

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 [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. An important area of dashboard design and usage is to visualize

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

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

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 <https://purr.purdue.edu/publications/2960/1> 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 [John Cromartie and USDA \(2020\)](#) 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.

4 We need to add words for the data that we will include from the Census that will tell use the age of towns (Are they aging?)

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.

5 Methods

5.1 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.

5.2 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 \(2016\)](#).

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. 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.

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 \(2006\)](#).

The Rural Shrink Smart EDA dashboard incorporates two major components: Data Visualization Interface component with GDL principles; and 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 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.

5.3 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.

6 Code

The document class `jds` provides several commands to decorate

- inline code, such as `print("Hello world!");`
- programming language, such as R, Python, and C++;
- software package, such as *stats*, *utils*.

```
## Dobson (1990) Page 93: Randomized Controlled Trial :
counts <- c(18, 17, 15, 20, 10, 20, 25, 13, 12)
outcome <- gl(3, 1, 9)
treatment <- gl(3, 3)
glm.D93 <- glm(counts ~ outcome + treatment, family = poisson())
summary(glm.D93)
```

Call:

```
glm(formula = counts ~ outcome + treatment, family = poisson())
```

Deviance Residuals:

1	2	3	4	5	6	7	8
-0.67125	0.96272	-0.16965	-0.21999	-0.95552	1.04939	0.84715	-0.09167
9							
-0.96656							

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.045e+00	1.709e-01	17.815	<2e-16 ***
outcome2	-4.543e-01	2.022e-01	-2.247	0.0246 *
outcome3	-2.930e-01	1.927e-01	-1.520	0.1285
treatment2	1.338e-15	2.000e-01	0.000	1.0000
treatment3	1.421e-15	2.000e-01	0.000	1.0000

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 10.5814 on 8 degrees of freedom
 Residual deviance: 5.1291 on 4 degrees of freedom
 AIC: 56.761

Number of Fisher Scoring iterations: 4

7 A Placeholder Section

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque

vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

8 Citing References

The citations are in the author-year format with the `jds` bibstyle.

Citations can be in either text or parenthesis format style with `\citet` or `\citep`, respectively. For example, ? is a seminal work on quantile regression; The Laplace distribution has applications in many fields (?).

9 Discussion

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

References

- John Cromartie and USDA (2020). Rural-urban commuting area codes. <https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/>. Accessed: 2022-04-01.
- DeDonno MA (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.
- Few S (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.
- Sarikaya A, Correll M, Bartram L, Tory M, Fisher D (2019). What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics*, 25(1): 682–692.