1. Introduction

- This survey was conducted to determine the extent to which the students played video games, and what aspects of video games the students find most and least fun. Video games were targeted to design labs that offer interactive learning environments, and video games provide interactive spaces for users to learn and cooperate. Out of 314 students, 95 students were selected at random to participate in the survey, and 91 participants completed their surveys. In addition, all the participants were given an exam a week prior to completing the survey. The surveys asked general questions, such as the number of hours spent on video games, how often the participants played, grades expected in the course, types of games played and why, et cetera. This paper will explore the participants' preferences through their interactions with video games and determine which information they provided will be most useful for designing interactive labs in the future.

2. Analysis

I. Scenario 2.1

Methods:

- For finding the point estimate for the fraction of students who played video games in the week prior to the survey, the dataset was filtered so that the "time" column, or the number of hours played in the week prior was greater than 0. This would provide us with the number of participants that actually did play video games. Then, this number (34 participants) were divided with the 91 selected participants for the survey, resulting in a fraction of 0.374. To find the interval estimate, the standard error needed to be calculated. However, we first needed to prove that the distribution of the sampled participants were normal, and this was achieved through the central limit theorem, which states that provided that n (91 participants) is large and n/N (91 / 314) is small, the distribution is roughly normal. The standard error was calculated using the formula given from the lectures. After the standard error was found, it was added or subtracted to the population parameter (0.374), which resulted in the confidence intervals.

Analysis:

- The point estimate resulted in a fraction of 0.374, meaning that 37.4% of the 91 participants actually played video games in the week prior to the survey. The interval estimate calculated the 95% confidence interval of the fraction of students that played video games, which were (0.289, 0.458). This means that with 95% confidence, the fraction of participants that played video games a week prior is between 28.9% and 45.8%. Although this estimate seems accurate, we constructed the interval estimate in a slightly different way, by calculating the margin of error. In *Appendix A*, the margin of error was calculated by multiplying 1.96 to the square rooted value of the point estimate multiplied by 1 - point estimate, divided by the number of participants in the survey (91). After this margin of error was calculated, it was added and subtracted to the point estimate, resulting in a confidence interval of (0.274, 0.473). Although these intervals are slightly different from the intervals calculated above, the similarities lead us to confirm that the interval estimated calculated above is an accurate measure of the fraction.

Conclusion:

- It is important to know the functions of the point and interval estimates. The point estimate allows us to find the approximate value of the given population parameter (314) from the random samples (91) of that population. It serves as a good estimate of the unknown parameter of the population. It is also important that the point estimate gives us a single estimate of the parameter. The interval estimate is important because it gives a range around an estimated measurement that conveys how precise that measurement actually is. If an estimate is calculated multiple times and is in the range of the confidence interval, it tells us how stable that estimate is. Both of the estimated functions are important in estimating the value of a given parameter in the population from a sample of that population.

II. Scenario 2.2

Methods:

- In order to check to see how the amount of time spent playing video games in the week prior to the survey compares to the reported frequency of play (daily, weekly, etc.), the “time” and “freq” columns of the dataset were extracted. Then, the rows with “freq” column value of 99 were also excluded from the dataset, because they bring no insight to the investigation. Then, the participants with “freq” column values of 1 (daily users) and 2 (weekly users) were investigated, because monthly and semesterly users will not bring significant differences in video game playing time due to an exam the week prior. After this process, we found the number of daily users (freq = 1) that played at least 7 hours of video games, and the number of weekly users (freq = 2) that played at least 1.5 hours of video games. For the daily users, only 2 out of 9 users played at least 7 hours of video games the week prior. For weekly users, only 17 out of 28 users played at least 1.5 hours of video games the week prior.

Analysis:

- The main reason why only daily and weekly video game users were included in this investigation, was due to the fact that for the monthly and semesterly users, their playtime could be altered by other confounding factors outside of the exam in the week prior to the survey. However, for the daily and weekly users, a decrease in play time may be evidence that the exam in the week prior to the survey did have an effect on the discrepancies between the reported time and actual frequency of game play. We believed that for the daily video game users, they would participate in at least one hour of gameplay for each day of that week, resulting in at least 7 hours of participation per week. By using this standard, the dataset of daily users were filtered, and only 2 out of the 9 daily users satisfied this requirement. For the weekly users, we believed that these users would participate in at least 1.5 hours of gameplay, and through this standard, only 17 out of 28 users satisfied this requirement. We believe that an existence of an exam in the week prior to the survey will have an effect on the video game playtime of the daily and weekly participants. If a participant receives a lower exam score then expected, they will try to find the factors that lead to this result of achieving that low score. This might be due to the lack of studying, illness, unexpected difficulties of the exam, et cetera. However, if this low score was due to the lack of studying, they might blame their participation in video games, and as a result decrease their playing time. This might explain the discrepancies shown in the statistics provided above. Shown in *Appendix B*, the possible effects of an exam the week prior is shown in full effect. The average time spent on video games for that week was only 4.4 hours for the 9 participants, showing very little play time for the participants that identified themselves as daily users.

Conclusion:

- By looking at the discrepancies between time spent on playing video games and how often the participant plays video games, the possible effects of an exam the week prior the survey is shown. Although this is not a confounding factor, because the data collectors knew that the students had taken an exam before the survey was conducted, it can result in a biased survey. If this survey was taken in a period of time where the students did not take an exam, say during week 1 of an academic quarter, the discrepancies shown in the dataset may not be existent. Therefore, we cannot exclude the fact that the exam might have altered the true nature of participants playing video games. The point estimate and interval estimates found above might not be an accurate estimate if an exam did not exist before the participants took the surveys.

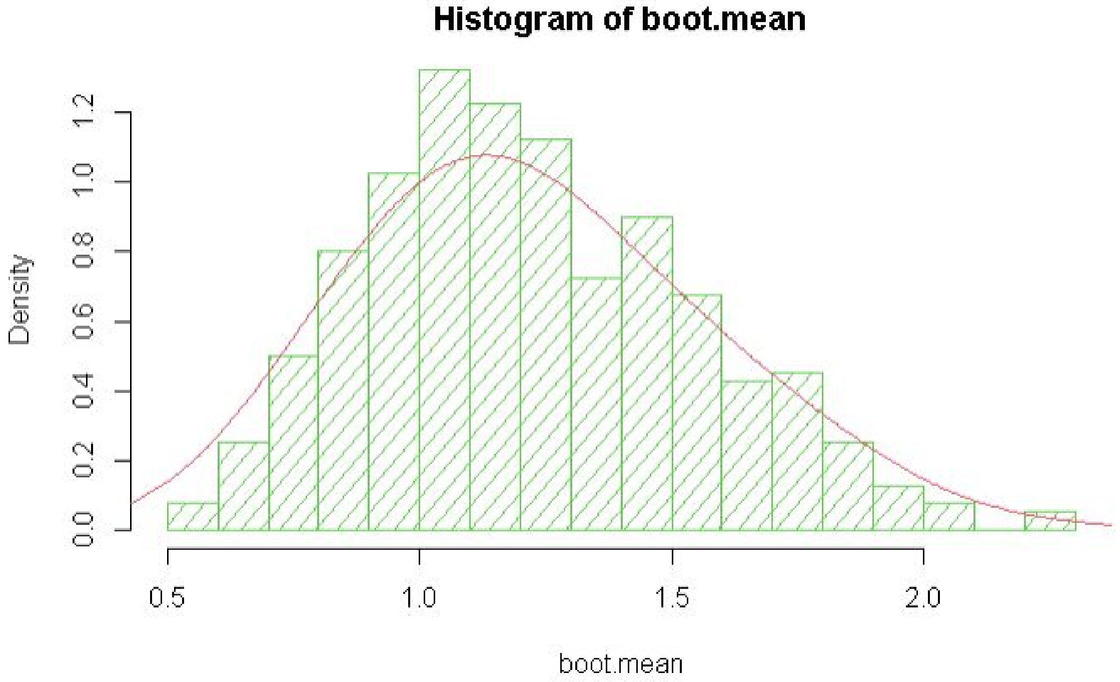
III. Scenario 2.3

Methods:

- In order to find the interval estimate for the average amount of time spent playing video games, the original survey's average amount was calculated. We calculated this by finding the mean of the "time" column in the dataset. The average amount of time spent was 1.243 hours. To find the interval estimate for average play time, we first created the sample population of the "time' column from the given sample data set. Bootstrapping was used, where 400 bootstrap samples were created from the given sample dataset, and the mean was calculated from the 400 bootstrap samples. After, the interval estimate was created by finding the standard deviation of the 400 bootstrap samples created, and this standard deviation value was multiplied by 1.96 then subtracted and added to the mean value of the 400 bootstrap samples, and the interval estimates were calculated. The interval estimates were (0.58, 1.85), and the original average amount of time spent was inside the 95% confidence interval.

Analysis:

- First, we needed to make sure that the dataset and sample size of 91 participants was enough to clearly determine if the distribution is normal or not. To test this, the sample dataset was used to create a bootstrap population of 314, which was the original population used to choose the 91 participants. We then simulated the bootstrap samples 400 times, and found the sample mean to find the interval estimates. A histogram was drawn from the 400 bootstrap samples, where the values are the average playtime for each sample.



- The histogram above shows that the simulated bootstrap average playtime shows a normal distribution. Because we know that the distribution is normal, we found the confidence intervals using the standard deviation of the 400 samples, and found that the 95% confidence interval values are (0.58, 1.85). This means that with 95% confidence, the average playtime of the sampled participants a week prior to the survey is between 0.58 hours and 1.85 hours. The actual average playtime calculated from the survey is above 1.243 hours, which is a value between the intervals.

Conclusion:

- The bell shaped curve of the normal distribution and the bootstrapping estimates show that the interval estimate is an appropriate range. Through simulating the bootstrap samples from the bootstrap population created, we were able to find the average playtime of the 400 simulations, and used these averages to find the mean playtime of the bootstrap simulations. Because the actual average playtime calculated is well between the calculated interval estimate, we can assume that the estimate is a reliable source.

IV. Scenario 2.4

Methods:

- From the dataset, we first selected the columns that can determine whether a student likes or dislikes video games. We figured that the “grade” column was the most appropriate because they are common factors of the students’ life that are usually prioritized (or supposed to be prioritized). So if the student’s response was “like” or “dislike” (possibly at its extreme level of 2 or 4) regardless of the value of “grade”, it would clearly display the attitude the student obtains toward video games. We also grouped data with respect to the “busy” column as well since, just like “grade”, it evaluates the student’s level of fondness or attachment for video games regardless of time or other values, priorities in his or her life. But first, we filtered out the rows where the values for column “like" were 1 or 5; that is, students who had never played video games before or do not like video games at all skipped attitude questions asking whether or not he or she likes video games. We also filtered out the unique value 99 in the “like” column. Then, we divided the data into students with an expected grade of C or lower and students with an expected grade of B or higher. The mean value of “like” would then be calculated. Same process repeats for the “busy” column (students who play even if busy, who don’t play if busy).

Analysis:

- After dividing the dataset with respect to “grade”, the mean of the “like” column was 2.75(like very much) when the expected grade of students was C or lower and 4.20(not really) when the expected grade of students was B or higher. We believed that this was simply because those who care and study hard for their grades tend to not like or refrain from playing video games, while those who don’t study hard obviously like playing video games. For the “busy” column, the mean was 2.98 when its value was 1 (play even if busy) and 2.47 when its value was 0 (don’t play if busy). Although the mean was only slightly higher when the value was 1, the students who didn’t play when they were busy generally liked playing video games, which was a little surprising. We thought that these students are the ones who don’t let video games bother them from doing other things in their life, although they like playing video games (not quite at the addiction stage). On the other hand, students who would play video games even if they are busy undoubtedly liked playing video games (2.98 -> somewhat).

Conclusion:

- By looking at the mean values of the “like” column (like to play video games) dependent on the values of “grade” and “busy” , the attitude question: “does a student like or dislike video games?” was answered. Some results were conventional or “as expected” (students with a lower expected grade like playing video games, students with a higher expected grade dislike playing video games.), while others were a little puzzling (students who don’t play games when they are busy like video games slightly more than students who play).

V. Scenario 2.5

Methods:

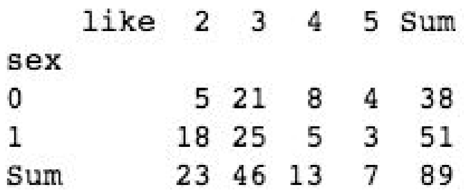
- To know each group’s differences of likes, We used cross-tabulation and histogram.

  We created the cross-tabulation, containing counts of likes, by using *Appendix C* resources. Thus, we are going to find the proportion of the likes to compare between two values. Also, we plot the probability density histogram to compare. To wrangle the data, we filtered out 1 and 99 since 1 is never played, and 99 is irregular data. Lastly, we categorized the work time. 0 is 0-10 hours, 1 is 11-20 hours, 2 is over 21 hours.

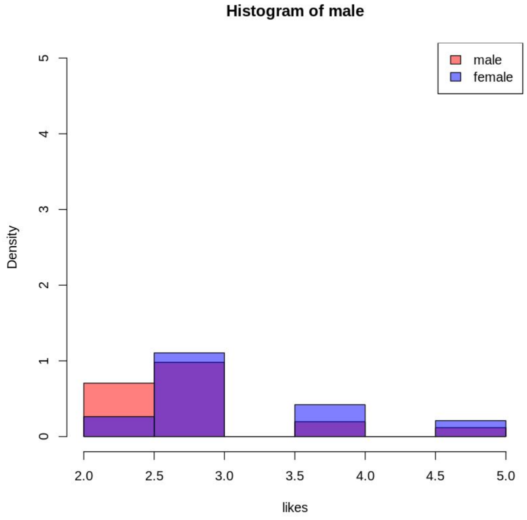
Analysis:

- **Male and Female**

  In our data, 0 represents female, and 1 represents male. As we can see the cross tabulation, male’s counts of like are larger than female. When we convert to the percentage, The probability of females who like is approximately 68% ((5+21)/3). The probability of males who like is approximately 84% (18+25)/51).When we also look at the density histogram, females are higher in the dislike part. On the other hand, males' density of probability is significantly higher in "very much". Thus, we can conclude that male more prefer to play a game than female.

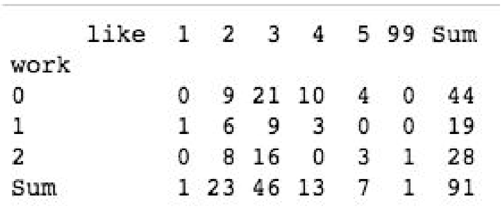


```
    like  2  3  4  5 Sum
sex
0         5 21  8  4  38
1        18 25  5  3  51
Sum      23 46 13  7  89
```

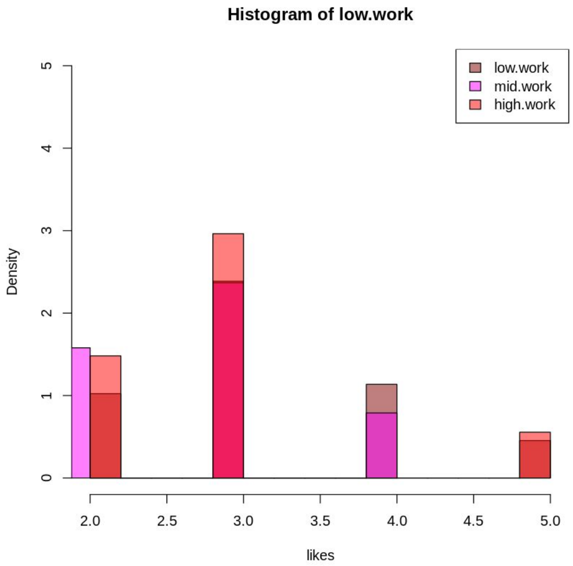


- Work hours

We divided work hours into three intervals. 0 represents 0 hours, 1 represents 0-10, 2 represents over 10 hours. The probabilities of the like people through each interval are 68%(30/44 in 0), 85%(15/19), 85%(24/28). In the probability density histogram, density of like also increases when work hour increases, and density of dislike increases when work hour decreases. Thus, we can guess what high work hours could cause stress or other factors, and make people play a game. However, very much part in 0-10 is little higher than over 10 hours. Through this fact, we also can assume that too much work hours could affect the decreasing preference of playing a game.

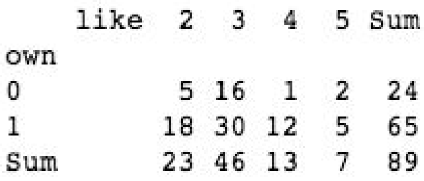


```
      like  1  2  3  4  5 99 Sum
work
0           0  9 21 10  4  0  44
1           1  6  9  3  0  0  19
2           0  8 16  0  3  1  28
Sum         1 23 46 13  7  1  91
```

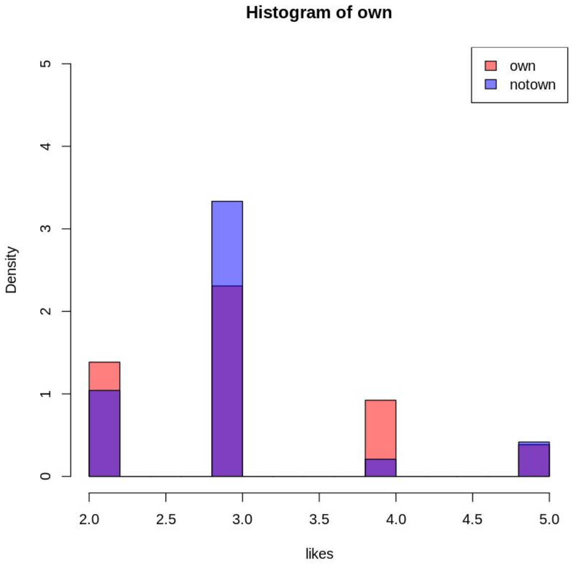


- Own computer

The counts of people who like games and own PCs are significantly higher than people who like games and don't have their own PCs. Nevertheless, the probability is the opposite. The probability of people who like games and have their own PC is 87.5. The probability of people who like games and don’t have their own PC is 73. By using this data, we could assume that PC owners prefer to play a game then non-owner. However, this data is not reliable since the number of people who do not have computers is too small. Therefore, the number of two groups is too different to compare two data.



```
    like  2  3  4  5 Sum
own
0         5 16  1  2  24
1        18 30 12  5  65
Sum      23 46 13  7  89
```



Conclusion:

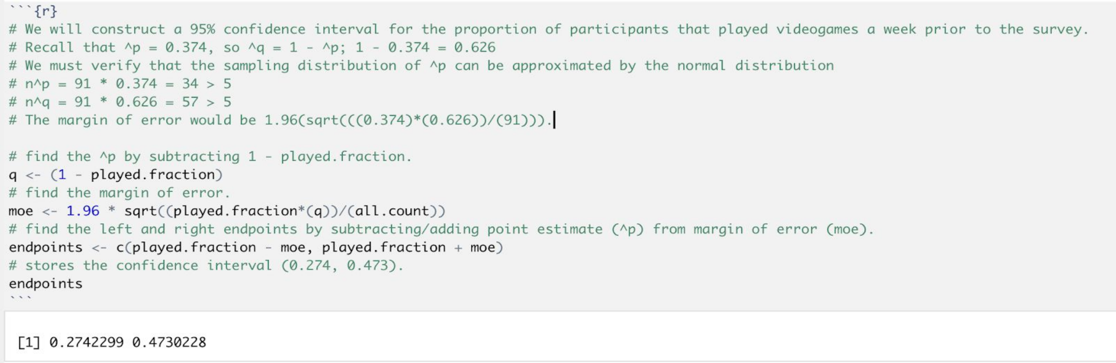
- In conclusion, there are three factors in comparing differences between groups: gender, work hour, and own PC. We discovered that male and working students prefer to play a game, based on comparing the probability of “like”. For the PC owner category, we found that students who do not have their own PC like playing video games more than students who have their own PC, but the number of data on each group was too different for a good comparison.

3. Final Conclusion

- By investigating the different scenarios, we have found useful information and statistics to give to the designers that will build new computer labs in the future. By randomly sampling participants in the advanced statistics courses, we were able to find an estimate of how many students actually participate in video game usage, how frequently the participants play and also found confounding factors such as the effects of an exam that might conflict with video game usage. We also found which demographics of the students actually enjoyed video games, and if they did, why they opted in playing video games. We used many different features to estimate these results, such as bootstrapping, cross tabulation, and data visualizations such as histograms to find different probabilities and estimates using the sampled dataset. We believe that there was lots of important information to give to the designers, such as the fact that many participants with higher grades tend to dislike video game usage, more males preferred playing video games compared to females, and the fact that as work hours increased, the preference of video game usage increased as well. This may reflect that playing video games helps the participants that worked many hours to relieve their stress through video games. However, we must recognize that the sample dataset given was very small, and there were statistics that did not logically make sense, such as participants with personal computers disliked video games compared to participants that did not have personal computers. In conclusion, we believe that the information and statistics extracted from the dataset will provide useful information for the designers.

4. Appendix

- Appendix A:



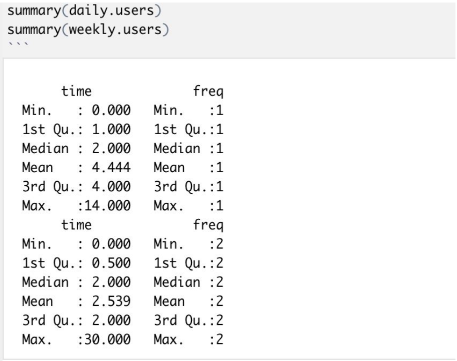
~~~
```{r}
# We will construct a 95% confidence interval for the proportion of participants that played videogames a week prior to the survey.
# Recall that ^p = 0.374, so ^q = 1 - ^p; 1 - 0.374 = 0.626
# We must verify that the sampling distribution of ^p can be approximated by the normal distribution
# n^p = 91 * 0.374 = 34 > 5
# n^q = 91 * 0.626 = 57 > 5
# The margin of error would be 1.96(sqrt(((0.374)*(0.626))/(91))).

# find the ^p by subtracting 1 - played.fraction.
q <- (1 - played.fraction)
# find the margin of error.
moe <- 1.96 * sqrt((played.fraction*(q))/(all.count))
# find the left and right endpoints by subtracting/adding point estimate (^p) from margin of error (moe).
endpoints <- c(played.fraction - moe, played.fraction + moe)
# stores the confidence interval (0.274, 0.473).
endpoints
```

[1] 0.2742299 0.4730228
~~~

- R programming language code that calculates the 95% confidence interval for the proportion of participants that played video games week prior to the survey. The margin of error was calculated by using the qˆ (0.374 from point estimate), and the pˆ (1 - point estimate). After finding the margin of error above, the endpoints (left and right) of the confidence interval were found (27.4%, 47.3%).

- Appendix B:



~~~
summary(daily.users)
summary(weekly.users)
```

      time            freq
 Min.   : 0.000   Min.   :1
 1st Qu.: 1.000   1st Qu.:1
 Median : 2.000   Median :1
 Mean   : 4.444   Mean   :1
 3rd Qu.: 4.000   3rd Qu.:1
 Max.   :14.000   Max.   :1
      time            freq
 Min.   : 0.000   Min.   :2
 1st Qu.: 0.500   1st Qu.:2
 Median : 2.000   Median :2
 Mean   : 2.539   Mean   :2
 3rd Qu.: 2.000   3rd Qu.:2
 Max.   :30.000   Max.   :2
~~~

- R programming language code that shows the summary statistics of participants that identified themselves as “daily” and “weekly” users. The mean time spent for daily users was 4.44 hours, and 2.54 hours for weekly users.

- Appendix C:

- Source ("http://pcwww.liv.ac.uk/~william/R/crosstab.r")
  imported cross-tabulations function to compare each groups

5. Author Contribution Statements

- Seokmin Hong developed the code and analytical reasoning (specifically method, analysis, conclusion in the report) for scenario 2.1~2.3. Yeongjae Kim and Youngseo Do encouraged Seokmin Hong to study the concept of "interval estimate" by referring to the slides and discussion labs provided in class and supervised the findings of his work.
- Youngseo Do developed the code and analytical reasoning (specifically method, analysis, conclusion in the report) for scenario 2.4. Yeongjae Kim helped Youngseo Do to interpret the meanings of "grade" and "busy" in the context of determining whether or not a student likes or dislikes playing video games.
- Yeongjae Kim developed the code and analytical reasoning (specifically method, analysis, conclusion in the report) for scenario 2.5. Yeongjae Kim provided visualizations such as cross tabulations and histograms to aid the analysis.
- Seokmin Hong wrote the introduction and final conclusion of the paper. Youngseo Do and Yeongjae Kim assisted by thinking of ways to write down the contextual background of this investigation and summarizing the critical findings from each scenario.
- Youngseo Do edited grammar and code (if it can be made more efficient) throughout the code part of the homework as well as the written report.
- Seokmin Hong and Yeongjae Kim added to the appendix where statistical information was useful or researched from outside sources/information.