Exploratory Data Analysis

- Student ID: 22000457

- Name: Sihyeon Yoo

- 1st Major: Ict Convergence

- 2nd Major: Data science

Suggest project to identify and compare the divorce rate according to work type and spouse's work type. Topic is about finding out whether the types of labor affect divorce.

1. Data overview

Data topic : Interview history information for 59600 people.
Sample size : 59600 * 11232
Number of variables : 11232
Data type : Structured data(data frame), It has both qualitative data and quantitative data.
Data range : 59600 rows

2. Univariate analysis

2.1 LABOR.FORCE.STATUS_labels

1. Data type : Qualitative Data
2. Data meaning : Respondent's labor force status
3. Value types : (1) Working Fulltime (2) Keeping House (3) retired (4)Working Parttime (5) Unempl,Laid off (6) School (7) Temp not working (8) Other
4. Frequency Distribution Table

	LABOR.FORCE.STATUS_labels	frequency
1	WORKING FULLTIME	29437
2	KEEPING HOUSE	9650
3	RETIRED	8102
4	WORKING PARTTIME	6115
5	UNEMPL, LAID OFF	1977
6	SCHOOL	1841
7	TEMP NOT WORKING	1253
8	OTHER	1208
9		1

Figure 1. Respondent' labor force status Frequency distribution Table.
The 'working fulltime' is overwhelmingly the highest frequency. The majority of the respondents were 'working fulltime' or 'keeping house' and 'retired'.

2.2 SPOUSE.LABOR.FORCE.STATUS_labels

1. Data type : Qualitative Data
2. Data meaning : Spouse's labor force status
3. Value types : (1) Working Fulltime (2) Keeping house (3) Retired (4)Working Parttime (5) Unempl,Laid off (6) School (7) Temp Not Working (8) Other (9) IAP
4. Frequency Distribution Table

	SPOUSE.LABOR.FORCE.STATUS_labels	frequency
1	IAP	27659
2	WORKING FULLTIME	17425
3	KEEPING HOUSE	5648
4	RETIRED	3758
5	WORKING PARTTIME	2861
6	TEMP NOT WORKING	622
7	UNEMPL, LAID OFF	610
8	OTHER	583
9	SCHOOL	347
10		1

Figure 2. Spouse Frequency distribution Table
The IAP was regarded as an answer that did not have meaning, and excluding this, the spouse's job frequency was the same as the respondent, with working full time the most, and the second was keeping house. The first to the fourth frequency was the same result with respondent's labor force status.

2.3 EVER.BEEN.DIVORCED.OR.SEPARATED

1. Data type : Quantitative Data
2. Data meaning : Respondent's divorce experience
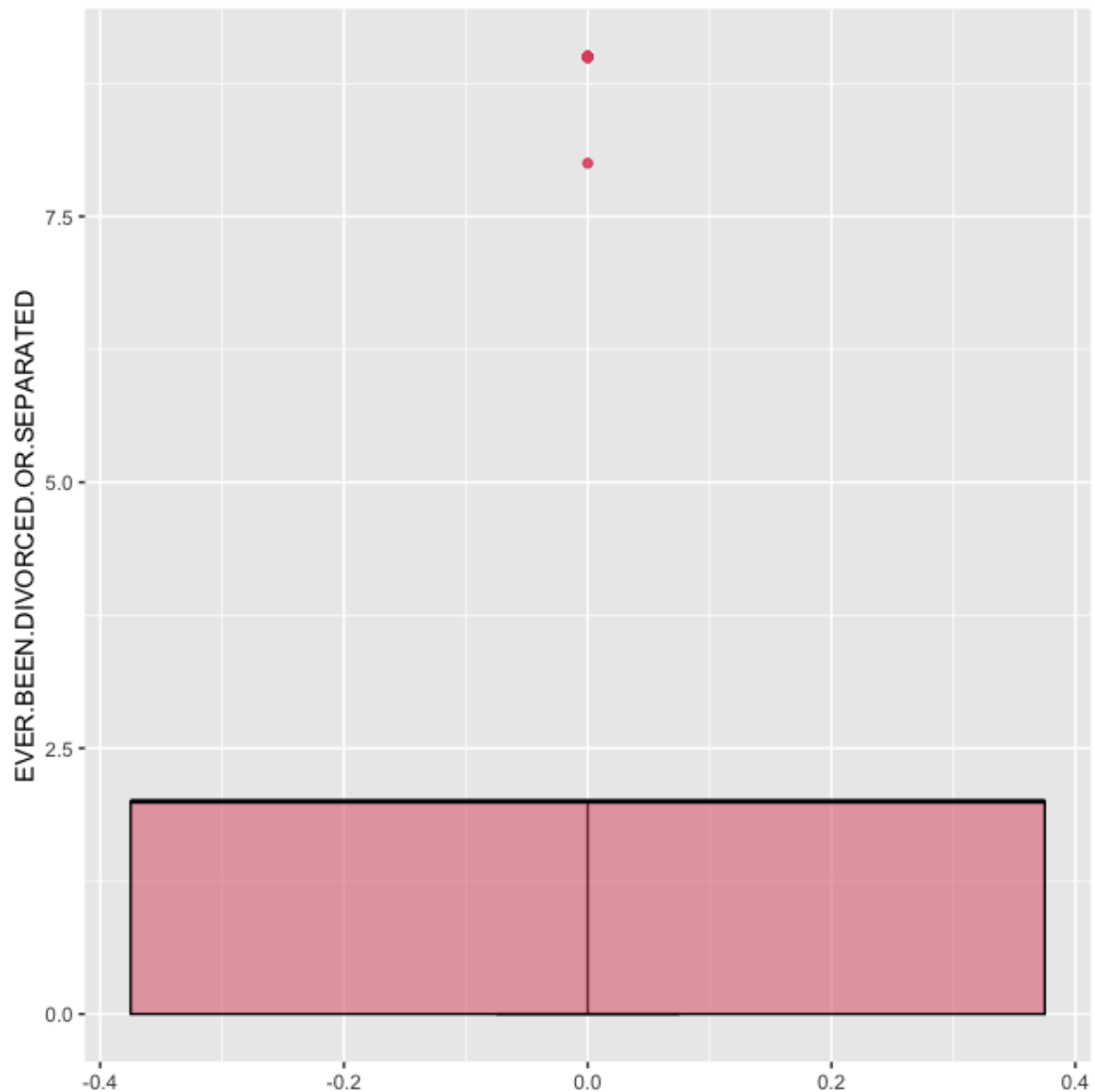3. Value type : integer(0~)
4. Box plot

Figure 3. Divorce experience box plot

Since the minimum value is 0 and the third quartile is 2.5 or less, it may be seen that most of the values are between 0 and 2, and a value of 7 or more, which is an outlier, was found.

3. Multivariate analysis

3.1 Distribution

1. Comparison of the number of divorces by labor force status (using Box plot)
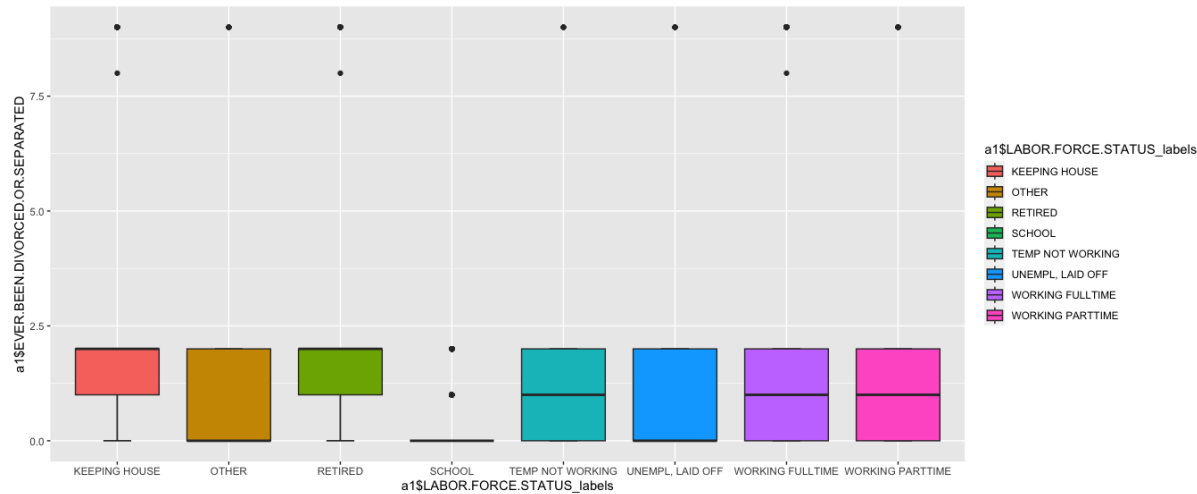
Figure 4. Divorce experience number group by labor force status.
It represents the variance of the number of divorce experiences according to the respondent's labor force status. The first quartile and the third quartile are generally similar, but the first quartile of keeping house and retired is higher than that of other groups. In the case of school, it can be seen that most of the number of divorces is distributed at zero. And every group except school contains an outlier of 7 or more.

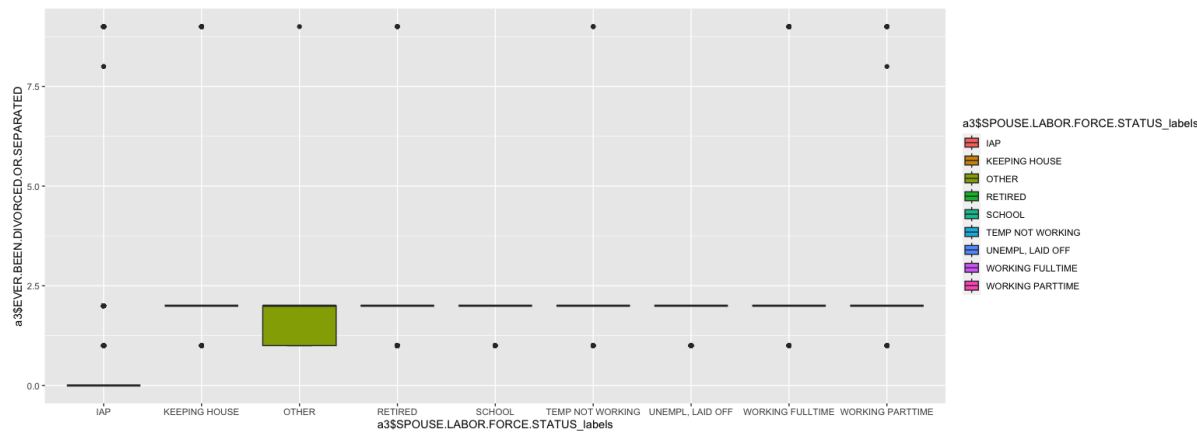2. Comparison of the number of divorces by spouse's labor force status (using Box plot)

Figure 5. Divorce experience number group by spouse's labor force status.
It can be seen that all groups have a median value of about 2.3. In addition, except for 'school' and 'unempl,laid off', it can be seen that they have outliers of 7 or more and 1 or less.

3.2 Correlation

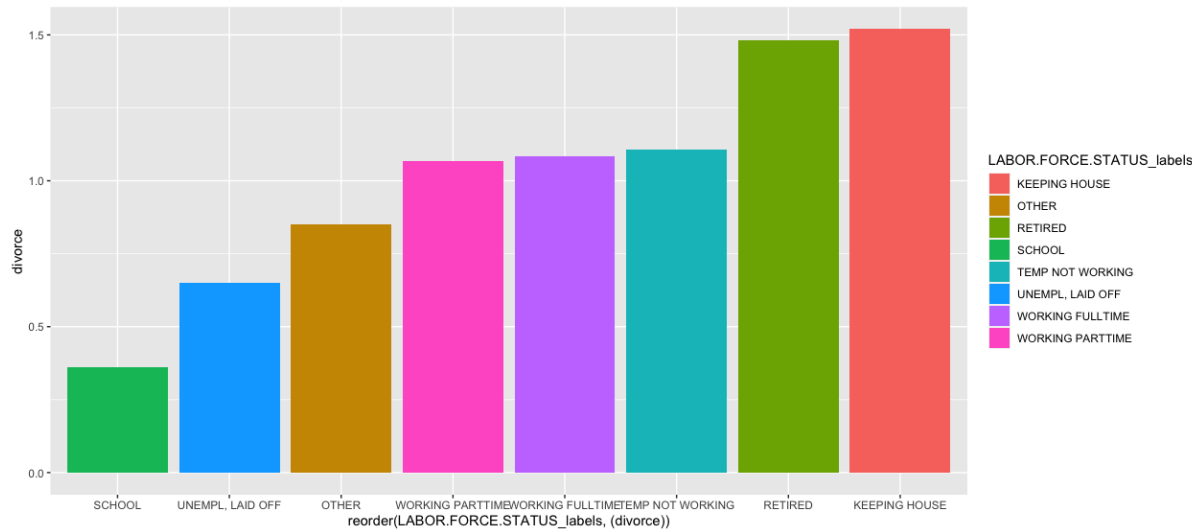1. Correlation between labor force status and divorce experience.

Figure 6. Mean of divorce experience number group by labor force status

The relationship was confirmed by summing the number of divorces according to the respondent's labor force status. Due to the influence of frequency within the labor force status, the number of divorces was high due to the frequency of labor force status. Therefore, the average, not the sum of the number of divorces, was calculated, it was confirmed that keeping house or registered had a large number of divorces.

2. Correlation between spouse's labor force status and divorce experience.
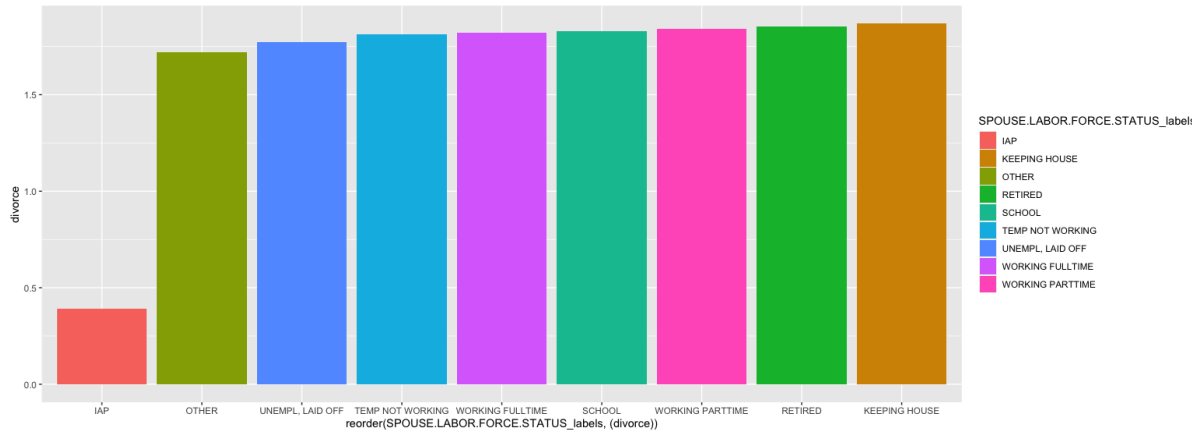

Figure 7. Mean divorce experience number group by spouse's labor force status.

The results of respondents and spouses were almost the same, and the number of divorces of spouses was not more different than that of respondents. It was judged that the relationship between labor force status and divorce could be meaningful in that labor force status, which has a high number of divorces, was similar to the respondents.

4. Suggestion

Project Idea : Identify the labor status with the highest divorce rate and analyze how it affects divorce when the spouse and respondent's labor status are same or not.

Description : After determining whether the divorce is affected by the labor force status of the respondent and the respondent's spouse, the correlation between the labor force status and the divorce is analyzed, and whether the number of divorces varies depending on the same or different cases (e.g., both working full time). It also examines which labor status combination couples have the lowest number of divorces.