#1.1 Number of households surveyed in 2007

10498

#1.2 Number of households with a 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.

I assumed between 25 and 35, exclusive.

2237

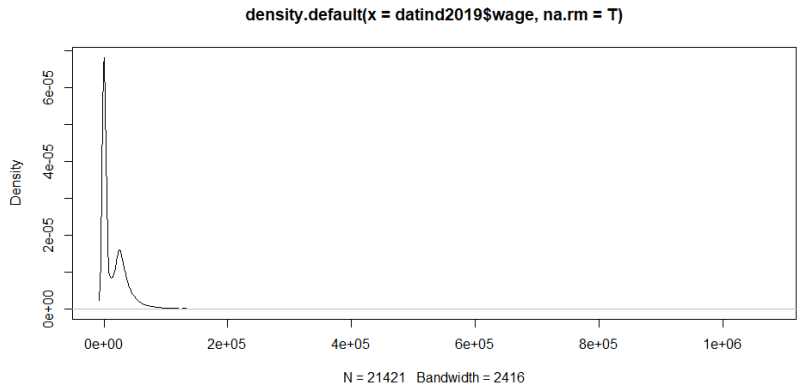
#1.5 Cross-table gender/profession in 2009

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

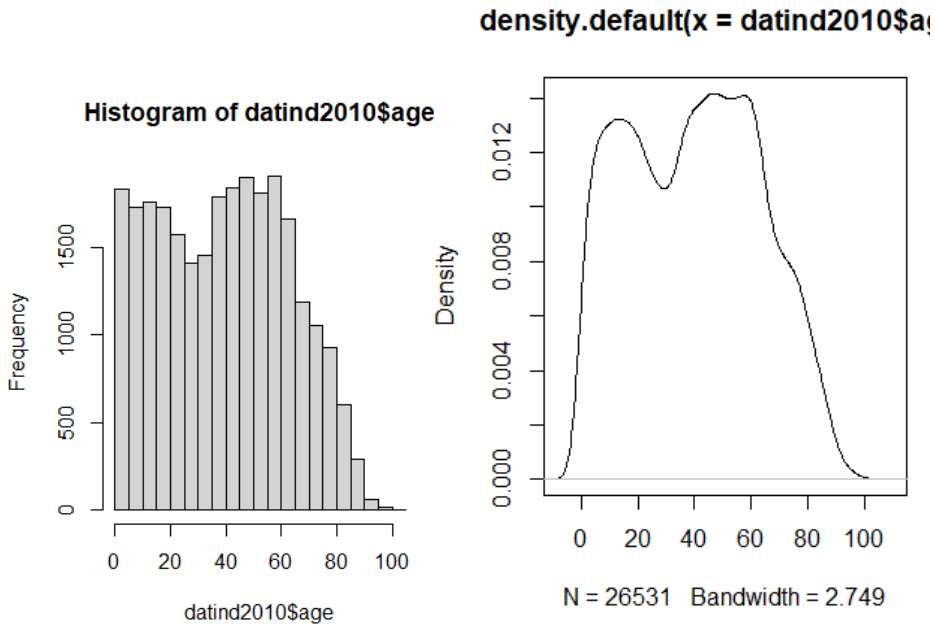
	45	46	47	48	52	53	54	55	56	62	63	64	65	67	68	69
Female	153	410	82	22	782	27	584	353	696	64	35	29	19	147	120	40
Male	95	340	429	215	169	182	98	101	74	443	520	246	159	237	177	82

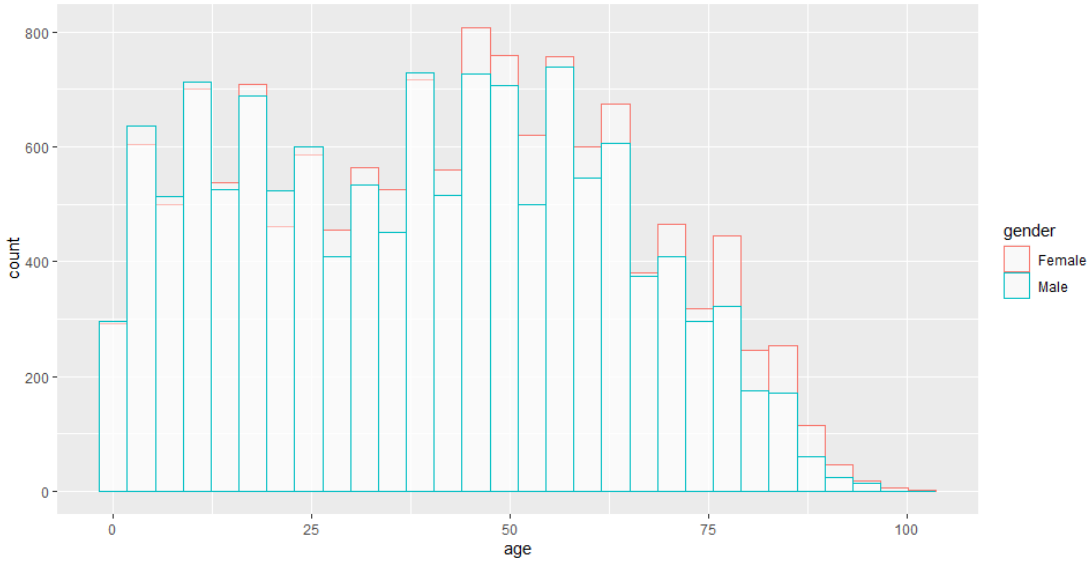
#1.6 Distribution of wages in 2005 and 2019. Report the mean, the standard deviation, the inter-decile ratio D9/D1 and the Gini coefficient. Write a function to compute.

2005	mean	sd	D9/D1	Gini
	11992.26	17318.56	Inf	0.6671654
2019	mean	sd	D9/D1	Gini
	15350.47	23207.18	Inf	0.6655301



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





Roughly speaking, there is no gender difference in wage by this graph.

#1.8 Number of individuals in Paris in 2011.

3514

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

Here, I only showed the first six rows of my appended data.

V1	idind	idmen	year	empstat	respondent
1:	1 1120001001293010001	1200010012930100	2004	Employed	1
2:	2 1120001004058010001	1200010040580100	2004	Employed	1
3:	1 1120001004058010001	1200010040580100	2005	Inactive	1
4:	3 1120001004058010002	1200010040580100	2004	Inactive	0
5:	2 1120001004058010002	1200010040580100	2005	Inactive	0
6:	4 1120001006663010001	1200010066630100	2004	Employed	1

	profession	gender	age	wage
1:	67	Male	31	19187
2:	56	Female	30	11586
3:		Female	31	12334
4:		Female	9	NA
5:		Female	10	NA
6:	38	Male	31	44656

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

V1	idmen	year	datent	myear	mstatus	move	location
1	1200010012930100	2004	2000	2000	Single	*NA*	Paris
2	1200010040580100	2004	2001	2001	Single Parent	*NA*	Paris
1	1200010040580100	2005	2001	2001	Single Parent	*NA*	Paris
3	1200010066630100	2004	2000	2000	Couple, No kids	*NA*	Paris
2	1200010066630100	2005	2005	2005	Couple, No kids	*NA*	Paris
4	1200010082450100	2004	1957	1957	Single	*NA*	Paris
3	1200010082450100	2005	1957	1957	Single	*NA*	Paris
5	1200010086440100	2004	2001	2001	Couple, No kids	*NA*	Paris
4	1200010086440100	2005	2001	2001	Couple, No kids	*NA*	Paris

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

"V1" "idmen" "year"

#2.4 Merge the appended individual and household datasets.

	idmen	year	datent	myear	mstatus	move	location	idind	empstat	respondent	profession	gender	age	wage
1:	1200010012930100	2004	2000	2000	Single	NA	Paris	1120001001293010001	Employed	1	67	Male	31	19187
2:	1200010040580100	2004	2001	2001	Single Parent	NA	Paris	1120001004058010001	Employed	1	56	Female	30	11586
3:	1200010040580100	2004	2001	2001	Single Parent	NA	Paris	1120001004058010002	Inactive	0		Female	9	NA
4:	1200010040580100	2005	2001	2001	Single Parent	NA	Paris	1120001004058010001	Inactive	1		Female	31	12334
5:	1200010040580100	2005	2001	2001	Single Parent	NA	Paris	1120001004058010002	Inactive	0		Female	10	NA
6:	1200010066630100	2004	2000	2000	Couple, No kids	NA	Paris	1120001006663010001	Employed	1	38	Male	31	44656

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

year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
number	745	814	862	874	814	810	821	785	816	748	783	763	753	703	647	692

#2.6 Number of households in which at least one member is unemployed

year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
number_unemp	950	1039	1030	975	909	1045	1109	1071	1205	1177	1187	1227	1137	1103	991	1086

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

year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
num	445	496	480	490	459	452	475	491	517	457	475	468	473	457	454	497

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

year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
number_allcoupwkids	11993	13217	13637	13963	13481	13286	13726	13801	14403	13097	13228	13008	12967	11963	11444	12151

#2.9 Number of individuals in the panel that are from Paris.

year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
number_allParis	3494	3734	3658	3735	3559	3524	3607	3514	3679	2288	2576	3033	2946	2836	2797	2924

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

Over 16 years, the most number of family members is 14 in 2007 and 2010

Idmen in 2007 is 2207811124040100

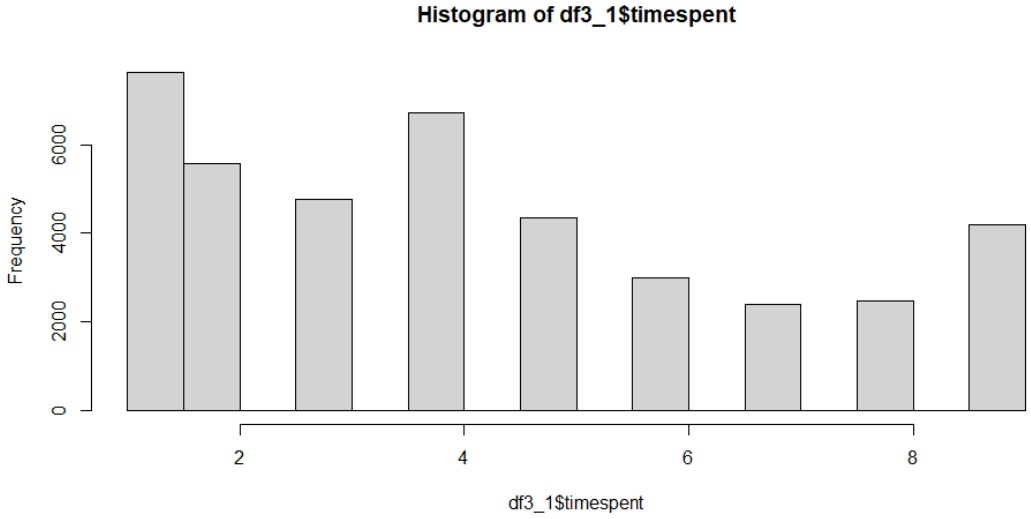
Idmen in 2010 is 2510263102990100

#2.11 Number of households present in 2010 and 2011.

8984

#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.

Here, I only listed enter and exit year of first 9 households in my data.

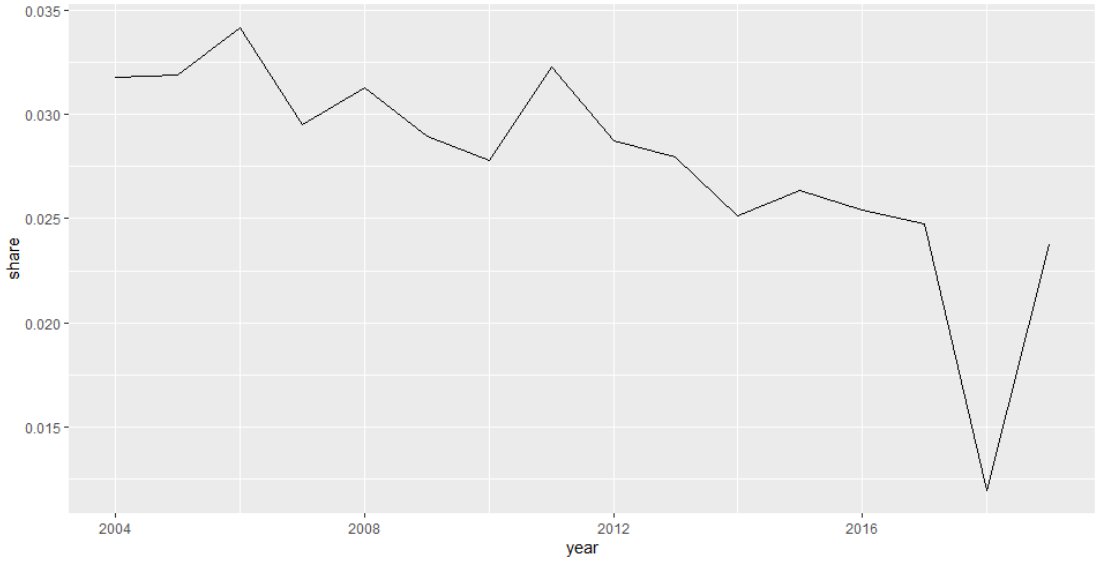


	idmen	enter	exit
1	1200010012930100	2004	2004
2	1200010040580100	2004	2005
3	1200010066630100	2004	2005
4	1200010082450100	2004	2005
5	1200010086440100	2004	2005
6	1200010102990100	2004	2005
7	1200010118450100	2004	2005
8	1200020012930100	2004	2005
9	1200020017390100	2004	2005

#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	sum(year == datent)	TF
1	1200010012930100	2004	0	0
2	1200010040580100	2004	0	0
3	1200010040580100	2005	0	0
4	1200010066630100	2004	0	0
5	1200010066630100	2005	2	1
6	1200010082450100	2004	0	0
7	1200010082450100	2005	0	0
8	1200010086440100	2004	0	0
9	1200010086440100	2005	0	0
10	1200010102990100	2004	0	0

TF=0 indicates year is not equal to datent, in other words, a household didn't move into its current dwelling at the year of the survey. For the first 10 rows of merged data, only the household with idmen 1200010066630100 in year 2005 satisfies the condition.

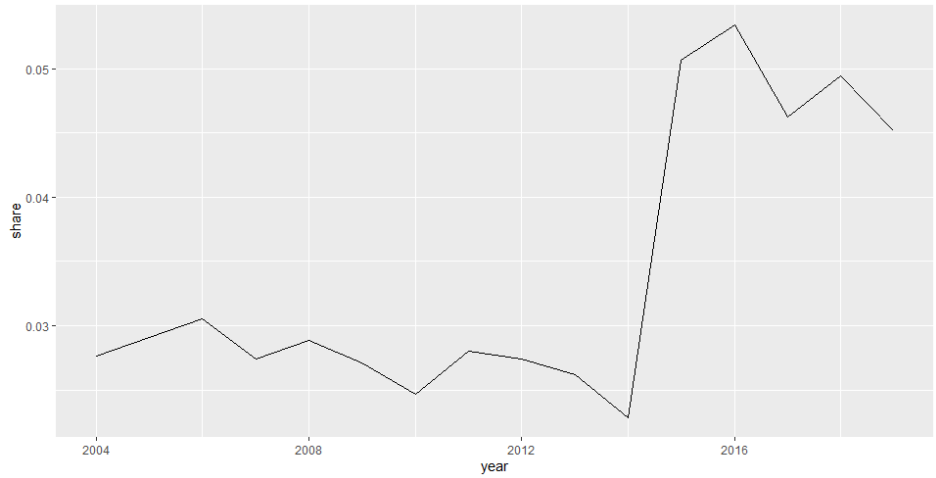


#3.3 Based on myear and move, identify whether or not household migrated at the year of survey. Report the first 10 rows of your result and plot the share of individuals in that situation across years.

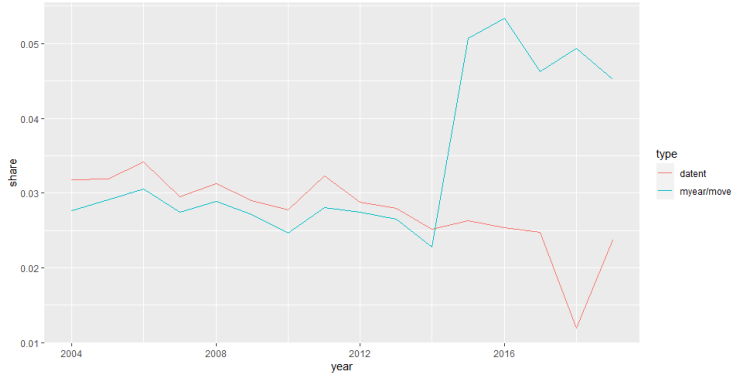
```
              idmen year movebefore2014 moveafter2014
 1: 1200010012930100 2004              0            NA
 2: 1200010040580100 2004              0            NA
 3: 1200010040580100 2005              0            NA
 4: 1200010066630100 2004              0            NA
 5: 1200010066630100 2005              1            NA
 6: 1200010082450100 2004              0            NA
 7: 1200010082450100 2005              0            NA
 8: 1200010086440100 2004              0            NA
 9: 1200010086440100 2005              0            NA
10: 1200010102990100 2004              0            NA
```

Owing to limited width, I showed my first 10 rows of result selecting key variables only, idmen, year,

movebefore2014, and moveafter2014.



#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



By my observation of "move" variables,

Suppose an individual enters the panel in 2013, and he didn't do a survey in 2014 and 2015. In 2016, he reentered the panel.

Year Datent move

2013

2016 2014 2

I prefer the method defined in 3.2 (i.e. Datent) over that in 3.3 since I care more about which exact year does an individual move rather than whether one moves or not during the period.

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

Consider two years as a period, i.e. (2004,2005), (2005,2006)…e.t.c.

Case1: We include the condition: NA x (where x is some number) as a change.

```
> change
      [,1] [,2]
 [1,] 2005   85
 [2,] 2006  102
 [3,] 2007   85
 [4,] 2008  105
 [5,] 2009  108
 [6,] 2010   78
 [7,] 2011  108
 [8,] 2012  111
 [9,] 2013   88
[10,] 2014   73
[11,] 2015   88
[12,] 2016   79
[13,] 2017   64
[14,] 2018   32
[15,] 2019   69
```

Case2: We don't include the condition: NA x (where x is some number) as a change. Obviously, the number of households in Case 2 will be lower.

```
> ncp
      [,1] [,2]
 [1,] 2005   81
 [2,] 2006   97
 [3,] 2007   80
 [4,] 2008   80
 [5,] 2009   91
 [6,] 2010   77
 [7,] 2011  100
 [8,] 2012  103
 [9,] 2013   79
[10,] 2014   64
[11,] 2015   79
[12,] 2016   71
[13,] 2017   55
[14,] 2018   31
[15,] 2019   61
```

#Exercise4

#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.

```
      [,1]      [,2]
 [1,] 2005 0.1352962
 [2,] 2006 0.2000743
 [3,] 2007 0.1787089
 [4,] 2008 0.2266955
 [5,] 2009 0.2056056
 [6,] 2010 0.1837882
 [7,] 2011 0.1936226
 [8,] 2012 0.1698866
 [9,] 2013 0.2557651
[10,] 2014 0.2201301
[11,] 2015 0.2192481
[12,] 2016 0.2172722
[13,] 2017 0.2507599
[14,] 2018 0.2441934
[15,] 2019 0.2433395
```