sim.R

EdwardMcneil

2022-11-18

#  
# R file to import Hong Kong census data ("hk census\_household size and age\_Ddate20210812.xlsx".  
# Uses data in sheet D108e.

# Generates the distributional proportions for various parameters:  
# Namely,   
# household size (HH\_size),   
# household type by size (HH\_type),   
# age of household head (HH\_head),   
# age of household member,   
# number of children (HH\_child) by type and size of household,  
# number of elderly (HH\_elder) by type and size of household.  
#  
  
# Load required R packages  
library(readxl)   
library(tidyr)   
library(dplyr)

library(readr)   
library(ggplot2)   
library(forcats)   
  
rm(list=ls())   
setwd("F:/Edward/Simulation")   
  
  
# Import the age distribution of Hong Kong (from 2016 By-Census) by HH size and HH type   
HK <- read\_excel("hk census\_household size and age\_Ddate20210812.xlsx",   
 sheet="D108e",   
 range="C9:AD122",   
 col\_names = c('Age', paste("Sex",3:20), 1:8,'n')) %>%   
 select(Age, `1`:`8`, n) %>%   
 filter(Age != "Sub-Total") %>%   
 mutate(hh\_size=rep(1:6, each=18),   
 age=fct\_collapse(Age,   
 '0 - 9'=c('0 - 4','5 - 9'),   
 '10 - 19'=c('10 - 14','15 - 19'),   
 '20 - 29'=c('20 - 24','25 - 29'),   
 '30 - 39'=c('30 - 34','35 - 39'),   
 '40 - 49'=c('40 - 44','45 - 49'),   
 '50 - 59'=c('50 - 54','55 - 59'),   
 '60 - 69'=c('60 - 64','65 - 69'),   
 '70 - 79'=c('70 - 74','75 - 79'),   
 '80+'=c('80 - 84','85+')))   
print(HK, n=Inf)

## # A tibble: 108 × 12  
## Age `1` `2` `3` `4` `5` `6` `7` `8` n hh\_size age   
## <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <dbl> <int> <fct>  
## 1 0 - 4 - - - - - - 570 - 570 1 0 - 9  
## 2 5 - 9 - - - - - - 1031 - 1031 1 0 - 9  
## 3 10 - 14 - - - - - - 269 - 269 1 10 -…  
## 4 15 - 19 - - - - - - 1375 - 1375 1 10 -…  
## 5 20 - 24 - - - - - - 7449 - 7449 1 20 -…  
## 6 25 - 29 - - - - - - 21479 - 21479 1 20 -…  
## 7 30 - 34 - - - - - - 28556 - 28556 1 30 -…  
## 8 35 - 39 - - - - - - 27628 - 27628 1 30 -…  
## 9 40 - 44 - - - - - - 32653 - 32653 1 40 -…  
## 10 45 - 49 - - - - - - 39943 - 39943 1 40 -…  
## 11 50 - 54 - - - - - - 49717 - 49717 1 50 -…  
## 12 55 - 59 - - - - - - 50118 - 50118 1 50 -…  
## 13 60 - 64 - - - - - - 45691 - 45691 1 60 -…  
## 14 65 - 69 - - - - - - 43364 - 43364 1 60 -…  
## 15 70 - 74 - - - - - - 27015 - 27015 1 70 -…  
## 16 75 - 79 - - - - - - 29715 - 29715 1 70 -…  
## 17 80 - 84 - - - - - - 26727 - 26727 1 80+   
## 18 85+ - - - - - - 25715 - 25715 1 80+   
## 19 0 - 4 - - 5596 - - 1182 - 223 7001 2 0 - 9  
## 20 5 - 9 - - 7667 - - 2001 - 95 9763 2 0 - 9  
## 21 10 - 14 - - 11257 - - 1302 - 71 12630 2 10 -…  
## 22 15 - 19 249 - 21612 - - 2494 - 423 24778 2 10 -…  
## 23 20 - 24 1825 - 24999 - - 6109 - 4926 37859 2 20 -…  
## 24 25 - 29 21883 - 25756 - - 11489 - 15608 74736 2 20 -…  
## 25 30 - 34 59861 - 23639 - - 11756 - 17485 112741 2 30 -…  
## 26 35 - 39 52626 - 21191 - - 10009 - 13792 97618 2 30 -…  
## 27 40 - 44 53562 - 26975 - - 10367 - 11130 102034 2 40 -…  
## 28 45 - 49 57374 - 29233 - - 12279 - 7886 106772 2 40 -…  
## 29 50 - 54 66544 - 33510 - - 15684 - 6067 121805 2 50 -…  
## 30 55 - 59 77252 - 30922 - - 13272 - 4110 125556 2 50 -…  
## 31 60 - 64 88038 - 24029 - - 11822 - 2825 126714 2 60 -…  
## 32 65 - 69 92810 - 18296 - - 9757 - 2745 123608 2 60 -…  
## 33 70 - 74 56905 - 10770 - - 5383 - 1928 74986 2 70 -…  
## 34 75 - 79 49806 - 12209 - - 5791 - 3498 71304 2 70 -…  
## 35 80 - 84 31676 - 11578 - - 6045 - 6190 55489 2 80+   
## 36 85+ 15439 - 11311 - - 8550 - 10986 46286 2 80+   
## 37 0 - 4 47 48212 7606 - - 4357 - 224 60446 3 0 - 9  
## 38 5 - 9 21 39756 11918 - - 4663 - 25 56383 3 0 - 9  
## 39 10 - 14 10 39295 14309 - - 4972 - 37 58623 3 10 -…  
## 40 15 - 19 25 58534 23015 - - 5210 - 117 86901 3 10 -…  
## 41 20 - 24 635 63840 33331 253 - 8568 - 1427 108054 3 20 -…  
## 42 25 - 29 3307 70437 31718 1587 - 15116 - 2394 124559 3 20 -…  
## 43 30 - 34 5917 88409 25027 4124 - 18986 - 1098 143561 3 30 -…  
## 44 35 - 39 6519 89623 18858 4250 - 17444 - 1211 137905 3 30 -…  
## 45 40 - 44 4686 98230 19752 4448 - 14407 - 1174 142697 3 40 -…  
## 46 45 - 49 4119 1039… 21375 4755 - 13771 - 774 148726 3 40 -…  
## 47 50 - 54 4672 1253… 23892 4754 - 17768 - 918 177394 3 50 -…  
## 48 55 - 59 5532 1313… 19254 3974 - 21098 - 645 181843 3 50 -…  
## 49 60 - 64 6153 98089 12079 3792 - 21666 - 527 142306 3 60 -…  
## 50 65 - 69 7231 67762 6650 3735 - 20371 - 249 105998 3 60 -…  
## 51 70 - 74 4597 29707 3676 2022 - 11925 - 119 52046 3 70 -…  
## 52 75 - 79 5410 21596 4017 2289 - 10392 - 100 43804 3 70 -…  
## 53 80 - 84 4758 11735 4873 2731 - 8068 - 261 32426 3 80+   
## 54 85+ 4155 5299 6737 4113 - 9620 - 871 30795 3 80+   
## 55 0 - 4 20 73459 3221 - 5781 8672 - - 91153 4 0 - 9  
## 56 5 - 9 - 82799 4484 8 3957 7138 - 26 98412 4 0 - 9  
## 57 10 - 14 - 80871 5217 9 3597 6427 - 30 96151 4 10 -…  
## 58 15 - 19 9 1118… 7732 8 4047 7205 - 25 130841 4 10 -…  
## 59 20 - 24 176 1426… 10440 294 3486 12485 - 781 170339 4 20 -…  
## 60 25 - 29 458 1218… 10203 3469 3710 21583 - 700 161998 4 20 -…  
## 61 30 - 34 803 1086… 7196 6053 6461 25444 - 259 154871 4 30 -…  
## 62 35 - 39 818 1161… 6533 4928 7471 18937 - 166 155044 4 30 -…  
## 63 40 - 44 637 1166… 5842 3698 7898 14520 - 210 149477 4 40 -…  
## 64 45 - 49 525 1220… 6357 3450 8074 12699 - 178 153319 4 40 -…  
## 65 50 - 54 646 1446… 6907 3110 8021 16940 - 65 180292 4 50 -…  
## 66 55 - 59 496 1287… 4725 4702 6928 20962 - 123 166699 4 50 -…  
## 67 60 - 64 543 68543 2841 5371 5072 20538 - 33 102941 4 60 -…  
## 68 65 - 69 672 31563 1330 4826 3816 17525 - 41 59773 4 60 -…  
## 69 70 - 74 599 13113 722 2294 3395 8223 - 61 28407 4 70 -…  
## 70 75 - 79 437 8196 875 2425 3959 7174 - 11 23077 4 70 -…  
## 71 80 - 84 524 4741 1001 2144 4173 5230 - 52 17865 4 80+   
## 72 85+ 425 2672 1406 2787 5238 6050 - 95 18673 4 80+   
## 73 0 - 4 - 47104 878 7 13968 8579 - - 70536 5 0 - 9  
## 74 5 - 9 - 58857 1261 - 12171 6662 - - 78951 5 0 - 9  
## 75 10 - 14 - 43804 982 66 8733 5324 - - 58909 5 10 -…  
## 76 15 - 19 - 44192 2085 18 9576 5892 - - 61763 5 10 -…  
## 77 20 - 24 22 51011 2694 149 7898 9301 - 45 71120 5 20 -…  
## 78 25 - 29 216 50166 2379 550 8299 15666 - 411 77687 5 20 -…  
## 79 30 - 34 185 47843 1665 1234 14417 17755 - 206 83305 5 30 -…  
## 80 35 - 39 198 65766 1258 681 15457 11999 - 204 95563 5 30 -…  
## 81 40 - 44 144 64567 1311 666 15053 10047 - 53 91841 5 40 -…  
## 82 45 - 49 50 49034 1378 732 14453 9758 - 69 75474 5 40 -…  
## 83 50 - 54 188 45193 1156 463 12889 10574 - - 70463 5 50 -…  
## 84 55 - 59 180 34015 932 669 9808 12100 - 65 57769 5 50 -…  
## 85 60 - 64 74 16858 355 745 8714 11033 - 10 37789 5 60 -…  
## 86 65 - 69 61 6283 219 709 9684 8145 - 56 25157 5 60 -…  
## 87 70 - 74 95 2819 127 373 7242 4522 - 12 15190 5 70 -…  
## 88 75 - 79 160 1631 156 352 8686 4040 - 90 15115 5 70 -…  
## 89 80 - 84 75 889 108 302 6431 2940 - - 10745 5 80+   
## 90 85+ 97 518 211 524 6301 2773 - 84 10508 5 80+   
## 91 0 - 4 - 14293 400 - 14882 17230 - - 46805 6 0 - 9  
## 92 5 - 9 - 16754 700 - 15258 9660 - - 42372 6 0 - 9  
## 93 10 - 14 16 12662 433 - 10028 5826 - 40 29005 6 10 -…  
## 94 15 - 19 16 13615 528 - 7588 5799 - 24 27570 6 10 -…  
## 95 20 - 24 - 14990 402 12 6321 9829 - 9 31563 6 20 -…  
## 96 25 - 29 61 14140 347 62 8068 19127 - 438 42243 6 20 -…  
## 97 30 - 34 121 13791 311 69 12423 20728 - 192 47635 6 30 -…  
## 98 35 - 39 112 16370 457 78 17282 15561 - 176 50036 6 30 -…  
## 99 40 - 44 30 15941 444 92 15086 10398 - 24 42015 6 40 -…  
## 100 45 - 49 39 12268 389 75 10584 8744 - 6 32105 6 40 -…  
## 101 50 - 54 31 10648 299 43 7516 9949 - 153 28639 6 50 -…  
## 102 55 - 59 4 7239 191 47 6246 10717 - 177 24621 6 50 -…  
## 103 60 - 64 15 3173 134 152 8312 10106 - 119 22011 6 60 -…  
## 104 65 - 69 73 1538 79 225 9531 7767 - 82 19295 6 60 -…  
## 105 70 - 74 33 464 39 7 6486 4191 - 6 11226 6 70 -…  
## 106 75 - 79 31 339 22 8 6637 2964 - 13 10014 6 70 -…  
## 107 80 - 84 38 90 34 20 4828 2372 - 4 7386 6 80+   
## 108 85+ 8 75 83 138 4054 2853 - 5 7216 6 80+

# Import the age distribution of Hong Kong by age group  
HH\_age <- read\_excel("hk census\_household size and age\_Ddate20210812.xlsx",   
 sheet="D108e",   
 range="AD123:AD140",   
 col\_names = 'n') %>%   
 mutate(age=rep(1:9, each=2)) %>%   
 group\_by(age) %>%   
 summarise(n=sum(n), .groups='drop') %>%   
 mutate(p=n/sum(n))   
HH\_age

## # A tibble: 9 × 3  
## age n p  
## <int> <dbl> <dbl>  
## 1 1 563423 0.0792  
## 2 2 588815 0.0828  
## 3 3 929086 0.131   
## 4 4 1134463 0.159   
## 5 5 1117056 0.157   
## 6 6 1234916 0.174   
## 7 7 854647 0.120   
## 8 8 401899 0.0565  
## 9 9 289831 0.0407

# 1. Household sizes (sheet D103ae)  
HH\_size <- read\_excel("hk census\_household size and age\_Ddate20210812.xlsx",   
 sheet="D103ae",   
 range="F6:F11",   
 col\_names = 'n') %>%   
 mutate(hh\_size=1:6, p=n/sum(n)) %>%   
 select(hh\_size, p)   
  
HH\_size

## # A tibble: 6 × 2  
## hh\_size p  
## <int> <dbl>  
## 1 1 0.183   
## 2 2 0.265   
## 3 3 0.244   
## 4 4 0.195   
## 5 5 0.0803  
## 6 6 0.0327

# 2. Age range of household head (sheet D121e)  
HH\_head <- read\_excel("hk census\_household size and age\_Ddate20210812.xlsx",   
 sheet="D121e",   
 range="D293:J302",   
 col\_names = c('Age', 1:6)) %>%   
 pivot\_longer(2:7, names\_to="hh\_size") %>%   
 mutate(age=recode(Age,   
 "0 - 24"=1,   
 "25 - 29"=2,   
 "30 - 34"=3,   
 "35 - 39"=4,   
 "40 - 44"=5,   
 "45 - 49"=6,   
 "50 - 54"=7,   
 "55 - 59"=8,   
 "60 - 64"=9,   
 "65+"=10,   
 )) %>%   
 group\_by(hh\_size) %>%   
 mutate(p=value/sum(value)) %>%   
 arrange(hh\_size)   
  
HH\_head

## # A tibble: 60 × 5  
## # Groups: hh\_size [6]  
## Age hh\_size value age p  
## <chr> <chr> <dbl> <dbl> <dbl>  
## 1 0 - 24 1 10694 1 0.0233  
## 2 25 - 29 1 21479 2 0.0468  
## 3 30 - 34 1 28556 3 0.0622  
## 4 35 - 39 1 27628 4 0.0602  
## 5 40 - 44 1 32653 5 0.0711  
## 6 45 - 49 1 39943 6 0.0870  
## 7 50 - 54 1 49717 7 0.108   
## 8 55 - 59 1 50118 8 0.109   
## 9 60 - 64 1 45691 9 0.0995  
## 10 65+ 1 152536 10 0.332   
## # … with 50 more rows

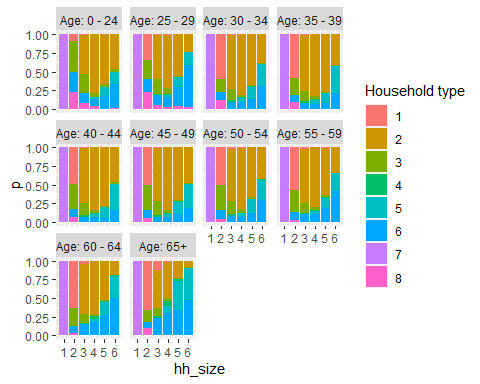
# 3. Age range of HH head by household type   
HH\_type <- read\_excel("hk census\_household size and age\_Ddate20210812.xlsx",   
 sheet="D121e", range="C7:J269",   
 col\_names = c('Sex','Age', 1:6)) %>%   
 fill(Sex) %>%   
 filter(Sex=="Sub-Total" & Age!="Sub-Total") %>%   
 mutate(hh\_type=rep(1:8, each=10)) %>%   
 select(-Sex) %>%   
 pivot\_longer(2:7, names\_to="hh\_size", values\_to="n") %>%   
 mutate(n=ifelse(n=="-", 0, n)) %>%   
 type\_convert() %>%   
 group\_by(Age, hh\_size) %>% # probabilities in each age/hh\_size must add to 1   
 mutate(p=if\_else(n>0, n/sum(n), 0))

##   
## ── Column specification ────────────────────────────────────────────────────────  
## cols(  
## Age = col\_character(),  
## hh\_size = col\_double(),  
## n = col\_double()  
## )

HH\_type

## # A tibble: 480 × 5  
## # Groups: Age, hh\_size [60]  
## Age hh\_type hh\_size n p  
## <chr> <int> <dbl> <dbl> <dbl>  
## 1 0 - 24 1 1 0 0   
## 2 0 - 24 1 2 1109 0.0857   
## 3 0 - 24 1 3 55 0.00443  
## 4 0 - 24 1 4 0 0   
## 5 0 - 24 1 5 0 0   
## 6 0 - 24 1 6 0 0   
## 7 25 - 29 1 1 0 0   
## 8 25 - 29 1 2 12599 0.347   
## 9 25 - 29 1 3 207 0.00878  
## 10 25 - 29 1 4 88 0.00526  
## # … with 470 more rows

# hh\_types (according to census)  
# 1=couple, 2=couple+child, 3=parent+child, 4=couple+parent,   
# 5=couple+parent+child, 6=other, 7=single, 8=complex   
  
  
# Visualize the distribution of household types and size by age group  
ggplot(HH\_type, aes(x=hh\_size, y=p, fill=factor(hh\_type))) +   
 facet\_wrap(~Age, labeller='label\_both') +   
 geom\_col() +   
 scale\_x\_continuous(breaks=1:6) +   
 labs(fill="Household type")



# Proportion of household types in Hong Kong  
HH\_types <- group\_by(HH\_type, hh\_type) %>%   
 summarise(n=sum(n), .groups='drop') %>%   
 mutate(p=n/sum(n))   
  
HH\_types # type 2 (couple+child) is the most common in Hong Kong

## # A tibble: 8 × 3  
## hh\_type n p  
## <int> <dbl> <dbl>  
## 1 1 492561 0.168   
## 2 2 1122558 0.383   
## 3 3 327294 0.112   
## 4 4 36074 0.0123  
## 5 5 106233 0.0363  
## 6 6 314841 0.108   
## 7 7 459015 0.157   
## 8 8 70062 0.0239

# 4. Age of partner, if household type = {2, 4, 5}   
# AFAIK, the census does not provide this data, so we assume that the difference in age between the household head and partner is normally distributed with a mean of 0 and sd 2 yrs  
  
  
# 5. Age of child(ren)   
# Use census data to determine the distribution of ages for each combination of hh type and size  
HH\_child <- read\_excel("hk census\_household size and age\_Ddate20210812.xlsx",   
 sheet="D105e", range="C7:H75",   
 col\_names = c('hh\_type', 'hh\_size', 0:3)) %>%   
 fill(hh\_type) %>%   
 filter(hh\_size!="Sub-Total" & hh\_type!="Sub-Total") %>%   
 mutate(hh\_type=sub('.+\\((\\d)\\)', '\\1', hh\_type)) %>%   
 pivot\_longer(3:6, names\_to="child", values\_to="n") %>%   
 mutate(n=ifelse(n=="-", 0, n),   
 hh\_size=recode(hh\_size, "6 and over"='6')) %>%   
 type\_convert() %>%   
 filter(n>0) %>%   
 group\_by(hh\_type, hh\_size) %>%   
 # probabilities in each hh\_type/hh\_size must add to 1   
 mutate(p=if\_else(n>0, n/sum(n), 0)) %>%   
 select(-n)

##   
## ── Column specification ────────────────────────────────────────────────────────  
## cols(  
## hh\_type = col\_double(),  
## hh\_size = col\_double(),  
## child = col\_double(),  
## n = col\_double()  
## )

print(HH\_child, n=Inf)

## # A tibble: 86 × 4  
## # Groups: hh\_type, hh\_size [32]  
## hh\_type hh\_size child p  
## <dbl> <dbl> <dbl> <dbl>  
## 1 1 2 0 1   
## 2 1 3 0 0.997   
## 3 1 3 1 0.00345  
## 4 1 4 0 0.990   
## 5 1 4 1 0.0103   
## 6 1 5 0 1   
## 7 1 6 0 0.955   
## 8 1 6 3 0.0455   
## 9 2 3 0 0.679   
## 10 2 3 1 0.321   
## 11 2 4 0 0.563   
## 12 2 4 1 0.233   
## 13 2 4 2 0.204   
## 14 2 5 0 0.373   
## 15 2 5 1 0.133   
## 16 2 5 2 0.428   
## 17 2 5 3 0.0660   
## 18 2 6 0 0.302   
## 19 2 6 1 0.125   
## 20 2 6 2 0.288   
## 21 2 6 3 0.286   
## 22 3 2 0 0.860   
## 23 3 2 1 0.140   
## 24 3 3 0 0.752   
## 25 3 3 1 0.143   
## 26 3 3 2 0.105   
## 27 3 4 0 0.666   
## 28 3 4 1 0.126   
## 29 3 4 2 0.157   
## 30 3 4 3 0.0513   
## 31 3 5 0 0.597   
## 32 3 5 1 0.131   
## 33 3 5 2 0.150   
## 34 3 5 3 0.123   
## 35 3 6 0 0.382   
## 36 3 6 1 0.0776   
## 37 3 6 2 0.111   
## 38 3 6 3 0.430   
## 39 4 3 0 1   
## 40 4 4 0 0.999   
## 41 4 4 1 0.00137  
## 42 4 5 0 0.973   
## 43 4 5 1 0.00910  
## 44 4 5 2 0.0176   
## 45 4 6 0 1   
## 46 5 4 0 0.439   
## 47 5 4 1 0.561   
## 48 5 5 0 0.306   
## 49 5 5 1 0.468   
## 50 5 5 2 0.225   
## 51 5 6 0 0.199   
## 52 5 6 1 0.215   
## 53 5 6 2 0.508   
## 54 5 6 3 0.0770   
## 55 6 2 0 0.948   
## 56 6 2 1 0.0421   
## 57 6 2 2 0.00983  
## 58 6 3 0 0.831   
## 59 6 3 1 0.156   
## 60 6 3 2 0.0116   
## 61 6 3 3 0.00139  
## 62 6 4 0 0.684   
## 63 6 4 1 0.260   
## 64 6 4 2 0.0549   
## 65 6 4 3 0.00140  
## 66 6 5 0 0.515   
## 67 6 5 1 0.330   
## 68 6 5 2 0.141   
## 69 6 5 3 0.0141   
## 70 6 6 0 0.226   
## 71 6 6 1 0.394   
## 72 6 6 2 0.306   
## 73 6 6 3 0.0732   
## 74 7 1 0 0.996   
## 75 7 1 1 0.00407  
## 76 8 2 0 0.993   
## 77 8 2 1 0.00707  
## 78 8 3 0 0.951   
## 79 8 3 1 0.0276   
## 80 8 3 2 0.0214   
## 81 8 4 0 0.936   
## 82 8 4 1 0.0504   
## 83 8 4 2 0.0140   
## 84 8 5 0 1   
## 85 8 6 0 0.965   
## 86 8 6 3 0.0346

# Households with members aged 65 and over   
HH\_elder <- read\_excel("hk census\_household size and age\_Ddate20210812.xlsx",   
 sheet="D106e", range="C7:H75",   
 col\_names = c('hh\_type', 'hh\_size', 0:3)) %>%   
 fill(hh\_type) %>%   
 filter(hh\_size!="Sub-Total" & hh\_type!="Sub-Total") %>%   
 mutate(hh\_type=sub('.+\\((\\d)\\)', '\\1', hh\_type)) %>%   
 pivot\_longer(3:6, names\_to="elder", values\_to="n") %>%   
 mutate(n=ifelse(n=="-", 0, n),   
 hh\_size=recode(hh\_size, "6 and over"='6')) %>%   
 type\_convert() %>%   
 filter(n>0) %>%   
 group\_by(hh\_type, hh\_size) %>% # probabilities in each hh\_type/hh\_size must add to 1   
 mutate(p=if\_else(n>0, n/sum(n), 0)) %>%   
 select(-n)

##   
## ── Column specification ────────────────────────────────────────────────────────  
## cols(  
## hh\_type = col\_double(),  
## hh\_size = col\_double(),  
## elder = col\_double(),  
## n = col\_double()  
## )

print(HH\_elder, n=Inf)

## # A tibble: 119 × 4  
## # Groups: hh\_type, hh\_size [32]  
## hh\_type hh\_size elder p  
## <dbl> <dbl> <dbl> <dbl>  
## 1 1 2 0 0.600   
## 2 1 2 1 0.120   
## 3 1 2 2 0.280   
## 4 1 3 0 0.362   
## 5 1 3 1 0.123   
## 6 1 3 2 0.511   
## 7 1 3 3 0.00416   
## 8 1 4 0 0.266   
## 9 1 4 1 0.123   
## 10 1 4 2 0.593   
## 11 1 4 3 0.0185   
## 12 1 5 0 0.278   
## 13 1 5 1 0.109   
## 14 1 5 2 0.550   
## 15 1 5 3 0.0630   
## 16 1 6 0 0.0455   
## 17 1 6 2 0.784   
## 18 1 6 3 0.170   
## 19 2 3 0 0.772   
## 20 2 3 1 0.113   
## 21 2 3 2 0.115   
## 22 2 3 3 0.000295   
## 23 2 4 0 0.882   
## 24 2 4 1 0.0722   
## 25 2 4 2 0.0452   
## 26 2 4 3 0.0000919  
## 27 2 5 0 0.928   
## 28 2 5 1 0.0475   
## 29 2 5 2 0.0240   
## 30 2 5 3 0.000254   
## 31 2 6 0 0.930   
## 32 2 6 1 0.0491   
## 33 2 6 2 0.0177   
## 34 2 6 3 0.00274   
## 35 3 2 0 0.641   
## 36 3 2 1 0.352   
## 37 3 2 2 0.00701   
## 38 3 3 0 0.738   
## 39 3 3 1 0.255   
## 40 3 3 2 0.00765   
## 41 3 3 3 0.000125   
## 42 3 4 0 0.769   
## 43 3 4 1 0.220   
## 44 3 4 2 0.00915   
## 45 3 4 3 0.00234   
## 46 3 5 0 0.795   
## 47 3 5 1 0.196   
## 48 3 5 2 0.00914   
## 49 3 6 0 0.771   
## 50 3 6 1 0.169   
## 51 3 6 2 0.0418   
## 52 3 6 3 0.0179   
## 53 4 3 0 0.187   
## 54 4 3 1 0.725   
## 55 4 3 2 0.0350   
## 56 4 3 3 0.0530   
## 57 4 4 0 0.252   
## 58 4 4 1 0.364   
## 59 4 4 2 0.350   
## 60 4 4 3 0.0345   
## 61 4 5 0 0.268   
## 62 4 5 1 0.160   
## 63 4 5 2 0.504   
## 64 4 5 3 0.0680   
## 65 4 6 0 0.0727   
## 66 4 6 1 0.176   
## 67 4 6 2 0.345   
## 68 4 6 3 0.406   
## 69 5 4 0 0.180   
## 70 5 4 1 0.785   
## 71 5 4 2 0.0239   
## 72 5 4 3 0.0112   
## 73 5 5 0 0.180   
## 74 5 5 1 0.636   
## 75 5 5 2 0.177   
## 76 5 5 3 0.00682   
## 77 5 6 0 0.191   
## 78 5 6 1 0.472   
## 79 5 6 2 0.329   
## 80 5 6 3 0.00701   
## 81 6 2 0 0.559   
## 82 6 2 1 0.393   
## 83 6 2 2 0.0481   
## 84 6 3 0 0.426   
## 85 6 3 1 0.362   
## 86 6 3 2 0.204   
## 87 6 3 3 0.00782   
## 88 6 4 0 0.474   
## 89 6 4 1 0.319   
## 90 6 4 2 0.198   
## 91 6 4 3 0.00924   
## 92 6 5 0 0.497   
## 93 6 5 1 0.303   
## 94 6 5 2 0.191   
## 95 6 5 3 0.00936   
## 96 6 6 0 0.473   
## 97 6 6 1 0.312   
## 98 6 6 2 0.193   
## 99 6 6 3 0.0212   
## 100 7 1 0 0.668   
## 101 7 1 1 0.332   
## 102 8 2 0 0.555   
## 103 8 2 1 0.428   
## 104 8 2 2 0.0163   
## 105 8 3 0 0.652   
## 106 8 3 1 0.301   
## 107 8 3 2 0.0468   
## 108 8 4 0 0.772   
## 109 8 4 1 0.165   
## 110 8 4 2 0.0126   
## 111 8 4 3 0.0504   
## 112 8 5 0 0.464   
## 113 8 5 1 0.199   
## 114 8 5 2 0.284   
## 115 8 5 3 0.0536   
## 116 8 6 0 0.662   
## 117 8 6 1 0.273   
## 118 8 6 2 0.0173   
## 119 8 6 3 0.0476

# Load the algorithm to simulate households  
source("simulate.R")   
  
# Run based on above probabilities generated by the census (takes around 8-10 minutes for n=10,000)  
HK\_HH <- simulate(n=10000, HH\_size, HH\_type, HH\_head, HH\_child, HH\_elder, HH\_age, age.threshold=0.005)   
  
  
# Save to a file  
saveRDS(HK\_HH, file=*"Simulate\_2022.Rds"*)  
  
  
# Or can load a previously saved simulation result ####   
HK\_HH <- readRDS("Simulate\_2022.Rds")   
  
  
# We can check marginal proportions and compare simulation results with the actual population   
# Household sizes (population proportions are in HH\_size$p)   
p\_size <- tabulate(rowSums(HK\_HH$hh)) / sum(tabulate(rowSums(HK\_HH$hh)))   
tibble(HH\_size, p\_size, diff=round(p\_size-HH\_size$p, 4))

## # A tibble: 6 × 4  
## hh\_size p p\_size diff  
## <int> <dbl> <dbl> <dbl>  
## 1 1 0.183 0.178 -0.0044  
## 2 2 0.265 0.299 0.0333  
## 3 3 0.244 0.228 -0.0154  
## 4 4 0.195 0.185 -0.01   
## 5 5 0.0803 0.0805 0.0002  
## 6 6 0.0327 0.029 -0.0037

# Sum of squared residuals  
sum((HH\_size$p - p\_size)^2)

## [1] 0.001479533

# Household member ages (population proportions are in HH\_age$p)   
p\_age <- colSums(HK\_HH$hh) / sum(colSums(HK\_HH$hh))   
tibble(HH\_age, p\_age, diff=round(p\_age-HH\_age$p, 3))

## # A tibble: 9 × 5  
## age n p p\_age diff  
## <int> <dbl> <dbl> <dbl> <dbl>  
## 1 1 563423 0.0792 0.0792 0  
## 2 2 588815 0.0828 0.0828 0  
## 3 3 929086 0.131 0.130 0  
## 4 4 1134463 0.159 0.159 0  
## 5 5 1117056 0.157 0.157 0  
## 6 6 1234916 0.174 0.174 0  
## 7 7 854647 0.120 0.120 0  
## 8 8 401899 0.0565 0.0565 0  
## 9 9 289831 0.0407 0.0408 0

sum((HH\_age$p - p\_age)^2)

## [1] 4.490981e-08

# Household types   
p\_type <- tabulate(unlist(sapply(HK\_HH$type, unlist)))   
p\_type <- p\_type/sum(p\_type)   
tibble(HH\_types, p\_type, diff=round(p\_type-HH\_types$p, 3))

## # A tibble: 8 × 5  
## hh\_type n p p\_type diff  
## <int> <dbl> <dbl> <dbl> <dbl>  
## 1 1 492561 0.168 0.162 -0.006  
## 2 2 1122558 0.383 0.364 -0.019  
## 3 3 327294 0.112 0.116 0.004  
## 4 4 36074 0.0123 0.0121 0   
## 5 5 106233 0.0363 0.0294 -0.007  
## 6 6 314841 0.108 0.108 0.001  
## 7 7 459015 0.157 0.178 0.022  
## 8 8 70062 0.0239 0.0291 0.005

sum((HH\_types$p - p\_type)^2)

## [1] 0.0009498571

# Assess most common age groups in each household of sizes 1-6   
# Households of size 1   
hh\_1 <- colSums(HK\_HH$hh[rowSums(HK\_HH$hh)==1,])   
hh\_1\_p <- hh\_1/sum(hh\_1)   
  
filter(HK, hh\_size==1) %>%   
 group\_by(age) %>%   
 summarise(n=sum(n), .groups = 'drop') %>%   
 mutate(p\_pop=n/sum(n), p\_sample=hh\_1\_p, diff=p\_sample-p\_pop) %>% print() %>%   
 summarise(SS=sum(diff^2))

## # A tibble: 9 × 5  
## age n p\_pop p\_sample diff  
## <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 0 - 9 1601 0.00349 0 -0.00349  
## 2 10 - 19 1644 0.00358 0 -0.00358  
## 3 20 - 29 28928 0.0630 0.0751 0.0120   
## 4 30 - 39 56184 0.122 0.155 0.0328   
## 5 40 - 49 72596 0.158 0.173 0.0150   
## 6 50 - 59 99835 0.217 0.213 -0.00405  
## 7 60 - 69 89055 0.194 0.220 0.0262   
## 8 70 - 79 56730 0.124 0.114 -0.00931  
## 9 80+ 52442 0.114 0.0487 -0.0655

## # A tibble: 1 × 1  
## SS  
## <dbl>  
## 1 0.00655

# The most common age groups for households of size 1 in HK are 60-69 (22.0%) and 50-59 (21.3%), which are similar to the population proportions (19.4% and 21.7%).   
  
  
# Households of size 2   
hh\_2 <- colSums(HK\_HH$hh[rowSums(HK\_HH$hh)==2,])   
hh\_2\_p <- hh\_2/sum(hh\_2)   
  
filter(HK, hh\_size==2) %>%   
 group\_by(age) %>%   
 summarise(n=sum(n), .groups = 'drop') %>%   
 mutate(p\_pop=n/sum(n), p\_sample=hh\_2\_p, diff=p\_sample-p\_pop) %>% print() %>%   
 summarise(SS=sum(diff^2))

## # A tibble: 9 × 5  
## age n p\_pop p\_sample diff  
## <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 0 - 9 16764 0.0126 0.0226 0.0100   
## 2 10 - 19 37408 0.0281 0.0305 0.00238  
## 3 20 - 29 112595 0.0846 0.0916 0.00704  
## 4 30 - 39 210359 0.158 0.166 0.00764  
## 5 40 - 49 208806 0.157 0.165 0.00847  
## 6 50 - 59 247361 0.186 0.175 -0.0104   
## 7 60 - 69 250322 0.188 0.194 0.00593  
## 8 70 - 79 146290 0.110 0.0988 -0.0111   
## 9 80+ 101775 0.0764 0.0564 -0.0200

## # A tibble: 1 × 1  
## SS  
## <dbl>  
## 1 0.000952

# The most common age groups for households of size 2 are 60-69 (19.4%) and 50-59 (17.5%), which are similar to the population proportions (18.8% and 18.6%)   
  
  
# Households of size 3   
hh\_3 <- colSums(HK\_HH$hh[rowSums(HK\_HH$hh)==3,])   
hh\_3\_p <- hh\_3/sum(hh\_3)   
  
filter(HK, hh\_size==3) %>%   
 group\_by(age) %>%   
 summarise(n=sum(n), .groups = 'drop') %>%   
 mutate(p\_pop=n/sum(n), p\_sample=hh\_3\_p, diff=p\_sample-p\_pop) %>% print() %>%   
 summarise(SS=sum(diff^2))

## # A tibble: 9 × 5  
## age n p\_pop p\_sample diff  
## <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 0 - 9 116829 0.0637 0.0846 0.0209   
## 2 10 - 19 145524 0.0793 0.0818 0.00247   
## 3 20 - 29 232613 0.127 0.126 -0.000596  
## 4 30 - 39 281466 0.153 0.171 0.0173   
## 5 40 - 49 291423 0.159 0.165 0.00649   
## 6 50 - 59 359237 0.196 0.187 -0.00900   
## 7 60 - 69 248304 0.135 0.111 -0.0239   
## 8 70 - 79 95850 0.0522 0.0443 -0.00799   
## 9 80+ 63221 0.0345 0.0288 -0.00569

## # A tibble: 1 × 1  
## SS  
## <dbl>  
## 1 0.00153

# The most common age group for households of size 3 is 50-59 (18.7%), which is similar to population proportion (19.6%)   
  
  
# Households of size 4   
hh\_4 <- colSums(HK\_HH$hh[rowSums(HK\_HH$hh)==4,])   
hh\_4\_p <- hh\_3/sum(hh\_4)   
  
filter(HK, hh\_size==4) %>%   
 group\_by(age) %>%   
 summarise(n=sum(n), .groups = 'drop') %>%   
 mutate(p\_pop=n/sum(n), p\_sample=hh\_4\_p, diff=p\_sample-p\_pop) %>% print() %>%   
 summarise(SS=sum(diff^2))

## # A tibble: 9 × 5  
## age n p\_pop p\_sample diff  
## <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 0 - 9 189565 0.0967 0.0782 -0.0186   
## 2 10 - 19 226992 0.116 0.0756 -0.0403   
## 3 20 - 29 332337 0.170 0.117 -0.0530   
## 4 30 - 39 309915 0.158 0.158 -0.000371  
## 5 40 - 49 302796 0.155 0.153 -0.00173   
## 6 50 - 59 346991 0.177 0.173 -0.00445   
## 7 60 - 69 162714 0.0830 0.103 0.0200   
## 8 70 - 79 51484 0.0263 0.0409 0.0146   
## 9 80+ 36538 0.0186 0.0266 0.00794

## # A tibble: 1 × 1  
## SS  
## <dbl>  
## 1 0.00547

# The most common age group for households of size 4 is 50-59 (17.3%), which is similar to population proportion (17.7%)   
  
  
# Households of size 5   
hh\_5 <- colSums(HK\_HH$hh[rowSums(HK\_HH$hh)==5,])   
hh\_5\_p <- hh\_5/sum(hh\_5)   
  
filter(HK, hh\_size==5) %>%   
 group\_by(age) %>%   
 summarise(n=sum(n), .groups = 'drop') %>%   
 mutate(p\_pop=n/sum(n), p\_sample=hh\_5\_p, diff=p\_sample-p\_pop) %>% print() %>%   
 summarise(SS=sum(diff^2))

## # A tibble: 9 × 5  
## age n p\_pop p\_sample diff  
## <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 0 - 9 149487 0.148 0.125 -0.0233   
## 2 10 - 19 120672 0.120 0.131 0.0110   
## 3 20 - 29 148807 0.148 0.155 0.00764  
## 4 30 - 39 178868 0.177 0.153 -0.0244   
## 5 40 - 49 167315 0.166 0.146 -0.0197   
## 6 50 - 59 128232 0.127 0.134 0.00693  
## 7 60 - 69 62946 0.0625 0.0678 0.00537  
## 8 70 - 79 30305 0.0301 0.0393 0.00919  
## 9 80+ 21253 0.0211 0.0484 0.0274

## # A tibble: 1 × 1  
## SS  
## <dbl>  
## 1 0.00262

# The sample proportions are not too dissimilar to the population proportions  
  
  
# Households of size 6   
hh\_6 <- colSums(HK\_HH$hh[rowSums(HK\_HH$hh)==6,])   
hh\_6\_p <- hh\_6/sum(hh\_6)   
  
filter(HK, hh\_size==6) %>%   
 group\_by(age) %>%   
 summarise(n=sum(n), .groups = 'drop') %>%   
 mutate(p\_pop=n/sum(n), p\_sample=hh\_6\_p, diff=p\_sample-p\_pop) %>% print() %>%   
 summarise(SS=sum(diff^2))

## # A tibble: 9 × 5  
## age n p\_pop p\_sample diff  
## <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 0 - 9 89177 0.171 0.117 -0.0543   
## 2 10 - 19 56575 0.108 0.114 0.00594  
## 3 20 - 29 73806 0.141 0.163 0.0212   
## 4 30 - 39 97671 0.187 0.138 -0.0493   
## 5 40 - 49 74120 0.142 0.129 -0.0133   
## 6 50 - 59 53260 0.102 0.123 0.0209   
## 7 60 - 69 41306 0.0792 0.0885 0.00934  
## 8 70 - 79 21240 0.0407 0.0563 0.0156   
## 9 80+ 14602 0.0280 0.0718 0.0439

## # A tibble: 1 × 1  
## SS  
## <dbl>  
## 1 0.00872

# The sample proportions are not too dissimilar to the population proportions  
  
  
# The sum of squared differences between the sample and population age proportions for households of size 1-6 are 0.007, 0.001, 0.002, 0.005, 0.003, and 0.008, respectively.

# Age mixing patterns ####   
Sim <- tibble(as.data.frame(HK\_HH$hh))   
head(Sim)

## # A tibble: 6 × 9  
## `0 - 9` `10 - 19` `20 - 29` `30 - 39` 40 - 4…¹ 50 - …² 60 - …³ 70 - …⁴ 80 - …⁵  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 0 0 0 0 0 0 0 2 0  
## 2 0 0 2 0 0 0 0 0 0  
## 3 0 0 1 1 0 2 0 0 0  
## 4 1 0 0 1 1 0 0 0 0  
## 5 0 0 0 0 1 0 0 0 0  
## 6 0 0 0 1 0 0 0 0 0  
## # … with abbreviated variable names ¹​`40 - 49`, ²​`50 - 59`, ³​`60 - 69`,  
## # ⁴​`70 - 79`, ⁵​`80 - 89`

hh\_hk <- mutate(Sim,   
 hh\_size=rowSums(across(everything())),   
 hid=1:nrow(Sim)) %>%   
 pivot\_longer(-c(hid, hh\_size), names\_to="age") %>%   
 filter(value>0) %>%   
 uncount(weights=value) %>%   
 group\_by(hh\_size, hid) %>%   
 mutate(pid=row\_number()) %>%   
 pivot\_wider(id\_cols=c(hid, hh\_size), names\_from=pid, names\_prefix="age\_", values\_from=age)   
hh\_hk

## # A tibble: 10,000 × 8  
## # Groups: hh\_size, hid [10,000]  
## hid hh\_size age\_1 age\_2 age\_3 age\_4 age\_5 age\_6  
## <int> <dbl> <chr> <chr> <chr> <chr> <chr> <chr>  
## 1 1 2 70 - 79 70 - 79 <NA> <NA> <NA> <NA>   
## 2 2 2 20 - 29 20 - 29 <NA> <NA> <NA> <NA>   
## 3 3 4 20 - 29 30 - 39 50 - 59 50 - 59 <NA> <NA>   
## 4 4 3 0 - 9 30 - 39 40 - 49 <NA> <NA> <NA>   
## 5 5 1 40 - 49 <NA> <NA> <NA> <NA> <NA>   
## 6 6 1 30 - 39 <NA> <NA> <NA> <NA> <NA>   
## 7 7 4 0 - 9 10 - 19 50 - 59 50 - 59 <NA> <NA>   
## 8 8 4 0 - 9 20 - 29 30 - 39 60 - 69 <NA> <NA>   
## 9 9 5 10 - 19 10 - 19 10 - 19 50 - 59 50 - 59 <NA>   
## 10 10 1 60 - 69 <NA> <NA> <NA> <NA> <NA>   
## # … with 9,990 more rows

# Households of size 2   
filter(hh\_hk, hh\_size==2) %>%   
 group\_by(age\_1, age\_2) %>%   
 summarise(n=n(), .groups='drop') %>%   
 mutate(p=n/sum(n)) %>%   
 arrange(-n)

## # A tibble: 40 × 4  
## age\_1 age\_2 n p  
## <chr> <chr> <int> <dbl>  
## 1 60 - 69 60 - 69 345 0.116   
## 2 50 - 59 50 - 59 238 0.0797  
## 3 30 - 39 30 - 39 217 0.0727  
## 4 40 - 49 40 - 49 210 0.0703  
## 5 70 - 79 70 - 79 152 0.0509  
## 6 40 - 49 50 - 59 147 0.0492  
## 7 50 - 59 60 - 69 133 0.0445  
## 8 20 - 29 30 - 39 126 0.0422  
## 9 30 - 39 40 - 49 108 0.0362  
## 10 30 - 39 50 - 59 98 0.0328  
## # … with 30 more rows

# The most common age pattern in households of size 2 is 60-69{2} (11.6%)   
  
  
# Households of size 3  
filter(hh\_hk, hh\_size==3) %>%   
 group\_by(age\_1, age\_2, age\_3) %>%   
 summarise(n=n(), .groups='drop') %>%   
 arrange(-n) %>%   
 mutate(p=n/sum(n))

## # A tibble: 131 × 5  
## age\_1 age\_2 age\_3 n p  
## <chr> <chr> <chr> <int> <dbl>  
## 1 20 - 29 50 - 59 50 - 59 134 0.0587  
## 2 0 - 9 30 - 39 30 - 39 129 0.0565  
## 3 30 - 39 50 - 59 50 - 59 124 0.0543  
## 4 10 - 19 40 - 49 40 - 49 106 0.0465  
## 5 20 - 29 40 - 49 40 - 49 103 0.0451  
## 6 10 - 19 50 - 59 50 - 59 84 0.0368  
## 7 0 - 9 30 - 39 40 - 49 75 0.0329  
## 8 10 - 19 30 - 39 30 - 39 75 0.0329  
## 9 30 - 39 60 - 69 60 - 69 74 0.0324  
## 10 40 - 49 60 - 69 60 - 69 68 0.0298  
## # … with 121 more rows

# The most common pattern in households of size 3 is 50-59{2} + 20-29{1} (5.9%)   
# i.e., a married couple (presumably) aged 50-59 living with a single child aged 20-29 yrs   
  
  
# Households of size 4  
filter(hh\_hk, hh\_size==4) %>%   
 group\_by(age\_1, age\_2, age\_3, age\_4) %>%   
 summarise(n=n(), .groups='drop') %>%   
 arrange(-n) %>%   
 mutate(p=n/sum(n))

## # A tibble: 241 × 6  
## age\_1 age\_2 age\_3 age\_4 n p  
## <chr> <chr> <chr> <chr> <int> <dbl>  
## 1 20 - 29 30 - 39 50 - 59 50 - 59 180 0.0972  
## 2 10 - 19 20 - 29 40 - 49 40 - 49 106 0.0572  
## 3 10 - 19 20 - 29 50 - 59 50 - 59 105 0.0567  
## 4 0 - 9 10 - 19 40 - 49 40 - 49 71 0.0383  
## 5 0 - 9 0 - 9 30 - 39 30 - 39 67 0.0362  
## 6 30 - 39 30 - 39 50 - 59 50 - 59 59 0.0319  
## 7 0 - 9 10 - 19 30 - 39 30 - 39 58 0.0313  
## 8 20 - 29 20 - 29 50 - 59 50 - 59 57 0.0308  
## 9 30 - 39 40 - 49 60 - 69 60 - 69 57 0.0308  
## 10 10 - 19 20 - 29 40 - 49 50 - 59 55 0.0297  
## # … with 231 more rows

# The most common pattern in households of size 4 is 50-59{2} + 30-39{1} + 20-29{1} (9.7%)   
  
  
# Households of size 5  
filter(hh\_hk, hh\_size==5) %>%   
 group\_by(age\_1, age\_2, age\_3, age\_4, age\_5) %>%   
 summarise(n=n(), .groups='drop') %>%   
 arrange(-n) %>%   
 mutate(p=n/sum(n)) # %>% print(n=Inf)

## # A tibble: 240 × 7  
## age\_1 age\_2 age\_3 age\_4 age\_5 n p  
## <chr> <chr> <chr> <chr> <chr> <int> <dbl>  
## 1 20 - 29 20 - 29 30 - 39 50 - 59 50 - 59 42 0.0522  
## 2 10 - 19 20 - 29 20 - 29 40 - 49 40 - 49 37 0.0460  
## 3 0 - 9 0 - 9 10 - 19 30 - 39 30 - 39 35 0.0435  
## 4 10 - 19 10 - 19 20 - 29 40 - 49 40 - 49 34 0.0422  
## 5 20 - 29 30 - 39 30 - 39 50 - 59 50 - 59 34 0.0422  
## 6 0 - 9 10 - 19 10 - 19 40 - 49 40 - 49 24 0.0298  
## 7 0 - 9 0 - 9 10 - 19 30 - 39 40 - 49 23 0.0286  
## 8 0 - 9 0 - 9 0 - 9 20 - 29 30 - 39 20 0.0248  
## 9 0 - 9 0 - 9 10 - 19 40 - 49 40 - 49 19 0.0236  
## 10 10 - 19 10 - 19 20 - 29 50 - 59 50 - 59 19 0.0236  
## # … with 230 more rows

# Households of size 6  
filter(hh\_hk, hh\_size==6) %>%   
 group\_by(age\_1, age\_2, age\_3, age\_4, age\_5, age\_6) %>%   
 summarise(n=n(), .groups='drop') %>%   
 arrange(-n) %>%   
 mutate(p=n/sum(n)) #%>% print(n=Inf)

## # A tibble: 204 × 8  
## age\_1 age\_2 age\_3 age\_4 age\_5 age\_6 n p  
## <chr> <chr> <chr> <chr> <chr> <chr> <int> <dbl>  
## 1 20 - 29 20 - 29 20 - 29 30 - 39 50 - 59 50 - 59 8 0.0276  
## 2 10 - 19 10 - 19 20 - 29 20 - 29 40 - 49 40 - 49 7 0.0241  
## 3 20 - 29 20 - 29 30 - 39 30 - 39 50 - 59 50 - 59 7 0.0241  
## 4 10 - 19 20 - 29 20 - 29 20 - 29 40 - 49 40 - 49 5 0.0172  
## 5 0 - 9 0 - 9 0 - 9 10 - 19 30 - 39 40 - 49 4 0.0138  
## 6 0 - 9 0 - 9 10 - 19 10 - 19 40 - 49 40 - 49 4 0.0138  
## 7 0 - 9 10 - 19 10 - 19 10 - 19 40 - 49 40 - 49 4 0.0138  
## 8 0 - 9 10 - 19 30 - 39 30 - 39 70 - 79 70 - 79 4 0.0138  
## 9 10 - 19 20 - 29 20 - 29 20 - 29 50 - 59 50 - 59 4 0.0138  
## 10 10 - 19 20 - 29 40 - 49 40 - 49 80 - 89 80 - 89 4 0.0138  
## # … with 194 more rows

# Compare with the UK data  
  
# Read in the UK data   
UK <- read\_excel("Simulate 10000 households with HK demographic data\_ABM\_2.xlsx",   
 sheet="baseline\_household\_demographic",   
 range="B1:K10001")   
  
UK <- mutate(UK, hid=1:10000); UK

## # A tibble: 10,000 × 11  
## a\_0\_9 a\_10\_19 a\_20\_29 a\_30\_39 a\_40\_49 a\_50\_59 a\_60\_69 a\_70\_79 a\_80 hh\_size  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 0 0 0 0 2 0 0 0 0 2  
## 2 0 1 0 0 0 0 0 0 0 1  
## 3 0 0 0 1 0 0 2 0 0 3  
## 4 0 2 0 0 1 0 0 0 0 3  
## 5 0 0 1 0 1 0 0 0 0 2  
## 6 0 0 2 0 0 0 0 0 0 2  
## 7 0 0 0 0 1 0 0 0 0 1  
## 8 0 0 0 0 0 0 0 1 0 1  
## 9 0 0 0 0 1 0 1 1 0 3  
## 10 1 2 0 2 0 0 0 0 0 5  
## # … with 9,990 more rows, and 1 more variable: hid <int>

filter(UK, hh\_size==1) %>%   
 pivot\_longer(starts\_with("a"), names\_to="age") %>%   
 group\_by(age) %>%   
 summarise(n=sum(value), .groups='drop') %>%   
 mutate(p=n/sum(n))

## # A tibble: 9 × 3  
## age n p  
## <chr> <dbl> <dbl>  
## 1 a\_0\_9 0 0   
## 2 a\_10\_19 19 0.00624  
## 3 a\_20\_29 234 0.0768   
## 4 a\_30\_39 368 0.121   
## 5 a\_40\_49 446 0.146   
## 6 a\_50\_59 456 0.150   
## 7 a\_60\_69 517 0.170   
## 8 a\_70\_79 494 0.162   
## 9 a\_80 511 0.168

# The most common age group for hh of size 1 in the UK is 60-69 (17.0%)