

# ukpopulation: unified national and subnational population estimates and projections, including variants

Andrew P Smith<sup>1</sup> and Tom Russell<sup>2</sup>

<sup>1</sup> School of Geography and Leeds Institute for Data Analytics, University of Leeds <sup>2</sup> Environmental Change Institute, School of Geography and the Environment, University of Oxford

DOI: [10.21105/joss.00803](https://doi.org/10.21105/joss.00803)

## Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 19 June 2018

Published: 04 August 2018

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

## Summary

UK statistics agencies (ONS, StatsWales, NRSotland and NISRA) produce mid-year population estimates (MYE), national population projections (NPP), and subnational population projections (SNPP), by age and gender. Mid-year estimates and subnational population projections are at local authority district (or equivalent) geography; national projections are produced for England, Wales, Scotland and Northern Ireland respectively.

At the time of writing (June 2018), mid-year estimates are available from 1991 to 2016. Subnational population projections typically have a 25-year horizon and in most cases cover the period 2016-2041. National projections have a longer horizon, extending from 2016 to 2116.

The datasets are published online in different places and in different formats, depending on the originating agency. The data may require significant preprocessing/reformatting before it can be used programmatically and, crucially, reproducibly.

Population projection variants (e.g. high fertility) are extremely useful for scenario analyses and a number of population projection variants are available at national scale. However, the availability of variants at subnational scale is patchy at best.

This python package (Smith 2018) aims to unify the retrieval of UK-wide data, providing consistent interfaces for each of the three (MYE, SNPP, NPP) datasets. For efficiency all data is cached locally.

The package includes functionality to filter data by geography (e.g. for analysis of a single local authority), and/or by age and by gender, e.g. for analysis of working-age population.

It also provides a simple methodology for extrapolating the shorter-term SNPP data using the longer-term NPP data. By extrapolating the SNPP independently for each age and gender enables the age-gender structure of the original population to be captured. Aggregation only takes place on the extrapolated age-gender specific values. This means that the trends shown by SNPP geographies with different age-gender structures, will differ, even though they both extrapolate using the same NPP data. (If aggregation was performed before the extrapolation it would not be possible to compute different extrapolated trends for regions with different initial age-gender structures.)

This functionality enables researchers and policymakers to examine population growth trajectories and construct plausible projections, both principal and variant, at subnational scale, consistently for anywhere in the UK over longer time horizons than the official data permit. By automating the process of obtaining and consistently formatting the data, the package also makes it much easier to create reproducible analyses.

## References

Smith, Russell, Andrew P. 2018. “ukpopulation: Unified UK National and Subnational Population Projections.” <https://github.com/nismod/ukpopulation>.