ECON 613

Yuqi Zhou

Exercise 1 Basic Statistics

• Number of households surveyed in 2007. 10498

• Number of households with marital status “Couple with kids” in 2005. 3374

• Number of individuals surveyed in 2008. 25510

• Number of individuals aged between 25 and 35 in 2016. 2765

• Cross-table gender/profession in 2009.

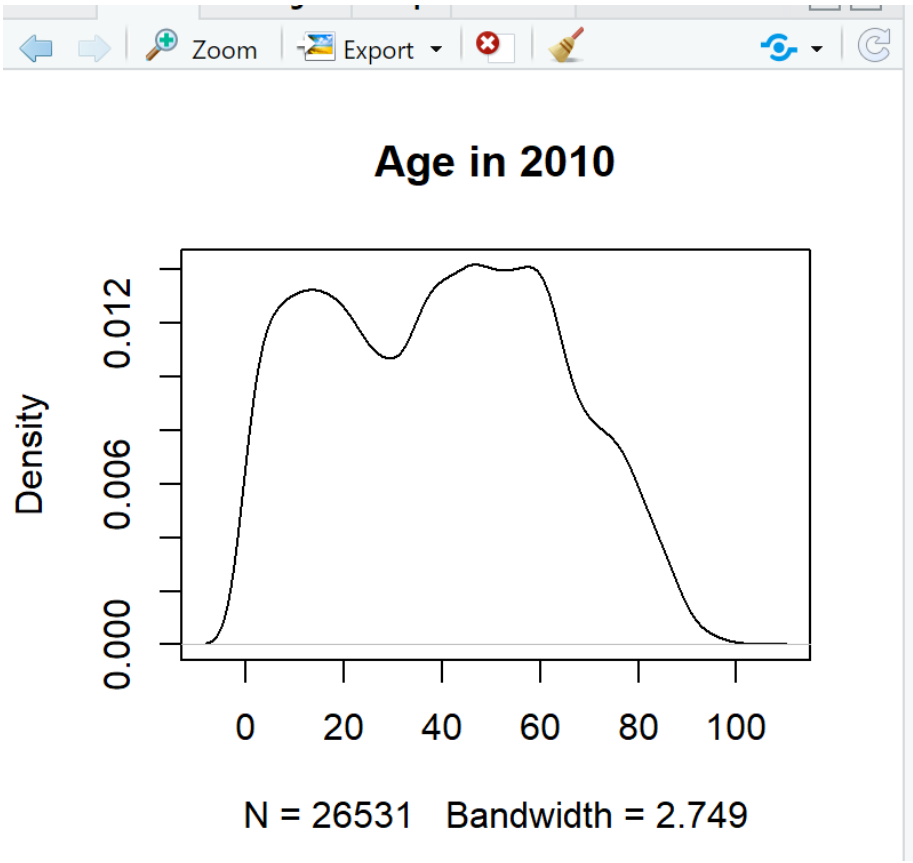
```
         0  11  12  13  21  22  23  31  33  34  35  37  38  42  43  44  45  46  47  48  52  53  54  55
Female  11  30   8  29  63  65   8  68  85 184  50 179  78 258 437   1 153 410  82  22 782  27 584 353
Male    19  57  19  78 213 114  48  98 107 142  59 260 368 110 117   2  95 340 429 215 169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

         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
```

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

They are mean, Standard deviation, inter decile ratio, and Gini coefficients in 2005 and 2019.

```
> data_of_wages_in_2005 <- as.character(print(data_function(wages_in_2005)))
[1] 2.244303e+04 1.807671e+04 8.896525e+00 3.771135e-01
> data_of_wages_in_2019 <- as.character(print(data_function(wages_in_2019)))
[1] 2.757884e+04 2.510719e+04 1.386230e+01 3.990875e-01
```

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


Zoom
Export
Age in 2010
Density
0.000
0.006
0.012
0
20
40
60
80
100
N = 26531 Bandwidth = 2.749

- Number of individuals in Paris in 2011 3514

Exercise 2 Merge Datasets

• Read all individual datasets from 2004 to 2019. Append all these datasets. Presented in R

• Read all household datasets from 2004 to 2019. Append all these datasets. Presented in R

• List the variables that are simultaneously present in the individual and household datasets. V1, idmen, year

• Merge the appended individual and household datasets. In the second part, we use the newly created dataset from the previous to answer the following questions: Presented in R

• Number of households in which there are more than four family members 4201

• Number of households in which at least one member is unemployed 1240

• Number of households in which at least two members are of the same profession 31296

• Number of individuals in the panel that are from household-Couple with kids 11376

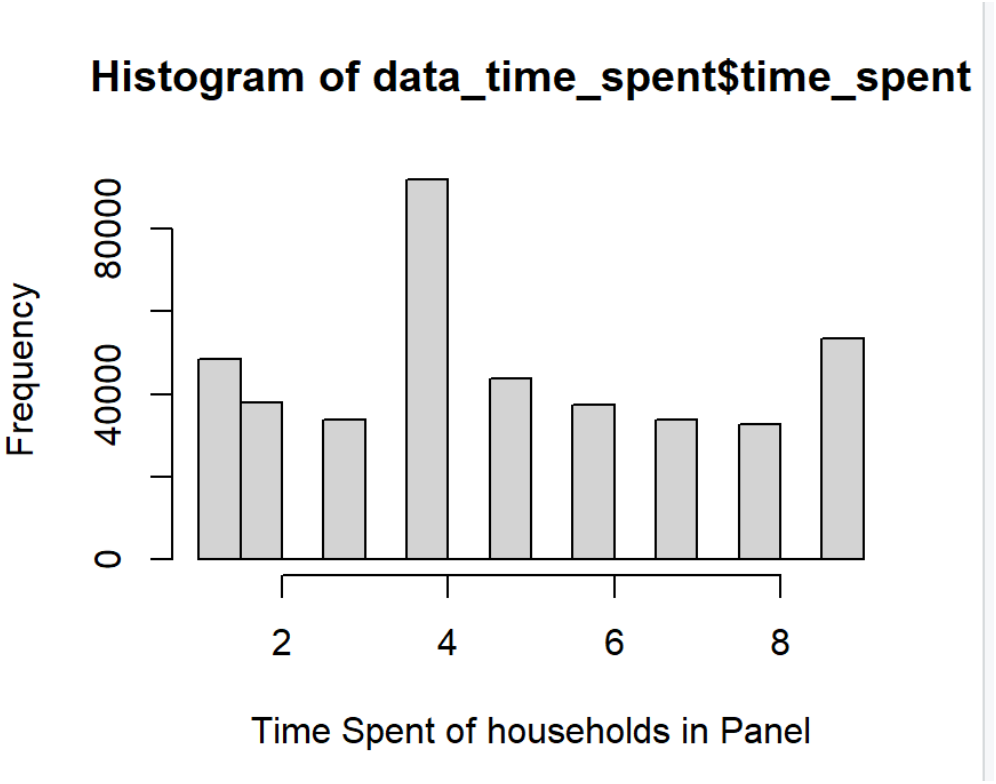
• Number of individuals in the panel that are from Paris. 4665

• Find the household with the greatest number of family members. Report its idmen. n = 14; household reported in 2004 and in 2010; the idmen is shown in R,

• Number of households present in 2010 and 2011 13426

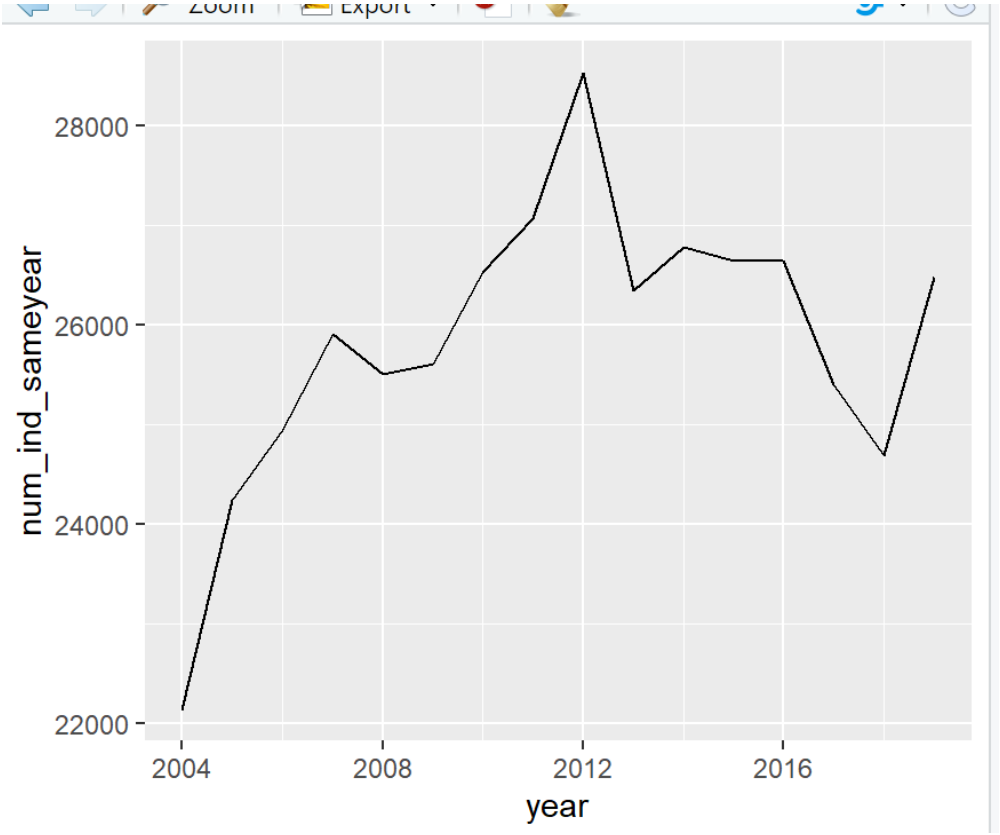
Exercise 3 Migration

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



• 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.

```
> head(data_moved_in, 10)
         idmen        idind year moved_same_year
1  1.50001e+15 1.150001e+18 2008            TRUE
2  1.50001e+15 2.150001e+18 2008            TRUE
3  1.50001e+15 1.150001e+18 2008           FALSE
4  1.50001e+15 2.150001e+18 2008           FALSE
5  1.50001e+15 1.150001e+18 2008           FALSE
6  1.50001e+15 1.150001e+18 2008           FALSE
7  1.50001e+15 1.150001e+18 2008           FALSE
8  1.50001e+15 1.150001e+18 2008           FALSE
9  1.50001e+15 1.150001e+18 2008           FALSE
10 1.50001e+15 1.150001e+18 2008           FALSE
>
```



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

• 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

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. Hint: Construct a year of entry and exit for each individual.