

---

---

CHAPTER IV. VISUAL ANALYSIS —  
A MEANS OF EXPLORATION IN  
STATISTICAL RESEARCH

---

---

## CONTENTS

	Page
Visual analysis—A means of exploration in statistical research-----	227
ILLUSTRATIONS	
	Page
Figure 1.—Three stages in the preparation of visual-analysis cards—blank form ready for transcription, transcribed card, and coded and slotted card-----	229
Figure 2.—Three cards each representing a farm and farm family. All of the information used in the study concerning the farm and the operator's family is carried on one card-----	230
Figure 3.—Pack of slotted cards before beginning exploratory studies of relationships-----	231
Figure 4.—Sorting slotted cards by means of a needle-----	232
Figure 5.—Cards sorted to show the relationships existing between size of farm and selected characteristics-----	233
Figure 6.—Card designed primarily for visual-analysis purposes-----	235
Figure 7.—Cards sorted first by size of farm, then each size group halved to illustrate the stability of the relationships visible-----	235
Figure 8.—Cards sorted to show the effect of tenure on size of farm; also, that some characteristics related to size of farm are not affected by tenure-----	236
Figure 9.—Graphic illustration of the changes in number of farms by tenure which accompany changes in the size of farms used in these examples-----	237
Figure 10.—Farms reporting automobiles compared with those not reporting automobiles. In these examples characteristics associated with large farms rarely appear on cards representing farms not reporting automobiles-----	239
Figure 11.—Graphic illustration of the proportions shown in figure 10-----	239
Figure 12.—Cards representing farms reporting trucks are shown below the heavy line-----	241
Figure 13.—Relative lengths of bars indicate the changing proportions in these examples of farms reporting trucks in six size-of-farm groups-----	241
Figure 14.—Cards sorted to separate farms reporting tractors from those not reporting tractors. Cards used in these examples show most of the smaller farms in the nontractor group-----	243
Figure 15.—Bar graph showing, for the farms represented by the cards in figure 14, the increase in proportion reporting tractors as the size of farms increases-----	243
Figure 16.—Progressive selection. Final group of cards, shown at top, represents farms reporting automobiles, tractors, and trucks, and having electricity available-----	245

## VISUAL ANALYSIS — A MEANS OF EXPLORATION IN STATISTICAL RESEARCH

A representative of the research division of a national farm magazine made a visit to the Bureau of the Census for the purpose of discussing tabulation of material which would be of special value to advertisers using his magazine as a medium. The Census had a few months earlier made a subscriber study for the company and as a result had on hand all of the material previously prepared.

The representative was, of course, familiar with the methods used in making that study. His purpose in making a personal visit to the Census was to determine, as a preliminary to further tabulating, what the results would be. Having knowledge of the preparation for the earlier study, he knew that the Census was in position to furnish, in a matter of minutes, a preview of the results to be expected in any suggested tabulation of the material used. After having examined the results of several proposed tabulations, he arranged for payment of the costs and ordered selected tables prepared.

The representative was enabled to "see" the results to be accomplished because of the use in his study of a special card which provided a means of visually analyzing the material on hand. This visual analysis enabled the Census not only to present the probable results of several specified tabulations, but actually to show the representative approximately what those results would be when tabulations of his material had been made.

One of the questions asked of the Bureau of the Census was "Among our subscribers farms, what are some of the characteristics which distinguish those on which tractors are used as a source of farm power from those on which tractors are not used?"

Previously, no tables had been prepared which could answer this question, but within just a few minutes the information, applicable to the subscribers farms included in the sample, was supplied to the representative:

1. On most of the farms using tractors, the major source of income is from field crops; next in importance are those with major source from livestock sold or traded; and third, those with major source of income from dairy products.
2. Late model tractors were found among livestock and dairy farmers more often than among those whose major source of income was from crops.
3. Late model automobiles tended to accompany late model tractors.
4. Livestock and field crop farms reporting tractors ranged in size upward from 70 acres. Dairy farms ran smaller.
5. The majority of farms of over 140 acres reported tractors.
6. Only a very small proportion of farms of less than 70 acres reported tractors.
7. Farms without tractors reported fewer automobiles than did the farms with tractors.

With this information the representative ordered three separate tabulations which would show the numerical extent of the principal relationships brought out in this preliminary analysis.

<sup>1</sup>The system of visual analysis described in this chapter was developed by Carl R. Nyman and Orville C. Demaree of the Division of Agriculture, Bureau of the Census. In a routine search of the Patent Office records made in the interest of the Government, expired Letters Patent, No. 1,150,795, were found issued to Charles H. Tallmadge on August 17, 1915. In the letters patent was described a card embodying a system of classification by means of peripheral slots similar to the cards used in the illustrations accompanying this article.

The researcher is very often faced with the problem of determining which way to proceed. Given a wide and detailed knowledge of the subject with which to set up hypotheses the researcher can proceed great distances into uncharted fields. As the development of the project proceeds, however, problems arise. One of these is that of wasted effort. A long series of tabulations and computations completed and found somewhat short of supporting the hypothesis may prove, for example, that certain factors do not influence the final result to a significant extent. This information may be valuable, but often must be considered as wasted effort to the extent that it could have been avoided. In other words, if a shorter and more efficient means of determining the lack of influence of these factors had been used, the same results would have been known more quickly and at less cost.

Another problem arising is one involved simply with the completeness of the analysis. In proceeding with research and analysis with a given amount of data, when is the problem finished, or, when have all important factors been considered? With only three or four items or characteristics involved, the answer to this question can very often be obtained in a relatively short time. But in any analysis involving Census data when dozens of items of information bearing on certain subjects are available, the question is very definite. The selection, in advance, of all of the characteristics which will be required to complete the study, or the selection of those which have the most bearing on the problem in question, are of paramount importance. The funds available for a given project very often limit the study to the completion of a single planned attack. When, upon nearing the completion of the project, the analyst discovers other characteristics which look promising and he is unable to use them in place of some which proved of little value to his study, he is likely to feel as Moses must have felt when gazing over the land of Canaan.

What is needed then is a means of determining at low cost in time and effort the answers to these two questions: (a) What are the factors which are likely to be of influence in the study, and (b) which are of greatest importance. When the answers to these questions are known, the analyst can proceed with the knowledge that, when the project is completed, he will have thoroughly covered the subject within the limits prescribed.

With such a means at hand and used to its utmost capabilities, exact factors of influence can be determined, leaving for tabulations and further study the determination of the extent of such influence under the various restrictions imposed in the study. The purpose of this article is to describe a means of obtaining in advance of tabulations a thorough understanding of the factors involved and a general knowledge of their relative importance. The Census has been using for the past year a system which has amazing possibilities when used in many research problems. The system provides a means of keeping "visible" at all times all items involved in the study, and also a means of noting relationships existing between these items. Relationships between, and the relative importance of, the various characteristics can actually be seen on the edges of a pack of visual-analysis cards.

The principles involved in the visual-analysis system are simple. Using the individual farm as an example, all possible characteristics descriptive of each farm are coded and slotted in edges of cards. The cards may be of any convenient size and shape. On each card is carried a prepared form to which is transcribed from all sources all of the information to be used or considered in a particular study.

Figure 1 shows the three stages of development of the cards. The card at the top shows the prepared form ready for transcription, the center card is a sample of the transcribed data, and the lower card shows the same data coded and slotted in the margins of the card. Note that holes are provided for each code, but only the appropriate code is slotted for each item.

The particular cards used in this and all following illustrations were used in a special study mentioned in the opening paragraphs of this chapter. Cards designed specifically for visual analysis would be somewhat different. A layout for a strictly exploratory visual-analysis card expected to be used in the Census is shown later in figure 6.

For purposes of simplification, only a few items or characteristics of farms will be discussed. The purpose is to explain exactly how the system works rather than to work out any particular problem. Since it will be well to become familiar with these particular items in order to more readily understand the descriptions and illustrations to follow, a brief listing and a description of each of the selected items are here given.

Referring to the upper edge of the top card, at the left is a block of 7 holes labeled "Tenure." A card slotted in the first hole represents a farm operated by a full owner, or in other words, the operator owns all of the land he operates. Slotted in the second hole the card represents a farm operated by a part owner or a farm of which the operator owns a part and rents a part from others. Code 3 refers to managed farms and codes 4, 5, 6, and 8 to various forms of tenancy. Further explanation concerning tenure will not be necessary.

In the block of 6 holes labeled "Size of farm" the first hole on the left refers to farms of less than 30 acres. The second hole refers to farms of 30 to 69 acres, the third to 70 to 139 acres, the fourth to 140 to 179 acres, the fifth to 180 to 259 acres, and the last to farms of 260 acres and over. Question 12 in the upper section of the printed form on the card carries the acres in farm.

In the block marked "Mortgage" the left of the 3 holes, with the word "Yes" printed underneath, is slotted when the reply to Question 22 is "Yes" and indicates that a farm is operated by a full owner or part owner and carries a mortgage on the land and buildings of the farm. The center hole is slotted if the reply to Question 22 is "No," indicating that the farm is operated by a full owner or part owner and is free from mortgage. The hole at the right marked "Amt." is slotted when the amount of the mortgage is shown in reply to Question 23 on a full-owner or part-owner farm.

Holes are slotted in the block marked "Co-op" in accordance with replies to Questions 34, 35, and 36, a reply of "Yes" indicating that the farm operator did business with or through the kind of cooperative organization shown. The 3 types of cooperatives shown refer to selling, buying, and service cooperatives.

Holes in the blocks for "Autos," "Trucks," and "Tractors," are slotted in accordance with replies to these questions in the body of the card. In each group the hole to the left is slotted if the year of latest model is 1936 or later, the center hole is slotted if the year of latest model is 1931 to 1935, and the right hand hole is slotted if the year of latest model is earlier than 1931. In addition, to conserve space by not using an additional hole in each set, the right hand hole was slotted whenever a machine of any year model was reported. Thus, every card showing a machine reported was slotted in the right hand hole and if the year of latest model was before 1931, no other holes in the block were slotted. If the year of latest model was 1931 or later, one of the other two holes in the appropriate block was slotted.

In this discussion, only the left hand hole of the block marked "Electric" will be noted. This hole is slotted whenever the reply to Question 47 indicates that there was an electric distribution line within 1/4 mile of the operator's dwelling. The other two holes indicate the source of current when the operator's dwelling was lighted by electricity. They are "power line" and "home plant," respectively.

The hole marked "Tel" is slotted whenever the reply to Question 49 indicated that a telephone was reported.

All other holes on the card were notched in a similar manner in accordance with information shown on the card. For

the purpose of this discussion, however, a clear understanding is necessary of only those described.

After the material has been transcribed and coded and the appropriate slots made in the edges of the card, the slots on any individual card may be said to form a "pattern" for the farm represented. Another farm having the same characteristics would, therefore, have the same pattern of slots and, conversely, any difference in the pattern for individual farms would indicate corresponding differences in their characteristics. In figure 2 are shown such cards for three individual farms. Examination of these will show that as the characteristics of the farms change so do the patterns of slots.

When a group of cards, each card representing an individual farm, are placed together in a pack, the patterns of the individual farms are, of course, not visible. Only the edges are exposed. Figure 3 shows a pack of approximately 550 cards representing that number of individual farms. These particular farms are located for the most part in the Corn Belt. While many of these farms with respect to any one characteristic are similar, there are probably no two farms in the entire group which have exactly the same individual pattern of characteristics. But, as will be seen shortly, when two or more farms have one similar characteristic as indicated by slots in the same position, they are very likely to be similar in one or more other characteristics, and as a consequence have corresponding slots in similar positions. It is this coincidence of slots representing similar characteristics of farms which forms the basis for the visual system of exploration being described.

When a number of contiguous cards possess the same characteristic, this fact becomes evident in the appearance of a groove formed by the several coincident slots. Any process which places together all cards showing identically placed slots has automatically selected from the pack all cards representing farms similar in the characteristic represented by the particular slot being used in the selection. When these cards are placed in a separate pack by themselves, the slots representing this characteristic form a continuous groove across the pack. Figure 4 illustrates the manner in which slotted cards are sorted.

Note that the "needle" has been inserted entirely through the pack in holes provided in that position, and, upon being raised, all cards that have been slotted in that position fall and can be readily assembled with others having a similar slot. By simply sorting the cards, all farms possessing a specified characteristic are placed together. When the various characteristics are described by proper notation on the cards themselves, as has been done in these examples, the sorting of farms by characteristics is simple and rapid.

Earlier in this chapter it has been stated that when farms are similar with respect to one characteristic, they are likely to be similar with respect to others. In varying degrees, certain characteristics are accompanied by certain other characteristics and any change in the one is likely to be accompanied by a corresponding change in others. The closer the relationship between items, the more likely are these accompanying changes to occur. If the correlation were perfect between two items, any change in one would always be accompanied by a similar change in the other. Such perfect correlation, however, is hardly to be expected in any considerable number of cards.

When a pack of cards, sorted according to any given characteristic, is viewed from the edge, the slots representing that particular characteristic, as has been said, form a continuous groove across the pack. Slots representing other characteristics most closely related to the one used in sorting will tend to assemble in a similar group. In case of perfect correlation, another groove would be formed across the pack. While this rarely happens, it is not uncommon for as many as 90 percent of the cards representing a group of farms possessing one characteristic to also possess the same second characteristic. Less closely related characteristics will be represented by fewer continuous adjacent slots forming grooves. By this means the relative relationship between items can be judged by the relative frequency of coincidence of slots—the greater the frequency of coincidence, the closer the relationship.