The Relationship Between Mental Health and the Number of Cigarettes Consumed

Authors and Affiliation:
Aidan Medina, and Yichen Pan
STA 371G
McCombs School of Business

Introduction:
The co-occurrence of smoking habits and related mental health problems has attracted more and more public awareness. Therefore, we intend to investigate the relationship between mental health conditions and how many cigarettes a person consumes per day. By analyzing this relation, we can determine whether mental health symptoms or affective measures are related to smoking behavior. Discovering this relationship is important because it can help us understand the influence of affect on smoking and it is the requisite of future cigarette smoking intervention. In this study, we include nine different variables to research the relationship between cigarette smoking and mental health conditions, such as depression and anxiety.

Data:
The Add Health Wave IV survey is created to study various developmental and health problems from adolescent to adult. Following up the first interview in 1995 for the Wave I survey, this dataset covered over 15,000 participants, but around 10000 participants were unable to be located. Therefore, data provided was collected from an in-home interview in 2008-2009, which contains 5,114 respondents aged from 24 to 32.

Quantitative variables:

- Num_cigar was measured with the question "During the past 30 days, on the days you smoked, how many cigarettes did you smoke each day?" (H4TO6) Responses ranged from 1 (1 cigarette) to 100 (cigarette), with 6-90 omitted. This is the study response variable. The sample size is 1847.
- Age_first_smoke was measured with the question "How old were you when you first smoked cigarettes regularly, that is, at least one cigarette every day for 30 days?" (H4TO4) Responses ranged from 5-31, with 10-29 omitted. This sample size is 1649.
- Times_quit_smoke was measured with the question "How many times have you tried but been unable to quit smoking or using tobacco for at least one month?" (H4TO30) Responses ranged from 1 (1 time) to 25 (25 times). The sample size is 341.

Categorical variables:

- Feel_isolated (H4MH2) is a categorical variable which was measured with the question "How often do you feel isolated from others?" Responses ranged from 0 (never) to 3 (often). The sample size is 1847.

- Feel_depressed (H4MH22) is a categorical variable which was measured with the question "(During the past seven days:) You felt depressed." Responses ranged from 0 (never or rarely) to 3 (most of the time or all of the time). This sample size is 1847.
- Feel_tired (H4MH23) is a categorical variable which was measured with the question "(During the past seven days:) You felt that you were too tired to do things." Responses ranged from 0 (never or rarely) to 3 (most of the time or all of the time). The sample size is 1847.
- Anxiety (H4ID5J) is a categorical variable with only two levels:
  A diagnosis of anxiety was measured with the question "Has a doctor, nurse or other health care provider ever told you that you have or had: anxiety or panic disorder?" The sample size is 1846.
- Stress (H4PE22) is a categorical variable which was measured with the question "I get stressed out easily". Responses ranged from 1 (strongly agree) to 5 (strongly disagree). The sample size is 1845.
- Try_quit_smoke (H4TO27) is a categorical variable which was measured with the question "Have you ever tried to quit or cut down on smoking or using tobacco?". Responses ranged from 0 (no) to 1 (yes). The sample size is 1465.

Literature Review:

Link to the first article: https://jamanetwork.com/journals/jama/article-abstract/193305

The article *Smoking and Mental Illness: A Population-Based Prevalence Study* published on JAMA Network concluded that "persons with mental illness are about twice as likely to smoke as other persons but have substantial quit rates" (Lasser et al., 2000). This study includes a random population sample of 4411 respondents from the National Comorbidity Survey aged between 15 to 54 years, and this survey was conducted from 1991 to 1992. The main difference between this study and our study is that we use the daily cigarette consumption in the past 30 days as our response variable, but the data source of this study did not ascertain the number of cigarettes smoked. This study instead used the number of cigarettes consumed during the smoking most period as its response variable. Another difference is that this research gave detailed definitions of mental illnesses and smoking rates, and used several severe diagnosed mental illnesses as the explanatory variables. However, our study includes different mental health conditions and feelings, such as depression and anxiety, as our explanatory variables.

Link to the second article: https://www.bmj.com/content/348/bmj.g1151.full

This article *Change in Mental Health After Smoking Cessation: Systematic Review and Analysis* published by the BMJ concludes that "smoking cessation is associated with reduced depression, anxiety, and stress and improved positive mood and quality of life compared with continuing to

smoke" (Taylor et al., 2014). Data for this study was extracted from related publications with a data extraction form. 219 full-text articles were eligible after the initial screening, but 166 were excluded before data extraction. Finally, this study included 26 studies in the meta-analyses and acquired additional data from six of them. One of the differences between our work and this study is that this study examined six different measures of mental health, including anxiety, depression, mixed anxiety and depression, positive affect, psychological quality of life, and stress. In contrast, our study only includes psychological isolation, depression, tried, and anxiety. Therefore, we plan to add more explanatory variables to our study to better investigate the relationship between mental health and smoking. Compared to our study, this research also recorded participants' motivation to quit smoke in addition to times people tried to quit smoking.

Link to the third article:
https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-9-285

The article, *Smoking and Mental Illness: results from population surveys in Australia and the United States,* concludes that "smokers with high levels of psychological distress smoked a higher average number of cigarettes per day" (Lawrence, Mitro, and Zubrick, 2009). This article goes over a study from the US National Comorbidity Survey-Replication (the study is called the Survey of Mental Health and Wellbeing or SMHWB) which included a sample of 8,841 adults aged 16-85 who were personally interviewed based on the prevalence of three major groups of disorders: anxiety disorders, affective disorders, and substance use disorders. Another survey conducted, the National Health Interview Survey (NHIS) measured the level of psychological distress in families and was used with the SMHWB and one more study by the Australian Health Interview Survey to investigate the relationship between current smoking and ICD-10 mental disorders. One of the main differences between this study and our study is that the NHIS measures the psychological distress of smokers and also family members who live with smokers. Another difference is shown in the way that the quantitative variable is studied, as in our project we will be looking at the age first smoked, while these studies break down the prevalence between smoking and mental illness into age groups. The studies also use a Kessler 6 scale to measure a surveyor's distress which measures distress over four weeks. Finally, the article measures how the amount of cigarettes a person smokes per day influences their Kessler 6 score. According to the article, "The average number of cigarettes smoked per day by current smokers increased with the level of psychological distress, from 12 per day among smokers with no psychological distress to 19 per day among smokers with serious psychological distress" (Lawrence et al., 2009).

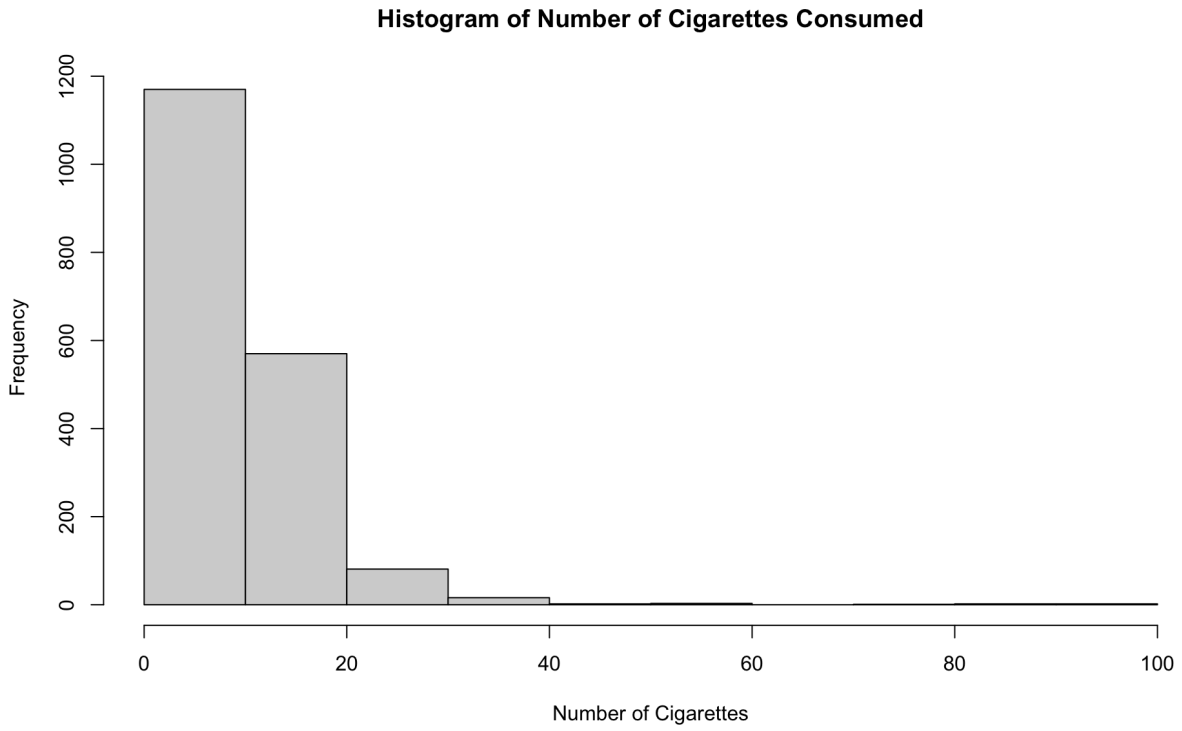
Conclusion:
Based on our review of related work, we have modified our variables to include an additional categorical variable H4PE22: "I get stressed out easily". By including this new variable, we intend to determine whether stress levels influence smoking frequency. After researching more

articles regarding this topic, we concluded that our selected explanatory variables are important when exploring and analyzing how the smoking rate is impacted by emotional problems. We aspire to provide a more accurate and comprehensive analysis regarding the relationship between smoking and mental health problems, leading to a better understanding of how to combat such a prominent issue.

Methods:

Analysis for the Response Variable (num_cigar):
Our response variable for this project is the number of cigarettes consumed per day during the past 30 days. We analyzed the response variable based on its frequency distribution. Figure 1 shows a histogram of the response variable's distribution.

Figure 1
Distribution of the Number of Cigarettes



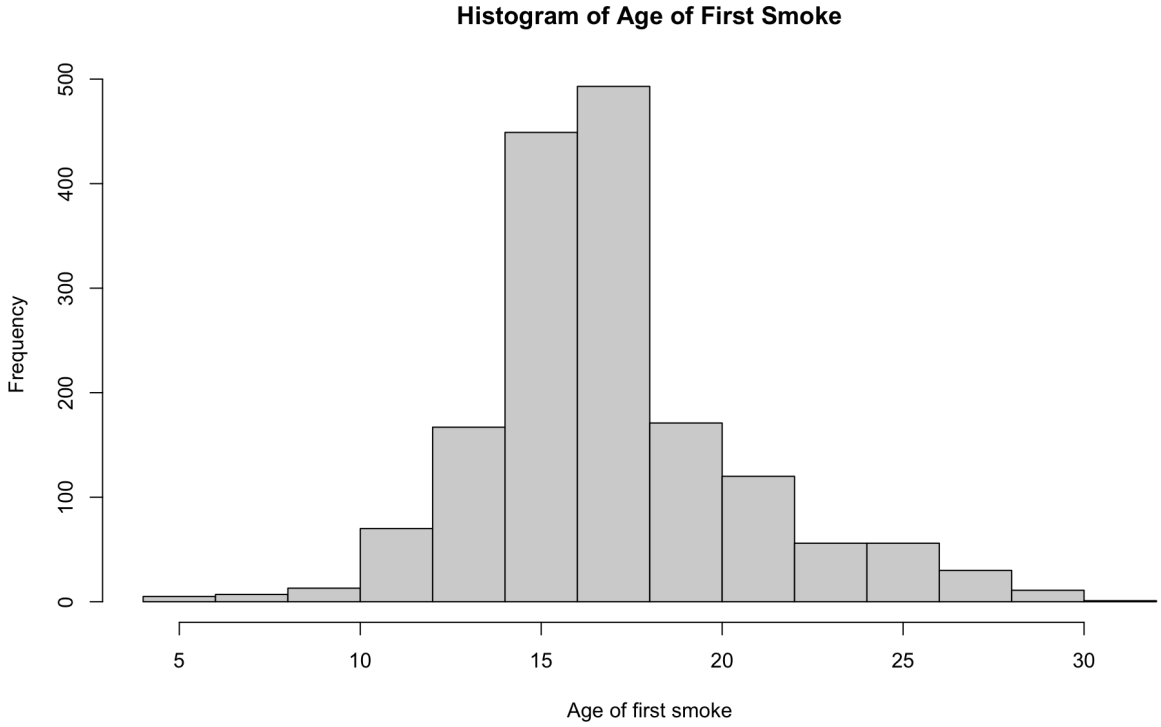
Note. The histogram above illustrates the frequency distribution for the number of cigarettes consumed per day during the past 30 days.

There are a total of 1847 observations in the dataset. From the mean of the response variable, we understood the population participating in this survey consumed approximately 10.517 cigarettes each day. The standard deviation is 9.304. The median is 10, which is slightly lower than the

mean. The histogram is skewed to the right (negatively skewed), and it is an unimodal distribution. Most people in the sample intend to smoke 0-10 cigarettes per day, and there are also a large number of people smoking 10-20 cigarettes each day. The data for the response variable ranges from 0 to 100. No outliers were found in the dataset. Missing data is completely missing at random.

Analysis for the First Explanatory Variable (age_first_smoke):
We analyzed the frequency distribution for the quantitative variable called age_first_smoke. Figure 2 shows a histogram for the distribution of this variable.

Figure 2
Distribution of the Age of First Smoke



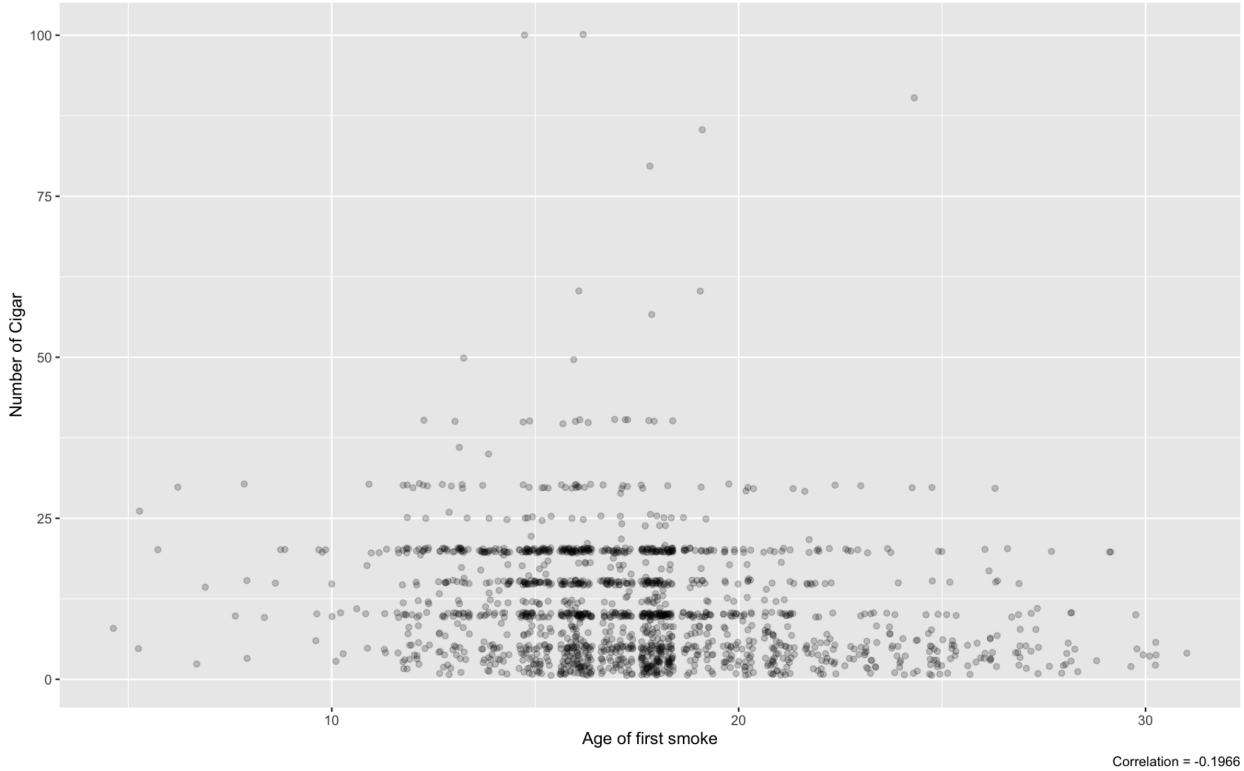
Note. The histogram above shows the frequency distribution for the age of first smoke.

This quantitative variable has a total of 1649 observations in the dataset. The sample size is slightly smaller than the sample size for the response variable. No outlier was found in the dataset. Missing data is completely missing at random. The data from this explanatory variable range from 5 to 31. The mean age respondents formed a smoking habit at 17.47 years old. The standard deviation is 3.64. The histogram is symmetrical, and it is a normal unimodal distribution. Data concentrates around the center of the histogram, and then we know that most smokers tend to smoke their first cigarette between the ages of 13 and 18.

The relationship between the response variable (number of cigarettes consumed) and the quantitative explanatory variable (the age of first smoke) has been analyzed. Figure 3 below shows a scatterplot between the two variables.

Figure 3

Scatterplot Illustrating the Relationship Between the Number of Cigarettes Consumed and the Age of First Smoke



Note. Scatterplot showcasing the relationship between the response variable (number of cigarettes consumed) and the quantitative explanatory variable (the age of first smoke). The correlation is -0.1966.

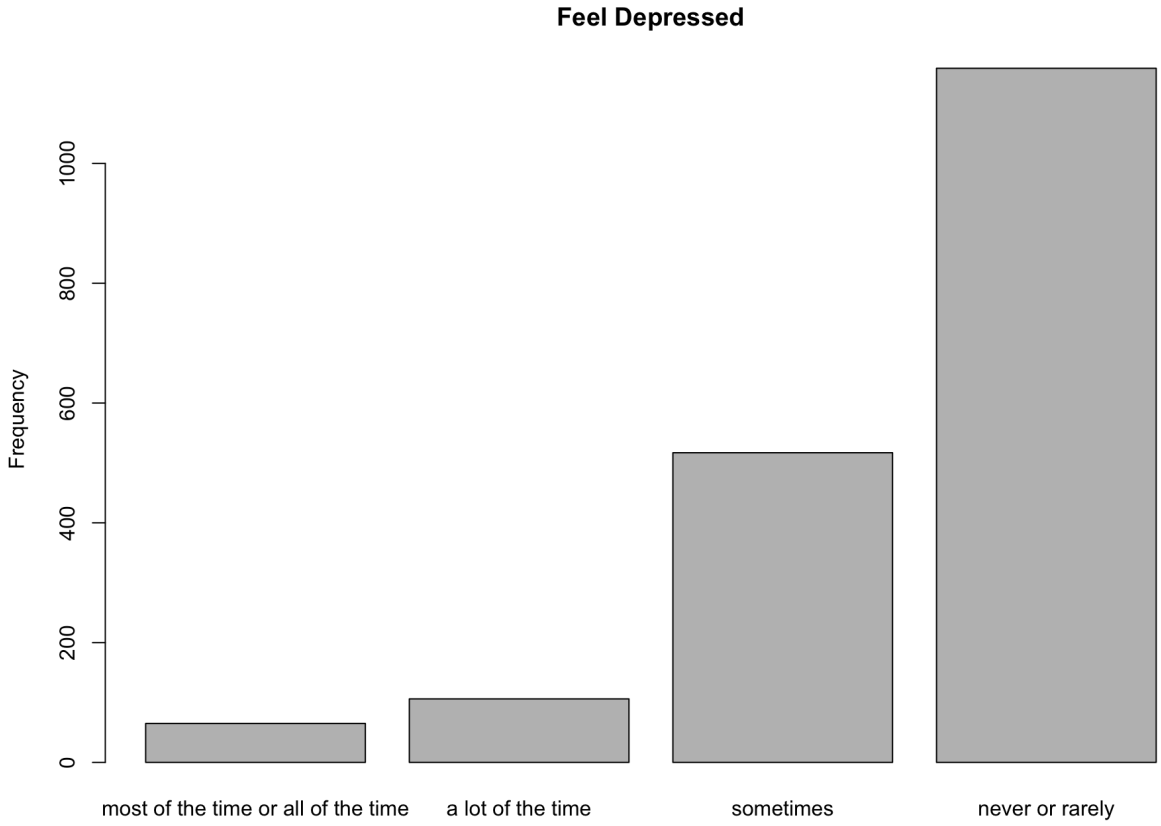
According to the scatterplot above, there is no linear relationship shown. The correlation between those two variables is approximately -0.1966, which is close to 0. Therefore, the relationship between them is weak and hard to detect.

Analysis for the Second Explanatory Variable (feel_depressed):

We analyzed the frequency distribution for the categorical variable called feel_depressed. Missing data was completely at random either due to refusing to answer or don't know. Figure 4 shows a bar chart for the distribution of this variable.

Figure 4

Distribution of Frequency for feeling depressed



Note. Bar chart illustrating the frequency distribution for the four categories (a lot of the time, most of the time or all the time, never or rarely, and sometimes) of the frequency for feeling depressed.

From the bar chart, most people answered this question with “never or rarely” and less people answered “most of the time or all the time”. In table 1, the precise number of respondents for four categories are shown.

Table 1

Numerical Summary for the Frequency of Feeling Depressed

Most or all of the time	A lot of the time	Sometimes	Never or rarely	**TOTAL**
65	106	517	1159	**1847**

Note. Counts for four categories of the Frequency of Feeling Depressed. Among a total of 1847 observations in the dataset, 106 respondents feel stress a lot of the time, 65 respondents feel

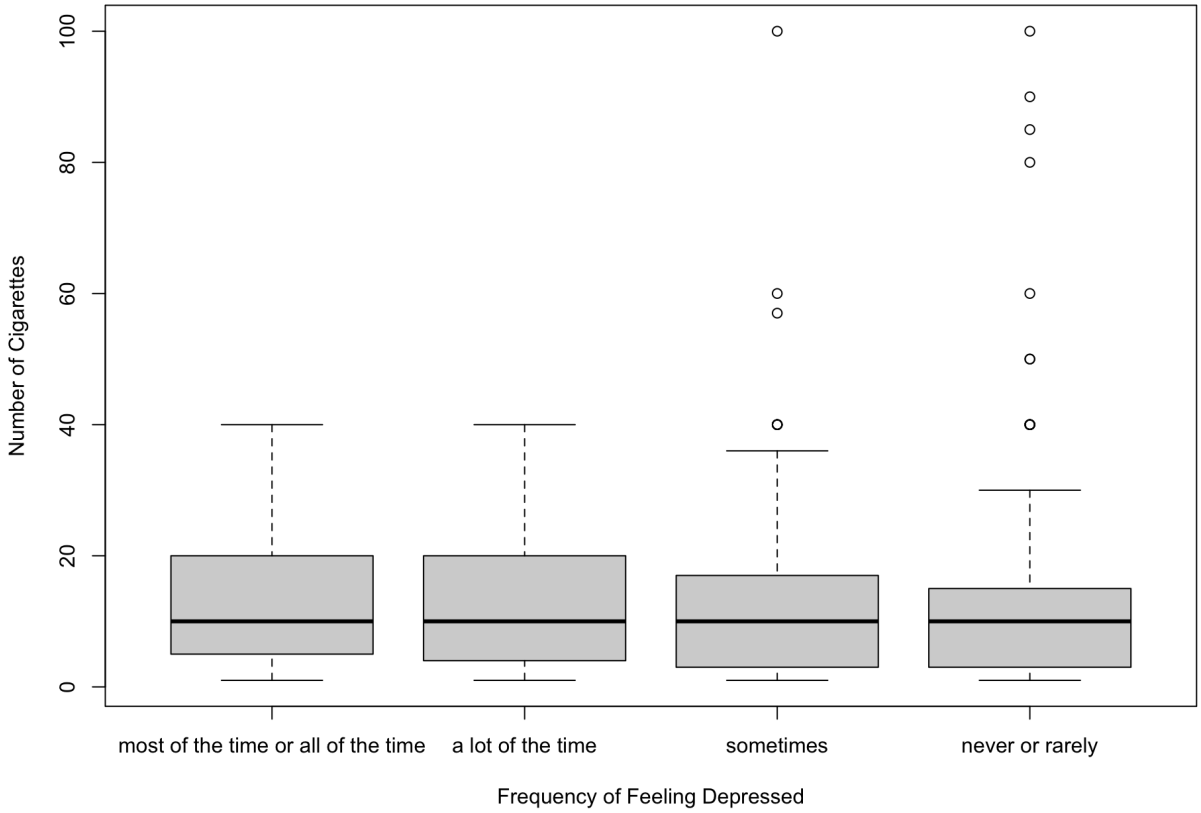
stress most or all of the time, 1159 respondents never or rarely feel stress, and 517 respondents sometimes feel stress.

Calculating the percentage of those four categories, we know that 5.74% of the respondents answered “a lot of the time”, 3.52% of the respondents answered “most of the time or all of the time”, 62.75% of the respondents answered “never or rarely”, and 27.99% of the respondents answered “sometimes”.

The relationship between the frequency for feeling depressed and the number of cigarettes consumed has been analyzed. The boxplot analyzing the relationship between two variables is presented in Figure 5.

Figure 5

Comparison for the Number of Cigarettes Consumed and the Frequency of Feeling Depressed



Note. Boxplot comparing the number of cigarettes consumed and the frequency of feeling depressed.

A few outliers were found in the dataset, specifically in the “never or rarely” category. There is no relationship found in the result. The lower quartile seems to be the same for different four categories, but the upper quartile varies: people who feel depressed more frequently tend to have a higher upper quartile. The precise numbers for the average number of cigarettes consumed in four different categories are recorded in Table 2.

Table 2

Numerical Summary for the Average Number of Cigarettes Consumed and the Frequency of Feeling Depressed

Most of the time or all of the time	A lot of the time	Sometimes	Never or rarely
12.01538	11.54717	10.63056	10.28818

Note. Numerical summaries for the number of cigarettes consumed and the four different frequencies (a lot of the time, most or all of the time, never or rarely, and sometimes) of feeling depressed.

The average number of cigarettes consumed per day among people who feel depressed most or all of the time is 12.02, among those who feel depressed a lot of the time is 11.55, among those who feel depressed sometimes is 10.63, and among those who feel depressed never or rarely is 10.29. Based on those data, people who feel depressed more frequently are likely to consume more cigarettes per day. The standard deviation for the number of cigarettes consumed per day for four different frequencies of feeling depressed are presented in Table 3.

Table 3

Numerical Summary for the Standard Deviation of the Number of Cigarettes Consumed and the Frequency of Feeling Depressed

Most of the time or all of the time	A lot of the time	Sometimes	Never or rarely
9.274448	8.807067	9.651529	9.188007

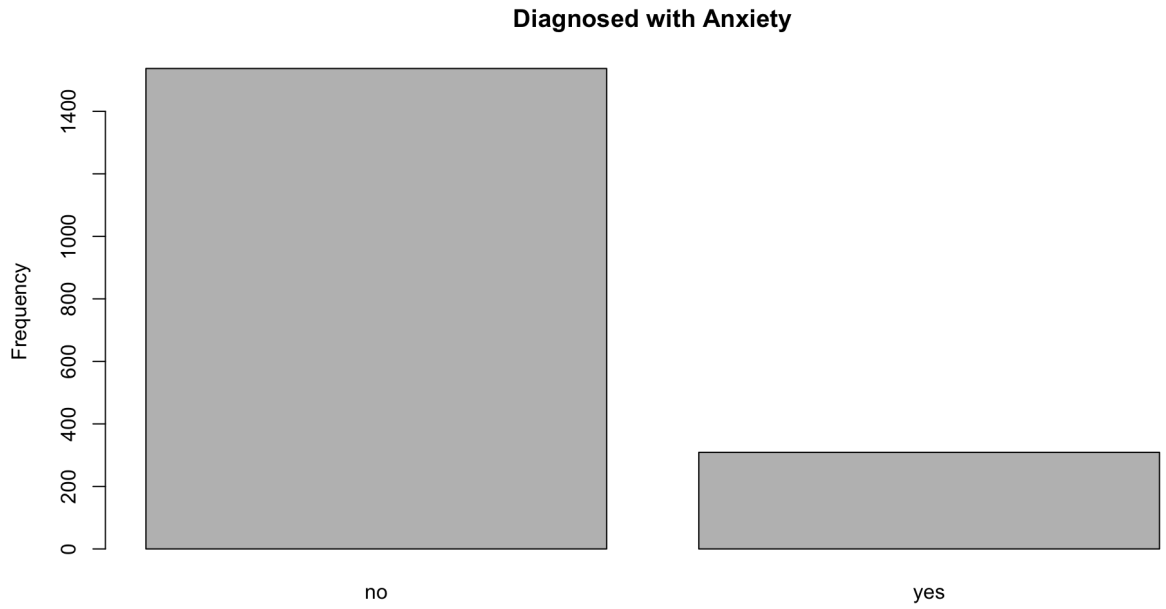
Note. Numerical summaries for the standard deviation of the number of cigarettes consumed and the four different frequencies (a lot of the time, most or all of the time, never or rarely, and sometimes) of feeling depressed.

The standard deviation for the number of cigarettes consumed per day among people who feel depressed most or all of the time is 9.27, among those who feel depressed a lot of the time is

8.81, among those who feel depressed sometimes is 9.65, and among those who feel depressed never or rarely is 9.19.

Analysis for the Third Explanatory Variable (anxiety):
We analyzed the frequency distribution for the two-level categorical variable called anxiety. Missing data was completely at random either due to refusing to answer or don't know. Figure 6 shows a bar chart for the distribution of this variable.

Figure 6
Distribution of Whether a Person Diagnosed with Anxiety



Note. Bar chart illustrating the frequency distribution for the two categories (no, yes) of whether a person is diagnosed with anxiety or panic disorder by doctors.

From the bar chart, most people answered this question with "no" and least people answered "yes". In table 4, the precise number of respondents for four categories are shown.

Table 4
Numerical Summary for the Frequency of Being Diagnosed with Anxiety

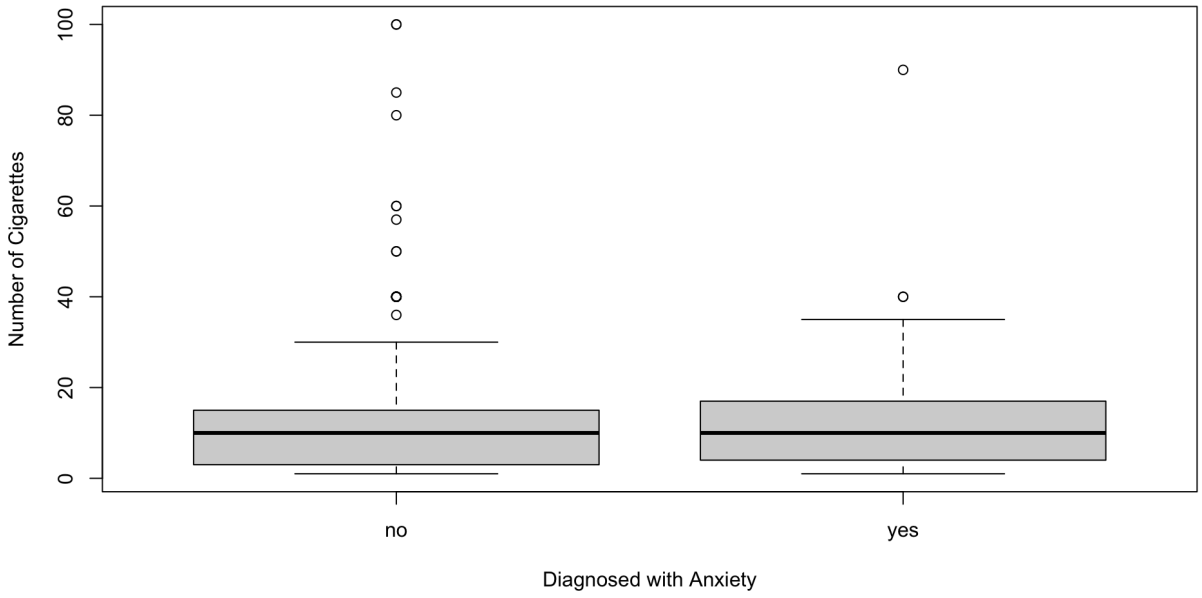
No	Yes	**Total**
1537	309	**1846**

Note. Counts for four categories of the Frequency of being Diagnosed with Anxiety. Among a total of 1846 observations in the dataset, 1537 respondents were not diagnosed with anxiety, while 309 respondents were. Calculating the percentage of these categories, we know that 83.26% of respondents answered “no” and 16.74% of respondents answered “yes”.

The relationship between whether a person is diagnosed with anxiety and the number of cigarettes consumed has been analyzed. The boxplot analyzing the relationship between two variables is presented in Figure 7.

Figure 7

Comparison for the Number of Cigarettes Consumed and whether a Person is Diagnosed with Anxiety



Note. Boxplot comparing the number of cigarettes consumed and whether a person is diagnosed with anxiety or not.

There are a few outliers, specifically in the “no” category. There is no relationship found in the result. The lower quartile seems to be the same for different four categories, but the upper quartile is higher for the category “yes”. This means people usually consume more cigarettes if they are diagnosed with anxiety. The precise numbers for the average number of cigarettes consumed in four different categories are recorded in Table 5.

Table 5

Numerical Summary for the Average Number of Cigarettes Consumed and whether being Diagnosed with Anxiety

No	Yes
10.36630	11.23625

Note. Numerical summaries for the number of cigarettes consumed for two categories (yes or no) of whether a person is diagnosed with anxiety.

The average number of cigarettes consumed per day among people diagnosed with anxiety is 11.24, while among those who are not diagnosed with anxiety is 10.37. Based on those data, people diagnosed with anxiety are likely to consume more cigarettes per day. The standard deviation for the number of cigarettes consumed per day for two categories of whether a person is diagnosed with anxiety are presented in Table 6.

Table 6

Numerical Summary for the Standard Deviation of the Number of Cigarettes Consumed for People not Diagnosed with or Diagnosed with Anxiety

No	Yes
9.335918	9.123566

Note. Numerical summaries for the standard deviation of the number of cigarettes consumed for two categories (yes or no) of whether a person is diagnosed with anxiety.

The standard deviation for the number of cigarettes consumed per day among people diagnosed with anxiety is 9.34, while among those who are not diagnosed with anxiety is 9.12.

Summary for Exploratory Data Analysis:

Through our data analysis we sought to determine the strength (or weakness) of the relationships between the explanatory variables and our response variable. Through our analysis in R, we have concluded that all eight of our chosen explanatory variables have either a weak or no relationship with the response variable.

References

Lasser, K., Boyd, W., Woolhandle, S., Bor, D. H., McCormick, D., & Himmelstein, D. U. (2000, November 22). Smoking and Mental Illness. JAMA. https://jamanetwork.com/journals/jama/article-abstract/193305.

Lawrence, D., Mitrou, F., & Zubrick, S. R. (2009, August 7). Smoking and mental illness: results from population surveys in Australia and the United States. BMC Public Health. https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-9-285.

Taylor, G., McNeill, A., Girling, A., Farley, A., Lindson-Hawley, N., & Aveyard, P. (2014, February 13). Change in mental health after smoking cessation: systematic review and meta-analysis. The BMJ. https://www.bmj.com/content/348/bmj.g1151.full.