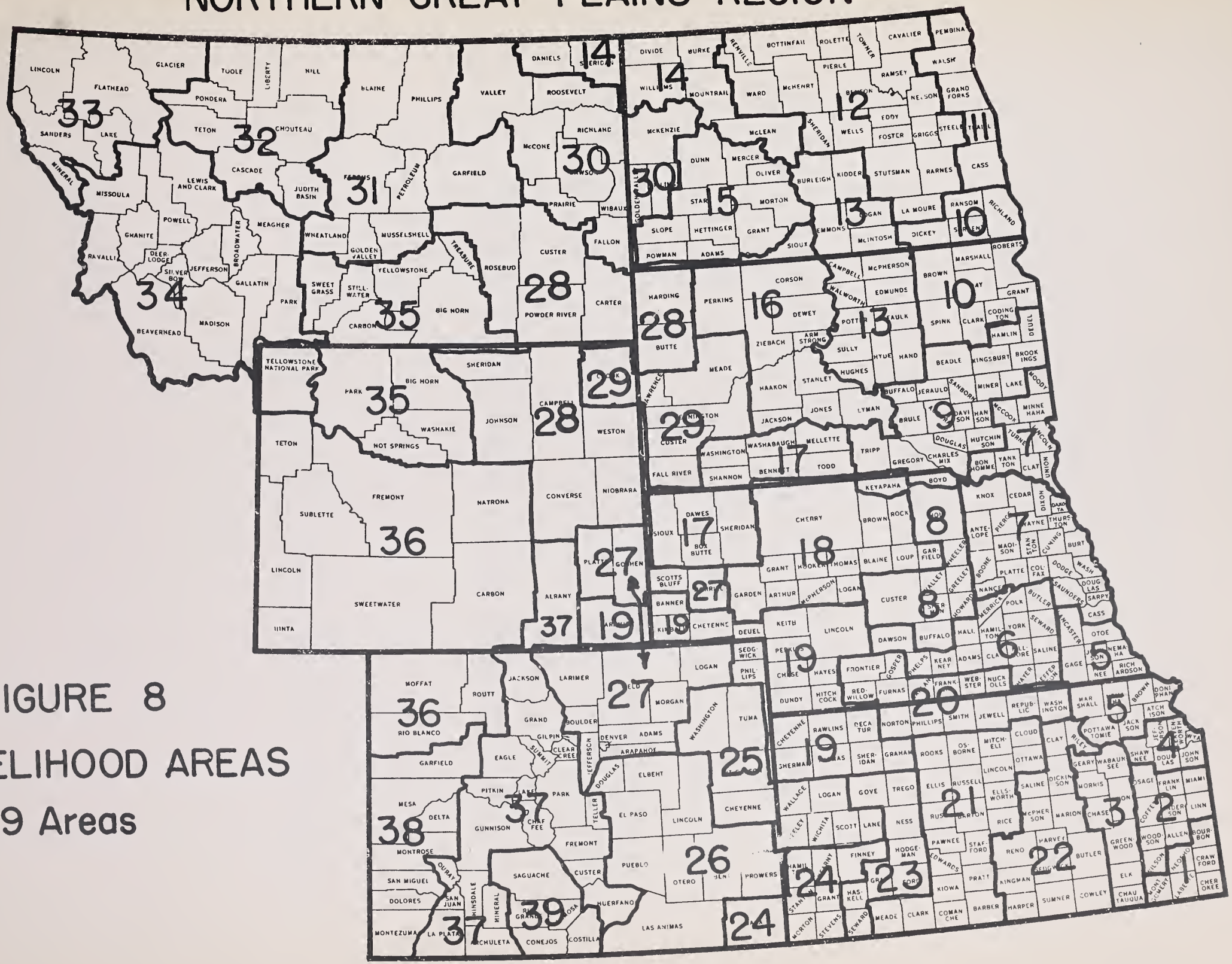

FIGURE 8

LIVELIHOOD AREAS

39 Areas

U.S. DEPARTMENT OF AGRICULTURE

BUREAU OF AGRICULTURAL ECONOMICS

APPENDIX

SELECTED BIBLIOGRAPHY

Because the relative efficiencies of alternative methods of stratification frequently depend upon factors outside of the scope of this paper, such as the method of selecting a county within a stratum, the method of expansion of sample indications, etc., the reader's attention is called to the following selected bibliography:

E. W. Burgess and T. C. McCormick, The Selection of Cases for Small Social Science Studies, Social Science Research Council, New York, August 1942.

Morris H. Hansen and William N. Hurwitz, A New Sample of the Population - Sampling Principles Introduced in the Bureau's Monthly Report on the Labor Force, Bureau of the Census, September 1944.

A. J. King and R. J. Jessen, The Master Sample of Agriculture, Journal of the American Statistical Association, Vol. 40, pp. 38-56, March 1945.

R. J. Jessen and E. E. Houseman, Statistical Investigations of Farm Sample Surveys Taken in Iowa, Florida, and California, Research Bulletin 329, Iowa Agricultural Experiment Station, Ames, Iowa, June 1944.

Jerzy Neyman, On Statistical Methods in Social and Economic Research, Lectures and Conferences on Mathematical Statistics, United States Department of Agriculture Graduate School, Washington, D. C., 1937.

George W. Snedecor, Design of Sampling Experiments in the Social Sciences, Journal of Farm Economics, Vol. XXI, No. 4, November 1939.

Norman V. Strand and R. J. Jessen, Some Investigations on the Suitability of the Township as a Unit for Sampling Iowa Agriculture, Research Bulletin 315, Iowa Agricultural Experiment Station, Ames, Iowa, May 1943.