

Institute of Educational Statistics  
National Center for Education Statistics

NATIONAL INSTITUTE OF STATISTICAL SCIENCES  
TECHNICAL EXPERT PANEL REPORT

INNOVATIVE GRAPHICS FOR  
NCES ONLINE REPORTS

## TABLE OF CONTENTS

Executive Summary .....	3
Preface.....	5
Innovative Graphics for NCES Online Reports.....	6
I. INTRODUCTION .....	6
Background.....	6
Goals .....	6
Panel Purpose .....	6
Report Format .....	6
II. GOALS FOR A REVAMPED WEB GRAPHICS PLAN .....	7
A Path to Better Data Visualization .....	8
III. KNOWING YOUR AUDIENCE .....	9
Types of Audience .....	9
Audience Needs.....	10
Types of Materials Created.....	10
Ask the Audience .....	11
Other Strategies for Building an Audience .....	12
IV. WORKING WITH CONTRACTORS .....	12
V. CREATING ENGAGING GRAPHICS .....	14
Techniques.....	14
Considerations for Interactive Graphics.....	17
VI. SHOWING UNCERTAINTY .....	17
VII. TECHNOLOGY.....	18
Programmable, GUI, and Design Tools.....	19
<i>Types of Tools and Pros and Cons</i> .....	19
Other Considerations.....	20
Types of Developer Systems.....	21
VIII. ETHICAL PRINCIPLES .....	22
Transparency .....	22
Applying Equity Awareness in Data Visualization.....	22
Making Data more Accessible .....	24
IX. SUMMARY OF FINDINGS .....	25
Knowing Your Audience.....	26
Working with Contractors .....	26
Creating Engaging Graphics.....	26
Showing Uncertainty .....	27
Technology .....	27
Ethical Principles.....	28
APPENDICES.....	29

# NATIONAL INSTITUTE OF STATISTICAL SCIENCES

## INNOVATIVE GRAPHICS FOR NCES ONLINE REPORTS

### EXECUTIVE SUMMARY

Well-designed statistical graphics can grab the attention of the viewer and invite curiosity while providing statistically accurate, relevant, and suitably detailed information. NCES seeks to expand how they communicate their data analysis from basic, static maps and graphs to “Living Graphics” that will encourage viewer engagement and navigation for its online reports, with special emphasis on accessibility for the full spectrum of its audiences. NCES charged the National Institute of Statistical Sciences (NISS) with convening a panel of technical expert to address three aspects to this transition: Technology, Information, and Form/Design. Specifically, the panel was asked to identify key decisions and a program for transition. The panel framed the context for their deliberations in terms of the following questions:

- What technology decisions are needed for implementation of a system with a 5- to 10-year lifespan?
- What kinds of statistical graphics will make increased, relevant, and interesting information accessible to viewers?
- What are the design opportunities for interactive graphics and maps to engage viewers?

The panel met virtually for four conferences between January 6 and January 22, 2021, with additional *ad hoc* video conferences during preparation of this report.

The panel proposed a path forward in three phases from static to dynamic custom graphics.

**Phase 1 consists of immediate goals for NCES and its contractors.** Initiate work on improving static graphics to communicate data more effectively by utilizing data visualization best practices, including a greater variety of static graph types, and incorporating consistent styles and branding. At the same time, NCES can immediately explore creating static “infographics” that combine charts, text, images, and more into single images. First steps are to build a data visualization resource library and provide training to NCES staff.

**Phase 2 for NCES and its contractors.** Advance graphics to include basic interactive graphics such as drag and drop tools to the NCES websites. Introducing interactive capabilities would encourage users to interact with the data in ways that help them gain insights more easily. Once these capabilities are in place, NCES staff and contractors need only modest training to include interactive graphics in their reports and webpages.

**Phase 3 for NCES.** Invest in creating more complex graphics--like those that are commonly published by major media outlets. These typically require custom coding and programming languages, requiring additional staff or working with contractors.

### Major Strategic Recommendations include:

1. Identifying the target audience at the start of each project is a top priority. Data visualizations have limited impact without understanding what the target audience wants and what they will use. The desired level of NCES directed data story versus user-directed exploration and local versus national focus may vary greatly by target audience. Information about the audiences and whether efforts to engage them are successful should be collected.
2. A more strategic approach is needed for the communication efforts between NCES and their contractors. In particular, RFPs should ask for:
  - a) A portfolio of the contractors' past creation of standard types of charts as well as more novel, engaging charts.
  - b) Specification of a data communication team such as data visualization specialists, graphic designers, and media experts.
  - c) A full dissemination and social media plan for the final product.
3. Two products are needed for NCES staff and contractors:
  - a) A branded style guide, particularly for charts, graphs, and tables.
  - b) Templates for standard packages that can be reused by agency staff and contractors.
4. Choice of static or interactive format for each visualization depends on the particular use of the graphic and the audience. While interactive graphics often get high-profile attention, they also can interact badly with some platforms and can be expensive to create.
5. Each visualization should be coupled with easy access to a data file, spreadsheet or other easily accessible format that contains both the visualized measurements<sup>1</sup> and their associated uncertainties displayed.
6. By the final phase of implementation of the project, most visualizations produced by contractors should be designed so that NCES can assure long-term maintenance of the code and the capability to add new features without returning to the contractor for minor revisions. For more for complex and customized projects that require the use of expensive tools or tools with steep learning curves, NCES may need to consider staffing changes that may not be cost-effective.

Additional specific recommendations are included in the report.

---

<sup>1</sup> NCES needs a data file containing information required to reconstitute the graphic, for purposes of maintenance of the online report and graphic. Data files for public access are limited to data points that could be extracted from the graphic, such as coordinates for points or equations for lines displayed.

## Innovative Graphics

### NATIONAL INSTITUTE OF STATISTICAL SCIENCES EXPERT PANEL REPORT

## PREFACE

The National Center for Education Statistics (NCES) has moved the distribution of its reports from paper to online. A benefit of this change is that new ways of displaying data in the reports are more easily available. For a start, the need for terseness is lessened, so that graphical displays can be more abundant and need not be limited to the most efficient representation or only the most important findings. The new technology for information delivery can also provide a way to deliver a greater variety of data displays. For example, dynamic graphics and those that invite viewer interaction become possible when a report is not limited to paper. Such a complete change in reporting methods provides an opportunity to rethink the role of graphics in the information that NCES provides to all its constituencies. Therefore NCES asked the National Institute of Statistical Sciences (NISS) to convene a panel of technical experts to advise them on what issues should be considered when making the change, including guidance on technology, statistical, and design issues. This report is the result of that panel's work.

# NATIONAL INSTITUTE OF STATISTICAL SCIENCES EXPERT PANEL REPORT

## INNOVATIVE GRAPHICS FOR NCES ONLINE REPORTS

### I. INTRODUCTION

#### BACKGROUND

The National Center for Education Statistics (NCES), being a federal statistical agency, presents many of its findings about education in the United States as data summaries, either as tables or graphs of statistics. NCES disseminates all its reports on its website, rather than as paper documents. This provides more flexibility in the types of graphics feasible to display, and it also makes it possible to present them in a way that is more personalized to the audience's interest and need for detail. This new way of disseminating information also makes it accessible to people who may not have even been aware of the work of NCES or familiar with its reports; that is, to audiences beyond its mandated constituencies. The performance of the U.S. education system and outcomes of its students are topics that could be interesting and/or useful to many people in our country, from parents to strategic planners in business. NCES sees the change to web-based information dissemination as a way to reach these new audiences to ensure that they have access to accurate information.

#### GOALS

A long-term goal of NCES is to make its website and its data summaries the first stop for questions about the U.S. education system, for audiences of all backgrounds and experiences. One way to do that is to improve the way it uses graphical data displays in its reports and briefs, including with "living graphics." By this, they mean graphs that are engaging, accessible, and dynamic. For example, they would like to provide some interactive graphics, so that viewers can explore or reveal interesting relationships for themselves.

#### PANEL PURPOSE

NCES therefore asked the National Institute of Statistical Sciences (NISS) to convene a panel to make recommendations on how to best accomplish the goal of reaching and engaging diverse audiences with graphics. (The Charge to the panel appears in Appendix A.) The resulting panel contained experts in statistical graphics, design, computing technologies, and human perception. This report contains the results of the panel's deliberations and recommendations, both about graphics and how the agency might navigate its path to change.

#### REPORT FORMAT

The report is divided into 2 parts. The first part (Sections 2 through 4) discusses considerations for NCES as it attempts to engage new audiences and develop a revamped website.

## Innovative Graphics

- Section 2 covers how NCES can **clarify its goals** for a revamped graphical communication plan for its website.
- Section 3 discusses why **identifying the audience** is so important to the choices made in graphical design.
- Section 4 is a guide to using the **expertise of NCES's own staff and contractors** for producing graphical content.

The second part of the report, including Sections 5 through 8, describes recent thinking about effective graphical displays and technical advice about how to produce them.

- Section 5 summarizes some **techniques for making graphics engaging and informative**. Appendices provide examples of graphical displays that illustrate the techniques.
- Section 6 describes current understanding of how to most **effectively display uncertainty**.
- Section 7 summarizes considerations that should be made when selecting **technologies** to produce graphics, and some recommendations about choices.
- Section 8 raises some considerations of **ethics and access for data display**.

The **findings** from these two parts are summarized in Section 9.

## II. GOALS FOR A REVAMPED WEB GRAPHICS PLAN

It is realistic to expect that more engaging graphics and visuals to tell interesting stories would bring more viewers to the NCES website. Large media outlets, such as the *New York Times*, *Washington Post*, and *The Guardian*, offer models of how effective visual storytelling can drive attention and page views. Making data, code, and graphics easier to find and understand would likely also drive greater coverage from the media.

The most effective types of data visualizations for NCES and its contractors to create in the future are, first, a function of the target audiences. Data visualizations will have limited impact without understanding what the target audience wants and will use. Do users/readers need to select, filter, and interact with the data? Do they need visualizations that **explain** concepts with simple, clear representations of the data? Do they want to **explore** the data, either through more complex/dense static graphics or through interactive visualizations?<sup>2</sup> Honestly assessing and answering these questions will help guide the agency to the types of data visualizations that will be most useful for its target audience. We discuss this in more detail in the next section.

Second, creators of data visualization programs must now consider the format of its visualizations, meaning the choice of static or interactive content. While interactive data visualizations often get the most high-profile attention, there has been a recent widespread pull back from interactive data visualizations, for two reasons:

1. A large and increasing percentage of users consume information on websites by scrolling through content and not clicking or hovering. With the increased use of mobile platforms, the way people interact with online content has changed dramatically.

---

<sup>2</sup> See more detail on these dichotomies, along with examples, in [Schwabish, 2014](#) and [Schwabish, 2021](#).

## Innovative Graphics

2. Creating interactive online content is often more expensive than static content. Even with free, open-source tools like the D3 JavaScript library and the R programming language (see Section 7 for more on data visualization technology tools), there still needs to be a staff member with expertise to load the visualization into the agency's website platform.

Thus, before pursuing interactive dashboards and data visualizations, NCES should carefully consider whether each target audience needs such interactivity.<sup>3</sup>

Third, changes in data visualizations used in reports can be made in small steps to build momentum. That momentum can be generated from staff members and contractors interested in creating more, new, and different graphic types, as well as externally from stakeholders seeking different kinds of content, data, and research.

### A PATH TO BETTER DATA VISUALIZATION

One way for NCES to proceed is to plan for three different stages of development for their data visualization implementation, taking place in the immediate-, middle- and long-term.

**IMMEDIATE:** The panel recommends that NCES and its contractors start immediately creating a more diverse set of graphics, adopt effective data visualization principles, and incorporate consistent styles and branding. For example, the agency can create effective data visualizations that accompany many of the tables found throughout the NCES website by employing basic line charts, dot plots, slope charts, and other simple graphic types. These will more effectively show the arguments or patterns in the data and do not require much time, technology, or advanced expertise. These charts, while simple, are the bread-and-butter of most data graphics and can be created in any statistical software that the staff already uses.

This can be supplemented by creating static "infographics" that combine charts, text, and images into single images. While these sorts of graphics often require skilled graphic or information designers, they can be effective ways to visually summarize information, especially for NCES's audience of parents and students. They can also be useful "leave-behinds" for policymakers and other stakeholders. Some organizations make these accessible and engaging types of graphics available as part of press releases.<sup>4</sup>

To help accomplish these goals, the panel recommends that NCES offer data visualization training to its staff. Appendix B provides a list of resources to include in a library for a training and resource toolkit for staff. Creating an internal learning group to help foster greater collaboration within NCES is another option for building staff expertise.<sup>5</sup> Ways to get started on developing consistent styles and branding are discussed in Section 3. Suggestions for how to contract for skills the staff does not yet have (such as, perhaps, creating infographics) is discussed in Section 4. Finally, basic principles to follow for creating effective data visualizations are discussed in Section 5.

---

<sup>3</sup> See [Tse](#) (2016) and [Aisch](#) (2016) for the reasons for the move from interactive to static graphics by dataviz teams in journalism.

<sup>4</sup> For an example of the use of infographics in a press release, see <https://www.norc.org/NewsEventsPublications/PressReleases/Pages/how-medicare-recipients-responded-to-covid-19.aspx>.

<sup>5</sup> Many high-quality data visualization courses with training in various software are available through MOOC platforms, such as EdX or Coursera (e.g., Data Visualization in R).



## Innovative Graphics

**MEDIUM TERM:** In the medium-term, NCES and its contractors would add some basic interactive graphics to the NCES websites. By “basic interactivity” we do not mean adding tooltips or hovers on top of what are essentially static graphics, but instead ones that encourage users to interact with the data in ways that help them gain real insights.

Such interactivity can be implemented on the NCES website by using drag-and-drop tools like Tableau, Datawrapper, Flourish, or others (see Section 7 for more on data visualization tools). While these tools are limited in some respects, they would allow NCES and their contractors to create interactive elements on their website(s) without substantial investment in technologies or additional staff. This approach might require some additional funding (and though tools like Datawrapper and Flourish can be free to use, they might require additional website support) and training.

**LONG-TERM:** For the long-term, NCES would consider the benefits of creating more complex and custom-made interactive data visualization products. These kinds of graphics--like those that are commonly published by major media outlets like the *New York Times* and the *Washington Post*--would likely require custom coding and programming languages like JavaScript/D3 (See Section 7 for more on the differences among data visualization tools and coding required). To produce such graphics in-house would likely require more staff than are currently available. Alternatively, NCES could work with contractors with established portfolios that include this type of work.

The panel recommends that across all three time horizons, NCES encourage (or require) their contractors to take a more strategic approach to their communication efforts. For example, NCES can include a “Dissemination Strategy” requirement as part of any issued RFPs. Contractors would need to establish their expertise and experience in effectively and visually communicating the results of their analyses. Contractors would be required to show past products and metrics of success, such as page views, time on page, and media mentions. This “Dissemination Strategy” could become part of the score for all proposals and be judged on a series of metrics deemed important by NCES.

### III. KNOWING YOUR AUDIENCE

It is important to know one’s audience to create materials targeted towards them specifically and make sure their needs are met. Some websites use Google Analytics to examine the characteristics of the people who visit their website. This is, however, different from the targeted audience; Google Analytics can inform one about the website’s present audience, but the desired audience and its needs also need to be defined.

#### TYPES OF AUDIENCES

There are many potential types of audiences, such as the general public, policy makers, practitioners (e.g., state administrators), researchers, and journalists. Each group will seek different information from the website, and data presentation should be different for each. For example, the general public and journalists may be less familiar with certain topics, so they need clear definitions of certain phrases and less jargon or statistics that may be confusing to them. Researchers, on the other hand, will be more interested in the methodology and statistical inferences made from the data, and will require less explanation of terminology. Consequently, knowing the audience will help frame the types and level of information needed for the data and graphics to be useful to the targeted audience.

### AUDIENCE NEEDS

Another way to think about this is: what does the audience need? For example, journalists may only want the bottom-line, the main points to a data story, while researchers want all the data to use for themselves. The general public, on the other hand, may want to “find themselves in the data” and need more focused graphics to their state, age group, or occupation. Policymakers may want to better understand the context of the data and how it can be applied to their constituencies. Some audience types, such as members of Congress, may want to know data at the national and state level, while state administrators may be only interested in their local region.

The audience can be classified into three groups, according to their needs: (1) show me the numbers (e.g. journalists who want the bottom-line and don’t want to do digging, or parents, who are only interested in data that affects them or their children); (2) compare the numbers (e.g. a state administrator who wants to compare across regions), and (3) I can do it myself (e.g., researchers who want to do their own investigations). Having a clear picture about the target audience and their needs helps better plan the deliverables.

Identifying the target audience is best done before work begins on a data visualization project and should shape all decisions made during development. The materials and the design should both differ by audience type. For example, researchers may want data tables while the general public requires easy-to-understand infographics. Each type (researcher/table vs. general public/infographic) requires a different perspective in terms of:

- The level of data included (all the data vs. the main points)
- The level of supporting documentation (standard errors vs. no standard errors)
- The mode of delivery (downloadable csv vs. pdf)
- The design (individual level data in cells vs. icons and pictures)
- The graphical design (hardly any for the data tables vs. branded colors and iconography).

The problem is, however, when many audience types are targeted simultaneously. In this case, a pyramid method is useful. That is, starting at the top, include only the main points and bottom-line of the data (this will help the journalists and most members of the general public). This level is more focused towards infographics. Next add layers of detail by including more data and methodological facts, by including a methodology page or footnotes. The bottom level includes a link to downloadable data and supporting documentation. Each new layer does not hamper the one above it but rather supports it with progressively more detail. This method ensures that each type of audience gets what it needs and also encourages individuals to dig a little deeper. See Appendix C for an example.

### TYPES OF MATERIALS CREATED

Other ways to help discern the needs of audiences is to consider where the materials you create fall on the following scales:

1. Does the audience want presentation or exploration? Here the difference is in the depth of the data presented.

## Innovative Graphics

- If presentation is desired, the goal should be to present the data attractively in a form easy to consume. In these cases there is a predefined data story presented in a (usually) static graphic.
  - If exploration is desired, some audience types may want to engage with data and reach their own conclusions. For these audience members, a data tool that allows them to dig into the data, such as an interactive dashboard, is useful.
  - When audiences can pick and choose the data they examine, there is a risk of inadvertently allowing them to reach a false conclusion. This can occur if summaries from small samples or non-significant findings are displayed without commentary. Best practice is to have a clear story, suppress uninterpretable findings (based on statistical considerations), and use multiple reviewers to ensure viewable results are not misleading.
2. Does the visualization replace or append data tables? Some audience members want exact numbers; in that case, provide a table that can be read into software without pre-processing. Others want to see patterns and trends; for them, create summary graphics, such as those discussed in Section 5.
  3. How does the audience engage with the data? That is, how do they access the data (web, mobile, csv, static, or interactive)? Does the graphic require a server, and if so does it need data security or can anyone access it? Do they need to print it?
  4. Some audience types can only engage in more traditional static graphics. Others want interactive dashboards. Each method requires its own method of delivery (pdf vs. online tool platform).

### ASK THE AUDIENCE

The most direct way to examine whether your efforts in engaging your (potential) audience with website content are successful is to collect data about their behaviors and opinions. For audiences that you already have, such as the former recipients of your paper reports or those currently visiting your website, there are simple ways to collect this data. For the audience that you don't yet have but are targeting, the most obvious approaches, such as surveys, would be expensive, and best left to the late stage of development, if needed.

For current visitors to your website, there are several ways to collect data about what types of graphics or what topic/graphic combinations they prefer. One way is to directly observe their behavior by carrying out an experiment on the website. In e-commerce, this is known as A/B testing, which is simply the implementation of a randomized control trial on a website, where the response variable is some desired behavior. A visitor to the site would be randomized to view one type of page or another (e.g., perhaps one with a bar chart showing school start time versus other types of graphics such as those discussed in Section 5.) Then a response behavior of interest is selected and measured for each visitor, and when a sufficient number of visitors have accumulated, the summary statistics for the groups compared. In this case, the response could be a discrete event, such as downloading a report, or a measure of time spent interacting or viewing the graphics page. Software to implement such experiments on websites is readily available.<sup>6</sup>

Opinions of visitors to your website can be collected easily using user experience surveys. Like software for A/B testing, there are a variety of commercial options, and many have free versions (e.g., Google Surveys,

---

<sup>6</sup> Google Optimize has a free version, and there are many other software options with more features, costing from none to several hundred dollars per month. The statistical testing features available in some of the more expensive options should be unnecessary for NCES staff.

## Innovative Graphics

Qualtrics), though they may not allow you to customize to the extent you prefer. For example, you may want to ask visitors for feedback on a particular graphical display, or to compare their opinions of alternatives or to elicit their understanding of a display.

### OTHER STRATEGIES FOR BUILDING AN AUDIENCE

More engaging graphics are partly a means to the end of more use of NCES data by the public, including those who do not currently access the data frequently. One approach to increase usage among these audiences is to build a community of users. Some federal statistical agencies provide models for implementation of such a site. One example of these is the Data User's Group for the American Community Survey.<sup>7</sup> This website is not maintained by the Census Bureau, but rather by the [Population Reference Bureau](#). The users themselves provide much of the content. For example, a message board for users contains discussions about how to find and use ACS data, as well as a way to share code for producing data summaries. A substantial number of posted questions concern how to produce maps and displays of data. Users also produce and share presentations on projects they have completed using ACS data for high profile questions (like health and economic disparities from COVID in their regions).

The threaded discussions or other content on a user's group website can draw people to the site through search engines like Google<sup>8</sup>. Use of social media (e.g., Clubhouse, Instagram, Twitter, TikTok) for displaying engaging graphics is also a low-cost entry-level tool for reaching potential audiences who can be drawn to your site. Both of these are ways to increase traffic to the website, where they will be exposed to the new graphical content and support and be more likely to return.

## IV. WORKING WITH CONTRACTORS

Using contractors for skills not available from NCES staff can be a cost-effective approach in all three stages of development for the new data visualization program. RFPs can include a requirement that contractors specify a data communication team, such as data visualization specialists, graphic designers, and media experts. These requirements could also request a dissemination and social media plan with help from communications teams. Contractors should be able to create standard types of charts as well as more novel, engaging charts. RFPs can ask for a portfolio of contractors' work and specification of a full dissemination plan for the final product(s). Contractors should be flexible in the tools they can use and able to determine the right tool for the job. Finally, they should be aware of the types of IT support needed for online tools (such as clouds and servers). Contractors should also be flexible in their timing of deliverables. That is, different levels of templates require different levels of effort. Contractors should be able to show a breadth of skills and understanding in how long it takes to create each type of materials.

**GRAPHIC DESIGN:** There are differences between the analytical design of a graphic or interactive visualization and the graphic design. The former refers to the decisions made regarding the types of data shown (e.g., continuous or categorical), the method (e.g., comparison, change over time), and type of chart (e.g., bar, line, dot plot), as well as how to include weighting and standard errors. Graphic design, however, refers to graphical elements such as font type and size, spacing and layout, colors, branding, user experience, and

---

<sup>7</sup> See the ACS User's Group website at <https://acsdatacommunity.prb.org>.

<sup>8</sup> This is how one panel member discovered and began using ACS data tools.

## Innovative Graphics

508 compliance. In the case of analytical design, decisions should be made by statisticians or methodologists. For graphical design, however, it is sometimes necessary to work with a graphic designer, especially for highly stylized or custom visualization products. (Note: It is possible for statisticians and methodologists to do a very good job at graphical design, with appropriate training.)

In an ideal world, organizations would have a designer on site, though this can be cost-prohibitive. In that case, there are ways to enhance the graphic design of deliverables without having to hire a full-time designer. We recommend NCES develop two products for NCES staff and contractors to use in their visualization work:

1. *A branded style guide especially for charts and graphics*<sup>9</sup>: This should include suggestions for all formatting and design elements. For example, the designer could help determine positions for the chart headers, legend, and other text; determine a color palette for chart objects; select a font and set font sizes; and create a page layout for entire reports and briefs.<sup>10</sup> This guide could also include details about making the charts 508 compliant so that more items are consistent with the regulation. This would ensure compliance and reduce burden over time. Moreover, a guide should also be designed to set guidelines for social equity (see Section 8) and maximize accessibility for those with visual, physical, or intellectual impairments.
2. *Templates that can be reused by agency staff and contractors*: These provide a guide for all graphics, reports, presentation slides, infographics and briefs. The designer would create a template and train users on how to use them. This will help materials be well-designed and formatted, as well as branded.
  - Icon library: Designers can create a library of commonly used icons, such as “Education”, “Schools”, “Data” that can be applied to all projects.
  - Templates: Templates should be accessible to staff in the formats they use most frequently, such as Excel, Word, and PowerPoint. Designers could also inform templates created in programs that allow replication, such as SAS and R.
  - Training: Template creators could train staff on how to use the templates.

The difference between a grammar guide and a data visualization guide is that many of our data style decisions are subjective. While the word *their* is objectively different than *they’re*, and the use of the word *one* in a particular case is either correct or incorrect, there is no objectively correct or incorrect line thickness for a chart. There are, however, certain best practices to consider for specific chart types as well as color, font, and layout considerations. But for the most part, chosen styles will reflect the organization’s preferences and goals.

An effective, comprehensive data visualization style guide is best developed at the organizational level. If possible, the organization should bring together the design and data teams to determine branding guidelines that meet the needs of the organization. If the organization does not have these divisions, contractors or other experts may be best suited to create the guide. Referencing existing style guides - for example, [this one from the National Cancer Institute](#) can also be useful to begin the development process.

---

<sup>9</sup> See a data visualization specialist’s view on why organizations need a data visualization style guide: [policyviz.com/2021/03/16/why-your-organization-needs-a-data-visualization-style-guide](https://policyviz.com/2021/03/16/why-your-organization-needs-a-data-visualization-style-guide)

<sup>10</sup> See examples: [policyviz.com/2016/11/30/style-guides](https://policyviz.com/2016/11/30/style-guides)

### V. CREATING ENGAGING GRAPHICS

There is no one-to-one mapping between data types and graph types. A column chart, for example, can be used to show changes over time or to compare differences between categories. There are a variety of techniques and strategies analysts can use to guide their selection of a graph, chart, or diagram, but there will rarely be a “right” graph; in the end, it will depend on the goal(s) of the graph creator and the needs of the audience.

#### TECHNIQUES

With that in mind, we discuss seven techniques that can help guide the design of a graphical display that will hold the viewers’ interest, draw them to engage with the data, and meet their needs.

1. *Provide an Entry Point*

A comparative technique for both supporting and inviting meaningful comparisons is that of providing entry points: personally meaningful data points that are familiar and relatable. This enables the viewer to develop a broader understanding of the data set as a whole. For example, in a popular visualization of trends in babies’ names over the last 100+ years,<sup>11</sup> a prominent interaction feature is a search box that lets a viewer put in a name. By doing so, the visualization updates to show the trends of that name. The first name most people put in is their own, or the names of their family members. In this way, the visualization allows viewers to personalize the view, engage the viewer, and provide contextual information useful for comparison when further exploring the data.

Entry points can be explicit in a visualization - such as providing mechanisms to highlight or filter the data set based on a viewer’s input - or they can be implicit. Implicit entry points are provided through views of the data that invite a viewer to quickly find a personally relevant data point, such as a map that lets a viewer see data about the state/county/municipality where they live, or about a demographic group that they fall into. When designing a visualization to support entry points, it is important consider what attributes of a data set will allow for personal comparisons to ensure the visualization makes those apparent to the viewer. See Appendix E.1 for an example illustrating this technique.

2. *Provide Explanations*

Do not shy away from unfamiliar graphics that require explanation. Visualizations revealing interesting stories or complex comparisons are rarely self-evident. Titles and labels are often not enough to cue viewers into the most interesting messages that a visualization can reveal. Any graphic that is meant to communicate a specific message or illustrate a point should include explanations that help the viewer know where to look and how to interpret what they see. At a minimum, these explanations could be narrated in associated text that describes which aspects of the visualization support the point; but more effective is to annotate the visualization directly. These annotations can include both visual and textual elements, highlighting for the viewer where

---

<sup>11</sup> <https://www.babynamewizard.com/voyager#prefix=&sw=both&exact=false>

## Innovative Graphics

to look and providing an interpretive description about why that is an interesting part of the data. See Appendix E.2 for illustration of this technique.

### 3. *Provide a Story*

One approach to engage the viewer for any type of display, but especially if it is a complex or customized graphical form, is to provide them the “story” to be found in the data. This goes beyond providing explanations of how to understand the graphic, as discussed above. Rather, this approach provides a graphical display designed to support provocative, illuminating, and/or timely storylines revealed by the data. Of course this requires that the data provider anticipate and find the story first. These stories could be found within the mandated reporting statistics, or they could involve other statistics, such as the school starting time data which were provided to our panel for experimentation. If the latter, such activities may not be on the critical path for data production of NCES staff, and so better left to contractors. See Appendix E.3 for illustration of this technique.

### 4. *Provide Graphics that Allow Comparisons*

Making sense of data values is often a comparative task. A test score of 18 is meaningless unless it is known to be relative to a maximum score of 20; a school’s start time of 8:05 am provides limited insight unless it is known to be 15 minutes earlier than the national average. Thus one advantage of data visualizations is that they provide the viewer with the ability to put specific data information into a bigger contextual picture to support meaningful comparisons. They also make this comparison task easier for the viewer than it is for other forms of data summary. They can provide opportunities to answer a larger number of questions than one showing, for example, the distribution of a single characteristic. As the number of variables to be examined or number of comparisons to be made increases, however, the task of understanding the relationships from graphical displays requires more experience and therefore may be less engaging for the casual viewer unless thorough explanations are provided. Nevertheless, there are several relatively simple graphical displays that are especially designed for comparisons, and for which software is readily available, such as scatterplot matrices, side-by-side boxplots, and slope charts. See Appendix E.4 for illustration of this technique.

### 5. *Use Small, Multiple Charts*

One way to address the packed, single chart is to break it into smaller parts. Known as grid charts or panel charts (also called facets, trellis charts, or, most commonly, small multiples). These are smaller charts that use the same scale, axes, and scope but spread the data across multiple visuals. In other words, instead of putting all the data on one graph, create multiple, smaller versions with variations on the basic data. Examples of small multiples are the micromap in Appendix E.2, the boxplots in Appendix E.3, and the histograms in Appendix E.4. These graphics are alternatives to interactive displays that simply reveal one chart at a time using a dropdown or similar feature, and actually provide the viewer an easier method for comparison.

### 6. *Provide Summary Data used in the Graph in an Easily Manipulated Format*

Some audiences, such as researchers or visual journalists pursuing a story, will want to access the summary statistics used to make a graph. Though NCES frequently provides files of its graphs’ underlying data, they are often inconvenient to use. Specifically, labels and titles are sometimes

## Innovative Graphics

interspersed throughout the file, making it difficult to read into statistical software that expects the “row is a record, column is a variable” format. Data used in the graphic should be provided in formats that can be read into software (such as SAS) with simple input statements.

### 7. *Provide a Limited Number of Choices in Interactive Graphics*

Interactivity is engaging, but too many choices can overwhelm the viewer. This is particularly important if the target audience is the general public (e.g., parents or students). Such an audience may quickly leave a page if they encounter “dead ends,” or uninteresting results when interacting with the graphic. (This is not such a risk for the audience who needs the data for a specific well-defined purpose, such as a state NAEP coordinator who is examining her state’s performance.) When the graphic is interactive, it is useful to have the viewer’s entry point reveal a carefully chosen and meaningful overview. Then the viewer should be allowed a limited number of choices for filtering. By comparison, the NAEP Data Explorer is an example of an NCES interactive graphic (and data summary and retrieval) tool, that allows the viewer the opportunity to produce a large number of comparisons. However, the entry point for this interactive tool does not provide a data summary, and the novice user cannot anticipate the form of the outcome in advance of experimentation. Because of the large number of choices that must be made prior to reaching a result, the casual user is not likely to maintain his or her interest long enough to produce a result. On the other hand, the NAEP data explorer is excellent for an experienced user who has a specific goal, or who needs the same summary each year.

The screenshots of interactive displays in Appendix E.1 and E.3 are examples of graphics that provide the user with several well-defined choices of how to view different aspects of the data, but do not overwhelm the user with complexity.

Here are three additional practical ideas for helping NCES staff just beginning to produce new types of graphics in the initial phase of the visualization development plan.

### 8. *Start with Gray*

Whenever creating a graph, start with all gray data elements. By doing so, the chart creator is forced to be purposeful and strategic in their use of color, labels, and other elements.

### 9. *Reduce the Clutter*

The use of unnecessary visual elements distracts the reader from the central data and clutters the page. There are many different types of chart clutter: heavy tick marks and gridlines; data markers like squares, circles, and triangles; textured or filled gradients; unnecessary dimensions that distort the data; and too much text or too many labels.

### 10. *Don’t Stress about Selecting THE Right Display for a Dataset*

There is not necessarily a single solution to any data visualization challenge. One of the pleasures of data visualization development is that it can provide an opportunity for originality. Even simple graphical displays can be made more engaging for certain audiences (for example, students or their teachers) by surprising or delighting them with design features that are fun.

See Appendix E.5 for demonstration of these ideas for several basic graphical displays.



### CONSIDERATIONS FOR INTERACTIVE GRAPHICS

For a time, interactive data visualization developers followed what is commonly called the “Shneiderman Mantra” after Ben Shneiderman, the long-time director of the Human-Computer Interaction Lab at the University of Maryland in 1997: “Overview first, zoom and filter, then details-on-demand.”<sup>12</sup> This idea was used as an organizing principle for interactive visualization designers. Users are provided first with an overview of the data, then given the opportunity to filter and zoom further into the data, and then through tooltips or other labels to view more details on demand.

The way people use online content has changed dramatically since 1997, however. By 2016, mobile internet use surpassed desktop use and the smaller screens, different bandwidth requirements, and different apps and platforms changed the way people used the internet. As mobile use increased, developers were creating their own updated versions of the Shneiderman Mantra. In 2016, Gregor Aisch, then at the *New York Times*, specified a new version of the Shneiderman Mantra as “Details First, No Zoom and Filter, Overview on Desktop Only.”<sup>13</sup> This approach, he argued, was more consistent with how people actually used the internet—users want the story and they don’t want to filter or find more information in tooltips. Similarly, in that same year, Archie Tse, Graphics Director at the *New York Times*, argued against using tooltips and rollovers in interactive visualizations: “If you make a tooltip or rollover, assume no one will ever see it”. If content is important for readers to see, don’t hide it. He also argued that “if you make the reader click or do anything other than scroll, something spectacular has to happen.”<sup>14</sup>

What we have seen over the past 10 to 15 years is a change in how interactive data visualizations are built and how they are used to provide content to users. There has been a pull-back from interactive data visualizations at major media companies because of how people use them and their higher cost than static visuals. Thus, as NCES and its contractors consider how and whether to create interactive visualizations, it is important to understand whether people will work with online content, how they will do so, and who will actually use those visualizations. As we have noted throughout this report, identifying the audience is a first step in creating effective content.

The [NAEP Data Explorer](http://www.naep.edu/dataexplorer/) is an example of an exploratory data visualization that cannot be used by all audiences equally. It does not persuade the viewer to dig deeper with an initial view of any summary data. There are so many choices to make, that the casual user is likely to end up without any visualization at all, since views are suppressed when sample sizes are too small. For users of mobile technology to browse the internet, the data explorer will be difficult to use. On the other hand, the data explorer is valuable for data analysts and researchers who know exactly what summaries they need and the path to it, and those who are more likely to be conducting such work on desktop computers.

## VI. SHOWING UNCERTAINTY

While the consideration of uncertainty is a critical component of data analysis, it is a notoriously difficult measurement to communicate and interpret. Research studies have shown that both novice and expert

---

<sup>12</sup> <http://www.ifp.illinois.edu/nabhc/abstracts/shneiderman.html>.

<sup>13</sup> <https://vimeo.com/182590214>.

<sup>14</sup> <https://github.com/archietse/malofiej-2016/blob/master/tse-malofiej-2016-slides.pdf>.

analysts misinterpret standard uncertainty visual encodings,<sup>15</sup> due in part to the abstract nature of probabilities. Complicating efforts to standardize ways to visualize uncertainty is the myriad of measurements that the term uncertainty refers to: accuracy, error, precision, completeness, consistency, credibility, noise, missingness, and subjectivity.<sup>1617</sup>

There is no consensus on a best way(s) to visualize uncertainty. Each technique has strengths and weaknesses with respect to ease of interpretability and expressiveness. Padilla et al.<sup>18</sup> provide a concise overview of best practices, and ground their review of the most common uncertainty visualization techniques in empirical findings about their effectiveness. They conclude with a recommendation to avoid interval techniques such as error bars or the cone-of-uncertainty as these have been shown to be the most challenging for viewers to correctly interpret, and to instead use more representative techniques such as a scatterplot, violin plot, gradient plot, ensemble plot, or quantile dotplot.

Despite the importance of uncertainty in rigorous data analysis, many visualization professionals consciously omit uncertainty when designing visualizations for the public. In a recent survey of 90 professionals,<sup>19</sup> Hullman found a variety of reasons for the omissions, from not having access to uncertainty measurements, to not wanting the data and data collection practices to appear untrustworthy. But the most common reason to omit uncertainty in visualizations was to avoid confusing or overwhelming viewers. This latter reason is particularly important when considering an audience that may be difficult to engage.

Thus, while visualizing uncertainty may be a priority from a policy perspective, the challenges of visualizing and explaining uncertainty should not be overlooked. How and when to include visual representations of uncertainty measurements is an open research question that will continue to be studied and debated. New methods are being used, such as displaying simulated potential outcomes, but to date, it is not clear if these improve understanding of uncertainty for the general public. If including uncertainty in a visualization may detract from the goals of engagement and understandability, we recommend alternative ways of communicating this information, such as annotations, captions, or footnotes that alert the viewer to the existence of uncertainty, coupled with easy access to a data file that contains both the visualized measurements and their associated uncertainties.

## VII. TECHNOLOGY

Technology for visualization, like most other computer technology in the last few decades, has evolved at a revolutionary pace. The challenge of using this technology is to build applications that do not become obsolete more quickly than the technology itself. This means content producers must develop designs and choose development environments that facilitate implementing new requirements as well as ones that encourage effective maintenance. Unfortunately, visualization enthusiasts often are attracted more to novel aesthetics than to the underlying architecture and code.

---

<sup>15</sup> S. Belia, F. Fidler, J. Williams, and G. Cumming, "Researchers misunderstand confidence intervals and standard error bars.", *Psychological methods*, vol. 10, no. 4, p. 389, 2005.

<sup>16</sup> Brodlić K, Osorio RA, Lopes A. A review of uncertainty in data visualization. *Expanding the frontiers of visual analytics & visualization*. 2012:81-109.

<sup>17</sup> Boukhelifa N, Perrin ME, Huron S, Eagan J. How data workers cope with uncertainty: A task characterisation study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* 2017 May 2 (pp. 3645-3656).

<sup>18</sup> Padilla, Lace, Matthew Kay, and Jessica Hullman. 2020. "Uncertainty Visualization." *PsyArXiv*. April 27. doi:10.31234/osf.io/ebd6r.

<sup>19</sup> Hullman J. Why authors don't visualize uncertainty. *IEEE transactions on visualization and computer graphics*. 2019 Aug 19;26(1):130-9.

## Innovative Graphics

The following summary covers important issues that any organization should consider before choosing one or more visualization platforms for its web site and, concurrently, before negotiating conditions with a contractor responsible for contributing to that website. Each issue involves a tradeoff with others; no single visualization program has so far been able to avoid these tradeoffs. Consequently, we do not recommend one or more preferred programs and instead will point to alternatives.

### PROGRAMMABLE, GUI, AND DESIGN TOOLS

#### *Types of Tools and Pros and Cons*

This refers to the way that the person creating the graphics interacts with the data. Programming tools include packages like JavaScript, R and Python. GUI's are software packages like Tableau or Excel/PowerPoint that are point-and-click. Programming languages often have a steeper learning curve but are more flexible in what can be created. Point-and-click tools have a lower barrier to entry in that anyone can easily open and insert a chart but are limited in their flexibility.

- PROGRAMMABLE
  - R, Python, SAS/Stata/SPSS (traditional statistical packages), D3 (a JavaScript library)
  - Pros
    - Some, like R, Python, and D3 are open-source, which are free to download and use.
    - Because these are code based, they are good for replication across projects, thus reducing cost and labor.
    - Can also do data analysis, in addition to graphics (excluding D3).
  - Cons
    - These require a knowledge of that programming language and typically have a steep learning curve.
    - Interactive graphics require considerations around HTML and websites, and often around servers and data security.
    - Not all government agencies are able to use open-source tools.
- GUI (POINT-AND-CLICK)
  - Desktop: Tableau, Power BI, Excel/PowerPoint
  - Browser-based: Datawrapper, Flourish, RAW Graphs
  - Some can be used to conduct data analysis, in addition to graphics
  - Pros
    - Easy to learn, share, and implement.
      - Drag-and-drop menus enable anyone to create graphs.
      - Most people are familiar with the Microsoft Excel software package.
    - Possible to create both static and interactive visualizations.
      - Browser-based tools, especially, are designed to incorporate basic interactivity and be embedded on websites.
  - Cons
    - Cost

## Innovative Graphics

- Many of these tools are not open source and therefore require a license or paid subscription package.
- Styling and graphics libraries are more limited in terms of what can be created and how they can be styled.
- Replicability is more difficult without step-by-step tutorials/directions.
- DESIGN TOOLS
  - Adobe Illustrator and InDesign, Canva, and PowerPoint.
  - These are often used to create more customized or add detailed annotations/design to visualizations.
  - Pros: allow greater customization.
  - Cons: They lack the easy replicability of programs and are often not open-source.

### OTHER CONSIDERATIONS

#### *Interactive vs. Static*

Static visualizations lend themselves to being stored in PDF or static HTML documents. Using them avoids the worst complexities of security, deployment, and quality control. As the web becomes more interactive and engaging, however, users can regard static visualizations as quaint. Designing visualizations for both interaction as well as presentation is a useful and feasible practice. Consequently, in the short-term, we recommend static plotting programs (e.g., Adobe Illustrator or statistical package default graphics) for in-house development, and interactive platforms (e.g., Vega-Lite, eCharts, Plotly) for contractor development. Moving forward, as we have noted above, NCES staff could become familiar with some of these interactive visualization tools, especially some of the browser-based tools like Datawrapper and Flourish.

#### *Vector vs. Bitmap*

Bitmap files (e.g., BMP, PNG) became popular originally because of browser and hardware limitations. Their static sizes made embedding them in HTML documents relatively easy. Vector files (e.g., SVG), by contrast, are resizable without sacrificing resolution and more easily adaptable to different form factors (laptop, tablet, phone, etc.). We recommend that consultants provide vector files when possible.

#### *Multi-Device vs. Single-Device*

As of late 2020, phone users outnumbered laptop and tablet users accessing web pages. Because of this, many commercial companies now design their web sites to work on multiple devices. In addition, these companies test their site on multiple browsers. This trend is likely to continue in the future. We recommend that contractors program web sites for multiple devices.

#### *Open-Source vs. Closed-Source*

The trend toward open-source analytic software is accelerating, both in companies and in government installations.<sup>20</sup> The monetary advantages of open-source are obvious. The disadvantages include lack of support from a single provider and uncertainty about testing and quality control. It is possible for potential

---

<sup>20</sup> For information on measuring popularity, see <http://r4stats.com/articles/popularity/>.

## Innovative Graphics

users to evaluate the latter issue; the former is not as easily solved. Therefore we recommend that developers should use either open-source or commercial software for which they already have licenses.

### *Full-Stack vs. Proprietary Platform*

Visualizations can be embedded in full-stack applications (JavaScript, TypeScript, Go or equivalent on the front-end, a middle layer to handle handshaking, and a database back-end for data reading and transformation). Full-stack applications offer the best potential for handling big data, security, and unusual data sources. Proprietary platforms must be loaded and maintained continuously. Commercial proprietary platforms risk obsolescence. We recommend that consultants provide full-stack implementations.

## TYPES OF DEVELOPER SYSTEMS

There are several types of *visualization developer systems* available today.

- **Atomistic.** These tools contain primitive functions for drawing elements (points, lines, etc.) as well as scales, legends, and annotations. These are largely programming languages such as the D3 library in JavaScript. These tools offer the most customizability--many are free and open-source--but they tend to have a steeper learning curve.
- **Chart-Based.** These programs contain internal procedures for drawing each type of chart. Chart-based programs offer the simplest methods for creating charts, although the types of charts offered limit these programs to a simpler visualization repertoire. This class of tools enables anyone to create a graph quickly and (relatively) easily but are often limited in what they can do. This is probably the widest collection of tools because it can appeal to the largest audience. Tools in this class that are installed on the user's desktop include, for example, Microsoft Excel, Tableau, and Qlik. Tools in this class that are used through the web browser include Datawrapper, Flourish, and Vega-Lite.
- **Grammar of Graphics (GG).** These programs offer functions for each representative marker and the scales, statistics, and coordinate systems on which values are plotted. Each type has its own set of tradeoffs. These programs offer the largest range of chart types, but require a relatively more experienced user to exploit their capabilities. Examples include GGPlot2 in R and graphics.py in Python.

The tradeoffs among these types of programs affect both functionality, and maintainability. The agency should be aware of these tradeoffs for its own development process as well as for contractors'.

Many agencies rely on external contractors to develop visualizations and websites for end users. If contractors are engaged in the immediate stage of NCES's data visualization implementation plan, it is likely that one set of tools will be used for in-house development and another for consultant development. The disadvantage of this is that NCES staff will be unable to maintain the code and add new features without returning to the contractor for minor revisions. In later stages when staff has acquired skills, there should be more commonality in the tools available to both staff and contractors. This will facilitate both the back-and-forth that often happens between the two when developing a product, and also allow staff to maintain and adapt contractor-developed visualizations. On the other hand, contractors should not be restricted so tightly that the resulting work product is uninspiring. Thus for the most advanced and customized projects, contracting for developers using tools like D3 or JavaScript is valuable and likely more cost effective than hiring new staff.

## VIII. ETHICAL PRINCIPLES

### TRANSPARENCY

Trust with your target audience can be built by being transparent about your work and methods. This is especially relevant for government agencies who are tasked with collecting and processing personal information for people around the country.

Data and data visualization transparency can come in many forms, but four factors can work as guiding principles:

1. **Produce Technical Documents.** General communication material such as technical documents, guides, and blog posts that describe the approaches and modeling decisions are important for maintaining transparency and establishing trust with audiences. Technical documents relating to data collection efforts and data dictionaries should also be made public. The tone of these documents should vary based on perspectives and audience.
2. **Define the Metrics.** Intermediate and final measures should be clearly defined and described. Glossaries of terms sitewide might also be valuable. Definitions should consider the level of expertise of the audience.
3. **Give the Source of the Data.** Data sources should be disclosed including, if applicable, specific URLs for downloadable data. Questionnaires for original surveys should be made available. Exposition for any limitations and rules for redactions or omissions should also be provided with public releases.
4. **Provide Underlying, Machine-Readable Code and Data Files.** Good, reproducible code - especially if using open-source software like R or Python - should be provided, even if confidential information (e.g., variable names) are redacted. Corrections, change logs, and code versions of both data sets and code should also be provided. If the underlying data is confidential, the agency should at least try to provide the underlying methodology used to process and analyze the data (i.e., code) or other data information so that it can be potentially replicated by other users.

### APPLYING EQUITY AWARENESS IN DATA VISUALIZATION

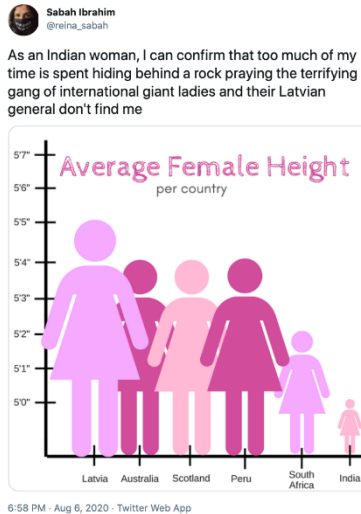
Visual content can influence how people use data and make decisions. Thus it is important for data communicators to provide careful and objective content, and also to recognize when visualizations and data presentation may have underlying bias or error, or even implicit bias that data creators may not even be aware of.

Sabah Ibrahim provides an example<sup>21</sup> of how the scaling of icons used in visualizations might be offensive and inaccurate. Ibrahim's tweet is a simple demonstration of how not being equitable and careful in how we present people and communities in our work can have a harmful effect on our readers and users.

---

<sup>21</sup> See [https://twitter.com/reina\\_sabah/status/1291509085855260672](https://twitter.com/reina_sabah/status/1291509085855260672).

## Innovative Graphics



Often the data we use may be biased or unrepresentative. D'Ignazio and Klein<sup>22</sup> describe how standard practices in data science reinforce existing power inequalities. They explore how data has been used for both good and evil - to expose injustice and improve health and policy outcomes, for example, but also to surveil and discriminate. By asking who is producing the data and for whom it is being produced, we can be better stewards of our own data and our own visualizations.

Many fields are squarely built on a model of the world in which men are the only - or maybe just the most important - participants. In *Invisible Women*<sup>23</sup>, Perez reveals the hidden places where inequality resides in even the most basic data. There are straightforward examples, like how the average smartphone is 5.5 inches long - too big for most women's hands and pants pockets. Or how the average temperature in many office buildings is five degrees too cold for women because the formula to determine the ideal temperature was developed in the 1960s based on the metabolic resting rate of a forty-year-old, 150-pound man. There are more insipient examples as well, like how women in Britain are 50 percent more likely to be misdiagnosed following a heart attack, or how car crash test dummies are based on the male body, so even though men are more likely to get into car accidents, women involved in collisions are almost 50 percent more likely to be seriously injured.

When it comes to data visualization specifically, we must be mindful of the underlying biases and inequality in how we present our results. As just one example of how data and visualizations have been used to discriminate, consider this map of Richmond, Virginia, produced in 1937 by the Home Owners' Loan Corporation (HOLC), a federal agency tasked to appraise home values and neighborhoods across the United States. As Rothstein writes in [The Color of Law](#),<sup>24</sup> "The HOLC created color-coded maps of every metropolitan area in the nation, with the safest neighborhoods colored green and the riskiest colored red. A neighborhood earned a red color if African Americans lived in it, even if it was a solid middle-class neighborhood of single-family homes." Systematic discrimination is and can be generated by how data are used and misused.

Another issue to consider is how different groups are referred to. In many cases, NCES has mandated reporting groups that are carefully defined. If this is not the case, however, terms for specific groups

<sup>22</sup> Data Feminism (2020) Catherine D'Ignazio and Klein, L.F., The MIT Press.

<sup>23</sup> Invisible Women: Data Bias in a World Designed for Men (2019), Caroline C. Perez, Abrams Press.

<sup>24</sup> The Color of Law: A Forgotten History of how our Government Segregated America (2017), Richard Rothstein, Liveright Press, NY.

## Innovative Graphics

summarized in tables and graphs, such as nationalities or ethnicities, should be those recognized by the target audience. Also consider using “people-first” language, such as “people with disabilities” instead of “disabled people.” Diversity in organizations and cultures can extend to diversity in design as people and communities are represented.

This also applies to the layout of graphs and the language used. How should bars or lines in tables and graphs be ordered? Is it alphabetical, based on sample size, or is it based on some unknown, arbitrary decision?

There are not many answers to these questions, but they are worthy of explicit consideration. Some practitioners are beginning to investigate appropriate practices.<sup>25</sup> The conscientious data visualizer will take the time to consider approaches and strategies that make their work more accessible and inclusive.

### MAKING DATA MORE ACCESSIBLE

Nearly 28 million people in the United States have an issue with web accessibility. Data and data visualization creators should consider the needs of people who may have vision, physical, or intellectual impairments or limitations that affect their ability to read or use online content. Making data and data visualization accessible also extends to considering the technology limitations some users may face when trying to connect to the internet - large data files or interactive data visualizations, for example, may be difficult for people in low-bandwidth areas to access.

With respect to static graphs or images on a website, content creators should consider implementing “alternative text” (or “alt text”). Alt text succinctly describes the content of images on the internet. Using alt text with images, graphs, and other visual elements will help make website content more accessible to people with different impairments.

Alt text can take different forms and should be applied strategically to different visual content. For example, it is generally recommended that alt text not be attached to decorative images like header photos or photos within the body of a blog post. These kinds of images might be visual styling elements such as borders, spacers, and corners; illustrative of adjacent text but do not necessarily contribute information; or convey information that is identified and described in the surrounding text.

Other images—including icons, buttons, tables, graphs, charts, and diagrams—should include alt text. In these cases, best practice is to provide a two-part text alternative: The first part is the short description to identify the image and the second part is the long description—text that represents the essential information in the chart.

The short description should include:

- The chart type (e.g., bar chart, line chart, etc.).
- What data are shown in the chart. This might be taken from the x- and y-axis titles.
- Short description alt text for the bar chart below would say something like, “Bar chart that shows the unemployment rate for 5 states in June 2018, June 2019, and June 2020.”

The long description can include elements such as:

- Scales (i.e., what are the units of the data or the range it spans).

---

<sup>25</sup> See <https://osf.io/x8tbw/>.

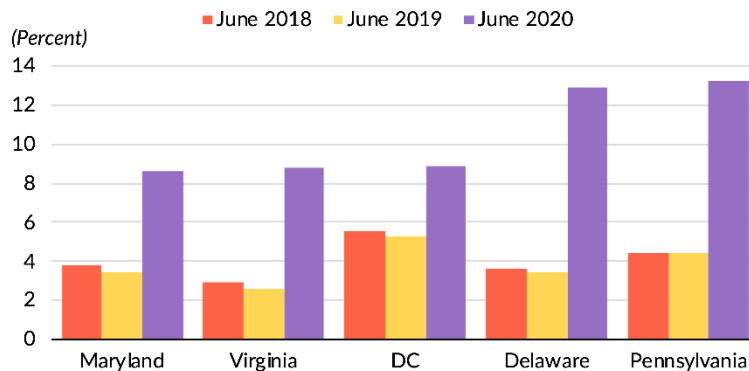


## Innovative Graphics

- Specific values worth highlighting.
- Relationships or trends that reflect the main takeaway of the chart. This might come from the title of the chart itself.
- Long description alt text for the bar chart below would say something like, “Bar chart that shows the unemployment rate for Maryland, Virginia, Delaware, DC, and Pennsylvania in June 2018, June 2019, and June 2020. The unemployment rate spiked for all five areas in June 2020 and the unemployment rate for DC was elevated for all three months.”

Content creators should consult with their technology and back-end web departments for more information on how to include alt text in their images. More references are included in the Appendix C.

**Unemployment rates in the DMV area spiked in June 2020**



Source: Bureau of Labor Statistics.

## IX. SUMMARY OF FINDINGS

The panel proposed a path forward in three stages from static to dynamic custom graphics.

Immediate Goals for NCES and its contractors: Initiate work on improving static graphics to communicate data more effectively by utilizing data visualization best practices and including a greater variety of static graph types, and by incorporating consistent styles and branding. At the same time, NCES can immediately explore creating static “infographics” that combine charts, text, images, and more into single images.

First steps are to build a data visualization resource library and to provide training to NCES staff.

Phase 2 for NCES and its contractors: Advance graphics to include basic interactive graphics such as drag and drop tools to the NCES websites. Introducing interactive capabilities would encourage users to interact with the data in ways that help them gain real insights easily.

Once these capabilities are in place, NCES staff and contractors need only very modest training to be able to include interactive graphics in their reports.

Phase 3 for NCES: Invest in creating more complex graphics--like those that are commonly published by major media outlets. These typically require custom coding and programming languages, requiring additional staff or working with contractors.

The other specific findings and recommendations of the panel correspond to Sections 3 to 8 of the report.

### KNOWING YOUR AUDIENCE

Identifying the target audience at the start of each project is a top priority. Data visualizations have limited impact without understanding what the target audience wants and what they will use. The desired level of NCES directed data story versus user directed exploration and local versus national focus may vary greatly by target audience. Information about the audiences and whether efforts to engage them are successful should be collected by NCES and contractors.

### WORKING WITH CONTRACTORS

A more strategic approach is needed for the communication efforts between NCES and their contractors. In particular, RFPs should ask for

- A portfolio of the contractors' past creation of standard types of charts as well as more novel, engaging charts. This should include a discussion of the tools used to create the graphics and the availability/ flexibility of platforms and servers. Also, how graphics will be shared with/transitioned to NCES.
- Specification of a data communication team such as data visualization specialists, graphic designers and media experts.
- A full dissemination and social media plan for the final product.

Two products are needed for NCES staff and contractors

- A branded style guide particularly for charts and graphics.
- Templates that can be reused by agency staff and contractors in tools easily accessible and familiar to NCES staff.

### CREATING ENGAGING GRAPHICS

Choice of static or interactive format for each visualization depends on the particular use of the graphic and on the audience. While interactive graphics often get high-profile attention, they also can interact badly with some very commonly used platforms (such as those on phones and tablets) and can be expensive to create.

Seven techniques that can help guide the design of a graphic display that will hold the viewers' interest, draw them to engage with the data, and meet their needs:

- 1) Provide an entry point – Give a way for the user to focus on data related to what is of particular importance to them (such as seeing their state or city).
- 2) Provide explanations – An appropriate explanation of how to read the graphic (going beyond titles and labels) can allow for use of graphics readers are unfamiliar with and ensure that if there is a specific message or point that it isn't lost.
- 3) Provide a story – Provide a graphical display designed to support the provocative, illuminating and/or timely storylines revealed by the data.
- 4) Provide graphics that allow comparisons – The real advantage of graphics over other forms of data summary is the ease of showing comparisons.

## Innovative Graphics

- 5) Use small, multiple charts – Grid or panel charts can alleviate the problem of an overpacked single chart. This can allow for the use of more variables while still being readable (such as micromaps).
- 6) Provide summary data used in the graph in an easily manipulated form – Allow those users with the skill and desire to try other graphs to do so.
- 7) Provide a limited number of choices in interactive graphics – Too many choices can overwhelm the viewer, and dead ends or uninteresting results can encourage an audience to leave a page. Have the graphic start at carefully chosen and meaningful overview.

### SHOWING UNCERTAINTY

Each visualization should be coupled with easy access to a data file (in .csv or other easily usable format) that contains both the visualized measurements and their associated uncertainties as displayed.

Considering uncertainty is a critical component of data analysis, but it is notoriously a notoriously difficulty measurement to communicate and interpret appropriately. Including the uncertainty information in annotations, captions, or footnotes may be more effective than attempting to directly incorporate them in the graphic.

### TECHNOLOGY

By the final phase of implementation of the project, most visualizations produced by contractors should be designed so that NCES can assure long-term maintenance of the code and the capability to add new features without returning to the contractor for minor revisions. For more complex and customized projects that require the use of expensive tools or tools with steep learning curves, this may not be possible. For example, inhouse NCES staff may not have the capability to modify the code or the work required to make needed changes may not be cost-effective.

The various classes of programming tools, packages, programs, or options within them have a variety of implications:

- Programmable packages (e.g., Python, R, SAS, etc.) allow for replication across projects using the initial code, can also do data analysis, and may be open-source. Such packages, however, also require a knowledge of the programming language and syntax. Furthermore, interactive graphics may require web-based programming languages such as HTML and CSS and solving other website issues. Finally, while open-source tools are free and increasingly popular, not all agencies allow use of those open-source tools.
- Graphical User Interface (GUI or point-and-click) packages such as Datawrapper, Excel/PowerPoint, Flourish, Tableau, and others can be easier to learn, share and implement. Most are not open-source, are more limited in styling and graphics choices, and are more difficult to replicate for subsequent usage without step-by-step directions.
- Design Tools e.g. Adobe Illustrator and InDesign, Figma, PowerPoint, allow the customization of graphics or the addition of additional annotation/design. They lack the easy replicability of using programs to do so or even using step-by-step directions, and are often not open-source.

## Innovative Graphics

A variety of other decision points also lead to very different outcomes and implications that may not be obvious if not considered in advance: whether the graphic is static or interactive, whether vector or bitmap files are produced, whether it is multidevice or single-device usable, whether it is open-source or closed source, whether it is full-stack or a proprietary platform, and whether it is atomistic-, chart-, or grammar of graphics-based.

### **ETHICAL PRINCIPLES**

- **Transparency.** Building trust with the target audience is especially relevant for government agencies involved in data collection. Four guiding principles of transparency can greatly impact how information is produced and target audiences respect and use the agency's data: the production of technical documents, the definition of the metrics, the providing of the data source, and providing machine-readable code and data files.
- **Equity Awareness.** Data visualizations can suffer from underlying biases and inequality in how results are presented if they are not consciously addressed. The use of colors, word choice, and layout can greatly affect the interpretations made by and about different groups.
- **Accessibility.** Making data and data visualizations reach their entire audience includes issues such as the format (not usable on some devices), file size (requiring high bandwidth), and appropriate use of alternative text (for screen readers).

### **APPENDICES**

Appendix A: Agenda

Appendix B: Charge to Panel

Appendix C: A Reference Library for Graphics and Design

Appendix D: Examples of Tailoring Data to the Audience

Appendix E: Examples of Engaging Graphics

Appendix F: Expert Panel Biosketches

**Appendix A: Agenda**

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

NCES EXPERT PANEL ON INNOVATIVE STATISTICAL GRAPHICS

January 6 & 7 & 11 & 22, 2021

AGENDA

**Wednesday, January 6**

1:00 pm – 3:00 pm	Open Session
1:00 pm – 1:20 pm	Welcome and Introduction of Panel and NCES staff
1:20 pm – 1:30 pm	Review of Charge and Presentation of Work at NCES
1:30 pm – 3:00 pm	Questions from the Panel and Discussion
3:00 pm – 3:20 pm	Break
3:20 pm – 5:00 pm	Panel Executive Working Session (closed)

**Thursday, January 7**

1:00 pm – 5:00 pm	Panel Executive Working Session (closed)
1:00 pm – 2:50 pm	Panel Deliberations (closed)
2:50 pm – 3:10 pm	Break
3:10 pm – 5:00 pm	Panel Deliberations (closed)

**Monday, January 11**

1:00 pm – 5:00 pm	Panel Executive Working Session (closed)
1:00 pm – 2:50 pm	Panel Deliberations (closed)
2:50 pm – 3:10 pm	Break
3:10 pm – 5:00 pm	Panel Deliberations (closed)

**Friday, January 22**

11:00 am – 12:00 pm	Panel Executive Working Session (closed)
12:00 pm – 1:00 pm	Break
1:00 pm – 4:00 pm	Open Session
1:00 pm – 3:30 pm	Review and Presentation of Findings by Panel
3:30 pm – 4:00 pm	Questions from NCES Staff and Discussion

### Appendix B: Charge to Panel

#### **CHARGE TO PANEL ON INTERACTIVE STATISTICAL GRAPHICS IN NCES ONLINE REPORTS**

Well-designed statistical graphics grab the attention of the viewer and invite curiosity while providing statistically accurate, relevant information, suitably detailed. NCES seeks to move from static maps and graphs to “Living Graphics” that will encourage viewer engagement and navigation for its online reports, with special emphasis on accessibility for the full spectrum of its audiences. The panel is asked to address three aspects to this transition: Technology, Information, Form/Design.

- What technology decisions are needed for implementation of a system with a 5-10 year [SL1] lifespan?
- What kinds of statistical graphics will make increased, relevant and interesting information accessible to viewers?
- What are the design opportunities [SL2] for interactive graphics and maps to engage viewers?

The panel is charged with identifying key decisions and a programme for the transition.

### Appendix C: A Reference Library for Graphics and Design

#### General Graphics Theory and Technique

- Carr, D. B. (1994). Using gray in plots. *Statistical Computing & Statistical Graphics Newsletter*, 5, 11–14.
- Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). *Graphical methods for data analysis*. Monterey, CA: Wadsworth Publishing.
- Cleveland, W. S. (1984). Graphs in scientific publications. *The American Statistician*, 38, 19–26.
- Cleveland, W. S. (1994). *The elements of graphing data* (Rev. Ed.). Summit, NJ: Hobart Press.
- Cleveland, W. S. (1995). *Visualizing data*. Summit, NJ: Hobart Press.
- Few, S. (2009). *Now you see it: Simple visualization techniques for quantitative analysis*. Oakland, CA: Analytics Press.
- Fienberg, S. (1979). Graphical methods in statistics. *The American Statistician*, 33, 165–178.
- Inbar, O., Tractinsky, N., & Meyer, J. (2007). Minimalism in information visualization: Attitudes towards maximizing the data-ink ratio. In *Proceedings of the 14th European Conference on Cognitive Ergonomics* (pp. 185-188). New York: ACM Press.
- Kosslyn, S. M. (1994). *Elements of graph design*. New York: W. H. Freeman.
- Norman, D. A. (2007). Simplicity is highly overrated. *Interactions*, 14, 40-41.
- Rangecroft, M. (2003). As easy as pie. *Behavior & Information Technology*, 22, 421-426.
- Robbins, N. B. (2005). *Creating more effective graphs*, Hoboken, NJ: Wiley.
- Schwabish, J. (2021). *Better data visualizations*. New York: Columbia University Press.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Wilkinson, L. & Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *The American Psychologist*, 54, 594-604.

#### Specific Graph Types

- Carr, D. B., Littlefield, R. J., Nicholson, W. L., & Littlefield, J. S. (1987). Scatterplot matrix techniques for large N. *Journal of the American Statistical Association*, 82, 424–436.
- Carr, D.B. & L. W. Pickle (2010). *Visualizing data patterns with micromaps*. Chapman and Hall/CRC.
- Monmonier, M. (1996). *How to lie with maps*. Chicago: University of Chicago Press.
- Spence, I., & Lewandowsky, S. (1991). Displaying proportions and percentages. *Applied Cognitive Psychology*, 5, 61-77.
- Wilkinson, L. (1999). Dot plots. *The American Statistician*, 53, 276–281.
- Wilkinson L. (2006). *Revising the Pareto chart*. *The American Statistician*, 60, 332-334.



### Technology for Graphics

Wilkinson, L. (2005). *The grammar of graphics* (2nd. Ed). New York: Springer-Verlag.

Wilkinson, L. (2009). *SYSTAT, Version 13*. Chicago: Systat Software Inc.

### History of Graphics & Example Graphs

Beniger, J. R., & Robyn, D. L. (1978). Quantitative graphics in statistics: A brief history. *The American Statistician*, 32, 1–11.

Playfair, W. (2005). *The commercial and political atlas and statistical breviary*. Original edition (1786) edited and republished by H. Wainer and I. Spence. Cambridge: Cambridge University Press.

Spence, I. (2005). No humble pie: The origins and usage of a statistical chart. *Journal of Educational and Behavioral Statistics*, 30, 353–368.

Tufte, E. R. (1983) *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Tufte, E. R. (1990). *Envisioning data*. Cheshire, CT: Graphics Press.

Tufte, E. R. (1997). *Visual explanations*. Cheshire, CT: Graphics Press.

Wainer, H. (1997). *Visual revelations: Graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot*. New York: Springer-Verlag.

Wainer, H. (2009). *Picturing the uncertain world*. Princeton, NJ: Princeton University Press.

Wainer, H., & Spence, I. (1997). *Who was Playfair?* *Chance*, 10, 35-37.

### Human Perception

Bertin, J. (1981). *Graphics and graphic information—processing*. English translation by W. J. Berg and P. Scott. New York: Walter de Gruyter.

Cleveland, W. S., & McGill, R. (1983). A color-caused optical illusion on a statistical graph. *The American Statistician*, 37, 101–105.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79, 531–554.

Gillan, D. J., Richman, E. H., (1994). Minimalism and the syntax of graphs. *Human Factors*, 36 (4)

Gillan, D. J., & Sorensen, D. (2009). Effects of graph backgrounds on visual search. *Proceedings of the HFES 53rd Annual Meeting*, 1096–1100.

Lewandowsky, S., & Spence, I. (1989). The Perception of Statistical Graphs. *Sociological Methods & Research*, 18, 200-242.

Lewandowsky, S. (1999). Statistical graphs and maps: Higher level cognitive processes. In M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, & R. Tourangeau (Eds.), *Cognition and survey research* (pp. 349-362). New York: Wiley.

McNamara, T. P., Ratcliff, R., & McKoon, G. (1984). The mental representation of knowledge acquired from maps. *Journal of Experimental psychology: Learning, Memory, and Cognition*, 10, 723-732.

Simkin, D., & Hastie, R. (1987). An information processing analysis of graph perception. *Journal of the American Statistical Association*, 82, 454–465.

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 683–692.

## Innovative Graphics

Spence, I. (2004). The apparent and effective dimensionality of representations of objects. *Human Factors*, 46, 738-747.

Tversky, B., & Schiano, D. (1989). Perceptual and conceptual factors in distortions in memory for maps and graphs. *Journal of Experimental Psychology: General*, 118, 387-398.

Tversky, B., Agrawala, M., Heiser, J., Lee, P. U., Hanrahan, P., Phan, D., Stolte, C., Daniele, M.-P. (2007). Cognitive design principles for generating visualizations. In G. Allen (Ed.). *Applied spatial cognition: From research to cognitive technology*, (pp. 53–73). Mahwah, NJ: Lawrence Erlbaum.

Tversky, B., Zacks, J., Lee, P. U., & Heiser, J. (2000). Lines, blobs, crosses, and arrows: Diagrammatic communication with schematic figures. In M. Anderson, P. Cheng, & V. Haarslev (Eds.), *Theory and application of diagrams*, (pp. 221–230). Berlin: Springer-Verlag.

Wainer, H., & Velleman, P. F. (2001). Statistical graphs: Mapping the pathways of science. *The Annual Review of Psychology*, 52, 305-335.

Ware, C. (2008). *Visual thinking for design*. Burlington, MA: Morgan Kaufman.

### Accessibility in Graphics

Cesal, A. (2018, Jul 26). 5 easy ways to make your data visualization more accessible. *Storytelling with Data*. <https://www.storytellingwithdata.com/blog/2018/6/26/accessible-data-viz-is-better-data-viz>.

Cesal, A. (2020, Jul 23). *Writing Alt Text for Data Visualization*. Nightingale. <https://medium.com/nightingale/writing-alt-text-for-data-visualization-2a218ef43f81>.

Digital Services Georgia (n.d.). *Making Your Documents Accessible, 2 Adding Alt Text to Images & Tables*. Retrieved April 4, 2021 from <https://digitalservices.georgia.gov/2-adding-alt-text-images-tables>.

Eggert, E. and Abou-Zahra, Shadi, eds. (2019, July 27). *Web Accessibility Tutorials, An alt Decision Tree*. The W3C Web Accessibility Initiative (WAI). <https://www.w3.org/WAI/tutorials/images/decision-tree/>.

Eggert, E. and Abou-Zahra, Shadi, eds. (2019, July 27). *Web Accessibility Tutorials, Complex Images*. The W3C Web Accessibility Initiative (WAI). <https://www.w3.org/WAI/tutorials/images/complex/>.

Fossheim, S.L. (2020, Jan 12). An intro to designing accessible data visualizations. *Sarah L. Fossheim*. <https://fossheim.io/writing/posts/accessible-dataviz-design/>.

MacArthur Foundation (n.d.) *Inclusive Design: Bring Web Accessibility to Your Nonprofit, Things You Can Do Today*. Retrieved April 4, 2021 from <https://www.macfound.org/pages/get-started/things-you-today>.

Schepers, D. (2020). *Why Accessibility is at the Heart of Data Visualization*. Nightingale. <https://medium.com/nightingale/accessibility-is-at-the-heart-of-data-visualization-64a38d6c505b>.

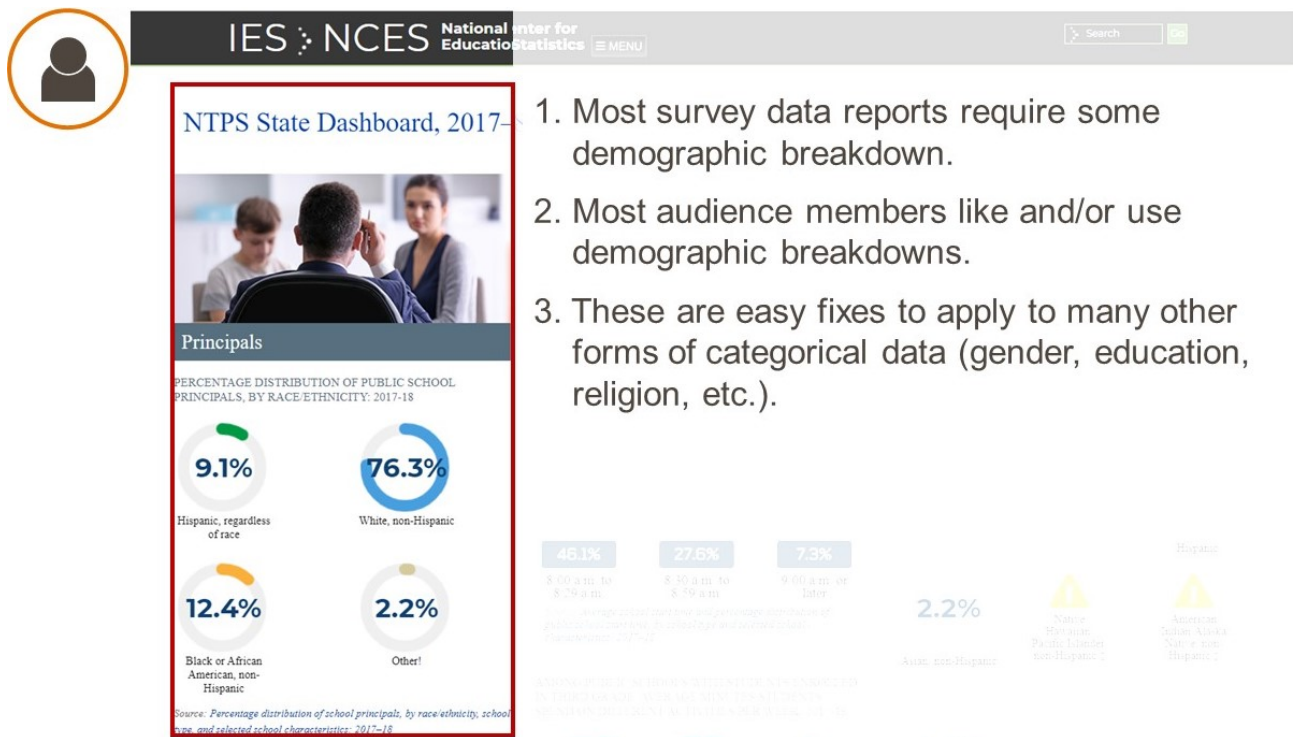
## Appendix D: Examples of Tailoring Data to the Audience



Who is your audience?

And how are they using the data?


Nola du Toit




1. Most survey data reports require some demographic breakdown.
2. Most audience members like and/or use demographic breakdowns.
3. These are easy fixes to apply to many other forms of categorical data (gender, education, religion, etc.).

We used data from this site: <https://nces.ed.gov/surveys/ntps/ntpsdashboard/Dashboard/TX>.

Consider, what type of audience do you have?

 Show me my numbers!!

 Compare the numbers!!

 I can do it myself!!

Dr. Leak's boss wants her to write a report on data from Illinois.

 Show me my numbers!!



**DR. DONNA S. LEAK**  
VICE CHAIRPERSON  
Illinois State Board of Education

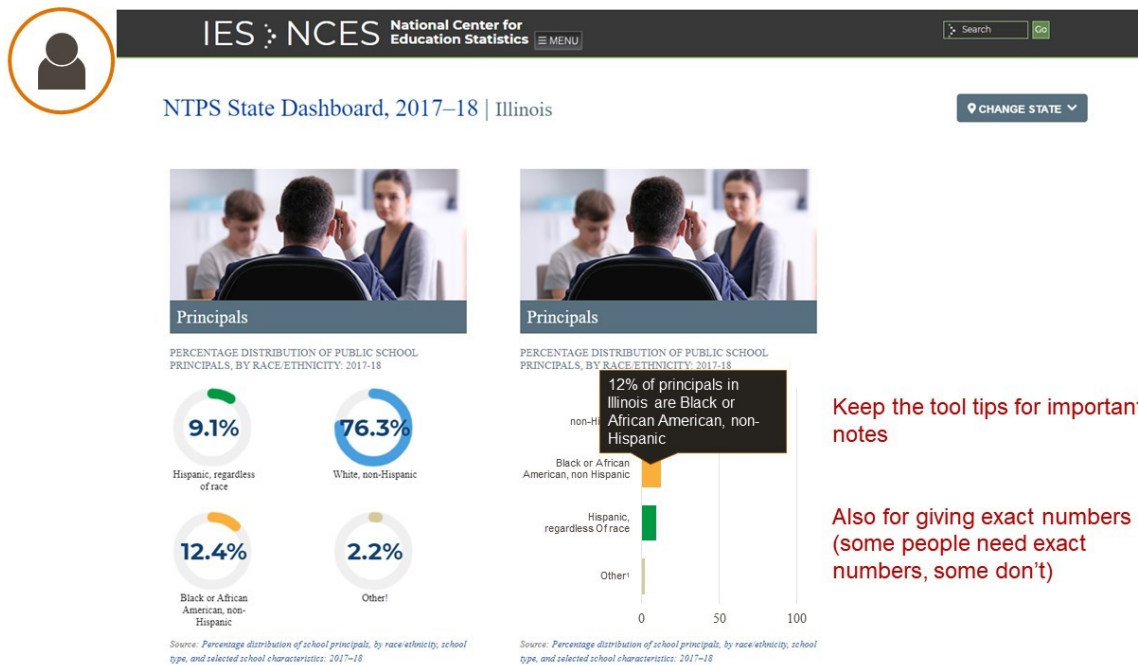
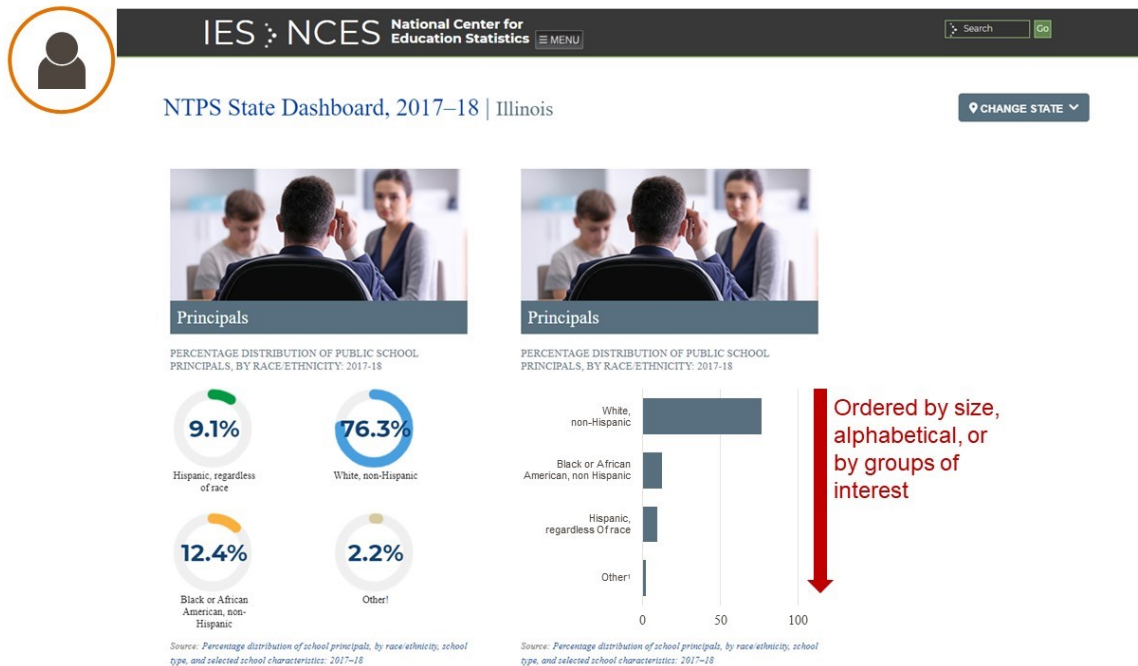


**DR. CARMEN I. AYALA**  
STATE SUPERINTENDENT OF EDUCATION  
Illinois State Board of Education

Write a report on the demographic  
breakdown of principals in Illinois.

# Innovative Graphics

She would find the pie charts useful, though the bar chart, in order of size, may be easier to read.



That works for Dr. Leak because it is targeted to her state. But what if someone wants to compare state data?

## Innovative Graphics

Consider Bill Gates as the audience. He wants to compare states to find one that has the highest proportion of minorities.



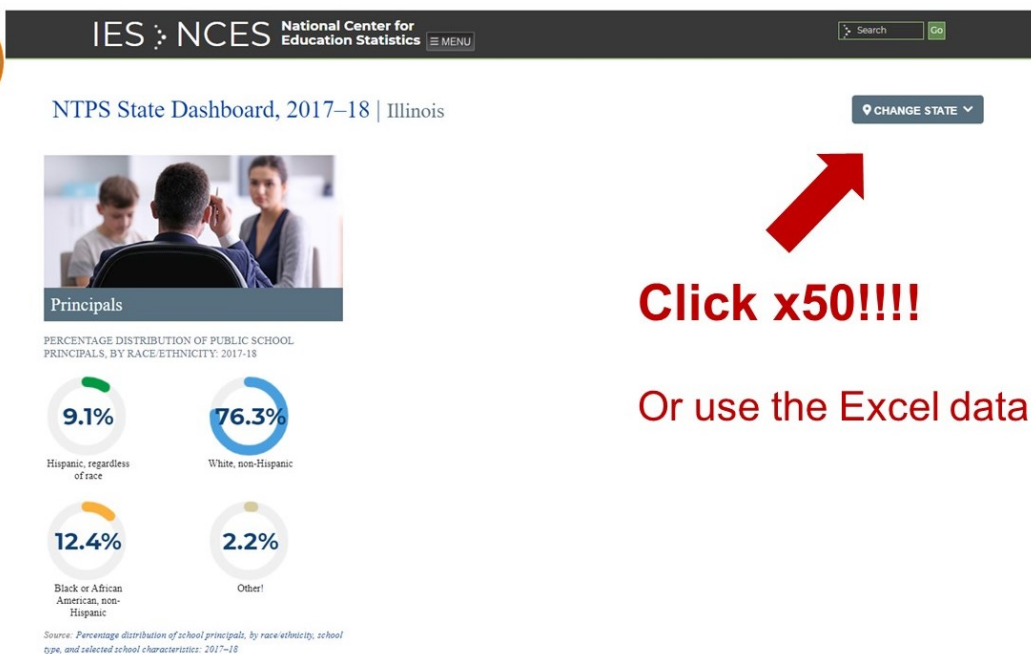
Compare the numbers!!



BILL GATES  
CHAIRPERSON  
The World

Wants to funnel \$\$\$ into states with schools with a large number of minorities.

In its current format, Bill would have to do a lot of work to get what he needs from this site.



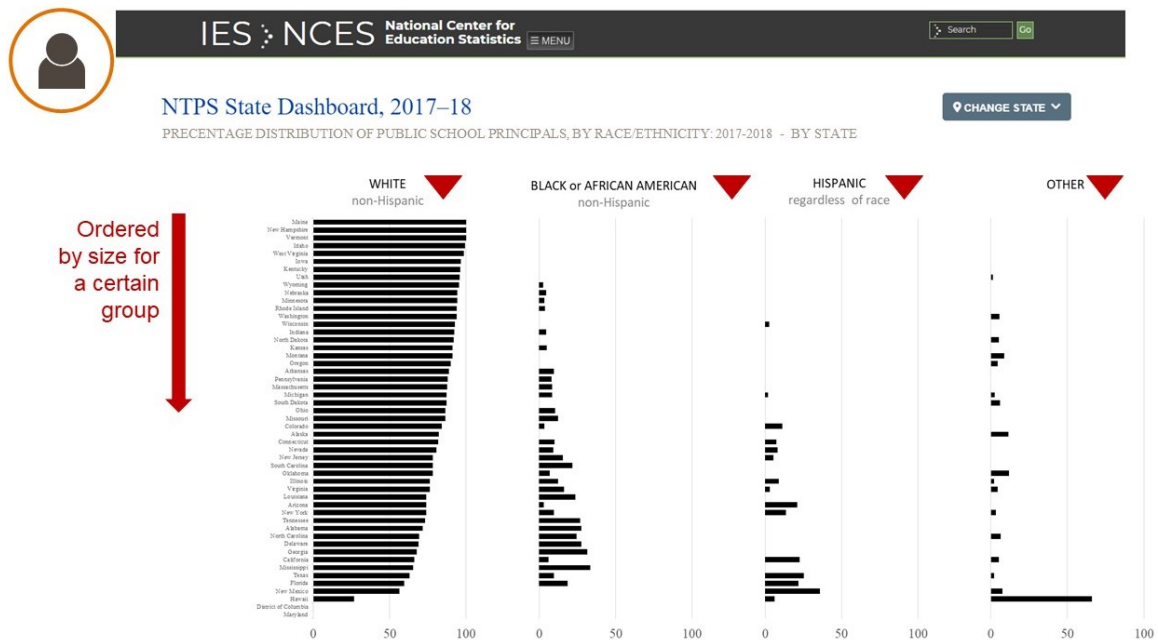
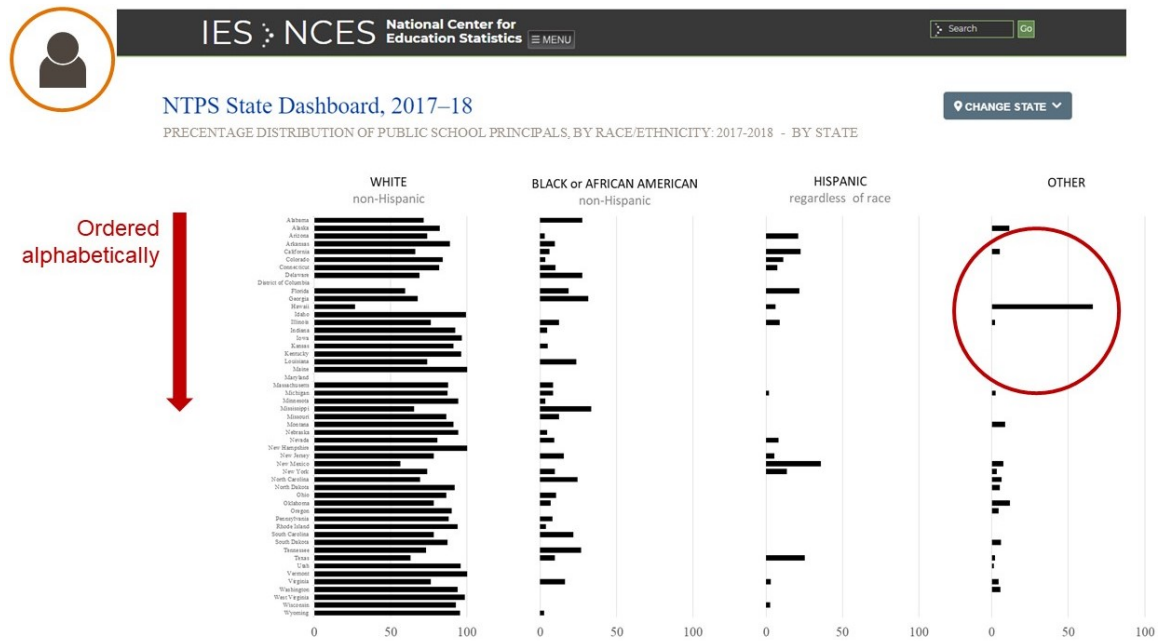
**Click x50!!!!**

Or use the Excel data

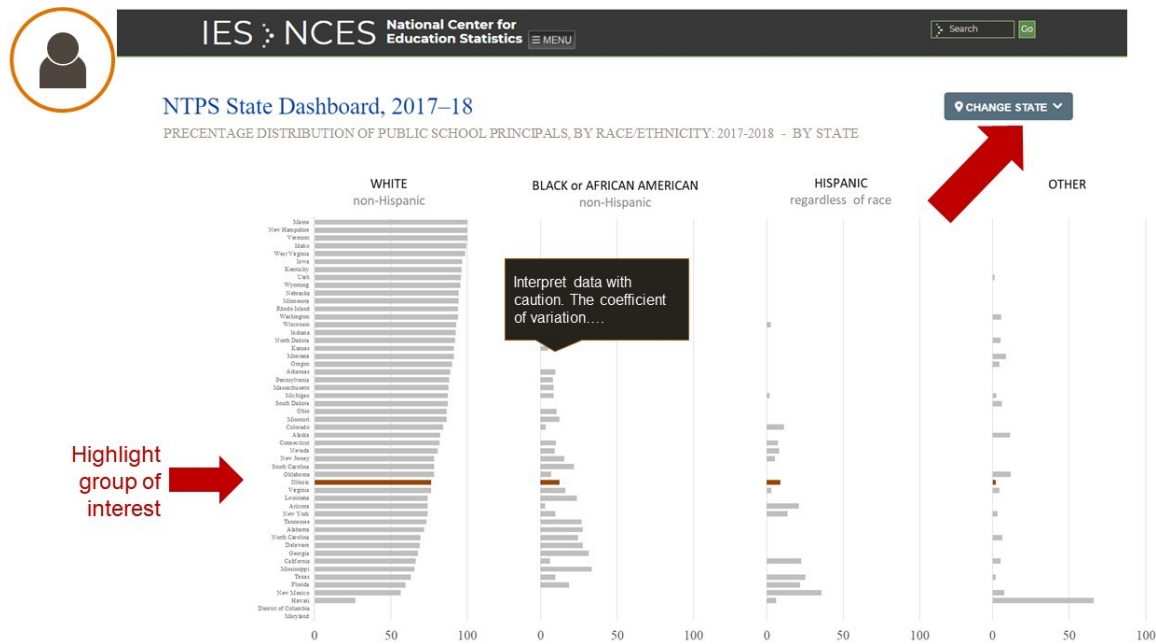


# Innovative Graphics

A simple bar chart with all the states' data would allow for comparisons.



## Innovative Graphics



A third audience type is the research who just wants to access the data.



I can do it myself!!



LYNNE STOKES  
STATISTICIAN

Southern Methodist University

Just wants the data!



# Innovative Graphics

Even though the site lets users get the data, again, getting ALL the data is cumbersome.



IES NCES

National Center for Education Statistics

MENU

Search

Go

Were you contacted about the 2020-21 NTPS? Please contact [ntps@census.gov](mailto:ntps@census.gov) or 1-888-595-1338 for more information.



National Teacher and Principal Survey (NTPS)

List of Tables

Download Excel (16 KB)

Download SE Excel (16 KB)

Percentage distribution of school principals, by race/ethnicity, school type, and selected school characteristics: 2017–18

School type and selected school characteristic	Hispanic, regardless of race <sup>1</sup>	White, non-Hispanic	Black or African American, non-Hispanic	Other <sup>2</sup>
All schools	8.0	79.6	9.4	3.1
All public schools	8.9	77.7	10.5	2.9
School classification				
Traditional public	8.6	78.6	10.0	2.7
Charter school	12.3	66.5	16.3	4.8
Community type				
City	14.9	60.8	20.1	4.2
Suburban	10.2	78.5	9.4	1.9
Town	4.6	87.6	5.0	2.8
Rural	3.1	89.7	4.3	2.7
School level				
Primary	9.6	76.7	10.8	2.9
Middle	8.8	77.9	10.3	3.0
High	8.2	78.9	10.6	2.4
Combined	6.6	80.8	8.9	3.7
Student enrollment				
Less than 100	6.6	84.4	6.7	2.3
100–199	7.8	79.1	9.8	3.3
200–499	6.3	79.2	11.6	3.0
500–749	11.4	75.2	10.5	2.8
750–999	14.0	73.3	9.9	2.8
1,000 or more	10.5	76.7	10.0	2.8

Percent of K-12 students who were approved for free or reduced-price lunches

x50!!!!

Also, the formatted tables in the data downloads are not conducive to analysis (i.e. footnotes, blank lines, superscripts, symbols, etc.).



Percentage distribution of school principals, by race/ethnicity, school type, and selected school characteristics: 2017–18

School type and selected school characteristic	Percent of principals by race/ethnicity			
	Hispanic, regardless of race <sup>1</sup>	White, non-Hispanic	Black or African American, non-Hispanic	Other <sup>2</sup>
All schools	8.0	79.6	9.4	3.1
All public schools	8.9	77.7	10.5	2.9
School classification				
Traditional public	8.6	78.6	10.0	2.7
Charter school	12.3	66.5	16.3	4.8
Student enrollment				
Less than 100	4.7	82.7	8.3	4.3
100–199	6.3	87.0	2.8	3.9
200–499	4.4	90.6	2.6	2.3
500–749	†	95.1	†	†
750 or more	2.0	92.5	†	4.4

<sup>1</sup> Interpret data with caution. The coefficient of variation (CV) for this estimate is between 30 percent and 50 percent (i.e., the standard error is at least 30 percent and less than 50 percent of the estimate).

<sup>†</sup> Reporting standards not met. The coefficient of variation (CV) for this estimate is 50 percent or greater (i.e., the standard error is 50 percent or more of the estimate).

<sup>1</sup> Hispanic includes Latino.

<sup>2</sup> Other includes American Indian/Alaska Native, non-Hispanic; Asian, non-Hispanic; Native Hawaiian or Other Pacific Islander, non-Hispanic; and two or more races, non-Hispanic.

NOTE: Detail may not sum to totals because of rounding and because some data are not shown.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Teacher and Principal Survey (NTPS), "Public School Principal and Private School Principal Documentation Data Files," 2017–18.

## Innovative Graphics

Simplifying the format would make it easier for researchers to analyze the data or feed into statistical programs.



**Percentage distribution of school principals, by race/ethnicity, school type, and selected school characteristics: 2017–18 (Percent of principals by race/ethnicity)**

School type	School characteristic	Hispanic, regardless of race <sup>1</sup>	White, non-Hispanic	Black or African American, non-Hispanic	Other <sup>2</sup>
All schools	All schools	8.0	79.6	9.4	3.1
All public schools	All public schools	8.9	77.7	10.5	2.9
All public schools	Traditional public	8.6	78.6	10.0	2.7
All public schools	Charter school	12.3	66.5	16.3	4.8
Student enrollment	Less than 100	4.7	82.7	8.3	4.3
Student enrollment	100–199	6.3	87.0	2.8	3.9
Student enrollment	200–499	4.4	90.6	2.6	2.3
Student enrollment	500–749		95.1		
Student enrollment	750 or more	2.0	92.5		4.4
! Interpret data with caution. The coefficient of variation (CV) for this estimate is between 30 percent and 50 percent (i.e., the standard error is at least 30 percent and less than 50 percent of the estimate).					
† Reporting standards not met. The coefficient of variation (CV) for this estimate is 50 percent or greater (i.e., the standard error is 50 percent or more of the estimate).					
<sup>1</sup> Hispanic includes Latino.					
<sup>2</sup> Other includes American Indian/Alaska Native, non-Hispanic; Asian, non-Hispanic; Native Hawaiian or Other Pacific Islander, non-Hispanic; and two or more races, non-Hispanic.					
NOTE: Detail may not sum to totals because of rounding and because some data are not shown.					
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Teacher and Principal Survey (NTPS), "Public School Principal and Private School Principal Documentation Data Files," 2017–18.					

**Keep it simple.  
There is no perfect way.  
Limit the work Lynne has to do to get the data ready to be  
processed.**

Here are a few others ways to think about it...



Where does your audience fall?



**PRESENTATION**

**EXPLORATION**

You present the data in visual form

The users find what they think is interesting

Tables are also “visualization”

Show ALL the data

You can shape the experience, the story

Can be a rabbit hole with design

## How do they access the data?



PDFs, images

More traditional formats

More limited for exploration

Filters, selectors, tool-tips

Depending on the tools and platform

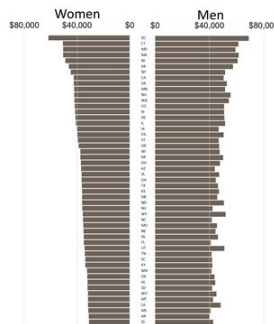
How does the user access the data?

Mobile

## How much detail do they need?



The bottomline  
Can be biased; does not tell the whole picture



## ALL THE DATA

State	Women	Men
AK	46,288	57,318
AL	32,136	44,245
AR	31,161	39,916
AZ	36,916	43,945
CA	42,486	50,539
CO	41,690	50,898
CT	50,706	61,385
DC	61,718	68,932
DE	41,278	50,976
FL	34,768	40,971
GA	36,468	44,623
HI	40,162	46,786
IA	36,522	47,202

Tables and data files



Clear picture of the audience

Clear picture of how they will be using the data

Clear design and consistency

Reduce clutter

## Appendix E: Examples of Engaging Graphics

### E.1. Provide an Entry point

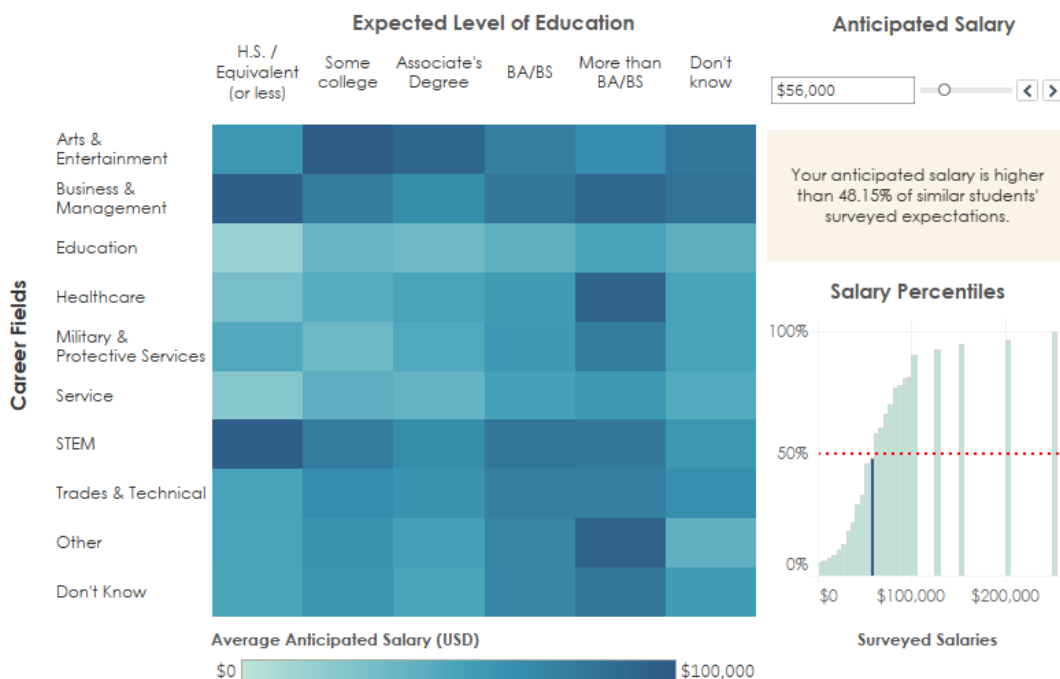
The screenshots below are from an interactive graphic created from data used for the NCES Stats in Brief “Education, Employment, and Earnings Expectation of 2009 9<sup>th</sup> Graders” in 2016.” It invites the viewer to start by entering their own expected salary. Prior to their own entry, however, the view provided is for a pre-selected salary value, to help the user understand the range of information available.

As a second entry point, the user may enter his or her own anticipated occupation or level of education. The display will provide the relative standing of the viewer’s own salary expectation among all others with those same occupation/education levels.

These graphics excerpts produced in Tableau, from Bryant, K. and Moriondo, D. (2021) NISS SAID in Graphics Data Visualization Competition entry.

### Education, Employment & Earnings: Comparing Expectations

Enter your anticipated salary in the box below to see how your expectations compare to surveyed students. Explore career options by selecting career fields and degree goals.

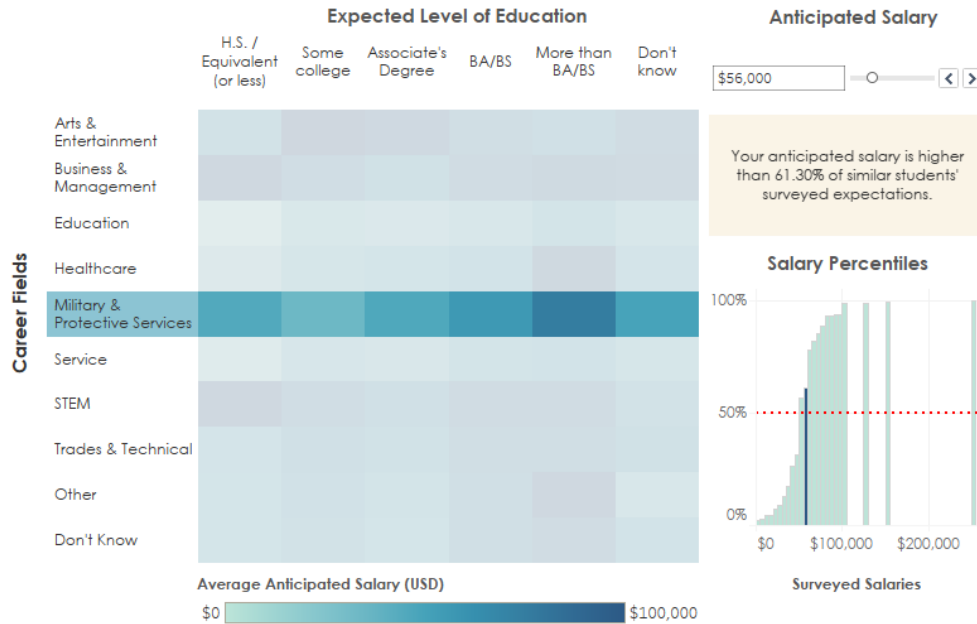


**Entry Point** – An anticipated salary is pre-selected and the indicated percentile and place in distribution are over the entire sample.

## Innovative Graphics

### Education, Employment & Earnings: Comparing Expectations

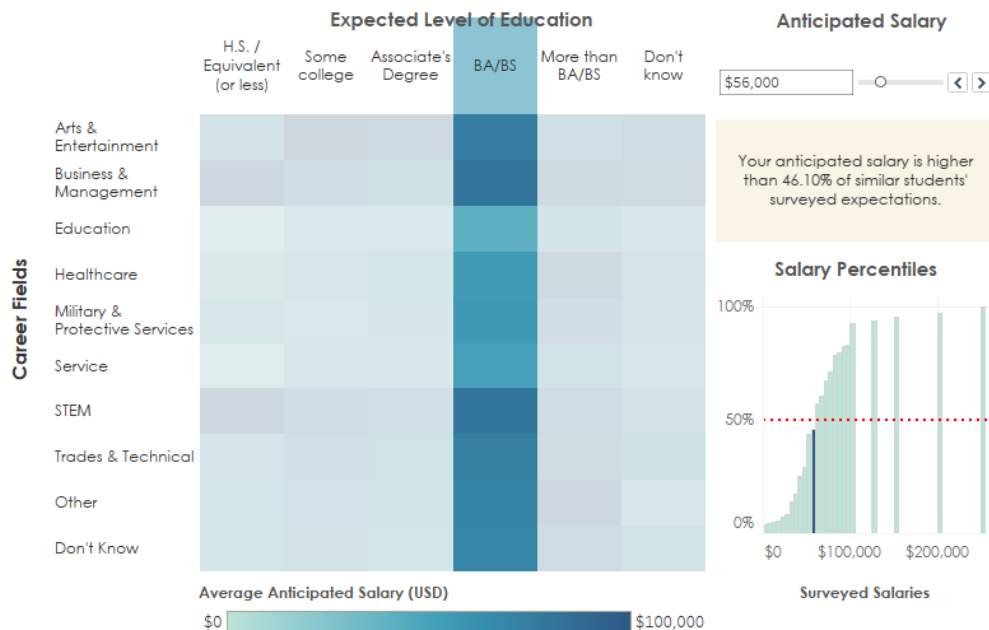
Enter your anticipated salary in the box below to see how your expectations compare to surveyed students. Explore career options by selecting career fields and degree goals.



**User Choice of Career** – The user is invited to choose their chosen career field and/or education level. Selecting “Military & Protective Services” shows that the percentile for the anticipated salary is higher, suggesting that expectations for this career field are lower than overall.

### Education, Employment & Earnings: Comparing Expectations

Enter your anticipated salary in the box below to see how your expectations compare to surveyed students. Explore career options by selecting career fields and degree goals.

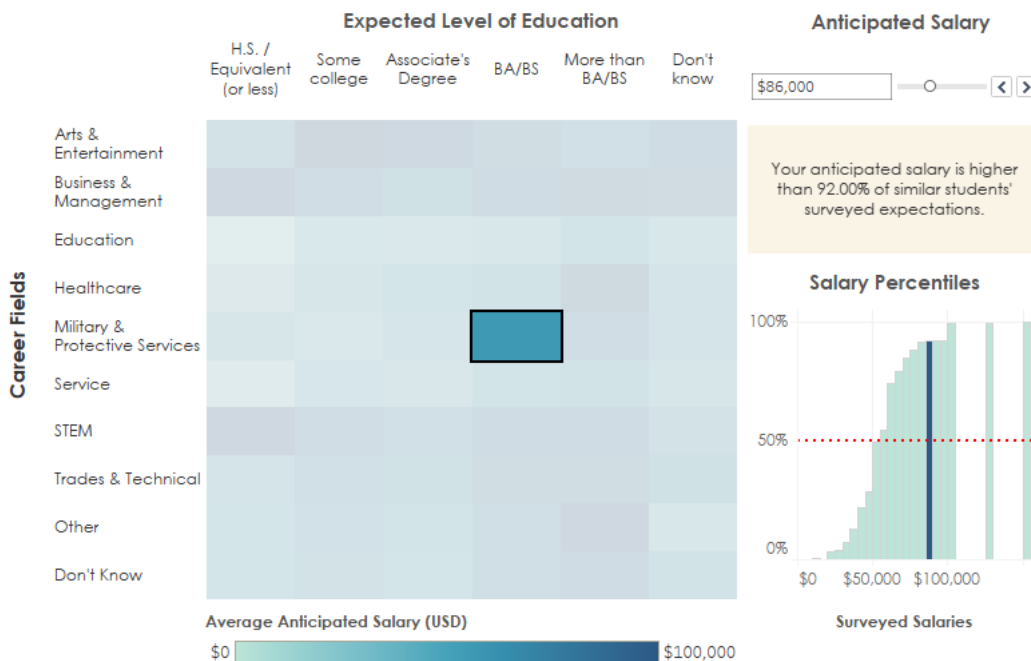


**User Choice of Degree**

## Innovative Graphics

### Education, Employment & Earnings: Comparing Expectations

Enter your anticipated salary in the box below to see how your expectations compare to surveyed students. Explore career options by selecting career fields and degree goals.



#### User Choice of Career, Degree, and Target Salary

#### E.2. Provide Explanations

The graphic below is called a micromap. It exemplifies several of the principles for engaging graphics discussed in Section 5. First, because of its complexity, it is the type of graphic that explanations can make more accessible for viewers (see the attached video *MicromapNarration.mp4*). A micromap is likely to be unfamiliar to most of IES's audiences and requires more visual work from the viewer. Cues on the graphic itself can ease the task.

The micromap is also an example of a small multiple graphic. Small multiples are designed to ease the task of comparisons, and can also provide an entrypoint (for example the viewer's own state or gender), that provides a basis for interest and a framework for providing opinions.

Finally, the graphic below is also an example of a pyramid method of display; i.e., a graphical display that provides useful information to a variety of audiences. For example, parents may simply be interested in how their state's performance compares to other states. This is easy to find with the map cue. More technical audiences may need to perform inferences, which the confidence intervals provide.

Micromaps can also be made interactive or linked, and software is available for implementation. Here are examples of R-packages that can produce micromaps.

Payton, Q.C., McManus, M.G., Weber, M.H., Olsen, A.R. and Kinaid, T.M. (2015) "[micromap: A package for linked micromaps](#)," *Journal of Statistical Software*, 63(2),1-16.

Pickle, L.W., Pearson, J.B., Carr, D., (2015) "[micromapST: Exploring and Communicating Geospatial Patterns in US State Data](#)," *Journal of Statistical Software*, 63(3),1-25.

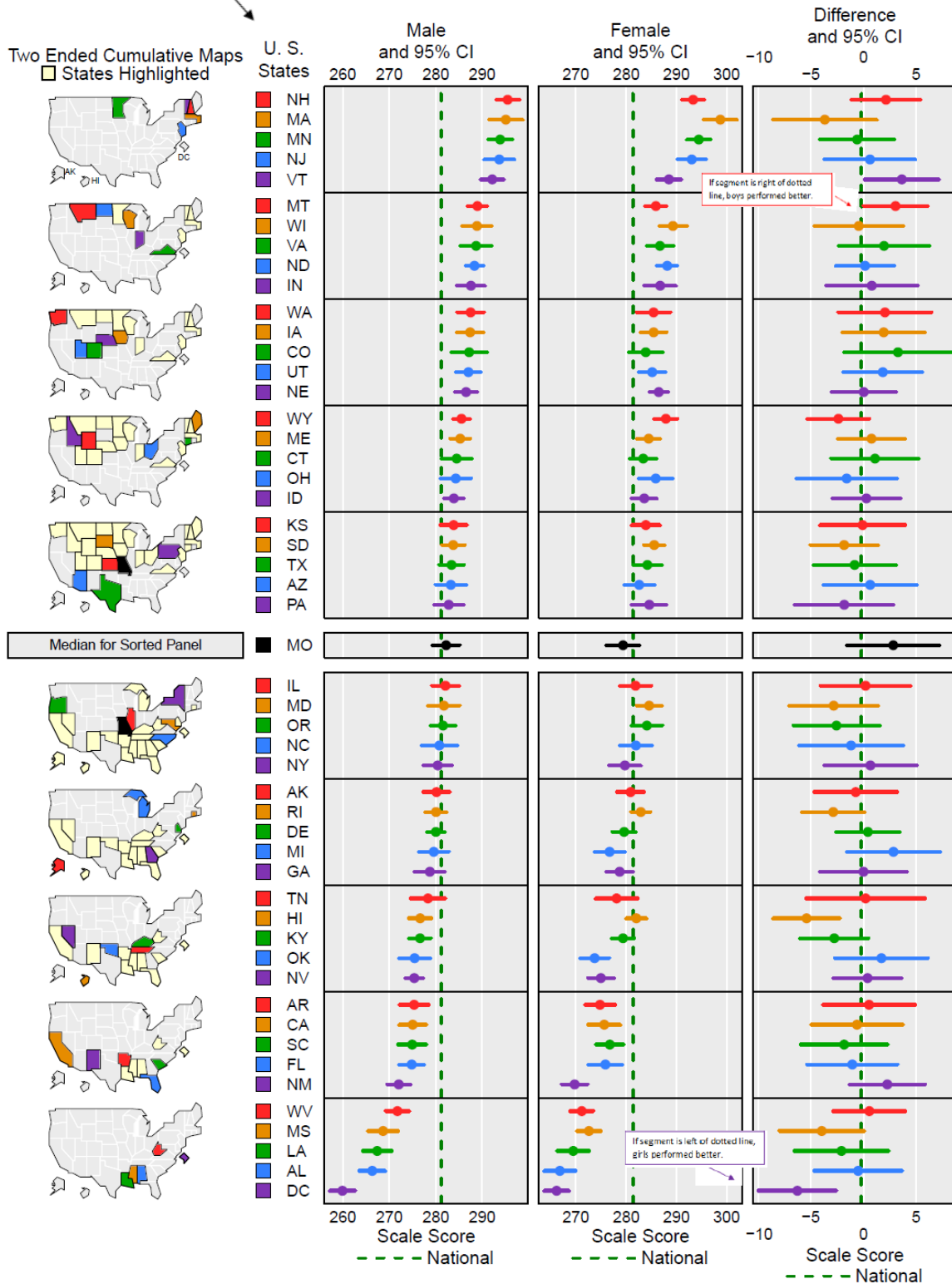
More discussion of micromaps is available in:

Carr, D. B. and L.W. Pickle. *Visualizing Data Patterns with Micromaps* (2010), Chapman & Hall/CRC Press.



## The Nations Report Card for 2015 Gender and 4th Grade Math Scores

States are ranked from best to worst on how boys performed in 4<sup>th</sup> grade math.





## Innovative Graphics

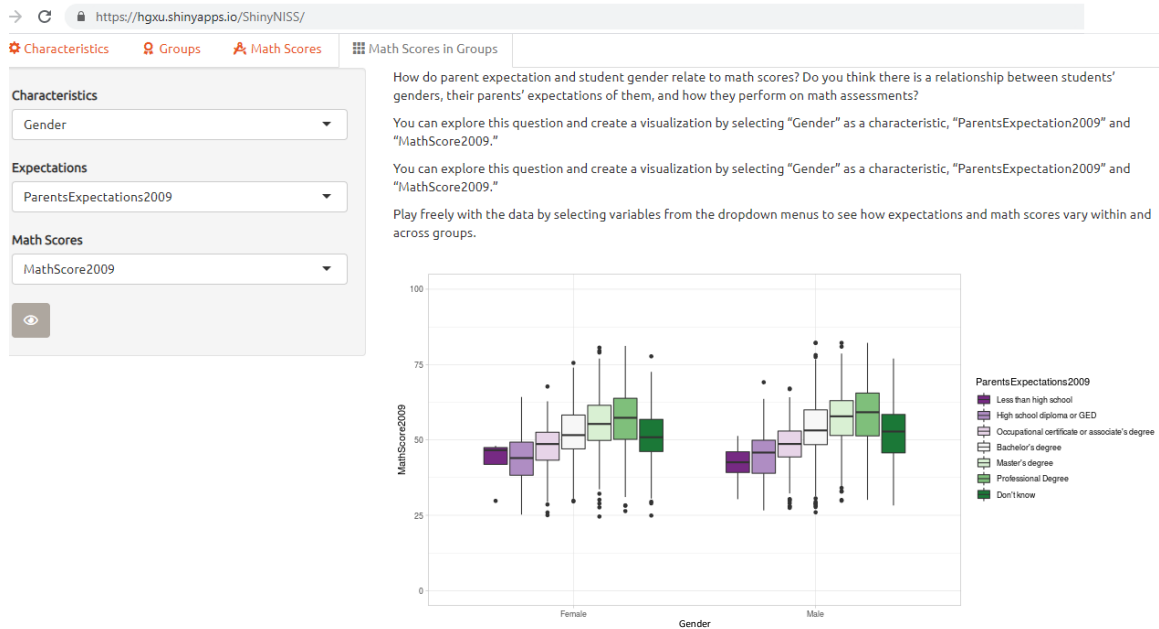
### E.3. Provide a Story

The screenshot below is based on data from NCES Stats in Brief “Education, Employment, and Earnings Expectation of 2009 9<sup>th</sup> Graders” in 2016.” The simple graphic supports the following story:

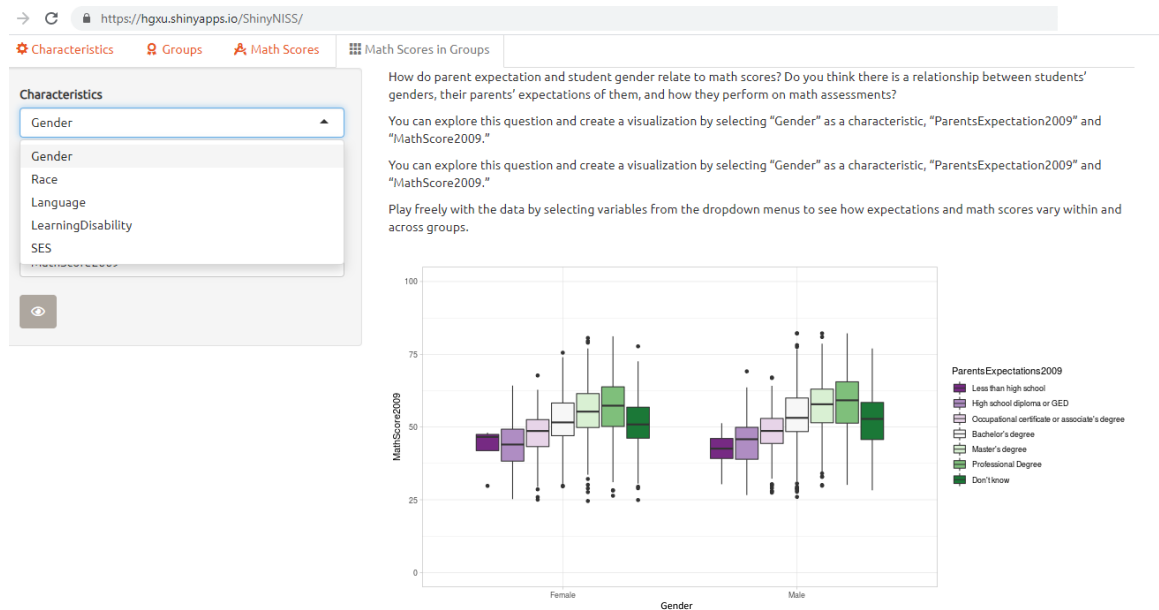


## Innovative Graphics

Another way to draw a viewer into a graphic is by leading them to discover the story. In the screenshots below, this interactive graphic first provides instruction to the viewer on how to answer a specific question about parent's expectation for their children based on gender:



Then the viewer is invited to explore and discover their own story by using the options provided:



These graphics excerpts produced in RShiny, from Haji-Georgi, M., Treis, F., Rosca, O., Yu, C.-Y., and Xu, X. (2021) NISS SAID in Graphics Data Visualization Competition entry.

### E.4. Provide Graphics that Allow Comparisons

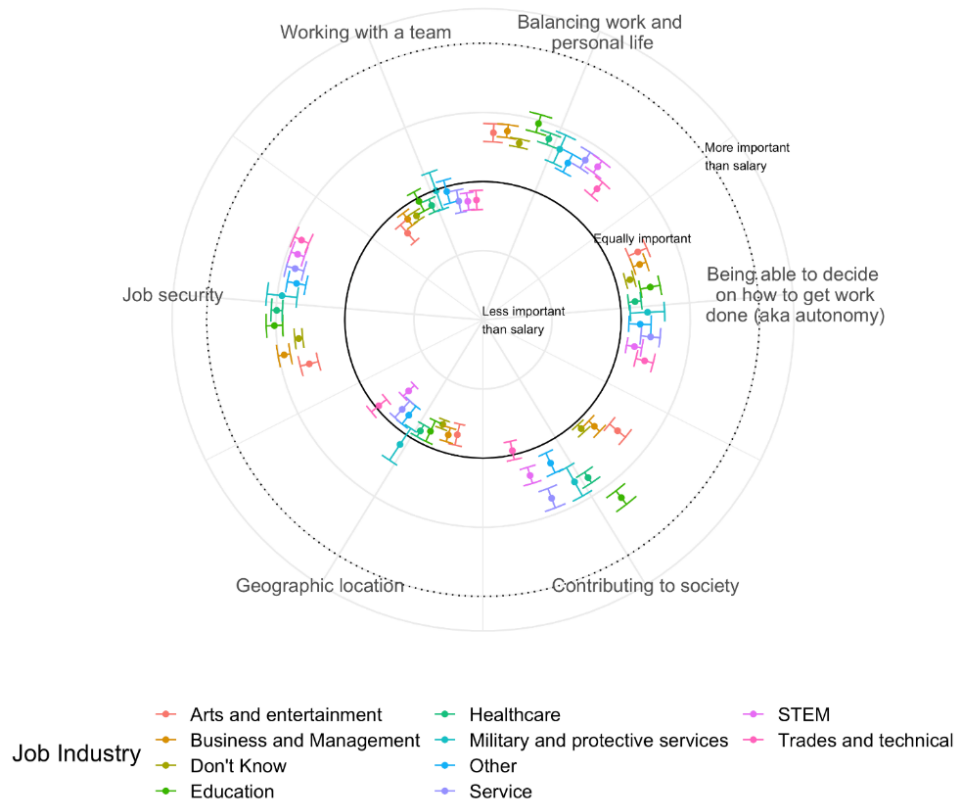
Simple charts that are widely available in commonly used software such as R or Tableau are very useful for allowing the viewer to make comparisons, even among several variables at a time. Side-by-side boxplots make it easy to compare three variables at a time, even without interactivity. For example, the screenshots

## Innovative Graphics

of the boxplots in Appendix E.3 allow comparisons of Parents Expectations and Math Scores by Gender. Comparisons by others demographic groups are available on the dropdown menu.

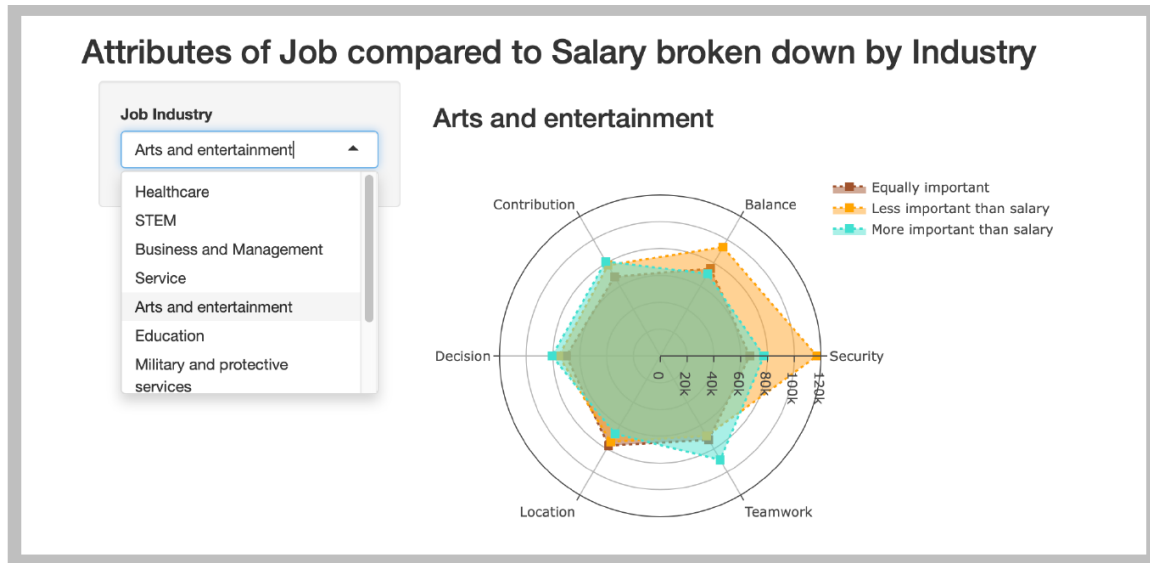
The chart below is another simple and visually appealing graphic, known as a radar chart, that is also useful for making comparisons of several variables simultaneously. It also displays three variables from the “Education, Employment, and Earnings Expectation of 2009 9<sup>th</sup> Graders” study: Students’ expected job/industry; how important various aspects of the job are to them; and whether or not that aspect is more or less important to them than salary. The graphic also encodes uncertainty measures in the form of confidence intervals for each proportion, though these are subtle enough to be ignorable to those for whom this is not of interest.

This graphic allows the viewer to immediately see that “Contributing to society” varies in importance by industry, and is seen as much more important than salary to some (e.g., Education fields) than others (e.g., Trades and Technical). On the other hand, job security is seen as more important than salary almost uniformly by field (except those in the arts).

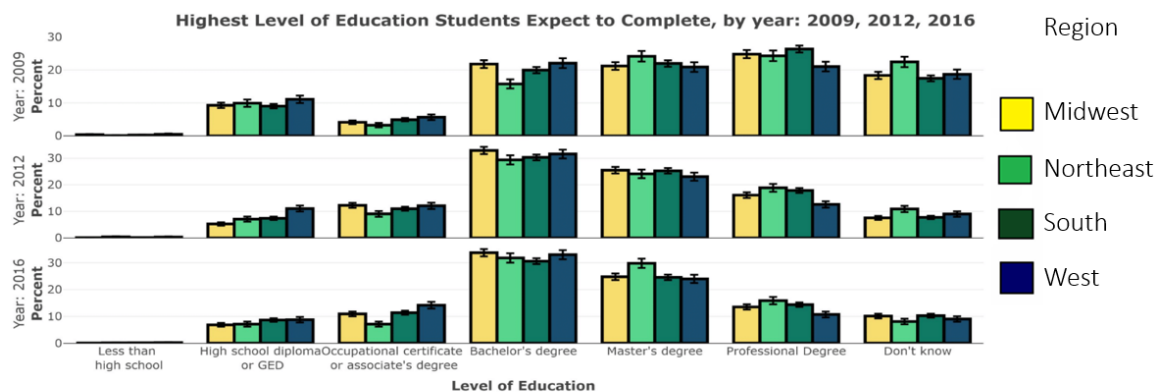


This graphic excerpt produced in Rshiny, from Yarger, A. (2021) NISS SAID in Graphics Data Visualization Competition entry

Another radar chart of the same data, shown below, encodes a fourth variable, Salary. This makes it both more interesting and still attractive, but requires a little more effort to interpret.



This graphic excerpt produced in RShiny, from An, H. (2021) NISS SAID in Graphics Data Visualization Competition entry. Finally, the graphic below is composed of what may be the simplest of all charts, which are bar charts of student expectations for their highest educational outcome. However, because they are small multiples, they allow easy comparisons of these outcomes across two additional dimensions: time and geography.



This graph was constructed in Python.

This graphic excerpt produced in RShiny, from Ozment, S., Becker, R., Edison, A., and Hunnicutt, J. (2021) NISS SAID in Graphics Data Visualization Competition entry.

## E.5. Implementing the Guidelines: An NCES Example

An example using NCES data seems especially useful to demonstrate how there is not necessarily a single solution to any data visualization challenge. The National Teacher and Principal Survey (NTPS) from NCES provides data on public and private elementary and secondary schools across the country. As part of our work, we explored the [average public school starting time](#) across each U.S. states. NCES publishes these data as a single long table, which may be useful for data analysts and researchers, but is likely less useful for parents and elementary- and secondary-school educators.

# Innovative Graphics

Average public school start time and percentage distribution of school start time, by state: 2017–18

State	Average a.m. start time	Percentage distribution of start times				
		Before 7:30 a.m.	7:30 a.m. to 7:59 a.m.	8:00 a.m. to 8:29 a.m.	8:30 a.m. to 8:59 a.m.	9:00 a.m. or later
United States	8:10	4.3	27.5	40.3	19.8	8.2
Alabama	7:49	12.9	49.8	31.4	5.9 <sup>†</sup>	#
Alaska	8:30	#	8.2 <sup>†</sup>	29.9	34.9	27.0
Arizona	8:14	5.4 <sup>†</sup>	24.1	41.7	21.3	7.5 <sup>†</sup>
Arkansas	7:56	#	54.2	45.2	3.4 <sup>†</sup>	#
California	8:10	#	19.5	58.7	18.1	3.3
Colorado	8:07	2.8 <sup>†</sup>	25.8	48.4	19.0	3.9 <sup>†</sup>
Connecticut	8:31	5.1	22.9	24.1	33.8	14.1 <sup>†</sup>
Delaware	8:06	11.8	25.9	27.8	27.7	6.9 <sup>†</sup>
District of Columbia	8:30	#	5.2 <sup>†</sup>	25.5	64.5	4.8 <sup>†</sup>
Florida	8:15	8.0	18.4	35.0	22.4	16.2
Georgia	8:03	5.1 <sup>†</sup>	40.2	35.2	13.2	6.3
Hawaii	7:58	#	54.0	40.3	#	#
Idaho	8:19	#	9.9	56.7	21.8	11.6
Illinois	8:16	#	18.5	46.1	27.6	7.3
Indiana	8:07	6.6 <sup>†</sup>	25.7	47.2	13.3	7.3 <sup>†</sup>
Iowa	8:18	#	12.2	61.1	23.6	#
Kansas	8:11	#	20.6	57.8	18.9	#
Kentucky	8:06	#	42.5	33.4	8.0 <sup>†</sup>	13.0 <sup>†</sup>
Louisiana	7:53	16.2	46.5	25.3	10.8 <sup>†</sup>	#
Maine	8:03	#	33.0	39.3	20.1	#
Maryland	#	#	#	#	#	#
Massachusetts	8:14	7.2	30.6	21.8	28.0	12.4 <sup>†</sup>
Michigan	8:08	3.0 <sup>†</sup>	30.6	38.6	24.4	3.3
Minnesota	8:26	#	14.8	36.3	31.5	16.9
Mississippi	7:48	8.3 <sup>†</sup>	60.8	27.3	#	#
Missouri	8:09	4.4 <sup>†</sup>	25.3	45.8	19.1	5.5 <sup>†</sup>
Montana	8:18	#	#	79.2	15.2	#
Nebraska	8:07	#	15.2	74.8	7.8 <sup>†</sup>	#
Nevada	8:20	9.9	17.4	26.3	20.8	25.6
New Hampshire	8:06	7.3	33.6	26.8	27.4	4.9 <sup>†</sup>
New Jersey	8:18	3.5 <sup>†</sup>	20.0	32.9	39.0	4.5 <sup>†</sup>
New Mexico	8:01	2.4 <sup>†</sup>	43.6	41.1	12.2	#
New York	8:15	5.5	21.4	37.3	23.6	12.1
North Carolina	8:00	9.3	38.9	32.5	15.1	4.2 <sup>†</sup>
North Dakota	8:24	#	8.8 <sup>†</sup>	39.6	50.6	#
Ohio	8:15	7.1	27.5	24.2	17.8	23.4
Oklahoma	8:08	#	21.6	56.2	14.1	6.4 <sup>†</sup>
Oregon	8:15	#	25.1	46.4	17.9	10.0 <sup>†</sup>
Pennsylvania	8:14	5.5	28.1	24.0	28.8	13.6
Rhode Island	8:21	7.4 <sup>†</sup>	12.3	33.2	30.5	16.5
South Carolina	8:00	8.5 <sup>†</sup>	39.7	31.6	17.3	#
South Dakota	8:17	#	2.2 <sup>†</sup>	64.2	32.2	#
Tennessee	8:04	9.1	31.4	36.1	14.3	9.2 <sup>†</sup>
Texas	7:59	4.9	45.6	37.6	8.7	3.2
Utah	8:20	#	17.4	45.7	22.8	14.2
Vermont	7:59	#	36.5	56.1	#	#
Virginia	8:23	1.3 <sup>†</sup>	16.5	40.3	25.3	16.7
Washington	8:25	#	17.5	31.7	27.4	22.4
West Virginia	7:53	14.4	35.1	40.9	9.6	#
Wisconsin	8:06	2.3 <sup>†</sup>	26.6	50.9	20.1	#
Wyoming	8:12	#	23.5	44.5	29.6	2.5 <sup>†</sup>

# Rounds to zero.

<sup>†</sup> Interpret data with caution. The coefficient of variation (CV) for this estimate is between 30 percent and 50 percent (i.e., the standard error is at least 30 percent and less than 50 percent of the estimate).

<sup>‡</sup> Reporting standards not met. The coefficient of variation (CV) for this estimate is 50 percent or greater (i.e., the standard error is 50 percent or more of the estimate) or the response rate is below 50 percent.

NOTE: Detail may not sum to totals because of rounding and because some data are not shown.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Teacher and Principal Survey (NTPS), "Preliminary Public School Restricted-Use Data File," 2017–18.

For each state, the table includes the overall average start time and the distribution of schools within 29-minute bands (i.e., 7:30 a.m.–7:59 a.m.; 8:00 a.m.–8:29 a.m.; and so forth). In this post, we show four different ways to visualize the average starting time, all constructed in the Microsoft Excel software package.

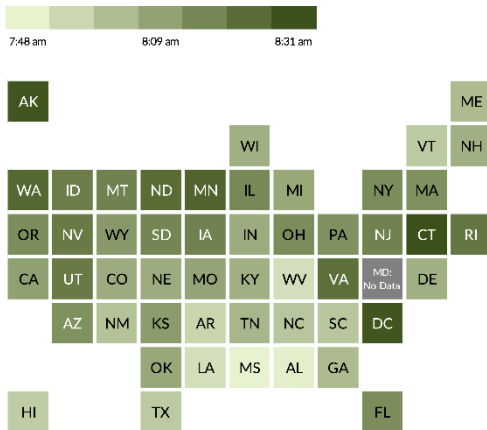
## Choropleth and Tile Grid Maps

As is probably the case with many people who are visualizing geographic data, the first instinct is likely to create a map. It's a fairly basic approach and can be combined with other chart types, such as a bar chart or even a line chart to show changes over time. One of the challenges with maps, however, is that it can be difficult to extract exact differences between the values or even to quickly see the highest value (here it is Connecticut which starts at 8:31 a.m., with Alaska and DC at 8:30) and lowest values (Mississippi at 7:48 a.m., with Alabama at 7:49), but the familiarity of the map can offset those concerns.

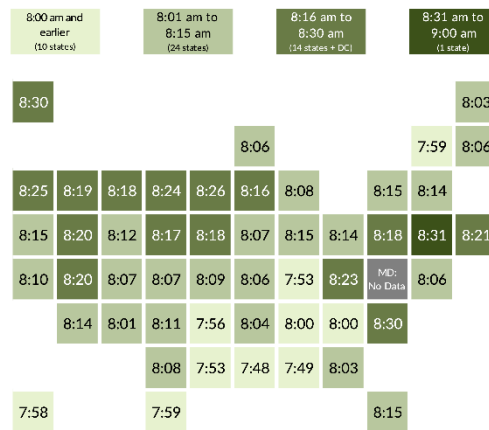
In this tile grid map, which uses a single square for states, is a tradeoff between geographic accuracy (the states are placed in arbitrary positions) and geographic distortions caused by the typical data-driven map (that is, the size of the state may not correspond to its importance in the data).

# Innovative Graphics

Average public school start time



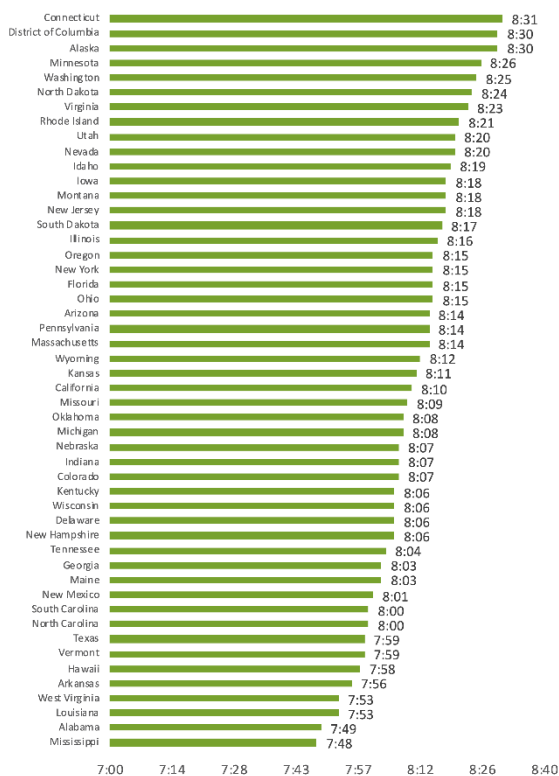
Average public school start time



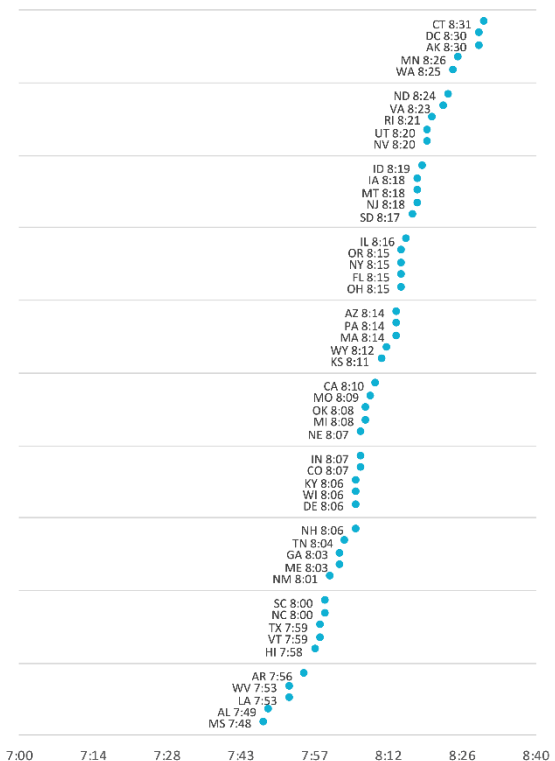
## The Dot Plot

We could create a standard bar chart with these data (on the left), but instead a dot plot (sometimes called a dumbbell chart, barbell chart, or gap chart; shown on the right) can be used, which reduces the amount of ink on the page and can enable clearer labeling. The data values correspond to one axis and the groups to the other, which do not necessarily need to be ordered in a specific way, though sorting can help.

Average public school start time



Average public school start time  
(with earliest and latest start times)



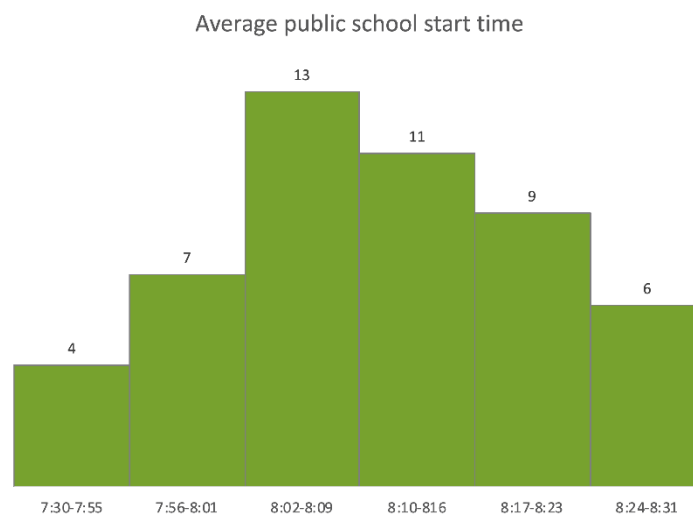
## Innovative Graphics

When bar charts include too many bars, the graph looks cluttered. By contrast, the dot plot shows the same data with a dot at each data value connected by a line to show the range or difference. The circles use less ink than the bars, which lightens the visual by adding more empty space.

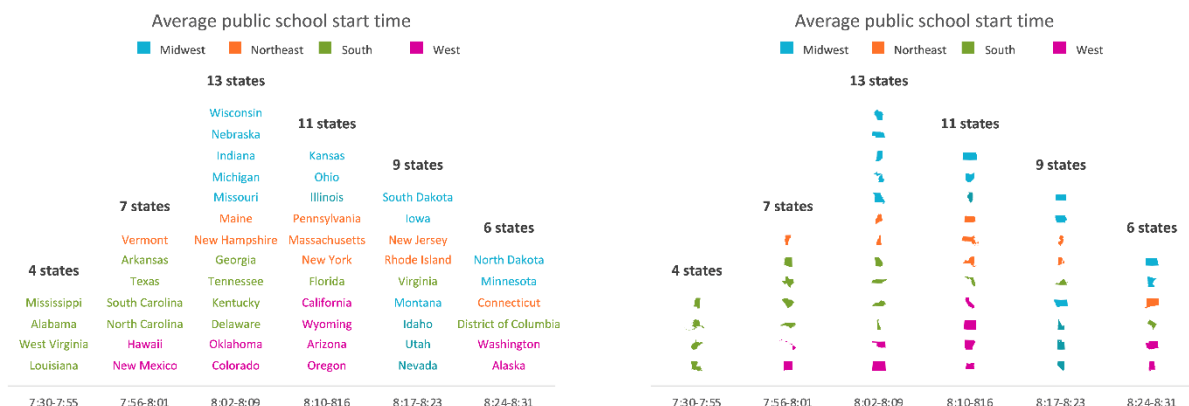
### The Histogram

The histogram is the most basic graph type for visualizing distributions. It is a kind of bar chart that presents the tabulated frequency of data over distinct intervals, called bins, that sum to the total distribution. The entire sample is divided into these bins, and the height of each bar shows the number or proportion of observations in each interval. Histograms can show where values are concentrated within a distribution, where extreme values are, and whether there are any gaps or unusual values.

In this case, a basic histogram might look like this.



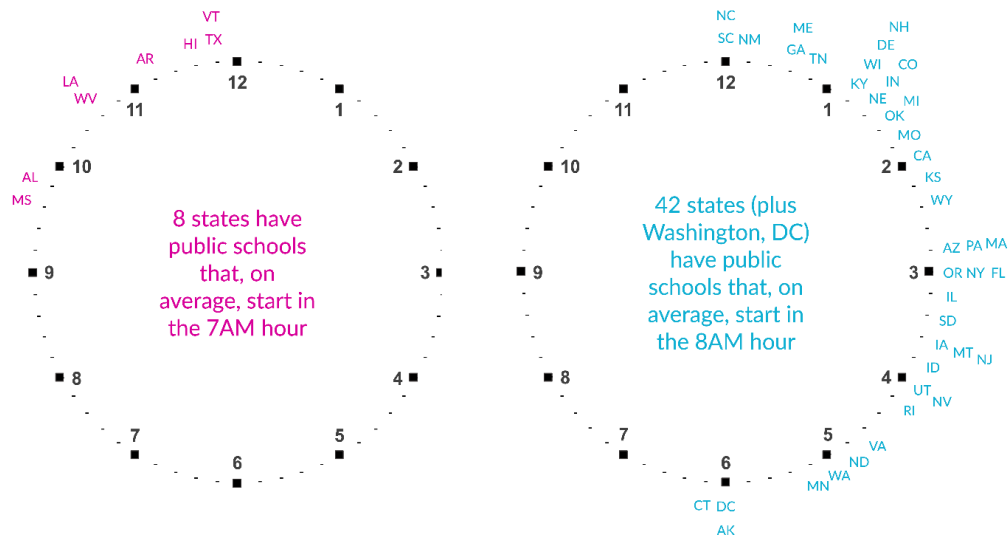
This graph can be modified in various ways, both to make it more informative and more interesting. Instead of just using bars—which aggregate the data into an abstract shape—each state name could be listed separately. The version on the left employs this approach and adds color to each state name based on its region of the country. The version on the right uses a similar approach but uses icons (using [ProPublica's StateFace font](#)) instead of text.



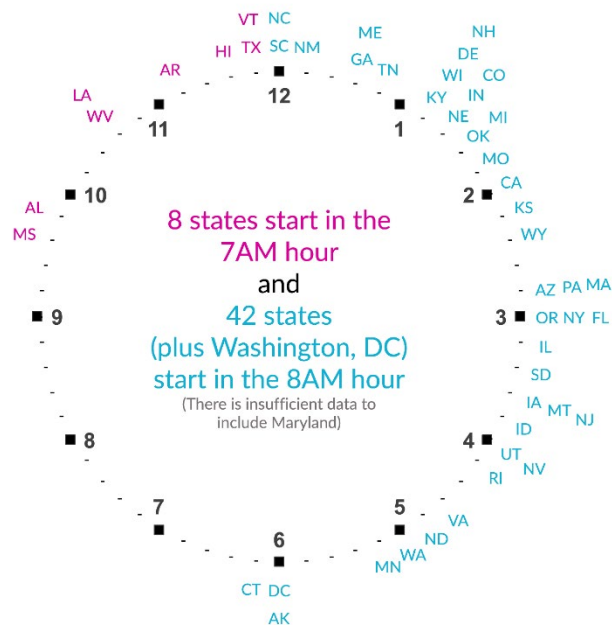
## Innovative Graphics

### The Clock

One additional option is to embrace the idea of *time* in the data and utilize the familiar shape of a clock. The state abbreviations could be listed in two separate clocks (top) or around a single clock (bottom). These visualizations are built in the Microsoft Excel software package and can thus be created and shared by nearly all users.



All three clocks show the minute hand of starting times for each state.



To sum up, even with a relatively simple data set like the average starting time of public schools around the country, there are many visualization options. None of these are necessarily right or wrong, but each may be more appropriate for a certain audience or for communicating a certain level of detail from the basic story to the precise data.



### Appendix F: Expert Panel Biosketches

**Daniel B. Carr, Ph.D.**

*Title: Professor of Statistics at George Mason University and a Fellow of the American Statistical Association*  
Daniel Carr's expertise in the field of statistical graphics has led to ground-breaking work on statistical graphics and maps. At Pacific Northwest National Laboratory in the 1980s he was a member of the analysis of large data sets team that produced interactive and dynamic graphics that including stereo graphics. Since joining GMU in 1990, he has statistical graphics for classes at different levels, reaching over a thousand students; and he has directed eleven Ph.D. dissertations addressing statistical graphics. He has held many research grants in the area of statistical graphics, collaborated with more than a dozen federal agencies and served on federal agency advisory boards. His work with Linda Pickle and others while on sabbatical at NIH, was recognized with a National Institutes of Health Award of Merit that said: *For exceptional scientific leadership in the conceptualization and development of a web-based cancer control planning tool, State Cancer Profiles*. His book with Linda entitled *Visualizing Data Patterns with Micromaps* is a standard in the field.

**Nola du Toit, Ph.D.**

*Title: Senior Research Methodologist and Data Visualization Specialist at NORC at the University of Chicago*  
Nola Dutoit uses methods of data and information visualization to support the research process and enhance the value of data findings. Data means nothing if people cannot understand it. Her goal is to shed light on analytical findings and make data and research accessible to more people. Nola's expertise draws upon more than a decade of experience in data collection, survey research, and analysis, as well as both qualitative and quantitative methods. The visuals and graphics that she creates explore and illustrate concepts, ideas, and data across a broad range of areas of interest, including public health, education, economics, and public opinion.

**Miriah Meyer, Ph.D.**

*Title: Associate Professor, School of Computing at the University of Utah; and faculty member in Scientific Computing and Imaging Institute*

Miriah Meyer also co-directs the Visualization Design Lab, which focuses on the design of visualization systems for helping people make sense of complex data, and on the development of methods for helping visualization designers make sense of the world. Prior to joining the faculty at University of Utah, Miriah was a postdoctoral research fellow at Harvard University and a visiting scientist at the Broad Institute of MIT and Harvard.

Miriah is the recipient of multiple awards and grants, including an NSF CAREER grant, a Microsoft Research Faculty Fellowship, and a NSF/CRA Computing Innovation Fellow award. She was named a University of Utah Distinguished Alumni, both a TED Fellow and a PopTech Science Fellow, and included on MIT Technology Review's TR35 list of the top young innovators. She has served as the papers chair for IEEE VIS, an associate editor for IEEE TVCG, and as a member of the Visualization Executive Committee.

## Innovative Graphics

### **Jonathan Schwabish, Ph.D.**

*Title: Senior Fellow, Urban Institute*

Jonathan Schwabish is an economist, writer, teacher, and creator of policy-relevant data visualizations. His research agenda includes disability insurance, retirement security, data measurement, and the Supplemental Nutrition Assistance Program. He is also the founder of PolicyViz, a consulting firm that helps clients improve how they work with and communicate data and analysis. Dr. Schwabish is considered a leading voice for clarity and accessibility in how researchers communicate their findings. His book [\*Better Presentations: A Guide for Scholars, Researchers, and Wonks\*](#) helps people improve the way they prepare, design, and deliver data-rich content and his edited book, [\*Elevated the Debate: A Multilayered Approach to Communicating Your Research\*](#), helps people develop a strategic plan to communicating their work across multiple platforms and channels. His latest book, [\*Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks\*](#) details essential strategies to create more effective data visualizations.

### **Leland Wilkinson, Ph.D.**

*Title: Chief Scientist at H2O and Adjunct Professor of Computer Science at the University of Illinois Chicago*

Leland Wilkinson has a long history of creating software for statistical computing and publication of research and development of visualization systems. First known for writing the SYSTAT statistical package and founding SYSTAT Inc. in 1984, he also holds patents in visualization and distributed analytic computing. He is the author (with Grant Blank and Chris Gruber) of *Desktop Data Analysis with SYSTAT*, and *The Grammar of Graphics*, the foundation for several commercial and opensource visualization systems (IBM-RAVE, Tableau, ggplot2, and Python-Bokeh).

Wilkinson is a Fellow of the American Statistical Association, an elected member of the International Statistical Institute, and a Fellow of the American Association for the Advancement of Science. He has won best speaker award at the National Computer Graphics Association and the Youden prize for best expository paper in the statistics journal *Technometrics*. He has served on the Committee on Applied and Theoretical Statistics of the National Research Council and is a member of the Boards of the National Institute of Statistical Sciences (NISS) and the Institute for Pure and Applied Mathematics (IPAM).

### ***Panel convened by National Institute of Statistical Sciences***

### **S. Lynne Stokes, Ph.D.**

*Title: Professor, Department of Statistical Science and Director, Data Science Institute, Southern Methodist University, and Senior Fellow at NISS*

Lynne Stokes is an expert in surveys, polls and sampling, as well as in non-sampling survey errors, such as errors by interviewers and respondents. She is a Fellow of the American Statistical Association and a recipient of ASA's Founder's Award. She recently has conducted research on evaluating the accuracy of contest judges and on improving estimates of marine fishery yields by the National Oceanic and Atmospheric Administration. She also serves on a technical advisory committee for the National Assessment of Educational Progress, or "Nation's Report Card." Stokes served on the faculty at Vanderbilt and University of Texas at Austin, as well as in government positions with the U.S. Fish and Wildlife Service and the Census Bureau. She recently became the inaugural Director of the new Data Science Institute at SMU.

## Innovative Graphics

Research Interests: Surveys, Polling and Sampling, Voter Exit Polling, Sampling Methods, Non-Sampling Errors, Non-Disclosure Methodology, Measurement Error, Order Statistics, and Mark and Recapture Methods.

### **Brian Habing, Ph.D**

*Title: Associate Director for Education Research at NISS, and Associate Professor of Statistics at University of South Carolina*

Brian Habing is Associate Director for Education Activities and Research working with the DC Office of the National Institute of Statistical Sciences (NISS) and Associate Professor of Statistics at the University of South Carolina. His research has focused on psychometrics and scale construction, with a particular emphasis on multidimensional item response theory. His research focus includes analysis of education statistics and his interests also extend to statistical education, including work with AP Statistics and the development of new courses at the undergraduate and graduate level.

### **Alexi Albert**

*Title: Research Assistant, National Institute of Statistical Science*

Alexi Albert is a Research Assistant at NISS, working with the National Center for Education Statistics (NCES). She is currently completing her MS in Mathematics and Statistics from Georgetown University. Prior to graduate school she taught sixth grade math for two years in Eastern Washington State.

### **Ya Mo, Ph.D.**

*Title: Research Fellow, National Institute of Statistical Sciences; Assistant Professor, Boise State University*

Dr. Ya Mo is a research fellow at the National Institute of Statistical Sciences and an assistant professor of Curriculum, Instruction, and Foundational Studies at Boise State University. She received a dual major Ph.D. in Measurement and Quantitative Methods and Curriculum, Instruction, and Educational Policy Programs, and an M.S. in Statistics from Michigan State University, as well as an Ed.M in TESOL from Boston University. She researches quantitative methods, psychometric measures, and survey statistics; she also applies quantitative research methods to study substantive topics in education, especially large-scale assessments.