ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Paper ID #45015

Statistical Investigation of the Relationship Between Mental Health & Marijuana Use from NHANES Database

Zuyi Huang, Villanova University

Zuyi (Jacky) Huang is an Associate Professor in the Department of Chemical Engineering at Villanova University. He teaches Chemical Process Control (for senior students) and Systems Biology (for graduate students) at Villanova. He is enthusiastic in apply

Ryan Jacob Forte, Villanova University
Rebecca Goehring, Villanova University
Michael Fogliano, Villanova University

Statistical Investigation of the Relationship Between Mental Health & Marijuana Use from NHANES Database

Ryan Forte, Rebecca Goehring, Michael Fogliano, and Zuyi (Jacky) Huang

Department of Chemical and Biological Engineering, Villanova University, Villanova, PA, 19085

Background & Motivation

- Mental health advocacy and awareness have entered the mainstream of culture and society in recent years [1].
- Understanding the relationship between individuals' lifestyle choices and their mental well-being can provide valuable insights into coping mechanisms for mental health challenges.
- One such lifestyle factor worthy of examination is drug use among individuals.
- With the legalization of marijuana in an increasing number of states over the past decade, its prevalence in our culture and society has surged [2].
- People of different ages may choose different ways to cope with bad mental health [1].

We hypothesize that a correlation exists between mental health and marijuana use, particularly among younger adults: individuals with poorer mental health use more marijuana than individuals with better mental health. It is expected that this relationship has become more apparent since the onset of marijuana legalization.

Materials

Data analysis programs: R & RStudio

- R is a free programming language used in many statistical applications.
- R is especially useful because it is equipped with powerful statistical analysis packages.
- Because of its statistical prowess and free al, R has been widely used in data mining & machine learning.
- RStudio provides a friendly user-interface to implement R script.

Data from CDC's National Health and Nutrition Examination Survey (NHANES)

- The NHANES database provides comprehensive and nationally representative health data, enabling critical insights into public health trends, risk factors, and disparities.
- 3,893 Samples were extracted from NHANES in two timepoints: 2009-2010 & 2017-2018 (pre- & post-legalization onset), with the inputs for following questions [3-4]:
  - "Over the last two weeks, how often [have you felt] down, depressed, or hopeless?"
  - "Have you had at least several days with bad mental health?"
  - "Have you ever used marijuana during the past 30 days"
  - "During the past 30 days, on how many days have you used marijuana?"
  - "What is your age?" (Age at screening: ranged from 18-59)
  - "What is the number of household members?"
  - "What is your income?" (i.e., income relative to poverty threshold)

Data Extraction in R

- The survey responses were extracted and digitized into a table format, in which each column corresponds to a specific question, while each row represents an individual sample.
- The two timepoint datasets were combined and filtered to only include subjects who answered each question
- Subjects who never used marijuana were assigned "0 days" before filtering.
- Table 1 summarizes the data extracted from NHANES in this study.

Table 1. Extracted Samples from Surveys for Data Analysis

	Surveyed Subjects	Excluded by criteria questions	Active Marijuana Use	No Current Marijuana Use	Good Mental Health	Poor Mental Health
Pre-legalization	2097	4	505	1588	1531	562
Post-legalization	1796	0	560	1236	1386	410
Total	3893	4	1065	2824	2917	972

Methods

Data Organization in R

- For each timepoint, subjects were divided into six groups based on age and mental health status. The selected age ranges (shown in Table 2 below) allowed our group to focus specifically on younger people (18-25) and to capture differences with age.

Table 2. Age and mental health status for each numbered group in the statistical analysis.

Group	Age	Mental Health
1	18-25	Good
2	26-40	
3	41-59	
4	18-25	Bad
5	26-40	
6	41-59	

Principal Component Analysis (PCA)

- It reduces the dimension of the data for visualization.
- Data points get projected in the directions of highest variance.
- Input is represented a matrix organized with subjects and properties to be analyzed.
  - Subjects include Age & Mental Health Groups
  - Properties were focused on Marijuana Use
- Matrix tabulates number of subjects from each group that used marijuana a particular number of days in the last month.

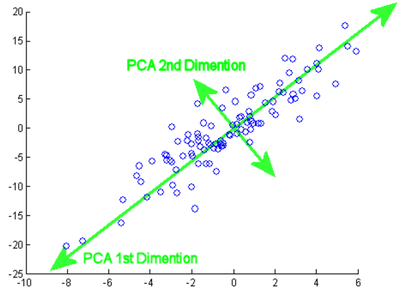


Figure 1. Illustration of how PCA allows for multi-dimensional data to be visualized in the two directions of highest variance [5].

Hierarchical Clustering

- It allows for similarity analysis of the groups used for PCA.
- It does not need a cluster number specified beforehand.
- Distances between different groups were calculated based on PCA projections.
- Groups in the same/closely-linked clusters have more similar marijuana use than groups in distant clusters.

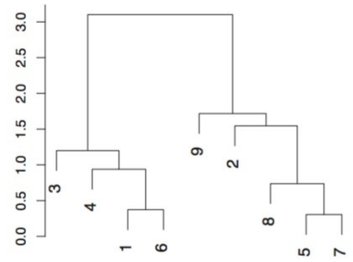


Figure 2. Hierarchical clustering results can be visualized in a dendrogram, with similar groups (based on a particular characteristic) being lumped together.

T-tests

- It quantifies the statistical differences in marijuana use based on mental health status.
- Groups of people of the same age range were compared to assess difference due to mental health

Dominance Analysis

- It is based upon quantitative assessment using linear regression modeling.
- It identifies which input factors contributed the most to marijuana use and how their influence may have changed after the onset of marijuana legalization.

Table 3. Input and output factors analyzed in the dominance analysis.

Variable Names	Meaning of Variables
age	Age (years) of person at time of screening
mental_health	Mental health status of person (0-3)
income_poverty	Ratio of family income to poverty threshold
people_house	Total number of people in the household
mj_days_used	Number of days in last month that a person used marijuana

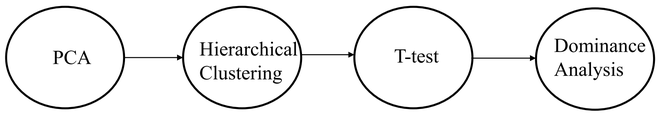


Figure 3. Flow chart summarizing the statistical analyses used.

Results

Principal Component Analysis, Hierarchical Clustering & T-tests:

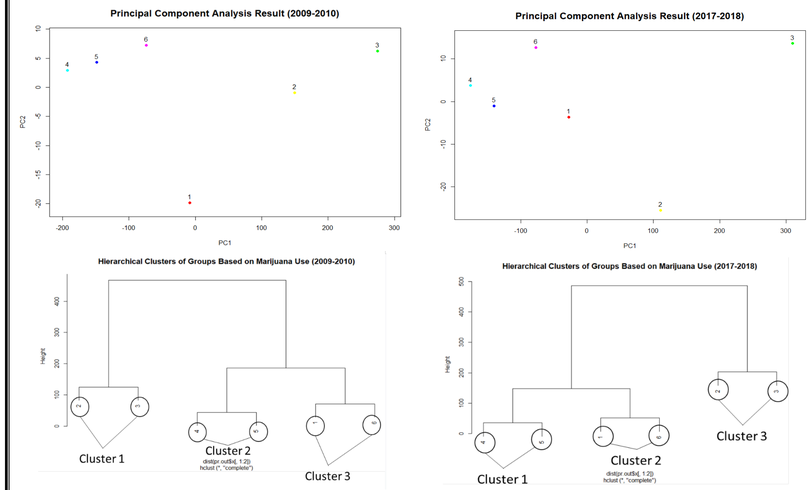
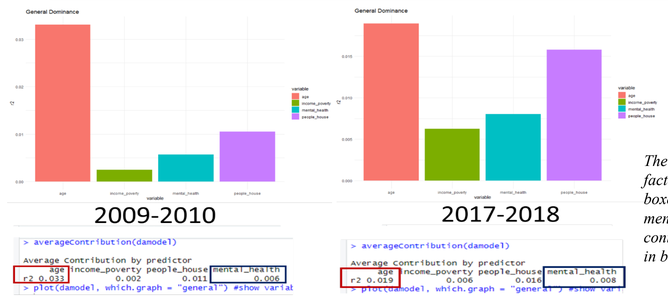


Table 4. T-test results assessing significant differences in monthly marijuana use between people of good and bad mental health. Bolded values indicate significance.

Group	Age	Mental Health	2009-10 Monthly Marijuana Use (Days)	p-value	2017-18 Monthly Marijuana Use (Days)	p-value
1	18-25	18-25	3.79	**0.001594**	4.88	**0.00486**
4	18-25	18-25	6.79		8.24	
2	26-39	26-39	2.62	**0.000355**	5.31	**0.00425**
5	26-39	26-39	5.28		8.42	
3	40-59	40-59	1.58	0.07161	2.56	**0.03591**
6	40-59	40-59	2.30		3.77	



The most-dominant factor contribution is boxed in red; the mental health factor contribution is boxed in blue.

Conclusions

- Individuals with bad mental health tend to have higher levels of marijuana use than individuals with good mental health and young people tend to use more marijuana than older people
- No causal link yet proven between heightened marijuana use in younger adults and worsening mental health levels over the analyzed period (before and after legalization).

Reference

[1] Venters, N. (n.d.-a). NHS choices. https://digital.nhs.uk/blog/transformation-blog/2018/the-past-present-and-future-of-innovation-in-mental-health
[2] Speights, K. (n.d.). Where is marijuana legalized?. The Motley Fool. https://www.fool.com/investing/stock-market/market-sectors/healthcare/marijuana-stocks/marijuana-legalization/#:~:text=In%202014%2C%20voters%20approved%20Measure%2091%20to%20legalize%20recreational%20marijuana.&text=Medical%20marijuana%20is%20legal%20in%20Pennsylvania.,3%20to%20legalize%20medical%20marijuana
[3] Centers for Disease Control and Prevention. (n.d.-a). Nhanes 2009-2010: Drug Use Data Documentation, codebook, and frequencies. Centers for Disease Control and Prevention. https://wwwn.cdc.gov/Nchs/Nhanes/2009-2010/DUQ_F.htm
[4] Centers for Disease Control and Prevention. (n.d.). NHANES 2017-2018 Drug Use. Centers for Disease Control and Prevention. https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/DUQ_J.htm
[5] Raghavan. Principal component analysis (PCA): Explained and implemented. Medium. Online. Available from: https://medium.com/@raghavan99o/principal-component-analysis-pca-explained-and-implemented-eeab7cb73b72