Econ 613 - Applied Econometrics - 2022 Spring
Homework 1

Yican Liu

January 21, 2022

1

1.1

Number of households surveyed in 2007.

- 10498

1.2

Number of households with marital status "Couple with kids" in 2005.

- 3374

1.3

Number of individuals surveyed in 2008.

- 25510

1.4

Number of individuals aged between 25 and 35 in 2016.

- 2765

1.5

Cross-table gender/profession in 2009.

- The cross table is:

```
          0   11   12   13   21   22   23   31   33   34   35   37   38   42   43   44   45   46
Female   11   30    8   29   63   65    8   68   85  184   50  179   78  258  437    1  153  410
Male     19   57   19   78  213  114   48   98  107  142   59  260  368  110  117    2   95  340

         47   48   52   53   54   55   56   62   63   64   65   67   68   69
Female   82   22  782   27  584  353  696   64   35   29   19  147  120   40
Male    429  215  169  182   98  101   74  443  520  246  159  237  177   82
```

1.6

Distribution of wages in 2005 and 2019. Report the mean, the standard deviation, the inter-decile ratio D9/D1 and the Gini coefficient.

- For 2005, the mean is 11992, the standard deviation is 17318.56, the inter-decile ratio is 32340.4/ 0.0 , and the Gini coefficient is 0.67.

- For 2019, the mean is 15350, the standard deviation is 23207.18, the inter-decile ratio is 40267/0.0, and the Gini coefficient is 0.67.

1.7

Distribution of age in 2010. Plot an histogram. Is there any difference between men and women?

Distribution of Age

Distribution of Men Age

Frequency

Men Age, 2010

Distribution of Women Age

Frequency

Women Age, 2010
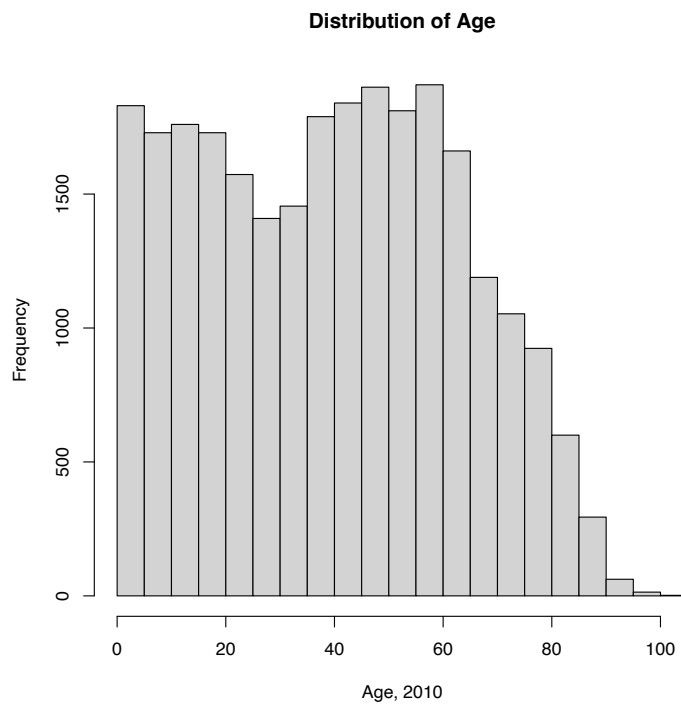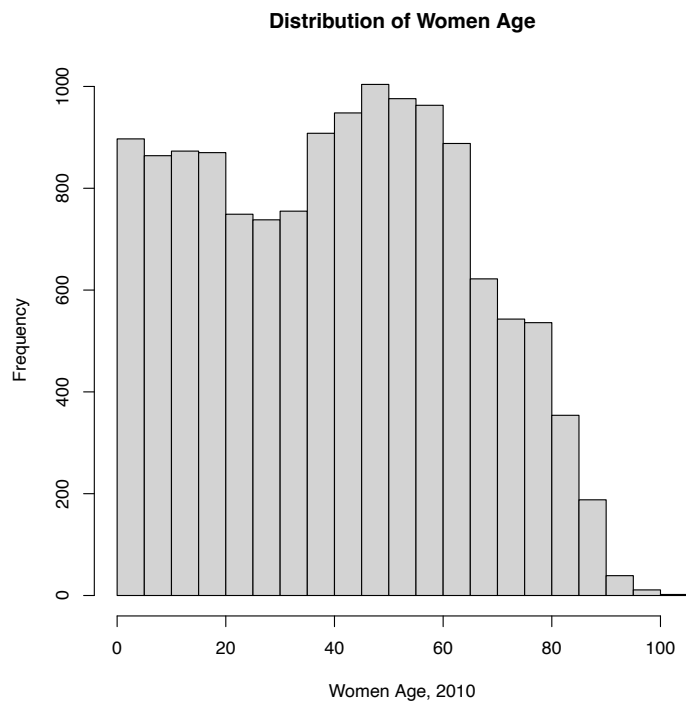
- I plot the histogram of the distribution of age in the above three figures. The results show that there is no significant difference between the distribution of men and women's age.

1.8

Number of individuals in Paris in 2011.

- 3531

2

2.1

Read all individual datasets from 2004 to 2019. Append all these datasets.

- Please see my code.

2.2

Read all household datasets from 2004 to 2019. Append all these datasets.

- Please see my code.

2.3

List the variables that are simultaneously present in the individual and household datasets.

- Including *idmen* and *year*. I dropped the "X" when merging because it will not be used in the later process.

2.4

Merge the appended individual and household datasets.

- Please see my code.

2.5

Number of households in which there are more than four family members

- 12005 household-year observations. That is, in this question, some households might be counted for several times.

2.6

Number of households in which at least one member is unemployed

- 17241 household-year observations.

2.7

Number of households in which at least two members are of the same profession

- 7586

2.8

Number of individuals in the panel that are from household-Couple with kids

- 208468

2.9

Number of individuals in the panel that are from Paris.

- 52601

2.10

Find the household with the most number of family members. Report its idmen.

- 2207811124040100
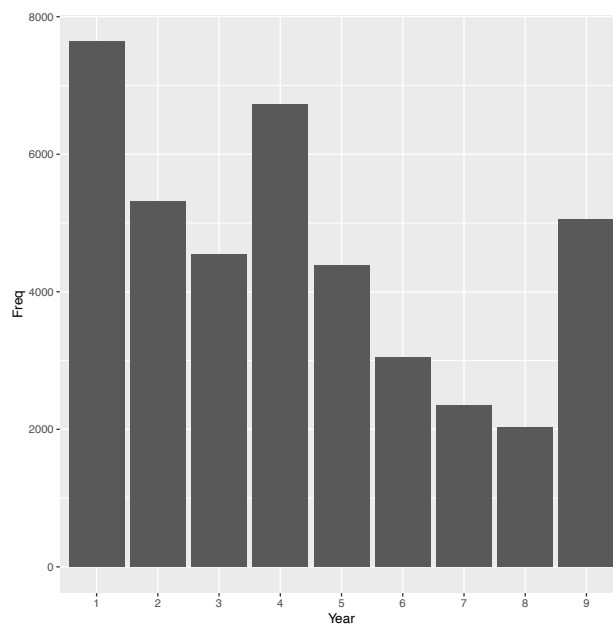
2.11

Number of households present in 2010 and 2011.

- 22408

3

3.1

Find out the year each household enters and exit the panel. Report the distribution of the time spent in the survey for each household.
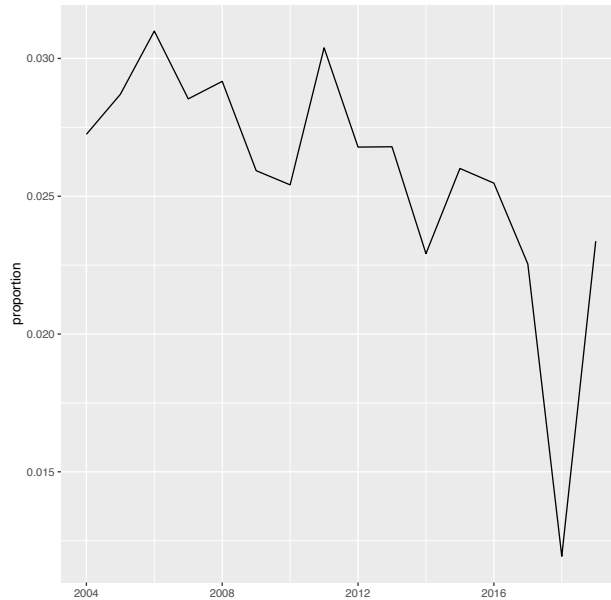
- The figure is as follows

3.2

Based on *datent*, identify whether or not a household moved into its current dwelling at the year of survey. Report the first 10 rows of your result and plot the share of individuals in that situation across years.

	idmen	year	idind	empstat	respondent	profession	gender	age
8	1.200010e+15	2005	1.120001e+18	Employed		0	45	Female	28
9	1.200010e+15	2005	1.120001e+18	Employed		1	38	Male	32
87	1.200177e+15	2005	2.120018e+18	Employed		1	43	Female	22
88	1.200177e+15	2005	1.120018e+18	Employed		0	63	Male	27
152	1.200339e+15	2005	1.120034e+18	Employed		0	54	Female	36
153	1.200339e+15	2005	1.120034e+18	Inactive		0		Male	6
154	1.200339e+15	2005	1.120034e+18	Employed		0	21	Male	38
155	1.200339e+15	2005	1.120034e+18	Inactive		0		Male	11
157	1.200339e+15	2005	1.120034e+18	Inactive		0		Male	0
158	1.200339e+15	2005	1.120034e+18	Employed		1	54	Female	21

	wage	datent	myear	mstatus	move	location	startyear
8	19231	2005	2005	Couple, No kids	NA	Paris	2004
9	50659	2005	2005	Couple, No kids	NA	Paris	2004
87	8798	2005	2005	Couple, No kids	NA	Urban 100000 to 199999	2005
88	19688	2005	2005	Couple, No kids	NA	Urban 100000 to 199999	2005
152	13237	2005	2005	Couple, with Kids	NA	Urban 10000 to 19999	2005
153	NA	2005	2005	Couple, with Kids	NA	Urban 10000 to 19999	2005
154	0	2005	2005	Couple, with Kids	NA	Urban 10000 to 19999	2005
155	NA	2005	2005	Couple, with Kids	NA	Urban 10000 to 19999	2005
157	NA	2005	2005	Couple, with Kids	NA	Urban 50000 to 99999	2004
158	15521	2005	2005	Couple, with Kids	NA	Urban 50000 to 99999	2004

|	endyear	surveyyear	startwithsurvey
8	2005	2	TRUE
9	2005	2	TRUE
87	2005	1	TRUE
88	2005	1	TRUE
152	2005	1	TRUE
153	2005	1	TRUE
154	2005	1	TRUE
155	2005	1	TRUE
157	2005	2	TRUE
158	2005	2	TRUE

3.3

Based on *myear* and *move*, identify whether or not household migrated at the year of survey. Report the first 10 rows of your result and plot the share of individuals in that situation across years.

```
       idmen year         idind  empstat respondent profession gender age   wage
1  1.20001e+15 2004 1.120001e+18 Employed          1         67   Male  31  19187
2  1.20001e+15 2004 1.120001e+18 Employed          1         56 Female  30  11586
3  1.20001e+15 2004 1.120001e+18 Inactive          0            Female   9     NA
4  1.20001e+15 2005 1.120001e+18 Inactive          1            Female  31  12334
5  1.20001e+15 2005 1.120001e+18 Inactive          0            Female  10     NA
6  1.20001e+15 2004 1.120001e+18 Employed          1         38   Male  31  44656
7  1.20001e+15 2004 1.120001e+18 Employed          0         45 Female  27  20413
8  1.20001e+15 2005 1.120001e+18 Employed          0         45 Female  28  19231
9  1.20001e+15 2005 1.120001e+18 Employed          1         38   Male  32  50659
10 1.20001e+15 2004 1.120001e+18  Retired          1            Female  89      0
   datent myear        mstatus move location migrated
1    2000  2000          Single   NA    Paris    FALSE
2    2001  2001   Single Parent   NA    Paris    FALSE
3    2001  2001   Single Parent   NA    Paris    FALSE
4    2001  2001   Single Parent   NA    Paris    FALSE
5    2001  2001   Single Parent   NA    Paris    FALSE
6    2000  2000 Couple, No kids   NA    Paris    FALSE
7    2000  2000 Couple, No kids   NA    Paris    FALSE
8    2005  2005 Couple, No kids   NA    Paris     TRUE
9    2005  2005 Couple, No kids   NA    Paris     TRUE
10   1957  1957          Single   NA    Paris    FALSE
```
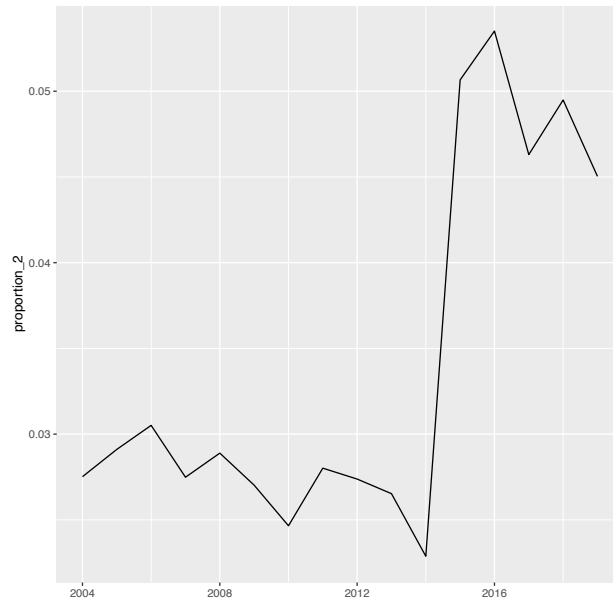
3.4

Mix the two plots you created above in one graph, clearly label the graph. Do you prefer one method over the other? Justify.
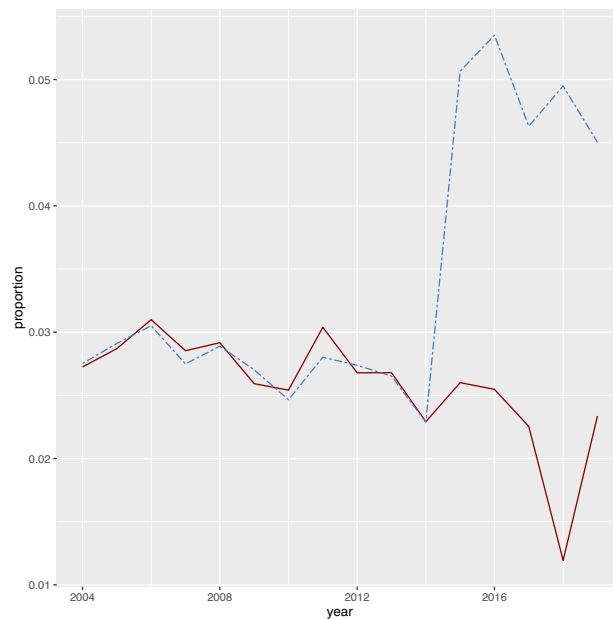
I prefer the first estimation. This is because the value is estimated from the same variable before and after 2014. That is, there is no "sudden change" before and after 2014.

3.5

- I use the year + 1 to get the information of "lagged *empstat*." The results is 6547.

4

Compute the attrition across each year, where attrition is defined as the reduction in the number of individuals staying in the data panel. Report your final result as a table in proportions.

According to the hint, I estimate the attrition of each year in the following table:

	endyear	n
1	2004	1137
2	2005	1757
3	2006	1592
4	2007	2131
5	2008	1878
6	2009	1698
7	2010	1850
8	2011	1723
9	2012	2416
10	2013	2006
11	2014	2076
12	2015	2084
13	2016	2277
14	2017	2227
15	2018	2495
16	2019	11737