The Puzzle of the Distinct Solitude Experiences in Our Daily Life*

Greater Social Quality, Higher Level of Self-Efficacy and Less Anxiety are Keys to A Better Solitude Experiences.

Qiao Zhou

05 April 2022

Abstract

Solitude is a double-edged sword; people have quite distinct solitude experiences; whether positive or negative. This report utilizes the data from an experiment to:1)figure out which factors correlate with the positive solitude experiences and negative Solitude experiences. And 2)discuss the potential bias in our data. According to the data and analysis, a better social quality, higher self-efficacy, and greater age will lead to better solitude experiences. In contrast, a higher level of anxiety will lead to worse solitary experiences. However, as our dataset is too small and not representative to some extent, our conclusion may get less accurate than we expect.

Keywords: Solitude Experience , Linear Regression, Experiment, BIC Selection

1 Introduction

Solitude is a ubiquitous experience in our daily lives(Lay et al. 2019), especially during the covid pandemic. And solitude is a double-edged sword(Lay et al. 2019). Some people prefer time alone over time spent with others, while others feel very depressive and lonely when they experience solitude. This fact made me curious about how solitude can be both lonely and nourishing. What's more, I have learned that many people misunderstand this word as loneliness and most psychological research on solitude emphasizes the negative correlates and consequences of loneliness(Long and Averill 2003). This fact may imply that most people feel inadequate and lonely when they experience solitude, which makes it even worthwhile for me to investigate the factors impacting solitude experiences. Hopefully, my research paper will help people get better solitude experiences.

This paper will investigate what factors will impact our solitude experiences. I divided the solitude experiences into two types- one reflecting negative experiences such as loneliness and complex thoughts(negative solitude experiences), and the other reflecting positive experiences such as calm effects and pleasant thoughts(positive solitude experiences)(Lay et al. 2019). For further investigation, I aim to explore the data to discuss the potential correlation between the selected factors(age, gender, the rate of overall well being, social network size, social relationship quality, social self-efficacy, the self-reflection, solitude time, and anxiety level) and both the positive and negative solitude experiences. Based on my common sense and prior research, my initial hypothesis is that all of the selected variables correlate with both solitude experiences. However, the results are not the same as my expectation. The results show that people with better social quality and greater age will have more positive solitude experiences and fewer negative solitude experiences. Also, our results have shown some other factors determining the solitude experiences, such as that people with higher anxiety levels will experience more serious negative solitude experiences. And the people with a higher rate of self-reflection will experience more positive solitude. Besides, all the variables we didn't mention in the above results have little or no effect on both solitude experiences.

*Code and data are available at:https://github.com/kzhou1999/the__puzzle__of__solitude__experiences

The rest of the paper is organized as follows: In the Data section (section 2), I explain the data background, the data overview, the cleaning process, and analysis of the potential factors influencing the solitude experiences. In the Model section (section 3), I first introduce the methodology I use for this section. Then, I fit two linear models using all of the selected variables(one for the positive solitude experiences and the other one for the negative solitude experiences). And finally, I made a variables selection for each model and determined the final models. In the Result section(section4), I interpret both the final models I got from the last section to show the impact of the final selected variables corresponding to both positive and negative solitude experiences. In the discussion (section 5), I comment on the potential bias in the paper, such as the limitation in data, p-hacking, and other possible variables with correlation to the solitude experiences not included in the dataset. And I will suggest what we should do to further investigate this topic in the future.

2 Data

The research uses the R language(R Core Team 2021) as its foundation. I have used packages such as tidyverse(Wickham et al. 2019), here(Müller 2020), car(Fox and Weisberg 2019) and readr(Wickham, Hester, and Bryan 2022) to prepare data and models for this project. I then used dplyr(Wickham et al. 2021), KableExtra(Zhu 2021) and knitr(Xie 2014) to generate table ; patchwork(Pedersen 2020) and ggplot2(Wickham 2016) to generate and organize plots.

2.1 Data Source and the Experiment

The original dataset was gathered during a experiment at the University of British Columbia. One hundred community-dwelling adults aged 50–85 years (64 percent female; 56 percent East Asian, 36 percent European, 8% other/mixed heritage) and 50 students aged 18–28 years (92 percent female; 42 percent East Asian, 22 percent European, 36 percent other/mixed heritage) each completed approximately 30 daily life assessments on their current and desired social situation, thoughts, and affect over ten days for this experiment(Lay et al. 2019). A baseline session, a time-sampling phase, and an exit session were all part of this experiment. Participants completed questionnaires evaluating individual differences (e.g., trait self-reflection) and were trained to use portable electronic devices during the baseline session. Then, starting the day after the baseline session, individuals were beeped three times daily (once in the morning, once in the afternoon, and once in the evening) over a 10-day time sampling period. On each occasion, participants used a touch screen interface on an iPod or iPad mini to answer a brief questionnaire about their thoughts, feelings, and existing and ideal social situation (Lay et al. 2019).

2.2 Data Overview and Data Cleaning process

There are 4571 rows and 51 variables in the original dataset. It contains all the experimental survey results from 150 observations; each participant completed the same survey around 30 times over ten days (3 to 4 times a day). As a result, each survey is recorded in one row. Thus each participant's response is stored in 30 rows. Individual differences (person-level) and temporal sampling (momentary level) variables are included in this dataset. Individual difference variables are collected before and after the trials; this variable contains participants' personal information such as age, gender, and social network size. Every participant's different variables in each row are the same.On the other hand, the time sampling variables contain the data gathered during the experiments. The sampling variables' results are unique in every single row. Every row has a different set of results for the sampling variables. Every participant's number of surveys (from 1 to 30) was recorded in a variable named Assessment Number(Lay et al. 2019).
I did some data cleaning on the original data set because I intend to investigate the potential correlation between the selected factors (age, gender, rate of overall well-being, social network size, social relationship

Table 1: Definitions of Variables

Variables	Definition
sub_health	The rate of the overall wellbeing on the scale of 5
age	Age at baseline session in years
gender	1 = female, 0 = male
network_size	Total network size
social quality	The social relationship quality of participants on the scale of 5
self_efficacy	Social self-efficacy score on the scale of 5
self_reflection	Self-reflection score on the scale of 5
solitude	Overall time in solitude in proportion
anxiety	Personal anxiety score for their social relationship on the scale of 5
mean_pos_affect	The average positive solitude experience in 10 days
mean_neg_affect	The average negative solitude experience in 10 days

quality, social self-efficacy, self-reflection, solitude time, and anxiety level) and both positive and negative solitude experiences. First, I calculate the mean value of each participant's positive affect (average of the high-arousal positive affect and low-arousal positive affect) and negative affect((average of the high-arousal negative affect and low-arousal negative affect) and save them into these two variable columns for further use. Then, I filter out all rows whose assessment number does not equal 1. This method allows me to obtain individual differences and the mean value of positive and negative solitude emotion scores over ten days, with each row representing a unique participant. Then I remove all of the empty cells. Finally, I chose all of the required variables to create the final dataset I need for this study. The definitions of the variables in the cleaned data are shown in the above table(Table 1).

2.3 Descriptive Analysis

2.3.1 Overall distribution of solitude experiences

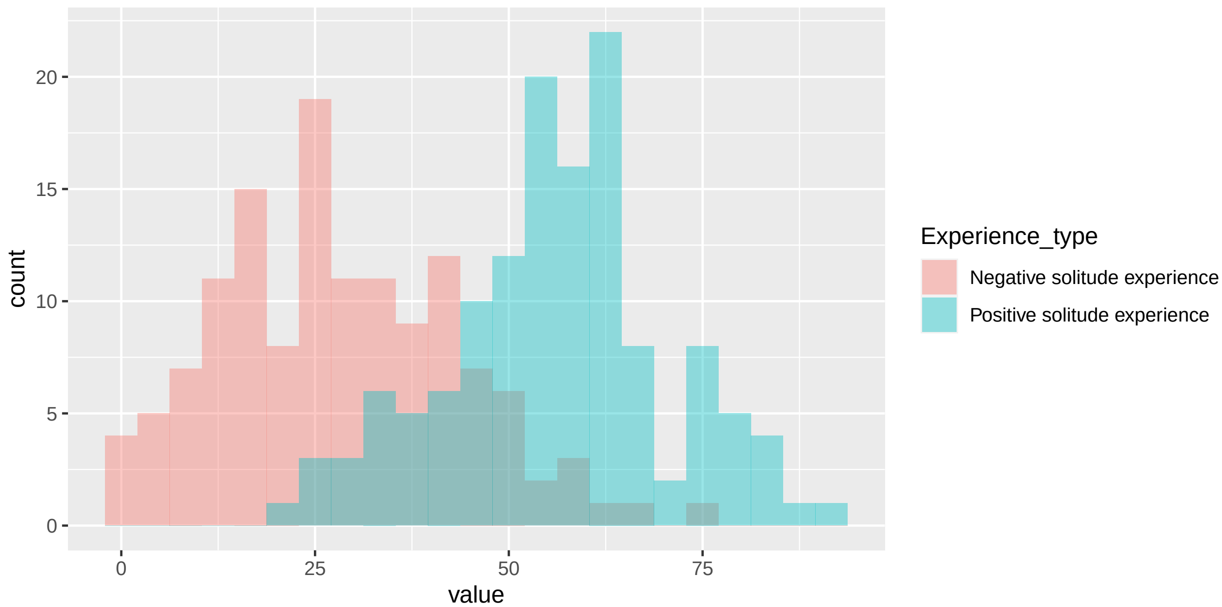

Figure 1: Distribution of solitude experiences

Let's look at the above graph(Figure 1). For the negative solitude experiences painted in pink, we can observe that the distribution follows a normal distribution approximately where the peak is located at around 25 scores. As for the positive solitude experiences painted in blue, its distribution is more likely to be right-skewed and bimodal. We observe that the first peak is around 50 and then decreases sharply. When the score is around 65, the graph reaches the second peak. From this figure (Figure 1), we can conclude that the negative experiences distribute more evenly than the positive experiences. And there is only a small amount of people are having delightful solitude experiences, which further enhances the need to investigate what contributes to good solitude experiences. Fortunately, the average score on the negative solitude experience is lower than the average score on the positive solitude experience. This fact implies that people are not getting severe negative experiences in solitude.

2.3.2 Solitude experiences and the social quality

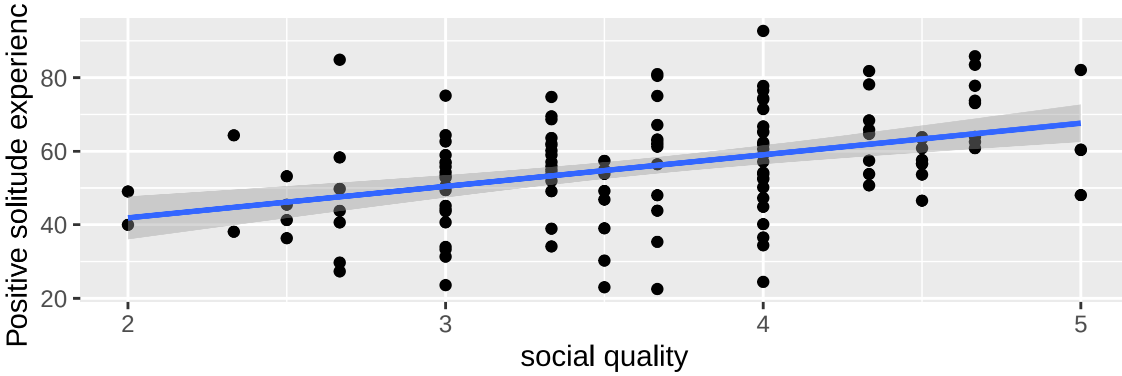

Figure 2: Relationship between the social quality and the positive solitude experiences

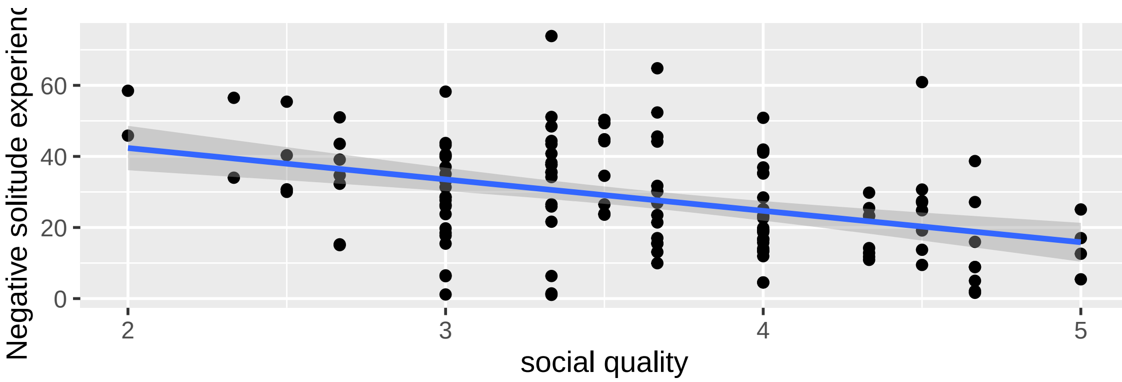

Figure 3: Relationship between the social quality and the positive solitude experiences

From the above graphs(Figure 2)(Figure 3), we can observe linear trends in both plots. Better social quality will lead to a higher score on the positive solitude affect(Figure 2), which means a positive linear relationship between the positive solitude experience and the social quality. As for another graph, on the other hand, we can notice that as the social quality becomes better, the score on negative solitude experiences gets lower(Figure 3), which means that there is a negative linear relationship between the negative solitude experiences and the social quality. These facts imply that better quality is helpful for people to get a better solitude experience.

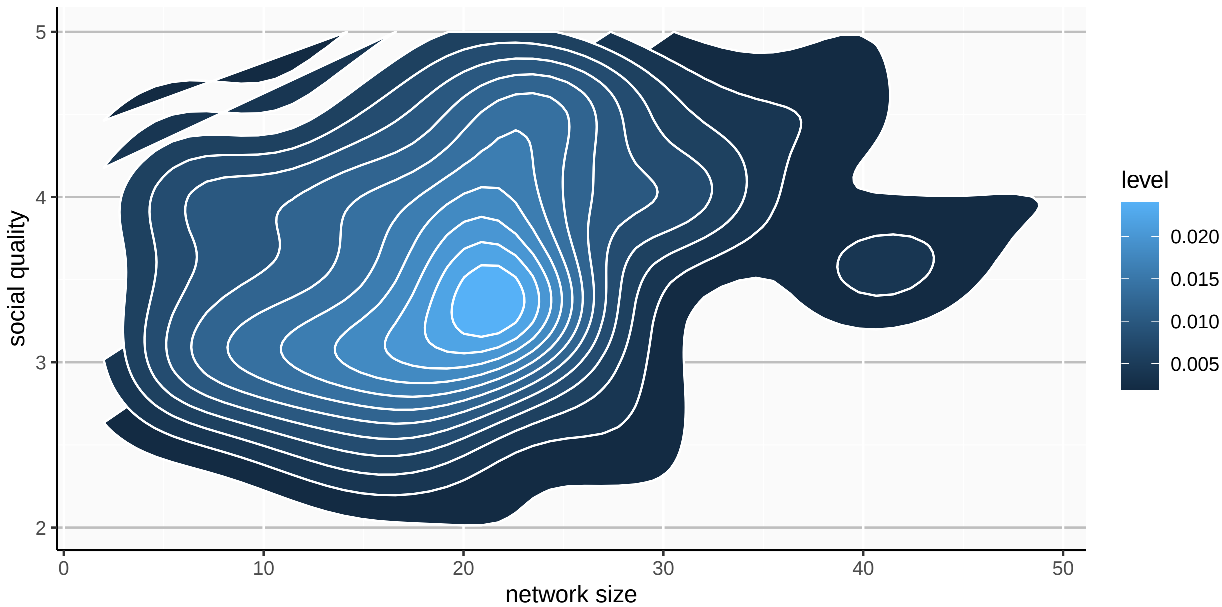

Figure 4: Relationship between the social quality and the network size

Due to the rapid development of technology, social media has become a considerable component of people's social life. Thus, I assume primarily that the network size will influence the social quality to a great extent. However, I am not sure whether this influence will be good or bad. Because spending too much time on the social media will be beneficial for some people since social media helps them to keep in touch with their friends more quickly, but it will also harm people's social quality since people will have less time to hang out and communicate with their friends in real life. To investigate further the relationship between the social quality and the network size, I use our trimmed dataset to plot a density graph(Figure 4). From the above plot(Figure 4), I observed that participants with high network size(total network size greater than 30) tend to have better social quality(with a score on social media more significant than 3). As for the people with relatively smaller network sizes, we can see that their scores on social quality range from the lowest to the highest. But we can still observe that most people with relatively smaller network sizes are still getting a pretty satisfactory score on the social quality at around 3.5. According to the above observations, we can conclude that the network size may have a little bit positive effect on social quality. Still, there is no sign that it dramatically impacts social quality.

2.3.3 Solitude experiences and anxiety

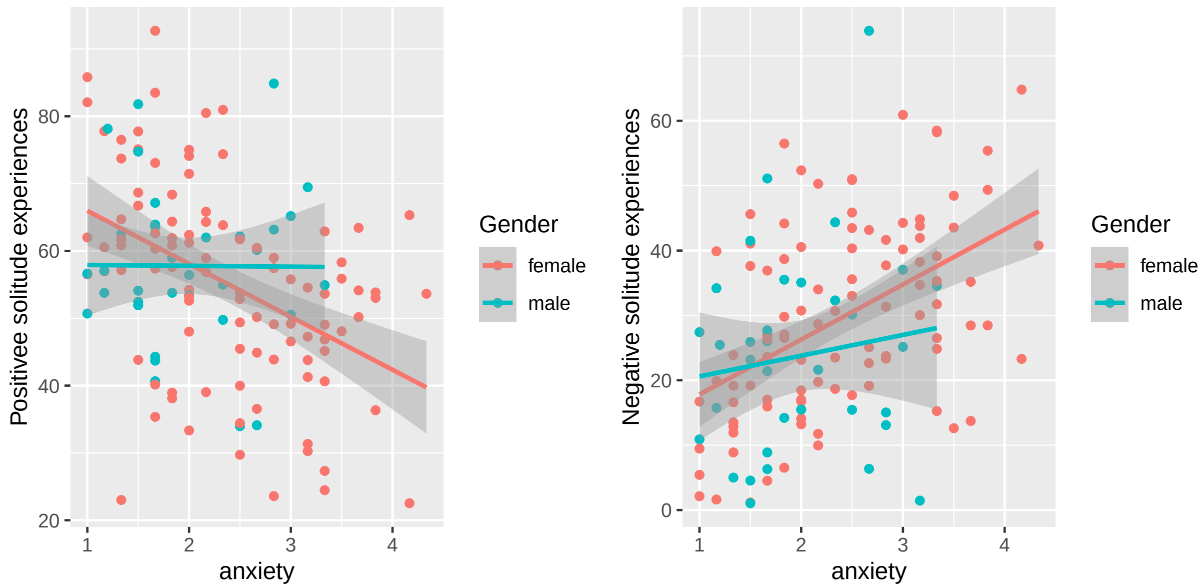

Figure 5: Relationship between the anxiety and the solitude experiences based on gender

The graph above(Figure 5) shows the relationships between the anxiety level and the solitude experiences based on gender. We can see a negative linear trend between the positive solitude affect and the anxiety level. At the same time, there is a positive linear trend between the negative solitude affect and the anxiety level. However, if we take a deep dive into the two plots above(Figure 5), we can notice the trend in both plots differs much in groups of the gender. For the plot on the left, we can see that line(blue line) indicating the impact the anxiety level brings to the men's positive solitude experience is almost horizontal, which suggests that the anxiety level is not affecting the men's positive solitude experience much. However, let's look at the blue line in the left plot. We can notice that the higher anxiety level significantly decreases the score on the positive solitude experience since this line is rather steep. As for the right plot, similarly, we can observe that the red line is more vertical than the blue line, which further proves that the anxiety level affects women's negative solitude more severely than men. Also, for both plot(Figure 5), we can see that all of the people with a high level of anxiety(anxiety level greater than 3.5) are women; this further suggest the differences in the thinking patterns of men and women. Our data concludes that women are more likely to have high anxiety levels and are more easily affected by the anxiety.

2.3.4 Solitude experiences and age

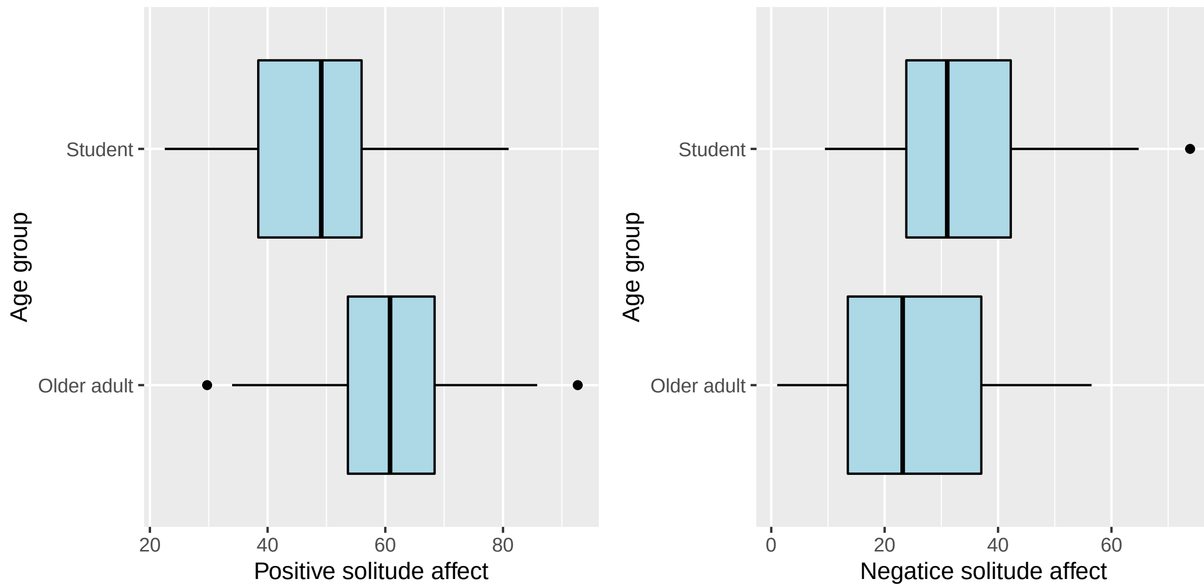

Figure 6: Relationship between the age group and the solitude experiences

In the original experiment, all participants were divided into two groups based on age. The participants are divided into one hundred community-dwelling adults aged 50–85 years and 50 undergraduate students aged 18–28 years. I did the same grouping on age for my dataset. I plot two boxplots(Figure 6) to compare the score on positive and negative solitude experiences between different age groups. Based on my common sense, I will claim that the older group will have a better solitude experience because they are more mature than the students. They can handle the negative affect when being isolated. Now, I will look at the above graphs(Figure 6) to check whether my assumptions are correct. For the positive solitude affect(plot on the left), we notice that the older adult group has a mean score of around 60, which is much higher than the student group. As for the negative solitude affect, on the other hand, the student group has a higher score on the negative solitude affect than the older adult group. These facts match my assumption well.

3 Model

Before starting to fit models, I would like to introduce the methods I will use to fit and choose models.

3.1 Methods

In this paper, I will fit linear models to determine the relationships between the selected factors and the solitude experiences. Thus, I would like first to introduce the linear regression and the linear model. After initially fitting the model, I will check the VIF values and the hypothesis tests for the model to select the predictors, and I will use backward selection(BIC) to determine the final model. Thus, I will introduce the VIF values and the BIC backward selection in the following.

3.1.1 Linear Regression

By fitting a linear equation to observed data, linear regression models the connection between many variables. The linear regression model can be used to predict a particular variable. Independent variables are the variables we utilized to predict care. We expect these variables to have no repeating information, as their name suggests. The dependent variable is the variable we're trying to forecast.
Here is the general linear regression model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_i x_i + \epsilon_i$$

x_i represents the independent variable for the above equation, while y_i represents the dependent variable.β_0 represents the y-intercept of the multilinear regression model. It demonstrates the value of the dependent variable when x_i is zero. This variable can be meaningless at times. β_i is known as the regression coefficient. These numbers indicate how much the dependant value is expected to vary as x_i changes. ϵ_0 is the residual, which is the genuine deviations between the observed values and their means.
The following are the assumptions for the linear model(CFI 2022):

Linearity We assume all the explanatory variable x_i have a linear relationship to the dependent variable y_i

Constant Variance The error should have constant variance. In other words, this assumption suggests the amount of error in the residual for each point is the same and equal to σ^2

Normality All the errors should follow a normal distribution. The mean value should be 0, and the variance is σ^2.

Independence All the errors should be independent of each other. It should be independently and identically distributed. If errors are dependent, it will cause the predictive power of the model will be worse somewhere.

Independence of observations The observations in our model should be independent. It means there should be any relation between observations. We cannot find a pattern or relationship between observations by simply looking at the data frame. Every observation cannot affect others.

3.1.2 VIF Values

The VIF is used to detect multicollinearity in a linear regression model. It will estimate how much a linear regression coefficient variance is inflated due to multicollinearity. The range of the variance inflation factors range starts from 1. The value of VIF interpreted what percentage of the error is inflated for each coefficient. Thus, we want the VIF as small as possible, and we will remove the variables with high VIF values(Stephanie 2020).

3.1.3 BIC and backward selection

First, I will explain the process of backward selection:

1. Starts with the model that contains all the variables under previous consideration (also called the Full Model).

2. Removing the minor significant variables from the model.

3. Stop when the stopping rule is reached out. I use the BIC value as my stopping rule in the Backward selection. I want to find the model with the best AIC value using the "step()" function. The BIC (Bayesian information criterion) evaluates whether the model fits the data. We use the Backward selection function to find the lowest possible BIC.

3.1.4 Hypothesis test

The hypothesis test can test a claim about a proper parameter in a population. We will do the hypothesis test for every single coefficient for the multilinear regression model. Here is how the procedure:

1. First, We need to state the hypothesis, alternate hypothesis, and significance level. In this model, I am interested in the value of coefficients.
2. Then, we can calculate the p-value for every coefficient.
3. we can compare the p-value with a significant level and get our conclusion.
4. If the significant level is greater than the p-value, we can conclude that we have no evidence to support our null hypothesis.

Note: In our paper, the significance level I use is $alpha$=0.05

3.2 Model fitting and selection

I fit two models using all of the selected variables, one for the positive solitude affect and the other one for the negative solitude affect. The variables are age, gender, the rate of overall well being, social network size, social relationship quality, social self-efficacy, the self-reflection, solitude time, and anxiety level.
From the below table(Table 2, we can see that all of the VIF values are relatively small(Note: the vif values for both models are the same because the predictors in both models are the same). So, we decided to keep all those variables in the full models. Then, I use the backward selection to help me select the best model. Our full model for positive solitude affect contains nine variables: age, gender, the rate of overall well-being, social network size, social relationship quality, social self-efficacy, self-reflection, solitude time, and anxiety level. The backward selection model removes six variables from my full model. Thus, the backward selection model for the positive solitude experiences has only three variables; age, social quality, and self-efficacy. And those variables are concluded as having the most correlation with the positive solitude experiences. Our full model for the negative solitude affect contains nine variables as well. The backward selection model also removes six variables from my full model. Thus, the backward selection model for the negative solitude affect has three variables; age, social quality, and anxiety level. And those variables are concluded as having the most correlation with the negative solitude affect.

Table 2: VIF value for the full linear models

	variables of positive solitude model	vif value	variables of negative solitude model	vif value
age	age	1.453489	age	1.453489
gender	gender	1.156494	gender	1.156494
network_size	network_size	1.121330	network_size	1.121330
social_quality	social_quality	1.370842	social_quality	1.370842
self_efficacy	self_efficacy	1.351759	self_efficacy	1.351759
self_reflection	self_reflection	1.272974	self_reflection	1.272974
solitude	solitude	1.125516	solitude	1.125516
anxiety	anxiety	1.407440	anxiety	1.407440
sub_health	sub_health	1.291094	sub_health	1.291094

Final model(Positive solitude affect):

mean positive solitude affect = $\hat{\beta}_0+\hat{\beta}_1 age+\hat{\beta}_2\text{social\_ quality}+\hat{\beta}_3\text{self\_efficacy}+\epsilon$

Final model(Negative solitude affect):

mean negative solitude affect = $\hat{\beta}_0+\hat{\beta}_1\text{age}+\hat{\beta}_2\text{social\_ quality}+\hat{\beta}_3\text{anxiety}+\epsilon$

4 Results

Now, we can finally analyze our linear regression models. Let us first look at the model for the positive solitude affect. The following table includes the estimated result for this model(Table 3).

Table 3: The estimations and the p values for the positive solitude affect model

variables	estimations	p_values
intercept	0.74659	0.915136
age	0.24844	0.000000
social_quality	5.96085	0.000221
self_efficacy	5.78392	0.001318

We first look at $\hat{\beta}_0$. This value is the intercept of our linear regression models. It means when the other variable equals 0. Accordingly, our score on positive solitude is 0.74659, which is almost 0. This result is meaningless in an actual situation since the score for the positive solitude can hardly be 0. Also, we found out that the p-value for the intercept is more significant than the significance level of 0.05, which further proves my statement that this intercept estimate is meaningless. Then, for all of the predictors, we can find out that all p values are less than the significance level of 0.05. Therefore, I conclude that we have strong evidence to support our null hypothesis; thus, the positive solitude affect is highly correlated with age, social quality, and self-efficacy.
Now, let's discuss the estimations for each predictor. The estimate for the age is 0.24844, which means that one year increase in age while holding other variables constant will lead to a 0.24844 increase in the score for the positive solitude experience. This result shows that age has a positive relationship with the positive solitude affect, as we expect, but the impact on the positive solitude affect is relatively tiny. As for the social quality, the estimation for the social_quality is 5.96085, which means that one score increase in the social_qualoity while holding other variables constant will lead to a 5.96085 increase in the score for the positive solitude experience. This result shows that the social_quality do a positive relationship with the positive solitude affect, as we expect, and the impact on the positive solitude affect is significant. Lastly, for the self-efficacy, since we have not dived into this concept earlier in this paper, I will first introduce what self-efficacy refers to. Self-efficacy refers to an individual's belief in their capacity to execute behaviors necessary to produce specific performance attainments("Self-Efficacy Teaching Tip Sheet," n.d.). The estimation for the social_quality is 5.78392, which means that one score increase in the self_efficacy while holding other variables constant will lead to a 5.78392 increase in the score for the positive solitude experience. This result shows that the self_efficacy has a positive relationship with the positive solitude affect and significantly impacts the positive solitude affect.

Let's also look at the model for the negative solitude affect. The following table includes the estimated results for this model(Table 4).

Table 4: The estimations and the p values for the negative solitude affect model

variables	estimations	p_values
intercept	50.09895	3.00e-07
age	-0.14323	6.95e-03
social_quality	-6.83339	1.40e-04
anxiety	4.33072	7.43e-03

We first look at $\hat{\beta}_0$. This value is the intercept of our linear regression models. It means when the other variable equals 0. Accordingly, our score on negative solitude is 50.09895. This result is significant since the p-value is very small. Then, for all of the predictors, we can find out that all p values are less than the significance level of 0.05. Therefore, I conclude that we have strong evidence to support our null hypothesis;

thus, the positive solitude affect is highly correlated with age, social quality, and anxiety.
Now, let's discuss the estimations for each preditor. The estimation for the age is -0.14323, which means that one year increase in age while holding other variables constant will lead to a 0.14324 decrease in the score for the negative solitude experience. This result shows that age has a negative relationship with the negative solitude affect, as we expect, but the impact on the positive solitude affect is also tiny. As for the social quality, the estimation for the social_quality is -6.83339, which means that one score increase in the social_qualoity while holding other variables constant will lead to a 6.83339 decrease in the score for the negative solitude experience. This result shows that the social_quality has a negative relationship with the negative solitude affect as we expect, and the impact on the negative solitude affect is significant. Lastly, for the anxiety, the estimation for the anxiety is 4.33072, which means that one score increases the anxiety level while holding other variables constant will lead to a 4.33072 increase in the score for the positive solitude experience. This result shows that anxiety does a positive relationship with the negative solitude affect, and the impact on the positive solitude affect is significant. (i.e., a higher level of anxiety will lead to more server negative solitude affect, which is not good)
To conclude these two models, we have learned that only four variables significantly correlate with the solitude experiences. Age is associated with both types of solitude experiences, and a greater age will lead to better solitude, though the impact of age on solitude experience is tiny. Social quality is correlated with both types of solitude experiences as well, as a better social quality will to a great extent, make the solitude experience better. Although self-efficacy is only shown to be related to the positive solitude affect, having a higher self-efficacy level can help improve the solitude experience a lot. As for the anxiety level, it is positively related to the negative solitude affect. Thus we have learned that we should try not to be anxious too much to have better solitude experiences.

5 Discussion

5.1 Conclusion

During the covid pandemic, especially in my home country China, we spend a lot of time doting quarantine to prevent the virus from spreading. Many people get too depressed when they are forced to be isolated. Thus, I need to investigate what can help people have a better solitude experience, and I have done so. The final result of my paper implies that the social relationship quality strongly relates to positive and negative solitude experiences. By increasing their social relationship quality, participants can quickly reduce their negative solitude feeling in their daily life. Thus, for people trying to have an enjoyable solitude experience, it might be better for them to improve their social quality. However, I know it is tough to improve one's social quality in a short time. Our result shows a more straightforward way for people to have better solitude experiences: trying not to be anxious. As anxiety level has a significantly positive relationship to the negative solitude experiences, trying to be less anxious can directly reduce the negative solitude experience and help us feel better.

5.2 Limitations

Our dataset has a few limitations that may affect our final result. First of all, the sample size of our dataset is too small, which makes our dataset not as representative as we expect. Second, the original experiments have more female participants than male participants and more dwelling adults than students. This setting may result in potential bias in our conclusions; for example, we conclude that females are more likely to have a high level of anxiety and are easily affected by the anxiety. However, this conclusion may occur simply due to the more significant number of female participants. Third, we notice that the age gap between the students and adult are pretty high. As a result, we don't gain any information for that younger adult, which may significantly affect our final results. Lastly, I want to mention the potential bias when doing the data cleaning. I choose only the arousal positive and negative affect and use them to represent positive and negative solitude experiences, which is not considered.

Then, I want to discuss the potential limitation of fitting, choosing, and interpreting the models. For my paper, I use linear regression as the primary method; however, as I mentioned in the previous section, linear regression requires multiple assumptions to be accurate. I did not check strictly on those assumptions; thus, the final result we get will be insignificant if the dataset fails to meet any assumptions. Also, as we interpret the model, we regard our development as significant only by saying the p-values are less than the significance level, which may bring up the potential risk of p-hacking.

5.3 Next step

For our future study, I think we first need a greater sample size for our research to become more representative. Secondly, we must balance the number of female and male participants and ensure that we have similar data on different age groups. Third, I think it will be better to collect additional information for our dataset; for example, I think we can generate information about the participant's educational background, job position, and family background... Forth, we need more than the simple linear regression since the variables in our dataset are not entirely independent of each other. Then using the linear model to get the final result will not be that accurate. More complex models are needed. Lastly, I want to mention the relationship between social quality and the total network size. I have studied this for a bit in this paper, and I feel that this topic is worth digging into since it has been a very confusing problem for a long time.

Appendix

Data Sheet:

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*
    - The original dataset was gathered during a experiment at the University of British Columbia. One hundred community-dwelling adults aged 50–85 years (64 percent female; 56 percent East Asian, 36 percent European, 8% other/mixed heritage) and 50 students aged 18–28 years (92 percent female; 42 percent East Asian, 22 percent European, 36 percent other/mixed heritage) each completed approximately 30 daily life assessments on their current and desired social situation, thoughts, and affect over ten days for this experiment. It was used to investigate the predictors of distinct types of solitude experiences in daily life.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
    - Jennifer C. Lay, Theresa Pauly, Peter Graf, Jeremy C. Biesanz and Christiane A. Hoppmann from department of Psychology inUniversity of British Columbia created this dataset.
3. *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*
    - Vancouver Foundation grant to Christiane Hoppmann, Sandra Petrozzi, Atiya Mahmood, and Peter Graf; University of British Columbia Faculty of Arts grant to Christiane Hoppmann; University of British Columbia Alpha Mater Society grant to Jennifer Lay; support from Michael Smith Foundation for Health Research and the Canada Research Chairs Programme to Christiane Hoppmann; support from Social Sciences and Humanities Research Council of Canada (Vanier CGS Program) to Jennifer Lay.
4. *Any other comments?*
    - No

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*
    - The only instance is people
2. *How many instances are there in total (of each type, if appropriate)?*
    - 150
3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*
    - The dataset is a sample of instances from a larger set and it is not randomly sampled. The sample is not a representative of the larger set, because this experiment does not randomly sample from the community using some random sampling methods like Simple Random Sampling or Stratified Random Sampling, and this may lead to the selection bias.

4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*

   - Individual differences (person-level) and temporal sampling (momentary level) variables are included in this dataset. Individual difference variables are collected before and after the trials; this variable contains participants' personal information such as age, gender, and social network size. Every participant's different variables in each row are the same.On the other hand, the time sampling variables contain the data gathered during the experiments. The sampling variables' results are unique in every single row.

5. *Is there a label or target associated with each instance? If so, please provide a description.*

   - There is no label associated with each instance because this is not a machine learning task.

6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*

   - The socres on anxiety level and self-efficacy; the total network size are missing for some participants.

7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*

   - No

8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*

   - There is no recommended data split since this task is to do statistical analysis on an experiment, rather than a machine learning task that needs to split the data into training and test data.

9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*

   - There are few missing values in the dataset and there are some extreme values for the response variable and explanatory variables.

10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*

    - The dataset can be downloaded from the https://can01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdataverse.scholarsportal.info%2Fdataset.xhtml%3FpersistentId%3Ddoi%3A10.5683%2FSP2%2FBUPAJK&data=04%7C01%7Ckatherine.zhou%40mail.utoronto.ca%7Cebd0555947bb486ef22a08da11c255a0%7C78aac2262f034b4d9037b46d56c55210%7C0%7C0%7C637841823591320507%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI
3D%7C3000&sdata=5fIalUC9FTnFimlBSFELLyM4XLBdhR1rSuZUgD4RBnA%3D&reserved=0 website freely.

11. *Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*

    - The dataset does not contain the confidential information, because each instance is represented using a unique id number and there is no demographic information included.

12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*

- No

13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*

    - The dataset identifies two subgroup by age-one group is composed of students aged 18-28 years old and another is group of dwelling adults aged 50-88 years old.

14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*

    - It is possible to identify individuals directly by the unique id number.

15. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*

    - No

16. *Any other comments?*

    - No

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*

    - One hundred community-dwelling adults aged 50–85 years (64 percent female; 56 percent East Asian, 36 percent European, 8% other/mixed heritage) and 50 students aged 18–28 years (92 percent female; 42 percent East Asian, 22 percent European, 36 percent other/mixed heritage) each completed approximately 30 daily life assessments on their current and desired social situation, thoughts, and affect over ten days for this experiment. Participants completed questionnaires evaluating individual differences (e.g., trait self-reflection) and were trained to use portable electronic devices during the baseline session. Then, starting the day after the baseline session, individuals were beeped three times daily (once in the morning, once in the afternoon, and once in the evening) over a 10-day time sampling period. On each occasion, participants used a touch screen interface on an iPod or iPad mini to answer a brief questionnaire about their thoughts, feelings, and existing and ideal social situation

2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*

    - The data was recorded by experimenters manually. There may exist measurement error, but how this data collection procedure was validated was unknown.

3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*

    - No random sampling method is used.

4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*

- Students and adults. Older adults were reimbursed with up to $100 or the iPad mini they had used in the study. Students were reimbursed with 3 course credits and up to $30 (differences in compensation between the two samples reflect that older adults also attended a 6-month follow-up session, whereas students did not).

5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*

   - Data were collected year-round (August2014–May 2016). This timeframe matches the creation timeframe of the data associated with the instances, because this data was collected during the implementation of the experiment to record its results.

6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*

   - The study was approved by the university behavioral research ethics board.

7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*

   - Unkown

8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*

   - Unkown

9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*

   - Unknown

10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*

    - Unkown

11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*

    - Yes. Available at https://www.researchgate.net/publication/326371343_By_Myself_and_Liking_It_Predictors_of_Distinct_Types_of_Solitude_Experiences_in_Daily_Life

12. *Any other comments?*

    - No

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*

- Since most data in the dataset is categorical/discrete/continuous data and a few missing values exist in the dataset, there was no preprocessing/cleaning/labeling of the data done. But I cleaned the data by removing the missing values and muating a few columns.

2. *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.*
   - No need to answer since the data itself is raw data.
3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*
   - No need to answer since the data itself is raw data.
4. *Any other comments?*
   - No

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*
   - Yes. Available at https://www.researchgate.net/publication/326371343_By_Myself_and_Liking_It_Predictors_of_Distinct_Types_of_Solitude_Experiences_in_Daily_Life
2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*
   - Unkown
3. *What (other) tasks could the dataset be used for?*
   - Unknown
4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*
   - I don't think so.
5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*
   - Unknown
6. *Any other comments?*
   - No

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*
   - Yes, it can be download freely online.
2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*
   - On website
3. *When will the dataset be distributed?*

- The dataset was distributed on 2019

4. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*
   - Unknown
5. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*
   - I did not see any restrictions.
6. *Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*
   - I did not see any restrictions.
7. *Any other comments?*
   - No

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
   - The original authors who performed the experiment and collected the data are supporting/hosting/ maintaining the dataset.
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
   - You can email the Scholars Portal Dataverse Support
3. *Is there an erratum? If so, please provide a link or other access point.*
   - Unknown
4. *Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?*
   - Unknown. The dataset was downloaded from the website.
5. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.*
   - Unknown
6. *Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*
   - Unknown
7. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*
   - Unknown
8. *Any other comments?*
   - No

References

CFI, Education inc. 2022. *Multiple Linear Regression. Corporate Finance Institute.* https://corporatefinanceinstitute.com/resources/knowledge/other/multiple-linear-regression/.

Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression.* Third. Thousand Oaks CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Lay, Jennifer C., Theresa Pauly, Peter Graf, Jeremy C. Biesanz, and Christiane A. Hoppmann. 2019. "By Myself and Liking It? Predictors of Distinct Types of Solitude Experiences in Daily Life." *Journal of Personality* 87 (3): 633–47. https://doi.org/10.1111/jopy.12421.

Long, Christopher R., and James R. Averill. 2003. "Solitude: An Exploration of Benefits of Being Alone." *Journal for the Theory of Social Behaviour* 33 (1): 21–44. https://doi.org/10.1111/1468-5914.00204.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files.* https://CRAN.R-project.org/package=here.

Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots.* https://CRAN.R-project.org/package=patchwork.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

"Self-Efficacy Teaching Tip Sheet." n.d. *American Psychological Association.* American Psychological Association. https://www.apa.org/pi/aids/resources/education/self-efficacy.

Stephanie. 2020. *Variance Inflation Factor. Statistics How To.* https://www.statisticshowto.com/variance-inflation-factor/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2022. *Readr: Read Rectangular Text Data.* https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.