Issue Salience and Data Visualization

Erin Coates

Honors Thesis Prepared to the School of Journalism and Communication

University of Oregon

Winter 2017

Approved by: _____Adviser: Kim Sheehan

Approved by: _____Second Reader: Nicole Dahmen

Acknowledgements

Thank you to Kim Sheehan for all of your love and support over these past few years, and especially for serving as my advisor for this project. None of this would have been possible without your encouragement to chase my dreams. Thank you to Nicole Dahmen as well for getting me interested in Data Visualization and providing extensive knowledge on the subject. I would also like to thank all of the School of Journalism Faculty and Staff who have shaped me as a journalist and person over the past four years.

Thank you also to my family for believing in me and to Connor Mann, my constant support. I love you all.

Table of Contents

Introduction:

Interactive data visualizations have created a new way for online journalists to tell a story. With the overwhelming amount of data available, these visualizations make it easier to compare similar data. The data available on different controversial topics from vaccination rates to mass shootings in the United States could actually make a real impact on comprehension of complex topics. This is where interactive data visualizations come into play. Journalists, designers, and computer programmers team up to create new ways to tell stories based on data so the audience can learn from it and draw their own conclusions. This topic is important to look into because with more newspapers and magazines moving into online platforms, there is a growing need for different story telling techniques to grasp the attention of readers. The goal of this project is to see if the salience of controversial topics are affected upon seeing a data visualization, a traditional story or a mix of both. Whether or not interactive data visualizations are making a real difference has not been investigated

There are many different terms for what is deemed in this project interactive data visualization, but for this project, they all are similarly linked and intend to mean interactive data visualization. Some of these terms include: interactive information graphics, visual data storytelling, narrative visualization, narrative information visualizations, InfoVis, and data visualization. Researchers use different names based on the specifications of what they are researching, but for this project they can all be interpreted as the same construct.

Objective of Research

Previous research on interactive data visualization has covered what it is and how to create these newer forms of storytelling. Seth Lewis and Oscar Westlund's article *Big Data and Journalism: Epistemology, expertise, economics, and ethics* discuss how journalists have increasingly been using big data to report and present news in creative and different ways, one of which is data visualization. Wibke Weber and Hannes Rall discuss in their study *Data Visualization in Online Journalism and Its Implications for the Production Process* how data journalism today allows for the data to be analyzed and visualized with little knowledge of information technology. Jessica Hullman and Nicholas Diakopoulos the different ways to us data visualizations to tell stories in *Visualization Rhetoric: Framing Effects in Narrative Visualization.* However, there has been little research on why it is important to create interactive data visualizations. The specific research question of this project is: Does the use of data visualization in online media affect the salience of the story from the audience's perspective? Journalists and data collectors alike can draw from this research to determine whether or not audiences will pay more attention to story forms where they are able to draw their own conclusions based on what they see in front of them.

Literature Review

The Visual Teaching Alliance, a team of researchers dedicated to bringing unique instructional approaches and up-to-date technology to the field of education, quotes a few different studies to show the importance of using visuals. Approximately 65 percent of the human population are visual learners, (Mind Tools 1998 quoted by Visual Teaching

Alliance). “The brain processes visual information 60,000 times faster than text,” (3M Corporation 2001 quoted by Visual Teaching Alliance) and “our eyes can register 36,000 visual messages per hour,” (Jensen 1996 quoted by Visual Teaching Alliance). These facts are important to understand because interactive data visualizations are made up of visual information, which the general population perceives quicker than words.

Mark Coddington’s article *Clarifying Journalism’s Quantitative Turn* distinguishes between the three different types of quantitative journalism practices, the potential of each to develop stories, and the ethics behind using big data sets for stories and visuals. Computer assisted reporting, data journalism, and computational journalism are professional quantitative forms of journalism that have each been used and relied on to increase the interaction between programmers and journalists (Coddington, 2014). Coddington specifically said, quantitative journalism “has great potential to broaden journalism’s ability to make democratic institutions more responsive and legible to the public, but even within this sub-area of journalism, views of the public and the journalistic process are broadly disparate” (Coddington 332, 2014).

Computer-assisted reporting, or precision journalism, became popular in the 1950s and uses surveys, content analysis, and statistical analysis to answer and validate different journalistic questions, (Houston, 1996 as cited in Coddington, 2014). Data journalism, or data-driven journalism, is the analysis and presentation of different types of data. It differs from computer assisted reporting because of its emphasis on visualization design and journalistic values (Gordon, 2013 as cited in Coddington, 2014). According to Coddington, computational journalism is “a strand of technologically oriented journalism centered on the application of computing and computational thinking

to the practices of information gathering, sense-making, and information presentation, rather than the journalistic use of data or social science methods more generally," (Coddington 335, 2014).

The Role of Data Journalists and Big Data

Many data journalism projects allow audience access to data and develop tools for them to explore and personalize that data. In this manner, data journalists allow the audience to draw meaning from data in their own way (Parasie & Dagiral, 2013 as cited in Coddington, 2014). One of the headlines of the Society of Professional Journalists Code of Ethics is "Be accountable and transparent," further stating "Ethical journalism means taking responsibility for one's work and explaining one's decisions to the public," (Society of Professional Journalists, 2014). According to Coddington, "transparency of both process and product are a core element of data journalism." This is not only ethical, but audiences want to see the raw data as well as the analysis from other people (Stray, 2010 as cited in Coddington, 2014).

Data journalists use what is known as big data to gather and analyze information to reach different conclusions. According to Gartner Inc, the world's leading information technology research company, "Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms or information processing that enable enhanced insight, decision making, and process automation," (Gartner IT). Basically, big data is the data that can be collected from daily online activities such as searching the web are gathered together. With big data, the reporter's task shifts from finding and collecting data to processing it, and the analysis shifts from being hypothesis-driven to an inductive and exploratory approach. This means that

instead of in-depth analyses, this new form of journalism focuses on producing the analyses as quickly as possible (Coddington, 2014). Drawing from big data, data journalism, as well as the other quantitative journalism practices, requires a more active audience (Coddington, 2014). "The goal of data journalism is to allow the public to analyze and draw understanding from data themselves, with the data journalist's role being to access and present the data on the public's behalf" (Coddington, 2014).

Seth Lewis and Oscar Westlund's article *Big Data and Journalism: Epistemology, expertise, economics, and ethics* discusses big data, the future of journalism, and the ethics of using such data. "For journalism, big data embodies emerging ideas about, activities for, and norms connected with data sets, algorithms, computational methods, and related processes and perspectives tied to quantification as a key paradigm of information work," (449). This big data is neither good nor bad for journalism, but it all depends on its purpose and whose intentions are behind it (Lewis & Westlund, 2014). Data and databases have been used for news work and journalistic evidence for years in the form of computer assisted reporting and old forms of information visualization (Cox, 2000 as cited in Lewis & Westlund, 2014). This leads to the newer question for journalists of "not whether data, computers, and algorithms can be used by journalists in the public interest, but rather how, when, where, why, and by whom," (Howard, 2014 as cited in Lewis & Westlund, 2014).

In terms of economics, journalists have increasingly been using big data to report and present news in creative and different ways because audiences are interested in new types of news (Lewis & Westlund, 2014). This idea is called method journalism and it moves from the area of coverage, such as a beat or topic, to the method of coverage, such

as data visualization (Madrigal, 2014 as quoted in Lewis & Westlund, 2014). The ethics of big data have developed past the traditional standard such as honesty, accuracy, transparency, and public service (Singer, 2007 as quoted in Lewis & Westlund, 2014). Big data poses its own ethical questions about issues such as user privacy, information security, and data manipulations that need to be taken into consideration (Crawford, 2014 as quoted in Lewis & Westlund, 2014). The first issue discussed is the process of publishing data or making large data sets publicly available online. Even though many journalists want to make complete data sets open for public scrutiny and exploration, there are underlying problems with publishing such data regardless if it was provided by governments or political institutions, or assembled in the newsroom by data-scraping (Schudson, 2010 as cited in Lewis & Westlund, 2014). Data-scraping involves taking public data to use for stories without weighing "the benefits of open data against the risks of personal harm that may come with publication," (460). These problems could be overlooked because of the size of the data set or the attractiveness of making it publicly available (Lewis & Westlund, 2014). However, it is necessary for journalists to weigh the benefits of the open data set against the potential personal harm that could come with the publication of certain information (Howard, 2014 as quoted by Lewis & Westlund, 2014). "Just because certain content is publicly accessible does not mean that it was intended to be made public to everyone," (Boyd & Crawford, 2012 as quoted by Lewis & Westlund, 2014).

Types of Data Visualization

Wibke Weber and Hannes Rall conducted a study called *Data Visualization in Online Journalism and Its Implications for the Production Process* in 2012. They

interviewed different media companies about data visualization and found that "the crucial success factor . . .of data-based visualization in journalism is the attitude that everyone in the team acts as a journalist – no matter whether programmer, designer or statistician," (Weber & Rall, 2012). Statistical graphs and thematic cartography were both classical examples of visuals used to discover and display data (Weber & Rall, 2012).

Duke University Libraries and Angela Zoss created an easy to navigate guide of some of the many different types of data visualizations. 1D/Linear data visualizations are organized by a single feature and are usually a list of data rather than a visual. 2D/Planar data visualizations include visuals such as dot distribution maps, cartograms such as the Electoral College election maps and proportional symbol maps. From there, they get more complicated and include things such as computer simulations in 3D/Volumetric and things that use time as a visual (Zoss, 2017).

Data journalism today allows for the data to be analyzed and visualized with little knowledge of information technology (Weber & Rall, 2012). Interactive information graphics are visual representations of information with verbal and visual elements. The main characteristics of these graphics are interactivity and multimedia, and its value is based on its ability to display complex information (Weber & Rall, 2012). According to the different media companies the researchers interviewed, every person from reporters to programmers and designers sees him or her self as a journalist who belongs to the team in the newsroom and defines his tasks as journalistic tasks (Weber & Rall, 2012). This attitude allows for the collaboration that is necessary for creating data visualizations (Weber & Rall, 2012).

The challenge that needs to be addressed in the media is "how to create visually appealing images without compromising journalistic integrity?" (351) Data visualization requires time to collect data and to organize it, but time is the limiting factor for news coverage (Weber & Rall, 2012). In this big data age, accuracy must always come before speed (Weber & Rall, 2012). In an interview Charles Blow, a former graphics director, he said, "Show only what you know. Which is like we're not going to – if there's a breaking news of the Osama bin Laden's capture. We're not going to reproduce what happened inside that complex; we don't know it. But we know where it was, we know the shape of the complex, we can point to the general areas where we know that something happened, but we don't show that thing happening," (Blow quoted by Weber & Rall, 2012).

Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale argue against the use of the overarching phrase data visualization "without a clear consensus or discussion on what a visual data story encompasses,"(84) in their 2015 article *More Than Telling a Story: Transforming Data into Visually Shared Stories.* Data visualizations can include simple charts created from data, but are not visual data stories. These researchers narrow the definition of visual data stories as including a set of story pieces backed up by data that are presented in a meaningful order to support one or more intended messages (Lee, Riche, Isenberg, & Carpendale, 2015).

The Process of Data Visualization

Lee et al's proposed visual data storytelling process "summarizes the main roles and activities that visualization storytellers engage in as they turn raw data into a visually shared story, along with the types of artifacts that result from these activities," (Lee, Riche, Isenberg, & Carpendale, 2015). The first component of the visual data storytelling

process is exploring and analyzing data to choose a collection of data excerpts to create the story. Making the story is the second component and assembles the data into a storyline that is "interesting, illuminating, and compelling," (Lee, Riche, Isenberg, & Carpendale, 2015). In this step, it is important to draw connections between the different data pieces and formulate the message that will be conveyed through the visual data story. The third component is the process of delivering the story while keeping the target audience in mind. What the audience understands through the storytelling experience is the perceived story from the visual data story (Lee, Riche, Isenberg, & Carpendale, 2015).

In *Narrative Visualization: Telling Stories with Data*, Edward Segel and Jeffrey Heer investigate various case studies of narrative visualization and discuss how to use data to tell stories. These researchers talk about how storytellers, especially online journalists are "increasingly integrating complex visualization into their narratives" and the most sophisticated of these "focus on data exploration and analysis," (Heer & Segel, 2010). The case studies chosen for research mainly come from online journalism, and identify a few different techniques that can be applied when creating these visualizations. One is called visual highlighting and uses the color, size, or boldness of the visual elements to tell the viewer where to look. Multi-messaging is another tool that provides related but different information to the viewer in elements such as frames or panels. The presentation in multi-messaging guides the viewer through the data explaining certain patterns and highlighting key events. Single-frame interactivity pertains to the interactive items in a single frame without taking the viewer to a different "visual scene." Basically

it encourages the user to explore the data through the single view provided for them (Heer & Segel, 2010).

From these case studies, Herr and Segel also come up with seven genres of narrative visualization: "magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show, and film/video/animation," (Heer & Segel, 2010). All these genres can be used as building blocks, but each works better for different types of stories based on whether the story is author-driven or reader-driven. "A purely author-driven approach has a strict linear path through the visualization, relies heavily on messaging, and includes no interactivity," (Heer & Segel, 2010). "A purely reader-driven approach has no prescribed ordering of images, no messaging, and a high degree of interactivity," (Heer & Segel, 2010). Most narrative visualizations use a mix of the two.

In *Visualization Rhetoric: Framing Effects in Narrative Visualization*, Jessica Hullman and Nicholas Diakopoulos look into aspects of explorative and communicative visualization and the different techniques used to convey a story. Narrative information visualizations typically "convey an intended story" and "dialectic strategies aimed at providing the user with control over the insights she gains from interaction," (Diakopoulos & Hullman, 2011). "Researchers in InfoVis can benefit from a holistic understanding of visualization interpretation capable of providing insight into how particular interpretations arise as a result of interactions between a visualization, user mental models, and other external representations," (Diakopoulos & Hullman, 2011). Past investigations have shown evidence that "how data is framed or presented can significantly affect interpretation," (Diakopoulos & Hullman, 2011).

Some of the nomenclature used in narrative information visualizations includes terms such as bias, rhetoric, and framing. All these terms "describe how an interpretation arises from the interaction of representational, individual, and social forces," (Diakopoulos & Hullman, 2011). Bias in terms of narrative information visualization can be any other factors or preconceived notions an audience may have before viewing the visualization. Rhetoric is used "to refer to the set of processes by which intended meanings are represented in the visualization," (Diakopoulos & Hullman, 2011). And framing is how the topic is presented in the visualization. The framing theory is important in this discussion because "Information phrasing can influence interpretation in diverse ways," (Diakopoulos & Hullman, 2011).

Diakopoulos and Hullman also identify four editorial layers that can convey meaning. These layers are the data, visual representation, textual annotations, and interactivity, but it is also possible for omissions and ambiguity to be present at each level (Diakopoulos & Hullman, 2011). In terms of data, the creator of the visualization chooses what to represent, what variables to use, and which ones to omit. The visual representation is usually based on the abilities of human perception and also contains traces of data decisions. Textual, graphical, or social annotations focus the audience's attention on a specific area of the graph. The interactivity of the visual can constrain a viewer to explore certain subsets of data and ignore others (Diakopoulos & Hullman, 2011).

The researchers also gathered a sample of fifty-one professionally produced narrative visualizations to look into different visualization rhetoric techniques and how the decisions of visualization designers can manipulate an audience. The first decision is

what data will be represented, as well as the use of omission techniques to keep irrelevant information out of the visual. It is important to look out for misleading visuals that can obscure the data into creating a relationship that is not there. Typographic rhetoric, such as italicizing or bolding information, can also mislead the audience into drawing one certain conclusion from the data (Diakopoulos & Hullman, 2011).

"Viewing codes are the cultural, perceptual, cognitive, and psychological lenses that guide how an end-user interprets a representation," (Diakopoulos & Hullman, 2011). These viewing codes cannot be manipulated by the visualization and instead are a set of prejudices and preconceived notions that the audience has prior to viewing the data visualization. Cultural codes are the "social norms and wider beliefs of a culture," "Perceptual codes constrain what is salient to the user given human visual perception tendencies," and designers can target each of these to suggest a particular interpretation of the visualization (Diakopoulos & Hullman, 2011).

Joanna Wolfe in *Teaching Students to Focus on the Data in Data Visualization* discusses different interpretations of data and ways to visualize these interpretations. The interpretative level is basically the decision of how to present the data, for example using averages or percentages versus using raw counts (Wolfe, 2015). Wolfe quotes Perelman and Olbrechts-Tyteca noting "by foregrounding one interpretation, writers push other interpretations into the shadows," (Wolfe, 2015). The choices made about what data to present in the visualization can change the narrative of the story as well as the different patterns the data set could represent. Because of this, data visualization does not merely represent a given set of numbers; it involves selecting and rethinking the data on which the visualization is based. Graphics are able to distort data when they rely on the wrong

numbers or cause audiences to make misleading conclusions about the data. So data visualizers must reconsider the interpretive level being worked at and go back to the numbers and think about different way to represent them (Wolfe, 2015).

In the article *Visual Representation: Implications for Decision Making* Nicholas Lurie and Charlotte Mason discuss using data visualizations in marketing, but many of their points pertain to journalistic interactive data visualization as well. "Information visualization offers a way to shift cognitive load to the human perceptual system through graphics and animation," (Lurie & Mason, 2007, 160). There are two aspects of visual perspective that refer to how a given visual representations can change the relationship between the information and the audience. Interactivity is the user's ability to change the perspective and representation of the information. "By giving users increased control over the information flow, interactive visualization tools have important implication for decision making," (Lurie & Mason, 2007). With interactivity, viewing multiple factors rather than just a traditional report can reduce the effort required to make decisions. The other visual perspective is the depth of field, which "is likely to affect how information is accessed and evaluated," (Lurie & Mason, 2007). The depth of field can allow the audience to view specific data points, more context, or even the relationships between different data points. Another aspect of visual representations is the ease at which information can be assessed and compared. "Making information easier to compare is likely to lead to increased acquisition, weighting, and processing of this information," (Lurie & Mason, 2007).

Approach/Methodology

1. Does story mode (text, data visualization, text and data visualization) affect salience of the topic?
2. Does story mode affect whether people think their salience changed?
3. Does story mode affect perceptions of who has responsibility for vaccinations?

To answer the research questions, I created an online experiment using Qualtrics with three different randomly assigned treatments. One third of the subjects read a story from the Guardian on Measles and vaccinations in the United States. Another third viewed an online data visualization of the same topic. And the final third had both the data visualization and the story available for them to draw information from.

Using Angela Zoss and Duke University's classification of data visualizations, the one used for my study is a proportional symbol map, using colored circles to display the information. The proportion of red circles to yellow circles to blue circles shows how many kids are infected, susceptible, and vaccinated. Herr and Segel's narrative visualization genres would categorize the interactive data visualization as more of a flow chart, showing the affect of the herd immunity over time, but since it was a screenshot of the original visual, the interactive abilities of the graphic were omitted. The decision of what data will be represented and the omission of irrelevant data, as discussed by Diakopoulos and Hullman was solved by taking away the interactive abilities. This way, other information could not be explored and the viewer would not be as confused by the amount of data.

Interactivity was also discussed by Laurie and Mason and defined as the ability for the viewer to change the perspective of the data. By confining the viewer to one view

of the data, they could focus on the salience of one issue rather than multiple. The depth of field available to the viewer was much narrower and would not affect how the information is assessed and evaluated.

Before the data visualization or the story or both were presented, the survey participants were asked a series of questions about the topic to understand any bias they might have coming into the survey, as was discussed as viewing codes by Diakopoulos and Hullman. The viewing codes are the preconceived notions or prejudices that an audience has prior to viewing the data visualization. Each person came in with their own cultural codes, social norms, they grew up with, so the pre-exposure questions were important to understand these different thoughts and beliefs. The data visualization itself was also chosen because the way the information was presented, or the framing, seemed to be easier to interpret whether or not the story accompanied the visual or not. The graphical and textual annotations also don't force the viewers attention to one part of the visual more than another, as also discussed by Diakopoulos and Hullman, because all of the clusters of circles are the same size. The visual does not include its interactivity, but this ensures that all of the data is explored equally and the viewer does not accidentally ignore some of the data.

I had originally designed the project to be based on an interactive data visualization, however, the Guardian (and many other publications) does not open source the code for its visualizations, even for a study such as mine. Even with this drawback, the screenshot of this data visualization is relatively easy to understand and draw information from without the movement and it is still pertinent to study because of the unknown salience data visualizations could create about an issue.

After the reader gathered information via one of these three methods, he or she answered a series of questions based on the salience of the topic. I tested college students from the University of Oregon, and the survey was spread through professors at the School of Journalism and Communication as well as through social media. There were 76 respondents total, with 26 respondents exposed to the data visualization, 26 respondents exposed to the text story and 24 respondents exposed to both. close to equal numbers of responses from each treatment. From there I analyzed the data to decide whether or not one treatment made the issue appear more salient than the other.

Results:

During the pre-screening portion of the survey, before the information was displayed, there were no significant differences among the three groups of people who were exposed to the three conditions. This means that each group was equal going into the experiment in terms of their knowledge about vaccinations and the importance of vaccinations.

As outlined in Table 1, the group exposed to the text story only had a higher degree of opinion change than the people who were exposed to just the data visualization or the data visualization and the story. Additionally, the group exposed to just the text story thought that the importance of the issue was higher than the group that was exposed to both the data visualization and the story or just the data visualization. When asked the second time how important they thought the subject was, there was no significant difference in the data visualization only or the data visualization and the story, but there was a difference after the text story only.

Even though many survey participants said their opinions were not changed based on the information presented, their answers to the post survey questions show otherwise. Twelve different respondents changed their answer to the question of who is responsible for preventing outbreaks of contagious diseases after exposure to the data, but they said their opinions were not changed. Furthermore, four people from each treatment denied that their opinions of the importance of vaccines were changed.

The numbers for each option of the survey question, "who should be responsible for preventing outbreaks of contagious diseases" also changed after exposure for each treatment, but there was not a significant difference in the counts. It is important to note that each respondent could choose multiple answers for this question. Before exposure to the data visualization, 23 people said that doctors should be responsible, 24 people said that parents should be responsible, 22 people said individuals should be responsible and19 people said the government should be responsible. After exposure to the data visualization, the count for individuals to be responsible increased to 23 and the count for government to be responsible increased to 21. Before exposure to the text story, 21 people said that doctors should be responsible, 19 people said that parents should be responsible, 18 people said individuals should be responsible, and 18 people said the government should be responsible. After exposure to the text story, the count for doctors to be responsible decreased to 17, the count for parents to be responsible decreased to 18, the count for individuals to be responsible decreased to 16 and the count for government to be responsible decreased to 16. Before exposure to both the data visualization and the text story, 16 people said that doctors should be responsible, 19 people said that parents should be responsible, 21 people said individuals should be responsible and 15 people

said the government should be responsible. After exposure to both the data visualization and text story, the count for doctors to be responsible decreased to 14, the count for parents to be responsible decreased to 16, the count for individuals to be responsible decreased to 16, and the count for the government to be responsible decreased to 12.

The time spent on each treatment were not statistically different, however, the average was over a minute longer when the data visualization was present, in data visualization only and in data visualization and story.

Discussion:

Going into the study, I had thought that the groups exposed to the data visualization would have a higher amount of opinion change than the people exposed to the story because of the digital age students are living in, where they are used to reading shorter texts and seeing visualizations of information. However, the data shows that written stories are still relevant for people to gather information from and many prefer to do so. One interesting aspect of the data to note is that even though many survey participants said their opinions were not changed based on the information presented, their answers to the post survey questions show otherwise. This could be because people don't realize when their opinions change or don't care to admit it.

Another interesting note is that the average time spent on the survey when the data visualization was present was over a minute longer than the other treatments. This means that there is a kind of data visualization illiteracy among many people and if there is going to be an increase in the use of data visualizations to tell stories, there must first be a way to help people become data visualization literate. One way to test this would be

to use an interactive data visualization that walks a viewer through the information instead of having them draw conclusions from the data themselves. The main issue with this is the untold stories in the data that Wolfe warns about in her study, as discussed above (Wolfe, 2015).

Another question to consider is what would happen if the data were presented in a different way? Maybe the visualization could tell the story of one family who was affected by herd immunity or a community exposed to disease because other people refused to vaccinate their kids. This brings us back to the study done by Weber and Rall saying the importance of keeping journalistic integrity and only reporting what is known.

Expansion on this study would include conducting focus groups to understand how people are interacting with data visualizations. This is important to the future of data visualizations and journalism because "the goal of data journalism is to allow the public to analyze and draw understanding from data themselves, with the data journalist's role being to access and present the data on the public's behalf" (Coddington, 2014). By understanding how different people interact with the data, journalists will be able to create better visualizations and communicate in visually in a way their audience can understand.

Conclusion:

To test the research question "does data visualization in online media affect the salience of the story from the audience's perspective?" I created a survey with three different treatments to expose people to a data visualization, a data visualization and

story, or just a story. I tested 80 University of Oregon students and analyzed their answers to a series of questions based on salience.

The results showed that the group who was exposed to just the story had a greater amount of opinion change about the subject discussed and they also thought the issue was more salient than the groups exposed to either the data visualization or the data visualization and story. The amount of time each group spent on the survey could suggest data visualization illiteracy among people in general.

In answer to the research question does story mode affect the salience of the topic, no, the data visualization we showed did not increase the salience of the issue, however, further research using different visuals could change that. The story mode also didn't affect whether people thought their salience of the issue changed, but people were not always aware of their changing opinions. And finally, the story mode did not affect perceptions of who has responsibility for vaccinations, but I would argue that the information presented does.

For journalists and editors thinking about using data visualizations in the future, this research points out a few different aspects to be aware of and consider. To start, there needs to be context for the data visualization provided by labels or captions to tell the viewer how to interact with the data. The kind of data visualization illiteracy shown through this study needs to be acknowledged and shows that an audience will need to be taught or clearly instructed on how to interact with the data before they can start drawing their own conclusions from it. It is also important to understand any preconceived notions a viewer might have before engaging with data visualizations and know that those notions can affect how they interpret the data.

Tables and Charts:

Question	Score Guide	Both	DV Only	Story Only
1. How much do you know about vaccinations?	1= "A great deal" 2= "A lot" 3= "Some" 4= "A little" 5= "None at all"	3.125/0 .741	3.038/0.916	2.962/0.599
2. How much do you know about herd immunity?	1= "A great deal" 2= "A lot" 3= "Some" 4= "A little" 5= "None at all"	3.958/1 .160	3.808/1.1667	3.654/1.093
3. How important do you think it is for people to be vaccinated	1= "Extremely Important" 2= "Very Important" 3= "Moderately Important" 4= "Slightly Important" 5= "Not at all important"	1.458/0 .658	1.346/0.562	1.423/0.504

4. Based on the information presented, has your opinion of the importance of vaccines changed?	1 = "Yes" 2 = "No" 3 = "Maybe"	2.158/0 .501	1.923/0.392	1.462/0.989
5. Now, how important do you think it is for people to be vaccinated?	1= "Extremely Important" 2= "Very Important" 3= "Moderately Important" 4= "Slightly Important" 5= "Not at all important"	1.3/0.6 57	1.346/0.623	1/0.748

"Who should be responsible for preventing outbreaks of contagious diseases? (Check all that apply"

	Doctors	**Parents**	**Individuals**	**Government**	**Other**
Before Both	16	19	21	15	0
After Both	14	16	16	12	0

Before DV	23	24	22	19	1
After DV	23	24	23	21	1
Before Story	21	19	18	18	3
After Story	17	18	16	16	2

	Both	**DV Only**	**Story Only**
Time Spent On Survey (in seconds)	241.5/494.788	259.154/725.485	154.846/225.346

Appendix 1: Online Experiment

Interactive Data Visualization Thesis

Q1 How much do you know about vaccinations?

A great deal A lotSomeA little None at all

Q5 How much do you know about herd immunity?

A great deal A lotSomeA little None at all

Q2 How important do you think it is for people to be vaccinated?

Extremely important Very important Moderately important Slightly important Not at all important

Q3 Who should be responsible for preventing outbreaks of contagious diseases? (Check all that apply)

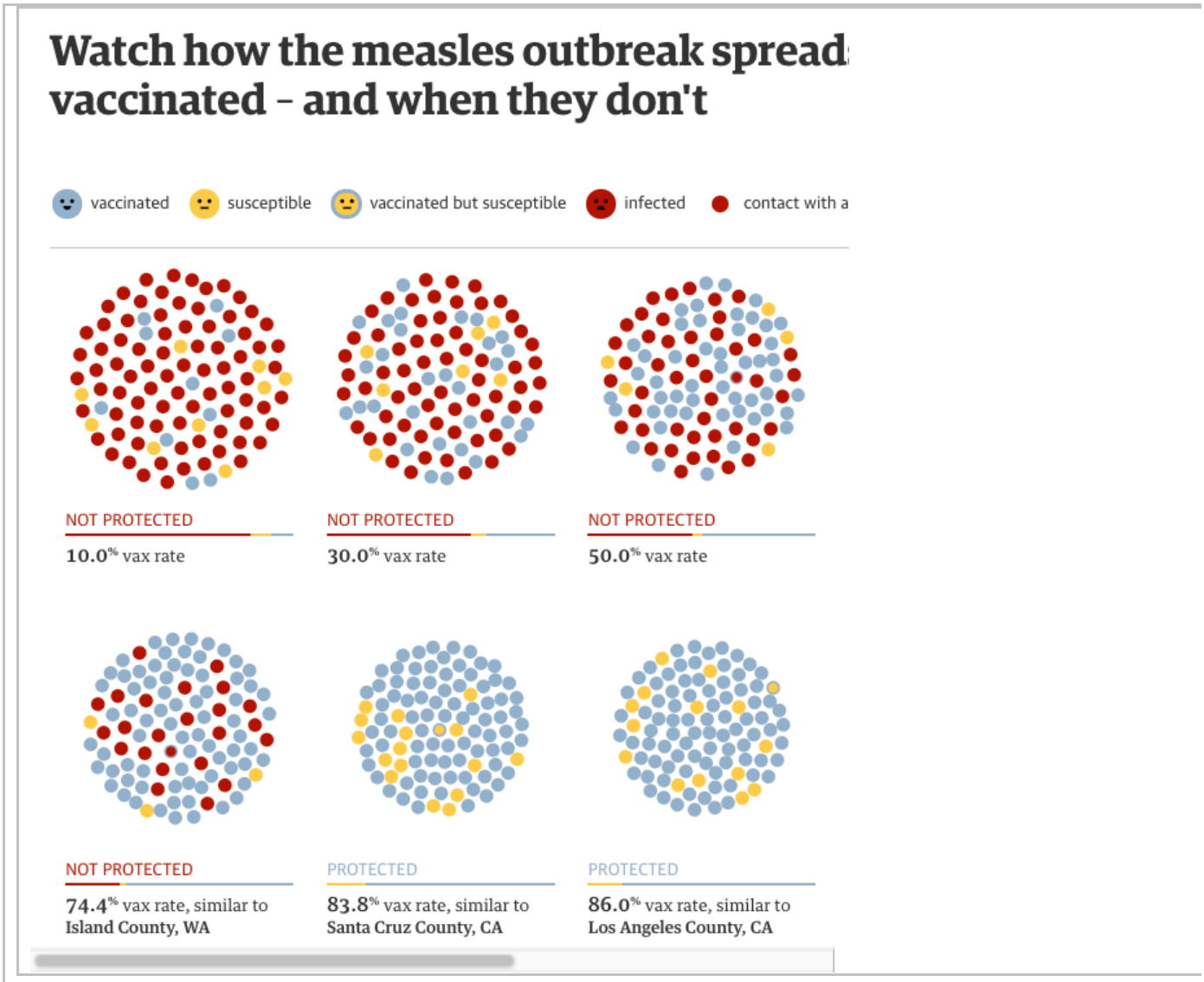
Doctors Parents Individuals Government Other

☐ ☐ ☐ ☐ ☐

Block 1

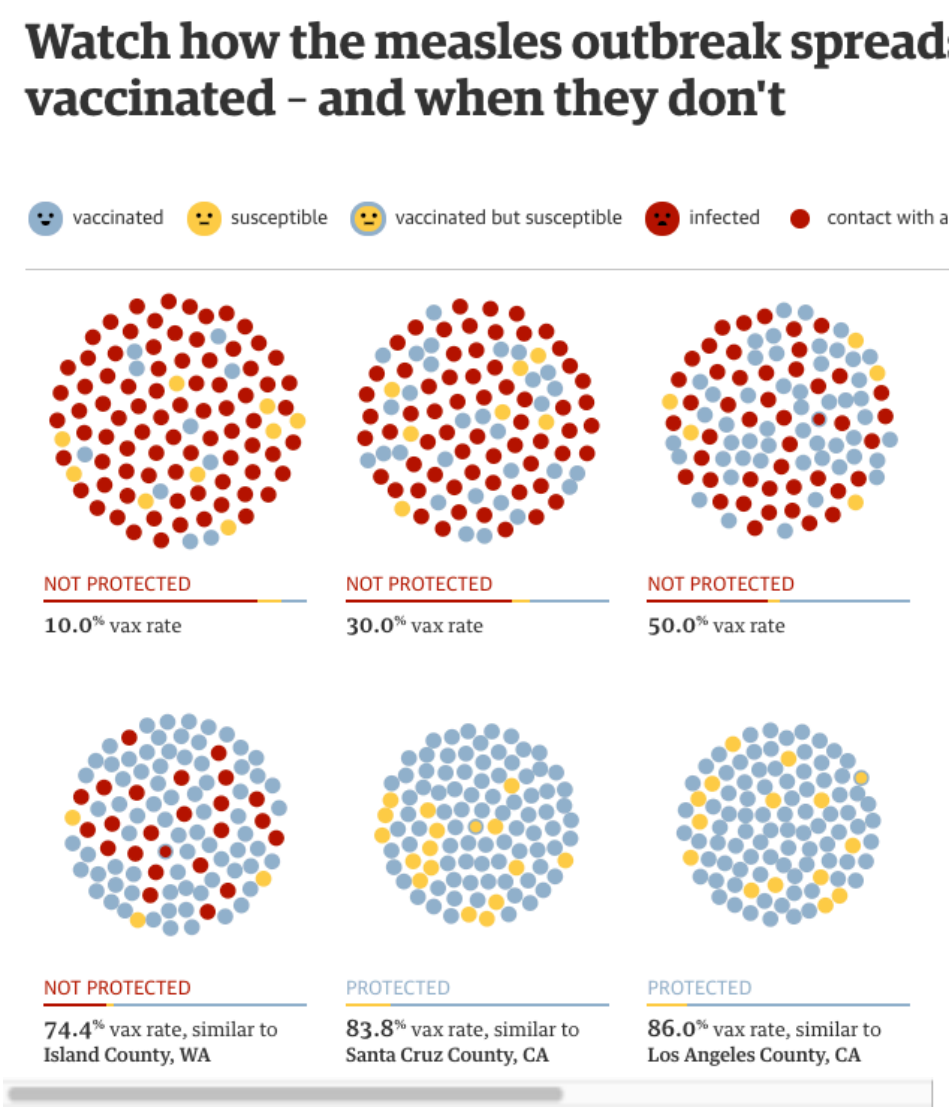
Please gather information from the graphic presented below:

Q4



Block 2

Please gather information from the graphic and story presented below:



Block 3

Measles is back in the US – and it's spreading. More than 100 cases across 14 states and

Washington DC have been confirmed by US health officials since an outbreak began at Disneyland last December. With a majority of those infections in unvaccinated people, widespread blame – from Washington to the rest of the world – has fallen on parents who chose not to vaccinate their children.

Part of the problem, according to Dr Elizabeth Edwards, professor of pediatrics and director of the Vanderbilt Vaccine Research Program, is just that: vaccination is understood by many as an individual choice, when science makes clear that the choice – to vaccinate or not to vaccinate – can affect an entire community.

“When you immunize your child, you’re not only immunizing your child. That child’s immunization is contributing to the control of the disease in the population,” Edwards explained.That sheltering effect is called herd immunity: a population that is highly immunized makes for a virus that can’t spread easily, providing protection to the community – or the herd – as a whole.

Despite the high overall measles vaccination rate in the US, vaccine skeptics – and their unimmunized kids – often congregate in like-minded communities, creating pockets of under-immunization.California, where the bulk of current measles cases can still be found, is a prime example.

It's one of 20 states that allow parents to skip vaccination based on their personal, philosophical beliefs – even though legislators introduced a bill on Wednesday that would ban such an opt-out provision.In populations where a large enough proportion of children are not immunized, everyone has a greater risk of catching the disease

But California remains home to communities with some of the highest vaccination opt-out rates in the country. Santa Cruz County, for example, has a personal belief exemption rate of 9.35%, nearly three times the state average. Some California school districts see exemption rates higher than 10%. That's enough to put a dent in herd immunity and fuel local outbreaks of measles.

Experts recommend that 92-95% of Americans be vaccinated against measles to protect everyone in the community, especially those who can't get the shot: babies under one year old, people born before the measles vaccine was introduced in 1963 and have never had measles themselves, and immunocompromised kids and adults like Rhett Krawitt, a young boy who recently went through chemotherapy.

That's a high threshold for herd immunity, but it's needed because of measles' extreme spreadability. As James Colgrove, professor of sociomedical sciences at Columbia University's Mailman School of Public Health, explained, "the more quickly a disease spreads, the higher level of herd immunity you need."Just how infectious is measles? The virus is highly airborne; it can stay on surfaces for up to two hours; and infectivity begins four days before a rash, so you can feel healthy but spread the disease. Measles is so contagious that "if one person has it, 90% of the people close to them who are not immune" – we'll call them susceptibles – "will also become infected," according to the CDC.Luckily, the measles vaccine – administered in the form of the MMR for measles, mumps and rubella – is very effective. If delivered fully (two doses), it will protect 99% of people against the disease. But, like all vaccines, it's not perfect: 1% of cases are likely to result in vaccine failure, meaning recipients won't develop an immune response to the given disease, leaving them vulnerable. Even with perfect vaccination, one of every 100 people would be susceptible to measles, but that's much better than the alternative.

The bottom line: in populations where a large enough proportion of children are not immunized, everyone has a greater risk of catching the disease – the unprotected, but also those who are vaccinated, Edwards told the Guardian."You're putting other children at risk by deciding not to immunize your own," she added.

Block 4

Please gather information from the story presented below:
Measles is back in the US – and it's spreading. More than 100 cases across 14 states and

Washington DC have been confirmed by US health officials since an outbreak began at Disneyland last December. With a majority of those infections in unvaccinated people, widespread blame – from Washington to the rest of the world – has fallen on parents who chose not to vaccinate their children.

Part of the problem, according to Dr Elizabeth Edwards, professor of pediatrics and director of the Vanderbilt Vaccine Research Program, is just that: vaccination is understood by many as an individual

choice, when science makes clear that the choice – to vaccinate or not to vaccinate – can affect an entire community.

“When you immunize your child, you’re not only immunizing your child. That child’s immunization is contributing to the control of the disease in the population,” Edwards explained.That sheltering effect is called herd immunity: a population that is highly immunized makes for a virus that can’t spread easily, providing protection to the community – or the herd – as a whole.

Despite the high overall measles vaccination rate in the US, vaccine skeptics – and their unimmunized kids – often congregate in like-minded communities, creating pockets of under-immunization.California, where the bulk of current measles cases can still be found, is a prime example.

It's one of 20 states that allow parents to skip vaccination based on their personal, philosophical beliefs – even though legislators introduced a bill on Wednesday that would ban such an opt-out provision.In populations where a large enough proportion of children are not immunized, everyone has a greater risk of catching the disease

But California remains home to communities with some of the highest vaccination opt-out rates in the country. Santa Cruz County, for example, has a personal belief exemption rate of 9.35%, nearly three times the state average. Some California school districts see exemption rates higher than 10%. That’s enough to put a dent in herd immunity and fuel local outbreaks of measles.

Experts recommend that 92-95% of Americans be vaccinated against measles to protect everyone in the community, especially those who can’t get the shot: babies under one year old, people born before the measles vaccine was introduced in 1963 and have never had measles themselves, and immunocompromised kids and adults like Rhett Krawitt, a young boy who recently went through chemotherapy.

That’s a high threshold for herd immunity, but it’s needed because of measles’ extreme spreadability. As James Colgrove, professor of sociomedical sciences at Columbia University’s Mailman School of Public Health, explained, “the more quickly a disease spreads, the higher level of herd immunity you need.”Just how infectious is measles? The virus is highly airborne; it can stay on surfaces for up to two hours; and infectivity begins four days before a rash, so you can feel healthy but spread the disease. Measles is so contagious that “if one person has it, 90% of the people close to them who are not immune” – we’ll call them susceptibles – “will also become infected,” according to the CDC.Luckily, the measles vaccine – administered in the form of the MMR for measles, mumps and rubella – is very effective. If delivered fully (two doses), it will protect 99% of people against the disease. But, like all vaccines, it’s not perfect: 1% of cases are likely to result in vaccine failure, meaning recipients won’t develop an immune response to the given disease, leaving them vulnerable. Even with perfect vaccination, one of every 100 people would be susceptible to measles, but that’s much better than the alternative.

The bottom line: in populations where a large enough proportion of children are not immunized, everyone has a greater risk of catching the disease – the unprotected, but also those who are vaccinated, Edwards told the Guardian.“You’re putting other children at risk by deciding not to immunize your own,” she added.

Q9 Based on the information presented, has your opinion of the importance of vaccines changed?

Yes Maybe No

Q10

Extremely important Very important

Moderately important Slightly important Not at all important

Now, how important do you think it is for people to be vaccinated?

Q11 Who should be responsible for preventing outbreaks of contagious diseases? (Check all that apply)

Doctors Parents Individuals Government Other

Any other comments or opinions about vaccinations, herd immunity or anything just

Q12 presented?

Q16

Freshman Sophomore

Junior Senior Other

What year are you?

Add Block

Q17

Which gender do you identify with?

MaleFemale Trans-gender Other

Q18

What is your major?

Story source: https://www.theguardian.com/society/ng-interactive/2015/feb/05/-sp-watch-how-measles-outbreak-spreads-when-kids-get-vaccinated

References

"Why Visual Teaching." *Visual Teaching Alliance*. Professional Development for Primary, Secondary & University Educators/Administrators, 2008. Web. 04 Dec. 2016. <http://visualteachingalliance.com/>.

Coddington, Mark. "Clarifying Journalism's Quantitative Turn." *Digital Journalism* 3.3 (2014): 331-48.

Hullman, J., and N. Diakopoulos. "Visualization Rhetoric: Framing Effects in Narrative Visualization." *IEEE Transactions on Visualization and Computer Graphics* 17.12 (2011): 2231-240.

Lee, Bongshin, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. "More Than Telling a Story: Transforming Data into Visually Shared Stories." *IEEE Computer Graphics and Applications* 35.5 (2015): 84-90.

Lewis, Seth C., and Oscar Westlund. "Big Data and Journalism." *Digital Journalism* 3.3 (2014): 447-66.

Lurie, Nicholas H., and Charlotte H. Mason. "Visual Representation: Implications for Decision Making." *Journal of Marketing* 71.1 (2007): 160-77.

Segel, E., and J. Heer. "Narrative Visualization: Telling Stories with Data." *IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010): 1139-148.

"SPJ Code of Ethics." *SPJ Code of Ethics | Society of Professional Journalists | Improving and Protecting Journalism since 1909*. Society of Professional Journalists, 6 Sept. 2014. Web. <http://www.spj.org/ethicscode.asp>.

Weber, Wibke, and Hannes Rall. "Data Visualization in Online Journalism and Its Implications for the Production Process." *IEEE Xplore*. International Conference on Information Visualisation, 2012. Web. 16 May 2016.

Wolfe, J. "Teaching Students to Focus on the Data in Data Visualization." *Journal of Business and Technical Communication* 29.3 (2015): 344-59.

"What Is Big Data? - Gartner IT Glossary - Big Data." *Gartner IT Glossary*. 19 Dec. 2016. Web. <http://www.gartner.com/it-glossary/big-data/>.

Zoss, Angela. "Introduction to Data Visualization: Visualization Types." *Duke University Libraries*. Duke University Libraries, 1 Feb. 2017. Web. <http://guides.library.duke.edu/datavis/vis_types>.