Exploring Indiana's education and workforce data through a regional map-based dashboard interface

Author 1, Author 2, *Member, IEEE* and Author 3

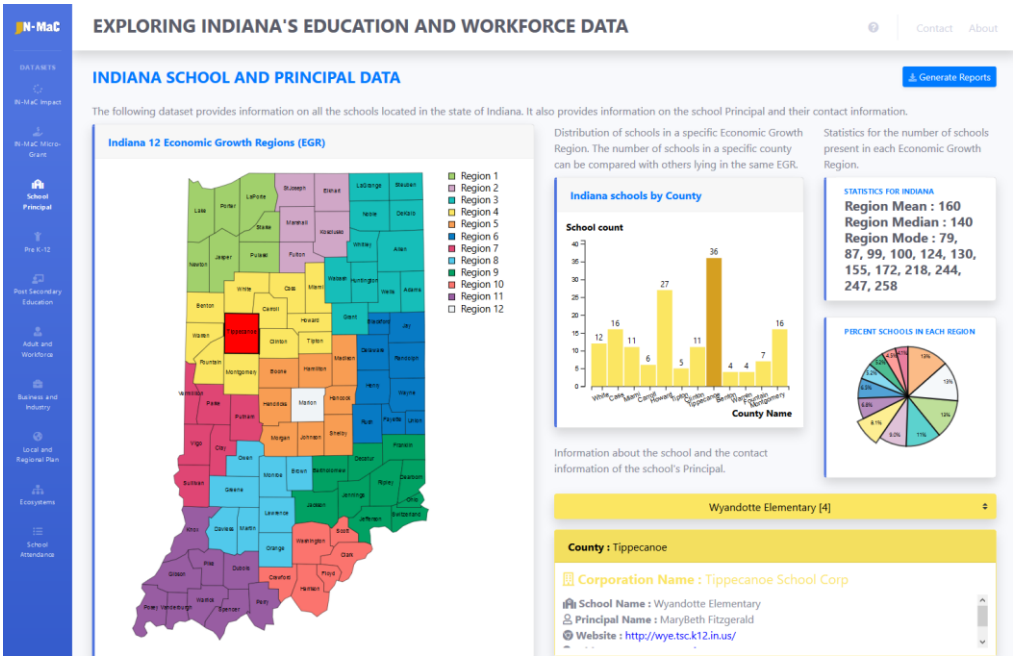

Fig. 1. Data dashboard interface based on Indiana's 12 Economic Growth Regions

Abstract— In this paper, we present our collaborative work with IN-MaC (Indiana Manufacturing Competitiveness Center), in which we develop a data visualization dashboard for education and workforce data for the state of Indiana. This dashboard uses the map of the state of Indiana as an interface to access the various data fields present in several education and workforce related datasets. The map is designed to mimic the 12 economic growth regions (EGR) in which the state of Indiana is divided as per the economic and social ties. The end users of such data view the state through the lens of these 12 regions. All the other visualizations on the dashboard are linked to this interactive map and they automatically update as per user interaction through the map visualization. Our dashboard allows users to view data about schools and principals, pre-k through 12, post-secondary education, attendance, businesses and industries, adult workforce and other stakeholders through this common interface. The visual encoding used in our dashboard is consistent with the theme of the 12 regions to ensure ease of use and visual interpretation from the graphs. Further, we propose that accessing various datasets for different states through a regional map based design can help to quickly and easily view and analyze large data for other domains with similar data structures.

Index Terms—Education data, workforce data, data dashboard design, visualization evaluation, Indiana data, map visualization, regional map, economic regions, data visualization

1 Introduction

The education and workforce domain consist of a variety of datasets such as schools and principals, pre K-12 and post-secondary education data, businesses and industries, enrolment, attendance, etc. Such data about the state of Indiana is collected every year and is available publicly on the Indiana Department of Education [1-2], the Indiana Department of Workforce Development [3] and other websites [4-5]. This data is largely present in excel sheets and PDF files and does not have a single central repository. The users of such data are government officials, education and workforce employees, research institutes, parents and other stakeholders in education. Thus, decisions about visualization and interactivity mechanisms must be made to address the perspectives and work processes of the target users. To address this problem, this research aims to create an ecosystem of the data represented through a data visualization dashboard which will allow users to make sense of and navigate through these large datasets. Various datasets in this domain are related to each other and finding relationships and dependencies between them can greatly aid decision making. Though the end user can be diverse, the dashboard design is primarily targeted towards a specific end user that views the state of Indiana through the lens of the 12 Economic Growth Regions [6].

- *Author 1, Author 2 and Author 3 are with some Midwestern University in the United States. E-mail: {Author1,Author2,Author3}@SomeMidwesternUniv.edu*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org.
Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx/.

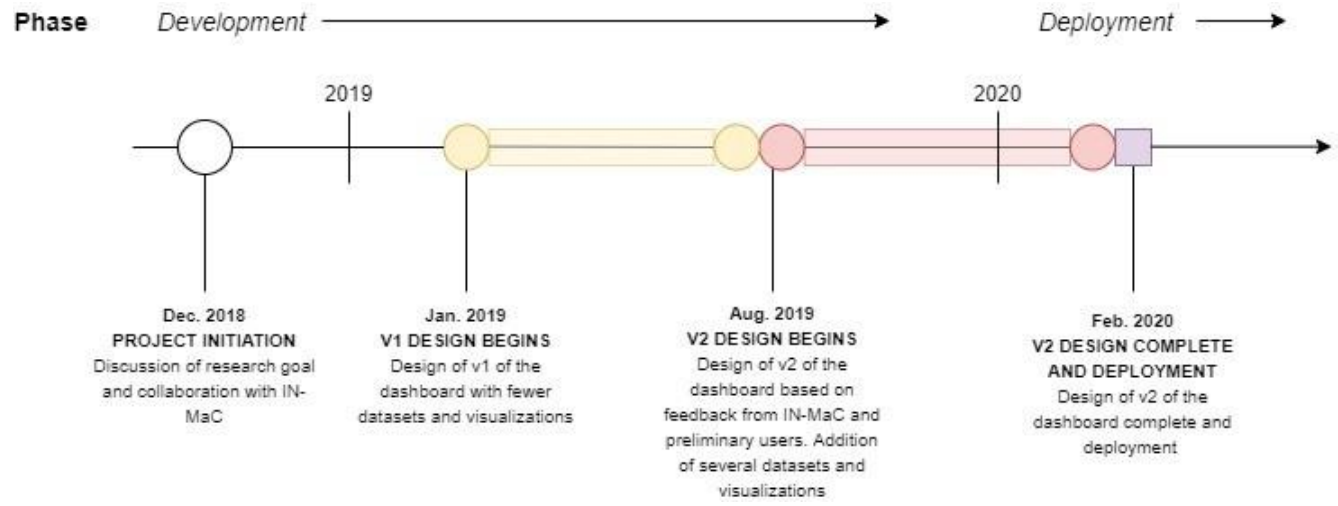

Fig. 2. Timeline of the dashboard's design and deployment phases: design of V1 is shown by yellow circles, design of V2 is shown by red circles, deployment is shown by the purple square.

The state of Indiana is divided into 12 Economic Growth Regions (EGR) based on economic and social ties [6]. As stated in [7], "The Economic Growth Region came from the Indiana Department of Workforce Development's 2005 Strategic Skills Initiative (SSI), designed to align the state's economic and workforce development efforts and address anticipated workforce shortages." This paper reports on the design and development of a data dashboard (deployed at www.palceholderwebsite.com) that utilizes this regional breakdown characteristic as a way to access and filter large datasets related to education and workforce for the state of Indiana. A timeline that shows the development and deployment of the dashboard is shown in Figure 2. After discussing the research goals, use case of the dashboard and learning about the perspectives of the target users, we developed an initial version of the dashboard (v1). After the design and regional breakdown strategy was approved, we created another version with multiple datasets and visualizations. This was then deployed on the web to be used by the stakeholders and end-users.

A dashboard template provided by startbootstrap.com [8] was used as a boilerplate and modified as per the need to create this dashboard (v2). As shown in Figure 1, the data dashboard is driven by various datasets that can be selected from the navigation pane (on the left). A map-based visualization of the 92 counties that make up the state of Indiana is used, that represents the 12 EGRs. The counties are color coded based on the region that they fall in. The user can click on a county to view data related to that county and region. Other visualizations on the dashboard, such as the region bar graph and pie chart, show data about the current selected county. Another bar graph is present (Figure 5) which shows data about all 92 counties at the same time to provide an overview of the entire state. Other detailed data about a specific organization/institute (such as a specific school) in that county can be viewed by using the drop down feature, located below the region bar graph, to select the school of interest.

Contributions: We present our work as a visualization design study in which we create a data dashboard specifically addressing challenges of Indiana's education and workforce domain and work with collaborators and users in that domain. Our contribution consists of the dashboard developed to address the needs of the users, the regional map based design choice as a suitable method to access and drill down data and the data pre-processing required to allow creation of the visual encoding and interaction schemes. We also discuss the uses of our data dashboard from preliminary discussions with the users at IN-MaC.

Outline: Section 2 consists of a review of the related literature in the domain of visualization of education and workforce data, data dashboards and dashboard design principles. In section 3, we discuss our overall motivation and dashboard development goal. The methodology of the design of the dashboard is presented in section 4. Lastly, we discuss some preliminary feedback from the end users at IN-MaC in section 5. In section 6, we discuss the limitations, scope for future work and user studies to evaluate the data dashboard. In section 7, we summarize our contribution and its applicability to visualizations in other domains.

2 Related Literature

Several methods have been used to visualize education and workforce data. Indiana specific data has also been visualized by the Indiana Department of Education (IDOE) and the Department of Workforce Development (DWD). We review such visualizations, various types of dashboards and dashboard design principles.

Education and workforce data visualization: The datasets from Indiana's Department of Education are the underpinnings of the visualizations in the dashboard created in this work. These visualizations will be tailored to address various goals and needs of a specific end user. Currently, some visualizations of some of these datasets are available on the internet [9]; however, most do not have a consistent visualization dashboard. Some others are created as individual visualizations on a webpage with different interaction mechanisms [10]. They do not follow visualization principles described by [11]: for example, unjustified 3D, chart junk, incorrect color scales, no labels for the x and y axes, etc. Also, the filtering mechanisms on the website are broken, leaving the user with limited functionality and unable to explore, and unable to find relationships and inter-dependencies within the data. Hence, there is a clear need to have a better visualization design that makes the interaction with and interpretation of the data easy, and insights from them clearer.

In our research, we found that an educational dashboard exists for schools in California [12]. California is one of the largest school districts in the United States by enrolment (based on a Wikipedia search). The dashboard provides an accessible location for its data, but it dashboard lacks functionality that support ease of use. First, the dashboard does not provide any overview or map-based visualization to select a specific region or area. The users need to enter the name of the school or city in order to find information about that school. The information about suspension rate, graduation rate, college/career, etc. is visualized as a speed gauge meter with a rainbow color scale which is highly unrecommended [11], [13]. Further, selecting one of these overviews takes the user to the bottom of the screen which explains what the color scale means and provides a general overview in percentages. No other context is provided. There is no easy way to view a different field, the user must

manually scroll back to the top of the page to view a different field. Also, there is only one gauge visualization used for all the data types which seems highly unsuitable [13]. Clearly, given the importance and far-reaching impact of such education data, it is imperative that the data should be visualized in such a way that fosters insight.

Data Dashboard: Data dashboards have been used for a long time to display information at a glance for various domains. Dashboards are designed to also allow users to interact with and change the granularity of the data displayed. It can be used for real-time data monitoring or can be used to present information from static datasets. With such a diverse use case, there have been discussions on what qualifies as a dashboard. Obviously, there is no fixed definition of what a dashboard is, and several data presentation applications are now called dashboards as well. In our research, we define the dashboard to have 5 main features: it must present the data with multiple views, it must have interactivity, multiple datasets should be visualized, it must be consistent across the datasets and it must satisfy multiple use cases.

Recent surveys in data dashboards have categorized dashboard designs into two types, functional genre and visual genre [14]. The visual genre consists of "a structured tiled layout of simple charts and/or large numbers" [14, p. 683]. On the other hand, the functional genre is "an interactive display that enables real-time monitoring of dynamically updating data" [14, p. 683]. This research focuses on producing a combination of these design choices by having a functional backend that allows the clear representation of data along with control and navigation mechanisms (interactivity) while having a visually effective frontend with various visualizations, text elements and relevant statistics laid out in a structured format. The dashboard was designed by following some best UX design principles given in [15].

Data dashboards can be designed by following several different guidelines/principles such as "organizing information to support meaning and use, maintaining consistency, putting supplementary information within reach, preventing excessive alerts, accommodating real-time monitoring", etc. as is mentioned in [13, p.183]. Similar precautionary guidelines are also mentioned in [13] such as "to avoid exceeding boundaries of a single screen, supplying inadequate context to the data, displaying excess detail, cluttering the display with visual effects" and so on. While these are recommended guidelines, they may not be applicable in the design of all dashboards and choices have to be made based on the purpose of the application.

Data dashboards have been used to address the needs in various domains such as risk management in the insurance industry [16], decision support and management control [17], ventilator management systems [18], radiology [19], blood pressure monitoring [20]. Such applications of dashboards are countless. A data dashboard was created to address the issue of data representation in the domain of urban education [21]. In order to reduce cognitive load during clinical decision making, a data dashboard (MIVA 2.0 – Medical Information Visualization Assistant 2.0) was developed. The goal of this human centric dashboard was to support rapid analysis of real-time clinical data-trends [22]. Health records visualized through dashboards can be combined with physiological vital sign monitor screen to provide near-real-time surveillance. Such an adaptation of a medical dashboard has been done in [23]. The data analytics performed through these dashboards guide nursing practice and research.

Some dashboard designs also facilitate end user customizations and social sharing. A dashboard application was developed along those lines for real-time energy monitoring [24] to study whether customization ability would influence users' interaction with the system and consumption patterns.

A dashboard to visualize and analyze public safety data has also been created with a variety of overlays with the goal of improving the community [25]. This dashboard uses public datasets and is designed to provide a one stop shop. Several siloed datasets were combined to support examining relationships among the data. The main themes that emerged from thematic analysis on learning dashboards designed for students [26] consisted of: "equal learning opportunities for everyone, choice of whether to compare performance with other students in the class, privacy, automatic alerts and customization ability" [26, p. 319].

Thus, static dashboards for organizations, dashboards designed for strategic and operational decision making, communication, quantified self, operations/alerting [14] and many more are being developed every day to address the problems in various domains.

Dashboard design principles: Standard references for general dashboard design principles are [13] and [27]. "Organize information to support its meaning and use, maintain consistency to enable quick and accurate interpretation, put supplementary information within reach, make the experience aesthetically pleasing, expose lower-level conditions, prevent excessive alerts, keeps viewers in the loop, accommodate real-time monitoring" [13, p.183], "do not exceed boundaries of a single screen, supply adequate context to the data, choose appropriate display media, highlight important information effectively, avoid misusing or overusing color" [13, p. 35] and so on are some of the commonly used and widely applicable dashboard design guidelines [15].

In the domain of medical dashboards, the main design principles are: supporting real-time visualizations, minimization of text and number use to reduce cognitive load, low failure rate, pattern and outlier detection support and rapid information assimilation [22]. Support for multiple datasets, consistency with ICU clinical practice, easily understandable icons for lab work, x-rays, meds, etc. and a mobile platform were other guidelines that were important. Dashboards designed in the medical domain must benchmark the patient's vital rates with normal standards, provide notification and warning, predict trends and summarize the information for quick monitoring [23]. Some dashboards allow users to choose their own datasets and support custom visualizations [24]. Further, adding a social sharing feature to the dashboard makes it more enjoyable as users would like to share their personalized version with others. Both a mobile and web application seem suitable for this purpose. These design principles are applicable in contexts where dashboards are designed for personal use but may not be applicable in other domains.

Student dashboards must have features that can direct them to supplementary study material, provide data on course and study habits, personalized feedback and suggestions for improvement [26]. This entails a dashboard design with linked references to external sources, an AI agent that can suggest study habits and provide personalized feedback. This also guides the choice of visualizations that will be used to represent the data. Encoding inconsistencies across multiple views can be avoided by using two design guidelines: automatic inconsistency detection and resolution with the help of a design assistant and using a mixed-initiative interface [28]. Six design guidelines for developing information systems are given in [29]. They are: "a comprehensive information model, functions to better analyze and process information, easy-to-use IS handling, a more flexible IS architecture and data model, a proper information management, and fast prototype implementation" [29, p. 16]. A prototype was developed for a large international chemicals group by applying the given principles.

Even though a large number of dashboard design principles have been established and defined, not all of them are applicable while designing effective dashboards for representing education and workforce data.

3 MOTIVATION

This project began in December 2018 with the discussion of finding ways to show how IN-MaC (Indiana Next Generation Manufacturing Competitiveness Centre), as an organization working closely in the education and workforce domain, impacts the current state of development through its various initiatives and activities. Here, we provide a discussion about IN-MaC and the motivation presented by

them for the development of the dashboard as a way to address specific problems in the education and workforce domain in the state of Indiana. Since multiple related datasets and organizations are presented under one dashboard umbrella, this dashboard is referred to as an 'ecosystem dashboard' to indicate the links and inter-dependencies between them. This is different from the ecosystem datasets that have been visualized as explain in further sections.

3.1 IN-MaC

As state in [30], "Indiana Next Generation Manufacturing Competitiveness Center (IN-MAC) creates a stronger, more competitive manufacturing ecosystem for the state of Indiana and the nation. IN-MaC does so by mobilizing its resources, expertise and network by linking and leveraging assets to strengthen the relationship between workforce education and manufacturing research to elevate Indiana as the manufacturing destination of choice." The Center's goal is to foster partnerships with local universities and community colleges by addressing a broad range of socio-economic challenges such as: "shortage of trained workers capable of filling open positions; integration of modern technologies and practices throughout the value chain; and smart investments in research with near-term applications to strengthen our manufacturing ecosystem" [30].

3.2 Dashboard purpose

The purpose of this dashboard is to guide participatory, national level analyses of the social service workforce. Workforce mapping is based on the premise that when the right number of workers are in the right positions and locations and have the right training, people will be able to access more effective and appropriate levels of care and support. As a result of completing this ecosystem mapping process, organizations and groups will have data and information about the current status of the education and workforce across the state of Indiana and will be better able to link and leverage assets, align and identify recommendations and strategies to improve the workforce going forward, thereby improving the quality of programming available for K-12 students, educators, workforce entities, industry, and community organizations.

Through the ecosystem dashboard we will eventually be able to compare education and workforce distribution in a more quantitative way. Tracking a core set of metrics consistently will provide a deeper understanding of the education and workforce ecosystem taking place over time:

1. Key enablers could identify where movement is progressing quickly and where it is lagging.
2. Creating an ecosystem that aligns and identifies services to add and facilitate distribution across the state of Indiana.

4 Dashboard Design

This dashboard is designed to visualize education and workforce data. The end users who access such data have a specific lens while looking at the development in the state of Indiana. We discuss this user perspective and then discuss how our dashboard is designed to main visual consistency and data access with that perspective.

4.1 Development tools

We use D3.js (Data Driven Documents) [31] which is a JavaScript library for creating visualizations for the web. Several other visualization tools are also available such as Tableau, Bokeh, Cytoscape, Microsoft Excel, etc [32]. However, we chose D3.js because of its flexibility, clean API design, documentation, ease of use and because we wanted to deploy the dashboard on the web. For the dashboard layout, we used a dashboard template from Startbootstrap.com [8] to speed up and simplify the design process. The template uses HTML and Bootstrap for structure and styling and JavaScript for simple navigational bar, scroll and other interactivity.

4.2 End user perspective

The state of Indiana is broken down into 12 Economic Growth Regions as shown in Figure 3 [6]. As [7] states, "An EGR defines economic regions in Indiana based on economic and social ties". The end users of the data refer to this regional map while looking at the various datasets about employment, schools, commuting patterns, etc. Currently, some visualization websites [9] use an interactive map but doesn't use the same visual encoding. Also, the other visualizations follow a different colour scheme not consistent across the datasets. As you can see in Figure 1, our dashboard follows the same visual encoding across all the visualizations and datasets so that the end user can use the dashboard by applying his prior perspective. Since the dashboard visualizes several datasets with a lot of entries, it can quickly become overwhelming to understand and interpret the visuals and the data. We overcome this problem by providing the user with a dashboard design that seems familiar to him, thus lowering his cognitive load while accessing and using the dashboard.

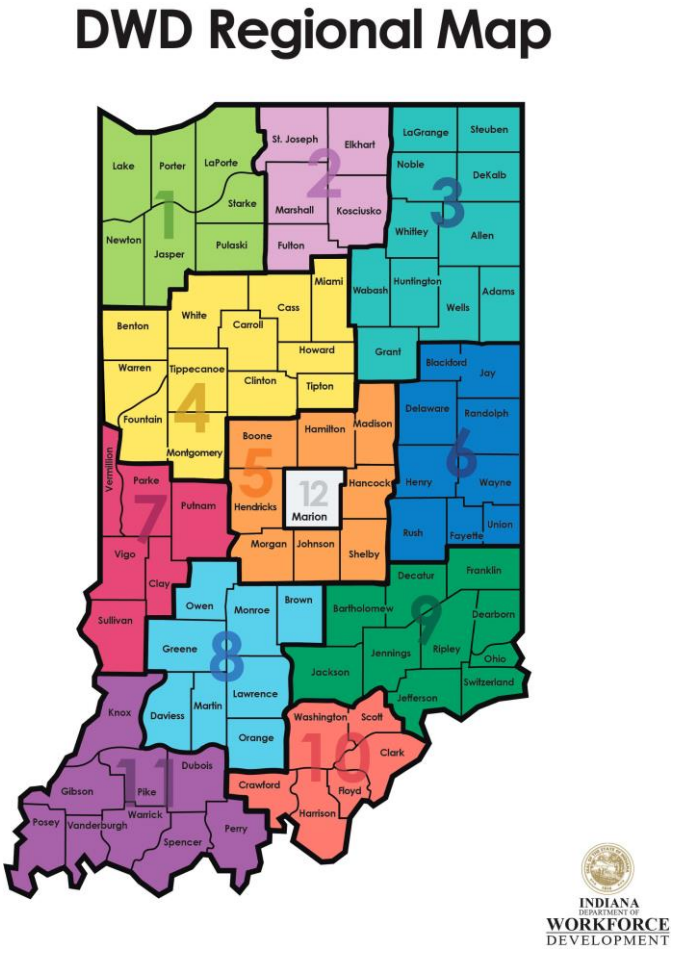

Fig. 3. Indiana 12 Economic Growth Regions (from the Department of Workforce Development), Source: [6]

4.3 Datasets visualized

The dashboard currently visualizes 10 datasets about Indiana's education and workforce development. 7 of these datasets are publicly available through the Indiana Department of Education (IDOE) and the Department of Workforce Development (DWD) and 3 are provided by IN-MaC. The datasets are:

1. IN-MaC impact data (IN-MaC)
2. IN-MaC micro-grant data (IN-MaC)
3. Indiana schools and principal data (IDOE)
4. Indiana pre k-12 data (DWD)
5. Indiana post-secondary education data (DWD)
6. Indiana adult and workforce data (DWD)

7. Indiana business and industry data (DWD)
8. Indiana local and regional plan data (DWD)
9. IN-MaC ecosystems data (IN-MaC)
10. Indiana school attendance data (IDOE)

The IN-MaC impact data consists of the count of the total people impacted by the activities and initiatives that were undertaken by IN-MaC in the state of Indiana. One such initiative is the micro-grant program in which IN-MaC provides funding to organizations that implement manufacturing initiatives that enhance all 12 Indiana Economic Development regions. Ranging from $1,000 - $2,000, the Micro-Grants are designed to support innovation and encourage organizations to dedicate funds towards development and program implementation that impact industry awareness. The IN-MaC ecosystems dataset consists of a different organization of the micro-grant data in order to be able to view the hierarchical relationship in the data. The Indiana schools and principal dataset consists of all the public schools across the state of Indiana along with the contact information for the principal of that school. Similarly, the Indiana pre k-12 and Indiana post-secondary datasets consist of data about the pre k-12 and post-secondary institutions in the state respectively. The Indiana school attendance dataset consists of the percentage attendance of all the public schools from 2005-06 to 2017-18. Further, we have the adult and workforce data that provides information about organizations that help to connect employees with the suitable employers to bridge the unemployment gap. The business and industry dataset consist of organizations that conduct innovative programs, networking events, vocal business advocacy and provide business growth opportunities. The local and regional plan dataset consists of organizations that conduct education and skill development activities at a regional level to enhance the life of the community.

4.4 Data pre-processing

The datasets mentioned above were originally present as excel files with multiple sheets. These were combined to form one CSV (Comma Separated Value) file to be used with D3.js. After converting from excel to CSV, the file was checked and formatting errors that occurred during conversion were corrected. In order to be able to link the data entry to the map, two data fields were important: the name of the county to which the organization belonged and the region number of that county. Most of the datasets consisted of these column entries. For some datasets such as the school attendance and micro-grant data, this data was added manually. This data pre-processing also helped in generation of other visualizations such as the bar graph in which the total number of organizations/institutes in a county and region were calculated based on these entries. Some other statistics such as mean, median and mode were also computed based on them. For some datasets, the information about the contact person, email, location, web address, etc. were missing or were broken. This was only present in the publicly available datasets. Since these were generated by the government, they could not be corrected manually since the data itself was not collected by them. Thus, those fields were left as it is so that the data collectors can use this dashboard and easily find out which data fields were missing so that they could collect them. This would also allow them to avoid collecting the same data again and again.

Some other errors in the data were corrected after creating the visualizations. For example, after a specific county was selected, all the entries in the drop down and the information display text box (shown in Figure 8) was formatted as per the colour of the region to which the data field belonged. Thus, if a field had an incorrect region number or county name, it would be colour coded differently, thus showing visual inconsistency. We thus corrected several entries in the dataset through this method. One drawback was that we had to manually select each entry in the drop down and verify it, but this technique was full-proof and proved to be very effective in correcting the data fields.

4.5 Visualization design

Several visualizations were created while developing the dashboard for finding ecosystems in the education and workforce data. Along with these visualizations, there were additional elements such as text boxes, sliders, drop downs to aid in the successful navigation of the ecosystem. The visualizations created along with example use cases are as follows:

1. 12 EGR and county map – for selecting the counties to view corresponding data
2. Bar graph (region and 92 county) – to compare various counties
3. Pie chart – to view proportions of schools in each region
4. Line chart – to view the attendance of students in a certain school over a period of several years
5. Bubble chart – to compare student impact goal vs. micro-grant amount requested by various organizations
6. Node link diagram – to view ecosystems
7. Proportional symbol map – to view IN-MaC impact across Indiana

Several elements to aid in the exploration of the dashboard were:

1. Text fields and text boxes – to show large textual data such as project information
2. Drop downs – to aid searching for a specific field
3. Sliders/filters – to filter out unwanted fields and select numerical ranges
4. Tooltips – To view information on hovering over the various elements of the visualizations

We now discuss the various visualizations, their design and purpose in the dashboard.

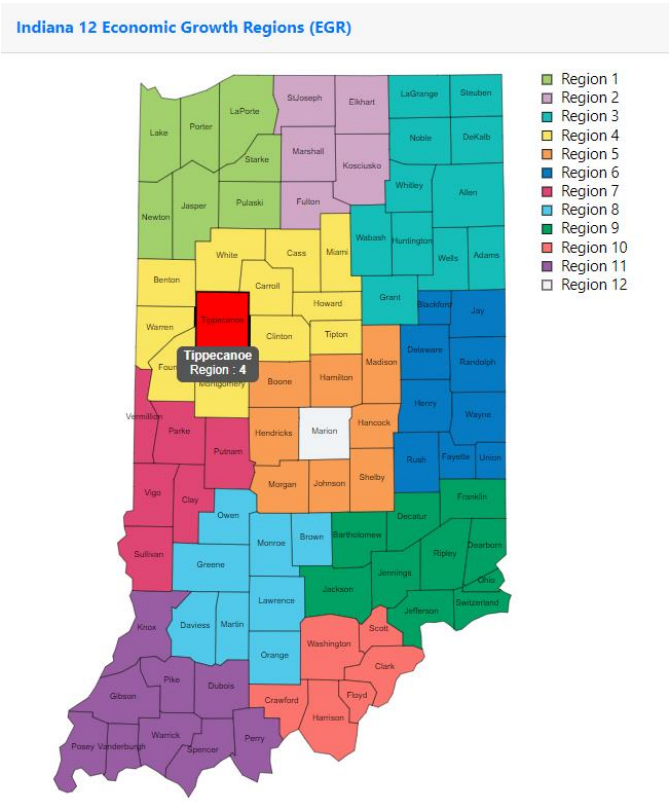

Fig. 4. 12 EGR and county map

4.5.1 12 EGR and county map

As shown in Figure 4, this map is interactive and acts as the main navigation mechanism to view the data fields associated with the current dataset. It is color coded as per the 12 EGR of Indiana. By clicking on a specific county on the map (Tippecanoe county is selected in the figure), the data gets updated to show the entries for

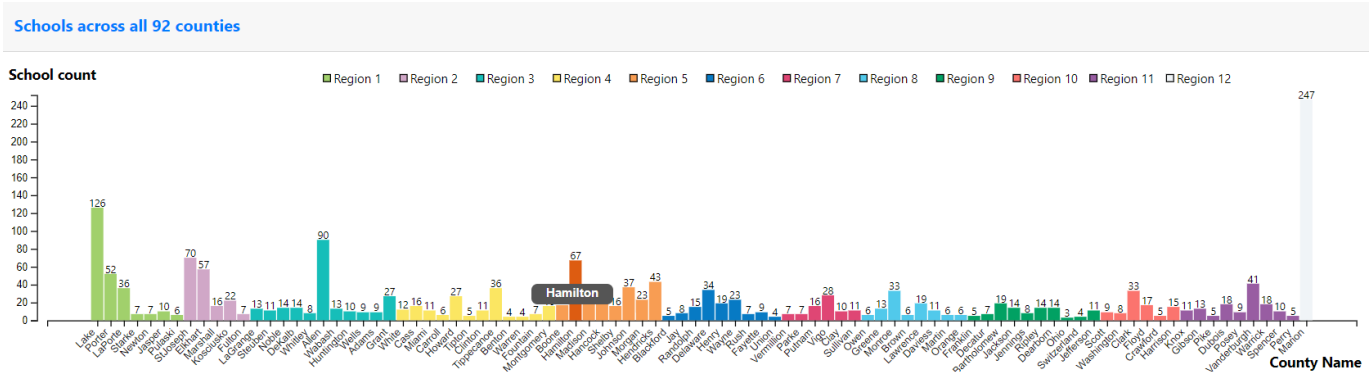

Fig. 5. 92 county bar graph

that county. Also, the selected county turns red and the borders are made thicker to ensure that the selection stands out. By default, Lake county is selected to serve as a cue to new users that the map is interactive. Also, the mouse cursor turns to a pointer (hand) icon on hovering over the county to provide another visual cue. Each county outline on the map has a county name which becomes larger on hovering over it. A tooltip also appears when hovering over the county area which provides information about the county name and region number. This is provided as a reinforcing mechanism. A region legend is present at the top right corner of the map so that users can easily refer to it while searching for a specific county. After clicking on the county, the region bar graph, pie chart, line chart, etc. update with the values corresponding to that county and region.

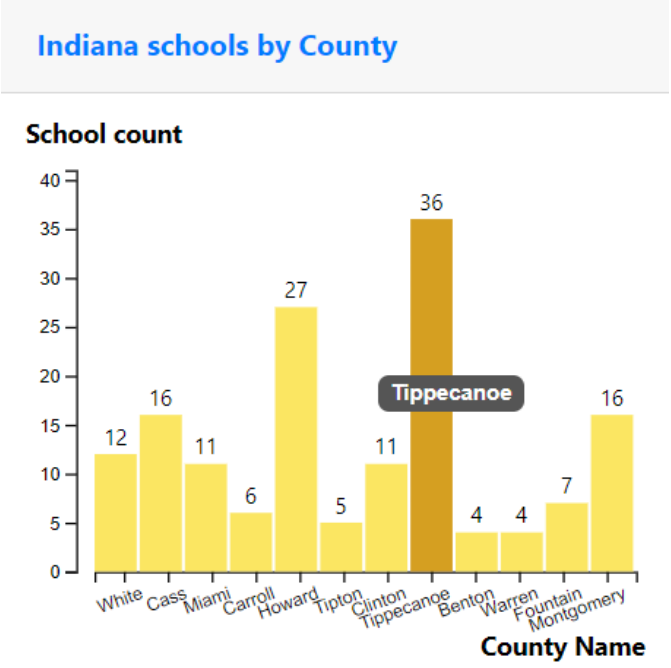

Fig. 6. Region 4 bar graph

4.5.2 Bar graphs

Two bar graphs are visualized for the data. The first one is a region bar graph which shows the counties falling under the selected region on the x-axis and the total organization/institute count in each county under that region on the y-axis. As shown in Figure 6, Tippecanoe county which falls in region 4, is currently selected on the map. Thus, the region bar graph shows all the counties under region 4 along with the total number of schools in that county. The current selected county is highlighted in a slightly darker shade of the region colour (dark yellow in this case). All other county bars are coloured in yellow to match the colour of region 4 on the map. If a county from a different region is selected, this region bar graph will transition smoothly to show the data for that region and county. Its colour will also change as per the region colour. A tooltip with the county name appears on hovering over the bar in the bar graph to ease readability.

There is another bar graph below the map which shows the data for all the 92 counties at once. This allows the users to view data about the state at once and compare all the 92 counties at the same time. Starting from region 1 to region 12, the bars are arranged from left to right respectively as shown in Figure 5.

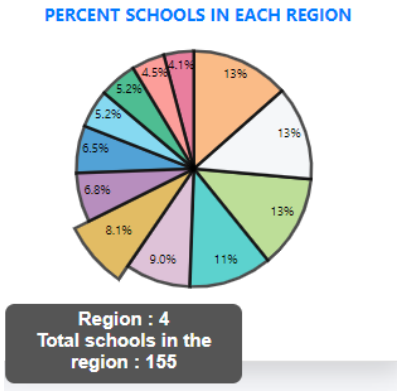

Fig. 7. Pie chart with region 4 selected and associated tooltip

4.5.3 Pie chart

A pie chart besides the region bar graph shows the percentage distribution of schools for all 12 regions across the state. This allows the users to view the total count of the schools from the tooltip and compare this as a percentage of all schools in the entire state. As shown in Figure 7, since Tippecanoe county lies in region 4, the pie for that county is pushed out from the pie chart to allow the users to easily reference the data with respect to the current selection.

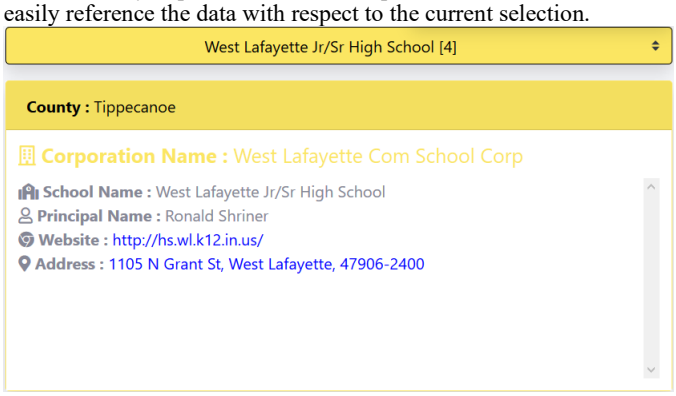

Fig. 8. Information text box and drop down

4.5.4 Data display

After selecting the county, the users can select the organization/institute of interest from the drop down below the region bar and pie charts. This drop down is also colour-coded based on the region that the county lies in. The region number for that organization is displayed in square braces after the name to enforce the data. After selecting the organization, the information text box updates with the data about that organization. As shown in Figure 8, West Lafayette Jr/Sr. High School is selected from the list and the text box shows the principal name, website and location of the school. This information varies depending on the dataset that is visualized. The text box heading is encoded in the same colour as that of region 4.

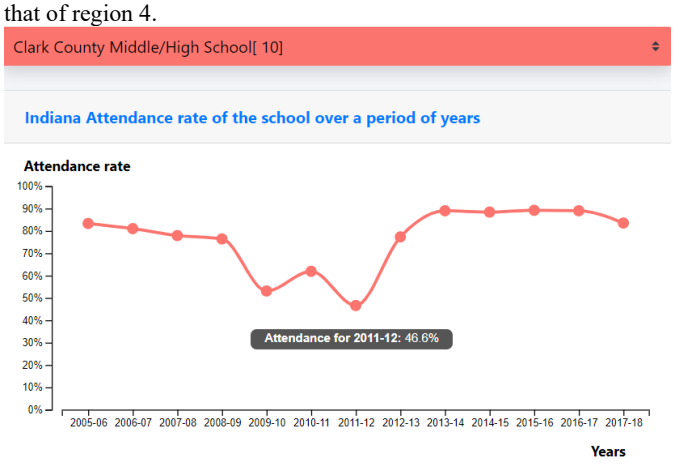

Fig. 9. Attendance data line chart

4.5.5 Line chart

As shown in Figure 9, a line chart is used to view the change in attendance rate over a period of several years for a specific school. Since this is change over time and can be very small, a line chart is preferred over a bar graph. The attendance rate is displayed on a tooltip that appears when hovering over the circle. The attendance rate is also displayed in the form of a table below the line chart to give the users an alternative method to view the data.

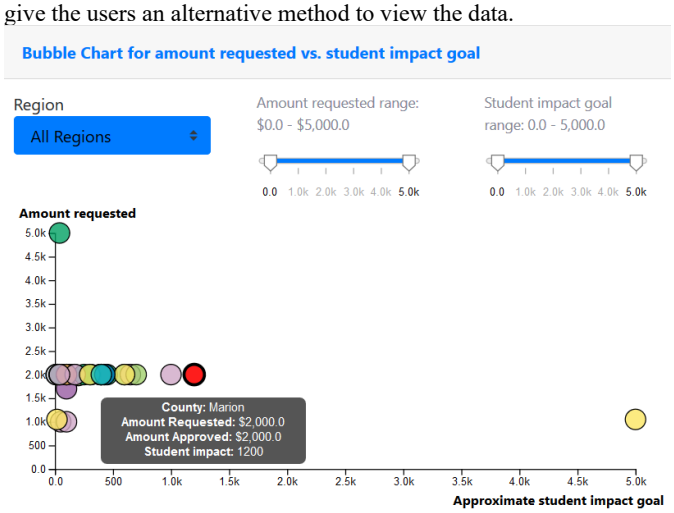

Fig. 10. IN-MaC micro-grant bubble chart

4.5.6 Bubble chart

This chart is present only in the micro-grant dataset visualization page. The bubble chart shows the 'student impact goal' on the x-axis and the 'amount (in dollars) requested' on the y-axis for the micro-grant (shown in Figure 10). Since IN-MaC receives applications from various organizations for small micro-grants, this helps them to evaluate which organization and activity will have the highest impact on the student and the corresponding amount they have requested. An associated information box also shows the activity that will be conducted using the micro-grant along with other information. This helps them to make informed decisions while selecting the organizations to which the grant will be approved.

As shown in Figure 10, a drop down is present to filter out the micro-grants by region. Two sliders are present that allows them to filter out the micro-grants based on the 'amount requested' and 'student impact goal' so that they can look at the data more thoroughly. On clicking on the bubble, the information box updates with the data about that micro-grant. The selected micro-grant circle turns red to show selection. All the circles are again colour coded based on the region that the organization falls under. Lastly, the tooltip contains information about the amount requested and approved, impact goal and county in which the organization falls.

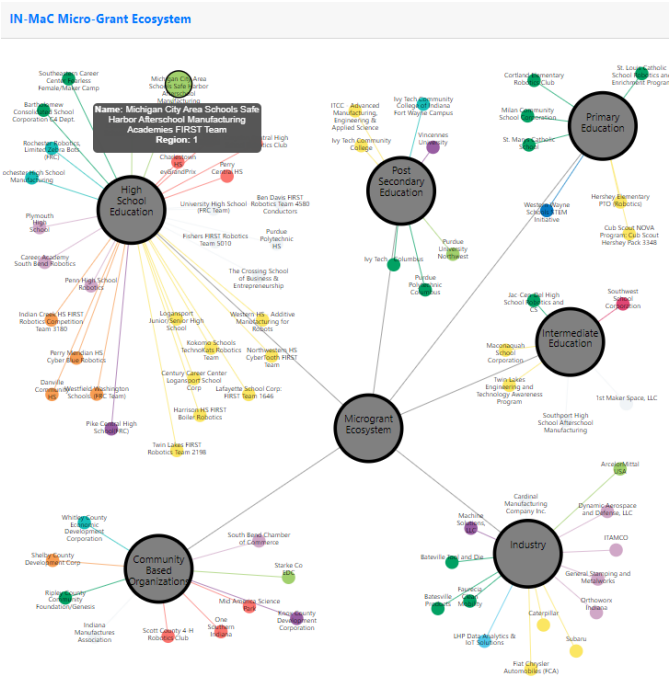

Fig. 11. IN-MaC micro-grant ecosystem

4.5.7 Node link diagram

Figure 11 shows the distribution of micro-grants awarded to organizations based on the category they fall under. This is useful to IN-MaC to ensure that they award micro-grants in all sectors and have a balanced reach. On clicking on the nodes, information about that micro-grant is displayed on a text box adjacent to this node-link graph. This mapping also allows other organizations to learn about the activities conducted by others in the community and the contact information provided helps them to reach out to connect, or find out more details about the activity. Such an ecosystem mapping is also done for the organizations that fall under the local and regional plan dataset.

4.5.8 Proportional symbol map

A proportional symbol map was used to show the impact of IN-MaC across Indiana through its various education and workforce development activities. This map helps IN-MaC track its activities and reach across various regions in the state. It allows them to plan out and direct their efforts in creating manufacturing awareness in other untouched areas. As shown in Figure 12, the impact is shown by the circle. The impact is broken down into 4 ranges as shown by the colour legend. Impact between 0 to 500 is shown in red, 501 to 1000 is shown in green, 1001 to 1500 is shown in blue and 1501 and above is shown in black. The circle size shows the value of impact within that rage. Thus, IN-MaC has greater impact in Starke county compare to Lake county in region 1 since the circle for Starke county

is green in colour and lies in a range higher than the red colour. These colours are unrelated to the 12 colours of the Economic Growth Regions and are chosen because of their visual clarity on overlay over the counties. Also, the opacity of the colour for the counties and circles have been reduced so that the impact circles are clearly visible. The tooltip shows the region number, county and the impact due to the 'Technology Adoption' and 'Education and Workforce Development' activity conducted by IN-MaC.

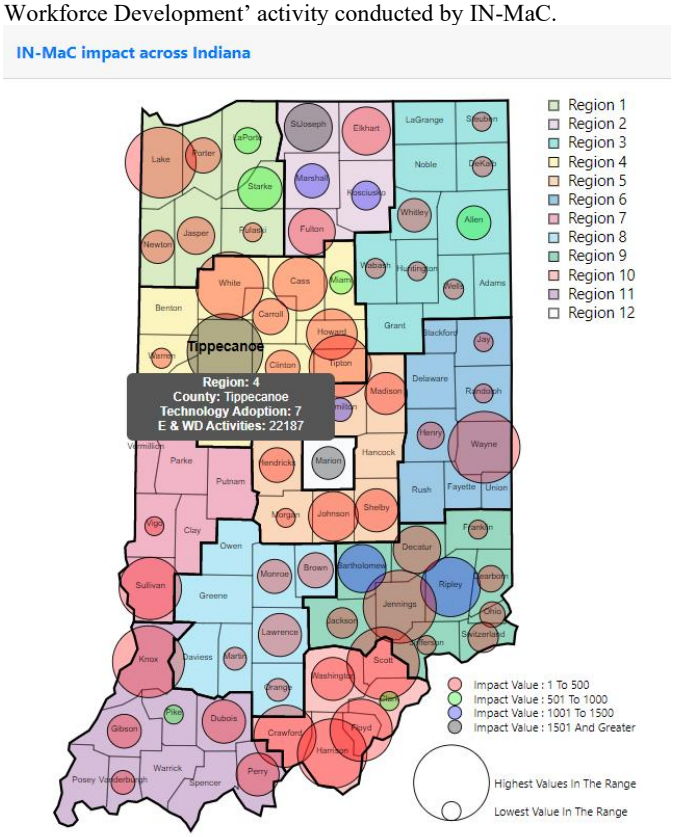

Fig. 12. IN-MaC impact across Indiana map

4.6 Other dashboard features

The dashboard provides textual description of the datasets at various places. A statistics text box is also present which shows basic statistics such as mean, median and mode of the dataset. These calculations are based on the total count of organizations in each region. Based on the level of statistical analysis required by the user, different statistics could be added. At present, only simple stats are shown to address the requirements of the current user.

The dashboard also allows the users to generate a report based on the data that they are currently looking at. After selecting the county and organization of interest, clicking on the 'Generate reports' button will create a snapshot of the current selection along with the timestamp and generate a downloadable PDF. This report can be used by the users to share information with each other or use it to prepare for a meeting, take notes, etc. Figure 13 shows a one-page report that is generated for West Lafayette Jr/Sr High School.

The dashboard also provides buttons for 'About', 'Contact' and 'Help'. The 'About' button provides the user with a basic explanation and description of the dashboard as a whole. It is intended for new users. The 'Contact' button provides the users with contact information in case they want to provide suggestions or get in touch with the developers. Finally, the 'Help' button links to a YouTube video that provides an explanation/tutorial on how to view, navigate and interpret the visualizations on the dashboard. This is intended for on-boarding, in order to accommodate new users and those unfamiliar with the features of the dashboard to train themselves to access the various functionalities.

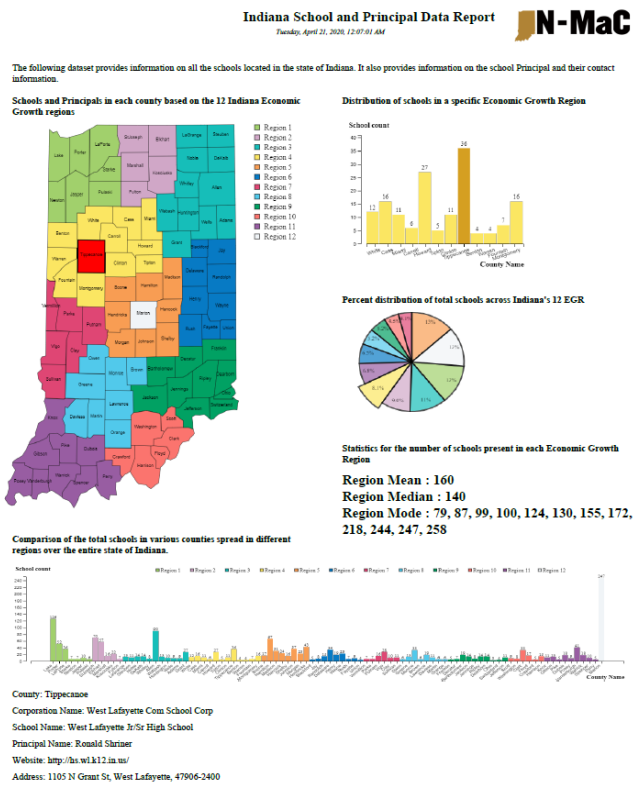

Fig. 13. Report generation feature of the dashboard

5 Preliminary Domain Expert Feedback

Our data dashboard was used by 3 stakeholders in education and workforce development at IN-MaC. In this section, we summarize the feedback from some informal discussions with them. First of all, the breakdown by the 12 Economic Growth Regions was extremely intuitive and visually pleasing to them. They found the interactivity and consistency across the various datasets to be very useful. The most important factor of the dashboard was that it provided a "one stop shop" to access all the datasets. This prevented the users from spending large amounts of time trying to hunt for data in various databases and search engines. Finding contact information was extremely crucial. The ecosystem mapping helped to view the various organizations all laid out in a hierarchical representation. The report generation feature was useful to save the current state of development and email it to various community partners. Also, the instructional video tutorial helped the users quickly get on board with the various features of the dataset and the dashboard functionality.

Some other datasets that were not visualized would be useful such as employment rates, wages, salary data, student count, teacher information and so on. However, regardless of the dataset, the users preferred to view and access it through the map-based visualization. The biggest concern of the users was to maintain and update the datasets. Since the datasets consisted of missing entries and were updated yearly, it was important to ensure that the data was up to date so that the users can use it reliably to make decisions. Some description about the datasets would also be helpful since the dataset names could mean different things for different users.

6 Limitations and Future Work

Currently, the dashboard does not allow the users to manipulate and move the visualizations around to suit their own layout. One of the users preferred to have the bar graphs and pie chart at the bottom and have the information box at the top adjacent to the map. Thus,

each user may find a different layout to be suitable for their use case and the dashboard must be able to accommodate for that. The datasets used are static. They are not being pulled from the original data source at runtime. Setting up such functionalities to load and clean the data will improve the issues with updating and maintaining the data. Support for other statistical analysis such as ANOVA, clustering, t-tests, etc. need to be added as they may be needed by certain end users. Additional datasets also need to be visualized to give the users access to a variety of data through a common interface.

To validate the usability and functionality of the dashboard with users and stakeholders, two user studies will be conducted. A usability test will be conducted on participants from Placeholder University in which they will be asked to perform tasks to evaluate the various features and functionalities of the dashboard. A second study will include semi-structured interviews conducted with stakeholders in education and workforce development in the state of Indiana. The results of the analysis from these two studies will help us to evaluate the usability of the dashboard and find out how the dashboard is being used in the work processes of stakeholders involved in education and workforce development in the state of Indiana. These two studies will further guide how our dashboard will be improved and new features will be added to address the challenges faced by the users.

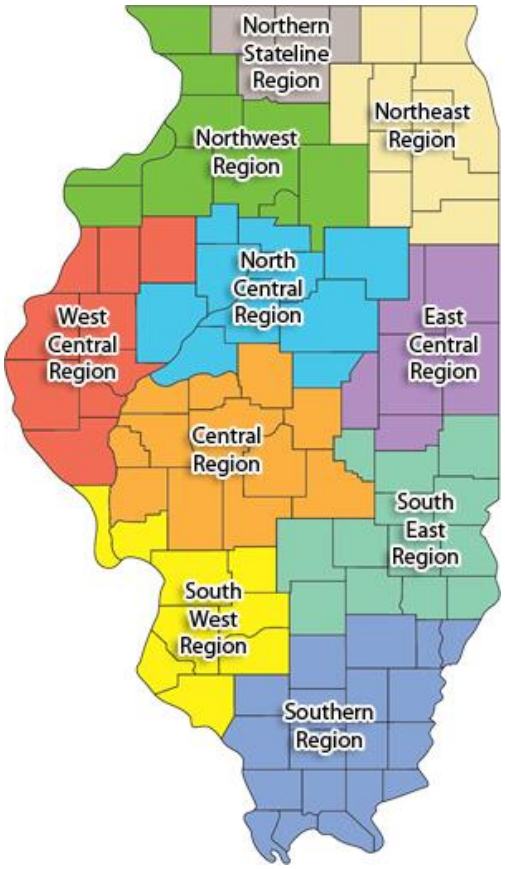

Fig. 14. Illinois Department of Commerce Economic Development Regions, Source: [33]

7 Conclusion

In this research project, we have developed a data dashboard to visualize education and workforce data for the state of Indiana. The dashboard uses a map-based visualization as the primary mechanism to access the various data fields in the dataset. This map is designed to mimic the Economic Growth Regional map that divides the state of Indiana into 12 regions based on social and economic ties. The color coding of these 12 regions is consistently maintained while designing all the other visualizations and elements of the data dashboard. This consistency of representation allows the users to smoothly view and navigate the various datasets thus simplifying their data analysis procedures.

We have found that several states have chosen to design and use such a map based on regional breakdown. As shown in figure 14 [33], the Illinois Department of Commerce also uses a regional map to divide the counties based on Economic Development Regions. Thus, datasets in this domain can be visualized through a map-based interface similar to ours. Though not all maps are broken down by economic regions, such divisions based on logical entities are present for different areas. Developers can take advantage of them while visualizing and designing dashboards for those domains. As shown in figure 15 [34], the state of Florida is divided into 8 Nursing Regions. Thus, the data on nurse supply, demand and education can be visualized through a regional map-based dashboard designed to visually mimic the map shown in figure 15 [34].

Thus, the applicability of this design principle is wide spread and can be used to design map-based data access interfaces and dashboards.

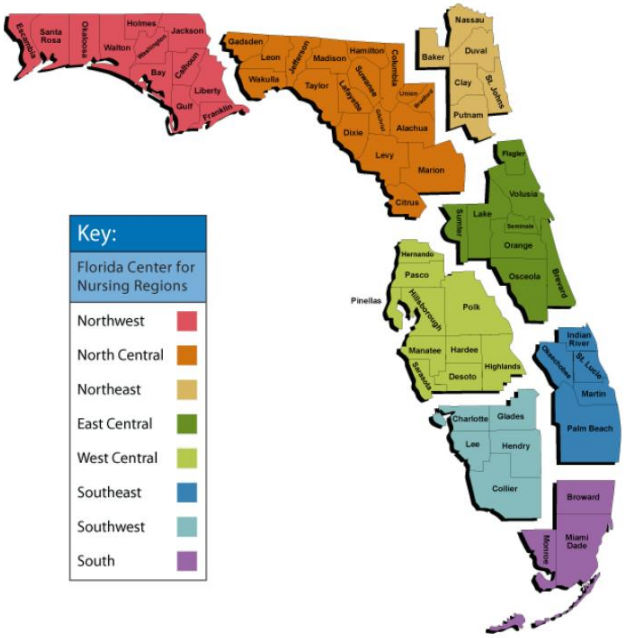

Fig. 15. Florida Center for Nursing Regions, Source: [34]

Acknowledgments

The authors would like to thank X, Y, and Z for their sponsorship and contributions to the research.

References

[1] *Find School and Corporation Data Reports*, Indiana Department of Education, Mar. 26, 2020. Accessed on: Sept. 30, 2019. [Online]. Available: https://www.doe.in.gov/accountability/find-school-and-corporation-data-reports

[2] *IDOE Data*, Indiana Department of Education, Mar. 23, 2020. Accessed on: Aug. 20, 2019. [Online]. Available: https://www.doe.in.gov/idoe/idoe-data

[3] *Workforce and Economic Data*, Indiana Department of Workforce Development, 2020. Accessed on: Aug. 20, 2019. [Online]. Available: https://www.in.gov/dwd/2329.htm

[4] *Hoosiers by the Numbers*, Hoosiers by the Numbers, 2020. Accessed on: Sept. 30, 2019. [Online]. Available: http://www.hoosierdata.in.gov/

[5] *STATS Indiana: Indiana's Public Data Utility*, STATS Indiana, 2020. Accessed on: Sept. 30, 2019. [Online]. Available: http://www.stats.indiana.edu/

[6] *Regional Maps*, Indiana Department of Workforce Development, 2020. Accessed on: Jan. 20, 2019. [Online]. Available: https://www.in.gov/dwd/2653.htm

[7] *Economic Growth Regions*, Indiana Nonprofits Project, 2020. Accessed on: Feb. 14, 2019. [Online]. Available: https://nonprofit.indiana.edu/community-profiles/egr.html

[8] D. Miller, *SB Admin 2*, Start Bootstrap, 2020. Accessed on: Aug. 24, 2019. [Online]. Available: https://startbootstrap.com/themes/sb-admin-2/

[9] Department of Workforce Development Research and Analysis division, *INfographics*, Hoosiers by the Numbers, 2020. Accessed on: Sept. 30, 2019. [Online]. Available: http://www.hoosierdata.in.gov/infographics/index.asp

[10] J. McCormick, *IDOE Home*, DOE Compass, 2020. Accessed on: Sept. 30, 2019. [Online]. Available: https://compass.doe.in.gov/dashboard/overview.aspx

[11] T. Munzner, *Visualization analysis and design*. CRC Press, 2014.

[12] *Public Schools and Districts Data Files*, California Department of Education, May 28, 2019. Accessed on: Mar. 10, 2020. [Online]. Available: https://www.cde.ca.gov/ds/si/ds/pubschls.asp

[13] S. Few, *Information Dashboard Design: Displaying data for at-a-glance monitoring*. Burlingame, CA: Analytics Press, 2013.

[14] A. Sarikaya, M. Correll, L. Bartram, M. Tory, and D. Fisher, "What do we talk about when we talk about dashboards?" *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 1, pp. 682-692, Jan. 2019.

[15] W. Grant, 101 UX Principles: A definitive design guide, Packt Publishing, 2018.

[16] R. Ceniceros, "Dashboard technology helps select measure, respond to risks" *Business Insurance*, vol. 43, no. 16, pp. 21-21, 2009.

[17] D. Dowding, R. Randell, P. Gardner, G. Fitzpatrick, P. Dykes, J. Favela, S. Hamer, Z. Whitewood-Moores, N. Hardiker, E. Borycki et al., "Dashboards for improving patient care: review of the literature", *International journal of medical informatics*, vol. 84, no. 2, pp. 87-100, 2015.

[18] J. Starmer and D. Giuse, "A real-time ventilator management dashboard: toward hardwiring compliance with evidence-based guidelines", *AMIA Annual Symposium Proceedings*, vol. 2008, p. 702, 2008.

[19] M. B. Morgan, B. F. Branstetter, D. M. Lionetti, J. S. Richardson and P. J. Chang, "The radiology digital dashboard: effects on report turnaround time", *Journal of digital imaging*, vol. 21, no. 1, pp. 50-58, 2008.

[20] L. J. Stinson, J. H. McMurray and C. B. Eaton, "E-health blood pressure control program", *The Journal of medical practice management: MPM*, vol. 28, no. 2, p. 91, 2012.

[21] R. Crooks, "Representationalism at work: dashboards and data analytics in urban education", *Educational Media International*, vol. 54, no. 4, pp. 289-303, 2017.

[22] A. Faiola, P. Srinivas, and J. Duke, "Supporting clinical cognition: a human-centered approach to a novel ICU information visualization dashboard", *AMIA Annual Symposium Proceedings*, vol. 2015 pp. 560–569. American Medical Informatics Association, Sept. 2015

[23] B. A. Wilbanks and P. A. Langford, "A Review of Dashboards for Data Analytics in Nursing", *CIN: Computers, Informatics, Nursing*, vol. 32, no. 11, pp. 545–549, Nov. 2014.

[24] D. Filonik, R. Medland, M. Foth, and M. Rittenbruch, "A Customisable Dashboard Display for Environmental Performance Visualisations", *International Conference on Persuasive Technology*, vol. 7822, pp. 51–62. Springer Berlin Heidelberg, 2013.

[25] K. O'Connell, Y. Lee, F. Peer, S. Staudaher, A. Godwin, M. Madden, and E. Zegura, "Making public safety data accessible in the westside Atlanta data dashboard", *Bloomberg Data for Good*, pp. 1–7. New York, Sept. 2016.

[26] L. D. Roberts, J. A. Howell, and K. Seaman, "Give me a customizable dashboard: personalized learning analytics dashboards in higher education", *Technology, Knowledge and Learning,* vol. 22, no. 3, pp. 317–333, 2017

[27] S. Wexler, J. Shaffer, and A. Cotgreave, *The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios*. Hoboken, NJ, USA: Wiley, 2017.

[28] Z. Qu and J. Hullman, "Keeping multiple views consistent: constraints, validations, and exceptions in visualization authoring", *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 468–477, 2017

[29] F. Marx, J. H. Mayer, and R. Winter, "Six principles for redesigning executive information systems—findings of a survey and evaluation of a prototype", *ACM Transactions on Management Information Systems*, vol. 2, no. 4, pp. 1–19, 2012

[30] *Indiana Manufacturing Competitiveness Center*, Indiana Next Generation Manufacturing Competitiveness Center, 2019. Accessed on: Jan. 20, 2019. [Online]. Available: https://www.purdue.edu/in-mac/

[31] M. Bostock, V. Ogievetsky and J. Heer, "D^3: Data-Driven Documents" *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2301-2309, 2011.

[32] J. Liu, T. Tang and W. Wang, "A survey of scholarly data visualization" *IEEE Access*, vol. 6, pp. 19205-19221, 2018.

[33] *Illinois Department of Commerce Economic Development Regions*, Illinois Department of Commerce and Economic Opportunity, 2020. Accessed on: Mar. 10, 2020. [Online]. Available: https://www2.illinois.gov/dceo/SmallBizAssistance/RegionSpecificAssistance/Pages/default.aspx

[34] *Regional Data - FCN Regional Workforce Reports*, Florida Center for Nursing, 2020. Accessed on: Mar. 10, 2020. [Online]. Available: https://www.flcenterfornursing.org/RegionalData/FCNRegionalWorkforceReports.aspx