income\_lookup.R

lmwang

Mon Mar 28 14:20:11 2016

# This script sets up lookup table for income  
  
# hh.metro <- hh %>%   
# mutate(inc.level=cut(INCOME,  
# breaks=c(1, 6, 11, 13),  
# labels=c("lowInc", "midInc", "highInc"), #allow alternative household grouping  
# include.lowest=T, right=F)  
# )  
  
# Portland 1994 income classification  
# lowInc: $0 ~ $24,999; midInc: $25,000 ~ $49,999; highInc: $50,000 ~ or more   
  
# Value Label  
# 1 $0 ‐ $4,999  
# 2 $5,000 ‐ $9,999  
# 3 $10,000 ‐ $14,999  
# 4 $15,000 ‐ $19,999  
# 5 $20,000 ‐ $24,999  
# 6 $25,000 ‐ $29,999  
# 7 $30,000 ‐ $34,999  
# 8 $35,000 ‐ $39,999  
# 9 $40,000 ‐ $44,999  
# 10 $45,000 ‐ $49,999  
# 11 $50,000 ‐ $54,999  
# 12 $55,000 ‐ $59,999  
# 13 $60,000 or more  
# 14 DK/RF  
  
  
# http://data.bls.gov/cgi-bin/cpicalc.pl?cost1=100&year1=1993&year2=2010  
# $100 in 1993 has the same buying power as 150.90 in 2010.   
  
# Income cutoff for OHAS 2011  
# lowInc: $0 ~ 37,724; midInc: $37,725 ~ 75,450; highInc: 75,450 ~ or more   
  
# OHAS 2011 survey data, income field distionary   
# INCOME What is your total household income for 2010?  
# 1 $0 - $14,999  
# 2 $15,000 - $24,999  
# 3 $25,000 - $34,999  
# 4 $35,000 - $49,999  
# 5 $50,000 - $74,999  
# 6 $75,000 - $99,999  
# 7 $100,000 - $149,999  
# 8 $150,000 or more  
# 99 REFUSED  
  
require(dplyr)

## Loading required package: dplyr  
##   
## Attaching package: 'dplyr'  
##   
## The following objects are masked from 'package:stats':  
##   
## filter, lag  
##   
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

# inflation adjust  
 inflation.df <- data.frame(year=c(1992:2010), inflation=NA)  
 rownames(inflation.df) <- as.character(c(1992:2010))   
   
 # Input inflation ratio  
 # Current use the inflaton one year before the survey data year: 2010 used for 2011 OHAS survey data   
 inflation.df["1993", "inflation"] <- 0.9709 # For Salt Lake City 1993 survey   
 inflation.df["1996", "inflation"] <- 1.0858 # For Tampa Bay 1996 survey  
 inflation.df["2009", "inflation"] <- 1.4900 # For NHTS  
 inflation.df["2011", "inflation"] <- 1.5090 # For OHAS  
   
 # Original cutoff data frame based on Portland 1994 income classificaton   
 # lowInc: $0 ~ $24,999; midInc: $25,000 ~ $49,999; highInc: $50,000 ~ or more   
 cutoff.df <- data.frame(ic=c("lowInc", "midInc", "highInc"), low.bound=c(0, 25000, 50000), high.bound=c(24999, 49999, Inf))  
 rownames(cutoff.df) <- c("lowInc", "midInc", "highInc")  
   
 cutoff.df[, c(2,3)] <- cutoff.df[, c(2,3)]\*inflation.df["2011", "inflation"]  
   
# try lookup table   
 income.index <- c(1:8, 99)  
 income.low <- c(0, 15000, 25000, 35000, 50000, 75000, 100000, 150000, NA)  
 income.high <- c(14999, 24999, 34999, 49999, 74999, 99999, 149999, Inf, NA)  
   
 income <- data.frame(income.index, income.low, income.high, year=2011)  
 # Income reclassification   
 income <- income %>%   
 mutate(ic = ifelse(income.high <= cutoff.df["lowInc", "high.bound"], "lowInc", NA)) %>%  
 mutate(ic = ifelse(income.low >= cutoff.df["highInc", "low.bound"], "highInc", ic)) %>%  
 mutate(ic = ifelse(income.low >= cutoff.df["midInc", "low.bound"]&income.high <= cutoff.df[2, "high.bound"], "midInc", ic)) %>%  
 mutate(ic = ifelse(income.low <= cutoff.df["midInc", "low.bound"]&(cutoff.df["midInc", "low.bound"]-income.low) <= (income.high-cutoff.df["midInc", "high.bound"]), "midInc", ic))%>%   
 mutate(ic = ifelse(income.low <= cutoff.df["midInc", "high.bound"]&(cutoff.df["midInc", "high.bound"]-income.low) <= (income.high-cutoff.df["midInc", "high.bound"]), "midInc", ic)) %>%  
 mutate(ic=ifelse((!is.na(income.low))&(is.na(ic)), "midInc", ic))