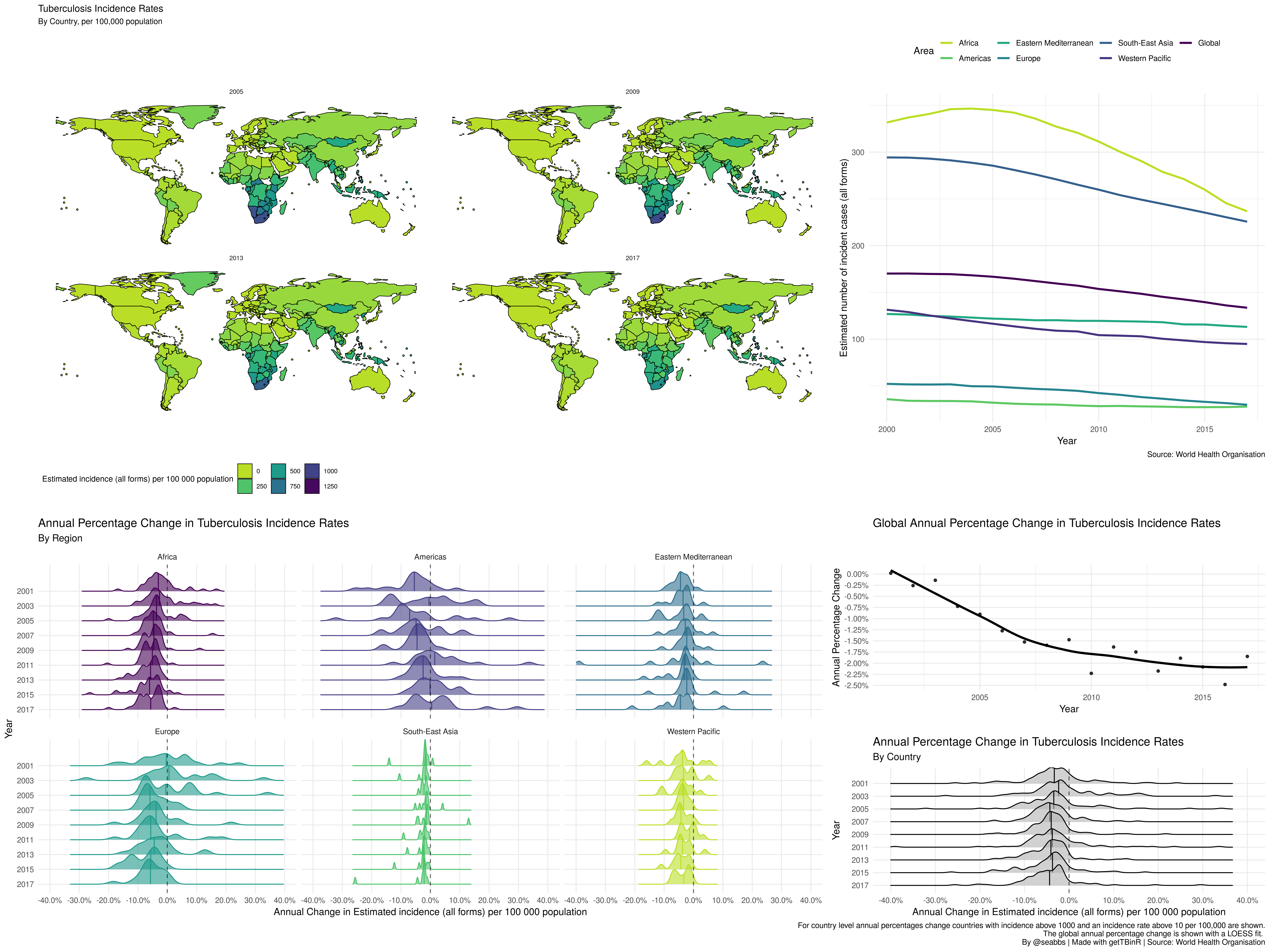
getTBinR: an R package for accessing and summarising World Health Organization Tuberculosis data



Global and regional Tuberculosis (TB) incidence rates, as well as trends in the annual percentage change of TB incidence rates. Generated using getTBinR.

## What is the issue?

Tuberculosis (TB) is one of the oldest human diseases, with recorded cases in ancient Egypt, renaissance Europe, and in the modern day across the globe. It is thought to infect over 1.7 billion people globally, of which 5-15% will develop symptomatic TB in their lifetime (World Health Organisation 2018). Of this number around 10% are likely to die from TB or TB related causes and globally TB remains the leading cause of death from infectious disease. TB is preventable and curable, but the majority of cases occur in less economically developed countries and are never diagnosed. Whilst global incidence rates have decreased year on year since the early 2000’s, global TB incidence remains above 134 per 100,000 people.

## Why is this package needed?

Developing tools for accessing and exploring data sets benefits the public health research community by enabling multiple data sets to be combined in a consistent manner, increasing their visibility, and providing a framework for exploring key questions. Tooling also reduces barriers to entry, allowing non-specialists to explore data sets that would otherwise be inaccessible. This widening of access may lead to new insights and wider interest for key public health issues.

## What does the package do?

getTBinR (Abbott 2019) is an R package (R Core Team 2019) to facilitate working with the data (World Health Organisation 2018) collected by the World Health Organization (WHO) on the country level epidemiology of Tuberculosis (TB). All data is freely available from the [WHO](https://www.who.int/tb/country/data/download/en/) and the package code is archived on Zenodo (Abbott 2019) and [Github](https://www.samabbott.co.uk/getTBinR/). The aim of getTBinR is to allow researchers, and other interested individuals, to quickly and easily gain access to a detailed TB data set and to start using it to derive key insights. It provides a consistent set of tools that can be used to rapidly evaluate hypotheses on a widely used data set before they are explored further using more complex methods or more detailed data. The functions provided in this package were developed to have sensible defaults to allow those new to the field too quickly gain key insights but also allow sufficient customisation so that experienced users may rapidly prototype new ideas.

## What data is available?

The data sourced by getBTinR is collected by the WHO, via member governments, and used to compile the yearly global TB report (World Health Organisation 2018). Data collection encompasses TB notifications, TB mortality rates, the proportion of drug resistant cases, case detection rates, treatment rates, interventions, outcomes, and planned intervention budgets. For a complete description of the data and data collection process, see (World Health Organisation 2018). These data are used by the WHO, governmental organisations and researchers to summarise country level TB epidemiology, as well as the wider global and regional picture.

## How can the package be used?

The figures at the top of this post were produced using a small subset of the functionality of getTBinR in R (R Core Team 2019). They show the global and regional trends in TB incidence rates and the annual change of those rates. They highlight the fact that TB reductions are no longer increasing annually and that regional trends differ (see [here](https://www.samabbott.co.uk/post/gettbinr-6-0/) for the code to generate these figures). Another example of a use of the package, to explore estimates of the TB case fatality ratio, can be seen [here](https://www.samabbott.co.uk/post/est-cfr-gettbinr/). In addition, getTBinR provides both a dashboard and an automated country level report that enables the global, regional, and country level picture to be quickly summarised. See [here](https://www.samabbott.co.uk/getTBinR/) for further examples and the package documentation.

## References

Abbott, Sam. 2019. “GetTBinR: An R Package for Accessing and Summarising the World Health Organisation Tuberculosis Data.” *Journal of Open Source Software* 4 (34): 1260. <https://doi.org/10.21105/joss.01260>.

Abbott, Sam. 2019. “Seabbs/getTBinR 0.6.0.” *Zenodo*. <https://doi.org/10.5281/zenodo.2547405>.

R Core Team. 2019. “R: A Language and Environment for Statistical Computing.” Vienna, Austria. <https://www.R-project.org/>.

World Health Organisation. 2018. “Global Tuberculosis Report.” <https://www.who.int/tb/country/data/download/en/>.