Descriptive analysis of harmonized broadband data

This document provides a brief descriptive overview of the harmonized German municipal broadband panel dataset (2005-2021).

Discontinuous changes in the data

The panel data is constructed from historical sources with changing methodologies. This results in two "breaks" in the time series that users must be aware of.

The 2009/2010 break

To create a continuous series for basic internet availability, the `share_broadband_baseline` variable was constructed from two different underlying sources:

- **For 2005-2009**: The variable is based *only* on historical data for basic DSL, defined as **>=0.128 Mbps**. This is the only variable that is present in pre-2010 data.
- **From 2010 onwards**: The variable is based on modern data for **>=1 Mbps** coverage. Since having >=1 Mbps implies having >=0.128 Mbps, this provides a consistent *lower bound* on broadband access.

To summarize, the `share_broadband_baseline` variable prior to 2010 is an exact share, while it is a lower bound starting in 2010.

The consequence is a "jump" in the data where the definition of the baseline metric changes. This is visible in the plots as a sharp jump between 2008 and 2010, when the data source switches. This is distinct from organic growth.

The 2015 break

As noted in the project documentation, 2015 marks a major structural break due to a change in the primary data provider and reporting standards. This affects all speed tiers and is visible as a very large, discontinuous jump in coverage levels. The `method_change_2015` dummy variable is included in the dataset to allow researchers to control for this event.

Descriptive plots

Average annual coverage

This plot shows the mean coverage for the baseline metric alongside higher speed tiers. The two key events are clearly visible: the **2010 jump** where the baseline definition changes, and the massive **2015 methodological break**.

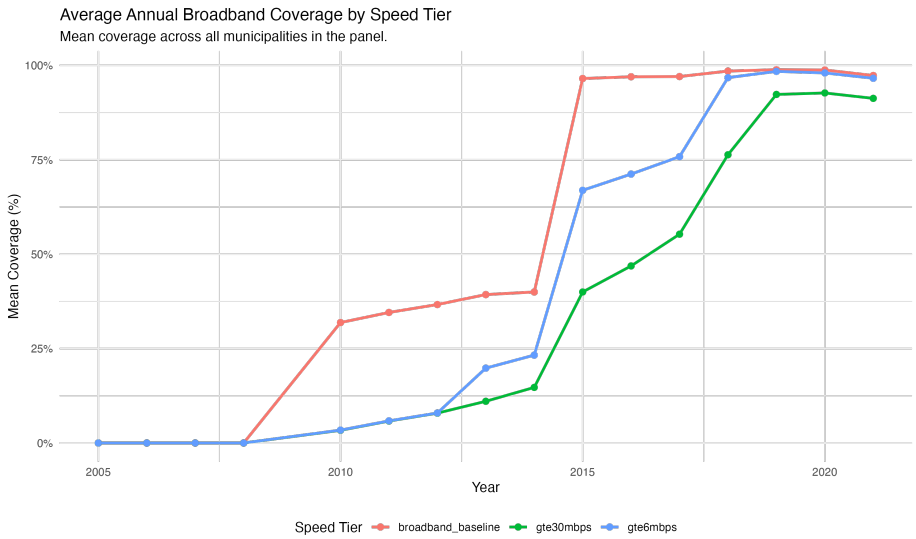

Figure 1: Average Annual Coverage

Year-over-year change in coverage

This plot highlights the magnitude of the two breaks. The jump in `baseline` coverage between 2008 and 2010 is substantial. However, it is dwarfed by the massive spike across all speed tiers in 2015.

Distribution of coverage levels

This set of histograms shows the distribution of coverage values. The `share_broadband_baseline` is heavily skewed towards 100%, indicating that most municipalities achieved full basic coverage relatively early. In contrast, higher speed tiers show more variation and a larger concentration near 0%.

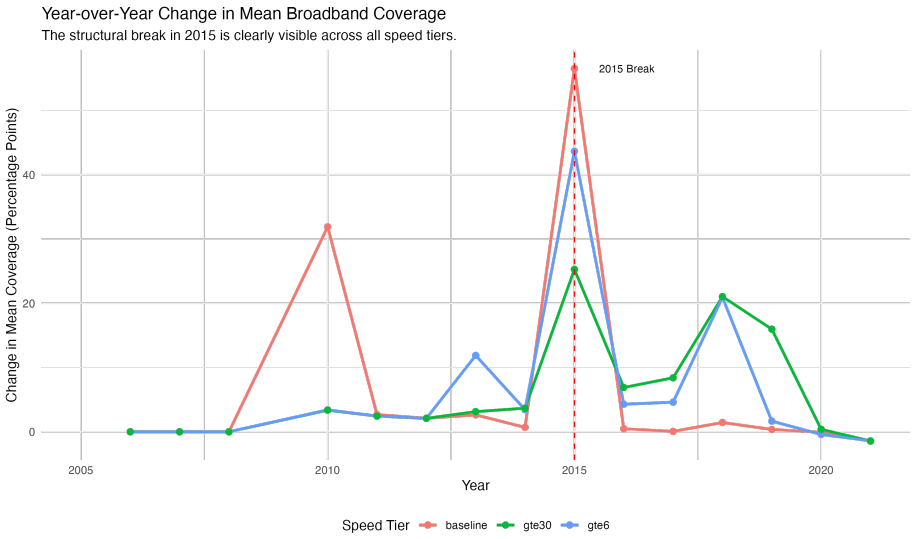

Figure 2: Year-over-Year Change in Coverage

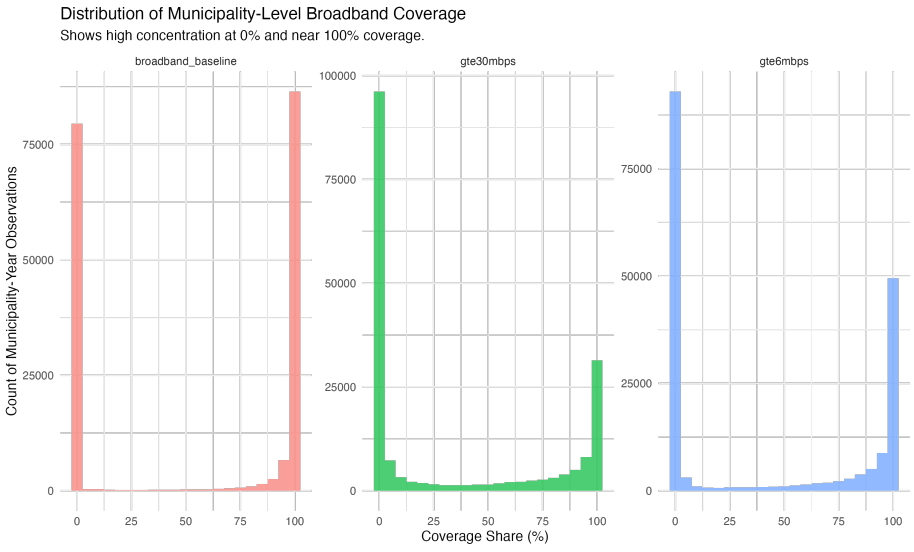

Figure 3: Distribution of Coverage Levels