5/83

THE ROLE OF A LARGE SCALE, AUTOMATED DATA BASE IN ON-GOING EXCAVATIONS: AN EXAMPLE FROM THE AMERICAN SOUTHWEST

Sylvia W. Gaines
Shereen Lerner
Department of Anthropology
Arizona State University

ABSTRACT

The development and utility of very large data bases for on-going excavations is still relatively new to archaeological endeavours. This paper focuses on the La Cuidad excavation project, located in downtown Phoenix. The dynamic nature of data base building from the initial to the final field phase is discussed and specific data sets are used to demonstrate various applications. Several graphics techniques are presented as well as the role they played in the field decision-making process. Factors of data flexibility, communication, documentation and data integrity are identified as the key to a successful application.

INTRODUCTION

Various types of computerized data bases have been employed by archaeologists during the past decade and their general efficacy and utility for planning and research have been adequately demonstrated. Our paper deals with a unique example - that of a very large, automated data base of Hohokam data from the extensive site of La Cuidad ("The City"). The site is located in central Phoenix, a major metropolitan center in Arizona with well over one million people.

BACKGROUND

La Cuidad is one of several major archaeological sites inhabited by the prehistoric Hohokam of the desert Southwest. The site was occupied for several hundred years, beginning ca. A.D. 700 and lasting through A.D. 1400. The major period of occupation appears to date between A.D. 900-1100. Architectural features characteristic of the Hohokam include pithouses, canals, trash pits, ball courts, roasting ovens, cremation and burial areas.

Approximately one million artifacts have been recovered at the site including both ceramic vessels and potsherds local to the area as well as a number of ceramic types indigenous to other areas in the American Southwest. A wide variety of lithic materials, shell from both the Gulf of California and the Pacific, paint palettes, effigies and figurines are also present. This data base will provide information regarding the extent of interregional exchange as well as the use and availability of local materials. Moreover, settlement data may be applied to community pattern studies and spatial information regarding accessibility to water. Ethnobotanical data recovered at the site will provide information regarding the types of plant foods utilized. Given the expansive data base and the range of research questions, computerization became the key to the success of the project.

Guidelines for Computerization

In our planning for the role of computerization in this project, we had few guidelines. Information on automated data bases from large scale excavation was not readily available. While we recognise that the excavation of every site is somewhat unique, our experience with developing computerized data bases from smaller scale projects did provide insight into alternative designs.

The advantages of computerized data processing are predicated on three critical factors: 1) rapid processing time, 2) capability of storing an extremely large amount of data, and 3) accuracy provided by computerized management and manipulation. In sum, time, size and accuracy are the cornerstone of the La Cuidad computer application.

The extensive data base anticipated for this project as well as the complexity of the research questions necessitated that we carefully consider the available computer resources. The Computer Center at Arizona State University maintains several computer configurations from micros to mainframes. Given the nature of our project we opted to utilize the IBM 3081 MVS which supports a variety of statistical packages, graphics programs and utility software. The archaeology computer laboratory on the university campus maintains a number of CRT terminals which were used for all phases of remote data processing.

The key factor in our initial planning for computerized procedures was the anticipated scope of work projected for field data recovery. Our field assessment suggested a five fold increase over the number of features estimated in an earlier investigation. Fortunately, our flexible, cost effective computerized data processing procedures allowed easy accommodation of this increased data base size.

The overall architecture or design of our data base and processing procedures was directly influenced by the excavation strategy. At each stage of field work, from initial trenching, to the excavation of features and random units, critical information was provided by the computerized data base. This capability necessitated a fairly rapid turn around time from field data recovery to preliminary analyses to data entry and data base updates to the computerized data analysis, searches and report generation.

EXAMPLES OF THE DATA SETS UTILIZED

As examples of the utility of our approach we have chosen to discuss three types of data sets (trenching data files, analysis files, and specimen files) from our integrated, computerized data base which illustrate computer generated information. The maps, graphs, charts, statistical tables, data searches and queries presented here, played a significant role in La Cuidad research questions concerning prehistoric lifeways in the desert Southwest.

Data Files

Briefly, the three data sets utilised in our examples are inventory information, artifactual data and a specimen number of file. The first data set is an inventory of features based on the profiles of trenches excavated during the trenching program on the site. Included in this inventory file are the feature numbers, feature types, feature locations in the site grid and feature depth below surface level.

The results of preliminary analyses of the lithic and ceramic artifactual material constitutes the second data file category. These analyses involved a 'rough sort' of the material into major artifact categories in anticipation of more specialized analyses at a later date. A separate file for ceramic and lithic data was developed.

(1) Lithics: the variables recorded include the level of decortification on a flake, number of cores, hammerstones, projectile points, and ground stone material as well as detailed proveniences.
(2) Ceramics: these data consisted of the number of plainwares, redwares, red-on buffwares (the Hohokam local decorated ware), intrusive ceramics (those non-local to the Hohokam area), and specific artifact classes such as spindle whorl, scoop etc. As with the lithic files, these data were entered into the data base with their corresponding feature number and provenience information.

The third type of computerized data set addressed in our discussion is a catalogue of specimen numbers including feature number, feature type, level, unit, guad and artifact type.

File Functions

Each of the above mentioned files was developed with a specific function in mind. These functions are briefly discussed here.

The inventory file of features was designed primarily to allow the spatial distribution of different types of features to be graphically depicted. The GIPSY computer mapping program and SAS graphics packages were used to generate maps of the site on which the spatial location of features was plotted. These maps provided information pertaining to preliminary patterns of feature locations and aided in the planning of future excavation strategies. For example, using the GIPSY grapics program we were able to plot any feature of artifact type to present a graphic display of spatial distributions. As seen in Figure 1, the locations of canals and as house floors have been plotted. From this map we were able to identify a generalised pattern of settlement with regard to access to available water at La Ciudad. Note how the majority of settlements are situated along the southern side of the canal. The orientation of the symbols depicted on this map are positioned in a vertical or horizontal manner, corresponding to the north-south or east-west placement of trenches on the site. The L-shape of the map indicates the portion of the site area actually investigated.

The SAS graphics software produced a choropleth map of floor features also identified in the trenching phase (Figure 2). This type of map depicts levels of density of a feature type throughout the site. The key on the bottom of the map illustrates different levels of artifact density, the darker patterns indicating higher concentrations. Another SAS graphics program, termed surface maps, allowed us to view a three-dimensional projection of the canal feature across the site. Figures 3 and 4 illustrate two capabilities of this technique, that of rotation and tilt. These two options provide different angle perspective of the plotted feature of the site. High peaks indicate high density of feature occurrence at particular loci. The break in the centre of the site area indicates the portion of the site not excavated.

An additional use of the inventory file was to provide information regarding the number and types of different features that occurred on the site. Figure 5 graphically depicts, in a pie chart format, the frequencies of some of the different types of features identified during the trenching phase. Notice the high frequency of trash pits and prepared surfaces, the latter of which were later re-classified as pithouse floors. It should also be noted that the excavation which followed the trenching phase located more features than previously identified and redefined some features which had been misinterpreted in the original profiles. Computerized data base management provided the capabilities to automatically redefine data and to restructure our files so that misclassification was not a problem.

The second file type, that of ceramic and lithic data, provided preliminary indications of the different artifact categories and densities of artifact classes that occur on the site. The breakdown of lithic categories is shown in Figure 6. This bar graph was produced using the SAS graphics package called HBAR. Some of these lithic categories have been combined to illustrate this particular graphic technique. The availability of a computerized data base at early stages of the analysis also permits predictions of the total number and types of artifacts that may occur on the site. In this regard, a more basic use of this file is to monitor the progress of analysis. From this file we are able to determine how many artifacts have been analysed to date. Using information on the number of

LA CIUDAD

GIPSY6 SEPTEMBER 1983

CANAL AND FLOOR FEATURES ON SITE

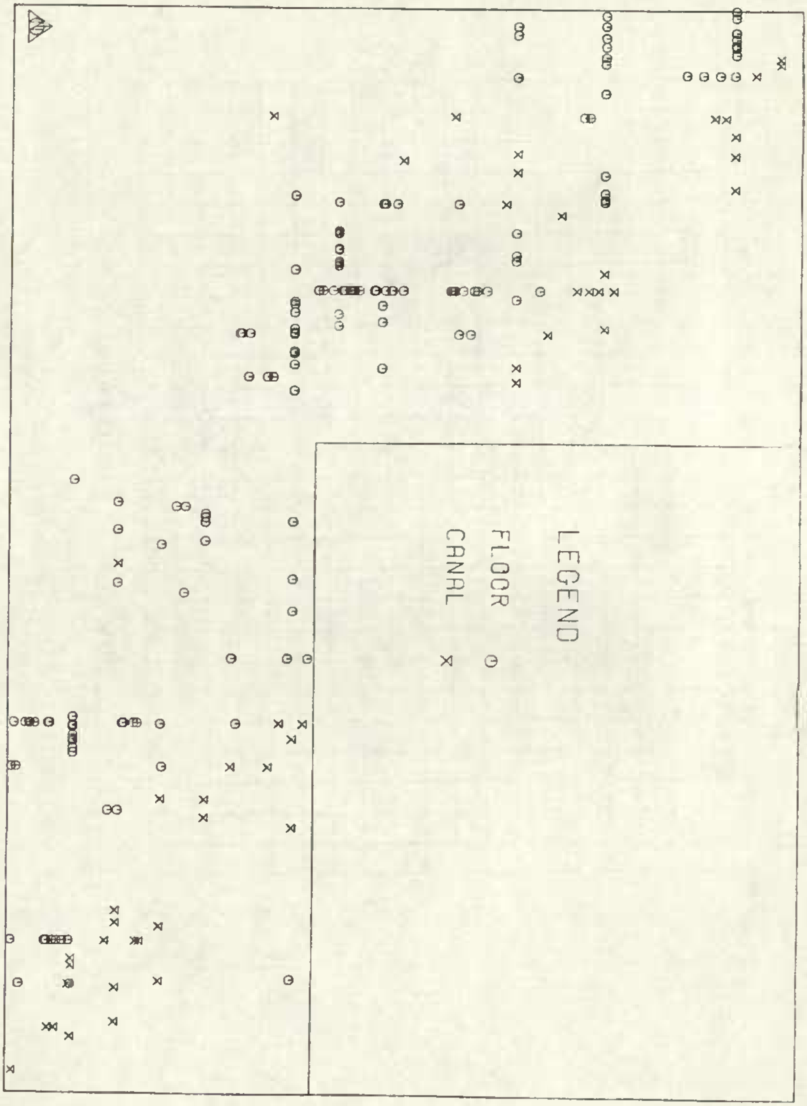


Figure 1

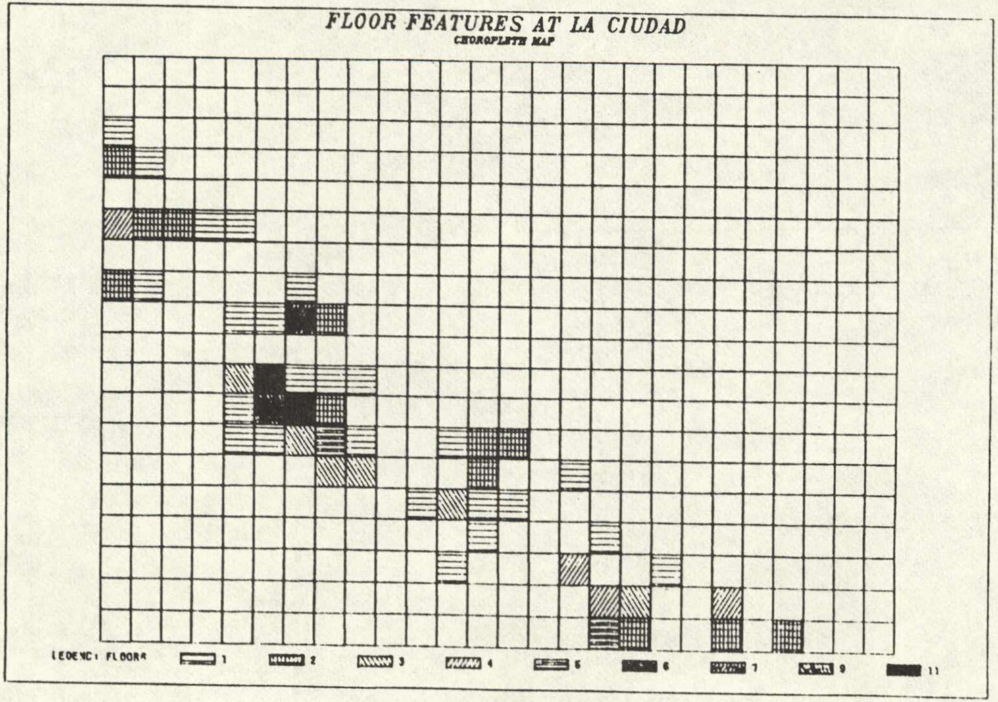
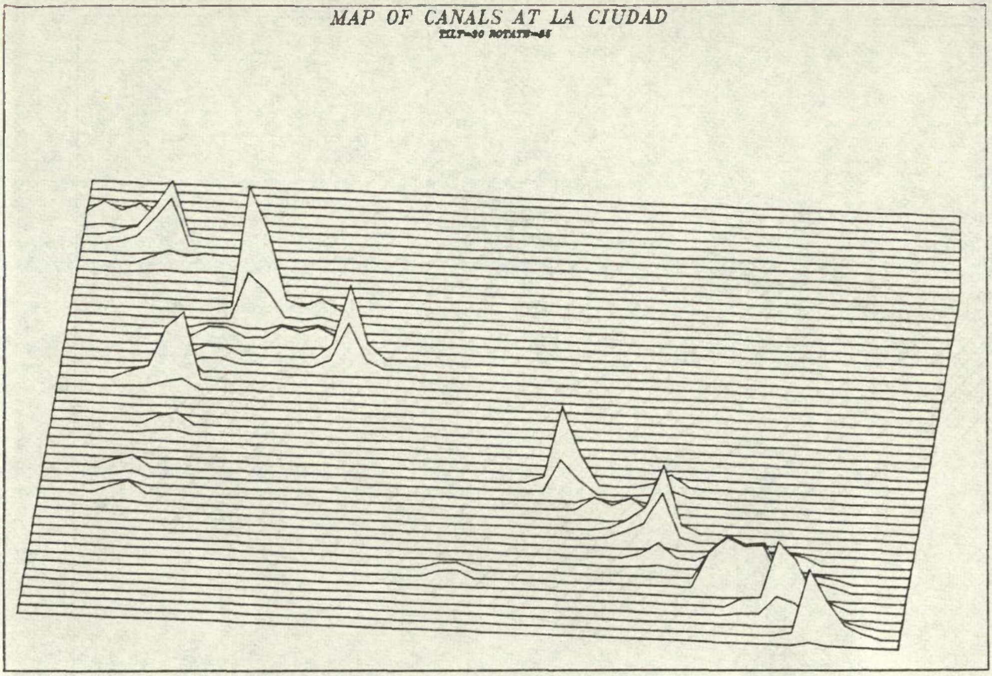

FLOOR FEATURES AT LA CIUDAD
CHOROPLETH MAP
LEGEND: FLOORS
1
2
3
4
5
6
7
9
11

Figure 2



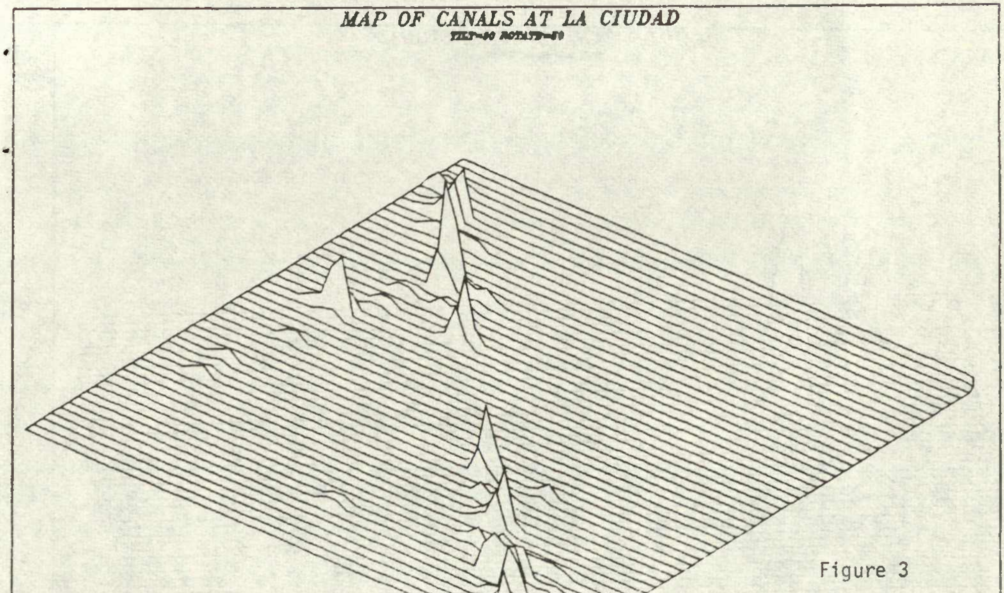
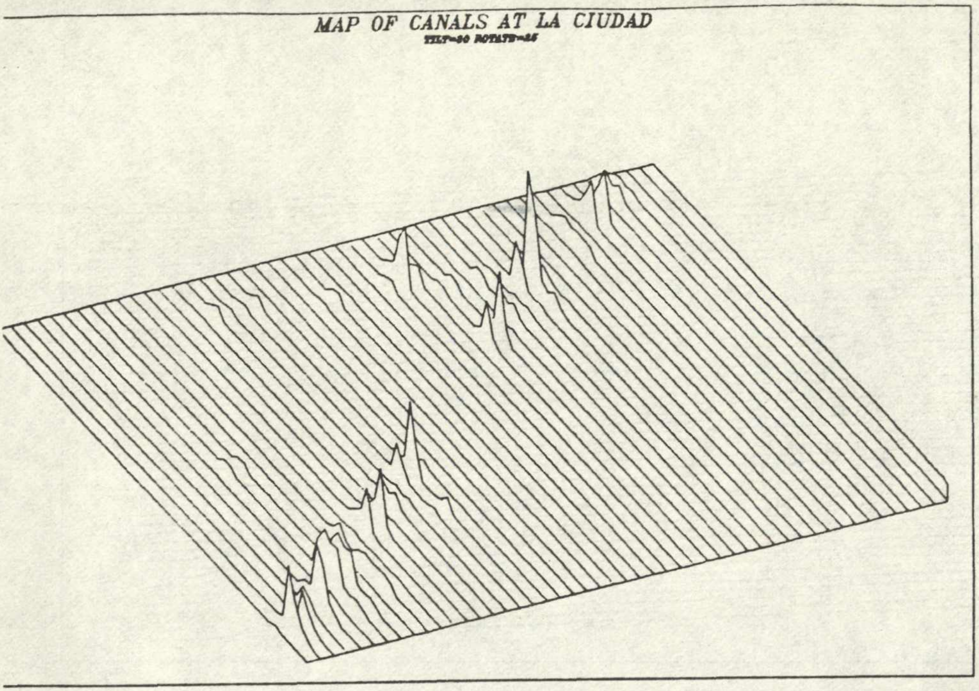


Figure 3
Surface Maps



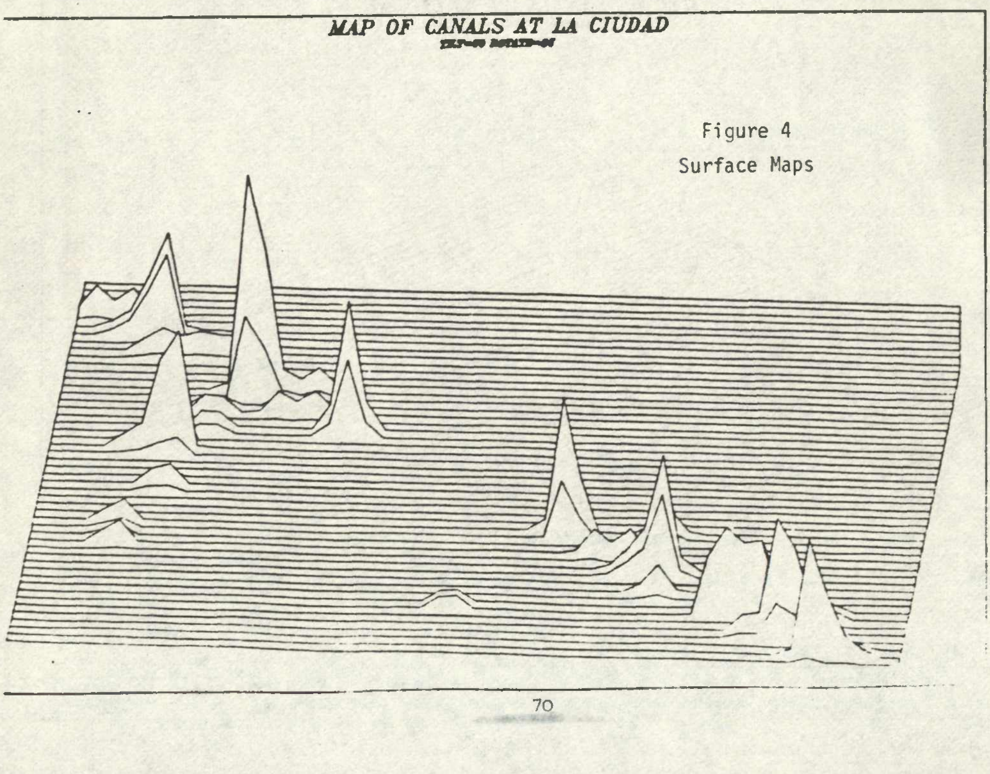
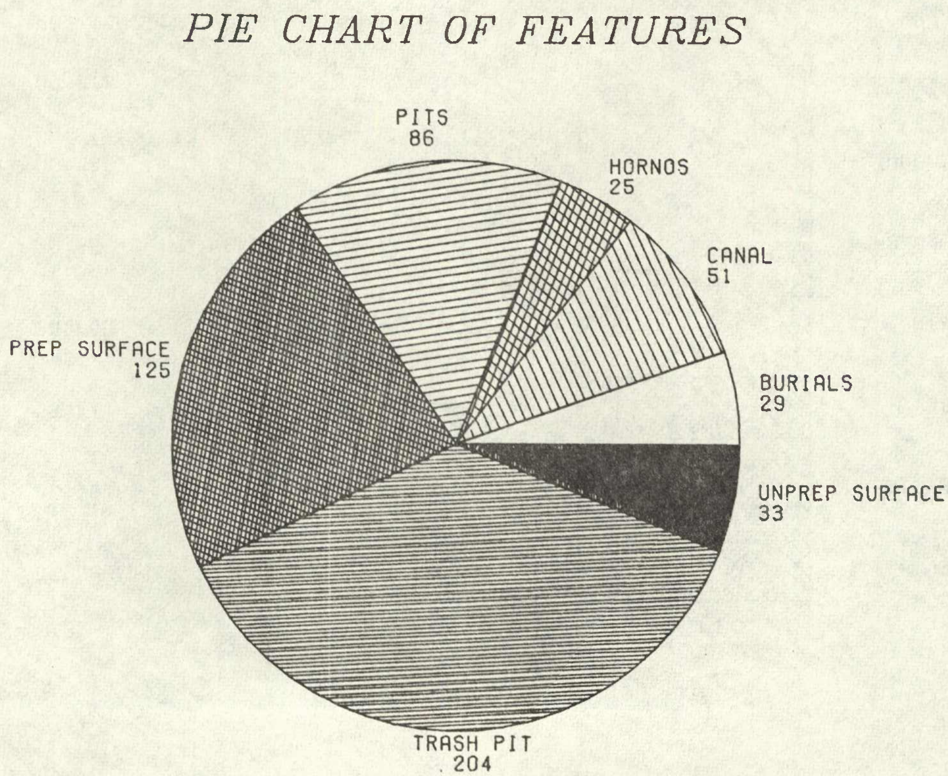


Figure 4

Surface Maps



PIE CHART OF FEATURES
PITS
86
HORNOS
25
CANAL
51
BURIALS
29
UNPREP SURFACE
33
TRASH PIT
204
PREP SURFACE
125

Figure 5

BREAKDOWN OF LITHIC CATEGORIES

LA CIUDAD

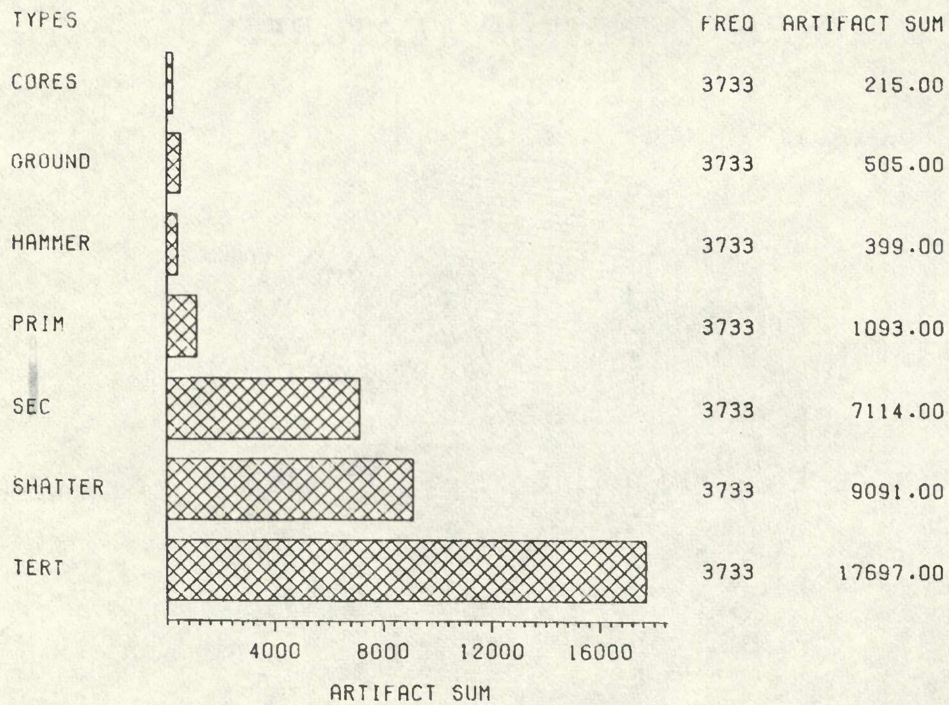


Figure 6

Bar Graph

remaining excavation units to be analysed it was possible to determine roughly how much additional time should be allocated to data analysis.

A second use of the ceramic and lithic data files is the potential for a spatial study of the distribution of particular artifact classes. For example, do certain types of artifacts occur in specific features on the site? In addition, these data files may be used as catalogues to aid in directing more detailed study of tne artifact classes. For instance, using the ceramic data file, an analyst may easily select those site features with high sample sizes of decorated sherds for a design attribute analysis of Hohokam ceramics.

The third type of file, a catalogue of specimen numbers, allows for the rapid selection of data for more detailed analyses. For instance, to select a sample of pollen data, user-specified parameters will allow the identification of a particular subset of data for further analyses. Table 1 depicts the diversity of artifact categories collected in all features across the site. Tables 2 and 3 illustrate artifacts found in pithouses and those artifacts found in trash pit context. Since these tables were compiled a number of artifact codes have been added to accomodate the wide variety of items recovered at the site. Additionally, although this table lists the first 12,00 cases, a final total of more than 50,000 specimen numbers has been attained.

EVALUATION OF THE COMPUTER APPLICATION

We will focus on five areas concerning our computerized procedures which have proven critical to the project: flexibility, communication, documentation, data integrity, and safeguards.

Foremost to the success of the project were the data management capabilities which allowed flexibility in terms of procedures and file structures. We were able to incorporate additional data categories and processing techniques as our feild work progressed. For example, our feature typology increased to include over 50 types as did our specimen categories. The data independence of our files assured that we could interface with multiple software packages as the project evolved. New SAS graphics techniques and utilisation of clustering programs are examples of this capability.

Communication between field, lab and computer components of the project, especially when rapid turn around time is required, is a prime consideration. When a number of individuals are involved in data capture and multiple data recording forms are utilised, errors increase. Specimen number forms and artifact analysis forms were used as computer data entry documents. In the course of the project many people have been responsible for recording and as a result, interpretive problems sometimes arose. We

ALL FEATURES

CATEGORY LABEL	CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	ADJUSTED FREQ (PCT)	CUM FREQ (PCT)
SHERDS	101.	2627	22.4	22.4	22.4
RECONSTRUCTIVE VESS	102.	6	0.1	0.1	22.4
WHOLE VESSEL	103.	4	0.0	0.0	22.5
SPINDLE WHORL	104	3	0.0	0.0	22.5
CLAY COIL	105.	28	0.2	0.2	22.7
CERAMIC FIGURINE	106.	17	0.1	0.1	22.9
CREMATION VESSEL	151.	1	0.0	0.0	22.9
WORKED SHERD	152.	6	0.1	0.1	22.9
UNKNOWN CERAMIC	199.	1	0.0	0.0	22.9
CHIPPED STONE	201.	2158	18.4	18.4	41.3
PROJECTILE POINT	202.	53	0.5	0.5	41.8
GROUND STONE	203.	478	4.1	4.1	45.9
GROUND AXE	204.	6	0.1	0.1	45.9
STONE VESSEL	205.	2	0.0	0.0	45.9
STONE PALLETTE	206.	13	0.1	0.1	46.0
FIRE CRACKED ROCK	207.	164	1.4	1.4	47.4
STONE ORNAMENT	250.	26	0.2	0.2	47.7
MISC TOOL	251.	94	0.8	0.8	48.5
SCHIST SLAB	252.	11	0.1	0.1	48.5
UNKNOWN STONE	299.	6	0.1	0.1	48.6
WORKED SHELL	301.	164	1.4	1.4	50.0
UNWORKED SHELL	302.	66	0.6	0.6	50.6
SHELL ORNAMENT	350.	83	0.7	0.7	51.3
UNKNOWN SHELL	399.	519	4.4	4.4	55.7
WORKED BONE	401.	34	0.3	0.3	56.0
FAUNAL	402.	19	0.2	0.2	56.1
HUMAN BONE	403.	20	0.2	0.2	56.3
BONE ORNAMENT	450.	2	0.0	0.0	56.3
UNKNOWN BONE	499.	1089	9.3	9.3	65.6
FLOTATION	501.	774	6.6	6.6	72.2
POLLEN	502.	1321	11.3	11.3	83.5
C14	503.	65	0.6	0.6	84.0
ARCHEOMAG	504.	4	0.0	0.0	84.0
SOIL	505.	125	1.1	1.1	85.1
MACROBOTANICAL	506.	123	1.0	1.0	86.2
DENDRO SAMPLE	507.	4	0.0	0.0	86.2
IMPRESSED DAUB	601.	22	0.2	0.2	86.4
DAUB	602.	222	1.9	1.9	88.3
PIT LINING	603.	1	0.0	0.0	88.3
UNKNOWN DAUB	699.	11	0.1	0.1	88.4
TURQUOISE	701.	8	0.1	0.1	88.4
MISC MINERAL	702.	525	4.5	4.5	92.9
UNKNOWN MINERAL	799.	7	0.1	0.1	93.0
HISTORIC MODERN	800.	693	5.9	5.9	98.9
WASP NEST	901.	6	0.1	0.1	98.9
SEA TURTLE-PLEISTOCE	902.	3	0.0	0.0	99.0
FISH SCALE	903.	1	0.0	0.0	99.0
ARTIFACT CLUSTER	998.	93	0.8	0.8	99.8
UNKNOWN MISC	999.	28	0.2	0.2	100.0
	TOTAL	11737	100.0	100.0	

FREQUENCIES OF ARTIFACT CATEGORIES IN SPECIMEN FILE

Table 1

PITHOUSES

CATEGORY LABEL	CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	ADJUSTED FREQ (PCT)	CUM FREQ (PCT)
SHERDS	101.	1272	20.9	20.9	20.9
RECONSTRUCTABLE VESS	102.	2	0.0	0.0	20.9
WHOLE VESSEL	103.	2	0.0	0.0	20.9
SPINDLE WHORL	104.	2	0.0	0.0	21.0
CLAY COIL	105.	10	0.2	0.2	21.1
CERAMIC FIGURINE	106.	6	0.1	0.1	21.2
WORKED SHERD	152.	3	0.0	0.0	21.3
UNKNOWN CERAMIC	199.	1	0.0	0.0	21.3
CHIPPED STONE	201.	1082	17.8	17.8	39.1
PROJECTILE POINT	202.	26	0.4	0.4	39.5
GROUND STONE	203.	232	3.8	3.8	43.3
GROUND AXE	204.	1	0.0	0.0	43.3
STONE VESSEL	205.	2	0.0	0.0	43.4
STONE PALLETTE	206.	6	0.1	0.1	43.5
FIRE CRACKED ROCK	207.	55	0.9	0.9	44.4
STONE ORNAMENT	250.	13	0.2	0.2	44.6
MISC TOOL	251.	50	0.8	0.8	45.4
SCHIST SLAB	252.	2	0.0	0.0	45.4
UNKNOWN STONE	299.	5	0.1	0.1	45.5
WORKED SHELL	301.	69	1.1	1.1	46.6
UNWORKED SHELL	302.	21	0.3	0.3	47.0
SHELL ORNAMENT	350.	33	0.5	0.5	47.5
UNKNOWN SHELL	399.	296	4.9	4.9	52.4
WORKED BONE	401.	22	0.4	0.4	52.7
FAUNAL	402.	5	0.1	0.1	52.8
BONE ORNAMENT	450.	2	0.0	0.0	52.9
UNKNOWN BONE	499.	520	8.5	8.5	61.4
FLOTATION	501.	426	7.0	7.0	68.4
POLLEN	502.	769	12.6	12.6	81.0
C14	503.	47	0.8	0.8	81.8
SOIL	505.	85	1.4	1.4	83.2
MACROBOTANICAL	506.	45	0.7	0.7	83.9
DENDRO SAMPLE	507.	3	0.0	0.0	84.0
IMPRESSED DAUB	601.	9	0.1	0.1	84.1
DAUB	602.	127	2.1	2.1	86.2
UNKNOWN DAUB	699.	6	0.1	0.1	86,3
TURQUOISE	701.	5	0.1	0.1	86.4
MISC MINERAL	702.	376	6.2	6.2	92.5
UNKNOWN MINERAL	799.	6	0.1	0.1	92.6
HISTORIC MODERN	800.	345	5.7	5.7	98.3
WASP NEST	901.	2	0.0	0.0	98.3
FISH SCALE	903.	1	0.0	0.0	98.4
ARTIFACT CLUSTER	998.	89	1.5	1.5	99.8
UNKNOWN MISC	999.	- 11	0.2	0.2	100.0
	TOTAL	6092	100.0	100.0	

Table 2

Pithouse frequencies

TRASHPITS

CATEGORY LABEL	CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	ADJUSTED FREQ (PCT)	CUM FREQ (PCT)
SHERDS	101.	252	17.6	17.6	17.6
CLAY COIL	105.	7	0.5	0.5	18.1
CERAMIC FIGURINE	106.	3	0.2	0.2	18.3
CHIPPED STONE	201.	223	15.6	15.6	33.9
PROJECTILE POINT	202.	4	0.3	0.3	34.1
GROUND STONE	203.	76	5.3	5.3	39.5
STONE PALLETTE	206.	1	0.1	0.1	39.5
FIRE CRACKED ROCK	207.	41	2.9	2.9	42.4
STONE ORNAMENT	250.	2	0.1	0.1	42.5
MISC TOOL	251.	14	1.0	1.0	43.5
SCHIST SLAB	252.	1	0.1	0.1	43.6
WORKED SHELL	301.	23	1.6	1.6	45.2
UNWORKED SHELL	302.	18	1.3	1.3	46.4
SHELL ORNAMENT	350.	11	0.8	0.8	47.2
UNKNOWN SHELL	399.	56	3.9	3.9	51.1
WORKED BONE	401.	3	0.2	0.2	51.3
FAUNAL	402.	1	0.1	0.1	51.4
UNKNOWN BONE	499.	129	9.0	9.0	60.4
FLOTATION	501.	176	12.3	12.3	72.7
POLLEN	502.	218	15.2	15.2	87.9
C14	503.	3	0.2	0.2	88.1
SOIL	505.	2	0.1	0.1	88.3
MACROBOTANICAL	506.	30	2.1	2.1	90.4
IMPRESSED DAUB	601.	2	0.1	0.1	90.5
DAUB	602.	34	2.4	2.4	92.9
MISC MINERAL	702.	54	3.8	3.8	96.6
HISTORICAL MODERN	800.	47	3.3	3.3	99.9
WASP NEST	901.	1	0.1	0.1	100.0
	TOTAL	1432	100.0	100.0	

VALID CASES 1432 MISSING CASES 0

Trashpit frequencies

Table 3

found that close communication with the lab. director was absolutely essential in coordinating the data capture efforts of so many individuals.

Documentation is closely allied to communication. We recommend clear documentation of procedures as they are developed during the course of a project. NEVER TRUST TO MEMORY is a cardinal rule of this type of application. As the project progressed new individuals were added to the staff and the documented procedures, particularly a procedural manual, proved to be most valuable. Additionally, all individuals involved in data entry and processing were encouraged to keep accurate logs of their computer work. Logs can often indicate sources of human errors and can serve as documents for reviewing work efforts.

The fourth point regards data integrity. A large scale field project can highlight the problem of human error. To assure maximum quality control of computerized data, automated data verification is ideal. If software is not available for this procedure, we highly recommend manual checks of the data file. Although such checking is time consuming, in the long run the benefits of error free data are worth the effort. All data stored in the La Cuidad computerized data base has been checked for error.

The final point concerns safeguards. Effective safeguard procedures are absolutely essential, especially in handling a data base as large as that of the La Cuidad Project. From the earliest stage of computer entry, backup systems should be employed with allow data recovery if the on-line system should malfunction. We periodically backup disk storage with magnetic tape updates. Human error can also readily destroy thousands of computer records and hundreds of manhours of work. As a safeguard, we maintain backups, or duplicate on-line files. Moreover, the 'raw data files', that is, data as they are initially entered, are always saved intact on separate disk files. This procedure provides a file which can always be accessed if other files are inadvertantly destroyed.

CONCLUSIONS

In conclusion, we would suggest that the maximum benefit derived from a computerized data base is directly dependent on quality control. As archaeologists, it is our responsibility to insure the most effective means of data capture available. The effort must extend from field to lab. to computerized data base development. Future archaeological research will rely even more heavily on the computerized data bases which we are currently developing.