NOT FOR QUOTATION
WITHOUT THE PERMISSION
OF THE AUTHORS

THE LEXIS PROGRAM FOR CREATING SHADED CONTOUR MAPS OF DEMOGRAPHIC SURFACES

Bradley A. Gambill
James W. Vaupel

December 1985
WP-85-94

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

Acknowledgments

We thank Nathan Keyfitz, Anatoli I. Yashin, Martina Joestl-Segalla, and Susanne Stock.

THE LEXIS PROGRAM FOR CREATING SHADED CONTOUR MAPS OF DEMOGRAPHIC SURFACES

Bradley A. Gambill and James W. Vaupel

INTRODUCTION

The LEXIS computer program, which was developed at the International Institute for Applied Systems Analysis (IIASA) and Duke University, is intended to aid demographers in the analysis of large arrays of data. Its application as a supplement to other methods of graphic display is demonstrated in "Thousands of Data at a Glance: Shaded Contour Maps of Demographic Surfaces" (Vaupel, Gambill, and Yashin, 1985) and will not be discussed here. This paper provides instructions on the use of the program, gives some hints concerning the art and craft of using the program in a creative way, and briefly describes the algorithm used in designing the program. A diskette containing a copy of the LEXIS program may be obtained from the authors or from IIASA. The program is copyrighted but the diskette is not protected against copying: please feel free to make and distribute copies. By making the program available to demographers and others interested in mapping the contours of surfaces, we hope to encourage the development of this method of data analysis. We would, of course, sincerely appreciate it if we and the International Institute for Applied Systems Analysis were acknowledged when the program or some modified version of it is used to produce maps for presentation or publication. Comments and suggestions are welcome!

ADDRESS COMMENTS TO: Bradley A. Gambill, 1315 Morreene Rd, Apt. 5H, Durham, N.C. 27705, USA or to James W. Vaupel, Humphrey Institute of Public Affairs, University of Minnesota, 301 19th Avenue South, Minneapolis, Minnesota 55455, USA.

ADDRESS REQUESTS FOR COPIES OF THE PROGRAM TO: either of the above or to Population Program, International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria.

USE OF THE PROGRAM

Important Preliminaries

We assume throughout this manual that the reader has some knowledge of the IBM personal computer and the DOS operating system. If this is not the case, we recommend that a potential user of the LEXIS program review the DOS manual provided with his or her IBM or IBM-compatible personal computer. Because the program does very little input/output checking, such knowledge may be quite helpful. We have not yet developed the program so that it can readily recover from various errors the user might make: it is best to be careful and if you do make an error, to start again.

We also assume that the reader has had some exposure to the use of shaded contour maps for demographic analysis. If not, the reader might peruse Vaupel, Gambill and Yashin (1985), Caselli, Vaupel and Yashin (1985), or Zeng, Vaupel and Yashin (1985).

The LEXIS program requires an IBM or IBM-compatible personal computer with graphics and with at least 128K of memory. The diskette accompanying this manual contains two files: LX.COM and USFERT. The use of the program will be discussed shortly, but it is important that the program be copied to a diskette formatted by the user's computer to include the DOS system and GRAPHICS for convenience. We recommend that the original diskette and a backup copy be stored in a safe place.

Printing Maps from the Screen

To print a LEXIS map from the computer screen, you must enter GRAPHICS at the > DOS prompt before beginning the LEXIS program. Once the surface has been completely drawn on the screen, enter the <shift> and <PrtSc> keys simultaneously to print the map.

Format of Input Data

The data to be used by the LEXIS program must be in ASCII format, with m repetitions of a pattern consisting of the value of a label (such as the current year) followed by n data items (such as mortality rates at various ages). The value of n must be less than 150 and the value of m must be less than 300. For example, to plot a surface of mortality rates $q(a,t)$ from age 0 to 99 and year 1900 to 1982 the first data entry in the input file would be 1900, the second entry $q(0,1900)$, the third entry $q(1,1900)$, and so on up to the 101st entry which would be $q(99,1900)$. The 102nd entry would be 1901, followed by $q(0,1901)$, and so on up through $q(99,1901)$. This pattern would be repeated up through the year 1982, the very last entry being $q(1982,99)$. The label does not have to be a year: it could be an age or a life-expectancy or the value of any other variable that describes the data. Similarly, the successive data values following the label do not have to pertain to successive years of age, but could pertain to successive values of any other variable that describes the second dimension of the data. In any case, with input data in this format, the map will have the n-variable (e.g., age) on the y-axis, the m-variable (e.g., year) on the x-axis, and the data values (e.g., the q's) as the third dimension represented in the map.

Zeroes may be used to represent missing values in the data, if the actual data range does not include zero. This use of zeros is discussed below.

Beginning the LEXIS Program

To begin the LEXIS program, simply type LX at the DOS prompt. A title screen will appear at the start of the program and disappear after a few seconds, leaving the Main Menu on the screen.

Main Menu Options

A: Initializing data files

All data to be mapped must be converted and stored in a file specially created by the initialization procedure of the LEXIS program. This procedure begins with the selection of option A from the Main Menu. The user is prompted to enter the disk drive and name of the input data file, as well as the disk drive of the output file. The initialization process creates two files, both with the same prefix as the input file, with the new data file given the extension .DAT, and another file, which contains information about the specifications of the data, given the extension .SPF. The information contained in the .SPF file is provided by the user of the LEXIS program, as described in the next step below. A list of the required information will appear on the screen, and the cursor will move from place to place to prompt the user to enter the appropriate numbers. The user is asked to provide information on both the "input file" (ASCII file) and "initialized file" (.DAT file).

The information provided about the input file must accurately describe that file: any error here will lead to program failure or strange output. The starting points for the initialized file can be greater than the starting points for the input file and the ending points can be less than the ending points for the input file. This option allows the user to create a .DAT file containing less information than is contained in the original ASCII file - a useful option if the ASCII file contains regions of numbers that will never be graphed, such as missing values or inaccurate numbers.

After the user has entered the requested information, the screen might appear as follows:

X-Axis information

	Input file	Initialized file
Starting point:	1900	1900
Ending point:	1982	1982
Label:	-	Year

Y-Axis information

Starting point:	0	0
Ending point:	99	99
Label:	-	Age

Notice that the x-axis information must be entered first, then the y-axis information. In our example of a mortality surface, we would enter the year-axis information, followed by the age-axis information. Once all necessary information has been entered, the program will begin the process of creating the .DAT file by printing the label (e.g., year) of the first column on the screen, followed by the label of each successive column. Checking the column labels as they appear on the screen is an excellent way to be sure that the data is in the proper format, and contains the proper number of entries in each data column. Once the initialized files are created, they are stored on the diskette and are available for all future mapping.

It is recommended that the user maintain two libraries of data - one with the original ASCII files, and the other containing LEXIS.DAT files. The extra effort required to maintain a second set of data will be rewarded with significant timesavings later, especially in cases of frequently mapped data.

B: Selecting Display Colors

The selection of pattern is a powerful option provided by the LEXIS program. Included in the program are two sets of color combination patterns or "palettes"— the "screen" patterns which fade from green to yellow to red, and the "print" patterns which use the colors necessary to produce appropriate shades of grey when a contour map is printed on an IBM or Epson printer or displayed on an appropriate black and white monitor. The screen patterns are probably most pleasing to the eye, and provide distinguishable shades of logically progressing colors on the screen; the print patterns provide the most easily distinguishable patterns when displayed on the printer. It is possible to alter these screen patterns (but not the print patterns) by entering 3 for pattern color and then placing a 1, 2 or 3 next to the appropriate region when prompted to do so; the value 4 denotes the default value such that the color will be the same as on the screen pattern. We have found it useful to shade ratio surfaces with just three colors, one representing the area below 1, one the region approximately equal to 1, and the last the region above 1. Other examples of use of this option are included in Vaupel, Gambill and Yashin (1985); experimentation will help a user decide which color patterns lead to the most informative maps.

Selection of display colors is optional, with background color zero, line color zero, and a standard screen pattern being the default settings. An experienced user of the LEXIS program may find the selection of display colors to be a useful option. By selecting this option, the user is able to select background color, line color, and pattern to be used in creating the map. The choice of background color is self-explanatory, and is a matter of personal taste for the user. The selection of line color depends on the data and the user's goals: if the lines are an important part of the map, they should be drawn in a color different from those used to create the patterns being used (yellow for the print patterns, background for the screen patterns). If the exact location of the lines is less important than the general trends of change within the surface, then it may be more informative to hide the lines by making them the color of the most dominant color of the patterns being used. Experimentation will allow the user to determine which combinations of color provides the most effective presentation of data.

C: Creating a map from a data file

Once all preliminary initializations and selections are completed, the map can be created by selecting the C option from the Main Menu. The surface creation menu will appear and prompt the user to select the type of file to be displayed. The map creation process is very similar for all options, with the main difference coming in the initial prompt for filenames.

Entering filenames

If only one file is to be used in creating the map, only one input prompt will be given. If two files are to be used, the initial prompt will be for the numerator or minuend filename, and the second prompt for the denominator or subtrahend file for ratio and difference surfaces, respectively.

Map size

The next step in creating a map is to enter the size of the surface to be displayed. The program will display the file specifications of each input file used in creating the surface, and then prompt the user to enter specifications for the map to be created. To map the entire data file, simply enter <return> for each starting and ending point. To map only part of a file, enter the appropriate numbers when prompted to do so. If the map being created is either a ratio or difference surface, it is important that the user take care in entering the size of the surface, so that the boundaries do not extend beyond those of either source data set.

Data with zeros

Once all of these numbers are entered, the user will be asked if the actual data include zeros. If the data range includes zero, then "Y" should be entered at this prompt. If, however, zeros are used only to denote missing values, in cohort data for example, then the user should enter "N" or <return>. For ratio surfaces, division by zero will be set equal to the background color.

Smoothing

Next, the user will be asked if the surface should be smoothed. If smoothing is desired, the user must enter an odd number less than or equal to twenty-one. The smoothing procedure simply replaces each point in the surface with the binomial weighted average of the points in the n by n square surrounding the point. For example, if n is five, 25 points surrounding the original point will be averaged using the following weights:

1	4	6	4	1
4	16	24	16	4
6	24	36	24	6
4	16	24	16	4
1	4	6	4	1

When a point is near a boundary, the weighted average is taken over all those points within the surrounding square for which data are available.

Line selection

Line selection is perhaps the most important decision that must be made in creating a shaded contour map. The LEXIS program provides seven different options for selecting the best lines:

1. "Mortality surface lines" begin with .000667 and then increase by 50% from one line to the next. We found these intervals to be useful in drawing contour maps of mortality surfaces. The contours are placed at the following levels (where B denotes "boundary"):

B1 = 0.000667	B5 = 0.003377	B9 = 0.017094	B13 = 0.086541
B2 = 0.001001	B6 = 0.005065	B10 = 0.025642	B14 = 0.129811
B3 = 0.001501	B7 = 0.007598	B11 = 0.038463	B15 = 0.194717
B4 = 0.002251	B8 = 0.011396	B12 = 0.057694	

2. "Fertility surface lines" are placed at intervals we found to be convenient when analyzing fertility surfaces:

B1 = 0.001	B5 = 0.05	B9 = 0.13	B13 = 0.21
B2 = 0.01	B6 = 0.07	B10 = 0.15	B14 = 0.23
B3 = 0.02	B7 = 0.09	B11 = 0.17	B15 = 0.25
B4 = 0.03	B8 = 0.11	B12 = 0.19	

3. "Ratio lines" are placed at levels we sometimes found useful when analyzing surfaces that represented the ratio of two surfaces (e.g., male vs. female mortality). Each contour is 10 percent greater than the previous contour, as follows:

B1 = 0.513158	B5 = 0.751315	B9 = 1.1	B13 = 1.61051
B2 = 0.564474	B6 = 0.826446	B10 = 1.21	B14 = 1.771516
B3 = 0.620921	B7 = 0.0909091	B11 = 1.331	B15 = 1.948717
B4 = 0.683013	B8 = 1	B12 = 1.4641	

4. The "difference lines" option allows the user to enter high and low values for a surface. The program then calculates the levels of 15 lines between the two points. These lines, depending on the user's choice, are placed either at even multiples or even intervals.
5. The "even multiple lines" option allows the user to enter a beginning point, and a multiplier. The first line is placed at the beginning point and the other lines are spaced at even multiples of the multiplier given. Thus, if 1.1 is the given multiplier, the lines increase by 10 percent, starting at the beginning point.
6. The "auto select lines" option takes the high and the low of the input file, and places 15 lines between the two points. The lines are placed at even multiples or even intervals, depending on the user's response to a prompt.
7. The "user entered lines" option allows the user to place the lines at any desired values. By placing more than one line at the same value, some of the contour lines and shades can be suppressed. In this way, a map with only five levels, for example, can be created.

Once the lines have been selected, the program will begin generating the surface, with no other user-entered information necessary until the surface is fully completed.

Surface Manipulation

When all calculations for the surface have been completed (drawing the border is the last step in the process), the user may view labeling, starting and ending points, and lines drawn on the map by entering <F1>. A new screen will appear with the pertinent information, and a new screen manipulation menu at the bottom.

<F1>View Map - This option allows the user to flip back and forth between the map and information screens with the touch of a key.

<F2>Add Grid - This option enables the user to add a grid to the screen map in order to get more precise x and y axis readings. Within this option the user may chose where to start the grid on each axis, the grid line increments, the color of the grid lines, and also a color to "exclude" from the grid. We have found that it is sometimes more informative to leave the contour lines intact, on top of the grid, because it relegates the grid to the background, rather than to the foreground of the map. By entering <return> for each option, the default numbers listed on the screen will automatically be chosen.

<F3>Save Screen Map - This option allows the user to save the screen image on the output diskette for later viewing with the Recall procedure included in the Main Menu.

<F4>View Key - This option will put the color key of specified size on the screen for viewing or printing.

<F5>Return to Main Menu - This option ends the map manipulation procedure, and puts the user back at the main menu level.

D: Recall screen file

As mentioned in the previous section, it is possible to save a map with the <F3> command, and then recall it from a diskette with this command. After the file name and disk drive is entered, the file will appear on the screen, and all map manipulation options mentioned in the previous section will again be possible.

E: Return to DOS

Selecting this option will end the LEXIS session and return the user to the DOS operating system.

ALGORITHM AND DESIGN DECISIONS

To more fully understand the shaded contour maps created by the LEXIS program, the user may find it helpful to understand the general nature of the algorithm used in LEXIS and to be aware of certain design decisions that we made over the course of the program's development. The program resulted from frequent interaction between a demographer with some knowledge of computers and a computer science student with a keen interest in demography. Working together to develop the program and simultaneously conducting some substantive research with it enabled us to make substantial improvements in the program in its eighteen months of development. This section outlines some of the more subtle features of the shaded contour maps created by the program.

It is important to first discuss the differences that exist between the LEXIS contour maps and the more conventional contour maps often used by geographers, architects, geologists, and others interested in studying surfaces. The most obvious difference is that we have chosen to shade the LEXIS maps to show the various levels of the surface, rather than to label the lines with numbers. This procedure allows us to indicate "cliffs" on a surface without drawing multiple contour lines closely spaced together. That is, sudden surface changes of more than one level are represented by a single contour line, with the rapid change in level indicated by a jump in color or shade between the adjacent regions.

Because the LEXIS program does no interpolation before creating a surface, the appearance of a contour map depends on the number of data points used in creating it. Maps created with few data points will tend to include strong rectangular patterns, whereas contours on maps made from larger arrays of data will appear more rounded. It is possible to change the appearance of the maps with a little creativity. Rounded maps will become more rectangular if enlarged and viewed in portions. Rectangular contours can be smoothed by interpolating the data to produce a larger array of data points.

The LEXIS program determines the location of contour lines by replacing actual data points with integers representing the tier, or region of height, into which the data point falls. With fifteen contour lines, there are sixteen tiers on the surface. Next, each point is evaluated in terms of its relationship to adjacent points. Various patterns among nine-point squares are included in the program, and each point, with the eight points that surround it, is checked for conformity with one of these patterns. If the actual pattern is so complex that there is more than one discontinuity, the middle point is surrounded by a box. We chose this simple method of drawing lines to be a reasonable compromise between the competing goals of minimizing computational time and maximizing the smoothness and accuracy of contour lines.

Because a surface depicted by the LEXIS program is based on a two-dimensional array of data, the surface can be thought of as a set of boxes, each box corresponding to a specific data point. The method used in LEXIS for drawing contour lines sometimes draws a line diagonally through a data box. The original data value for the box is always given by the value on the LEXIS map above the diagonal contour line.

For data pertaining to single years of time and age, each data box is one year by one year. The tick marks placed along the edges of the map correspond to the mid-points of the boxes that they describe. Thus, a tick mark at age zero, say, will appear offset slightly from the origin and a tick mark at year 1980, in a map that runs through 1980, will appear slightly before the end of the horizontal axis.

If you have any questions or suggestions concerning LEXIS, please write to one of us at the addresses given at the beginning of this paper.

References

Caselli, G., J.W. Vaupel, and Anatoli I. Yashin (1985) *Mortality in Italy: Contours of a Century of Evolution*. WP-85-24. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Vaupel, J.W., G.A. Gambill, and Anatoli I. Yashin (1985) *Thousands of Data at a Glance: Shaded Contour Maps of Demographic Surfaces*. Laxenburg, Austria: International Institute for Applied Systems Analysis. Forthcoming Research Report.

Zeng, Y., J.W. Vaupel, and Anatoli I. Yashin (1985) *Marriage and Fertility in China: A Lexis-Surface Analysis*. WP-85-70. Laxenburg, Austria: International Institute for Applied Systems Analysis.

APPENDIX 1: Example Data Map

In order to alleviate the frustration that often results from using an unfamiliar computer program, this appendix outlines the procedure for creating a LEXIS map of the U.S. fertility data contained on the LEXIS distribution diskette. We discuss, however, only basic steps and leave most of the options for users to learn at their convenience. Whenever we use the word "enter" we mean typing in whatever is requested and then hitting the <return> key.

Step 1: Beginning the LEXIS program:

Before beginning the procedure, create a work diskette by copying the files contained on the distribution diskette (LX.COM and USFERT) to a diskette formatted by your machine to include DOS. Put GRAPHICS on this diskette. Put this diskette into drive a: (the left side diskette drive). At the prompt type GRAPHICS; at the next prompt type LX.

Step 2: File Initialization

The data contained on the diskette are U.S. fertility rates and run from age of mother 14 to 49 and year 1945 to 1980. To initialize this data (i.e., to use this data to produce the kind of data file required by the LEXIS program), enter A from the main menu. The screen will then look like this :

File Initialization

Input File Disk Drive:
Input File Name:

Enter A as the input file disk drive, and USFERT as the input file name. Once this has been done, a prompt for the output file drive will appear, to which you should answer A.

Assuming all information has been entered correctly, the screen will now appear like this, with user responses included:

X-Axis information

	Data file	Initialized file
Starting point:	1945	1945
Ending point:	1980	1980
Label:	-	Year

Y-Axis information

Starting point:	14	14
Ending point:	49	49
Label:	-	Age

Once these numbers have been entered, the initialization process will put the year numbers on the screen as it processes the data, beginning with 1945 and ending with 1980. After the file has been completely initialized, the main menu will reappear.

The color pattern is set to the screen pattern by default, so we will ignore Main Menu option B and select C to create the LEXIS map. The Map Creation menu will now appear on the screen, giving the user the option to create a LEXIS map using one or two input files. Since there is only one data file on the diskette, enter A. The following input prompt will appear on the screen:

Initialized File Information

Initialized file prefix: ________

Disk drive: __

The initialized file prefix is again USFERT, and the disk drive is again A.

Next, the file dimensions will appear on the screen, in two places: on the top of the screen for the information of the user and on the bottom of the screen for the user to change as he or she wishes. We will look at the entire data set, so enter the <return> key at each prompt. Once this has been done, the question "Will actual data contain zero ?" will appear. Since U.S. fertility rates are zero at some older and younger ages, enter Y before continuing.

The next prompt will be: "Smooth the surface (Y/N) ?". Because this is the first analysis of fertility data, it may be most informative to look at a map of the actual rates first, without smoothing. Enter N at this prompt.

The last selection to be made is contour line levels. The Contour Line Menu contains seven options, with more than one appropriate in this case. However, option 2, "Fertility Lines", was designed specifically for maps of this data, so enter 2 for now. Once this last selection has been made, the map will begin to develop on the computer screen. It will take about one minute to complete, and will be entirely finished when a yellow border appears around the map.

After the entire map has been generated, pressing the <F1> key will reveal a text screen with line information, labeling and surface dimensions. Also, on the last line will appear a screen manipulation menu which outlines user options at this point in the mapping process. In order to see the color key used in creating the map, press <F4>. Because all 15 lines were assigned a different number, simply press <return> at the beginning and end point prompts. After the color key appears on the screen, hitting any key will return the user to the information screen. From here, press <F5> to end the process and return to the Main Menu.

This example should give the user a practical, but superficial, understanding of the LEXIS program capabilities. It is important that this example be used only as a supplement to, rather than a substitute for, the more extensive description contained in this manual.