# **XplainAct: Visualization for Personalized Intervention Insights**

Yanming Zhang \*

Krishnakumar Hegde †

Klaus Mueller‡

Department of Computer Science Stony Brook University

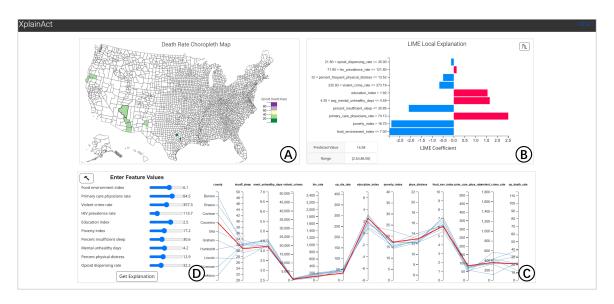


Figure 1: The XplainAct interface, illustrated here using the opioid dataset. (A) Choropleth map highlighting Coconino County, Arizona (bold black outline) and counties with similar opioid-related socioeconomic indicators with their respective opioid death rates colored according to the legend on the right; (B) Feature explanation panel using LIME to reveal the contribution of each socioeconomic factor to the opioid death rate in Coconino County; (C) Parallel coordinates plot comparing Coconino County's multidimensional profile (red line) with its socioeconomic peers (blue lines) and highlighted on the choropleth map (A); and (D) Slider group for defining contextual similarity criteria to Coconino County's profile.

#### **ABSTRACT**

Causality helps people reason about and understand complex systems, particularly through what-if analyses that explore how interventions might alter outcomes. Although existing methods embrace causal reasoning using interventions and counterfactual analysis, they primarily focus on effects at the population level. These approaches often fall short in systems characterized by significant heterogeneity, where the impact of an intervention can vary widely across subgroups. To address this challenge, we present XplainAct, a visual analytics framework that supports personalized causal analysis by enabling interventions at the individual level within subpopulations. We demonstrate the effectiveness of XplainAct through two case studies: investigating opioid-related deaths in epidemiology and analyzing voting inclinations in the presidential election.

**Index Terms:** Explainable AI, Causality, Personalization, Visual Analytics.

## INTRODUCTION

in recent years have created a growing need for tools that can ef-

\*e-mail: yanmzhang@cs.stonybrook.edu †e-mail: khegde@cs.stonybrook.edu

‡e-mail: mueller@cs.stonybrook.edu

The advancements of machine learning and artificial intelligence

fectively support the understanding and modification of complex systems. Traditional analytical methods, which rely on correlation, merely observe how variables tend to change together. In contrast, causal reasoning goes a step further by determining whether changes in certain variables can causally influence others. Here, independent and dependent variables are often viewed as treatment and outcome variables, respectively, with the effect commonly quantified as the average treatment effect (ATE).

The Average Treatment Effect (ATE) is commonly used in observational studies. In standard causal inference settings, each individual receives only one treatment or exposure, meaning that only one outcome is observed, while the outcome under the alternative treatment—the counterfactual—remains unobserved. The ATE estimates the causal effect at the population level by conceptually treating individuals exposed to the alternative treatment as standins for the unobserved counterfactuals. This estimand has proven effective in one-size-fits-all analyses, such as determining whether smoking causes cancer [3], sleep deprivation leads to cognitive impairment [1], or anthropogenic greenhouse gas emissions drive global warming [15].

Despite the usefulness of the ATE in uncovering general patterns, its averaging process can obscure critical variation in heterogeneous systems. A treatment that appears beneficial on average may be ineffective—or even harmful—for certain subgroups within the population. For instance, Manson et al. [13] found that hormone therapy benefited younger postmenopausal women but posed risks for older women. Similarly, Kravitz et al. [11] showed that the risk of gastrointestinal bleeding from aspirin use varied significantly depending on a patient's age and history of peptic ulcers.

The examples above underscore the need to account for variations in individual responses to interventions. Yet, existing approaches face two key challenges. First, due to regulations and ethical concerns, it is often infeasible to apply certain interventions—such as assigning students to underperforming schools to study the effects of poor educational environments—thus limiting the use of randomized trials. Second, explainability is essential for building practitioner trust. Many AI-based personalization tools operate opaquely, and practitioners often distrust recommendations that lack a clear rationale. This poses a barrier to adoption in high-stakes domains like clinical medicine [21]. These challenges underscore the need for systems that can personalize, explore, and interpret intervention effects—motivating the development of a visual analytics tool to support human-in-the-loop causal analysis.

In this work, we propose XplainAct, a visual analytics framework designed to support the formulation, estimation, and interpretation of intervention effects, with particular emphasis on response heterogeneity. We demonstrate how XplainAct can recommend the most effective interventions to shift outcome values toward desired targets through two illustrative usage scenarios.

#### 2 RELATED WORK

Causal inference is primarily approached through two frameworks: Structural Causal Models (SCM) by Pearl [16] and the Potential Outcomes (PO) framework by Rubin [19]. SCM uses graphical models to elucidate causal structures and interventions, while PO relies on statistical methods to approximate randomized trials using observational data. Based on the two frameworks, a growing number of methods has been developed focusing on heterogeneous effects in recent years. Methods such as Causal Trees [2] and Causal Forests [22] address heterogeneity by recursively partitioning the data to identify subgroups with distinct treatment responses. Hill et al. [9] offer a flexible probabilistic approach to estimate the conditional average treatment effect (CATE) using Bayesian Additive Regression Trees (BART). Although powerful, these methods face interpretability challenges, particularly in scenarios that require domain-specific user interpretation.

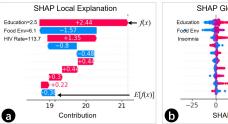
Visual analytics systems have emerged to enhance user understanding and interaction with causal inference processes [25, 10, 6]. Wang et al. [23] introduced one of the earliest interactive visual frameworks for causal reasoning, featuring causal networks and "what-if" simulations. This work was later extended to identify subgroup-specific causal networks in heterogeneous datasets [24]. Guo et al. [7] developed Causalvis, a Python-based visualization toolkit that supports confounder adjustment, cohort refinement, and exploration of heterogeneous treatment effects. Fan et al. [4] proposed a visual analytics approach for discovering and comparing causal graphs across multiple outcome variables in health data. CausalChat [26] expands causal reasoning chains with the help of large language models, while CausalPrism [27] introduces a novel algorithm to generate interpretable subgroups exhibiting significant treatment effects.

Despite these advances, existing visual analytics systems often fall short in supporting the analysis and interpretation of personalized interventions within subgroups, limiting their utility in highly heterogeneous contexts. XplainAct builds on insights from these prior methods to address this gap.

#### 3 XPLAINACT

During XplainAct's design phase, we collaborated with NIH public health experts to understand their needs for visualizing and interacting with personalized interventions. Their feedback informed the following four design goals (DG):

 DG1: Specify intervention and estimate outcomes. Xplain-Act should allow analysts to apply interventions to a unit of



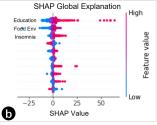


Figure 2: Explanations supported by SHAP. (a) Waterfall plot showing cumulative contributions from each feature. (b) Beeswarm plot showing SHAP values of each data point.

interest and estimate potential outcomes under various treatments. Besides, the system should also convey the uncertainty of these outcome estimates.

- DG2: Identify subgroups. Since individual responses vary based on feature attributes (covariates), XplainAct should enable users to identify relevant subgroups to account for heterogeneity during intervention analysis.
- **DG3: Explain intervention results.** To support reasoning and build trust, the framework should elucidate how each attribute contributes to the outcome and reveal the sources of variation in predicted results.
- DG4: Show feature-to-space relations. XplainAct is motivated by contexts where units are distributed geographically, thereby requiring visual tools that can help users to recognize familiar spatial patterns and intuitively link descriptive features to locations.

XplainAct uses Python on the backend, employing Flask for handling HTTP requests, DoWhy for causal inference, and additional libraries for machine learning algorithms. The front end is built with JavaScript, using D3.js to create interactive visualizations.

## 3.1 Visual Components

XplainAct has three views: the Choropleth Map view (A), the Explanation view (B), and the Intervention view (C) (see Fig. 1).

The Choropleth Map view visualizes outcomes across geographic units using two sequential color schemes to differentiate polarities. In the opioid dataset we use as a case study, shades of purple represent higher values of opioid-related deaths, while shades of green indicate lower values. A neutral or medium outcome value is depicted in light yellow. Hovering over a geographic unit reveals a tooltip showing the unit's precise outcome value. Typically, users explore areas exhibiting extreme values, notable spatial patterns, or regions with which they have high familiarity (**DG4**).

The Explanation view supports both local and global explainability for intervention results (**DG3**). It integrates two explainable AI tools—LIME and SHAP—to provide complementary interpretations from distinctive perspectives.

For local explainability using LIME (Local Interpretable Model-Agnostic Explanations [18]) (see Fig. 1(B)), the interface displays a vertically arranged diverging bar chart alongside a table in the bottom-left corner. The table shows the predicted outcome along with a variability interval derived from a set of perturbed samples (**DG1**). In the bar chart, each bar represents a feature, with its length indicating the magnitude of the feature's contribution to the prediction. Red bars indicate features that increase the prediction (e.g., education index > 1.92), while blue bars indicate features that lower the prediction (e.g., poverty index > 16.7). LIME approximates the model locally by learning a simple, interpretable model around the instance of interest—allowing users to understand which features are driving the prediction in that specific context.

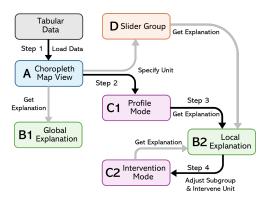


Figure 3: Workflow of XplainAct for analyzing personalized interventions. Each block represents a view, and each edge denotes an interaction between views.

Alternatively, users can switch to a waterfall plot generated using SHAP (SHapley Additive exPlanations [12]) for a different perspective on local explanations (see Fig. 2(a)). A SHAP waterfall plot explains a single prediction by decomposing it into additive contributions from each input feature. Starting from the model's baseline (i.e., the average prediction across the dataset), the plot shows how each feature pushes the prediction higher or lower, using bars whose lengths represent the magnitude of the contribution. These contributions are computed based on Shapley values, which quantify the average marginal effect of each feature across all possible subsets of features. The resulting plot highlights how much each feature contributed to the final prediction and in what direction, using a consistent color scheme to distinguish positive and negative influences.

Global explainability is visualized using SHAP beeswarm plots (see Fig. 2(b)), where each data (i.e., county) is shown as a scatter point. The horizontal axis represents the SHAP value, indicating how much a feature contributes to the model output for each instance. Features are listed vertically, ordered by their overall importance (i.e., the average absolute SHAP value across all instances). Each point is colored by the corresponding feature value, transitioning from low (blue) to high (red), thus revealing how different value ranges influence predictions. The beeswarm plot effectively combines feature importance with value distribution in a single view, for both granular and high-level insight into model behavior.

The Intervention view has two modes: profile mode and intervention mode. The profile mode features a slider group and a parallel coordinates plot (PCP) (see Fig. 1(c, d)). Slider values, reflecting the feature values of the unit  $u_i$ , correspond to a prominent red polyline in the PC plot. Additionally, the ten nearest neighbors of  $u_i$ , based on feature attributes, appear as thinner blue polylines. In intervention mode (see Fig. 4), users can adjust the subgroup via a dedicated slider on the far left (**DG2**). The PCP is enhanced with translucent red rectangles on each axis to indicate the value ranges defining the selected subgroup. In this mode, only the polyline of the selected unit (shown in red) and its counterfactual polyline (shown in blue) are displayed, with the intervention attribute highlighted by a blue axis title.

## 3.2 Interactions

Fig. 3 depicts the iterative process of *XplainAct*, operationalized through the dashboard interface shown in Fig. 1. Starting with a tabular dataset, the choropleth map visualizes the spatial distribution of the outcome variable, helping analysts identify areas of interest (Step 1). When a unit of interest  $u_i$  – a county – is selected, the map highlights  $u_i$  with a thicker black outline and shows its peers based on feature similarity. This is reflected in the *profile mode* of the intervention view, where the sliders and the PCP are updated

accordingly (Step 2). Subsequently, users can either customize feature attributes using the sliders or request a local model explanation for  $u_i$  by clicking the *Get Explanation* button. The resulting explanation is then displayed in the explanation view (Step 3).

In Step 4, with the insights gained from local explanations, users can switch to intervention mode to specify and simulate an intervention by clicking the axis of the desired intervention variable in the PCP (**DG1**) and setting the desired intervention value (see Fig. 4). The generated counterfactual unit  $u'_i$  represented by the blue polyline then reflects the impact of the selected intervention – the treatment – on all other variables downstream on the causal graph (see Section 3.3) (**DG1**). In this process, the values of the treatment's causal parents remain unchanged – only the values for the causal children are affected (**DG3**). Users can also adjust the subgroup by modifying the number of nearest neighbors (**DG2**).

XplainAct enables analysts to iteratively engage in a loop of proposing personalized interventions, estimating outcomes, and generating explanations through Steps 2 to 4. Once a unit is specified (Step 2), the feature importance shown in the explanation view provides insights that help analysts identify which feature attributes to target for intervention (Step 3). Analysts can then refine the subgroup and apply interventions accordingly (Step 4).

## 3.3 Algorithms

The subgroup to which  $u_i$  belongs is determined based on its K nearest neighbors, with a default value 10 when initialized. The neighbors are identified using Locality Sensitive Hashing (LSH) [5] for efficient similarity search.

All data associated with the identified subgroup is then used to compute counterfactual outcomes—i.e., predicted under alternative interventions. For this purpose, we use the fixed causal graph established during the initialization phase (see supplemental material), and estimate model coefficients using only the subgroup data involved in the current investigation. To support counterfactual inference, we fit an invertible Structural Causal Model (ISCM), defined by a set of structural functions  $\mathscr{F} = \{f_v | v \in V\}$  where each function models the causal mechanism for a node v:

$$X_{\nu} = f_{\nu} \left( X_{\text{pa}(\nu)}, N_{\nu} \right)$$

where  $X_{\nu}$  – an endogenous variable – is a child node in the causal graph,  $X_{\text{pa}(\nu)}$  – the exogenous variables – operate as parent nodes in the causal graph, and  $N_{\nu}$  denotes the noise term.

Aiming to fit an Interpretable Structural Causal Model (ISCM), we implement Pearl's three-step process [17]. For simplicity, we assume linear functional forms. Under this assumption, the resulting coefficients approximate local effects, capturing how individual responses vary in reaction to the intervention.

#### 4 CASE STUDIES

We demonstrate the capabilities of XplainAct by presenting two usage scenarios that employ real-world datasets. Visuals for these two investigations can be found in the supplementary material.

The Opioid Death Dataset focuses on the the relation between opioid-related deaths and social factors. It combines 10 key socioeconomic attributes sourced from the County Health Ranking database [20] with opioid death data from the CDC WONDER database [14]. It covers over 3,000 US counties. In the dataset, each county is characterized by 11 attributes: food environment index, primary care physicians rate, violent crime rate, HIV prevalence rate, education index, poverty index, percent insufficient sleep, average mental unhealthy days, percent frequent physical distress, opioid dispensing rate. Opioid death rate is the outcome variable and other variables are treatment variables.

The Presidential Election Dataset centers on the correlation between voting inclination and socioeconomic drivers. The dataset

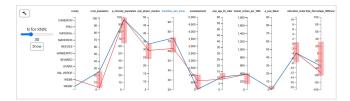


Figure 4: Bob simulates an intervention on increasing household homeownership.

includes 9 carefully selected socioeconomic attributes from a 2016 US election dataset [8] and each data point is a US county. The attributes are: percent rural population, percent minority population, percent population physically inactive, percent household own home, number of unemployment, percent population age 65 and older, violent crimes per 100k, education index, and percent population Black. The outcome variable is vote percentage difference with positive value leaning towards the Republican's candidate and negative value leaning towards the Democrat's candidate.

## 4.1 Case Study 1: Opioid-Related Death Dataset

We follow Taylor, a public health enthusiast who is deeply concerned about the opioid crisis in her home of Boone County, West Virginia. She turns to XplainAct in hopes of understanding why her county is struggling with high opioid mortality and identifying actionable steps that could help reduce the death rate.

She begins by analyzing the choropleth map to understand how opioid-related deaths vary across counties. By exploring several high-death-rate areas (shaded in purple) and low-death-rate areas (shaded in green), she compares their attribute distributions in the PCP. This exploration gives her an overview of how different attributes appear to correlate with opioid mortality, helping her form initial hypotheses about potential risk factors.

Next, she decides to focus on Boone County and its cluster by selecting it on the choropleth map. The map highlights Boone and its peer counties, and the PCP immediately reveals several concerning patterns shared across the cluster: high levels of insufficient sleep, a large number of mentally unhealthy days, and low education levels. These trends stand out as potential contributing factors to the cluster's alarmingly high opioid death rate.

To explore potential mitigating strategies, Taylor switches to the intervention mode of the PCP. In this mode, she interacts with the model to simulate the effects of various hypothetical interventions. After experimenting with different attributes, she discovers that the opioid death rate can be significantly reduced by lowering the number of mentally unhealthy days. Further exploration reveals that decreasing the percent of insufficient sleep also reduces the number of mentally unhealthy days, leading to an additional drop in the opioid death rate. This causal chain offers actionable insights for designing targeted interventions to address the crisis in her community.

To better understand Boone County's situation, Taylor turns to the explanation view. The waterfall plot generated using SHAP highlights key factors contributing to the high *opioid death rate*, including high *percent of insufficient sleep*, a high *number of mentally unhealthy days*, elevated *HIV prevalence rate*, and a high *violent crime rate*. The first two attributes confirm the trends she observed earlier, while all four collectively reveal the underlying challenges contributing to the opioid crisis in Boone County and its peers.

Based on these findings, Taylor recommends that Boone County promote better sleep habits through public health and workplace initiatives, expand access to health services, and address broader issues like crime and education. She also suggests that other counties with similar patterns could apply these priorities to reduce opioid-related challenges.

## 4.2 Case Study 2: Presidential Election Dataset

In this study, we follow Bob, a Republican campaign staffer, as he uses XplainAct to investigate why the Democratic Party secured a 53% lead in Webb County, Texas. His goal is to uncover factors influencing voter turnout and explore strategies to boost Republican support by mobilizing the "right" voters in future elections.

Bob begins his exploration by examining the choropleth map, which provides a geographic overview of election results across the country. In Texas, most counties lean Republican, except for South Texas, where Webb County is located. Upon clicking Webb County, he observes that counties with similar attributes are mostly concentrated in South Texas and California, and—except for Seward County, Kansas—nearly all are Democratic-leaning. He then proceeds to the intervention mode, where he adjusts the number of neighbors considered for Webb County. This helps him assess whether Webb's voting trend reflects a broader regional pattern or represents an outlier.

Meanwhile, the pattern shown in the PCP draws Bob's attention. The subgroup of Webb County is characterized by a low *rural population*, a low *percentage of residents aged 65 and older*, and a high *percentage of minority residents*. It also shows relatively low values for *unemployment rate*, *violent crime rate*, *percentage of Black population*, and *education index*. Bob agrees that many of these attributes align with his understanding of Webb County. However, some characteristics do not match common stereotypes of a Democratic-leaning county, prompting his curiosity about how each attribute influences the vote difference.

Bob then turns to the explanation view to examine the LIME and SHAP summary plots for Webb County. These plots decompose the impact of each attribute on the county's vote difference. Bob discovers that both LIME and SHAP identify the percentage of minority population as the most significant factor contributing to Webb's Democratic lean. Specifically, the SHAP waterfall plot shows that Webb County's 95% minority population alone contributes a -72.48 shift from the national average vote difference of 32.01. This places Webb as a Democratic stronghold, consistent with common expectations. Other attributes—such as *low rural population* and moderately low *homeownership rate*—further reinforce this alignment.

Conversely, Webb also exhibits characteristics typically associated with Republican support, such as a low *percentage of Black residents* and a moderately low *education index*. To his surprise, Bob notices that within Webb's subgroup, a low percentage of elderly residents pushes the county further toward the Democratic side—contradicting the common perception that the Republican Party performs strongly among older voters. This suggests that Webb County may follow a unique age-related voting pattern.

With this knowledge in mind, Bob now seeks to explore potential strategies for improving Republican performance. He toggles the PCP to intervention mode, where the interactive interface allows him to simulate changes and observe their impact on the vote difference. First, he adjusts the percentage of minority population and finds that decreasing this attribute predicts an increase in Republican votes. He also explores the effect of increasing the proportion of homeowners, which results in a positive shift toward the Republican Party. While population structure cannot be directly changed, the intervention suggests that policies or campaigns targeting homeowners and encouraging them to vote could be an effective way to gain traction for the Republican Party in Webb County. Armed with these insights, Bob is prepared to recommend targeted strategies to improve Republican performance in future elections.

#### 5 DISCUSSION AND CONCLUSION

We presented XplainAct, a visual analytics framework for exploring causal heterogeneity through individual-level interventions and subgroup analysis. Two case studies demonstrate its ability to provide interpretable, context-specific insights through an interactive

dashboard and explanation tools.

A current limitation of XplainAct is its use of a uniform causal graph across subgroups, which may miss subgroup-specific structures. Future work should explore dynamic graph generation and more flexible models, such as Causal Forests or Bayesian methods, to improve analytical depth. Scalability to larger datasets and robust validation of causal assumptions also remain important directions. Yet, overall, XplainAct advances visual frameworks for interpreting decision-making in complex, heterogeneous settings.

#### REFERENCES

- P. Alhola and P. Polo-Kantola. Sleep deprivation: Impact on cognitive performance. *Neuropsychiatric disease and treatment*, 3(5):553–567, 2007.
- [2] S. Athey and G. Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016. 2
- [3] R. Doll and A. B. Hill. Smoking and carcinoma of the lung. *British medical journal*, 2(4682):739, 1950.
- [4] M. Fan, J. Yu, D. Weiskopf, N. Cao, H.-Y. Wang, and L. Zhou. Visual analysis of multi-outcome causal graphs. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2
- [5] A. Gionis, P. Indyk, R. Motwani, et al. Similarity search in high dimensions via hashing. In Vldb, vol. 99, pp. 518–529, 1999. 3
- [6] G. Guo, M. Glenski, Z. Shaw, E. Saldanha, A. Endert, S. Volkova, and D. Arendt. Vaine: Visualization and ai for natural experiments. In 2021 IEEE Visualization Conference (VIS), pp. 21–25. IEEE, 2021. 2
- [7] G. Guo, E. Karavani, A. Endert, and B. C. Kwon. Causalvis: Visualizations for causal inference. In *Proceedings of the 2023 CHI conference on human factors in computing systems*, pp. 1–20, 2023. 2
- [8] B. Hamner. 2016 us election, 2016. Dataset available on Kaggle. 4
- [9] J. L. Hill. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240, 2011. 2
- [10] S. Kaul, D. Borland, N. Cao, and D. Gotz. Improving visualization interpretation using counterfactuals. *IEEE Transactions on Visualiza*tion and Computer Graphics, 28(1):998–1008, 2021. 2
- [11] R. L. Kravitz, N. Duan, and J. Braslow. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank Quarterly*, 82(4):661–687, 2004. 1
- [12] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. Advances in neural information processing systems, 30, 2017. 3
- [13] J. E. Manson, R. T. Chlebowski, M. L. Stefanick, M. A. K. Aragaki, J. E. Rossouw, R. L. Prentice, G. Anderson, B. V. Howard, C. A. Thomson, A. Z. LaCroix, et al. The women's health initiative hormone therapy trials: update and overview of health outcomes during the intervention and post-stopping phases. *JAMA: the journal of the American Medical Association*, 310(13):1353, 2013. 1
- [14] National Center for Health Statistics. Cdc wonder online database. http://wonder.cdc.gov/ucd-icd10-expanded.html, 2021. 3
- [15] R. K. Pachauri, M. R. Allen, V. R. Barros, J. Broome, W. Cramer, R. Christ, J. A. Church, L. Clarke, Q. Dahe, P. Dasgupta, et al. 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, 2014. 1
- [16] J. Pearl. Causality. Cambridge university press, 2009. 2
- [17] J. Pearl. 7.1 causal and counterfactual inference. The handbook of rationality, 427, 2021. 3
- [18] M. T. Ribeiro, S. Singh, and C. Guestrin. "why should i trust you?" explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1135–1144, 2016. 2
- [19] D. B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974. 2
- [20] University of Wisconsin Population Health Institute. County health rankings & roadmaps. www.countyhealthrankings.org, 2023. 3

- [21] A. Vallée. Envisioning the future of personalized medicine: Role and realities of digital twins. *Journal of Medical Internet Research*, 26:e50204, 2024. 2
- [22] S. Wager and S. Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Sta*tistical Association, 113(523):1228–1242, 2018. 2
- [23] J. Wang and K. Mueller. The visual causality analyst: An interactive interface for causal reasoning. *IEEE Transactions on Visualization* and Computer Graphics, 22(1):230–239, 2015. 2
- [24] J. Wang and K. Mueller. Visual causality analysis made practical. In IEEE VAST, pp. 151–161, 2017. 2
- [25] J. Wang and K. Mueller. Domino: Visual causal reasoning with timedependent phenomena. *IEEE Transactions on Visualization and Com*puter Graphics, 29(12):5342–5356, 2023. 2
- [26] Y. Zhang, A. Kota, E. Papenhausen, and K. Mueller. Causalchat: Interactive causal model development and refinement using large language models. arXiv preprint arXiv:2410.14146, 2024. 2
- [27] J. Zhou, X. Wang, K.-K. Wong, W. Zhang, X. Liu, J. Zhang, M. Zhu, and W. Chen. Causalprism: A visual analytics approach for subgroup-based causal heterogeneity exploration. arXiv preprint arXiv:2407.01893, 2024. 2