

# Independent time measures

Tim Riffe\*<sup>1</sup> and Joel E. Cohen<sup>2, 3</sup>

<sup>1</sup>Max-Planck-Institute for Demographic Research

<sup>2</sup>The Rockefeller University

<sup>2</sup>Columbia University

October 14, 2019

## Abstract

**Background** The demographic time identity consists in a linear relationship between all pairs of six fundamental time measures: age, period, cohort, time to death, length of life, and time of death. Certain subsets of time measures in this and other higher order identities consist in time measures that are independent of one another.

**Objective** We aim to describe the relationship between independent time measures and the identities within which they are nested, with special attention to the demographic time identity and its three independent time dyads. We also aim to determine whether data structured on such time dyads might be useful in demographic research.

**Data and Methods** We illustrate concepts visually based on data from the Colonial Québec Population Register and from the US Health and Retirement study.

**Results** We show that a Cartesian mapping of independent time dyads form a redundant mapping of the same demographic timespace. These add three previously undescribed demographic perspectives to the four Lexis-like perspectives, for a total of seven proposed standard perspectives.

**Conclusions** We show independent sets of time measures to be consistent with the temporal identities within which they are nested. We demonstrate that these perspectives can be useful for pattern detection and characterization in demographic research.

## 1 Background

Dated events on a timeline form the foundation of numerous empirical approaches to detect, estimate, and describe patterns in processes and outcomes in the social sciences. Process patterns might be called transitions, and there are several varieties of these (continuous/discrete, conditional/unconditional, forward/backward (Brouard 2019), and more). Outcome patterns refer to either things such as prevalence, distribution properties (e.g. expectancies), sequences, episode statistics, and others. Outcome patterns can

---

\*riffe@demogr.mpg.de

either be calculated from a process model or observed directly from data. Examples of process patterns include age-structured rates of fertility, mortality, and migration. Examples of outcome patterns include trends in length of life, healthy life expectancy, counts of births structured in some way, or employment trajectories.

This paper deals with the calendar relationships that underpin all such knowledge of the world. Most empirical work that describes processes and outcomes makes use of a limited subset of the temporal structure that is available, and often the implications of relationships between time measures are unrecognized. Previous work (Riffe et al. 2017) formalizes the relationship between events in time, and the durations between them. A given set of event types determines a further set of durations. Taken together, a set of events and their implied durations comprise a dense linear relationship that one might call a temporal *identity*. The most common of these in empirical demographic research is the *Lexis* identity between period (date), cohort(date), and age (duration). There are countless Lexis-like relationships between linearly dependent time measures, and these can be combined into higher-order Lexis identities.

A primary example of a higher order Lexis identity is the so-called *demographic time identity*, which defines a relationship between age (A), period (P), cohort (C), time until death (T), time of death (D), and length of life (L). In describing the demographic time identity, Riffe et al. (2017) give an almost exhaustive account of the so-called *informative* and *uninformative* dyads. Informative dyads are those sets of two time measures (from APCTDL) that imply a third time measure via identity. For example, AP implies C. Four such 3-part (triad) identities are described at length, with examples and diagram depiction (APC, TPD, CPD, TAL). Uninformative dyads are sets of two time measures that do not directly relate, and which therefore do not imply a third time measure. The three examples given are LP, AD, and CT, which are depicted in the authors' Tab. (2) and discussed in the text as follows:

Triad identities are more meaningful than uninformative dyads. This is so even in the absence of data, due to the underlying relationship between measures. Each of the triad identities can accommodate some version of a lifeline, for instance (p5 Riffe et al. 2017).

Contrary to the above, in the present note we aim to show how uninformative dyads can reveal meaningful relationships in data. For better semantics and precision, we suggest calling these *independent* dyads rather than uninformative dyads. We relate independent dyad diagrams to lifelines in two dimensions (2d). We relate 2d independent planes to the three-dimensional (3d) demographic time identity, showing how these are also cross-sections of the same 3d volume, but which follow a cubic rather than a tetrahedral structure. We show how data structured on independent time measures may still reveal meaningful patterns using historical data from Québec. We speculate on a way to view and interpret data structured by the demographic time identity using arbitrarily defined planes as cross-sections. We speculate on how data structured by independent time measures may be meaningful in general.

## 2 Geometric relationships

The demographic time identity can be represented as a tetrahedral graph, with time measures mapped to the six edges of a tetrahedron. The four Lexis-like identities are

found on the faces of the tetrahedron. The identity fills a 3d space because tetrahedra tessellate with octahedra in a 2:1 ratio. Riffe et al. (2017) considered neither independent pairs of time measures nor the significance of octahedra in this geometric construct. Figure 1 shows the tetrahedral graph of the demographic time identity between age (A), period (P), cohort (C), time to death (T), death cohort (D), and length of life (L). Independent dyads are those that do not share any vertex in the graph, and these appear as perpendicular but disjoint in the present graph representation: CT, LP, and AD.

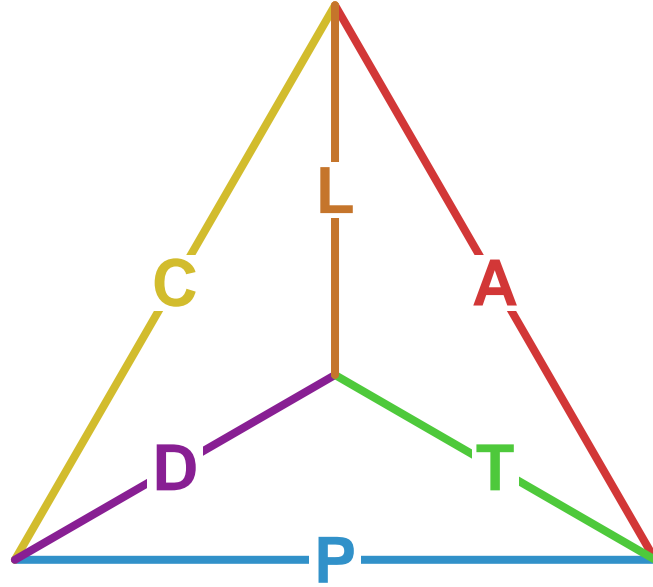


Figure 1: Tetrahedral graph of demographic time hexad identity, with edges labelled by the six time indices. Reproduced from Riffe et al. (2017).

Independent dyads can be identified in fully expanded time graphs such as Fig. 1 as those pairs of edges that do not share a vertex. In a triad identity there are none, but in the demographic time identity there are three, each consisting in a dated event paired with a duration defined as the elapsed time between the other two dated events. In general, in a time identity defined on the basis of  $n$  calendar-dated event types implies a graph with  $n + 1$  vertices, implying a total of  $\binom{n+1}{2}$  graph edges. Each edge in the graph represents a time measure. Since  $n$  of the edges are calendar-event measures,  $\binom{n+1}{2} - n$  are duration measures. The  $n$  calendar-dated measures are represented as graph edges sharing a single vertex. Triad identities are found wherever two edges share a vertex, for a total of three vertices, which means there must be  $\binom{n+1}{3}$  of them within an identity. Independent dyads are pairs of edges that do not share a vertex, meaning that they use four vertices. Since there are  $\binom{n+1}{2}$  ways to select one edge, there are  $\binom{n-1}{2}$  ways to do so after having removed the two vertices of the first edge, or  $\frac{1}{2}\binom{n+1}{2}\binom{n-1}{2}$  unique ways to select two unconnected edges, which simplifies to  $3\binom{n+1}{4}$  independent dyads. For a more complicated example, a multistate model of period, birth, marriage, divorce, and death ( $n = 5$ ) would therefore imply 45 independent dyads, and a total of 20 triad identities. As  $n$  increases, the fraction of possible temporal dyads that are independent

(a) The four view axes of the triad dependencies map to the medians of the tetrahedron. Rotate the tetrahedron so that a given cyan line is reduced to a point: The resulting view is then orthogonal to the outermost identity when flattened to 2d.

(b) The three view axes of the independent planes map to the bimedians of the tetrahedron. Rotate the tetrahedron so that a given magenta line is reduced to a point: The resulting view is then orthogonal to independent plane defined on the basis of the two perpendicular edges crossing in the middle of the figure.

Figure 2: The seven principal view axes of the demographic time identity. Figures are interactive if viewed in a recent version of Adobe Reader. Click each figure to activate. A mouse wheel (or right click and drag) can be used to zoom. Left click and drag to rotate.

approaches 1. To illustrate this point think of the universe of all dates and the potential durations between them: Pick two out of this massive urn and they will almost certainly be independent in this sense.

Any pair of quantitative variables can define the axes of a standard Cartesian plane. When Cartesian axes are defined by two independent time measures and in equal time units, the resultant plane is an isometric projection. Let's call this an independent plane for short. There are two ways that an independent plane may be treated: either (i) accounting for a larger temporal identity or (ii) not. Here we limit comments to the 3d case of the demographic time identity. For the first kind (i) planes are cross-sections of the same 3d demographic timespace previously defined in Fig. 1, and the independent dyads can also be conceived of as *axes of view* on the demographic timespace. This idea is described further in Sec. 2.1. For the second kind (ii), independent planes are the same as any standard Cartesian plane used in data analyses: patterns revealed could be due to real structure or due to sample heterogeneity in other time measures. That is, the larger temporal identity within which these independent measures are nested is *flattened*.

## 2.1 View axes

Imagine now the 3d projection of the graph from Fig. 1, a regular tetrahedron. An angle of view directly at any of the four faces of the demographic time tetrahedron reveals the plane-perspective of a Lexis-like identity. In other words, the viewing axis is set as *normal* to one of the faces of the tetrahedron. Interactive Fig. 2 may help visualize the notion of viewing axes—the view on a plane is orthogonal if an axis line is reduced to appear as a point in the centroid of the tetrahedron. The view axes of the four triad identities map to the four medians (axes of 3-fold symmetry) of the tetrahedron, depicted in Fig. 2a with cyan colored lines. Independent dyad planes are revealed when the viewing angle is such that the midpoints of opposite tetrahedral edges are aligned. There are three pairs of opposite edges (LP, TC, AD), and therefore three viewing angles of this kind, which map to the bimedians of the tetrahedron (axes of 2-fold symmetry), depicted in Figure 2b with magenta lines.

Given a 3d volume of data structured by the demographic time identity, a cross-section orthogonal to a magenta line from Fig. 2b is an independent plane. There is no reason not to do this for outcome measures, for example disease prevalence, in search of meaningful patterns. Further, when *slicing* through such a complex timespace, we are agnostic about cut angles in general, but *snapping* to axes defined on the basis of unit

time measures leads to easier interpretation.

## 3 Data

We illustrate concepts using two data sources: the *le registre de la population du Québec ancien* (RPQA) (Desjardins 1999), and RAND version P of the United States Health and Retirement Study (HRS) (HRS 2017, RAND 2018). The RPQA data are used to illustrate perspectives on population structure, whereas the HRS data are used to illustrate how independent time perspectives can be used to take cross-sections of phenomena with multiple time dimensions.

### 3.1 RPQA

[Provide a minimal description of RPQA data here, as it may be unfamiliar.]

### 3.2 HRS

Do we do this or not? If so, need to make game plan (see note at end). Maybe recycle prevalence from HLETTD, can cite WP?

## 4 Independent time surfaces of population structure

As a first pass to demonstrate concepts, we cross-tabulate Québec’s historical population structure by the three sets of independent demographic time measures. Each cross-tabulation is represented as a demographic surface using filled contour heat maps to create Lexis-like surfaces. These surfaces are distinct from Lexis surfaces in a few important ways. Lexis surfaces are drawn with diagonal lines at a slope equal to the aspect ratio between time units in  $y$  and time units in  $x$ , usually  $45^\circ$  (APC, LCD). Other triad time surfaces might be drawn with descending diagonals (TPD, TAL). Surfaces on independent planes might have features that run horizontally, vertically or in both  $45^\circ$  and  $-45^\circ$  diagonals. Features in population structure may arise in any of these orientations, depending on the case.

As mentioned, the three independent dyads in the demographic time identity each consist in one date measure and one duration measure. Our convention for orienting the Cartesian plane is to use the date measure as abscissa ( $x$ -axis) and the duration measure as ordinate ( $y$ -axis). Since a duration measure is the difference between two dated events, this means that independent dyads in this particular identity only arise when three dated events are known, which identifies the full demographic time identity (PCD implies TAL). Tabulations by two independent time measures can only be done in practice if each of the six demographic time measures are either observed or derivable.

### 4.1 LP: Lifespan and Period

Following this standard orientation, we draw an LP diagram with period in abscissa and lifespan in the ordinate axis. This is for the most part an ageless diagram. An individual with a lifespan of 80 years is represented as a horizontal line spanning from the time of birth until the time of death. That is, for each increment in P, L is identical within an individual. An individual lifespan, which determines  $y$  position, is only known in retrospect. On a single lifeline, one can trace the progression of age (and the other time measures) from left to right. For aggregate data tabulated by L and P, any given

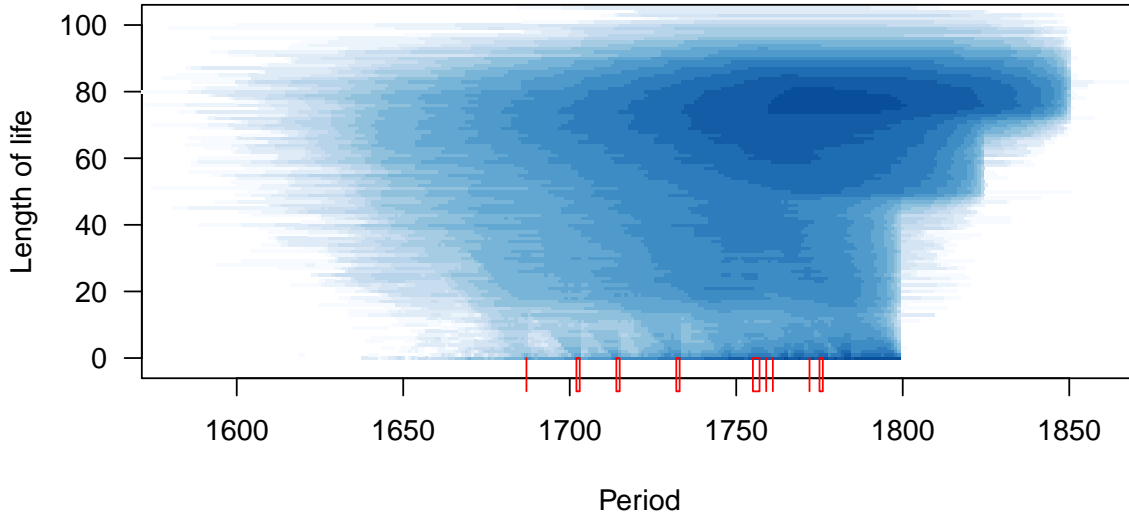


Figure 3: Population structure by length of life and year. Logged counts

combination of  $L$  and  $P$  is a mixture of ages. Heterogeneity of this kind is greater for longer lives than for shorter lives.

[Insert LP diagram about here, with illustrative lifelines]

Fig. 3 displays the Québec population structure on an LP surface. Higher  $y$  values correspond to longer and therefore more overlapped lives. The upper regions of the surface are thus smoother with respect to time than lower regions. Growth over time in to-be-long-lived individuals could occur due to population growth or mortality improvement. We know this pattern is not due to migration because this population was mostly closed in this time. For Québec, increasing contours over time are primarily due to population growth.

Shorter lives are found lower in  $y$ , and these reveal the rhythms of mortality. Mortality shocks therefore add peculiar relief to the surface. High-mortality years are marked with vertical cliffs in population structure: Lives over a broad range of ages that are simultaneously extinguished are right-aligned on the LP surface. Excess deaths in this case are spread over a range of ages, tracing back to births from a range of years, producing the  $-45^\circ$  feature  $\nabla$ . Mortality shocks therefore produce raised triangle relief pattern in population structure. A single isolated year with an extremely high number of births could produce a raised left-edge, but fertility shocks of this kind are far less prominent than mortality shocks in this population. One example might be the 1919 Swedish birth cohort.

Death info here comes from burial records, which had some age constraints in data collection: (1800-1825: deaths of 50+ only; 1826-1850: deaths of 75+ only). That explains the block edges on the right of LP, and the angle edges in TC. I'm also not very sure what accounts for the exceptions to these edges, but my guess would be that it owes to there being multiple potential sources for date of death. Here it is again, with known epidemics (i.e. main disease can be guessed) marked with red rectangles (if spanning more than 1 year) (cf Figure 1.1 Mazan 2011) Typhus, measles, smallpox, etc) were cross checked with other historical documents. These

match the right edge of most of the triangles, and the others may have just been lesser or unstudied epidemics, or ones that I just missed. The left oblique edge of each triangle is in general more visible than the vertical right edge, at least in the early part of the series, because growth in the early phase (pre 1663 if I recall) was largely from migration.

Missingness in LP: Data collection began in 1621, but lifelines in the LP figure begin well left of that. This is because a person dying at some old age right at the start of the series will have of course been born well before the series itself began. Anything to the left of 1621 is incomplete in a sense, because we're missing the lifelines that would overlap the pre 1621 lives we already have (e.g. a life extinguished at age 50 in 1620). So, we could of course crop the surface at 1621, no problem. However, we have the same problem on the right side of the series. Ignoring the block shapes for now. A person dying in 1850 at age 60 would clearly overlap with part of the blue we have, but would here be excluded. The right-side zone of still-incomplete observation in LP is the upper-right triangle of the whole surface; so, draw a line inward from the lower right corner at slope -1. Anything above that is still missing the early part of some lives. Those funny observation blocks offset said missingness to some degree, a happy accident. As a rule, missingness is higher in LP as we go up the y axis on the right side.

Now imagine we have an LP surface of the prevalence  $\pi(l,p)$  of some health condition from HRS. The observation window is much smaller, but we are located high up in the lifespan range. Most of the people whose prevalence should contribute to  $\pi(90,2015)$  have not died yet, because many of them are still toddlers. Others not yet included in  $\pi(90,2015)$  are those age 80 in 2015 that will not die until 2025. None of this will be a problem if we condition these planes on some other bounding time measure, I guess it's just something to be careful with. My first guess as to conditions that might vary in a regular way over LP would be prevalence of having ever smoked. Here's a screenshot of the TAL plane of the 1915 female birth cohort for ever-smoking prevalence:

I used this as an example in another recent paper (link). I found this striking. Lifespan contours are those with slope -1. One explanation would be that smoking determines length of life before age 70 and things simply don't change thereafter (you can't undo being an ever smoker). As I recall dental visits had contours with this slope, but the surface was not monotonic. It's just a guess that strong patterns like these would carry over into LP.

The Québec data has an incredible observation window, but no health except fertility (and extended family events). Lifeline surfaces like those produced for Québec can be inferred from any Lexis surface of death counts.

## 4.2 TC

The thanatological age distribution of a birth cohort specifies that birth cohort's life table (distribution of T given C). Conversely, the distribution of birth cohorts of people of a given thanatological age reveals mortality history and inequalities in survival. If more of the people who will die in 5 years are from recent cohorts than from cohorts long ago, this population has high childhood mortality and low survival to old ages.

A lifeline on the CT plane extends from 0 upward until the age of death. Ergo, someone born in 1700 dying at age 80 contributes 1 to the count in each thanatological age from 0 to 80, the entire countdown.  $T = 0$  is maximally heterogenous with respect to A (or L or D or P), and lifeline overlap decreases in higher T (long-lived people while they're young, i.e. in years closer to C itself). Here's that tabulation of this same data (also logged):

The first thing to pop out at me are birth dearths for particular years, possibly the same years that were the right edges of mortality shocks in the LP surface. Also a wider boom or two, which might just be natural growth. Here each vertical strip monotonically



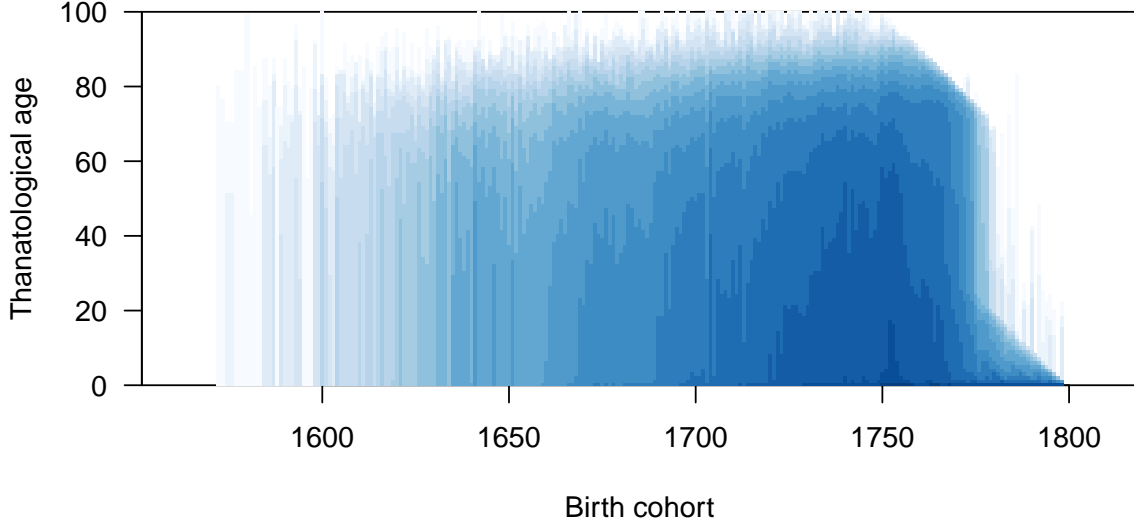


Figure 4: Population structure by time to death and birth cohort.

decreases as we go up to higher thanatological ages. I agree these are the stationary populations belonging to each birth cohort, each with a radix equal to the actual birth cohort size, so this is a bit more straightforward to wrap one's mind around than LP.

It seems that the upper envelope (that is, the thanatological age at which the lifelines terminate, as a function of cohort) is rising with increasing cohort, reflecting improvements in cohort life expectancy. These improvements are visible also in contour lines that are lower than the upper envelope. You suggested that growth happens in the contours mainly because the population was growing exponentially. If you wanted to remove the effect of population growth, you could plot a curve of births or a curve of population size or a curve of number of women in the population below the x axis with an inverted y axis and then normalize the counts by births or population size or number of women. That might reveal additional patterns. This kind of normalization by some measure of population size might also be applied to the next figure.

I now detect the affects of the right-side block observation of deaths. Had burial records for younger ages been captured in this data, we would have increasing contours as we go down towards the x axis on the right side. Here we just get the tips of the oldest deaths, each having needed to pass through young ages first, so they trace down to age 0.

### 4.3 AD

For a given age  $A$ , the distribution of the death cohort  $D$  of individuals gives the conditional life table, conditional on reaching the given age  $A$ , and hence gives, for example, the remaining life expectancy and the variability of remaining life, among other summaries of the conditional life table. Conversely, for a given death cohort  $D$ , the distribution of age  $A$  of individuals specifies the distribution of their birth cohorts and life lengths.

In the AD plane, we have the same sort of vertical overlap. Just as everyone experi-

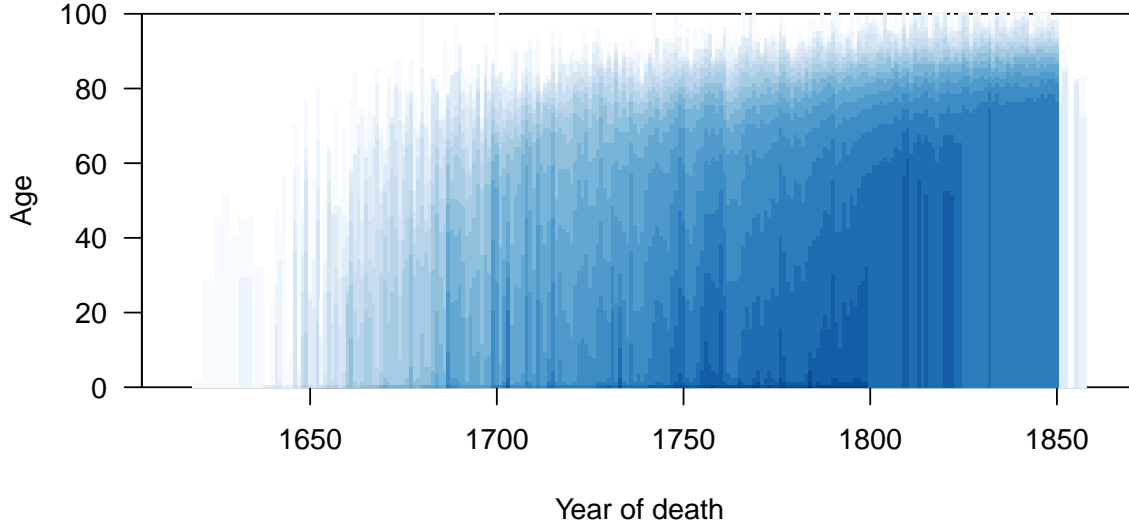


Figure 5: Population structure by age and year of death.

ences  $T = 0$ , everyone is also included in  $A = 0$  here. If this were a stationary population then a vertical strip here would be equal to one from the CT plane, but we have lots of non-stationary things happening. Vertical striation here is due to mortality shocks in particular years, and these are strictly positive. Anything dearthy-looking is either due to a 'harvest' or data collection oddities. At least I've never heard of a survival shock. So, to repeat, an individual born in 1700 destined to die at age 80 is located at 1780 in the x axes, and ages from the bottom up. Ergo, the bottom of the surface is young people, as we're used to, but they come from different birth cohorts. As with CT, the population is maximally mixed with respect to  $C$  (or  $L$  or  $P$  or  $T$ ) at  $A = 0$ . An 80 year old and an infant dying in 1780 both contribute to the count at age 0 in 1780.

We see here the same high density in death year 1705 that we detected in the LP plane, so maybe it is not a data artifact. Whether there are also high densities of deaths in the other years we identified in the LP plane is not so clear to me; perhaps yes. There appear to be little white rectangles between 1760 and 1800 at ages zero through perhaps 3 or 4. Every lifeline must extend to age zero, so something is wrong here, I think. Again we see the upper envelope increasing with the year of death and rising contour lines.

## 5 Independent time surfaces as cross-sections

Here use example of HRS prevalence surfaces produced at novel *cut* angles. I'm thinking maybe small multiples, since each would be a fly-through of cross sections in fixed intervals. There would be 3 independent fly-throughs, and 4 Lexis-like fly throughs. These would be like Fig. (7) of Riffe et al. (2017), which is a fly-through at the TAL viewing angle. So, maybe each fly-through is a column and each perspective is a row? Maybe 10-year jumps in columns to imply 3-4 columns? Total? For example, each TAL slice in that Fig 7 is held for a fixed  $C$ . It would have been possible to hold  $P$  or  $D$  fixed instead,

but I guess the actually slices would be identical, just the labels would change. Just need to be clear about it when labelling said panel figure.

## 5.1 LP

## 5.2 TC

## 5.3 AD

# References

- Brouard, N. (2019). Theory and applications of backward probabilities and prevalences in cross-longitudinal surveys. In A. Srinivasa Rao and R. C. R. (Eds.), *Handbook of Statistics*, Volume 40, Chapter Backward Probabilities Estimated From Cross-Longitudinal Surveys and How Age-Specific Backward Prevalences Could Help Analyzing Trends in Aging Societies, pp. 435–486. Elsevier.
- Desjardins, B. (1999). Le registre de la population du qubec ancien. *Annales de Dmographie Historique* 1998(2), 215–226.
- HRS (2017). Health and retirement study public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740).
- Mazan, R. (2011). *Analyzing Epidemics in New France: The Measles Epidemic of 1714-15*. Ph. D. thesis, University of Western Ontario.
- RAND (2018). RAND HRS Data, Version P. Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration.
- Riffe, T., J. Schöley, and F. Villavicencio (2017). A unified framework of demographic time. *Genus* 73(7).