ASSESSING MENTAL WELLBEING IN URBAN AREAS USING SOCIAL MEDIA DATA: UNDERSTANDING WHEN AND WHERE URBANITES STRESS AND DE-STRESS

A Dissertation
Presented to
The Academic Faculty

By

Florina Dutt

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
College of Design
School of City and Regional Planning

Georgia Institute of Technology

December 2021

ASSESSING MENTAL WELLBEING IN URBAN AREAS USING SOCIAL MEDIA DATA: UNDERSTANDING WHEN AND WHERE URBANITES STRESS AND DE-STRESS

Thesis committee:

Dr. Subhrajit Guhathakurta, Advisor
City and Regional Planning
Georgia Institute of Technology

Dr. Munmun De Choudhury
Interactive Computing
Georgia Institute of Technology

Dr. William J. Drummond
City and Regional Planning
Georgia Institute of Technology

Dr. Clio Andris
City and Regional Planning &
Interactive Computing
Georgia Institute of Technology

Prof. Ellen Dunham-Jones
Architecture
Georgia Institute of Technology

Date approved: 26th July, 2021

What attracts people most, it would appear, is other people.

William H. Whyte

For my grand-mother Shanti Sudha

ACKNOWLEDGMENTS

I would like to thank my thesis committee members, Dr. Subhrajit Guhathakurta, my advisor, for letting me dive into this research topic; Dr. Munmun De Choudhury, whose research on social computing is an inspiration and methodological push behind this work; Professor. Ellen Dunham-Jones for helping me contextualize the research in urban design problem setting; Dr. William J. Drummond, for helping me shape the interpretability of the methods used; and Dr. Clio Andris, for pushing me towards a strong spatial analytic framework - which forms the backbone of this research. Everyone's help and assistance are pivotal in the preparation of this work.

Special thanks to my husband, Dr. Subhajit Das, who helped me with computational thinking, besides the encouragement and hand-holding in every step of the Ph.D. life. Lifelong gratitude towards my mother, Dipika Dutta, and my father, Amarnath Datta, for inspiring me in every step of my education, and supporting me to come this far.

I gratefully acknowledge the Center of Spatial Analytics And Visualization (CSPAV), Imagine Lab, School of City and Regional Planning Georgia Tech for all the assistance and funding support I have received to materialize this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AADT Average Annual Daily Traffic

ACS American Community Survey

API Application Programming Interface

BE Built Environment

BPL Building Per Length

CP Cross-sectional Proportions

CSV Comma Separated Values

EFA Exploratory Factor Analysis

FAR Floor Area Ratio

FHA Federal Highway Administration

GA Georgia

GIS Geographic Information System

HC Hierarchical Clustering

HPMS Highway Performance Monitoring System

LBSM Location Based Social Media

LDA Latent Dirichlet Association

LEH Longitudinal Employer-Household Dynamics

LIWC Linguistic Enquiry and Word Count

LRM Linear Regression Model

LiDAR Ariel Light Detection And Ranging

MA Massachusetts

MWS Mental Wellbeing Score

NCEI National Centers for Environmental Information

NLP Natural Language Processing

NMF Negative Matrix Factorization

OSM Open Street Map

POI Point of Interest

QRM Quantile Regression Model

RFM Random Forest Model

SES Socio-Economic Status

SWC Street Wall Continuity

TA Tax Assessor's Data

TCC Tree Canopy Coverage

VIF Variance Inflation Factor

WAC Work Area Characteristics

SUMMARY

Are Americans more stressed out by living in dense, urbanized areas or less dense, car-oriented areas? To answer this question, can we use people's expressions of stress in different environments to understand what kinds of spaces help them de-stress? This study uses stress levels of geolocated tweets to help us answer such inquiries and resolve the longstanding disparities between the field of psychology and urban planning about mental health impacts of cities.

This is important because more than 75 percent of Americans are moderately stressed. Long-term stress is associated with mental health disorders, including sleeplessness, anxiety, and depression. Additionally, chronic stress is linked to physical ailments, including high blood pressure, cardiovascular diseases, and diabetes. The psychology literature claims that urban areas witness elevated levels of mental health problems, manifested as stress, mood disorders, and anxiety issues. Density, crowding, traffic, crime, and pollution are identified as stressors associated with urban living conditions. Contending this claim, the urban planning literature positions stress in the context of longer commutes, lack of accessibility, and social isolation that comes with suburban living conditions. Urban Planners and urban designers have advocated for density. With rapid urbanization, 60 percent of the world population will live in urban areas by 2030, making it crucial for urban planners to address these disparities to support the mental wellbeing of the urbanites.

This research uses multi-headed attention transformer model to classify tweets (token sequences), and assesses the stress levels of custom-defined assessment grids of ten acres within the city area of Atlanta and Boston. The assessed stress level of these assessment grids is called the mental wellbeing score (MWS). Mental wellbeing score is defined in this research as a measure of 'mental wellbeing' of any given grid (higher score is better). Using this measure, the research investigates the relationship between mental wellbeing and built environment characteristics in urban areas to uncover the impact of long-term

stress triggered by the conditions of the built environment in urban settings.

In summary, the results of the exploration shed light on three critical aspects:

1. Mental wellbeing score increases with increasing urbanness.

2. The mental wellbeing score increases with the increase in the diversity of escape facilities, including green parks, open spaces, and other points of interest.

3. The mental wellbeing score is positively impacted by accessible high-density spaces with high symbolic value.

The research also investigates the impact of safety perception and socio-economic status on mental wellbeing scores. The results show that addressing socio-economic disparity, crime, and investment in green infrastructure can improve mental wellbeing of urbanites. The methods and findings of the research show that 'urban areas' can positively impact mental health if designed appropriately. Furthermore, this study can empower urban planners and policymakers to develop tools to assess the mental wellbeing of urbanites, adjust infrastructure needs, and improve the urban amenities that support mental wellbeing.

CHAPTER 1
INTRODUCTION

1.1 Problem Statement

Chronic (long-term) stress is associated with mental health disorders, such as insomnia, anxiety, and depression [1].[1] The duration of stress is also associated with suppression of immune function, including severe health ailments such as cardiovascular diseases and autoimmune disorders. The American Psychological Association (APA) identified long-term stress as a looming public health crisis in 2010 [2]. The United States of America spends on average more than 201 billion dollars yearly on mental health disorders with additional overheads for treating physical ailments associated with stress [3]. With rising mental health disorders, and with a possibility of one in four people to get affected by mental health and neurological disorders globally, it is essential that urban planners and policymakers to become aware of the conditions in cities that may exacerbate people's stress levels [4].[2]

The environment around us influences our mental health and wellbeing. The incredible power of a well-designed built environment and how it can prevent physical ailments have been extensively studied by Richard Jackson and other scholars. Jackson's collaborative work with Frumkin and Dannenberg is credited for bringing a significant shift in public health, and urban planning [5].[3] Similar to physical health, mental wellbeing is greatly impacted by the built environment. Sheldon Cohen et al. described psychological stress

[1]Reported by the National Institute of Health, which is a part of the U.S Department of Health and Human Services. NIH is the nation's medical research agency invested in medical research and making discoveries.

[2]The World Health Organization's projection on mental health cases globally.

[3]Dr. Richard Jackson has described his aha moment, understanding the importance of the design of the built environment while driving to the Center for Disease Control (CDC) on Buford Highway in 1999 or 2000. He then rallied Frumkin at Emory's School of Public Health and others in public health to focus on the implication built environment on human health. He focused on obesity at the time.

as the excess of environmental demand on an individual's adaptive capacity [6]. While the word environment may include a wide range of concepts, including one's life events such as the experience of growing up, socialization, and others, in this research, I am interested in those aspects of the environment that is planned and designed by urban planners, architects, and urban geographers. Although physical form and shape are the by-products of a planned environment, they are primarily built for social concerns. Halpern states – "design of physical environment has social consequences whether intended or not, and even the humblest construction acquires a socially ascribed meaning to it" [7]. As stress levels and analogous health problems are on the rise globally, it is crucial for us to understand the impact of the built environment factors on mental wellbeing.

An initial assessment of multi-domain literature in urban planning and psychology shows a well-established relationship between the built environment and mental wellbeing. Digging deeper, a striking disparity resurfaces from the theories. Starting from the 1930s until recent years (2010 and later), the literature on mental health and built environment presented by the mental health experts (in the psychology domain) have reflected that urban areas are responsible for mental health issues (refer Chapter 2). The seminal psychology literature points out that people living in urban areas witness an elevated level of mental health disorders compared to their rural counterparts [8]. The density, crowding, crime, noise, traffic, pollution, and life quality are identified as key stressors [9].[4] There has been convergence and divergence of the ideas in psychology and urban planning in the context of mental health in cities. Mental health experts claimed the cities negatively impacted mental health. Studies published in the 1930s indicated that cases of psychopathologies are concentrated in large cities [10]. The thought gained traction with the post-world war - II ideas of the 'American Dream'. Living in the suburbs became popular, and center cities were seen as 'dirty' and 'crumbling'. People left the old, walkable, and livable neighbor-

[4] The literature on mental health and built environment presented by the mental health experts (in psychology domain) have described that urban areas are responsible for mental health issues. The content on urban and suburban have changed over time, starting from the 1930s until 1970s and in the later years (2010 and later). The change in the research trajectory is not so prominent in the psychology domain.

hoods to the suburbs. Owning a car and house in the suburb became the status quo of the high and middle-income groups. The center city experienced blight. In 1956, the Federal-Aid Highway Act allocated funding to build highways thereby, transforming cities from walkable neighborhoods to free-flowing conduits of traffic [11]. A change in the thought trajectory was observed in urban planning and design. In the 1960s urban designers, and theorists spent time to uncover how urbanness/urban conditions have supported spatial cognition [12], and attracted social gatherings [13, 14].

Urban designers and urbanists such as William Whyte, Jane Jacobs, Jan Gehl, and Kevin Lynch illustrated that the vitality of any urban space is essential for individuals to be excited and willing to spend time in the public realm. They discussed different approaches to measure vitality. Whyte conducted observational studies on the streets of New York. He found people tended to strike up conversations on busy sidewalks frequently, and they flocked around small plazas, sitting very close to each other. Through these observational studies, he was able to show that people were attracted to places with more people [15]. Jacobs emphasized the importance of smaller blocks and diversity of use that attracts people, "increases eyes on the streets", and also reduces the fear of crime [16]. Ghel talked about the importance of visual opportunities on the street-front that attracts people. His observations reflected that people lingered on urban streets with attractive ground floor facades [17]. Lynch mentioned the importance of landmarks and the role these landmarks play in imparting the city's imageability to a visitor or a city dweller [12].[5]

Kunstler described suburbs as lonely and "neurologically punishing".[6] He wrote – "Placed in such an environment, even a theoretically healthy individual would sooner or later succumb to the kind of despair and anomie that we have labeled depression." Planning theorists also resonated with this notion [18]. Suburban living is associated with long

[5] The quality in a physical object that gives it a high probability of evoking a strong 'image' in an observer.

[6] James Howard Kunstler is an American critic and social critic. His book "The Geography of Nowhere" 1994 is on the history of American suburbia and urban development. In one of his articles called 'Big and Blue in the USA' (2003), he wrote how suburbia could lead to a sense of isolation, loneliness, and despair. The finding from the article is documented by Howard Frumkin, Frank Lawrence, and Richard Jackson in "Urban Sprawl and Public Health: Designing, Planning and, Building Healthy Communities."

commutes, lack of access to amenities, and social isolation [19]. Putnam highlighted the growing disengagement from public life, decline of communities, and erosion of social capital. He reported a steady decline in social gatherings since the 1970s, when more time was spent at work, in daily commute, and in the use of technology for various needs [20]. In recent year's studies show that social isolation and solitary confinement has distressed people of different age groups. Vivek Murthy, the current Surgeon General's concerns about loneliness epidemic, is alarming. According to him, the health impact of being chronically lonely is almost equivalent to smoking 15 cigarettes a day [21]. He primarily focuses on loneliness in the workplace. Still, others like Holt-Lunstadt, see the loneliness epidemic manifesting most in seniors and teenagers who live in the suburbs and are vulnerable to isolation when they cannot drive [22].

Urban planning and mental health literature are dichotomous in expressing how the built environment affected mental wellbeing. The mental health literature highlighted that sprawl allowed people to escape crowding, get closer to nature, and provide sanctuary for the daily hassles of life. On the other hand, urban planning literature pointed out that sprawl might negatively affect a person's mental health. A disproportionate share of household responsibilities between men and women, travel time spent in long commutes, transporting children to school, and after-school activities might disengage people from forming their social identity. Besides, thousands of acres of land were developed as suburban housing with no policy in place for forest land, farmland, and parkland preservation. While suburban living may seem analogous to living in nature, the natural areas that remain accessible to people living in the suburb are relatively less. Moreover, highways, broad feeder roads, vast parking lots, and rows of big box stores constitute the landscape that makes driving a significant part of suburban experience [19].

Although driving through country roads has been reported to be a stress-releasing exercise in surveys, research shows extensive driving causes psychological arousal that negatively impacts mental health. More extended periods of behind-the-wheel experience inflicting agitation, anxiety, and overall elevated stress level to people [19]. Traveling twice a day to work at peak traffic hours can be a chronic source of stress on individuals. Avraham Kluger elicited that the stressors involved in longer commutes are *unpredictability* and *lack of control*. Commutes that vary widely due to heavy traffic, adverse weather conditions, and road crashes are unpredictable and give a sense of lack of control [23]. Studies showed that long-term stress due to driving has serious health consequences along with road aggression and accidents. Stress due to longer commutes has negative implications for both the work and family life of individuals. Apart from the stress due to driving, other causes that may lead to depression (from chronic stress) include limited opportunities for physical activities, reduced social contact, an unpleasant built environment, and urban sprawl. Recent studies on urban design and mental health explicitly talk about the importance of accessible urban green spaces, active spaces (spaces near transit), pro-social, and safe spaces [24].

There was a convergence of ideologies in psychology and urban planning around the 1970s when a few authors in psychology contributing to the topics of the built environment and mental health suggested that the relationship between mental health and built environment is more complex than presented [25]. Some attributed the higher number of mental health cases in the cities to '*social drift*'. Social drift is the change in the social constitution of the city in the post World War II era when people higher in the socio-economic status moved to the suburbs while the impoverished continued living in the unfavorable conditions of the city centers. The poor citizens living in the city centers usually had very little access to mental healthcare. Thus, the higher counts of mental health issues are attributed to social drift. Moreover, various scholars contended that the perception of density, crowding, and crime might vary between cultures and demographic groups, and those variations do not affect mental wellbeing as universally as presented. They also highlighted the possibility of reporting bias in the surveys that may affect the findings [25]. Nevertheless, these theories have not been empirically tested on an urban scale, and there have been few changes in the thought trajectory in psychology. As described, the research published in recent years has either diverged from the ideas of urban planners and urban designers or has left unanswered questions on the role of urban areas in the context of mental health.

Thus the debate over detrimental impacts and the benefits of urban living remains unresolved. It leaves us to investigate further the optimal built environment conditions in the urban areas that will allow any city's residents to de-stress. Given that 60% of the world population will be living in the urbanized world by 2030 [26], it is crucial for urban planners to enrich their understanding of current urban conditions and the characteristics of places that elevate the stress level of individuals in urban areas. In addition, knowing where individuals express stress or where they de-stress may allow urban planners to incorporate place-making strategies that support mental wellbeing.

1.2 Study Summary and Objectives

This research addresses the existing disparity between urban planning and psychology (or psychiatric geography). I utilized the theoretical constructs of both urban planning and psychology to examine how urban stressors impact people's mental wellbeing. First, using the methodological constructs of social computing, I measured the stress level of individuals using their social media micro-blogs or tweets; following that, I analyzed the association of mental wellbeing score with various urban built environment factors. I have focused on a diverse set of built environment factors that may act as long-term stressors, including lack of tree cover areas, lack of access to transit options, fear of crime, and others. Long-term stressors are usually difficult to identify, and we have a limited understanding of the influence of these stressors associated with urban living. My main research question is:

- *Despite the limitations of tweets/social media data, what can they tell us about the relationship between people's expression of stress and the proximate built environment characteristics in urban areas?"*

1.3 Chapter Outline

In Chapter 2, I provide a synopsis of the current literature on mental health and built environment. In Chapter 3, I enlist my research goals and questions. Following that, Chapter 4 delineates different data source, framework for computing built environment, and deep learning models that were constructed/trained to assess mental wellbeing spatially. In Chapter 5, I describe the statistical methods, and the results from the data analysis that answers my research questions. Finally, in Chapter 6, I discuss the findings, highlight the limitations in the methods adopted, and summarize potential future research that can stem from this work. The methods and findings of this research are likely to help urban planners and policymakers develop tools to assess the mental wellbeing of urbanites, make adjustments in the infrastructure needs, and empower decision-makers to improve the attractiveness of a place from a mental wellbeing perspective. In addition, I envisage that the study can guide incremental changes in urban places that lack spaces to support the resident's mental wellbeing. Although I have investigated urban areas (the city of Atlanta and Boston), in the future, I think similar methods can be applied to other geographic regions comprising of both urban and rural areas.

CHAPTER 2
LITERATURE REVIEW

This section highlights various studies in numerous domains exploring the relationship between mental health and built environment, the methodologies used, and their key findings. Refer Figure 2.1 to follow the timeline of the literature on the built environment and mental health. In this figure, a chronological order of the literature is provided to represent how it evolved over time. The black dots indicate studies that belong in the psychology/medicine/mental health domain, and the gray dots indicate that studies are in the urban design/urban planning domain. The section also includes a body of contemporary literature on social computing that uses state-of-the-art natural language processing and other machine learning techniques to determine the stress levels of individuals indicated by their social media microblogs.

2.1 Stress

Stress is a psychological reaction of individuals when they perceive the environmental demand exceeds their adaptive capacity [6, 27]. The study of psychological stress focuses on the occurrence of environmental events or on the individual's response to the events. Since the aetiology of mental wellbeing is too broad, we clarify that this research will be only focusing on stress as an indicator of mental wellbeing. Stress is different from mental illness or any serious psychological disorder. According to psychiatry, being profoundly unhappy and dissatisfied with a situation is not enough to provide a diagnostic label of mental illness. However, exposure to chronic stress is extremely harmful as it can result in permanent changes in the emotional, physiological, and behavioural responses that predispose an individual to both mental illness and physiological disorders.

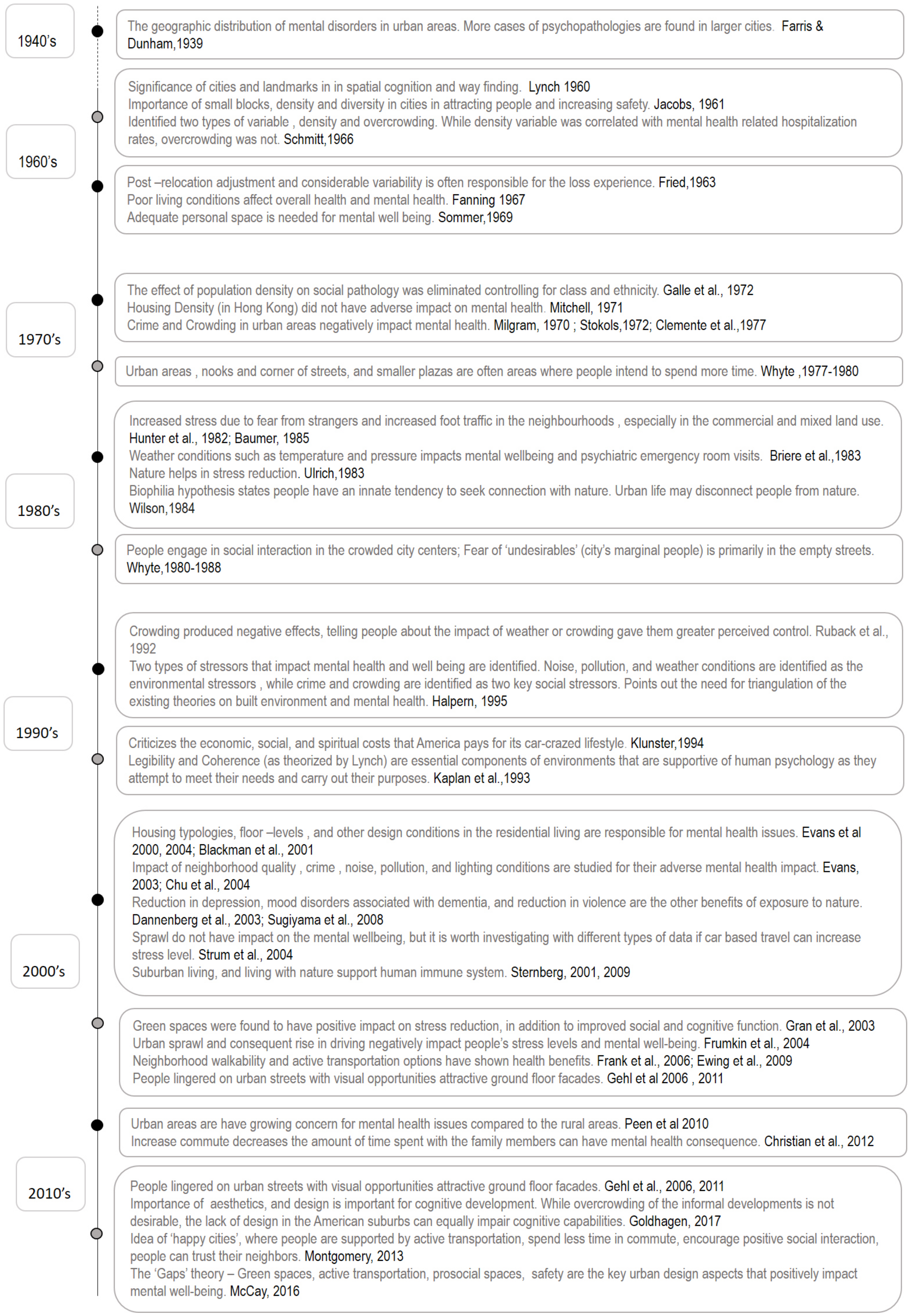

Figure 2.1: The figure shows chronological list of the seminal literature. Black dots for psychology and mental health. Gray dots for urban planning and design.

This research intends to address situational emotions (affect), feeling of anxiety, dissatisfaction, unhappiness, and even feelings of depression, causing a negative affective state. Negative affective state comes and goes periodically and is a normal part of everyday life. The negative affective state (measured by the length of the time) or stress determines whether a person is vulnerable to any serious mental health conditions. Halpern describes the stress and mental health issues as a maladaptive coping process to maintain an overall happy mental state [28] (refer Figure 2.2).

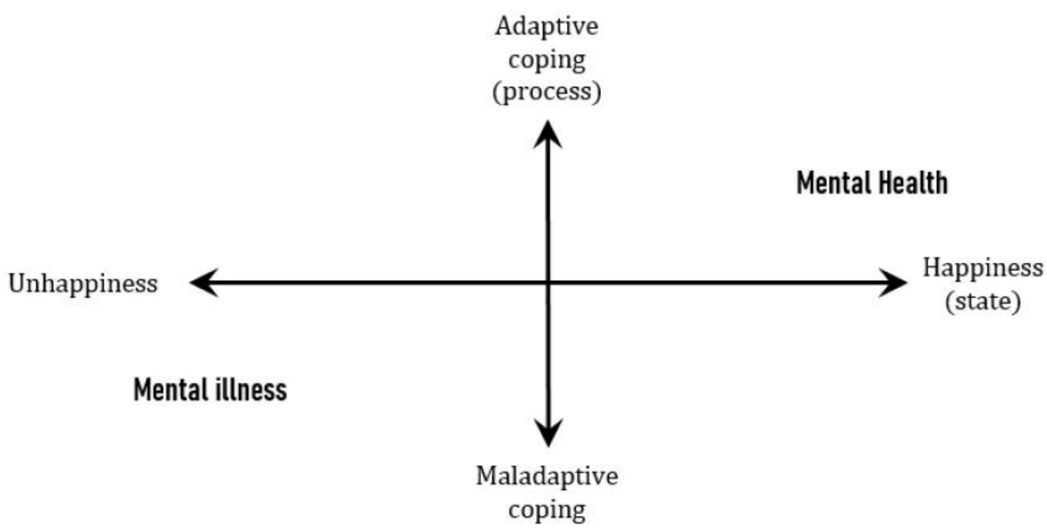

Figure 2.2: Diagram shows the relationship between psychological distress and mental wellbeing [9]. The horizontal axis shows the mental state, and the vertical axis shows the coping style. If an individual is unhappy and has a maladaptive coping style for prolonged periods, it could lead to mental illness.

The effectiveness of the coping strategies also depends on the genetic dispositions, life events, and environmental (or non-constitutional) variables. The debates on mental health and wellbeing revolve around the importance of the environmental factors, life events, and genetic factors in the onset of mental illness [29]. Studies show that environmental factors, such as daily hassles, can predict psychological symptom levels better than life events or genetic factors [30]. Although *life events* and absence of supportive *relationships* are identified as major causes of mental health issues, minor neurosis and psychological distress are also strongly associated with the socio-economic status, unemployment, weak or impoverished social network. Secondary factors, such as constitutional, social, and cultural variables, are responsible for determining the form and severity of the distress. As the broad range of etiologies of mental health includes many social and constitutional variables, the built environment variables only account for a narrow group of factors affecting an individual's mental health. Nonetheless, I think that the factors should be clearly understood and explored as complementary explanations, rather than alternative explanations of mental health issues [3].

2.2 Perspectives in Psychology

This section of the literature review presents a summary of the findings in the domain of psychology and closely related fields. The variables presented in the Table 4.1 are primarily those identified from the psychology literature.

2.2.1 Stress and Environment

Mental health and the built environment share a complex relationship. The environment can be a source of happiness as well as a source of irritation or annoyance. Stress forces an individual to adjust to the current state of the environment, and the *stress reaction* is the behavioural outcome of the adjustment process [3]. The physiological component of stress is described as the body's response by mobilizing its coping abilities [31]. The bodily changes triggered by the sympathetic *adreno-medullary system* inhibit certain non-prioritized activity in the body, such as digestive activity, and speed up metabolic processes preparing it to fight the condition. The longer the adaptation to the stress due to the threat from the environmental condition, the longer the body is driven by *hypothalamo – pituitary adrenocortical system*. In doing so, the body maintains a higher metabolic rate and blood glucose level, thus lowering the activity level of the body's immune system [32]. If the stress reaction keeps repeating for a prolonged period, the individual can enter a state of

Table 2.1: Different types of stress in the lifetime of a person [9].

	Minor	**Major**
Acute	Daily hassles, e.g., traffic, problem with work, etc.	Significant life events, e.g., death of close friend or family
Chronic	Environmental stressors e.g., noise, pollution, crowding, etc.	On-going difficulty, e.g., long-term health issues, poverty, etc.

exhaustion, and his/her adaptive reserve may deplete faster, causing psychological breakdown, making the body susceptible to other diseases [33]. This is also known as *general adaptation syndrome (GAS)* [31]. The stress outcomes are dissatisfaction, mood disorders, or anxiety based on the severity of the stressor. The stressors can vary in severity and can be chronic or acute. The more severe outcomes of stress are clinically recognized in psychiatric disorders (refer Table 2.1).

2.2.2 Classic Environmental Stressors

Halpern classifies environmental stressors into two types: type 1 is *classic environmental stressors* such as season, heat, wind, and pollution, and type 2 is *social environmental stressors*, such as crowding and fear of crime. The classic environmental stressors fall under the minor and chronic categories such as weather, daylight, temperature, pressure, wind, ongoing noise, and pollutants. They are primarily physical causes, devoid of any social meaning. Some of these stressors can be controlled or aggravated by the built environment design. It is hard to disentangle the effect of these variables individually. To that end, dimension reduction techniques such as *factor analysis* are useful to assess the impact of environmental stressors on individuals. For example, Briere et al. found a higher temperature, lower pressure, and lower sunlight were associated with poorer mental health conditions of individuals [34]. While the tolerance threshold for these environmental stressors between places may vary, the direction of the correlations does not change much. For example, a hot day in Atlanta may be unbearable for someone in Boston. Thus, the stress experienced by the individual living in Boston may be more than his counterpart living in

Atlanta. However, a hot day above the tolerance threshold for both individuals may equally increase their stress level [35] (refer Figure 2.3).

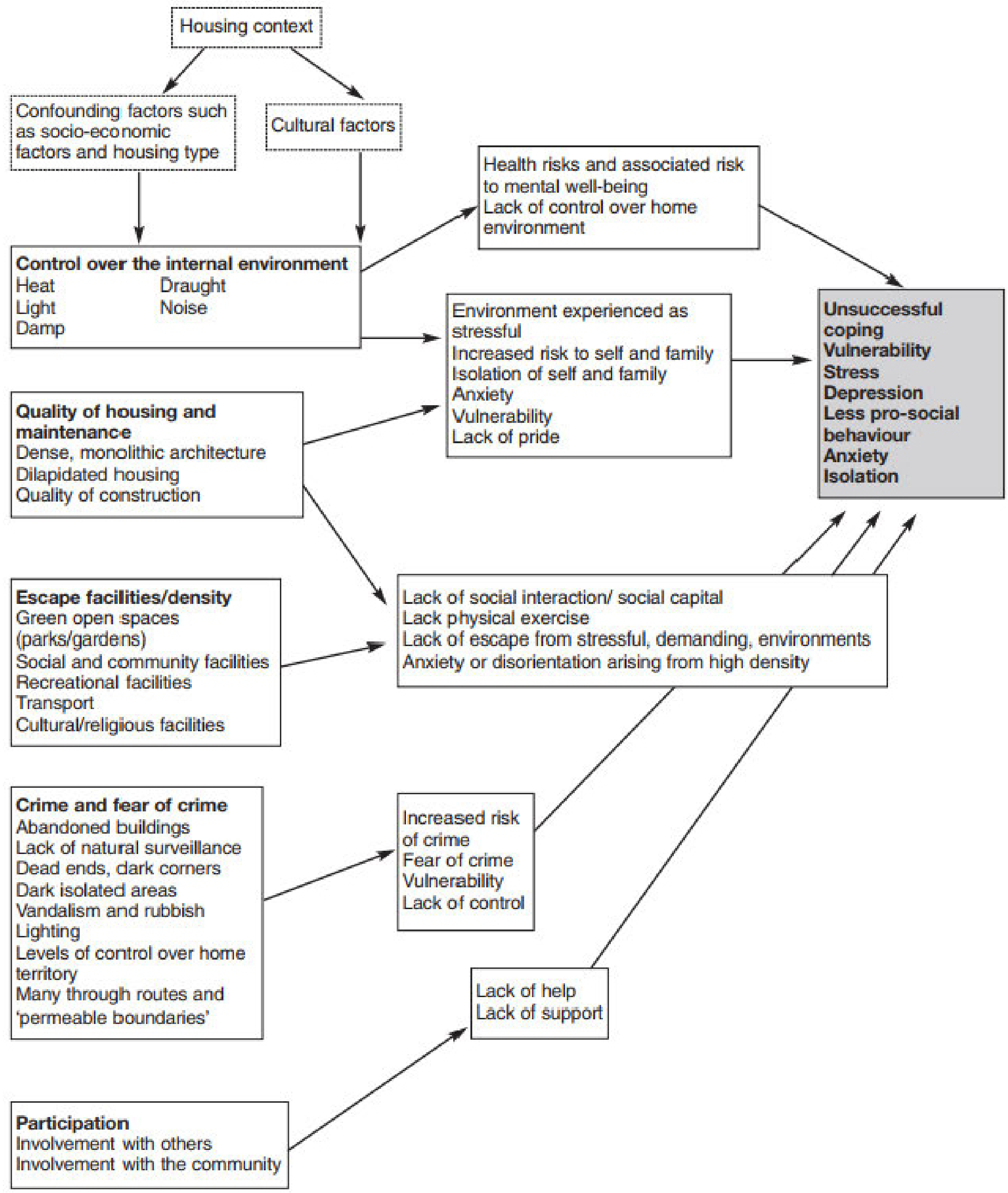

Figure 2.3: The summary of different literature showing the negative impact of built environment on mental health. The figure is presented in the review of urban mental health literature by [36].

2.2.3 Social Environmental Stressors

The second group of identified stressors is *social environmental stressors*. Halpern points out two major type of social stressors: *crowding* and *fear of crime*.

Crowding (density): Research on household density and mental health show many contradictory findings. Halpern [9] pointed out that the findings are contradictory due to the underlying assumption of linearity (that is, a linear relationship between density and mental wellbeing). Literature from the early 1900's such as, Faris and Dunham [10], Schmitt [37], indicates population density is positively correlated to the various measures of social pathology. Galle et al. [38], in their study of 75 community areas of Chicago, found that the effect of population density on social pathology was eliminated, controlling for class and ethnicity. Mitchell [39], in his study of Hong Kong, one of the most densely populated city in the world, found only a few adverse effects of population density. In contrast, studies from Britain showed that a higher population density was associated with psychosomatic disorders ($r=.123$, $p<.0001$), and controlling for age, sex, income, and noise sensitivity did not change the association. However, further assessment of the results suggests that most of the symptoms were mediated by the behaviour of neighbours (where neighbours were seen as less friendly). Another explanation is that the survey did not capture certain characteristics of dense urban neighbourhoods that can influence mental health. Milgram [40], pointed out that the fear of crime discourages people from engaging in interactions with a large number of people, including neighbours. Although there is an association between residential/population density and mental health, the pattern is not invariant and does not seem to hold across different cultures [41]. At a later time, researchers discovered that household density (number of members in a household) is the cause of mental health issues. Research on household density and mental health show many contradictory findings.

People living alone have a low household density, yet that is not an optimal condition for mental health. Being socially connected is not only influential for mental wellbeing, but studies show that it aids in dropping the mortality rate significantly. Holt-Lunstad et al. claim that social isolation results in a higher likelihood of mortality [42]. Culture also plays a vital role in the perception of crowding and the severity of mental health ailments that may have been caused by crowding. It is also important to note that a higher number of non-kin members may have an adverse influence on mental health, while a greater number of family members are not equally detrimental to mental health. Thus simple measures of association to examine the relationship between density and mental health can be misleading [25].

A study of crowding conducted in a dorm of a North American university shows that dormitory designs with a more socio-petal design called for unwanted social interaction, and the residents were found more hostile and withdrawn from each other [43].[1] Unpredictable and unwanted encounters with strangers usually lead to the perception of crowding and stress. Zukerman et al. [44] also found *perceived crowding*, and negative mood is highly correlated. Smith and Haythorn [43], 1972 found that the sense of crowding once again is related to a *lack of control on the environment* when incompatible personalities had to meet. On the other hand, a group of compatible personalities did not withdraw, and they increased their time spent together in recreational activities. Four different scenarios of crowding are: 1) lack of space, 2) division/dilution of resources, 3) proximity to others, and 4) unregulated interaction.

1. *Lack of physical space* may cause discomfort but may not cause crowding or crowding-related stress, e.g., an individual moving to a smaller but single room.

2. *Division/dilution of the environment's finite resources* due to a large number of users may increase the perception of crowding. In this case, crowding is associated with anxiety arising from frustration and aggression. Mutual understanding and cooperation play a big role in offsetting the impact of limited resources. Family is a good example where people may not feel the impact of crowding as there is mutual under-

[1] Socio-petal design encourage people to face each other and encourage more communication. Socio-fugal design tends to keep people apart and does not encourage communication˙ff

standing and cooperation, while it is different in a roommate scenario in a dorm.

3. *Close proximity to others* is a preferable condition with a close friend or family member, and this may not be considered as crowding. Closeness to unfamiliar or any hostile individual may generate negative feelings such as safety concerns and invasion of private space. Often in the desire for a safe space, people may show signs of withdrawal from any positive social interaction, thus causing a potential source of stress.

4. *Unregulated interactions* occur due to design fallacy, which may impart a sense of social strain/stress on individuals.

Fear of Crime: The other aspect of stress from the social environment is the fear of crime. Areas with a high concentration of crime tend to show different types of mental health disorders or social pathology [45]. Americans are advised to change their usual behaviour, such as staying off the street at night, avoiding strangers, reduce social activities [46]. According to Gelder et al. [47], a simple phobic neurosis has three components: first, the symptom of anxiety; the second, the anxious thoughts caused due to the anticipation of the situation the person may encounter; third, the habit of avoiding a situation that aggravates the fear. The main psychological consequence of an occurrence of crime is fear. The effect of fear can lead to a chronic state of anxiety. The constant fear in individuals may lead to compromises in their lifestyle choices. White et al. [48] found a negative association with the perceived crime level and mental health, in his study with 337 Black and Hispanic women, using different scales[2]. Halpern points out that the fear of crime can be a cause of social withdrawal, depriving people of positive experiences and undermining the local support network.

The fear of crime is due to three broadly classified reasons: the experience of crime,

[2] The measurement scales are: the Hopkins Symptom checklist, the Zung Depression Scale, and the Srole Anomia Scale.

individual vulnerability factors, and *incivilities* or degenerating condition of the neighbourhood. Individual physical and social vulnerability usually aggravates the fear of crime. The individual vulnerability factors include both physical and social vulnerability. Physical vulnerability factors include age and gender; social vulnerabilities include race, income, and even the size of the city. For example, studies show that women are more afraid of crime than younger males. Older people feel more vulnerable to crime. However, the results vary depending on the context and the specificity of the survey questions. Sometimes the size of age differentials is inflated in the crime surveys with vague questions [49]. In most cases, the two variables: age and sex, show statistical interactions. Income has an independent effect on the fear of crime, and poorer individuals are more fearful about the possibilities of crime [46, 50, 51]. Belonging to an ethnic minority group is highly correlated with low income and inner-city living.

Physical incivilities and signs of deterioration of an area are marked by the presence of vacant property, unmaintained housing stock, graffiti, etc. Social incivilities refer to the social deterioration signs characterized as unwanted gangs, vagabonds, and visible signs of criminal activities. Both physical and social incivilities are strongly correlated with the fear of crime. It is found that non-victims in an area with low perceived safety (or high incivility) have a substantial impact of fear in their life activity (behaviour), compared to the victims in the areas with high perceived safety (high incivility) [51]. Maxfield studied three contrasting neighbourhoods of San Francisco, where he found that the fear of crime is the most in areas where crime is perceived as a problem. Thus the impact of incivilities is marked in the areas where incivilities are associated as a visible decline and not necessarily associated with high crime [51, 52]. Studies also show that a high level of perceived incivilities is correlated with depression [53]. Thus, incivilities may not be a problem themselves; but they raise the fear of residents and thus increase their stress and anxiety. Besides the fear of crime, signs of physical incivilities may also raise concerns about gentrification in the low-income neighborhoods [54].

While Jacobs argued neighborhood safety perceptions depend on the active street fronts, and pedestrian's eyes on the streets [13], the sense of perceived safety hinges on factors such as length of residency, and social integration of the residents (here social integration is measured by their agreement on neighborhood issues).[3] Most researchers found an association between fear of strangers and increased foot traffic in the neighborhoods [55]. Social Integration mitigated such fear. As residents came to know each other better, they could recognize more faces in the neighborhood, reducing the number of strangers [56]. Unlike Jacobs [13], Fowler's [57] study shows the association between the concentration of use and fear of crime. The measure of concentration of use is the number of people using an area for residential, work, shopping, or recreation. The association of fear and concentration of use was found to be stronger (r=.47) than the association of crime and concentration of use (r= 0.26).

Halpern [9] talked about an alternative perspective where he proposed researchers to consider both perspectives of the residents and the pedestrians. For residents, the presence of strangers in their residential neighborhood may instigate the fear of crime, but from a pedestrian's perspective, a crowded street might increase their perception of safety. He explained that the distinction between *seeing a stranger* and *being a stranger* might explain the apparent controversy between Jane Jacobs's postulate of "*eyes on the street*" and the findings by Hunter and Baumer [55] or other researchers. If the survey questions are designed slightly differently, asking about fear of visiting places rather than fear of residing in their own neighborhoods, the findings will be different. Vrij and Winke [58] found that people perceived quiet, deserted, and poorly lit areas to be unsafe. Their second study demonstrated that increasing the level of lighting in a neighborhood with a high crime actually improved the perceived safety of the pedestrians. They also emphasized that the maintenance of the built environment in the neighborhood matter. Light played a big role in the perception of safety and in reducing the fear of crime (refer Figure 2.3).

[3] Jane Jacob used the term 'eyes on the street' to explain the importance of people on the street in improving the vigilance and consequently the sense of safety.

2.2.4 Built Environment of Housing and Neighbourhoods

The mental health and built environment literature specifically point out that built environment characteristics of housing and neighborhoods influence mental wellbeing. Evans and Chu et al. [36, 59, 60] summarizes how the built environment of housing and neighborhood affects mental health and wellbeing directly and indirectly. Direct effects of the built environment are the stress or relaxation experience, while the indirect effects are experienced when the psychological processes are altered, and people suffer from severe mental health issues such as anxiety, insomnia, or depression (refer Figure 2.3).

High-rise buildings: 'High rise' buildings are associated with stress. Particularly high-rise buildings for low-income families have insufficient resources to support spaces for the development and maintenance of a social network. The residents on the higher floor levels allegedly experience more significant mental health issues. For example, women and commonly single mothers with children report more loneliness, lack of territorial control over their surroundings compared to women with similar backgrounds living in other types of housing [59]. The incidence of higher mental health issues in high-rise buildings is presumably due to self-selection, i.e., people with pre-existing mental health conditions may choose to live in high-rise buildings to avoid social interaction. Another interesting aspect of high-rise living is, although it has been associated with stress yet they are highly sought after possession. It is gradually becoming a status quo for high-income living styles in the suburb [61].

Housing quality: Housing quality, such as its structural quality, maintenance and upkeep, and amenities (such as central heating, private bath), are positively correlated with mental wellbeing. Studies show that sense of insecurity is attached to poor quality housing [62, 36]. The residents living here are poor and worried about their tenure. Besides, housing is associated with an assigned valence or label (positive, negative, good, or bad). A label is given either by residents or by outsiders. The label is similar to stereotyping, which controls the behaviour, perception, and cognitive map of individuals living in that housing, as well as those of outsiders. The individual's stress level is dependent on the label of the area, or housing [9].

Neighbourhood quality: A growing body of literature exploring how neighborhood quality affects mental health. The neighborhood quality influences families irrespective of their socio-economic status (SES). However, Evans points out that it is hard to disentangle the quality of the housing (residential) units from the context of the neighborhood in which it is situated. Poor quality of housing is more likely to be situated in neighborhoods with multiple signs of urban decay. These studies do not use any of the physical attributes of neighborhoods but define neighborhoods as an index of social attributes. It is worth mentioning here that neighborhoods with access to well design green spaces scored higher in the social attributes [9].

Social participation: The amount of psychological comfort experienced by people is positively associated with improved community perception in the neighborhood [63]. The psychological variables, such as a sense of community, community satisfaction, and resident participation, play an essential role in shaping those perceptions. It is seen that a healthy neighborhood environment provides more opportunities for social integration than unhealthy environments that impose a threat to the safety and security of individuals [36].

2.2.5 Key Findings from Psychology

Previously, research in this domain has included weather conditions, indoor physical conditions, residential building conditions, residential building typologies (including the impact of a high-rise building), and a vague idea of neighborhood quality as key factors. There are very few studies identifying a clear pathway between urban built environment features constituting the public realm. The studies show no clear differentiation between the impact of the public and private realm while considering built environment variables. Besides, the measures of the built environment variables are not objective. They are usually a subjective measure of the perception of individuals surveyed as discussed in Figure 2.3.2. The subjective measures of crowding and the degree of blight in the built environment are usually biased and cannot be effectively used in creating design or policy guidelines. There is little in-depth research on the pathways of social interactions. While social spaces (escape facilities) allowing social interaction play an important role in social participation, the literature also point out that mixed land use and active street fronts have a negative impact on mental health [9, 64]. Measures of certain key variables causing stress, such as 'crowding', resurface in the stress studies from time to time and are often kept vague. Crowding is not differentiated from the phenomenon of overcrowding.

2.3 Perspectives in Urban Design and Planning

This section highlights the findings of the urban design and urban planning domain. Also, refer to the timeline in Figure 2.1 to understand how they evolved since the 1960s. Besides, this section touches upon the qualities of urban space that promote mental wellbeing. The controversies that connect urbanness to mental health in cities and the characteristics of urban space that support mental wellbeing are the two key aspects discussed here.

2.3.1 Impact of Urban Areas on Mental Health

The literature highlights that mental health issues are more common in cities compared to those in the countryside. Robert Gaupp reported that alcoholics, psychopaths, epileptics, paralytics, hysteria, organic brain patients are more frequently found in the cities than in rural places [65, 66]. The underlying causes of such mental health issues are tied to unhealthy living conditions in cities in the early 1900s. Karl Jaspers wrote in 1942, "the difficult urban living conditions with their harmful mental impact eventuate in a much greater frequency of psychopaths [67]". Increased mental health issues in today's cities are associated with increased traffic, pollution, noise, increased exposure to artificial light, and limited access to green areas. Homelessness and drug abuse are prevalent in cities that further aggravate mental health issues. Most city governments have not developed clear strategies for urban mental healthcare [66, 68]. The literature has primarily highlighted the defects of urban living over rural/countryside living. It fails to capture how runaways congregate in cities to escape abusive home conditions, only to then find themselves succumbing to addiction [69]. While cities may fail to cater for the runaway youths and homeless adults they are still a lucrative choice. Beyond allocated state budgets to support them, there are other system of social support.

Literature highlights that most people experience some symptoms of mental health issues in their lifetime, and one in every four may have diagnosable mental disorders [7]. Urban dwellers have 40% increased risk of depression, 20% more anxiety, and risk of schizophrenia [8]. However, blaming urban areas for their increased association with mental health disorders without further exploration is too naive [24]. Various authors have noted that increased mental health disorders in the urban areas can be attributed to the 'social drift' in the mid-1900s and post-world war II period [9]. People in the lower socioeconomic status (SES), unemployed, minority groups gravitated to the cities for the pursuit of better prospects and economic stability. Furthermore, the affluent urban residents moved to the suburb in the quest for a better quality of life. The blighted urban neighborhoods housed those in need of social support, healthcare, or mental health care facilities. Halpern [9], and McCay [24] both pointed out that the trend might be responsible for an increased baseline risk factor for mental health disorders in urban areas.

2.3.2 Urban Design Strategies to Improve Mental Health

With the current rate of urbanization (the world would be 68% urbanized by 2050, the United Nation's projection [70]), it is pertinent to focus on improving the urban living conditions. It is unreasonable to discard urban living considering the benefits outweigh the detriments [19, 20]. The literature on urban design and mental wellbeing highlights ways to improve the living conditions in the city through appropriate planning and design of urban areas. McCay L. describes the untapped potential of cities to improve mental health through various design exercises. The solutions highlighted by McCay extend beyond provision for urban green spaces [24]. According to him, promoting physical activities with sidewalks, bike paths, outdoor training facilities gives opportunities for stress reduction in individuals. Other examples of spaces that help in alleviating stress are places where people feel safe to experience nature, get to meet other people, and engage in social interaction.

Figure 2.4: Four key opportunity areas for good mental health, summarized by the acronym GAPS:
***G**reen places, **A**ctive places, **P**rosocial places, and **S**afe places.*

GAPS Theory: *Centre for Urban Design and Mental Health (UDMH) identifies* four key opportunity areas for good mental health in urban areas: ***G**reen places, **A**ctive places and, **P**ro-social places, and **S**afe places* (GAPS, refer Figure 2.4).[4] The four criteria provide a relatively clear guideline for urban designers and planning professionals. They help to identify the opportunity areas that may contribute to the mental wellbeing of the citizens.

Green places - Green spaces provide access to natural settings such as wilderness areas, parks, etc. Such spaces encourage participation in outdoor physical activities and social interactions. The higher the accessibility to green spaces, the greater is the possibility to integrate social activities in people's daily routines. Green spaces assist in strengthening social networks, cater to the biophilic nature in humans, have a restorative impact on the human mind, and positively impact physical health, all of which are necessary for mental

[4]Center for Urban Design and Mental Health is founded in 2015 by Laya McCay psychiatrist and adjunct professor of International Health at Georgetown University, https://www.urbandesignmentalhealth.com/how-urban-design-can-impact-mental-health.html

wellbeing.

Active spaces - Regular outdoor activities improve mood, general wellbeing and may cause positive mental health outcomes. There are plenty of opportunities to design cities that provide safe and active transport to activity areas and can integrate exercise and social interaction in people's daily life. Areas closer to transit and accessible by multimodal transportation are known as active spaces. Active spaces introduce a sense of agency in people's daily routines, promoting overall mental wellbeing.

Prosocial space - Opportunity to safe and natural interactions among residents, promotes a sense of community, integration, and belonging. The residents include vulnerable groups such as refugees, migrant workers, young and older adults, as well as other minority groups. These are areas that allow healthy social interactions (in addition to parks and green spaces), such as eateries, theaters, bowling alleys, etc.

Safe space - The sense of safety and security is integral to people's mental health and wellbeing. People are comfortable moving around in their neighborhoods if they feel safe and sense a minimal threat. Urban threats may include traffic, getting lost (way-finding difficulty), environmental pollutants, criminals, and 'undesirables'[5] (people perceived as a threat but are not harmful). Appropriate street lighting conditions, surveillance, distinct landmarks allow people to navigate in space and impart a sense of safety. A balanced approach between safety and activity opportunities is necessary for people to enjoy a sense of agency and de-stress. A risk-averse city space with no opportunities for physical activity reduces the sense of agency, which is not conducive to mental health and wellbeing.

Although the GAP's theory postulated by McCay is somewhat new, yet multiple authors support the core ideas. Some of the features of the GAPS theory are further explained below.

Emphasis on green spaces: McCay explains that the four themes/criterias can be inter-

[5] William Whyte uses the term undesirables to describe those who are harmless but perceived as threats by the residents. Undesirables are people who act strangely in public, such as hippies, street musicians, vendors, and poor/homeless.

related and applied to different types of built environment projects. White et al. conducted a survey of 10,000 people in UK, where they found communities living in greener urban areas were less likely to report mental distress and more likely to report higher levels of wellbeing [71]. An association between more trees per kilometer and a decrease in the antidepressant prescription was found [72]. Green spaces were also tested for their positive impact on stress reduction, in addition to improved social and cognitive function [73]. Reduction in depression, mood disorders associated with dementia, and reduction in violence are the other benefits of exposure to nature [74].

Researchers also found better opportunities for engagement in physical activity as one of the impacts of urban green spaces. The positive impact of exercise is associated with physical wellbeing, e.g., exercise is found to boost people's self-esteem, and mental wellbeing, thus moderately effective in reducing symptoms of stress and anxiety [75].

There are primarily three theories on how the association between green space and mental health may relate to one another. First, Edward Wilson's '*biophilia theory*' where Wilson argues that human beings have a close relationship with the natural world, and they subconsciously seek contact with other species, and they have a predetermined biological need that drives them [76]. Second, Ulrich's theory claims nature helps in stress reduction [77]. The theory asserts, recovery from stress happens through a range of physiological and psychological response to factors such as distance from everyday demands (such as facing workplace stress, meeting financial needs of the family, etc.) possibilities of aesthetic appreciation and activity is driven by individual interests [78, 79]. Third, Kaplan's theory proposes the idea of "attention restoration." It puts forward the idea that the natural environment relieves people's "attention fatigue" by distancing people from tasks that demand maintenance of prolonged attention. Instead, it facilitates them to use attention without the need for concentration [80]. As such, it is important to ensure that the natural spaces are encountered by people in the daily course of their life. If people need to make plans for traveling to the green spaces, they may risk their daily 'dose' of green places.

Irrespective of age, gender, and socioeconomics, the longer the exposure of people to outdoor green spaces less intense is their feeling of stress [81]. Sullivan's research shows that classroom views to green landscapes significantly improve performance on tests of attention and increase student's recovery from stressful experiences [82]. Similarly, studies show that having outdoor green spaces close to the workplace and home proves beneficial and acts as a de-stressor. The locations of these green spaces are specifically important. The distances from the people's home to the green spaces are found to be an independent predictor of stress [83].

Improving streetscapes: Streets are also crucial for people's stress reduction. McCay points out that streetscapes are vital opportunities for any city to improve the mental health of citizens. Being exposed to a well-designed streetscape daily can influence people's mental health positively. McCay also claims 'human scale facades' showcasing exuberant activities impart street character and break the monotony, thus encouraging people to engage in walking or lingering on the streets [24]. Similarly, Gehl recommends that there should be an interesting visual cue every five seconds of a street when an average walker is walking at 5 km. per hour [84]. Processing varied information while perceiving the external world prevents people from dwelling and ruminating on pessimistic thoughts, stopping them from getting bored. Boredom is primarily associated with stress and instigation of bad social behaviours such as addiction [85]. Teenagers and young adults who grew up in the suburbs routinely refer to them as boring. Matteau wrote in a New Jersey bulletin that "Young people find the suburbs very, very boring. They have suburban fatigue". They are seeking both visual and social stimulation. Interesting streetscapes not only break the boredom, they are also important in the cognition of space, way-finding, and providing a sense of safety to the pedestrian [86]. Tourists tend to seek out streetscapes that are anything, but boring.

Increasing land use diversity: Streetscapes are impacted by the nature of the surrounding land uses. While the mix of uses increases the diversity of activities and imparts a sense of community, single uses discourage activities and strip places of the sense of community. Locations with single, segregated land use promote driving between remote locations. Single-use discourages social trips, reduces social interactions. Driving between locations situated further apart requires roads and infrastructure for automobiles, downgrading the pedestrian or biking infrastructure like sidewalks and bike paths. Driving restrains people in their vehicles, reducing the opportunities for active and passive social interactions. Roadways prioritized for car traffic such as through roads and highways divide communities, reduce walkability between destinations, and create noise, pollution, and unsafe conditions. The long-distance commute is identified as a leading cause of urban stress and hurts overall mental health. Exposure to noise, traffic, and congestion for an extended period during commute may contribute to experiencing stress, anxiety, hostility, aggression, and other negative feelings, thus weakening mental health [87, 19].

Active Transportation: Commuting problems can also create fatigue at the workplace and aggravate work-related stress. Spending a long time in commute also means less time socializing, i.e., less social engagement with friends and family. Studies show a longer commute time has a substantial negative impact on social capital that is otherwise supportive of mental health [88, 89]. Long-distance commute by any mode (driving or public transportation) impacts sleep quality and maintenance of physical activity level essential for the upkeep of mental health. Interventions to improve mental health should include active transportation options. According to the Centers for Disease Control & Prevention (CDC), active transportation is defined as a self-propelled, human-powered mode of transportation, such as walking or bicycling.

While it is true that roadways often are a source of noise, pollution, and other safety issues that influence mental health negatively, we cannot design cities without them. Accessibility is one of the critical assets of any city. That being said, we can minimize driving by maximizing the use of active transportation options and public transportation for people. Promoting active transportation options such as walking and biking plays a vital role in linking communities, facilitating movement in and around the city, increasing opportunities for education, housing leisure activities, social interaction, and access to nature, all of which are important for mental wellbeing. Active transportation options reduce the overall cost burden and accessibility between places allowing for an increase in social interaction.

2.3.3 Key Findings from Urban Design and Urban Planning Literature

The urban design and urban planning literature on the 'built environment and mental health' point out some of the key aspects that overlap with the psychology literature, as well as add some additional perspective. Here, the focus is primarily on the public realm, the realm that imparts the experience of the city's culture to residents and visitors. This literature does not blame urbanness for mental health issues, rather highlights the importance of small blocks, diversity of land use, streetscapes, aesthetics, and, more importantly, the presence of people or pedestrians in the public realm. According to scholars, urban areas are visually stimulating. The urban density also supports the opportunity for social interaction and allows people to participate in various urban functions. The provision for active transportation is seen as a key contributor in maintaining social interaction and building social capital that is valuable for mental wellbeing. The notion of density is separated from the phenomenon of overcrowding, where people have very little control over their personal space. "Happy cities" are described as cities that allow residents to develop trust through easy communication among their neighbors [89].

2.4 Importance of Identifying a Causal Link

Both physical and social environmental stressors influence mental wellbeing. The impact of each stressor is different. For example, certain noise sources have a negligible impact on perceived stress levels. In contrast, others lead to heightened physiological arousal and a negative state of mind (experienced as irritability, annoyance, etc.). Besides, we have seen, the perception of crowding is lower when space is shared by kin members (close family or friends). The same size of space, if shared by non-kin members (strangers), tends to increase the perception of crowding and stress related to crowding. Halpern explains the phenomenon of differential perception, that is, an individual's insight about the causes. The insight is generally developed by his knowledge about the aetiology. Demonstrated in an experiment by Schachter and Singer [90], an individual's psychological arousal depends on their interpretation of the cause of the arousal. If individuals are uninformed of the cause of arousal, they show a higher degree of anger and other behavioural anomalies, which is very similar to the nature of stress experience. Different aspects of the environment may lead to a state of arousal, and the internal state of imbalance is then attributed to a cause. The attribution of the cause depends on an individual's world knowledge that comes from their culture, context, and upbringing. The interpretation or understanding of the cause relies on his worldview. Where the causal link was clear, the subjects could cope with ease (in the form of anxiety, stress, or nervousness) compared to the subjects whose causal link was unclear. For most stressors, other than noise, the cause of arousal remains unknown or wrongly attributed to some other cause. The wrong attribution of cause can lead to further anxiety or negative affective responses. Usually, the anxiety and negativity are heightened by a wrong coping process. Knowledge about the actual cause may allow individuals to use the right coping strategies. Similar conditions are observed in postpartum depression (lasting for about 2-4 weeks of depression). The cause of such depression is attributed to hormonal changes; however, there may be other factors that contribute to the depression. For example, the lack of support and difficultly in adapting to the role of motherhood is often the real cause of postpartum depression [91]. The incorrect attribution of the cause of stress and lack of knowledge deepens the problem of individuals than alleviating it. Halpern postulates that informing those who are affected and those who are vulnerable to the given stressors could offset the adverse effect of the stressors. The goal of this study is to draw attention to the built-environment stressors that heighten stress perception in the urbanites through an empirical approach. These built environment stressors may act independently or in conjunction with other stressors. Acknowledging the presence of these stressors and getting to know their impact in an urban scenario will not only help people manage stress but also reduce the level of stress perception.

2.5 Methods of Stress Measurement

2.5.1 Conventional Methods

Various quantitative methods for exploring the association between stress and the built environment are found in the psychology literature. Univariate, multivariate logistic regression and multivariate regression analysis are used to test the relationship between perceived stress and the built environment. In these regression models, demographic variables are used as controls for potential confounding factors, alongside key interest variables. A large number of studies did not use regression models. They have primarily used methods like partial correlation, Pearson correlation, analysis of variance (ANOVA), etc. The sample size of the largest dataset used in any of these studies is between 1100-2500, and they primarily focused on small areas (a few building blocks within a residential community). There are studies where interviews were limited to 100-500 people, and the interview results were used for both qualitative [92], and quantitative assessments [25]. Sturm and Cohen wrote a seminal piece of literature that addresses the association between sprawl and mental health disorders quantitatively on an urban scale [93]. To my knowledge, there are no other studies till date conducted on an entire urban area or city scale.[6] For the first time, they were able to integrate sprawl and mental health through a large-scale quantitative analysis.

In all of these studies, built environment variables are measured using people's perception/response. For example, a survey question may ask 'how safe do you feel in your

[6]Frumkin et al. [19], Putnam [20], and others wrote about the negative impact of sprawl on mental health. Most of their studies are backed by descriptive statistics and theoretical perspectives only.

neighborhood', and a scale (e.g., 1 to 7, higher is safer) is used to measure the safety of the neighborhood. No objective measurements for the built environment (which includes data, like the number of crimes or streetscape characteristics) are used in the study, i.e., the built environment measures used are not entirely independent of the mental health of an individual.

It is worth mentioning some studies measure mental health/wellbeing using a relevant scoring system. For example, Sort Study Form (SF-36) [7] mental health score [94], or Perceived Stress Scale (PSS) score [95] is used more universally. There are other scales used by NIH in the study of the mental wellbeing of individuals, such as the Social Readjustment Rating Scale (SRRS) [96, 42], Center of Epidemiologic Studies–Depression Scale (CES-D) [97].

The methods of data collection in the studies included a mailed survey questionnaire [98], and interviews (participants sampled using both random sampling and snowball sampling). People were asked to report their satisfaction level with elements of their environment. Peoples' satisfaction levels were measured by a score. They were asked to report their demographic information such as age, race, employment status, socio-economic status, details about their housing, and crime reports if any.

2.5.2 Issues with Conventional Methods

Halpern claims that the methodologies required in exploring the causal relationships between the built environment and mental health should be able to disentangle the compound association between social, psychological, and environmental factors. Social selection and response bias of the subjects [9] are two key problems to overcome when working with survey data (or interview data). Social selection is the process by which individuals are organized in a geography based on their social or individual characteristics. People with

[7] The SF-36 is a self-reported measure of quality of life related to health and wellbeing. The SF-36 has a 5-item mental health domain (MH) score. It is a 35-item component summary score that evaluates the general mental health status of individuals.

psychological impairment or poorer mental health may drift to the most unpleasant environment in the city. The phenomenon is called social drift / social selection and is caused due to a number of factors such as the likelihood of finding better mental health facilities in cities and the concentration of more affordable yet blighted neighborhoods in the city center [9]. Thus, the association between living in the city center, and having mental health issues can be a spurious causal claim. The concentration of individuals with a predisposition to mental ill-health in urban areas could be explained by the social and economic push-pull factors, where characteristics of urban living may not have a role to play.

Similarly, response bias is another central concern. Response bias occurs when the perception and reporting of the environmental variables are strongly influenced by the subject's mental state. If a strong association is found in some measure of an individual's mental health and their report of problems such as perceived noise or overcrowding, there can be two plausible explanations. One, that overcrowding or noise may have caused mental ill-health. Two, mental ill-health is causing the individual to perceive the environment as problematic.

2.5.3 New Approaches of Stress Measurements

There are three critical parts covered under new approaches of stress measurement: 1) ways to minimize biases in the conventional approach, 2) use of social media data in measuring stress, 3) ways to use the location-based data to augment the findings.

Minimizing problems or biases: There are various methods to minimize biases in the study. First, *identifying situations in which the link between two variables is broken, such as housing quality and income*. For example, subsidized housing or affordable housing is better than what an individual of low economic status could afford, weakening the link between income and household quality. If the comparison is restricted to a narrow income group or a very specific group, it is likely to reduce the range of housing quality or range of built environment variables they interact within their daily life. Second, *identifying* *variables that are causally relevant yet invisible to the subject, meaning the subjects are unaware of the influence*. A cross-sectional study can be used with more confidence. For example, if the influence of road networks on mental health is studied, then hypothesizing that the characteristics of the road influence children's safety in the neighborhood may not be obvious to a resident's mental health. Such type of hypothesis should be explored. The third approach is *adding relevant statistical controls*. Inappropriate controlling for covariates may conceal the effects of the environmental variables on mental health. The hidden variables may have a low association but can act as a meaningful statistical control. Fourth, the *longitudinal design should be considered where the subject does not change while the environment changes*. Halpern points out the strength of within-subjects design, eliminating variability introduced by the compounded covariates. The goal is to present a study that can isolate the effects of a very specific environmental change on mental wellbeing. For example, the impact of sudden onset of construction, noise can be studied by questioning the residents about their irritability, annoyance, and other psychosomatic symptoms, before and after the onset of the noise. The literature suggests different methodological adjustments and triangulation for minimizing bias before claiming causal relationships [9]. If they all lead to similar conclusions, then we may confidently establish a causal association.

Use of social media data: The past decade has witnessed a rise in the use of social media data in various disciplines (e.g., public health applications [99]). It is used for influenza surveillance, monitoring mass gathering, understanding public sentiments on health topics such as vaccination, and understanding depression. Scholars in urban planning and geography have used social media data to understand travel patterns [100], land use classification [101], and natural disaster detection [102, 103, 104]. Tweets[8] or Twitter microblogs, and Foursquare Check-in[9] are primarily used in the geography and planning domain compared

[8] Twitter is a popular social networking site, and Twitter microblogging has gained popularity in the last decade with the rapid growth of monthly active users of Twitter from 30 million in 2010 to about 330 million in 2017. The Tweets or status update messages (SUMs) are limited to 280 words, and they express valuable perspectives of users.

[9] Foursquare is also a popular social networking site that shares information about businesses and attrac-

to any other platform. These are preferred social media platforms firstly due to easy data availability through their public application programming interface (API), and secondly, the time and location information associated with Tweets and Foursquare check-in data allowed researchers to spatially contextualize the experience shared by individuals. The use of social media data or social computing facilitates social science disciplines such as planning to use information and communication technologies (ICT) to expand the domain knowledge [105]. Furthermore, using social media data, researchers can be naturalistic observers and get access to valuable information. In this section, I will specifically discuss the use of social media data in the context of predicting stress, affective responses, or mood, using location-based social media (LBSM) data in characterizing urban space, and using (LBSM) data in predicting user activity.

When we think about predicting stress or other mental health conditions using social media data, the argument we need to address is, are people really 'authentic' versions of themselves? An individual may present an "idealized" rather than an authentic version of herself [106]. However, the state of the art research shows social media platforms allow honest and candid expression of thoughts, experience, and belief, known as *self-disclosure* [99]. According to the literature self–disclosure has therapeutic value and is likely to enhance physical and mental wellbeing [107]. Self-disclosure results in disinhibition, which plays a positive role in psychological counseling. Research shows that the psychiatric interviews conducted by Computer-mediated communication (CMC) yielded honest and candid answers [108].

De Choudhury [27, 91, 109] pioneered a series of publications applying machine learning to investigate mental health issues from user posts shared on various social media platforms such as Twitter and Reddit.[10] For example, Reddit posts were used to measure the

tion venues. It is primarily used by people to share their location information with friends and acquaintances. People share their foursquare check-in information on Twitter.

[10]Reddit is a popular social media site. The platform is primarily used for content sharing, obtaining feedback, and information from diverse communities. These communities are known as subreddits. The subreddits communities are also geographically localized and dedicated to specific public or private organizations.

psychological impact of gun violence on college students [27]. Novel computational techniques are used to quantify and examine stress response after campus gun violence. Machine learning stress classifier has been developed to identify the stress level of the Reddit posts. Further, the temporal and linguistic changes in the posts are studied to characterize conversations on campus. Tweets were also used to quantify postpartum changes in mothers along the dimension of social engagement, emotion, social network, and linguistic style. The goal of this study was to use social media to identify mothers at risk of postpartum depression [91]. Other studies of De Choudhury et al. involve social media-based mental health index of college campuses [110], and identifying minority stress experiences on social media [109]. Such systems were developed with the goal of early detection of depression and promoting wellness. The studies may also help in rehabilitation efforts around the crisis events like gun violence [91] or in designing tools sensitive to the needs of minorities.

Use of location based social media data: Location-based social media (LBSM) platforms, such as Twitter, Foursquare, Instagram, etc. have allowed people to share their experiences and activity locations to their online social network [111]. Market analysis shows there are 3.5 billion smartphone users worldwide, meaning almost 45.5% of the world population has smartphones [112]. The rising numbers of smartphone users show that there will be greater potential to harness data from smartphone applications in the coming years. Although very few users (0.85% of Twitter users and in optimistic scenarios 1-2%) [113] share their location information on Twitter for security issues, there are still plenty of Tweets with geolocation. Augmenting the dataset with various location prediction algorithms has allowed researchers to capture larger than the anticipated size of datasets with location information [114, 115]. The use of LBSM data to measure wellbeing in an urban context is relatively new. To my knowledge, geolocated tweets are primarily used for sentiment analysis. Plunz et al. analyzed Twitter sentiments inside the parks in New York, presenting a comparison between aggregated sentiments in and outside New York City parks [116]. Tweets are also used to extract the spatial distribution of emotions such as anger, fear, happiness, and sadness using graph-based machine learning in different locations within the greater Boston area [117].

There are several other applications of LBSM data. The first use of LBSM data is seen in the "Livehoods Project". [118] The study used over 18 million Foursquare check-ins collected from users. The check-in data is used to build clusters called *Livehoods*. The spectral clustering method is used to account for spatial proximity (based on the distance between venues) and social proximity (number of times users checked into these venues). Frias-Martinez et al. used geolocated Tweets to determine land uses in specific urban areas based on people's Tweeting patterns and also in identifying points of interest in Manhattan, New York [119]. Hamstead et al. used geolocated Flickr and Twitter data to explore variation in park visitation across New York City. The study also models visitation based on criteria such as spatially-explicit park characteristics and facilities, neighborhood-level accessibility features, and neighborhood-level demographics. Noulas et al. used the semantic information associated with the foursquare data in the identification of user communities and categories in Manhattan, New York City. They are used to build user recommender systems [120, 121].

LBSM data is extensively used in land use and transportation planning. Land use classification framework at the level of traffic analysis zones is proposed for Guangzhou, China using remote sensing data and multi-source social media data [101]. LBSM data has been used in predicting the travel pattern of users. Hasan et al. use Twitter and Foursquare data to categorize activity patterns (into activity classes) of users based on the foursquare venue classification [122]. These activity classes are further refined using more features such as time of the tweets and tweet topics [111]. Through these explorations, Hasan et al. point out the limitations of the home and work location prediction in the models that are deployed to predict activity. They suggest the use of other relevant features such as user demographics, time of the tweets, and tweet topics to further augment the data set. The augmented data set, if used to train the models, performs significantly better in the home and work location prediction. Abbasi et al. used the Foursquare, and Twitter data as a substitute or an additional data source for travel surveys [123]. The comparison between the activity data from the traditional travel survey and the one from Twitter data showed that the model using Twitter data made more accurate predictions [124]. However, we need to acknowledge the repressiveness of the social media users and existing challenges such as the availability of user demographic information.

2.6 Critique of Literature Identifying Gaps

In this section, I have highlighted the gaps in the literature that this research seeks to address.

1. *First*, the literature on mental health and built environment primarily relies on survey data and mental health data available for counties or census tracts. The survey data are collected from residents in one or more residential communities in a city. The size of the study area, geography, and the population sample studied are too small. Also, the mental health data used in the urban-rural comparison are aggregated values for counties/ census tracts [93], making it difficult to understand the impact of built environment variables on the mental wellbeing of individuals.

2. *Second*, mental health issues primarily go on record when individuals are seeking help from mental health practitioners. Some of them (who qualify for the survey) have a greater proclivity to the disease due to life-changing events or other problems that may or may not be caused by built environment variables. Currently, the stress studies focus on those with severe mental health issues only, and there's a gap in looking at the influence of the built environment on otherwise relatively healthy individuals (i.e., the group is presumably included among those who are Tweeting).

3. *Third*, there is inadequate attention in the literature to boredom and boring urban

spaces. There's growing interest in this topic in the field of neuroscience and brain activity, but it still appears to be couched with lots of caveats. Psychologists like Esther Sternberg have focused on relaxing the brain and supported suburban living conditions [125], while the neuroscience-aestheticists have focused more on the benefits of stimulating the brain with compelling and complex design[86].

4. *Fourth*, people with existing conditions may lead to only negativity bias in their response. While it is extremely important to get a perspective of those who are suffering from mental health disorders on this topic – it may also help researchers to obtain an alternative perspective.

5. *Fifth*, some people may experience symptoms of anxiety and stress in their regular life events. They may experience acute stress due to daily hassles at work, traffic-related stress while commuting, etc. The chronic stress caused by environmental stressors such as ongoing noise, pollution, crowding, or even fear from crime or social isolation are also experiences that may have a serious impact on mental health if left unmanaged for a long period. Currently, no method for early detection of stress or depression is used in literature to decode the influence of the built environment.

6. *Sixth*, very few studies compare mental health issues on an urban scale and between two or more cities except the study by Sturm and Cohen [93]. This study aims to compare two cities, Atlanta and Boston.

7. *Seventh*, as discussed in the literature *urbanness* or urban characteristics of an area influence mental health problems in individuals. The literature reviewed has used aggregated data on mental health disorders for rural and urban areas, without much consideration of built environment variables. Some scholars point out that the higher number of mental health issues in the urban areas compared to their rural counterparts is because of homelessness, drug addiction, and poverty. Others claimed lack of work-life balance, a greater number of pollution sources, unaffordability, and density

are the leading cause of the mental health problem in cities, but there is an alternative perspective as well. Urbanists and social scientists have identified increased levels of stress due to social isolation, lack of opportunity for social activities [20], and longer commutes [19] in the less urban areas (such as suburban or rural areas). The disparity in literature further leaves us room for investigation.

8. *Eighth*, most studies in the past have focused on a single key environmental or social variable such as noise, weather, or a few socio-demographic variables as controls to test the impact of the built environment on mental wellbeing. Those studies are useful and form the basis of my future investigation, but Halpern and other scholars have pointed out that mental health has a complex relationship with the built environment, and more variables and methods of triangulation are necessary to disentangle any causal link. Using appropriate objective measures of built environment variables is likely to give a better understanding of these complex relationships.

9. *Ninth*, the built environment and mental health literature did point out the quality of design, quality of housing is key to good mental health. Presumably, people having access to good physical space are in the high-income group. Additionally, overcrowding and having a lack of access to adequate space can increase aggression and stress levels in an individual. Skeptics in the field of psychology also pointed how social-drift may cause cities to be the primary inhabiting place for those who are socio-economically disadvantaged. However, fiscal insecurity or poverty has not been adequately captured as a control variable in stress studies.

10. *Tenth*, the contemporary literature on using location-based social media data (LBSM) has focused on computational accuracy of location prediction of individuals, ranking places by choice, mapping land use, disaster prediction in urban areas. In addition, there is a rich body of literature that uses social media data to detect stress, depression, and other mental health ailments. To my knowledge, there is a dearth

of research that connects mental wellbeing indicators to characteristics of the built environment.

CHAPTER 3
RESEARCH GOALS AND RESEARCH QUESTIONS

This research aims to examine the relationship between mental wellbeing and the built environment in a city. Primarily focusing on the public realm, I use stress and/or de-stress as a mental wellbeing indicator. One of my key contributions is to widen the scope of empirical research from small urban areas and residential communities to the scale of a city. Here, I consider the administrative boundaries of large metropolitan areas of Atlanta and Boston. I contribute both methodologically and theoretically. In the methodological part, I specifically focus on predicting stress from Tweets and creating objective measures for built environment stressors, thus linking people's stress levels and the built environment stressors. Theoretically, I examine the claims of the *built environment and mental wellbeing* literature and subsequently address the disparities and gaps in the literature, as highlighted in Chapter 2.

Prior to explaining the research design, I would like to address the data collection issues for this research at an urban scale. Urban designers have worked on bolstering methodological aspects of quantifying perception in cities based on survey data. Photographs, videos, interviews, and survey questions were collected to quantify the perception of urban areas [14, 84]. While these are still a few of the most reliable ways of data collection, these methods are expensive, thus limiting the generalizability of the research findings. Studies show that mental wellbeing is governed by symbols and labels, which means people usually attach a symbolic label to a place as good or bad. As such, parameters of wellbeing vary across geography and culture [25]. Methods expanding the geographic scope of data collection and enabling the use of objective measures of built environment stressors are crucial to make these research claims generalizable. This research aims to bring a methodological perspective such that social media data or big data can be used to gauge mental wellbeing on an urban scale. Given the diverse populations and geographies, it might be a burning question 'why does this research focus on urban areas? I limit myself to urban areas to constrain the scope of this research. This research is exploratory and involves big data usage (both social media data and spatial data). The computing time to assess the fine-grain spatial data can limit the exploratory capabilities. In the future, I see the possibility of operationalizing this research beyond urban limits and using survey data to augment social media data. The use of an appropriate survey design to match the features obtained from social media data can reduce the demographic bias of social media and improve the population representativeness of the research.

My research is designed to identify people's stress and how they cope with the built environment 'stressors' in their daily life. Edward Bender coined the term pulses and presses in the 1980s [126]. He defined pulse disturbances as perturbations or short-term disturbances and described press disturbances as an effect that is persistent and long-term. Here, I conceive the built environment stressors as presses that may disrupt people's daily life and yet go unnoticed. For example, a busy street in front of a residential neighborhood may act as a stressor if it does not allow meaningful social engagement. A workplace with limited or no eateries within walking distance may add daily stress to employees. The knowledge of the impact of built environment characteristics on mental health and wellbeing can enable urban planners, urban designers, and policymakers to design/reinvent places. It may also find its use in prioritizing the place-making needs of a city, depending on the mental health vulnerability of certain areas. Also, from the literature, I observed that people's stress level and their ability to process stress improve if they recognize the cause of the stress [90]. This research aims to reveal relationships between stress and built environment stressors to people, such that it improves their ability to process stress. Refer Figure 3.1 which describes the concept map of my research.

By enabling the use of dynamic data sources of social media - that is the data that changes with time and space - I connect city planners with civilians such that city planners

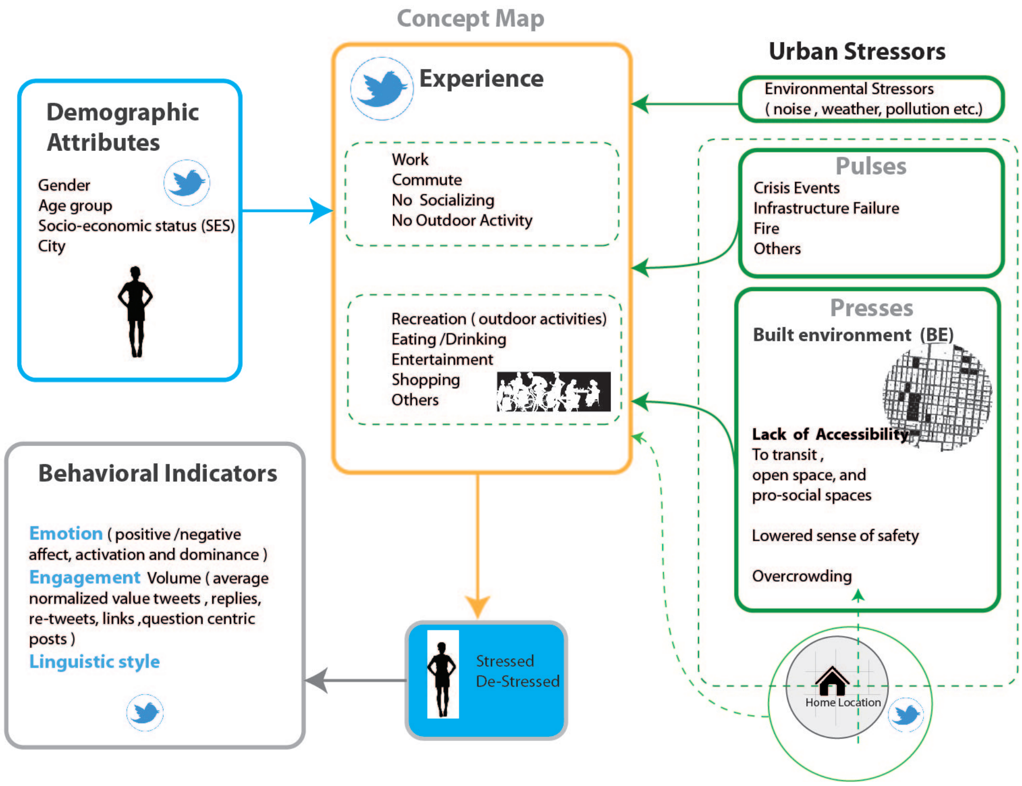

Figure 3.1: Concept map of the research showing human experience is dependent on exposure to urban stressors, as well as demographic attributes. The experience (in form of stress or de-stress) gets shared on social media platform which is then assessed to measure mental wellbeing.

could understand a civilian's perspective. Such understanding can empower planners to make effective policies supporting design strategies that would help minimize people's stress levels through appropriate built environment design [127].

My research addresses an overarching key research question and a few hypotheses to fulfill the research goals.

3.1 Research Question and Hypotheses

As mentioned earlier in this chapter, through my research, I intend to examine the claims of *built environment and mental wellbeing* literature and address disparities between psychology literature and urban theories. My key research question is:

RQ. Is there a relationship between people's stress expression and their proximate

built environment characteristics in urban areas?

The literature review shows that there are certain overlaps between urban theories and findings in the psychology literature. For example, the presence of *escape facilities* such as community engagement spaces and green spaces are beneficial for de-stressing. Besides, poor built environment quality and signs of blight may heighten the perception of crime and act as a built environment stressor. While I investigate those established claims in this research, I also address a few disparate claims in psychology and urban theory. These constitute my hypotheses.

H1. People are less stressed in areas with a higher degree of urbanness.[1]

This hypothesis (H1) contends that the urban built environment (including building density and high-rises) has an association with higher stress levels [59, 25, 36]. Urban theorists since the 1960s have highlighted the importance of urban qualities, including small blocks, block density, and accessibility in imparting a positive experience. Such spatial experiences are usually linked with a sense of satisfaction, improved memory, and cognitive development. In this research, I investigate whether people tend to discover spaces to de-stress in areas with a higher degree of urbanness and whether they are less stressed [2] compared to areas that lack density, lack diversity in block density, have larger block sizes, and more cul-de-sacs.

H2. People are less stressed in areas with a greater diversity of escape facilities. [3]

This hypothesis (H2) counters the claim of psychiatric geography that higher levels of anxiety and stress are associated with mixed land use [9]. The positive impact of mixed land uses is contested in psychology literature. Other than green spaces, the literature undermines how the various amenities in the city, such as restaurants, theaters, museums, and

[1]In this research, I define urbanness with qualities such as smaller block sizes, high diversity in block density (building density), the proportion of streets with no setback, the proportion of gridded streets, and access to active transportation.

[2]I measure the stress level of people in specific urban areas using a scoring system called 'mental well-being score'(MWS). MWS is explained further in the section 4.5

[3]Chu et al. describe escape facilities as destinations that allow social interactions. Green spaces, social and community facilities, recreational facilities, cultural and religious facilities [36].

urban plazas, are essential for de-stressing. In this research, I investigate how people effectively utilize mixed-land-use for de-stressing and examine whether they are less stressed in areas with mixed land use compared to areas with single land use.

H3. People are less stressed in active[4] high -density areas with high symbolic value[5]. The use of density as a proxy to measure crowding is somewhat misleading. Crowding is loosely defined as a measure of population density in the mental health literature. Although negatively connoted, crowding is considered neither good nor bad. Crowding may intensify certain existing social conditions, but rarely are urban problems caused by crowding itself [129]. In the mental health literature, crowding is not differentiated from *overcrowding*, which results from insufficient resources in a city. Resource limitation may result in a lack of control over one's personal space. With the inability to secure needed privacy due to factors including poverty, or rent burden, people experience extreme stress, anxiety, and other types of mental health issues. While the claim of scholars in psychology is that crowding negatively impacts mental health, it is only true for the conditions of overcrowding. The hypothesis (H3) assumes that crowding itself is not a stressor unless associated with social conditions of poverty, rent burden, and a consequent shortage of space. With this hypothesis, I examined whether places with a higher population density and footfall, that are accessible by active transportation, and have a higher symbolic value (due to economic affluence of residents, presence of landmarks, and others) have fewer individuals who are stressed.

The research design and methods that answer the aforementioned research questions and research hypothesis are further discussed in Chapter 4.

[4] The active spaces are places closer to transit and have bike- pedestrian infrastructure [128].

[5] Symbolic value is an immaterial value attributed to areas based on their historic significance, economic status, and perceived safety. Places with landmarks, rich neighborhoods, and places that invoke less fear of crime, have higher symbolic value or good label attached to them [25].

CHAPTER 4
ANALYTICAL FRAMEWORK TO ASSESS RELATIONSHIP BETWEEN STRESS AND BUILT ENVIRONMENT

This research focuses on identifying *stress levels of* tweets and characterizing the built environment where people are less stressed. I formulated an analytical framework that classifies the geolocated tweets as *stress* and *de-stress*. The framework aids in computing the fine grain built environment characteristics of a fixed area around a given tweet. In this fixed area, I compared how changes in the built environment factors (in the short and long distances) may impact the stress levels associated with a given location of interest. In that spirit, this chapter explains: (1) the details of the analytical framework of the research and (2) how the stress level for tweets (outcome variable) are assessed. In doing so, I explain the computational methods that quantifies the proximate built environment factors also called the explanatory variables. The following overarching research question and hypotheses are tested using this analytical framework:

RQ) Is there a relationship between people's mental wellbeing expressed on social media and their proximate built environment characteristics in urban areas?

I investigated the following key hypotheses to understand crucial aspects of places where people stress or de-stress in a city:

H1. People are less stressed in areas with a higher degree of urbanness.

H2. People are less stressed in areas with a greater diversity of escape facilities.

H3. People are less stressed in active high-density areas with a higher symbolic value.

As explained in Section 3.1, the hypotheses further supports me in clarifying the theoretical discrepancies in the urban theories and psychology/mental health literature.

4.1 Study Area: Rationale for Choice of Two Cities

To assess people's 'stress level' and understand how it varies with the built environment, I have identified two cities, Atlanta and Boston. The cities are located in the state of Georgia (GA) and Massachusetts (MA) respectively, where the proportion of the population living with severe mental health conditions are comparable. 3.71% of adults in MA and 3.6% of adults in GA are diagnosed with schizophrenia, bipolar disorder, and other major depression related ailments [130, 131]. However, the two states differ largely in the way access to mental health care services are available. Massachusetts tops the list of states in the year 2018 and 2020, in providing the highest rate of access to mental health care, while Georgia scores much lower, ranking 26^{th} [132].

Boston is a pre-colonial development. The city flourished as an international trade center. Boston experienced landscape expansion, development of park system, and construction of robust transit system in the late 1800's. In the urban design and planning literature Boston is identified as a city with adequate density. [1] and a city well-known for its historic urban core. Boston's bustling downtown area has smaller block sizes and historic neighbourhoods that are characterized by residential buildings with strong architectural characteristics. Other than these residential neighbourhoods, the city has abundant mixed-use blocks with greater land use diversity [134].

In contrast, Atlanta was founded as a railroad 'Terminus' and a Transportation hub. Post civil war, the population in Atlanta grew rapidly, retaining its role as a railway hub. The regulatory framework for land use in Atlanta has contributed to its sprawl. Atlanta is known as the poster child of sprawl as the city is characterized by its sprawled development

[1] Lloyd Alter (sustainable deign practitioner) defined the perfect density of urban areas as "dense enough to support vibrant main streets with retail and services for local needs, but not too high that people can't take the stairs in a pinch." However, planners argue there is no such concept as perfect density. As long as cities can operate maintaining multimodal transit-accessibility, and does not compromise their street enclosure, they can choose their optimal density level such that residents have adequate soft and hard infrastructure support. This means a large city has to build more to achieve adequate density, and a smaller city can achieve the same result building less [133]. Boston's transit system covers the city area, and in most parts, the city blocks are built to give the city an optimal density.

with large single-use blocks. The land use of Atlanta is less diverse than Boston. Boston is identified as a city that allows a closer fit between household transportation land-use preferences and offers a better choice of neighborhood for people in comparison to Atlanta [12]. Lynch's site selection criteria for understanding environmental images and cognitive map of people was Boston's uniqueness, vivid form, and complexity [135, p. 14]. To contrast Boston, Lynch selected Jersey City (New Jersey), which he described as formless, and Los Angeles, a completely different scale with a gridiron plan in the center and sprawled periphery. Although in the last 60 years much has changed in these cities.

I constrained my study area within the city boundary to ensure comparable built environment data for both cities. This also helped to optimize the computation time for the research. The differences in geographic scale of the two cities are adequate to test the impact of built environment changes on mental health. To compare further, the land area of the city of Atlanta is 133.2 square miles, almost 3 times larger than the city of Boston, with land area of 48.23 square miles. The average land parcel/block in Atlanta is almost 2.8 times the size of Boston's land parcels. The population density of Atlanta is 3740 people per square mile, and that of Boston is 14,400 per square mile, making Boston almost four times as dense. Atlanta has a gridiron center (downtown) and formless peripheral areas like Jersey City and Los Angeles.

Constraining my study boundary to the city boundaries does not constrain comparing urban measures. This is because the urban design literature does not characterize urban and suburban areas by administrative boundaries. The literature tends to focus on the formal characteristics, which I will call "urbanness". These include the physical characteristics of blocks, density (both population and building density), land use diversity, uniformity in density, block sizes, the proportion of dendritic road patterns, the proportion of streets with dead-ends to differentiate between urban and suburban areas (refer to the discussions in subsection 4.2.5) [136]. Inclusion of these two cities was prioritized overtaking a larger geographic area in a single city because literature review revealed that the impact of social and built environmental stressors depends on cultural values (distinct) of two cities/places [137]. Including fixed effects of the two cities has not only enabled me to capture the variability of the built environment but also capture certain immeasurable qualities of the two cities, such as culture, adaptation to weather, and other behavioral aspects of the residents. While I explored the built environment variables within the city boundary in this research, in future, I plan to expand it beyond the city limits to generalize the findings and improve the validity of my research[2].

Figure 4.1: To model mental wellbeing as a function of built environment stressors, I need to quantify mental wellbeing and the stressors. Hence, two analytical frameworks were needed - framework to compute mental wellbeing, and framework to compute stressors.

In the following two sections, I have outlined the analytical framework to conduct the research as explained in Figure 4.1. In Section 4.2, I described the datasets and analytical framework for analyzing the built environment and other stressors. In Section 4.3, I delineated the social media dataset (i.e, the set of tweets with their metadata), in addition to the analytical framework to classify tweets, and compute mental wellbeing.

4.2 Dataset and Analytical Framework: Built Environment Data

My research primarily emphasizes the exterior built environment attributes and their impact on mental health. In the past, researchers have addressed both interior and exterior built environment conditions (as discussed in Chapter 2). However, their study area was limited

[2]The construct validity refers to the degree with which we can legitimately operationalise my study on the theoretical constructs [138].

Table 4.1: List of all variables used for data exploration.

	Variables	Data Source
Outcome Variable	Mental Wellbeing Score (MWS)	Twitter data
Explainatory Variables		
Access to Facilities		
	Land use mix	Tax assessor's (TA) data
	Point of interest diversity	Google places
	Parks , green spaces , trails	TA data and Google places
	Social and community facilities	Google places
	Recreational facilities	Google places
	Cultural and religious facilities	Google places
	Health care facilities	SafeGraph and Google places
	Public transportation stop	City's publicly available database
Crime and Fear of crime		
	Vacant property	TA data
	Dead ends	Open Street Map (OSM)
	Street light	City's publicly available database
	Crime reports	Police reports
	Cross-sectional proportion	Building Height (LiDAR) and OSM
	Street wall continuity	Building footprint and OSM
	Building per 100m Segment	Building footprint
	Tree canopy data	City's publicly available database
Demographic and SES		
	Population density	ACS (2014-2018)
	Employment density	LEHD (2018)
	Percent white population	ACS (2014-2018)
	Poverty rate	Eviction Lab
	Eviction rate	Eviction Lab
	Rent burden	Eviction Lab
Urbaneness		
	Landmarks	City's publicly available database
	Intresection density	OSM
	Node ratio	OSM
	Setback	Building footprint and OSM
	Floor area ratio (range)	Building Footprint and TA data
Environmental Stressors		
	Daily temperature	Daily summaries (NCEI)
	Daily pressure	Daily summaries (NCEI)
	Daily precipitation	Daily summaries (NCEI)
	Traffic noise	FHA-HPMS Data
Participation level		
	Time of the day	Twitter data
	Weekday /weekend	Twitter data

to small housing communities. At the scale of the city, it is a computationally intensive task to analyze the interior conditions of the built environment. To account for insufficient interior variables, I assumed that the interior built environment qualities and conditions of overcrowding can be measured using socio-economic variables.

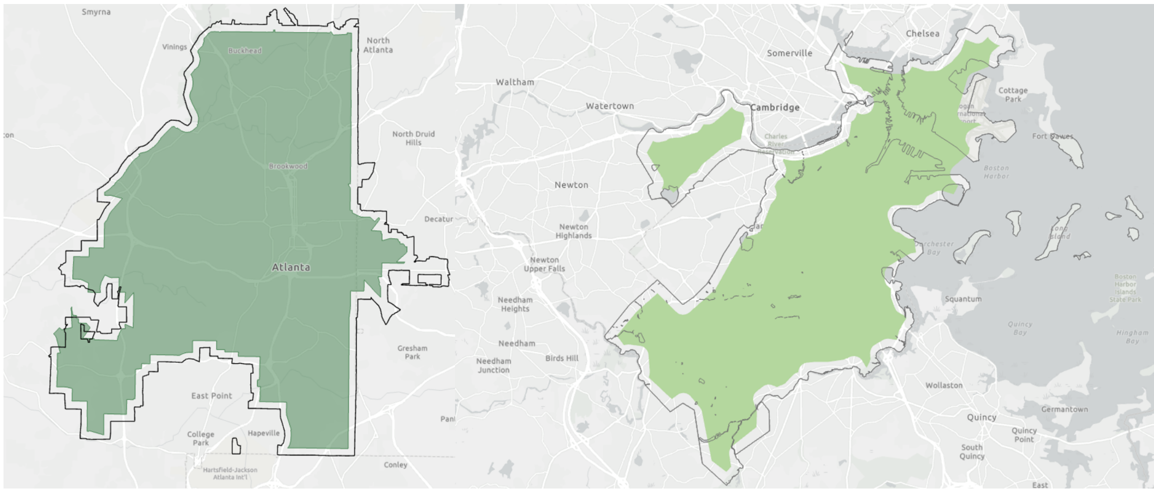

Figure 4.2: The figure shows areas of analysis in Atlanta (left) and Boston (right). The solid green area is drawn offsetting 1/4 miles. This was done to ensure all assessment grids used for the study have the built environment data available. Offsetting more than 1/4 miles significantly reduces the study area within the city boundaries of Atlanta and Boston

As discussed in Chapter 2, the qualities of the external built environment (BE) that primarily impacts mental wellbeing are: (1) access to escape facilities including open green parks, social and community facilities, cultural and religious facilities, transportation options, (2) opportunity for social engagement or participation, (3) and fear of crime [36]. I have considered built environment features within 1/4 miles, and 1 mile from a given tweet to test if the variation in the BE conditions have any impact on stress levels of the tweets. To measure these distances precisely, I have used the buffer function of ArcGisPro/Arcpy. Here the term buffer means a distance measuring function.

The 1/4 miles distance radius (buffer) is selected as it is a comfortable walking distance previously used in studies on transit-oriented development (TOD's) [139]. In addition, 1 mile distance radius (buffer) is used to assess the viability of non-motorized (active) modes of transportation, including walking and biking distance for studying travel behavior and mode choice [140, 141]. While some studies have used a 5 miles radius as the maximum analysis radius, I have used a 1 mile radius in the interest of optimizing time and computing resources required for this research. The studies that used 5 miles radius specifically compared mode choices between biking and walking. These studies used BE measures that were less granular, as such data availability was not a limitation. In this research, I estimated fine-grain BE measures. Furthermore, using any analysis distance above 1 mile radius would have excluded a substantial portion from the peripheral areas of the city due to unavailability of fine grain BE data outside the city boundary. The figure Figure 4.2 shows that larger distance buffers will significantly reduce the analysis space within the city boundary.

4.2.1 Assessment Grid (AG):

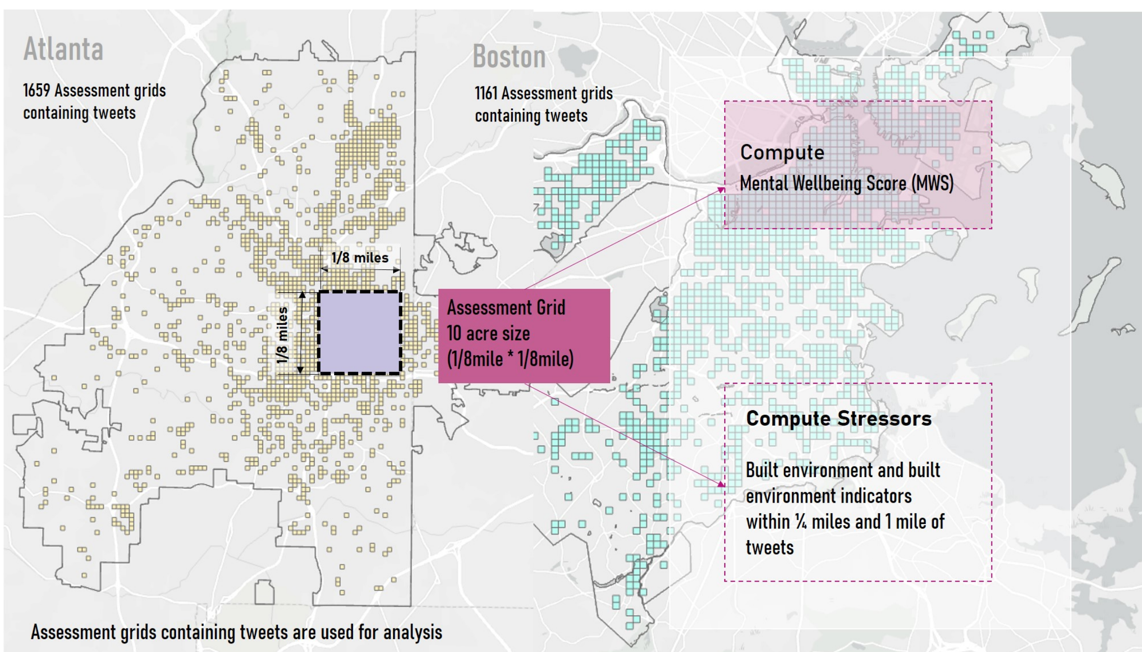

Figure 4.3: Assessment grids are roughly 10 ares in size (1/8 mile * 1/8 mile). Mental well being scores (MWS) are measured for these assessment grids.

Geospatial analysis is time-consuming, compute-intensive, and often does not scale efficiently with large data sets. In this research, the GIS (Geographic Information System) analysis involves handling large vector datasets to compute precise measurement of built environment variables including average building height, building set back distance, and others [142, 143]. Refer Table 4.1 to know all the variable used in this research. Arcpy, a Python-based GIS library, was used to compute the BE data using ArcGIS Pro's Python API. Estimating BE and demographic measures for various buffer distances (1/4 miles, 1 mile) for each tweet entailed running computational processes (for more than 100K tweets) twice (for each city). To give an estimate, computing variables that required vector data analysis, including drawing buffer, processing clip analysis based on specified distance radius, and calculating intersections, for over 100K data points, could take a few weeks or months. Given the scale, complexity, and computational load of the geospatial data, I developed a grid-based approach that optimizes for total compute time. I opted to use a square grid of sides 1/8th of a mile (660 feet). Here, I assumed that the BE change is minimal for 1/8th of a mile. Figure 4.3 shows assessment grids in Atlanta, and Figure 4.4 shows how the built environment variables are computed within 1/4 miles and 1 mile of those assessment grids.

Noteworthy to mention that pre-defined location tags available on Twitter show a varied level of accuracy, further explained in Section 4.3. For instance, location information from each tweet's metadata may not be exact. Such an error/noise was observed when I modeled built environment conditions and stress score/category for individual tweets. To mitigate this issue, measuring a normalized stress value for smaller grids improved the validity of the score compared to the same calculation performed for each tweet. Besides, the geolocated tweets might have a high bias in positive response due to the unbalanced distribution of the stress and de-stress tweets. For example, usually, people are more comfortable sharing blissful life events or sharing joyous stories on social media as opposed to sharing their sorrow. I observed the same pattern when I spot-checked the collected tweets. For instance, stress tweets were less than 10% of the total tweets identified as relevant for this study. A grid-based assessment of stress allowed to weight the proportion of stress tweets, such that they contribute significantly to determining the stress level of the location. Here, the stress level of tweets is a prediction based on the affective quality and situational response. Meaning, an individual's stress response could be either due to short-term stress caused by his/her immediate situation and the surroundings' built environment quality, or it could be due to long-term stress, including an unfortunate life event or an ongoing crisis.

Next, I have discussed the set of BE variables used to conduct this study. While an extensive description is provided on the rationale behind my choice in Chapter 2, in the following sections, I have provided brief explanations to substantiate my choice. I have also described the methodology adopted to compute various geospatial variables, such as tree cover area, building setbacks, grid-ness in streets, and others.

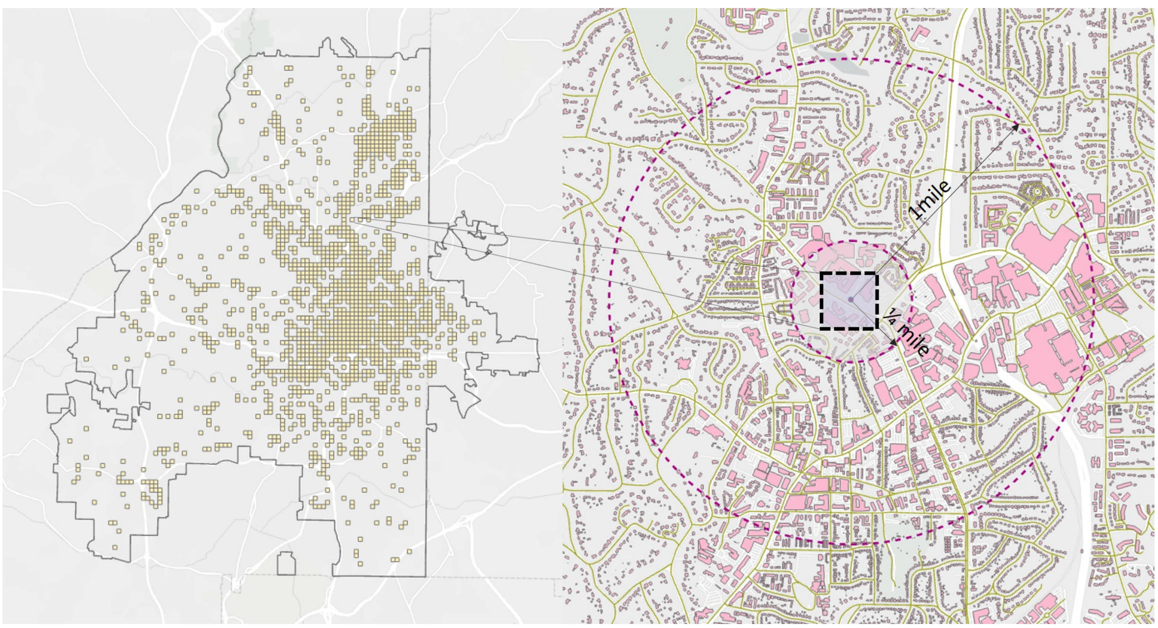

Figure 4.4: Built-environment indicators are computed within 1/4 miles and 1 mile of the assessment grids.

4.2.2 Measuring Access to Facilities

Land Use Mix: Land use mix measures the evenness in the distribution of land use [144, 145]. It also captures the spatial pattern and integration of complementary functions within an area [146, 147]. As juxtaposition of land uses is controversial in the context of mental health, I used land-use diversity as a variable to investigate its impact on peoples' stress levels. Land use typologies were obtained from the tax assessor's land parcel data for both Atlanta[3] and Boston.[4] The tax assessor's data contained complete information of the land parcel such as land use category, property price, number of stories, and age of the structure on the land parcel. There were more than fifteen land use categories in the dataset for both cities, out of which a few categories were dissimilar. This was justified, as the dataset was prepared and maintained by different administrative bodies with different zoning regulations. To ensure comparability, I simplified the land use categories to six - residential use, commercial use, mixed-use, government and institutional use, industrial use, park, and open spaces. A straight-line or (radial) buffer of size 1/4 miles and 1 mile was used to select the parcels around the assessment grids (AG). The land-use mix was calculated using an entropy measure as defined by Frank et al. [148].

$$Land\,use\,mix\,(Land\,use\,diversity) = -\,(A/\ln N) \tag{4.1}$$

Where $A = \sum_{i=1}^{n} b_i/a \ln(b_i/a)$

a = total square feet of land for all land uses present.

$b_1 b_2 ... b_i$ are - residential, commercial, Government and institutional, mixed-use, industrial , park and open spaces.

Points of Interest: Land use-mix alone was not adequate to capture the nuances of all the different categories of facilities that were available around a person's Tweet location (within, 1/4 miles, and 1 mile buffer).[5] A single land use such as *commercial* may house

[3]Link to land parcel data Atlanta:https://dpcd-coaplangis.opendata.arcgis.com

[4]Link to land parcel data Boston: https://bostonopendata-boston.opendata.arcgis.com

[5]In some cases, the tax assessor's data may assign a land parcel as tax-exempt land without assigning the specific use of the land parcel. A tax-exempt land could be a church, public school, or even land acquired by the city's infrastructure facilities such as transportation authority to house a shed. In this research, it is essential to differentiate between a church and a transportation shed. While the presence of a church is a facility for social gatherings, a transportation shed may act as a built environment stressor. A transportation shed does not promote constructive social interaction that can alleviate stress; instead, a transportation shed

different uses such as restaurants, churches, or grocery stores that are hard to capture by land-use data alone. POI data represented a finer grain snapshot of land use at the building level [149, 150]. In this research I have augmented the 'landuse diversity' variable by the use of POI diversity and the number of different categories of points of interest accessible within 1/4 miles and 1 mile buffer radius. The POI data used in this research was collected from the SafeGraph[6] and the Google Places[7] API. The reason for using two datasets was to create an exhaustive list of POIs. The individual data set from Safegraph or Google Places had some key places missing, while the two combined data sets gave us a comprehensive list of POI data. The POI data sets from SafeGraph has over 160 different categories. I used 75 other keywords to scrape the Google Places data set.

To scrape the data from Google Places, in addition to the keywords, I used a list of latitude and longitudes that was created using the centroids of the assessment grid.[8] Finally, I created a consolidated dataset of POI by keeping only the unique POI (i.e., dropped the duplicate POI's). The duplicates were detected using a combination of two fields: the name of the POI and the location address. The POI's from 200 plus categories were classified into facilities such as green escape facilities, social and community facilities, recreational facilities, cultural and religious facilities (refer Figure 4.5). The categories were derived from Chu et al. [151]. Refer Chapter 2. Shannon's diversity metric was used to assess the evenness of distribution of escape facilities and the opportunity of an individual engaging in social activity.

Shannon's diversity [152, 153] or entropy measure for assessment Gridpoint $_j$ is given

---

is an inactive land use (does not generate any human activity). Inactive and unmaintained land is identified as a stressor and is responsible for instilling fear of crime to individuals. The Point of interests (POI) data assisted in understanding the context where land use information was inadequate.

[6]Link to Safrgraph: https://www.safegraph.com/

[7]Link to Google Places API: https://developers.google.com/maps/documentation/places/web-service/overview

[8]Google Places API only provides 60 search result for a single keyword-based search. For a single keyword like 'restaurant' or 'bar', there were more than 60 search results expected; thus motivated by this theory, I used the grid centroids to maximize the search potential per query.

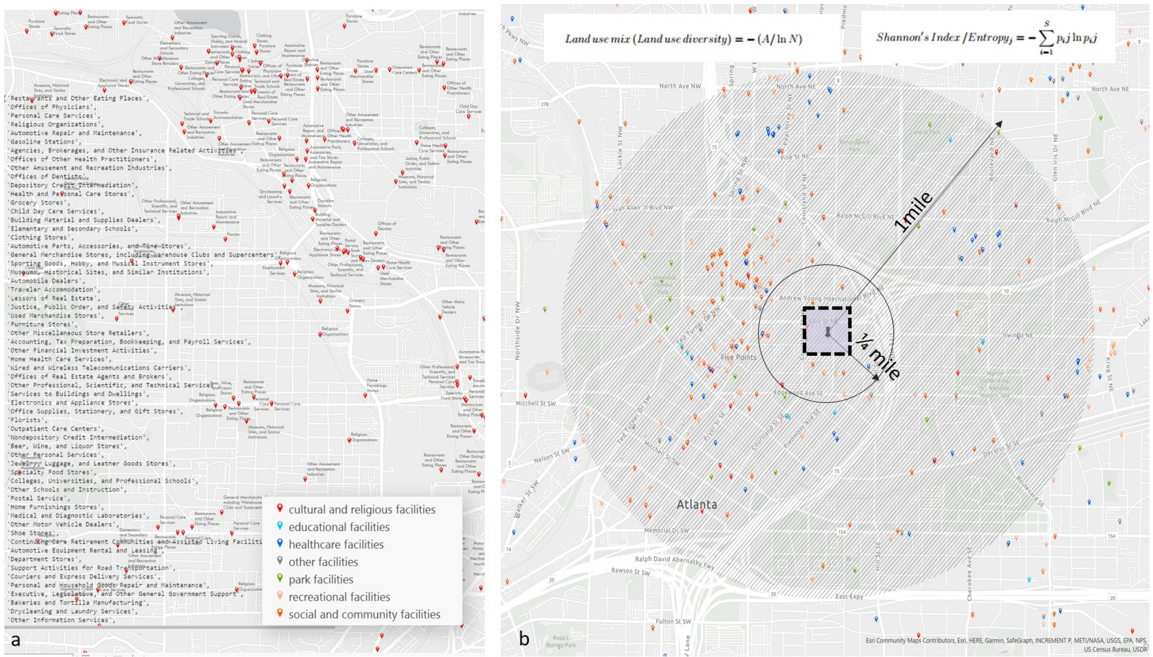

Figure 4.5: a) More than one 200 different categories of POI from Safegraph and Google Places are classified into green escape facilities, social and community facilities, recreational facilities, cultural and religious facilities. b) Diversity indices and accessibility to these facilities are measured within 1/4miles and 1mile of the assessment grids.

by Equation 4.2.

$$Shannon's\,Index\,/Entropy_j = -\sum_{i=1}^{S} p_i j \ln p_i j \tag{4.2}$$

Where $p_i j$ is the relative abundance (proportion) of point of interest of category $_i$ around the assessment grid point $_j$. Relative abundance is within distance d is measured by a_{ijd}/N, where $a_{ij0.25}$ and a_{ij1} are the numbers of point of interest of category $_i$ around 0.25 miles and 1miles of j^{th} assessment grid point respectively. N is the total number of points of interest of all categories within the city boundary.

Public Transportation: McCay highlighted that 'safe active transport' to the escape facilities and activity areas provides opportunities to de-stress [154]. On top of being proximate to the different facilities, ease of access to these facilities is an important criterion. Provision for multi-modal transportation options to reach the facilities may encourage people to visit the venue to relax and de-stress. To assess the opportunity to avail public transportation, I used the bus route and subway routes data for Atlanta (MARTA)[9] and Boston (MBTA)[10] from city's open source data platform. I computed two variables a) access to number of bus routes and b) access to the city's train transit systems within the buffer of 1/4 miles, and 1 mile of the assessment grid.

4.2.3 Measuring Crime and Indicators to Assess Crime Perception

'Fear of crime' is a leading cause of stress in people living or engaging in any type of urban activity [155, p 89]. Fear of crime, affects individuals more than the real crime reports, and it is usually determined by the quality of the built environment and people's perception of safety [156, 157].

Streetscape: Physical characteristics of urban streetscape contribute to people's perception of safety. *Cross-sectional proportion (CP), street wall continuity (SWC), buildings per length (BPL), and tree canopy coverage (TCC)* are the characteristics of streetscape that explains 46% of safety perception in the place pulse data generated through crowdsourcing by MIT Media Lab [158]. The *cross-sectional proportion (CP)* is the average height of the buildings in a street segment divided by the width of the street. Narrow streets enclosed by tall buildings show a larger cross-sectional proportion, while wider streets with relatively short buildings represent a smaller cross-sectional proportion. *Street wall continuity (SWC)* is the proportion of the street that is enclosed by building facades. *Buildings per length (BPL)* is the count of buildings along a street centerline segment per 100m. BPL additionally accounts for the sense of enclosure when people are out on the street. The *tree canopy coverage (TCC)* is the proportion of area between two opposite edges of the street that is covered by a tree canopy. Other variables added to the list of variables measuring percep-

[9] Link to Atlanta Regional Commission open data source for transit route, bus-stops location and station locations for MARTA: https://opendata.atlantaregional.com/

[10] Link to Massachusetts GeoDOT open data source for transit route, bus-stops location and station locations for MBTA: https://geo-massdot.opendata.arcgis.com/

tion of safety were, ‘the number of vacant lots’, street lights along street segments, and street segments with dead-ends (cul-du-sacs). Vacant lots are associated with an increased risk of violent crime [159]. Street lights provide a sense of safety to pedestrians [160]. While dead-ends are controversial when it comes to the perception of safety, one group of scholars claimed that cul-du-sacs are potentially safe areas for people to interact, compared to main thorough fares [28, 161, p122]. However, others have argued that dead-ends are secluded and unsafe [16].

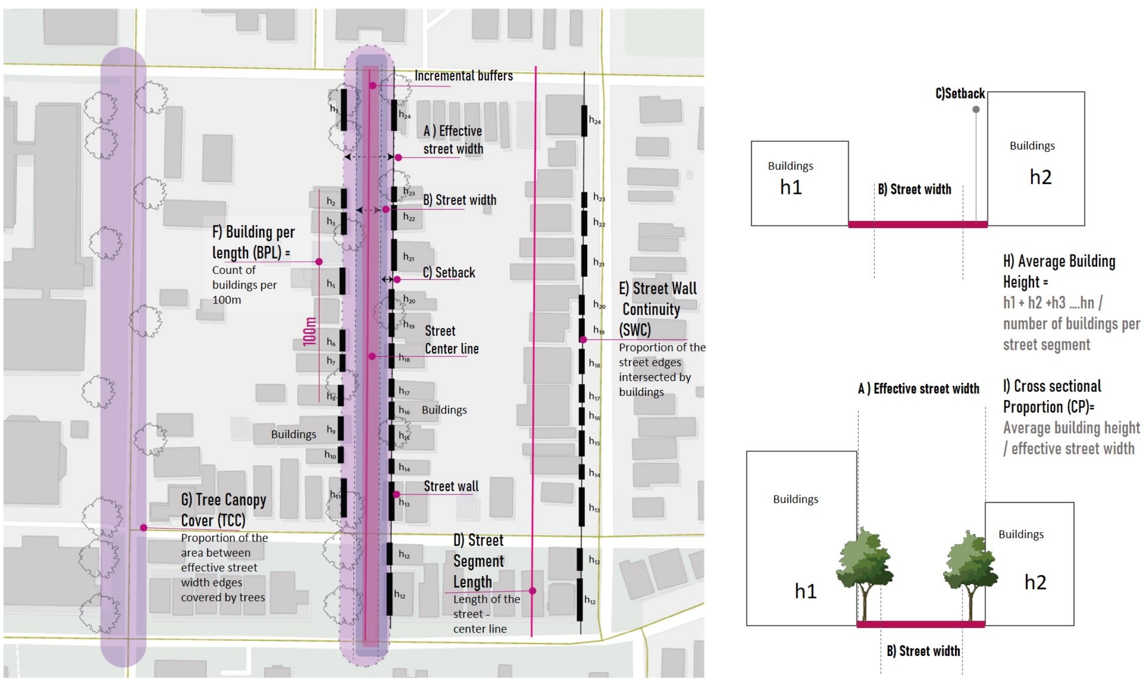

Figure 4.6: Diagram showing streetscape measures used. (A) Effective width of the street; (B) Street width; (C) Setback; (D) Length of the street segment; (E) Proportion of the street edge that has been intersected by buildings; (F) Building per 100m length of the street segment; (G) Proportion of tree canopy coverage; (H) Average height of the buildings on both sides of the street segments; (I) Cross sectional proportion of the street. The measures are modified from those developed by Harvey et al. [162].

The datasets used to measure the dimensions of urban streetscapes are - street segment (street center-line), building footprints, and tree canopy data. These are publicly available for both Atlanta and Boston. The street center-line data is obtained from the Open Street

Map using OSMnx Python package.[11] Atlanta has open-access building footprint data.[12] The building heights were obtained through aerial light detection and ranging (LiDAR) data processing using open source LiDAR data from Fulton County. High-resolution tree canopy image data was obtained on request from the Center for Spatial Analytics and Visualization (CSPAV), Georgia Tech. Boston has publicly available data for the building footprint with building height attribute and publicly available tree canopy image data.[13] Due to the unavailability of precise sidewalk data for Atlanta, I omitted it as one of the potential variables for assessing streetscape. The vacant lots data was obtained from the tax assessor's land parcel data; this data was used to calculate the land use mix. While Boston has open-access street light data, Atlanta's street light data were obtained on request from CSPAV, Georgia Tech. Furthermore, this data was obtained using object recognition techniques from google street view images, as such, it is a proportionate measure of street lights and not the accurate number of street lights along any street segment.

Cross-sectional proportion (CP), street wall continuity (SWC), buildings per length (BPL), and tree canopy coverage (TCC), number of vacant property, and dead ends were calculated for each street segment. I used the primary, secondary, and tertiary road segments and omitted the interstate major state highways. Center-line data for road segments were manually explored and edited to get rid of duplicates and streets with medians. Street segments less than 100 feet in length were either omitted or joined with adjacent street segments if they did not end in a dead-end. ArcGIS Pro was used to edit and prepare the road segments, and *Arcpy*, the Python package, was used to calculate the relevant variables for each road segment. A similar GIS-based approach to compute the streetscape was used by Harvey et al. [158]. Next, I drew incremental buffers for each of the segments, starting from a buffer width of 10 feet (20 feet actual width) until it intersected building footprint edges, either on both sides or at least on one side of the street's center-line. The effective

[11] Link for OSMnx Python package: https://osmnx.readthedocs.io/en/stable/
[12] Link to Atlanta open data source:https://opendata.atlantaregional.com
[13] Link to Boston open data source: https://bostonopendata-boston.opendata.arcgis.com

street width 'w' was recorded when the buffer intersected with the buildings. This idea of incremental buffer is used by Harvey et al. to obtain streetscape measures of New York city [158]. I have slightly modified this idea to obtain streetscape measures for Atlanta and Boston. Using *Arcpy*, a buffer for street center lines was drawn with increasing width in every iteration until it intersected the object of interest. The object of interest can be tax parcel or building footprint geometry, depending on the end-goal of the measurement. Once the intersection with the building or parcel was found, another buffer with a minimal fixed increment value was drawn to confirm the intersection and record the effective street width. This last increment was necessary when the streets were curvilinear and had irregular building setbacks to ensure precise measurements.

Depending on the street center line curvatures, and buildings alignment with street edges, there were several edge cases. The edge cases were programmatically addressed and fixed. For example, when buildings were not aligned to street edges, after finding the fist intersection, an extra incremental buffer was drawn from the intersecting buffer to find the actual number of buildings. If the number of buildings intersected remained the same, then the effective street width recorded was kept the same; else the effective street width was increased by 10%, and the number of buildings intersected and their average height was updated. The effective street width buffer was used to calculate the CP, and TCC. Here, $CP = \left(\sum_{i=0}^{n} h_i/n\right)/w$, h_i is the height of i^{th} buildings, n number of buildings along the street segment, and w is the effective street width. Furthermore, $TCC = \left(\sum_{i=0}^{n} Area_{treei}\right)$, $Area_{treei}$ is the area of the i^{th} polygon of the tree canopy cut by the effective street width buffer. In both cases i varies between 0 and n. To obtain SWC and BPL, the building edge length e and number of buildings n intersected by the increment buffer was calculated. Where, $SWC = \left(\sum_{i=0}^{n} e_i\right)/L$, e_i is the i^{th} building edge length intersected by the effective street width buffer, L is the street center line segment length. And, $BPL = n/L_{100m}$ is the number of buildings intersected by the effective street width or increment buffer and L_{100m} length of the street center line segment in 100 meters (refer Figure 4.6). Furthermore, the number of vacant lots per street segment and number of street lights was calculated using the effective street width buffer. The number vacant lot intersected by the effective street width buffer, and the number of street lights within the effective street width buffer were calculated for the respective measures. Dead-ends were found using ArcGIS pro and Arcpy Python package. In this algorithm, first, I filtered start and end points of the road segments, and then I removed the list of start and end points (from the filtered points) that matched the list of intersections (intersecting points with other road segments). The remaining points in the filtered points were the dead-ends.

The assessment grid points were used to draw buffers of 1/4 miles and 1 mile distance to select the street segment features, and clip the longer segments. The ratio of the clipped segment and the original segment CS_r was used as a factor to compute the aforementioned streetscape variables for each of the clipped segments. $C_r = L_c/L_o$, where L_c is length of the clipped segment, and L_o original length of the segment.

Actual Crime: Stress due to 'fear of crime' can be aggravated by actual crime reports. The term social pathology was used interchangeably in geographic literature to address crime, and mental health conditions [163]. Crime data was obtained from publicly available data for police reports of respective cities.[14,15] Police reports for the cities contain geocoded crime locations. I used crime reports for two years, i.e., between April 2018 and March 2020, the same time frame as the tweets used in the research. Only two year period was chosen as recent memory of crime incidents usually tend to be most impactful in generating stress [164]. Each city has slightly different crime categories enlisted in the dataset. For example, Boston's crime data included fine-grain crime categories like traffic rule violations, while Atlanta's crime data did not. I included the common crime categories from both cities to keep the data comparable. Furthermore, the crimes were classified as *misdemeanor* and *felony*; these categories were defined based on the aggression and seriousness

[14]Link for crime data Atlanta: https://www.atlantapd.org/i-want-to/crime-data-downloads

[15]Link for crime data Boston:https://data.boston.gov/dataset/crime-incident-reports-august-2015\-to-date-source-new-system

of the committed crime. For example, homicide, robbery, and aggressive assaults were classified as felonies, while thefts, burglary, and petty crimes were classified as a misdemeanor. The classification is adopted from criminal law [165]. Finally, the assessment grid points were used to calculate the number of crime incidents within 1/4 miles and 1 mile buffer distances.

4.2.4 Measuring Socio-economic and Overcrowding Indicators

The built environment and mental health literature used the term crowding interchangeably with density [28, p70]. To disentangle the impact of density from overcrowding on stress levels, I used different socioeconomic and demographic variables to assess the BE quality of tweet location. Population density, employment density, poverty rate, rent burden, and eviction rate were the indicators used in this research.

Population and Employment Density: Population data for census block groups were obtained from American Community Surveys' 5-year estimates of population data between 2015-2019. The key advantage of using multiyear estimates is primarily the statistical reliability of the data. The population data contained information on race and total area population. The number of people employed was obtained from the longitudinal employer-household dynamics (LEHD) data, origin-destination employment statistics (LODES) for 2018. [16] Origin is the home census block, and destination is the work census block. I used the workplace area characteristic (WAC) field to obtain the total number of jobs for each work census block in Atlanta and Boston [166]. Next, I computed the population and employment values within 1/4th mile and 1 mile buffer using proportion measures. Population is equal to $p_i * P(A_i)$, where p_i is the population of i^{th} census block A_i and PA_i is the proportion of the census block clipped by the 1/4miles and 1mile buffer. A similar calculation was done to obtain the employment measure, assuming equal distribution of population and employment across census block.

[16] Link to the LEHD, LODES dataset: https://lehd.ces.census.gov/data/

Socio-Economic Conditions: To account for overcrowding (which I have established is different from density measures in subsection 2.2.5), I used poverty rate, rent burden, and eviction rates as measures. The poverty rate defines households of given family size that are below the poverty threshold. Each state defines its threshold value for poverty. For example, for a 4 person household living in Georgia, the poverty threshold is 25,750 US dollars annually in the year 2019. That is for a 4 person household, if the annual household income is below 25,750, then they are considered below the poverty threshold. Rent burden is the percentage of household income paid in rent. Any household paying 30% or more of their income in rent is rent-burdened. In the data, that is used in this study, 50% is the maximum rent burden, which means if any household is 50% rent-burdened, they are paying 50% or more of their annual income in rent [167]. While most renters in the city of Boston and Atlanta area were rent-burdened, there were about 18-20% residents who were severely rent-burdened, disabling them to rent livable good quality spaces for each household member. The eviction rate defines the number of evictions per 100 rented households in an area. Any census block group above 16% eviction rates in Atlanta is considered among the top 1% in eviction filling rates, while in Boston, this statistic could be as high as 50%. Low-income potential and cost burden is stress generators, and households affected by poverty, cost burden, and evictions can face undesirable effects of overcrowding (refer subsection 2.2.3. The data used in this research was obtained from an open-access data source from Princeton eviction lab [168]. A proportionate measure of poverty rate, rent burden, and eviction rates was calculated for each assessment grid point for 1/4 miles and 1 miles buffer radius.

4.2.5 Measuring Urbanness

Urban areas and built environments are claimed to be precursors of stress in the mental health literature [169]. However, an important question that needs to be asked is: what constitutes urban? The census definition of cities is dependent on the population size or the administrative boundaries. Urban designers and planners have argued that the urban and suburban areas should be differentiated based on built environment characteristics as opposed to administrative boundaries [170]. Urban administrative boundaries do not capture the differences in streetscape (refer Figure 4.6, diversity of opportunities and amenities. To measure urbanness (metric computing, how urban an area is), we need to rely on specific built environment characteristics. These characteristics include measures to assess crime perceptions and measures to evaluate access to urban amenities like recreational facilities and parks. For example, a higher degree of urbanness can be associated with areas that have: (1) higher cross-sectional proportions (CP), (2) greater street-wall continuity (SWC) (refer Figure 4.6), (3) better access to urban facilities and (4) moderate to high public transportation options (refer Figure 4.5). These urbanness measures were additional features of BE that I added to ensure that the definition of urban is not vague, particularly in the context of its association with mental health and wellbeing. To add these measures of urbanness, I have adopted the definition of *urban form* from Dunham-Jones et al. [171].

According to Dunham-Jones et al., urban form is typically characterized by both mixed-use buildings and attached buildings, fronting sidewalks with either underground parking or parking behind the buildings. Single-use buildings surrounded by lawn or parking lots are typical of a suburban form. There is less auto dependency in areas of mixed-use and transit-oriented urban developments. Urban roads are laid in grid patterns, with interconnected networks as opposed to the dendritic suburban roads with dead-ends (cul-de-sacs). The block sizes in the urban areas are smaller compared to those in the suburban areas. Urban areas are denser than suburban areas, and the density is less uniformly distributed. Furthermore, urban areas are known for their qualities such as imageability and legibility [172]. In this work, these abstract qualities were added by the presence of historical landmarks, architectural structures, and other artistic, physical monuments.

Road connectivity: This measure was computed using intersection density and node ratio [173]. Intersection density is the number of street intersections per square miles (T_i/A_{mi}), where T_i is the total number of intersections including dead-ends within a buffer radius of i. Here, A_{mi} is the area of the buffer with radius i in square miles. Node ratio is the ratio between the number of road segments between intersections and the total number of intersections including dead-ends (N_s/T_i), where N_s is the number of road segments between intersections.

Setback: Setbacks are defined as the distance of the building from the street. Setbacks are part of zoning ordinances and building codes created for environmental protection/safety and to prevent property owners from crowding each other's property. However, setbacks if not designed properly, could alienate people from the streetscape and landscape. Larger setbacks reduce the spatial experience of being in a well-designed outdoor space and lack good cross-sectional proportions; increase the sense of pedestrian vulnerability; and replace the pleasures of window-shopping with big commercial signage. All that said, small setbacks filled with cafes or stoop gardens can increase the attractiveness of a street. Suburban property setbacks occupied by concrete driveways, arid grass, or wheelie bins appear to be monotonous, creating little or no memorability of the space [174]. I calculated setbacks for street segments using an incremental buffer approach. The incremental buffer approach has been explained in the subsection 4.2.3. From the previous measures, such as street cross-sectional proportion, I used the effective street-width buffer, which is essentially the street width, including the sidewalk area and the setback (refer Figure 4.6). To calculate the setback, I computed incremental buffers from the street segment until it intersected with the property line defined by tax parcels. I called it street width buffer. I ignored all the street segments that were within a large tax parcel and assigned them zero setback value. This was to make sure we do not include any erroneous edge cases for interior roads where roads were within a tax parcel. Figure 4.7a delineates setback diagrammatically. Noteworthy to mention that setback is the difference between the effective street width and the street width: Abs($w - s_w$), where w is effective street width, and s_w is the street width (Abs is the absolute value function).

Landmarks: Landmarks are physical structures of art, architecture, or landscape architecture that accentuates memorability of a place [12]. Landmarks not only impart historical relevance to cities but also act as way-finding instruments for people. The presence of a landmark attaches a high symbolic value to a place. To compute the presence of any landmarks at any point in a city, I used the dataset for: (1) historic landmarks and (2) landmark areas from the open-access data for both Boston and Atlanta. Additionally, I included a collection of landmark points of interest (POI) as crawled from the Google places API [175].

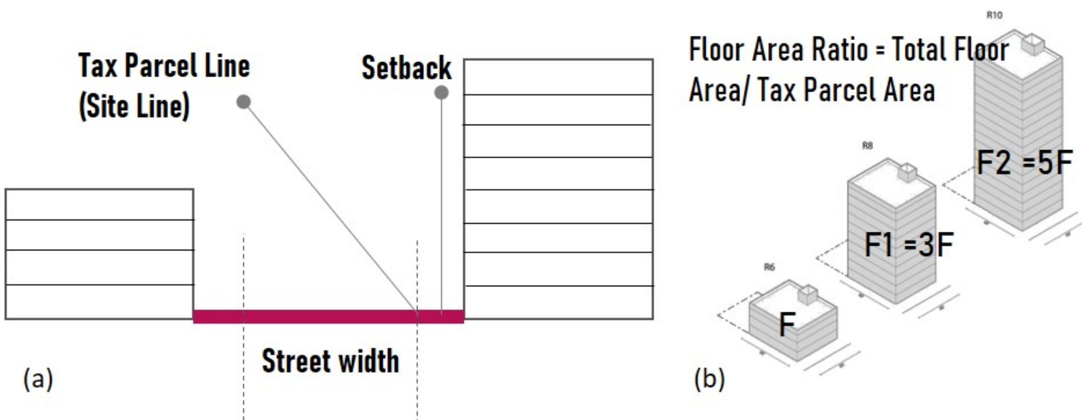

Figure 4.7: a) Diagram shows how setback is calculated using tax parcel and building footprint. b) Diagram shows how floor Area Ratio (FAR) is calculated. FAR can vary with the variation in height/number of floors of the building on the same site.

Density Distribution: Distribution of density is a collection of measures to quantify urbanness [171]. One such measure is floor area ratio (FAR), which is used to calculate building density. FAR is the ratio of total usable floor area in a building and area of the land parcel $\sum_{i=0}^{n} FA_i/PA$, where FA_i is the total usable floor area for the i^{th} floor and PA is land parcel area. Building footprint data with building height information was used to estimate the floor height of buildings.[17] Figure 4.7 show how FAR is calculated and how FAR can

[17]Accurate building heights were obtained from digital surface model (DSM) generated from Fulton county LiDAR data Atlanta. Digital surface models are a type of raster data generated from the first returned laser pulse used in optical remote sensing. This remote sensing technique is called LiDAR. DSM models contain building height information that was processed and joined to building footprint vector data or

vary with the increase in the building height and the number of floors.

The product of building story attribute and building footprint area was used to estimate the total usable floor area $\sum_{i=0}^{n} FA_i$ of the building. The FAR for the land parcels was obtained after spatial join of the building data and the tax parcel data. The final dataset created through spatial join contained the tax parcel data with the total building floor area attribute; this was used to calculate the FAR for each parcel. To measure density distribution, the difference between the 75th percentile FAR and 25th percentile FAR within a 1/4th mile and 1 mile buffer was used. The difference between percentile values are used to avoid outlier (very high or very small) FAR values computed due to the presence of any unrealistically large/small shapefile in the tax parcel data.[18] Density range closer to zero was considered to be less urban. Using the 'Arcpy' python package in ArcGIS Pro, various measures, including the road connectivity, setback, access to landmark, and density distribution, were proportionately calculated for both $1/4$ miles and 1 mile buffers around a persons' tweet location.

4.2.6 Measuring Environmental Stressors

Weather conditions can act as a stressor or de-stressor. For example, very high/low temperature, low pressure, or lack of sunlight are factors that are often associated with increased levels of depression. This effect was also seen in daily cases of psychiatric emergencies [177, p31-32]. To account for such environmental stressors, I sourced the climate data available online from the National Centers for Environmental Information [178]. This data contained daily average, maximum, and minimum statistics for variables including temperature, precipitation, and pressure. Utilizing each tweet's 'created' date-time record, I obtained the daily temperature, precipitation, and pressure from the aforementioned cli-

shapefile.

The building height attribute data was available for Boston's building footprint dataset [176].

[18]There were several tax parcels with very high and unrealistically high FAR. One of the reason is, these tax parcels are tiny rectangular data points to account for any co-ownership or planned unit development (PUD). FAR computed for these properties are extremely high.To avoid these outlier values the difference between percentile is assumed a reasonable mesure though trial and error.

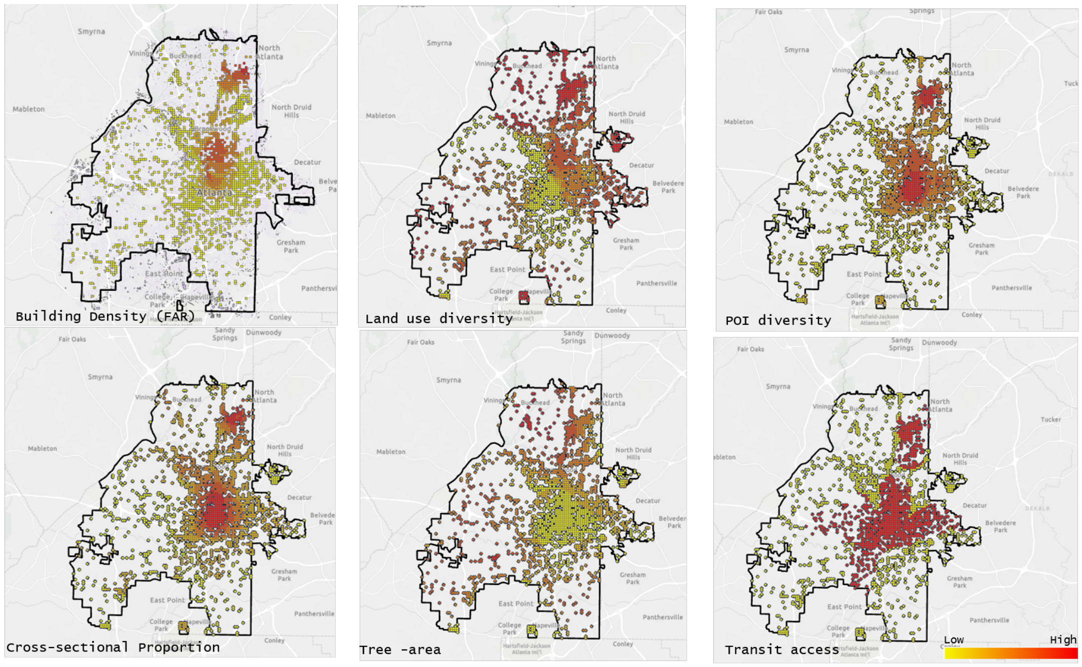

Figure 4.8: Assessment grid maps of Atlanta showing building density, land use diversity, POI diversity, street cross-sectional proportion, tree area, and transit access computed within 1 mile buffer radius. Only the assessment grids containing tweets are represented here and used for analysis.

mate data. Subsequently, I tagged tweets that occurred in comfortable temperature ranges, i.e., between $60°$ - $80°$, normal pressure ranges between 29.3 - 30.1 mmHg. Next, I computed the number of days without any precipitation and thunderstorm. These categorical values were converted to binary values (0, 1) to prepare a numerical dataset that a machine can understand for downstream tasks (explained in the following sections). For instance, days with precipitation and thunderstorm was coded as 1, and days without precipitation was coded as 0. The binary values were summed for each assessment grid using spatial join *field-mapping* technique using the *Arcpy* library.

Another key environmental stressor is noise. The literature points out that proximity to traffic-heavy streets and airports may induce higher levels of stress to the residents. In this research, variables related to traffic noise were estimated using the Federal Highway Administration's Highway Performance Monitoring System (HPMS) data [179]. The assumption made here was that the average annual daily traffic (AADT) count is correlated to noise in the major through streets. Evidently, I used AADT as a proxy for noise. Like all other spatial measures, I used the 'Arcpy' Python package to calculate the average annual daily traffic count for both $1/4$ miles and 1 mile buffers from a persons' tweet location.

4.2.7 Measuring Impact of Time on Social Engagement

The timestamp of social media posts was used for user-level stress detection. Research shows that users expressing stress and depression tend to be active during late nights and early mornings [180, 181]. Although in this research we are not predicting user-level stress, I was motivated to investigate if there were specific time windows when people stress or de-stress. To that end, I classified each tweet into time categories such as, early morning (4am-7am), morning (7am-12noon), afternoon (12noon-4pm), evening (4pm-10pm), night (10pm-12am), and late-night (12am-4am) based on the 'createdAt' variable. Similar to the technique used in computing environmental stressors in subsection 4.2.6, I transformed the time window categories into binary values (0,1). Finally, I summed these binary values for each assessment grid using the spatial join operation to retrieve what proportion of tweets in each grid were in different time windows.

4.3 Dataset: Text Data/Tweets

4.3.1 Data Overview

I used Twitter microblogs or tweets to identify stress experienced by people in urban spaces. Tweets are microblogs posted by Twitter users to record their thoughts, express views and share information using a maximum of 280 characters (since 2017). Twitter also allows cross-platform posts. Users can post on other social media accounts, such as Instagram, and choose to share the same post on Twitter. The concise format makes it easy for users to share and update content. Besides sharing personal views, users can communicate with one another via private messages, retweet (share) each other's tweets, reply, mention, and participate in conversation channels using hashtag identifiers [182]. Furthermore, the communication is streamlined in this platform through a *follower* and *followee* relationship. Users in Twitter can either follow any user or their interest, thus becoming the follower for the target user or followee. In that case, the user will be able to track the followee's tweets easily and engage either through retweeting, replying, or liking their posts.

Twitter has 69 million active users only in the United States [183]. In addition, Twitter population is generally found to be a younger, more democratic inclination, higher education, and higher income than the US adult population. Gender is equally distributed. Research shows that a Twitter user's opinion differs from the broader US population on a few social issues. They seem to favor immigration, and frequently voice their opinion against race and gender based inequalities. However, on other subjects, Twitter users are not dramatically different from those expressed by all US adults [184]. 10% of Twitter users are considered prolific users, accounting for 80% of the tweets, while the other 90% of the users are responsible for 20% of the tweets. The top 10% engage more than the bottom 20% of the users. Despite these differences in user activity, the top 10% active users are not different from the bottom 90% of the Twitter population in their social viewpoints (i.e., their views are identical) [185].

Even with the current limitations, researchers have found Twitter to be a platform that enables public disclosure, opens democratic exchanges on many controversial topics [186]. As mentioned in Chapter 2, Twitter data has been used to examine several topics, including climate change, lesbian-gay bisexuality, transgender rights, gun violence, abortion, postnatal depression in women, and many others. Twitter users have not only offered a second-hand opinion but also have made self-disclosures on these topics. Previous research and other relevant findings depict that Twitter data can be safely used to gauge the mental well-being of individuals in the context of an urban built environment (refer Chapter 2).

4.3.2 Data Collection

I programmatically scraped tweets over two years time period. To that end, Twitter 4J API [187] was used to crawl over 2 million geotagged and non-geotagged Tweets between 1^{st} of May 2018 and 31^{st} March 2020. I intentionally chose the end date of data collection as 31^{st} March 2020 to avoid overlap with the COVID-19 pandemic. The query radius specified was 100 miles around the city centers of Atlanta and Boston. The 100 miles radius is an arbitrary radius chosen to ensure that I could capture maximum tweets within a 30 minutes driving distance from the city center. To give an idea of scale, it is noteworthy to mention that, the distance from the city center to the city boundary of Atlanta is roughly 10 miles. For Boston, the average distance of the city boundary from the city center is 6 miles. The crawled Twitter data was stored as a collection of comma-separated values (csv) files each containing attributes such as, time of post (createdAt), place, hashtags, reTweets, replies, and favorites. Geolocated tweets could be represented as a set of point data in ArcGIS Pro with the tweets' location information embedded with latitude and longitude attributes. Figure 4.12 shows a blown-up version of geolocated tweets in a assessment grid. However, the location tags available on Twitter or other LBSM (location-based social media) posts shared on Twitter showed a varied level of accuracy. The most fine-grained Twitter location prediction was possible within 16 miles of the actual Tweet location [188]. As less than 1% of all tweets were geolocated, it was vital to retrieving enough relevant tweets for stress analysis.

In order to increase the sample size of geolocated tweets, I used two strategies: (1) collected tweets (using Twitter4J API, Java [189]) over a period of two years, and (2) assessed the location indicative metadata in the non-geolocated tweets to assign location attributes to them. Tweets crawled using Twitter API contained a 'place' object which encodes location. The 'place' object included information such as business name and street address. Some tweets had location cues in their text content, such as place names, place #hashtags, place @mention, or some combinations of these. Refer Figure 4.10 for a list

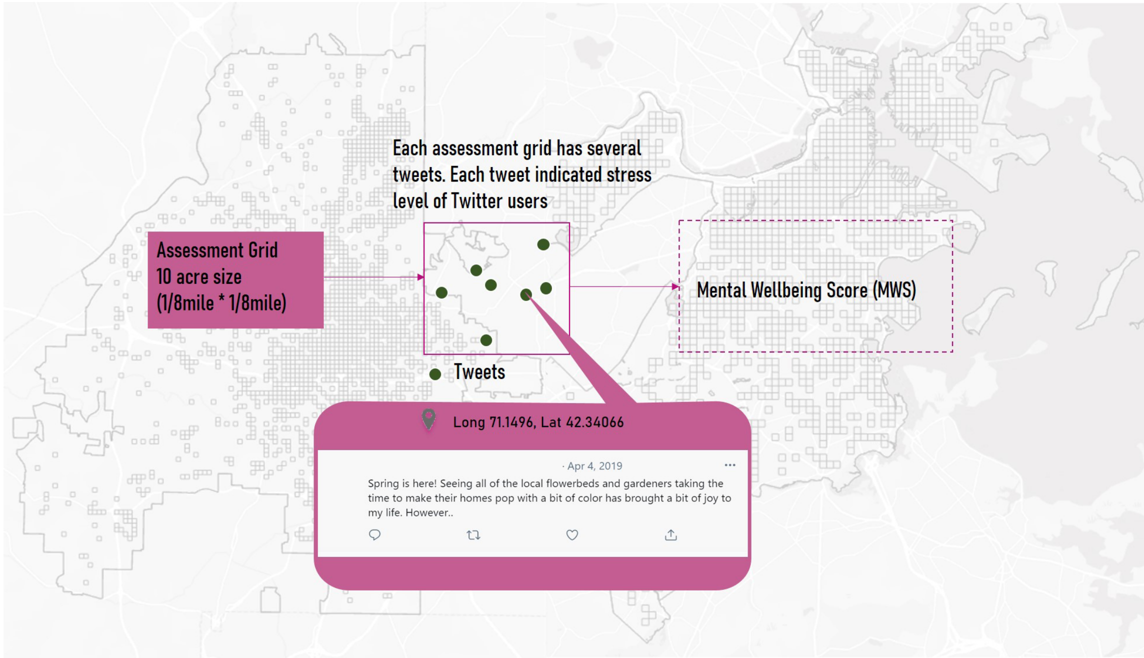

Figure 4.9: Each assessment grid has several geolocated tweets, and each tweet indicates stress level of twitter users which is used to compute the mental wellbeing score MWS of the grids. An example tweet showing tweet text, latitude, longitude and time stamp. For data privacy tweet handle has been removed.

of top hashtags. These metadata and cues were used to geolocate additional tweets. These additional geolocated tweets accounted for 0.1% of the non-geolocated tweets [190]. In this research, I selected the geolocated tweets within the city boundary of Atlanta and Boston. Once the geolocated tweets were collected, and a supplemental dataset from the non-geolocated tweets was programmatically geo-tagged, the next task was to explore this dataset and remove any noise from the data.

4.3.3 Cleaning Twitter Data

The tweets were collected in batches for two years (refer subsection 4.3.2 for the timeline of data collection). Batching the files into a set of relatively smaller CSV files ensured we could explore individual files without running out of memory or compute resources (e.g., each CSV file was 10 GB in size approximately). Spot-checking, a subset of these files, made it clear that the initial dataset had substantial noise in it. To maximize the search

Stress		De-Stress	
#mentalhealth	#wellness	#destress	#love
#depression	#today	#relax	#spa
#health	#care	#selfcare	#mindbodysoul
#anxiety	#parenting	#massage	#wellbeing
#ptsd	#mentalhealthmatters	#meditation	#happiness
#mentalillness	#thanks	#relaxation	#health
#work	#life	#mindfulness	#egai
#hope	#issues	#stressrelief	#breathe
#time	#stress	#wellness	#stressmanagement
#recovery	#read	#mentalhealth	#calm
#person	#adhd	#stress	#healthylifestyle
#alzheimers	#pain	#yoga	#holistichealth
#tumblr	#lonely	#unwind	#tired
#mood	#alobe	#selflove	#therapeuticmassage
#heartbroken	#sadness	#massagetherapist	#bhfyp

Figure 4.10: Top Twitter hashtags associated with Tweets expressing stress and de-stress.

space for data, I did not specify/adopt a keyword-based search that automatically filtered most of the irrelevant tweets. While some of these irrelevant tweets were bots, advertisements, and duplicate tweets, a large set of other tweets were second-hand experience or web links. I performed exploratory data analysis on a randomly sampled subset of 5000 tweets to identify irrelevant or noisy tweets. A list of keywords (both individual tokens and a set of n-grams[19], refer Table 4.2) frequently appearing in this subset was saved. Based on the keywords, I filtered irrelevant tweets from the data corpus. Furthermore, I used cosine distance and jacquard similarity metrics to measure the distance between the tweet corpus and a list of top-k frequently occurring irrelevant tweets. If the distance/similarity metric between the pair of tokenized tweets was above a pre-specified threshold distance value (higher value means more similar to each other), then I discarded them as irrelevant tweets. The optimal value of this threshold distance was obtained through an iterative trial and error process. In addition, I checked and removed all duplicate tweets based on the creation date/time and text content of tweets. Noteworthy to mention that this data cleaning process was exploratory and iterative. It was tested on multiple subsets of the data before implementing it on the entire dataset.

[19]n-gram is grouped words. Each word or token is called a ‘gram’. Creating a vocabulary of two-word pairs and three-word group is called a ‘bigram’ and ‘trigram’ respectively [191].

4.3.4 Classifying Stress

Stress Class Definition: The fundamental goal in this research is to assess the level of stress inferred through the text contents present in tweets. The categorical values of stress or de-stress assigned to the tweets was determined based on an individual's situational response. For example, a user might be momentarily de-stressing by meeting his/her friends in a restaurant after work hours, where he/she tweets - "I am super-excited to be in a friend's reunion here in Atlanta." It was classified as a de-stress tweet. Meaning, the classification label captured the situational stress level as'low'. The situational stress did not account for a persons' ongoing or long-term crisis. The premise of this finding holds for the situation, i.e., at the time the tweet was posted.

The classification method adopted in this work was inspired by Saha et al. [192], and Bagroy et al. [193], where binary classification models were trained to classify tweets into high-stress and low-stress categories. The choice of the two stress classes: (1) low stress (labelled as de-stress), and (2) high stress (labeled as stress), derived in the research was motivated by the psychometric measure of stress given by the perceived stress scale (PSS) [194]. PSS score identifies three categories: minimal stress for scores ranging between 0-13, moderately stressed for scores ranging between 14-26, and extremely stressed for those scoring between 27-40. Typically it is rare to find people in the extremely stressed category. People are either suffering from chronic stress or are affected by severe mental health disorders. Following a similar line of thought, I created two classes, 'stress' and 'de-stress'. Tweets with low level of stress, essentially shared relaxation or recreational activities were labeled as de-stress. Tweets that showed a moderate to high level of stress were classified as 'stress' tweets. To further ensure that only stress and de-stress tweets are captured in the final dataset the probability score was used. Tweets classified as stress and had stress probability score above 70% was retained as stress tweet. Tweets classified as de-stress and had stress probability score below 30% was retained as de-stress tweets.

For example, *"Traveling alone doesn't have to be boring. Today I was off work and*

Table 4.2: Examples of relaxation and stress n-grams retrieved from the tweets.

Stress	**De-stress**
traffic	sleep
begging	nap
bored	dinner
losing it	potlucks
feeling crap	TV watching
I am sick	meeting friends
stuck in traffic	took a shower
Atlanta traffic jam	baking for fun

took myself, and my tripod to the Walking Dead filming location ... " is an example of a 'de-stress' tweet. Similarly, *"I just drove up to this accident and found a mother screaming while getting one toddler out of the car. I ran over ... "* and *"Coming through Atlanta tonight during rush hour was a challenge. Stop and go bumper to bumper traffic during ... "* are a few examples of 'stress' tweets.

Training, Test, and Validation Data: To build the classifier, I utilized a set of *self-disclosure* tweets, where the user discloses their experience of stress or de-stress. The self-disclosure tweets were obtained by filtering them based on stress-related and de-stress related keywords and hashtag identifiers from the whole corpus of over 1 million tweets for each city (Atlanta and Boston). These keywords were recovered from frequently occurring keywords from urban planning related articles and text snippets from the web. This set was expanded after spot-checking the tweet corpus and finding relevant keywords for this research. Doan et al. [195] had adopted a similar approach in identifying the stress and relaxation tweets to develop their training dataset. I filtered the stress and de-stress tweets for both Atlanta and Boston from the entire dataset (irrespective of the availability of the geolocation information). These filtered tweets were used to build the training dataset. There were about 5000 tweets in the training set, out of which there were 1700 stress tweets and 3300 de-stress tweets. I used a set of most frequently occurring keywords related to stress and de-stress (comprising of uni-grams, bi-grams, and tri-grams) from these initial training set to iteratively filter and hand label 11000 more tweets (collected from the set of unlabeled tweets). The final training set comprised of 12000 annotated tweets, where 4500 were 'stress' and the rest were 'de-stress' tweets. In addition, to train the stress classifier (explained later in this section), we set aside 4000 labeled tweets as a test set to evaluate and validate our classification model.

Furthermore, I created a validation set with all the geolocated tweets within the city boundary (for both Atlanta and Boston) that were not used in the training set. As discussed in subsection 4.3.2, to maximize the number of geolocated tweets, the location indicative metadata in the non-geolocated tweets were matched with the POI dataset within the city boundary. The matches were found using n-grams from the tweets and the 'name' attribute present in the POI data (refer Table 4.2). The matched tweets were geolocated by assigning them the latitude-longitude information of the POI. Finally, the geolocated tweets within the boundary of the city of Atlanta and Boston were retained in the validation set for stress classification. For each city, I recovered a substantial collection of (approximately 61000 for Atlanta and 65000 for Boston) tweets as validation data. When the stress classifier was trained well to match the expected performance metrics (e.g., accuracy), I used it to machine label the validation set for both cities. Check Chapter 5 for the final models' performance results.

Text prepossessing: Prior to inputting the data to the modeling pipeline, the data was pre-processed and transformed in ways that improved the performance of the model. Each tweets' text was expressed as a list of words/tokens. I pre-processed these words using standard natural language processing methods, including stemming, lemmatization, removal of stop words and special characters, limiting text length (a hyperparameter that I iteratively adjusted to improve the models' test set performance).

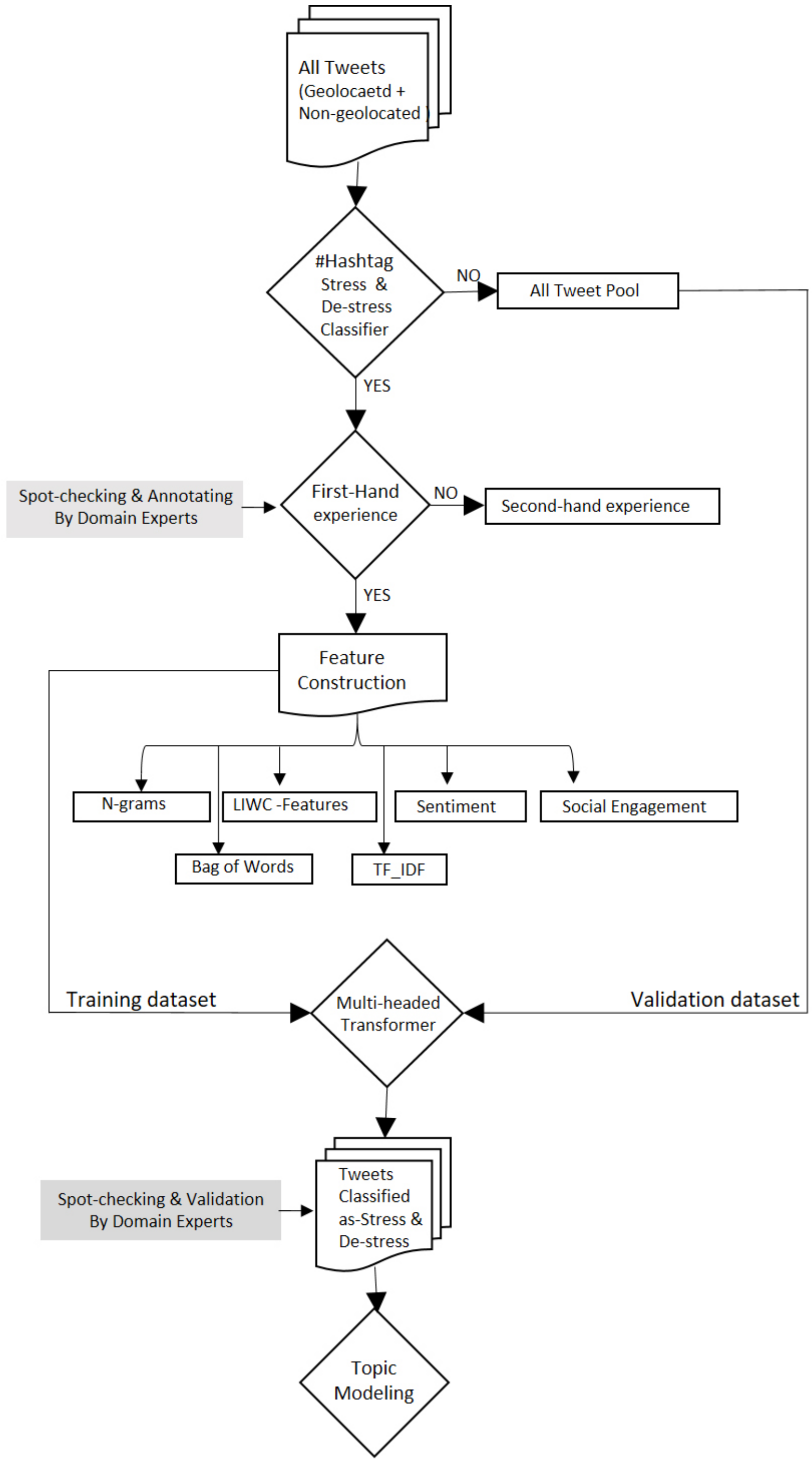

Figure 4.11: The schematic showing the process of building the stress classifier.

Sentiment Analysis: I utilized Stanford's Core NLP sentiment analysis package [196]. For each input tweet post, this package generated a sentiment score ranging between 0.0 to 1.0. I classified the score into 5 different levels, namely, [0.0, 0.2] very negative, [0.2- 0.4] negative, [0.4, 0.6] neutral, [0.6, 0.8] positive, and [0.8, 1.0] very positive [197]. I used the sentiments score and the corresponding labels as a feature for stress classification.

Stress Classifier: For stress classification, a binary dependent variable was created, which had values of: (1) Stress, and (2) De-stress, as labels. This classifier was designed based on a modified transformer model (with multi-headed attention layer) [198]. For many NLP tasks, including text classification, named entity recognition, language generation, and others, transformer models such as BERT are one of the best state-of-the-art models.

Their success in various NLP tasks has led to them being available pre-trained, where they are trained on millions of data samples (and often on multiple languages) by large corporations including Google, Open AI, and others. However, as promising as this is, for nuanced tasks such as sentiment analysis/classification, often these models need to be custom-designed based on the task at hand and the dataset. For the same reason, the stress classifiers' architecture was designed to be different from a conventional transformer model in the following ways: (1) I discarded the decoder from the transformer model, (2) I removed residual connections, layer normalization, and the layer masking (as its' a classification task, not a language modeling problem), and (3) I employed a multi-headed attention with position-wise feed-forward encodings. These changes were inspired by the implementation referred here [199], and upon multiple training iterations, this model design performed the best (in terms of test set accuracy metric). Since this was a classification task to predict stress labels, the final layer incorporated a cross entropy loss with soft-max activation function ($-\sum_0^n y_i log(J(f_0(x_i)))$, J is softmax function). I trained this model by tuning its' hyperparameters, including: batch-size, learning-rate, num-epochs, drop-out-rate, and others. Finally, the best performing model was trained using these hyperparameter settings, batch-size: 128, learning-rate: 0.0015, num-encoders: 5, embedding-dim:

512, division-factor: 64, num-epochs: 100. Furthermore, the model was optimized using the Adam optimizer [200, 201], which is known to perform better in many problem domains than other comparable optimizers such as *ada-grad*, *stochastic gradient descent*, etc. I reduced overfitting by setting a drop-out rate of 0.18, and by utilizing a learning rate decay function. For detailed model performance results please read Chapter 5. Noteworthy to mention that, I adopted a random search approach [202] to tune the hyperparameters of the model with the goal to maximize test sets' accuracy metric. In this approach, a search space of hyperparameter values is defined apriori, with a bounded domain of hyperparameter values. Next, randomly sampled points are used as hyperparameter combinations to train models. In the end, the best performing model was selected (refer Figure 4.11).

4.4 Building the Dataset for Analysis

Selecting Assessment Grids for Built Environment Computation: Once the final set of classified geolocated tweets for the city of Atlanta and Boston was recovered, the dataset was used to prepare the set of assessment grids (AG).[20] The selection operation was done using ArcGIS Pro's *select-features-by-location* operation using the default overlap type *intersect*.[21] The selected assessment grids were used to generate assessment grid points or centroids. The centroids were used to draw a $1/4$ mile and 1 mile buffer for the built environment variable computation. The built environment computation has been elaborated in Section 4.2.

Spatially Joining the Tweets and Built Environment Variables: Along with the stress classification, the tweets also contained classifications for the weather conditions and the time of their occurrence as explained in subsection 4.2.6 and subsection 4.2.7. I used *spatial join*[22] function using the *Arcpy* Python library to join the classified tweet dataset

[20]The assessment grids were drawn using the Autocad software and was converted to a GIS feature. Features were represented in a commonly used vector formatted data with spatial and geometric attributes.

[21]Link for select features by location:://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/select-layer-by-location.htm

[22]Link to spatial join function: ://pro.arcgis.com/en/pro-app/2.7/tool-reference/analysis/spatial-join.htm

to the assessment grids. The *spatial join* function joins attributes from one shapefile to another based on their spatial relationship. The different classification fields in the tweets were joined to the AG using field mapping operation to obtain a final grid dataset containing all the tweet attributes. The fields computed for each AG during the join operation were: number of tweets, number of stress tweets, number of de-stress tweets, the proportion of tweets posted in different time windows of the day, and proportion of tweets posted in different weather conditions. The built environment (BE) variables computed for the AG points were concatenated and joined to the AG feature. Subsequently, I tested if the selected AG had complete BE information. To ensure the same, I discarded the grids that were within a quarter (1/4th) miles of the city's periphery.

4.5 Computing the Mental Wellbeing Score

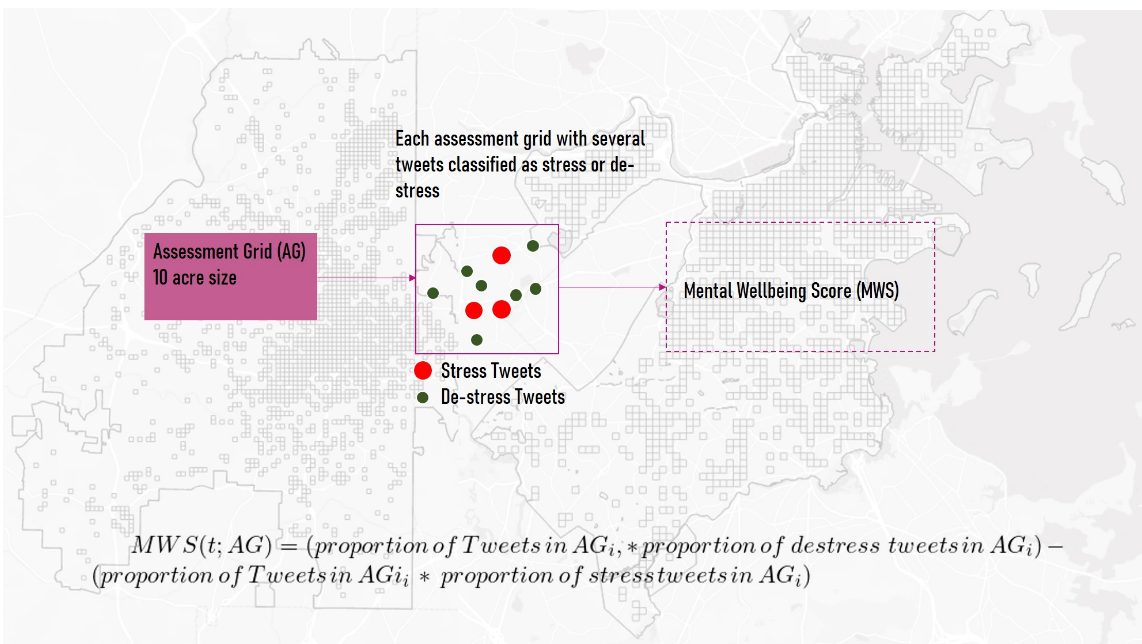

Figure 4.12: Figure shows a sample assessment grid with classified stress tweets and de-stressed tweets. The mental wellbeing score for each assessment grid is the difference between the normalized frequencies of the de-stress tweets and the stress tweets occurring between the time period May 2018 and 31st March 2020.

I constructed a composite normalized value called *Mental Wellbeing Score* (MWS).

This metric formulation is motivated by Bagroy et al. [193] where a mental wellbeing index was used to evaluate mental wellbeing of college campuses. MWS score is a metric defining the net stress level of each $1/8 * 1/8$ square miles grid in space. It is measured by the normalized difference in the frequency of stress tweets and de-stress tweets for a time period. For example, for a time period t and for assessment grid AG_i, the difference between the normalized frequencies of de-stress tweets $f_{ds}(t, AG_i)$ and the normalized frequencies of stress tweets $f_s(t, AG_i)$ is shown here:

$$MWS(t, AG_i) = f_{ds}(t, AG_i)/f_T(t, AG_x)_{max} - f_s(t, AG_i)/f_T(t, AG_x)_{max} \quad (4.3)$$

[23]

where $f_T(t, AG_x)_{max}$ is the maximum frequency of tweets in assessment grid AG_x, and t is the time period between 1st of May 2018 and 31st March 2020. The assumption made here was that, the mental wellbeing of any assessment grid is a proportionate share of the de-stress tweets and stress tweets. Framing the metric in this way allowed me to offset the bias of de-stress tweets or stress tweets. The MWS maps for both Atlanta and Boston are included in the appendix as Figure A.2 and Figure A.1.

4.6 Data Analysis Framework

I developed a data analysis framework that consisted of three key steps: data exploration, choice of modelling techniques, and finding insights from the text data to contextualize the modelling results. Refer Table 4.1, which shows entire set of explanatory variable for MWS.

4.6.1 Data Exploration

Dimension Reduction Techniques: The model to test the relationship between mental wellbeing and fine-grain built environment (BE) characteristics of cities is a novel inter-

[23] $MWS(t, AG) = (proportion\,of\,Tweets\,in\,AG_i * proportion\,of\,de-stress\,tweets\,in\,AG_i) - (proportion\,of\,Tweets\,in\,AG_i * proportion\,of\,stress\,tweets\,in\,AG_i)$

vention. Although the variables chosen for this study are backed by the theory of mental health and built-environment, the measures developed in this research, both outcome variable and explanatory variables, have not been tested before. To ensure the correct choices of variables in the models and to mitigate any multicollinearity effect, I utilized two dimension reduction techniques: (1) hierarchical clustering and (2) factor analysis.

Hierarchical Clustering (HC): Hierarchical clustering reveals similarities between rows or objects in a dataset. Here the rows can be considered vectors in a multidimensional space. Dimensionality is determined by the number of variables (i.e., columns in a table) for each data point [203]. I performed hierarchical clustering to find similarities between the independent variables using their correlation coefficients. Lets' say the objects or data points i, j are two vectors. The distance between the pairs of vector i, j was calculated at a time for similarity or dissimilarity measurements. In hierarchical clustering, the similar vector pairs are groups and compared with the new vector or data point. The vector pairs that are most similar are grouped first. Their similarity is represented by tree-like branches or dendrograms. Shorter or closest dendrogram represent more similar vector pairs or groups of vector. The longer or furthermost dendrogram represents dissimilar pairs or groups [204]. Examining similarities/dissimilarities between the explanatory variables was useful to group them. Grouping the independent variables based on their similarity helped me understand their underlying relationships and infer their performance in the model. I utilized the Python module Seaborn [205] for hierarchical clustering.

Exploratory Factor Analysis (EFA): This technique was utilized to reduce the number of variables used to model the data. This is especially helpful for building traditional models like linear and quantile regression, where multicollinearity between variables can be a cause of concern. Multicollinearity interferes with the robustness of a model and generates unstable regression coefficients [206]. Multicollinearity may also increase variance in the coefficient estimates and may render the coefficient estimates to be very sensitive to minor changes in the model. Research on the built environment and travel patterns uses factor analysis to resolve multicollinearity issues [207, 208]. Factor analysis simplifies the observed variables to a small number of unobserved underlying factors called latent variables. The explanatory variables such as streetscape, land use mix, socio-economic conditions, and some others were found to be highly correlated as observed in the correlation matrices. However, in theory, they constitute the built environment and impact mental wellbeing. To investigate the impact of highly correlated built-environment variables and socio-economic indicators on mental wellbeing, I used exploratory factor analysis. The standardized factors generated from EFA are used to replace the raw explanatory variables.

4.6.2 Relating MWS to Built Environment

After computing the MWS and the built environment variables for a total of 2820 assessment grids, I used the dataset to train a set of machine learning models to examine the relationship of MWS to the built environment variables. In this setting, I trained two models namely, a Quantile Regression model (QRM) and a Random Forest Classifier (RFM). Twitter data is a crowd-sourced real-world dataset from social media, leveraging a naturalistic observation of people. However, the key issue that I faced was in tackling the non-normal distribution of MWS and the non-linear relationship of MWS exhibited with the built environment variables. The distribution of MWS was studied using box plot analysis (refer Figure 5.8).

Keeping that in mind, I selected these two modeling techniques, Quantile Regression Model and Random Forest Model. While Random Forest is a relatively advanced model, which is more suitable to explain non-linear relationships, I believe Quantile Regression adds more insights in explaining the relationships between MWS and various built environment variables; in spite of limitations to explain non-linear relationships. Further reasoning is provided below:

Quantile Regression Model (QRM): MWS is a continuous unimodal variable with a long right tail Figure 5.8. MWS exhibited non-linear relationships to the explanatory variables (refer Table 4.1 to see the list of explanatory variables). A QRM model is reasonable for this type of data, as it uses independent variables to predict a specified percentile or quartile of the dependent variable. In this technique, the regression parameters are considerably less sensitive to outliers. The set of models for different quartiles such as (25^{th}, 50^{th}, or 75^{th}) can explain how the independent (explanatory) variables vary at different levels of the dependent (response) variable [209]. Given that Linear Regression Model (LRM) has strict assumptions on normality, homoskedasticity, and auto-correlation, QRM has some advantages. For instance, QRM works better in describing skewed distributions, especially if the response variable is uni-modal or bi-modal (i.e., more than one peak). Median or quantiles gives better estimates of the population than the mean as used in a typical ordinary least squares (OLS) regression model. QRM does not make any assumption on the distribution of residuals and, as such is less sensitive to spatial auto-correlation [210]. Another advantage of using quantile regression is that it provides an appropriate approximation to the non-linear relationship exhibited in the real-world dataset (e.g., Twitter data). In this research, I used Pythons' statsmodel quantile regression package for analysis [211].

Random Forest Model (RFM): Random forest is an ensemble-based modeling technique capable of performing both regression and classification tasks using multiple decision trees. The models' algorithmic logic is analogous to asking a series of questions to narrow down the possible values to make a single prediction [212]. RFM involves training a set of decision trees on different sample data (sampling with replacement). The key idea is to combine outputs from multiple decision trees in determining a final output (often determined using voting from individual decision trees). I used *Scikit Learn's* machine learning package to construct the RFM model [213]. Furthermore, I used the permutation feature importance and the drop column feature importance technique (pre-built in the Scikit Learn package) to find the set of relevant features (built environment variables) in predicting MWS [214, 215]. Since RFM's feature importance does not explain the direction of the impact, I used

Pythons' 'treeinterpreter,' and the 'Shap' package [216, 217] to explain how each feature contributes (i.e., explains the effect of increase/decrease of any feature value) to the MWS score.

4.7 Content Analysis Framework

In this research, the assessment of mental wellbeing was performed through the abstraction of tweets. Analyzing the content of the tweets motivated me to contextualize stress and de-stress [195]. This further helped me to reason the mental wellbeing scores beyond the models' specifications. In addition, it also led to a deeper understanding of the model. Here are a few examples of stressed Tweets: *'deadlines are keeping me busy at work # stressed*; *'stuck in traffic for the last 20 mins'#stuckintraffic #stressed #anxious"*. Looking at the Tweets, we can identify the cause of stress as work pressure and traffic congestion, respectively. Similarly, the tweet *'Depression said no to getting out of bed. Anxiety said no to meeting a friend for dinner in a public place. #Stress'*, is an example of a feeling of stress or depression without an assigned reason. I adopted two approaches for content analysis: psycholinguistic characterization and topic modeling.

4.7.1 Psycholinguistic Characterization

Psycholinguistic characterization is used primarily as a method of validation, and how stress tweets linguistically differ from de-stress tweets. I performed psycholinguistic characterization of stress and de-stress tweets using a lexicon called Linguistic Inquiry and Word Count (LIWC) [218], a dictionary that comprises psycholinguistic categories. This lexicon is well-validated and widely used. The LIWC counts give the percentage of word counts that can be utilized as features for any modeling purposes. I used the following attributes: 1) *affective attributes* (categories: anger, anxiety, negative and positive affect, sadness), 2) *cognitive attributes* (categories: negation), 3) *interpersonal focus* (categories: first person singular, second person plural, third person plural), 4) *personal concerns* (categories: achievement, home, money, religion), and 5) *social concerns* (categories: family, friends, humans, social). These values were not only utilized as features in the tweets but also were used to find any significant linguistic distinction in the stress and de-stress tweets using Welch's t-test [192].

4.7.2 Topic Modeling:

I was motivated to study the tweet topics and their spatial association to identify what topics are causing stress and if any urban planning intervention is possible. I prepared the input data (tweets) to be fed into a LDA (latent dirichlet association) based topic model. The input to train the LDA model was a bag of words[24] of term frequencies[25], where each row is a document, and each words' frequency in that document was recorded. LDA models each document as a distribution of topics and each topic as a distribution of words. LDA computes the distribution of words/tokens per tweet and then aims to fit the word frequency/occurrence distributions into a set of a pre-specified number of topics. There are a few assumptions in LDA, including each tweet can be part of multiple topics, and the output of this model is non-deterministic. For my use case, I utilized the Scikit-learns LDA module [219]. However, LDA is considered to be more performant for documents with longer text. As tweet posts are relatively short text segments, I investigated the performance of non-negative matrix factorization (NMF) based topic models on this dataset. In NMF, I input a matrix where each row is a document. Each row contains the Tfidf[26,27] values for the words present in that document. NMF decomposes this input matrix into a set of two matrices with non-negative values (non-negativity is a constraint in the objective function of NMF's algorithm). These two matrices say W and H respectively contain the

[24]The bag-of-words model is commonly used in methods of document classification where the (frequency of) occurrence of each word is used as a feature for training a classifier [191].

[25]It is the number of times a word occurs in a document with respect to the total number of words in the document.

[26]Tfidf is term frequency-inverse document frequency

[27]The inverse document frequency is a measure of how much information the word provides, i.e., if it's common or rare across all documents [191].

pre-specified number of topics and the respective weights for each topic. In this dataset, NMF seemed to perform better and faster with respect to the quality of topics that were retrieved.

As effective as this was, both LDA and NMF were based on a bag of words approach, where the input document matrix stored either term frequencies or Tfidf values. In doing so, it failed to capture the semantic and syntactic meaning and order of words, which are usually expressed through word embeddings and sentence embeddings. Motivated to utilize the semantics of words in each tweet post, I utilized the topic model proposed by Moody et al. that combines a conventional LDA model with ‘word2vec’ [220] embeddings. This model is designed to capture interpretable document mixtures through a non-negative constraint. To instantiate this model, I was inspired by the implementation (using Pythons’ Pytorch module) provided here [221]. Using this model, I observed slightly better results than with the NMF model. The final analysis results reported in Chapter 5 were obtained using this model.

CHAPTER 5

DATA EXPLORATION AND MODELING

In Chapter 4, I covered the analytical framework of this research, which explained the methods applied to compute the mental wellbeing score (MWS, from the tweets), the built environment (BE) variables, and other measures as potential stressors. In this chapter, I will discuss the results. I begin with the Twitter stress classification results, discussing the performance evaluation of the trained deep learning model. Then, I discuss and compare the distribution of MWS scores in Atlanta and Boston. Next, I discuss the results from stress classification and other statistical analyses conducted to explain the relationship between MWS and built-environment. Finally, I discuss the results from the topic modeling algorithm.

5.1 Inferring Stress and De-stress in Social Media:

A binary multi-headed transformer-based mental wellbeing classifier was trained to classify tweets into ‘stress’ and ‘de-stress’ categories. I used a k-fold (k=10) cross-validation technique to evaluate the model and achieved a mean accuracy score of 94.74%. Refer Table 5.1 and Table 5.2 for precision and recall statistics for Atlanta and Boston respectively. After iteratively tuning the hyperparameters of this model (using a random search based

Table 5.1: Precision and recall metrics on test set of the stress classifier for the city of Atlanta.

# of class	class name	Training set		Test set	
		Precision	Recall	Precision	Recall
2	stress	83.17	91.72	81.51	78.53
	de-stress	91.23	95.22	91.08	81.12
Average		87.20	93.47	86.30	79.83

Table 5.2: Precision and recall metrics on test set of the stress classifier for the city of Boston.

# of class	class name	Training set		Test set	
		Precision	Recall	Precision	Recall
2	stress	87.82	89.12	84.81	72.33
	de-stress	98.78	94.10	94.78	78.52
Average		93.30	91.61	89.76	75.43

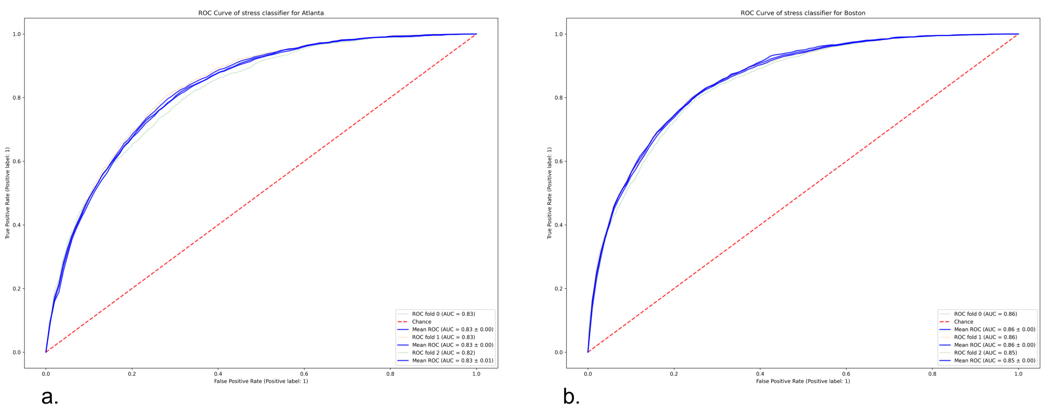

Figure 5.1: ROC curve performance chart of the trained stress classifier for the city: (a) Atlanta and (b) Boston.

approach as reported in subsection 4.3.4), the accuracy achieved was much better than the baseline accuracy of 62.34% which was tested on a chance model in this dataset. Table 5.3 and Table 5.4 reports the confusion matrix of the stress classifier for Atlanta and Boston respectively. Figure 5.2 shows the graph of accuracy and loss with the number of epochs the model was trained on the test set for the city of Atlanta. Figure 5.3 shows the same statistics for the city of Boston. Based on the reported performance indicators, I concluded that the mental wellbeing classifier was successfully able to classify tweets as stress and de-stress expressions. There were approximately $126,000$ classified tweets in the final dataset, $65,000$ for Atlanta and $61,000$ for Boston. About 8-9% of the tweets were in the stress category which is relatively small compared to the 20-22% of stress tweets in the randomly selected subset from non-geolocated tweets. To further probe the models' performance, I validated it by reviewing its area-under-curve (AUC) value, based on the receiver operating characteristic (ROC) curve. AUC for Atlanta was 0.843, while for Boston it was 0.832 as reported in Figure 5.1. AUC is considered a highly reliable metric as it compares the true positive rate with the false positive rate of the model.

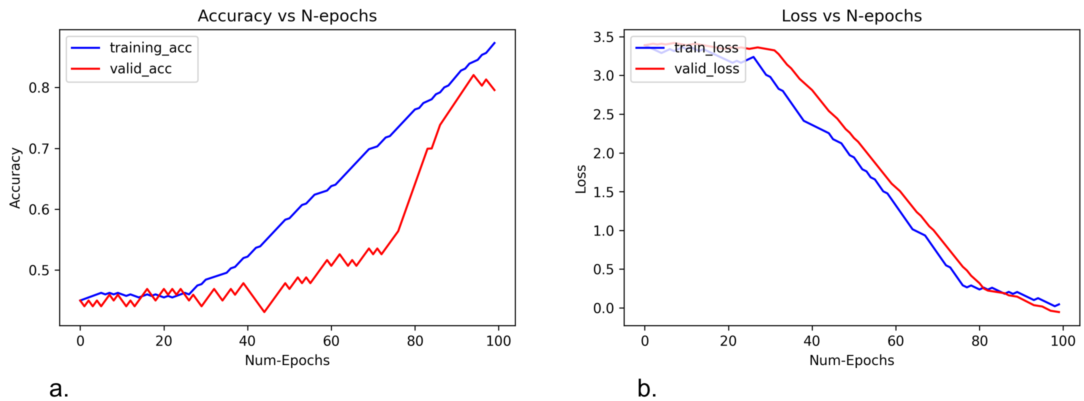

Figure 5.2: The figure shows the loss and accuracy statistics with respect to number of epochs the stress classifier was trained for the city Atlanta.

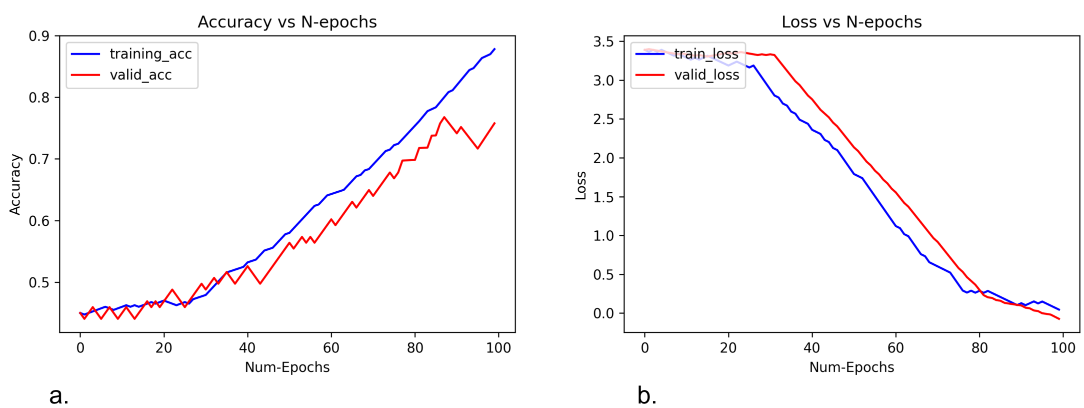

Figure 5.3: The figure shows the loss and accuracy statistics with respect to number of epochs the stress classifier was trained for the city Boston.

The top 20 features of the stress classifier is listed in Figure 5.4 and Figure 5.5. There were nouns and action based verbs in the list that were contextually related to stress and de-stress. For instance, frequently occurring words such as *traffic, working, boring, accident* were observed to be related to stress, and words such as *photo, drink, sunday, birthday,*

Table 5.3: Confusion matrix showing performance of the stress classifier on the test set for the city of Atlanta. The first row indicates - 874 stress tweets are predicted correctly (true-positive), and 18 de-stress tweets are predicted incorrectly as stress (false- positive). The second row indicates 543 stress tweets are predicted incorrectly as de-stress (false-negative), and 2565 de-stress tweets are predicted correctly (true-negative).

		Actual		
		Stress	De-stress	Total
Predicted	Stress	874	18	$892 = 874 + 18$
	De-stress	543	2565	$3108 = 543 + 2565$
	Total	$1417 = 874 + 543$	$2583 = 18 + 2565$	N

Table 5.4: Confusion matrix showing performance of the stress classifier on the test set for the city of Boston. 1338 tweets are predicted stress, while there are 1965 stress tweets. 2662 are predicted de-stress while 1935 are actual de-stress tweets.

		Actual		
		Stress	De-stress	Total
Predicted	Stress	1212	126	$1338 = 1212 + 126$
	De-stress	753	1909	$2662 = 753 + 1909$
	Total	$1965 = 1212 + 753$	$1935 = 126 + 1909$	N

family, eat, walk were related to de-stress. The words related to stress provided cues about urban activities that entail stress, such as traffic congestion, work pressure, fear from crime, etc. *'I thought today wasn't going to be stressful and then a rock hit my windshield and cracked it"* is an example of a classified stress tweet. Similarly, the words related to de-stress provided cues to relaxation activities like clicking photos, eating or drinking, and spending time with family. *No better way to end a saturday! falloutboy fob music musicmidtown musicmidtown2018 @ Piedmont Park* is an example of a classified de-stress tweet. Refer Table 5.5 and Table 5.6 for examples of classified de-stress and stress tweets in Atlanta and Boston respectively.

Psycholinguistic Characteriztion: To characterize the psycholinguistic characteristic cues in the stress and de-stress tweets, I extracted the normalized values of LIWC attribute categories [192]. This is used to assess whether the psycholinguistic difference between the stress and de-stress tweets was statistically different. I performed Weltch's t-test, followed by Bonferroni corrections. The results are presented in the Table 5.7. The results showed

Table 5.5: Example of classified de-stress and stress tweets in Atlanta.

Atlanta
De-stress Tweets
My wife has us at the megabus stop in Atl. We were about to patronize a local food truck when I caught an eye
Random walk turned into a 6 mile walk with no map ? or exact place in mind. @ Atlanta Georgia
Traveling alone doesn't have to be boring. Today I was off work and took myself & my tripod to the Walking Dead filming location
But on a serious note I need 10 of these flower walls ??? @ Shops at Buckhead
Out for a walk in my favorite park:). Aloha!#dog #dogsofinstagram #darwininatlanta #walk #piedmontpark #midtownatl
Stress Tweets
They say Trina accidentally bumped into her & the women called her a nigga bitch @ Atlanta Georgia
Idk when the last time I stayed at work past 3. I need to make up a little time and I'm out.
Last night was a GRIND! Got stuck in the infamous ATL traffic on the way to the station guest couldn't make it
Niggas still in the clubs fighting and shooting.. This shit ain't whats up.. Please stay safe for the holidays
Coming through Atlanta tonight during rush hour was a challenge. Stop and go bumper to bumper traffic

Table 5.6: Example of classified de-stress and stress tweets in Boston.

Boston
De-stress Tweets
Not sure how I haven't been to a parade before but there's only one thing I can say. Incredible. #Patriots...
Nothing like a walk along the water to refresh the soul. JFK Library & Harbor Islands in the background....
A teaser trailer for my most recent wedding at the MIT Chapel! @mit @ Cambridge Massachusetts
There's no better reward for a long day exploring a city quite like a cannoli! #Boston #bostonfoodies #foodies...
Spent the day goofing around the neighborhood at the @lawnond Pumpkin Palooza.
Stress Tweets
i got to the gym 2 hours ago a couple was arguing outside. just got out these niggas still arguing ?????
every time I have to wait more than5 minutes for the @mbta is a policy failure
Feeling a bit meh at work today. @ Allston Massachusetts
Story time folks...I was backing up this morning to let a car out of a tiny two way street and this dude starts beeping behind me
Very frustrating weekend first lap of the main I hit a bike in the sand and was stuck for an hour

under *affective attributes* stress tweets showed a higher occurrence of anger, anxiety, negative affect, and swear words. De-stress tweets showed a higher occurrence of positive affect words. This indicates that stress tweets communicate negative emotions, anger, anxiety, and sadness. The next measure used was *cognitive attributes*. Results showed a high occurrence of negation words in the stress tweets. Cognitive attributes or perceptual expressions in the language is associated with personal and first experience of real-world events and happenings [27, 222]. In this case, it represents that stress tweets communicate real-world perception of denial or negation. Under *Interpersonal focus attributes*, there is an increase in either first-person singular pronouns (I, me) or third-person singular and plural pronouns (he, she, they) in the stress tweets. An increase in the first person pronoun showed an increase in self-attention, and an increase in third person singular or plural pronouns showed a connection to the social realm. Results indicated the stress tweets either share personal experiences about their day-to-day emotion or express frustration about another person, organization, or community. The occurrence of first-person plural pronouns (we, us) was significantly higher in the de-stress tweets. Second person pronouns showed more participatory feeling and bonding [27, 222].

The last measure is *personal and social concern attributes*. Achievement, home, money, religion, family, and friend are some of the specific categories under personal and social concern. Results showed that words related to achievements occur significantly more in the de-stress tweets. This category included pride, fulfillment, confidence, self-esteem. Lower use of achievement words showed a decline in expressions related to career engagements and self-growth in the stress tweets. Words related to home, family, and friends were more frequent in the de-stress tweets. This showed the people are either showing bonding and engagement with their nearest social connections [27]. Religious words occurred more frequently in the de-stress tweets in Atlanta. My conjecture is that people show engagement and participation in religious ceremonies. Words related to money were discussed more frequently in the stress tweets in Boston. This showed that people express stress about their finances. In both Atlanta and Boston, the results showed similar psycholinguistic characteristics of stress and de-stress tweets. Minor differences were observed under personal and social concerns. In Atlanta, religious words were more frequent in the de-stress tweets, and in Boston, money-related words were more common in the stress tweets.

Expert validation: I validated the classified tweet labels from an expert who is experienced in analyzing social media data and has experience in psycholinguistic assessment. A random sample of 600 stress tweets and 600 de-stress tweets were utilized for validation. A coding menu that I developed for the purpose of hand labeling was shared with

Table 5.7: Welch's t-test comparing psycolinguistic attributes of stress and de-stress tweets. Here *** $p < 0.000$,**$p < 0.001$, * $p < 0.05$

Atlanta					**Boston**			
Category	Stress	De-stress	t-stat	p	Stress	De-stress	t-stat	p
Affective Attributes								
Anger	1.022	0.102	65.259	***	1.006	0.116	53.748	***
Anxiety	0.255	0.040	27.701	***	0.391	0.045	34.755	***
Positive Affect	0.558	3.002	-44.267	***	0.702	3.437	-41.683	***
Negative Affect	2.718	0.320	104.898	***	3.271	0.371	105.878	***
Sadness	0.558	0.069	46.273	***	0.718	0.091	47.630	***
Swear	0.742	0.087	47.681	***	0.629	0.076	36.684	***
Cognitive Attribute								
Negate	0.957	0.317	33.129	***	1.515	0.495	38.035	***
Interpersonal								
Ist P.Singular	2.165	1.747	8.824	***	2.210	1.801	7.627	***
1st P.Plural	0.513	0.592	-3.022	**	0.590	0.708	-3.787	***
3rd P.Singular	0.342	0.235	6.134	***	0.276	0.225	2.650	**
3rd P.Plural	0.192	0.122	5.968	***	0.163	0.120	3.428	**
Personal and Social Concerns								
Achievement	0.706	0.905	-6.476	***	0.819	1.046	-6.088	***
Home	0.282	0.310	-1.472	-	0.363	0.485	-4.721	***
Money	0.416	0.379	1.700	-	0.586	0.443	5.513	***
Religious	0.204	0.269	-3.262	**	0.235	0.257	-0.993	-
Family	0.208	0.273	-3.658	***	0.302	0.354	-2.344	*
Friend	0.158	0.295	-8.197	***	0.168	0.313	-7.478	***

the expert, and a few additional codes were added to the existing code set. We reached a high agreement in this task (Fleiss' = 0.86) and obtained an accuracy of 83% for the stress classification.

5.2 Understanding Patterns in the Data

5.2.1 Distribution of MWS

Next, the classified geolocated tweets were spatially joined with the assessment grids to build the final dataset with built-environment features and mental wellbeing scores. Refer section 4.4 for the details on spatial join and section 4.5 for MWS calculation. To understand the distribution of mental wellbeing scores, I visualized the data through box-plots (refer Figure 5.8). As can be seen, the MWS score is highly skewed towards the left, meaning MWS has very high outlier values.

Interestingly, I observed that the distribution was very similar for both Atlanta and

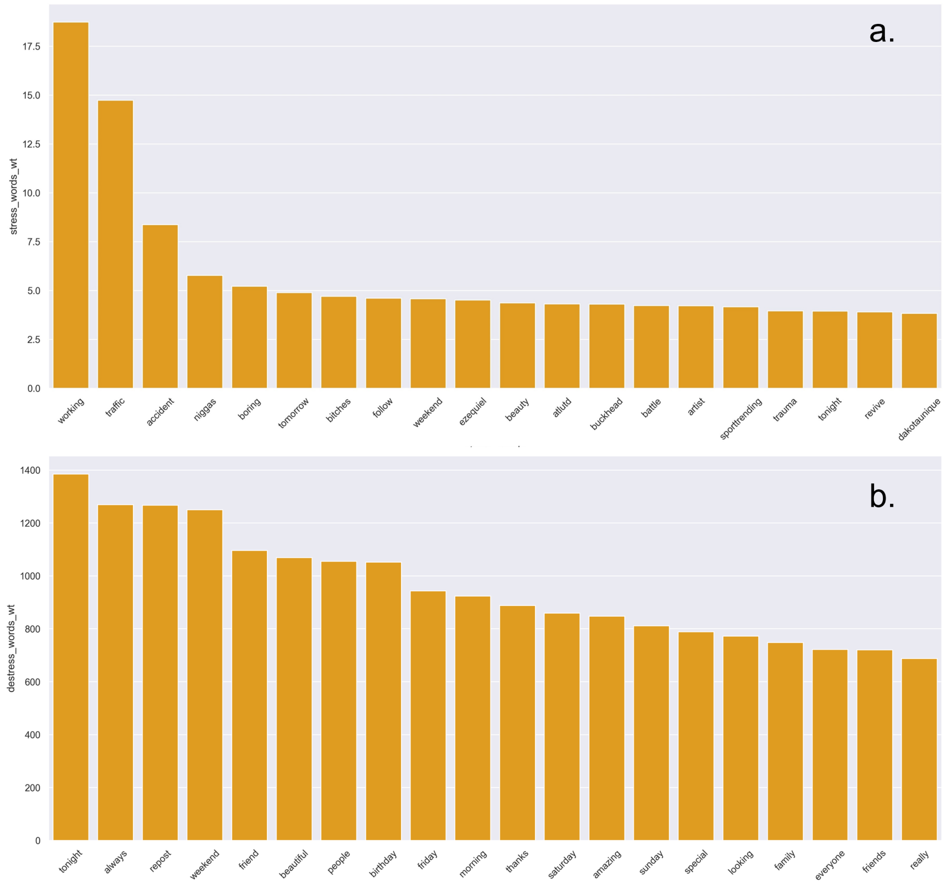

Figure 5.4: Most important features from the Tweet text in Atlanta for the label: (a) Stress and (b) De-stress.

Boston. The median value of the MWS score in both cities was 1.3; the 25^{th} quartile is 0.4. The 75^{th} quartile of Boston was 6.5, and for Atlanta, this statistic was 5.7. The outlier values are as high as 217 in Atlanta, and in Boston, outlier values are above 800.

5.2.2 Analysing the Explanatory (independent) Variables

Hierarchical clustering (HC): The tree-based representation or dendrograms shows the similarity between the explanatory variables based on their Pearson correlation coefficients. The shortest dendrograms represent more similar data points, while the longest ones encode more distant and dissimilar data points. For instance, Figure 5.6-b on the right shows hierarchical clustering of the variables computed within 1 mile of the assessment grids (a per-

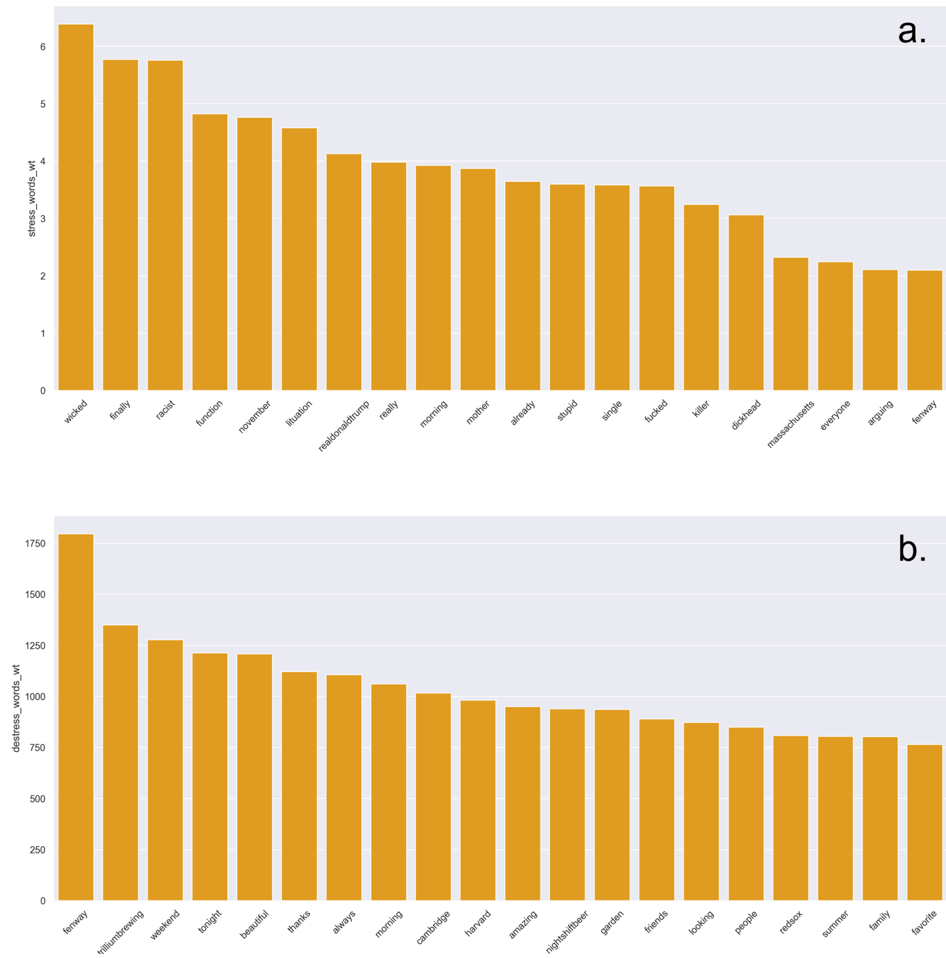

Figure 5.5: Most important features from the Tweet text in Boston for the label: (a) Stress and (b) De-stress.

son's tweet location) in Boston. Looking at the HC dendrograms, we can conclude that the proportion of tweets in the normal pressure conditions (*normalPressure*), no precipitation days (*noPrcip*), comfortable temperature conditions (*comfortable*), and on weekends(*sum weekend*) have the shortest dendrogram. This implies that the variables are most similar and thus are clustered in the hierarchy first. These variables also have high positive Pearson correlation coefficients among themselves (above 0.75 at p< 0.05).

Access to *social* and *recreational* facilities were second in the hierarchy. I also observed that these facilities were clustered with the point of interest (POI) diversity (*shannon2*) vari-

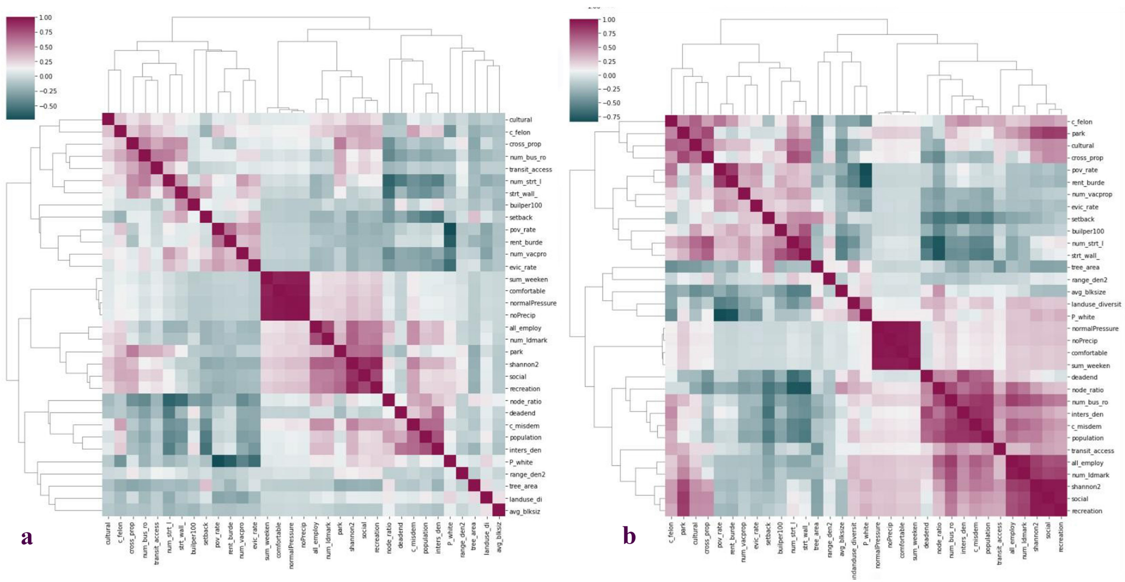

Figure 5.6: The figure shows the correlation matrix and hierarchical clusters of explanatory variable based on their correlation matrix. a) shows variables for Boston within 1/4 mile of tweet locations and, b) shows variables within 1 mile of tweet locations. Refer Appendix B for larger images.

able. The hierarchical structure and the Pearson correlation coefficient showed a person's tweet location with greater access to social and recreational facilities had a higher POI diversity. This cluster appeared to be similar to the cluster with access to landmarks (*num ldmark*) and employment density (*all employ*). These variables depicted a higher positive correlation coefficient (above 0.75 at p$<$ 0.05) among themselves and a moderate negative correlation to the poverty rate, rent burden, setback, eviction rate, and average block size variable. This finding is intuitive and is also supported in the literature [16]. An interesting observation was that the access to social, recreational facilities and a greater POI diversity was found to be negatively correlated to the street tree cover (*tree area*) variable. In the hierarchical cluster for Boston, within 1 mile of a person's tweet location, roughly three cluster groups were seen. The first cluster comprised of socio-economic status (SES) variables and related built environment indicators. Poverty rate, rent burden, eviction rate, and vacant properties showed a moderate positive correlation (ranging $0.25 - 0.5$ with p$<$ 0.05) effect and were placed closer in the cluster with a shorter dendrogram. Similarly, *percent*

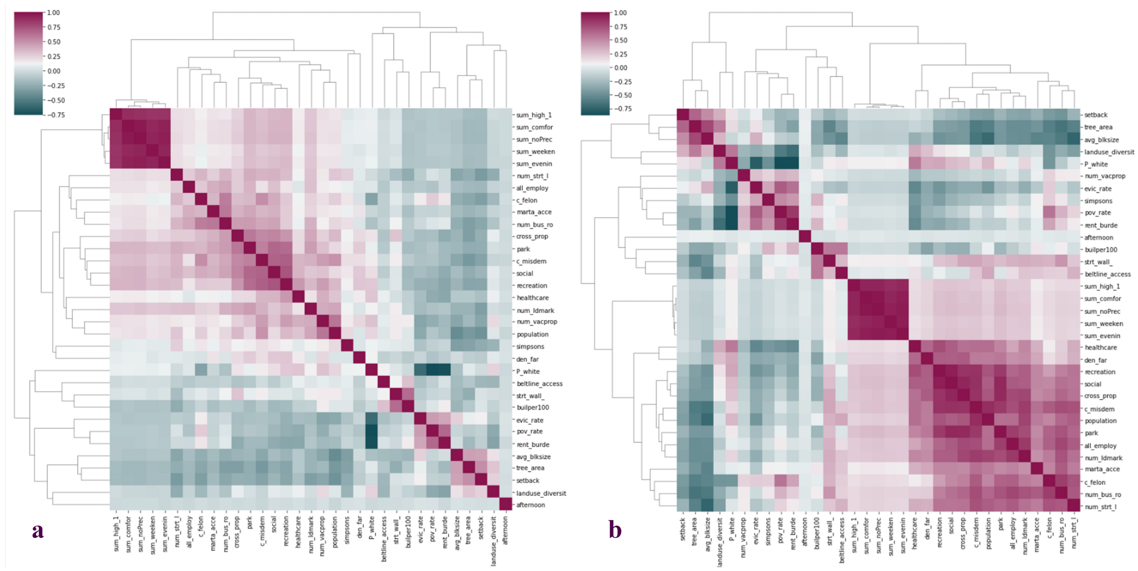

Figure 5.7: The figure shows the correlation matrix and hierarchical clusters of explanatory variable based on their correlation matrix. a) shows variables for Atlanta within 1/4 mile of tweet locations and, b) shows variables within 1 mile of tweet locations. Refer Appendix B for larger images.

white, *average block size*, *street tree cover* variables of the street segments were related and were placed closer together in the cluster. The second cluster consisted of variables pertaining to comfortable weather conditions and time availability on the weekend. The *weather condition* and *time availability* variables showed high positive correlation among themselves, and had shortest dendrograms. The third cluster was explained access to point of interests, landmarks, access to public transportation (*transit access*, *num bus ro*), and network connectivity (*inters den*, *node ratio*). The crime rate variable showed low to moderate positive correlation (ranging $0.25 - 0.5$ and p< 0.05) to variables measuring access to the point of interests and employment density variable. The HC using variables computed within 1/4 miles of tweet locations in Boston showed a slightly different arrangement of cluster groups. One interesting difference was seen in the position of access to parks in the HC. For instance, access to parks, access to landmarks, employment density, and POI diversity were placed close to each other in this cluster, which is not so in the HC generated with variables computed within a mile of tweet location. As can be seen in Figure 5.6,

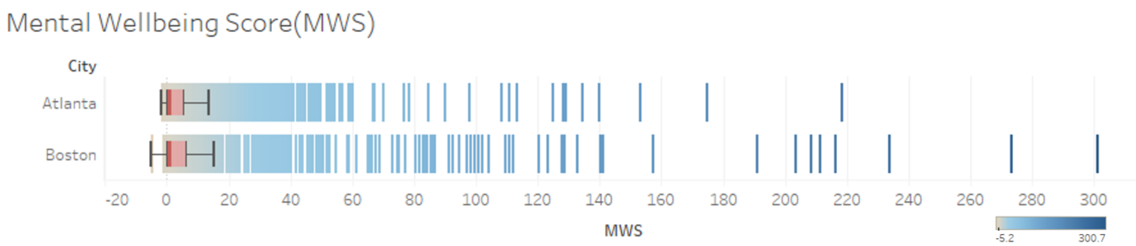

Figure 5.8: The figure shows the boxplot of mental wellbeing score (MWS) for Atlanta and Boston generated from tweets between May 2018, and March 2010.

the Pearson correlation coefficients of these variables (*access to parks*, *access to landmarks*, *employment density* and *poi diversity*) were high positive (above 0.5 at p$<$ 0.05). From this result, we can say that Tweet locations with greater access to parks within 1/4 miles were co-located with other facilities, landmarks, and employment centers in Boston. The Pearson correlation coefficients between the variables computed within 1/4 miles of tweet locations were lower than those calculated within a 1 mile radius. I found a similar relationship pattern among variables from the HC clusters in Atlanta. There were only subtle differences in the spatial arrangements and distances between the clusters, as seen in Figure 5.7. Like Boston, in Atlanta, the correlation coefficients between the variables computed within 1/4 miles of tweet locations showed lower Pearson correlation coefficients. For detailed images of Hierarchical clusters, refer Appendix B .

HC using Pearson correlation coefficients of the explanatory variable was a key step forward in exploring the relationship between explanatory variables. I identified three groups namely *Socio-economic status and Built Environment* group , *Access to facilities* group and *weather and time opportunity* group. However, the clusters were still fuzzy, with unexplained differences in cluster distances. To obtain consistent and discrete groups of explanatory variables, I conducted Factor Analysis, which I have explained next.

Exploratory Factor Analysis (EFA): The HC charts as seen in Figure 5.6 show that the independent variables have moderate to high correlation with each other, a condition that is known as multicollinearity. Multicollinearity results in wider confidence intervals and

Table 5.8: Descriptive statistics of the explanatory variables used. The table only contains values computed within 1 mile of the tweet locations (assessment grids) for both cities together.

Variables	mean	variance	std	min	max
Landuse diversity	0.52	0.04	0.19	0	0.88
POI Diversity	0.22	0.03	0.16	0.01	0.85
Access to parks and trails	16.4	436.65	20.9	0	98
Access to social facilities	47.2	2172.7	46.61	0	243
Access to recreational facilities	75.92	5724.3	75.66	0	404
Access to healthcare facilities	28.4	890.79	29.85	0	162
Access to cultural facilities	9.33	49.66	7.05	0	35
Access to educational facilities	16.91	188.15	13.72	0	69
Access to historic districts and landmarks	17.17	1161.8	34.08	0	189
Street cross-sectional proportion	0.98	0.67	0.82	0.12	4.11
Street intersection density	162.16	6279.1	79.24	11.78	367.65
Number of bus routes	15.49	152.83	12.36	0	55
Transit access	0.69	0.21	0.46	0	1
Street tree cover	2849.4	2.70×10^{6}	1642	526.1	13826.7
Street lights	16.69	176.21	13.27	0.3	44.41
Building density	15.74	967.85	31.11	0.12	320.92
Building setback	6.36	26.06	5.1	0	28.31
Buildings per 100 m of street segments	3.11	0.87	0.93	0.37	5.81
Street wall continuity	0.03	0	0.02	0.01	0.08
Deadend	74.96	1189.2	34.49	10	187
Average blocksize	1.56×10^{6}	3.96×10^{12}	1.99×10^{6}	2.04×10^{5}	3.15×10^{7}
Employment Density	31798	4.20×10^{9}	64844	0	352333
Population Density	30316	4.74×10^{8}	21764	680.4	97029.1
% White	49.36	673.92	25.96	0.53	96.9
% Black	35.9	916.89	30.28	0.67	97.91
Poverty	14.71	105.89	10.29	0	44.47
Rent burden	28.85	58.46	7.65	1.21	46.92
Eviction rate	3.67	18.84	4.34	0	91.29
Misdemeanor (rule violations, and thefts)	4601.3	2.90×10^{7}	5389	7	21852
Felony (homicides and robbery)	157.62	15343	123.9	0	507
Vacant properties along street	2.78	21.49	4.64	0	45
Noise (avergae traffic count)	4.22×10^{6}	2.75×10^{13}	5.24×10^{6}	0	2.45×10^{7}
Comfortable weather (%tweets)	0.44	0.11	0.34	0	1
Weekend days (%tweets)	0.32	0.1	0.32	0	1
Evening time (%tweets)	0.37	0.1	0.32	0	1
Sample size (n)	2820				

less reliable probabilities of rejecting null hypotheses in the regression models. To ensure minimum inter-correlation and minimize variance inflation factors (VIF) [223] between variables, I used factor analysis. I generated factors explaining the relationship between mental wellbeing score (MWS) and the built environment. These factors were based on my existing theoretical assumptions that mental wellbeing is dependent on access to facilities, urbanness, perception of crime, and socio-economic markers of the built environment. Some of these factors had overlapping features. For instance, urbanness included access to diverse facilities, high employment density. Assigning meaningful factor nomenclature based on their factor loading was critical in describing the relationship between mental wellbeing and the built environment. Factors were assigned for both cities (Atlanta and Boston) within 1/4th miles and 1 mile distances of tweet locations. Assignment of individual factors revealed unobserved features and relationships between variables. I recovered 7 to 8 factors with eigenvalues above 1 for each assessment category for both cities within 1/4 miles and 1 miles areas around tweet locations. After varimax rotation of the factors, and retaining values above ± 0.3, I recovered three explainable factors. Table 5.9 and Table 5.11 show the factor loading's, the variance explained by each of the factors, and their cumulative variance. There were few overlaps in loading on a few factors, and the cumulative variance for all the different factors explained was not high. However, concurrent loading's and low cumulative variance have been used in exploratory factor analysis (EFA) when they are explainable by underlying theory [224].[1] Moreover, I wanted to explore the best representative study area between the choices of 1/4 miles, and 1 mile distances from a person's tweet location. Below I have explained all the factors and their underlying assumptions.

I present the factors generated for 1/4 miles of a person's tweet location in Atlanta and

[1]It is noteworthy to mention that I had explored other variables such as weather conditions, time of the day for tweets, and time of the week in the exploratory factor analysis. This was intentional as I wanted to see if there were any underlying patterns in these factors. These variables were loaded in a pair of two or three based on their category. For instance, hot, cold, and comfortable temperature was loaded together. While hot and cold showed positive loading, comfortable showed negative loading. I discarded these factors.

Boston in Table 5.9. I retained two explainable factors- *urbanness*, and *fear of crime and poverty*. In *urbanness*, variables describing access to different facilities, building density are loaded, which followed my constructs for urbanness. A revealing observation is that, *misdemeanor* is loaded in this factor. This crime category captures less intense crimes such as rule violations, thefts, etc. The fact that it appears alongside other variables describing urbanness is somewhat aligned with the theory that crime is a major challenge for urban mental health [155]. The second factor was fear of crime and poverty, variables such as % black, poverty rate, eviction rate, rent burden are loaded along with felony. In this category of crime, armed robbery, homicides, and other intense crimes are included. Land use diversity has a negative loading in this factor which shows a lack of access to facilities. In Atlanta, urbanness explained 15.95% variance, and fear of crime and poverty explained 13.8%. The two factors together explained 29% variance (cumulative variance). For Boston, I found similar factors; however, the factor loadings were slightly different. Cultural facilities, employment density, block sizes were some additional variables in the urbanness factor. The factor signs were found to be intuitive. For example, average block sizes were smaller in urban areas, and it was loaded with a negative sign. Employment density was expected to be high in urban areas, and it was loaded with a positive sign. Interestingly, the proxy to the noise variable was loaded as positive in this factor. Like crime, *noise* was identified as a key urban stressor [225]. Given these variations between the two cities, I established that the *urbanness* factor included the underlying characteristics of urban areas, which are one of the key variables of interest in this research.

Table 5.10, shows the factors derived for 1 mile of a person's tweet location in both Atlanta and Boston. I named them as *urbanness factor*, *socio-economic deprivation factor*, and *street enclosure and connectivity factor*. Once again, here urbanness factor had positive loading for all variables explaining accessibility to urban facilities. In addition, it included additional urban density variables. Access to urban facilities included *Shannon diversity index, access to parks and trails, landmark destinations, social, recreational, cul-*

tural, religious, educational and healthcare facilities, and access to public transportation. The density variables included *density FAR, cross-sectional proportion of streets, population density, and employment density*. The *urbanness factor* showed positive loading for misdemeanor and noise. It also showed negative loading for the *street tree cover* variable. This reflected that dense urban areas lacked adequate street tree cover [226] in comparison to less urban areas. The second factor observed was fear of crime and poverty. This factor was not different from the one identified in the EFA for the study area within 1/4 miles of a persons' tweet. The third factor was named *urban rent burden*. Although this factor contributed less in explaining the variance, it showed interesting loadings. The built environment factors loaded in this factor explained characteristics such as compact urban street network (intersection density), transit access, less setback, street wall continuity; yet it reflected positive loading on two factors, namely, rent burden and eviction rate. As such, I named it urban rent burden. In Atlanta, the urbanness factor explained 34% variance, fear of crime and poverty explained 18.15% variance, and urban rent burden explained 9.26% variance. The total variance explained by the three factors was found to be 61.42%. In Boston, the urbanness, fear of crime and poverty factor, and urban rent burden factors explained 39.9%. 19.18% and 15.95% variance, respectively. The total variance explained by these factors was 64.54%. This highlighted the characteristics of the study area within 1 mile of a person's tweet location. With such observations, it can be concluded that this result explained more variance and captured more built environment qualities than an area within 1/4 miles of a person's tweet location.

I conducted a third set of factor analyses combining both Atlanta and Boston's data together. The results are presented in Table 5.11. Three factors that were extracted include *urbanness*, *street safety perception* and *socio-economic deprivation*. The urbanness factor loaded access to facilities, transit access, population, and employment density. Noise and misdemeanor variables loaded positive in this factor. The street safety factor had common loading from the urbanness factor, e.g., access to the park and cultural facilities. Alongside, the variables describing safety perception, such as high crossectional proportion, street-wall continuity, building per 100 meters of the segment, have high positive loading on this factor. Urbanness factor for areas within 1 mile of a person's tweet location, the misdemeanor showed a higher negative loading. The third factor was socio-economic deprivation. It reflected a high positive loading for percentage black population, poverty rate, rent burden, eviction rate, and felony. It had negative loading on land-use diversity. For areas within 1/4 miles of a person's tweet location, the urbanness factor explained 15% of the variance, the street safety perception explained 9.81% of the variance, and socio-economic deprivation explained 8.15% of the variance. The total variance explained by the three factors was 43.54%. For areas within 1 mile of a person's tweet location, the urbanness factor explained 17.91% of variance, the street safety perception explained 13.71% of the variance, and socio-economic deprivation explained 11.91% of the variance. The total variance explained by the three factors was observed to be 43.54%.

5.2.3 Urban Mental Wellbeing Model

Next, I utilised the factors extracted from the independent variables to build the 'urban mental-wellbeing model', a novel model investigating mental wellbeing at an urban scale using social media data. The model is designed to add insight in understanding mental wellbeing and its relationship to built environment. Given the complex nature of this relationship, I used two kinds of models to derive any conclusion. As explained in Chapter 4 (specifically in subsection 4.6.2), I trained multiple quantile regression models and random forest classification models to determine the nature of this relationship.

Quantile Regression Models (QRM): The quantile regression model empowered me to analyze full conditional distributional properties of the mental wellbeing score (MWS). The basic idea of quantile regression can be expressed through the following equation [227]:

$$Q^{(\tau)}(y_i|x_i) = \beta_0^{(\tau)} + \beta_1^{(\tau)} x_i + Q^{(\tau)}(\varepsilon_i) \tag{5.1}$$

Table 5.9: Factors generated with variables computed within 1/4 miles of a person's tweet location in Atlanta and Boston.

	Atlanta 1/4 miles		Boston 1/4 miles	
Variables	Urbanness	Fear of Crime and Poverty	Urbanness	Fear of Crime and Poverty
Land use diversity		-0.316		
Point of interest (POI) diversity	0.902		0.964	
Access to parks and trails	0.728		0.788	
Access to social facilities	0.836		0.905	
Access to educational facilities	0.416		0.755	
Access to recreational facilities	0.843		0.908	
Access to historic districts and landmarks	0.314		0.861	
Access to cultural facilities			0.569	
Access to healthcare facilities	0.541		0.805	
Street cross-sectional proportion	0.627		0.678	
Street intersection density			0.384	
Transit access	0.396		0.520	
Street tree cover				
Street lights			0.404	
Building density	0.411			
Building setback				
Buildings per 100 m of street segments				
Street wall continuity				
Deadends				
Average block size				
Employment density			-0.880	
Population			0.342	
% White		-0.940		-0.904
% Black		0.936		0.830
Poverty		0.826		0.780
Rent burden		0.784		0.540
Eviction rate		0.763		0.488
Vacant properties along street		0.518		0.409
Misdemeanor (rule violations, and thefts)	0.772		0.688	0.684
Felony (homicides and robbery)		0.416		0.616
Noise (annual average daily traffic)			0.587	
Variance	15.95%	13.83%	26.00%	11.80%
Cumulative variance	29.77%		38.61%	

Table 5.10: Factors generated with variables computed within 1mile of a person's tweet location in Atlanta and Boston.

	Atlanta 1mile			Boston 1mile		
Variables	Urbanness	Fear of Crime and Poverty	Urban Rent Burden	Urbanness	Fear of Crime and Poverty	Urban Rent Burden
Land use diversity		-0.724	-0.445			
Point of interest (POI) diversity	0.906			0.981		
Access to parks and trails	0.942			0.909		
Access to social facilities	0.900			0.967		
Access to educational facilities	0.879			0.912		
Access to recreational facilities	0.854			0.970		
Access to historic districts and landmarks	0.847			0.956		
Access to cultural facilities	0.721			0.832		
Access to healthcare facilities	0.576	-0.611		0.918		
Street cross-sectional proportion	0.920			0.931		
Street intersection density	0.754		0.521	0.430	0.819	
Transit access	0.643		0.313	0.311		0.608
Street tree cover	-0.624		-0.626	-0.317		-0.675
Street lights	0.811		0.427	0.723		
Building density	0.589			0.874		
Building setback	-0.379		-0.449			-0.762
Buildings per 100 m of street segments			0.607	0.303		0.419
Street wall continuity	0.356		0.765	0.731		0.483
Deadends			0.688		0.776	
Average block size	-0.559		-0.724	-0.350		
Employment density	0.868			0.974		
Population	0.816		0.442	0.531	0.651	
% White		-0.948			-0.771	
% Black		0.935			0.705	
Poverty		0.922			0.866	
Rent burden		0.895	0.305		0.807	0.315
Eviction rate		0.749	0.305			0.379
Vacant properties along street		0.428			0.599	
Misdemeanor (rule violations, and thefts)	0.947			0.648	0.684	
Felony (homicides and robbery)		0.554			0.759	
Noise	0.742			0.645		
Variance	34.01%	18.15%	9.26%	39.90%	19.18%	15.95%
Cumulative variance	61.42%			64.54%		

Table 5.11: Factors generated with variables computed within 1/4 miles, and 1 mile of a person's tweet location in both Atlanta and Boston.

	1/4mile			1 mile		
Variables	Urbanness	Street Safety Perception	Socio-Economic Deprivation	Urbanness	Street Safety Perception	Socio-Economic Deprivation
Land use diversity						
Point of interest (POI) diversity	0.905			0.951		
Access to parks and trails	0.589	0.592		0.785	0.410	
Access to social facilities	0.788			0.939		
Access to educational facilities	0.645			0.888		
Access to recreational facilities	0.799			0.934		
Access to historic districts and landmarks	0.792			0.859		
Access to cultural facilities	0.417	0.507		0.597	0.422	
Access to healthcare facilities	0.697			0.835		-0.307
Street cross-sectional proportion	0.311	0.756			0.760	
Street intersection density	0.357					
Transit access	0.712			0.315		
Street tree cover						
Street lights		0.872			0.873	
Building density					0.575	
Building setback						
Buildings per 100 m of street segments					0.446	
Street wall continuity		0.879			0.842	
Deadends						
Average block size						
Employment density	0.824			0.816		
Population	0.360					
% White			-0.906			-0.916
% Black			0.842			0.902
Poverty			0.812			0.923
Rent burden			0.583			0.869
Eviction rate			0.502			0.511
Vacant properties along street						0.470
Misdemeanor (rule violations, and thefts)	0.655			0.526	-0.737	
Felony (homicides and robbery)			0.555			0.611
Noise	0.573			0.418		
Variance	15.19%	9.81%	8.15%	17.91%	13.71%	11.91%
Cumulative variance	33.33%			43.54%		

where $Q^{(\tau)}(y_i|x_i)$ is the τth conditional *quantile* determined by the quantile specific value of $\beta_0^{(\tau)}$ and $\beta_1^{(\tau)}x_i$ and a specific value of x_i (i.e i^th value of explanatory variable x). The error terms in the quantile regression are constant and depend on the quantities and not on $i = 1, ...n$. In other words, they are independent and identically distributed (i.i.d). In my research y_i corresponded to the mental wellbeing score for i^{th} assessment grid (i.e., tweet location) and the x_i corresponded to all standardized factor values generated from the factor analysis. As MWS showed a very skewed distribution, I used a log transformed value of y_i. The transformed distribution is transformed quantiles of the original distribution. Additionally, log transforming the response variable improved the model performance as it yielded a better fit of the model to the data than the raw score [228, 227].

$$log(Q^{(\tau)}(y_i|x_i)) = \beta_0^{(\tau)} + \beta_1^{(\tau)}x_i + Q^{(\tau)}(\varepsilon_i) \tag{5.2}$$

Table 5.12 shows quantile estimates for mental wellbeing score on urbanness and fear of crime within 1/4 miles of a person's tweet location. The model can be conceptually represented as:

$$MWS^{\tau} => \beta_0^{(\tau)} + \beta_1 Comfortable\, weather\, conditions^{(\tau)} + \beta_2 Urbannes^{(\tau)} + \\ \beta_3 Fear\, of\, crime\, and\, poverty^{(\tau)}$$

where τ is the quantile value. I have used standard quartile values which is 0.25 quantile, 0.50 quantile, and 0.75 quantile. Coefficient values $(coeff)_t$ showed percentage change in MWS at the τ^{th} quantile. That is, for every one unit change in the factors, holding all other factors, and time window for comfortable weather condition constant, $(coeff)_t$ showed the percentage change in the MWS at the τ^{th} quantile.[2] The percentage change is calculated as $(exp(coeff)_{raw} - 1) * 100$ [227]. The pseudo R^2 value is a measure of 'goodness of

[2]Variables weekend days (% tweets), and evening time (%tweets) are considered to estimate social engagement measures are not used in the final regression model as they are highly correlated to the comfortable weather conditions (% tweets). Refer subsection 5.2.2 for the results of hierarchical clustering, and correlation matrix of explanatory variables.

fit', and unlike a linear regression model, it is only to be used to understand the model fit between different conditional-quantile models. The models I presented here show that the 'goodness of fit' is highest in all models at the '0.75' quantile. For Atlanta, results indicated that the urbanness factor within 1/4 miles of the tweet location was significant at 0.1% for all the quartiles of MWS. It means that MWS increases by 8.7% at 0.25 quantile, 14.69% at the 0.5 quantile, and then 8.7% at 0.75 quantile for one unit increase in the urbanness factor score. This can be reinstated as: the increase in the median value of MWS is 14.96% for a unit increase in the urbanness factor score, which is more than the change at the lower and the upper quartiles. The effect of urbanness was found to be strongest at the median MWS score. The fear of crime and poverty factor within 1/4 miles of the tweet location was significant at all quartiles of MWS. At 0.25 quantile of MWS, fear of crime and poverty was significant at 1% level, and for all other quartiles, it was significant at 0.1% level. MWS decreases by 2.8% at 0.25 quantile, 7.28% at the 0.5 quantiles, and then again 8.3%, showing the effect of fear of crime and poverty increases for a higher value of mental wellbeing score. On the other hand, for Boston, both factors were observed to be significant. The results showed that the increase in the MWS score with one unit increase in the urbanness factor was almost similar at all quartiles. The increase in MWS is highest at the median. A unit increase in the urbanness factor increases MWS by 18.63% keeping all other factors and variables constant. MWS decreases by 1.69% at the lowest quartile and decreases by 4.12% at the highest quartile for a unit increase in the fear of crime and poverty factor. Even in Boston, the effect size of fear of crime and poverty on MWS increases in the upper quartiles.

Table 5.13 shows quantile estimates for mental wellbeing score on urbanness, fear of crime and urban rent burden within 1 miles of a person's tweet location. The model can be

Table 5.12: Quantile Regression Models explaining mental wellbeing score (MWS), within 1/4 mile of a person's tweet location, for Atlanta and Boston.[a] [b]

	$\tau(0.25)$				$\tau(0.50)$				$\tau(0.75)$			
	$(coeff)_t$	$(coeff)_{raw}$	Std Er	P -value	$(coeff)_t$	$(coeff)_{raw}$	Std Er	P -value	$(coeff)_t$	$(coeff)_{raw}$	Std Er	P -value
Atlanta												
Intercept	875.129	2.2774	0.010	***	1066.508	2.4566	0.012	***	1379.368	2.6942	0.016	***
Comfortable weather	4.102	0.0402	0.001	***	5.422	0.0528	0.001	***	6.599	0.064	0.001	***
Factors												
(1) Urbanness	8.752	0.0839	0.008	***	14.694	0.1371	0.012	***	8.904	0.0853	0.016	***
(2) Fear of crime and poverty	-2.819	-0.0286	0.009	**	-7.281	-0.0756	0.011	***	-8.341	-0.0871	0.015	***
Pseudo R- squared	0.281				0.4518				0.5654			
No of assessment grids	1659				1659				1659			
Boston												
Intercept	106.308	0.7242	0.004	***	120.097	0.7889	0.005	***	747.144	2.1367	0.011	***
Comfortable weather	0.100	0.001	0.001	***	1.298	0.0129	0.005	***	4.050	0.0397	0.001	***
Factors												
(1) Urbanness	14.225	0.133	0.004	***	18.863	0.1728	0.005	***	14.970	0.1395	0.011	***
(2) Fear of crime and poverty	-1.695	-0.0171	0.004	***	-2.800	-0.0284	0.005	***	-4.123	-0.0421	0.011	***
Pseudo R- squared	0.1127				0.3222				0.5974			
No of assessment grids	1161				1161				1161			

[a] $(coeff)_t$ is transformed coefficient and gives $\%$ values, $(coeff)_{raw}$ is raw coefficient for $log(MWS)$, 'Std Er' is standard error.
[b] P-value $*** < 0.000$; $** < 0.001$;$* < 0.05$

conceptually represented as:

$$MWS^{\tau} => \beta_0^{(\tau)} + \beta_1 Comfortable\, weather\, conditions^{(\tau)} + \beta_2 Urbannes^{(\tau)} + \\ \beta_3 Fear\, of\, crime\, and\, poverty^{(\tau)} + \beta_4 Urban\, rent\, burden^{(\tau)}$$

By increasing the analysis area around a persons' tweet location, I retrieved one additional factor - *urban rent burden* for both Atlanta and Boston. Although the variance captured by this factor is low, yet it captured a vital characteristic that distinguishes each city. This character represented the parts of the city, which accommodated well-formed urban design characteristics and connectivity, but vulnerable population group. I added the factor in the model to explore its impact on MWS. For Atlanta, results indicate that the urbanness factor within 1 mile of the tweet location is significant for all the quartiles of MWS at 0.1% level. MWS increases by 5.61% at 0.25 quantile, 12.97% at the 0.5 quantile, and 11.71% at 0.75 quantile for a unit increase in the urbanness factor. The increase in the median value of MWS was 12.97% for a unit increase in the urbanness factor score, which was higher than the change seen at the lower and the upper quartiles. Thus, the effect size of the urbanness score on MWS was higher on the top quartile when I increased the analysis area from 1/4 miles to 1 mile. The effect of urbanness was still strongest at the median MWS score. The fear of crime and poverty factor within 1 mile of the tweet location was significant at all quartiles of MWS at 0.1% level. MWS decreases by 1.9% at 0.25 quantile, 7.59% at the 0.5 quantiles, and 7.29% at the 0.75 quantiles, for one unit increase in the fear of crime and poverty factor. This shows that the effect of fear of crime and poverty factor increases for median values of MWS and remains relatively constant at the upper quartiles of MWS. The urban rent burden factor was not found to be significant at the lowest and the topmost quartile; however, it was significant at 1% level for the median quartile. There is a 2.49% decrease in the MWS for a unit increase in the urban rent burden factor. This is true for areas around tweet locations with a median value of MWS; i.e., we may see a rise in the MWS score if rent burden and eviction rates decrease. For Boston, we see slightly different results as compared to Atlanta. The urbanness factor within 1 mile of tweet location was significant at 0.1% level for all quartiles. MWS increases by 9.5% at 0.25 quantile, 17.61% at 0.5 quantile, and 14.97% at 0.75 quantile for one unit increase in the urbanness score. Like Atlanta, the highest increase in MWS was seen at the median. Fear of crime and poverty factor was significant at 1% level for the median and 0.75 quantile. MWS decreases by 2.84% at 0.5 quantile and 4.13% at 0.75 quantile for one unit increase in the fear of crime and poverty factor. This effect size increases in the top quartile. The urban rent burden factor was not significant for any quartile of MWS.

Table 5.14 shows quantile estimates for mental wellbeing score on urbanness, street safety perception, and socio-economic deprivation within 1/4 miles and 1 mile of a person's tweet location. The model includes both Atlanta and Boston. *City* was included as a dummy variable to control any differences across cities in any observable or unobservable predictors. An unobservable predictor, in this case, could be people's cultural constructs. Thus, the model can be conceptually represented as:

$$MWS^{\tau} => \beta_0^{(\tau)} + \beta_1 Comfortable\, weather\, conditions^{(\tau)} + \beta_2 City^{(\tau)} + \beta_3 Urbannes^{(\tau)} + \\ \beta_4 Street\, safety\, perception^{(\tau)} + \beta_5 Socio-economic\, deprivation^{(\tau)}$$

I combined the dataset for two cities to examine if there are any commonalities between the two cities and the single city model. I was also interested in investigating if any new factors may surface in explaining the mental wellbeing score. One additional factor that gained prominence in the combined dataset is *street safety perception*. As explained previously, the factor had positive loading for variables that impacted street safety perception, and it had high negative loading for misdemeanor. These characteristics are essential for safety perception. The third factor had similar loading's as *fear of crime and poverty*, but I renamed it as socio-economic deprivation, as I already covered the concept of safety perception.

The results indicated that the urbanness factor within 1/4 miles and 1 mile of the tweet

Table 5.13: Quantile Regression Models explaining mental wellbeing score (MWS), within 1 mile of a peron's tweet location, for Atlanta and Boston.[a b]

	$\tau(0.25)$				$\tau(0.50)$				$\tau(0.75)$			
	(coeff)$_t$	(coeff)$_{raw}$	Std Er	P -value	(coeff)$_t$	(coeff)$_{raw}$	Std Er	P -value	(coeff)$_t$	(coeff)$_{raw}$	Std Er	P -value
Atlanta												
Intercept	866.683	2.269	0.010	***	1069.311	2.459	0.012	***	1396.479	2.706	0.016	***
Comfortable weather	4.248	0.042	0.001	***	5.475	0.0533	0.001	***	6.599	0.064	0.001	***
Factors												
(1)Urbanness	5.61	0.055	0.009	***	12.975	0.122	0.012	***	11.717	0.111	0.016	***
(2)Fear of crime and poverty	-1.961	-0.020	0.008	**	-7.596	-0.079	0.011	***	-7.291	-0.076	0.015	***
(3)Urban rent burden	-1.499	-0.015	0.008	-	-2.498	-0.0253	0.011	**	-0.965	-0.010	0.015	-
Pseudo R- squared	0.2789				0.4474				0.5649			
No of assessment grids	1659				1659				1659			
Boston												
Intercept	597.756	1.9427	0.006	***	677.179	2.0505	0.007	***	747.144	2.1367	0.011	***
Comfortable weather	1.704	0.0169	0.001	***	2.491	0.0246	0.001	***	4.050	0.0397	0.001	***
Factors												
(1)Urbanness	9.516	0.0909	0.006	***	17.610	0.1622	0.007	***	14.970	0.1395	0.011	***
(2)Fear of crime and poverty	-0.896	-0.009	0.006	-	-2.849	-0.0289	0.007	***	-4.123	-0.0421	0.011	***
(3)Urban rent burden	-0.050	-0.0005	0.006	-	0.280	0.0028	0.007	-	0.531	0.0053	0.011	-
Pseudo R- squared	0.2731				0.4423				0.5974			
No of assessment grids	1161				1161				1161			

[a] (coeff)$_t$ is transformed coefficient and gives percentage %values, (coeff)$_{raw}$ is raw coefficient for $log(MWS)$,'Std Er' is standard error.
[b] P-value $*** < 0.000$; $** < 0.001$;$* < 0.05$

location are significant for all the quartiles of MWS at 0.1% level. MWS increases by 8.75% at 0.25 quantile, 9.22% at the 0.5 quantile, and $8.51.71\%$ at 0.75 quantile for one unit increase in the urbanness factor within 1 mile of tweet location. The increase in the median value of MWS was 9.22% for a unit increase in the urbanness factor score; this value was higher than the change at the lower and the upper quartiles. MWS increases by 12.98% at 0.25 quantile, 13.52% at the 0.5 quantile, and 11.07% at 0.75 quantile for one unit increase in the urbanness factor within 1/4 miles of tweet location. From the results, it is clear that the maximum impact of the urbanness score is seen at the median value of the MWS score in most cases. Street safety perception factors within 1/4 miles and 1 mile of tweet location were significant at all quartiles. For within 1 mile of tweet location, the street safety perception factor was significant at 5% level of significance for 0.25 quantile, at 1% level of significance for the 0.5 quantile, and at 0.1% level of significance for 0.75 quantile. For within 1/4miles of tweet location, street safety perception was significant at 0.1% level for all quartiles. Within 1 mile of tweet location, if street safety perception score increases one unit, the MWS increases by 1.5% at 0.25 quantile, by 3.09% at the 0.25 quantile, and 5.68% at the 0.75 quantile. The increase was highest at the top quartile. Within 1/4 mile of tweet location, if street safety perception score increases one unit, the MWS increases by 7.8% at 0.25 quantile, by 8.63% at the 0.25 quantile, and 6.69% at the 0.75 quantile. Within 1/4 mile of tweet location, I observed that the impact of street safety perception score had almost the same impact at all quartiles of MWS score. The third-factor socio-economic deprivation was not significant at the lowest quartile of MWS within 1 mile of tweet location. The factor was significant at the level of 0.1% for both 0.5 quantile and 0.75 quantile within 1 mile of tweet location. MWS decreases by 1.6% at 0.5 quantile, 2.7% at 0.75 quantile for one unit increase in the socio-economic deprivation factor within 1 mile of the tweet location. For within 1/4 miles of tweet location socio economic deprivation was significant at the level of 1% for the lowest and the top quartile of MWS. At median MWS, socio economic deprivation factor was significant at the level of 0.1%. MWS decreases by 1.2% at 0.25 quantile, 2.34% at 0.5 quantile, and 2.93% at 0.75 quantile for one unit increase in socio economic factor within 1/4 miles of the tweet location.

The dummy variable for Atlanta is *is Atlanta*. The dummy variable is the categorical variable that reveals the impact of the city for different conditional quantiles of the MWS score. The result showed the variable was significant for all quartiles at 0.1% significance. This shows that cities have an impact on MWS score. For within 1/4 miles and 1 mile of tweet location, the MWS is less in Atlanta, compared to that of Boston. For within 1 mile of tweet location, MWS at 0.25 quantile was 3.6% less in Atlanta than in Boston. This decrease in MWS is 5% and 10% at 0.5 and 0.75 quartile, respectively. Meaning the difference in MWS within 1 mile of tweet location between Atlanta and Boston was highest at the top quartile. A similar trend was seen for areas within 1/4 miles of a tweet location. MWS at 0.25 quartile was 7.8% less in Atlanta than in Boston. At 0.5 quartile and 0.75 quartile, the decrease is 7.94% and 7.00% respectively. This shows the decrease is constant within 1/4 miles of tweet location.

Spatial Lag and Spatial Error Regression: I have previously described the reason for the choice of the quantile regression model (QRM) as my main explanatory model in Section 4.6.2. To recapitulate, QRM has reliable performance when there is a skewed distribution in the response variable.QRM does not make an assumption on error term distribution. Moreover, QRM makes it easier for us to measure the different effects of an explanatory variable on different points of distribution of response variable [229]. However, from the first law of geography, we know - "everything is related to everything else. But near things are more related than distant things" [230]. It is not uncommon for spatial variables to exhibit spatial auto-correlation. Spatial auto-correlation refers to the presence of systematic spatial variation in both explanatory and response variables. The spatial autocorrelation is positive when the adjacent values have similar values and negative when adjacent spatial values are contrasting [231]. Spatial auto-correlation is measured by *Moran's I*. To compute *Moran's I*, I created a spatial weight matrix which is an essential component in the

Table 5.14: Quantile Regression Models explaining mental wellbeing score (MWS), within 1/4 miles, and 1 miles of a person's tweet location, in both Atlanta and Boston.[a] [b]

	$\tau 0.25$				$\tau 0.50$				$\tau 0.75$			
	(coeff)$_t$	(coeff)$_{raw}$	Std Er	P -value	(coeff)$_t$	(coeff)$_rawt$	Std Er	P -value	(coeff)$_t$	(coeff)$_{raw}$	Std Er	P -value
1Mile												
Intercept	594.554	1.9381	0.010	***	658.521	2.0262	0.012	***	760.635	2.1525	0.019	***
Is Atlanta	-3.140	-0.0319	0.016	*	-5.710	-0.0588	0.019	**	-10.031	-0.1057	0.030	***
Comfortable weather	2.071	0.0205	0.001	***	3.345	0.0329	0.001	***	4.331	0.0424	0.001	***
Factors												
(1) Urbanness	8.752	0.0839	0.003	***	9.221	0.0882	0.004	***	8.513	0.0817	0.007	***
(2) Street Safety Perception	1.501	0.0149	0.007	*	3.097	0.0305	0.009	**	5.686	0.0553	0.015	***
(3) Socio-Economic Deprivation	-0.319	-0.0032	0.004	-	-1.676	-0.0169	0.005	***	-2.761	-0.028	0.007	***
Pseudo R- squared	0.3				0.4518				0.6253			
No of assessment grids	2820				2820				2820			
1/4Miles												
Intercept	624.057	1.9797	0.005	***	670.446	2.0418	0.007	***	751.220	2.1415	0.013	***
Is Atlanta	-7.818	-0.0814	0.006	***	-7.946	-0.0828	0.01	***	-7.003	-0.0726	0.017	***
Comfortable weather	1.898	0.0188	0.001	***	3.159	0.0311	0.005	***	4.050	0.0397	0.001	***
Factors												
(1) Urbanness	12.987	0.1221	0.004	***	13.519	0.1268	0.007	***	11.071	0.105	0.010	***
(2) Street Safety Perception	7.853	0.0756	0.005	***	8.632	0.0828	0.007	***	6.695	0.0648	0.012	***
(3) Socio-Economic Deprivation	-1.203	-0.0121	0.004	**	-2.342	-0.0237	0.006	***	-2.926	-0.0297	0.011	**
Pseudo R- squared	0.2848				0.4582				0.6067			
No of assessment grids	2823				2823				2823			

[a] (coeff)$_t$ is transformed coefficient and gives percentage $\%$ values, (coeff)$_{raw}$ is raw coefficient for $log(MWS)$, 'Std Er' is standard error.
[a] P-value $*** < 0.000$; $** < 0.001$;$* < 0.05$

computation of spatial autocorrelation statistics. I have used *Geoda* to create the spatial weight matrix of the assessment grids or spatial units used in this study. Rook contiguity approach was used to measure the spatial weights for the assessment grids [232].[3] Subsequently, I computed *Moran's I* for the explanatory variables (X's) and response variable (Y's). The *Moran's I* value for the response variable mental wellbeing score in both cities combined is 0.299 (p-value < 0.001). This shows a low to moderate spatial auto-correlation in the response variable. *Moran's I* values for the key independent variables (for both cities combined, and within 1 mile of a person's tweet location) such as urbanness, street safety perception, and socio-economic deprivation were 0.926, 0.916, and 0.789, respectively (p-value < 0.001). The spatial autocorrelation for the independent variables was found to be high. Figure D.1 shows the Moran's I scatter plots for response and explanatory variables.

Given such high values of spatial auto-correlation, it is highly probable that the error values of the regression will cluster and will not be random and normally distributed. Although quantile regression models do not make strict assumptions on error term distribution yet, scholars have pointed out that the model may not fully account for the spatial variation in the response variables [229]. Quantile regression model may exhibit arbitrary forms of heteroscedasticity and spatial dependence. There are a few techniques that have been tried and tested by [229, 210] to account for spatial lag and spatial error type dependence, recommended in the conservative estimation of the coefficients [233]. However, due to the unavailability of standard spatial quantile regression models, incorporating this model in my research is currently out of scope. Keeping this in mind, I have provided the spatial-lag and spatial-error models alongside ordinary least square (OLS) models to compare the regression coefficients of the explanatory variables as well as their statistical significance.

[3] To compute weights based on rook contiguity only common sides of the polygons are considered to define the neighbor relation (common vertices are ignored).

OLS Model can be conceptually represented as:

$$MWS => \beta_0 + \beta_1 Comfortable\, weather\, condition\ + \beta_2 City\ + \beta_3 Urbanness + \\ \beta_4\, Street\, safety\, perception\ + \beta_5 Socio - economic\ deprivation$$

Spatial error regression and spatial lag regression has extra term/coefficients. The spatial lag regression can be represented as:

$$MWS => \beta_0 + \beta_1 Comfortable\, weather\, condition\ + \beta_2 City\ + \beta_3 Urbanness + \\ \beta_4\, Street\, safety\, perception\ + \beta_5 Socio - economic\ deprivation + \\ \beta_6 Spatially\, correlated\, errors$$

Here the spatially correlated errors captures the the influence of the errors from nearby locations.

The spatial lag regression can be represented as:

$$MWS => \beta_0 + \beta_1 Comfortable\, weather\, condition\ + \beta_2 City\ + \beta_3 Urbanness + \\ \beta_4\, Street\, safety\, perception\ + \beta_5 Socio - economic\ deprivation + \\ \beta_6 Spatial\, lag$$

Here the spatial lag captures the influence of the MWS values in the nearby locations.

Table 5.19 and Table 5.20 show the results of OLS, spatial error, and spatial lag models explaining the mental wellbeing score within quarter miles and 1 mile of a person's tweet location for both Atlanta and Boston. The results of these models are still consistent with the results from the quantile regression model results. The spatial lag regression model is the better performing model, as we see a decrease in the *Akaike info criterion* and *Schwarz criterion* value alongside a slight increase in the R-Squared values. Urbanness, street safety perception, and socio-economic deprivation factors were significant after accounting for spatially correlated errors and spatial lag. The significance of the city-level differences remains (captured by the *is Atlanta* variable) when I consider factors within 1/4th miles of a tweet location. The city-level significance disappears when I consider factors within 1 mile of a person's tweet location after adjusting for spatial lag. I can thus conclude that built environment factors impact mental wellbeing even adjusting for spatial autocorrelation. Only the city level differences of mental wellbeing score disappear as I increase the geographical scale of the built environment factors from a 1/4th miles to 1 mile.

Random Forest Model (RFM): Next, a random forest classification model was tested to determine the impact of individual variables on MWS (the outcome variable). Random forest classifiers can model non-linearity in the data. These models are also more tolerant to multicollinearity and outliers. Since random forest classification models do not specify significance, I used bi-variate Pearson correlation coefficients and SHAP values to explain the relationship with MWS. Similar methods have been used in geography and planning, where there is often high multicollinearity between variables [234].

Table 5.15: Precision and recall metrics on test set of the Random forest classifier for the city of Atlanta.

# of class	class name	Training set		Test set	
		Precision	Recall	Precision	Recall
4	very low	85.21	86.12	58.12	55.19
	low	79.77	81.34	60.13	61.02
	high	84.64	82.32	58.12	61.11
	very high	83.32	84.51	59.65	57.32
Average		83.24	83.46	59.00	58.66

RFM predicts the outcome variable by the creation of many component decision trees (a hyperparameter that needs to be specified). First, RFM samples training data points with replacements and utilizes a subset of the features from the training set. Next, for each component decision trees, it supplies this sampled training set (also called bootstrapped subsets). Finally, it makes a decision on predicting the class label (the outcome variable) by a voting mechanism. This approach to train component trees on different bootstrapped

Table 5.16: Precision and recall metrics on test set of the Random forest classifier for the city of Boston.

# of class	class name	Training set		Test set	
		Precision	Recall	Precision	Recall
4	very low	81.01	89.12	61.22	57.13
	low	82.21	78.50	57.34	51.54
	high	88.78	77.33	62.43	61.11
	very high	89.08	82.21	56.23	58.02
Average		85.27	81.79	59.31	56.95

subsets is also called *bagging*, which stands for bootstrap aggregation [235]. For this analysis, I utilized Scikit-Learns' random forest classifier [213] module to instantiate and train the RFM model.

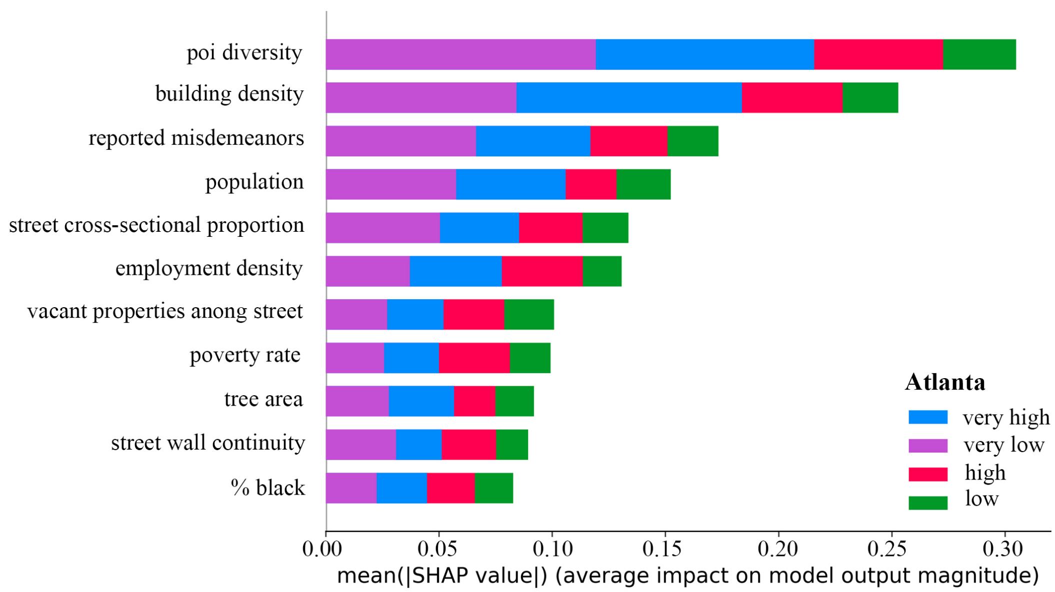

Figure 5.9: Built environment feature importance for predicting MWS using Random Forest model for Atlanta. The features are computed within 1 mile of a person's tweet location.

To build this classifier, I divided the mental wellbeing score into four classes based on quartiles. This was done to ensure the assessment was comparable to the QRM. Scores below the first quartile (Q1) were named *very low*. Scores below the second quartile (Q2) or median ($0.5\ quantile$) were named *low*. Scores below the third quartile (Q3) and above median were named *high*. Lastly, the scores above the third quartile were named *very high*. For the selection of built environment variables, I used the correlation matrix for each city. I retained a set of hand-chosen variables of interest, given there were multiple variables that were highly correlated among each other, with Pearson correlation coefficients above ± 0.8. ReferFigure C.1, Figure C.2, and Figure C.3 for correlation coefficients between explanatory variables. The labeled dataset was then used to train the RFM for class prediction. For each city, I divided the assessment grid points into training (70%) and test datasets (30%). Using a random-search based hyperparameter tuning approach, the random forest model with the best hyperparameter settings was selected for each city and then for both cities. Hyperparameters specified to this model included, num-estimators: 500, max-depth: 15, and max-num-leaf-split: 50. I used only variables computed within 1 mile of the tweet location for the task. The accuracy of MWS into four classes using only built-environment and socio economic variables was found to be 63% for Atlanta, 60.09% for Boston, and 59.78% for both cities together (measured on the test set). Please refer Table 5.16 and Table 5.15 for other performance indicators of this model. Noteworthy to mention that the accuracy increased to 82-85% when I used only two class labels *high* and *low*. However, to account for a fair comparison with the QRM model with four quartiles, I continued the analysis with the aforementioned four class labels.

As can be seen in Figure 5.9, in Atlanta, POI diversity, building density, reported misdemeanors, population, and street cross-sectional proportions within 1 mile of a person's tweet location were the top five most important features for predicting the MWS score. The impact of these features was highest on predicting very low MWS, which is the lowest quartile (Q1), and very high MWS which is the topmost quartile (above Q3). Relatively less impact of these features was observed in predicting the low and high MWS. Comparing the results using factor analysis and QRM results, in Atlanta, the top five most influential built environment features in predicting MWS were included in the *urbanness* factor. Pearson correlation coefficient of Atlanta showed that the top five features were significant and positively correlated to MWS.

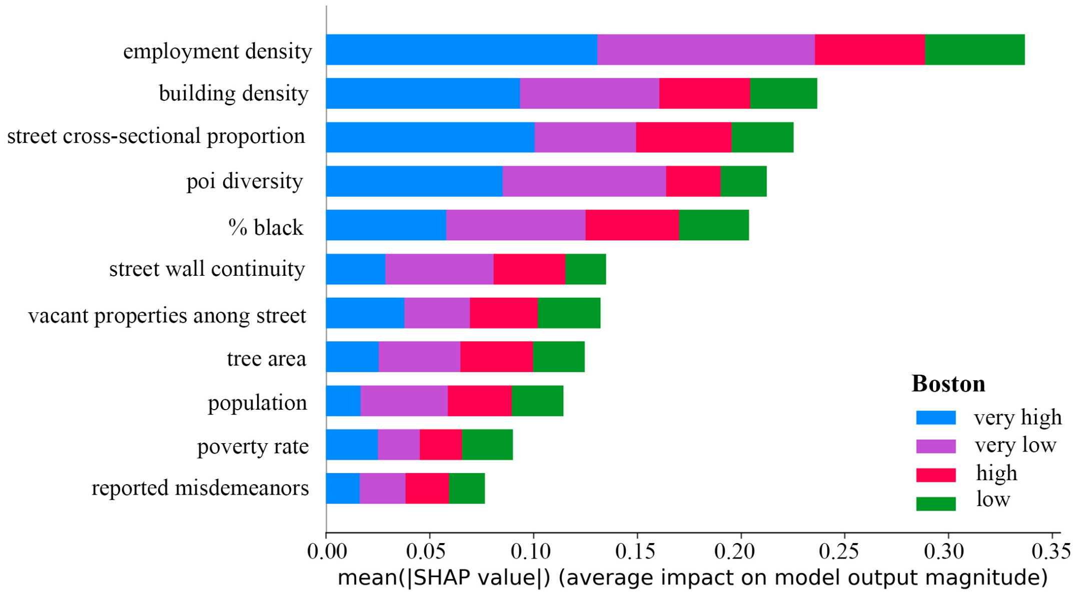

Figure 5.10: Built environment feature importance for predicting MWS using Random Forest model for Boston. The features are computed within 1 mile of a person's tweet location.

Figure 5.10 shows in Boston, employment density, building density, street cross-sectional proportion, poi diversity, and % black within 1 mile of a person's tweet location were the top five most important features for predicting MWS. Like the results in Atlanta, the impact of these features was highest in predicting very high MWS and very low MWS. Except for % black, all other features were included in the urbanness factor. The Pearson correlation coefficient of Boston showed that the top four features were significant and positively correlated to MWS. % black showed a low negative correlation to MWS, and it was found to be significant. This indicated that race and socio economic conditions within 1 mile of a tweet location in Boston had importance in predicting MWS.

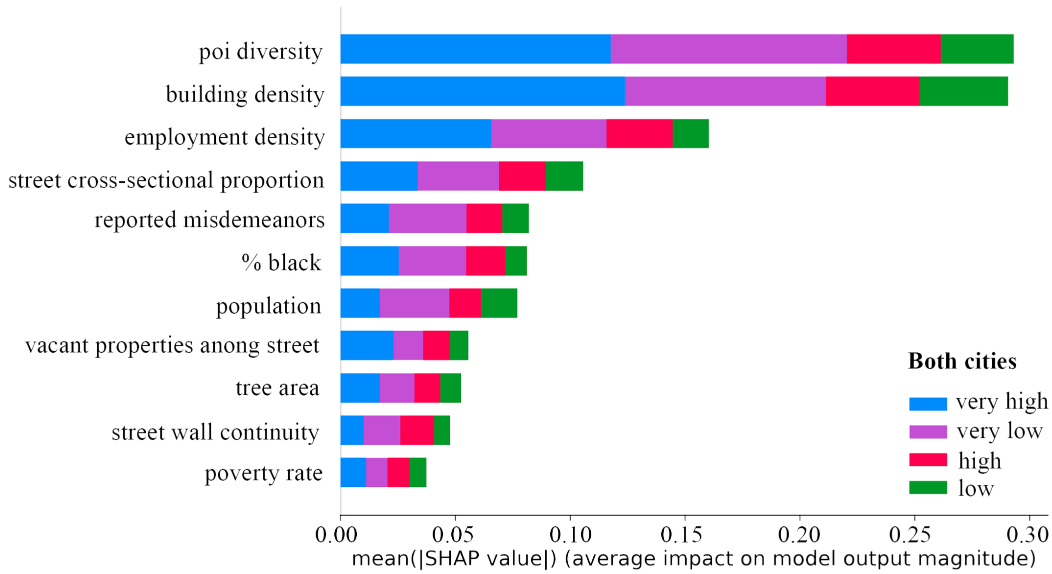

Figure 5.11: Built environment feature importance for predicting MWS using Random Forest model for both cities. The features are computed within 1 mile of a person's tweet location.

In addition, Figure 5.11 shows that for both cities, POI diversity, building density, employment density, street cross-sectional proportion, and reported misdemeanor within 1 mile of a person's tweet location were the top five features in predicting MWS. These features were most important in predicting the very high and very low MWS. The top five features here were also included in the *urbanness* factor.

The RFM results showed that individual built environment features had a low to moderate impact in predicting MWS. While the accuracy of prediction was higher for very low and very high MWS, the in-between classes (low and high) were harder to predict. This indicated that the built environment characteristics of areas with very high and very low MWS were different. *POI diversity*, *building density*, and *street cross-sectional proportion* within 1 mile of a person's tweet location were the three key features in the random forest classifier that differentiated areas with very high MWS from very low MWS. These features were common in the urbanness factor derived through factor analysis.

Topic models: I generated two sets of topics for Atlanta and Boston. The first set of topics were extracted using the entire tweet text corpus of the two cities. Table 5.17 shows the 5 topics from each city. The words under each topic were arranged based on their importance in generating the topics. For instance, Topic 1 in Atlanta is composed of the words *atlanta, georgia, chilling, hiphop, mixtape, downtown, record, feelings*. Among these words, *atlanta* showed the most important, and *feelings* showed the least importance in defining the topic. I discarded the set of words with a very low importance score, which did not impart any meaning to the topic. Next, I assigned a theme or name to each topic, which seemed most suitable in describing the topic.

Table 5.17: Topics, words and themes that emerged from the semantic analysis of tweets in Atlanta and Boston. The words are arranged in the order of their importance.

	Words in the order of importance	Themes
Atlanta		
Topic 1	atlanta, georgia, hiphop, mixtape, chilling, downtown, record, feelings	Music
Topic 2	buckhead, traffic, morning, school, airport , international, hartsfield, jackson, streets,	Traffic
Topic 3	drinking brewing, scofflaw, mondaynight, scofflawbrewing, garage, barrell, coffee, orange, funkybuddhabrew	Drinking
Topic 4	mercedes, stadium, bitches, aquarium, falcons, treated, hanging, learned	Events
Topic 5	people, saturday, trying, Friday, weekend, couple, Sunday, rapper, artist, moment,	Activity
Boston		
Topic 1	boston, station, accident , traffic, center, city, airport, international, dropping, school, seaport, difficult	Traffic
Topic 2	drinking, beer, company, trilliumbrewing, samuel, bitter, sitting, madness, double, samueladamsbeer	Drinking
Topic 3	employee, establishing, department, passed, enroll, student, retirement, mentalhealth, correction	Work
Topic 4	weather, delays, flightdelay, degrees, departure, suddenly, conditions, experiencing , currently, obsessed	Weather
Topic 5	marathon, repost, better, getting, running, Friday, family , coffee, episode, afternoon, couple , picture , harvard	Activity
Topic 6	fenway, redsox, season, terrible, minute, watching, studpid, night, baseball, selfie, chance, falling	Sports

In Atlanta, the five most prominent topics that appeared were *music, traffic, drinking, events, and activity*. The topic *music* described the music culture of the city and how people love music and use it as part of leisure. *Traffic* is related to traveling in general, traveling to the airport, dropping kids to school, stress related to morning traffic. *Drinking* is part of recreational activities that people frequently do in Atlanta. It may include drinking coffee or enjoying a beer in a brewery. *Event* topic includes words where people spend time infamous point of interest (POI), including a game stadium like Mercedes-Benz Arena or visiting the aquarium. The topic *Activity* includes leisurely weekend activity, people watching, visiting art-show, and others.

Table 5.18: Topics, words and themes that emerged from the semantic analysis of tweets from high mental wellbeing score and low mental wellbeing scores. The words are arranged in the order of their importance.

	Atlanta	**Boston**
High MWS		
Topic 1	fulton, stadium, mercedes, buckhead	boston, drinking, beautiful, weekend, tonight
Topic 2	drinking, modaynight, summer,pretty, dinner	roxbury, college, street , restaurants, school
Topic 3	friend, weekend, birthday, partytime	drinking, redsox, sameuladamsbeer, better, trilliumbrewing
Topic 4	buckhead, fitness, museum, community, center, yoga	massachusetts, fenway, brighton, allston
Topic 5	drinking, Saturday, coffee, really, westside, brewery	weekend, garden, looking, tomorrow, breakfast
Topic 6	beautiful, tonight, morning, weekend, happy	tonight, street, people, Saturday, morning,
Topic 7	niggas, wasted, wicked, problem, scared , killer	realdonaldturmp, racist, terrible, person
Topic 8	-	employee, department, engross, task
Low MWS		
Topic 1	church, school, community, restaurant, couple	brighton, street, couple, buffet, laundry
Topic 2	birthday, christmas, beautiful, special	neighborhood, dorchester, people, beautiful
Topic 3	morning, thanks, family, weekend, amazing	drinking, roxbury, franklin, rosindale
Topic 4	sagave, drinking, lowery, highland	school, boston, center
Topic 5	closed, fighters, killed, tonight, conversations	anxiety, accident, stress, people, recovered
Topic 6	hate, reason, brakes, applied	boston, violence, students, protests

In Boston, I observed a few similar topics and a few different. The six most prominent topics in Boston found were *traffic, drinking, work, weather, activity, and sports*. Work and weather were two new topics that emerged in Boston. *Work* topic included conversations related to employment or employee, student life, mental health. *Weather* was discussed in the context of flight delays, sudden or unexpected conditions due to weather. The emergence of these topics was expected in Boston, as it is the nation's one of the key employment centers, housing many educational institutions. Simultaneously, weather in Boston is significantly more unpredictable than in Atlanta, as such weather-related topics were expected. The sixth topic found in Boston was *sports*. It included the name of the baseball team (Redsox), stadium name (Fenway), and words describing the experience of watching games. The topic showed that participating in sports events might be a major part of life in Boston.

The second set of topics were generated using a spatial component. I divided the tweet text corpus into text corpus for areas with high mental wellbeing score (MWS), and low mental wellbeing score. Table 5.18 shows 14 topics that were generated from the corpus. Similar to the first set of topics, I arranged the words in these topics in order of their importance. In areas with high MWS, more events and activity topics were generated. Diverse words related to various relaxation/de-stressing activities emerged under each topic. For instance, topics 4 and 5 occurring in areas with high MWS Atlanta indicates people's engagement in fitness and drinking activities on weekends. In Boston, topic 1 described the experience of leisurely additives. There were stressed topic clusters, such as topic 7. In areas with high MWS in Atlanta, a topic cluster with racial slurs and swear words appeared. A possible reason could be that people use them while getting into a bar fight. In Boston, topic 7 primarily included words expressing dissatisfaction on political issues. Areas with low MWS showed topic clusters around schools, communities, family activities, and drinking in both cities. Here, the diversity of words describing different leisurely activities was limited. Topic 5 and 6 extracted for areas with low MWS were primarily stress word clusters. In Atlanta, words seemed to describe fights, gruesome incidents, hatred, etc. In Boston as well, Topics 5 and 6 described severe stressful conditions such as violence, student protests, etc.

Using these results, in Chapter 6 I have answered the research questions and discussed the implication of my findings for urban planners and designers.

Table 5.19: OLS, spatial error, and spatial lag regression models explaining mental wellbeing score (MWS), 1 mile of a person's tweet location, in both Atlanta and Boston.[a] [b]

	OLS				Spatial Error Model				Spatial Lag Model			
	(coeff)t	(coeff)raw	Std Er	P -value	(coeff)t	(coeff)raw	Std Er	P -value	(coeff)t	(coeff)raw	Std Er	P -value
1Mile												
Spatially correlated errors					11.136	0.106	0.029					
Spatial lag	-	-	-		-	-	-		7.591	0.073	0.010	***
Intercept	863.989	2.266	0.027	***	860.362	2.262	0.019	***	702.354	2.082	0.036	***
Is Atlanta	-10.532	-0.111	0.044	**	-10.342	-0.109	0.047	**	-5.337	-0.055	0.044	
Comfortable weather	1.319	0.013	0.001	***	1.299	0.013	0.001	***	1.298	0.013	0.001	***
Factors												
(1) Urbanness	19.115	0.175	0.009	***	19.595	0.179	0.010	***	15.817	0.147	0.009	***
(2) Street Safety Perception	6.972	0.067	0.021	**	6.847	0.066	0.022	**	4.369	0.043	0.021	*
(3) Socio Economic Deprivarion	-3.629	-0.037	0.010	***	-3.510	-0.036	0.011	*	-3.390	-0.034	0.010	***
No of assessment grids	2819				2819				2819			
Pseudo R- squared	**0.563**				**0.569**				**0.573**			
Adjusted R-squared	0.562											
Multicolinearity Condition Number	8.040											
Akaike info criterion	3468.380				3443.950				3412.510			
Schwarz criterion	3504.050				3479.630				3454.130			

Test for Spatial Dependence	**Value**	**P-Value**
Moran's I (error)	**4.864**	***
Lagrange Multiplier (lag)	59.335	***
Robust LM (lag)	38.186	***
Lagrange Multiplier (error)	22.755	***
Robust LM (error)	1.606	

[a] $(\text{coeff})_t$ is transformed coefficient and gives percentage $\%$ values, $(\text{coeff})_{raw}$ is raw coefficient for $log(MWS)$, 'Std Er' is standard error.
[b] P-value $*** < 0.000$; $** < 0.001$;$* < 0.05$

Table 5.20: OLS, spatial error, and spatial lag regression models explaining mental wellbeing score (MWS), 1/4 miles of a person's tweet location, in both Atlanta and Boston.[a] [b]

	OLS				Spatial Error Model				Spatial Lag Model			
	(coeff)t	(coeff)raw	Std Er	P -value	(coeff)t	(coeff)raw	Std Er	P -value	(coeff)t	(coeff)raw	Std Er	P -value
1/4 MIle												
Spatially correlated errors					13.598	0.127	0.020	***				
Spatial lag	-	-	-		-	-	-		9.555	0.091	0.009	***
Intercept	884.339	2.287	0.014	***	877.727	2.280	0.016	***	693.958	2.072	0.025	***
Is Atlanta	-11.257	-0.119	0.019	***	-11.796	-0.126	0.021	***	-7.865	-0.082	0.019	***
Comfortable weather	1.214	0.012	0.001	***	1.194	0.012	0.001	***	1.191	0.012	0.001	***
Factors												
(1) Urbanness	27.850	0.246	0.013	***	28.473	0.251	0.014	***	22.827	0.206	0.013	***
(2) Street Safety Perception	13.412	0.126	0.014	***	13.232	0.124	0.015	***	10.504	0.100	0.014	***
(3) Socio Economic Deprivarion	-5.599	-0.058	0.012	***	-5.403	-0.056	0.013	***	-5.065	-0.052	0.012	***
No of assessment grids	2818.000				2818.000				2818.000			
R- squared	**0.569**				**0.577**				**0.585**			
Adjusted R-squared	**0.568**											
Multicolinearity Condition Number	3.342											
Akaike Info Criterion	3474.310				3437.820				3376.990			
Schwarz Criterion	3509.990				3473.500				3418.620			

Test for Spatial Dependence	**Value**	**P-Value**
Moran's I (error)	5.977	***
Lagrange Multiplier (lag)	102.199	***
Robust LM (lag)	69.605	***
Lagrange Multiplier (error)	34.661	***
Robust LM (error)	2.067	

[a] (coeff)$_t$ is transformed coefficient and gives percentage $\%$ values, (coeff)$_{raw}$ is raw coefficient for $log(MWS)$, 'Std Er' is standard error.
[b] P-value $*** < 0.000$; $** < 0.001$;$* < 0.05$

CHAPTER 6

ARE CITIES DETRIMENTAL TO MENTAL HEALTH? : A DISCUSSION

"Living in cities is associated with a higher risk of some mental health problems – such as mood disorders, anxiety, and schizophrenia – when compared to living rural." - is highlighted by World Health Organization (WHO) in their web application on urban health 'Tackling mental health in cities' in the year 2020 [236, 169].

'Is it a reality, or the claim that cities are not supportive of mental health needs revisitation?'

The theory of mental health and built environment dates back to the 1940s. Farris Dunham's study showed there were more cases of psychopathology in large cities [10]. Since then, the theory has been carried forward with much ongoing debate. The question was, and it has always been: historically, were cities causing mental health issues? Or, people with mental illness were in cities, as they had better access to mental health facilities [9]? International and national health organizations (like the World Health Organization or National Institute of Health) have quoted time and again that 'urban living' isn't good for mental health. Multiple studies support the claim, which projected that cities are an accumulation of social and environmental stressors. These stressors act against mental health, increasing mood disorders, anxiety disorders, risk of schizophrenia [169, 237]. However, before blaming the urban, we need to investigate why urban living is bad for mental health? What can we do to create mental health-friendly living 'conditions'? In this chapter, I discuss how the new findings from this research can help planners or policy-makers to improve mental wellbeing in cities.

As explained in Chapter 2, past studies investigating the association and relationship between mental health and built environment were primarily based on three methodological practices. The first group of studies used highly aggregated datasets for counties. To give an idea of scale, counties are larger than cities. Usually, a county contains vast areas of both urban and less urban areas. The second group of studies relied on small-scale urbanized or suburban areas. These studies surveyed local communities, where people self-reported their happiness or frustrations in living there. The third group of studies conducted a meta-analysis of past studies reporting the urban-rural prevalence of mental health disorders. While these studies highlighted the problem with urban, how they differentiate and define urban is unclear. Fine-grain built environment characteristics were not controlled in these studies. Aggregated mental health data from counties are reliable indicators of the severity of the problem is in more urbanized counties, but it is not enough to explain what exactly built environment and social conditions might have inflicted mental health disorders on people. If we keep defaming ’urban’ without examining what makes the urban population vulnerable to mental health issues, we will never be able to save the urban population from this crisis.

Center for urban design and mental health, a start-up think tank, elicited that it is challenging to design research that can demonstrate the causal relationship between mental health and urban design [238]. Considering these methodological issues, this research was designed to address two key challenges: (1) find a relevant data source where people express their ‘affect’ naturalistically, and they have relatively precise location information 2) analyze built environment at a granular level around which people usually express their ‘affect’. Following along this approach enabled me to differentiate areas based on a quantitative measure of mental wellbeing, which I named mental wellbeing score (MWS). Very high or a high mental wellbeing score of a ’10 acre’ (represented by a 1/8th mile grid square) area grid indicates a proportionate measure of how well people are doing in terms of their mental health compared to areas with low or very low MWS. Additionally, the fine-grain geospatial analysis of the built environment (of fixed buffer areas of 1/4th miles and 1 mile) around these 10-acre blocks helped differentiate the degree of urban characteristics. I used the term *urbanness* in my research to represent the degree of urban characteristics a place may show. Assessment of a fixed area for their fine-grained urban characteristics made the comparison more meaningful to urban planners, rather than comparing areas defined by census administrative boundaries, such as counties.

The results from my research show that while there is a rationale behind the claim that urban areas are associated with a higher risk of mental health problems, it would be wrong to project that urban characteristics or *urbanness* of a place can aggravate people's mental health. I have explained it further by answering the research questions I initially posited in Chapter 3. This research has conveyed that the density and diversity of spaces in cities do not act as a detriment towards mental wellbeing. Mental wellbeing scores are higher in areas with high building density, high street cross-sectional proportions, easy access to various facilities. The results also show that safety perception plays an important role. Safety perception of areas increases with street enclosure (by buildings), street lights, proximity to urban parks, and cultural facilities such as museums, theater, galleries, etc. In other words, cities provide opportunities to replenish people's minds, improving their cognitive abilities [239]. Many studies show rates of dementia are lower among urban dwellers [240].[1]

6.1 Relationship between Built Environment and Mental Wellbeing

First, I would answer the overarching question: *Is there a relationship between people's mental wellbeing expression on social media and their proximate built environment characteristics in urban areas?*

The relationship between the built environment and mental health was established by psychologists, social scientists, urban designers, and planners decades ago. The datasets used in these studies were either from medical records, survey data, or naturalistic observation of urban spaces. The use of social media data to assess mental health is relatively new.

[1]Dementia is a type of neurocognitive disorder. It typically affects people's ability to learn, memorize, and carry out perceptual-motor functions. Besides, dementia impairs one's attentiveness and problem-solving capability.

The applicability of social media data for understanding human psychology started when it was used to generate automatic sentiment labels in the early 2000s [241]. Gradually, the prediction algorithms' performance improved through the use of hashtags and emojis to detect sentiments in the 2010s [242]. A more sophisticated affective classification and mental wellbeing detection was documented in [243, 27, 244]. These works showcase that it is possible to obtain authentic expressions from social media that can be used to gauge the mental wellbeing of individuals. Given the methods are relatively new and less used in establishing spatial relationships of mental wellbeing, the answer to this question may open doors to future research using large-scale location-based social media data.

Results from my research show that the mental wellbeing score (MWS) derived from Twitter data between May 2018 and March 2020 has a significant relationship with the built environment characteristics. I have conducted three levels of analysis. First, correlation study between the MWS and the built environment variables. Second, understanding the underlying relationship between built environment variables and socio economic indicators to obtain latent features, and subsequently testing their relationship to MWS. And third, individual feature relevance/importance in predicting MWS.

The correlation study Appendix C showed moderate to low significant correlation.POI diversity, access to facilities, and street features had a moderate correlation with MWS. A low yet significant correlation was observed between demographic and social-economic variables. The direction of correlation with individual built environment features was observed to be intuitive. For instance, MWS was negatively correlated with eviction rate in Atlanta, while it was positively correlated to % white and access to parks. Now the question is, can we establish anything from moderate to weak correlations? Interpretation of correlation coefficients depends on the topic or variable of interest. If the variable of interest is difficult to measure, including, in this case, mental wellbeing, the correlations are usually lower than easily countable variables [245]. Besides, we have to understand how much-built environment factors can impact a person's wellbeing compared to other day-today factors. As discussed in Chapter 2, a person's mental wellbeing is primarily defined by major acute and chronic stressors. Figure 6.1 diagrammatically illustrates an individual's stress level. The major acute stressor includes significant 'life events', such as the death of a close family member or any accidental loss suffered by an individual. The major chronic stressors include ongoing difficulties, like long-term health disorders, poverty, and others. The built environment apparently impacts the minor acute and chronic causes of stress. Examples of minor acute causes of stress include daily hassles like traffic problems, inability to get essential jobs done, like grocery shopping, going to the bank, etc. Environmental stressors cause minor chronic stress, including noise and pollution. These are not caused by the built environment itself, instead by the product of urban life. It is extremely difficult to isolate the impact of the built environment on mental wellbeing, given the major acute stressors play the most important role in determining a person's mood and situational response on social media as well. Given this knowledge, I summarize that it is not unusual for MWS to show moderate to low correlation with individually built environment factors [28].

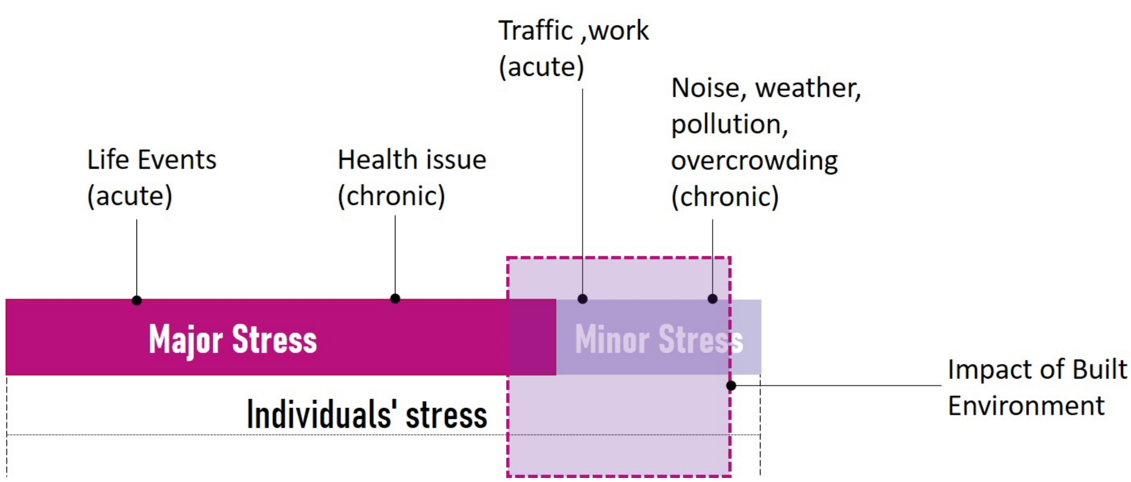

Figure 6.1: An individual's stress level can be roughly understood as sum total of major stress and minor stress. While major stress is a consequence of significant life events, or chronic health issues, minor stress can be resulted from traffic, work related issues, noise, pollution, or overcrowding. Built environment primarily impacts minor stress.

Furthermore, I conducted exploratory factor analysis (EFA) to unveil any latent relationships between built environment variables. Most built environment variables co-exist in a place; as such, understanding its impact on MWS in isolation is difficult. On the contrary, the factors *urbanness*, *fear of crime*, *street safety perception*, and *socio economic deprivation*, derived from EFA, did show a moderately strong explanatory power and theoretical support. For instance, the urbanness factor defined urban areas as places providing access to diverse POI, access to transit, high building density, high street enclosure, population density, and employment density. Concurrently, these areas have high noise, crime issues, and relatively fewer tree-lined streets than their less urban counterparts. This definition of urban is also aligned with current theories. Similarly, other factors like *fear of crime and poverty* and *socio-economic deprivation* are conditions of severe crime, like homicide or robbery in low-income, rent-boasted neighborhoods. People living in such conditions experience extreme uncertainty and stress from impending evictions [246]. Using these factors in quantile regression reduced redundancy and added a theoretical perspective to the findings. The urbanness factor was significant at all quartiles of MWS, and so were the other factors. This indicates the MWS score generated from tweets has a significant relationship with its proximate built environment and socio-economic indicators of the proximate built environment. Lastly, I was interested in isolating some individual built environment variables and socio-economic indicators that can explain MWS. The results showed that individual explanatory variables have moderate (60%) predictive power. Given that the built environment can only impact the minor acute or minor chronic stressors, I conclude that the predictive power of the model is substantial. Thus, through rigorous data exploration using multiple statistical models, I found that indeed there is a relationship between the MWS of a tweet location and its proximate built environment characteristics.

Next, I explain the three hypotheses in my research:

H1. People are less stressed in areas with a higher degree of urbanness.

The results showed that mental wellbeing score (MWS) increases between 8-9% at each quartile, for a unit increase in the urbanness factor, within 1 mile of a persons' tweet location after controlling for all other factors including city, weather conditions, safety perception, and socio-economic conditions. The increase in MWS is even higher, as much as 12-13% at each quartile with a unit increase in urbanness factor within 1/4 miles of a tweet location (Table 5.14). Both within 1/4th miles and 1 mile radius of a person's tweet location I observed, that the highest impact was at the median MWS, meaning areas with median MWS showed greater potential for improvement if the degree of urbanness increases. This trend was similar when both Atlanta and Boston were studied separately. Additionally, the urbanness factor for individual cities on Table 5.10 showed positive loading with features like building density (floor area ratio), cross-sectional proportions, smaller block sizes alongside access to diverse facilities. These findings showed urban built environment features contributed to improving the MWS of cities. These findings were contrary to the existing claim that urban areas are associated with higher mental health risks. Rather, I would claim that urban areas are supportive of mental health if developed with sensitivity and care.

There are some caveats when I claim urban areas support mental health. We need to be mindful of the positive loading of social and environmental stressors like crime and noise on the urbanness factor. Looking at the results, I can conclude that in areas with a higher degree of urbanness, people are less afraid of crime. The findings are aligned with the study results showing fear of crime is more of a concern in low-income areas than in high-income areas. Given two areas with the same number of crime reports, individuals living in low-income neighborhoods are highly stressed by crime, as opposed to those living in higher-income neighborhoods [51, 177]. Similarly, the tolerance level of noise is high among city dwellers. High traffic noise (considered proportionate to average traffic counts) in urban areas has little or no significant negative impact on the mental wellbeing of urbanites and where they choose to de-stress. Besides crime and noise, the street tree cover area was loaded negatively on the urbanness factor of individual cities (Atlanta and Boston). This is intuitive, as cities usually have fewer tree-covered streets compared to their suburban or rural counterparts. Lacking tree cover areas with building enclosure along streets may not show as much negative impact on MWS as areas lacking building enclosure. For instance, a neighborhood street with an aesthetically pleasing building facade, and lacking tree cover, will not negatively impact MWS as much as a street with large building setbacks, vacant lots, and no tree cover would.

The three stressors, crime, noise, and lack of green, need to be addressed by city authorities. While people may not be overly stressed by petty crimes in their neighborhood, a drastic increase in crime level will definitely have a negative impact on mental wellbeing. The same is true for noise and insufficient tree cover. A drastic increase in noise levels and air pollution due to high traffic volume can cause sleep disorders, irritation, and other stress disorders. Although cities do not need to match tree cover in rural areas, tree enclosure along streets, appropriate landscaping, and access to urban parks should not be compromised at the cost of urbanization.

H2. People are less stressed in areas with a greater diversity of escape facilities.

As discussed in the last section, the quantile regression results, as shown in Table 5.14, depict that MWS increases with an increase in the urbanness factor within 1/4th miles and 1 mile of a person's tweet location. As shown in Table 5.11 in both Atlanta and Boston, the urbanness factor has a high positive loading with access to different urban facilities. Furthermore, it showed that increasing access to diverse facilities will have a positive impact on MWS. Besides, the random forest model showed POI diversity within a mile of a person's tweet location is one of the key features predicting MWS for Atlanta, Boston and both cities overall (refer Figure 5.9, Figure 5.10, and Figure 5.11). Additionally, the topic models generated in the high and very high MWS areas showed there was a more diverse activity and an affective response. This implies that people tweeting in high MWS areas have greater access to diverse facilities, and more options to engage in different urban activities. Various themes including *drinking*, *fitness regime weekend* can be assigned to these topics (refer Table 5.18). In low MWS areas, the topics were less well-formed, the words defining the topics had low importance levels, and it was hard to assign any themes overall. This implies that people tweeting in low MWS areas might have fewer options to engage in diverse urban activities. As such, I can conclude that ease of access to diverse facilities increases the mental wellbeing score.

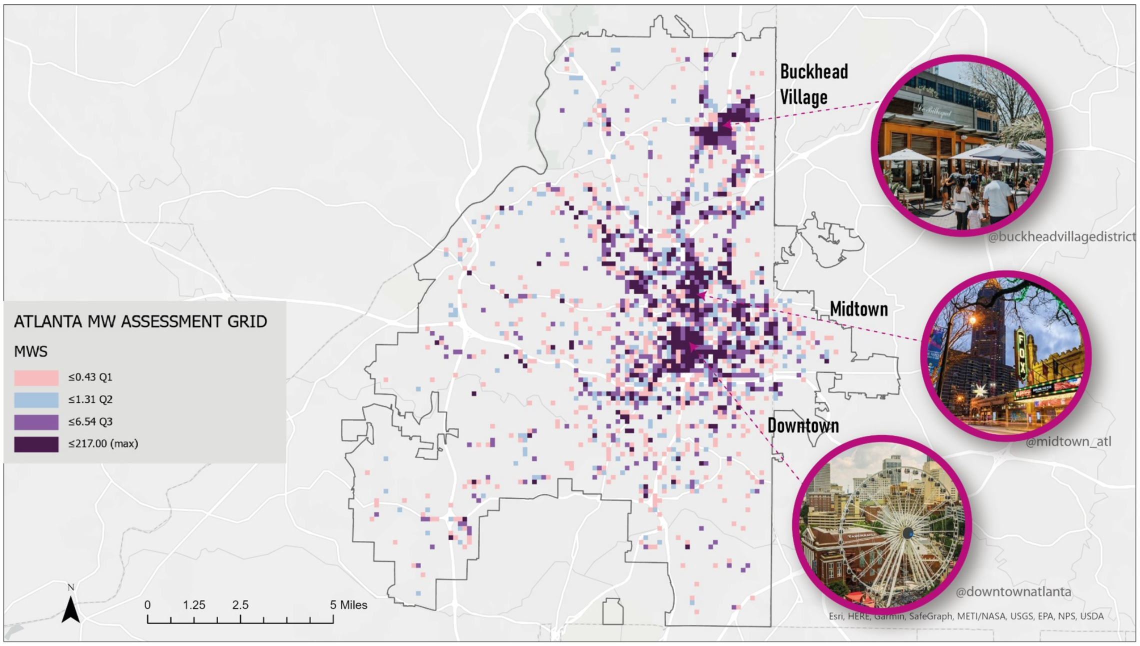

Figure 6.2: This Map shows the mental wellbeing scores for the assessment grids in Atlanta. The darkest purple shows highest mental wellbeing score, and the lowest mental wellbeing score is shown in light pink. The scores are color coded by their quartiles meaning highest scores are greater than or equal to the 75^{th} quantile and the lowest scores are less than or equal to the 25^{th} quantile.

In psychiatric geography, primarily 'escape facilities' including open spaces, parks, and trails were considered supportive of mental health. While in recent years, authors have pointed out that only green escape facilities are not enough to support mental health. Historically mixed-use is associated with higher levels of stress and anxiety. The claim is mixed-use developments cause unwanted, frequent social contacts. Spaces that would otherwise be used by neighbors only get shared with strangers, reducing ownership, and privacy [177, p. 128]. This claim needs revision. My findings show that access to diverse facilities, i.e., mixed-use areas with high POI diversity, increases MWS. Additionally, urban parks, those that were co-located with other facilities and closer to transit, showed higher MWS than stand-alone urban parks in remote locations. In recent years, public mental health literature addressed the need for diverse facilities Figure 2.3, with limited empirical evidence. Alongside mental wellbeing, mixed land-use has several advantages, including a reduction in dependence on car use, encouraging active transportation mode choice, activating deteriorating urban areas, bolstering social capital by creating social links and relationships [247]. For instance, studies show activities such as 'social drinking' has numerous de-stressing health benefits as opposed when drinking to alleviate stress and loneliness. People who have access to 'local' pubs they visit regularly and tend to feel more socially engaged [248]. As urban designers and planners, we need to be cautious about the pros and cons of proposing certain uses/programs. Certain uses like convenience stores and fast-food restaurants were perceived to be associated with higher crimes [249]. As such, neighborhood characteristics and residents' sentiment should be considered before laying out use/program specifications.

H3. People are less stressed in active high-density areas with high symbolic value.

Symbolic value is a non-monetary value that people often attach to places. This can be due to the presence of historical landmarks, scenic beauty, or aesthetic characteristics of the surrounding built environment. As such, rich and posh neighborhoods often enjoy a high symbolic value [155]. Places with urban characteristics or a higher degree of urbanness having access to historic districts and landmarks is one of the features conducive to mental health. Much of the symbolism is also attached to the cultural nuances of a city. For instance, in Atlanta, a city known for its hip-hop music, few low-income neighborhoods may take on a high symbolic value when they are mentioned in songs. There is a discussion of this - mapping 'the dirty south', the 'pink trap house'[2] as tourist location, Donald Glover's

[2] Atlanta-based rapper Chainz to promote his newest album, "Pretty Girls Like Trap Music", leased an ordinary house at 1530 Howell Mill Road in West Midtown and dumped a coat of paint on it that resembled Pepto Bismol. He etched the word "TRAP" in black lettering under the gable of the craftsman-style bungalow.

television show 'Atlanta', [3] etc. Some of these locations are now torn down public housing projects but celebrated alongside the musicians. Similarly, South End Boston has a rich cultural history, home to a strong art scene is one of the trendiest and vibrant places to be in Boston. Some of these places with high symbolic value may or may not have a rich community, however close to urban centers and accessible by public transportation. These are also closer to areas characterized by high population and employment density.

Then there are physical environments that adds to the symbolism of places. The *factor street safety perception* in Table 5.11 captures features including street wall continuity, street lights, high street cross-sectional proportion, access to parks, and cultural facilities (includes churches and religious places). These are not only reliable indicators of street safety [158], but they also create rich surroundings through textures and surfaces in the built environment. According to Goldhagen, the environmental stimuli reach the sensory system and thereby shape human cognition and interpretation. The cognition and interpretation act passively to shape people's response [239]. If it is a pleasant experience (stimuli), the response is de-stress, and if the experience is unpleasant, then the response is stress. The results in Table 5.14 show that at 0.75, quantile MWS increases by 5% with a unit increase in the factor of street safety perception within a mile of tweet location. Given these findings, I can conclude MWS increases in high-density areas with a higher symbolic value and heightened perception of safety. Figure 6.2 and Figure 6.3 show the spatial distribution of MWS scores in Atlanta and Boston. Additionally, the figures highlight high MWS is found in areas with the walkable street characteristics, high density areas, distinct architectural characters, and landmarks.

He parked a classic wide-body sedan in front of it, which was also covered in pink [250].

[3] Atlanta' is a television show that exhibits much of the daily life of the protagonist near downtown and the south side of Atlanta. Moreland Avenue, Little Five Points, North Avenue, Castle-berry hill Midtown, Infinite Energy Arena are some of the locations selected for shooting [251].

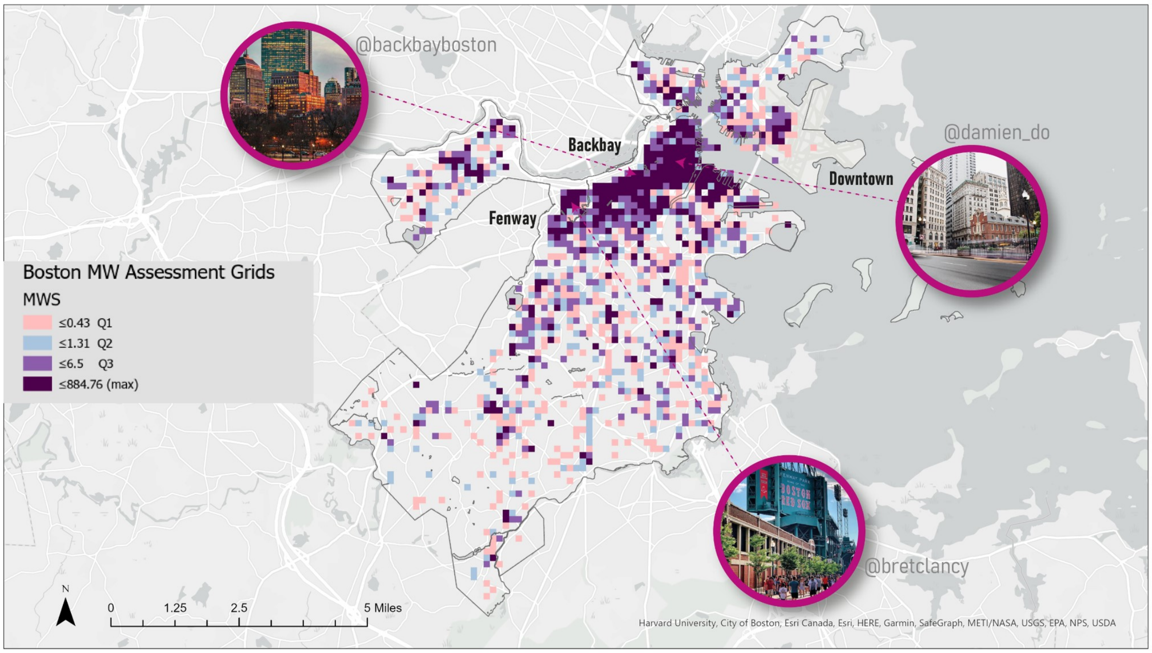

Figure 6.3: This Map shows the mental wellbeing scores for the assessment grids in Boston. The darkest purple shows highest mental wellbeing score, and the lowest mental wellbeing score is shown in light pink. The scores are color coded by their quartiles meaning highest scores are greater than or equal to the 75^{th} quantile and the lowest scores are less than or equal to the 25^{th} quantile.

Other aspects of my findings that need highlighting here are contrary to what constitutes high symbolic value in urban areas. The factors like *fear of crime and poverty*, or *socio-economic deprivation* have unique loadings like higher percentage black population, high poverty rate, high rent burden, high eviction rates, lack of land-use diversity, lack of access to health care facilities. These factors describe characteristics of an area that are an indication of socio-economic disparity. The factor *Urban rent burden* highlights characteristics of the urban areas that have good bones but have rent-burdened residents or residents living under the stress of impending evictions (refer Table 5.10). In Atlanta, the MWS at 0.75 quantile decreases by 7.29% for one unite increase in fear of crime of poverty within one mile of a tweet location, the same is true for Boston. In Boston, MWS at 0.75 quantile decreases by 4.13% with one unit increase in the factor of fear of serious crime and poverty within 1 mile of a person's tweet location. In both cities overall, we see a 2.76% decrease in the MWS at 0.75% quantile with a one-unit increase in the socio-economic factor. We see a fear of crime and poverty factor has more impact on the MWS score in Atlanta. Also, the urban rent burden within a mile of a person's tweet location is significant at 1% level in Atlanta and decreases the MWS by 2.49% at the 0.5 quantile. Additionally, the topic models show that the areas with low or very low MWS have some serious stress topics, which include words like anxiety, stress, violence, or protests. There was also frequent mention of health-related words like cancer, sick, dying, etc., but they did not form any defined topic as such. All these findings further establish that, unlike high-density affluent areas, the areas with a lack of amenities and resources have low mental wellbeing scores, and people living there are more vulnerable to mental health disorders.

City level differences: The rationale behind studying the two locations, Atlanta and Boston, was twofold: (1) increasing the generalizability of the research, that is, to see if the results are applicable for a broader population group and context and (2) to investigate if the differences in the characteristics of two cities have any impact on the mental wellbeing of people living there. The results in Table 5.14 show that MWS at all quartile levels in Atlanta is less compared to Boston, keeping all other factors and the proportion of comfortable days (in terms of weather) constant. At the topmost quartile (that is 0.75), when we consider built environment factors within 1 mile of a person's tweet location there is a 10% decrease in the MWS score of Atlanta than Boston. The decrease in MWS score is constant about 7% at all quartiles, when we consider built environment factors within 1/4 miles of a person's tweet location.

Here, it is worth considering the linear spatial lag regression results where the influence of the nearby locations' mental wellbeing score (MWS) is accounted for. The results in Table 5.19 show that the city level differences disappear when we consider built environment factors within 1mile of a person's tweet location. The results in Table 5.20 show the city level differences remain. Within quarter miles of a person's tweet location, there is a 7% decrease in MWS score of Atlanta than Boston, considering spatial lag. This could potentially indicate two things: 1) the city level differences exist when we evaluate built environment at a smaller geographic scale such as 1/4 miles and disappear at a larger geographic scale such as 1mile; 2) the city level differences are more prominent in areas with highest mental wellbeing score (above 75^{th} quartile).

The evidence that Boston has marginally higher mental welbeing can also be explained through topic model output. In Boston, more well-formed topics emerged, with equally distributed words (in terms of topic weights)describing each topic cluster. There were discrete topics for sports, work, weather, etc. I found unique and diverse words describing each topic. As reported in Table 5.17, topic 6 represented activity *sports* in Boston, explained by a set of unique words like *fenway, Redsox, season, terrible, minute, watching*. This shows people's involvement in sports events. In comparison, fewer words related to sports were found in Atlanta under topic 4 named *events*. Notably, the distribution of the keywords was uneven, which implied that the keywords describing the *events* topic were less explanatory than the ones describing the topic *sports*. Well-formed topics indicate a large group of people has been consistently discussing and describing those topics.

It is worth discussing why there may be city-level differences in the mental wellbeing score. One of the reasons is the built environment differences in both cities. Factor loading and correlation coefficients of Boston showed built environment characteristics of Boston are different from that of Atlanta, a result that was expected. For instance, a higher positive correlation is observed between built environment variables such as land-use diversity, poi diversity, building density. *Urbanness* factor has high loadings on features such as street wall continuity. As noted previously in Section 4.1, Atlanta overall is sprawled, with population density almost $1/5^{th}$ that of Boston. The average block size of Atlanta is 2.8 times that in Boston. According to the constructs of this research, the urbanness score of Boston is higher than that of Atlanta, which may explain the higher MWS in Boston.

A second reason to observe city-level differences in mental wellbeing scores is the plausible gap in the state-wide mental health support. Noteworthy to mention, Massachusetts has the highest rate of access to mental health care. In comparison, Georgia scores lower in mental healthcare access [130, 131]. We see in Table 5.10 that *fear of crime and poverty* factor has negative loading for healthcare in Atlanta, which implies that disadvantaged populations in the city have insufficient access to health care facilities in Atlanta, making them more vulnerable to stress and mental health issues.

The key takeaway in this research isn't Boston is better off than Atlanta in terms of mental wellbeing, but the key takeaway here is: *Boston is better off than Atlanta despite its built environment density, population, and employment density, as well as a high diversity of uses*. The findings follow my initial research hypothesis that increased density and diversity do not entail poor mental health, rather quite the opposite. Given people are not threatened by lack of resources, healthcare, and felonies, an urban built environment supports mental health.

1/4 mile vs 1 mile scale: The built environment measures of areas around one mile of a person's tweet location show more robust relationship with the Mental Wellbeing Score, although the built environment measure of quarter-mile around a person's tweet location adds more insight in the city level differences. Given that this research is a novel methodological contribution, I have explored the built environment features within 1/4th miles and 1 mile of tweet locations defined by the assessment grids (AG). The correlational matrix for hierarchical clustering (HC) shows moderate to low correlation between variables computed within 1/4 miles of tweet location, while there are stronger correlations between variables computed within 1 mile. In exploratory factor analysis, the explanatory variables computed within 1 mile of tweet location explained 61.42% and 64.54% variance for factors in Atlanta and Boston, respectively (refer Table 5.10). In contrast, those computed within 1/4 mile explained only 29.77% of factor variance in Atlanta - 38.61% of factor variance in Boston (refer Table 5.9). Looking at data for both cities together, we find that variables computed within 1/4 mile explain only 33% of factor variance, and those computed within

1 mile explain 43.54% factor variance. I used all independent variables, including weather conditions, time of the week, and time of the day, in the factor analysis, one of the reasons why the chosen factors explain low variance. However, looking at all the results, I could say that the 1 mile scale for individual cities showed the most robust model results. Analysis with variables computed within 1/4 mile shows a similar trend in the relationships of the variables to MWS, as it is with variables computed within 1 mile. For both Atlanta and Boston, the impact of urbanness factor within 1/4 mile of tweet location is marginally higher on mental wellbeing score at the lowest quantile. As urbanness factor within 1/4 miles of tweet location, primarily includes loadings on access to diverse facilities. This implies that if we improve access to facilities like parks, trails, grocery stores, restaurants etc., within 1/4 miles of the tweet locations with low or very low MWS, we may observe a greater increase in MWS compared to areas with high or very high MWS. Since the research was in the exploratory stage, there was value in looking at all scales and levels. Henceforth, it is established that 1 mile measure is a robust scale at which built environment variables' effect on MWS can be measured/assessed.

6.2 Research Implication in the Post COVID-19 Era

Mental health burdens are increasing every day. The assessments by WHO in the pre-COVID-19 pandemic showed one in four people get affected by mental health and neurological disorders. The psychological effects of COVID-19 are still under investigation, where experts have tagged this phase as a 'mental health tsunami'. A study by Saha et al. showed that the psychological expressions related to COVID-19 on Twitter increased by 14% (between March 2020 and May 2020). The topics discussed on social media during this period, other than treatment options and precautionary measures, were: job loss, closing of schools, feeling of loneliness, boredom, and the detriments of restricted lifestyle [252]. While the initial impact of COVID-19 might have died, it is still unclear if the new normal imposed by the restricted lifestyle will have any long-term impact. In the midst of all uncertainty, it is clear that public health interventions and policies supporting people's mental health should be prioritized. To that end, the findings in this research can contribute to developing urban planning policies supporting mental health in the following ways:

Developing Mental Wellbeing Index: In the future, the metrics identified in this study could be further tested and verified to develop the Mental Wellbeing Index nationwide. As we know, the COVID-19 pandemic has significantly reduced social interactions. Atlantic reports that some Americans are forced to work from home without necessarily wanting to.70% of working people said mixing work and home responsibilities have added additional stress and burnout [253]. To release the pressure, there were instances where people flocked to urban parks even when infections rates were high. Getting easy access to places may enable people to de-stress, a need that is more important than ever before in the post-covid world. However, there are currently no formal guidelines to index the mental wellbeing of places at a granular level. In this study, I used tweets to create mental wellbeing scores for assessment grids, essentially at locations where people tweeted. The tweet locations with high or very high mental wellbeing scores showed particular built environment characteristics like high POI diversity, access to a park, landmarks, healthcare facilities, high street cross-sectional proportions, street wall continuity, high employment density. This research also identified what could change and what could potentially improve in areas with high MWS. These include controlling crime occurrences, encouraging active transportation, and green infrastructure along the street. Areas can be compared using these metrics to assign mental wellbeing index. While each location does not need to have the same characteristic, ensuring people the bare minimum necessities for reducing their stress perception and improving their cognitive abilities can support their mental health.

Additionally, areas with low or very low mental wellbeing scores can be identified as high priority zones for addressing mental health vulnerabilities. These areas suffer from economic and racial disparity, low employment opportunities, poor access to facilities, and high crime rates. Many studies reported that the COVID-19 pandemic had exacerbated the socio-economic disparity. While the affluent group has either enjoyed or complained about stay-at-home orders, those with low incomes had little time at home and were threatened by the fear of eviction. Demarcating high-priority areas for mental health and wellbeing could help planners address mental health burdens in cities [234].

Urban Agnostic: The findings from this research help to isolate the built environment metrics that support mental health. While I argue MWS increases with the increase in the degree of urbanness, I do not confine urban within the city boundary. As Dunham-Jones et al. [171] argued that urban is defined by its physical characteristics, not by administrative boundaries. Evidently, this means that a place outside of the city boundary can have more urban characteristics, like enclosed streetscape, accessible green spaces, reduced automobile dependency in the suburbs. Currently, there are ample examples of retrofitting suburbia as well as retrofitting suburban forms and imparting urban characteristics to them. For instance, Meriden Green, a working-class city in Connecticut, executed the re-greening of a suburban-style dead retail mall into a storm water park. By removing 227 properties from the flood zone, the storm water park has triggered redevelopment that both enhances the city's urbanness while also providing the amenities of a park - and opportunities for people to socialize. Additionally, the redevelopment added one and a half miles of the walking path, pedestrian bridges, amphitheater, provisions for a farmers market, affordable housing, and enhance the existing transit center. Meriden Green has incorporated the built environment qualities described in the *urbanness* factor [254]. Williamson et al. described that people in the suburban setting lack access to meaningful human connection, not the greenery. Suburban forms designed to preserve privacy have led to an increase in people experiencing an epidemic of loneliness [255]. My findings not only substantiate this fact but also provide valid metrics to redesign and retrofit urban places that could make them conducive to mental health.

Incentives to Promote Mental Wellbeing: My research shows that mental wellbeing could be positively impacted through strategies that planners and urban designers have been preaching since the mid-1950s and 1960s. The strategies for sustainable neighborhood building, reclaiming the public realm to promote active mode choices, placemaking, and addressing social equity go hand in hand. What is essential here is re-instating which strategies support mental health and why so. Awareness of mental health is crucial in uplifting peoples' affect from a place. The awareness of policymakers, as well as the common public, is vital in understanding the needs and making investment decisions. Democratizing knowledge through policy guidelines and designing careful investment strategies can help communities. Understanding the incentive behind lowering the mental health burden is pivotal. Additionally, if urban planning and design projects addressing mental health are incentivized through funding support, such as tax increment financing (TIF), it may be easier to realize the strategies on site.

6.3 Limitations

Below I have laid out the limitation of this research in terms of data and findings:

Representativeness of Twitter Population: Twitter users are reasonably well distributed across gender, income, and education levels in the US, but most users are younger, more urban, and belong to middle-income households. The most active Twitter users are in the age group 18-29; those aged 30-49 have the most geolocated Tweets. The older (above 50) and youngest users (below 18) are underrepresented. These numbers/proportions vary across different cities and between urban and rural areas [256]. There is a positive correlation between the urban population and Twitter usage. For this work, I have restricted the analysis within city limits, as the built environment data availability is scarce outside city regions. Restricting my study area within the city boundary minimized the rural-urban location bias [257]. Furthermore, a household earning below 30K is underrepresented on Twitter [258]. Similarly, the Asian population is the least influential Twitter user. These biases in Twitter data show that the user population is not representative of the US population.

Limited Data Accessibility: The Twitter data scraped using Twitter API is just "1% of the Twitter stream" [259]. In addition, I could only use less than 1% of the sampled data that could be geolocated. Using two years' worth of data was one of the ways in which I could adjust for the loss of discarding non-geotagged tweets, but it is not enough to claim that the study represents all Twitter users or their Tweets. The other issue was the precision of the geolocation in Twitter data. Not every geolocation information included with the tweets was equally precise. Twitter allows users the option to opt for publishing precise location data. Moreover, not every tweet shared with location information has the same precision level. To resolve it, I discarded tweets below four decimal places of precision.

Non-availability of User Demographic Information: In this research, I did not use any user-level demographic information as control variables due to limitations in data availability. Moreover, if the demographic attributes for numerous Twitter users were inferred using machine learning tools, the data may suffer from limited accuracy. Readily available machine learning models (often pre-trained) to predict user information are still black-box models, often difficult to reason. Besides, limiting the research to publicly available user information would have required me to discard a large sample of Tweets. To account for this issue, I have used census data to draw socio-economic information, including relevant demographic information of the users. The assumption made here was that Twitter users are best represented by the local demographic trend.

Diagnostic claims: The mental wellbeing score generated in this research is a comparative measure of wellbeing in different areas of the city. The measure is derived from the normalized proportion of de-stress and stress tweets. It is not a metric to assess accurate measurements of mental health incidences in these areas. While through this research, I claim that areas with low MWS are vulnerable and need attention by urban planners and policy-makers (to improve mental health care), it does not project that a person living in the High MWS area needs any less mental health attention. I would be careful in making any causal claims in this research based on built environment attributes. While the relationship has been explored in past studies and research findings are robust, we still need to understand that mental health is primarily an outcome of major acute and major chronic stress in a person's lifetime. Major acute stress is derived from a person's life event, including the death of a family member, separation, or any major unpleasant event in a person's lifetime. Major chronic stress can be an outcome of a person's severe health disorder or a long-term workplace issue. While built environment factors can sometimes aggravate the mental breakdown, they can also support individuals in mental healing. As such, this research does not claim that built environment amendments can act as a cure.

Prevalence: I found the percentage of stress expression in the geolocated tweets is lower than the geolocated tweets. From my study, I found 5-7% (approximately) difference in the proportion of stress tweets. Studies show that users on Twitter exhibit signs of emotional contagion. This means people are susceptible to post content with the emotional valence they are exposed to. More importantly, people are more inclined to adopt positive valance than negative ones [260]. Another assumption is that people share geolocation when they find it relevant in sharing. For instance, geolocation is shared when someone celebrates a life event, like a birthday, meeting date, or friend. It is less common for people to share their geolocation when they tweet from their homes about non-exciting events. This causes the stress expressions to be less prevalent in the geolocated tweets.

Social Media as Stressor: Studies show social media itself acts as a source of stress for people. The recently published article by Penn Medicine reports 37% of social media users say they are worn out by the political content they see. Users feel stressed when they see views opposing their moral and political viewpoints get shared on social media. The social media-induced stress primarily ranges from misinformation about tending news to 'fear of missing out' when people witness they are not part of a group activity, or their life is less perfect than others[261]. Pew's research center reports, although the frequency of internet use and social media has no direct relationship to stress in men or women, the use of these technologies is tied to lowering their stress level. The more email exchanges they receive, the more frequently they use Twitter, the lower is their reported stress levels. The research also shows that social media users tend to perceive higher levels of social support in their network [262]. Although this is outside of the scope of my research, looking at the research findings, I can say as researchers, we need to be mindful of these facts.

6.4 Future Work

In the last two sections (Section 6.2, and Section 6.3), I have described the implication and some limitations of the current research. Drawing on those in the future, I would like to contribute in the following ways:

Changes in Post-COVID era: In the post-COVID-era, various restrictions in socializing may or may not have changed in people's behavior on space usage. Saha et al. have discovered two phases. [252], the initial phase when people showed heightened stress perception on Twitter, and the next phase when the stress perception on Twitter was stabilized with the new normal. Keeping this in mind, I would like to investigate if there is any spatial implication. My hypothesis is that in the peak stress perception period, Twitter users would be closer to their home location and tweeting relatively less from locations far away. As such, the investigation using data from the peak period in different cities may show us which locations or communities were resilient during the pandemic phase. Any difference in the observation may enlighten planners and designers in shaping places and programs suitable to combat the pandemic threat. Additionally, it would be insightful to see if people are using public spaces similarly or somewhat differently when they enter the phase of 'new normal'.

Triangulation: There are biases in social media data that can be addressed through a multiple triangulation approach. These triangulation approaches include survey designs where people can be asked to reveal their actual stress levels through popular measures of stress scales like PSS (perceives stress scales) [263]. Next, I can compare/correlate these survey feedbacks with people's home location and activities to understand if the findings using Tweets were similar or different. Given the time limitations, I could not include this in my current research agenda. This method of triangulation will not only help us compare urban areas but also rural and suburban locations where social media usage is sparse. This approach may successfully address the urban-rural divide.

Refine the MWS: Using the findings from this research and from the proposed future work, I would like to develop built environment metrics that can predict the mental wellbeing score of a place. I also plan to validate the findings through surveys and social media data. As previously noted, the built environment has only limited capacity to predict mental wellbeing. However, this metric is aimed at helping urban planners and policy-makers identify vulnerable communities.

Augmenting Twitter data: In the future, I want to address the limitation of using geolocated tweets by the use of secondary data source such as geolocated Instagram data. While tweet analysis is limited to text data, Instagram's data includes publicly available image data. Through Image classification and object detection techniques I can obtain useful cues about built environment characteristics where people tend to de-stress. Besides, I intend to add social network, groups, and more demographic info the data-set.

6.5 Conclusion

My study presents an approach for planners to utilize social media data to understand the mental health of urbanites. This research uncovers facts motivated to help experts in the public health domain by isolating the real problems previously overshadowed by the term *urban*. The real problems are poverty, rent burden, eviction threat, limited access to amenities, limited opportunity to make social connections, and fear of crime. These problems cannot be measured by density but through socio-economic and built environment indicators. Irrespective of urban or rural living conditions, the problems identified in this research can push communities to resolve/mitigate mental health risks. Knowing this may not only help urban planners and policy-makers address the mental health of urbanites but also help those living outside urban boundaries.

The research contributes methodologically in two ways. The first contribution is to unveil the potential of using point data sources to gauge mental wellbeing. This is implemented through methods commonly used in social computing to predict and unveil mental health issues. The second contribution is to use GIS-based analytical methods to compute fine-grain built environment measures within custom-defined spatial buffers to retrieve the built environment factors that explain the mental wellbeing of an area. In this research, I have used two cities, Atlanta and Boston, to compare whether differences in built environment characteristics have made a difference in the mental wellbeing score. The results show that the *urbanness* factor of cities positively impacts mental wellbeing scores.

I envision this is only the starting point, and in the future, methodologies from this research can be developed further to create area-specific guidelines and recommendations for improvements to residents' mental health.

Appendices

APPENDIX A

MENTAL WELLBEING SCORE (MWS) MAPS

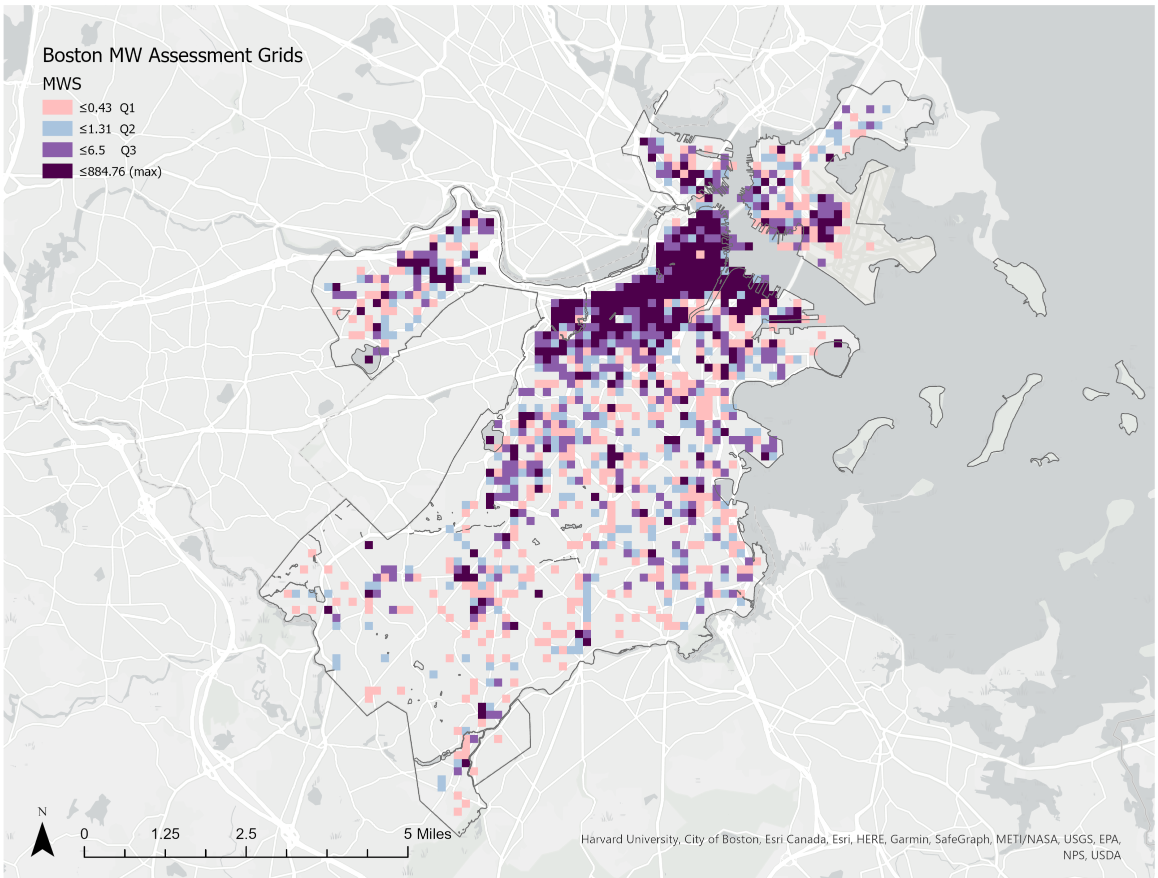

Figure A.1: The map shows mental wellbeing score (MWS) of Boston. MWS is calculated for square assessment grid of size 1/8 mile. Grids are colour coded based on MWS quartiles.

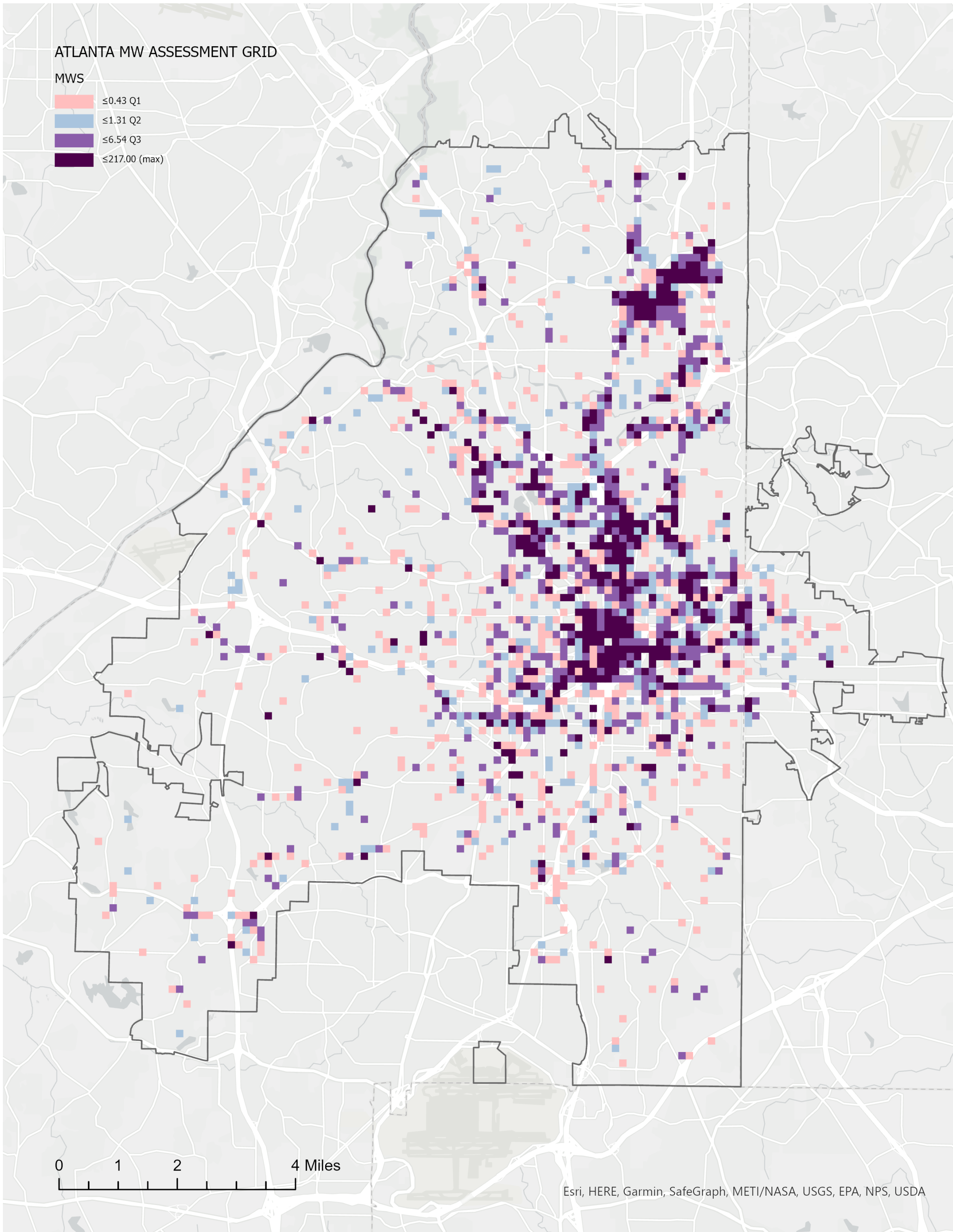

Figure A.2: The map shows mental wellbeing score (MWS) of Atlanta. MWS is calculated for square assessment grid of size 1/8 mile. Grids are colour coded based on MWS quartiles.

APPENDIX B

HIERARCHICAL CLUSTERING

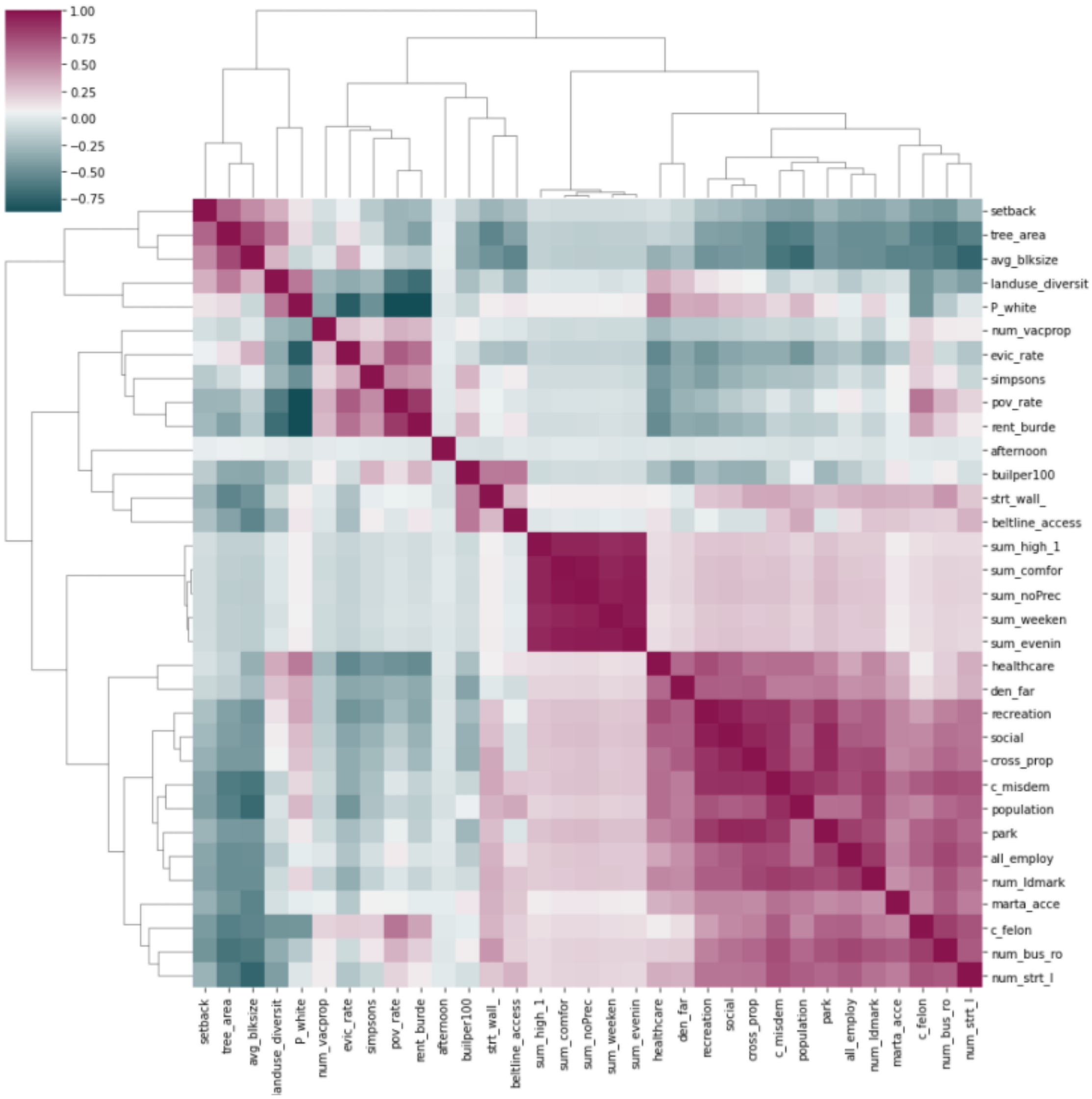

Figure B.1: Hierarchical Clustering for Atlanta, variables computed within 1/4 miles of a person's tweet location.

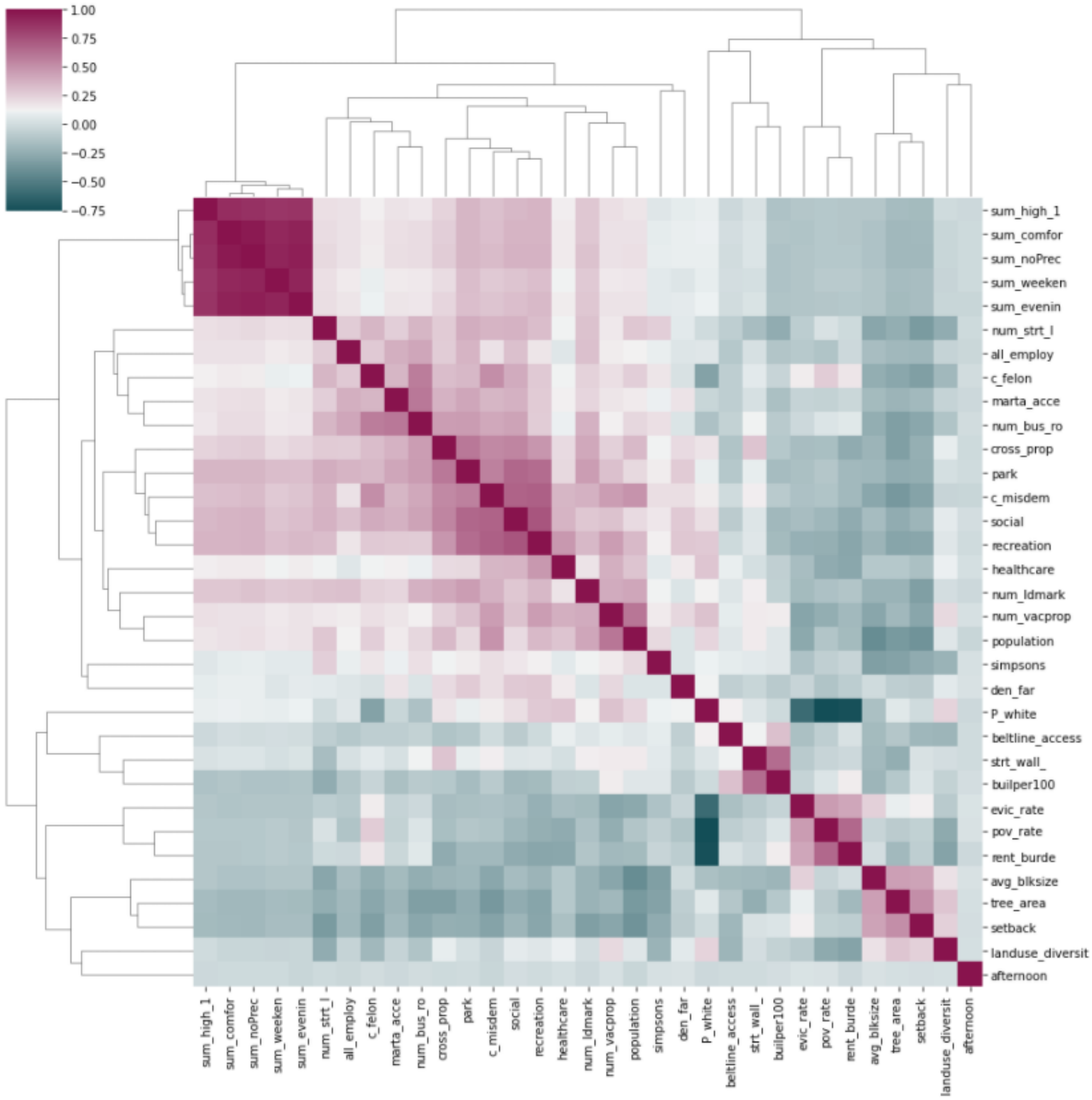

Figure B.2: Hierarchical Clustering for Atlanta, variables computed within 1 mile of a person's tweet location.

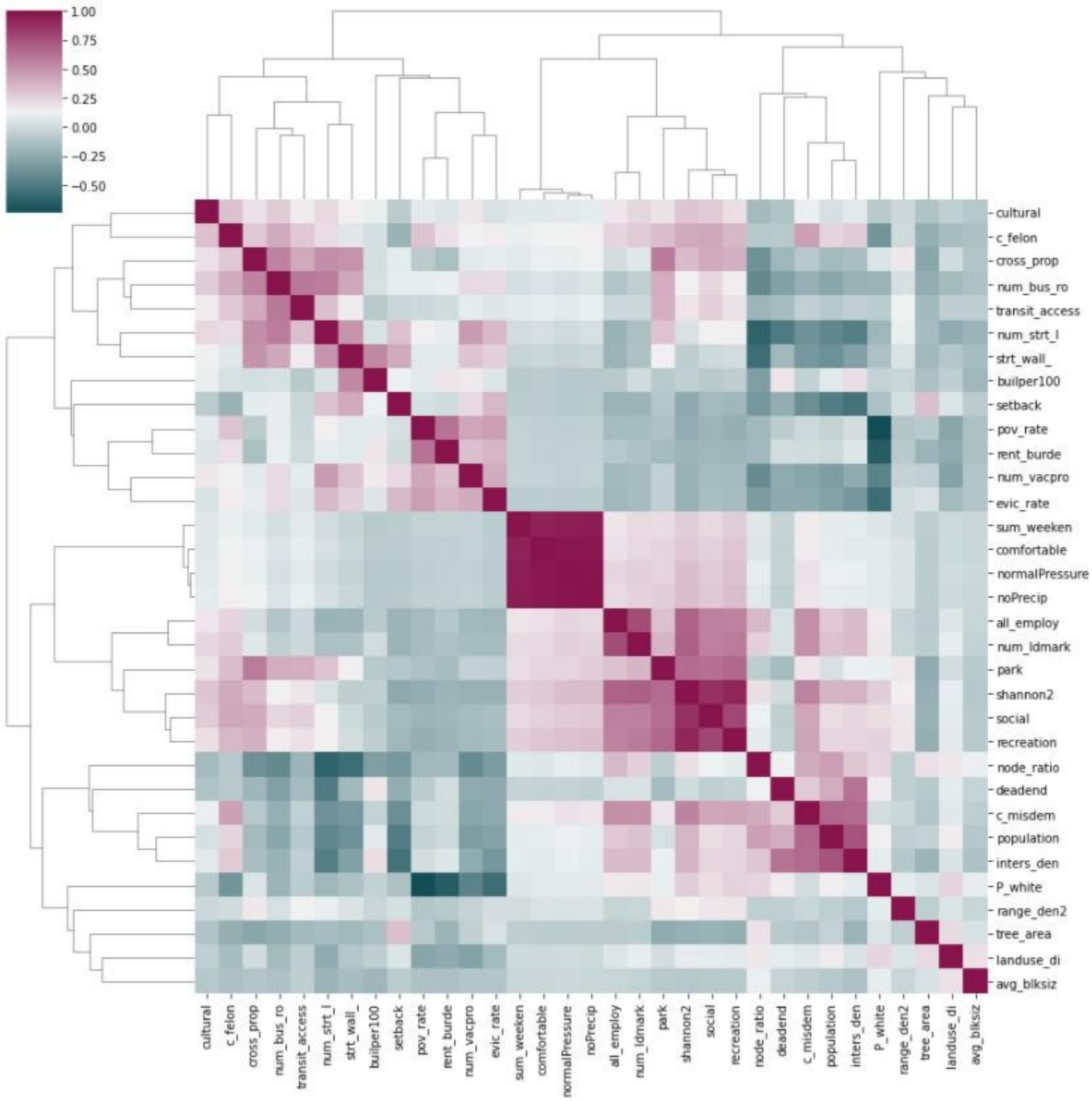

Figure B.3: Hierarchical Clustering for Boston, variables computed within 1/4 miles of a person's tweet location.

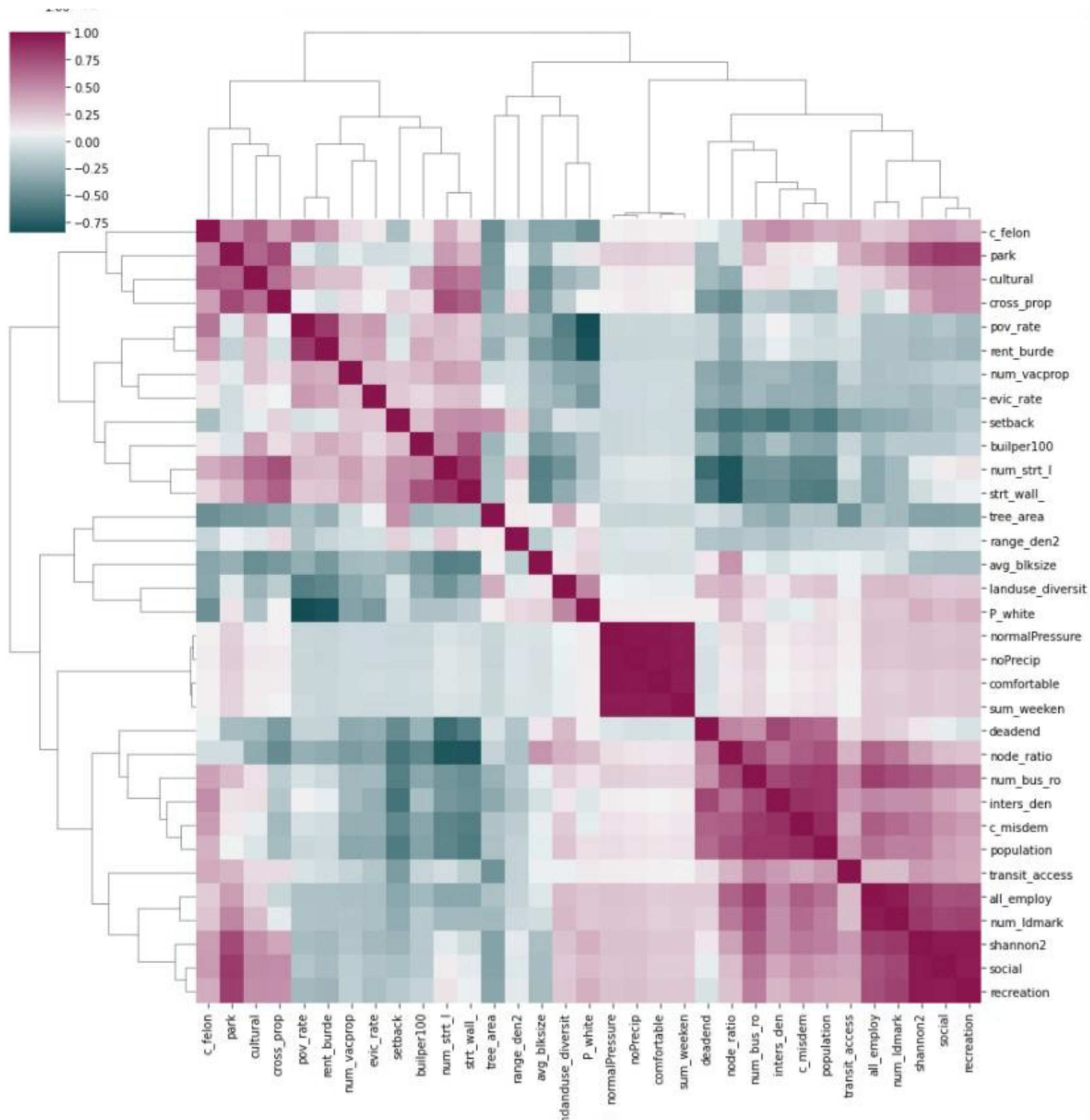

Figure B.4: Hierarchical Clustering for Boston, variables computed within 1 mile of a person's tweet location.

APPENDIX C

CORRELATION MATRIX

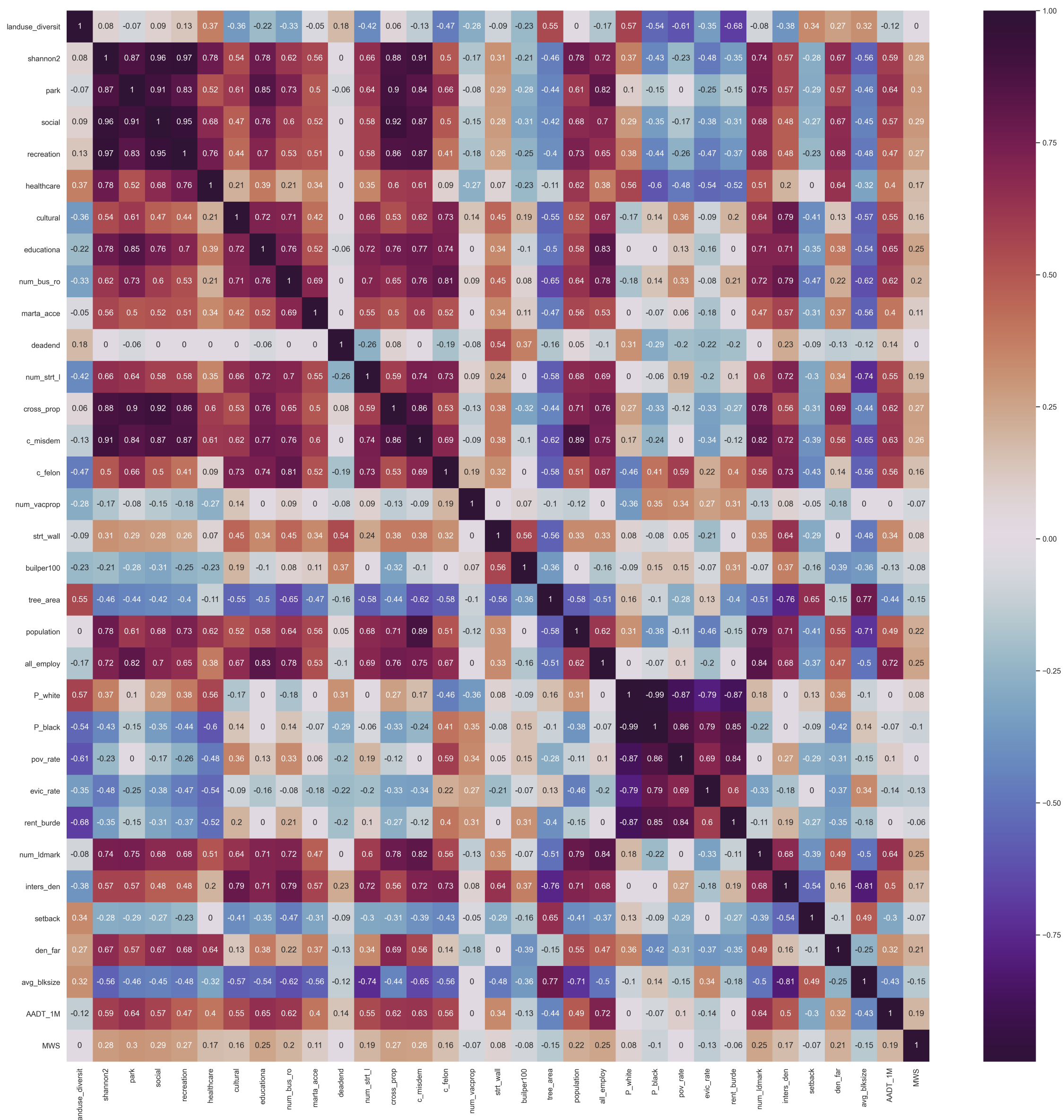

Figure C.1: Correlation matrix for Atlanta, variables computed within 1mile of a person's tweet location.

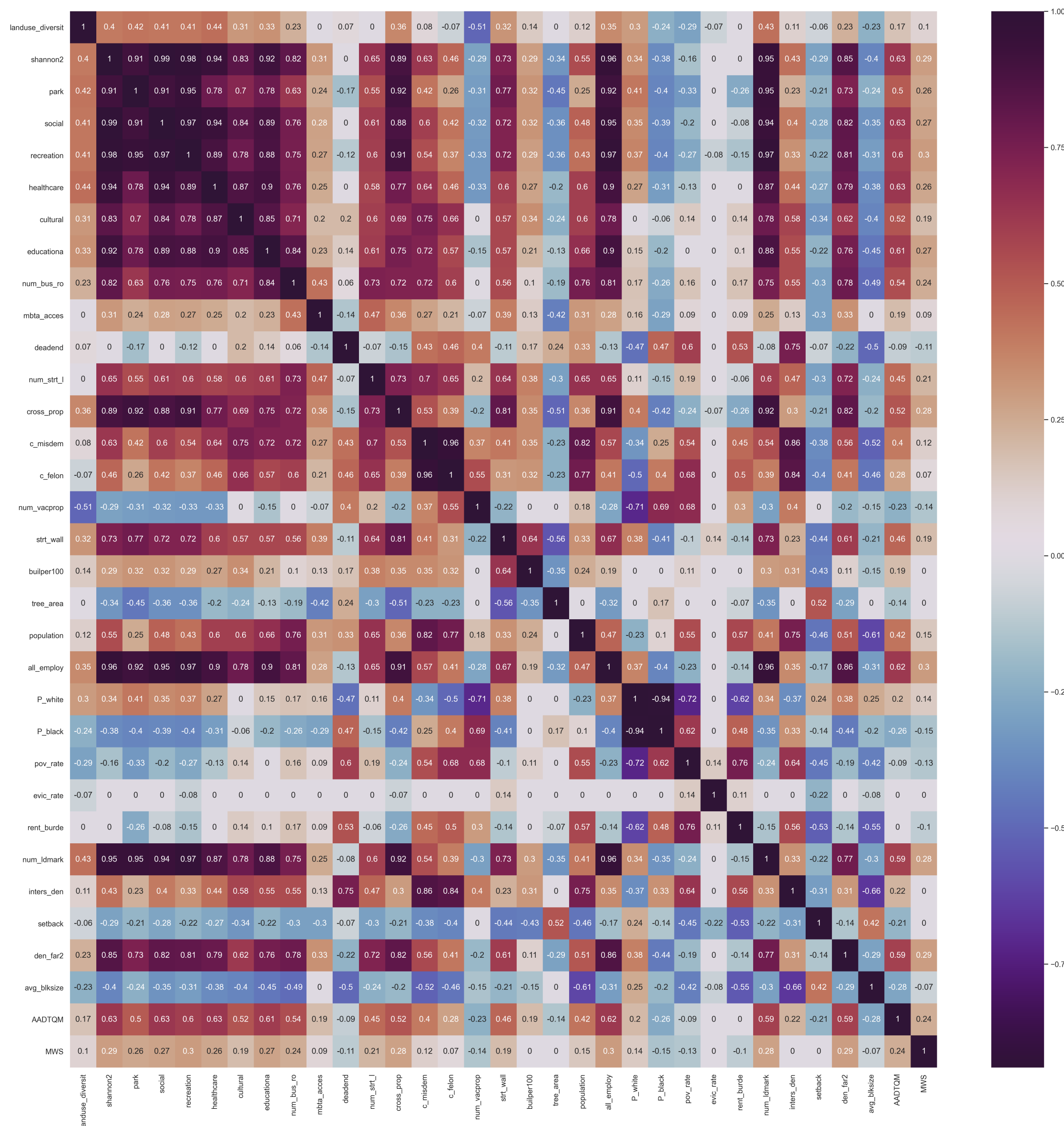

Figure C.2: Correlation matrix for Boston, variables computed within 1mile of a person's tweet location.

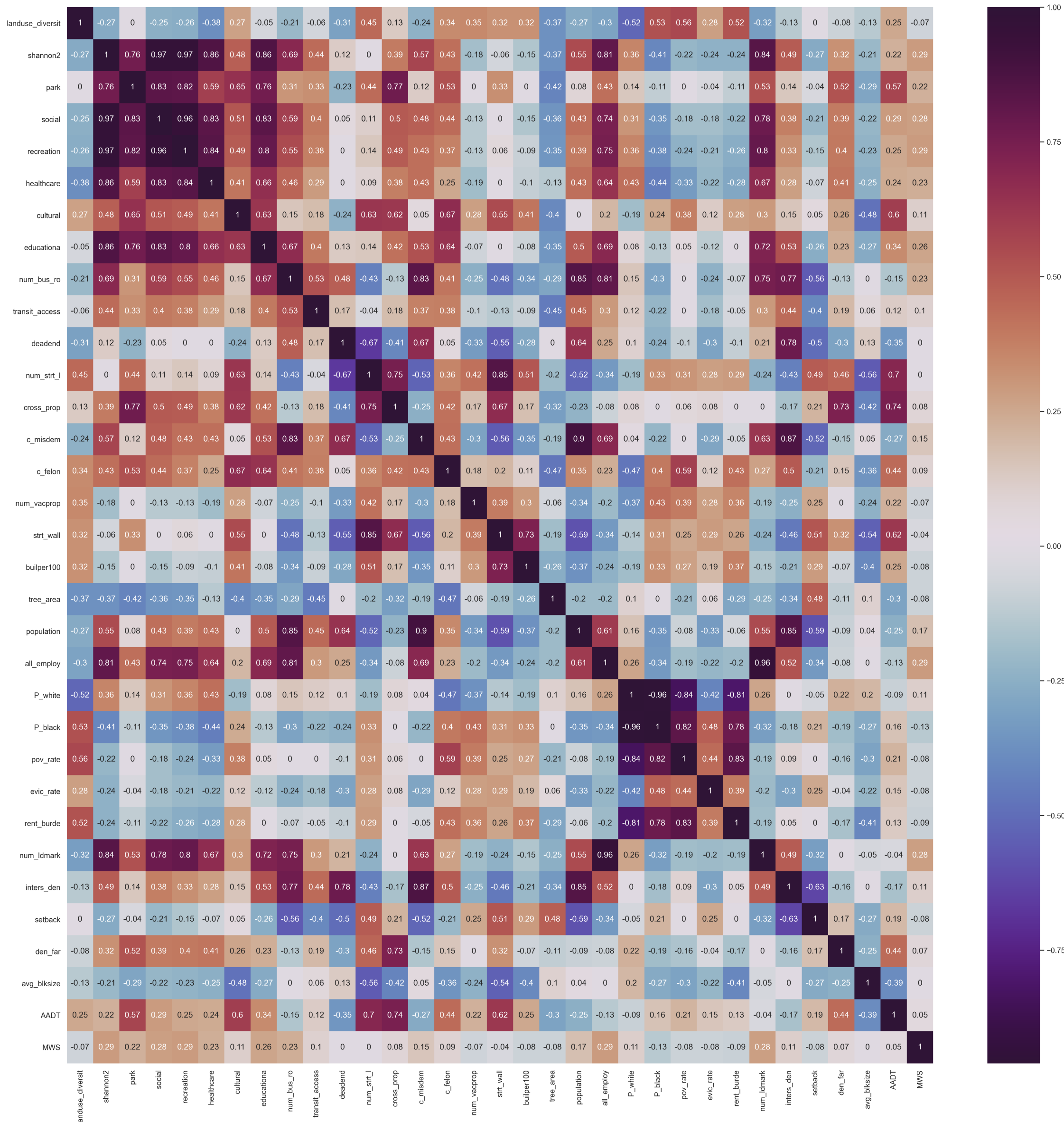

Figure C.3: Correlation matrix for both Atlanta and Boston, variables computed withing 1mile of a person's tweet location.

APPENDIX D

SPATIAL CLUSTERING

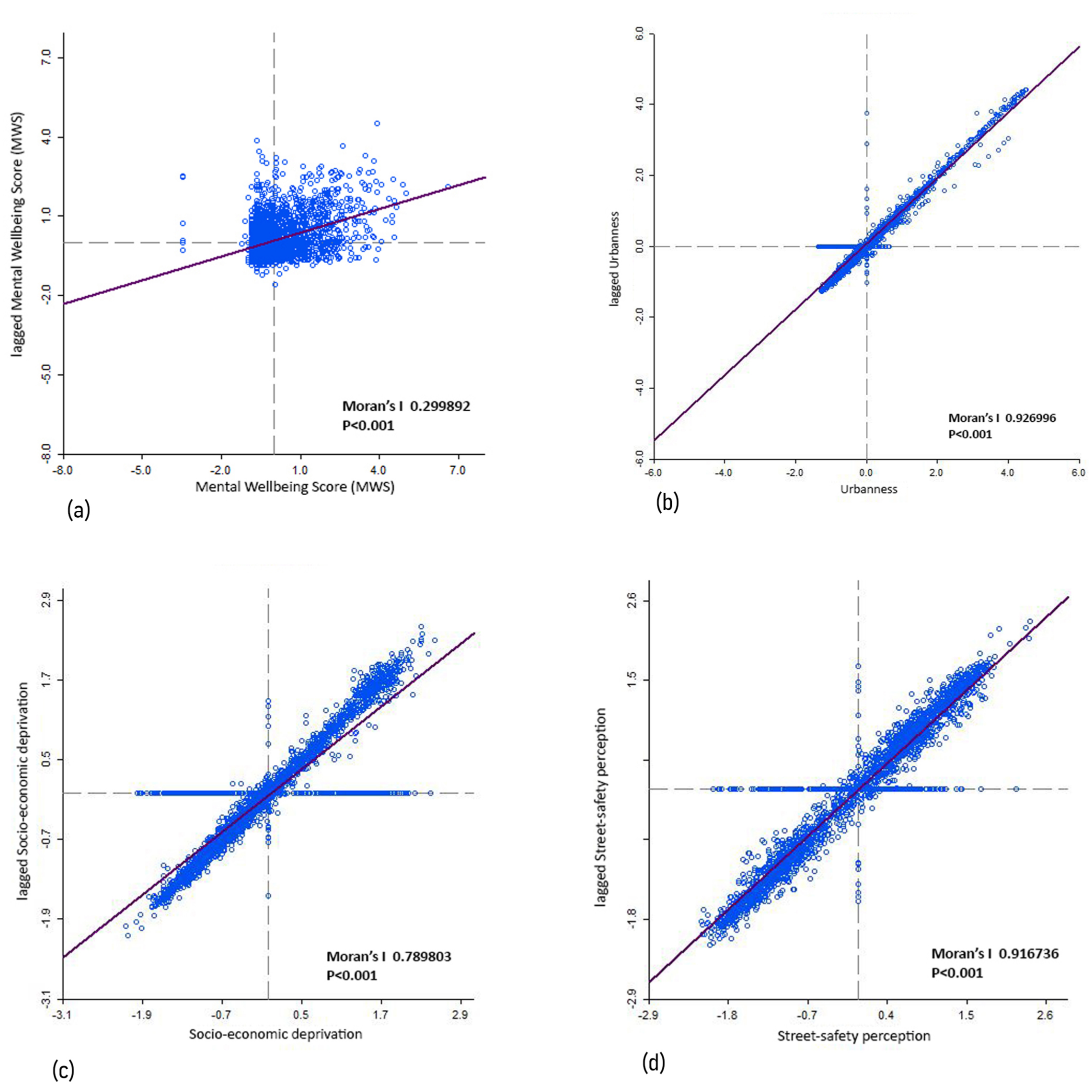

Figure D.1: The figure shows *Moran's I* values for a) mental wellbeing score (MWS); b) urbanness; c) socio-economic deprivation; and d) street-safety perception.

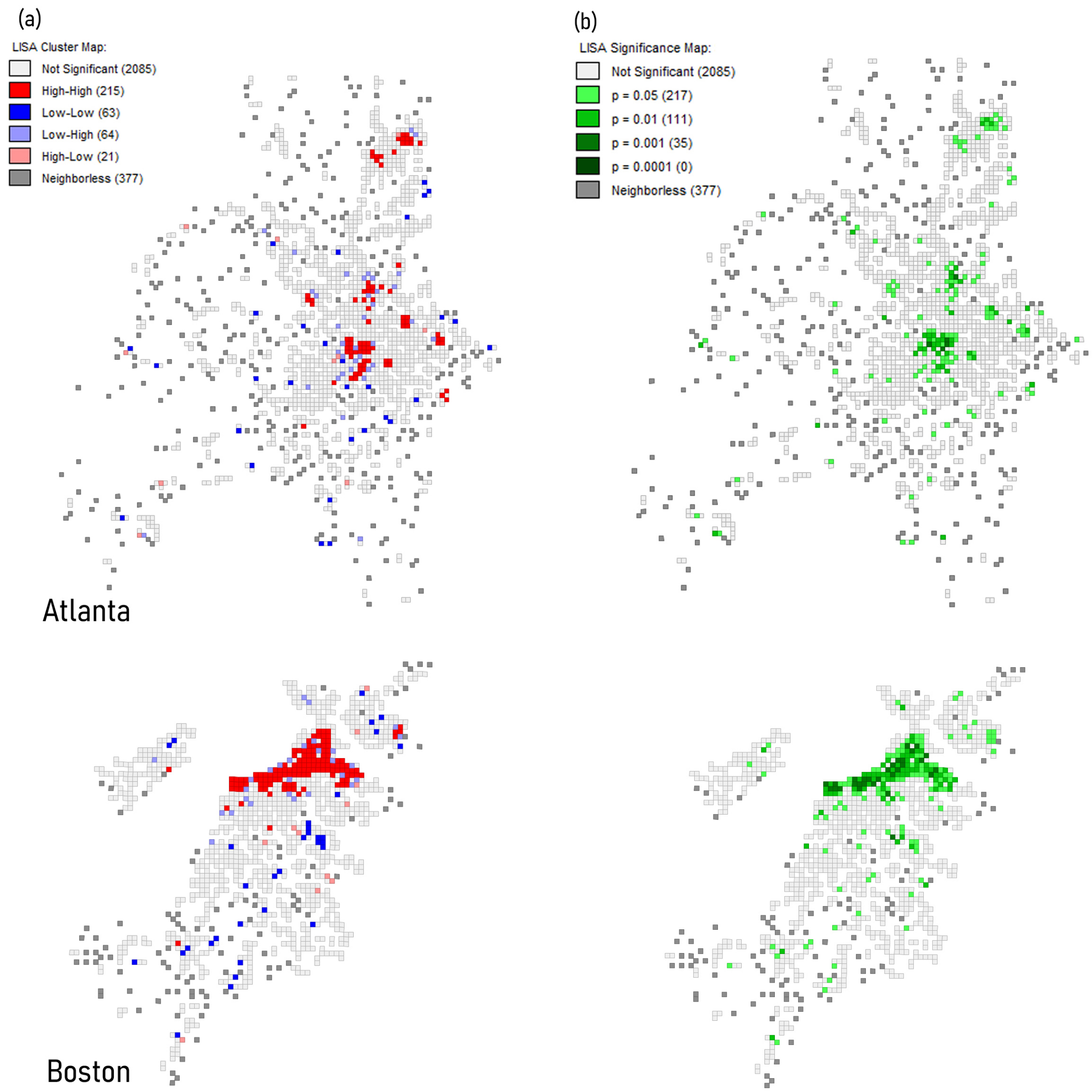

Figure D.2: The maps show local indicators of spatial clusters (LISA), and their significance . a) LISA cluster maps for mental wellbeing scores (MWS) b) right: maps show statistical significance level at which each region in Atlanta and Boston can be regarded as making a meaningful contribution to the global spatial autocorrelation outcome.

REFERENCES

[1] F. S. Dhabhar, “Effects of stress on immune function: The good, the bad, and the beautiful,” *Immunologic Research*, vol. 58, no. 2–3, pp. 193–210, 2014.

[2] *Apa survey raises concern about health impact of stress on children and families*, https://www.apa.org/news/press/releases/2010/11/stress-in-america, Accessed: 2020-07-13.

[3] C. Roehrig, “Mental disorders top the list of the most costly conditions in the united states: $201 billion$,” *Health Affairs*, vol. 35, no. 6, pp. 1130–1135, Jun. 2016.

[4] *Who — mental disorders affect one in four people*.

[5] A. Dannenberg, H. Frumkin, and R. J. Jackson, *Making healthy places : designing and building for health, well-being, and sustainability*. Washington, D.C: Island Press, 2011, ISBN: 9781597267274.

[6] S. Cohen, D. Janicki-Deverts, and G. E. Miller, “Psychological stress and disease,” *Journal of the American Medical Association*, vol. 298, no. 14, pp. 1685–1687, 2007.

[7] *Who — mental disorders affect one in four people*, https://www.who.int/whr/2001/media_centre/press_release/en/, Accessed: 2020-07-13.

[8] J. Peen, R. Schoevers, A. Beekman, and J. Dekker, “The current status of urban-rural differences in psychiatric disorders.,” *Acta Psychiatr Scand*, vol. 121, pp. 84–93, 2010.

[9] D. Halpern, *Mental Health and Built Environmet, More than Bricks and Mortar?* 1995.

[10] R. E. L. Faris and H. Dunham W., *Mental disorders in urban areas: an ecological study of schizophrenia and other psychoses*. University of Chicago Press, 1939.

[11] J. Massengale and V. Dover, *Street design : the art and practice of making complete streets*. Hoboken, New Jersey: John Wiley & Sons, 2014, ISBN: 978-1118066706.

[12] K. Lynch, “Three Cities,” in *Image of the City*, 1960, pp. 14–45.

[13] J. Jacobs, *The Death and Life of Great American Cities*. 1961.

[14] W. H.Whyte, *Social Life in Small Urban Spaces*, L. Šlusarski, Ed. Project for Public Spaces, 2001.

[15] ——, "The life of plazas," in *The Social Life of Small Urban Spaces*. 1980, pp. 16–23.

[16] J. Jacobs, "The death and life of great american cities," in 1961.

[17] J. Gehl, L. J. Kaefer, and S. Reigstad, "Close encounters with buildings," *Urban Design International*, vol. 94, no. 1, pp. 29–47, 2006.

[18] J. H. Kunstler, *The Geography of Nowhere: The Rise and Decline of America's Man-Made Landscape*. Simon Schuster, 1994.

[19] H. Frumkin, L. D. Jackson, and R. J. Frank, *Urban Sprawl and Public Health: Designing, Planning, and Building for Healthy Communities*. 2004.

[20] R. D. Putnam, *Bowling Alone: The Collapse and Revival of American Community*. Simon Schuster, 2000.

[21] V. Murthy, *Work and the loneliness epidemic*, https://hbr.org/2017/09/work-and-the-loneliness-epidemic, Accessed: 2021-08-15, 2017.

[22] J. Holt-Lunstad, "The Potential Public Health Relevance of Social Isolation and Loneliness: Prevalence, Epidemiology, and Risk Factors," *Public Policy Aging Report*, vol. 27, no. 4, pp. 127–130, Jan. 2018. eprint: https://academic.oup.com/ppar/article-pdf/27/4/127/23487069/prx030.pdf.

[23] A. N. Kluger, "Commute variability and strain," *Journal of Organizational Behavior*, vol. 19, no. 2, pp. 147–165, 1998.

[24] L. McCay, "Urban design and mental health," in *Mental Health and Illness in the City*, 9th. Springer, 2016, ch. 17.

[25] D. Halpern, "Improving mental health through the environment," in *Mental Health and Built Environment:More than Bricks and Morter*. Taylor Francis, 1995, pp. 173–203.

[26] U. N. F. for Population Activities2020, *The value of modelled population estimates for census planning and preparation*, https://www.pewresearch.org/internet/2013/12/30/social-media-update-2013/, Accessed: 2021-06-24, 2013.

[27] K. Saha and M. De Choudhury, "Modeling stress with social media around incidents of gun violence on college campuses," *Proceedings of the ACM on Human-Computer Interaction*, vol. 1, no. CSCW, 2017.

[28] D. Halpern, "Mental health and built environmet, more than bricks and mortar?" In 1995.

[29] G. Brown and T. Harris, *Social Origins of Depression: A study of psychiatric disorder in women*. Tavistock Publications Limited, 1978.

[30] A. D. Kanner, J. C. Coyne, C. Schaefer, and R. S. Lazarus, "Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events," *Journal of Behavioral Medicine*, vol. 4, no. 1, pp. 1–39, 1981.

[31] H. Selye, *The Stress of Life*. McGraw-Hill, 1978.

[32] S. Kennedy, J. K. Kiecolt-Glaser, and R. Glaser, "Social support, stress, and the immune system," in *Social support: An interactional view*, B. R. Sarason, I. G. Sarason, and G. R. Pierce, Eds., Wiley seri. John Wiley Sons., 1990, pp. 253–266.

[33] A. Baum, J. E. Singer, and C. S. Baum, "Stress and the environment," in *Environmrntal Stress*, G. Evans, Ed. Cambride University Press, 1982.

[34] J. Briere, A. Downes, and J. Spensley, "Summer in the city: Urban weather conditions and psychiatric emergency-room visits," *Journal of Abnormal Psychology*, vol. 92, no. 1, pp. 77–80, 1983.

[35] R. B. Ruback and J. Pandey, "Very hot and really crowded: Quasi-experimental investigations of indian "tempos"," *Environment and Behavior*, vol. 24, no. 4, pp. 527–554, 1992.

[36] A. Chu, A. Thorne, and H. Guite, "The impact on mental well-being of the urban and physical environment: An assessment of the evidence," *Journal of Public Mental Health*, vol. 3, no. 2, pp. 17–32, 2004.

[37] R. Schmitt, "Density, health and social disorganization," *Planners*, vol. 32, pp. 38–40, 1966.

[38] O. R. Galle, W. R. Gove, and J. M. McPherson, "Population density and pathology: What are the relations for man?" *Science*, vol. 176, no. 4030, pp. 23–30, 1972.

[39] R. E. Mitchell, "Some social implications of high density housing," *American Sociological Review*, vol. 36, no. 1, pp. 18–29, 1971.

[40] S. Milgram, "The experience of living in cities," *Science*, vol. 167, no. 3924, pp. 1461–1468, 1970.

[41] D. Stokols, "A social-psychological model of human crowding phenomena," *Journal of the American Planning Association*, vol. 38, no. 2, pp. 72–83, 1972.

[42] T. H. Holmes and R. H. Rahe, "The social readjustment rating scale," *Journal of Psychosomatic Research*, vol. 11, no. 2, pp. 213–218, 1967.

[43] S. Smith and W. W. Haythorn, “Effects of compatibility, crowding, group size, and leadership seniority on stress, anxiety, hostility, and annoyance in isolated groups,” *Journal of Personality and Social Psychology*, vol. 22, no. 1, pp. 67–79, 1972.

[44] M. Zuckerman, M. Schmitz, and A. Yosha, “Effects of crowding in a student environment,” *Journal of Applied Sicial Psycology*, vol. 7, no. 1, pp. 67–72, 1977.

[45] K. Wedmore and H. Freeman, “Social pathology and urban overgrowth,” in *Mental illness and Environment*. 1994.

[46] F. Clemente and M. Kleimen, “Fear of crime in the united states,” *Social Forces*, vol. 56, no. 2, pp. 519–531, 1977.

[47] M. Gelder, D. Gath, and R. Mayiu, *Oxford Text book of Psychiatry*. Oxford University Press, 1983.

[48] M. White, S. V. Kasl, G. E. P. Zahner, and J. C. Will, “Perceived crime in the neighborhood and mental health of women and children,” *Environment and Behavior*, vol. 19, no. 5, pp. 588–613, 1987.

[49] R. L. Lagrange and K. F. Ferraro, “Assessing age and gender differences in perceived risk and fear of crime,” *Criminology*, vol. 27, no. 4, pp. 697–720, 1989.

[50] T. L. Baumer, “Testing a general model of fear of crime: Data from a national sample,” *Journal of Research in Crime and Delinquency*, vol. 22, no. 3, pp. 239–255, 1985.

[51] S. Box, C. Hale, and G. Andrews, “Explaining fear of crime,” *The British Journal of Criminology*, vol. 28, no. 3, pp. 340–356, 1988.

[52] M. G. Maxfield, “The limits of vulnerability in explaining fear of crime: A comparative neighborhood analysis,” *Journal of Research in Crime and Delinquency*, vol. 21, no. 3, pp. 233–250, 1984.

[53] J. Birtchnell, N. Masters, and M. Deahl, “Depression and the physical environment: A study of young married women on a london housing estate,” *British Journal of Psychiatry*, vol. 153, no. 1, pp. 56–64, 1988.

[54] J. Michener, “Neighborhood disorder and local participation: Examining the political relevance of “broken windows”,” *Political Behavior*, vol. 35, no. 4, pp. 777–806, 2013.

[55] A. Hunter and T. L. Baumer, “Street traffic, social integration, and fear of crime,” *Sociological Inquiry*, vol. 52, no. 2, pp. 122–131, 1982.

[56] J. M. Darley and B. Latané, "Bystander intervention in emergencies: Diffusion of responsibility.," *Journal of personality and social psychology*, vol. 8 4, pp. 377–383, 1968.

[57] F. J. Fowler and T. W. Mangione, *Neighborhood Crime, Fear and Social Control:A Second Look at the Hartford Program*. National Institute of Justice (U.S.), 1982.

[58] A. Vrij and F. W. Winkel, "Characteristics of the built environment and fear of crime: A research note on interventions in unsafe locations," *Deviant Behavior*, vol. 12, no. 2, pp. 203–215, 1991.

[59] G. W. Evans, "The built environment and mental health," *Journal of Urban Health*, vol. 80, no. 4, pp. 536–555, 2003.

[60] G. W. Evans, N. M. Wells, and A. Moch, "Housing and mental health: A review of the evidence and a methodological and conceptual critique," *Journal of Social Issues*, vol. 59, no. 3, pp. 475–500, 2003.

[61] M. Rousseau, "'many rivers to cross': Suburban densification and the social status quo in greater lyon," *International Journal of Urban and Regional Research*, vol. 39, no. 3, pp. 622–632, 2015.

[62] G. W. Evans, N. M. Wells, H. Y. E. Chan, and H. Saltzmand, "Housing quality and mental health," *Journal of Consulting and Clinical Psychology*, vol. 68, no. 3, pp. 526–530, 2000.

[63] T. Blackman and J. Harvey, "Housing renewal and mental health: A case study," *Journal of Mental Health*, vol. 10, no. 5, pp. 571–583, 2001.

[64] A. Baum, J. Aiello, and G. Davis, "Neighborhood determinants of stress symptom perception," *American Psycological Association*, 1979.

[65] R. Gaupp, "Die klinischen besonderheiten der seelenstörungen unserer grossstadtbevölkerung," *1906*, vol. 53, pp. 1250–1252,

[66] N. Sartorius and E. Heok, *Mental Health and Illness in the City*. 2016, ISBN: 9789811023255.

[67] E. Shorter, "History of urban mental illness," in *Mental Health and Illness in the City*, N. Okkels, C. B. Kristiansen, and P. Munk-Jorgensen, Eds. Springer Singapore, 2017, pp. 1–9, ISBN: 978-981-10-0752-1.

[68] N. Sartorius, *Fighting for Mental Health: A personal view*. Cambridge University Press, 2002.

[69] J. Eskenazi, *'one false move and you're done': How us cities are changing for runaway kids*, https://www.theguardian.com/cities/2017/oct/09/one-false-move-and-youre-done-how-us-cities-are-changing-for-runaway-kids, Accessed: 2021-09-24, 2017.

[70] *2018 revision of world urbanization prospects — multimedia library - united nations department of economic and social affairs*.

[71] M. White, I. Alcock, B. Wheeler, and M. Depledge, "Would you be happier living in a greener urban area? a fixed-effects analysis of panel data.," *Psychol Sci*, vol. 24, no. 6, pp. 920–928, 2013.

[72] M. S. Taylor, B. W. Wheeler, M. P. White, T. Economou, and N. J. Osborne, "Research note: Urban street tree density and antidepressant prescription rates a cross-sectional study in london, uk," *Landscape and Urban Planning*, vol. 136, pp. 174–179, 2015, Cb8leTimes Cited:1Cited References Count:36.

[73] P. Grahn and U. A. Stigsdotter, "Landscape planning and stress," *Urban Forestry Urban Greening*, vol. 2, no. 1, pp. 1–18, 2003.

[74] A. L. Dannenberg *et al.*, "The impact of community design and land-use choices on public health: A scientific research agenda," *American Journal of Public Health*, vol. 93, no. 9, pp. 1500–1508, 2003.

[75] G. Stathopoulou, M. B. Powers, A. C. Berry, J. A. J. Smits, and M. W. Otto, "Exercise interventions for mental health: A quantitative and qualitative review," *Clinical Psychology: Science and Practice*, vol. 13, no. 2, pp. 179–193, 2006.

[76] T. Sugiyama, E. Leslie, B. Giles-Corti, and N. Owen, "Associations of neighbourhood greenness with physical and mental health: Do walking, social coherence and local social interaction explain the relationships?" *Journal of Epidemiology Community Health*, vol. 62, no. 5, e9–e9, 2008.

[77] R. S. Ulrich, "Aesthetic and affective response to natural environment," in *Behavior and the Natural Environment*, I. Altman and J. F. Wohlwill, Eds. Springer US, 1983, pp. 85–125, ISBN: 978-1-4613-3539-9.

[78] E. O. Wilson, *Biophilia*, H. U. Press, Ed. 1986.

[79] S. [Vries], S. M. E. [Dillen], P. P. Groenewegen, and P. Spreeuwenberg, "Streetscape greenery and health: Stress, social cohesion and physical activity as mediators," *Social Science Medicine*, vol. 94, pp. 26–33, 2013.

[80] S. Kaplan and C. Peterson, “Health and environment - a psychological analysis,” *Landscape and Urban Planning*, vol. 26, no. 1–4, pp. 17–23, 1993, Mj973Times Cited:8Cited References Count:35.

[81] Y. Fan, K. V. Das, and Q. Chen, “Neighborhood green, social support, physical activity, and stress: Assessing the cumulative impact,” *Health Place*, vol. 17, no. 6, pp. 1202–1211, 2011.

[82] D. Li and W. Sullivan, “Impact of views to school landscapes on recovery from stress and mental fatigue,” *Landscape and Urban Planning*, vol. 148, pp. 149–158, Apr. 2016.

[83] T. S. Nielsen and K. B. Hansen, “Do green areas affect health? results from a danish survey on the use of green areas and health indicators,” *Health Place*, vol. 13, no. 4, pp. 839–850, 2007.

[84] J. Gehl, *Life Between Buildings: Using Public Space - translated by Jo Koch.* 4. Edition, Ed. 2011, ISBN: 1610910230.

[85] M. Wittmann, *The science of time explains boredom and stress, and how to deal with them — huffpost*, 2017.

[86] S. W. Goldhagen, *Welcome to Your World: How the Built Environment Shapes Our Lives*. Harper, 2017, ISBN: 9780062199188.

[87] G. Fong, D. Frost, and S. Stansfeld, “Road rage: A psychiatric phenomenon?” *Social Psychiatry and Psychiatric Epidemiology*, vol. 36, no. 6, pp. 277–286, 2001.

[88] T. J. Christian, “Automobile commuting duration and the quantity of time spent with spouse , children , and friends,” *Preventive Medicine*, vol. 55, no. 3, pp. 215–218, 2012.

[89] C. Montgomery, *Happy City: Transforming Our Lives Through Urban Design*, First edit. Farrar, Straus and Giroux, 2013.

[90] S. Schachter and J. Singer, “Cognitive, social, and physiological determinants of emotional state,” *Psychological Review*, vol. 69, no. 5, pp. 379–399, 1962.

[91] M. De Choudhury, S. Counts, and E. Horvitz, “Predicting postpartum changes in emotion and behavior via social media,” *Conference on Human Factors in Computing Systems - Proceedings*, pp. 3267–3276, 2013.

[92] D. Kopec, “A case study of urban design and mental health in las vegas,” *Journal of Urban Design and Mental Health*, vol. 5, no. 7, 2018.

[93] R. Sturm and D. A. Cohen, "Suburban sprawl and physical and mental health," *Public Health*, vol. 118, no. 7, pp. 488–496, 2004.

[94] E. R. Pfoh *et al.*, "The sf-36 offers a strong measure of mental health symptoms in survivors of acute respiratory failure. a tri-national analysis.," *Annals of the American Thoracic Society*, vol. 13, no. 8, pp. 1343–1350, Aug. 2016.

[95] S. Cohen, "A global measure of perceived stress author (s): Sheldon cohen , tom kamarck and robin mermelstein source: Journal of health and social behavior , vol . 24 , no . 4 (dec ., 1983), pp . 385-396," *Journal of Health and Social Behavior*, vol. 24, no. 4, pp. 385–396, 2014.

[96] M. O'Dougherty, M. O. Hearst, M. Syed, M. S. Kurzer, and K. H. Schmitz, "Life events, perceived stress and depressive symptoms in a physical activity intervention with young adult women," *Bone*, vol. 23, no. 1, pp. 1–7, 2008.

[97] L. S. Radloff, "The ces-d scale: A self-report depression scale for research in the general population," *Applied Psychological Measurement*, vol. 1, no. 3, pp. 385–401, 1977.

[98] H. F. Guite, C. Clark, and G. Ackrill, "The impact of the physical and urban environment on mental well-being," *Public Health*, vol. 120, no. 12, pp. 1117–1126, 2006.

[99] S. Balani and M. De Choudhury, "Detecting and characterizing mental health related self-disclosure in social media," *Conference on Human Factors in Computing Systems - Proceedings*, vol. 18, pp. 1373–1378, 2015.

[100] E. D'Andrea, P. Ducange, B. Lazzerini, and F. Marcelloni, "Real-time detection of traffic from twitter stream analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2269–2283, Aug. 2015.

[101] X. Liu *et al.*, "Classifying urban land use by integrating remote sensing and social media data," *International Journal of Geographical Information Science*, vol. 31, no. 8, pp. 1675–1696, Aug. 2017.

[102] R. Arthur, C. A. Boulton, H. Shotton, and H. T. Williams, "Social sensing of floods in the uk," *PLoS ONE*, vol. 13, no. 1, pp. 1–18, 2018.

[103] T. Sakaki, M. Okazaki, and Y. Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 4, pp. 919–931, Apr. 2013.

[104] P. S. Earle, D. C. Bowden, and M. Guy, "Twitter earthquake detection: Earthquake monitoring in a social world," *Annals of Geophysics*, vol. 54, no. 6, pp. 708–715, 2011.

[105] F.-y. Wang, D. Zeng, C. Academy, and K. M. Carley, "Social computing: From social informatics," *IEEE Intelligent Systems*, vol. 22, no. December 2006, pp. 79–83, 2007.

[106] E. Goffman, "The presentation of self in everyday life: Introduction," in University of Edinburgh, Social Science Research, 1956.

[107] D. Ellis and J. Cromby, "Emotional inhibition: A discourse analysis of disclosure," *Psychology Health*, vol. 27, no. 5, pp. 515–532, 2012.

[108] M. Ferriter, "Computer aided interviewing and the psychiatric social history.," *Social Work and Social Sciences Review*, vol. 4, no. 3, pp. 255–263, 1993.

[109] K. Saha *et al.*, "The language of lgbtq+ minority stress experiences on social media," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, 2019.

[110] S. Bagroy, P. Kumaraguru, and M. D. Choudhury, "A social media based index of mental well-being in college campuses," in *Conference on Human Factors in Computing Systems - Proceedings*, 2017, pp. 1634–1646.

[111] S. Hasan and S. V. Ukkusuri, "Urban activity pattern classification using topic models from online geo-location data," *Transportation Research Part C: Emerging Technologies*, vol. 44, pp. 363–381, 2014.

[112] *Smartphone users worldwide 2020 — statista*.

[113] *Twitter by the numbers (2020): Stats, demographics fun facts*.

[114] B. Hecht, L. Hong, B. Suh, and E. H. Chi, "Tweets from justin bieber's heart: The dynamics of the "location" field in user profiles," *32nd International Conference on Infrared and Millimeter Waves and the 15th International Conference on Terahertz Electronics*, pp. 934–936, 2007.

[115] W. H. Chong and E. P. Lim, "Exploiting user and venue characteristics for fine-grained tweet geolocation," *ACM Transactions on Information Systems*, vol. 36, no. 3, 2018.

[116] R. A. Plunz *et al.*, "Twitter sentiment in new york city parks as measure of well-being," *Landscape and Urban Planning*, vol. 189, no. April, pp. 235–246, 2019.

[117] B. Resch, A. Summa, P. Zeile, and M. Strube, "Citizen-centric urban planning through extracting emotion information from twitter in an interdisciplinary space-time-linguistics algorithm," *Urban Planning*, vol. 1, no. 2, pp. 114–127, 2016.

[118] J. Cranshaw, J. I. Hong, and N. Sadeh, "The livehoods project: Utilizing social media to understand the dynamics of a city," *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, no. January, pp. 570–573, 2011.

[119] V. Frias-Martinez, V. Soto, H. Hohwald, and E. Frias-Martinez, "Characterizing urban landscapes using geolocated tweets," *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*, pp. 239–248, 2012.

[120] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "Exploiting semantic annotations for clustering geographic areas and users in location-based social networks," *AAAI Workshop - Technical Report*, vol. WS-11-02, pp. 32–35, 2011.

[121] Z. A. Hamstead, D. Fisher, R. T. Ilieva, S. A. Wood, T. McPhearson, and P. Kremer, "Geolocated social media as a rapid indicator of park visitation and equitable park access," *Computers, Environment and Urban Systems*, vol. 72, pp. 38–50, 2018.

[122] S. Hasan, X. Zhan, and S. V. Ukkusuri, "Understanding urban human activity and mobility patterns using large-scale location-based data from online social media," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2013.

[123] A. Abbasi, T. H. Rashidi, M. Maghrebi, and S. T. Waller, "Utilising location based social media in travel survey methods: Bringing twitter data into the play," *Proceedings of the 8th ACM SIGSPATIAL International Workshop on Location-Based Social Networks, LBSN 2015 - Held in Conjunction with ACM SIGSPATIAL 2015*, 2015.

[124] M. Saeedimoghaddam and C. Kim, "Modeling a spatio-temporal individual travel behavior using geotagged social network data: A case study of greater cincinnati," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 4, no. 4W2, pp. 207–214, 2013.

[125] E. Sternberg, *Healing spaces : the science of place and well-being*. Cambridge, Mass: Belknap Press of Harvard University Press, 2009, ISBN: 978-0674057487.

[126] T. J. . C. Edward A . Bender and M. E. . G. P. by : Wiley on behalf of the Ecological Society of America Stable URL : https://www.jstor.org/stable/1939452 REFERENCES Linked references are available on JSTOR for this article : You may

need to log in to JSTO, "Perturbation experiments in community ecology: Theory and practice," vol. 65, no. 1, pp. 1–13, 2020.

[127] L. Schweitzer, "Planning and social media: A case study of public transit and stigma on twitter," *Journal of the American Planning Association*, vol. 80, no. 3, pp. 218–238, 2014, Aw6cxTimes Cited:3Cited References Count:33.

[128] L. D. Frank, T. L. Schmid, J. F. Sallis, J. Chapman, and B. E. Saelens, "Linking objectively measured physical activity with objectively measured urban form: Findings from smartraq," *American Journal of Preventive Medicine*, vol. 28, no. 2 SUPPL. 2, pp. 117–125, 2005.

[129] J. L. Freedman, *Crowding and Behavior: The Psychology of High-Density Living*. 1975.

[130] *Mental health resources in georgia, resources to recover*, https://www.rtor.org/directory/mental-health-resources-in-georgia/, Accessed: 2020-07-13.

[131] *Mental health resources in massachusetts — resources to recover*, https://www.rtor.org/directory/mental-health-massachusetts/, Accessed: 2020-07-13.

[132] *Ranking the states 2018 — mental health america*, https://mhanational.org/issues/ranking-states-2018-0, Accessed: 2020-07-13.

[133] M. Lewyn, *Is there a perfect density? - blogs — planetizen*, https://www.planetizen.com/node/94240/there-perfect-density, Accessed: 2020-07-13, 2017.

[134] J. Gehl, L. J. Kaefer, and S. Reigstad, "Close encounters with buildings," *Urban Design International*, vol. 11, no. 1, pp. 29–47, 2006.

[135] A. N. Kluger, "Commute variability and strain," *Journal of Organizational Behavior*, vol. 19, no. 2, pp. 147–165, 1998.

[136] E. Dunham-Jones and J. Williamson, "Retrofitting suburbia: Urban design solutions for redesigning suburbs," in John Wiley Sons, Inc., 2008.

[137] R. E. Mitchell, "Some social implications of high density housing," *American Sociological Review*, vol. 36, no. 1, pp. 18–29, 1971.

[138] *Construct validity — research methods knowledge base*, https://conjointly.com/kb/construct-validity/, Accessed: 2020-07-13.

[139] P. Pongprasert and H. Kubota, "Tod residents' attitudes toward walking to transit station: A case study of transit-oriented developments (tods) in bangkok, thailand," *Journal of Modern Transportation*, vol. 27, no. 1, pp. 39–51, Mar. 2019.

[140] R. Cervero and M. Duncan, "Walking, bicycling, and urban landscapes: Evidence from the san francisco bay area," *American Journal of Public Health*, vol. 93, no. 9, pp. 1478–1483, 2003.

[141] Z. Guo and J. Ferreira, "Pedestrian environments, transit path choice, and transfer penalties: Understanding land-use impacts on transit travel," *Environment and Planning B-Planning Design*, vol. 35, no. 3, pp. 461–479, 2008, 314QBTimes Cited:7Cited References Count:36.

[142] Q. Huang and C. Yang, "Optimizing grid computing configuration and scheduling for geospatial analysis: An example with interpolating dem," *Computers Geosciences*, vol. 37, no. 2, pp. 165–176, 2011.

[143] M. Schwartz, "Human centric accessibility graph for environment analysis," *Automation in Construction*, vol. 127, p. 103 557, 2021.

[144] L. D. Frank, T. L. Schmid, J. F. Sallis, J. Chapman, and B. E. Saelens, "Linking objectively measured physical activity with objectively measured urban form: Findings from smartraq," *American Journal of Preventive Medicine*, vol. 28, no. 2 SUPPL. 2, pp. 117–125, 2005.

[145] B. B. Brown, I. Yamada, K. R. Smith, C. D. Zick, L. Kowaleski-Jones, and J. X. Fan, "Mixed land use and walkability: Variations in land use measures and relationships with bmi, overweight, and obesity," *Health and Place*, vol. 15, no. 4, pp. 1130–1141, 2009.

[146] S. Eom, T. Suzuki, and M.-H. Lee, "A land-use mix allocation model considering adjacency, intensity, and proximity," *International Journal of Geographical Information Science*, vol. 34, no. 5, pp. 899–923, 2020. eprint: https://doi.org/10.1080/13658816.2019.1687899.

[147] R. Cervero and M. Duncan, "Walking, bicycling, and urban landscapes: Evidence from the san francisco bay area," *American Journal of Public Health*, vol. 93, no. 9, pp. 1478–1483, 2003.

[148] L. D. Frank, J. F. Sallis, T. L. Conway, J. E. Chapman, B. E. Saelens, and W. Bachman, "Many pathways from land use to health: Associations between neighborhood walkability and active transportation, body mass index, and air quality," *Journal of the American Planning Association*, vol. 72, no. 1, pp. 75–87, 2006.

[149] Y. Yang, Z. Yan, J.-Y. Yeh Anthony G. O.and Xie, C.-L. Ma, and Q.-Q. Li, "Measurements of poi-based mixed use and their relationships with neighbourhood vibrancy," *International Journal of Geographical Information Science*, vol. 31, no. 4, pp. 658–675, 2017. eprint: https://doi.org/10.1080/13658816.2016.1220561.

[150] E. LOUW and F. BRUINSMA, "From mixed to multiple land use," *Journal of Housing and the Built Environment*, vol. 21, no. 1, pp. 1–13, Feb. 2006.

[151] A. Chu, A. Thorne, and H. Guite, "The impact on mental well-being of the urban and physical environment: An assessment of the evidence," *Journal of Public Mental Health*, vol. 3, no. 2, pp. 17–32, 2004.

[152] K. Manaugh and T. Kreider, "What is mixed use? presenting an interaction method for measuring land use mix," *Journal of Transport and Land Use*, vol. 6, no. 1, pp. 63–72, Apr. 2013.

[153] Y. Yue, Y. Zhuang, A. G. O. Yeh, J.-Y. Xie, C.-L. Ma, and Q.-Q. Li, "Measurements of poi-based mixed use and their relationships with neighbourhood vibrancy," *International Journal of Geographical Information Science*, vol. 31, no. 4, pp. 658–675, 2017.

[154] L. McCay, "Urban design and mental health," in *Mental Health and Illness in the City*, 9th, Springer, 2016, ch. 17.

[155] D. Halpern, "Social environmental stressors," in *Mental Health and Built Environment:More than Bricks and Morter*, Taylor Francis, 1995, pp. 70–108.

[156] F. Clemente and M. B. Kleiman, "Fear of Crime in the United States: A Multivariate Analysis*," *Social Forces*, vol. 56, no. 2, pp. 519–531, Dec. 1977. eprint: https://academic.oup.com/sf/article-pdf/56/2/519/6889264/56-2-519.pdf.

[157] J. Bannister and N. Fyfe, "Introduction: Fear and the city," *Urban Studies*, vol. 38, no. 5–6, pp. 807–813, 2001.

[158] C. Harvey, L. Aultman-Hall, S. E. Hurley, and A. Troy, "Effects of skeletal streetscape design on perceived safety," *Landscape and Urban Planning*, vol. 142, pp. 18–28, 2015.

[159] C. C. Branas, D. Rubin, and W. Guo, "Vacant properties and violence in neighborhoods," *ISRN Public Health*, vol. 2012, p. 246 142, Nov. 2012.

[160] K. Painter, "The Impact of Street Lighting on Crime, Fear, and Pedestrian Street Use," *Security J.*, vol. 5, no. 3, pp. 116–125, 1994.

[161] P. Willmott, "The evolution of a community," in Routledge Kegan Paul, 1963.

[162] C. Harvey, L. Aultman-Hall, S. E. Hurley, and A. Troy, "Effects of skeletal streetscape design on perceived safety," *Landscape and Urban Planning*, vol. 142, pp. 18–28, 2015.

[163] K. Wedmore and H. Freeman, "Social pathology and urban growth," in *Mental Illness and the Environment*, H. Freeman, Ed., London: Churchill Livingstone, 1984.

[164] *5 facts about crime in the u.s. — pew research center*, https://www.pewresearch.org/fact-tank/2019/10/17/facts-about-crime-in-the-u-s/, Accessed: 2020-07-13.

[165] *Criminal law : Classification of crimes*, https://open.lib.umn.edu/criminallaw/chapter/1-4-classification-of-crimes/, Accessed: 2021-07-11.

[166] *Lehd origin-destination employment statistics (lodes) dataset structure format version 7.5*, https://lehd.ces.census.gov/data/lodes/LODES7/LODESTechDoc7.5.pdf, Accessed: 2021-07-11.

[167] Y. Aratani, M. Chau, V. R. Wight, and S. Addy, *Rent Burden, Housing Subsidies and the Well-being of Children and Youth*, November. 2011.

[168] *The eviction lab*.

[169] J. Peen, R. Schoevers, A. Beekman, and J. Dekker, "The current status of urban-rural differences in psychiatric disorders.," *Acta Psychiatr Scand*, vol. 121, pp. 84–93, 2010.

[170] W. Airgood-Obrycki, B. Hanlon, and S. Rieger, "Delineate the u.s. suburb: An examination of how different definitions of the suburbs matter," *Journal of Urban Affairs*, vol. 00, no. 00, pp. 1–22, 2020.

[171] E. Dunham-Jones and J. Williamson, *Retrofitting suburbia: urban design solutions for redesigning suburbs*. John Wiley Sons, Inc., 2008.

[172] R. Ewing and S. Handy, "Measuring the unmeasurable: Urban design qualities related to walkability," *Journal of Urban Design*, vol. 14, no. 1, pp. 65–84, 2009.

[173] *Street network— congress for the new urbanism*, https://www.cnu.org/our-projects/street-networks/street-networks-101, Accessed: 2020-07-13.

[174] *Reinventing density: Overcoming the suburban setback*, https://theconversation.com/reinventing-density-overcoming-the-suburban-setback-66411, Accessed: 2020-06-24.

[175] *Google places api*, https://developers.google.com/maps/documentation/places/web-service/overview, Accessed: 2021-04-30, Apr. 2021.

[176] *Creating raster dems and dsms from large lidar point collections*, https://desktop.arcgis.com/en/arcmap/latest/manage-data/las-dataset/lidar-solutions-creating-

raster-dems-and-dsms-from-large-lidar-point-collections.htm, Accessed: 2021-07-11.

[177] D. Halpern, “Classic environmental stressors,” in *Mental Health and Built Environment:More than Bricks and Morter*, Taylor Francis, 1995, pp. 70–108.

[178] M. J. Menne *et al.*, *Global historical climatology network - daily (ghcn-daily), version 3*, 2012.

[179] *Hpms public release of geospatial data in shapefile format - policy — federal highway administration*.

[180] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, “Predicting depression via social media,” *Icwsm*, vol. 2, pp. 1–10, 2013.

[181] H. Lin *et al.*, “User-level psychological stress detection from social media using deep neural network,” *MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia*, pp. 507–516, 2014.

[182] T. H. McCormick, H. Lee, N. Cesare, A. Shojaie, and E. S. Spiro, “Using twitter for demographic and social science research: Tools for data collection and processing,” *Sociological Methods & Research*, vol. 46, no. 3, pp. 390–421, 2017, PMID: 29033471.

[183] *Leading countries based on number of twitter users as of january 2021*, https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/, Accessed: 2020-06-24.

[184] M. Duggan and A. Smith, *Social media update 2013*, https://www.pewresearch.org/internet/2013/12/30/social-media-update-2013/, Accessed: 2020-06-24, 2013.

[185] S. WOJCIK and A. HUGHES, *Sizing up twitter users*, https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/, Accessed: 2020-06-24, 2019.

[186] E. Sharma, K. Saha, S. K. Ernala, S. Ghoshal, and M. De Choudhury, “Analyzing ideological discourse on social media: A case study of the abortion debate,” in *Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas*, 2017, pp. 1–8.

[187] *Twitter 4j*, https://github.com/Twitter4J/Twitter4J, Accessed: 2020-07-13.

[188] M. Ghaffari, A. Srinivasan, and X. Liu, “High-resolution home location prediction from tweets using deep learning with dynamic structure,” *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2019*, pp. 540–542, 2019.

[189] *Twitter 4j api in java*, https://twitter4j.org/en/index.html, Accessed: 2021-04-30, Apr. 2021.

[190] L. Chi, K. H. Lim, N. Alam, and C. J. Butler, "Geolocation prediction in twitter using location indicative words and textual features," *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT'16)*, pp. 227–234, 2016.

[191] P. Singh, *Fundamentals of bag of words and tf-idf*, https://medium.com/analytics-vidhya/fundamentals-of-bag-of-words-and-tf-idf-9846d301ff22, Accessed: 2021-11-24, 2019.

[192] K. Saha and M. De Choudhury, "Modeling stress with social media around incidents of gun violence on college campuses," *Proceedings of the ACM on Human-Computer Interaction*, vol. 1, no. CSCW, 2017.

[193] S. Bagroy, P. Kumaraguru, and M. D. Choudhury, "A social media based index of mental well-being in college campuses," in *Conference on Human Factors in Computing Systems - Proceedings*, 2017, pp. 1634–1646.

[194] S. Cohen, "A global measure of perceived stress," *Journal of Health and Social Behavior*, vol. 24, no. 4, pp. 385–396, 2014.

[195] S. Doan, A. Ritchart, N. Perry, J. D. Chaparro, and M. Conway, "How do you relax when you're stressed? a content analysis and infodemiology study of stress-related tweets," *JMIR Public Health and Surveillance*, vol. 3, no. 2, e35, 2017.

[196] *Stanford corenlp – natural language software — stanford corenlp*.

[197] R. Socher *et al.*, "Recursive deep models for semantic compositionality over a sentiment treebank," *EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 1631–1642, 2013.

[198] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon *et al.*, Eds., vol. 30, Curran Associates, Inc., 2017.

[199] *Text classification pytorch*, https://github.com/prakashpandey9/Text-Classification-Pytorch, Accessed: 2021-02-25, May 2021.

[200] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2017. arXiv: 1412.6980 [cs.LG].

[201] S. Bera and V. K. Shrivastava, "Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification," *International Journal of Remote Sensing*, vol. 41, no. 7, pp. 2664–2683, 2020. eprint: https://doi.org/10.1080/01431161.2019.1694725.

[202] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, no. null, pp. 281–305, Feb. 2012.

[203] F. Murtagh and P. Contreras, "Algorithms for hierarchicalclustering: An overview," *WIREs Data Mining Knowledge Discovery*, vol. 2, pp. 86–97, 2012.

[204] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.

[205] *Seaborn: Statistical data visualization*, https://seaborn.pydata.org/index.html, Accessed: 2021-04-30, Apr. 2021.

[206] S. M. Bager, "Ridge parameter in quantile regression models. an application in biostatistics," *International Journal of Statistics and Applications*, vol. 8, no. 2, pp. 72–78, 2012.

[207] R. Cervero, *America's Suburban Centers: The Land Use -Transportation Link*. 1989.

[208] R. Ewing, R. Pendall, and D. Chen, "Measuring sprawl and its impact," *Smart Growth America*, p. 42, 2002.

[209] D. Kehler, "A gentle introduction to probability and likelihood for ecologists," *Frontiers in Ecology and the Environment*, vol. 1, no. 8, pp. 412–420, 2003.

[210] P. Kostov, "A spatial quantile regression hedonic model of agricultural land prices," *Spatial Economic Analysis*, vol. 4, no. 1, pp. 53–72, 2009. eprint: https://doi.org/10.1080/17421770802625957.

[211] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python," in *9th Python in Science Conference*, 2010.

[212] W. Koehrsen, *Random forest simple explaination*, https://williamkoehrsen.medium.com/random-forest-simple-explanation-377895a60d2d, Accessed: 2020-06-24, 2017.

[213] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[214] M. N. Wright, A. Ziegler, and I. R. König, "Do little interactions get lost in dark random forests?" *BMC bioinformatics*, vol. 17, p. 145, 2016.

[215] L. Brieman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2002.

[216] *Tree interpreter: Package for interpreting scikit-learn's decision tree and random forest predictions.* https://pypi.org/project/treeinterpreter/, Accessed: 2021-04-30, Apr. 2021.

[217] *Shap (shapley additive explanations)*, https://shap-lrjball.readthedocs.io/en/docs_update/index.html, Accessed: 2021-04-30, Apr. 2021.

[218] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods," *Journal of Language and Social Psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[219] *Topic extraction with non-negative matrix factorization and latent dirichlet allocation*, https://scikit-learn.org/stable/auto_examples/applications/plot_topics_extraction_with_nmf_lda.html, Accessed: 2021-04-30, Apr. 2021.

[220] C. Moody, "Mixing dirichlet topic models and word embeddings to make lda2vec," *ArXiv*, vol. abs/1605.02019, 2016.

[221] *Pytorch implementation of moody's lda2vec*, https://github.com/TropComplique/lda2vec-pytorch, Accessed: 2021-04-30, Apr. 2021.

[222] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer, "Psychological aspects of natural language use: Our words, our selves," *Annual Review of Psychology*, vol. 54, no. 1, pp. 547–577, 2003, PMID: 12185209. eprint: https://doi.org/10.1146/annurev.psych.54.101601.145041.

[223] M. Kutner, C. Nachtsheim, and J. Neter, "Applied linear regression models- 4th edition," in McGraw-Hill Irwin, 2004.

[224] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, "Multivariate data analysis," in Pearson Education, 2014, ISBN: 9788131715284.

[225] D. Halpern, "Improving mental health through the environment," in *Mental Health and Built Environment:More than Bricks and Morter*, Taylor Francis, 1995, pp. 173–203.

[226] S. Pauleit, "Urban street tree plantings: Identifying the key requirements," *Proceedings of the Institution of Civil Engineers - Municipal Engineer*, vol. 156, no. 1, pp. 43–50, 2003. eprint: https://doi.org/10.1680/muen.2003.156.1.43.

[227] L. Hao and D. Q. Naiman, "Quantile regression model and estimation," in *Quantile Regression*, 2007.

[228] M. Geraci and M. C. Jones, “Improved transformation-based quantile regression,” *Canadian Journal of Statistics*, vol. 43, no. 1, pp. 118–132, 2015. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cjs.11240.

[229] D. P. McMillen, *Quantile Regression for Spatial Data*, 1. Aufl., ser. SpringerBriefs in Regional Science. Berlin, Heidelberg: Springer-Verlag, 2013, ISBN: 3642318150.

[230] H. J. Miller, ““tobler’s first law and spatial analysis.”,” *Annals of the Association of American Geographers*, vol. 11, no. 2, pp. 284–289, 2004.

[231] R. Haining, “Spatial sampling,” in *International Encyclopedia of the Social Behavioral Sciences*, N. J. Smelser and P. B. Baltes, Eds., Oxford: Pergamon, 2001, pp. 14 822–14 827, ISBN: 978-0-08-043076-8.

[232] L. Auslin, *Contiguity-based spatial weights*, https://geodacenter.github.io/workbook/4a_contig_weights/lab4a.html#rook-contiguity, Accessed: 2021-09-24, 2020.

[233] P. Kostov, “Empirical likelihood estimation of the spatial quantile regression,” *Journal of Geographical Systems*, vol. 15, Jan. 2012.

[234] X. Huang, J. Lu, S. Gao, S. Wang, Z. Liu, and H. Wei, “Staying at home is a privilege: Evidence from fine-grained mobile phone location data in the united states during the covid-19 pandemic,” *Annals of the American Association of Geographers*, vol. 0, no. 0, pp. 1–20, 2021. eprint: https://doi.org/10.1080/24694452.2021.1904819.

[235] T. Hastie, R. Tibshirani, and J. Friedman, *Random forests*.

[236] *Mental health in cities*, https://apps.who.int/health-topics/urban-health/cities-spotlight/resources-for-health-in-cities, Accessed: 2021-20-11.

[237] E. Vassos, C. B. Pedersen, R. M. Murray, D. A. Collier, and C. M. Lewis, “Meta-Analysis of the Association of Urbanicity With Schizophrenia,” *Schizophrenia Bulletin*, vol. 38, no. 6, pp. 1118–1123, Oct. 2012. eprint: https://academic.oup.com/schizophreniabulletin/article-pdf/38/6/1118/16975253/sbs096.pdf.

[238] *Center for urban design and mental health*, https://www.urbandesignmentalhealth.com, Accessed: 2020-07-13.

[239] S. W. Goldhagen, “Welcome to your world: How the built environment shapes our lives,” in Harper, 2017, ISBN: 9780062199188.

[240] R. N. Robbins, T. Scott, J. A. Joska, and H. Gouse, “Impact of urbanization on cognitive disorders,” *Current opinion in psychiatry*, vol. 32, no. 3, pp. 210–217, May 2019, PMC6438716[pmcid].

[241] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? sentiment classification using machine learning techniques,” in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, ser. EMNLP ’02, USA: Association for Computational Linguistics, 2002, pp. 79–86.

[242] D. Davidov, O. Tsur, and A. Rappoport, “Enhanced sentiment learning using twitter hashtags and smileys,” in *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, ser. COLING ’10, Beijing, China: Association for Computational Linguistics, 2010, pp. 241–249.

[243] M. De, C. Michael, and G. S. Counts, *Happy, nervous or surprised? classification of human affective states in social media*, 2012.

[244] P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss, and C. M. Danforth, “Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter,” *PLOS ONE*, vol. 6, no. 12, pp. 1–1, Dec. 2011.

[245] T. Shortell, *An introduction to data analysis presentation*, http://academic.brooklyn.cuny.edu/soc/courses/712/chap18.html, Accessed: 2021-07-11.

[246] *Measuring evictions during the covid-19 crisis*, https://www.clevelandfed.org/en/newsroom-and-events/publications/community-development-briefs/db-20200717-measuring-evictions-during-the-covid-19-crisis.aspx, Accessed: 2021-07-11.

[247] N. A. Nabil and G. E. A. Eldayem, “Influence of mixed land-use on realizing the social capital,” *HBRC Journal*, vol. 11, no. 2, pp. 285–298, 2015.

[248] R. I. M. Dunbar *et al.*, “Functional benefits of (modest) alcohol consumption,” *Adaptive Human Behavior and Physiology*, vol. 3, no. 2, pp. 118–133, Jun. 2017.

[249] A. P. Askey, R. Taylor, E. Groff, and A. Fingerhut, “Fast food restaurants and convenience stores: Using sales volume to explain crime patterns in seattle,” *Crime & Delinquency*, vol. 64, no. 14, pp. 1836–1857, 2018. eprint: https://doi.org/10.1177/0011128717714792.

[250] *2 chainz brings pink trap house back to atlanta for holiday season*, https://www.ajc.com/news/world/chainz-brings-pink-trap-house-back-atlanta-for-holiday-season/efxxANsGbvjKQ3zsNIQvRL/, Accessed: 2021-11-24.

[251] *2 chainz brings pink trap house back to atlanta for holiday season*, https://www.exploregeorgia.org/things-to-do/list/top-film-locations-for-fx-networks-atlanta, Accessed: 2021-11-24.

[252] K. Saha, J. Torous, E. D. Caine, and M. De Choudhury, "Psychosocial effects of the covid-19 pandemic: Large-scale quasi-experimental study on social media," *J Med Internet Res*, vol. 22, no. 11, e22600, Nov. 2020.

[253] A. C. Brooks, *The hidden toll of remote work*, https://www.theatlantic.com/family/archive/2021/04/zoom-remote-work-loneliness-happiness/618473/, Accessed: 2021-07-11, 2021.

[254] J. Williamson and E. Dunham-Jones, "Case stidies in retrofitting suburbia," in John Wiley and Sons, Inc, 2021, ISBN: 9781119149194.

[255] J. Holt-Lunstad, T. B. Smith, M. Baker, T. Harris, and D. Stephenson, "Loneliness and social isolation as risk factors for mortality: A meta-analytic review," *Perspectives on Psychological Science*, vol. 10, no. 2, pp. 227–237, 2015.

[256] A. E. Kim, H. M. Hansen, J. Murphy, A. K. Richards, J. Duke, and J. A. Allen, "Methodological considerations in analyzing twitter data," *Journal of the National Cancer Institute - Monographs*, no. 47, pp. 140–146, 2013.

[257] I. Johnson, C. McMahon, J. Schöning, and B. Hecht, "The effect of population and 'structural' biases on social media-based algorithms - a case study in geolocation inference across the urban - rural sprectrum," *CHI*, pp. 1167–1178, 2017.

[258] M. M. Malik, H. Lamba, C. Nakos, and J. Pfeffer, "Population bias in geotagged tweets," *AAAI Workshop - Technical Report*, vol. WS-15-18, pp. 18–27, 2015.

[259] J. Pfeffer, K. Mayer, and F. Morstatter, "Tampering with twitter's sample api," *EPJ Data Science*, vol. 7, no. 1, p. 50, Dec. 2018.

[260] E. Ferrara and Z. Yang, "Measuring emotional contagion in social media," *PLOS ONE*, vol. 10, no. 11, pp. 1–14, Nov. 2015.

[261] P. Medicine, *The facebook effect: How is social media impacting your stress levels?* https://www.chestercountyhospital.org/news/health-eliving-blog/2020/march/how-is-social-media-impacting-your-stress-levels, Accessed: 2021-09-24, 2020.

[262] K. Hampton, L. Rainie, W. Lu, I. Shin, and K. Purcell, *Psychological stress and social media use*, https://www.pewresearch.org/internet/2015/01/15/psychological-stress-and-social-media-use-2/, Accessed: 2021-09-24, 2015.

[263] S. C. Guntuku, A. Buffone, K. Jaidka, J. C. Eichstaedt, and L. H. Ungar, "Understanding and measuring psychological stress using social media," *Proceedings of the 13th International Conference on Web and Social Media, ICWSM 2019*, no. Icwsm, pp. 214–225, 2019.

VITA

Florina Dutt was born and raised in Kolkata, known as the 'City of Joy' in West Bengal, India. She has lived in Pittsburgh, Philadelphia, Guangzhou, and Shanghai before moving to Atlanta to pursue a life as a researcher at the Georgia Institute of Technology. Her interest in place-making and how the built environment influenced people started taking shape when she studied architecture at Jadavpur University, Kolkata. After making a brief stop at the Center for Building Performace and Diagnostics (CBPD) at Carnegie Mellon University Pittsburgh, she began exploring the ideas of design computation at the University of Pennsylvania, where she acquired her Master's degree in Architecture. Later she worked as an architect and urban designer in Guangzhou and Shanghai, China.

She moved to Atlanta as a Ph.D aspirant in the School of City and Regional Planning at Georgia Institute of Technology. Given her experience, she always wanted to use her computation and urban design skills in her research at Georgia Tech. Her current research interest primarily focuses on assessing the influence of the built environment on people's mental health and wellbeing. In the future, she would want to incorporate her research findings and ideas in designing and developing policies and guidelines for smart and sustainable cities (in the context of mental health). Her work with the Center for Spatial Planning Analytics and Visualization (CSPAV) at Georgia Tech involves human-centered computing, applied machine learning techniques to process large-scale social media data, and GIS automation techniques to analyze fine-grain built environment data.