

Synthesize population

Chris Hoover et al

4/23/2021

Synthesize population

```
ct_synth <- function(ct_fips, h_seed, p_seed, h_tgt, p_tgt){
  ct_ipu <- ipu(h_seed, h_tgt, p_seed, p_tgt, primary_id="SERIALNO")
  ct_syn_h <- synthesize(ct_ipu$weight_tbl, primary_id="SERIALNO")
  ct_syn_p <- left_join(ct_syn_h, p_seed, by="SERIALNO") %>%
    rename(house_id=new_id) #>% select(house_id, indiv_id, hhsize, hhincome, sex, age, occp, race)

  return(ct_syn_p)
}

# cts_synth_list <- lapply(synth_cts, function(ct){
#   puma_ct <- cts_to_pumas %>% filter(CTFP == ct) %>% pull(PUMA5CE)
#   ct_p_seed <- SF_p_seed %>% filter(PUMA == puma_ct) %>% dplyr::select(-c(PUMA, weight))
#   # Make sure all households have corresponding individuals in seed data
#   p_ids <- ct_p_seed$SERIALNO
#   ct_hh_seed <- SF_hh_seed %>% filter(PUMA == puma_ct & SERIALNO %in% p_ids) %>% dplyr::select(-c(PUMA, weight))
#   #
#   ct_hh_tgt <- lapply(SF_hh_tgt, function(i){
#     i %>% filter(geo_tract == ct) %>% dplyr::select(-geo_tract)
#   })
#   #
#   ct_p_tgt <- lapply(SF_p_tgt, function(j){
#     j %>% filter(geo_tract == ct) %>% dplyr::select(-geo_tract)
#   })
#   #
#   cat(ct, "\n")
#   #
#   ct_synth(ct_fips = ct,
#             h_seed = ct_hh_seed,
#             p_seed = ct_p_seed,
#             h_tgt = ct_hh_tgt,
#             p_tgt = ct_p_tgt) %>%
#     mutate(GEOID = ct)
#   #
# })

# Run synthesis across census tracts in parallel
clust <- makeCluster(detectCores()-2)
clusterExport(clust, c("ct_synth", "synth_cts",
```

```

      "SF_hh_seed", "SF_p_seed",
      "SF_hh_tgt", "SF_p_tgt"))
clusterEvalQ(clust, devtools::load_all())

```

```

## [[1]]
## [[1]]$env
## <environment: namespace:Rsynthpops>
##
## [[1]]$data
## [1] "acs_grade_lookup"          "acs_hhincome_lookup"
## [3] "acs_occip_lookup"         "acs_race_eth_lookup"
## [5] "acs_sex_by_age_lookup"    "acs_sex_by_age"
## [7] "acs_vars_hhincome_lookup" "cts_to_pumas"
##
## [[1]]$code
## [1] "/Users/chrishoover/Documents/Research/Rsynthpops/R/All_Functions.R"
## [2] "/Users/chrishoover/Documents/Research/Rsynthpops/R/Gen_Pop.R"
## [3] "/Users/chrishoover/Documents/Research/Rsynthpops/R/data.R"
##
## [[1]]$dll
## list()
##
##
## [[2]]
## [[2]]$env
## <environment: namespace:Rsynthpops>
##
## [[2]]$data
## [1] "acs_grade_lookup"          "acs_hhincome_lookup"
## [3] "acs_occip_lookup"         "acs_race_eth_lookup"
## [5] "acs_sex_by_age_lookup"    "acs_sex_by_age"
## [7] "acs_vars_hhincome_lookup" "cts_to_pumas"
##
## [[2]]$code
## [1] "/Users/chrishoover/Documents/Research/Rsynthpops/R/All_Functions.R"
## [2] "/Users/chrishoover/Documents/Research/Rsynthpops/R/Gen_Pop.R"
## [3] "/Users/chrishoover/Documents/Research/Rsynthpops/R/data.R"
##
## [[2]]$dll
## list()
##
##
## [[3]]
## [[3]]$env
## <environment: namespace:Rsynthpops>
##
## [[3]]$data
## [1] "acs_grade_lookup"          "acs_hhincome_lookup"
## [3] "acs_occip_lookup"         "acs_race_eth_lookup"
## [5] "acs_sex_by_age_lookup"    "acs_sex_by_age"
## [7] "acs_vars_hhincome_lookup" "cts_to_pumas"
##
## [[3]]$code
## [1] "/Users/chrishoover/Documents/Research/Rsynthpops/R/All_Functions.R"

```

```

## [2] "/Users/chrishoover/Documents/Research/Rsynthpops/R/Gen_Pop.R"
## [3] "/Users/chrishoover/Documents/Research/Rsynthpops/R/data.R"
##
## [[3]]$dll
## list()
##
##
## [[4]]
## [[4]]$env
## <environment: namespace:Rsynthpops>
##
## [[4]]$data
## [1] "acs_grade_lookup"          "acs_hhincome_lookup"
## [3] "acs_occip_lookup"         "acs_race_eth_lookup"
## [5] "acs_sex_by_age_lookup"     "acs_sex_by_age"
## [7] "acs_vars_hhincome_lookup" "cts_to_pumas"
##
## [[4]]$code
## [1] "/Users/chrishoover/Documents/Research/Rsynthpops/R/All_Functions.R"
## [2] "/Users/chrishoover/Documents/Research/Rsynthpops/R/Gen_Pop.R"
## [3] "/Users/chrishoover/Documents/Research/Rsynthpops/R/data.R"
##
## [[4]]$dll
## list()
##
##
## [[5]]
## [[5]]$env
## <environment: namespace:Rsynthpops>
##
## [[5]]$data
## [1] "acs_grade_lookup"          "acs_hhincome_lookup"
## [3] "acs_occip_lookup"         "acs_race_eth_lookup"
## [5] "acs_sex_by_age_lookup"     "acs_sex_by_age"
## [7] "acs_vars_hhincome_lookup" "cts_to_pumas"
##
## [[5]]$code
## [1] "/Users/chrishoover/Documents/Research/Rsynthpops/R/All_Functions.R"
## [2] "/Users/chrishoover/Documents/Research/Rsynthpops/R/Gen_Pop.R"
## [3] "/Users/chrishoover/Documents/Research/Rsynthpops/R/data.R"
##
## [[5]]$dll
## list()
##
##
## [[6]]
## [[6]]$env
## <environment: namespace:Rsynthpops>
##
## [[6]]$data
## [1] "acs_grade_lookup"          "acs_hhincome_lookup"
## [3] "acs_occip_lookup"         "acs_race_eth_lookup"
## [5] "acs_sex_by_age_lookup"     "acs_sex_by_age"
## [7] "acs_vars_hhincome_lookup" "cts_to_pumas"

```

```
##
## [[6]]$code
## [1] "/Users/chris Hoover/Documents/Research/Rsynthpops/R/All_Functions.R"
## [2] "/Users/chris Hoover/Documents/Research/Rsynthpops/R/Gen_Pop.R"
## [3] "/Users/chris Hoover/Documents/Research/Rsynthpops/R/data.R"
##
## [[6]]$dll
## list()

cts_synth_list <- parLapply(clust, synth_cts, function(ct){
  puma_ct <- cts_to_pumas %>% filter(CTFP == ct) %>% pull(PUMA5CE)
  ct_p_seed <- SF_p_seed %>% filter(PUMA == puma_ct) %>% dplyr::select(-c(PUMA, weight))
  # Make sure all households have corresponding individuals in seed data
  p_ids <- ct_p_seed$SERIALNO
  ct_hh_seed <- SF_hh_seed %>% filter(PUMA == puma_ct & SERIALNO %in% p_ids) %>% dplyr::select(-c(PUMA,

  ct_hh_tgt <- lapply(SF_hh_tgt, function(i){
    i %>% filter(geo_tract == ct) %>% dplyr::select(-geo_tract)
  })

  ct_p_tgt <- lapply(SF_p_tgt, function(j){
    j %>% filter(geo_tract == ct) %>% dplyr::select(-geo_tract)
  })

  ct_synth(ct, ct_hh_seed, ct_p_seed, ct_hh_tgt, ct_p_tgt) %>%
    mutate(GEOID = ct)

})

stopCluster(clust)

sf_pop <- bind_rows(cts_synth_list)

# Add unique individual identifier
sf_pop <- sf_pop %>%
  mutate(indiv_id=rownames(.),
         house_id = paste0(GEOID, "_", house_id))

head(sf_pop)

## # A tibble: 6 x 14
##   house_id SERIALNO hhincome hhtype hhsize grade sex occupation age_cat race
##   <chr>      <chr>    <fct>    <dbl>  <dbl> <chr> <chr> <chr>      <int> <chr>
## 1 06075010~ 2015001~ 2          1      2 bb    1    1010          3 2
## 2 06075010~ 2015001~ 2          1      2 16    2     0009          4 1
## 3 06075010~ 2017000~ 1          1      6 bb    2     4810          4 6
## 4 06075010~ 2017000~ 1          1      6 bb    1     0040          4 6
## 5 06075010~ 2017000~ 1          1      6 01    1     0009          1 6
## 6 06075010~ 2017000~ 1          1      6 bb    2     0009          1 6
## # ... with 4 more variables: hispanic <dbl>, occ_group <dbl>, GEOID <chr>,
## #   indiv_id <chr>

saveRDS(sf_pop,
        file = here::here("Tutorial/data/synth_pop.rds"))
```