Bremen, January 2021

Helen Seitzer
Michael Windzio

Development of Historical Enrolment Trends

WeSIS — Technical papers
No. 6

Global Dynamics of Social Policy CRC 1342

Gefördert durch
DFG Deutsche Forschungsgemeinschaft

Helen Seitzer, Michael Windzio
Development of Historical Enrolment Trends
SFB 1342 Technical Paper Series, 6
Bremen: SFB 1342, 2021

in alphabetical order
Helen Seitzer https://orcid.org/0000-0003-1918-6637
Michael Windzio https://orcid.org/0000-0002-7681-5893

University of Bremen

SFB 1342 Globale Entwicklungsdynamiken von Sozialpolitik / CRC 1342 Global Dynamics of Social Policy

Postadresse / Postaddress:
Postfach 33 04 40, D - 28334 Bremen

Website:
https://www.socialpolicydynamics.de

[DOI https://doi.org/10.26092/elib/2085]
[ISSN 2700-0389]

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) Projektnummer 374666841 – SFB 1342

Helen Seitzer
Michael Windzio

Development of Historical Enrolment Trends

SFB 1342
No. 6

Development of Historical Enrolment Trends

Helen Seitzer and Michael Windzio

Index

Introduction

Ensuring universal access to schooling can be considered a crucial part of a welfare state's responsibilities (Ramirez & Boli, 1987). In fact, universal access to formalized education is often defined as an important goal in establishing a modern welfare state in the first place, and multiple organizations including International Organizations (IOs) are working on ensuring that goal by declaring 'universal primary education' as an important target for agreements like the Millennium Development Goals (MDGs). Education is similarly credited to be a crucial determinant of economic development (Hanushek & Woessmann, 2011) and is considered an important tool for providing citizens with equal opportunities and for promoting equality and social mobility (Deary et al., 2005). Enrolment in state-organized formalized schooling is therefore often used to measure a country's industrial and social standards (Benavot & Riddle, 1988). Consequently, enrolment rates are a recurring measure in comparative studies on welfare states, focusing especially on the development of welfare policy and compliance. For these types of time-series analyses it is crucial to have complete measures for many countries over a long period of time. Historical data on enrolment, however, is often either scarce or ambiguous. There are multiple possibilities for calculating enrolment in the first place, for example, by focusing only on non-repeaters or by separating enrolments according to age group, gender, level or type of education, and so forth. In addition, the basis of calculation is not always sufficiently explained, especially in historical records. Furthermore, there are issues regarding existing data reliability: Depending on the source, national comparative data could be collected through population surveys or reported from statistical office of the respective country, causing suspicion over the accuracy of the data. A source for consistent enrolment statistics is UNESCO's Institute of Statistics, as they provide enrolment rates from the 1970s onwards for a multitude of countries. Recent changes in their calculation methods reduced the number of available enrolment measures so that there are currently only enrolment rates calculated by two different methods, as opposed to the previously available four measures (which are included here). This measure for enrolment rates combines multiple historical and recent datasets on enrolment into one consistent trendline for up to 192 countries from 1820 to 2018. The strength of this trendline is not the historical accuracy of every single countries' statistics, but the overall trend, completeness and consistency. In contrast to other datasets focusing on the accuracy of their numbers, this dataset includes correction factors to obtain a consistent trendline representing a probable, complete, trend and rise of enrolment over time.

Previous efforts to consolidate enrolment rates for longer time periods were conducted by Lee and Lee (2016). This dataset combines multiple enrolment measures and previously collected data on schooling to provide a more accurate measure of historical enrolment rates, similar to our approach. However, this dataset only contains 110 countries and provides data every five years. The approach is much more focused on the historical accuracy of the enrolment trend than its completeness. Another collection of historical enrolment rates is Benavot and Riddle's (1988) dataset for primary enrolment rates. This dataset is also only available for a limited set of countries. Other collections of historical enrolment are embedded in encyclopedias such as Mitchell's International Historical Statistics (Mitchell, 2007a, 2007b, 2007c). Most historical sources are mostly included in Lee and Lee's dataset and were therefore not further considered in this calculation. Additionally, measures from statistical institutes like the World Bank or Eurydice for European countries are mostly based on UNESCO's datasets (UIS, 2019) and were therefore also not considered any further. UNE-

SCO's data is deemed the most accurate, as the countries report these numbers themselves; however, only UNESCO's processed data, not the raw data, is available to the public.

Most previous datasets are mainly concerned with the historical accuracy of the data provided. For this dataset, however, we chose a slightly different approach than that of previous efforts: Since most historical records on enrolment can be contested somehow, we wanted to provide a more comprehensive estimation of a possible enrolment trend development. This dataset unites multiple sources on enrolment and therefore provides a trendline, describing how enrolment over time most likely evolved without too much focus on the historical accuracy of single datapoints.

Our estimation includes a total of 219 countries and provides data on enrolment from 1820 to 2018. The significance of this contribution therefore lies not in its historical representativeness but in its comprehensiveness. The strength of the trend as presented here allows for the analysis of school enrolment, including its influence on other welfare policies and vice versa, over a longer time span.

Data

Five data sources were included into this trendline, four of which were available from the UNESCO Institute of Statistics as of September 2020. Three measures from UNESCO were included in the dataset: 'net enrolment rates', 'adjusted net enrolment rates', and current 'total enrolment rates'. Also included were Lee and Lee's dataset as well as Benavot and Riddle's dataset. All these measures defined enrolment differently. While gross enrolment ratios are the standard in most publications, they are disaggregated by level of education but not by age and therefore include all over and under school-age children, skewing the distribution depending on the number of children repeating classes. Lee and Lee define their data as "…adjusted ratios are a modification of the gross enrolment ratios in that they account for the repetition of grades in primary and secondary schools" (Lee & Lee, 2016; 148), while gross enrolment ratios "[are] constructed by dividing the number of all persons enrolled in a given level of schooling by the population of the age group that should be enrolled at that level according to national regulations or customs" (Lee & Lee, 2016; 148).

The UNESCO statistics are calculated slightly differently: "Adjusted net enrolment is the number of pupils of the school-age group for primary education, enrolled either in *primary or secondary* education, expressed as the percentage of the total population in that age group". Total net enrolment in contrast is "[The] total number of students of the official age group for a given level of education who are enrolled in *any level* of education, expressed as a percentage of the corresponding population". Net enrolment rate on the other hand is "the ratio of children of official school age who are enrolled in school to the population of the corresponding official school age". Benavot and Riddle's measures are described as "unadjusted enrolment ratio is found by dividing enrolments at each level by the population of a constant school-age category for all countries" (Benavot & Riddle, 1988; 195). While all these measures sound very similar at first glance, they either include different age groups, do not distinguish between primary and secondary schooling, or do distinguish between the levels of schooling while ignoring age groups. The measures are all highly correlated though and are therefore predictive of the resulting combination trend.

Methods

All datasets were treated similarly before the main estimation procedure. Missing datapoints had to be interpolated where possible. Lee and Lee's dataset, for example, has only datapoints for every five years and so its gaps had to be filled. Benavot and Riddle's dataset is only available for primary education, boys and girls combined, and was therefore only used for this specific series. Data from UNESCO is not available consistently, so if the original series for one country only contained five datapoints or less, the series was excluded.

If there was no data available in 1820, the data for each respective country was set to 0.001 at that time to ensure the interpolation would not go below zero.

To ensure the trendline would not be linear, the interpolation was performed with a logit transformation. First, to ensure values within a 100% range, the original enrolment values were divided by 100. The fractions π were then transformed onto the logit scale.

$$y = \ln\left(\frac{\pi}{\pi - 1}\right)$$

These values were then interpolated linearly to complete the time span from 1820 to 2018 and transformed back to their original scale with the inverse logit to get the share or probability p.

$$P = \frac{e^y}{1 + e^y}$$

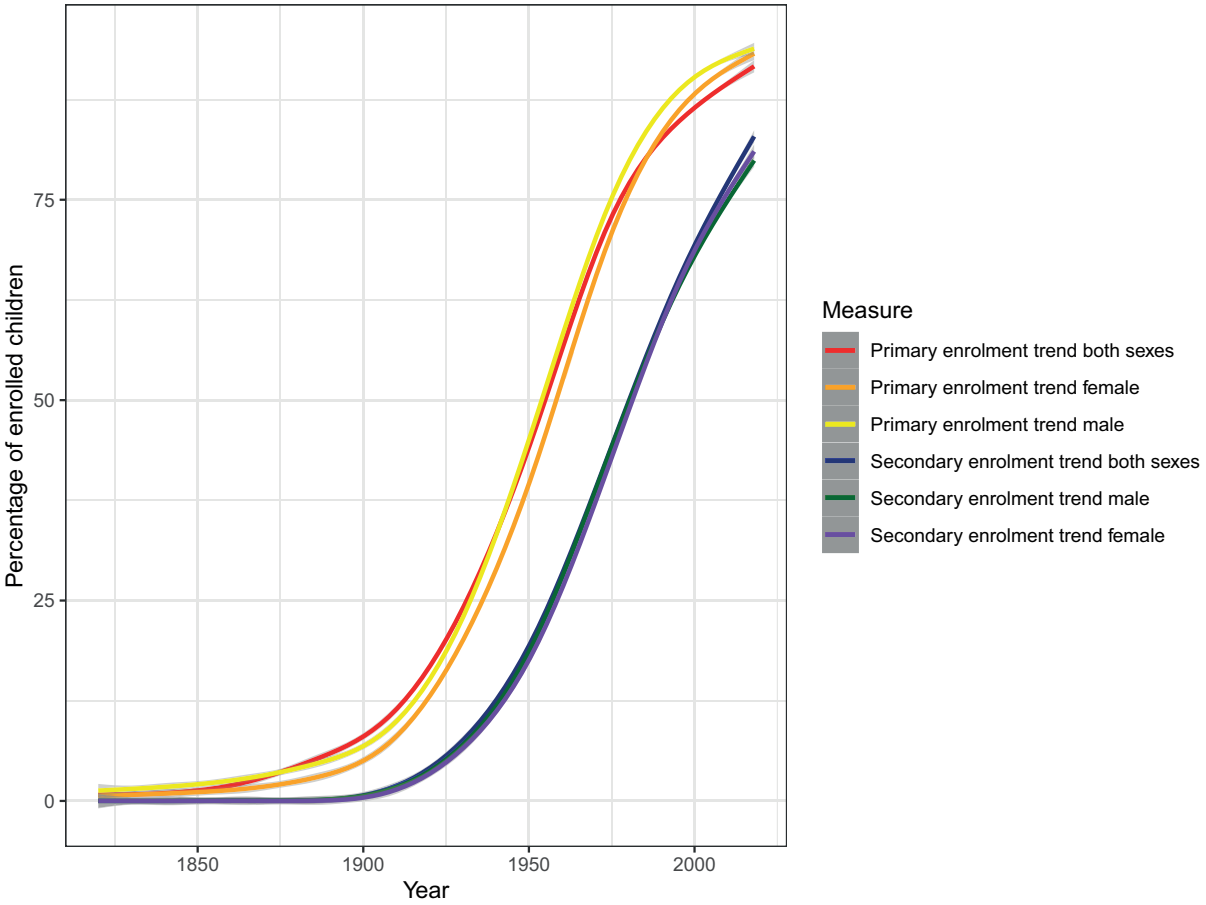
This ensures that the values are kept within range. The values were then multiplied by 100 to restore the original scale. This transformation results in a logistic growth curve of the data instead of a linear or growth curve. As demonstrated in Figures 2.1 to 2.6, the curve is also of a similar shape to previously estimated growth curves for enrolment trends.

The last but most crucial step of the analysis consisted of performing a Principal Component Analysis (PCA) and adding the resulting value to the mean of all values per indicator, i.e. country and year. Factor analyses are used to reduce multidimensional data to a 'digestible' format by projecting the covariance matrix of the involved variables onto the first few principal components while maximizing their variance. The results are one or more components representing the eigenvectors of the covariance matrix. This allows the combination of all the different values included and reduces them to one or two components representing the variance between the involved variables. Here, we only included the first resulting factor. This results in a correction of the row-wise values by a general factor to be added to the row-wise mean of all values to smooth out any irregularities and discrepancies of the empirical data. This is displayed in the following equation with i = country, t = time, and x being the different variables included.

$$enrolment\ trend_{i,t} = PCA_i + mean(x_{1;i,t}, x_{2;i,t}, \dots x_{n;i,t})_{i,t}$$

If one of the previously collected trend values constantly overestimated the empirical values, this is corrected not only through the mean of the empirical values but also though the added factor. The resulting trendlines can be seen in Figure 1.

Figure 1. Enrolment trend separated by gender and level of education



Despite this correction, it was impossible to prevent all abnormalities, since the actual data still contained a great number of irregularities. This can be seen in Figures 2 to 7 in the Appendix, which compare the interpolated values with the produced trendline. A total of six trends were estimated, whereby the trends for each gender were estimated both separately and once combined for both, primary and secondary education. The individual trends for each country disaggregated by gender as well as by education level (primary and secondary) can be found in the Appendix as well. Figures 8 to 13 show all six trends for all countries as well as their means and medians, showing the variance of trends between the countries. These Figures clearly show, that our trendline represents a relatively accurate approximate measure for the empirical data. They also show the variability of the produced trendline over the sample. To better understand this variability, the supplementary material available consists fof individual trendlines per country. To test the accuracy of the produced trendline we correlated the produced trend with the empirical data, which resulted in correlations ranging from .83 to .99. Overall, this is a very high correlation, as expected.

Table 1 shows descriptive statistics for the six outcome trends for all years and countries combined. The shape of the distribution must be taken into account when inspecting these statistics. The growth curve is relatively slow at first, resulting in the relatively low mean values seen in Figure 1. This table also shows, that not all trends are available for all countries due to the lack of data in the included datasets. Figure 14 in the Appendix therefore shows the data availability for the outcome trendline.

Overall, this dataset provides a good measurement for enrollment trends over a large country sample for a long period of time. Due to the unique combination of existing data, this trendline offers a solution to include enrolment measures into historical country-level

analyses. Unlike other measurements we do not claim to provide absolute accurate data, but rather the most probably trend development in enrolment rates.

References

Benavot, A., & Riddle, P. (1988). The Expansion of Primary Education, 1870-1940: Trends and Issues. *Sociology of Education*, *61*(3), 191–210. https://www.jstor.org/stable/2112627

Deary, I. J., Taylor, M. D., Hart, C. L., Wilson, V., Smith, G. D., Blane, D., & Starr, J. M. (2005). Intergenerational social mobility and mid-life status attainment: Influences of childhood intelligence, childhood social factors, and education. *Intelligence*, *33*(5), 455–472. https://doi.org/10.1016/j.intell.2005.06.003

Hanushek, E. A., & Woessmann, L. (2011). The Economics of International Differences in Educational Achievement. In *Handbook of the Economics of Education* (1st ed., Vol. 3). Elsevier B.V. https://doi.org/10.1016/B978-0-444-53429-3.00002-8

Lee, J. W., & Lee, H. (2016). Human capital in the long run. *Journal of Development Economics*, *122*, 147–169. https://doi.org/10.1016/j.jdeveco.2016.05.006

Mitchell, B. R. (2007a). *International historical statistics / B. R. Mitchell: Africa, Asia & Oceania : 1750 - 2005* (5. ed.). Palgrave Macmillan. https://suche.suub.uni-bremen.de/peid=B55121309

Mitchell, B. R. (2007b). *International historical statistics / B. R. Mitchell: Europe : 1750 - 2005* (6. ed.). Palgrave Macmillan. https://suche.suub.uni-bremen.de/peid=B55121362

Mitchell, B. R. (2007c). *International historical statistics / B. R. Mitchell: The Americas : 1750 - 2005* (6. ed.). Palgrave Macmillan. https://suche.suub.uni-bremen.de/peid=B55121343

Ramirez, F. O., & Boli, J. (1987). The Political Construction of Mass Schooling: European Origins and Worldwide Institutionalization. *Sociology of Education*, *60*(1), 2–17. https://doi.org/10.2307/2112615

UNESCO Institute for Statistics (UIS). (2019). *Gross enrolment ratio by level of education*. http://data.uis.unesco.org/index.aspx?queryid=142

Appendix

Table 1. Descriptive Statistics

Statistic	N	Min	Median	Mean	St. Dev.	Max
Primary enrolment trend both sexes	192	0	28.2	40.228	38.424	102
Primary enrolment trend male	194	0	28.0	41.669	39.708	102
Primary enrolment trend female	194	0	21.4	39.146	39.880	102
Secondary enrolment trend both sexes	191	0	6.4	25.950	32.581	102
Secondary enrolment trend male	184	0	6.9	25.794	32.064	102
Secondary enrolment trend female	184	0	5.4	25.517	32.869	102

Figure 2. Primary enrolment trend female and interpolated empirical data

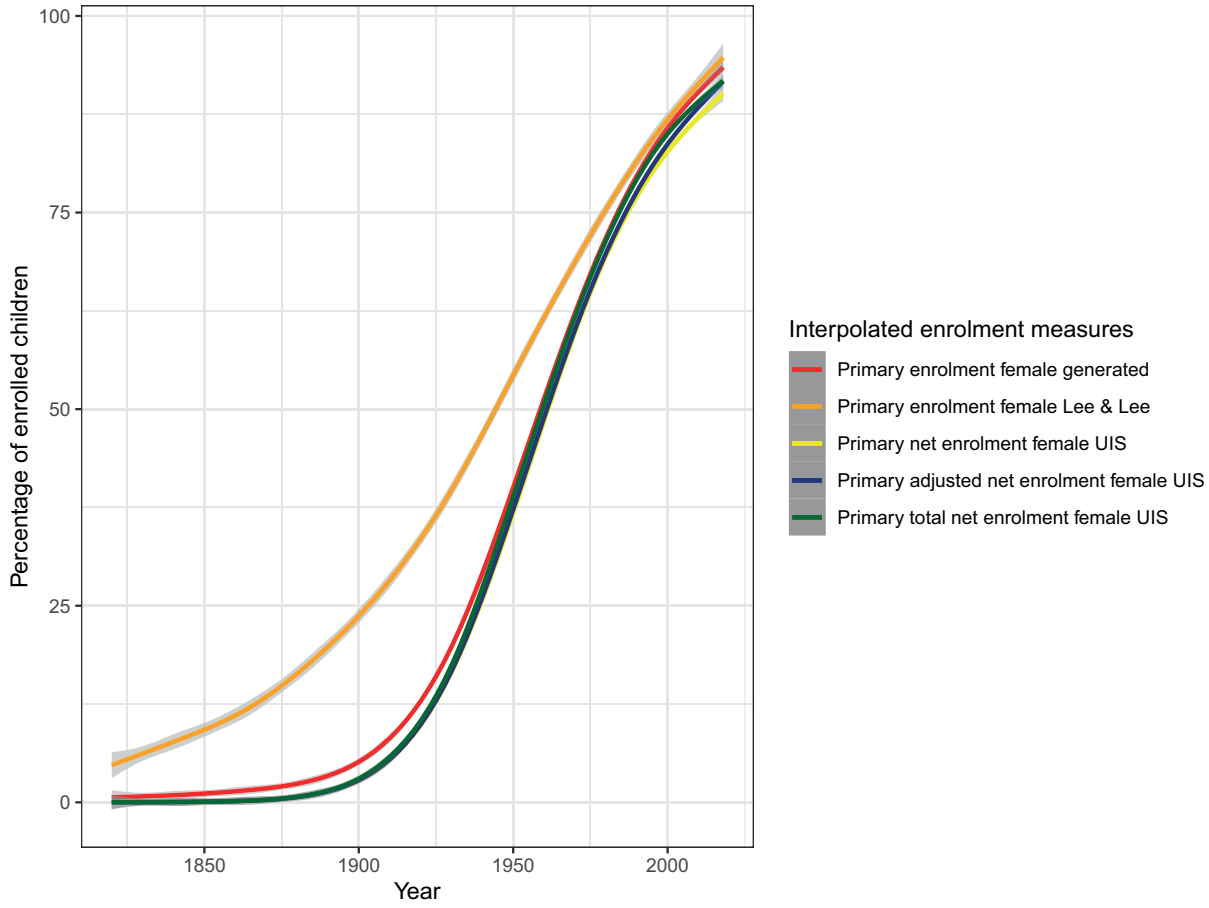


Figure 3. Primary enrolment trend male and interpolated empirical data

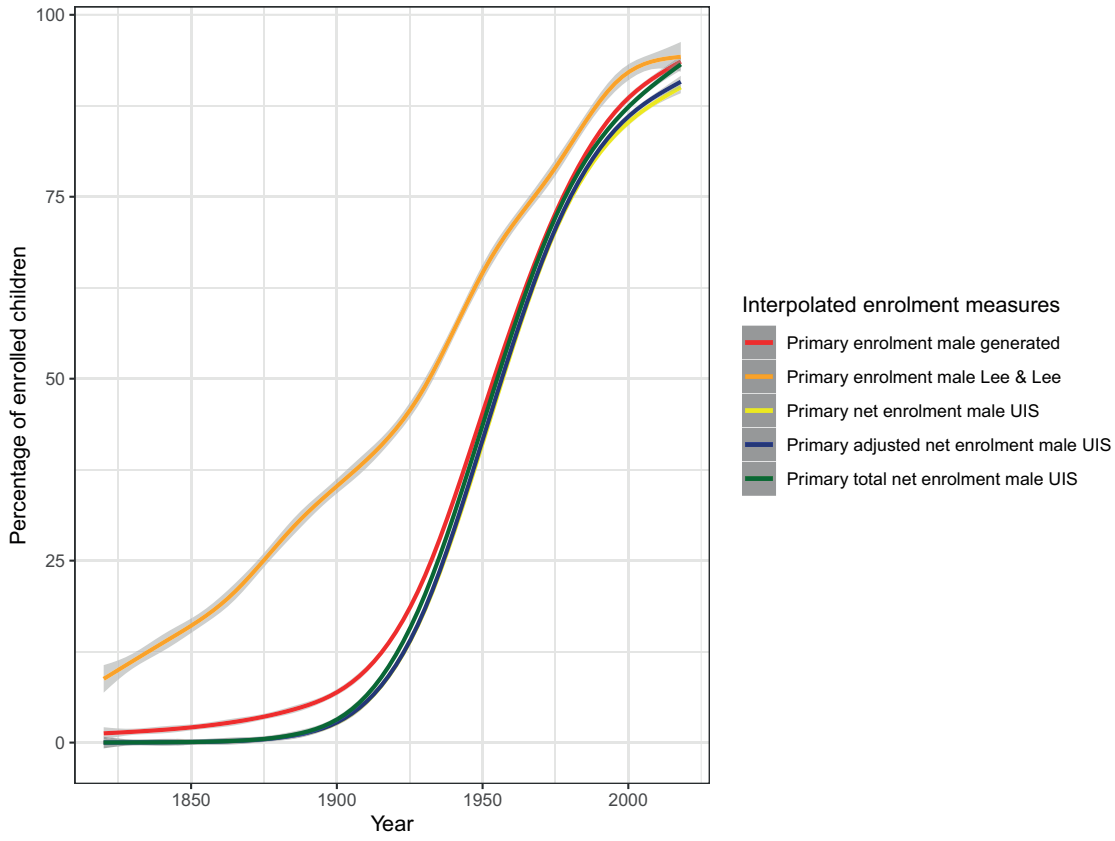


Figure 4. Primary enrolment trend both sexes and interpolated empirical data

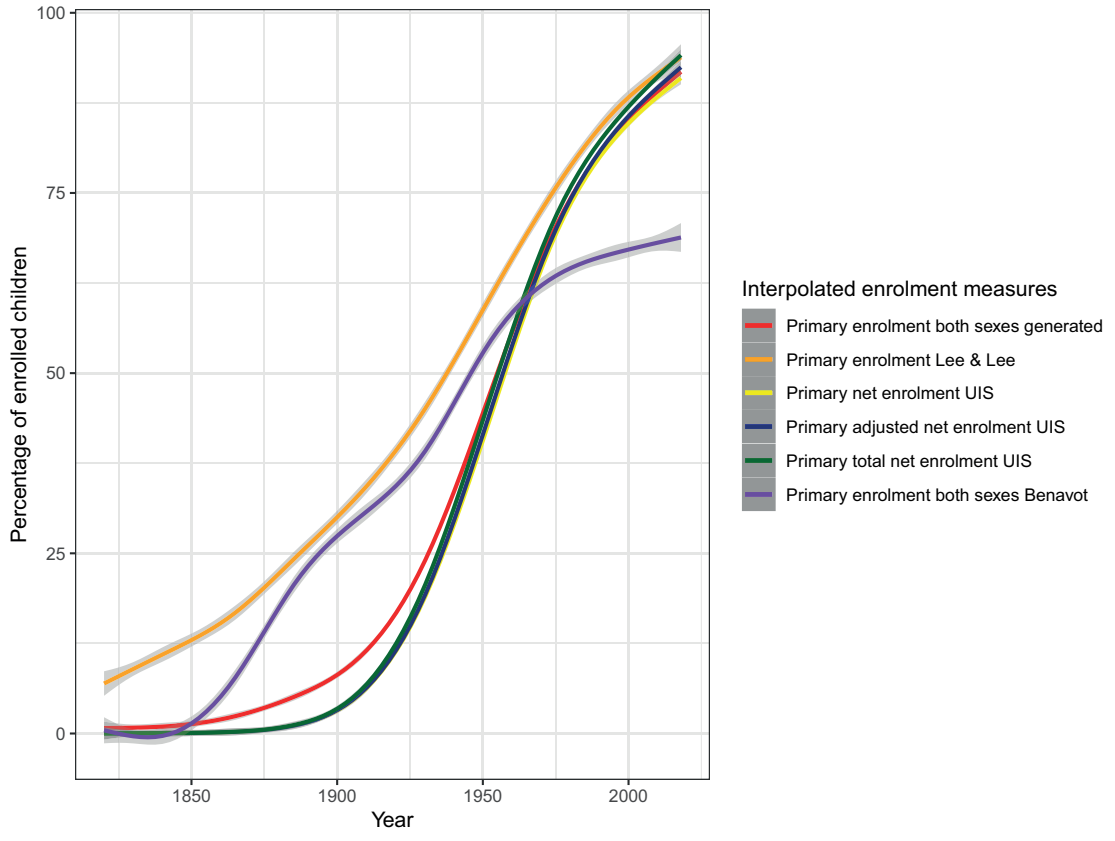


Figure 5. Secondary enrolment trend female and interpolated empirical data

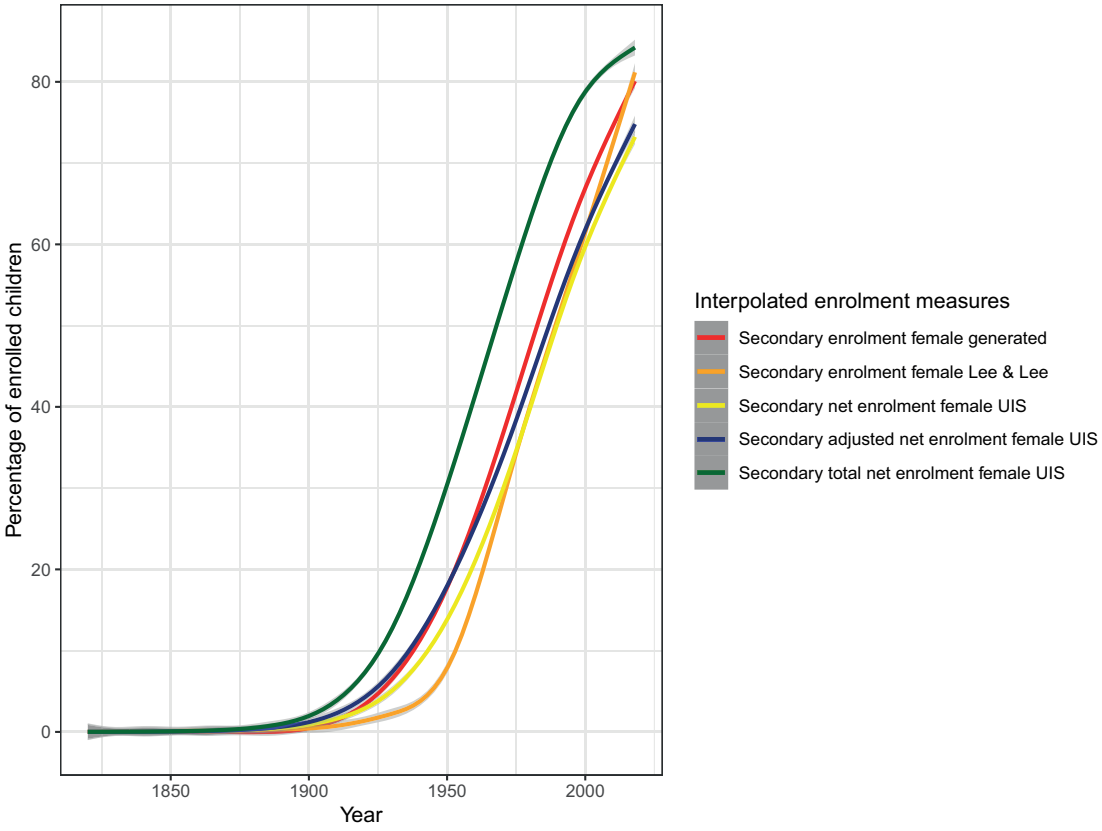


Figure 6. Secondary enrolment trend both sexes and interpolated empirical data

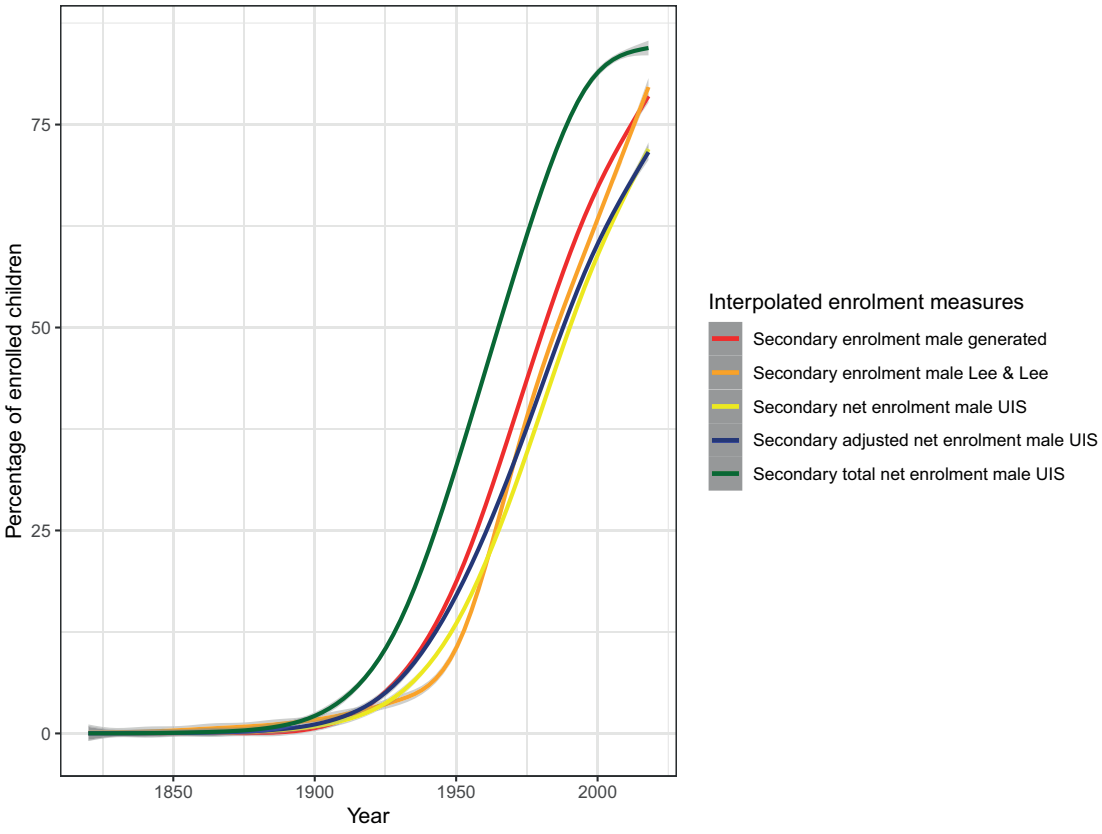


Figure 7. Secondary enrolment trend male and interpolated empirical data

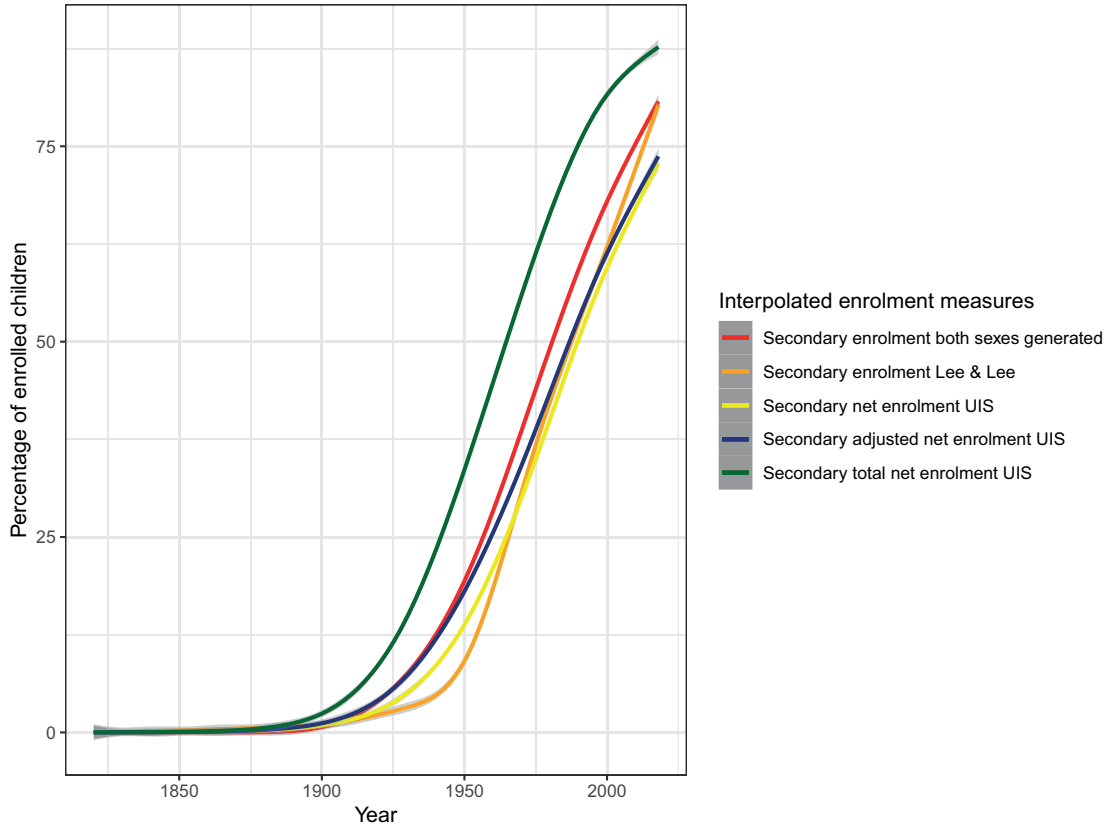


Figure 8. Primary enrolment trend for females per country

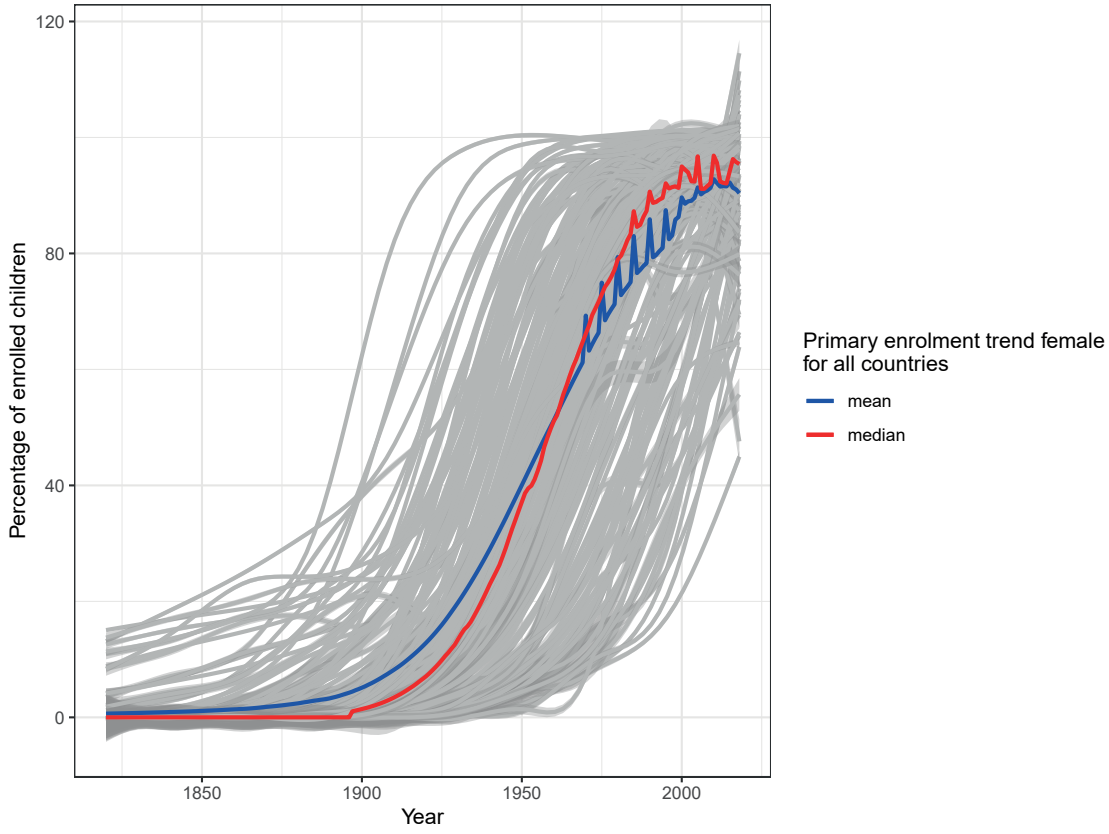


Figure 9. Primary enrolment trend for males per country

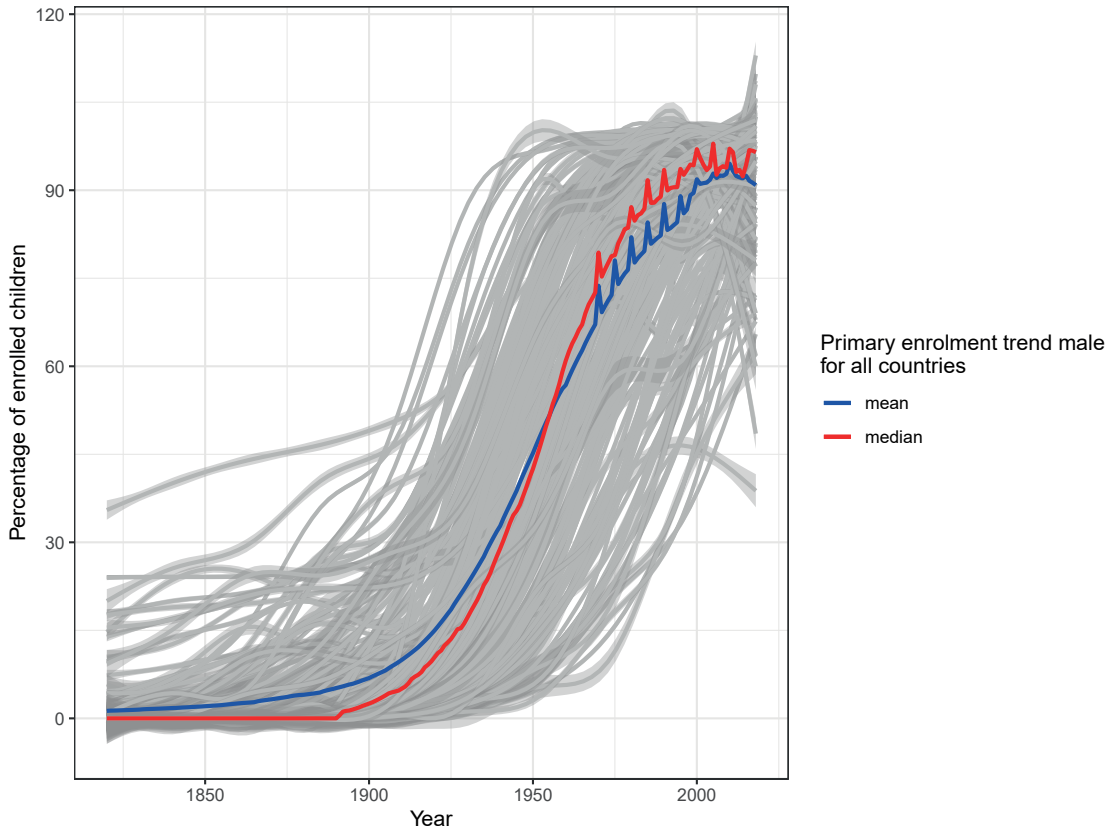


Figure 10. Primary enrolment trend for both genders per country

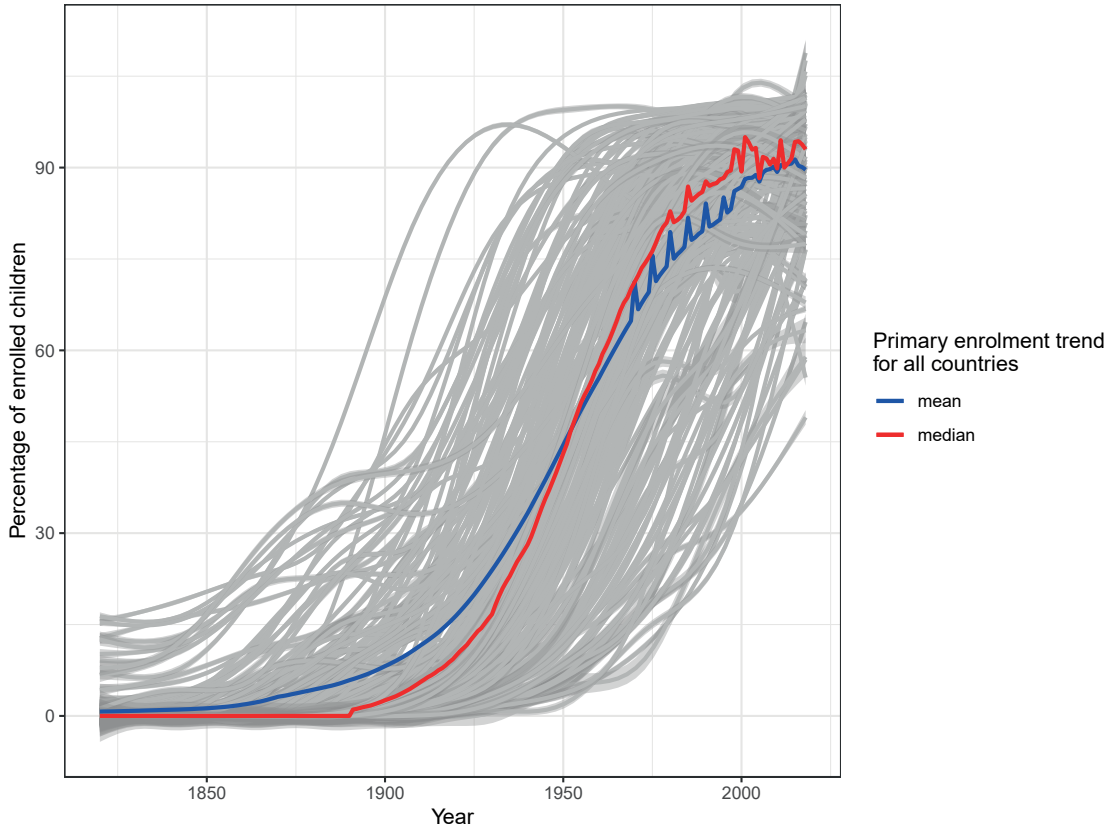


Figure 11. Secondary enrolment trend for females per country

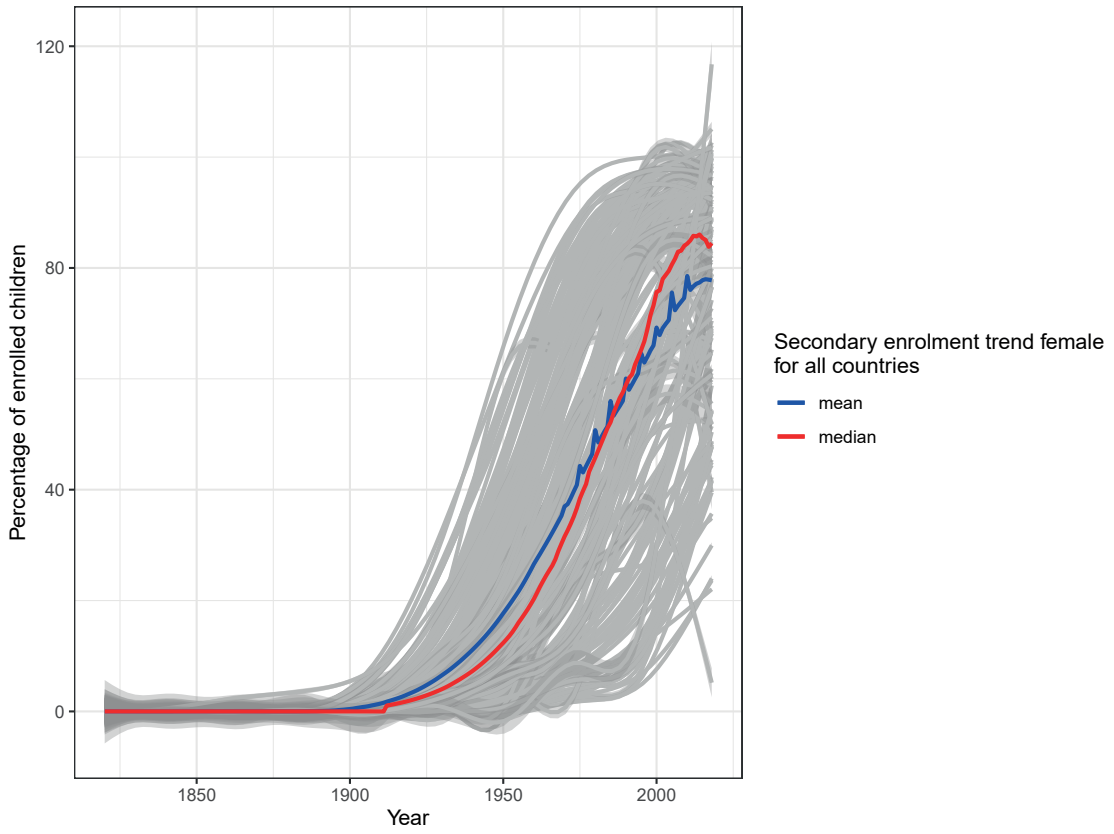


Figure 12. Secondary enrolment trend for males per country

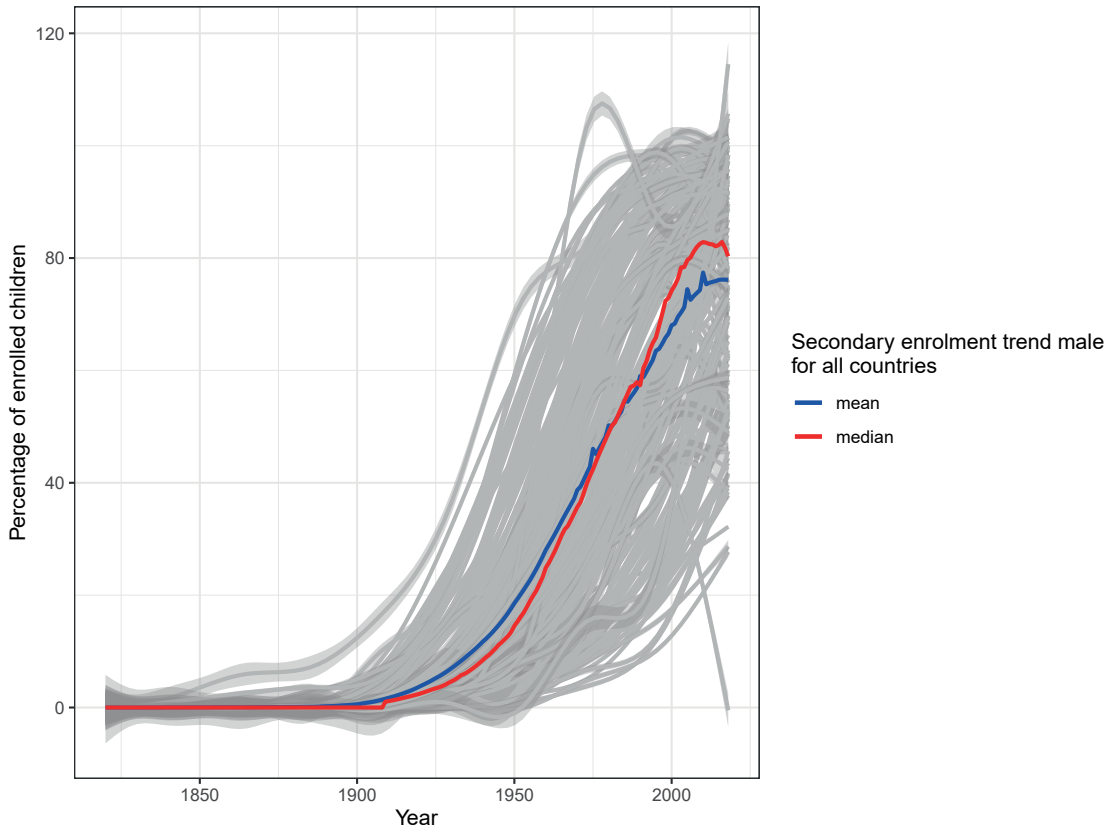


Figure 13. Secondary enrolment trend for both genders per country

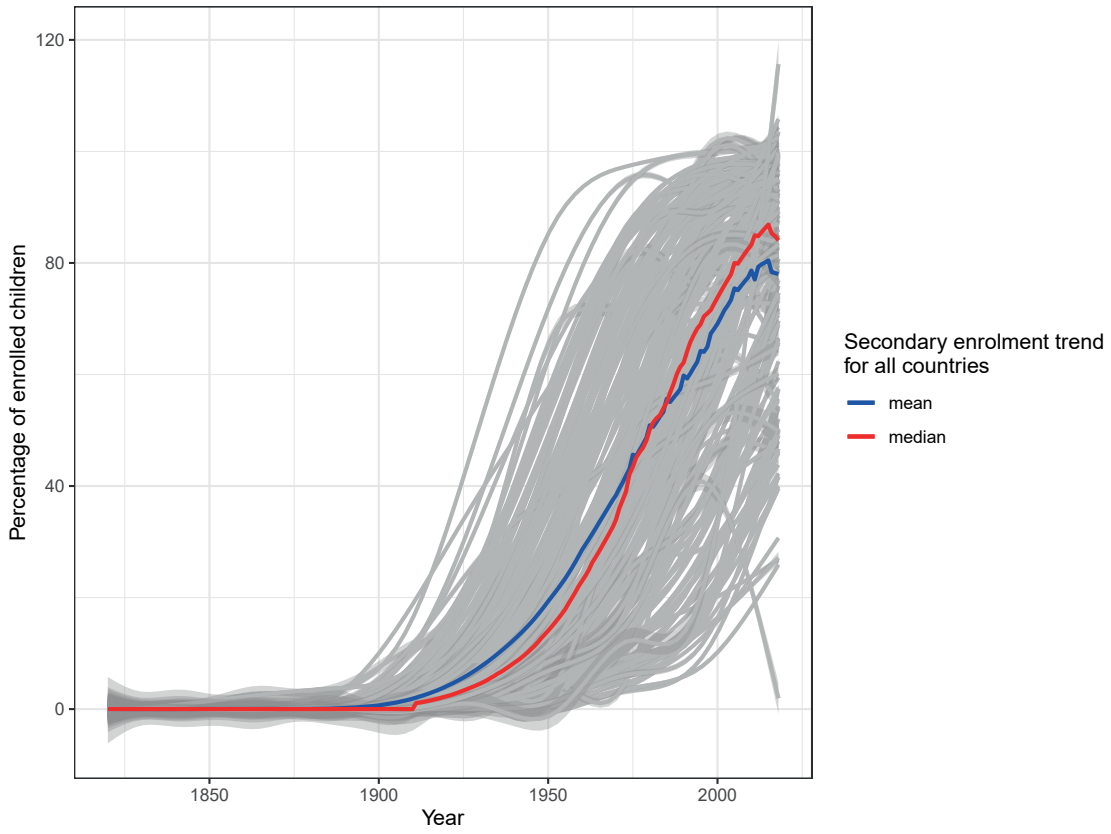


Figure 14. Data Availability

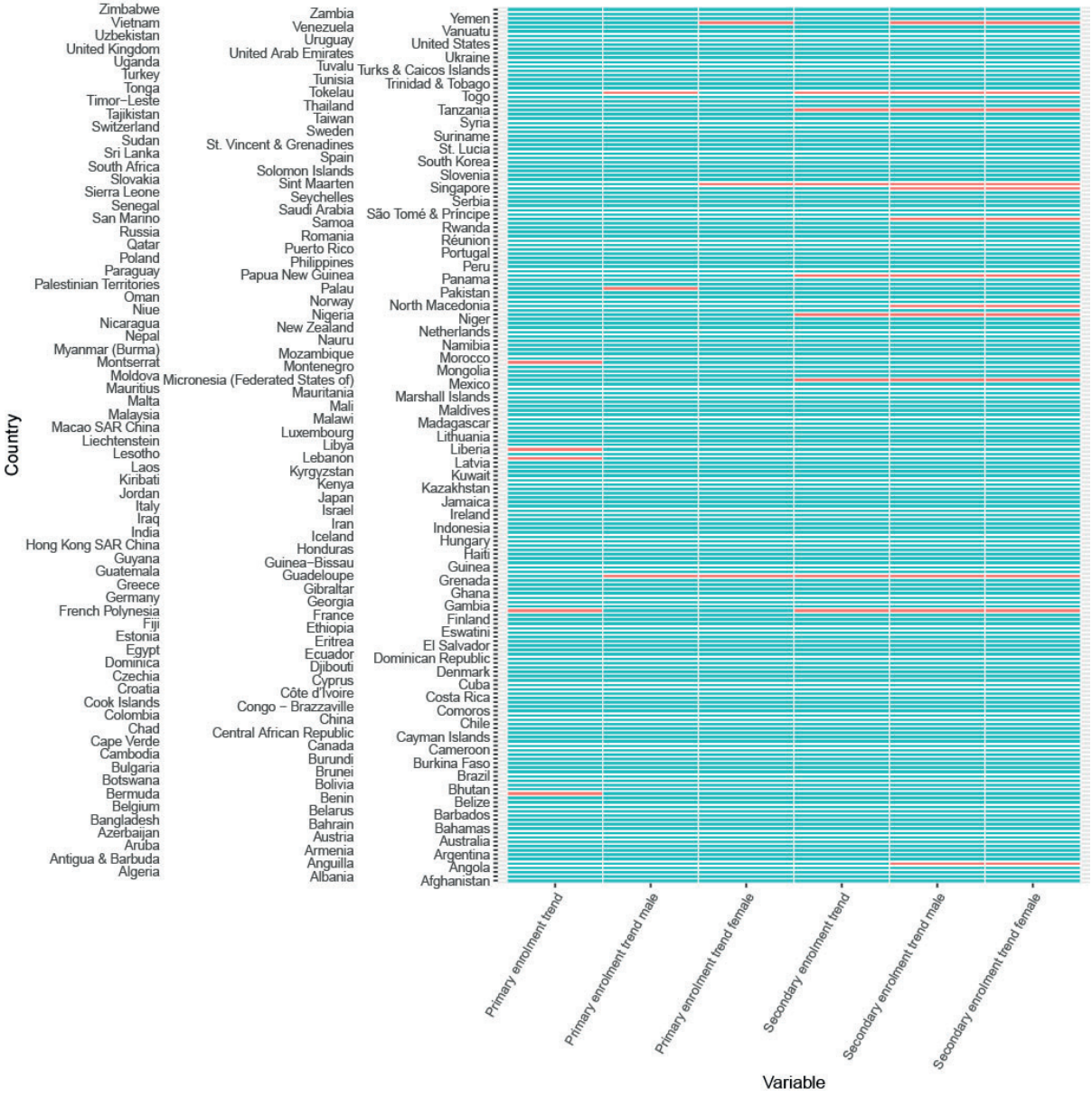
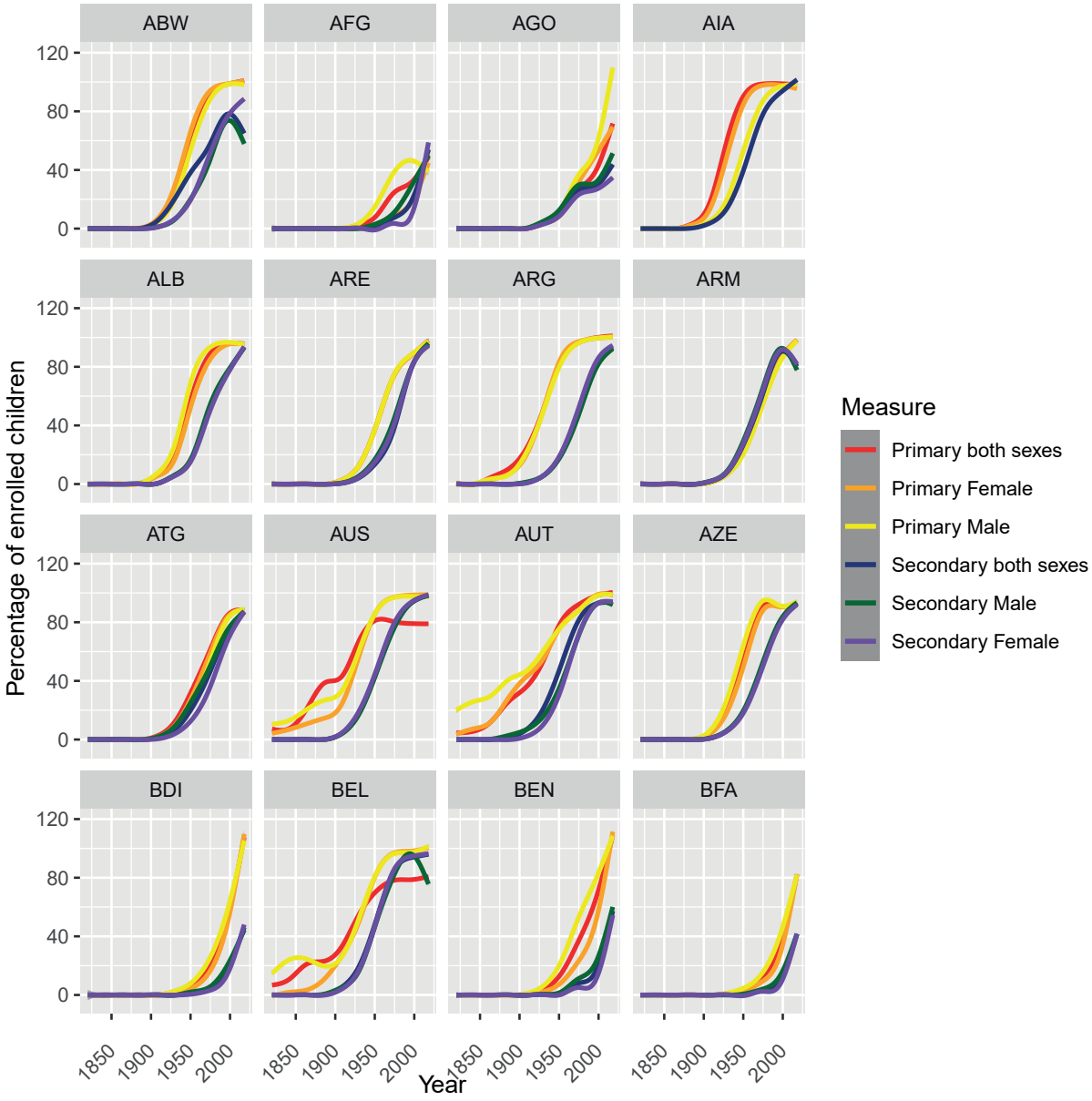
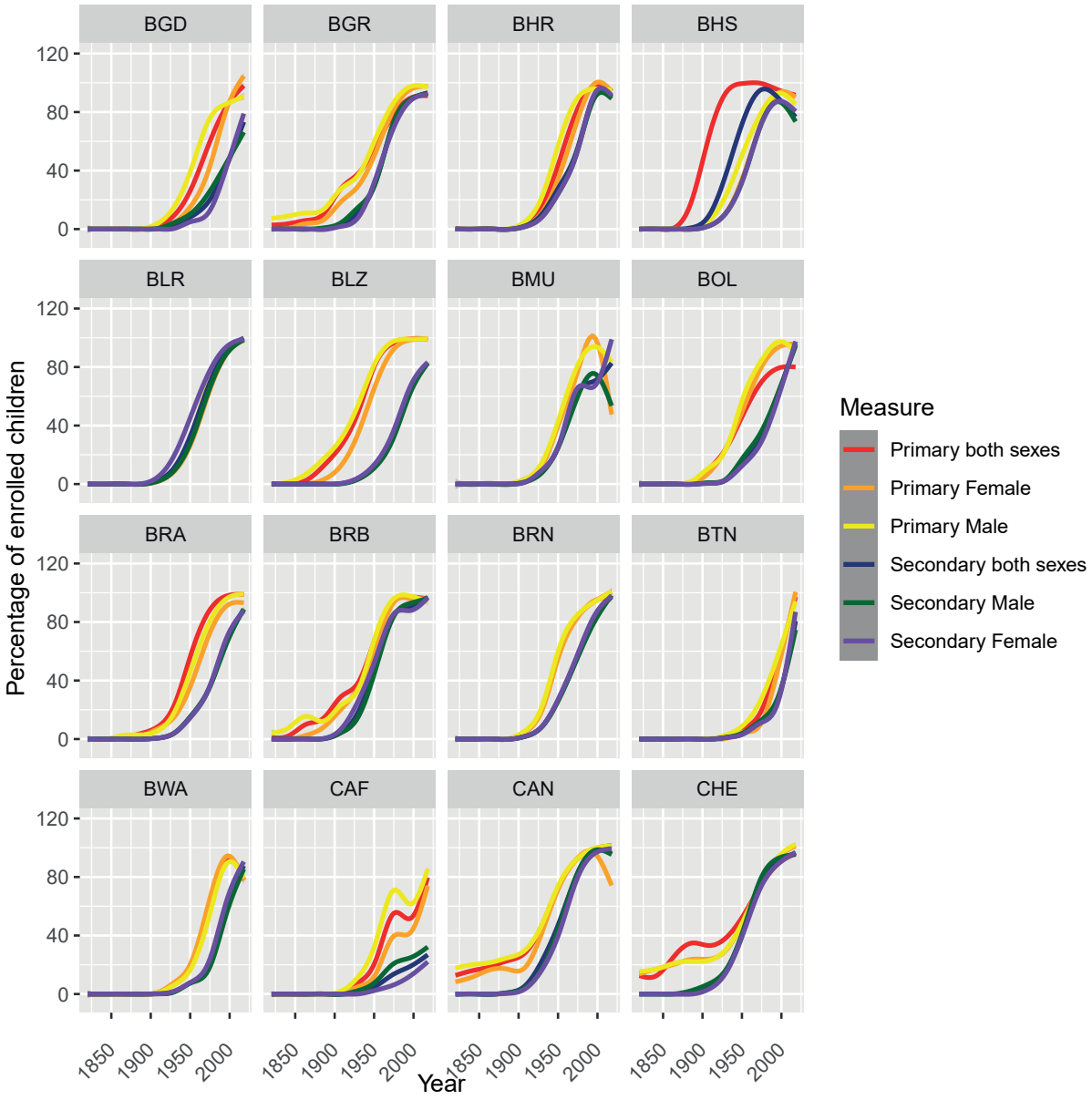
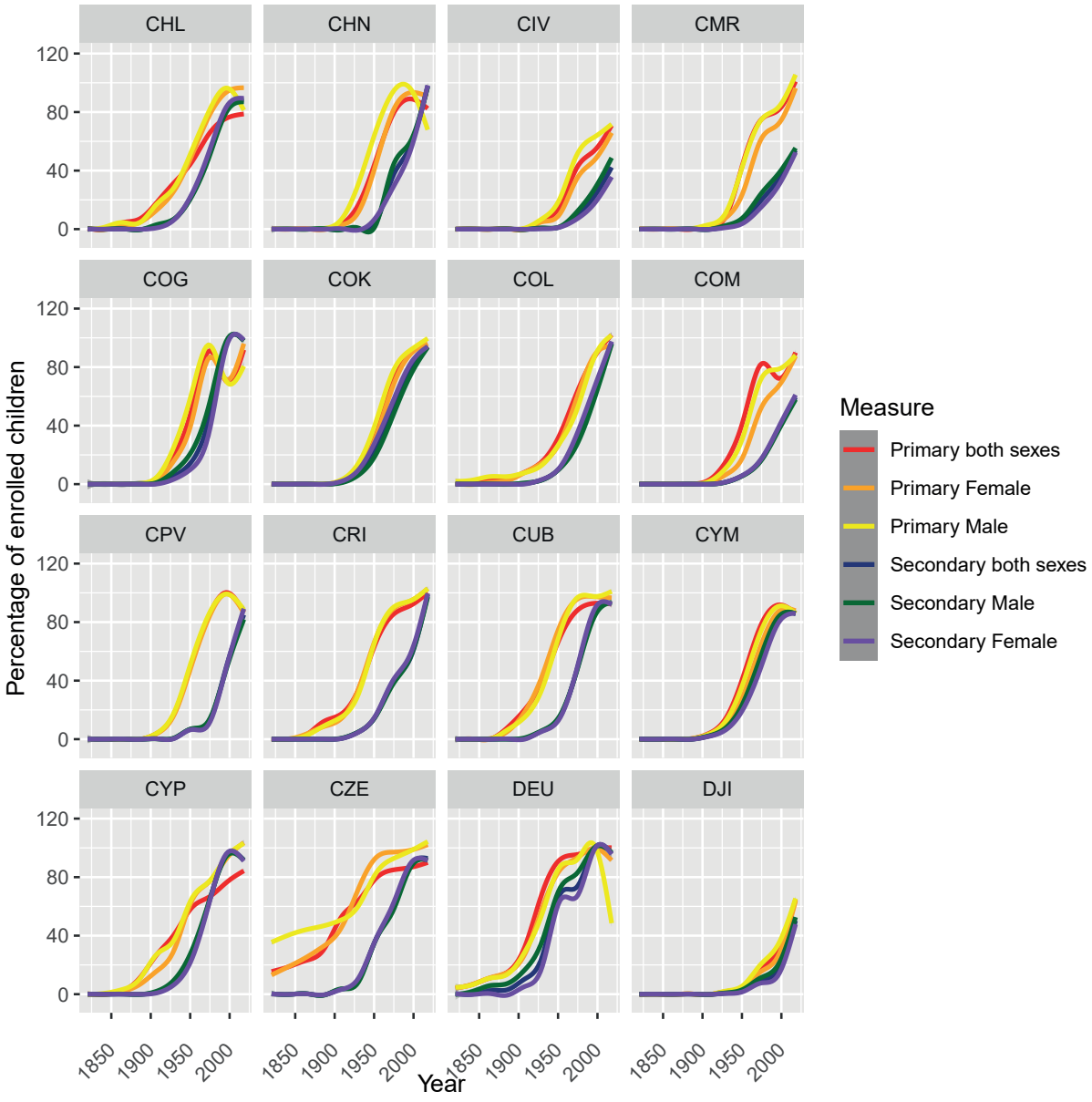
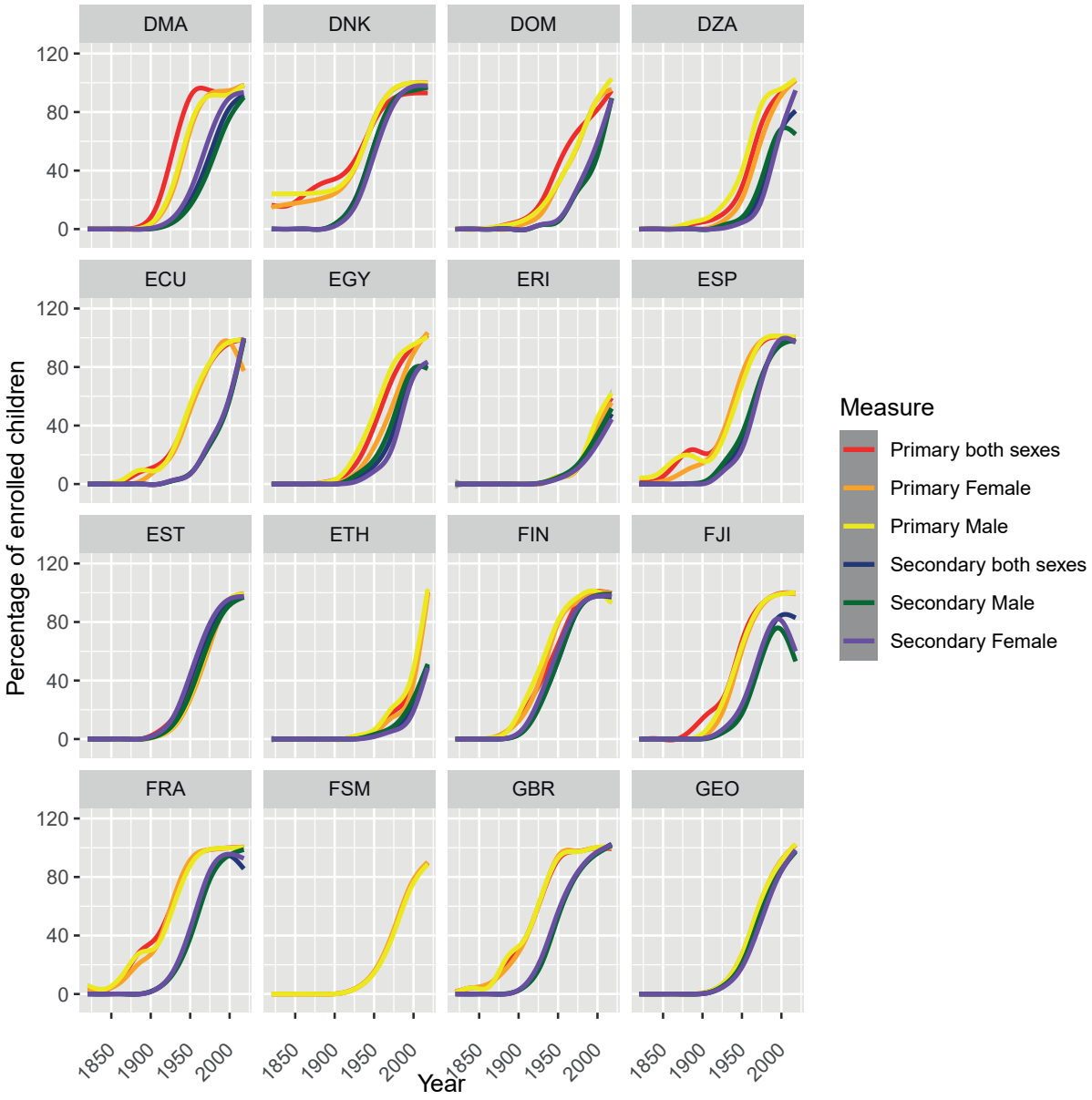


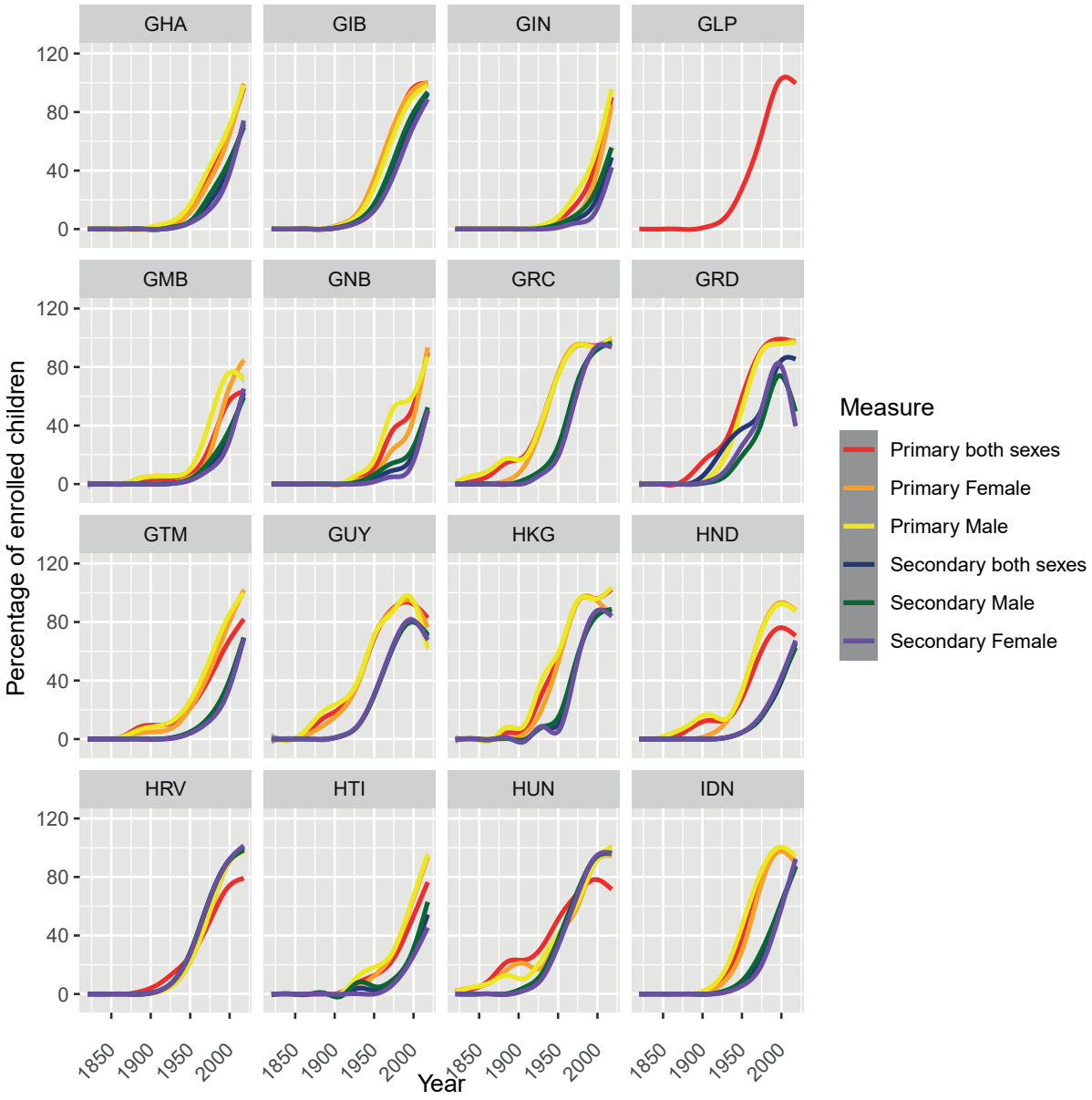
Figure 15. Trends by Country

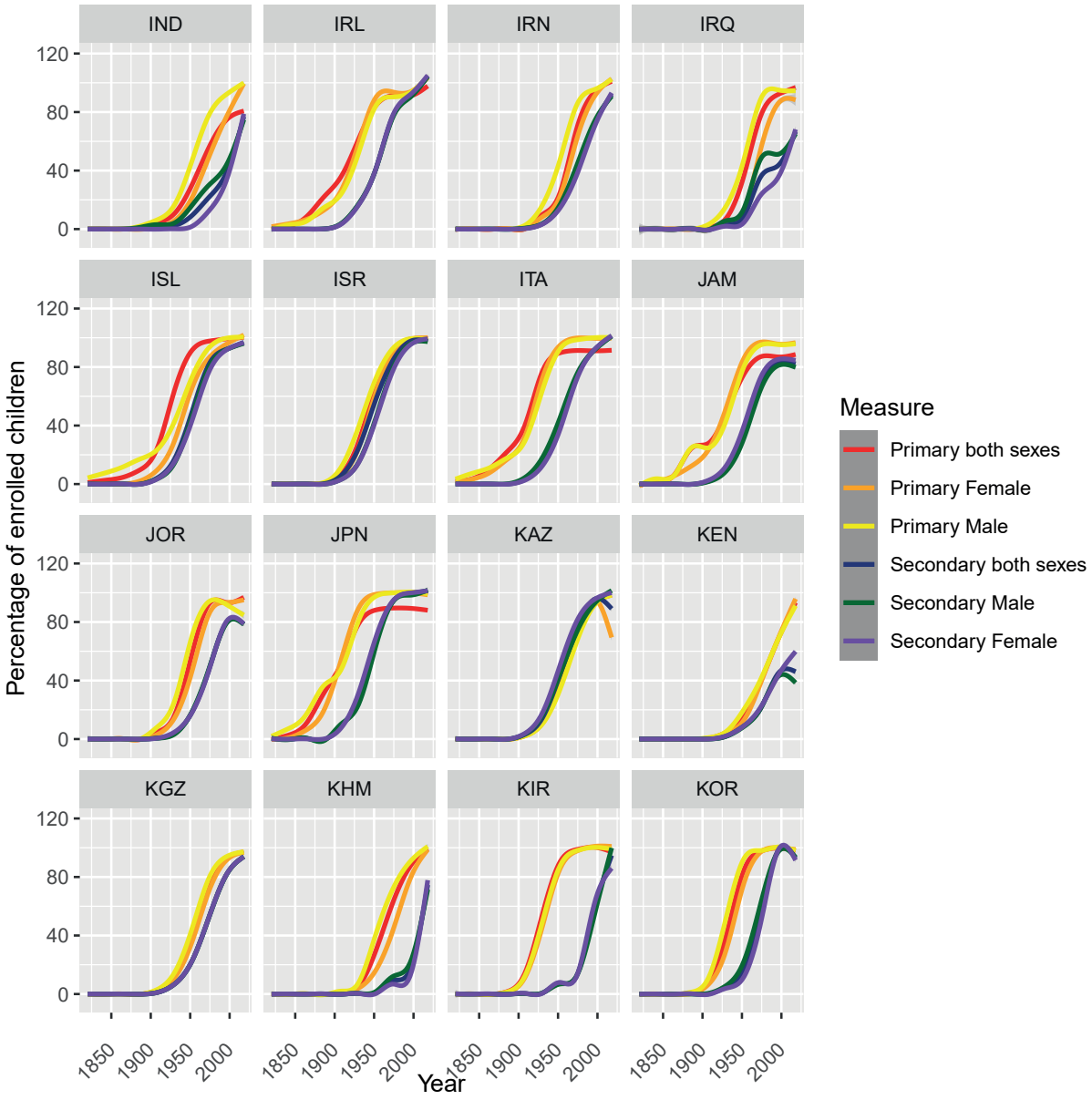


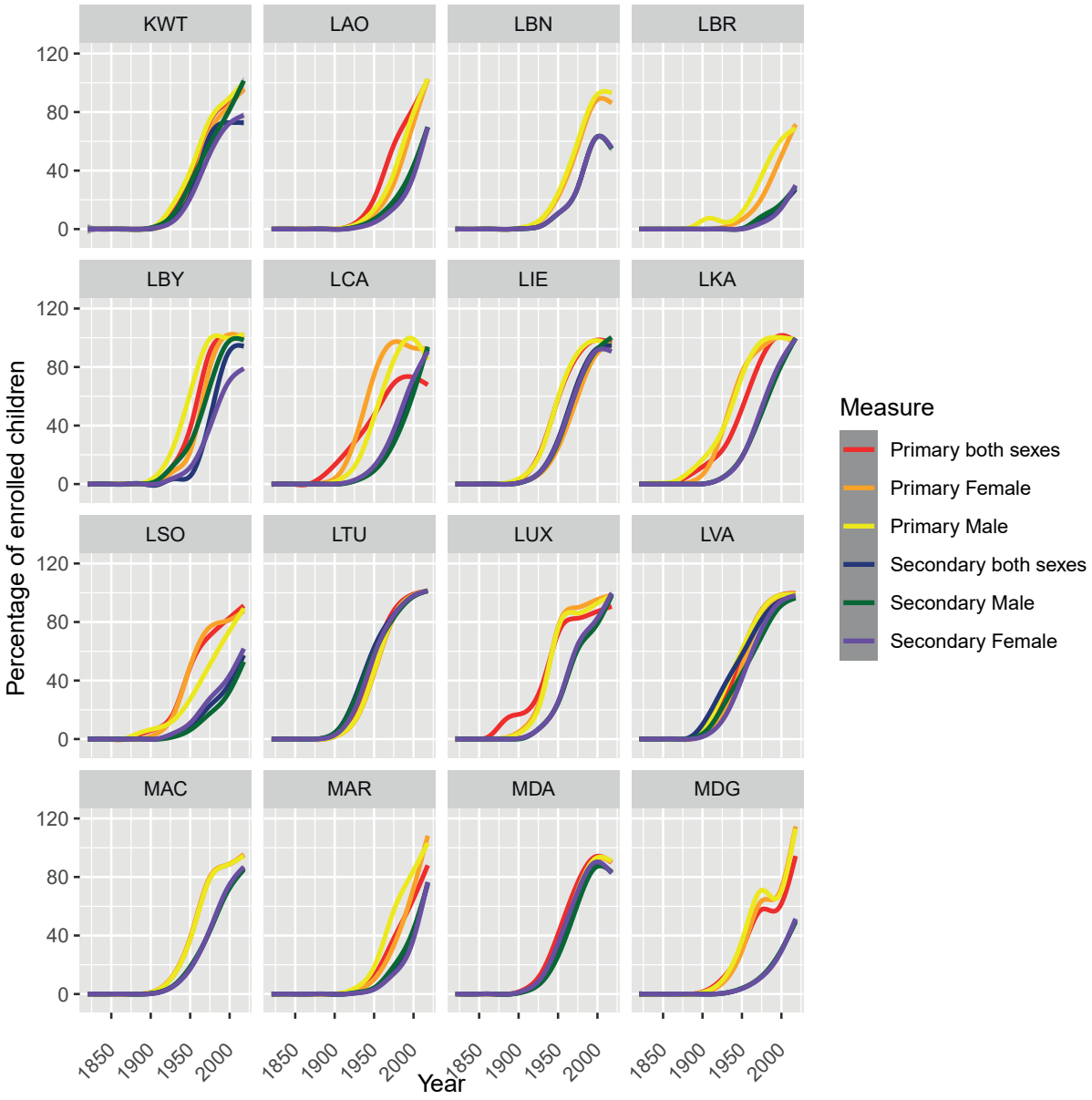

BGD
BGR
BHR
BHS
BLR
BLZ
BMU
BOL
BRA
BRB
BRN
BTN
BWA
CAF
CAN
CHE
120
80
40
0
1850
1900
1950
2000
Percentage of enrolled children
Year
Measure
Primary both sexes
Primary Female
Primary Male
Secondary both sexes
Secondary Male
Secondary Female

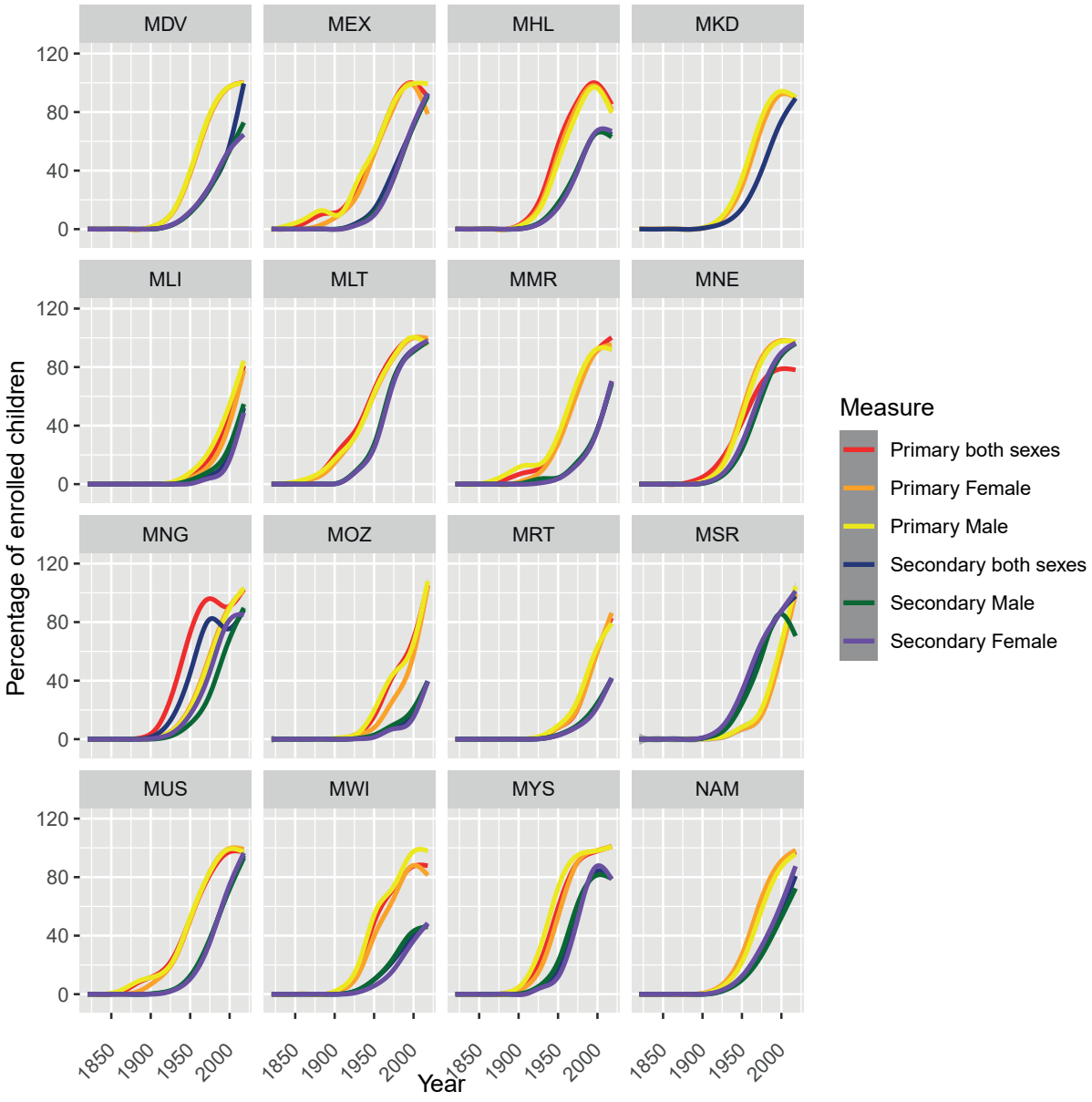

CHL
CHN
CIV
CMR
COG
COK
COL
COM
CPV
CRI
CUB
CYM
CYP
CZE
DEU
DJI
120
80
40
0
1850
1900
1950
2000
Percentage of enrolled children
Year
Measure
Primary both sexes
Primary Female
Primary Male
Secondary both sexes
Secondary Male
Secondary Female

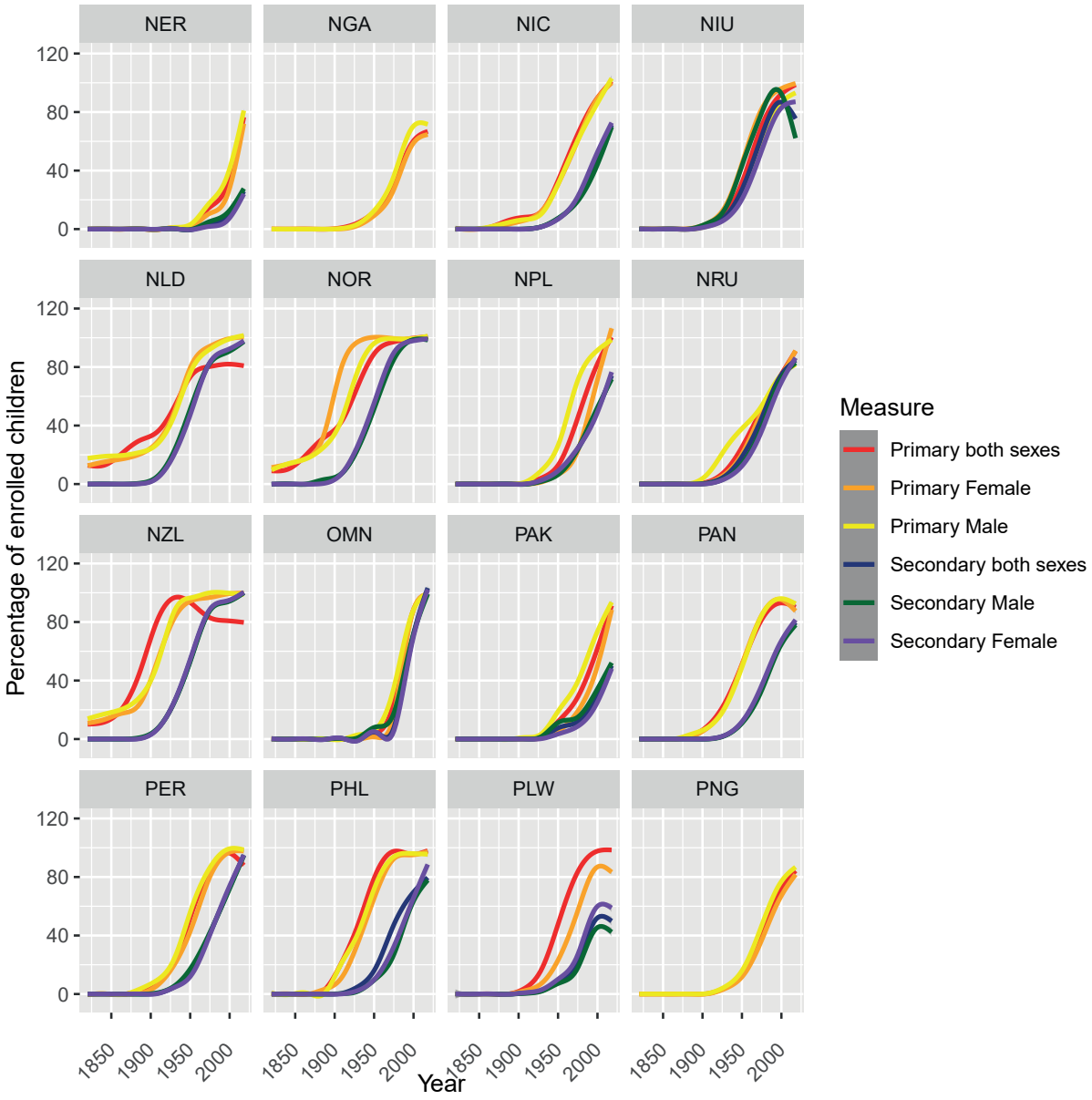

DMA
DNK
DOM
DZA
ECU
EGY
ERI
ESP
EST
ETH
FIN
FJI
FRA
FSM
GBR
GEO
Percentage of enrolled children
0
40
80
120
1850
1900
1950
2000
Year
Measure
Primary both sexes
Primary Female
Primary Male
Secondary both sexes
Secondary Male
Secondary Female

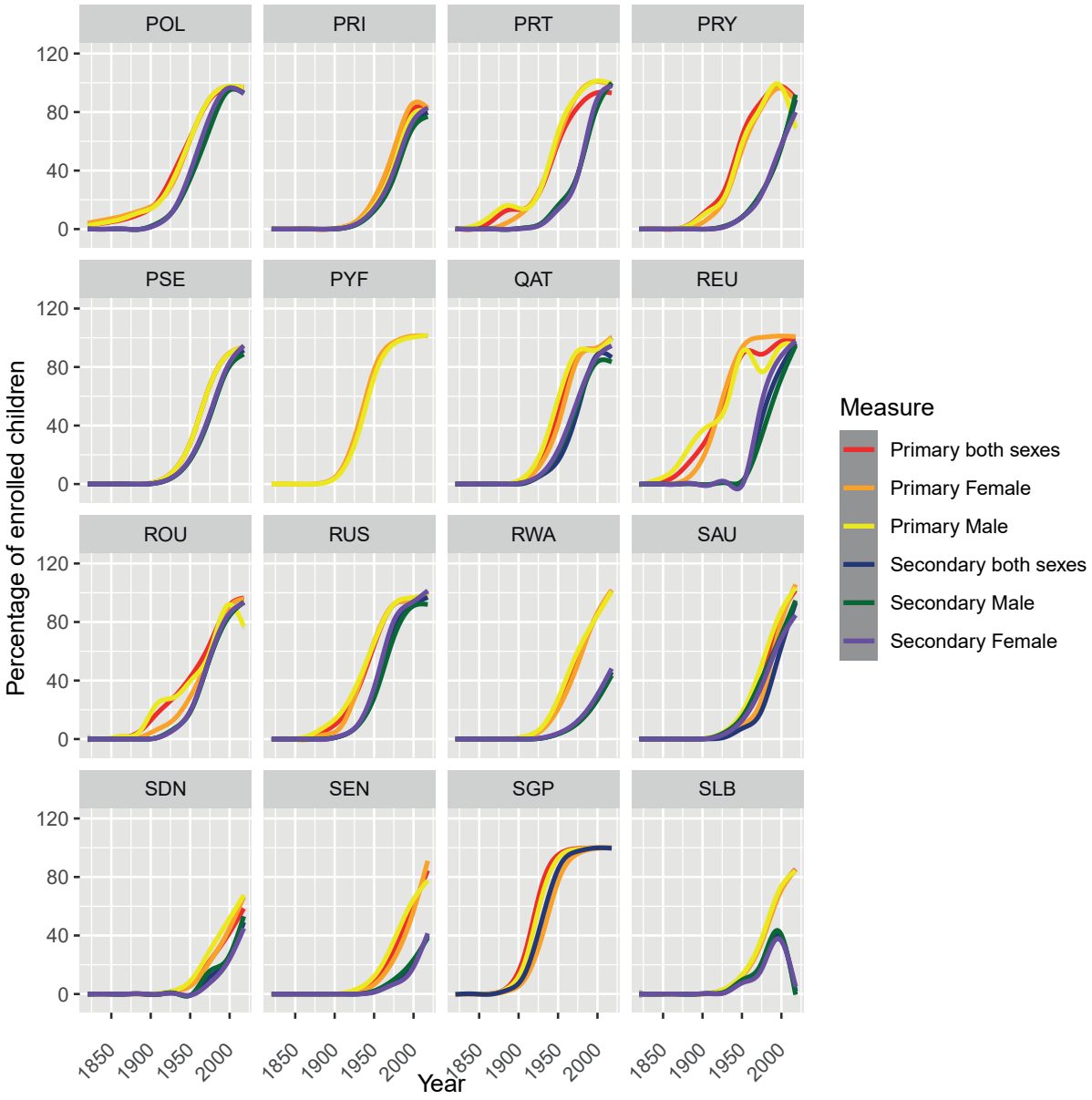

GHA
GIB
GIN
GLP
GMB
GNB
GRC
GRD
GTM
GUY
HKG
HND
HRV
HTI
HUN
IDN
120
80
40
0
1850
1900
1950
2000
Percentage of enrolled children
Year
Measure
Primary both sexes
Primary Female
Primary Male
Secondary both sexes
Secondary Male
Secondary Female

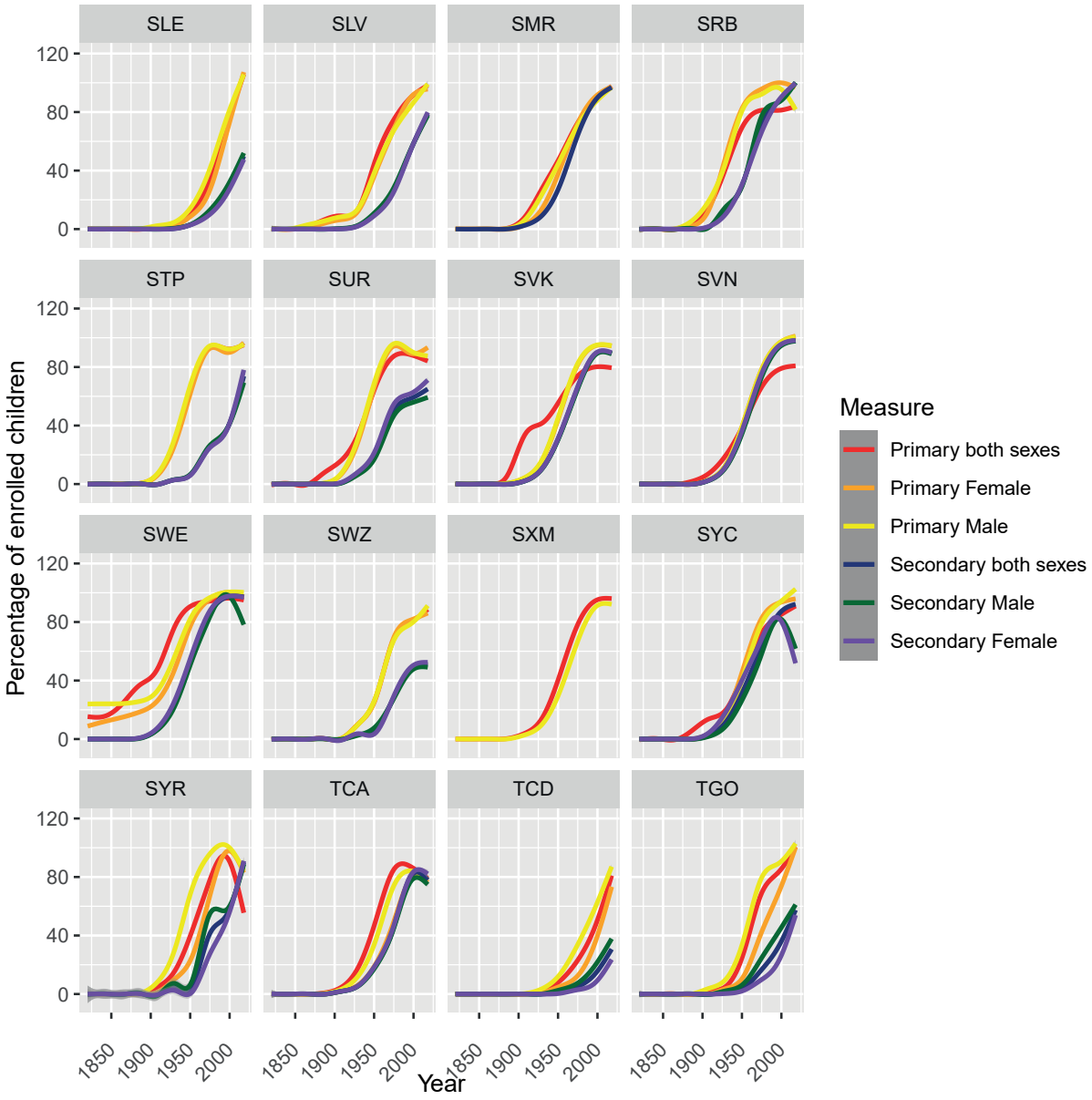

IND
IRL
IRN
IRQ
ISL
ISR
ITA
JAM
JOR
JPN
KAZ
KEN
KGZ
KHM
KIR
KOR
120
80
40
0
1850
1900
1950
2000
Percentage of enrolled children
Year
Measure
Primary both sexes
Primary Female
Primary Male
Secondary both sexes
Secondary Male
Secondary Female

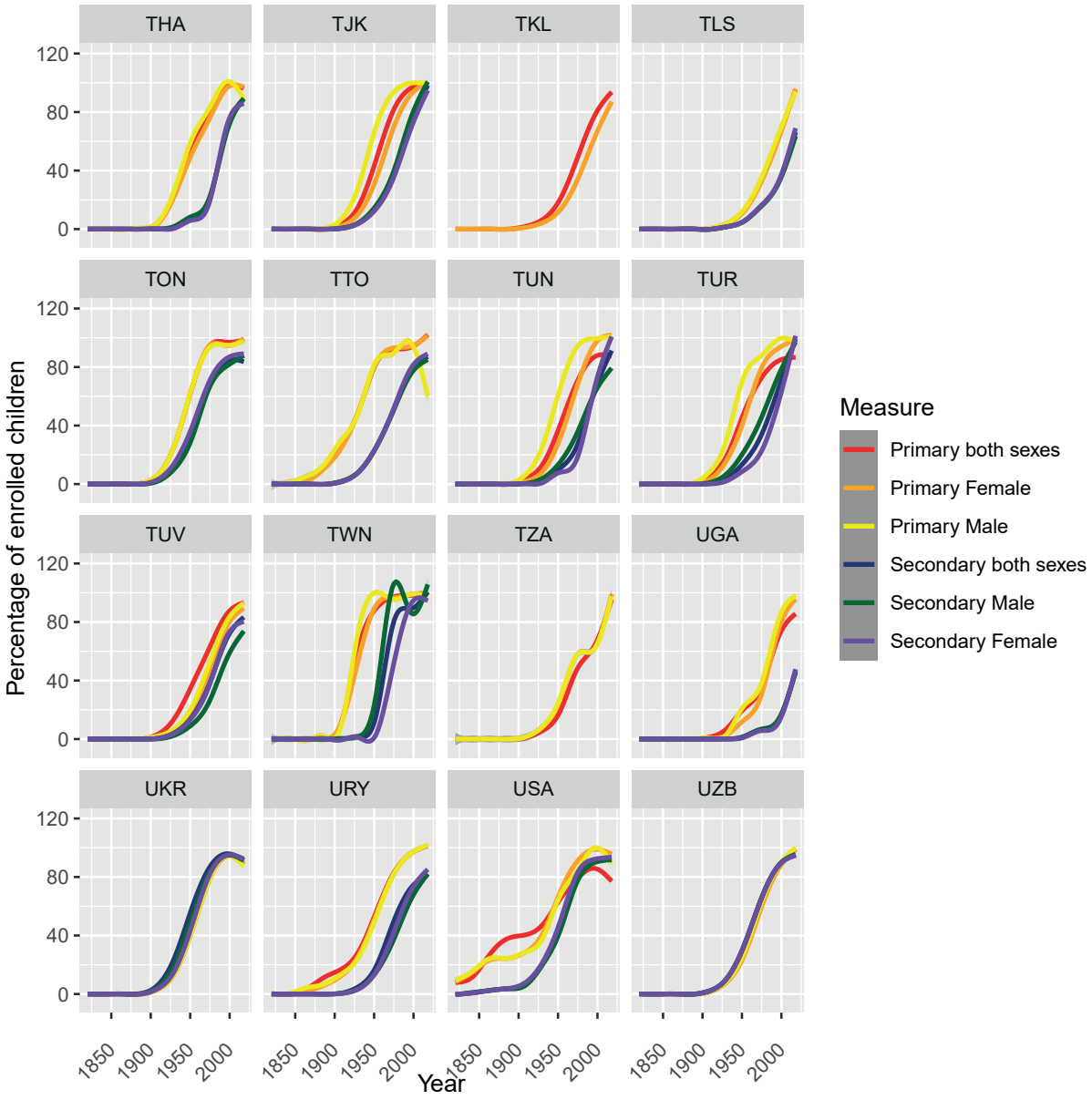

KWT
LAO
LBN
LBR
LBY
LCA
LIE
LKA
LSO
LTU
LUX
LVA
MAC
MAR
MDA
MDG
120
80
40
0
1850
1900
1950
2000
Year
Percentage of enrolled children
Measure
Primary both sexes
Primary Female
Primary Male
Secondary both sexes
Secondary Male
Secondary Female

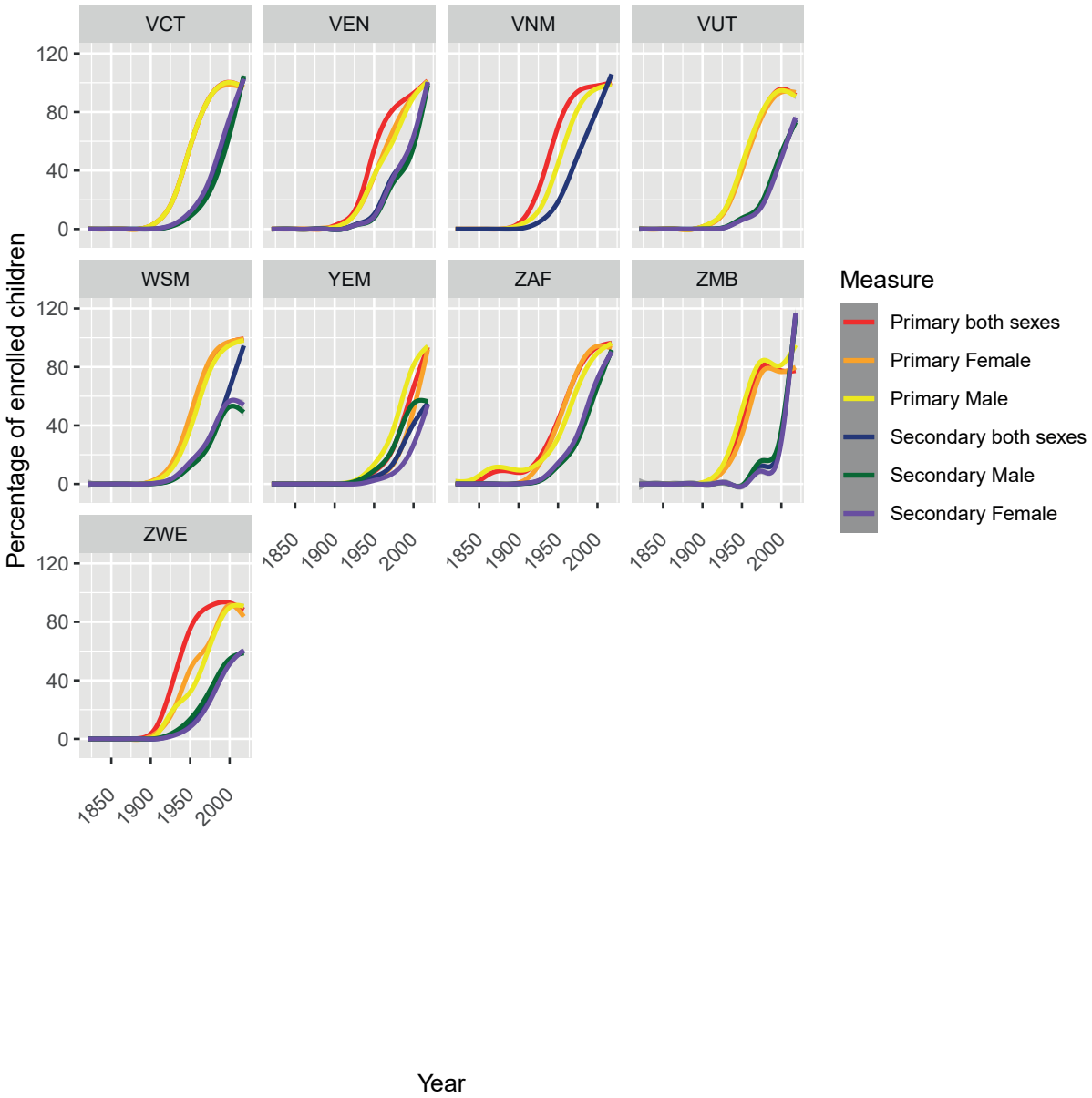

MDV
MEX
MHL
MKD
MLI
MLT
MMR
MNE
MNG
MOZ
MRT
MSR
MUS
MWI
MYS
NAM
120
80
40
0
1850
1900
1950
2000
Percentage of enrolled children
Year
Measure
Primary both sexes
Primary Female
Primary Male
Secondary both sexes
Secondary Male
Secondary Female


NER
NGA
NIC
NIU
NLD
NOR
NPL
NRU
NZL
OMN
PAK
PAN
PER
PHL
PLW
PNG
120
80
40
0
1850
1900
1950
2000
Year
Percentage of enrolled children
Measure
Primary both sexes
Primary Female
Primary Male
Secondary both sexes
Secondary Male
Secondary Female


POL
PRI
PRT
PRY
PSE
PYF
QAT
REU
ROU
RUS
RWA
SAU
SDN
SEN
SGP
SLB
120
80
40
0
1850
1900
1950
2000
Percentage of enrolled children
Year
Measure
Primary both sexes
Primary Female
Primary Male
Secondary both sexes
Secondary Male
Secondary Female


SLE
SLV
SMR
SRB
STP
SUR
SVK
SVN
SWE
SWZ
SXM
SYC
SYR
TCA
TCD
TGO
120
80
40
0
1850
1900
1950
2000
Year
Percentage of enrolled children
Measure
Primary both sexes
Primary Female
Primary Male
Secondary both sexes
Secondary Male
Secondary Female


THA
TJK
TKL
TLS
TON
TTO
TUN
TUR
TUV
TWN
TZA
UGA
UKR
URY
USA
UZB
120
80
40
0
1850
1900
1950
2000
Percentage of enrolled children
Year
Measure
Primary both sexes
Primary Female
Primary Male
Secondary both sexes
Secondary Male
Secondary Female


VCT
VEN
VNM
VUT
WSM
YEM
ZAF
ZMB
ZWE
Percentage of enrolled children
120
80
40
0
1850
1900
1950
2000
Year
Measure
Primary both sexes
Primary Female
Primary Male
Secondary both sexes
Secondary Male
Secondary Female