From Post to Policy: Using Social Media Data to Inform Decision-Making

by

Nicolas Guetta-Jeanrenaud

Diplôme d'Ingénieur, École des Mines de Paris (2017)

Submitted to the Institute for Data, Systems, and Society
in partial fulfillment of the requirements for the degree of

Technology and Policy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2022

© Massachusetts Institute of Technology 2022. All rights reserved.

Author .....
Institute for Data, Systems, and Society
January 31, 2022

Certified by .....
Siqi Zheng
STL Champion Professor, Department of Urban Studies and Planning
Faculty Director, Sustainable Urbanization Lab
Thesis Supervisor

Certified by .....
Juan Palacios
Post-Doctoral Researcher, Sustainable Urbanization Lab
Thesis Supervisor

Accepted by .....
Noelle Eckley Selin
Professor, Institute for Data, Systems, and Society and
Department of Earth, Atmospheric and Planetary Sciences
Director, Technology and Policy Program

From Post to Policy: Using Social Media Data to Inform Decision-Making

by

Nicolas Guetta-Jeanrenaud

Submitted to the Institute for Data, Systems, and Society
on January 31, 2022, in partial fulfillment of the
requirements for the degree of
Technology and Policy

Abstract

While researchers and policy-makers traditionally rely on survey methods as they seek to understand preferences, user-generated data on social media—coupled with advanced methods of Natural Language Processing—can, in certain cases, serve as a valid alternative. In this thesis, I introduce a novel data set of global social media content and present a multilingual algorithmic method of text analysis which provides valuable insights into population well-being and public opinion at a global scale. I conduct three validation tests to assess the extent to which metrics computed from social media data are consistent with more traditional methods of measurement such as census population counts, well-being surveys, and political polls.
I go on to present two case studies which rely on social media-based metrics. In the first, we evaluate the effect of temperatures on subjective well-being worldwide. We find a non-linear, inverse U-shaped relationship and estimate high-temperature damages in a large selection of countries. In the second, we connect subjective perception of climate events with real estate market outcomes. We find that while objective temperature stress is consistently associated with lower location value, regions where sentiment is most sensitive to climate discomfort are also the ones where these shocks are the strongest. Both empirical studies confirm the strong potential of social media data for policy-makers and researchers alike.

Thesis Supervisor: Siqi Zheng
Title: STL Champion Professor, Department of Urban Studies and Planning
Faculty Director, Sustainable Urbanization Lab

Thesis Supervisor: Juan Palacios
Title: Post-Doctoral Researcher, Sustainable Urbanization Lab

Acknowledgments

First and foremost, thank you to Professor Siqi Zheng for over two years of support and guidance. Her ambitious vision and leadership, as well as her openness to new ideas and techniques, have inspired and empowered me during my time at MIT.

Juan Palacios was instrumental in helping me bring this thesis together. I've consistently benefited from both his insightful professional feedback and informal mentorship. If (and when) I become a skilled negotiator, I'll know who to credit!

Thank you to Jianghao Wang for giving me a chance to get involved in the bold "Global Sentiment" project, and for always placing my learning first in our collaborations. I extent my gratitude to the entire MIT Sustainable Urbanization Lab: it's been a privilege working with such a diverse, complementary, and talented group of people. Special thanks to Yuchen and Yichun for contributions to so many projects I was involved in, and for fostering such a great vibe in the RA lounge.

Many collaborators have supported the work I present in this thesis. Thank you to Harvard CGA—and especially to Devika Kakkar, Ben Lewis, and Wendy Guan—for hosting me as an affiliate, and for your incalculable contributions to our data collection and analysis efforts. Thank you to Miguel Neto, Fátima Neves, and Mauro Pereira for supporting a semester-long research sprint with such vigor and good humour.

None of this would have been possible without the Technology & Policy Program. A special thank you to Noelle Selin, Frank Field, and Barb DelaBarre for making TPP such a welcoming and formative experience.

A number of people, from MIT and elsewhere, have made these last few years especially thrilling and rewarding. Thank you to Axelle, Boyu, Patrick, Tristan, and many other TPPers; Domdom; Rick and Jen; Ugo, Alison, and more; Verity... I'm so grateful for their encouragement, stimulating conversation, and exceptional generosity. Thank you to Léa, for supporting me so far as to discovering a recent passion for BERT; but especially for making New York, and then Boston, your second home.

Beyond my interests in technology and social sciences, my parents have instilled in me confidence, optimism, and open-mindedness which, I hope, permeate my work. Finally, thank you to my brother Lio, my biggest source of inspiration and laughter.

Contents

List of Figures

List of Tables

Chapter 1

The Advent of Social Media: Opportunities and Challenges for Policy-Makers

The rise of social media worldwide has radically altered the ways in which we communicate, interact, and inform ourselves. Every day, active social media users share content on their favorite platforms, thereby generating data which reflects their thoughts at a specific point in time. Simultaneously, advances in the field of Machine Learning (ML) and Natural Language Processing (NLP) have made the automated analysis of large quantities of text data possible.

Policy-makers and researchers alike have always been interested in understanding public opinion, measuring well-being and satisfaction, and collecting feedback to events or interventions. And while they have traditionally relied on surveys to do so, these methods present important limitations. As social media is increasingly used across the world, the digital data generated on these platforms has become a valuable, alternative resource to understand public opinion and subjective well-being.

This introductory chapter frames the extent to which social media data can be used for impactful research that informs policy decisions. I start by describing limitations of traditional data collection approaches (Section 1.1.1), then present the relative advantages of social media data (Section 1.1.2.1). Social media data is not immune to problems of its own (Section 1.1.2.2)—but important validation studies have confirmed its relevance as a proxy measure for satisfaction and well-being (Section 1.1.2.3). Finally, I review the growing computational social science literature that uses social media data for analyses (Section 1.2), and present the contributions of my thesis to the field (Section 1.3).

1.1 Social media data as a source of public opinion

1.1.1 Limitations of traditional survey methods

Policy-makers and researchers have always looked to measures of subjective well-being or satisfaction to assess the value of public goods or evaluate policy alternatives [78, 139, 48]. Traditionally, they have relied on self-reported metrics collected by surveys. Survey methodology relies on representative samples to collect a comprehensive overview of the opinions shared by a population of interest [57]. However, surveys face a number of challenges today. They are an expensive enterprise, and are lengthy to run. As technological shifts have progressively replaced telephone landlines with mobile devices, and have democratized caller screening, response rates have declined sharply [34]. Non-response rates make it hard to construct representative samples, potentially biasing the survey results [58, 82].

Since surveys are an intentional data collection process, they are constrained by the questions they pose and the ways in which they are formulated [57]. Survey wording has been found to influence survey results [119]. Desirability bias—where respondents answer survey questions in ways that they believe agree with the organization that conducts them—also has the potential of skewing results, especially when it comes to sensitive topics such as mental and physical health or political preferences [127, 16].

More fundamentally, surveys can provide poor measures of real-time well-being, which can be difficult to disentangle from long-term happiness when analyzing self-reported responses [71]. Traditional surveys only reach out to a respondent pool once, making them inadequate at tracking individual-level changes in sentiment, behavior, and opinion [33]. Even longitudinal surveys, which offer a valid alternative for individual-level analysis, can't necessarily accurately time data collection phases around sometimes unpredictable events.

1.1.2 Social media as an alternative data source

1.1.2.1 Advantages of social media data

Given these limitations in survey methodology, social media offers an appealing alternative as a measure of public opinion. Data acquisition is cheap[1] [109], and analysis can be quasi-instantaneous. Social media platforms are "always on", continuously col-

[1]This is true notwithstanding the fixed costs to set up social media data harvesting systems, and the computational costs linked to transforming data for analysis.

lecting user-generated information [117]. Instead of requiring researchers to specify when and where to conduct surveys, social media facilitates the study of a wide range of unexpected events after-the-fact. Social media data is by nature unsolicited, and therefore not prone to the same response and question-wording biases as survey answers [31]. Add to that the fact that data collection can be quasi-instantaneous, and social media can be seen as a real-time sensor of human activity and conversations.

The most appealing aspect of social media data lies in the sheer quantity of data provided. Twitter, for instance, has over 330 million monthly active users across the world[2] [128]. Weibo, a microblog platform used mostly in China, reported 462 million monthly active users in 2018 [138]. A significant subset of these posts are geolocated, meaning that precise location information is collected at the time of the post (for more information about geolocation, see Section 2.1). The underlying data sets of social media posts have a geographical and temporal density far superior to even the largest-scale surveys. This level of coverage allows for studies with precise granularity, all the while remaining relevant at a high level as well.

While this scale of data might have been previously impossible to analyze, recent developments in the fields of computer science, ML, and NLP have made it possible to automatically extract information from social media posts. For instance, quantifying expressed sentiment in text—a field known as sentiment analysis—can be conducted by dozens of openly available models [83].

1.1.2.2 Limitations to the use of social media data

Beyond the important question of privacy and data ownership [12], social media data is not necessarily a perfect substitute for surveys. First off, it presents the risk of being inaccurate: logics of expansion have pushed social media platforms to prioritize ease and ergonomics over accuracy in user sign-up. Users can usually provide false names, location, or demographic information. According to one report, 46% of social media users have multiple accounts on at least one platform [116].

While traditional surveys strive to construct representative samples of the population, social media user pools guarantee nothing similar. Giant social networks bring together millions of users across the world, but accessibility issues still restrict access for specific population groups. At a global scale, developed countries are overrepresented on Twitter [90], and there is a significant bias towards urban regions [62]. One recent review of Twitter users in Italy found them to be on average younger and more highly educated [131]. Similarly, Pew Research Center found US Twitter users

[2]As reported in Q4 of 2018.

to be more wealthy than the average [140]. People with disabilities have also been historically disenfranchised by the internet and digital platforms [114].

Just as surveys rely on important theoretical frameworks for developing questions, conducting interviews, and aggregating information [57], extracting information from social media data requires adequate technical tools and statistical methods. While much hope is placed on the use of NLP for automatic analysis of social media content, these methods might not capture subtleties of speech. Ceron and Negri (2016), for example, assert that natural language on social media can evolve over time, differ across topics, and make use of irony or other nuances that are not necessarily captured by NLP algorithms. Human input is regularly required to calibrate sentiment analysis model [26].

1.1.2.3 Validating the use of social media

Despite these challenges, social media's relevance for research is supported by a number of validation studies. Global mobility patterns observed using geotagged Twitter posts have been found to closely match global tourism statistics [61]. Within urban areas, spatial distributions of Twitter users reflect those observed with census or cell-phone data, according to Lenormand et al. (2014). Their validation is conducted using data from the municipalities of Barcelona and Madrid, at a high level of spatial (up to the square-kilometers) and temporal granularity. High correlation levels are observed between the three data sources on population concentration and temporal distribution patterns, and the three data sources yield similar mobility networks [77]. In Section 2.3, we use a similar approach to test spatial coverage and distribution of our social media data.

NLP-based sentiment analysis on social media data has also been shown to be a valid measure of subjective well-being. By collecting Facebook profile information from survey respondents, Settanni and Marengo (2015) find that automated methods of emotion detection[3] applied to publicly shared content yield similar results to self-reported levels of depression, anxiety, and stress [120]. At the aggregate level in the United States, county-level happiness surveys correlate strongly with NLP-based measures of sentiment [68]. We conduct a similar validation test using state-level happiness-survey results in Section 2.3.

Combining topic modeling with sentiment analysis also provides similar measures of subject-specific public opinion. In the field of politics, social media-based sentiment

[3]Settanni and Marengo (2015) rely on the LIWC software [106] to compute emotion scores to social media content.

on candidates closely tracks preference levels captured by traditional polls [97] or survey-based approval ratings [25]. Here too, Section 2.3 presents a similar analysis using 2020 US presidential election data to validate our sentiment measure.

1.2 Existing applications of social media research

Therefore, while the limitations of social media data are important to keep in mind—especially when disenfranchised groups are a critical component of a research project's agenda—validation work has confirmed the potential of social media data as a strong and reliable proxy of population well-being and public opinion. Subsequent research in the fields of computational social science and economics has used social media data to study a wide variety of phenomena, with direct insight for policy-makers [76].

The network structure of social media is at the source of important findings in social sciences. Park et al. (2018) use Twitter data to study the strength of long-range ties, for instance, with important implications regarding the spread of culture or epidemics [104]. Social media structures have also contributed to the understanding of political polarization [6]. Tracking content diffusion on social networks has been used to study structural virality [54], the spread of misinformation [134], or the diffusion of lexical changes [43].

Building on validation studies that find that social media is an accurate reflection for population concentration and movement, urban science research frequently relies on geotagged posts of users over time. Jurdak et al. (2015), for instance, examine the mobility patterns of social media users in Australia, characterizing both inner-city and between-city movement [69]. While socioeconomic indicators are not available on Twitter, Huang and Wong (2015) match user profiles to census data and compare (social media-based) activity patterns by (census-based) income levels [66]. Their empirical study in the Washington, DC, area highlights that poorer populations travel the greater distances between home and work. Llorente et al. (2015) use social media mobility patterns—as well as diurnal rhythm and communication styles—to predict regional unemployment rates, highlighting low-cost (and public) alternatives to traditional economic indicators [80].

An important body of research has focused on topic prevalence using classification techniques. Qian et al. (2015) use supervised Latent Dirichlet topic modeling to examine the prevalence of social events—such as the Occupy Wall Street movement or the Syrian Civil War—on Flickr [108]. Driss et al. (2019) collect data from Tunisian citizen-group Facebook pages and classify content into policy-relevant areas

[10]. Finally, Paul and Dredze (2014) use topic modeling to uncover public health discussions on Twitter, and track the prevalence of the seasonal influenza or allergies across the United States [105].

Other analyses focus on extracting sentiment from social media data [40]. In their seminal paper on Twitter-based sentiment, Golder and Macy (2011) use millions of public posts to emphasize high-level daily and seasonal trends and point out cultural differences [56]. Mitchell et al. (2013) correlate sentiment from geotagged Twitter data with dozens of demographic and health characteristics in the United States [91]. Sentiment analysis techniques on Twitter data have been used to track the spread of racism and hate speech [21] or levels of political trust [24].

Social media-based sentiment can inform government on reactions to policy implementations: in Italy, Ceron and Negri (2016) use Supervised Aggregated Sentiment Analysis to measure citizen satisfaction during two important policy reforms led by the Renzi government in 2014 and 2015 [26]. Wang et al. (2022) examine sentiment reactions to the first COVID-19 wave and to lockdown policies that were implemented at the time [135].

One application area of particular interest is the use of social media sentiment to measure well-being responses to climate amenities. Kryvasheyeu et al. (2016) use social media data to assess the damage of Hurricane Sandy, for instance [74]. Using sentiment measured on Twitter, Bailys (2020) finds that extreme temperatures are associated with significant drops in population well-being [9]. Wang, Obradovich, and Zheng (2020) find similar results using Weibo data in China [136]. Again in the context of China, Zheng et al. (2019) use social media data to measure the impact of air pollution on subjective well-being in China [146].

1.3 Thesis Contributions

The current thesis contributes to this growing body of work. First, I introduce a novel data set of social media content which combines geotagged Twitter data (with global coverage from 2015 to 2021) [79] and Weibo data (covering China between 2018 and 2021) [27]. Text from social media posts is encoded into machine-readable embeddings using a transformers model for text representation. My main methodological contribution is the elaboration of a multilingual sentiment imputation algorithm—which supports over 50 of the most common languages worldwide—to assign a sentiment score to every social media post. Post-level sentiment scores are aggregated to construct spatial and temporal indices. To support external validity of my subsequent

results, I conduct three tests assessing the extent to which metrics computed from social media data are consistent with more traditional methods of measurement such as census population counts, well-being surveys, and political polls. The data and methods, along with these validation studies, are presented in Chapter 2.

I go on to present two applications of the use of social media data. In Chapter 3, I use global social media data to examine the impact of extreme temperatures on well-being across the world. This study is made possible by the multilingual aspect of our sentiment imputation algorithm, and contributes to the literature by assessing well-being costs in less wealthy countries that are traditionally excluded from similar analyses. In Chapter 4, I examine the impact of subjective perception of temperature stress on real estate value. Here, we contribute to the literature by using sentiment damages—measured on social media—as an input to real estate value models, and by examining whether subjective perceptions drive real-world outcomes. I discuss the different results and conclude in Chapter 5.

Chapter 2

Natural Language Processing on a Novel Social Media Data Set

This chapter presents the rich social media data set that I use for the subsequent case studies of my thesis. Seven years of Twitter data (2015–2021) and four years of Weibo data (2018–2021), with full text, user information, and precise geolocation were incorporated into the different case studies. The full data is presented in Section 2.1. Relevant features (in particular, sentiment) are extracted from the social media posts using advanced methods of NLP, and aggregated to spatial and temporal indices—I present these methods in Section 2.2. Finally, before presenting our case studies, we would be remiss not to test the validity of our data and features as accurate proxies of public opinion. I present three frameworks for comparing our social media-based sentiment index to more traditional outcomes in Section 2.3.

2.1 A global data set of social media content

The Twitter data set was collected by the Harvard Center for Geographic Analysis (CGA) as part of the Geotweet Archive [79], a large scale project aimed at collecting geolocated posts, or tweets, over time. It was generously shared with the MIT Sustainable Urbanization Lab in the context of an ongoing collaboration. The Weibo data was collected by the MIT Sustainable Urbanization Lab [27].

Both data sets come with a variety of fields, summarized in Table 2.1. The main content field is the full post text, which includes emojis, hashtags, and URL links posted by the user. Twitter text fields are limited to 280 characters[1] (Weibo has no

[1]Until 2018, Twitter's character count limit was 140 characters. See Gligorić, Anderson, and West (2018) for a study of how doubling the character limit affected post quality [52].

such limitation). The unique user ID—which we have for both Weibo and Twitter—tracks users over time in the data, allowing us to conduct longitudinal studies or include user-level fixed effects in our analyses.

Since our data set is made up of only geolocated posts, each social media entry is assigned a precise latitude and longitude position. On Weibo, this information reflects the user's exact location at the time of the post, and is available for users who consent to sharing their location. Geolocated posts reflect 1% of all Weibo content shared in China [41]. On Twitter, precise location information was collected until 2019 for users who had enabled background GPS data collection. Starting in 2019, however, GPS tracking was disabled. Current location information on Twitter results from one of three possible mechanisms: (1) the user tags themselves at a point of interest (POI), thereby generating an approximate location; (2) the user shares a geolocated photo taken through the Twitter mobile application; and (3) the tweet is posted through a third-party application (such as Instagram) that collects precise geolocation. Hu and Wang (2020) provide a detailed analysis of the impacts of Twitter's policy change for researchers. While the new POI geotagging mechanism does increase noise on location, it is precise enough when data is aggregated to a unit of analysis larger than the city [65]. Based on a sample of 2018–2019 data, Baylis (2020) finds that the share of posts on Twitter where location is provided represents only 4.5% of all tweets, but also that this content is typically very similar to non-geotagged posts [9].

Variable	Description	Notes
`post_id`	Unique ID of the post	
`user_id`	Unique ID of the user	
`lat`	Latitude of user at time of post	For Twitter, location is approximate since 2018
`lon`	Longitude of user at time of post	For Twitter, location is approximate since 2018
`text`	Full post text	

Table 2.1: Social media data core fields

Based on the information provided by users, additional features are generated to facilitate subsequent analysis of the data. These imputed features are detailed in Table 2.2. Reverse geocoding techniques on the GPU-based OmniSci software[2] are used to locate the post within administrative bodies (country, administrative-1 region equivalent to the largest sub-national division, administrative-2 region equivalent to the second-largest sub-national division). From the text field, we impute the language of the content, and we restrict to the 50 languages most used in the data (covering more than 99% of the content for which language is imputable). We also impute a

[2] `https://www.omnisci.com/`

sentiment score based on an NLP method described in the following section (Section 2.2).

Variable	Description	Notes
`language`	Post text language	Imputed based on post text
`country`	Country of user	Imputed based on latitude and longitude
`admin1`	Administrative-1 region of post	Imputed based on latitude and longitude
`admin2`	Administrative-2 region of post	Imputed based on latitude and longitude
`sentiment`	Sentiment score	Imputed based on post text and language

Table 2.2: Social media data imputed features

The main novelty of these data sets is the extent of geographic coverage they allow. As indicated in the introductory chapter, social media data has already been used for research—but these projects have mostly focused on data-rich countries like the United States [91, 9], Europe [26], or China [146, 136]. Here, we collect observations from across the world into a single, comprehensive, global data set. Fig. 2-1 illustrates the data coverage using a sample of data from 2020. Content is posted across the world, and dense areas of coverage can be found in all inhabited continents worldwide. However, regions with lower social media penetration are also clearly visible: rural, developing, and non-English speak countries all present lower Twitter and Weibo use. A more extensive analysis the geographic coverage of our data set is provided in Section 2.3.1.

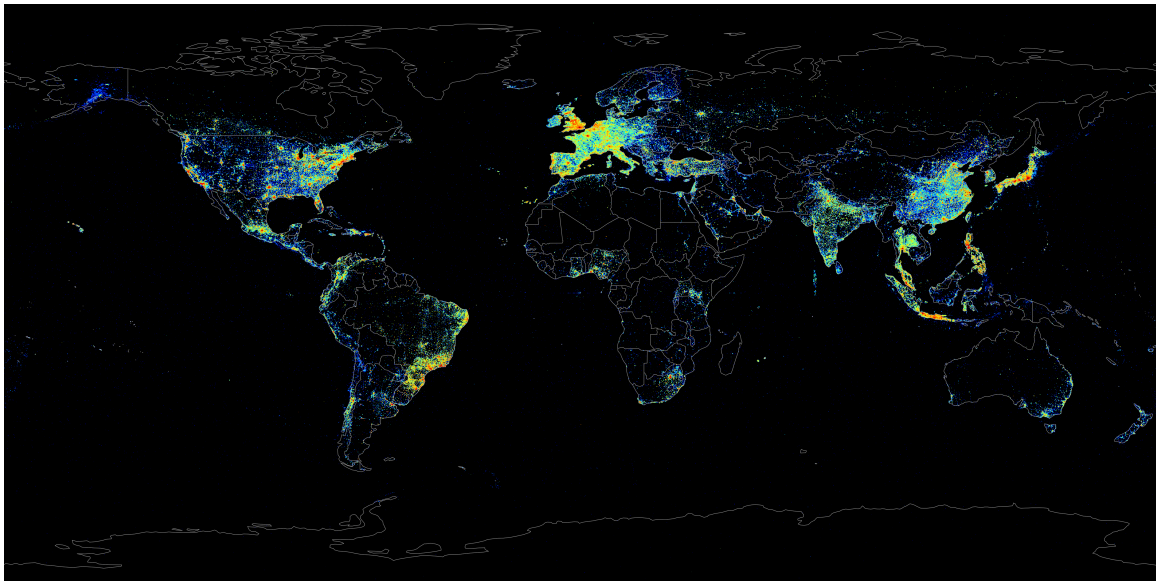

Figure 2-1: Social media data set geographic coverage (2020)

2.2 Natural Language Processing and sentiment imputation

A rich data set of social media offers important research possibilities. Tracking sentiment and topic distribution over time can inform on population well-being or on public opinion. Given the quantity of data at hand, however, these analyses require computational methods of NLP, which extract semantic characteristics of text at a large scale. The first step is to convert the text into machine-readable representations which are indicative of semantic meaning: this is done in Section 2.2.1 using the Transformer model BERT. From these representations we can extract topical information, but also sentiment scores by training a classifier on labeled data (Section 2.2.2). Finally, we aggregate sentiment features to relevant spatial and temporal units of analysis using methods detailed in Section 2.2.3.

2.2.1 Semantic representations of social media posts using the BERT Transformer model

One of the main objectives of NLP is converting human-readable text into a machine-readable numerical sequence. This process is known as text representation and can be done in a variety of ways. Trivial text representation, for instance, is conducted based on word occurrence in the text: a sentence might be encoded into a list of 1's and 0's based on whether a list of given words are in the sentence or not. Resulting representations can be thought of as a vector of coordinates for the text, within a high dimensional space called the embedding space.

Representations are useful for classification only to the extent that they reflect semantic meaning. Texts that have similar representations (i.e., are nearby in the embedding space) ought to have similar semantic meanings, and vice versa. While the word-occurrence based model described in the previous paragraph might seem like an appealing choice (sentences that are made up of the same words will have the same coordinates in the embedding space), there are important shortcomings: the method does not account for synonyms, word order, or sentence construction.

Training relevant representation models for classification has become an important field of research in NLP [3]. Moving beyond word-occurrence models, the *word2vec* framework first used neural networks to learn word representations which account for synonymy and semantic similarities. The neural network is trained on a large text corpus, and representations are based on a words "context", or surrounding words

[88, 89]. Transformer models, which have emerged in recent years, allowed text representation models to increase the width of the context window, vastly improving performance by accounting for long-range interactions between words [132]. In these models, word representations depend on their entire sentence, not just their nearest neighbors.

Bidirectional Encoder Representations from Transformers, or BERT, is a word representation method developed by Google in 2018 [38]. BERT reads in sentences as sequences of words, preceded by a start-of-sentence token referred to as the [CLS] token. A Transformer architecture assigns contextual embeddings to words based on how frequently they appear in the same sentences as other words within the training corpus[3]. Dozens of pre-trained BERT-based models are hosted (and openly shared) on the *HuggingFace* Model Hub [141]. The Python *SentenceTransformers* framework greatly facilitates the use of BERT for text representation by sequencing input sentences, imputing BERT representations, and pooling these representations into a single high-dimensional embedding representative of the entire initial text [112]. An illustrative diagram of the BERT-based text representation process, which we apply to each social media post, is provided in Fig. 2-2.

[3]BERT models are usually pre-trained on millions of data entries from BooksCorpus and Wikipedia.

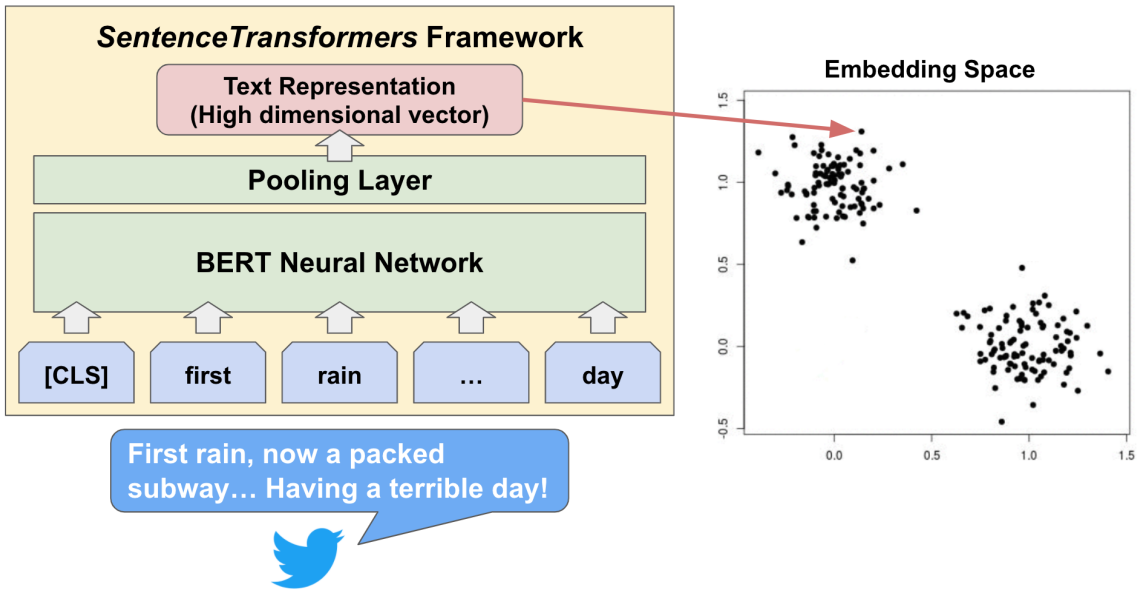

Figure 2-2: Text representation using BERT and the *SentenceTransformers* framework

Social media posts (bottom left of the diagram) are converted into a sequence of word tokens, starting with a [CLS] token. The BERT neural network reads in the sequence of tokens, and imputes embeddings to each. A pooling layer combines the different embeddings into a single representation. The representation can be seen as the coordinates of the post within the embedding space. Within the embedding space, similar-meaning posts appear nearby.

One of the main characteristics of BERT models (and therefore of the *Sentence-Transformers* framework) is that it allows for multilingual text encoding, meaning that direct translations across different languages are encoded similarly by multilingual modules of BERT [107]. Therefore, once text is converted into the BERT embedding space, a single classification model can be used for prediction regardless of the underlying language. Currently, BERT is known to support 50 languages.

Using a pre-trained multilingual BERT model, we create representations of every social media post in our data set. These representations are stored as a set of post-level features, and feed into models that we use throughout this thesis. Within the embedding space, clustering based on representation coordinates groups together content of similar semantic meaning—this can be used for topic modeling, for instance. Coordinates in the embedding space can also serve as inputs to train a supervised classifier: we use this property to train the sentiment analysis model described in the next section.

2.2.2 Sentiment analysis of social media data

Our sentiment analysis model builds on the post-level BERT representations that we previously computed, using the dimensions of the representations as a set of features for supervised classification. We train our classifier on the representations of a set of labeled Twitter posts, then predict sentiment scores of the unlabeled social media posts of our data set.

The training data we use is *Sentiment140*, an English-language sentiment-labeled data set of 1.6 million Twitter posts [53]. The data set was constructed by imputing the sentiment of every tweet based on the occurrence of positive or negative emojis with the text—allowing the imputation of sentiment scores at a large scale with limited human intervention[4]. We compute the S-BERT embeddings of every observation of the *Sentiment140* data set, then train a classifier on 80% of the data (training set). Our classifier is a Machine Learning pipeline made up of a dimensionality reduction step (we keep the first 100 principal components of the representations), and a logistic regression model. The output can take the form of either a binary prediction (0 for negative, 1 for positive), or of a continuous score (between 0 and 1, equal to the estimated probability that the post is positive). For validation, we run binary predictions on the training set, as well as on the remaining 20% of the data (test set). Performances are assessed using simple accuracy[5], and provided in Table 2.3.

Data set	Nb of Observations	Accuracy
Training set	1,280,000	0.81
Test set	320,000	0.80

Table 2.3: Sentiment analysis model performance on the training and test sets

To check for potential overfitting of the classifier on the *Sentiment140* data set, we also evaluate our model's performance on an alternative English-language labeled data set provided by CrowdFlower[6]. Although our model is trained on English-language data, it predicts in over 50 languages supported by multilingual BERT. We test for potential linguistic bias by evaluating the performance of our model in

[4]This training data set, constructed by Go, Bhayani, and Huang (2009) relies on the assumption that sentiment in the text will reflect sentiment in the emojis. After being used to impute the sentiment, emojis are removed from the data set's text field, and the sentiment score is associated to the remaining text entry. We too remove them from our data as part of our data cleaning steps.

[5]Model accuracy is defined as the number of correct predictions over the total number of predictions.

[6]We use the "Sentiment Analysis: Emotion in Text" data set.

different languages, including on sentiment-labeled Portuguese data [17], and similar data in European languages [94]. Model performance on the alternative English data set and on foreign languages is presented in Table 2.4.

Language	Source	Nb of Observations	Accuracy
ALBANIAN	Mozetič et al., 2016	1,866	0.722
BOSNIAN	Mozetič et al., 2016	1,872	0.788
BULGARIAN	Mozetič et al., 2016	1,101	0.723
CROATIAN	Mozetič et al., 2016	6,890	0.819
ENGLISH	CrowdFlower	3,152	0.840
GERMAN	Mozetič et al., 2016	1,865	0.813
HUNGARIAN	Mozetič et al., 2016	4,071	0.767
POLISH	Mozetič et al., 2016	11,049	0.766
PORTUGUESE	Brum & Nunes, 2017	15,047	0.750
RUSSIAN	Mozetič et al., 2016	3,592	0.748
SERBIAN	Mozetič et al., 2016	345	0.643
SLOVAK	Mozetič et al., 2016	6,154	0.819
SLOVENIAN	Mozetič et al., 2016	5,927	0.761
SPANISH	Mozetič et al., 2016	9,247	0.720
SWEDISH	Mozetič et al., 2016	2,590	0.725

Table 2.4: Sentiment analysis model performance by language

Traditionally, sentiment or emotion indices are conducted using dictionaries-based approaches, such as LIWC [106]. In these models, lists of words that are pre-established as positive or negative are matched to words in the text content, and scores are established based on match counts. These methods are highly dependent on the aforementioned word lists and on the specific language of analysis, and are therefore not adapted to global, multilingual studies. Prior research has also found that, when analyzing sentiment and measuring well-being, ML-driven models [68] and BERT-based models in particular [126] perform better than more trivial word-based methods such as LIWC.

However, as a measure of validation and as a means of building on more traditional sentiment-analysis research, we also compute LIWC-based sentiment scores on a sample of English-language posts in our Twitter data set. We test the correlation between BERT-based and LIWC-based sentiment indices in Appendix Section A.2. We find that both indices are strongly correlated, with a Pearson's coefficient of

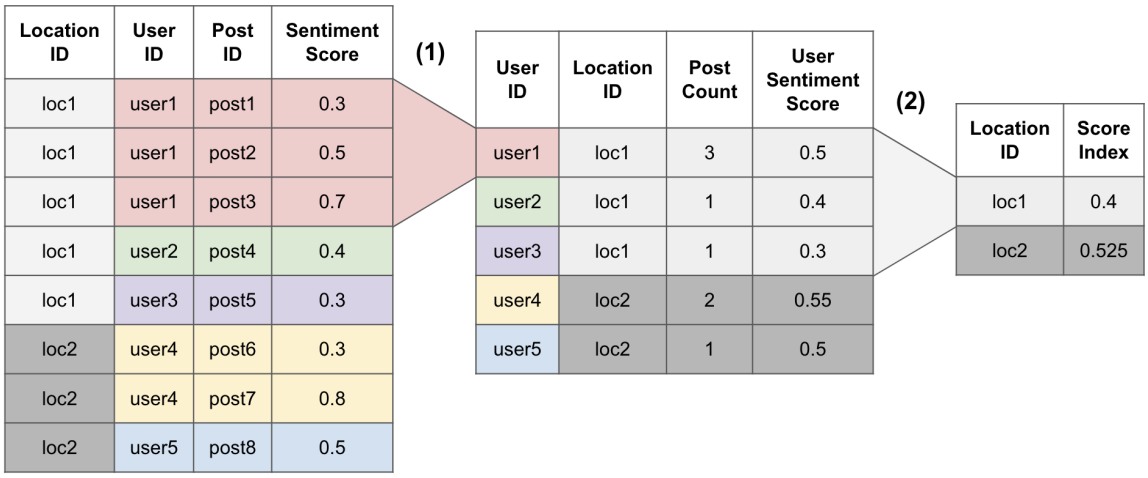

Figure 2-3: Two-step sentiment aggregation diagram

In step (1), we average the sentiment scores of every individual user into a user sentiment score. In step (2), we average user scores regardless of the number of posts shared by the user. This two-step aggregation method equally weighs social media *users* instead of *posts*, and avoids over-emphasizing frequent users.

$rho = 0.74$ ($p < 0.001$). Given the broad acceptance of LIWC as an accurate measure of expressed sentiment, our BERT-based index is a valid measure as well.

2.2.3 Aggregating post-level sentiment scores to sentiment indices

The trained classifier predicts a sentiment value on each post of our social media data set. Post-level sentiment scores are then aggregated into spatial and temporal indices. Since individuals are not equally active on social media, however, a simple average would overweight the most frequent posters in our sample. This becomes especially problematic with the rise of social media robots (or "bots"). These programmed mega-posters are usually driven by political motives, and have been found to spread misinformation [121] and disrupt democratic processes [13].

Instead, aggregation is conducted by weighting each post by the inverse-frequency of the user within the time-space unit of aggregation. Our aggregation mechanism can be understood as a two-step process where, for a given location l and time t, we (1) average each individual user's sentiment based on their posts during that time; and (2) compute the unweighted sentiment average of all users in that location. The process is illustrated in Fig. 2-3.

The two-step process can be formalized by the following formula:

$$sent_{l,t} = \frac{\Sigma_{j \in U_{l,t}} \frac{\Sigma_{i \in T_j} sent(tweet_{i,j})}{T_j}}{U_{l,t}} \tag{2.1}$$

where $sent_{l,t}$ is the sentiment index value of location l at time t. $U_{l,t}$ is the number of users in a location l during time t, T_j is the number of tweets by user j, and $tweet_{i,j}$ is the i-th social media post by user j.

2.3 Validating our social media data set and sentiment analysis approach

The generalization power of our social media data-based findings depends on the extent to which our measures proxy actual public opinion and well-being. While our results are indicative of the behavior of social media users, these might not reflect the general public. Verifying whether our results align with alternative, more traditional methods (i.e., census data, surveys, or polls) is essential for our results to be of use to policy-makers and researchers.

Here, I offer three potential avenues to test the generalizability of our results. First, I compare the coverage of our data set to population numbers across the world. I find strong correlations both globally (between different countries) and within countries at regional level. The last two validation approaches focus on validating our sentiment score by comparing it to other measures of satisfaction. First, I compare aggregate social media-based measures of well-being to survey-based "happiness scores" in the United States. Strong correlation between the two measures indicates that our sentiment measure is in fact capturing a similar measure of population well-being. Second, I compare topical aggregates of our sentiment score to election polls in the United States in 2020. State-level sentiment on content mentioning specific candidates ("Biden" or "Trump") is again found to be highly correlated to traditional public opinion polling measures.

2.3.1 Validating the spatial distribution of social media data

One important limitation of social media data is that it does not guarantee spatial representativity. The digital divide has historically led to disparities in internet access and digital platform penetration rates, especially between developing and developed nations or urban and rural areas [62]. However, the last decade has seen significant increases in social media use in developing countries [1].

We test the representativity of our social media data set by comparing the average number of daily social media users to official population statistics. Country-level population numbers are provided by the World Bank, based on national Census reports [143]. Our social media-based user count is the number of daily summed over the entire year users in each country. Since Twitter and Weibo user numbers are not necessarily comparable, we only use Twitter data for this analysis (we therefore exclude China from the results) and restrict to the year 2019.

Country-level population and social media user numbers are plotted in Fig. 2-4. The regression line, computed on the non-logged measures and plotted in black, highlights a positive relationship between the two measures. Disparities do emerge, however, between different continents. Observations above the regression line are countries where the population is disproportionately large compared to the social media user base: this group is almost entirely made up of developing countries from Africa and Asia. The furthest outliers include Ethopia (ETH), the People's Republic of North Korea (PRK), Bangladesh (BGD), and India (IND). In Ethiopia, for instance, current social media penetration rate is a whooping 983 times lower than in the United States (USA). However, the two measures are overall highly correlated, with a strong and significant Pearson correlation coefficient ($\rho = 0.34$, $p < 0.001$).

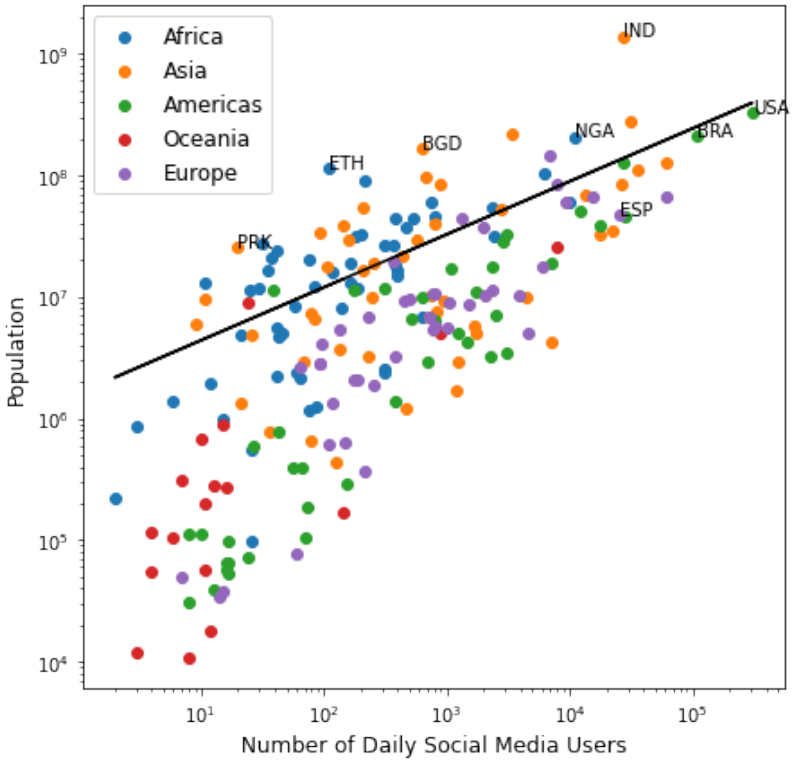

Figure 2-4: Number of social media users by country population, worldwide

Both measures are logged. A regression line is plotted in black. Observations are colored by the countries continent to facilitate the observation of geographic trends. The Pearson correlation coefficient between the two non-logged measures is $\rho = 0.34$ ($p < 0.001$).

To test for the prevalence of an urban-rural divide in our data, we run similar analyses within four case countries (United States, Spain, Brazil, and Nigeria). Here, we compare regional population to the number of social media users we observe in those sub-national divisions. Results are provided in Fig. 2-5. In highly developed, western countries, such as the United States and Spain, the correlation between social media penetration and population is almost perfect. Rare outliers, such as the District of Columbia, attest of exceptional situations—in this case, the prevalence of social media in American political communication. The relationship between social media users and population is noisier in developing countries like Brazil and Nigeria. Some outliers, such as the state of Bahia in Brazil, highlight poorer regions where digital penetration rates are still low. However, these countries still present high and significant correlation rates ($\rho = 0.60$ and $\rho = 0.56$ in Brazil and Nigeria, respectively). Comparing these rates to the global, country-level correlation rate we found above ($\rho = 0.34$) illustrates that the within-country digital divide is less of an ob-

stacle to homogeneous coverage than between-country inequalities. This informs our subsequent analyses, where we consistently estimate global results with, at minimum, country-level fixed effects.

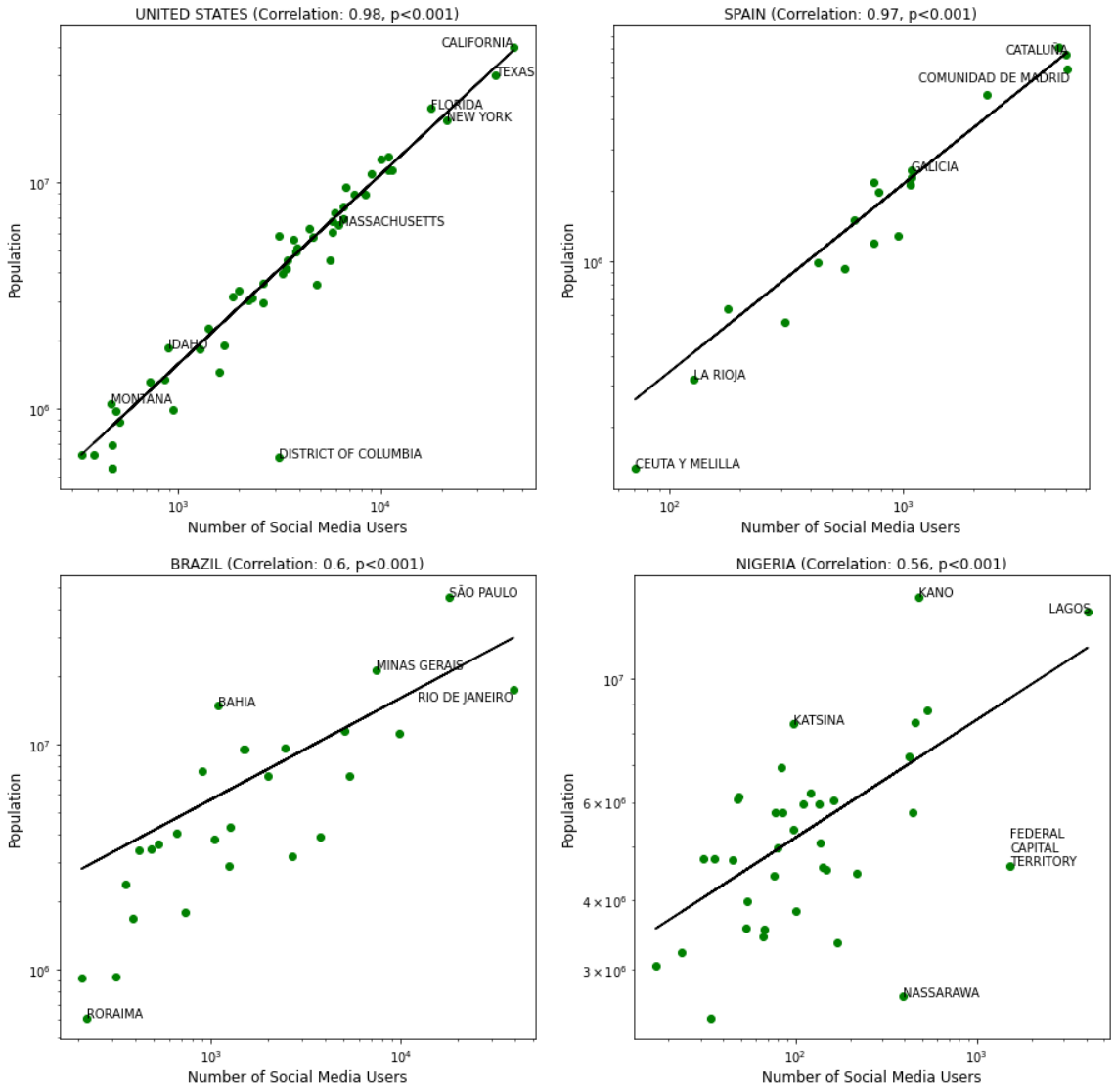

Figure 2-5: Number of daily social media users by regional population in four case countries

2.3.2 Validating sentiment scores based on happiness survey results

In addition to verifying the geographic coverage of our data, we attempt to validate the sentiment index we compute based on our NLP algorithm. The first generalizability test consists in comparing social media-based sentiment measures to traditional

surveys of happiness and satisfaction. The conceptual framework is presented in Fig. 2-6. Since we are interested in overall well-being for this validation, we aggregate sentiment scores from all social media data, regardless of the content's topic.

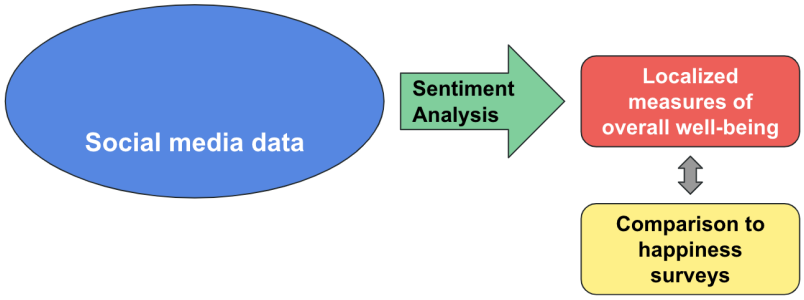

Figure 2-6: Conceptual framework for validating social media data with happiness surveys

Prior research has already documented a positive relationship between the two: Jaidka et al. (2020), for instance use the county-level Gallup-Sharecare Well-Being Index in the United States to test the accuracy of Twitter-based emotions using a number of different NLP models [68].

Here, we use the Gallup annual "State of the States" poll (2019 data), which assigns a score to US states according to their reported well-being. We create a Twitter-based alternative sentiment measure by aggregating the sentiment scores of every 2019 Twitter post at the state level. Consistently with the aggregation method described in the previous section, we avoid frequent user oversampling by weighting post scores by the inverse posting frequency of the user. Fig. 2-7 plots, for every US state, the survey-based well-being score by the Twitter-based sentiment measure. We also report the regression line in black: full regression results are provided in Appendix Section A.3.

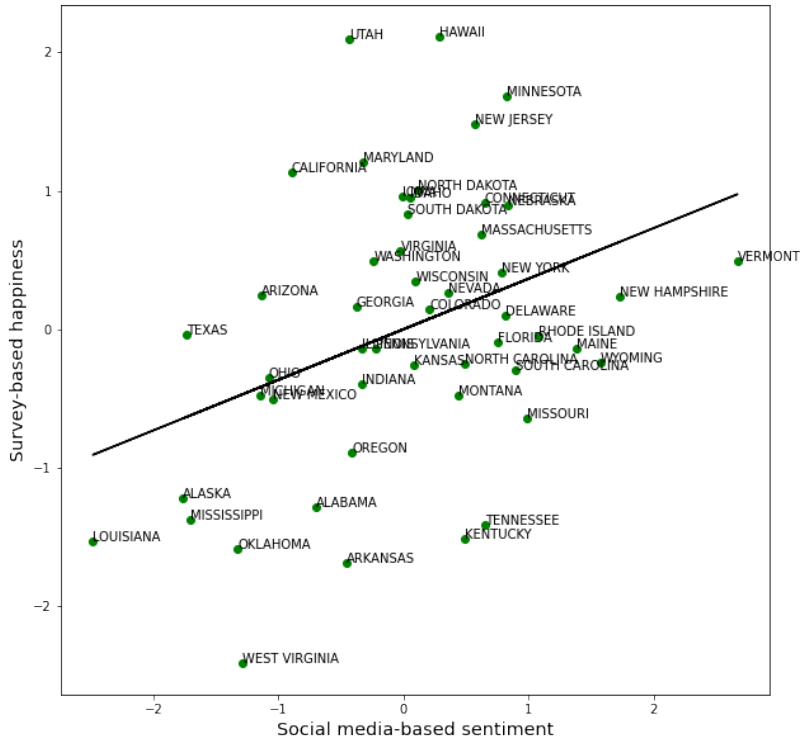

Figure 2-7: Social media-based well-being score by survey-based well-being score

Both sentiment scores are standardized. We find a strong and significant correlation between the two (Pearson correlation coefficient $r = 0.37$, $p < 0.01$).

We characterize the correlation between the two by calculating the Pearson correlation coefficient of the two scores. The coefficient we obtain is positive ($r = 0.37$) and significant ($p < 0.01$). Some of the outliers in the data (Maine, Alaska, West Virginia) are among the most rural states of the country. Others (Utah) are among the states where we noted the lowest number of social media users per capita in the previous section. Further analysis on these results could shed light on other factors that might influence the precision of our social media estimates. Overall, however, these results support the idea that our sentiment analysis model and the underlying social media data are a close approximate of overall well-being in the United States. The correlation we obtain is also similar to the results of Jaidka et al. (2020): using state-of-the-art NLP models, they achieve a Pearson correlation of $r = 0.51$ between US county-level survey and social media happiness [68].

2.3.3 Validating sentiment scores based on election polling

An alternative way of validating our social media sentiment score is by comparing it to topical surveys or public opinion polls. In this case, social media sentiment aggregates can be built off of topical content—or based on a subset of posts mentioning a specific topic. Aggregate sentiment of topic restrictions can be interpreted as measures of support for specific topics, instead of overall measures of well-being. An illustrative pipeline of the process is provided in Fig. 2-8.

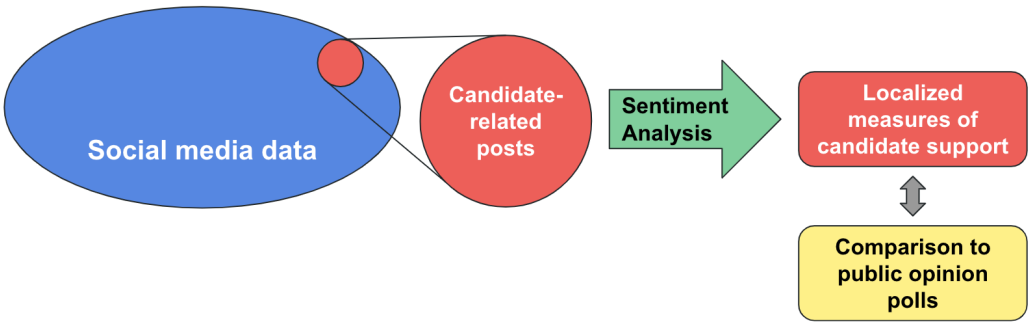

Figure 2-8: Conceptual framework for validating social media data with topical polls

In the context of the US presidential elections, hundreds of state-level polls are conducted regarding citizen preferences for candidates. These polls are referenced and aggregated by analytics websites like FiveThirtyEight[7]. FiveThirtyEight's presidential election forecast, which is based on polling aggregation from all of the largest US pollsters, has become something of a reference in the political world [30]. During the 2020 election year, the election forecast provided readers with national predictions, as well as final polling average for every state of the country [122]. We collect final polling averages, and compute a polling-based preference measure in state s, marked $polling_s$, such that:

$$polling_s = polling(Biden)_s - polling(Trump)_s \tag{2.2}$$

where $polling(Biden)_s$ and $polling(Trump)_s$ are final polling number for Biden and Trump, respectively. A positive polling score indicates that the state has more support for Biden than Trump. A negative polling score, on the other hand, indicates that Trump is polling above Biden in that state.

We construct an alternative measure of public preference by aggregating the candidate-specific sentiment of Twitter users in every US state over a 3-month period

[7] https://fivethirtyeight.com/

preceding the US elections (August 1st–October 31, 2020). Topical samples specific to a candidate are constructed using string matching in the raw tweet text field: in our sample period, 832,653 tweets (posted by 127,083 users) mention "Biden" and 2,935,239 tweets (posted by 236,754 users) mention "Trump". The distribution of users discussing Biden or Trump during our sample period—hereafter referred to as "topical users"—by state is provided in Appendix Section A.4 (Fig. A-2). State-level topical sentiment is constructed for both Biden-related and Trump-related tweets by aggregating the sentiment scores, with $sent(Biden)_s$ (respectively, $sent(Trump)_s$) the aggregated sentiment score of Biden (respectively, Trump) in state s. Again to avoid oversampling of frequent posters, we use the two-step aggregating method described in Section 2.2.3 and Equation 2.1. We finally compute a social media-based preference measure in every state s, marked $socialmedia_s$, such that:

$$socialmedia_s = sent(Biden)_s - sent(Trump)_s \tag{2.3}$$

A positive social media score indicates that Biden-related discourse is more positive than Trump in a given state. A negative social media score, on the other hand, indicates that social media users in that state are more positive in their discussions about Trump.

The two preference scores ($polling_s$ and $socialmedia_s$) are mapped on in Fig. 2-9. To allow for more intuitive comparison, we standardize both measures and examine the correlation by plotting the two scores in Fig. 2-10.

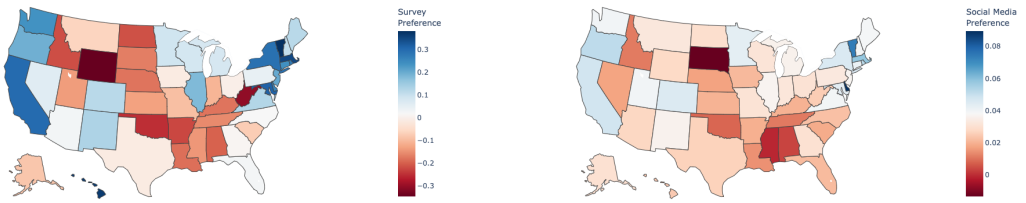

Figure 2-9: Mapping 2020 election preference scores

Survey-based 2020 election-preference score (left) and the social media-based 2020 election-preference score (right).

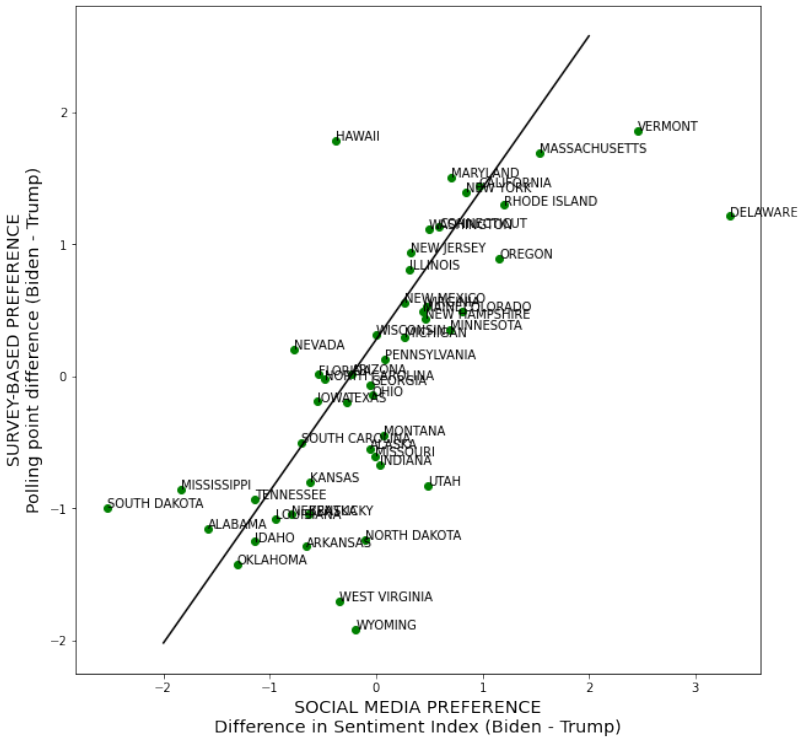

Figure 2-10: Social media-based preference by survey-based preference

Scatter plot of the social media-based preference score by the survey-based preference score. We find a strong correlation between the two (Pearson correlation coefficient $r = 0.72$).

We obtain a strong positive correlation of $r = 0.72$, and highly significant ($p < 0.001$). Some of the outliers we obtain are states with particularly low social media coverage—Delaware, Wyoming, South Dakota, and Hawaii, for instance, all have fewer than than 1000 topical users during our sample period. Other outliers we obtain are similar to the ones we had in our previous validation study (Utah, West Virginia). Again, poorer estimations in these states align with existing literature on the digital divide between urban and rural areas—however, further research is required to fully interpret these results.

To limit the impact of our noisy estimates on the overall results, we conduct our regression analysis by weighting the observations by the number of topical users in each state. The regression line is plotted in black Fig. 2-10, and full regression results can be found in Appendix Section A.4. The main coefficient of interest, reported in Table 2.5, is highly significant and the regression has an R^2 value of 0.74.

	Baseline 1: Tweet Counts	Baseline 2: Tweet Sentiment	Main: Sentiment Index
Description	Share of tweets on each candidate	Simple average of sentiment scores	Two-step weighted average of sentiment scores
Correl. coef.	-0.06	0.42**	0.72***
Reg. coef.	-0.23	0.73***	1.01***
Reg. R^2	0.02	0.50	0.74

Note: *p<0.01; **p<0.005; ***p<0.001

Table 2.5: Comparing social media-based preference methods

We use the Pearson coefficient for correlation. The regression model is run by weighting each state by the number of topical users. Baseline 1 (a non-NLP method which relies on counting topical tweets regardless of sentiment) is not correlated with the survey results. Baseline 2 (simple aggregation of sentiment scores that therefore overemphasizes frequent users) has lower correlation with survey results and lower regression R^2 than our main, two-step aggregation method.

Baseline levels of correlation can be provided by non-NLP methods of social media preference analysis—based on topical tweet counts, for instance. We define a baseline preference metric as the share of tweets in each state that are on Biden, among all tweets that are on Biden or Trump. The correlation between this metric and the survey-based preference is basically null (see the first column of Table 2.5, and see Section A.4 for full results). The sentiment component that our scores provide is therefore the main driver of the correlation between social media content and survey results.

We can also test the relevance of our two-step aggregation method: we compute a trivial sentiment aggregation on the same data (average of all tweet scores), which does not account for user posting frequency, and calculate the correlation with the survey-based preference scores. Results are also presented in Table 2.5 (second column). Here, the correlation is positive and significant ($r = 0.42, p < 0.01$), but lower than what we obtain using the two-step aggregation method. Regressions using this more trivial aggregation are presented in Appendix Section A.4: we obtain an R^2 coefficient of only 0.50, signaling that the two-step aggregation method reflects more accurately the state of public opinion and political preferences.

2.4 Chapter conclusion

This chapter presents the social media data we collect, the methods of feature extraction that we implement to analyse it, and makes the case for its use as a proxy for public opinion and well-being. We harness a novel data set of Twitter and Weibo

content, covering billions of posts globally, over several years. Our main methodological contribution is the NLP sentiment extraction algorithm, which relies on the state-of-the-art Transformer model BERT to provide sentiment scores to our social media posts in 50 languages. We aggregate the post-level sentiment scores to spatio-temporal indices using a two-step method that better accounts for user representation. Finally, we validate our approach by successfully correlating our social media-based indices to traditional survey measures of well-being and preferences.

The sentiment score data set that we constructed provides, we hope, future researchers with a relevant additional feature when using social media data. As part of our collaboration with Harvard CGA, we intend to make the sentiment scores available to researchers using the Geotweet Archive, and we have shared a procedure to access the data on the Dataverse [59]. In the following chapters, we lead the way in using the data and sentiment index in a series of case-study applications.

Chapter 3

Case Study: The global effect of temperature on subjective well-being

This chapter is derived from a manuscript co-authored with Jianghao Wang, Juan Palacios, Yichun Fan, Devika Kakkar, Nick Obradovich, and Siqi Zheng.

Abstract

Existing studies assessing the impact of climate change on well-being usually center on wealth, data-rich countries. However, these countries may differ substantially in how they cope with climate events. Here, we use a novel data set of 1.2 billion social media posts from 157 countries, coupled with daily meteorological data, to estimate how sentiment is affected by extreme temperatures. We rely on multilingual NLP to compute expressed sentiment in 50 identifiable languages. Combining these metrics in a fixed effect time-series regression model, we find that extreme temperature produce substantial drops in expressed sentiment: temperatures above 35°C reduce sentiment by 18.2% of a standard deviation (relative to moderate temperatures). The trend holds in most countries across the world, despite strong heterogeneity in the magnitude of the effect. Overall, our results corroborate the idea that that climate change harms mental well-being. Assessing these psychological factors and well-being costs is critical to properly adopting and targeting effective resiliency policies.

3.1 Introduction

Continued global warming increases the pervasiveness, frequency, duration, and intensity of extreme weather events, putting both earth systems and human societies in jeopardy [67]. A wealth of evidence demonstrates that warmer temperatures due to climate change have significant social impacts [22, 5]. Exposure to abnormally warm conditions increases the mortality burdens [137, 8, 133], lowers economic growth [20, 73], reduces agricultural productivity [95, 100], threatens food security [49, 60], and increases domestic violence [32]. Together, these disruptions to health, economics, and social stability may induce stress and anxiety, and cause damages to mental well-being.

To encourage policy action mitigating the threat of climate change, it is necessary to understand how it is subjectively perceived by people [39, 28, 123]. Though the impacts of abnormal climatic conditions on mental health and psychology are accumulating evidence [11, 29, 19], there remains limited quantitative evidence linking extreme temperatures to human well-being.

Recent research, already cited in Chapter 2, has harnessed social media data and NLP to quantify the effects of climate amenities on sentiment with unprecedented spatial and temporal resolution. Zheng et al. (2019) quantified the well-being impacts of air pollution in China [146]. Baylis (2020) and Wang, Obradovich, and Zheng (2020) find that higher temperatures worsen expressed sentiment in the US and China, respectively [9, 136]. For reasons of data availability, however, these studies have focused mostly on wealthier, data-rich countries with high social media penetration rates.

The global nature of climate change means that all countries are impacted by rising temperatures—not just developed nations. Numerous anthropological, cultural, political, and socioeconomic factors may also lead different regions to have dissimilar perceptions of climate change [64, 93, 55]. Generalizing estimates from data-rich societies to inform the potential impact of climate change globally is thus unsuitable. And while the existing literature is unanimous in its assessment of the negative impact of climate change on expressed sentiment, heterogeneities between countries on a global level—which might inform projected well-being costs—have yet to be investigated.

Here, we conduct the first global analysis of the impact of ambient temperature on expressed sentiment. We harness a sample of our rich geotagged social media data set, composed of 1.2 billion Twitter and Weibo posts shared between January 1, 2019, and December 31, 2019, to investigate how daily maximum temperatures alter

individual's expressed sentiment at a regional, country, and global scale. Social media posts, shared in 50 identifiable languages and in 156 different countries worldwide, are assigned sentiment score based on the sentiment imputation model described in Section 2.2.2. Sentiment scores are then aggregated to nearly 3000 daily location indices, corresponding to the largest sub-national units (or administrative-1 regions) of those countries. We couple this data with corresponding-level measures of daily temperature, as well as other environmental variables, in a fixed-effect regression estimator. The daily change of expressed sentiment in each locality is modeled as a non-parametric function of temperature exposure, and we include other environmental variables as flexible controls. Our data and model allow us to observe large spatial and temporal variations in temperature, and we assess how responses differ across different climate zones and countries.

3.2 Data

3.2.1 Temperature and environmental data

The temperature and environmental data are extracted from the NASA Modern-Era Retrospective Analysis for Research and Applications, or MERRA-2, project [50]. Based on atmospheric reanalysis of satellite data, this model provides hourly grid-level measures of a number of environmental variables. The native spatial resolution of MERRA-2 is 0.5° latitude x 0.625° longitude (approximately equivalent to 50km x 50km), meaning that the entire globe is divided into 576 x 361 grid boxes.

The MERRA-2 model provides three hourly measurements of temperature: maximum temperature, minimum temperature, and mean temperature. We maintain this distinction as we aggregate data to daily level: daily minimum temperature is defined as the lowest of the hourly minimum temperatures in that grid unit, daily maximum temperature is the highest of the hourly maximum temperature measures, and daily mean temperature is the average of the hourly mean temperature measures. For the purpose of this study, maximum temperature is most likely to reflect temperatures during the day-time, and these are the temperatures individuals are most likely to be exposed to. We therefore use the maximum daily temperature as our main measure of temperature, and we include the difference between the daily maximum and minimum temperature (or temperature range) as a control. We run robustness checks using minimum and maximum temperature as our main independent variable, to similar results (see Appendix Section B.3.6).

In addition, we collect a series of hourly environmental controls: precipitation, humidity, cloud coverage, wind speed, and $PM_{2.5}$ air pollution (particulate matter with an aerodynamic equivalent diameter of less than 2.5 μm). All are aggregated to daily grid-level, either by averaging over the hourly measures (humidity, cloud coverage, wind speed, $PM_{2.5}$ air pollution) or by summing them (precipitation).

Daily grid-level measurements are finally aggregated to our relevant geographic unit of analysis (global administrative-1 regions) by averaging the values of the grid boxes comprised in these regions. Overall, we obtain 930,033 daily regional meteorological measurements. Summary statistics of the administrative-1 level daily environmental measures are provided in Table 3.1.

Table 3.1: Daily Environmental Data Summary Statistics

Variable	Period	Obs.	Mean	Std. Dev.	Min.	Max.
Max Temperature (in °C)	January-December 2019	920033	21.80	10.29	-39.8	47.5
Temperature Range (in °C)	January-December 2019	920033	6.59	3.40	0.2	25.5
Cloud Coverage (in %)	January-December 2019	920033	0.53	0.29	0.0	1.0
Air Pollution ($PM_{2.5}$, in $\mu g/m^3$)	January-December 2019	920033	23.15	25.85	0.3	1411.9
Wind Speed (in m/s)	January-December 2019	920033	5.25	2.22	0.6	30.8
Humidity (in g/m^3)	January-December 2019	920033	10.24	5.26	0.1	23.6
Precipitation (in mm)	January-December 2019	920033	0.03	0.07	0.0	3.5

There is important global heterogeneity when it comes to temperature exposure. Daily maximum temperatures range from almost -40°C (in the Russian region of Tomsk on February 1, 2019) to 47.5°C (in Dhi-Qar, Irak, on August 27, 2019). Fig. 3-1 (top panel) plots the yearly average of the daily maximum temperature in each administrative-1 region. Consistent with prior literature [4], we define extreme-warm days as days where the maximum daily temperature is above 30°C, and Fig. 3-1 (bottom panel) visualizes the annual number of "extreme-warm" days per administrative-1 region during our study period. While most of North America, Europe, and Asia witness fewer than 100 extreme-warm days a year, some parts of Africa experience over 300.

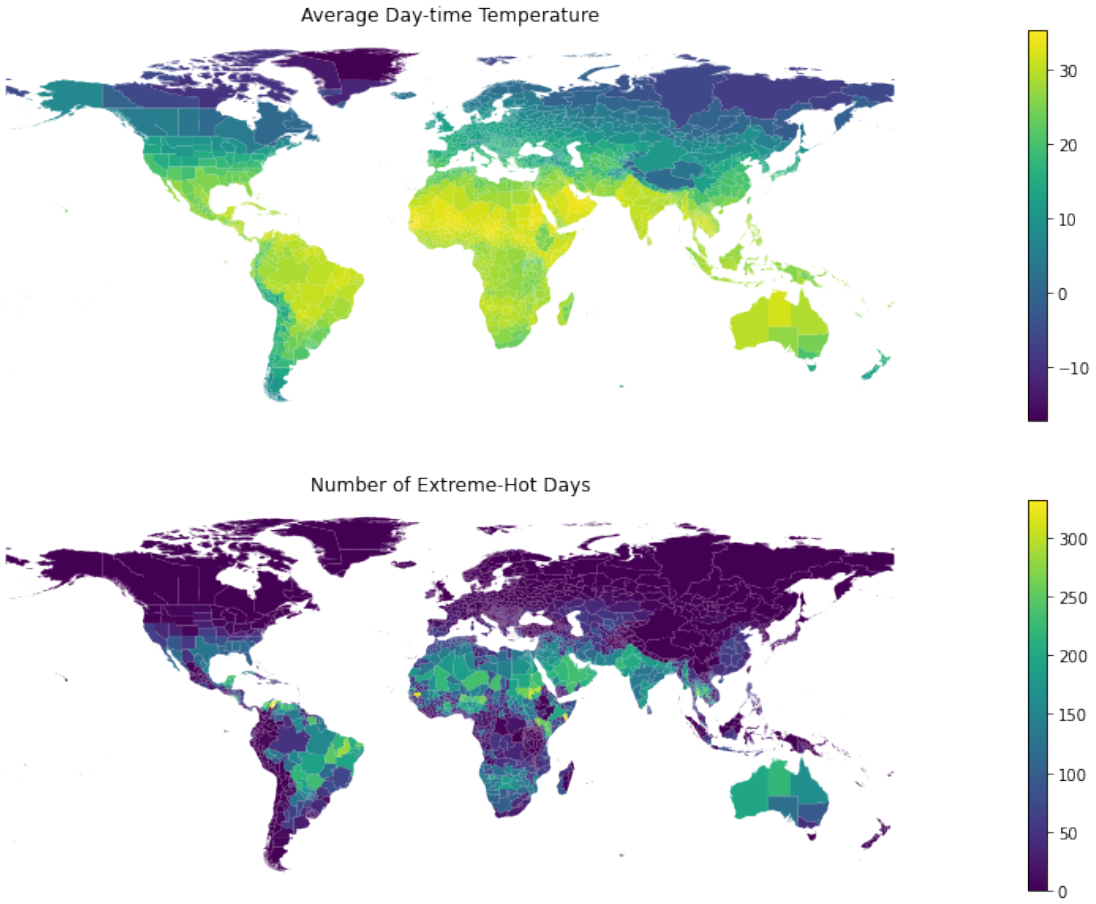

Figure 3-1: Global temperature distribution (2019)

Average day-time temperature is calculated as the average of daily maximum temperatures by region. The number of extreme-hot days is the number of days with maximum temperatures over 30°C.

3.2.2 Social media data

The social media data we base this study on is a subset of the geolocated social media data set that I present in Chapter 2. Only post from Twitter and Weibo shared between January 1, 2019, and December 31, 2019 are included in the sample. We exclude data from countries for which we have fewer that 100 daily observations on average, and are left with 157 analysis countries (the complete list is provided in Appendix Section B.2.1). Over 1.2 billion geotagged Twitter and Weibo posts are included in the study period. Fig. 3-2 plots the number of social media posts collected every day—while we usually obtain around 3.5 million posts every day, occasional issues in the data harvesting process result in lack of data for short period (part of June 2019, for instance).

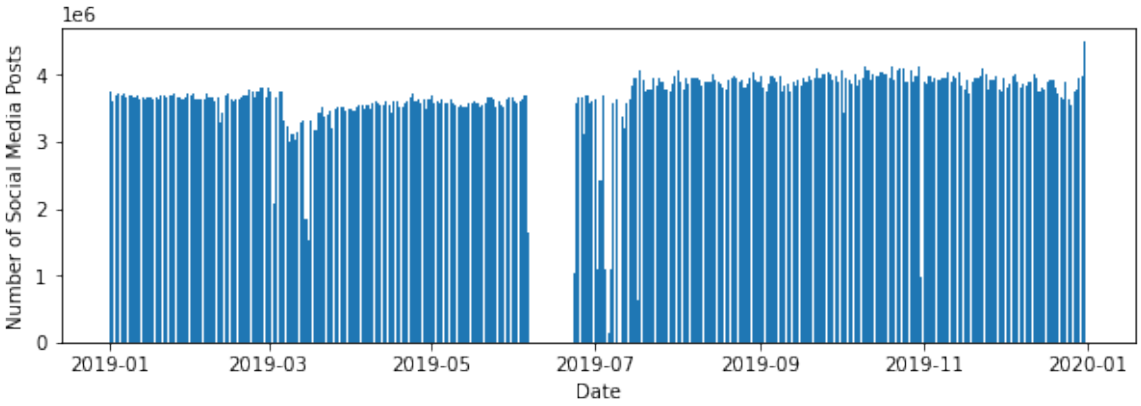

Figure 3-2: Number of social media posts collected by day (2019)

The data is usually composed of around 3.5 million daily social media posts. Occasional periods where no data is collected are due to data harvesting issues.

Sentiment scores are imputed on every post using the NLP-based sentiment analysis method described in Section 2.2.2, and aggregated to administrative-1 level using the two-step aggregation mechanism formalized in Section 2.2.3. At that level of analysis, we have 872,705 observations overall, each comprised of, on average, 1473 posts and 412 users. The daily regional sentiment index ranges from 0.0 to 1.0 (see Table 3.2), although the bulk of the distribution is between sentiment scores of 0.5 and 0.75 (see Fig. 3-3).

Table 3.2: Social Media and Sentiment Data Summary Statistics

Variable	Period	Obs.	Mean	Std. Dev.	Min.	Max.
Number of Posts	January-December 2019	872705	1473.00	7991.75	1.0	298326.0
Unweighted Sentiment Score	January-December 2019	872705	0.62	0.09	0.0	1.0
Number of Users	January-December 2019	872705	411.86	2122.59	1.0	65657.0
Weighted Sentiment Score (main)	January-December 2019	872705	0.63	0.09	0.0	1.0

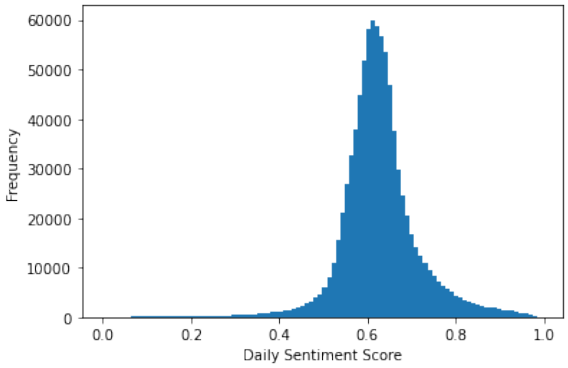

Figure 3-3: Histogram of sentiment index values

The sentiment index is created by aggregating sentiment scores to daily, regional values. The aggregation process is detailed in Section 2.2.3. 872,705 daily regional sentiment index values are computed. The index ranges from 0.0 to 1.0, although most of the scores are between 0.5 and 0.75 (10th percentile: 0.54; 90th percentile: 0.72).

Fig. 3-4 illustrates the geographic coverage of our social media data set. The overall yearly number of social media posts in each region is provided in the top panel, and ranges from fewer than 100 posts to over 50 million in the administrative-1 regions of California (USA), England (UK), and Rio de Janeiro (Brazil). Our data set largely covers all habitable continents, and while user distribution does not entirely reflect population, most administrative-1 regions have at least 10,000 posts overall during our study period.

The bottom panel of Fig. 3-4 plots the yearly sentiment score average for each administrative-1 region globally. While most scores fall in a relatively narrow range of 0.56–0.70, there is important global heterogeneity in sentiment scores with some countries (like Brazil or Russia) displaying consistently lower sentiment scores than others (like India). While this could be interpreted as inherent differences in population happiness, we do not make this claim. Instead, different cultural and linguistic norms might influence our NLP sentiment analysis algorithm in ways that are hard to correct. Therefore, we prefer to leverage changes in sentiment *within* specific regions following exposure to climate events, instead of *between* regions exposed to different climate conditions. We standardize sentiment by administrative region and include regional fixed effects in all of the models we run.

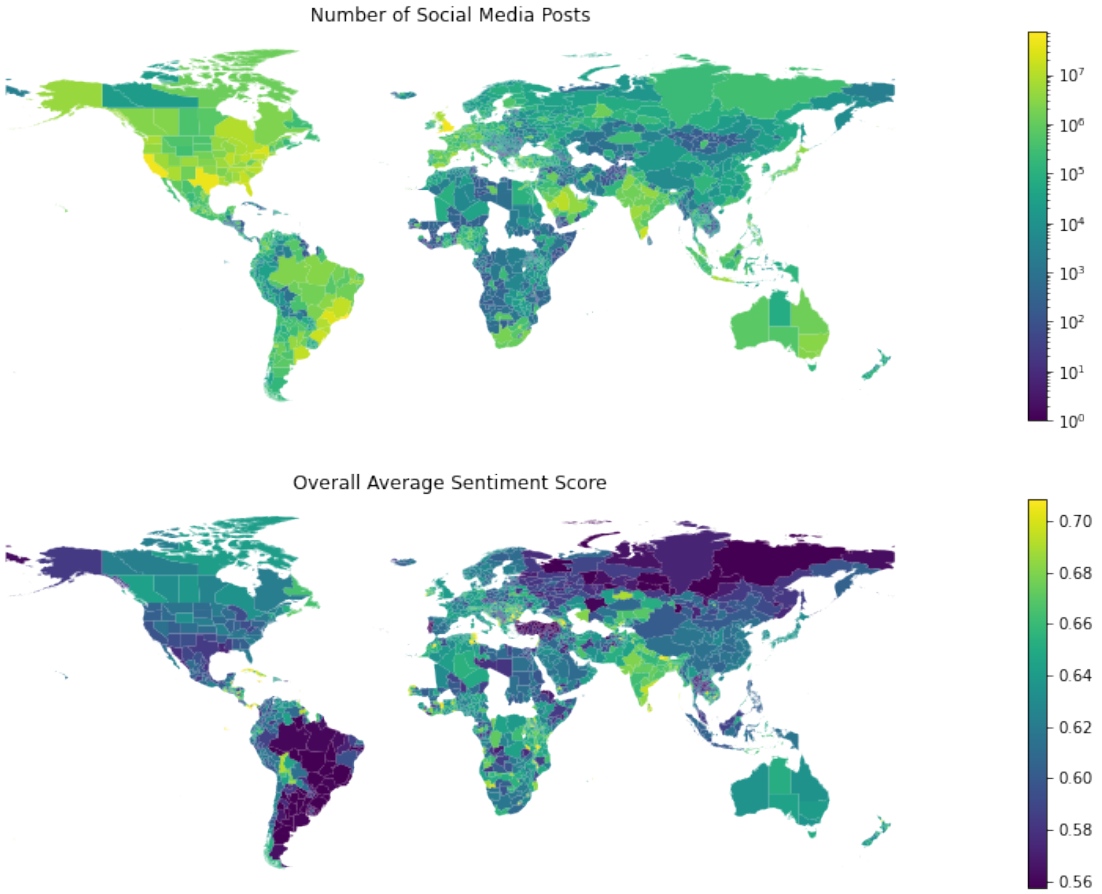

Figure 3-4: Geographic distributions of the social media data

Only countries with over 100 daily social media posts are included in the analysis. Overall average sentiment scores are calculated by averaging the daily sentiment index in each region over the entire year. Sentiment index values are weighted by the number of social media users in that region that day. For ease of visualization, scores are winsorized at the 5th and 95th percentiles.

3.3 Econometric Modeling

The impact function of temperature on expressed sentiment is estimated by applying a fixed effect time-series regression model to our social media and climatic data. Equation 3.1 formalizes our econometric model.

$$sentiment_{it} = \alpha_0 + \alpha_1 f(Tmax_{it}) + \alpha_2 X_{it} + T_t + \gamma_i + \epsilon_{it} \tag{3.1}$$

The dependent variable $sentiment_{it}$ stands for our measure of sentiment at location i and on calendar date t, respectively. The independent variable of interest,

$Tmax_{it}$, represents the daily maximum temperature of each location, measured in °C (we also run robustness checks using the daily mean temperature $Tmean_{it}$ and the daily minimum temperature $Tmin_{it}$; results are robust with both measures; see Appendix Section B.3.6).

The function f transforms continuous values of temperature into discrete 5°C bins, and α_1 is a set of coefficients associated with each one of those bins. In each regression, one temperature bin is omitted and serves as the reference sentiment measure (in the global results of Fig. 3-5, for instance, the reference temperature bin is 15°C-20°C). Therefore, the α_1 coefficients can be interpreted as the marginal effect of a given temperature bin relative to the reference bin. This method allows for a flexible, non-linear estimation of the relationship between temperature and expressed sentiment.

X_{it} are environmental controls including temperature range, precipitation, wind speed, cloud coverage, humidity, and $PM_{2.5}$ air pollution levels. Unobserved factors specific to locations and calendar dates might affect sentiment in ways that correlate with temperature and environmental measures. Different locations might have inherent differences in sentiment levels linked to economic or cultural factors. National holidays or seasonality might also alter expressed well-being. We account for these spatial and temporal cofounders by including date (T_t) and location (γ_i) fixed effect. Finally, the main regression is run weighting location-days by the number of social media users recorded within that location and on that day. Different sets of controls, fixed effects, and weighting schemes are tested in robustness checks, and presented in Appendix Section B.3.

3.4 Results

3.4.1 The global effect of temperature on expressed sentiment

The panel regression model presented in Section 3.3 is estimated on our entire data set to assess the global effect of 5-degree temperature bins on expressed sentiment. In Fig. 3-5, we report a non-linear relationship between daily maximum temperature and the sentiment index. The 95% confidence interval (CI) for each bin is provided in shaded blue around the bin's estimate—highlighting that our results are strongly significant. The figure also includes a histogram underneath the plot to visualize the temperature distribution, and the 95th percentile is marked with a vertical dotted red line.

The relationship between temperature and sentiment resembles an inverse-U curve. Sentiment increases when temperatures increases from cold to temperate and peaks between 15°C and 20°C (we this temperature bin temperatures as our omitted reference bin). Subsequent increases in temperature are associated with sharp drops in sentiment. Daily maximum temperatures of 35°C–40°C reduce global sentiment by 18.2% of a standard deviation (95% CI: 17.3%–19.1%).

The magnitude of the effect is similar to the results found by Baylis (2020) in the United States, where extreme-warm temperatures are associated with sentiment drops of 21% of a standard deviation [9]. As a measure of interpretation, in Appendix Section B.3.3, we find in our data set that sentiment is on average 17.2% of a standard deviation (95% CI: 16.9%–17.4%) lower on weekdays (Monday–Friday) than on weekends (Saturday and Sunday). The sentiment drop associated with temperatures over 35°C is therefore equivalent to 105% of the average weekend-to-weekday difference. Similarly, Wang et al. (2020) find the sentiment drop associated with temperatures over 35°C in China to be equivalent to over 89% of the average Sunday-to-Monday sentiment difference [136].

Full regression results are provided in Table B.2 of Appendix Section B.3.1. Estimates for our environmental controls are also provided, indicating that air pollution, precipitation, and higher wind speed all significantly worsen sentiment as well.

3.4.2 Robustness by weighting scheme

In our main regression, daily-regional observations are weighted by the number of social media users in a given location on a given day. However, social media penetration rates may differ greatly by region. Chapter 2—and more specifically Section 2.3.1—highlights that while social media use is overall significantly correlated to population, the digital divide still results in important disparities, especially between countries worldwide. This may lead to our results being mostly driven by highly-connected, data-rich nations (like the United States and China), as well as urban areas. To address this, we test out four distinct weighting schemes: (1) weighting all regions equally regardless of social media activity or population; (2) weighting by social media posts (instead of users); (3) weighting by social media users (our default); and (4) weighting by regional population. Apart from the weighting scheme, the four estimations are run with the same model specifications (see Equation 3.1). Results are presented in Fig. 3-6, with full results provided in Appendix Section B.3.4.

We find that the results are mostly robust by weighting scheme. All four esti-

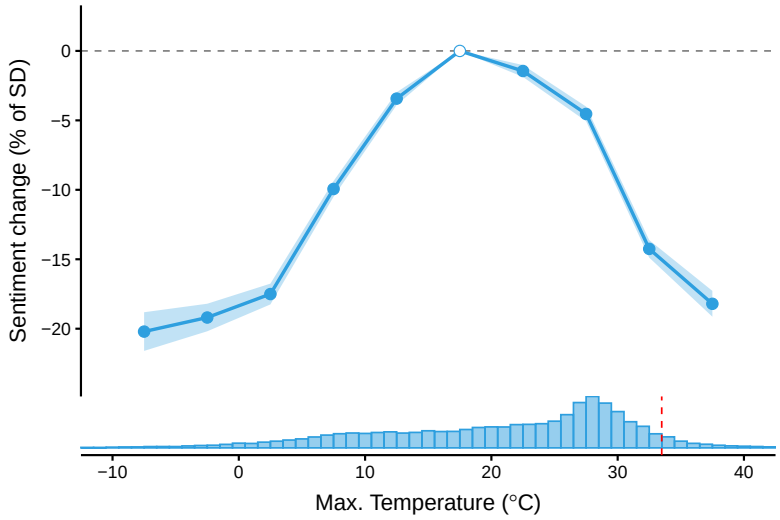

Figure 3-5: The global effect of temperature on expressed sentiment

The temperature measure we use is daily maximum temperature, measured in 5-degree bins. The regression model includes environmental controls (temperature range, air pollution, wind speed, cloud coverage, humidity, and precipitation), and location and date fixed effects. The temperature bin coefficients are plotted in blue, along with their 95% CI (in light blue). The 15°C-20°C bin serves as the omitted temperature bin, and the coefficients of the other bins are expressed as relative changes (in percentages of a standard deviation) compared to the omitted bin. The bottom histogram plots the temperature distribution in countries of our sample, weighted by the number of social media users. The 95th percentile of temperature is marked with a vertical dotted red line.

mations yield a similar inverse-U curve, with significant sentiment drops associated to both high and low temperatures relative to the omitted 15°C-20°C bin. Unsurprisingly, weighting by social media users (in blue) and posts (in red) produce very similar results. The unweighted estimate presents less variation and larger confidence intervals—which are likely the result of increased noise due to the weight of very small regions with scarce data in this specification. The population-weighted estimates are similar to our main estimates in both extreme-low and extreme-high temperatures; however, sentiment peaks in the 25°C–30°C bin, instead of the omitted reference bin of 15°C–20°C. While extreme temperatures are consistently associated with sentiment drops, results of the population-weighted estimate indicate that there exists a wide range of temperatures that might be regionally considered as "most comfortable". In country-level regressions that we present in Section 3.4.6, we account for this by allowing a wide range of temperature bins (any of the four bins between 10°C and 30°C)

to serve as the country's reference bin.

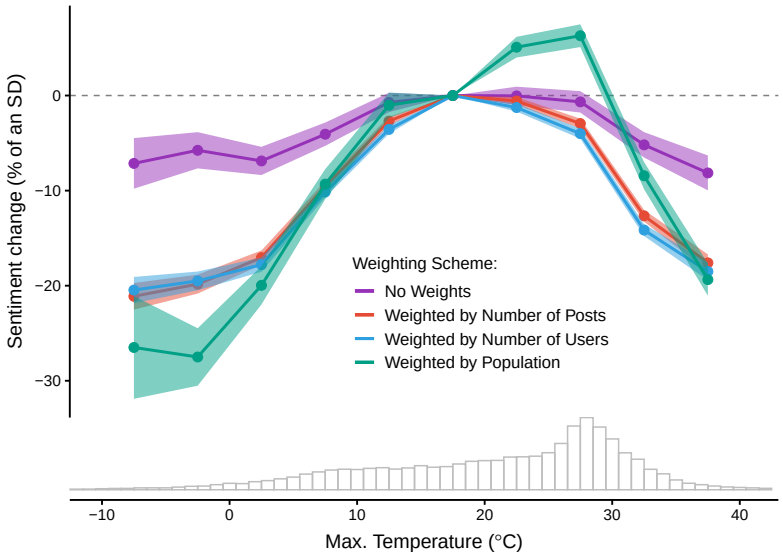

Figure 3-6: Robustness by weighting scheme

Equation 3.1 is estimated using four distinct weighting schemes. In purple, observations are unweighted. In red, they are weighted by the number of social media posts. In blue, they are weighted by the number of social media users (as in the main results of Fig. 3-5). Finally, in green, they are weighted by regional population, regardless of social media usage.

3.4.3 Individual-level results

We recognize that the relationship between temperature and sentiment could be attributed to compositional changes in the user pool: for instance, individuals more sensitive to extreme temperatures post on social media more frequently on days when the weather is exceptionally cold or warm. We address this potential endogeneity issue by estimating the model restricted to a random sample of 100,000 frequent social media users[1] and by including user fixed effects. Results are presented in Figure 3-7 (with full results provided in Appendix Section B.3.2).

[1]Frequent users are selected by randomly sampling users from the raw post-level data. Therefore, the probability of a given user being selected is proportional to the overall number of posts from that user.

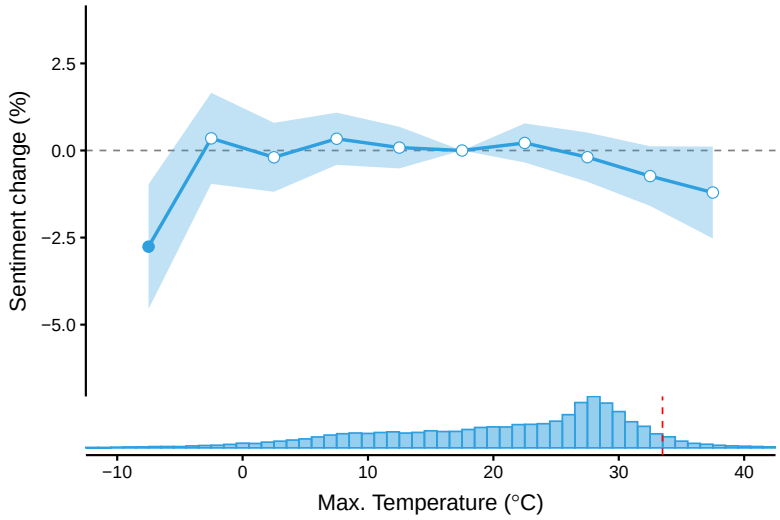

Figure 3-7: Individual-level effect of temperature on expressed sentiment

The model specifications are the same as those of Fig. 3-5, with additional user fixed effects. The sample is restricted to a sample of 100,000 frequent social media users.

We find similar patterns, with sentiment peaking for mild temperatures. Temperatures below -5°C are associated with significant sentiment drops of 3.7% of a standard deviation (95% CI: 1.1%–6.3%), and temperatures above 35°C result in an (insignificant) sentiment drop of 1.9% of a standard deviation (95% CI: 0.0%–3.8%). The magnitudes are smaller than for the original sample, consistent with prior findings using user-level fixed effects. Baylis (2020), for instance, finds a high-temperature sentiment drop of 10% with user fixed-effects, down from 21% for the overall sample [9]. This could support the hypothesis of some compositional change during extreme temperatures, but could also be the result of increased noise due to the constraining user-level fixed effects.

3.4.4 Climate zone heterogeneous effects

We go on to investigate how the effects of extreme temperatures on sentiment vary based on a region's climate characteristics. We sort our nearly 3000 regions into climate zones based on a generalization of the Köppen-Geiger map [72]. Each region is assigned a main climate feature which can take one of six values: equatorial, arid, warm, snow, or polar. We estimate our econometric model in each global climate zone separately, and present the results in Fig. 3-8.

Here, we find general trends consistent with our global results, but regional hetero-

geneities start to emerge in the magnitude of sentiment damage. All regions display a most-comfortable temperature of 15°C–20°C (arid, warm, snow, polar) or 25°C–30°C (equatorial). In colder temperature ranges, sentiment drops are steeper in Arid, Equatorial, and Warm climate zones. Snow zones, on the other hand, show only limited sentiment declines even as temperatures reach -15°C. High temperatures above 25°C are associated with significant sentiment drops in all non-polar climate zones, but the decrease is more gradual in Arid regions where these high temperatures are more common. These results suggest reference-based utility patterns in which the emotional impacts of temperature are contingent on a region's usual temperature range.

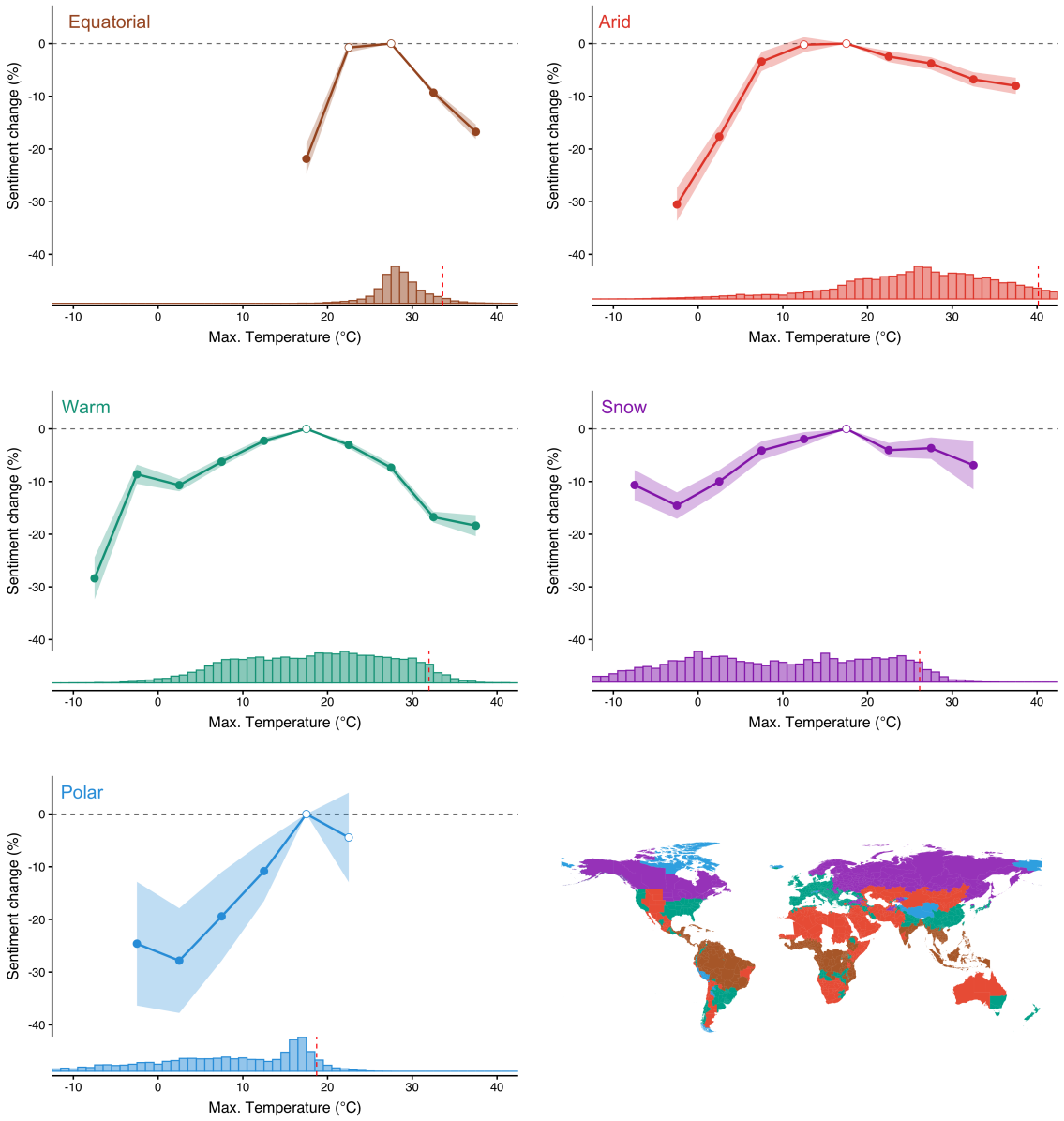

Figure 3-8: Climate zone-specific effect of temperature on expressed sentiment

The effect of maximum daily temperature on expressed sentiment (Equation 3.1) is estimated separately out each different climate zone: (a) Equatorial climate zone; (b) Arid climate zone; (c) Warm climate zone; (d) Snow climate zone; and (e) Polar climate zone. Density plots under the coefficient plots display the distribution of temperatures for each climate zone. The mapping of these climate zones is provided in the bottom right-hand corner.

3.4.5 Development stage heterogeneous effects

Subjective perception of temperature stress may differ largely between regions at different development stages. Disparities might be due to availability of mitigation and adaptation technologies, such as heating or air conditioning, to infrastructure resiliency, or to expected economic losses resulting from climate events. In a seminal paper, Mendelsohn, Dinar, and Williams (2006) argue that poor countries will suffer the most from climate change damages [86].

To shed light on these potential disparities, we investigate heterogeneity by regional wealth. Using the grid-level GDP measures developed by Kummu et al. (2018) [75], we compute an administrative-1 regional GDP measure. Regions are then split into "High GDP" and "Medium-to-Low GDP"[2], and independently estimate the effect of temperature on sentiment in both sets of regions.

The results are provided in Fig. 3-9. We obtain inverse-U-shaped curves in both High and Medium-to-Low GDP regions, with sentiment peaking around 15°C-20°C in both zones. However, sentiment drops associated with high and low temperatures are significantly larger in Medium-to-Low GDP regions. Temperatures below 0°C reduce sentiment by over 35% of a standard deviation (95% CI: 33.2%–36.0%) in less wealthy regions—the drop is of only 8.2% of a standard deviation (95% CI: 6.1%–10.2%) in richer regions. Similarly, temperatures above 35°C are associated with 25.3% (95% CI: 24.1%–26.5%) and 5.2% (95% CI: 2.9%–7.4%) sentiment standard-deviation drops in low GDP and high GDP regions, respectively.

These results are consistent with prior literature on wealth-dependent adaptation mechanisms. Wang et al. (2020) find that poorer cities display sharper sentiment drops under cold temperatures than rich cities, and that this could be explained in part by winter-heating ownership. They also find positive—if insignificant—effects of AC unit ownership on sentiment during high temperature days [136].

[2]We split administrative-1 regions based on a threshold value of $40,000 of GDP per capita.

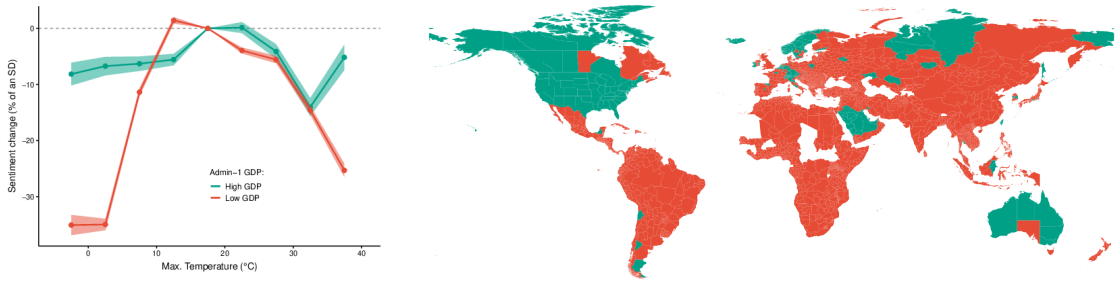

Figure 3-9: Development zone-specific effect of temperature on sentiment

The effect of maximum daily temperature on expressed sentiment (Equation 3.1) is estimated separately in high GDP-per-capita regions (above $40,000, in green) and low GDP-per-capita regions (below $40,000, in red). The mapping of the administrative-1 regions by GDP level is provided on the left.

3.4.6 Heterogeneous effect across countries

Analyses conducted in the previous sections have highlighted that both climate characteristics and development stage play a role in determining a region's temperature damage function. These factors point to expected differences in country-level damage functions. To assess country-level heterogeneity, the baseline regression function (defined in Section 3.3) is estimated separately in the 157 countries where our data coverage allowed for modeling.

Fig. 3-10 presents the damage functions we obtain in eight example countries—covering all inhabited continents and a wide range of climatic and economic characteristics. As in previous result visualizations, sentiment levels are expressed as marginal changes relative to an omitted reference temperature bin. Here, and the reference bin is chosen in each country as the bin associated with the highest value of sentiment, within a range of moderate temperatures[3]. In the United States, for instance, the reference temperature bin is 20°C–25°C; in Argentina, maximum sentiment is achieved under temperatures of 15°C–20°C and in Great Britain, the reference bin is 10°C–15°C. South Africa presents the highest reference bin, with temperatures between 25°C and 30°C considered most comfortable.

General trends in the damage curves are consistent with the inverse-U shape of global results: all countries present negative sentiment impacts associated with their highest temperature bins. However, the value of the damage varies considerable. In the United Kingdom, temperatures in the 25°C–30°C bin are already associated to

[3]We define moderate temperatures as comprised between 10°C and 30°C. Therefore, the reference bin is either the 10°C–15°C bin, the 15°C–20°C bin, the 20°C–25°C bin, or the 25°C–30°C bin.

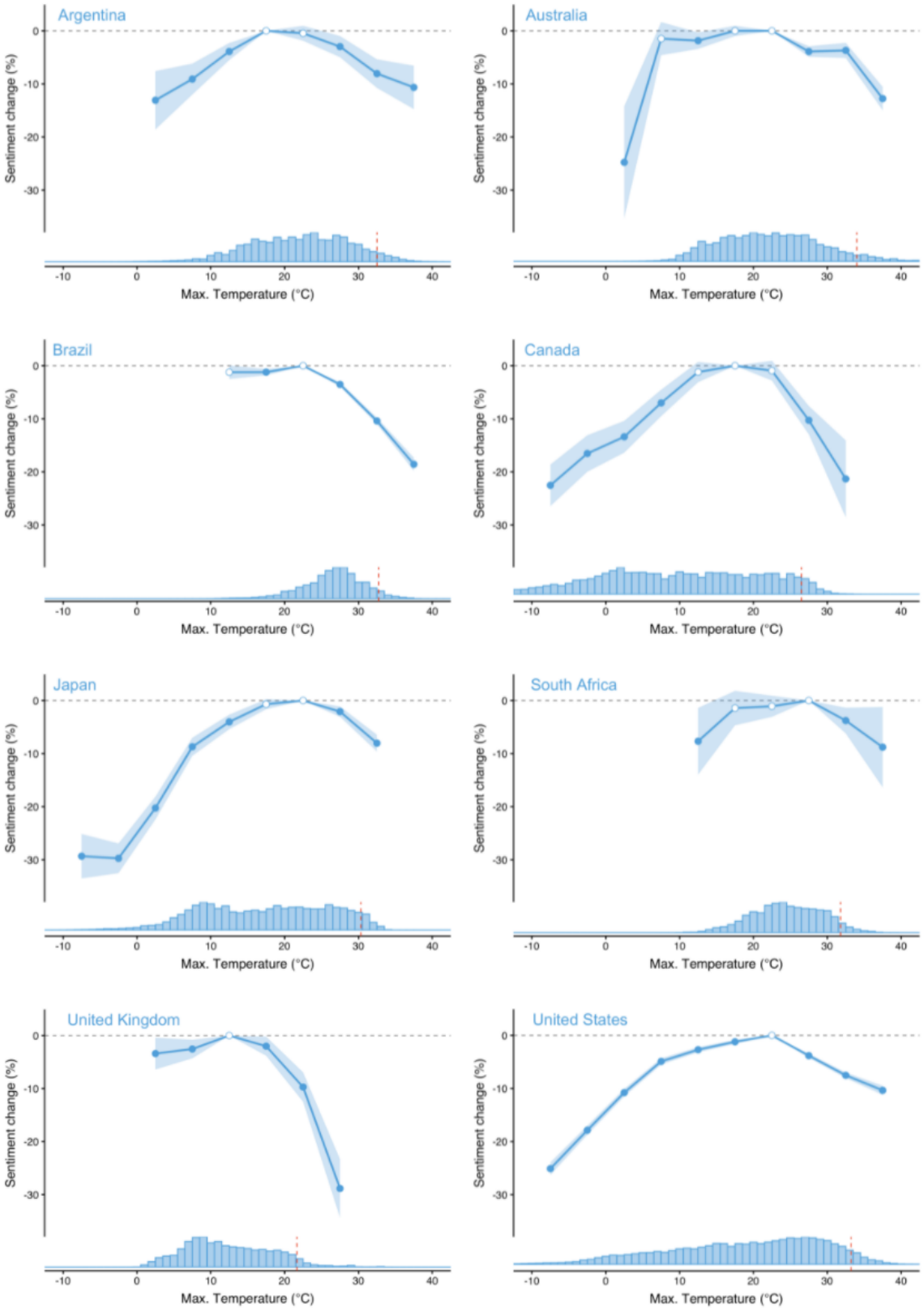

Figure 3-10: Country-specific effect of temperature on expressed sentiment

sentiment drops of 28.8% of a standard deviation (95% CI: 23.3%–34.3%). In most of the other countries however, sharp drops in sentiment: temperatures above 35°C in sentiment drops of 18.5% of a standard deviation in Brazil (95% CI: 17.3%–19.8%), and of 12.7% of a standard deviation in Australia (95% CI: 10.5%–14.9%). Country-specific temperature histograms are provided in the lower section of the plot, and may provide a first intuition to the heterogeneous results we observe: while the 95th percentile of temperature is equal to only 21.5°C in the United Kingdom, its value is above 34°C in Australia.

Country-level high-temperature damage coefficients are constructed by estimating the value of the coefficient associated with the temperature bin of the 95th temperature percentile. A negative damage coefficient implies that relatively high temperatures in that country are associated with a sentiment drop, while a positive damage coefficient indicates higher sentiment under relatively hotter temperatures. In Japan, for instance, the 95th temperature percentile is 30.4°C, and damage coefficient—the sentiment change associated with the 30°C–35°C bin—is -8% of a standard deviation. Under relatively hot temperatures, most countries display negative sentiment reactions: 139 out of the 157 countries of our analysis have negative damage coefficients. The bulk of the countries have a damage coefficient between -25% of a standard deviation (25th percentile) and -8% of a standard deviation (75th percentile). Global heterogeneity of these coefficients is summarized in Fig. 3-11. The upper panel plots the map of the damage coefficient value by country, and the lower panel plots the damage coefficient values for a selection of countries, along with the 95% CI to inform on statistical significance of the results.

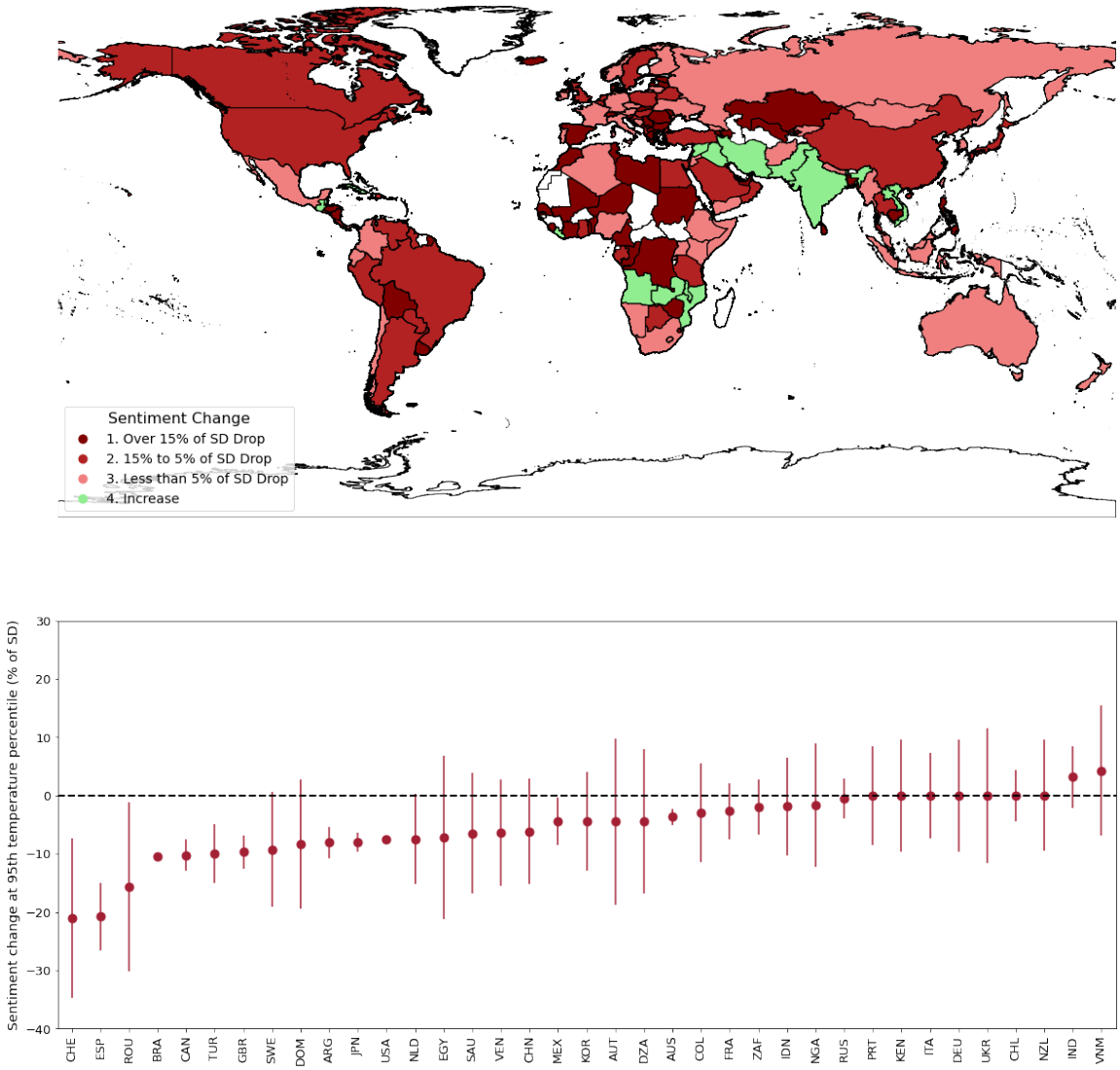

Figure 3-11: Country-specific sentiment change at 95th temperature percentile

Top panel: Mapping the effect size across countries. The effect size is defined by sentiment change (in % of a SD) between the country's reference bin and their 95th percentile of temperature. Bottom panel: Comparing the sentiment changes of a select countries. The dots represent the point estimates, the lines are the 95% CI.

3.5 Chapter conclusion

To the best of our knowledge, this study represents the largest and broadest-scale investigation of potential subjective well-being impacts associated with extreme ambient temperatures. We observe that high temperatures substantially depress expressed sentiment on social media—days above 35°C are associated with a drop of

18.2% of a standard deviation compared with pleasant temperatures. The trends hold in most countries across the world, although the magnitude of the sentiment drop varies significantly.

The trends observed in the main findings are consistent with previous studies, especially results obtained in data-rich countries such as the United States or China [9, 136]. However, our use of a novel social media data set—which provides abundant observations at a global level—and of multilingual sentiment analysis models allows us to generalize these findings to country-specific damages, and to highlight important heterogeneity between regions and nations.

Some assumptions and limitations should be acknowledged. First, only 2019 data feeds into our analysis. This allows us to include more countries in our analysis—since data quality decreases with earlier years of data—but also prohibits us from studying longer term trends. Therefore, we are unable to properly investigate adaptation mechanisms that alter reactions to extreme temperatures, and we are unable to confidently make future projections of climate damages. Subsequent studies could incorporate additional years of data into the analysis, and assess how damage coefficients change over time.

Second, our sentiment measure is derived from individuals who post on social media. Although this fraction represents a substantial portion of global residents, it is not necessarily representative of overall population (see Section 1.1.2.2 for more information about the demographic characteristics of social media users). Our results might reflect this sample selection bias. Elders and children, who are less likely to use social media, are also the most vulnerable to extreme temperatures—potentially leading us to underestimate the overall negative effect of adverse temperatures on individual sentiment. Here, further research could look into correcting selection biases in social media data—by combining it with more traditional survey data, for instance.

Finally, given the range of sentiment drops in different countries associated with the 95th percentile of temperature, in depth heterogeneity analysis is required. Climate and development factors might be at play, and we explore these two potential sources of variation in Sections 3.4.4 and 3.4.5. However, we would also expect more unconventional attributes—such as belief in climate change or anticipated future climate damages—to affect sentiment reaction to extreme temperatures.

In summary, this study offers a comprehensive characterization of extreme-temperature impacts on expressed sentiment. Results are subject to a number of implications and considerations. We highlight a global emotional toll to climate change. This additional well-being cost is important for policymakers to account for when considering

resource allocation and policy alternatives on the topic. More broadly, given the growing use of social media worldwide, monitoring communication on these platforms can provide researchers and policymakers with insight into levels of public support or dissatisfaction. In this context, studies such as this one should incentivize governments all over the world to act on climate change.

Chapter 4

Case Study: Temperature stress perception and location attractiveness

This chapter is derived from a manuscript co-authored with Fátima Trindade Neves, Mauro F. Pereira, Juan Palacios, Miguel de Castro Neto, and Siqi Zheng.

Abstract

As anthropogenic climate change disrupts cities worldwide, increasingly severe weather events and temperature discomfort pose potential harm to location attractiveness, and therefore to real estate value. However, the extent to which climate events affect markets depends highly on the subjective perception of these events by local inhabitants. This study proposes a model examining the impacts of temperature stress on real estate value which incorporates subjective measures of these climate events, in the form of a social media-based sentiment index. We run an empirical study in the largest metropolitan areas of Portugal, and find that temperature discomfort has a significant, negative impact on housing prices. Different regions are impacted at different scales though, and municipalities with the strongest subjective sentiment reactions to temperature discomfort also witness the sharpest drops in real estate value. These result reinforce the importance of including subjective perceptions when assessing the impacts of climate change, and the relevance of social media data to do so.

4.1 Introduction

The frequency and intensity of climate events have soared in recent years, and extreme temperature are increasingly affecting cities across the world. Between 2000 and 2016, the number of people subject to heatwaves increased by 125 million worldwide, and it is projected to reach over 1.6 billion people by 2050 [129]. This growing exposure is introducing substantial burdens to societies, and has been linked to increased health risks and criminality, as well as lower food security, water supply, and economic growth [22]. Understanding the full array of societal costs, as well as the effectiveness of interventions, is essential for policymakers to develop effective mitigation and adaptation strategies that cope with these events.

Real estate markets are key when it comes to assessing the economic impacts of weather events on cities. The long-term nature of real estate assets means these properties are exposed to the long-term changes in the distribution of temperature. Real estate asset value is also closely tied with household financial stability: in the European Union, for instance, 75% of the population live in cities [142] and 70% are homeowners [45][1]. Temperature discomfort is already challenging real estate markets, as urban development and planning struggle to mitigate the impacts on well-being and infrastructure [2]. Costly adaptation measures incur a burden on urban households: extreme temperatures are have increased dependence on residential air conditioning units [8], increased residential energy consumption [37], and increased electricity demand [36][2]. These additional costs, coupled with subjective costs to well-being, may make locations less desirable altogether, depreciation real estate value in the area.

This chapter investigates the impact of temperature stress on real estate markets in urban settings. We contribute to the literature by accounting for heterogeneous perceptions of these events: we explore the role of subjective measures of temperature discomfort in explaining observed price drops on the real estate market. Using a rich data set of social media posts combined with NLP methods, we estimate the damage to extreme-hot and extreme-cold weather on expressed sentiment locally[3] and use it as a factor of price changes associated with temperature stress.

[1] In the United States, over 80% of the population is urban [130] and 65% owns property [18].

[2] Lower income individuals are at risk of being excluded from these adaptation mechanisms: Kahn (2016) estimates that it would cost at least $120 per year for poorer households to operate an AC unit [70].

[3] Social media expressed sentiment indices have been increasingly used by computer and social scientists as indicator of expressed well-being, and validated with traditional survey measures of happiness [68]

We conduct our empirical study in the context of Portugal, one of the most vulnerable European countries with regard to extreme heat in particular, and climate change in general [44]. One hundred and thirty heatwaves were detected in the country between 1981 and 2010, and the frequency is expected to increase further by the end of the 21st century [103]. High temperatures are specially detrimental in metropolitan areas: a July 2006 heatwave resulted in significant excess mortality in Porto, for instance [92]. Exposure to extreme weather—as well as flooding and wildfires—already has adverse effects on the economy, health and well-being of the country's citizens [101]. Our unique, multi-layer data set provides novel estimates of the impact of temperature stress on the four largest urban areas in Portugal, covering around 50% of the population and 25% of the housing units of the country: Greater Lisbon, Greater Porto, Coimbra, and Braga[4].

The remainder of this chapter is organized as follows. Section 4.2 provides an overview of the literature examining the impacts of weather events on real estate value and introduces the role of subjective well-being in shaping price discounts in housing markets. Section 4.3 introduces the data that we use in our analyses. In Sections 4.4 and 4.5, we describe the methodology and results of our three empirical analyses: the direct impact of temperature discomfort on real estate value (Section 4.5.1), the direct impact of temperature discomfort on well-being (Section 4.5.2), and the integrated model combining the effects of sentiment damages and weather events on real estate value (Section 4.5.3). In the last section (Section 4.6), we discuss our results and conclude.

4.2 Literature Review: The impact of temperature stress on real estate value and well-being

To understand the effect of heterogeneous subjective damages of temperature stress on real estate prices and guide our empirical investigation, we build a conceptual model of the housing market in Fig. 4-1. Extreme temperature events have a direct impact on location value (arrow 1). Temperature stress also has an impact on well-being: however, households can differ in their perception of temperature stress—leading to heterogeneous well-being damages across different municipalities (arrow 2). These damages capture the direct emotional impact of temperature stress on individuals as well as the costs for associated behavioral change triggered by temperature stress

[4]In Portugal, only Greater Lisbon and Greater Porto are qualified as metropolitan areas. Coimbra and Braga are next largest municipalities.

(e.g., not leaving the house to avoid extreme heat), and can be higher in municipalities with poorer mitigation options[5]. Given these disparities, the direct impact of temperature stress on the real estate market can be seen as a coarse approximation. In our integrated model (arrow 3), the impact of temperature stress on location value is moderated by the subjective perception these events by the inhabitants of that municipality.

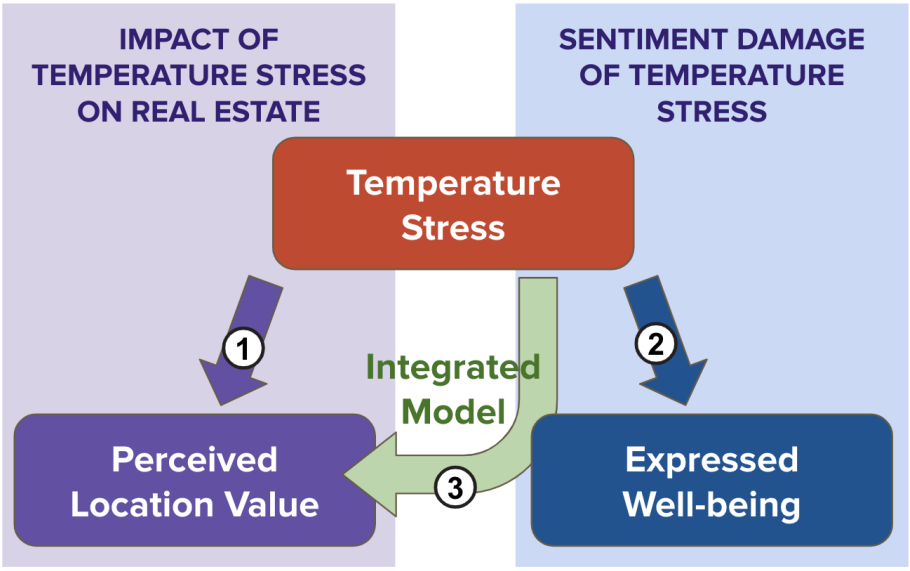

Figure 4-1: Conceptual Framework

Temperature stress can be modeled as having a direct impact on location value (arrow 1). Temperature stress also has a heterogeneous, localized effect on expressed well-being, or sentiment (arrow 2). In our integrated model, the effect of temperature stress on location value is moderated by sentiment damages (arrow 3).

Prior research has documented the case of temperature as an urban amenity, often finding a negative link between temperature stress and real estate value (arrow 1 of Fig. 4-1). An overview of the literature is provided in Appendix Section C.2 (Table C.1). In the United States, temperatures have been found to influence both wages and house prices [113, 14, 35]. Fan, Klaiber, and Fisher-Vanden (2016) find that extreme temperature drives residential sorting, with household willingness to pay (WTP) increasing by $144 and $91 to avoid an additional extreme-hot and extreme-cold day, respectively[6] [46]. Still in the context of the United States, Albouy et al.

[5]A higher sentiment damage can indicate, for instance, that a municipality is less prepared to handle extreme temperature: for example, the municipality has poor housing infrastructure (no central heating or air conditioning) or lacks green areas to mitigate heat island effects [125].

[6]Estimates are even more important for college-educated households, with are slightly lower for the overall population, with WTP decreases of $203 per warm day and $326 per cold day.

(2016) examine American households' WTP to live in different areas depending on climate characteristics and find that Americans have a preferred temperature of 65°F (equivalent to 18°C), and that extreme-high temperatures have a stronger negative impact on WTP than extreme-cold temperatures [4]. Smaller scale studies in European countries—such as Italy, Germany, and Great Britain—have also yielded similar results [81, 111, 85].

The mechanism by which climate or weather events impacts real estate prices depends highly on information availability and interpretation. This is consistent with the idea that climate risks are often inappropriately reflected in asset pricing: Hong, Li, and Xu (2019), for instance, show that stock markets are ineffective at discounting drought-related risk [63]. Baldauf, Garlappi, and Yannelis (2020) claim that local subjective perception of climate change is a key driver in the poor forecasting of climate risks on assets. They find that increased climate risks impact real estate prices only to the extent that buyers believe in climate change and expect future climate events to have a negative effect on their well-being. Their hedonic model estimates that a one standard-deviation increase in the share of climate change believers is associated with a 7% decrease in the prices of houses projected to be affected by inundation [7].

Therefore, real estate value discounts due to temperature discomfort ought to depend on the subjective perception of these events by local inhabitants. Prior literature has attempted to quantify these subjective impacts, paying a particular focus to the effect of climate events on expressed well-being, or happiness (arrow 2 of Fig. 4-1). Studies have traditionally relied on surveys to measure the subjective responses of individuals to weather or climate events. Rehdanz and Madison (2005), for instance, evaluate the impact of temperature and precipitation on self-reported levels of happiness in 67 countries, as measured by the "World Database of Happiness" survey [110]. More recently, researchers have looked to social media as a valuable resource in understanding public perception and behavior of climate events. Its spatial and temporal granularity largely exceeds that of survey data, meaning the derived insights are localized and relevant for real estate studies. Since social media platforms are used globally, results can be compared between countries with a large degree of consistency (see Section 1.1.2).

Measures of expressed well-being are extracted from social media text using sentiment analysis methods of NLP[7]. Based on textual analysis of social media data from

[7]These can range from dictionary-based approaches, like LIWC [106] or Hedonometer [40], to more complex algorithms like the one use here (described in Section 4.3.3).

Twitter in the United States, Baylis (2020) documents a relationship between temperature and sentiment in the form of an "upside-down U shape": sentiment initially rises with temperature, plateaus around 20°C–25°C, then decreases sharply. Temperatures above 40°C are associated with a drop of up to 20% of a standard deviation compared to days with 20°C–25°C temperature [9]. Using data from Weibo, Wang, Obradovich, and Zheng (2020) also finds that extreme temperatures have a negative impact on expressed sentiment in China: temperatures above 35°C (compared to 20°C–25°C) reduce sentiment by over 89% of what they find as the typical Sunday-to-Monday difference. They also find important heterogeneity in the impact of extreme temperatures (especially extreme-heat)—based on genders and access to air conditioning units, for instance [136]. Finally, in Chapter 3 of this thesis, we conduct a global study that compares the damages of extreme temperatures on climate events across the world. While most countries exhibit significant drops in expressed well-being under extreme temperatures, perception of temperature stress also differs based on climate zone and income level.

Our integrated model (arrow 3 of Fig. 4-1) describes how heterogeneity in these well-being damages may impact urban real estate market value. Similarly to climate, happiness is an urban amenity that can be offset by real estate value or wages [51]. Therefore, while the hedonic preference for moderate temperature ranges expressed by urban populations guarantee that temperature stress commands a price discount everywhere, this ought to especially be the case in regions with poor adaptation or mitigation mechanisms. Sentiment, which captures subjective perception of temperature amenities, can serve as a proxy for adaptation capacity. Localized sentiment damage measures can moderate the impact of temperature stress on real estate value, or serve to identify regions where real estate value is especially at risk from changing climate.

4.3 Data

This section describes the data sources of each variable included in our analysis, and the construction of the different indices to measure temperature stress, and monitor real estate and sentiment changes in the Portuguese municipalities that are part of our sample. We construct a unique data set combining three layers of geocoded information: weather data, real estate data, and social media-based sentiment data. We also include a rich set of municipality-level characteristics as controls for our analyses to ensure that our estimates are not driven by differences in observable

attributes. Table 4.1 contains the summary statistics of the different variables of the analysis.

Table 4.1: Data Summary Statistics

Variable	Period	Obs	Mean	SD	Min	Max
Panel A: Daily Weather Data						
Maximum Temperature (in °C)	2015-2019	537884	19.36	6.75	1.4	42.8
High-Temperature Day (above 30°C, dummy)	2015-2019	537884	0.08	-	-	-
Low-Temperature Day (below 10°C, dummy)	2015-2019	537884	0.06	-	-	-
Air Pollution ($PM_{2.5}$, in mug/m^3)	2015-2019	537884	18.46	43.69	1.2	4138.9
Precipitation (in mm)	2015-2019	537884	0.02	0.06	0.0	1.3
Humidity (in g/m^3)	2015-2019	537884	64.11	1.31	60.1	68.7
Panel B: Yearly-Aggregated Weather Data						
Average Daily Humidity (in g/m^3)	2015-2019	1540	64.11	0.39	63.3	65.3
Annual Precipitation (in mm)	2015-2019	1540	8.45	3.84	0.1	18.0
Annual Air Pollution ($PM_{2.5}$, in $\mu g/m^3$)	2015-2019	1540	6445.92	2591.61	179.4	16017.1
Average Daily Maximum Temperature (in °C)	2015-2019	1540	19.39	1.61	16.0	26.6
Number of High-Temperature Days (above 30°C)	2015-2019	1540	27.40	23.59	0.0	100.0
Number of Low-Temperature Days (below 10°C)	2015-2019	1540	20.76	24.97	0.0	118.0
Average Winter Temperature (in °C)	2015-2019	1540	13.05	2.14	8.1	17.9
Average Summer Temperature (in °C)	2015-2019	1540	26.36	2.91	20.6	33.0
Summer Discomfort Index	2015-2019	1540	5.56	2.91	0.0	12.4
Winter Discomfort Index	2015-2019	1540	4.31	2.11	0.0	8.9
Temperature Discomfort Index	2015-2019	1540	7.31	2.99	0.0	12.8
Panel C: Real Estate Data						
Total Number of Dwellings Sold	2015-2019	141	2839.03	2770.74	62.0	14179.0
Average Price per Square Meter (in Euros)	2015-2019	141	1171.62	521.56	595.1	3772.0
Mean Absorption Time (in months)	2015-2019	114	7.91	2.64	4.2	15.8
Panel D: Social Media Data						
Number of posts	2015-2019	228070	34.66	147.57	1.0	5498.0
Sentiment Score	2015-2019	315784	0.57	0.14	0.0	1.0
Panel E: Municipality Characteristics						
Population	Time-invariant	219	160964.29	104007.19	19148.0	544851.0
Share of Green Areas (in 2015)	Time-invariant	219	10.52	7.95	0.0	31.9
Share of Adults with Higher Education	Time-invariant	219	17.78	6.25	9.1	33.6
Airport (dummy)	Time-invariant	219	0.22	-	-	-
Port (dummy)	Time-invariant	219	0.29	-	-	-
Beach (dummy)	Time-invariant	219	0.78	-	-	-
Median Income (in 2015)	Time-invariant	219	1110.82	230.48	856.3	1775.9

4.3.1 Weather data

Similarly as in Chapter 3, the weather data is retrieved from the NASA MERRA-2 project [50] (see Section 3.2.1 for more details). We construct daily aggregates for each municipality from the raw grid-level and hourly MERRA-2 data[8]. Panel A in Table 4.1 includes the Summary statistics of these variables. Municipality-level daily temperatures range from 1.4°C to 42.8°C, and the 5th and 95th percentiles of temperature are reached at 9°C and 31°C, respectively. The sample period includes 472 days with temperatures above 30°C (which we define as warm days) and 513 days

[8]Hourly humidity is averaged to daily level. Precipitation and $PM_{2.5}$ levels are summed. We construct three measures of temperature: maximum, minimum, and mean. Maximum temperature, however, is most likely to reflect the day-time temperature that people actually interact with. Therefore, we use maximum temperature as our main daily temperature measure.

with maximum daily temperatures below 10°C (which we define as cold days)[9].

For our real estate analysis, we build annual measures of temperature stress from the daily data. Panel B in Table 4.1 describes the distribution of the annual measures of temperature stress in Portugal during the sample period. We compute annual averages of humidity, and aggregate measures of precipitation and air pollution. Fig. 4-2 illustrates the average annual number of warm days by municipality (with temperatures above 30°C, left), and the average annual number of cold days by municipality (with maximum temperatures below 10°C, right). Yearly counts of warm days range from fewer than 5 in some municipalities of Leiria, to 100 in the municipality of Serpa during the year 2017. Most municipalities in the region of Faro are never exposed to cold temperatures, while the municipality of Bragança has over 90 cold days per year on average.

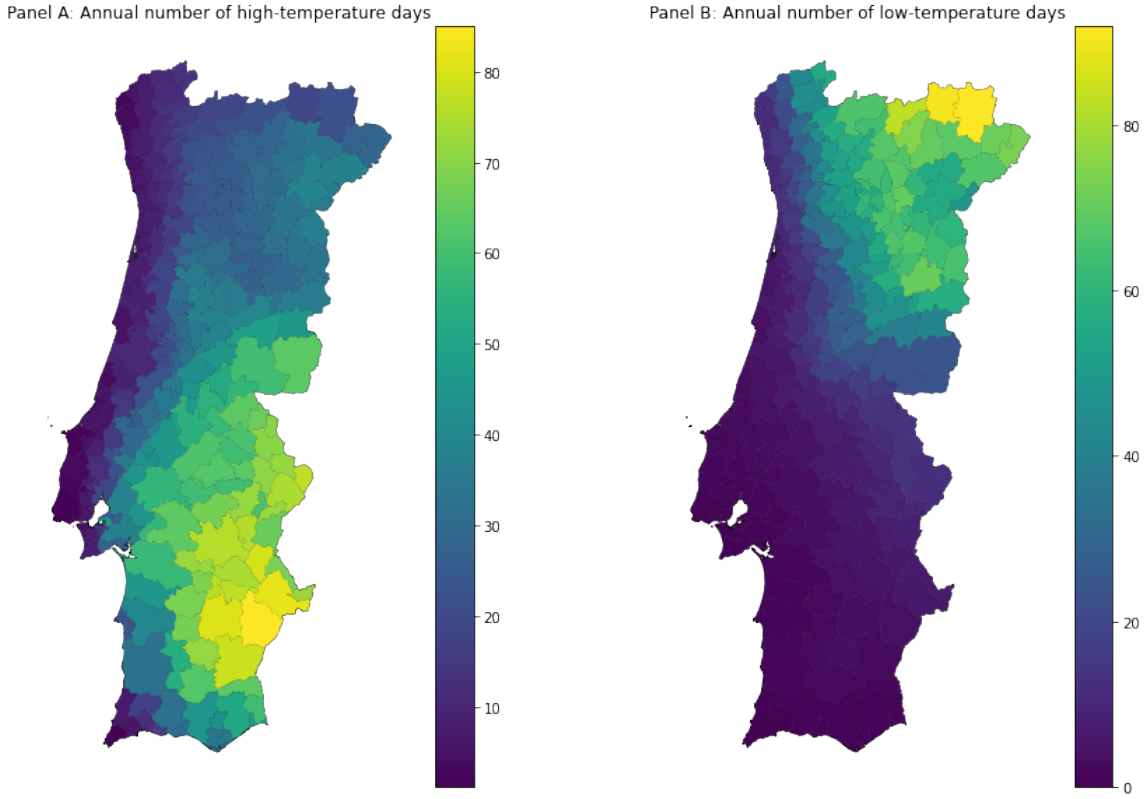

Figure 4-2: Weather data in Portugal

Left, the number of annual days with maximum temperatures above 30°C. Right, the number of annual days with maximum temperatures below 10°C. Numbers are averaged over the five years of our analysis (2015–2019).

[9]These definitions are similar to those used by Albouy et al. (2016) in a USA setting, where temperature is split into four bins with breakpoints at 45°F (7°C), 65°F (18°C), and 80°F (27°C) [4].

In addition, we construct yearly measures of summer and winter temperatures, by averaging temperature in summer-months (June to August) and winter-months (December to February), respectively. Summer temperatures range from 20.6°C to 33°C in the country, while winter temperatures are comprised between 8.1°C and 17.9°C. These ranges highlight substantial heterogeneity within the country in terms of exposure to extreme temperatures.

Building on the temperature discomfort index proposed by Zheng, Fu, and Liu (2009) [144], and used in prior literature on the impact of extreme temperatures on real estate value [145], we define season-specific "summer discomfort" and "winter discomfort" indices. The summer discomfort index (SDI_{it}) quantifies how warm a given municipality's summer is compared to the municipality with the coolest summer in the country that year. Therefore, a municipality with a summer discomfort index value of 2 denotes that the average summer temperature in that municipality that year were 2°C above the the average summer temperature in the municipality with the lowest average summer temperature. Similarly, the winter discomfort index (WDI_{it}) describes how cold their winter is compared to the municipality with the most temperate winter in the country that year. Season discomfort indices take a high values when the municipalities have especially harsh summers (for SDI_{it}) or winters (for WDI_{it}). These municipalities make for more unpleasant places to live during those seasons.

More formally:

$$SDI_{it} = \sqrt{(summerT_{it} - min_i(summerT_{it}))^2} \tag{4.1}$$

$$WDI_{it} = \sqrt{(winterT_{it} - max_i(winterT_{it}))^2} \tag{4.2}$$

where $summerT_{it}$ (respectively, $winterT_{it}$) is the mean temperatures in the summer (respectively, winter) of municipality i in year t.

We also compute an "overall temperature discomfort index" (DI_{it}, at location i and in year t), defined by Zheng, Fu, and Liu (2009) [144], which combines the two seasonal indices, such that:

$$DI_{it} = \sqrt{(winterT_{it} - max_i(winterT_{it}))^2 + (summerT_{it} - min_i(summerT_{it}))^2} \tag{4.3}$$

This overall temperature discomfort index accounts for municipality-level discomforts due to relatively-warm summers and relatively-cold winters. High levels of overall temperature discomfort mean harsher seasonal temperatures, and denotes

a municipality as a less desirable place to live throughout the year.

Distributions of the summer and winter discomfort indices are provided in Fig. 4-3 (left). Consistent with season temperature ranges, the summer discomfort index presents more variation between different municipalities. A large part of the municipalities is concentrated around index values of 4 to 8, indicating summer temperatures 4°C–8°C higher than the coolest summer temperatures in the country that year. The distribution of the winter discomfort index is concentrated around several distinct peaks, highlighting different winter regimes in different parts of the country. The distribution of the overall temperature discomfort index is provided on the right-hand side panel of Fig. 4-3. A standard deviation change in the overall discomfort index is equivalent to a value change of 3—which represents a significant change in environmental conditions for average seasonal temperatures.

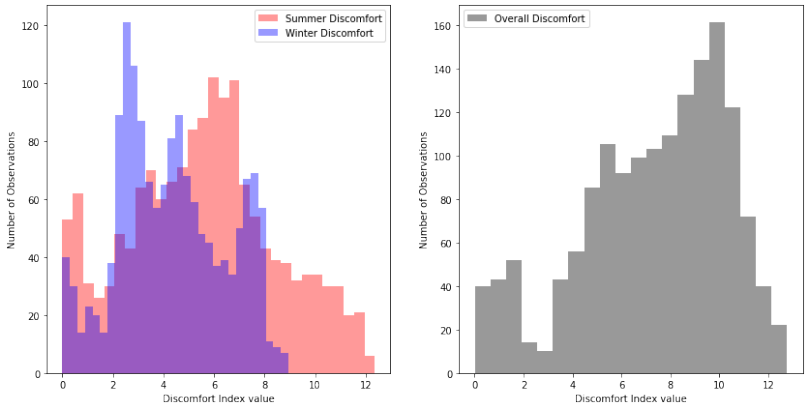

Figure 4-3: Distribution of temperature discomfort indices

Left, the distributions of the summer and winter discomfort indices, calculated based on Equations 4.1 and 4.2, respectively. Right, the distribution of the overall temperature discomfort index, calculated based on Equation 4.3.

4.3.2 Real estate data

Several characteristics of Portugal's real estate market make it a specifically relevant country for our analysis. The population of Portugal is highly urbanized: 61% of the population in Portugal live in cities [23], and around 50% of the population is concentrated in the four urban areas of our analysis. Portugal is among the European countries with the highest housing stock per 1,000 citizens [99][10].

[10]National policy-making has encouraged the rehabilitation of the housing stock, facilitated access to bank credit and low-interest rates on mortgages, and promoted affordable housing [47, 118]. While

Portuguese residential real estate market data is retrieved from the Residential Information System (SIR), provided by Confidencial Imobiliário[11]. SIR data aggregates information reported by more than 600 real estate agents, developers, investors, and credit institutions in the country. SIR provides statistical data by location, typology, and market range. It is the reference database for housing asset transaction prices in Portugal, and is used by all major Portuguese banks to analyze financing opportunities and to monitor the market. Our final data set consists of yearly aggregates covering the geographical areas of Great Lisbon, Great Porto[12], Braga, and Coimbra, collected between the first quarter of 2015 and the last quarter of 2019. A map of the data coverage is provided in Fig. 4-4. According to the preliminary results of the Portuguese 2021 Census, our data set contains the ten most populous municipalities of Portugal and accounts for 50% of the residential population of the country. The raw data is pooled at a municipality level—the second-largest subnational tier in Portugal (or administrative-2 level). Our final data set accounts for 29 municipalities in our four urban areas.

The real estate data set includes a number of annual indicators describing the state of the market that year. The total number of sold dwellings is an estimation of the total number of housing transactions in the municipality that year: it ranges from 62 in the municipality of Espinho (Greater Porto) in 2017 to 14,179 in central Lisbon, also in 2017. While Greater Lisbon and Porto dominate the housing market, all four urban areas saw increases in the number of sales in the study period. We also collect the yearly mean price per square meter of real estate transactions in each municipality. This metric is constructed by calculating the ratio between the offer value of the dwelling (in euros) and its private gross area (in square meters). The mean price per square meter rose in all four metropolitan areas between 2015 and 2019, with Lisbon displaying the highest values (see Fig. C-1 in Appendix Section C.3.1). Finally, we include the absorption time—defined as the time passed between the initial dwelling offer and the transaction—as a control in our models. Absorption time is accounted for in months, and municipality-level average yearly values for that variable range from 4 months to 16 months. More detailed summary statistics of all three variables are included in Table 4.1 (Panel C).

this has stimulated house ownership in the country—77% of the population are homeowners [45], well above the 70% European Union average—it also make the Portuguese population highly exposed to the financial risks of weather events on real estate.

[11] `https://www.confidencialimobiliario.com/`

[12] For reasons of data availability, we only have data for 10 of the 17 municipalities in Greater Porto, including the most populous ones of Porto and Vila Nova de Gaia

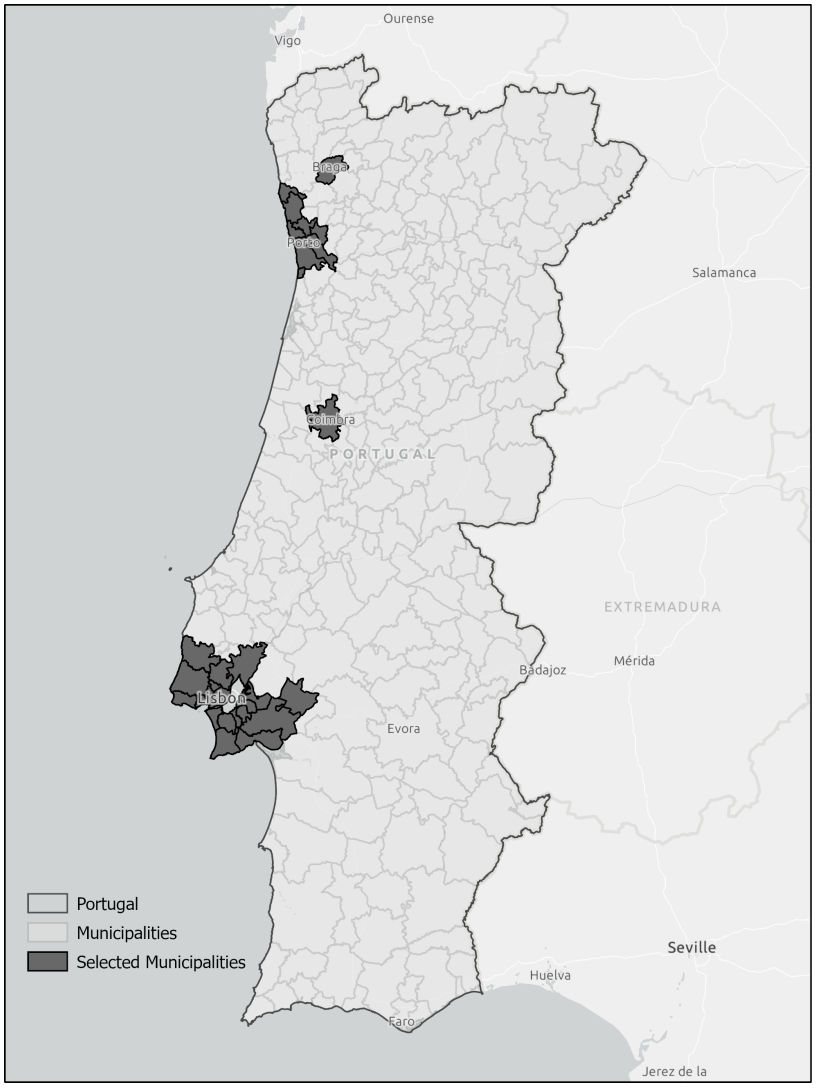

Figure 4-4: Real estate data coverage

Real estate data covers the four largest metropolitan areas in Portugal: Greater Lisbon, Greater Porto, Braga, and Coimbra.

4.3.3 Social media data and sentiment imputation

We rely on a sample of the novel social media data set presented in Chapter 2 to study perception of temperature discomfort. This data captures changes in individual sentiment at high frequency and high spatial resolution, and provides flexibility when matching to the real estate and weather data. Our data extraction covers Twitter posts shared in Portugal between January 1, 2015, and December 31, 2019. In the study period, nearly 8 million geotagged posts were collected.

Geographic distribution of Twitter users (aggregated to municipality-level) is pro-

vided in Fig. 4-5 (left). We test the spatial representativeness of our sample by comparing, at municipality level, the sample size of social media users[13] with the population size reported in the 2021 census of the Portuguese population. The sample size of Twitter users closely tracks the overall population, as indicated in Fig. 4-5 (right): the Pearson correlation coefficient between the two is $\rho = 0.81$ ($p < 0.001$).

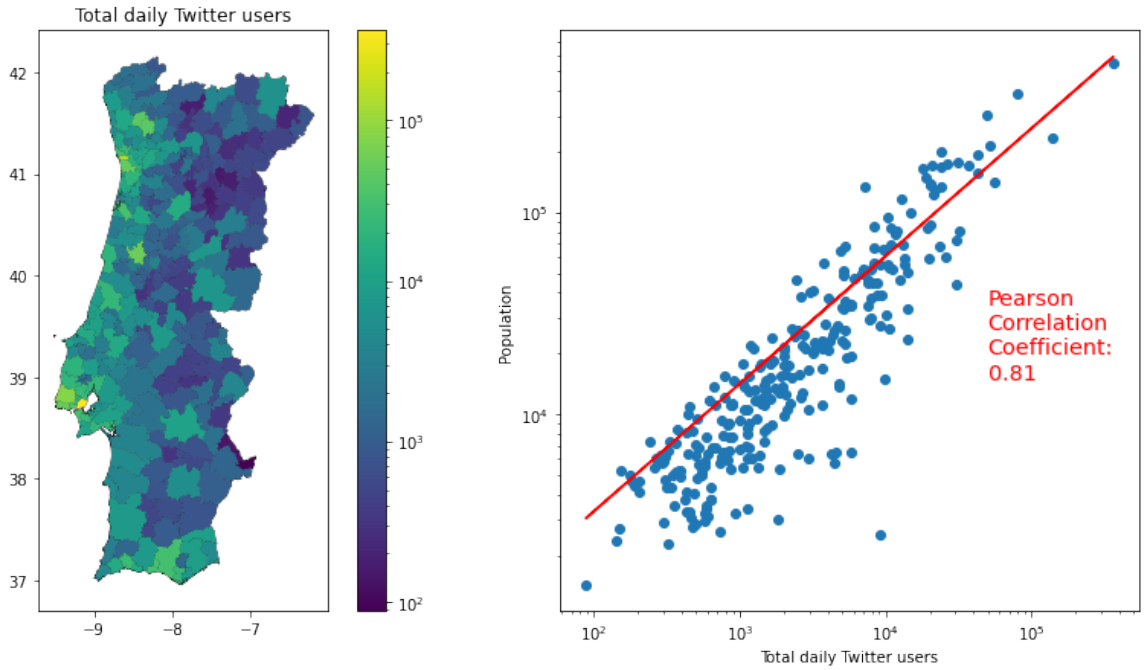

Figure 4-5: Social media data coverage

Logged-Number of daily tweets by municipality (left) and logged-Number of daily tweets by logged-population (right). Data coverage is restricted to Portugal, over the 2015-2019 period. The distribution of social media data is concentrated around the main coastal cities of the country. The Pearson correlation coefficient of the two variables in the right-hand panel is ρ=0.81 ($p < 0.001$).

The sentiment of every social media post is imputed using the NLP method described in Chapter 2 (Section 2.2.2). We aggregate our post scores to the daily-municipality level using the two-step aggregation method described in Section 2.2.3. Summary statistics are provided in Table 4.1 (Panel D). At the municipality level, the average daily sentiment score is 0.57, based on an average of 35 daily posts by municipality. Fig. 4-6 provides some additional explanatory trends in the sentiment index we obtain. On the left, the distribution of daily municipality-level sentiment index values indicates that most of the scores are concentrated between 0.5 and 0.7

[13]The sample size of social media users is the sum of unique users every day, so represents an inflated number of overall users.

(on a scale from 0.0 to 1.0)[14]. The right-hand side panel illustrates important weekly cyclicity in the data. Sentiment drops 6.2% of a standard deviation between an average Sunday and an average Monday, and 16.0% of a standard deviation between an average Sunday and an average Tuesday. Seasonal cyclicity is also present; month and day-of-week fixed effects are included in our weather and sentiment models to avoid cyclical trends from impacting our results.

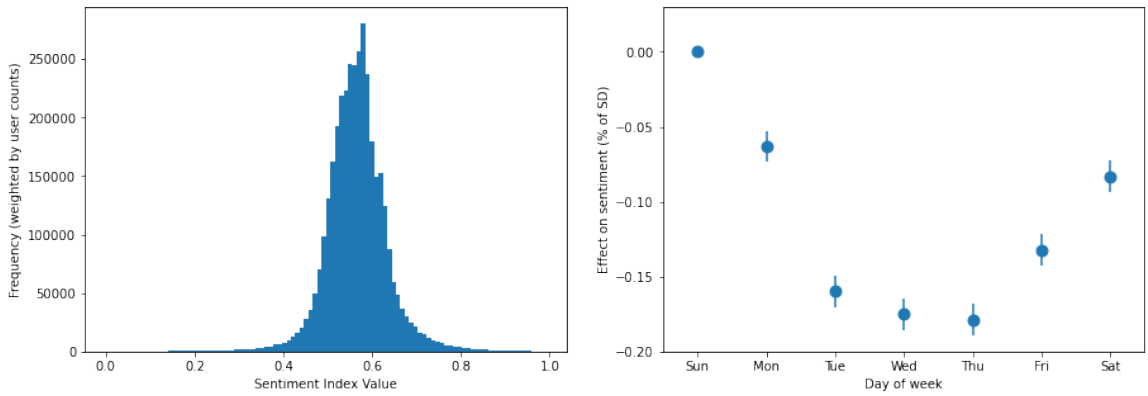

Figure 4-6: Daily, municipality-level sentiment score description

Left: Distribution of the social media-based sentiment index, weighted by the number of social media users. Most sentiment index values are concentrated between 0.5 and 0.7. Right: Effect of weekday on sentiment levels. This reveals high levels of weekly cyclicity in the data.

4.3.4 City-level controls data

The real estate market is highly influenced by location-specific attributes. To account for this, we collect an extensive set of municipalities characteristics for our real estate model. The set of controls we include replicates—to the extent that the data was available—the conceptual framework of Zheng, Kahn, and Liu (2010)[15] [145]. The presence in the municipality of a beach (dummy), airport (dummy), and port (dummy) are all collected from the OpenStreetMap portal[16]. The availability of green areas by municipality is collected from the "Observatório do Ordenamento territorial e urbanismo"[17]. Information relative to employment and income (in 2015) is provided

[14]The distribution is weighted by the number of social media users, consistently with all of the modeling performed in the latter sections of this chapter.

[15]Foreign direct investment per municipality, which was used as a control in Zheng, Kahn, and Liu (2010), was unavailable for Portugal.

[16]`https://www.openstreetmap.org`

[17]`https://www.dgterritorio.gov.pt/`

by the "Pordata" platform[18]. Finally, we use data from the national Portuguese census to include the municipality-level share of population with higher education as a control. A detailed description of how these features were extracted is provided in Appendix Section C.3.2, and summary statistics are provided in Table 4.1 (Panel E).

4.4 Methods

Our econometric analysis is conducted in two steps, following the logic of our conceptual framework (Fig. 4-1). In the first part, we estimate the direct damage function of temperature discomfort on real estate prices by applying a regression model (see Section 4.4.1).

In the second part, we construct an integrated model to estimate how perceived (sentiment) damages of temperature stress moderate changes in real estate prices induced by climate events (see Section 4.4.2). In this part of the analysis, we first estimate the direct damage of temperature discomfort on sentiment; then, we construct municipality-specific sentiment-damage coefficients; and finally, we inject these coefficients into the real estate regression model.

4.4.1 Direct impact of temperature stress on real estate value

We first estimate the impact of temperature discomfort on real estate value. Given the temporal granularity of the real estate data, these models are run at a yearly, municipality-level scale:

$$value_{it} = \alpha_0 + \alpha_1 D_{it} + \alpha_2 X_{it} + T_t + \epsilon_{it} \tag{4.4}$$

The dependent variable $value_{it}$ stands for the natural logarithm of the average price per square meter in municipality i and at year t. The independent variable of interest D_{it} is a measure of temperature discomfort for municipality i at year t. We measure temperature discomfort in multiple ways to ensure the robustness of our results. In particular, we define it either with the overall Discomfort Index (DI_{it}), with the Summer and Winter Discomfort Indices (SDI_{it} and WDI_{it}), or as the number of hot and cold days (Hot_{it} and $Cold_{it}$). We standardize all explanatory variables to facilitate coefficient comparisons across different definitions of temperature stress.

Our coefficient of interest, α_1, characterizes the nature of the relationship between temperature discomfort and real estate value. A negative value of α_1 means that, all

[18] https://www.pordata.pt

other things being equal, an increase of one standard deviation in the temperature discomfort index is associated with a percentage drop in the mean square-meter real estate price of that value.

The observed relationship between temperature discomfort and real estate value might be due to omitted factors, such as concurrent weather events or intrinsic city characteristics. For example, cities that are exposed to extreme weather might be poorer and offer fewer amenities. To mitigate these effects, we include several controls (X_{it}): yearly weather aggregates (air pollution, precipitation, and humidity, described in Section 4.3.1), municipality-level characteristics (green area, share of population with higher education, presence of airports, ports, and beaches, population, and income level, described in Section 4.3.4), and additional real estate controls (total number of dwellings sold and mean absorption time, described in Section 4.3.2). All non-dummy controls are standardized.

To account for overall trends in the real estate market, we also include year fixed effects (T_t). The standard errors are clustered at the municipality level to account for the correlation of transactions within the municipality.

4.4.2 Incorporating subjective perceptions of temperature discomfort in real estate value modeling

In a second step of our analysis, we quantify to what extent perceived temperature damages on individual sentiment in each municipality serve as a predictor of the damages in real estate value. To build this model, we start by estimating the main effect of temperature on sentiment, then construct municipality-specific sentiment-damage coefficients. These coefficients are then incorporated into updated real estate value estimations.

4.4.2.1 Effect of temperature on sentiment

To study the main effect of temperature on expressed sentiment, we estimate Equation 4.5 using a fixed effect time-series regression model similar to that used in Baylis (2020) [9] or Wang, Obradovich, and Zheng (2020) [136]:

$$sentiment_{it} = \alpha_0 + \alpha_1 f(Temp_{it}) + \alpha_2 X_{it} + T_t + \gamma_i + \epsilon_{it} \tag{4.5}$$

The dependent variable $sentiment_{it}$ stands for our standardized[19] measure of sen-

[19]Standardization is conducted at the municipality-level to account for inherent differences in

timent in municipality i and on date t. $Temp_{it}$, represents the maximum daily temperature measured in a given municipality, and the function f transforms the continuous values into discrete 5°C bins. Each bin's impact on sentiment is estimated by the set of coefficients α_1. This method allows for a flexible, non-linear estimation of the relationship between temperature and expressed sentiment. In each regression, one temperature bin is omitted and serves as the reference sentiment measure (in the results of Fig. 4-8, for instance, the reference temperature bin is 15°C–20°C). Therefore, the coefficients of interest α_1 can be interpreted as the marginal effect of a given temperature bin relative to the reference bin.

X_{it} is a set of daily and municipality-level environmental controls including air pollution ($PM_{2.5}$), precipitation, wind speed, cloud coverage, and humidity. Unobserved factors specific to locations and time periods might affect sentiment in ways that correlate with temperature and environmental measures. Different municipalities might have inherent differences in sentiment levels linked to economic or cultural factors. Seasonality might also alter expressed well-being. We account for these spatial and temporal cofounders by including temporal (T_t for year, month, and day-of-week) and location (γ_i, at municipality level) fixed effects.

Here too, standard errors are clustered at municipality level. Finally, the main regression is run by weighting each observation by the number of social media users recorded within that municipality and on that day.

4.4.2.2 Heterogeneous effects of temperature on well-being: Damage coefficients at municipality-level

Estimating the binned regression model presented in Equation 4.5 on our data set provides a non-linear response function linking temperature stress to sentiment in the entire country. However, this might overlook important heterogeneity between regions—due to differences in adaptation and mitigation mechanisms, for instance. We construct a single heat-related damage coefficient in each municipality by characterizing the sentiment shock associated with temperatures beyond a given threshold of 30°C. Equation 4.6 is estimated in each of the 29 municipalities of our real estate data coverage, using daily measures of sentiment and temperature.

$$sentiment_{it} = \alpha_0 + \alpha_1 Hot_{it} + \alpha_2 X_{it} + T_t + \gamma_i + \epsilon_{it} \tag{4.6}$$

The left side of the equation is the sentiment index described above for Equa-

happiness levels between cities.

tion 4.5. For the main independent variable, we replace temperature bins with a dummy variable (Hot_{it}) equal to 1 if the maximum temperature of location i on day t is above 30°C[20]. Similarly to Equation 4.5, we include environmental controls, year, month, day-of-week fixed effects, and municipality-level location fixed effects. The α_1 coefficient we obtain for each region is the sentiment drop associated with extreme warm temperatures in that municipality, and we interpret this coefficient as the location-specific sentiment damage of extreme warm temperatures.

The distribution of the coefficients for the 29 municipalities of our real-estate analysis is provided in Fig. 4-7. All but 3 of the damage coefficients are strictly negative, indicating that high-temperature days are consistently associated with a drop in expressed well-being[21]. A sentiment-damage dummy is defined based on the median sentiment damage value (the vertical red line in Fig. 4-7): the dummy is equal to 0 in low-damage municipalities, where α_1 less negative than the median (or above the median, right-hand side of the graph), and to 1 in high-damage municipalities, where the coefficient is more negative than the median (or below the median, left-hand side of the graph).

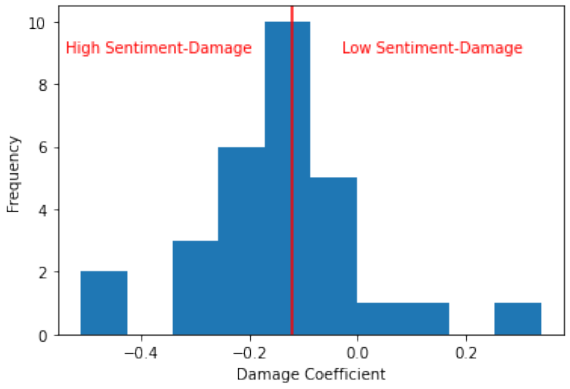

Figure 4-7: Histogram of the sentiment damage coefficient

The red line marks the median value. Municipalities with damage coefficients more negative than the median are classified as "high sentiment-damage", whereas municipalities with damage coefficients above the median are flagged as "low sentiment-damage".

[20]The choice of 30°C is based on the temperature distribution described in Section 4.3.1)

[21]Of the three municipalities where the damage coefficient is positive, two have very high P-values (above 0.8), so are insignificant under any threshold. Only one municipality has a significant positive effect.

4.4.2.3 Adding sentiment to real estate modeling: Moderated effect of temperature stress on location value

The core analysis of our study investigates whether perceived sentiment responses to temperature discomfort contribute to the price drop we observe in the real estate market. We integrate the municipality-level sentiment damage dummy into our real estate model in two ways: first as a moderator (by interacting it with the value of temperature discomfort), and secondly as a classifier (by splitting our municipalities based on the value of the sentiment-damage dummy and running separate estimations on the high- and low-damage groups).

To test the relevance of our sentiment damage as a moderator, we define a new real estate value model (Equation 4.7) where municipality-level sentiment damage is interacted with temperature discomfort.

$$value_{it} = \alpha_0 + \alpha_1 SD_i * DI_{it} + \alpha_2 X_{it} + T_t + \gamma_i + \epsilon_{it} \tag{4.7}$$

Equation 4.7 is a slight modification of Equation 4.4: the additional independent variable SD_i is the value of the sentiment-damage dummy in location i—or whether extreme warm temperatures in municipality i are associated with a higher-than-median sentiment drop. The coefficient α_1 is an updated estimation of the impact of temperature discomfort on real estate which accounts for heterogeneous subjective perception of these climate events.

Finally, we test whether the real-estate value drops associated with temperature discomfort are more substantial in *specific* municipalities where the sentiment-damage coefficient is higher. We split the sample into high sentiment-damage and low sentiment-damage municipalities, based on the value of the sentiment-damage dummy variable. We estimate weather damages separately in the two samples using the model defined in Equation 4.4 (direct effect of temperature discomfort on real estate value). We obtain two separate estimate equations:

$$value_{it} = \alpha_{0,low} + \alpha_{1,low} D_{it} + \alpha_{2,low} X_{it} + T_t + \epsilon_{it} \tag{4.8}$$

$$value_{i't} = \alpha_{0,high} + \alpha_{1,high} D_{i't} + \alpha_{2,high} X_{i't} + T_t + \epsilon_{i't} \tag{4.9}$$

Equation 4.8 is run only on municipalities i that are in the half of the sample where the sentiment-damage dummy is equal to 0—i.e., with a sentiment-damage below the sample median. Similarly, Equation 4.9 is only run on the other half of the municipalities i', where the sentiment-damage dummy is equal to 1—i.e., with a

sentiment-damage above the sample median. Comparing the values and significance of $\alpha_{1,low}$ and $\alpha_{1,high}$ informs on the relative impact of weather events on real estate value in each subset of municipalities.

4.5 Results and discussion

This section presents the results of our econometric analysis. We first describe the estimates of the damages of temperature on real estate value, and then incorporate the role of sentiment damages as predictor of real estate value drops.

4.5.1 Temperature stress and real estate value

Table 4.2 presents the results linking objective measures of temperature discomfort to real estate prices varying the sets of controls included in the regression model. Column (1) reflects the baseline effect of temperature discomfort on the mean square-meter value. Column (2) includes additional weather controls: yearly aggregates of PM-2.5 air pollution and precipitation, and average humidity. In Column (3), we replace the weather controls with time-invariant municipality characteristics: the share of green areas in the district, the logged share of the professional population with a higher education degree, the logged population, and the median income of the municipality in 2015. Column (4) introduces a final set of controls linked to the location's real estate market that year: overall number of dwellings sold, and average absorption time (in months). Finally, Column (5) includes all the controls.

The results are presented in the form of three panels for the three measures of temperature discomfort. Panel A uses the Temperature Discomfort Index (DI) as the main independent variable. In Panel B, we decompose the overall index into Summer and Winter Discomfort indices (SDI and WDI) to understand if the effect is driven by one of the two seasons exclusively. In Panel C, we opt for a combination of the number of hot days and the number of cold days (Hot and $Cold$) as a robustness check.

Table 4.2: Impact of Temperature Discomfort on Mean Square Meter Price

	Dependent variable:				
	Logged Price per Square Meter				
	(1)	(2)	(3)	(4)	(5)
Panel A					
Discomfort Index	-0.20^{***}	-0.20^{***}	-0.15^{***}	-0.16^{***}	-0.14^{***}
Air Pollution		-0.06^{***}			0.005
Precipitation		0.10			0.01
Humidity		0.11^{*}			0.03
Logged Green Area			-0.03		-0.02
Logged Higher Ed			0.18^{***}		0.21^{***}
Airport (dummy)			-0.02		-0.02
Port (dummy)			0.03		0.01
Beach (dummy)			-0.03		-0.03
Logged Population			0.01		-0.06
Median Income			0.07		0.002
Total Dwellings Sold				0.14^{***}	0.11^{**}
Mean Absorp. Time				-0.05	-0.01
Panel B					
Summer Discomfort	-0.19^{***}	-0.20^{***}	-0.14^{***}	-0.15^{**}	-0.14^{***}
Winter Discomfort	-0.18^{***}	-0.20^{*}	-0.12^{***}	-0.16^{***}	-0.33^{**}
Air Pollution		-0.05^{**}			0.01
Precipitation		0.07			-0.002
Humidity		0.02			-0.21
Logged Green Area			-0.03		-0.02
Logged Higher Ed			0.18^{***}		0.21^{***}
Airport (dummy)			-0.01		-0.02
Port (dummy)			0.03		0.01
Beach (dummy)			-0.03		-0.03
Logged Population			0.02		-0.07
Median Income			0.05		-0.003
Total Dwellings Sold				0.15^{***}	0.12^{**}
Mean Absorp. Time				-0.06	-0.02
Panel C					
Nb of Hot Days	-0.12^{***}	-0.14^{***}	-0.09^{***}	-0.07^{*}	-0.08^{***}
Nb of Cold Days	-0.16^{***}	-0.14^{***}	-0.11^{***}	-0.16^{***}	-0.15^{***}
Air Pollution		-0.04			0.0001
Precipitation		0.05			0.02
Humidity		0.06			-0.02
Logged Green Area			-0.05		-0.03
Logged Higher Ed			0.21^{***}		0.22^{***}
Airport (dummy)			-0.02		-0.03
Port (dummy)			0.01		-0.01
Beach (dummy)			-0.03		-0.03
Logged Population			0.04		-0.05
Median Income			0.04		-0.01
Total Dwellings Sold				0.17^{***}	0.12^{**}
Mean Absorp. Time				-0.08^{**}	-0.02
Weather Controls	No	Yes	No	No	Yes
City Controls	No	No	Yes	No	Yes
Housing Controls	No	No	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	156	156	156	120	120
Note:					*p<0.1; **p<0.05; ***p<0.01

The estimates of Panel A describe that the Temperature Discomfort Index has a significant and negative impact on real estate value in Portugal. One standard deviation increase in the discomfort increase is associated with a 14% to 20% drop in mean square-meter price, depending on the controls that are included (the coefficient is significant at the 1% level regardless). Air pollution is also associated with a significant drop in real estate value (5% drop for an increase of one standard deviation) when included with other weather controls—however, this effect disappears when including additional city and real estate market controls.

In Panel B, both summer and winter discomfort are associated with significant drops in real estate value, although the summer discomfort coefficient is more robust to different model specifications. With the full set of controls, a standard deviation

increase in summer discomfort is associated with a 14% drop in real estate value. These results hold in Panel C: a 1 standard deviation increase in the number of hot (equivalent to 3 days, based on an average of 5) or cold (equivalent to over 2 days, based on an average of 4) days is also associated with significant drops in real estate value. In sum, the results consistently indicate that experiencing higher temperature distress is associated with price discounts in real estate markets.

The magnitude of these results are in line with the findings of Albouy et al. (2016) in the United States: individuals are willing to pay 0.1 standard deviations more to avoid extreme-cold days and 0.2 standard deviations more to avoid extreme-hot days, relative to moderate days [4]. When we express our results in terms of standard deviations, we find that a standard deviation increase in the number of extreme-warm days (above 30°C) is associated with a drop of 0.14 of a standard deviation in real estate value (standardized estimation results are presented in Appendix Section C.4.1). While the magnitudes seem large in terms of absolute changes in real estate value, the increase in the temperature discomfort indices also represents a significant change in weather conditions. For the summer discomfort index, for instance, it would be equivalent to an increase of 3°C in average summer temperature—a substantial change even compared to the current projections of climate change. Similarly, when considering the extreme-temperature days, a standard deviation increase represents a 60% increase in the number of hot days compared to the mean.

4.5.2 Temperature stress and subjective well-being

Incorporating subjective perceptions of climate events in our analysis, we start by evaluating the direct effect of temperature on sentiment in Portugal. Fig. 4-8 documents the sentiment response to each temperature bin using Equation 4.5. Because the 15–20°C bin is used as our reference temperature bin, estimates can be interpreted as a change in sentiment relative to having the same day under 15–20°C temperatures. The figure also includes a histogram underneath the plot to visualize the temperature distribution, and the 95th and 99th percentiles of temperature are marked with vertical two dotted red lines. Since our sentiment measure is standardized, the values are expressed in standard deviation (SD) percentages. The 95% confidence intervals of the estimates are visualized in shaded blue.

Consistent with prior literature on the impact of temperature on sentiment [9, 136], we find that high temperatures are associated with a significant drop in expressed well-being compared to the reference temperature bin of 15-20°C. The difference between

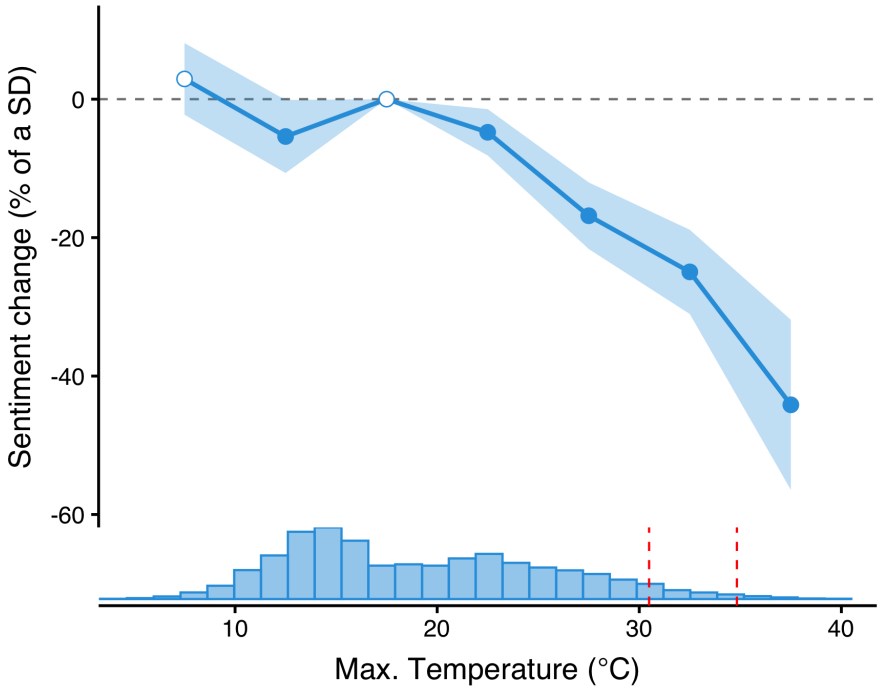

Figure 4-8: Impact of temperature on sentiment

Based on data from 2015-2019.

the warmest days (above 35°C) and the days in the reference temperature bin is around 45% of a standard deviation. These results are highly significant, despite confidence intervals widening slightly for the highest temperatures.

The magnitude of the results is also of the same order of magnitude–if slightly more substantial—that results from the prior literature. Using the same model specifications, Baylis (2020) finds temperatures above 40°C are associated with a sentiment drop of 21% of a standard deviation in the United States, and a drop of between 10% and 30% of a standard deviation in Australia, India, and South Africa [9]. Wang, Obradovich, and Zheng (2020) express their sentiment drop in absolute terms, and find a 2% drop in sentiment when temperatures are above 35°C [136]: when expressing our results in absolute terms, we find a slightly larger drop of 3% in sentiment for the same temperature bin (see Appendix Section C.4.2).

We also examine the impact of other weather events on sentiment in Portugal, such as air pollution. We find that increases in $PM_{2.5}$ levels are also associated with decreased well-being. These results are presented in the Appendix Section C.4.3.

4.5.3 Integrated model of temperature discomfort, sentiment, and real estate value

The full-sample analysis we conduct in Section 4.5.2 highlights important sentiment damages associated with high temperatures in Portugal. However, these damages differ by municipality, and we estimate municipality-level damages in Section 4.4.2.2. Here, these municipality-level damages feed into an updated real estate analysis: first as moderators of the direct impact of temperature stress on real estate value, and then as classifiers distinguishing high sentiment-damage from low sentiment-damage municipalities.

4.5.3.1 Sentiment damage as a moderator

Table 4.3 presents the results when we incorporate the municipality-level sentiment-damage dummy as a moderator in the real estate estimation model (Equation 4.7). Column (1) reproduces the first column of Table 4.2 and estimates the baseline effect of temperature discomfort of mean square-meter value. Column (2) includes the additional moderation term, and reflect the baseline effect of our sentiment-moderated temperature discomfort variable on real estate prices. Columns (3) and (4) add weather, city, and real estate market controls to estimations (1) and (2), respectively. Similarly as for Table 4.2, the results are split into 3 panels for different definitions of temperature discomfort: the Temperature Discomfort Index (Panel A), Summer and Winter Discomfort (Panel B), and the number of hot and cold days (Panel C).

Table 4.3: Impact of Temperature Discomfort and Sentiment Damage on Mean Square Meter Price

	Dependent variable:			
	Logged Price per Square Meter			
	(1)	(2)	(3)	(4)
Panel A				
Discomfort Index	-0.20^{***}		-0.14^{***}	
High Sentiment Damage (dummy):Discomfort Index		-0.27^{***}		-0.10^{**}
Air Pollution			0.005	0.06^{*}
Precipitation			0.01	-0.04
Humidity			0.03	0.04
Logged Green Area			-0.02	-0.01
Logged Higher Ed			0.21^{***}	0.23^{***}
Airport (dummy)			-0.02	-0.04
Port (dummy)			0.01	-0.02
Beach (dummy)			-0.03	-0.02
Logged Population			-0.06	-0.03
Median Income			0.002	-0.03
Total Dwellings Sold			0.12^{**}	0.10
Mean Absorp. Time			-0.01	-0.06^{**}
Panel B				
Summer Discomfort	-0.19^{***}		-0.14^{***}	
Winter Discomfort	-0.18^{***}		-0.33^{**}	
High Sentiment Damage (dummy):Summer Discomfort		-0.24^{***}		-0.10^{*}
High Sentiment Damage (dummy):Winter Discomfort		-0.17^{***}		-0.08
Air Pollution			0.01	0.06^{*}
Precipitation			-0.002	-0.05
Humidity			-0.21	0.02
Logged Green Area			-0.02	-0.01
Logged Higher Ed			0.21^{***}	0.23^{***}
Airport (dummy)			-0.02	-0.04
Port (dummy)			0.01	-0.01
Beach (dummy)			-0.03	-0.02
Logged Population			-0.07	-0.04
Median Income			-0.003	-0.02
Total Dwellings Sold			0.13^{**}	0.10
Mean Absorp. Time			-0.02	-0.06^{**}
Panel C				
Nb of Hot Days	-0.12^{***}		-0.08^{***}	
Nb of Cold Days	-0.16^{***}		-0.15^{***}	
High Sentiment Damage (dummy):Nb of Hot Days		-0.22^{**}		-0.08
High Sentiment Damage (dummy):Nb of Cold Days		-0.15^{***}		-0.08^{*}
Air Pollution			0.0001	0.05
Precipitation			0.02	-0.04
Humidity			-0.02	0.02
Logged Green Area			-0.03	-0.01
Logged Higher Ed			0.22^{***}	0.23^{***}
Airport (dummy)			-0.03	-0.04
Port (dummy)			-0.01	-0.02
Beach (dummy)			-0.03	-0.02
Logged Population			-0.04	-0.04
Median Income			-0.01	-0.03
Total Dwellings Sold			0.13^{**}	0.11^{*}
Mean Absorp. Time			-0.02	-0.05^{*}
Weather Controls	No	No	Yes	Yes
City Controls	No	No	Yes	Yes
Housing Controls	No	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	156	156	120	120
Note:				*p<0.1; **p<0.05; ***p<0.01

Prior to adding the controls, moderated estimates of the temperature stress impact (Column 2) are similar to direct estimates (Column 1). Moderated overall discomfort is associated with a 27% drop in real estate value, up from 20% in the direct model. When decomposing the temperature stress by hot and cold, the moderated impact of high temperatures is consistently higher than the unmoderated version (24% up from 19% for summer discomfort, 22% up from 12% for number of hot days), and the moderated impact of low temperatures is slightly lower (17% compared to 18% for winter discomfort, 15% down from 16% for number of cold days). In the presence

of our complete set of controls, however, the moderated discomfort index loses in magnitude and significance.

4.5.3.2 Splitting the sample based on sentiment damage level

In the final step of the analysis, we explore the potential of our sentiment-damage dummy as a classifier, to detect municipalities most exposed to the negative impacts of extreme temperatures. Splitting the sample into high and low sentiment-damage municipalities allows us to estimate the effects of temperature stress on real estate separately for municipalities with high and low sentiment-damage.

The full results are presented in Appendix Section C.4.4 (Table C.4). As in previous sets of results, Panel A uses the overall discomfort index as a measure of temperature stress. In Panel B, temperature stress is expressed by both summer discomfort and winter discomfort. In Panel C, we use counts of extreme warm (temperatures over 30°C) and cold (maximum temperatures under 10°C) as measures of temperature discomfort. Column (1) estimates the baseline impact of the temperature discomfort metric on real estate value in low sentiment-damage municipalities. Column (2) runs the same estimation with the complete set of controls. In Columns (3) and (4), we include the same estimations on the high sentiment-damage municipalities. A visual comparison of the coefficients is also provided in Fig. 4-9.

Overall, the results indicate that temperature stress days have a negative impact on real estate value in both municipality groups. However, the magnitude of the drop is more sizeable in municipalities with high sentiment-damage: for instance, an increase of a standard deviation in the number of hot days (with no additional controls) is associated with a 25% drop in real estate value in these municipalities, while the same change results in only a 7% drop in low sentiment-damage regions. In low sentiment-damage regions, the significance of the coefficients associated with temperature discomfort also reduces as additional controls are added to the model. In high sentiment-damage municipalities, when all weather, location, and real estate market controls are included, a standard deviation increase in the number of days over 30°C is associated with a 9% drop in real estate value ($p < 0.01$).

These results highlight the value of our integrated approach. The overall results, connecting climate amenities to real estate market value, can be decomposed here based on subjective perception of these climate events. Regions where temperature stress is most damaging to well-being (or, high sentiment-damage municipalities) also face the sharpest drops in real estate value.

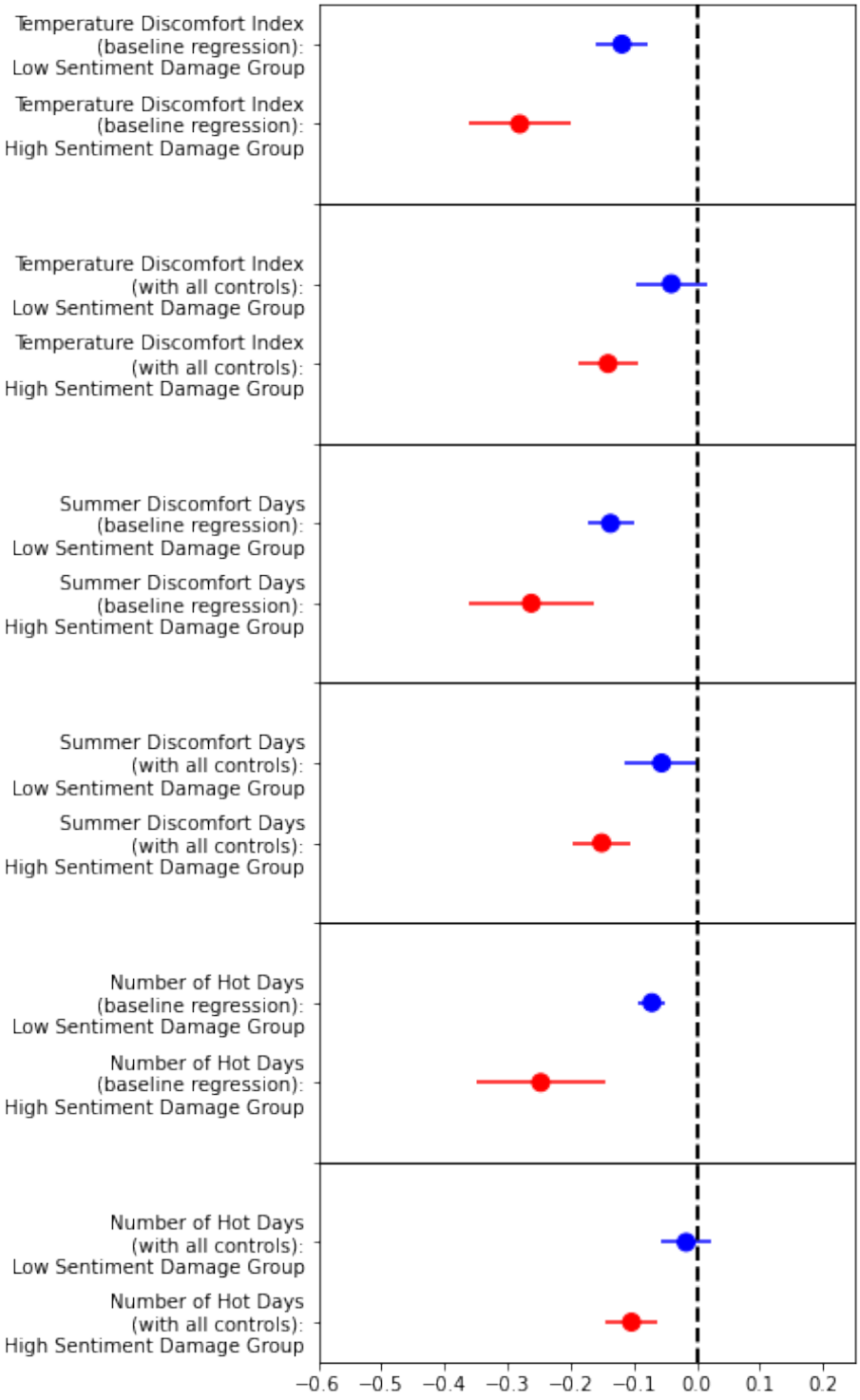

Figure 4-9: Impact of temperature discomfort on real estate value

4.6 Chapter conclusion

Extreme weather events incur important damages on urban environments, with dire impacts on citizens. Prior research has documented damages on psychological well-being, health outcomes, and mortality [9, 19, 37]. As weather conditions in cities become more unbearable, inhabitants are increasingly encouraged to move elsewhere, influencing local real estate markets in both locations.

Novel data sets offer researchers the opportunity to model and measure such trends [96]. For example, satellite weather data produces granular maps of exposure to weather events; geolocated social media data from platforms like Twitter provide local measures of preferences and subjective reactions to climate events. Prior research, including Chapter 3 of this thesis, has connected these two data sources to assess the impact of high temperatures on expressed well-being [136]. In this study, we replicate these results in the context of Portugal. We also add a layer of real estate data at the municipality level, and findings suggest that both weather events and subjective perception of these events are useful predictors of changes in housing value within Portugal's largest metropolitan areas. Indeed, we find temperature discomfort to be associated with significant drops in real-estate value, even after controlling for urban and real estate market characteristics. Social media-based measures of discomfort can serve as a useful (leading) indicator for policymakers as they seek to project the effect of rising temperatures on the real estate market.

The methodological approach of our study contains certain limitations. Contrarily to data collected through traditional surveys, our novel data set contains no socio-demographic information about Twitter users, and therefore about the representativeness of our sample. As detailed in Chapter 2, existing research has both validated the use of social media as a proxy of public opinion [25] and pointed out its biases [131]. The coverage of our other data sets also restrains us: our real estate data covers only four urban areas over a period of five years, prohibiting city fixed effects, for instance. Moreover, as the data we collect only tracks transactions, we do not have information regarding housing appraisals, raising an ongoing debate about how best to assess real estate value (see Pagourtzi et al. (2003) for a review of the literature on this issue [102]).

Our study has important theoretical implications. It clarifies the relevant role of subjective perception measures of weather events: these act as moderating factors when assessing the impact of these same events on real estate prices. Our study also raises practical implications for urban environments facing increasingly frequent

climate events, especially heat waves. We contribute further to the efforts to quantify the effects of these events on real estate markets, and propose our social media index as a relevant indicator for policy-makers to track.

Future work could extend the scope of this study to other countries and examine longer-term effects over more years. Additional indices of real estate value, for example, could also be considered (including some based on non-transactional home appraisals). Sentiment analysis is just one way of measuring subjective perception: survey-based metrics—like levels of belief in climate change in Portugal—or additional NLP-based analyses of social media data (such as topic modeling) could provide future researchers with a richer understanding of subjective reactions to temperature discomfort and its impact on real estate value.

Chapter 5

Conclusion

Social science research that resonates with policy-makers requires a granular understanding of individual well-being, preferences, and behavior. Technology shifts have challenged traditional ways of collecting public opinion (such as surveys and polls) but have also enabled the widespread adoption of new platforms where people share content. In this context, user-generated data from social media has emerged as a rich alternative to understand social dynamics. Further still, these data sets can produce estimates of policy-relevant metrics—such as real-time well-being—that were previously immeasurable. Building on this, my thesis has highlighted two case studies that harness a novel data set of social media content and advanced, multilingual methods of NLP.

In the first (Chapter 3), we examine well-being damages associated with high temperatures across the world. In this specific case, relying on global social media data allows us to study countries where traditional data is scarce, and that were previously excluded from similar research endeavors. We find important spatial heterogeneity in reaction to extreme temperature—evidence of how informative global data can be when addressing global issues like climate change.

In the second (Chapter 4), we go one step further and connect well-being shocks to real-world outcomes. The sentiment drops we find associated with temperature stress also drive drops in real estate market location value. Beyond well-being and mental health costs, social media sentiment is found to be indicative of economic outcomes—making it all the more relevant for governments and decision-makers to track it.

Social media data should not be seen as a golden bullet. Inherent biases in the user pool, as well as more insidious biases in the methods of analysis, make it an imperfect resource. Algorithmic vetting and validation studies, like the ones I conduct

in Chapter 2, are essential to better understanding when social media analysis is a relevant proxy for public opinion, and when it is not. We find that while there are strong disparities in social media use between countries worldwide, the data is more spatially representative within countries—aggregated global analyses should therefore systematically explore country-level heterogeneity as well. In highly connected countries like the United States, we find that social media-based sentiment closely track more traditional measures of public opinion and well-being. Moving forward, subsequent research in this field will also need to systematically confront new social media metrics with results from experiments, surveys, and real-world outcomes.

Social media data is becoming an important asset for academic research, especially in the burgeoning field of computational social science [76, 42]. With the growing imperative that policy-making be evidence-based [115], it seems likely that user-generated data—including from social media platforms—will increasingly make its way into government decisions as well, complementing more conventional data sources in impact assessment and policy evaluation pipelines. Global data sets, and multilingual text-analysis algorithms necessary to their analysis, are especially relevant in countries where traditional data infrastructure is weak.

It should finally be noted that mining user-generated text data is part of a larger trend incorporating digital platforms and technologies in government operations. More hands-on approaches—such as online polls, petitioning systems, and large-scale civic consultations—can also source policy-relevant information from a wide population base [15]. Such initiatives are flourishing across the world, on topics like Artificial Intelligence governance [87], public transport planning [84], or constitutional reform [98].

All of these mechanisms present some risks that governments will have to grapple with in the years to come—representativity, algorithmic bias, and data privacy, to name a few. However, they also provide policy-makers with the chance to account for increasingly large and diverse groups, as well as previously immeasurable outcomes, in their decision process. Democratic checks and balances will be necessary to guarantee that these new technologies carry out their promise of a more inclusive approach to policy-making.

Appendix A

Supplementary Material to Chapter 2

A.1 Acknowledgements

We gratefully thank Harvard Center for Geographic Analysis (CGA) for providing the Twitter data used in this study [79]. We thank Devika Kakkar and the Harvard CGA team for the support and consultation for scaling the sentiment and geography computation on this big data set using Harvard's High Performance Computing Cluster, and for developing the scripts for Geography computation on this data set using GPU based database OmniSci.

A.2 Comparing BERT-based and LIWC-based sentiment scores

To test whether our BERT-based sentiment results are consistent with more traditional sentiment imputation methods, we also compute sentiment on a sample of English-language content using a dictionary-based sentiment imputation method which relies on the LIWC emotion dictionaries. Results are provided in Fig. A-1. We find a strong correlation between the two measures ($\rho = 0.74$, $p < 0.001$).

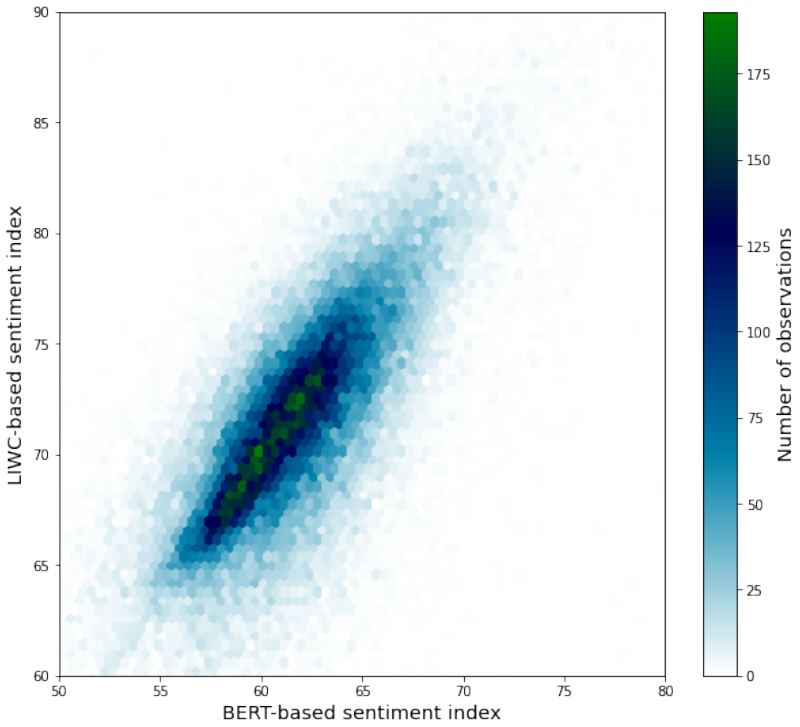

Figure A-1: Correlation between BERT-based and LIWC-based sentiment scores

We find a Pearson correlation coefficient of $\rho = 0.74$ ($p < 0.001$).

A.3 Validating sentiment scores based on happiness surveys: Full results

Table A.1: Social media-based sentiment and survey-based well-being

	Dependent variable:
	Survey-based happiness
Social media-based sentiment	0.37*
Constant	0.00
Weights	None
Observations	50
R^2	0.13
Adjusted R^2	0.12
Note:	*p<0.01; **p<0.005; ***p<0.001

A.4 Validating sentiment scores based on election polling: Full results

Since the state-level accuracy of our social media sentiment score will depend considerable on the number of users who post on the different candidates, we check the distribution of these users by state in Fig. A-2.

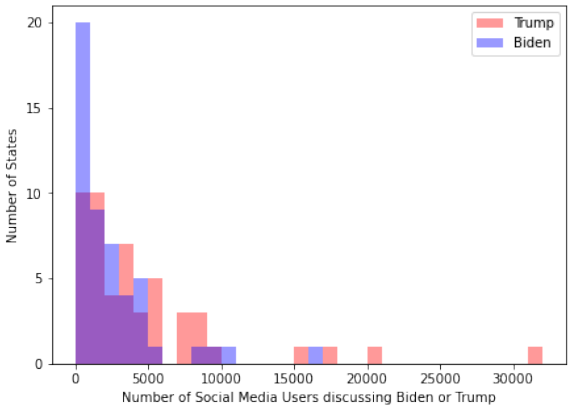

Figure A-2: Number of social media users discussion presidential candidates by state

As a measure of interpretation, we create two alternative social media-based preference metrics to compare with polling outcomes. The first does not rely on NLP, simply comparing the share of Twitter content posted on Biden to the share of content posted on Trump. The second uses our sentiment index, but aggregates the post-level sentiment scores using a simple mean, instead of our two-step aggregation process. Results for the two benchmarks, and for the main social media-based preference index presented in Section 2.3.3, are presented below.

Table A.2: Social media-based sentiment and Polling-based preference

	Dependent variable:		
	Difference in polling numbers		
	(1)	(2)	(3)
Difference in share of posts	−0.23		
Difference in tweet score		0.73^{***}	
Difference in sentiment index			1.01^{***}
Constant	0.41^{**}	0.28^{**}	0.21^{**}
Weights	Nb Users	Nb Users	Nb Users
Observations	50	50	50
R^2	0.02	0.50	0.74
Adjusted R^2	0.004	0.49	0.73
Note:		*p<0.01; **p<0.005; ***p<0.001	

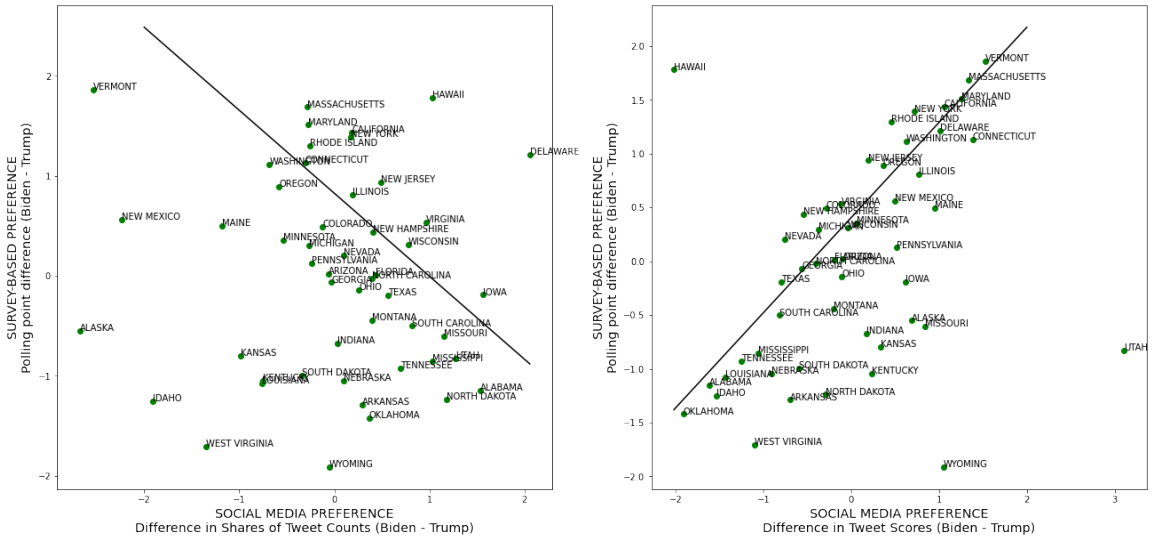

Figure A-3: Social media-based preferences and polling results

Left: comparing social media post counts to polling results. We find a negative, insignificant correlation. Right: comparing (non-weighted) average sentiment scores to polling results. We find a significant, positive correlation—but below the correlation level obtained using the two-step aggregation method.

Appendix B

Supplementary Material to Chapter 3

B.1 Acknowledgements

We gratefully thank Harvard Center for Geographic Analysis (CGA) for providing the Twitter data used in this study [79]. We thank Devika Kakkar and the Harvard CGA team for the support and consultation for scaling the sentiment and geography computation on this big data set using Harvard's High Performance Computing Cluster, and for developing the scripts for Geography computation on this data set using GPU based database OmniSci.

B.2 Data information

B.2.1 Countries of analysis

Only countries for which we have more than 100 daily sentiment-imputed social media posts (from Twitter or Weibo) are included in our analysis. The 157 countries for which that is the case are listed in Table B.1.

Country Code	Country Name	Country Code	Country Name	Country Code	Country Name
AFG	Afghanistan	GGY	Guernsey	OMN	Oman
ALB	Albania	GUY	Guyana	PAK	Pakistan
DZA	Algeria	HND	Honduras	PAN	Panama
AND	Andorra	HKG	Hong Kong	PRY	Paraguay
AGO	Angola	HUN	Hungary	PER	Peru
ARG	Argentina	ISL	Iceland	PHL	Philippines
ARM	Armenia	IND	India	POL	Poland
AUS	Australia	IDN	Indonesia	PRT	Portugal
AUT	Austria	IRN	Iran	PRI	Puerto Rico
AZE	Azerbaijan	IRQ	Iraq	QAT	Qatar
BHR	Bahrain	IRL	Ireland	COG	Republic of Congo
BGD	Bangladesh	IMN	Isle of Man	REU	Reunion
BRB	Barbados	ISR	Israel	ROU	Romania
BLR	Belarus	ITA	Italy	RUS	Russia
BEL	Belgium	JAM	Jamaica	RWA	Rwanda
BLZ	Belize	JPN	Japan	SAU	Saudi Arabia
BEN	Benin	JEY	Jersey	SEN	Senegal
BOL	Bolivia	JOR	Jordan	SRB	Serbia
BIH	Bosnia and Herzegovina	KAZ	Kazakhstan	SLE	Sierra Leone
BWA	Botswana	KEN	Kenya	SGP	Singapore
BRA	Brazil	XKO	Kosovo	SVK	Slovakia
BRN	Brunei	KWT	Kuwait	SVN	Slovenia
BGR	Bulgaria	KGZ	Kyrgyzstan	SOM	Somalia
KHM	Cambodia	LAO	Laos	ZAF	South Africa
CMR	Cameroon	LVA	Latvia	KOR	South Korea
CAN	Canada	LBN	Lebanon	ESP	Spain
CHL	Chile	LSO	Lesotho	LKA	Sri Lanka
CHN	China	LBR	Liberia	SDN	Sudan
COL	Colombia	LBY	Libya	SWZ	Swaziland
CRI	Costa Rica	LTU	Lithuania	SWE	Sweden
CIV	Côte d'Ivoire	LUX	Luxembourg	CHE	Switzerland
HRV	Croatia	MAC	Macao	SYR	Syria
CUB	Cuba	MKD	Macedonia	TWN	Taiwan
CYP	Cyprus	MWI	Malawi	TZA	Tanzania
CZE	Czech Republic	MYS	Malaysia	THA	Thailand
COD	Democratic Republic of the Congo	MLI	Mali	TGO	Togo
DNK	Denmark	MTQ	Martinique	TTO	Trinidad and Tobago
DOM	Dominican Republic	MUS	Mauritius	TUN	Tunisia
ECU	Ecuador	MEX	Mexico	TUR	Turkey
EGY	Egypt	MDA	Moldova	UGA	Uganda
SLV	El Salvador	MNG	Mongolia	UKR	Ukraine
EST	Estonia	MNE	Montenegro	ARE	United Arab Emirates
ETH	Ethiopia	MAR	Morocco	GBR	United Kingdom
FIN	Finland	MOZ	Mozambique	USA	United States
FRA	France	MMR	Myanmar	URY	Uruguay
GUF	French Guiana	NAM	Namibia	UZB	Uzbekistan
GAB	Gabon	NPL	Nepal	VEN	Venezuela
GMB	Gambia	NLD	Netherlands	VNM	Vietnam
GEO	Georgia	NZL	New Zealand	YEM	Yemen
DEU	Germany	NIC	Nicaragua	ZMB	Zambia
GHA	Ghana	NER	Niger	ZWE	Zimbabwe
GRC	Greece	NGA	Nigeria		
GTM	Guatemala	NOR	Norway		

Table B.1: Countries included in Chapter 3 analysis

B.3 Supplementary results

Full result tables, as well as additional robustness checks are presented in this section.

B.3.1 Baseline regression full results

To assess the overall impact of temperatures on sentiment worldwide, we estimate Equation 3.1 on the entire data set. We use the standardized sentiment score as the dependent variable. The full results are presented here in Table B.2. For each temperature bin and control variable, we report the estimated coefficient, the statistical significance of the coefficient (with the superscript stars), and the standard errors (in parentheses). The results are discussed in Section 3.4.1 and visualized in Fig. 3-5.

Table B.2: Max. Temperature and Sentiment

	Dependent variable:
	Standardized Sentiment Score
Temperature ∈(-10,-5]	-0.20^{***}
	(0.01)
Temperature ∈(-5,0]	-0.19^{***}
	(0.01)
Temperature ∈(0,5]	-0.18^{***}
	(0.004)
Temperature ∈(5,10]	-0.10^{***}
	(0.003)
Temperature ∈(10,15]	-0.03^{***}
	(0.002)
Temperature ∈(20,25]	-0.01^{***}
	(0.002)
Temperature ∈(25,30]	-0.05^{***}
	(0.003)
Temperature ∈(30,35]	-0.14^{***}
	(0.003)
Temperature ∈(35,40]	-0.18^{***}
	(0.005)
trange	0.01^{***}
	(0.0004)
Air pollution	-0.001^{***}
	(0.0000)
Wind speed	-0.01^{***}
	(0.0003)
Cloud coverage	-0.001
	(0.003)
Humidity	6.96^{***}
	(0.35)
Precipitation	-604.74^{***}
	(9.06)
Weights	Nb Users
Date FE	Yes
Admin-1 FE	Yes
Observations	850,834
R^2	0.34
Adjusted R^2	0.34
Note:	*p<0.01; **p<0.005; ***p<0.001

B.3.2 Individual-level regression full results

To account for potential composition changes in our sample that might affect our results, we run an additional set of regressions on a sample of frequent social users. Estimates are generated based on Equation 3.1, with additional user fixed effects. Observations are aggregated to individual user-level instead of geography, and are unweighted in the regression model. Results are presented below in Table B.3, and discussed in Section 3.4.3.

Table B.3: Max. Temperature and Sentiment

	Dependent variable:
	score_stdz
Temperature ∈(-10,-5]	-0.03^{**}
Temperature ∈(-5,0]	0.003
Temperature ∈(0,5]	−0.002
Temperature ∈(5,10]	0.003
Temperature ∈(10,15]	0.001
Temperature ∈(20,25]	0.002
Temperature ∈(25,30]	−0.002
Temperature ∈(30,35]	−0.01
Temperature ∈(35,40]	−0.01
trange	0.001^{**}
Air pollution	-0.0001^{*}
Wind speed	−0.001
Cloud coverage	-0.01^{*}
Humidity	0.70
Precipitation	-56.93^{***}
Weights	Nb Users
Date FE	Yes
Admin-1 FE	Yes
Observations	1,785,825
R^2	0.27
Adjusted R^2	0.23
Note:	*p<0.01; **p<0.005; ***p<0.001

B.3.3 Weekday relative to weekend sentiment regression

As a measure of comparison to interpret sentiment damages due to temperature, we also estimate the impact of weekdays and weekends on sentiment. We use a fixed effect time-series regression model:

$$sentiment_{it} = \alpha_0 + \alpha_1 weekday(t) + T_t + \gamma_i + \epsilon_{it} \tag{B.1}$$

where $sentiment_{it}$ is the value of the standardized sentiment index at location i and on date t, and $weekday(t)$ is a dummy variable equal to 1 if t is a weekday (Monday-Friday) and 0 if t is a weekend (Saturday or Sunday). To account for seasonal trends and location-specific differences, we also include month (T_t) and location (γ_i) fixed effects.

Table B.4 report the α_1 coefficient estimating the sentiment change on weekdays relative to weekends. We find a significant drop associated with weekdays equal to 17% of a standard deviation of the sentiment index.

Table B.4: Weekdays (relative to Weekends) and Sentiment

	Dependent variable:
	Standardized Sentiment Score
Weekday	-0.17^{***}
Weights	Nb Users
Month FE	Yes
Admin-1 FE	Yes
Observations	863,052
R^2	0.16
Adjusted R^2	0.16
Note:	*p<0.01; **p<0.005; ***p<0.001

B.3.4 Robustness by weighting scheme

Our main results, presented in Section 3.4.1, are estimating weighting daily, regional observations by the number of social media users in that location on that day. However, social media coverage is not representative of global population, and such a weighting scheme might skew our results by overemphasizing data-rich countries and urban areas. To address this, we test out four weighting schemes: (1) weighting all regions equally regardless of social media activity or population; (2) weighting by social media posts (instead of users); (3) weighting by social media users (default); and (4) weighting by regional population. Apart from the weighting scheme, the four estimations are run with the same model specifications (see Section 3.3 and Equation 3.1). Results are discussed in Section 3.4.2, and full results are provided in Table B.5.

Table B.5: Robustness by Weighting Scheme

	Dependent variable:			
	Standardized Sentiment Score			
	(1)	(2)	(3)	(4)
Temperature ∈(-10,-5]	-0.20^{***}	-0.21^{***}	-0.03^{***}	-0.06^{***}
Temperature ∈(-5,0]	-0.19^{***}	-0.20^{***}	-0.01	-0.06^{***}
Temperature ∈(0,5]	-0.17^{***}	-0.17^{***}	-0.02^{**}	-0.07^{***}
Temperature ∈(5,10]	-0.10^{***}	-0.09^{***}	-0.02^{***}	-0.04^{***}
Temperature ∈(10,15]	-0.03^{***}	-0.02^{***}	0.03^{***}	-0.01
Temperature ∈(20,25]	-0.02^{***}	-0.01^{***}	0.02^{***}	-0.001
Temperature ∈(25,30]	-0.05^{***}	-0.04^{***}	0.01^{**}	-0.01^{*}
Temperature ∈(30,35]	-0.14^{***}	-0.13^{***}	-0.05^{***}	-0.05^{***}
Temperature ∈(35,40]	-0.18^{***}	-0.17^{***}	-0.15^{***}	-0.08^{***}
trange	0.01^{***}	0.01^{***}	0.01^{***}	0.001^{*}
Air pollution	-0.001^{***}	-0.001^{***}	-0.001^{***}	-0.001^{***}
Wind speed	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}
Cloud coverage	-0.001	0.004^{*}	-0.01^{***}	-0.02^{***}
Humidity	7.26^{***}	4.21^{***}	0.62	-1.01
Precipitation	-605.10^{***}	-608.46^{***}	-318.08^{***}	-255.88^{***}
Weights	Nb Users	Nb Posts	Population	None
Date FE	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes
Observations	845,212	845,212	837,846	845,212
R^2	0.34	0.34	0.05	0.03
Adjusted R^2	0.34	0.34	0.05	0.02
Note:				*p<0.1; **p<0.05; ***p<0.01

B.3.5 Robustness by fixed effects

Our main results are also estimated using date and location fixed effects. To assess the robustness of our results, we test out four fixed-effect schemes: (1) only location fixed effects; (2) location and date fixed effects (default); (3) location and year, month, and day-of-week effects; and (4) no fixed effects at all. Apart from the changes in fixed effects, all four estimations are run with the same model specifications (see Section 3.3 and Equation 3.1). Results are presented in Fig. B-1 and Table B.6.

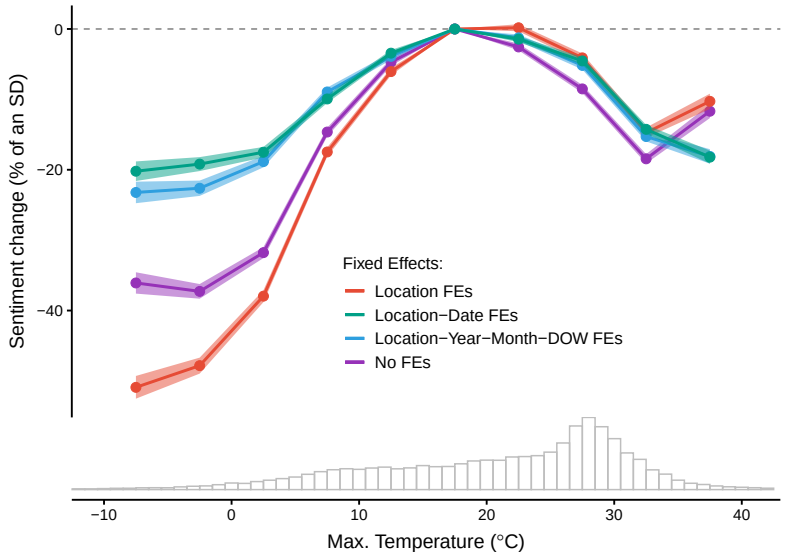

Figure B-1: Robustness by fixed effects

We find that the results are robust by fixed effects. All four estimations yield a similar inverse-U curve, with significant sentiment drops associated to both high and low temperatures relative to the omitted 15°C-20°C bin. The magnitude of the sentiment drop associated with low temperatures is larger for estimations without temporal fixed effects (Location FEs and No FEs): this reflects strong seasonal sentiment trends, and it is consistent with a great amount of existing literature on the negative impact of winter on subjective well-being [56]. We observe similar-magnitude drops associated with high temperatures, regardless of the fixed effect specification chosen.

Table B.6: Robustness by Fixed Effects

	Dependent variable:			
	Standardized Sentiment Score			
	(1)	(2)	(3)	(4)
Temperature ∈(-10,-5]	-0.36^{***}	-0.51^{***}	-0.23^{***}	-0.20^{***}
Temperature ∈(-5,0]	-0.37^{***}	-0.48^{***}	-0.23^{***}	-0.19^{***}
Temperature ∈(0,5]	-0.32^{***}	-0.38^{***}	-0.19^{***}	-0.18^{***}
Temperature ∈(5,10]	-0.15^{***}	-0.17^{***}	-0.09^{***}	-0.10^{***}
Temperature ∈(10,15]	-0.05^{***}	-0.06^{***}	-0.04^{***}	-0.03^{***}
Temperature ∈(20,25]	-0.03^{***}	0.002	-0.01^{***}	-0.01^{***}
Temperature ∈(25,30]	-0.09^{***}	-0.04^{***}	-0.05^{***}	-0.05^{***}
Temperature ∈(30,35]	-0.18^{***}	-0.15^{***}	-0.15^{***}	-0.14^{***}
Temperature ∈(35,40]	-0.12^{***}	-0.10^{***}	-0.18^{***}	-0.18^{***}
trange	0.01^{***}	0.01^{***}	0.01^{***}	0.01^{***}
Air pollution	-0.0002^{***}	-0.001^{***}	-0.001^{***}	-0.001^{***}
Wind speed	-0.01^{***}	-0.02^{***}	-0.01^{***}	-0.01^{***}
Cloud coverage	0.02^{***}	0.01	-0.01^{***}	-0.001
Humidity	1.03^{**}	3.80^{***}	7.47^{***}	6.96^{***}
Precipitation	-627.33^{***}	-591.39^{***}	-606.99^{***}	-604.74^{***}
Weights	Nb Users	Nb Users	Nb Users	Nb Users
Location FE	No	Yes	Yes	Yes
Year-Month-DOW FE	No	No	Yes	No
Date FE	No	No	No	Yes
Observations	850,834	850,834	850,834	850,834
R^2	0.03	0.06	0.19	0.34
Adjusted R^2	0.03	0.05	0.18	0.34
Note:				*p<0.01; **p<0.005; ***p<0.001

B.3.6 Robustness by temperature measure

The MERRA-2 data set from which we collect environmental variables provides three measures of regional hourly temperature: minimum temperature, maximum temperature, and mean temperature [50]. We maintain this distinction when we aggregate the hourly MERRA-2 data to daily level, computing the three temperature measures: (1) the maximum daily temperature is the largest value of all the maximum hourly temperatures; (2) the minimum daily temperature is the smallest value of all the minimum hourly temperatures; and (3) the mean daily temperature is the average of the mean hourly temperatures. As detailed in Section 3.4.1, we use the maximum temperature in our main results, as this is most likely the temperature during the daytime that individuals are likely to interact with. In this section, we test the robustness of our results by changing the temperature variable we use for estimating Equation 3.1. Apart from the temperature variable, the model specifications are the same as for our main results (see Section 3.3. Results are presented in Fig. B-2 and in Table B.7.

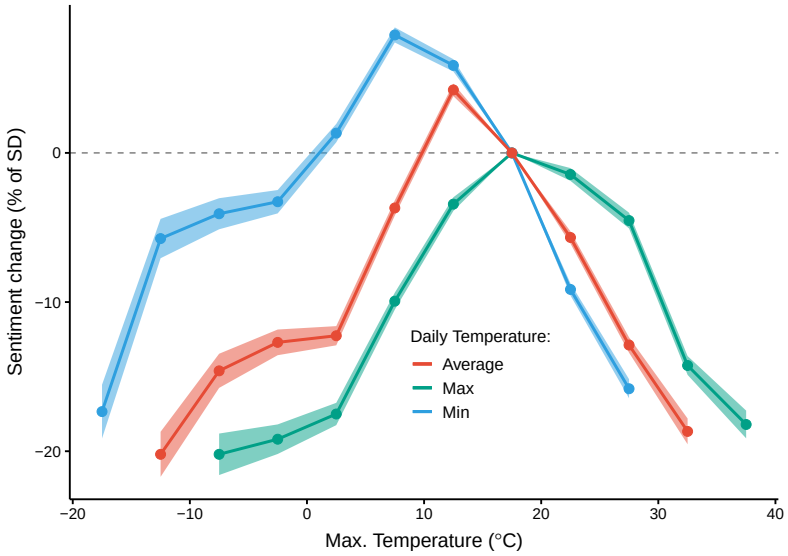

Figure B-2: Robustness by temperature measure

We find robust results by choice of temperature variable. For maximum, minimum, and average temperature, we observe a U-shaped curve of subjective sentiment. While sentiment peaks for a maximum temperature of 15°C–20°C, the most comfortable mean temperature bin is 10°C–15°C and the most comfortable minimum temperature bin is 5°C–10°C. The magnitude of the sentiment drops, associated to both high and

low temperatures, are also similar.

Table B.7: Robustness by measure of temperature

	Dependent variable:		
	Standardized Sentiment Score		
	(1)	(2)	(3)
Temperature ∈(-20,-15]	-0.17^{***}		
Temperature ∈(-15,-10]	-0.06^{***}	-0.20^{***}	
Temperature ∈(-10,-5]	-0.04^{***}	-0.15^{***}	-0.20^{***}
Temperature ∈(-5,0]	-0.03^{***}	-0.13^{***}	-0.19^{***}
Temperature ∈(0,5]	0.01^{***}	-0.12^{***}	-0.18^{***}
Temperature ∈(5,10]	0.08^{***}	-0.04^{***}	-0.10^{***}
Temperature ∈(10,15]	0.06^{***}	0.04^{***}	-0.03^{***}
Temperature ∈(20,25]	-0.09^{***}	-0.06^{***}	-0.01^{***}
Temperature ∈(25,30]	-0.16^{***}	-0.13^{***}	-0.05^{***}
Temperature ∈(30,35]		-0.19^{***}	-0.14^{***}
Temperature ∈(35,40]			-0.18^{***}
trange	0.01^{***}	0.01^{***}	0.01^{***}
Air pollution	-0.001^{***}	-0.001^{***}	-0.001^{***}
Wind speed	-0.01^{***}	-0.01^{***}	-0.01^{***}
Cloud coverage	-0.01^{***}	-0.01^{**}	-0.001
Humidity	14.66^{***}	10.26^{***}	6.96^{***}
Precipitation	-574.33^{***}	-579.37^{***}	-604.74^{***}
Weights	Nb Users	Nb Users	Nb Users
Location FE	Yes	Yes	Yes
Date FE	Yes	Yes	Yes
Observations	853,793	852,561	850,834
R^2	0.34	0.34	0.34
Adjusted R^2	0.33	0.34	0.34
Note:		*p<0.01; **p<0.005; ***p<0.001	

Appendix C

Supplementary Material to Chapter 4

C.1 Acknowledgements

This work was supported by national funds through FCT (Fundação para a Ciência e a Tecnologia) under research grant FCT UIDB/04152/2020 - Centro de Investigação em Gestão de Informação (MagIC), and by MIT Portugal seed grand funding.

This work was also supported by an MIT Portugal seed funding grant.

We gratefully thank Harvard Center for Geographic Analysis (CGA) for providing the Twitter data used in this study [79]. We thank Devika Kakkar and the Harvard CGA team for the support and consultation for scaling the sentiment and geography computation on this big data set using Harvard's High Performance Computing Cluster, and for developing the scripts for Geography computation on this data set using GPU based database OmniSci.

We gratefully thank Confidencial Imobiliário for providing the real estate transaction data used in this study.

C.2 Literature review: Temperature stress and real estate value

Study	Market	WTP for Warm Temperatures	WTP for Cold Temperatures
Roback (1982) [113]	USA	Not provided.	$200 per 1000 heating degree days
Blomquist, Berger, and Hoehn (1988) [14]	USA	$360 per 1000 cooling degree days	$80 per 1000 heating degree days
Cragg and Kahn (1997) [35]	USA	$984 per 1°F decrease in July	$1182 per 1°F increase in February
Maddison and Bigano (2003) [81]	Italy	737,000 Lira per year to avoid an increase in July temperatures	Not given.
Rehdanz and Maddison (2009) [111]	Germany	Up to 985€ per year for a decrease in July temperature	Up to 802€ per year for an increase in January temperature
Sinha and Cropper (2013) [124]	USA	2% of household income to avoid summer temperature increases	Not provided.
Fan, Klaiber, and Fisher-Vanden (2016) [46]	USA	$144 to avoid additional heat day	$91 to avoid additional heat day
Albouy et al. (2016) [4]	USA	1.9% of income per additional cooling day	0.8% of income per additional heating day
Meier and Rehdanz (2017) [85]	Great Britain	Marginal WTP decrease of 478£ by cooling degree day	Marginal WTP decrease of 17£ by heating degree day

Table C.1: Literature Review on the Impact of Temperature Stress on WTP

C.3 Data

C.3.1 Real Estate Data

Descriptive statistics of the real estate data are provided below.

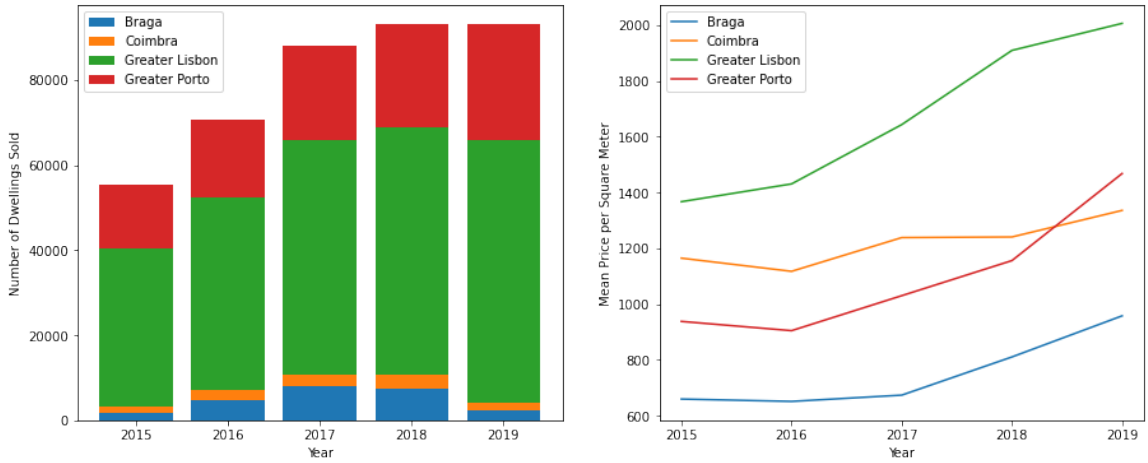

Figure C-1: Descriptive statistics of the real estate data

C.3.2 City-level controls

The first set of city-level controls were collected from the OpenStreetMap portal[1]. These features were extracted from the using the tags key-value selection of Table C.2.

Key	Value	Description
aeroway	terminal	An airport passenger building
amenity	ferry_terminal	Ferry terminal/stop. A place where people/cars/etc. can board and leave a ferry.
leisure	beach_resort	A managed beach, including within the boundary any associated facilities. Entry may also require payment of a fee.
natural	beach	landform along a body of water which consists of sand, shingle or other loose material

Table C.2: Openstreetmap tags selection

We also consider the availability of green areas in the municipalities. For that we use the proportion of green area considering the urban area in municipalities master

[1] https://www.openstreetmap.org

plans. All relevant information is collected from the "Observatório do Ordenamento territorial e urbanismo"[2].

C.4 Supplementary Results

C.4.1 Impact of Weather on Real Estate Value Results, in Standard Deviations

We also run the main results changing the variable transformations to express the result in terms of standard deviations (instead of percentage changes). We standardize the dependant variable, and run the same estimation (see Equation 4.4). Results are provided in Table C.3.

[2] https://www.dgterritorio.gov.pt/

Table C.3: Impact of Temperature Discomfort on Mean Square Meter Price

	Dependent variable:				
	pvm2mean				
	(1)	(2)	(3)	(4)	(5)
Panel A					
Discomfort Index	-0.44^{***}	-0.41^{***}	-0.31^{***}	-0.35^{**}	-0.31^{***}
Air Pollution		-0.05			0.06
Precipitation		0.005			-0.15
Humidity		0.10			-0.02
Logged Green Area			-0.05		-0.002
Logged Higher Ed			0.32^{**}		0.40^{***}
Airport (dummy)			0.03		0.01
Port (dummy)			0.08		0.03
Beach (dummy)			-0.05		-0.07
Logged Population			0.08		-0.20
Median Income			0.22		0.01
Total Dwellings Sold				0.47^{***}	0.43^{**}
Mean Absorp. Time				-0.004	0.07
Panel B					
Summer Discomfort	-0.41^{***}	-0.43^{***}	-0.27^{***}	-0.31^{**}	-0.31^{***}
Winter Discomfort	-0.37^{***}	-0.13	-0.28^{***}	-0.35^{***}	-0.64
Air Pollution		-0.03			0.06
Precipitation		-0.04			-0.17
Humidity		0.21			-0.48
Logged Green Area			-0.06		0.01
Logged Higher Ed			0.31^{**}		0.40^{***}
Airport (dummy)			0.06		0.01
Port (dummy)			0.07		0.03
Beach (dummy)			-0.03		-0.07
Logged Population			0.12		-0.22
Median Income			0.18		-0.003
Total Dwellings Sold				0.48^{***}	0.44^{***}
Mean Absorp. Time				-0.03	0.06
Panel C					
Nb of Hot Days	-0.26^{***}	-0.29^{***}	-0.16^{**}	-0.13	-0.14^{**}
Nb of Cold Days	-0.33^{***}	-0.20^{*}	-0.24^{***}	-0.37^{***}	-0.36^{**}
Air Pollution		0.003			0.04
Precipitation		-0.08			-0.10
Humidity		0.08			-0.17
Logged Green Area			-0.08		-0.01
Logged Higher Ed			0.36^{**}		0.43^{***}
Airport (dummy)			0.03		-0.01
Port (dummy)			0.03		-0.03
Beach (dummy)			-0.03		-0.07
Logged Population			0.17		-0.17
Median Income			0.17		-0.03
Total Dwellings Sold				0.52^{***}	0.47^{***}
Mean Absorp. Time				-0.07	0.04
Weather Controls	No	Yes	No	No	Yes
City Controls	No	No	Yes	No	Yes
Housing Controls	No	No	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	156	156	156	120	120
Note:					*p<0.1; **p<0.05; ***p<0.01

C.4.2 Impact of Temperature on Non-Standardized Sentiment

While our main results are presented using a standardized measure of sentiment, non-standardized sentiment is used below as a robustness check.

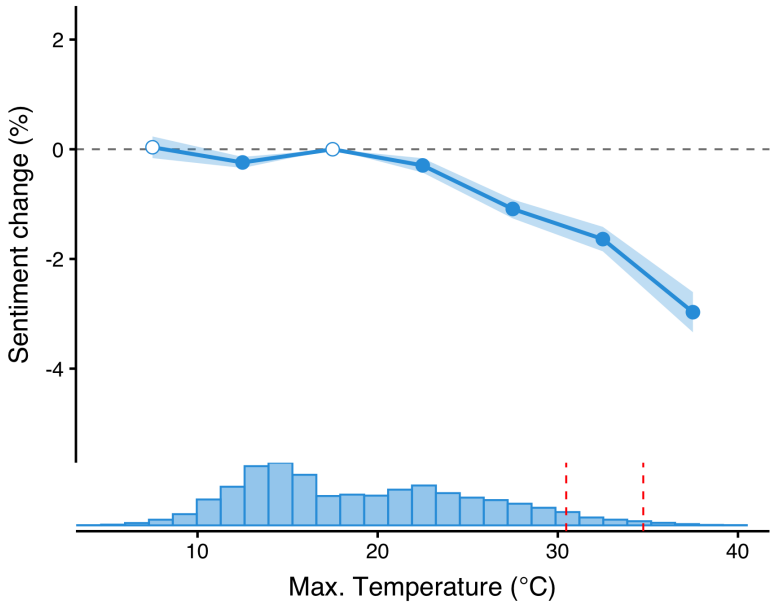

Figure C-2: Impact of Temperature on Non-Standardized Sentiment

C.4.3 Impact of PM2.5 Air Pollution on Sentiment

In order to assess the impact of other weather events, we estimate the impact of air pollution on sentiment in Portugal below.

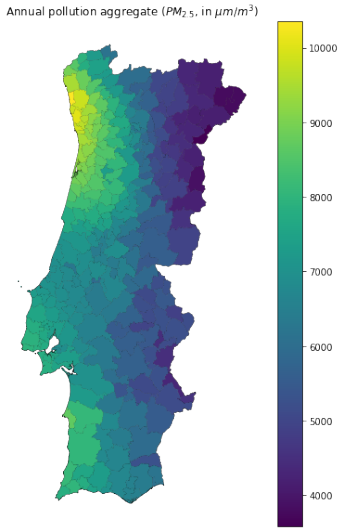

Figure C-3: Annual aggregate air pollution, as measured by $PM_{2.5}$.

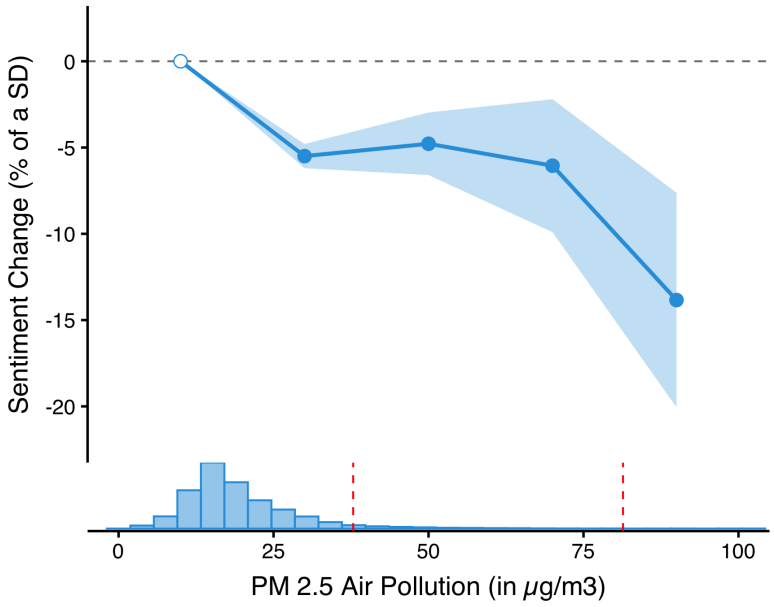

Figure C-4: Impact of air pollution level on sentiment

C.4.4 Full results splitting the sample based on sentiment damage level

Full results for the sample-split regression, visualized in Fig. 4-9 of Section 4.5.3.2, are presented below.

Table C.4: Impact of Temperature Discomfort on Mean Square Meter Price in Low and High Sentiment-Damage Regions

	Dependent variable:			
	Logged Price per Square Meter			
	Low Sentiment-Damage		High Sentiment-Damage	
	(1)	(2)	(3)	(4)
Panel A				
Discomfort Index	-0.12^{***}	-0.04	-0.28^{***}	-0.14^{***}
Air Pollution		0.09		0.01
Precipitation		-0.12		-0.11
Humidity		-0.01		-0.06
Logged Green Area		-0.08^{***}		0.05
Logged Higher Ed		0.26^{**}		0.14
Airport (dummy)				0.03
Port (dummy)		-0.09^{***}		0.05
Beach (dummy)		-0.0003		0.08^{**}
Logged Population		0.18^{***}		-0.03
Median Income		0.04		0.07
Total Dwellings Sold		-0.25^{**}		0.09^{***}
Mean Absorp. Time		-0.03		-0.001
Panel B				
Summer Discomfort	-0.14^{***}	-0.06	-0.26^{**}	-0.15^{***}
Winter Discomfort	-0.19^{***}	-0.03	-0.17^{**}	-0.09
Air Pollution		0.08		0.01
Precipitation		-0.11		-0.13
Humidity		-0.01		-0.08
Logged Green Area		-0.08^{**}		0.05
Logged Higher Ed		0.26^{**}		0.14
Airport (dummy)				0.03
Port (dummy)		-0.08^{**}		0.05
Beach (dummy)		-0.001		0.09^{**}
Logged Population		0.17^{**}		-0.06
Median Income		0.05		0.07
Total Dwellings Sold		-0.22		0.10^{***}
Mean Absorp. Time		-0.03		-0.01
Panel C				
Nb of Hot Days	-0.07^{***}	-0.02	-0.25^{**}	-0.10^{**}
Nb of Cold Days	-0.18^{***}	-0.02	-0.16^{***}	-0.09^{*}
Air Pollution		0.10		0.01
Precipitation		-0.14		-0.10
Humidity		-0.02		-0.07
Logged Green Area		-0.08^{***}		0.03
Logged Higher Ed		0.27^{**}		0.16^{*}
Airport (dummy)				0.04
Port (dummy)		-0.09^{***}		0.03
Beach (dummy)		0.0001		0.09^{**}
Logged Population		0.19^{***}		0.01
Median Income		0.04		0.06
Total Dwellings Sold		-0.26^{*}		0.08^{**}
Mean Absorp. Time		-0.04		0.01
Weather Controls	No	Yes	No	Yes
City Controls	No	Yes	No	Yes
Housing Controls	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	71	54	85	66
Note:				*p<0.1; **p<0.05; ***p<0.01

Bibliography

[1] Social Media Use Continues to Rise in Developing Countries, June 2018.

[2] Scorched: Extreme Heat and Real Estate. Technical report, Urban Land Institute, August 2019.

[3] Charu C. Aggarwal and ChengXiang Zhai. A Survey of Text Classification Algorithms. In Charu C. Aggarwal and ChengXiang Zhai, editors, *Mining Text Data*, pages 163–222. Springer US, Boston, MA, 2012.

[4] David Albouy, Walter Graf, Ryan Kellogg, and Hendrik Wolff. Climate Amenities, Climate Change, and American Quality of Life. *Journal of the Association of Environmental and Resource Economists*, 3(1):205–246, March 2016. Publisher: The University of Chicago Press.

[5] Maximilian Auffhammer. Quantifying Economic Damages from Climate Change. *Journal of Economic Perspectives*, 32(4):33–52, November 2018.

[6] Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37):9216–9221, September 2018. Publisher: National Academy of Sciences Section: Social Sciences.

[7] Markus Baldauf, Lorenzo Garlappi, and Constantine Yannelis. Does Climate Change Affect Real Estate Prices? Only If You Believe In It. *The Review of Financial Studies*, 33(3):1256–1295, March 2020.

[8] Alan Barreca, Karen Clay, Olivier Deschenes, Michael Greenstone, and Joseph S. Shapiro. Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century. *Journal of Political Economy*, 124(1):105–159, February 2016. Publisher: The University of Chicago Press.

[9] Patrick Baylis. Temperature and temperament: Evidence from Twitter. *Journal of Public Economics*, 184:104161, April 2020.

[10] Olfa Belkahla Driss, Sehl Mellouli, and Zeineb Trabelsi. From citizens to government policy-makers: Social media data analysis. *Government Information Quarterly*, 36(3):560–570, July 2019.

[11] Helen L. Berry, Thomas D. Waite, Keith B. G. Dear, Anthony G. Capon, and Virginia Murray. The case for systems thinking about climate change and mental health. *Nature Climate Change*, 8(4):282–290, April 2018. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 4 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Climate-change impacts;Interdisciplinary studies;Scientific community Subject_term_id: climate-change-impacts;interdisciplinary-studies;scientific-community.

[12] John Carlo Bertot, Paul T. Jaeger, and Derek Hansen. The impact of polices on government social media usage: Issues, challenges, and recommendations. *Government Information Quarterly*, 29(1):30–40, January 2012.

[13] Alessandro Bessi and Emilio Ferrara. Social Bots Distort the 2016 US Presidential Election Online Discussion. SSRN Scholarly Paper ID 2982233, Social Science Research Network, Rochester, NY, November 2016.

[14] Glenn C. Blomquist, Mark C. Berger, and John P. Hoehn. New Estimates of Quality of Life in Urban Areas. *The American Economic Review*, 78(1):89–107, 1988. Publisher: American Economic Association.

[15] Abdelhamid Boudjelida, Sehl Mellouli, and Jungwoo Lee. Electronic Citizens Participation: Systematic Review. In *Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance*, ICEGOV '15-16, pages 31–39, Montevideo, Uruguay, March 2016. Association for Computing Machinery.

[16] Andy Brownback and Aaron Novotny. Social desirability bias and polling errors in the 2016 presidential election. *Journal of Behavioral and Experimental Economics*, 74:38–56, June 2018.

[17] Henrico Bertini Brum and Maria das Graças Volpe Nunes. Building a Sentiment Corpus of Tweets in Brazilian Portuguese. *arXiv:1712.08917 [cs]*, December 2017. arXiv: 1712.08917.

[18] U. S. Census Bureau. Housing Vacancies and Homeownership.

[19] Marshall Burke, Felipe González, Patrick Baylis, Sam Heft-Neal, Ceren Baysan, Sanjay Basu, and Solomon Hsiang. Higher temperatures increase suicide rates in the United States and Mexico. *Nature Climate Change*, 8(8):723–729, August 2018. Number: 8 Publisher: Nature Publishing Group.

[20] Marshall Burke, Solomon M. Hsiang, and Edward Miguel. Climate and Conflict. *Annual Review of Economics*, 7(1):577–617, 2015. _eprint: https://doi.org/10.1146/annurev-economics-080614-115430.

[21] Pete Burnap and Matthew L. Williams. Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making. *Policy & Internet*, 7(2):223–242, 2015.

[22] Tamma A. Carleton and Solomon M. Hsiang. Social and economic impacts of climate. *Science*, September 2016. Publisher: American Association for the Advancement of Science.

[23] D. Carvalho, H. Martins, M. Marta-Almeida, A. Rocha, and C. Borrego. Urban resilience to future urban heat waves under a climate change scenario: A case study for Porto urban area (Portugal). *Urban Climate*, 19:1–27, January 2017.

[24] Andrea Ceron. Internet, News, and Political Trust: The Difference between Social Media and Online Media Outlets. *Journal of Computer-Mediated Communication*, 20(5):487–503, September 2015. Publisher: Oxford Academic.

[25] Andrea Ceron, Luigi Curini, Stefano M Iacus, and Giuseppe Porro. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society*, 16(2):340–358, March 2014. Publisher: SAGE Publications.

[26] Andrea Ceron and Fedra Negri. The "Social Side" of Public Policy: Monitoring Online Public Opinion and Its Mobilization During the Policy Cycle. *Policy & Internet*, 8(2):131–147, 2016.

[27] Yuchen Chai. Weibo User Historical Geotagged Posts Dataset. 2021.

[28] Susan Clayton. Climate anxiety: Psychological responses to climate change. *Journal of Anxiety Disorders*, 74:102263, August 2020.

[29] Susan Clayton, Patrick Devine-Wright, Paul C. Stern, Lorraine Whitmarsh, Amanda Carrico, Linda Steg, Janet Swim, and Mirilia Bonnes. Psychological research and global climate change. *Nature Climate Change*, 5(7):640–646, July 2015. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Psychology;Research data Subject_term_id: psychology;research-data.

[30] Stephanie Clifford. Finding Fame With a Prescient Call for Obama. *The New York Times*, November 2008.

[31] Emily M. Cody, Andrew J. Reagan, Lewis Mitchell, Peter Sheridan Dodds, and Christopher M. Danforth. Climate Change Sentiment on Twitter: An Unsolicited Public Opinion Poll. *PLOS ONE*, 10(8):e0136092, August 2015. Publisher: Public Library of Science.

[32] Ellen G. Cohn. The prediction of police calls for service: The influence of weather and temporal variables on rape and domestic violence. *Journal of Environmental Psychology*, 13(1):71–83, March 1993.

[33] Michael Coughlan, Patricia Cronin, and Frances Ryan. Survey research: Process and limitations. *International Journal of Therapy and Rehabilitation*, 16(1):9–15, January 2009. Publisher: Mark Allen Group.

[34] National Research Council, Division of Behavioral and Social Sciences Education, , Committee on National Statistics, and Panel on a Research Agenda for the Future of Social Science Data Collection. *Nonresponse in Social Science Surveys: A Research Agenda*. National Academies Press, October 2013. Google-Books-ID: mg51AgAAQBAJ.

[35] Michael Cragg and Matthew Kahn. New Estimates of Climate Demand: Evidence from Location Choice. *Journal of Urban Economics*, 42(2):261–284, September 1997.

[36] Lucas W. Davis and Paul J. Gertler. Contribution of air conditioning adoption to future energy use under global warming. *Proceedings of the National Academy of Sciences*, 112(19):5962–5967, May 2015. Publisher: National Academy of Sciences Section: Social Sciences.

[37] Olivier Deschênes and Michael Greenstone. Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US. *American Economic Journal: Applied Economics*, 3(4):152–185, October 2011.

[38] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv e-prints*, 1810:arXiv:1810.04805, October 2018.

[39] Delavane Diaz and Frances Moore. Quantifying the economic risks of climate change. *Nature Climate Change*, 7(11):774–782, November 2017. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 11 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Climate-change impacts;Climate-change policy;Economics;Environmental economics Subject_term_id: climate-change-impacts;climate-change-policy;economics;environmental-economics.

[40] Peter Sheridan Dodds, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth. Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter. *PLOS ONE*, 6(12):e26752, December 2011. Publisher: Public Library of Science.

[41] Jiexiong Duan, Weixin Zhai, and Chengqi Cheng. Crowd Detection in Mass Gatherings Based on Social Media Data: A Case Study of the 2014 Shanghai New Year's Eve Stampede. *International Journal of Environmental Research and Public Health*, 17(22):8640, January 2020. Number: 22 Publisher: Multidisciplinary Digital Publishing Institute.

[42] Achim Edelmann, Tom Wolff, Danielle Montagne, and Christopher A. Bail. Computational Social Science and Sociology. *Annual Review of Sociology*, 46(1):61–81, 2020. _eprint: https://doi.org/10.1146/annurev-soc-121919-054621.

[43] Jacob Eisenstein, Brendan O'Connor, Noah A. Smith, and Eric P. Xing. Diffusion of Lexical Change in Social Media. *PLOS ONE*, 9(11):e113114, November 2014. Publisher: Public Library of Science.

[44] "European Environment Agency". The European environment: state and outlook 2015 — Portugal country briefing. Briefing, European Environment Agency, February 2015.

[45] Eurostat. Distribution of population by tenure status, type of household and income group - EU-SILC survey. Technical report, Eurostat, December 2021.

[46] Qin Fan, H. Allen Klaiber, and Karen Fisher-Vanden. Does Extreme Weather Drive Interregional Brain Drain in the U.S.? Evidence from a Sorting Model. *Land Economics*, 92(2):363–388, May 2016. Publisher: University of Wisconsin Press.

[47] Sofia F. Franco and Carlos Daniel Santos. The impact of Airbnb on residential property values and rents: Evidence from Portugal. *Regional Science and Urban Economics*, 88:103667, May 2021.

[48] Bruno S. Frey and Alois Stutzer. The use of happiness research for public policy. *Social Choice and Welfare*, 38(4):659–674, April 2012. Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 4 Publisher: Springer-Verlag.

[49] Shinichiro Fujimori, Tomoko Hasegawa, Volker Krey, Keywan Riahi, Christoph Bertram, Benjamin Leon Bodirsky, Valentina Bosetti, Jessica Callen, Jacques Després, Jonathan Doelman, Laurent Drouet, Johannes Emmerling, Stefan Frank, Oliver Fricko, Petr Havlik, Florian Humpenöder, Jason F. L. Koopman, Hans van Meijl, Yuki Ochi, Alexander Popp, Andreas Schmitz, Kiyoshi Takahashi, and Detlef van Vuuren. A multi-model assessment of food security implications of climate change mitigation. *Nature Sustainability*, 2(5):386–396, May 2019. Number: 5 Publisher: Nature Publishing Group.

[50] Ronald Gelaro, Will McCarty, Max J. Suárez, Ricardo Todling, Andrea Molod, Lawrence Takacs, Cynthia A. Randles, Anton Darmenov, Michael G. Bosilovich, Rolf Reichle, Krzysztof Wargan, Lawrence Coy, Richard Cullather, Clara Draper, Santha Akella, Virginie Buchard, Austin Conaty, Arlindo M. da Silva, Wei Gu, Gi-Kong Kim, Randal Koster, Robert Lucchesi, Dagmar Merkova, Jon Eric Nielsen, Gary Partyka, Steven Pawson, William Putman, Michele Rienecker, Siegfried D. Schubert, Meta Sienkiewicz, and Bin Zhao. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-

2). *Journal of Climate*, 30(14):5419–5454, July 2017. Publisher: American Meteorological Society Section: Journal of Climate.

[51] Edward L. Glaeser, Joshua D. Gottlieb, and Oren Ziv. Unhappy Cities. *Journal of Labor Economics*, 34(S2):S129–S182, April 2016. Publisher: The University of Chicago Press.

[52] Kristina Gligorić, Ashton Anderson, and Robert West. How Constraints Affect Content: The Case of Twitter's Switch from 140 to 280 Characters. *Proceedings of the International AAAI Conference on Web and Social Media*, 12(1), June 2018. Number: 1.

[53] Alec Go, Richa Bhayani, and Lei Huang. Twitter Sentiment Classification using Distant Supervision. page 7, 2009.

[54] Sharad Goel, Ashton Anderson, Jake Hofman, and Duncan J. Watts. The Structural Virality of Online Diffusion. *Management Science*, 62(1):180–196, January 2016. Publisher: INFORMS.

[55] Matthew H. Goldberg, Abel Gustafson, Seth A. Rosenthal, and Anthony Leiserowitz. Shifting Republican views on climate change through targeted advertising. *Nature Climate Change*, pages 1–5, June 2021. Bandiera_abtest: a Cg_type: Nature Research Journals Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Climate change;Communication;Psychology;Psychology and behaviour Subject_term_id: climate-change;communication;psychology;psychology-and-behaviour.

[56] Scott A. Golder and Michael W. Macy. Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures. *Science*, September 2011. Publisher: American Association for the Advancement of Science.

[57] Robert M. Groves, Floyd J. Fowler Jr, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. *Survey Methodology*. John Wiley & Sons, September 2011. Google-Books-ID: ctow8zWdyFgC.

[58] Robert M. Groves and Emilia Peytcheva. The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly*, 72(2):167–189, January 2008.

[59] Nicolas Guetta-Jeanrenaud and Siqi Zheng. Geotweet Archive Sentiment Score. Harvard Dataverse, 2021.

[60] Tomoko Hasegawa, Gen Sakurai, Shinichiro Fujimori, Kiyoshi Takahashi, Yasuaki Hijioka, and Toshihiko Masui. Extreme climate events increase risk of

global food insecurity and adaptation needs. *Nature Food*, 2(8):587–595, August 2021. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 8 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Climate-change impacts;Environmental impact Subject_term_id: climate-change-impacts;environmental-impact.

[61] Bartosz Hawelka, Izabela Sitko, Euro Beinat, Stanislav Sobolevsky, Pavlos Kazakopoulos, and Carlo Ratti. Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41(3):260–271, May 2014. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/15230406.2014.890072.

[62] Brent Hecht and Monica Stephens. A Tale of Cities: Urban Biases in Volunteered Geographic Information. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1):197–205, May 2014. Number: 1.

[63] Harrison Hong, Frank Weikai Li, and Jiangmin Xu. Climate risks and market efficiency. *Journal of Econometrics*, 208(1):265–281, January 2019.

[64] Peter D. Howe, Ezra M. Markowitz, Tien Ming Lee, Chia-Ying Ko, and Anthony Leiserowitz. Global perceptions of local temperature change. *Nature Climate Change*, 3(4):352–356, April 2013. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 4 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Climate change;Psychology Subject_term_id: climate-change;psychology.

[65] Yingjie Hu and Ruo-Qian Wang. Understanding the removal of precise geotagging in tweets. *Nature Human Behaviour*, 4(12):1219–1221, December 2020. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 12 Primary_atype: Comments & Opinion Publisher: Nature Publishing Group Subject_term: Cultural and media studies;Science, technology and society Subject_term_id: cultural-and-media-studies;science-technology-and-society.

[66] Qunying Huang and David W. S. Wong. Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? *International Journal of Geographical Information Science*, 30(9):1873–1898, September 2016. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/13658816.2016.1145225.

[67] IPCC. Global Warming of 1.5 ºC. Technical report, Intergovernmental Panel on Climate Change, 2018.

[68] Kokil Jaidka, Salvatore Giorgi, H. Andrew Schwartz, Margaret L. Kern, Lyle H. Ungar, and Johannes C. Eichstaedt. Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences*, 117(19):10165–10171, May 2020. ISBN: 9781906364113 Publisher: National Academy of Sciences Section: Physical Sciences.

[69] Raja Jurdak, Kun Zhao, Jiajun Liu, Maurice AbouJaoude, Mark Cameron, and David Newth. Understanding Human Mobility from Twitter. *PLOS ONE*, 10(7):e0131469, July 2015. Publisher: Public Library of Science.

[70] Matthew E. Kahn. The Climate Change Adaptation Literature. *Review of Environmental Economics and Policy*, 10(1):166–178, January 2016. Publisher: The University of Chicago Press.

[71] Daniel Kahneman and Alan B. Krueger. Developments in the Measurement of Subjective Well-Being. *Journal of Economic Perspectives*, 20(1):3–24, March 2006.

[72] Markus Kottek, Jürgen Grieser, Christoph Beck, Bruno Rudolf, and Franz Rubel. World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, pages 259–263, July 2006. Publisher: Schweizerbart'sche Verlagsbuchhandlung.

[73] Maximilian Kotz, Leonie Wenz, Annika Stechemesser, Matthias Kalkuhl, and Anders Levermann. Day-to-day temperature variability reduces economic growth. *Nature Climate Change*, 11(4):319–325, April 2021. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 4 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Climate change;Climate-change impacts;Economics;Environmental economics;Environmental impact Subject_term_id: climate-change;climate-change-impacts;economics;environmental-economics;environmental-impact.

[74] Yury Kryvasheyeu, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian. Rapid assessment of disaster damage using social media activity. *Science Advances*, March 2016. Publisher: American Association for the Advancement of Science.

[75] Matti Kummu, Maija Taka, and Joseph H. A. Guillaume. Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015. *Scientific Data*, 5(1):180004, February 2018. Bandiera_abtest: a Cc_license_type: cc_publicdomain Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Developing world;Economics;Environmental social sciences Subject_term_id: developing-world;economics;environmental-social-sciences.

[76] David Lazer, Alex (Sandy) Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. Life in the network: the coming age of computational social science. *Science (New York, N.Y.)*, 323(5915):721–723, February 2009.

[77] Maxime Lenormand, Miguel Picornell, Oliva G. Cantú-Ros, Antònia Tugores, Thomas Louail, Ricardo Herranz, Marc Barthelemy, Enrique Frías-Martínez,

and José J. Ramasco. Cross-Checking Different Sources of Mobility Information. *PLOS ONE*, 9(8):e105184, August 2014. Publisher: Public Library of Science.

[78] Arik Levinson. Valuing public goods using happiness data: The case of air quality. *Journal of Public Economics*, 96(9):869–880, October 2012.

[79] Benjamin Lewis and Kakkar Devika. Harvard CGA Geotweet Archive v2.0. Harvard Dataverse, 2016.

[80] Alejandro Llorente, Manuel Garcia-Herranz, Manuel Cebrian, and Esteban Moro. Social Media Fingerprints of Unemployment. *PLOS ONE*, 10(5):e0128692, May 2015. Publisher: Public Library of Science.

[81] David Maddison and Andrea Bigano. The amenity value of the Italian climate. *Journal of Environmental Economics and Management*, 45(2):319–332, March 2003.

[82] Douglas S. Massey and Roger Tourangeau. Where Do We Go from Here? Nonresponse and Social Measurement. *The ANNALS of the American Academy of Political and Social Science*, 645(1):222–236, January 2013. Publisher: SAGE Publications Inc.

[83] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4):1093–1113, December 2014.

[84] Abid Mehmood and Muhammad Imran. Digital social innovation and civic participation: toward responsible and inclusive transport planning. *European Planning Studies*, 29(10):1870–1885, October 2021. Publisher: Routledge _eprint: https://doi.org/10.1080/09654313.2021.1882946.

[85] Helena Meier and Katrin Rehdanz. The amenity value of the British climate. *Urban Studies*, 54(5):1235–1262, April 2017. Publisher: SAGE Publications Ltd.

[86] Robert Mendelsohn, Ariel Dinar, and Larry Williams. The distributional impact of climate change on rich and poor countries. *Environment and Development Economics*, 11(2):159–178, April 2006. Publisher: Cambridge University Press.

[87] Nicolas Miailhe. A Global Civic Debate on Governing the Rise of Artificial Intelligence. Technical report, The Future Society, 2018.

[88] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. *arXiv:1301.3781 [cs]*, September 2013. arXiv: 1301.3781.

[89] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed Representations of Words and Phrases and their Compositionality. *arXiv:1310.4546 [cs, stat]*, October 2013. arXiv: 1310.4546.

[90] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Rosenquist. Understanding the Demographics of Twitter Users. *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1):554–557, 2011. Number: 1.

[91] Lewis Mitchell, Morgan R. Frank, Kameron Decker Harris, Peter Sheridan Dodds, and Christopher M. Danforth. The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place. *PLOS ONE*, 8(5):e64417, May 2013. Publisher: Public Library of Science.

[92] Ana Monteiro, Vânia Carvalho, Teresa Oliveira, and Carlos Sousa. Excess mortality and morbidity during the July 2006 heat wave in Porto, Portugal. *International Journal of Biometeorology*, 57(1):155–167, January 2013.

[93] Frances C. Moore, Nick Obradovich, Flavio Lehner, and Patrick Baylis. Rapidly declining remarkability of temperature anomalies may obscure public perception of climate change. *Proceedings of the National Academy of Sciences*, 116(11):4905–4910, March 2019. Publisher: National Academy of Sciences Section: Social Sciences.

[94] Igor Mozetič, Miha Grčar, and Jasmina Smailović. Multilingual Twitter Sentiment Classification: The Role of Human Annotators. *PLOS ONE*, 11(5):e0155036, May 2016. Publisher: Public Library of Science.

[95] Gerald C. Nelson, Hugo Valin, Ronald D. Sands, Petr Havlík, Helal Ahammad, Delphine Deryng, Joshua Elliott, Shinichiro Fujimori, Tomoko Hasegawa, Edwina Heyhoe, Page Kyle, Martin Von Lampe, Hermann Lotze-Campen, Daniel Mason d'Croz, Hans van Meijl, Dominique van der Mensbrugghe, Christoph Müller, Alexander Popp, Richard Robertson, Sherman Robinson, Erwin Schmid, Christoph Schmitz, Andrzej Tabeau, and Dirk Willenbockel. Climate change effects on agriculture: Economic responses to biophysical shocks. *Proceedings of the National Academy of Sciences*, 111(9):3274–3279, March 2014. Publisher: National Academy of Sciences Section: Social Sciences.

[96] Fátima Trindade Neves, Miguel de Castro Neto, and Manuela Aparicio. The impacts of open data initiatives on smart cities: A framework for evaluation and monitoring. *Cities*, 106:102860, November 2020.

[97] Brendan O'Connor, Ramnath Balasubramanyan, Bryan Routledge, and Noah Smith. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. *Proceedings of the International AAAI Conference on Web and Social Media*, 4(1), May 2010. Number: 1.

[98] OECD. Innovative Citizen Participation and New Democratic Institutions: Catching the Deliberative Wave. 2020. Publisher: OECD Publishing.

[99] OECD. Housing market - OECD, 2021.

[100] Ariel Ortiz-Bobea, Toby R. Ault, Carlos M. Carrillo, Robert G. Chambers, and David B. Lobell. Anthropogenic climate change has slowed global agricultural productivity growth. *Nature Climate Change*, 11(4):306–312, April 2021. Number: 4 Publisher: Nature Publishing Group.

[101] R. K. Pachauri, M. R. Allen, V. R. Barros, J. Broome, W. Cramer, R. Christ, J. A. Church, L. Clarke, Q. Dahe, P. Dasgupta, N. K. Dubash, O. Edenhofer, I. Elgizouli, C. B. Field, P. Forster, P. Friedlingstein, J. Fuglestvedt, L. Gomez-Echeverri, S. Hallegatte, G. Hegerl, M. Howden, K. Jiang, B. Jimenez Cisneroz, V. Kattsov, H. Lee, K. J. Mach, J. Marotzke, M. D. Mastrandrea, L. Meyer, J. Minx, Y. Mulugetta, K. O'Brien, M. Oppenheimer, J. J. Pereira, R. Pichs-Madruga, G.-K. Plattner, Hans-Otto Pörtner, S. B. Power, B. Preston, N. H. Ravindranath, A. Reisinger, K. Riahi, M. Rusticucci, R. Scholes, K. Seyboth, Y. Sokona, R. Stavins, T. F. Stocker, P. Tschakert, D. van Vuuren, and J.-P. van Ypserle. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva, Switzerland, 2014. Pages: 151 Publication Title: EPIC3Geneva, Switzerland, IPCC, 151 p., pp. 151, ISBN: 978-92-9169-143-2.

[102] Elli Pagourtzi, Vassilis Assimakopoulos, Thomas Hatzichristos, and Nick French. Real estate appraisal: a review of valuation methods. *Journal of Property Investment & Finance*, 21(4):383–401, January 2003. Publisher: MCB UP Ltd.

[103] J. Parente, M. G. Pereira, M. Amraoui, and E. M. Fischer. Heat waves in Portugal: Current regime, changes in future climate and impacts on extreme wildfires. *Science of The Total Environment*, 631-632:534–549, August 2018.

[104] Patrick S. Park, Joshua E. Blumenstock, and Michael W. Macy. The strength of long-range ties in population-scale social networks. *Science*, 362(6421):1410–1413, December 2018. Publisher: American Association for the Advancement of Science Section: Report.

[105] Michael J. Paul and Mark Dredze. Discovering Health Topics in Social Media Using Topic Models. *PLOS ONE*, 9(8):e103408, August 2014. Publisher: Public Library of Science.

[106] James W. Pennebaker, Martha E. Francis, and Roger J. Booth. Linguistic inquiry and word count: LIWC 2001, 2001.

[107] Telmo Pires, Eva Schlinger, and Dan Garrette. How multilingual is Multilingual BERT? *arXiv:1906.01502 [cs]*, June 2019. arXiv: 1906.01502.

[108] Shengsheng Qian, Tianzhu Zhang, Changsheng Xu, and M. Shamim Hossain. Social Event Classification via Boosted Multimodal Supervised Latent Dirichlet Allocation. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 11(2):1–22, January 2015.

[109] Taha H. Rashidi, Alireza Abbasi, Mojtaba Maghrebi, Samiul Hasan, and Travis S. Waller. Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies*, 75:197–211, February 2017.

[110] Katrin Rehdanz and David Maddison. Climate and happiness. *Ecological Economics*, 52(1):111–125, January 2005.

[111] Katrin Rehdanz and David Maddison. The amenity value of climate to households in Germany. *Oxford Economic Papers*, 61(1):150–167, January 2009.

[112] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv:1908.10084 [cs]*, August 2019. arXiv: 1908.10084.

[113] Jennifer Roback. Wages, Rents, and the Quality of Life. *Journal of Political Economy*, 90(6):1257–1278, December 1982. Publisher: The University of Chicago Press.

[114] Elizabeth Rust. How the internet still fails disabled people. *the Guardian*, June 2015.

[115] Paul D. Ryan. H.R.4174 - 115th Congress (2017-2018): Foundations for Evidence-Based Policymaking Act of 2018, January 2019. Archive Location: 2017/2018.

[116] Fadi Salem. The Arab social media report 2017: Social media and the internet of things: Towards data-driven policymaking in the Arab World. Technical Report 7, MBR School of Government, Dubai, 2017.

[117] Matthew J. Salganik. *Bit by Bit: Social Research in the Digital Age*. Princeton University Press, August 2019. Google-Books-ID: 58iXDwAAQBAJ.

[118] Sanam Samadani and Carlos J. Costa. Forecasting real estate prices in Portugal : A data science approach. In *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6, June 2021. ISSN: 2166-0727.

[119] HOWARD SCHUMAN, STANLEY PRESSER, and JACOB LUDWIG. Context Effects on Survey Responses to Questions About Abortion. *Public Opinion Quarterly*, 45(2):216–223, January 1981.

[120] Michele Settanni and Davide Marengo. Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. *Frontiers in Psychology*, 6:1045, 2015.

[121] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kaicheng Yang, Alessandro Flammini, and Filippo Menczer. The spread of low-credibility content by social bots. *Nature Communications*, 9(1):4787, December 2018. arXiv: 1707.07592.

[122] Nate Silver. 2020 Election Forecast, August 2020.

[123] Nicholas P. Simpson, Katharine J. Mach, Andrew Constable, Jeremy Hess, Ryan Hogarth, Mark Howden, Judy Lawrence, Robert J. Lempert, Veruska Muccione, Brendan Mackey, Mark G. New, Brian O'Neill, Friederike Otto, Hans-O. Pörtner, Andy Reisinger, Debra Roberts, Daniela N. Schmidt, Sonia Seneviratne, Steven Strongin, Maarten van Aalst, Edmond Totin, and Christopher H. Trisos. A framework for complex climate change risk assessment. *One Earth*, 4(4):489–501, April 2021.

[124] Paramita Sinha and Maureen L. Cropper. The Value of Climate Amenities: Evidence from US Migration Decisions. Working Paper 18756, National Bureau of Economic Research, February 2013. Series: Working Paper Series.

[125] T. Susca, S. R. Gaffin, and G. R. Dell'Osso. Positive effects of vegetation: Urban heat island and green roofs. *Environmental Pollution*, 159(8):2119–2126, August 2011.

[126] Michael J. Tanana, Christina S. Soma, Patty B. Kuo, Nicolas M. Bertagnolli, Aaron Dembe, Brian T. Pace, Vivek Srikumar, David C. Atkins, and Zac E. Imel. How do you feel? Using natural language processing to automatically rate emotion in psychotherapy. *Behavior Research Methods*, March 2021.

[127] Roger Tourangeau and Ting Yan. Sensitive questions in surveys. *Psychological Bulletin*, 133(5):859–883, 2007. Place: US Publisher: American Psychological Association.

[128] Twitter. Q1 2019 Letter to Shareholders. Technical report, 2019.

[129] UCCRN. The Future We Don't Want. Technical report, Urban Climate Change Research Network, February 2018.

[130] {US Census Bureau}. Urban Areas Facts. Section: Government.

[131] Cristian Vaccari, Augusto Valeriani, Pablo Barberá, Richard Bonneau, John T. Jost, Jonathan Nagler, and Joshua Tucker. Social media and political communication: A survey of Twitter users during the 2013 Italian general election. *Rivista italiana di scienza politica*, (3/2013), 2013.

[132] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. *arXiv:1706.03762 [cs]*, December 2017. arXiv: 1706.03762.

[133] A. M. Vicedo-Cabrera, N. Scovronick, F. Sera, D. Royé, R. Schneider, A. Tobias, C. Astrom, Y. Guo, Y. Honda, D. M. Hondula, R. Abrutzky, S. Tong, M. de Sousa Zanotti Stagliorio Coelho, P. H. Nascimento Saldiva, E. Lavigne, P. Matus Correa, N. Valdes Ortega, H. Kan, S. Osorio, J. Kyselý, A. Urban, H. Orru, E. Indermitte, J. J. K. Jaakkola, N. Ryti, M. Pascal, A. Schneider,

K. Katsouyanni, E. Samoli, F. Mayvaneh, A. Entezari, P. Goodman, A. Zeka, P. Michelozzi, F. de'Donato, M. Hashizume, B. Alahmad, M. Hurtado Diaz, C. De La Cruz Valencia, A. Overcenco, D. Houthuijs, C. Ameling, S. Rao, F. Di Ruscio, G. Carrasco-Escobar, X. Seposo, S. Silva, J. Madureira, I. H. Holobaca, S. Fratianni, F. Acquaotta, H. Kim, W. Lee, C. Iniguez, B. Forsberg, M. S. Ragettli, Y. L. L. Guo, B. Y. Chen, S. Li, B. Armstrong, A. Aleman, A. Zanobetti, J. Schwartz, T. N. Dang, D. V. Dung, N. Gillett, A. Haines, M. Mengel, V. Huber, and A. Gasparrini. The burden of heat-related mortality attributable to recent human-induced climate change. *Nature Climate Change*, 11(6):492–500, June 2021. Number: 6 Publisher: Nature Publishing Group.

[134] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, March 2018. Publisher: American Association for the Advancement of Science Section: Report.

[135] Jianghao Wang, Yichun Fan, Juan Palacios, Yuchen Chai, Nicolas Guetta-Jeanrenaud, Nick Obradovich, Zhou Chenghu, and Siqi Zheng. Global evidence of expressed sentiment alterations during the COVID-19 pandemic. *Nature Human Behaviour*, 2022.

[136] Jianghao Wang, Nick Obradovich, and Siqi Zheng. A 43-Million-Person Investigation into Weather and Expressed Sentiment in a Changing Climate. *One Earth*, 2(6):568–577, June 2020.

[137] Nick Watts, W. Neil Adger, Paolo Agnolucci, Jason Blackstock, Peter Byass, Wenjia Cai, Sarah Chaytor, Tim Colbourn, Mat Collins, Adam Cooper, Peter M. Cox, Joanna Depledge, Paul Drummond, Paul Ekins, Victor Galaz, Delia Grace, Hilary Graham, Michael Grubb, Andy Haines, Ian Hamilton, Alasdair Hunter, Xujia Jiang, Moxuan Li, Ilan Kelman, Lu Liang, Melissa Lott, Robert Lowe, Yong Luo, Georgina Mace, Mark Maslin, Maria Nilsson, Tadj Oreszczyn, Steve Pye, Tara Quinn, My Svensdotter, Sergey Venevsky, Koko Warner, Bing Xu, Jun Yang, Yongyuan Yin, Chaoqing Yu, Qiang Zhang, Peng Gong, Hugh Montgomery, and Anthony Costello. Health and climate change: policy responses to protect public health. *The Lancet*, 386(10006):1861–1914, November 2015. Publisher: Elsevier.

[138] Weibo. Weibo Reports Fourth Quarter and Fiscal Year 2018 Unaudited Financial Results, March 2019.

[139] Heinz Welsch. Environment and happiness: Valuation of air pollution using life satisfaction data. *Ecological Economics*, 58(4):801–813, July 2006.

[140] Stefan Wojcik and Adam Hughes. Sizing up Twitter users. Technical report, Pew Research Center, April 2019.

[141] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz,

Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. HuggingFace's Transformers: State-of-the-art Natural Language Processing. *arXiv:1910.03771 [cs]*, July 2020. arXiv: 1910.03771.

[142] World Bank Data Team. Urban population - European Union, 2020.

[143] World Bank Data Team. Population, 2021.

[144] Siqi Zheng, Yuming Fu, and Hongyu Liu. Demand for Urban Quality of Living in China: Evolution in Compensating Land-Rent and Wage-Rate Differentials. *The Journal of Real Estate Finance and Economics*, 38(3):194–213, April 2009.

[145] Siqi Zheng, Matthew E. Kahn, and Hongyu Liu. Towards a system of open cities in China: Home prices, FDI flows and air quality in 35 major cities. *Regional Science and Urban Economics*, 40(1):1–10, January 2010.

[146] Siqi Zheng, Jianghao Wang, Cong Sun, Xiaonan Zhang, and Matthew E. Kahn. Air pollution lowers Chinese urbanites' expressed happiness on social media. *Nature Human Behaviour*, 3(3):237–243, March 2019.