Mental Health: Depression & Anxiety

ANALYSING DIRECT AND INDIRECT CAUSES

Surabhi Summi
Caitlin Salerno
Baishakhi Mitra
Nimesh Gala
USF

UNIVERSITY of SOUTH FLORIDA

Contents

Executive Summary

In recent times, the importance of mental health and its various far-reaching effects on people has become more well-known. Depression and anxiety are two most commonly experienced mental health issues that could arise at any point in life. In this project we analyze what factors reduce/control depression or anxiety by altering some of the basic aspects/habits or our lives or improving treatments for other health issues.

Our analysis is intended for anyone looking to decrease their anxiety and depression levels, but we are especially interested in reaching people with various ailments (such as diabetes, heart disease and cancer) and the doctors treating said ailments. Some of these diseases cause major distress in an individual, and through our analysis, certain actions can be recommended by doctors to patients with diseases that are especially prone to anxiety and depression.

Our data is obtained from the National Health Interview Survey in 2020. This survey results have been instrumental in providing data to track health status, health care access and progress towards achieving national health objectives. Our data set contains 618 columns related to the quality of life, practices, family structure and income, health issues and health insurance. We have ultimately chosen to analyze anxiety and depression levels with 21 carefully chosen predictor variables.

In our analysis we have considered aspects of quality of life and different lifestyle choices as determinants of depression and anxiety level in that individual. We have created linear regression models that will help us understand the effects of these factors which will in turn help in reducing or controlling depression and anxiety.

Problem Definition & Significance

Depression and anxiety are two of the most common mental illnesses. The National Institute of Mental Health has estimated that approximately 8.4 percent of the US population suffers from depression (“Major Depression”), and 19.1 percent of the US population suffers from an anxiety disorder (“Any Anxiety Disorder”). These mental health problems impair people’s work, careers, family and personal life. Depression and anxiety are quite different in their symptoms. People suffering from anxiety have symptoms such as fear, feelings of panic, restlessness, shortness of breath, and heart palpitations (“Anxiety Disorders”), whereas people in depression show feelings of guilt, worthlessness, helplessness, hopelessness, insomnia, and even physical pains (“Depression”). People going through severely stressful situations, especially including those battling other diseases, can have difficulty controlling their mental health.

While anxiety and depression are very common mental disorders, they are also treatable via several methods, including several lesser-known factors that are not as well known as medication and mental therapy. Through this project we aim to analyze the various factors leading to these medical conditions.

We will focus on actionable items that can be used to lesson the severity of these mental disorders, and also focus on health variables so that doctors can know to also prescribe mental health treatment (possibly involving

our actionable variables) along with other treatments for a disease that has a strong effect on developing anxiety or depression. Through our models, we aim to come to conclude to what extent these factors are responsible for these medical conditions and how we can reduce the risk.

Prior Literature

While researching papers and journals on mental health and depression, we came across various informative and interesting research paper published before. We read more than 8 papers which helped us find some interesting trends and patterns and fueled our project.

One of those papers focused on the "Psychosocial job quality and mental health among young workers" where it analyzed how young worker are more likely to get depressed in a workplace because they are more vulnerable to bullying and clashes with superior and other colleagues. Various important factors such as age, level of education, presence of long-term health condition, household structure, income etc. has been studied to analyzed depression. This paper worked with fixed-effects regression analysis using 13 waves of annual data to analyze depression. It helped us to understand depression as a problem and how bullying at workplace and health issues can affect people. Therefore, we included some long-term health issue variables in our model such as high cholesterol, heart disease, heart attack, stroke, asthma, cancer, diabetes, lung disease, defect in hearing, vision and arthritis.

Some of our papers did research on how Quarantine Experiences and Attitudes Towards COVID-19 Affect the Distribution of Mental Health. These research papers gave a fuller picture of psychological impacts of COVID-19 including how COVID-19 impacts quality of life in human beings. Here, an approach of quantile regression and multiple regression were used to analyze the potential multiformity in relationship to mental health in quarantine during COVID-19. These uses data from snowballing sampling technique and other survey technique to gather data from the users. This analysis found that home self-quarantine is more likely to cause decrease in depression while community level quarantine is more likely to increase depression. During these studies we came up with some more important variable to be considered such as age, exercise, family member and marital status. However, on our analysis we did not get enough responses pertaining to COVID and hence we could not consider it now and this could be a future area of analysis.

In the past researchers have emphasized on the importance of the quality of life of people and the different lifestyle choices they make on their mental health. We adhere to that and also agree that while the demographic aspects of an individual's life cannot be modified, we can focus on the choices and make recommendations on those so that depression and anxiety can be controlled before it starts to impact family and society at large.

Data Source/Preparation

We are taking data from the National Health Interview Survey (NHIS) from 2020. This data takes a sample of 31,568 individuals, asking them questions about all aspects of their health and socio-economic conditions. There are 618 variables in this dataset, and we have deemed 202 of these variables to be important in our analysis

towards factors that could affect anxiety and depression levels, our two target variables. We have excluded all the health insurance related variables as our analysis is focusing on that. Other variables in this study were either not important factors of our outcome variables or were heavily related to the 21 chosen variables and therefore, had too much multicollinearity to include.

Data Engineering Explanation and Description of Variables:

DEP_LEVEL and ANX_LEVEL (variables created the same way) = Depression level and Anxiety Level, from 0-3. Taken from a combination of multiple variables. We used the original DEPLEVEL_A/ANXLEVEL_A in the dataset, which people typically only responded to if they had ever had depression/anxiety. For unknown depression/anxiety level values which answered 'no' to the question related to the "Ever had depression/anxiety?" variable (DEPEV_A/ANX_A), they were given a depression/anxiety level 0. For missing depression/anxiety levels past this point, if someone answered 'never' to the question related to the "How often depressed/anxious" variable (DEPFREQ_A/ANXFREQ_A), they were listed as depression/anxiety level 0. There were very few cases left that were still 0s for DEP_LEVEL/ANX_LEVEL, so we dropped those remaining rows.

Nicotine = Nicotine usage, categorical (Never, Formerly, Currently). We started with the SMKCIGST variable, giving information about cigarette smoking status. The multiple answers associated with current use were labeled as such, as were answers associated with former use and never used. Those who did not give a definitive answer were labeled as 'Unknown'. We then overlapped the nicotine variable with current status of e-cigarettes (SMKECIGT_A), where current users of e-cigarettes would have any status of nicotine now be 'Current' and any people who never smoked would have their status changed to 'Former' if they formerly smoked e-cigarettes. Additionally, for any Unknown values, the nicotine status changed to match the status of the e-cigarettes, if there was one. We repeated this process for the variables associated with the status of smokeless tobacco (SMOKELSCUR_A), pipes (PIPECUR_A), and cigars (CIGARCUR).

Sleep = (SLPHOURS_A) The number of hours of sleep the subject gets in a 24 hour period, a numeric variable. Those that gave answers related to unknowns (97=Refused, 98=Not Ascertained, 99=Don't Know) were taken out of the dataset.

Exercise = How active the subject is, a categorical variable (Not Active, Lightly Active, Moderately Active, Very Active, Not Physically Able). This is a combination of multiple variables related to exercise frequency per week, which are moderate exercise frequency (MODFREQ_A), vigorous exercise frequency (VIGFREQ_A) and strength-training exercise frequency (STRFREQ_A). Subjects that had a value of 0 (we changed to represent less than once per week or never) in all three variables were listed as 'Not Active'. Those who exercised 1-2 times a week (had a value of 1 or 2) in any category were listed as 'Lightly Active'. Those who exercised 3-6 times a week (had a value of 3-6) were listed as 'Moderately Active'. Those who exercised 7 times or more a week were listed as 'Very Active'. Any category representing more activity than before would override previous categories if the subject met the requirements for the higher category. If the subject had a value of -1 (we inserted this placeholder as not physically able) for all 3 variables, the exercise was listed as 'Not Physically Able". Lastly, if the subject had unknown values (placeholder was -2) for all three variables, the exercise was listed as "Unknown" and these rows were taken out of the dataset.

Alcohol = (DRKSTAT_A) The alcohol drinking status of the subject, a categorical variable (Never, Former, Infrequent, Light, Moderate, Heavy). Responses related to never used (1=Lifetime abstainer) were marked as such, as was former use (2, 3, 4), infrequent use (5) light use (6), moderate use (7, 9-frequency unknown-we are assuming moderate), and heavy use (8). Those who answered 10, which meant 'drinking status unknown', were deleted from the dataset.

BMI = (BMICAT_A) The BMI of the subject, categorical (Underweight, Healthy Weight, Overweight, Obese). The original BMI of the subjects is not known since the only bmi variable given was categorical. We changed the factors to what they represented to be more readable, since we cannot read it as ordinal. Those who answered 9 which is unknown were taken out of the dataset.

FoodSecurity = The food security of the subject, as an ordinal factor, taken from the FDSCAT4_A variable, which is a re-coding of multiple other variables in the survey data. The original data had the variables listed in descending order so we reversed the order to be ascending, where 1 is the lowest food security. We also took all unknown rows (8=Not Ascertained) out of the dataset.

For the following variables, which are binary; the original layout of all variables was 1=yes, 2=no, 7=refused, 8=Not Ascertained, 9=Don't know. We changed this into a binary format (0=no, 1=yes), deleting any rows with unknowns (7-9). These are the variables with this transformation:

Walk = (WALKLEIS_A) Whether or not someone walked for leisure in the past week, binary.
Weightloss = (NOWWGTPRG_A) Whether the subject is in a weight loss program, binary.
Hypertension = (HYPEV_A) If the subject has ever had hypertension.
High_Cholesterol = (CHLEV_A) If the subject has ever had high cholesterol.
Heart_Disease = (CHDEV_A) If the subject has ever had heart disease.
Angina = (ANGEV_A) If the subject has ever had angina.
Heart_Attack = (MIEV_A) If the subject has ever had a heart attack.
Stroke = (STREV_A) If the subject has ever had a stroke.
Asthma = (ASEV_A) If the subject has ever had asthma.
Cancer = (CANEV_A) If the subject has ever had cancer.
Diabetes = (DIBEV_A) If the subject has ever had diabetes.
Hospital_Overnight = (HOSPONGT_A) If the subject has been hospitalized overnight in the last 12 months.
Therapy_Not_Mental = (THERA12M_A) If the subject has received physical/speech/ rehabilitative/occupational therapy in the last 12 months (not mental therapy).
Lung_Disease = (COPDEV_A) If the subject has ever had COPD, emphysema or chronic bronchitis
Arthritis = (ARTHEV_A) If the subject has ever had arthritis.
Dementia = (DEMENEV_A) If the subject has ever had dementia.

Variable Choice

Predictor	Sign	Rational
Nicotine	+	Nicotine usage can cause anxiety or depression in terms of being dependent on a drug and needing it to feel normal
Walking	-	Walking for leisure can decrease stress and therefore decrease anxiety and depression levels.

Sleep	-/+	Getting more sleep can help anxiety levels. However, too much sleep can increase depression levels.
Exercise	- for active	Being moderately active and very active can cause endorphins to be produced naturally, decreasing stress, anxiety and depression
Weight Loss	-	Being in a weight loss program means that someone will be eating healthier and less junk food, so this can help decrease anxiety and depression levels.
Alcohol	+	More frequent drinkers are dependent on alcohol or use it as an unhealthy crutch, which can lead to higher anxiety and depression levels
BMI	+	Higher bmi can lead to lower body positivity along with an unhealthy lifestyle, increasing anxiety and depression levels.
Food Security	-	People who have higher food security are less likely to be in financial distress and are also more likely to be regularly eating, which will decrease anxiety and depression levels.
Hypertension	+	Having hypertension can add extra stress in a person's life leading to anxiety and depression levels increasing.
High Cholesterol	+	Having high cholesterol can add extra stress in a person's life (they have to go on a specific diet) leading to anxiety and depression levels increasing.
Heart Disease	+	Having heart disease can add extra stress in a person's life leading to anxiety and depression levels increasing.
Heart Attack	+	Having a heart attack can add extra stress in a person's life leading to anxiety and depression levels increasing.
Stroke	+	Having a stroke can add extra stress in a person's life leading to anxiety and depression levels increasing.
Asthma	+	Having asthma can prevent someone from doing activities that would help reduce stress (being outdoors, vigorous exercise) leading to higher anxiety and depression levels.
Cancer	+	Having cancer would add lots of extra stress into someone's life lading to anxiety and depression levels increasing.
Diabetes	+	Having diabetes adds stress and also controls someone's diet and lifestyle, so this can lead to higher anxiety and depression levels.
Hospital Overnight	+	Being hospitalized overnight in the last 12 months indicates serious illness or injury and also indicates a very high medical bill, increasing anxiety and depression levels.
Therapy, Not Mental	-	Having therapy (not mental) in a different area of life can help decrease mental stress and therefore decrease anxiety and depression levels.
Lung Disease	+	Having lung disease can increase someone's level of stress and therefore increase someone's level of depression and anxiety.
Arthritis	+	Having arthritis can limit someone's life, adding stress and increasing anxiety and depression levels.
Dementia	+	Having dementia can increase anxiety and depression levels due to the coming to terms of this uncurable and mostly untreatable disease.

Descriptive Analysis and Data Visualization

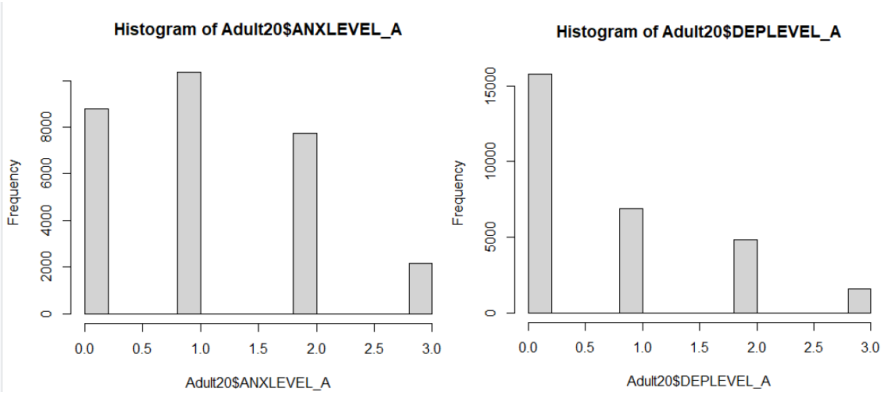


Fig 1.1

Fig 1.1 shows histogram of the dependent variables (anxiety and depression levels). Due to the variables being ordinal factors, this distribution is more generalized, but overall it does seem that depression is skewed to the right. The anxiety variable looks relatively normal.

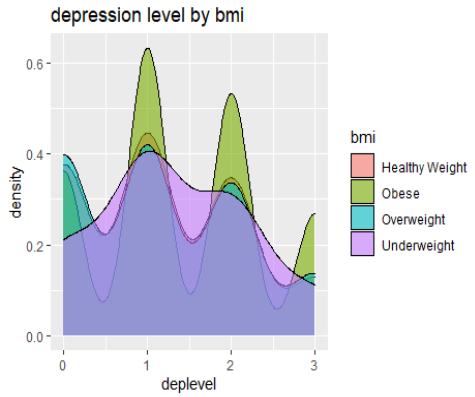


Fig 2.1

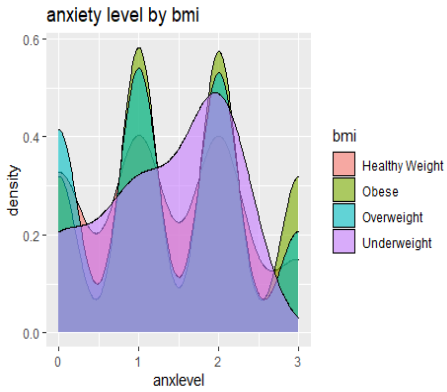


Fig 2.2

Fig2.1 and Fig2.2 depicts the relation between body mass index (BMI) and depression and anxiety.
Obese people are most likely to suffer from depression followed by overweight and underweight people whereas depression is noticed in less people with healthy BMI.
Anxiety is noticed a lot in obese and overweight people and lesser in underweight people. People with healthy BMI do not suffer with anxiety nearly as much.

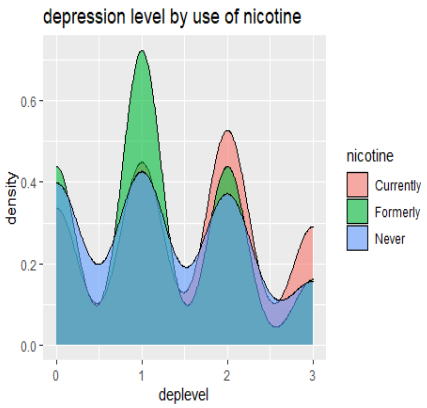


Fig3.1

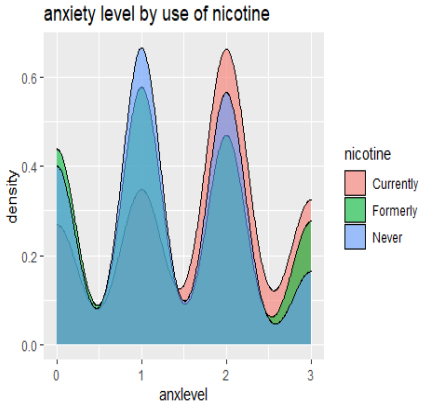


Fig3.2

Fig 3.1 and 3.2 depicts the relation between nicotine use and depression and anxiety.
We notice that people who smocked in the past has lower depression level whereas current smokers have higher levels of depression.
Higher levels of anxiety is noticed in current smokers and anxiety level is low in people who smoked in the past or have never smoked.

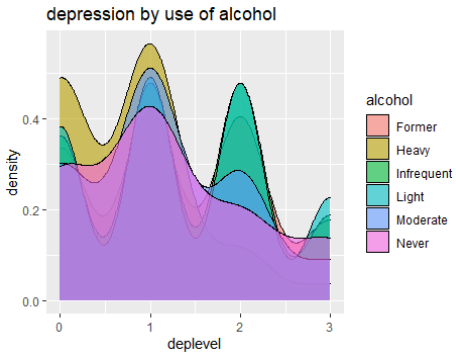


Fig 4.1

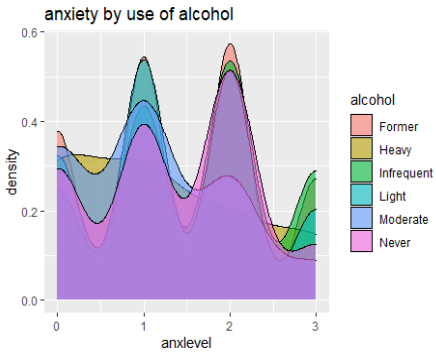


Fig4.2

Fig 4.1 and 4.2 depicts how alcohol consumption affects levels of depression and anxiety.
We see here that people who consume alcohol infrequently have higher levels of depression as compared to heavy consumer of alcohol. Depression level however is lower in people who never had alcohol.
Anxiety level in lower in infrequent alcohol consumers and higher in people who never had alcohol as well as in people who consumed alcohol in the past.

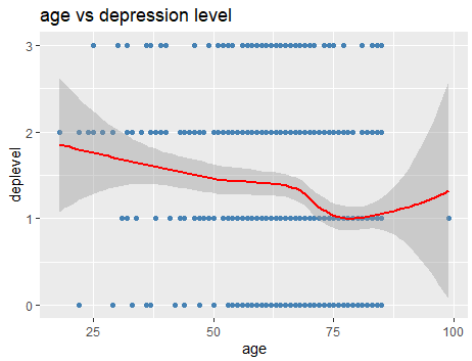


Fig5.1

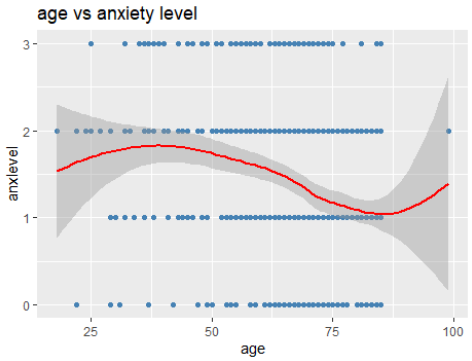


Fig5.2

Fig5.1 and 5.2 shows how depression and anxiety levels can alter with age.
Depression level lowers gradually from 18 years till around 72 years and then it increases as age increase.
Anxiety level slowly increases from 18 years and peaks around 45 years and decreases till a person reaches 75 years, After that we notice anxiety level to increase again.

Models

Actionable Variable Models

```
m_AnxietyAction=lm(ANXLEVEL_A~relevel(nicotine,ref="Never")+walk+sleep+relevel(Exercise,ref="NotActive")+
weightloss+relevel(alcohol,ref="Never")+bmi+FoodSecurity, data=Adult20Actionable)
```

```
m_DepressionAction=lm(DEPLEVEL_A~relevel(nicotine,ref="Never")+walk+sleep+relevel(Exercise,ref="NotActive")+
weightloss+relevel(alcohol,ref="Never")+bmi+FoodSecurity, data=Adult20Actionable)
```

	Dependent variable:	
	ANXLEVEL_A (1)	DEPLEVEL_A (2)
relevel(nicotine, ref = "Never")Currently	0.139*** (0.015)	0.209*** (0.015)
relevel(nicotine, ref = "Never")Formerly	0.076*** (0.013)	0.119*** (0.013)
walk1	0.070*** (0.012)	0.014 (0.012)
sleep	-0.042*** (0.004)	-0.020*** (0.004)
relevel(Exercise, ref = "Not Active")Lightly Active	0.042** (0.017)	-0.017 (0.017)
relevel(Exercise, ref = "Not Active")Moderately Active	0.014 (0.015)	-0.060*** (0.015)
relevel(Exercise, ref = "Not Active")Not Physically Able	0.350*** (0.080)	0.407*** (0.079)
relevel(Exercise, ref = "Not Active")Very active	-0.080*** (0.016)	-0.120*** (0.016)
weightloss1	0.137*** (0.023)	0.117*** (0.023)
relevel(alcohol, ref = "Never")Former	0.177*** (0.021)	0.160*** (0.021)
relevel(alcohol, ref = "Never")Heavy	0.359*** (0.027)	0.337*** (0.027)
relevel(alcohol, ref = "Never")Infrequent	0.251*** (0.022)	0.199*** (0.022)
relevel(alcohol, ref = "Never")Light	0.245*** (0.019)	0.175*** (0.019)
relevel(alcohol, ref = "Never")Moderate	0.205*** (0.021)	0.142*** (0.021)
bmiObese	-0.021 (0.014)	0.050*** (0.014)
bmiOverweight	-0.092*** (0.013)	-0.065*** (0.013)
bmiUnderweight	0.087* (0.046)	0.090** (0.045)
FoodSecurity	-0.226*** (0.009)	-0.258*** (0.009)
Constant	2.010*** (0.047)	1.670*** (0.047)
Observations	29,028	29,028
R2	0.050	0.063
Adjusted R2	0.050	0.062
Residual Std. Error (df = 29009)	0.901	0.895
F Statistic (df = 18; 29009)	85.363***	107.910***
Note:		*p<0.1; **p<0.05; ***p<0.01

Disease Related Models

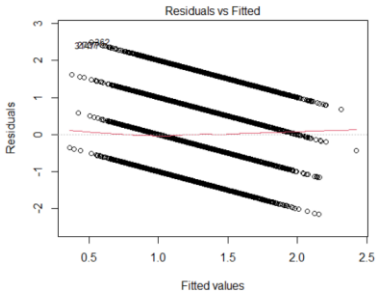
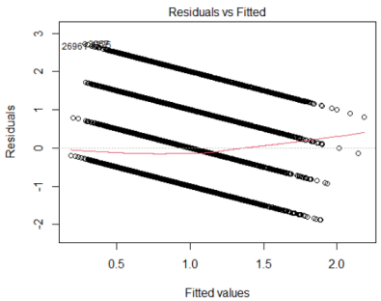
```
m_AnxietyHealth=lm(ANXLEVEL_A~HYPERTENSION+HIGH_CHOLESTEROL+HEART_DISEASE+ANGINA+HEART_ATTACK
+STROKE+ASTHMA+CANCER+DIABETES+LUNG_DISEASE+ARTHRITIS+DEMENTIA+HOSPITAL_OVERNIGHT+
THERAPY_NOT_MENTAL, data=Adult20Health)
```

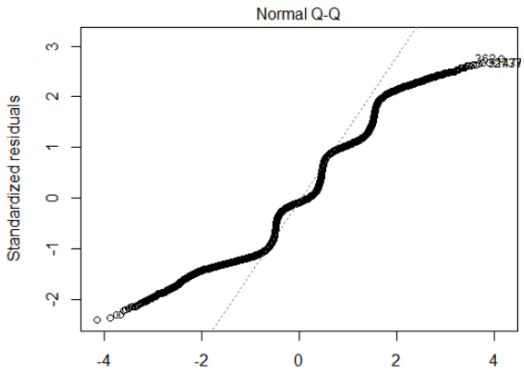
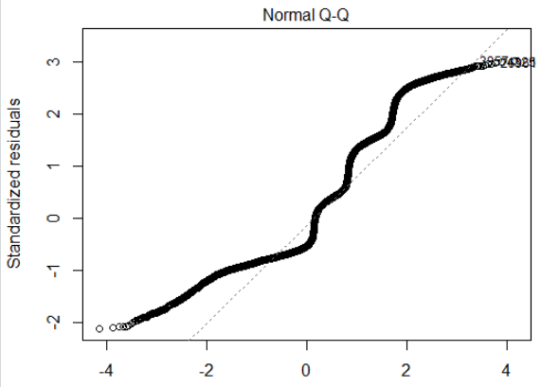
```
m_DepressionHealth=lm(DEPLEVEL_A~HYPERTENSION+HIGH_CHOLESTEROL+HEART_DISEASE+ANGINA+
HEART_ATTACK+STROKE+ASTHMA+CANCER+DIABETES+LUNG_DISEASE+ARTHRITIS+DEMENTIA+
HOSPITAL_OVERNIGHT+THERAPY_NOT_MENTAL, data=Adult20Health)
```

```
================================================================
                                Dependent variable:
                            ---------------------------------
                                ANXLEVEL_A        DEPLEVEL_A
                                    (1)               (2)
----------------------------------------------------------------
HYPERTENSION                  -0.051*** (0.013)  -0.014 (0.013)
HIGH_CHOLESTEROL               -0.003 (0.013)    0.025* (0.013)
HEART_DISEASE                 -0.156*** (0.028) -0.140*** (0.028)
ANGINA                        0.137*** (0.044)  0.157*** (0.043)
HEART_ATTACK                   -0.027 (0.034)     0.030 (0.034)
STROKE                         -0.023 (0.032)    0.055* (0.032)
ASTHMA                        0.236*** (0.016)  0.256*** (0.016)
CANCER                        -0.071*** (0.017) -0.046*** (0.016)
DIABETES                      -0.050*** (0.019)   0.013 (0.019)
LUNG_DISEASE                  0.153*** (0.025)  0.211*** (0.025)
ARTHRITIS                     0.109*** (0.013)  0.160*** (0.013)
DEMENTIA                      0.228*** (0.060)  0.235*** (0.060)
HOSPITAL_OVERNIGHT            0.076*** (0.020)  0.099*** (0.020)
THERAPY_NOT_MENTAL            0.149*** (0.017)  0.145*** (0.017)
Constant                      1.058*** (0.008)  0.618*** (0.008)
----------------------------------------------------------------
Observations                      28,784            28,784
R2                                 0.021             0.031
Adjusted R2                        0.021             0.031
Residual Std. Error (df = 28769)   0.914             0.909
F Statistic (df = 14; 28769)     44.785***         66.737***
================================================================
Note:                                *p<0.1; **p<0.05; ***p<0.01
```

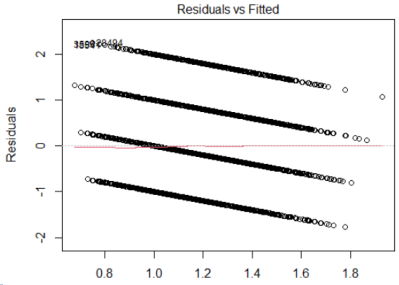
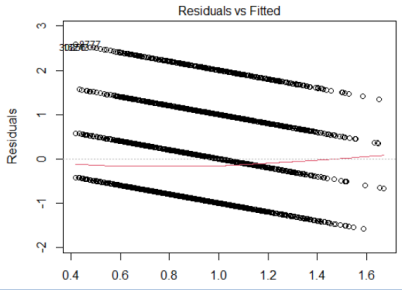
Quality Checks

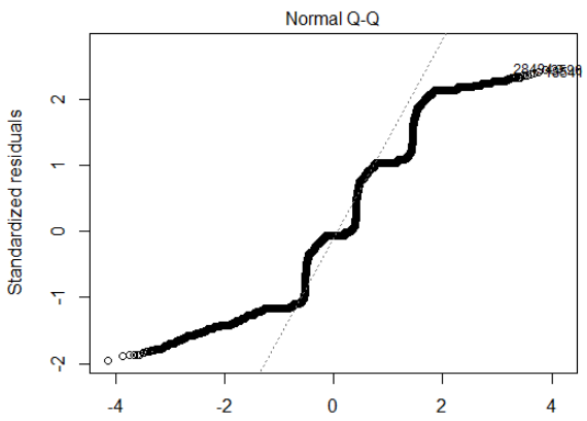
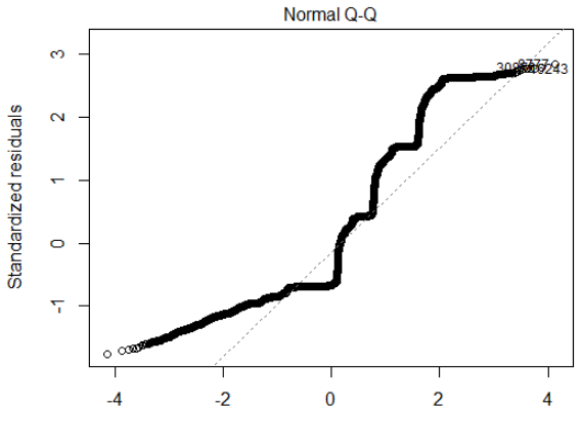
Actionable Variable Models

Assumption	m_AnxietyAction	m_DepressionAction
Linearity	Residual plot slows downward linear bias This data doesn't seem to be true Gaussian. 	Residual plot slows downward linear bias This data doesn't seem to be true Gaussian. 

Normality	The Q-Q plot shows the data is relatively normal, but there are definitely some limitations showing from being an ordinal factor treated as a continuous integer. Pass? 	The Q-Q plot shows the data is relatively normal, but there are definitely some limitations showing from being an ordinal factor treated as a continuous integer. Pass? 
Homoskedasticity	Pass; See residual plot above. Variance remains the same throughout all distributions.	Pass; See residual plot above. Variance remains the same throughout all distributions.
Multicollinearity	Passed `> vif(m_AnxietyAction)` GVIF Df GVIF^(1/(2*Df)) relevel(nicotine, ref = "Never") 1.135095 2 1.032186 walk 1.173487 1 1.083276 sleep 1.013570 1 1.006762 relevel(Exercise, ref = "Not Active") 1.235195 4 1.026755 weightloss 1.026442 1 1.013135 relevel(alcohol, ref = "Never") 1.170337 5 1.015854 bmi 1.074059 3 1.011979 FoodSecurity 1.047342 1 1.023397	Passed `> vif(m_DepressionAction)` GVIF Df GVIF^(1/(2*Df)) relevel(nicotine, ref = "Never") 1.135095 2 1.032186 walk 1.173487 1 1.083276 sleep 1.013570 1 1.006762 relevel(Exercise, ref = "Not Active") 1.235195 4 1.026755 weightloss 1.026442 1 1.013135 relevel(alcohol, ref = "Never") 1.170337 5 1.015854 bmi 1.074059 3 1.011979 FoodSecurity 1.047342 1 1.023397
Independence	Durbin-Watson test: Passed DW = 1.9524, p-value = 2.495e-05	Durbin-Watson test: Passed DW = 1.9422, p-value = 4.14e-07

Disease Related Models

Assumption	m_AnxietyHealth	m_DepressionHealth
Linearity	Residual plot slows downward linear bias This data doesn't seem to be true Gaussian. 	Residual plot slows downward linear bias This data doesn't seem to be true Gaussian. 
Normality	The Q-Q plot shows the data is relatively normal, but there are definitely some limitations showing from being an ordinal factor treated as a continuous integer. Pass?	The Q-Q plot shows the data is relatively normal, but there are definitely some limitations showing from being an ordinal factor treated as a continuous integer. Pass?

		
Homoskedasticity	Pass; See residual plot above. Variance remains the same throughout all distributions.	Pass; See residual plot above. Variance remains the same throughout all distributions.
Multicollinearity	Passed `> vif(m_AnxietyHealth)` HYPERTENSION 1.261987 HIGH_CHOLESTEROL 1.232975 STROKE 1.062196 ASTHMA 1.035412 ARTHRITIS 1.198364 DEMENTIA 1.012022 HEART_DISEASE 1.476039 ANGINA 1.157413 CANCER 1.061722 DIABETES 1.122891 HOSPITAL_OVERNIGHT 1.090108 THERAPY_NOT_MENTAL 1.082510 HEART_ATTACK 1.354775 LUNG_DISEASE 1.098403	Passed `> vif(m_DepressionHealth)` HYPERTENSION 1.261987 HIGH_CHOLESTEROL 1.232975 STROKE 1.062196 ASTHMA 1.035412 ARTHRITIS 1.198364 DEMENTIA 1.012022 HEART_DISEASE 1.476039 ANGINA 1.157413 CANCER 1.061722 DIABETES 1.122891 HOSPITAL_OVERNIGHT 1.090108 THERAPY_NOT_MENTAL 1.082510 HEART_ATTACK 1.354775 LUNG_DISEASE 1.098403
Independence	Durbin-Watson test: Passed DW = 1.9545, p-value = 5.633e-05	Durbin-Watson test: Passed DW = 1.9436, p-value = 8.546e-07

Recommendations

Actionable Variables

- Smoking, whether currently or formerly, increases a subject's expected anxiety and depression levels. People wanting to have better mental health should never smoke, and if they are smoking, former smokers don't increase in the anxiety and depression levels as much as current smokers do – so quitting can help mental health in the long run.
- Walking for leisure actually tends to increase anxiety and depression levels, which is interesting but goes against most health professionals so we will not use this for recommendations.
- Being lightly or moderately active actually seems to give a slight increase in anxiety levels. Only being very active seems to decrease it. On the other hand, any amount of physical activity decreases depression levels. If one is suffering from depression, he or she should try working out more.

- Those in weight loss programs have higher levels of anxiety and depression. This is most likely due to the stress of lower caloric intake on the body. If one is suffering from anxiety and depression already, signing up for a weight loss program might not be a good idea.
- Any amount of alcohol use increases anxiety and depression levels, including former use. Heavy alcohol drinkers should reduce their usage or better yet, stop drinking alcohol altogether to decrease anxiety and depression levels.
- People who are underweight have higher levels of anxiety and depression, while people who are obese have higher levels of only depression. Overall, having a healthy bmi is important for mental health.
- Having higher food security decreases expected anxiety and depression levels, so those with low food security should seek out food pantries and charities to avoid mental health issues.

Disease Related Variables

- Angina, Asthma, Lung Disease, Arthritis and Dementia are all diseases that increase anxiety levels. If a doctor is treating a patient for these diseases, he or she should also recommend mental health treatment to the patient, or changing their habits based on the actionable variables mentioned above.
- High Cholesterol, Angina, Stroke, Asthma, Diabetes, Lung Disease, Arthritis, and Dementia are all diseases or conditions that increase depression levels. If a doctor is treating a patient for these diseases, he or she should also recommend mental health treatment to the patient, or changing their habits based on the actionable variables mentioned above.
- Staying in a hospital overnight in the last year increases both anxiety and depression levels. For those staying in a hospital overnight, perhaps having a mental health expert visit them in their treatments might be beneficial.
- Having therapy that is not mental therapy also increases aniety and depression levels. This could be related to other variables, but nevertheless, if someone has any type of therapy, perhaps he or she should also have mental therapy to have their mind as well as their body heal.

References:

1) Rayner L, Hotopf M, Petkova H, Matcham F, Simpson A, McCracken LM. Depression in patients with chronic pain attending a specialised pain treatment centre: prevalence and impact on health care costs. Pain. 2016 Jul;157(7):1472-9. doi: 10.1097/j.pain.0000000000000542. PMID: 26963849; PMCID: PMC4912238.

2) Cruz, Luciane & Fleck, Marcelo & Polanczyk, Carisi. (2009). Depression as a determinant of quality of life in patients with chronic disease: Data from Brazil. Social psychiatry and psychiatric epidemiology. 45. 953-61. 10.1007/s00127-009-0141-2.

3) Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. Lancet. 2007 Sep 8;370(9590):851-8. doi: 10.1016/S0140-6736(07)61415-9. PMID: 17826170.

4) Farren GL, Zhang T, Gu X, Thomas KT. Sedentary behavior and physical activity predicting depressive symptoms in adolescents beyond attributes of health-related physical fitness. J Sport Health Sci. 2018 Oct;7(4):489-496. doi: 10.1016/j.jshs.2017.03.008. Epub 2017 Mar 22. PMID: 30450259; PMCID: PMC6226422.

5) Lu, H., Nie, P. & Qian, L. Do Quarantine Experiences and Attitudes Towards COVID-19 Affect the Distribution of Mental Health in China? A Quantile Regression Analysis. Applied Research Quality Life 16, 1925–1942 (2021). https://doi.org/10.1007/s11482-020-09851-0
6) Kuwato, M., Hirano, Y. Sense of coherence, occupational stressors, and mental health among Japanese high school teachers in Nagasaki prefecture: a multiple regression analysis. BMC Public Health 20, 1355 (2020). https://doi.org/10.1186/s12889-020-09475-x
7) Schilling, E.A., Aseltine, R.H. & Gore, S. Adverse childhood experiences and mental health in young adults: a longitudinal survey. BMC Public Health 7, 30 (2007). https://doi.org/10.1186/1471-2458-7-30
8) Psychosocial job quality and mental health among young workers: a fixed-effects regression analysis using 13 waves of annual data by Allison Milner, PhD,1 Lauren Krnjacki, MPH,2 Anthony D LaMontagne, ScD12 | https://www-jstor-org.ezproxy.lib.usf.edu/stable/pdf/26386160.pdf?refreqid=excelsior%3A079549e7a065c0d450786a64c476eea7&ab_segments=&origin=
9) "Anxiety Disorders." National Institute of Mental Health, 2019, www.nimh.nih.gov/health/topics/anxiety-disorders.
10) "Any Anxiety Disorder." National Institute of Mental Health, 2017, www.nimh.nih.gov/health/statistics/any-anxiety-disorder.
11) "Depression." National Institute of Mental Health, Feb. 2018, www.nimh.nih.gov/health/topics/depression.
12) "Major Depression." National Institute of Mental Health (NIMH), Jan. 2022, www.nimh.nih.gov/health/statistics/major-depression#:~:text=disorders%2C%20or%20medication.-.

Appendix: R Code

```
rm(list=ls())
setwd("C:/Users/caitl/Downloads")
Adult20 <- read.csv("adult20.csv")
head(Adult20)
names(Adult20)
View(Adult20)


#Depression Level: One of our response variables

Adult20$DEPLEVEL_A <- ifelse(Adult20$DEPLEVEL_A == 1 ,'1',ifelse(Adult20$DEPLEVEL_A == 2,'3',
                ifelse(Adult20$DEPLEVEL_A == 3,'2',ifelse(Adult20$DEPLEVEL_A == 7,'99',ifelse(Adult20$DEPLEVEL_A == 8,'99',
                ifelse(Adult20$DEPLEVEL_A == 9,'99','99'))))))

table(is.na(Adult20$DEPLEVEL_A))
#Lots of missing values...can use other depression variables to help us.

Adult20$DEPLEVEL_A[(is.na(Adult20$DEPLEVEL_A)==TRUE)&(Adult20$DEPFREQ_A==5)] = 0
Adult20$DEPLEVEL_A[(is.na(Adult20$DEPLEVEL_A)==TRUE)&(Adult20$DEPEV_A==2)] = 0
```

```
Adult20$DEPLEVEL_A[(is.na(Adult20$DEPLEVEL_A))==TRUE] = 99

#Take out the unknowns, there are less than 100 and we want to use DEPLEVEL_A as
#a numeric variable.
table(Adult20$DEPLEVEL_A)
Adult20 = subset(Adult20, (DEPLEVEL_A != 99))
table(Adult20$DEPLEVEL_A)
Adult20$DEPLEVEL_A = as.numeric(Adult20$DEPLEVEL_A)

#Anxiety Level: One of our response variables

Adult20$ANXLEVEL_A = ifelse(Adult20$ANXLEVEL_A==1, '1', ifelse(Adult20$ANXLEVEL_A==2, '3',
                     ifelse(Adult20$ANXLEVEL_A==3, '2', ifelse(Adult20$ANXLEVEL_A==7, '99',
                     ifelse(Adult20$ANXLEVEL_A==8, '99', ifelse(Adult20$ANXLEVEL_A==9, '99', '99'))))))

table(is.na(Adult20$ANXLEVEL_A))
#Lots of missing values...can use other anxiety variables to help us.

Adult20$ANXLEVEL_A[Adult20$ANXFREQ_A==5] = 0
Adult20$ANXLEVEL_A[(is.na(Adult20$ANXLEVEL_A)==TRUE)&(Adult20$ANXEV_A==2)] = 0
Adult20$ANXLEVEL_A[(is.na(Adult20$ANXLEVEL_A))==TRUE] = 99

#Take out the unknowns, there are less than 100 and we want to use ANXLEVEL_A as
#a numeric variable.
table(Adult20$ANXLEVEL_A)
Adult20 = subset(Adult20, (ANXLEVEL_A != 99))
table(Adult20$ANXLEVEL_A)
Adult20$ANXLEVEL_A = as.numeric(Adult20$ANXLEVEL_A)


#Plotting / Data Exploration
#Subset to only known data

#Depression vs. Anxiety
onlyknownAnxDep = subset(Adult20, (ANXLEVEL_A != 99)&(DEPLEVEL_A!=99))
counts = table(onlyknownAnxDep$ANXLEVEL_A, onlyknownAnxDep$DEPLEVEL_A)
barplot(counts, main="Anxiety Level vs. Depression Level",
    xlab="Anxiety Level", col=c("darkblue","red", "green", "yellow"),
    legend = rownames(counts))
onlyknownAnxDep$ANXLEVEL_A = as.numeric(onlyknownAnxDep$ANXLEVEL_A)
onlyknownAnxDep$DEPLEVEL_A = as.numeric(onlyknownAnxDep$DEPLEVEL_A)
cor.test(onlyknownAnxDep$ANXLEVEL_A, onlyknownAnxDep$DEPLEVEL_A)
#There is a 53% correlation between anxiety level and depression level.


#Anxiety vs. Sleep
```

```
onlyknownAnxSleep = subset(Adult20, (ANXLEVEL_A!= 99)& (SLPHOURS_A!=97)& (SLPHOURS_A!=98) &
(SLPHOURS_A!=99))
counts = table(onlyknownAnxSleep$ANXLEVEL_A, onlyknownAnxSleep$SLPHOURS_A)
barplot(counts, main="Anxiety Level vs. Hours of Sleep",
      xlab="Sleep Hours", col=c("darkblue","red", "green", "yellow"),
      legend = rownames(counts))
onlyknownAnxSleep$ANXLEVEL_A = as.numeric(onlyknownAnxSleep$ANXLEVEL_A)
onlyknownAnxSleep$SLPHOURS_A = as.numeric(onlyknownAnxSleep$SLPHOURS_A)
cor.test(onlyknownAnxSleep$ANXLEVEL_A, onlyknownAnxSleep$SLPHOURS_A)
#No correlation between hours of sleep and anxiety


#Depression vs. Sleep
onlyknownDepSleep = subset(Adult20, (DEPLEVEL_A!= 99)& (SLPHOURS_A!=97)& (SLPHOURS_A!=98) &
(SLPHOURS_A!=99))
counts = table(onlyknownDepSleep$DEPLEVEL_A, onlyknownDepSleep$SLPHOURS_A)
barplot(counts, main="Depression Level vs. Hours of Sleep",
      xlab="Sleep Hours", col=c("darkblue","red", "green", "yellow"),
      legend = rownames(counts))
onlyknownDepSleep$DEPLEVEL_A = as.numeric(onlyknownDepSleep$DEPLEVEL_A)
onlyknownDepSleep$SLPHOURS_A = as.numeric(onlyknownDepSleep$SLPHOURS_A)
cor.test(onlyknownDepSleep$DEPLEVEL_A, onlyknownDepSleep$SLPHOURS_A)
#No correlation between hours of sleep and depression



#Actionable variables

#Nicotine. Creating column for use of nicotine related products (cigarettes, e-cigarettes, smokeless
tobacco, pipes, cigars)
Adult20$nicotine = 3
table(Adult20$nicotine)
Adult20$nicotine[Adult20$SMKCIGST_A==1 | Adult20$SMKCIGST_A==2 | Adult20$SMKCIGST_A==4 |
Adult20$SMKCIGST_A==5] = 'Currently'
Adult20$nicotine[Adult20$SMKCIGST_A==4] = 'Never'
Adult20$nicotine[Adult20$SMKCIGST_A==3] = 'Formerly'
Adult20$nicotine[Adult20$SMKCIGST_A==9] = 'Unknown'
table(Adult20$nicotine, Adult20$SMKCIGST_A)
table(Adult20$nicotine, Adult20$SMKECIGST_A)
Adult20$nicotine[Adult20$SMKECIGST_A==1 | Adult20$SMKECIGST_A==4] = 'Currently'
Adult20$nicotine[(Adult20$SMKECIGST_A==2)&(Adult20$SMKCIGST_A==4)] = 'Formerly'
table(Adult20$nicotine, Adult20$SMOKELSCUR_A)
Adult20$nicotine[Adult20$SMOKELSCUR_A==1 | Adult20$SMOKELSCUR_A==2] = 'Currently'
table(Adult20$nicotine, Adult20$PIPECUR_A)
Adult20$nicotine[Adult20$PIPECUR_A==1 | Adult20$PIPECUR_A==2] = 'Currently'
```

```
table(Adult20$nicotine, Adult20$CIGARCUR_A)
Adult20$nicotine[Adult20$CIGARCUR_A==1 | Adult20$CIGARCUR_A==2] = 'Currently'
Adult20 = subset(Adult20, nicotine!='Unknown')
Adult20$nicotine <- as.factor(Adult20$nicotine)


#Walking
colnames(Adult20)[which(names(Adult20) == "WLKLEIS_A")] <- "walk"
table(Adult20$walk)
table(is.na(Adult20$walk))
Adult20$walk[Adult20$walk ==2] = 0
Adult20$walk[Adult20$walk ==7 | Adult20$walk ==8 | Adult20$walk ==9 | is.na(Adult20$walk) ==TRUE]
= 'Unknown'
Adult20 = subset(Adult20, walk!='Unknown')



#Sleep
colnames(Adult20)[which(names(Adult20) == "SLPHOURS_A")] <- "sleep"
Adult20$sleep[Adult20$sleep==97 | Adult20$sleep==98] =99
table(Adult20$sleep)
Adult20 = subset(Adult20, sleep!=99)



#Exercise - Create new column which measures how physically active
colnames(Adult20)[which(names(Adult20) == "MODFREQW_A")] <- "ModExerciseFreq"
colnames(Adult20)[which(names(Adult20) == "VIGFREQW_A")] <- "VigExerciseFreq"
colnames(Adult20)[which(names(Adult20) == "STRFREQW_A")] <- "StrExerciseFreq"


Adult20$ModExerciseFreq[Adult20$ModExerciseFreq==94]=0
Adult20$ModExerciseFreq[Adult20$ModExerciseFreq==95]=29
Adult20$ModExerciseFreq[Adult20$ModExerciseFreq==96]=-1
Adult20$ModExerciseFreq[Adult20$ModExerciseFreq>96]=-2
table(Adult20$ModExerciseFreq)

Adult20$VigExerciseFreq[Adult20$VigExerciseFreq==94]=0
Adult20$VigExerciseFreq[Adult20$VigExerciseFreq==95]=29
Adult20$VigExerciseFreq[Adult20$VigExerciseFreq==96]=-1
Adult20$VigExerciseFreq[Adult20$VigExerciseFreq>96]=-2
table(Adult20$VigExerciseFreq)

Adult20$StrExerciseFreq[Adult20$StrExerciseFreq==94]=0
Adult20$StrExerciseFreq[Adult20$StrExerciseFreq==95]=29
Adult20$StrExerciseFreq[Adult20$StrExerciseFreq==96]=-1
```

```
Adult20$StrExerciseFreq[Adult20$StrExerciseFreq>96]=-2
table(Adult20$StrExerciseFreq)


#Exercising 0 times a week can be considered not active...
Adult20$Exercise[Adult20$ModExerciseFreq==0 | Adult20$VigExerciseFreq==0 |
Adult20$StrExerciseFreq==0] = 'Not Active'
#Exercising 1-2 times a week can be considered lightly active...overrides Not Active
Adult20$Exercise[Adult20$ModExerciseFreq==1 | Adult20$ModExerciseFreq==2 |
Adult20$VigExerciseFreq==1 | Adult20$VigExerciseFreq==2 | Adult20$StrExerciseFreq==1 |
Adult20$StrExerciseFreq==2] = 'Lightly Active'
#Exercising 3-6 times a week can be considered moderately active...overrides Not Active and Lightly Active
Adult20$Exercise[((Adult20$ModExerciseFreq>2)&(Adult20$ModExerciseFreq<7)) |
((Adult20$VigExerciseFreq>2)&(Adult20$VigExerciseFreq<7)) |
((Adult20$StrExerciseFreq>2)&(Adult20$StrExerciseFreq<7))] = 'Moderately Active'
#Exercising 7 times or more a week can be considered very active...overrides Not Active, Lightly Active and
Moderately active
Adult20$Exercise[Adult20$ModExerciseFreq>6 | Adult20$VigExerciseFreq>6 |
Adult20$StrExerciseFreq>6] = 'Very active'
#Physically unable to to exercise...only this factor if labeled physically unable for all three types
Adult20$Exercise[(Adult20$ModExerciseFreq==-1)&(Adult20$VigExerciseFreq==-
1)&(Adult20$StrExerciseFreq==-1)] = 'Not Physically Able'
#Unknown...only this factor if unknown for all three types
Adult20$Exercise[(Adult20$ModExerciseFreq==-2)&(Adult20$VigExerciseFreq==-
2)&(Adult20$StrExerciseFreq==-2)] = 'Unknown'

Adult20 = subset(Adult20, Exercise!='Unknown')
table(Adult20$Exercise)
Adult20$Excercise = as.factor(Adult20$Exercise)



#Weight Loss
colnames(Adult20)[which(names(Adult20) == "NOWWGTPRG_A")] <- "weightloss"
Adult20$weightloss[Adult20$weightloss==2] = 0
Adult20$weightloss[Adult20$weightloss==7 | Adult20$weightloss==8 | Adult20$weightloss==9] =
'Unknown'
table(Adult20$weightloss)
Adult20 = subset(Adult20, weightloss!='Unknown')


#Drinking Status
colnames(Adult20)[which(names(Adult20) == "DRKSTAT_A")] <- "alcohol"
table(Adult20$alcohol)
Adult20$alcohol[Adult20$alcohol==1] = "Never"
Adult20$alcohol[Adult20$alcohol==2 | Adult20$alcohol==3 | Adult20$alcohol==4] = "Former"
```

```
Adult20$alcohol[Adult20$alcohol==5] = "Infrequent"
Adult20$alcohol[Adult20$alcohol==6] = "Light"
Adult20$alcohol[Adult20$alcohol==7 | Adult20$alcohol==9] = "Moderate"
Adult20$alcohol[Adult20$alcohol==8] = "Heavy"
Adult20$alcohol[Adult20$alcohol==10] = "Unknown"
table(Adult20$alcohol)
Adult20 = subset(Adult20, alcohol!='Unknown')
Adult20$alcohol = as.factor(Adult20$alcohol)


#BMI
colnames(Adult20)[which(names(Adult20) == "BMICAT_A")] <- "bmi"
Adult20$bmi[Adult20$bmi==1] = "Underweight"
Adult20$bmi[Adult20$bmi==2] = "Healthy Weight"
Adult20$bmi[Adult20$bmi==3] = "Overweight"
Adult20$bmi[Adult20$bmi==4] = "Obese"
Adult20$bmi[Adult20$bmi==9] = "Unknown"
table(Adult20$bmi)
Adult20 = subset(Adult20, bmi!='Unknown')
Adult20$bmi = as.factor(Adult20$bmi)

#Food Security
colnames(Adult20)[which(names(Adult20) == "FDSCAT4_A")] <- "FoodSecurity"
#Placeholders...
Adult20$FoodSecurity[Adult20$FoodSecurity==4] = -1
Adult20$FoodSecurity[Adult20$FoodSecurity==3] = -2
Adult20$FoodSecurity[Adult20$FoodSecurity==2] = -3
Adult20$FoodSecurity[Adult20$FoodSecurity==1] = -4
Adult20$FoodSecurity[Adult20$FoodSecurity==-1] = 1
Adult20$FoodSecurity[Adult20$FoodSecurity==-2] = 2
Adult20$FoodSecurity[Adult20$FoodSecurity==-3] = 3
Adult20$FoodSecurity[Adult20$FoodSecurity==-4] = 4
Adult20 = subset(Adult20, FoodSecurity!=8)
table(Adult20$FoodSecurity)

#Creating a new subset with all actionable variables...
Adult20Actionable = Adult20[, c("nicotine", "walk", "sleep", "Exercise", "weightloss", "alcohol", "bmi",
"FoodSecurity", "ANXLEVEL_A", "DEPLEVEL_A")]
View(Adult20Actionable)
str(Adult20Actionable)
Adult20Actionable$walk = as.factor(Adult20Actionable$walk)
Adult20Actionable$Exercise = as.factor(Adult20Actionable$Exercise)
Adult20Actionable$weightloss = as.factor(Adult20Actionable$weightloss)
```

```
#Health Related Variables - Unchanging.

Adult20Health = Adult20[, c("HYPEV_A", "CHLEV_A", "CHDEV_A", "ANGEV_A", "MIEV_A", "STREV_A",
                            "ASEV_A", "CANEV_A", "DIBEV_A", "COPDEV_A", "ARTHEV_A", "DEMENEV_A",
                            "HOSPONGT_A", "THERA12M_A", "DEPLEVEL_A", "ANXLEVEL_A")]
names(Adult20Health)
colnames(Adult20Health)[which(names(Adult20Health) == "HYPEV_A")] <- "HYPERTENSION"
colnames(Adult20Health)[which(names(Adult20Health) == "CHLEV_A")] <- "HIGH_CHOLESTEROL"
colnames(Adult20Health)[which(names(Adult20Health) == "CHDEV_A")] <- "HEART_DISEASE"
colnames(Adult20Health)[which(names(Adult20Health) == "ANGEV_A")] <- "ANGINA"
colnames(Adult20Health)[which(names(Adult20Health) == "MIEV_A")] <- "HEART_ATTACK"
colnames(Adult20Health)[which(names(Adult20Health) == "STREV_A")] <- "STROKE"
colnames(Adult20Health)[which(names(Adult20Health) == "ASEV_A")] <- "ASTHMA"
colnames(Adult20Health)[which(names(Adult20Health) == "CANEV_A")] <- "CANCER"
colnames(Adult20Health)[which(names(Adult20Health) == "DIBEV_A")] <- "DIABETES"
colnames(Adult20Health)[which(names(Adult20Health) == "HOSPONGT_A")] <-
"HOSPITAL_OVERNIGHT"
colnames(Adult20Health)[which(names(Adult20Health) == "THERA12M_A")] <-
"THERAPY_NOT_MENTAL"
colnames(Adult20Health)[which(names(Adult20Health) == "COPDEV_A")] <- "LUNG_DISEASE"
colnames(Adult20Health)[which(names(Adult20Health) == "ARTHEV_A")] <- "ARTHRITIS"
colnames(Adult20Health)[which(names(Adult20Health) == "DEMENEV_A")] <- "DEMENTIA"


Adult20Health$HYPERTENSION[Adult20Health$HYPERTENSION==2]=0
Adult20Health$HYPERTENSION[Adult20Health$HYPERTENSION==7
|Adult20Health$HYPERTENSION==8 | Adult20Health$HYPERTENSION==9]=2
table(is.na(Adult20Health$HYPERTENSION))
table(Adult20Health$HYPERTENSION)

Adult20Health$HIGH_CHOLESTEROL[Adult20Health$HIGH_CHOLESTEROL==2]=0
Adult20Health$HIGH_CHOLESTEROL[Adult20Health$HIGH_CHOLESTEROL==7
|Adult20Health$HIGH_CHOLESTEROL==8 | Adult20Health$HIGH_CHOLESTEROL==9]=2
table(is.na(Adult20Health$HIGH_CHOLESTEROL))
table(Adult20Health$HIGH_CHOLESTEROL)

Adult20Health$HEART_DISEASE[Adult20Health$HEART_DISEASE==2]=0
Adult20Health$HEART_DISEASE[Adult20Health$HEART_DISEASE==7
|Adult20Health$HEART_DISEASE==8 | Adult20Health$HEART_DISEASE==9]=2
table(is.na(Adult20Health$HEART_DISEASE))
table

Adult20Health$ANGINA[Adult20Health$ANGINA==2]=0
Adult20Health$ANGINA[Adult20Health$ANGINA==7 |Adult20Health$ANGINA==8 |
Adult20Health$ANGINA==9]=2
table(is.na(Adult20Health$ANGINA))
```

```
Adult20Health$HEART_ATTACK[Adult20Health$HEART_ATTACK==2]=0
Adult20Health$HEART_ATTACK[Adult20Health$HEART_ATTACK==7
|Adult20Health$HEART_ATTACK==8 | Adult20Health$HEART_ATTACK==9]=2
table(is.na(Adult20Health$HEART_ATTACK))

Adult20Health$STROKE[Adult20Health$STROKE==2]=0
Adult20Health$STROKE[Adult20Health$STROKE==7 |Adult20Health$STROKE==8 |
Adult20Health$STROKE==9]=2
table(is.na(Adult20Health$STROKE))

Adult20Health$ASTHMA[Adult20Health$ASTHMA==2]=0
Adult20Health$ASTHMA[Adult20Health$ASTHMA==7 |Adult20Health$ASTHMA==8 |
Adult20Health$ASTHMA==9]=2
table(is.na(Adult20Health$ASTHMA))

Adult20Health$CANCER[Adult20Health$CANCER==2]=0
Adult20Health$CANCER[Adult20Health$CANCER==7 |Adult20Health$CANCER==8 |
Adult20Health$CANCER==9]=2
table(is.na(Adult20Health$CANCER))

Adult20Health$DIABETES[Adult20Health$DIABETES==2]=0
Adult20Health$DIABETES[Adult20Health$DIABETES==7 |Adult20Health$DIABETES==8 |
Adult20Health$DIABETES==9]=2
table(is.na(Adult20Health$DIABETES))

Adult20Health$HOSPITAL_OVERNIGHT[Adult20Health$HOSPITAL_OVERNIGHT==2]=0
Adult20Health$HOSPITAL_OVERNIGHT[Adult20Health$HOSPITAL_OVERNIGHT==7
|Adult20Health$HOSPITAL_OVERNIGHT==8 | Adult20Health$HOSPITAL_OVERNIGHT==9]=2
table(is.na(Adult20Health$HOSPITAL_OVERNIGHT))

Adult20Health$THERAPY_NOT_MENTAL[Adult20Health$THERAPY_NOT_MENTAL==2]=0
Adult20Health$THERAPY_NOT_MENTAL[Adult20Health$THERAPY_NOT_MENTAL==7
|Adult20Health$THERAPY_NOT_MENTAL==8 | Adult20Health$THERAPY_NOT_MENTAL==9]=2
table(is.na(Adult20Health$THERAPY_NOT_MENTAL))

Adult20Health$LUNG_DISEASE[Adult20Health$LUNG_DISEASE==2]=0
Adult20Health$LUNG_DISEASE[Adult20Health$LUNG_DISEASE==7 |Adult20Health$LUNG_DISEASE==8 |
Adult20Health$LUNG_DISEASE==9]=2
table(is.na(Adult20Health$LUNG_DISEASE))

Adult20Health$ARTHRITIS[Adult20Health$ARTHRITIS==2]=0
Adult20Health$ARTHRITIS[Adult20Health$ARTHRITIS==7 |Adult20Health$ARTHRITIS==8 |
Adult20Health$ARTHRITIS==9]=2
table(is.na(Adult20Health$ARTHRITIS))
```

```
Adult20Health$DEMENTIA[Adult20Health$DEMENTIA==2]=0
Adult20Health$DEMENTIA[Adult20Health$DEMENTIA==7          |Adult20Health$DEMENTIA==8          |
Adult20Health$DEMENTIA==9]=2
table(is.na(Adult20Health$DEMENTIA))


Adult20Health = subset(Adult20Health, HYPERTENSION!=2)
Adult20Health = subset(Adult20Health, HIGH_CHOLESTEROL!=2)
Adult20Health = subset(Adult20Health, HEART_DISEASE!=2)
Adult20Health = subset(Adult20Health, ANGINA!=2)
Adult20Health = subset(Adult20Health, HEART_ATTACK!=2)
Adult20Health = subset(Adult20Health, STROKE!=2)
Adult20Health = subset(Adult20Health, CANCER!=2)
Adult20Health = subset(Adult20Health, DIABETES!=2)
Adult20Health = subset(Adult20Health, HOSPITAL_OVERNIGHT!=2)
Adult20Health = subset(Adult20Health, THERAPY_NOT_MENTAL!=2)
Adult20Health = subset(Adult20Health, LUNG_DISEASE!=2)
Adult20Health = subset(Adult20Health, ARTHRITIS!=2)
Adult20Health = subset(Adult20Health, DEMENTIA!=2)
table(Adult20Health$HYPERTENSION)
table(Adult20Health$STROKE)



#Data Visualization
#Histograms of Dependent Variables
hist(Adult20$ANXLEVEL_A)
hist(Adult20$DEPLEVEL_A)
hist(Adult20Actionable$ANXLEVEL_A)
hist(Adult20Actionable$DEPLEVEL_A)
hist(Adult20Health$ANXLEVEL_A)
hist(Adult20Health$DEPLEVEL_A)


#Correlations
#All variables are factors or binary - so we cannot do any correlation charts.
#We will be using ordinal factors for some variables though, and treating
#them as numerical in the model.

#We can still check for multicollinearity using the vif function and taking out
#variables showing mulicollinearity.
cor(Adult20Actionable$ANXLEVEL_A, Adult20Actionable$DEPLEVEL_A)
cor(Adult20Health$ANXLEVEL_A, Adult20Health$DEPLEVEL_A)
#Correlation between variables is relatively high, so we will not be able to use
```

```
#one DV as a predictor for the other.

#Boxplots
boxplot(Adult20Actionable$ANXLEVEL_A~Adult20Actionable$nicotine)
boxplot(Adult20Actionable$ANXLEVEL_A~Adult20Actionable$walk)
boxplot(Adult20Actionable$ANXLEVEL_A~Adult20Actionable$sleep)
boxplot(Adult20Actionable$ANXLEVEL_A~Adult20Actionable$Exercise)
boxplot(Adult20Actionable$ANXLEVEL_A~Adult20Actionable$weightloss)
boxplot(Adult20Actionable$ANXLEVEL_A~Adult20Actionable$alcohol)
boxplot(Adult20Actionable$ANXLEVEL_A~Adult20Actionable$bmi)
boxplot(Adult20Actionable$ANXLEVEL_A~Adult20Actionable$FoodSecurity)

boxplot(Adult20Actionable$DEPLEVEL_A~Adult20Actionable$nicotine)
boxplot(Adult20Actionable$DEPLEVEL_A~Adult20Actionable$walk)
boxplot(Adult20Actionable$DEPLEVEL_A~Adult20Actionable$sleep)
boxplot(Adult20Actionable$DEPLEVEL_A~Adult20Actionable$Exercise)
boxplot(Adult20Actionable$DEPLEVEL_A~Adult20Actionable$weightloss)
boxplot(Adult20Actionable$DEPLEVEL_A~Adult20Actionable$alcohol)
boxplot(Adult20Actionable$DEPLEVEL_A~Adult20Actionable$bmi)
boxplot(Adult20Actionable$DEPLEVEL_A~Adult20Actionable$FoodSecurity)



boxplot(Adult20Health$ANXLEVEL_A~Adult20Health$HYPERTENSION)
boxplot(Adult20Health$ANXLEVEL_A~Adult20Health$HIGH_CHOLESTEROL)
boxplot(Adult20Health$ANXLEVEL_A~Adult20Health$HEART_DISEASE)
boxplot(Adult20Health$ANXLEVEL_A~Adult20Health$ANGINA)
boxplot(Adult20Health$ANXLEVEL_A~Adult20Health$HEART_ATTACK)
boxplot(Adult20Health$ANXLEVEL_A~Adult20Health$STROKE)
boxplot(Adult20Health$ANXLEVEL_A~Adult20Health$ASTHMA)
boxplot(Adult20Health$ANXLEVEL_A~Adult20Health$CANCER)
boxplot(Adult20Health$ANXLEVEL_A~Adult20Health$DIABETES)
boxplot(Adult20Health$ANXLEVEL_A~Adult20Health$LUNG_DISEASE)
boxplot(Adult20Health$ANXLEVEL_A~Adult20Health$ARTHRITIS)
boxplot(Adult20Health$ANXLEVEL_A~Adult20Health$DEMENTIA)
boxplot(Adult20Health$ANXLEVEL_A~Adult20Health$HOSPITAL_OVERNIGHT)
boxplot(Adult20Health$ANXLEVEL_A~Adult20Health$THERAPY_NOT_MENTAL)

boxplot(Adult20Health$DEPLEVEL_A~Adult20Health$HYPERTENSION)
boxplot(Adult20Health$DEPLEVEL_A~Adult20Health$HIGH_CHOLESTEROL)
boxplot(Adult20Health$DEPLEVEL_A~Adult20Health$HEART_DISEASE)
boxplot(Adult20Health$DEPLEVEL_A~Adult20Health$ANGINA)
boxplot(Adult20Health$DEPLEVEL_A~Adult20Health$HEART_ATTACK)
boxplot(Adult20Health$DEPLEVEL_A~Adult20Health$STROKE)
boxplot(Adult20Health$DEPLEVEL_A~Adult20Health$ASTHMA)
boxplot(Adult20Health$DEPLEVEL_A~Adult20Health$CANCER)
```

```
boxplot(Adult20Health$DEPLEVEL_A~Adult20Health$DIABETES)
boxplot(Adult20Health$DEPLEVEL_A~Adult20Health$LUNG_DISEASE)
boxplot(Adult20Health$DEPLEVEL_A~Adult20Health$ARTHRITIS)
boxplot(Adult20Health$DEPLEVEL_A~Adult20Health$DEMENTIA)
boxplot(Adult20Health$DEPLEVEL_A~Adult20Health$HOSPITAL_OVERNIGHT)
boxplot(Adult20Health$DEPLEVEL_A~Adult20Health$THERAPY_NOT_MENTAL)

#Regression models. Since we are treating our ordininal variable as continuous,
#we are doing OLS regression here.
m_AnxietyAction = lm(ANXLEVEL_A~relevel(nicotine, ref="Never")+walk+sleep+relevel(Exercise,
ref="Not Active")+weightloss+relevel(alcohol, ref="Never")+bmi+FoodSecurity, data=Adult20Actionable)
summary(m_AnxietyAction)

m_DepressionAction = lm(DEPLEVEL_A~relevel(nicotine, ref="Never")+walk+sleep+relevel(Exercise,
ref="Not Active")+weightloss+relevel(alcohol, ref="Never")+bmi+FoodSecurity, data=Adult20Actionable)
summary(m_DepressionAction)

m_AnxietyHealth =
lm(ANXLEVEL_A~HYPERTENSION+HIGH_CHOLESTEROL+HEART_DISEASE+ANGINA+HEART_ATTACK+
STROKE+ASTHMA+CANCER+DIABETES+LUNG_DISEASE+ARTHRITIS+DEMENTIA+HOSPITAL_OVERNIG
HT+THERAPY_NOT_MENTAL, data=Adult20Health)
summary(m_AnxietyHealth)

m_DepressionHealth =
lm(DEPLEVEL_A~HYPERTENSION+HIGH_CHOLESTEROL+HEART_DISEASE+ANGINA+HEART_ATTACK+
STROKE+ASTHMA+CANCER+DIABETES+LUNG_DISEASE+ARTHRITIS+DEMENTIA+HOSPITAL_OVERNIG
HT+THERAPY_NOT_MENTAL, data=Adult20Health)
summary(m_DepressionHealth)

#We are going to split the stargazer into two, since there is a total split between
#variables analyzed.
library(stargazer)
stargazer(m_AnxietyAction, m_DepressionAction, type="text", single.row=TRUE)
stargazer(m_AnxietyHealth, m_DepressionHealth, type="text", single.row=TRUE)

#Testing assumptions
plot(m_AnxietyAction, 1)
plot(m_DepressionAction, 1)
plot(m_AnxietyHealth, 1)
plot(m_DepressionHealth, 1)

plot(m_AnxietyAction, 2)
plot(m_DepressionAction, 2)
plot(m_AnxietyHealth, 2)
plot(m_DepressionHealth, 2)
```

```
library(lmtest)
dwtest(m_AnxietyAction)
dwtest(m_DepressionAction)
dwtest(m_AnxietyHealth)
dwtest(m_DepressionHealth)

library(car)
vif(m_AnxietyAction)
vif(m_DepressionAction)
vif(m_AnxietyHealth)
vif(m_DepressionHealth)
```