

Exploring Rural Shrink Smart Through Guided Discovery Dashboards

DENISE BRADFORD^{1,*} AND SUSAN VANDERPLAS^{1,*}

¹*Department of Statistics, University of Nebraska - Lincoln, United States of America*

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); Parallel Coordinate Plots (PCP); Guided Discovery Learning (GDL).*

1 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.

2 Introduction

With the amount of publicly open-source data, a proliferation of visualization dashboards has increased in nearly every industry `fisher`. 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. An important area of dashboard design and usage is to visualize many different attributes that are useful to display key performance metrics of importance for stakeholders.

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 `scc`. 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

*Corresponding author Email: denise.bradford@huskers.unl.edu.

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.

3 Data Description

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

Data collected from ELSI from <https://nces.ed.gov> were used to show the distance to any private or public school. The National Center for Education Statistics (NCES) is the primary federal entity for collecting and analyzing data related to education. These data are used for details are a town level, that will provide information about the quality of life.

Data collected from the Index of Relative Rurality (IRR) <https://www.ers.usda.gov/topics/rural-economy-population/rural-classifications/> were used in the SCC dashboard to help classify the towns. The Index of Relative Rurality (IRR) is a continuous, threshold-free, and unit-free measure of rurality. It is an alternative to the traditional discrete threshold-based classifications. The IRR ranges between 0 (low level of rurality, i.e., urban) and 1 (most rural). Four steps are involved in its design:

1. Identifying the dimensions of rurality: population size, density, remoteness, and built-up area.
2. Selecting measureable variables to adequately represent each dimension:
 - Size: logarithm of population size
 - Density: logarithm of population density.
 - Remoteness: network distance.
 - Built-up area: urban area (as defined by the US Census Bureau) as a percentage of total land area.
3. Re-scaling the variables onto bounded scales that range from 0 to 1.
4. Selecting a link function: unweighted average of the four re-scaled variable.

Data collected from Rural Urban Commuting Area Codes [usda](https://www.usda.gov) were used to help identify towns with commuting behaviors in our rural areas. The rural-urban commuting area (RUCA) codes classify U.S. census tracts using measures of population density, urbanization, and daily commuting. This data is on a zipcode-level that will help identify those communities that commute to more urban areas. The most recent RUCA codes are based on data from the 2010 decennial census and the 2006-10 American Community Survey. The classification contains two levels. Whole numbers (1-10) delineate metropolitan, micropolitan, small town, and rural commuting areas based on the size and direction of the primary (largest) commuting flows.

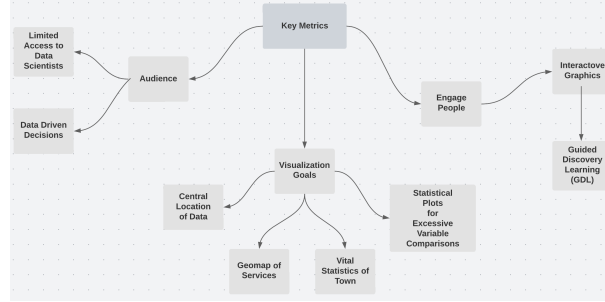
Using this data, we created a dashboard which allows communities to explore these data and compare and contrast their local community to other communities of similar size and location. In addition to manual comparisons created by the user, we will use statistical clustering methods to identify groups 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.

4 Methods

4.1 Identifying Key Metrics for our Dashboard Design



Our audience being small rural towns, traditionally, do not have access:

- Data scientists
- 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.

While data is important for data-driven decisions, we need to help minimize the amount of cognitive load for the town-leaders to digest the amount of data will be conveyed, as a result we will create a central place for the excess of data to be conveyed in a single comprehensive visual. We will utilize Parallel Coordinate Plot (PCPs). Create a map feature that will allow for town leaders to understand the distance of essential QoL components (e.g. Fire Dept, Schools, etc.). Lastly, we will add an overview of town general basics related to the seven QoL factors.

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.

4.2 Guiding Dashboard 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.

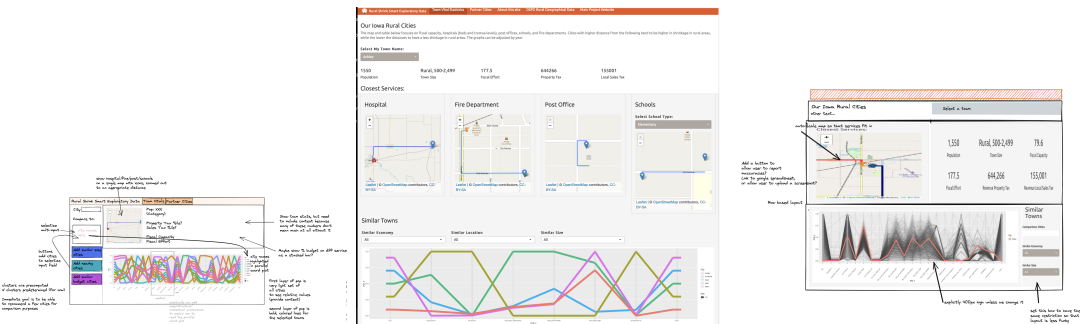
4.3 Dashboard Design Workflow

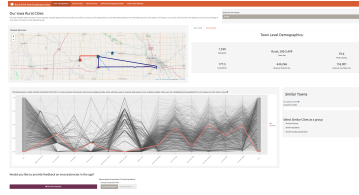
We provide users with a town-centric approach: their town is at the center of our application, and comparisons to other, similar towns are secondary. As it can be extremely difficult to predict which towns are optimal for comparison purposes (similar may involve population, region, economic indicators, 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. The suggested comparison towns are generated using unsupervised clustering methods on different sets of variables, but as we acknowledge that the problem of what towns should be used for comparisons is a tough one even for humans, we still want to allow for user input in this space.

Secondly in our workflow, we will look into using Unsupervised Clustering Methods to help our towns make decisions that are informed in comparison to similar towns, for budget size, population size and location. Our research implemented unsupervised statistical methods, such as K-means, Principal Component Analysis (PCA), and Hierarchical Clustering, to determine the next five to ten 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. We will take into account that towns will make sure to help towns not make comparisons to towns that may be a bit out of reach and not similar in make-up. We will discuss this more in detail in the discussion, our original goal at using the clustering methods to determine similar towns actually led us to using the methods to determine the data quality of our open-source data.

In our final stages of the Dashboard Design Workflow, we will look at two important aspects including the Layout and feedback. Firstly, we will look at the dashboard layout that will allow the user to engage that most without overwhelming their curiosity and understanding of the delivered information. We will conclude with feedback on what elements on the dashboard are useful, confusing, misleading or incorrect based on two modes - clicks on the UI (Passive) and UI survey (Active).

An aspect of our dashboard design, we will iterate over designs, allowing for the design of dashboards to be a spectrum from dream design of the design team to a more functional design from the town leaders feedback.





4.3.1 Dashboard Components

The large set of publicly available data (primarily from data.iowa.gov) we have assembled is useful, but we must be 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](#).

While our dashboard layout is vitally important into making sure that the town leaders are using the dashboards, we should discuss the dashboard components that we used along with the philosophy of those key components. Starting with our parallel coordinate plot, which will allow for objective measures related to the QoL Survey. As our primary objective is to provide users with a town-centric approach: their town is at the center of our application, and comparisons to other, similar towns are secondary. We believe that this will allow the town leaders the ability to see as much data as possible with relation to their towns and others. The framework mentioned in [Lee](#), ‘How do People Make Sense of Unfamiliar Visualizations?: A Grounded Model of Novice’s Information Visualization Sensemaking,’ allowed our work to determine that a novice user seeing a PCP will be able to determine two basic components: Visual Object (textual objects and non-textual objects) and Frame (frame of content and frame of visual encoding). The PCP in this study were found to be promising for novice users, when comparing to the Treemap and Chord Diagram.

Furthermore, the large set of publicly available data we have assembled is useful, but we must be careful with how we present this data because it would be easy to overwhelm the user with small details that mask the bigger picture.

4.4 Guided Discovery Learning (GDL)

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](#).

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 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.

4.5 User Testing

In this paper, we present the results from testing this application with users, and examine the implications for the feedback we received on our application design and on the more general design of dashboards for statistical novices. We examine which aspects of dashboard design “common knowledge” hold up in practice, and assess why these principles may not hold in our use-case.

Complex visualizations are difficult to assess in vivo: it is fairly easy to design an experiment to test a single graphic, but much harder to test the interactive dashboards which consist of systems of displays and user interactions. As a result, it is important to continuously assess which “common knowledge” design principles stemming from more controlled experiments hold up in real-world situations.

Using the framework from “Creating design guidelines for building city dashboards from a user’s perspectives,” young, we will help summarize a users concerns utilizing six main critiques, which are framed into questions. We will adapt these questions to better help the town leaders communicate clear feedback for our team to use. We will collect how the users are experiencing the app with these set questions:

- Is the information the dashboard presents clear? (*Scope*)
- Is the dashboard easy to navigate? (*Usability*)
- Is the dashboard used in a socially responsible manner? (*Ethics*)
- Do you believe that you can trust the Rural Shrink Smart dashboard? (*Validity*)
- Do you find value in the Rural Shrink Smart dashboard? (*Utility*)

4.5.1 Beta Testing

Our Beta testing of a design dashboard before deploying the product to our town leaders, we allowed users from the following groups:

- Graphics Group Members (Users know Statistical Graphics)
- Rural Shrink Smart Project Members (Users know the project with little background in Statistical Graphics)
- Others (Member of doctoral community that has background in educational adaptivity)

While this may not be a large sample size, our range of testers helped with changing approaches in data collection and dashboard changes.

5 Discussion and Future Research

5.1 Data Collection Revisited

As mentioned above, we wanted to use Unsupervised Clustering Methods to determine similar towns, however we were able to use this methodology to determine data quality in our open data sources. This work empowered our time to look for other open data sources to validate our vital

statistics and maps. Since the clustering methods were used differently, we decided to use the USDA's version of rural area definitions. This allowed for a more transparent look into towns that may be similar based on their rurality index. Similar to the Census, the rurality metrics are county based and updated every ten years. This could later cause a problem for towns that may be shrinking faster than the collection of data. It could be worth looking into developing a proxy from our data collected that would be reliable for the years in between the USDA's collected years.

In a related idea of making sure that the towns are identified correctly, we should look into the aging of our towns. This may be important to identify the reasons for shrinking. In the event, that the towns are aging, will we be able to identify how fast the younger generation is moving out of the towns. This could be the most important piece of truly identifying the use of data in towns with an aging population.

5.2 Visualization Revisited

In our statistical graphs, we used Parallel Coordinate Plots (PCPs) to allow the user to look at a variety of variables, while the Guided Discovery Learning was used to make sure that the users were informed of the information that was presented, it did not allow for the town leaders any more ease in understanding. As a result, we decided to look for easier ways to display data, such as boxplots and barplots. This direction will be limited but may be useful for the sake of ease on the town leaders.

In the event that we want to continue to have the town leaders explore using a different statistical graphic, we would like to introduce the flower plot. This method may not represent that data in the way that could be misleading for understanding, we believe that the flower plot will allow for town leaders to engage in being curious as they explore more dimensions of data related to their Quality of Life (QoL) survey.

5.3 Feedback

We will continue to engage with town leaders to ensure that we are developing a product that will help them make better data-driven decisions that are useful to increasing their quality of life, which may lend itself to help less shrinkage in these rural areas.