

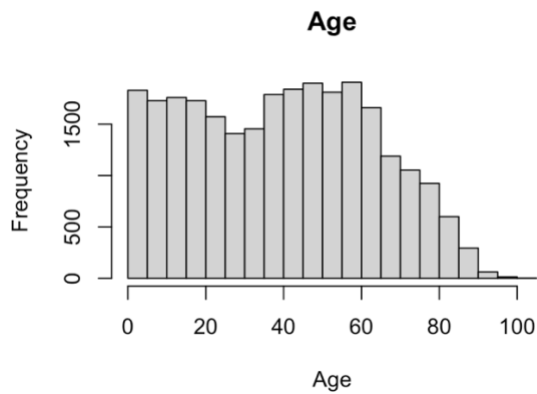
Econ 613 Homework 1
Hannah Marsho

Exercise 1: Basic Statistics

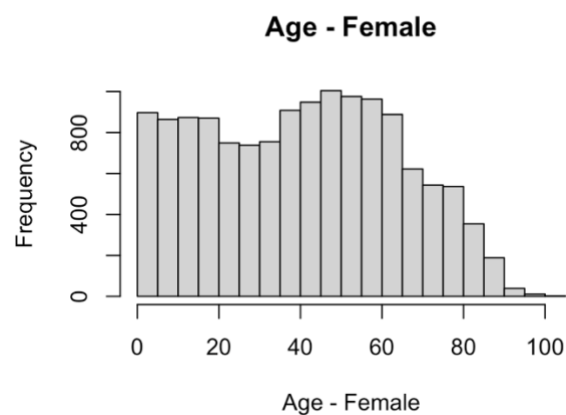
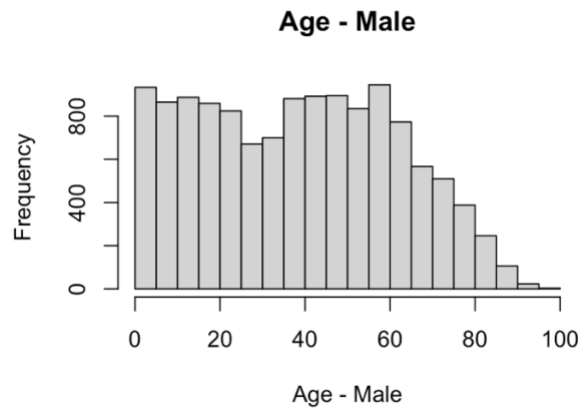
- a) 10,498
- b) 3,374
- c) 25,510
- d) 2,765
- e)

	56	62	63	64	65	67	68	69
Female	696	64	35	29	19	147	120	40
Male	74	443	520	246	159	237	177	82

- f) 2005:
 Mean – 21,219.94
 St. Dev. – 18,300.66
 D9/D1 Ratio – 23.2218
 Gini Coef. – 0.8
 2019:
 Mean – 26,550.06
 St. Dev. – 25,182.74
 D9/D1 Ratio – 26.31411
 Gini Coef. – 0.8



- g)



There isn't a strong difference in the 2010 age distribution when we consider gender. However, the 2010 age distribution for males has slightly stronger humps for ages 0-20 and 40-70 than the female distribution. Also, the 2010 age distribution for females has a more gradual drop-off for old age, so women tend to live longer.

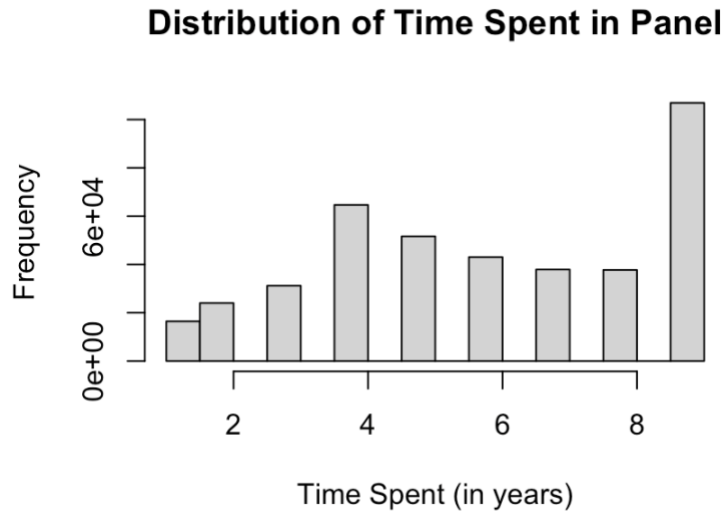
h) 3,514

Exercise 2: Merge Datasets

- a) Appended dataset available through code. Too large to input here.
- b) Appended dataset available through code. Too large to input here.
- c) "...1", "idmen", and "year"
- d) Merged dataset available through code. Too large to input here.
- e) 3,622
- f) 8,162
- g) 3,022
- h) 15,646
- i) 6,186
- j) There are two households that are tied for having the most family members. They both have 14 members and are located at index positions 1524 and 2114. The idmens are 2.21e15 and 2.51e15, respectively.
- k) 8,984

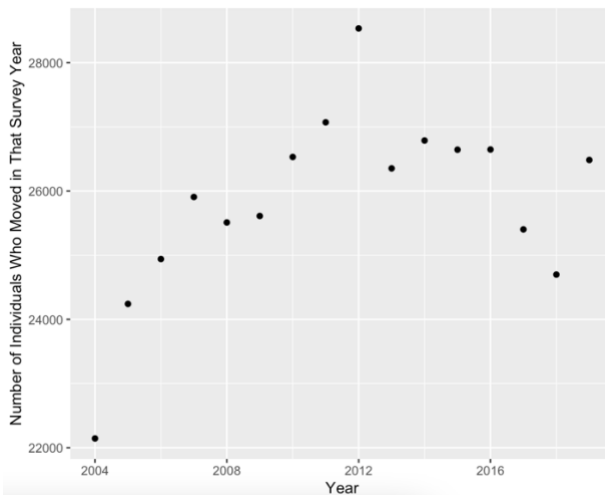
Exercise 3: Migration

- a) Dataframes which provide the year each household enters and exits the panel are available through code. Too large to input here. Distribution shown through histogram.



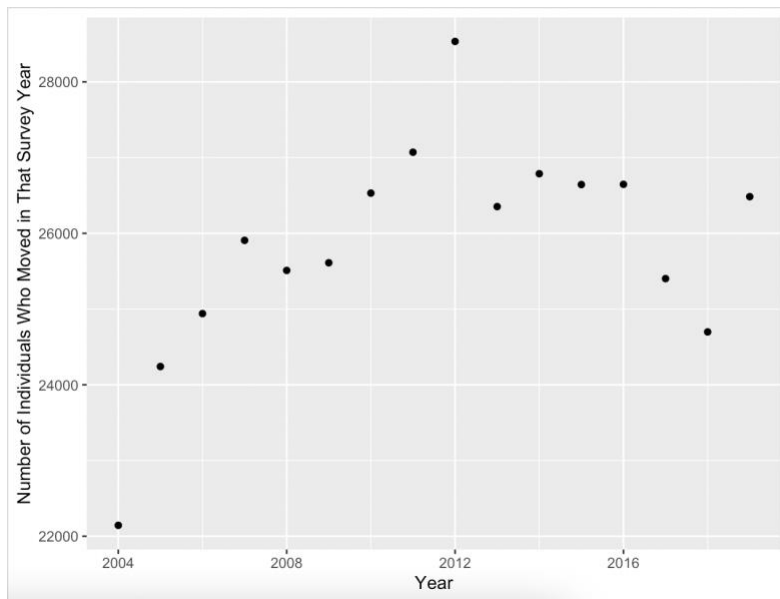
- b) Dataframe which provides whether or not a household moved into its current dwelling at the year of survey are available through code. Too large to input here.

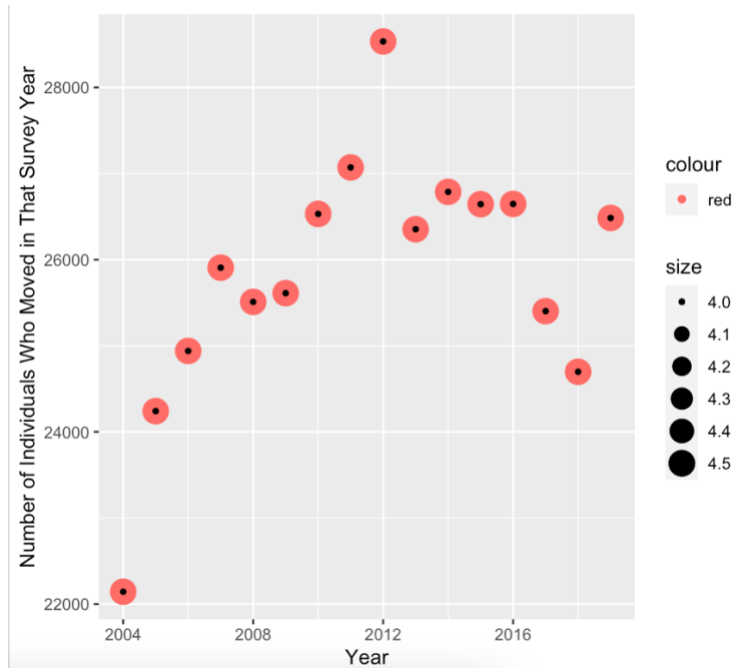
	idmen	idind	year	sameyear
1	1.20001e+15	1.120001e+18	2004	0
2	1.20001e+15	1.120001e+18	2004	0
3	1.20001e+15	1.120001e+18	2004	0
4	1.20001e+15	1.120001e+18	2005	0
5	1.20001e+15	1.120001e+18	2005	0
6	1.20001e+15	1.120001e+18	2004	0
7	1.20001e+15	1.120001e+18	2004	0
8	1.20001e+15	1.120001e+18	2005	1
9	1.20001e+15	1.120001e+18	2005	1
10	1.20001e+15	1.120001e+18	2004	0



- c) Dataframe which provides whether or not a household moved into its current dwelling at the year of survey are available through code. Too large to input here.

	idmen	idind	year	sameyear2
1	1.20001e+15	1.120001e+18	2004	0
2	1.20001e+15	1.120001e+18	2004	0
3	1.20001e+15	1.120001e+18	2004	0
4	1.20001e+15	1.120001e+18	2005	0
5	1.20001e+15	1.120001e+18	2005	0
6	1.20001e+15	1.120001e+18	2004	0
7	1.20001e+15	1.120001e+18	2004	0
8	1.20001e+15	1.120001e+18	2005	1
9	1.20001e+15	1.120001e+18	2005	1
10	1.20001e+15	1.120001e+18	2004	0





d)

I prefer the method used in part b. This is because you can simply use one variable, datent, in order to find our solution. With the method in # part c, you are required to use two variables, myear and move, in order to obtain the same result. With this second method, you must create the solution # conditionally on which of the two, myear and move, is available for a given data point. This makes the process a bit more cumbersome and increases the likelihood # of error since there are two variables for participants and researchers to interpret and use. So, it is simpler to use the method in part c.

e) No final answer available. Code returns NA.

Exercise 4: Attrition

No final answer available. Code returns NA.