Out-of-School Rate Estimation

Agenda

- Out-of-School Model
  - Modelling Strategy
  - Sample Results
- Data Issues
  - Identifying Bad Data
  - Managing Partial Data
  - Data Sourcing
  - Survey Data
  - Theoretical School Ages
  - High Priority Countries
- Dissemination
  - VIEW
  - Documentation

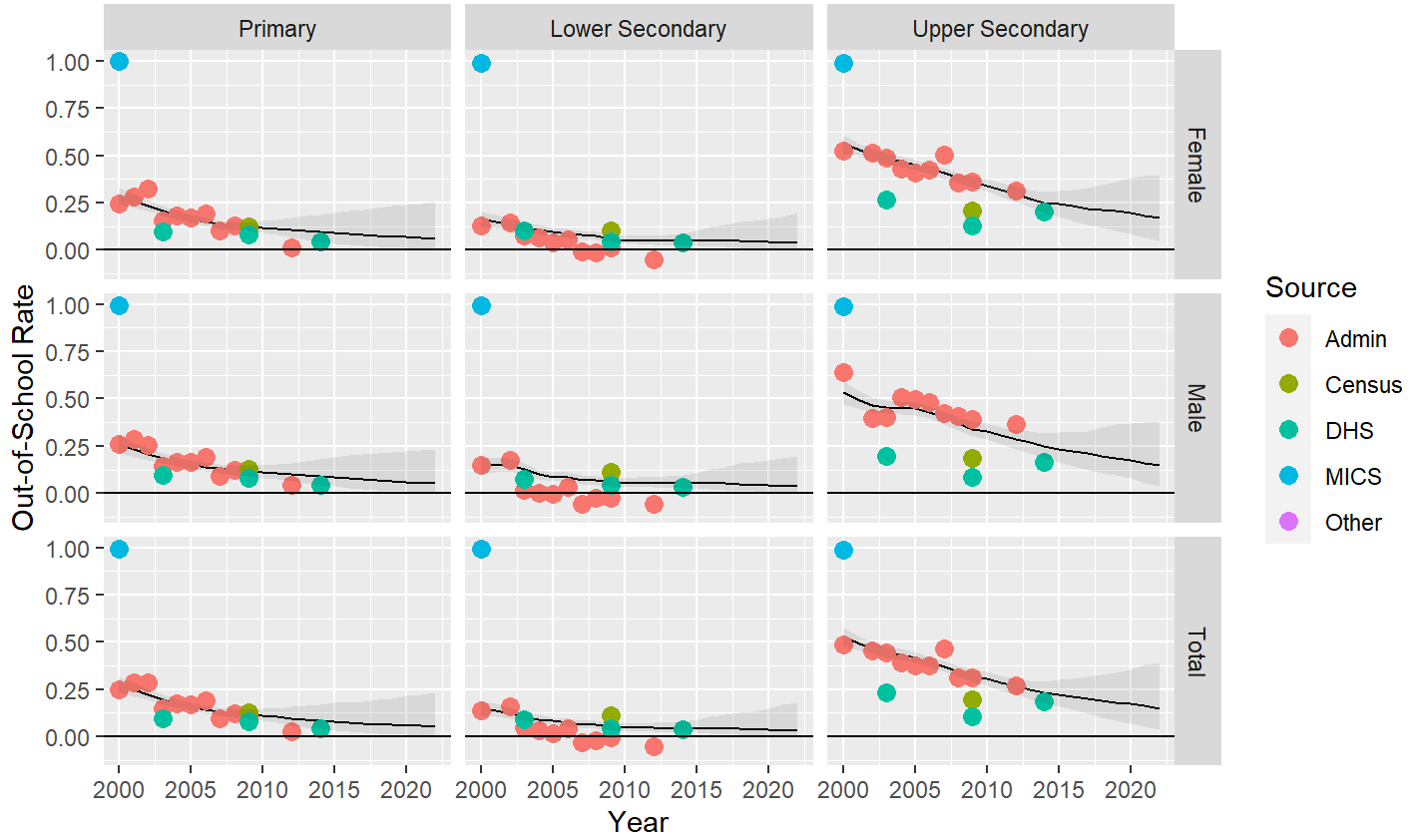
Out-of-School Model

Modelling Strategy

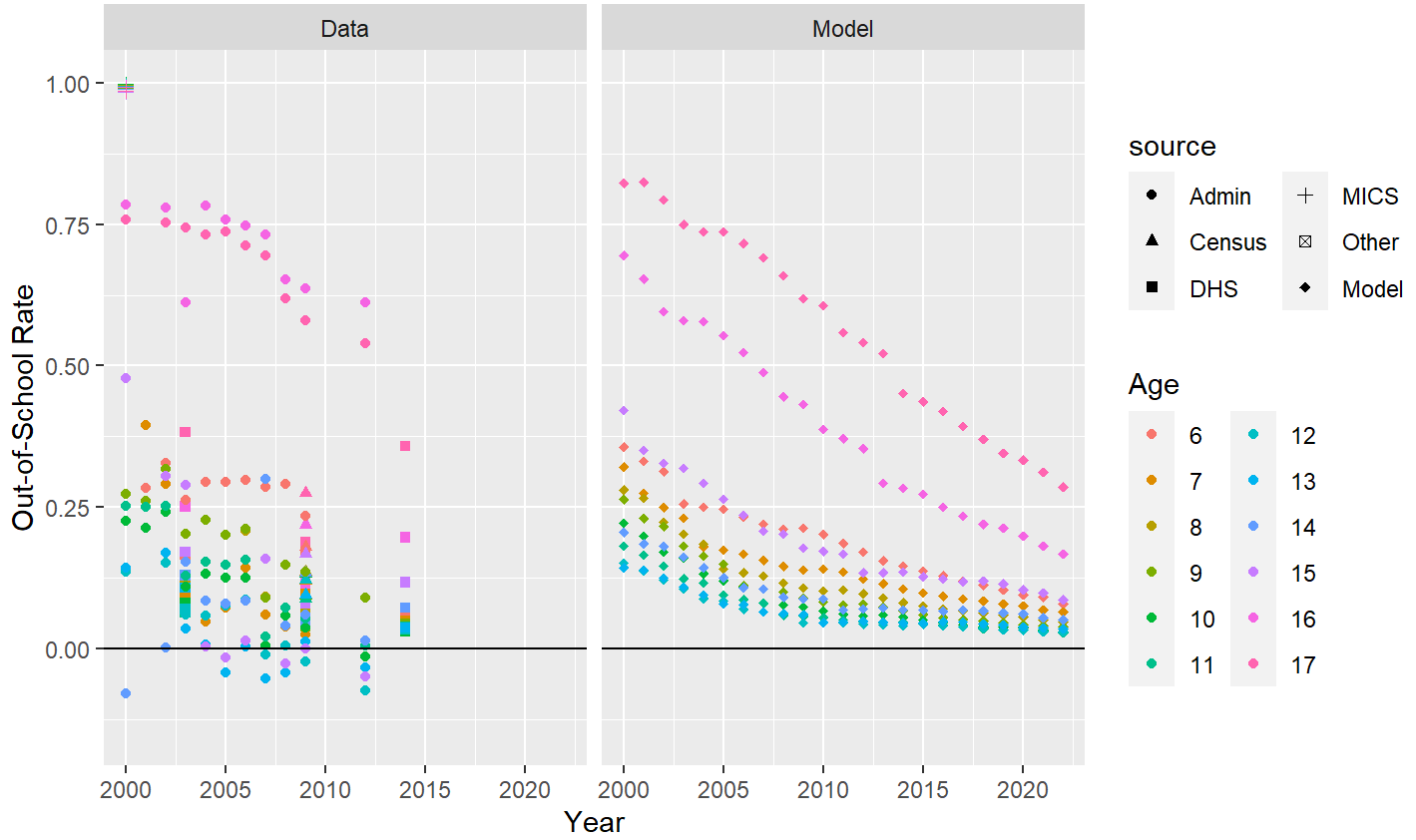
The following are the key observations and ideas underpinning our modelling approach

- Out-of-School (OOS) rates can be sourced either from administrative data or household survey data. The two types of sources have distinct challenges:
  - Administrative data is not subject to sampling errors but due to independent sourcing of enrollment and population counts, invalid observations are present. Data is also not always complete.
  - Survey data is infrequent, subject to sampling errors, and can be subject to large bias.
- While typically reported by level, OOS rates are not constant within a level, rather they are fluid and often form smooth curves that capture late entry/dropout patterns.
- There is a natural cohorting structure to exploit - students that dropout in one year and still absent the next.

Sample Results - Kenya

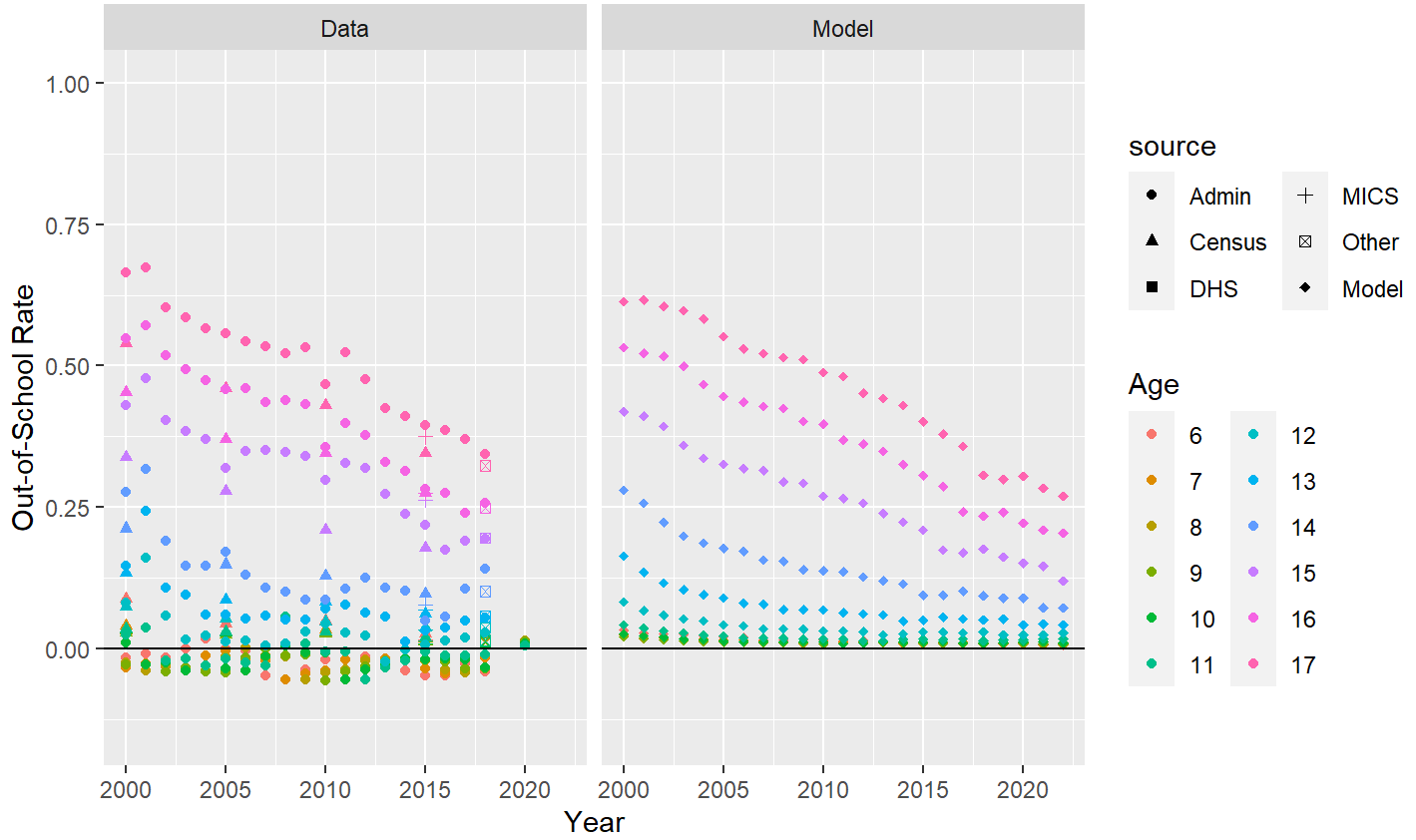


Sample Results - Kenya

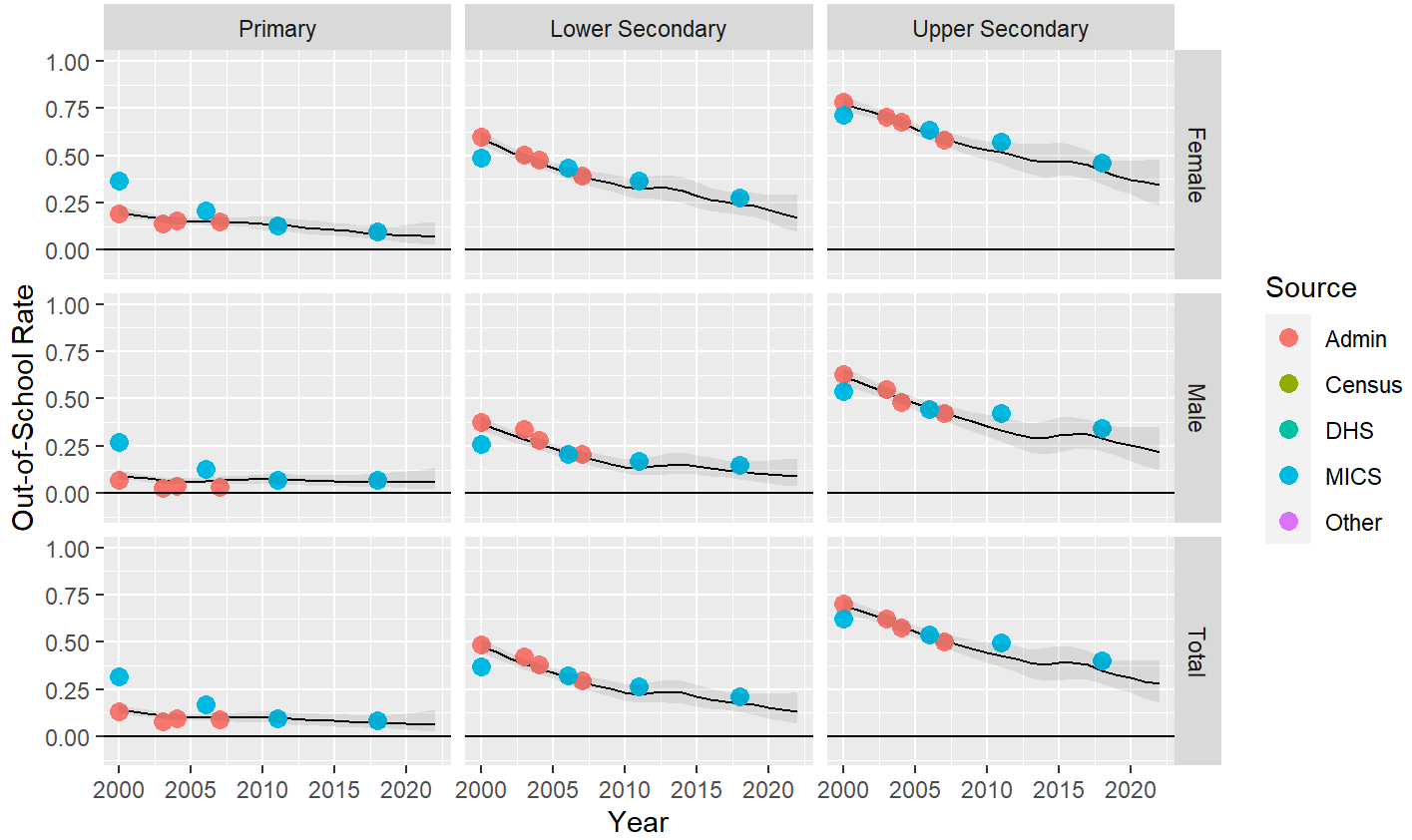


Sample Results - Mexico

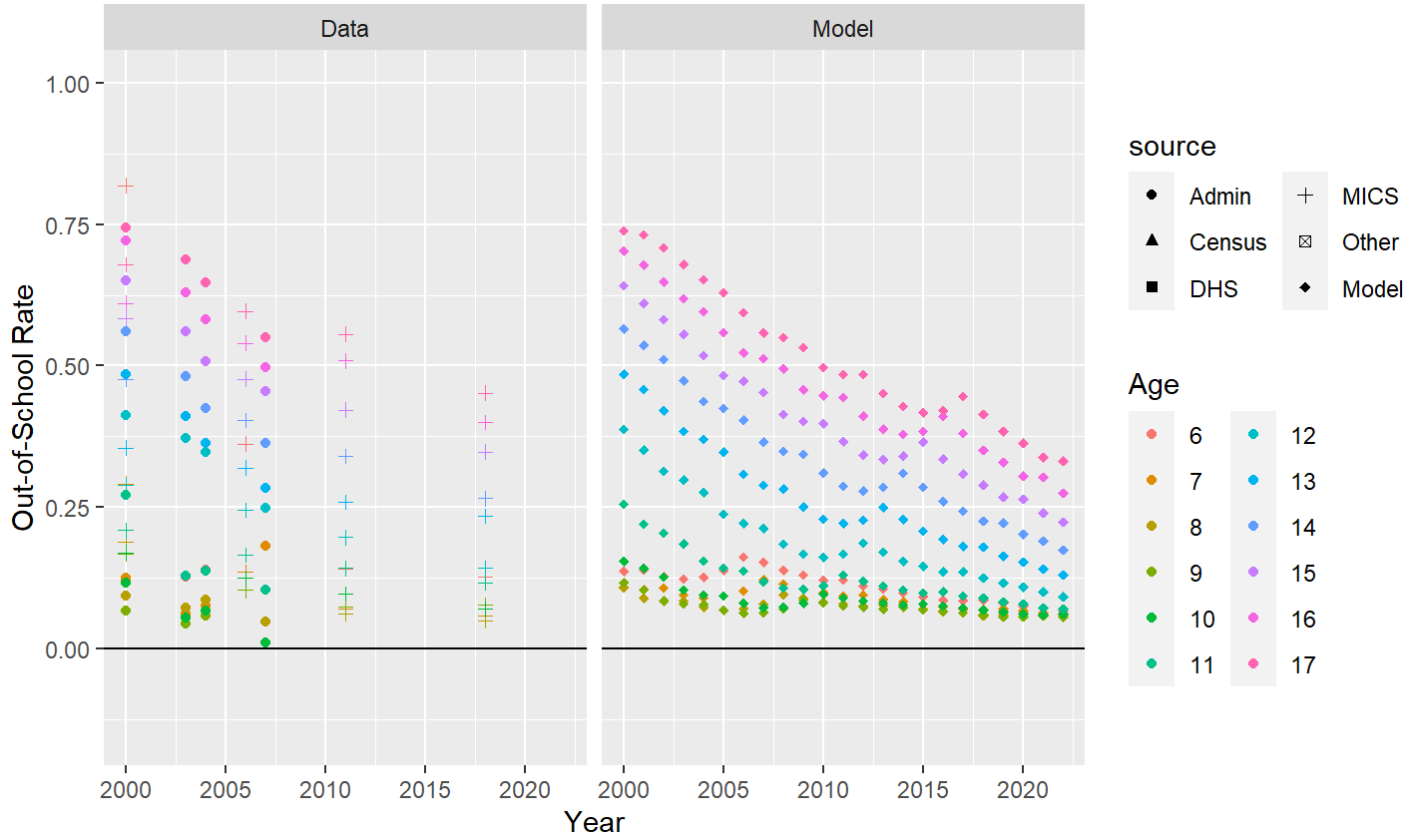
Sample Results - Mexico



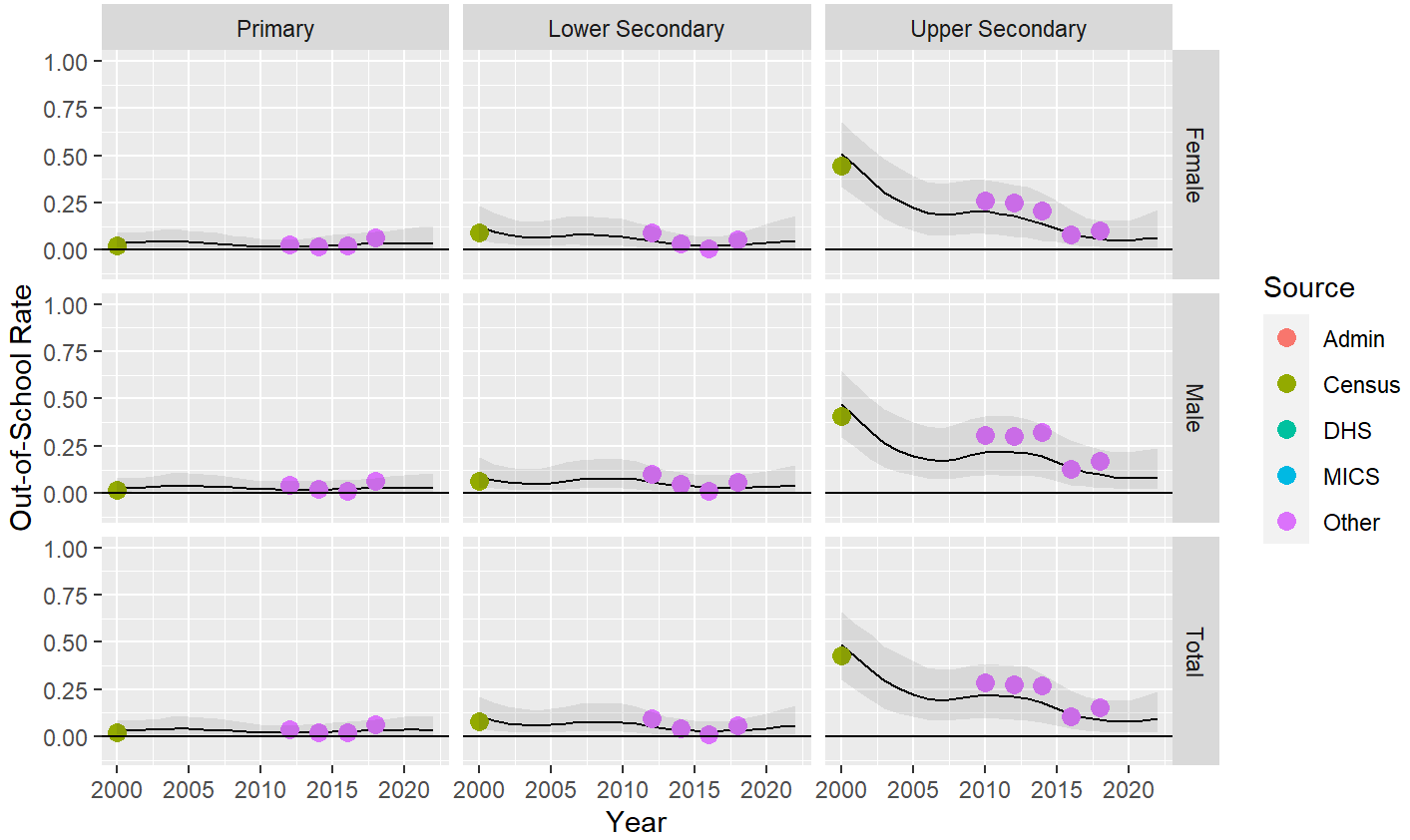
Sample Results - Iraq



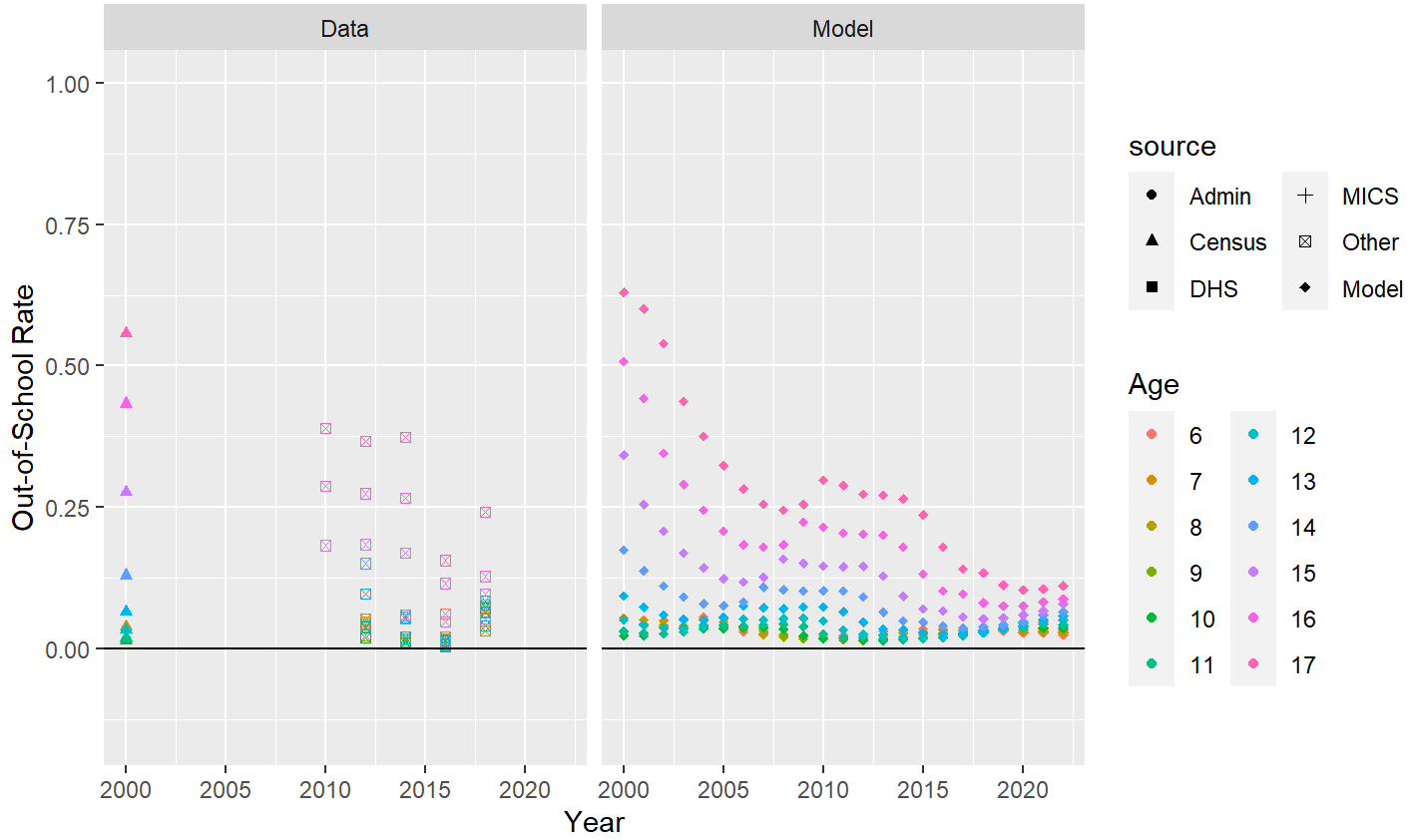
Sample Results - Iraq



Sample Results - China



Sample Results - China



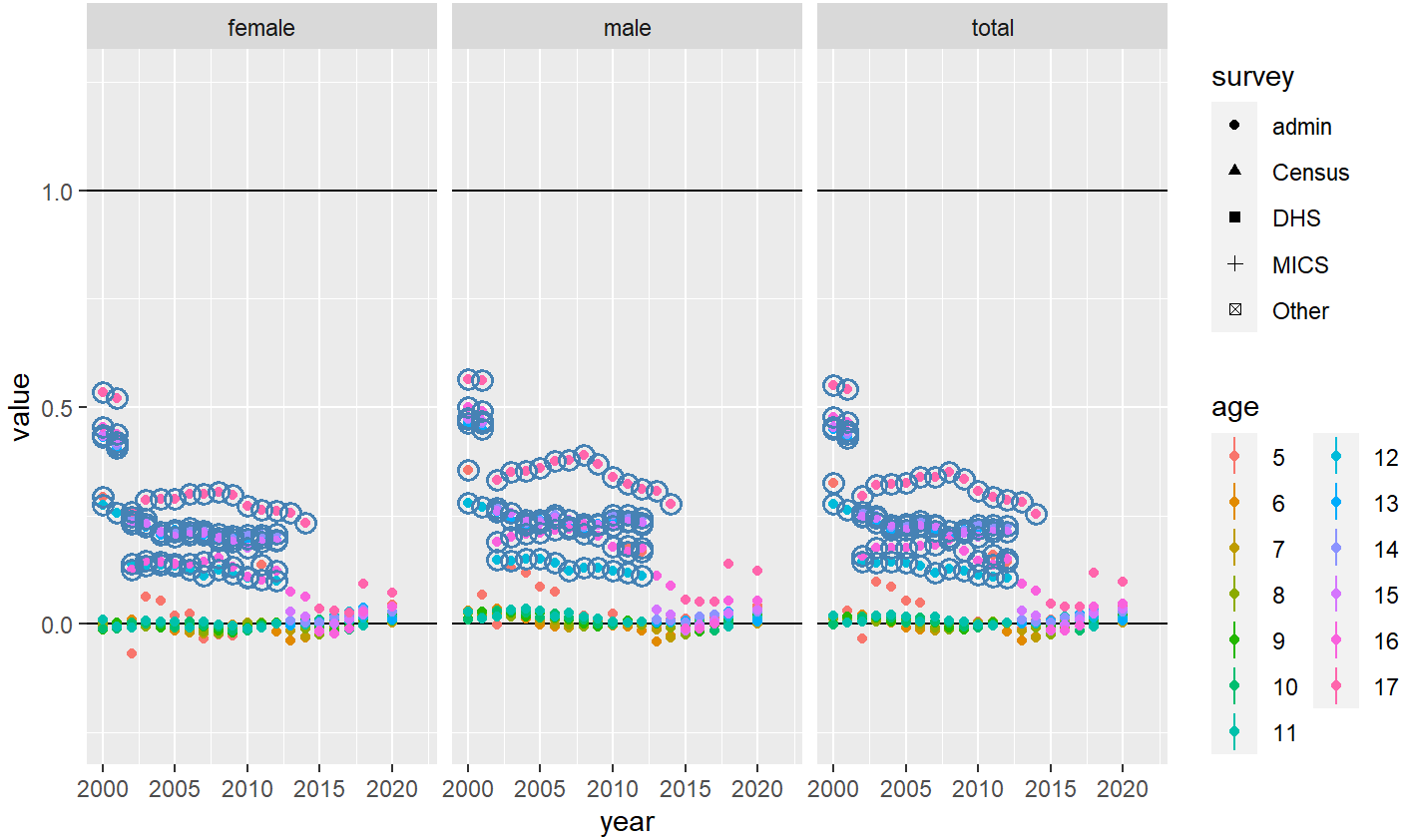
Data Issues

Our Process

- Administrative Data
  - For each age, aggregate enrollment counts for all levels.
  - Flag points as invalid if the count for the corresponding level is missing (ex. 7 year-olds must have primary enrollment).
  - If primary is missing, propagate missingness into lower and upper secondary ages.
  - Remove country-specific bad data identified by inspection and consultation.
  - Divide total enrollment for each age by the corresponding WPP population estimate and subtract from 1.
- Survey Data
  - Accumulate the observed attendance by age using observation weights.
  - Compute sampling variance using clustered jackknife.
  - No major material adjustments made.

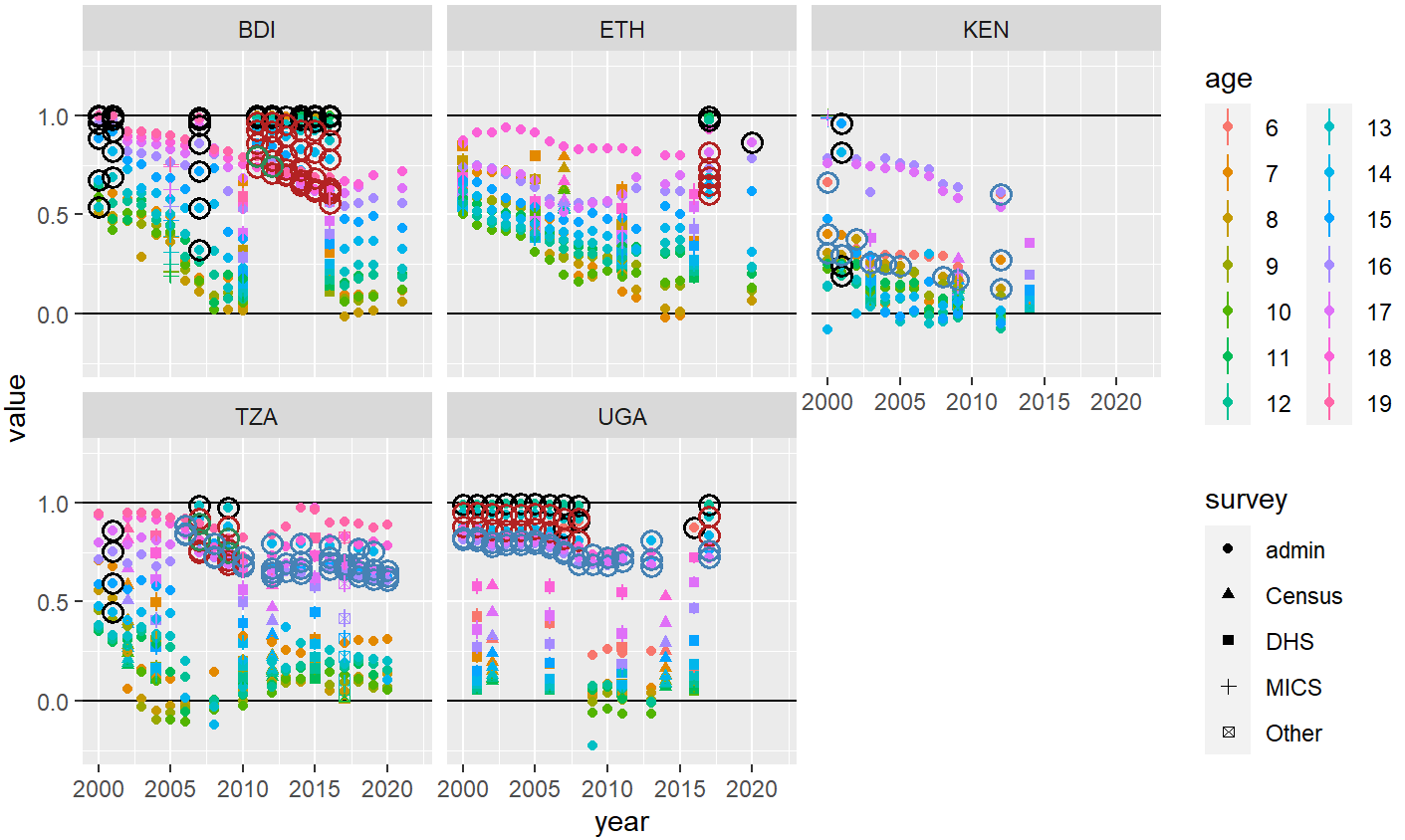
Identifying Bad Data

Frequently, for a given age, only some of the relevant counts are present. In Australia, pre-primary and upper secondary are absent in certain years.



Managing Partial Data

This problem is more difficult with lateness. When does missing primary stop being problematic? Burundi/Ethiopia report primary for all ages, the others do not.



Data Sourcing

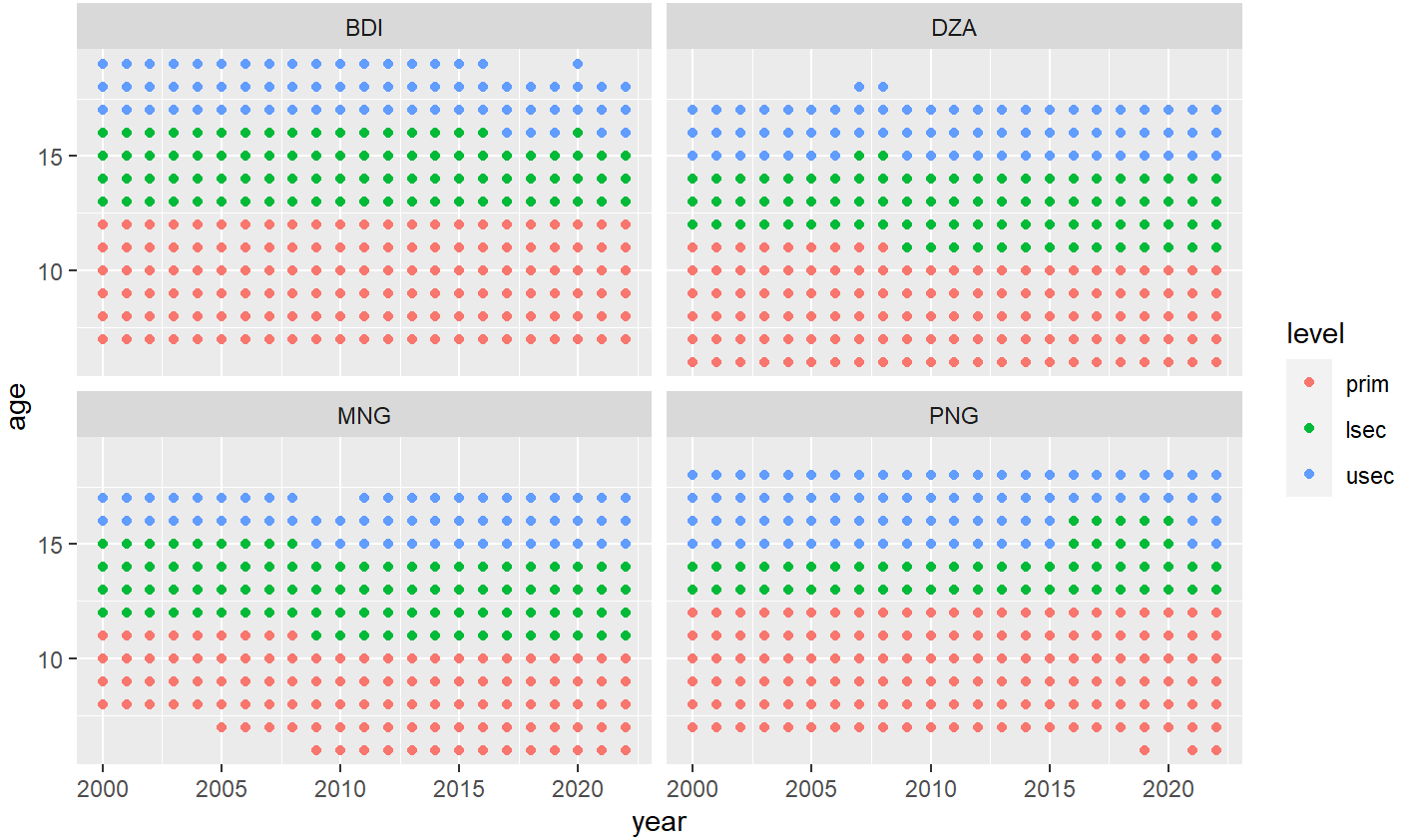
- Enrollment counts and population data are sourced separately resulting in frequent discrepancies.
  - 23.59% of the data are negative due to enrollment counts exceeding population estimates.
  - These points are present in 165 countries out of 190 total.
  - How can we do better going forward?
- Some observations in the UIS database seem to be sourced as rates directly from the countries. For example, Singapore has a 0.1% OOS rate in the database but that is not consistent with the enrollment counts/WPP data. The rates appear more reliable though more clarity on directly sourced rates would be helpful.

Survey Data

- Our approach to survey data does not make any major adjustments to the survey data before computing observations.
- We understand UIS does make some adjustments and would like to talk through them.
- The difference in approach leads to some substantial differences.
    - For example, in the most recent 2019 Zimbabwe MICS survey, we have a 51% upper secondary OOS whereas UIS has 72%. We note that the adminstrative data is consistently ~42%.

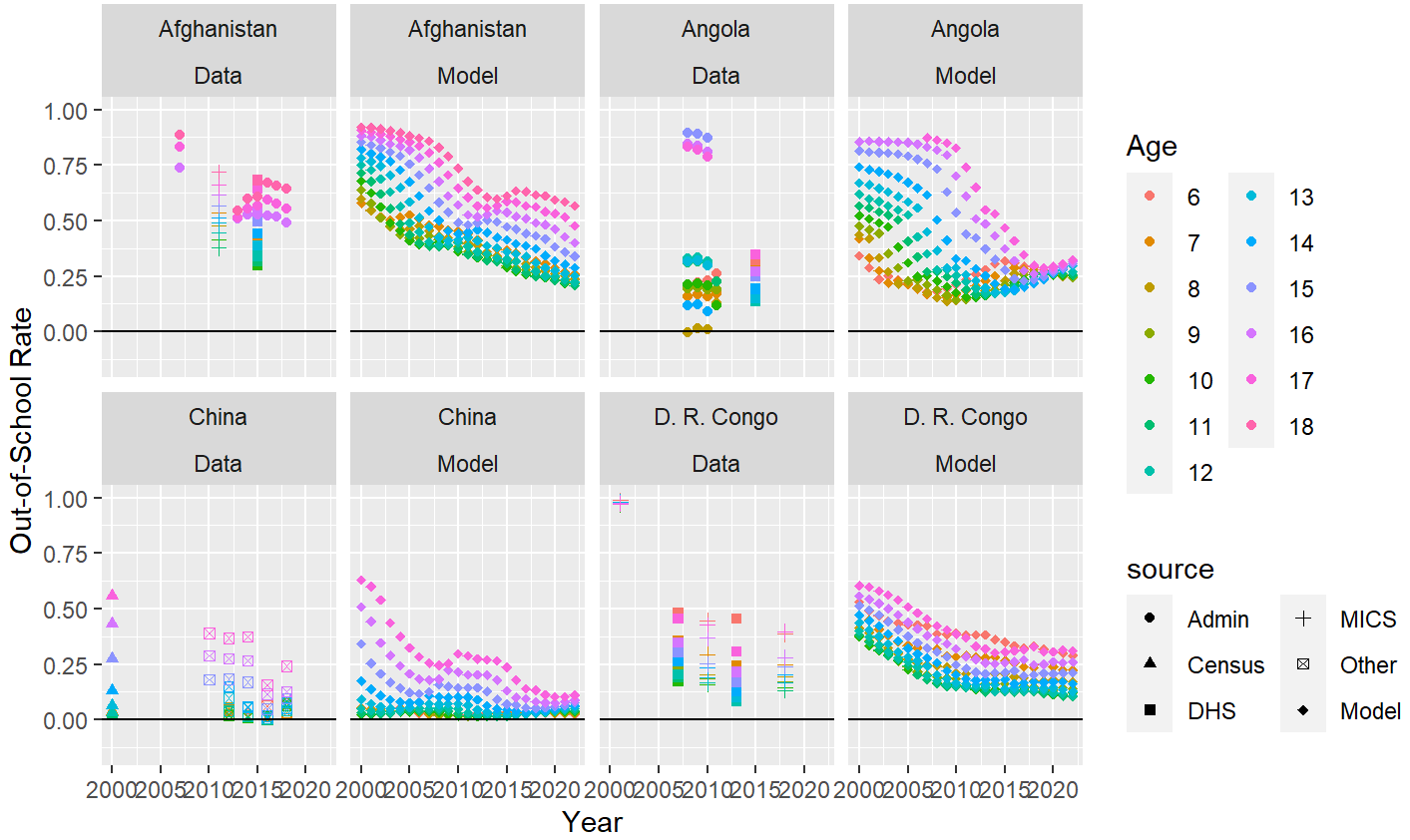
Theoretical School Ages

Occasionally, countries have frequent changes in school ages causing jittery aggregate results. Refreshed data from UIS smoothed most but not all cases.

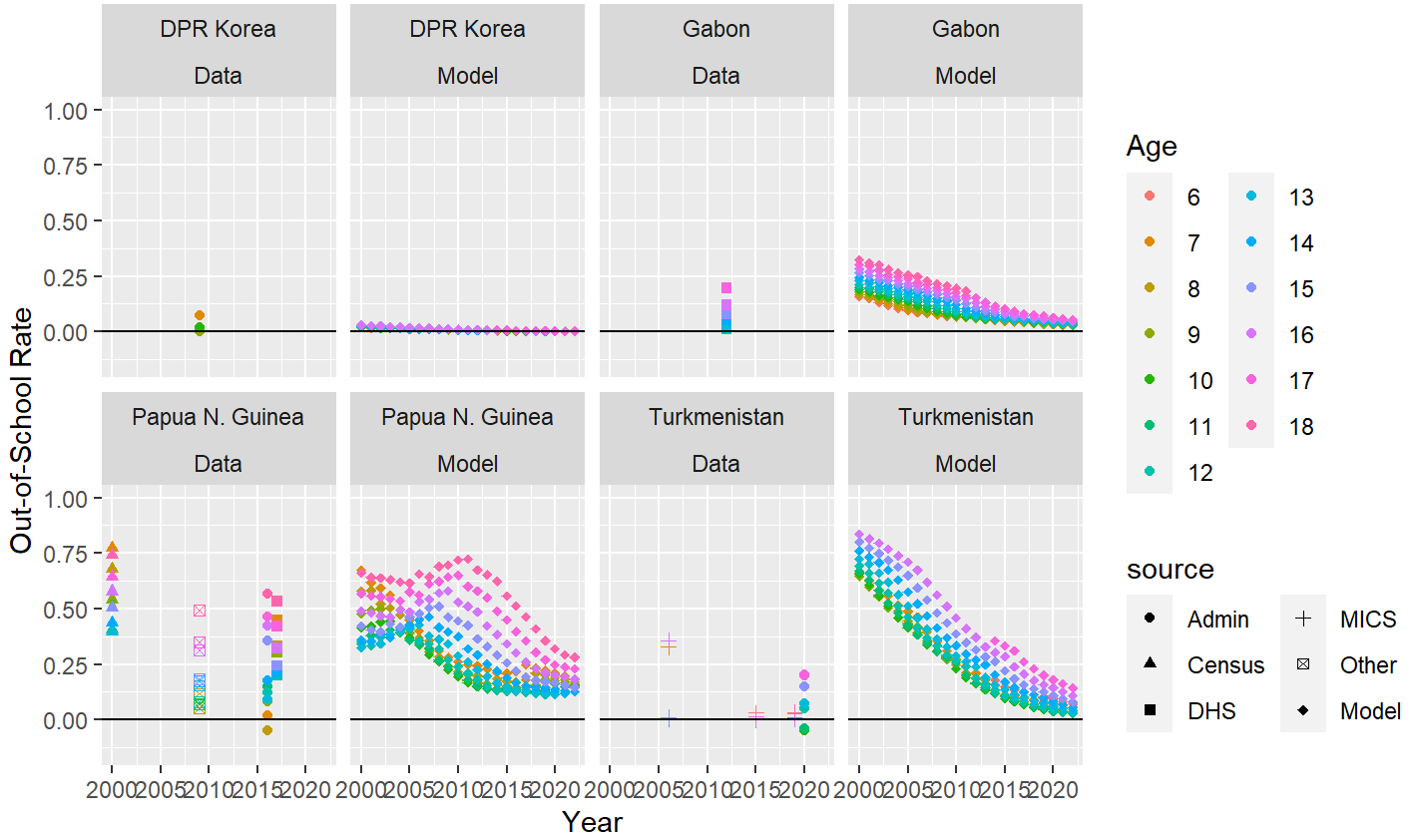


Theoretical School Ages - Algeria

High Priority Countries



High Priority Countries



Dissemination

Dissemination

Draft visualizations will be available at: https://education-estimates.org/oos-apps/

This link is not live yet but can be enabled at any time while remaining inacessible from the VIEW website.

Our intention is to prepare associated documentation describing the modelling process and guides to faciliate interpreting visualizations. This breaks down into two parts:

- Technical paper targeted at a statistics audience similar to the completion paper
- VIEW documentation for each set of visualizations

Discussion