PASH

Pace & Shape Analysis in R

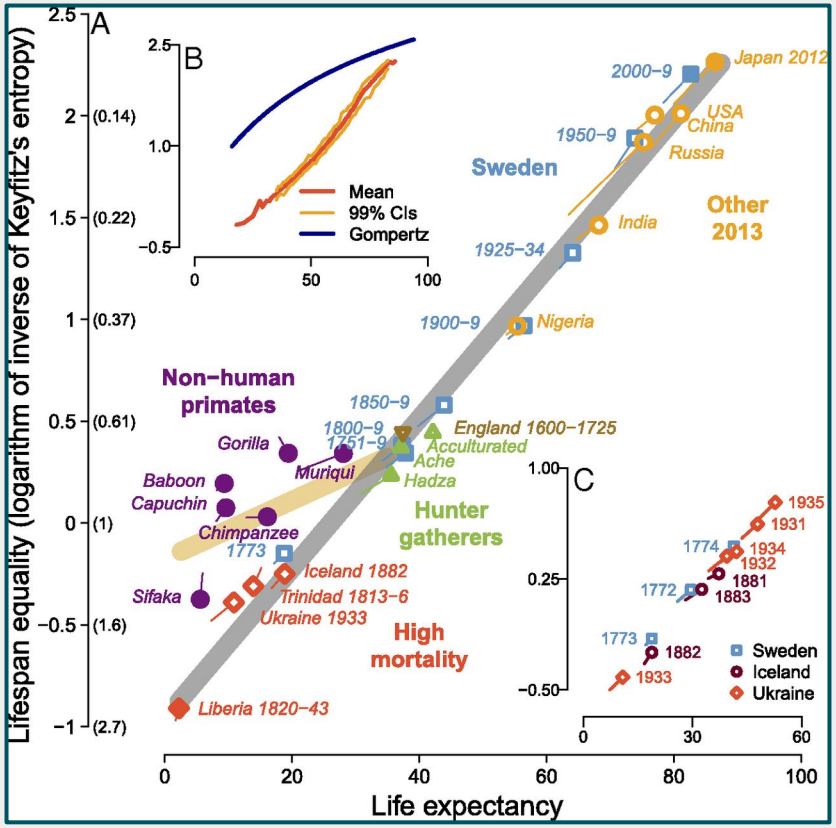
Jonas Schöley etal.

jschoeley@health.sdu.dk

Max-Planck Odense Center on the
Biodemography of Aging

Department of Public Health
University of Southern Denmark

The logistics of comparative mortality research



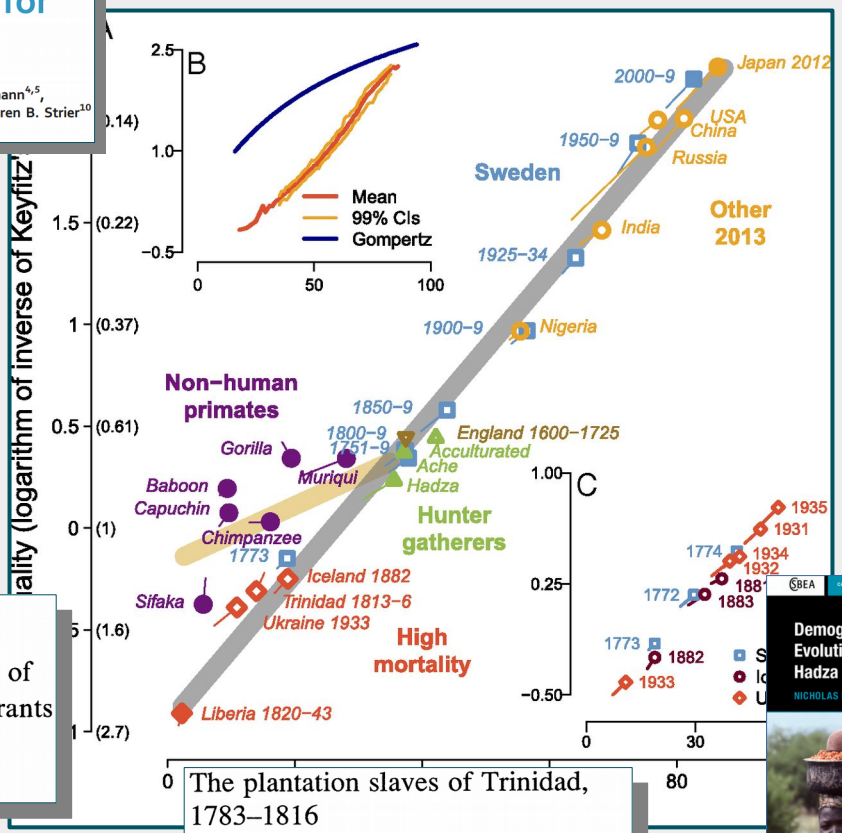
The logistics of comparative mortality research

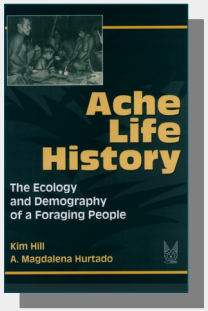
The Human Mortality Database

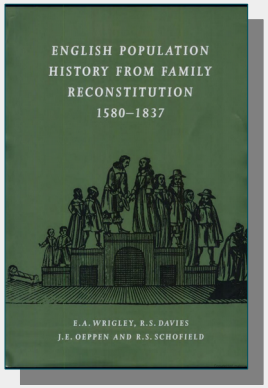
World Health Organization

Female and male life tables for seven wild primate species

Anne M. Bronikowski[1,*], Marina Cords[2], Susan C. Alberts[3,4], Jeanne Altmann[4,5], Diane K. Brockman[6], Linda M. Fedigan[7], Anne Pusey[8], Tara Stoinski[9], Karen B. Strier[10] & William F. Morris[3,*]







Demography, Vol. 29, No. 4, November 1992

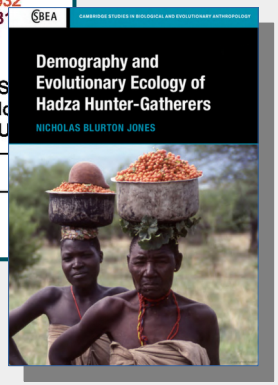
Extreme Mortality in Nineteenth-Century Africa: the Case of Liberian Immigrants*

Antonio Mc Daniel
Population Studies Center
University of Pennsylvania
Philadelphia, PA 19104

Population Studies, **48** (1994), 99–115
Printed in Great Britain

Patterns of Mortality by Age and Cause of Death Among Nineteenth-Century Immigrants to Liberia*

ANTONIO MCDANIEL AND SAMUEL H. PRESTON†



The plantation slaves of Trinidad, 1783–1816

A mathematical and demographic enquiry

A. MEREDITH JOHN
Stanford University

A software solution

Magic

Gini, Entropy, CV, IQR, Mortality ratio, Life-expectancy ratio

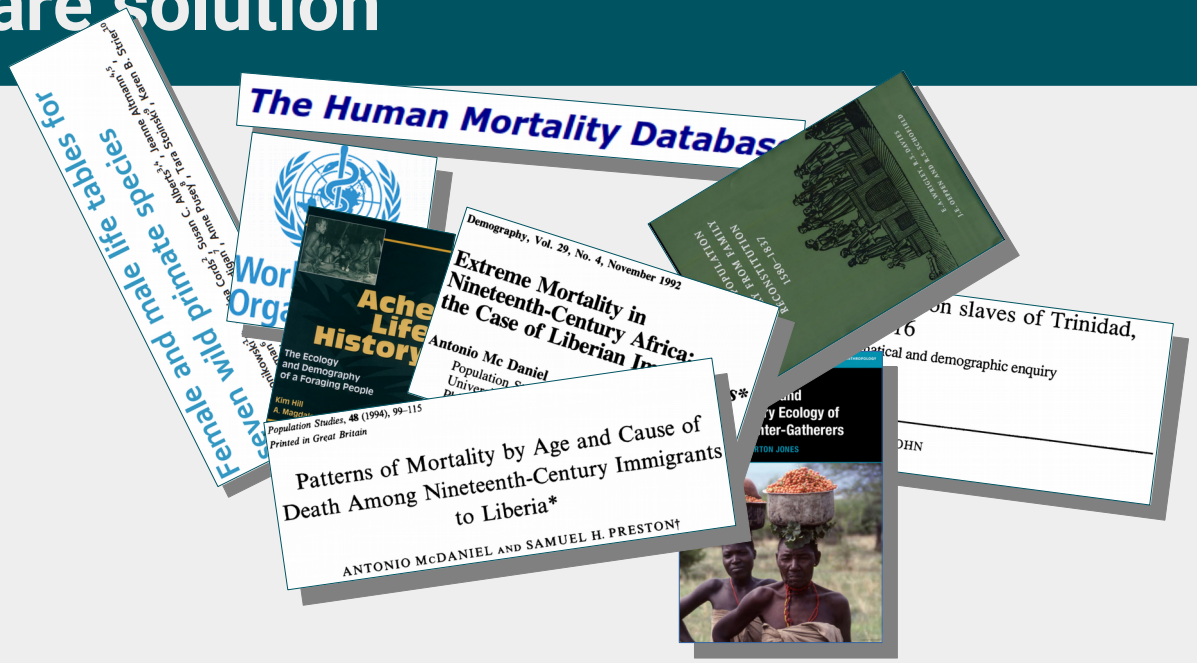
Life-expectancy, Survival quantiles, LT death rate

Pace standardized life-tables

Harmonized life-tables

...

A software solution



~~Magic~~ PASH

Gini, Entropy, CV, IQR, Mortality ratio, Life-expectancy ratio

Life-expectancy, Survival quantiles, LT death rate

Pace standardized life-tables

Harmonized life-tables

...

PASH: Pace & shape analysis in R

Friendly to casual & expert users

Complete **implementation** of Annette Baudischs **Pace & Shape** methodology

Support for many **different data formats**

Works with data **across the tree of life**

A pace & shape implementation

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution doi: 10.1111/j.2041-210X.2010.00087.x

The pace and shape of ageing

Annette Baudisch*

Max Planck Research Group on Modelling the Evolution of Ageing, Max Planck Institute for Demographic Research, D-18057 Rostock, Germany

Pace measures

Life-expectancy, life-table death rate, survival quantiles

Shape measures

Keyfitz's entropy, Gini coefficient, CV, IQR, Mortality ratio, Life-expectancy ratio, Probability to survive to mean age at death, Average change of mortality over age

Life-table standardization

Support for many different data situations

Life-tables

- ${}_nm_x$ ${}_nq_x$ ${}_nl_x$ ${}_nd_x$ ${}_nL_x$
- open/closed
- single-year, abridged,
- irregular abridged
- constant hazard, uniform distribution of deaths
- user specified ${}_na_x$
- "continuous" cohort life-tables

Survival times

- censoring, truncation, late-entry
- choice among Kaplan-M. or Nelson-Aalen estimator for LT construction
- leverage existing functionality of "`survival`" package

Population matrices

- continuous and discrete time Markov chain models

PASH

Gini, Entropy, CV, IQR, Mortality ratio, Life-expectancy ratio

Life-expectancy, Survival quantiles, LT death rate

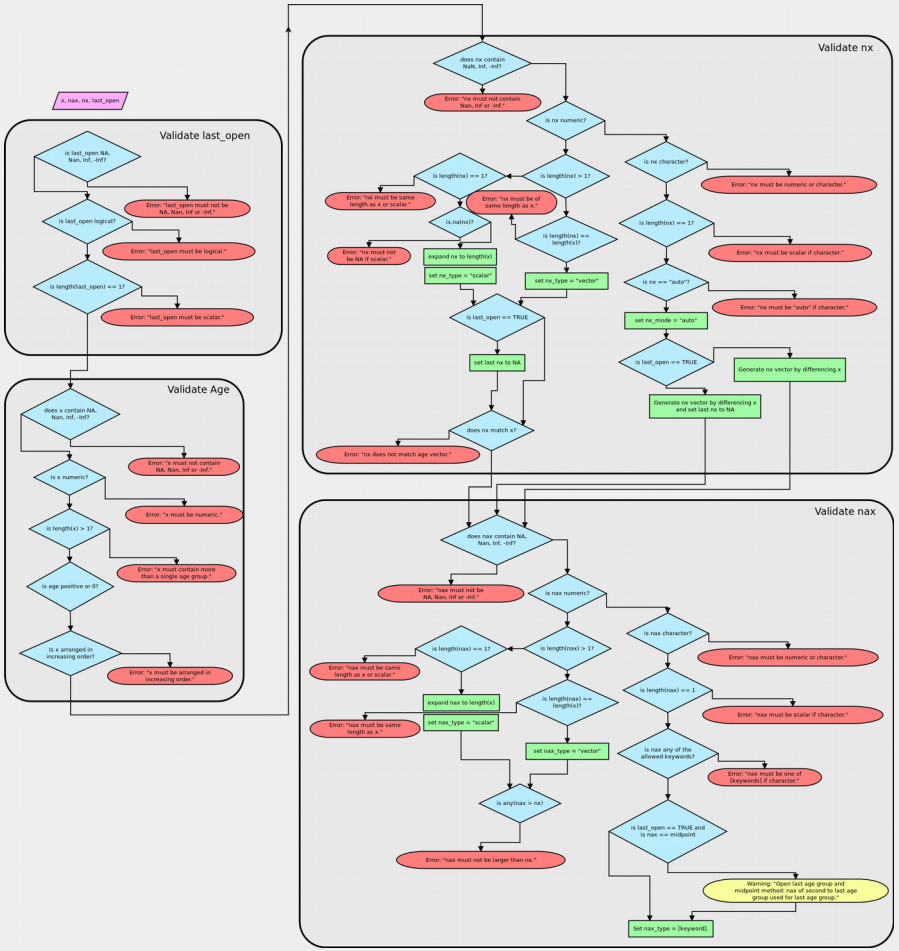
Pace standardized life-tables

Harmonized life-tables

...

Extensive error detection

	x	lx
1	0	100000
2	1	99133
3	5	99986
4	10	98910
5	15	98815
6	20	98334
7	25	97704
8	30	97151
9	35	96492
10	40	95588
11	45	94195
12	50	91937
13	55	88711
14	60	83845
15	65	76377
16	70	66225
17	75	53803
18	80	37441
19	85	21134



Error: lx function not monotonically decreasing

Extensive documentation and method reports

Notation

X	Continuous random variable "age at death"
x	Lower bound of an age group
θ	Starting age of life-table (usually $\theta = 0$)
ω	Lower bound of last life-table age group
	(discrete setting) or highest observed value
	of X (continuous setting)
n_x	Width of age group x as $[x, x + n_x)$
$f(x) = P(X = x)$	Probability density function of ages at death
$\ell(x) = \ell_x = P(X > x)$	Probability of surviving up until x
${}_nd_x = \int_x^{x+n} f(x)\,\mathrm{d}x$	Expected number of deaths in the life-table
	population over age interval $[x, x + n)$
$\mu(x)$	Force of mortality at age x (hazard of death)
${}_nm_x$	Expected mortality rate over age interval $[x, x + n_x)$
$e(x) = e_x = \int_x^\omega \ell(x)\,\mathrm{d}x$	Remaining life-expectancy at age x
${}_na_x$	Avg. time spent in $[x, x + n)$ when dying there
${}_nA_x = \frac{{}_na_x}{n_x}$	Avg. fraction of interval alive when dying there
Let $k \in \mathbb{R}$, then	
$\lfloor k \rfloor$	Integer part of k.
$\lceil k \rceil$	$\lfloor k \rfloor + 1$
$\mathrm{frac}(k)$	$k - \lfloor k \rfloor$

Average change in force of mortality wrt. ℓ_x, `acfm`, ACFM

A rescaled variant of the average lifetime change of the hazard of death divided by average mortality. Suggested by [7].

Continuous

$$\mathrm{ACFM} = \int_{x=\theta}^{\omega} \frac{\mu(x) - \mu(\theta)}{\bar{\mu}} f(x)\,\mathrm{d}x$$

Discrete

$$\widehat{\mathrm{ACFM}} = e_\theta \sum_{x=\theta}^{\omega} \left[({}_nm_x - {}_nm_\theta)\, {}_nd_x \right]$$

Harmonizing Transformation

$$\mathrm{ACFM}^* = 1 - e^{-\mathrm{ACFM}}$$

***pash* implementation**

```
# Average of Change in Force of Mortality with respect to lx
ACFM <- function(nmx, ndx, ex, harmonized){
  acfm_x = (nmx - nmx[1L]) * ndx
  D = ex[1L] * sum(acfm_x)
  if (!isTRUE(harmonized)) {S = D}
  if (isTRUE(harmonized)) {S = 1-exp(-D)}
  return(S)
}
```

Probability to survive up to mean age at death, `psmad`, ℓ_{e_θ}

The probability to survive up to the mean age at death, suggested by [3].

Continuous

$$\ell_{e_\theta} = \mathrm{P}(X > e_\theta)$$

Discrete

Using linear interpolation of the $\ell(x)$ function we get

$$\widehat{\ell}_{e_\theta} = \frac{1}{n_{\lfloor e_\theta \rfloor}} \left[\ell_{\lfloor e_\theta \rfloor} (n_{\lfloor e_\theta \rfloor} - \mathrm{frac}\, e_\theta) + \ell_{\lfloor e_\theta \rfloor} \mathrm{frac}\, e_\theta \right].$$

Harmonizing Transformation

$$\ell^*_{e_\theta} = 1 + \log \ell_{e_\theta}$$

***pash* implementation**

```
# Probability to Survive up to the Mean Age at Death
PSMAD <- function(x, nx, lx, ex, harmonized){
  l_e0  = approx(x = x, y = lx, xout = ex[1L])[["y"]]
  if (!isTRUE(harmonized)) {S = l_e0}
  if (isTRUE(harmonized)) {S = 1 - log(l_e0)}
  return(S)
}
```

Life expectancy ratio, `ler`, LER

The ratio between remaining life expectancy at the age of total life expectancy and total life expectancy. Suggested by [7].

Continuous

$$\mathrm{LER} = \frac{e(e_\theta)}{e(\theta)}$$

Discrete

$$\widehat{\mathrm{LER}} = 1 - \frac{\hat{e}_{e_\theta}}{e_\theta}$$

where, using linear interpolation of the $e(x)$ function, we have

$$\hat{e}_{e_\theta} = \frac{1}{n_{\lfloor e_\theta \rfloor}} \left[e_{\lfloor e_\theta \rfloor} (n_{\lfloor e_\theta \rfloor} - \mathrm{frac}\, e_\theta) + e_{\lfloor e_\theta \rfloor} \mathrm{frac}\, e_\theta \right].$$

Harmonizing Transformation

$$\mathrm{LER}^* = 1 - \mathrm{LER}$$

***pash* implementation**

```
# Life Expectancy Ratio
LER <- function(x, nx, ex, harmonized){
  e_e0 = approx(x = x, y = ex, xout = ex[1L])[["y"]]
  ler = e_e0/ex[1L]
  if (!isTRUE(harmonized)) {S = ler}
  if (isTRUE(harmonized)) {S = 1-ler}
  return(S)
}
```

Mortality ratio, `mr`, MR

The ratio between mortality at age 0 and mortality at the age of total life-expectancy. First suggested by [1], further developed by [7].

Continuous

$$\mathrm{MR} = \frac{\mu(\theta)}{\mu(e_\theta)}$$

Discrete

$$\widehat{\mathrm{MR}} = \frac{{}_nm_\theta}{\hat{\mu}_{e_\theta}}$$

where, using linear interpolation of the ${}_nm_x$ data, we have

$$\hat{\mu}_{e_\theta} = \frac{1}{n_{\lfloor e_\theta \rfloor}} \left[{}_nm_{\lfloor e_\theta \rfloor} (n_{\lfloor e_\theta \rfloor} - \mathrm{frac}\, e_\theta) + {}_nm_{\lfloor e_\theta \rfloor} \mathrm{frac}\, e_\theta \right]$$

Harmonizing Transformation

$$\mathrm{MR}^* = 1 - \mathrm{MR}$$

***pash* implementation**

```
# Mortality Ratio
MortalityRatio <- function(x, nx, nmx, ex, harmonized){
  m0   = nmx[1L]
  m_e0 = approx(x = x, y = nmx, xout = ex[1L])[["y"]]
  MR   = m0/m_e0
  if (!isTRUE(harmonized)) {S = MR}
  if (isTRUE(harmonized)) {S = 1 - MR}
  return(S)
}
```

Life-table Gini coefficient, `gini`, G

The Gini coefficient of the age-at-death distribution represents the mean absolute difference in the ages at death between any two individuals of a population.

We base our discretization of the Gini coefficient on the formulation by [3] because it allows for a straightforward inclusion of ${}_na_x$ and doesn't require special considerations in case of an open-ended age group. An alternative formulation in terms of the life-table survivorship curve proposed by [2] and discretized by [8, 9] is in widespread use in demography.

Continuous

$$\mathrm{G} = \frac{1}{2e_\theta} \int_{x=\theta}^{\omega} \int_{y=\theta}^{\omega} |x - y| \cdot f(x) f(y)\,\mathrm{d}x\,\mathrm{d}y$$

Discrete

Let i, j be indices of life-table age groups $1 \ldots k$. Let $\overline{x}_i = x_i + a_i$ be the average age at death for those who die in age group i. Then

$$\widehat{\mathrm{G}} = \frac{1}{2e_\theta} \sum_{i,j=1}^{k} d_i \cdot d_j \cdot |\overline{x}_i - \overline{x}_j|$$

This expression can easily evaluated using the matrix facilities available in R. We therefore reformulate it in matrix algebra notation. Let $\mathbf{D}_{k \times k} = \mathbf{d} \otimes \mathbf{d}$ and let $\overline{\mathbf{X}} = |\mathbf{e}\overline{\mathbf{x}}^\top - \overline{\mathbf{x}}\mathbf{e}^\top|$, then

$$\begin{aligned} \widehat{\mathrm{G}} &= \frac{1}{2e_\theta} \sum_{i,j=1}^{k} \mathbf{D} * \overline{\mathbf{X}} \\ &= \frac{\mathbf{e}^\top \left[\mathbf{D} * \overline{\mathbf{X}} \right] \mathbf{e}}{2e_\theta} \end{aligned}$$

Harmonizing Transformation

$$\mathrm{G}^* = 1 - 2\mathrm{G}$$

***pash* implementation**

```
# Life Table Gini-Coefficient
LifetableGini <- function (x, nax, ndx, ex, harmonized) {
  e = rep(1, length(x))
  D = outer(ndx, ndx)
  x_ = x+nax
  X_ = abs(e%*%t(x_) - x_%*%t(e))
  G = sum(D*X_)/(2*ex[1L])
  if (!isTRUE(harmonized)) {S = G}
  if (isTRUE(harmonized)) {S = 1-2*G}
  return(S)
}
```

Life-table entropy, `entropy`, H

Life-table entropy, also known as Keyfitz's entropy [4], is the average remaining life-expectancy lost upon death ($e^\dagger$ [5]) divided by the total life-expectancy.

Continuous

$$\mathrm{H} = \frac{e^\dagger}{e(\theta)} = \frac{\int_{x=\theta}^{\omega} e(x) f(x)\,\mathrm{d}x}{e(\theta)}$$

Discrete

$$\widehat{\mathrm{H}} = \frac{\sum_{x=\theta}^{\omega} {}_nd_x \left[{}_nA_x e_{x+n_x} + (1 - {}_nA_x) e_x \right]}{e_\theta},$$

with

$${}_nd_\omega \left[{}_nA_\omega e_{\omega+n_\omega} + (1 - {}_nA_\omega) e_\omega \right] = e_\omega$$

for the final age group.

Harmonizing Transformation

$$\mathrm{H}^* = 1 - \mathrm{H}$$

***pash* implementation**

```
# Average Life-Expectancy in Age x
EDaggerx <- function(nax, nx, ex) {
  nAx = nax/nx
  edx = (nAx*C(ex[-1L], 0) + (1-nAx)*ex)
  edx[length(edx)] = ex[length(ex)]
  return(edx)
}

# Total Life Years Lost due to Death
EDagger <- function(nax, nx, ndx, ex) {
  edx = EDaggerx(nax, nx, ex)
  ed  = sum(ndx*edx)
  return(ed)
}

# Life Table Entropy
LifetableEntropy <- function(nax, nx, ndx, ex, harmonized) {
  ed = EDagger(nax, nx, ndx, ex)
  H  = ed/ex[1L]
  if (!isTRUE(harmonized)) {S = H}
  if (isTRUE(harmonized)) {S = 1-H}
  return(S)
}
```

Life-table coefficient of variation, `cv`, CV

The coefficient of variation of the age at death is the standard deviation in the age at death divided by the total life-expectancy.

Continuous

$$\mathrm{CV} = \frac{\sqrt{\mathrm{Var}(X)}}{\mathrm{E}(X)} = \frac{\sqrt{\int_{x=\theta}^{\omega} f(x)(x - e_\theta)^2\,\mathrm{d}x}}{e_\theta}.$$

Discrete

$$\widehat{\mathrm{CV}} = \frac{\sqrt{\sum_{x=\theta}^{\omega} {}_nd_x (\overline{x} - e_\theta)^2}}{e_\theta} = \frac{\sqrt{\sum_{x=\theta}^{\omega} {}_nd_x (x + {}_na_x - e_\theta)^2}}{e_\theta}.$$

Harmonizing Transformation

$$\mathrm{CV}^* = 1 - \mathrm{CV}$$

***pash* implementation**

```
# Life Table Variance
LifetableVar <- function(x, ndx, nax, ex) {
  var = sum(ndx*(x+nax-ex[1L])^2)
  return(var)
}

# Life Table Coefficient of Variation
LifetableCV <- function(x, ndx, nax, ex, harmonized) {
  var = LifetableVar(x, ndx, nax, ex)
  CV  = sqrt(var)/ex[1L]
  if (!isTRUE(harmonized)) {S = CV}
  if (isTRUE(harmonized)) {S = 1-CV}
  return(S)
}
```

Demography across the tree of life

Never assume a human origin!

PASH: Data input

```
> prestons_lx
    x     lx
1   0 100000
2   1  99133
3   5  98986
4  10  98910
5  15  98815
6  20  98334
7  25  97704
8  30  97151
9  35  96492
10 40  95588
11 45  94195
12 50  91937
13 55  88711
14 60  83845
15 65  76377
16 70  66225
17 75  53803
18 80  37441
19 85  21134
```

Preston, Samuel H., Patric Heuveline, and Michel Guillot. 2001. Demography. Oxford, UK: Blackwell, p. 51 and United Nations 1994.

PASH: Data input

```
> lt = Inputlx(x = prestons_lx$x, lx = preston_lt$lx,
               last_open = TRUE)

Inputlx() and last_open = TRUE: nmx of open age group
log-linearly extrapolated based on preceding two nmx.

> summary(lt)

A life table with 19 age groups.
Source: lx

Average life expectancy : 72.96 years
Avg. e0 lost upon death : 12.79 years
Life table entropy      : 0.175
Gini Coefficient        : 0.117
```

PASH: Pace & shape measures

```
> GetPace(lt)

        e0          qlx          ldr
72.95989256 76.16214399  0.01370616

> GetShape(lt, harmonized = FALSE)

  entropy      gini        cv        mr        ler
0.1752712 0.1168986 0.2265215 0.1467121 0.1422102
     acfm     psmad
4.9759710 0.5887144

> GetShape(lt, harmonized = TRUE)

  entropy      gini        cv        mr        ler
0.8247288 0.7662027 0.7734785 0.8532879 0.8577898...
```

PASH: Pace standardization

```
> StandardizeLT(lt)
```

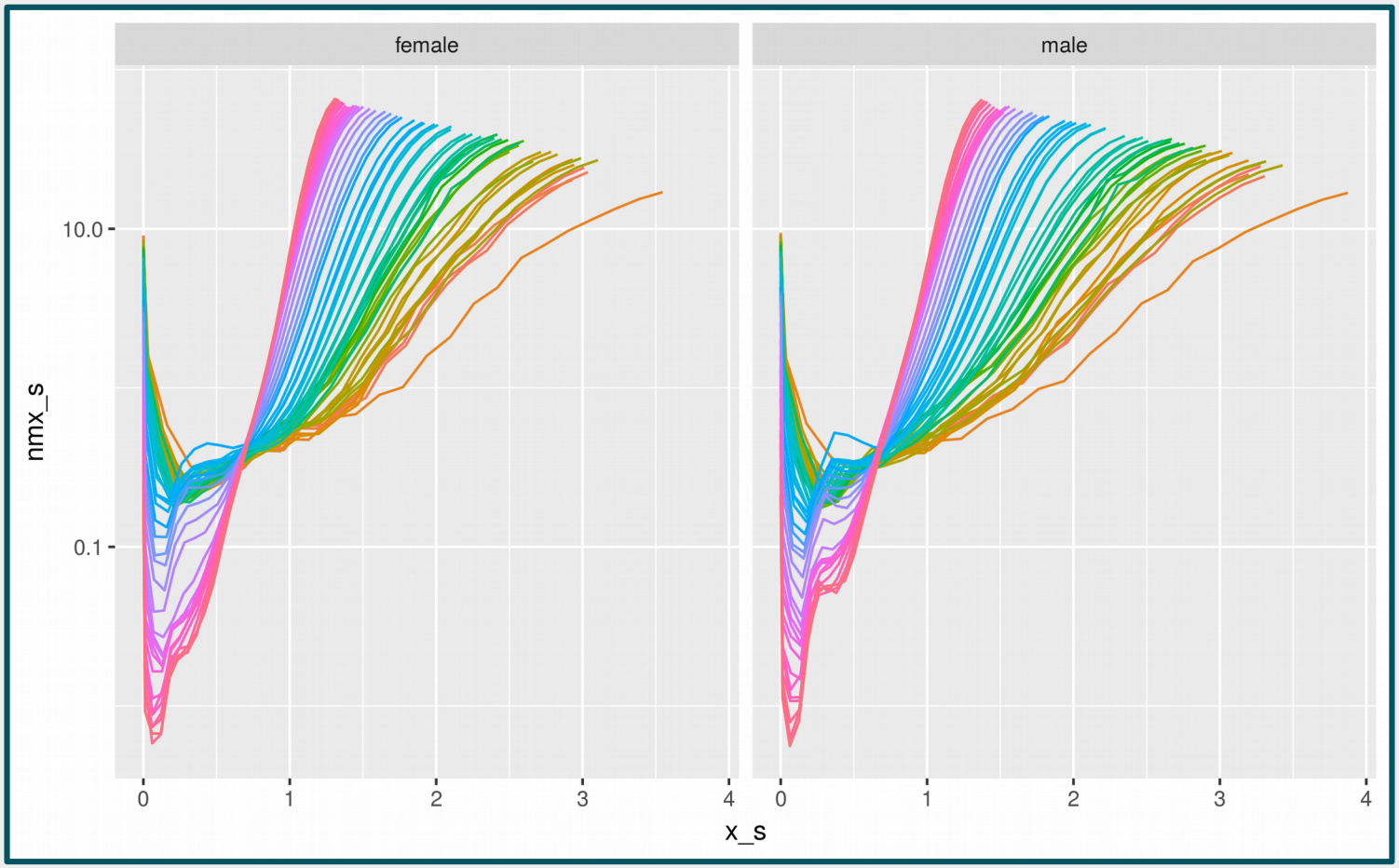
	x	nx	x_s	nx_s	nmx_s	lx_s	ex_s
1	0	1	0.00000000	0.01370616	0.63531636	1.00000	1.00000000
2	1	4	0.01370616	0.05482464	0.02706733	0.99133	0.99497973
3	5	5	0.06853080	0.06853080	0.01120781	0.98986	0.94159198
4	10	5	0.13706161	0.06853080	0.01402188	0.98910	0.87375835
5	15	5	0.20559241	0.06853080	0.07120241	0.98815	0.80603463
6	20	5	0.27412321	0.06853080	0.09378739	0.98334	0.74127893
7	25	5	0.34265401	0.06853080	0.08282430	0.97704	0.67730698
8	30	5	0.41118482	0.06853080	0.09931796	0.97151	0.61243648
9	35	5	0.47971562	0.06853080	0.13735057	0.96492	0.54785434
10	40	5	0.54824642	0.06853080	0.21420913	0.95588	0.48418068
11	45	5	0.61677722	0.06853080	0.35403571	0.94195	0.42230344
12	50	5	0.68530803	0.06853080	0.52116517	0.91937	0.36330297
13	55	5	0.75383883	0.06853080	0.82297419	0.88711	0.30673770
14	60	5	0.82236963	0.06853080	1.36027381	0.83845	0.25402001
15	65	5	0.89090043	0.06853080	2.07763939	0.76377	0.20697640
16	70	5	0.95943124	0.06853080	3.02032121	0.66225	0.16492143
17	75	5	1.02796204	0.06853080	5.23330745	0.53803	0.12655641
18	80	5	1.09649284	0.06853080	8.12467413	0.37441	0.09835747
19	85	NA	1.16502365	NA	12.61350119	0.21134	0.07928013

Case study: Pace & shape analysis of HMD data

```
> sweden5x5
# A tibble: 2,496 × 5
      sex    period     x      nmx   nax
   <fctr>    <fctr> <int>    <dbl> <dbl>
1  female 1755-1759     0 0.23517  0.35
2  female 1755-1759     1 0.04076  1.52
3  female 1755-1759     5 0.01256  1.88
4  female 1755-1759    10 0.00658  2.53
5  female 1755-1759    15 0.00622  2.44
6  female 1755-1759    20 0.00720  2.62
7  female 1755-1759    25 0.00929  2.57
8  female 1755-1759    30 0.01230  2.53
9  female 1755-1759    35 0.01108  2.49
10 female 1755-1759    40 0.01622  2.59
# ... with 2,486 more rows
```

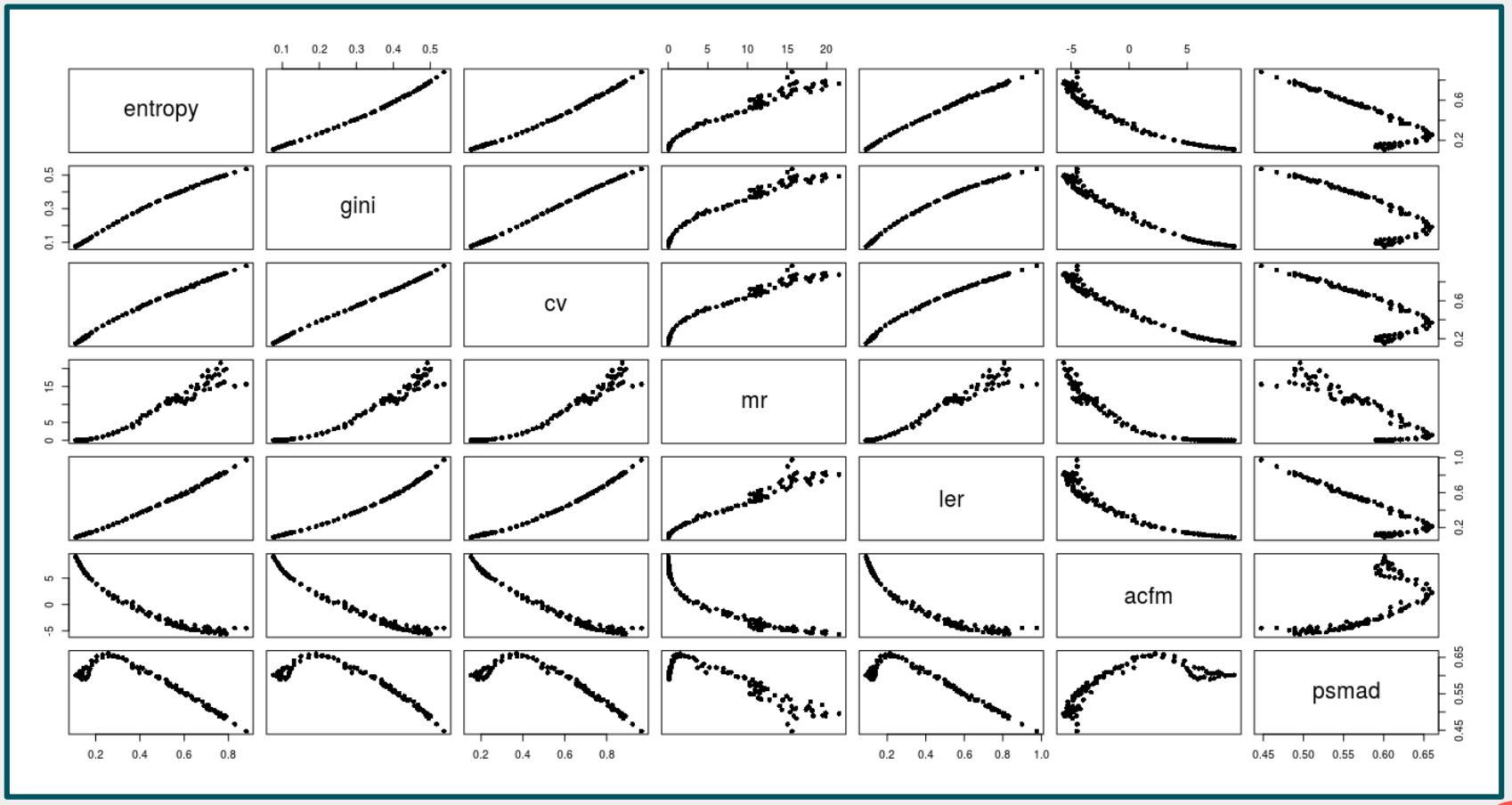
Pace-standardized age-specific mortality

```
sweden5x5 %>% group_by(period, sex) %>%
  do(Inputnmx(x=.$x, nmx=.$nmx) %>% StandardizeLT()) %>%
  qplot(x=x_s, y=nmx_s, color=period, geom="line", facets=~sex, data=., log="y") +
  guides(color = FALSE)
```



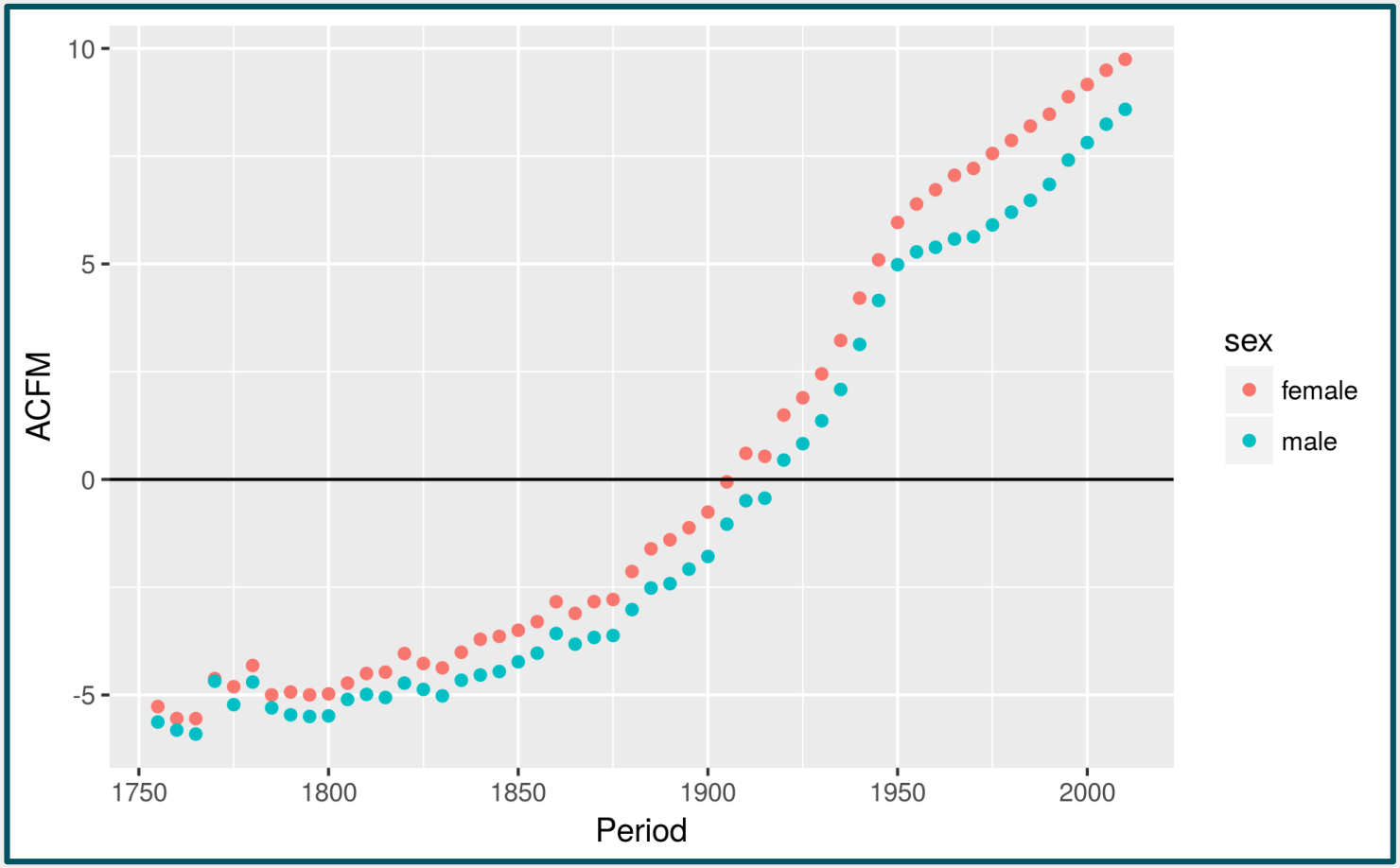
Correlation among shape measures

```
sweden5x5 %>% group_by(period, sex) %>%
  do(Inputnmx(x = .$x, nmx = .$nmx, last_open = TRUE) %>%
       GetShape(harmonized = FALSE) %>% t() %>% as.data.frame()) %>%
  pairs(x = .[3:9], data = ., pch = 16)
```

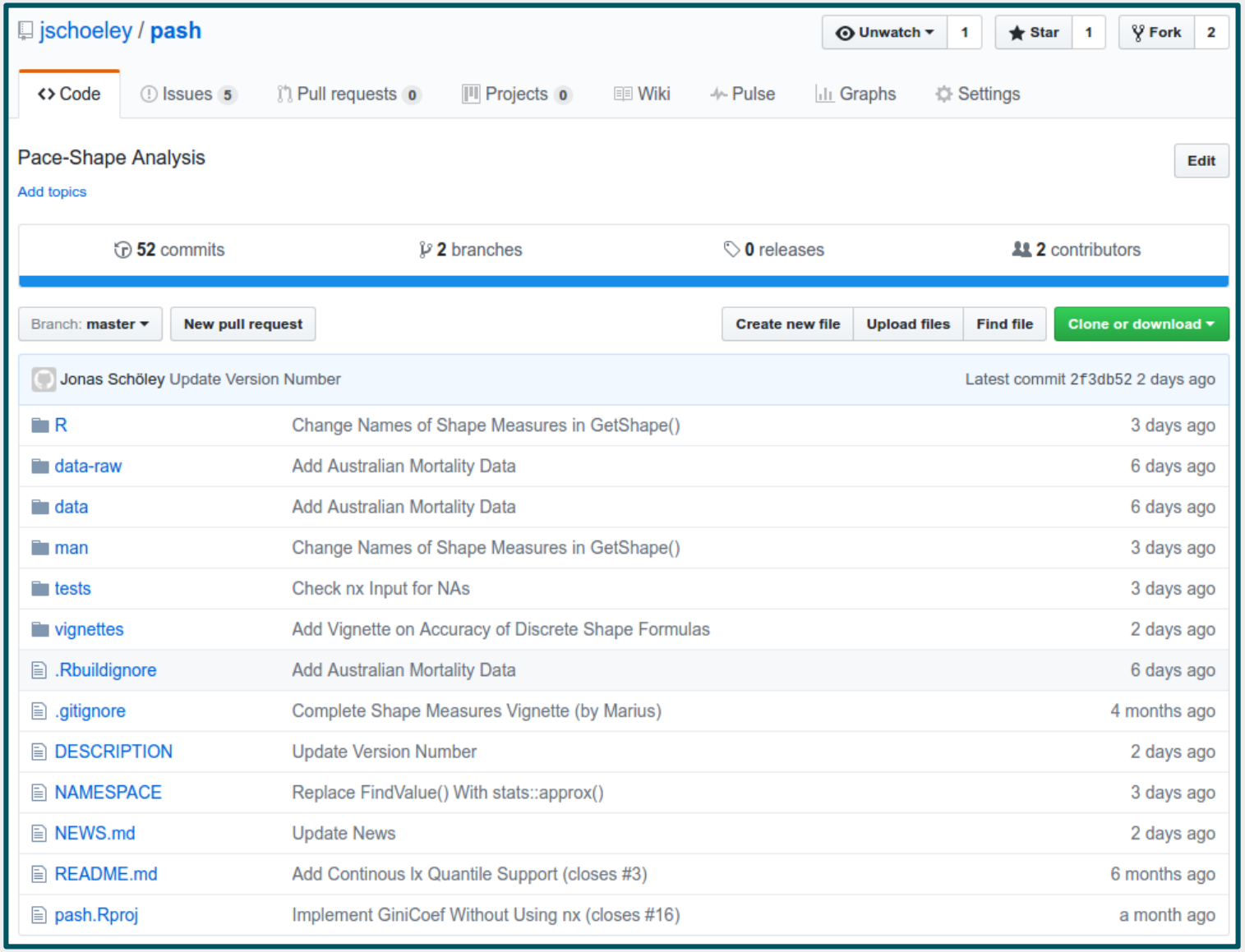


The emergence of senescence

```
sweden5x5 %>% group_by(period, sex) %>%
  do(Inputnmx(x=.$x, nmx=.$nmx, nax = .$nax, last_open=TRUE) %>%
       GetShape(type="acfm", harmonized=FALSE) %>% as.data.frame()) %>%
  qplot(x=as.numeric(substr(period, 1,4)), y=., color=sex, data=.,
        ylab="ACFM", xlab="Period") + geom_hline(yintercept=0)
```



Join us on **github.com/jschoeley/pash**



Jonas Schöley

Marius Pascariu

Francisco Villavicencio

Maciej Danko

Iain Stott

Annette Baudisch

Reproducible analysis

github.com/jschoeley/pash

Jonas Schöley

jschoeley@health.sdu.dk

Twitter: @rettungstweet