Econ 613 A1
Peilin Wang

Exercise 1:

(1) Number of households surveyed in 2007: 10498

(2) Number of households with marital status “Couple with kids” in 2005: 3374

(3) Number of individuals surveyed in 2008: 25510

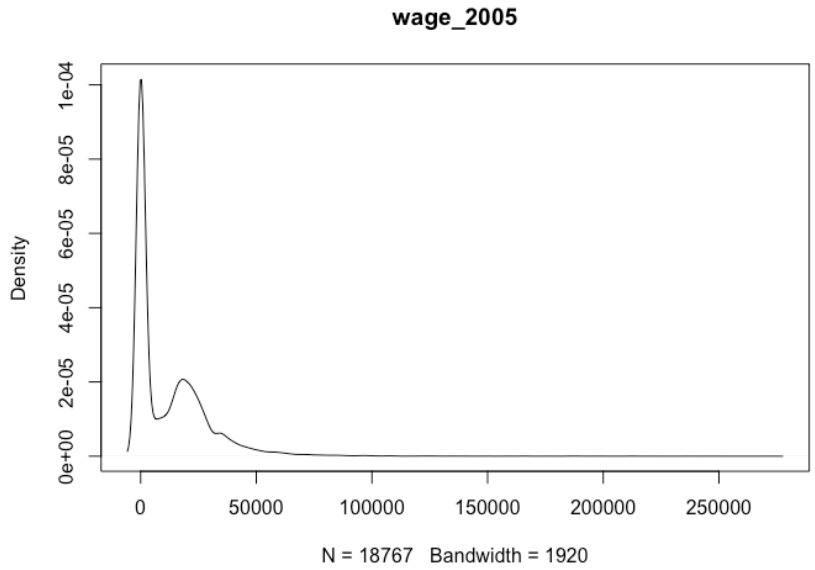
(4) Number of individuals aged between 25 and 35 in 2016: 2765

(5) Cross-table gender/profession in 2009:

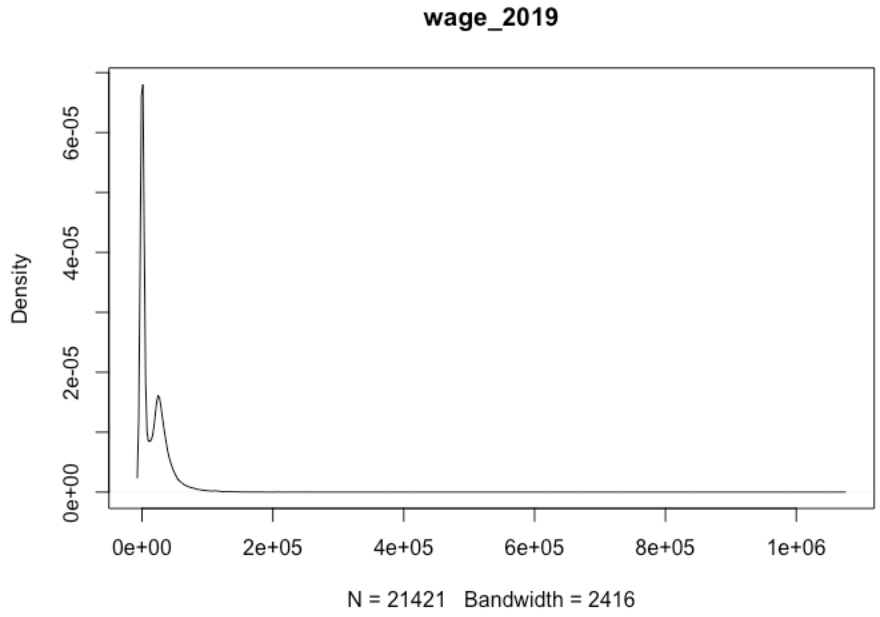
```
          0  11  12  13  21  22  23  31  33  34  35  37  38  42  43  44  45  46  47  48  52
Female   11  30   8  29  63  65   8  68  85 184  50 179  78 258 437   1 153 410  82  22 782
Male     19  57  19  78 213 114  48  98 107 142  59 260 368 110 117   2  95 340 429 215 169

          53  54  55  56  62  63  64  65  67  68  69
Female    27 584 353 696  64  35  29  19 147 120  40
Male     182  98 101  74 443 520 246 159 237 177  82
>
```

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

Distribution of wages in 2005/ the distribution is lognormal

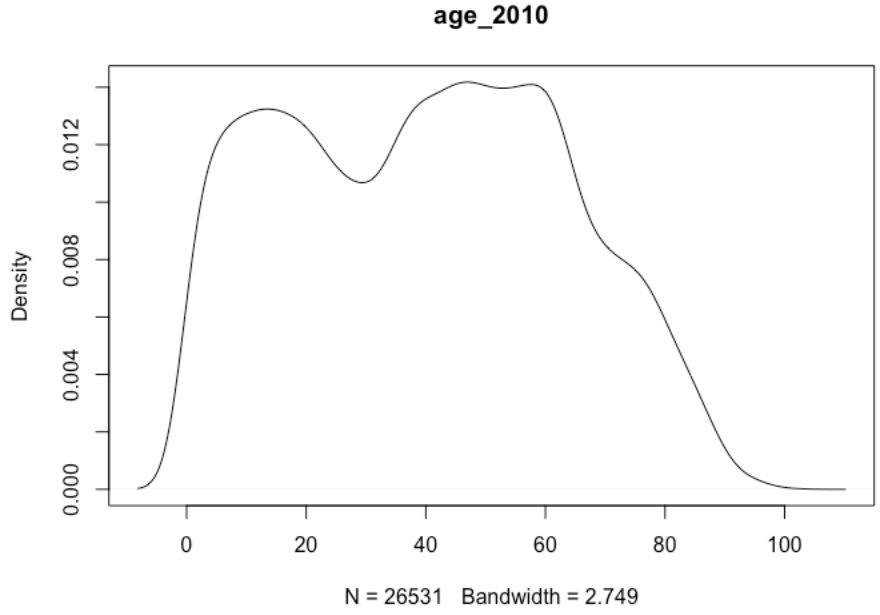


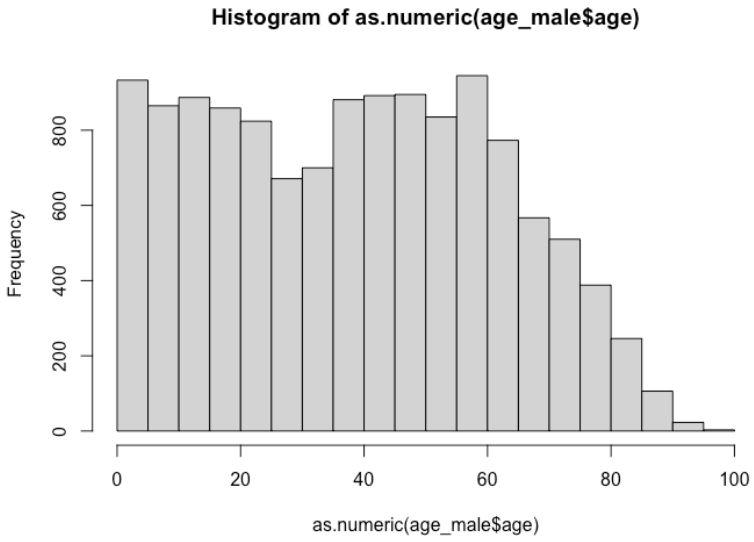
Distribution of wages in 2019/the distribution is also lognormal.

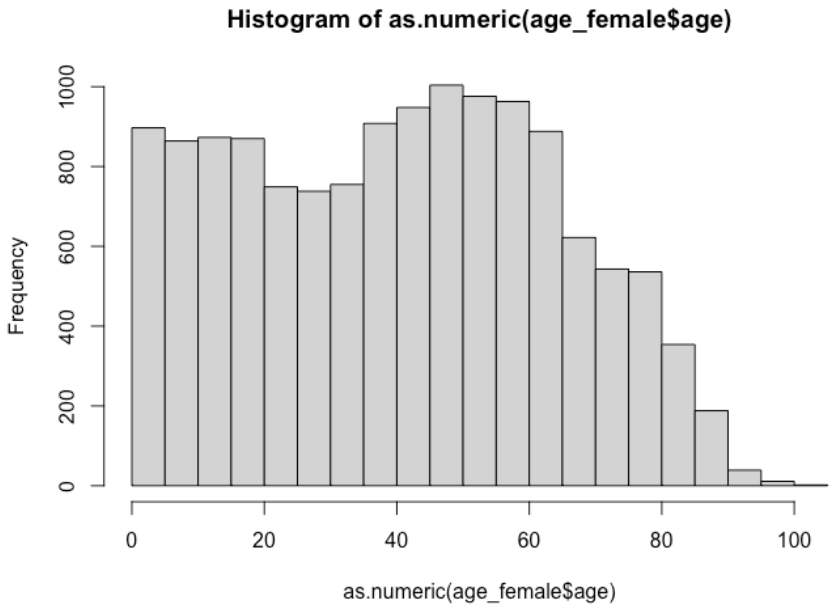


	mean	Sd	D9/D1	Gini coefficient
2005	22443.0291184683	18076.7088817948	8.89652518143831	0.377113473381286
2019	27578.839302189	25107.1871955391	13.862300495322	0.399087488264856

(7) Distribution of age in 2010. Plot an histogram. Is there any difference between men and women?
Distribution of age in 2010/it's likely to be uniform distribution.







By comparing two histograms, we can see that these are quite similar. However, there exists some elder females which makes the second histogram skew to the left a little bit.

(8) Number of individuals in Paris in 2011: 3514

Exercise 2:

Please check (1)-(2), (4) in R
(3) List the variables that are simultaneously present in the individual and household datasets:
"idmen" "V1" "year"

(5) Number of households in which there are more than four family members:12436

(6) Number of households in which at least one member is unemployed: 17241

(7) Number of households in which at least two members are of the same profession:7615

(8) Number of individuals in the panel that are from household-Couple with kids:209382

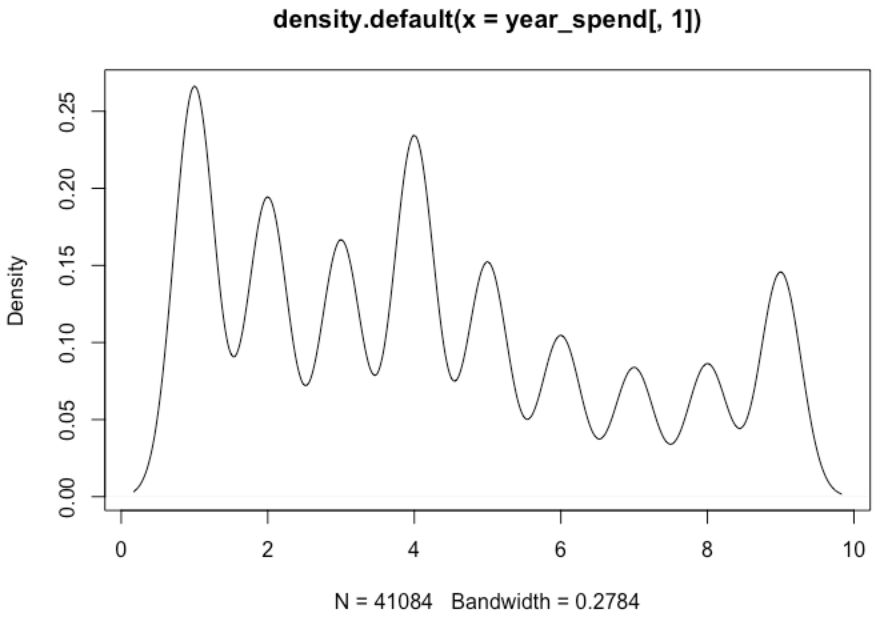
(9) Number of individuals in the panel that are from Paris: 51904

(10) Find the household with the greatest number of family members. Report its idmen: Two households have the 14 family members which are the most. These idmens are 2207811124040100 from year 2007 and 2510263102990100 from year 2010.

(11) Number of households present in 2010 and 2011: 8984

Exercise 3:

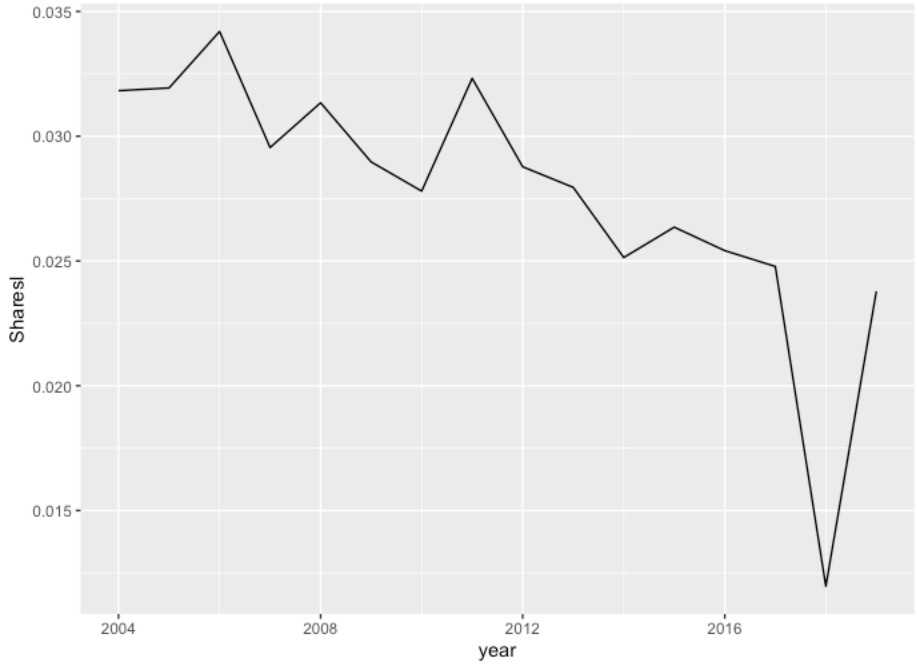
(1) Find out the year each household enters and exit the panel. Report the distribution of the time spent in the survey for each household



(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:

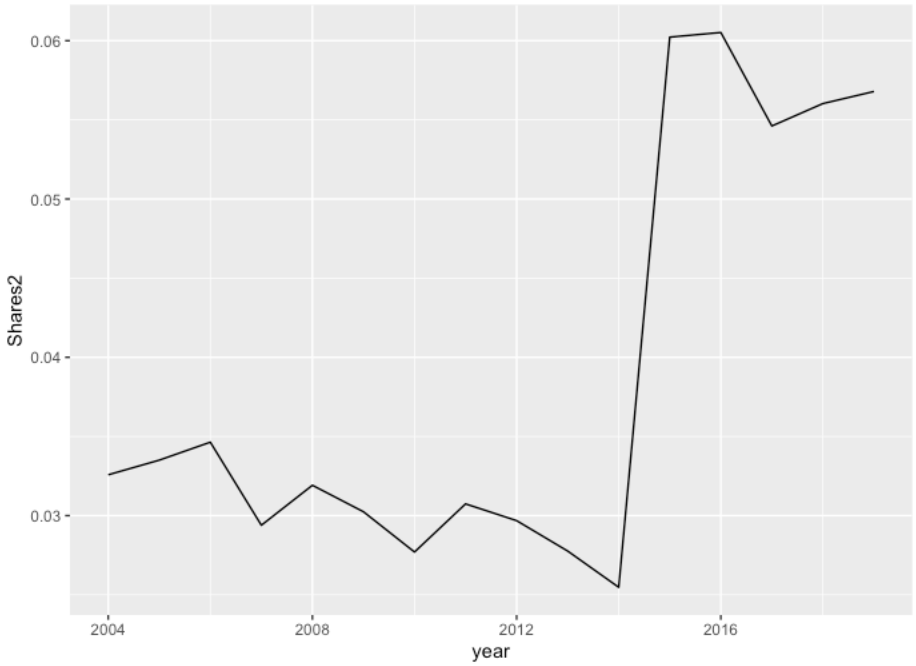
.

	V1	idmen	year	datent	myear	mstatus	move	location	dwelling_moved
1:	1	1200010012930100	2004	2000	2000	Single	NA	Paris	FALSE
2:	2	1200010040580100	2004	2001	2001	Single Parent	NA	Paris	FALSE
3:	3	1200010066630100	2004	2000	2000	Couple, No kids	NA	Paris	FALSE
4:	4	1200010082450100	2004	1957	1957	Single	NA	Paris	FALSE
5:	5	1200010086440100	2004	2001	2001	Couple, No kids	NA	Paris	FALSE
6:	6	1200010102990100	2004	1990	1990	Single Parent	NA	Paris	FALSE
7:	7	1200010118450100	2004	2000	2000	Couple, No kids	NA	Paris	FALSE
8:	8	1200020012930100	2004	1948	1988	Other	NA	Rural	FALSE
9:	9	1200020017390100	2004	1979	1979	Single	NA	Rural	FALSE
10:	10	1200020026420100	2004	1984	1981	Other	NA	Rural	FALSE

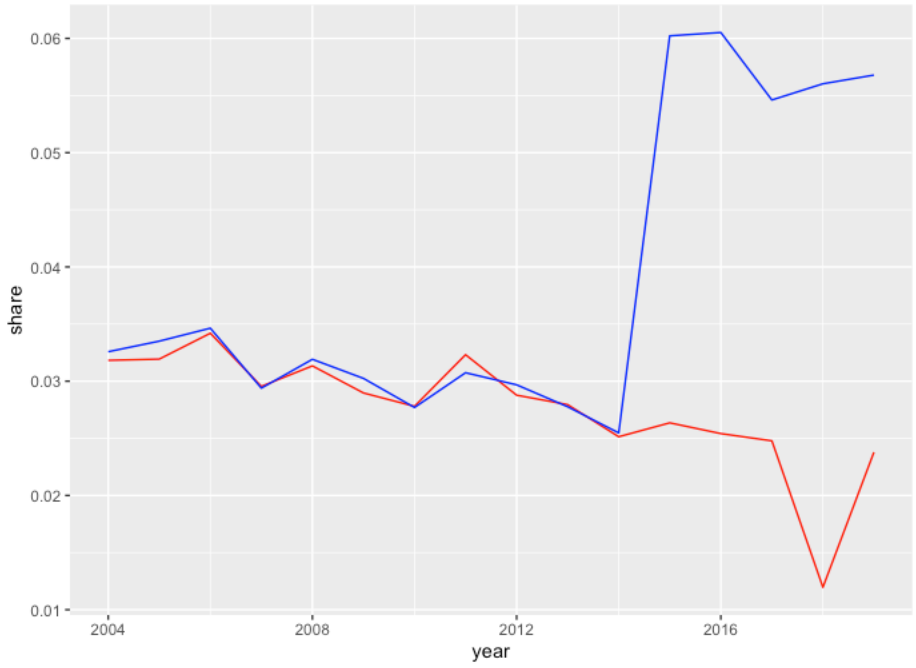


(3) Based on myear and move, identify whether or not household migrated at the year of survey. Report the first 10 rows of your result and plot the share of individuals in that situation across years:

	V1	idmen	year	datent	myear	mstatus	move	location	migrated
1:	1	1200010012930100	2004	2000	2000	Single	NA	Paris	FALSE
2:	2	1200010040580100	2004	2001	2001	Single Parent	NA	Paris	FALSE
3:	3	1200010066630100	2004	2000	2000	Couple, No kids	NA	Paris	FALSE
4:	4	1200010082450100	2004	1957	1957	Single	NA	Paris	FALSE
5:	5	1200010086440100	2004	2001	2001	Couple, No kids	NA	Paris	FALSE
6:	6	1200010102990100	2004	1990	1990	Single Parent	NA	Paris	FALSE
7:	7	1200010118450100	2004	2000	2000	Couple, No kids	NA	Paris	FALSE
8:	8	1200020012930100	2004	1948	1988	Other	NA	Rural	FALSE
9:	9	1200020017390100	2004	1979	1979	Single	NA	Rural	FALSE
10:	10	1200020026420100	2004	1984	1981	Other	NA	Rural	FALSE



(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 think the method based on datent is better. There is a huge jump in the plot based on myear and move. I guess it may not accurate. By comparing, the plot based on datent is relatively smooth.

(5) For households who migrate, find out how many households had at least one family member change his/her profession or employment status: 1655

Exercise 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.

	indi_2004_2018	indi_2005_2019	vec4	exit	attrition
1	22144	24241	19148	2996	13.52962
2	24241	24940	19391	4850	20.00743
3	24940	25907	20483	4457	17.87089
4	25907	25510	20034	5873	22.66955
5	25510	25611	20265	5245	20.56056
6	25611	26528	20904	4707	18.37882
7	26528	27071	21392	5136	19.36068
8	27071	28534	22472	4599	16.98866
9	28534	26321	21268	7266	25.46436
10	26321	26787	20530	5791	22.00144
11	26787	26644	20914	5873	21.92481
12	26644	26647	20855	5789	21.72722
13	26647	25402	19965	6682	25.07599
14	25402	24698	19199	6203	24.41934
15	24698	26484	18688	6010	24.33395