Econ 613 - Assignment 1

Zixuan Qi

January 19, 2022

Exercise 1

Question 1

The number of households surveyed in 2007 is 10498. Because the household identifier is *idmen*, I can delete the repeated *idmen* (if exist) and then calculate the number of rows in *dathh2007*.

Question 2

The number of households with martial status "Couple with kids" in 2005 is 3374.

Question 3

The number of individuals surveyed in 2008 is 25510.

Question 4

The number of individuals aged between 25 and 35 in 2016 is 2765.

Question 5

The part of cross table of gender/profession in 2009 is as follows. Please see the code for the full cross table.

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

Figure 1: Cross Table

Question 6

The mean of wage in 2005 is 11992.26, the standard deviation of wage in 2005 is 17318.56, the inter-decile ratio is infinity, and the Gini coefficient is 0.6319074. The mean of wage in 2019 is 15350.47, the standard deviation of wage in 2019 is 23207.18, the inter-decile ratio is infinity, and the Gini coefficient is 0.6294939.

Question 7

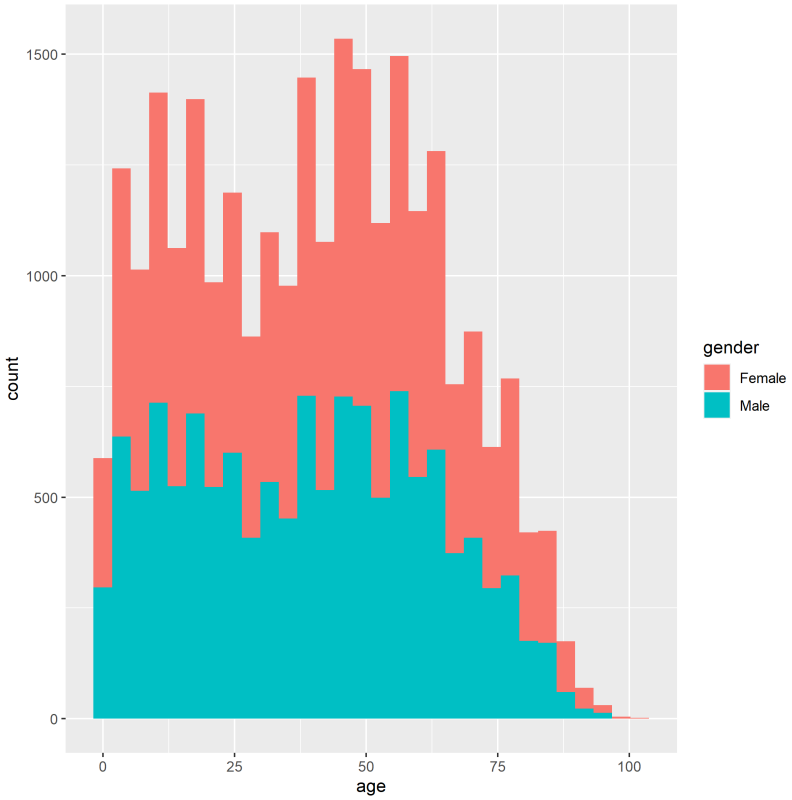

Figure 2: Merged Histogram

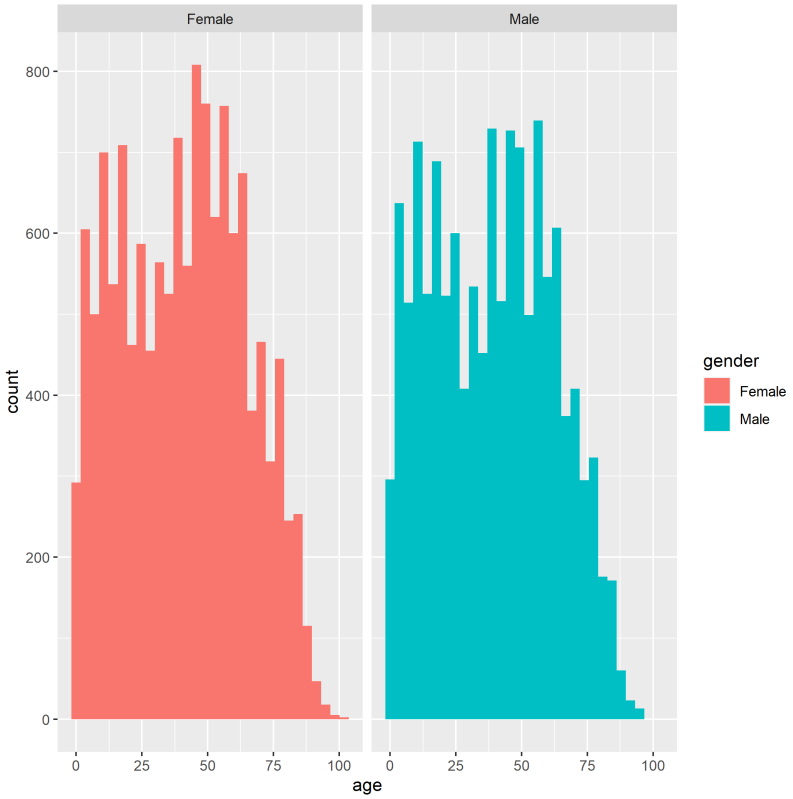

Figure 3: Separable Histogram

In the figure 2 and 3, the histograms of age between male and female are similar.

Question 8

The number of individuals in Paris in 2011 is 3514.

Exercise 2

Question 1

I use datind_all as the name of combined individual dataset. The part of dataset is as follows. Please see the full version in the code.

	Rows	idind	idmen	year	empstat	respondent	profession	gender	age	wage
1:	1	1120001001293010001	1200010012930100	2004	Employed	1	67	Male	31	19187
2:	2	1120001004058010001	1200010040580100	2004	Employed	1	56	Female	30	11586
3:	3	1120001004058010002	1200010040580100	2004	Inactive	0		Female	9	NA
4:	4	1120001006663010001	1200010066630100	2004	Employed	1	38	Male	31	44656
5:	5	1120001006663010002	1200010066630100	2004	Employed	0	45	Female	27	20413

Figure 4: First 5 Rows of Datind

Question 2

I use *dathh_all* as the name of combined household dataset. The part of dataset is as follows. Please see the full version in the code.

	Rows	idmen	year	datent	myear	mstatus	move	location
1:	1	1200010012930100	2004	2000	2000	Single	NA	Paris
2:	2	1200010040580100	2004	2001	2001	Single Parent	NA	Paris
3:	3	1200010066630100	2004	2000	2000	Couple, No kids	NA	Paris
4:	4	1200010082450100	2004	1957	1957	Single	NA	Paris
5:	5	1200010086440100	2004	2001	2001	Couple, No kids	NA	Paris

Figure 5: First 5 Rows of Dathh

Question 3

Note that I change the name of row column as Rows. The same variables are *Rows*, *idmen*, and *year*

```
c("Rows", "idmen", "year")
```

Figure 6: Same Variables

Question 4

The first two rows of the complete data is as follows.

	Rows	idind	idmen	year	empstat	respondent	profession	gender	age	wage
1:	1	1120001001293010001	1200010012930100	2004	Employed	1	67	Male	31	19187
2:	2	1120001004058010001	1200010040580100	2004	Employed	1	56	Female	30	11586

	datent	myear	mstatus	move	location
1:	2000	2000	Single	NA	Paris
2:	2001	2001	Single Parent	NA	Paris

Figure 7: Complete Data

Question 5

The number of households in which there are more than four family members is 3622.

Question 6

The number of households in which at least one member is unemployed 8162.

Question 7

The number of households in which at least two members are of the same profession is 8752.

Question 8

The number of individuals in the panel that are from household - Couple with kid is 55094.

Question 9

The number of individuals in the panel that are from Paris is 14563.

Question 10

The household with the most number of family members is *2202243098040100*

Question 11

The number of households present in 2010 and 2011 is 22410.

Exercise 3

Question 1

The distribution of the time spent in the survey for each household is as follows.

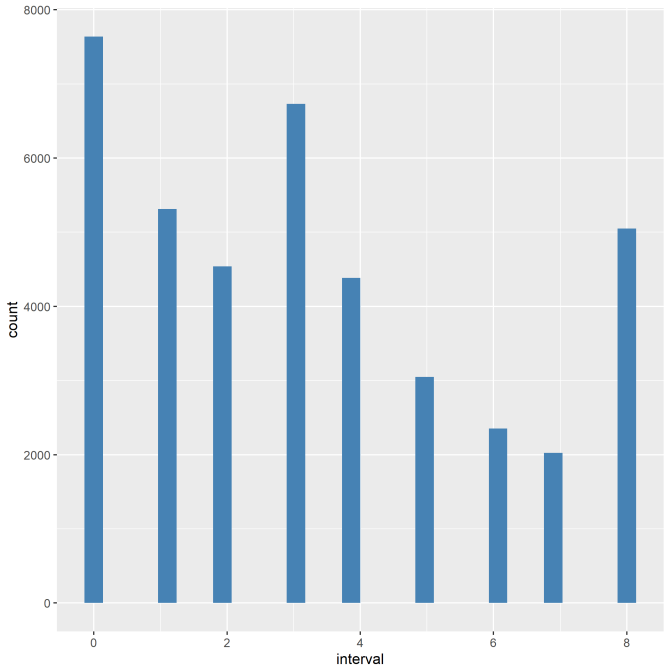

Figure 8: The Distribution of Time Spent

Question 2

Whether or not the household moved into its current dwelling at the year of survey is as follows.

	idind	idmen	year	datent	dwelling	share
	<chr>	<chr>	<int>	<int>	<dbl>	<dbl>
1	1120001001293010001	1200010012930100	2004	2000	0	0.0273
2	1120001004058010001	1200010040580100	2004	2001	0	0.0273
3	1120001004058010002	1200010040580100	2004	2001	0	0.0273
4	1120001006663010001	1200010066630100	2004	2000	0	0.0273
5	1120001006663010002	1200010066630100	2004	2000	0	0.0273
6	1120001008245010001	1200010082450100	2004	1957	0	0.0273
7	1120001008644010001	1200010086440100	2004	2001	0	0.0273
8	1120001008644010002	1200010086440100	2004	2001	0	0.0273
9	1120001010299010001	1200010102990100	2004	1990	0	0.0273
10	1120001010299010002	1200010102990100	2004	1990	0	0.0273

Figure 9: Dwelling

The plot of shares of dwelling in each year is as follows. There is not any household which did not move into its current dwelling at the year of survey.

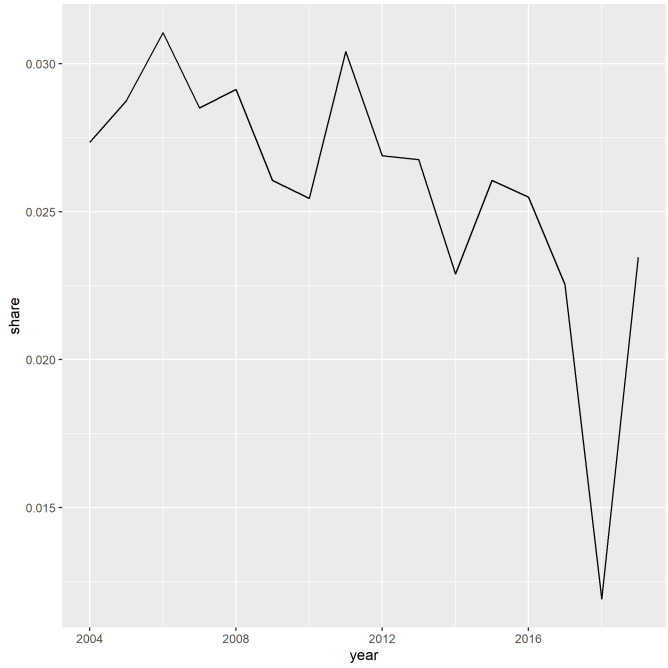

Figure 10: Shares of Current Dwelling

Question 3

Whether or not household migrated at the year of survey is as follows.

	idind	idmen	year	myear	move	share
	<chr>	<chr>	<int>	<int>	<dbl>	<dbl>
1	1120001001293010001	1200010012930100	2004	2000	1	0.0282
2	1120001004058010001	1200010040580100	2004	2001	1	0.0282
3	1120001004058010002	1200010040580100	2004	2001	1	0.0282
4	1120001006663010001	1200010066630100	2004	2000	1	0.0282
5	1120001006663010002	1200010066630100	2004	2000	1	0.0282
6	1120001008245010001	1200010082450100	2004	1957	1	0.0282
7	1120001008644010001	1200010086440100	2004	2001	1	0.0282
8	1120001008644010002	1200010086440100	2004	2001	1	0.0282
9	1120001010299010001	1200010102990100	2004	1990	1	0.0282
10	1120001010299010002	1200010102990100	2004	1990	1	0.0282

Figure 11: Move or Not

The plot of shares of moving at the year of survey in each year is as follows.

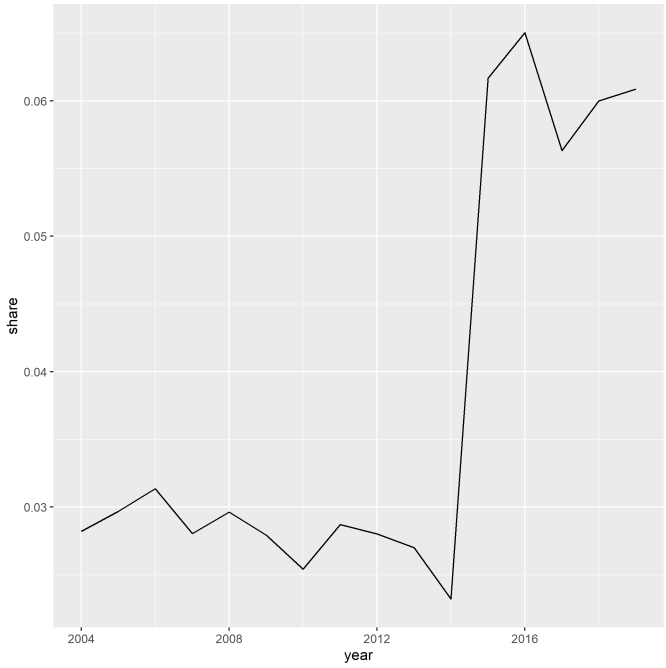

Figure 12: Shares of Moving

Question 4

I prefer the first method because it uses the same standard to assess whether the household migrate at the year of survey. However, the second method uses two variables, and I cannot know whether these two variables have consistent judgement to move. In the figure, you can see, there exists a rapid increasing in 2014, which shows that the judgment to move of these two variables may be different.

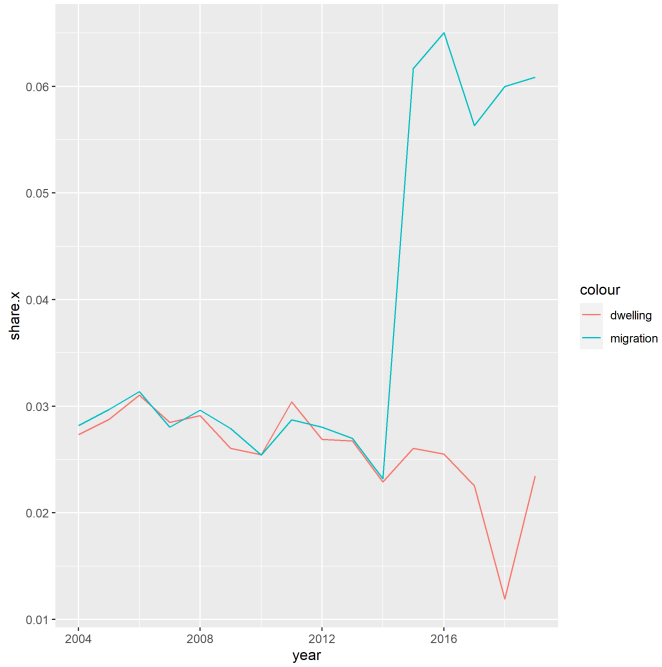

Figure 13: Shares of Moving

Question 5

The number of households which had at least one family member changed his/her profession or employment status is 2245.

Exercise 4

The attribution dataset is as follows.

	year	proportion
	<int>	<dbl>
1	2005	0.296
2	2006	0.374
3	2007	0.698
4	2008	0.689
5	2009	0.704
6	2010	0.784
7	2011	0.733
8	2012	1.13
9	2013	1.28
10	2014	1.31
11	2015	1.44
12	2016	1.80
13	2017	2.29
14	2018	3.55

Figure 14: Proportion