Exploring the demographic history of populations with enhanced Lexis surfaces

Jorge Cimentada
Sebastian Klüsener
Tim Riffe

This publication is part of the Special Collection on “Data Visualization,” organized by Guest Editors Tim Riffe, Sebastian Klüsener, and Nikola Sander.

© 2020 Jorge Cimentada, Sebastian Klüsener & Tim Riffe.

This open-access work is published under the terms of the Creative Commons Attribution 3.0 Germany (CC BY 3.0 DE), which permits use, reproduction, and distribution in any medium, provided the original author(s) and source are given credit. See https://creativecommons.org/licenses/by/3.0/de/legalcode.
## Contents

1. **Motivation**  
2. **Data and implementation**  
   2.1 Standardization of cohort line width  
   2.2 Dimensions to be visualized  
   2.3 Color categorization  
3. **Visualization examples**  
4. **Discussion and conclusion**  

References
Exploring the demographic history of populations with enhanced Lexis surfaces

Jorge Cimentada¹
Sebastian Klüsener²
Tim Riffe³

Abstract

BACKGROUND
Lexis surfaces are widely used to analyze demographic trends across periods, ages, and birth cohorts. When used to visualize rates or trends, these plots usually do not convey information about population size. The failure to communicate population size in Lexis surfaces can lead to misinterpretations of mortality or other conditions that populations face. For example, high mortality rates at very high ages have historically been experienced by only a small proportion of a population or cohort.

OBJECTIVE
We propose enhanced Lexis surfaces that include a visual representation of population size. The examples we present demonstrate how such plots can give readers a more intuitive understanding of the demographic development of a population over time.

METHODS
Visualizations are implemented using an R-Shiny application, building upon perception theories.

RESULTS
We present example plots for enhanced Lexis surfaces that show trends in cohort mortality and first-order differences in cohort mortality developments. These plots illustrate how adding the cohort size dimension allows us to extend the analytical potential of standard Lexis surfaces.

CONTRIBUTION
Our enhanced Lexis surfaces improve conventional depictions of period, age, and cohort trends in demographic developments of populations. An online interactive

¹ Max Planck Institute for Demographic Research, Rostock, Germany. Email: cimentada@demogr.mpg.de.
² Federal Institute for Population Research, Wiesbaden, Germany; Max Planck Institute for Demographic Research, Rostock, Germany; and Vytautas Magnus University, Kaunas, Lithuania.
³ Max Planck Institute for Demographic Research, Rostock, Germany.
visualization tool based on Human Mortality Database data allows users to generate and export enhanced Lexis surfaces for their research. The R code to generate the application (and a link to the deployed application) can be accessed at https://github.com/cimentadaj/lexis_plot.

1. Motivation

Lexis surfaces are widely used to visualize life table and other kinds of demographic data in order to detect and analyze regularities and anomalies in demographic patterns across periods, ages, and birth cohorts (Vaupel, Gambill, and Yashin 1987; Schöley 2016; Schöley and Willekens 2017; Rau et al. 2018). These surfaces are built upon the Lexis diagram representation developed by Wilhelm Lexis (1875), based on earlier work by Gustav Zeuner, Otto Brasche, and Karl Becker, among others (Vandeschrick 2001). The Lexis diagram can be represented in several different ways (Keiding 2011); we follow the most common orientation, where age is displayed on the ordinate and period on the abscissa (also referred to as AP). Lexis surfaces usually focus on a single aspect of population development, such as mortality rates. However, such single layer representations have limitations. For example, with a color scheme for mortality rates in which higher mortality is indicated by a more saturated or more prominent color (such as dark red), the upper part of a Lexis surface becomes visually very dominant, as mortality is particularly high at the highest ages. Moreover, because the area representing high mortality rates usually extends over several ages, this spatial cluster in the upper part of the Lexis diagram forms a strong perceptual group (see Duncan 1984; Merikle 1980). However, it is often the case that at very high ages, only a few members of a population remain alive. Thus, this visually dominant part of the Lexis surface is of limited relevance to the survival experiences of a population or cohort.

The enhanced Lexis surfaces we propose are an attempt to overcome this limitation by integrating into a single plot at least two dimensions of demographic development: In our examples, mortality rates are the first dimension, and population size is the second dimension. Enhanced Lexis surfaces could also include dimensions related to fertility, migration, or other phenomena that vary over age and time. In constructing such figures, one of our objectives is to avoid moving to a 3D representation, as 2D representations continue to be easier for readers to grasp and more convenient to include in scientific manuscripts. Our enhanced Lexis surfaces take a cohort perspective, with birth cohorts plotted as diagonal lines so that readers can follow cohorts as they age. These lines are formed by a succession of age-cohort Lexis parallelograms, which we call AC cells. In our plots, the width of the lines varies
depending on the size of the cohort over time. Each age-cohort combination is represented by an AC cell, where we expand or shrink the width of these lines to represent standardized values of the population, with color used to inform the reader about the value of the indicator (e.g., the mortality rate) in each cell. AC cells are placed in succession to form a diagonal line for each cohort. Figure 1 shows a simple example of the differences between the classical Lexis surface and our enhanced Lexis surface.

**Figure 1:** Conventional Lexis surface and enhanced Lexis surface

A conventional Lexis surface graph is depicted in Figure 1a, where each combination of age and cohort is represented by a Lexis parallelogram. The color varies by an indicator of interest, in this case cohort mortality rates. The area covered by each AC cell in this conventional Lexis surface is set to be the same across all cells, both within a given cohort but also between cohorts. Our enhanced Lexis surface in Figure 1b adapts the width of each AC cell. In this example, this is done proportional to the mean population size of that specific cohort in a certain year, standardized by the biggest size ever recorded for that cohort (see below for more details and options). Such enhanced Lexis surfaces allow us to (1) visualize how the size of a cohort is subject to variation as it ages and/or (2) show how different cohorts vary in size. We believe that such plots provide users with a more intuitive understanding of the ages at which mortality levels are or were particularly relevant for cohorts or populations living at specific time periods. In addition, enhanced Lexis surfaces allow us to visually explore data to investigate whether cohort size differences relate to variation in mortality or other demographic phenomena across cohorts. For example, these surfaces enable users to trace how disparities change depending on cohort size (also called Easterlin effects; Easterlin 1987) or to grasp the degree to which a cohort was subject to potentially selective in- and out-migration. An accompanying online visualization tool allows scientific users and the interested public to explore these new analytical potentials independently.
2. Data and implementation

Data for the visualization tool come from the Human Mortality Database (HMD 2019), which provides high-quality cross-country comparative life table mortality data for about 40 countries. We use cohort death rates and exposure to risk, extracted from HMD files cMx_1x1 and cExposures_1x1, respectively. The visualization tool is implemented using an R-Shiny application. The main advantage of relying on the R-Shiny app rather than other visualization programs is that it allows us to access existing R-libraries that support the analysis of HMD data in Lexis surfaces. For example, we use the HMDHFDplus package (Riffe 2015) to scrape HMD data directly from the HMD web page (HMD 2019), thereby gaining instant access to new data when the HMD is updated. Our visualizations are created using functions from R base graphics (R Core Team 2018).

2.1 Standardization of cohort line width

A key challenge in creating enhanced Lexis surfaces is to ensure that valid and consistent procedures are used to scale line widths to the cohort size at each point in time. The HMD contains data for populations that vary greatly in size. For example, the United States has a population of more than 300 million, while Iceland’s population is almost 1,000 times smaller. For some populations, the data span more than two centuries, including periods of substantial growth. We provide users four ways to standardize line widths. The first option (1) is the conventional Lexis surface, in which the AC cells that form the cohort lines are plotted next to each other, with no space between them (see, e.g., Figure 2 or Figure 4). The other three options shrink the cells so that cohort lines vary by the mean size of a cohort at a specific age. The second option (2) enables users to standardize by a reference cohort of their choice (see, e.g., Figure 5b, where we take the cohort born in 1960 as the reference cohort), while the third option (3) allows users to standardize by a reference year. The fourth option (4) enables users to standardize each cohort relative to its own maximum size (see, e.g., Figure 3).

In the second option, the application detects the biggest size recorded for the reference cohort at any age, whereas in the third option, the application detects the biggest cohort of any age in the reference year. This reference number is then used to standardize the size of the cohort line segments by age for the whole plot. A limitation

4 In this standardization, we divide all the recorded cohort sizes by the derived reference number and multiply these numbers by 0.9. As a result, the line for the cohort/year combination used as a reference is plotted at a width of 0.9 (a width of 1 would imply that we were plotting a whole surface).
of this approach is that it could end up generating overlapping lines if the cohort/year chosen as the reference has much lower population numbers than the biggest sizes ever recorded for the studied population. To avoid overlapping lines, a scaling factor is applied if the biggest line is above a threshold. In the fourth approach, the application treats each cohort separately. For each cohort we detect the biggest cohort size ever recorded at any age and use this as the base for standardization of the AC cell widths of that cohort. For this operation we consider only cohorts that can be followed from birth onward. Otherwise, the standardization might result in misleading patterns, particularly for those cohorts that can be tracked for only a short period at high ages with high mortality levels. Thus, in this fourth option, we lose the upper left corner of the Lexis surface, in which the cohorts that cannot be followed from birth onward are displayed. Unlike the second and third approaches, the fourth approach does not display differences in cohort sizes. Instead, it allows the reader to explore the data to determine at which age cohorts reported their highest population numbers, and to detect differences in cohort attrition at higher ages. Of course, the size of a cohort over time is affected not only by mortality events but also by migration. However, given that international migration is usually undertaken between the ages of 20 and 40, the interpretation of variation in the development of the line widths at higher ages is mostly driven by mortality events. This fourth option thus allows the reader to explore the data to find out long-term shifts in which proportions of cohorts were able to reach high ages.

2.2 Dimensions to be visualized

The application currently provides data on three indicators: (1) cohort mortality rates, (2) gender differences in cohort mortality rates, and (3) first-order differences in cohort mortality rates. As previously stated, the indicators are not necessarily limited to mortality indicators, as we could potentially plot fertility (e.g., from the Human Fertility Database) and migration indicators (e.g., implied migration balances).

The application provides users with the choice to visualize raw rates or smoothed rates. Smoothed rates aim to make broad patterns clearer where random variation may otherwise mask the patterns. Such random variation is more likely to occur in smaller populations and might especially play a role at (1) young ages at around 10, where death rates are very low, and (2) older ages, where death rates are high but where few survivors remain to be exposed to risk. Smoothing may also have the undesirable effect of eliminating real ruptures, and for this reason it is advised to compare smoothed and

5 If the standardization returns for any AC cell values above 0.95, we rescale all polygons by the biggest AC cell value displayed in the plot. The latter AC cell is then displayed with a line width of 0.9.
raw rate surfaces. Our application uses one-dimensional (1D) P-splines, implemented in the MortalitySmooth R package (Camarda 2012). This method assumes that death (or birth) counts are Poisson distributed. We opted to 1D smooth over age within calendar years to preserve period mortality shocks, which appear frequently in these historical data.

All three currently implemented indicators are available for men and women. A specificity of the second indicator, on gender differences in mortality, is that users can standardize these mortality differences for either women in reference to men (with the cohort line size depicting women) or for men in reference to women (with the cohort line size depicting men). In addition, users must select the country for which they would like the data to be displayed. Finally, the background color can be black, gray, or white. The white background is useful if the graph is intended to be used in a manuscript, while the gray and black backgrounds might be the preferred choice for computer screen presentations to avoid glaring effects.

2.3 Color categorization

Currently, color schemes are chosen automatically based on choices made by the user. For the cohort mortality rates, colors are derived from the magma palette of the viridis library (Garnier 2018). This palette is designed to ensure that most color-impaired readers can detect the color variation. To maximize contrast, we use a gamma distribution to cut out colors from the given range of colors in the palette. For the plots displaying gender mortality differences and first-order differences in mortality over time, we use a purple-green diverging palette (PRGn) from the RColorBrewer library (Neuwirth 2015). This palette is also distinguishable for most color-impaired readers. In all derived plots, a legend provides the range of colors and a density curve of the displayed data. In deriving the density curve we treat each plotted AC cell as a unit of observation.

For cohort mortality rates, the legend refers to deaths per 1,000 persons, while for the gender mortality differences, it depicts the ratio of male deaths per 1,000 males over female deaths per 1,000 females multiplied by 100 (the default option; this can be reversed to females over males in the Shiny application). For the first-order differences, the legend displays the change in death rates over time within each age. This last indicator is calculated as the natural log of the difference between a given year and the preceding year, and it can be interpreted as the growth rate change in percentage between years.
3. Visualization examples

In our visualization examples, we focus on graphs displaying cohort mortality rates and those showing first-order differences in cohort mortality rates. We first turn to Sweden, the country with the longest available time series, to plot male cohort mortality rates (\( cmx \)). To demonstrate how enhanced Lexis surfaces can improve our understanding of mortality conditions, we plot first a conventional Lexis surface in Figure 2. We then present the same data with an enhanced Lexis surface in Figure 3.

**Figure 2: Conventional Lexis surface of cohort mortality trends among males in Sweden**

If readers are interested in age-period-cohort variation of mortality, the conventional Lexis surface is likely the preferred choice. An advantage of the conventional Lexis surface is that period fluctuations in cohort mortality rates come out very clearly. Mortality crises such as the famine of 1772–1773, the Dano-Swedish War of 1808–1809, and the Spanish flu of 1918 clearly stand out. In addition, smaller mortality crises are detectable. However, the conventional Lexis surface is not able to provide an intuitive understanding of which ages were particularly important for the mortality experience of cohorts and populations over time. In the 18\(^{th}\) century, a large fraction of children did not survive beyond age 5. However, in the left part of the graph,
the high mortality rates at age 80 are much more dominant, even though they were of little relevance for these cohorts. To gain a deeper understanding of what this variation implies for the studied population at risk, the enhanced Lexis surface depicted in Figure 3 offers additional insights. In this example surface, the cohort line width has been standardized by the biggest size ever recorded for each cohort (standardization 4).

**Figure 3:** Enhanced Lexis surface of cohort mortality trends among males in Sweden

![Enhanced Lexis surface of cohort mortality trends among males in Sweden](image)

*Source:* HMD, own calculations.

In comparison to Figure 2, the areas displaying mortality at ages above 80 years are less dominant, particularly in the upper left part of Figure 3. As a result, the high infant mortality in the 18th and 19th centuries becomes more prominent. Consequently, readers are better able to grasp the high relevance of infant mortality for the mortality experience of cohorts born in this period. We can also observe that the proportion of a cohort that reached ages above 80 years increased among the cohorts born in the early 19th century, while this tendency stalled among the cohorts born between 1850 and 1880. However, this trend seems to be driven by migration, as these cohorts already thin out at ages with high migration intensities (20–30 years). This is particularly visible in the period 1871–1900. While survival improvements at young ages continued through the first half of the 20th century, massive increases in the share of cohorts reaching age 80 are detectable only after 1960, particularly with the onset of the
A decrease in cardiovascular-related deaths in the 1970s (Mensah et al. 2017). In addition, our enhanced Lexis surface demonstrates that in recent years, high mortality rates at older ages have become increasingly relevant for men in Sweden, as higher shares of male cohorts in Sweden are able to reach these ages.

In addition to providing a general perspective, the enhanced Lexis surface invites readers to zoom in to study other details. In particular, the Shiny application provides high-resolution images (Scalable Vector Graphics) with a bigger plot size; the original resolution is retained regardless of the size of the image. This property, combined with the interactivity of the application and the option to download the high-resolution images, provides much more analytical potential than just having access to static plots in a manuscript. For example, in the case of Sweden, the option to zoom in allows us to explore patterns in the late 19th century, when the country experienced a large emigration wave. According to the healthy migrant hypothesis (e.g., Razum, Zeeb, and Rohrmann 2000), healthier people were more likely to leave the country. Readers can identify the cohorts that were most affected by emigration by zooming in to study the plot. They can then investigate the question of whether cohorts that were thinning out in their very early adult years due to emigration (e.g., the cohorts born 1850–1880) had different mortality experiences in their later years than those cohorts with lower emigration levels. Clear indications of such a pattern are not visible in the graph.

In a second example, we show cohort mortality data for males in France. In the conventional Lexis surface (Figure 4), similar to the Swedish example, mortality crises in France, such as World War I and World War II, are clearly visible. However, the area of the plot displaying mortality rates above age 80 is again perceptually dominant. The enhanced Lexis surface (Figure 5a), in which each cohort has been standardized by the biggest size ever recorded for that cohort (standardization 4), demonstrates the shifting relevance of mortality by age and also offers a better understanding of the tremendous effects World War I had on the size of cohorts of prime drafting ages. As with Sweden, the high importance of mortality early in life for the mortality experience of populations during the 19th century is more apparent in the enhanced Lexis surface. For interpretation of the effects of World War I, it is important to note that France experienced territorial changes that had an impact on the total population size (Glei et al. 2017). For example, Alsace-Lorraine was excluded from the French dataset from 1871 to 1919 due to its temporary secession to the German Empire. The distorting effects of such territorial changes are most visible in the trends during World War I, as the data for France during this period exclude the data for Alsace-Lorraine as well as for the territories most affected by war activities (the departments of Aisne, Ardenne, Marne, Meurthe-et-Moselle, Meuse, Nord, Oise, Pas-de-Calais, Somme, and Vosges). The existence of distortions due to territorial changes is a limitation when constructing
enhanced Lexis surfaces. However, unlike in conventional depictions, drastic changes in the covered population are immediately visible in such plots.

Figure 4: Conventional Lexis surface of cohort mortality trends among males in France

An even different perspective is offered by an alternative enhanced Lexis surface (Figure 5b) in which we standardize the line widths by the biggest cohort size ever recorded for the cohort born in 1960 (standardization 2). For this particular example (Figure 5b), the highest cohort size for the cohort born in 1960 was recorded at an age of 11 years. With this alternative standardization we are able to see links between World War I and World War II. As a result of World War I, the French cohorts born between 1914 and 1919 were particularly small. At the beginning of World War II, these small cohorts were of prime drafting age, which put France demographically in a weak position. On the other hand, the German Empire, despite a drastic fertility bust during World War I, was in a better demographic position, as the overall fertility level had been higher than that in France in the beginning of the 20th century (see Vandenbroucke 2014).
Figure 5: Enhanced Lexis surfaces of cohort mortality trends among males in France

a) Cohort line width standardized by the biggest size ever recorded for each cohort

b) Cohort line width standardized by biggest size recorded for cohort born in 1960

Notes: In Figure 5b line widths are set relative to the biggest size recorded for the cohort born in 1960. In this example, the biggest size was recorded at the age of 11 years. See standardization 2 in the data and implementation section.

Source: HMD, own calculations.
Finally, in Figure 6 and Figure 7, we provide an example of first-order differences in mortality, with the focus on baby boomers in the United States. Figure 6 shows the conventional Lexis surface, while in the enhanced Lexis surface in Figure 7 we standardize the line widths relative to the cohort born in 1960 (standardization 2). The biggest size of this cohort was recorded at age 1. The color green denotes a mortality rate reduction and purple denotes a mortality rate increase relative to the preceding year. Turning first to the conventional Lexis surface in Figure 6, we can detect in the period between 1975 and 2000, at ages up to 50 years, green and purple clusters of improvement and deterioration. These appear to be linked to certain periods and seem to follow a diagonal cohort pattern. However, only when we turn to our enhanced Lexis surface in Figure 7 does it become visible that this cohort pattern is centered on the large baby boom cohorts. This finding is suggestive of period events having a particular impact on the mortality of the large baby boomer cohorts. Such visual explorations with enhanced Lexis surfaces can then be further explored with more detailed data and analyses.

**Figure 6:** Conventional Lexis surface of changes in cohort mortality rates among males in the United States (log-scaled difference compared to preceding year)

Source: HMD, own calculations.
Figure 7: Enhanced Lexis surface of changes in cohort mortality rates among males in the United States (log-scaled difference compared to preceding year)

Notes: Line widths are set relative to the biggest size recorded for the cohort born in 1960 (standardization 2). In this example, the biggest size was recorded in 1961, when this cohort turned 11. Source: HMD, own calculations.

4. Discussion and conclusion

Our aim in this paper is to motivate and present enhanced Lexis surfaces and to demonstrate how they allow us to overcome limitations in the interpretation of conventional Lexis surfaces. Adjusting the Lexis parallelograms according to cohort size information at a specific age can provide the reader with a better understanding of which parts of a population or a cohort were experiencing specific mortality levels at certain ages. The use of this approach gives the reader a more intuitive understanding of why high mortality levels in the first years of life have historically been much more relevant to the overall mortality experience of a population than high mortality levels at higher ages. In addition, we demonstrate that adding the cohort size dimension to the plot helps the reader better understand how differences in cohort sizes can affect the mortality experiences of cohorts (e.g., Easterlin effects). Our visualization tool with enhanced Lexis surfaces will allow researchers to explore these dimensions in a more
intuitive way than is possible using tools based on conventional Lexis surfaces. In the future, we intend to expand the available indicators to cover fertility and migration data as well.
References


Human Mortality Database (2019). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at: www.mortality.org or www.humanmortality.de (data downloaded in July, 2019).


