EXPLORING ECOSYSTEMS IN INDIANA'S EDUCATION AND WORKFORCE DEVELOPMENT USING A DATA VISUALIZATION DASHBOARD

by
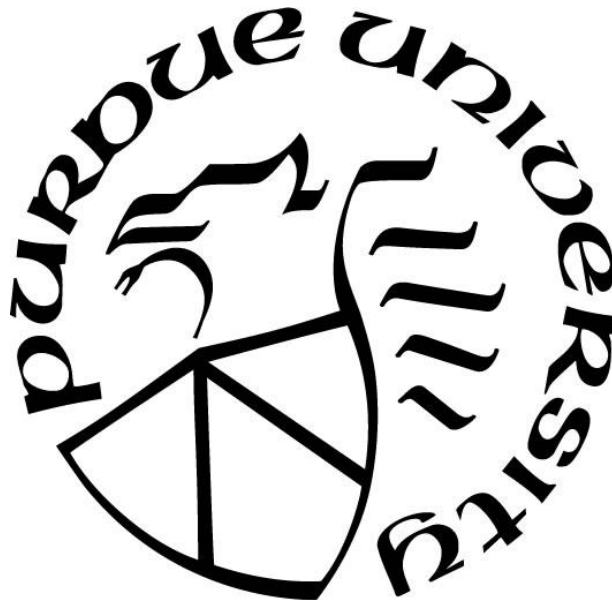
Yash Gugale

A Thesis

Submitted to the Faculty of Purdue University
In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Computer Graphics Technology

West Lafayette, Indiana

May 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Vetria Byrd, Chair

Department of Computer Graphics Technology

Dr. Tim McGraw

Department of Computer Graphics Technology

Dr. Paul Parsons

Department of Computer Graphics Technology

Approved by:

Dr. Nicoletta Adamo-Villani

ACKNOWLEDGMENTS

I would like to thank all my committee members, Dr. Vetria Byrd, Dr. Tim McGraw and Dr. Paul Parsons who have helped me by providing valuable feedback about my research topic, methodology and user studies and directed me to valuable resources that have helped me to conduct this research and write this thesis document. I would also like to thank Sascha Harrell, Education and Workforce Development Director, IN-Mac, for giving me an opportunity to work on this research project. Lastly, I would like to thank my parents and family back in India and my Purdue family consisting of faculty, staff and friends who have provided constant guidance and support during this period. I dedicate all my work to your tireless efforts.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CSS   Cascading Style Sheets

DWD   Department of Workforce Development

D3   Data Driven Documents

ERIC   Education Resources Information Center

HTML   Hyper Text Markup Language

IDOE   Indiana Department of Education

IEEE   Institute of Electrical and Electronics Engineers

IN-MaC   Indiana Manufacturing Competitiveness Center

SVG   Scalable Vector Graphics

VLAT   Visualization Literacy Assessment Test

ABSTRACT

Large datasets related to Indiana's Education and Workforce development are used by various demographics such as stakeholders and decision makers in education and government, parents, teachers and employees of various companies to find trends and patterns in the data to better guide decision-making through statistical analysis. However, most of this data is scattered, textual and available in the form of excel sheets which makes it harder to look at the data from different perspectives, drill down and roll up the data and find trends and patterns in the data. Such data representation does not take into account the inherent characteristics of the user which can affect how well the user understands, perceives and interprets the data.

Information dashboards used to view and navigate between visualizations of different datasets, provide a coherent, central access to all data, and make it easy to view different aspects of the system. The purpose of this research is to create a new data visualization dashboard for visualizing education and workforce data and find which design principles are applicable while designing such a dashboard for the target demographic in the education and workforce domain. This study also aims at assessing how the introduction of such a data dashboard affects the work processes and decision making of stakeholders involved in education and workforce development in the state of Indiana.

User studies consisting of usability testing and semi-structured interviews with the stakeholders in education and workforce development in the state of Indiana is conducted to test the effectiveness of the dashboard. Finally, this research proposes how a regional map-based dashboard can be used as an effective method to design a data dashboard for education and workforce data for other states and other domains as well.

CHAPTER 1.   INTRODUCTION

This chapter introduces the research questions, problems addressed, significance and the purpose of this research study. It also defines the scope and limitations of the research.

1.1. Background

Education and workforce data consist of a variety of datasets such as schools and principals, pre k-12 and post-secondary education data, businesses and industries, enrollment, attendance, etc. Such data about the state of Indiana is collected every year and is available publicly on the Indiana Department of Education (Find School and Corporation Data Reports, 2019) (IDOE Data, 2019), The Indiana Department of Workforce Development (Workforce and Economic Data, 2019) and other websites. Various datasets in this domain are related to each other and finding relationships and dependencies between them can greatly aid decision making.

Data visualization tools are used to find patterns, trends and outliers in large datasets (Parsaye & Chignell, 1993). A variety of such tools are available ranging from simple tools that require minimum to no programming effort such as Microsoft Excel, Tableau, to tools that require significant programming effort such as D3.js, Bokeh and others. To make interactive visualizations on the web, D3.js and Tableau are the preferred tools (Liu, 2018). However, D3.js provides much more flexibility and lower level control to produce interactive visualizations.

In this research, the goal is to create and test the design of a new information dashboard to best represent education and workforce data through a combination of visualizations, text fields and interaction mechanisms to extract meaning from the data. The design of the dashboard is influenced by best practices from the relevant literature (Grant, 2018) and evaluated by conducting studies (Shneiderman & Plaisant, 2006) with the target demographic.

1.2. Problem Statement

There is a huge amount of data about Indiana's education (IDOE Data, 2019) and workforce (division, n.d.) development such as student enrollment, attendance, end-of-course assessment, general school information, graduation, SAT scores, ACT scores, Indiana college readiness, commuting patterns, employment status and many more. 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. The problem addressed by this research study is to be able to see the ecosystem (relationships and dependencies) that exists in these large datasets that are represented through a data visualization dashboard and enable the users to make sense of and navigate through these large datasets.

There is no visual aid to help the users find patterns, trends, outliers, etc. in the data. Further, they are not able to drill down the data to find a specific field that they are interested in. Managing large excel sheets is hard. If the data updates, the sheets need to be downloaded again and distributed among all the stakeholders. This makes managing the data extremely difficult. Further, these stakeholders make use of this data to take decisions for education and workforce development. These decisions greatly affect the lives of individuals and their jobs. Lastly, since data collectors do not know what information is already recorded and what is not, they end up collecting the same data repeatedly, thus wasting time, effort and money.

The ecosystem in the dashboard will help to identify how data is related and influences each other. The users will be able to view the data visualized in the form of maps, bar charts, scatterplots and other visualization techniques needed to make sense of the data. They can find patterns, perform searches and view statistics about the data and perform other relevant operations as well. Further, since this dashboard will be accessible from anywhere on the

internet, data collectors will be aware of the data that is already recorded, thus solving the issue of data redundancy and duplicity.

Lastly, the dashboard design is targeted towards a specific end user and so decisions about visualization and interactivity mechanisms must be made to address the work processes and perspective of that user. This research aims to conduct usability testing (Rubin & Chisnell, 2008) and qualitative user studies consisting of semi-structured interviews (Shneiderman & Plaisant, 2006) to test the dashboard design and better understand the abilities, work processes and uses of the target end users respectively.

1.3. Significance of the problem

Education and workforce are significant areas that contribute to an individual's economic and personal growth and improve his/her standards of living (A skilled workforce for strong, sustainable and balanced growth: A G20 training strategy, 2010) . Decision makers that influence the future state of the development of these domains need to have an accurate and multi-perspective view of the current state and need to see how such data is inter-related. This can help them in making important decisions such as directing funds, providing training and educational resources, merit-based scholarships, etc. that improve the development of these fields. Since a lot of this data is textual and scattered, it makes the decision-making process tedious and limited.

Different data visualization techniques can help us address this issue. This research aims to visualize the education and workforce domain specific information through a dashboard so that the users can see distributions, part-to-whole and hierarchical relationships, correlations, trends, connections, overlays and compare fields. Since such datasets and end users are common, the research proposes that the regional map-based dashboard design and guidelines from the user

study can serve as a model on which other states can build similar frameworks to solve the challenges in education, workforce and other domains.

1.4. Research questions

Through this research, the aim is to develop a data dashboard designed specifically to address the challenges with education and workforce data. The following questions are addressed through this research:

1. Which data visualization principles/guidelines are applicable while designing a regional map-based data dashboard for stakeholders in education and workforce domain in the state of Indiana?

2. How does an information visualization dashboard affect the work processes and decision making of stakeholders involved in education and workforce development in the state of Indiana?

1.5. Definitions

1. Data visualization: "The use of computer generated, interactive, visual representations of data to amplify cognition". (Card, 1999, p. 6)

2. Data dashboard: "A visual display of data used to monitor conditions and/or facilitate understanding". (Wexler, 2017, p. xiv)

3. Visualization literacy: "Visualization literacy is the ability and skill to read and interpret visually represented data in and to extract information from data visualizations". (Lee, Kim, & Kwon, 2017, p. 552)

4. Ecosystem (Operational definition): A complex network of interconnected entities which are related, depend on and influence each other and operate within the boundaries defined by their nature and characteristics.

1.6. Assumptions

The proposed research makes several assumptions which can be categorized as:

1.6.1. Assumptions about the user

This research assumes that the user of the system can use and navigate a website. They know how to use the basic interaction mechanisms such as drop-downs, sliders, checkboxes, scrolling, clicking and so on. Further, it is assumed that the user can understand and interpret statistical analysis terms such as mean, median, mode, etc. and has a basic understanding and background about the datasets that are visualized. The user also knows how to interpret simple visualizations like bar charts, maps, scatterplots, grouped bar charts, etc. This dashboard also aims to provide a video that shows the user how to view, navigate the ecosystem and how to interpret the results and visualizations. So, the research assumes that the user is familiar with watching videos.

1.6.2. Assumptions about the data

The data is assumed to be in CSV, JSON or text format for creating the visuals. Even though the visualizations will be able to work with data that is updated, it is assumed that the datasets are static. There is a need to clean, pre-process and add additional fields in the datasets so that they can be used for creating the required visualizations. Some additional datasets will be created using the existing ones to support different visualizations that can aid the development of the system. Most of the data is publicly available but a few datasets will be provided by IN-MaC (research sponsor) which will be used to demonstrate how other organizations can impact education and workforce in a state. In order to show data with missing fields, the display text boxes will show the entries as is (or they will be replaced with a suitable placeholder such as 'Not Readily Available'), so that the users will be able to easily find out which data needs to be collected for evaluation.

1.6.3. Assumptions about the user study

This research will perform usability tests (Rubin & Chisnell, 2008) on the design of the dashboard with students from Purdue University. Qualitative studies (Shneiderman & Plaisant, 2006) consisting of semi-structured interviews will be conducted with the target demographic consisting of stakeholders in education and workforce and assumes that it represents the ideal users (population) of the system. The data collected using interviews and usability tests will be assumed to have no error and the questions proposed in the user studies will in no way direct the users towards a specific answer.

1.7. Limitations

The dashboard cannot be dynamically updated and customized. Thus, the users will not be able to move the positions of the different elements nor can they add additional elements on the fly. The datasets cannot be updated through the interface as well. Further, the user studies do not represent the entire demographic population. This research will only use a small sample from it. This will introduce some bias in the studies. The dashboard design will not be ideal and further improvements in its design can be made based on future research and understanding of user centric design. Even though the research aims to build a stand-alone dashboard, references to external data and reports may be needed while making decisions and gaining insights which depends on the task the user is performing.

Further, the dashboard design and visualizations created to address the current problems are selected based on the literature study (Kirk, 2019) (Few, 2013). After analyzing the data collected from usability studies and interviews, recommendations for design guidelines and areas requiring further investigation are made and further studies need to be conducted to test them.

The results and guidelines are based on the results of this study only. There may also be other designs that may be suitable to address the problem.

1.8. Delimitations

To narrow down the scope of the study, this research only uses 10 datasets related to education and workforce from among the many ones that are available. Further, the research uses D3.js as the sole visualization tool since it provides the maximum flexibility and ability to create new visualizations. The qualitative user studies are conducted on the stakeholders in education and workforce. This study used only 8 participants for conducting the interviews (Yin, 2015). The usability testing part of the study is conducted on 20 users which are students from Purdue University (Rubin & Chisnell, 2008).

The research does not design multiple visualizations to represent the same data and make comparisons between their effectiveness in performing the designated tasks. Such studies will significantly broaden the scope and goal of this research.

1.9. Summary

This chapter explained the motivation and purpose behind this research study. The assumptions on various levels and limitations of the research are also mentioned. Key terms and concepts necessary to understand the topic have also been defined. The significance of the research and the results is also mentioned in brief while addressing the various research questions.

CHAPTER 2.   REVIEW OF RELEVANT LITERATURE

This chapter discusses the various tools and libraries to be used to create interactive visualizations. It also introduces previous work done in the visualization of education and workforce data and the various dashboard designs and principles. Various tests created to evaluate visualizations consisting of quantitative and qualitative studies and the effectiveness of several visualizations to solve certain challenges and perform tasks as also discussed here.

2.1. Methodology

IEEE Transactions on Visualization and Computer Graphics and IEEE Access were the primary source for the literature on visualization techniques, literacy and design tests (Lee, Kim, & Kwon, 2017) (Sarikaya, 2018) (Liu, 2018). Different visualization methods and techniques such as bar graphs, scatterplots, line charts, treemaps, node-link diagrams, pie charts were also studied from (Munzner, 2014), (Kirk, 2019) and (Murray, 2017). Other literature was obtained through databases such as Engineering Village, Abi/Inform Collection, IDOE, ERIC by using key terms related to education and workforce data visualization. Literature on development of dashboards was obtained through (Sarikaya, 2018) other books (Few, 2013), (Wexler, 2017). (Carpendale, 2008) and (Shneiderman & Plaisant, 2006) were the primary sources for user studies in information visualization.

2.2. Data Visualization tools

Several data visualization tools are available to aid the user in creating visual representations of the data (Liu, 2018). These tools vary in terms of the programming proficiency needed to create different visualizations (and the programming languages used to create them)

and the types of visualizations that can be created. Some of the most popular ones are (Liu, 2018) Google Chart Tools, D3.js, Chart.js, JavaScript InfoVis Toolkit, Dimple, Tableau, Microsoft Excel, IBM Watson Analytics, Many Eyes, Plotly and Datawrapper. A few common ones are presented here.

"D3 is a novel representation-transparent approach to visualization for the web." (Bostock, Ogievetsky, & Heer, 2011, p. 2301) D3 is a JavaScript library that uses SVG elements to create visual marks (Munzner, 2014) and sets various properties on those elements to represent their channels.

Tableau is another data visualization library mostly suitable for beginners and non-programmers. However, the ability to create new innovative visualizations in Tableau is limited.

Cytoscape is an open-source software platform originally designed for visualization of biological research and is now commonly used for network analysis and visualization. (Shannon, 2003)

Microsoft excel provides the functionality of creating static charts using the data in excel sheets. Since they are not interactive and the visualizations cannot be deployed on the web, this is a poor choice for this research.

Bokeh is another interactive visualization library that uses Python for creating different visualizations. It is targeted for the web. However, since D3 allows us to create visualizations directly for the web, this research prefers it over Bokeh.

2.3. Visualization of Education and Workforce data

The datasets from Indiana's Department of Education will be used to create some visualizations. Several of these visualizations will be created to address various goals and needs of a specific end user. Currently, some visualizations of some of these datasets is available on the

internet (division, n.d.). Most of these do not have a consistent visualization dashboard. Some

others are created as individual visualizations on a webpage with different interaction

mechanisms (McCormick, n.d.). They are poorly designed (unjustified 3D, chart junk, incorrect

color scales, no labels for the x and y axes) based on the visualization principles mentioned in

(Munzner, 2014) Also, the filtering mechanisms are broken. Here, the user is not able to find

relationships and inter-dependencies within the data as well. Hence, there is a clear need to have

a better visualization design that makes the interpretation of the data easy and clear.

2.4. Data visualization dashboards

Recent surveys in data dashboards have categorized dashboard designs into two types,

functional genre and visual genre (Sarikaya, 2018). The visual genre consists of "a structured

tiled layout of simple charts and/or large numbers" (Sarikaya, 2018, p. 683). On the other hand,

the functional genre is "an interactive display that enables real-time monitoring of dynamically

updating data" (Sarikaya, 2018, 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

(Grant, 2018).

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 (Few, 2013). Similar precautionary guidelines are also

mentioned in (Few, 2013) 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 (Ceniceros, 2009), decision support and management

control (Dowding, 2015), ventilator management systems (Starmer & Giuse, 2008), radiology

(Morgan, 2008), blood pressure monitoring (Stinson, 2012). Such applications of dashboards are

countless. A data dashboard was created to address the issue of representationalism in the

domain of urban education (Crooks, 2017). 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 also to support rapid analysis

of real-time clinical data-trends (Faiola, 2015). 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 (Wilbanks, 2014). 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 (Filonik,

2013) 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 (O'Connell, 2016). 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 (Roberts, 2017) 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.

Thus, static dashboards for organizations, dashboards designed for strategic and operational decision making, communication, quantified self, operations/alerting (Sarikaya, 2018) and many more are being developed everyday to address the problems in various domains.

2.5. Dashboard design principles

Standard references for general dashboard design principles are (Few, 2013) and (Wexler, 2017). "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, 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" and so on are some of the commonly used and widely applicable dashboard design guidelines (Few, 2013).

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 (Faiola, 2015). 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 (Wilbanks, 2014). Some dashboards allow users to choose their

own datasets and support custom visualizations (Filonik, 2013). 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 (Roberts, 2017). 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 (Qu & Hullman, 2017). Six design guidelines for developing information systems are given in (Marx, 2012). They are: "design a comprehensive information model, tailor EIS functions to better analyze and process information, improve easy to use IS handling, setup a more flexible IS architecture and data model, design for more accurate and consistent information with proper information management and implement prototypes faster" (Marx, 2012). A prototype is 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. In this research, one of our goals is to identify the relevant principles and find the context in which they are applied.

2.6. User studies

It is important to know the ability of the users to understand and interpret the different types of visualizations (Lee, Kim, & Kwon, 2017). This will help the research to make sound choices while selecting the best visual representation to depict the underlying data. Previous research has developed tests to evaluate this literacy (Lee, Kim, & Kwon, 2017). Three levels of graph comprehension are used to test the users understanding about primitive graphs (Börner, 2016) (Carswell, 1992) (Wainer, 1992). The first is "the elementary level in which the user can read a specific value in the graph like the height of a bar in the bar chart. The second is the intermediate level in which the user can compare two values and lastly, there is the advanced level in which the reader is able to predict a future trend in the data" (Lee, Kim, & Kwon, 2017). The design of a data dashboard must take into consideration the users work processes and his characteristics as well as ability to use the dashboard. Methods to evaluate data visualizations and tools are mainly classified as quantitative and qualitative methods (Carpendale, 2008). Quantitative methods consist of controlled experiments where the effect of a single feature of the visualization is tested for effectiveness among two or more samples. The same set of users can also be given two or more treatments and the differences can be compared. This requires a careful design to ensure that the observed effect on the dependent variables is due to independent variables and not influenced by confounding variables. Also, a large sample size is needed to validate and generalize the results of such a study. Further, since such experiments are conducted in controlled environments, they lack realism. Interviews, observations and diary studies are conducted for qualitative methods where the main goal is to gain a better understanding of usefulness of the visualization tool in real life scenarios (Carpendale, 2008). Usability tests are also conducted to study a user's interaction with the tool (Rubin & Chisnell, 2008). Such tools are tested with actual users and the impact of the tool in the work processes are of prime

importance. The protocols for usability testing were developed with the help of (Bank, 2014). An in-depth understanding of the end user, his work practices, collaboration and decision making is needed to design such tools. Since this research aims to design a data dashboard for stakeholders in education and workforce, qualitative evaluation methods are best suited.

2.7. Summary

Thus, the various types of visualizations used to represent large datasets are discussed and the tools to create them are mentioned. A brief explanation of information dashboards and their design guidelines is also provided. However, since this research is focusing on a specific demographic, deployment platform and interests of the stakeholders, there is a need to create a data dashboard that addresses the challenges faced in these domains. The Indiana Department of Workforce Development categorizes the state of Indiana into 12 regions to help identify the counties that comprise the Economic Growth Regions (EGR) (Regional Maps, 2019). The choice of marks, channels and their properties for creating visualizations and various elements of the dashboard will be made to aid this end user in better understanding the visual representation of the data and to maintain a consistency in their working space.

CHAPTER 3.   METHODOLOGY

In this chapter, the focus is on the tools that will be used in the research, the research type and the development procedure. Then, it is followed by a discussion on the experiments and user studies, the population and sampling. Further, the datasets to be used and the variables in this research are specified. The main purpose of the research is to assess the impact of introducing an information visualization dashboard in the work processes of stakeholders involved in education and workforce development in the state of Indiana. Based on the data collected from user studies, guidelines applicable for the design of a data dashboard, interaction and filtering mechanisms and visualizations will be found which can be used to create dashboards for similar data concerning other states and domains as well.

3.1. Development tools
3.1.1. D3.js

D3 (Data Driven Documents) (Bostock, Ogievetsky, & Heer, 2011) is a data visualization library create by Mike Bostock. D3 is written in JavaScript and makes use of SVG elements to represent the various visual marks. Hence, the quality of the visualizations does not decrease on increasing the resolution of the display. Further, there is an extensive and comprehensive documentation about the D3 API (Bostock M. , 2019). Lastly, there is a huge support community and several examples available online for creating common visualizations which greatly contributed in the development of this research project.

3.1.2. Bootstrap

Bootstrap (Bootstrap, 2019) is a CSS styling and formatting library developed to create beautiful layouts and elements that follow material design guidelines. This research uses a dashboard template from startbootstrap.com (Miller, 2019) as a starting point to simplify the designs.

3.2. Summary of methods

This study will perform thematic analysis (Braun & Clarke, 2006) which is used in evaluation of information visualization tools where the research questions will be answered through qualitative data analysis. This research is divided into two parts. The aim of the first part is to design a data dashboard based on the best practices of design obtained through the relevant literature on dashboard design (Few, 2013), visual perception and data visualization (Kirk, 2019). It follows by an evaluation of the design through usability testing (Rubin & Chisnell, 2008) and semi-structured interviews (Shneiderman & Plaisant, 2006). The insights gained from both the phases will then be distilled as design guidelines that can be used as references for designing data visualization dashboards for education and workforce data for other states as well. Figure 3.1 shows the methodology for conducting the research. The details for studies conducted and data collected in each study is shown in Table 3.1 and discussed below.

Literature review
- Viz principles
- Tools
- UI/UX guidelines
- Dashboard design

Dashboard design based on literature on recommended tools, viz principles, UI/UX guidelines

Dashboard design

Outcome

1. Final design, recommendations and guidelines/principles for dashboard design
2. Effectiveness of Info Viz dashboard in work and collaboration

Usability testing for each aspect of the dashboard

Semi-structured interviews to understand use of the data dashboard for making decisions and collaboration by the users

Protocols

Usability testing (screen + voice capture)

30 tasks

Semi-structured Interviews (voice capture)

20 Students from Purdue University

8 Stakeholders in Education & Workforce

Research plan

Figure 3.1. Research methodology

Table 3.1: *Overview of Study Design, Data Collection Methods and sub-parts*

Research Design	Study Design	Data Collection Methods	Sub-parts
Data dashboard design and analysis	**Study 1**: Design of dashboard and analysis of its usability for performing tasks	Usability testing	Protocols for usability testing
			Perform usability tests on various pages of the dashboard
			Analyze
	Study 2: Gaining insights of stakeholder's work processes, degree of familiarity with information visualization tools, and use of the dashboard for performing tasks, decision making and collaboration	Semi-Structured Interviews with stakeholders	Design questions to gain demographic insights and current use of information visualization tools
			Conduct interviews
			Analyze and codify data

3.3. PART 1: Data dashboard design and development

3.3.1. Dashboard template

A dashboard template was used from startbootstrap.com (Miller, 2019) as a starting point to simplify the design. The template provided is for an admin for a website. It was modified to make it suitable for the current use. The template consists of a combination of CSS, HTML and JavaScript files that were then modified to use in this project. Though the template was modified for the need of this research, it is still the backbone of the entire dashboard.

3.3.2. Dataset selection

Several datasets for this research were primarily obtained from the Indiana Department of Education (Find School and Corporation Data Reports, 2019), The Indiana Department of

Workforce Development (Workforce and Economic Data, 2019) and IN-Mac (Key Impacts, 2019) and other datasets provided by IN-MaC. The datasets consist of:

1. IN-MaC Impact data
2. IN-MaC Micro-grant data
3. Schools and Principals data
4. Pre K-12 data
5. Post-Secondary Education data
6. Adult and Workforce data
7. Business and Industry data
8. Local and Regional Plan data
9. Ecosystems data
10. School Attendance data

These datasets were selected to show the impact of an organization such as IN-MaC in the field of education and workforce through the micro-grants that it offers to various organizations for conducting education related activities. Also, the other datasets from schools, pre k-12 and so on until attendance data were selected to show the pipeline from education to workforce that individuals need to go through and how it affects their career.

3.3.3. Visualizations and other interaction mechanisms

In this research several visualizations were created to develop 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 were:

1. Proportional symbol map – to view IN-MaC impact
2. Region and county map – for selecting the counties to view corresponding data

3. Bar graph (region and 92 county)– to compare various fields

4. Line chart – to view the attendance of students in a certain school

5. Bubble chart – to compare youth impacted vs amount requested

6. Pie chart – to view proportions of schools in each region

7. Node link chart – to view ecosystems

Several elements to aid in the exploration of the ecosystem will be:

1. Text fields – 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

4. Tooltips – To view information on hovering over the various elements of the visualizations

Dashboard homepage

By default, the homepage of the dashboard shows the Impact of IN-MaC in the state of Indiana through its various activities. Thus, the visitors can immediately get an idea of the involvement of IN-MaC in various activities across the state. The home page is shown in the figure 3.2 below.
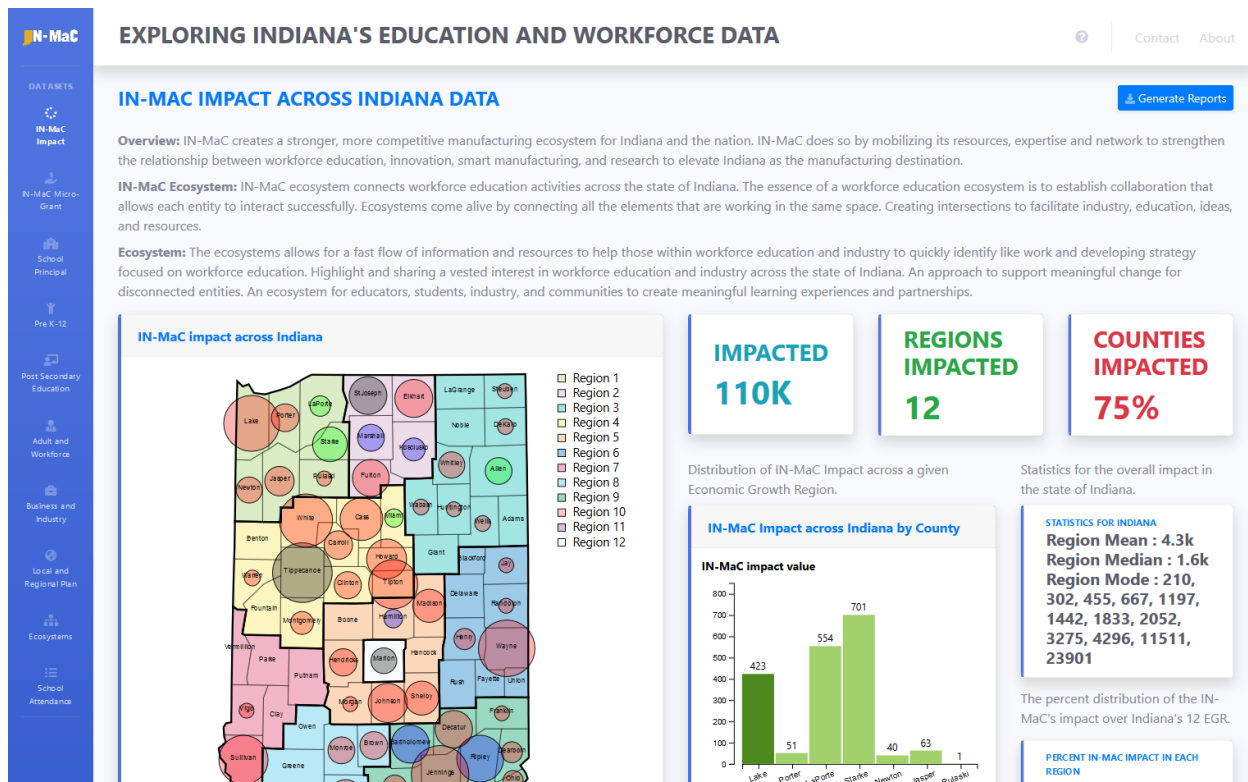
Figure 3.2. Data dashboard home page

Map visualization

The stakeholders involved in education and workforce development use the state map consisting of the 12 Economic Growth Regions (EGR) (Regional Maps, 2019) of the state of Indiana. This map is used to refer to the various development regions and areas while making certain decisions. Hence, as shown in Figure 3.3 below, the map of the state of Indiana is visualized which shows several counties. These counties are further grouped into 12 regions based on the Economic Growth Regions (EGR) (Regional Maps, 2019). Since most of the data is associated using regions and counties, it made more sense to have a map-based visualization. It uses the color channel to represents these regions. Since the stakeholders are trained to look at static maps that follow this setup (Regional Maps, 2019), this research chooses to create the visuals along the same lines to maintain visual connectivity and to help the user maintain

consistent perception of the system. They must feel familiarity with the map visualization that they see. All the other visualizations created later maintain this color-coding to ensure a consistency in the representation (Few, 2013). Another aspect is to enable the user to quickly look for information with the help of color hue, shape and spatial region channels (Munzner, 2014). The map is interactive and acts as the main navigation mechanism to view the data associated with that dataset. By clicking on a specific county on the map, the data gets updated to show the entries for that county. As show in Figure 3.3, the Tippecanoe county is currently selected, and the visualizations and information boxes show the data associated with that county.

Other visualizations

Other visualizations on the map such as the region bar graph, pie chart and 92 county bar graphs are also created to provide the user with more insights on the data. The selection of a county on the map by clicking on it with a mouse will update all the corresponding visuals with the information about the information of that county and region. The region bar graph allows the user to compare the data for a specific region. The 92 county bar graph allows the user to look at the overall distribution across the entire state. The pie chart shows the region wise data in terms of percentages. Based on the kind of dataset, additional visualizations are also created such as bubble chart, node-link diagram and line chart to aid the user in performing various tasks and gaining additional insights. These visualizations are shown in figures 3.4, 3.5 and 3.6 below.

Other features

The webpage consists of tabs as follows:

1. **Data drop down:**

   The data drop down allows the user to select the specific data entry that they are interested in. For example, for the 'School and Principals' dataset, the user can select the school that he is interested in from among the county selected (Tippecanoe county in Figure 3.3) and find the contact information of the school's principal.

2. **Information text box:**

   On selecting the particular entry from the drop down, the information card below it updates with the information about that data entry. For the 'School and Principals' dataset, it will update data about the contact information of the principal of the school that was selected.

3. **Stats text box:**

   It will contain the statistical information about the dataset such as the mean, median and mode values for each region. For the 'School and Principals' dataset, it will calculate the total number of schools in each region. Then, compute the mean, median and mode for the 12 regions and display the statistics here. This statistics box is to show statistics about the data that the user might be interested in so that it can aid him to make faster decisions. Of course, additional statistics can be added here to suit the needs of the user.

4. **Descriptive text:**

   Descriptive text boxes and titles are provided at various locations to provide more information about the data and the visualizations.

5. **About tab:**

   It provides the user with information about the dashboard and the data that is being visualized.

6. **Contact tab:**

   It provides the contact information about the developers of the dashboard, the sponsor and the supervisor, so that the user can contact them for more information or to provide feedback.

7. **Information tab:**

   The help icon (?) on the top right of the dashboard consists of a video tutorial that explains the various aspects of the dashboard to the user so that they can navigate and use it better. It is also designed to help a new user make themselves familiar to using the dashboard.

8. Generate reports button:

This allows the user to generate a report based on the current selection on the data dashboard so that they can use it to create a summary of the data or use it to prepare for a meeting. The report consists of a timestamp and all the visuals and information present on the dashboard at the time the report was generated. The report is show in Figure 3.7 below.

9. Dataset navigation pane:

The various datasets can be selected and viewed by using the navigation pane to the left side of the dashboard. The current dataset is highlighted so that the user can quickly be aware of what he is looking at. Also, navigating between the various datasets becomes easy.
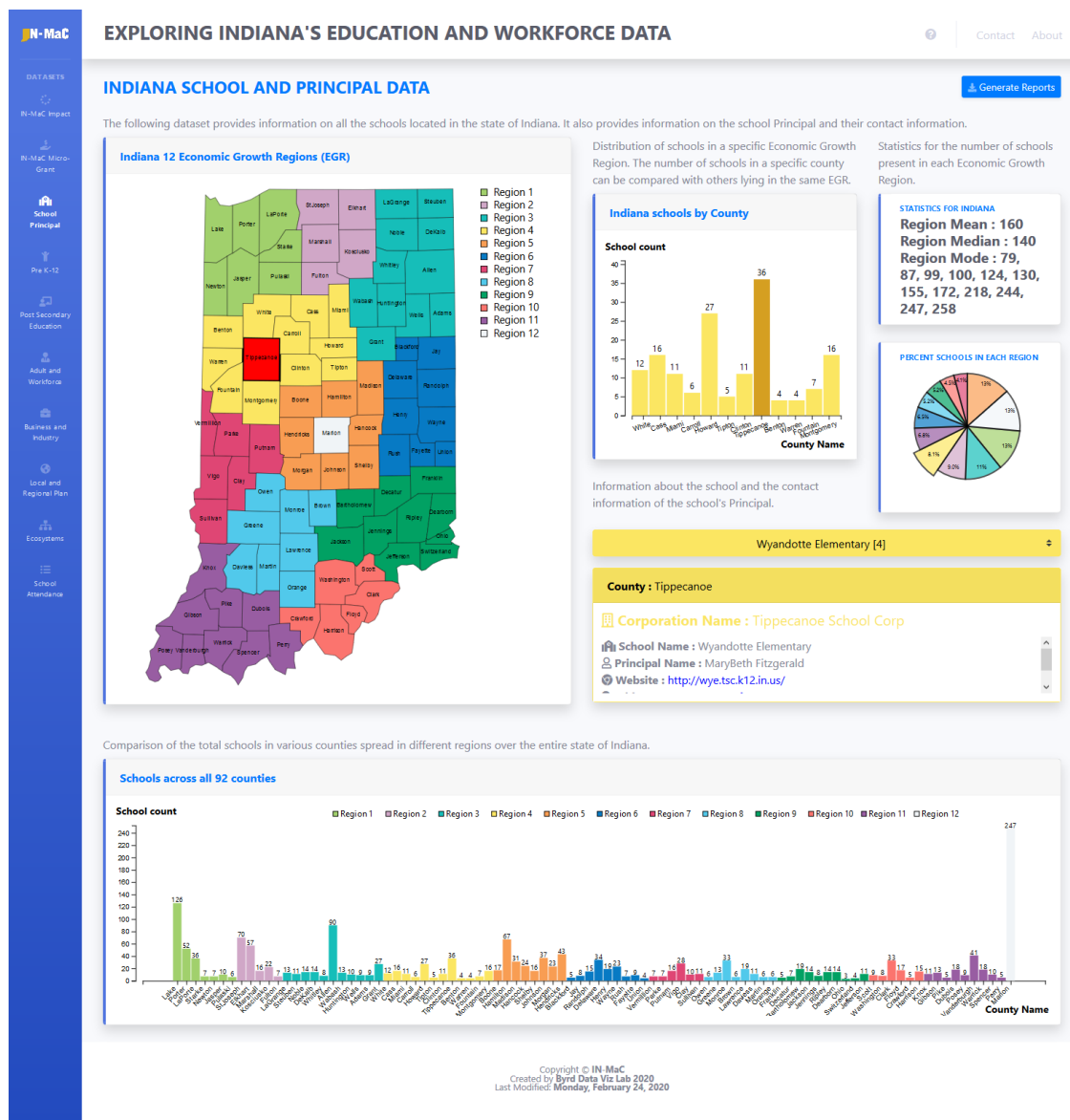
Figure 3.3. Schools and Principal dataset page

Scatterplot graph

Region

All Regions ⇕

Amount requested range:
$0.0 - $5,000.0

0.0 1.0k 2.0k 3.0k 4.0k 5.0k

Student impact goal
range: 0.0 - 5,000.0

0.0 1.0k 2.0k 3.0k 4.0k 5.0k
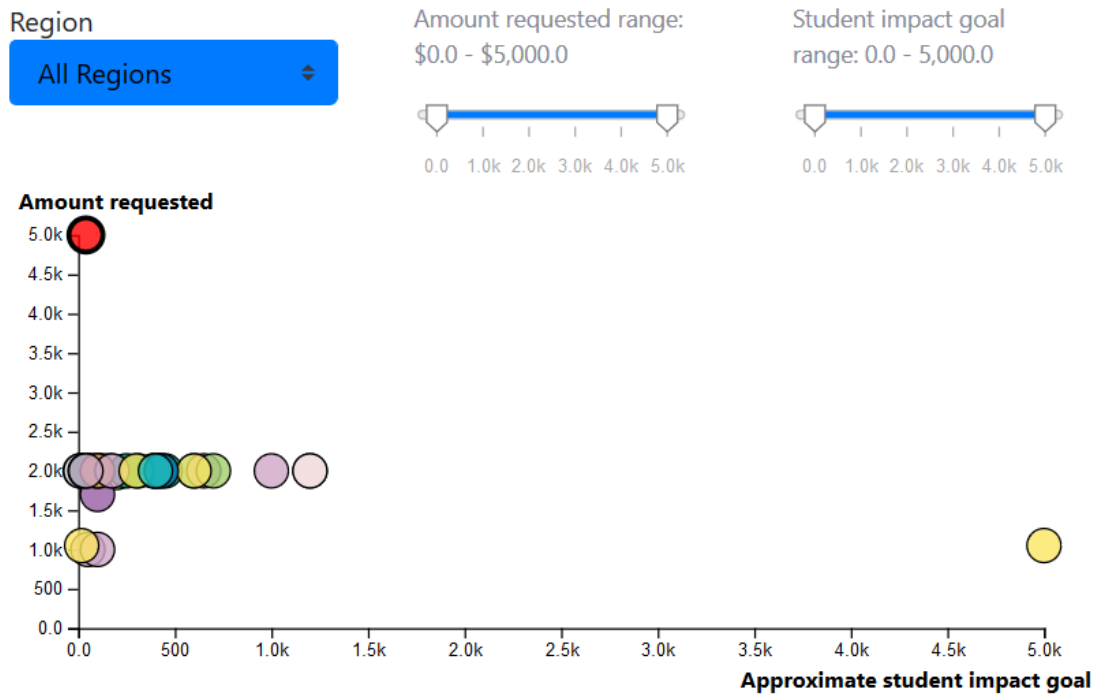

Figure 3.4. IN-MaC Micro-grant scatterplot

Indiana Attendance rate of the school over a period of years
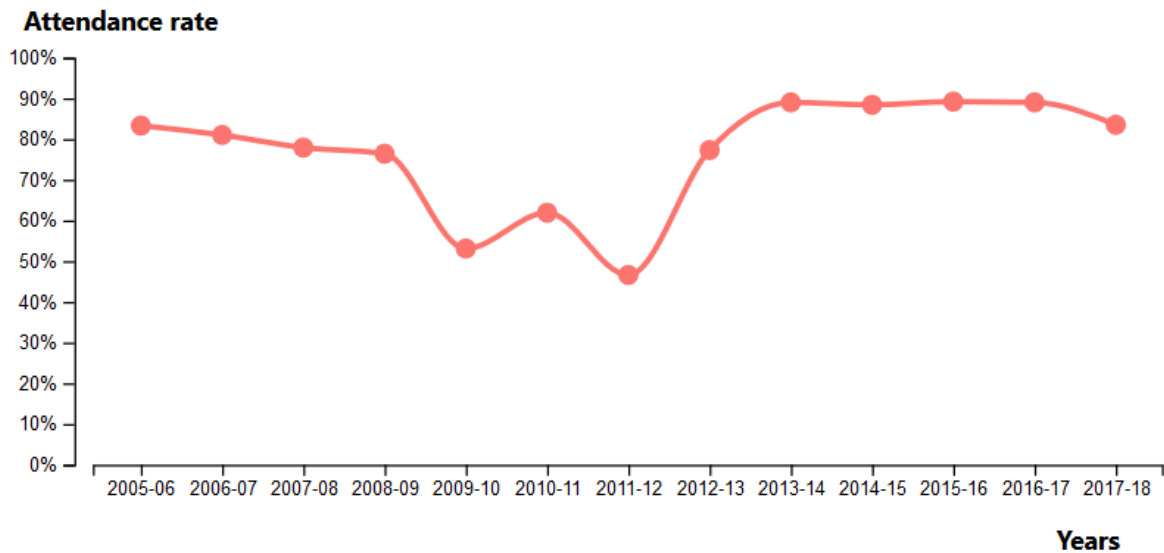
Figure 3.5. Indiana school attendance rate from 2005-06 to 2017-18 for Clark County
Middle/High School

Figure 3.6. IN-MaC Micro-grant ecosystem

The following dataset provides information on all the schools located in the state of Indiana. It also provides information on the school Principal and their contact information.

Schools and Principals in each county based on the 12 Indiana Economic Growth regions
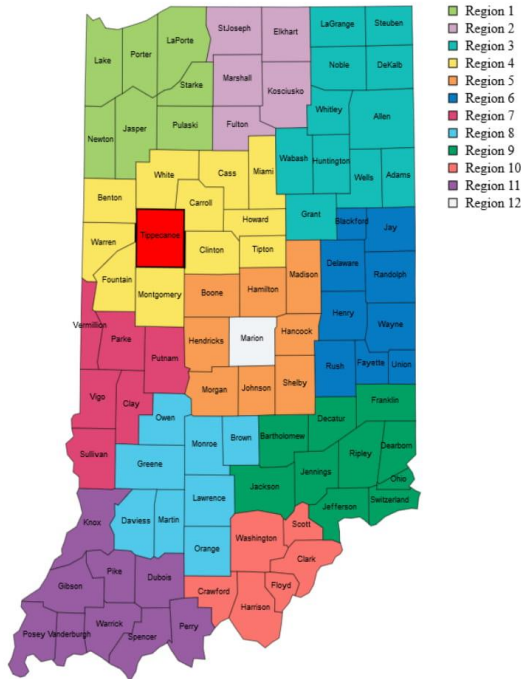
Distribution of schools in a specific Economic Growth Region
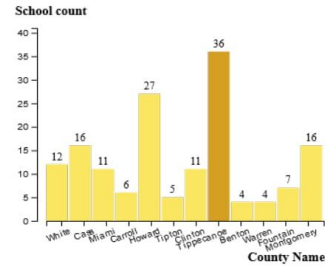
Percent distribution of total schools across Indiana's 12 EGR
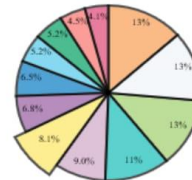
Statistics for the number of schools present in each Economic Growth Region

Region Mean : 160
Region Median : 140
Region Mode : 79, 87, 99, 100, 124, 130, 155, 172, 218, 244, 247, 258

Comparison of the total schools in various counties spread in different regions over the entire state of Indiana.
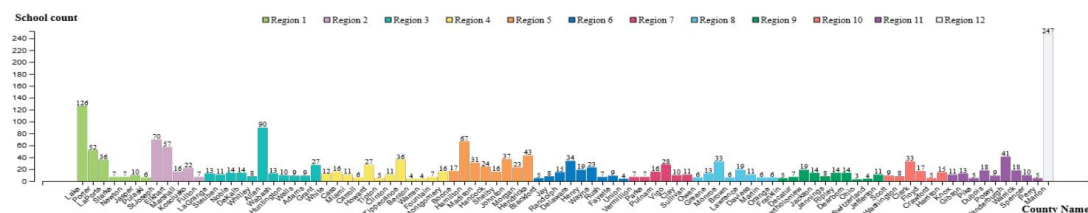
County: Tippecanoe

Corporation Name: Tippecanoe School Corp

School Name: Wyandotte Elementary

Principal Name: MaryBeth Fitzgerald

Website: http://wye.tsc.k12.in.us/

Address: 5865 E 50 S, Lafayette, 47905-0000

Figure 3.7. 1 Page report generated for Wyandotte Elementary School from School and Principal Dataset

3.4. PART 2: Data dashboard evaluation and user studies

The qualitative user studies are designed based on the in-depth analysis approach recommended in (Shneiderman & Plaisant, 2006) and (Carpendale, 2008). After the dashboard has been designed, it was evaluated by conducting the following two studies:

3.4.1. Usability testing

A usability test was conducted on 20 participants to test the design of the dashboard. The various aspects of this study are discussed below.

Goal of usability test

The goal of the usability test was to evaluate the design of the data dashboard to allow the users to perform various tasks and evaluate the design in terms of navigation ability, interaction and understanding of the elements and their use. It will also help in gaining an understanding users familiarity with the visualizations and their interpretations. Though the dashboard was designed by following good UI/UX principles (Grant, 2018), it was necessary to test its functionality with live users. The interactive map was used as the main mechanism to access and view the data. Other mechanisms such as filters, drop downs, tooltips and sliders were also used to allow the user to control the granularity of data that he/she may be interested in.

Aspects to test

The following aspects of the data dashboard were tested during the usability tests:

1. Information presentation (text)

2. Navigation

3. Visualizations

4. Filtering mechanisms

5. Interaction mechanisms

6. Miscellaneous aspects

Usability tasks

The usability test consisted of 30 tasks to be performed by the user. The tasks were created to test the usability of the various aspects as mentioned above. These tasks were sequenced in a random order so that the participant does not recognize the groups and avoids doing the tasks based on memory from previous tasks. The complete list of tasks and the order in which they were asked are given in appendix C. The testing category to which each task belongs and the rationale behind each task is also mentioned. Some tasks may fall in more than one category.

Study participants

20 students above the age of 18 years were the participants for the usability test. These participants were chosen through a voluntary sample method. Flyers and emails about the study were sent to the students from the Department of Computer Graphics Technology at Purdue University. Flyers were also printed and put up on the notice boards across various departments throughout the campus from which more participants were recruited.

Usability test study procedure

There were three main phases to the usability test as follows:

Pre-test

In this phase, the entire study procedure was explained to the participant. The participant was asked to sign the consent form, giving their consent to participate in the study. The participant's voice and actions on the screen were captured and recorded to help with the analysis of the session later. The participant was asked to fill in a pre-test questionnaire that asked about some basic demographic information and their prior familiarity with data visualization. This questionnaire is given in appendix B. Since the participants were using the dashboard for the first time, they were given 5 minutes to make themselves familiar with the navigation and use of the dashboard. The researcher did not guide the participant in any way, and they were free to use the dashboard as they liked. This was done to mimic the real-world scenario of using a new website for the first time. The participants were asked to 'think out loud' as they used the website. The entire walk through procedure is given in appendix A.

Usability test

In the usability testing phase, the tasks were read out one by one by the researcher. The participant was asked to use the dashboard to perform the task. The use of a mouse and keyboard as interaction mechanisms was allowed. For some tasks that were longer, the participant could ask the researcher to repeat the task or read it out slowly in parts. As the participant performed each task, the researcher would take notes on the successful completion of the task, places were the participant was stuck or any odd behavior that was not expected by the researcher. Such observations were recorded on another document as given in appendix D. If some questions were unclear, they were clarified by the researcher.

Post-test

After the test, the participant was thanked for their time and effort. Each participant was enrolled in a raffle for a Starbucks gift card worth $20 for their participation in the user study. After completing the tasks, the participants were asked if they had any further comments, feedback to give and notes were taken on those as well. Finally, the recording devices were stopped, and the participant was escorted out of the test lab.

Time required for the usability test

The brief and introduction to the usability test lasted for about 5 minutes. Each participant needed 5 to 10 minutes to fill in the pre-test questionnaire and about 30 to 45 minutes for performing the tasks in usability testing. Thus, the total time required was about 1 hour for each participant.

Data collected

During the usability test, the participants voice and actions on the screen were captured and recorded. The researcher also took notes as the participant performed the task. The recordings were further used to analyze the session in detail and to verify the notes and check for errors. All the voice recordings were transcribed using Otter.ai (Otter Voice Meeting Notes, 2020) which is an online transcribing software. Demographic information about the participant was collected through a pre-test questionnaire (given in appendix B).

3.4.2. Semi-structured interviews

In order to test the usefulness of the dashboard and its visualizations in decision making, gaining insight, collaboration and suitability for the target end users, semi-structured interviews

were conducted with 8 participants who were the stakeholders in education and workforce development in the state of Indiana. The various aspects of the semi-structured interviews are explained below.

Goals of semi-structured interviews

The goals of the semi-structured interviews were to gain an insight into the use of the data dashboard in real world scenarios by the actual end users of the tool. Various stakeholders in education and workforce currently use various data sources and tools to perform their tasks. With the help of this dashboard, these stakeholders can better equip themselves for decision making by making use of the various datasets and functionalities of the data dashboard. To evaluate the usefulness of the dashboard in their work processes, in depth semi-structured interviews were conducted with these users. Analysis of this data helps to answer the two research questions of this study. The results from this analysis is given in the Results section of this document.

Aspects to study

The questions on the semi-structured interviews were designed to gain an understanding of the following aspects:

1. Industry (role and industry of the user)

2. Tasks (Tasks the user performs on a daily basis and those performed using this dashboard)

3. Tools (Data visualization, analysis and other tools that the user uses in his work)

4. Data (Datasets that the user is primarily interested in)

5. Dashboard (Use of the data dashboard by the user)

6. Data visualization and analysis (Use of visualizations and analysis on the dashboard)

7. Platform (Platform that the user prefers)

8. Open Ended (Open ended discussion of any other aspect that was not covered)

Interview questions

The semi-structured interview consisted of a variety of questions to cover important aspects as mentioned in the section above. Other clarification and follow up questions were also asked to the participants to get a better understanding. The complete list of questions is given in appendix G. These questions were designed to keep the researcher on track with the aspects that were important. Some clarification was also provided to the participants if they did not understand what exactly was being asked.

Semi-structured interview participants

Semi-structured interviews were conducted on 8 participants who were stakeholders in education and workforce development in the state of Indiana. These participants were recruited with the help of Sascha Harrell, Director of Education and Workforce Development, IN-MaC (IN-MaC, 2019). Each participant was chosen by sending an invitation to participate in the study. The participants were free to choose if they wished to participate in the study or not.

Semi-structured interview study procedure

The semi-structured interviews consisted of three phases as explained below:

Pre-interview

The participants were given access to the data dashboard and were asked to use it for a minimum period of 1 week before conducting the semi-structured interview. They were not given any explanation on how to use the data dashboard. They were simply given access to the

weblink (the dashboard was deployed through GitHub pages) and asked to use it in their daily work to perform their tasks. They were free to use it for as long as they liked and for any purpose. It was necessary for the users to use the dashboard for a minimum period of time before the interview was conducted so that they could better respond to the questions on the interview. On the day of the interview, the participants were given a brief on the interview procedure. The script of the interview procedure is given in appendix E. They were asked to sign a consent form to give their consent for participation in the study. Their voice was recorded as the interview was conducted in order to analyze the sessions further. They had access to the data dashboard if they needed to explain or show something. However, the screen was not captured during this process. Before the interview, the participants were given a pre-interview questionnaire that asked about some basic demographic information and their prior familiarity with data visualization. This questionnaire is given in appendix F.

Semi-structured interview

The participants were asked various questions about their daily work processes, tasks and use of the data dashboard for solving problems. They were also asked about their current tools and what they liked and disliked about the data dashboard and so on. The entire list of questions is given in appendix G.

Post-interview

After the interview, the participant was thanked for their time and effort. Each participant was enrolled in a raffle for a Starbucks gift card worth $20 for their participation in the user study. After completing the interview, the recording devices were stopped, and the participant was escorted out of the interview room.

Time required for the study

The brief and introduction to the semi-structured interview lasted for about 5 minutes. Each participant was interviewed for about 45 minutes to 1 hour and needed 5 to 10 minutes to fill in the pre-interview questionnaire. Thus, the total time commitment was about 1 hour for each participant.

Data collected

The participants voice was recorded during the entire study procedure. This helped the researcher to analyze the session further. This data was transcribed using Otter.ai (add citation here) to perform further analysis. Thematic coding methods (Braun & Clarke, 2006), which is a widely known method for evaluating qualitative studies was used. The research questions of this study were answered using the themes that emerged from this analysis. The outcome of this analysis is explained in the Results section. Demographic information about the participant was collected through a pre-interview questionnaire (given in appendix F).

3.5. Data analysis procedure

The data collected from usability testing procedures and interviews need to be analyzed and evaluated to validate the dashboard design and answer the research questions. The various analysis methods are described here.

3.5.1. Analysis of usability testing data

After collecting data from the usability test, it will be evaluated to gain insights on the tasks completed successfully by the user, problems in usability and interaction and suggestions

for improvement. After a careful understanding of the feasibility of the implementation of these features, suggestions for future development will made.

The following statistics and analysis will be performed on the data:

Performance data

1. Task accuracy (Participants performing successfully including those who required assistance)

   a. Performance of participants on each individual task.

      i. Number of participants that performed the task successfully

      ii. Percentage of the participants that completed the task successfully

   b. Performance of participant on all 30 tasks.

      i. Number of tasks completed successfully by each participant.

      ii. Percentage of tasks completed successfully by each participant.

      iii. Average of percentage of tasks completed successfully by each participant.

      iv. Median of percentage of tasks completed successfully by each participant.

      v. Mode of percentage of tasks completed successfully by each participant.

      vi. Range of percentage of tasks completed successfully by each participant.

Preference data

1. Free-form comments and questions

2. Suggestions

The following analysis will be performed on the results from above to make further conclusions and make recommendations for improvement:

1. Identify tasks that did not meet success criteria

2. Conduct a "Source of error analysis"

3. Prioritize problems

    a. Problem severity ranking

    b. Frequency of occurrence ranking

Since, there is no comparison between two or more versions of the dashboard or two participant groups, performing inferential statistics is not necessary.

3.5.2. Analysis of semi-structured interview data

The data collected from interviews was in the form of voice recordings that were transcribed using Otter.ai. This data was analyzed to find out the themes in the data (Braun & Clarke, 2006) i.e. to perform thematic analysis. The original transcripts were first coded through open coding methods (Saldaña, 2015). This comprised of the first cycle. Then, focused coding (Saldaña, 2015) was used to further categorize the data and find the themes. The two research questions were answered using the analysis performed of the semi-structed interviews. Analysis memos were also written during different phases of the analysis to help the researcher track to track their thought process and note any useful observations.

The analysis procedure is given below:

Transcription of interviews and preliminary reading

The interviews were transcribed using Otter.ai (Otter Voice Meeting Notes, 2020). These transcriptions were then edited through the Otter.ai interface online before downloading the word files for analysis. After transcribing the data, the transcripts were read several times to ensure familiarity with the data. Some initial ideas and themes were also noted to give the researcher some idea during further analysis.

The analytic memo (Saldaña, 2015) after the preliminary reading is given below:

19th March 2020

PRELIMINARY READING: IDEAS AFTER INITIAL READING OF INTERVIEW TRANSCRIPTS

I am reading the interview transcripts to increase my familiarity with the data. I can see some key ideas, trends in several places which are related to the end users' characteristics, data and its use, dashboard design and scope for future work. They are listed as follows:

1. The end users work in different roles and positions and they perform a wide variety of tasks and some do not have a fixed day to day.

2. Having access to data in a centralized location is really helpful. It's kind of a "One stop shop". The biggest challenge is to collect and verify data from multiple sources and find contact information, which is in a way addressed by the dashboard.

3. Making sure that the data is updated and not incorrect is the biggest concern of the users. If there were problems with the validity and reliability of the data, they would start to lose trust in the dashboard and might not rely on it too much.

4. A website to view on the computer will be good to look for detailed data and a mobile app to get an overview or snapshots would be great.

5. The users are not necessarily interested in all the datasets that are currently present. There is some specific data that is useful for their use case and work which is not present on the dashboard.

6. Decision making, finding contact and activity info and using it to collaborate/partner with others (other organizations or schools/institutions), keeping up to date with the current state, and using data to back their presentation, is what the users will primarily use the dashboard and its data for.

7. The breakdown by counties and regions makes it very easy to access data.

8. Users don't want to deal with raw data. This is where the visualization helps them; to grasp and access the information, to find gaps, areas in need of development, compare data.

9. Users are interested in different levels of data - from overview to granular level.

10. Description of the dataset would also be needed since not every user will be familiar with what it means and may interpret the name of the dataset in a different way. Also, this will help to determine what the data includes and what it doesn't.

First cycle coding

The transcripts were initially coded using open coding methods (Saldaña, 2015). This coding was done manually on hard copies of the transcripts. About 300+ codes were developed during this open coding phase. After these codes were generated, they were sorted alphabetically to find codes that were similar in meaning but phrased differently so that they could be combined. After removing the duplicates and merging some codes, these codes were grouped together into similar categories. Some more codes were merged in this phase. The research questions were referred to from time to time to ensure that the focus of coding was maintained.

After the first round of coding, these codes were used to create a coding manual (codebook) to be used for subsequent coding process. The codebook was used to transfer the codes to a CAQDAS software for a more thorough analysis. NVivo (Bazeley & Jackson, 2013) was used for coding using software. The codebook is given in appendix H.

The analytic memo at this stage of coding is:

19th March 2020

SELECTING APPROPRIATE CODING METHODS

Since I am not familiar with the coding methods, I think it is better to go over all of them. After having a better grasp of the method and their use, I will create a list of methods that can help to code my data. In making this selection, it will be important to consider my research questions. I need to avoid choosing too many methods.

However, in the book (Saldaña, 2015), the author recommends using In Vivo Coding, Initial Coding, and/or Values Coding for first cycle coding of interview data. For the second cycle, Pattern Coding and/or Focused Coding is recommended (though I think Axial coding

might help). Thus, if I find any other useful coding method after going through the list, I will use it, or else I will stick with the authors recommendations.

Potential coding methods that can be useful:

1. Attribute coding.

2. Magnitude coding.

3. Descriptive coding.

4. In Vivo coding.

5. Process coding.

6. Initial coding (open coding).

7. Emotion coding.

8. Holistic coding

20th March 2020

CODEBOOK CREATION

Codebook is created after manual coding of all interviews. Then I just use 3 interviews and pick up codes from them to create the codebook. These codes are then refined. While transferring the codes to NVivo, I will use this codebook (after its refinement) to code the interview data. Since a lot of codes seem to repeat, it makes sense to create a codebook first. Also, the number of codes has grown a lot. So, condensing them will be better.

There were about 300+ codes generated from the 3 interviews. A lot of these were different ways of phrasing the same information. So, I grouped them into ideas that emerged through the preliminary readings and categories that developed after manual coding. Now, I will combine these codes to form more sophisticated and ideal codes and then use them to code the interviews in NVivo.

Second cycle coding

Focused coding method (Saldaña, 2015) was used to categorize the codes into groups to further answer the research questions and find out emerging themes in the data. The codes in the coding manual were further segregated into these categories. Again, the research questions were referred to during this process to ensure that the codification process was well informed. Codes were moved around into different categories and the categories were renamed, readjusted and modified.

Presenting the themes and categories

Themes were generated based on the categories that were developed. There were two sets of categories each addressing the individual research questions. These categories were refined to ensure that they did not overlap, and each category was distinct from the other. Categories within the same group were compared with each other and then also compared with categories in the other group to ensure that they were distinct.

The themes and categories that address the research questions are given in the Results section of this document. The codebook was rearranged as per the research questions and themes that emerge in the data so that it is easier for readers to follow the codes.

3.6. Population and sampling
3.6.1. Description of Population

For conducting the usability studies, students from Purdue University above 18 years of age were recruited through a voluntary response method. Further, to test the design of the dashboard on the end users, stakeholders in education and workforce were recruited by invitation

through IN-MaC. Such stakeholders were employees in government, no-profit organizations, research institutes, etc.

3.6.2. Sampling methods

This research used 8 users for performing semi-structured interviews. Such users will be provided through collaboration with IN-Mac, which invited users on request to be a part of the study. Usability testing for the data dashboard design consists of 20 participants from Purdue University. Since the scale of this study and potential samples were small, a voluntary response method was chosen for usability studies. Though this method introduces some bias in the results, this method was chosen due to the constraints of time and availability of the demographic.

3.6.3. Sample size

In discussions with the Statistical Consultation Service at Purdue and review of the relevant literature on qualitative studies on information visualization (Yin, 2015) (Isenberg, 2008, April), 8 participants were selected for conducting interviews. Since the user studies are qualitative in nature, this sample size is enough. For usability testing 20 participants were recruited (Rubin & Chisnell, 2008). Since there is only one phase of usability testing and the entire dashboard is being tested, such a large sample size was chosen. Further, the researchers wanted to have a higher degree of confidence in their results.

3.7. Summary

In this chapter, the development tools and research frameworks were mentioned. Information on the datasets, development procedure, research methods, user studies involved, and population and sampling were also provided.

CHAPTER 4.   RESULTS

In this chapter, the results for the research are presented. First, the results from usability testing are presented to show the evaluation of the data dashboard for use. Then, the results of the analysis from semi-structured interviews are given that help to answer the two research questions that were proposed earlier.

4.1. Usability testing results

The usability testing consisted of 30 tasks performed by the participants to test the various aspects of the data dashboard. The results from this testing are presented below.

4.1.1. Demographic information

The participants for usability testing were students from Purdue University and were from various majors, age groups and industry experience. As shown in Figure 4.1, the participants lie in the age range from 20 to 32 years with the highest density between 22 to 26 years. The average age of the participants was 24.15 years. Figure 4.2 shows the distribution of gender of the participants which shows that majority of the participants were male. The participants are from a variety of majors as can be seen from Figure 4.3. Figure 4.4 shows that majority of the participants have a bachelor's degree. These participants also have very little to no industry and data visualization experience as can be seen from Figures 4.5 and 4.6. Lastly, the relevance of education and workforce data lies between 'neutral' to 'absolutely relevant' for most participants. This can be seen from Figure 4.7. These statistics give us a general understanding of the demographic of the participants to further understand the results of our usability tests.

The choice of visualizations used to represent education and workforce data were selected based on previous studies of the commonly used visualizations (Lee, Kim, & Kwon, 2017) and from ease of visual interpretation (Kirk, 2019). Through the pre-test questionnaire (as given in appendix B), an attempt was made to understand the participants' previous familiarity with visualizations and interpretation ability. Figure 4.8 shows the various commonly used visualizations and participants ability to interpret them. Most of the visualizations with a high count are used in the dashboard. Hence, the choice of those visualizations was appropriate. Based on the meaning that needs to be extracted from the data, some more visualization types can be used from the ones that have a high popularity among the users.
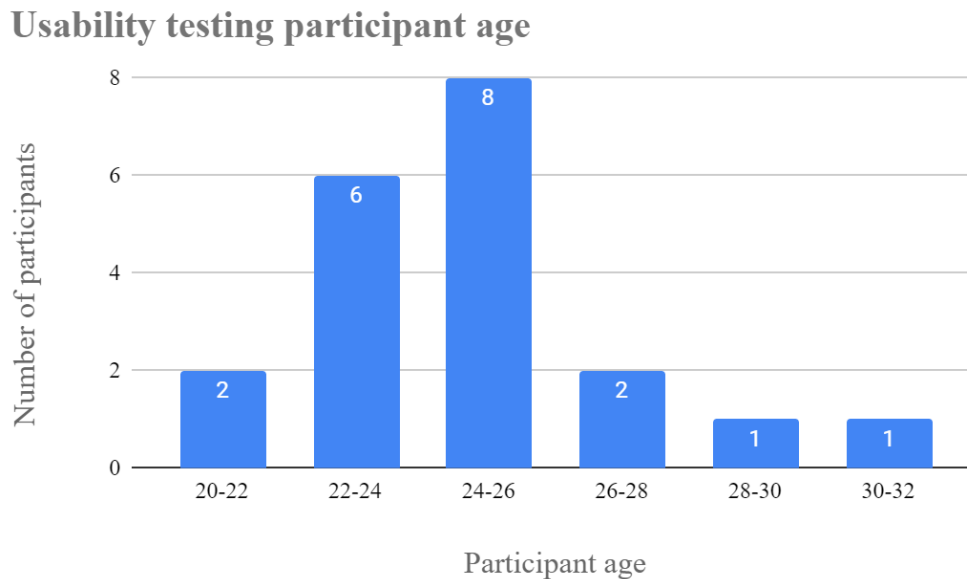

Figure 4.1. Usability testing participant age

Usability testing participant gender

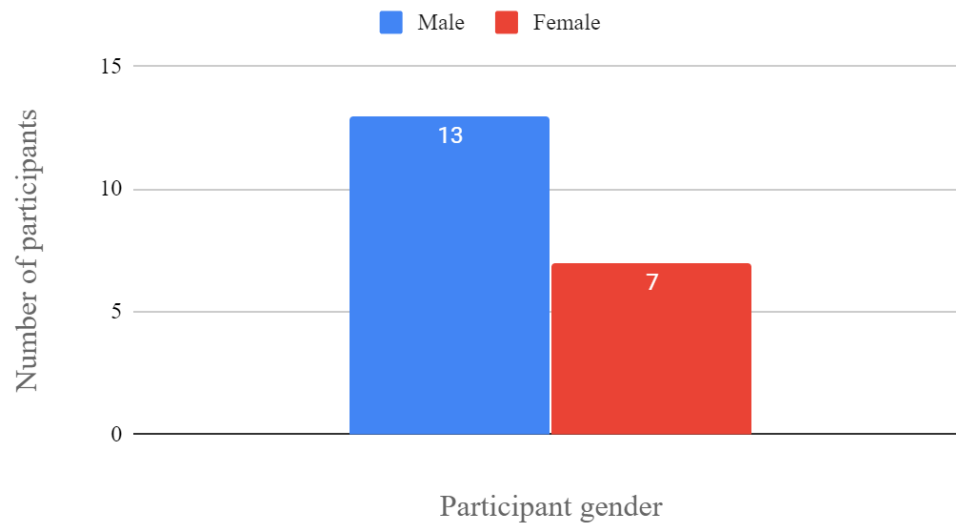Figure 4.2. Usability testing participant gender

Usability testing participant majors
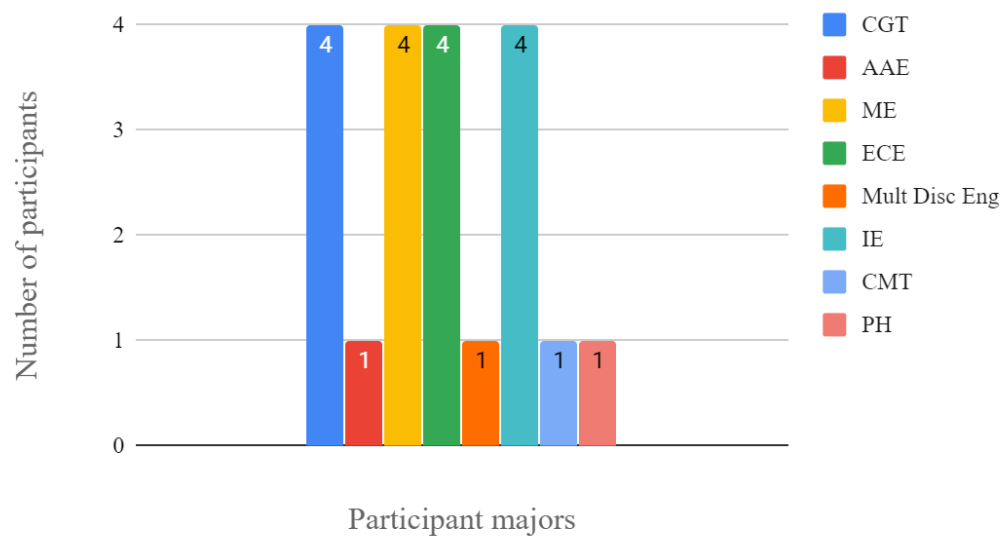
Figure 4.3. Usability testing participant majors

Usability testing participant highest level of education completed
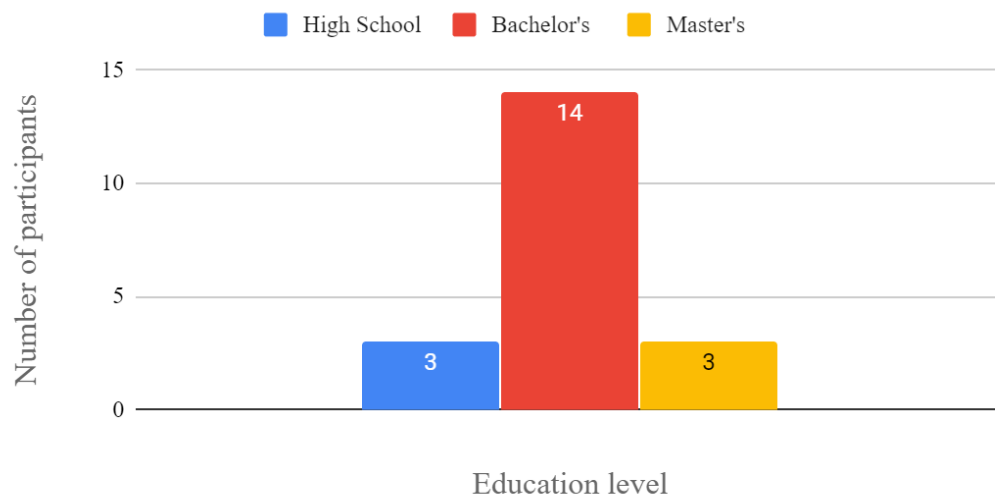

Figure 4.4. Usability testing participant highest level of education completed

Usability testing participant industry experience
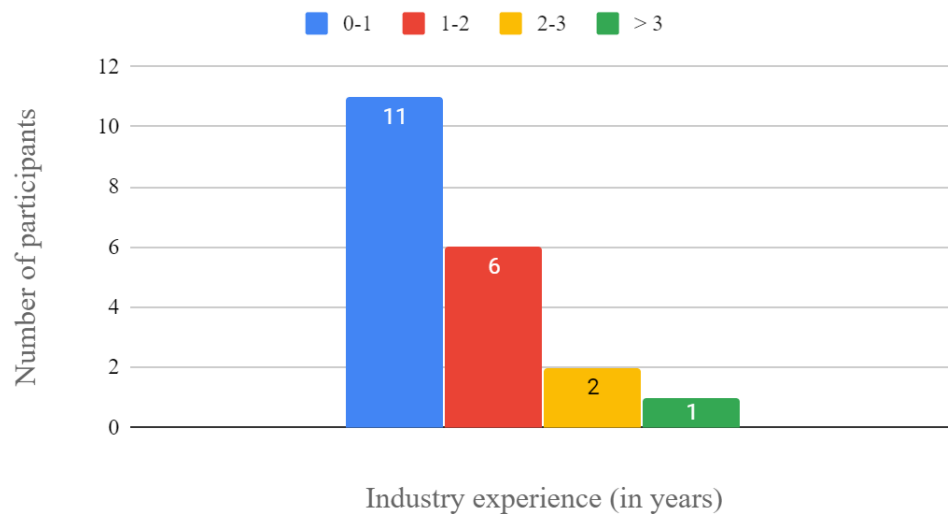

Figure 4.5. Usability testing participant industry experience

Figure 4.6. Usability testing participant data visualization experience

Figure 4.7. Usability testing participant relevance to education and workforce data
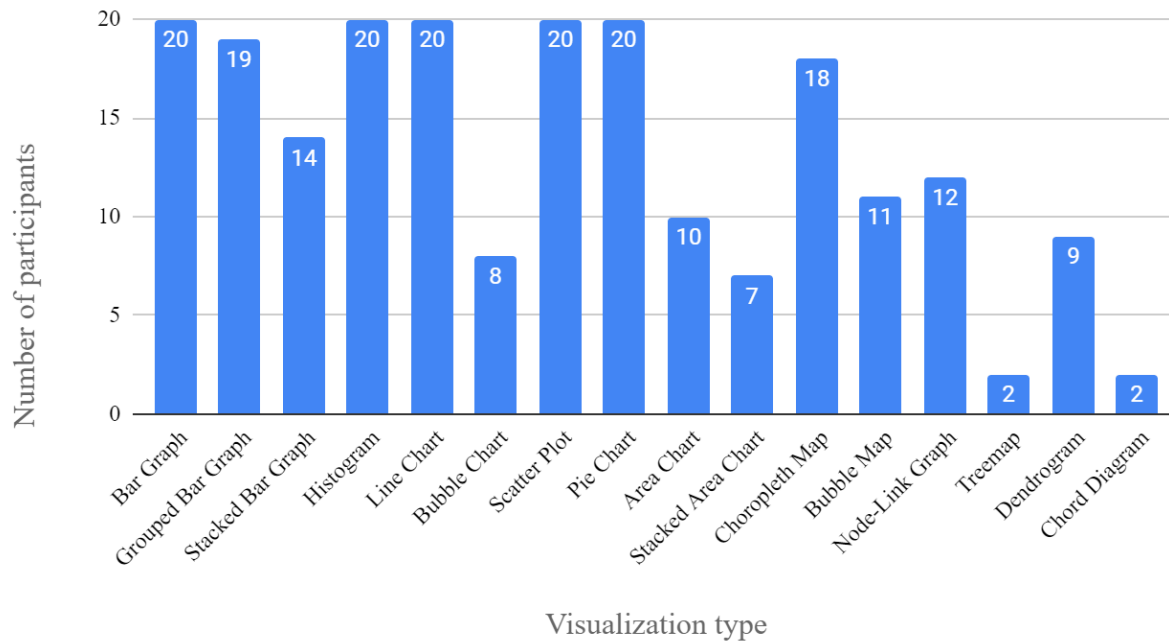
Figure 4.8. Usability testing participant familiarity with different data visualizations

4.1.2. Analysis results

As mentioned in the methodology, the following analysis was performed on the usability testing data:

Performance data

1. **Task accuracy (Participants performing successfully including those who required assistance)**

   a. **Performance of participants on each individual task.**

      i. Number of participants that performed the task successfully:
         Since the usability testing was performed on the first design of the data dashboard, the task accuracy benchmark was 70% (Rubin & Chisnell, 2008). Figure 4.9 shows the number of participants that performed the tasks successfully. Participants had the most difficulty in performing tasks 2 and 3. Task 2 focused on retrieving the number of datasets currently visualized through the dashboard and task 3 focused on using the tooltip to retrieve important

information. The cause of error by the participants in performing these tasks are further analyzed in the later sections.

ii. Percentage of the participants that completed the task successfully:
For tasks 2 and 3, the task accuracy percentage is below 70%. For all other tasks, the percentage is above 70%. The distribution can be seen in the Figure 4.10 below.

Usability testing task accuracy
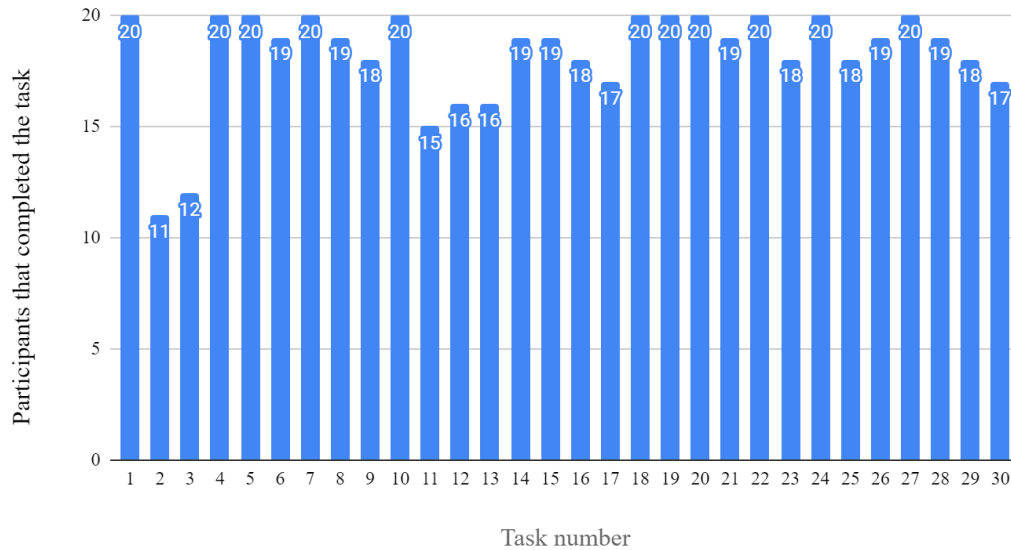
Figure 4.9. Usability testing task accuracy

Usability testing participant task accuracy percentage
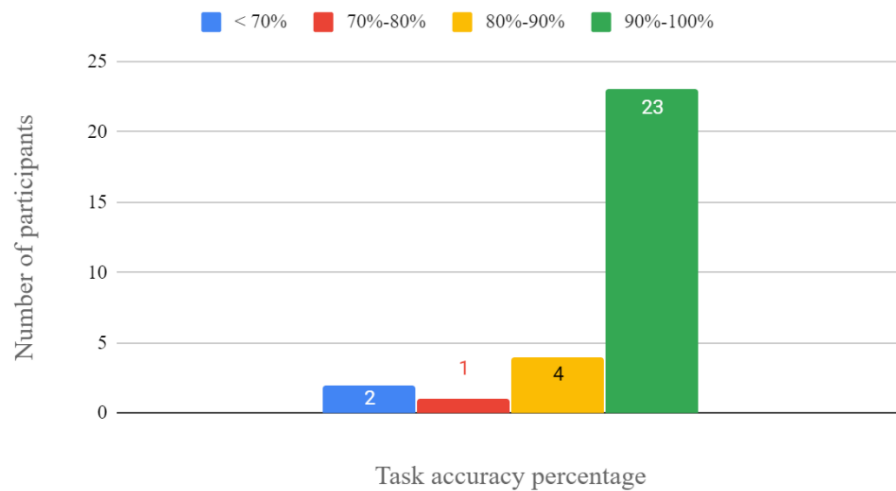
Figure 4.10. Usability testing participant task accuracy percentage

b. Performance of participant on all 30 tasks.

The number of tasks completed successfully by each participant is given in Figure 4.11. The percentage of tasks completed successfully by each participant is given in Figure 4.12. The average percentage of tasks performed successfully by the participants is 91.16%. The median of percentage of tasks completed successfully the participants is 93.33%. The mode of percentage of tasks completed successfully by the participants is 96.66%. The range of percentage of tasks completed successfully by the participants is 23.34%. Thus, from the distribution above, we can see that the general performance of the participants on the tasks was quite high ($> 70\%$). Only one participant completed 76.66% of the tasks, while all others completed 80% or more of the tasks. Two participants completed all the tasks successfully. Hence, we can infer that the dashboard is highly usable.

Usability testing tasks completed successfully be each participant
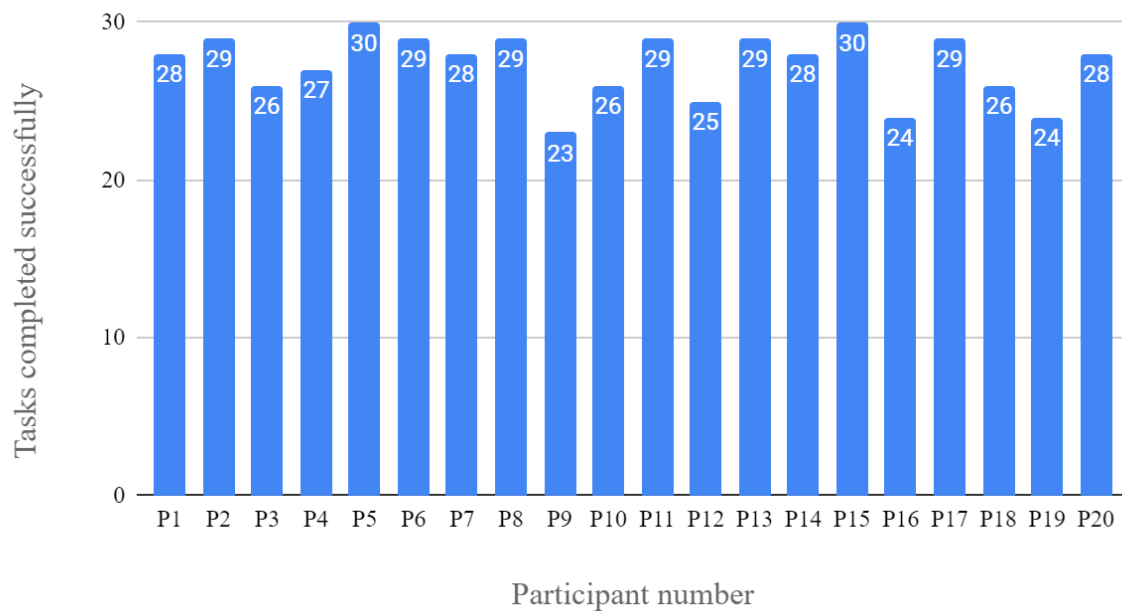
Figure 4.11. Usability testing tasks completed successfully by each participant

Figure 4.12. Usability testing percentage of tasks completed successfully by the participants

Preference data

Notes and observations were made during the usability testing procedures. Depending on how the user performed each task, some description was provided along with a label that indicates the overall conclusion about that task. The labels along with their meanings are given below:

A = Mis-interpreted and didn't know how to perform task

B = Understood but didn't know how to perform task

C = Needed some help to perform task

AE = Performed as expected

U = Performed the task correctly but in an unexpected way.

The labels above helped to segregate the tasks and perform further analysis. These are presented as follows:

1. **Free-form comments and questions:**

    a. **Features that users liked (positive comments):**

i.   Ecosystem layout is really helpful and helps to understand the grouping

ii. Interactivity is great since the data is linked to the county and all is in one place. Thus, no need to go hunting for information. Also, it is fun.

iii. Data is up to date (see Feb 2020 at the Footer)

iv. 'About tab' is straight to the point

v.   'How to' video helps since there is information overload.

vi. Report generation is great, and the report is short and to the point.

vii. Visualizations are great

viii. Highlighting the bar on the bar graph helps to know which county is currently selected.

ix. Tooltip is also nice.

x.   Uniform design and everything is visually pleasing.

xi. Colors are appealing and color coding helps.

xii. It was easy to discover that data is missing (from the info card).

xiii. Visual bubble sizes on impact map are helpful for comparison

xiv. The number in the drop down next to the name i.e. region number is good/useful

xv. Use experience is good and intuitive.

xvi. Comparison between counties in same region is easy to find out

xvii. Typeface is nice

xviii. Not very cluttered

xix. Graphs are explicit and show information clearly

xx. Choosing information is easy

xxi. Navigation is pretty good

b. **Challenges faced and features disliked (negative comments):**

i.   Close bubbles on the scatterplot are hard to read. How to select the bottom one? Thus, zooming or separating them will help.

    ii. Drop down not very intuitive.

    iii. Need to look at 3 legends to understand the circles for the impact map and it is not very clear from just looking at the map visualization.

    iv. 'Datasets' - word is a little confusing as to what exactly is being referred to.

    v. '>' icon on the attendance info card before the 'School ID' is counter-intuitive since it makes you think that something will open up.

    vi. Text on the map is small, so the tooltip helped to read.

    vii. Pie chart for Impact is too cluttered.

    viii. 92 county bar graph x-axis labels are a little difficult to read.

2. Suggestions (areas for future development and testing):

a. A search box to find the data (data can be a county name or even an entry in the drop down or possibly other pieces of information in the dataset).

    i. The popup tooltip for the Impact map should stay there for a while after hovering out.

    ii. The bar graph counties must be in alphabetical order so that it is easy to find the county.

    iii. Use of the keyboard to select multiple counties (ctrl and shift key press) and show data for them.

    iv. Randomizing the default county that is selected so that users know easily that it is clickable.

    v. Listing the ecosystem nodes on the right-hand side of the page and putting corresponding numbers on the circles.

    vi. Report generation in landscape orientation.

    vii. The text box for ecosystem organization information that scrolls up and down with your mouse.

    viii. Option to input value for range along with sliders (for bubble chart).

    ix. Animations that tell interactivity on the page for first time users.

    x. Highlight the About, Contact and Help tabs, since grey means that you cannot choose it in different software.

The following analysis was performed on the results from above to make further conclusions and make recommendations for improvement:

1. **Identify tasks that did not meet success criteria:**

    a. **Task 2: How many datasets are currently being visualized through this dashboard?**

    The percentage of task accuracy for task 2 was 55% which was below the benchmark of 70%.

    b. **What is the total impact of IN-Mac in "Tippecanoe" county and how many Technology Adoption (TA) and Education & Workforce Development (E &WD) activities were conducted there?**

    The percentage of task accuracy for task 2 was 60% which was below the benchmark of 70%.

2. **Conduct a "Source of error analysis"**

    a. **Task 2:** The main problem was participant's misinterpretation and lack of understanding of the term 'dataset'. Several users were not sure what exactly was meant by a dataset. They answered by sometimes counting the number of visualizations on the page or the total dashboard, or guessed some number based on their understanding, etc. The solution would be rephrasing the question or provide the user with information on what a dataset means in this context.

    b. **Task 3:** Some users are unable to perform this task since they do not read the tooltip properly. Some methods that may help overcome this problem are:

        i.   Make the tooltip appear while hovering over the Impact map county areas as well as the county name. Currently it appears only on hovering on top of the county name.

        ii.   On hover out over the name text, the tooltip disappears immediately. Increase the duration for which the tooltip stays for a little longer after the user hovers out

        iii.   Place this information somewhere else or find another way to present this information.

            However, it is found that in other tasks, the user easily finds information within the tooltip. So further analysis of the cause of this issue needs to be performed.

c. **Tasks 6 and 11:** For these tasks that ask how many micro-grants were awarded, some participants answered the amount and not the count. So, rephrasing of the question would be required.

d. **Tasks 16, 23 and 29:** The upper right-hand corner of the dashboard that consists of the About, Contact and Help buttons is not noticed by a few users. 2 users completely missed these and couldn't perform the tasks associated with them. Hence, they need to be redesigned and further tests need to be conducted.

3. **Prioritize problems**

   a. **Problem severity ranking:**

      i. Task 3 needs to be addressed first since users were not able to use the tooltip to find the information.

      ii. Tasks 16, 23 and 29 ask for information regarding the 'About', 'Contact' and 'Help' tabs. These are important sources for the user and hence must be addressed next.

      iii. Task 2, 6 and 11 mainly failed because the users were unable to understand the question. There are some other tasks that need to be rephrased to add clarity. These problems are not due to the lack of usability, but due to poorly designed testing procedures. Hence, they are least severe.

   b. **Frequency of occurrence ranking:**

      i. Tasks 2 – 9 users failed (low priority)

      ii. Task 3 – 8 users failed (very high priority)

      iii. Task 6 – 1 user failed (high priority)

      iv. Task 11 – 5 users failed (low priority)

      v. Task 16 – 2 users failed (high priority)

      vi. Task 23 – 2 users failed (high priority)

      vii. Task 29 – 2 users failed (high priority

Even though in some tasks the users failed more, they were not very high priority i.e. did not stop the user from performing a major functionality. Also, for some of these tasks, the testing aspects were covered in different tasks (such as using sliders and drop downs to filter data).

Hence, even though they might have failed in one such task, they successfully completed the

same functionality in another task. Thus, giving this task a low priority.

c. Other interesting observations:

i. For the task that tests the use of the ^ icon at the bottom right corner to go back up to the top of the webpage, the user is asked to not use any scroll bar on the page or the middle mouse scroll button to perform the task. The expectation is that the user will use the ^ icon on the web page to perform the task. While most users were able to do this as expected, some users used the Up-arrow key on the keyboard to perform the task which was not anticipated by the researcher.

ii. For the tasks that ask about the attendance data, the goal was to test if the users were able to navigate properly and use the filters to select the required data and then interpret the values from the line chart. However, the attendance dataset also consists of a table that displays the attendance of the school that the user is currently looking at. Now, one of the main purposes of data visualization is to avoid users from looking at large datasets. But it may be beneficial if the users look at small subsets of textual information from the large dataset. Hence, in this task, the researchers wanted to test whether the users will read the data from the line chart only or will use the table as well. It was found that users generally use both to perform the task and sometimes use one or the other. Hence, adding a small table with raw textual information might also be useful while displaying such data on a dashboard.

iii. For the tasks that ask to find the region in which Brown county lies, some participants used the Ctrl + F keys to search for the name by typing in the keyboard and used the browser's search box to highlight the county on the map. These same users also used the alphabet key to look for a specific school in the attendance data drop down by name (since there were a lot of options in the drop down) and navigate there to find further information.

iv. One user was unable to notice the nav pane on the left side of the dashboard. During development, the screen size of the laptop was medium and so the dashboard elements were placed in such a way that the nav pane was close to the main content. However, during usability testing, a much wider screen was used which introduced a small gap between the contents in the center and the nav pane to the side. This may have caused the user to only focus on the center of the screen and not look at the edges.

v. For some users it was clear that the map is clickable since Lake county was highlighted in red. Their assumption was further verified when they clicked on a different county and saw the data update.

4.2. Semi-structured interview results

The semi-structured interviews were conducted on 8 stakeholders in education and workforce development in the state of Indiana. Each interview lasted from about 30 minutes to 1.15 hours. The results from this study helped to answer the research questions and are presented below.

4.2.1. Demographic information

The participants for semi-structured interviews were experts with various levels of industry experience. As shown in Figure 4.13, the participants lie in the age range from 25 to 65 years (one participant preferred not to fill in the Age information on the pre-interview questionnaire). The mean age of the participants was 38.42 years and the median was 36 years. Figure 4.14 shows the distribution of gender of the participants which shows that majority of the participants were female. But the difference is not too big. The participants are from a variety of majors such as engineering, management, public relations, etc. Majority of the participants have a bachelor's degree as can be seen from Figure 4.15. These participants also have very high industry experience as shown in Figure 4.16. The data visualization experience is almost uniformly distributed as can be seen from Figure 4.17. The relevance of education and workforce data lies at 'absolutely relevant' for all participants. These statistics give us a general understanding of the demographic of the participants to further understand the results of our interviews.

The participants are also familiar with most of the visualizations as can be seen from Figure 4.18. Participants have low familiarity with the Treemap, Dendrogram and the Chord diagram and so further studies need to be conducted to find whether or not to use them while presenting the data.

Semi-structured interview participant age
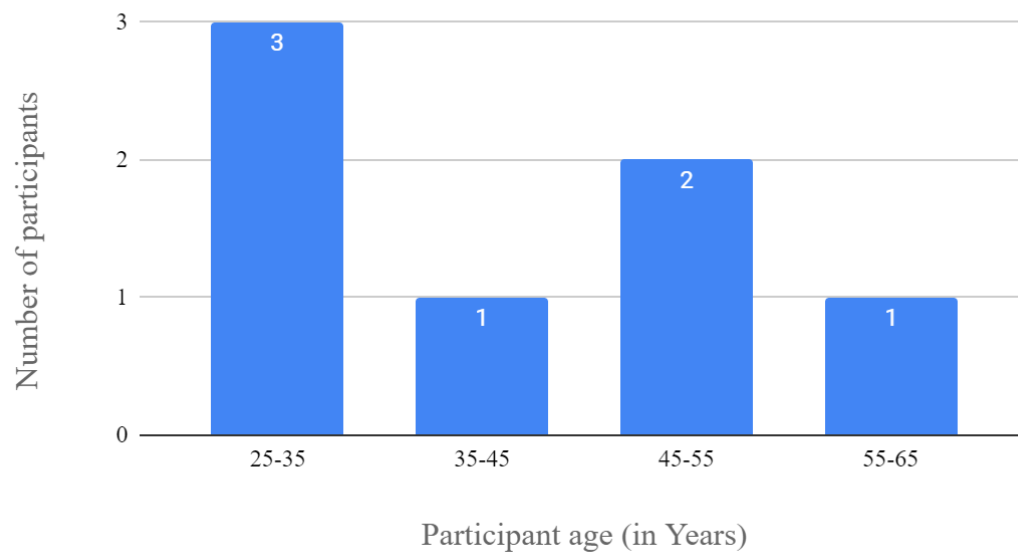
Figure 4.13. Semi-structured interview participant age

Semi-structured interview participant gender
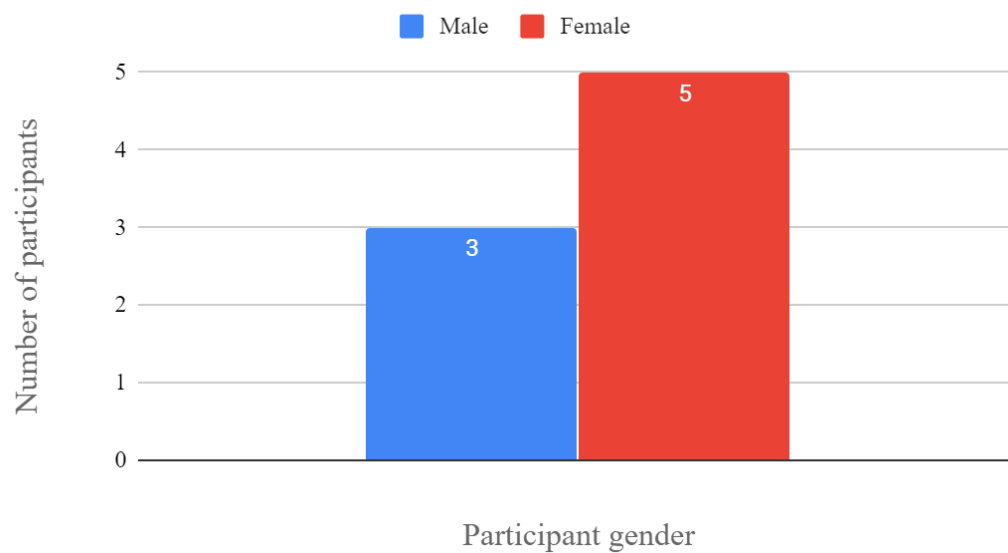
Figure 4.14. Semi-structured interview participant gender

Figure 4.15. Semi-structured interview participant highest level of education completed
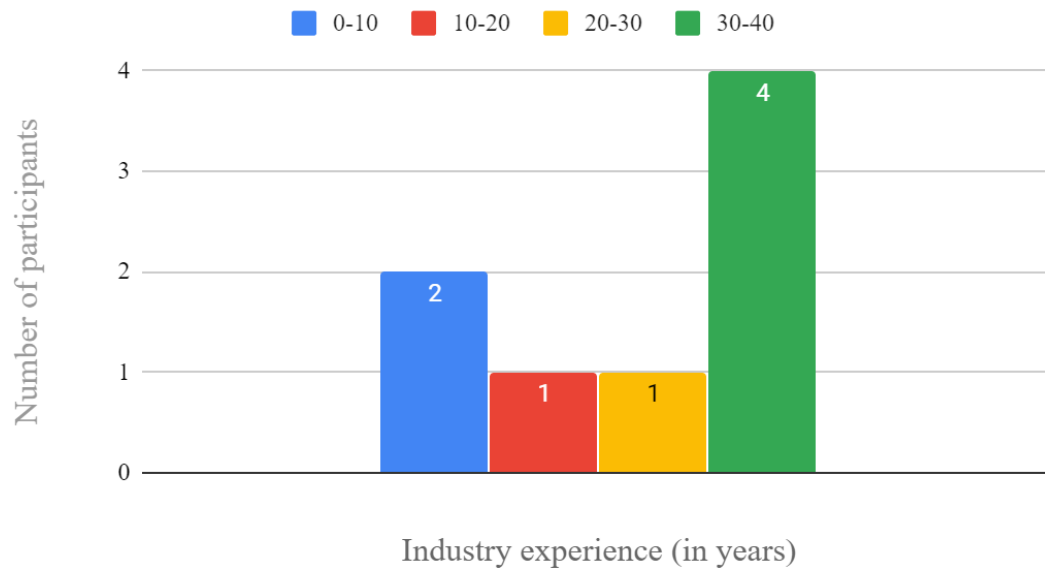

Figure 4.16. Semi-structured interview participant industry experience

Semi-structured interview participant data visualization experience
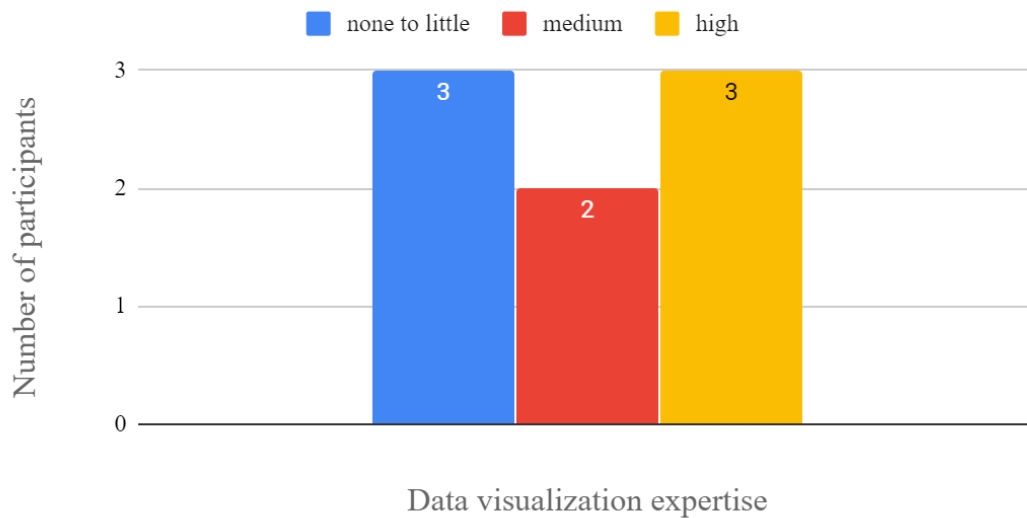

Figure 4.17. Semi-structured interview participant data visualization experience

Usability testing participant familiarity with different visualizations
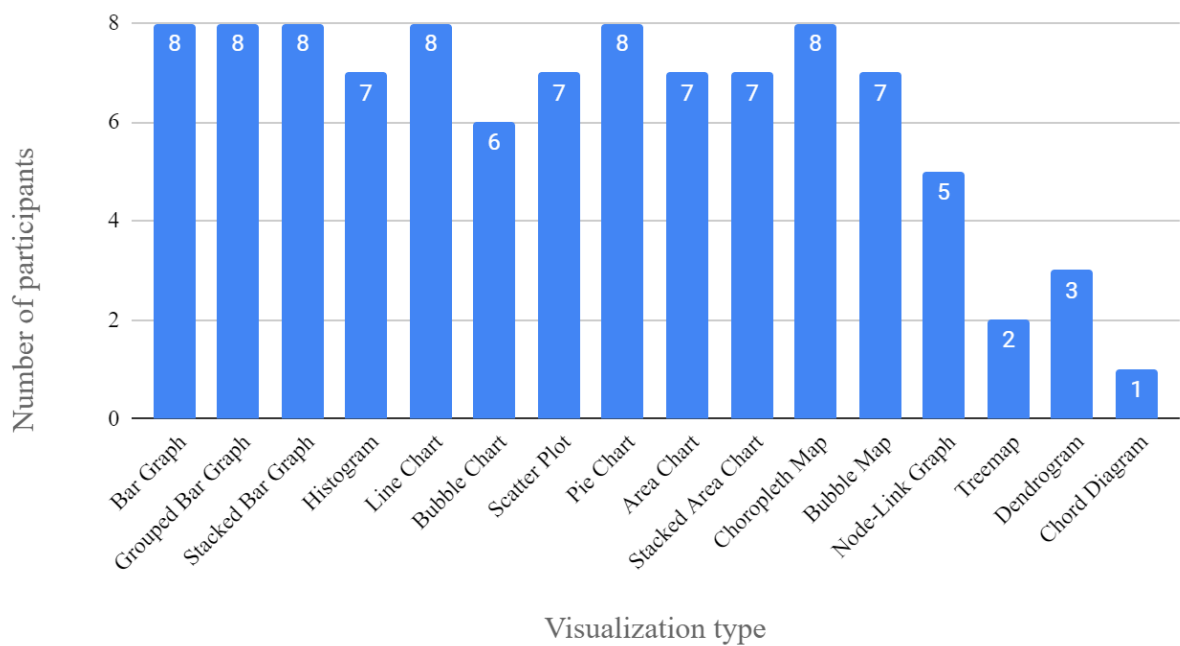
Figure 4.18. Semi-structured interview participant familiarity with visualizations

4.2.2. Analysis results

After performing level 1 and level 2 coding of the data, a general understanding of the work processes of the user was developed. The dashboard is used by the user for various activities throughout this process. This work process and use of the dashboard is shown in Figure 4.19 below and explained further. This process will help to better understand the context in which the research questions are addressed.
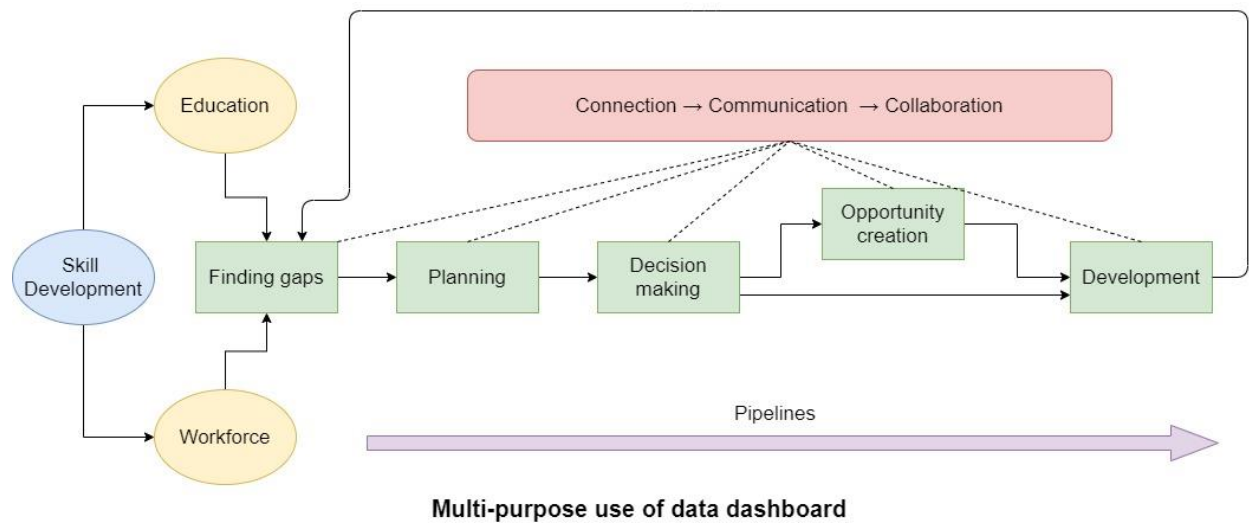
Figure 4.19. User work process and use of data dashboard

1. Skill development

Various activities (such as competitions, camps, tours, etc.) are organized by the users to improve the education level of the students. The students are given knowledge on computer programming, manufacturing practices, business development, VR, etc. through these activities. They are taught how to build a robot, and then program the robot to do the things that the employees do. They work on the manufacturing floor and they do different supply chain activities. So, they will take parts from a supplier to the manufacturer who's going to build something. They may pick up wiring to take it to a manufacturer that builds wiring harnesses, to take it to Caterpillar to put on the engine and then take it to the hospital as a generator to restore power. Students are learning the entire process of how the community works together, and why it's important that we have all of these things in the community. They learn about leadership by playing different games. They are learning how to communicate and how to work together. Current educators are also trained in the emerging technologies by partnering with various institutes. The users also help new students with onboarding processes. Gaps in education and workforce are found through the data on this dashboard.

2. **From finding gaps to development**

Gaps in the industry, education, employment, etc. are found across the various regions in the state of Indiana. The users have data that breaks down to a granular level to find out the number of accounting jobs, for example, and the number of graduates. They pull from their own system to have actual data on, say, 'this is the number of people who have completed in the last five years that you put out in the community, and these are the ones needed, but based on the enrollment, there is a gap'. Thus, in this way, the users are able to find out various gaps in the community and industry.

Bridging these gaps require creating the necessary curriculum, training, marketing the skills of graduating students, etc. The education and workforce data help to find the gaps in demand versus supply and leads to better planning activities. The map helps the user, or their bosses get up to speed with what's going on in a given initiative and who's involved. They mainly plan for equity - making sure that different economic sectors are getting a fair shake out and that not all of the statewide appropriations are hit toward a specific region (say region four). The stakeholders plan where to assign funds, grants and what steps need to be taken to address the gaps.

Further, it leads to decision making based on data rather than opinion or bias. This creates opportunities for development in those weaker areas. This cycle is started again by finding gaps in the state of education and workforce.

3. **Three C's**

Throughout the entire cycle from finding gaps to development, the dashboard helps to connect, communicate and collaborate with various organizations and institutes. It provides the necessary contact information, statistics and other relevant data to guide decision making. By clicking on the counties, they are able to see under one drop down, all the institutes that are present. This is huge to be able to see what's out there. Then they can find information about the activities of these organizations. Thus, connecting and communicating with them becomes easy. They would like to see which organizations are close by and if they are collaborating and supporting each other instead of working in silos. If not, then they can connect to form collaborations on certain events.

4. **Pipelines**

The entire ecosystem applies to various pipelines such as education pipeline, workforce pipeline, talent pipeline, manufacturing pipeline, school to job pipeline, etc. The pipeline is essentially the entire process from starting a desired goal to fulfilling that goal. For example, the stakeholders are helping the students in k through 12 systems with awareness on manufacturing, and awareness and career exploration in many other career pathways, not always just in manufacturing. Since Indiana is primarily a manufacturing state, they're trying to explain to people that there's a whole city in a manufacturer and so if they want to go into nursing, that's great. They can go into nursing and here's the pipeline. But there are opportunities in a local community in nursing, not only in a hospital, or the traditional places where one would see nursing, but also in manufacturing.

Answering research question 1

The first research question deals with finding which guidelines/visualization design recommendations are applicable and need to be followed while creating a data dashboard for visualizing education and workforce data. After several rounds of thematic coding (Braun & Clarke, 2006) of the interviews, two main themes emerged as shown in Figure 4.19.

Themes that address RQ 1
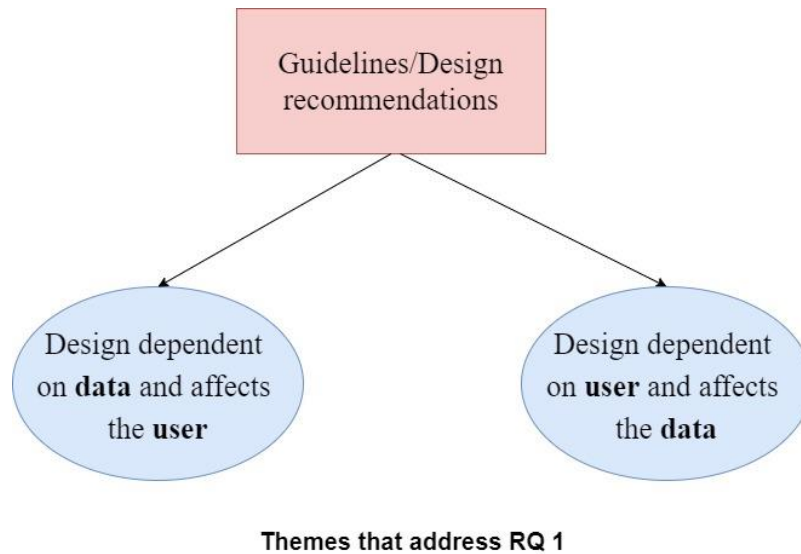
Figure 4.20. Themes that address research question 1

These themes are further comprised of the main categories that form the guidelines to design the data dashboard. These guidelines are given below:

1. **Dashboard design dependent on the characteristics of the user and affects the data:**

   a. Consider the end user's frame of reference while designing the visualizations and dashboard elements (try to find the frameworks, visualizations that the user refers to frequently by conducting a study on the user.)

   → The users of the data dashboard often refer to the regional map available on the Department of Workforce Development while addressing certain areas. They are familiar with the regional colors and the counties that lie in the given region. Thus, the dashboard uses a map that is designed similar to this regional map (Regional Maps, 2019) in order to allow users to access the data. When asked whether the regional map-based representation was helpful while accessing the data, one of the users mentioned: *"That's a given. For me, that's how I work. You know, when I first*

started in this job, I had that map in the back of one of my notebooks so that I can refer to it. Because again, I didn't grow up here. It's not second nature to me. But now I think that's good. It's easily recognizable. And the way it's structured, going from the region down to the county, everything I think, it makes it much easier to highlight what you want to find."

b. Maintain consistency of view in the dashboard pages and visualizations across all the datasets.

→ Since there are several datasets with their own characteristics, it can quickly become overwhelming to grasp all the information while the users move from one dataset to the other. Across all the datasets, the visualizations and their features such as color, size, position, etc. are kept consistent so that once users learn how to use one dataset, they can transfer the knowledge across all the other datasets. One user was able to correctly point out his preference for this consistency: *"I like that it's consistent across all its tabs."*

c. Provide users with the ability to generate reports from the data.

→ Users found the "Generate reports" feature to be extremely useful. One user mentions this: *"And then being able to pick a county and generate a report. I think that's really great too."* The other aspect was that they can quickly lose track of what they were looking at since there is a lot of data. The report generation feature was useful in that case as well. Lastly, they felt that other users are going to want to customize the reports that are generated and so one user pointed that out by saying: *"I think people are going to start asking you for more reports. I think they're going to want to say, 'how can I customize this report?' Yes, I think they're going to be like this is amazing. How do you customize these reports?"*

d. Design the data and visualization interfaces to support additional datasets that may be required by the user.

→ Some users wished to see information about wages, employment rates, graduation rates and so on. But currently, the dashboard only visualizes a few datasets. Thus, being able to support the visualization of other datasets would be necessary. There were a lot of datasets that the users were interested in: *"It would be interesting to be able to click on this (map) and be able to look at like, what are the main industries for like that either region or county? To me, that'd be an interesting data set to have."* Other datasets the users were interested in were: *"I'll pull the data to see, you know, what the unemployment rate is, how many jobs are open, what is the graduation rate for the school systems? The other data could be you know, free and reduced lunch data. So those are some of the data sets that I would look at."*

e. Use desktop web applications for detailed analysis of data and mobile for overview or quick snapshots

→ Since the users mostly used the dashboard at work in their offices, they would use a computer to access it. In that case, they would like to perform a thorough analysis using a larger screen. One user says: *"I feel like this is a lot of data and you're messing with it and going back and forth. I think it's hard to pull up and look on your phone. So definitely a website that I can use on my desktop."* They would also not use it outside of their work and would prefer to use it on their phone only for quick summaries or to view snapshots of the data as one user states: *"Mobile, I wouldn't use as much. I mean, it's nice to see quick impacts."*

f. Ensure stability, maintainability and ease of use.

→ The dashboard must be stable i.e. it must not fail too often due to server crashes. Since there is a lot of data, it must be well structured, and the dashboard must be easy to maintain and update. The dashboard must be easy to use, and it should not be confusing. One user summarizes such statements from several others into one sentence as: *"So those three things though, you know, something that's stable, aesthetically pleasing and easy to manage."* Another user addresses the ease of use by saying: *"Yeah, so I think it's really helpful because I like the color of it and everything that you've got, the different colors for the regions, I think that's super helpful. People are using more of that language. So, people are like what region? So, I think that's really helpful."* Another user found the dashboard to be so useful that they said that they would add it to their bookmark tab so that they can use it more frequently because of how easy it was to use it to find the relevant data: *"But the website is what I could add to my bookmark bar."*

g. Design for users with varying work roles (from management, planning, technical, etc.)

→ Since the users of the dashboard work in various roles, they have different requirements in terms of use, datasets of interest and what decisions they would make as was explained in the previous section on users' work processes. A prior knowledge of this can help to make better design decisions. One user found the statistical information and graphs to be very useful: *"Certainly, you know, we use a lot of bar graphs, process diagrams and those kinds of things as well for the technical work that we do"*, whereas another did not find that to be very useful: *"The bar graph and pie chart are interesting statistics but I don't know what I would use them for in my role."* It also helps to know the user's characteristics such as: *"Well, because I'm a visual person anyway."* Knowing about the user's work focus can further help to guide how certain aspects of the dashboard are designed: *"Half of what I do is address the workforce pipeline concerns for our community in Tippecanoe, but also in our region, which covers 10 counties. So, it's the footprint of the western side of region for the workforce area."*

2. **Dashboard design dependent on the characteristics of the data and affects the user:**

a. Provide enough description of the dataset, what the data contains and what it doesn't.

→ Since the datasets have names which can mean multiple things in different contexts, the users wished to find more information about what exactly the dataset contained. For example, one dataset visualizes the "Businesses and Industries" in Indiana. But they do not contain data about every single business such as a restaurant, shop or manufacturing industry. They contain data about the main groups/organizations that allow local businesses to connect, provide them with resources, etc. Such description of the datasets would be necessary as the users have pointed: *"But like, what is business and industries? What does that make up? What exactly is that? Because business and industry could be a lot of things to a lot of people.", "I want to know exactly what the definition is, that it is not for me to assume. I know what, for instance, on this particular one, we're in the 'Post-secondary education'. So exactly what does that, what does that mean when you're looking at this particular county."*

b. Ensure clarity is conveyed:

   i. Through the visualizations

   ii. About the datasets

   → The visualizations should not be confusing and must clearly show what the data entails. These visualizations can quickly help the user to gain a understanding of the data as one user states: *"I think when I'm looking at the institutions, as I've said previously, you know, I as I look at the 'post-secondary institutions for all 92 counties (refers to 92 county bar graph), my first response is, who knew there were that many?"* There should be enough clarity about the data presented through the graphs: *"Yeah, you know, but when you're getting into some details, there's a lot more nuance. And that's what's great about a visualization such as this, is that, you know, a lot of nuance can be communicated in a very concise image."*

c. Allow the user ability to control the granularity of data that he is interested in looking at.

   → The dashboard should give the users an ability to look at the data in detail such as the name and contact information of a school principal, but it should also give them an overview such as by showing how many schools are present in a specific county or region. The dashboard supports this ability. The users support this by saying: *"And then not only did you give information about what that is, then it can also link to the website to let me know what that does cover and all those different things. So, there's a layer of information that I can go and dig with, without having to do that own work on my own. I think it's nice that that snapshot is there together"* and *"Just a breakdown by the county and then being able to access all of that information in one place, that's really helpful because then you've got to go through the resources, you've got to drill down within there and see how you can get connected."*

d. Ensure data is valid (avoid old data), reliable (avoid missing data and broken fields) and updated frequently (quarterly to yearly)

→ The users showed concerns with the reliability and validity of the data. Currently, some data fields were missing, some data was old, and some links were broken. The datasets are available publicly and the data is collected by the government. Thus, those need to be fixed otherwise users will be skeptical of the reliability of the dashboard to inform decisions. One stakeholder expressed their concerns as: *"And so, you know, for example, I pulled my county, I work remotely, so in my county which is Decatur, I pull up the Greensburg/Decatur County Economic Development Corporation, and Marc Coplon, he hasn't been the EDC director for almost three years. So my concern is when we're pulling this data, is either the DWD isn't updating their data on a timely basis, and is that something that we have to go in and refresh or do ourselves overtime?"*

e. Provide data source and last update information.

→ Along the same lines of concerns of validity and reliability of the data, it would be useful for the users to know when the data and dashboard was updated. The users want to know where the data was pulled from so that they can trust the contents while making decisions. A user states: *"Because, you know, if it's 2018, then, you know, we need to somewhere when you click on that data set, there needs to be an inset that says this data is generated from DWD, last updated 2016, because then then they'll know okay, that's where they're pulling that data."*

f. Use area breakdown (by county/region) as a starting point to access the data.

→ The map is interactive and clicking on a county updates the visualizations and other text boxes, drop down with information about that county. Users find this navigation mechanism very useful while accessing the data and comparing different counties: *"I think it's going to be fantastic if I can come to this and click a specific county, go straight to the person who I need to contact or what I need to retrieve and do it quickly."* Also, they mention: *"And especially in this area, where Floyd and Harrison are really close to Louisville, what are they doing in that particular space? And just to benchmark against things. So I would look at Allen county and say, 'What's going on in Fort Wayne that they're doing that, you know, is possibly leading to that, and comparing those particular programs that may be in Allen county that we should look at here.' That would be kind of a contrast and compare between, like communities (similar communities) across Indiana."*

Answering research question 2

The second research question deals with finding how an information visualization dashboard affects the work processes and decision making of the users involved in the education and workforce development for the state of Indiana. Again, after several rounds of thematic

coding (Braun & Clarke, 2006) of the interviews, three main themes emerged as shown in Figure 4.20.
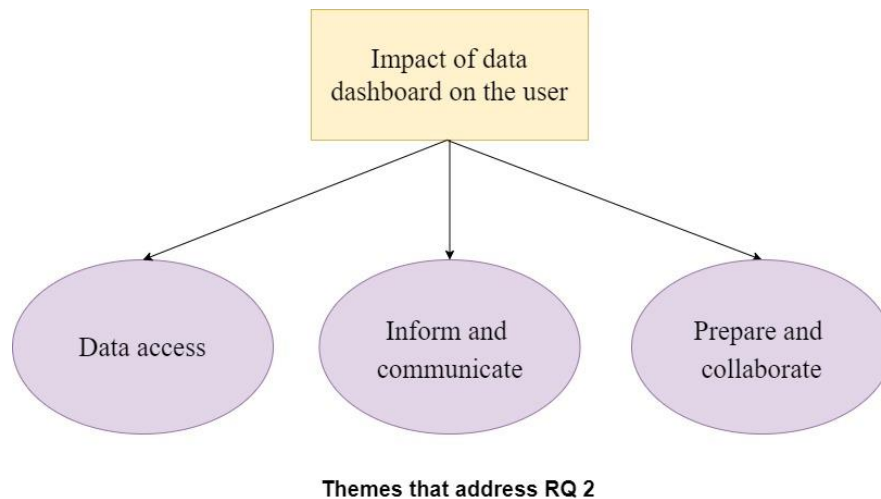

Figure 4.21. Themes that address research question 2

These themes describe how the data dashboard affects the user in accessing data, informs and allows them to communicate and enables them to prepare and collaborate with others. These themes consist of categories as given below:

1. **Data access:**

   a. Dashboard provides centralized access to data.

   → All the users had one thing in common to say about the dashboard: *"It is a one stop shop".* For most of the users, they had to look for several resources and websites online to find relevant information. They had to sometimes do a Google search to find the information they were looking for. But the dashboard was very useful to them in that it was the only place they had to go to in order to find most of the information they were looking for. Other users said: *"And so finding this, having this information with this website really helps because it is consolidated into one place.", "Now for me, the value is, is having everything in one site, especially school attendance, the local and regional plans and the pre k and all of those things together, so that I can just move within this one site, instead of having to move between, for instance, 10 different counties, or 10 different economic development websites or 10 Chamber of Commerce websites.", " it's kind of hard to find that one stop shop with that information."* Even while considering the fact that the data needs to be updated, the user said: *"I would have just one location that I have to worry about."* Thus, the

80

dashboard proved to be a very useful resource to be able to find most of the information in one place.

b. Dashboard provides easy access to various datasets through common interface.

→ Since all the datasets have a similar page layout, it is very easy to access the data and navigate across the dashboard as a user mentions in this context: *"I think the ease of use, you got the ease of use, you're already far down that pathway and probably ahead of most of what we have now".*

c. Dashboard is frequently accessed to obtain relevant information (daily to a minimum of weekly access).

→Most of the users said that they would use this dashboard minimum on a weekly basis if not daily. Even though the users' day-to-day is not fixed and they perform a variety of tasks, they would still use this dashboard for some purpose or the other. One user says: *"My role is pretty ill defined, like what I actually do day to day is pretty ill defined. So, I would say, probably weekly."* When asked about how frequently they expected the data to be updated, the users gave a variety of answers ranging from real time update, monthly, quarterly to yearly. But, a few of them felt that quarterly or 4 times a year is a practical number: *"I would say, I mean quarterly. I think that's fair or at least every six months. It's a lot of information."*

d. Dashboard helps avoid dealing with raw (dry) data directly.

→ The users preferred to use the dashboard instead of looking at raw data in Excel sheets. It was hard for them effectively draw conclusions from the data and also know where to look for a specific field in the raw data. They had trouble looking at data from other websites as well since the data organization was different from what the users expected. In these contexts, the users say: *"But I think the fact that I would much rather look at this than just row upon row of data which is dry.", "I'm not going to DOE (Department of Education), which is a terrible website to try to find any data from.", "Sometimes on 'Hoosiers by the numbers' it's hard to drill down to what I'm really looking for."*

2. Inform and communicate:

a. Dashboard helps to find information to connect, communicate and collaborate with user's clients and partners.

→ The dashboard was a great point for finding contact information to the users: *"So for me, I may go in here to find a contact, you know, within my county service area, which is great because that's not always easy to find.", "I'm working in the post secondary education. I think that's probably good for contacts, just seeing what's available in those areas."* The users used the dashboard to collaborate and communicate with their partners and clients: *"So we can, more or less be*

communicating with this image. 'Look what we've done elsewhere across the state. We've not been participating in this area yet. We'd love to.'", "It's to develop collaborations, mainly. To communicate upward and develop opportunities."

b. Dashboard helps to find gaps, plan work and create more opportunities.

→ The dashboard helps the users to find gaps in education and employment. They can then direct their funds, plan their activities to reach out to others and create more opportunities for the students' growth. A user states his use as: *"Just to see what's out there, what are we missing? Is there a place where we need to focus for additional camps or where could we hit a pocket of students that may not be able to be reached through traditional education?"*. New business can also be attracted with the help of the data on the dashboard: *"So for us, especially when we're looking at an attraction, they're going to determine whether or not we can show that we have the workforce here, to be able to come here, which is our number one concern with site selectors and new businesses, being able to show that we have the workforce or being able to show that we can train the workforce. And so bring in that data of qualified personnel or that we have students that are training here are ready, like at Purdue or Ivy Tech, that they're going to go into those".*

c. Dashboard helps to generate awareness.

→ A lot of users found that they were not even aware of the number of schools or organizations in their areas of interest and counties that they work in. Through this dashboard, they could search for it. Navigating across the various datasets, they were able to see which organizations were involved in which activities, what others were doing in specific domains and so on. Also, the various visualizations helped to visually see what the ecosystem looked like. In this context, the users say: *"What I do need to know is which schools are doing something and what are they doing and who's the contact. The contact information is important and also what kind of activity they have been doing.", "So like, I'm looking at the 'IN-MaC micro-grants', I want to see what did this school do with that $2,000.", "As I look at the 'post-secondary institutions for all 92 counties (refers to 92 county bar graph), my first response is, who knew there were that many?"*

3. **Prepare and collaborate:**

a. Dashboard is used to support or back conversations with facts.

→ The data from the dashboard and the various visualizations help the users to back their conversations and statements with facts. One user points this out by saying: *"You can talk about, well, in our region, we have blah, blah, blah, well, you don't have to come without any specifics or data or any visualization. You can bring that with you and a one-piece snapshot and then people can go back and look at the more in-depth information. So, I think that would be one of the main things for me is to be able to, you know, create fact around what I'm saying, I'm not just speaking out of 'this is what I think'."* It can also help them for marketing purposes: *"You know, this*

helps get us in the door, we can fit a nice visualization like this on a on a trifold flyer or in a presentation, you know, as a lead developer, as a business development in."

b. Dashboard is used to prepare for meetings and presenting data.

→ The dashboard allows the user to look at data that is currently of interest to them. They can take notes from the dashboard and also generate small reports of information that they want to present. They can prepare well in advance for meetings which helps them build trust with other partners: *"Part of what I want to show is, you know, 'hey, we've done our diligence, we've done our homework, and we've already been participating in this area.' You know, instead of just, 'Hi, I'm from Purdue, I'm here to help. What can I do?' You know, it helps build that trust, it reinforces that trust.".* While addressing how they would use it to present data, one user mentions: *"And I would use this for I mean, presentations too, to tell other people, you know, here's what we're currently impacting."*

c. Dashboard allows to maintain a consistent view about the data between all users to avoid confusion and inconsistencies.

→ The way raw data is present in the current form, various users can be looking at different versions of the data. This can lead to inconsistencies in their perception. Since the dashboard is deployed on the web every user that pulls it up, will be looking at the same information irrespective of their location. Thus, it helps to maintain a consistent perspective across all users as one user comments: *"So we're all seeing the same thing, with the same information. And we don't have different versions of things."* It also helps them to make decisions quickly: *"Well, on a practical level, if you're all looking at the same screen with the same information, it's easy to come to an agreement I think on different things."*

4.3. Summary

The thematic coding on semi-structured interviews helped to address the two main research questions of this study. The design guidelines and dashboard influence on work processes was further strengthened by the results of the usability tests conducted as part of the study. These guidelines and effects can further be analyzed, improved and added upon by conducting further studies in this domain.

CHAPTER 5.   DISCUSSION

Here we discuss how this study makes new contributions to the field and its contrast with existing work. We also discuss the applicability of the research findings in the design of data dashboards for other states and domains as well.

5.1. Dashboard design: comparison with existing work

In this research, we developed a data dashboard that uses a regional map-based visualization as a mechanism to view and access data in the education and workforce domain for the state of Indiana. Through the user studies we performed, we found which design principles (from the many previously known) are applicable for developing dashboards in this domain. However, not all guidelines from the literature were applicable *as is* and some guidelines had to be overlooked to address the needs for this domain. Such comparison with existing design principles and dashboards is discussed here.

As (Few, 2013) mentions, the dashboard must not exceed a single screen. The rationale is that the user cannot hold large chunks of information in memory at a time. In our dashboard design, after interacting with the map, most of the information and visualizations (region bar graph, pie chart, information text box) that update fit in a single screen. There are other visuals in the bottom region of the screen, but them can be looked at from a different context and does not require information present in the upper part of the screen. Thus, the dashboard applies the essence of this principle in the required context. Also, (Sarikaya, 2018) mentions that the context and dashboard use are expanding beyond simple monitoring and single screen reports. Since all

our dashboard pages have a common layout, choice of visualization and color coding, we avoid introducing meaningless variety (Few, 2013) that confuses the user. Also, the map-based interactivity and animation during visualization update that follows, is simple and does not distract the user. Other useless embellishments are also not present. Our regional map-based design and visual encoding ensures contextually aware representations (Sarikaya, 2018) and uses the visual language of 12 EGRs as used by the end users.

However, our dashboard is not real-time. Such real-time behavior is required for financial dashboards (Stríček & Andrisková, 2015), healthcare dashboards (Faiola, 2015) and the more recent dashboard for monitory covid-19 cases (Dong, 2020). Since the data for our dashboards is pulled from public datasets, which are themselves updated once a year, the data for our dashboard is not real-time. Further, our datasets consist of structured textual fields and does not consist of video, audio, images, click streams, log files, etc. as is present in organizational data (Lempinen, 2012) and (Laney, 2001).

Guidelines for designing performance dashboards, dashboards that measure performance in organizations, have also been developed in (Lempinen, 2012). As stated here, the decision makers are employees at all levels in the organization and not just the executives, and their decision-making process is semi-structured to unstructured. Thus, it is difficult to predefine what the users may need (Lempinen, 2012). Similarly, the users of our data dashboard are stakeholders in various positions and their decision making is based on viewing data from various datasets, some being unique to the user and not being useful to others. In order to measure the KPIs, two design principles were established: "to define a general performance measurement framework" and "to follow a structured process for selecting key measures" (Lempinen, 2012). Though these have been generalized by them, they are not applicable to our

dashboard as we are not trying to measure the performance. We are trying to view the current

state of development in the education and workforce domains and simple statistical

visualizations satisfy our needs.

As we have seen, there is a huge variation in dashboard design based on specific use

cases. Examples of several strategic, tactical, operational and social dashboards are given

(Sarikaya, 2018) that do not provide contextual information, are extremely cluttered, do not

consider design from the end user perspective and so on, thus overlooking dashboard design

principles. Such principles are followed in the regional map-based dashboard presented here. A

dashboard that shows the crime rates across the US over a period of 50 years (Pulikathara, 2015)

uses a heatmap calendar to represent the states and access the data. This is difficult to navigate

since we now need to hover over every single entry to find the state we are looking for. A

regional map-based interface like ours would have made this navigation extremely simple and

easy. Our dashboard does not allow the user to modify or change the composition of

views/layout as can be done for some other dashboards (Roberts, 2017). One of the goals for

future work is to overcome this limitation. Users also cannot write back or modify the database

through the dashboard since such requirement is not necessary and may lead to violations in the

validity and reliability of the data. Thus, such design principles are not applicable to the

dashboard presented here.

Our dashboard shares common features with other strategic and operational dashboards

(Albuquerque, 2015) (Company, 2017) that support decision-making such as multiple consistent

views, contextual information, filtering mechanisms, simple graphs, etc. However, they still have

other drawbacks such as inconsistent and repetitive color scales. These dashboard types allow the

users to interact with the visualizations to focus on data relevant to them. These dashboards are generally used for business purposes or for monitoring school data (Education, 2017).

An educational dashboard exists for the schools in California (Public Schools and Districts Data Files, 2019). We see several differences between our dashboard and theirs. First of all, it 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 (Munzner, 2014) (Few, 2013). Further, selecting one of these overviews takes us to the bottom of the screen which just explains what the color scale means and provides a general overview in percentages. It does not provide us with any other context. Also, we need to scroll all the way back to the top to view a different field. Also, there is only one gauge visualization used for all the data types which seems highly unsuitable (Few, 2013). In contrast, our dashboard provides an interactive map-based selection mechanism in which we can go to a specific county and then pull up information about schools there. The drop down allows us to select a specific school of interest and then we can see the contact information and other details about that school. All the linked information is within the same view and avoids the user from going back and forth. Also, after scrolling down, we can move back up with the click of a single button that appears in the lower right-hand corner of the webpage. Further, in the California school dashboard, to view more details, we need to click on the information card that takes us to another webpage with lots of information. This is definitely an information overload (Few, 2013) and difficult to navigate. One common feature is the ability to generate reports from the data. However, the report generation takes several seconds and does not consist of the date and timestamp as is present in

the report generation feature for our dashboard. However, the report is clean and well formatted and structured similar to ours. The dashboard for the schools in California can make use of the regional map (Services, 2020), (also shown in Figure 5.3) and adapt the strategies for design similar to our dashboard to improve upon the drawbacks mentioned above.

The data dashboard presented in this research serves as a "one stop shop" for the users. This is not the case with other dashboards where only one or two datasets are visualized (Pulikathara, 2015). Further, the dashboard also provides a video tutorial that explains how to use the dashboard and interpret all the visualizations. This may not be needed for simple dashboards with a few datasets or general knowledge domain. However, even dashboards with domain specific language and data such as fire incidents and cancer reports (Fire Incident Dashboard, 2019) (Bullock, 2016), did not use such help tutorials that could help new users quickly get onboard with the information and visuals. Users can be taught how to use the dashboard through explicit training such as one-on-one training, classroom training, online learning software, virtual classrooms, help desk, etc. (Eckerson, 2010) or guidance built into the dashboard itself (Sarikaya, 2018) though mouse-over prompts and online help documentation (Malik, 2005). For a large user base, a webinar or interactive tutorial (such as ours) is an effective choice (Malik, 2005).

Similar to the dashboard presented here, the public safety dashboard (O'Connell, 2016) uses publicly available datasets that are combined to find relationships in the data and is designed to serve as a one stop shop for the desired information.

5.2. Uniqueness of the datasets

The education and workforce datasets visualized on the dashboards are primarily structured text. They do not contain images, audio, video, log information, user clicks, etc. Thus,

there is a representation challenge to this abstract information. Also, the datasets consist of the region number and county information to which the organization/institute belongs. Our dashboard makes use of this inherent quality of the datasets while creating the visualizations and multi-linked views.

Further, since all the datasets have region number and county name, they can be used as secondary keys along with other primary keys (such as school/corporation ID) to link various datasets together. This can significantly help to go from one dataset to the other. For example, using the school ID, the users can move from the principal contact information dataset to the attendance and graduation rate data for the same school without having to go to those datasets individually and use all the filtering mechanisms to retrieve that data. This can also be used to pull such information from all the other datasets at the same time and generate a full-fledged report. Such development falls under the future work of this research project. Additional data processing will be needed to incorporate such features with other financial or weather forecast data, for example.

5.3. Contributions

In this research, we developed a data dashboard that uses a regional map-based visualization as a mechanism to access and navigate various datasets. Currently, we focused on the education and workforce domain and stakeholders in the state of Indiana who will access this dashboard for various activities as was mentioned in the Results section. By considering the prior user perspective of referring to the regional map (Regional Maps, 2019), we were able find that the usability of the dashboard based on such a design was high. Also, the regional breakdown as a way to access data was intuitive to the users and made it extremely easy for them to find relevant information. Other features of the dashboard such as report generation, use of simple

graphs, layout, video tutorial, etc. was also found to be useful. Several principles can be used while designing the dashboard as is mentioned in the literature review part of this document. However, not all of them were relevant for the desired domain and demographic. Through one of the outcomes of this research, we were able to find the relevant principles that are applicable while visualizing education and workforce data for the stakeholders in the state of Indiana.

This study also provides information on the work processes of the users and how the dashboard affects their work and decision making. These results can help us to further improve the dashboard design and opens up possibilities for further investigation. Further, since we visualized multiple large datasets on a single dashboard, we found that users preferred to use this "one stop shop" instead of looking at various data sources to find information. Not all the data dashboards use multiple datasets in their visualizations (Sarikaya, 2018). This we found was a useful practice and should be explored further.

Our research focuses on the requirements of the stakeholders in education and workforce domain that are specific to the state of Indiana. Thus, through this regional map-based dashboard that utilizes the visual encoding of the 12 EGRs in all its visualizations and interactivity mechanisms, we recommend that further dashboard designs can benefit from such a design consideration.

5.4. Applicability to other states and domains

A regional map-based dashboard is developed in this research that can be used to access and view the education and workforce data. Such a map view is used by the stakeholders in education and workforce domain for the state of Indiana. We found that other states and domains also use a regional map to breakdown the counties in order to make the decision-making process simple.
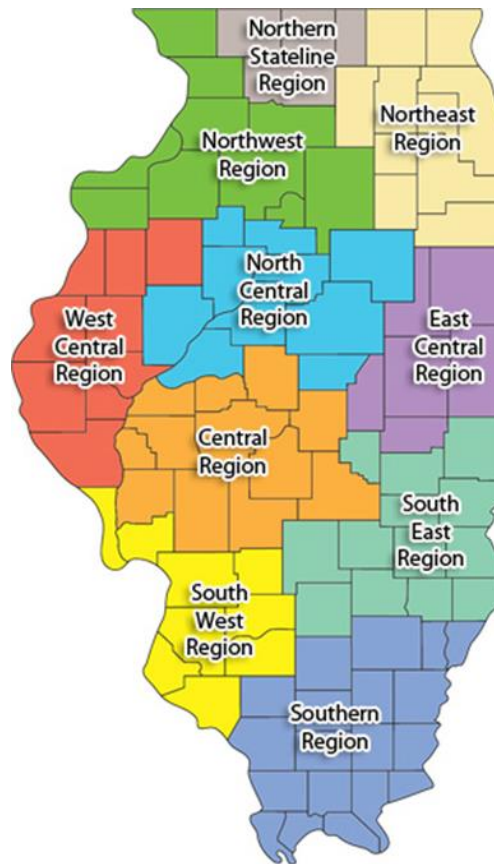
Figure 5.22. Illinois Department of Commerce Economic Development Regions. From *Illinois Department of Commerce and Economic Opportunity,* by DCEO, 2020, (DCEO, 2020), https://www2.illinois.gov/dceo/SmallBizAssistance/RegionSpecificAssistance/Pages/default.aspx.

As shown in Figure 5.1., 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 5.2., 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 5.2.
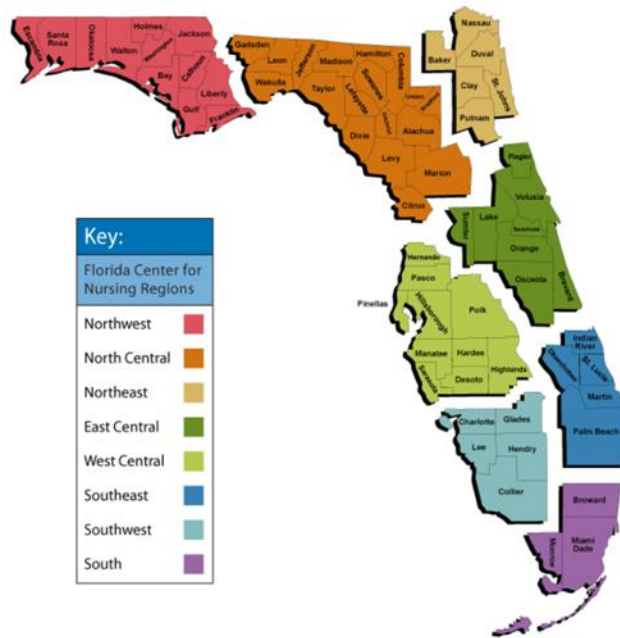
Figure 5.23. Florida Center for Nursing Regions. From *Florida Center for Nursing,* by University of Central Florida, 2020, (Florida, 2020), https://www.flcenterfornursing.org/RegionalData/FCNRegionalWorkforceReports.aspx.

Also, as mentioned earlier, the state of California is also divided into 11 regions based on the LEC (local Education Consortia). This can be seen from the map given in Figure 5.3. The education dashboard developed for California has several drawbacks and they can be overcome by using a regional map-based dashboard based on the guidelines and outcomes presented in this research study. Thus, the applicability of these design principle are widespread and can be used to design new map-based data access interfaces and dashboards.
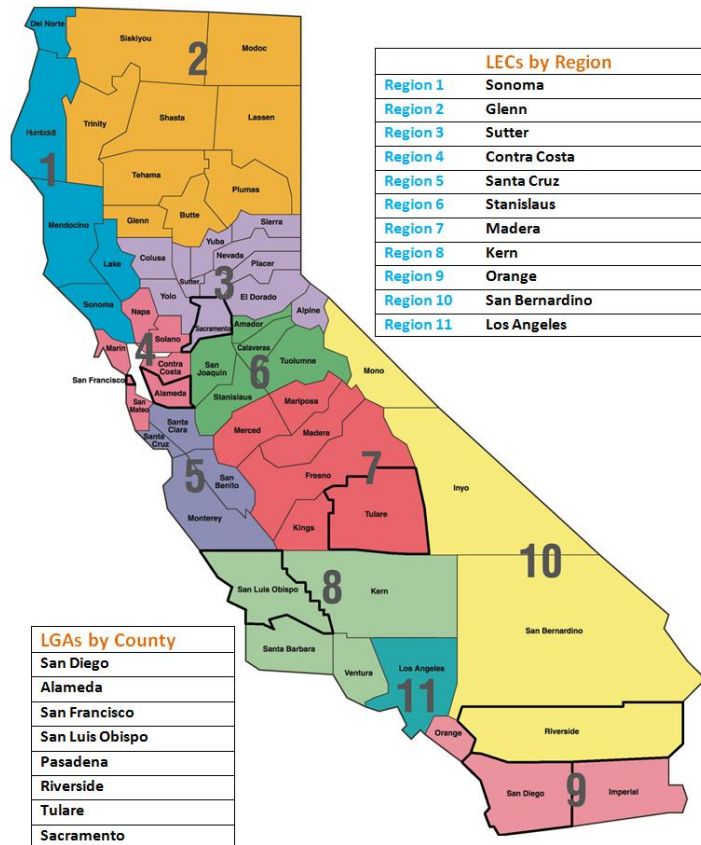
Figure 5.24. California Local Educational Consortia (LEC) and Local Governmental Agencies (LGA). From *Service Regions Map and Contact Information,* by Department of Healthcare Services, 2020, (Services, 2020), https://www.dhcs.ca.gov/provgovpart/Pages/MapLECsLGAs.aspx.

CHAPTER 6.   CONCLUSION

In this chapter, this study is summarized, and a discussion is provided on the results that address the two main research questions. The chapter concludes with a discussion of the scope for future work.

6.1. Summary of this research

In this research, a data visualization dashboard was developed for education and workforce development for the state of Indiana. The usability of this dashboard for performing various tasks was tested by conducting a usability test on 20 participants. Most of the participants performed the tasks accurately, with just two tasks getting a low accuracy rate as was mentioned in the results section of this study. This dashboard template (Miller, 2019) was created by following good UX design principles (Grant, 2018) and was further modified for this specific use. The link to the source code repository for this dashboard is given in Appendix I.

Further, the dashboard was used by stakeholders in education and workforce for a period of 1 week before interviews were conducted with them. These interviews covered various aspects such as tasks performed by the users, tools currently used, dashboard use, challenges faced, likes and dislikes, etc. Thematic analysis on the interview data further led to the generation of design principles/guidelines that should be used to design such data dashboards. This analysis also helped to understand the impact of the dashboard in the various work processes of the users.

Thus, the insights gain from this study will allow other developers and researchers to develop dashboards for visualizing education and workforce data for other states as well.

6.2. Discussion of research question 1

This research question is focused on finding guidelines/design principles to develop an effective data dashboard for visualizing massive sets of education and workforce data. The guidelines presented in this research are in no way complete. There are several other guidelines that can emerge from further studies. The primary goal of these guidelines is to help to developers to design better data dashboards, faster. Some of these guidelines may be not be effective in certain designs based on the goal of the data dashboard. Such cases need further studies and evaluation.

6.3. Discussion of research question 2

This research question is focused on finding how a data dashboard that visualizes massive datasets related to education and workforce affects the work processes of the stakeholders. Since the users of such a dashboard are from various educational backgrounds and perform a variety of roles, the dashboard is used in a variety of ways. Several datasets currently not visualized in this dashboard are relevant to different users. Accommodating the needs of all such individual users is not possible. These users are interested in various aspects of the data. They perform different levels of analysis on the data. Thus, the dashboard can never be developed as a 'one design fits all' and its requirement may change depending on the user.

The main purpose here was to enable the user to find ecosystems in the datasets through a regional map-based data access dashboard interface. Thus, the design of the dashboard and user studies were conducted to test the effectiveness of the dashboard to fulfil that requirement. Dashboard designs can vary in terms of content, platform, data and navigation and this will thus affect how it impacts the users' work processes.

6.4. Future work

The dashboard design can be further improved and tested as follows:

1. Add functionality to allow the user to dynamically update and load the data, modify visualizations and customize the dashboard elements.

2. Visualize several different datasets related to education and workforce.

3. Test the applicability of the guidelines for developing data dashboards for other states and countries.

4. Generate several versions of the designs and perform comparisons between them through controlled quantitative studies.

Through the user studies, several features were developed that can help to improve the usability of the dashboard. Implementing these features and testing them would be the next step.

Further, performing long term user studies that last for several months and consist of various types (such as interviews, diary studies, observation, etc.) (Carpendale, 2008), and are conducted with large number of end users, will further help to develop and validate the dashboard design principles.

6.5. Summary

In this chapter, the research questions were summarized and scope of future work that can add on this research was mentioned.

REFERENCES

(2010). *A skilled workforce for strong, sustainable and balanced growth: A G20 training strategy.* Geneva, Switzerland: International Labour Office.

Albuquerque, C. o. (2015, August 24). *Economic Dashboard.* Retrieved from Tableau Public: https://public.tableau.com/profile/olivia.padilla.jackson#!/vizhome/EconomicDashboard_2/CABQEconomicDashboard

Bank, C. C. (2014). *The Guide to Usability Testing* . UXPin.

Bazeley, P., & Jackson, K. (2013). *Qualitative data analysis with NVivo.* SAGE publications limited.

Bootstrap. (2019). Retrieved from Bootstrap: https://getbootstrap.com/

Börner, K. M. (2016). Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, 15(3), 198-213.

Bostock, M. (2019). *D3 API Reference.* Retrieved from https://github.com/d3/d3/blob/master/API.md

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D³ Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics, 17*(12), 2301-2309. Retrieved 10 11, 2019, from http://vis.stanford.edu/files/2011-d3-infovis.pdf

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.

Bullock, C. (2016, September 19). *Cancer Moonshot.* Retrieved from Tableau Public: https://public.tableau.com/profile/chris.bullock5961#!/vizhome/CancerMoonshot_0/CancerMoonshot2016

Card, M. (1999). *Readings in information visualization: using vision to think.* Morgan Kaufmann.

Carpendale, S. (2008). Evaluating information visualizations. In S. Carpendale, *Information Visualization* (pp. 19-45). Berlin, Heidelberg: Springer.

Carswell, C. M. (1992). Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human factors*, 34(5), 535-554.

Ceniceros, R. (2009). Dashboard technology helps select measure, respond to risks. *Business Insurance*, 21-21.

Company, T. B. (2017, June 23). *Boeing Current Market Outlook*. Retrieved from Tableau Public: https://public.tableau.com/en-us/gallery/boeing-current-market-outlook?gallery=featured

Crooks, R. (2017). Representationalism at work: dashboards and data analytics in urban education. . *Educational Media International*, 289-303.

DCEO. (2020). *Illinois Department of Commerce Economic Development Regions*. Retrieved from Illinois Department of Commerce and Economic Opportunity: https://www2.illinois.gov/dceo/SmallBizAssistance/RegionSpecificAssistance/Pages/default.aspx

division, D. o. (n.d.). *INfographics*. Retrieved from Infographics - Hoosiers by the Numbers: http://www.hoosierdata.in.gov/infographics/index.asp

Dong, E. D. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases*.

Dowding, D. R. (2015). Dashboards for improving patient care: review of the literature. *International journal of medical informatics*, 87-100.

Eckerson, W. W. (2010). *Performance dashboards: measuring, monitoring, and managing your business.* . John Wiley & Sons.

Education, C. D. (2017). *Los Angeles High School of the Arts*. Retrieved from California School Dashboard: https://www.caschooldashboard.org/reports/19647330117747/2019#graduation-rate-card

Faiola, A. S. (2015). Supporting clinical cognition: a human-centered approach to a novel ICU information visualization dashboard. *AMIA Annual Symposium Proceedings* (p. 560). American Medical Informatics Association.

Few, S. (2013). *Information Dashboard Design: Displaying Data for At-a-Glance Monitoring.* Analytics Press.

Filonik, D. M. (2013). A customisable dashboard display for environmental performance visualisations. . *International Conference on Persuasive Technology* (pp. 51-62). Berlin, Heidelberg: Springer.

Find School and Corporation Data Reports. (2019, September 24). Retrieved from Indiana
Department of Education: https://www.doe.in.gov/accountability/find-school-and-
corporation-data-reports

Fire Incident Dashboard. (2019, June). Retrieved from ArcGIS:
https://www.arcgis.com/apps/opsdashboard/index.html#/d97231c35d9b424d85e1bb110c
45991c

Florida, U. o. (2020). *REGIONAL DATA - FCN REGIONAL WORKFORCE REPORTS*.
Retrieved from Forida Center for Nursing:
https://www.flcenterfornursing.org/RegionalData/FCNRegionalWorkforceReports.aspx

Grant, W. (2018). *101 UX Principles: A definitive design guide.* Packt Publishing.

IDOE Data. (2019, August 9). Retrieved from Indiana Department of Education:
https://www.doe.in.gov/idoe/idoe-data

IN-MaC. (2019). Retrieved from Indiana Next Generation Manufacturing Competitiveness
Center: https://www.purdue.edu/in-mac/

Isenberg, P. Z. (2008, April). Grounded evaluation of information visualizations. *Proceedings of
the 2008 Workshop on BEyond time and errors: novel evaLuation methods for
Information Visualization*, (p. 6). ACM.

Key Impacts. (2019). Retrieved from IN-Mac: https://www.purdue.edu/in-mac/

Kirk, A. (2019). *Data Visualisation: A Handbook for Data Driven Design (Second Edition).*
SAGE Publications Ltd.

Laney, D. (2001, February 6). *3D data management: Controlling data volume, velocity and
variety*. Retrieved from https://blogs.gartner.com/: https://blogs.gartner.com/doug-
laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-
and-Variety.pdf

Lee, S., Kim, S.-H., & Kwon, B. C. (2017). VLAT : Development of a Visualization Literacy
Assessment Test. *IEEE Transactions on Visualization and Computer Graphics, 23*(1),
551-560. Retrieved 10 11, 2019, from https://ncbi.nlm.nih.gov/pubmed/27875171

Lempinen, H. (2012). Constructing a design framework for performance dashboards.
Scandinavian Conference on Information Systems (pp. 109-130). Berlin, Heidelberg:
Springer.

Liu, J. T. (2018). A survey of scholarly data visualization. *IEEE Access, 6*, 19205-19221.

Malik, S. (2005). *Enterprise dashboards: design and best practices for IT.* . John Wiley & Sons.

Marx, F. M. (2012). Six principles for redesigning executive information systems—findings of a survey and evaluation of a prototype. *ACM Transactions on Management Information Systems (TMIS)*, 1-19.

McCormick, J. (n.d.). *IDOE - Home*. Retrieved from IDOE Website: https://compass.doe.in.gov/dashboard/overview.aspx

Miller, D. (2019, September 16). *SB Admin 2*. Retrieved from Startbootstrap: https://startbootstrap.com/themes/sb-admin-2/

Morgan, M. B. (2008). The radiology digital dashboard: effects on report turnaround time. *Journal of digital imaging*, 50-58.

Munzner, T. (2014). *Visualization Analysis and Design.* A K Peters/CRC Press.

Murray, S. (2017). *Interactive Data Visualization for the Web, Second Edition.* O'Reilly Media, Inc.

O'Connell, K. L. (2016). Making public safety data accessible in the westside Atlanta data dashboard. *arXiv preprint arXiv:1609.09756*.

Otter Voice Meeting Notes. (2020, April 28). Retrieved from Otter.ai: https://otter.ai/

Parsaye, K., & Chignell, M. (1993). *Intelligent Database Tools and Applications: Hyperinformation access, data quality, visualization, automatic discovery.* John Wiley & Sons, Inc.

Public Schools and Districts Data Files. (2019, May 28). Retrieved from California Department of Education: https://www.cde.ca.gov/ds/si/ds/pubschls.asp

Pulikathara, S. (2015, April 27). *50 Years of Crime in the US*. Retrieved from Tableau Public: https://public.tableau.com/en-us/gallery/50-years-crime-us

Qu, Z., & Hullman, J. (2017). Keeping multiple views consistent: Constraints, validations, and exceptions in visualization authoring. *IEEE transactions on visualization and computer graphics*, 468-477.

Regional Maps. (2019). Retrieved from Indiana Department of Workforce Development: https://www.in.gov/dwd/2653.htm

Roberts, L. D. (2017). Give me a customizable dashboard: Personalized learning analytics dashboards in higher education. *Technology, Knowledge and Learning*, 317-333.

Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing: how to plan, design and conduct effective tests.* John Wiley & Sons.

Saldaña, J. (2015). *The coding manual for qualitative researchers.* Sage Publications.

Sarikaya, A. C. (2018). What do we talk about when we talk about dashboards? *IEEE transactions on visualization and computer graphics*, 25(1), 682-692.

Services, D. o. (2020, April 9). *Service Regions Map and Contact Information*. Retrieved from Department of Health Care Services: https://www.dhcs.ca.gov/provgovpart/Pages/MapLECsLGAs.aspx

Shannon, P. M. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11), 2498-2504.

Shneiderman, B., & Plaisant, C. (2006). Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, 1-7. ACM.

Starmer, J., & Giuse, D. (2008). A real-time ventilator management dashboard: toward hardwiring compliance with evidence-based guidelines. *American Medical Informatics Association*, 702.

Stinson, L. J. (2012). E-health blood pressure control program. *The Journal of medical practice management: MPM*, 91.

Stríček, I., & Andrisková, I. (2015). Dashboard Usability In Financial Modeling. *CBU International Conference Proceedings*, 020-027.

Wainer, H. (1992). Understanding graphs and tables. *Educational researcher*, 21(1), 14-23.

Wexler, S. S. (2017). *The big book of dashboards: visualizing your data using real-world business scenarios.* John Wiley & Sons.

Wilbanks, B. A. (2014). A review of dashboards for data analytics in nursing. *CIN: Computers, Informatics, Nursing*, 545-549.

Workforce and Economic Data. (2019). Retrieved from Indiana Department of Workforce Development: https://www.in.gov/dwd/2329.htm

Yin, R. K. (2015). *Qualitative research from start to finish.* Guilford publications.

APPENDIX A

Indiana Education and Workforce Data Visualization Dashboard - Usability Testing Script
Test conducted by Yash Gugale

Date

(Welcome the participant and have them seated. Don't show the website to be used yet)

Hello {participant name}, thank you for coming. My name is Yash Gugale and I'm a graduate researcher in the department of Computer Graphics Technology here at Purdue University. I am working in the Byrd Data Visualization Lab for visualizing education and workforce data for the state of Indiana – I will walk you through the session today.

Before we start, I have some information for you.

I am currently testing the Indiana Education and Workforce Data Dashboard, that I created to learn as much as I can about the way people use it. Every product is intended to work in a certain way by its creators, but as you know – the reality might be quite different. The goal of this research is to get as close to reality as possible.

The session will take about 45 minutes to an hour.

Please remember, during the next hour, that I am testing the data dashboard, not you. Don't worry at all about mistakes. If they happen, it's the fault of the product. Finding about it is essential and gets me closer to creating a great product.

During the whole test, **please try to think out loud**. Share anything that's in your head. Tell me what you are looking at on the screen, what your thoughts are, what you like and what don't you like, etc. I want to learn about your honest reactions while using the dashboard.

Let me inform you that you are free to withdraw from the session any time you want without giving us an explanation.

Do you have any questions so far?

(Explain that the 'thinking out loud' protocol is very important for this research)

If you have any doubts during the test, feel free to ask me any questions. However, please remember that I might not be able to answer them during the test. I want to learn about your experience with using the website, so I can't guide you. I will be glad to answer all your questions after the test.

If you don't mind, I'd like to ask you to sign this consent form. It simply states that you give me permission to record the session and that I am going to secure this data and make it available only to the researchers working on the data dashboard which includes me and my advisor, Dr. Vetria Byrd. Please go over it carefully and thoroughly. Also, please ask me about any questions that you may have before giving your consent to participate in this study.

(hand over the consent form)

With your permission, I'm going to record the session. I will record your voice and actions on the screen – that will give me the full picture of the session for further analysis. I will not share these recordings outside of this research.

Also, I hope you don't mind, but I'm going to take notes. It will help me in analyzing the session later.

Before we begin, could you please look at this quick pre-test questionnaire? It will ask you to fill in some basic demographic information and familiarity with data visualization.

(Provide the pre-test questionnaire to the participant)

If you feel ready, let's begin with the usability testing module.

I will first give you 5 minutes to use the website. After the time is over, we will begin with the tasks.

(Open the webpage and hand it over to the user to use it. Start the recording device. Start the timer once they begin. After the time is up, begin with the tasks)

There are 30 tasks that I would like you to perform today. I will read the first task now. If you have any questions about it – please ask them right away.

(Read the first task. Also, hand over a printed version to them in case they need a reminder on the tasks later on)

Start whenever you feel ready and once again – please share your thoughts with me out loud.

(Do each task one by one)

Thank you, that was a very helpful session. Your feedback is extremely valuable to me.

(Stop the recordings and ensure that they are saved properly. Ask the user to step away from the computer)

Thank you for participating in the usability test. You will now be enrolled in a raffle for winning one $20 Starbucks gift card. Since there are 20 participants for the usability test, your probability of winning the gift card is 5%. I will contact you by email if you win the gift card. Thank you again for your time and contribution in this research. Have a good day.

(Escort the participant carefully out of the lab)
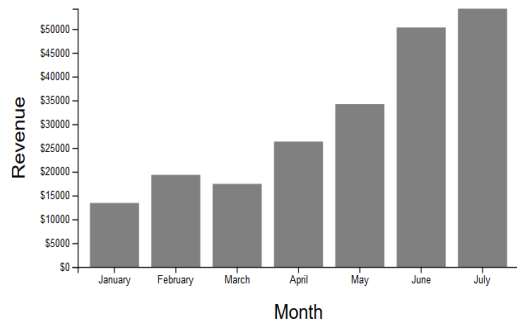
APPENDIX B

Usability Testing – Pre-test questionnaire

Please fill in the demographic information in the given sections:

1. Name: _____
2. Age: _____
3. Gender: _____
4. Major: _____
5. Highest level of education completed: _____
6. Industry experience (in years): _____
7. Data visualization experience (creating data visualizations, evaluating, teaching, programming, using tools, courses, etc.): _____
8. How relevant is education and workforce data to you?
☐ Absolutely relevant
☐ Somewhat relevant
☐ Neutral
☐ Somewhat irrelevant
☐ Absolutely irrelevant

Please check the visualizations that you are familiar with (seen before and know how to interpret them):
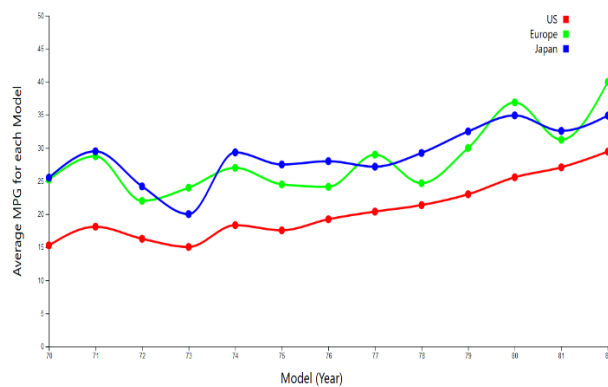
☐ Bar graph
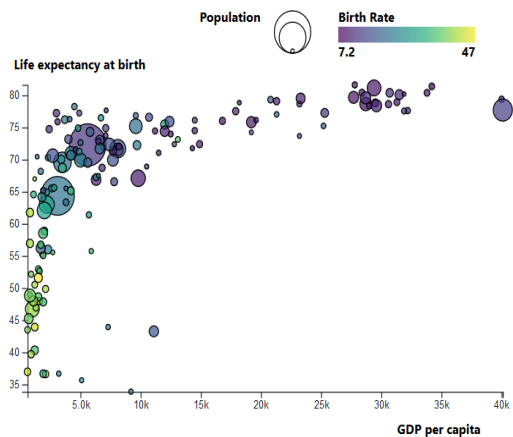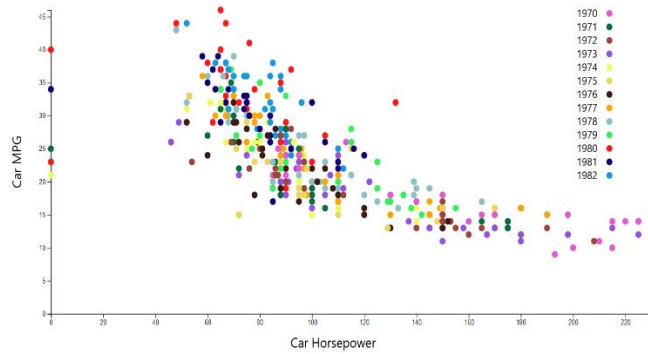
☐ Grouped bar graph
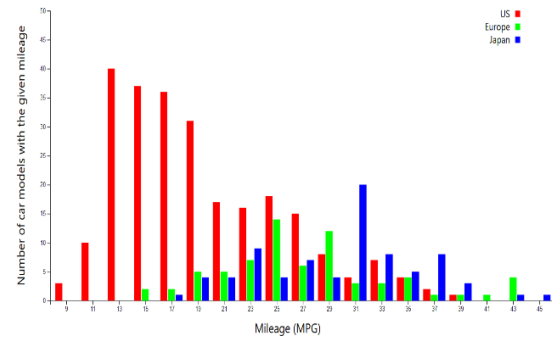
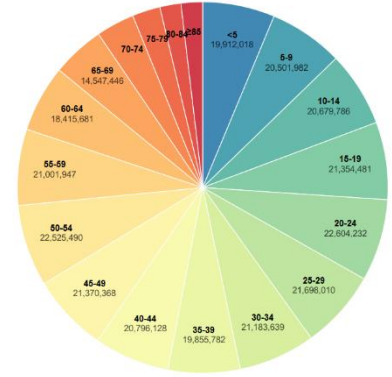
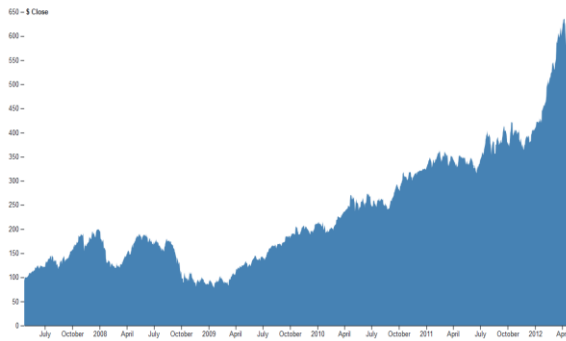

☐ Stacked bar graph
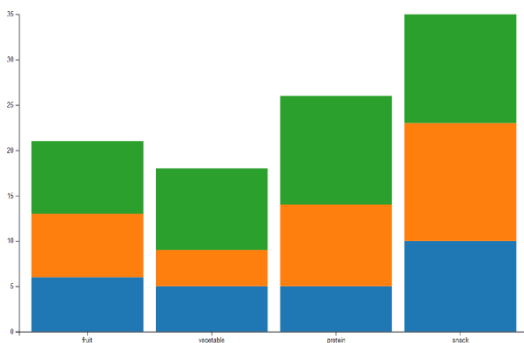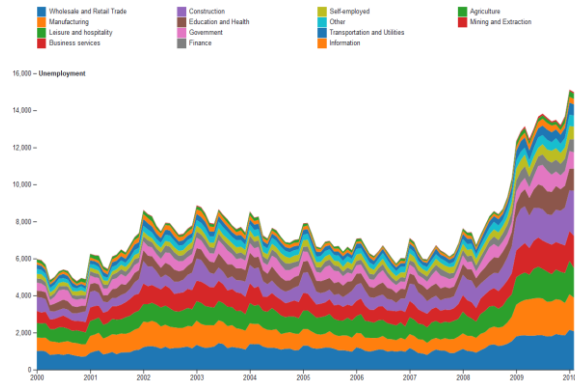
☐ Histogram
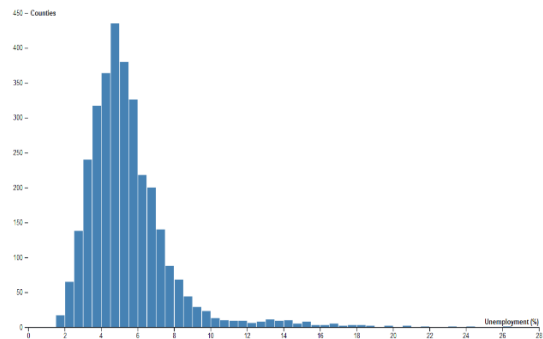
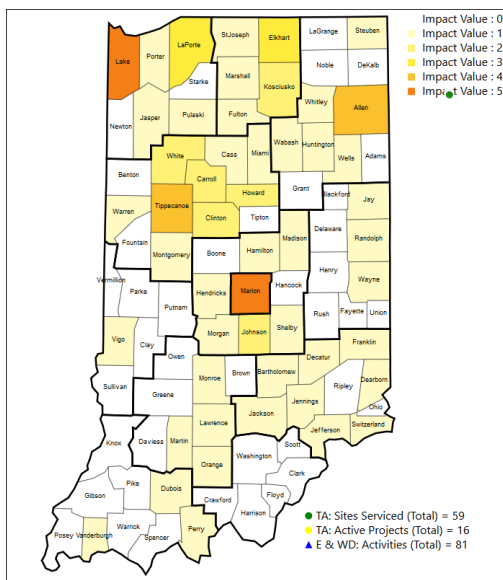

☐ Line chart

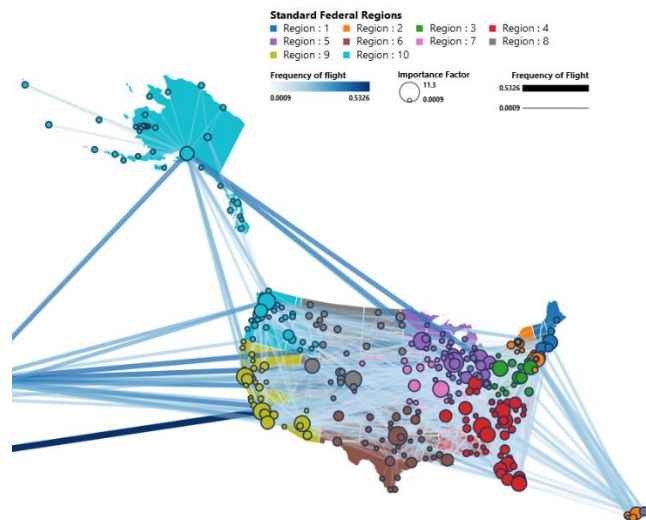☐ Bubble chart

☐ Scatter plot

☐ Pie chart
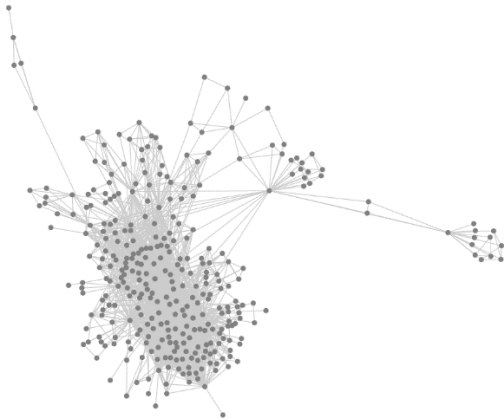

☐ Area chart

☐ Stacked Area chart

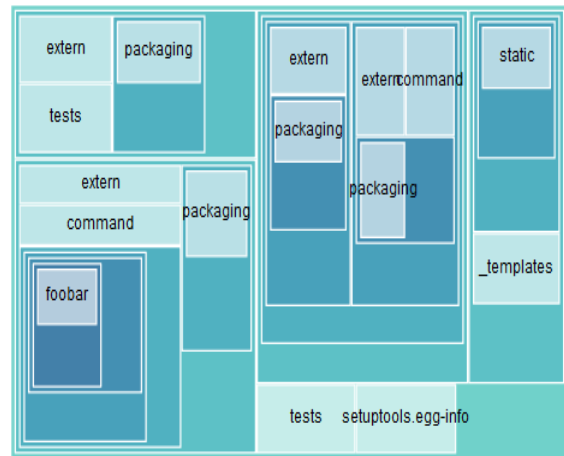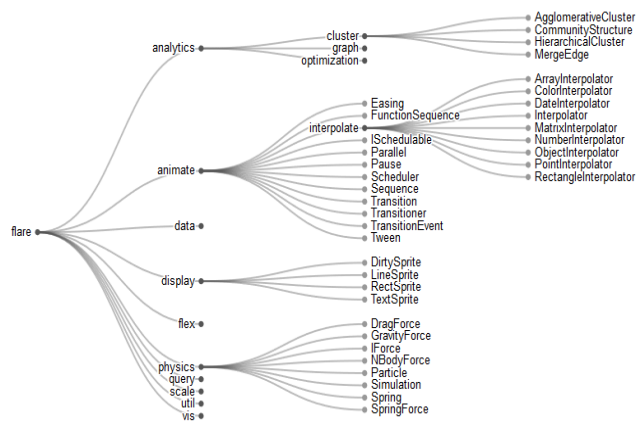☐ Choropleth map

☐ Bubble map
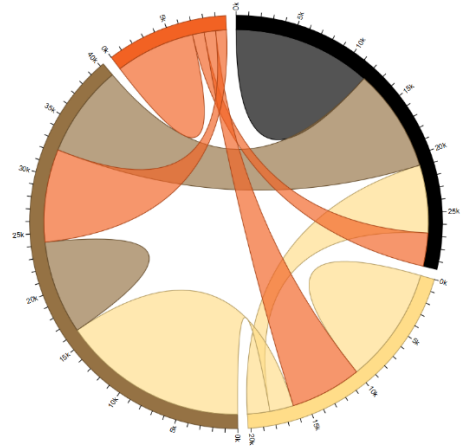
☐ Node-link graph

☐ Treemap

☐ Dendrogram

☐ Chord diagram

APPENDIX C

Usability Test – 30 Tasks

Please complete the following tasks one by one as guided by the researcher.

1. Which dataset are you currently looking at?
    (Information presentation – To test the highlighting of the current dataset on the navigation panel on the left of the dashboard and user's ability to read the title of the page)

2. How many datasets are currently being visualized through this dashboard?
    (Navigation – To test the user's ability to find and scroll through all the datasets on the navigation panel)

3. What is the total impact of IN-Mac in "Tippecanoe" county and how many Technology Adoption (TA) and Education & Workforce Development (E &WD) activities were conducted there?
    (Information presentation – To test the effectiveness of the tooltip to present important information about the data)

4. Navigate to the "Micro-Grant" dataset.
    (Navigation – To test the navigation panel on the left)

5. In which Economic Growth Region does "Brown" county fall in?
    (Visualization – To test the ability of using the map and legend to find/search for a specific county)

6. What are the highest and lowest microgrants awarded in region 9?
    (Visualization, interaction mechanism – To test the action of clicking on the county to

view the region bar graph or use the 92 county bar graph. The user's ability to interpret the bar graph visualization was also tested)

7. You want to contact a personnel from the "Caterpillar" organization in "Tippecanoe" county, region 4 to ask about their microgrant. How would you find the details about their name, email and location?
   (Interaction, filtering mechanism, navigation – To test the use of map for selection, then drop down for filtering and information box to view the data)

8. What are the statistics about this data for the state of Indiana?
   (Information presentation – To test the interpretation of the statistics information box)

9. You want to find out details on the activity that was conducted using the microgrant awarded to "1st Maker Space, LLC" from Marion county, region 12. How would you find this information?
   (Interaction, filtering mechanism, navigation – To test the user's readability of important information from the information box)

10. What is the Micro-grant program about?
    (Information presentation – To test the readability and placement of the dataset description and title)

11. What are the highest and lowest Micro-grants awarded overall for the entire state?
    (Visualization – To test the use of the 92 county bar graph for comparing data fields)

12. Scroll to the bottom of the page. Now without scrolling back up, tell me how many Micro-grants were awarded in region 11?
    (Filtering mechanism – To test the use of drop down on the scatterplot to filter the data shown)

13. How many Micro-grants have an impact of more than 500 students across all regions?
    (Filtering mechanism – To test the use of the slider on the scatterplot to filter the data shown)

14. Who created this project? Who owns the copyright?
    (Miscellaneous aspects – To test the readability and accessibility of the footers)

15. For an impact between 0 to 1500, amount between $1,200 to $4,000 and region 4, how many organizations are there?
    (Filtering mechanism – To test the use of all filtering mechanisms (sliders and drop down) on the scatterplot together)

16. You work in the Education department in the state of Indiana. You want to make some suggestions to the developers of this dashboard. How will you get in touch with them?
    (Miscellaneous aspects – To test the visibility and use of the 'Contact' tab on the top right corner of the dashboard to contact the project developers.)

17. What is the name of organization in region 10 that requested $2,000 and had an impact of 400 students?
    (Filtering mechanism – To test the use of filtering mechanisms, tooltip and ability to click on the circle to view and update the data on the corresponding text box on the right side of the page)

18. You are an employee at IN-Mac. You want to visit the website of 'Ivy Tech-Columbus' in Jennings county, region 9 to find out if they posted the information about the Micro-grant that was awarded to them. How will you find this information?
    (Navigation, interaction mechanism – To test the ability to click on the weblink to go to the organization's website)

19. Navigate to the "Schools and Principals" dataset. What percentage of the total schools in the state of Indiana are in region 3?

(Navigation, visualization – To test the user's ability to view and interpret the pie chart)

20. Scroll to the bottom of the screen. Now without using the scrollbar or scroll button, navigate back to the top of the page.
    (Navigation – To test the use of the ^ button in the bottom right corner of the page to go back to the top)

21. How many Economic Growth Regions (EGR) are there in the state of Indiana?
    (Information presentation – To test the readability of the title for the map).

22. Navigate to the "Ecosystems" dataset. For the Micro-grants, how many organizations fall under the 'Primary Education' Ecosystem? Name them.
    (Visualization, information presentation – To test the usability of the node link graph and tooltip to view data. Also, test the readability of the text on top of the nodes)

23. You are intrigued by this project. How can you find more information about it?
    (Miscellaneous aspects - To test the visibility and use of the 'About' tab on the top right corner of the dashboard to find more information about the dashboard.)

24. In which county does the "Orthoworx Indiana" organization from the 'Industry Ecosystem' belong to?
    (Visualization – To test the interactivity with the node link graph to select an organization and view details about it in a text box)

25. Imagine that you are the principal of "Clark County Middle/High School" located in "Clark" county, region 10. You want to find in which year the attendance was the lowest so that you can investigate further into the cause. How will you find this information?
    (Visualization, interaction, filtering – To test the use of the line chart and tooltip to find information.)

26. Now you want to know the attendance for the latest year i.e. 2017-18. What is the attendance?

    (Visualization, interaction, filtering – To test the use of the line chart and tooltip to find information. Also test if the user uses the table instead to retrieve the information)

27. Suppose you work in the administration department of "Paul Harding Jr. High School" located in "Allen" county, Region 3. You are responsible for collecting attendance data for the school. Do you have data for all the years from 2005-06 to 2017-18? If not, for which years is the data missing?

    (Visualization, interaction, filtering – To test the use of the visualization to find missing data)

28. Suppose you work in the administration department of one of the schools in the state of Indiana and are visiting a potential client who is willing to invest to create new facilities in your school. To show how much students love coming to school, you want to show the excellent attendance record over the years. How will you create such a document that shows the current selection on the data dashboard and the graphs and tables related to your attendance data?

    (Miscellaneous aspects – To test the use of the 'Generate reports' button to generate a report from the current selection of the user.

29. Imaging that you are having difficulty in using the dashboard. Is there any guide or tutorial on this website that can help you to understand the dashboard better? If there is, where can you find access to it?

    (Miscellaneous aspect – To test the use of the help icon '(?)' on the top right corner of the dashboard to find a video tutorial on how to use the dashboard.)

30. Imagine that you have been using this dashboard for a while and you seem to be lost. You want to go back to the home page/first page to start over. How will you go there?

    (Navigation – To test the use of the 'IN-MaC' icon to navigate back to the home page)

APPENDIX D

Usability Test Notes

Participant: {Name of the participant}

Date of the test: {Date}

Researcher: Yash Gugale

Task	Observation
1	
2	
3	
4	
5	

6	
7	
8	
9	
10	
11	
12	
13	

14	
15	
16	
17	
18	
19	
20	
21	

22	
23	
24	
25	
26	
27	
28	
29	

30	

Indiana Education and Workforce Data Visualization Dashboard – Semi-structured interview study brief
Data collected by Yash Gugale

Date

(Welcome the participant and have them seated)

Hello {participant name}, thank you for coming. My name is Yash Gugale and I'm a graduate researcher in the department of Computer Graphics Technology at Purdue University. I am working in the Byrd Data Visualization Lab for visualizing education and workforce data for the state of Indiana – I will walk you through the session today. We will be conducting a semi-structured interview today. This study is designed to give the researchers a better understanding of the end user's behavior, skills, thought processes, requirements and work practices so that we can improve our data dashboard.

Before we start, I have some information for you.

I am currently testing the Indiana Education and Workforce Data Dashboard, that I created to learn as much as I can about the way people use it. Every product is intended to work in a certain way by its creators, but as you know – the reality might be quite different. The goal of this research is to get as close to reality as possible.

Let me inform you that you are free to withdraw from the session any time you want without giving us an explanation. If you are not comfortable answering any questions, please inform us and we will skip that part.

Do you have any questions so far?

If you don't mind, I'd like to ask you to sign this consent form for the semi-structured interview. It simply states that you give me permission to record the session and that I am going to secure this data and make it available only to the researchers working on the data dashboard which includes me and my advisor, Dr. Vetria Byrd. Please go over it carefully and thoroughly. Also, please ask me about any questions that you may have before giving your consent to participate in this study.

(hand over the consent form)
With your permission, I'm going to record the session. I will record your voice as we conduct the interview. This will give me the full picture of the session for further analysis. I will not share these recordings outside of this research. Also, I hope you don't mind, but I'm going to take notes. It will help me in analyzing the session later.

The main purpose of this interview is to get an in-depth understanding of the end user of the system, so that we can redesign the dashboard so that it allows you to do all your tasks well and use the dashboard effectively. This interview session will last for 45 minutes to an hour.

Before we begin, could you please look at this quick pre-interview questionnaire? It will ask you to fill in some basic demographic information and familiarity with data visualization.

(Provide the pre-test questionnaire to the participant)

Let me remind you again that you that you are free to withdraw from the session any time you want without giving us an explanation.

If you are ready, let's begin with the interview.

(Start the recording device. Conduct the interview as per the questions listed)

Thank you for providing me with all the information. I sincerely appreciate your contribution.

(Stop the recording device and ensure that the data is saved properly)

With all your valuable feedback, we should be able to create a better data visualization dashboard for you. You will be enrolled in a raffle for a Starbucks gift card worth $20. Since there will be 10 participants who will be interviewed, your chance of winning the gift card is 10%. We will inform you if you win the gift card. Thank you again for your participation in the semi-structured interview.

(Escort the participant out of the room)
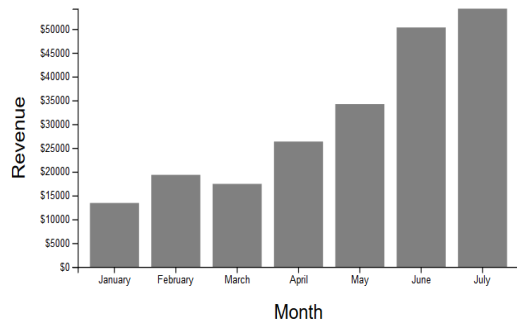
APPENDIX F

Semi-structured interview – Pre-interview questionnaire

Please fill in the demographic information in the given sections:

9. Name: _____
10. Age: _____
11. Gender: _____
12. Major: _____
13. Highest level of education completed: _____
14. Industry experience (in years): _____
15. Data visualization experience (creating data visualizations, evaluating, teaching, programming, using tools, courses, etc.): _____
16. How relevant is education and workforce data to you?
☐ Absolutely relevant
☐ Somewhat relevant
☐ Neutral
☐ Somewhat irrelevant
☐ Absolutely irrelevant

Please check the visualizations that you are familiar with (seen before and know how to interpret them):
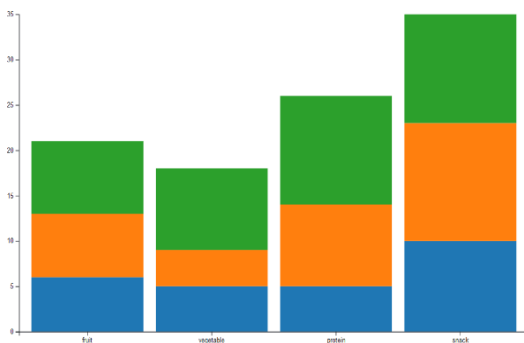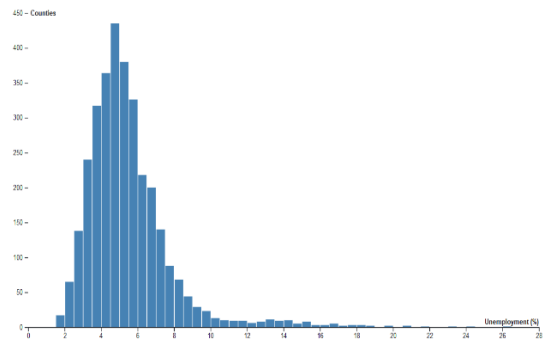
☐ Bar graph

☐ Grouped bar graph
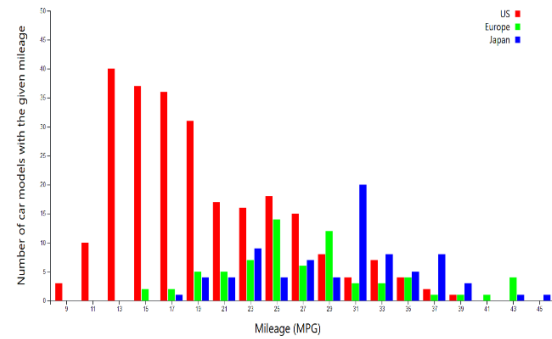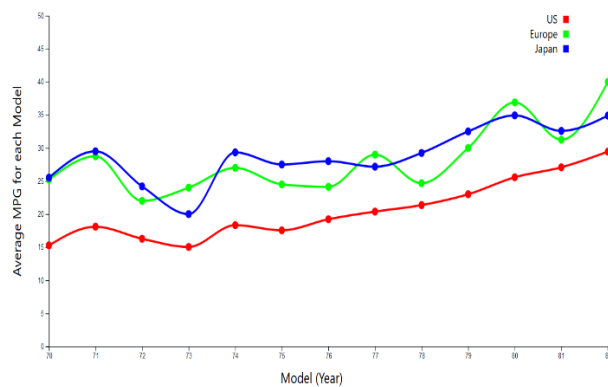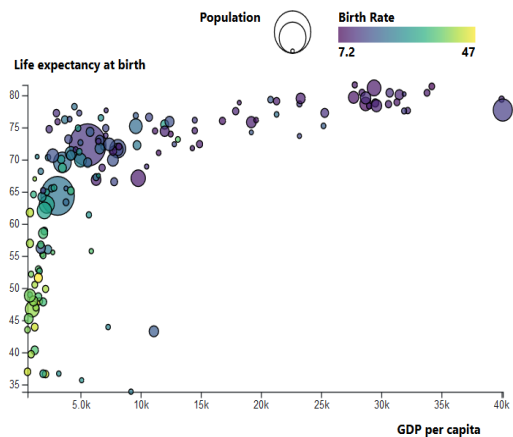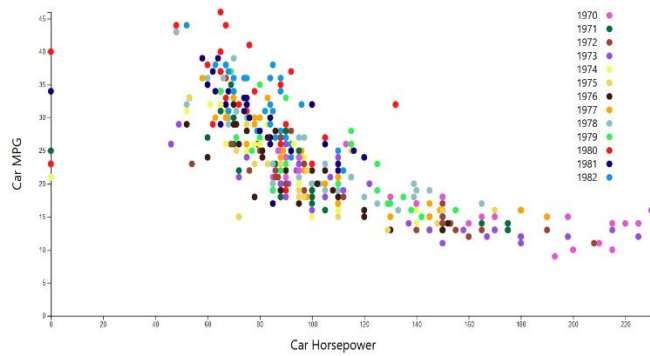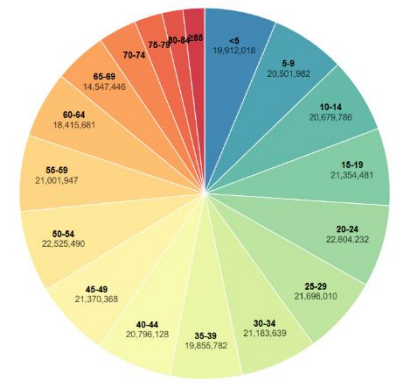

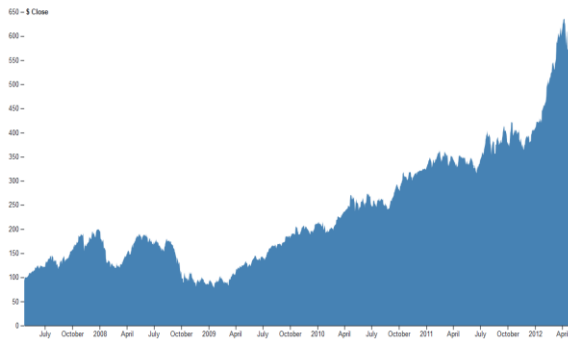☐ Stacked bar graph
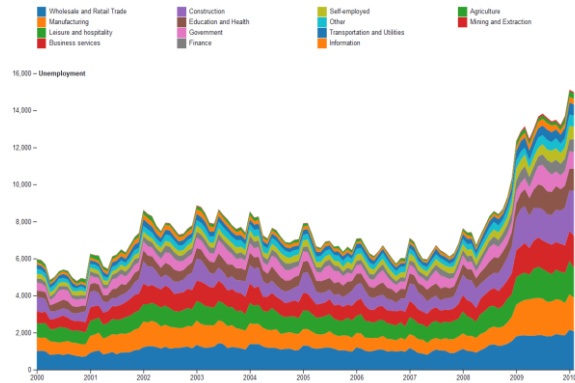

☐ Histogram

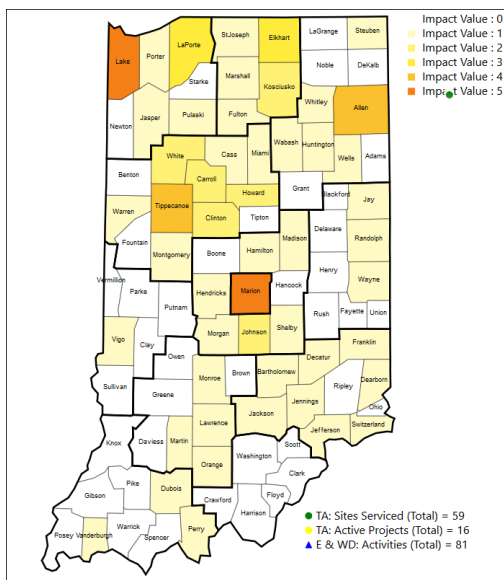

☐ Line chart

☐ Bubble chart

☐ Scatter plot

☐ Pie chart

☐ Area chart

☐ Stacked Area chart

☐ Choropleth map

☐ Bubble map
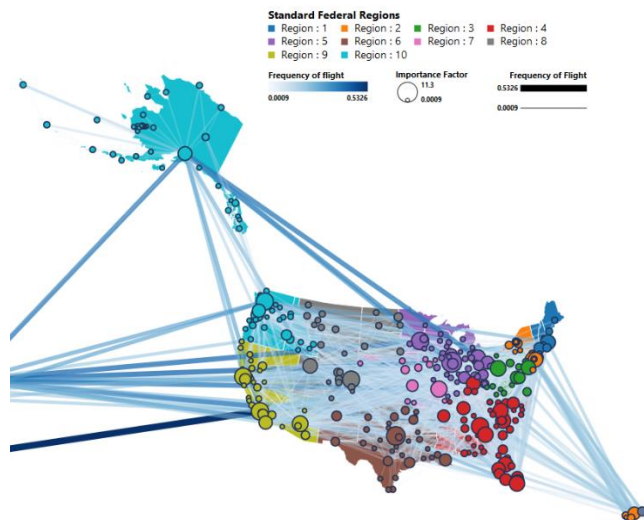
☐ Node-link graph
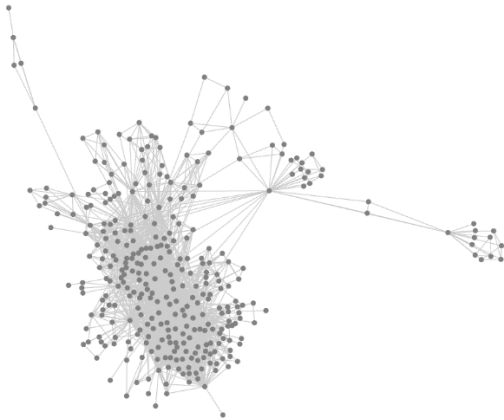
☐ Treemap
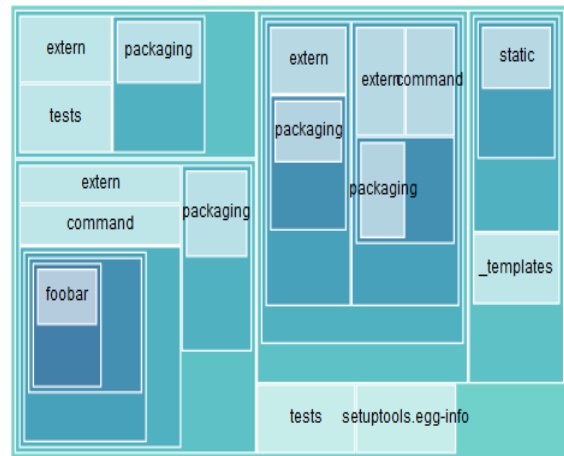
☐ Dendrogram
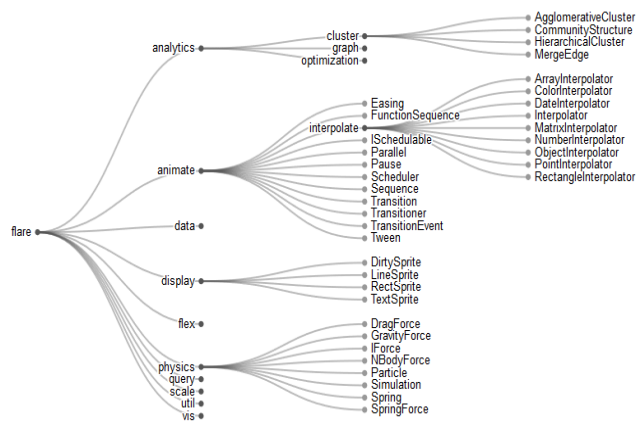
☐ Chord diagram
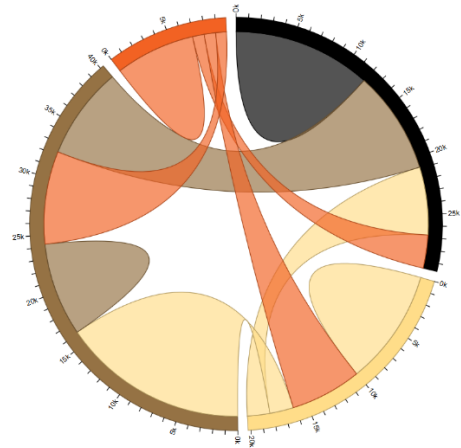
APPENDIX G

Semi-Structured interview questions

The main purpose of this interview is to collect information about the target end user of the data dashboard that has been created to assist with workforce and education identifiers and work practices. Identifying the tasks they perform, the tools they use, their behavior, and familiarity with using data visualization tools and technology, will greatly help in designing the data dashboard for the intended users.

The interview will last for 45 minutes to an hour. Through the semi-structured interviews, the following information will be collected:

Industry:
1. Which industry do your work in?
2. What is your primary role/position?

Tasks:
1. What aspect of your daily work involves dealing with education and workforce data?
2. What kinds of tasks do you perform on a daily basis?
3. What kinds of tasks did you expect to perform with this data visualization dashboard?
4. How useful is this dashboard for you in performing your tasks?
5. What prompted you to use the dashboard?

Tools:
1. Which data visualization and analysis tools do you currently use?
2. What do you like and dislike about the tools provided?
3. What role does the tools play in decision making?
4. What are the characteristics of the ideal tool that will help you to perform your tasks?

5. What are the challenges you face in your work for which you think that this dashboard can help?

Data:

1. Which datasets are you primarily interested in?
2. Why is that data relevant/important to you?
3. What decisions do you plan to make using this data?
4. How frequently would you access that data (hourly, daily, weekly, monthly)?
5. How frequently do you want the data to be updated?

Dashboard:

1. What do you like about this dashboard and why?
2. What do you dislike about this dashboard and why?
3. What limitations/challenges did you face while using the data dashboard?
4. What features would you like to have in this dashboard and why?
5. How did you use this dashboard to collaborate with others?
6. How do you see this dashboard to be a major contributor in decision making?

Data visualization and analysis:

1. Were you able to understand and interpret all the visualizations currently on the dashboard?
2. How do you think data visualization will help you to solve your problems?
3. What analysis would you like to perform with the help of this dashboard?
4. What statistics should this dashboard compute for you?

Platform:

1. Which platform would you prefer - A website, mobile or desktop application and why?

Open ended:

1. Is there anything else you want to tell me about your experience that we haven't already discussed?

APPENDIX H

Thematic analysis - Codebook

RQ 1: What are the guidelines/visualization design recommendations to be followed while creating a data dashboard for visualizing education and workforce data?

Themes:

1. Design dependent on characteristics of the user and affects the data

a. Consider the end user's frame of reference while designing the visualizations and dashboard elements.

1. Prior user frame of reference:
Definition: Use of color coding and breakdown by county and the 12 Economic Growth Regions (which replicates the design of the state map) while designing the map of the dashboard, helps the user to quickly adapt to using the data dashboard.
Example: I have that map in the back of one of my notebooks so that I can refer to it.

b. Maintain consistency of view in the dashboard pages and visualizations across all the datasets.

1. Dashboard design for visual consistency:
Definition: Consistent dashboard design and layout across all datasets to ensure ease of use
Example: I like that it's consistent across all its tabs.

c. Provide users with the ability to generate reports from the data.

1. Generate report feature:

Definition: Generate reports from the dashboard for analysis and sharing with others. Also, ability to customize the report and related formatting and alignment of content

Example: So if I generated a report from Allen County, I'd like to see what schools we're working with, what programs, what companies we worked with, what companies we did Tech Adoption on, that's just a very useful tool.

d. Design the data and visualization interfaces to support additional datasets that may be required by the user.

1. Missing relevant data:

Definition: Datasets that the user is interested in, but are currently not visualized on the dashboard. There are several datasets since each user was interested in looking at a different aspect of the education and workforce data.

Ex:

1. Employer data

2. (Un)employment data

3. Free and reduced meals data

4. Graduation rate data

5. Job data

6. Major industries

7. Manufacturer data

8. Salary data

Example: I think including our graduation data, so that employers or someone that's trying to do a site visit can pull up the data real easy, because we have this data.

e. Use desktop web applications for detailed analysis of data and mobile for overview or quick snapshots

1. Dashboard platform:

Definition: Preferred choice of platform (mobile or desktop) for the dashboard based on the consideration of use and challenge of each,

Example: There's a lot of data on here. So I think desktop is good for me as a user, preferred for me as a user.

f. Ensure stability, maintainability and ease of use

1. Dashboard maintainability:

Definition: Ensure that the dashboard is easy to maintain by allowing easy management of data and visualizations

Example: Again, you get into so much data that it gets hard to maintain.

2. Dashboard usability:

Definition: Ensure that the dashboard is easy to use with simple interactivity and intuitive navigation mechanisms

Example: I've got this and I can quickly navigate to something else.

g. Design for users with varying work roles (from management, planning, technical, etc.)

1. User prior experience level:

Definition: Experience of the user of the data dashboard with respect to:

1. Data literacy (how much the user knows about the data)

2. Visualization literacy (how well the user knows how to interpret and reason about different visualizations)

3. Statistical literacy (what statistics can the user compute and understand)

4. Technology use (which technology platforms can the user use comfortably)

Example: I'm not a statistician, so that part is not relevant to me.

2. Characteristics of users work:

Definition: User's daily tasks, work activities, role, primary work, work challenges and work relevant data

Example: Engineering, technical, project management, program management, administrative support, business planning, program planning, strategic development, a lot of support for researchers as well.

3. User's work focus:

Definition: Industry or field that is the focus of the user's work (such as manufacturing, education, curriculum development, target area or county of focus, etc.)

Example: And then I'm in charge of our biggest events, so I do a lot of planning with vendors for those.

4. User:

Definition: User of the data visualization dashboard. They can be from industry, education, government, non-profit or any entity (including common public) working with or interested in education and workforce data at several levels.

Example: Like we have a new chancellor in South Bend, this would be so valuable for him, right, to sit there and say this is what's going on in this workforce region.

5. User's client:

Definition: The people, organizations, community or any entity that will be influenced by the actions or decisions made by the user (E.g. STEM educators, STEM partners)

Example: My goal is to help employers build customized training solutions for things like - to help them increase their skills in their workforce, increase retention, decrease turnover

6. User's partner:

Definition: The people, community, organization, industry or any entity that will work in collaboration with the user to further the cause of education and workforce development in the state of Indiana.

Example: And again, I look to partner with IN-MaC a lot.

2. Design dependent on the characteristics of the data and affects the user

a. Provide sufficient description of the dataset, what the data contains and what it doesn't.

1. Data description:
Definition: Description and meaning of the datasets and individual data fields that it contains. Explanation of the data's content - what's included and what's excluded in the data and criteria for making that decision.
Example: I don't think we have to, but maybe some more information in regards to business and industry would be helpful.

2. Data type:
Definition: Types of data in the data fields. It includes categorical, ordinal, numeric and descriptive data
Example: what the number of industry people are in a specific county.

b. Ensure clarity is conveyed:
A. Through the visualizations
B. About the datasets

1. Visual appeal:
Definition: Use of marks and channels (colors, shapes, etc.) helps to aid the affinity and reduces the time and effort required in interpreting the visualizations.
Example: Certainly, visually and aesthetically pleasing

2. Convey clarity:
Definition: Clarity in the visualizations helps to find trends, outliers in data and increases its utility and avoids overwhelming the user
Example: But the figures should always be clear, because it's going to have many users.

c. Allow the user ability to control the granularity of data that he is interested in looking at.

1. Data granularity:

Definition: Presenting the data at various levels and allowing the user to control the granularity - from detailed data to overview of the data. This allows for filtering the data and ensures that the data does not overwhelm the user.

Example: So it really breaks down to a granular level

2. Data organization:

Definition: Combining data from various sources and organizing them into clusters and intuitive groups (such as by area) so that the user can control its granularity.

Example: Because maybe I don't use the same language that they use, or the same ideas or they group them in, you know, they group them different than how I would think about grouping them.

d. Ensure data is valid (avoid old data), reliable (avoid missing data and broken fields) and updated frequently (quarterly to yearly)

1. Data validity concern:

Definition: Concerns about the presence of invalid content in the data fields.

Example: But if you start getting too many errors, or too many web links not found, people will be like, it's great, but it's not accurate.

2. Data reliability concern:

Definition: Concerns about missing data and broken fields or incorrect data, which can cause the user to lose trust in the reliability of the dashboard for decision making.

Example: So working forecasts that are coming from data sets that are pretty reliable.

3. Data update concern:

Definition: Concerns about ensuring that the datasets are up to date and avoiding outdated information.

Example: One of my biggest concerns is making sure that we have the updated information and I know we're pulling that from a source that we have no control over.

e. Provide data source and last update information.

1. Data source information:
Definition: Provide information on the source that the data was collected from and validity and reliability of that source.
Example: Because then then they'll know okay, that's where they're pulling that data.

2. Data update information:
Definition: Provide information on when the data was last updated.
Example: When you click on that data set, there needs to be an inset that says this data is generated from DWD, last updated 2016.

f. Use area breakdown (by county/region) as a starting point to access the data.

1. Area breakdown:
Definition: Breakdown of the state of Indiana in 92 counties and 12 Economic Growth Regions (EGR) and using that as a navigation and filtering mechanism while accessing data. Since users perceive and access data by area, maintaining the area-data relationship helps the user. Color coding based on the 12 EGR (which matches the state map) further acts as a visual aid during data access.
Example: I think they want to say, boom (clicking on a county to get the results), okay, top five jobs (info presented from clicking on the county), this and so on.

RQ 2: What impact does an information visualization dashboard have in the work processes and decision making of stakeholders involved in education and workforce development in the state of Indiana?

Themes:

1. Data access

a. Centralized access to data

1. One stop shop:
Definition: Having access to data in one centralized location in a composite format makes it easy to find data, helps avoid accessing multiple datasets from different sources, increases the data access speed and reduces the time spent in finding relevant data. It is a "one stop shop".
Example: It's also kind of a one stop shop, because all the entities we deal with really are on here.

b. Easy access to various datasets through common interface

1. Consistent data access interface:
Definition: Dashboard helps to ease the access of data since all the pages have similar layout for visualizations, features, interactivity and method of data access and navigation. This helps the user avoid confusion as he moves through the different datasets.
Example: I think the ease of use, you got the ease of use, you're already far down that pathway and probably ahead of most of what we have now.

c. Frequently accessed to obtain relevant information (daily to a minimum of weekly access)

1. Dashboard access frequency:

Definition: The number of times the dashboard will be accessed for various use cases (hourly, daily, weekly, monthly or yearly basis)

Example: I definitely think it's something that I personally would access at a minimum, monthly, but probably even weekly.

2. Expected frequency of data update:

Definition: The period of time after which the data is expected to be updated (weekly, monthly, quarterly, yearly basis)

Example: I think you definitely have to do a yearly. I think quarterly might be good enough. That's a practical number.

d. Dashboard helps avoid dealing with raw (dry) data directly

1. Experience with current tools:

Definition: Use case, challenges faced and overall experience while using current data visualization and analysis tools.

Example: Well as with the current tools that we use, part of the challenge is, you know, in developing, in creating any graph and creating any chart, part of the dialogue with your participant, with your audiences, is you need to get them up to speed on what your intentions are.

2. Experience using raw data sources:

Definition: Difficulty in finding relevant data and the necessary contact information is the main challenge. Datasets are related to education and workforce. They are scattered and publicly available (such as IDOE and DWD data) and collected through other organizations (EMSI, Burning Glass, Conexus, etc.)

Example: That we have to collect everything and kind of put it in our own framework. It takes a lot of time and energy, and we would much prefer just to have a one stop shop.

2. Inform and communicate

a. Find information to connect, communicate and collaborate with other people, organizations and partners.

1. Connection use:
Definition: Use of dashboard to develop connections between users, users and their partners or users and their clients. They can consist of employer-employee connections, school-industry connections, etc.
Example: Or just try other routes, maybe other people I know who may have connections there, to get to the right people.

2. Communication use:
Definition: Use of dashboard to communicate and reach out to clients, partners, etc. This activity consists of sharing knowledge and information, selling training, showing statewide development, etc. and doing this communication through relevant data and visualizations.
Example: You know, just keeping on top, keeping up with things and what this can help us do is communicate upward, you know, to both inside the university and outside the university of what we're doing, of what we've done, of where we're going and why.

3. Collaboration use:
Definition: Use of dashboard for collaboration and co-creation between the users, partners, community and industry.
Example: Specifically, in region 4, we have huge collaboration between, you know, elected officials and education, and community organizations.

b. Find gaps, plan work and create more opportunities.

1. Finding gaps:
Definition: The use of dashboard to find problems, needs and gaps within the industry and education, such as unemployment, employee retention, educational qualification, etc.

Example: This is the number of people who have completed in the last five years that you put out in the community, and these are the ones needed, but based on your enrollment, you have a gap.

2. Planning use:

Definition: Use of data dashboard to plan activities and initiatives to fulfill gaps in workforce and education. It consists of planning the design of education pipeline, workforce pipeline, talent pipeline, manufacturing pipeline and more broadly, the education-to-job pipeline.

Example: And so a lot of that is planning, programming for workforce pipeline activities for students, K through 12.

3. Decision making use:

Definition: Use of dashboard for making fast data-driven decisions that expedite work processes.

Example: That's very valuable, because these people that are really making decisions, they're trying to make that decision within as little time as possible.

4. Create opportunity:

Definition: Develop and grow opportunities by creating initiatives for business attraction.

Example: So, what is going on in Fountain county, and how can I help to bring students to help them understand career exploration, and to understand their opportunities within their community for employment after graduation, or even to look at options for robotics and coding.

5. Development use:

Definition: Use dashboard for launching initiatives and activities to help support and develop the community. This helps the overall economy and development of the state.

Example: So, do we, as a regional entity and for our economic development, do we look at putting things in, to help to come alongside the public education in that community?

6. Seek funds:

Definition: Seek financial support through grants and funds for training and other activities

involved in education and workforce development.

Example: So, we don't necessarily pay them but we need funding sources to be able to support bringing them to the camp, feeding them at the camp and helping them to move forward in the robotics team.

7. Education and workforce development activity:

Definition: It consists of diverse activities that enhance learning through collaboration, teamwork, games and industry exposure through expert talks, manufacturing training, management training, career exploration, etc.

Example: So we're going back to some of those things that are really hard to teach unless they're working in a group setting like teamwork and problem solving, and, you know, who rises to be the leader and who is, you know what your roles are.

8. Skill development:

Definition: Developing the relevant technical skills in emerging technologies such as programming, robotics, supply chain, etc. all related to the manufacturing sector.

Example: If we don't know those emerging skills, then we just keep training to existing skills, that's the key that's missing on a lot of these, is emerging competencies.

c. Generate awareness.

1. Increase awareness about development:

Definition: Use of dashboard to know the current state of development in various counties, EGR and overall across Indiana. This also consists of comparing data across different areas. The use can be for work or even personal interest.

Example: You guys should get the bump from that but this would let me know, one, you are heavily involved in Sullivan county, so I might reach out to you if we don't have anything going on.

2. Increase awareness about partners and clients:

Definition: Use dashboard to find current partners, clients and look for potential partners and

clients to work with.

Example: As I look at the 'post secondary institutions for all 92 counties (refers to 92 county bar graph), my first response is, who knew there were that many?

3. Increase awareness about activity and connections:

Definition: Find and show opportunities for collaboration by viewing the activity and connections between organizations and partners.

Example: And then, you know, when you break it down even more, you get to see how all of us are existing with one another and co-creating this space together.

3. Prepare and collaborate

a. Dashboard use to support or back conversations with facts.

1. Support communication with data:

Definition: Use of dashboard to provide data that will support the conversations with facts and serve as an entity to back presentation content and suggestions for development. This can help to change the perception of the user's clients and partners.

Example: And they'll be like, 'they run one posting for 40 jobs that are the same'. Well, that's true but still matches up with the workforce data. So looking at the two data points, it still looks pretty accurate.

2. Dashboard for marketing:

Definition: Use of dashboard by an organization (such as IN-MaC) to market their work, development and future goals. This helps to save time and efforts and eases communication. It can also be used to market the skilled workforce to get them employed.

Example: This helps get us in the door, we can fit a nice visualization like this on a trifold flyer or in a presentation, you know, as a lead developer, as a business development in.

b. Prepare for meetings and presenting data.

1. Prepare for meeting:

Definition: Using dashboard to make notes and prepare for a meeting. This consists of analyzing data, using visualizations to aid understanding and plotting activities.

Example: And I could quickly make notes to go in. And if I'm meeting with economic development or with the DWD working group for a grant, I could come in with so much information based on the partners I know and what they do.

2. Presenting data:

Definition: Present data during meetings to other colleagues, partners, bosses and influencers in the state using the dashboard.

Example: It's really more presenting what we've done.

c. Maintain a consistent view about the data between all users to avoid inconsistencies.

1. Maintain consistent data perception:

Definition: Use of a common version of the data dashboard for communication and information sharing. This allows all partners involved in the conversation to look at the same version of data and visualizations to avoid gaps and inconsistencies in communication.

Example: Well, on a practical level, if you're all looking at the same screen with the same information, it's easy to come to an agreement I think on different things.

d. Dashboard helps to raise questions about the data from the interpretation of visualizations

1. Investigation from visual interpretation:

Definition: Using the visualizations and data presented on the dashboard to raise questions about the data, its meaning, correctness (accuracy) and what it conveys.

Example: I'd love to find out what DOE (Department of Education) defines attendance 97.3%?

Other miscellaneous codes:

a. User expected features.

1. Expected feature:

Definition: A feature, behavior that the user is interested in or expects the dashboard to support.

Example: It'd be nice to have a header page (on the report) that had the same header here (on the webpage) that we could type up, like just a box that you could type up, say, 'hey, here's what we want to do with this data'.

APPENDIX I

Source code for the data dashboard (GitHub): https://github.com/yashgugale/IN-MaC-Data-Dashboard