Data Journalism, Data Literacy and Data Visualizations: A Quantitative Study

A Project
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree
Master of Journalism

by STEVEN PERRIN
Dr. Kathleen Rose, Project Advisor
NOVEMBER 2022

Acknowledgements

I would like to thank my Advisor, Dr. Kathleen Rose. This degree has taken over 13 years and many false starts to come to fruition. I know firsthand the importance of a dedicated and supportive advisor, and I can say with great certainty that I would not have finished without Dr. Rose's expertise and encouragement.

I would also like to thank my committee members, Dr. Beverly Horvit and Dr. Zachary Massey. Thank you both for all the support along the way.

Finally, I would like to thank my wife, Suzanne Greenberg, for encouraging me to finally complete this degree.

Table of Contents

List of Tables

Abstract

As data becomes increasingly important in contemporary society, data journalism and data literacy also become more important. This project explores these concepts and examines the role each can play in writing about and understanding data intensive information. To test the effects of data visualizations and data literacy on comprehension, this project uses a quantitative experimental design where subjects read different versions of an article followed by a comprehension test. The article treatments include a text-only version, a version with a bar graph and a version with a data table. In addition, subjects were classified as data literate and non-data literate based on a survey. As hypothesized, the results showed a significant comprehension benefit for both groups of subjects with access to a data visualization, with the text-only group scoring lowest in comprehension. The results also showed significant comprehension differences based on data literacy in the bar graph test condition. These results can be used to inform future study, as well as to inform best practices in data journalism and in data science education.

Keywords: data journalism, data literacy, data visualization, data science education, statistics education, multimedia learning

Data Journalism, Data Literacy and Data Visualizations: A quantitative study

Introduction

The world is in the midst of what the United Nations Secretary-General's Independent Expert Advisory Group has called a "data revolution" (2014, p. 2). Exponential increases in data volume, new data analysis technologies and new kinds of data have the potential to transform society, but inequalities in access to data also present major equity challenges (UN Data Revolution Group, 2014). Massive amounts of data available online, sophisticated data analysis software, and the tools to publish all this information allow "more people to work with more data more easily than ever before" (Bounegru, 2021, p. 3).

The amount of data generated worldwide in 2010 was 2 zettabytes; by 2022 it is estimated that will increase to 97 zettabytes (Statista, 2022). This growing mountain of data, in turn, has exposed the importance within journalism of analyzing, interpreting and presenting this data (data journalism) and the importance within society to better understand it (data literacy).

From my experiences as a journalist and high school data science teacher, it occurs to me that a common goal of both journalist and teacher is to effectively engage and convey information to an audience. The similarities are even more pronounced in the specific realms of data journalism and data literacy. The increased societal importance of data in recent years has, in turn, been reflected in both the newsroom and the classroom. Data journalism (the analysis of data to uncover newsworthy information and the creation of data visualizations to convey information to readers) is of ever-increasing importance to news organizations (Kennedy et al., 2021). Meanwhile, data science education is being added at all levels of the education system, from universities to elementary schools (Martinez & LaLonde, 2020). These parallel developments are important and symbiotic, as the promise of data journalism will not be fully realized without increased data literacy in the broader population. Data journalists play a vital role in interpreting data and making it accessible to a wider audience (Bounegru, 2021). More and better data journalism will not only inform readers but will also increase their data literacy.

This is in keeping with the broader goal of journalism in general. “The purpose of journalism is to give people the information they need to make better decisions about their lives and society” (Kovach & Rosenstiel, 2021, p. 7). Some of that needed information is now hidden within a vast expanse of data. Journalists and their audiences may need to co-evolve to deal with the situation – data journalists developing creative ways to present complex information, while readers develop new skills for interpreting the information. Ideally these factors can form a positive feedback loop – journalists providing more useful information, audiences developing more data literacy and the capability to handle still more complex information going forward.

“A picture is worth a thousand words” is an adage in multiple languages, dating to the early 20th century and even earlier than that in similar forms (Rogers, 1985). It is such a well-known saying in part because it feels obvious, and journalists and teachers alike often rely on words and pictures to impart meaning.

It has long been proposed that student comprehension is increased when instruction addresses multiple learning modalities, including words and pictures. Emphasizing different learning styles, such as those in the VARK model (visual, auditory, read-write and kinesthetic learners) has been prevalent in pedagogy for many years (Cherry, 2019). Similarly, Mayer's (1997) Multimedia Learning Theory posits that learners comprehend better when information is delivered through words and images than when the information is conveyed with words alone.

Data journalism (or at least some aspects of data journalism) relies on a similar premise. If people comprehend better through words and pictures, then it follows that a news article accompanied by charts or graphs may be more informative and more compelling than the article alone.

The primary intent of this project is to examine whether a data visualization element increased comprehension of data intensive information. To explore this question, I conducted an experiment using different versions of a data intensive news article – one version with text only, one version with text and a bar graph, and one with text and a data table. My main hypothesis is that the subjects reading the article with accompanying data visualizations will demonstrate greater comprehension of the information than those with the text-only version.

In addition, the project explores potential relationships between increased data literacy and comprehension of data intensive information. In a data intensive world, data literacy is an increasingly important skill, and data journalists must understand their audiences’ level of data literacy to effectively communicate information.

Literature Review

This project examines the basic premise of Multimedia Learning Theory, that comprehension is improved through words and images as compared to words alone, as well as the role of data literacy in comprehension of data intensive information. This research is important to data journalism as it addresses one of the basic tenets of the field, that the addition of data visualizations makes information more accessible and comprehensible. The project also underscores the importance of a data literate society, addressing questions about the comprehension benefits for data visualizations in data intensive information and the effects of educational background in data literacy.

Is a picture "worth a thousand words?" Do pictures in fact add measurable value in terms of comprehension? There is a vast body of research concerning words and pictures as it relates to communication and persuasion. In the area of health communication, incorporating images with text has shown to be more persuasive than words alone as it pertains to warning labels (e.g., Kees et al., 2010; Lochbuehler et al., 2019) or combating disinformation (Dixon et al., 2015). This research has found that the images need to be appropriate to the message and congruent with the message (Lochbuehler et al., 2019) to be most effective.

As for the comprehension of new material, in education or journalism, certain ideas seem obvious, like the adage that "A picture is worth a thousand words." One such notion is that individuals have different styles of learning and learn best when the instruction is delivered in their preferred style. Another notion is that information will be retained more readily when it is conveyed with both words and pictures. However, the empirical evidence for these assumptions is mixed in the literature (e.g., Coleman et al., 2018; Guo et al., 2020; McTigue, 2009; Segers et al., 2008; Vekiri, 2002), which tends to show that the right picture and the right audience are both necessary to achieve the anticipated comprehension benefits.

My project requires a review of data journalism and data visualizations to help define the concepts, as well as a review of the importance of data literacy. This review will also examine the existing literature on multimedia learning and learning styles and, finally, the research on visual displays.

Data Journalism and Data Visualizations

Data journalism is a broad term that has been used to refer to many different things. The use of large data sets, public records, analysis tools and visual presentation of data are all aspects of data journalism. Indeed, writing in *The Data Journalism Handbook*, Cohen (2021) includes all of the above areas of expertise as elements of the discipline, and goes so far as to define data journalism as “anything in a newsroom that requires the use of numbers, or … computer programming” (p. 1) – which is quite a lot.

One specific element of data journalism in the 21st century is uncovering the story hidden in the data: The availability of public data sets and the tools to analyze them make it possible to find correlations, significance, patterns and important information that were previously hidden. One prominent example of this form of data journalism would be *The Guardian*’s analysis of Edward Snowden’s leaked National Security Agency (NSA) files (MacAskill & Dance, 2013). The massive volume of data in the NSA leaks precluded a simple understanding of any significance. Data journalists at *The Guardian* used analysis and visualizations to add vital context to the data, making important information accessible to the public, and their work helped to popularize the term "data journalism" (Bounegru, 2021).

After the important work of analysis, the data journalist needs to do something equally important: Present the information to the reader. That can be accomplished through words alone, but it is also increasingly done through data visualizations (Kennedy et al., 2021). This data journalism is at the same time traditional, telling a story with words and pictures, and modern, utilizing interactive graphics or new data visualizations like heat maps where data values are represented by colors. This form of data journalism involves the presentation of data and information in a manner that attempts to be more engaging and informative. For the purposes of this project, we will focus on this narrower example of data journalism: The inclusion of data visualizations within text articles for the purpose of enhancing understanding.

The project defines a data visualization as "a visual representation of data, created to amplify the cognitive processing and social application of data represented" (Engebretsen et al., 2018, p. 4). In interviews with data journalists in European news organizations, interviewees told Kennedy et al. (2021) that the use of data visualizations is growing and that they are becoming "central elements in journalistic storytelling" (p. 1). The same series of interviews found that a data visualization can be "the driving force of a story, even when it is a simple graphic or diagram" (p. 1). In other words, data visualizations are increasing both an use and in importance.

Data visualizations appear to increase engagement, as well. Through reader statistics, news organizations have determined that the inclusion of a data visualization causes readers to remain on a page longer (Kennedy et al., 2021). This additional engagement indicates a reader who is cognitively active, one who is actively trying to make sense of information (Mayer, 2020).

Segel and Heer (2010) examined data journalism and data visualizations to identify the crucial elements in telling a story with data. Among their findings were that the uses of data visualizations should be limited and strategic: "Data stories appear to be most effective when they have constrained interaction at various checkpoints within a narrative, allowing the user to explore the data without veering too far from the intended narrative" (Segal & Heer, 2010, p. 1147). The implication is that too much data, or data that is not directly related to the narrative story, is ineffective.

So, while there are many possible forms and definitions of data journalism, it is this telling of a story through data or supplemented with data that is my emphasis in this project. If a visualization does indeed cause the reader to "stay on the page a little longer" (Kennedy et al., 2021, p. 1) does that help improve comprehension of the information? Does the visualization itself contribute additional comprehension?

Data Literacy

Ridsdale et al. (2015) define data literacy as “the ability to collect, manage, evaluate, and apply data, in a critical manner” (p. 3). If literacy is to evaluate and interpret the significance of words, then data literacy is to do the same with data. The promise of data journalism is at least partially tied to the data literacy of the readership lest the information be lost. The goal of a data visualization is to convey information quickly, but even the most effective visualization requires a level of familiarity on the part of the reader – some level of data literacy. The proliferation of data in the modern world has led not only to a rise in data journalism, but also to a need for increased data literacy in society. Frank et al. (2016) equate the increased importance of data literacy in the internet age to other types of literacy. In a sense, data literacy is to data journalism as literacy is to journalism.

Though a lack of data literacy among many in society is by no means a new problem, the COVID pandemic has shone a light on the issue. As any consumer of news in 2020 saw, “In the pandemic, data are everywhere” (Nguyen, 2021, p. 210). Has there ever been a line graph more prevalent across society than the one stressing the need to “flatten the curve”? However, the ability to comprehend an x-axis of time and a y-axis of cases might have tested the limits of society’s data literacy. It is clear at this point that few understood the data or the implications behind *keeping* the curve flat, and a logarithmic scale on a y-axis was almost certainly expecting too much of the audience (Nguyen, 2021). The need for data literacy was exposed during the pandemic, but there are also limits to the level of competency journalists can expect the public to possess in this area.

As the importance of data literacy grows, the education system has attempted to respond and help develop this new competency with data science curriculum (as distinct from more traditional statistics courses). Colleges and universities have offered data science classes for some time, but there is also an active movement to provide data literacy education in secondary and even elementary schools (Wolff et al., 2016).

D'Ignazio (2017) views the challenge of increased data literacy as a broader societal problem that goes beyond the need for additional education in schools. As with other opportunity gaps that leave some in society at a disadvantage, D'Ignazio (2017) argues that without increased data literacy, societies will see an ever-increasing "gap between the data haves and data have-nots" (p. 6).

To address this need for a higher level of data literacy in society, some educators feel that it is past time to reexamine mathematics education in the US to deemphasize pure math such as calculus in favor of more statistics education (Benjamin, 2009). Some school districts are taking concrete steps to include data science courses. Los Angeles Unified, the second largest school district in the nation, offers an Introduction to Data Science course, which is now recognized by the University of California system as a statistics course that can be substituted for Algebra 2 as an admissions requirement (Boaler & Levitt, 2019).

It is perhaps to early to know if these updates to the education system are successfully increasing data literacy in American society, but the effort is ongoing.

Multimedia Learning

A fundamental premise of data journalism is that including visualizations with text is beneficial to comprehension. Educational research supports this premise, particularly Mayer's Multimedia Learning Theory (2020). Mayer's Multimedia Principle states that "people learn better from words and pictures than from words alone" (p. 4). This is the primary hypothesis this project will examine: that the inclusion of pictures with words will increase comprehension as compared to words alone.

Modality has always been an important concept within journalism. Valkenburg (2016) lists modality as a fundamental property, pointing out that it has long been common to examine the effects of different presentation modes on audiences. The seminal media theorist McLuhan coined the expression "the medium is the message" in 1964 to emphasize the preeminence of modality in media.

The pertinent question to this research is not one of specific modality, but the potential for increasing comprehension through conveying information both through words and through visuals, engaging multiple modalities with the same message.

Several theories inform this idea. Dual coding theory (Paivio, 1986) posits that the brain has one method for processing images and a separate method for processing words. Comprehension is thought to be reinforced when it is processed by both methods. Visual argument theory (Waller, 1981) holds that graphical representations are less taxing to process than text because, as compared to texts, graphics remove a layer of transformation involved in assigning meaning to the words. Building on these two theories, Mayer (2020) developed the Multimedia Learning Theory that states, “Human understanding occurs when learners are able to mentally integrate corresponding pictorial and verbal representations” (p. 7). Put simply, Mayer believes that words (verbal representations) and images (pictorial representations) convey information more effectively than words alone.

While Mayer’s theory was developed addressing the pedagogical implications of multimedia learning, the similarities between a teacher presenting information to students and a journalist presenting information to readers imply that Multimedia Learning Theory is applicable. Whether the “learner” is a newspaper reader or a student in a classroom, the premise of Multimedia Learning Theory is the same: Include a picture (or a chart, graph, or data table) with the text to help convey information. For the purposes of this research, we will use the lens of Multimedia Learning Theory.

A prevailing view in pedagogy holds that students learn best when information is tailored to their preferred learning-style. This view is influential in the education field and examples of its importance can be found at all levels of education (Pashler et al., 2008). However, multiple reviews of learning-styles research have found little evidence to support the validity of the *meshing* *hypothesis*, which holds that comprehension is highest when instruction is tailored to mesh with the predominant or preferred learning style of the individual student (e.g., Kirschner, 2017; Pashler et al., 2008). In fact, among studies that have attempted to measure the effect, it was more common that the data contradicted the meshing hypothesis (Pashler et al., 2008).

The insufficiency of the meshing hypothesis of learning styles does not, however, invalidate the principle of Multimedia Learning Theory as students benefit from different kinds of instruction that address multiple modalities of learning (Yale Poorvu Center, n.d.*)*. So, while there is little empirical evidence that adapting instruction to a specific learning style is effective, that does not preclude the premise that addressing multiple modalities will be effective for most learners.

Research on Visual Displays

The research regarding the impact of visual displays in education is inconclusive considering the conventional wisdom that the inclusion of visuals should always provide a comprehension benefit. In their meta-analysis on the use of visuals, Guo et al. (2020) concluded that among the 23 studies they examined six failed to find a positive comprehension effect with the inclusion of visuals while the other 17 studies reported varying degrees of benefit when including visuals.

The effectiveness of increasing comprehension by adding visual information to text seems to vary by age group as well as by visual complexity and expertise in the specific discipline (e.g., Elia et al., 2007; McTigue, 2009;

Renkl & Scheiter, 2017). For example, a simple picture depicting the events in a story can aid comprehension in young readers (Segers et al., 2008). However, in one study the arrows on a water-cycle diagram were often misinterpreted by elementary school students as simply pointing at elements of the diagram as opposed to indicating a sequence of events (McTigue & Flowers, 2011). Renkl and Scheiter (2017) found that insufficient content knowledge and lack of competence in representational models were barriers to achieving the desired learning outcomes of some visual displays.

In an experiment involving science material, middle school students were found to have little or no additional comprehension when given a text with accompanying diagrams compared to those given the text only (McTigue, 2009). In other studies, students given information with text and images sometimes understand less than those who receive text-only versions of the information (e.g., Coleman et al., 2018; Vekiri, 2002). This may be due to the additional information overwhelming the students' processing capacity. In another experimental study, fourth graders were given a science text with three different kinds of diagrams and a control with no diagram: The researchers found that the diagrams added minimal or no additional comprehension as compared to the text-only version and that the diagrams created a cognitive overload for some students (Coleman et al., 2018). These results imply that simply adding charts and graphs without consideration of the context and the level of the audience will not provide a comprehension benefit and may be detrimental to understanding.

In other words, contrary to the effect proposed by Multimedia Learning Theory, presenting information through both visuals and text can sometimes impede comprehension. Multiple forms of information should be beneficial if they can be processed but are problematic if the information is overwhelming.

For these reasons the research on visual learning is mixed, showing that the inclusion of visuals does not consistently provide the anticipated comprehension benefits (e.g., Coleman et al., 2018; McTigue, 2009; Renkl & Scheiter, 2017). In cases where the comprehension benefits are missing or limited, researchers have surmised that a lack of familiarity with the subject matter, a lack of expertise in interpreting the visuals or cognitive overload may be limiting factors (e.g., McTigue, 2009; Renkl & Scheiter, 2017, Vekiri, 2002). These findings suggest the necessity for audiences to have data literacy skills for data journalism to achieve its broadest possible impact.

It seems reasonable that prior training on the proper interpretation of specific visualizations will impact the effective use of visualizations to enhance comprehension. In this project, I test for comprehension benefits of the inclusion of a data visualization across subjects with varying prior education in statistics and/or data science (see Research Questions and Hypotheses below).

Finally, prior research has specifically addressed how incorporating data visualizations in journalism can impact reader comprehension. Ward (1992) used five different versions of a news article – the article alone, with a bar graph, with an illustrated bar graph, with a data table, and with a sidebar story – to test comprehension in college students. The experiment showed that the students retained information best with the sidebar story, and worst with the article alone. The two graphical representations and the data table did positively impact information retention (Ward, 1992). More recently, a similar study (Parodi & Julio, 2017) examined the importance of subject matter expertise (in this case, economics) in a college student's ability to gain a comprehension benefit from the inclusion of graphics in reading material, finding that the role of disciplinary knowledge was crucial in comprehending a specialized text and specific data visualizations.

See Table 1 for a summary of studies examining comprehension benefits of visual displays, and the potential limiting factors identified.

Table 1 *Summary of comprehension effects and limiting factors*

	Comprehension benefit	Test subject grade(s)	*Lack of Prior Knowledge*	*Overload*	*Cognitive development*	
Study			*Limiting factors*			Display type(s)
Mayer & Gallini (1990)	Yes	UG				diagrams
Ward (1992)	Yes	UG				bar graph
Mayer (1997)	Yes	UG				diagrams
Segers, Verhoeven & Hulstijn-Hendrikse (2008)	Yes	5				pictures
Elia, Gagatsis & Demetriou (2007)	Mixed	1,2,3			x	pictures, number lines
McTigue (2009)	Mixed	6,7,8		x	x	diagrams
Roberts, Norman, & Cocco (2015)	Mixed	3	x			diagram, flowcharts, tables and timelines
Parodi & Julio (2017)	Mixed	UG	x			graphs
Roberts & Brugar (2017)	Mixed	3,4,5	x		x	images, maps, tables, and timelines
Coleman, McTigue & Dantzler (2018)	Mixed	4		x	x	diagrams

Purpose

The overall purpose of this research is to test the impacts of addressing multiple learning modalities with multimedia learning techniques on comprehension of data intensive information. This research provides empirical evidence for data journalists as to the benefits of including data visualizations to convey data intensive information to readers.

The research also examined the role of data literacy in the comprehension of data intensive information. As data becomes more and more important in the modern world, data literacy must keep pace. Does prior education in statistics and/or data science increase comprehension of data intensive information?

The project framed the research through the lens of Richard Mayer's Multimedia Learning Theory (1997) and examines:

- The basic validity of Multimedia Learning Theory (Mayer, 1997) – that as compared to words alone, comprehension is improved through words and graphs as would be present in data journalism.
- The role of data literacy in comprehension of data intensive information.

Prior research has often shown the comprehension benefits of including visuals with text, which this project hopes to replicate. However, where the results of earlier studies have been mixed, the absence of a comprehension benefit is often attributed to a lack of specific prior knowledge. This study helps to fill this research gap, examining general data literacy as a necessary component for achieving comprehension benefits from the inclusion of data visualizations with data intensive text.

Hypotheses

This research is important to data journalism as it addresses one of the basic tenets of the field, that the addition of data visualizations makes information more accessible and comprehensible. The project underscores the importance of a data literate society as I address questions about the comprehension benefits for data visualizations in data intensive information and the effects of educational background in data literacy.

Based on the prior research conducted in this area, I proposed the following hypotheses:

H_1: Subjects with a data visualization supplementing their reading material will demonstrate better comprehension of the material than those without a data visualization.

H_2: Subjects who have had a course in statistics and/or data science will demonstrate better comprehension of the material than those who have not had statistics and/or data science.

H_3: Subjects with higher levels of data literacy (as measured through a data literacy test) will demonstrate better comprehension of the materials than those who have lower data literacy scores.

Analysis and Findings

Methods

To address the effects of data visualization and data literacy on information comprehension, I gathered data using an online experiment conducted with a final sample of N=286 undergraduate students from a large public university in the Midwest.

The experiment consisted of three conditions manipulating the inclusion of a data visualization in a text-based article. The different versions of the article presented to the subjects were as follows:

- Version 1 – Text-only
- Version 2 – Text with a bar graph
- Version 3 – Text with a data table

Within this sample of students, participants were randomly and evenly assigned to one of the three reading conditions. The complete survey, including all three versions of the reading, is included in Appendix A. The original article is from the New York Times (Zraick, 2019) and is about American teens and the issues that concern them. The original article features research from the Pew Research Center and includes a bar graph of teens' views on a variety of problems they face. All survey versions of the article were edited for length. Version 1 was edited to remove references to a bar graph and the bar graph itself. Version 1 was also edited to include enough information to answer all the comprehension questions without the graph or the data table. Version 1 is 257 words long. Versions 2 and 3 were edited only for length (254 words). Version 2 retained the original bar graph, while Version 3 replaced the bar graph with a sorted table of data.

As defined earlier, a data visualization is simply “a visual representation of data, created to amplify the cognitive processing and social application of data represented” (Engebretsen et al., 2018, p. 4). While the bar graph includes graphical elements like magnitude of a bar, a sorted list of tabular data is also a visual way of representing multiple points of data.

The survey was delivered to students online, via their smartphone or a web browser. Subjects were asked to read the assigned version of the article and prompted to proceed to the quiz questions after they completed the reading. They were not allowed to return to the article once they proceeded. Time spent on the reading was recorded. Identical quizzes to determine comprehension of the material were administered to each group of subjects.

The dependent variable, comprehension, was determined by a test based on the material covered in the articles (see Measures below). Subjects were also asked to take a data literacy pre-test consisting of three questions measuring how comfortable they are with data

Collecting and Preparing the Data

Prior to launching the final survey, a pilot survey was made available to a small number of students (N=15) during the summer 2022 session at a large public university in the Midwest. Owing to the small total size of the pilot, only the text-only condition and the bar graph condition were included in the summer pilot. The pilot was used as a proof of concept to ensure that 1) the Qualtrics survey was coded correctly and that students could access it properly and 2) there were no issues with the experimental conditions themselves. The pilot revealed no major issues with the survey and experimental conditions and also suggested there were meaningful differences in comprehension based on experimental conditions, albeit with a very small sample size.

For the full experiment, I distributed the research survey among multiple sections of Journalism and Math courses during the first 8-week session of the Fall 2022 semester at a large public university in the Midwest.

The survey was active from August 30, 2022, through October 13, 2022. During that time the survey was distributed in 26 class sections. Students were given a small amount of extra credit for their course grade for completing the survey. Information for providing extra credit was collected via a separate survey, while the survey data used in the research was stored without specific identifiers of participants.

A total of 404 surveys were collected. Of the 404 surveys, 42 were incomplete. Incomplete surveys were discarded, leaving 362 complete surveys. Of the 362 complete surveys, some subjects spent very little time on the reading that was the basis for the comprehension questions (reading time mean=144 seconds; SD=195 seconds; median=91 seconds). For a 250-word reading, based on the assumption that 400 words per minute is the high end of speed reading with comprehension, it would take a minimum of 37.5 seconds to read the article. Allowing for skimming which could yield some comprehension and the possibility that data visualizations might enhance comprehension even without close reading, the decision was made to consider only those surveys where the subject spent at least 25 seconds on the reading page. Based on this, 76 responses were identified as speeding and removed from the sample. The final sample included in the research was N=286.

Of the final 286 completed surveys, respondents were still distributed evenly among the three experimental conditions, with 95 receiving the text-only version, 96 receiving the version with a bar graph and 95 receiving the version with a data table.

Measures

Demographics. *Gender* was included as a dichotomous variable (72% female).

Data Literacy. Data literacy was measured in two ways: 1) a data literacy score and 2) self-reported math, statistics and data science education in high school and college.

A *data literacy score* was calculated based on subject responses to a data literacy pretest designed by the News Literacy Project (n.d.), an education nonprofit concerned with news literacy throughout American society. The pre-test asked two multiple choice questions and one “choose all that apply” question. The first multiple choice question, worth two points, involved correctly reading a bar graph (M=.94, SD=1). The second multiple choice question, also worth two points, was to compute a percentage based on a headline (M=.38, SD=.79) and the “select all that apply” question (worth a half point for each correct selection, up to two points total) was to critique a pie chart (M=.89, SD=.42). Correct responses were then summed to create a new index on a scale of 0-6 (M=2.2, SD=1.5) and then split into Low (0-0.7), Mid (0.7-3.7) or High (3.7-6) data literacy, with the Mid data literacy category encompassing plus or minus one standard deviation of the mean.

High school math experience was determined by self-reported high school math courses. A combined measure was created from the data separating subjects into those who had taken neither statistics/data science nor calculus (45%), those who had taken statistics/data science (44%), and those who had taken calculus but not statistics (11%). The reasoning behind these groupings is the assumption that pure math courses might develop data literacy less than a statistics course or a specific data science course. Data science courses are distinct from traditional statistics in that they are more hands-on, working more directly with data. Note that few subjects (N=13) reported having had a data science course.

College math experience was also determined by self-reported coursework. A combined measure was created from the data separating subjects into those who had taken neither statistics nor mathematics courses in college (29%), those who had taken at least one statistics class (29%), and those who had taken math but not statistics (42%).

Dependent Variable: Comprehension. A *comprehension score* was calculated based on subject responses to the post-reading comprehension test. All subjects received the same comprehension test, regardless of the version of the article they read. All three formats of the article contained sufficient information to correctly answer all the questions on the comprehension test. In other words, there were no questions specific to data displayed in the table or in the bar graph that were not also referenced in the text. Comprehension was measured by an 18-question quiz based on the reading: three multiple choice questions, a series of eight questions about the “Major Problem” categories listed in the reading, and a series of seven True/False questions. Correct responses were then summed to create a new index on a scale of 0-18 (M=12.3, SD=2.8).

Analysis

To test my hypotheses, I conducted a series of Analysis of Variance (ANOVAs) and other tests examining comprehension and the effects of data literacy across the experimental conditions.

A correlation was run looking for a relationship between comprehension score and other data collected, including the reading condition, data literacy, prior education, and gender.

Given the importance within the context of the research of the randomization process and assignment of the three reading conditions, I analyzed the data to determine if there were any factors that might have skewed the results by overrepresenting a group within a reading condition.

After confirming that respondents were randomly distributed across conditions, I conducted further analysis related to my hypotheses.

Results

The reading condition (text-only, bar graph, data table) and data literacy as measured by the data literacy pre-test were the only variables shown to have significant correlation with comprehension (see Table 2).

Table 2 *Correlations with comprehension scores*

Pearson correlation scores					
	Variables				
Measure	**Data Lit Score**	**Gender**	**College Math**	**HS Math**	**Reading condition**
Comprehension	.160**	.026	.080	.057	.423**
Sig. (2-tailed)	.007	.659	.179	.337	<.001
** Correlation is significant at the 0.01 level (2-tailed)					

For all the variables examined, including gender, prior math education and data literacy, ANOVAs showed no anomalies in the assignment of reading condition (see Table 3). All variables examined were reasonably distributed among all three reading conditions. Note that even for nominal variables such as gender, the data were assigned numeric values internally, and although the means do not have inherent meaning, a consistent mean across conditions represents an even distribution

Table 3 *ANOVA test checking the randomization of variables that may impact comprehension across the reading conditions*

	Reading condition							
Variable	**text-only**		**bar graph**		**data table**			
	M	***SD***	***M***	***SD***	***M***	***SD***	***F***	***p***
Gender (female=high)	1.72	0.45	1.75	0.44	1.76	0.43	0.246	0.782
HS math education	0.79	0.82	0.77	0.99	0.83	0.95	0.108	0.898
College math education	1.59	1.00	1.44	1.10	1.39	1.10	0.891	0.412
Data literacy (pre-test)	2.30	1.57	2.20	1.53	2.17	1.37	0.192	0.826

The first hypothesis (H_1) predicted that subjects with a data visualization supplementing their reading material would demonstrate better comprehension of the material than those without a data visualization. Subjects with the text only version of the reading scored lowest on the comprehension test, with a mean comprehension score of 10.8. Subjects with the bar graph version of the reading had a mean comprehension score of 12.53, 16% higher. Subjects with the data table version of the reading scored higher still, with a mean comprehension score of 13.68, 27% higher than those in the text-only condition. The differences in the mean comprehension scores by version of the reading were significant [$F(2,283)=31.3$, $p<.001$]. A post-hoc analysis (with Tukey HSD correction) found that the differences were significant between each version (i.e., text-only to bar graph, text-only to data table, and bar graph to data table all showed significance -- see Table 4).

Table 4 *Mean comparison (ANOVA) and post-hoc analysis for comprehension scores across the reading conditions*

Comprehension by reading condition								
	text-only		**bar graph**		**data table**			
	M	***SD***	***M***	***SD***	***M***	***SD***	***F***	***p***
Comprehension	10.8 [a]	2.5	12.5 [a]	2.4	13.7 [a]	2.7	2.65	<.001
[a] Post-hoc sig. difference, *p*<.05 (Tukey HSD)								

The results provide support for H_1 – those in a condition with a data visualization scored significantly higher on comprehension compared to those who received the text-only version.

Turning to the second hypothesis (H_2), I predicted that subjects who have had a course in statistics and/or data science will demonstrate better comprehension of the material than those who have not had statistics and/or data science.

Based on self-reported high school math courses, high school students who had taken calculus recorded a higher, though not statistically significant, mean comprehension than those who had taken statistics, contrary to H_2. An ANOVA did not find any comprehension differences based on high school math background to be significant [$F(3, 282)=0.29$, $p=0.75$].

As for college math courses, a *T*-test found no significant difference in comprehension for students who had taken a statistics course in college versus those who had not. Likewise for non-statistics college math, a *T*-test found no significance. An ANOVA, examining comprehension by those who had taken no college math, those who had taken statistics, and those who had taken math but not statistics, found that the students who had taken a statistic course had the highest mean comprehension, but the ANOVA did not find the difference to be significant [$F(3, 282)=0.58$, $p=0.56$].

To control for the potential impact of the significance of the reading version, separate ANOVAs were run for both high school and college math experience for each reading condition. Neither high school math background, nor college math background, had a significant influence on comprehension in any of the three reading conditions (see Tables 5, 6).

H_2 is not supported. I did not find a significant difference in mean comprehension based on prior statistical education.

Table 5 *Significance of high school math experience on comprehension (ANOVA) for each reading condition*

Comprehension by high school math/stats											
	No calc/stats			**Statistics**			**Calculus**				
	M	***SD***	***N***	***M***	***SD***	***N***	***M***	***SD***	***N***	***F***	***p***
text-only	10.83	2.26	40	10.53	2.67	43	11.67	2.39	12	0.989	0.376
bar graph	12.47	2.36	51	12.71	2.75	34	12.27	2.05	11	0.160	0.852
data table	13.51	2.46	43	13.67	2.63	43	14.56	2.46	9	0.567	0.569

Table 6 *Significance of college math experience on comprehension (ANOVA) for each reading condition*

Comprehension by college math/stats											
	No math/stats			**Statistics**			**Math**				
	M	***SD***	***N***	***M***	***SD***	***N***	***M***	***SD***	***N***	***F***	***p***
text-only	10.82	2.40	22	10.59	2.66	27	10.91	2.43	46	0.141	0.868
bar graph	12.14	1.92	28	12.38	2.74	26	12.88	2.59	42	0.821	0.443
data table	12.94	3.33	32	14.50	2.19	30	13.67	2.13	33	2.767	0.068

In addition to stats/math education, I tested another measure of data literacy, the data literacy score. H_3 states that subjects with higher levels of data literacy will demonstrate better comprehension of the materials than those who have lower data literacy scores. To account for potential confounding effects due to the different reading conditions, I tested the impact of data literacy on comprehension using a separate ANOVA for each reading condition (three tests in total). Literacy was split into three groups: Low, Mid and High data literacy.

When examined separately for each reading condition, comprehension was lowest for the text-only reading condition, and low data literacy yielded the lower comprehension than other levels of data literacy (see Table 7). However, the differences were only significant [$F(3, 282)=6.9$, $p=0.002$] within the bar graph condition. Tukey HSD found significance for the mean difference in the bar graph condition from low data literacy to high data literacy.

Table 7 *Significance of data literacy level on comprehension (ANOVA) for each reading condition*

Comprehension by data literacy											
	Low data lit			**Mid data lit**			**High data lit**				
	M	**SD**	**N**	**M**	**SD**	**N**	**M**	**SD**	**N**	***F***	***p***
text-only	10.14	2.60	14	10.79	2.48	66	11.47	2.29	15	1.529	0.222
bar graph	11.38 [a]	1.98	13	12.49	2.42	69	13.79 [a]	2.61	14	6.916	0.002
data table	13.00	3.35	6	13.85	2.48	79	12.80	3.61	10	1.668	0.194

[a] Post-hoc sig. difference, p<.05 (Tukey HSD)

An ANOVA of data literacy by reading condition did not find any significance, indicating that the inclusion of low data literacy individuals in the text-only condition did not skew the comprehension by reading condition findings.

It is worth noting that the highest mean comprehension score across reading conditions by data literacy was for the mid data literacy subjects in the data table condition, not for any group in the high data literacy group. These results would seem to indicate that a minimum level of data literacy contributes to comprehension of data intensive information, but more research is required in this area.

H_3 is partially supported. Low data literacy subjects demonstrated lower comprehension of the information, at least for those in the bar graph condition.

Discussion

The result that comprehension was higher for those in the data table condition compared to those in the bar graph condition was unexpected, but the results are perhaps even more important for the field of data journalism. If the research was originally conceived to show that data visualizations increase comprehension, a more fundamental concept may be illustrated by these results. Data – access to data in any form – increases comprehension, full stop.

It is worth noting that, owing in part to the constraints of a research survey in which test subjects receiving extra credit might be likely to drop out of a complicated or time-consuming survey, the reading and data within this experiment were relatively simple. It is possible that the data table condition was more successful in this case because of the straightforward nature of the data and that another visualization might have more significant comprehension benefits given a more complicated scenario. This is an area of study that requires further research.

What might explain the lack of support for H_2? The most straightforward explanation of course is simply that it is not supported, and prior math education does not influence the comprehension of data intensive information. The results may also be related to flaws in the experimental design. The subjects for the survey were a convenience sample from Journalism courses and required general education mathematics courses such as Math 1100, college algebra. In other words, the subjects may have lacked the variety of mathematics background required to demonstrate differences. It may also be that the self-reported data gathered to determine math experience were insufficient to identify any differences.

The results for H_3 seem to indicate that a baseline level of data literacy is beneficial in interpreting data intensive information. It is also worth noting that data literacy level was only truly significant as a factor in the bar graph condition in which a graphical representation of the data is utilized as opposed to just raw numbers. This is another area that would benefit from further research.

Limitations

Several factors may have limited the research findings. It is possible that participants may not have spent sufficient time to analyze and comprehend the experimental device, the reading. If participants rushed through the reading or answered the questions randomly the results would be unreliable. It is for this reason that surveys where participants spent under 25 seconds on the reading page were omitted. It is, of course, still possible that students lingered on the reading page but did not spend sufficient time on the questions themselves, but presumably a student taking the survey just for the extra credit would save the most time on the reading page and could be identified at that point.

It is also possible that the methods chosen failed to accurately identify a data literate cohort. I attempted to distinguish data literacy level in two ways, by educational background and by means of a pre-test. For educational background, no comprehension differences were noted. This may be because the information I collected regarding math and statistics education did not properly reflect data literacy levels based on educational background.

The data literacy pre-test did show a significant benefit to comprehension in the bar graph condition, but the designation may lack merit. The data literacy pre-test utilized came from the Data Literacy Project, but there does not appear to be a generally accepted means of testing data literacy. Measuring the data literacy of the subject is clearly desirable in this experimental design, but it is unclear whether this specific pre-test did so effectively.

Professional Work Product

The professional component of this project is a lesson plan for my high school Introduction to Data Science class.

The lesson will occur over the course of two days during the second semester. The timing is significant. It should be late enough in the Introduction to Data Science (IDS) curriculum that a data literacy effect would be expected. It should also occur near the beginning of the final unit of Algebra 2, a unit on statistics and probability, so that it fits with the flow of the Algebra 2 class as well.

On day one of the lesson, students in IDS, as well as those in Algebra 2 classes, will be given a warmup that consists of the reading and comprehension quiz from this project. Versions of the article will be assigned to students randomly. The IDS students will be presumed to be the data literate group and the Algebra 2 students will be the non-data literate group. This portion of the lesson will only encompass the warmup for day one. The remainder of the day will be within the normal flow of the curriculum.

The second part of the lesson will occur a week later, after analysis of the data. This lesson, for the IDS students only, will consist of a discussion of the project they have now taken part in. We will discuss the way that research is designed and the way data is collected and analyzed. Assuming the hypotheses are supported, the lesson will also demonstrate the role of data visualizations and data literacy in comprehension of complex information which will help to validate for the students the importance of the class they are taking.

Reflection and Conclusion

Reflection

This project is a continuance of the research outlined above. It was intended to verify that the inclusion of a data visualization will enhance comprehension of data intensive information for readers, as Multimedia Learning Theory would suggest. My research lends strong support to these ideas, as the comprehension benefits were significant for participants with access to data.

It was unexpected, however, that a simple data table provided even more comprehension benefit than a bar graph, even though both contained identical information. The lesson, and it would be an important lesson for data journalists, may be that less is more – the data is what is important, and the simpler the presentation of that data, the more easily it can be understood by readers.

The relatively straightforward nature of the bar graph in this experiment may explain the results to some extent. A more complicated data visualization might convey more information to a data literate reader. It is worth noting as well that the bar graph condition showed significantly higher comprehension for the highly data literate cohort. It may be that with a simple ordered list of data, a table is sufficient and even preferable, but with more sophisticated visualizations data literacy will become increasingly important. These questions would be interesting opportunities for future research.

I had also hoped to validate the importance of data science education as a means of increasing data literacy, a vital life skill in the modern world as data becomes increasingly important. My research does indicate that a certain level of data literacy is important in the comprehension of data intensive information. However, the hypothesis that education in statistics will increase data literacy more than others is not supported in this research.

Although I did not see any effect of data literacy as identified by prior education, that may well be because I did not measure prior education in a way that properly reflected educational experiences or that the cohort of test subjects did not include the necessary backgrounds. The experiment did however show significance of data literacy as measured by a pre-test, which showed that data literacy can be important, especially for more specific data visualizations such as a bar graph. So the basic premise remains that data literacy will be of increasing importance as data journalism grows, and the best way to develop data literacy in society remains via the education system.

It should be stated that only 14 of the 286 participants in this study had taken a course in data science in high school. Traditional statistics courses were not shown to develop data literacy any more effectively than other math courses based on my research. However, as data science specific courses continue to grow in popularity, it is possible that those courses may help produce more data literate students. More research is definitely needed in this area.

As a teacher and a journalist, I have always believed strongly in the promise of data journalism and in the importance of data literacy. As a journalist, I feel that conclusions should always be supported by data. As a math teacher, one of the most important things I can hope to do is to help students become discerning consumers of information.

This project has validated these beliefs. Journalists should base conclusions on data where possible but should also SHOW the data to their audience in some form (preferably the simplest form possible). Educators should be striving to increase the data literacy of their students. Still, many questions remain, not the least of which being, how exactly do we increase data literacy?

If this research were to continue, I'd want to see a more robust means of identifying a data literate cohort, both through prior education and through a data literacy test. In fact, as originally conceived, this research would have used high school students taking the Introduction to Data Science course and would have tested whether they scored higher in comprehension than those who had not taken the course. That design was abandoned owing to the complexities of using minors as test subjects, but the idea certainly has merit. Not only would it provide a test condition that was unavailable in the existing experiment, the results could also be used to validate the importance of the data science curriculum directly, if indeed the curriculum is beneficial.

Conclusion

Data journalism has much to gain if we can better understand effective ways to include data with text and thereby provide better information to audiences. Society has much to gain if its members are more data literate.

As has been discussed, the amount of data available has exploded in recent years. The importance of data and of data literacy will only continue to grow. The important work of data journalists can only keep pace if broader society has the data literacy to understand the information data journalists produce.

We have seen during the COVID pandemic, during political campaigns and in many other instances, the use and misuse of information. Data journalists have a responsibility to try to accurately present information, and data visualizations can be a key tool in doing so.

At the same time, part of being an informed consumer of information in modern society is having the data literacy to correctly interpret the data intensive information available. Data journalism and data literacy both have vital roles to play in maintaining a well-informed society in the future.

References

Benjamin, A. (2009, February). *Teach statistics before calculus!* [Video]. TED. https://www.ted.com/talks/arthur_benjamin_teach_statistics_before_calculus/transcript?language=en

Boaler, J. & Levitt, S. (2019, October 23). Modern high school math should be about data science — not Algebra 2*. Los Angeles Times.* https://www.latimes.com/opinion/story/2019-10-23/math-high-school-algebra-data-statistics

Bounegru, L. (2021). Data journalism in perspective. In L. Bounegru & J. Gray (Eds.). *The data journalism handbook: Towards a critical data practice*. Amsterdam University Press.

Cherry, K. (2019, November 27). *Overview of VARK learning styles*. https://www.verywellmind.com/vark-learning-styles-2795156. Accessed March 2, 2022

Cohen, S. (2021). Using visualizations to tell stories. In L. Bounegru & J. Gray (Eds.). *The data journalism handbook: Towards a critical data practice*. Amsterdam University Press.

Coleman, J., McTigue, E., & Dantzler, J. (2018). What makes a diagram easy or hard? The impact of diagram design on fourth-grade students' comprehension of science texts. *The Elementary School Journal, 119*(1). DOI:10.1086/698819

D'Ignazio, C. (2017). Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal, 23*(1), 6–18. https://doi-org.proxy.mul.missouri.edu/10.1075/idj.23.1.03dig

Dixon, G. N., Mckeever, B. W., Holton, A. E., Clarke, C., & Eosco, G. (2015). The power of a picture: Overcoming scientific misinformation by communicating weight-of-evidence information with visual exemplars. *Journal of Communication, 65*(4), 639–659. https://doi-org.proxy.mul.missouri.edu/10.1111/jcom.12159

Elia, I., Gagatsis, A., & Demetriou, A. (2007). The effects of different modes of representation on the solution of one-step additive problems. *Learning and Instruction, 17*, 658-672.

Engebretsen, M., Kennedy, H., & Weber, W. (2018). Data visualization in Scandinavian newsrooms: Emerging trends in journalistic visualization practices. *NORDICOM Review, 39*(2), 3–18. https://doi-org.proxy.mul.missouri.edu/10.2478/nor-2018-0007

Frank, M., Walker, J., Attard, J., & Tygel, A. (2016). Data literacy - What is it and how can we make it happen? *The Journal of Community Informatics, 12*(3), 4-8.

Guo, D., McTigue, E., Matthews, S., & Zimmer, W. (2020). The impact of visual displays on learning across the disciplines: A systematic review. *Educational Psychology Review, 32*(3), 627–656.

Kees, J., Burton, S., Andrews, J. C., & Kozup, J. (2010). Understanding how graphic pictorial warnings work on cigarette packaging. *Journal of Public*

Policy & Marketing, 29(2), 265–276. https://doi-org.proxy.mul.missouri.edu/10.1509/jppm.29.2.265

Kennedy, H., Allen, W., Hill, R., Engebretsen, M., Kirk, A., & Weber, W. (2021). Data visualizations: Newsroom trends and everyday engagements. In L. Bounegru & J. Gray (Eds.). *The data journalism handbook: Towards a critical data practice*. Amsterdam University Press.

Kirschner, P. A. (2017). Stop propagating the learning styles myth. *Computers & Education, 106,* 166-171.

Kovach, B., & Rosenstiel, T. (2021). *The elements of journalism: What newspeople should know and the public should expect* (4th edition). New York: Crown.

Lochbuehler, K., Wileyto, E. P., Mercincavage, M., Souprountchouk, V., Burdge, J. Z., Tang, K. Z., Cappella, J. N., & Strasser, A. A. (2019). Temporal effects of message congruency on attention to and recall of pictorial health warning labels on cigarette packages. *Nicotine & Tobacco Research, 21*(7), 879–886. https://doi-org.proxy.mul.missouri.edu/10.1093/ntr/nty124

MacAskill, E., & Dance, G. (2013, November 1). NSA files: Decoded*. The Guardian.* https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded#section/1

Martinez, W., & LaLonde, D. (2020). Data science for everyone starts in kindergarten: Strategies and initiatives from the American Statistical

Association. *Harvard Data Science Review, 2*(3). https://doi.org/10.1162/99608f92.7a9f2f4d

Mayer, R. (1997) Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1-19. DOI: 10.1207/s15326985ep3201_1

Mayer, R. (2020). *Multimedia learning* (3rd edition). Cambridge: Cambridge University Press. doi:10.1017/9781316941355

McTigue, E. (2009). Does multimedia learning theory extend to middle-school students? *Contemporary Educational Psychology, 34*(2), 143-153 DOI:10.1016/j.cedpsych.2008.12.003

McTigue, E. M., & Flowers, A. C. (2011). Science visual literacy: Learners' perceptions and knowledge of diagrams. *Reading Teacher, 64*(8), 578–589. https://doi-org.proxy.mul.missouri.edu/10.1598/RT.64.8.3

Murphy, S. J. (2009). *The power of visual learning in secondary mathematics education*. Pearson: Research into Practice.https://assets.savvas.com/asset_mgr/current/201033/VisualLearningResearch.pdf

Nguyen, D. (2021). Mediatisation and datafication in the global COVID-19 pandemic: On the urgency of data literacy. *Media International Australia, 178*(1), 210–214. https://doi-org.proxy.mul.missouri.edu/10.1177/1329878X20947563

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.

Parodi, G., & Julio, C. (2017). More than words: Contending semiotic systems and the role of disciplinary knowledge in specialized text comprehension. *Iberica, 33*, 11–35.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest*, *9*(3), 105–119. https://doi.org/10.1111/j.1539-6053.2009.01038.x

Renkl, A., & Scheiter, K. (2017). Studying visual displays: How to instructionally support learning. *Educational Psychology Review, 29*(3), 599–621. https://doi.org/10.1007/s10648-015-9340-4

Ridsdale, C., Rothwell, J., Smit, M., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., Wuetherick, B., & Ali-Hassan, H. (2015). Strategies and best practices for data literacy education. *Knowledge Synthesis Report*. doi:10.13140/RG.2.1.1922.5044.

Roberts, K. L., & Brugar, K. A. (2017). The view from here: Emergence of graphical literacy. *Reading Psychology, 38*(8), 733–777. https://doi-org.proxy.mul.missouri.edu/10.1080/02702711.2017.1336661

Rogers, J. (1985). *The dictionary of clichés.* New York: Ballantine Books.

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics, 16*(6), 1139-1148. doi: 10.1109/TVCG.2010.179.

Segers, E., Verhoeven, L., & Hulstijn-Hendrikse, N. (2008). Cognitive processes in children’s multimedia text learning. *Applied Cognitive Psychology, 22*(3), 375–387. https://doi-org.proxy.mul.missouri.edu/10.1002/acp.1413

Statista Research Department. (2022, March 18). *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025*. Statista. https://www.statista.com/statistics/871513/worldwide-data-created/

The Data Literacy Project. (n.d). *Assessment*. Accessed April 23, 2022. https://thedataliteracyproject.org/assessment

The News Literacy Project. (n.d). *News lit quiz: Can you make sense of data?* Accessed April 23, 2022. https://newslit.org/tips-tools/can-you-make-sense-of-data/

UN Data Revolution Group, A World That Counts. (n.d). undatarevolution.org. Accessed March 12, 2022. https://www.undatarevolution.org/report/

Valkenburg, P. M., Peter, J., & Walther, J. B. (2016). *Media effects: Theory and research. Annual Review of Psychology, 67.*

Vekiri, I. (2002) What is the value of graphical displays in learning? *Educational Psychology Review, 14*(3).

Waller, R. (1981, April). *Understanding network diagrams* [Conference presentation]. Annual Meeting of the American Educational Research Association, Los Angeles.

Ward, D. B. (1992). The effectiveness of sidebar graphics. *Journalism Quarterly, 69*(2), 318–328. https://doi-org.proxy.mul.missouri.edu/10.1177/107769909206900207

Wolff, A., Gooch, D., Cavero Montaner, J.J., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics, 12*(3), 9-26.

Yale Poorvu Center. (n.d.). *Learning styles as a myth*. The Yale Poorvu Center for Teaching and Learning. Accessed April 28, 2021. https://poorvucenter.yale.edu/LearningStylesMyth

Zraick, K. (2019, February 20). Teenagers say depression and anxiety are major issues among their peers*. New York Times.* https://www.nytimes.com/2019/02/20/health/teenage-depression-statistics.html

Appendix A - Data Journalism Survey

Welcome and Consent

Project Title: Data Journalism, Data Literacy and Data Visualizations – A quantitative study
Principal Investigator/Researcher: Steven Perrin; Dr. Kathleen Rose, Advisor
IRB Reference Number: 2091433

You are being invited to take part in a research project. You must be 18 years of age or
older. Your participation is voluntary, and you may stop being in this study at any time. The purpose of the research is to better understand how data journalists rely on and developdata literacy in their audience. This research is conducted on the Internet using an online survey. You will be asked to read a news article and then complete an online survey. Your participation should take about 15 minutes. The information you provide will be kept confidential and only the research team will have access. Only your pawprint and the course ID will be provided to your professor as proof of your participation in this study in order to receive extra credit for compensation. These identifiers are recorded on a separate survey to which you will be redirected after completing the research survey. Your identity will not be connected to your responses in the research survey.

If you do not wish to participate in the survey but would still like to receive the extra credit, you may complete an alternative assignment of sharing 100-200 words on your "thoughts on the importance of data literacy in society." If you intend to complete the alternative assignment, please inform your course professor for further instructions.

If you have questions about this study, you can contact Steve Perrin at 562-826-2646,
sfprm8@missouri.edu. If you have questions about your rights as a research participant,
please contact the University of Missouri Institutional Review Board (IRB) at 573-882-3181 or muresearchirb@missouri.edu. The IRB is a group of people who review research studies to make sure the rights and welfare of participants are protected. If you want to talk privately about any concerns or issues related to

your participation, you may contact the Research Participant Advocacy at 888-280-5002 (a free call) or email
muresearchrpa@missouri.edu.

You can ask the researcher to provide you with a copy of this consent for your records, or you can save a copy and/or print this consent if it has already been provided to you. We appreciate your consideration to participate in this study.

Your response below indicates that you have read this consent sheet, had an opportunity to ask any questions about your participation in this research, and voluntarily consented to participate.

○ I agree to participate in this study and have read the consent sheet above.

○ I do not wish to participate in this study (you will be directed out of the survey).

Learning style

Which is your preferred learning style (select all that apply)?

▢ Visual (I learn and retain information best when I see it)

▢ Read-write (I learn and retain information best when I read it or write it down)

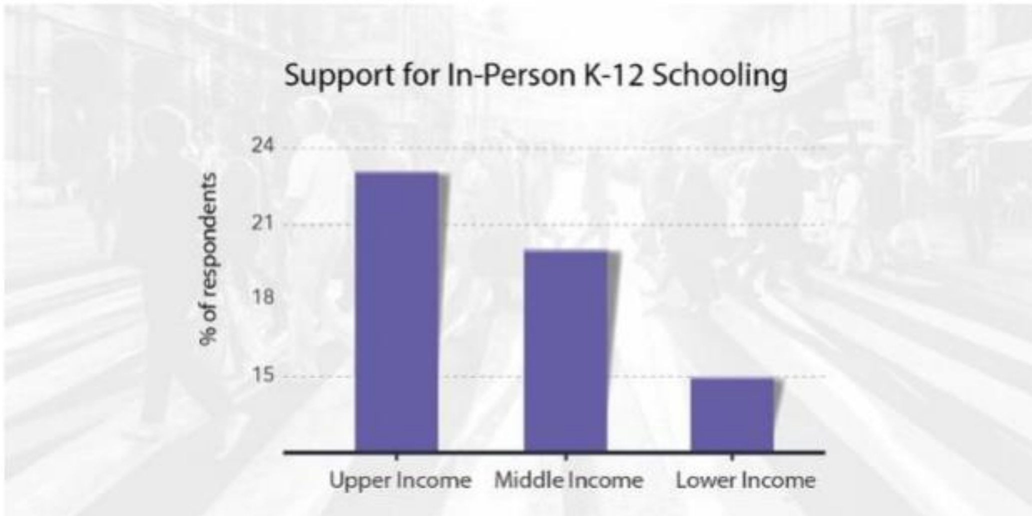
▢ Auditory (I learn and retain information best when I hear it)

▢ Kinesthetic (I learn and retain information best when I can touch or manipulate something about it) (4)

▢ I don't know.

Data Literacy

Which of these news headlines is appropriate, given the data presented in the image?



- ◯ Most high income families support in-person K-12 schooling.
- ◯ Lower income families are less likely to support in-person instruction for K-12 schooling.
- ◯ Middle- and upper-income families are twice as likely to support in-person schooling compared to low-income families.
- ◯ Lower income families are more fearful of in-person instruction.

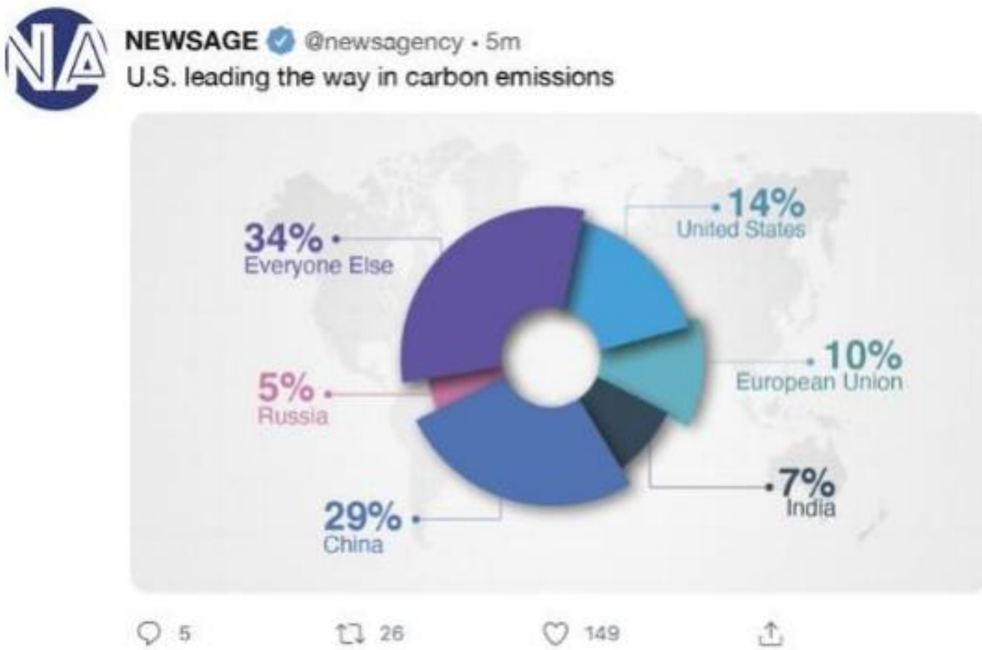
What was the unemployment rate in April according to the news clipping?


The Daily
Thursday, September 10, 2020

The U.S. unemployment rate has plummeted 30% from its peak in April. The rate as of July was 10.2%

- ◯ 40.2%
- ◯ 14.6%
- ◯ 3.2%
- ◯ 22.8%

What is problematic about this social media post? Select all that apply.



- [ ] The title does not match the graph.
- [ ] The size of the pie slices exaggerates the differences.
- [ ] The chart does not consider the populations of the countries or groups of countries.
- [ ] The percentages only add up to 99%.

Reading

We're now going to ask you to read a newspaper article. Please read the article closely, as we are going to ask you questions about it later and you will not be able to return to the article once you continue!

Text Only Article

Teenagers Say Depression and Anxiety Are Major Issues Among Their Peers

Most American teenagers — across demographic groups — see depression and anxiety as major problems among their peers, a new survey by the Pew Research Center found.

The survey found that 70 percent of teenagers saw mental health as a big issue. All but 4 percent saw it as at least a minor problem. A majority of teens also listed bullying and drug addiction as major problems. Fewer teenagers cited, in descending order, alcohol, poverty, teen pregnancy and gangs as major problems.

The consistency of the responses about mental health issues across gender, race and income lines was striking, said Juliana Horowitz, an associate director of research at the center.

The survey of 920 teenagers ages 13 to 17 in the United States was conducted online and by phone in the fall.

Some psychologists have tied a growth in mental health issues among teenagers to increased social media use, academic pressure and frightening events like terror attacks and school shootings.

Teenagers who grew up in the post-9/11 era, and amid many school shootings, may have anxiety tied to an environment filled with dire warnings about safety, said Philip Kendall, director of the Child and Adolescent Anxiety Disorders Clinic at Temple University in Philadelphia.

Another major stressor is constant surveillance by peers on social media, and the “fear of missing out” it can generate, he added.

A study released in 2017 found that the number of children and adolescents admitted to children's hospitals for thoughts of self-harm or suicide had more than doubled from 2008 to 2015, echoing trends in federal data.

Article with bar graph

Teenagers Say Depression and Anxiety Are Major Issues Among Their Peers

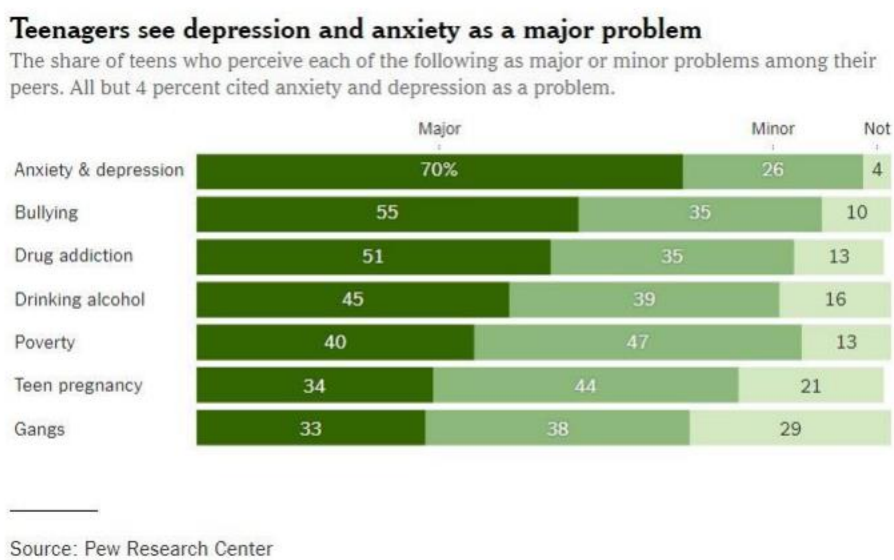
Most American teenagers — across demographic groups — see depression and anxiety as major problems among their peers, a new survey by the Pew Research Center found.

The survey found that 70 percent of teenagers saw mental health as a big issue. All but 4 percent saw it as at least a minor problem. A majority of teens also listed bullying and drug addiction as major problems. Fewer teenagers cited, in descending order, alcohol, poverty, teen pregnancy and gangs as major problems.

The consistency of the responses about mental health issues across gender, race and income lines was striking, said Juliana Horowitz, an associate director of research at the center.

The survey of 920 teenagers ages 13 to 17 in the United States was conducted online and by phone.

Some psychologists have tied a growth in mental health issues among teenagers to increased social media use, academic pressure and frightening events like terror attacks and school shootings.



Teenagers who grew up in the post-9/11 era, and amid many school shootings, may have anxiety tied to an environment filled with dire warnings about safety, said Philip Kendall, director of the Child and Adolescent Anxiety Disorders Clinic at Temple University in Philadelphia.

Another major stressor is constant surveillance by peers on social media, and the “fear of missing out” it can generate, he added.

A study released in 2017 found that the number of children and adolescents admitted to children’s hospitals for thoughts of self-harm or suicide had more than doubled from 2008 to 2015, echoing trends in federal data.

Article with data table

Teenagers Say Depression and Anxiety Are Major Issues Among Their Peers

Most American teenagers — across demographic groups — see depression and anxiety as major problems among their peers, a new survey by the Pew Research Center found.

The survey found that 70 percent of teenagers saw mental health as a big issue. All but 4 percent saw it as at least a minor problem. A majority of teens also listed bullying and drug addiction as major problems. Fewer teenagers cited, in descending order, alcohol, poverty, teen pregnancy and gangs as major problems.

The consistency of the responses about mental health issues across gender, race and income lines was striking, said Juliana Horowitz, an associate director of research at the center.

The survey of 920 teenagers ages 13 to 17 in the United States was conducted online and by phone in the fall.

Some psychologists have tied a growth in mental health issues among teenagers to increased social media use, academic pressure and frightening events like terror attacks and school shootings.

Teenagers see depression and anxiety as a major problem

The share of teens who perceive each of the following as major or minor problems among their peers. All but 4 percent cited anxiety and depression as a problem.

	Major	Minor	Not
Anxiety & depression	70%	26%	4%
Bullying	55%	35%	10%
Drug addiction	51%	35%	13%
Drinking alcohol	45%	39%	16%
Poverty	40%	47%	13%
Teen pregnancy	34%	44%	21%
Gangs	33%	38%	29%

Source: Pew Research Center

Teenagers who grew up in the post-9/11 era, and amid many school shootings, may have anxiety tied to an environment filled with dire warnings about safety, said Philip Kendall, director of the Child and Adolescent Anxiety Disorders Clinic at Temple University in Philadelphia.

Another major stressor is constant surveillance by peers on social media, and the “fear of missing out” it can generate, he added.

A study released in 2017 found that the number of children and adolescents admitted to children’s hospitals for thoughts of self-harm or suicide had more than doubled from 2008 to 2015, echoing trends in federal data.

Comprehension

Answer the following questions based on the reading.

What percentage of teens surveyed considered anxiety and depression to be a **major problem**?

- ○ 70%
- ○ 4%
- ○ 96%
- ○ This information isn’t provided

What percentage of teens surveyed considered anxiety and depression to be a **problem at some level**?

- ○ 70%
- ○ 4%
- ○ 96%
- ○ This information isn’t provided

What percentage of teens surveyed considered bullying to be a **major problem**?

- 55% (1)
- 10% (2)
- 90% (3)
- This information isn't provided (4)

Based on the information provided in the reading, indicate whether more than half of the teens surveyed thought the following were major problems, or if this was not addressed in the reading. How confident are you in your response?

	Was this a **major problem** according to more than half of teens?			How confident are you in your answer?				
	Yes	No	Not Addressed	Not at all confident	A little confident	Somewhat confident	Confident	Very confident
Anxiety & depression	○	○	○	○	○	○	○	○
Bullying	○	○	○	○	○	○	○	○
COVID	○	○	○	○	○	○	○	○
Drinking alcohol	○	○	○	○	○	○	○	○
Drug addiction	○	○	○	○	○	○	○	○
Gangs	○	○	○	○	○	○	○	○
Poverty	○	○	○	○	○	○	○	○

Teen pregnancy	○	○	○	○	○	○	○	○

Based on the reading, please indicate whether these statements are true or false, or if there isn’t enough relevant information provided.

	Is the statement definitely true, definitely false, or is there no way to be certain based on the reading?			How confident are you in your answer?				
	Definitely True	Definitely False	No way to be certain	Not at all confident	A little confident	Somewhat confident	Confident	Very confident
The primary cause of teen depression is "fear of missing out"	○	○	○	○	○	○	○	○
Most American teens consider poverty to be a problem at some level	○	○	○	○	○	○	○	○
More teens consider drinking alcohol to be a problem than any other issue in the survey	○	○	○	○	○	○	○	○
Fewer teens consider gangs to be a problem than any other issue in the survey	○	○	○	○	○	○	○	○

Self harm and suicide are growing problems among children and adolescents	○	○	○	○	○	○	○	○
Most American teens do not consider teen pregnancy to be a problem at any level	○	○	○	○	○	○	○	○
The mental health findings in the study were consistent across demographic groups	○	○	○	○	○	○	○	○

Helpful text

Display This Question:

If the subject read the text only version of the article

How helpful was this element in answering the survey questions?

	Not at all helpful				Extremely helpful
	1	2	3	4	5
The text of the article					

Helpful bar graph

Display This Question:

If the subject read the bar graph version of the article

How helpful were the following elements in answering the survey questions?

	Not at all helpful				Extremely helpful
	1	2	3	4	5
The text of the article					
The bar graph					

Helpful data table

Display This Question:

If the subject read the data table version of the article

How helpful were the following elements in answering the survey questions?

	Not at all helpful Extremely helpful 1 2 3 4 5
The text of the article	
The table	

Post test

Now we want to know a little bit about your background with statistics and data. We're going to ask you questions about your comfort level and also some questions about your prior education.

To what extent do you agree or disagree with the statements below regarding data visualizations such as bar graphs, line graphs and histograms?

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Data visualizations help me to interpret data on my own.	○	○	○	○	○
I get nervous when first encountering new types of data visualizations.	○	○	○	○	○
Much of the content in data visualizations is too technical for me to understand.	○	○	○	○	○

Rate your confidence in each of the following areas.

	No at all confident				Very confident
	1	2	3	4	5

How confident are you that you could interpret and understand a data visualization? ()	
How confident are you in your knowledge of statistics? ()	
How confident are you in your knowledge of data science? ()	

Which **high school math** courses have you taken previously (select all that apply)?

Do your best to select a comparable course from the list. If you have taken a math course that you feel is not represented on the list, please include the course name in the “Other course” field(s).

- ☐ Algebra 2
- ☐ Statistics
- ☐ Pre-calculus
- ☐ Calculus
- ☐ Data science course
- ☐ Click to write other course ________________
- ☐ Click to write other course ________________

Have you ever taken any **college-level math** courses (excluding statistics and data science courses)?

- ○ Yes
- ○ No
- ○ Not sure
- ○ Prefer not to state

Display This Question:

If Have you ever taken any college-level math courses (excluding statistics and data science courses)? = Yes

How many **college-level math** courses have you taken (excluding statistics and data science courses)?

- ○ 1-2
- ○ 3-5
- ○ 6-10
- ○ More than 10

Have you ever taken any **college-level statistics or data science** courses (including data visualization, research methods)?

- ◯ Yes
- ◯ No
- ◯ Not sure
- ◯ Prefer not to state

Display This Question:

If Have you ever taken any college-level statistics or data science courses (including data visualiz... = Yes

Q12.8 How many **college-level statistics or data science** courses have you taken (including data visualization, research methods)?

- ◯ 1-2
- ◯ 3-5
- ◯ 6-10
- ◯ More than 10

Demographics

Almost done! Now we'd just like to find out a little bit about you.

What year were you born?

____

What is your gender?

- ○ Male
- ○ Female
- ○ Non-binary
- ○ Prefer to self describe ____
- ○ Prefer not to state

Appendix B - Original Project Proposal

Data Journalism, Data Literacy and Data Visualizations – A quantitative study

The world is in the midst of what the United Nations Secretary-General's Independent Expert Advisory Group calls a "data revolution" (2014, p. 2). Exponential increases in data volume, new data analysis technologies and new kinds of data have the potential to transform society, but inequalities in access to data also present major equity challenges (UN Data Revolution Group, 2014). Massive amounts of data available online, sophisticated data analysis software, and the tools to publish all of this information allow "more people to work with more data more easily than ever before" (Bounegru, 2021, p. 3).

The amount of data generated worldwide in 2010 was 2 zettabytes; by 2022 it is estimated that total will be 97 zettabytes (Statista, 2022). This growing mountain of data in turn has exposed the importance within journalism of analyzing, interpreting and presenting all of this data (data journalism) and the importance within society to better understand it all (data literacy).

From my experiences as a journalist and high school data science teacher, it occurs to me that a common goal of both journalist and teacher is to effectively engage and convey information to an audience. The similarities are even more pronounced in the specific realms of data journalism and data literacy. The increased societal importance of data in recent years has in turn been reflected in both the newsroom and the classroom. Data journalism (the analysis of data to uncover newsworthy information and the creation of data visualizations to convey information to readers) is of ever-increasing importance to news organizations (Kennedy et al., 2021). Meanwhile, data science education is being added at all levels of the education system, from universities to elementary schools (Martinez & LaLonde, 2020). These parallel developments are important and symbiotic, as the promise of data journalism will not be fully realized without increased data literacy in the broader population. Data journalists play a vital role in interpreting data and making it accessible to a wider audience which will in turn help to increase the data literacy of readers (Bounegru, 2021).

"The purpose of journalism is to give people the information they need to make better decisions about their lives and society" (Kovach & Rosenstiel, 2021, p. 7). Some of that needed information is now hidden within a vast expanse of data. Journalists and their audience may need to co-evolve to deal with the situation – data journalists developing creative ways to present complex information, while readers develop new skills for interpreting the information. Ideally these factors can form a positive feedback loop – journalists providing more useful information, audiences developing more data literacy in order to handle still more complex information going forward.

"A picture is worth a thousand words" is an adage in multiple languages, dating to the early 20th century and even earlier than that in similar forms (Rogers, 1985). It is such a well-known saying in part because it feels obvious, and journalists and teachers alike often rely on words and pictures to impart meaning.

It has long been proposed that student comprehension is increased when instruction addresses multiple learning modalities, including words and pictures.

Emphasizing different learning styles, such as those in the VARK model (visual, auditory, read-write and kinesthetic learners) has been prevalent in pedagogy for many years (Cherry, 2019). Similarly, Mayers' (1997) Multimedia Learning Theory posits that learners comprehend better when information is delivered through words and images than when the information is conveyed with words alone.

Data journalism (or at least some aspects of data journalism) relies on a similar premise. If people comprehend better through words and pictures, then it follows that a news article accompanied by charts or graphs may be more informative and more compelling than the article alone.

The primary intent of this project is to examine whether a data visualization element increases comprehension of data intensive information. To explore this question, I will conduct an experiment using different versions of a data intensive news article – one version with text only, one version with text and tabular data, and one with text and data visualizations. The hypothesis is that the subjects reading the article with accompanying data visualizations will demonstrate greater comprehension of the information than the other two groups.

In addition, the project explores potential dependencies between increased data literacy and comprehension of data intensive information. In a data intensive world, data literacy is an increasingly important skill and data journalists must understand the level of data literacy in their readership in order to effectively communicate information.

Literature Review

Is a picture "worth a thousand words?" Do pictures in fact add measurable value in terms of comprehension? There is a vast body of research concerning words and pictures as it relates to communication and persuasion. In the area of health communication, incorporating images with text has shown to be more persuasive than words alone as it pertains to warning labels (e.g., Kees et al., 2010; Lochbuehler et al., 2019) or combating disinformation (Dixon et al., 2015). This research on messaging has found that the images need to be appropriate to the message and congruent with the message (Lochbuehler et al., 2019) to be most effective.

As regards to the comprehension of new material, in education or journalism, there are certain ideas that seem obvious, like the adage that "A picture is worth a thousand words." One such notion is that individuals have different styles of learning and learn best when the instruction is delivered in their preferred style. Another notion is that information will be retained more readily when it is conveyed with both words and pictures. However, the empirical evidence for these assumptions is mixed in the literature (e.g., Coleman et al., 2018; Guo et al., 2020; McTigue, 2009; Segers et al., 2008; Vekiri, 2002), which tends to show that the right picture and the right audience are both necessary to achieve the anticipated comprehension benefits.

The proposed project requires a review of data journalism and data visualizations to help define the concepts as well as a review of the importance of data literacy. This review will also examine the existing literature on multimedia learning and learning styles and finally the research on visual displays.

Data Journalism and Data Visualizations

Data journalism is a broad term that has been used to refer to many different things. The use of large data sets, public records, analysis tools and visual presentation of data are all aspects of data journalism. Indeed, writing in *The Data Journalism Handbook*, Cohen (2021) includes all of the above areas of expertise, and goes so far as to define data journalism as "anything in a newsroom that requires the use of numbers, or … computer programming" (p. 1) – which is to say, essentially everything.

One specific element of data journalism in the 21st century is uncovering the story hidden in the data: The availability of public data sets and the tools to analyze them make it possible to find correlations, significance, patterns and importance that were previously hidden. One prominent example of this form of data journalism would be *The Guardian*'s analysis of Edward Snowden's leaked National Security Agency (NSA) files (MacAskill & Dance, 2013). The massive volume of data in the NSA leaks precluded a simple understanding of any significance. Data journalists at *The Guardian* used analysis and visualizations to add vital context to the data, making important information accessible to the public and their work helped to popularize the term "data journalism" (Bounegru, 2021)

After the important work of analysis, the data journalist needs to do something equally important: Present the information to the reader. That can be accomplished through words alone, but it is also increasingly done through data visualizations (Kennedy et al., 2021). This data journalism is at the same time traditional, telling a story with words and pictures, and modern, utilizing interactive graphics or new data visualizations like heat maps where data values are represented by colors. This form of data journalism involves the presentation of data and information in a manner that attempts to be more engaging and informative. For the purposes of this project, we will focus on this narrower definition of data journalism: The inclusion of data visualizations within text articles for the purpose of enhancing understanding.

Similarly this project defines data visualization as “A visual representation of data, created to amplify the cognitive processing and social application of data represented” (Engebretsen et al., 2018, p. 4). In interviews with data journalists in European news organizations, interviewees told Kennedy et al. (2021) that the use of data visualizations is growing and that they are becoming “central elements in journalistic storytelling” (p. 1). The same series of interviews found that a data visualization can be “the driving force of a story, even when it is a simple graphic or diagram” (p. 1).

Data visualizations appear to increase engagement, as well. Through reader statistics, news organizations have determined that the inclusion of a data visualization causes readers to remain on a page longer (Kennedy et al., 2021). This additional engagement indicates a reader who is cognitively active, one who is actively trying to make sense of information (Mayer, 2020).

Segel & Heer (2010) examined data journalism and data visualizations to identify the crucial elements in telling a story with data. Among their findings were that the uses of data visualizations should be limited and strategic. "Data stories appear to be most effective when they have constrained interaction at various checkpoints within a narrative, allowing the user to explore the data without veering too far from the intended narrative" (Segal & Heer, 2010, p. 1147).

This is the data journalism that is of interest in this project: the inclusion of data in the story, the telling of the story with data. If a visualization does indeed cause the reader to "stay on the page a little longer" (Kennedy et al., 2021, p. 1) does that help improve comprehension of the information? Does the visualization itself contribute additional comprehension?

Data Literacy

Ridsdale et al. (2015) define data literacy as "the ability to collect, manage, evaluate, and apply data, in a critical manner" (p. 3). The promise of data journalism is at least partially tied to the data literacy of the readership. The goal of a data visualization is to convey information quickly, but even the most effective visualization requires a level of familiarity on the part of the reader – some level of data literacy. The proliferation of data in the modern world has led not only to a rise in data journalism, but also to a need for increased data literacy in society. Frank et al. (2016) equate the increased importance of data literacy in the internet age to other types of literacy. In a sense, data literacy is to data journalism as literacy is to journalism.

Though it is by no means a new problem, the COVID pandemic has shone a light on the lack of data literacy among many in society. “In the pandemic, data are everywhere” (Nguyen, 2021, p. 210). Has there ever been a line graph more prevalent across society than the one stressing the need to “flatten the curve”? However, the ability to comprehend an x-axis of time and a y-axis of cases might have tested the limits of society’s data literacy. It is clear at this point that few understood the data or the implications behind *keeping* the curve flat, and a logarithmic scale on a y-axis was almost certainly expecting too much of the audience (Nguyen, 2021).

As the importance of data literacy grows, the education system needs to respond and help develop this new competency. Colleges and universities have offered data science classes for some time but now there is an active movement to provide data literacy education in secondary and even elementary schools (Wolff et al., 2016).

D’Ignazio (2017) views the challenge of increased data literacy as a broader societal problem that goes beyond the need for additional education in schools. As with other opportunity gaps that leave some in society at a disadvantage, D’Ignazio (2017) argues that without increased data literacy, societies will see an ever-increasing “gap between the data haves and data have-nots” (p. 6).

To address this need for a higher level of data literacy in society, some educators feel that it is past time to reexamine mathematics education in the US to deemphasize pure math like calculus in favor of more statistics education

(Benjamin, 2009). Some school districts are taking concrete steps to include data science courses. Los Angeles Unified, the second largest school district in the nation, offers an Introduction to Data Science course, which is now recognized by the University of California system as a statistics course that can be substituted for Algebra 2 as an admissions requirement (Boaler & Levitt, 2019).

The challenge of increasing data literacy is significant, and while many are working to meet the challenge, the need will continue to grow as well.

Multimedia Learning

A fundamental premise of data journalism is that including visualizations with text is beneficial to comprehension. Educational research supports this premise, particularly Mayer's Multimedia Learning Theory (2020). The guiding principle of Mayer's Multimedia Learning Theory is the Multimedia Principle which states that "people learn better from words and pictures than from words alone" (p. 4). This is the primary principle this project will examine, that the inclusion of pictures with words will increase comprehension as compared to words alone.

Modality has always been an important concept within journalism. Valkenburg (2016) lists modality as a fundamental property, pointing out that it has long been common to examine the effects of different presentation modes on audiences. The seminal media theorist McLuhan coined the expression "the medium is the message" in 1964 to emphasize the preeminence of modality in media.

The pertinent question to this research is not one of specific modality, but the potential for increasing comprehension through conveying information both through words and through images, addressing multiple modalities with the same message.

Several theories inform this idea. Dual coding theory (Paivio, 1986) posits that the brain has one method for processing images and a separate method for processing words. Comprehension is thought to be reinforced when it is processed by both methods. Visual argument theory (Waller, 1981) holds that graphical representations are less taxing to process than text because, as compared to texts, graphics remove a layer of transformation involved in assigning meaning to the words. Building on these two theories, Mayer (2020) developed the Multimedia Learning Theory that states “Human understanding occurs when learners are able to mentally integrate corresponding pictorial and verbal representations” (p. 7).

While Mayer’s theory was developed addressing the pedagogical implications of multimedia learning, the similarities between a teacher presenting information to students and a journalist presenting information to readers imply that Multimedia Learning Theory is applicable. Whether the “learner” is a newspaper reader or a student in a classroom, the premise of Multimedia Learning Theory is the same: Include a picture (or a chart or a graph or a diagram) with the text to help convey information. For the purposes of this research, we will use the lens of Multimedia Learning Theory.

A prevailing view in pedagogy holds that students learn best when information is tailored to their preferred learning-style. This view is influential in the education field and examples of its importance can be found at all levels of education (Pashler et al., 2008).

However, multiple reviews of learning-styles research have found little evidence to support the validity of the *meshing hypothesis*, which holds that comprehension is highest when instruction is tailored to mesh with the predominant or preferred learning style of the individual student (e.g., Kirschner, 2017; Pashler et al., 2008). In fact, among studies that have attempted to measure the effect, it was more common that the data contradicted the meshing hypothesis (Pashler et al., 2008)

The insufficiency of the meshing hypothesis of learning styles does not however invalidate the principle of Multimedia Learning Theory. In fact, students benefit from different kinds of instruction that address multiple modalities of learning (Yale Poorvu Center, n.d.*)*. So, while there is little empirical evidence that adapting instruction to a specific learning style is effective, that does not preclude the premise that addressing multiple modalities will be effective for most learners.

Research on Visual Displays

The research regarding the impact of visual displays in education is inconclusive considering the conventional wisdom that the inclusion of visuals should always provide a comprehension benefit. In their meta-analysis on the use of visuals, Guo et al. (2020) concluded that among the 23 studies they examined six failed to find a positive comprehension effect with the inclusion of visuals while the other 17 studies reported varying degrees of benefit when including visuals.

The effectiveness of increasing comprehension by adding visual information to text varies by age group as well as by visual complexity and expertise in the specific discipline (e.g., Elia et al., 2007; McTigue, 2009; Renkl & Scheiter, 2017). For example, a simple picture depicting the events in a story can aid comprehension in young readers (Segers et al., 2008). However, in one study the arrows on a water-cycle diagram were often misinterpreted by students as simply pointing at elements of the diagram as opposed to indicating a sequence of events (McTigue and Flowers, 2011). Renkl and Scheiter (2017) found that insufficient content knowledge, as well as a lack of competence in representational models were barriers to achieving the desired learning outcomes of some visual displays. This would seem to correspond with the importance of data literacy in the audience for data journalism to achieve its broadest possible impact.

In an experiment involving science material, middle school students were found to have little or no additional comprehension when given a text with accompanying diagrams compared to those given the text only (McTigue, 2009). Some researchers have found an effect opposite to the one expected: Students given information with text and images sometimes understand less than those who receive text-only versions of the information (e.g., Coleman et al., 2018; Vekiri, 2002). This may be due to the additional information overwhelming the students' processing capacity. In one experimental study, fourth graders were given a science text with three different kinds of diagrams and a control with no diagram: The researchers found that the diagrams added minimal or no additional comprehension as compared to the text-only version and that the diagrams created a cognitive overload for some students (Coleman et al., 2018).

Contrary to the effect proposed by Multimedia Learning Theory, presenting information through both visuals and text can sometimes impede comprehension. Multiple forms of information should be beneficial if they can be processed, but are detrimental if the information is overwhelming. Because of the demands on cognitive processing, it is possible that some students are unable to integrate the information coherently, limiting the effectiveness of the multimedia presentation (Vekiri, 2002).

Overall, the research on visual learning is mixed, showing that the inclusion of visuals does not consistently provide the anticipated comprehension benefits (e.g., Coleman et al., 2018; McTigue, 2009; Renkl & Scheiter, 2017). In cases where the comprehension benefits are missing or limited, researchers have surmised that a lack of familiarity with the subject matter or a lack of expertise in interpreting the visuals may be limiting factors (e.g., McTigue, 2009; Renkl & Scheiter, 2017).

It seems reasonable that prior training on the proper interpretation of specific visualizations will impact the effective use of visualizations to enhance comprehension. In this project we will test for comprehension benefits of the inclusion of a data visualization if the subjects have had prior education in statistics and/or data science (see Hypotheses).

Finally, prior research has specifically addressed how incorporating data visualizations can impact reader comprehension. Ward (1992) used five different versions of a news article – the article alone, with a bar graph, with an illustrated bar graph, with a data table, and with a sidebar story – to test comprehension in college students. The experiment showed that the students retained information best with the sidebar story, and worst with the article alone. Importantly, the two graphical representations tested and the data table did positively impact information retention (Ward, 1992). More recently, a similar study (Parodi & Julio, 2017) examined the importance of subject matter expertise (in this case, economics) in a college student's ability to gain a comprehension benefit from the inclusion of graphics in reading material, finding that the role of disciplinary knowledge was crucial in comprehending a specialized text and specific data visualizations.

Table 1 summarizes the broad findings of some of the research on the impact of visual displays on comprehension. In the studies where comprehension benefits were not achieved or not clear, the authors posited several potential limiting factors which are summarized in the table. In addition to the individual studies listed, Renkl and Scheiter's (2017) meta-analysis also found mixed results for comprehension, concluding that cognitive overload was a limiting factor.

Table 1. Summary of comprehension effects and limiting factors

	Comprehension benefit	Test subject grade(s)	*Lack of Prior Knowledge*	*Overload*	*Cognitive development*	
Study			*Limiting factors*			Display type(s)
Mayer & Gallini (1990)	Yes	UG				diagrams
Ward (1992)	Yes	UG				bar graph
Mayer (1997)	Yes	UG				diagrams
Segers, Verhoeven & Hulstijn-Hendrikse (2008)	Yes	5				pictures
Elia, Gagatsis & Demetriou (2007)	Mixed	1,2,3			x	pictures, number lines
McTigue (2009)	Mixed	6,7,8		x	x	diagrams
Roberts, Norman, & Cocco (2015)	Mixed	3	x			diagram, flowcharts, tables and timelines
Parodi & Julio (2017)	Mixed	UG	x			graphs
Roberts & Brugar (2017)	Mixed	3,4,5	x		x	images, maps, tables, and timelines
Coleman, McTigue & Dantzler (2018)	Mixed	4		x	x	diagrams

Building on this prior research, the proposed experiment will test the impacts on comprehension of overall data literacy (developed through prior education in statistics and/or data science) as opposed to specific subject matter expertise.

Purpose

The purpose of this research is to test the impacts of addressing multiple learning modalities with multimedia learning techniques on comprehension of data intensive information. This research will provide empirical evidence for data journalists as to the benefits of including data visualizations to convey data intensive information to readers.

The research also intends to examine the role of data literacy in the comprehension of data intensive information. As data becomes more and more important in the modern world, data literacy must keep pace. Does prior education in statistics and/or data science increase data literacy and prepare individuals for understanding data intensive information more effectively than more traditional non-data focused mathematics courses, such as Algebra 2 and Calculus?

The project will frame the research through the lens of Richard Mayer’s Multimedia Learning Theory (1997).

This project will examine the following:

- The basic validity of Multimedia Learning Theory (Mayer, 1997) – that as compared to words alone, comprehension is improved through words and graphs as would be present in data journalism;
- The role of data literacy in comprehension of data intensive information and the importance of data literacy education as part of a more comprehensive literacy for all members of society in a data intensive future.

Prior research has often shown the comprehension benefits of including visuals with text, which this project hopes to r

eplicate. However, where the results of earlier studies have been mixed, the absence of a comprehension benefit is often attributed to a lack of specific prior knowledge. If there is a research gap which we hope to fill, it is in examining general data literacy as a necessary component for achieving comprehension benefits from the inclusion of data visualizations with data intensive text.

Methods

Research Questions and Hypotheses

This project intends to examine the basic premise of Multimedia Learning Theory that comprehension is improved through words and images as compared to words alone as well as the role of data literacy in comprehension of data intensive information. This research is important to data journalism as it addresses one of the perceived benefits of the field, that the addition of data visualizations makes information more accessible and comprehensible. The research questions underscore the importance of a data literate society.

These ideas inform the following research questions (RQs):

> **RQ1:** What is the effect of the inclusion of a data visualization on comprehension of data intensive information?
>
> **RQ2:** What is the effect of education in statistics and/or data science on comprehension of data intensive information?

Further, based on the prior research conducted in this area, I propose the following hypotheses:

> **H1:** Subjects with a data visualization supplementing their reading material will demonstrate better comprehension of the material than those without a data visualization, controlling for other variables.
>
> **H2:** Subjects who have had a course in statistics and/or data science will demonstrate better comprehension of the material than

those who have not had statistics and/or data science, controlling for other variables.

Experimental Design

To address the effects of data visualization and data literacy on information comprehension, I propose to use data gathered from an online experiment conducted with a sample of N=180 undergraduate students. The experiment will consist of a 3X2 factorial design manipulating the inclusion or exclusion of a data table or a data visualization in a text-based article and the effect of data literacy. One experimental treatment will therefore be the format of the information, presented to the subjects in three different versions:

- Version A – Text-only
- Version B – Text with tabular data
- Version C – Text with a data visualization (a line graph)

The three versions of the reading are included in Appendix 1. The original article is from the CNBC website and includes the charts. All the articles were edited identically for length. Version A was edited to remove references to charts and the charts themselves. Version A is 349 words long. Version B was also edited to remove references to charts while replacing the charts with data tables. Version B is 340 words long. Version C was edited only for length and is also 340 words long.

The second experimental condition is whether students have prior education in statistics and/or data science. Prior educational experience will be determined in a post-test survey.

The dependent variable will be comprehension, as determined by a test based on the material. All three formats will contain sufficient information to correctly answer all the questions on the test. In other words, there will be no questions specific to data displayed in the table or in the graph that is not also referenced in the text. In addition, participants will be asked to self-report their preferred “learning style”: Visual, Auditory, Read-write or Kinesthetic. Additional questions will ask about participants’ gender, race and prior math and statistics education. Note that based on the literature would indicate that the self-reported learning style will not be a factor in comprehension, but the project will examine it as a possible confounding factor.

The study participants will consist of N=180 undergraduate students from the University of Missouri. Within this sample of students, participants will be randomly assigned to a version of the reading material.

To minimize the risk of identifiable information attached to study participants, any identifiable information will be removed and responses will be anonymized. Minimal demographic information will be collected. The data will be stored without specific identifiers of participants. Any results from the research will be reported in aggregate.

Materials will be presented to students online. Participants will be given time to read the assigned version of the article and prompted to proceed to the quiz questions after they have completed reading. They will not be allowed to return to the article once they proceed. Time spent on the reading will be recorded. Identical quizzes to determine the students' comprehension of the material will be administered to each group of students.

Analysis of results

After students have read the material and taken the quiz, the data will be analyzed. Mean test scores will be compared by conducting two ANOVA tests – one for the version of the article and for prior education in statistics and/or data science – as well as post-hoc analysis.

H1 will be supported if participants given Version C have statistically significant ($p<.05$) higher mean comprehension scores than those given other versions.

H2 will be supported if participants with statistics/data science education have statistically significant ($p<.05$) higher mean test scores than participants without statistics/data science education.

Limitations

There are a number of factors which may limit the findings of the proposed project. First, the results of this study are highly dependent on the experimental device selected (the news article and graphic). It is possible that the device chosen will be too simplistic, such that it is easily comprehended with or without a data visualization. In this case, the comprehension test might show no effect due to the entire cohort scoring very high. Likewise, it is possible that the device chosen may be too complex, such that it is too difficult to comprehend with or without a data visualization. In this case, the comprehension test might show no effect due to the entire cohort scoring very low. To minimize the chances of this happening, the device will be pilot tested with a small cohort of students. Second, it is possible that participants may not spend sufficient time to analyze and comprehend the device. Should participants rush through the reading or answer the questions randomly the results will be unreliable. To limit the possibility of this occurring, a quality check tracking time spent on the online survey page with the device will be included and participants who are more than 3 standard deviations below the average time spent on the page will be removed from the final sample. Finally, it is possible that the methods chosen will fail to accurately identify a data literate cohort. There is no generally accepted means of testing data literacy, and education in statistics and/or data science may or may not be a valid proxy. To minimize the possibility of this impacting the results, two separate measures of data literacy are included.

Professional Component

The professional component of this project will be a lesson plan for my High School Introduction to Data Science class.

The lesson will occur over the course of two days during the second semester. The timing is significant. It should be late enough in the Introduction to Data Science (IDS) curriculum that a data literacy effect would be expected. It should also occur near the beginning of the final unit of Algebra 2, a unit on statistics and probability, so that it fits with the flow of the Algebra 2 class as well.

On day one of the lesson, students in IDS, as well as those in Algebra 2 classes, will be given a warmup which consists of the reading and comprehension quiz from this proposal. Versions of the article will be assigned to students randomly. The IDS students will be presumed to be the data literate group and the Algebra 2 students will be the non-data literate group. This portion of the lesson will only encompass the warmup for day one. The remainder of the day will be within the normal flow of the curriculum.

The second part of the lesson will occur a week later, after analysis of the data. This lesson, for the IDS students only, will consist of a discussion of the project they have now taken part in. We will discuss the way that research is designed and the way data is collected and analyzed. Assuming the hypotheses are supported, the lesson will also demonstrate the role of data visualizations and data literacy in comprehension of complex information which will help to validate for the students the importance of the class they are taking.

Implications

This project is a continuance of the research outlined in this proposal. It hopes to verify that the inclusion of a data visualization will enhance comprehension of data intensive information for readers, as Multimedia Learning Theory would suggest.

Furthermore, it hopes to validate the importance of data science education as a means of increasing data literacy, a vital life skill in the modern world as data becomes increasingly important. Broadening the understanding of the role of data literacy in the audience can inform the use of data journalism going forward.

Data journalism has much to gain if we can better understand effective ways to include visualizations with text and measure comprehension benefits. Society has much to gain if its members are more data literate.

As has been discussed, the amount of data available has exploded in recent years. The importance of data and of data literacy will only continue to grow. The important work of data journalists can only keep pace if broader society has the data literacy to understand the information data journalists produce.

We have seen during the COVID pandemic, during political campaigns and in many other instances the use and misuse of information. Data journalists have a responsibility to try to accurately present information, and data visualizations can be a key tool in doing so.

At the same time, part of being an informed consumer of information in modern society is having the data literacy to correctly interpret the data intensive information available. Data journalism and data literacy both have vital roles to play in maintaining a well-informed society in the future.

Appendix A

Article Treatments

Version A

How much does it cost to charge an EV vs, refueling a gas vehicle?

It has been true for years: Mile for mile, it's cheaper — generally much cheaper — to recharge an electric vehicle than it is to refuel one with an internal-combustion engine.

That has been a key selling point for Tesla and other EV makers, particularly in times when gas prices have soared, such as now. But this time there's a wrinkle: While gas prices have indeed soared in the wake of Russia's invasion of Ukraine, so have electricity prices — particularly in some parts of the U.S. that have been big markets for Tesla's EVs.

That raises a question: Is it still true that it's much cheaper to "refuel" an EV? The charts below help us find the answer.

We looked at three sets of data, the first being nationwide figures, which provides a baseline. The other comparisons were for data specific to Boston and San Francisco, two markets where EVs are popular — and where electricity tends to be more expensive than the national average.

The answer in all three cases is that — even with regional surges in the price of electricity — it's still quite a bit more expensive to fill your gas tank than it is to charge your EV's battery.

Electricity rates have roughly kept pace with gas price increases in Boston and San Francisco. Yet, on average across the U.S., adding 100 miles of range in your internal-combustion vehicle has become more expensive, relative to charging an EV an equivalent amount, over the last couple of months.

Is that likely to change? While oil prices are nearly certain to fall in coming months as producers increase output, it's unlikely that the price of electricity will rise enough to make EVs less affordable over their life cycles than internal-combustion alternatives.

Using February data, Jeffries analyst David Kelley recently calculated that the total lifetime cost of ownership of an EV is about $4,700 less than that of an internal-combustion vehicle. He said that cost difference is likely to increase as more EVs come to market — and as battery prices continue to fall — over the next couple of years.

Version B

How much does it cost to charge an EV vs, refueling a gas vehicle?

It has been true for years: Mile for mile, it's cheaper — generally much cheaper — to recharge an electric vehicle than it is to refuel one with an internal-combustion engine.

That has been a key selling point for Tesla and other EV makers, particularly in times when gas prices have soared, such as now. But this time there's a wrinkle: While gas prices have indeed soared in the wake of Russia's invasion of Ukraine, so have electricity prices — particularly in some parts of the U.S. that have been big markets for Tesla's EVs.

That raises a question: Is it still true that it's much cheaper to "refuel" an EV? The data below help us find the answer.

The first table, using nationwide figures, provides a baseline. The others use data specific to Boston and San Francisco, two markets where EVs are popular — and where electricity tends to be more expensive than the national average.

U.S. national averages

	02/19	08/19	02/20	08/20	02/21	08/21	02/22
Gas	$9.33	$10.56	$9.88	$8.86	$10.09	$12.69	$14.08
Electricity	$4.72	$4.82	$4.65	$4.75	$4.75	$5.00	$5.14

Source: U.S. Bureau of Labor Statistics for the electricity rates and U.S. Energy Information Administration for the gas prices

Boston gas and electricity averages

	02/19	08/19	02/20	08/20	02/21	08/21	02/22
Gas	$9.61	$10.84	$10.07	$8.55	$9.89	$12.21	$14.04
Electricity	$8.09	$7.43	$7.98	$7.25	$7.98	$7.91	$9.33

Source: U.S. Bureau of Labor Statistics for the electricity rates and U.S. Energy Information Administration for the gas prices

San Francisco gas and electricity averages

	02/19	08/19	02/20	08/20	02/21	08/21	02/22
Gas	$13.06	$14.20	$13.77	$12.94	$13.67	$17.41	$18.78
Electricity	$7.15	$7.70	$7.95	$8.33	$8.40	$9.13	$9.68

Source: U.S. Bureau of Labor Statistics for the electricity rates and U.S. Energy Information Administration for the gas prices

The answer in all three cases is that — even with regional surges in the price of electricity — it's still quite a bit more expensive to fill your gas tank than it is to charge your EV's battery.

Electricity rates have roughly kept pace with gas price increases in Boston and San Francisco. Yet, on average across the U.S., adding 100 miles of range in your internal-combustion vehicle has become more expensive, relative to charging an EV an equivalent amount, over the last couple of months.

Is that likely to change? While oil prices are nearly certain to fall in coming months as producers increase output, it's unlikely that the price of electricity will rise enough to make EVs less affordable over their life cycles than internal-combustion alternatives.

Using February data, Jeffries analyst David Kelley recently calculated that the total lifetime cost of ownership of an EV is about $4,700 less than that of an internal-combustion vehicle. He said that cost difference is likely to increase as more EVs come to market — and as battery prices continue to fall — over the next couple of years.

Version C

How much does it cost to charge an EV vs, refueling a gas vehicle?

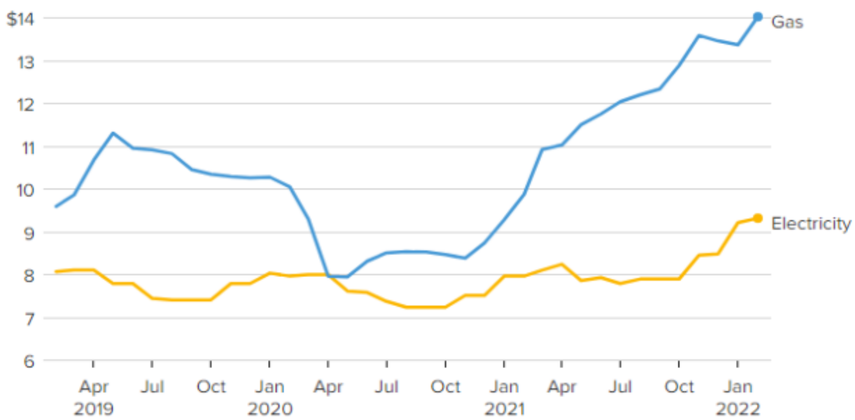
It has been true for years: Mile for mile, it's cheaper — generally much cheaper — to recharge an electric vehicle than it is to refuel one with an internal-combustion engine.

That has been a key selling point for Tesla and other EV makers, particularly in times when gas prices have soared, such as now. But this time there's a wrinkle: While gas prices have indeed soared in the wake of Russia's invasion of Ukraine, so have electricity prices — particularly in some parts of the U.S. that have been big markets for Tesla's EVs.

That raises a question: Is it still true that it's much cheaper to "refuel" an EV? The charts below help us find the answer.

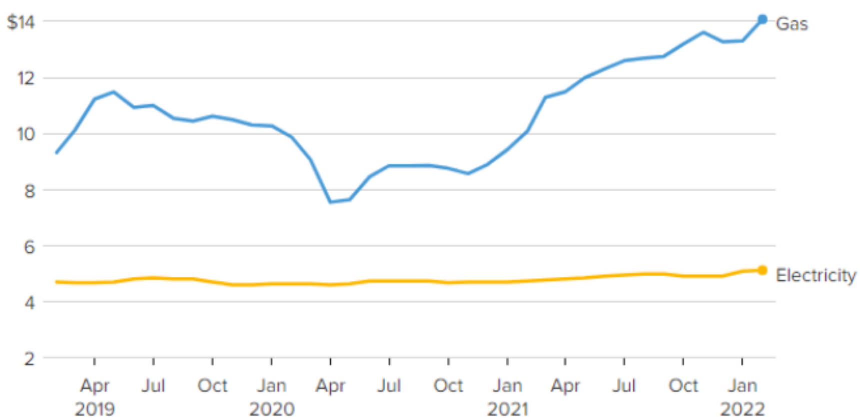
The first chart, using nationwide figures, provides a baseline. The others use data specific to Boston and San Francisco, two markets where EVs are popular — and where electricity tends to be more expensive than the national average.

Boston gas and electricity averages



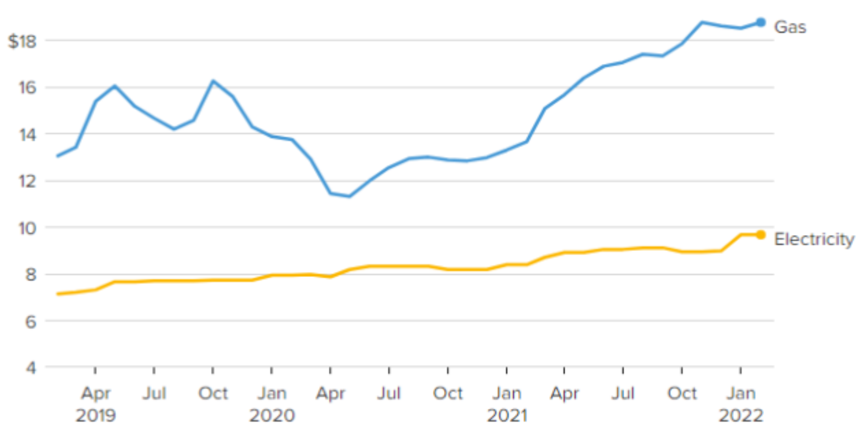
Source: U.S. Bureau of Labor Statistics for the electricity rates and U.S. Energy Information Administration for the gas prices

U.S. national averages



Source: U.S. Bureau of Labor Statistics for the electricity rates and U.S. Energy Information Administration for the gas prices

San Francisco gas and electricity averages



Source: U.S. Bureau of Labor Statistics for the electricity rates and U.S. Energy Information Administration for the gas prices

The answer in all three cases is that — even with regional surges in the price of electricity — it's still quite a bit more expensive to fill your gas tank than it is to charge your EV's battery.

Electricity rates have roughly kept pace with gas price increases in Boston and San Francisco. Yet, on average across the U.S., adding 100 miles of range in your internal-combustion vehicle has become more expensive, relative to charging an EV an equivalent amount, over the last couple of months.

Is that likely to change? While oil prices are nearly certain to fall in coming months as producers increase output, it's unlikely that the price of electricity will rise enough to make EVs less affordable over their life cycles than internal-combustion alternatives.

Using February data, Jeffries analyst David Kelley recently calculated that the total lifetime cost of ownership of an EV is about $4,700 less than that of an internal-combustion vehicle. He said that cost difference is likely to increase as more EVs come to market — and as battery prices continue to fall — over the next couple of years.

Appendix B

Surveys/Tests

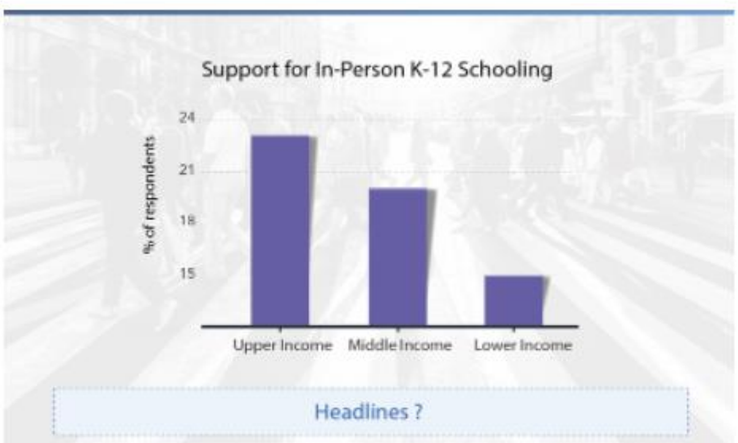
Section 1: Pre-test Survey

Which is your preferred learning style?

- Visual (I learn and retain information best when I see it)
- Auditory (I learn and retain information best when I hear it)
- Read-write (I learn and retain information best when I read it or write it down)
- Kinesthetic (I learn and retain information best when I can touch or manipulate something about it)

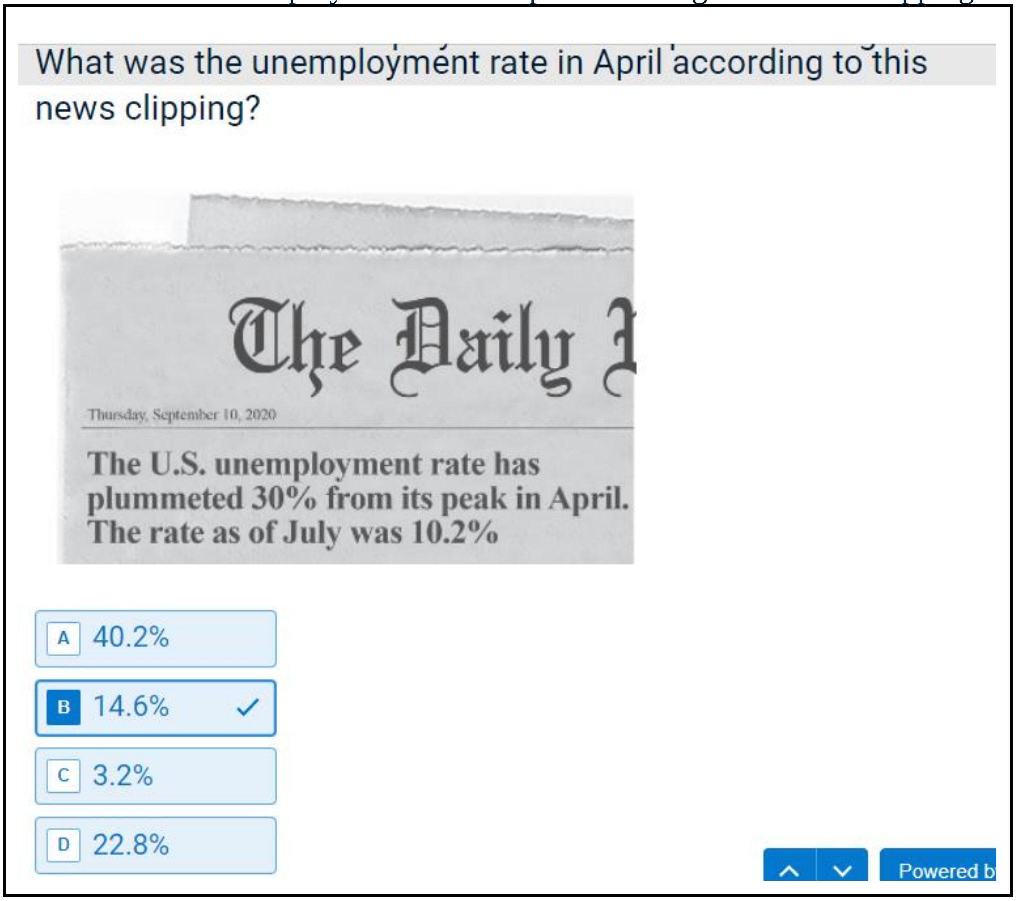
Section 2: Data literacy

1. Which of these news headlines is appropriate, given the data presented below?

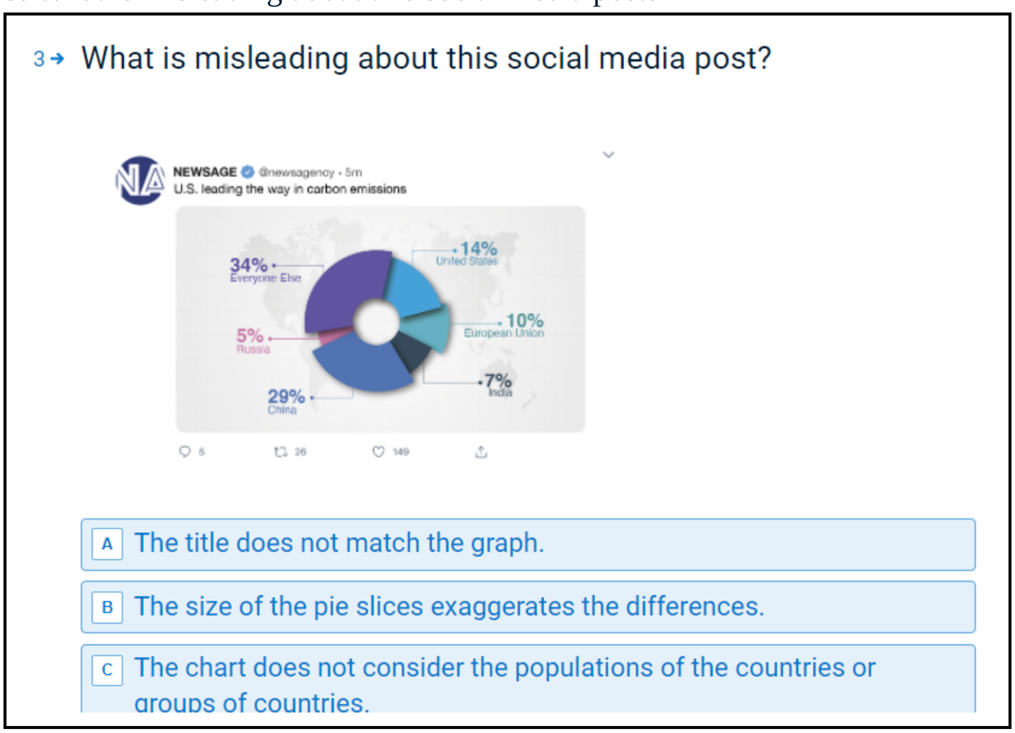


A. Most high income families support in-person K-12 schooling.
B. Lower income families are less likely to support in-person instruction for K-12 schooling.
C. Middle- and upper-income families are twice as likely to support in-person schooling compared to low-income families.
D. Lower income families are more fearful of in-person instruction.

2. What was the unemployment rate in April according to the news clipping?



3. What is misleading about this social media post?



Section 3: Experiment - readings

Experimental conditions:

- Version A – Text-only
- Version B – Text with tabular data
- Version C – Text with a data visualization (a line graph)

Section 4: Experiment - comprehension

Answer the following questions based on the reading.

1) As gas prices in the US went up following Russia's invasion of Ukraine, what happened to electricity prices in the US?

a. Electricity prices went up just as much as gas prices in some markets
b. Electricity prices went up equally across the nation
c. Electricity prices stayed low across the nation
d. Electricity prices went up on average, but more in some markets than in others

2) According to the analysis in the article, which statement best summarizes the difference between the cost to drive 100 miles in an average electric car versus 100 miles in an average gas engine car?

a. It is always less expensive to drive 100 miles in an electric car
b. It is always less expensive to drive 100 miles in a gas engine car
c. It is only less expensive to drive 100 miles in an electric car in certain cities
d. On average, it is about the same to drive 100 miles in an electric car as it is to drive 100 miles in a gas engine car.

3) Which of the following statements is true about the change in the prices for gas and electricity from Feb 2019 (before the pandemic) to March 2022?

a. The prices of gas and electricity have both risen by a similar amount.
b. The price of gas has risen more quickly than the price of electricity
c. The price of electricity has risen more quickly than the price of gas
d. The price of gas has risen while the price of electricity has dropped

4) Which of the following statements about the popularity of Teslas is supported by information in the article?

a. Teslas are more popular in Boston than they are in San Francisco
b. Teslas are popular in cities where electricity rates are low
c. Teslas are popular in Boston and San Francisco
d. Teslas are popular in all US cities

5) Which of the following statements best describes the main point of the article?

a. The war in Ukraine made everything more expensive.
b. It is significantly more expensive to fill the tank of a gas engine car than to charge the battery of an electric car
c. It is slightly more expensive to fill the tank of a gas engine car than to charge the battery of an electric car
d. Regional differences in electricity prices make it impossible to know whether it is more expensive to fill the tank of a gas engine car or to charge the battery of an electric car

----
Section 5: Post-test

Now we want to know how comfortable you are with statistics and data. How much do you agree with the following statements?

When I see a data visualization, such as a line graph or bar graph…

Strongly disagree	*Disagree*	*Somewhat disagree*	*Neither disagree nor agree*	*Somewhat agree*	*Agree*	*Strongly agree*
1	*2*	*3*	*4*	*5*	*6*	*7*

A. I would know how to interpret the data on my own.
B. I would get nervous when first encountering new types of data visualizations.
C. Much of the content would be too technical for me to understand.

How confident are you that you could understand and interpret a data visualization?

Not at all confident	*Slightly confident*	*Moderately confident*	*Confident*	*Very confident*
1	*2*	*3*	*4*	*5*

How would you rate your knowledge of statistics?

Not at all knowledgeable	*Slightly knowledgeable*	*Moderately knowledgeable*	*Knowledgeable*	*Very knowledgeable*
1	2	3	4	5

Which high school math courses have you taken previously (select all that apply)? Do your best to select a comparable course from the list. If you have taken a math course that you feel is not represented on the list, please include the course name and the institution in the "Other course" field(s).

- Algebra 2
- Statistics
- Pre-calculus
- Calculus
- Data science course
- Other course ____________________________
- Other course ____________________________

Have you ever taken any college-level math courses (including calculus)?
A. Yes
B. No
C. Don't know
D. Prefer not to say
(If yes--)

How many college-level math courses have you taken (including calculus)?
A. 1-2 courses
B. 3-5 courses
C. 6-10 courses
D. 11-20 courses
E. More than 20 courses

Have you ever taken any college-level statistics or data science courses (including data visualization, research methods)?
A. Yes
B. No
C. Don’t know
D. Prefer not to say
(If yes--)

How many college-level statistics or data science courses have you taken (including data visualization, research methods)?
A. 1-2 courses
B. 3-5 courses
C. 6-10 courses
D. 11-20 courses
E. More than 20 courses

Section 6: Demographics

What year were you born?

What is your gender?

- Male
- Female
- Non-binary
- Prefer not to state