Decomposing and recomposing the population pyramid by remaining years of life

John MacInnes[1], Tim Riffe[2], and Jeroen Spijker[1]

[1]School of Social and Political Science, University of Edinburgh
[2]Department of Demography, University of California, Berkeley

November 15, 2013

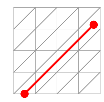
Abstract

We apply a simple decomposition method to age-classified population counts in order to derive a probabilistic distribution of remaining lifetimes within age-classes. This decomposition method allows us to invert the population pyramid so as to be classified by remaining years of life. This paper is a graphical exploration of the kinds of information that such an image can yield. We think that this age-specific remaining-years decomposition should become part of the basic demographer's toolkit for many reasons. We also suggest a rich visualization for consideration as a basic product for consumers of demographic information.

The population pyramid evokes graphical inference about several aspects of demography that go beyong population structure itself. A top-heavy pyramid  suggests that the population is aged and aging, while a pyramidal pyramid  suggests that the population had and has high fertility. By association with high fertility we might infer high mortality. From other details in the profile we infer past wars and famines . Age-heaping  suggests an innumerate or superstitious population, and so we silently conclude that the population also has a lower average educational attainment, and expect to see a wide base. Such judgements are of course subject to the bias of the beholder, but we must admit that these are the kinds of thoughts that a pyramid evokes — the kinds of things we look for.

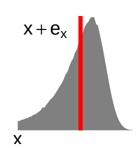
Age (and sex) is of technical use as the primary classifying variable for demographic rates. Rates are segmented by age because demographic phenomena show stable and consistent patterns over age; in order to make a statement about the population as a whole, it is held, one must purge rates of age structure. Of course, if age-specific rates themselves depend on age-structure, then we have a vicious circle, as Stolnitz and Ryder (1949) so eloquently pointed out. Age virtually always results significant in well-specified statistical models of demographic phenomena. We conclude that age is paramount to both parlour demography and serious demography.

It therefore behooves the demographer to be explicit about the meaning of age. We use age to explain other phenomena, and these demographic forces lend phenomenological meaning back to age. Here we are just interested in the integer value itself. We all know that age is counted as the number of years that have passed since the event of birth. Age is in this way backward-looking. For this reason we consider the pyramid to be a reflection of the past. Age-structured demographic models therefore have this implicit connection to the past, as do statistical inferences on age. That the lifeline traversing a Lexis diagram also has an endpoint signifying death is all too obvious, as is the idea that the length of this lifeline is proportional to the length of life. Time of birth is known, but time of death is not known ? (until it's too late), and so we cannot measure lifelines directly for living populations.



The most widely applied approximation to lifeline endpoint is the so-called remaining life expectancy, e_x, which is an average remaining lifetime for the population of a given age, x, according to some particular assumptions about future mortality in the aggregate. Hersch (1944), Ryder (1975) and Sanderson and Scherbov (2007) have made good use of the remaining life expectancy column of the lifetable to gain insight into a populations' potential future. The latter authors have elaborated a valid demographic perspective, that of remaining years of life, as a way of looking at age. Instead of counting up from birth, one may conceive of counting down until death — one's *prospective* age as opposed to chronological age. At least as early as 2001, Ken Wachter coined the term *thanatological* age[1] to refer to this manner of forward-looking age. These two terms are referring to the same thing: "Prospective" highlights the projective or potential nature of this view on age, while "thanatological" highlights the orientation along the lifeline.

Miller (2001) offers another tack on approximating remaining years of life for a given age group. Rather than assigning a single value, e_x, to all members in an age group, we recognize the full distribution over potential future death times, as inferred using information from the lifetable. This makes sense in general because it gives us more information, but also because the distribution around e_x of other possible death times is usually not symmetrical. When making generalizations about population aging based on remaining years of life, we obtain a closer approximation in summing over specific thanatological age-classes than by splitting population counts over e_x.



Formulas are cleaner when written in continuous form. Say we have population counts of exact age a, $P(a)$, as well as the hazard function, $\mu(a)$. The survival function, $l(a)$, is related to $\mu(a)$ by $l(a) = e^{-\int_0^a \mu(b)\, \mathrm{d}b}$. We then define the population of exact age a with

[1] Thanatos was the Greek god of death.

exact remaining years of life, y, $P(a, y)$ as

$$P(a, y) = P(a)\mu(a + y)\frac{l(a + y)}{l(a)} \quad , \tag{1}$$

where $\mu(a + y)$ is the instantaneous probability of death at age $a + y$ given survival to age $a+y$ and $\frac{l(a+y)}{l(a)}$ is the probability of surviving to age $a+y$ given survival to age a. Summing (1) over a for a given y will lead us to $P(y)$, the population with exact thanatological age, y. Under fixed mortality conditions, this relationship is exact.

To make use of (1) with real data we need a decent discrete approxiamtion, and so must compromise on aesthetics. Given single-age data, where x stands for the single interval $[x, x+1)$, take age-classified population counts, P_x, and a radix-1 single-age lifetable that describes the mortality experience of this population well-enough. Start either with survivorship, l_x, or the death distribution, d_x[2] and decompose P_x into the population cross-classified by chronological age, x, and thanatological age, y, $P_{x,y}$:

$$P_{x,y} = P_x \cdot \frac{d_{x+y} + d_{x+y+1}}{l_x + l_{x+1}} \tag{2}$$

Other discretizations are imaginable, but (2) gives useful enough results. In carrying out (2) for each x and all possible values of y we have already decomposed P_x by remaining years of life. The result of this exercise is a matrix with population counts cross-classified by chronological age and thanatological age. One can blend out chronological age altogether

$$P_y = \sum_{x=0}^{\omega} P_x \cdot \frac{d_{x+y} + d_{x+y+1}}{l_x + l_{x+1}} \quad , \tag{3}$$

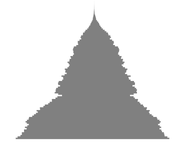
which can be thought of as *recomposing* what we have just decomposed by this new variable, y. If our assumptions about future mortality are not too far off, we have a fairly good approximation of thanatological age structure, P_y. If adhering to the synthetic cohort approach, the resulting thanatological age structure is also in a way synthetic — an indicator of the year t mortality conditions combined with the year t population structure.

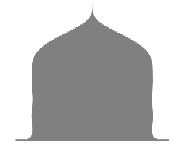
A great many practical and theoretical results will likely fall out of playful interaction with thanatologically transformed data, some of which have been spelled-out in Riffe (2013). Here we interest ourselves directly with the visual inspection of this decomposition-recomposition — in how remaining lifetimes are distributed over age and how chronological age is distributed over remaining lifetimes — via the population pyramid.

Our objective is to build the same kind of demographic associations with thanatological age-structure that we derive from inspection of traditional population pyramids. At a glance,

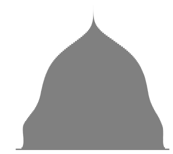
[2] That is to say, derive l_x from d_x or vice versa, rather than grabbing each from a published lifetable. This way the redistribution is free of rounding error.

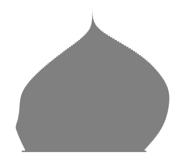
what does a given profile suggest about the past and future of mortality and fertility for a given population? The two population structure profiles in the margin, A and B, compare our two orientations of age for a particular year and population. One recognizes thanatological population structure first for its relative smoothness. An artifact of the transformation method is that roughness is blended out due to the staggering of ages. Of the two pyramids in the margin, B is much smoother than A. The traditional pyramid, A, represents a population that either is growing very rapidly, or that has high fertility and high attrition, or in this case some of both. The extruding bottom rung of B indeed contains all ages, but this kind step only arises when the first year or two of life has high mortality. From B, we see that if vital conditions persist, the relative size of attrition (at least for those beyond infancy) does not really change from year to year, despite the aging of large cohorts at the bottom of the age pyramid, A. If conditions are roughly constant, we conclude that the growth rate for this population has been a bit higher than zero. Of course, such intuition can lead us astray, as a population's vital rates may be in flux. A and B both represent 1900 Sweden, with male and female life expectancies at birth of 50.79 and 53.62, mean remaining life expectancies of 46.36 and 43.12 (the weighted average of thanatological age), and mean (chronological) ages of 34.74 and 36.32, respectively. This was a young and growing population, at least according to its period indicators.

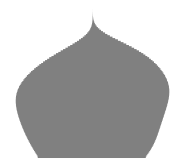

A


B

A mortality crisis year, such as 1918, will cause an abrupt sag in the thanatological age structure with respect to surrounding years. WWI and WWII caused the male (left) side to sag more than the female side in some countries, for a rather lop-sided pyramid. In recent years, the typical profile of contemporary thanatological age-structure for most low-mortality countries resembles the average. Populations that have recently undergone a rapid mortality decline, but are still young will have very narrow bases and very high centers of gravity. All contemporary (postwar) populations have maintained such a taper at the base — instead mortality gains have tended to both slow and disperse the advancement of large birth cohorts toward the lowest thanatological ages. These profiles can be enriched to yield more information.


SE, 1918

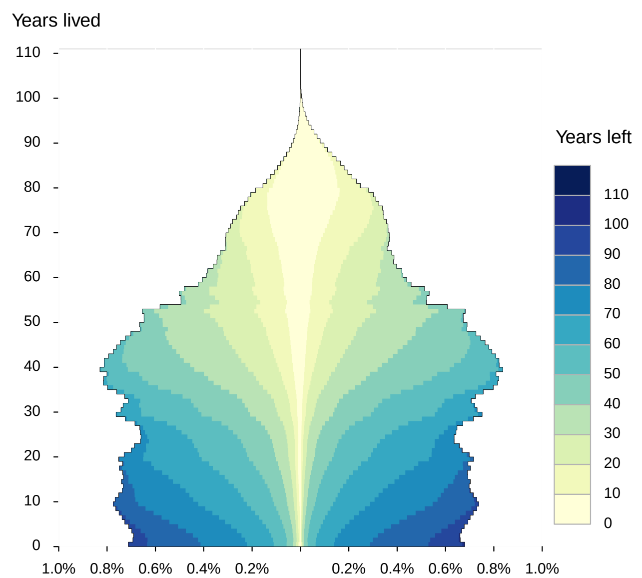

ENW, 1944


Avg. contemp.

The outer profile of Figure 1a is the year 2000 US population, as typically depicted by age and sex. By following (2) and then aggregating y into 10-year groups within each x, we can add a layer of information to this figure. One gains an overview of the heterogeneity of remaining lifetimes within each age. This excercise offers a manifold gain in resolution for the calculation of indicators on population aging, health, indirect indicators of dependency, and even just a simple idea of the likely distribution of future lifetimes for a given age-group. According to the color gradient used here, a darker pyramid indicates a population with many years to live and vice versa. If temporal proximity to death can be thought of as a useful indicator of morbidity or disability, one might here gauge how these states are distributed over age.

Figure 1: US population structure, 2000 (HMD)

(a) Chronological age structure with years left indicated by shades



(b) Thanatological age structure with years lived indicated by shades

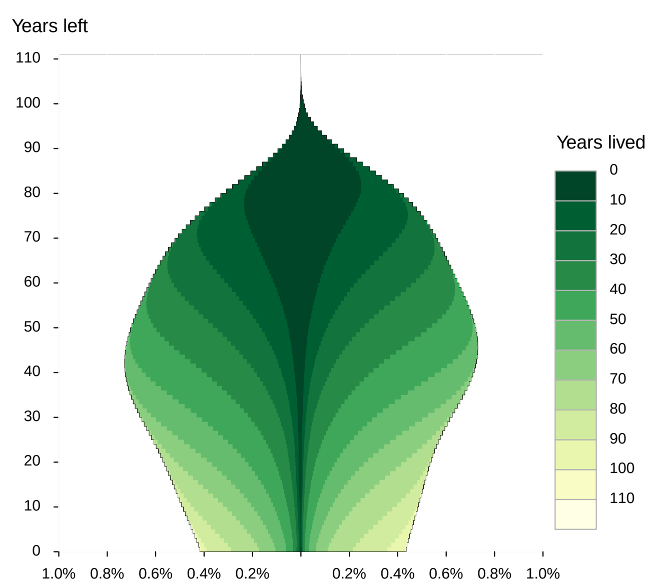


Figure 1b is based on the transformation equation (2) as well, but instead we have turned the population on its side and aggregated chronological age into 10-year groups. Now we see single-age resolution of thanatological age-structure (purely probabilistic), and an idea of the time-since-birth heterogeneity within each remaining lifetime. The bottom layer in this depiction, which resembles a leaf[3], is expected to decrement within the year. A top-heavy leaf indicates a population whose members have many years ahead. According to our gradient, a dark leaf is young and a light leaf is older. Population leaves are sensitive to mortality conditions, and history has known them to take different characteristic shapes.

The best way to develop some visual heuristics for inferring vital conditions from popualtion structure is to examine some standard population transformed under different fertility and mortality assumptions. To be more systematic about it, we derive the stable population structure, $c(a)$, which can be defined easily in terms of survival and a growth rate, r:

$$c(a) = \frac{e^{-ra} l(a) \, \mathrm{d}a}{\int_{a=0}^{\infty} e^{-ra} l(a) \, \mathrm{d}a} \tag{4}$$

In this case r stands in imperfectly for fertility, but the main conclusions will still follow. To derive the same for both sexes at once, we share the same r, but assign sex-specific survival and then rescale each sex to account for the sex ratio at birth[4]. Figure 2 displays some pyramids that will likely be familiar to the demographer,[5] as they represent chronological age strucuture. In growing populations, newer generations at the base are larger than older generations further up the pyramid, no matter what the mortality. High mortality squuezes the pyramid down, and low mortality stretches it higher. Most of the World's populations will fit into these ranges of r and e_0, although they will not necessarily approximate the stable form.

One way to derive the stable thanatological age structure, $c(y)$, for a given r and a given set of mortality conditions is to simply do the thanatological redistribution described earlier on the stable chronological age structure, $c(a)$.[6] Figure 3 displays the results of this exercise for the same standard population under the same constraints as Figure 2. The conclusion is perhaps obvious: in a stable population, the thanatological age structure is approximately equal to the chronological age structure after switching the sign of r. It turns out that this relationship is only approximate, but it is still a good rule of thumb. The middle row, a stationary population, is in fact exactly equal for chronological and thanatological populations, but we already knew that this would be the case (Vaupel 2009).

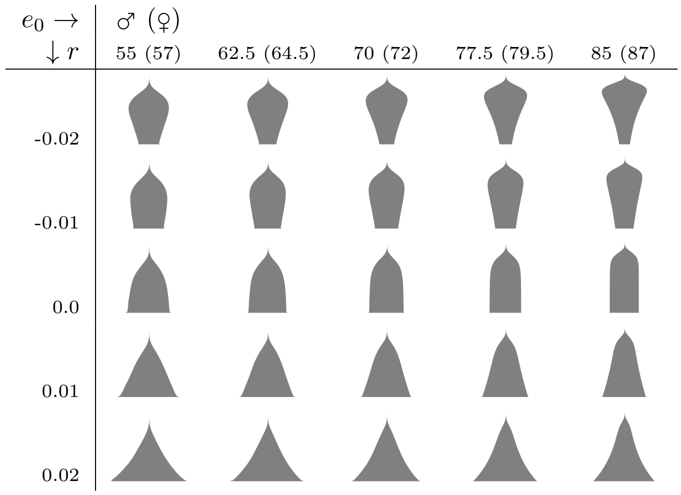
[3]For the botanically-minded, the thanatological age-sex-structure of low-mortality populations almost always takes the shape of a truncate deciduous leaf with entire margins, and, when chronological age is also highlighted in the leaf, as here, arcuate veination.

[4]There are more cumbersome details here that are not worthy of taking center-stage, but we do intend to put all code in a public repository.

[5]The reference male and female survival functions are for 1950 Sweden, for no particular reason. Data come from the HMD.

[6]$c(y)$ can also be derived directly using a thanatological variant of (4), discussed elsewhere.

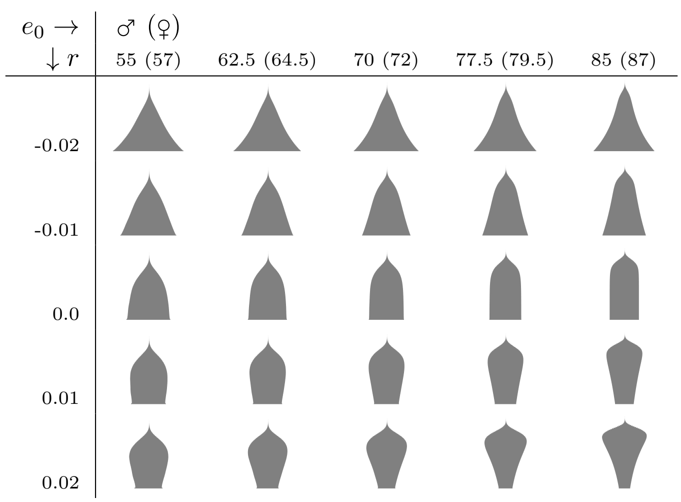
Figure 2: Stable chronological age structure under different mortality conditions and growth rates.



This mirror relationship is a good reference, although it is only exact in the case of a stationary population, and the relationship weakens as r moves away from 0. There is still a direct relationship between the two age structures, as described above, and it is reversible if the population is stable.[7] If the population is not stable, as is the case for vitrtually all human populations, chronological age structure cannot be inferred from thanatological age

[7] The transformation from thanatological to chronological stable age structure is rather ugly, and we've chosen to omit it.

Figure 3: Stable thanatological age structure under different mortality conditions and growth rates.



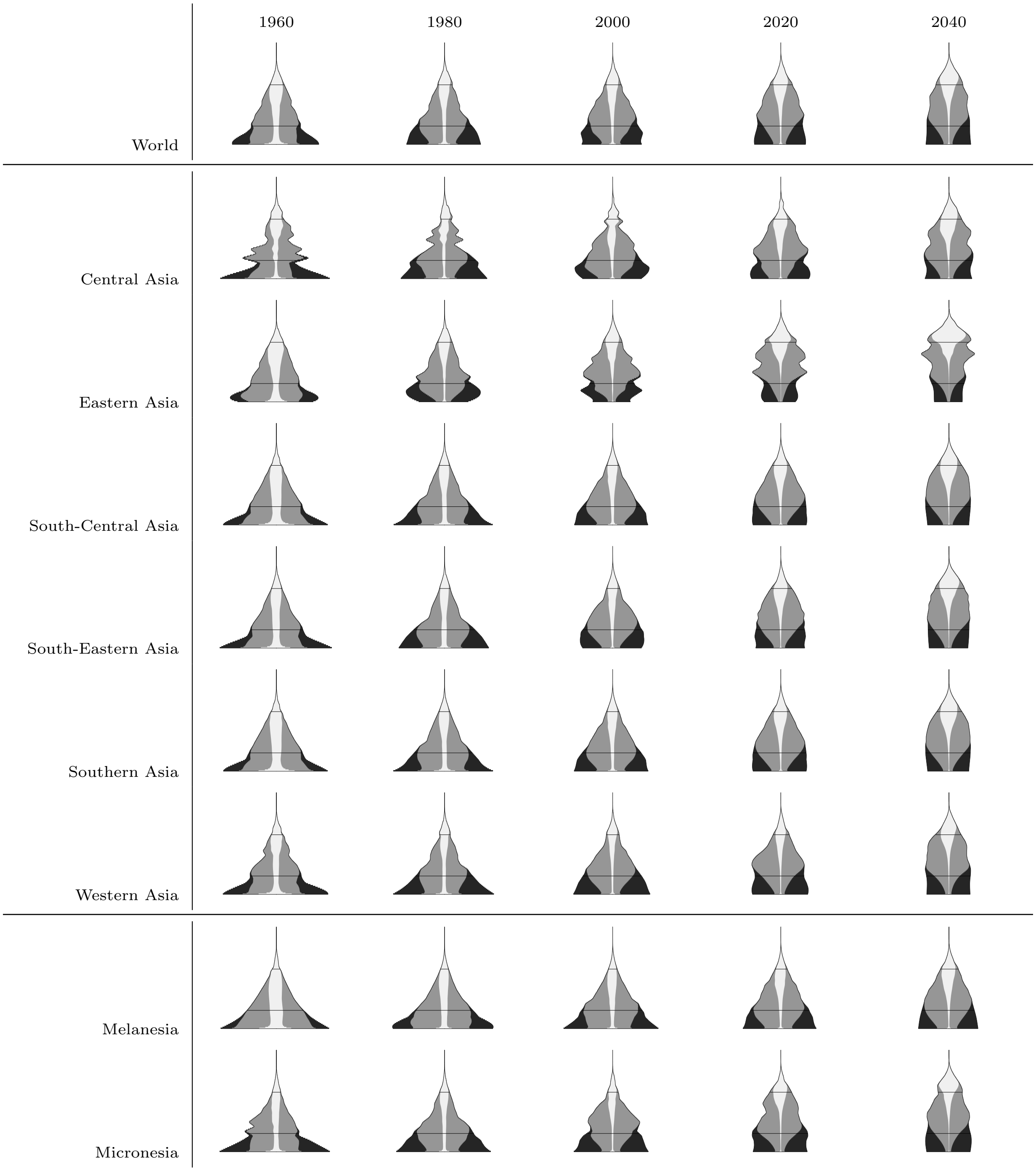
structure.

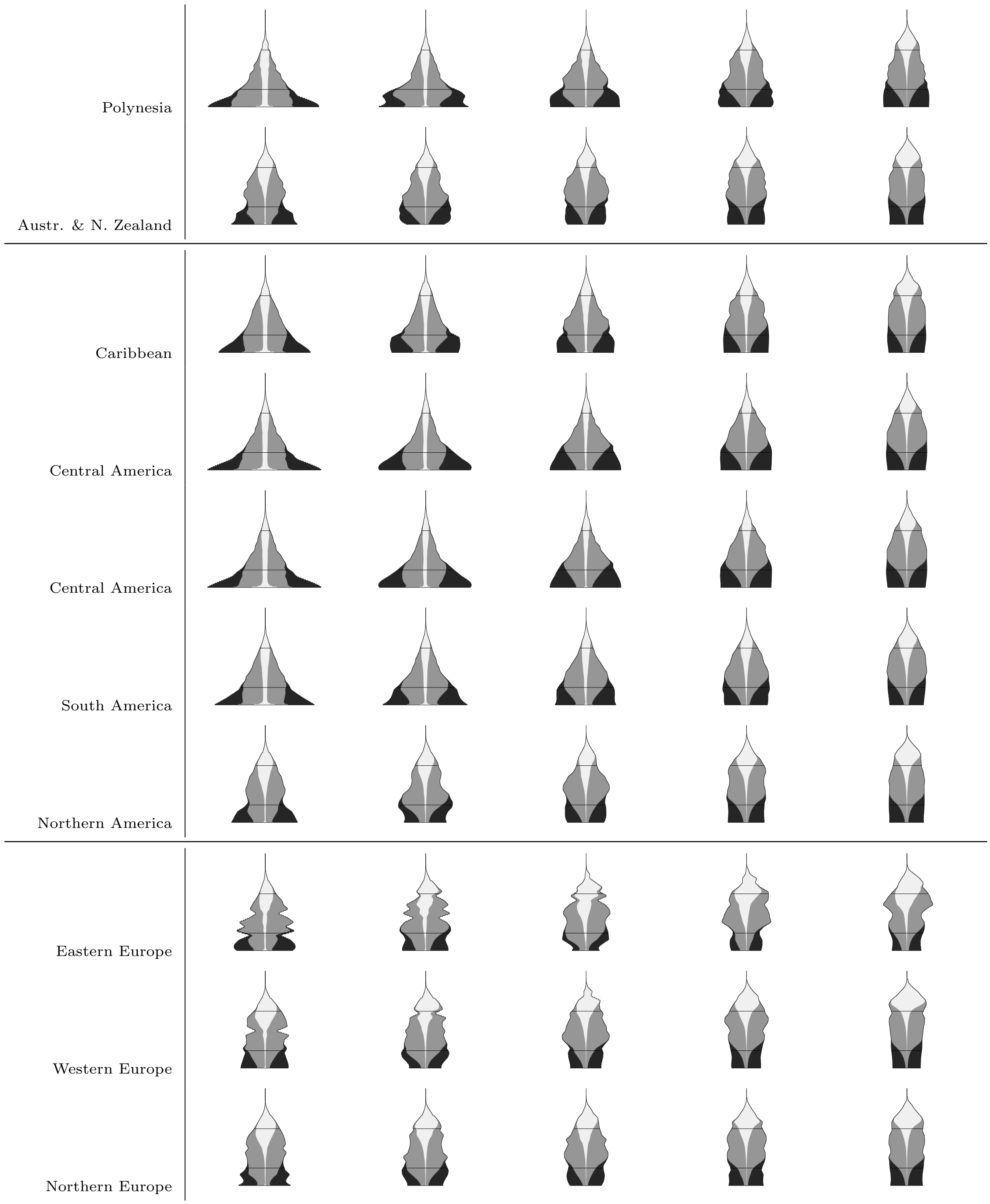
The rules of thumb based on stable population structure are more didactic than handy, given the vital changes that human populations underwent in past decades, or are projected to go through in the coming decades. Different stages of the demographic transition — their order, magintude, spacing, and velocity — combined with abrupt period shocks and the full breadth of human diversity, will create a large stock of potential age-structures that human populations have or are likely to obtain. At present, the most comprehensive and up-to-date source of data for population stocks and mortality data is the 2012 UN Population Prospects.

*This is a work in progress. We'll relate the following pyramid tables with time-series of fertility and mortality indices to help build associations and generalize — still being designed.

* these annexes are not final, as they lack legends and explanation.

Table 1: UN regional groupings, chronological age-sex pyramids




Polynesia
Austr. & N. Zealand
Caribbean
Central America
Central America
South America
Northern America
Eastern Europe
Western Europe
Northern Europe

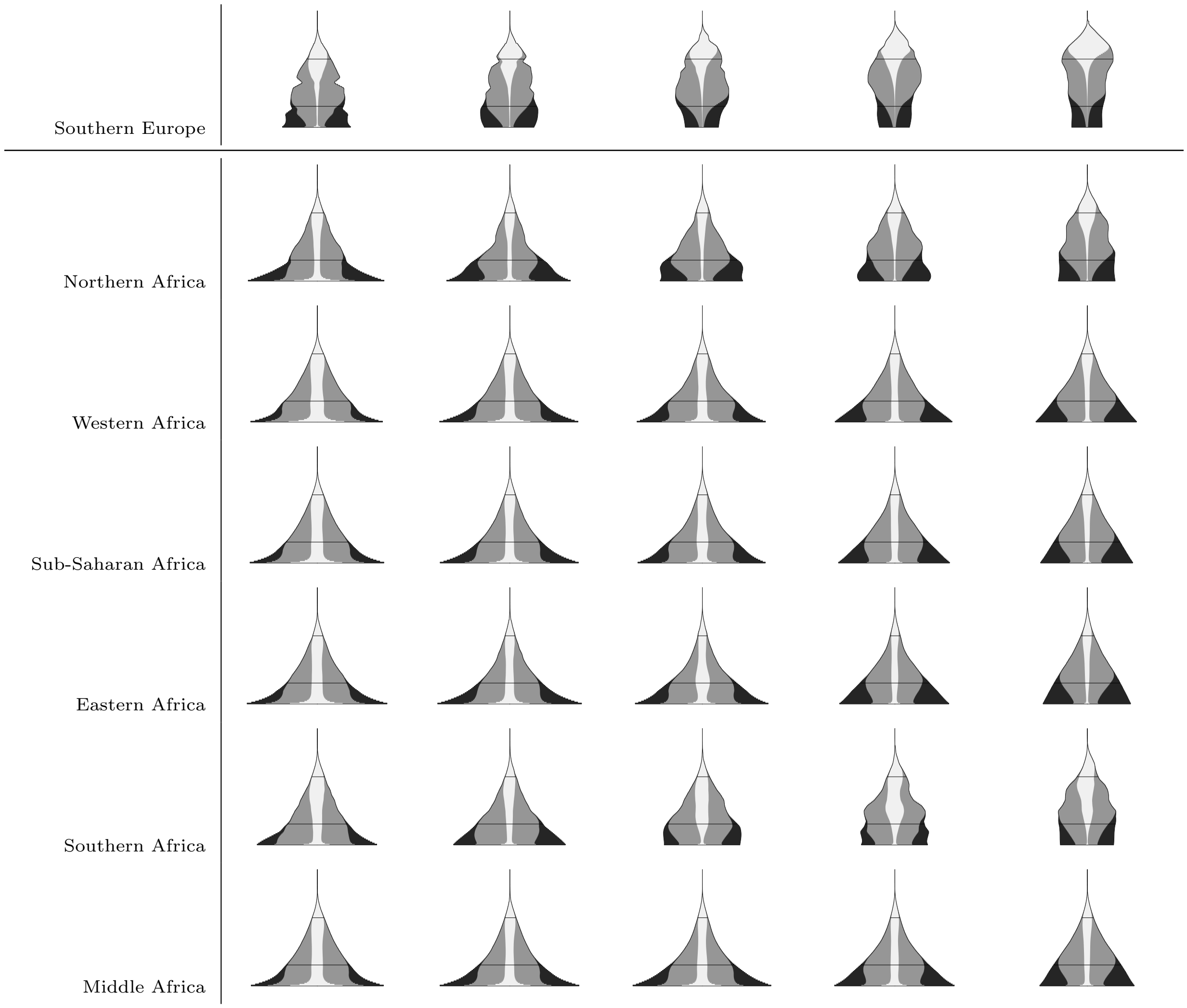
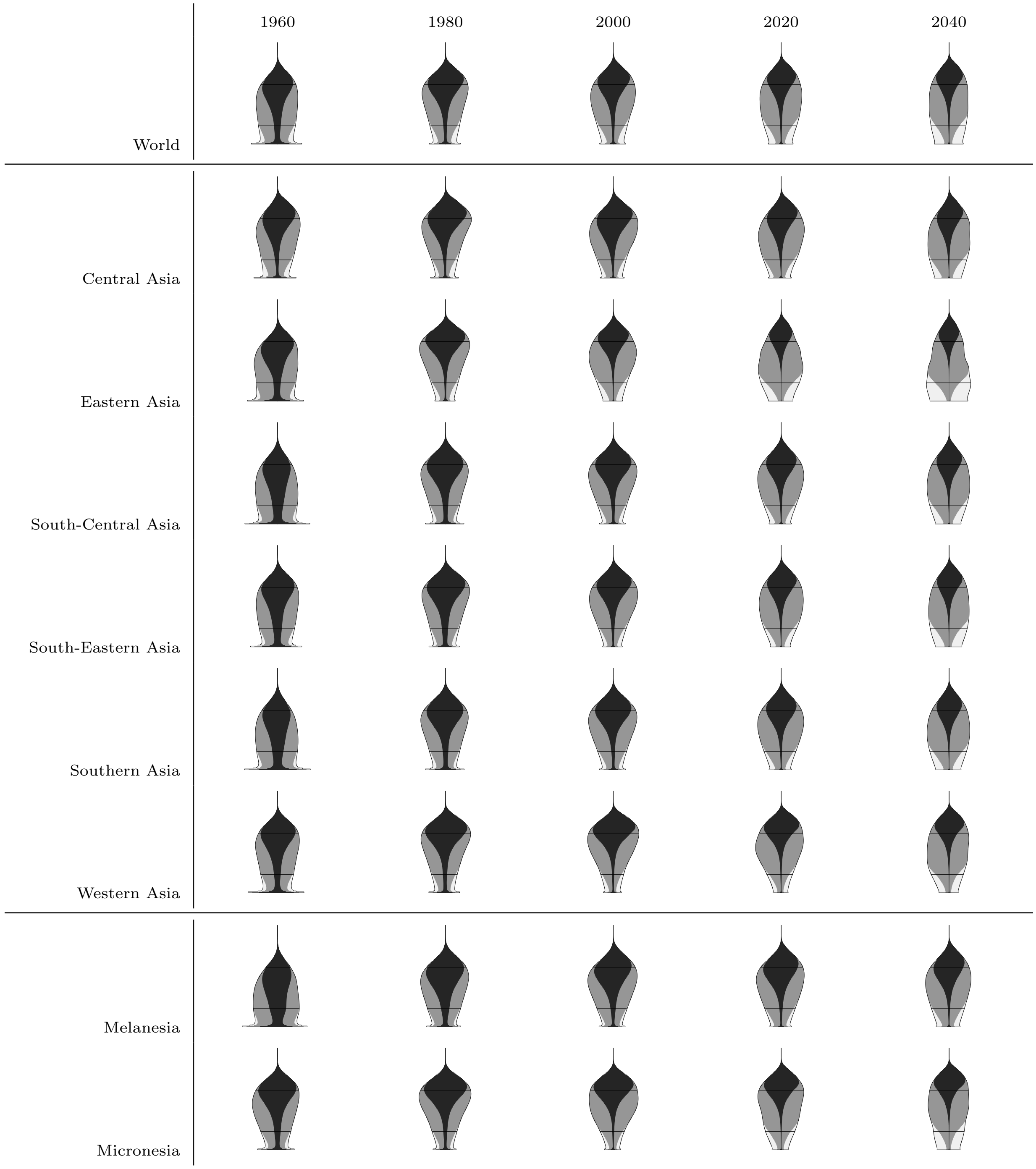
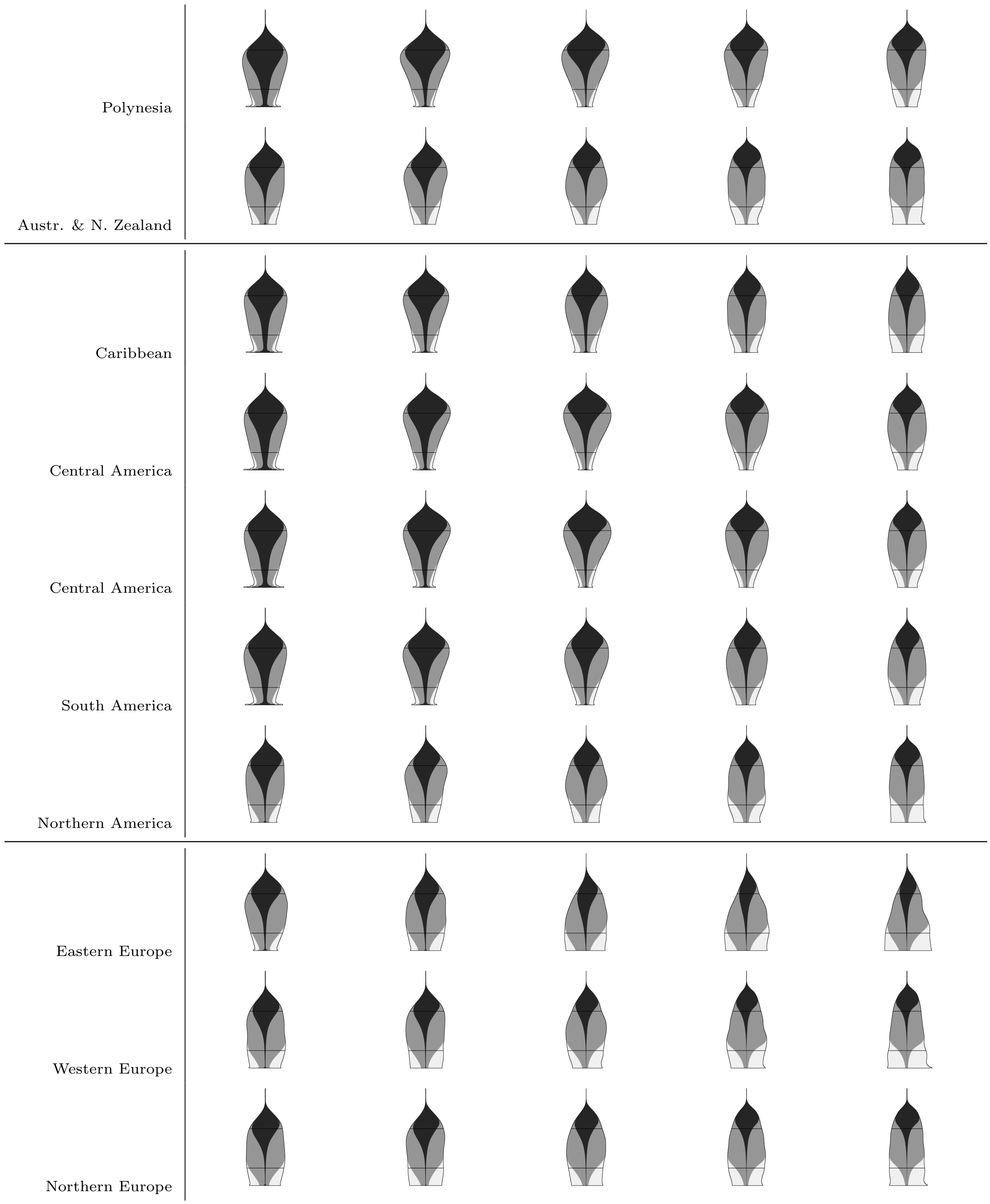
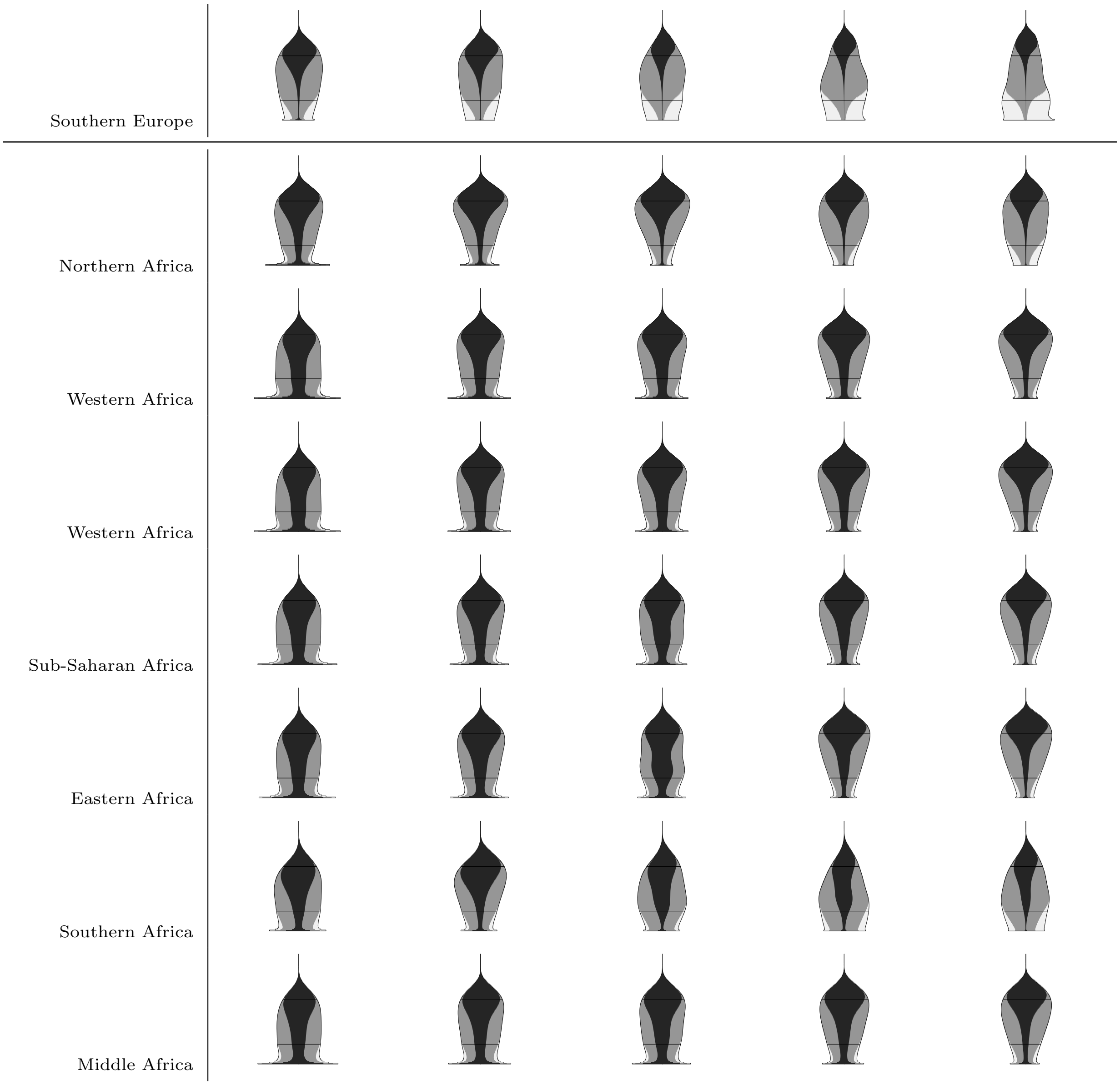

Southern Europe
Northern Africa
Western Africa
Sub-Saharan Africa
Eastern Africa
Southern Africa
Middle Africa

Table 2: UN regional groupings, thanatological age-sex pyramids




Polynesia
Austr. & N. Zealand
Caribbean
Central America
Central America
South America
Northern America
Eastern Europe
Western Europe
Northern Europe


Southern Europe
Northern Africa
Western Africa
Western Africa
Sub-Saharan Africa
Eastern Africa
Southern Africa
Middle Africa

References

Liebmann Hersch. *De la démographie actuelle à la démographie potentielle.* Georg & C SA, Librairie de l'université Genève, 1944. Extrait des Mélanges d'Etudes économiques et sociales offerts à M. William E. Rappard.

Tim Miller. Increasing longevity and medicare expenditures. *Demography*, 38(2):215–226, 2001.

Timothy LM Riffe. *The Two-Sex Problem in Populations Structured by Remaining Years of Life.* PhD thesis, Universitat Autònoma de Barcelona, 2013.

Norman B Ryder. Notes on stationary populations. *Population Index*, 41(1):3–28, 1975.

Warren Sanderson and Sergei Scherbov. A new perspective on population aging. *Demographic research*, 16(2):27–58, 2007.

G.J. Stolnitz and N.B. Ryder. Recent discussion of the net reproduction rate. *Population Index*, 15(2):114–128, 1949. ISSN 0032-4701.

James W Vaupel. Life lived and left: Carey's equality. *Demographic Research*, 20(3):7–10, 2009.