Sense-making of Big Data

Spring Break 2013 – Visualization recognition and meaning making

November 2013

Prepared by:
Joe E. Heimlich, Ph.D.
Victor Yocco, Ph.D.
Sasha Palmquist, Ph.D

Prepared for:
Sense-making of Big Data Team

This project was completed with support from the National Science Foundation(1223698)

LifelongLearningGroup.org
COSI | 333 West Broad Street | Columbus, Ohio 43215

Contents

Visualization recognition and meaning making

Introduction

The Sense Making of Big Data project is designed to study how audiences in public spaces, in this case those in a museum setting, relate to and make sense of representations of large data sets. As the test case for complex big data representations, the project will use science maps. These representations provide an overview of science generally as well as specific areas of STEM, charting and exploring the history and future of science and technology. The data collection portion of the project is conducted at the New York Hall of Science, the Marian Koshland Science Museum, COSI in Columbus, Ohio, and WonderLab Museum in Bloomington, Indiana. Findings from this project will inform the development of a traveling, hands-on exhibition that will enable visitors to create and utilize representations of big data such as maps and charts. In addition, the project hopes to create a foundation for the design of informal learning experiences that encourage participants to explore, engage and make better sense of big data. This project is potentially transformative because big data is becoming ubiquitous and making sense out of data representations is necessary in order to understand and begin to utilize big data.

A pilot study revealed a significant gap in visitor understanding of reference systems and a lack of ability to explain what various representations were communicating. Given the findings in the pilot study, it was determined that the next study needed was to determine what types of visual representations visitors recognize, and how they make meaning (or not) of various visuals.

Method

Population

For this study, each museum initially was intending to conduct 25 adult only, 25 youth only, and 25 adult with youth respondents. If data saturation was not met in the 25, the data collection would then continue through 35 in the categories where saturation was incomplete. After the first round of data collection at COSI, it was determined there was too much 'noise' in the adult with youth respondents (inconsistent means of interaction and process) and that the youth only and adult only categories provided more meaningful data.

Instrumentation

Each of the four participating museums had a set of laminated data representations. These representations were coded and preset in a sequence that has a selection across the types of representations. The 20 visualizations were collected by the project leaders, had labels and legends removed, and were printed in color and laminated so that they had a finished, bright look and would last through the study. The purpose in removing labels and legends was so that individuals responded to the data representation and would then attempt to explain how it would be interpreted without the visual cues provided by the actual legends, titles, and keys. The images were coded on the back by letter for identification. The cards were organized into four discreet sets with each set having five visuals covering different representations. These sets were used in constant rotation. After approach and agreement/verbal consent to participate, the data collector told the participant:

> I'm going to show you a series of figures that show data in different ways. For each, I'm going to ask you a few questions. Are you ready?

They were then asked the following four questions:

> Does this type of data presentation look at all familiar? (probe: where might you have seen images that look like this?)
>
> How do you think you read this type of data presentation?
>
> What would you call this type of data presentation?
>
> What types of data do you think would makes sense for this type of presentation?

Each individual/group did this for five reference systems while the data collector captured comments on the data collection sheet coded for each of the five figures. Perceived sex was also noted to make sure there was no bias in who was asked to participate.

Findings

Respondents

A total of 247 individuals responded: 130 adults and 115 youth. Of those whose sex was recorded, there were 43 adult males, 69 adult females, 55 male youth, and 52 female youth. Youth ages ranged from 9 – 14. The distribution for response to each series follows in Table 1

Table 1: Respondents by series

	Adult		Youth		Sex not recorded		No info	total
	M	F	B	G	Adults	Youth		
1	10	16	12	15	4	3		60
2	13	17	11	14	6	1	1	63
3	10	16	14	12	3	4	1	60
4	10	20	18	11	3	2		64
TOTAL	43	69	55	52	16	10	2	247

One of the anchor comparisons for the series of studies is a gauge of interest in math, science, and art. Participants are asked to rank on a scale of their interest from 1 (not at all interested) to 10 (very interested) in each of the three areas. Science interest was clearly dominant for adult and youth across all four series. Interestingly, only in 3 of the eight categories was there more interest in math than in art. The one outlier in the data is the youth interest in art in series two which had an extremely low 1.12

response. As this response is averaged across the four museums in this study, this is just an oddity in the randomness of research. The findings presented in Table 2 below do not appear to have been influenced or altered by this particular aberration in the data.

Table 2: Interest in science, math, and art by respondents to each series

	Series 1		Series 2		Series 3		Series 4		Overall		Total
	Youth	Adult	Youth	Adult	Youth	Adult	Youth	Adult	Youth	Adult	
Science	8.18	8.22	8.27	8.19	7.41	8.07	7.29	7.86	7.79	8.09	7.94
Math	7.57	6.63	7.50	6.97	6.97	5.93	6.74	6.34	7.20	6.47	6.84
Art	8.45	7.72	1.12	7.33	5.83	5.83	7.17	6.77	5.64	6.91	6.28

Detailed analysis of visual presentations of data by group

The visuals presented five types of data representations: charts (labeled C), graphs (G), maps (M), tables (T) and other visuals (O). Twenty visuals were selected by the data visualization experts on the team they felt would be reasonable representatives of each type of graphic. Each set of five had one visual from each category randomly assigned to the set. A respondent was shown one set of visuals with the five visuals in that set randomly ordered each time.

Series One Visuals

C1

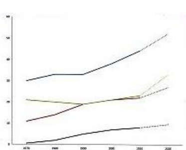
Overall, participants expressed a high level of familiarity with this visual; 47 out of 59 (80%) participants stated they had previously seen a similar image with 21/29 youth and 27/31 adults expressing the affirmative. The locations where participants stated they had seen this image included their school computer lab, in newspapers and magazines, in an art museum, and online.

Participants made comments reflecting that the size of the word is in order of importance i.e. larger words are of greater importance. Some participants felt the words are all related, while others noted that the color of the word represented words that are related.

A number of titles were given to this visual, including: wordle, word cloud, word art, word cluster, words, marketing, and advertising.

Suggestions for types of data that would make the most sense for this type of presentation all focused on words. This included suggestions around themes or topics, and descriptions of something.

G1



Almost all participants had encountered a similar visual before; 56 out of 58 (97%) with 27/29 and 29/29 adults who stated they were familiar with this visual. Participants stated they had seen this image in places such as school— math class in particular—presentations, work, training, and newspapers.

Participants stated you read this visual from left to right along the x-axis and that y-axis represents an increase or decrease in what is being measured. Children and adults noted the visual was missing labels or a key for what was being shown.

Titles given were almost unanimous between line graph and line chart.

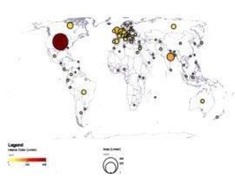
Suggestions for types of data that would make the most sense for this type of presentation all centered on quantifying or measuring the amount of something.

T1

Just over half, 33 out of 60 (55%), of all participants stated this visual looked familiar, though slightly less than half of children participants 14 out of 29 (48%) said so. The places said they had seen this image included school and on the news for children. Adults (17/29) familiar with this type of visual said it was through encounters such as in family trees, sports brackets, work, and statistical courses.

Participants were nearly unanimous in stating that this visual is to be read from left to right, following the lines that connect the terms. Some noted that the dots and shading represented a change or a break, while others noted that all of the words could somehow be related back to the original word on the left.

Suggestions for the types of data that would make the most sense to be presented in this for of visual display focused on tracing the relationship between different things, e.g. family tree, history, main topic and subtopics.



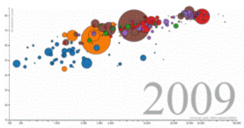
M1

Forty-three out of 60 (72%) participants stated they were familiar with this visual. Only eight out of 28 (29%) of youth and nine out of 32 (28%) of adults were unfamiliar with this type of visual.

Youth had encountered this type of visual in school. Most adults recalled encountering this type of visual at work. Other responses given from adults included online, in books, and in classes.

Participants noted you would read this type of data presentation by using the legend and looking at the colors and the sizes of the dots. Some noted the dots represented countries. Very few participants had difficulty articulating how they would go about reading the data presented in the visual.

Examples of what type of information participants thought would make sense to be included in this type of visual presentation focused on quantifiable data and comparisons, especially population data.

01 

Few participants were familiar with this visual, with 19 out of 58 (33%) stating they had previously seen something similar. Notably, only 5 out of 29 (17%) children stated they had encountered a similar visual presentation of data in the past while 13 out of 29 adults had.

Youth who were familiar with this type of visual stated they had encountered similar visuals in school, online, or in a magazine. Adults stated they were familiar with this visual from work, books and magazines, online, school, and work.

Both youth and adults noted you would read the data presentation by utilizing the x and y-axes and that the color and size of the circle meant something. Many participants noted that the circles were labeled with the names of countries.

Titles given to this visual included bubble graph/chart, chart, circle chart, graph, and scatter plot. Seven (24%) youth and four (14%) adults were unable to give a title.

Suggestions for the type of data that would be best presented with this type of visual focused on quantitative data, e.g. how much of something, demographic information, distributions, and comparing groups or showing relationships.

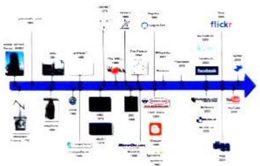
Ease of Series One Visuals

Eight out of 29 (34%) youth participants stated they found the activity easy, seven out of 29 (24%) youth participants stated they found the activity difficult, and 14 out 29 (48%) youth participants stated a combination of both easy and difficult.

Fifteen out of 29 (52%) adult participants stated they found the activity easy, four out of 29 (14%) participants stated they found the activity difficult, and 10 out of 29 (34%) adult participants stated a combination of both easy and difficult.

Regardless of age, the reason behind finding the activity easy or difficult had to do with whether participants were familiar with the visuals (or similar) or not.

Series Two Visuals



C2

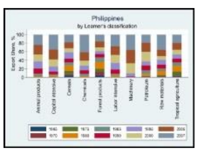
Sixty out of 63 (95%) participants were familiar with this type of data presentation. Thirty five out of 36 (97%) of the adults were familiar with type of data presentation, while 25 out of 27 (93%) of the youth expressed familiarity.

Both youth and adults stated they have encountered this type of data presentation in school, online, and (adults) at work.

Both youth and adults stated you would read this type of data presentation from the left to the right, with the events on the left having occurred prior to the events on the right.

Participants were near unanimous in calling this type of presentation a “timeline.”

Participants stated that information related to dates or the progression of something would be appropriate to display with this type of visual presentation. A few adults used the term “chronological.”



G2

Forty-seven out of 61 (77%) participants were familiar with this type of data presentation. Twenty-nine out of 35 (83%) adults and 18 out of 26 (69%) youth also noted being familiar with this type of data presentation. Youth participants were familiar with this type of data presentation from school, specifically math class. Adults stated they were familiar with this type of data presentation from school, work, and magazines.

Youth explained they would read this type of data presentation by looking at the key and the colors, and comparing the years. Adults noted the presence of the key, colors, and the x and y-axes as being critical components of how they would read this type of data presentation.

Most participants called this type of data presentation a “bar graph.” Some youth participants also called this type of data presentation a “chart,” “table,” or “time plot.” Other names given by adult participants included, “timeline,” “flow chart,” and “proportion chart.”

Generally speaking, both youth and adult participants suggested quantities, percentages, comparisons, and other quantitative data would be the best types of information to be presented with this type of data presentation.

T2

Thirty-five of 61 (59%) participants were familiar with this type of visual. Twenty-three out of 35 (66%) adults were familiar with this type of visual, while 12 out of 26 (46%) youth expressed familiarity. Most participants stated they were familiar with this type of data presentation through something they encountered online. News websites, Yahoo, and RSS feeds were examples provided by participants for where they have encountered similar presentations of data.

Youth and adult participants gave responses suggesting that this type of visual presentation would be read by reading the words, the size of the words, and the colors associated with the words. Some mentioned "news" items that could potentially be clicked on and a larger story would appear. Adults typically gave more in depth answers, including that the colors represented specific categories of news, and that the words themselves were grouped into categories/topics.

There were almost as many suggestions for what participants would name this type of presentation as there were participants. Many of the names given contained the term "news." Examples of names given by youth participants include: "online newspaper," "news chart," and "news map." Examples of responses given by adult participants include: "News map," "News block," and "News feed."

Participants from both age groups suggested news and current events were the types of data that made the most sense to be presented with this type of visual. Adult participants were more likely to give additional descriptive terms to their answers such as "Important headlines."

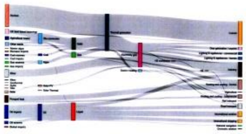
M2

Fifty-nine out of 63 (94%) participants were familiar with this type of data presentation. Thirty-five out of 37 (95%) adults and 24 out of 26 (92%) youth recognized this type of visual representation. Youth participants were primarily familiar with this type of data presentation from school, including history, geography, and social studies classes, as well as TV such as the weather channel. Adult participants stated they were familiar with this type of data presentation from school, online, news/newspapers, as well as TV.

Participants stated they would use the legend/key, shapes, and color to read this presentation. Many noted the shade or level of coloring represented a percentage of whatever was being measured. Adults used the word "scale" in relation to the level of color; youth participants did not.

Participants often used the word map or graph in what they would call this type of data presentation. Examples that were given included chart, data chart, map, data map, population chart, and population graph.

Participants felt information that could be presented by geographical location would be most appropriate to display with this type of visual presentation. Population based data was mentioned frequently by both adult and youth participants. Weather and political information were also noted by a number of participants as data that would best be presented with this type of visual display.

O2 

Fifteen out of 61 (25%) participants expressed familiarity with this type of visual representation. Ten out of 35 (29%) adults and five out 26 (14%) youth were familiar with this visual. Adult participants stated they had encountered this type of presentation online, in a news article, in academic articles, "in everyday life," and in presentations. Youth noted they had seen this type of data presentation at school (history), on their parents work papers, and on the Wii video game system.

Participants noted that you would read this type of presentation based on the colors, lines, and labels of the presentation. A number of youth and adult participants noted the data would be read from left to right. Adults often gave a deeper description of how they would read this type of data presentation including making comparisons.

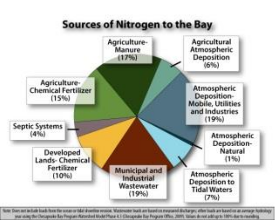
Youth participants provided a wide variety of names they would give to this type of data presentation, including: bar graph, line graph, graph, flow charge, wave chart, and scale. Adult participants gave a wide variety of responses similar to the youth, including: flow chart, bar graph, pictorial graph, and graph.

Youth participants provided vague responses to what type of data would be most appropriate to present with this type of visual. Many youth took cues from the information given on the visual itself: "something to do with energy" and "resources and transportation." Adults were often more general with what they would think was appropriate to present with this type of visual: "connecting ideas," "something with a start and an ending," and "comparison of different things," are examples of responses given by adults.

Ease of Series Two Visuals

Seven out of 26 (28%) youth participants stated they found the activity easy, four out of 26 (16%) of youth participants stated they found the activity difficult, and 15 out of 26 (56%) of youth participants stated they found the activity both difficult and easy. Fifteen out of 34 (44%) adults stated they found the activity easy, three out of 34 (9%) adults stated they found the activity difficult, and 16 out of 34 (47%) adults stated they found the activity both easy and difficult. Regardless of age, the reason behind finding the activity easy or difficult had to do with whether participants were familiar with the visuals (or similar) or not.

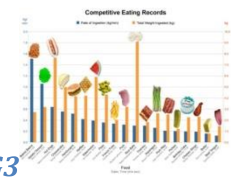
Series Three Visuals

C3 

Fifty-six out of 60 (93%) participants were familiar with this type of visual display of information. Thirty out of 30 adults were familiar with this type of visual display while 26 out of 30 (87%) youth were familiar. Youth stated they had encountered similar visual displays of information in school, particularly math class, online, and in textbooks. Adult participants were familiar with this type of visual display of information from work and school. A few adults noted they have created similar displays for their work.

Both youth and adult participants were able to easily describe how they would read this type of presentation. Most participants noted the word “percentages” as it related to the colored “slices” as well as making up parts of a whole represented by the entire circle. Participants were near unanimous in calling this visual a “Pie” chart or graph. A few participants called this a circle chart.

Participants stated the types of data that would make the most sense to be presented like this were percentages, comparisons, parts of a whole, and things that can be broken down categorically.

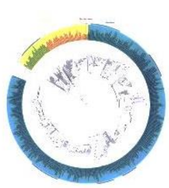


G3

Fifty-six out of 60 participants were familiar with this type of data presentation. Twenty-eight out of 30 adults and 28 out of 30 youth (93% of each group) expressed familiarity. The vast majority of participants stated they had encountered this type of visual display of data at school or work. Math and science class were frequently mentioned as specific classes this type of visual display of data is used.

Participants stated they would read this type of visual presentation by looking at the colors and length of the bars, the pictures on the bars, and reading the labels of the x and y-axes. The vast majority of participants called this a bar graph. Table graph and column graph were other names given to this visual.

Participants stated “how many” and “how much” of something would be make the most sense displayed with this type of visual. Adult participants were more frequently likely to note that comparisons between two or more things would make sense to be displayed like this.

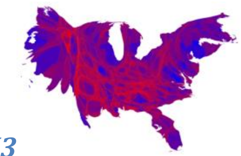


T3

Twenty-four out of 60 (40%) participants were familiar with this type of visual display of information. Fourteen out of 30 (47%) adults and10 out of 30 (40%) youth claimed familiarity with this type of visual display of information. Youth familiar with this type of visual stated they had encountered similar visuals in school. Two youth cited math class and one cited science as the specific classes they encountered similar visuals. Three adults noted they had encountered similar visuals in school, other adults each stated a different place they had encountered similar visuals, including maps, family tree, and a book.

Youth stated they would read this type of data presentation by looking at the words, the colors, and where they are located on the circle. Youth had difficulty articulating beyond this how the visual should be read, although some youth noted that where/how far along the circle a word was had some type of significance. Adults were more likely to note that there was a starting point along the circle, and the words started there and the brackets signified how the words were related to this starting point.

Adults and youth had difficulty providing a name for this type of presentation. Those who did attempt to name this visual were likely to include something about the shape (circle) or the timeline aspect of the visual. As with thinking of a name, both youth and adults had difficulty describing what type of data would be most appropriate to display with this type of visual. There was little consistency among answers given by youth, with the exception of multiple youth making comments around "how much" of something. Adults noted data representing something with a start and finish, how something has progressed, and hierarchical data as being appropriate to display with this type of visual.

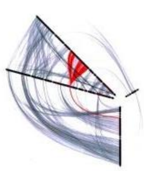


M3

Thirty-eight out of 60 participants were familiar with this type of visual display of information. Twenty-one out of 30 adults were familiar with this type of visual display of information, and 17 out of 30 youth were familiar with this type of visual display of information.

Adults familiar with this type of visual stated they had encountered similar visuals in the news (paper), maps, and weather. Adults also noted this visual was similar to Rorschach tests. Youth that were familiar with this type of visual stated they had encountered similar version in school, particularly geography and social studies classes. Both youth and adult participants stated they would look at the colors and the key to read this type of data presentation. Four adults noted the shape of the data presentation looked like the U.S., while two youth did so.

Nearly all participants who attempted to provide a name for the visualization used the term "map" in their name. Youth participants most often simply called it a "map," while adult participants added descriptive terms such as demographics and distributions. Examples of other terms given this visualization included graph and cardiogram. In providing examples of what type of information would be most appropriately presented with this type of data visualization, participants used language around shapes, colors, and geography. Many participants noted weather or temperatures would be appropriate to display with this type of visualization.



O3

Ten out of 60 participants were familiar with this type of visual display of information. Seven out of 30 (23%) adults compared to 3 out of 30 (10%) youth who expressed familiarity.

Adults provided a range of answers as to how they were familiar with this type of visual, including books, Scientific America, art, and drawings from school. Two of the youth stated they were familiar with this visual from math class, and one other stated "the eye doctor."

Most youth were unable to explain how they would read this type of visual. Some noted the shapes and colors of the lines would be relevant to interpreting this visual. Adults were more likely to note the length, thickness, and distance of the lines were important for interpreting the data presentation, although many adults were unable to explain how they would read this visual.

Many participants did not attempt to name the presentation. Those who did took contextual clues from the shape of the visual, providing names such as scribbles, seismic chart, spiral graph, and swirl. Again, most participants were unable to provide an answer to what type of data would make most sense to be presented using this type of visual. Some suggestions made by youth included colors and shapes. Some adults suggested comparisons, movement, and how things are interconnected.

Ease of Series 3 Visuals

One out of 30 (3%) youth participants stated they found the activity easy, 16 out of 30 (53%) youth participants stated they found the activity difficult, and 11 out of 30 (37%) youth participants stated they found the activity both easy and difficult.

Eight out of 27 (30%) adult participants stated they found the activity easy, three out of 27 (11%) adult participants stated they found the activity difficult, and 16 out of 27 (59%) adult participants stated they found the activity both easy and difficult. Regardless of age, the reason behind finding the activity easy or difficult had to do with whether participants were familiar with the visuals (or similar) or not.

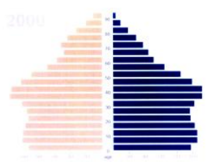
Series Four Visuals

C4 

Fifty-nine out of 71 (83%) participants were familiar with this approach to data presentation. Twenty-four out of 34 (71%) youth and 34 out of 37 (92%) adults stated they were familiar with this type of data presentation. A majority of both youth and adult participants stated they had encountered this type of data presentation on the subway or on a map.

Participants stated they would read this type of data presentation like a map, use the legend or key, find out where they are and where they wanted to go, and follow the path. Most participants included the term "map" in the name they would give this type of data presentation. Often a descriptor was present such as transit, or subway.

Participants stated they type of data that would make the most sense to include in this type of visual was locations, places, and maps. Both youth and adult participants noted the importance of including a key and compass for this type of data presentation.

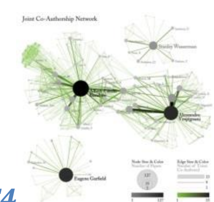


G4

Forty-nine out of 71 (69%) participants were familiar with this type of data presentation. Seventeen out of 34 (50%) youth and 32 out of 37 (86%) adults stated they were familiar with this type of data presentation.

Youth participants stated they had encountered this type of data presentation in school, specifically math class, social studies, and in textbooks. Most adults that were familiar, had encountered this type of data presentation at school, work, or on the news. Two adults noted being familiar with this type of data presentation from the "census." Responses from both youth and adult participants reflected they would read this type of data presentation by reading the labels and comparing the size of the different bars.

A majority of participants stated they would call this type of data presentation a "bar graph. Participants stated the type of data that would make the most sense to include in this type of visual was quantitative data, and data that compares (at least) two different things.

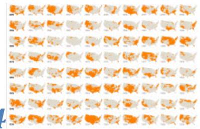


T4

Only 14 out of 71 participants (20%) were familiar with this type of data presentation. Three out of 34 youth (9%) and 11 out of 37 (30%) adults stated they were familiar with this type of data presentation. The three youth stated they had encountered this visual either on the computer or Internet. Adult responses were varied, and included work, scientific paper, TV, chemistry, and a spider web.

Most youth participants were unable to elaborate in detail how they would read this type of data presentation. Many stated they would use the key, color, and words. Adult participants gave similar responses reflecting the use of a key, the colors, and the words, however many adults were more descriptive in noting that sizes and shades mattered, as well as there being connections between the different nodes.

"I don't know" was a frequent response by both adult and youth participants for what name they would give this type of data presentation. Those that did attempt to give a name often based it on the shape of the presentation and used the word "web" e.g. spider web and web graph. There were a wide variety of responses as to what type of data would make the most sense to present with this type of visual. Responses reflecting the relationship or connections between things were the most frequently overlapping theme among all participants.

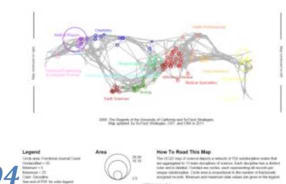
M4 

Twenty-three out of 33 (70%) youth participants were familiar with this type of visual presentation of data while thirty-two out of 37 (86%) adults stated they were familiar for an overall familiarity by 79% of the respondents. Youth participants were familiar with this type of visual presentation of data from school, social studies an math class specifically, the Internet, an atlas, and on TV. Four youth specifically noted a popular commercial showing cell phone coverage across the U.S. Adults gave similar responses as youth, with many adding they had seen this type of presentation on the news. Three adults specifically noted the cell phone coverage map commercial.

Both youth and adult participants gave similar responses to how they would read this type of data presentation. Most participants responded that the data were to be read as different years represented by the maps and that the color and concentration of color represented the distribution/spread of something.

There was little consistency between responses participants gave to what they would call this type of data presentation. A number of responses included the term chart or graph. Pictograph, color graph, time graph and population map are examples of these types of responses. Some other responses included population growth and timeline.

Both youth and adult participants felt the most appropriate type of data to display with this type of visual would be location based or geographical data that changes over time. Specific responses related to political data such as lawmaker representation and election results, weather, and changes in population.

O4 

Sixteen out of 71 (23%) participants were familiar with this type of data visualization. Five out of 34 youth (15%) and 11 out of 37 (30%) adult participants expressed familiarity.

Youth participants familiar with this type of presentation stated they had seen this type of presentation online, maps of human body (veins), in school, and math class. Adult participants familiar with this type of data presentation stated they had seen similar presentations of data in presentations at work, in school, and in medical journals. Eight youth and two adults were unable to describe how they would read this type of data presentation. Youth providing a response noted they would look at the legend, key, size of circles, and the colors. Some of the youth stated the circles represented different subjects. Adults responses mirrored those of youth, with slightly more sophisticated language such as “academic subjects,” “Venn Diagram,” and “perimeter” being used by some of the adult participants.

Ten youth and 11 adults did not attempt to provide a meaningful response to what they would name this type of data presentation. Both youth and adult participants providing a response gave names such as “chart,” “graph,” “map,” and “line graph.”

Youth participants stated quantitative data and data comparing things would make the most sense to display with this type of data presentation. Twelve youth gave no meaningful response to this item, or couldn't think of anything to display. Adult participants were able to give greater detail as to what data they would display with this type of visual presentation. Adult responses included relationships, distributions, and things that are connected. Some adults took their cue from the data on the actual visual and suggested academic subjects and topics would be most appropriate.

Ease of Series 4 Visuals

Four out of 32 (13%) youth participants stated they found the activity easy, 13 out of 32 (41%) youth participants stated they found the activity difficult, and 15 out of 32 (47%) youth participants stated they found the activity both easy and difficult.

Eleven out of 30 (37%) adult participants stated they found the activity easy, four out of 30 (13%) adult participants stated they found the activity difficult, and 15 out of 30 (50%) adult participants stated they found the activity both easy and difficult.

Regardless of age, the reason behind finding the activity easy or difficult had to do with whether participants were familiar with the visuals (or said they were familiar).

Comparison of Types of Visual Representations

The visuals were selected to represent a range of complexity for each basic type of data representation. Both youth and adult familiarity followed the same patterns (see Table 2) on the ranking of familiarity among the five types of representations, and also on their familiarity within each of the types of representations. Overall, the greatest familiarity was with charts followed by graphs. Both these types of visuals, and the maps, had a consistent majority of adults and youth being claiming some level of familiarity with the type of image. Tables were somewhat lower in recognition, with none of the tables being familiar to over half the youth, and only two approaches to tables being familiar to more than half the adults. Other visuals were by far the most unfamiliar with an average of only 14% of youth stating any familiarity and just over a third (34%) of adults having familiarity overall.

Charts

The most familiar of data representations were the charts. These were also the representations that had the least consistent rankings of the types of visuals. Adults had the pie chart as the most familiar whereas youth had the bar chart (time chart of communications media) as their most familiar. The two were flipped for the second place ranking. Likewise, the third ranked visualization for the adults was the subway stop visual with the fourth being the Wordle; these were reversed for the youth. Even so, there are clear distinctions between the first two and the second two, suggesting that there is far more familiarity with the top two than with any other visual representation.

Table 2: Familiarity with types of data representations

	Adult Familiarity			Youth			Overall
	n/N	N	%			%	%
C1	27	33	82	21	29	72	80
C2	35	36	95	25	27	93	95
C3	30	30	100	26	30	87	93
C4	34	37	92	24	34	71	83
			92.3			**80.75**	**87.75**
G1	29	29	100	27	29	93	97
G2	29	35	83	18	26	69	77
G3	28	30	93	28	30	93	93
G4	32	37	86	17	34	50	69
			90.5			**76.25**	**84.00**
T1	17	29	59	14	29	48	55
T2	23	35	66	12	26	46	59
T3	14	30	47	10	30	40	40
T4	11	37	30	3	34	9	20
			50.5			**35.75**	**43.5**
M1	23	32	72	20	28	71	72
M2	35	37	95	24	26	92	94
M3	21	30	70	17	30	57	63
M4	32	37	86	23	33	70	79
			80.75			**72.5**	**77.00**
O1	16	29	55	5	29	17	33
O2	10	35	29	5	26	14	25
O3	7	30	23	3	30	10	17
O4	11	37	30	5	34	15	23
			34.25			**14.00**	**24.50**

Graphs

Of the graphs, the bar graph was recognized by all adults and most of the youth. The more complex bar graph of competitive eating foods by weight and ingestion was in a strong second in terms of being familiar and describable by the respondents (though there were many misinterpretations of the graph). The population pyramid was the third most familiar for adults, but the fourth for the youth with a large difference in familiarty between this one and the stacked bar graph for the youth.

Tables

The tables used for the representations were challenging to most of the respondents. They were not recognized as tables, and the familiarity with the T2, a newsmap, seemed to be cued more by the words than the structure. As the colors generally had different sized titles with related content, some individuals did seem able to make connections between size, color, and placement. The second most familiar was a visual that respondents noted looked like a flow chart or a decision tree, even though they did not make sense of the specific topic. The circle tree was far less familiar and the relationship visual was most confusing of this type of representation.

Maps

The maps included a density map of the U.S. (most recognized and comfortable). The second most familiar was a time series drought map of the U.S. The world map with counts of density was recognized as a world map, but there was confusion as to the circles. A distorted map of proportions of U.S. voters by party vote was the least familiar and respondents had the most difficult time explaining how it would be read or used.

Other data representations

The health and wealth of nations (O1) was familiar to slightly more than half the adults and a third of the youth as a line graph with some other things being measured. There was some recognition of the relationships among sciences by a third of the adults and a sixth of the youth as representing something about relationships and because science areas were present on the visual, there were comments about "science." The other relationships represented by the final two graphics were confusing to the majority of respondents, many of whom did not even try to explain how these visuals would be interpreted.

Conclusions

Overall perceptions and recommendations

Familiarity with a visual display of data was most often based on events/activities occurring in daily life, such as attending school, work, online, and reading material such as magazines and books. Experience/familiarity with more complex visuals increased with age.

Participants utilized anchor points for their understanding of a chart. Familiar visuals were often easily understood. When encountering an unfamiliar visual, participants utilized the anchor points. For example, if a visual looked like a certain shape, participants would note that. If a visual had a pattern in it they would note that. Participants were familiar with looking for things like a key and labels. When this information is missing, participants expressed disorientation as to what the information might be, as well as noted that this information is critical and missing, making the visual useless to them.

Both youth and adults had the least familiarity and most difficulty explaining how to interpret the tables and other labeled visuals.

Large takeaway message: Context and previous experience matters. Participants of all ages are familiar with a wide variety of visual displays of data. If a participant encounters a visual display of data with which they are unfamiliar, they attempt to make a connection to something with which they *are* familiar. When there is no context or previous history with a visual representation, participants would default to searching for clues such as title, key, and legend. When these items were not present, OR when they were unclear, participants expressed frustration.

Finding	Recommendation

• Participants were familiar with visual displays of data from daily life activities
  ○ For children, school and specific classes such as math, as well as online are the most common cited places they encountered visual displays of data
  ○ For adults, work, college, online, and reading newspapers and magazines are the most common ways they encountered visual displays of data
	• Connecting data representations to familiar experiences and visuals may be valuable and worth testing

• Familiarity and understanding of visuals increased with age and experience	• Scaffolding from less complex visuals to more complex visuals, drawing the connection between less complex displays of data, may also facilitate understanding of more complex data representations
• Participants expressed frustration/confusion over visuals that did not seem intuitive to them	• There may be value in attempting to help individuals see how even complex representations are based on consistent structures and types of visuals
• Participants expect titles, keys, and labels on all visual displays of data	• The first study found that many individuals did *not* read titles, keys, and labels in making meaning of the visual
• Many participants, especially children, utilize the shape of the visual display to provide guidance in naming the type of visual display of data	• To understand visuals, it may be important to determine if meaning is made by building on familiar, or deconstructing the unfamiliar representation
• Participants have a set vocabulary around how to interpret visual displays of data (e.g. X and Y axes, legend, key, comparison, quantify), seemingly influenced by age/experience	• It might be valuable to conduct additional testing to determine the specific terminology visitor types use, as well as how they would like more complex terminology introduced to them
• Participants were more familiar with the purpose of visuals designed to compare quantitative data	• There may be a need to understand why and how qualitative data are displayed visually
• Similarly, participants were better able to suggest types of data that could be displayed on the quantitative visuals	• Respondents recognize basic representations, but have difficulty when the representations are more complex with multiple layers of data