Methods for Subnational Estimation of Child Mortality

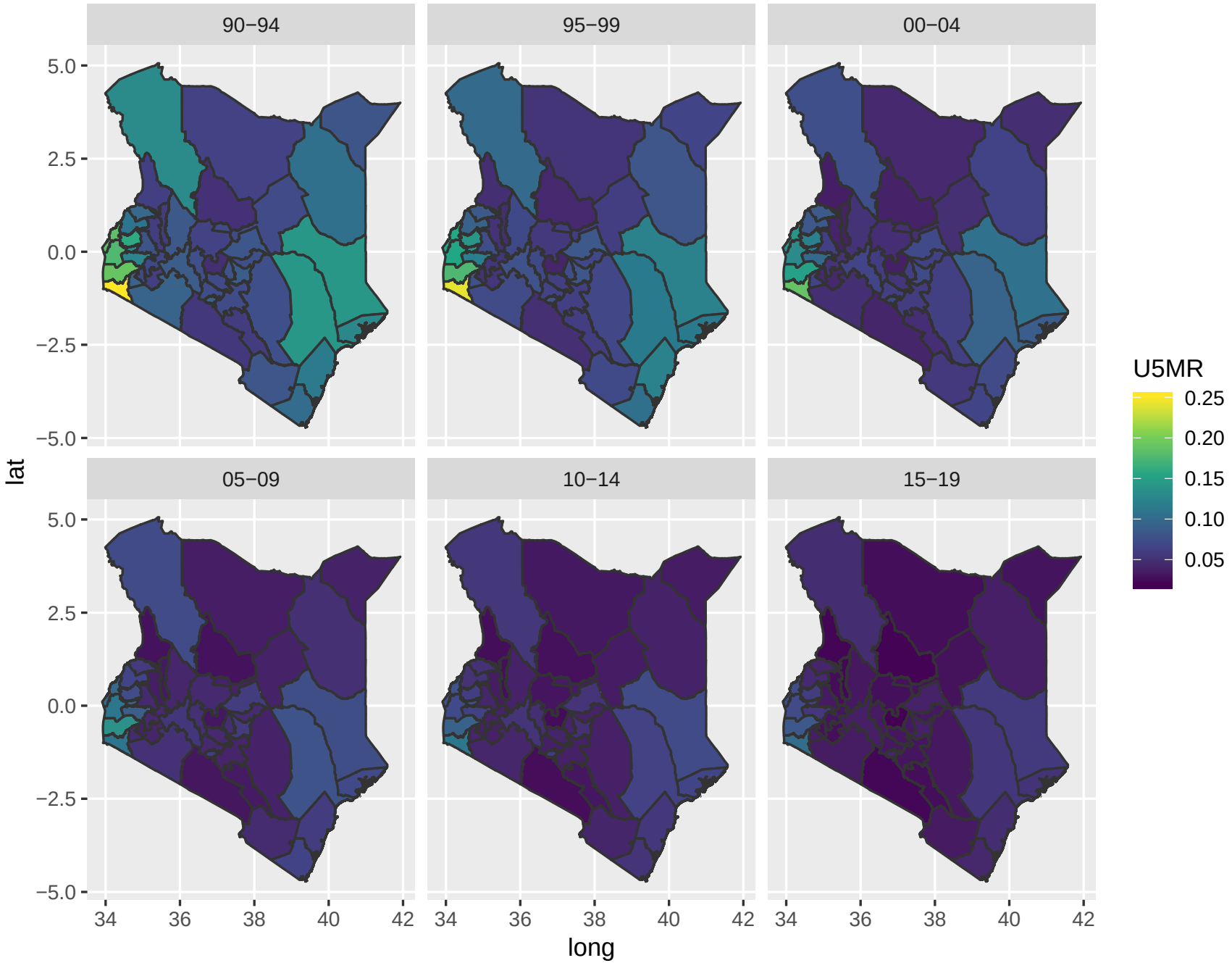
Hands-on Demo

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

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

Introduction to the Package

Overview


90–94
95–99
00–04
05–09
10–14
15–19
lat
long
5.0
2.5
0.0
−2.5
−5.0
34
36
38
40
42
U5MR
0.25
0.20
0.15
0.10
0.05

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	

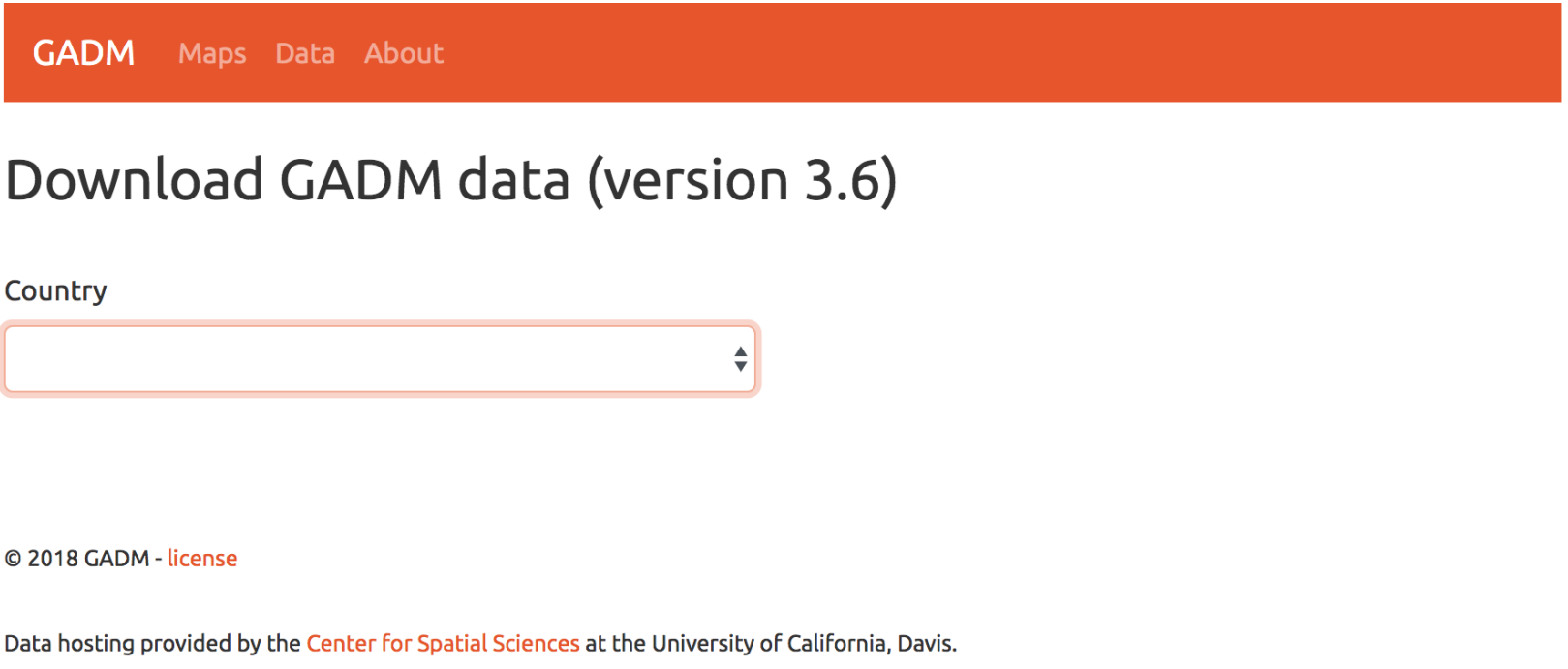
`fitINLA()`
`projINLA()`

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

Kenya

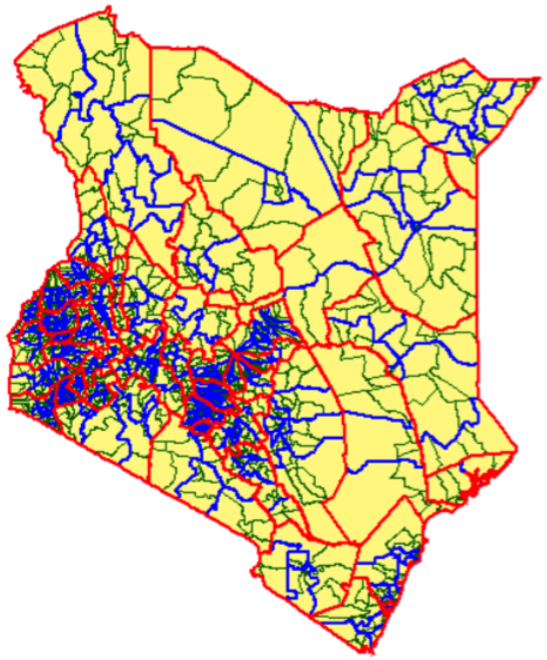
Geopackage

Shapefile

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

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

KMZ: level-0, level1, level2, level3



The coordinate reference system is longitude/latitude and the WGS84 datum.

Description of file formats.

The Kenya shapefiles are also available at
http://faculty.washington.edu/jonno/UNICEF-WORKSHOPS/gadm36_KEN_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)
library(INLA)
```

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:

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

Example: Subnational Direct Estimates

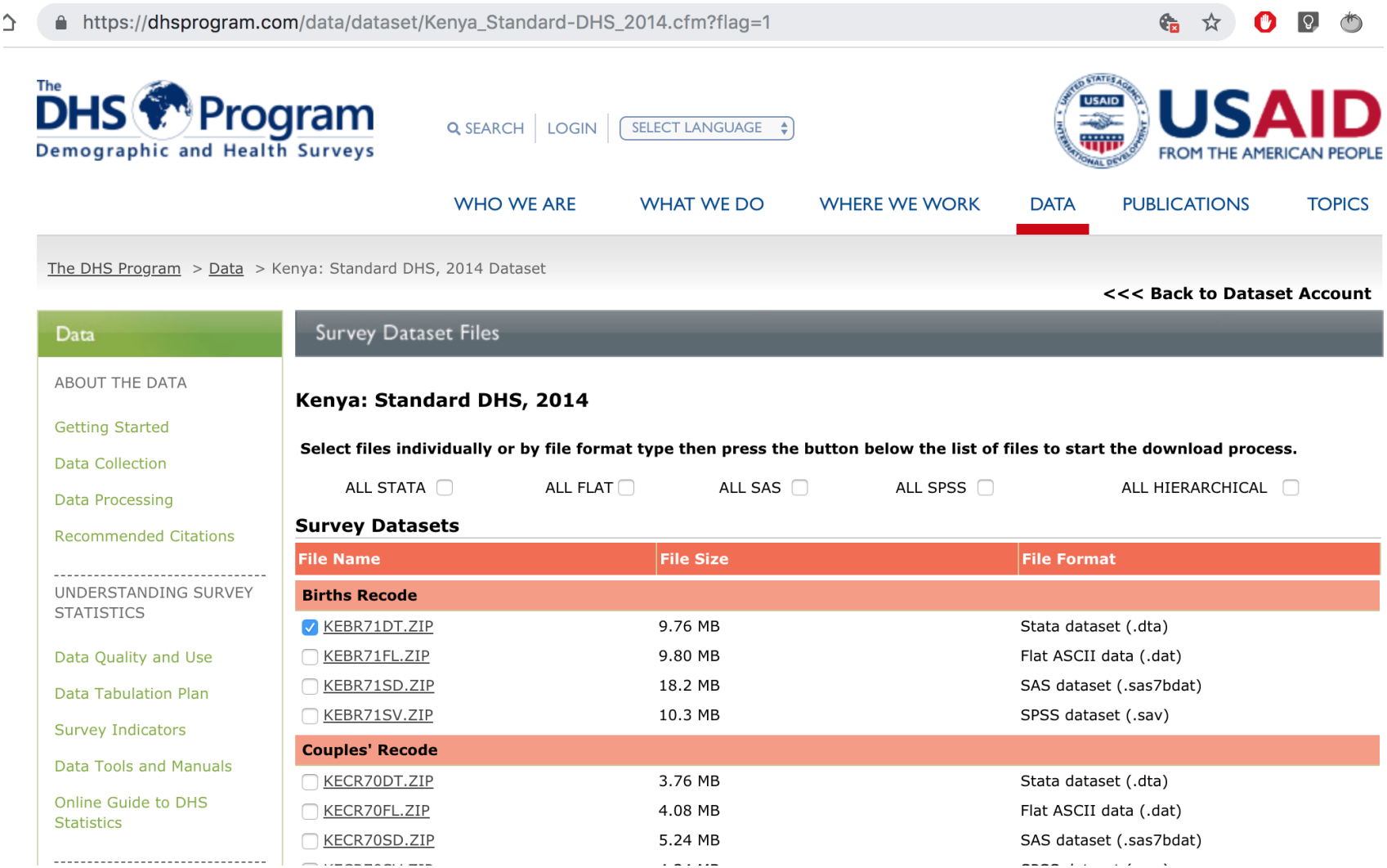
Preparing data

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 Kenya 2014 DHS data as an example.



Reading in the data and reorganizing into person-month format

```
filename <- "Data/KE_2014_DHS_03292019_1342_132518/KEBR71DT/KEBR71FL.DTA"
births <- getBirths(filepath = filename,
    surveyyear = 2014, strata = "v023", year.cut = seq(1990,
        2015, by = 5))
```

- v001: cluster ID
- v002: household ID
- v005: weight
- v008: interview century month code (CMC)
- b3: dob CMC
- b7: age at death (months)
- v023: strata
- v025: urban/rural

Add in county of residence

```
Kenya_pts <- readOGR("~/Dropbox/kenya14_ge/")
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/kwilson/Dropbox/kenya14_ge", layer: "KE
## with 1594 features
## It has 20 fields
```

```
births$DHSCLUST <- births$v001
births <- merge(births, Kenya_pts)
```

Getting to know person-month format

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

```
##   DHSCLUST  dob survey_year died id.new
## 1        1 1313        2014    0      1
## 2        1 1313        2014    0      1
##            caseid v001 v002 v004
## 1      0001006 02    1    6    1
## 2      0001006 02    1    6    1
##      v005
## 1 5476381
## 2 5476381
```

Getting to know person-month format

```
table(births$ADM1NAME, births$age)[1:5, ]
```

```
##
##                       0  1-11 12-23
##   Baringo          1442 14927 15359
##   Bomet            1877 19507 20459
##   Bungoma          2140 22289 22786
##   Busia            1886 19277 19378
##   Elgeyo Marakwet  1535 16176 16791
##
##                   24-35 36-47 48-59
##   Baringo         14475 13570 12818
##   Bomet           19404 18469 17526
##   Bungoma         21606 20419 19277
##   Busia           18109 16926 15910
##   Elgeyo Marakwet 15886 15099 14141
```

Getting to know person-month format

```
table(births$ADM1NAME, births$time)[1:5,
    ]
```

```
##
##                   90-94 95-99 00-04
##   Baringo          5390  9438 16033
##   Bomet            7898 14383 21784
##   Bungoma          9090 16569 25091
##   Busia            7950 12173 20370
##   Elgeyo Marakwet  5172 10917 19256
##
##                   05-09 10-14
##   Baringo         21935 19795
##   Bomet           29218 23959
##   Bungoma         31122 26645
##   Busia           27312 23681
##   Elgeyo Marakwet 24278 20005
```

Direct estimates

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

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

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

```
years <- c(years, "15-19")
u5m <- countrySummary(births = births, years = years,
    regionVar = "ADM1NAME", timeVar = "time",
    clusterVar = "~v001 + v002", ageVar = "age",
    weightsVar = "v005", geo.recode = NULL)
```

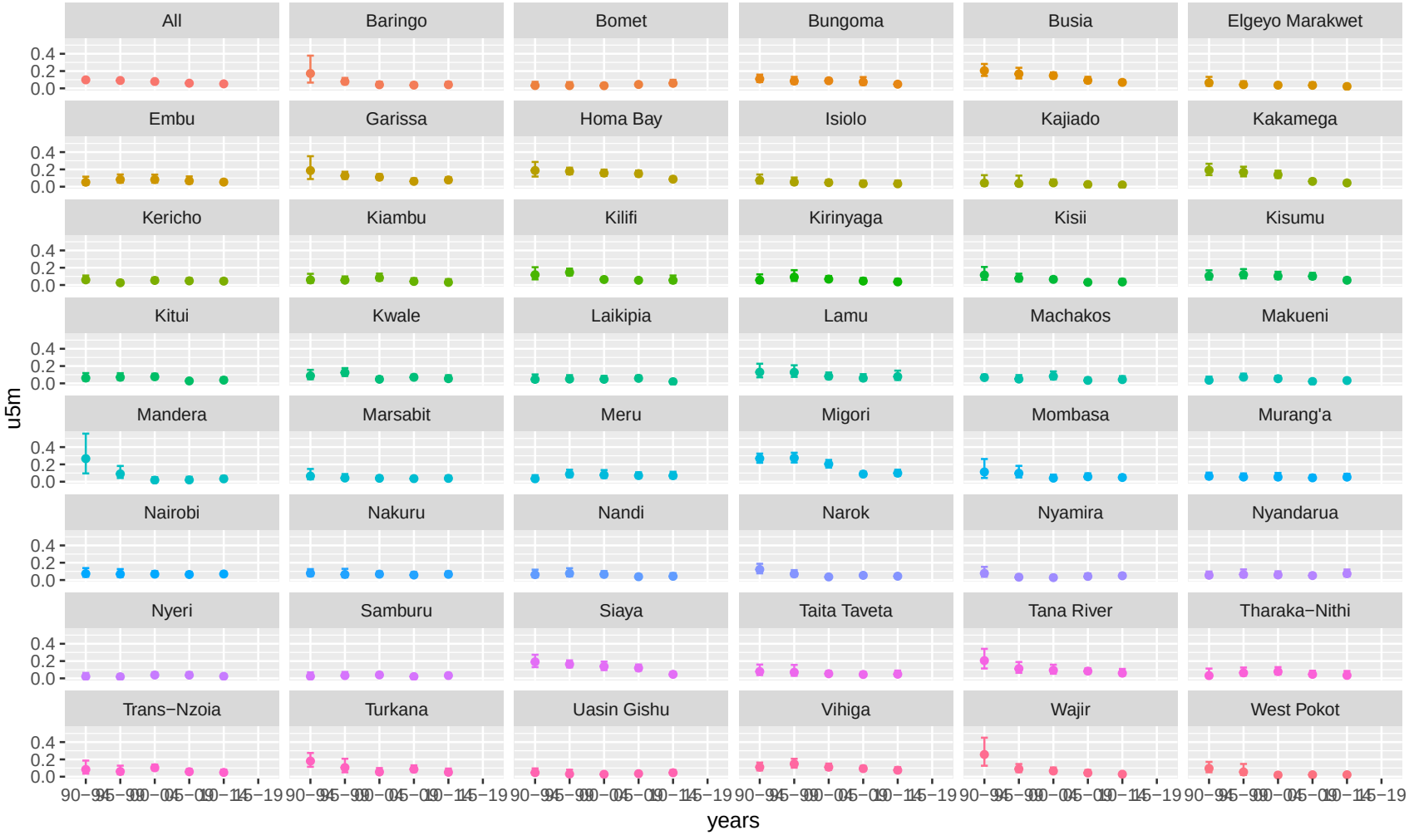
Read in Direct estimates

The direct estimates from the 2014 Kenya DHS are available at http://faculty.washington.edu/jonno/UNICEF-WORKSHOPS/direct-estimates-Kenya.csv

```
url <- "http://faculty.washington.edu/jonno/UNICEF-WORKSHOP
u5m <- read.csv(url)
head(u5m)
```

```
##   X region years        u5m      lower
## 1 1    All 90-94 0.09831390 0.08895894
## 2 2    All 95-99 0.09103547 0.08362311
## 3 3    All 00-04 0.07901263 0.07344286
## 4 4    All 05-09 0.05963244 0.05501516
## 5 5    All 10-14 0.05219889 0.04772870
## 6 6    All 15-19         NA         NA
##        upper logit.est     var.est
## 1 0.10853543 -2.216101 0.003167741
## 2 0.09903386 -2.301057 0.002253956
## 3 0.08496608 -2.455839 0.001629974
```

Visualize direct estimates



Example: Subnational Smoothed Estimates

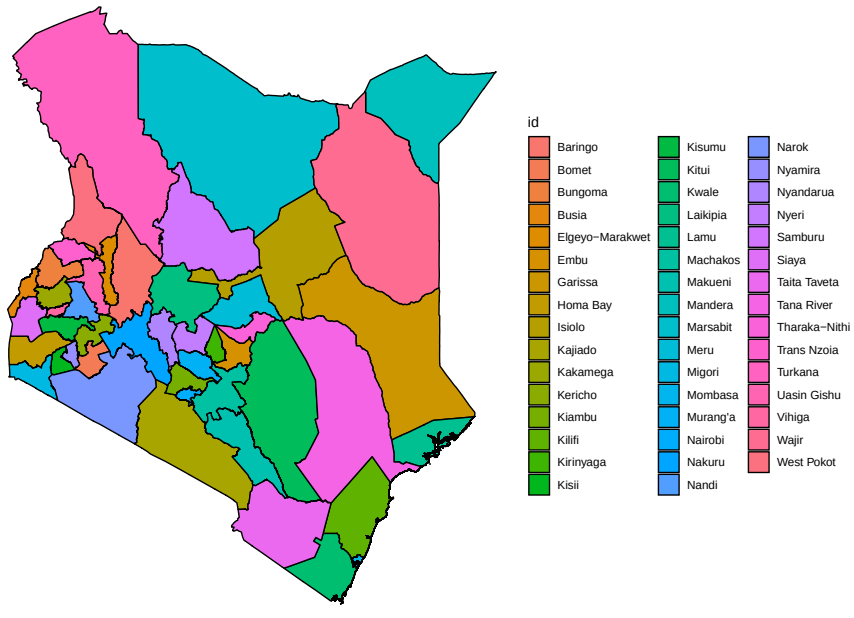
Getting to know the map

```
geo <- readOGR("Data/gadm36_KEN_shp", layer = "gadm36_KEN_1
    verbose = F)
print(geo$NAME_1)
```

```
##  [1] Baringo          Bomet
##  [3] Bungoma          Busia
##  [5] Elgeyo-Marakwet  Embu
##  [7] Garissa          Homa Bay
##  [9] Isiolo           Kajiado
## [11] Kakamega         Kericho
## [13] Kiambu           Kilifi
## [15] Kirinyaga        Kisii
## [17] Kisumu           Kitui
## [19] Kwale            Laikipia
## [21] Lamu             Machakos
## [23] Makueni          Mandera
## [25] Marsabit         Meru
## [27] Migori           Mombasa
```

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]
```

```
##                   Baringo Bomet Bungoma
## Baringo                 0     0        0
## Bomet                   0     0        0
## Bungoma                 0     0        0
## Busia                   0     0        1
## Elgeyo-Marakwet         1     0        0
##                   Busia Elgeyo-Marakwet
## Baringo               0                1
## Bomet                 0                0
## Bungoma               1                0
## Busia                 0                0
## Elgeyo-Marakwet       0                0
```

Bayesian smoothing

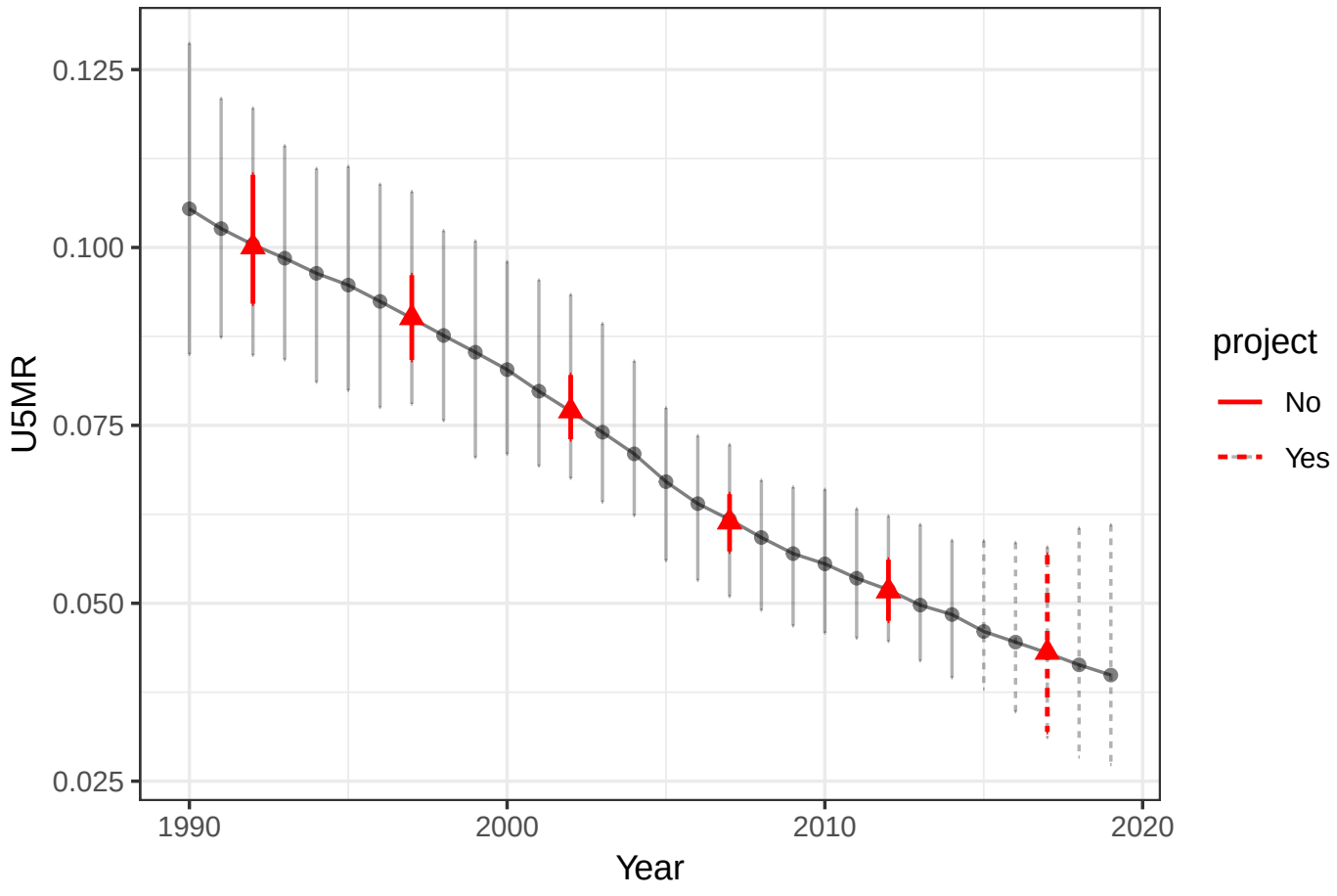
We will use INLA for the main workhorse to compute the Bayesian smoothing model.

```
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(1990, 2019), rw = 2,
                        is.yearly=TRUE, m = 5)
proj.national <- projINLA(fit.national, is.yearly = TRUE,
                          year_range = c(1990, 2019), year_label = years)
plot(proj.national, is.subnational = FALSE,
     is.yearly = TRUE, year_label = years,
     year_med = seq(1992, 2017, by=5),
     proj_year = 2015)
```

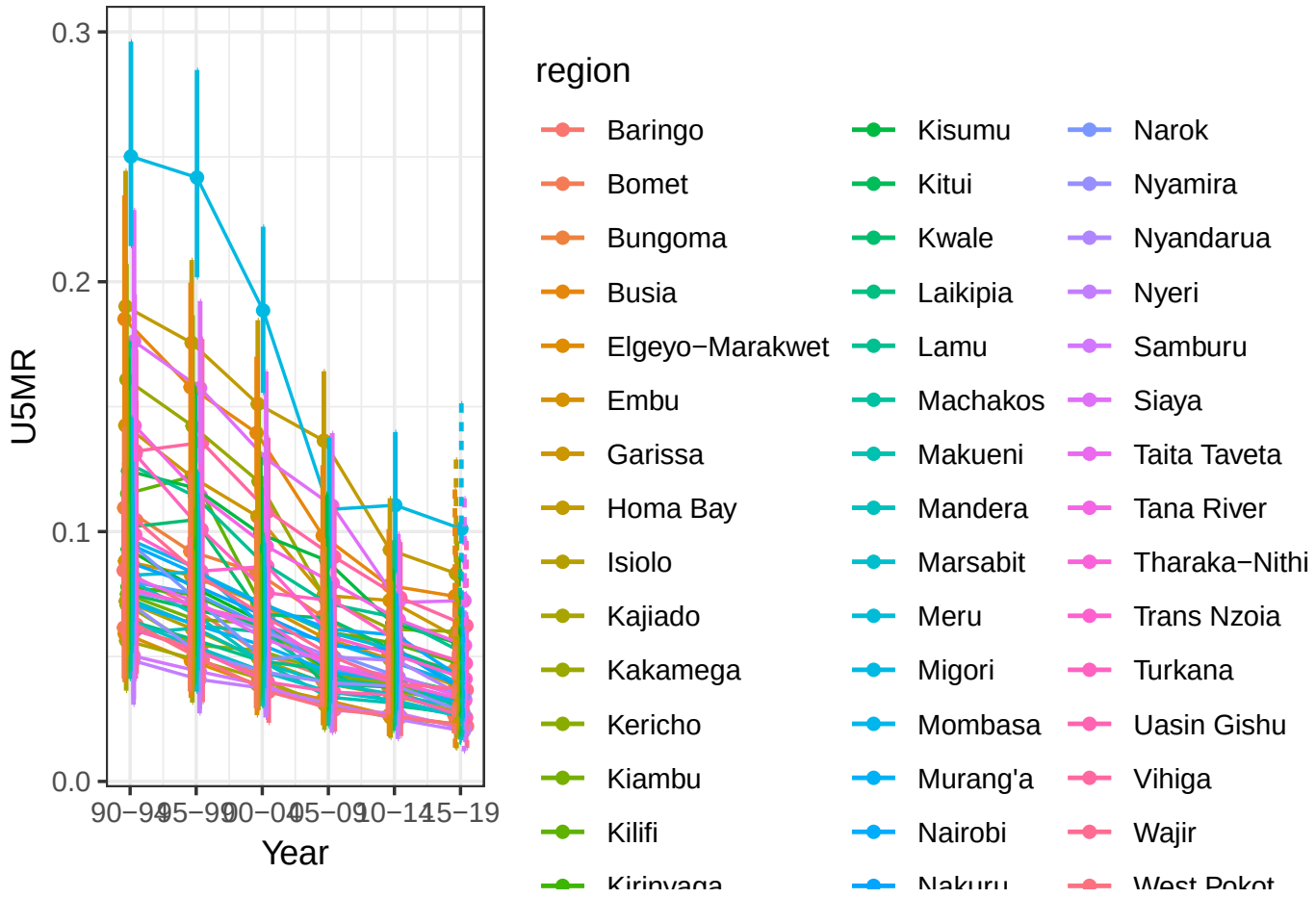
Temporal smoothing of national estimates



Spatial-temporal smoothing of subnational estimates

```
fit <- fitINLA(data = u5m, geo = geo, Amat = mat,
    year_names = years, year_range = c(1990,
        2019), rw = 2, is.yearly = FALSE,
    m = 5)
proj <- projINLA(fit, is.yearly = FALSE,
    year_range = c(1990, 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 Baringo 90-94  -1.917067  -2.803602
## 2   Bomet 90-94  -2.318369  -3.158403
##   logit.med       q975       q025
## 1 -2.390166 0.12256737 0.05740969
## 2 -2.737708 0.08718964 0.04111012
##          med is.yearly Year.num
## 1 0.08441738     FALSE       NA
## 2 0.06156300     FALSE       NA
```

Spatial-temporal smoothing of subnational estimates

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

```
##          region years logit.q975
## 281       Wajir 15-19  -2.748666
## 282 West Pokot 15-19  -3.209363
##     logit.q025 logit.med       q975
## 281  -3.775875 -3.261268 0.06252304
## 282  -4.320402 -3.789351 0.03892237
##           q025        med is.yearly
## 281 0.02307112 0.03667085     FALSE
## 282 0.01339442 0.02208990     FALSE
##     Year.num
## 281       NA
## 282       NA
```

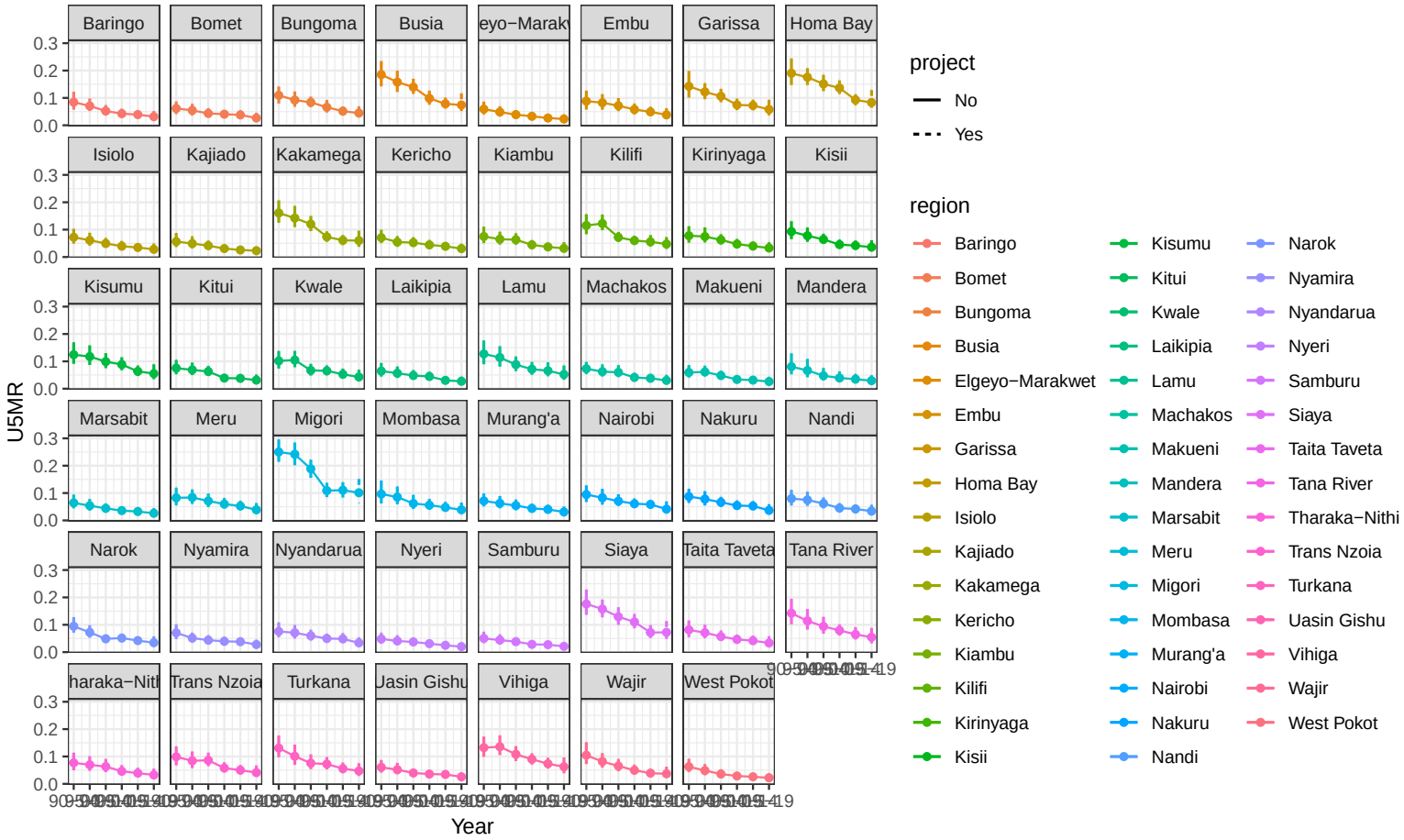
Table of estimates and 95% uncertainty intervals

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

```
##              region years        med
## 1           Baringo 90-94 0.08441738
## 2             Bomet 90-94 0.06156300
## 3           Bungoma 90-94 0.10951727
## 4             Busia 90-94 0.18505534
## 5 Elgeyo-Marakwet 90-94 0.05907401
## 6              Embu 90-94 0.08820910
##         q025       q975
## 1 0.05740969 0.12256737
## 2 0.04111012 0.08718964
## 3 0.07953978 0.14158177
## 4 0.14239367 0.23457546
## 5 0.03971313 0.08639814
## 6 0.05737825 0.12643837
```

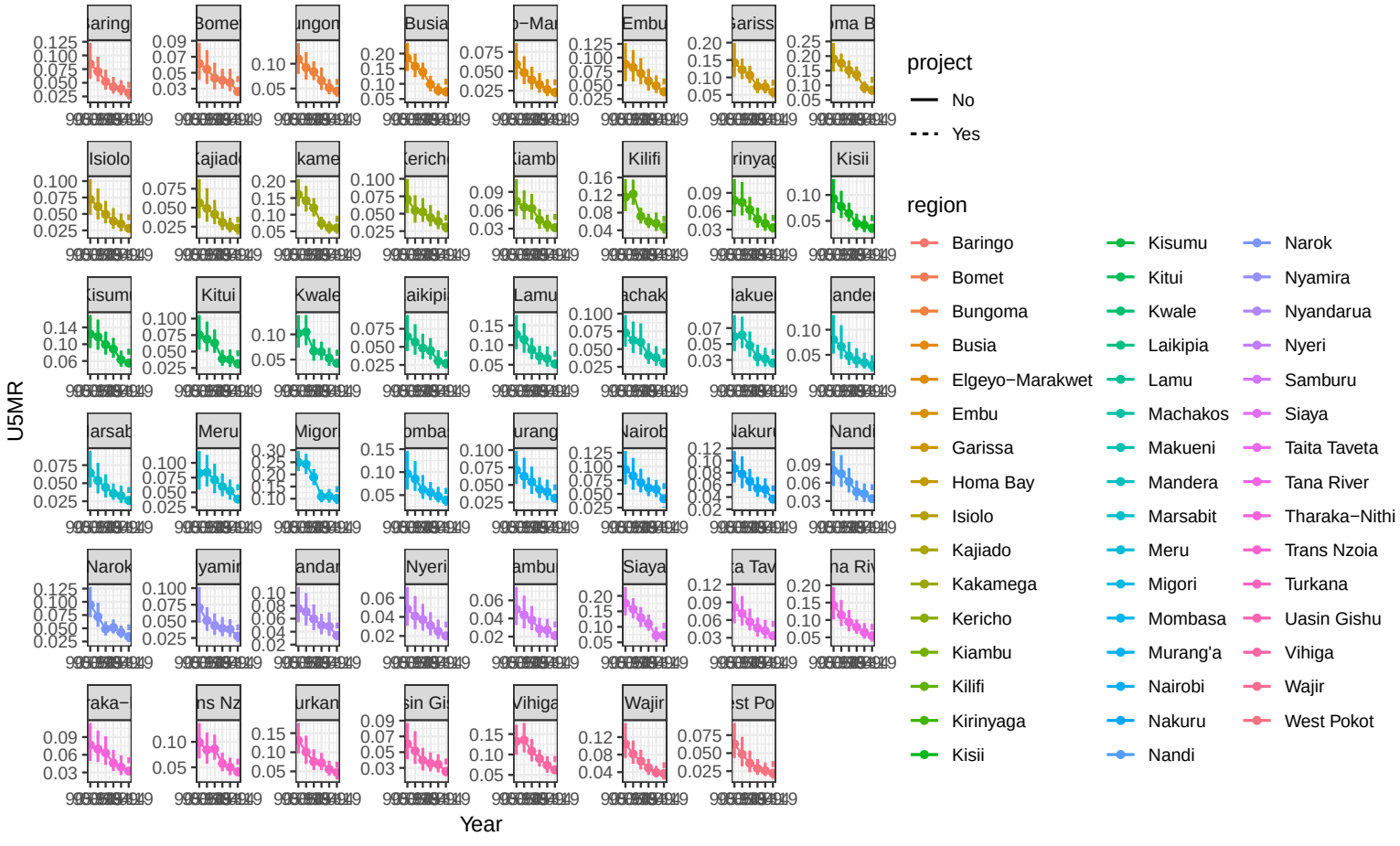
Spatial-temporal smoothing of subnational estimates

```
g + facet_wrap(~region, ncol = 8)
```



Spatial-temporal smoothing of subnational estimates

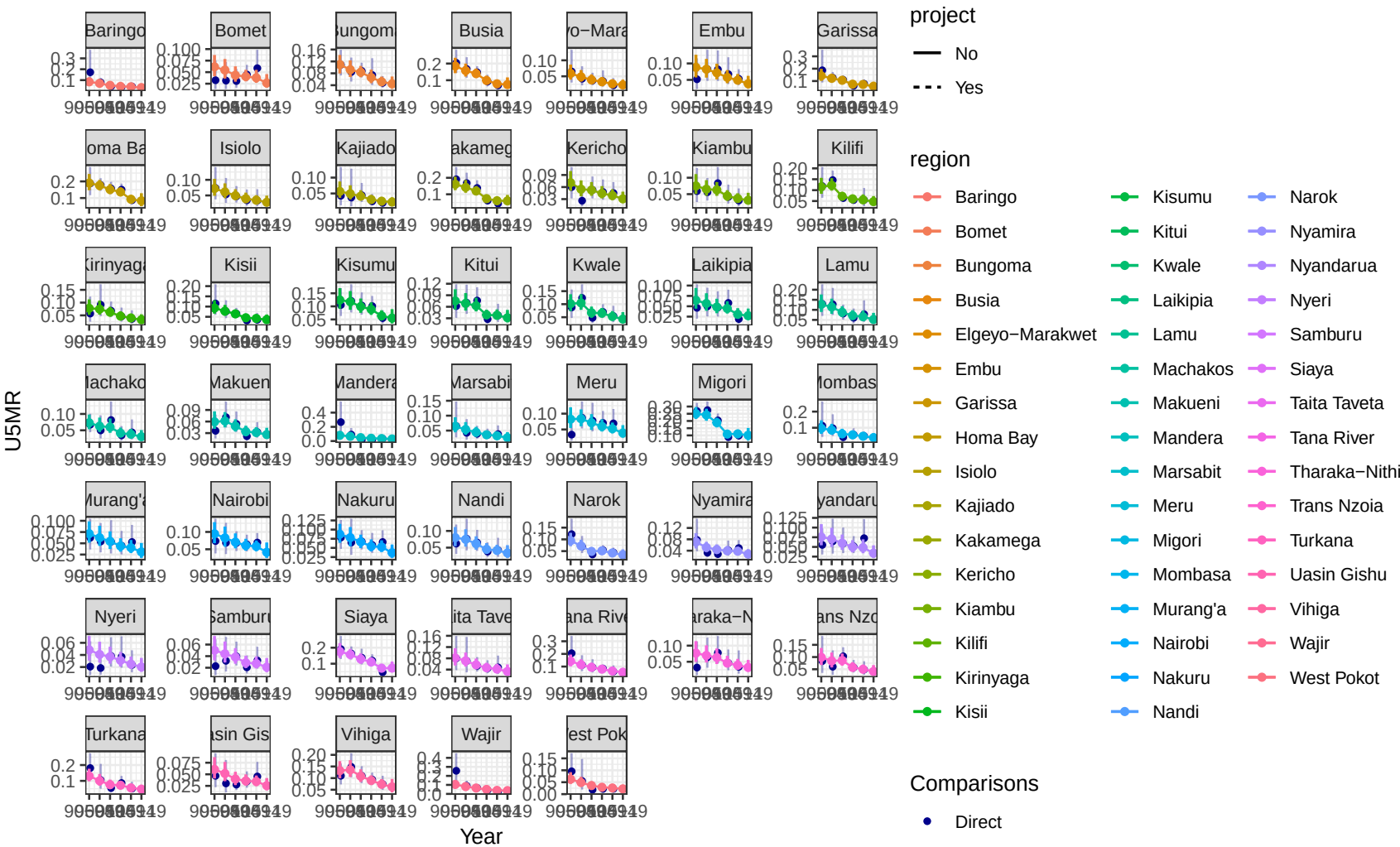
```
g + facet_wrap(~region, ncol = 8, scales = "free")
```



Add back the direct estimates for comparison

```
g <- plot(proj, is.yearly = FALSE, is.subnational = TRUE,
    year_label = years, year_med = seq(1992,
        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")
```

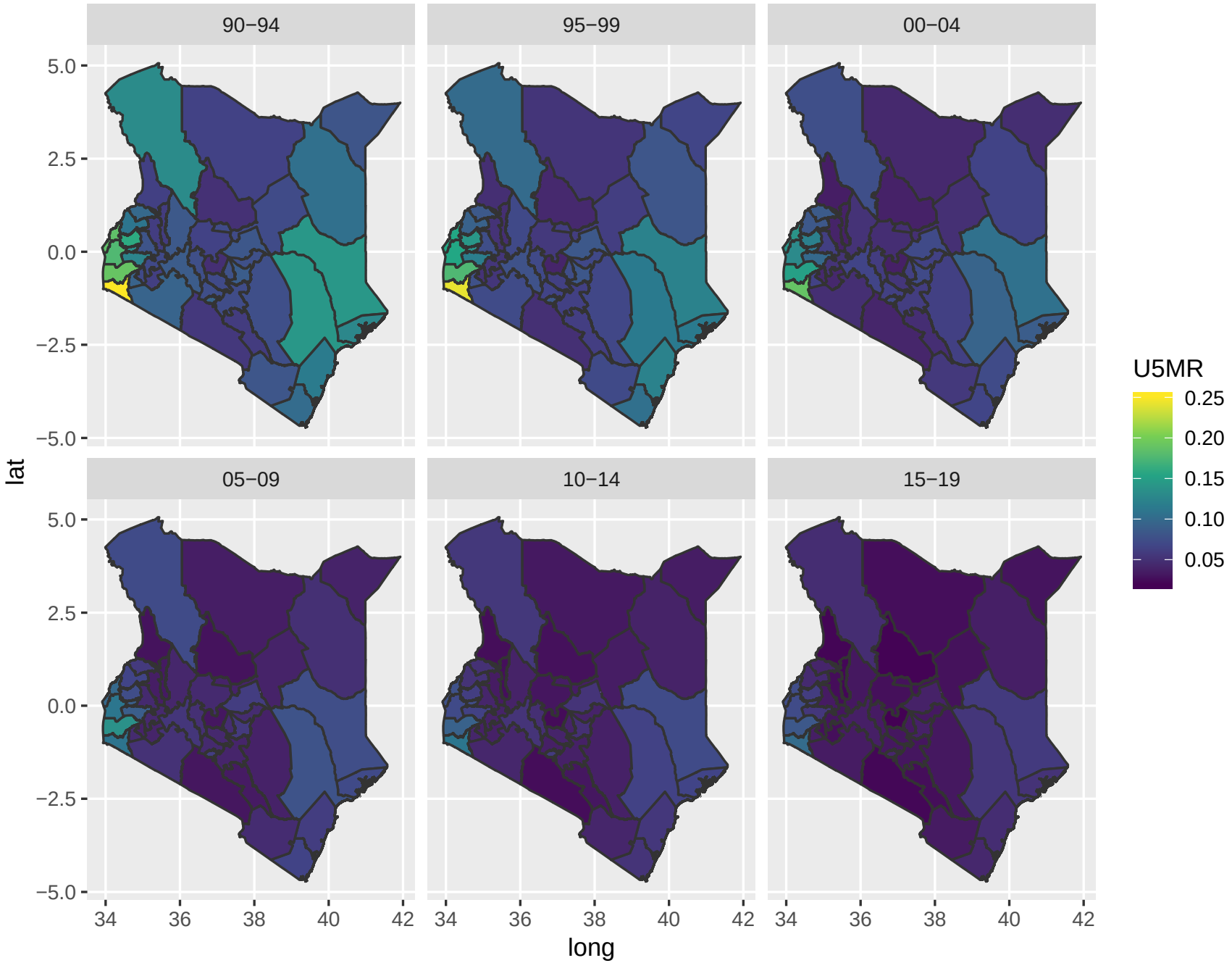
Add back the direct estimates for comparison



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 = 3)
g <- g + coord_map() + scale_fill_viridis_c("U5MR")
g
```

Visualize on the map



Appendix

HIV adjustment

```
url <- "http://faculty.washington.edu/jonno/UNICEF-WORKSHOP
HIVadj <- read.csv(url)
head(HIVadj)
```

```
##   province years_prior     c
## 1  central         0-4 0.977
## 2  central         1-5 0.970
## 3  central         2-6 0.959
## 4  central         3-7 0.946
## 5  central         4-8 0.931
## 6  central         5-9 0.913
```

HIV adjustment

An adjustment for HIV can be made by adjusting the direct estimates

```
adjust <- function(p, v, c) {
    # p: 5q0; v: logit variance; c:
    # adjustment factor
    f.prime <- 1 - (c - 1) * (p/(1 - p))/(c +
        (c - 1) * p/(1 - p))
    p <- p/c
    v <- v * f.prime^2
    return(c(p, log(p/(1 - p)), v, 1/v))
}
```

HIV adjustment

```
url <- "http://faculty.washington.edu/jonno/UNICEF-WORKSHOP
region_to_province <- read.csv(url)
u5m <- merge(u5m, region_to_province)
u5m$years_prior <- c("20-24", "15-19", "10-14",
    "5-9", "0-4", NA)
u5m.hiv <- merge(u5m, HIVadj, all.x = T)
u5madj <- matrix(adjust(u5m.hiv$u5m, u5m.hiv$var.est,
    u5m.hiv$c), ncol = 4)
u5m.hiv$u5m.unadj <- u5m.hiv$u5m
u5m.hiv$logit.est.unadj <- u5m.hiv$logit.est
u5m.hiv$var.est.unadj <- u5m.hiv$var.est
u5m.hiv$logit.prec.unadj <- u5m.hiv$logit.prec
u5m.hiv$u5m <- u5madj[, 1]
u5m.hiv$logit.est <- u5madj[, 2]
u5m.hiv$var.est <- u5madj[, 3]
u5m.hiv$logit.prec <- u5madj[, 4]
```

HIV adjustment

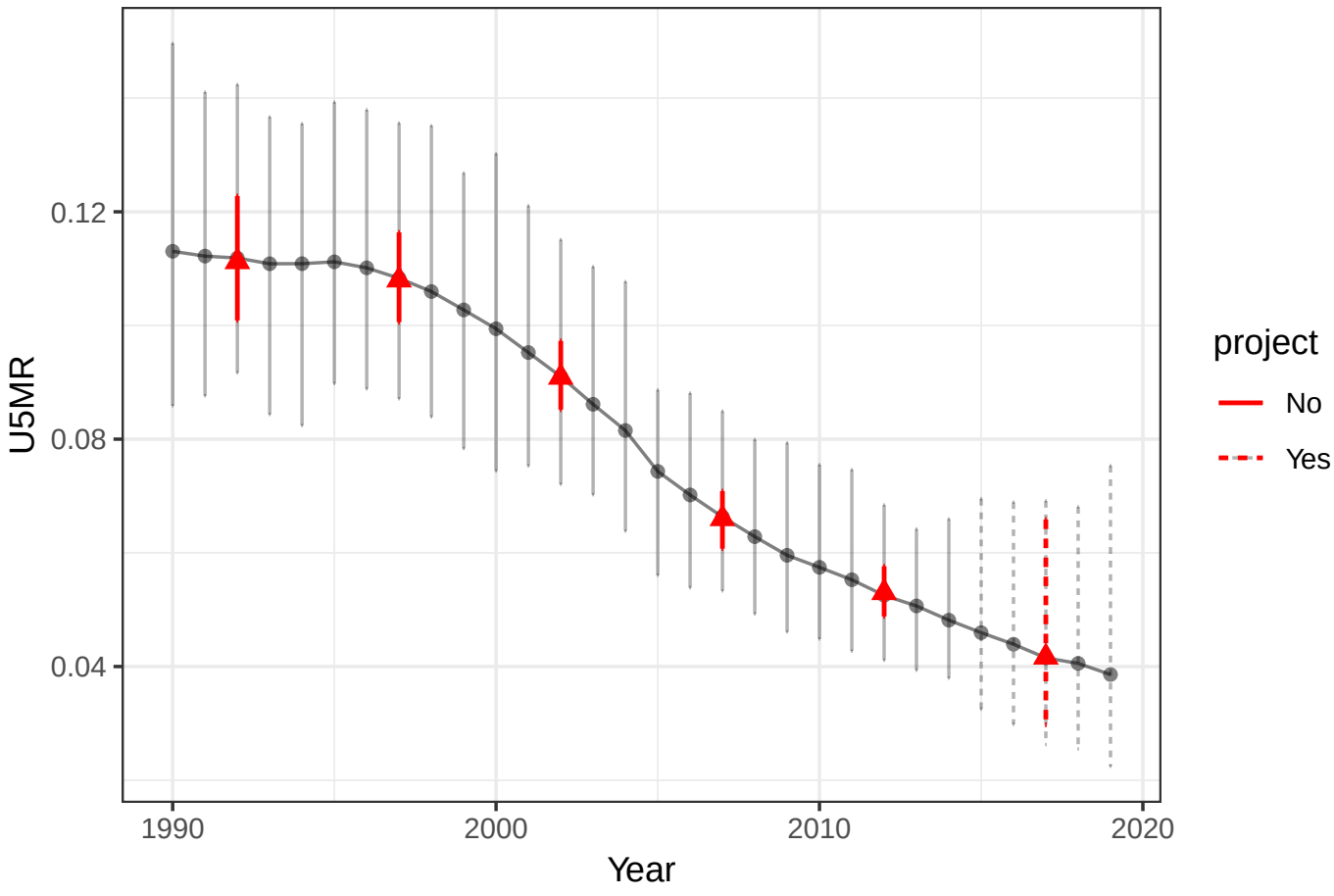
```
head(u5m.hiv)
```

```
##   province years_prior region X years
## 1      All         0-4    All 5 10-14
## 2      All       10-14    All 3 00-04
## 3      All       15-19    All 2 95-99
## 4      All       20-24    All 1 90-94
## 5      All         5-9    All 4 05-09
## 6      All        <NA>    All 6 15-19
##          u5m       lower       upper
## 1 0.05331858 0.04772870 0.05706267
## 2 0.09273783 0.07344286 0.08496608
## 3 0.10941764 0.08362311 0.09903386
## 4 0.10887475 0.08895894 0.10853543
## 5 0.06418993 0.05501516 0.06461075
## 6         NA          NA          NA
##   logit.est     var.est region_num
## 1 -2.876678 0.002317092          0
```

HIV adjustment

```
fit.national.adj <- fitINLA(data = u5m.hiv,
    geo = NULL, Amat = NULL, year_names = years,
    year_range = c(1990, 2019), rw = 2, is.yearly = TRUE,
    m = 5)
proj.national.adj <- projINLA(fit.national.adj,
    is.yearly = TRUE, year_range = c(1990,
        2019), year_label = years)
```

HIV adjustment



Subpopulation (Males) - Person-months

```
dat <- readstata13::read.dta13(filename,
    generate.factors = TRUE)
table(dat$b4)
```

```
##
##   male female
##  42337  41254
```

```
dat_male <- dat[dat$b4 == "male", ]
births_male <- getBirths(data = dat_male,
    surveyyear = 2014, strata = "v023", year.cut = seq(199
        2015, by = 5))
```

Subpopulation (Males) - Direct estimates

```
births_male$DHSCLUST <- births_male$v001
births_male <- merge(births_male, Kenya_pts)
```

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 = dat, surveyyear = 2014,
    strata = "v023", year.cut = seq(1990,
        2015, by = 5), month.cut = c(1, 12))
births_infant$DHSCLUST <- births_infant$v001
births_infant <- merge(births_infant, Kenya_pts)
```

Infant Mortality - Direct estimates

```
u1m <- countrySummary(births = births_infant,
    years = years, regionVar = "ADM1NAME",
    timeVar = "time", clusterVar = "~v001 + v002",
    ageVar = "age", weightsVar = "v005",
    geo.recode = NULL)
```

Neonatal Mortality - Person-months

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

```
births_neonates <- getBirths(data = dat,
    surveyyear = 2014, strata = "v023", year.cut = seq(199
        2015, by = 5), month.cut = c(1))
births_neonates$DHSCLUST <- births_neonates$v001
births_neonates <- merge(births_neonates,
    Kenya_pts)
```

Neonatal Mortality - Direct estimates

```
nmr <- countrySummary(births = births_neonates,
    years = years, regionVar = "ADM1NAME",
    timeVar = "time", clusterVar = "~v001 + v002",
    ageVar = "age", weightsVar = "v005",
    geo.recode = NULL)
```

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)
```

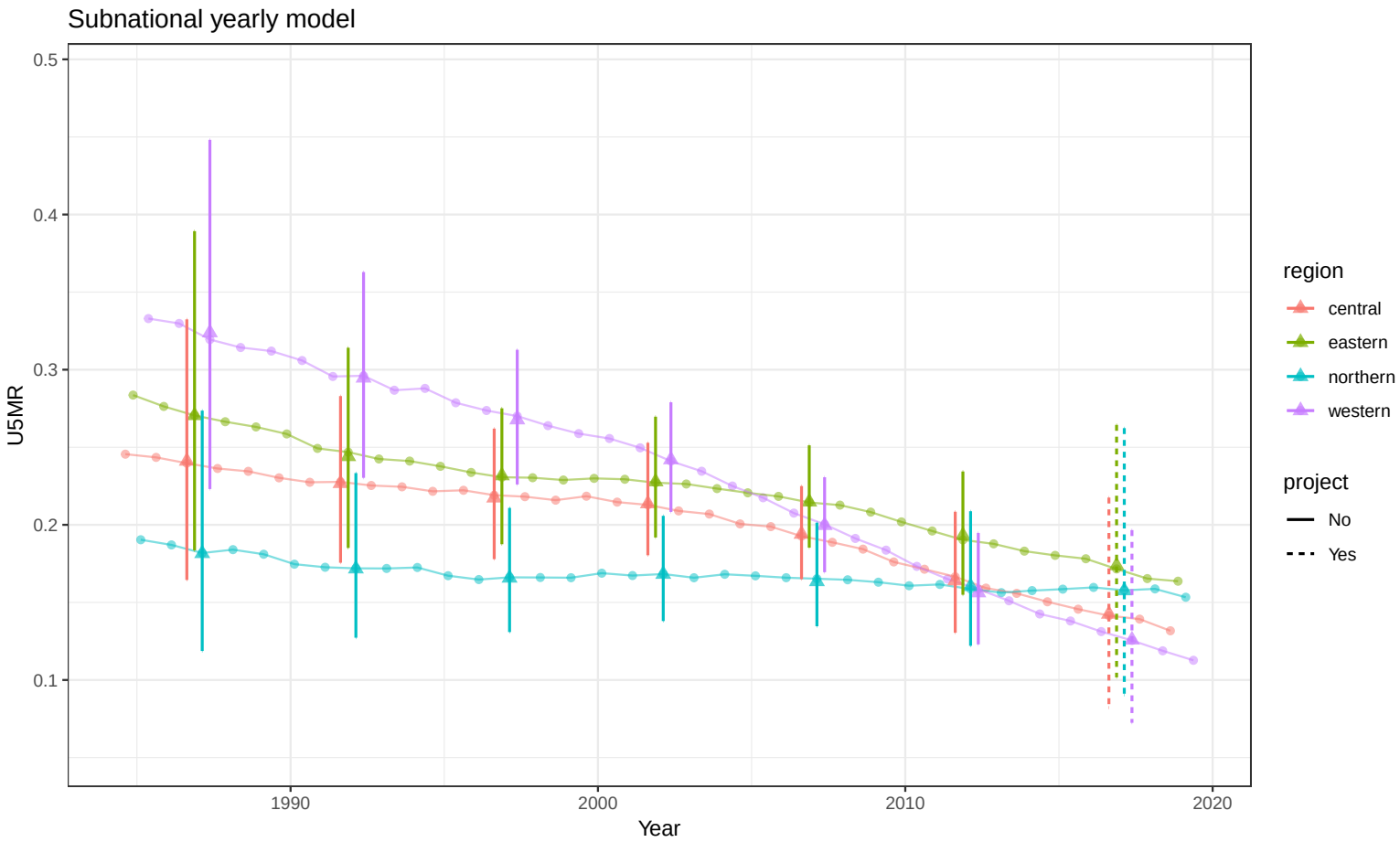
```
##   region years       u5m     lower
## 1    All 85-89 0.2410139 0.3267986
## 2    All 90-94 0.2082129 0.2729283
## 3    All 95-99 0.1929170 0.2338386
## 4    All 00-04 0.2169683 0.2519708
## 5    All 05-09 0.1890490 0.2158057
## 6    All 10-14 0.1589555 0.1876288
##       upper logit.est     var.est
## 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)
```

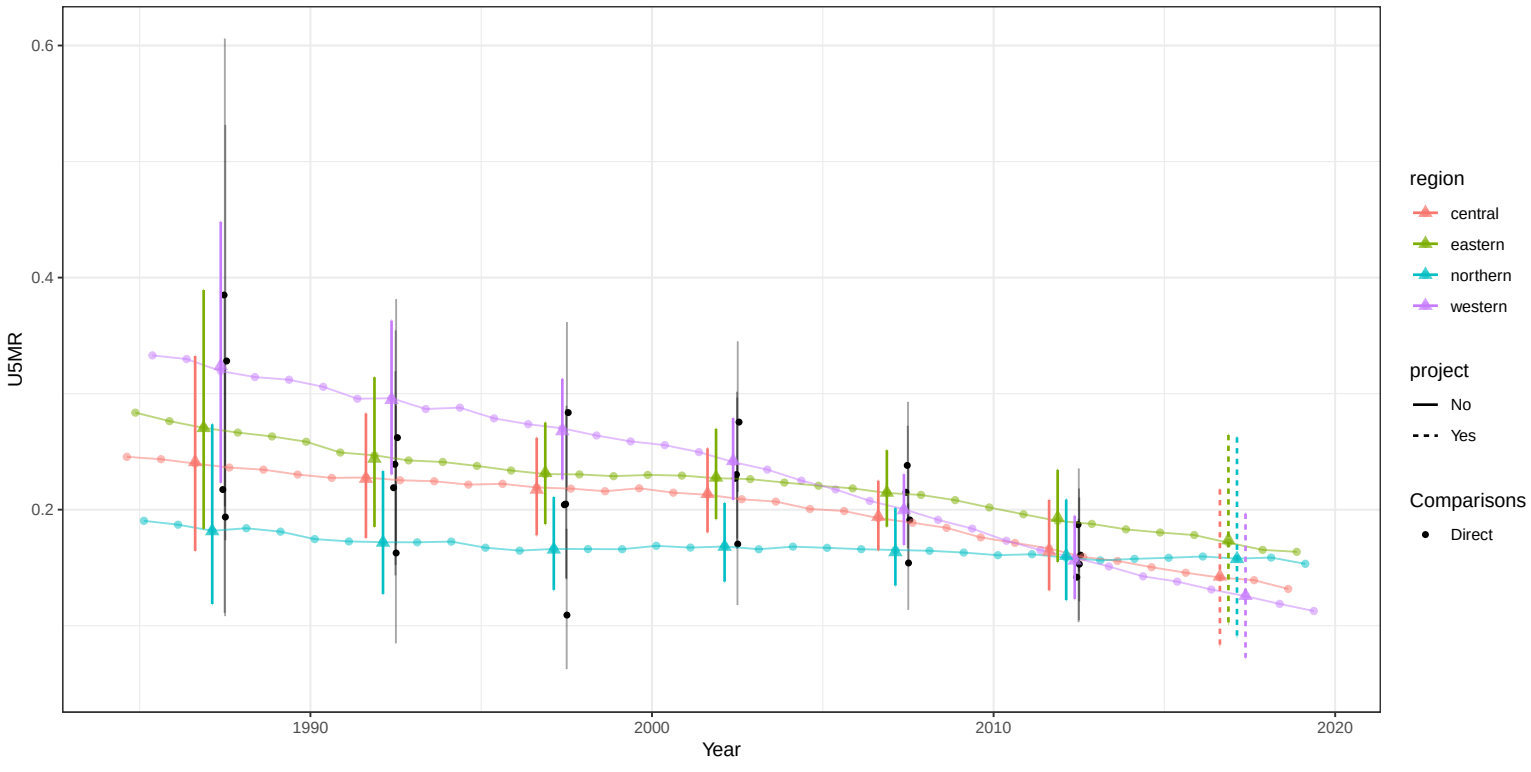
Multiple Surveys - Time-series plot

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



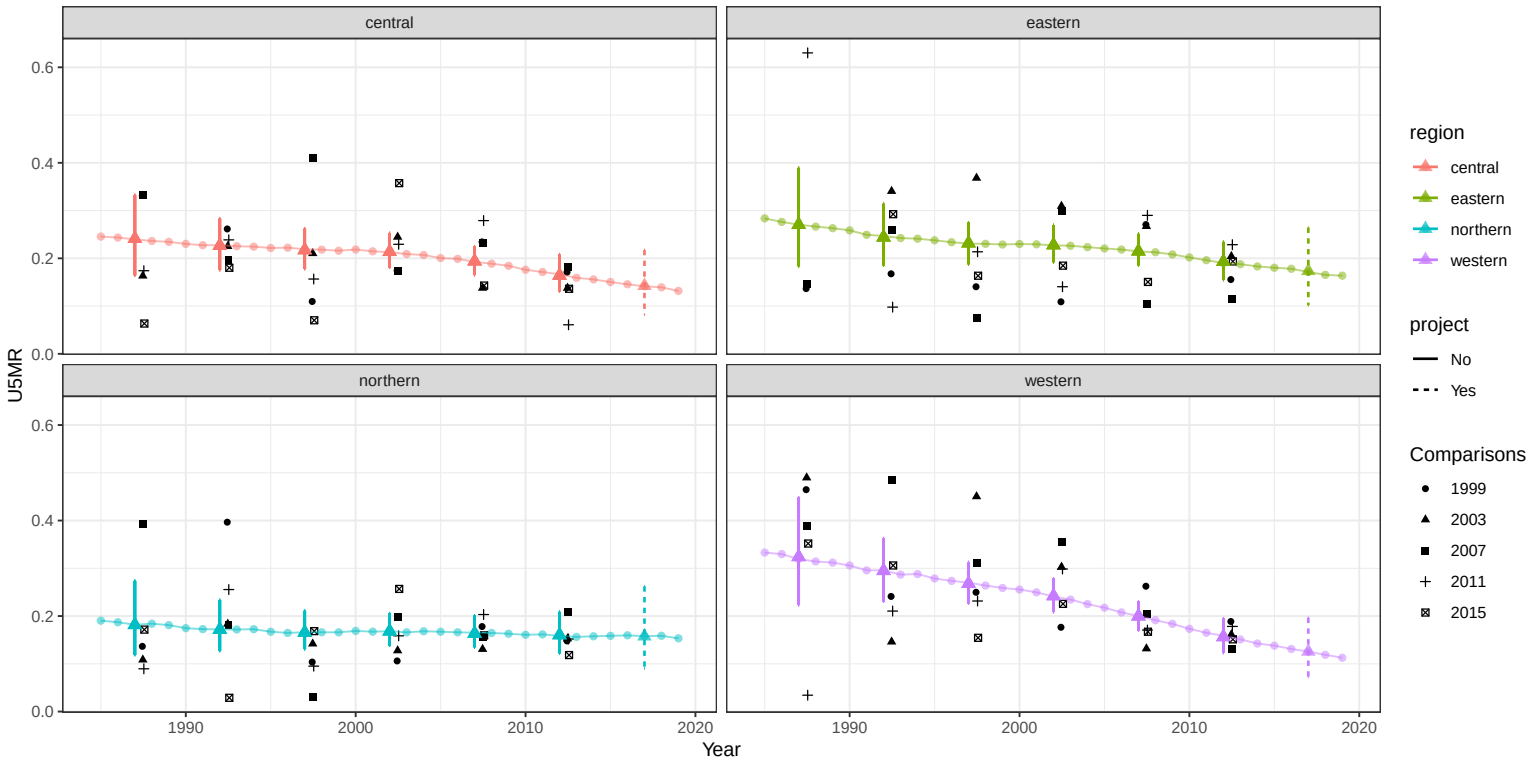
Multiple Surveys - Time-series plot

```
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()
```