Chapter 15 Internal Migration in Iceland

John Bryant and Junni L. Zhang

2018-09-20

Setup

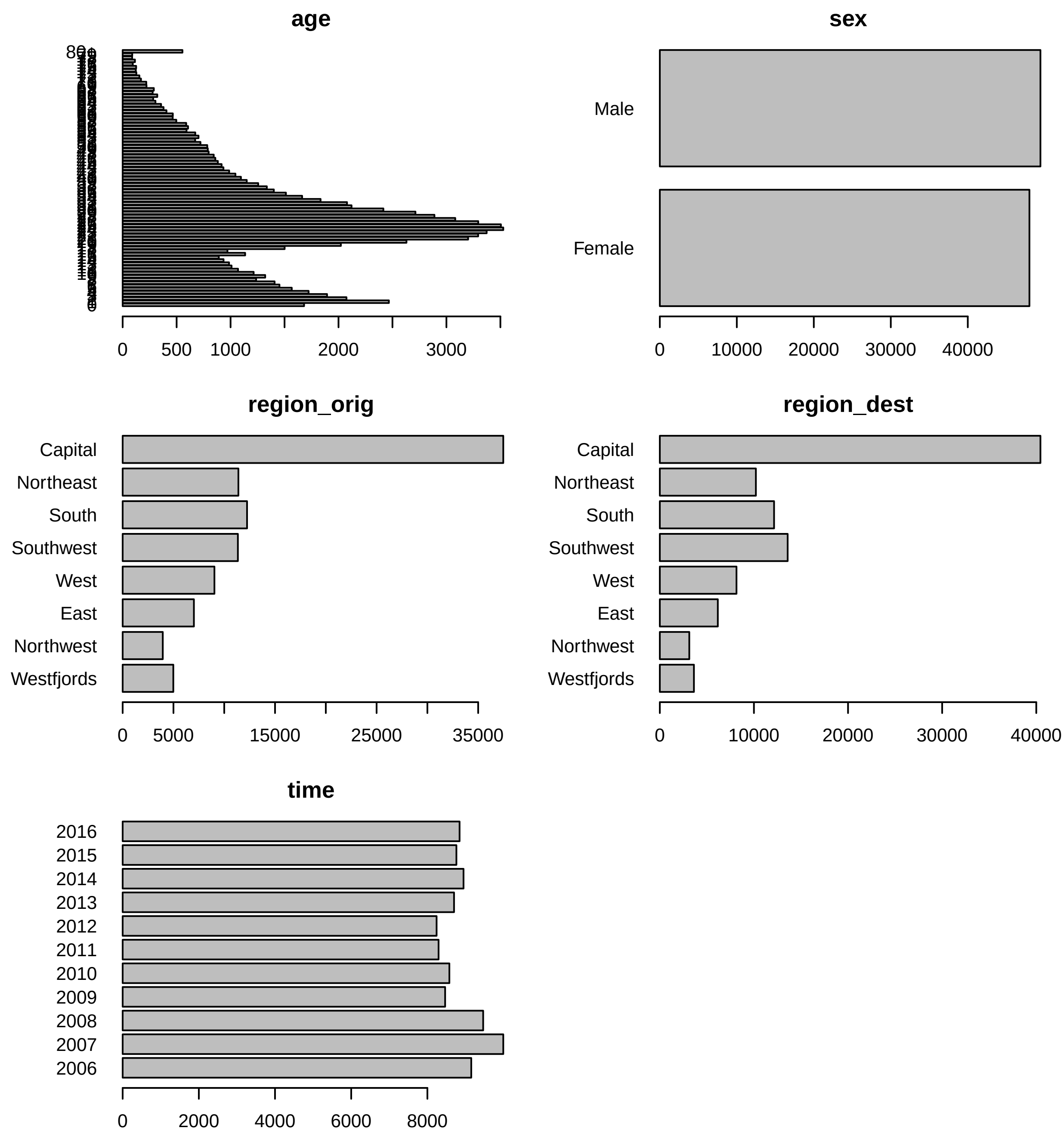
```r
library(methods)
library(bdefdata)
library(demest)
library(dplyr)
library(tidyverse)
library(latticeExtra)
```

Get data

```r
#Unconfidentialized migration counts
migration <- bdefdata::iceland_migration %>%
    Counts(dimscales = c(time = "Intervals"))
summary(migration)
```

```
##
## name:     age       sex    region_orig region_dest time
## length:   81        2      8           8           11
## dimtype:  age       sex    origin      destination time
## dimscale: Intervals Sexes  Categories  Categories  Intervals
## first:    0         Female Westfjords  Westfjords  2006
## last:     80+       Male   Capital     Capital     2016
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.8544  1.0000 34.0000
```

```r
plot(migration)
```

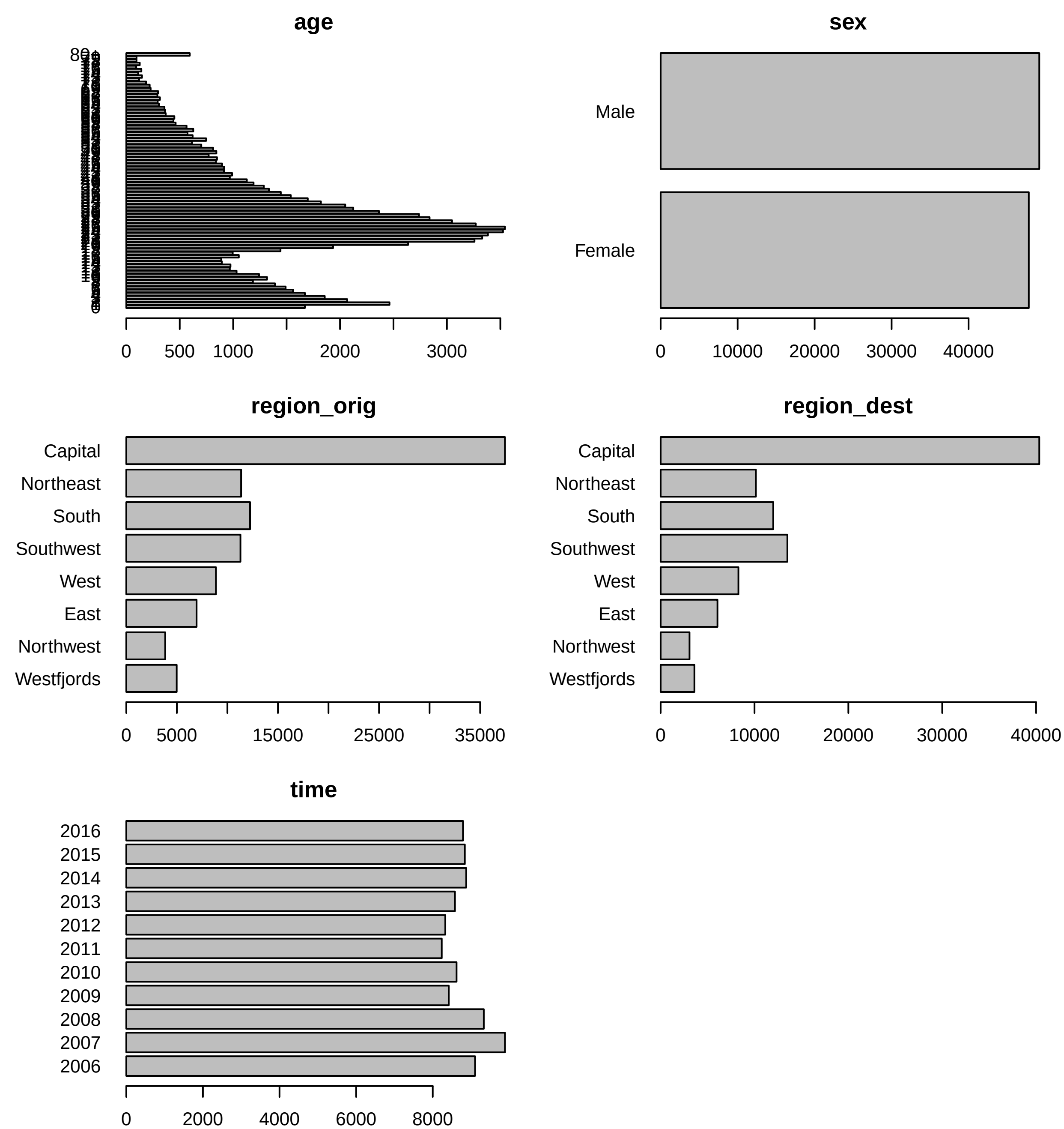



```r
#Confidentialized migration counts
set.seed(0)
migration_conf <- migration %>%
    round3()
summary(migration_conf)
```

```
##
## name:     age       sex    region_orig region_dest time
## length:   81        2      8           8           11
## dimtype:  age       sex    origin      destination time
## dimscale: Intervals Sexes  Categories  Categories  Intervals
```

```
## first:     0          Female Westfjords  Westfjords  2006
## last:      80+        Male   Capital     Capital     2016
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.8506  0.0000 33.0000
```

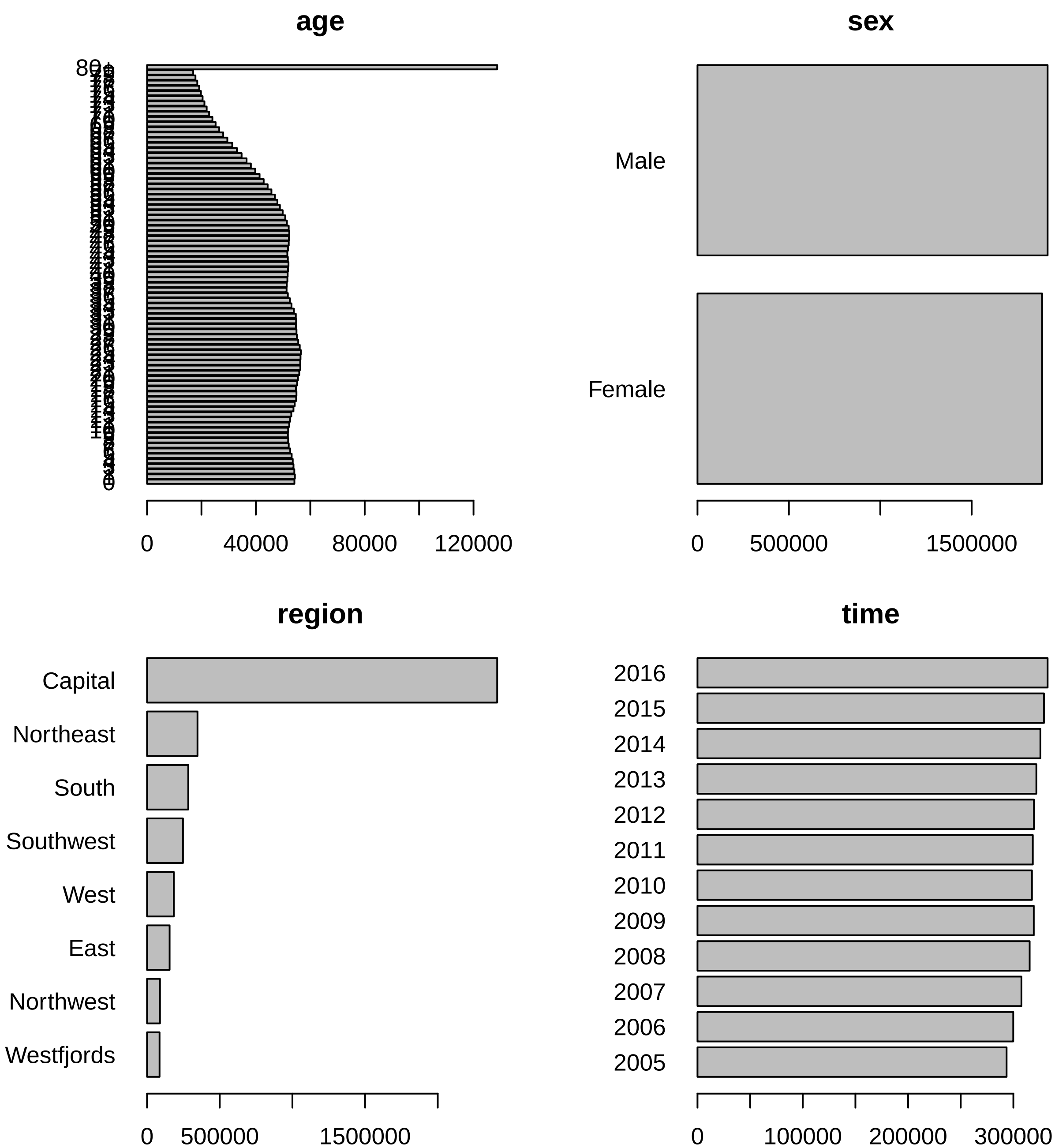
```r
plot(migration_conf)
```




```r
population <- bdefdata::iceland_population %>%
  Counts(dimscales = c(time = "Points"))
summary(population)
```

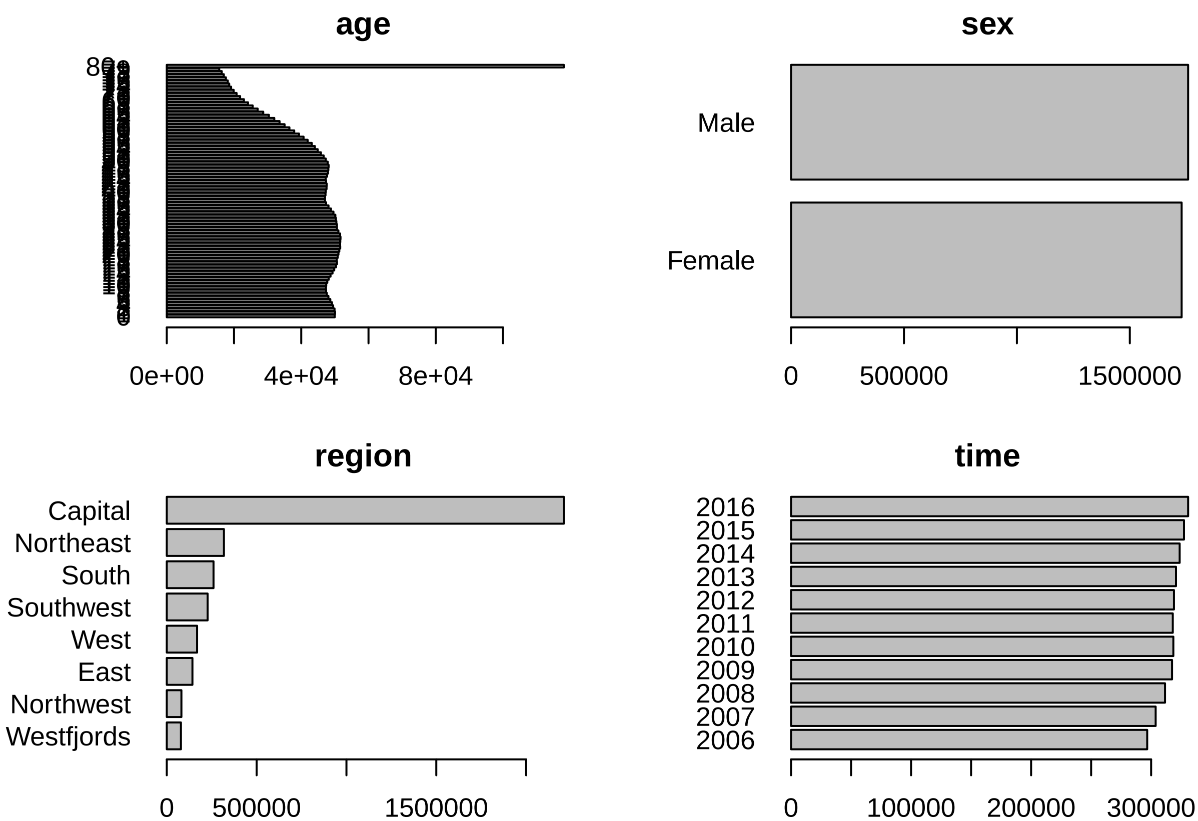
```
##
## name:     age       sex    region     time
## length:   81        2      8          12
## dimtype:  age       sex    state      time
## dimscale: Intervals Sexes  Categories Points
## first:    0         Female Westfjords 2005
## last:     80+       Male   Capital    2016
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    11.0    54.0   103.0   244.4   173.0  4716.0
```

```r
plot(population)
```

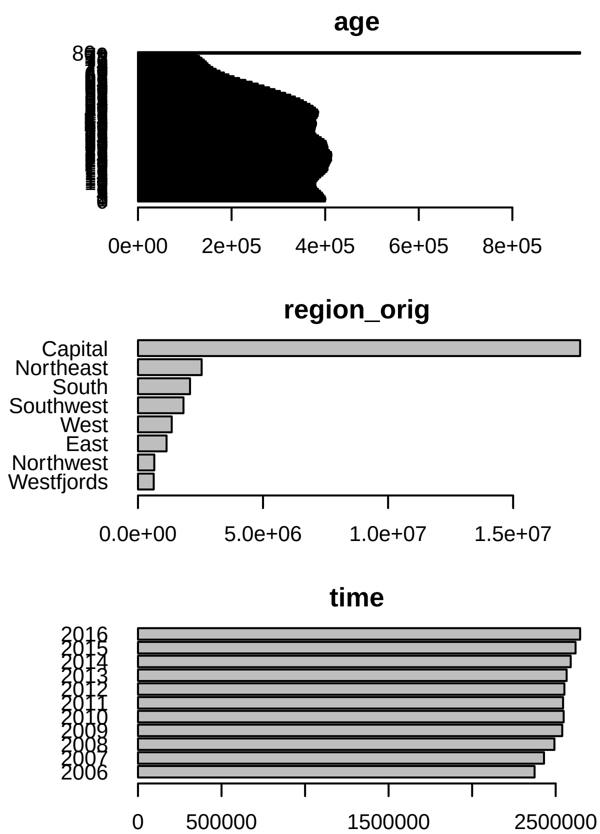


Calculate exposure

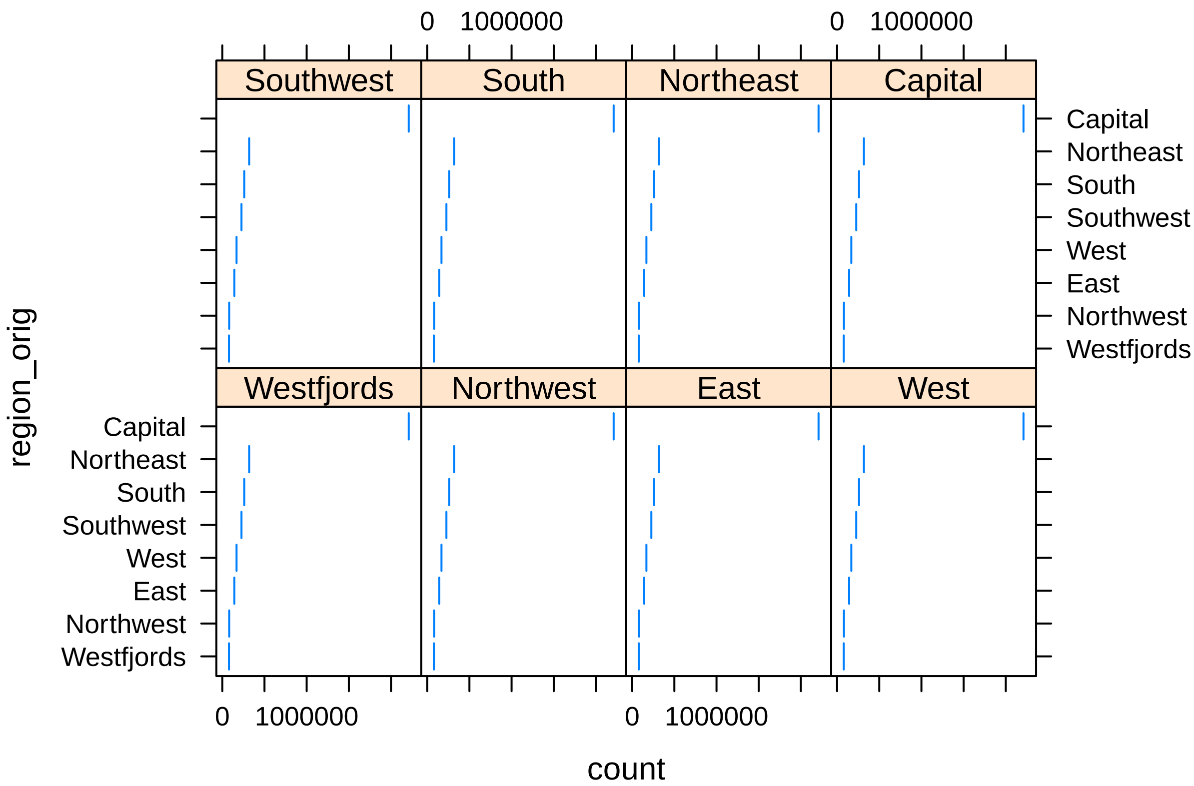
```
exposure <- population %>%
  exposure()
plot(exposure)
```




```
#`addPair` adds a `region_dest` dimension, with each destination having the same exposures
exposure <- exposure %>%
  addPair(base = "region")
plot(exposure)
```




```
dplot(~ region_orig | region_dest,
      data = exposure,
      horiz = TRUE)
```



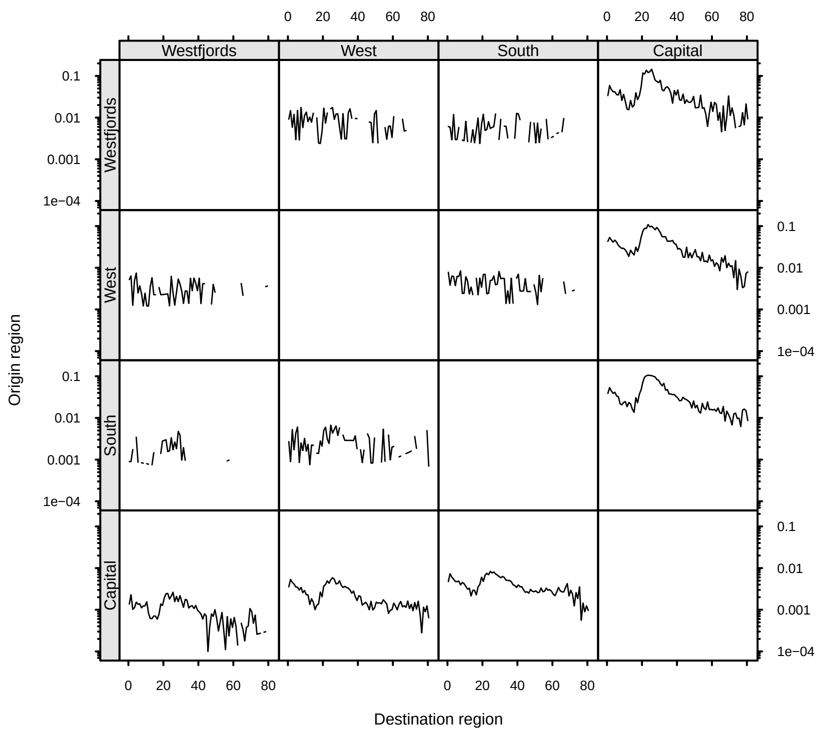
Plot data

```
#Direct estimates of migration rates (based on confidentialized counts),
#between four selected regions of Iceland, for the period 2006-2015.
rate <- (migration_conf / exposure) %>%
    collapseDimension(margin = c("age", "region_dest", "region_orig"),
                      weights = exposure)
rate[rate == 0] <- NA

sample_regions <- c("Westfjords", "West", "South", "Capital")

p <- dplot(~ age | region_dest + region_orig,
           data = rate,
           subarray = (region_dest %in% sample_regions) & (region_orig %in% sample_regions),
           midpoints = "age",
           as.table = TRUE,
           xlab = "Destination region",
           ylab = "Origin region",
           col = "black",
           lwd = 0.7,
           scales = list(tck = 0.3,
                         y = list(log = TRUE)),
           yscale.components = yscale.components.log10ticks,
           par.settings = list(fontsize = list(text = 6),
                               strip.background = list(col = "grey90")))
```

```r
p <- useOuterStrips(p)
plot(p)
```

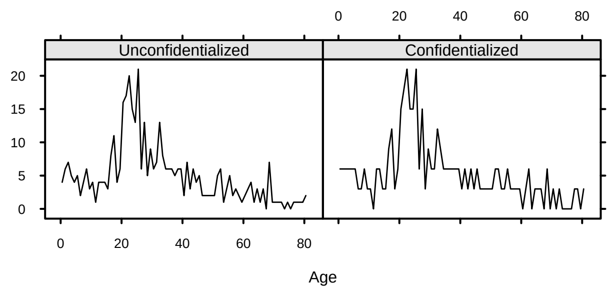



```r
#Data on internal migration from South Region to Capital Region, for females in
#2015---with and without #confidentialization by random rounding to base 3
migration_df <- dbind(Unconfidentialized = migration,
                      Confidentialized = migration_conf,
                      along = "variant") %>%
    as.data.frame(direction = "long", midpoints = "age") %>%
    filter(region_orig == "South") %>%
    filter(region_dest == "Capital") %>%
    filter(time == 2015) %>%
    filter(sex == "Female")

if (max(migration_df$count) > 3) {
    scales <- list(tck = 0.3)
} else {
    scales <- list(tck = 0.3, y = list(at = 0:3))
}

p <- xyplot(count ~ age | variant,
            data = migration_df,
            groups = variant,
            type = "l",
            xlab = "Age",
            ylab = "",
            lwd = 0.7,
            col = "black",
            scales = scales,
```

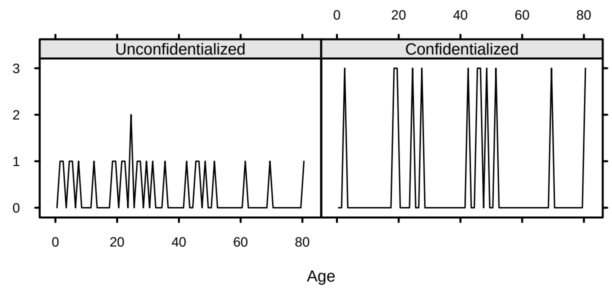
```
            par.settings = list(fontsize = list(text = 6),
                                strip.background = list(col = "grey90")))
plot(p)
```




```
#Data on internal migration from South Region to West Region, for females in
#2015---with and without confidentialization by random rounding to base 3
migration_df <- dbind(Unconfidentialized = migration,
                      Confidentialized = migration_conf,
                      along = "variant") %>%
    as.data.frame(direction = "long", midpoints = "age") %>%
    filter(region_orig == "South") %>%
    filter(region_dest == "West") %>%
    filter(time == 2015) %>%
    filter(sex == "Female")

if (max(migration_df$count) > 3) {
    scales <- list(tck = 0.3)
} else {
    scales <- list(tck = 0.3, y = list(at = 0:3))
}

p <- xyplot(count ~ age | variant,
            data = migration_df,
            groups = variant,
            type = "l",
            xlab = "Age",
            ylab = "",
            lwd = 0.7,
            col = "black",
            scales = scales,
            par.settings = list(fontsize = list(text = 6),
                                strip.background = list(col = "grey90")))
plot(p)
```



Fit the main model

```
filename <- "iceland_model_main.est"
if(!file.exists(filename)) {
    #Get datasets
    datasets <- list(migration = migration_conf)

    #Set up data models
    dataModels <- list(Model(migration ~ Round3()))

    #Set up system models
    model <- Model(y ~ Poisson(mean ~ age * sex + region_orig * region_dest + time,
                           structuralZeros = "diag"),
                   age ~ DLM(damp = NULL),
                   time ~ DLM(level = Level(scale = HalfT(scale = 0.05)),
                              trend = NULL,
                              damp = NULL),
                   age:sex ~ DLM(trend = NULL,
                                 damp = NULL),
                   jump = 0.5)

    #Estimation
    set.seed(0)

    n_burnin <- 225000
    n_sim <- 25000
    n_chain <-4
    n_thin <- 250

    estimateCounts(model,
                   y = migration_conf,
                   exposure = exposure,
                   dataModels = dataModels,
                   datasets = datasets,
                   filename = filename,
                   nBurnin = n_burnin,
                   nSim = n_sim,
                   nChain = n_chain,
                   nThin = n_thin)
}
fetchSummary(filename)
```

```
## --------------------------------------------------
## model:
## y ~ Poisson(mean ~ age * sex + region_orig * region_dest + time,
##     structuralZeros = "diag"),
## age ~ DLM(damp = NULL),
## time ~ DLM(level = Level(scale = HalfT(scale = 0.05)), trend = NULL,
##     damp = NULL),
## age:sex ~ DLM(trend = NULL, damp = NULL),
## 0.5
## dimensions: age, sex, region_orig, region_dest, time
## --------------------------------------------------
## y:
## dimensions: age, sex, region_orig, region_dest, time
## n cells: 103680
## --------------------------------------------------
## Data models:
## *migration*
## migration ~ Round3()
## dimensions: age, sex, region_orig, region_dest, time
## --------------------------------------------------
## Datasets:
## *migration*
## Object of class "Counts"
## dimensions: age, sex, region_orig, region_dest, time
## n cells: 103680,  n missing: 0, integers: TRUE, n zeros: 85009, median: 0
## --------------------------------------------------
## MCMC statistics:
## nBurnin: 225000,  nSim: 25000,  nChain: 4,  nThin: 250,  nIteration: 400
##
## Metropolis-Hastings updates:
##                        jump acceptance autocorr
## model.likelihood.rate  0.5      0.414    0.135
##
## parameters:
##                                             Rhat     2.5%     50%
## model.likelihood.rate                       1.09 0.000238 0.00192
## model.prior.mean                          . 1.23    -8.28   -6.26
## model.prior.sd                              1.07    0.215   0.228
## model.hyper.age.scaleLevel                  1.06   0.0023  0.0303
## model.hyper.age.scaleTrend                  1.03   0.0329  0.0468
## model.hyper.age.scaleError                  1.01  0.00623  0.0347
## model.hyper.region_orig.scaleError        . 1.14   0.0251  0.0672
## model.hyper.region_dest.scaleError          1.05    0.753    1.25
## model.hyper.time.scaleLevel                 1.01   0.0217  0.0445
## model.hyper.time.scaleError                 1.01 0.000887    0.02
## model.hyper.age:sex.scaleLevel              1.09   0.0439  0.0544
## model.hyper.age:sex.scaleError              1.07  0.00166 0.00913
## model.hyper.region_orig:region_dest.scaleError   1.02    0.297   0.378
## y                                           1.06        0       0
##                                            97.5% length
## model.likelihood.rate                     0.0463 103680
## model.prior.mean                           -3.08 103680
## model.prior.sd                             0.235      1
## model.hyper.age.scaleLevel                0.0883      1
```

```
## model.hyper.age.scaleTrend                         0.072       1
## model.hyper.age.scaleError                        0.0573       1
## model.hyper.region_orig.scaleError                 0.231       1
## model.hyper.region_dest.scaleError                  1.71       1
## model.hyper.time.scaleLevel                       0.0899       1
## model.hyper.time.scaleError                       0.0871       1
## model.hyper.age:sex.scaleLevel                    0.0639       1
## model.hyper.age:sex.scaleError                    0.0275       1
## model.hyper.region_orig:region_dest.scaleError      0.45       1
## y                                                      7  103680
## --------------------------------------------------
```

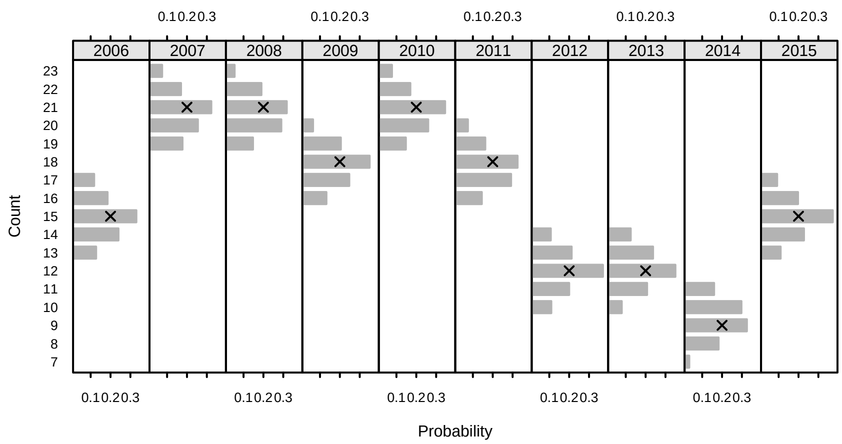
Results for Unconfidentialized Migration Counts

```
#Unconfidentialized migration counts from South Region to Capital Region
est_counts <- fetch(filename,
                    where = "y") %>%
    subarray(region_orig == "South") %>%
    subarray(region == "Capital") %>%
    subarray(age == "20") %>%
    subarray(sex == "Female") %>%
    as.data.frame(direction = "long") %>%
    mutate(count = factor(count, levels = 0:max(count))) %>%
    group_by(time, count) %>%
    summarise(n = n()) %>%
    mutate(prob = n / sum(n))

conf_counts <- fetch(filename,
                     where = c("datasets", "mig")) %>%
    subarray(region_orig == "South") %>%
    subarray(region == "Capital") %>%
    subarray(age == "20") %>%
    subarray(sex == "Female", drop = FALSE) %>%
    as.data.frame(direction = "long") %>%
    mutate(prob = 0.2) %>%
    mutate(count = count - min(as.integer(est_counts$count)) + 2)

p <- barchart(count ~ prob | time,
              data = est_counts,
              horiz = TRUE,
              xlab = "Probability",
              layout = c(NA, 1),
              ylab = "Count",
              scales = list(tck = 0.3),
              col = "grey70",
              border = "grey70",
              par.settings = list(fontsize = list(text = 6, points = 5),
                                  strip.background = list(col = "grey90")))
p <- p + as.layer(xyplot(count ~ prob | time,
                         data = conf_counts,
                         col = "black",
                         cex = 0.9,
                         pch = 4,
```

```
                          layout = c(NA, 1)))
plot(p)
```

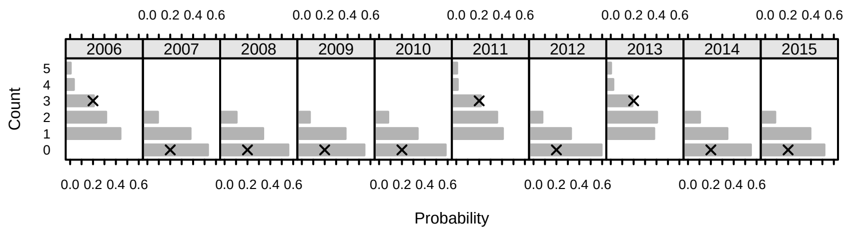



```
#Unconfidentialized migration counts from South Region to West Region
est_counts <- fetch(filename,
                    where = "y") %>%
    subarray(region_orig == "South") %>%
    subarray(region == "West") %>%
    subarray(age == "20") %>%
    subarray(sex == "Female") %>%
    as.data.frame(direction = "long") %>%
    mutate(count = factor(count, levels = 0:max(count))) %>%
    group_by(time, count) %>%
    summarise(n = n()) %>%
    mutate(prob = n / sum(n))

conf_counts <- fetch(filename,
                     where = c("datasets", "mig")) %>%
    subarray(region_orig == "South") %>%
    subarray(region == "West") %>%
    subarray(age == "20") %>%
    subarray(sex == "Female", drop = FALSE) %>%
    as.data.frame(direction = "long") %>%
    mutate(prob = 0.2) %>%
    mutate(count = count - min(as.integer(est_counts$count)) + 2)

p <- barchart(count ~ prob | time,
              data = est_counts,
              horiz = TRUE,
              xlab = "Probability",
              layout = c(NA, 1),
              ylab = "Count",
              scales = list(tck = 0.3),
              col = "grey70",
              border = "grey70",
              par.settings = list(fontsize = list(text = 6, points = 5),
                                  strip.background = list(col = "grey90")))
p <- p + as.layer(xyplot(count ~ prob | time,
```

```
                          data = conf_counts,
                          col = "black",
                          cex = 0.9,
                          pch = 4,
                          layout = c(NA, 1)))
plot(p)
```



Fit the model that ignores the fact that the data have been confidentialized.

```
filename_alt <- "iceland_model_alt.est"
if(!file.exists(filename_alt)) {
    #Set up model
    model <- Model(y ~ Poisson(mean ~ age * sex + region_orig * region_dest + time,
                                structuralZeros = "diag"),
                   age ~ DLM(damp = NULL),
                   time ~ DLM(level = Level(scale = HalfT(scale = 0.05)),
                              trend = NULL,
                              damp = NULL),
                   age:sex ~ DLM(trend = NULL,
                                 damp = NULL),
                   jump = 0.5)

    #Estimation
    set.seed(0)

    n_burnin <- 25000
    n_sim <- 25000
    n_chain <-4
    n_thin <- 250

    estimateModel(model,
                  y = migration_conf,
                  exposure = exposure,
                  filename = filename_alt,
                  nBurnin = n_burnin,
                  nSim = n_sim,
                  nChain = n_chain,,
                  nThin = n_thin)
}
fetchSummary(filename_alt)
```

```
## --------------------------------------------------
## model:
## y ~ Poisson(mean ~ age * sex + region_orig * region_dest + time,
```

```
##     structuralZeros = "diag"),
## age ~ DLM(damp = NULL),
## time ~ DLM(level = Level(scale = HalfT(scale = 0.05)), trend = NULL,
##     damp = NULL),
## age:sex ~ DLM(trend = NULL, damp = NULL),
## 0.5
## dimensions: age, sex, region_orig, region_dest, time
## --------------------------------------------------
## y:
## Object of class "Counts"
## dimensions: age, sex, region_orig, region_dest, time
## n cells: 103680,  n missing: 0, integers: TRUE, n zeros: 85009, median: 0
## --------------------------------------------------
## MCMC statistics:
## nBurnin: 25000,  nSim: 25000,  nChain: 4,  nThin: 250,  nIteration: 400
##
## Metropolis-Hastings updates:
##                       jump acceptance autocorr
## model.likelihood.rate  0.5      0.829    0.071
##
## parameters:
##                                                Rhat      2.5%    50%
## model.likelihood.rate                        . 1.24 0.0000136 0.0007
## model.prior.mean                               1.03     -9.95  -7.31
## model.prior.sd                                 1.04      1.42   1.43
## model.hyper.age.scaleLevel                     1.02    0.0107 0.0593
## model.hyper.age.scaleTrend                     1.01    0.0214 0.0596
## model.hyper.age.scaleError                     1.00    0.0015 0.0344
## model.hyper.region_orig.scaleError             1.05    0.0185  0.148
## model.hyper.region_dest.scaleError             1.01     0.751   1.35
## model.hyper.time.scaleLevel                    1.00   0.00722 0.0592
## model.hyper.time.scaleError                    1.02   0.00699 0.0462
## model.hyper.age:sex.scaleLevel                 1.07     0.036 0.0529
## model.hyper.age:sex.scaleError                 1.00    0.0137 0.0341
## model.hyper.region_orig:region_dest.scaleError   1.01      0.37  0.439
##                                              97.5% length
## model.likelihood.rate                        0.0561 103680
## model.prior.mean                              -3.45 103680
## model.prior.sd                                 1.46      1
## model.hyper.age.scaleLevel                    0.151      1
## model.hyper.age.scaleTrend                   0.0795      1
## model.hyper.age.scaleError                   0.0827      1
## model.hyper.region_orig.scaleError             0.48      1
## model.hyper.region_dest.scaleError             2.01      1
## model.hyper.time.scaleLevel                  0.0965      1
## model.hyper.time.scaleError                   0.139      1
## model.hyper.age:sex.scaleLevel               0.0759      1
## model.hyper.age:sex.scaleError               0.0606      1
## model.hyper.region_orig:region_dest.scaleError  0.573      1
## --------------------------------------------------
```

Results for Migration Rates

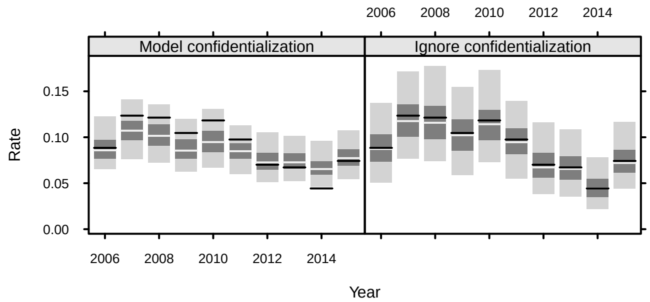
```
#Migration rates from South Region to Capital Region
rates_main <- fetch(filename,
                    where = c("model", "likelihood", "rate"))
rates_alt <- fetch(filename_alt,
                    where = c("model", "likelihood", "rate"))

rates <- dbind("Model confidentialization" = rates_main,
               "Ignore confidentialization" = rates_alt,
               along = "variant") %>%
    subarray(region_orig == "South") %>%
    subarray(region == "Capital") %>%
    subarray(age == "20") %>%
    subarray(sex == "Female")

mig <- fetch(filename_alt,
             where = "y")
exposure_draws <- fetch(filename_alt,
                  where = "exposure")
direct <- dbind("Model confidentialization" = mig / exposure_draws,
                "Ignore confidentialization" = mig / exposure_draws,
                along = "variant") %>%
    subarray(region_orig == "South") %>%
    subarray(region == "Capital") %>%
    subarray(age == "20") %>%
    subarray(sex == "Female")


at <- match(seq(2006, 2014, 2), dimnames(rates)$time)
p <- dplot( ~ time | variant,
           data = rates,
           groups = variant,
           xlab = "Year",
           ylab = "Rate",
           col = "grey",
           ylim = c(-0.005, NA),
           scales = list(tck = 0.3, x = list(at = at)),
           par.settings = list(fontsize = list(text = 6),
                               strip.background = list(col = "grey90")),
           overlay = list(values = direct, col = "black"))

plot(p)
```



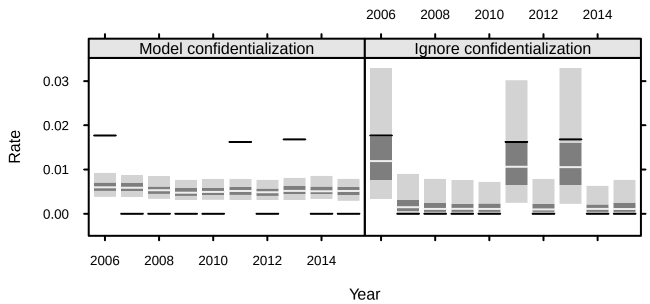

```
#Migration rates from South Region to West Region
rates_main <- fetch(filename,
                    where = c("model", "likelihood", "rate"))
rates_alt <- fetch(filename_alt,
                   where = c("model", "likelihood", "rate"))

rates <- dbind("Model confidentialization" = rates_main,
               "Ignore confidentialization" = rates_alt,
               along = "variant") %>%
    subarray(region_orig == "South") %>%
    subarray(region == "West") %>%
    subarray(age == "20") %>%
    subarray(sex == "Female")

mig <- fetch(filename_alt,
             where = "y")
exposure_draws <- fetch(filename_alt,
                        where = "exposure")
direct <- dbind("Model confidentialization" = mig / exposure_draws,
                "Ignore confidentialization" = mig / exposure_draws,
                along = "variant") %>%
    subarray(region_orig == "South") %>%
    subarray(region == "West") %>%
    subarray(age == "20") %>%
    subarray(sex == "Female")


at <- match(seq(2006, 2014, 2), dimnames(rates)$time)
p <- dplot( ~ time | variant,
           data = rates,
           groups = variant,
           xlab = "Year",
           ylab = "Rate",
           col = "grey",
           ylim = c(-0.005, NA),
           scales = list(tck = 0.3, x = list(at = at)),
           par.settings = list(fontsize = list(text = 6),
                               strip.background = list(col = "grey90")),
           overlay = list(values = direct, col = "black"))

plot(p)
```



Forecast migration counts

```
filename_pred <- "iceland_model_main.pred"
if(!file.exists(filename_pred)) {
    exposure_all <- exposure %>%
        extrapolate(along="time",
                    labels = c("2017", "2018", "2019", "2020")) %>%
        subarray(time %in% c("2016", "2017", "2018", "2019", "2020"))

    predictCounts(filenameEst = filename,
                  filenamePred = filename_pred,
                  n = 5,
                  exposure = exposure_all)
}
```

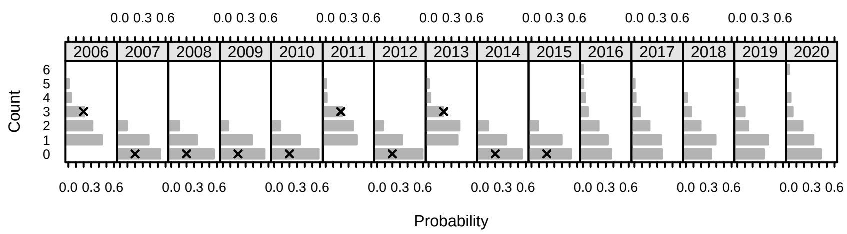
```
#Estimated and forecasted unconfidentialized migration counts between South Region and West Region
counts <- fetchBoth(filenameEst = filename,
                    filenamePred = filename_pred,
                    where = "y") %>%
    subarray(region_orig == "South") %>%
    subarray(region == "West") %>%
    subarray(age == "20") %>%
    subarray(sex == "Female") %>%
    as.data.frame(direction = "long") %>%
    mutate(count = factor(count, levels = 0:max(count))) %>%
    group_by(time, count) %>%
    summarise(n = n()) %>%
    mutate(prob = n / sum(n))

conf_counts <- fetch(filename,
                     where = c("datasets", "mig")) %>%
    subarray(region_orig == "South") %>%
    subarray(region == "West") %>%
    subarray(age == "20") %>%
    subarray(sex == "Female", drop = FALSE) %>%
    as.data.frame(direction = "long") %>%
    mutate(prob = 0.2) %>%
    mutate(count = count - min(as.integer(counts$count)) + 2) %>%
    right_join(select(counts, time))
```

```
## Joining, by = "time"
```

```
## Warning: Column `time` joining factors with different levels, coercing to
## character vector
```

```
p <- barchart(count ~ prob | time,
              data = counts,
              horiz = TRUE,
              xlab = "Probability",
              layout = c(NA, 1),
              ylab = "Count",
              scales = list(tck = 0.3),
              col = "grey70",
              border = "grey70",
              par.settings = list(fontsize = list(text = 6, points = 5),
                                  strip.background = list(col = "grey90")))
p <- p + as.layer(xyplot(count ~ prob | time,
                         data = conf_counts,
                         col = "black",
                         cex = 0.7,
                         pch = 4,
                         layout = c(NA, 1)))
plot(p)
```




```
#Forecasts of confidentialized migration counts between South Region and West Region
forecasts <- fetch(filename = filename_pred,
                   where = "y") %>%
    subarray(region_orig == "South") %>%
    subarray(region == "West") %>%
    subarray(age == "20") %>%
    subarray(sex == "Female") %>%
    round3() %>%
    as.data.frame(direction = "long", stringsAsFactors = FALSE) %>%
    group_by(time, count) %>%
    summarise(n = n()) %>%
    mutate(prob = n / sum(n)) %>%
    mutate(count = factor(count, levels = seq(from = 0, to = max(count)))) %>%
    xtabs(prob ~ time + count, data = .) %>%
    as.data.frame.table(responseName = "prob") %>%
    mutate(prob = ifelse(prob == 0, NA, prob))

p <- barchart(count ~ prob | time,
              data = forecasts,
              horiz = TRUE,
              xlab = "Probability",
              layout = c(NA, 1),
              ylab = "Count",
              scales = list(tck = 0.3),
```

```r
                col = "grey70",
                border = "grey70",
                par.settings = list(fontsize = list(text = 6, points = 5),
                                    strip.background = list(col = "grey90")))

plot(p)
```