

Are the Poor Are More Vulnerable to Climate Impacts on Mental Health?

Matthew Cooper

September 21, 2020

We explored the impact of temperature and precipitation on expressed sentiment on twitter. This is an initial overview.

This work is heavily influenced by previous research finding a link between weather and twitter sentiment, especially work by Patrick Baylis (See [here](#), [here](#)). However, this work builds on that work in several ways:

- We more precisely measure temperature
 - We take into account the Heat Index, which is more indicative of temperature stress than raw temperature
 - We infer the hourly temperature, rather than using the daily maximum temperature
 - We get more local temperature, at each PRISM grid cell, rather than at CBSA
- We run tweet-level models without aggregating the data (although this may get harder as we better account for fixed effects)
- We examine factors that moderate the effect of weather, to explore heterogeneities in vulnerability.

All of this work was done in the cloud and this would not have been possible without a generous grant from Microsoft's [AI for Earth](#) Program.

Data

Twitter Data

We use about 240 million tweets from the University of Vermont. This is every geolocated tweet that they have from 2009-2019 in the lower 48. This only includes tweets where the location is specifically supplied by the user, and does not include tweets where location is inferred from the user profile. Because twitter has made it easier for users to opt-out of geolocation in recent years, the data is mostly from the early-mid 2010s.

For each tweet, we calculated the sentiment in the text of the tweet using the following methods:

- `afinn`
- `textblob`
- `hedonometer`
- `vader`
- `sen`
- `wkwsci`

We also tagged tweets as including weather terms or not, so those tweets could be excluded.

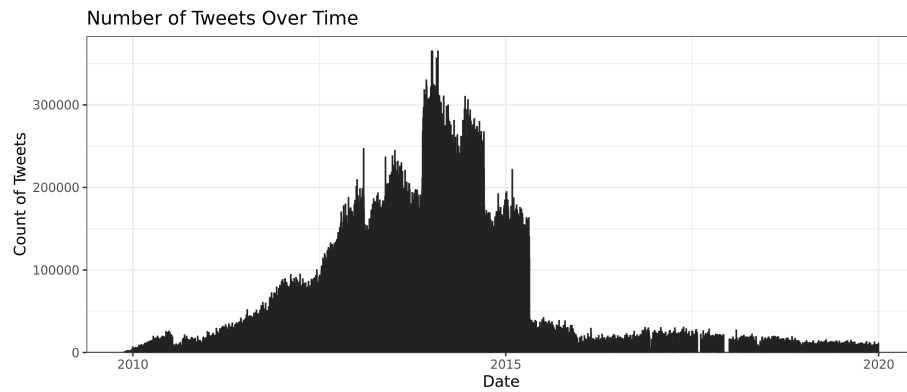


Figure 1: Time

Climate Data

We use climate data from [PRISM](#), including precipitation and daily highs and lows of temperature and vapor pressure deficit. We extract data for the day of each tweet from PRISM, taking into account the different time zones between tweet location and UTC, using Google Earth Engine.

We then calculate relative humidity from temperature and vapor pressure deficit using an equation from Dennis Hartman “Global Physical Climatology” (p 350), which I got from [this](#) Stack Exchange answer. Using relative humidity, we can get the Heat Index, using `weathermetrics::heat.index()` in R, which is based on the National Weather Service official equations for the Heat Index.

Then, we use the hour of the tweet and estimated hours of the daily minimum and maximum temperature to get the heat index temperature at the time of the tweet, based on [Linville 1990](#). Basically, we assume the minimum temperature occurs at sunrise and increases sinusoidally, peaking two hours after solar noon, and then falls logarithmically after sunset. In some cases the minimum daily temperature can occur after sunset and there’s really no way we can correct for

this, but I think this is a good way to proxy temperature at the time of the tweet and makes more sense than only using the daily maximum temperature.

Local income per capita

Finally, we use year census data at the census block group scale to estimate income per capita, which each year's data converted to 2019 dollars. The available data ends in 2018, so we use 2018 data for 2019.

Initial Results

These initial results are run for simple models where we estimate a relationship between the hedonometer score and maximum temperature (for > 20 degrees C), minimum temperature (for < 20 degrees C) and precipitation (for all tweets). We estimate six different slopes and intercepts for each income quantile.

So far, we have only looked at hedonometer scores.

Across every weather metric, there is a clear relationship between income and sentiment. The pattern with intercepts is much clearer than the pattern with slopes.

Heat

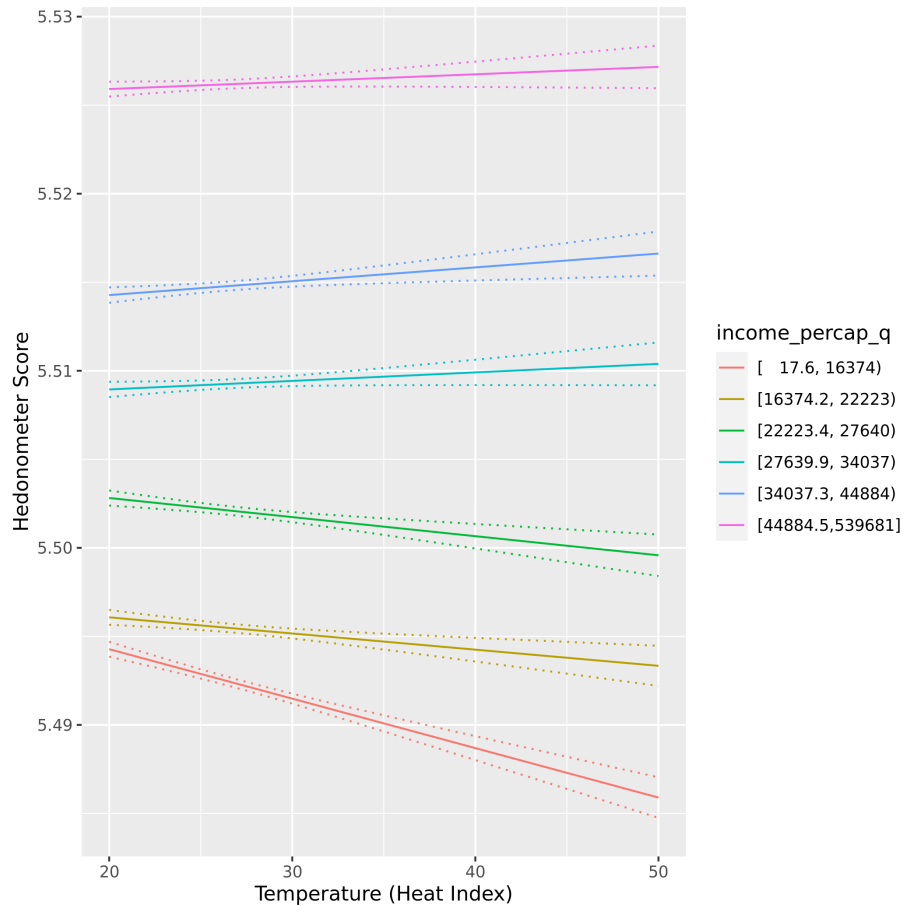


Figure 2: Heat Index Temperature and Sentiment

It looks like heat only matters in the bottom half of the income distribution, and wealthy people might even enjoy heat?

Cold

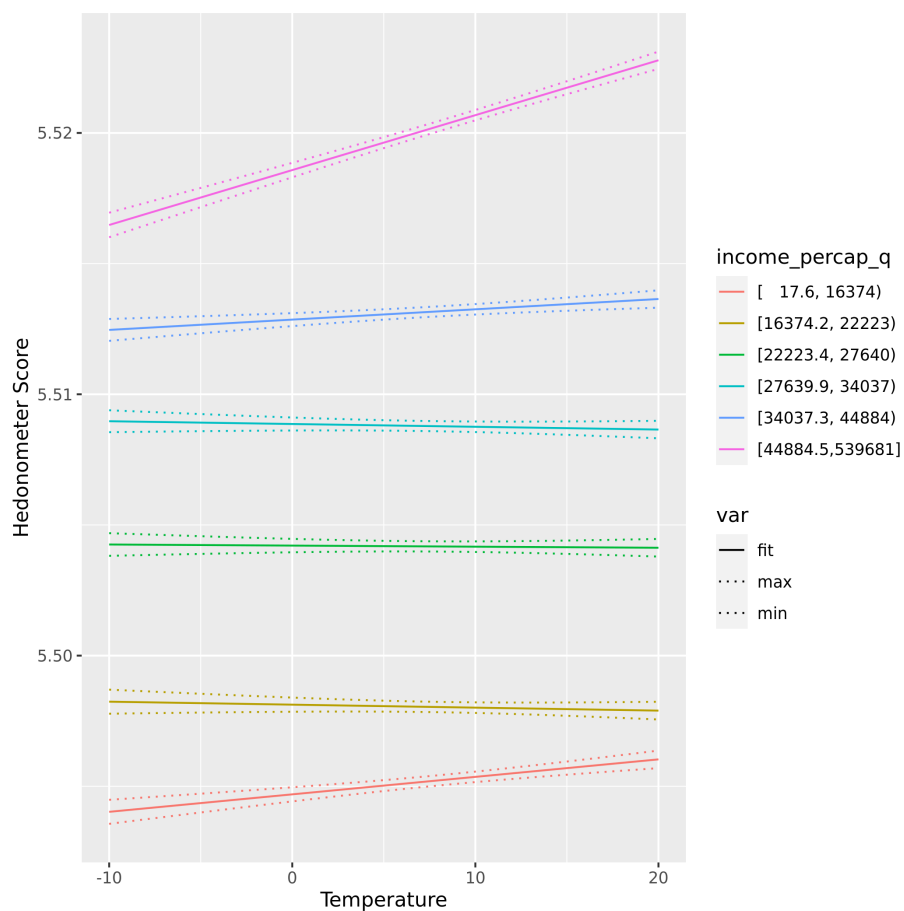


Figure 3: Cold Temperatures and Sentiment

Kinda bizarre, cold only affects the richest and the poorest? Maybe there is some endogeneity with income, where the top quantile is mostly found in colder places, like the northeast.

Precipitation

We all hate rain and it seems to affect us all equally, regardless of income bracket. Hmmm.

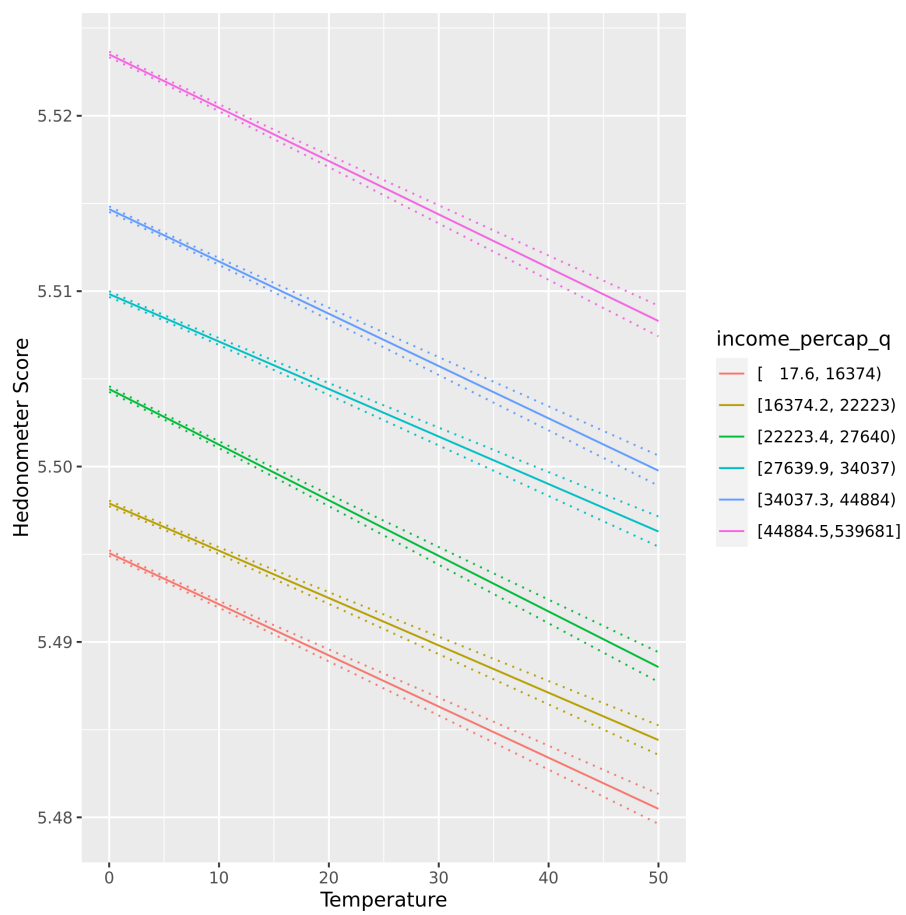


Figure 4: Precipitation and Sentiment

Challenges and Next Steps

We need to control for fixed effects. That might explain why we are getting interesting-but-puzzling temperature effects: there is a lot more spatial endogeneity with temperature.

- Maybe use spatial & temporal splines?
- Maybe use Stochastic Gradient Descent instead of Maximum Likelihood? This wouldn't give p-values or standard errors, though.

Look at other factors influencing vulnerability, including:

- Race, other census factors?
- Green Spaces? (worried about endogeneity with temperature)
- Infrastructure/permeable surfaces? (can we tell a concrete jungle from a hyper-developed, hyper-rich urban center? Does Manhattan look different than Harlem to a satellite?)