Homework 1

Jiaxin Li

Exercise 1

1.1

```
[1] 10498
```

1.2

```
[1] 3374
```

1.3

```
[1] 25510
```

1.4

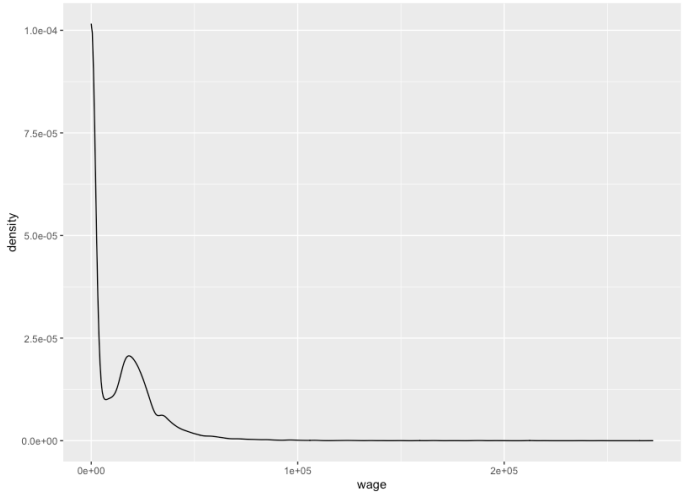
```
[1] 2765
```

1.5

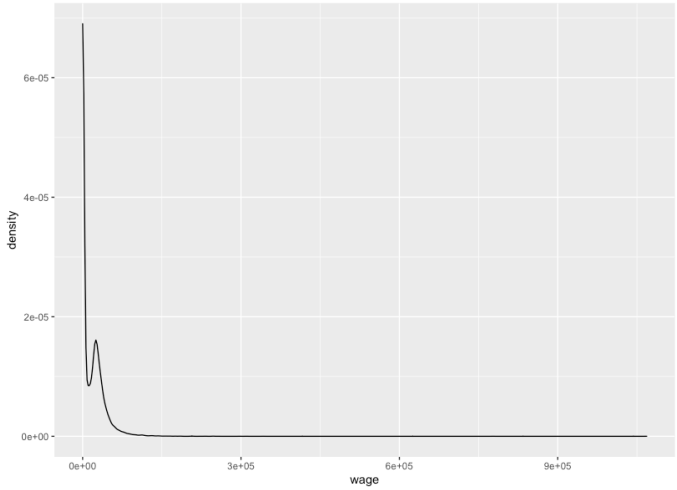
```
# A tibble: 66 x 3
# Groups:   profession [33]
   profession gender     n
        <int> <chr>  <int>
 1          0 Female    11
 2          0 Male      19
 3         11 Female    30
 4         11 Male      57
 5         12 Female     8
 6         12 Male      19
 7         13 Female    29
 8         13 Male      78
 9         21 Female    63
10         21 Male     213
# … with 56 more rows
```

1.6

Distribution of wage in 2005



Distribution of wage in 2019



Summary of wage in 2005

```
      MEAN      STD     IDR
1 11992.26 17318.56 32340.4

[1] 0.6671654
```

Summary of wage in 2019

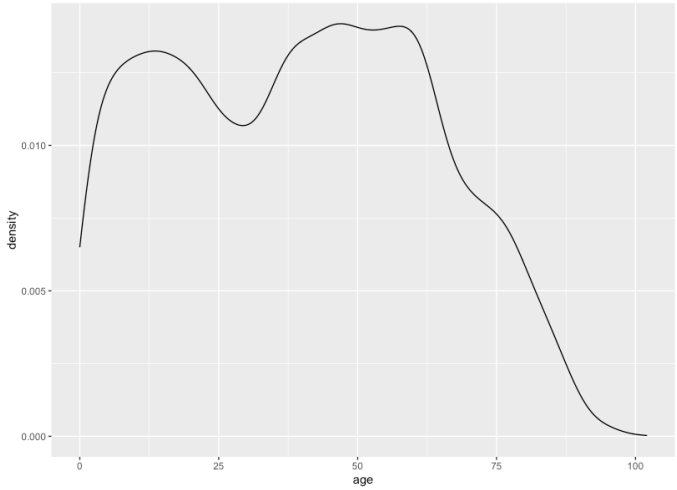
```
      MEAN      STD   IDR
1 15350.47 23207.18 40267

[1] 0.6655301
```

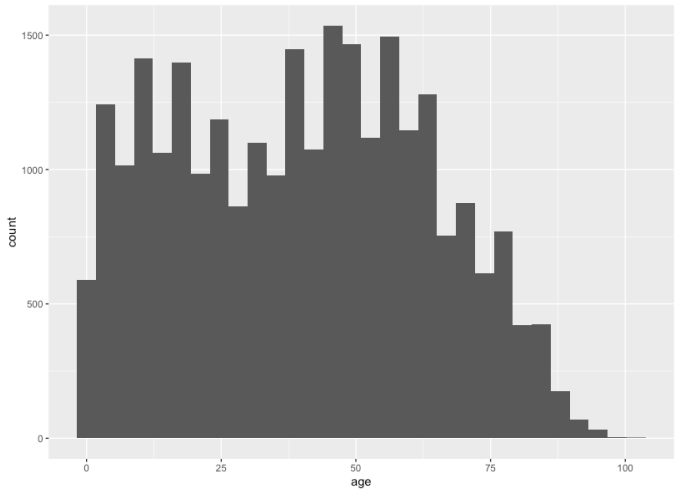
where IDR = D9 - D1, since cannot calculate D9/D1, D1=0

1.7

Distribution of age in 2010

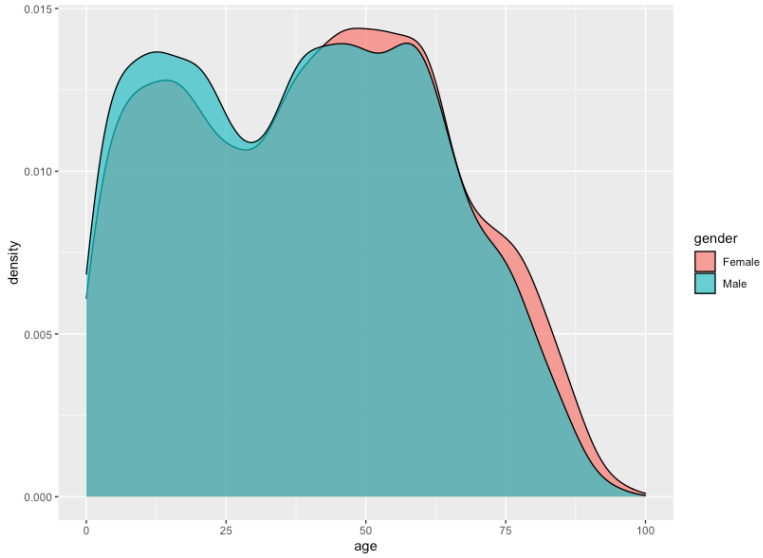


Histogram of age in 2010



There are differences between men and women.

In 2010, the number of men who is about 50 and older than 70 is greater than women. Meanwhile, the number of man who is younger than 30 is less than woman.



1.8

```
[1] 3514
```

Exercise 2

2.1.1 & 2.1.2

Files reading is completed in R programming.

2.1.3

```
[1] "X"     "idmen" "year"
```

2.1.4

Merging is completed in R programming.

2.2.1

```
[1] 18103
```

```
   year    n
1  2004 6412
2  2005  814
3  2006  862
4  2007  874
5  2008  814
6  2009  810
7  2010  821
8  2011  785
9  2012  816
10 2013  754
11 2014  783
12 2015  763
13 2016  753
14 2017  703
15 2018  647
16 2019  692
```

2.2.2

```
[1] 17242
```

```
   year    n
1  2004  950
2  2005 1039
3  2006 1030
4  2007  975
5  2008  909
6  2009 1045
7  2010 1110
8  2011 1071
9  2012 1205
10 2013 1177
11 2014 1187
12 2015 1227
13 2016 1137
14 2017 1103
15 2018  991
16 2019 1086
```

2.2.3

```
[1] 15681
```

```
   year    n
1  2004 8511
2  2005  497
3  2006  485
4  2007  492
5  2008  460
6  2009  453
7  2010  477
8  2011  492
9  2012  517
10 2013  460
11 2014  477
12 2015  469
13 2016  475
14 2017  459
15 2018  457
16 2019  500
```

2.2.4

```
[1] 58670
```

```
   year    n
1  2004 3060
2  2005 3522
3  2006 3668
4  2007 3820
5  2008 3752
6  2009 3727
7  2010 3899
8  2011 3922
9  2012 4136
10 2013 3770
11 2014 3796
12 2015 3744
13 2016 3729
14 2017 3436
15 2018 3264
16 2019 3425
```

2.2.5

```
[1] 22208
```

```
   year    n
1  2004 1429
2  2005 1540
3  2006 1507
4  2007 1557
5  2008 1499
6  2009 1526
7  2010 1575
8  2011 1552
9  2012 1605
10 2013  946
11 2014 1088
12 2015 1325
13 2016 1305
14 2017 1252
15 2018 1223
16 2019 1279
```

2.2.6

```
[1] "1208045118450100" "1607839058220100" "1610263040580100" "1804363114960100"
```

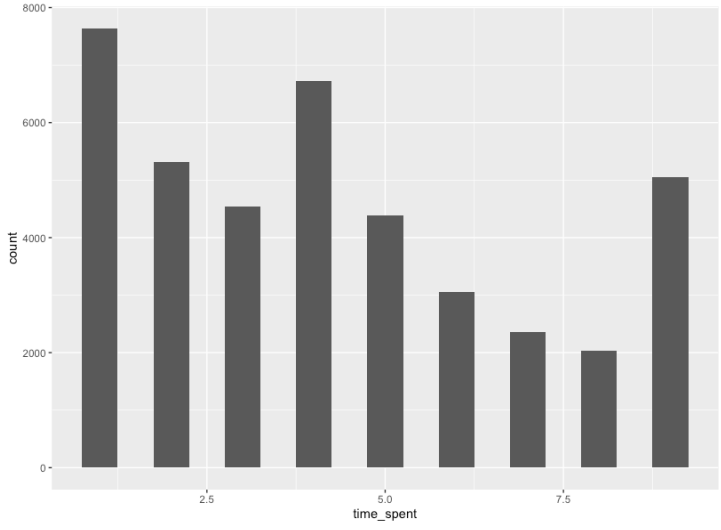
2.2.7

```
[1] 22410
```

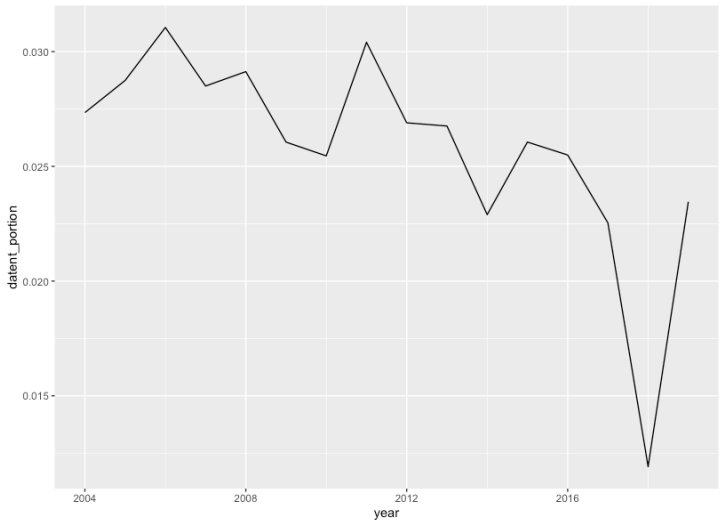
```
  year     n
1 2010 11050
2 2011 11360
```

Exercise 3

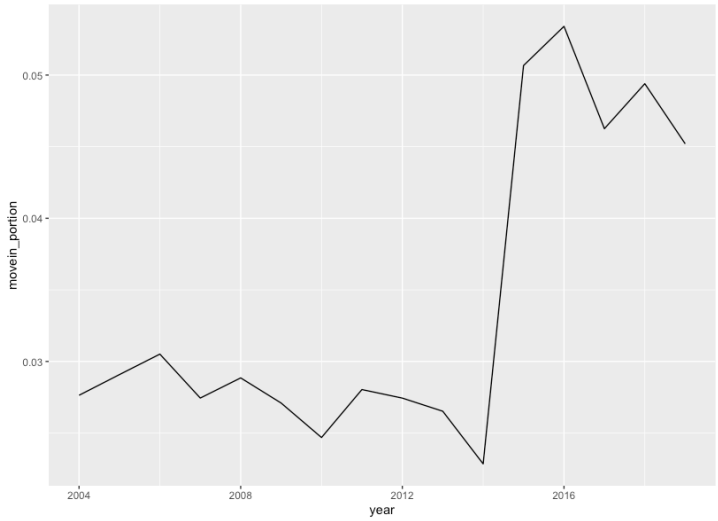
3.1



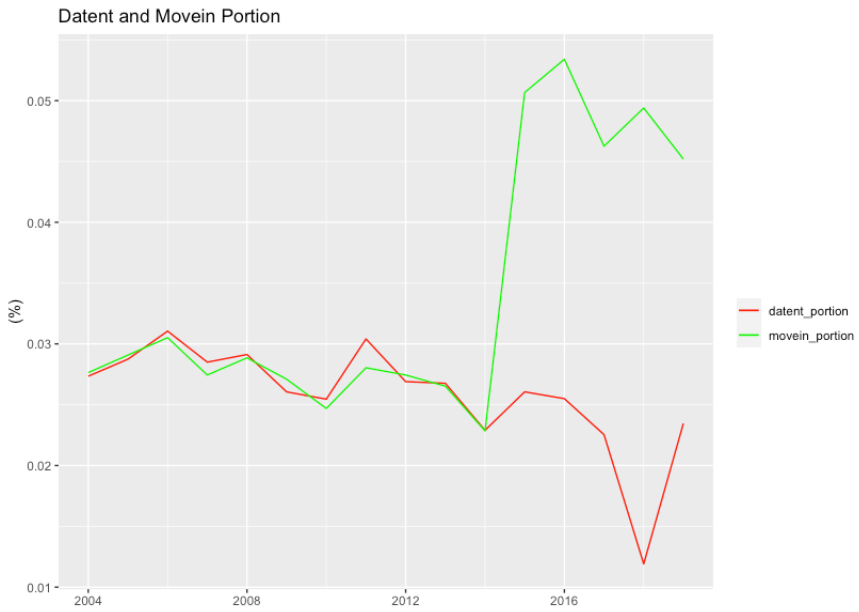
3.2


datent_portion
0.030
0.025
0.020
0.015
2004
2008
2012
2016
year

3.3


movein_portion
0.05
0.04
0.03
2004
2008
2012
2016
year

3.4


Datent and Movein Portion
(%)
0.05
0.04
0.03
0.02
0.01
2004
2008
2012
2016
datent_portion
movein_portion

3.5

```
[1] 11231
```

Exercise 4

	year	attrition_rate
	<int>	<dbl>
1	2004	0.108
2	2005	0.169
3	2006	0.152
4	2007	0.202
5	2008	0.179
6	2009	0.159
7	2010	0.162
8	2011	0.145
9	2012	0.197
10	2013	0.180
11	2014	0.180
12	2015	0.181
13	2016	0.203
14	2017	0.202
15	2018	0.230