demdescript

Christopher Huong

2023-04-14

library(rio)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.4.0 ✔ purrr 0.3.5   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1   
## ✔ readr 2.1.3 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

wave1 <- import('data/wave1.rda')  
wave11 <- import('data/wave1\_1.rda')  
wave5 <- import('data/wave5.rda')  
  
  
  
  
dat <- wave1 %>%  
 select(AID, H1GI1Y, IYEAR,  
   
   
   
   
 H1FS1, H1FS2, H1FS3, H1FS4, H1FS5, H1FS6,  
 H1FS7, H1FS8, H1FS9, H1FS10, H1FS11, H1FS12,  
 H1FS13, H1FS14, H1FS15, H1FS16, H1FS17,   
 H1FS18, H1FS19)  
  
dat5 <- wave5 %>%  
 select(AID, H5ID6G, H5ID6GA)  
  
#clean up age data to compute age, then remove B\_YR and I\_YR  
dat <- dat %>%  
 mutate(H1GI1Y = case\_when(H1GI1Y == "(74) (74) 1974" ~ 1974,  
 H1GI1Y == "(75) (75) 1975" ~ 1975,  
 H1GI1Y == "(76) (76) 1976" ~ 1976,  
 H1GI1Y == "(77) (77) 1977" ~ 1977,  
 H1GI1Y == "(78) (78) 1978" ~ 1978,  
 H1GI1Y == "(79) (79) 1979" ~ 1979,  
 H1GI1Y == "(80) (80) 1980" ~ 1980,  
 H1GI1Y == "(81) (81) 1981" ~ 1981,  
 H1GI1Y == "(82) (82) 1982" ~ 1982,  
 H1GI1Y == "(83) (83) 1983" ~ 1983))  
  
  
dat <- dat %>%  
 mutate(IYEAR = case\_when(IYEAR == "(94) (94) 1994" ~ 1994,  
 IYEAR == "(95) (95) 1995" ~ 1995))  
  
dat <- dat %>%  
 mutate(AGE = IYEAR - H1GI1Y, .after = IYEAR)  
  
dat <- dat %>%  
 select(-c(H1GI1Y, IYEAR))  
  
  
  
  
dat <- inner\_join(dat, wave11, by = "AID")  
dat <- inner\_join(dat, dat5, by = "AID")  
  
  
  
#   
# a <- c("ID", "AGE",   
# "D1", "D2", "D3", "D4", "D5",  
# "D6", "D7", "D8", "D9", "D10",   
# "D11", "D12", "D13", "D14", "D15",  
# "D16", "D17", "D18", "D19",  
# "MATCH", "MOVER",   
# "URBAN", "RACE", "RACE0", "HISP",  
# "SEXCOMP", "MED\_AGE", "DISP1", "MARITAL", "DISP2",  
# "UNDER5", "MIGRATE", "DISP3", "HOUSEHOLD", "DISP4",  
# "MEDHHINCOME", "DISP5", "MEDFAMILYINCOME", "DISP6",  
# "PROPPOV",  
# "HHEDU", "DISP7", "HHFEMLABOR", "HHUNEMPLOY", "HHOCCUP",  
# "DISP8", "HHTENURE", "PROPOCCHOUSE", "MEDIANVAL", "DISP9",  
#   
#   
#   
# "DEPR", "AGEDEPR")  
  
  
  
  
descript <- dat %>%  
 select(AGE, BST90P02, BST90P05, BST90P06, BST90P08,  
 BST90P10, BST90P13, BST90P15, BST90P19,  
 BST90P20, BST90P23, BST90P24)  
  
colnames(descript) <- c("age", "race", "hh\_sex", "hh\_median\_age",  
 "hh\_maritalstatus", "childrn\_under\_5",  
 "hh\_type", "med\_hh\_income", "prop\_below\_pov",  
 "hh\_edu", "unemployment", "hh\_occupation")  
   
  
library(table1)

##   
## Attaching package: 'table1'  
##   
## The following objects are masked from 'package:base':  
##   
## units, units<-

table1(~age + race + hh\_sex + hh\_median\_age + hh\_maritalstatus + childrn\_under\_5 + hh\_type + med\_hh\_income + prop\_below\_pov + hh\_edu + unemployment + hh\_occupation, data = descript, overall=c(left="Total"), caption = "Population descriptives", footnote = "\* Mean and SD presented")

## Get nicer `table1` .docx output by simply installing the `flextable` package

## Total  
## 1 (N=4196)  
## 2 age   
## 3 Mean (SD) 16.0 (1.77)  
## 4 Median [Min, Max] 16.0 [12.0, 21.0]  
## 5 race   
## 6 (1) (1) White 3420 (81.5%)  
## 7 (2) (2) Black 595 (14.2%)  
## 8 (3) (3) Other 139 (3.3%)  
## 9 Missing 42 (1.0%)  
## 10 hh\_sex   
## 11 (1) (1) Heavily male 372 (8.9%)  
## 12 (2) (2) Balanced 3415 (81.4%)  
## 13 (3) (3) Heavily female 372 (8.9%)  
## 14 Missing 37 (0.9%)  
## 15 hh\_median\_age   
## 16 Mean (SD) 33.7 (5.75)  
## 17 Median [Min, Max] 33.0 [16.0, 72.0]  
## 18 Missing 42 (1.0%)  
## 19 hh\_maritalstatus   
## 20 (1) (1) Never married 375 (8.9%)  
## 21 (2) (2) Married, spouse present 3758 (89.6%)  
## 22 (3) (3) Separated or divorced 11 (0.3%)  
## 23 Missing 52 (1.2%)  
## 24 childrn\_under\_5   
## 25 (1) (1) Low 468 (11.2%)  
## 26 (2) (2) Medium 3236 (77.1%)  
## 27 (3) (3) High 450 (10.7%)  
## 28 Missing 42 (1.0%)  
## 29 hh\_type   
## 30 (1) (1) Married couple family household 3506 (83.6%)  
## 31 (2) (2) Other family household 174 (4.1%)  
## 32 (3) (3) Non-family household 352 (8.4%)  
## 33 Missing 164 (3.9%)  
## 34 med\_hh\_income   
## 35 Mean (SD) 30900 (14400)  
## 36 Median [Min, Max] 29000 [5000, 100000]  
## 37 Missing 164 (3.9%)  
## 38 prop\_below\_pov   
## 39 (1) (1) Low 2379 (56.7%)  
## 40 (2) (2) Medium 931 (22.2%)  
## 41 (3) (3) High 849 (20.2%)  
## 42 Missing 37 (0.9%)  
## 43 hh\_edu   
## 44 (1) (1) No high school or equivalency 632 (15.1%)  
## 45 (2) (2) High school degree/no college degree 3084 (73.5%)  
## 46 (3) (3) College degree or more 423 (10.1%)  
## 47 Missing 57 (1.4%)  
## 48 unemployment   
## 49 (1) (1) Low 2246 (53.5%)  
## 50 (2) (2) Medium 985 (23.5%)  
## 51 (3) (3) High 855 (20.4%)  
## 52 Missing 110 (2.6%)  
## 53 hh\_occupation   
## 54 (1) (1) Managerial or professional 934 (22.3%)  
## 55 (2) (2) Technical/sales/administrative support 2150 (51.2%)  
## 56 (3) (3) Service occupations 235 (5.6%)  
## 57 (4) (4) Farming/forestry/fishing 28 (0.7%)  
## 58 (5) (5) Production/craft/repair 57 (1.4%)  
## 59 (6) (6) Operators/fabricators/laborers 641 (15.3%)  
## 60 Missing 151 (3.6%)