

A Study of The Relationship Between Screen Time And Life Satisfaction of Different Age Groups In Taiwan

Fausto Urrutia (B11303091)

Chintara Sahelangi (B12303097)

Yanisa Saengcharoensuklert (B11303085)

May 25, 2024

1 Methodology

1.1 Data Collection

The data is collected from the Center for Survey Research, RCHSS, Academia Sinica and was provided to us in various data formats. The data to be used is the Digital Development Survey data which has been collected from 2002 to 2023. For the purposes of this research question we will use the data provided for 2022.

Along with the relevant data, the original questionnaire used by the research was provided to us for use as reference (Where each question is one variable in the dataset).

From the questionnaire, the questions most relevant to the research question are as follows (translated from Chinese):

- 4. How old are you? (In ranges)
- 7. How many days a week do you use the Internet? (In categories)
- 33. If 0 points means very dissatisfied and 10 points means very satisfied, how would you rate your current life?

Hereafter, we will only refer to this questions as variables in the dataset, named by our research group. The questions are renamed as variables respectively as follows: `age_range`, `internet_usage`, and `life_satisfaction`. We can extract only these columns from the dataset to use in our analysis. Using a subset of the data allows us to focus on the most relevant variables to our research question and reduce the amount of data we need to analyze.

1.2 Data Cleaning

What follows is the data cleaning process, where missing or irrelevant data is removed from the dataset. Firstly, the questionnaire provided is answered using numbers for each question (for example, (1) is Yes and (2) is No), the values in each column are all numerical in format. This means that, excluding `life_satisfaction`, we aligned each number value to its corresponding answer in the questionnaire to make the data more readable and understandable. To illustrate, this is Question 4 and 7 from the Academia Sinica questionnaire:

4. How old are you?
- (01) 12-14 years old
 - (02) 15-19 years old
 - (03) 20-29 years old
 - (04) 30-39 years old

- (05) 40-49 years old
- (06) 50-59 years old
- (07) 60-64 years old
- (08) Over 65 years old
- (98) Refuse to answer [Terminate interview]

7. [Ask those who have used the Internet in the past three months] How many days a week do you use the Internet?

- (01) Almost every day, and the Internet time is long or the frequency is high every day
- (02) Almost every day, but the Internet time or frequency is not high every day
- (03) Four to six days a week
- (04) One to three days a week
- (05) Internet access only once in more than a week
- (98) Don't know/refuse to answer

Cleaning was done by removing any rows with missing data (Don't Know/Refusal to answer) in the variables `age_range`, `internet_usage`, and `life_satisfaction`.¹

This removal and subsequent renaming was done using the Python programming language and the Pandas library for data manipulation and analysis. The original dataset contains 15,142 observations. After removing all non-answered observations we are left with 13,535 useful observations for our next step, data analysis and visualization.

1.3 Methology of the Visualization

The visualization was created using the Python programming language and the Matplotlib with Seaborn library for data visualization.

Using these tools we can group each data observation in order of age group, internet usage, and reported life satisfaction. That allows to analyze the averages of life satisfaction for each age group and internet usage category. The results from this analysis will be further disussed in our Preliminary Findings section.

For our visualization the most useful type of plot would be a boxplot. Yet, for size reasons more properly discussed in the Discussion of the Design section, we will use a multi-line chart to represent the data. In this case, each line will be individuallly made according to each age group. Internet usage will serve as a descending X-axis, while *mean* life satisfaction will be the Y-axis. Internet usage can also be renamed in-graph with their in-questionnaire name, instead of its corresponding number, from lowest to

¹This is because respondents are always given the option of refusal to answer, yet, answering *all* relevant questions is necessary for the analysis of the data.

highest usage. and life satisfaction will be the Y-axis from 0 to 10 (Lowest to highest). The resulting visualization can be seen in the Preliminary Findings section.

2 Preliminary Findings

Age and Life Satisfaction

Using the data provided and simple data summarization techniques we are able find the following preliminary findings:

- The mean life-satisfaction reported for *all* age-groups is: 7.049.

Dividing by age groups the mean life-satisfaction is as follows²:

Age Range	Mean Life Satisfaction	Generation
12-14 y/o	7.604	Generation Z
15-19 y/o	7.210	⋮
20-29 y/o	7.166	Early Gen-Z to Late Millenials
30-39 y/o	7.156	Millenials
40-49 y/o	7.125	Early-Millenials to Late-GenX
50-59 y/o	7.032	Gen X
60-69 y/o	7.155	Baby Boomers
70-79 y/o	7.263	⋮

To understand whether there is a significant difference for life satisfaction among different age groups, we will conduct a one-way ANOVA (*Analysis of Variance*) test. The test is conducted using the R programming language as follows:

```
model <- aov(life_satisfaction_0to10 ~ age_range, data = data)
summary(model)
>
>           Df Sum Sq Mean Sq  F-value Pr(>F)
> age_range    1      1  0.7955    0.272  0.602
> Residuals 11615 34015  2.9286
```

We can see from the results above that there is not sufficient evidence to suggest that there is a significant difference in life satisfaction among different age groups.

Screen Time and Life Satisfaction

Still, the motivation of this research is to find a relationship between screen time and life satisfaction among different age groups. Dividing the data into screen time usage groups, we can see the following preliminary findings:

²The *Generation* column is an addition for referencing and consistency with our literature review

Reported screen time	Mean Life Satisfaction
Everday frequent usage	7.171
Everday infrequent usage	7.137
4-6 times a week	7.015
2-3 times a week	7.202
Once a week	6.72

We can, as well, run a one-way ANOVA test to see if there is a significant difference in life satisfaction among different screen time usage groups.

```
model <- aov(life_satisfaction_0to10 ~ internet_usage, data = data)
summary(model)
```

```
>               Df Sum Sq Mean Sq F value Pr(>F)
> internet_usage    1      4   4.222   1.442   0.23
> Residuals       11615  34012   2.928
```

The results above show that there is not sufficient evidence to suggest that there is a significant difference in life satisfaction among different screen time usage groups. Even when, at a glance, the results from the least usage group (Once a week) seem to be significantly lower than the rest.

The least usage group inabilities to provide a significant difference in life satisfaction could be due to the small sample size of the group. After all, the group with the least usage of screen time is expected to be the smallest at only 54 samples. This affects its ability to significantly influence the results of the ANOVA test.

Relationship between Screen Time and Life Satisfaction per Age Group

Finally, we can look at the relationship between screen time and life satisfaction among different age groups. Given the three-dimensionality of the data we initially proposed this analysis to be conducted through visualization, as it allows for an easier approach to understanding the data.

As follows, the graph is a multi-line graph that shows the relationship between screen time and life satisfaction with a separate line for different age groups. Highlighted in brighter colors are the two youngest age groups, 12-14 and 15-19 years old, as well as our second most populous age group (40-49 years old) for a generational comparison.

The general findings through this visualization are as follows:

- There exists significant variation in life satisfaction among different age groups when they use the internet only once a week.

40 years old and above have a lower than average life satisfaction than the younger age groups at this usage level.

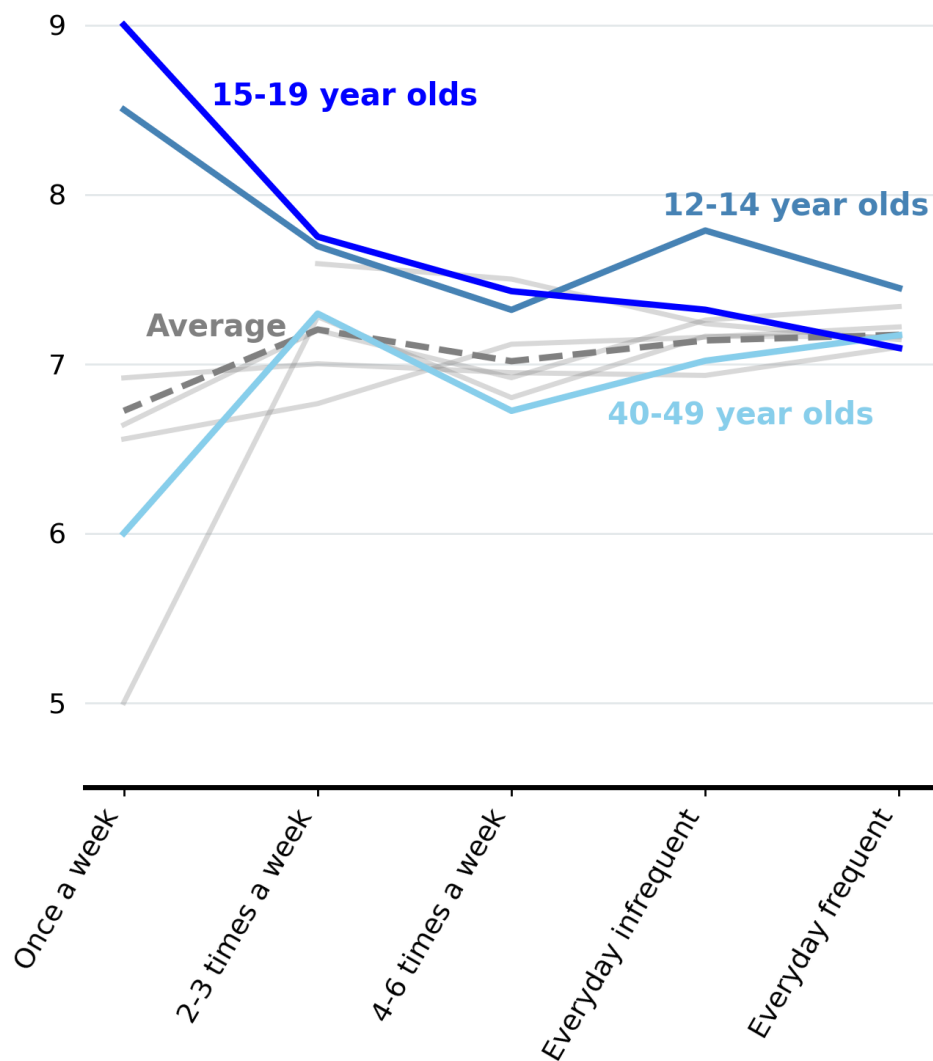
Yet, people younger than 20 years old have a higher than average life satisfaction at this usage level.

- As weekly usage increases life satisfaction decreases for those below 20 years old, but the opposite is true for those above 40 years old.

Doomscrollers

Average Reported Life Satisfaction by Weekly Usage, 2023

* Lower is worse



Source: Academia Sinica (2023)

Figure 1: Initial visualization of the relationship between screen time and life satisfaction by age group.

Hypothesis testing of effect of internet usage on life satisfaction

This paper's purpose being the effect of one variable on another, there is value in conducting a simple linear regression model to test whether there exists a significant effect of

internet usage on life satisfaction. The model is run for each age group and the p-value of the coefficient for internet usage is extracted.

$$\widehat{\text{LifeSatisfaction}}_{i, \text{age}} = \hat{\beta}_0 + \hat{\beta}_1 \times \text{Internet Usage}_{i, \text{age}}$$

```
for(age in unique(data$age_range)){
  #Filter the data in each age group
  data_filter <- data[data$age_range == age,]

  #Run a simple linear regression model
  model <- lm(life_satisfaction_0to10 ~ internet_usage, data = data_filter)
  summ <- summary(model)
  print(summ$coefficients[8])) #For p-value
```

The results of the hypothesis testing are as follows:

Selected group	P-value test result	Significant/Not Significant at $\alpha = 0.1$
For all age groups	0.229	No
12-14 y/o	0.762	No
15-19 y/o	0.113	No
20-29 y/o	0.035	Yes
30-39 y/o	0.123	No
40-49 y/o	0.097	Yes
50-59 y/o	0.041	Yes
60-69 y/o	0.016	Yes
70-79 y/o	0.501	No

3 Disussion of the Design

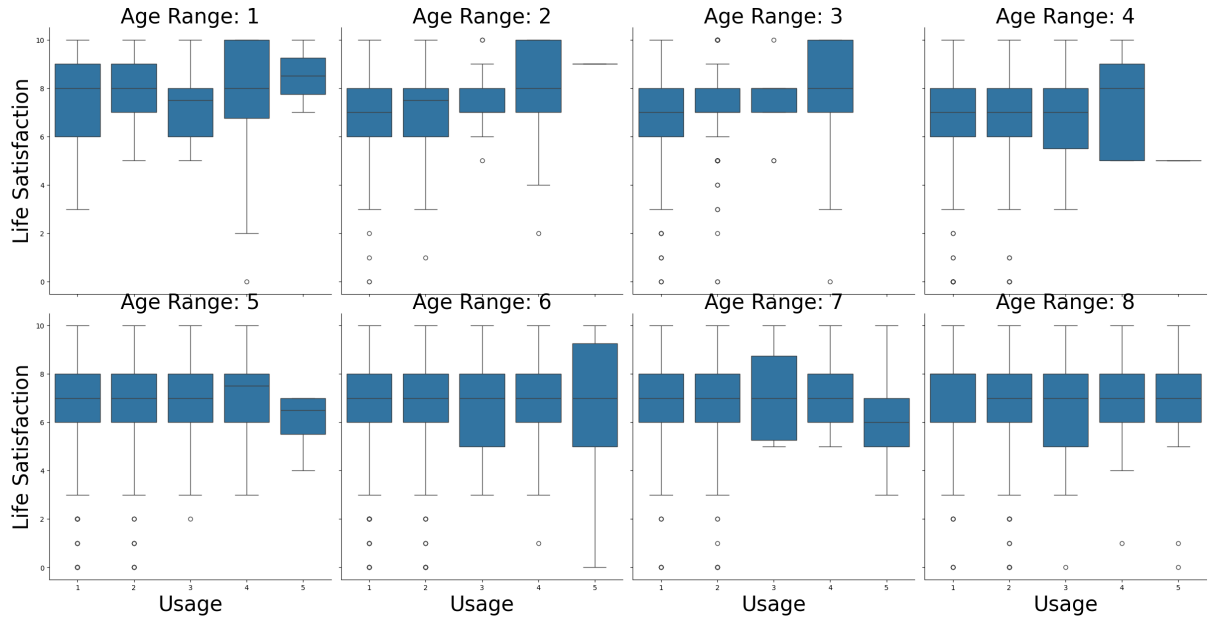


Figure 2: An initial look at the structure of the visualization.

3.1 Purpose and Objective

The purpose of this paper, and subsequently this graph, is to reveal a relationship between screen time and life satisfaction: a relationship that differs from age group to age group in Taiwan.

Hence, we aim for our graph to also enhance our understanding of the different age demographics in Taiwan and their diverging attitudes towards using electronic devices and going online. This is motivated by the common conception that the older generation has a distinctively negative attitude towards use of electronic devices as compared to the younger generation. Our hypothesis would then be that the older generation's relationship between screen time and life satisfaction would be more negatively correlated than the younger generation.

However, from the line chart, we can succinctly conclude that that is not the case. Consequently, this graph integrates seamlessly with our report narrative, going against the preconceived notion that the older generation are less happy due to increased average screen time.

3.2 Target Audience

The target audience for our visualization is the average reader with little to no specialized expertise on the paper subject and information visualization. Hence, we focus on making our graph as accessible and as unconvoluted as possible. Additionally, this also means minimizing the number of graphs to represent our paper's topic (correlation between screen time and life satisfaction) and consolidating all important information into one

graph that readers can immediately discern the conclusion meant to be understood.

Data Representation

Our graph focuses on two main variables: average weekly usage of electronic devices and reported life satisfaction. The former is reflected in the x-axis of the graph and the latter, the y-axis of the graph. We interpret average weekly usage of electronic devices as synonymous with screen time and take life satisfaction reported on the survey as a true reflection of the respondent's feelings.

We chose to represent our data visually through a modified line chart. Conventionally speaking, line charts are not the first choice when it comes to representing a relationship using a graph (this point will be elaborated on in 5.4 Visual Design Principles), however keeping in mind our target audience, we have decided to modify it to suit our data best.

To overcome the limitations of a line chart when it comes to representing a relationship between variables, the data points plotted out on the graph are the average life satisfaction (from a scale of x to x) of each weekly electronic device usage (that ranges from once a week up to every day) by different age groups. Therefore, we bypass the need for a linear regression line as we have approximated the relationship through use of the average responses and plot out an estimate of the two variables' relationship.

3.3 Visual Design Principles

As mentioned previously, the use of a line chart to show a relationship between two variables is not a conventional choice when considering it under a visual design lens. Utilizing scatter plots or bubble charts are far more common in this category while line charts are used for comparison of variables.

However, one thing we have to consider is the fact that the x-axis of this graph is not of quantifiable data, it is qualitative. It is not a continuous range of numbers but instead similar to a scale of frequency. It would not be feasible to graph such a relationship with a scatter plot. That being the case, we decided to use a line chart for our visualization, following the next objective of our graph: to compare the relationship between screen time and life satisfaction by age group.

Additionally, taking into consideration the need for clarity and interpretability, we utilized color coding to highlight the different age demographics taken into account in this research.

3.4 Comparison With Alternatives

When comparing our chosen information visualization design with alternative approaches, several key considerations emerged.

Firstly, while we successfully conveyed the relationship between screen time and life satisfaction through the line chart, there are several other charts that also could have been considered. Using a box plot or a violin plot could provide more insight into the spread of life satisfaction at every screen time category. However, when we add our second goal of comparing the relationship by age group, we would need to make small multiples of these charts which could overload the reader and complicate data analysis for the average reader. Another chart alternative would be a bar chart that utilizes the same x-and-y-axes that we used. However, it wouldn't be intuitive to draw a relationship from a bar chart, especially when we want to then compare it by age as compared to the line chart we used.

Secondly, there is an argument to be made in the usefulness of utilizing small multiples with our paper's research topic as opposed to consolidating all our findings in one graph. Nevertheless, we deem them unnecessary due to the straightforward analysis required by our target audience.

Consequently, while alternative information visualization designs were examined, the chosen graph effectively communicates the desired insights while maintaining accessibility and clarity for the reader.

4 Conclusion

Appendix

Example of Data Cleaning

Pre-cleaning Dataset

Age Range	Internet Usage Category	Life Satisfaction Level
03	01	8
05	01	7
05	98	6
04	02	7

⋮

⇓

Cleaned Dataset

Age Range	Internet Usage Category	Life Satisfaction Level
20-29 Years old	Almost every day (High frequency)	8
40-49 Years old	Almost every day (High frequency)	7
30-29 Years old	Almost every day (Low frequency)	7

⋮

Note: The red highlighted row is removed from the dataset due to missing data.