Personal preference of Youtube content depending on one's time and place

Minseok Kim[*]
Bo-Gyu Jeong[*]
minseok@mrl.snu.ac.kr
cloudy5939@gmail.com
Seoul National University
Seoul, South Korea

Hyung-Kwon Ko[*]
hkko@hcil.snu.ac.kr
Hanyang University
Seoul, South Korea

[*]The authors contributed equally to this research.

ABSTRACT

Watching Youtube videos is one of the biggest parts of daily life for young people. We can find one's interest by tracking one's Youtube usage since their interest indirectly affects their every day thinks and behaviors like looking for the interesting Youtube video. We can also easily see a video at any time or any place using our smartphone. We tracked Youtube usage data which asks for the video information, time duration, and place. We collected 134 data from 6 participants whose ages are the 20s and work in a company or laboratory. The total length of watching YouTube is 1809 minutes. We analyzed the data using descriptive and inferential statistics. As a result, we found some relationship between video length and place using a t-test and some other qualitative factors through various graph analysis.

CCS CONCEPTS

• **Human-centered computing** → Ubiquitous and mobile computing; Ubiquitous and mobile computing systems and tools.

Human-Computer Interaction Team Project, June 21, 2019

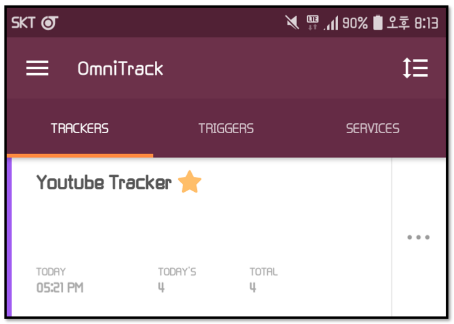

Figure 1: Youtube tracker

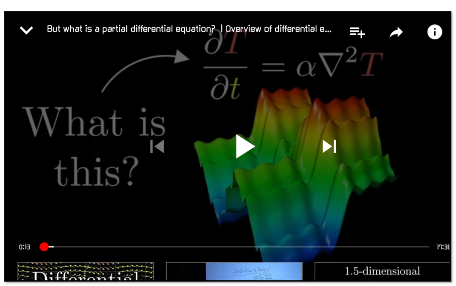

Figure 2: Captured image of a video

KEYWORDS

Self-tracking, self-monitoring, semi-automated tracking, personal informatics, tracking apps, mobile apps, personalized recommendation.

ACM Reference Format:

Minseok Kim, Bo-Gyu Jeong, and Hyung-Kwon Ko. 2019. Personal preference of Youtube content depending on one's time and place. Human-Computer Interaction Team Project, June 21, 2019. Seoul, South Korea, 12 pages.

1. INTRODUCTION AND MOTIVATION

These days, people cannot live without a cell phone, and Youtube either. In usual case, a user selects one of the contents on the recommendation list considering where she/he is, and what she/he is doing at that moment. For example, it is hard to find the one watching entertainment programs in the workplace however she/he likes it. Therefore we thought that the contents of videos would be depended on the place and the time that we are watching it. To understand the reasoning behind it on a deeper level, we set it as our research topic and try to dive into what may have caused it. We expect to find the relation between her/his Youtube content and her/his time and place of watching it. If so, we think it is possible to utilize our results in a personalized recommendation of Youtube content. By collecting and analyzing the in-situ data with the knowledge acquired during the class, we try to understand the behavior of human beings and the world around us better. Our project is done using a well-made self-tracking mobile application, Omnitrack [3], by gathering data from targeted users and analyzing them with additional software.

2. EXPERIMENT DESIGN

1) Information of participants

At first, we left the possibility of being a participant open to anyone, but soon we narrowed down the options with the constraints we set in the study. For example, to keep the validity and reliability of the study, we planned to recruit people that do not bring about any unexpected dependency between one another. Furthermore, we tried to collect an equal number of people in both sexes with diverse interests and personalities to make it general as much as possible.

2) Study period

The data collection period is set to 2 weeks considering the recommended period of research in the project documentation. We thought it is appropriate for both the researchers and the participants.

3) Tracker design & data collecting procedure

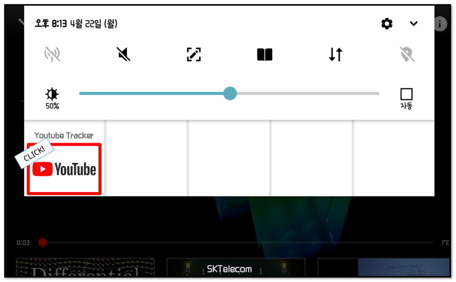

Figure 3: Cell phone shortcut

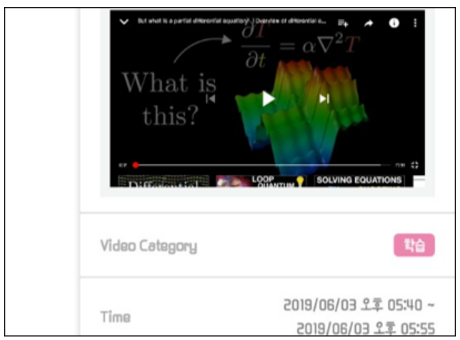

Figure 4: Documented data

Unlike our first proposal, we urged the participants to record all cases whether she/he is watching Youtube on a desktop or a cell phone. The tracking procedure followed this order:

(1) When a participant finishes watching a video, she/he can capture the video with the name of it included (Figure 2).
(2) To run the application, the participant clicks the shortcut of tracker (Figure 3).
(3) The participant fills the components and posts on the tracker (Figure 4).

In Omnitrack mobile application, We made a tracker(Figure 1) with six components.

- Video URL: URL of Youtube video (Short text, optional)
- Video Capture: Screenshot of Youtube video (Image, optional)
- Video Category: Category of Youtube video (Single choice, optional)
- Time: Time when she/he watched a video (default: current time, mendatory)
- Place: Place where she/he watched a video (Single choice, mendatory)
- Score: How much did she/he enjoy the video (5-point Likert scale, optional)

Although it is best to fill out all of the first three components about the video, the participant can choose one from Video Category if she/he does not want to give detailed information of the video. Furthermore, it is more convenient to fill in Video URL than Video Capture when one is watching Youtube on a desktop. But for most cases of watching Youtube on cell phones, Video Capture is the easiest way to go. With the video information, we only required two components that are compulsory. With this simplified process, it only took about 5 seconds for a single post when we tested. We think it is a reasonable time that does not give much pressure to the participants for posting. Researchers filled out other requirements if necessary after the experiment(e.g. Video Category in case of not written). It was okay since we tracked 2 weeks of 6 people. The data tracked are sent to the Amazon cloud server automatically.

3. RECRUITMENT

In the actual study, we recruited 6 participants in total with two main groups; graduate students(Group1) and office workers(Group2). Because we had assumed that the current status of a participant can have a large impact on her/his consuming pattern and type of videos, we evened out the number of participants in both groups.First, we investigated them as a whole, and then analyzed the effect of each independent variables with statistical inference. All of the participants are the 20s for which our project is intended. As we have a small number of participants, we thought it is more suitable to focus on the same age group for comparing one another. In the additional information tab, we put the department of a graduate student or whether a company is private or public.

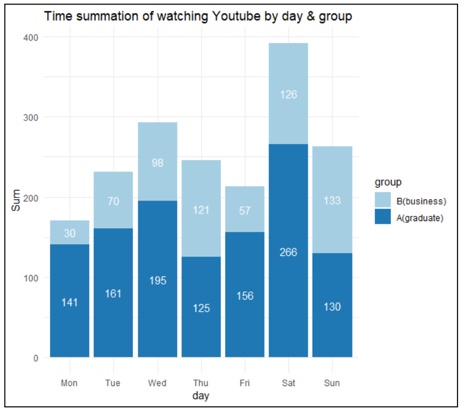

Figure 5: Time sum of watching Youtube by day & group

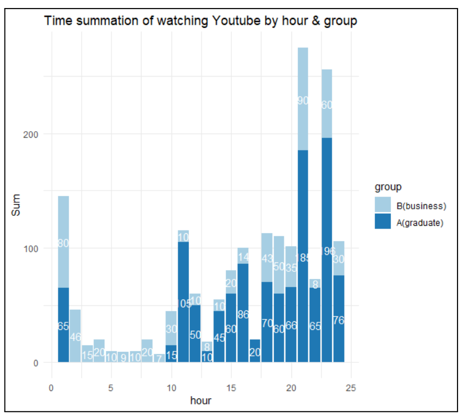

Figure 6: Time sum of watching Youtube by Time & group

	Age	Sex	Status	Additional Information
P1	27	Male	Graduate Student	Dept. of Statistics
P2	26	Male	Graduate Student	Dept. of CSE
P3	24	Male	Graduate Student	Dept. of CSE
P4	27	Male	Office Worker	Public enterprise
P5	27	Male	Office Worker	Private enterprise
P6	27	Male	Office Worker	Private enterprise

Table 1: Participants information

3. SUMMARY OF THE COLLECTED DATASET

Participant	Group	Category	Day	Length	Place	Score	Time	Work
P4	A	music	Sat	16	home	1	16	N
P1	A	movie review	Thu	10	public transportation	0.7	22	N
P6	B	sport	Thu	5	work/lab/school	0.5	11	Y
...								

Table 2: Glimpse of CSV file

We collected 134 rows of data with each of 9 columns and converted it as a CSV file (Table 2) and we used R software for statistic analysis. We separated the Time data to Day(Sun, Mon, etc.), Time(the time she/he started to watch) and Length(the duration time that she/he watched) which was a tracking component that represents start time and end time of watching. A detailed description of each column is given below.

- Participant: P1, P2, P3, P4, P5, P6 (categorical)
- Group: A: graduate student, B: office worker (categorical)
- Category: music, movie review, comedy, sport, game, study, etc (categorical)
- Day: Mon, Tue, Wed, Thu, Fri, Sat, Sun (categorical)
- Length: 1-120 (numeric)
- Place: public transportation, home, cafeteria, work/lab/school, etc (categorical)
- Score: 0-1 (numeric)
- Time: 1, 2, ..., 24 (categorical)
- Work: Y, N (categorical)

When looking at the summary statistics of two groups (Table 3), the total counts of watching Youtube videos were the same. But the sum of watching length of Group A was almost a double to Group B's. The average time of watching a single video was longer in Group A, along with the larger standard deviation.

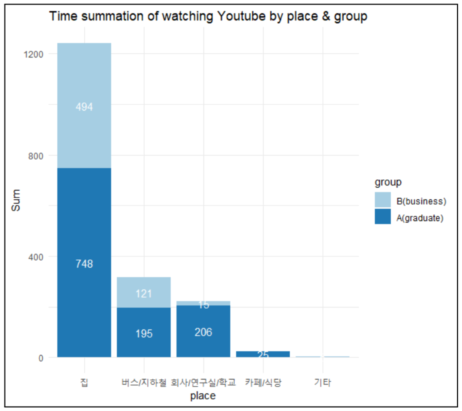

Figure 7: Time sum of watching Youtube by place & group

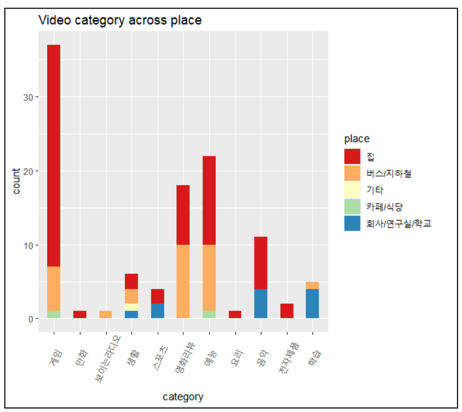

Figure 8: Video category across place

Group	Count	Sum	Mean	SD	Min	Max
Total	134	1809	13.5	17.1	1	120
A(Graduate Student)	67	1174	17.5	22.8	1	120
B(Office Worker)	67	635	9.48	6.1	3	40

Table 3: Statistics of groups

Participant	Count	Sum	Mean	SD	Min	Max
Total	134	1809	13.5	17.1	1	120
P1	34	356	10.5	6.9	1	30
P2	20	481	24.0	36.4	1	120
P3	33	278	8.42	4.5	3	20
P4	13	337	25.9	18.5	5	61
P5	18	167	9.28	1.5	5	11
P6	16	190	11.9	10.5	5	40

Table 4: Statistics of participants

To analyze individual participants, we calculated the same statistics for each of them (Table 4). We could find some interesting facts when looking at this table. First, the total count does not always proportional to the sum of watching Youtube videos. P2 has the highest sum of time but was the third with respect to the total count. Standard deviation varied from participant to participant as P2 has shown the highest as 36.4(min) while P5 has only 1.5(min) even though their total counts were almost the same(20 and 18).

Place	Count	Sum	Mean	SD	Min	Max
Total	134	1809	13.5	17.1	1	120
Home	88	1242	14.1	19.7	1	120
Public Transportation	32	316	9.88	5.00	5	25
Workplace/Lab/School	11	221	20.1	18.4	5	61
Cafeteria	2	25	12.5	10.6	5	20
etc	1	5	5	NaN	5	5

Table 5: Statistics of place

Home was the most popular place to watch Youtube videos. Its total count and summation time are even bigger than the sum of other places. The mean time of watching video was the smallest in 'Public Transportation' category when 'etc' category is not considered.

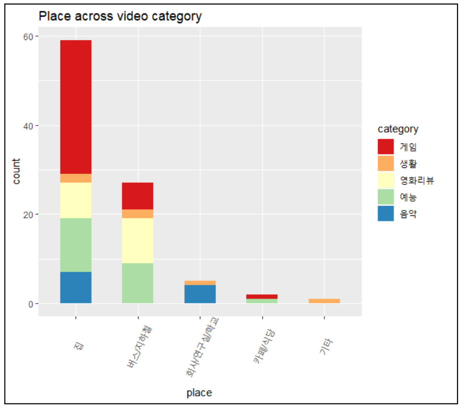

Figure 9: Place across top5 video category

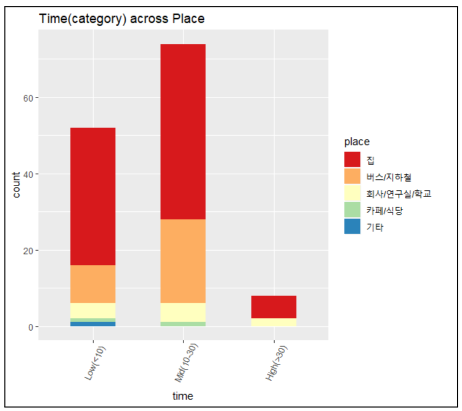

Figure 10: Time(category) across place

Among many variables, we could not extract some of the category data from our raw data because of some technical issues. We tagged them as 'x' and they were excluded in the actual analysis process. Some videos had too vague subjects to tag one category, we put them in 'etc' and not used as well. The 'Game' category was most popular in terms of the total count, but the 'Music' has the biggest sum of watching time with the highest mean time per single video(51 minutes).

Category	Count	Sum	Mean	SD	Min	Max
Total	134	1809	13.5	17.1	1	120
Game	37	411	11.1	7.14	1	35
Comedy	22	168	7.64	2.40	5	10
Movie review	18	176	9.78	5.23	5	20
Music	11	562	51.1	40.4	5	120
Living	6	35	5.83	2.04	5	10
Study	5	60	12	10.4	5	30
Sport	4	60	15	16.8	5	40
Animation	1	20	20	NaN	20	20
Radio	1	11	11	NaN	11	11
Cooking	1	5	5	NaN	5	5
Electronic devices	2	18	9	1.41	8	10
etc	7	91	13	10.6	1	30
x	19	192	10.1	6.50	3	25

Table 6: Statistics of category

Finally, we went through visualization for the dataset, delving for a valuable relationship that we thought was important to figure out. Plots enumerated on the left side per se are explaining some concise but forceful meanings.

(1) Figure 5: Saturday was the most popular day to watch a Youtube video, while Monday was the opposite, especially for the Group B. Friday was the second from the bottom for the Group B.
(2) Figure 6: We could see the high concentration on the night time (18-01).
(3) Figure 7: Regardless of the group, home was the most popular place to watch Youtube videos. The gap was almost triple even for the second popular place which is public transportation. An interesting fact is that Group A has 17.5% (206/1174) of watching time in Workplace/Lab/School category, but Group B has only 2.3% (15/635) of watching time in the same category.
(4) Figure 8: In this figure, we noticed that no one has watched a game, movie review and comedy – the top three categories in terms of its total count – in the workplace/lab/school. Fortunately,

this result was congruous to what we expected as the research had begun. We colored the category 'Home' as red and the category 'workplace/lab/school' as blue to see the difference more clearly.

(5) Figure 9: Subsequent to the previous figure, we could find an obvious distinction between place to place.

(6) Figure 10: We divided the watching time into three categories — Low, Mid, High — where Low is less than 10 minutes, Mid is between 10 and 30, and High is more than 30 minutes for a single video. Then, we draw a bar plot of time across watching place to see their relationship. As we expected, we could check public transportation does not have any count for videos in the High(>30) category.

4. PROCESS OF ANALYSIS

1) Research questions and associated hypothesis

Before entering into the main analytics, we set the research question and associated hypothesis of our study:

RQ1. Does one's place of watching Youtube video cause any difference in the personal preference of content?
- H1.1 Both groups will not watch videos at work/laboratory if the content is not work/study-related.
- H1.2 The watching length of video will be different when watched in public transportation and at home. (It will be shorter in public transportation than at home)

RQ2. Does one's time of watching Youtube video cause any difference in the personal consumption pattern of content?
- H2.1 Graduate students will watch more videos at night(after 8PMs) than office workers.
- H2.2 On weekends, both groups will have flatter distribution of watching videos by Time than the weekdays.

As we have suggested at first, we wanted to know the correlation between one's Youtube content and her/his time and place of watching it. RQ1 suggests that the place matters when choosing a Youtube content, H1.1 and H1.2 are made to backup whether our topic is guaranteed properly. In RQ2, we expanded the definition of time to day of the week, to capture the difference between weekends and weekdays. Since both groups don't go to the workplace/laboratory on weekends, we set the hypothesis as H2.2 In the usual case, graduate students have more flexible working time which helped us to set H2.1.

2) Defining variables

As a next step, we set the independent, dependent, control, and random variables of our research to investigate the required hypothesis. We listed a number of controlled variables, but most of them were kept among participants because we could not manipulate their workplace/school and home as well as their daily routine. Since all of them had no car, only no self-driving was abided by among participants.

H1.1 Both groups will not watch videos at work/laboratory if the content is not work/study-related.
- Independent variables: group
- Dependent variables: video contents at work/laboratory

H1.2 The watching length of video will be different when watched in public transportation and at home.
- Independent variables: place
- Dependent variables: watching length of a single video

H2.1 Graduate students will watch more videos at night(after 8PMs) than office workers.
- Independent variables: group
- Dependent variables: the sum of watching time at night

H2.2 On weekends, both groups will have flatter distribution of watching video by Time than the weekdays.
- Independent variables: day(weekday or weekend)
- Dependent variables: flatness of distribution on watching Time

Common
- Controlled variables: the amount of time using public transportation a day, number of times taking public transportation a day, forbidden self-driving to get to workplace/laboratory/school.
- Random variables: personal preference of content

3) Statistical approach

We went through both inferential and descriptive statistics to prove and analyze each hypothesis:

- Descriptive statistics(H1.1, H2.1, H2.2): For each of the participants, her/his contents is shown as graphs and tables with the time and place being considered(e.g. bar plot, table).
- Inferential statistics(H1.2): Once the assumptions like normality and the same variance are met, the T-test is done for two groups.

Figure 11: Post-observation interview

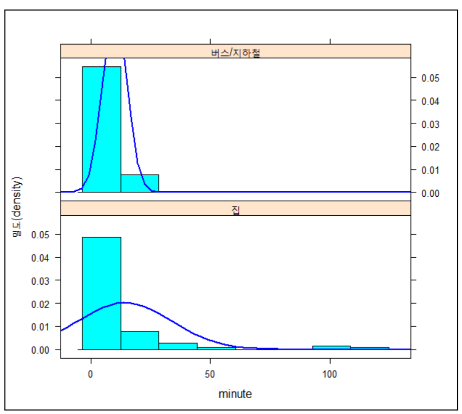

Figure 12: Normality check on 'Place'

With the help of R software, we draw a bat plot and pie chart for visual analytics of the collected dataset.

4) Additional analytics

Finally, we went through a post-observation interview(Figure 11) to understand participants better using Google Form that has five open-ended and closed-ended questions:

- How much do you care about other people when you are in public place? (5-point Likert scale)
- Under what circumstances do you find YouTube? (Long text)
- How is your daily routine of weekdays? (Long text)
- How is your daily routine of weekend? (Long text)
- How long do you take public transportation per day on average? (Long text)

As we can learn from [1], the p-value should be analyzed with the firm logic hidden underneath it, not as just the number to make a dichotomized decision on. We tried to go further by analyzing each participant with the information we got from the questionnaire.

5. RESULT

Analysis on H1.1 : Both groups will not watch videos at work/laboratory if the content is not work/study-related.

As the Figure 8 shows, there was no one watching 'Game', 'Movie' and 'Movie review' at workplace/labortory/school. According to the post-interview, two of our participants in group B(office worker) – P3 and P5 – were employed in 6 months and 8 months ago respectively. They were busy learning new tasks and being modest in the workplace, not to lose favor in senior workers' eyes. Even though our dataset is rather small, we provisionally agreed upon this result since it suits well with our common sense. In short, **We accepted H1.1.**

Analysis on H1.2 : The watching length of video will be different when watched in public transportation and at home.

To show that the mean time of watching video is not the same in public transportation and at home, we went through an unpaired two-samples t-test. We defined the null hypothesis(H_0) and the alternative hypothesis(H_1) as follow:

$$H_0 : m_P = m_H$$
$$H_1 : m_P \neq m_H$$

where 'P' represents public transportation and 'H' does home.

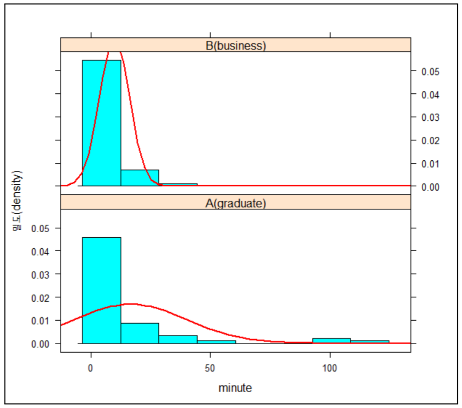

Figure 13: Normality check on 'Group'

Assumption 1: Are the two samples independent? Yes, since they are not related to one another.

Assumption 2: Are the data from each of the 2 groups follow a normal distribution? No, the p-value that we got from Shapiro-Wilk normality test which tests normality. In group P(public transportation) we got p-value 3.586e-06, and in group H(home) we got 4.103e-16 which are both too small. So we rejected the null hypothesis; the data follows normal distribution. However we moved forward since t-test is known to be robust although the samples do not follow normal distribution. The Figure 12 shows the overview of them. We believe it will be relatively solved by gathering many participants.

Assumption 3: Do the two populations have the same variances? No, the p-value we got from the F-test is 1.15e-12 which rejects the null hypothesis that both groups have the same variances. In fact, student's t-test is not robust to the two samples with different variances. However, as Welch's t-test, which is used by default in R, does not assume equal variances, we used it instead.

$$t = 1.8621,\ df = 110.84,\ p - value = 0.06524$$
$$mean\ time\ of\ Home = 14.11364$$
$$mean\ time\ of\ Public\ Transportation = 9.87500$$
$$95\ percent\ confidence\ interval = (-0.2719658, 8.7492385)$$

With the p-value from the result, we accepted the null hypothesis with the standard value 0.05, and concluded that the mean length of watching in different places are the same. That is to say, **we rejected H1.2.** Furthermore, we did the same work on two different groups, and it showed that the mean time of watching a single video is different in group A and group B. The same variances and normality check are not met in this case as well (Figure 13).

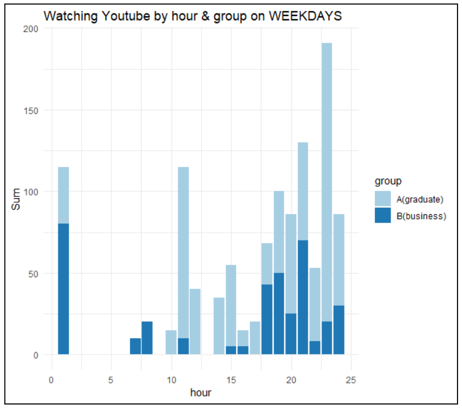

Figure 14: Watching Youtube by Time & group on weekdays

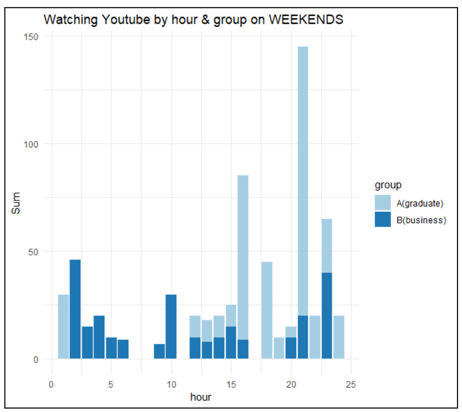

Figure 15: Watching Youtube by Time & group on weekends

Analysis on H2.1 : Graduate students will watch more videos at night(after 8PMs) than office workers.

As we have seen from the Figure 6, we could find that group B has more tendency to watch video after night, especially at midnight. Soon we realized that it is because of P3 that he was watching many videos at midnight on weekends. Except for him, the general tendency of others was like we expected. As we need more data to generalize this tendency, **we would like to defer the decision on H.2.1.**

Analysis on H2.2 : On weekends, both groups will have flatter distribution of watching the video by Time than the weekdays.

We would like to analyze this statistically but were in short of enough dataset since we have done the research only for 2 weeks in total. We had no choice but to go with descriptive statistics. When looking at the bar charts, weekends have more horizontal form except 16 to 21 pm, while on weekdays it showed more regular shape. We also can see that the watching amount of time peeks at 0 to 1 am on weekdays, but on weekends, it is evenly distributed in 0 to 6. The tendency was stronger in group B(office worker) than group A. Because we need statistical analysis, **we partially accepted H.2.2.**

Summary of the result

Since only limited hypotheses are accepted, we need more analysis as future work to make a conclusion on our research questions. But some meaningful results were found during the process, we believe in more data will make it rich and colorful.

6. DISCUSSION

We collected 134 tracking data from 6 participants for 2 weeks. Since every participant tracked at least one time for most of the 2 weeks, we can get lost of data from them. We attempted to find the influence or relation of their Youtube contents on their time and place. But it was not easy because 1) the content is mostly dependent on the person's interest, 2) there are so many kinds of content and the way of showing the content is too diverse. For example, the participant P2 mostly watched game category videos in the 2 weeks, without reference to his time or place.

One thing that we have found about content is that people in their work/laboratory do not watch a video that is not related to work/study. But this factor does not mean that people prefer other content in their work/laboratory. Instead, we attempted to find some relationship about watching time or length of the video and their time and place. We expect that people would like to watch short clip video in their public transportation while they would like to see long video at their home. But the hypothesis(H1.2) was rejected. Maybe we need more samples or need to add some more control variables to see the tendency. We found some descriptive statistics about watching time and days. We can see that people watched Youtube videos until late at night on weekends, while on weekdays they do not watch video after 1 am at all. We can easily understand this result because people have to go to work tomorrow on the weekdays.

There are so many kinds of interest and there are so many kinds of videos on Youtube. We expected to find some factors so that we can make some recommendation algorithms about the Youtube video. However, to analyze content, we may need much more data from many people to reduce the strong dependency on people and the contents they prefer. We wish we can find some more parametric factors and meaningful factors in further research which we will gather much more tracking data.

REFERENCES

[1] Amrhein, V., Greenland, S. and McShane, B. Scientists rise up against statistical significance. *Nature*, 567(7748):305-307. 2019. doi: 10.1038/d41586-019-00857-9

[2] Choe, E. K. et al. Semi-Automated Tracking: A Balanced Approach for Self-Monitoring Applications. In *IEEE Pervasive Computing*, 16(1):74-84, Jan.-Mar. 2017. doi:10.1109/MPRV.2017.18

[3] Kim, Y. H., Jeon, J. H., Lee, B., Choe, E. K., and Seo, J. OmniTrack: A Flexible Self-Tracking Approach Leveraging Semi-Automated Tracking. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3), 67. 2017.