

TRI-Direct: Interactive Visual Analysis of TRI Data

David Burlinson, Kalpathi Subramanian, Aidong Lu; The University of North Carolina at Charlotte, Charlotte Visualization Center, Dept. of Computer Science; Charlotte, North Carolina, USA

Abstract

In this work, we present *TRI-DIRECT*, an interactive visual analytic system with capabilities to analyze spatio-temporal data for both professional and novice users. The system is motivated by the Toxic Release Inventory (TRI) program of US Environmental Protection Agency (EPA) and its associated datasets; the TRI program was created in 1986 to track toxic chemical usage, which includes release, recycling, treatment and recovery, and its impact on the environment. The design of the system is motivated for ease of use and its future transition to mobile platforms, so as to have the widest possible impact across users possessing a range of skills/interests. We describe *TRI-DIRECT*'s design, implementation and capabilities, and present two detailed use cases with the system, (1) Texas vs. Louisiana's usage and processing of toxic chemicals, and, (2) comparing an urban region, Raleigh, NC, vs. the state of North Carolina.

Introduction

The Toxic Release Inventory (TRI) program was created in 1986 to track the management of certain toxic chemicals that can potentially threaten human health and the environment. The program was created partly in response to earlier industrial disasters of toxic chemical releases. Managed by the Environmental Protection Agency (EPA), the TRI program's goals of improved environmental protection and public health safety are based on disseminating toxic chemical release information as widely as possible into the public domain. While facilities that meet TRI criteria are required to report to the TRI program, the public access to the TRI data provides a strong incentive for improved environmental performance [6]. Data on over 650 chemicals have been submitted by roughly 47,000 US facilities to the TRI program.

Over the past two decades, a number of spatio-temporal analytical tools have been developed to explore and analyze the TRI and related datasets. The TRI National Analysis tools [5] provide maps of TRI chemical usage across the United States. Tools are provided to search by state, city, county or zipcode, for more detailed information. My Right To Know (MyRTK) [7] is another tool that addresses a critical application of the TRI program: the ability to search for facilities around a specific location. This is important for people who might be considering moving to a particular neighborhood, locating new businesses, etc. TRI.NET [9] is an application that also allows search by radius for facilities around a particular location. It also has the capability of integrating additional layers, for instance, environmental justice, chemical toxicity data. The Pollution Prevention (P2) tool [8] collects information related to waste generation and waste management by industrial facilities. The P2 tool reports on chemical reductions for selected industry, facility, or state. Data for individual facilities can be extracted via drill-down operations, and comparisons made between facilities. The Toxic Trends Mapper [19] integrates

a Risk Screening Environmental Indicators (RSEI) measure to the map of facilities, and adds health risk related information to TRI data.

Closely related is the research on *geo-visualization and geo-visual analytics*, which has been employed to analyze a broad range of socio-economic data. Researchers from both geology and visualization fields have developed a large number of approaches to represent various geo-spatial datasets. The core of geovisualization has been the generation of data maps [18, 1] and the use of cartograms for displaying geo-statistical information [15]. Panse et al. [20] combined cartogram layouts with PixelMaps for exploring dense geo-spatial data. The Wordmapper project [4] used cartograms to depict comparisons between different territories for communicating quantitative geographic data. Hotmap [11] exploited the mapping system's imagery pyramid to overlay a heatmap of the log files over the original maps. For analysis of spatial point patterns, like the TRI data, bivariate kernel densities have also been applied to estimate the chemical distribution of the site locations with a cloud-like visualization [14]. Approaches for spatial point visualization were developed by Wong et. al. [21] for visual analytics of geographically-based electric power grids.

The challenges of visual analytics of TRI style geo-spatial data come from complex data relationships spanning across multiple data and spatial dimensions, and time [16]. Recent visual analytics approaches for geo-spatial datasets often contain a choropleth map and additional information visualization tools suited for special application requirements. BirdVis [10] was developed to analyze bird distribution models to identify relevant correlations and patterns. Kim et al. [17] developed Bristle Maps for aggregation, abstraction, and stylization of spatiotemporal data that enabled multi-attribute visualization, exploration, and analysis.

In this work, we present *TRI-DIRECT*, a visual analytics system for spatio-temporal datasets. We demonstrate its application for analyzing the trends of chemical releases, recycling, recovery and treatment levels from related facilities, using the TRI dataset as our primary example. The goals of this work are two-fold: (1) build an interactive visualization application that is extremely easy to use and can cater to *users possessing a range of skills and experience*, and (2) make the application adaptable to use on mobile and touch style displays/platforms, so that the application can be deployed in public spaces, such as museums, libraries, and other venues of public interest, and across a range of computing devices. A key factor in the design is to keep even novice users engaged through an intuitive interface.

Current EPA and other data exploration tools are multi-page web interfaces or traditional software packages. Our experience in experimenting with many of these TRI tools has been mixed; they provide valuable information with some effort, however, the interfaces need considerable improvement for larger public use.

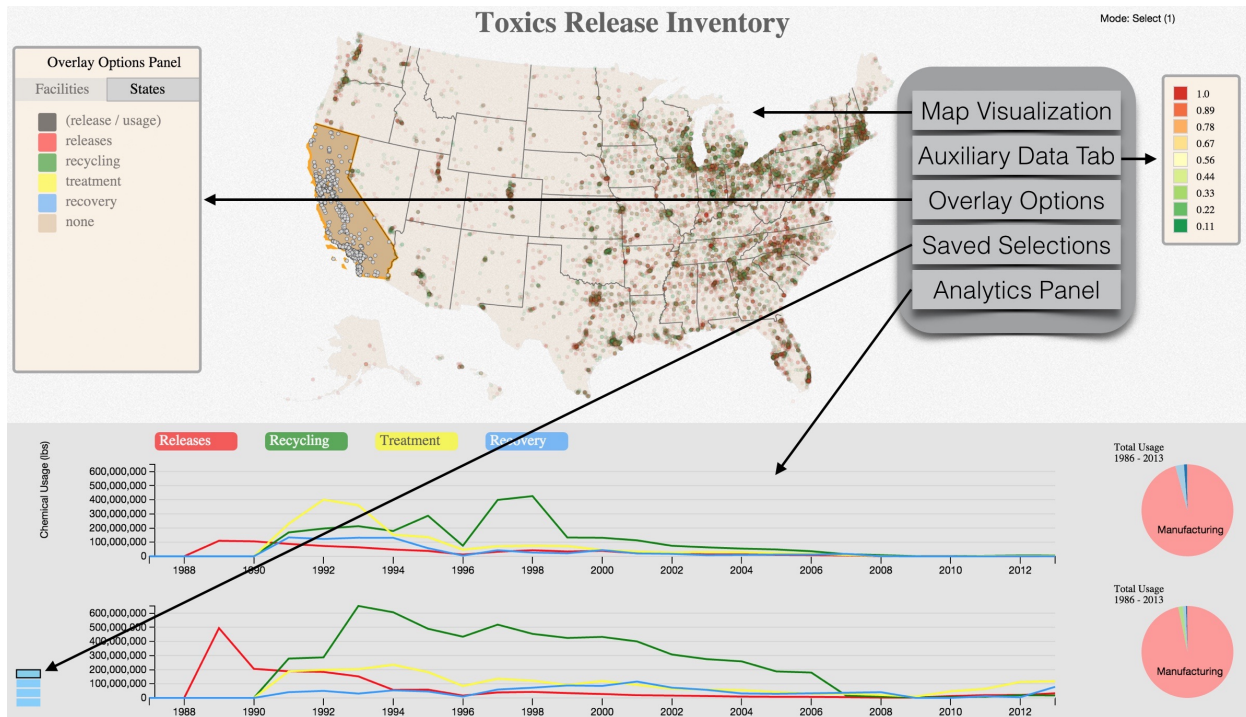


Figure 1: TRI-DIRECT interface: The interface consists of two main components, (1) *Map*, which is the primary area through which the user interacts and make selections of subsets of the data for further analysis, and (2) the *Analytics Panel*, which contains charts and graphs relevant to the user's selections. In addition, there is an overlay options panel for modifying visual properties and an auxiliary data tab that provides pertinent or quantitative information relating to the selections.

Our goal is to engage novice users of all ages and educational backgrounds, so as to comprehend high level TRI information (map level visualization, TRI in their own community). Our approach incorporates data exploration and visualization features into a single application (as seen in Fig. 1), thereby simplifying navigation, rather than stepping through a potentially confusing sequence of pages and visualizations. Our interface makes the process of comparison and exploration much smoother, as users can seamlessly create selections, examine resultant visualizations, and refine parameters with a series of intuitive interactions. In the following sections, we describe *TRI-DIRECT*, its design, features and its application to the TRI data. We also present two use cases: a state level comparison between Texas and Louisiana, and comparing a large urban region (Greater Raleigh) to the rest of the state of North Carolina.

TRI Dataset

The TRI datasets contain about 102 attributes covering the categories of facility information, chemical identification/ classification, on-site release quantities, treatment transfer quantities (public works and off-site), and pollution prevention quantities. The dataset we use in our application contains relevant attributes from all the TRI datasets released since the program began in 1987, including latitude and longitude values, facility names, NAICS and SIC codes to identify facilities by industry sector, and aggregations of total release, recycling, recovery, and treatment of toxic chemicals (in pounds) for each year. As the data is based on self-reporting by each facility, the reports for each year may be incomplete or contain gaps in the reporting periods. Close to

51,300 facilities have reported toxic chemical usage since 1987. Of these, around 4400 facilities were removed from the dataset for having no relevant usage reported in any of the available years or containing inappropriate values for some of the aforementioned attributes. Approximately 1000 of the facilities in the trimmed dataset were not able to be identified by industry sector.

Visualization Design

Our goal was to build a highly responsive and intuitive application with the potential to quickly yield meaningful insights into toxic chemical usage, processing or activities in any geographical region. A second goal was to make the tool applicable to users with a range of skills, experience and expertise to improve their understanding and the resulting environmental impacts. As such, we have been careful to design the system interface around fundamentally simple interactions, without sacrificing detail of data or depth of potential exploration that would be useful for more experienced users. Finally, our interface considers principles of visual perception for layout design and choice of visual metaphors so that we present the TRI data in a clear, salient, and accessible way.

Layout Design.

Fig. 1 illustrates the overall layout of TRI-DIRECT. The visualization layout and the interface to our application is divided into two main interactive components: a map visualization and an analytics panel. The map is the primary area of the screen with which the user interacts, providing a *direct* interface to the usage data. The analytics panel contains charts and graphs relevant to

the users' selections, which are built as the user interacts with the map visualization. In addition to these main regions, our tool features an overlay options panel to give users control over the map's visual properties and an auxiliary data tab to display pertinent information and raw values related to the selections. The layout design tries to balance a simple easy-to-use interface that can accommodate a range of users, using a drill-down approach to more detailed information and analysis; novice users might choose to look at high level information - national, state or their own region of interest, in terms of facilities or chemical usage, while expert users can drill down to perform more detailed and comparative analysis across a range of scales.

Color Choices.

Our application uses specific color schemes to (1) maintain an effective and organized representation of the data, (2) consistency with the underlying domain and related tools and, (3) wherever possible, use perceptually proven color scales to optimally represent the structure of the underlying data with perceptual dimensions of color. We represent the four chemical usage variables to be consistent with other EPA tools [8]: chemical releases are in red, recycling is green, treatment is yellow, and recovery is blue. Maintaining color uniformity across the visual metaphors and data panels of our tool is important for minimizing confusion and expediting the process of interpreting the visual values presented. Our tool also maps usage variables and other derived variables to color scale overlays as part of the spatio-temporal analysis. We rely on sequential, diverging, and qualitative ColorBrewer [3, 13] scales to accurately represent the nature of each overlay's data.

Visualization Features

As described earlier and illustrated in Fig. 1, the main application interface is via spatial selections that trigger updates to the analytics components of the interface. As such, our system, though currently developed specifically to analyze the TRI dataset, is a tool that permits spatio-temporal analysis of multivariate datasets. The primary interaction operations begin in the spatial domain and via a user-defined brushing operation.

Brushing

Our application supports two levels of brushing operations: in 'Selection' mode, a user can select individual facilities and individual or groups of states (this will encompass all the facilities in each selected state). This mode permits comparisons between specific states or groups of states, and similarly between facilities. In the 'Brushing' mode, a more traditional selection is supported, allowing the user to select a region of interest (currently limited to rectangular regions); this will simply integrate the data across the facilities that are inside the selection region. Thus, a quick comparison between toxic releases of facilities in two different regions of the country can be performed using this type of selection. Fig. 2 shows an example of selecting regions surrounding Lake Michigan.

Multi-scale Visualization

TRI-DIRECT provides spatio-temporal visual exploration of the multivariate data at different scales; for the TRI data set, exploration can be at the national, state or facility level. Orthogo-

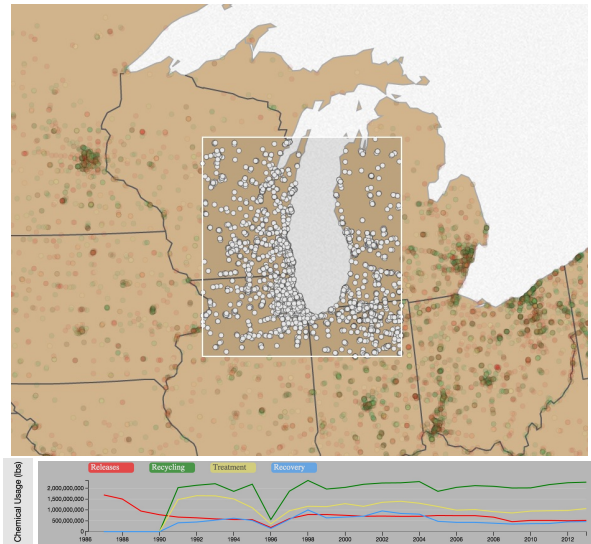


Figure 2: Brushing a region. Resultant line graph showing usage variables (release, recycling, treatment, recovery) for all facilities in the brushed region.

nally, at any of these levels temporal filters can restrict the selection. At the lowest levels, comparative analysis can be performed to explore or compare activities at individual facilities.

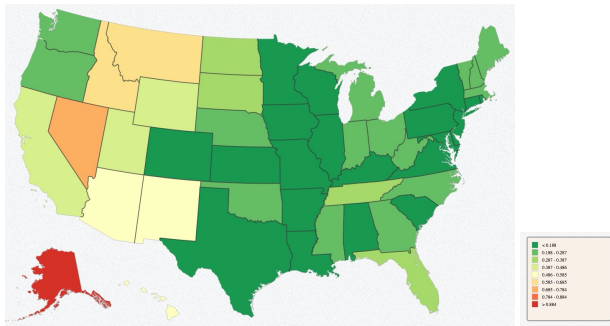
Level 0: Map Level.

The map metaphor in TRI-DIRECT is an Albers projection of the United States with state boundaries clearly defined, and is the only interface element displayed when the application is loaded. The map region supports panning and zooming to allow users to focus on whichever segment of the spatial region is most relevant to their current exploration. In this preliminary implementation, we have not integrated the traditional street map structure in our system in order to maintain the simplicity of our interface; we do recognize that users may benefit from overlays such as terrain, watersheds, cities, and roads during the exploration process. These can be brought in as additional layers into the application on an application specific basis, thereby not computationally overburdening the application, which will sacrifice interactivity.

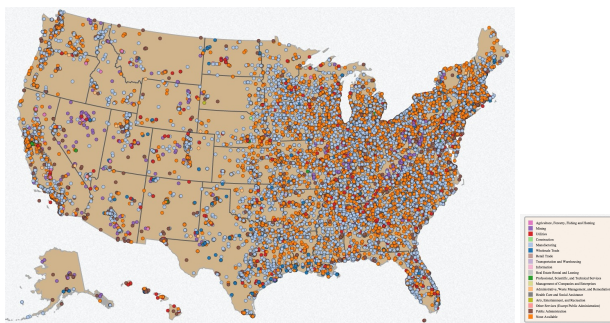
With the overlay options panel (see Fig. 1) we can overlay the geographical region (map, in this case) with different metrics to reflect high level spatial assessments of toxic chemical usage. Each attribute can be individually visualized; in addition we also support a ratio metric that represents the ratio of the toxic release to the sum of the total chemical usage. This is seen in Fig. 3a.

Level 1: State Level.

A user may also select a state or a group of states. When a state is selected, all the facilities in that state become part of the selection, and their accumulated release, recycling, treatment, and recovery values are added into the current selection. All other states and their facilities fade out to highlight the selected state(s), and the analytics panel is updated to display a multiline graph for the current selection's usage variables. In Fig. 4 a group of states has been selected in the Great Lakes region. The temporal plots below show the aggregate of the usage variables for these states.



(a) Ratio Overlay. Ratio is defined as $\frac{Release}{Release+Recycling+Treatment+Recovery}$



(b) Industry Sector Overlay.
Figure 3: Map Overlays.

Level 2: Facility Level.

In the facility level, users can overlay facilities on the map to explore *usage variables*, examine geo-spatial distribution, and draw inferences about the performance of an individual facility, facility group or facilities covering a region. Each facility is represented by a small colored circle projected onto the map using the reported latitude and longitude positions. Only facilities that reported *usage variables* to the TRI at some point are included in this overlay, so once we have trimmed out any facilities that do not report any relevant values, we are left with a little under 47,000 facilities. Hovering over a facility causes the auxiliary data tab to update, displaying germane attributes such as facility name, raw values for *usage variables*, location, etc. Clicking on a facility functions are similar to clicking on a state: the current selection is modified to contain the facility’s usage variables and the analytics panel updates to reflect the new selection.

The color attribute is the main variable used to highlight different trends among facilities and across the map. Research in visual perception indicates that humans have no trouble abstracting information from sets of objects grouped spatially or by features such as color. In fact, we can ‘preattentively’ estimate higher-order statistics such as numerosity and average value with reasonable accuracy and reliability, especially when given plenty of time to examine the display [12]. Our tool takes advantage of this ‘ensemble coding’ when applying various color scales to the facility representations; users can gain insight about the categories or relative values applied in any given scale regardless of the number of facilities represented. Finally, users can examine various color overlays for the facility representations by exploring the options in the overlay options panel.

Visual Analysis

Analytics Panel.

The analytics panel fills the lower portion of the application’s interface and contains interactive line graph and pie chart visualizations that dynamically update to represent the user’s current selection. The line plots represent a temporal variation of the *usage variables* for all the years for which TRI data is available, or a user selected period (using brushing). The line graphs are drawn as colored lines for each variable. The pie chart is populated with segments representing either the same variables or a breakdown of the selection’s facilities by industry sector. These visual metaphors complement each other well; the line graph allows examination of the releases, recycling, treatment, and recovery over time, whereas the pie chart presents aggregates of these values for quick, salient comparison between variables.

Comparative Analysis:

Since one of the main goals of this visualization tool is to facilitate comparison between facilities, regions, states, and groups of states, the tool allows users to save selections. Saving the current selection takes a snapshot of the facilities, states, or regions involved, as well as all the relevant usage variables. A saved selection causes a new line graph and pie chart to be populated directly beneath the original visualizations in the analytics panel. Successive interactions with the map metaphor update the current selection, which in turn modifies the upper row of visualizations in the analytics panel. During user interaction, the lower row is either empty or representing a saved selection. Any time either graph is

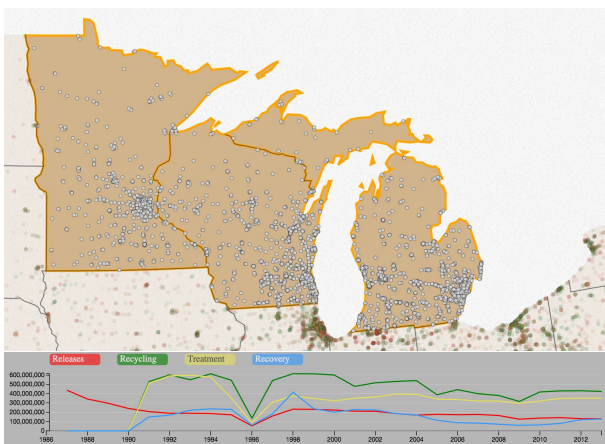


Figure 4: Multi-scale Visualization: State Group. A group of states in the Great Lakes region selected; the temporal plots below indicate the variation of the 4 usage variables over time.

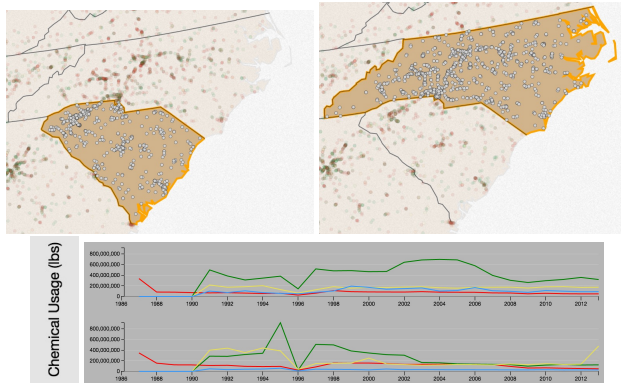


Figure 5: Comparative Analysis of South and North Carolina. South Carolina is the upper line graph and North Carolina is the lower graph.

given new data to display - by building or modifying the current selection or by selecting a saved selection - both line graphs are scaled to the larger of the two ranges. This facilitates an accurate comparison between the usage variables of both selections.

Saved comparisons are controlled by a stack of rectangular buttons at the left edge of the analytics panel. When a selection is saved, it is added to this stack, and users can click on any of the buttons to repopulate the relevant visualizations with the saved comparison selection data. Whenever a saved selection is reloaded, the relevant state(s), facility, or brushed region are highlighted on the map.

Finally, the analytics panel also supports temporal brushing, i.e. the comparisons between two selections can be further constrained in time, as can be seen in Fig. 6 where the shaded regions represent the temporal brushing in the top panel, while the bottom panel only shows the recovery variable across a short period. The pie-charts on the right are immediately updated, depending on the current temporal subset that is selected by the user.

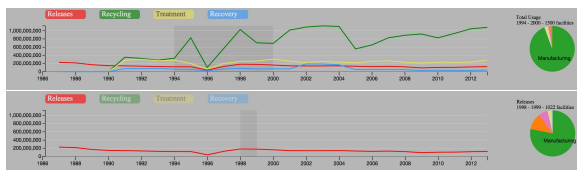


Figure 6: Comparative Analysis of Illinois showing (1) all usage variables and (2) just releases, both brushed for a subset of years. The pie charts represent the breakdown of industry sectors contributing to each graph's selected usage variables during the defined span of time.

Results

Use Case 1: Texas vs. Louisiana

Using the state overlay options panel, we can quickly paint a high-level picture of chemical usage across the country. As illustrated in Fig. 7, upon applying and examining the overlay options for each of the usage variables and the combinations thereof, it becomes apparent that a subset of states are involved in a large amount of chemical usage, as can be seen in Fig. 7 that displays the 4 variables for all states. In terms of volumes of toxic releases, Texas, Nevada, Alaska, California, Ohio, and Louisiana are the frontrunners, all categorized into one of the darker color buckets.

Recycling is dominated by Texas, with Louisiana and Illinois not far behind. The treatment overlay shows Texas and Louisiana at a much darker color than all the rest of the states to an even greater extent than the recycling overlay, and recovery is even more disproportionate than all the other options. It is clear that Texas has the highest chemical usage by a significant margin, and Louisiana has been prominent in a few of the overlays, so it is no surprise that these two are colored far darker in the overlay option for total usage.

Adjusting the upper boundary for the ColorBrewer scale causes the states' corresponding usage variables to be re-binned into the discrete color slots for the updated scale. This way, even if a particular state skews the scale's domain (looking at Texas, for instance), differences between states with less usage can be teased out and examined.

Texas and Louisiana's chemical usage stood out in our initial examination, so let's dig a bit deeper into their performance over time.

Texas:

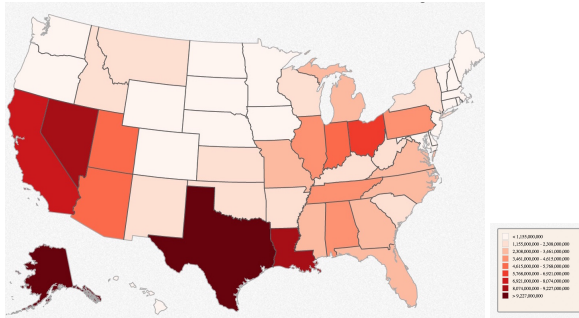
Texas has the highest overall values in every single category of the toxic chemical usage, since TRI began collecting data in 1987. Looking at the line graph in the analytics panel in Fig. 9, we see that Texas' facilities reported releases more than 2.4 billion pounds of chemicals in that first year alone. Releases were down below 1 billion pounds in the next year, and have been below 450 million pounds every year since 1990. While the rate of release is still quite high, the amounts of treatment, recycling, and recovery significantly outweigh releases after 1990. In fact, Texas is among the best states in the country when considering the proportion of releases to total usage, as seen with the ratio overlay in the overlay options panel (Fig. 3a). Chemical usage remained fairly steady between 1990 and 2013, with treatment fluctuating between 600 million and 2.5 billion pounds, recycling between 500 million and 1.5 billion pounds, recovery between 400 million and 1 billion pounds, and releases between 80 and 400 million pounds. Texas' reported treatment in 2004 was its highest usage variable for any year at around 2.5 billion pounds.

Louisiana:

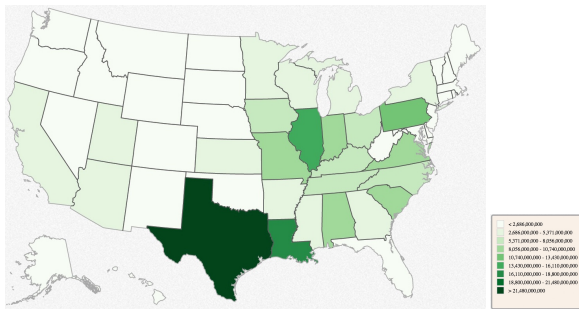
Louisiana's release values parallel Texas' closely across all 27 years. Louisiana started in a similar fashion to Texas, with releases the only form of usage reported in 1987 at just over 1.7 billion pounds. Neither Louisiana nor Texas reported any recycling in the first four years, then maintained a fairly consistent rate between around 500 million and 1.3 billion pounds per year. Both states reported a similar degree of consistency for chemical recovery, again starting in 1990, with Louisiana recovering between 200 and 330 million pounds per year and Texas recovering generally between 3 to 4 times as much. Louisiana's treatment spiked anomalously in 2000, jumping from around 1.6 billion pounds in 1999 to over 4.3 billion pounds before returning below 1.7 billion pounds for the remainder of the reported years.

Comparative Analysis: Texas vs. Louisiana

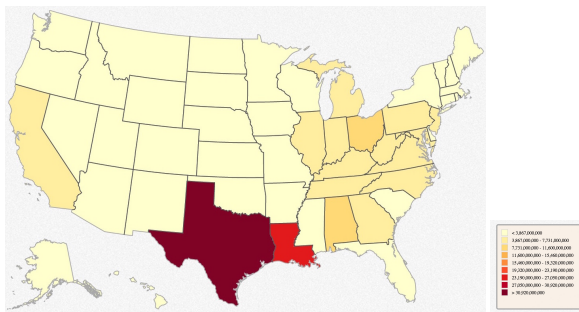
Judging by the line graphs in the analytics panel, Texas and Louisiana are quite comparable in terms of consistency, values, and trends over time for each usage variable, aside from Louisiana's burst of reported chemical treatment in 2000, Texas'



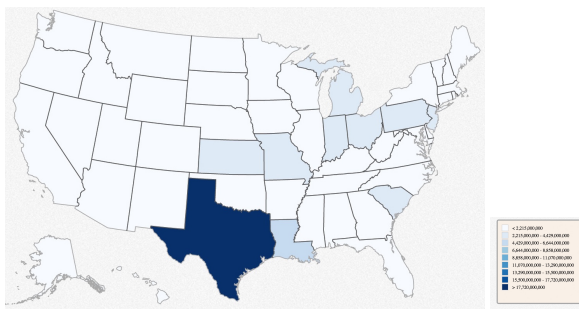
(a) Toxic Releases.



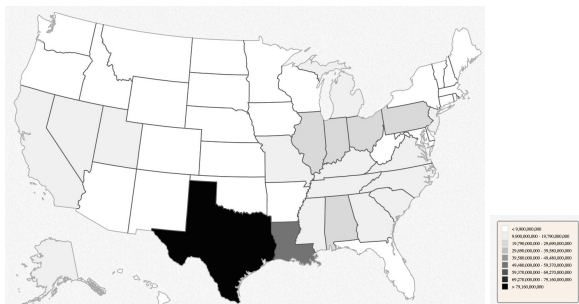
(b) Recycling.



(c) Treatment.

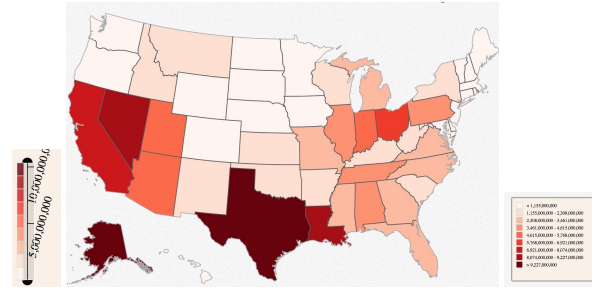


(d) Recovery.

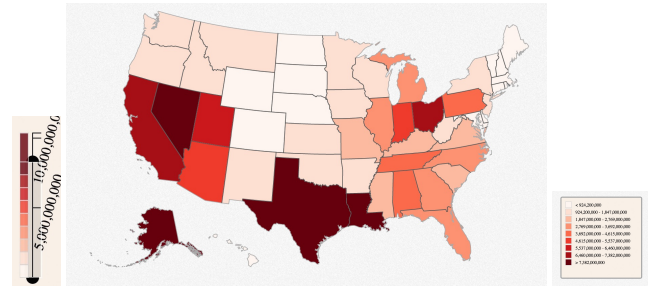


(e) All Usage.

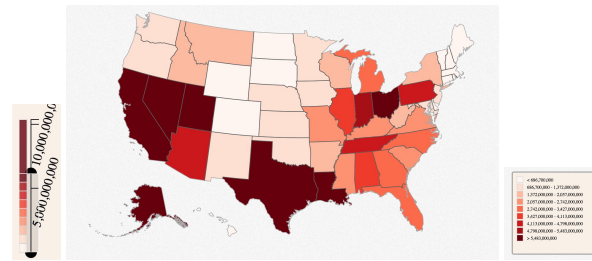
Figure 7: State overlays for each usage variable across the USA. These plots give a quick overview of the states that dominate for the respective usage variable. All values are in pounds.



(a) Range: 1.1-9.2B lbs



(b) Range: 0.9-7.4B lbs



(c) Range: 0.7-5.5B lbs

Figure 8: State overlays for releases with different user-defined ranges(in billions of pounds) mapped to color. Notice clamping the low and high ends of the range brings a better distribution across a large number of states, and highlights states with lower values.

smaller spike of treatment in 2004, and Louisiana's slightly lower amounts of recovery across the board. It is interesting to note that despite the similarity of the reported values for chemical usage in both states, the auxiliary data tab indicates that Texas has approximately 2900 facilities, more than 4 times more than Louisiana's approximately 650 facilities. While Texas is responsible for slightly higher usage variables than Louisiana, we can conclude that the latter's facilities are, on a rough average, each responsible for more chemical usage. The chemical usage is explained similarly within each state, as shown by the respective pie charts in the analytics panel: over 95% of both states' data are attributed to facilities from the manufacturing industry. Around 89% of Texas' facilities and 85% of Louisiana's facilities are categorized as manufacturing industries, as shown in the auxiliary data tab when either state is selected, so this result is not surprising. The industry sector with the second-highest contribution to the usage variables is Utilities, which accounts for between 5% and 10% of Texas' releases and between 1% and 3% of Louisiana's chemical releases reported to TRI.

Texas' facilities are about 6.4% of all reporting facilities in the country (2969 out of 46,542 facilities, according to the analytics panel), yet are shown by the auxiliary data tab to be responsible for about 15.4% of the overall data when the line graphs are brushed for the entire range of years. Louisiana represents a smaller 1.4% of facilities in the country (686 out of 46,542), yet accounts for close to 9.5% of usage data. Between these two states, around 8% of the country's facilities release, recycle, treat, and recover almost 25% of the overall chemical usage.

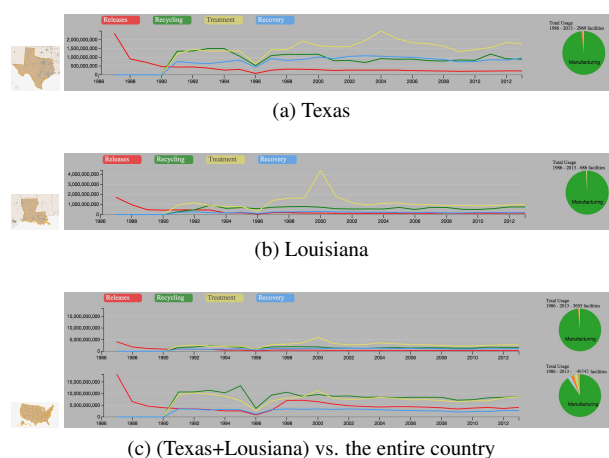


Figure 9: Toxic Chemical Usage: Comparative analysis of Texas and Louisiana

Use Case 2: Greater Raleigh vs. North Carolina

Our second case considers chemical usage at a lower level than state-by-state comparisons, choosing instead to examine brushed regions corresponding to a few cities and their metropolitan areas. These selections are not quite as straightforward as those from the previous case. State color overlays, selection, and comparison are easy single-click actions, but selecting a desired region requires brushing, which involves a little more finesse when the selection involves specific boundaries and irregularly-shaped areas. On the other hand, brushing allows quick selections of two regions of the country (for instance, the northeast vs. the California region) without going through each of the states in these regions. Our tool currently supports brushing rectangular regions; future versions of the application will support more sophisticated selection tools, as well as pre-defined regions considered to be important. Creating and comparing brushed regions of facilities provides a comparable degree of insight to the state comparison; areas can easily be characterized with respect to the usage variables and industry sectors, and compared with other regions of the country.

For this exercise, we examine some of the major cities in North Carolina, beginning with its capital, Raleigh. Figure 10 illustrates the comparative graphs between Raleigh and the state of North Carolina. Raleigh and its surrounding regions contain somewhere around 100 unique facilities that have reported toxic chemical usage since TRI began collecting data. Facilities in Raleigh reported only chemical release until 1990, and then reported all four usage variables steadily in successive years. Releases ranged between 1 and 5 million pounds per year prior to 1998 and remained around 500,000 pounds for the rest of the

years in the TRI reports. Treatment levels increased gradually from just under 4 million pounds in 1990 to a sudden relative maximum of around 107 million pounds in 2000, then returned to between 4 and 5 million pounds through 2013. Recovery stayed between 1 and 2 million pounds per year from 1990 all the way through the yearly TRI reports. By far the most interesting usage variable for facilities in Raleigh is recycling, which was reported at 3 to 5 million pounds per year from 1990 to 1995, at which point it spiked anomalously to almost 600 million pounds before dropping back below 5 million pounds from 1996 to 2013. As can be seen in Fig. 10, this spike of activity dominates the line graph for the region. Examination of the state as a whole reflects this trend for recycling in 1995, at which point North Carolina's recycling values tripled from around 300 million to over 900 million pounds, the highest yearly amount for any usage variable in the state for any of the TRI reports. Raleigh's facilities explain a significant portion of this increase, contributing more than two thirds of the state's overall recycling that year. Brushing some of the state's other major cities, particularly those with many facilities such as Charlotte and Greensboro, doesn't really shed much light onto the anomalous results of 1995. The line graph based on the general Charlotte region appears to show recycling as the highest usage variable for most of the years TRI collected data, but no usage variable exceeds 25 million pounds in any given year. Greensboro's facilities also perform chemical recycling far more than any other form of usage, but only exceed 100 million pounds in 1991 and remain between 25 and 45 million pounds per year through 2013. Returning to Raleigh and the surrounding facilities, closer examination shows that a single facility accounted for more than 90% of the region's recycling values in 1995; a facility called 'Cargill Inc Raleigh Facility' recycled 590 million pounds, according to the line graph. This single facility contributed over 64% of North Carolina's recycled chemicals in 1995. It and the other facilities around Raleigh contributed almost 44% of the state's chemical usage in 1995 and just under 12% of the state's usage across all the years of TRI reports.

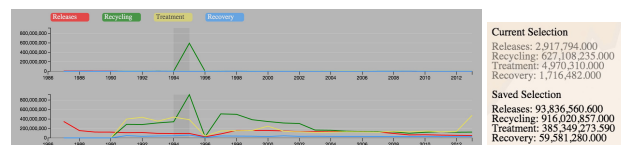


Figure 10: Comparative analysis: Greater Raleigh region vs. North Carolina. The selection is brushed to show values from 1995, which displays an anomalously large recycling, and the corresponding raw values that appear in the auxiliary data tab.

While the brushing technique in our tool does not currently provide options for a more granular selection than a simple rectangle region, the level of detail is still certainly sufficient to characterize the performance of selections, investigate usage variables and their change over time, and drill deeper into noteworthy or interesting segments of time. By applying state and facility color overlays via the overlay options panel, exploring the geo-spatial interface, and refining selections to examine a particular usage variable or segment of time, users with any range of experience can quickly and easily navigate the TRI data and derive a deeper understanding of the toxic chemical usage in America, between states and facilities, and in their home town.

Implementation

TRI Direct is built using web technologies: HTML, CSS, Javascript, and SVG - brought together with D3 [2], a Javascript library commonly used in visualization applications. Many of the tool's visual elements such as the Albers USA projection, line and pie charts, and brushing functionality are based on D3 examples that have been extended and modified for our particular use cases and interactivity goals. The simplicity and power of D3 facilitates rapid prototyping of interface elements, allowing us to determine reasonable and useful components by playing with the data in various ways. The facility and usage data is parsed from the original TRI csv files and combined into a single JSON file with attributes relevant to this application. The JSON data is bound to DOM elements with D3 when the application loads, and every visual representation of the data either directly reads or aggregates properties of these elements, and maps the resulting values to visual elements.

Conclusions

We have presented TRI-DIRECT, a visual analytic application to explore spatio-temporal datasets. While the tool design has been motivated by the TRI dataset, the design of the system is generally applicable to similar spatio-temporal datasets. We have taken care in designing the interface so that high level information is easily accessible to novice users, while drill-down operations are available for deeper insights and for more experienced users. We have illustrated two use-cases using TRI-DIRECT, (1) comparing Texas and Louisiana, and (2) contrasting a few large urban areas within the state, viz., the greater Raleigh region in relation to the state of North Carolina. In each case, we walk through the (fully interactive) analytic steps to understanding the activity of facilities in these regions across the 4 usage variables: toxic releases, recycling, treatment and recovery.

We are in the process of beginning a year-long project with the EPA through the TRI Challenge; as part of this effort, we will be benefiting from direct feedback on TRI-DIRECT and associated tools that we will disseminate as part of the project. Future work on this project will involve deployment of the system to both novice and expert users, and adapting the system to mobile platforms, include devices with touch interfaces (tablets, smart phones). Implementation will also focus on performance and scalability, especially as we consider bringing in additional layers of data to correlate TRI data with other dimensions such as environmental policies, health risk data, etc.

References

- [1] *Exploring Geovisualization*. Elsevier Science, 2005.
- [2] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011.
- [3] C. Brewer. Colorbrewer, a web tool for selecting colors for maps. <http://www.ColorBrewer.org>, Accessed 8/15.
- [4] D. Dorling, A. Barford, and M. Newman. Worldmapper: The world as you've never seen it before. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):757–764, Sept 2006.
- [5] EPA. "2013 tri national analysis". <http://www2.epa.gov/toxics-release-inventory-tri-program/2013-tri-national-analysis-introduction>.
- [6] EPA. "learn about the toxics release inventory". www2.epa.gov/toxics-release-inventory-tri-program/learn-about-toxics-release-inventory.
- [7] EPA. "my right-to-know application". <http://www2.epa.gov/toxics-release-inventory-tri-program/my-right-know-application>.
- [8] EPA. "pollution prevention(p2)". <http://www2.epa.gov/toxics-release-inventory-tri-program/pollution-prevention-p2-and-tri>.
- [9] EPA. Tri.net. <http://www2.epa.gov/toxics-release-inventory-tri-program/trinet>.
- [10] N. Ferreira, L. Lins, D. Fink, S. Kelling, C. Wood, J. Freire, and C. Silva. Birdvis: Visualizing and understanding bird populations. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2374–2383, Dec 2011.
- [11] D. Fisher. Hotmap: Looking at geographic attention. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1184–1191, Nov 2007.
- [12] M. Gleicher, M. Correll, C. Nothelfer, and S. Franconeri. Perception of average value in multiclass scatterplots. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2316–2325, Dec 2013.
- [13] M. A. Harrower and C. A. Brewer. Colorbrewer.org: An online tool for selecting color schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003.
- [14] C. houh Chen, W. K. Härdle, and A. Unwin. *Handbook of Data Visualization*. Springer Science & Business Media, 2007.
- [15] D. Keim, S. North, and C. Panse. Cartodraw: a fast algorithm for generating contiguous cartograms. *Visualization and Computer Graphics, IEEE Transactions on*, 10(1):95–110, Jan 2004.
- [16] D. Keim, C. Panse, M. Sips, and S. North. Visual data mining in large geospatial point sets. *Computer Graphics and Applications, IEEE*, 24(5):36–44, Sept 2004.
- [17] S. Kim, R. Maciejewski, A. Malik, Y. Jang, D. Ebert, and T. Isenberg. Bristle maps: A multivariate abstraction technique for geovisualization. *Visualization and Computer Graphics, IEEE Transactions on*, 19(9):1438–1454, Sept 2013.
- [18] M.-J. Kraak and F. Ormeling. *Cartography: Visualization of Geospatial Data*. Routledge, 2009.
- [19] M. E. Kraft, M. Stephan, and T. D. Abel. "toxic trends mapper". toxictrends.org.
- [20] C. Panse, M. Sips, D. Keim, and S. North. Visualization of geospatial point sets via global shape transformation and local pixel placement. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):749–756, Sept 2006.
- [21] P. C. Wong, K. Schneider, P. Mackey, H. Foote, G. Chin, R. Gutromson, and J. Thomas. A novel visualization technique for electric power grid analytics. *Visualization and Computer Graphics, IEEE Transactions on*, 15(3):410–423, May 2009.