Methods for Subnational Estimation of Child Mortality

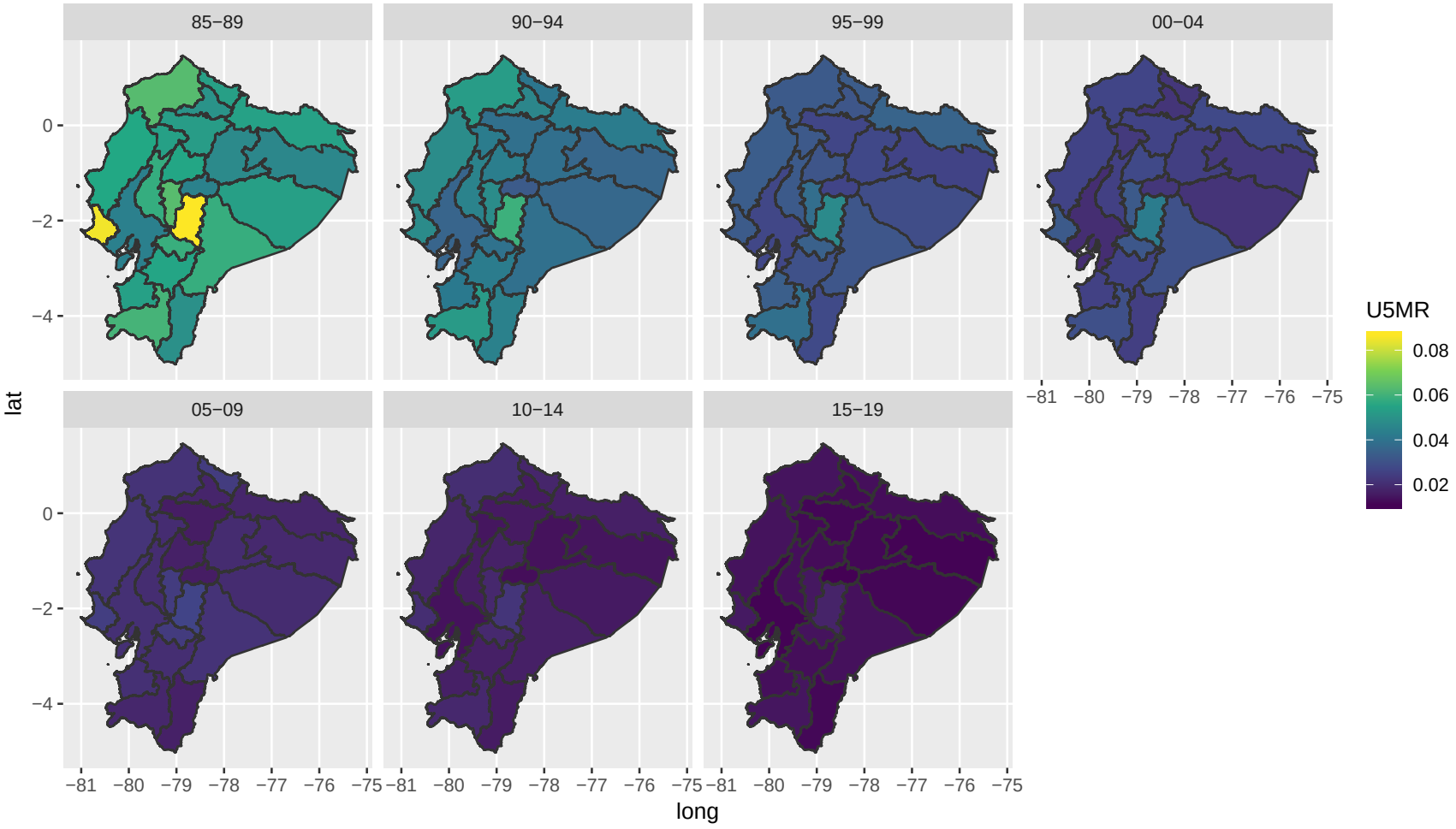
Hands-on Demo

Jon Wakefield[1,2] and Katie Wilson[1]

[1]Department of Biostatistics
[2]Department of Statistics
University of Washington

Introduction to the Package

Overview


85–89
90–94
95–99
00–04
05–09
10–14
15–19
lat
long
0
−2
−4
−81
−80
−79
−78
−77
−76
−75
U5MR
0.08
0.06
0.04
0.02

Overview

Full birth history data

	idhog	province	V005	V008	B3	B4	alive	B7	V021	V024	V025
1	1.015000e+13	Azuay	106.83381	1351	1320	Male	yes	*NA*	1	Azuay	1
2	1.015000e+13	Azuay	106.83381	1351	1131	Female	yes	*NA*	1	Azuay	1
3	1.015000e+13	Azuay	160.18701	1351	1339	Female	yes	*NA*	1	Azuay	1

`getBirths()`

Person-month format

	dob	survey_year	died	id.new	idhog	province	V005	V021	V024	V025	agemonth	obsStart
1	1320	2012	0	1	1.015e+13	Azuay	106.8338	1	Azuay	1	0	132
2	1320	2012	0	1	1.015e+13	Azuay	106.8338	1	Azuay	1	1	132
3	1320	2012	0	1	1.015e+13	Azuay	106.8338	1	Azuay	1	2	132

`countrySummary()`

Direct estimates

	region	years	u5m	lower	upper	logit.est	var.est	reg
1	All	85-89	0.049421740	0.0372606076	0.065282880	-2.956680	0.02267965	
2	All	90-94	0.043608351	0.0346791306	0.054706382	-3.087919	0.01479550	
3	All	95-99	0.028046933	0.0228419803	0.034396179	-3.545428	0.01154791	

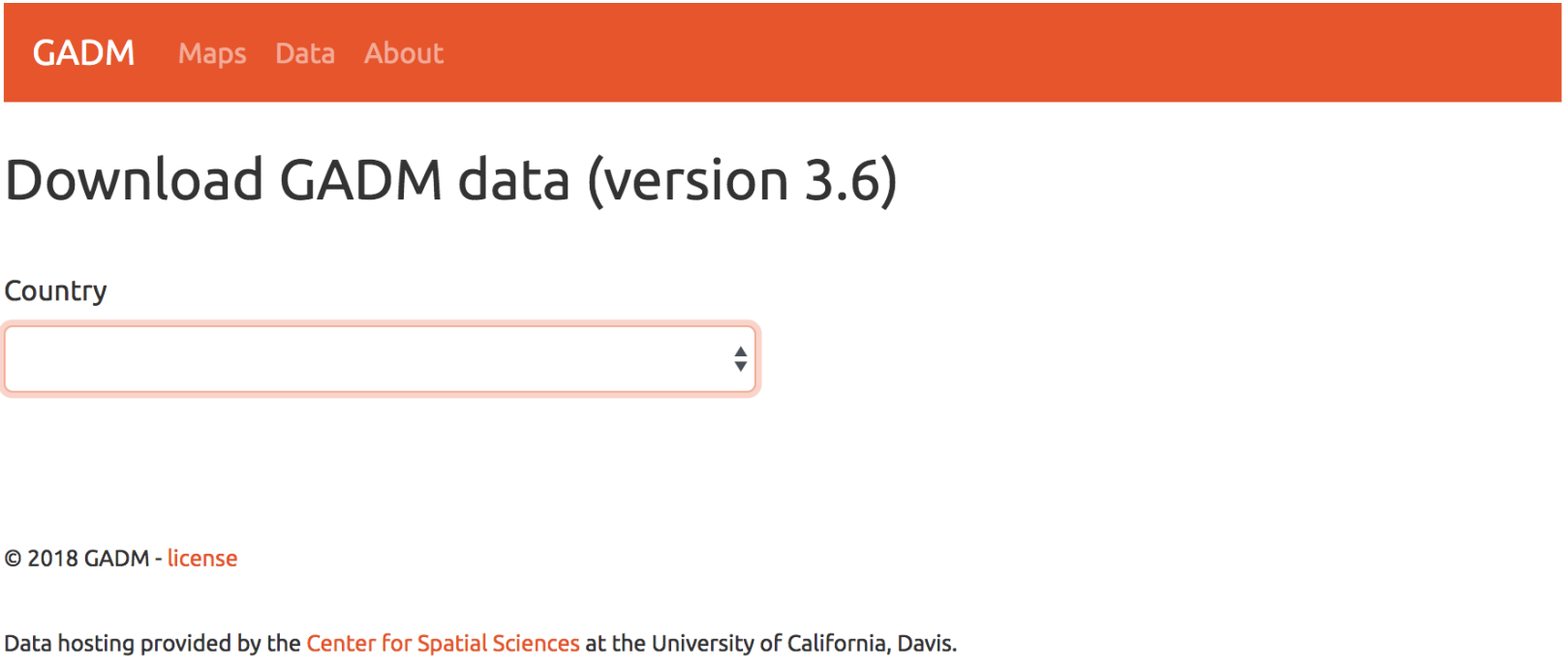
`fitINLA()`
`projINLA()`

Smoothed estimates

	region	years	logit.q975	logit.q025	logit.med	q975	q025	med
1	Azuay	1985	-2.0242780	-3.645622	-2.850483	0.11528930	0.024524320	0.054904
2	Bolivar	1985	-1.9009213	-3.479034	-2.707441	0.13193127	0.030647734	0.065774
3	Cañar	1985	-2.7320866	-4.936635	-3.882309	0.06187711	0.006283163	0.020821

Downloading the map data

There are 5 files needed for the map. They can be downloaded from: https://gadm.org/download_country_v3.html



Downloading the map data

GADM Maps Data About

Download GADM data (version 3.6)

Country

Ecuador

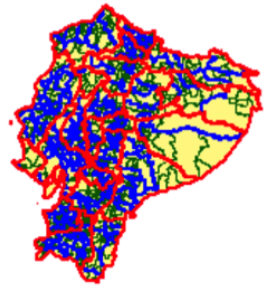
Geopackage

Shapefile

R (sp): level-0, level1, level2, level3

R (sf): level-0, level1, level2, level3

KMZ: level-0, level1, level2, level3



The Ecuador shapefiles are also available at http://faculty.washington.edu/jonno/UNICEF-WORKSHOPS/gadm36_ECU_shp.zip

R and R packages

We will be using the R package SUMMER and several other packages that will help us read in maps and maniuplate the data.

```
library(rgdal)
library(spdep)
library(SUMMER)
library(ggplot2)
library(gridExtra)
```

Overview of SUMMER

SUMMER can be used to obtain direct estimates of under-five mortality from full birth history data.

Smoothed estimates can then be obtained using SUMMER.

First, we will go over obtaining direct estimates of under-five mortality from full birth histories. Next, we will show code for obtaining the smoothed estimates from the direct estimates. An appendix contains additional code for obtaining estimates:

- Multiple surveys
- Subpopulations (e.g., males)
- Infant or neonatal mortality

Example: Subnational Direct Estimates

Preparing data

The `SUMMER` package can be used to obtain direct estimates from full birth histories. The data will need to organized such that every row corresponds to a birth and columns that correspond to

- Indicators corresponding to survey design (e.g., strata, cluster, and household)
- Survey weight
- Date of interview in century month code (CMC) format
- Date of child's birth in CMC format
- Indictor for death of child
- Age of death of child in months

Getting to know the data

Throughout, we will use the Ecuador ENSANUT 2012 data as an example.

Reading in the data

```
url <- "http://faculty.washington.edu/jonno/UNICEF-WORKSHOPS/ensanut2012fbh.csv
fbh <- read.csv(url)
head(fbh, n = 2)
```

```
##       idhog province     V005 V008   B3
## 1 1.015e+13    Azuay 106.8338 1351 1320
## 2 1.015e+13    Azuay 106.8338 1351 1131
##       B4 alive B7 V021  V024 V025
## 1   Male   yes NA    1 Azuay    1
## 2 Female   yes NA    1 Azuay    1
```

- idhog: household ID
- V005: weight
- V008: interview century month code (CMC)
- B3: birth CMC
- B7: age at death (months)
- V021: cluster ID
- V024: strata
- V025: urban/rural

Reorganize data into person-month format

We then reformat the data into person-months.

```
births <- getBirths(data = fbh, surveyyear = 2012,
    variables = c("idhog", "province", "V005",
        "V021", "V024", "V025"), strata = "V024",
    dob = "B3", alive = "alive", age = "B7",
    date.interview = "V008", year.cut = seq(1985,
        2015, by = 5))
```

Getting to know person-month format

```
births[1:2, 1:10]
```

```
##    dob survey_year died id.new
## 1 1320        2012    0      1
## 2 1320        2012    0      1
##       idhog province     V005 V021
## 1 1.015e+13    Azuay 106.8338    1
## 2 1.015e+13    Azuay 106.8338    1
##    V024 V025
## 1 Azuay    1
## 2 Azuay    1
```

Getting to know person-month format

```
table(births$V024, births$age)[1:5, ]
```

```
##
##                   0  1-11 12-23 24-35
##   Azuay        1282 13493 13766 12942
##   Bol\xedvar   1702 17817 18479 17562
##   Ca\xf1ar     1210 12738 13237 12460
##   Carchi       1500 15856 16217 15014
##   Chimborazo   1640 16992 17573 16685
##
##                36-47 48-59
##   Azuay        11934 11030
##   Bol\xedvar   16679 15875
##   Ca\xf1ar     11787 11180
##   Carchi       13995 13042
##   Chimborazo   15822 14873
```

Getting to know person-month format

```
table(births$V024, births$time)[1:5, ]
```

```
##
##                85-89 90-94 95-99 00-04
##   Azuay         2815  7143 10358 15138
##   Bol\xedvar    3792 10815 18114 22744
##   Ca\xf1ar      3534  7640 12411 15368
##   Carchi        2937  6657 12248 19024
##   Chimborazo    3914  9401 16841 21721
##
##                05-09 10-14
##   Azuay        20162  8831
##   Bol\xedvar   23086  9563
##   Ca\xf1ar     16403  7256
##   Carchi       24359 10399
##   Chimborazo   22702  9006
```

Direct estimates

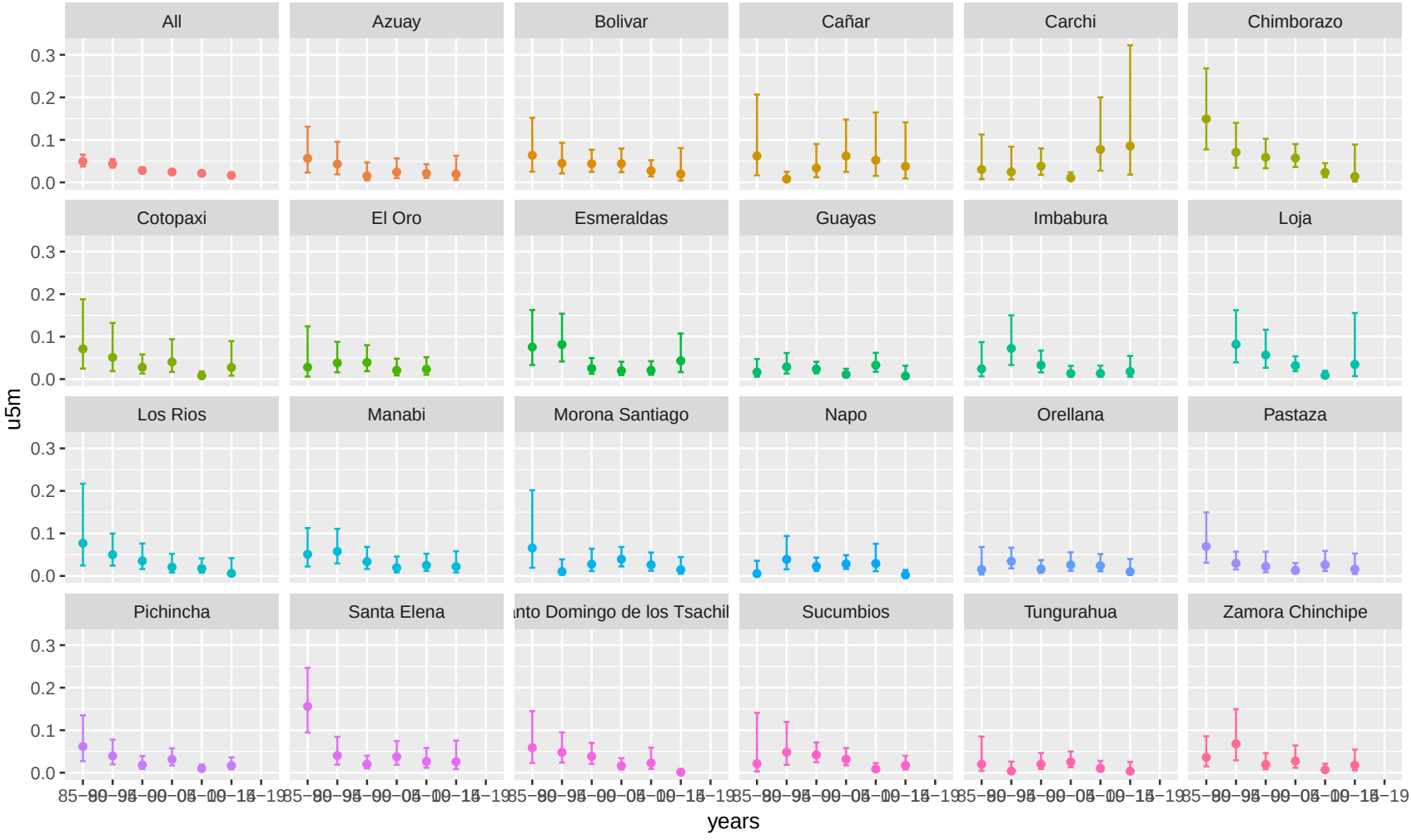
Now we are ready to calculate the direct estimates for each region and time period.

```
years <- levels(births$time)
print(years)
```

```
## [1] "85-89" "90-94" "95-99" "00-04"
## [5] "05-09" "10-14"
```

```
years <- c(years, "15-19")
u5m <- countrySummary(births = births, years = years,
    regionVar = "province", timeVar = "time",
    clusterVar = "~V021+idhog", ageVar = "age",
    weightsVar = "V005", geo.recode = NULL)
```

Visualize direct estimates



Example: Subnational Smoothed Estimates

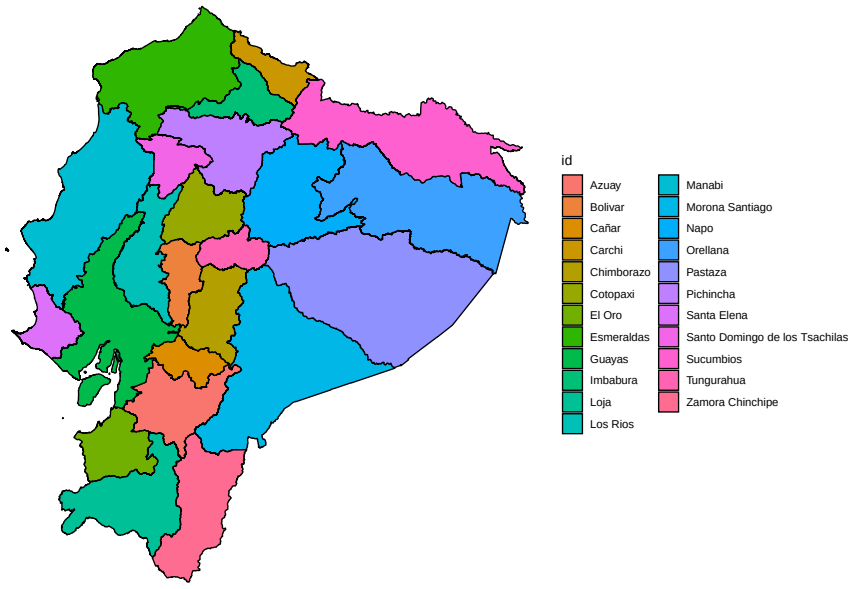
Getting to know the map

```
geo <- readOGR("gadm36_ECU_shp", layer = "gadm36_ECU_1",
    verbose = F)
geo <- geo[geo$NAME_1 != "Galápagos", ]
geo$NAME_1 <- factor(geo$NAME_1)
print(geo$NAME_1)
```

```
##  [1] Azuay
##  [2] Bolivar
##  [3] Cañar
##  [4] Carchi
##  [5] Chimborazo
##  [6] Cotopaxi
##  [7] El Oro
##  [8] Esmeraldas
##  [9] Guayas
## [10] Imbabura
## [11] Loja
## [12] Los Rios
```

Getting to know the map

```
g <- ggplot(fortify(geo, region = "NAME_1"))
g <- g + geom_polygon(aes(x = long, y = lat,
    group = group, fill = id), color = "black")
g <- g + theme_void() + coord_map()
print(g)
```



View map as an adjacency matrix

```
nb.r <- poly2nb(geo, queen = F, row.names = geo$NAME_1)
mat <- nb2mat(nb.r, style = "B", zero.policy = TRUE)
regions <- colnames(mat) <- rownames(mat)
mat <- as.matrix(mat[1:dim(mat)[1], 1:dim(mat)[1]])
nreg <- length(regions)
```

View map as an adjacency matrix

```
mat[1:5, 1:5]
```

```
##                Azuay Bolivar Cañar Carchi
## Azuay              0       0     1       0
## Bolivar            0       0     0       0
## Cañar              1       0     0       0
## Carchi             0       0     0       0
## Chimborazo         1       1     1       0
##                Chimborazo
## Azuay                   1
## Bolivar                 1
## Cañar                   1
## Carchi                  0
## Chimborazo              0
```

Bayesian smoothing

We will use INLA for the main workhorse to compute the Bayesian smoothing model.

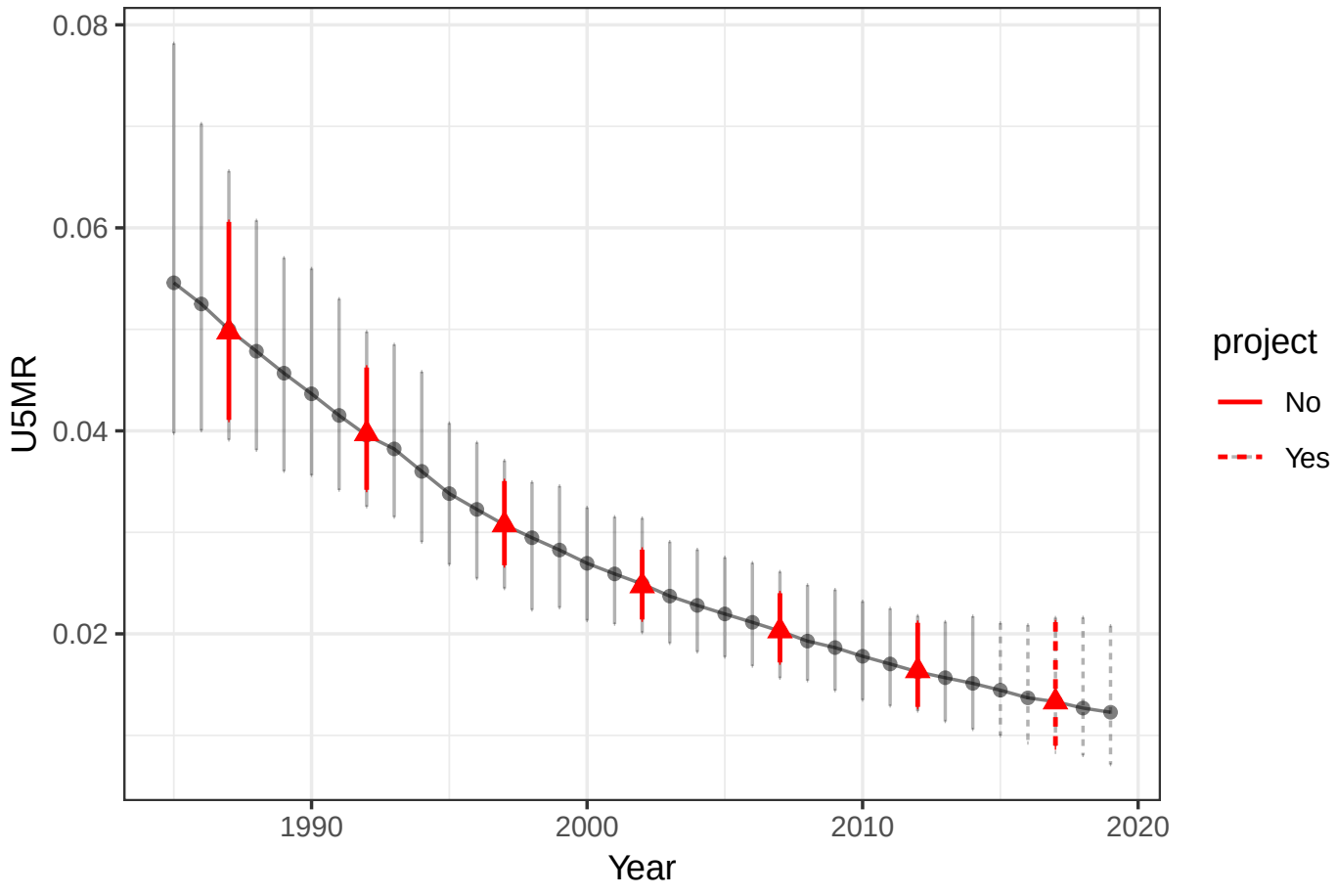
```
install.packages("INLA", repos = c(getOption("repos"),
    INLA = "https://inla.r-inla-download.org/R/stable"),
    dep = TRUE)
```

```
library(INLA)
```

Temporal smoothing of national estimates

```
years <- as.character(unique(u5m$years))
fit.national <- fitINLA(data = u5m, geo = NULL,
                        Amat = NULL, year_names = years,
                        year_range = c(1985, 2019), rw = 2,
                        is.yearly=TRUE, m = 5)
proj.national <- projINLA(fit.national, is.yearly = TRUE,
                          year_range = c(1985, 2019), year_label = years)
plot(proj.national, is.subnational = FALSE,
     is.yearly = TRUE, year_label = years,
     year_med = seq(1987, 2017, by=5),
     proj_year = 2015)
```

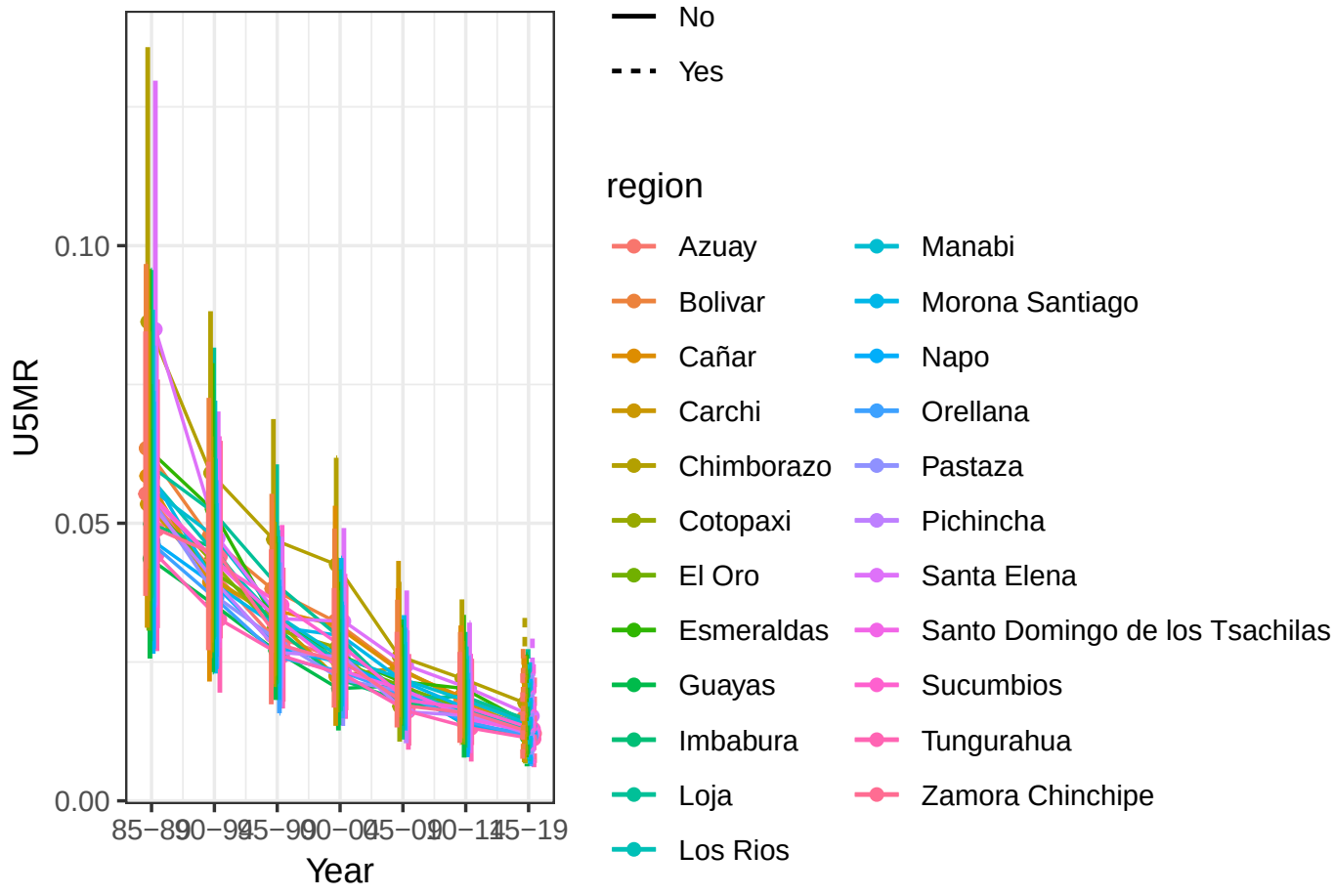
Temporal smoothing of national estimates



Spatial-temporal smoothing of subnational estimates

```
fit <- fitINLA(data = u5m, geo = geo, Amat = mat,
    year_names = years, year_range = c(1985,
        2019), rw = 2, is.yearly = FALSE,
    m = 5)
proj <- projINLA(fit, is.yearly = FALSE,
    year_range = c(1985, 2019), year_label = years,
    Amat = mat)
g <- plot(proj, is.yearly = FALSE, is.subnational = TRUE)
g
```

Spatial-temporal smoothing of subnational estimates



Spatial-temporal smoothing of subnational estimates

```
head(proj, n=2)
```

```
##    region years logit.q975 logit.q025
## 1   Azuay 85-89  -2.386433  -3.277468
## 2 Bolivar 85-89  -2.229712  -3.150055
##   logit.med        q975       q025
## 1 -2.846857 0.08455655 0.0369093
## 2 -2.695450 0.09669245 0.0418097
##          med is.yearly Year.num
## 1 0.05528506     FALSE       NA
## 2 0.06349848     FALSE       NA
```

Spatial-temporal smoothing of subnational estimates

```
tail(proj, n = 2)
```

```
##                region years logit.q975
## 160       Tungurahua 15-19  -3.861206
## 161 Zamora Chinchipe 15-19  -3.744801
##     logit.q025 logit.med       q975
## 160  -5.161025 -4.492337 0.02107856
## 161  -4.974727 -4.401348 0.02220299
##              q025        med is.yearly
## 160 0.006052648 0.01114943     FALSE
## 161 0.006421347 0.01213466     FALSE
##     Year.num
## 160       NA
## 161       NA
```

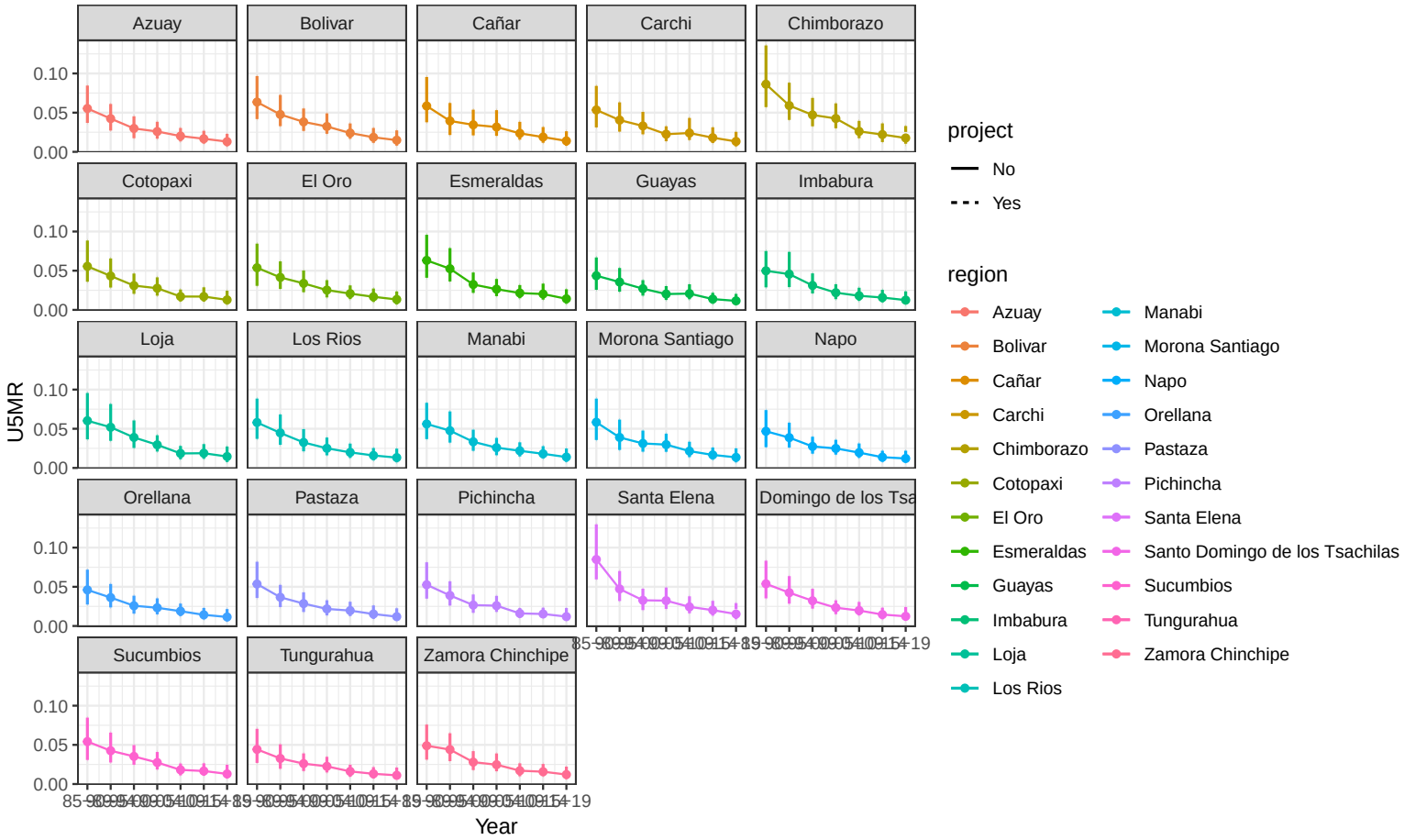
Table of estimates and 95% uncertainty intervals

```
estimates <- proj[, c("region", "years",
    "med", "q025", "q975")]
head(estimates)
```

```
##       region years        med
## 1      Azuay 85-89 0.05528506
## 2    Bolivar 85-89 0.06349848
## 3      Cañar 85-89 0.05851979
## 4     Carchi 85-89 0.05343966
## 5 Chimborazo 85-89 0.08627933
## 6   Cotopaxi 85-89 0.05558013
##          q025       q975
## 1 0.03690930 0.08455655
## 2 0.04180970 0.09669245
## 3 0.03771657 0.09547393
## 4 0.03117528 0.08412454
## 5 0.05707986 0.13575267
## 6 0.03594549 0.08851390
```

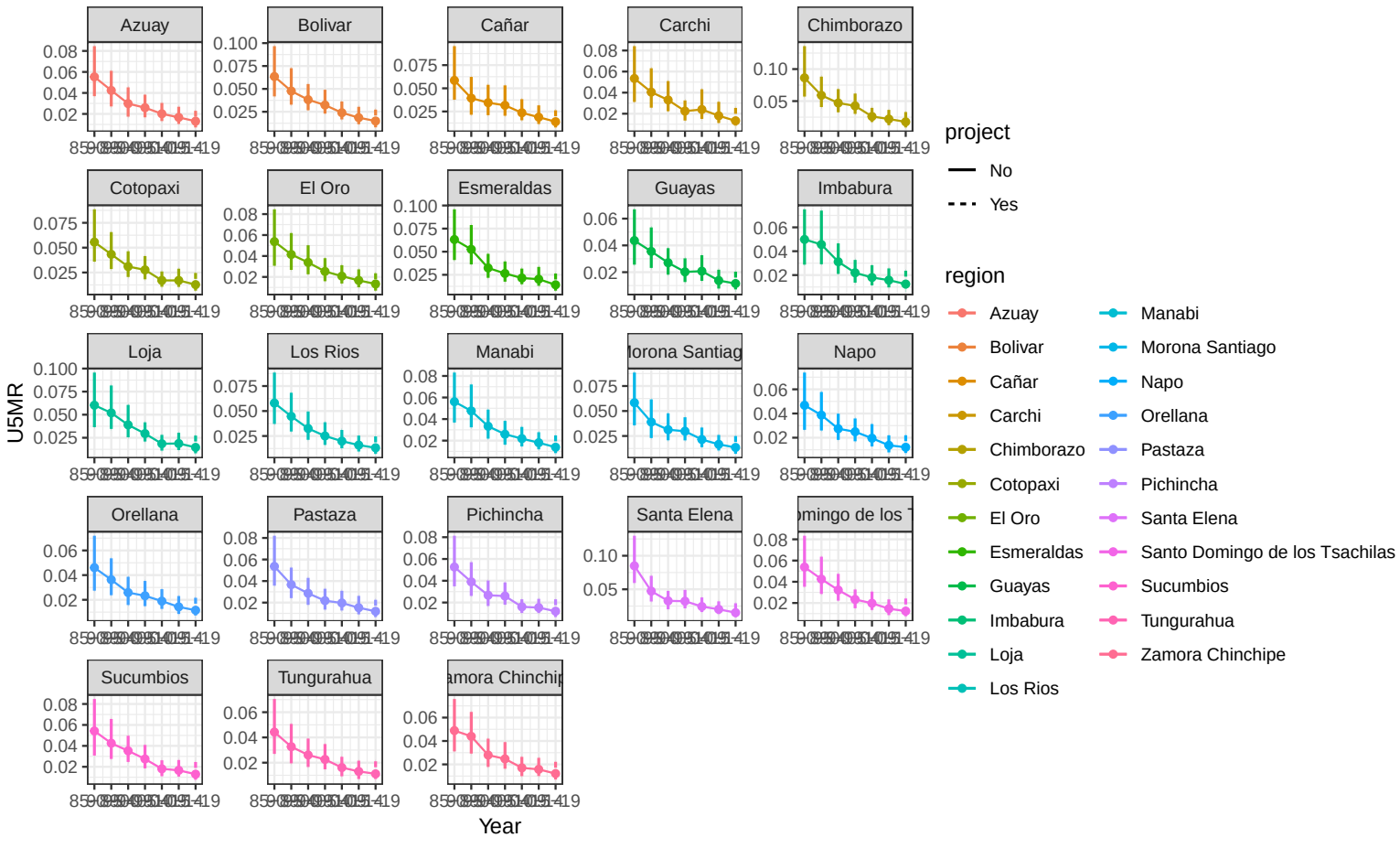
Spatial-temporal smoothing of subnational estimates

```
g + facet_wrap(~region)
```



Spatial-temporal smoothing of subnational estimates

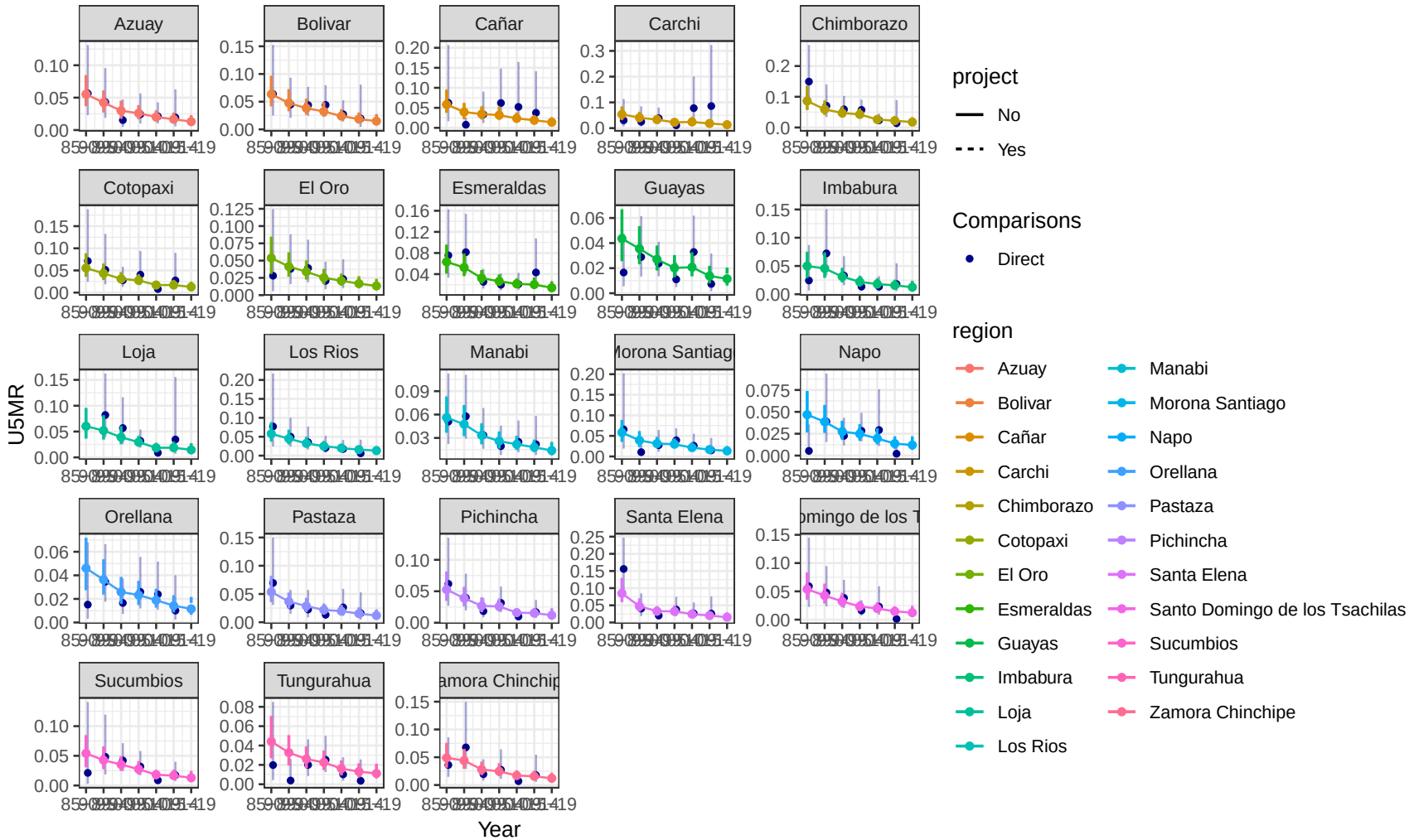
```
g + facet_wrap(~region, scales = "free")
```



Add back the direct estimates for comparison

```
g <- plot(proj, is.yearly = FALSE, is.subnational = TRUE,
    year_label = years, year_med = seq(1987,
        2017, by = 5), proj_year = 2015,
    data.add = u5m, option.add = list(point = "u5m",
        by = "survey", lower = "lower", upper = "upper"),
    color.add = "darkblue")
g + facet_wrap(~region, scales = "free")
```

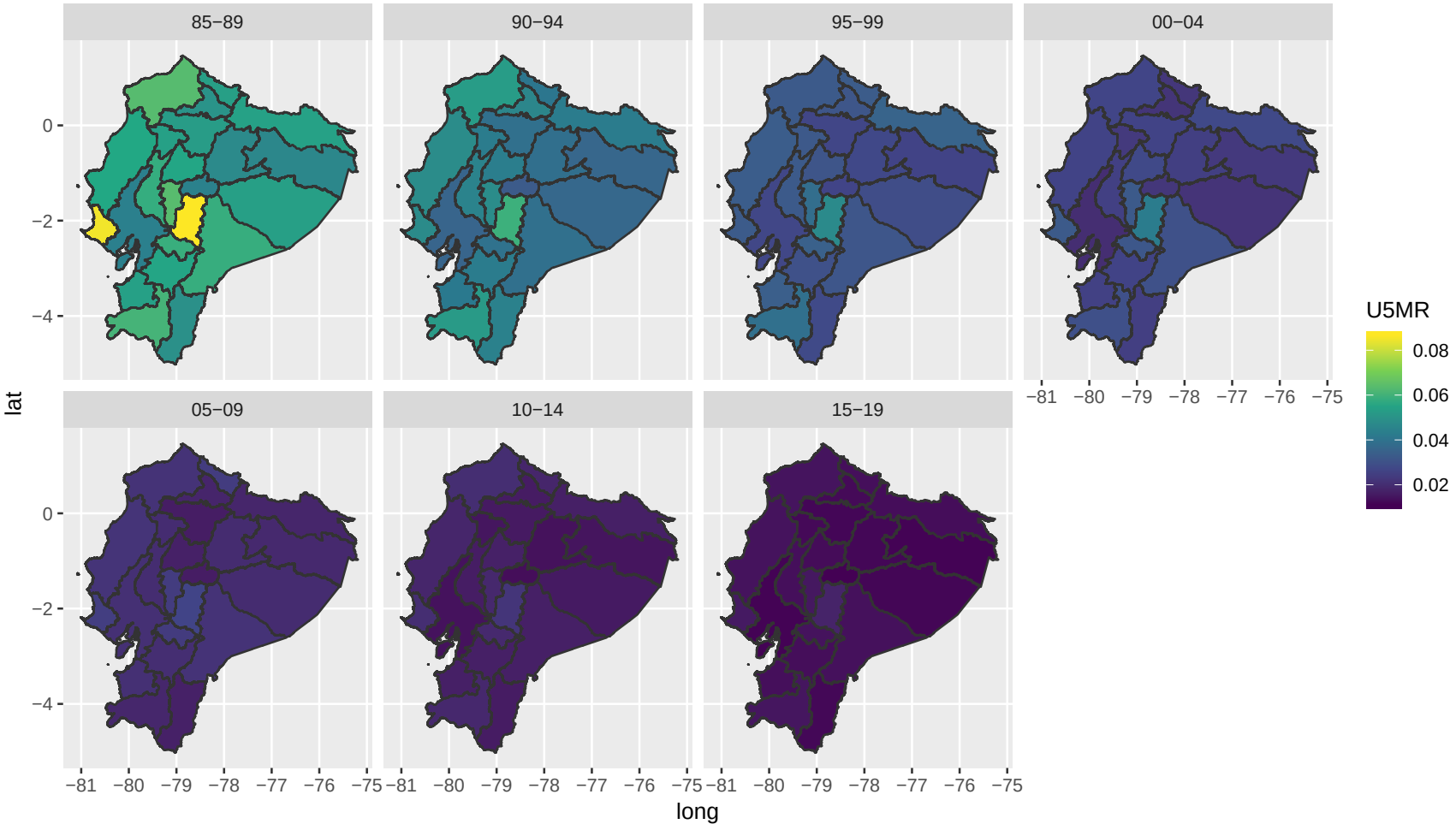
Add back the direct estimates for comparison



Visualize on the map

```
g <- mapPlot(data = subset(proj, is.yearly ==
    F), geo = geo, variables = c("years"),
    values = c("med"), by.data = "region",
    by.geo = "NAME_1", is.long = TRUE, ncol = 4)
g <- g + coord_map() + scale_fill_viridis_c("U5MR")
g
```

Visualize on the map



Appendix

Subpopulation (Males) - Person-months

```
dat_male <- fbh[fbh$B4 == "Male", ]
births_male <- getBirths(data = dat_male,
    surveyyear = 2012, variables = c("idhog",
        "province", "V005", "V021", "V024",
        "V025"), strata = "V024", dob = "B3",
    alive = "alive", age = "B7", date.interview = "V008",
    year.cut = seq(1985, 2015, by = 5))
```

Subpopulation (Males) - Direct estimates

```
u5m_male <- countrySummary(births = births_male,
    years = years, regionVar = "province",
    timeVar = "time", clusterVar = "~V021+idhog",
    ageVar = "age", weightsVar = "V005",
    geo.recode = NULL)
```

Infant Mortality - Person-months

We can change the discrete survival model into having only two bins: $[0, 1)$ and $[1, 12)$.

```
births_infant <- getBirths(data = fbh, surveyyear = 2012,
    variables = c("idhog", "province", "V005",
        "V021", "V024", "V025"), strata = "V024",
    dob = "B3", alive = "alive", age = "B7",
    date.interview = "V008", year.cut = seq(1985,
        2015, by = 5), month.cut = c(1, 12))
```

Infant Mortality - Direct estimates

```
u1m <- countrySummary(births = births_infant,
    years = years, regionVar = "province",
    timeVar = "time", clusterVar = "~V021+idhog",
    ageVar = "age", weightsVar = "V005",
    geo.recode = NULL)
colnames(u1m)[4] <- "IMR"
```

Neonatal Mortality - Person-months

We can change the discrete survival model into having only bin: $[0, 1)$.

```
births_neonates <- getBirths(data = fbh,
    surveyyear = 2012, variables = c("idhog",
        "province", "V005", "V021", "V024",
        "V025"), strata = "V024", dob = "B3",
    alive = "alive", age = "B7", date.interview = "V008",
    year.cut = seq(1985, 2015, by = 5), month.cut = c(1))
```

Neonatal Mortality - Direct estimates

```
nmr <- countrySummary(births = births_neonates,
    years = years, regionVar = "province",
    timeVar = "time", clusterVar = "~V021+idhog",
    ageVar = "age", weightsVar = "V005",
    geo.recode = NULL)
colnames(nmr)[4] <- "NMR"
```

Multiple Surveys - Setup

We will use the example dataset derived from the DHS model data to calculate U5MR as an example.

We will use the following commands to load the built-in dataset, and the following time periods.

```
data(DemoData)
years <- levels(DemoData[[1]]$time)
years.all <- c(years, "15-19")
```

Multiple Surveys - Person-months

```
data <- countrySummary_mult(births = DemoData,
    years = years, regionVar = "region",
    timeVar = "time", clusterVar = "~clustid+id",
    ageVar = "age", weightsVar = "weights",
    geo.recode = NULL)
head(data)
```

```
##   region years       u5m      lower
## 1    All 85-89 0.2373033 0.11954261
## 2    All 90-94 0.3250755 0.13332676
## 3    All 95-99 0.1484044 0.08851810
## 4    All 00-04 0.1410587 0.09294438
## 5    All 05-09 0.2319795 0.17516469
## 6    All 10-14 0.1639098 0.11747183
##       upper  logit.est    var.est
## 1 0.4162288 -1.1675213 0.17900781
## 2 0.6012732 -0.7305432 0.33909046
## 3 0.2382176 -1.7471707 0.08899408
```

Multiple Surveys - Map

```
data(DemoMap)
geo <- DemoMap$geo
mat <- DemoMap$Amat
```

Multiple Surveys - Combine

We have not seen this function in the example before. It basically calculates a weighted average for each area and time from the multiple surveys.

```
data_agg <- aggregateSurvey(data)
head(data_agg)
```

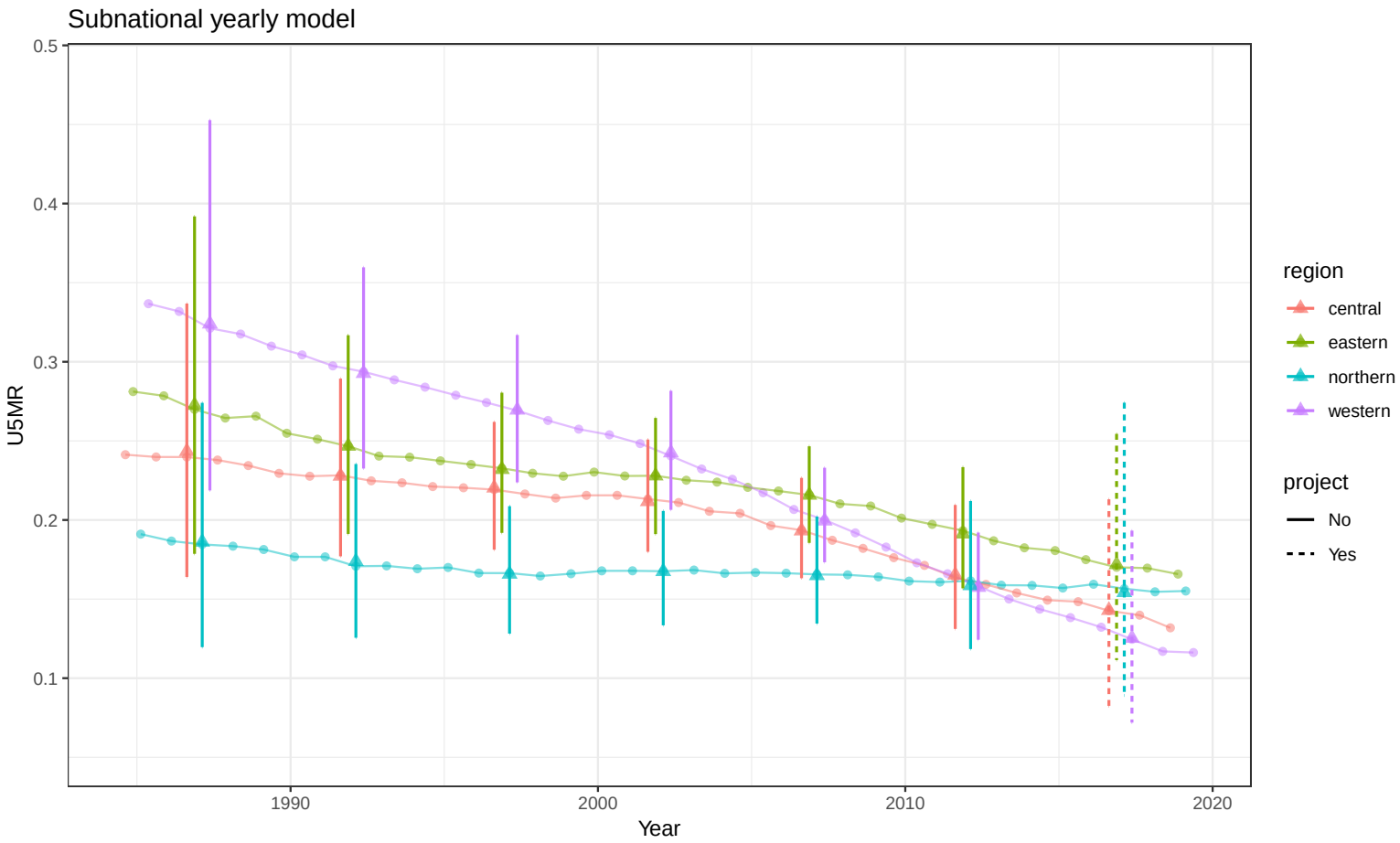
```
##   region years       u5m     lower
## 1    All 85-89 0.2410139 0.3267986
## 2    All 90-94 0.2082129 0.2729283
## 3    All 95-99 0.1929170 0.2338386
## 4    All 00-04 0.2169683 0.2519708
## 5    All 05-09 0.1890490 0.2158057
## 6    All 10-14 0.1589555 0.1876288
##       upper logit.est     var.est
## 1 0.1719946 -1.147129 0.046893454
## 2 0.1555594 -1.335731 0.032975917
## 3 0.1576828 -1.431166 0.015549751
## 4 0.1856215 -1.283422 0.009928410
```

Multiple Surveys - Subnational smoothed estimates

```
fit <- fitINLA(data = data_agg, geo = geo, Amat = mat,
               year_names = years.all,
               year_range = c(1985, 2019),
               rw = 2, is.yearly=TRUE, m = 5,
               type.st = 4)
out <- projINLA(fit, Amat = mat, is.yearly = TRUE)
```

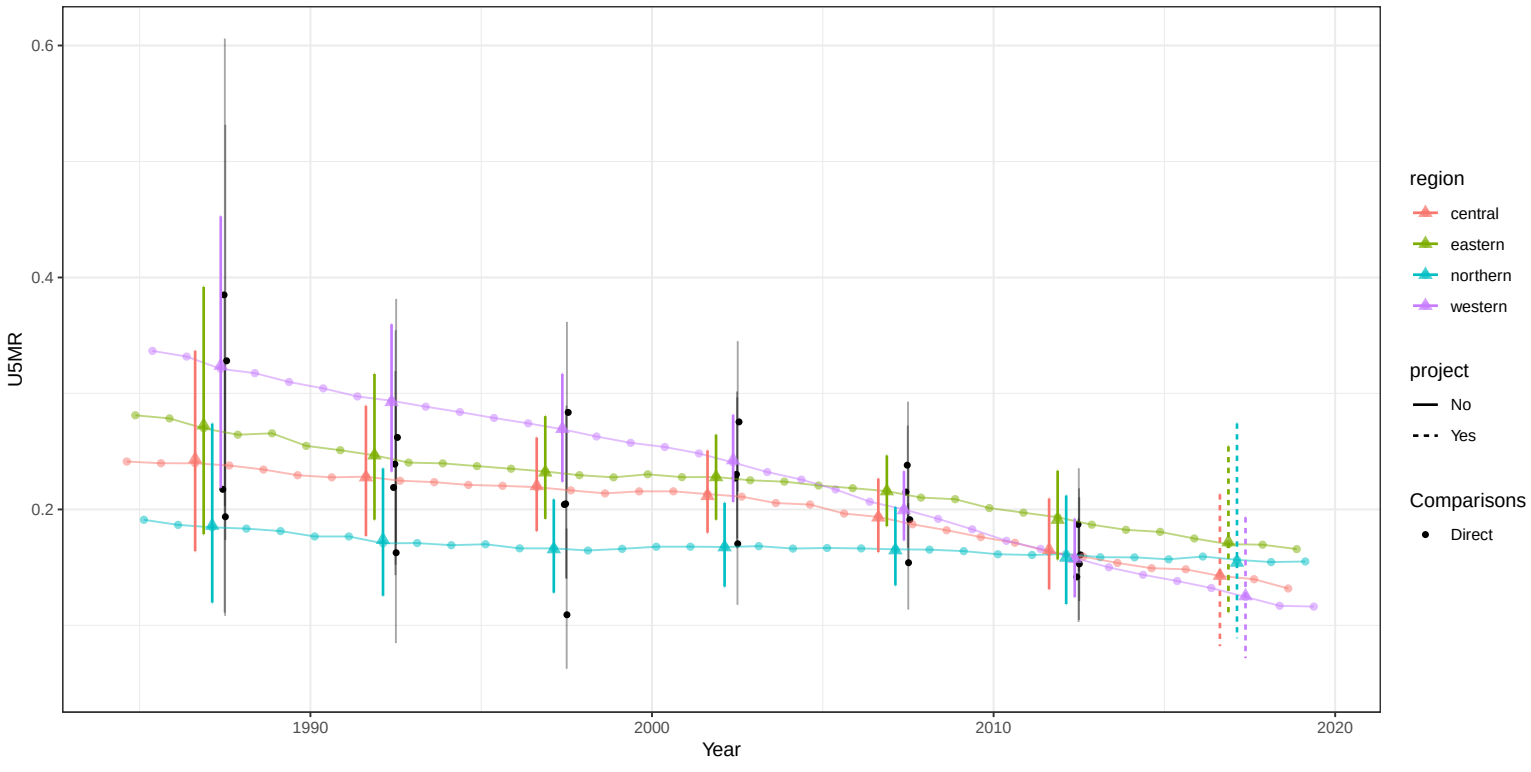
Multiple Surveys - Time-series plot

```
plot(out, is.yearly = TRUE, is.subnational = TRUE) +
    ggplot2::ggtitle("Subnational yearly model")
```



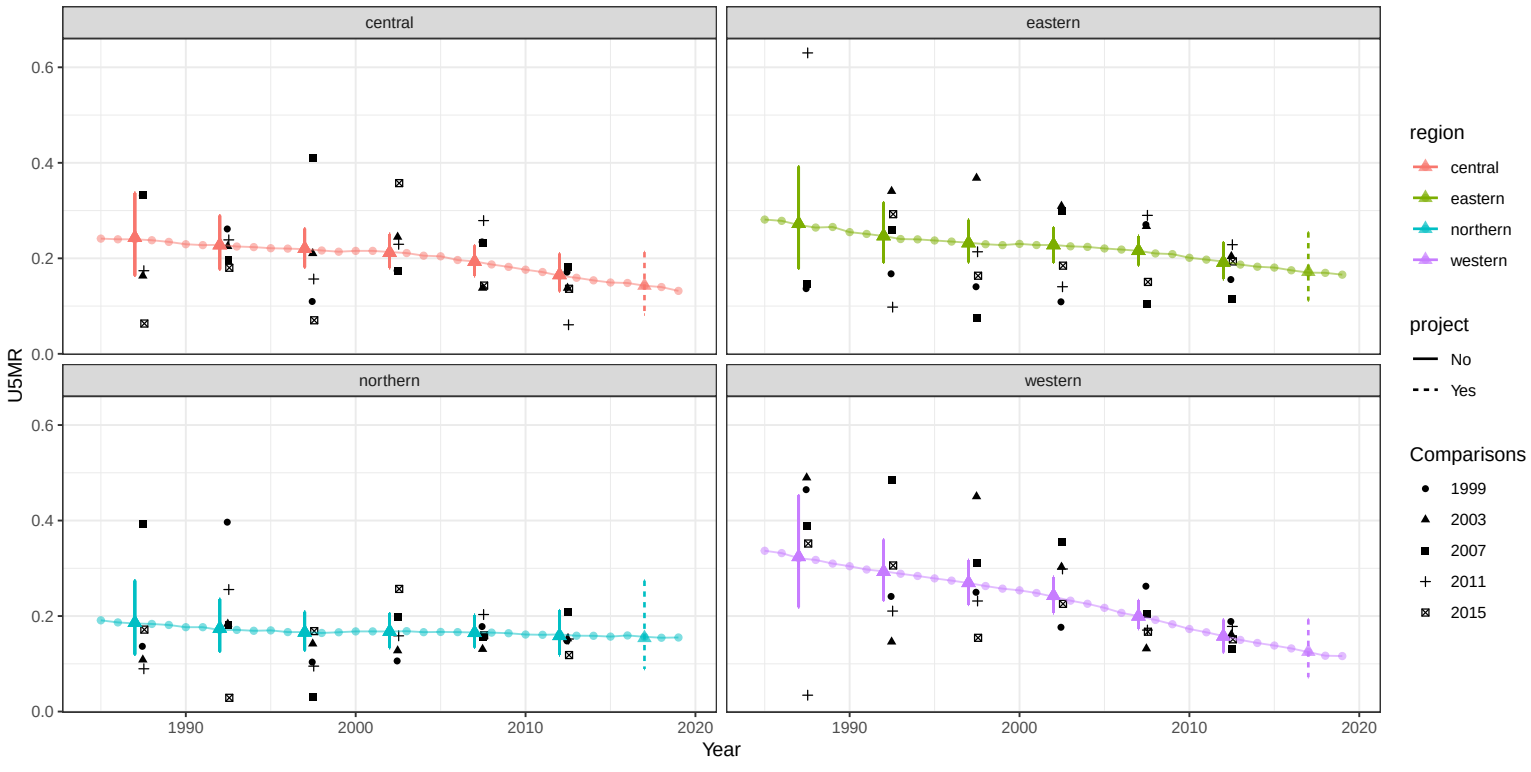
Multiple Surveys - Time-series plot

```
plot(out, is.yearly = TRUE, is.subnational = TRUE,
     data.add = data_agg, option.add = list(point = "u5m",
          lower = "lower", upper = "upper"))
```



Multiple Surveys - Time-series plot

```
plot(out, is.yearly = TRUE, is.subnational = TRUE,
     data.add = data, option.add = list(point = "u5m",
          by = "surveyYears")) + facet_wrap(~region)
```



Multiple Surveys - Map plot

```
mapPlot(data = subset(out, is.yearly == F),
     geo = DemoMap$geo, variables = c("years"),
     values = c("med"), by.data = "region",
     by.geo = "NAME_final", is.long = TRUE,
     ncol = 4) + coord_map()
```