

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 24, 2024

1 Methodology

1.1 Data

1.1.1 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 and 2023. This is because the data for 2022 and 2023 is the most recent data available and will provide the most up to date information on the relationship between screen time and mental well-being among age-groups.

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). Therefore, we will first declare which questions from the questionnaire are the most relevant to our research question.

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 these 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 will allow us to focus on the most relevant variables to our research question and reduce the amount of data we need to analyze.

1.1.2 Data Cleaning

What follows is the data cleaning process, where we will remove any missing or irrelevant data 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 will need to align 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 will be 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 will be 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.

1.1.3 Data Analysis

After the data cleaning process, we will proceed to the data analysis process using the Matplotlib and Seaborn libraries in Python for data visualization. This will allow us to quickly visualize the data given its size and the number of variables we are working with.

From here we can follow to group each data observation in order of age group, internet usage, and reported life satisfaction. For this visualization the most useful type of plot will be a boxplot. More specifically, the plots will be individually made according to each age group. Internet usage will serve as a descending X-axis, while life satisfaction will be the Y-axis. For example a quick plot of the data can be seen in the following boxplot grid:

In the case of our data, there will be a total of 8 boxplots, one for each age group. Internet usage will descend from 1 to 5 (Highest to lowest), and life satisfaction will be the Y-axis from 0 to 10 (Lowest to highest).

A quick visualization of the data can be seen in the following figure:

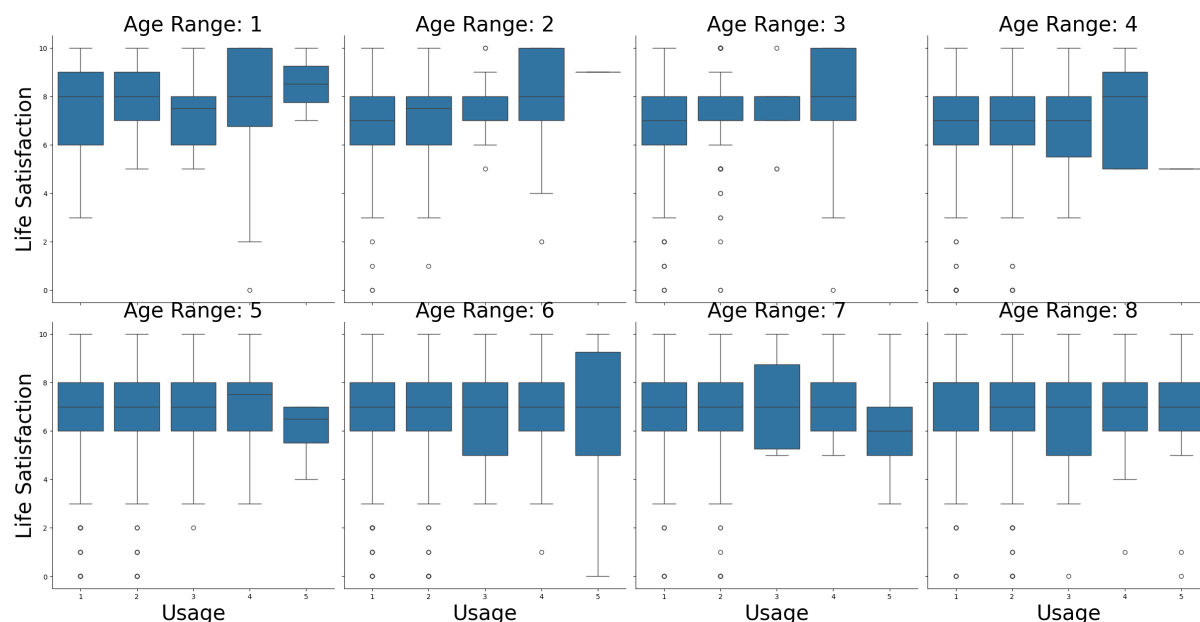


Figure 1: An initial look at the structure of the visualization.

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.

1.2 Goal of Visualization Design

As seen in the previous figure there is not much to be said about the visualization design as it is still in its early stages. The visualization must be clear, concise, and easy to understand. To achieve this, design principles such as color, size, and shape will be used to highlight the most important information. For example in the following plot, a good use of color can be seen to differentiate between the different categories

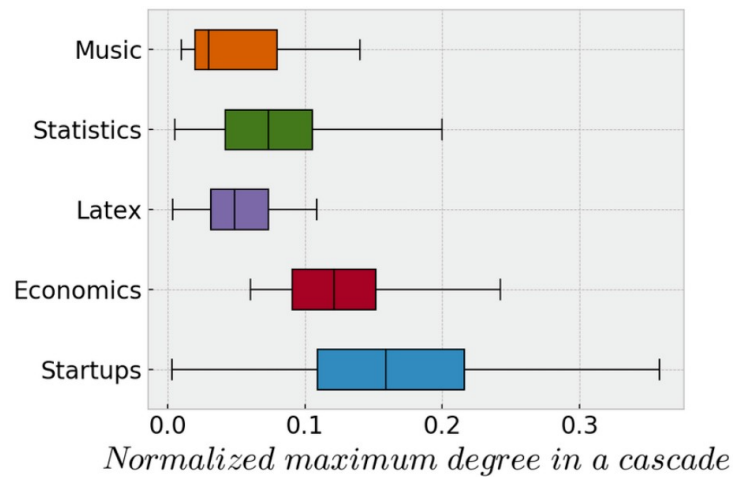


Figure 2: The use of color is of great visual aid and eases understanding of the data.

A clean and uncluttered layout is also a goal of the final visualization.

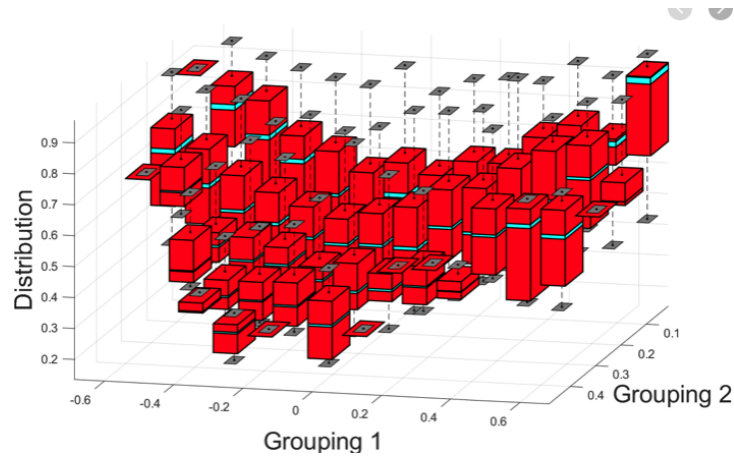


Figure 3: Designs with no clear structure or layout can be confusing and hard to read.

2 Preliminary Findings

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 follows:

| Age Range | Mean Life Satisfaction |
|-----------|------------------------|
| 12-14 y/o | 7.604 |
| 15-19 y/o | 7.210 |
| 20-29 y/o | 7.166 |
| 30-39 y/o | 7.156 |
| 40-49 y/o | 7.125 |
| 50-59 y/o | 7.032 |
| 60-69 y/o | 7.155 |
| 70-79 y/o | 7.263 |

As well, The preliminary findings show that there is a clear relationship between screen time and life satisfaction among different age groups. Yet the difference is not equal for all age-groups.