

## Visualizing Lives: New Pathways for Analyzing Life Course Trajectories

Michael D. Maltz  
Department of Criminal Justice  
and  
Department of Information and Decision Sciences  
University of Illinois at Chicago  
1007 W. Harrison Street (M/C 141)  
Chicago, Illinois 60607-7140  
Voice: 312-413-2473  
Fax: 312-996-8355  
E-mail: [mikem@uic.edu](mailto:mikem@uic.edu)

and

Jacqueline Klosak-Mullany  
School of Public and Environmental Affairs  
Indiana University Northwest  
3400 Broadway  
Gary, Indiana 46408  
Voice: 219-980-6842  
Fax: 219-980-6737  
E-mail: [jklosak@iunhaw1.iun.indiana.edu](mailto:jklosak@iunhaw1.iun.indiana.edu)

*Acknowledgments:* The research for this paper was supported by Grant 95-BJ-CX-0001 from the Bureau of Justice Statistics, U.S. Department of Justice (Visiting Fellowship, "Development of Graphical and Geographical Methods of Analyzing Data"), as well as by a sabbatical from the University of Illinois at Chicago, to the first author. The opinions expressed herein are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. A version of this paper was presented at the Life History Research Society Conference, September 25, 1999, Kauai, Hawaii. The authors thank John Laub, Mindie Lazarus-Black, Joan McCord, Daniel Nagin, John Richters, Robert Sampson, and Leland Wilkinson for their comments on an earlier draft.

DRAFT: DO NOT CITE OR QUOTE WITHOUT PERMISSION FROM THE AUTHORS

## ABSTRACT

The goal of statistical analysis is to find patterns in data. Most statistical methods rely on analyzing the effect of the same set of variables on the population under study, i.e., *nomothetic* analysis. Therefore, when data are collected in the social sciences, most often they are put in a framework that resembles a spreadsheet: each row represents a separate individual, and each column represents a separate characteristic (or variable) that pertains to that individual.

However, not all individuals in the study are affected by the same set of variables: each individual may have his/her own individual set of relevant variables, suggesting that methods be developed that consider them individually, i.e., *idiographic* analysis. Moreover, lives are lived chronologically, and are often best described in narrative form. These narratives usually have to be condensed, or abridged in other ways, in order to fit the data framework and permit what one might call "algorithmic analysis." Each set of methods has its advantage: nomothetic methods generate general laws that apply to all, while idiographic methods trace the putative causal relationships that are unique to each individual.

This paper describes another data collection and analytic framework, one that (a) is chronological; (b) recognizes that different people may have experienced entirely different events and thus may need different "variables" to understand their behavior; (c) recognizes that, even if people experience similar events, they may have entirely different reactions to them; and (d) can be studied (and patterns inferred) using an exploratory graphical analysis that is more free-form than algorithmic analysis. Examples of this type of analysis used in different medical and criminal justice contexts are given, and suggested directions of research in this area are described.

**Key words:** longitudinal analysis, graphical analysis, life course, data visualization, exploratory data analysis

The overall goal of this paper is to describe an alternative means of collecting, coding, and analyzing life course data that permits a greater understanding of life course dynamics. The paper first describes the restrictions that the current array of analytic tools place on the collection of data, particularly life course data. It then describes an alternative data collection framework that has fewer such limitations. Subsequent sections describe how data collected using this alternative framework can be depicted and analyzed using visual data analysis techniques, and provides examples that have been reported in the literature. Examples of extensions of these techniques are then provided, and the paper closes with a list of attributes that a set of graphically-based analytic tools would need to have in order to accommodate the analysis of data using the suggested data collection framework.

## **1. Life Course Analysis**

Socialscience has as its primary goal the understanding of how and why people behave the way they do. Life course analysis is one of the means (and from our standpoint, one of the best means) of attaining this understanding, since individuals' behavior is conditioned on their earlier life experiences.

Individual behavior can be seen as a sequence of outcomes. Some of the outcomes may be due primarily to structural conditions in society, while others may be due primarily to the individual's initial conditions or subsequent experiences. In criminology, one school of thought (Gottfredson & Hirschi, 1990) attributes the outcomes largely to the individual's initial conditions; another school of thought suggests that initial conditions can largely be negated by subsequent experiences (e.g., Nagin & Paternoster, 1991; Sampson & Laub, 1993); and a third (e.g., Moffitt, 1993, Nagin, 1999) suggests that both schools are correct to some extent; i.e., that some individuals cannot change their behaviors easily while others can.

Regardless of the school of thought, similar techniques are used to investigate the relationship between behaviors and their correlates. A group of individuals is studied, a set of variables is specified, and a set of data is collected. Although we know that different things happen to different people over their life course, in order to analyze patterns we collect the same data for everyone. The reason for this is supposedly because we want to have a level playing field, to examine the characteristics of each individual on the basis of the same criteria. We suggest that this is not really the reason (or is not the sole reason), and that new tools need to be developed to examine individuals' lives in better ways.

For example, we know that in many cases the same things happen to different people, but each may have a different reaction to the same stimulus (Farrington, 1993; McCord, 1990, 1993; Maltz, 1994). That is, a divorce may adversely affect one family, while it frees another family from the stresses of an abusive parent and spouse. In the first case the family (or some of its members) may become dysfunctional, in the second case they (or some of them) may thrive. Or a move to a different neighborhood may improve one youth's life prospects but adversely affect another's. Yet patterns as complex as these may become masked by virtue of the analytic tools that are normally brought to bear on the data. This may be due in large part to the way data are collected so that these analytic tools can be applied, and on the types of data collected for analysis.

## **2. The Standard Data Collection Framework**

The standard framework for data collection in the social sciences is predicated on the use of hypothesis tests. The research question is usually cast as an inquiry into the effect of a specific variable (or set of variables) on an outcome (or set of outcomes). If the magnitude of the effect is sufficiently large that

it is unlikely to be attributable to chance, then the effect is said to exist, to be statistically significant.<sup>1</sup>

But hypothesis tests are rather crude analytic tools (Loftus, 1993; Maltz, 1994). In fact, consider the data framework in which these tests are applied. In general, a data base is analyzed; that data base contains a list of variables (the columns) and a number of "observations" of those variables, where each observation (i.e., each row) represents the value of each variable that obtains to each separate individual.

The term "observation" comes from the natural sciences, wherein outcomes of an experiment are observed and tabulated. Each observation may represent a replication of the experiment under both similar and different conditions, in order to determine the extent of variation in the outcomes and how they are affected by different conditions. For example, different observations of the pressure, temperature, and volume of a gas will lead a researcher to Boyle's Law. And Fisher, in *The Design of Experiments* (1935), observed the yield of different strains of barley seed in different years and at different sites to determine the relationship linking yield to site and seed type.

But observations are different in the social sciences. Observing people is not like observing gases or crop yield; the situation with respect to data collection in the social sciences is much more reflexive than in a physics or agricultural experiment (e.g., Briggs, 1986). Moreover, use of this term in the social sciences implies that there is only one mode of behavior, leading to a single "law," a single relationship between the dependent and independent variables – yet we know that people often react quite differently to the same stimulus.<sup>2</sup> And even when many individuals react similarly to the same stimulus, their reasons for doing so may be quite different from one another (Kagan, 1998: 76).

Although social science researchers recognize that different things happen to different people (which would almost seem to require a different set of variables for every individual in the data set), the number of variables is limited by the analyst: if the number of variables is not limited, the data set becomes so sparsely populated and its analysis so unwieldy and idiosyncratic that the analysis may not come up with (statistically significant) findings of any utility whatever. Thus the collection of data is usually restricted to only those variables that are common to most of the individuals under study, so that a sufficiently large number of individuals' records can be analyzed to provide statistically significant findings. This procedure has been used to great benefit in the social sciences, to uncover *nomothetic* relationships (i.e., general laws; *nomos* is Greek for law) among variables (Maxfield & Babbie, 1998: 49). However, it is also limiting; such studies may just "round up the usual variables" (Maltz, 1994: 451). In other words, the need for sufficient data restricts the view – and, we would suggest, the vision – of the researchers. In fact, the rectangular shape of most data sets (Figure 1) can be seen as a confinement of sorts – our assessment of people's lives is based only on the variables they hold in common, ignoring the very real (and important) individual differences they may have that do not fit in the restricted data frame.

[Figure 1 about here]

This situation exists in part because of the way the data are analyzed. As described earlier, standard analyses test hypotheses about the effect of variables on outcomes of interest. This is a *variable-based* analysis, in which the effect of the variable is assumed *a priori* to be univalent – that is, a weight is attached

---

<sup>1</sup>Statistical significance only make sense when the data represent a random sample, but the use of a random sample seems to be honored more in the breach than the observance (Maltz & Zawitz, 1998).

<sup>2</sup>For example, when considering the effect of divorce or of residential mobility on youths, a single coefficient is obtained for each variable, masking the fact that some are removed from risky situations and some are put at greater risk.

to that variable that is either positive or negative, implying that the variable either increases or decreases the outcome. However, as we noted earlier, effects are not that straightforward in studying human behavior.

Others (e.g., Magnussen, 1990; Kagan, 1997) have written about the advantages of *person-based* analyses over variable-based analyses in developmental research. Person-based analyses focus on developmental sequences and chronology, which can show the effect of not just the variables but how they are sequenced over time. It is possible to employ the standard data framework to investigate developmental and chronological aspects of life courses; for example, Nagin (1999) describes an innovative procedure developed to deal with time-sequenced data. This method is predicated on the same data elements being collected repeatedly from each individual in the study, again in the search for general laws of human behavior. But even person-based analyses may fall short of their potential if consideration is not given to:

- permitting variables to have multivalent effects,
- lifting the restriction on the number of variables on which data are collected, and
- retaining the narrative aspects of the collected data.

A person-based analysis that considers these factors requires a different framework for collecting data. We propose the following alternative data collection frame that takes these factors into consideration.

### 3. An Alternative Data Collection Framework

When people fill out forms or answer structured questionnaires, they provide information on variables that have been selected according to the criterion mentioned earlier – variables that are relevant to a reasonably large number of subjects. They would ordinarily not include all of the factors or incidents that were important in each individual's life, because so many of them may in fact be idiosyncratic. Yet even in structured interviews people may give narratives and tell stories, with all the "thick description" (Geertz, 1973: 6) and contextualization that that implies. Much of this information may be lost, because it does not fit into the restrictive data frame based on common information.

This need not be the case. One can envision an alternative data framework that augments common information with subject-defined variables based on the narratives given by the subjects. It is not suggested that all such information will be useful, or that the subjects all know what the salient events in their lives were, or that they will trust the interviewers sufficiently to provide such information; however, insofar as such information is obtained, one can develop an alternative data collection framework that would use subject-defined variables based on these narratives.

The narratives will, of course, all be different. Moreover, the *number* of different events that occur to different people may vary considerably: Mike may have led a relatively placid existence while Jackie's life may have a great number of significant turning points.

The question then becomes how to achieve the goal described earlier: to obtain an understanding of how people are affected by their experiences. We suggest that this alternative data collection framework can be used to draw *individual time line trajectories* showing life course events, based on the narratives. Events and processes would be arrayed on the time line, a separate time line for each individual in the study. This has a number of advantages: First, it portrays events in their chronological sequences, which often suggests possible causal connections. Second, with each person's time line(s) being portrayed individually, the fact that some lives are more complex than others can easily be seen and grasped. Third, it not only shows the sequencing of events (which event came first) but *when* (the date) the events occurred, which may be important from three standpoints:

- a developmental standpoint: child abuse takes on a different importance if it occurred at age 3 or at age 15;
- a historical standpoint: a person's unemployment takes on a different meaning if it was during a period

- of prosperity or during an economic downturn; and
- a cohort standpoint: those born in the 1930's and experienced unemployment at age 20 will have different trajectories than those born in the 1940's and experienced unemployment at age 20.

Thus, it may be possible to distinguish age and period and cohort effects in this kind of framework. Moreover, it is much easier to deal with the effect of event sequencing in this framework than using standard analytic methods.

The framework within which these kinds of data are collected is quite different from the standard data collection framework. Rather than there being a column for every variable and row for every respondent (Figure 1), it includes a separate *page* of data (as in a spreadsheet) or *table* of data (as in a database) for every subject. The page has a row for every relevant event or process (e.g., date of birth, school, a new job). The first column contains the dates of the events or processes, and subsequent columns contain descriptors of the events and assessments of their impact; an example is shown in Table 1. The number of entries for an individual depends upon the amount of activity the individual generated over his/her life course, and not upon the number of variables selected by the researcher as being important. [In fact, sufficient information about these events may already exist in the data collected as part of a study, but not used as coders try to fit the rich information into the restrictive standard data collection framework.]

[Table 1 about here]

This type of analysis is termed *idiographic* analysis (Maxfield & Babbie, 1998: 49), since it looks at each individual's trajectory separately to infer causal relationships rather than across all individuals. The problem with idiographic analyses, however, is that it is difficult to generalize from them. Many believe that nomothetic methods are the best way to gain an understanding of a process, and cite Meehl (1954) as justification; they interpret his primary finding to be "statistical prediction trumps clinical prediction."

However, as he noted (Meehl, 1973: 83), there are times when we "should use our heads instead of the formula," when the algorithm is not as good as the brain. Granted, when the task at hand is applying a known algorithm to a set of data in the service of prediction or classification, the human brain can't compete with the computer – or, more generally, with a mechanical application of the algorithm. But when the algorithm is unknown, the tables are turned (Richters, 1997: 224). We suggest that this may be the case in life course analysis, and that computer-based visual techniques are more suited to the analysis of life course trajectories than statistical algorithms.

That is, computers can be used to do more than just compute. They can be used to organize data in ways that make it easier to discern patterns visually; i.e., by using them as "power steering" devices, where the driver decides on the route, rather than as "autopilots," where the decision-making is taken from the driver's hands (Maltz, Gordon & Friedman, 1991: 46). The next section describes different techniques currently used to analyze data visually. Instead of an approach that is geared toward algorithmic analysis and *testing* hypotheses, this framework is geared toward exploratory data analysis (Tukey, 1977), specifically graphical analysis, and the *generation* of hypotheses.

#### 4. Visual Data Analysis

We are accustomed to using computer algorithms to analyze data, to provide us with specific answers (i.e., whether a variable is or is not statistically significant). But data can be analyzed visually as well, to infer patterns of a different sort.

Graphical analysis has its roots in the work of Playfair's *Political Atlas*, published in 1786 (Wainer, 1997). His techniques were brought into the (early) computer age by Tukey in his classic book, *Exploratory*

*Data Analysis* (1977). Tukey distinguished between confirmatory data analysis – of which significance testing is a primary example – and exploratory data analysis, in particular, the depiction of data in a way that "forces us to notice **what we never expected to see**" (p. vi, emphases in the original). Although his book is somewhat dated – he devotes a great deal of attention to pencil-and-paper methods, many of which have become superfluous since the widespread use of computers – it contains a great number of examples of the use of graphical methods to find patterns in data.

Six years later, Tufte approached the subject from the other side. Instead of looking at how data might be graphed, he concerned himself with how graphs of data convey information. In *The Visual Display of Quantitative Information* (1983), he criticized the way data were displayed in graphs; he coined the term "chartjunk" to refer to graphs that used kitschy symbols to draw attention to them, but that were basically "data-thin." His two subsequent books (Tufte, 1990, 1997) reinforced this concern and provide additional examples of graphing techniques that enhance the understanding of data..

Cleveland moved the field of graphical data analysis a few steps further. In his books, *Visualizing Data* (1993) and *The Elements of Graphing Data* (1994), he provided detailed examples of how data might be explored visually as an analytic tool. The use of some of these techniques can be found in Maltz (1998).

Cleveland (1993: 328) provides a telling example that shows the benefit of graphical analysis over standard statistical analysis. He used a data set that was first analyzed by Fisher in the 1930s (on the yield of different strains of barley in different sites in 1931 and 1932) and subsequently found its way into a number of statistics texts. However, *none of the earlier statisticians actually plotted the data* until Cleveland did so – and he found that someone had mistakenly switched the 1931 and 1932 data for one of the sites! Although we do not expect to uncover similar problems with life history data, it does point out the benefit of actually looking at the data before analyzing them with a statistical algorithm.

Two general strategies have been used for depicting life course trajectories, depending on the number of variables depicted in each trajectory. When there are only a few variables, many trajectories can be plotted simultaneously; however, when many variables are to be plotted (in the examples given below, more than three), only one trajectory can be plotted in a single chart.

#### 4.1 Few Variables

Three examples of trajectory depiction are given. Two are from the medical literature and are based on survival analysis, the third describes family dynamics.

Goldman (1992) uses "eventcharts" to display survival data; the data were taken from a bone marrow transplant database to study the prevention of cytomegalous virus infection. In Figure 2 the vertical axis represents date of an intervention, in this case a bone marrow transplant. The further up the vertical axis the line starts, the more recently the transplant took place. The horizontal axis represents time since intervention; the length of each line represents the length of time each individual was observed (or stayed alive – the vertical bar at the end of a horizontal line represents time of death). To give some perspective to the time available for observation, a "now" line (the diagonal dashed line) indicates the extent to which survival data are available – those who entered the treatment program later in time have a shorter maximum follow-up time. Markers on the line represent characteristics of the patient's medical history; the open circles represent when Graft-Versus-Host Disease (GvHD) occurred. Goldman (1992:14) notes that "GvHD occurs early and although a serious complication, is seldom followed by relapse or death;" as can be seen in the figure, most of the trajectories with open circles traverse the chart all of the way to the now line, indicating that they did not die, at least during the time they were observed. The closed squares in the figure represent time of relapse.

[Figure 2 about here]

Lee, Hess, and Dubin (forthcoming) extend Goldman's concept and display a number of different ways of representing such data. Using a different data set (patients with head and neck cancer), they give examples (Figure 3) of displaying life course trajectories of patients, but rearrange them ("stack" them) according to (a) the patient ID, (b) their date of registration (with a "now" line), (c) their total exposure time, and (d) the presence (dashed line) or absence (dotted line) of an important indicator, the p53 protein. The open circles represent the date of registration in the program (start time) and the open squares represent the date of last follow-up. Note that, unlike the data in the previous figure, some people were lost to follow-up before the end of the observation period – they may have moved or even died (from some other cause, like a traffic accident). An X represents time of death, a filled triangle the date of recurrence, and a filled circle the date of second primary tumor. The software that generates this type of display, "event.chart," is written in the language S-Plus and archived at <http://lib.stat.cmu.edu/S/>.

[Figure 3 about here]

Another type of life course trajectory is based on the Lexis plot, named after its creator, the 19th century German demographer Wilhelm Lexis. Francis & Fuller (1996) developed software (also written in S-Plus) to plot the life course trajectories of individuals. Examples can be found at the URL <http://www.cas.lancs.ac.uk/alcd/visual/>. A figure at this website depicts "Lexis pencils" showing the employment history of 188 married couples in Kirkcaldy, Scotland, as a function of the presence of children of varying ages. Figure 4 is a black-and-white perspective representation of one of their color figures, showing only four of the 188 trajectories.<sup>3</sup>

[Figure 4 about here]

The trajectories are canted at 45 degrees from the horizontal, and the base plane is "anchored" according to the woman's age at marriage. The (Z) axis coming toward the reader is the year of marriage, the (X) axis moving to the right is age of the woman, and the vertical (Y) axis is time since marriage. For example, the trajectory "closest" to the reader is of a couple who married in 1969 when the woman was about 19 years of age. Five years later (moving vertically five years) she was 23 years of age, so the trajectory has moved diagonally, up five years and to the right five years.

Each trajectory has three facets, like the three visible faces of a hexagonal pencil, which is why the trajectories are called "Lexis pencils." The upper facet represents the employment history of the husband and the middle facet the employment history of the wife: light shading represents working, dark shading unemployed. The third (lowest) facet represents the age of the youngest child; the first change in color (shading) occurs upon that child's birth. Note that the woman in the closest trajectory stopped working after about a year of marriage (middle facet), followed soon afterwards by the birth of their first child (bottom facet).

The full figure, not included here (it can be found at the cited URL), shows how the employment of women has changed over the forty years represented in the figure, and how more women returned to work (and returned earlier) after giving birth in the more recent years. That figure can also be rotated to show different perspectives on the data. Although this finding is hardly surprising, such age, period and cohort

---

<sup>3</sup>Since color is essential to understanding this figure, the reader is encouraged to access the cited website.

behavior cannot be characterized easily using common statistical methods.

These three examples show that it is possible to plot many trajectories in a single plot when only a few variables are included in/on the trajectories. Both symbols and line weights can be used to portray variables, depending on whether they represent events or long-term processes. Trajectories can be stacked in different ways, which can provide insight into relationships. In addition, Figure 4 shows that different domains can be portrayed in a single trajectory by including different "faces" on the trajectory line.

## 4.2 Many Variables

Three examples of trajectories with many variables are given. One is from the medical literature and two are from the social sciences.

Cohen (1999) shows how a number of variables representing different domains of an individual's life can be portrayed in a single figure (Figure 5). Each panel represents a different individual, and the figure depicts the extent (0 to 100 percent) to which each individual has made the transition to adulthood and independence in the depicted domains. The figure shows how such variables can be portrayed over the life course, permitting comparisons of the individuals' development and permitting inferences as to how the variables interact in different individuals.

[Figure 5 about here]

Post, Roy-Byrne and Udhe (1988; see also Leverich & Post, 1993) use graphical techniques to analyze life course data of patients with affective disorder (Figure 6). The authors point to the benefits of using graphic representation of the life course of patients, such as more accurate tracking and the ability to identify which combination of factors (medication, treatment, contextual factors, etc.) has been effective in dealing with the disorder throughout the life of a patient.

[Figure 6 about here]

There are three levels of annotation on the graph. The relevant **events** that occurred during the patient's life course are listed *beneath the time line* (which is segmented to permit it to fit in the journal article); the **symptoms** are arrayed *on the line*, manic episodes above the line and depressive episodes below the line; and the levels and durations of the **treatments** are given *above the line*. Although the figure appears complicated, it has been found useful in organizing the information about life course events, symptoms that may have been produced by these events, and treatments that may (or may not) have ameliorative effects on the symptoms. Thus, the data are organized in a way that preserves the chronological order of the events, permitting tentative hypotheses to be generated about events that may have triggered episodes and the effectiveness of different treatments.

An initial attempt to plot a criminal justice life course trajectory similar to that in Figure 6 is shown in Figure 7 (from Maltz, 1995). This figure is based on information obtained from one source, narratives written by Cook County (Illinois) juvenile probation officers in the case jackets of their charges. It uses rather simple symbols to depict different events in a youth's life. In places where symbols were inadequate to describe the events, a text box was used to convey the necessary information. While this figure is admittedly crude – it would benefit from better symbols and/or icons, varying line weights, color, and other graphical embellishments – it depicts patterns in a way that would not be possible with the methods and data that are normally used in studying delinquency. Little detail is shown in Figure 7, because it shows data only from a

single source of information, Cook County Juvenile Court records.

[Figure 7 about here]

## 5. Extensions of these Methods

As Henry (1999) noted, there appears to be a tradeoff in these trajectory depictions between being extensive (in terms of number of subjects depicted) and intensive (in terms of number of variables for each subject). This may appear to be a limitation to the use of graphical methods; however, we think that one can overcome this limitation to a great extent. One can "squeeze down" the portrayal of complex individual trajectories so that many of them can be stacked on the same figure or computer screen, to permit a search for patterns and commonalities among the different trajectories. This is the direction we have been moving toward in this research.

Our initial attempts to develop graphical data displays are presented. It should be noted that these are preliminary steps in developing a system for presenting life course data. One of the strongest differences between our explorations and those already presented is the number of different event types we portray, related to different domains of the subject's life. Our ultimate goal is to squeeze these multiple-trajectory life course pictures down to portray all information on a single trajectory (perhaps with more than one facet, as in Figure 4), so that many trajectories can be depicted and compared on the same chart or computer screen. In this way we hope to be able to spot patterns in the subjects' lives, patterns that are undetectable using standard variable-based techniques.

For example, Klosak (1999) uses different icons, colors, and other graphical symbols in depicting life course data from field notes obtained through personal interviews with 15 adult women on probation in Cook County (Chicago) Illinois. Her analysis involves the development of ways of portraying life course trajectories of each woman through graphical presentations of all the relevant domains of their lives.<sup>4</sup> Figure 8 is based on a life course trajectory depicted in color in Klosak (1999). This means of presenting the data allowed all of the variables to be seen at one time and highlighted the sets of factors, and their time order, that influence a woman's life choices.

[Figure 8 about here]

Note the many different icons used, with different levels representing different domains of activity and different colors (in this case, shades) representing varying impacts of these events (favorable or unfavorable). For example, educational achievement is represented by a scroll: positive aspects are represented by white scrolls and negative aspects (e.g., moving from one school to another in the middle of the school year, dropping out) are represented by gray scrolls. Good relationships are depicted with normal white hearts, abusive relationships with inverted grey hearts. The domains represented in this graph include family, education, employment, relationships, children, alcohol and drugs, and criminal justice involvement.

Figure 9 uses the same set of icons as Figure 8. "Christy's" present status appears to be reflective of activities outlined on the drug and family trajectories. She began taking drugs at an early age (seven); both parents used drugs. At various stages of her life, she experienced losses of those close to her. As indicated

---

<sup>4</sup>Since the events are based on interviews, it cannot be expected that the interviewees (a) remember and (b) report all events in all relevant domains of their lives. In other words, the extent of the data portrayed depends on the memory and truthfulness of the respondents, their understanding of the nature of the inquiry, and their ability to respond appropriately to it.

by the black down arrows, she has had a lot of stress in her life: her parents separated when she was an infant, her grandmother died when she was a young girl, her mother was hospitalized in March 1987, her brother was sentenced to jail in May 1987, her sister was killed in June 1987, and her mother died in July 1992. Further review of her chart also shows that since she has been on probation, she has demonstrated signs of progress. She has obtained her first job and, perhaps most importantly, she has refrained from further substance abuse.

[Figure 9 about here]

Analyzing the data in this manner can reveal the contributing factors associated when a client is either heavily involved or least involved in criminal activity. In this way, one can better understand the interrelations of the different aspects of their lives and how these aspects may have affected the decisions that ultimately led to their current criminal status. Although the figures are fairly complicated, this is not necessarily a problem. As Tufte (1990:37) says, "Simplicity of reading derives from the context of detailed and complex information, properly arranged. A most unconventional design strategy is revealed: *to clarify, add detail*" (emphasis in the original).

These two life courses can be compared by confining their respective trajectories to a thin horizontal strip and "stacking" them on the same page. This is shown in Figure 10. An additional trajectory is included for comparison, one depicting an individual who experienced what one might consider "ideal" life course events. Code-named "Barbie," this trajectory includes grade school, middle school, high school, and college; few moves; use of marijuana during college; a few relationships, the last resulting in marriage and two children; and employment after college (including a change of employers) and until the first birth.

[Figure 10 about here]

Macroscopically, we can contrast the relatively spare trajectory of Barbie with the rather congested trajectories of Megan and Christy. Moreover, the nature of the events is considerably different, suggesting that an analysis using the same variables for each of the subjects might miss important phenomena in each person's life.

We recognize that these figures are inadequate; they should be taken more as representative of what might be done than as a finished product. First, the figures are in black and white, since this journal does not print in color. Second, the figures were created by a spreadsheet program (Excel) that has limited graphics capability and a limited set of ready-made symbols—some of the symbols were drawn by the authors. Third, the symbols that represent different variables are not intuitive, and there is no control over which ones are on top when they overlap. Fourth, long-term processes (such as parenting or education), which might be represented by lines or horizontal bars whose thickness and/or color change over time, could not be represented in this way in Excel.

Our expectation is that a mature graphical analysis tool, one more fully developed than this, would produce a number of benefits.

- After the initial interview, it could be shown to interviewees to give them an idea of how their own lives have been shaped, and might elicit more information from them about other salient events and processes in their lives.
- By putting all (or much) of the information in a readily accessible form, it would also help the analyst in understanding all of the many events that shape a person's life.
- In encapsulating events from so many domains of a person's life, it may assist in ferreting out hidden patterns, for example, whether the nature of stressful events is more important than their intensity,

their number or their timing.

- It could be of benefit in developing new categories of individuals, in terms of behavioral responses that are not strictly linear and amenable to existing algorithms, and may also help in the development of new algorithms to study such behaviors. In fact, a visual display of data "offers us a way to do perceptual clustering" (Wilkinson, 1999: 111).

In other words, we believe that the figures do show the promise inherent in graphical representation of life courses. They also suggest the directions that must be taken to develop an analytic tool that can be used to understand life course trajectories.

## 6. Requirements for a Graphical Tool for Life Course Analysis

This brief review of the pros and cons of the various means of depicting trajectories suggests the following desiderata for a new set of graphical analysis tools:

- The **data frame** used to collect and store the data should have a separate page or table for each individual. Each page/table should have separate rows for each event or process, with as many columns as needed to describe that individual's events. The events themselves may have a number of descriptors associated with them (e.g., location, color, size, etc.) that can be detailed in different columns.
- There should be a fairly large **repertoire of symbols** to represent different phenomena. These symbols should be readily interpretable; i.e., they should bring to mind the type of events they represent. The symbols collected by Liungman (1994) are a possible starting point, since they are in some sense standardized. Standardization makes interpretation easier; for example, in regression analyses we interpret the  $x_i$  as independent variables and the  $\beta_i$  as coefficients, and would be confused were we to read a paper that reversed their meaning. The tools should be sufficiently versatile to allow an investigator to create new symbols as needed and incorporate them into the analysis.
- Since the events represented by the symbols may have different meanings for different subjects, they should be distinguished by **color and size** as to their importance and valence, i.e., whether they have a positive or negative effect – in fact, graphing permits one to rethink causal relationships and determine the effect of different events *retrospectively*.
- Not only should **events** be depicted, but **processes** as well. For example, educational attainment is not just a static outcome but a process that may be affected by family dynamics, residential mobility, peer group involvement, or other events in an individual's life. To depict such processes, the lines themselves should take on different colors, thicknesses, and textures, and it should be possible to vary these characteristics over the life course. That is, if an individual begins to fall behind in school, the line representing educational attainment may change color or thickness.
- **Different domains** of a person's life should be represented on the same graph. Rather than doing so by separating the domains vertically (as in Figure 8), they should be packed together (as in the "Lexis pencils" of Figure 4), in order to permit the comparison of many trajectories simultaneously.
- It should be possible to **stack different trajectories** in the same graph, and use different criteria to determine the order of the stacking (as in Lee *et al.*, forthcoming). For example, it should be possible to change the horizontal axis from calendar time to time since birth or time since some other event (marriage, as with Francis & Fuller, 1996, or birthdate, or release from custody, etc.).
- Since some events are fairly complicated, it should be possible to describe their characteristics by attaching a **text box** to the symbols representing them, that can be opened by clicking on them (or

by passing the cursor over them). This, therefore, implies that the methodology should be dynamic, i.e., should be **interactive**. That is, not only should we go beyond numbers, tables, and text to present (and represent) data, but we should go beyond paper as well.

- It should be possible to "explode" or magnify the trajectories either horizontally (to inspect a sequence of closely-packed events) or vertically (to inspect the different life course domains individually). This also implies an interactive user interface.

Although it was not known at the time this paper was originally written, a product that includes virtually all of these features is now available. Wilkinson (1999) and his colleagues have developed a software package, Graphics Production Language, in Java that can be used to depict processes and events on individual trajectories. We have not yet employed it in depicting and analyzing trajectories, but feel that it will prove to be a very useful adjunct to the person-based, idiographic analytic approach described in this paper.

## 6. Conclusion

The goal of statistical analysis is to find patterns in data. Most statistical methods rely on analyzing the effect of the same set of variables on the population under study, i.e., *nomothetic* analysis. However, not all individuals in the study are affected by the same set of variables: each individual may have his/her own individual set of relevant variables, suggesting that methods be developed that consider them individually, i.e., *idiographic* analysis. Each set of methods has its advantage: nomothetic methods generate general laws that apply to all, while idiographic methods trace the putative causal relationships that are unique to each individual. Our goal has been to explore how (or whether) the best features of these two different ways of looking at data can be combined.

This paper provides a rationale for and some desired characteristics of a new methodology for analyzing data in non-traditional ways. In particular, the new methodology focuses on suggested requirements for analyzing life course data, in an attempt to combine nomothetic and idiographic methods. It may be argued that some lives are so complicated that the amount of detail would swamp anyone's attempt to depict those trajectories, and that some measure of selective editing needs to be done, but it is premature to make such a judgment. As a counterexample, consider the amount of information conveyed by a modern map, showing, in different colors and overlays, the topography, location and approximate size of cities, jurisdictional boundaries, altitude, points of interest, land use, and other useful information as well.

Old coding procedures seem to squeeze the juice out of personal histories in an effort to collect comparable data. One has the feeling, even after reading some prize-winning analyses of life courses, that the researchers wished that there were other, more appropriate ways of handling individual life histories and finding patterns, but they were constrained by a set of algorithms that didn't fit the type of data they had. They did the best they could, but in the end had to distill their data to fit into the strait-jacket of the rectangular data frame.

It may be said that the methods outlined herein are somewhat crude and limited. It was suggested that using graphical techniques to analyze and compare more than 20 trajectories may not be possible. This may in fact be the case; however, once these techniques are in use, it may well be that newer methods of analyzing the data are developed, in much the same way that LISREL and HLM and other statistical methods have been added to the repertoire of the social scientist.

This is more than just a new way of collecting data; one has to reconceptualize from collecting "data" to collecting stories. It also means that we have to unlearn some of the strictly ingrained (variable-based) habits of data analysis, and to permit consideration of different ways of finding patterns in data. In sum, we feel that this exploration into graphical analysis of life course data is an approach well worth the effort. We have just begun to scratch the surface in moving toward the realization of Wild's (1994, p. 168) prediction,

that in the future "the primary language for promoting human understanding of data will be sophisticated computer graphics rather than mathematics."

## 7. References

- Briggs, Charles L. (1986). *Learning How to Ask: A Sociolinguistic Appraisal of the Role of the Interview in Social Science Research*. Cambridge University Press, Cambridge, England.
- Cleveland, William S. (1993). *Visualizing Data*. Hobart Press, Summit, New Jersey.
- Cleveland, William S. (1994). *The Elements of Graphing Data*. Hobart Press, Summit, New Jersey.
- Cohen, Patricia (1999). Presentation materials prepared for the Life History Research Society Conference, Kauai, Hawaii, September 22-25, 1999.
- Farrington, David P. (1993). "Interactions between Individual and Contextual Factors in the Development of Offending." In Michael D. Rutter, Ed., *Studies in Psychosocial Risk: The Power of Longitudinal Data*. Cambridge University Press, Cambridge, England.
- Fisher, Ronald A. (1935). *The Design of Experiments*. Oliver and Boyd, Edinburgh, Scotland.
- Francis, Brian, and Mark Fuller (1996). "Visualization of Event Histories." *Journal of the Royal Statistical Society A*, Vol. 199, Part 2, pp. 301-308.
- Geertz, Clifford (1973). *The Interpretation of Culture*. Basic Books, Inc., New York, NY.
- Goldman, Anne I. (1992). "Eventcharts: Visualizing Survival and Other Timed-Events Data." *The American Statistician*, Vol. 46, No. 1, February, pp. 13-18.
- Gottfredson, Michael R., and Travis Hirschi (1990). *A General Theory of Crime*. Stanford University Press, Stanford, California.
- Henry, David (1999). Presentation at the Methods Workshop, Life History Research Society Conference, Kauai, Hawaii, September 23, 1999.
- Kagan, Jerome (1997). "Conceptualizing Psychopathology: The Importance of Developmental Profiles." *Development and Psychopathology*, 9, 321-334.
- Kagan, Jerome (1999). *Three Seductive Ideas*. Harvard University Press, Cambridge, MA.
- Klosak, Jacqueline (1999). *The Course of their Lives: Women Offenders on Probation*. Unpublished Ph.D. dissertation in Public Policy Analysis, University of Illinois at Chicago.
- Lee, J. Jack, Kenneth R. Hess, and Joel A. Dubin (forthcoming), "Extensions and Applications of Event Charts." *The American Statistician*.

- Leverich, Gabriele S., and Robert M. Post (1993). "The NIMH Life Chart Manual for Recurrent Affective Disease: the LCM." Biological Psychiatry Branch Monograph, National Institute of Mental Health.
- Liungman, Carl G. (1994). *Dictionary of Symbols*. Norton, New York, NY.
- Loftus, Geoffrey R. (1993). "A Picture Is Worth a Thousand p-Values: On the Irrelevance of Hypothesis Testing in the Microcomputer Age." *Behavior Research Methods, Instruments, & Computers*, Vol. 25, No. 2, pp. 250-256.
- Magnusson, David, and L. R. Bergman (1990). "A Pattern Approach to the Study of Pathways from Childhood to Adulthood." In Lee N. Robins and Michael Rutter, Eds. (1990). *Straight and Devious Pathways from Childhood to Adulthood*. Cambridge University Press, Cambridge, England.
- Maltz, Michael D. (1994). "Deviating from the Mean: The Declining Significance of Significance," *Journal of Research in Crime and Delinquency*, Vol. 31, No. 4, pp. 434-463.
- Maltz, Michael D. (1995). "Criminality in Space and Time: Life Course Analysis and the Micro-Ecology of Crime," in John Eck and David Weisburd, Eds., *Crime And Place*, Criminal Justice Press, Monsey, New York.
- Maltz, Michael D. (1998). "Visualizing Homicide: A Research Note." *Journal of Quantitative Criminology*, Vol 15, No. 4, pp. 397-410.
- Maltz, Michael D. and Marianne W. Zawitz (1998). *Displaying Violent Crime Trends Using Estimates from the National Crime Victimization Survey*. Bureau of Justice Statistics Technical Report NCJ 167881, June 1998.
- Maxfield, Michael G. and Earl Babbie (1998). *Research Methods in Criminal Justice and Criminology*. Second Edition, Wadsworth, Belmont, California.
- McCord, Joan (1990). "Long-Term Perspectives on Parental Absence." In Lee N. Robins and Michael Rutter, Eds., *Straight and Devious Pathways from Childhood to Adulthood*. Cambridge University Press, Cambridge, England.
- McCord, Joan (1993). "Descriptions and Predictions: Three Problems for the Future of Criminological Research." *Journal of Research in Crime and Delinquency*, Vol. 30, pp. 412-425.
- Meehl, Paul E. (1954). *Clinical Versus Statistical Prediction: a Theoretical Analysis and a Review of the Evidence*. University of Minnesota Press, Minneapolis, Minnesota.
- Meehl, Paul E. (1973). *Psychodiagnosis: Selected Papers*. University of Minnesota Press, Minneapolis, Minnesota.
- Moffitt, Terri E. (1993). "Adolescence-Limited and Life-Course-Persistent Antisocial Behavior: A Developmental Taxonomy." *Psychological Review*, Vol. 100, No. 4, 674-701.

- Nagin, Daniel S. (1999). "Analyzing Developmental Trajectories: A Semiparametric, Group-Based Approach." *Psychological Methods*, Vol. 4, No. 2, 139-157.
- Nagin, Daniel S., and Raymond Paternoster (1991). "On the Relationship of Past and Future Participation in Delinquency." *Criminology*, Vol. 29, pp. 163-190.
- Post, Robert M., Peter Roy-Byrne, and Thomas W. Uhde (1988). "Graphic Representation of the Life Course of Illness in Patients with Affective Disorder." *American Journal of Psychiatry*, Vol. 145, No. 7, July, pp. 844-848.
- Richters, John E. (1997). "The Hubble Hypothesis and the Developmentalist's Dilemma." *Development and Psychopathology*, 9, 193-229.
- Sampson, Robert J., and John Laub (1993). *Crime in the Making: Pathways and Turning Points through Life*. Harvard University Press, Cambridge, Massachusetts.
- Tufte, Edward R. (1983). *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut.
- Tufte, Edward R. (1990). *Envisioning Information*. Graphics Press, Cheshire, Connecticut.
- Tufte, Edward R. (1997). *Graphical Explanations*. Graphics Press, Cheshire, Connecticut.
- Tukey, John W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, Massachusetts.
- Vaupel, James W., James R. Carey, Kaare Christensen, Thomas E. Johnson, Anatoli I. Yashin, Neils V. Holm, Ivan A. Iachine, Väinö Kannisto, Aziz A. Khazaeli, Pablo Liedo, Valter D. Longo, Yi Zeng, Kenneth G. Manton, and James W. Curtsinger (1998). "Biodemographic Trajectories of Longevity." *Science* Vol. 280, pp. 855-860. See also James W. Vaupel, Wang Zhengliao, Kirill F. Andreev, and Anatoli I. Yashin (1998). *Population Data at a Glance: Shaded Contour Maps of Demographic Surfaces over Age and Time*. Odense University Press, Odense, Denmark.
- Wainer, Howard (1997). *Visual Revelations: Graphical Tales of Fate and Deception from Napoleon Bonaparte to Ross Perot*. Copernicus (Springer-Verlag), New York, New York.
- Wild, C. J. (1994). "Embracing the 'Wider View' of Statistics." *The American Statistician*, Vol. 48, pp. 163-171.
- Wilkinson, Leland (1999). *The Grammar of Graphics*. Springer-Verlag, New York, New York.

## "Megan's" Domains and Events

Dates	DOB	Siblings' DOB	Change of Guardian	Moves	Grade Schools	Sister's Death	Relationships	Childrens' DOB	High Schools	Use of Alcohol	Use of Marijuana	Use of Cocaine	Employment	Children Imprisoned	Date of Probation
05/04/61	0.25														
08/24/95															4.25
05/04/61			0.75												
03/02/79			0.75												
03/02/85			0.75												
03/02/89			0.75												
03/01/49		0.5													
04/01/50		0.5													
05/01/53		0.5													
06/01/55		0.5													
04/01/56		0.5													
07/01/57		0.5													
08/01/58		0.5													
09/01/62		0.5													
05/01/64		0.5													
06/01/66		0.5													
07/01/68		0.5													
05/01/78															
05/01/70						1.5									
05/04/61				1											
05/02/66				1											
03/02/79				1											
03/02/85				1											
03/02/89				1											
04/02/95				1											
09/01/68					1.25										
06/07/70					1.25										
09/01/70					1.25										
06/07/73					1.25										
09/01/73					1.25										
10/04/74					1.25										
10/05/74					1.25										
06/07/75					1.25										
09/01/75									2.25						
12/15/78									2.25						
03/01/97													3.75		
05/02/97													3.75		
10/03/94													3.75		
01/01/95													3.75		
05/01/95							1.75								
12/31/97							1.75								
05/01/92							1.75								
05/05/92							1.75								
02/01/82							1.75								
08/01/82							1.75								
03/01/74							1.75								
12/01/78							1.75								
06/25/75								2							
09/01/78								2							
03/01/83								2							
04/01/95														4	
12/31/97														4	
04/05/96														4	
10/04/97														4	
06/01/86										3					
06/27/95										3					
06/01/86											3.25				
06/27/95											3.25				
05/01/88												3.5			
06/27/95												3.5			

Numbers in the cells represent the y-values of the graphed variables. For example, "Megan" used cocaine between May 1988 and June 1995, and this is represented in Figure 8 as a line 3.5 units above the x-axis.

**Table 1:** Example of a Chronological Data Collection Framework

1991 U.S. General Social Survey - SPSS for Windows Data Editor

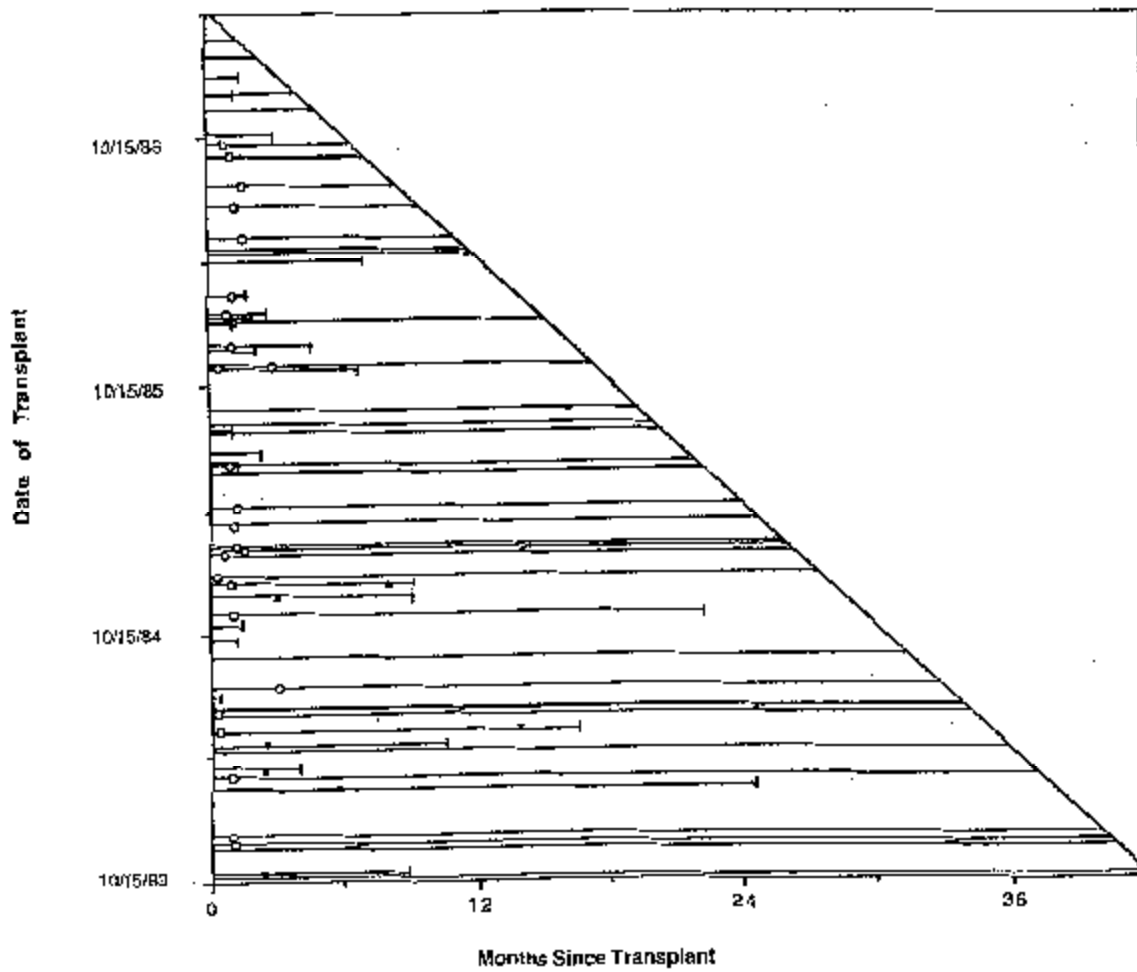
File Edit View Data Transform Analyze Graphs Utilities Window Help

1:sex

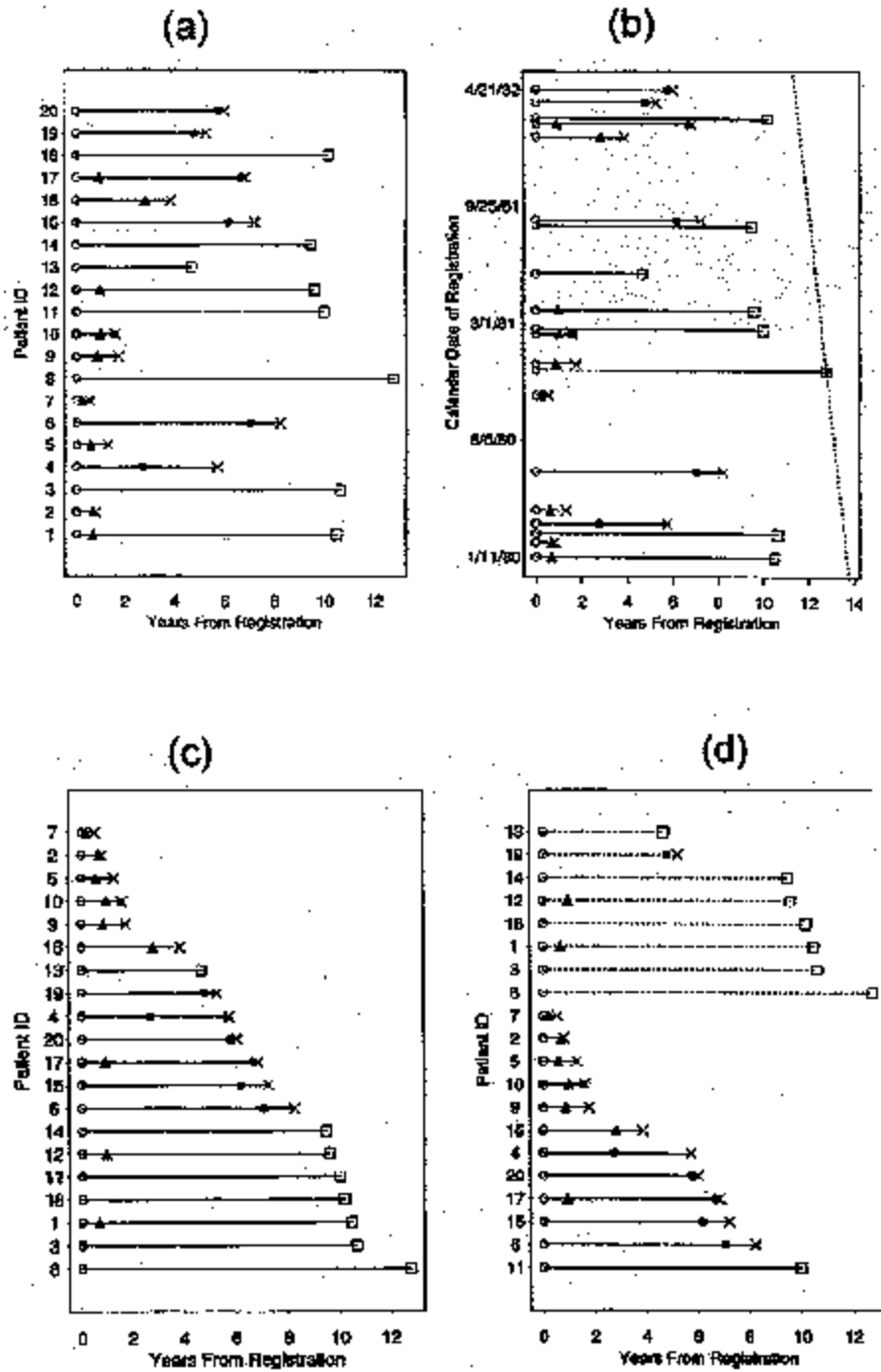
	sex	race	region	happy	life	sibs	childs	age	educ	paeduc	maeduc	speduc	prestg80	occcat80	tax	usintl
1	2	1	1.00	1	1	1	2	61	12	97	12	97	22	3.00	1	1
2	2	1	1.00	2	1	2	1	32	20	20	18	20	75	1.00	1	0
3	1	1	1.00	1	0	2	1	35	20	16	14	17	59	1.00	0	1
4	2	1	1.00	9	2	2	0	26	20	20	20	97	48	1.00	1	0
5	2	2	1.00	2	1	4	0	25	12	98	98	97	42	3.00	1	1
6	1	2	1.00	2	0	7	5	59	10	8	6	97	0	.	0	2
7	1	2	1.00	1	1	7	3	46	10	8	98	97	0	.	8	0
8	2	2	1.00	2	0	7	4	99	16	5	6	97	60	2.00	0	8
9	2	2	1.00	2	2	7	3	57	10	6	5	97	0	.	8	0
10	2	1	1.00	2	1	1	2	64	14	8	12	20	38	6.00	1	1
11	1	1	1.00	2	1	6	0	72	9	12	98	97	36	6.00	1	1
12	2	1	1.00	1	0	2	5	67	12	8	8	13	28	6.00	0	1
13	1	1	1.00	2	0	1	0	33	15	11	12	14	65	1.00	0	1
14	1	3	1.00	2	2	2	1	23	14	12	12	97	49	2.00	1	2
15	2	1	1.00	2	2	7	1	33	12	12	12	97	50	2.00	1	0
16	2	1	1.00	1	2	6	2	59	12	8	98	12	0	.	1	1
17	1	1	1.00	2	0	4	1	60	14	6	6	97	32	2.00	0	1
18	1	1	1.00	1	2	6	2	77	9	0	0	8	36	6.00	2	2
19	2	2	1.00	2	0	12	2	52	14	8	12	8	51	2.00	0	1
20	1	2	1.00	1	3	5	1	55	7	98	98	16	42	3.00	1	0
21	2	2	1.00	1	2	2	1	37	14	12	12	97	42	3.00	1	1

SPSS for Windows Processor is ready

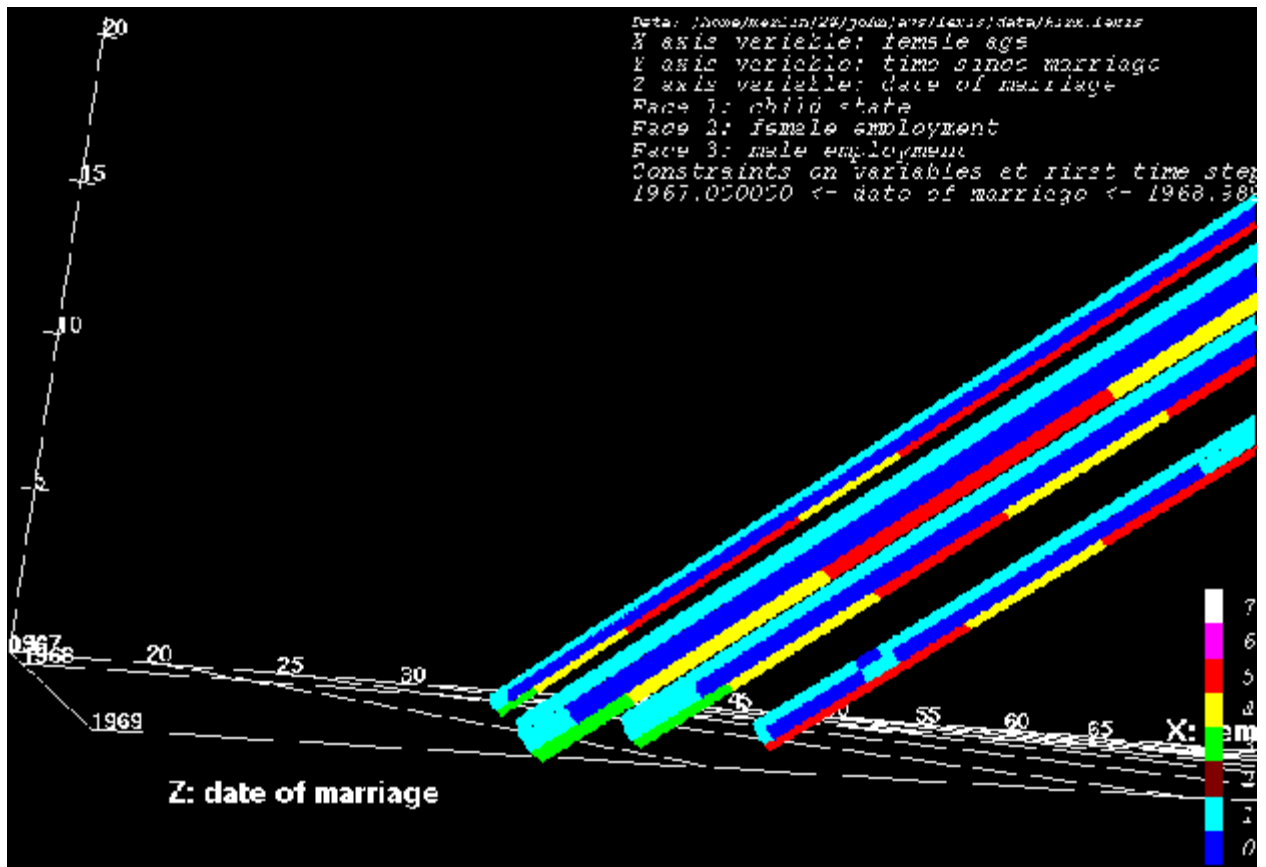
Figure 1: A Standard Data Collection Framework in SPSS



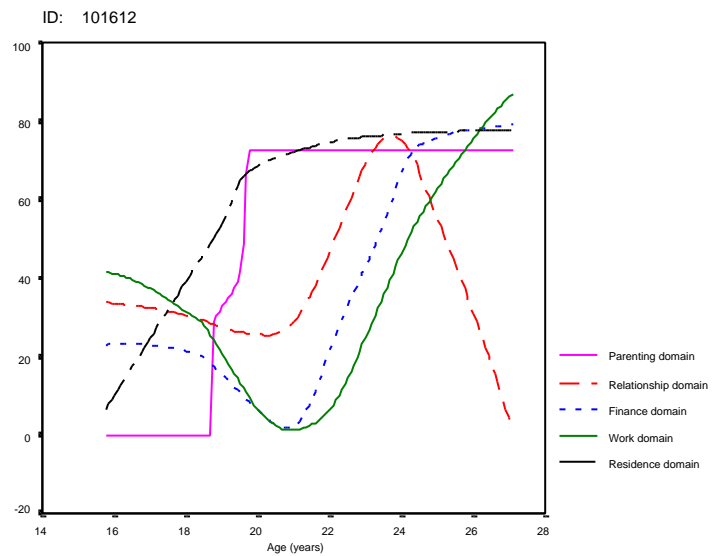
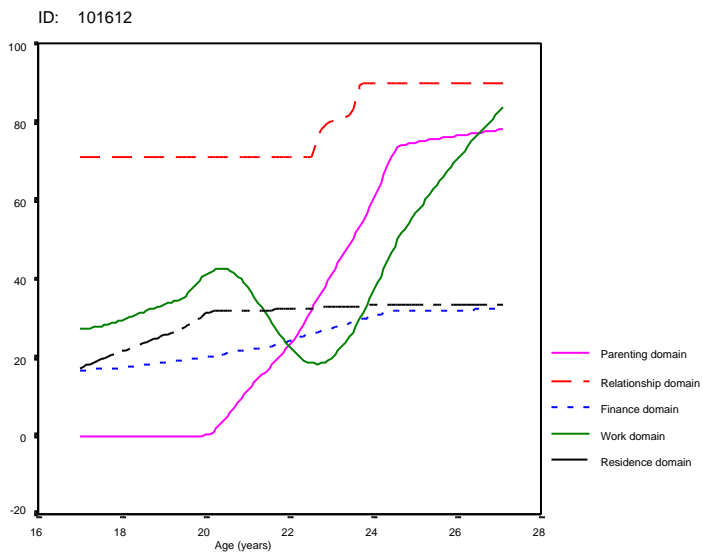
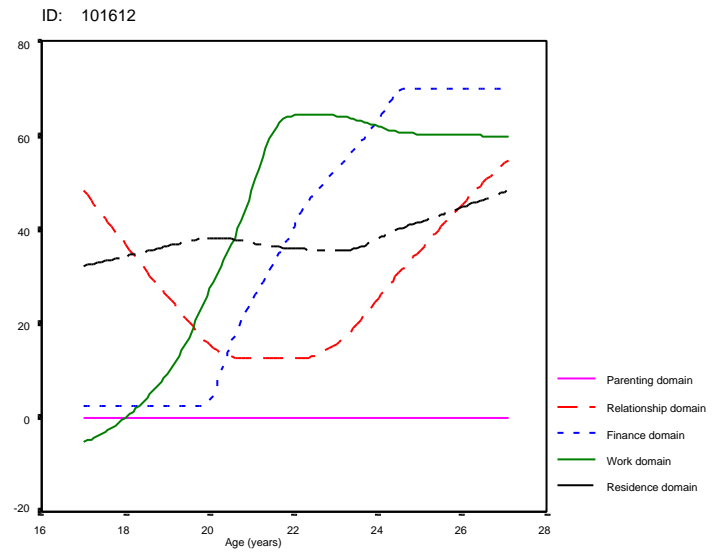
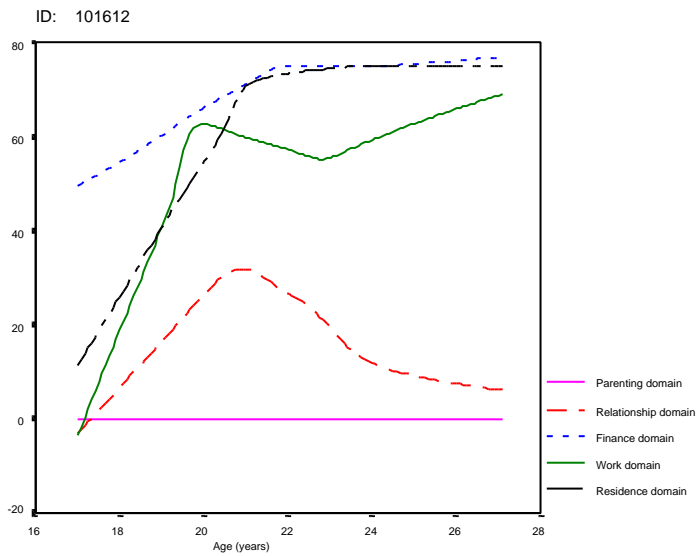
**Figure 2:** Eventchart of Bone Marrow Transplantation Data. Lines represent patients' event records, with open circles showing time of GvHD, black squares showing time of relapse, vertical lines showing time of death, and the diagonal line representing the "now line" (the limit to time available for observation). Source: Goldman (1992). Reprinted with Permission from *The American Statistician*. Copyright 1992 by the American Statistical Association. All rights reserved.



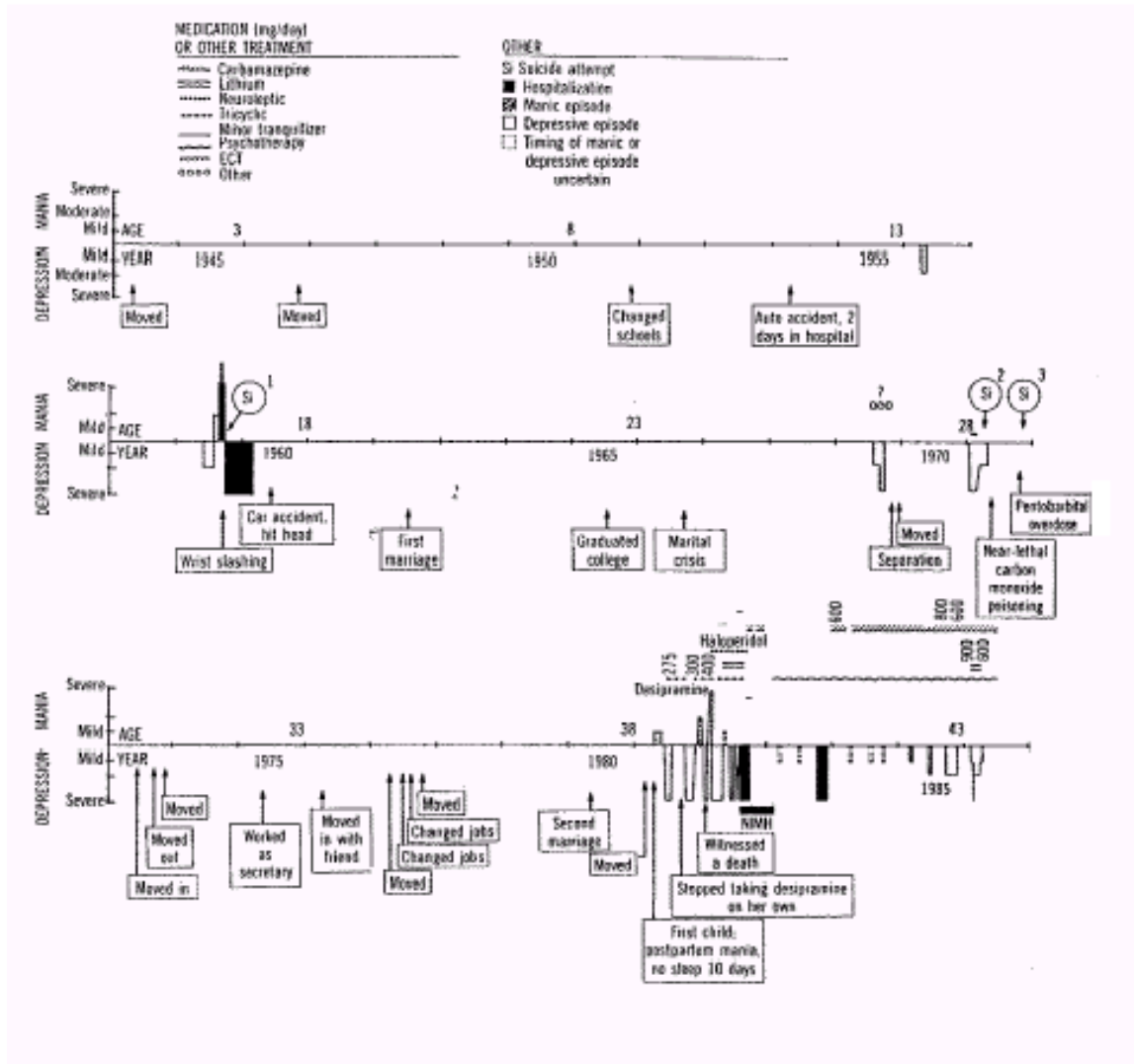
**Figure 3:** Event Charts Showing Different Ways of Displaying Survival Data.  
 Source: Lee *et al* (forthcoming).



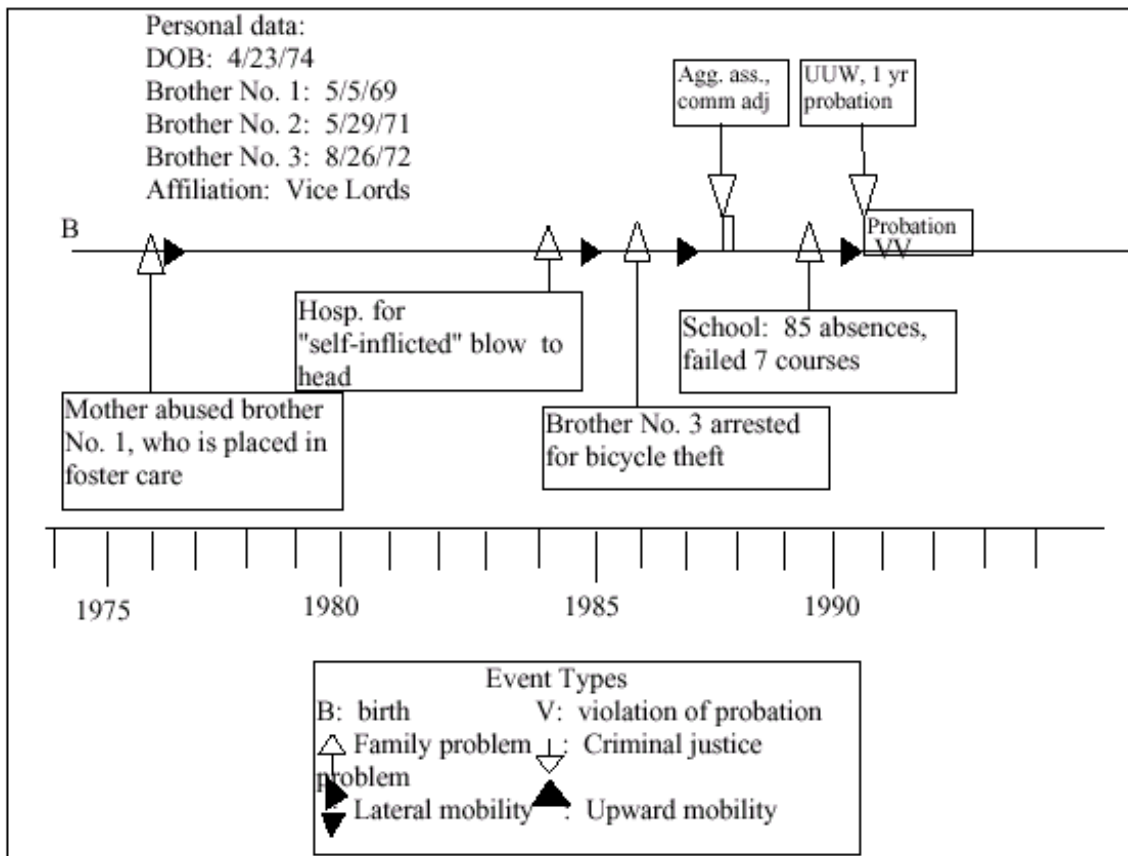
**Figure 4:** Four "Lexis Pencils" Showing the Relationship among Male Employment, Female Employment, and Age of the Youngest Child



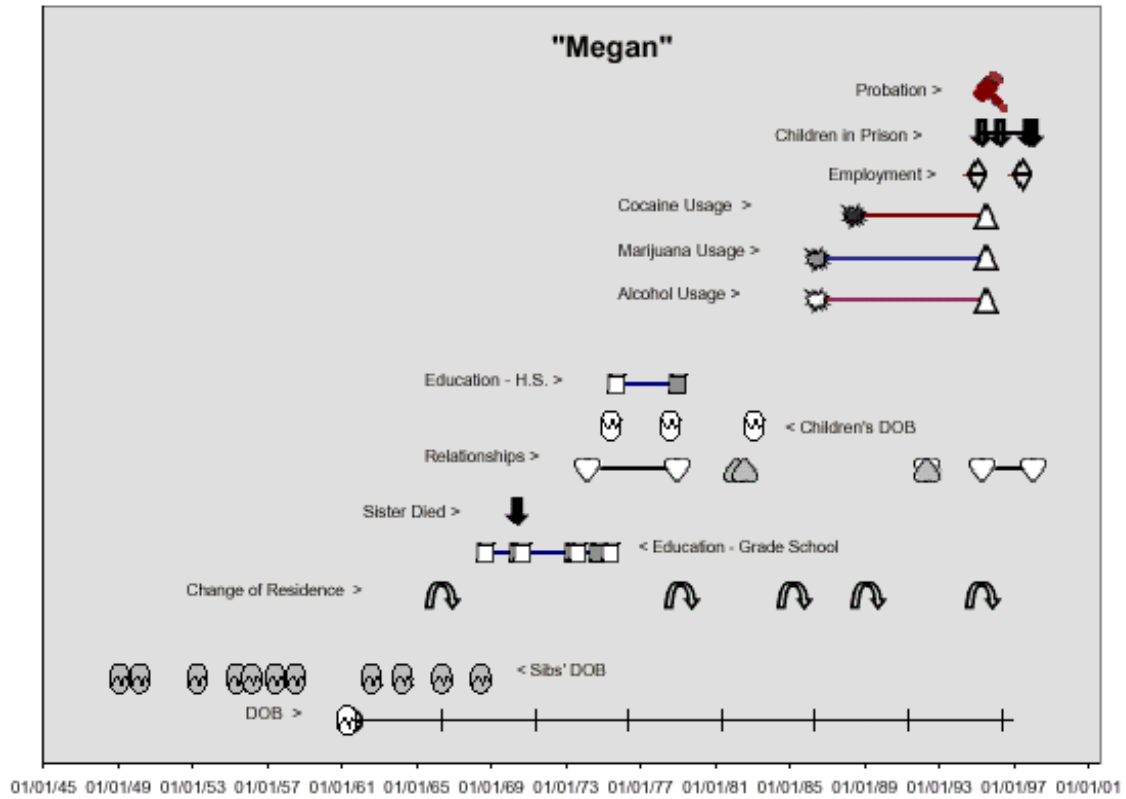
**Figure 5:** Different Paths of Four Individuals Making the Transition between Childhood and Adulthood.  
Source: Cohen (1999)



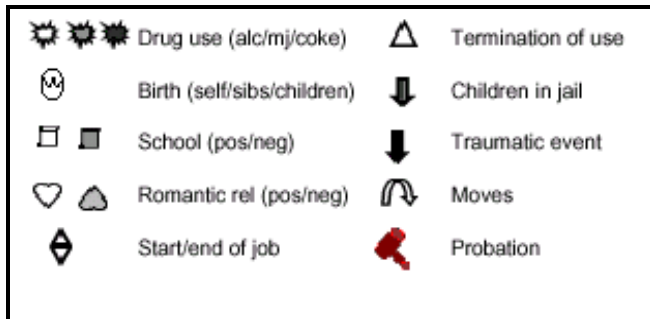
**Figure 6:** Life Chart of the Course of Illness of a Manic-Depressive Woman.  
Source: Post, Roy-Byrne, and Uhde (1988)



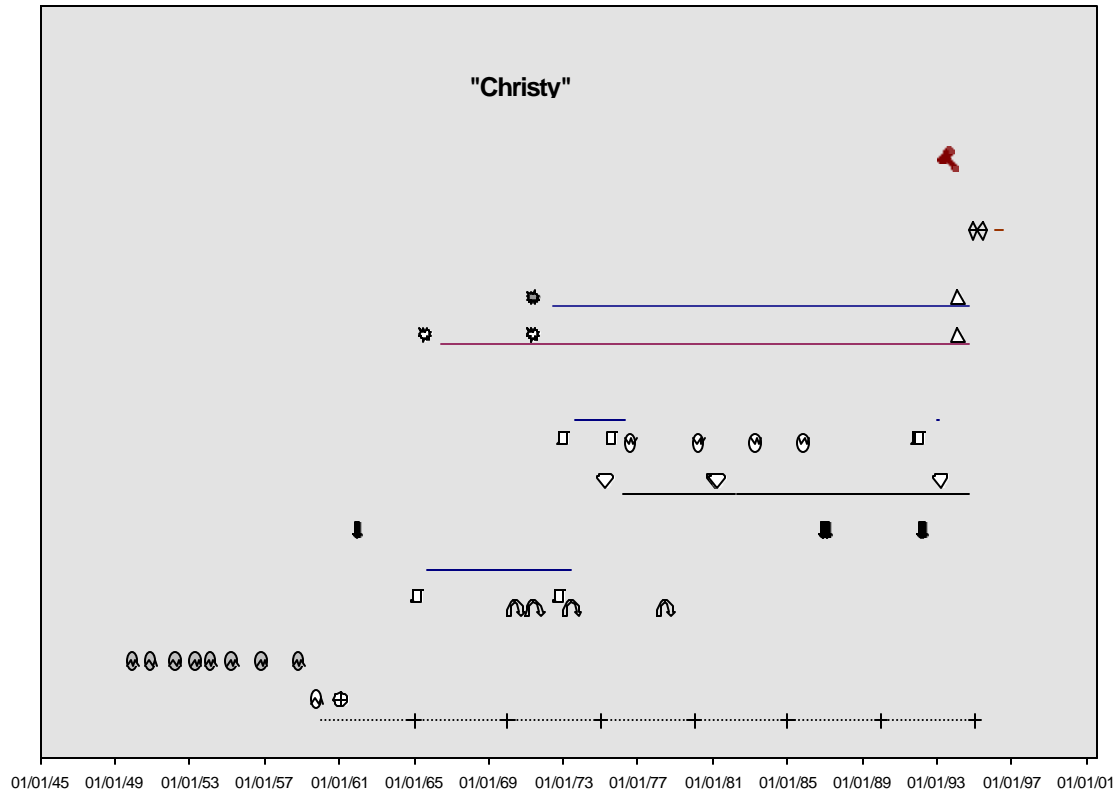
**Figure 7:** A Single Youth's Juvenile Record, Based Only on Data from Juvenile Court.  
 Source: Maltz (1995).



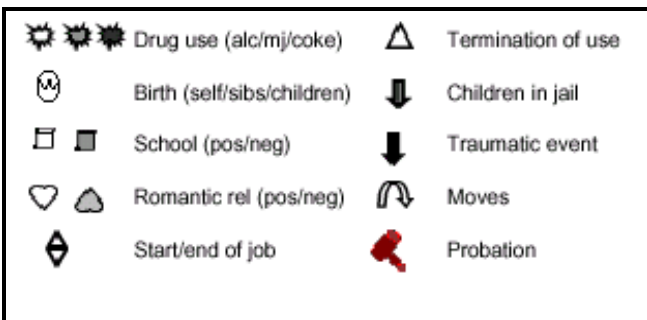
Legend:



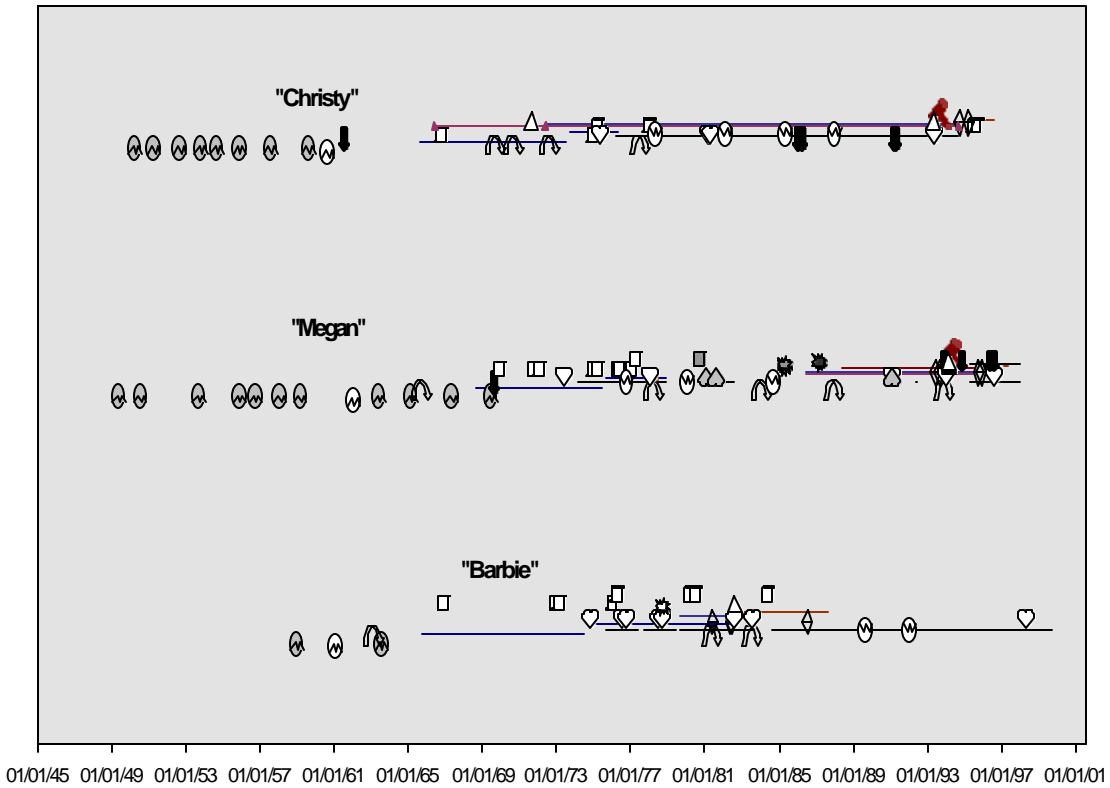
**Figure 8:** Life Course Events of a Female Probationer Code-Named "Megan."  
Source: Klosak (1999).



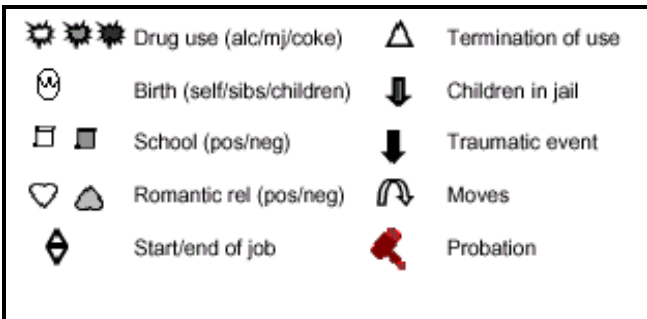
Legend:



**Figure 9:** Life Course Events of a Female Probationer Code-Named "Christy."  
Source: Klosak (1999).



Legend:



**Figure 10:** Life Course Events of "Megan," "Christy," and "Barbie," Compressed on the Same Page.