CPIE: A Spatiotemporal Visual Analytic Tool to Explore the Impact of Coal Pollution

Sichen Jin*
School of Interactive Computing
Georgia Institute of Technology

Lucas Henneman †
Sid and Reba Dewberry Department of Civil, Environmental, and Infrastructure Engineering
College of Engineering and Computing
George Mason University

Jessica Roberts ‡
School of Interactive Computing
Georgia Institute of Technology

ABSTRACT

This paper introduces CPIE (Coal Pollution Impact Explorer), a spatiotemporal visual analytic tool developed for interactive visualization of coal pollution impacts. CPIE visualizes electricity-generating units (EGUs) and their contributions to statewide Medicare deaths related to coal $PM_{2.5}$ emissions. The tool is designed to make scientific findings on the impacts of coal pollution more accessible to the general public and to raise awareness of the associated health risks. We present three use cases for CPIE: 1) the overall spatial distribution of all 480 facilities in the United States, their statewide impact on excess deaths, and the overall decreasing trend in deaths associated with coal pollution from 1999 to 2020; 2) the influence of pollution transport, where most deaths associated with the facilities located within the same state and neighboring states but some deaths occur far away; and 3) the effectiveness of intervention regulations, such as installing emissions control devices and shutting down coal facilities, in significantly reducing the number of deaths associated with coal pollution.

Index Terms: Coal Pollution, Spatiotemporal Visualization, Public Health.

1 INTRODUCTION

Coal remains a primary energy source globally, providing essential electricity but at a significant environmental cost. Coal pollution is a major contributor to poor air quality [4, 21], with exposure to pollutants like $PM_{2.5}$ linked to adverse health effects and increased mortality [7]. This pressing issue necessitates ongoing efforts to understand and mitigate the impacts of coal pollution. Previous work has developed complex computational models to estimate the number of deaths associated with exposure to coal $PM_{2.5}$ from electricity-generating units (EGUs) [8, 12, 13, 5, 2]. However, an interactive visualization tool that can effectively analyze and communicate the multidimensional spatiotemporal impacts of coal pollution on public health is still missing. Such tools are crucial for policymakers, researchers, and the public to better understand pollution patterns, assess the effectiveness of regulatory measures, and identify areas requiring intervention.

In this paper, we present CPIE (Coal Pollution Impact Explorer), a spatiotemporal visual analytic tool designed to explore the impacts of coal pollution on mortality. This work is part of a scientific study on mortality risks from U.S. coal electricity generation [8]. The motivation behind CPIE is to provide an interactive platform for identifying the impact of coal pollution on excess death patterns across U.S. states and to serve as an open-source tool for communicating scientific evidence and findings to a broad audience.

Collaborating with experts in civil, environmental, and infrastructure engineering, biostatistics, and public health, we iteratively developed CPIE to visualize the estimated mortality risks associated with exposure to coal $PM_{2.5}$ from 1999 to 2020 for 480 U.S. EGUs. Users can explore the estimated death coefficients in each state, associated with or attributed to pollution emission coefficients from each state or facility over this period. Additionally, CPIE incorporates data on pollution reduction interventions such as scrubber installations and unit retirements, allowing users to observe the impacts of these interventions on excess deaths over time. This tool aims to enhance the understanding and communication of coal pollution's health impacts, ultimately supporting efforts to improve air quality and public health. CPIE is deployed as an open-source tool and is freely accessible by general public at here. Notably, CPIE has also drawn the attention of policymakers at the U.S. Environmental Protection Agency. Furthermore, journalists have used CPIE to gather information for reporting on specific power plants in their news articles.

2 RELATED WORK

Visualizations are an effective method for analyzing air pollution. Air pollution visualizations often involve spatiotemporal data visualization designs, with maps being widely used to present spatial distributions [14, 3, 17, 20]. Utilizing geographic maps to visualize air pollution data not only presents the spatial distribution but also helps the public better relate it to their own locations, allowing them to envision the real impact of pollution on their health [1]. In addition, charts such as line charts [14, 17], calendar views [14, 20, 6], and circular heatmaps [16, 15] are widely used in air pollution visualization to present temporal patterns with both overview contexts and fine-grained details. Deng et al. [3] developed AirVis to visualize the uncertain propagation patterns of air pollution using graph visualizations. Park et al. [17] created an interactive dashboard incorporating maps, time-series plots, and bar charts to identify high-concentration areas, temporal trends, and pollutant comparisons. Similarly, AirLens was developed to analyze air quality evolution trends [20]. Yue et al. developed AirPollutionViz [22] to explore both long-term and short-term spatiotemporal sequential patterns in air pollution evolution.

In addition to complex visual analytic systems designed for domain experts in air quality monitoring and pattern extraction, visualizations also serve as tools for data communication to educate the public, increase awareness, and promote discourse. For instance, Proma et al. [19] developed an interactive map, CleanAirNowKC, to help community members in Kansas City monitor air quality by reporting and tracking industrial emissions or toxic releases, thereby increasing community participation and engagement. Park

*sichenjin@gatech.edu
†lhennem@gmu.edu
‡jessica.roberts@cc.gatech.edu

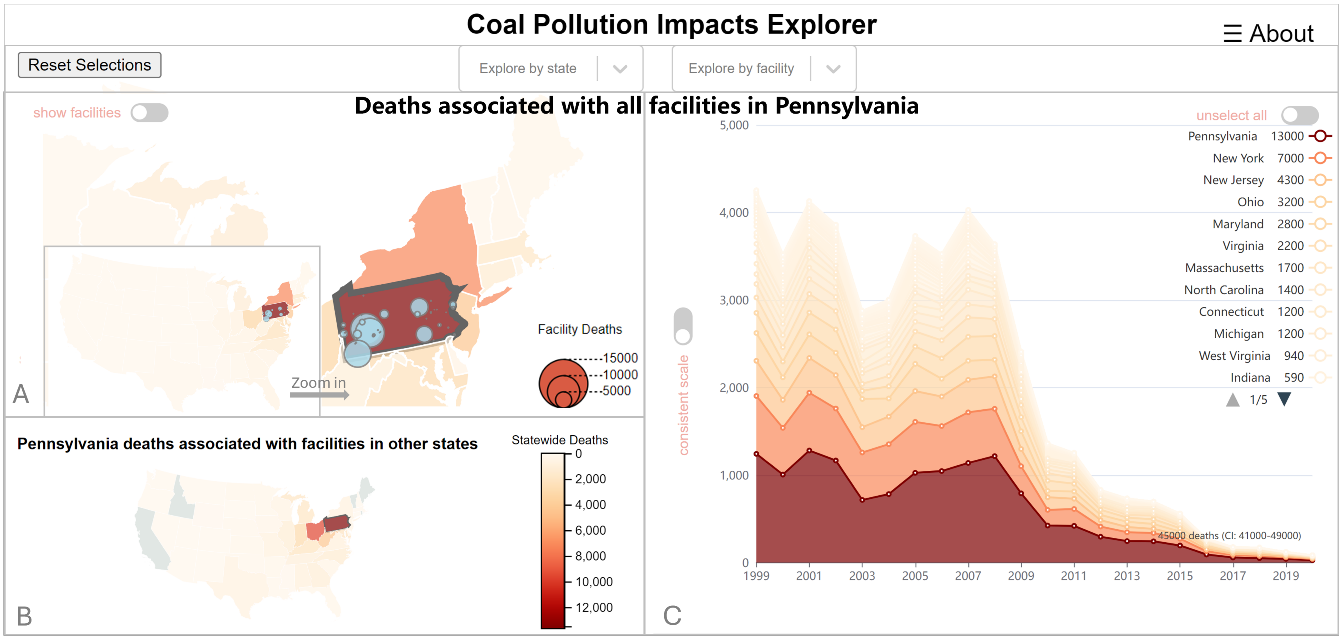

Figure 1: The user interface of CPIE shows the coal pollution impacts when Pennsylvania is selected. It consists of (A) a choropleth map view highlighting facilities in Pennsylvania and showing statewide deaths associated with all facilities in Pennsylvania, (B) a choropleth map displaying the number of deaths in Pennsylvania attributable to facilities in other states, and (C) a stacked line chart showing the changes in deaths associated with all Pennsylvania facilities from 1999 to 2020.

[17] created a web-based interactive visualization to provide non-experts with easy access to scientific research findings and knowledge, specifically showing predicted spatial and temporal concentrations by a pollution prediction model in South Korea. Carro et al. [1] evaluated the effectiveness of different visualizations (e.g., chart types, color usages) in translating complex air quality indices into comprehensible information for stakeholders.

3 Background

CPIE was built as part of a scientific study on mortality risks from U.S. coal electricity generation [8]. Prior work in this study has led to three major findings: 1) coal $PM_{2.5}$ was more harmful and linked to more deaths (nearly 500,000) than originally thought; 2) U.S. regulations and a shift from coal to natural gas reduced annual deaths from coal power plants since 1999; 3) the number of deaths from each coal power plant depends on both its emissions and its proximity to large population centers.

CPIE is designed to make the results of this research accessible to general audiences because of coal's coal's important, yet controversial status in our society; it has enabled progress by generating cheap electricity while also extracting substantial tolls on the environment and human health. Over the 22 years of study, coal pollution's impact on the environment and public health was reduced substantially with regulations and shifts to cleaner forms of energy; these industry-wide impacts were realized through actions at hundreds of individual power plants. CPIE was intended to enable users to explore the data and draw conclusions about the evolution of negative health impacts from coal power plants.

The dataset visualized in CPIE includes estimated coal emissions from each facility and the estimated number of excess Medicare deaths attributable to air pollution emissions from each coal power plant annually in each state. Here, we briefly describe the models and methods used to derive the data. For more detailed information on the methods and models, please refer to this paper about the full study [8].

Henneman et al. first collected coal EGU location and emissions information from the United States Environmental Protection Agency's Clean Air Markets Program Data tool for the years 1999-2020 [1]. They then used the HYSPLIT air pollution trajectory and dispersion tool to track how the emissions traveled through the air from each facility; in total, they ran HYSPLIT over 30 million times. They combined the results of the HYSPLIT modeling with another atmospheric chemical transport model to estimate how each coal EGU contributed to fine particulate matter ($PM_{2.5}$).

The authors then linked the exposure data with Medicare health records of people aged 65 and older, and used statistical models to estimate the increased risk of death associated with elevated coal $PM_{2.5}$. Using the combined results of the epidemiological modeling and the atmospheric modeling, they calculated the number of excess Medicare deaths attributable to air pollution emissions from each coal power plant in each year (i.e., the number of deaths that would not have occurred if that coal power plant were not operating).

4 Visual Design

The CPIE platform was iteratively designed by an HCI research team led by one of the paper authors. Preliminary work conducted in 2020 explored multiple views and formats for data representation. The HCI team created a series of interactive prototypes using the online prototyping tool Figma, which were then presented back to the paper authors and tested with users in a small user study. This preliminary work elicited the three main views in the final CPIE platform: (Figure 1): a choropleth map displaying statewide deaths associated with the selected state or facility, a choropleth map showing deaths in selected states attributable to facilities in other states, and a stacked line chart illustrating the deaths associated with facilities in each state over time. These views are interconnected through selection and highlighting interactions.

[1] https://campd.epa.gov/

4.1 Map of Statewide Deaths Associated with the Selected State or Facility

This map view (Figure 1A) comprises two layers: a choropleth map displaying the estimated statewide deaths in each state, and a dot density map indicating the locations of all 480 U.S. EGUs (facilities), with dot size representing the total deaths caused by each facility. The map supports zoom and pan interactions. Users can hover over facilities to view a tooltip showing the deaths they caused, with values in parentheses representing 95% confidence intervals (CIs). Reported deaths are rounded up to the nearest multiple of 10. Users can select a facility by clicking on a dot or searching in the 'Explore by Facility' dropdown menu. The selected facility will be highlighted in light blue, while other facilities appear in low opacity to reduce visual clutter. A 'show facilities' switch allows users to hide or show the dot density layer. Selecting a specific facility updates the choropleth map to show deaths in each state associated with the selected facility.

Users can also explore at the state level by selecting a state in the 'Explore by State' dropdown menu. This highlights facilities within the selected state in light blue and hides facilities in other states. The choropleth map updates to show deaths in each state associated with the selected state.

4.2 Map of Deaths in Selected State Associated with Facilities in Other States

This map (Figure 1B) shows how deaths in selected states can be attributed to facilities in other states (e.g., identifying the sources of air pollution causing deaths in a selected state). The darker the red, the greater the impact of a state's facilities on the selected state. For instance, we can see in Figure 1B that Pennsylvania's deaths are mainly attributable to facilities in Pennsylvania and partially attributable to facilities in Ohio. This map updates as users select a state from the 'Explore by State' dropdown menu.

4.3 Stacked Line Chart

The stacked line chart (Figure 1C) displays deaths in each state associated with the selected state or facility from 1999 to 2020. The legend and lines are sorted by the number of deaths. Hovering over a line highlights the state on the map and shows a tooltip with the number of deaths. A text annotation in the bottom right corner of the chart shows the total deaths associated with the current selection, with values in parentheses representing 95% confidence intervals. Reported deaths are rounded up to the nearest multiple of 10. The legend lists states with the most deaths at the top. Users can click state names in the legend to select and deselect specific states to hide or display them in the chart. When 'consistent scale' is enabled, the y-axis value range remains unchanged when users select different states or facilities, aiding comparison. Otherwise, the y-axis maximum updates to the extreme value of the current selection.

4.4 Implementation

CPIE is implemented using D3.js and React.js for front-end visualization and interaction. CPIE is deployed as an open-source tool and is freely accessible at here.

5 Case Studies

5.1 Overall Distribution and Trends

The landing page of the tool shows the overall spatial distributions of all 480 facilities on the map and the estimated total deaths in each state due to coal pollution exposure (Figure 2A), and the total estimated deaths attributed to coal pollution emissions from 1999 to 2020 in each state (Figure 2B). The data reveals that Pennsylvania, Ohio, New York, Illinois, and North Carolina are the top five states with the highest number of deaths attributable to coal pollution. Pennsylvania, with its extensive history of coal mining and numerous coal-fired power plants, tops the list. The state's significant coal production and use for electricity generation result in substantial pollutant emissions, adversely affecting public health. The stacked line chart, summarizing data over the study period, indicates that annual excess deaths attributable to coal PM2.5 were worst between 1999 and 2007, which adds up to 390,000 (95% Confidence Interval: 360,000 to 430,000) and averages over 43,000 deaths per year. The annual excess deaths decreased significantly after 2007, dropping below 1,600 (95% Confidence Interval: 1,400 to 1,700) in 2020.

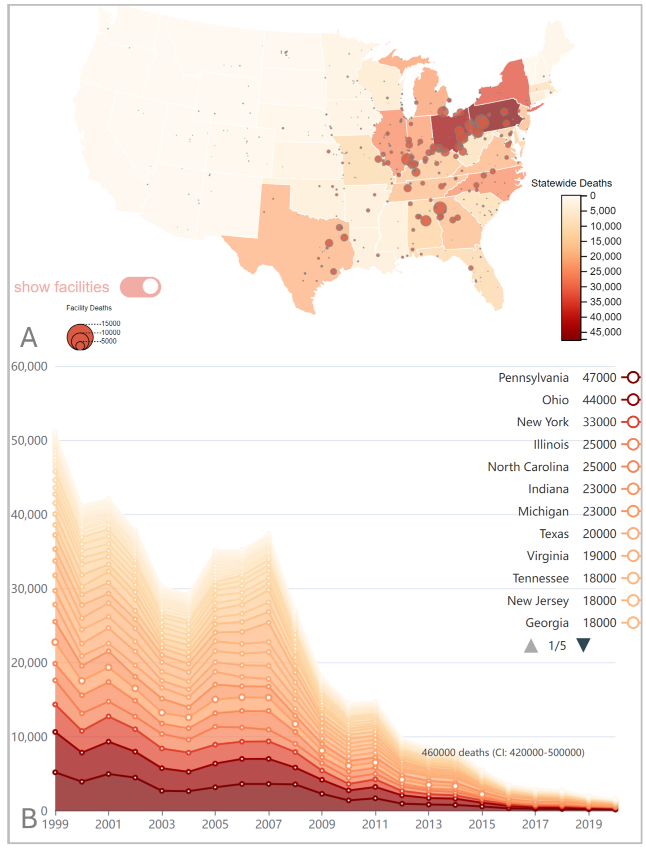

Figure 2: (A) An overview of the spatial distributions of all 480 facilities and the deaths associated with them in each state. (B) The total estimated deaths attributed to coal pollution emissions from 1999 to 2020 in each state.

5.2 Localized Impact

Air pollution from coal plants consistently has the most significant impact on nearby regions. Emissions from these plants, including particulate matter, sulfur dioxide, and nitrogen oxides, tend to disperse locally before spreading further afield, through poor air quality, water pollution, and soil degradation, etc. As a result, communities close to coal plants are the first and most severely affected by the harmful pollutants. This localized impact is evident in the data: for almost every state or facility, the majority of deaths associated with coal plant emissions occur within the same state or in neighboring states. We take the top most influential facilities in Michigan (Figure 3A), Texas (Figure 3B), Georgia (Figure 3C), Florida (Figure 3D) as examples to show the effect. These facilities represent that around 50% of deaths occur within the same state and neighboring states. Similarly, at the state level, deaths are mostly attributable to facilities within the same state (Figure 3E, F, G, H).

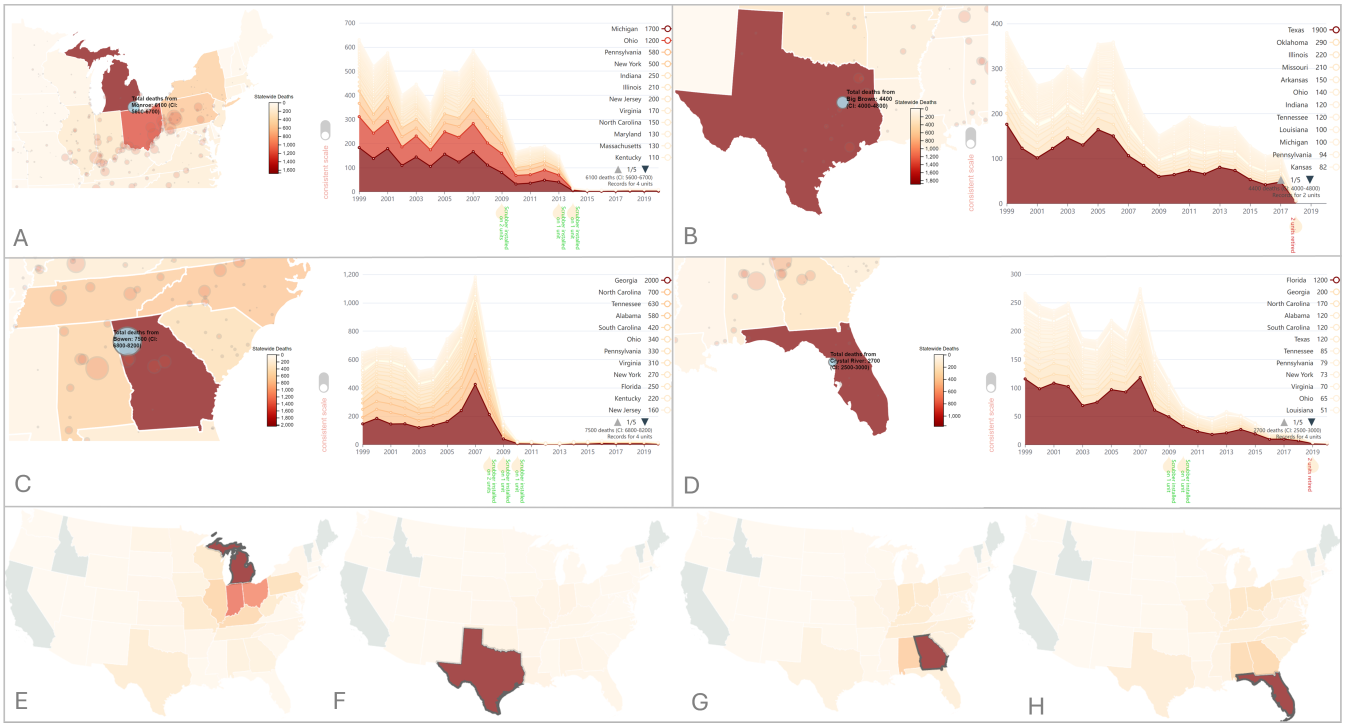

Figure 3: The deaths associated with selected facilities: (A) the Monroe facility (Michigan), (B) Big Brown facility (Texas), (C) Bowen facility (Georgia), and (D) Crystal River facility (Florida). The stacked line charts mark the timing of regulatory interventions on the x-axis (e.g., scrubber installations, unit retirements). Deaths in (E) Michigan, (F) Texas, (G) Georgia, and (H) Florida attributed to pollution from facilities in other states.

This pattern underscores the direct correlation between the concentration of coal plants and the incidence of pollution-related health issues.

5.3 Effective Regulatory Interventions

In coal pollution control, scrubbers play a crucial role in cleaning the gases that pass through the smokestacks of coal-burning power plants. Scrubbers are devices that use a liquid, usually a mixture of water and various chemicals, to remove pollutants from the exhaust gases before they are released into the atmosphere. Installing scrubbers helps mitigate the harmful environmental and health effects associated with sulfur dioxide emissions [10].

The timeline on the stacked line chart highlights the installation time of scrubbers and the retirement time of coal plant units. This timeline allows us to examine the impact of these regulatory measures on public health. We found that both installing emissions control devices and shutting coal facilities completely have a visible and immediate effect on reducing the number of deaths associated with coal pollution. This indicates the effectiveness of regulatory measures in improving air quality and protecting public health.

In Figure 3, we selected some influential facilities as examples to illustrate this effect. The stacked lines chart for each facility shows a significant drop in associated deaths following scrubbers' installation or units' retirement. For instance, Facility Monroe (Michigan), which installed scrubbers in 2009, shows a marked decline in mortality rates in the subsequent years. The total number of deaths associated with the facility dropped to single digits after two additional scrubbers were installed in 2013 (Figure 3A). Similarly, the 'Big Brown' facility in Texas, which retired two of its units in 2018, shows a significant decrease in deaths related to its emissions (Figure 3B). These examples highlight the essential role of regulatory interventions in mitigating the health impacts of coal pollution.

6 Limitations and Future Work

We note the following improvements for future iterations of CPIE. First, representing uncertainties in the estimated dataset properly is critical for enhancing the tool's accuracy and reliability. For instance, current estimations include confidence intervals for excess deaths associated with each coal facility, presented in CPIE through tooltips with values noted in parenthesis. To further refine this, future versions of CPIE could consider adopting glyphs, such as error bars or box plots, instead of the dot symbol to show the confidence intervals on the map. These enhancements would not only provide clearer visual cues but also convey the uncertainty range more intuitively to users, thereby improving the tool's interpretability and usefulness in decision-making processes related to public health and environmental policy [18, 9, 11].

Second, we will work with users to refine the color mapping choices in the tool. Our intention with the light yellow to dark red gradient was to visually represent the magnitude of deaths across regions, using prominent red to highlight the most impacted areas at a glance. However, we recognize that variations in saturation could affect how readers perceive the importance of data points, potentially leading to misinterpretation. Therefore, in future work, we will consider a color mapping strategy that adjusts the hue while limiting changes to the saturation range.

Additionally, the current map displays most values aggregated by time or by state. With a finer granularity dataset, we aim to implement an animated simulation to visualize the daily dispersion process of coal pollution. Figure 4 shows a screenshot from previous formative design studies (Section 4) that visually simulates the dispersion of pollution from specific facilities. This dynamic simulation will offer a more nuanced view of the pollution's impact, illustrating how pollution disperses under varying weather conditions and across different geographic regions. This animated dispersion process will further introduce uncertainties inherent in estimated locations and timestamps. Potential ways to address these uncertainties include shading along the dispersion traces to represent the range of possible pollution levels and outlining affected areas with dashed lines to show regions estimated to be impacted.

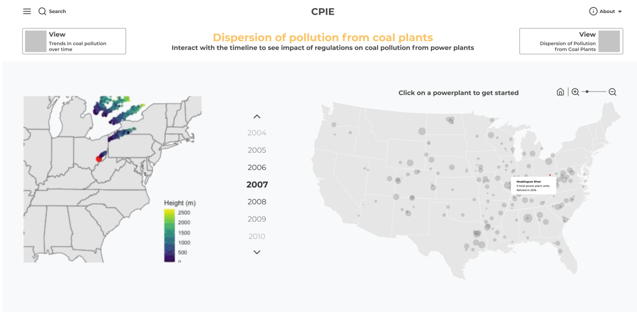

Figure 4: A visual simulation design for coal pollution dispersion, based on previous formative design studies for future work.

7 Conclusion

Visualizations are widely adopted in air pollution analysis; however, few interactive tools have been developed to illustrate the impact of coal pollution on public health data. In this work, we present an interactive visualization tool to explore the estimated excess statewide deaths associated with 480 electricity-generating units (EGUs). We employed choropleth maps and stacked line charts to allow users to interactively explore the spatial distribution and temporal trends of coal pollution impacts. Using this tool, we demonstrated three key findings. We hope this work will make scientific findings more accessible to the general audience and increase awareness among stakeholders and policymakers.

Acknowledgments

We wish to thank Sue Reon Kim and Srijan Jhanwar for their work on user interviews and ideation, which led to the development of multiple versions of the pioneering CPIE prototype through an iterative process. This work was supported in part by a grant from the U.S. Environmental Protection Agency (EPA) EmPOWER Air Data Challenge; The Health Effects Institute (HEI) (grant R-82811201). The manuscript contents are solely the responsibility of the grantee and do not necessarily represent the official views of the EPA. Further, the EPA does not endorse the purchase of any commercial products or services mentioned in the publication. Research described in this article was conducted under contract to the HEI, an organization jointly funded by the EPA and certain motor vehicle and engine manufacturers. The contents of this article do not necessarily reflect the views of HEI, or its sponsors, nor do they necessarily reflect the views and policies of the EPA or motor vehicle and engine manufacturers.

References

[1] G. Carro, O. Schalm, W. Jacobs, and S. Demeyer. Exploring actionable visualizations for environmental data: Air quality assessment of two belgian locations. *Environmental Modelling & Software*, 147:105230, 2022. 1, 2

[2] I. C. Dedoussi, S. D. Eastham, E. Monier, and S. R. Barrett. Premature mortality related to united states cross-state air pollution. *Nature*, 578(7794):261–265, 2020. 1

[3] Z. Deng, D. Weng, J. Chen, R. Liu, Z. Wang, J. Bao, Y. Zheng, and Y. Wu. Airvis: Visual analytics of air pollution propagation. *IEEE transactions on visualization and computer graphics*, 26(1):800–810, 2019. 1

[4] J. M. Godowitch, G. A. Pouliot, and S. T. Rao. Assessing multi-year changes in modeled and observed urban nox concentrations from a dynamic model evaluation perspective. *Atmospheric Environment*, 44(24):2894–2901, 2010. 1

[5] A. L. Goodkind, C. W. Tessum, J. S. Coggins, J. D. Hill, and J. D. Marshall. Fine-scale damage estimates of particulate matter air pollution reveal opportunities for location-specific mitigation of emissions. *Proceedings of the National Academy of Sciences*, 116(18):8775–8780, 2019. 1

[6] F. Guo, T. Gu, W. Chen, F. Wu, Q. Wang, L. Shi, and H. Qu. Visual exploration of air quality data with a time-correlation-partitioning tree based on information theory. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 9(1):1–23, 2019. 1

[7] M. Hendryx, K. J. Zullig, and J. Luo. Impacts of coal use on health. *Annual review of public health*, 41:397–415, 2020. 1

[8] L. Henneman, C. Choirat, I. Dedoussi, F. Dominici, J. Roberts, and C. Zigler. Mortality risk from united states coal electricity generation. *Science*, 382(6673):941–946, 2023. 1, 2

[9] Z. Huang, Y. Lu, E. A. Mack, W. Chen, and R. Maciejewski. Exploring the sensitivity of choropleths under attribute uncertainty. *IEEE transactions on Visualization and Computer Graphics*, 26(8):2576–2590, 2019. 4

[10] Y.-T. Hung and I. Ashner. Sulfur dioxide emission and mitigation. In *Handbook of Environment and Waste Management: Acid Rain and Greenhouse Gas Pollution Control*, pp. 627–658. World Scientific, 2020. 4

[11] C. Kinkeldey, A. M. MacEachren, and J. Schiewe. How to assess visual communication of uncertainty? a systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51(4):372–386, 2014. 4

[12] J. Lelieveld, J. S. Evans, M. Fnais, D. Giannadaki, and A. Pozzer. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 525(7569):367–371, 2015. 1

[13] J. Lelieveld, K. Klingmüller, A. Pozzer, R. Burnett, A. Haines, and V. Ramanathan. Effects of fossil fuel and total anthropogenic emission removal on public health and climate. *Proceedings of the National Academy of Sciences*, 116(15):7192–7197, 2019. 1

[14] H. Li, H. Fan, and F. Mao. A visualization approach to air pollution data exploration—a case study of air quality index (pm2. 5) in beijing, china. *Atmosphere*, 7(3):35, 2016. 1

[15] J. Li, Z. Xiao, H.-Q. Zhao, Z.-P. Meng, and K. Zhang. Visual analytics of smogs in china. *Journal of Visualization*, 19:461–474, 2016. 1

[16] W. Lu, T. Ai, X. Zhang, and Y. He. An interactive web mapping visualization of urban air quality monitoring data of china. *Atmosphere*, 8(8):148, 2017. 1

[17] Y. Park, I. Song, J. Yi, S.-J. Yi, and S.-Y. Kim. Web-based visualization of scientific research findings: national-scale distribution of air pollution in south korea. *International journal of environmental research and public health*, 17(7):2230, 2020. 1, 2

[18] A. Preston and K.-L. Ma. Communicating uncertainty and risk in air quality maps. *IEEE Transactions on Visualization and Computer Graphics*, 29(9):3746–3757, 2022. 4

[19] R. A. Proma, M. Sumpter, H. Lugo, E. Friedman, K. T. Huq, and P. Rosen. Cleanairnowkc: Building community power by improving data accessibility. In *2021 IEEE Workshop on Visualization for Social Good (VIS4Good)*, pp. 1–5. IEEE, 2021. 1

[20] D. Qu, C. Lv, Y. Lin, H. Zhang, and R. Wang. Airlens: Multi-level visual exploration of air quality evolution in urban agglomerations. In *Computer graphics forum*, vol. 41, pp. 223–234. Wiley Online Library, 2022. 1

[21] L. Xu, H. Guo, C. M. Boyd, M. Klein, A. Bougiatioti, K. M. Cerully, J. R. Hite, G. Isaacman-VanWertz, N. M. Kreisberg, C. Knote, et al. Effects of anthropogenic emissions on aerosol formation from isoprene and monoterpenes in the southeastern united states. *Proceedings of the National Academy of Sciences*, 112(1):37–42, 2015. 1

[22] X. Yue, D. Feng, D. Sun, C. Liu, H. Qin, and H. Hu. Airpollutionviz: visual analytics for understanding the spatio-temporal evolution of air pollution. *Journal of Visualization*, pp. 1–19, 2024. 1