Report and Suggestions from IPEDS Technical Review Panel #58: Exploring Best Practices in Data Visualization

SUMMARY: The Technical Review Panel discussed ways in which other organizations effectively use data visualization in order to inform NCES as it considers its approach to implementing and utilizing data visualization tools with IPEDS data. This summary provides an overview of the information presented to the panel and the primary themes and takeaways that emerged during the associated discussion. Comments from interested parties are due to Janice Kelly-Reid, IPEDS Project Director at RTI International, at ipedsTRPcomment@rti.org by June 10, 2019.

On March 26 and 27, 2019, RTI International, the contractor for the Integrated Postsecondary Education Data System (IPEDS) web-based data collection system, convened a meeting of the IPEDS Technical Review Panel (TRP) in Washington, DC. IPEDS TRP meetings are conducted by RTI to solicit expert discussion and suggestions on a broad range of issues related to postsecondary education and the IPEDS collection. IPEDS Technical Review Panel meetings are conducted by RTI to solicit expert discussion and suggestions on a broad range of issues related to postsecondary education and the IPEDS collection. As the postsecondary education industry evolves, IPEDS TRP meetings are increasingly critical in addressing changes to ensure that IPEDS data remain relevant, informative, and on the forefront of industry advancements and legislative needs. To this end, IPEDS TRP meetings are designed to foster public discourse and enhance IPEDS data collection, products, data quality, and system user-friendliness. The TRP does not report to or advise the U.S. Department of Education.

RTI's specific purpose for TRP 58 was to learn from other organizations (e.g., federal agencies, postsecondary institutions, research institutes, and other non-education groups) that effectively use data visualization tools with their data. Through this TRP, NCES aims to gain insight from others who have quickly delivered non-technical information to a larger audience, as well as increase the public's usage of and access to its data. The panel consisted of 47 individuals representing institutions, researchers, state governments, the federal government, higher education associations, and other experts.

Background

IPEDS collects and makes publicly available a vast amount of data on postsecondary institutions and their students to a variety of stakeholders. This diverse community represents a broad spectrum ranging from non-technical (e.g., students, parents, media) to technical (e.g., data providers, researchers, policy analysts) data consumers. Some IPEDS data are disseminated online and in somewhat dynamic environments, such as the IPEDS Data Center and the IPEDS Trend Generator, but the majority of IPEDS data are distributed in the form of relatively static tables and datasets. While these data are available for users to download and perform further analysis, doing so requires investment of user resources, can be cumbersome, and can often add to the time it takes to ultimately arrive at meaningful analysis of IPEDS data.

Using data visualization techniques to distribute IPEDS data in new ways could offer numerous advantages such as enhancing accessibility of the data, making it available to wider audiences,

facilitating new types of analyses of IPEDS data, reducing the time it takes to release data, and increasing public usage of the data.

This TRP meeting was organized into four series of short presentations delivered by representatives from various organizations including research institutes, media outlets, postsecondary institutions, higher education associations, and federal agencies. Each presenter shared his or her organization's experiences developing, implementing, and maintaining data visualization products. Following each series of presentations, the panel engaged in a moderated discussion as well as collaborative brainstorming.

Overview of Presentations

Session One. The first presentation session included representatives from a higher education media outlet as well as a large research organization that conducts a wide variety of research in many different areas.

The Data Editor and the News Applications Developer representing the media outlet provided an overview of the scope of their work developing data tools and visualizations using IPEDS data in close collaboration with the organization's reporters and news staff. These data visualizations are designed to enhance reader engagement and also to provide tools that can assist staff in their development of news and editorial content. The presenters described a combination of manual and automated processes used to acquire, store, and prepare IPEDS data for integration into their tools, followed by an outline of the conceptual process of moving from static data to visualizations, an evolution that progresses from asking questions, defining scope and parameters, and finally creating and testing the visualizations.

A Data Scientist from the research organization then delivered a presentation detailing their work with a city police department on the development and implementation of a data visualization and analytic tool that enables real-time decision-making using 911 emergency call data. The research organization brought data science expertise while the police department provided data and law enforcement insight. The two groups collaborated closely to design a product that was user-oriented but also comprehensive enough to offer certain specific data and functionality critical to effective police department operations. Furthermore, the application was designed to offer flexibility and the ability to quickly and efficiently respond to change.

Session Two. Presenters in the next group included representatives from three postsecondary institutions of varying size, level, and control, including a private nonprofit institution, a large public 4-year institution, and a small community college.

The Associate Vice Provost for Institutional Research and Planning representing the private nonprofit institution described the process of transitioning her institution's university factbook from a printed resource to an interactive data visualization tool. Prior to the transition, large amounts of data were presented in table format, which made it difficult to extract useful information. The printed factbook could not provide all possible combinations of information, and examining cross-cutting concepts that spanned multiple sections of the data was especially challenging. Providing data in a dynamic, visual tool has enabled more useful analyses by allowing users to more easily extract and display relevant information and make important connections across the dataset that were not previously possible.

Next, the Director of Institutional Research at the public 4-year institution presented an overview of the ways in which his institution uses data visualization, work which involves numerous departments and functional groups at the institution, including institutional research, academic analytics, business intelligence, and information technology. The resulting data visualizations are used for a variety of internal and public facing data products. In addition to dynamic dashboards that use IPEDS data to describe its educational mission and illustrate comparisons with peer institutions, the institution also uses data visualization extensively to present academic program review data to internal stakeholders as well as to the public. This information was formerly contained in large amounts of tabular data but has since been transitioned to a dynamic set of data visualization tools that has increased accessibility and understandability of the data.

The Vice President for Strategic Initiatives and Assessment at the small community college then described her institution's successful implementation of data visualization tools despite limited resources, a poor data infrastructure, and an institutional culture that did not value the use of data as a tool for decision making. By transitioning from tabular data to more visually appealing offerings while maintaining the simplicity and accessibility required by the institution's users and data culture, the institution has improved the way data are communicated both internally and externally. This has resulted in more data discussion and use; better transparency and a more informed community; and stronger partnerships within the institution. Key elements to successful implementation included a deep understanding of the audiences and how they interpret data along with providing comprehensive training using a teaching approach.

Session Three. The third collection of presentations included individuals representing a higher education association, two state longitudinal data systems (SLDS'), and a private nonprofit organization that focuses on improving data and decision-making for states and institutions.

The Director of Research and Data Policy for the higher education association first presented an overview of a subscriber-based data tool designed to support postsecondary education leaders' use of data for strategic planning and decision making. This tool combines multiple national data sources into an institution-level data set for analysis and reporting. The comprehensive underlying data, which span nearly a decade and include hundreds of variables, are downloadable along with a user guide and codebook, but making use of this raw data is difficult and requires a steep learning curve. To make the information more accessible, the organization utilizes data visualization tools to deliver a variety of reports that include data, features, and functionality that are most relevant to higher education leadership.

Next, the Data Communications Coordinator for an organization that houses its state's longitudinal data system described leveraging data visualization to link data and facilitate research across sectors (e.g., examining how former early learning participants fare in kindergarten and beyond) with the goal of informing policymaking. Because the primary audiences of these data are the general public and policymakers, developing tools that are accessible to these non-researcher groups was critical. To help address this priority, the organization hired two new staff members, one with specialized skills in data visualization and another with a background in teaching and instructional design. Approaching data visualization with this intentionally mixed focus enabled the development of tools that provided the necessary functionality along with user-friendly design in order to make the data accessible by the nontechnical audiences.

A Senior Research Analyst representing a second SLDS with similar goals presented an overview of his organization's efforts to move from static reports to more dynamic and interactive tools that can more effectively address previously unanswerable questions. In order to ensure that a focused research agenda drives the development of the proper data tools, stakeholders from all levels of the state's education spectrum are given representation on the organization's board. This philosophical foundation helped guide the transition from comprehensive collections of PDF reports to more dynamic data visualization tools that enabled users to gain access to the specific data they needed in a timelier manner.

Finally, a Research Associate from a private nonprofit organization that focuses on improving state- and system-level decision making through data described her group's work using data visualization to help clients build their own data-driven decision-making capacities. By not only facilitating access to data, but also providing analytic support through data visualization, the organization is able to assist clients who often have limited resources to most efficiently utilize their existing data to address their research needs.

Session Four. The final group of presentations included representatives from a variety of different federal agencies offering their experience using data visualization in their respective organizations.

The first presenter provided an overview of how his agency has used data visualization to illuminate data related to the earnings, employment, and other outcomes of postsecondary students over the course of time after graduation. By linking their own employment data with additional data from postsecondary institutions, the agency aims to evaluate how students are doing after graduation with regard to earnings. This organization possesses extensive historical experience with data visualization and as a result is able to use mapping and other sophisticated techniques to translate complex raw datasets into tools that users can take advantage of even if they are not experts in data analysis and manipulation.

The next presenter—representing the communications department of another federal agency—described how her organization conducts outreach work in order to shape the development of their applications. This outreach includes using surveys, customer analytics, and other engagement tools to inform the development of user personas and shape the nature of their work. The resulting data visualization products have been extremely successful and have helped earn the organization high levels of customer satisfaction.

The last presentation featured a Management Analyst within a federal agency's policy office whose work on open data infrastructure initiatives focuses on making federal data more accessible, discoverable, and useful. The presenter provided an overview of the philosophical guidelines that shape the development of data visualizations. These concepts include building at the intersection of design and technology, which helps to not only make the resulting products interesting, but to also feel intuitive to the user; drawing inspiration from other successful data visualization work in order to fuel creativity; and keeping the user at the center of one's approach.

Overarching Themes and Takeaways

Throughout the course of the presentations and subsequent discussions, a number of broad themes, patterns, and best practices emerged, which can serve as a set of guidelines and suggestions for NCES as they explore next steps related to data visualization.

Maintaining and sustaining over time is critical. Panelists reinforced throughout the meeting the importance of advance planning as it relates to the maintenance of data visualization tools.

Panelists noted that once a data visualization is developed and made available to users, it can sometimes create an automatic expectation that the product will be maintained in perpetuity, especially if it turns out to be effective and useful (but sometimes even if not). Furthermore, once a visualization exists, it can be difficult to discontinue—even if there is a legitimate reason to do so—because of the demand and expectation that have already been created as a result of the tool.

Panelists agreed that maintenance often accounts for a significant portion of the investment in data visualization efforts. Problems can arise if this maintenance requires an inordinate amount of effort and resources that ends up outweighing the advantages of the tool. Assembling a team with diverse skills and specializations and building with potential maintenance in mind from the start can serve to maximize efficiency and reduce maintenance requirements later. Panelists also emphasized that implementing automatic functionality can help mitigate potential issues of unsustainability and suggested automating processes and procedures as much as possible when designing products.

Panelists noted that thorough documentation throughout the life cycle of development and implementation can help maximize maintenance efficiencies.

Communication is key. Because one of the ultimate goals of data visualization is to provide useful, accessible information to one's audiences, panelists emphasized the need for constant communication with those audiences and other stakeholders throughout the process in order to ensure that the resulting products meet their needs.

Panelists noted out that there can sometimes be multiple groups of audiences for a particular product, each of whom may have different data visualization needs that require varying levels of effort to develop, so clear communication throughout the life cycle of the process is important.

In addition to recurring conversations with audiences and stakeholders, collaborative communication among team members is also critical during the development process. Panelists pointed out that communication styles can vary widely among individuals, so assembling a team with this in mind can help facilitate effective collaboration.

Begin with informed and intentional research questions. Thinking in advance about the specific research questions can serve to guide the development of successful data visualization products. Panelists pointed out that while this forethought is necessary, one research question can often lead to subsequent, deeper questions; so continuous collaboration with key constituents is critical. Ensuring that these conversations occur during the development phase and continue through implementation can lead to improvement of the product and its effectiveness over time.

Panelists agreed that starting with intentional research questions can help inform the creation of effective data visualizations. They also pointed out that facilitating discovery and allowing users to explore the data on their own are equally important capabilities that a statistical resource can offer.

Purposeful storytelling is important. Panelists noted the importance of making conscious decisions up front about how data visualization is intended to be used, including choosing whether one is attempting to tell a particular story or simply presenting data for users' exploration. Determining the audience is another example that panelists provided of a decision that should be considered early in the process of developing a data visualization.

Panelists acknowledged that presenting interesting and compelling data while avoiding bias and perceptions of specific agendas can be a delicate balance and another argument for thorough forethought and planning.

Maintain neutrality in the presentation of data. Data visualization can be an effective tool for presenting information in such a way that consumers and users can easily access and understand the data. However, because data visualizations often present complex data in a simple manner, there must be a balance between making data understood and introducing bias in the presentation of data. Furthermore, the degree to which data must be presented with neutrality depends largely on the organization developing the visualizations. Panelists noted that a certain amount of bias is sometimes unavoidable, especially when data are used to communicate a particular story, but there are certain measures that can be taken to minimize the introduction of bias.

Panelists agreed that it is important for a statistical agency such as NCES to present data in a manner that is as neutral and as free from bias as possible. Panelists pointed out that presenting a large dataset with a wide scope can serve to minimize the effect bias. This approach would then allow end users to determine how themes are communicated with the specific data and methods chosen for their particular analyses.

Keep it simple. One of the advantages of using data visualization to present information is the ability to communicate complex data in a more user-friendly manner. Panelists noted that because of this potential power inherent with data visualization, it is important to exercise caution and avoid getting too carried away with features and functionality, which could have an effect opposite to one of the primary goals of using data visualization, which is to make data more accessible and understandable to users.

Panelists agreed that simple, clear, and crisp visualizations tend to be the most effective. Panelists further emphasized the importance of keeping the intended audience in mind when making decisions on design, layout, and color choice. Panelists also noted that plain and direct language can serve to make visualizations more understandable and stressed the need to be conscious of the potential perceptions of the chosen language in order to avoid unintended connotation or misinterpretation.

Panelists also pointed out that while data visualization is designed to take advantage of the dynamic and interactive capabilities of technology, many users still prefer print format in certain situations, so designing visualizations with printing in mind is prudent. This could entail either making the entire visualization printable or incorporating functionality that enables printing of a more basic version.

Technology should be the tool, not the driver. Data visualization by nature leverages technology to facilitate new or enhanced access to data. While the particular technology tool one chooses is an important factor in developing data visualizations, panelists agreed that the specific technology itself should not necessarily drive one's approach to developing data visualizations. Instead, the intended audience, the data themselves, and the research questions being investigated should propel and guide the development of data visualization tools, while the role of technology should be to facilitate the interaction between the data and the users thus increasing access to information.

Panelists did point out, however, that more powerful technology can generally enable more complex visualizations and thus answer more specific research questions, so choosing the proper technology is an important consideration. Panelists noted that each technology tool will have its own particular strengths and limitations which should considered when choosing a technology. Furthermore,

panelists again emphasized the importance of thorough documentation, which can help ease the transition between technologies if changing tools is ever necessary.

Prioritize the needs of IPEDS data users. Panelists consistently reinforced the importance of remaining conscious of and focusing on the needs of data users when exploring data visualization and other channels for facilitating access to IPEDS data. Panelists generally agreed that the ability to connect data across IPEDS survey components is an important need of the community and should be prioritized as data visualization is explored. Similarly, the ability to link IPEDS data with other external data sources is another desire that should be considered.

Panelists suggested top priorities for NCES should be to get IPEDS data into the hands of users and to do so in a timely manner. While data visualization is certainly a potential avenue to consider for future data dissemination, it would be also prudent for NCES to investigate other approaches to increasing access to IPEDS data. Panelists noted that providing IPEDS data via an application programming interface (API) would facilitate greater access to data and provide users with unprecedented flexibility to integrate IPEDS data into their own systems, websites, and tools.

Because the community of IPEDS data users is so diverse, panelists emphasized the importance of balancing the needs of a variety of users regardless of the channels used for data dissemination. For example, smaller institutions and other stakeholders with fewer resources at their disposal may require more assistance accessing and making use of IPEDS data and may therefore benefit greatly from the user advantages offered by data visualization techniques. On the other hand, organizations with more staff and a more extensive technology infrastructure may be more equipped to utilize more complex data without the need for assistance through data visualization tools and may instead prefer to have enhanced access to the underlying data to use as they see fit, such as through an API.

Next Steps

Once the TRP summary comment period has closed, RTI will review the comments and outline recommendations for NCES based on the outcome of the TRP meeting and subsequent public comment period. NCES will review the recommendations to determine next steps.

Comments

RTI is committed to improving the quality and usefulness of IPEDS data as well as strategies that might be helpful in minimizing additional reporting burden. We encourage interested parties to send any comments or concerns about this topic to Janice Kelly-Reid, IPEDS Project Director, at ipedsTRPcomment@rti.org by **June 10, 2019.**