

Final Project Reflection and Data Documentation

Jiaojianjie Yu

1. Description of Data Collection and Research Purpose

In this project, I collected both objective and subjective data to analyze the relationship between my phone usage behavior and emotional state during toilet visits. The core purpose of this data collection is to explore whether phone attractiveness, emotional state, and content type influence toilet duration.

My hypothesis is: **As phone attractiveness level increases, toilet duration tends to increase. Pre-toilet emotion level and phone content type also have an influence on toilet duration.**

The objective data include:

Date

Duration of each toilet visit (minutes)

Time spent on the phone during each toilet visit (minutes)

Daily number of toilet visits

Total daily water/beverage intake (milliliters)

Daily dietary fiber intake (grams)

Phone distraction level (1–10)

Phone content type (games / social media / news / no phone use)

The subjective data include:

Subjective attractiveness level of phone content during each toilet visit (1–10)

Emotion score before each toilet visit (1–10)

In previous visualization exercises, I experimented with multiple combinations of variables. In the end, I decided to focus on four core variables: **toilet duration, phone content type, phone distraction level, and emotion score** (although phone usage duration was initially treated as one of the core variables, further exploration showed that it was highly overlapping with toilet duration; therefore, it was replaced by phone attractiveness level in the final visualization). These variables most directly reflect the relationship between behavior and psychological state and are the main focus of this project's analysis.

2. Data Collection Frequency and Assessment of Sufficiency

In terms of data collection methods, since it was difficult to rely on smart applications or automated devices for precise recording, I initially used paper forms to manually

record the start and end time of each toilet visit, emotion scores, and phone usage. After about three days of trial recording, I found obvious issues with paper-based tracking, including difficulty in merging tables, repeated information entry (such as the date), and a relatively high risk of errors.

Therefore, I switched to using a digital spreadsheet for data management, and merged and standardized daily-level data (such as daily number of toilet visits and water intake), which ultimately led to the current data structure.

The data were continuously recorded from October 5 to October 22. Except for an interruption from October 14 to 16 due to travel, the rest of the period was recorded continuously. I believe this dataset has a high degree of authenticity and naturalness, because it is derived from my everyday behavior without deliberate manipulation or intervention, and therefore reflects my actual lifestyle patterns relatively well.

However, as the amount of data increased, I realized that high-frequency daily recording had exceeded the practical needs of visualization analysis. Therefore, starting from October 23, I adjusted the data collection frequency to 2–3 times per week, and excluded data from Tuesdays and Wednesdays (since going out for classes would significantly affect the naturalness of my toilet behavior), and continued recording until December 1.

All raw data are included in the appendix at the end of this report in text-table format.

3. Explanation of Dataset Fields (Before Cleaning)

The following explains the fields in the dataset before cleaning and visualization:

Date: the date on which toilet behavior occurred

Toilet Visit Start Time / End Time: start and end time of each toilet visit

Duration (minutes): duration of each toilet visit; if less than 1 minute, it is uniformly recorded as 1 minute

Time Spent on Phone (minutes): time spent using the phone during each toilet visit

Daily Toilet Visit Times: total number of toilet visits on that day, in “counts”

Daily Water/Beverage Intake (milliliters): total daily liquid intake (including only water, coffee, tea, and juice; excluding alcohol)

Daily Dietary Fiber Intake (grams): estimated daily fiber intake based on food nutrition labels, with a certain margin of error

Phone Content Category (phone content type):

Games (mainly *Honor of Kings*)

Social media (Weibo, Xiaohongshu)

News (various news websites)

Phone Attractiveness Level (0–10):

0 indicates no phone use

10 indicates extremely strong attraction, making it very difficult to put the phone down

Emotion Score Before Toilet Visit (1–10):

1 indicates a low mood

5 indicates a neutral or stable mood

10 indicates a highly aroused emotional state (including both positive and negative high arousal)

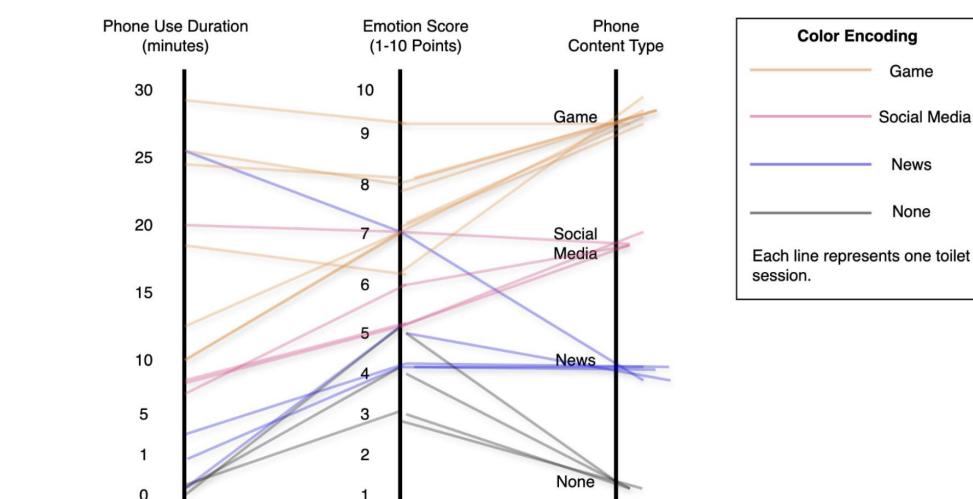
4. Alternative Visualization Attempts and Comparison

Before deciding to use a scatterplot as the final visualization form, I tried multiple alternative visualization approaches, including a parallel coordinates plot, radar chart, heatmap, and cluster bubble map. These forms are theoretically capable of representing multivariate data and thus had some exploratory value in the early stages.

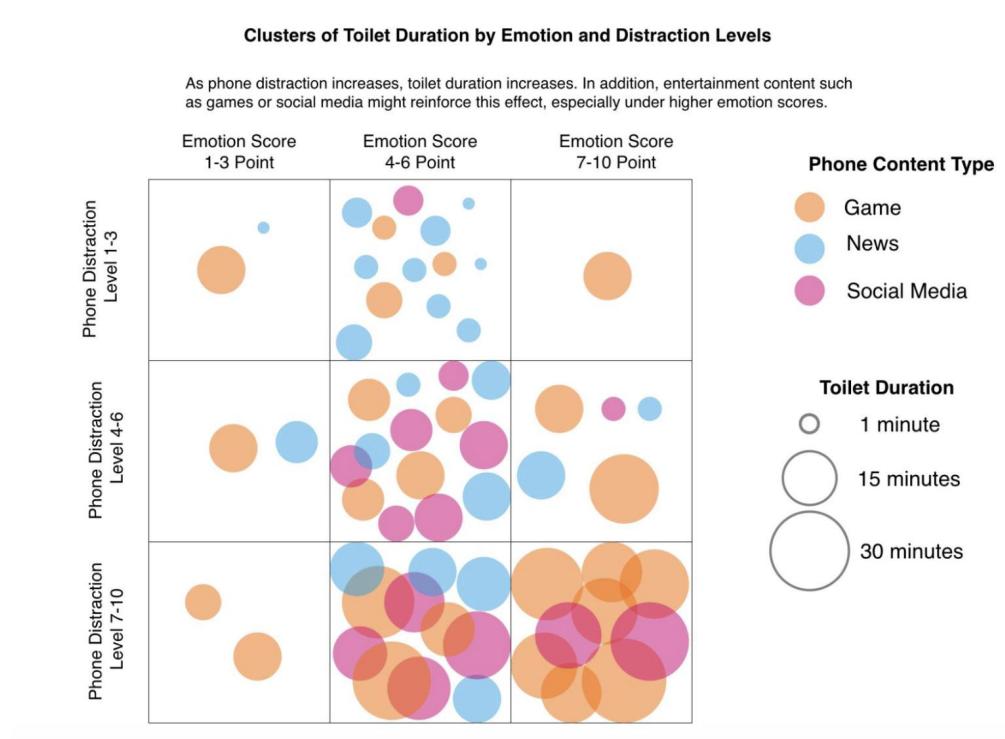
The parallel coordinates plot is methodologically capable of displaying multiple variables simultaneously and is considered suitable for high-dimensional data analysis. However, under the data conditions of this project, the relatively limited value ranges for emotion scores and phone content types led to heavy overlap of records along the axes. When multiple data points share the same emotion score and content type, their corresponding polylines completely overlap between axes and are visually misinterpreted as a single record. This severe line overlap makes individual differences difficult to identify and significantly reduces the interpretability of the chart in practice. Therefore, although this method is theoretically feasible, it did not effectively convey information with this dataset and was ultimately discarded.

Phone Use Time VS Emotion Score by Content Type

As emotion score increases, phone use duration tends to increase. In addition, higher emotion scores are more often associated with the use of game content, which typically corresponds to longer phone use than other content types



In the attempt with the cluster bubble map, I intended to distribute different types of records into different areas of the chart through sorting and spatial arrangement, thereby reducing occlusion and enhancing structural layering. However, due to strong concentration patterns across several variables, a large number of bubbles still accumulated in only a few regions, causing serious visual occlusion. In addition, without relying on AI or complex automatic layout algorithms, I was unable to implement a bubble distribution that was both visually coherent and faithful to the data. This approach thus presented considerable difficulties in both implementation and readability and was deemed infeasible.



After comprehensively evaluating multiple visualization options, I ultimately chose the scatterplot as the primary visualization of this project. This decision is based on the characteristics of the data structure, the core focus of the research question, and the comparative strengths and weaknesses of different visualization forms in conveying information.

The project data simultaneously include several continuous variables (such as toilet duration, phone attractiveness level, and emotion score) as well as categorical variables (phone content type). The research focus is not on the mean or distribution of a single variable, but on observing the relationships among multiple variables at the level of individual events, and how these variables jointly influence changes in toilet duration. Under these circumstances, the scatterplot preserves each individual record, allowing individual behavioral differences to be fully represented without losing detail through statistical aggregation.

Compared with other visualization types, the scatterplot is better suited to exploring potential correlations between continuous variables and identifying possible positive or negative trends. At the same time, through visual encodings (such as using color to distinguish phone content types and using point size or opacity to represent phone

attractiveness level), the scatterplot can integrate multiple dimensions of information in a single view without significantly reducing readability. This is particularly important for this project, because behavior and emotional state often occur simultaneously and influence each other.

Furthermore, compared with more complex multivariate visualization forms such as parallel coordinates or Cluster Bubble Maps, scatterplots impose a lighter cognitive load on the viewer. For personal behavior data with a limited sample size and overlapping value ranges, overly complex visualizations tend to produce line or shape occlusion, thereby reducing the efficiency of information perception. The scatterplot strikes a better balance between information density and readability, allowing both overall trends and outliers to be clearly observed.

In summary, in this project the scatterplot effectively represents the relationships among multiple variables, avoids detail loss caused by aggregation, and maintains a high level of readability and interpretability. Therefore, after weighing multiple feasible options, I consider the scatterplot to be the visualization form that best supports my research question and most closely matches the characteristics of the dataset.

5. Conclusions Based on the Visualization

Based on the final scatterplot, I observe a clear positive relationship between phone attractiveness level and toilet duration. As the phone content becomes more attractive to me, the time I spend on the toilet generally increases.

Emotional state also exerts a noticeable influence: higher emotion scores are more often associated with longer toilet durations. Phone content type plays a moderating role in this process—when my emotion level is relatively high, I am more likely to play games or browse social media; when my emotion level is lower, I tend to read news or engage in low-intensity social media browsing. News content is usually associated with shorter toilet durations, while gaming behavior appears more frequently in situations characterized by high phone attractiveness and long toilet durations.

Overall, when phone attractiveness is low, I am more likely to choose news or social media, with shorter toilet visits; when phone attractiveness is high, I tend to play games, accompanied by longer toilet durations and higher emotion scores. These visualization results support the hypothesis I initially proposed.

6. Self-Analysis and Reflection

This project is built around the following hypothesis: **As phone attractiveness level increases, toilet duration tends to increase. Pre-toilet emotion level and phone content type also have an influence on toilet duration.**

Learning Goals and Focus of Exploration

I hoped to explore the key factors influencing toilet duration through long-term recording of my everyday behavior. Ultimately, I focused my analysis on the

interaction between phone usage behavior and emotional state, because these variables exhibited the clearest associative patterns in the data.

Additional Findings from Unvisualized Data

By analyzing the temporal distribution of toilet visits, I discovered a clear time-series pattern: in the early part of the day, my emotion level is generally lower, I tend to read news during toilet visits, and the duration is relatively short; as the day progresses, my emotion level gradually rises, and especially at night I am more likely to play games during toilet visits, which significantly extends the time I stay in the bathroom.

Unexpected Findings

What surprised me was that my reliance on games was much stronger than I had expected. I gradually realized that I had developed a habit of playing games on the toilet, and this behavior not only affects time management but may also have negative implications for my physical health.

Changes in Personal Ways of Perceiving Behavior

Before starting this project, I did not have a systematic habit of recording personal behavioral data. Although I occasionally checked scattered usage statistics in mobile applications, these data usually lacked context and were not used for long-term analysis. Therefore, this project represents my first attempt to observe my own behavior in a continuous and structured way, and it has provided foundational experience for subsequent data analysis.

This project has changed how I observe my daily behavior. Through data-driven recording, I realized that many seemingly trivial everyday actions actually follow clear patterns. Data visualization allows me to examine my own behavior from a more rational perspective, rather than relying solely on subjective impressions.

Future Directions for Data Collection

In the future, I could attempt to introduce deliberate behavioral interventions (for example, strictly prohibiting myself from playing games on the toilet) and continue collecting data, then compare the differences before and after the behavioral change. This would help further distinguish correlation from causality and provide a deeper understanding of the mechanisms underlying the data.

In addition, the current dataset still has certain limitations, such as the lack of detailed records on the specific time of day when toilet visits occur, physical conditions, or overall stress level for that day. Incorporating these factors in future data collection would contribute to a more comprehensive understanding of the relationships among emotion, behavior, and environment, rather than remaining solely at the level of correlation.

7. Data Record

Hypothesis: As my phone use during toilet time increases, my time spent on the toilet increases:										
Objective Data									Subjective Date	
Date	Toilet Visit Start time	Toilet Visit End time	Duration(min)	Time spent on my phone during each toilet visit	Daily toilet visits times	Daily amount of water/beverage intake(ml)	Daily amount of dietary fiber intake(g)	(New)phoneContent Category (game/news/social media)	Phone Attractiveness Level during each toilet visit(1-10)	Emotion before each toilet visit(1-10)
5 Oct	09:31	09:44	13min	12min	7	2100	50	game	4	3
	12:31	12:33	2min	0min				/	0	5
	14:20	14:22	2min	2min				news	5	4
	15:35	15:50	15min	15min				game	8	8
	17:55	17:58	3min	1min				news	1	5
	19:01	19:05	4min	1min				news	1	5
	22:07	22:36	19min	18min				game	10	9
6 Oct	00:23	00:40	17min	17min	9	1300	30	social media	10	4
	02:12	02:21	9min	9min				social media	5	5
	09:59	10:01	2min	0min				/	0	5
	11:07	11:20	13min	10min				news	5	6
	14:26	14:28	1min	0min				/	0	5
	15:03	15:25	22min	22min				game	9	9
	17:55	17:56	1min	0min				/	0	5
7 Oct	20:36	20:36	1min	0min	10	2400	30	/	0	5
	23:44	23:56	12min	10min				social media	7	4
	02:07	02:08	1min	0min				/	0	2
	04:58	04:58	1min	0min				/	0	2
	10:55	11:13	18min	18min				game	10	7
	12:45	12:55	10min	10min				social media	6	6
	14:57	14:58	1min	0min				/	0	5
8 Oct	18:00	18:02	2min	0min	7	1000	70	/	0	3
	19:10	19:10	1min	0min				/	0	5
	20:01	20:02	1min	0min				/	0	5
	21:23	21:26	3min	1min				news	1	5
	22:55	22:58	3min	3min				news	1	6
	08:11	08:12	1min	1min				news	1	2
	09:06	09:07	1min	0min				/	0	3
9 Oct	12:46	12:47	1min	1min	9	2200	30	news	1	5
	13:33	13:36	3min	0min				/	0	7
	16:00	16:11	11min	10min				game	4	9
	21:16	21:35	19min	19min				social media	10	9
	23:28	23:35	7min	7min				game	6	8
	00:17	00:18	1min	1min				news	1	4
	09:22	09:23	1min	0min				/	0	5
10 Oct	09:55	10:17	23min	23min	6	1500	40	social media	10	7
	13:47	13:50	3min	3min				news	1	5
	15:55	15:56	1min	0min				/	0	4
	18:02	18:03	1min	0min				/	0	4
	19:45	19:46	1min	0min				/	0	6
	20:57	21:12	15min	15min				social media	7	6
	23:45	00:12	27min	27min				game	10	8
11 Oct	10:54	11:06	12min	12min	6	1700	50	news	8	6
	13:04	13:05	1min	0min				/	0	5
	14:29	14:37	8min	8min				game	5	5
	17:56	18:02	6min	6min				social media	5	7
	21:23	21:23	1min	0min				/	0	8
	23:45	00:08	23min	23min				game	10	8
	09:50	09:51	1min	0min				/	0	5
11 Oct	12:39	12:55	16min	16min	6	1700	50	social media	7	6
	14:10	14:13	3min	2min				news	1	6
	17:25	17:26	1min	0min				/	0	7
	20:07	20:30	23min	23min				game	10	7
	23:12	23:15	3min	2min				news	4	7

12 Oct	09:50	10:02	12min	10min	8	2100	20	news	7	5
	11:59	12:15	16min	16min				social media	9	7
	12:40	12:43	3min	0min				/	0	7
	15:33	15:33	1min	0min				/	0	6
	19:26	19:27	1min	0min				/	0	6
	20:43	20:43	1min	0min				/	0	5
	21:55	22:14	19min	19min				game	10	7
	23:48	23:52	4min	0min				/	0	5
13 Oct	02:30	02:35	5min	5min	7	1700	40	news	2	6
	11:13	11:14	1min	0min				/	0	5
	14:29	14:29	1min	0min				/	0	6
	15:55	16:01	6min	5min				social media	1	6
	19:24	19:33	9min	9min				news	5	7
	19:51	19:52	1min	0min				/	0	6
	22:47	22:50	3min	0min				/	0	5
17 Oct	08:31	08:32	1min	0min	8	2500	30	/	0	2
	10:52	10:54	2min	0min				/	0	4
	11:36	11:37	1min	0min				/	0	4
	13:03	13:10	7min	7min				social media	5	5
	16:49	17:15	26min	26min				game	10	7
	19:17	19:18	1min	0min				/	0	7
	20:05	20:16	9min	9min				social media	5	8
	22:19	22:37	18min	18min				game	8	8
18 Oct	00:41	00:42	1min	0min	8	2000	70	/	0	8
	10:32	10:58	26min	20min				social media	10	4
	12:33	12:35	2min	0min				/	0	5
	15:57	14:05	8min	8min				game	6	5
	18:11	18:21	10min	10min				social media	6	6
	19:23	19:24	1min	0min				/	0	6
	22:09	22:15	6min	0min				/	0	7
19 Oct	23:34	23:55	21min	20min	6	1800	60	game	10	7
	10:35	10:45	10min	10min				social media	5	5
	13:13	13:14	1min	0min				/	0	6
	14:47	14:50	3min	0min				/	0	5
	18:52	18:55	3min	3min				news	2	5
	20:16	20:17	1min	0min				/	0	6
20 Oct	23:19	23:31	12min	12min	6	1700	10	game	6	6
	00:27	00:28	1min	0min				/	0	7
	10:50	11:03	13min	13min				social media	6	4
	14:58	15:19	21min	19min				game	8	6
	17:16	17:22	6min	6min				news	4	6
	20:03	20:15	12min	10min				social media	5	6
21 Oct	23:22	23:30	8min	5min	8	1300	10	social media	3	7
	09:57	09:58	1min	0min				/	0	3
	11:08	11:34	16min	16min				social media	8	4
	14:56	14:57	1min	0min				/	0	5
	17:15	17:16	1min	0min				/	0	7
	17:57	18:02	4min	0min				/	0	7
	21:15	21:23	8min	3min				game	8	6
24 Oct	22:19	22:21	2min	0min	6	1000	30	/	0	5
	23:28	23:44	16min	15min				game	7	7
	10:43	10:48	5min	3min				news	2	4
	12:56	13:07	11min	5min				news	7	5
	15:12	15:16	4min	0min				/	0	5
	19:24	19:27	3min	0min				/	0	5
27 Oct	22:19	22:46	27min	25min	5	1700	60	game	3	
	23:26	23:55	29min	25min				social media	10	5
	11:31	11:42	11min	7min				news	7	4
	14:44	14:52	8min	8min				social media	5	5
	17:11	17:14	3min	3min				social media	4	7
29 Oct	20:32	20:38	6min	3min	5	1500	60	social media	2	5
	22:16	22:21	5min	0min				/	0	6
	10:32	10:39	7min	7min				news	6	3
	13:07	13:34	27min	20min				social media	10	7
	16:52	17:03	11min	7min				news	6	5
29 Oct	21:41	21:55	14min	10min	5	1500	60	social media	8	6
	23:31	23:47	16min	16min				game	10	7

	09:58	10:07	9min	4min	5	1200	40	social media	3	3
Nov 3	12:42	13:03	21min	20min				game	10	6
	16:16	16:21	5min	0min				/	0	5
	19:33	19:41	8min	7min				game	7	3
	23:05	23:09	4min	0min				/	0	5
	11:21	11:27	6min	6min				social media	4	3
Nov 7	13:08	13:21	13min	10min	5	800	20	news	5	6
	17:27	15:36	9min	5min				news	7	5
	20:43	21:04	21min	20min				game	9	7
	23:15	23:29	14min	14min				social media	8	5
	08:52	08:57	5min	5min				/	6	5
Nov 12	12:18	12:27	9min	9min	5	1200	30	social media	4	6
	16:35	16:43	8min	8min				social media	6	7
	20:21	20:24	3min	0.00				/	0	3
	23:04	23:27	23min	23min				game	9	8
	09:45	10:01	16min	16min	5	1000	50	social media	8	2
Nov 17	14:22	14:28	6min	/				/	0	5
	18:17	18:41	24min	20min				game	9	6
	21:25	21:45	20min	20min				game	0	5
	10:27	10:32	5min	5min				news	3	2
	14:09	14:16	7min	7min				social media	4	4
Nov 21	17:34	17:47	13min	13min	5	950	60	social media	5	6
	20:25	20:34	9min	9min				news	5	7
	22:43	23:01	22min	19min				game	8	5
	10:12	10:23	9min	9min				news	4	3
	14:58	15:07	9min	/				/	0	5
Nov 26	20:07	20:24	17min	17min	4	1200	30	game	8	6
	23:11	23:24	13min	13min				game	6	5
	10:57	11:02	5min	5min				news	3	4
	13:21	13:28	7min	7min				news	5	4
	18:25	18:31	6min	6min				news	4	5
Dec 1	20:34	20:57	23min	23min	5	1100	40	game	9	4
	23:19	23:27	8min	8min				news	4	4

8. Works Cited

"Area Graph." The Data Visualisation Catalogue,
https://datavizcatalogue.com/methods/area_graph.html. Accessed Nov. 2025.

"Circle Packing." Observable, D3, <https://observablehq.com/@d3/pack/2>. Accessed Nov. 2025.

"d3.mean() Function." GeeksforGeeks, <https://www.geeksforgeeks.org/d3-js-d3-mean-function/>. Accessed Nov. 2025.

Northeastern University, DGM 6108/6109/6501: Course Materials. Fall 2025.

"The 52 Best — and Weirdest — Charts We Made in 2016." FiveThirtyEight, Nov. 2016, <https://fivethirtyeight.com/features/the-52-best-and-weirdest-charts-we-made-in-2016/>.

Van De Wiel, Marco J. "Clustered Bubble chart of Circular Economy and Nexus." ResearchGate, June 2021, https://www.researchgate.net/figure/Clustered-Bubble-chart-of-Circular-Economy-and-Nexus-top-20-most-frequent-meaningful_fig4_352083307.