Lexis fields

Tim Riffe*[1] and José Manuel Aburto †[2]

[1]Max Planck Institute for Demographic Research
[2]Epidemiology, Biostatistics and Biodemography, Department of Public Health, University of Southern Denmark

May 24, 2018

Abstract

Background Lexis surfaces are an established visualization technique to show how a given value changes over age and time. Vector fields are a 2d or 3d representation of flow variables such as direction and speed (or force).

Objective We aim to increase the information density of patterns shown on the Lexis surface by placing a vector field on the Lexis surface.

Results We show Lexis fields using different combinations of visual encodings, such as color, contour layering, and angle, length, and thickness of field elements. These instruments enable information layering that is not common practice on standard Lexis surfaces.

Conclusions Lexis fields extend the analytic power of the Lexis surface, and these can be rendered to display information at higher densities than standard Lexis surfaces.

Introduction

Lexis surfaces are a graphical tool used to display data on the Lexis coordinate plane, a Cartesian plane that is also a simplex relationship between age, period, and cohort. Surfaces are often displayed as heat maps, contour maps, perspective plots, or variants of these things. Various kinds of quantities, such as raw magnitudes, differences, ratios, intensities, proportions, derivatives, and even compositions (Schöley and Willekens 2017)) can be displayed on Lexis surfaces in order to put age, period, cohort, or other patterns in relief.

Maps in general combine layers of categorical, continuous, and symbolic information on a common spatial projection. Lexis surfaces in contrast almost exclusively display one visual layer at a time. Even the composite surfaces of Schöley and Willekens (2017), which display layered information are rendered as a single visual layer. Small multiples of Lexis surfaces on the other hand constitute a de-layering, as these are spatially disjoint. We propose to enrich Lexis surfaces by adding a visual layer of quantitative information coded symbolically as a vector field, and we liken to cartographic information layering.

Vector fields are a graphical form generally used to display variation in speed, direction, or force over a plane. Point estimates of these quantities on the plane are often represented with segments or arrows, where length may be proportional to a function of magnitude (force, speed), and angle indicates direction, potentially disambiguated with an arrowhead or articulated as a curve. We propose a fusion of Lexis surfaces and vector fields, *Lexis fields*, as a tool to display variation in relationships between variables over age and time. A Lexis field may either be rendered atop on a Lexis surface, representing two map layers — a true Lexis map — or as a single-layer alone visualization.

*riffe@demogr.mpg.de
†jmaburto@health.sdu.dk

We give an overview of constructing a Lexis field, and an appplication. Our example explores the relationship between remaining life expectancy and the standard deviation of remaining lifespan over age and time based on all available populations in the Human Mortality Database from 1950 onward. Other potential applications are discussed.

1 Lexis field construction

It makes sense to plot a Lexis field if data contain a relationship that can be summarized with a line, a simple curve, or similar, and that varies by age and/or time. Constructing a Lexis field involves several desgrees of designer freedom, which can however be codified into four basic steps, which we summarize in Fig. 1. The steps to do so are outlined in the following steps, referenced to regions of Fig. 1.

A Determine the Lexis reference grid size for data selection. For example, a five-year grid implies 5×5 Lexis cells. Data may be selected from multiple populations on the same reference grid or may consist in a selection of attributes from a single underlying Lexis surface. Presumbably two variables are required.

B Abstract a model from the data, such as a linear, parabolic, or similarly simple relationship. We consider the case of a bivariate linear model that produces a result of the form $y = a + b * x$. An example of a nonlinear relationship is discussed.

C Translate the model fit to the characteristics of a line segment, or field *pointer*. Treat each grid cell on the Lexis diagram as a plot area, by default with equal year units in x and y, for example as implied by a 5×5 cell. In our implementation, the pointer always passes through the centroid of the Lexis cell. The pointer angle or slope may be taken as-is from a linear model, or exaggerated by the same multiplier for the entire field to increase definition. Within the Lexis cell a circle tangent to the four cell borders standardizes the length of the pointer, where the radius of the standard circle can be adjusted by defining an inner margin width (pad) to the Lexis cell. In the simplest case, all pointers may be of the same length, irrespective of slope, as determined by this reference circle. Otherwise, length may be proportional to some other data or model characteristic, such as the observed range of x, the goodness of model fit, or similar. Likewise, other segments characteristics, such as color, or width, may also map to data characteristics.

D Render the segment in the corresponding Lexis grid cell, repeating all steps for each cell in the diagram. Variations over the Lexis plane in pointer aesthetics

2 Application

We select all HMD data available for females after 1950. For each lifetable we calculate two additional columns: the standard deviation of remaining lifespan $sd(x)$, and the coefficient of variation of remaining lifespan $CV(x)$. Each Lexis field element is based on the relationship between $sd(x)$ and remaining life expectancy $e(x)$ in 1×5 Lexis cells[1], as summarized by bivariate linear regression over the data points in each cell. The regression results used for each Lexis cell in resulting Lexis fields are identical, but the translation of regressions to field pointers varies between designs. We offer four examples of Lexis field designs, displayed in Fig. 2, each rendered on 5×5 Lexis subplots.

The first of these, Fig. 2a, is a bare-bones Lexis field that serves to illustrate the underlying concept. This display renders each regression slope as a line segment of equal length (4 "years" long) and centered on each Lexis cell. The slope of each pointer is rendered identical to regression slopes, which may be justified in this

[1] Data points included in regressions are single ages evenly divisible by five for each of the five years included in a Lexis reference cell.

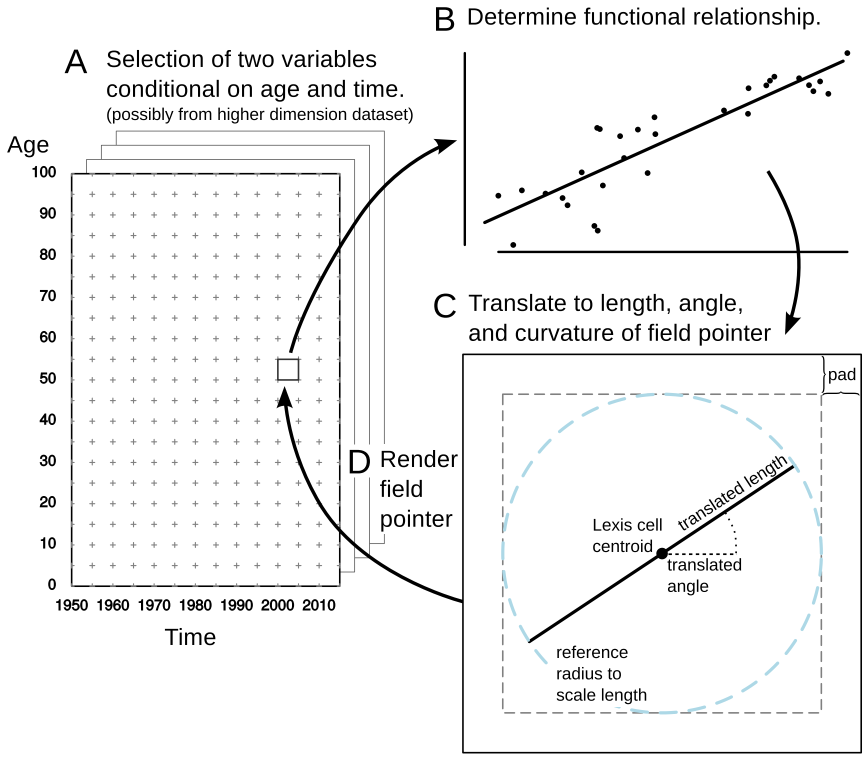
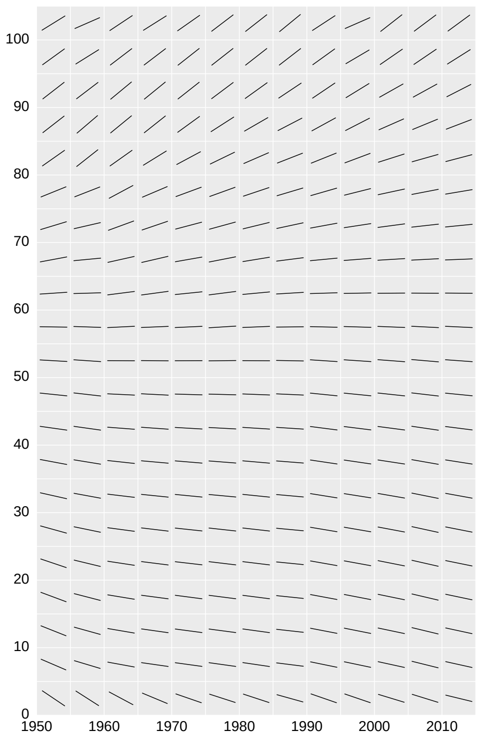


Figure 1: A diagram depicting the translation of functional relationships in data conditioned on age and time to visual encoding on a Lexis field. **A**: Condition data selection on age and time. **B**: model the functional relationship in data subset. **C**: Translate the model to field elements, 'pointers', using angle, and possibly also length, curvature, thickness, etc to encode model qualities. **D**: Populate the Lexis plane with field pointers to create a Lexis field.

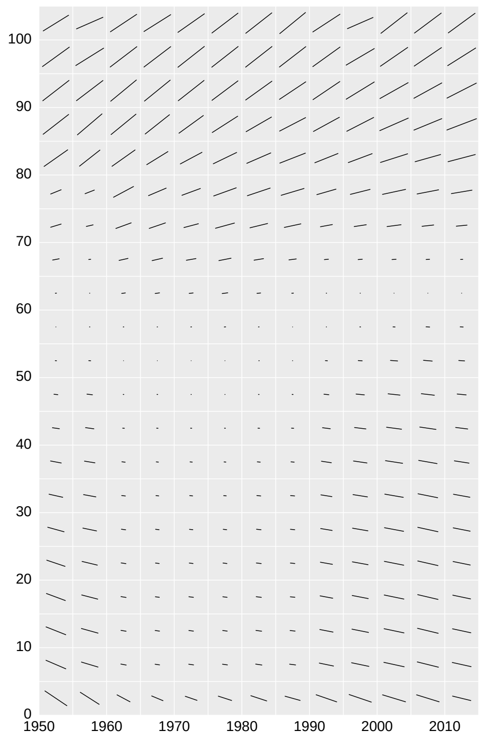
case, since Lexis cells fix a 1:1 year aspect ratio, while $e(x)$ and $sd(x)$ are also in year units. This is the truest and most literal depiction of how these regression slopes vary over age and time among HMD females, and nothing more. From this figure we can see, for example, that there is some age where the relationship turns from negative to positive, which increased slightly over time. Slopes dampened in younger ages around the 1980s, but have since increased (except infants).

The second version, Fig. 2b, also renders slopes on a unity aspect ratio, but lengths are proportional to the regression r^2. Ergo, longer segments represent more "linear" fits and shorter segments, almost approaching points, indicate areas where the relationship is not well described by a line. Other measures of model fit could be used to similar effect to scale some characteristic of the field pointers. For example, Fig. 2c renders slopes exaggerated by a factor of 2, with pointer lengths scaled proportional to the interquartile range of $e(x)$ used in regression, and line weight and grayscale "proportional" to the r^2 of the regression fit. Segment lengths are therefore indicative of the spread in the data, while higher r^2 results in more contrast in the field.

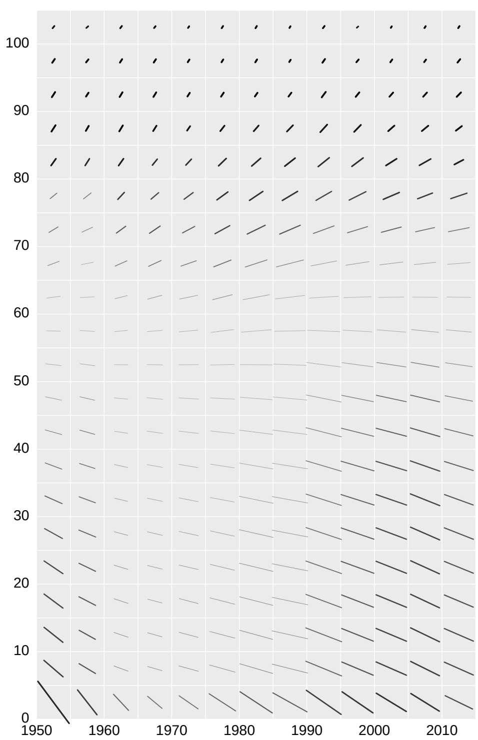
The final example, Fig. 2d is a true Lexis map. The base of the map is a filled contour plot of the average coefficient of variation $CV(x)$ in single ages. This map is redundantly coded with a sequantial color palette and labelled contours, which liberates the surface from an explicit color legend. The same field from Fig. 2d is layered atop the CV surface, acheiving a true layered map. In principle, one could represent variation in the slope of some other regression over age and time as a contour plot, with the present field atop, thereby layering comparable information. However, the present example serves to illustrate layering with the field.



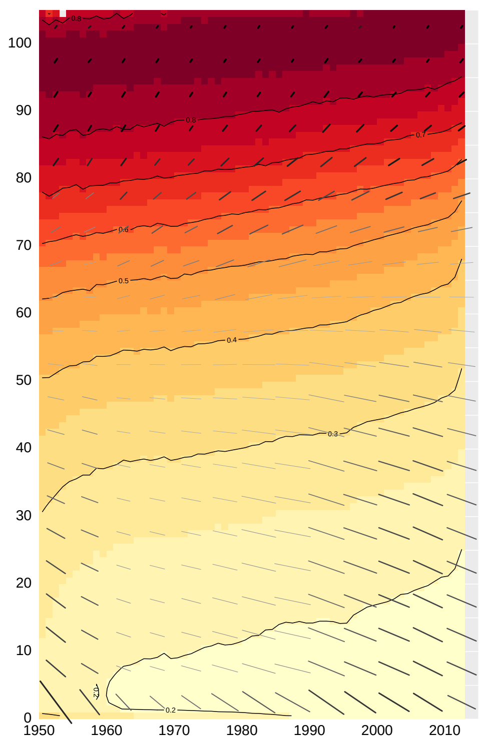
(a) Lexis field: $sd(x)$ by $e(x)$ linear fits, with slopes drawn directly. Pointer lengths are all equal.



(b) Lexis field: $sd(x)$ by $e(x)$ linear fits, with slopes drawn directly. Pointer length is proportional to the r^2 of the fit.



(c) Lexis field: $sd(x)$ by $e(x)$ linear fits, with slopes exaggerated by 2. Pointer length is proportional to the IQR of $e(x)$, while grayscale and segment width are proportional to r^2 of the fit.



(d) A Lexis map: Lexis surface of mean CV as a filled contour plot plus a Lexis field of $sd(x)$ by $e(x)$ linear fits, slopes exaggerated by 2. Pointer length is proportional to the IQR of $e(x)$, while grayscale and segment width are proportional to r^2 of the fit.

Figure 2: Four versions of Lexis fields displaying the linear relationship between the standard deviation and mean of remaining lifespan, females (HMD).

3 Discussion

We suggest the use of vector fields on the Lexis surface, introducing the notion of a Lexis field, which is simple a vector field on a regular Lexis grid over age and time. We demonstrate some of the designer degrees of freedom in translating data into the elements of a Lexis field, as well as an instance of Lexis map layering. These examples serve to illustate the construction of Lexis fields, but do not pretend to be "best practice" Lexis surface in terms of visual design or legibility. It is our sense that the patterns revealed in Figures 2a-2c are accessible to the viewer and lend themselves to substantive interpretation. This visual instrument indeed arose in practice in an attempt to investigate the apparently mechanical relationship between lifespan variation and average length of life with a macro view. Fig. 2c simply serves to illustate that Lexis fields can be layered with traditional Lexis surfaces that are color coded, increasing the information and pattern density on the Lexis plane with little drawback in terms of legibility.

Although patterns in data may be much more complex than can be expressed with linear models, these simple model fits can be thought of as regular samples from the complex space implied by data, such that the pattern revealed on the Lexis field is still revealing. On the other hand a field may be derived from a single underlying pattern rather than a series of regressions on different populations or variables. For example, Shang (2018) recommends the use of phase diagrams to represent the rate of change of the hypothetical lifecourse implied by period fertility curves. In these phase diagrams, first derivatives (velocity) of age-specific fertiltiy rates (ASFR) are assigned to x and second derivatives (acceleration) to y on a plane, such that fertility plots a clockwise a spiral path around the origin. The information contained in a fertility phase diagram derives from a vertical strip of a fertility Lexis surface, yet fertility surfaces are typically rendered as a single layer consisting in absolute ASFR, which is not an effective way to put derivatives into perspective. The phase diagram fills this gap but loses the intuition of levels and the temporal orientation of the Lexis surface. It would be possible to draw a Lexis map with ASFR as a standard surface with an extra field layer, where field pointers are rendered tangent to the corresponding age on the phase path, and length proportional to the "distance" travelled on the path between adjacent ages around this point. Other variants can be imagined that would make use of segment length and direction to characterize surface gradient, with second derivatives mapped to color (acceleration heating up, deceleration cooling off). Certainly variants of vector fields could be used to intuitively render demographic flows, and these do not need to adhere to the Lexis grid.

4 Conclusions

We describe the construction and use of vector fields on the Lexis plane. We argue that this technique can increase the information density and scope displayed on the Lexis surface. We also argue that the visual encoding of fields is easily rendered compatible with Lexis surfaces rendered as filled contour plots, which lends itself to map layering. We provide examples of various visual encodings of Lexis fields, and a single example of a Lexis map consisting in a filled contour surface base layer with a field overlay. We emphasize some aspects of designer degrees of freedom. We suggest alternative visual encodings, as well as an approach to recodify phase planes as fields in the Lexis space. In sum, we think that displaying a larger variety of demographic quantities on the Lexis plane and increasing the information density on the Lexis plane using techniques such as fields should broaden the scope of demographic exploration and sharpen the instruments of demographic pattern detection.

References

Human Mortality Database. University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 17 April, 2017).

Jonas Schöley and Frans Willekens. Visualizing compositional data on the lexis surface. *Demographic Research*, 36:627–658, 2017.

Han Lin Shang. Visualizing rate of change: an application to age-specific fertility rates. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2018.

James W Vaupel, Bradley A Gambill, and Anatoli I Yashin. Thousands of data at a glance: shaded contour maps of demographic surfaces. 1987.

Daniel Weiskopf. *GPU-Based Interactive Visualization Techniques*, chapter Vector Field Visualization, pages pp. 81–159. Springer, 2007.