

Exploring Rural Shrink Smart Through Guided Discovery Dashboards

Denise Bradford

University of Nebraska - Lincoln
Lincoln, Nebraska
denise.bradford@huskers.unl.edu

Susan VanderPlas

University of Nebraska - Lincoln
Lincoln, Nebraska
susan.vanderplas@unl.edu

Abstract

Many small and rural places are shrinking. Interactive dashboards are the most common use cases for data visualization and context for exploratory data tools. In our paper, we will explore the specific scope of how dashboards are used in small and rural area to empower novice analysts to make data-driven decisions. Our framework will suggest a number of research directions to better support small and rural places from shrinking using an interactive dashboard design, implementation and use for the every day analyst.

Keywords: Interactive Dashboards, Exploratory Data Analysis (EDA), Guided Discovery

1 Research Problem

With the amount of publicly open-source data, a proliferation of visualization dashboards has increased in nearly every industry (Sarikaya et al., 2019). A dashboard in its fundamental form, a dashboard supports a way of presenting and making sense of complex data to better enable and support decision making.

Some communities continue to thrive as they lose population because they adapt, maintaining quality of life and community services for residents while investing in the future. This process, *smart shrinkage*, is important for rural areas who have experienced shrinking populations for decades. As small rural towns do not have access to data scientists or even the ability to easily leverage data collected locally to support decisions, our research team will provide communities with data about services in small town Iowa in order to assist with developing strategies to improve quality of life for their residents amid shrinking populations (Rural Shrink Smart Team, 2022). We hope to allow towns to discover their own data and compare to other similar towns, centering decision-making on data in the context of small-town Iowa life. In the process, we will assess our visualizations to determine which strategies for user interface and interactive graphics design are most useful to empower town leaders to make discoveries in publicly available data assembled with a focus on items that impact rural quality of life.

2 Data Description

Data collected from data.iowa.gov were used to create the SCC dashboard. Most of these datasets are col-

lected on a town/city or county level, requiring us to carefully join data accounting for differences in spatial resolution. data.iowa.gov contains unique information about residents, including local liquor sales, school building locations, town budgets and expenditures, hospital beds, Medicaid reimbursements, and other details that may provide information about local quality of life. Using this data, we created a dashboard which allows communities to explore these data and compare and contrast their local community to other communities of similar size and location. In addition to manual comparisons created by the user, we will use statistical clustering methods to identify groups of towns which employ similar strategies to maintain resident quality of life.

One of the interesting features of this assembled dataset is that missing data can be missing for multiple reasons: not all state data is complete, but data about certain services may also be missing because towns do not offer that service. Thus, in addition to the usual challenges of working with real-world data that is "messy" in a variety of ways, we also have to contend with missing data that is missing due to the size of the community or the lack of services. This makes both visualization and statistical analysis more complicated.

3 Guiding Design Principles

An additional challenge is that research on dashboard creation and interactive visualization tends to be very task-specific and not generalizable. That is, it is relatively easy to create a dashboard that works for a particular task, but it is hard to generalize from that process what will work for the next dashboard. With this in mind, we have clearly documented our intentions at each stage of the design and evaluation process, with the goal of gathering some useful information about general dashboard design from the process of creating this specific dashboard. Thus, our initial set of dashboard design principles is as follows:

- The town leaders are the focus audience; thus, the town itself should be the central focus of the app.
- Facilitate comparisons with other towns in order to allow the user to explore other potential solutions to offering services that enhance resident quality of life.
- Present the user with peer comparisons in order to widen the scope of exploration beyond the initial set of obvious peers in the local region.

- Allow for more detailed data and feature requests to improve the dashboard design over time.

4 Current Progress

We leverage the framework of Guided Discovery Learning (GDL) to guide the town leaders to make discoveries using our interactive visualization. This framework leverages hints, feedback, and other helpful information to guide users in interactive exploration (DeDonno, 2016).

We provide users with a town-centric approach: their town is at the center of our application, and comparisons to other, similar towns are secondary. As it can be extremely difficult to predict which towns are optimal for comparison purposes (similar may involve population, region, economic indicators, sports rivalries, and any number of other variables), we allow users to modify a set of suggested comparison towns to indicate other towns of interest.

As part of this project, we have assembled a large set of publicly available data (primarily from `data.iowa.gov`), but we are careful with how we present this data because it would be easy to overwhelm the user with small details that mask the bigger picture. We select a small subset of towns (out of the 999 towns in Iowa) and a small subset of variables of interest to start with, and then allow the user to increase the complexity of the display in accordance with their interest. This avoids some of the pitfalls of dashboard design that can easily lead to user overload (Few, 2006).

The Rural Shrink Smart EDA dashboard incorporates two major components:

1. Data Visualization Interface component with GDL principles.
2. Exploratory Data Analysis component utilizing statistical clustering methods.

The Data Visualization component with GDL principles is a single page which populates the information in maps related to necessary services, including directions and distance to the fire department, schools, post offices, and hospitals. Value Boxes populate vital statistics sections that have information about the town's QoL Metrics and financial metrics, followed by a parallel coordinate plot that allows the town to see five towns that are similarly based on the most common variables in the towns, such as similar economy or population size. The graphical representations including a parallel coordinate plot and geographic maps. [Figure 1]. The Rural Smart EDA Dashboard map visuals were built using osrm route functions in R to amplify the accuracy of the distances from necessary services in town-centric point.

The Exploratory Data Analysis component utilizing statistical clustering methods enables the town leaders to interact with the parallel coordinate plots at various levels of granularities of the data, with constraints, to similar towns. The Rural Smart EDA Dashboard features offer geospatial and structural data to assist town leaders in gaining a comprehensive understanding of the town in which they support.

Each component provides the town leaders with analytical aspects related to the performance of their town. As town

leaders interact with the tool more helps build the knowledge and understanding different dimensions and measures related to the shrinkage of the town to a particular indicator.

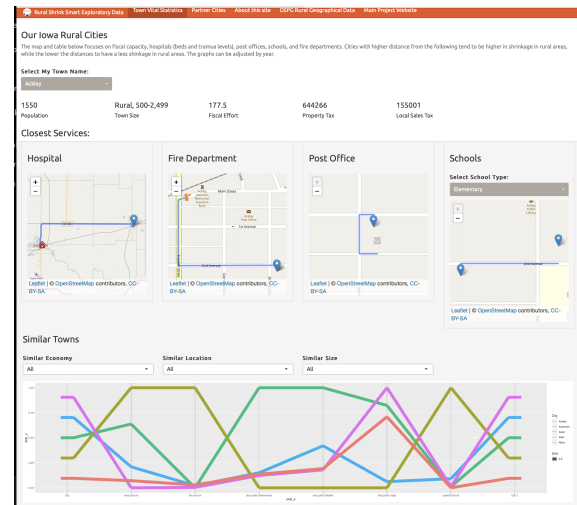


Figure 1: Figure 1: Rural Smart EDA Dashboard Design

4.1 Statistical Methods

We have implemented unsupervised statistical methods, such as K-means, Principal Component Analysis (PCA), and Hierarchical Clustering, to determine the five similar towns, based on distances and services available. We continue to use clustering methods to identify the towns that will be best compared when the analyst interacts with our dashboard. Using GDL, our dashboard makes sure that the focus small town is compared to other towns to find ways to improve their town using data-driven decisions. Our research partners have worked with five towns directly, who have agreed to partner with the project to help identify the best practices that are useful in the small and rural Iowa towns. As a result, we found out that the small towns do not want to be told what to do but would like to make decisions for those who have been a part of the community for generations. Our dashboard will hopefully make sure that the town's analysts are guided in the direction that will help the town's population shrink smart.

To restate, our dashboard incorporates an extensive number of variables used to ensure a valuable tool to town leaders, we implemented parallel coordinate plots to show comparisons between the user's small town with the five similar towns. The user has an option to select similarities, such as economy, location and size, allowing the town leaders the flexibility to guided their interests in the small town. As stated by DeDonno (DeDonno, 2016), our dashboard leverages hints for the town leaders to learn through feedback on the next steps to take in their towns.

A parallel (coordinates) plot allowed our town analyst to compare the essential features of several individual towns on a set of numeric and categorical variables collected. Historically, parallel coordinate plots have been used in air traffic control to help with collision avoidance algorithms but have been adapted in the statistical data visualization space, emphasizing three critical considerations: the order, the rotation,

and the scaling of the axes. Our coordinate plots will display variables categorically and numerically by scaling the data.

Our dashboard development has become more focused on the GDL such that a continual feedback loop from the town analyst in these five towns has helped provide our team with valuable notes that will be adaptive to all of the small towns in Iowa. Our team has captured valuable metrics on using the product allowing the quantitative and qualitative data collection to better user experience. This can promote challenges in towns feeling "studied" by people in cities outside of their own.

5 Acknowledgements

We would like to thank our sponsors and grant provider NSF, LSAMP Inspire Program and Iowa League of Cities. We would like to thank our partners and team members at Iowa State University and University of Nebraska - Lincoln. We would like to thank the State of Iowa for providing Data along with our partners in the Rural Iowa communities.

References

- Michael A. DeDonno. 2016. The influence of iq on pure discovery and guided discovery learning of a complex real-world task. *Learning and Individual Differences*, 49:11–16.
- Stephen Few. 2006. *Information dashboard design: The effective visual communication of data*. O'Reilly Media, Inc., Newton, MA.
- Rural Shrink Smart Team. 2022. Rural shrink smart. <https://ruralshrinksmart.org/>. Accessed: 2021-12-01.
- Alper Sarikaya, Michael Correll, Lyn Bartram, Melanie Tory, and Danyel Fisher. 2019. What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics*, 25(1):682–692.