# **Data and Methods**

## **Data Used**

### ***General Social Survey***

The main dataset used in this analysis is the General Social Survey, a nationally representative survey of the attitudes and behaviors of U.S. adults, pulled from the official National Opinion Research Center (NORC) website. The survey is taken by NORC every one to two years, with cumulative cross-sectional datasets available from the year 1972 to 2021[[1]](#footnote-1). The specific data employed will come from the 2016-2020 Panel (Davern et al., 2022) updated in April 2022, which sampled 6,200 housing units in 2016, with a final tally of 2,867 completed individual interviews, and 5,200 housing units in 2018, with a final tally of 2,348 respondent individuals. While the total number of recorded observations was 5,215, considering that only a fraction of each year’s respondent also participated to the 2020 survey[[2]](#footnote-2), the total number of identifiable observations was 1,823 (34.95%). To track the same respondents across waves, only those who participated in the 2018 survey were considered (1,014).

### ***American National Elections Survey***

All respondents who completed the 2020 wave of the GSS and were U.S. citizens at the time of study were then offered to answer a second survey administered by the American National Elections Study (ANES). The ANES is one of the oldest continuous series of survey data of electoral behavior and general attitudes in the United States and it is used here to track social participation determinants as mediators between loneliness and wellness. By tracking political involvement within the United States, the variables extracted from the ANES will give us a clearer view of the role of community participation within patterns of online and offline communication. In fact, the surveys are taken before and after presidential and national congressional elections (Howell, 2022), granting an additional panel observation year to 635 GSS respondents. In order to utilize this dataset and its information in the analysis, the model sample will be limited to these 635 observations.

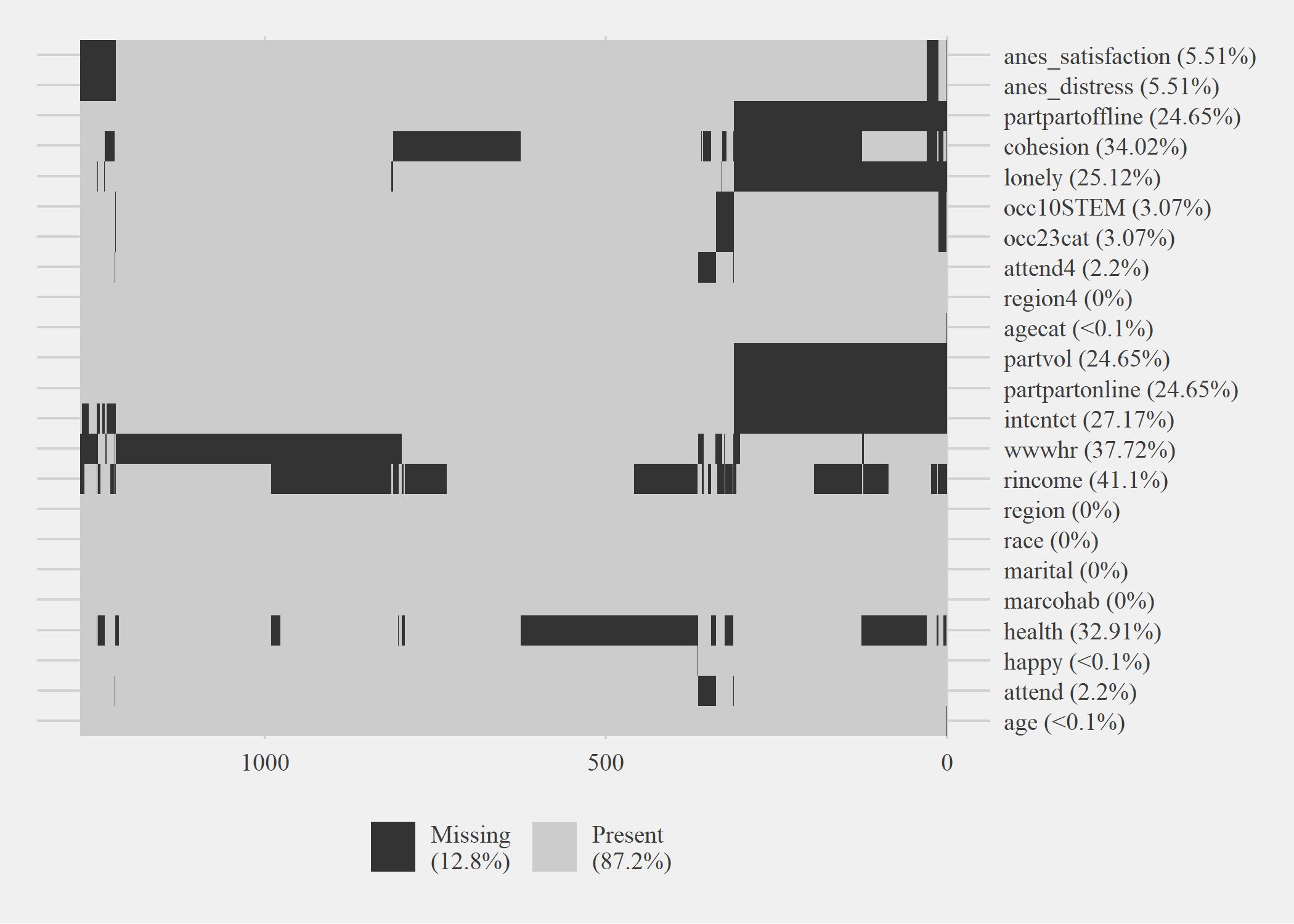
**Table 1. Descriptive Statistics of Selected Sample (Demographic)**

**Table

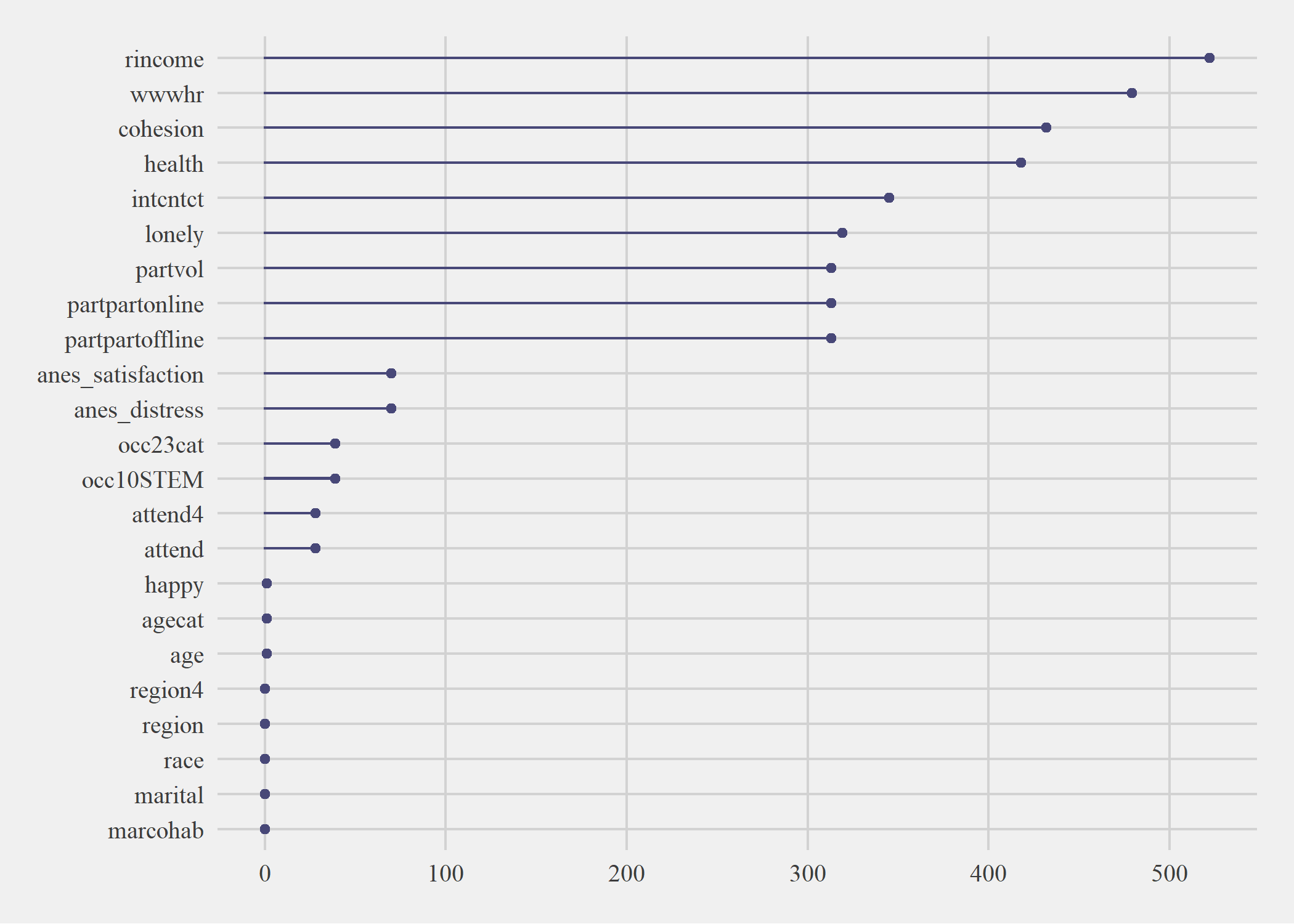
Description automatically generated**

## **Addressing Missingness**

Panel data is unfortunately prone to missingness due to attrition or methodological changes across panel years. In fact, even in the 2016-2020 panel, though wording remains relatively identical across included variables, certain questions are either omitted or changed from the 2016 and 2018 waves to the 2020 wave. Although non-response negatively affects the explanatory power of certain variables, the 2016-2020 panel was selected for its relative stability in questionnaire variation and representativeness compared to other panels. In fact, a previous analysis by Smith and Son (2010) on the patterns of missingness within the 2006-2008 survey panel indicates that more complex questions tend to be the ones to attract missingness, prompting us to utilize simpler questions to avoid attrition. Furthermore, certain questions with similar prompts across years were used to compensate for the possible lack of continuity between 2018 and 2020. Figure 4 and 5 show clusters of missingness and the variables with the highest number of missing values.



**Figure 4. Clustered Missingness Patterns in the Data**

****

**Figure 5. Variables in the Dataset Ordered by Missingness**

## **Dependent Variable**

Self-Perceived happiness (*happy*) was measured through a 3-point scale answering the following: “Taken all together, how would you say things are these days—would you say that you are very happy, pretty happy, or not too happy?” The scale adopted the Very Happy, Pretty Happy, and Not Too Happy distinction, which was not modified for easier categorization. An alternative happinessvariable (*anes\_satisfaction*) was obtained through the ANES addendum, which asked the respondents to answer the following: “All things considered, how satisfied are you with your life as a whole these days?” The answers were collapsed from their original 5-point format to the 3-point scale of the first happiness variable to maintain consistency.

## **Independent Variables**

The calculation of the independent variables will vary across years, as not all elements of the 2018 questionnaire were included in the 2020 questionnaire. However, the wording of differentiated variables was similar across the two waves, and the ANES addendum helps complement the missing aspects of certain omitted variables. To this latter point, certain variables were combined to complete the missing information.

### ***Social Cohesion***

Measures for social cohesion in the GSS came separated into three dummies with three distinct set of responses: *trust* (“Generally speaking, would you say that most people can be trusted or that you can't be too careful when dealing with others?”), *fair* (“Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?“), and *helpful* (Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?”). Certain authors like Mewes, Fairbrother, Giordano, Wu, & Wilkes (2021) took the mean of the three variables to obtain a generalized dependent scale going from 0 to 1, with 0 indicating negative connotations of trust, while numbers closer to 1 pointed at higher trust and social cohesion. On the other hand, others like Glanville, Andersson, and Paxton (2013) created a composite index of latent generalized trust, which was based on a previous analysis by Paxton (1999) confirming the stability of these parameters when aggregated together[[3]](#footnote-3).

In this paper, the parameters were standardized into a 0 and 1 dichotomy, collapsing the “Depends” answer common of all three into the 0 level (Carl & Billari, 2014), and a scale was created measuring from 0 (“Not Fair, Not Helpful, Not Trustworthy”) to 3 (Fair, Helpful, Trustworthy. The scale permits an easier indication of degrees of social cohesion, while treating all components with similar weight of importance.

### ***Loneliness***

Measures of loneliness were identified by the *lonely3* parameter (“How often in the past 4 weeks have you felt that you are left out?”) in 2018, and by the *lonely1* and *lonely2* parameters (“How often in the past 4 weeks have you felt that you lack companionship?”; “How often in the past 4 weeks have you felt that you are isolated from others?”) in 2020. While *lonely3* seems to directly capture the aspect of emotional loneliness we want to express (Prohaska et al. 2020), *lonely1* and *lonely2* seem to distinguish emotional loneliness with physical loneliness, the latter as a consequence of concrete physical isolation (Holt-Lunstad & Steptoe, 2022). The variable *conwkday* in 2018 measures the number of people a respondent contacts in a typical weekday (“Please indicate about how many people do you have contact with on a typical weekday irrespective of whether you know them or not. Include anyone you chat with, talk to, or text, either face-to-face, by phone, internet or any other communication device.”), which was used as a way to pair the *lonely3* variable in 2018 to mirror the physical-emotional dichotomy available in 2020. As such we would obtain one merged measure of both emotional and physical loneliness, which gives equal importance to physical and mental distress (Lee & Lee, 2010; Vacchiano & Bolano, 2021). Previous literature predicts that just physical loneliness should have a small, though still negative, effect on wellness if we take into account a person’s preferred method of communication (Digital or Physical; Steafnone et al., 2011). The influence of including mental loneliness will be determined in this study.

**Table 2. Descriptive Statistics of Selected Sample (Dependent and Indipendent Variables**Table

Description automatically generated

### **Behavioral Mediators**

### ***Communication Methods***

Measures of communication were determined by the variables *conf2f* (“About how many of these people do you see face-to-face on a typical weekday?”) and *intcntct* (“Think now of your contact with all of your family members and close friends. How much of it is through text messages, mobile phones, or other communication devices that use the internet?), which were combined to consider the number of physical interactions and digital interactions in 2018. The variables V202541a to V20254f, which track use of various social media sites, were used to follow up on online usage in 2020. Finally, the resulting *intcntct* variable was collapsed into a binary differencing high levels of online presence (1) and low to mid-levels of online presence (0). It is expected that the effect of social digital interactions will be driven by the level of social real-life interactions (Filiposka et al., 2017), and that placing online communication frequency as a mediator across the relationship with political participation will damper the positive effect of the latter on wellness.

### ***Political Participation***

Differing from previous research, the behavioral aspect of social trust is placed upon a person’s willingness to participate in political and community activities. This is not the same as spending time with friends and acquaintances, but rather shows a level of connectedness with the individual’s surroundings that is strong enough to warrant collaboration and participation. Questions that tracked a person’s political participation in 2018 were included in the GSS *partpart* and *partvol* variables (“In the past 12 months, how often, if at all, have you taken part in the activities? Of political parties, political groups or political associations?” and “In the past 12 months, how often, if at all, have you taken part in the activities? Of charitable or religious organizations that do voluntary work?”). In 2020, the ANES annex contained 9 different questions tracking political participation, not including direct contact with governmental institutions. These asked for an individual’s participation in political arguments, marches, religious organizations, money donations, online discussions, community problem-solving, school management, and volunteering.

In a similar fashion as the social cohesion independent variable, two grouped variables (one tracking online political participation and one offline political participation) were created indicating if a person had participated to at least three of the following activities:

|  |  |
| --- | --- |
| **Online Political Participation** | **Offline Political Participation** |
| Online Political Meetings and Events | Physical Political Meetings and Events |
| Posting Political Issue Comments Online | Working with Others in Community Issues |
| Signing Internet Petition | Attending Meetings on Community Issues |
| Political Arguments | Political Arguments |
| Donating to Social Organizations | Donating to Social Organizations |

Political participation is expected to have a positive effect on wellness, but its effect will be reduced by the higher levels of online communication (Moy et al., 2005), at either very high or very low levels of social contact.

## **Table 3. Descriptive Statistics of Selected Sample (Behavioral Variables)**

## **Controls**

While the FE model does deal with time-invariant variable effects, certain time-variant aspects of a person’s daily life can indirectly affect the role that online communication has on happiness (marriage happiness, cohabitation status), social cohesion (religious attendance, volunteering activities), and online use (digital occupations).

To account for volunteering activities, the 2018 *partvol* variablewas associated with a grouped binary indicating if respondents had done any volunteering work or given money to a religious organization. Accounting for marriage happiness, *marcohab* tracks if a person is both married *and* living with their spouse, considering the elevated importance of a partner’s intimate relationship within the perception of social connectedness (Prohaska et al., 2020; DeMaris, 2018)). However, some individuals might already have high levels of social connectedness through religion. In fact, attendance at religious events allows people to feel more connected to one another and gives them additional reasons to get together and avoid isolation (Whitehead & Stroope, 2015; Hastings, 2016). Other papers further point at the role of religious participation on an increased sense of wellbeing, connectedness and social participation (Dunbar, 2021; Lewis et al., 2013; Lim & Putnam, 2010). We assume that religiosity is time-invariant within our sample selection, and we only use *attend* as a tracker of religious service attendance (“How often do you attend religious services?).

Finally, we also take into account the role that the digital world has in the individual’s life, as people whose work directly involves heavy internet use might skew simple tracking of hours spent on the web (McClain et al., 2021). As such, *occ10,* which uses the census’ occupational coding,will be adapted into an *occSTEM* dummy placing individuals whose job is labelled as STEM at 1 and others at 0[[4]](#footnote-4).

A picture containing table

Description automatically generated

**Figure 6. Initial Correlation Analysis with Variables in the Model**

## **Model Specification and Testing**

The effect of loneliness on our variables of interest was measured at a high to low categorical scale, expecting increased effects on the extreme ends of the distribution, while the majority near the mean either does not experience significant correlations or sees relationships different from the tails. As such, the model used was a Multinomial Logit Fixed Effects, with the dependent variable being coded as to distinguish the three distinct categories (High, Medium, Low).

### ***Fixed Effects vs. Random Effects***

The choice of Fixed Effects over Random Effects arises from the presence of similar individuals across waves, which allows us to account for time-invariant observed characteristics without the need for further controls. In the case of the Fixed Effects, the individual heterogeneity is included in the intercept and allowed to be correlated with other regressors, while the Random Effects model places the assumption that the individual effect is not correlated with the regressors, estimating the error variance (Park, 2011).

Given that we assume that the observations are unique from each other, and we want to draw conclusions from the differences among each individual, the Within-Effect Fixed Effects model is more appropriate, especially since we have a limited number of respondents and time periods to analyze (Yang & Land, 2008). Further, the limitation placed on the use of only two years of data could make it so that using a First Difference (FD) model would be better:

However, FD assumes that the idiosyncratic error term is serially uncorrelated with each period, which can be problematic with vulnerable population analysis during the COVID-19 period (Especially seniors; Fawcett & Karastoyanova, 2022).

While still maintaining unique categorization, and the difference between extremes and middle values, estimation of the Multinomial Logit model can be performed through a variety of methods that employ the reduction of the model to a binary estimation without loss of information. However, most, like *femlogit* and *feologit* (Baetschmann et al., 2015; 2020) do not allow survey setting of our datasets. In order to guarantee robust standard errors, and assuming that the command works well in small sample sizes (Riedl & Geishecker, 2014),the native *xtmlogit* was used instead.

**References**

American National Election Studies. (2022). *ANES-GSS 2020 Joint Study*[dataset and documentation]. April 8, 2022 version. [www.electionstudies.org](http://www.electionstudies.org/)

Baetschmann, G., Ballantyne, A., Staub, K. E., & Winkelmann, R. (2020). feologit: A new command for fitting fixed-effects ordered logit models. *The Stata Journal*, *20*(2), 253-275.

Baetschmann, G., Staub, K. E., & Winkelmann, R. (2015). Consistent estimation of the fixed effects ordered logit model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *178*(3), 685-703.

Carl, N., & Billari, F. C. (2014). Generalized trust and intelligence in the United States. *PloS one*, *9*(3), e91786.

Davern, M., Bautista, R., Freese, J., Morgan, S. L., & Smith, T. W. (Release 1a, 2022, April). General Social Survey Panel Data (2016-2020). <https://doi.org/10.17605/OSF.IO/HACZV>

Dunbar, R. I. (2021). Religiosity and religious attendance as factors in wellbeing and social engagement. *Religion, Brain & Behavior*, *11*(1), 17-26.

Fawcett, B., & Karastoyanova, K. (2022). Older people, loneliness, social isolation and technological mitigations: utilising experiences of the Covid-19 pandemic as we move forward. *The British Journal of Social Work*.

Filiposka, S., Gajduk, A., Dimitrova, T., & Kocarev, L. (2017). Bridging online and offline social networks: Multiplex analysis. *Physica A: Statistical Mechanics and its Applications*, *471*, 825-836.

Glanville, J. L., Andersson, M. A., & Paxton, P. (2013). Do social connections create trust? An examination using new longitudinal data. *Social Forces*, *92*(2), 545-562.

Hastings, O. P. (2018). Less equal, less trusting? Longitudinal and cross-sectional effects of income inequality on trust in US States, 1973–2012. *Social Science Research*, *74*, 77-95.

Holt-Lunstad, J., & Steptoe, A. (2022). Social isolation: An underappreciated determinant of physical health. *Current Opinion in Psychology*, *43*, 232-237.

Howell, D. (July 14, 2022). The American National Election Studies (ANES) awarded $14 million to study 2024 elections. *ANES* [Press Release]. Accessed on November 17, 2022. <https://electionstudies.org/anes-2024-award/>

Lee, J., & Lee, H. (2010). The computer-mediated communication network: Exploring the linkage between the online community and social capital. *new media & society*, *12*(5), 711-727.

Lewis, V. A., MacGregor, C. A., & Putnam, R. D. (2013). Religion, networks, and neighborliness: The impact of religious social networks on civic engagement. *Social science research*, *42*(2), 331-346.

Lim, C., & Putnam, R. D. (2010). Religion, social networks, and life satisfaction. *American sociological review*, *75*(6), 914-933.

McClain, C., Vogels, E. A., Perrin, A., Sechopoulos, S., & Rainie, L. (2021). The Internet and the pandemic. *Pew Research Center*. Retrieved from <https://www.pewresearch.org/internet/2021/09/01/the-internet-and-the-pandemic/>

Mewes, J., Fairbrother, M., Giordano, G. N., Wu, C., & Wilkes, R. (2021). Experiences matter: A longitudinal study of individual-level sources of declining social trust in the United States. *Social Science Research*, *95*, 102537.

Moy, P., Manosevitch, E., Stamm, K., & Dunsmore, K. (2005). Linking dimensions of Internet use and civic engagement. *Journalism & Mass Communication Quarterly*, *82*(3), 571-586.

Park, H. M. (2011). Practical guides to panel data modeling: a step-by-step analysis using stata. *Public Management and Policy Analysis Program, Graduate School of International Relations, International University of Japan*, *12*, 1-52.

Paxton, P. (1999). Is social capital declining in the United States? A multiple indicator assessment. *American Journal of sociology*, *105*(1), 88-127.

Prohaska, T., Burholt, V., Burns, A., Golden, J., Hawkley, L., Lawlor, B., ... & Fried, L. (2020). Consensus statement: loneliness in older adults, the 21st century social determinant of health?. *BMJ open*, *10*(8), e034967.

Riedl, M., & Geishecker, I. (2014). Keep it simple: estimation strategies for ordered response models with fixed effects. *Journal of Applied Statistics*, *41*(11), 2358-2374.

Smith, T. W., & Son, J. (2010). *An analysis of panel attrition and panel change on the 2006-2008 General Social Survey Panel*. NORC/University of Chicago.

Steafnone, M. A., Huang, Y. C., & Lackaff, D. (2011, January). Negotiating social belonging: Online, offline, and in-between. In *2011 44th Hawaii International Conference on System Sciences* (pp. 1-10). IEEE.

Vacchiano, M., & Bolano, D. (2021). Online and offline leisure, relatedness and psychological distress: A study of young people in Switzerland. *Leisure Studies*, *40*(3), 338-351.

Whitehead, A. L., & Stroope, S. (2015). Small groups, contexts, and civic engagement: A multilevel analysis of United States Congregational Life Survey data. *Social Science Research*, *52*, 659-670.

Yang, Y., & Land, K. C. (2008). Age–period–cohort analysis of repeated cross-section surveys: fixed or random effects?. *Sociological methods & research*, *36*(3), 297-326.‌

1. Information obtained from the GSS website: <https://gss.norc.org/About-The-GSS> [↑](#footnote-ref-1)
2. Panstat tracks whether a respondent was selected and reinterviewed (1 = Selected, Eligible, and Re-Interviewed, 2 = Not Selected, 3 = Selected, but not re-interviewed, 4 = Selected, but not eligible and not re-interviewed because R was deceased, 5 = Selected, but not eligible and not re-interviewed because R was permanently incapacitated, outside the U.S., or otherwise out of scope) [↑](#footnote-ref-2)
3. Paxton’s research also looked into elements of trust in institutions and of social connections, which relates back to the use of social participation as an independent variable (Figure 1 and Figure 2) [↑](#footnote-ref-3)
4. STEM classification was taken directly from the U.S. Bureau of Labor Statistics, and the *occ10* was first divided into the 23 original SOC 10 categories before being turned into a binary. [↑](#footnote-ref-4)