Chapter 2
The Lexis Diagram

To look at 20,000 numbers and draw out their meaning
is a major research enterprise in itself.
Yet on the methods used in [Vaupel et al. (1985a)]
all that information is contained in a single contour map.

Nathan Keyfitz in his foreword of Vaupel et al. (1985a).

Any dynamics in vital events such as births and deaths involve change over calendar time, age, and/or cohort. The so-called Lexis diagram represents the ideal canvas to illustrate such dynamics. The Lexis Diagram as we use it today consists of a Cartesian coordinate system where calendar time ("period") is depicted on the x-axis and age on the y-axis (see Fig. 2.1 on page 6).[1] We added horizontal and vertical reference lines to facilitate orientation.

Birth cohorts move in such a diagram along the 45° line since a person is 1 year later 1 year older. Expressed differently: The current age of a person can be calculated if we subtract the birth date from the current calendar date. We used the example of three eminent demographers of the twentieth century in Fig. 2.1 to illustrate this relationship: William Brass, Ansley Coale, and Nathan Keyfitz. To be able to follow the cohorts on the 45° line, we made sure in Fig. 2.1—as well as in all other figures in this monograph—that the aspect ratio maps the length of one calendar year to exactly one age year.

Of course, we are not restricted to depict individuals on the Lexis plane. The standard approach is, indeed, to use population level data. It is obvious that we can not draw lines for every individual in that case. Colors are used instead to indicate the same value for the chosen statistic. While most figures in the remaining chapters show (smoothed) age-specific mortality or its time derivative, we opted to illustrate the basic approach of Lexis surface maps by depicting the population size of the

[1]It should be noted that the Lexis diagram can be considered to represent an example of "Stigler's law of eponymy" that states "No scientific discovery is named after its original discoverer." Please see Vandeschrick (2001) for a discussion about the problem of calling the diagram used in this book a "Lexis diagram".

R. Rau et al., *Visualizing Mortality Dynamics in the Lexis Diagram*,
The Springer Series on Demographic Methods and Population Analysis 44,
DOI 10.1007/978-3-319-64820-0_2

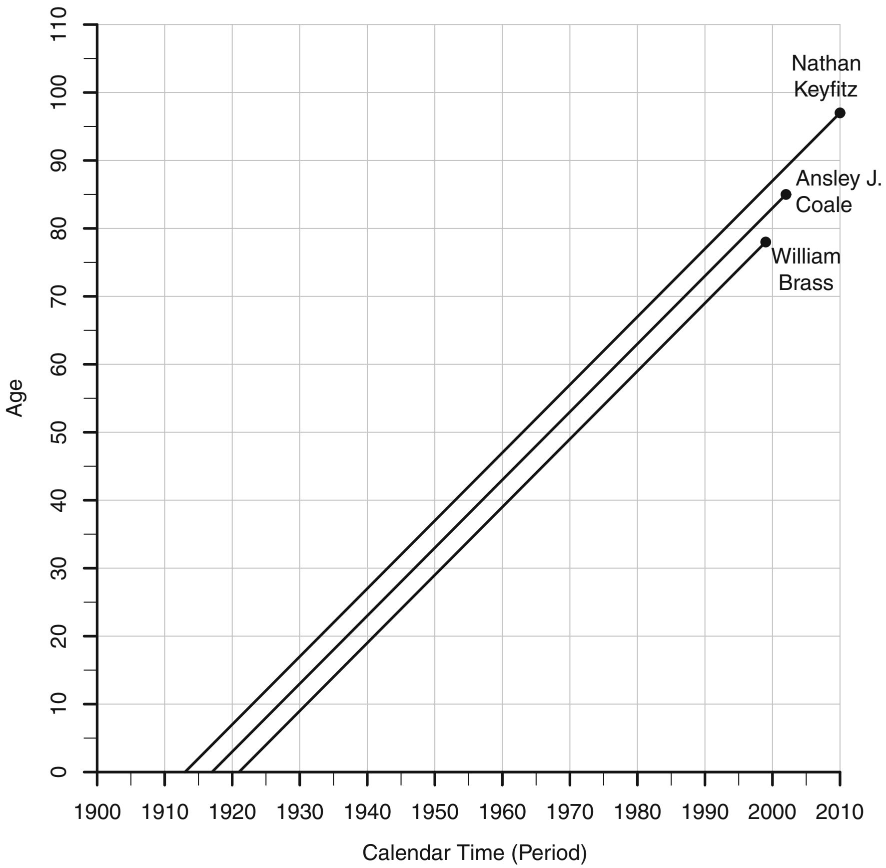

Fig. 2.1 An example of a Lexis diagram with individual life lines for William Brass, Ansley J.Coale, and Nathan Keyfitz

United States for women and men combined from 1900 until 2010 for ages 0–110 in Fig. 2.2. Thus, we have $111 \times 111 = 12{,}321$ individual datapoints. They are less than the 20,000 mentioned by Keyfitz in Vaupel et al. (1985a) but considerably more than the median number of entries in data matrices for statistical graphics found by Tufte (2003) in various scientific and non-scientific publications. Tufte—who was described as the “da Vinci of Data” by The New York Times (Deborah 1998)—states in a related book (Tufte 2001, p. 166): “Data graphics should often be based on large rather than small data matrices and have a high rather low data density. More information is better than less information, especially when the marginal costs of handling and interpreting additional information are low, as they are for most graphics.”

In our Lexis maps we employed a color scheme reminiscent of geographic maps where green colors indicate lower values and brown colors are used for high “altitudes”. Analogously to standard maps, we added contour lines to emphasize areas of equal elevation, which translates to the same number of people in our figure.

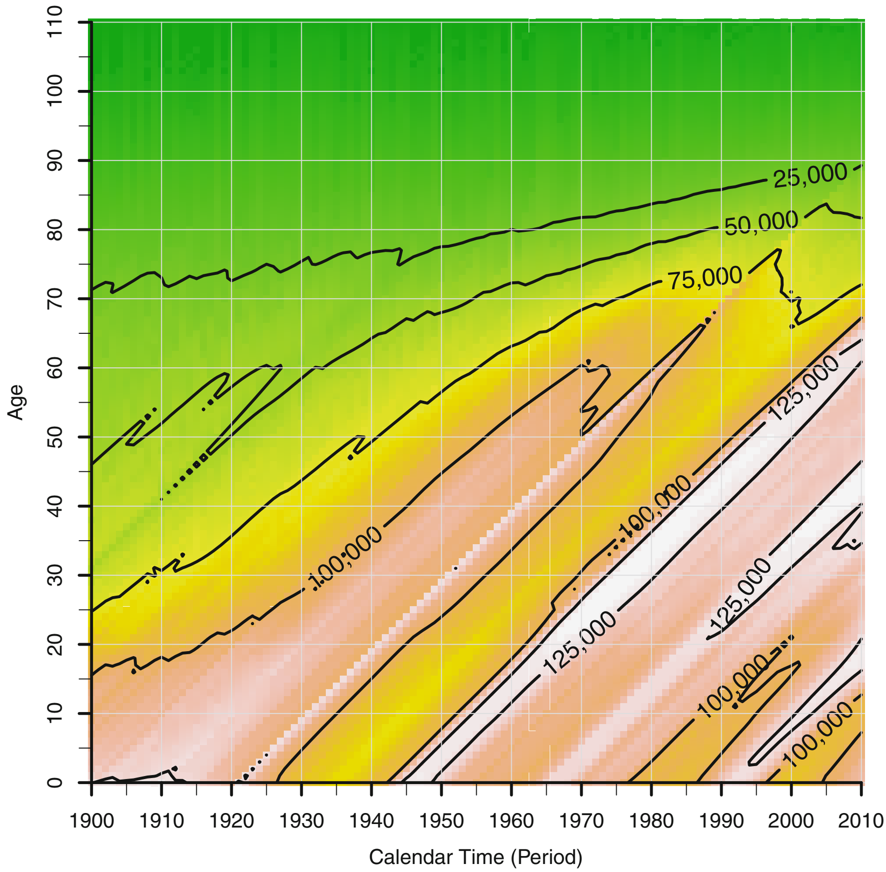

Fig. 2.2 An example of a Lexis surface depicting the population size of the United States by calendar year and age (Source: Own illustration based on data from the Human Mortality Database 2017)

Depicting mortality, fertility or other population characteristics in the Lexis diagram provides a useful framework to analyze data for the presence of age-, period-, and cohort- ("APC") effects. The major problem of standard statistical approaches (e.g., regression analysis) in this area is the so-called *Identification Problem*, which refers to the perfect correlation of age plus cohort equaling period. Various methods have been introduced (e.g., constraining the parameters in a regression setting) but "there is no magic solution" (Wilmoth 2006, p. 235).[2] With our surface maps, we suggest instead a graphical approach that can be used for questions such as "[w]hether mortality improvements takes place by cohorts or by periods" (Keyfitz in Vaupel et al. 1985a, p. ix).

[2]Please refer to this article also for a systematic overview of APC models used in demographic research.

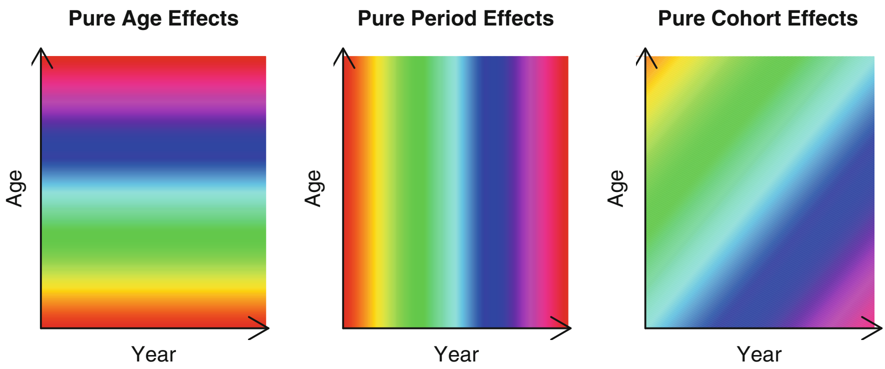

Fig. 2.3 "Ideal" age-, period-, and cohort-effects on the Lexis surface

Figure 2.3 gives an overview how age-, period-, and cohort effects would ideally look like on the Lexis surface. The same color indicates the same value in the variable of interest (e.g., death rates). The left panel represents "pure" age effects. That means that the only variation in the variable of interest takes place across the age dimension, regardless of calendar year or cohort. The panel in the middle denotes "pure" period effects, i.e., the same values are measured at all ages but they differ along the calendar time/period dimension ("Year"). Finally, the panel on the right illustrates how a surface map would like if (birth) cohorts alone were driving the development in the variable of interest. The same color along the 45° line shows that each cohort has their own characteristic value of the variable of interest, which does not change throughout their life course. Obviously, those are idealized and simplified representations. We expect to find rather interactions of these three forces than such "pure" effects. Furthermore, we should acknowledge the biggest drawback of our method: In contrast to other methods of APC analysis, our visual approach does not attribute any numerical value to each of those effects. Hence, one can neither compare various effects with each other nor is it possible to conduct significance tests that are typical of regression analyses and other standard methods in statistics.

We are not the first to illustrate demographic phenomena in three dimensions, i.e., *either* on the Lexis plane using colors to indicate the third dimension *or* by wireframe plots. An interesting overview of the history of such "Frequency Surfaces and Isofrequency Lines" is given in Caselli and Vallin (2006). They cite the example of Luigi Perozzo's depiction of the change in the Swedish age pyramid in 1880, based on a diagram by Gerard Van Den Berg (1860), as one of those earliest examples. We have reproduced Perozzo's diagram in Fig. 2.4. About 60 years later, Pierre Delaporte used such wireframes to depict French mortality (1938) and contour lines for European mortality (1942).

An explicit case of using such plots to separate age-, period-, and cohort-effects from each other can be found in Thomas Pullum's article on US fertility published in 1980. A few years later, the population program at the International Institute

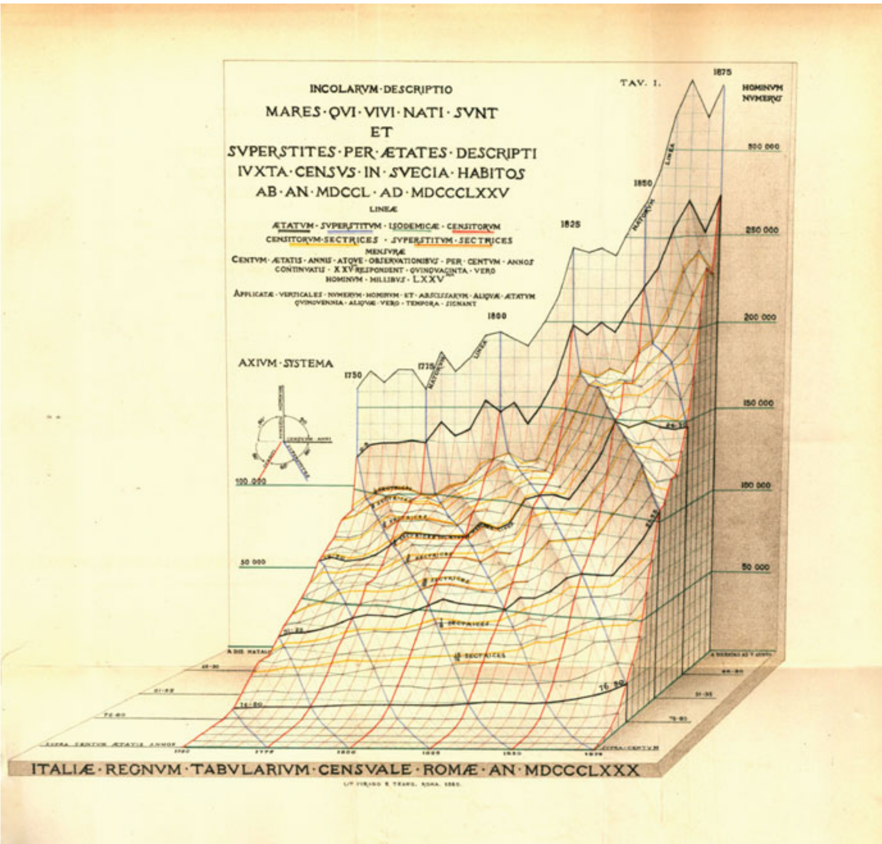

Fig. 2.4 Change in the Swedish age pyramid as depicted by Luigi Perozzo in 1880 (Source: Timothy Riffe, with kind permission)

for Applied Systems Analysis (IIASA) in Laxenburg in Austria turned out to be an incubator for advancing the display of population dynamics on the Lexis plane in the 1980s. Vaupel, Yashin, Caselli, and others introduced colored/shaded contour maps to depict, for example, population size, mortality, or birth rates (e.g., Vaupel et al. 1985a,b, 1987; Caselli et al. 1985; Gambill and Vaupel 1985). The "democratization" effort described in the introductory chapter was also mirrored in the late 1990s for Lexis surfaces: Kirill Andreev developed not only the user-friendly software Lexis to analyze demographic trends in Denmark and other highly developed countries (Vaupel et al. 1997; Andreev 2002). He also shared it freely with anyone interested.[3] Despite being a milestone for the creation of Lexis surface

[3]While writing his Master's thesis, the first author of this monograph received the Lexis software from Kirill Andreev simply via email in early 2000.

maps, almost no one is using it anymore. The aforementioned specialized languages such as Matlab (Mathworks 2017) or R (R Development Core Team 2015) have become the favorite tools nowadays along with Python (van Rossum 1995). With the exception of the reproduction of Perozzo's plot all figures in this monograph were created with R as we will explain in Sect. 3.2 and in the appendix, starting on page 161.