Are Environmental Regulations Working? A Visual Analytic Approach To Answering Their Impact on Toxic Emissions

David Burlinson[1] and Kara Koehrn[2] and Kalpathi Subramanian[1] and Aidong Lu[1]

[1]Dept. of Computer Science, The University of North Carolina, Charlotte, North Carolina, USA
[2]Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention/Toxics Release Inventory Program, Washington, DC, USA

Abstract

Acquisition and analysis of environmental data are essential to the quality, safety and health of human life, and typically consist of heterogeneous, multivariate, and spatio-temporal information collected from different resources. In the United States, the Environmental Protection Agency (EPA) is responsible for protecting human health and the environment; it does so by developing and enforcing regulations to monitor the environment and reduce pollution. In this work, our goal is to obtain a deeper understanding of the effectiveness and impact of EPA regulations; we present preliminary work on an interactive visualization system that attempts to answer specific questions as a function of the dimensions of interest. We demonstrate our system on toxic chemical release reductions resulting from the Clean Air Interstate Rule (CAIR), focusing on the electric utilities sector. We have received positive feedback from our domain expert (toxicologist at the US EPA), indicating that our initial results can more effectively communicate the impacts of environmental regulations, through providing strong visual correlation tools that are currently not available.

Categories and Subject Descriptors (according to ACM CCS): Information Interfaces and Presentation [H.5.2]: —User Interfaces

1. Introduction

In the United States, the Environmental Protection Agency (EPA) is responsible for protecting human health and the environment. One of the key ways it does this is by developing and enforcing regulations that, among other things, set standards, issue permits, monitor the environment, and reduce pollution. These regulations, among other aspects, affect the air we breathe, the water we consume, and the use of harmful chemicals in commerce.

One example of a program at EPA that monitors the use of toxic chemicals is the Toxics Release Inventory (TRI) program [EPAd]. TRI is a mandatory reporting program that gathers detailed data on toxic chemical releases from industrial facilities in a number of sectors, including manufacturing, electric power generation, and hazardous waste management. TRI also collects information on how facilities manage toxic chemicals through recycling, treatment, or burning for energy. The TRI program's goal is to disseminate this data to the public. In the past, the TRI data itself has been an incentive for companies to reduce their releases of toxic chemicals to the environment [KSA11]. However, TRI data also provides a way to measure the effectiveness of other environmental policy and regulations.

The overall goal of this project is to use data visualization to get a better understanding of the effectiveness and impact of environmental regulations on human health and the environment; a deeper understanding could pave the way towards more effective use of regulations and incentives for industry compliance. Specifically, we are interested in using visualization tools to understand the complex interplay between regulations, industry response, and chemical release levels. We believe visualization tools will also allow environmental agencies to see the impacts of their own regulations in an interactive and intuitive way that could provide a more nimble source for data-driven justification for certain types of regulations.

In this short communication, we demonstrate a specific instance of the larger problem: we look at the electric utility industry, which is responsible for a significant part of the toxic chemical releases in the TRI data. Using interactive visual analytic tools, we demonstrate the impact of the Clean Air Interstate Rule (CAIR) and other related regulations on air releases from this industry sector. CAIR was created under the Clean Air Act (CAA) to reduce sulfur dioxide (SO2) and nitrogen oxides (NOX) air emissions from power plants in eastern states (EPA, 2011). We examined chemicals likely to be affected by these regulations during a specific period. Our results show the potential of this approach to understanding the complex relations across these dimensions and their scalability (spatial and temporal).

2. Related Work

The challenge in this work involves analyzing heterogeneous information from different data sources, that are typically multivariate and spatio-temporal. An example is illustrated by Castro et. al. [CS15], where the reductions in Hg emissions from power plants were shown to be strongly correlated with multiple chemicals and other environmental factors. To the best of our knowledge, there are no visualization or visual analytics approaches devoted to study the impact on EPA regulations through combining all the data.

While a number of geo-visualization tools have been developed to explore and analyze the TRI and related datasets, they often concentrate on smaller subsets of data than what is required to comprehensively study the impacts of EPA regulation. The most common tool provided in these systems [EPAa,EPAb,EPAg,KSA] maps TRI chemical usage across the country, where users can search data such as chemicals or facilities with interactive selection methods, by regions (state, city, county or zipcode), or by radius (facilities around a particular location). Some systems allow the integration of several layers, such as the environmental justice and chemical toxicity data in TRI.NET [EPAg]. The Toxic Trends Mapper [KSA] integrates a *Risk Screening Environmental Indicators (RSEI)* measure to the map of facilities and adds health risk related information to TRI data. Another popular approach is a drill-down interface for extracting and comparing data based on a combination of search criteria, such as the Pollution Prevention (P2) tool [EPAc] which collects information related to waste generation and waste management by industrial facilities.

The challenges of visual analytics of TRI style geospatial data also come from complex data relationships spanning across multiple data types and temporal-spatial dimensions [KPSN04, JF06]. Recent visual analytics approaches for geospatial datasets often contain a choropleth map and additional information visualization tools suited for special application requirements. The work by Goodwin et. al. [GDST16] explores geo-visual parameter spaces, comparing multiple variables that are sensitive to scale and geography to varying degrees. Semantics of movement patterns of social media users are studied with an interesting visual analytics procedure based on an uncertainty model [CYW*16]. In addition, such complex combinations of data resources have been explored in visualization applications from other domains, such as social media [CTB*12], geotagged photo collections [KKK*10], climate study [LZM14], movement behaviors [KTE14], and crowd sourced crisis mapping [AIA14].

3. Data and Application Requirements

To study the the impact and effectiveness of air regulations on the environment, and consequently, on public health and safety, we need to analyze heterogeneous data mainly from the following three resources.

TRI. EPA has been collecting TRI data since 1987. TRI data provides information for more than 650 toxic chemicals and chemical categories that are being used, manufactured, treated, transported, or released into the environment. Also included are the facility name, address, location (lat/long); chemical CAS number, regulatory classification, on-site release quantities; publicly owned treatment works (POTW), transfer quantities; off-site transfer quantities for release/disposal, waste management; summary pollution prevention quantities; and scrubber information.

TRI Chemical Groupings. In order to understand the health impact of chemicals, TRI also publishes chemical groupings and their relationship to health impacts [EPAe]. Our application uses this dataset to relate specific categories of chemicals (carcinogens, cardiovascular) to other dimensions of the data.

Scrubber Data. Industrial facilities are required to report information on the type of pollution controls (scrubbers, for instance) they use in complying with environmental regulations. These are reported on the TRI Form R section 7 [EPAf] and are part of the TRI full dataset [EPA11]. Each facility might report a sequence of different treatment methods for a particular chemical, or might combine a number of different waste streams into one treatment process. Each waste treatment reported for TRI chemicals indicates the type of waste stream used (gaseous, wastewater, liquid waste streams, solid waste streams), the particular type of treatment (condenser, scrubber, mechanical separation, oxidation, absorption, evaporation, etc), and the waste treatment efficiency estimate.

3.1. Data Organization

We combine elements of these datasets and organize the data to facilitate the types of queries we intend to perform. We combine the yearly TRI reports on a facility-by-facility basis so that each facility has at least the following attributes: facility name, facility address, lat/long coordinates, industry sector, total yearly usage for each subtype of releases, recycling, treatment and recovery, a list of chemicals and their yearly usage metrics, and a list of scrubbers with their efficiency, identification codes, and waste streams. To understand the data from a chemicals' view, we include the list of chemical grouping and their health impacts. The speed and flexibility of extant APIs for TRI and related data sources are currently not conducive to a responsive interactive application, therefore implementation of such an API and integration of the various data sources into an optimized MongoDB database remain as secondary focuses to facilitate extensibility to a wide range of questions and performant interactions. In its prototypal stage, our application relied upon static files containing parsed data from the TRI data sources. As we continue to iterate upon our design, we are transitioning to a RESTful API to interact with our database.

3.2. Application Implementation

Our tool is a single-page web application built on a MEAN Stack and uses D3.js [Bos] to create all visualizations. Users interact with a GUI to filter the data dimensions and guide the exploratory process, asynchronously obtaining data relevant to their selections. An internal model maps known tasks to specific data sources and attributes. As users interact with the application, relevant visualizations are automatically generated to provide additional perspective to the current task. The visual encoding strategies depend on the shape and nature of the data returned during this process. . Where necessary, users can remove any visualizations deemed irrelevant, and can select a particular visualization to interactively filter and highlight relevant trends.

3.3. Application Requirements

The application requires strong visual analytics tools to make data comparison and correlations in both spatial and temporal dimensions. The final conclusion is often drawn upon the results from a sequence of studies in smaller scales. For example, to address the question of “Are EPA regulations affecting the change in TRI air releases in the electric utilities sector?", we need to study several sub-problems:

1. How are air releases in TRI changing over time for the electric utilities sector? And how does this relate to when EPA regulations were implemented?
2. Is the change from just one facility or from a broader change in that sector happening at many facilities?
3. What chemical or chemicals are driving that overall downward change?
4. Is the change due to a change in production?
5. Is the change due to implementation of pollution controls (like installing scrubbers)? How many scrubbers and when were they installed, especially of chemicals of interest? Does this correlate with the downward change? What EPA regulations does this correspond with?
6. Is the change in air releases due to a change in emissions factors that the facility used to estimate their TRI air releases?

4. Visualization Design

Our visualization design is driven by two goals, (1) we want the visualization to be highly interactive, so that the user can directly query, filter or group on the various dimensions in answering the questions stated above (or some variation of them), and, (2) have means to compare within each dimension (by facilities, facility groups, chemical groups, scrubber types, manufacturing process, etc.). In the long term, the full application will support all of these operations seamlessly to provide the most flexibility in answering specific research questions.

Since the main question here is to study the impact of environmental regulations on emissions over time, our visualization approach uses a combination of temporal plots and a general purpose matrix style visualization (see Fig. 2) that displays specific attributes given a pair of the dimensions of interest.

The matrix metaphor is one of a few visual metaphors users will interact with to explore and filter the TRI data. This panel can prompt the user to refine and direct their exploration. It might show the yearly aggregates of scrubbers, treatment/recycling/release amounts, etc, and this data might be presented for each facility, groups of facilities, for each chemical, chemical categories, states, industry sectors, etc.

Rows can be reordered and regrouped to focus on different elements. Users can then brush or highlight certain subsets of the matrix to drive filtering and modifications in the other panels of the application.

Analysts can explore individual chemicals or subsets of chemicals to examine their overall contributions to the Electric Utilities' air releases. In Fig. 3, we present the results of filtering and examining the top 3 chemicals by emissions. We observe that these top 3 chemicals contribute a significant portion of this sector's air releases, correlating the number of scrubbers with the chemicals most commonly released into the atmosphere. Additional visualization panels (currently in progress) will provide information to either augment the main visualization or provide mechanisms for brushing. For instance, the map panel provides a mechanism for selecting groups of facilities.

5. Results

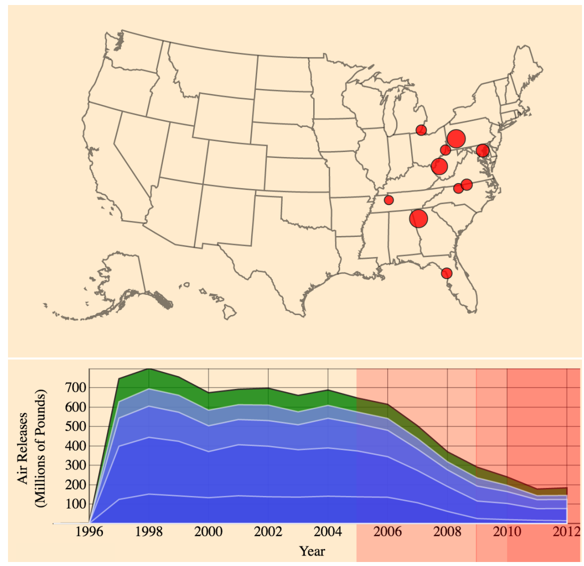

Figure 1: *The top 150 electric plant facilities (by total emissions, in millions of pounds) over time. Upper panel (map) indicates the locations of the top 10 facilities (by total emissions). In the lower panel, the filled regions (bottom to top) shows the contributions of emissions by 10, 50, 100, 150 facilities, in comparison to all 648 facilities (in green). The red rectangular regions indicate the three phases of EPA air regulations, corresponding roughly to years 2009, 2010, 2011. Notice the significant reductions in emissions corresponding to the implementation of the EPA regulations.*

We have begun implementation of a visual analytic system to understand the impact of environmental regulations on the emissions, by various industry sectors. In this work, we focus only on electric facilities. Our application filters the 648 facilities and their related attributes from the original dataset. All visualizations were generated using D3.js [Bos]. We next present the results from the point of view of the questions raised in section 3.3.

Fig. 1 illustrates the variation of toxic emissions (by total emissions over all facilities) over time. The filled bluish plots show (bottom to top) of including respectively 10, 50, 100 and 150 facilities, overlaid on top of the green region that displays the emissions of all facilities. In all, the 150 facilities constitute about 80% of the total emissions of this industry sector. The 3 increasingly opaque red rectangles illustrate the period of the various EPA CAIR regulations: rules promulgated in 2005, phase I cap for NOX in place in 2009 and phase 1 cap for SO2 in place in 2010. It can be clearly seen that the overall emissions (green region) reduction in these

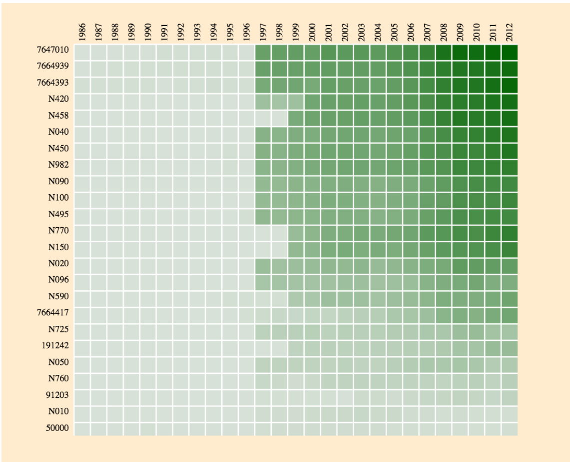

Figure 2: *A matrix style visualization, that illustrates the number of scrubbers used in electric facilities; plot lists the number of scrubbers (light to dark green) for the top 28 chemicals (Y axis) vs. time (1986-2012).*

years, mimicked by the top 10 plants (shown in the lowest panel of Fig. 1). The top panel shows the location of the top 10 facilities (by total emissions) in the US.

Fig. 2 uses our matrix visualization metaphor to illustrate the number of scrubbers used for each of the top 28 chemicals as a function of time (1986-2012). The mapping of the number of scrubbers (left to right) is from light to dark green. The chemicals (Y axis) are sorted by total emissions over all years. Fig. 3 illustrates the contributions of the top 3 chemicals (hydrogen chloride/hydrochloric acid, sulphuric acid/sulphuric acid solution, and hydroflouric acid) to the total emissions (green background). It can be clearly seen that the top 3 chemicals are responsible for almost all of the emissions.

Based on these initial results, referring to section 3.3, we are able to answer questions 1, 2, 3 and 5 at this stage of our implementation.

6. Conclusions and Future Work

This work presents a visual analytics system for studying the effectiveness and impact of environmental regulations, which involves heteroegenoeous spatio-temporal data from multiple sources. We presented the application of this tool by looking at the emissions from the electric utilities sector, and correlated that with the CAIR regulations, viz., a reduction in emissions with the onset and enforcement of the regulations. We plan to continue our work towards developing a complete data query and analysis pipeline that can be integrated with multiple online databases. We will investigate visualization approaches to support flexible data correlation functions beyond existing TRI data visualization and analytics systems. We also plan to collaborate with expert analysts to evaluate our approach on studying the impacts of EPA regulations.

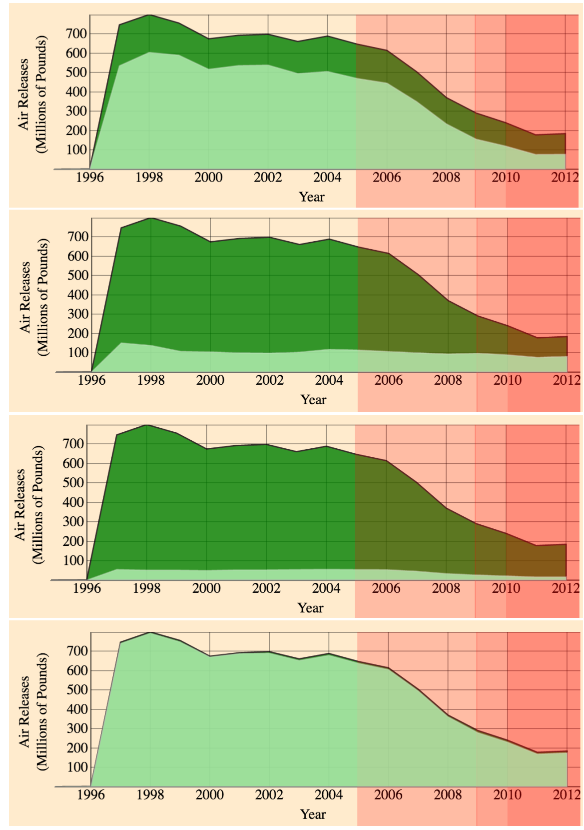

Figure 3: *Toxic chemical emissions(in millions of pounds) over time. The top 3 plots show the contribution of the top 3 chemicals (hydrogen chloride/hydrochloric acid, sulphuric acid/sulphuric acid solution, and hydro-flouric acid) to the total emissions (green overlay). The bottom plot shows the sum of their contributions, dominating the total toxic chemical emissions, as the green area has almost disappeared from the plot.*

References

[AIA14] AMAN H., IRANI P., AMINI F.: Revisiting crisis maps with geo-temporal tag visualization. In *Visualization Symposium (PacificVis), 2014 IEEE Pacific* (2014), pp. 177–184.

[Bos] BOSTOCK M.: d3js. d3js.org, Accessed 2/16.

[CS15] CASTRO M. S., SHERWELL J.: Effectiveness of emission controls to reduce the atmospheric concentrations of mercury. *Environmental Science & Technology* (2015).

[CTB*12] CHAE J., THOM D., BOSCH H., JANG Y., MACIEJEWSKI R., EBERT D., ERTL T.: Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on* (2012), pp. 143–152.

[CYW*16] CHEN S., YUAN X., WANG Z., GUO C., LIANG J., WANG Z., ZHANG X., ZHANG J.: Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *Visualization and Computer Graphics, IEEE Transactions on 22*, 1 (2016), 270–279.

[EPAa] EPA: 2013 tri national analysis. http://www2.epa.gov/toxics-release-inventory-tri-program/2013-tri-national-analysis-introduction.

[EPAb] EPA: My right-to-know application. http://www2.epa.gov/toxics-release-inventory-tri-program/my-right-know-application.

[EPAc] EPA: Pollution prevention (p2). http://www2.epa.gov/toxics-release-inventory-tri-program/pollution-prevention-p2-and-tri.

[EPAd] EPA: "Toxic Release Inventory (TRI) program". http://www.epa.gov/toxics-release-inventory-tri-program.

[EPAe] EPA: Toxic Release Inventory(TRI) chemical groupings. https://inventory.data.gov/dataset/531ede9c-40b9-437f-8742-01ee2d59b39f/resource/9bf438c4-1041-4987-91ef-8aa5a5cc4386.

[EPAf] EPA: Toxics release inventory tri assistance library (section 7). http://www3.epa.gov/trimewebhelp/WebHelp/welcome_to_usepa_s_toxic_release_inventory_assistance_library/welcome_to_toxics_release_inventory_tri_assistance_library.htm.

[EPAg] EPA: Tri.net. http://www2.epa.gov/toxics-release-inventory-tri-program/trinet.

[EPA11] EPA: United states environmental protection agency website. http://www.epa.gov/toxics-release-inventory-tri-program/tri-basic-plus-data-files-calendar-years-1987-2014, 2011.

[GDST16] GOODWIN S., DYKES J., SLINGSBY A., TURKAY C.: Visualizing multiple variables across scale and geography. *Visualization and Computer Graphics, IEEE Transactions on 22*, 1 (2016), 599–608.

[JF06] JERN M., FRANZEN J.: "geoanalytics" - exploring spatio-temporal and multivariate data. In *Information Visualization, 2006. IV 2006. Tenth International Conference on* (2006), pp. 25–31.

[KKK*10] KISILEVICH S., KRSTAJIC M., KEIM D., ANDRIENKO N., ANDRIENKO G.: Event-based analysis of people's activities and behavior using flickr and panoramio geotagged photo collections. In *Information Visualisation (IV), 2010 14th International Conference* (2010), pp. 289–296.

[KPSN04] KEIM D., PANSE C., SIPS M., NORTH S.: Visual data mining in large geospatial point sets. *Computer Graphics and Applications, IEEE 24*, 5 (Sept 2004), 36–44.

[KSA] KRAFT M. E., STEPHAN M., ABEL T. D.: "toxic trends mapper". toxictrends.org.

[KSA11] KRAFT M., STEPHAN M., ABEL T.: *Coming Clean: Information Disclosure and Environmental Performance*. The MIT Press, Cambridge, MA., 2011.

[KTE14] KRUEGER R., THOM D., ERTL T.: Visual analysis of movement behavior using web data for context enrichment. In *Visualization Symposium (PacificVis), 2014 IEEE Pacific* (2014), pp. 193–200.

[LZM14] LI J., ZHANG K., MENG Z.-P.: Vismate: Interactive visual analysis of station-based observation data on climate changes. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on* (2014), pp. 133–142.

———————————————————————-