Mapping Digital Mental Health Sentiment Across U.S. Cities

Zhiyang Cheng, Jiayi Li, Cheng Gao

Georgetown University, McCourt School of Public Policy

December 2025

Abstract

Cities across the United States face growing challenges related to population mental health, with psychological distress unevenly distributed across communities. Traditional measures of mental health often rely on surveys or administrative records, which are infrequently collected and not up to date. This report used social media data as a complementary digital indicator to examine patterns of negative emotional expression across U.S. cities. We seek to better understand how emotional distress varies across cities and over time, and how these patterns relate to mental health service availability and broader socioeconomic conditions. Specifically, we focus on eleven large U.S. cities and constructed city-level measures of negative emotional expression based on posts and comments from city-specific subreddits. In addition, we extracted demographic and socioeconomic data that is publicly available. We find that negative sentiment differs persistently across cities and reflects meaningful themes such as social isolation, financial pressure, and city-specific challenges. We also find correlation between mental health service availability and negative sentiment, with evidence of non-linear patterns.

Keywords — *Data Science, Emotional Distress, Sentiment, Psychology, City-level*

1 INTRODUCTION

Mental health has become an increasingly crucial public concern across countries, including the United States. From the National Institute of Mental Health (NIMH), it shows that there was an estimate of 23.1% of adults aged 18 or older to have observed prevalence of AMI (any mental illness) in the United States in 2022. Young adults aged 18-25 years had the highest prevalence of AMI (36.2%) compared to adults aged 26-49 years (29.4%) and aged 50 and older (13.9%). This statistics not only underscores the scale of mental health challenge in the United States, but also illustrates the uneven distribution of psychological distress across groups.

Existing research works suggest that such disparities are shaped not only by individual

characteristics, but also by broader economic and social policy environments. In particular, Donnelly and Farina's study "Mapping Mental Health Across U.S. States: The Role of Economic and Social Support Policies" examines how variation in state-level social and economic support policies—such as income assistance, health insurance coverage, and labor-market protections—relates to mental-health outcomes across the United States. Their study indicates that states with stronger social safety nets tend to experience lower levels of psychological distress, highlighting the importance of policy context as a determinant of mental health. While the study provides strong evidence in showing the relationship between people's mental health status and policy environment, it relies primarily on traditional measures for mental health, such as surveys and administrative records. It may lead to infrequent data collection and delayed release, restricting the usefulness for monitoring up-to-date emotional conditions of people. It also focuses on a broader level — state level, which may overlook meaningful within-state heterogeneity across cities, where local health service support or broader socioeconomic conditions can differ substantially.

Based on the previous work, this project extends the focus from state level to city level and incorporates digital sentiment data as an alternative and complementary indicator on population mental health. It captures finer geographic variation and more immediate signals of mental-health conditions than traditional approaches. Specifically, by focusing on city-level variation, our project aims to generate descriptive insights into how psychological distress is expressed online, including identifying the dominant themes that appear in mental-health discussions across different cities. In addition to textual analysis on mental health, we also plan to further analyze the relationship between digital mental-health sentiment and city-level factors, including mental-health service availability and broader seriocomic conditions. By combining sentiment analysis with regression, this project provides a clearer picture of how mental-health experiences are expressed online across cities and how the expressions vary across cities with different levels of mental-health support. Overall, this study helps us assess whether digital sentiment could serve as an indicator for identifying timely emotional distress and informing more responsive resource allocation.

2 DATA COLLECTION

2.1 Data Acquisition

This study utilized data from three main sources: Reddit, American Community Survey (ACS), and CountyHealthRankings.org.

2.1.1 Reddit

To measure negative sentiment across U.S. cities, we used the Reddit API to scrape relevant data. To get data on a city-level, since Reddit no longer provides IP address or precise location information, we relied on city-specific subreddits as a proxy for geographical location and extracted data from communities like r/AskChicago, r/AskLosAngeles, r/NYC, etc. These subreddits serve as city-specific discussion forums that include user's posts and comments on various topics, including personal experience, life challenge, concerns, etc. To scrape the negative sentiment online, we extracted both submissions and their associated

comments and aggregated them by city and year. Due to the large size of Reddit data and computational constraints, we limited our analysis to 11 large cities in the U.S., which are Boston, Seattle, Portland, Austin, New York City, Houston, Los Angeles, Chicago, Philadelphia, San Francisco and Atlanta. Data extracted covers the period from 2009 to 2024. To quantify the sentiment, our calculation occurred in two steps. From the raw text data, we first calculated the total number of submissions, total number of comments, and total number of observations (submissions plus comments) for each city annually. Then we applied keyword-filtering and machine-learning–based text classification approaches to determine the number of submissions or comments that are negative. Finally, we could get the percentage of negative posts and comments out of all observations within a given city-year as the measure of negative sentiment across U.S. cities. We eventually had 183 units of observation for the dataset on a city-year level. Details of the keyword filtering and classification procedures are described further in the Methods section.

2.1.2 ACS

The American Community Survey (ACS) is a survey administered by the U.S. Census Bureau that provides annual estimates on demographic, economic, and social characteristics of the U.S. population. To include broader socioeconomic conditions that may affect digital emotional expressions, we extracted multiple social and economic variables of ACS from Census. For each metric, we searched for related terms like ‘income’, ‘population’, ‘education’ and then selected the most relevant metric for the model. The total metrics we selected include median age, sex ratio, average years of schooling, household income, unemployment rate, mean work hours and population density.By considering the issue of multicollinearity, we eventually selected four variables related to either economic or social factors, which are median household income, average work hours, average years of schooling, and population density. Median household income and average work hours reflects overall economic condition and labor market intensity for a certain city, which has an impact on people’s emotional expressions. These two metrics are directly extracted from the database, where we limited the time range from 2021 to 2024 across 11 large U.S cities. Since there are no direct data for education and population density, we applied feature engineering for these two metrics. To quantify education, we calculate average years of schooling as a weighted average of assigned years of education for each attainment category, using the population share of each category as the weight. To calculate population density, we extracted both the total population per city and the land area of each city from ACS directly. We finally aggregated the data to the city-year level to align with our Reddit dataset, making the unit of observation a city-year. In total, we got 44 units of observation for socioeconomic controls.

2.1.3 CountyHealthRanking

CountyHealthRankings.org is a data platform developed by the Robert Wood Johnson Foundation that provides annual measures of health outcomes and health-related resources across U.S. counties. It is widely used by professionals to evaluate access to care and local health system capacity. In this project, we utilized se County Health Rankings data to capture variation in mental-health service availability across U.S. cities. In specific, we extracted data on the number of mental-health providers per 100,000 residents, which serves as our key independent variable. This variable includes licensed professionals such as psychiatrists,

psychologists, clinical social workers, and counselors, and serves as a proxy for the strength of a city's mental-health safety net. Higher number of mental-health providers indicates greater access to state-specific mental health support, which may affect emotional distress and psychological sentiment online. Since the data are only reported in either state-level or county-level, we chose the state-level and then matched each city in our sample to its corresponding state and aggregated the data accordingly. The unit of observation is therefore a state-year, resulting in 44 state-year observations for mental-health service availability.

2.2 Data Merging: Final Dataset

After conducting cleaning, feature engineering, filtering and classifying the data, we merged each of the datasets together, resulting in a final dataset with city, state, year, negative rate, the mental health (MH) provider rate, average years of schooling, median household income, mean work hours and population density. The table is shown below.

City	State	Year	Negative Rate	Avg. Years of Schooling	Median Income	MH Provider Rate	Mean Work Hours	Population Density
Atlanta	Georgia	2021	0.014844545	14.87	74107	144.77148	39.4	3670.750372
Atlanta	Georgia	2022	0.014486142	14.95	83251	156.97454	39.8	3690.276741
Atlanta	Georgia	2023	0.012975745	14.91	85880	167.71970	39.4	3776.818259
Atlanta	Georgia	2024	0.012077368	15.05	88165	178.97207	39.2	3845.134674
Austin	Texas	2021	0.017981820	14.72	79542	120.86889	39.7	2953.748104

2.3 Data Limitation

One data limitation concerns the consistency of the geographic level of the data. While most variables in the analysis are measured at the urban level, the number of mental health service providers is only available at the state level. As a result, this variable must be mapped from the state level to the urban level for use in the analysis. This mismatch may introduce measurement error.

3 METHODS

3.1 Classification of Negative Sentiment

The goal of this section is to calculate the count of observations (posts and comments) that express negative feelings.

3.1.1 Gathering Primary Dataset

The first step was to extract relevant observations from the data for each city's subreddits. We utilized scripts provided by the original database to decompress and extract observations from .zst files using keyword matching. We applied a specific rule for the keyword search: observations were included if they contained either (1) one appearance of a direct word, such

as “depression” or “suicide,” or (2) two appearances of indirect words, such as “alone” or “tired.” (The full word list is available in the data folder) We presume that posts and comments filtered by these rules capture most content expressing negative feelings. However, because this method also captures many unrelated posts, a secondary filtering step was necessary.

3.1.2 Filtering Unrelated Posts

We filtered out unrelated posts using a RoBERTa-based machine learning model. First, we used the gpt-5-mini-2025-08-07 model via the OpenAI API to label a subset of 8,000 observations. After verification, we found the accuracy of the LLM labeling to be higher than 90%. We then built a machine learning model using RoBERTa-base. We chose the base model because it performed well on our data and was more resource-efficient than other pretrained models. As shown in the results, the model achieved an accuracy of over 90% and a recall rate of 97%. We used this model to label the primary dataset and filter out unrelated observations. From an initial 1,591,072 observations, we filtered out 617,486 unrelated posts, preserving 973,586 observations for analysis.

3.2 Analysis Methods

The analysis consists of regression modeling and structural topic modeling (STM).

3.2.1 Regression Analysis

We use OLS regression to reveal the causal relationship between our variables of interest. The dependent variable is the rate of posts expressing negative feelings for each city. The primary variable of interest is the doctor-to-population ratio for the state in which each city is located. We also controlled for several state-level variables: Average Income, Population Density, Education (years of schooling), and Mean Work Hours.

The specification is as follows:

$$\text{NegativeAffectRate}(i) = \beta 0 + \beta 1\ \text{ProviderRate}\ (i) + \beta 2\ \text{AverageIncome}\ (i) + \beta 3\ \text{PopulationDensity}\ (i) + \beta 4\ \text{Education(years)}(i) + \beta 5\ \text{MeanWorkHours}\ (i) + \epsilon i$$

3.2.2 Structural Topic Model Analysis

To investigate the reasons behind negative feelings, we analyzed the corpus of posts containing the word “depression” using a Structural Topic Model (STM). STM allows us to use regression to analyze the prevalence of each topic, showing how topics are distributed across different variables. We followed standard preprocessing procedures but preserved emojis, as they carry crucial emotional information in internet discourse.

The formula for the STM is:

$$\text{Topic Prevalence} \sim \text{source\_type} + \text{city} + s(\text{date\_numeric})$$

The source type variable indicates whether the observation comes from comments or posts. After testing we found 15 topics to be best extracting needed semitic information and thus it is picked as our STM's topic number.

4 ANALYSIS & RESULT

This section reports the core empirical findings of the study, integrating descriptive analyses with regression-based inference. We begin by documenting the thematic composition and temporal evolution of negative sentiment using Structural Topic Modeling and visualization techniques. In addition, we conducted the regression models to assess how mental health provider availability and key socioeconomic factors are associated with cross-city differences in negative sentiment.

4.1 Thematic Content of Negative Sentiment Posts

To better understand what our negative sentiment measure captures, we utilized Structural Topic Model (STM) to examine substantive themes from both the submissions and comments on Reddit. Figure A presents us with the top terms associated with each identified topic.

A large share of the topics reflect general psychological distress. Relevant topics include Clinical Treatment/Meds, Casual Reactions, Transgender Mental Health, having words like depression, anxiety,depressed, reality, suicide and risk. Specifically, several topics also involve help-seeking behavior, including discussion about medication, treatment, suicide resources, illustrating that people often reflect actual mental health challenges instead of casual dissatisfaction or complaints. Other topics point to specific reasoning. Themes such as Loneliness & Socializing and Suicide Philosophy emphasize social isolation and personal life challenges, specifically about difficulty finding friends, feelings of disconnection, existential reflections, family issue, etc. In addition, we also identify strong themes on financial pressure, topics such as Cost of Living/Housing and Politics & Economy include frequent mentions of money, rent, jobs, and economic conditions. These patterns suggest that financial pressures are salient contributors to negative emotional expression online, reinforcing the idea that mental health is shaped by broader socioeconomic contexts. Finally, there are also several topics related to city-specific issues. Topics include Seasonal Affective Disorder and Urban Planning & Transit, mentioning cities with long winters, bad weather or city's issues with neighborhood and downtown.

Overall, the STM results show us that the negative post we extracted did consistently reflect emotional distress,which is mainly about mental health struggles, social isolation, financial stress, and city-specific challenges.

Figure A: Themes of Negative Sentiment Posts from STM

4.2 City-Level Trends in Negative Sentiment

This project also examines the variation in negative sentiment across cities over time. Figure B shows substantial and persistent differences in negative sentiment rates by city and year. Darker colors indicate higher negative sentiment rate, which is consistently concentrated in a subset of cities, while lighter colors show a relatively lower level of negative rate. Across the entire period, the heat map shows a broad increase in negative sentiment beginning around 2020, with darker shading appearing across nearly all cities. This pattern indicates a shared shock affecting all locations, consistent with nationwide social and economic disruptions during this period (e.g. the pandemic). Specifically, Figure C shows persistent cross-city differences in negative sentiment rates by year, presenting the trends from top and bottom three cities. Cities such as Seattle and Portland consistently exhibit higher levels of negative sentiment, while cities such as Atlanta and Houston remain comparatively lower and more stable throughout the period. Even though all cities experience multiple fluctuations year to year, the overall ranking stays persistent. This persistence therefore suggests that negative sentiment is not driven by short-term shocks, but reflects more enduring city-level conditions that shape online emotional expression.

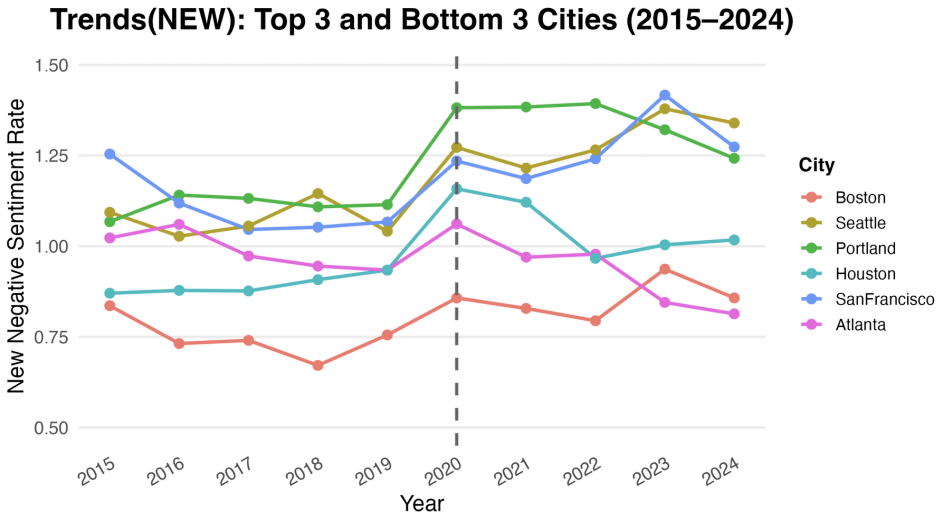

Figure B: Variation in negative rate for top & bottom three U.S. cities

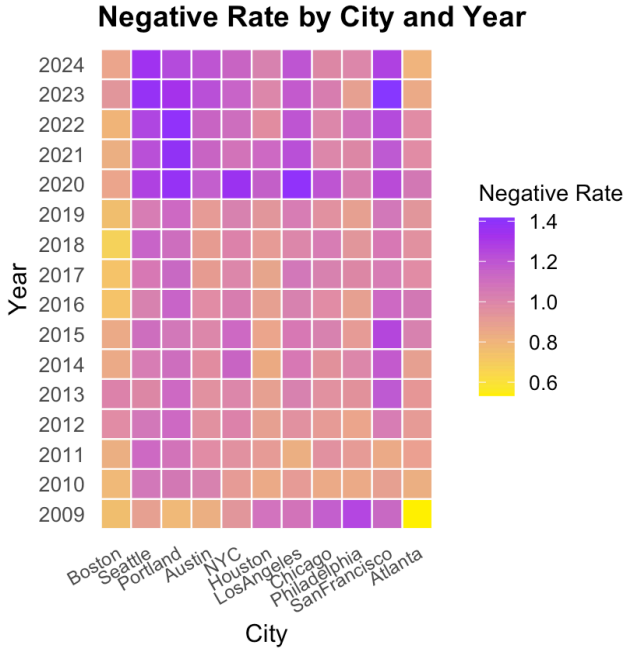

Figure C: Negative Rate for 11 large U.S. cities

4.3 Topic Network Structure and Thematic Clusters

From the network visualization in Figure D, we can see that different categories of topics form several clearly separated clusters. This suggests that discussions related to depression are distributed across several interrelated thematic areas instead of being concentrated on one core issue.

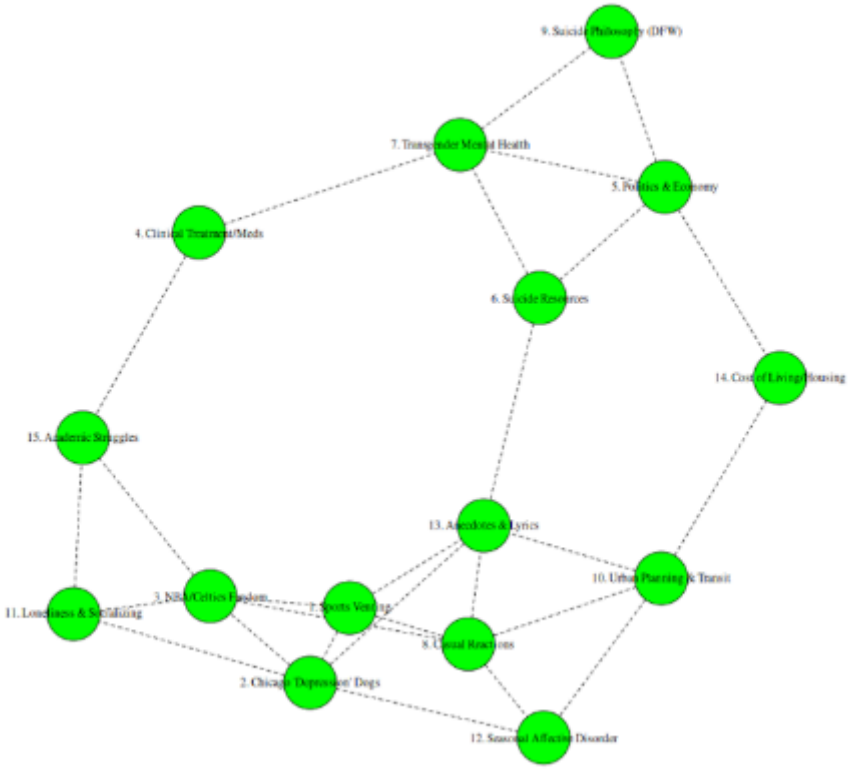

Figure D

A high-risk mental health cluster is in the upper region of the network, comprising topics such as Suicide Philosophy, Transgender Mental Health, and Suicide Resources. The dense connections among these nodes suggest strong thematic overlap between suicide-related reflection, population-specific mental health concerns, and help-seeking resources.
The second cluster covers socio-economic and environmental factors, including cost of living / housing, urban planning and transit, and seasonal affective disorder. The connections among these topics are intuitive and grounded in everyday experience: economic pressure, urban conditions, and environmental factors share a common semantic basis in the discourse and jointly reflect the role of external living conditions in shaping emotional distress.
In contrast, a cluster centered on everyday emotional expression is located in the lower portion of the network. This cluster includes topics such as Sports Venting, Casual Reactions, and Anecdotes & Lyrics.

Although these topics are closely connected to one another, they remain spatially separated from the high-risk mental health cluster, highlighting a distinction between informal emotional catharsis and more severe mental health discussions. This distribution of topics also indicated that our STM successfully captured the pattern of the topics.

4.4 Topic Dynamics over Time (2011–2022)

Several topics in the figure display distinctly non-linear temporal patterns. Politics & Economy fluctuates substantially over time, with visible peaks and declines, indicating episodic attention rather than a stable long-term trend. Seasonal Affective Disorder follows a clear U-shaped trajectory, declining during the middle of the period and rising sharply after 2019. Loneliness & Socializing and Academic Struggles exhibit similar dynamics: both decrease slightly in the early years but increase markedly in the later period, reaching their

highest proportions between 2019 and 2021.
In contrast to these late-emerging topics, suicide-related themes show earlier prominence followed by decline. Suicide Resources rises in the early years but steadily decreases afterward, while Suicide Philosophy (DFW) peaks earlier in the timeline and gradually diminishes.

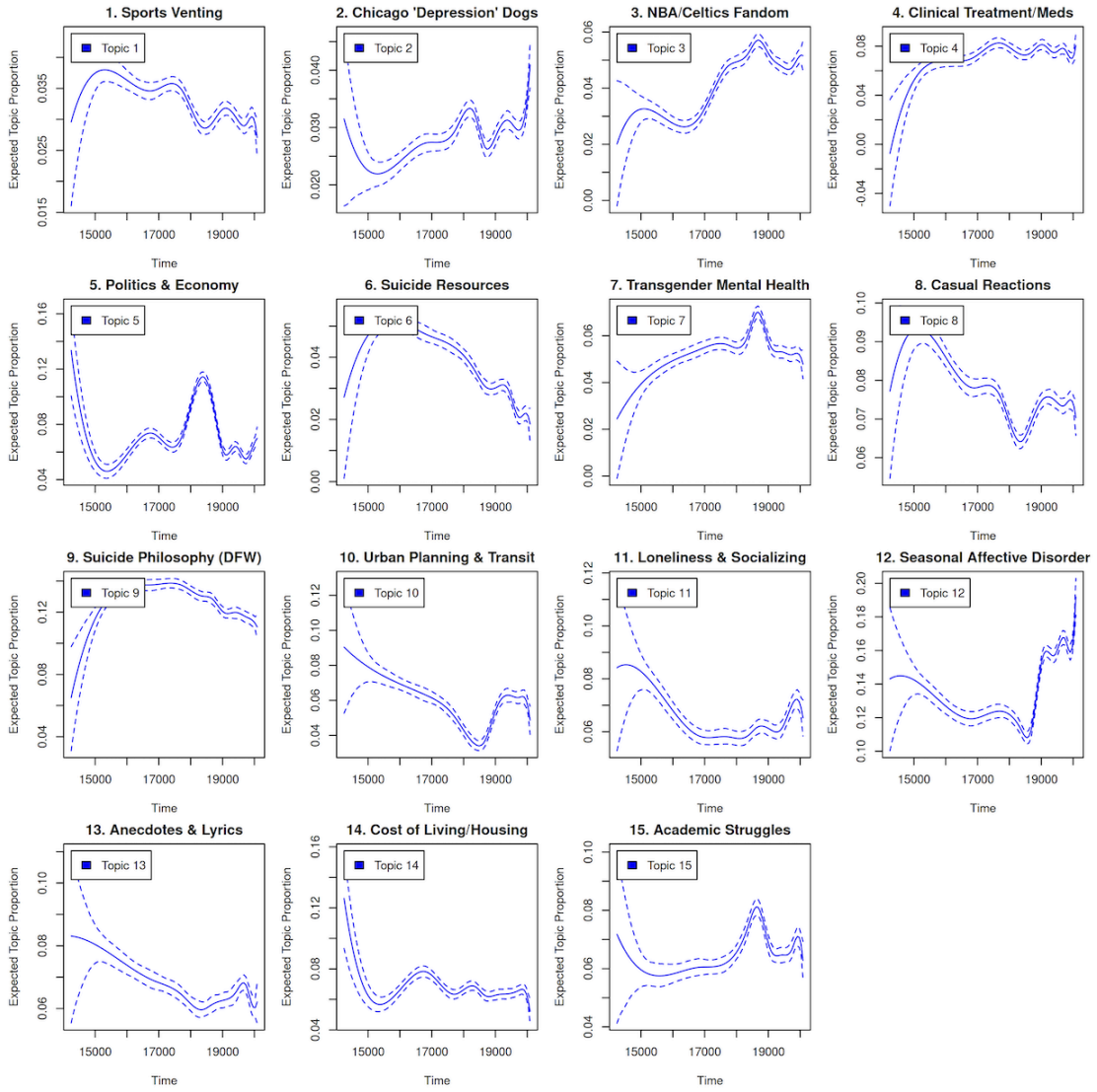

Figure E: Non-linear temporal patterns of different themes

4.3 REGRESSION RESULT

4.3.1 Regression Estimates

To move beyond descriptive analysis, we estimate an OLS regression that models the city-level rate of Reddit observations expressing negative feelings as a function of state-level mental health provider availability and key socioeconomic controls. Under this specification, the coefficient on the provider rate captures the association between mental health service availability and negative sentiment, conditional on socioeconomic characteristics. Most importantly, the coefficient on the provider rate, our main independent variable of interest, is negative and statistically significant.

Regression Results on Negative Sentiment

	(1)					
	coef	**std err**	**t**	**P > |t|**	**[0.025**	**0.975]**
const	9.7291	2.5273	3.8496	0.0004	4.6129	14.8453
density	-0.0000	0.0000	-0.4588	0.6490	-0.0000	0.0000
income	0.0000	0.0000	5.0700	0.0000	0.0000	0.0000
average_years_schooling	-0.1000	0.0577	-1.7327	0.0913	-0.2169	0.0168
mean_work_hours	-0.1919	0.0607	-3.1625	0.0031	-0.3148	-0.0691
mental_health_provider_rate	-0.0006	0.0003	-2.1100	0.0415	-0.0011	-0.0000

Regression Results on New Negative Sentiment

	(1)					
	coef	**std err**	**t**	**P > |t|**	**[0.025**	**0.975]**
const	8.6581	1.7224	5.0266	0.0000	5.1712	12.1450
density	-0.0000	0.0000	-2.5858	0.0137	-0.0000	-0.0000
income	0.0000	0.0000	5.5532	0.0000	0.0000	0.0000
average_years_schooling	-0.0838	0.0393	-2.1294	0.0398	-0.1634	-0.0041
mean_work_hours	-0.1783	0.0414	-4.3104	0.0001	-0.2620	-0.0946
mental_health_provider_rate	-0.0006	0.0002	-3.0995	0.0036	-0.0010	-0.0002

In regression analysis, the measurement method of dependent variables has an important impact on the estimation results. In the initial analysis, negative emotions are mainly identified based on keywords, but this method may misjudge semantically non-negative expressions as negative emotions in practical applications, such as being classified as negative only because of words like "depression" or "anxiety". This error will introduce additional noise in the dependent variable, thus affecting the stability of the regression result.

In order to reduce this measurement error, we introduce an emotional classification model based on BERT to identify negative comments more strictly. The reclassification results show that only about 60% of the comments that were originally marked as negative emotions do reflect negative emotions semantically. Based on this stricter definition of emotions, we recalculated the proportion of negative emotions and estimated the regression model again.

After adopting the improved emotional indicators, the regression results showed higher consistency. Some variables did not show a significant relationship in the original model, but after the emotional measurement error was alleviated, the estimated results became clearer.

4.3.2 Non-linear Patterns in Negative Sentiment

This figure highlights a clear non-linear relationship between mental health provider availability and the negative sentiment rate on Reddit. Each point is a city-year observation, and the smoothed curve summarizes the average pattern with a confidence band. At low provider levels (about 150–250 per 100,000 residents), negative sentiment is relatively high (around 1.8–2.0%). As provider availability increases into a mid-range (roughly 250–450),

negative sentiment does not decline monotonically and instead rises slightly. After provider availability reaches higher levels (above about 450–500), negative sentiment falls clearly and consistently. Overall, the data suggests that negative sentiment remains high when provider availability is low, with less change in the medium term, and a more dramatic decline in negative sentiment once provider density is high enough.

5 DISCUSSION

5.1 Conclusion

This project uses Reddit data to examine patterns of negative emotional expression across U.S. cities and over time. Using a topic model, we first document persistent city-level differences in negative sentiment. Although all cities experience a noticeable increase beginning around 2020, the relative ranking of cities remains largely stable over time. We then discover that depression-related discussions form distinct but interconnected thematic clusters: high-risk mental health topics are separated from everyday emotional venting, while socio-economic and environmental issues constitute a coherent cluster of their own. Finally, topic prevalence evolves in non-linear ways over time, indicating that mental health–related discussions on Reddit shift across periods rather than following a simple monotonic trend. About our regression, it is constrained by a limited sample size and endogeneity, and its results should therefore be interpreted with caution.

5.2 Limitation

The first and most important issue relates to the effective sample size used in the regression analysis. While Reddit data were scraped at the city level covering posts from 2009 to 2024 and allowing us to construct negative sentiment measures over many years, the key independent variables are only available for the period from 2021 to 2024. As a result, the regression analysis is restricted to these four years, yielding a total of 44 observations across 11 cities. In addition, we treat all negative emotions equally, although in fact, the intensity of different emotions varies greatly, and some negative expressions are obviously stronger than others. Ideally, negative emotions of different types or intensities should be given different weights.

5.3 Next steps

Looking ahead, future research should address two separate limits on the regression sample size. First, the current analysis covers only 11 cities, so we should scrape and include additional cities to expand the cross-city dimension and increase the number of city-year observations. Second, the regression is restricted to 2021–2024 because the key independent variables are only available for those years. A priority is to locate or construct comparable independent variables with longer time coverage (or alternative proxies), so the regression period can be extended beyond 2021–2024 and the effective sample size can increase substantially. In addition, sentiment measurement should be refined by classifying negative sentiment by type or intensity rather than treating all negative expressions equally. These steps would strengthen statistical inference and provide a more precise understanding of how mental health resources relate to online emotional expression.

Citation

[1] DONNELLY, RACHEL. Mapping Mental Health across US States: The Roleof Economic and Social Support Policies - Donnelly - the Milbank Quarterly -Wiley Online Library,onlinelibrary.wiley.com/doi/abs/10.1111/1468-0009.70015. Accessed 18 Dec. 2025.

[2] "Mental Illness." National Institute of Mental Health, U.S. Department of Health and Human Services,www.nimh.nih.gov/health/statistics/mental-illness. Accessed 17 Dec. 2025.