Creating an Interactive Visualization of General Social Survey Data: From Start to Finished Product

Nola du Toit, Edward Mulrow[1], Rene Bautista, and Mu-Hsien Lee

NORC at the University of Chicago, 55 E. Monroe, Suite 31, Chicago, IL 60603

[1]NORC at the University of Chicago, 4350 East-West Highway, F8, Bethesda MD 20814

Abstract

"It takes a village" to create a successful interactive data visualization. We draw upon our recent experience creating an interactive graphic for the General Social Survey (GSS) to examine the different roles and perspectives that come together in making the visualization a success. For decades, the GSS has examined social change and complexity in the United States. The interactive graphic allows users to delve into the data and examine how the attitude of Americans towards government spending has changed over time. Roles involved in producing the graphic include the subject matter expert, who contributed an understanding of the research purpose, as well as the visualization expert and programmer who designed and implemented the interactive tool. In addition, the graphic designer contributed visual balance and appeal to the graphic, while the writer aided with interpretation and messaging. Finally, the statistician worked to ensure that the visualization accurately reflected the underlying data. We present the benefits and obstacles that go along with this expanded team that ultimately resulted in a comprehensive data visualization experience for users of the GSS.

Key Words: Interactive, data graphic, General Social Survey.

1. Creating an Interactive

Data graphics and visualizations are becoming increasingly common ways of presenting information and telling data stories. Specifically, interactive graphics present users with an engaging way to examine and interact with the data. However, the process of creating an effective interactive can be complicated and often involves many different perspectives. In this paper, we present our experiences in creating an effect interactive graphic. We detail the different aspects of the process that should be considered in order to make the graphic present a good data story in an appealing and user friendly format.

2. Background

NORC at the University of Chicago is a non-profit, nonpartisan research institution that collects and analyzes objective and reliable data to help inform policy and business decisions. NORC works with clients from the government and private sectors, and covers a range of research areas, including Health and Well-being; Economics, Markets, and the Workforce; Education, Training, and Learning; Global Development; and Society, Media and Public Affairs.

Every two years, NORC staff are engaged in creating a biennial report. The report presents clients and stakeholders, as well as the general public, with highlights of projects and NORC studies. Staff from around NORC are tasked with working on the report. The work brings together subject matter experts, web designers, writers and editors, graphic

designers, statisticians, methodologists, and communications experts. The report process creates a microcosm in which these sometimes disparate groups needs to work together for one purpose: to create an engaging and informative report.

The report also showcases graphics created using data collected by NORC (read the most recent report: https://biennialreport.norc.org/). In the past, the graphics were static as the report was published in hardcopy and pdf formats. Recently, however, the report moved to an online version, making possible the addition of interactive data graphics.

One of NORC's flagship projects is the General Social Survey (GSS). The GSS collects attitudinal data from U.S adults in order to track societal complexity and change over time. NORC has been collecting data for the GSS since 1972 (about once every two years; see the project website for more information: https://gssdataexplorer.norc.org/). Considering the wealth of data captured by the GSS, the staff on the biennial report proposed creating an interactive graphic to showcase the GSS.

2. The Process

In order to pull together a successful data story and interactive on the GSS, the report team had to draw upon people with various expertise, including the Data Visualization Expert, the Subject Matter Expert, the Statistician, the Graphic Designer, the Programmer, and the Writer.

2.1 Creating the idea: The Data Visualization Expert

The Data Visualization Expert was tasked with the design of an interactive graphic using GSS data. This part of the process involved examining the data and coming up with a data story that was appealing to the report's various readers. In order to perform this task, the expert had to have a solid background in data, as well as a good knowledge of best practices for creating graphics. Drawing upon sources that provide guidance on best practices for data graphics, including, for example, Robbins (2013), Cleveland and McGill (1984, 1987), and Cairo (2019), the Data Visualization Expert had to consider the scope of the GSS data and the range of graphics that would be appropriate. This allowed the expert to come up with potential ideas for the interactive. However, while well-versed in the best graphics for the data, the Data Visualization Expert was not the most knowledgeable person about the GSS. Consequently, in order to create the best interactive, the Data Visualization Expert had to engage the Subject Matter Expert.

2.2 Knowledge about the topic: The Subject Matter Expert

The Subject Matter Expert brings to the team a wealth of knowledge about the data. This expert has worked with the data for years and is familiar with its scale and limitations. Moreover, the expert has produced many data products and reports that can be used to inform the interactive.

One such report outlined changes in the national spending priorities of US adults. The GSS includes questions on government spending. It asks respondents if government spending on a variety of topics is too much, too little or about right. There are over 20 topics including, for example, space exploration, welfare, foreign aid, science, social security, and so on. Using these data, the Subject Matter Expert created a measure for net spending. That is, the difference between the percent saying "too little" and "too much" is being spent.

This created a relative ranking of the topics within year. The relative ranking of items produced an intriguing data story that shows how attitudes towards these topics have changed in relative importance over time (see Figure 1). The team decided to focus on this information in the interactive.

Figure 1: Example data from the GSS report on government spending

C. Improving and Protecting the Nation's Health/Health							
	1973	1974	1975	1976	1977	1978	1980	1982
Too little	63.4	66.9	65.6	63.0	59.2	57.2	57.9	59.3
About right	31.8	28.8	29.1	32.4	34.0	35.9	34.3	33.8
Too much	4.8	4.3	5.4	4.6	6.8	6.9	7.8	6.9
Net (TL -TM)	+58.6	+62.6	+60.2	+58.4	+52.4	+50.3	+50.1	+52.4
	1983	1984	1985	1986	1987	1988	1989	1990
Too little	59.5	58.8	57.8	60.6	68.5	70.2	70.4	72.2
About right	34.7	33.4	34.5	34.1	25.3	25.5	25.0	24.1
Too much	6.9	5.8	7.8	7.7	5.5	6.1	4.6	3.7
Net (TL -TM)	+52.6	+53.0	+50.0	+52.9	+63.0	+64.1	+65.8	+68.5
	1991	1993	1994	1996	1998	2000	2002	2004
Too little	70.7	73.2	66.0	65.6	69.6	71.9	73.7	78.3
About right	26.0	16.2	23.0	24.9	23.5	23.0	20.7	15.8
Too much	3.3	10.7	11.0	9.4	6.9	5.0	5.5	5.9
Net (TL -TM)	+67.4	+62.5	+55.0	+56.2	+62.7	+66.9	+68.2	+72.4
	2006	2008	2010	2012	2014	2016	2018	
Too little	73.5	76.2	60.3	61.2	57.6	63.6	71.6	
About right	19.4	15.7	20.8	23.9	25.7	23.6	21.3	
Too much	7.1	8.1	18.9	14.9	16.8	12.9	7.1	
Net (TL -TM)	+66.4	+68.1	+41.4	+46.3	+40.8	+50.7	+64.5	

D1. Solving the Problems of the Big Cities							
	1973	1974	1975	1976	1977	1978	1980	1982
Too little	55.4	58.6	56.5	48.0	47.7	45.0	45.9	49.8
About right	31.0	29.0	29.5	29.9	29.6	33.6	29.5	26.9
Too much	13.7	12.3	14.0	22.0	22.7	21.5	24.6	23.4
Net (TL -TM)	+41.7	+46.3	+42.5	+26.0	+25.0	+23.5	+21.3	+26.4
	1983	1984	1985	1986	1987	1988	1989	1990
Too little	47.3	51.9	43.3	49.2	44.4	54.7	53.1	59.0
About right	34.5	33.3	38.0	33.3	39.2	34.4	34.0	30.4
Too much	18.3	14.7	18.8	17.5	16.4	10.9	12.9	10.7
Net (TL -TM)	+29.0	+37.2	+24.5	+31.7	+28.0	+43.8	+40.2	+48.3
	1991	1993	1994	1996	1998	2000	2002	2004
Too little	52.5	62.1	60.8	59.1	51.6	51.9	44.7	43.3
About right	33.5	25.3	25.0	25.9	34.5	35.8	40.6	42.5
Too much	13.9	12.5	14.1	14.9	13.9	12.3	14.7	14.2
Net (TL -TM)	+38.6	+49.6	+46.7	+44.2	+37.7	+39.6	+30.0	+29.1
	2006	2008	2010	2012	2014	2016	2018	
Too little	48.1	48.6	42.8	42.3	45.7	51.9	54.7	
About right	39.1	38.5	37.3	39.4	38.2	34.5	34.2	
Too much	12.9	12.9	19.9	18.2	16.2	13.6	11.1	
Net (TL -TM)	+35.2	+35.7	+22.9	+24.1	+29.5	+38.3	+43.6	

Working with the report data provided by the Subject Matter Expert, the Data Visualization Expert was able to create the initial design for the interactive. The idea was to present the relative rankings in order from highest ranked (or most important) to lowest ranked by year. The interactive would allow users to see how the ranking have changed over time. However, in order to move forward with creating the interactive, the next step involved pulling in the expertise of the Programmer.

2.3 Creating the interactive: The Programmer

Working within the vision of the Data Visualization Expert, the programmer was tasked with putting together the interactive. For the GSS interactive, the Data Visualization Export had found some open source coding that could be applied.[1] The Programmer was able to adapt this code to produce an initial graphic.
Programmers bring a wealth of knowledge relevant to creating interactives. There are decisions to be made about the best coding language, the platform, usability, and access.

---

[1] Open source code: http://www.brightpointinc.com/2015-nfl-predictions/

Programmers can help with issues related to where the data live, platforms, and help with other technical concerns.

Figure 2 shows the initial interactive. There are nodes for each category within a year. The nodes are ranked in order of relative ranking, from most important to least important. Figure 3 shows a screen shot of how the user engages with the interactive to show the relative ranking tracts over time. The user can hover over a topic to see the ranking over time in the form of highlighted nodes.

Figure 2: Interactive of GSS spending prioritizes ranked within each year
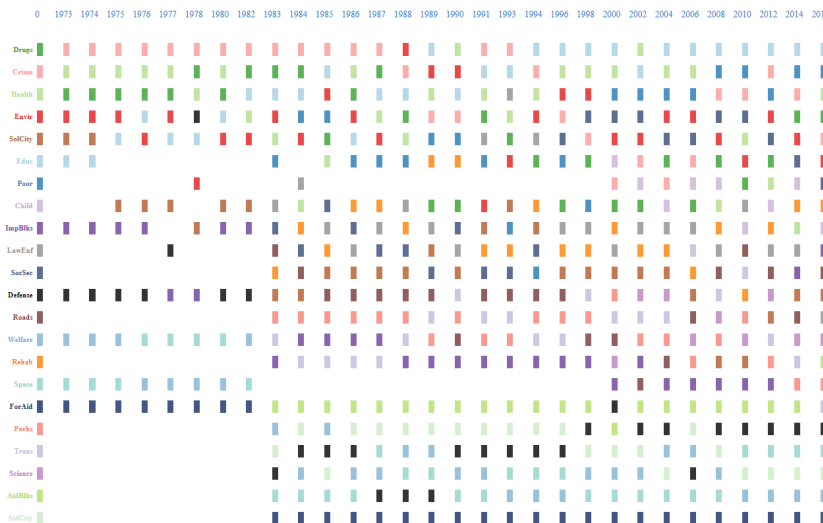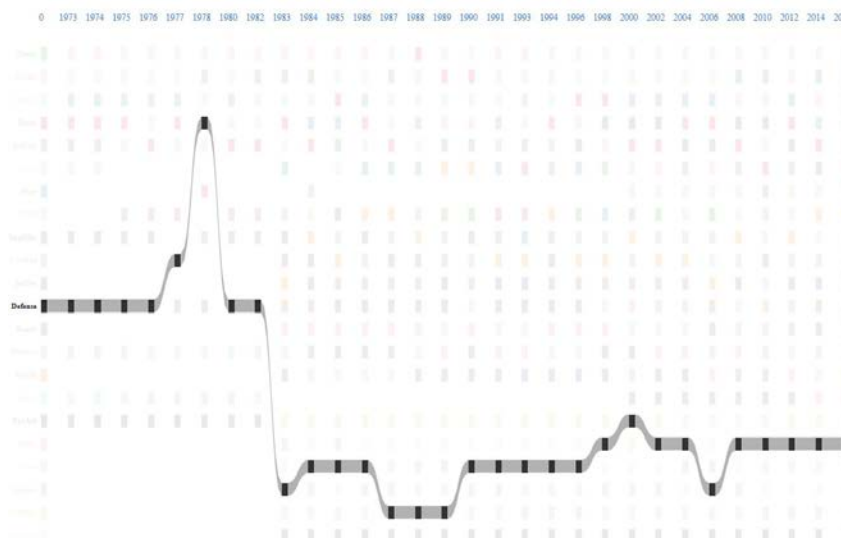

Figure 3: Interactivity by hovering over the nodes

The interactive was off to a good start. However, there were issues with categories that were not collected at each period. Over time, categories have been added to the GSS. Figure

4 shows what the interactive tracts looked like for categories that had incomplete data. That is, there were gaps in the tracts that made the pathway difficult to interpret.

Figure 4: Tracts for categories with incomplete data
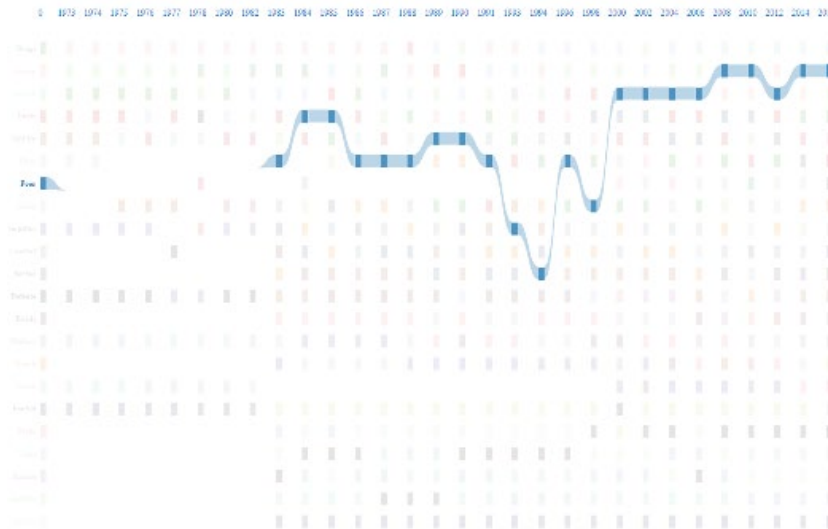
This was problematic for both the visual display and the data story. On the one hand, in order to correct the visual display, topics not included in each year could be removed. On the other hand, removing topics could paint an incomplete data story. That is, if categories were removed, it may affect the relative ranking. In order to check that the data integrity would not be compromised by removing certain topics, the Data Visualization Expert needed to consult with a data expert: the Statistician.

2.4 Maintaining data integrity: The Statistician

Throughout the process, a central tenet of the interactive and data story was data integrity. That is, at all times, the data, methods, and statistics were robust, valid, and reliable. While the Data Visualization and Subject Matter Experts were familiar with data principles, it was up to the Statistician to make sure that the removal of any category would not affect the relative ranking of the items or skew the data story in any way. The statistician suggested that the data story would be more complete if the data were limited to set period of time (the past 30 years, 1998 to 2018). In addition, the exclusion of some categories would not affect data quality. The Statistician also pointed out the need for better category titles.

Figure 5 shows the interactive with new specifications. Consequently, drawing upon years of experience working with data and statistics, the Statistician was able to ensure that the changes did not impede NORC's commitment to data quality.

These changes definitely improved the interactive. However, it lacked visual appeal. The next step was to engage the Graphic Designer.

Figure 5: Narrower time period with limited categories and new category titles
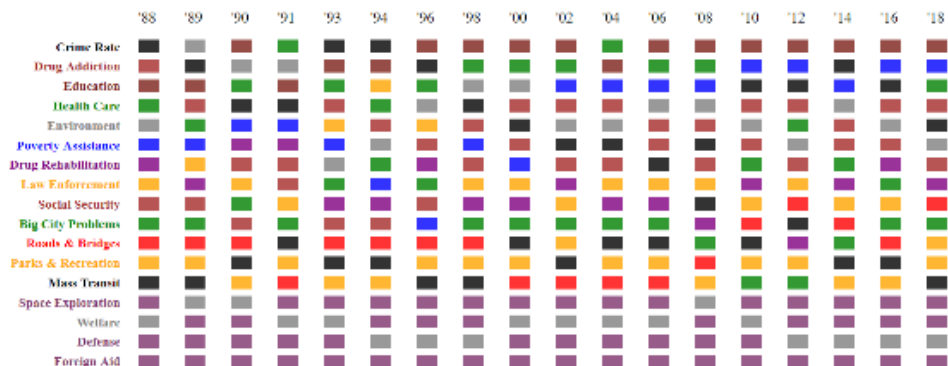
2.5 Visual appeal and aesthetics: The Graphics Designer

While the interactive was functional and told a good data story, it lacked visual appeal. The Graphic Designer was employed to make the visual more compelling by creating balance between the form and function. With an eye for detail, the Graphic Designer had years of

experience with design programs, audience needs, as well as research insights on design. This knowledge was applied to the interactive in a variety of ways. First, the Graphic Designer updated the color scheme to match NORC's branding for the biennial report. Second, the categories were coded into the colors for each topic area; Health and Well-being; Economics, Markets, and the Workforce; Education, Training, and Learning; Global

Figure 6: Visual balance and branded colors added to the interactive

Development; and Society, Media and Public Affairs. This change made patterns in the data more apparent.

Finally, while the interactive was both functionally and visually appealing, it still lacked context in terms of headers and subject lines. It was time to pull in the Writer.

2.6 Improving the language: The Writer

Writers play a crucial role in creating any data graphic. Too often project team members can get caught up in jargon or forget important details that provide context. The writer provides the team an opportunity to see the interactive through the eyes of the target audience by framing the language and verbiage in words that would be understood by the general public. The Writer cleaned up the language and added the header and study description.

Figure 7: Added headers and project description
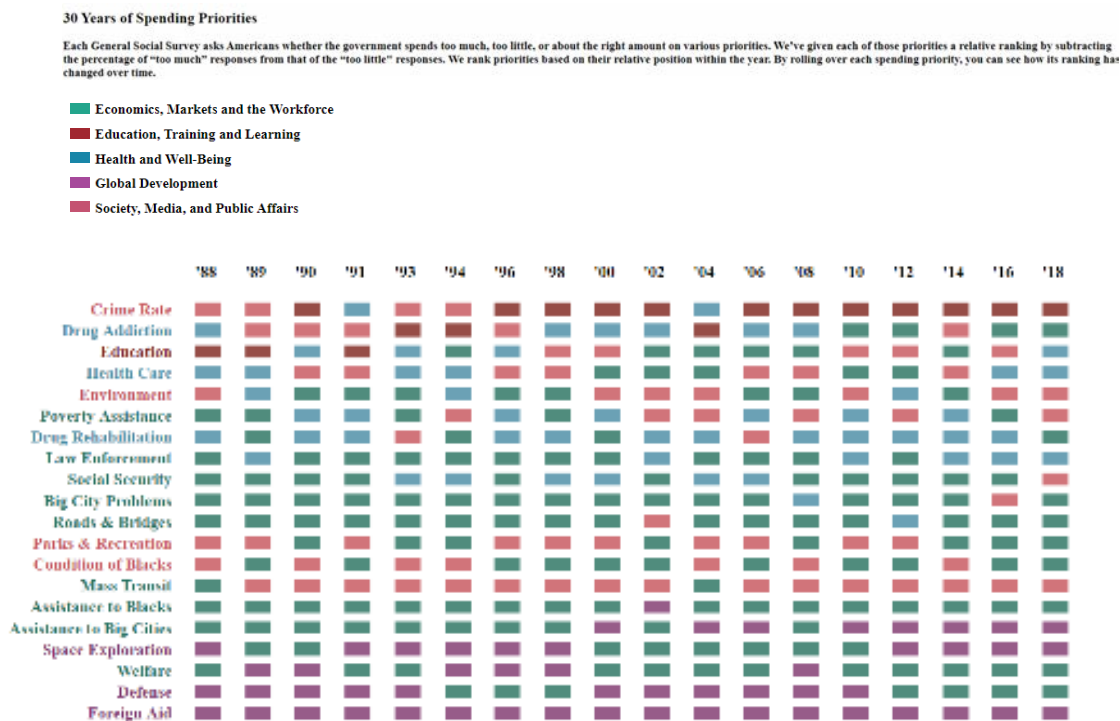
3. The Outcome

The process details the important roles that were needed in order to complete the task of creating an interactive data graphic for the biennial report using GSS data. While it is possible for one person to create an interactive, this project highlights the important roles that are needed in order to create a fully developed interactive data graphic.

Figures 8 and 9 show the final version of the interactive. There were many more iterations until the report was published online (see the final version here: https://biennialreport.norc.org/infographics-gss.html). The path to the end product included many people, many roles, and many skills. There was a great deal of

collaboration. While the primary participants are listed in this paper, there were many more voices heard on the project, including the Communications Director, website designers, and other project team members.

The project was a collaboration and the outcome a success. However, including that many people on a project takes time and costs money. That is an obstacle to this process. It also involved a great deal of management; that is, keeping all the different players moving in the same direction. However, the end result was a good data story presented with an engaging interactive tool.
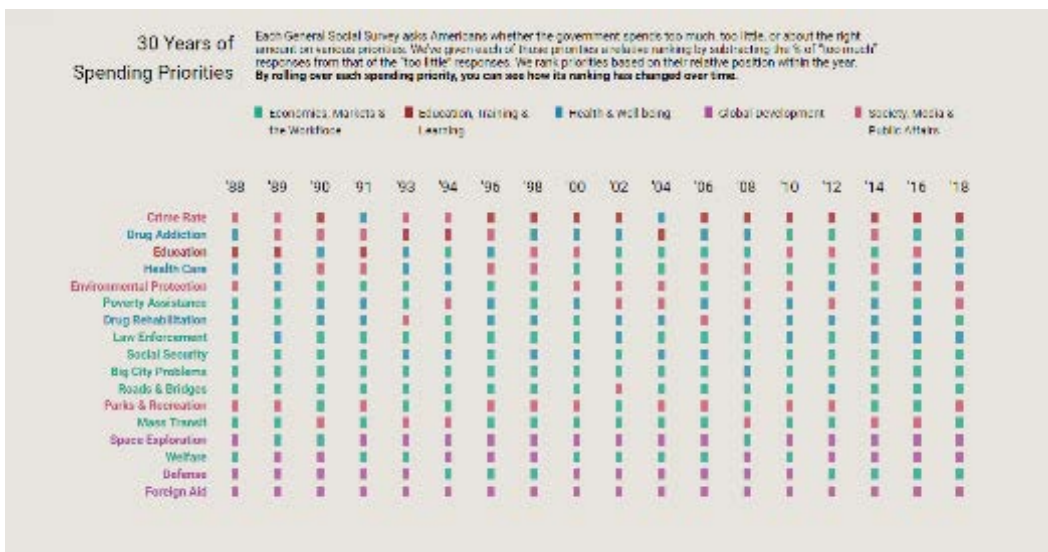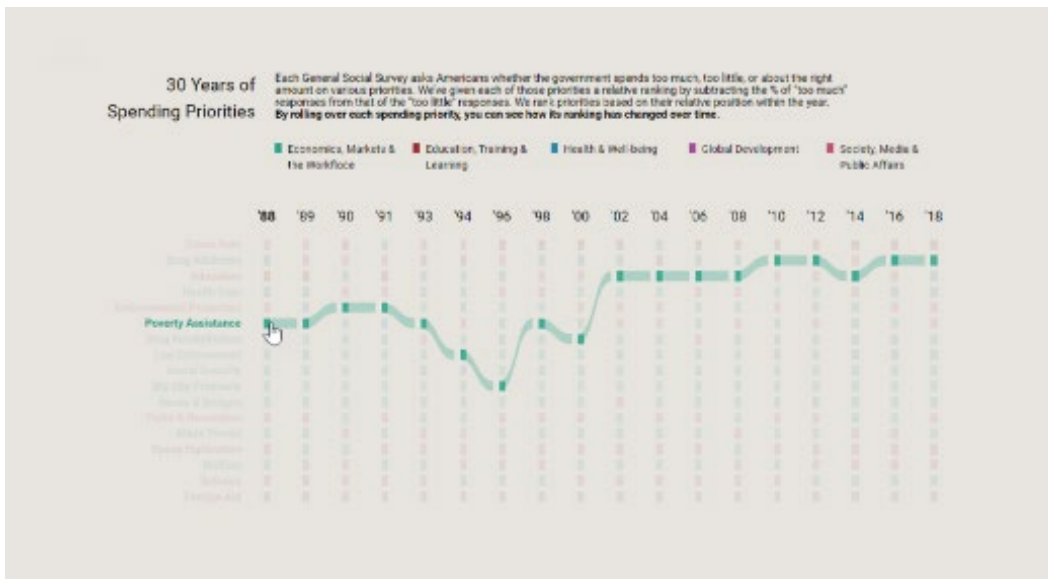
Figure 8: The final version

Figure 9: The final version interactivity

References

Cairo, A. (2013). *The Functional Art: An Introduction to Information Graphics and Visualization.* New Riders, www.newriders.com.

Cairo, A. (2016). *The Truthful Art: Data, Charts and Maps for Communication.* New Riders, www.newriders.com.

Cleveland, W. and McGill, R. (1984). "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods," *Journal of the American Statistical Association,* 79:807-822.

Cleveland, W. and McGill, R. (1987). "Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data," *Journal of the Royal Statistical Society, Series A,* 150:192-229.

Robbins, N. (2013). *Creating More Effective Graphs.* Chart House, Wayne, NJ.