# Econ 613 Homework 1

Lu Liu

## Exercise 1 Basic Statistics

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

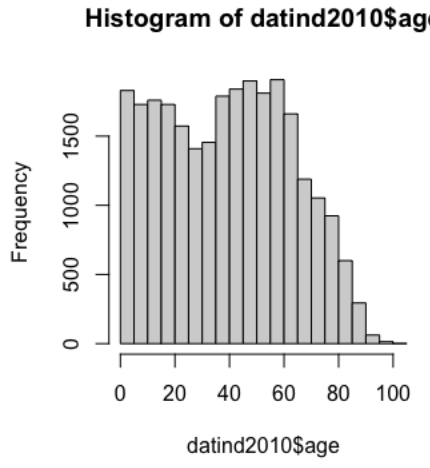
5. Cross-table gender/profession in 2009:

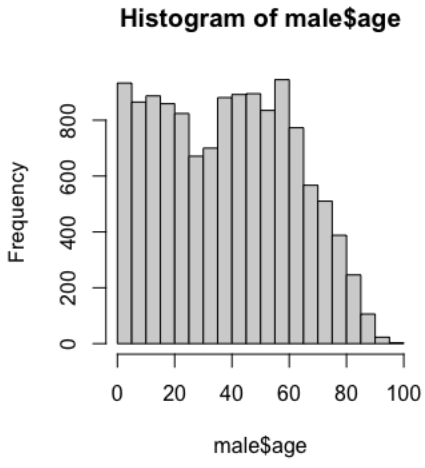
```
          0  11  12  13  21  22  23  31  33  34  35  37  38  42  43  44  45  46  47  48  52  53  54  55  56
Female  11  30   8  29  63  65   8  68  85 184  50 179  78 258 437   1 153 410  82  22 782  27 584 353 696
Male    19  57  19  78 213 114  48  98 107 142  59 260 368 110 117   2  95 340 429 215 169 182  98 101  74

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

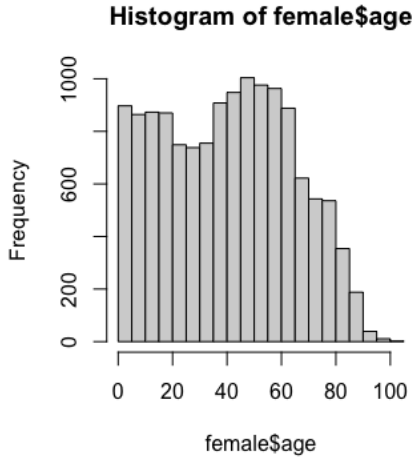
6. Distribution of wages in 2005 and 2019. Report the mean, the standard deviation, the inter-decile ratio D9/D1 and the Gini coefficient:

- 2005: Mean: 22443.03; Standard deviation:18076.71; Inter-decile ratio D9/D1: 8.896525
  Gini coefficient: 0.3771135
- 2019: Mean: 27578.84; Standard deviation:25107.19; Inter-decile ratio D9/D1: 13.8623
  Gini coefficient: 0.3990875

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






Yes, there is a difference between male and female. From age 35 to 65, the number of Females is more than the number of males.

8. Number of individuals in Paris in 2011: 3531

## Exercise 2 Merge Datasets

1. Read all individual datasets from 2004 to 2019. Append all these datasets. Check the R code.

2. Read all household datasets from 2004 to 2019. Append all these datasets. Check the R code.

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

“x” , “idmen”, “year”

4. Merge the appended individual and household datasets.

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: 8702

7. Number of households in which at least two members are of the same profession:7651

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

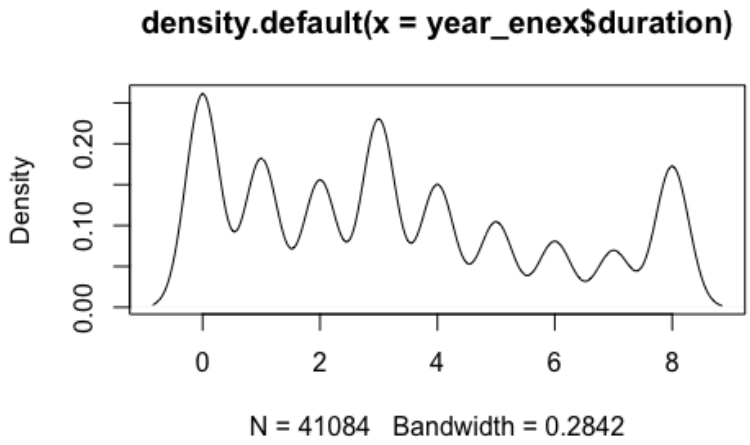
10. Find the household with the most number of family members. Report its idmen.

2.207811e+15 in year 2007, 2.510263e+15 in year 2010

11. Number of households present in 2010 and 2011:8984

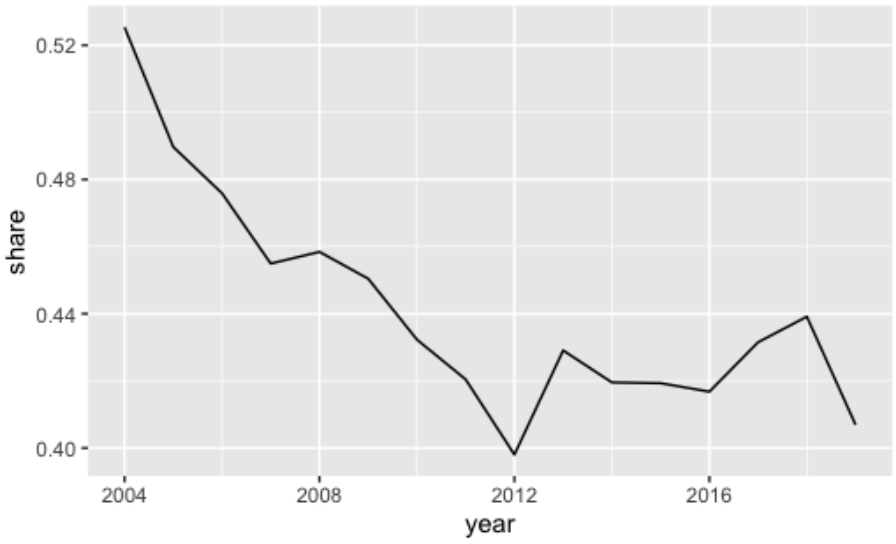
## Exercise 3 Migration

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.

| | X | idmen | year | datent | myear | mstatus | move | location | move_atyear |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1.20001e+15 | 2004 | 2000 | 2000 | Single | NA | Paris | 0 |
| 2 | 2 | 1.20001e+15 | 2004 | 2001 | 2001 | Single Parent | NA | Paris | 0 |
| 3 | 3 | 1.20001e+15 | 2004 | 2000 | 2000 | Couple, No kids | NA | Paris | 0 |
| 4 | 4 | 1.20001e+15 | 2004 | 1957 | 1957 | Single | NA | Paris | 0 |
| 5 | 5 | 1.20001e+15 | 2004 | 2001 | 2001 | Couple, No kids | NA | Paris | 0 |
| 6 | 6 | 1.20001e+15 | 2004 | 1990 | 1990 | Single Parent | NA | Paris | 0 |
| 7 | 7 | 1.20001e+15 | 2004 | 2000 | 2000 | Couple, No kids | NA | Paris | 0 |
| 8 | 8 | 1.20002e+15 | 2004 | 1948 | 1988 | Other | NA | Rural | 0 |
| 9 | 9 | 1.20002e+15 | 2004 | 1979 | 1979 | Single | NA | Rural | 0 |
| 10 | 10 | 1.20002e+15 | 2004 | 1984 | 1981 | Other | NA | Rural | 0 |




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.

| | X | idmen | year | datent | myear | mstatus | move | location | house_migrated |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1.20001e+15 | 2004 | 2000 | 2000 | Single | NA | Paris | 0 |
| 2 | 2 | 1.20001e+15 | 2004 | 2001 | 2001 | Single Parent | NA | Paris | 0 |
| 3 | 3 | 1.20001e+15 | 2004 | 2000 | 2000 | Couple, No kids | NA | Paris | 0 |
| 4 | 4 | 1.20001e+15 | 2004 | 1957 | 1957 | Single | NA | Paris | 0 |
| 5 | 5 | 1.20001e+15 | 2004 | 2001 | 2001 | Couple, No kids | NA | Paris | 0 |
| 6 | 6 | 1.20001e+15 | 2004 | 1990 | 1990 | Single Parent | NA | Paris | 0 |
| 7 | 7 | 1.20001e+15 | 2004 | 2000 | 2000 | Couple, No kids | NA | Paris | 0 |
| 8 | 8 | 1.20002e+15 | 2004 | 1948 | 1988 | Other | NA | Rural | 0 |
| 9 | 9 | 1.20002e+15 | 2004 | 1979 | 1979 | Single | NA | Rural | 0 |
| 10 | 10 | 1.20002e+15 | 2004 | 1984 | 1981 | Other | NA | Rural | 0 |

4. Mix the two plots you created above in one graph, clearly label the graph. Do you prefer one method over the other? Justify.

5. For households who migrate, find out how many households had at least one family member changed his/her profession or employment status.

Exercise 4 Attrition

1. 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.