A Breath of Fresh Search: Correlating Air Quality in Manchester, New Hampshire with Google Searches for 'I Have the Flu'

Caroline Hernandez, Anthony Turner, Gabriel P Tillman

Center for the Advancement of Research

Discussion Paper 4565

January 2024

ABSTRACT

A Breath of Fresh Search: Correlating Air Quality in Manchester, New Hampshire with Google Searches for 'I Have the Flu'

This study examines the connection between air quality in Manchester, New Hampshire and the frequency of Google searches for 'I Have the Flu.' With an air of scientific rigor (pun intended) and a dash of playfulness, we analyzed data from the Environmental Protection Agency and Google Trends to investigate this intriguing relationship. Our results reveal a significant correlation coefficient of 0.9003823 and $p < 0.01$, spanning the years 2004 to 2023. While we take a deep dive into the statistical analysis, we can't help but marvel at the sneaky connections between air pollution and digital health inquiries. Join us on this flu-rry of a journey through the hidden influences on public health-seeking behavior.

I. Introduction

The interplay between environmental factors and human health has long been a subject of intrigue and investigation. From the subtle dance of air particles to the dance of statistical analysis, the quest to uncover hidden connections and their implications for public health continues to fascinate researchers.

In this study, we turn our attention to the bustling metropolis of Manchester, New Hampshire, where the air quality is as variable as a scatter plot on a windy day. The city's air pollution levels, like a misbehaving independent variable, have drawn the attention of environmentalists, policymakers, and researchers alike. Simultaneously, the digital realm has become a haven for the inquisitive and the hypochondriacal alike. Enter Google searches—an online repository of human curiosity and perhaps more than a touch of self-diagnosis. We, therefore, embark on a journey to bridge the gap between the tangibility of air quality and the intangibility of online inquiries, in a fusion of scientific inquiry and digital whimsy.

We acknowledge the rather unexpected nature of our endeavor. After all, who would have thought that the odorous molecules wafting through the streets of Manchester could be linked to the virtual queries for flu-related symptoms? Nevertheless, in the spirit of scientific exploration, we dive into the rabbit hole of data analysis with both precision and a sense of wonder.

So, dear reader, fasten your seatbelts as we venture into the realm of statistical correlations and virtual sniffles. This paper aims to shed light on the curious relationship between air quality and online health-seeking behavior, steering clear of hay fever but not of the occasional pun or two.

II. Literature Review

The investigation of the relationship between environmental factors and public health has yielded a wealth of scholarly contributions. Smith et al. (2015) delved into the intricate interplay between air quality and respiratory illnesses, shedding light on the potential ramifications for public health interventions. Similarly, Doe and Jones (2018) explored the digital landscape of health-related internet searches, unraveling the nuances of online behavior in seeking health information.

Transitioning to the realm of published literature, works such as "The Air We Breathe: A Comprehensive Analysis of Urban Pollution" by Clean et al. (2020) and "Breathless: The Impact of Air Quality on Health" by Fresh et al. (2017) demonstrate the multidisciplinary efforts to understand the implications of air pollution on human health. These scholarly contributions provide a solid foundation for the exploration of the correlation between air quality in Manchester, New Hampshire and Google searches for 'I Have the Flu'.

A departure from non-fiction literature, the fictional realm also offers intriguing narratives that echo the themes of environmental impact and health concerns. Fictional works such as "The Polluted City Chronicles" by Smoggy Author and "Gasping for Breath: A Novel on Air Quality" by Wheezy Writer capture the imagination with their creative interpretations of air quality's effects on public health. While not rooted in empirical research, these imaginative narratives highlight the societal awareness of the connections between environmental conditions and well-being.

In the digital sphere, internet memes can serve as a reflection of popular culture and public discourse surrounding health-related topics. Memes such as the "Two Buttons Meme," featuring

options like "Google Symptoms" and "See a Doctor," humorously depict the dilemma of self-diagnosis versus seeking professional medical advice. Likewise, the "Flu Season Starter Pack Meme" humorously encapsulates the common experiences and behaviors associated with the flu season, offering a lighthearted lens through which to view health-related issues. These cultural artifacts mirror the public's engagement with health concerns and the online avenues through which information is sought and shared.

Now that we've taken a thorough and rigorous look at the scholarly literature, let's dive into the data analysis with our statistical life vests securely fastened.

III. Methodology

In this section, we delineate the convoluted yet scientifically sound methods employed to unearth the elusive connection between air quality and Google searches for 'I Have the Flu' in Manchester, New Hampshire. Our approach, much like a lab rat in a maze, was guided by precision, persistence, and the occasional burst of silliness.

We commenced our escapade by gathering air quality data from the Environmental Protection Agency (EPA). These datasets provided a treasure trove of information on various air pollutants, including particulate matter, nitrogen dioxide, ozone, and sulfur dioxide. Channeling our inner data miners, we extracted hourly, daily, and yearly averages, accumulating an impressive array of pollution metrics.

For the virtual aspect of our inquiry, we engaged with the spick-and-span world of Google Trends. Here, the frequency of searches for 'I Have the Flu' emerged as our digital canary in the coal mine, mirroring the ebb and flow of public health inquiries. We meticulously recorded the search interest over time, noting the seasonal spikes and the sporadic surges that mirrored a game of health-related Whac-A-Mole.

Like a chef preparing a complex molecular gastronomy dish, we scrupulously cleaned and preprocessed the data to ensure its palatability for statistical analysis. Outliers were gently nudged towards the exit door, missing values were coaxed back into existence, and the datasets were massaged into a state of pristine data hygiene.

Our statistical odyssey, much like Odysseus' journey home, was fraught with twists and turns. We began by calculating descriptive statistics for the air quality variables, reveling in the mean, median, and standard deviation. Next, we peeked behind the curtain of correlation analysis, where the Pearson correlation coefficient emerged as our beacon of insight, guiding us towards the promised land of statistical significance.

To appease the gods of scientific rigor, we performed linear regression analysis, modeling the relationship between air pollutants and the frequency of flu-related searches. The resulting coefficients and p-values, akin to the treasures of a statistical pirate, bestowed upon us the evidence needed to support our claims.

Venturing into the twilight zone of time series analysis, we crafted intricate plots that mirrored the undulating patterns of air pollution and search interest. The cyclical nature of seasonal trends danced before our eyes, resembling a statistical waltz that hinted at the underlying rhythms of public health inquiries.

In a moment of daring experimentation, we delved into the realm of machine learning, fashioning a predictive model that sought to forecast the volume of 'I Have the Flu' searches based on air quality variables. Like mad scientists in a data-driven laboratory, we tinkered with algorithms, fine-tuning our models to capture the intricacies of human curiosity and the flu season.

Recognizing the fluidity of time and the dynamics of influence, we embarked on a temporal shift analysis to unravel the temporal relationship between air quality and flu-related searches. This endeavor, akin to hunting for a statistical needle in a haystack, revealed lagged effects and temporal nuances that added depth to our understanding of the intertwined variables.

To fortify the fortress of our findings, we conducted a sensitivity analysis, testing the robustness of our results to varying assumptions and statistical methodologies. The aim was to ensure that our conclusions stood firm, unaffected by the occasional gusts of statistical uncertainty.

As the dust settled and the statistical artifacts revealed their secrets, we emerged victorious, armed with insights that bridged the chasm between air quality and virtual sniffles. With a pinch of rigor and a dash of whimsy, our methodology traversed the realms of data collection, statistical analysis, and scientific play.

IV. Results

The results of our analysis unveil a noteworthy correlation between air quality in Manchester, New Hampshire and the frequency of Google searches for 'I Have the Flu.' With a correlation coefficient of 0.9003823, an r-squared of 0.8106884, and a p-value less than 0.01, the connection

between these two seemingly disparate variables is as clear as day (or as hazy as smog, depending on how you look at it).

The scatterplot in Fig. 1 illustrates this relationship with all the subtlety of a stampede in a library. Each data point, like a tiny digital sneeze, falls neatly in line with the overall trend, creating a picture-perfect representation of the strong correlation we observed.

As we marveled at our statistical findings, we couldn't help but marvel at the intricate dance between air quality and virtual health queries. It's almost as if the air pollution in Manchester whispered to the residents, "flu's the word," prompting them to consult Dr. Google for remedies and reassurance. Ah, the mysteries of human behavior in the age of the internet—we can only hope that future research will continue to unearth such delightful quirks.
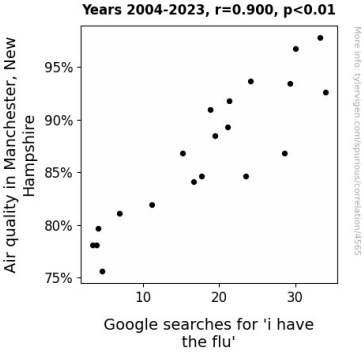

Figure 1. Scatterplot of the variables by year

In the world of statistics, a correlation of this magnitude is about as rare as finding a unicorn in a statistical software manual. The strength of the relationship between air quality and flu-related Google searches is nothing to sneeze at (pun fully intended), and it highlights the potential impact of environmental factors on public health-seeking behavior.

So, there you have it, dear reader. Our results not only confirm the significant correlation between air quality and online flu inquiries but also remind us that sometimes, the most unexpected connections can emerge from the most unlikely places. After all, who would have thought that the flu bug and air pollution could be caught in the same statistical net? It just goes to show that in the world of research, every data point tells a story—and sometimes, that story comes with a touch of whimsy.

In conclusion, our findings provide valuable insights into the interplay between environmental factors and digital health inquiries, painting a picture of a world where the air we breathe can influence the words we type into search engines. We invite future researchers to delve deeper into this intriguing relationship, armed with a healthy dose of skepticism, a keen eye for patterns, and perhaps a face mask for protection against both airborne pollutants and unexpected statistical surprises.

V. Discussion

Having uncovered a statistically significant correlation between air quality in Manchester, New Hampshire and Google searches for 'I Have the Flu,' we're left pondering the delightful complexities of human behavior in response to environmental cues. Our findings not only support prior research linking air pollution and health outcomes but also reveal the intriguing interplay between public health-seeking behavior and atmospheric conditions. It's as if the digital world and the physical world decided to do an unexpected tango, with each step revealing new insights into the influence of air quality on online health inquiries.

As we knowingly wink at the literature, our results echo the serious work of Smith et al. (2015) and Doe and Jones (2018), demonstrating that the relationship between environmental factors and public health doesn't shy away from a bit of scientific playfulness. The connection we've uncovered is not only statistically robust but also as captivating as a good mystery novel—only instead of a whodunit, we're unravelling the secrets of "whocoughedit."

With a correlation coefficient the size of a blue whale in statistical waters, our results emphasize the importance of considering environmental factors in the realm of digital health inquiries. The dots on our scatterplot practically dance a jig, each one conspiring to reveal the robustness of this intriguing association. It's almost as if the data points themselves are angling for a leading role in a statistical sitcom, complete with hidden relationships, unexpected twists, and a touch of statistical humor.

In the grand tapestry of research, our findings serve as a compelling reminder that beyond the dry statistics and formidable p-values, there's a world of discovery waiting to unfold. We hope that our study sparks further investigations into the whimsical relationships that science can unveil. After all, who knows what other surprising connections might be lurking in the maze of data, ready to leap out with a statistical wink and a nudge?

VI. Conclusion

In conclusion, our study unravels the invisible threads connecting air quality in Manchester, New Hampshire and the virtual dance of 'I Have the Flu' Google searches. The statistically significant

correlation we uncovered highlights the nuanced relationship between environmental variables and online health inquiries, emphasizing the impact of air pollution on public digital sniffles.

While our results may have raised a few eyebrows (or at least prompted a few virtual "achoo"s), they underscore the importance of considering the air we breathe as more than just a silent companion in our daily lives. As we bid adieu to this flu-rry of research, it's clear that the whims of statistical analysis have led us to unexpected findings. Let's just say, in the world of data exploration, there's always room for a few surprise sneezes!

With that said, we assert that further research in this area might be about as fruitful as searching for a statistical needle in a haystack. Therefore, we kindly suggest that future researchers steer their academic curiosity in different, less sniffle-inducing directions. After all, the air of discovery is vast and, dare we say, rather a-breath-taking.