



---

Undergraduate Honors Theses

---

2021-03-18

## Effects of Actual and Perceived Air Pollution on U.S. Twitter Sentiment

George R. Garcia III  
*Brigham Young University*

Follow this and additional works at: [https://scholarsarchive.byu.edu/studentpub\\_uht](https://scholarsarchive.byu.edu/studentpub_uht)

---

### BYU ScholarsArchive Citation

Garcia, George R. III, "Effects of Actual and Perceived Air Pollution on U.S. Twitter Sentiment" (2021).  
*Undergraduate Honors Theses*. 184.  
[https://scholarsarchive.byu.edu/studentpub\\_uht/184](https://scholarsarchive.byu.edu/studentpub_uht/184)

This Honors Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

Honors Thesis

EFFECTS OF ACTUAL AND PERCEIVED AIR  
POLLUTION ON U.S. TWITTER SENTIMENT

by  
George R. Garcia III

Submitted to Brigham Young University in partial fulfillment  
of graduation requirements for University Honors

Economics Department  
Brigham Young University  
April 2021

Advisor: C. Arden Pope III

Honors Coordinator: John E. Stovall



## ABSTRACT

EFFECTS OF ACTUAL AND PERCEIVED AIR  
POLLUTION ON U.S. TWITTER SENTIMENT

George R. Garcia III

Economics Department

Bachelor of Science

**Objective**

This study examines the associations between actual and perceived air pollution (PM<sub>2.5</sub>, AQI, and ground visibility), weather information, and expressed sentiment via US Twitter. Heterogeneity in the associations across date and county characteristics are also explored.

**Methods**

A sentiment index was constructed using 27,827,828 geotagged U.S. tweets posted between May 31 and November 30, 2015. Associations between AQI category changes and the sentiment index were estimated using multi-cutoff regression discontinuity models. Associations between same-day and lagged PM<sub>2.5</sub>, ground visibility, and the sentiment index were estimated using weighted linear regression models. Models include weather variables and county and date fixed effects. Stratified analyses by county type (MSA, urban, rural) and date characteristics (holiday or non-holiday, weekday or weekend) were performed.

## Results

Being in the AQI category of Moderate rather than Good is estimated to predict a 1.5 percentage point decrease in the sentiment index. A 1-mile increase in ground visibility is estimated to predict roughly a 0.34 percentage point increase in the sentiment index, while increasing PM<sub>2.5</sub> is found to predict a very small increase of about 0.02 percentage points per 1  $\mu\text{g}/\text{m}^3$ . Temperature, pressure, wind speed, and precipitation were all found to significantly affect sentiment. Heterogeneous results were observed across both date and county characteristics.

## Conclusion

The findings suggest that air pollution has a short-term psychological effect on expressed sentiment via U.S. Twitter but not a physiological effect. Weather variables are also found to be significantly associated with expressed sentiment.



## ACKNOWLEDGEMENTS

I am especially grateful for Dr. Arden Pope III, who has allowed me to be his research assistant and collaborator over the last year and a half. He has helped me understand not only how to be a better researcher but a better person in general. I would like to thank Dr. Rob Reynolds for advising me on the twitter sentiment analysis and for his patience with me as I struggled to understand the difference between CPU and GPU. I would also like to thank Dr. John Stovall for the time he has dedicated as my honors coordinator.

Dr. Darren Hawkins deserves acknowledgement for instilling in me a love for social science research. Zac Pond—my friend, co-worker, collaborator, and classmate—must be acknowledged for the many hours he spent discussing this thesis with me. Lastly, I wish to acknowledge and thank my wife Helen Barton, whose love and support have made the world a brighter place.





## TABLE OF CONTENTS

Title.....	i
Abstract.....	iii
Acknowledgments.....	vi
Table of Contents.....	viii
List of Tables.....	x
List of Figures.....	xi
Introduction .....	1
Data .....	6
Models .....	13
Results .....	16
Discussion .....	26
Conclusion .....	32
References .....	34
Supplementary Material .....	38



## LIST OF TABLES

Table 1: Subsamples for multi-cutoff regression discontinuity models .....	14
Table 2: Summary Statistics by county and date characteristics .....	17
Table 3: Estimated coefficients from multi-cutoff regression discontinuity .....	19
Table 4: Estimated WLS coefficients of air pollution and weather variables .....	22
Table 5: Estimated WLS coefficients for lagged $PM_{2.5}$ .....	23
Table 6: Estimated WLS coefficients for lagged visibility .....	24

## LIST OF FIGURES

Figure 1: Diagram depicting tweet sentiment prediction model .....	7
Figure 2: Heat maps of average number of tweeters, sentiment index, PM <sub>2.5</sub> , and AQI per county.....	11
Figure 3: Time series plots depicting average sentiment index, PM <sub>2.5</sub> , and AQI per day, May 31 to November 30, 2015 .....	18
Figure 4: Scatter plots of relationship between AQI and sentiment index residuals.....	19
Figure 5: Estimated coefficients from multi-cutoff regression discontinuity at the Good to Moderate AQI threshold across date and county stratifications .....	20
Figure 6: Estimated WLS coefficients of PM <sub>2.5</sub> and weather variables across date and county stratifications .....	25

## Introduction

Over the past five decades, an ongoing body of research has documented the adverse effects of air pollution on human health outcomes such as cardiopulmonary and cardiovascular morbidity, asthma attacks, and infant mortality (1). This research has been vital in shaping our understanding of the societal costs of air pollution and in establishing public health policies throughout the United States and the world.

Without disputing the immense importance of this line of research, studies on the health effects of pollution make up only one part of a larger whole. A relatively younger strand of research suggests that air pollution's effects on individual and societal life extends beyond health factors to other factors such as cognition, wellbeing, and behavior. Specifically, pollution has been shown to predict decreases in happiness, productivity, and rational decision-making; increases in suicide rates, crime rates and avoidance behavior (individuals changing their behavior to avoid the pollution, for example, by staying indoors and purchasing air masks); and more negative perceptions of the government (2). These non-health related findings are important because they allow for a more complete understanding of the societal costs of pollution, and, accordingly, are vital for the creation of more effective public policy.

As noted by Jackson Lu, one key limitation currently marks this strand of research: a lack of theory as to how pollutants affect non-health outcomes (2). Air pollutants differ in considerable ways such as size, odor, and toxicity. For example, while fine particulate matter (PM<sub>2.5</sub>, fine particulate matter of size less than 2.5 micrometers) can enter indoor spaces, Ozone (O<sub>3</sub>) breaks down rapidly indoors (3-4). Thus, the effects of PM<sub>2.5</sub> and O<sub>3</sub> on non-health outcomes may differ widely both in terms of magnitude

and mechanistic pathway. Furthermore, it is largely unknown whether it is perceived air pollution, rather than actual air pollution, that is affecting these non-health outcomes (2). Might decreases in happiness be due to the appearance of smog rather than actual exposure to air pollutants? On the other hand, might increases in avoidance behavior be caused by upticks in pollution-related illness rather than advisory warnings to remain indoors?

Disentangling the psychological effects from the physiological effects is difficult as they tend to occur simultaneously. The sensory cues (such as the appearance of smog, the issuance of air pollution reports and warnings, and the enforcement of air pollution related restrictions) that affect us psychologically coincide, albeit imperfectly, with actual levels of air pollution that can enter our bodies and affect us physiologically. Moreover, parsing out the differential effects via experiment (for example, by issuing warnings despite low air pollution levels or, conversely, not issuing warnings despite high air pollution levels) is both practically infeasible and unethical. Accordingly, we must rely on quasi-experimental methods.

As of writing, there exist relatively few studies that have used quasi-experimental methods to parse out the actual and perceived effects of air pollution on non-health outcomes (2). To consider two examples, Fehr et al. had participants in Wuhan, China write down appraisals of air pollution each day (5). The authors then estimated the effects of actual and perceived air pollution on rates of unethical work behavior by including both the appraisals and objective measurements of air pollution in their models (5). It was found that the appraisals, but not the objective measurements, significantly predicted upticks in unethical behavior (5). Neidall regressed Los Angeles Zoo and Griffith Park

Observatory attendance on an indicator for whether a smog alert was issued while controlling for local ozone levels and found that the issuance of a smog alert predicted a decrease of roughly 13 and 6 percentage point decrease in attendance, respectively (6).

These two studies, along with others, offer some evidence that perceived air pollution might matter more than actual air pollution in affecting human and social behavior (see also 7-9). But does this carry over to human happiness? This question has remained largely unanswered. Past studies have found evidence for a variety of pollutants as well as self-reported pollution having a negative effect on life satisfaction and happiness levels but have not attempted to parse out the psychological from the physiological (see 10-11 for examples).

Accordingly, this study is intended to build on past research by considering both actual and perceived effects of air pollution on happiness via the expressed sentiment of U.S. Twitter users. Using Twitter posts allows for a live, real-time update on individuals' thoughts and mood, thus mitigating ambiguity and bias that may arise from self-reported measures (12). This study is similar to the work of Zheng et al. who consider the effects of air pollution on happiness by constructing a sentiment metric based on posts from the Chinese social media platform Sina Weibo (13). They find that  $PM_{2.5}$  concentration and Air Quality Index (AQI) have a comparably negative effects on same-day sentiment, with heightened suffering on holidays and weekends (13). One goal of this study is to determine whether these effects can be observed in the United States. This study will also attempt to parse out whether the effects are more physiological or psychological.

To do this, I consider a variety of models including AQI,  $PM_{2.5}$ , ground visibility, and the predicted polar sentiment (positive or negative) of 27,827,828 U.S. Tweets. AQI

is chosen primarily for its salience as a measure of air quality in the United States. It is an index ranging from 0 to 500 constructed using the measurements of six air pollutants: O<sub>3</sub>, Carbon Monoxide (CO), Sulfur Dioxide (SO<sub>2</sub>), Nitrogen Dioxide (NO<sub>2</sub>), PM<sub>10</sub>, and PM<sub>2.5</sub> (14). In the U.S., the range of AQI is divided into six levels, each associated with a particular color: Good (0-50; Green), Moderate (51-100; Yellow), Unhealthy for Sensitive Groups (101-150; Orange), Unhealthy (151-200; Red), Very Unhealthy (201-300; Purple), and Hazardous (301-500; Maroon; 14). The Environmental Protection Agency (EPA) has required that metropolitan statistical areas (MSAs) with a population of at least 350,000 report the most current AQI level at least five days of the week, with the AQI level, category, and associated color (14). Moreover, non-MSA regions, as well as weather and map apps, often report the AQI level as a voluntary service (14-15).

According to AirNow, color is included because it “makes it easy for people to quickly determine whether air quality is reaching unhealthy levels in their communities” (16). Thus, it is reasonable to assume that when people view an AQI report, they pay more attention to the color, and accordingly the AQI category, than to the actual AQI value. Finding differences in tweet sentiment at the AQI category thresholds would suggest that the AQI category, rather than the air pollution level itself, has an effect on sentiment. This would offer some evidence for a psychological or perceived effect over a physiological or actual effect. As air quality worsens, I would expect the sentiment to decrease. Hence, my first hypothesis is that:

**H1:** Expressed sentiment will be more positive just below an AQI cutoff point than just above it.



PM<sub>2.5</sub> was chosen due primarily to its ability to enter indoor spaces and deeply penetrate the lungs, potentially leading to decreases of oxygen flow to the brain and other physiological problems (3). It is widely considered to be one of the most dangerous pollutants. Thus, there is reason to believe that increases in PM<sub>2.5</sub> exposure may have a negative physiological effect on individuals' expressed sentiment.

Of course, finding a negative effect for PM<sub>2.5</sub> does not necessarily imply that the effect is physiological, as high PM<sub>2.5</sub> might affect sentiment via non-physiological channels such as the appearance of smog. Any attempt to parse out these non-physiological effects will likely be imperfect; nevertheless, including information on ground visibility does offer a partial solution. Ground visibility (henceforth just visibility) is the distance one can see into the horizon at ground level (17). It is affected by the presence of air pollutants that scatter and absorb light, as well as by weather variables such precipitation and fog (17).

Thus, considering PM<sub>2.5</sub> in models alongside visibility might allow for an imperfect estimation of air pollution's physiological effect while controlling for its visual effect. If PM<sub>2.5</sub> has a short-term physiological effect on sentiment, it is expected that higher PM<sub>2.5</sub> would predict lower sentiment even after controlling for visibility and other weather variables. Accordingly, my second hypothesis is that:

**H2:** PM<sub>2.5</sub> will have an inverse relationship with expressed sentiment even after controlling for weather information.

Moreover, considering visibility in a model while controlling for key weather information might offer an imperfect estimation of air pollution's psychological effects via the appearance of smog. As visibility increases (i.e., smog decreases), I would expect sentiment to increase. Hence, my third hypothesis:

**H3:** Visibility will have a direct relationship with expressed sentiment even after controlling for other weather information.

In the following sections, I consider the data and analytical methods used to test these hypotheses. I then report and discuss the results.

## **Data**

### *Twitter Sentiment Data*

For this study, IDs of geotagged tweets were taken with permission from the collection “Geotagged Twitter posts from the United States: A tweet collection to investigate representativeness” (18). The collection includes the IDs of tweets posted between May 31 and November 30, 2015, with around 160,000 tweets per day. The tweet IDs represent a random sample of all geotagged tweets posted (i.e., tweets posted with public data on the posting location) in the U.S. for each day in the sampling period.

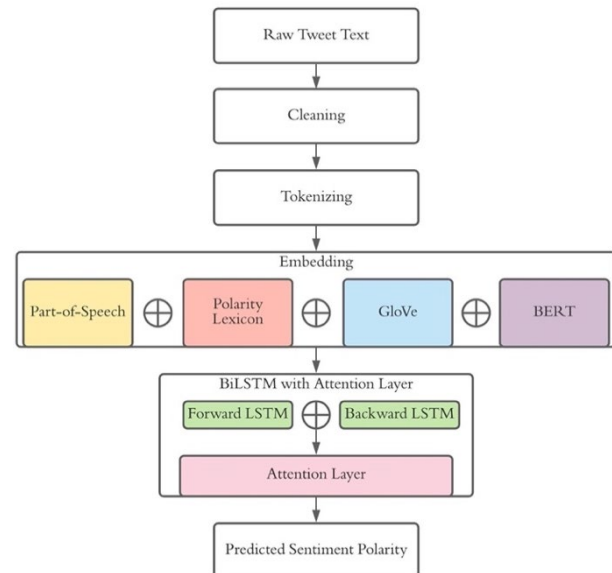
To acquire the text of each tweet, as well as information on each tweet and the tweet's poster, the IDs were “rehydrated” using the Twython package in Python (19). Tweet and user information includes the language of the tweet, date and time of when the tweet was posted, the county wherein the tweet was posted, and the ID of the poster. For

the purpose of this study, the tweet collection was limited to only tweets written in English that were not posted by users affiliated with a news outlet. A user was considered to be affiliated with a news outlet if its ID was found in the “News Outlet Tweet Ids” dataset (20). After exclusions, 27,827,828 Tweets remained. See Figure 3 below for the distribution of tweeters by county.

To predict the expressed sentiment of each tweet, this study incorporates a machine learning model similar to that proposed by Naseem et al. (21). The model begins with text preprocessing (cleaning and tokenizing), followed by text embedding, and ends with a Bidirectional Long-term Short-term Memory model with Attention Layer (BiLSTM) to obtain the predicted probability of a given tweet being categorized as either positive or negative. A graphical representation of this model can be seen in Figure 1.

Tweet text often contains a considerable amount of noise as well as a lack of structure, which can drastically decrease the accuracy of tweet sentiment classification. Accordingly, tweets were preprocessed by removing URLs, user tags, special characters, and punctuation; correcting spelling mistakes; expanding contractions (eg, “couldn’t” to “could not”); segmenting hashtags (eg,

**Figure 1:** Diagram depicting tweet sentiment prediction model



GloVe stands for the Global Vectors for Word Representation model. BERT stands for the Bidirectional Encoder Representation from Transformers model. LSTM stands for the Long Short-Term Memory model.

“#GOODJOB” to “good job”); replacing emojis and emoticons with descriptions (eg, “😊” to “smile”); and lowercasing all text. Finally, tweets were tokenized using the auto-tokenizer from the pretrained BERTweet model discussed below (22). Text cleaning was conducted using the ekphrasis and emoji packages in Python (23-24). Emoticon descriptions used to replace emoticons were adopted from Wikipedia’s List of Emoticons (25).

Preprocessed tweet text was then embedded into numerical matrices each of dimension  $i \times 978$ , where  $i$  represents the number of words (or tokens) in a given tweet, with each of the 978 columns representing a distinct piece of information about word  $i$ . The first 4 columns represent the part-of-speech of a word (noun, verb, adjective, adverb, or neither). Part-of-speech tagging was conducted using the Stanford parser in the nltk package in Python (26). The next 6 columns represent the sentiment polarity of a word according to six sentiment lexicons (SenticNet 6.0, VADER, Bing Liu Opinion Lexicon, SemEval Twitter English Lexicon, NRC Sentiment140 Lexicon, and TS-LEX; 27-32). In each of these lexicons, words and phrases are assigned sentiment values ranging from -1 to 1, with negative values and positive values representing negative and positive words, respectively. As each lexicon contains a different set of words/phrases and was created using a distinct method, combining the lexicons allows for a more complex understanding of a word’s sentiment polarity.

The remaining 968 columns represent the contextual information of a word as generated by Global Vector (GloVe) and Bidirectional Encoder Representations from Transformers (BERT) models (33-34). GloVe is an unsupervised machine learning model designed by Pennington, Socher, and Manning that offers a quantitative depiction of how

often a given word appears in context with other words (33). The GloVe model used in this study was pretrained on 2 billion tweets and generates a vector of 200 dimensions to represent the global context of a word. BERT, designed by Devlin, Chang, Lee, and Toutanova is also an unsupervised machine learning model that offers contextual information for a word (34). But unlike GloVe, which generates only a single embedding per word, BERT generates an embedding for each instance of a word, including contextual information both to the left and right (34). For example, GloVe represents the word “punch” as used in the phrase “I drank punch” and “he will punch me” using the same 200-dimensional vector. BERT, on the other hand, produces two distinct 768-dimensional vectors for the word “punch” as it is used in each of the distinct phrases. The BERT model used in this study, titled BERTweet, was pretrained on 850 million English tweets (22).

To predict the sentiment polarity, the embedded tweets were then input to a trained Bidirectional Long Short-term Memory model with an Attention layer (BiLSTM). A description of the BiLSTM model can be found elsewhere (21). The model was built and trained using the Keras API via the Tensorflow package in Python (35). See Table S1 in the supplementary material for a description of the model build.

To train the BiLSTM model, a set of labeled tweets (i.e., tweets labeled as either positive or negative) were constructed by aggregating tweets from the Stanford Twitter Sentiment Test Set, the Stanford Twitter Sentiment-Gold set, the SemEval 2013 Task 2 Training Set, and the Twitter US Airline Sentiment set (36-39). Combining these four sets allowed the model to consider a rich collection of tweets (41,984 in total) as well as many unique words (19,842 in total) in varying contexts, and to consider different

approaches to categorizing tweet sentiment polarity (the first two sets were labeled by researchers, while the latter two were labeled through the crowdsourcing platforms such as MTurk; 36-39). 5-fold cross validation was used to determine regularization parameters for the BiLSTM model. The model achieved a validation accuracy of 88.03%, which is comparable with accuracy achieved by state-of-the-art twitter sentiment classification models (21). See Table S1 in the supplementary material for the model build and a list of the hyperparameters selected through the cross validation.

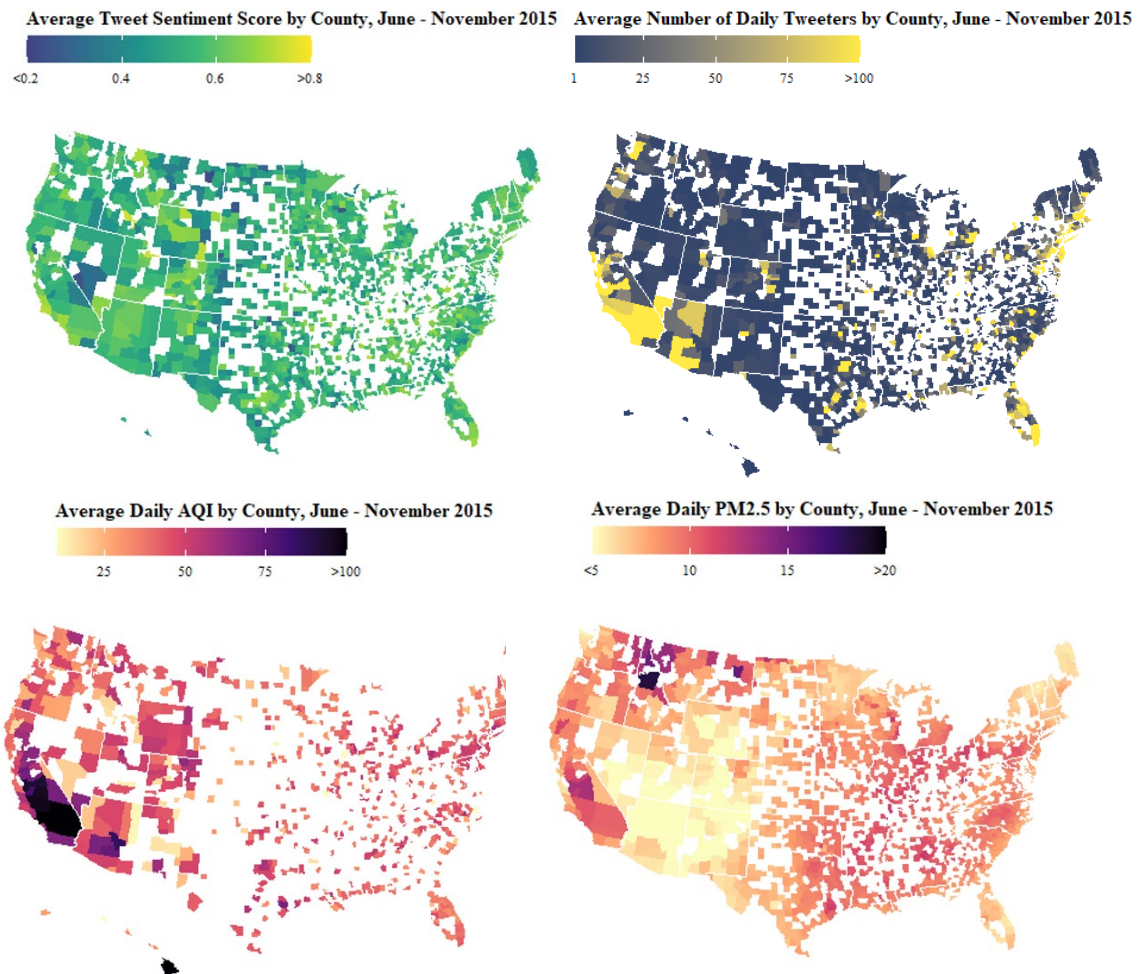
For this study, rather than classifying tweets as either positive or negative, the probability of a tweet being classified as positive is used as a continuous sentiment index, where values near 1 represent tweets more likely to be classified as positive and values near 0 represent tweets more likely to be classified as negative by the trained BiLSTM model. The sentiment index was then aggregated to the county-day level by first taking the average sentiment of each tweet per user-day, then taking the median of each user's average sentiment score per county-day. Thus, throughout this paper, the sentiment index for each county-day can be defined as the probability that a tweet posted by the median tweeter is positive. See Figures 2 and 3 below for the sentiment index by county and over time, respectively.

### *Pollution Data*

For this study, I consider AQI and PM<sub>2.5</sub> estimates at the county-day level. AQI estimates were produced by the EPA and were calculated each day for each monitor for the criteria gases of O<sub>3</sub>, CO, SO<sub>2</sub>, NO<sub>2</sub>, as well as PM<sub>10</sub> and PM<sub>2.5</sub> (40). More information on AQI can be found elsewhere (14). Daily PM<sub>2.5</sub> estimates were retrieved from the

Centers for Disease Control and Prevention website and were produced by the EPA's Downscaler model, which combines estimates from the Community Multi-scale Air Quality Model (CMAQ) with measurements from station monitors (41-42). Estimates within a county's borders were aggregated to the county level by taking the population-weighted average (42). See Figures 2 and 3 below for AQI and  $PM_{2.5}$  measurements by county and over time, respectively.

**Figure 2:** Heat maps of average number of tweeters, sentiment index,  $PM_{2.5}$ , and AQI per county



The sentiment index, at the county-day level, refers to the probability that a tweet posted by the median tweeter is positive. AQI stands for Air Quality Index.  $PM_{2.5}$  refers to fine particulate matter of size less than 2.5 micrometers and is measured in micrograms per cubic meter.

### *Weather Data*

Daily weather data used in this study was produced by the Integrated Surface Database (ISD) and was collected through the National Oceanic and Atmospheric Administration website (17). Variables include the average temperature in degrees Fahrenheit (the arithmetic mean of the maximum and minimum daily temperature), station pressure in millibars, average maintained wind speeds in miles per hour, precipitation and snow depth in inches, and visibility in miles (censored at 10 miles; 17).

The data was originally reported at the station-day level. To produce county-level estimates, an approach similar to that of Menne et al was used (43). ISD stations tend to be placed in the most heavily populated areas of the U.S.; hence, averaging the measurements of stations within or near a county should produce estimates similar to a population-weighted county average (17). To determine which stations are located within a given county, Menne et al. suggest taking the county centroid latitude and longitude and, supposing for a moment that each county is circular, determining the “radius”  $r$  of the county using its area (in square kilometers). But because most counties tend to be more rectangular-shaped than circular, in this study, I extend the radius by multiplying  $r$  by  $\sqrt{\frac{\pi}{2}}$ , which essentially extends the radius to include the corners of a square with the same area and centroid of the supposed circle. The information from all ISD stations within this extended radius of a county centroid are then averaged to produce county-day level estimates. ISDs within the bounds of a county’s extended radius were found to have similar measurements across all weather variables (see Table S2). Information on county



centroids and areas were produced by the Census Bureau and collected through the countyweather package in R (44).

#### *County and Date Characteristics*

County characteristics, used for providing summary statistics and for stratified analyses, were produced by the Department of Agriculture's Economic Research Service through information from the 2010 Census (45). Variables include racial and ethnic makeup (White non-Hispanic, Black non-Hispanic, and Hispanic), median household income, and county classification (MSA: Metropolitan Statistical Area, Urban, and Rural).

Date characteristics include whether a given date is a holiday or non-holiday, weekend or weekday. For this study, holidays include all federal holidays as well as Halloween. Weekends include Saturday and Sunday.

### **Models**

The purpose of this study is to consider the short-term effects of perceived and actual air pollution on expressed sentiment via U.S. Twitter. To do this, I consider two distinct modeling approaches: a multi-cutoff regression discontinuity and a weighted least squares (WLS) regression.

#### *Multi-Cutoff Regression Discontinuity*

To test the hypothesis that expressed sentiment will be more positive just below an AQI cutoff point compared to just above it, I consider the multi-cutoff regression

discontinuity model proposed by Cattaneo et al. (46). A multi-cutoff regression discontinuity model is, in all practical senses, similar to a standard regression discontinuity model but applied to  $j$  subpopulations of the data, where each subpopulation is limited to observations with an AQI at or near a given cutoff point  $c_j$ . Following Cattaneo et al., I split the dataset into 3 subpopulations such that observation of county  $i$  at date  $t$  is in subpopulation  $j$  if  $c_{j-1} + k_{j-1} < AQI_{i,t} < c_{j+1} - k_{j+1}$ , where  $c_0 = 0$ ,  $c_1 = 51$ ,  $c_2 = 101$  and  $c_3 = 151$ ; and  $k_0 = 0$  and  $k_1 = k_2 = k_3 = 25$  (i.e., I split the dataset at the AQI middle point between the cutoff values; See Table 1; 46). I do not consider

**Table 1:** Subsamples for multi-cutoff regression discontinuity models

AQI Category Threshold	AQI Lower Bound	AQI Cutoff	AQI Upper Bound
Good to Moderate	0	51	75
Moderate to Unhealthy for Sensitive Groups	76	101	125
Unhealthy for Sensitive Groups to Unhealthy	126	151	175

the Unhealthy to Very Unhealthy and Very Unhealthy to Hazardous AQI categories due to small n-size (only 82 county-days in this dataset have an AQI above 175).

For each subpopulation  $j$ , I then consider the standard local polynomial model:

$$\begin{aligned}
 SR_{j,i,t} = & \alpha_j + \tau_j D_{j,i,t} + \beta_{j,1}(AQI_{j,i,t} - c_j) + \beta_{j,2}D_{j,i,t}(AQI_{j,i,t} - c_j) \\
 & + \beta_{j,1}(AQI_{j,i,t} - c_j)^2 + \beta_{j,2}D_{j,i,t}(AQI_{j,i,t} - c_j)^2 + \varepsilon_{j,i,t}
 \end{aligned}$$

where  $c_j$  is the cutoff,  $D_{j,i,t}$  is an indicator variable for whether  $AQI_{j,i,t} \geq c_j$ , and  $\varepsilon_{j,i,t}$  is the error term for county  $i$ , date  $t$ . The term  $SR_{j,i,t}$  is the residual of the sentiment index regressed on weather variables and county and date fixed effects, which effectively allow the models to control for weather, time-invariant, and geographically invariant information (47). Weights for each observation are assigned using the triangle kernel, and

only observations within a bandwidth  $h_j$  are considered, where  $h_j$  is chosen to minimize the mean squared error. A polynomial of order 2 was chosen to consider non-linearity of the association without accounting for excess noise, though models with polynomials of order 1 and of order 3 are also employed to check for robustness. Throughout all models, robust standard errors are clustered at the county level. Coefficients with p-values below 0.1 are considered statistically significant.

Because the EPA only require MSAs with populations of at least 350,000 to report AQI, models are restricted to only consider observations from counties within such MSAs (14). I also consider replacing  $AQI_{j,i,t}$  with a one-day lag  $AQI_{j,i,t-1}$  in the model, as some areas reports AQI information from the previous day rather than, or along with, the current day (14). Lastly, to consider heterogeneous effects, stratified analyses are conducted by weekend or weekday, holiday or non-holiday, as well as whether the county resides in an MSA with a population of at least 350,000 or not.

### *Weighted Least Squares*

To test hypotheses 2 and 3, WLS models are conducted with sentiment index as the dependent variable. WLS models are considered over ordinary least squares (OLS) models to account for the fact that the county-day observations have an unequal number of underlying tweeters that make up the sentiment index (see Figure 2). Weighting observations according to the number of tweeters thus allows county-days with more tweeters—and hence more precise sentiment index calculations—to have more influence on the model. In this study, the weight  $w_{i,t}$  for county  $i$  date  $t$  is equal to  $\sqrt{n_{i,t}}$ , where  $n$  is

the number of unique tweeters. Unweighted OLS models are also performed to check for robustness.

To test hypothesis H2—that  $PM_{2.5}$  has an inverse relationship with sentiment—the sentiment index is regressed on  $PM_{2.5}$  with and without county and date fixed effects, with and without visibility, and with and without all other weather variables. These models are also conducted using AQI rather than  $PM_{2.5}$  to check for consistency. To consider short-term latent effects, one- and two-day lags of  $PM_{2.5}$  are considered independently as well as jointly in a distributed lag model with a 2-degree polynomial. The difference between  $PM_{2.5}$  and  $PM_{2.5}$  lagged by one day is also considered. A similar approach is taken to test H3—that visibility has a direct relationship with sentiment—but replacing  $PM_{2.5}$  with visibility. Stratified analyses are performed to test for heterogeneous effects across date type (nonholiday, holiday, weekday, and weekend) and county type (MSA, urban, or rural). For all models, robust standard errors are clustered at the county level. Coefficients with p-values below 0.1 are considered statistically significant.

## Results

In total, the sentiment information of 27,827,828 tweets posted in the United States from May 31 to November 30, 2015 were aggregated to form a sentiment index across 3,339 counties and 422,467 county-days. Excluding observations without weather information left 1,699 counties and 254,862 county-days. Of the remaining counties, 748 are contained in MSAs (all of which have populations greater 350,000), 773 are classified as urban, and 161 as rural.

**Table 2:** Summary Statistics by county and date characteristics

Characteristic	N	Mean Sentiment (SD)	Mean AQI (SD)	Mean PM <sub>2.5</sub> (SD)
All	254,862	0.56 (0.22)	41.44 (22.29)	8.52 (5.07)
<b>County Type</b>				
MSA	132,527	0.57 (0.18)	43.33 (21.48)	8.82 (4.82)
Urban	111,980	0.54 (0.26)	35.81 (23.92)	8.30 (5.18)
Rural	14,605	0.52 (0.30)	36.40 (20.87)	7.53 (6.24)
<b>Majority Race/Ethnicity</b>				
White	224,898	0.56 (0.23)	40.17 (21.04)	8.54 (5.13)
Black	8,281	0.52 (0.23)	43.75 (16.72)	9.50 (3.89)
Hispanic	10,417	0.51 (0.23)	48.83 (28.90)	7.55 (4.00)
No Majority	17,641	0.56 (0.19)	50.33 (28.93)	8.52 (4.98)
<b>Median Household Income</b>				
≤\$40,000	23,216	0.53 (0.27)	37.06 (17.73)	8.52 (4.68)
\$40,001 to \$60,000	156,766	0.55 (0.24)	40.71 (22.41)	8.60 (5.15)
\$60,001 to \$80,000	60,200	0.57 (0.19)	42.79 (22.96)	8.38 (5.08)
>\$80,000	18,930	0.59 (0.14)	42.89 (21.19)	8.43 (4.79)
<b>Date Type</b>				
Weekend	74,779	0.59 (0.22)	40.80 (21.88)	8.55 (5.62)
Weekday	180,083	0.54 (0.23)	41.70 (22.44)	8.51 (4.82)
Holiday	8,349	0.62 (0.23)	40.47 (20.38)	9.82 (7.16)
Non-holiday	246,513	0.55 (0.22)	41.47 (22.35)	8.48 (4.98)

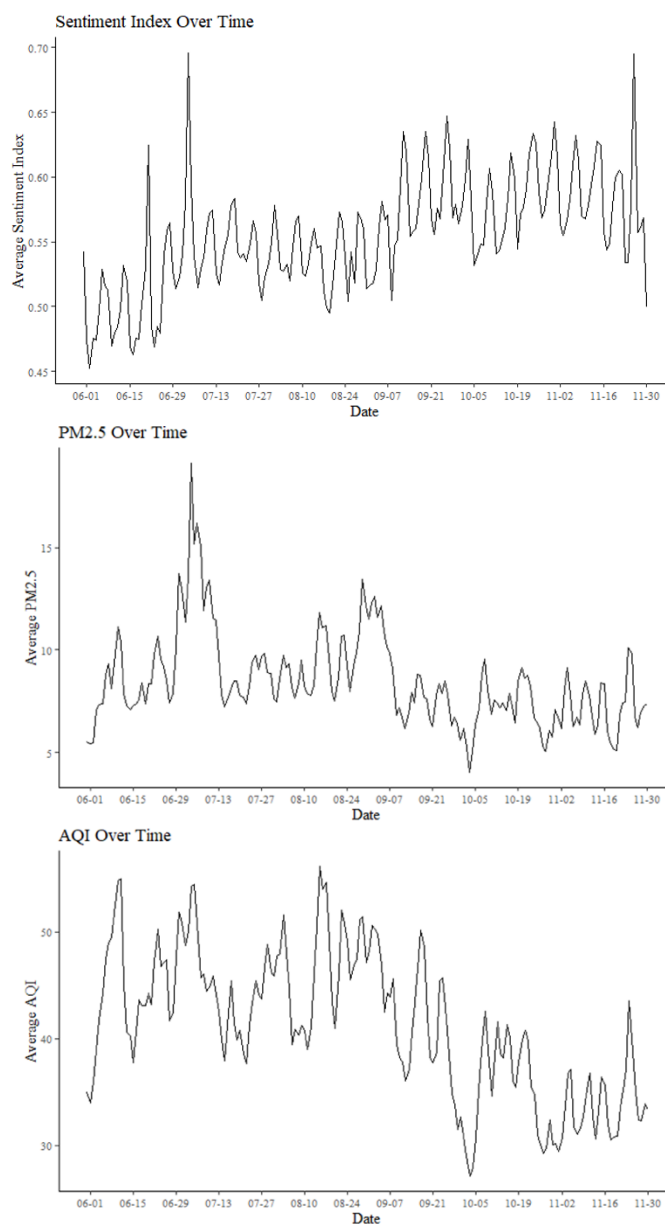
Observations are at the county/day level. Sentiment refers to the median tweeters' probability of posting a positive tweet given that they post a tweet. AQI refers to air quality index. PM<sub>2.5</sub> refers to fine particulate matter of size less than 2.5 micrometers and is measured in micrograms per cubic meter. MSA stands for Metropolitan Statistical Area. All county information is from 2010. Weekends include Saturday and Sunday. Holidays include all federal holidays in the given time period as well as Halloween

Sentiment is found to vary systematically with both county and date characteristics. Counties contained in MSAs have a higher mean sentiment than urban counties, and urban counties have a higher mean sentiment than rural counties (Table 2). Sentiment also tends to be lower in Hispanic-majority and Black-majority counties than in White-majority counties and increase with median household income (Table 2). Holidays and weekends tend to experience higher sentiment than nonholidays and weekdays, respectively (Table 2; Figure 3). Both AQI and PM<sub>2.5</sub> tend to be higher in summer than in autumn, with PM<sub>2.5</sub> peaking on July 4 (Figure 3).

Table 3 shows the results of the multi-cutoff regression

discontinuity model for both AQI and AQI lagged by one day, considering only counties that reside in MSAs with a population of at least 350,000. A coefficient of -0.015 is estimated for the same-day AQI threshold of Good to Moderate (cutoff value of 51) and is significant with a p-value less than 0.01. This would suggest that being in the AQI category of Moderate rather than Good predicts a 1.5 percentage point decrease in the sentiment index after controlling for weather and county and date fixed effects. See Figure 4 for a scatter plot depicting this relationship. This result is similar when considering a polynomial of order 3 with a

**Figure 3:** Time series plots depicting average sentiment index, average PM2.5, and average AQI per day, May 31 to November 11, 2015



The sentiment index, at the county-day level, refers to the probability that a tweet posted by the median tweeter is positive. AQI stands for Air Quality Index. PM<sub>2.5</sub> refers to fine particulate matter of size less than 2.5 micrometers and is measured in micrograms per cubic meter.

coefficient of -0.016 and smaller when considering a polynomial of order 1, with a statistically insignificant coefficient of -0.05 (Table S3 Column 1).

The estimates for the same-day AQI Moderate to Unhealthy for Sensitive Groups threshold (cutoff value of 101) and the Unhealthy for Sensitive Groups to Unhealthy

**Table 3:** Results from multi-cutoff regression discontinuity

AQI Category Threshold	Same day			Lagged one day		
	Coefficient (SE)	N	Estimated Bandwidth	Coefficient (SE)	N	Estimated Bandwidth
Good to Moderate	-0.015 (0.006)**	82,965	7.156	-0.003 (0.004)	81,685	17.819
Moderate to Unhealthy for Sensitive Groups	0.032 (0.043)	5,135	4.505	0.005 (0.013)	5,118	38.107
Unhealthy for Sensitive Groups to Unhealthy	0.256 (0.179)	782	4.47	-0.032 (0.179)	782	28.492

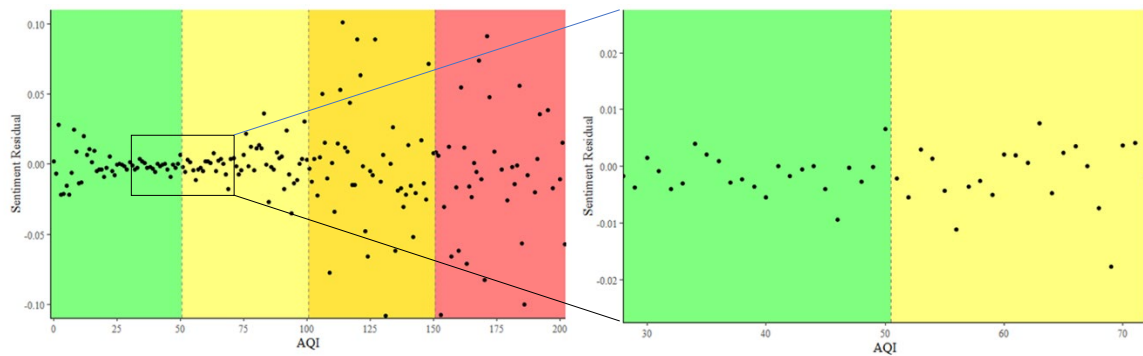
\*  $p\text{-value} < 0.1$

\*\*  $p\text{-value} < 0.05$

\*\*\*  $p\text{-value} < 0.01$

AQI stands for air quality index. AQI values between 0 and 50 are categorized as Good, between 51 and 100 as Moderate, between 101 and 150 as Unhealthy for Sensitive Groups, and between 151 and 200 as Unhealthy. Dependent variable is residual of sentiment index regressed on weather variables, county fixed effects, and date fixed effects, population-weighted by the square root of the number of unique tweeters per county/day. The models include only information from counties that reside in MSAs with populations of at least 350,000. The coefficients were estimated using a local polynomial of degree 2. Robust standard errors are clustered at the county level. Bandwidths were estimated so as to optimize the common mean-squared error on both sides of the given cutoff point.

**Figure 4:** Scatter plots depicting relationship between AQI and sentiment index residuals



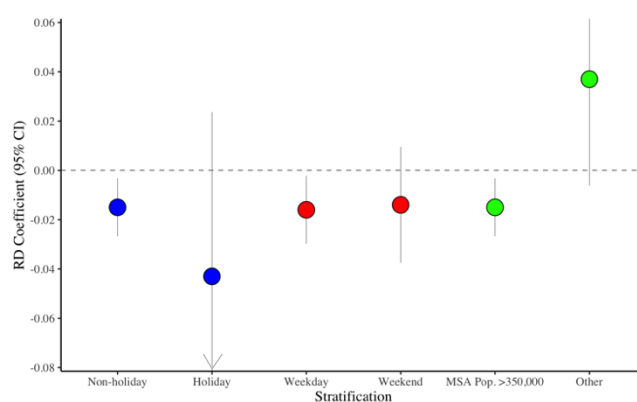
The sentiment index, at the county-day level, refers to the probability that a tweet posted by the median tweeter is positive. AQI stands for Air Quality Index. AQI values between 0 and 50 are categorized as Good (green), between 51 and 100 as Moderate (yellow), between 101 and 150 as Unhealthy for Sensitive Groups (orange), and between 151 and 200 as Unhealthy (red). Sentiment residual is the residual of sentiment index regressed on weather variables, county fixed effects, and date fixed effects, population-weighted by the square root of the number of unique tweeters per county/day. The plot include only information from counties that reside in MSAs with populations of at least 350,000.

threshold (cutoff value of 151) are both found to be positive but insignificant, with p-values larger than 0.1. Table S4 shows results for same-day AQI with the sentiment residuals predicted from slightly different models—with and without visibility, and with and without population weighting. The results are similar across each of these model specifications. The estimated coefficients for AQI lagged by one day are generally insignificant, with a negative coefficient at the Good to Moderate threshold and the Unhealthy for Sensitive Groups to Unhealthy threshold, and a positive coefficient at the Moderate to Unhealthy for Sensitive Groups threshold (Tables 3, S3).

Figure 5 shows regression discontinuity results stratified by date type (holiday, non-holiday, weekday, and weekend) and by whether the given county resides in

an MSA with a population of at least 350,000 and so is required to report daily AQI. As can be seen, the estimated effect of being above the 51 AQI threshold (Good to Moderate) is similar for weekdays, weekends, and nonholidays but is considerably larger

**Figure 5:** Estimated coefficients from multi-cutoff regression discontinuity at the Good to Moderate AQI threshold across date and county stratifications



AQI values between 0 and 50 are categorized as Good and between 51 and 100 as Moderate. The coefficients were estimated using a local polynomial of degree 2. Robust standard errors are clustered at the county level. Bandwidths for the regression discontinuity were estimated so as to optimize the common mean-squared error on both sides of the given cutoff point. The dependent variable is the residual of sentiment index regressed on weather variables, county fixed effects, and date fixed effects, population-weighted by the square root of the number of unique tweeters per county/day. Weekends include Saturday and Sunday. Holidays include all federal holidays in the given time period as well as Halloween. Stratified models by date characteristics include only information from counties that reside in MSAs with populations of at least 350,000.



on holidays with a coefficient of -0.043 (Table S5). A positive coefficient of 0.032 was estimated for counties outside large MSAs with a p-value less than 0.1 (Table S5).

Table 4 shows the results of WLS models with sentiment regressed on PM<sub>2.5</sub>, AQI, and/or visibility while controlling for weather variables and county and date fixed

**Table 4:** Estimated WLS coefficients of air pollution and weather variables

Variable	1	2	3	4	5
PM <sub>2.5</sub>	0.00007 (0.00008)	0.00022 (0.00008)***	--	--	--
AQI	--	--	0.00004 (0.00002)*	0.00005 (0.00002)***	--
Visibility	--	0.00353 (0.00035)***	--	0.00294 (0.00042)***	0.00338 (0.00035)***
Temperature	0.00226 (0.00029)***	0.00235 (0.00029)***	0.00190 (0.00035)***	0.00196 (0.00035)***	0.00230 (0.00028)***
Temperature <sup>2</sup>	-0.00002 (0.00000)***	-0.00002 (0.00000)***	-0.00002 (0.00000)***	-0.00002 (0.00000)***	-0.00002 (0.00000)***
Pressure	0.00050 (0.00009)***	0.00039 (0.00009)***	0.00034 (0.00012)***	0.00025 (0.00011)***	0.00039 (0.00009)***
Wind Speed	-0.00084 (0.00015)***	-0.00090 (0.00015)***	-0.00057 (0.00018)***	-0.00065 (0.00018)***	-0.00099 (0.00014)***
Precipitation	-0.01123 (0.00108)***	-0.00809 (0.00109)***	-0.00876 (0.00121)***	-0.00620 (0.00121)***	-0.00850 (0.00107)***
Snow Depth	0.00157 (0.00222)	0.00299 (0.00218)	0.00109 (0.00230)	0.00230 (0.00224)	0.00291 (0.00218)
County FE	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes
Pop Weighted	Yes	Yes	Yes	Yes	Yes
N	254,850	254,850	119,907	119,907	254,850
R <sup>2</sup>	0.23024	0.23061	0.32206	0.32244	0.23060

\* *p*-value < 0.1

\*\* *p*-value < 0.05

\*\*\* *p*-value < 0.01

AQI stands for air quality index. FE stands for fixed effects. Pop stands for population. The dependent variable is the sentiment index, which can be described as the median tweeters' probability of posting a positive tweet given that they post a tweet. All models include county fixed effects and date fixed effects and are population-weighted by the square root of the number of unique tweeters per county/day. Robust standard errors are clustered at the county level. Observations are at the county-day level.

effects.  $PM_{2.5}$  is estimated to have a statistically insignificant coefficient of 0.00007 when visibility is not included in the model (Table 4 Column 1), and a significant coefficient of 0.00022 when controlling for visibility (p-value less than 0.01; Table 4 Column 2). Similarly, AQI is estimated to have a positive coefficient that is amplified when including visibility (Table 4 Columns 3-4). Across all models, visibility has an estimated coefficient hovering around 0.0035, with a p-value well below 0.01 (Table 4 Columns 2, 4, 5). All other weather variables excluding snow depth are observed to have robust significant estimates across all models, with average temperature being positive, the square of average temperature being negative, station pressure being positive, and wind speed and precipitation being negative.

Table S6 in the supplementary material includes the results from other model specifications and shows the estimated coefficients for  $PM_{2.5}$  and all weather variables, including visibility. When not controlling for date fixed effects nor weather variables,  $PM_{2.5}$  is estimated to have a significantly negative coefficient (Table S6 Columns 1-2). Regressions with population weighting tend to have slightly smaller but similar estimates for all variables compared to regressions without population weighting. No evidence was found for  $PM_{2.5}$  having a quadratic effect. Table S7 shows that the estimated coefficient of visibility remains largely consistent across all model specifications, with a coefficient generally between 0.003 and 0.004. Figure S1 in the supplementary material shows that the association between visibility and sentiment is highly linear despite the censoring of the visibility variable at 10 miles.

Table 5 reports the results of WLS models considering  $PM_{2.5}$  lagged by one to two days. In all models, weather variables and county and date fixed effects are included as

**Table 5:** Estimated WLS coefficients for lagged PM<sub>2.5</sub>

Variable	1	2	3	4	5	6
PM <sub>2.5</sub>	0.0022 (0.00008)***	--	--	0.00030 (0.00009)***	--	0.00016 (0.00009)*
PM <sub>2.5, t-1</sub>	--	0.00002 (0.00008)	--	-0.00015 (0.00009)	--	--
PM <sub>2.5, t-2</sub>	--	--	0.00001 (0.00008)	0.00000 (0.00001)	--	--
(PM <sub>2.5</sub> – PM <sub>2.5, t-1</sub> )	--	--	--	--	0.00022 (0.00007)***	0.00014 (0.00009)*
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Pop Weighted	Yes	Yes	Yes	Yes	Yes	Yes
N	254,850	254,850	254,850	254,850	254,850	254,850
R <sup>2</sup>	0.23061	0.23059	0.23059	0.23062	0.23061	0.23062

\*  $p$ -value < 0.1\*\*  $p$ -value < 0.05\*\*\*  $p$ -value < 0.01

FE stands for fixed effects. Pop stands for population. The dependent variable is the sentiment index, which can be described as the median tweeters' probability of posting a positive tweet given that they post a tweet. All models include weather variables, county fixed effects, and date fixed effects and are population-weighted by the square root of the number of unique tweeters per county/day. Robust standard errors are clustered at the county level. Observations are at the county-day level. Column 4 reports the results of a distributed lag model. Standard errors were solved for using the delta method.

controls. No significant coefficient is estimated for either the one-day or two-day lag (Table 5 Columns 2-4). This is seen both in the models that include the one-day and two-day lags separately (Columns 2 and 3) and in the quadratic distributed lag model (Column 4). Columns 5 and 6 suggest that an increase in the current day's PM<sub>2.5</sub> level compared to the previous day's PM<sub>2.5</sub> level predicts higher sentiment, with a statistically significant coefficient of 0.00022 ( $p$ -value less than 0.05). This holds even after controlling for the current day's PM<sub>2.5</sub> level, although the coefficient is slightly smaller (0.00014).

Table 6 reports the results for lagged visibility. Significant positive coefficients are observed in models where the one-day lag and two-day lags are considered separately (0.00066 and 0.00059, respectively; Columns 2-3). However, in the distributed lag

**Table 6:** Estimated WLS coefficients for lagged visibility

Variable	1	2	3	4	5	6
Visibility	0.00338 (0.00035)***	--	--	0.00345 (0.00031)***	--	0.00298 (0.00043)***
Visibility <sub>t-1</sub>	--	0.00066 (0.00030)**	--	-0.00063 (0.00031)**	--	--
Visibility <sub>t-2</sub>	--	--	0.00059 (0.00032)*	0.00078 (0.00025)***	--	--
(Visibility – Visibility <sub>t-1</sub> )	--	--	--	--	0.00196 (0.00025)***	0.00046 (0.00030)*
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Pop Weighted	Yes	Yes	Yes	Yes	Yes	Yes
N	254,850	233,780	220,749	220,749	233,780	233,780
R <sup>2</sup>	0.23060	0.24258	0.24931	0.24967	0.24274	0.2493

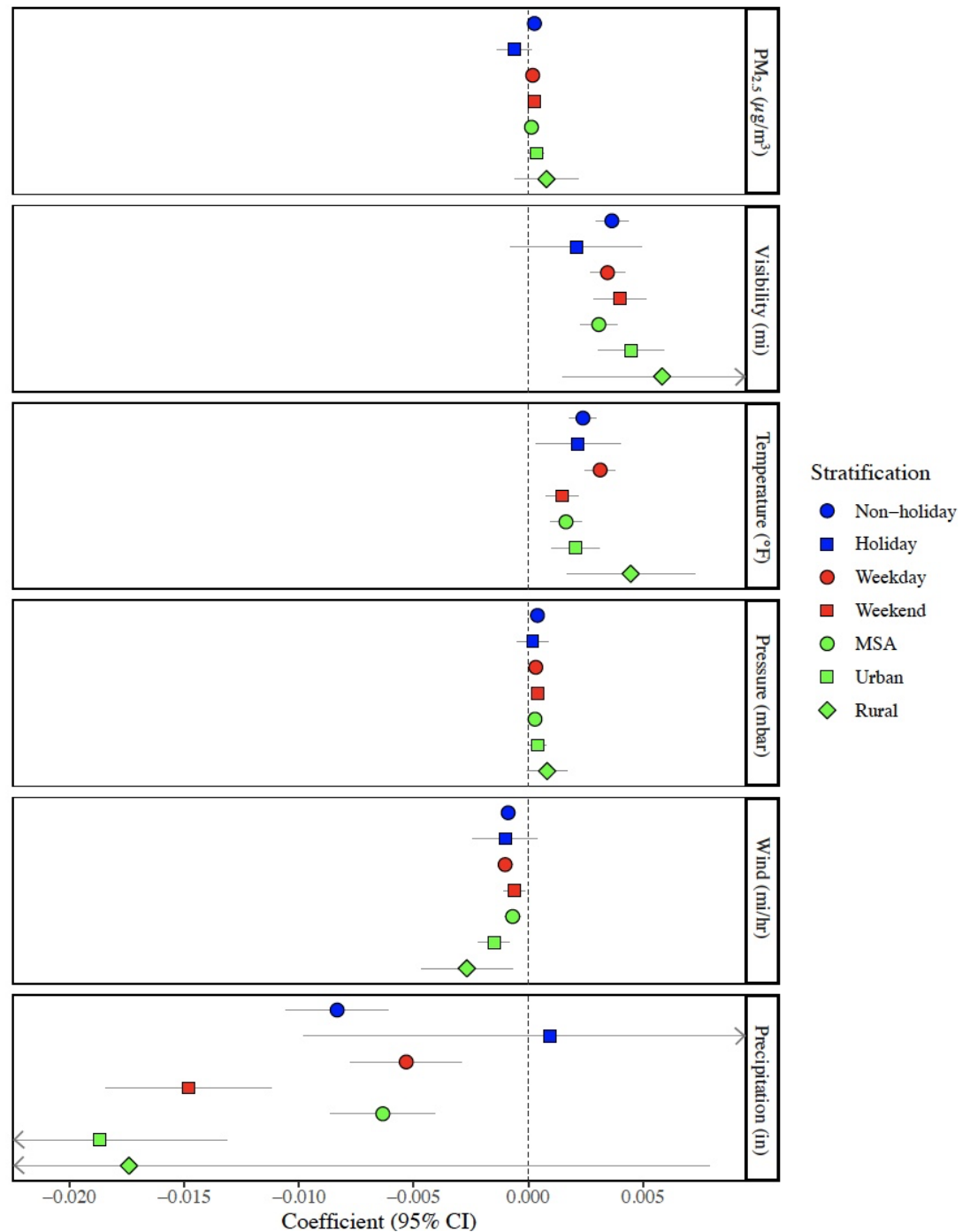
\*  $p$ -value < 0.1\*\*  $p$ -value < 0.05\*\*\*  $p$ -value < 0.01

FE stands for fixed effects. Pop stands for population. The dependent variable is the sentiment index, which can be described as the median tweeters' probability of posting a positive tweet given that they post a tweet. All models include weather variables, county fixed effects, and date fixed effects and are population-weighted by the square root of the number of unique tweeters per county/day. Robust standard errors are clustered at the county level. Observations are at the county-day level. Column 4 reports the results of a distributed lag model. Standard errors were solved for using the delta method.

model, the one-day lag has a significant negative coefficient of -0.00063 (p-value less than 0.1) while the two-day lag remains positive with a coefficient of 0.00078 (p-value less than 0.01; Column 4). Columns 5 and 6 suggest that an increase in the current day's visibility compared to the previous day's visibility level predicts higher sentiment, with a statistically significant coefficient of 0.00196 (p-value less than 0.01). This holds even after controlling for the current day's visibility, although the coefficient is considerably smaller (0.00046).

Figure 6 and Table S8 show the results of stratified analyses for PM<sub>2.5</sub> and all weather variables excluding snow depth. Across all stratifications (holidays, non-holidays, weekends, and weekdays, and MSA, urban, and rural counties), visibility, temperature, and barometric pressure are found to have a direct relationship with the

**Figure 6:** Estimated WLS coefficients of PM<sub>2.5</sub> and weather variables across date and county stratifications



The dependent variable is the sentiment index, which can be described as the median tweeters' probability of posting a positive tweet given that they post a tweet. Estimates are from weighted least squared models that include county fixed effects and date fixed effects and are population-weighted by the square root of the number of unique tweeters per county/day. Robust standard errors are clustered at the county level. MSA stands for Metropolitan Statistical Area. Weekends include Saturday and Sunday. Holidays include all federal holidays in the given time period as well as Halloween

sentiment index, with slightly amplified effects in rural counties compared to MSA and urban counties. Similarly, wind is estimated to have a consistent inverse relationship with the sentiment index with amplified effects in rural counties. PM<sub>2.5</sub> is estimated to have a positive effect across all county types and across nonholidays, weekdays, and weekends. However, a statistically insignificant negative coefficient was estimated for PM<sub>2.5</sub> on holidays (-0.000627). Precipitation is found to be negatively associated with sentiment except on holidays, with an amplified effect on weekends (0.00184 compared to 0.00124 on weekdays).

## Discussion

### *Perceived air pollution*

In this study, evidence was found to suggest that perceived air pollution is negatively associated with expressed sentiment via U.S. Twitter. It was estimated that going from the Good AQI category to the Moderate category leads to a 1.5 percentage point decrease in the sentiment index; however, positive coefficients—all insignificant—were found at the other AQI category thresholds (Table 3). Thus, the first hypothesis that expressed sentiment will be more positive just below an AQI cutoff point than just above it is only partially confirmed. Moreover, viewing the relationship between AQI and the sentiment index after controlling for weather variables and county and date fixed effects does not reveal any drastic differences in the relationship around the cutoff value of 51 (the threshold between the Good and Moderate categories) apart from a relatively high average sentiment at the AQI value of 50 (Figure 4). Is this simply noise? It is difficult to say.

On the one hand, there are reasons to believe that the results are revealing an important effect of air quality on sentiment. The estimate at the Good to Moderate cutoff was found to be amplified on holidays compared to non-holidays, which would be expected if the AQI category affects expressed sentiment by influencing recreational behavior choices (see 8). Moreover, the negative effect is found only in counties contained in MSAs that have populations of at least 350,000 and thus are required to report the AQI (Figure 5, Table S5). In contrast, a positive effect is found in all other counties, which matches the general trend seen that higher air quality predicts higher sentiment (discussed more below). This would be expected if the AQI category affects sentiment, as those in counties with more consistent access to AQI information are more likely to be primed to be affected by the AQI category.

On the other hand, it should be noted that the population of Twitter users, particularly Twitter users with activated geotagged settings, tend to be younger than the general population with 18-29-year-olds overrepresented (48). As has been documented by both Pew and The Media Insight Project, adults aged 18-29 are less likely than older adults to follow the news (49-50). Specifically, The Media Insight Project's 2014 survey of 1,492 adults found that an estimated 69% of 18-29-year-olds follow news on the environment and natural disasters and 71% to traffic and weather, compared to 78% and 93% of 30-39-year-olds and 74% and 81% of 40-49-year-olds, respectively (50). This would suggest that the tweeters who make up the sentiment index may be relatively likely less to be aware of their area's AQI category, hence diminishing any effect that the AQI category may have on sentiment.

Robust evidence was found to support the hypothesis that visibility has a direct relationship with expressed sentiment. Overall, a coefficient roughly between 0.003 and 0.004 was estimated, suggesting that one mile of increased visibility leads to a 0.3 to 0.4 percentage point increase in the probability that a tweet posted by a county-day's median tweeter is positive (Tables 4,6, S6-S8). The association is observed to be linear despite the censoring of the visibility variable at 10 miles. That being said, the censoring requires that caution be exercised when interpreting this estimate.

The results found in Table 6 Columns 4-6 suggest that the association between visibility and sentiment is dependent on the visibility from previous days. Specifically, it would appear that when today's visibility is higher than yesterday's visibility, sentiment will tend to be higher. Conversely, higher visibility yesterday compared to today predicts a decrease in today's sentiment. Put in other words, the importance of visibility appears to be relative: opaque skies are worse when yesterday's skies were clear.

While keeping in mind that visibility is an imperfect proxy for smog or haze, these results do offer evidence that air pollution has a psychological or perceived effect on U.S. Twitter sentiment. They suggest that the appearance of air pollution can decrease human happiness. This should come as no surprise to those who have lived in areas prone to hazy days. However, to the best of my knowledge, this study is the first to document a link between visibility and expressed sentiment (see 2).

A number of other studies have shown that measurements of perceived air pollution, as well as objective measurements of air pollutants, are negatively associated with happiness (10, 51-52). It is quite plausible that the results of these other studies are picking up the visible component of air pollution observed in this study. For



measurements of perceived air pollution are likely, to some degree, measurements of smog, as the appearance of smog (or lack thereof) can act as a simple and accessible indicator of air quality for study participants. Moreover, studies considering objective measurements of air pollution will inevitably capture the visible component of air quality unless it is specifically controlled for. Future studies on the effects of perceived and actual air pollution on non-health factors should more fully consider the important visible component of air pollution.

#### *Actual air pollution*

In this study, no significant evidence was found to support the hypothesis that PM<sub>2.5</sub> has an inverse relationship with expressed sentiment. In fact, it was observed that the association between PM<sub>2.5</sub> and sentiment is mainly *positive* (Tables 4-5, S6, S8; Figure 6). This contradicts the findings of other studies, particularly Zheng et al. who considered the association between PM<sub>2.5</sub> and expressed sentiment of Chinese urbanites via the social media platform Sina Weibo (13). In that study, a one standard deviation increase in PM<sub>2.5</sub> was observed to predict a roughly 0.045 standard deviation increase in the sentiment index (13).

A few points may offer some insight into why nonnegative estimates are found in this study. First, it has been well documented among the epidemiology literature that the effect of PM<sub>2.5</sub> exposure on health outcomes is substantially larger when considering exposure over the long-term (decades) rather than over the short-term (days; 1). This suggests that the physiological effects of PM<sub>2.5</sub> on psychological outcomes such as sentiment might be best understood by considering long-term average concentrations

rather than short-term variation. In this study, only same-day and one-to-two day lagged estimates of  $PM_{2.5}$  were considered. Future studies may wish to instead consider the effects of long-term  $PM_{2.5}$  on expressed sentiment; however, controlling for important time-invariant information in the analyses will prove a challenge.

Second, it is important to note that estimates for  $PM_{2.5}$  are found to be negative when date fixed effects are not included in the model (Table S6). This may be interpreted to signify that increases in local  $PM_{2.5}$  levels are loosely associated with nationwide Twitter trends. For example, forest fires in California—prevalent throughout the sample period of May 31 to November 30, 2015—caused both increases in  $PM_{2.5}$  levels throughout California and increases in nationwide conversations about fires (likely resulting in decreased Twitter sentiment). Date fixed effects would largely capture these trends in national sentiment.

Third, it was observed that lagged  $PM_{2.5}$  has no significant association with sentiment while same-day  $PM_{2.5}$  minus previous-day  $PM_{2.5}$  has a significant positive association even after controlling for same-day  $PM_{2.5}$  (Table 5). This provides further evidence that short-term variation in  $PM_{2.5}$  has no measurable physiological effect on expressed sentiment. However, these results would be expected if some behaviors that increase people's expressed happiness also increase levels of  $PM_{2.5}$  (such as setting off fireworks and driving to recreational areas). To better understand this, future studies may consider analyzing the association between air pollution and tweet topic trends. This will allow for a more holistic understanding of how air pollution and tweeting behavior are related, beyond simple sentiment polarity.

### *Weather*

Throughout this study, robust estimates were found for average temperature, barometric pressure, precipitation, and average wind speed. Average temperature (the arithmetic mean of the daily maximum and minimum temperatures) was found to have a quadratic association with sentiment, increasing to a maximum at around 55° Fahrenheit and decreasing afterwards (Tables 4, S6-S8; Figure 6). Barometric pressure—which is commonly associated with improving weather—was found to have a positive association. Wind speed and precipitation were found to have a negative association. These results are similar to those found by past studies that analyzed the association between weather and expressed sentiment via social media (53).

### *Strengths and Limitations*

This study has several important limitations. First, the sample period considered in this study covers only six months of a single year (May 31 to November 30, 2015). It is possible that the results observed in this study would differ had I considered a different or longer time period. Future studies should check the external validity of the results presented in this study by analyzing the association between air pollution and sentiment across different time periods.

Second, the sentiment index used in this study is imperfect. While the machine learning model used to predict tweet sentiment had a testing accuracy comparable to state-of-the-art models, the accuracy never rose above 88.03% (21). The resulting index likely suffers from idiosyncratic errors and systematic failures at detecting complex emotional patterns such as sarcasm. On a similar note, it is not necessarily true that the

sentiment of a tweet matches the emotion had by the tweeter. Thus, the sentiment index and, more generally, the results of this paper, should be interpreted with caution.

Third, the weather and air pollution estimates used in this study may suffer from measurement error. Of particular importance is the censoring of the visibility variable at 10 miles. Unfortunately, censoring ground visibility at 10 miles is a common approach taken by weather stations in reporting weather information (17). The production of high-quality weather data at the county-date level will prove exceptionally helpful to future studies on the association between air pollution, weather, and non-health factors.

Fourth and lastly, the results of this study should not be generalized to the broader U.S. population. The tweeters who make up the sentiment index are self-selected as Twitter users with activated geotagged settings. And, as discussed above, this subpopulation of tweeters is not representative of the U.S. population (49-50).

Strengths of this study include a large number of tweets from which to create the sentiment index and a large number of observations, allowing for precise estimates. A rich set of covariates (including county and date fixed effects) allowed for the limitation of omitted variable bias in the analytical models.

## **Conclusion**

Using data on 27,827,828 U.S. tweets, evidence was found for air pollution having a psychological but not a physiological effect on expressed sentiment. The marginal effect of going from the Good to the Moderate AQI category was estimated to predict a 1.5 percentage point decrease in the sentiment index, and a 1-mile increase in ground visibility was estimated to predict roughly a 0.34 percentage point increase. PM<sub>2.5</sub>

was found to predict a small increase of about 0.02 percentage points in the sentiment index. Significant evidence was also found for temperature, pressure, wind speed, and precipitation affecting expressed sentiment.

## References

1. Pope III CA, Dockery DW. Health Effects of Fine Particulate Air Pollution: Lines that Connect. *Journal of the Air & Waste Management Association*. 2006; 56: 709-742.
2. Lu JG. Air pollution: A systematic review of its psychological, economic, and social effects. *Current Opinion in Psychology*. 2019; 32: 52-65.
3. Wilson WE, Suh HH. Fine Particles and Coarse Particles: Concentration Relationships Relevant to Epidemiologic Studies. *Journal of the Air & Waste Management Association*. 1997; 47: 1238-1249.
4. Chang LC, Koutrakis P, Catalano PJ, Suh HH. Hourly personal exposures to fine particles and gaseous pollutants. *Journal of the Air & Waste Management Association*. 2000; 50: 1223-1235.
5. Fehr R, Yam KC, He W, Chiang JT, Wei W. Polluted work: A self-control perspective on air pollution appraisals, organizational citizenship, and counterproductive work behavior. *Organizational Behavior and Human Decision Process*. 2017; 143: 98-110.
6. Neidell M. Air quality warnings and outdoor activities: evidence from Southern California using a regression discontinuity design. *Journal of Epidemiology & Community Health* 2010; 64: 921-926.
7. Claeson AS, Lidén E, Nordin M, Nordin S. The role of perceived pollution and health risk perception in annoyance and health symptoms: a population-based study of odorous air pollution. *International Archives of Occupational and Environmental Health*. 2013; 86: 367-374.
8. Saberian S, Heyes A, Rivers N. Alerts work! Air quality warnings and cycling. *Resource and Energy Economics*. 2017; 49: 165-185.
9. Lu H, Yue A, Chen H, Long R. Could smog pollution lead to the migration of local skilled workers? Evidence from the Jing-Jin-Ji region in China. *Resource, Conservation, and Recycling*. 2018: 130; 177-187.
10. Mackerron G, Mourato S. Life satisfaction and air quality in London. *Ecological Economics*. 2009; 68: 1441-1453.
11. Levinson A. Valuing public goods using happiness data: the case of air quality. *Journal of Public Economics*. 2012; 96: 869-880.
12. Kahneman D, Krueger AB. Developments in the Measurement of Subjective Well-Being. *Journal of Economic Perspectives*. 2006; 20: 3-24.
13. Zheng S, Wang J, Sun C, Zhang X, Kahn ME. Air pollution lowers Chinese urbanites' expressed happiness on social media. *Nature Human Behavior*. 2019; 3: 237-243.
14. AirNow. Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI). <https://www.airnow.gov/sites/default/files/2020-05/aqi-technical-assistance-document-sept2018.pdf>. Updated September 2018. Accessed February 2021.
15. Meyers J. View Air Quality in Apple Maps to See How Polluted Cities & Destinations Are. *Gadget Hacks*. <https://ios.gadgethacks.com/how-to/view-air->

- quality-apple-maps-see-polluted-cities-destinations-are-0193020/. Updated January 2019. Accessed February 2021.
16. AirNow. Air Quality Index (AQI) Basics. <https://www.airnow.gov/aqi/aqi-basics/#:~:text=Think%20of%20the%20AQI%20as,300%20represents%20hazardous%20air%20quality>. Accessed February 2021.
  17. National Oceanic and Atmospheric Administration. Local Climatological Data (LCD) Dataset Documentation. [https://www.ncei.noaa.gov/data/local-climatological-data/doc/LCD\\_documentation.pdf](https://www.ncei.noaa.gov/data/local-climatological-data/doc/LCD_documentation.pdf). Updated March 2020. Accessed February 2021.
  18. Jürgen P, Morstatter F. Geotagged Twitter posts from the United States: A tweet collection to investigate representativeness, *Gesis*. <https://data.gesis.org/sharing/#!/Detail/10.7802/1166>. Updated 2016. Accessed October 2020.
  19. McGrath R. Twython 3.8.0 documentation. <https://twython.readthedocs.io/en/latest/>. Updated 2019. Accessed October 2020.
  20. Littman J, Wrube L, Kerchner D, Gaber YB. News Outlet Tweet Ids. *Harvard Dataverse*. 2017; V4.
  21. Naseem U, Razzak I, Musial K, Imran M. Transformer based Deep Intelligent Contextual Embedding for Twitter sentiment analysis. *Future Generation Computer Systems*. 2010; 113: 58-69.
  22. Nguyen DQ, Vu T, Nguyen AT. BERTweet: A pre-trained language model for English Tweets. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. 2020; 9-14.
  23. Baziotis C, Pelekis N, Doukeridis C. DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. 2017; 747-754.
  24. Kim T, Wurster K. emoji. GitHub. <https://github.com/carpedm20/emoji/>. Updated January 2021. Accessed January 2021.
  25. Wikipedia. List of Emoticons. [https://en.wikipedia.org/wiki/List\\_of\\_emoticons](https://en.wikipedia.org/wiki/List_of_emoticons). Updated November 2020. Accessed November 2020.
  26. Bird S, Loper E, Klein E. *Natural Language Processing with Python*. Newton: O'Reilly Media Inc.; 2009.
  27. Cambria E, Poria S, Hazarika D, Kwok K. SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*. 2018; 1795–1802.
  28. Hutto CJ, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*. 2014; 216-225.
  29. Hu M, Liu B. Mining and summarizing customer reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2004; 168-177.
  30. Mohammad S, Kiritchenko S, Zhu X. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh*

- International Workshop on Semantic Evaluation, SemEval 2013, Association for Computational Linguistics*. 2013; 321-327
31. Kiritchenko S, Zhu X, Cherry C, Mohammad S, NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. *SemEval@COLING*. 2014.
  32. Tang D, Wei F, Qin B, Zhou M, Liu T. Building large-scale Twitter-specific sentiment lexicon: A representation learning approach. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, Dublin City University and Association for Computational Linguistics*. 2014; 172-182
  33. Pennington J, Socher R, Manning CD. *GloVe: Global Vectors for Word Representation*. 2014.
  34. Devlin J, Chang M., Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Association for Computational Linguistics. 2019; 4171-4186.
  35. Abadi M, Agarwal A, Barham, et al. TensorFlow: A System for Large-Scale Machine Learning. *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*. 2016; 265-283.
  36. Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*. 2009.
  37. Saif H, Fernandez M, He Y, Alani H. Evaluation datasets for twitter sentiment analysis. *Proceedings, 1st Workshop on Emotion and Sentiment in Social and Expressive Media (ESSEM) in conjunction with AI\*IA Conference*. 2013.
  38. Nakov P, Rosenthal S, Kozareva Z, Stoyanov V, Ritter A, Wilson T. SemEval-2013 Task 2: Sentiment Analysis in Twitter. *Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. 2013; 312-320.
  39. Kaggle. Twitter US Airline Sentiment V4.  
<https://www.kaggle.com/crowdflower/twitter-airline-sentiment>. Updated October 2019. Accessed November 2020.
  40. US Environmental Protection Agency. Air Quality System Data Mart.  
[https://aqs.epa.gov/aqsweb/airdata/download\\_files.html#AQI](https://aqs.epa.gov/aqsweb/airdata/download_files.html#AQI). Updated 2021. Accessed January 2021.
  41. Centers for Disease Control and Prevention. Daily PM2.5 Concentrations All County, 2001-2016. <https://catalog.data.gov/dataset/daily-pm2-5-concentrations-all-county-2001-2016>. Updated November 2020. Accessed January 2021.
  42. Berrocal VJ, Gelfand AE, Holland DM. Space-time data fusion under error in computer model output: an application to modeling air quality. *Biometrics*. 2012; 68: 837-848.
  43. Menne MJ, Durre I, Vose RS, Gleason BE, Houston TG. An Overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology*. 2012; 29: 897-910.



44. Severson R. countyweather. *GitHub*.  
<https://github.com/leighseverson/countyweather>. Updated December 2020.  
 Accessed January 2021.
45. U.S. Department of Agriculture Economic Research Service. County-level  
 Datasets. <https://www.ers.usda.gov/data-products/county-level-data-sets/documentation/>. Updated 2021. Accessed January 2021.
46. Cattaneo MD, Keele L, Titiunik R, Vasquez-Bare G. Interpreting Regression  
 Discontinuity Designs with Multiple Cutoffs. *The Journal of Politics*. 2016; 78:  
 1229-1248
47. Hausman C, Rapson DS. Regression discontinuity in time: Considerations for  
 empirical applications. *Annual Review of Resource Economics*. 2018; 5: 533-  
 552.
48. Jiang Y, Li Z, Ye X. Understanding demographic and socioeconomic biases of  
 geotagged Twitter users at the county level. *Cartography and Geographic  
 Information Science*. 2019. 46: 228-242.
49. Pew Research Center. State of the News Media 2015.  
<https://assets.pewresearch.org/wp-content/uploads/sites/13/2017/05/30142603/state-of-the-news-media-report-2015-final.pdf>. Updated April 2015. Accessed February 2021.
50. The Media Insight Project. The Personal News Cycle: How Americans choose to  
 get their news. *American Press Institute*.  
<https://www.americanpressinstitute.org/publications/reports/survey-research/personal-news-cycle/single-page/>. Updated March 2014. Accessed  
 February 2021.
51. Rehdanz K, Maddison D. Local environmental quality and life satisfaction in  
 Germany. *Ecological Economics* 2008; 64: 787-797.
52. Xu J, Li J. Tax payment, social contribution for pollution prevention and  
 happiness. *Problemy Ekorozwoju-Problems of Sustainable Development*. 2016;  
 11: 59-64.
53. Baylis P, Obradovich N, Kryvasheyeu Y, Chen H, Coviello L, Moro E, Cebrian  
 M, Fowler JH. Weather impacts expressed sentiment. *PloS one*. 2018; 13:  
 e0195750.

## Supplementary Material

**Table S1:** BiLSTM Design and Specification for Twitter Sentiment Prediction

### **A. Sequential Model Build using the Keras package in Python**

```

model = Sequential()
model.add(GaussianNoise(gauss_sd))
for i in range(layers):
    model.add(Bidirectional(layer = LSTM(hidden_layer_dim,
                                         dropout = dropout,
                                         batch_input_shape = (batch_size, n_tokens, n_dimensions),
                                         return_sequences = True),
                           backward_layer = LSTM(hidden_layer_dim,
                                                  dropout = dropout,
                                                  batch_input_shape = (batch_size, n_tokens, n_dimensions),
                                                  activation= 'relu',
                                                  go_backwards = True,
                                                  return_sequences = True)))
model.add(Dropout(dropout))
model.add(Aattention(return_sequences = False))
model.add(Dense(2,
                activation='softmax',
                kernel_regularizer = l2(l2 = l_2)))
model.compile(loss = 'categorical_crossentropy',
              optimizer = Adam(learning_rate = l_r),
              metrics = ["accuracy"])

```

### **B. Hyperparameters**

<b>Parameter (variable name)</b>	<b>Value</b>
Dropout (dropout)	0.25
Gaussian standard deviation (gauss_sd)	0.3
Batch size (batch_size)	128
Dimension of each hidden layer (hidden_layer_dim)	150
L2 regularization value (l_2)	0.001
Learning rate (l_r)	0.001
Number of hidden layers (layers)	2

Hyperparameter values were selected through 5-fold cross validation

**Table S2:** Comparison of ISD weather station information within an extended county radius

<b>Estimate</b>	<b>Temperature (°F)</b>	<b>Station Pressure (mbar)</b>	<b>Ground Visibility (mi)</b>	<b>Wind Speed (mi/hr)</b>	<b>Precipitation (in)</b>	<b>Snow Depth (in)</b>
Pearson Correlation	0.98	0.93	0.70	0.80	0.58	0.56
Euclidean Distance	2.39	21.72	0.50	1.78	0.09	0.02

ISD refers to the Integrated Surface Database. A description of the extended county radius can be found in the methodology section

**Table S3:** Results from multi-cutoff regression discontinuity across different polynomial order

	AQI Category Threshold	Same day			Lagged one day		
		Coefficient (SE)	N	Estimated Bandwidth	Coefficient (SE)	N	Estimated Bandwidth
<b>A. Polynomial Order 1</b>	Good to Moderate	-0.005 (0.003)	82,96 5	8.998	-0.002 (0.003)	81,68 5	17.243
	Moderate to Unhealthy for Sensitive Groups	-0.011 (0.010)	5,135	8.025	0.003 (0.011)	5,118	29.329
	Unhealthy for Sensitive Groups to Unhealthy	0.066 (0.055)	782	3.478	-0.024 (0.021)	782	24.421
<b>B. Polynomial Order 3</b>	Good to Moderate	-0.016 (0.009)**	82,96 5	8.44	-0.005 (0.005)	81,68 5	18.603
	Moderate to Unhealthy for Sensitive Groups	0.028 (0.027)	5,135	9.633	0.010 (0.014)	5118	42.026
	Unhealthy for Sensitive Groups to Unhealthy	0.066 (0.060)	782	10.501	-0.064 (0.037)*	782	25.436

\*  $p\text{-value} < 0.1$ \*\*  $p\text{-value} < 0.05$ \*\*\*  $p\text{-value} < 0.01$ 

AQI stands for air quality index. AQI values between 0 and 50 are categorized as Good, between 51 and 100 as Moderate, between 101 and 150 as Unhealthy for Sensitive Groups, and between 151 and 200 as Unhealthy.

Dependent variable is residual of sentiment index regressed on weather variables, county fixed effects, and date fixed effects, population-weighted by the square root of the number of unique tweeters per county/day.

The coefficients were estimated using a local polynomial of degree 1 or 3. Robust standard errors are clustered at the county level. Bandwidths were estimated so as to optimize the common mean-squared error on both sides of the given cutoff point.

**Table S4:** Results from multi-cutoff regression discontinuity with the residuals from distinct model specifications

	<b>AQI Category Threshold</b>	<b>Coefficient (SE)</b>	<b>N</b>	<b>Estimated Bandwidth</b>
<b>Unweighted, Excluding Visibility</b>	Good to Moderate	-0.015 (0.007)**	82965	7.085
	Moderate to Unhealthy for Sensitive Groups	0.037 (0.043)	5135	4.505
	Unhealthy for Sensitive Groups to Unhealthy	0.262 (0.186)	782	4.295
<b>Unweighted</b>	Good to Moderate	-0.015 (0.007)**	82965	7.089
	Moderate to Unhealthy for Sensitive Groups	0.030 (0.042)	5135	4.502
	Unhealthy for Sensitive Groups to Unhealthy	0.251 (0.180)	782	4.318
<b>Weighted, Excluding Visibility</b>	Good to Moderate	-0.015 (0.006)**	82965	7.156
	Moderate to Unhealthy for Sensitive Groups	0.032 (0.043)	5135	4.505
	Unhealthy for Sensitive Groups to Unhealthy	0.256 (0.179)	782	4.47
<b>Weighted</b>	Good to Moderate	-0.015 (0.006)**	82965	7.229
	Moderate to Unhealthy for Sensitive Groups	0.027 (0.043)	5135	4.528
	Unhealthy for Sensitive Groups to Unhealthy	0.248 (0.175)	782	4.436

\*  $p\text{-value} < 0.1$ \*\*  $p\text{-value} < 0.05$ \*\*\*  $p\text{-value} < 0.01$ 

AQI stands for air quality index. AQI values between 0 and 50 are categorized as Good, between 51 and 100 as Moderate, between 101 and 150 as Unhealthy for Sensitive Groups, and between 151 and 200 as Unhealthy.

Dependent variable is residual of sentiment index regressed on weather variables (possibly excluding visibility), county fixed effects, and date fixed effects, possibly population-weighted by the square root of the number of unique tweeters per county/day.

The coefficients were estimated using a local polynomial of degree 2. Robust standard errors are clustered at the county level. Bandwidths were estimated so as to optimize the common mean-squared error on both sides of the given cutoff point.

**Table S5:** Results from multi-cutoff regression discontinuity with the residuals from distinct model specifications

	<b>AQI Category Threshold</b>	<b>Coefficient (SE)</b>	<b>N</b>	<b>Estimated Bandwidth</b>
<b>Holiday</b>	Good to Moderate	-0.043 (0.034)*	2,660	7.493
	Moderate to Unhealthy for Sensitive Groups	0.006 (0.070)	172	12.265
	Unhealthy for Sensitive Groups to Unhealthy	N/A	N/A	N/A
<b>Non-holiday</b>	Good to Moderate	-0.015 (0.006)**	80,305	7.591
	Moderate to Unhealthy for Sensitive Groups	0.040 (0.043)	4,963	4.420
	Unhealthy for Sensitive Groups to Unhealthy	0.236 (0.172)	762	4.95
<b>Weekend</b>	Good to Moderate	-0.014 (0.012)	23,969	7.79
	Moderate to Unhealthy for Sensitive Groups	-0.053 (0.077)	1,375	4.98
	Unhealthy for Sensitive Groups to Unhealthy	0.190 (0.130)	224	6.935
<b>Weekday</b>	Good to Moderate	-0.016 (0.007)**	58,996	8.47
	Moderate to Unhealthy for Sensitive Groups	0.062 (0.050)	37,60	4.619
	Unhealthy for Sensitive Groups to Unhealthy	-0.031 (0.033)	558	9.927
<b>Outside MSAs with pop. &gt; 350,000</b>	Good to Moderate	0.037 (0.022)*	29,725	6,762
	Moderate to Unhealthy for Sensitive Groups	-0.033 (0.070)	1,003	6,856
	Unhealthy for Sensitive Groups to Unhealthy	0.241 (0.243)	184	7.778

\*  $p\text{-value} < 0.1$ \*\*  $p\text{-value} < 0.05$ \*\*\*  $p\text{-value} < 0.01$ 

AQI stands for air quality index. AQI values between 0 and 50 are categorized as Good, between 51 and 100 as Moderate, between 101 and 150 as Unhealthy for Sensitive Groups, and between 151 and 200 as Unhealthy.

Dependent variable is residual of sentiment index regressed on weather variables (possibly excluding visibility), county fixed effects, and date fixed effects, possibly population-weighted by the square root of the number of unique tweeters per county/day.

The coefficients were estimated using a local polynomial of degree 2. Robust standard errors are clustered at the county level. Bandwidths were estimated so as to optimize the common mean-squared error on both sides of the given cutoff point. MSA stands for Metropolitan Statistical Area, and pop. stands for population. All county information is from 2010. Weekends include Saturday and Sunday. Holidays include all federal holidays in the given time period as well as Halloween

**Table S6:** Estimated WLS coefficients of air pollution and weather variables

Variable	1	2	3	4	5	6	7	8	10	11
PM <sub>2.5</sub>	-0.00041 (0.00015)***	-0.00032 (0.00011)***	0.00023 (0.00017)	0.00035 (0.00012)***	0.00011 (0.00012)	0.00028 (0.00013)**	0.00028 (0.00016)	0.00007 (0.00008)	0.00022 (0.00008)***	0.000213 (0.00011)*
PM <sub>2.5</sub> <sup>2</sup>	--	--	--	--	--	--	0.00000 (0.00002)	--	--	0.00000 (0.00000)
Visibility	--	--	--	--	--	0.00425 (0.00049)***	0.00425 (0.00049)***	--	0.00353 (0.00035)***	0.00354 (0.00035)***
Temperature	--	--	--	--	0.0031 (0.00036)***	0.00321 (0.00036)***	0.00321 (0.00036)***	0.00226 (0.00029)***	0.00235 (0.000285)***	0.002345 (0.00029)***
Temperature <sup>2</sup>	--	--	--	--	-0.00002 (0.00000)***	-0.00003 (0.00000)***	-0.00003 (0.00000)***	-0.00002 (0.00000)***	-0.00002 (0.00000)***	-0.00002 (0.00000)***
Pressure	--	--	--	--	0.00069 (0.00015)***	0.00058 (0.00015)***	0.00058 (0.00015)***	0.00050 (0.00009)***	0.00039 (0.00009)***	0.00039 (0.00009)***
Wind Speed	--	--	--	--	-0.00141 (0.00022)***	-0.00151 (0.00022)***	-0.00151 (0.00022)***	-0.00084 (0.00015)***	-0.00090 (0.00015)***	-0.00090 (0.00015)***
Precipitation	--	--	--	--	-0.01499 (0.00163)***	-0.01138 (0.00167)***	-0.01138 (0.00167)***	-0.01123 (0.00108)***	-0.00809 (0.00109)***	-0.00809 (0.00109)***
Snow Depth	--	--	--	--	0.00058 (0.00413)	0.00149 (0.00413)	0.00149 (0.00413)	0.00157 (0.00222)	0.00299 (0.00218)	0.00299 (0.00218)
County FE	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pop Weighted	No	No	No	No	No	No	No	Yes	Yes	Yes
R <sup>2</sup>	0.00009	0.08783	0.03838	0.12543	0.12685	0.12718	0.12718	0.23024	0.23061	0.23061

\*  $p\text{-value} < 0.1$ \*\*  $p\text{-value} < 0.05$ \*\*\*  $p\text{-value} < 0.01$ 

AQI stands for air quality index. FE stands for fixed effects. Pop stands for population. The dependent variable is the sentiment index, which can be described as the median tweeters probability of posting a positive tweet given that they post a tweet. All models include county fixed effects and date fixed effects and are population-weighted by the square root of the number of unique tweeters per county/day. Robust standard errors are clustered at the county level. Observations are at the county-day level.

**Table S7:** Estimated WLS coefficients of visibility and other weather variables

Variable	1	2	3	4	5	6	7
Visibility	0.00325 (0.00064)***	0.00394 (0.00045)***	0.00397 (0.000683)***	0.00481 (0.00046)***	0.00408 (0.00035)***	0.00405 (0.00048)***	0.00338 (0.00035)***
Temperature	--	--	--	--	--	0.00316 (0.00036)***	0.00230 (0.000283)***
Temperature <sup>2</sup>	--	--	--	--	--	-0.00002 (0.00000)***	-0.00002 (0.00000)***
Pressure	--	--	--	--	--	0.00059 (0.00015)***	0.00039 (0.00009)***
Wind Speed	--	--	--	--	--	-0.0016 (0.00022)***	-0.00099 (0.00014)***
Precipitation	--	--	--	--	--	-0.01192 (0.00165)***	-0.00850 (0.00107)***
Snow Depth	--	--	--	--	--	0.00146 (0.00413)	0.00291 (0.002178)
County FE	No	Yes	No	Yes	Yes	Yes	Yes
Date FE	No	No	Yes	Yes	Yes	Yes	Yes
Pop Weighted	No	No	No	No	Yes	No	Yes
R <sup>2</sup>	0.00026	0.08817	0.03871	0.12589	0.22952	0.12719	0.23060

\*  $p$ -value < 0.1\*\*  $p$ -value < 0.05\*\*\*  $p$ -value < 0.01

AQI stands for air quality index. FE stands for fixed effects. Pop stands for population. The dependent variable is the sentiment index, which can be described as the median tweeters probability of posting a positive tweet given that they post a tweet. All models include county fixed effects and date fixed effects and are population-weighted by the square root of the number of unique tweeters per county/day. Robust standard errors are clustered at the county level. Observations are at the county-day level.



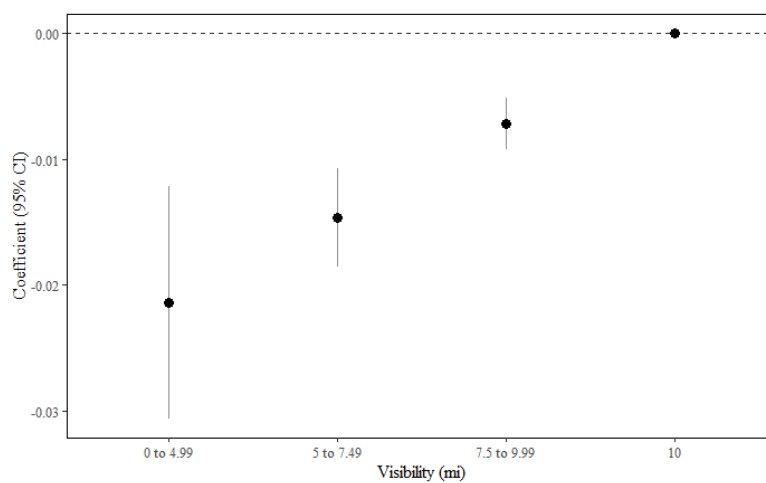
**Table S8: Stratified WLS results for PM<sub>2.5</sub>, weather variables**

Variable	Non-holiday	Holiday	Weekday	Weekend	MSA	Urban	Rural
PM <sub>2.5</sub>	0.00027 (0.00008)***	-0.00069 (0.00041)	0.00019 (0.00009)**	0.00026 (0.00011)**	0.00014 (0.00009)	0.000352 (0.00017)**	0.00079 (0.00071)
Visibility	0.00358 (0.00036)***	0.00298 (0.00175)	0.0034 (0.00039)***	0.00400 (0.00058)***	0.00307 (0.00041)***	0.00448 (0.00073)***	0.00582 (0.00220)***
Temperature	0.00239 (0.00029)***	0.00229 (0.00098)**	0.00313 (0.00033)***	0.001456 (0.00035)***	0.00164 (0.00035)***	0.00207 (0.00053)***	0.00446 (0.001412)***
Pressure	0.00040 (0.00009)***	0.00015 (0.00037)	0.00032 (0.00001)***	0.00040 (0.00012)***	0.00029 (0.00011)***	0.00040 (0.00019)***	0.00082 (0.000445)***
Wind Speed	-0.00087 (0.00015)***	-0.00101 (0.00079)	-0.00101 (0.00017)***	-0.00060 (0.00023)**	-0.00068 (0.00017)***	-0.00149 (0.00034)***	-0.002678 (0.00101)***
Precipitation	-0.00831 (0.00112)***	0.00121 (0.00627)	-0.00532 (0.00124)***	-0.01477 (0.001834)***	-0.0063 (0.00116)***	-0.01867 (0.00284)***	-0.01743 (0.01290)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pop Weighted	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	246,502	8,348	180,076	74,774	130,392	109,853	12,480
R <sup>2</sup>	0.22368	0.41542	0.22086	0.22397	0.32147	0.12526	0.16027

\*  $p\text{-value} < 0.1$ \*\*  $p\text{-value} < 0.05$ \*\*\*  $p\text{-value} < 0.01$ 

FE stands for fixed effects. Pop stands for population. The dependent variable is the sentiment index, which can be described as the median tweeters' probability of posting a positive tweet given that they post a tweet. All models include county fixed effects and date fixed effects and are population-weighted by the square root of the number of unique tweeters per county/day. Robust standard errors are clustered at the county level. Observations are at the county-day level. MSA stands for Metropolitan Statistical Area. All county information is from 2010. Weekends include Saturday and Sunday. Holidays include all federal holidays in the given time period as well as Halloween

**Figure S1:** Estimated shape of visibility and sentiment index association



Visibility refers to ground visibility censored at 10 miles. Coefficients were estimated using a weighted linear regression including  $PM_{2.5}$ , weather variables, and county and date fixed effects. Robust standard errors are clustered at the county level