The Subnational CRVS R Package

Jeremy Roth

Contents

1	Set	ıp	1			
2	asets Included with the Package: Tabulations from Ecuador	2				
3	Example of Subnational Analysis: Provinces in Ecuador					
	3.1	Demographic Data Quality Assessment (DDQA) $\ \ldots \ \ldots \ \ldots \ \ldots \ \ldots \ \ldots$	3			
	3.2	DDM Estimation of Death Registration Completeness	13			
4	Example of National Analysis: Ecuador					
	4.1	Plot of sex ratios	17			
	4.2	Plot of age ratios	18			
	4.3	Plots related to age heaping	19			
	4.4	Plot of DDM estimates	22			
$\mathbf{R}_{\mathbf{c}}$	efere	nces	24			

1 Setup

The SubnationalCRVS package, which is still under active development, is hosted on my GitHub page (www.github.com/jroth-unfpa/SubnationalCRVS) rather than on CRAN. As a result, SubnationalCRVS must be installed with the install.github() function from the devtools package instead of the usual install.packages() function. The required dependency DemoTools (Riffe et al. 2019) is also hosted on GitHub instead of CRAN and thus must also be installed with install.github().

```
library(devtools)
install_github("timriffe/DemoTools") # install the DemoTools dependency
install_github("jroth-unfpa/SubnationalCRVS") # install the SubnationalCRVS package
```

Now we can load SubnationalCRVS, specify the name of the local folder in which we will save the plots produced by the package, and create that local folder if it does not already exist.

```
library(SubnationalCRVS)
my_plots_dir <- "Plots/" # local folder where the plots should be saved
dir.create(my_plots_dir) # create the folder if it does not already exist</pre>
```

For this demonstration, we will also load the dplyr package to customize the display of some of the tables returned by SunbationalCRVS.

```
library(dplyr)
```

2 Datasets Included with the Package: Tabulations from Ecuador

The SubnationalCRVS package comes with tabulations of population and registered deaths in Ecuador – disaggregated by age, sex, and province – created from publicly available datasets from Ecuador (Peralta et al. 2019; INEC 2010) based on the Ecuador's 2001 Census, 2010 Census, and annual counts of registered deaths from 2001 through 2010. The structure of the included example datasets is inspired by the data requirements of the ddm() function from the DDM package (Riffe, Lima, and Queiroz 2017).

The ecuador_single_year_ages_combined dataset reports the estimated populations in Ecuador by single-year ages (0, 1, 2, ... in the age column) from both the 2001 Census (pop1 column) and the 2010 Census (pop2 column), separately for males and females (m and f in the sex column) and province (in the province column).

```
head(ecuador_single_year_ages)
```

```
##
     province_name sex age pop1 pop2
                                            date1
                                                        date2
## 1
                          0 6086 6750 2001-11-25 2010-11-28
             Azuay
                      m
             Azuay
## 2
                          1 6555 6984 2001-11-25 2010-11-28
## 3
                          2 7232 7090 2001-11-25 2010-11-28
             Azuay
                      m
## 4
                          3 7101 7095 2001-11-25 2010-11-28
             Azuay
                      \mathbf{m}
## 5
             Azuay
                          4 7083 6961 2001-11-25 2010-11-28
                      m
## 6
             Azuay
                          5 6583 6895 2001-11-25 2010-11-28
```

The ecuador_single_year_ages_combined dataset appends rows to ecuador_single_year_ages_combined that report the sex- and single-year-age- disaggregated population estimates for the entire country (as opposed to a single province).

```
pop2
##
                             pop1
                                                          date2
     province_name sex age
                                               date1
## 1
          National
                     m
                         0 117897 123013 2001-11-25 2010-11-28
## 2
          National
                         1 133280 141602 2001-11-25 2010-11-28
                     m
## 3
          National
                         2 141869 141719 2001-11-25 2010-11-28
## 4
          National
                         3 134793 143418 2001-11-25 2010-11-28
                     m
                         4 139268 143830 2001-11-25 2010-11-28
## 5
          National
                     m
## 6
          National
                         5 131491 136479 2001-11-25 2010-11-28
```

The ecuador_five_year_ages and ecuador_five_year_ages_combined datasets have the same variables as ecuador_single_year_ages_combined and ecuador_single_year_ages_combined with two exceptions: (1) the age variable now represents five-year age groups (in the age column, with 0-4 coded as 0, 5-9 coded as 5, 10-14 coded as 10, etc.) instead of single-year ages; and (2) there is an additional column called deaths that reports the registered deaths collected between 2001 and 2010.

```
head(ecuador_five_year_ages)
##
     province_name sex age
                            pop1 pop2 deaths
                                                     date1
                                                                 date2
## 1
             Azuay
                          0 34101 34886
                                            772 2001-11-25 2010-11-28
                     m
## 2
                         10 34946 38125
                                            223 2001-11-25 2010-11-28
             Azuay
## 3
             Azuay
                         15 32387 37611
                                           416 2001-11-25 2010-11-28
                     m
## 4
             Azuay
                     m
                         20 25634 33665
                                            480 2001-11-25 2010-11-28
## 5
             Azuay
                         25 18606 28376
                                            475 2001-11-25 2010-11-28
                     m
## 6
             Azuay
                        30 16193 22026
                                            456 2001-11-25 2010-11-28
head(ecuador_five_year_ages_combined)
##
     province_name sex age
                            pop1 pop2 deaths
                                                     date1
                                                                 date2
## 1
                                            772 2001-11-25 2010-11-28
             Azuay
                     m
                          0 34101 34886
                                            223 2001-11-25 2010-11-28
## 2
                         10 34946 38125
             Azuay
                     m
## 3
                         15 32387 37611
                                           416 2001-11-25 2010-11-28
             Azuay
                     m
## 4
                         20 25634 33665
                                            480 2001-11-25 2010-11-28
             Azuay
## 5
                         25 18606 28376
                                            475 2001-11-25 2010-11-28
             Azuav
                     m
                                            456 2001-11-25 2010-11-28
## 6
                        30 16193 22026
             Azuay
                     \mathbf{m}
ecuador_five_year_ages_combined %>% filter(province_name == "National") %>%
                                     head()
##
                                     pop2 deaths
                                                                   date2
     province_name sex age
                                                       date1
                              pop1
## 1
          National
                          0 678280 743576
                                           22338 2001-11-25 2010-11-28
## 2
                        10 679067 782559
                                             4290 2001-11-25 2010-11-28
          National
## 3
          National
                        15 616725 712878
                                            8293 2001-11-25 2010-11-28
## 4
                                           13412 2001-11-25 2010-11-28
          National
                        20 569964 638042
                     m
## 5
          National
                         25 456230 585652
                                            13301 2001-11-25 2010-11-28
## 6
          National
                        30 422307 519493
                                           11811 2001-11-25 2010-11-28
```

3 Example of Subnational Analysis: Provinces in Ecuador

In this section, we describe the key functions of SubnationalCRVS in the context of visualizing outputs from a demographic data quality assessment (DDQA) and estimates of death registration completeness (Riffe, Lima, and Queiroz 2017) within provinces of Ecuador. Later in this tutorial, we also show how SubnationalCRVS provides corresponding visualizations for national-level results.

3.1 Demographic Data Quality Assessment (DDQA)

3.1.1 Sex ratios: PlotSexRatios()

A key step in this demographic data quality assessment (DDQA) process is to use the PlotSexRatios() function to compute and plot sex ratios within each combination of province, sex, and single-year age for the 2001 and 2010 data stored in ecuador_single_year_ages_combined.

To use PlotSexRatios(), we are required to provide our tabulated data frame in the data argument and a few additional required arguments that describe the variable names and values in data. These arguments and the expected format of the specified dataset are motivated by the data structure enforced in the DDM package (Riffe, Lima, and Queiroz 2017).

- name.disaggregations is the name of variable representing the subnational disaggregation (apart from sex, which is required,) in data. Here we specify name.disaggregations="province name".
- name.sex is the name of variable representing sex. Here we specify name.sex="sex"
- name.age is the name of variable representing age. Here we specify name.age="age"
- name.date1 is the name of variable that provides the date of the earlier of the two time periods represented in data. Here we specifyname.date1="date1"(the value of this variable, "2001-11-15" is the date of Ecuador's 2001 Census)
- name.date2 is the name of variable that provides the date of the earlier of the two time periods represented in data. Here we specify name.date2="date2" (the value of this variable, "2010-11-28" is the date of Ecuador's 2010 Census)
- name.population.year1 is the name of variable representing the population in the earlier of the two time periods represented in the dataset. Here we specify name.population.year2="pop1"
- name.population.year2 is the name of variable representing the population in the earlier of the two time periods represented in the dataset. Here we specify name.population.year2="pop2"
- name.male is the name of value of the name.sex variable that represents males. Here we specify name.males="m"
- name.female is the name of value of the name.sex variable that represents females. Here we specify name.females="f"

In addition, we specify two optional arguments, plots.dir and name.national. We set plots.dir=my_plots_dir so that the plots will save in the local folder Plots; if the plots.dir argument is omitted, the plots will be saved in the same working directory of the R script. We also set name.national="National" to reflect the fact that national-level datasets are provided in rows where the name.disaggregations variable has the value "National". Specifying name.national produces visualizations catered specifically to national-level analysis (instead of accommodated further disaggregation) that will be presented later in the tutorial. If name.national is not provided, those national-level visualizations are simply not produced.

The plots of sex ratios are saved in the Plots/sub-folder we specified with the argument plots.dir=my_plots_dir; plots.dir is an optional argument and, if we do not specify a value for it, the plots will be saved the working directory. We also specified label.subnational.level="Province" so that the disaggregations are labeled Province instead of the less clear province_name.

The sex ratios for all levels of subnational disaggregation are overlaid in the following "combined" plots separately for each data year.

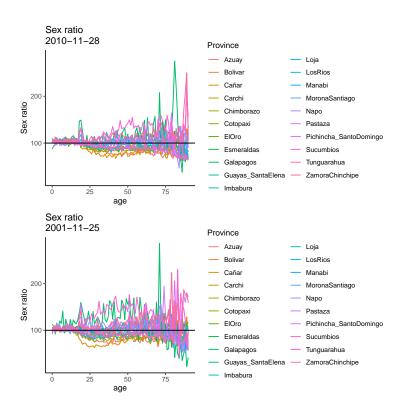


Figure 1: Sex ratios in Ecuador by province, combined plot

Additionally, the sex ratios are plotted in separate figures for each level of subnational disaggregation in the following "disaggregated" plots.

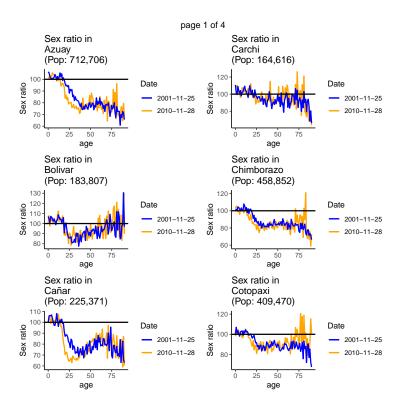


Figure 2: Sex ratios in Ecuador by province, disaggregated plots (only first page shown)

3.1.2 View sex ratios in table

The object returned by PlotSexRatios() is a table that shows us the sex ratios for each combination of province, sex, and single-year age in the sex_ratio_1 column (for the 2001 Census) and the sex_ratio_2 column (for the 2010 Census).

```
s %>% select(province_name, age, pop1, pop2, sex_ratio_1, sex_ratio_2) %>%
head()
```

```
pop1 pop2 sex_ratio_1 sex_ratio_2
##
     province_name age
## 1
              Azuay
                      0 12073 13092
                                            101.7
                                                         106.4
## 2
                      1 13060 13596
                                            100.8
                                                         105.6
              Azuay
              Azuay
                                            103.9
## 3
                      2 14195 14014
                                                         102.4
                                             99.8
                                                         100.9
##
  4
              Azuay
                        14217 14124
## 5
              Azuay
                        14012 13911
                                            102.2
                                                         100.2
                                            105.6
## 6
              Azuay
                      5 12815 13736
                                                         100.8
```

3.1.3 Age ratios: PlotAgeRatios()

Another step in our demographic data quality assessment is using the PlotAgeRatios() function in the SubnationalCRVS package to compute and plot age ratios within each combination of province and sex for the 2001 and 2010 data stored in ecuador_five_year_ages_combined. The arguments we provide to PlotAgeRatios() are actually identical to those we specified for PlotSexRatios(), except now we are using the tabulation with five-year age groups (ecuador_five_year_ages_combined) instead of the tabulation with single-year ages.

3.1.4 View age ratios in combined plot

The following "combined" plots, saved in the Plots/ folder, show the age ratios for all levels of subnational disaggregation, separately for males and females in each data year.

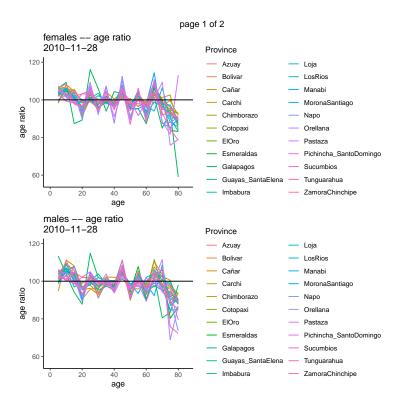


Figure 3: Age ratios in Ecuador by province, combined plot

In addition, PlotAgeRatios() also creates the following "disaggregated" plots, saved in the Plots/ folder,

where the age ratios for each level of disaggregation are shown in separate plots, with different sexes and data years overlaid within each plot.

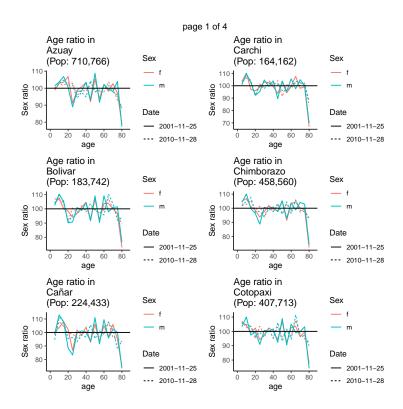


Figure 4: Age ratios in Ecuador by province, disaggregated plots (only first page shown)

Just as the PlotSexRatios() function returns a table of disaggregated sex ratios, PlotAgeRatios() returns a table of disaggregated age ratios.

```
a %>% select(province_name, age, pop1, pop2, age_ratio_1, age_ratio_2) %>% head()
```

```
##
                                pop2 age_ratio_1 age_ratio_2
     province_name age
                         pop1
## 1
                                                NA
              Azuay
                      0 33491 33876
                                                             NA
                               35701
## 2
              Azuay
                      5 33817
                                             98.8
                                                         100.2
## 3
              Azuay
                      10 34975 37366
                                            102.9
                                                         102.5
## 4
                                            103.6
                                                         101.8
              Azuay
                      15 34181 37215
                                            106.9
                                                         103.2
## 5
              Azuay
                     20 31000 35753
## 6
                     25 23844 32054
                                             91.2
                                                         102.9
              Azuay
```

3.1.5 Potential age heaping: PlotPotentialAgeHeaping()

To give us a sense of whether "age-heaping" is occurring within the levels of disaggregation present in our ecuador_single_year_ages_combined dataset, we turn to the PlotPotentialAgeHeaping() function and

actually provide the same arguments we used in the PlotSexRatios() function.

3.1.6 View potential age heaping in combined plot

The following "combined" plots show us estimated population counts by single-year ages with different provinces represented with different overlaid colors, and separate plots for each sex and data year.

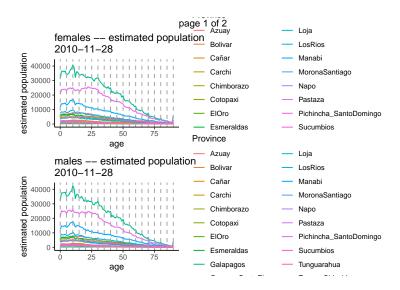


Figure 5: Population counts in Ecuador by single-year age, combined plot

PlotPotentialAgeHeaping() present separate plots of population counts for each province in Ecuador, with different sexes and data years overlaid within each plot.

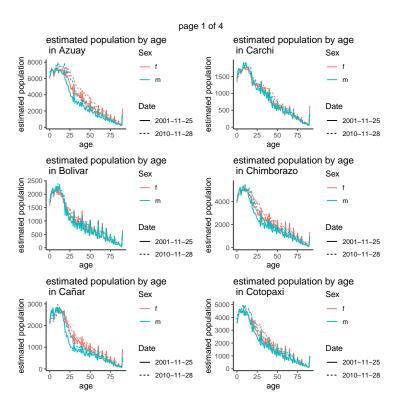


Figure 6: Population counts in Ecuador by single-year age, disaggregated plots

3.1.7 Age heaping indices: PlotAgeHeapingScores()

As a more concise summary of potential age-heaping suggested by the visualizations from PlotPotentialAgeHeaping, we now use the PlotAgeHeapingScores function with the same arguments we provided to the PlotAgeRatios function.

One set of plots returned by PlotAgeHeapingScores() shows the values of three age-heaping indices within each combination of province, sex, and data year: Roughness, Whipple, and Myers. The indices are computed with the check_heaping_roughness(), check_heaping_Whipple(), and check_heaping_myers() functions, respectively, from the DemoTools package (Riffe et al. 2019).

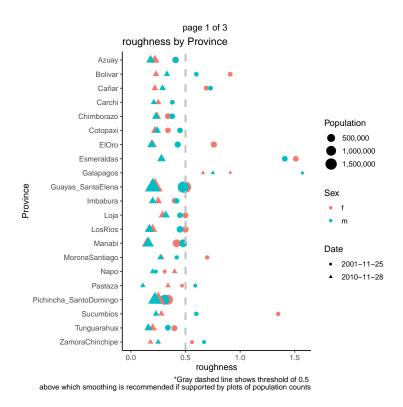


Figure 7: Roughness, Whipple, and Myers indices in Ecuador by province (only first page shown)

Further, PlotAgeHeapingScores() also creates plots showing the Noumbissi index in each province computed for each terminal digit from 0-9, using the check_heaping_Noumbissi() function from the DemoTools package (Riffe et al. 2019).

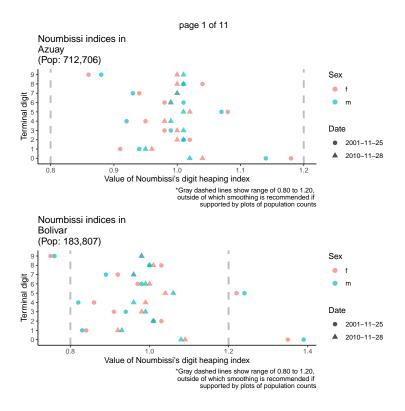


Figure 8: Noumbissi indices in Ecuador by province (only first page shown)

We can also view the age-heaping indices in the table returned by PlotAgeHeapingScores()

head(ageheaping)

##		province_name	date	total_pop	sex	roughness	Whipple	Myers	Noumbissi_0
##	1	Azuay	2001-11-25	319983	f	0.41	1.18	4.21	1.18
##	2	Bolivar	2001-11-25	86256	f	0.91	1.37	7.39	1.35
##	3	Cañar	2001-11-25	112041	f	0.69	1.22	4.89	1.22
##	4	Carchi	2001-11-25	77172	f	0.38	1.18	3.75	1.15
##	5	Chimborazo	2001-11-25	213106	f	0.34	1.25	5.44	1.23
##	6	Cotopaxi	2001-11-25	180328	f	0.34	1.27	5.99	1.25
##		Noumbissi_1 No	oumbissi_2 N	Noumbissi_3	3 Noi	umbissi_4 N	Voumbissi	i_5 Noi	umbissi_6
##	1	0.91	1.02	0.98	3	0.95	1.	.08	0.98
##	2	0.84	1.03	0.91	L	0.86	1.	. 22	0.97
##	3	0.90	1.02	0.98	3	0.91	1.	. 11	0.98
##	4	0.91	1.05	0.96	3	0.92	1.	. 11	0.98
##	5	0.88	1.01	0.97	7	0.89	1.	.12	0.98
##	6	0.88	1.02	0.95	5	0.90	1.	. 18	0.98
##		Noumbissi_7 Noumbissi_8 Noumbissi_9							
##	1	0.94	1.04	0.86	3				
##	2	0.92	1.03	0.75	5				
##	3	0.95	1.03	0.80)				
##	4	0.96	1.04	0.86	3				
##	5	0.94	1.05	0.82	2				
##	6	0.91	1.06	0.80)				

3.2 DDM Estimation of Death Registration Completeness

The structure of the ecuador_five_year_ages dataset is inspired by the requirements for the ddm() function from the DDM package (Riffe, Lima, and Queiroz 2017), which uses established Death Distribution Methods (DDM) to estimate death registration completeness for adults (aged 15+) between two consecutive Censuses. Essentially, the SubnationalCRVS package offers its EstimateDDM() function only as a convenient wrapper to DDM::ddm() to perform DDM estimation of death registration completeness using the same pipeline used to perform the DDQA.

3.2.1 Compute DDM estimates: EstimateDDM()

[1] "performing GGB-SEG estimation within each of 20 possible age ranges..."

We note EstimateDDM() uses exactly the same arguments as PlotAgeRatios(), for example, from the DDQA with two additional required arguments:

- name.deaths, which provides the name of the variable representing the count of registered deaths between the two dates represented in name.date1 and name.date2
- deaths.summed, which should be set to TRUE when the name.deaths variable represents the total number of registered deaths name.date1 and name.date2 and set to FALSE when the name.deaths variable represents the average number of registered deaths between the two dates.

3.2.2 Plot DDM estimates: PlotDDM()

We can plot the estimated adult death registration completeness (using the "hybrid" GGB-SEG method) with the PlotDDM function:

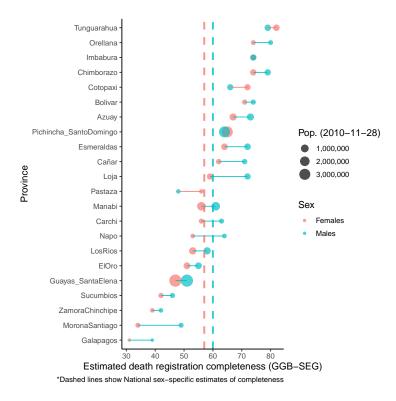


Figure 9: Point estimates of death registration completeness in Ecuador from 2001-2010, using the GGB-SEG method

The EstimateDDM function also returns a list, in which the ddm_estimates object contains the estimated death registration completeness with the GGB-SEG approach

```
head(ddm_results$ggbseg_estimates)
##
                  sex ggbseg lower_age_range upper_age_range total_pop1 total_pop2
                                                                                710766
## 1
       Azuay Females
                        0.67
                                            15
                                                             60
                                                                    599313
  2
                        0.73
                                                                    599313
                                                                                710766
##
       Azuay
                Males
                                            15
                                                             65
## 3 Bolivar Females
                        0.71
                                            15
                                                             65
                                                                    170696
                                                                                183742
                        0.74
## 4 Bolivar
                Males
                                            15
                                                             65
                                                                    170696
                                                                                 183742
## 5
                        0.62
                                                             65
                                                                                224433
       Cañar Females
                                            15
                                                                    206346
## 6
       Cañar
                Males
                        0.71
                                            15
                                                             65
                                                                    206346
                                                                                224433
```

Additionally, PlotDDM presents a visualization

of the point estimates of completeness and the corresponding root mean squared errors (RMSEs) for each of the permitted values of the age range that is selected DDM::ddm() as part of its fitting procedure.

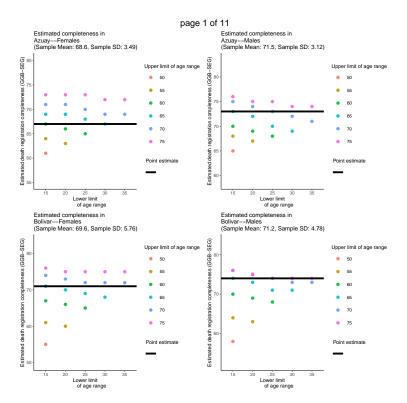


Figure 10: Sensitivity of point estimates of death registration completeness in Ecuador from 2001-2010 to choice of age-range parameter in the GGB-SEG method

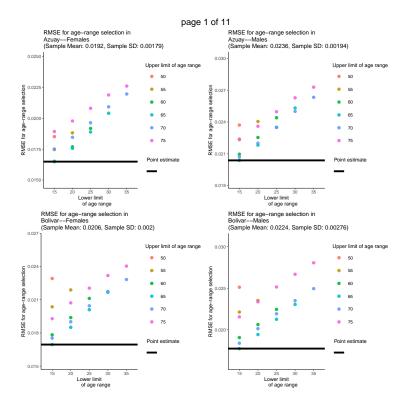


Figure 11: Sensitivity of point estimates of death registration completeness in Ecuador from 2001-2010 to choice of age-range parameter in the GGB-SEG method

The sensitivity estimates are also returned in table by EstimateDDM, in the sensitivity_ddm_estimates element of its list.

```
head(ddm_results$sensitivity_ggbseg_estimates, n=5)
##
       cod
                sex ggbseg lower_age_range upper_age_range total_pop1 total_pop2
## 1 Azuay Females
                      0.61
                                         15
                                                          50
                                                                  599313
                                                                              710766
## 2 Azuay Females
                      0.64
                                                                              710766
                                         15
                                                          55
                                                                  599313
## 3 Azuay Females
                      0.63
                                         20
                                                          55
                                                                  599313
                                                                              710766
## 4 Azuay Females
                      0.67
                                         15
                                                          60
                                                                  599313
                                                                              710766
## 5 Azuay Females
                      0.66
                                         20
                                                          60
                                                                  599313
                                                                              710766
##
        RMSE
## 1 0.01851
## 2 0.01747
## 3 0.01882
## 4 0.01649
## 5 0.01769
```

4 Example of National Analysis: Ecuador

As discussed, specifying the argument name.national="National" during the previous DDQA and estimation of death registration completeness permitted SubnationalCRVS to save plots specifically designed to display national-level (as opposed to subnational-level) visualizations. These

national-level plots are can also be produced by specifying a national-level-only dataset, for example ecuador_five_year_ages_national (shown below), as long as the corresponding name.national argument is also specified (e.g. name.national="National" here)

```
head(ecuador_five_year_ages_national)
```

```
##
                                   pop2 deaths
    province_name sex age
                            pop1
                                                    date1
                                                               date2
## 1
         National
                    m
                        0 678280 743576
                                        22338 2001-11-25 2010-11-28
## 2
         National
                                          4290 2001-11-25 2010-11-28
                    m 10 679067 782559
                    m 15 616725 712878
## 3
         National
                                          8293 2001-11-25 2010-11-28
## 4
         National
                    m 20 569964 638042
                                         13412 2001-11-25 2010-11-28
                                         13301 2001-11-25 2010-11-28
## 5
         National m 25 456230 585652
                                         11811 2001-11-25 2010-11-28
## 6
                       30 422307 519493
         National
```

tail(ecuador_five_year_ages_national)

```
##
      province_name sex age
                              pop1
                                     pop2 deaths
                                                      date1
                                                                  date2
## 31
                         60 149508 203789
                                           12500 2001-11-25 2010-11-28
           National
                      f
## 32
           National
                         65 126310 166559
                                           14635 2001-11-25 2010-11-28
## 33
                         70
                             99469 123647
                                           18139 2001-11-25 2010-11-28
           National
                      f
## 34
           National
                      f
                         75
                             73819
                                    86449
                                           21616 2001-11-25 2010-11-28
                                           23629 2001-11-25 2010-11-28
## 35
           National
                      f
                         80
                             52335
                                    62304
                                           52329 2001-11-25 2010-11-28
## 36
           National
                         85 72515
                                    55087
```

4.1 Plot of sex ratios

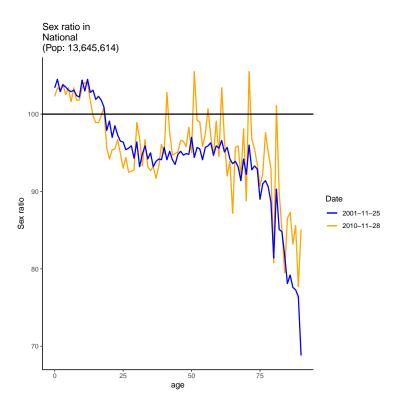


Figure 12: Sex ratios in Ecuador

4.2 Plot of age ratios

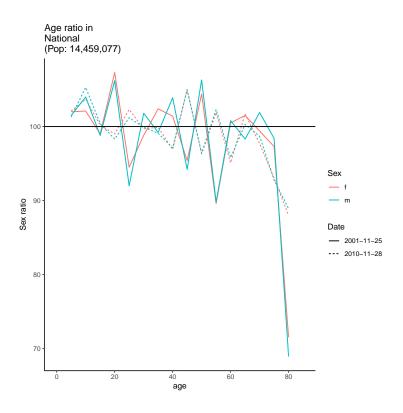


Figure 13: Age ratios in Ecuador

4.3 Plots related to age heaping

4.3.1 Plot of potential age heaping

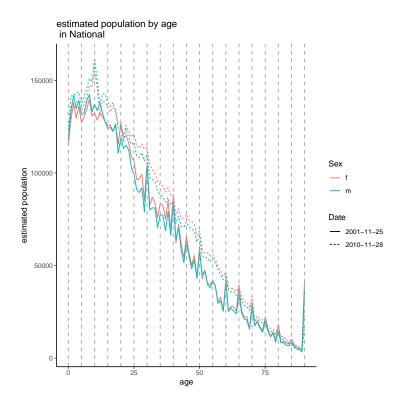


Figure 14: Population counts in Ecuador by single-year age

4.3.2 Plot of age heaping indices

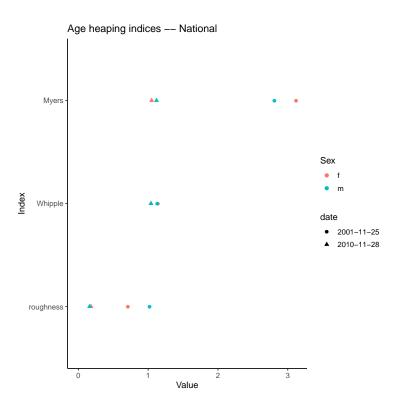


Figure 15: Roughness, Whipple, and Myers indices in Ecuador

4.3.3 Plot of Noumbissi indices

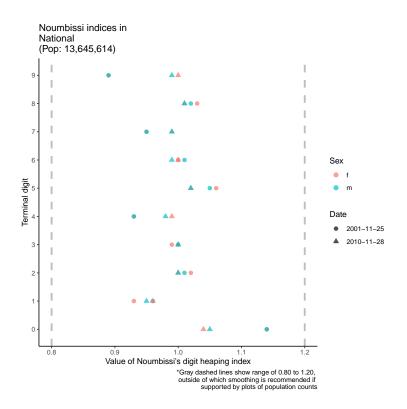


Figure 16: Noumbissi indices in Ecuador

4.4 Plot of DDM estimates

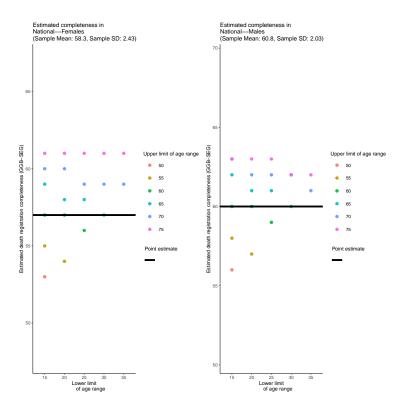


Figure 17: Sensitivity of point estimates of death registration completeness in Ecuador from 2001-2010 to choice of age-range parameter in the GGB-SEG method

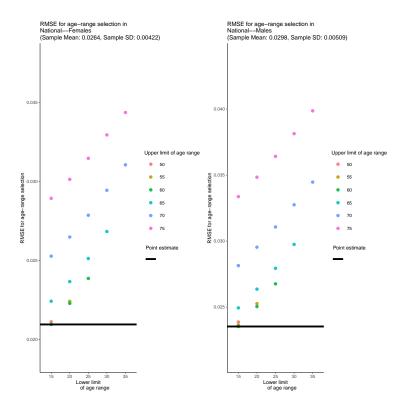


Figure 18: Sensitivity of point estimates of death registration completeness in Ecuador from 2001-2010 to choice of age-range parameter in the GGB-SEG method

References

INEC. 2010. "Censo de Población Y Vivienda." https://www.ecuadorencifras.gob.ec/censo-de-poblacion-y-vivienda//.

Peralta, Andrés, Joan Benach, Carme Borrell, Verónica Espinel-Flores, Lucinda Cash-Gibson, Bernardo L Queiroz, and Marc Marı'-Dell'Olmo. 2019. "Evaluation of the Mortality Registry in Ecuador (2001–2013)—Social and Geographical Inequalities in Completeness and Quality." *Population Health Metrics* 17 (1). Springer: 3.

Riffe, T, JM Aburto, M Alexander, S Fennell, I Kashnitsky, M Pascariu, and P Gerland. 2019. "DemoTools: An R Package of Tools for Aggregate Demographic Analysis."

Riffe, Tim, Everton Lima, and Bernardo Queiroz. 2017. "DDM: Death Registration Coverage Estimation."