Exploratory analysis

We obtained the primary dataset from the open source of the Programme for the International Assessment of Adult Competencies (PIAAC). From August 2011 to March 2012, the organization collected this survey data set from 24 countries. The individuals’ age participated in this survey range from 16 to 56. We focus on employers in South Korea whose age is from 50 to 65 based on the research purpose. Also, we averaged the variables measured by Likert scale for assessing each of the literacy usage and the numeracy usage. The new averaged Likert scale for the numeracy usage and literacy usage have values ranging from 1 to 5, including decimal numbers. We then rounded off the decimal points and aggregated the values to four integers:1,2,3 and 4. The averaged Likert scale value 5 was aggregated into 4 for the simplicity and holding model assumptions. Table 1 is the data description which has been modified for the analysis.

Table 1: Data description

|  |  |  |  |
| --- | --- | --- | --- |
| Type | Variable | Scale | Explanation |
| Demographic  Factor | AGE\_R | Continuous | Age of individual. 50 to 65. |
| Years\_wk | Number of years at work |
| GENDER\_R | Categorical | Gender. 1-male, 2-female |
| ED\_Level | Education Level. 1-middle, 2-high, 3-college, 4-graduate |
| Full\_part | Employment type. 1-fulltime, 2-part-time |
| Organizational Contexts | Work\_flexM | Continuous | Work flexibility. |
| Work\_lrnM | Learning opportunity. |
| pub\_priv | Categorical | Public vs. private sector. 1-private, 2-public |
| Learning/  Education | NFE12 | Categorical | Participation in non-formal education |
| act\_lrn | Continuous | Active learning strategies.  Averaging 6 sub-items of likert scale(1,2,3,4,5) |
| Response  variable | lit\_use\_new | Categorical  (Ordinal) | Literacy skills usage (1,2,3,4) |
| num\_use\_new | Numeracy skills usage (1,2,3,4) |

Visualization of the data is presented in Figures below. Figure 1 is the histogram of response variables. We can see the values are not evenly distributed. There are many values of 1 and the number of each value is decreasing.

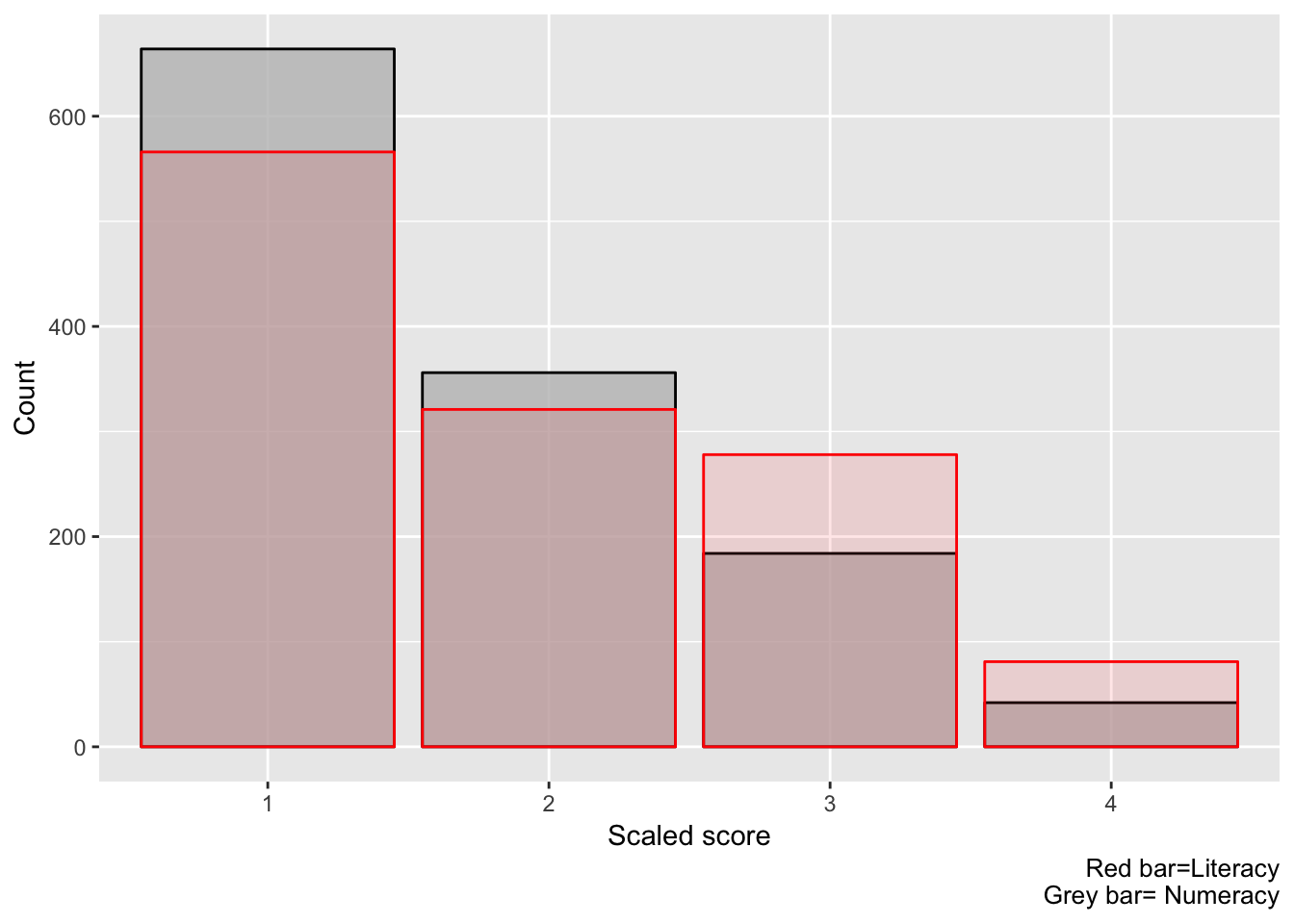


Figure 1: Histograms of response variables

All of the boxplots in Figure 2, 3, and 4 show that the three categorical variables(ED\_level, NFE12, and pub\_priv) are significant of explaining literacy skills usage variances. The other variables are not significant of literacy skills. For the numeracy skills usage, none of the variables shows significance from their boxplots.

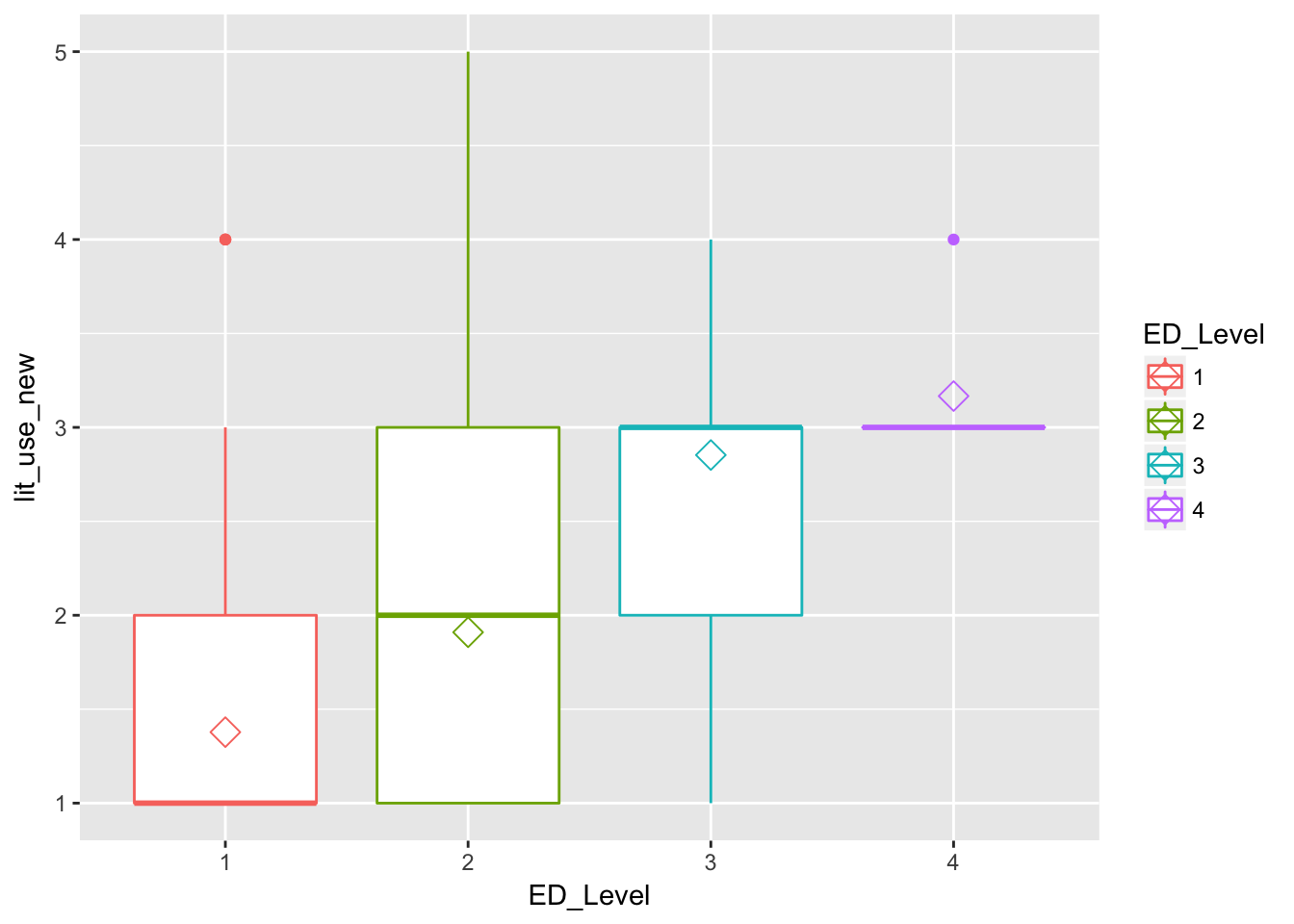


Figure 2: Boxplot of ED\_level and lit\_use\_new

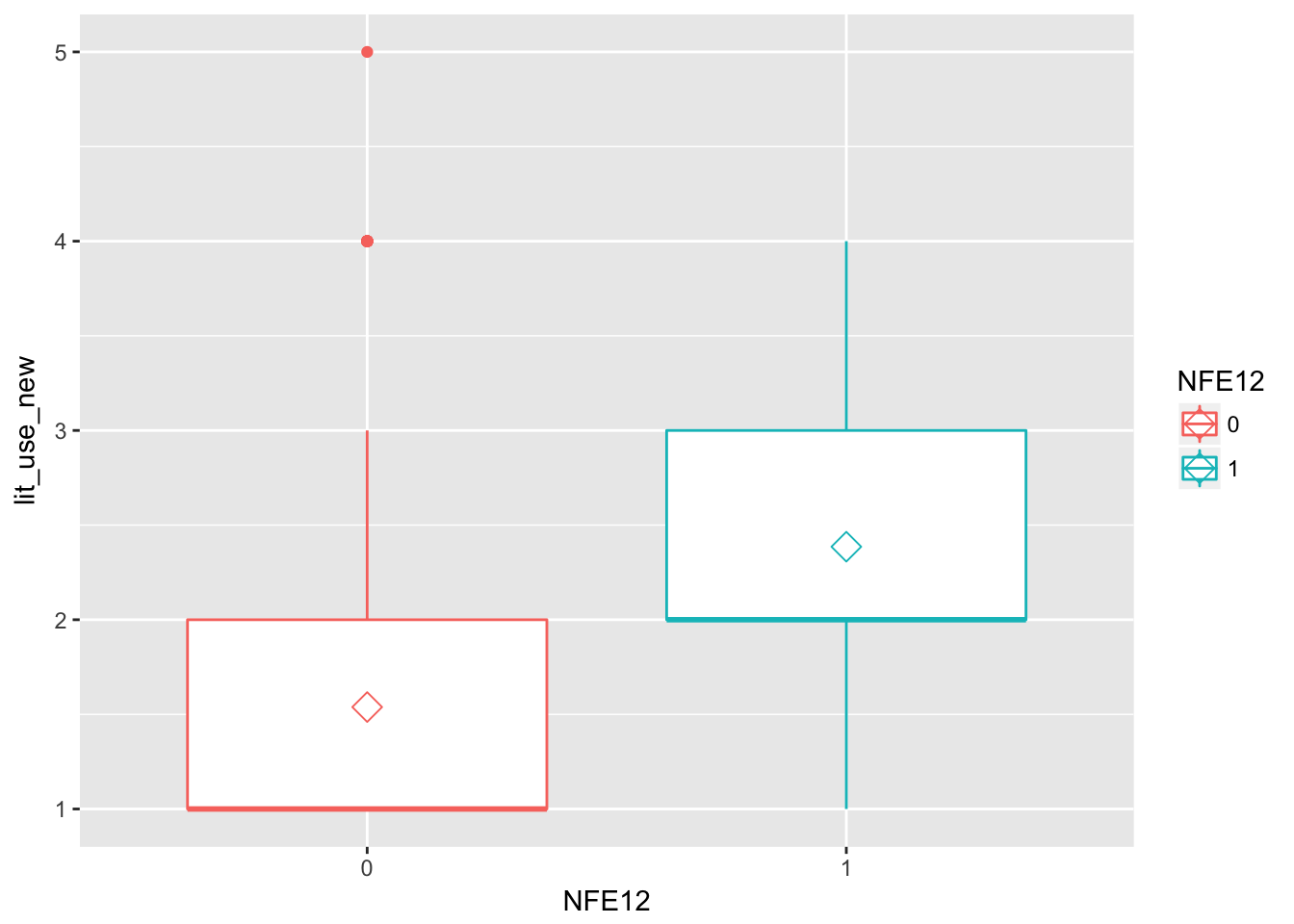


Figure 3:Boxplot of NFE12 and lit\_use\_new

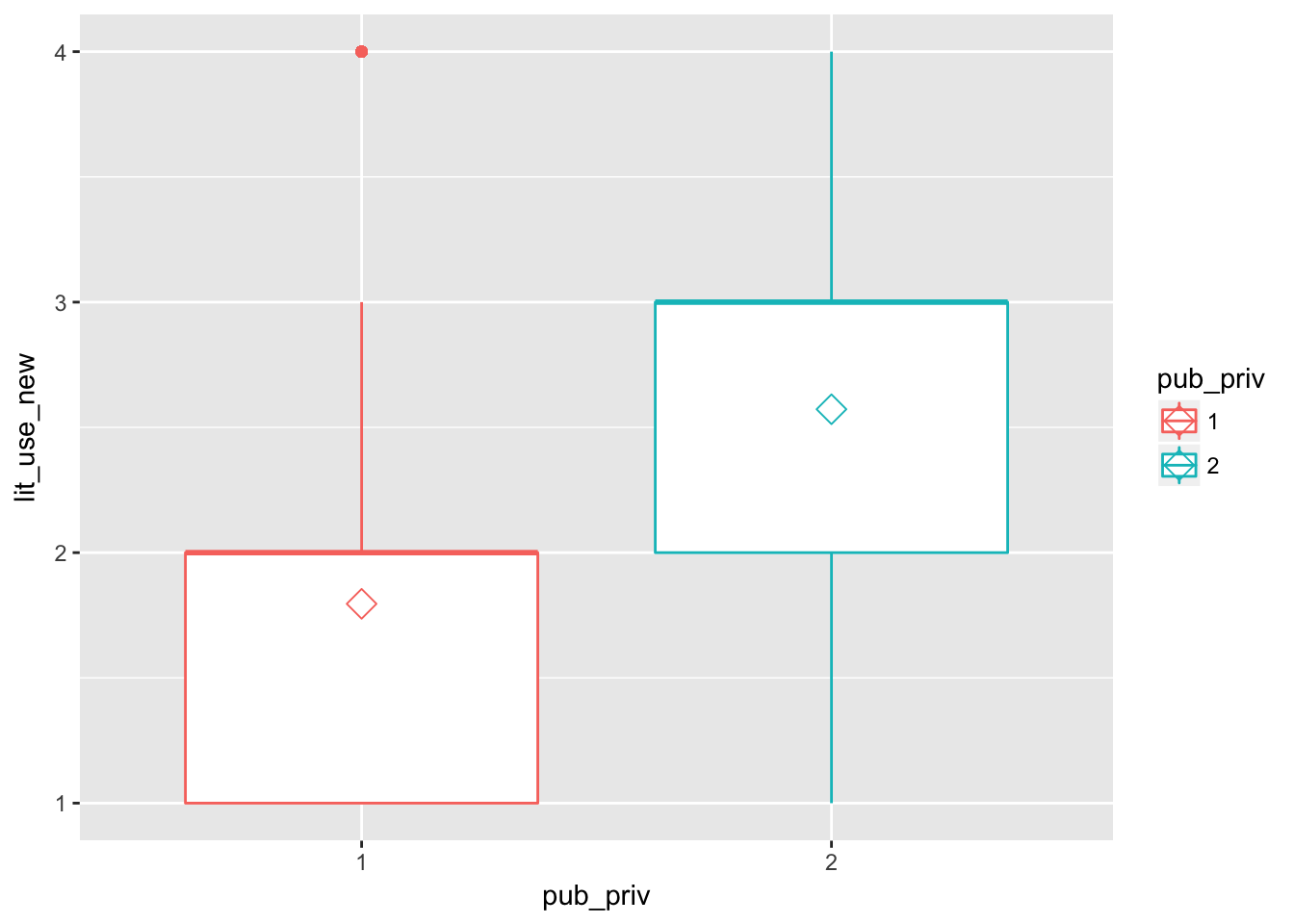


Figure 4:Boxplot of pub\_priv and lit\_use\_new