Methods in Digital Humanities Course, 2016
Final Project

Sezin Gizem Yaman
University of Helsinki
Student no: 014036523

1. Introduction

This project includes the initial analysis of a survey with 108 responses. The survey aims to investigate how software companies involve users in their software development. The survey is part of a larger study that aims to improve capabilities for user involvement in software development. The ultimate aim is to help to develop better methods for making software that is useful for users. This knowledge can also be used in the software industry and academia.

This document presents initial analysis of the survey data, which is conducted with three Finnish software companies. While Section 2 gives portrays the survey and the data collection method, Section 3 elaborates on data preprocessing and analysis. Section 4 provides several visualizations on the results and Section 5 presents the future work with the final remarks on the initial results.

2. The survey and Data Collection

The survey was designed and conducted in autumn, 2016 with three different software companies and the aggregated responses are initially analyzed in this report.

The survey included two parts as: 1) background, 2) current state with user involvement. In the background part, demographics of the respondents were collected, whereas part 2 aims at determining the current state with user involvement in the companies. Background questions included the current roles of the respondents (e.g., developing software, management etc.), their age and the duration of how long they have been working in the company. Part 2 included questions in order to evaluate in which stages of software development the users are involved (e.g., specifying requirements gathering, testing etc.) and general information about accessing and communicating users (e.g., direct communication, up to date and enough information etc.). (The survey questions can be seen in the Survey_Questions.pdf document.)

The survey was provided to the companies through a single link that was generated by one of the survey tools owned by University of Helsinki.

3. Data preprocessing and Analysis

The data was collected and the aggregated data from the three companies was extracted from the tool as a one csv file. R language with RStudio was used in the analysis.

```
data <- read.csv("~/Desktop/Methods Project/raportti.csv")
```

Preprocessing stage included reading the file and creating necessary data structures to store the data. For instance, as can be seen in the code snippet below, the responds related to job functions were stored in a factor, as it suits well with nominal values as a vector of integers.

```
data$jobfunction<- factor(data$X1.1.Which.of.the.following.most.closely.matches.your.primary.job.function..,
                          levels = c(0:6),
                          labels = c("Developing software", "Testing software", "UX Design", "Management", "Operations", "Architecture", "Other"))
```

After the preprocessing stage, the responds were analyzed with descriptive statistics methods using functions such as *summary(), table(), crosstable(), cor()* and they were plotted as can be seen in Section 4.

4. Visualization

In this section, a number of visualizations extracted from the data responses through RStudio are presented. As the demographics portray (see Figure a), majority of the respondents are software developers (36 respondents), followed by managers (24) and people from operations (14) out of total 108. The results show that the *activities performed after a software release* is the most common stage indicated by the respondents where the users are involved. *Specifying the requirements* is second most common activity where the users are involved, and it is followed by *implementing the software*.

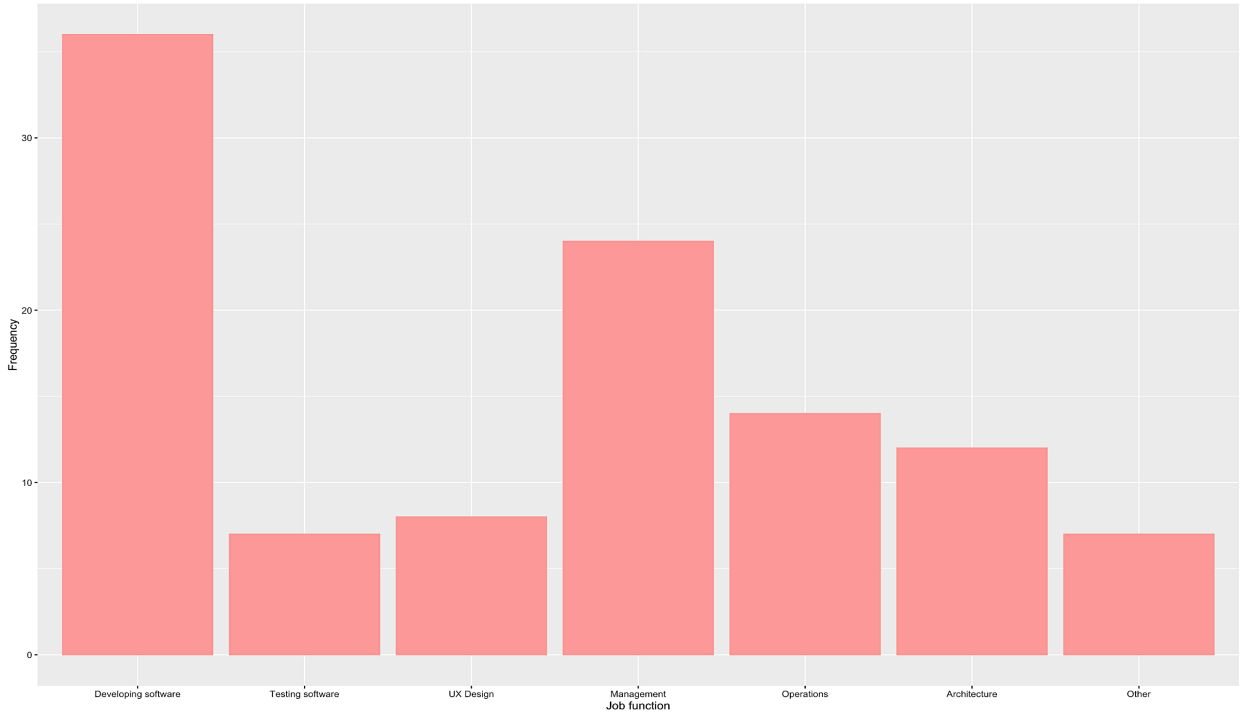

Figure a – Job functions frequency

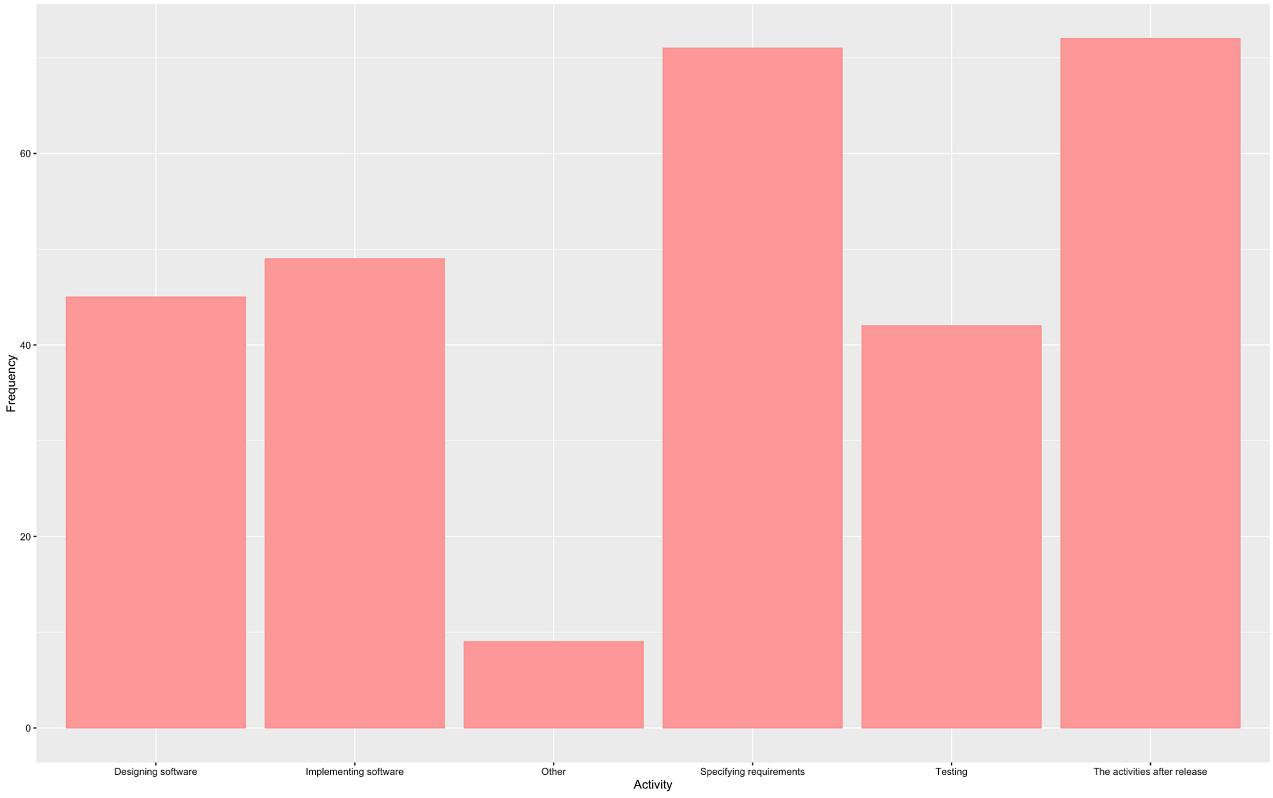

Figure b. Software activity frequencies where the users are involved

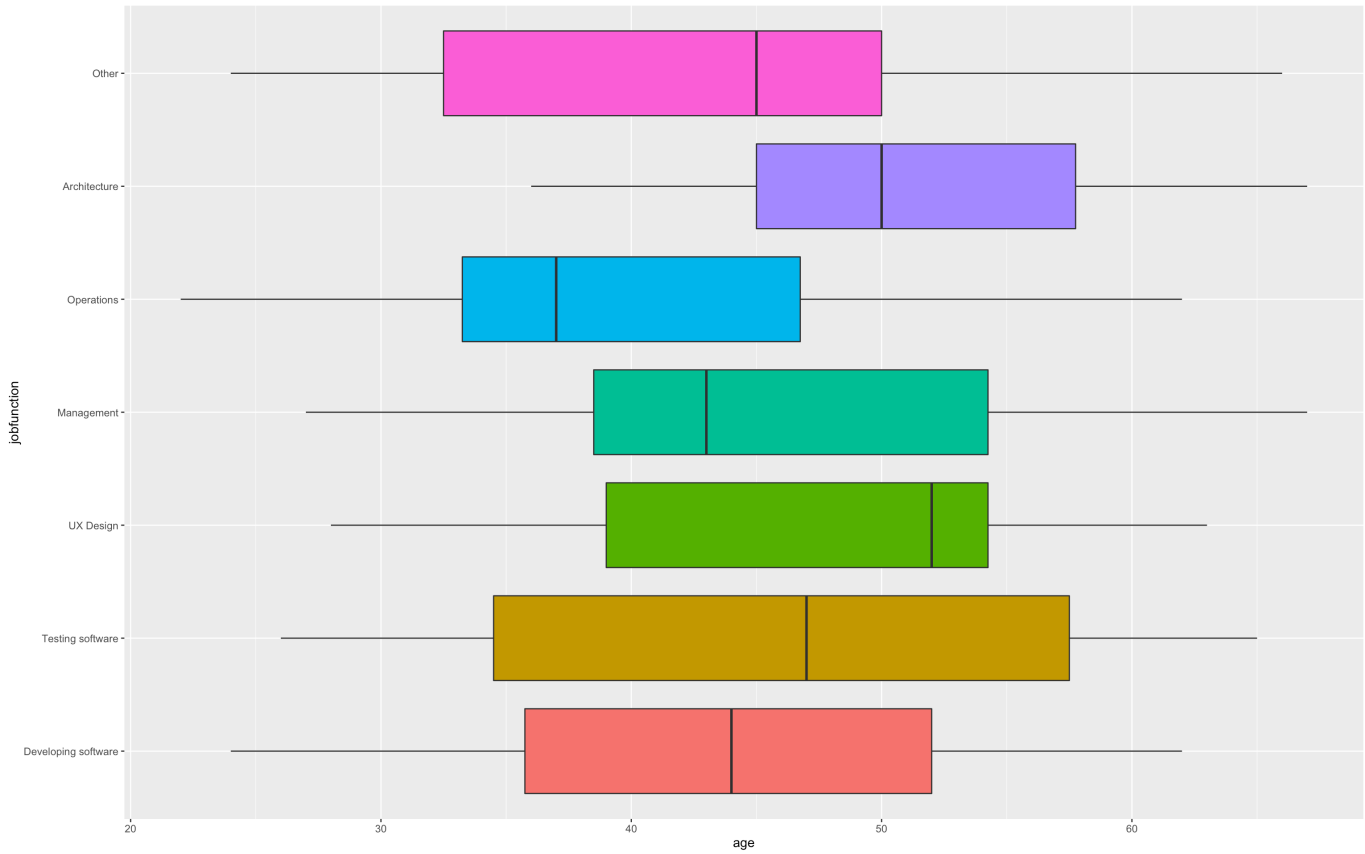

Figure c. People from different job functions vs. their age distribution

As one interesting cross-analysis example of the different roles over the companies versus people's age, we see above in the Figure c that people who work as operators have are generally the youngest judging by the median. The average age for the managements in around 45, where as UX designers seem to have a median of over 50 years old.

Figure d below shows that the respondents are mostly agreed with the statement *"I know who uses the software I contribute to in my work"*. The least agreed option is *"I need to ask permission to contact users."*. When we further look in the distribution of different job functions that requires people to ask permission to contact users, we see that (Figure e) testers are remarkably the most common ones.

Lastly, another interesting plot (Figure f) shows that the respondents who thought that they have "frequently direct contact with users" are mostly people from software architecture and UX design jobs. Software developers and testers are seemed to disagree with the statement that they have direct contact with the users.

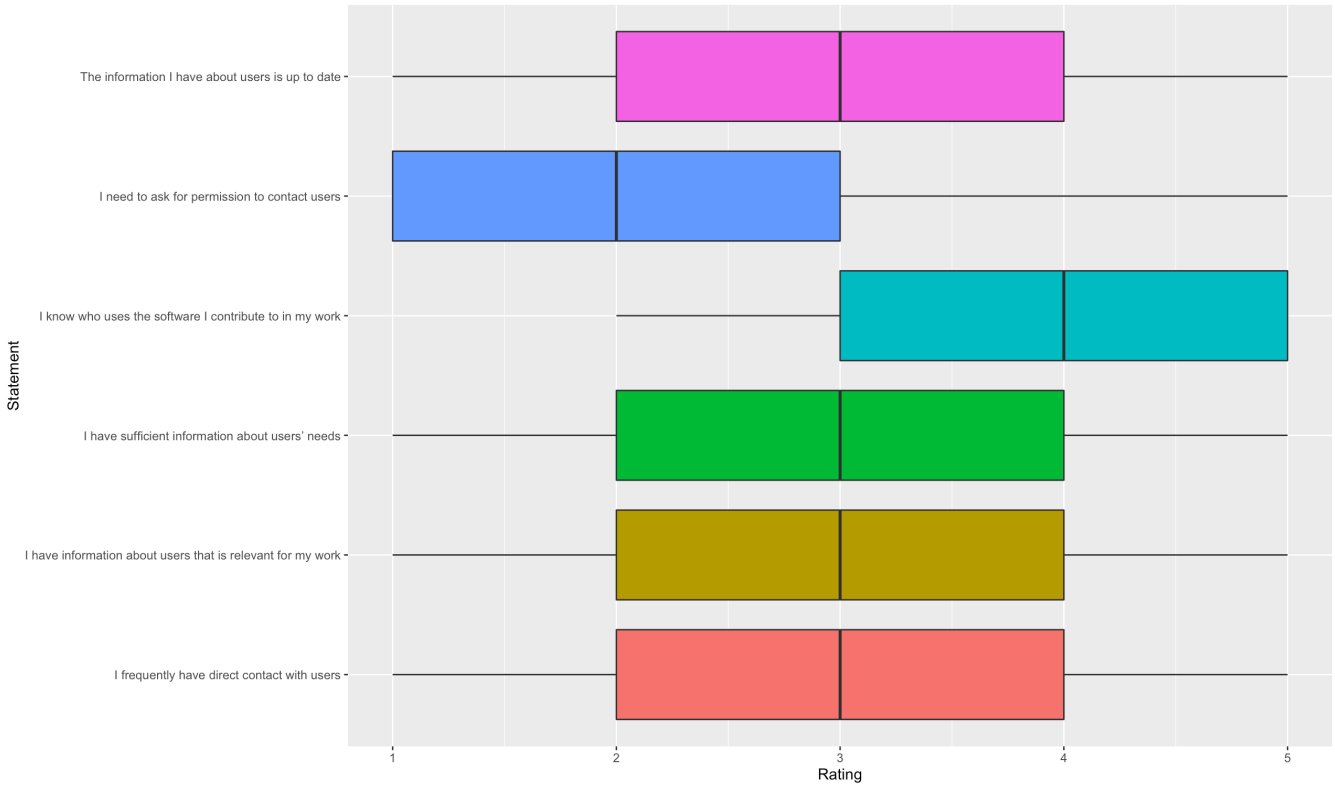

Figure d. Statements about general user involvement in the companies vs. the ratings. (Note: 1 is completely disagree, 5 in completely agree)

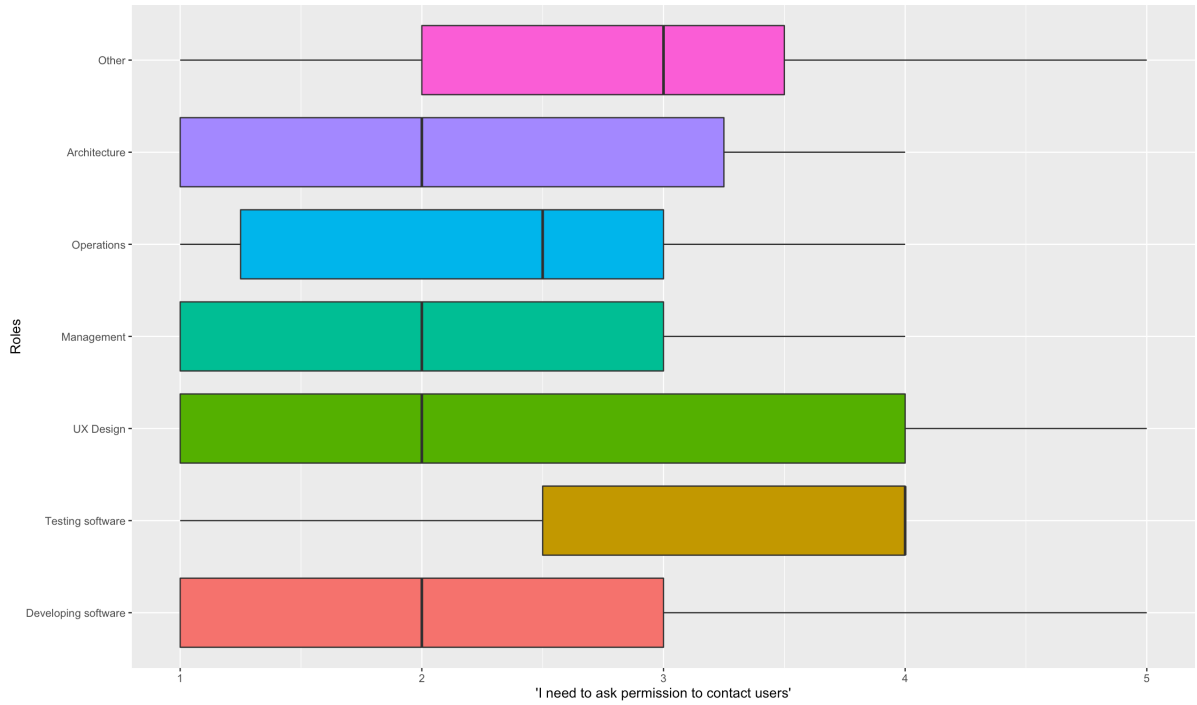

Figure e. The job function distribution of the people who ranked for "I need to ask permission to contact users" (Note: 1 is completely disagree, 5 in completely agree)

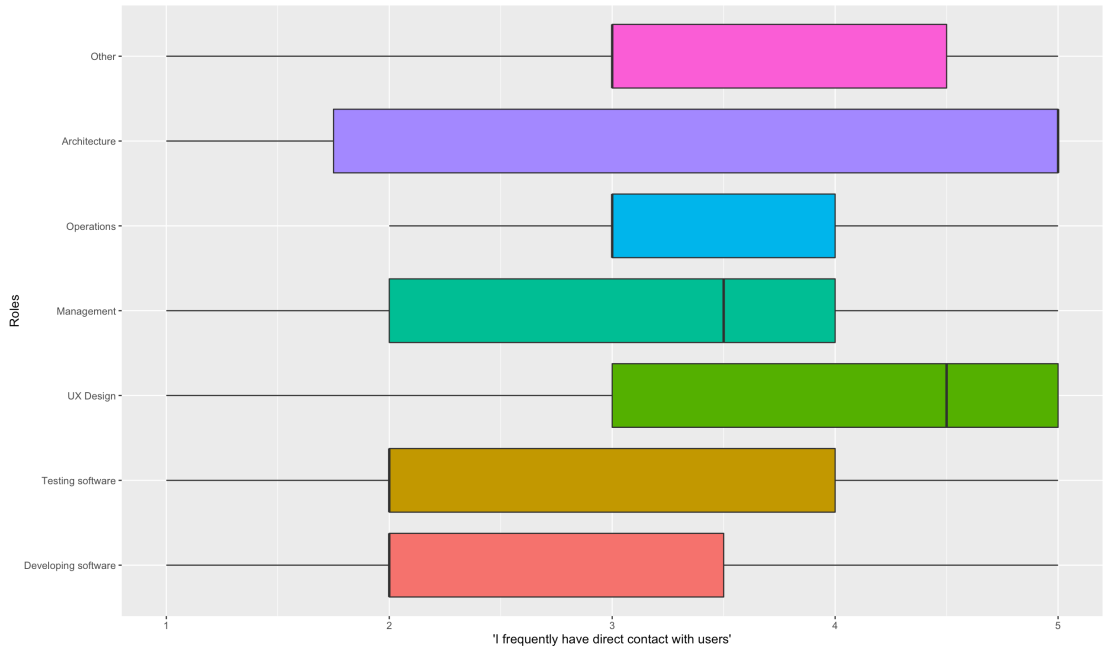

Figure f. The job function distribution of the people who ranked for "I frequently have direct contact with users" (Note: 1 is completely disagree, 5 in completely agree)

5. Final Remarks

Even though the complete analysis of the survey is not done yet, the initial results already demonstrates interesting remarks to examine more with cross analysis.

For instance, activities after release and specifying the requirements are the most common activities indicated by the respondents where the users are involved in software development. On the other hand, different roles/job functions show different distributions over the statements regarding, for instance, how often users are contacted, how much information is collected and the relevancy of the information. The people's agreement with these statements further should be cross-checked with the software development activities where the users are involved.

The information such as testers and operators who have stated that they need permission to contact users could be used to improve communication strategies within the companies. This inference already seems to be

consistent with another outcome showed by Figure f, that is testers and developers indicate that they do not have frequent access to the users. However, the correlations will be checked further for validation.

As the survey involved additional questions that were not yet analyzed yet, the initial results will also provide basis for cross-analysis of the whole survey. Besides, checking the correlations will validate the consistency of the respondents' rankings as the future research. Furthermore, at the moment the aggregated results from three companies were analyzed, however, in addition to that, responds from each company will be analyzed separately, as well. This way, the user involvement culture for different companies can be compared.