Exploring Ecosystems in Indiana's Education and Workforce Development Using a Data Visualization Dashboard

Yash Gugale
MS CGT

Academic Advisor: Dr. Vetria Byrd
Committee member: Dr. Tim McGraw
Committee member: Dr. Paul Parsons

INTRODUCTION

Background

- Education and workforce data about the state of Indiana consisting of a variety of datasets such as schools and principals, pre k-12 and postsecondary education data, businesses and industries, enrollment, attendance, etc. is collected every year and is available publicly on the Indiana Department of Education, The Indiana Department of Workforce Development and other websites.
- In this research, the goal is to create and test the design of an 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 and evaluated by conducting studies with the target demographic.

Problems

- There are large datasets covering various aspects of education and workforce
- They are largely textual data in excel sheets, PDF files, etc. and there is no central repository
- There is no visual aid to help the users find patterns, trends, outliers, etc. in the data
- They are not able to drill down the data to find a specific field that they are interested in
- There are issues with updating data and it leads to repetitive collection of the same data and difficulty in performing statistical analysis
- There is no coherent ecosystem
- There is an inability to collaborate or co-create

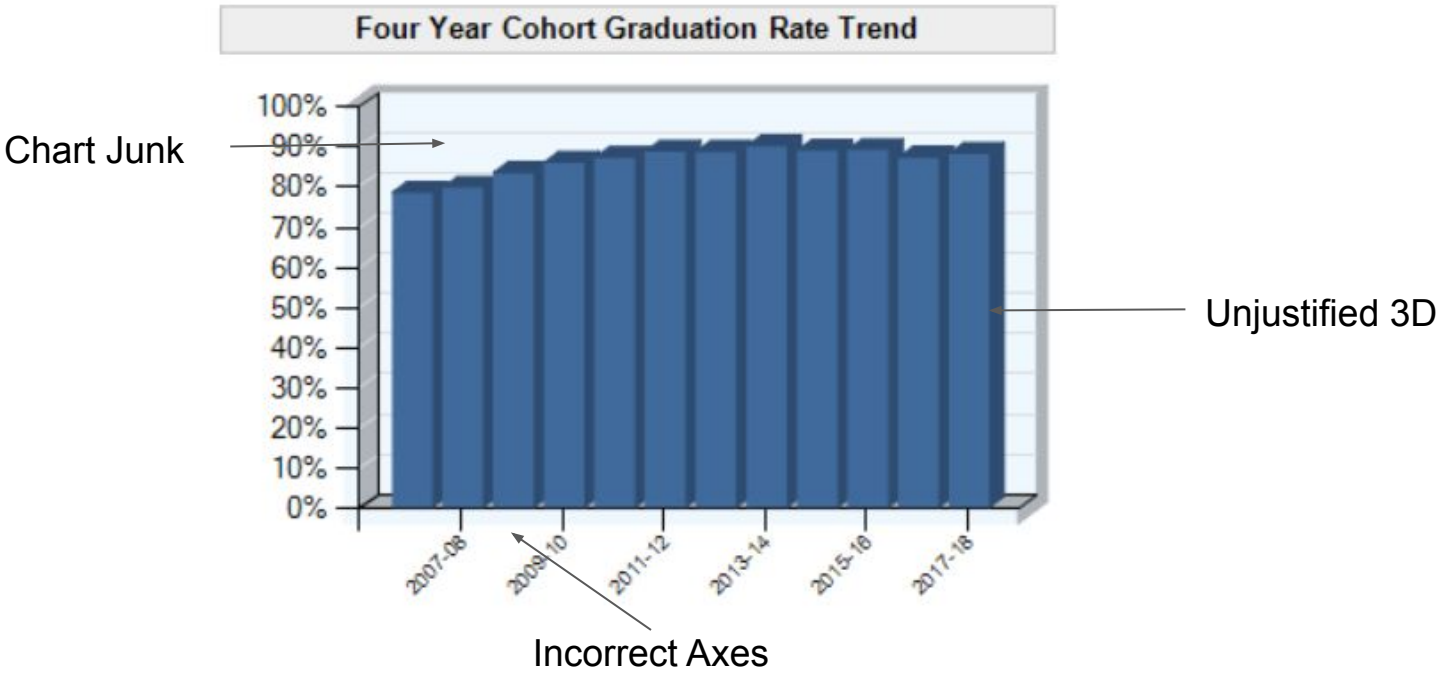
→ **All lead to poor decision making!**

Significance

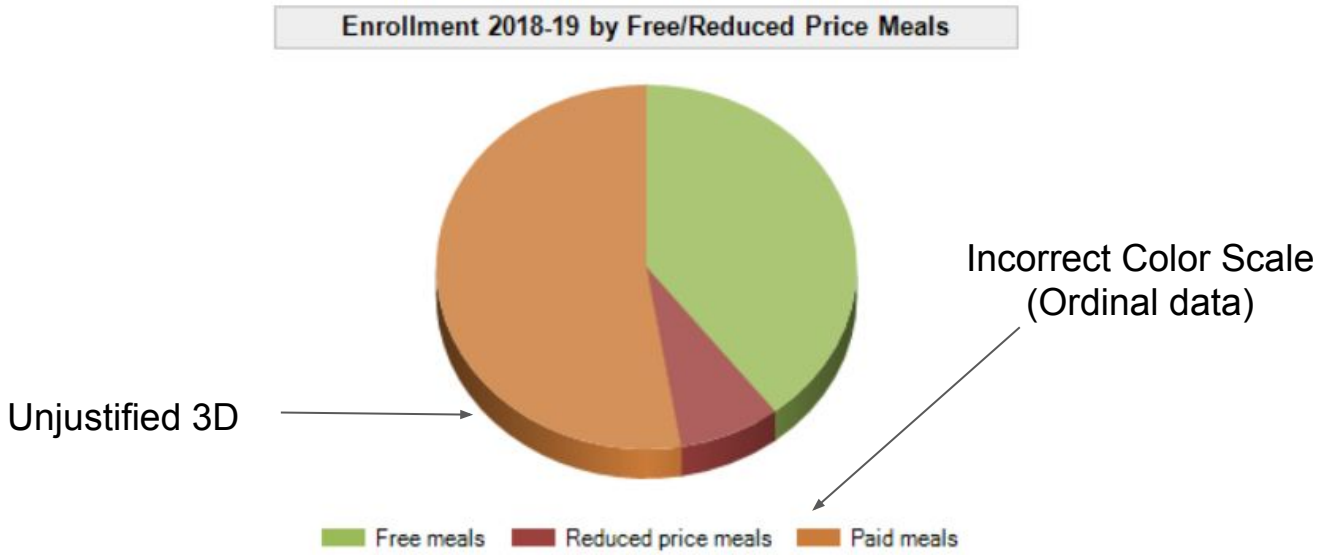
- Decision makers that influence the future state of the development of these domains need to have an accurate and multi-perspective view of the current state and need to see how such data is inter-related.
- This can help them in making important decisions such as directing funds, providing training 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.
- Education, workforce - significantly affect an individual's life and growth
- Thus, important to have a better understanding of the data

→ **Data visualization to the rescue!**

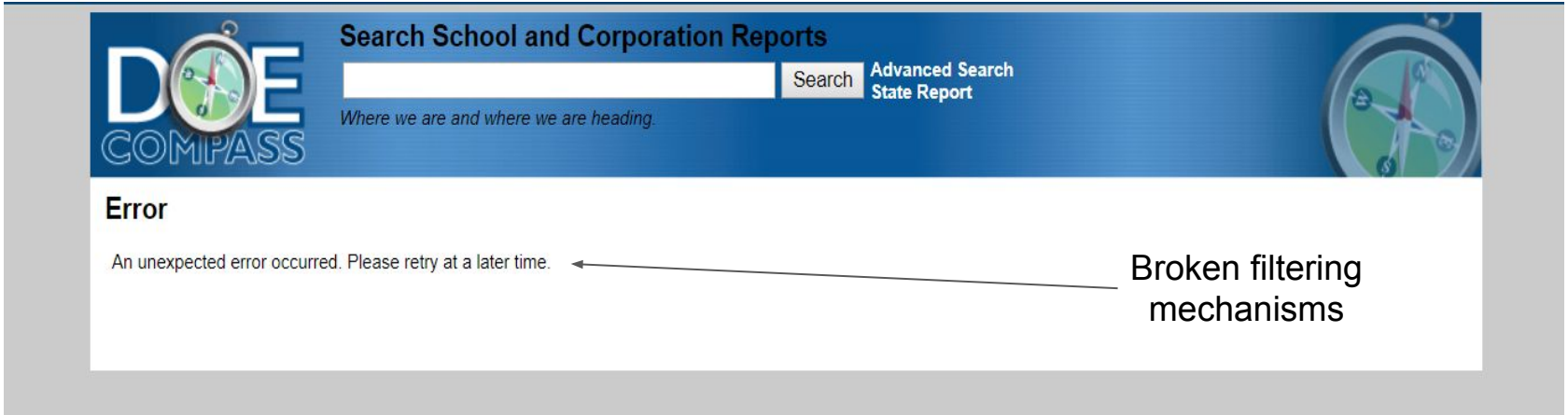
Failures!


Four Year Cohort Graduation Rate Trend
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
2007-08
2009-10
2011-12
2013-14
2015-16
2017-18
Chart Junk
Unjustified 3D
Incorrect Axes

Failures!


Enrollment 2018-19 by Free/Reduced Price Meals
Free meals
Reduced price meals
Paid meals
Unjustified 3D
Incorrect Color Scale
(Ordinal data)

Failures!



- No links to other relevant datasets
- Poor dashboard design
- How to explore/interpret the data?

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

Assumptions

User:

1. User has prior experience with using a website
2. User has a basic understanding and background about the datasets that are visualized.
3. User’s visualization literacy is unknown
4. User can understand and interpret statistical analysis terms such as mean, median, mode, etc.

Assumptions

Data:

1. Data is in JSON/CSV/Text format
2. Datasets are static
3. The data has been cleaned, pre-processed and additional fields have been added in the datasets so that they can be used for creating the required visualizations

Assumptions

User studies:

1. Sample represents the population
2. Questions on interviews are unbiased and do not direct the study participant towards a specific answer
3. 30 Tasks are enough and test all aspects of the dashboard design during usability testing

Limitations

- Dashboard is not dynamically updateable
- Not all education and workforce datasets are visualized
- Further studies on user centered design may suggest improvements
- Cannot use the entire population for user studies
- Users may wish to perform other tasks based on use case

Delimitations

- Dashboard visualizes 10 datasets
- D3.js is used for visualization
- Usability testing consists of 30 tasks
- There is no comparison of multiple visualizations for the same task
- Data dashboard is web based
- Datasets are specific to the state of Indiana

REVIEW OF RELATED LITERATURE

(Showing only the primary sources)

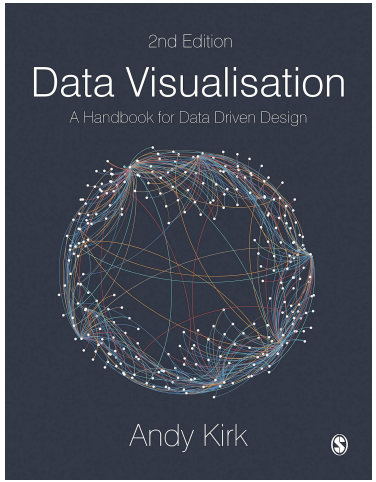
Tools

1. D3.js
2. Tableau
3. Cytoscape
4. Microsoft Excel
5. Bokeh

Visualizations

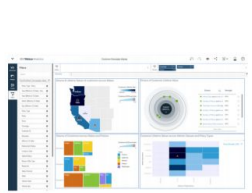
Best visualizations for the design and tasks were selected from 3 main sources:

1. Data Visualisation: A Handbook for Data Driven Design - *Kirk* (Visualization choice)
2. D3 website (https://d3js.org/) - *Bostock* (Feasibility for implementation)
3. VLAT: Development of a Visualization Literacy Assessment Test - *Lee* (User literacy)



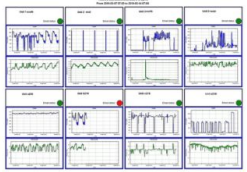
Data dashboards

1. What Do We Talk About When We Talk About Dashboards? - *Sarikaya*
2. Information Dashboard Design: Displaying Data for At-a-Glance Monitoring - *Few*
3. The big book of dashboards: visualizing your data using real-world business scenarios - *Wexler*



1 Strategic decision-making (DB052)

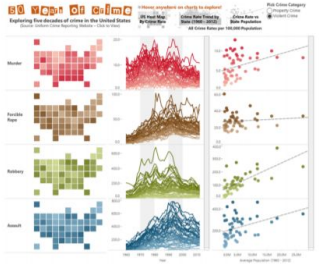
2 Quantified Self (DB021)



3 Static Operational (DB034)

4 Static Organizational (DB101)

5 Operational decision-making (DB036)



6 Communication (DB005)

7 Dashboards Evolved (DB045)

User studies

Usability testing:

1. Handbook of usability testing: how to plan, design and conduct effective tests - *Rubin*

Semi-structured interviews:

1. Qualitative research from start to finish - *Yin*
2. Evaluating information visualizations - *Carpendale*
3. Grounded evaluation of information visualizations - *Isenberg*

Thematic coding:

1. The coding manual for qualitative researchers - *Saldaña*

METHODOLOGY

Development Tools

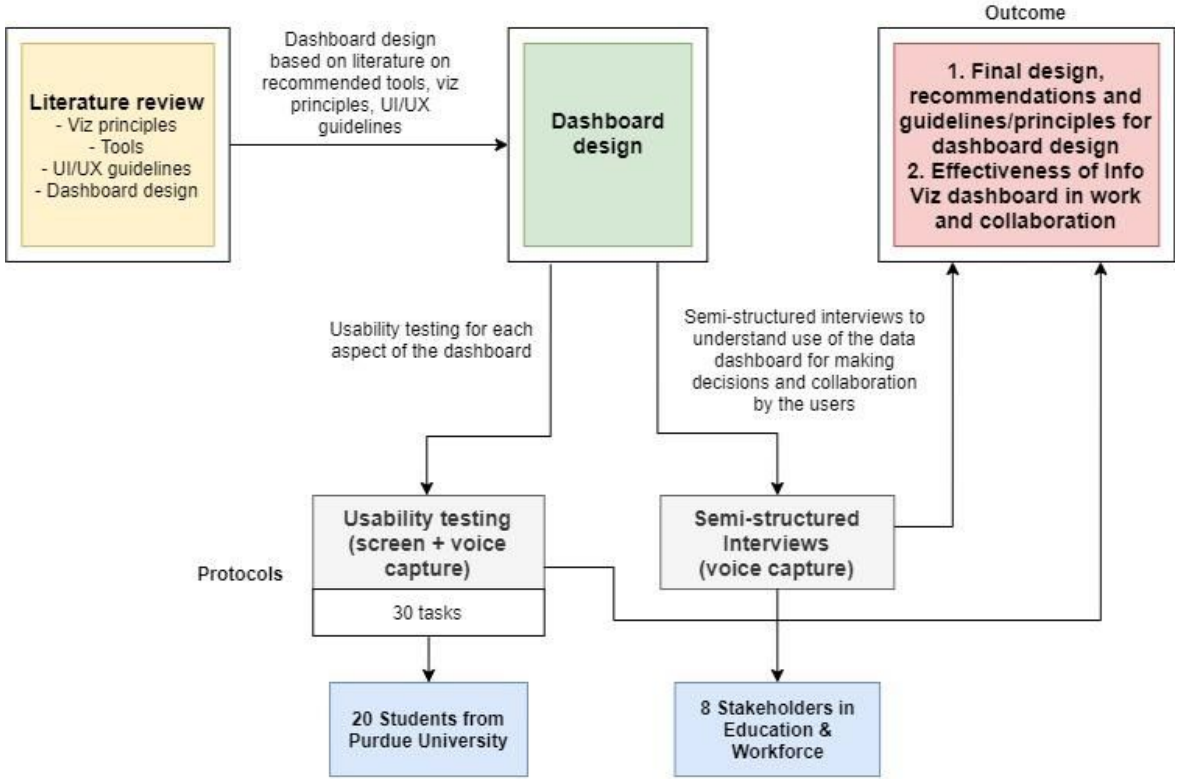
1. D3.js - For creating visualizations
2. Bootstrap (startbootstrap.com) - For layout and styling

Methods

Part 1: Data dashboard design and development

Part 2: Data dashboard evaluation through user studies

Research methodology



Research plan

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

Part 1: Data dashboard design and development

1. Dashboard template:

- A dashboard template was used from Startbootstrap.com 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.

2. Dataset selection:

Several datasets for this research were primarily obtained from the Indiana Department of Education, The Indiana Department of Workforce Development and IN-Mac. They were:

1. IN-MaC Impact data
2. IN-MaC Micro-grant data
3. Schools and Principals data
4. Pre K-12 data
5. Postsecondary 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

3. Visualizations and other interaction mechanisms:

The visualizations created were:

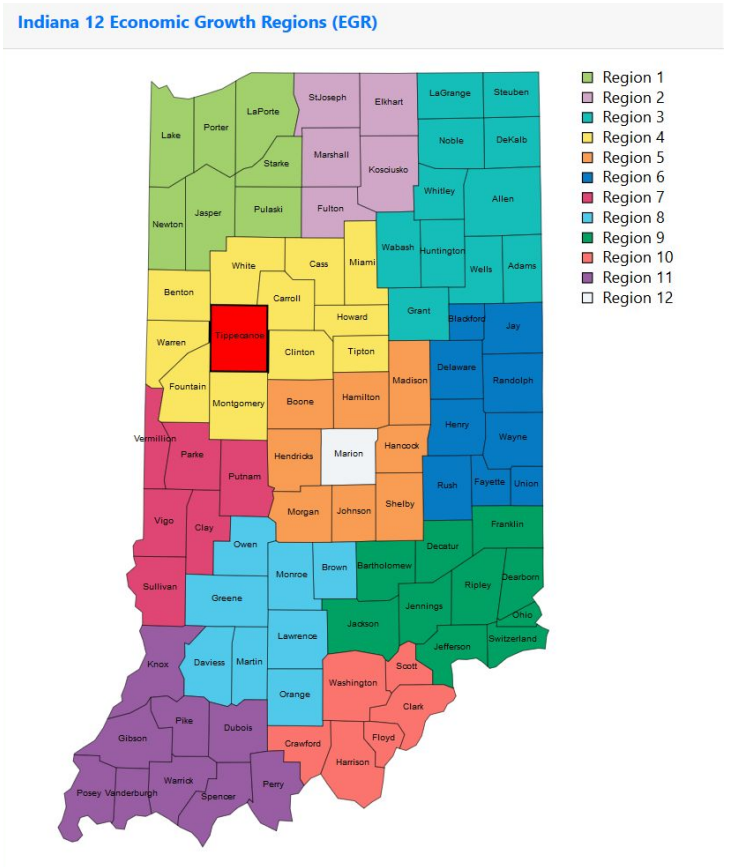
1. Proportional symbol map – to view IN-MaC impact
2. Region and county map – for selecting the counties to view corresponding data
3. Bar graph (region and 92 county)– to compare various fields
4. Line chart – to view the attendance of students in a certain school
5. Bubble chart – to compare youth impacted vs amount requested
6. Pie chart – to view proportions of schools in each region
7. Node link chart – to view ecosystems

3. Visualizations and other interaction mechanisms:

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

1. Text fields – to show large textual data such as project information and dataset description
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

From the government website



Visualization created using D3

Final data dashboard - www.yashgugale.com

- Dashboard homepage
- Map visualization
- Other visualizations
- Other features:
    1. Data drop down
    2. Information text box
    3. Stats text box
    4. Descriptive text
    5. About tab
    6. Contact tab
    7. Information tab
    8. Generate reports button
    9. Dataset navigation pane

Part 2: Data dashboard evaluation through user studies

1. Usability testing:

A usability test was conducted on 20 participants from Purdue University to test the usability of the dashboard.

Goal of usability testing

- 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 also helped in gaining an understanding users familiarity with the visualizations and their interpretations
- Analysis of the data collected helps to strengthen the validity of the design principles recommended from the evaluation of semi-structured interview data

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.
- 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.
- Sample task:

  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.)

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

Three main phases:

Pre-test:

- Consent form and voice and screen recording
- Pre-test questionnaire (data visualization familiarity and demographic information)
- 5 minutes to use the dashboard

Usability test:

- 30 tasks as guided by the researcher

Post-test:

- Further comments, suggestion, feedback
- Raffle for $20 gift card

Time required for the usability test

- Brief - 5 minutes
- Pre-test questionnaire and consent form - 5 to 10 minutes
- Tasks - 30 to 45 minutes
- **Total time: ~1 hour**

Time required for the usability test

- Brief - 5 minutes
- Pre-test questionnaire and consent form - 5 to 10 minutes
- Tasks - 30 to 45 minutes
- **Total time: ~1 hour**

Data collected

- Participant’s voice and actions on the screen were recorded
- Observations and notes by the researcher
- Demographic information (through pre-test questionnaire)

Part 2: Data dashboard evaluation and user studies

2. Semi-structured interviews:

Semi-structured interviews were conducted with 8 participants who were the stakeholders in education and workforce development in the state of Indiana.

Goal 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.
- 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.

Aspects to study

The interview was 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

- These questions were designed to keep the researcher on track with the aspects that were important.
- Other clarification and follow up questions were also asked to the participants to get a better understanding.
- Sample question:

  Q. How did you use this dashboard to collaborate with others?

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.
- 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

Three main phases

Pre-interview:

- Use the dashboard for a minimum period of 1 week for their daily work and tasks.
- They were not given any explanation on how to use the data dashboard.
- Consent form and voice recording
- Pre-interview questionnaire (data visualization familiarity & demographic information)

Interview:

- Interview questions (and follow up questions for clarification)

Post-interview:

- Raffle for $20 gift card

Time required for the semi-structured interview

- Brief - 5 minutes
- Pre-interview questionnaire and consent form - 5 to 10 minutes
- Interview- 45 minutes
- **Total time: ~1 hour**

Data collected

- Participant's voice was recorded for transcribing the interview
- Demographic information (through pre-interview questionnaire)

Time required for the semi-structured interview

- Brief - 5 minutes
- Pre-interview questionnaire and consent form - 5 to 10 minutes
- Interview- 45 minutes
- **Total time: ~1 hour**

Data collected

- Participant’s voice was recorded for transcribing the interview
- Demographic information (through pre-interview questionnaire)

Data analysis procedure

1. Analysis of usability testing data:

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

Performance data

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

1. Performance of participants on each individual task.
2. Performance of participant on all 30 tasks.

Preference data

1. Free-form comments and questions
2. Suggestions

(More details in the results section)

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.
- The original transcripts were first coded through open coding methods. This comprised of the first cycle.
- Then, focused coding was used to further categorize the data and find the themes.
- The two research questions were answered using the analysis performed on the semi-structured interviews.
- Analytic memos were also written during different phases of the analysis to help the researcher track their thought process and note any useful observations.

Transcription of interviews and preliminary reading

- The interviews were transcribed using Otter.ai.
- 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.

First cycle coding

- The transcripts were initially coded using open coding methods.
- 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 was used for coding using software.

Second cycle coding

- Focused coding method 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 codebook was rearranged as per the research questions and themes that emerged in the data so that it is easier for readers to follow the codes.

Population and sampling

1. Description of Population:

- Usability testing: Students from Purdue University above 18 years of age
- Semi-structured interviews: Stakeholders in education and workforce development in the state of Indiana

2. Sampling methods:

- Usability testing: Voluntary response method recruited through flyers and email
- Semi-structured interviews: Voluntary response method recruited through invitation by IN-MaC

3. Sample size:

- Usability testing: 20 students
- Semi-structured interviews: 8 stakeholders

3. Sample size:

- Usability testing: 20 students
- Semi-structured interviews: 8 stakeholders

RESULTS

1. Usability Testing Results

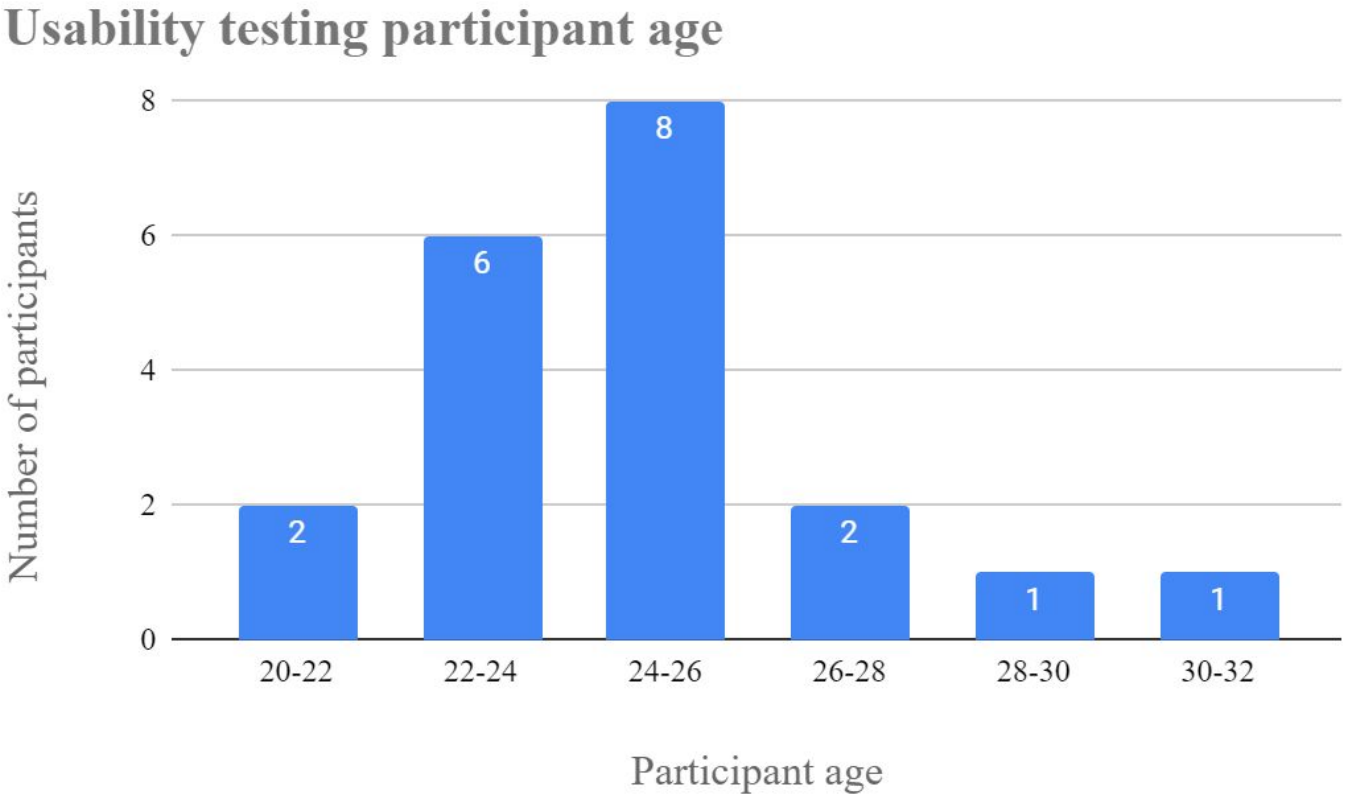
Usability testing results

1. Demographic information:

These statistics give us a general understanding of the demographic of the participants to further understand the results of our usability tests.

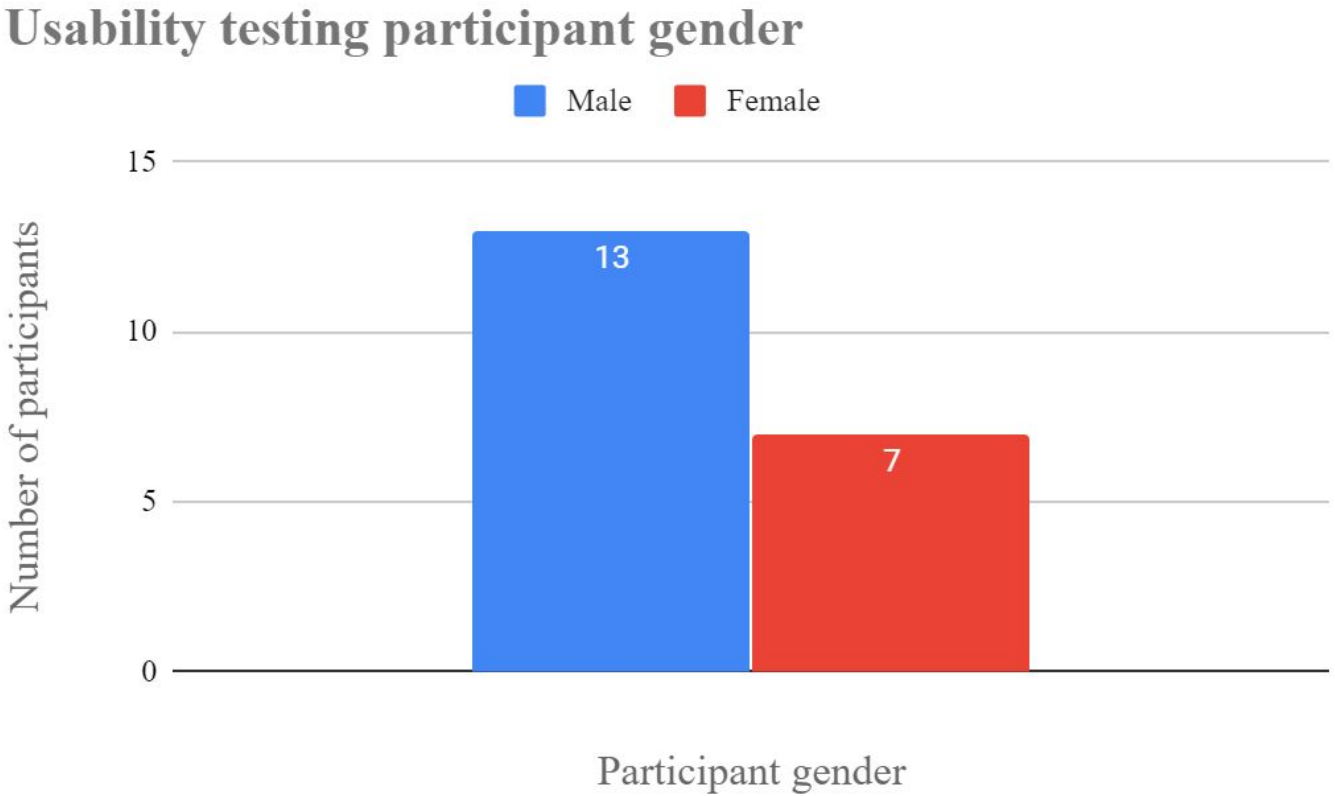
1. Participant age

(N = 20)



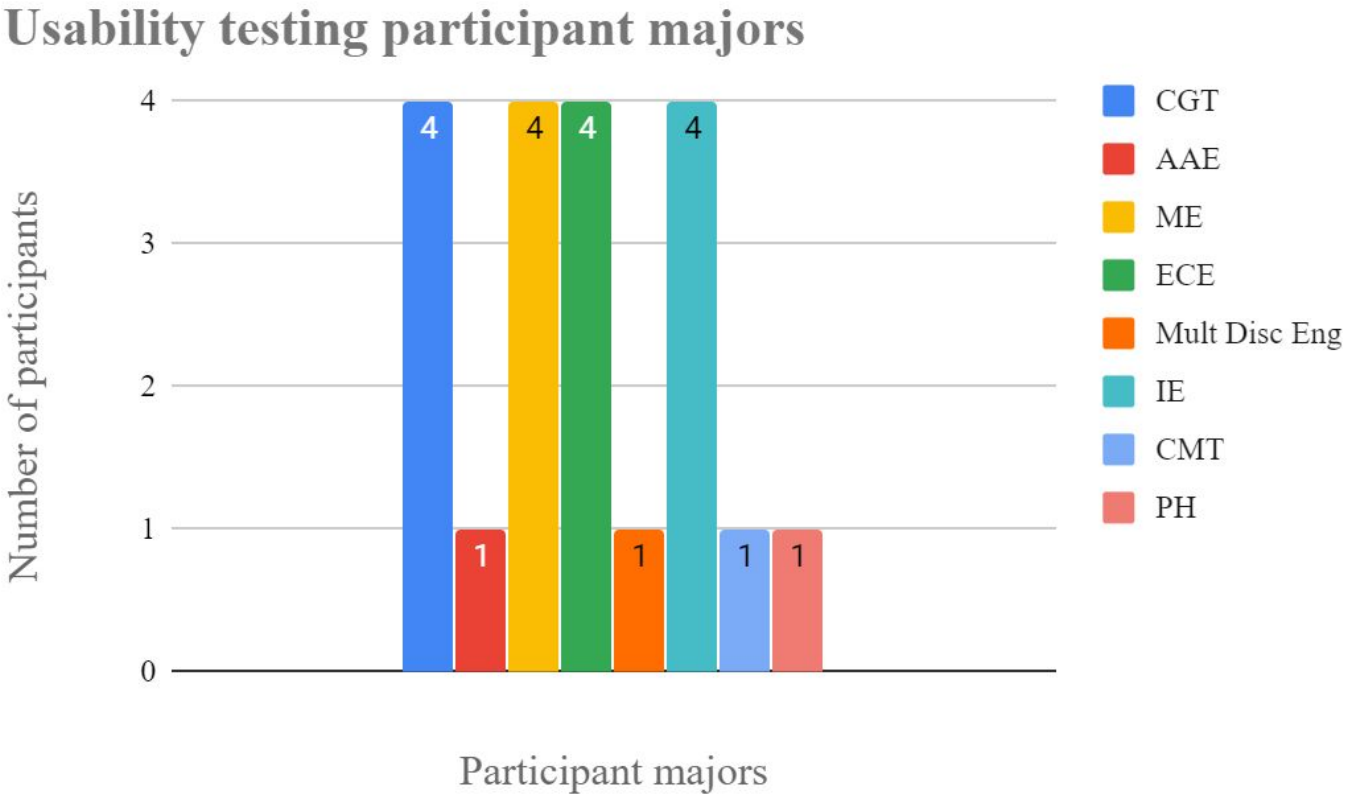
2. Participant gender

(N = 20)



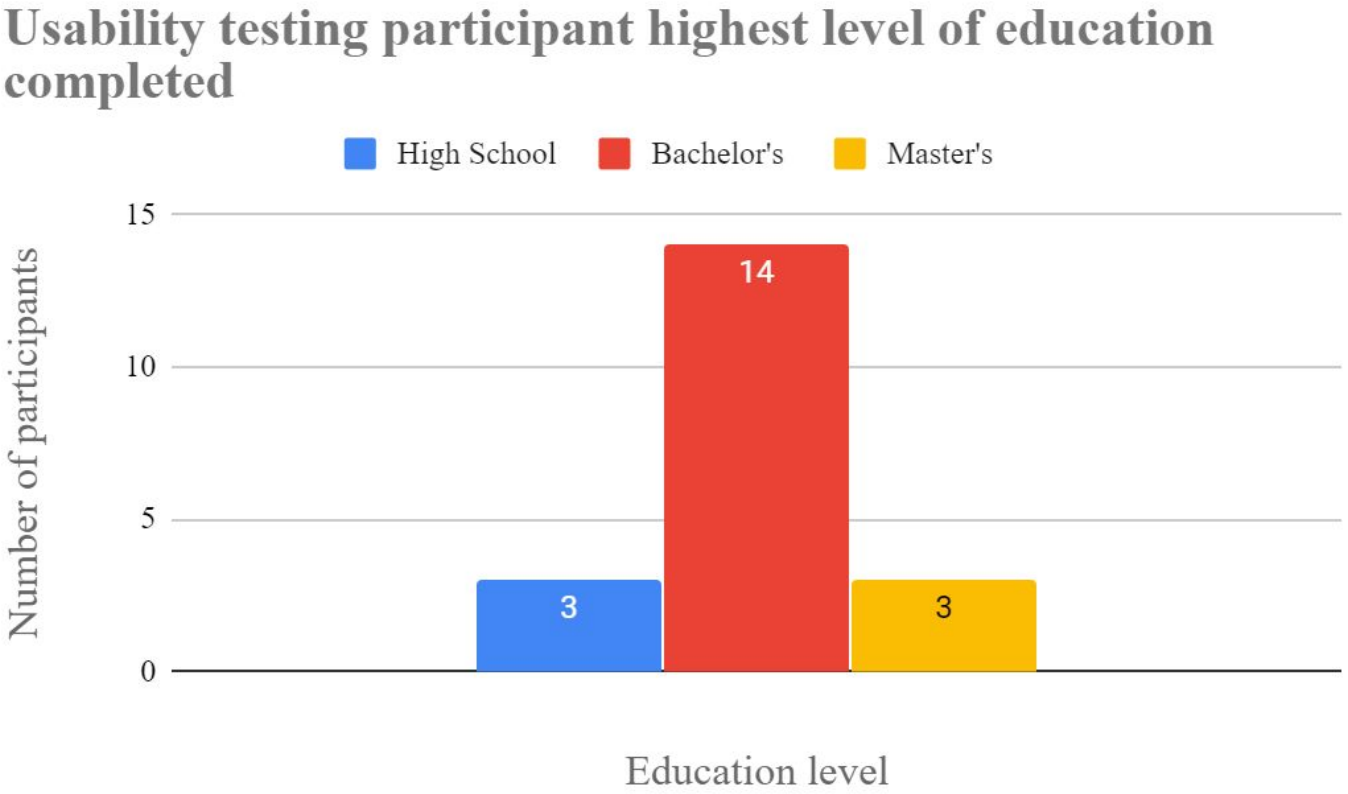
3. Participant majors

(N = 20)



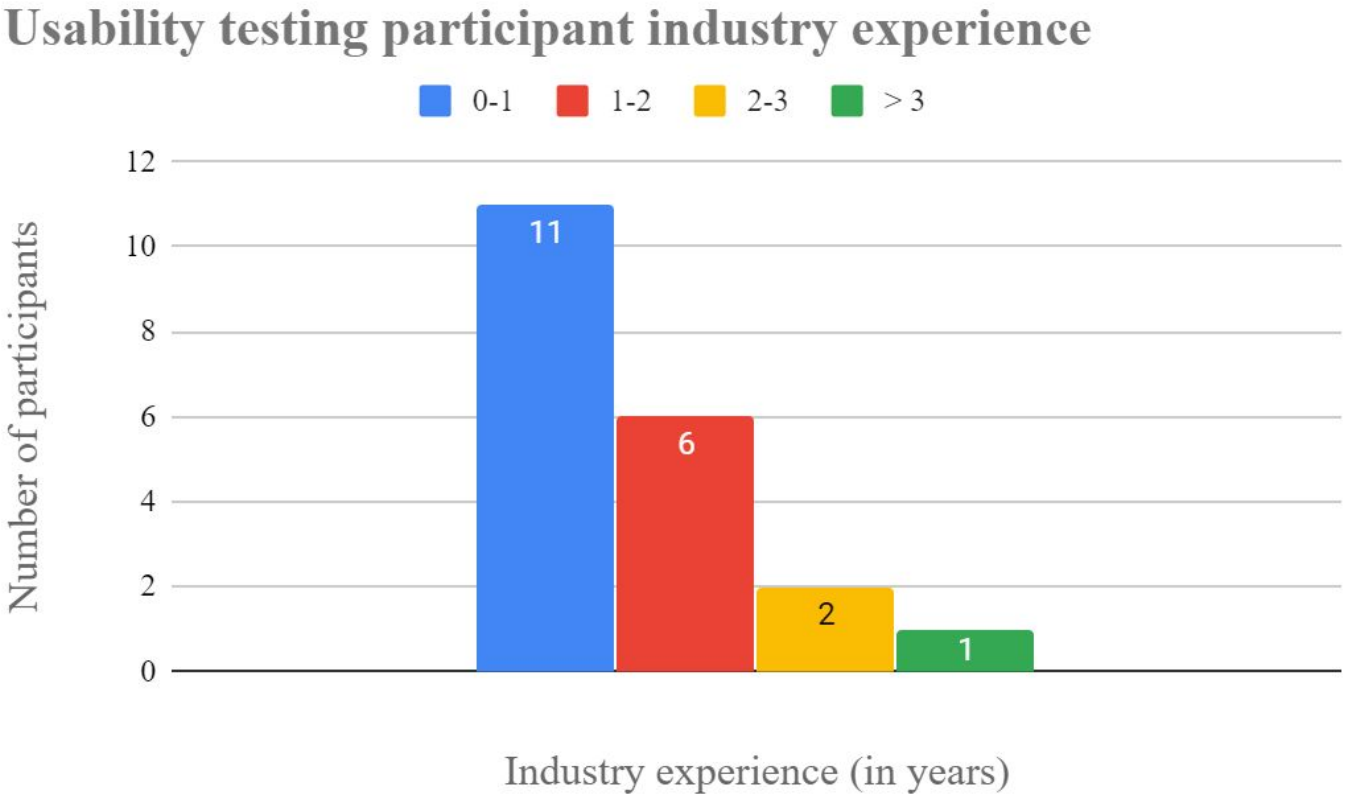
4. Participant education level

(N = 20)



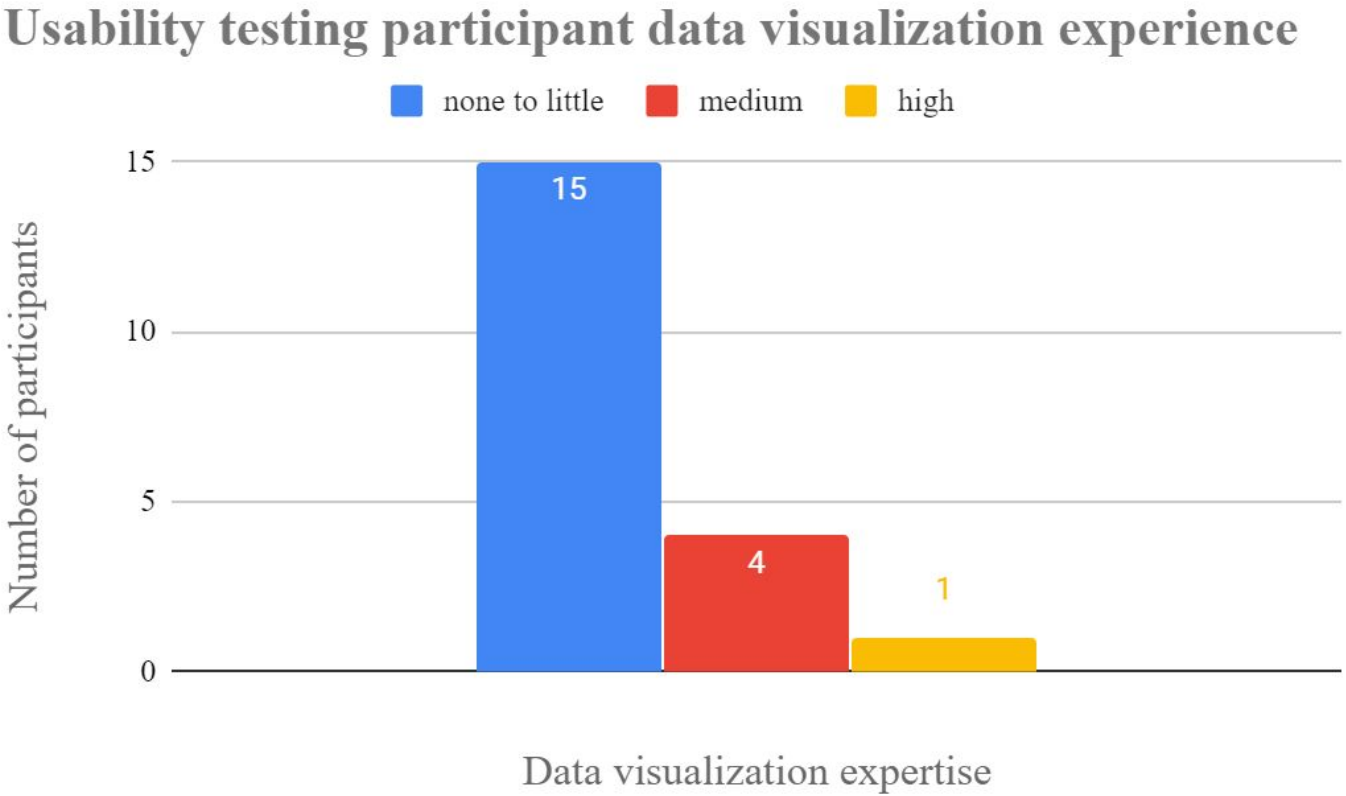
5. Participant industry experience

(N = 20)



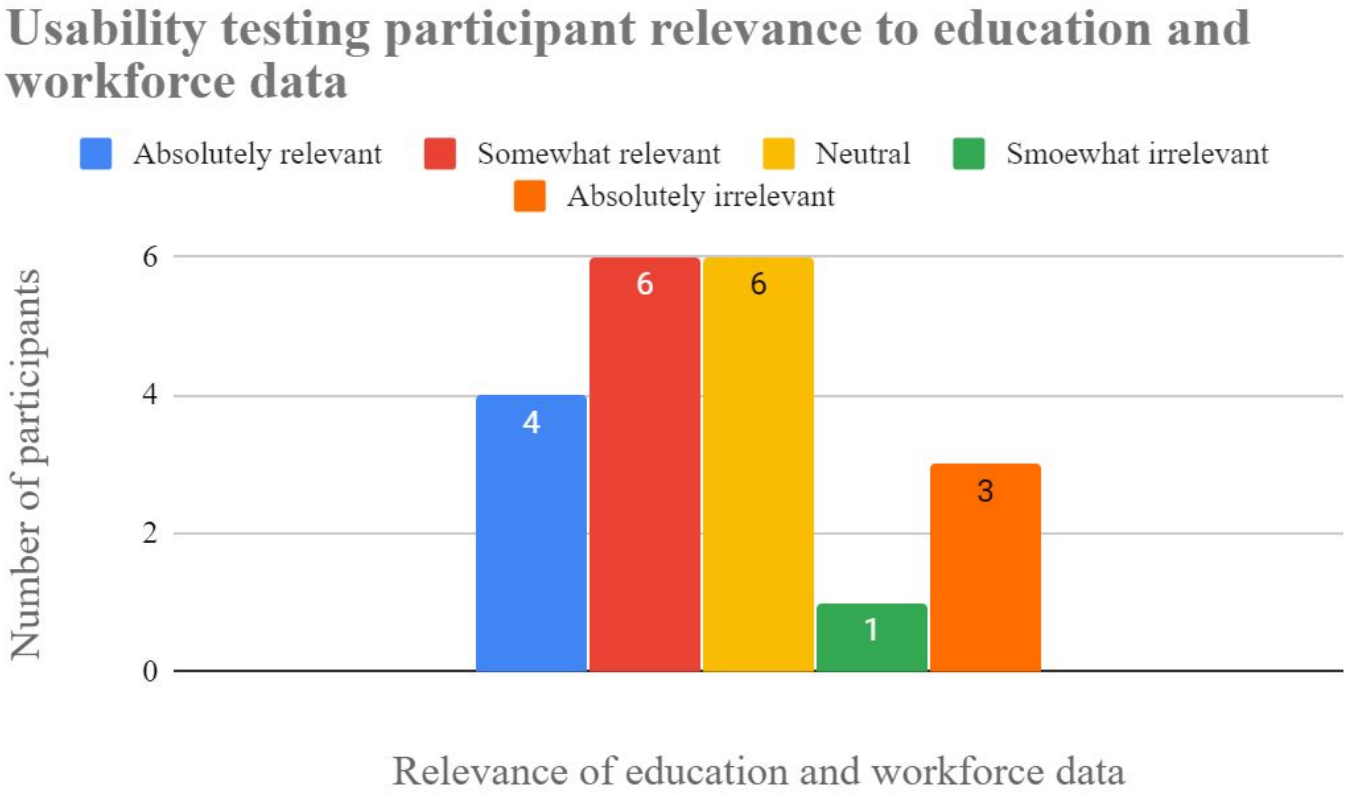
6. Participant data visualization experience

(N = 20)



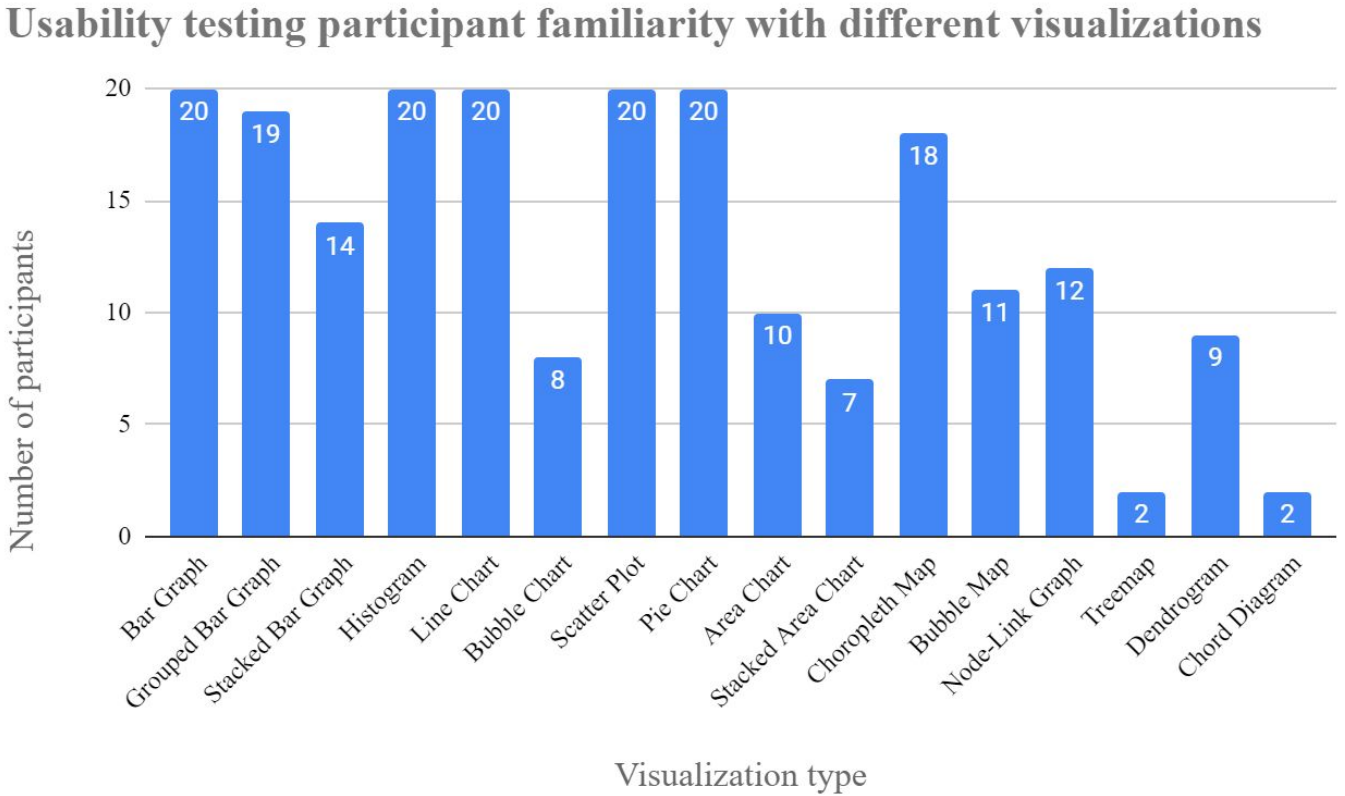
7. Relevance of education and workforce data to participant

(N = 20)



8. Participant familiarity with different visualization types

(N = 20)



2. Analysis results:

A. Performance data:

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

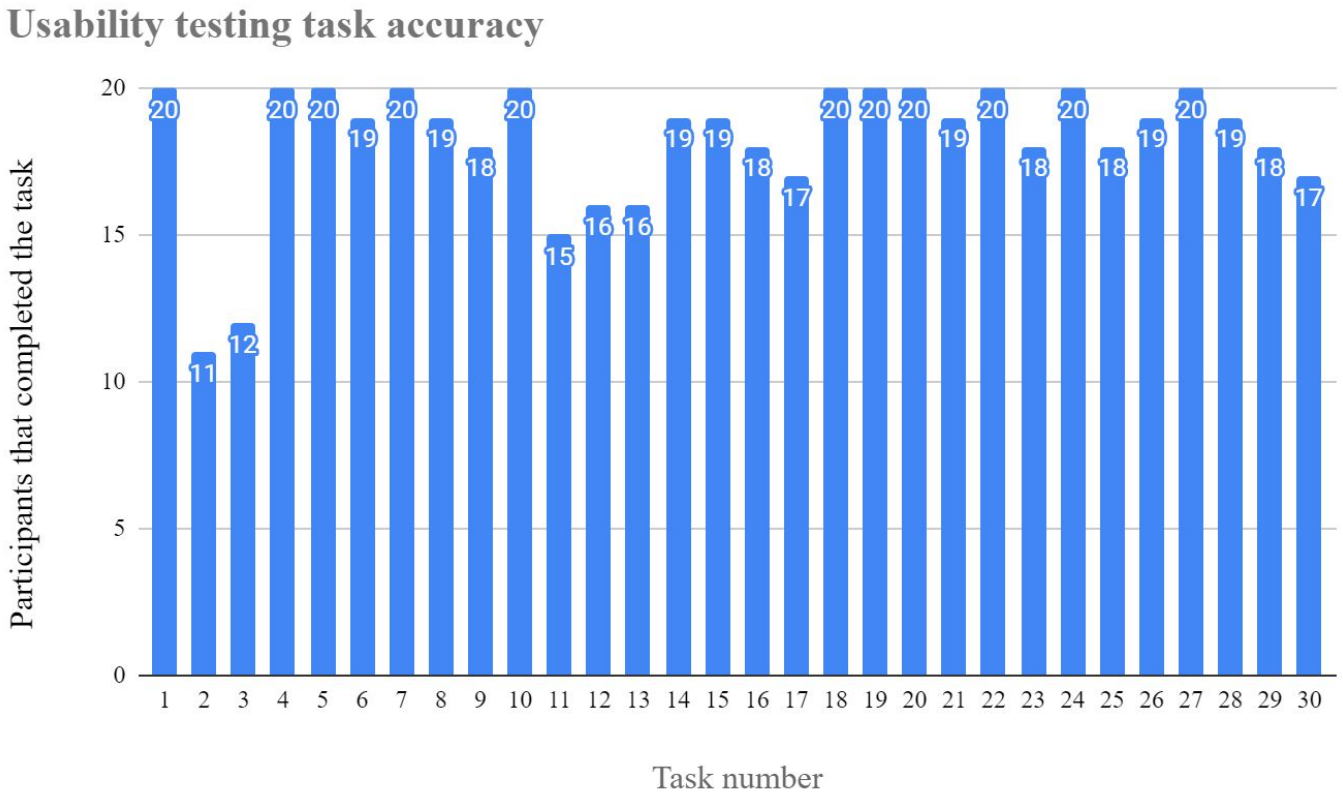
1. Performance of participants on each individual task:

a. Number of participants that performed the task successfully:
   - Participants had the most difficulty in performing tasks 2 (only 11 completed successfully) and 3 (only 12 completed successfully)

a. Percentage of the participants that completed the task successfully (N=20):
   - The task accuracy benchmark was 70%
   - For tasks 2 and 3, the task accuracy percentage is below 70%

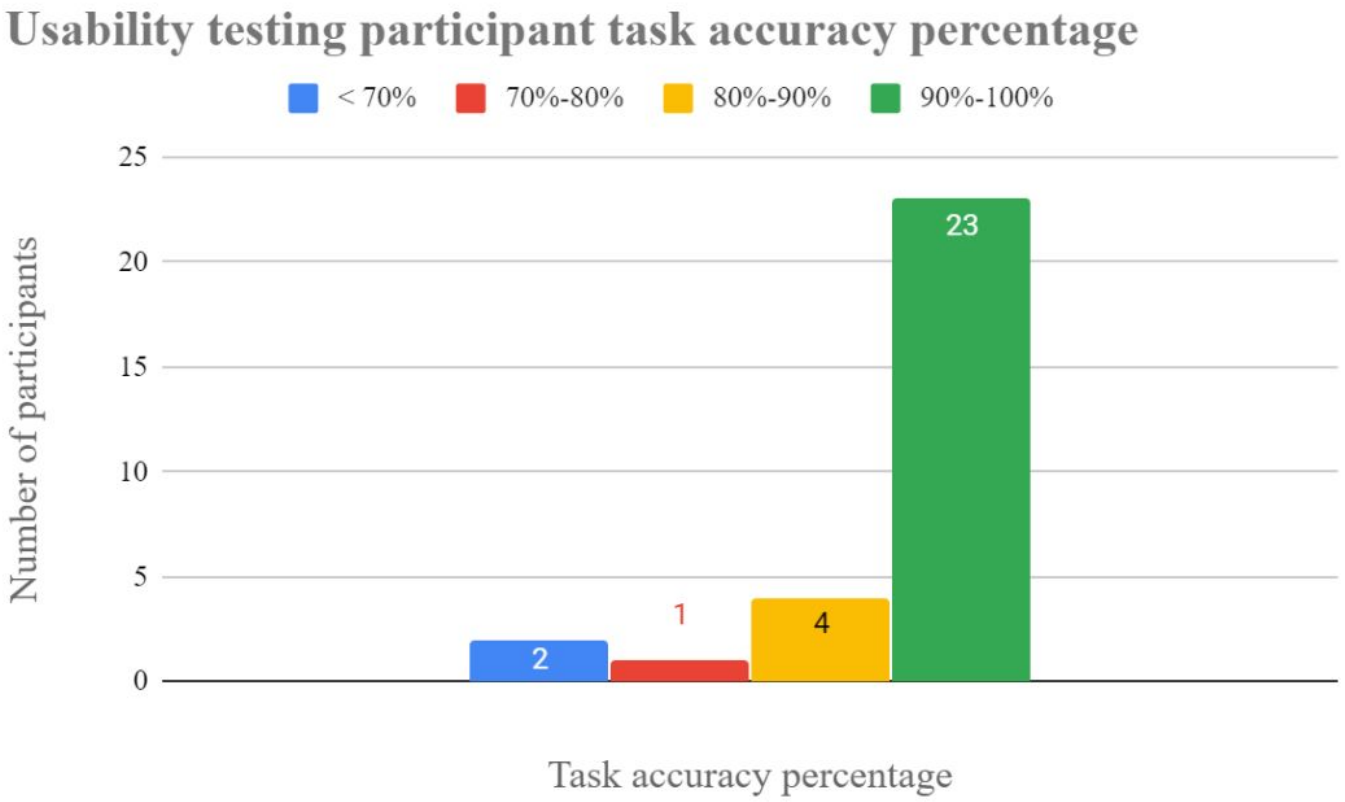
a. Number of participants that performed the task successfully:

(N = 20)



b. Percentage of the participants that completed the task successfully:

(N = 20)



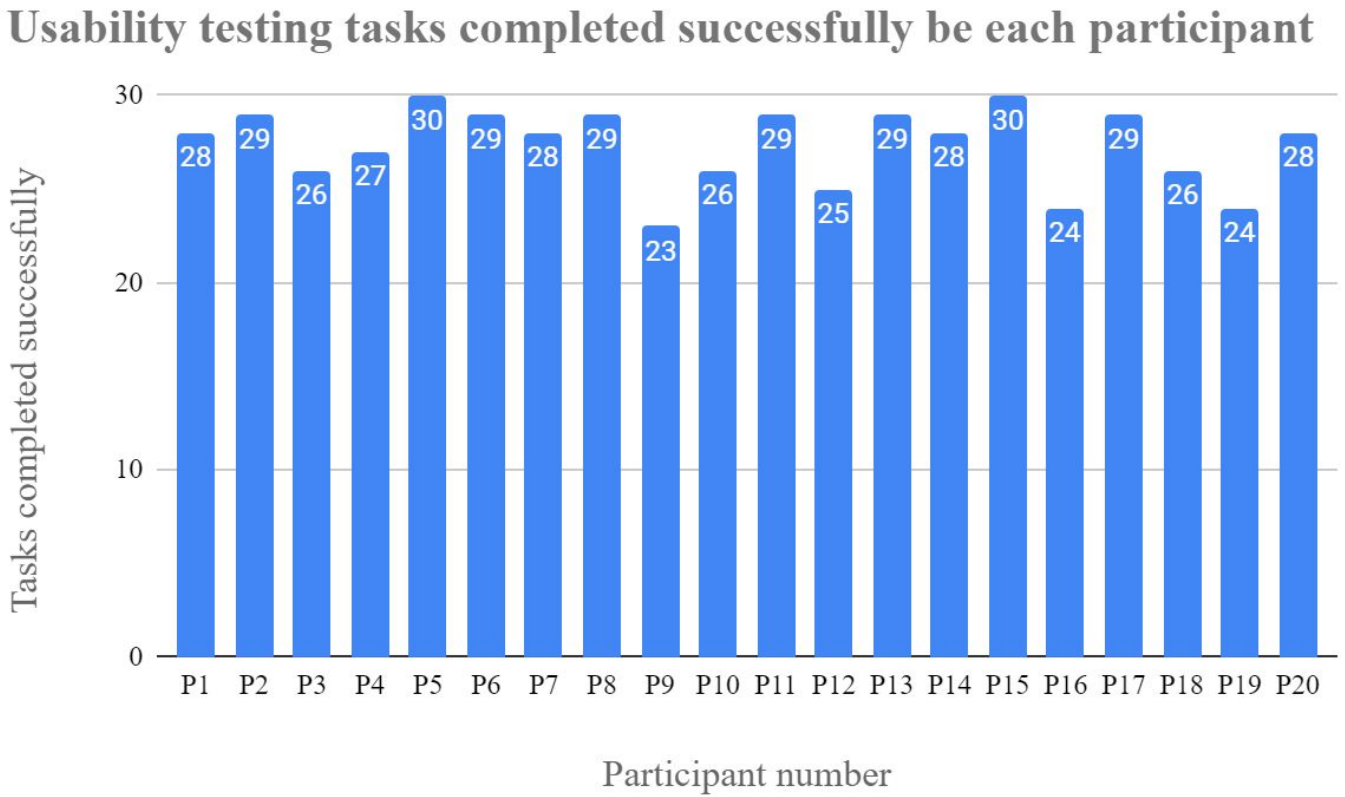
2. Performance of participant on all 30 tasks (N=20):

a. Number of tasks completed successfully by each participant:
   - Minimum is 23 and maximum is 30.

a. Percentage of tasks (N=30) completed successfully by each participant:
   - The task accuracy benchmark was 70%
   - All participants completed more than 70% of the tasks
   - Mean: 91.16%
   - Median: 93.33%
   - Mode: 96.66%
   - Range: 23.34%

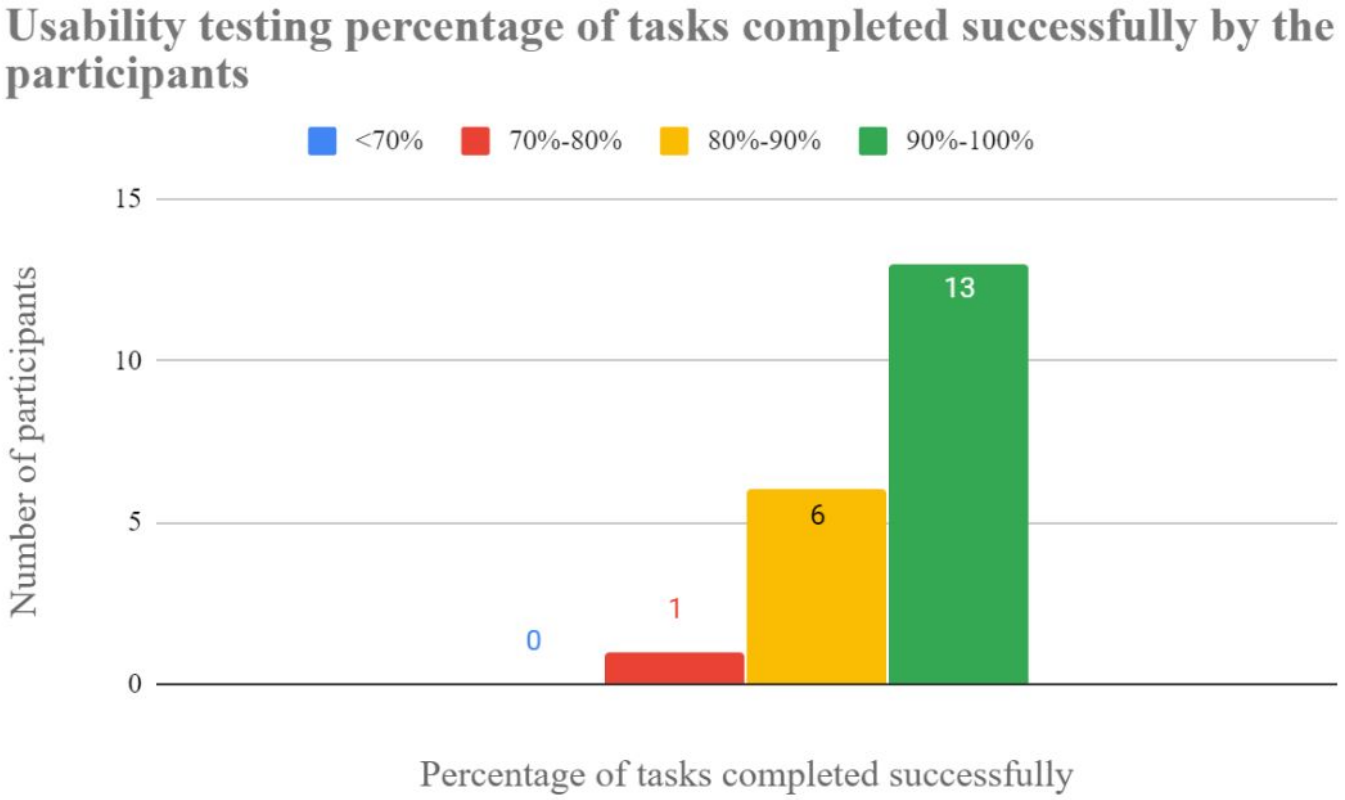
a. Number of tasks completed successfully by each participant:

(N = 20)



b. Percentage of tasks completed successfully by each participant:

(N = 20)



In general:

- Participants had most difficulty in performing tasks 2 and 3.
- For all other tasks, most participants (>70%) were able to complete them (N = 20) .
- All participants completed more than 70% of the tasks (more than 23/30)
- Thus, the dashboard had a good usability overall.

B. 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.

1. Free-form comments and questions:

a. Features that users liked:

- “Interactivity is great”, “Data is up to date”, “*How to* video helps since there is information overload.”
- “Uniform design and everything is visually pleasing”, “Use experience is good and intuitive”, “Graphs are explicit and show information clearly”, “Typeface is nice”

b. Features that users disliked:

- “Drop down not very intuitive”, “Datasets - word is a little confusing as to what exactly is being referred to”, “92 county bar graph x-axis labels are a little difficult to read”
- “Close bubbles on the scatterplot are hard to read. How to select the bottom one? Thus, zooming or separating them will help.”

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

- 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).
- The popup tooltip for the Impact map should stay there for a while after hovering out.
- Use of the keyboard to select multiple counties (ctrl and shift key press) and show data for them.
- Report generation in landscape orientation.
- Option to input value for range along with sliders (for bubble chart).
- Highlight the About, Contact and Help tabs, since grey means that you cannot choose it in different software.

Other analysis:

a. Tasks that did not meet success criteria:

- Task 2: How many datasets are currently being visualized through this dashboard?
- Task 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?

b. Conduct a “Source of error analysis”:

- 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.

- 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:
    1. 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.
    2. On hovering out of the name text, the tooltip disappears immediately. Increase the duration for which the tooltip stays for a little longer after the user hovers out
    3. Place this information somewhere else or find another way to present this information.
    4. 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.

- 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.
- 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.

(Note: The tasks are listed in the thesis document and not mentioned here to avoid visual clutter)

c. Prioritize problems:

a. Problem severity ranking:

1. Task 3 needs to be addressed first since users were not able to use the tooltip to find the information.
2. 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.
3. 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 probably due to poorly designed testing procedures. Hence, they are least severe.

b. Frequency of occurrence ranking:

1. Task 2: 9 users failed (low priority) - Count datasets on nav panel
2. Task 3: 8 users failed (very high priority) - Use tooltip to retrieve info
3. Task 6: 1 user failed (high priority) - Compare data using region bar graph
4. Task 11: 5 users failed (low priority) - Compare data using state bar graph
5. Task 16: 2 users failed (high priority) - Contact tab
6. Task 23: 2 users failed (high priority) - About tab
7. Task 29: 2 users failed (high priority) - Help tab with video explanation

c. Other interesting observations:

- For  icon use , the users instead using the keyboard UP arrow button
- For attendance data, users used the tooltip as well as the table to look up the information
- Ctrl + F keys to search for the county name
- One user was unable to notice the nav pane on the left side of the dashboard (Account for screen size)
- It was clear that the map is clickable since Lake county was highlighted in red.

2. Semi-structured interview Results

Semi-structured interview results

1. Demographic information:

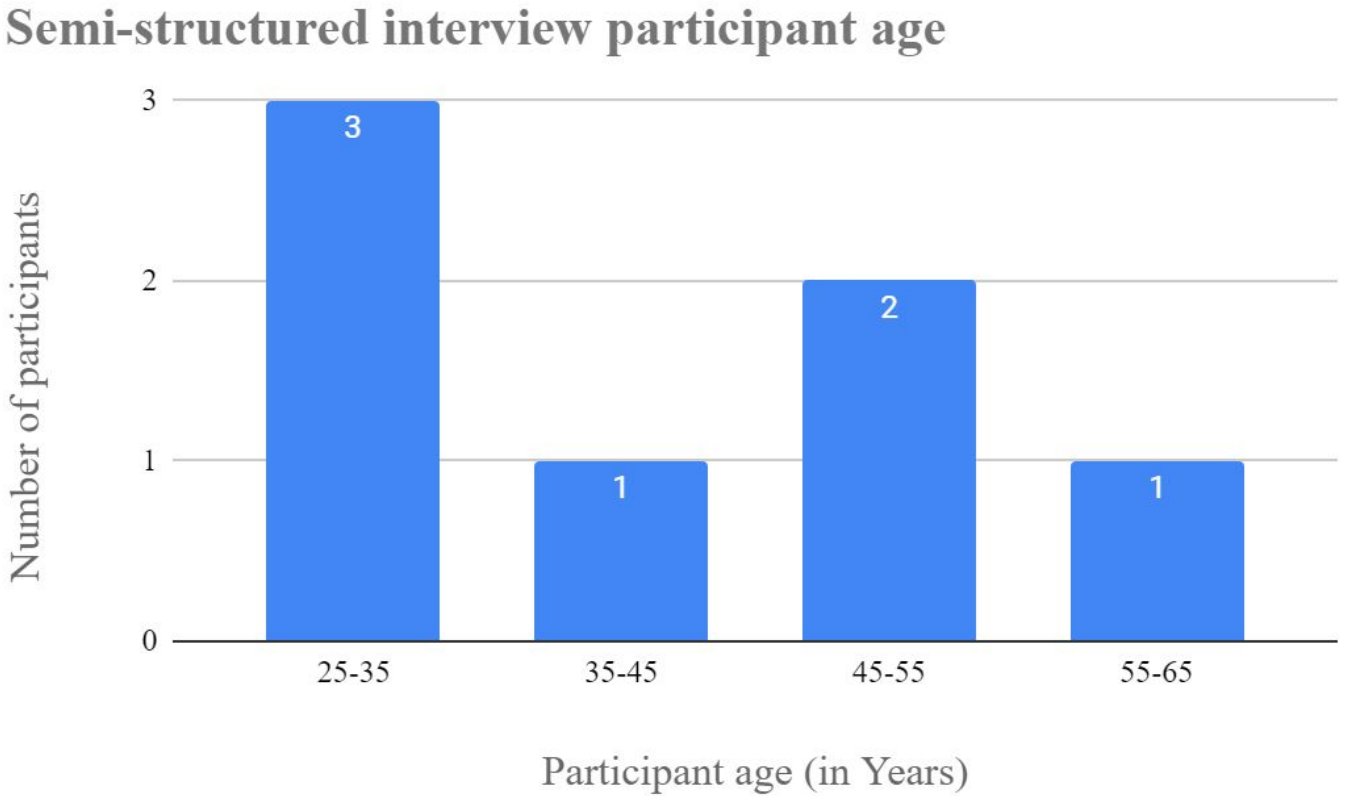
- These statistics give us a general understanding of the demographic of the actual end users of the data dashboard to further understand the results of our interviews.
- For all these participants, the ‘Education and workforce data’ was **Absolutely relevant** as is expected.
- The participants are from a variety of majors such as engineering, management, public relations, etc.

How relevant is education and workforce data to you?

□ *Absolutely relevant*
□ *Somewhat relevant*
□ *Neutral*
□ *Somewhat irrelevant*
□ *Absolutely irrelevant*

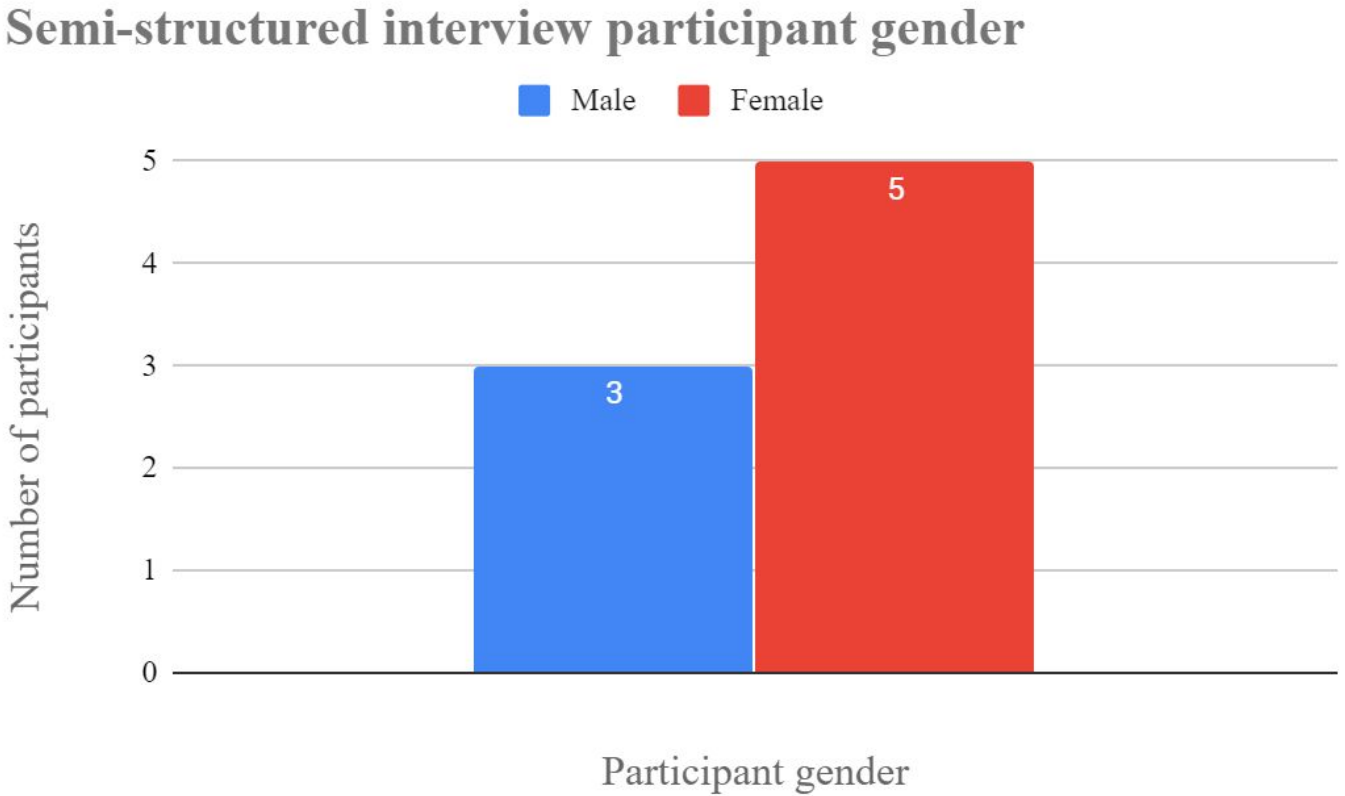
1. Participant age

(N = 8)



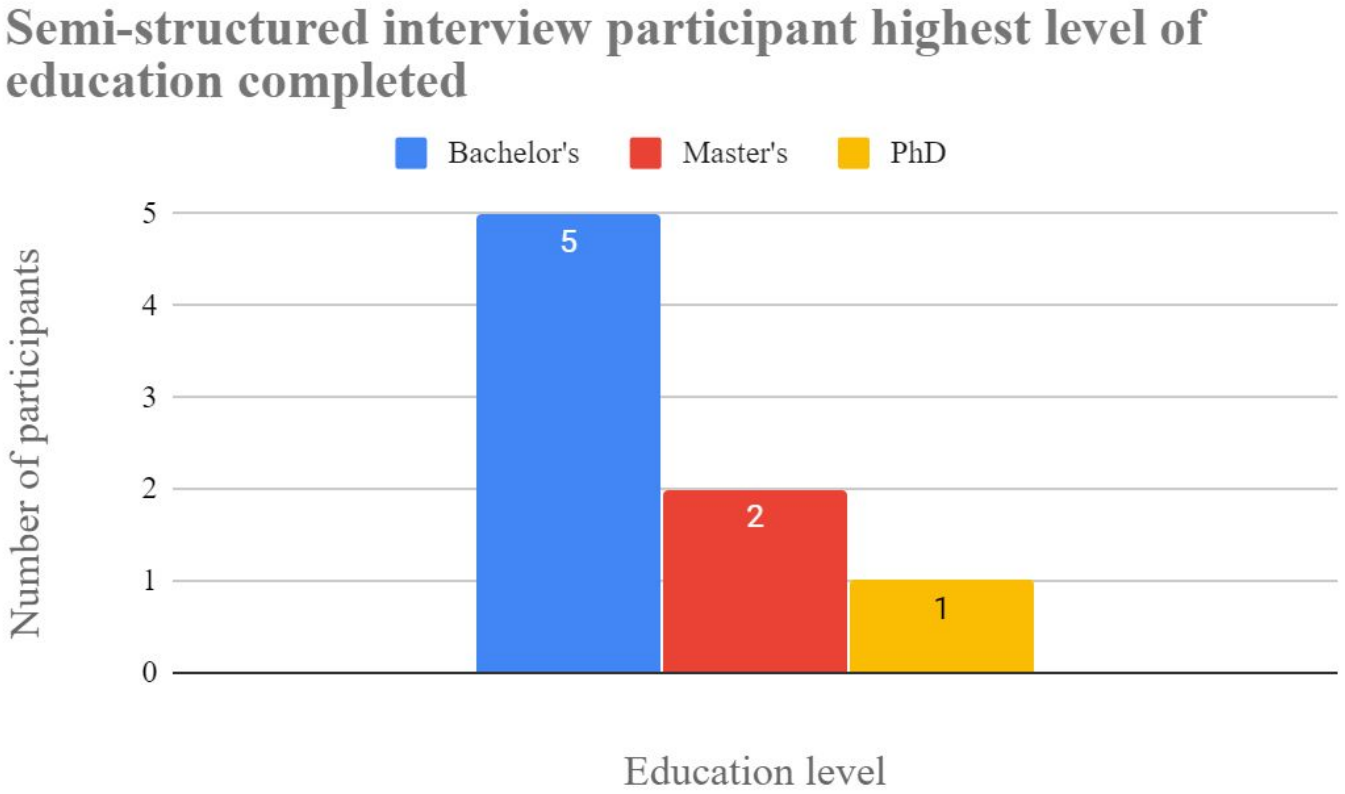
2. Participant gender

(N = 8)



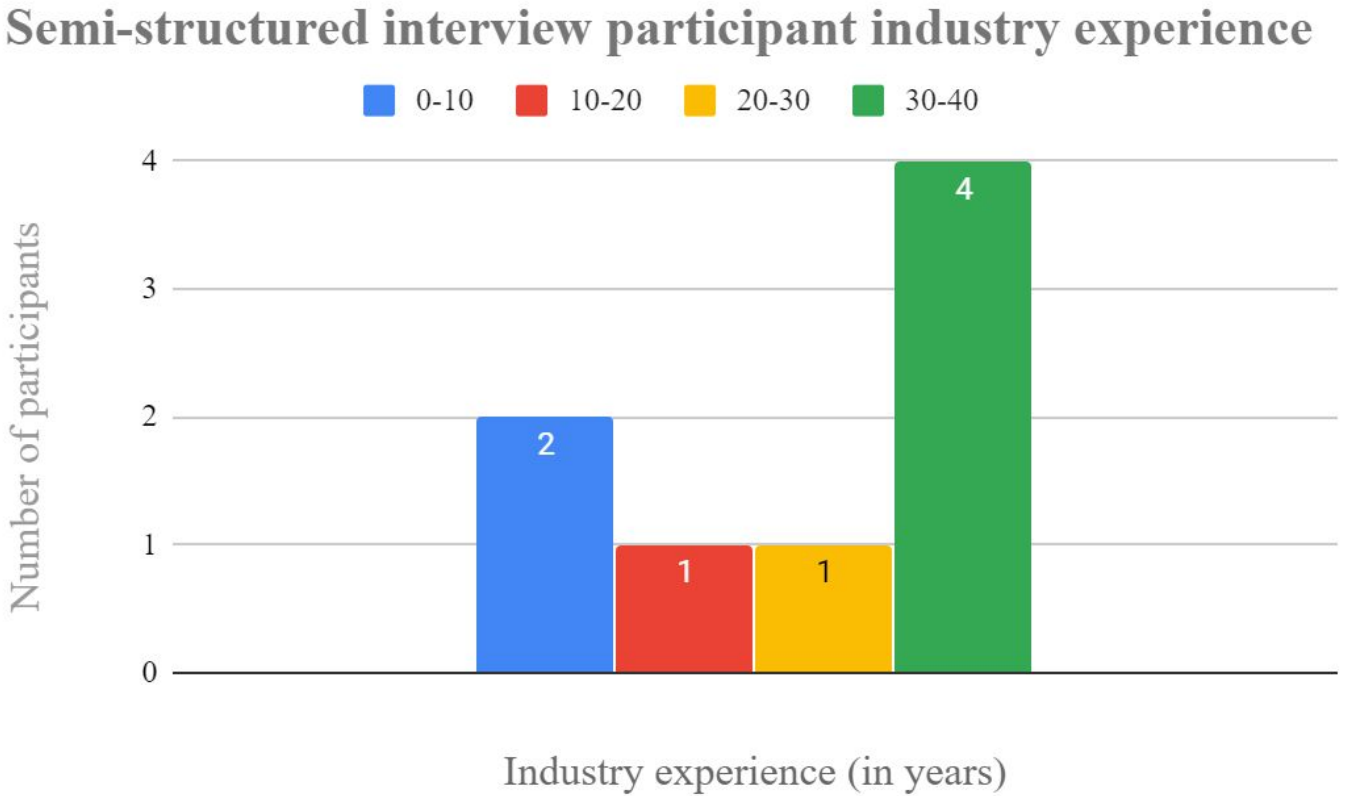
3. Participant education level

(N = 8)



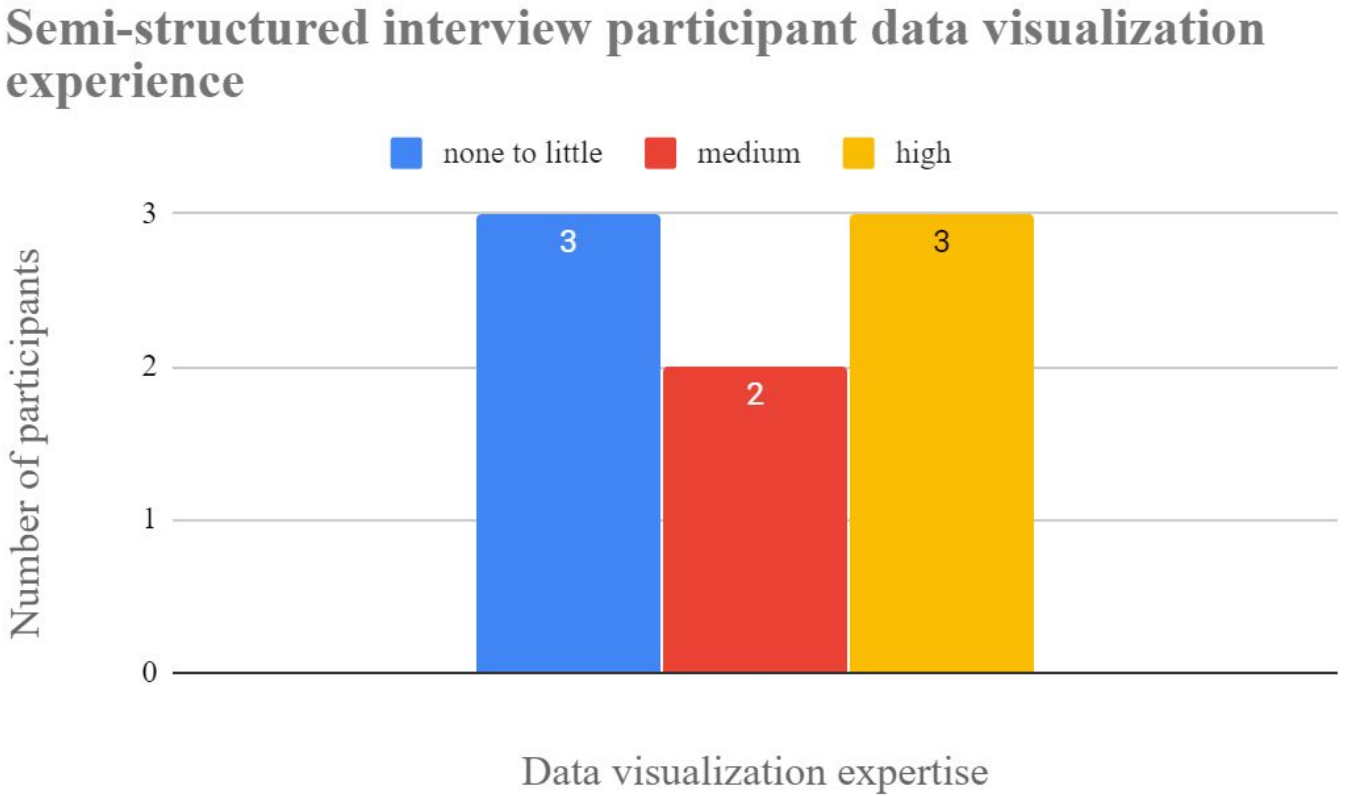
4. Participant industry experience

(N = 8)



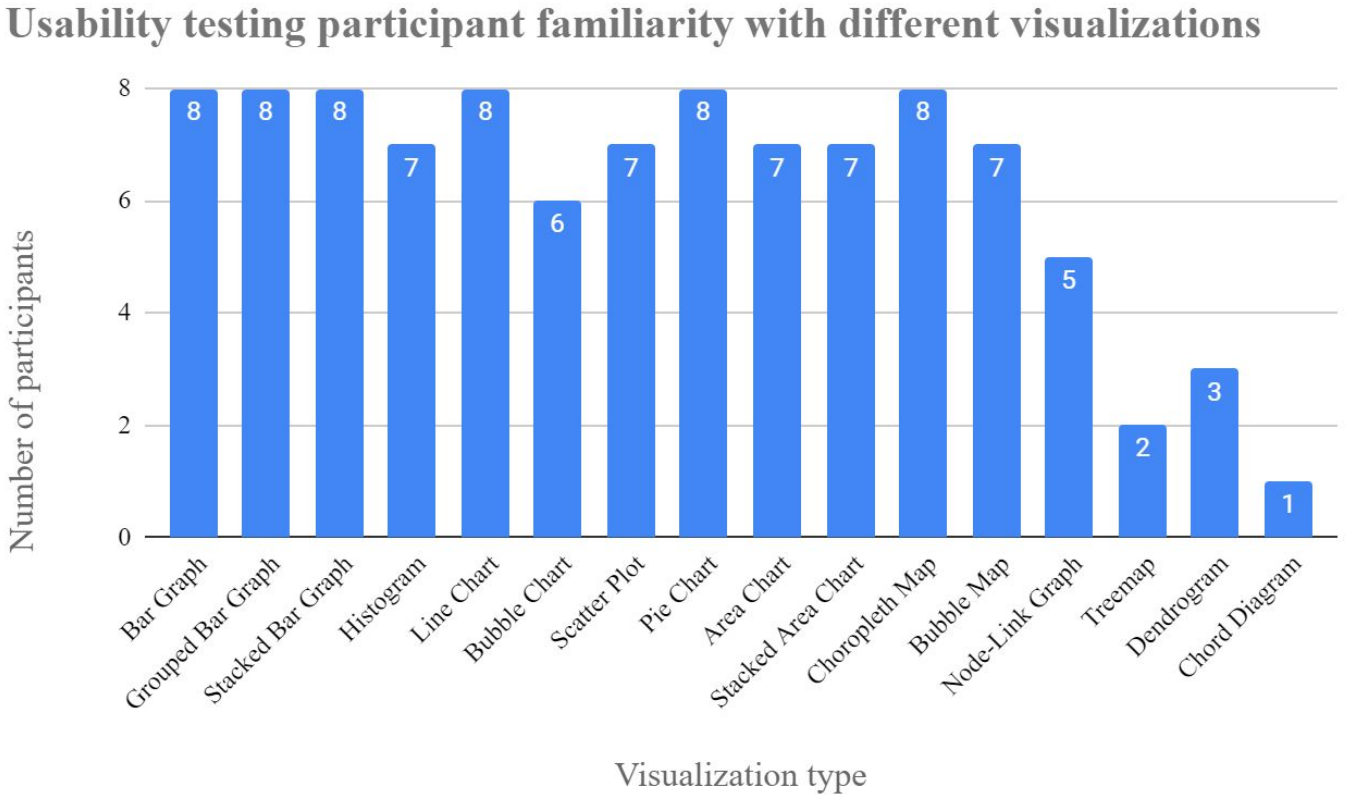
5. Participant data visualization experience

(N = 8)



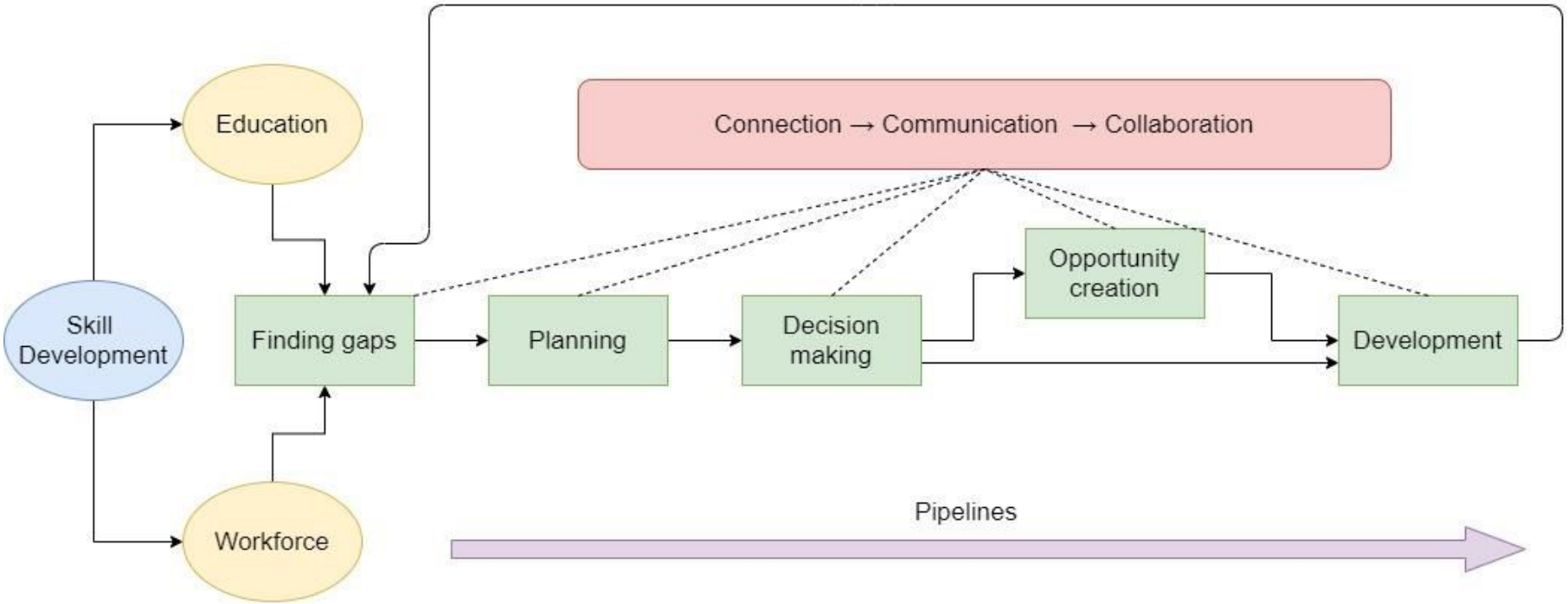
6. Participant familiarity with different visualization types

(N = 8)



2. Analysis results:

A. Work processes of the users:



Multi-purpose 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.
- Current educators are also trained in the emerging technologies by partnering with various institutes.
- 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.
- 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 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.

4. Pipelines:

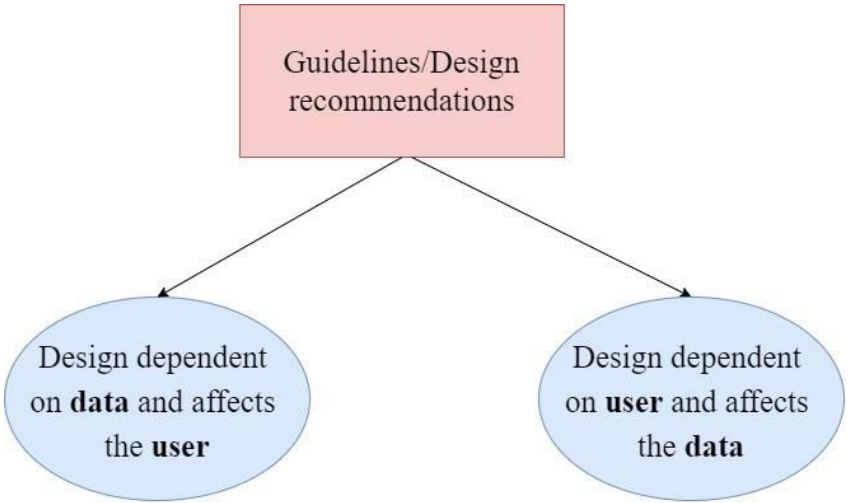
- The entire ecosystem applies to various pipelines such as education pipeline, workforce pipeline, talent pipeline, manufacturing pipeline, school to job pipeline, etc.
- The pipeline is essentially the entire process from starting a desired goal to fulfilling that goal.

Review research question 1

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

B. Answering research question 1:

- The first research question deals with finding the guidelines/visualization design recommendations to be followed while creating a data dashboard for visualizing education and workforce data.
- After several rounds of thematic coding of the interviews, two main themes emerged:



Themes that address RQ 1

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.)
b. Maintain consistency of view in the dashboard pages and visualizations across all the datasets.
c. Provide users with the ability to generate reports from the data.
d. Design the data and visualization interfaces to support additional datasets that may be required by the user.
e. Use desktop web applications for detailed analysis of data and mobile for overview or quick snapshots
f. Ensure stability, maintainability and ease of use.
g. Design for users with varying work roles (from management, planning, technical, etc.)

RQ 1

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.
b. Ensure clarity is conveyed:
   1. Through the visualizations
   2. About the datasets.
c. Allow the user ability to control the granularity of data that he is interested in looking at.
d. Ensure data is valid (avoid old data), reliable (avoid missing data and broken fields) and updated frequently (quarterly to yearly)
e. Provide data source and last update information.
f. Use area breakdown (by county/region) as a starting point to access the data.

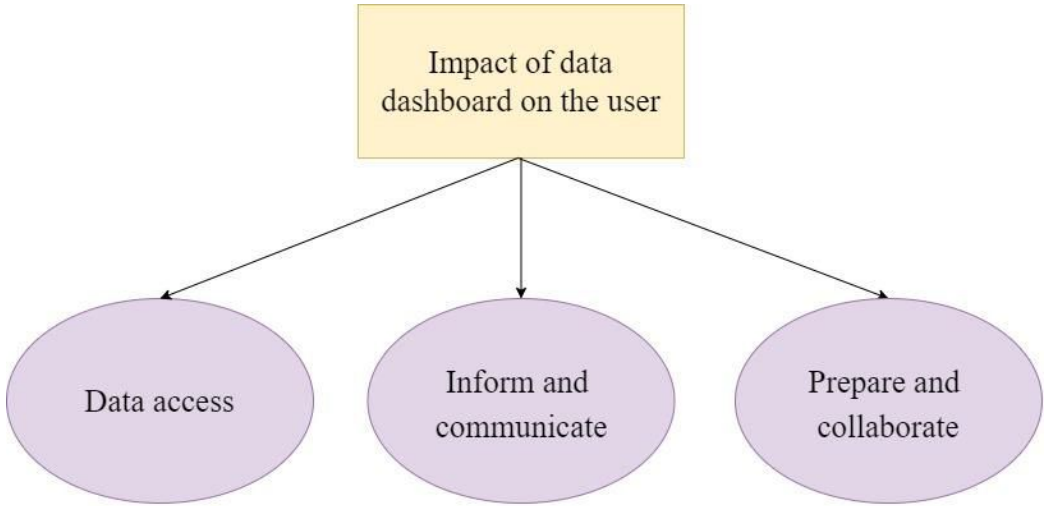
RQ 1

Review research question 2

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?

C. Answering research question 2:

- The second research question deals with finding the impact that an information visualization dashboard has in the work processes of the users involved in the education and workforce development for the state of Indiana.
- After several rounds of thematic coding of the interviews, 3 main themes emerged:



Themes that address RQ 2

1. Data access:

RQ 2

a. Dashboard provides centralized access to data.
b. Dashboard provides easy access to various datasets through common interface.
c. Dashboard is frequently accessed to obtain relevant information (daily to a minimum of weekly access).
d. Dashboard helps avoid dealing with raw (dry) data directly.

2. Inform and communicate:

a. Dashboard helps to find information to connect, communicate and collaborate with user's clients and partners.
b. Dashboard helps to find gaps, plan work and create more opportunities.
c. Dashboard helps to generate awareness.

3. Prepare and collaborate:

RQ 2

a. Dashboard is used to support or back conversations with facts.
b. Dashboard is used to prepare for meetings and presenting data.
c. Dashboard allows to maintain a consistent view about the data between all users to avoid confusion and inconsistencies.

CONCLUSION

Summary:

- In this research, a data visualization dashboard was developed for education and workforce development for the state of Indiana.
- The dashboard usability was very high since 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.
- Thematic analysis on the interview data further led to the generation of design principles/guidelines that should be used to design such data dashboards (addressing RQ 1). This analysis also helped to understand the impact of the dashboard in the various work processes of the users (addressing RQ 2).
- 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.

Future work:

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

1. Improve the dashboard 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 (more than 10).
3. Test the applicability of the guidelines for developing data dashboards for other states and countries.
4. Generate several versions of the designs and performing comparisons between them through controlled quantitative studies.

THANK YOU

Q and A