

Layers of Reproducibility

Jonas Schöley & Egor Kotov



@jschoeley



0000-0002-3340-8518



schoeley@demogr.mpg.de



MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH

Benefits of reproducibility and open science

Some personal experiences

LogrankA

Together with



Roland Rau
@Demographie

CRAN Task View: Survival Analysis

Maintainer: Arthur Allignol, Aurelien Latouche
Contact: arthur.allignol at gmail.com
Version: 2022-03-07
URL: <https://CRAN.R-project.org/view=Survival>
Source: <https://github.com/cran-task-views/Survival/>

Testing

- The `survdiff` function in [survival](#) compares survival curves using the Fleming-Harrington G-rho family of test. [NADA](#) implements this class of tests for left-censored data.
- The [maxcombo](#) package compares survival curves using the max-combo test, which is often based on the Fleming-Harrington G-rho family of tests and is designed to have higher power than the logrank test in the scenario of non-proportional hazards such as those resulting from delayed treatment effects.
- [clinfun](#) implements a permutation version of the logrank test and a version of the logrank that adjusts for covariates.
- The [exactRankTests](#) implements the shift-algorithm by Streitberg and Roehmel for computing exact conditional p-values and quantiles, possibly for censored data.
- `survTest` in the [coin](#) package implements the logrank test reformulated as a linear rank test.
- The [maxstat](#) package performs tests using maximally selected rank statistics.
- The [interval](#) package implements logrank and Wilcoxon type tests for interval-censored data.
- Three generalised logrank tests and a score test for interval-censored data are implemented in the [glt](#) ([archived](#)) package.
- [survcomp](#) compares 2 hazard ratios.
- The [TSHRC](#) implements a two stage procedure for comparing hazard functions.
- The [FHTest](#) package offers several tests based on the Fleming-Harrington class for comparing survival curves with right and interval censored data.
- The [LogrankA](#) ([archived](#)) package provides a logrank test for which aggregated data can be used as input.
- The short term and long term hazard ratio model for two samples survival data can be found in the [YPmodel](#) package.
- The [controlTest](#) implements a nonparametric two-sample procedure for comparing the median survival time.
- The [survRM2](#) package performs two-sample comparison of the restricted mean survival time
- The [emplik2](#) package permits to compare two samples with censored data using empirical likelihood ratio tests.
- The [KONPsurv](#) package provides powerful nonparametric K-sample tests for right-censored data. The tests are consistent against any differences between the hazard functions of the groups.

tricolore

Together with



Ilya Kashnitsky
@ikashnitsky

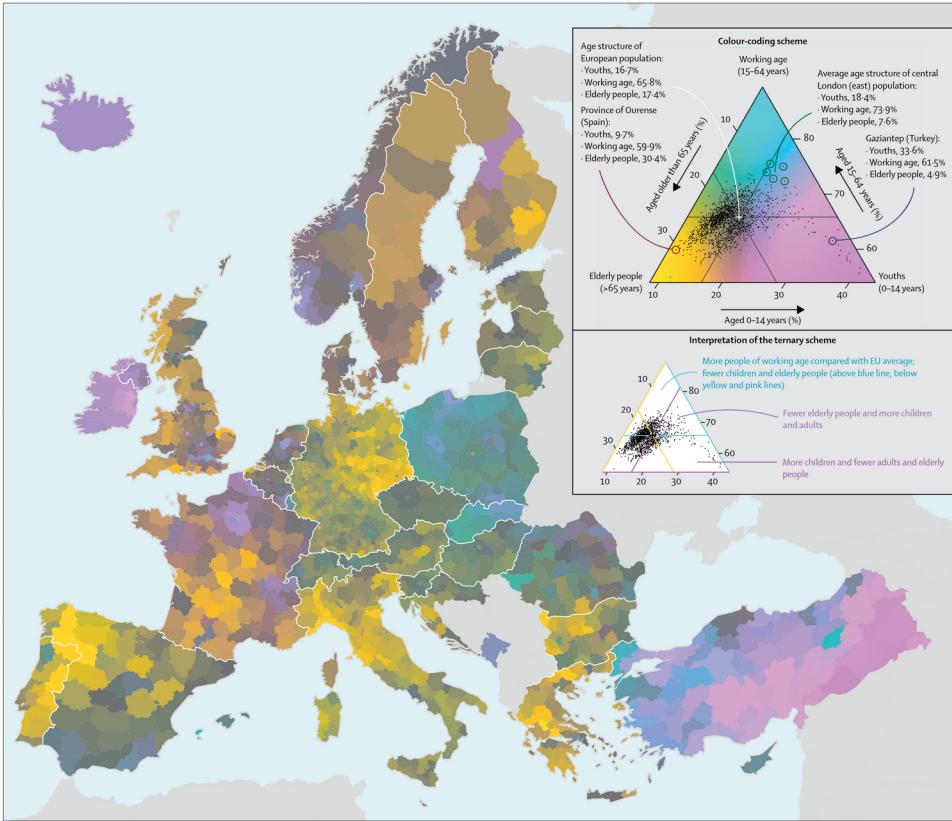


Figure: Colour-coded map of population structures in European Nomenclature of Territorial Units for Statistics 3 regions in 2015
Each population composition in the European Nomenclature of Territorial Units for Statistics 3 regions is uniquely colour coded. Colours show direction and magnitude of deviations from the centre point, which represents the average age of the European population, and is dark grey. The hue component of a colour encodes the direction of deviation: yellow indicates an elderly population (>65 years), cyan indicates people of working age (15–64 years), and magenta indicates children (0–14 years). Chroma and lightness components signify the distance from the centre ranging from desaturated and dark colours near the centre to vivid and bright colours at the corners. We provide R code to fully reproduce this map.¹⁰

Kashnitsky & Schöley (2018). Regional population structures at a glance. [10.1016/S0140-6736\(18\)31194-2](https://doi.org/10.1016/S0140-6736(18)31194-2)

tricolore

Together with



Ilya Kashnitsky
@ikashnitsky

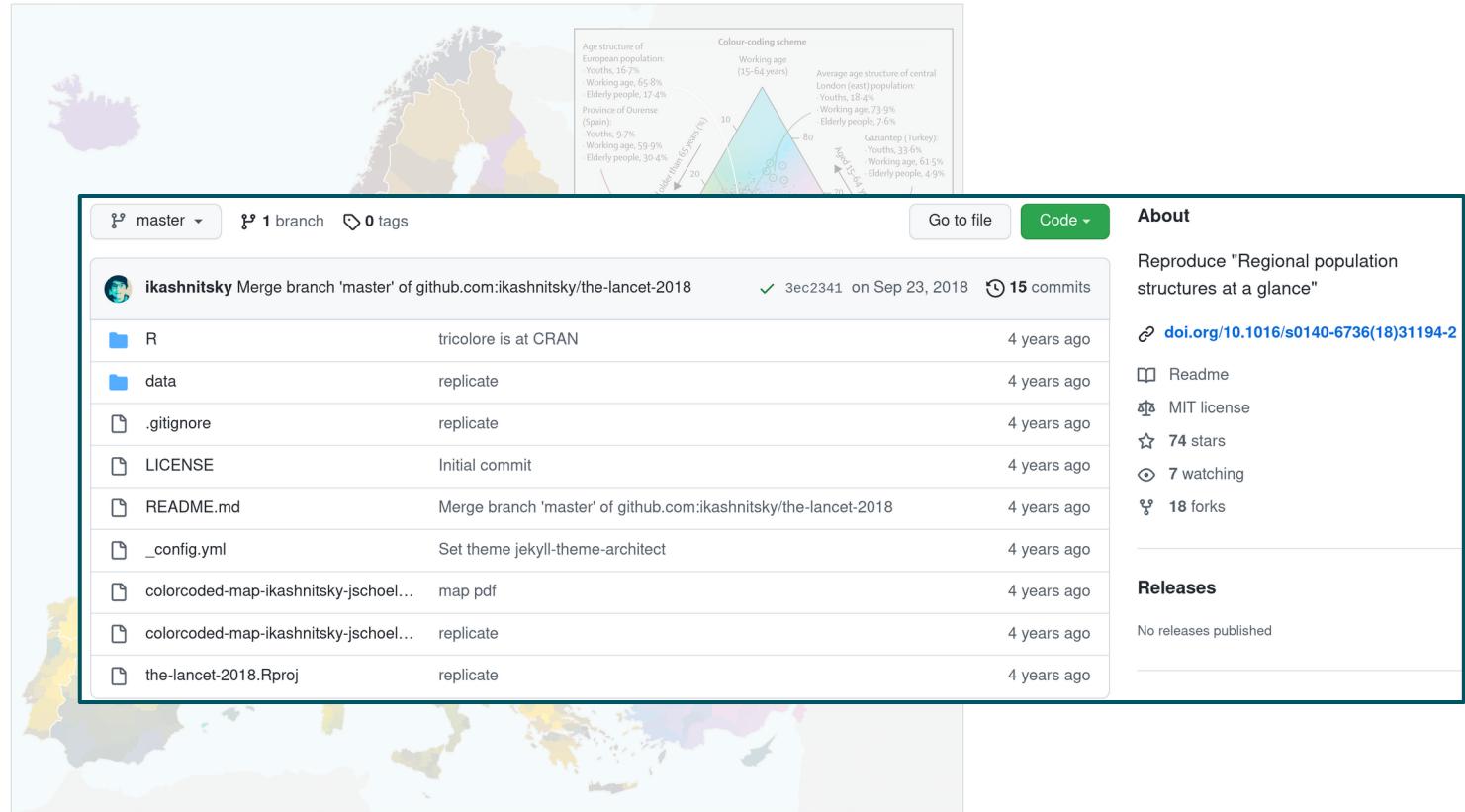
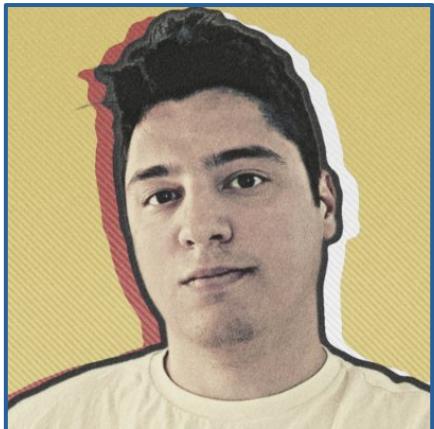


Figure: Colour-coded map of population structures in European Nomenclature of Territorial Units for Statistics 3 regions in 2015. Each population composition in the European Nomenclature of Territorial Units for Statistics 3 regions is uniquely colour coded. Colours show direction and magnitude of deviations from the centrepoint, which represents the average age of the European population, and is dark grey. The hue component of a colour encodes the direction of deviation: yellow indicates an elderly population (>65 years), cyan indicates people of working age (15–64 years), and magenta indicates children (0–14 years). Chroma and lightness components signify the distance from the centre ranging from desaturated and dark colours near the centre to vivid and bright colours at the corners. We provide R code to fully reproduce this map.¹⁰

Kashnitsky & Schöley (2018). Regional population structures at a glance. [10.1016/S0140-6736\(18\)31194-2](https://doi.org/10.1016/S0140-6736(18)31194-2)

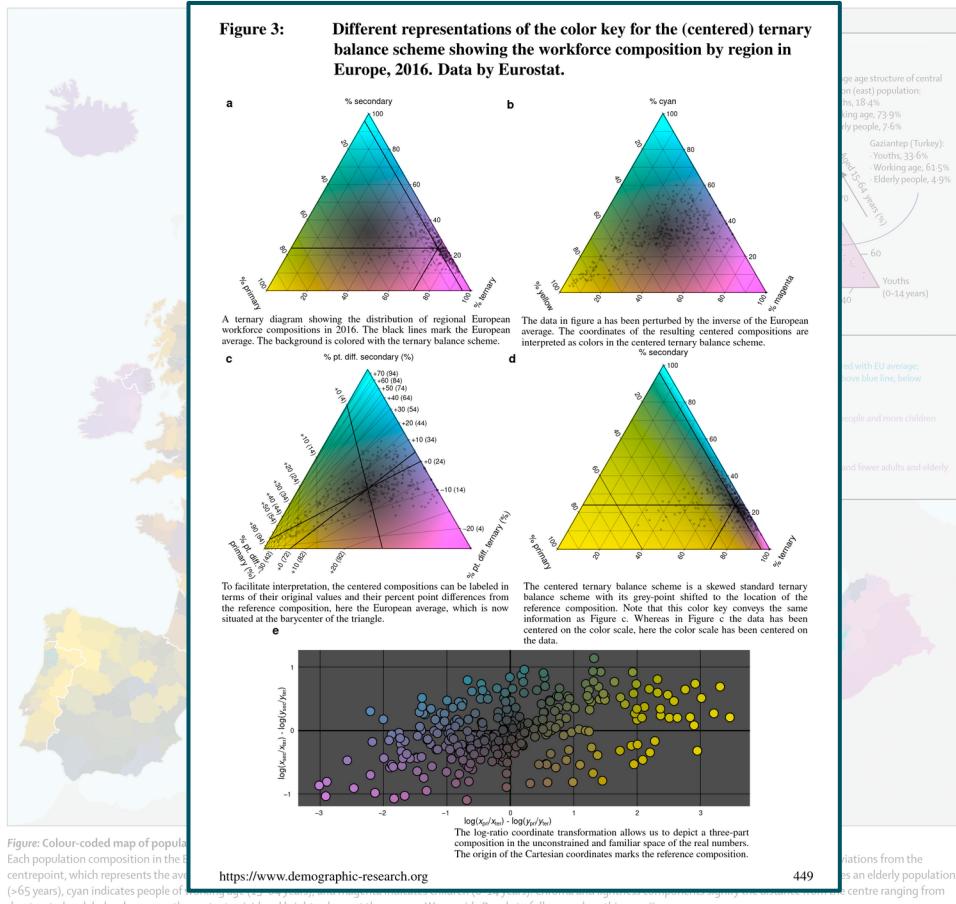
tricolore

Together with



Ilya Kashnitsky

@ikashnitsky



OpenScience 25

Schöley & Kotov – Layers of Reproducibility

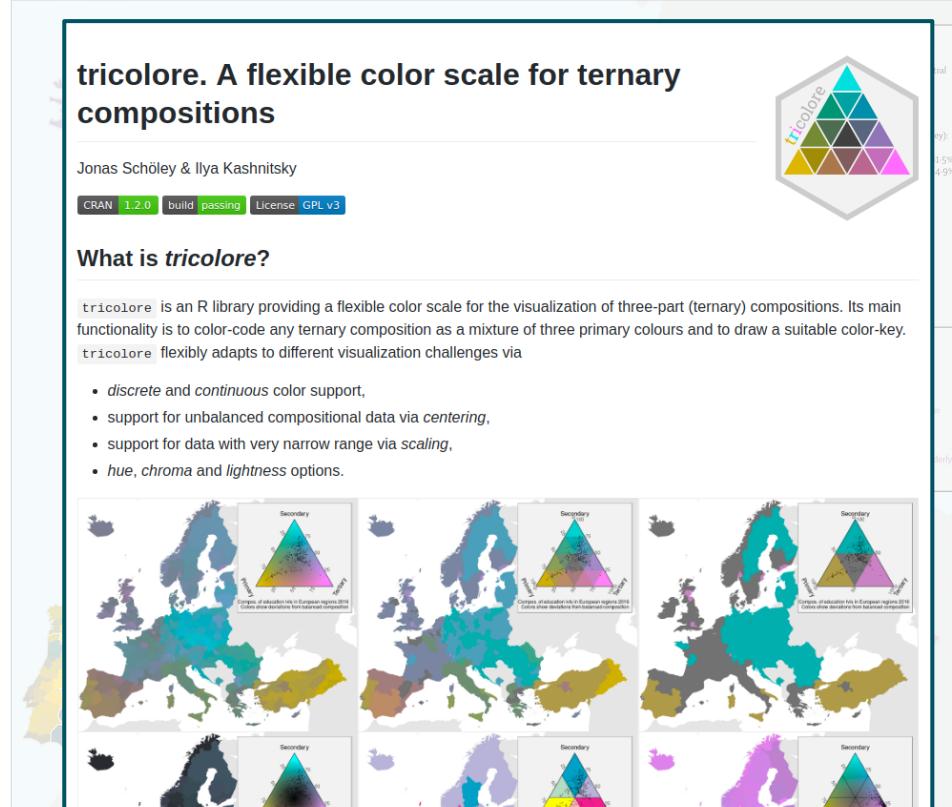
6

tricolore

Together with



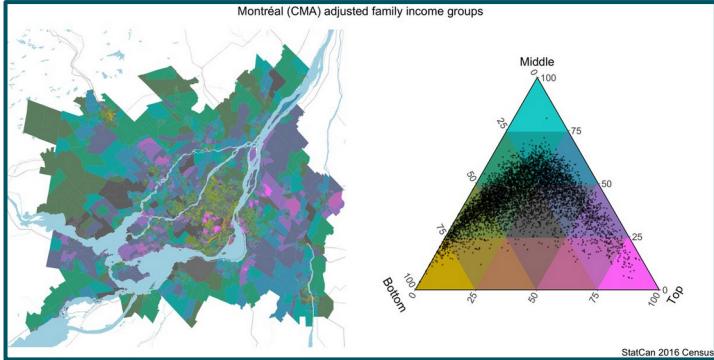
Ilya Kashnitsky
@ikashnitsky



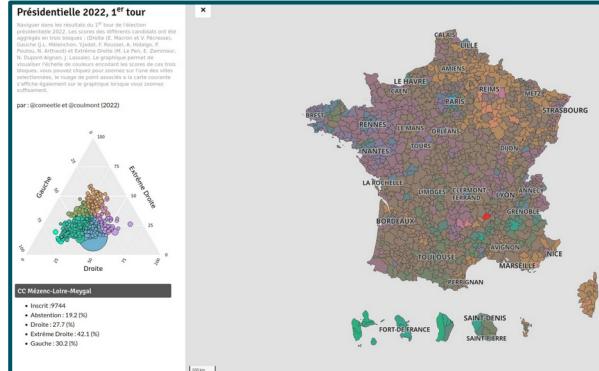
Kashnitsky & Schöley (2018). Regional population structures at a glance. [10.1016/S0140-6736\(18\)31194-2](https://doi.org/10.1016/S0140-6736(18)31194-2)

tricolore

Income distribution in Canadian cities



French election results



LMIC education disparity

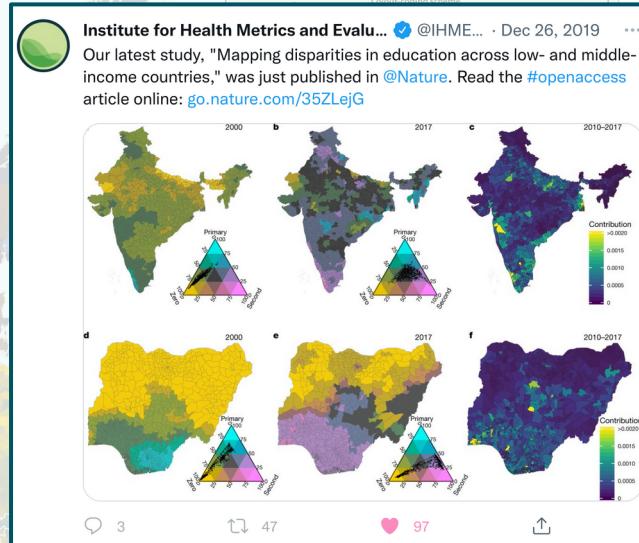
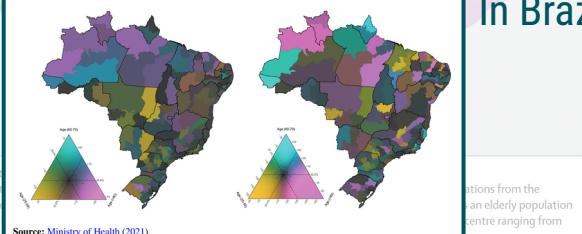
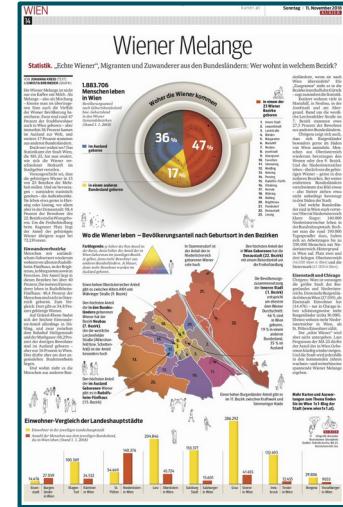


Figure 3:
Spatial distribution of deaths from COVID-19 by age group in Brazilian meso-regions, males – first wave (left) and second wave (right)

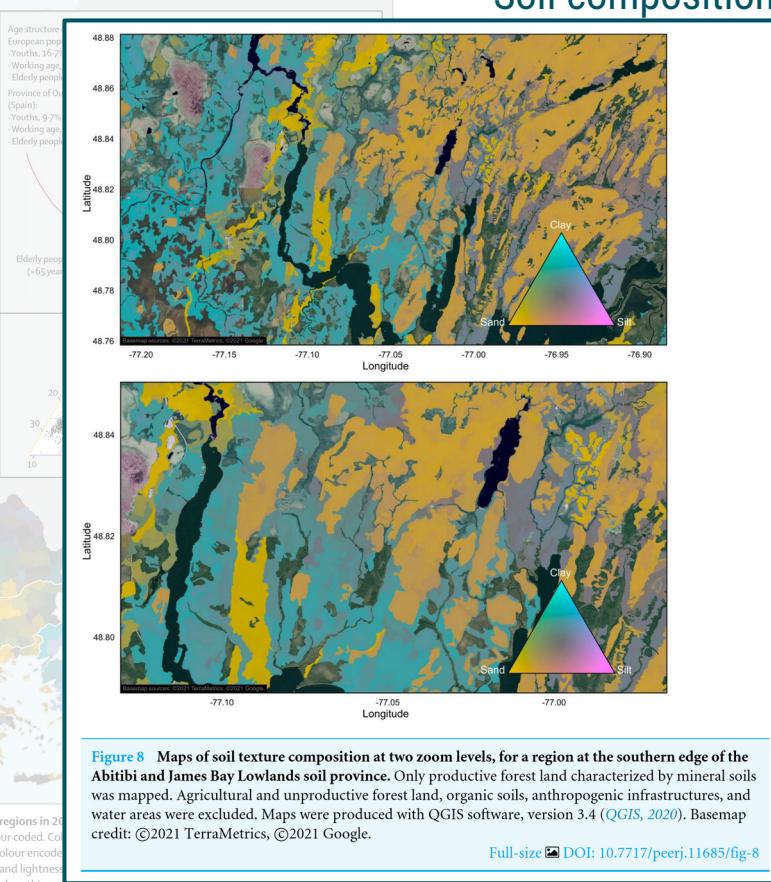
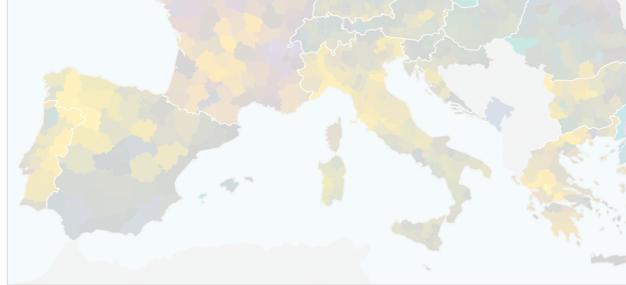
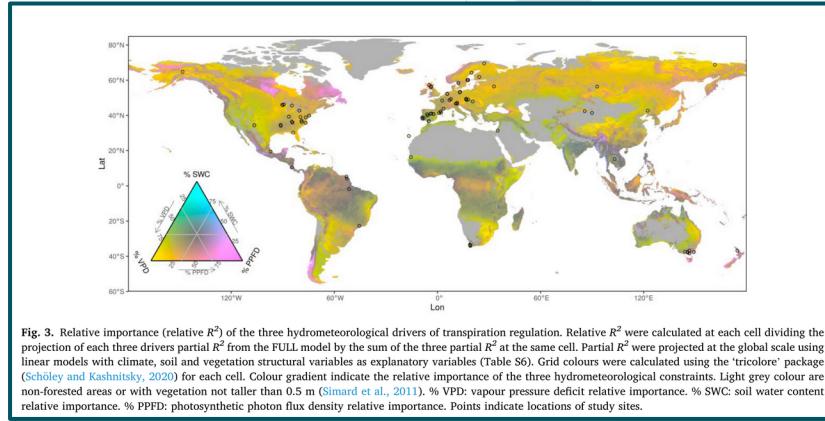


Vienna's population by origin



Regional age distribution of COVID deaths In Brazil

Agricultural and Forest Meteorology



Kashnitsky & Schöley (2018). Regional population structures at a glance. [10.1016/S0140-6736\(18\)31194-2](https://doi.org/10.1016/S0140-6736(18)31194-2)

Reproducible Demographers

Tim Riffe @timriffe1 & Enrique Acosta @Acosta_Kike_ &
Manal Elzalabany & Maxi Kniffka @MaxiKniffka & Jessica Donzowa @jdonzowa

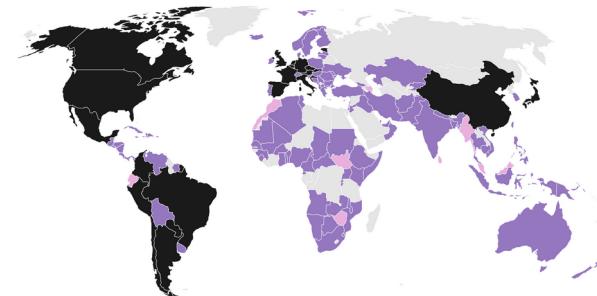


Created a fully reproducible data base of age specific COVID-19 statistics.

Data availability

You can get the most up-to-date data at the OSF site that we mirror to: <https://osf.io/mpwjq/>.

Here's an overview of global coverage as of now. A country marked as *forthcoming* means we've identified a source, but that collection is pending for one reason or another. Are you from one of the countries not yet in the collection and want to pitch in? Are you interested in adopting collection for one of our time series that has fallen behind? We're in need of more support. Please reach out, if so by emailing us at `coverage-db <at> demogr.mpg.de`.



Reproducible Demographers

Christina Bohk-Ewald



- Method00_FreezeRates
- Method01_Hadwiger1940
- Method02_CoaleMcNeil1972
- Method03_CoaleTrussell1974
- Method04_Brass1974
- Method05_Evans1986
- Method06_Chandola1999
- Method07_Schmertmann2003
- Method08_PeristeraKostaki2007M1
- Method09_PeristeraKostaki2007M2
- Method10_MyrskylaGoldstein2013
- Method11_Saboi1977
- Method12_WillekensBaydar1984
- Method13_deBeer1985and1989
- Method14_Lee1993Log
- Method16_HyndmanUllah2007
- Method17_ChengLin2010
- Method18_Myrskyla2013
- Method22_LiWu2003

Bohk et al. (2018). Forecast accuracy hardly improves with methods complexity when completing cohort fertility.
[10.1109/5.771073](https://doi.org/10.1109/5.771073)

Implemented, shared and compared 22 fertility forecasting methods

[github.com/fertility-forecasting/validate-forecast-methods
/tree/master/basic-scripts-forecast-methods](https://github.com/fertility-forecasting/validate-forecast-methods/tree/master/basic-scripts-forecast-methods)

Reproducible Demographers

Rob Hyndman



Author of countless R packages widely applied in demography.



`demography`: Forecasting Mortality, Fertility, Migration and Population Data

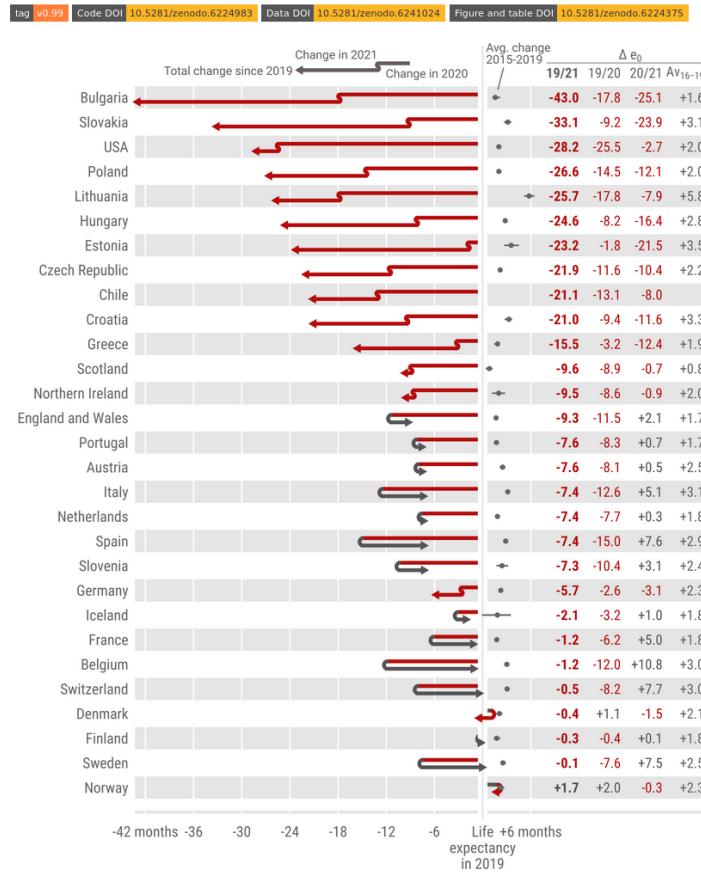
Functions for demographic analysis including lifetable calculations; Lee-Carter modelling; functional data analysis of mortality rates, fertility rates, net migration numbers; and stochastic population forecasting.

Version: 1.22
Depends: R (\geq 3.4), `forecast` (\geq 8.5)
Imports: `ftsa` (\geq 4.8), `rainbow`, `cobs`, `mgee`, `strucchange`, `RCurl`
Published: 2019-04-22
Author: Rob J Hyndman with contributions from Heather Booth, Leonie Tickle and John Maindonald.
Maintainer: Rob J Hyndman <Rob.Hyndman@monash.edu>
BugReports: <https://github.com/robjhyndman/demography/issues>
License: [GPL-2](#) | [GPL-3](#) [expanded from: GPL (\geq 2)]
URL: <https://github.com/robjhyndman/demography>
NeedsCompilation: no
Materials: [README](#) [ChangeLog](#)
CRAN checks: [demography results](#)

github.com/robjhyndman

Consuming open science

Life expectancy changes since COVID-19



2.2 Data quality. Data: I see that all data were downloaded on February 18, 2022. Can you provide more information about the level of completeness or comparability of completeness across countries as of this download date? I am only familiar with the US data, which have a very long lag time for all-cause mortality and so any counts obtained are likely an underestimate as of this date. Even if they technically cover deaths through the end of 2021, many deaths get reported and processed at later dates. I would also recommend updating these estimates with the most recently available data when doing the next revision.

Consuming open science

akarlinsky Local Mortality Update		d3a6d38 11 hours ago	579 commits
 local_mortality	Local Mortality Update		11 hours ago
 preliminary_mortality	Preliminary Mortality update		15 days ago
 .gitignore	Update .gitignore		5 months ago
 LICENSE	Create LICENSE		11 months ago
 README.md	2022-06-07 Update		9 days ago
 coverage_map_title.png	Update coverage_map_title.png		2 months ago
 world_mort_plot_all.png	2022-06-10 Update		6 days ago
 world_mortality.csv	2022-06-10 Update		6 days ago

Karlinsky & Kobak (2022). World Mortality Database. github.com/akarlinsky/world_mortality

Consuming open science

main ▾ 1 branch 0 tags

akarlinsky Local Mortality Update

local_mortality	Local Mortality Update
preliminary_mortality	Preliminary Mortality update
.gitignore	Update .gitignore
LICENSE	Create LICENSE
README.md	2022-06-07 Update
coverage_map_title.png	Update coverage_map_title.p
world_mort_plot_all.png	2022-06-10 Update
world_mortality.csv	2022-06-10 Update

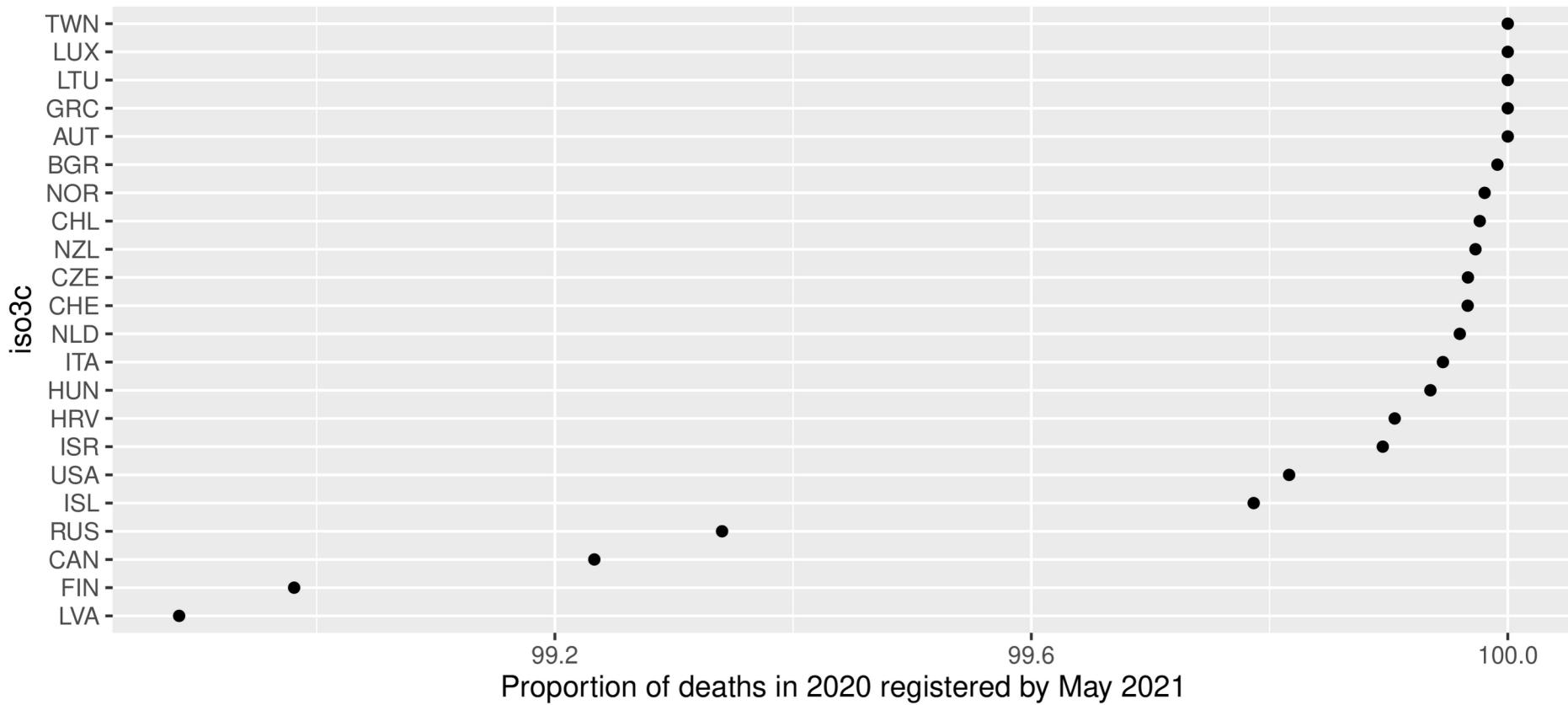
- o Commits on May 26, 2022
 - 2022-05-26 update ...
akarlinsky committed 22 days ago
- o Commits on May 24, 2022
 - 2022-05-24 update ...
akarlinsky committed 24 days ago
- o Commits on May 23, 2022
 - 2022-05-23 Update ...
akarlinsky committed 24 days ago
- o Commits on May 19, 2022
 - Update world_mortality.csv
akarlinsky committed 29 days ago
 - 2022-05-19 Update ...
akarlinsky committed 29 days ago
- o Commits on May 15, 2022
 - 2022-05-15 Update ...
akarlinsky committed on May 15

Go to file Code ▾

579 commits

Karlinsky & Kobak (2022). World Mortality Database. github.com/akarlinsky/world_mortality

Consuming open science



Derived from Karlinsky & Kobak (2022). World Mortality Database.
github.com/akarlinsky/world_mortality

Given **your data** and **your analysis**
I arrive at **your results**

Given **your research question,**
my data and **my analysis**
I arrive at **your results**

4 Layers

Layer 1: Aspirational

Layer 2: Personal

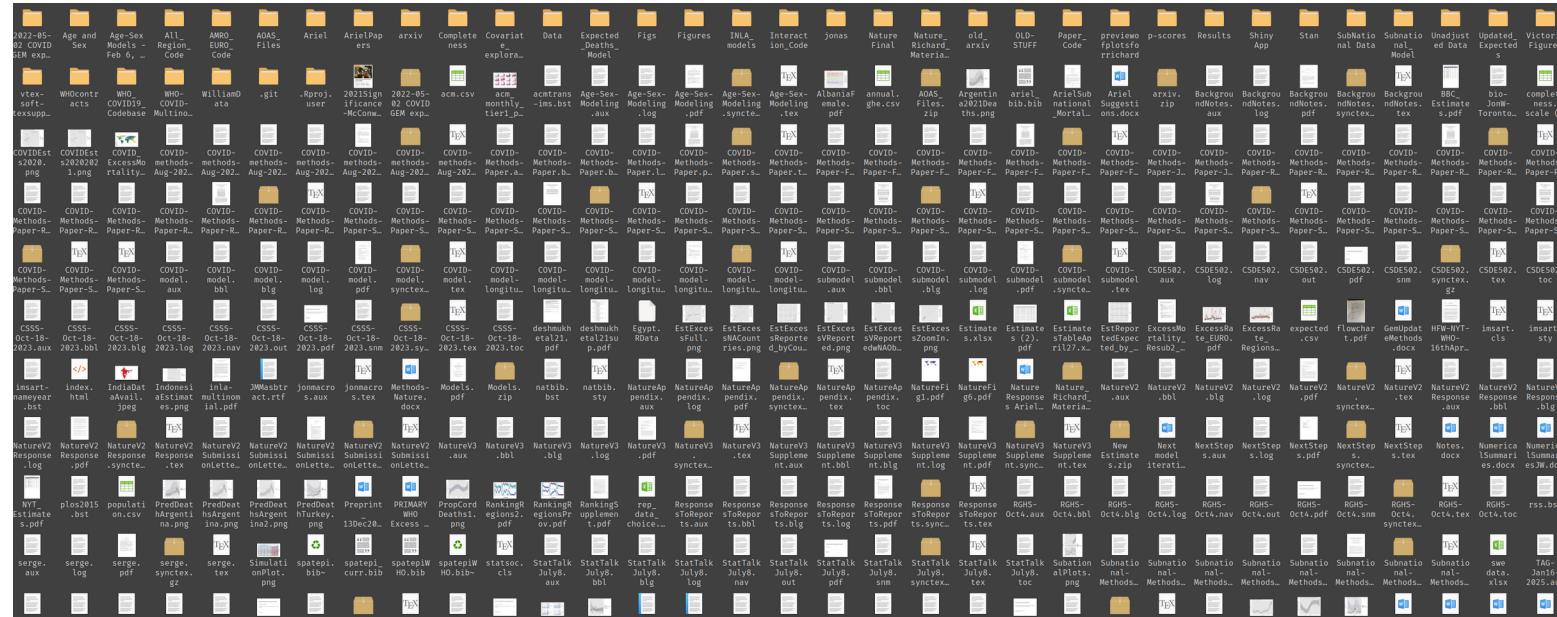
Layer 3: Communal

Layer 4: Computational

4 Layers

Layer 1: Aspirational

Realize you're having a problem



Layer 1: Aspirational

Reproduce your own work

Project structure
Analysis pipeline
Documentation

Layer 2: Personal

Layer 3: Communal

Layer 4: Computational

Layer 1: Aspirational

Layer 2: Personal

Layer 3: Communal

Layer 4: Computational

Let others reproduce your work



git



GitHub

zenodo

4 Layers

Layer 1: Aspirational

Layer 2: Personal

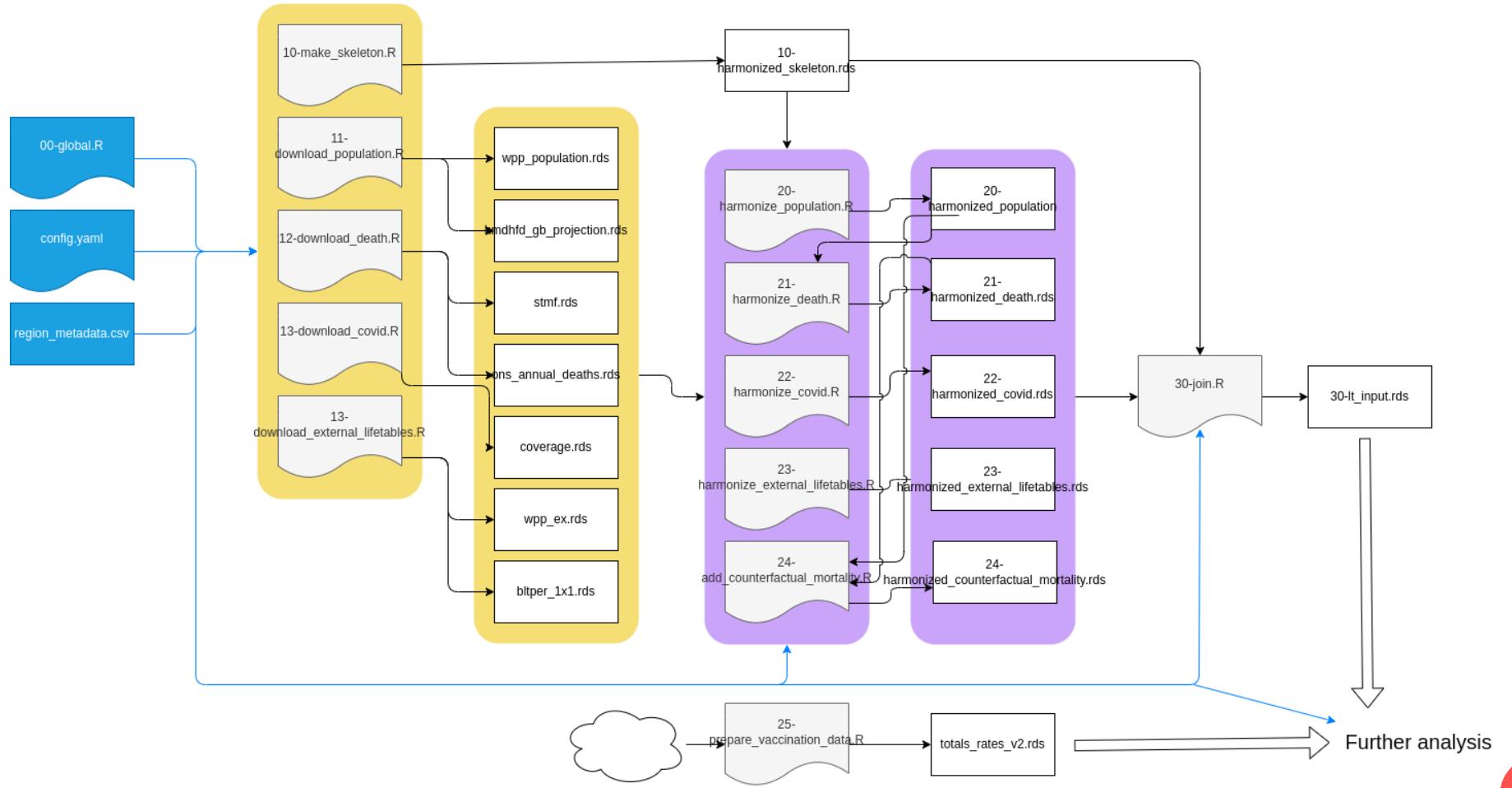
Layer 3: Communal

Layer 4: Computational

Automatize the reproduction of your work



The dependency graph



Reproducible analysis

github.com/jschoeley/openscience25

Jonas Schöley

 @jschoeley

 0000-0002-3340-8518

 schoeley@demogr.mpg.de



MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH