



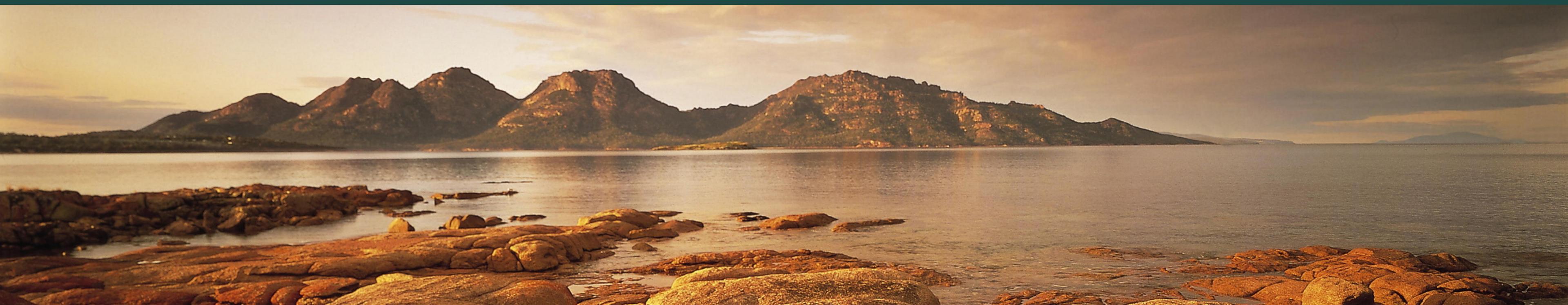
Does Nature Know Best? Analyzing Activity in Nature & Mental Health

Lisa Schreiber

Final Project | BDS-522-675 2021C

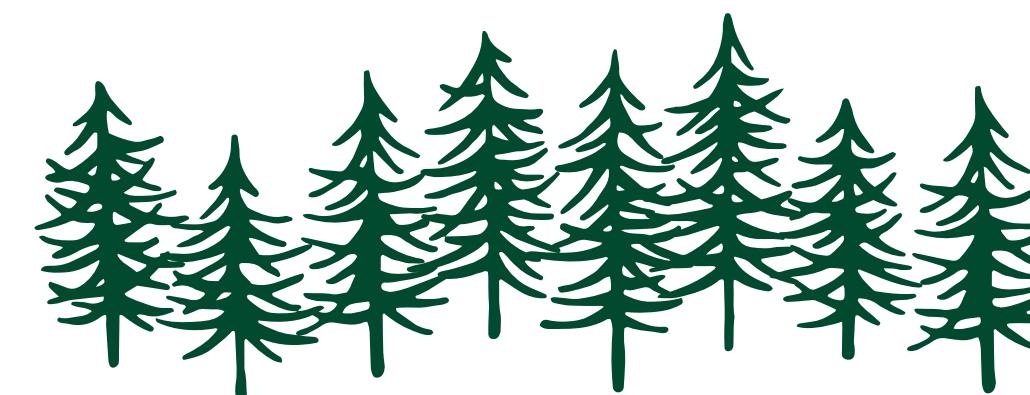
Hypothesis

Individuals who are active in nature are more likely to report better mental health.



Why Does This Matter?

In the last year, many people ventured outdoors for walks in the park or hikes in national parks as a respite from the pandemic¹. Even prior to the pandemic, research has shown nature has benefits for mental wellbeing². A pre-Covid study found that walking 90 minutes in nature could lead to a reduced risk in depression³. Another study conducted during the Covid-19 pandemic concluded that reduced park use underscored the importance of park-based recreation in promoting emotional health for US college students⁴. If there is a significant association between the two variables, then outdoor activities could potentially be leveraged to improve mental health.



1. The New York Times. (2021, July 8). How crowded are America's National Parks? see for yourself. The New York Times. Retrieved December 21, 2021, from <https://www.nytimes.com/2021/07/08/travel/crowded-national-parks.html>

2. American Psychological Association. (2020, April 1). Nurtured by Nature. Monitor on Psychology. Retrieved December 21, 2021, from <https://www.apa.org/monitor/2020/04/nurtured-nature>

3. University of Stanford. (2015, June 30). Stanford researchers find mental health prescription: Nature. Stanford News. Retrieved December 21, 2021, from <https://news.stanford.edu/2015/06/30/hiking-mental-health-063015/>

4. Larson, L. R., Mullenbach, L. E., Browning, M. H. E. M., Rigolon, A., Thomsen, J., Metcalf, E. C., Reigner, N. P., Sharaievska, I., McAnirlin, O., D'Antonio, A., Cloutier, S., Helbich, M., & Labib, S. M. (2021, November 10). Greenspace and park use associated with less emotional distress among college students in the United States during the COVID-19 pandemic. Environmental Research. Retrieved December 21, 2021, from https://www.sciencedirect.com/science/article/pii/S0013935121016686?casa_token=HZ-XY76jqRwAAAAA%3AZtuQLp6m5UEI6wl5Pv7JmhN4utLcjKsIxed_pN4OF8j-1gpM6e-WFb2ycgdgpGN85dF_0S6lg

General Social Survey Data

The General Social Survey (GSS) measures opinions, attitude, and behaviors for adults 18+ in the United States through yearly survey responses.

The GSS panel has shifted from the traditional in-person interview to a web self-administered questionnaire for 2021 in light of the pandemic.



GSS 2021 Data

The GSS 2021 data was collected December 2020-May 2021 as a cross-section survey

Population

Individuals (P18+) who answered survey questions about activity in nature and mental health

Sample

1,765 US respondents (P18+) from the GSS survey

Methods

Chi-Squared Test

A Chi-Squared test was conducted to examine if there is a significant relationship between the **categorical variables related to activity in nature and mental health**

Three Logistic Regressions

Three logistic regression models were run to further investigate if there is an association between the variables.

1. A **logistic regression tested if activity in nature impacted mental health**, with mental health converted into a binary variable.
2. Further analysis **converted activity in nature into a continuous variable (appendix)**, re-running it with the binary mental health.
3. **Two confounding variables** (physical health and income) were added to the model with activnat as categorical variable.

VARIABLES

HLTHMNTL: "In general, how would you rate your mental health, including your mood and your ability to think?"

- Scale 1-5 (1 = excellent, 5 = poor)
- Converted to binary for logistic regression (1-3 grouped as good, 4-5 as bad)

ACTIVNAT: "In the last twelve months how often, if at all, have you engaged in any leisure activities outside in nature, such as hiking, bird watching, swimming, skiing, other outdoor activities or just relaxing?"

- Scale 1-5 (1 = daily, 5 = never)
- Converted to continuous variable based on frequency for logistic regression (see assumptions about & appendix)

Assumptions

Data Variable Assumptions

- The self-reported variable "activnat" representing activity in nature (on categorical scale of 1-5) is a normal distribution of active vs. passive outdoor activities.
 - Since activity in nature does not indicate how long individuals are in nature or what activity they undertook, all activity duration in nature was not considered.
 - The "activnat" was converted into a continuous variable for analysis. It was assumed "several" indicated 4 times. E.g. "*several times a week*" would indicate a person was active in nature about 4 times that week.
- Variable "mntlhlt" was self-reported on a scale of 1-5, which was converted into a binary variable by pairing "poor" & "fair" as bad mental health vs. the rest as good mental health.
- Variable "hlthphys" was converted into a binary scale of bad health and good health (from a 1-5 ordinal scale) with "poor" & "fair" being labelled as bad health.
- Variable "income16" is categorical variable that functions as proxy for respondent's income and affluence, this was converted to a continuous variable by taking the mean of each category.

Chi-Squared and Logistic Regression Assumptions

- GSS data is representative of US population and sample size is sufficient
- Assume a normal data distribution, no extreme outliers
- Chi-Squared: variables are categorical & independent, and respondents' answers are mutually exclusive
- Logistic Regression: dependent variable is binary & ordinal, observations are independent from each other

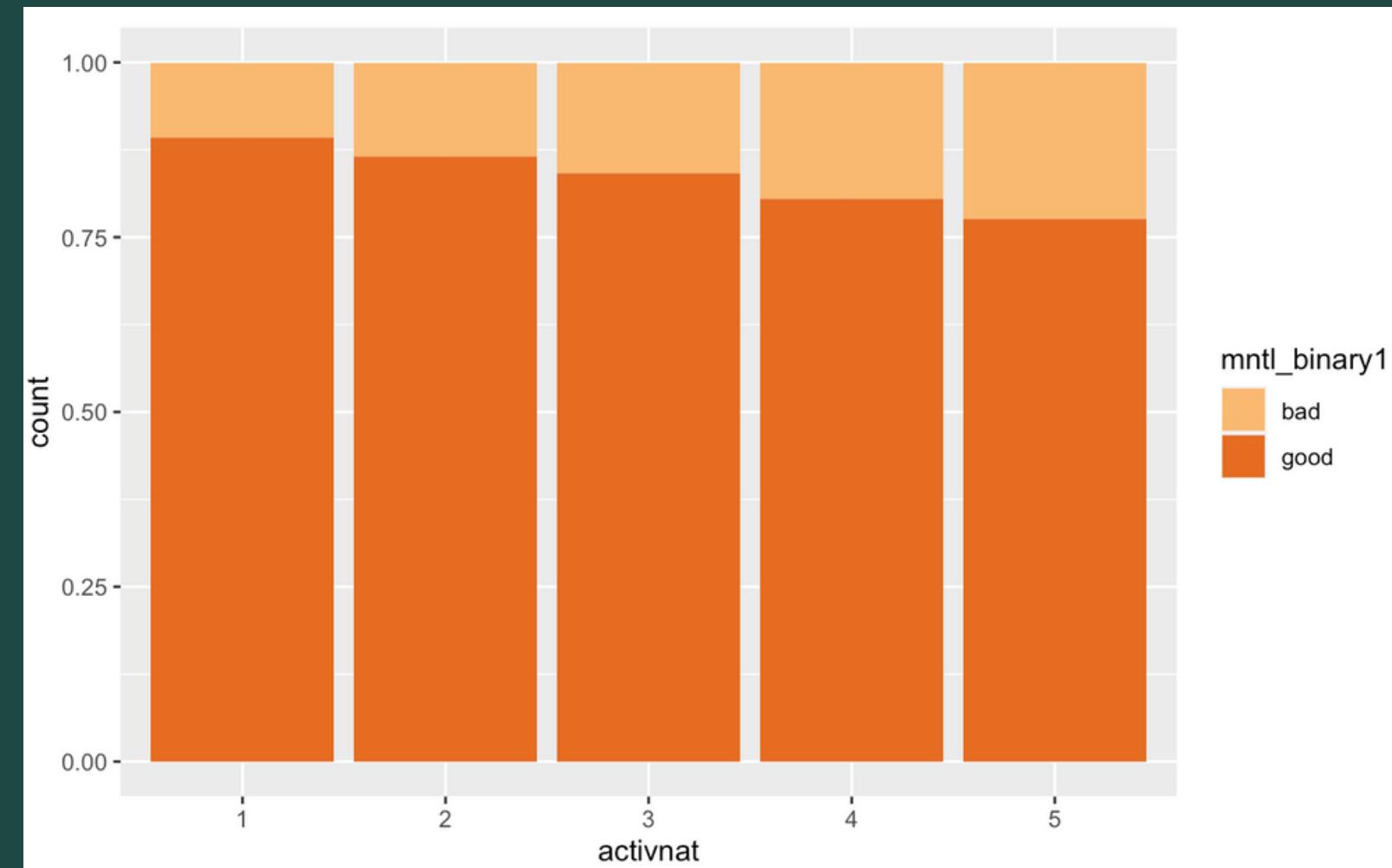


Initial Findings

Chi-Squared Test: A relationship between the categorical variables activity in nature in the past year and mental health (1 = excellent, 5 = poor) was found to be highly significant ($p=0.00001308037$).

Proportional Stacked Bar Chart: The less active in nature one is (e.g. 5 = never), the incrementally higher proportion of “bad” mental health was recorded. However, notably at least over 75% for each category of active in nature reported “good” mental health.

Proportional Stacked Bar Chart
activnat & mntl_binary1



Caveat: for activnat, category 5's sample ("never") is 73.9% smaller than the next category (cat. 1) or rather 85 vs. 326 respondents.

Findings: Logistic Regression #1

After testing for a significant association with a logistic regression, activity in nature was found to be significantly associated with mental health for its categories 1, 3, 4, and 5. There was no significance for category 2 (several times a week) in activnat.

This model suggests that the more outdoor activity a person reports (excluding "several times a week"), the higher likelihood that the person reports good mental health. The probability of reporting good mental health for someone that is active in nature daily (cat. 1) is 89.3%, whereas for someone that is active in nature several times a year (cat. 4) is 80.5% and never (cat. 5) is 77.6%. **It could be argued that these gaps are not substantive, with the most active category being only +15% larger than the least active.**

```
Call:  
glm(formula = mntl_binary ~ factor(activnat), family = binomial,  
     data = data_recode)  
  
Deviance Residuals:  
    Min      1Q  Median      3Q     Max  
-2.1126  0.4766  0.5385  0.5869  0.7113  
  
Coefficients:  
              Estimate Std. Error z value Pr(>|z|)  
(Intercept)   2.1180   0.1789 11.839 < 2e-16 ***  
factor(activnat)2 -0.2604   0.2197 -1.185 0.23609  
factor(activnat)3 -0.4465   0.2187 -2.041 0.04122 *  
factor(activnat)4 -0.6996   0.2236 -3.128 0.00176 **  
factor(activnat)5 -0.8728   0.3159 -2.763 0.00573 **  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
Null deviance: 1506.8 on 1764 degrees of freedom  
Residual deviance: 1492.2 on 1760 degrees of freedom  
AIC: 1502.2  
  
Number of Fisher Scoring iterations: 4
```

Logistic Regression #2 (appendix): activnat was converted to continuous variable. It was found that activity in nature was significantly associated with mental health ($p= 0.000487$), suggesting that with each additional day spent doing an activity in nature, the likelihood of reporting good mental health increases.

Confounders

HLTHPHYS: "In general, how would you rate your physical health?"

- Scale 1-5 (1 = excellent, 5 = poor)
- Converted to binary for logistic regression (1-3 grouped as good, 4-5 as bad)



Often mental health and physical health go hand in hand. Additionally, a person in good physical health may engage in more strenuous outdoor activities such as hiking or swimming.

INCOME16: "In which of these groups did your total family income, from all sources, fall last year before taxes, that is."

- Income Scale 1-26, converted into a continuous variable for analysis



Someone from a higher family income may tend to have fewer problems and better mental health. Higher income also allow someone to live in a more expensive area closer to nature or the ability to afford to visit multiple national parks across the US.

Findings: Logistic Regression #3 with Confounders

The association between being active in nature and mental health is insignificant in this model.

Accounting for confounders, this model indicates that good physical health provides a highly significant p-value, showing that good physical health has an exceedingly significant and substantive association with good mental health.

Income is also significantly associated with mental health.

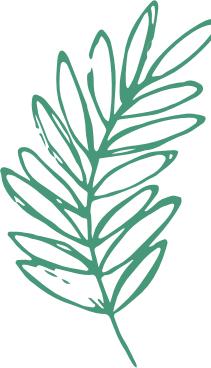
Someone active daily in nature but with bad physical health has a 59.3% probability of reporting good mental health when this person earns \$40k, yet 70.6% when earning \$100k. All other variables the same, the gap is less substantive when the incomes are closer (e.g. \$50k, 61.3% vs. \$60k, 63.2%).

All activnat categories were insignificant in this model.

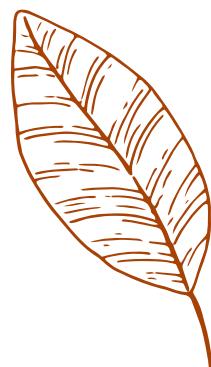
Note: The first logistic regression's AIC value has decreased from 1502.2 to 1121.6 when controlling for confounders here. This 380.6 value decline suggests that this model is better fitted relative to the model without confounders. **Likelihood Ratio Test in Appendix.**

```
Call:  
glm(formula = mntl_binary ~ factor(activnat) + factor(phys_binary) +  
     income, family = binomial, data = data_confounder)  
  
Deviance Residuals:  
    Min      1Q  Median      3Q      Max  
-2.7379  0.2612  0.3938  0.4839  1.3248  
  
Coefficients:  
              Estimate Std. Error z value Pr(>|z|)  
(Intercept) 0.042762183 0.239600460 0.178 0.858  
factor(activnat)2 -0.200270706 0.253961705 -0.789 0.430  
factor(activnat)3 -0.362870271 0.253641163 -1.431 0.153  
factor(activnat)4 -0.420111723 0.260097550 -1.615 0.106  
factor(activnat)5 -0.247759494 0.390021600 -0.635 0.525  
factor(phys_binary)1 2.017554933 0.159551612 12.645 < 0.00000000000002 ***  
income          0.000008319 0.000001673 4.974 0.00000656 ***  
---  
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
Null deviance: 1358.7 on 1586 degrees of freedom  
Residual deviance: 1107.6 on 1580 degrees of freedom  
(13 observations deleted due to missingness)  
AIC: 1121.6  
  
Number of Fisher Scoring iterations: 5
```

*Interaction coefficients were tested and found to be insignificant so there are not included in this model.



Conclusions

- 
- When considering the association between only activity in nature over the year and mental health, the model implies that more activity or each additional day spent in nature, makes it more probable that someone would report good mental health, but this difference is not particularly noteworthy.
 - After adding in physical health and income, an individual's physical health is strongly related to mental health, and activity in nature's relationship to mental health is not. Thus, Individuals in good physical health were more likely to report good mental health. A considerable difference in income makes a difference in the likelihood of reporting good mental health, yet when incomes are similar the gap is not considerable.
 - While outdoor activities are sometimes physical, it is not a replacement for physical health. Outdoor activities include sedentary bird-watching, while physical health can be maintained at the gym. Additional research would be required to fully understand how the activity in nature is recorded (e.g. taking into consideration the respondent's interpretation of the question to ensure that it is truly separate from physical health).
- 

Limitations

- One does not know how long someone spent in nature for each reported instance. The number of hours spent in nature would be a more interesting variable to explore than activnat.
- Regional 2021 data for was not available this year. To further examine these variables, adding location could enhance the model as location may determine people's accessibility to nature.
- For activity in nature and mental health, the more extreme categories had substantively fewer data points than other categories (e.g. only 85 participants were never active in nature). An increased sample size would provide more confidence in the results.
- Activnat's conversion to continuous numbers were incremental numbers based on the categories. The future analysis would be improved with a true continuous scale.
- 2021 may prove to be an outlier year in how people rated their mental health as there is an uptick in mental health issues during the pandemic⁵. A time-series analysis including pre-pandemic responses might reveal different result and significant associations.
- Variable activnat does not specify all activities in nature. For instance, bird watching and swimming are both considered activities in nature. Further research would benefit from segmenting the types of activities in nature (e.g. active vs. passive activities) to see if a type of activity in nature is acting as a proxy for fitness levels and mental health or if it is truly being in nature.
- The mental health variable is limited by the binary grouping created for the logistic regression and does not fully capture nuances, such as depression or anxiety which should be considered.



5. Parker-Pope, T., Caron, C., & Sancho, M. C. (2021, December 16). Why 1,320 therapists are worried about mental health in America right now. The New York Times. Retrieved December 23, 2021, from <https://www.nytimes.com/interactive/2021/12/16/well/mental-health-crisis-america-covid.html>

Summary

Hypothesis: Increased activity in nature throughout the year tends to improve people's mental health.

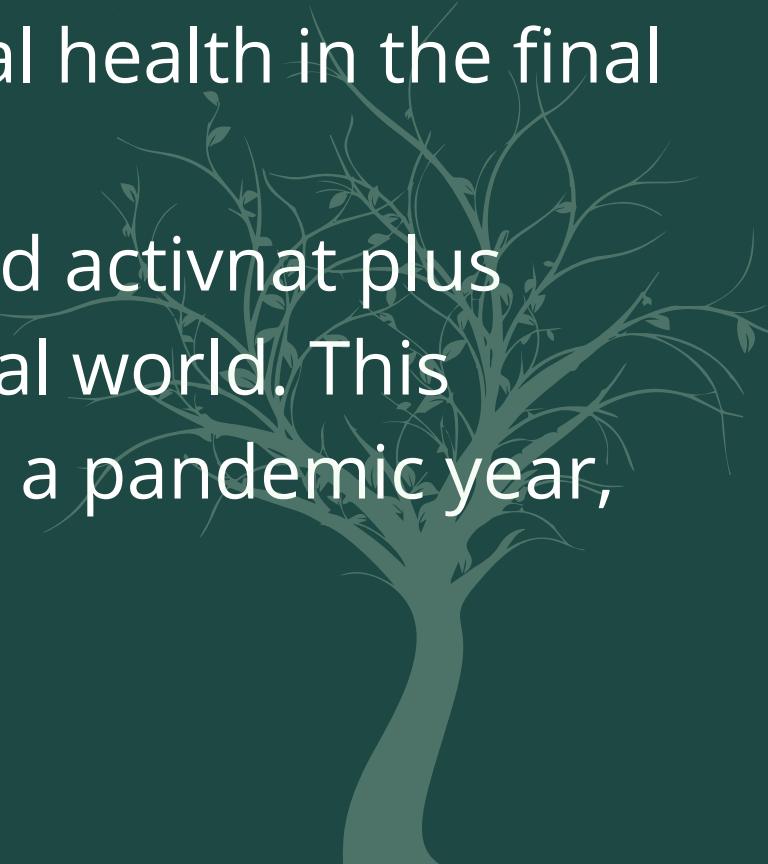
Sample: 1,765 US respondents (P18+) from the GSS survey

Methods: A Chi-Squared test checked for a significant association between variables, then three logistic regression analyses to further explore significance and extent of the relationship.

Findings: Activity in nature and mental health had a significant association in a chi-squared test and logistic regression. However, when controlling for confounders physical health and income, activity in nature and mental health's correlation was insignificant.

Conclusion: Although there was significance in the initial findings, the differences were not substantive and physical health as well as income proved to be significantly associated with mental health in the final model, while activity in nature was insignificant.

Limitations: Small sample size for extreme categories, assumed binary groupings, and activnat plus income16's categorical conversion to continuous made it less representative of the real world. This model could also be further improved by adding a time series analysis to look beyond a pandemic year, time spent in nature, and regional data to control for accessibility to nature.



Appendix.

- Likelihood Ratio Test
- Findings: Logistic Regression (continuous activnat)
- Original Variable Categorization



Likelihood Ratio Test

A likelihood ratio test compared logistic regression #1 (model without confounders) and logistic regression #3 model, producing a highly significant p-value (below).

Since p-value is less than .05, we reject the null hypothesis that there is no difference in the models. This result suggests the model with cofounding variables (logistic regression #3) provides a significantly improved fit over the model the one without them.

Likelihood ratio test

```
Model 1: mntl_binary ~ factor(activnat) + phys_binary + income
Model 2: mntl_binary ~ factor(activnat)
#Df LogLik Df Chisq      Pr(>Chisq)
1   8 -468.82
2   6 -622.98 -2 308.31 < 0.000000000000022 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Findings: Logistic Regression #2

After testing for a significant association in using logistic regression, it was found that activity in nature (continuous variable) had a highly significant p-value of 0.000487, indicating that activity in nature is significantly associated with mental health in this model.

This model suggests that with each additional day spent doing an activity in nature, the likelihood of reporting good mental health increases.

```
Call:  
glm(formula = mntl_binary ~ nature_cont, family = binomial, data = data_nature)  
  
Deviance Residuals:  
    Min      1Q  Median      3Q     Max  
-2.1332  0.4657  0.5433  0.6161  0.6421  
  
Coefficients:  
            Estimate Std. Error z value      Pr(>|z|)  
(Intercept) 1.4743232  0.0918602 16.050 < 0.000000000000002 ***  
nature_cont  0.0018971  0.0005439   3.488      0.000487 ***  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
Null deviance: 1506.8 on 1764 degrees of freedom  
Residual deviance: 1494.0 on 1763 degrees of freedom  
AIC: 1498  
  
Number of Fisher Scoring iterations: 4
```

Original Variable Categorization

Activity in Nature

Variable: ACTIVNAT Type: Numeric				
Label: In the last twelve months how often, if at all, have you engaged in any leisure activities outside in nature, such as hiking, bird watching, swimming, skiing, other outdoor activities or just relaxing?				
Notes:				
PCT Excl. Reserve Codes				
LABEL	VALUE	COUNT	PCT	
DAILY	1	328	8.1%	18.5%
SEVERAL TIMES A WEEK	2	530	13.1%	29.9%
SEVERAL TIMES A MONTH	3	475	11.8%	26.8%
SEVERAL TIMES A YEAR	4	356	8.8%	20.1%
NEVER	5	85	2.1%	4.8%
SUBTOTALS:		1774	44.0%	100.0%
RESERVED CODES:				
CAN'T CHOOSE	D	52	1.3%	n/a
NO ANSWER	N	191	4.7%	n/a
NOT APPLICABLE	I	2003	49.7%	n/a
SKIPPED ON WEB	S	12	0.3%	n/a
TOTALS:		4032	100.0%	100.0%

Variable: HLTHMNTL Type: Numeric				
Label: In general, how would you rate your mental health, including your mood and your ability to think?				
Notes: Handcard F1 contained punches 1-5.				
PCT Excl. Reserve Codes				
LABEL	VALUE	COUNT	PCT	
EXCELLENT	1	579	14.4%	15.9%
VERY GOOD	2	1313	32.6%	36.1%
GOOD	3	1147	28.4%	31.5%
FAIR	4	476	11.8%	13.1%
POOR	5	122	3.0%	3.4%
SUBTOTALS:		3637	90.2%	100.0%
RESERVED CODES:				
NO ANSWER	N	1	0.0%	n/a
NOT APPLICABLE	I	380	9.4%	n/a
SKIPPED ON WEB	S	14	0.3%	n/a
TOTALS:		4032	100.0%	100.0%