8 Deconstructing data visualizations: What every teen should know

Susan Smith

Emily wakes up to the sound of her cell phone alarm, thinking about what lies ahead for her at school as a high school senior. Her phone screen lights up with today's celebrity news, texts from friends, and calendar reminders of school assignments. Before she puts her feet on the floor, she posts to her Facebook page and checks the day's news, laden with images, graphs, and charts. In her first period environmental science class, the textbook requires her to decipher a chart depicting climate change in order to project the consequences of continued warming. During her free period she decides to research her father's recently diagnosed heart disease, as she's feeling anxious about the family discussion last evening. After school she's in charge of reporting her lacrosse team's stats to the school newspaper, then remembers she needs more acne medicine, and checks online reviews

and data to see which one is most effective. Before she leaves school, friends reminded her to register to vote in the upcoming primary, but she is unconvinced that young voters can really make a difference, so she consults her state's polling data.

Each of these activities involves data analysis and reading data in visualized form, some of it rather sophisticated. Today's information universe is vast, and there is too little time to scrutinize data, especially when it is embedded in colorful graphics with a convincing message. Data visualization has long been used by academic researchers to summarize their findings, but more recently publishers and commercial marketing firms have embraced the use of visualization to impart information in new ways that are attractive and space-efficient (Centaur Communications 2010). These graphics draw the reader in, provide an instant takeaway, and are easy to share on social media. In order to *read* these charts, graphs, and statistics, today's teen needs a toolkit for data literacy that requires a critical eye and a keen understanding of the creator's intentions.

How do we move beyond a simple check of the sponsoring website that published the information? When data and graphics have different creators, how do we evaluate the infographic – holistically or piecemeal?

As educators, we must model what it means to be a critical consumer of data visualizations and other visual representations of data while at the same time giving our students a few rules of thumb to help simplify the process in their everyday lives. We can promote healthy skepticism by helping students learn how to effectively *question* what they view. Students love to critique things; we can teach them a framework to do this constructively. By doing so, we can help them grapple with uncertainty and use their questions to spur further investigation. Evaluating data visualizations across the curriculum as well as providing opportunities for students to create their own visual depictions of real

world data are important additions to 21st-century information literacy skills.

You may be thinking, *Why can't I just leave data literacy to math and science teachers?* We know post-secondary students will likely contribute to the creation of data in their discipline and must become proficient at content-specific data visualizations. The challenge in high school is to both scaffold for these advanced academic skills as well as to prepare students to intelligently consume visualized data in their everyday lives. To this end, analyzing graphs and charts "in the wild" teaches them to identify misleading or ambiguous representations. Surveys of college faculty reveal that while students master the use of chart- and map-making tools quickly, they do not know what makes a *good* visualization (Carlson and Johnston 2015).

In a 2012 Pew study on teens and technology, media expert Sam Punnet observed, "[A]ll communications [in the teen world] must be short, visual, and distracting/entertaining" (in Anderson and Rainie 2012). Data visualizations fit this bill; they tell a story in a compact combination of words and images. Data are often embedded in infographics to lend credibility; after all, numbers do not lie … do they? Complicated infographics containing data visualizations require students to evaluate graphical techniques. In a world where visual information is preferred by consumers and marketers alike, students need to leave high school able to *read* data visualizations as powerful sources of information across disciplines, both in and out of the classroom.

How can you help students gain data literacy skills if the discipline you teach doesn't typically include data? Keep an eye out for teaching materials, collect subject-relevant examples embedded in media coverage of current events, scrutinize your textbook with a fresh eye for data and charts, offer visuals as an alternative to textual information, and consult resources in this book for professional development in this area. In the meantime, we can

identify some simple evaluation tactics to add to our toolkit. This chapter will help you think more like a designer, extend basic information literacy skills, and add some new heuristics so you can de-mystify data visualizations for your students.

Think like a designer

Data visualization expert Edward Tufte (2006) suggests that every designer makes moral and intellectual choices in the creation of an infographic, and in consuming them we must hold the creator(s) responsible for those decisions. Every visualization involves a series of intentional choices, and while we can never fully appreciate each, it is helpful to *read* all graphics with this in mind. Every visualization is a picture with a message; students must think in terms of the designer's agenda or perspective, e.g., *Why was this data, and the story it tells, depicted in a series of bar graphs rather than a pie chart, and why the use of bright, bold colors?* As consumers, we must actively engage with each graphic; as teachers we need to model this in each discipline.

Infographics and data visualizations represent a burgeoning new field of marketing. Companies spend 62% of their content marketing dollars on infographics, up 15% over a year ago (Content Marketing Institute 2016). An estimated 84% of Internet communication will be visual by 2018, according to *Reuters* (Lopes 2014). In the world of marketing, designers work with their clients to craft a visual message consistent with the company's mission. Similarly, political campaigns depend on infographics to quickly communicate their track record to voters on issues like job creation. In this case, graphics are often designed to make the reader feel aligned with the candidate, and every detail — from the range and unit of data selected, to the color palette used, to the font type used for text labels — is carefully selected to support the message. A graph of the same data may look very different if produced by BarackObama.com rather than WashingtonPost.com.

One site is intended to persuade you to agree with a candidate's policies, and the other's goal is primarily to inform from an objective standpoint. As readers, it's our job to decipher the message, given the perspective of its creator(s), rather than to merely react to the image and accept the data visualization at face value. As teachers, we need to help students to develop this fluency.

Extending information literacy skills

Information literate students already learn how to evaluate text sources. They can leverage this knowledge to deconstruct infographic data visualizations. Using techniques similar to those employed for the evaluation of text sources (in which we prompt students to consider who created it, on what authority, over what period of time, and reflecting what point of view) teachers must consider the following when teaching about visual information:

- **Provide students with time to look and think about the message, and begin to generate their own questions.** Unlike a purely textual source, there are many layers to visualized data, including text, graphics, data, proportion, and color. What is seen initially, changes with time and focus given to each element. When students generate their own questions about the data, they engage with the graphic.
- **Graphics, and the datasets within, may originate from different (or multiple) sources.** Students need to interrogate how and when data was collected, analyzed, and ultimately displayed; the answers are essential to judge the veracity of an infographic.
- **Look for the source of the data, not the URL where the infographic may have been found.** Basic website evaluation techniques usually include deconstructing URLs and investigating the domain. Such common practices give rise to black-and-white *rules* for students such as "All .gov sites can be trusted" or ".edu sites are all

created by scholars." However, where a data visualization is published (or re-published) on the web may be unrelated to who generated the data or the agenda they suggest. For example, while respected universities often produce datasets and infographics, they may be funded by an outside interest or be republished on blogs or online journals. "Following the money" may uncover the story behind the story.

A few basic rules of thumb

While mastering all the design and statistical principles necessary to evaluate visualizations is daunting, we can lead students through their deconstruction.

A key question to ask is, "What is the story, and who are the storytellers?" Every infographic seeks to tell a story using images or icons and sparse text, laid out cleverly to convey a message. It is important to consider both your first impression (the thesis) as well as deeper questions about each element. Question the story: Are the claims clear, reasonable, and accurate?

What can we ask about the graphs, charts, and images? How can we better understand the choice of visualization?

What does the data really mean?

Having examined the story and those telling it, we now turn our attention to the data itself. Americans often accept data at face value; we – and our students! – tend to accept numbers as irrefutable "facts." When students don't know much about the claim being made, it seems overwhelming to try to evaluate the data. Like a textual search, there is a need to build context in order to evaluate. The graphs and charts contained in an infographic

Who

- Who has created the visual components (often a graphic designer)? What can you find out about his or her credentials that suggests expertise in the issue being described?
- Who collected, analyzed, and published the data?
- What perspective or objectives might the creator or organizational sponsor represent?

What

- Try to summarize the thesis — or big takeaway — of this graphic. What does the creator want you to think? Is the goal to persuade, sell, or inform? Who is the storyteller?
- Is there an alternative explanation?
- What information is missing?
- Is the claim plausible based on what you know? Can you do an Internet search and find a trusted source to corroborate the findings?

When

- When was the chart or graph created? Is it different from the publication date of the visualization and/or the article/post in which the visualization appears?
- When was the source data collected? Is it reasonably current? If not, consider the implication of using old data to make new claims.

Where

- Where did the data come from? Was the data collected by a research team at a university? A non-profit? What else have they published on this topic? Does knowing the source change the context?
- Where is the data visualization published; does it suggest an agenda?

Why/How

- How might I better understand this topic? What context or background information is required?
- How does the author or organization's perspective affect the interpretation of the data?

WHAT IS THE STORY? WHO ARE THE STORYTELLERS?

reflect a constellation of choices, and this is where we must focus our evaluation. Page XXX provides some questions to discuss with students, and Appendix B provides a sample infographic accompanied by a sample student conversation using those questions.

Numerical information

When analyzing data, we can ask many questions of the numbers themselves. Here are some useful strategies to keep in mind:

- **Correlation of two sets of data should not be used to imply cause and effect.** For example, the fact that autism and ice cream consumption both increase over some period of time does not imply that ice cream consumption causes autism. This may be the simplest yet one of the most valuable data literacy skills.
- **Can you quickly verify some benchmark statistics** (Best 2013), e.g., population, distance, birth or death rates that can help provide context for the data and numbers in a visualization? Providing students with a short list of statistics related to the subject of the infographic can make this easier.
- **Do the numbers seem too big or too conveniently rounded?** Big, round numbers are often inexact guesses (Best 2013).
- **What is being measured over what period of time?** Are the units of measure consistent, or are different units used to describe a single phenomenon? Does it seem implausible based on what you know? Odd units or preposterous claims should be questioned.
- **Check the y-axis on the graph.** We are used to seeing graphs in math class where the intersection of the x- and y-axes begin at zero. In the wild, this may not be the case, so students need to consciously check the labels

on graphs, particularly the y-axis. A y-axis that does not begin at zero is not inherently incorrect – it could be that truncating the range of numbers is space-efficient or that there are no data points that correspond with numbers close to zero. On the other hand, graphics with non-traditional or unexpected y-axes can raise questions: Was the range of data selectively chosen, perhaps to cherry-pick a short-term trend to support an argument? Would you feel differently about the trend if a longer time period was depicted, or a wider range of numbers? Were the intervals selected to support the creator's intended conclusion?

» **Be cautious when looking at percentages, percentile, and percent change.** As noted in Chapter 1, these terms sound similar but have distinct meanings.

Methods

An infographic's raw data and the organization or researchers that produced it may not always be included, but some additional searching may lead to these kinds of details.

» **Can you locate the sample size (expressed as n=x)?**

» **Is the methodology for data collection available, especially information on how the sample was gathered?** Was it randomized, meaning the researcher's data collection accurately depicts a larger population or dataset? Beware of generalizations made using small samples (n<25). For more on sampling, see Chapter 1.

» **Are the intervals regularly or irregularly spaced?** *Logarithmic graphs* are used to depict change over time, and feature an accelerated rate of change using unevenly spaced intervals. They can be useful when a few data points are skewed away from the bulk of the data or

when depicting percent change (rather than absolute). Absolute graphs feature regularly-spaced intervals (the same distance is present between each). Beware of charts comparing data graphed on an absolute scale with those using a logarithmic scale.

» **Is the data cherry-picked?** Are date ranges or time boundaries specifically chosen because they demonstrate a desired result? If you looked at a different or broader interval, would the data look very different? For more on intervals, see Chapter 6.

» **Beware of the word *average*.** Many people assume that average means the arithmetic mean (adding up all numbers and dividing by the quantity of numbers), but the author may mean median or mode, which can distort the raw data. For more discussion on averages, see Chapter 1.

» **Ask questions of text and labels**

- Does the title convey a thesis or put forth an argument? Does this match your "read" of the data?
- Can you spot emotionally-laden words: rhetoric (words like freedom-fighter or terrorist) and hyperbole (claims like the most, the largest, or the best) or is the language neutral and purely descriptive?
- Do font size or color connote importance (size) or caution (red, orange)? Were either used to intentionally mislead? Read the fine print for important details that the creator may have tried to de-emphasize.
- Are all values clearly labeled using consistent units? Do the labels clarify or confuse? Are the terms defined, e.g., the definition of *family* versus *household* when used to collect income data.

Layout

Helping students understand the organization and flow of elements in an infographic can aid in improving comprehension.

Here are some questions you may ask of your students.

- **Is this the best way to visualize this data?** If not, can you find it graphed another way or find the original data set and lay it out differently? See Figure 2 for some common charts and when to use them (Maguire 2016).
- **What does the color choice tell us?** How do the colors used make you feel? Red usually signals trouble, or at least emphasis. Are the colors communicating emotion that is supported by the data? Or distracts from or over-emphasizes it?
- **Is the graphic layout designed to emphasize or de-emphasize some of the data?**
- **Does the relative size of elements/icons accurately represent the data?** Respective icon size should reflect data not emphasis; the area of two-dimensional shapes should be proportional to values.

For a sample conversation about layout, see Appendix C.

Practice, practice, practice

Short lessons, peppered throughout the high school curriculum, will make evaluation of visualized data a habit of mind. Plan collaborative lessons with other teachers to scaffold learning throughout the year, for example:

- **Conduct a whole-class evaluation of a single infographic;** assign groups to evaluate one of the categories.
- **Display a lesson-relevant image as a warm-up exercise,** coupled with a checklist to complete or a required post to an online class discussion. Introduce a small subset of the criteria shown in this chapter at a single sitting so that

students gain evaluative experience and confidence over time.

- **Assign groups to evaluate one of a set of related infographics on a single topic** and have each share what their evaluation revealed. Be sure to include "what's missing?" as a prompt.
- **Include an infographic requirement to a research paper** and require a critical annotation to explain why it was chosen as a source.
- **Post the chart in Appendix D and remind students that the selection of a visualization style can be critical in assisting readers' comprehension.**

Conclusion

Every opportunity we provide students to critically evaluate infographics improves their data literacy. Ultimately – and with practice – students will transfer the critical eye we require in the classroom to the myriad graphics they confront in their everyday lives.

Resources

Anderson, Janna, and Lee Rainie. 2012. *Main Findings: Teens, Technology, and Human Potential in 2020*. Washington, D.C.: Pew Research Center. Accessed May 3, 2016. http://www.pewinternet.org/2012/02/29/main-findings-teens-technology-and-human-potential-in-2020/.

Best, Joel. 2013. *Stat-spotting: A Field Guide to Identifying Dubious Data*. Berkeley: University of California Press.

Carlson, Jake, and Lisa Johnston. 2015. *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers.* Purdue Information Literacy Handbooks. West Lafayette, IN: Purdue University Press.

Centaur Communications Limited. 2010. "Data Visualization: Facts and Figures." *New Age Media,* September 2.

Content Marketing Institute. 2016. *B2C Content Marketing: Benchmarks, Budgets, and Trends— North America.* Cleveland, OH: Content Marketing Institute. Accessed May 1, 2016. http://contentmarketinginstitute.com/wp-content/uploads/2015/10/2016_B2C_Research_Final.pdf.

Lopes, Marina. "Videos May Make Up 84 Percent of Internet Traffic by 2018: Cisco." Edited by Andre Grenon. Reuters.com. Last modified June 10, 2014. Accessed May 1,

2016. http://www.reuters.com/article/us-internet-consumers-cisco-systems-idUSKBN0EL15E20140610.
Maguire, Sara. 2016. "No Fuss Chart Design." Business2Community.com. March 26. Accessed April 4, 2017. http://www.business2community.com/content-marketing/no-fuss-chart-design-cheat-sheet-01494593
Tufte, Edward R. *Beautiful Evidence*. 2006. Cheshire, Conn.: Graphics Press.

Appendix A: Tips for integrating visualization analysis into your classroom

If you have five minutes:

- **Create a warm-up exercise with an infographic related to the content you are teaching.** Post a journal prompt for students: "What do you think the thesis is for this infographic?"
- **Screen an infographic for 1 minute, and then have students describe in writing or orally how the color, font, or layout of the infographic made them feel.** Have them describe their first impressions.
- **Screen an infographic from a news article or popular science topic.** List 4 possible "messages" and place them in a 4-Corners Game (see Figure 1). Ask students to move to the corner that best represents the message they took away from the infographic.

If you have thirty minutes:

- **Ask students to compare two infographics on the same topic.** Break into groups to discuss which is more effective and why.
- **Ask students to select an effective infographic from among 3-5 chosen for evaluation (representing varying levels of accuracy).** Focus the evaluation on 3 criteria, e.g., data visualization (chart choice), color, and text labels.

If you have one class period:

- **Hold a debate on a topic where only infographic evidence is allowed.** Model how to search for infographics using "Images" in a search engine. Break the class into groups and have them debate both sides of an issue in pairs, using infographics they find and evaluate, then present as evidence of their stance.
- **Using the criteria outlined in this chapter,** ask students to present a critical summary of a single infographic, either found by them or selected by you.

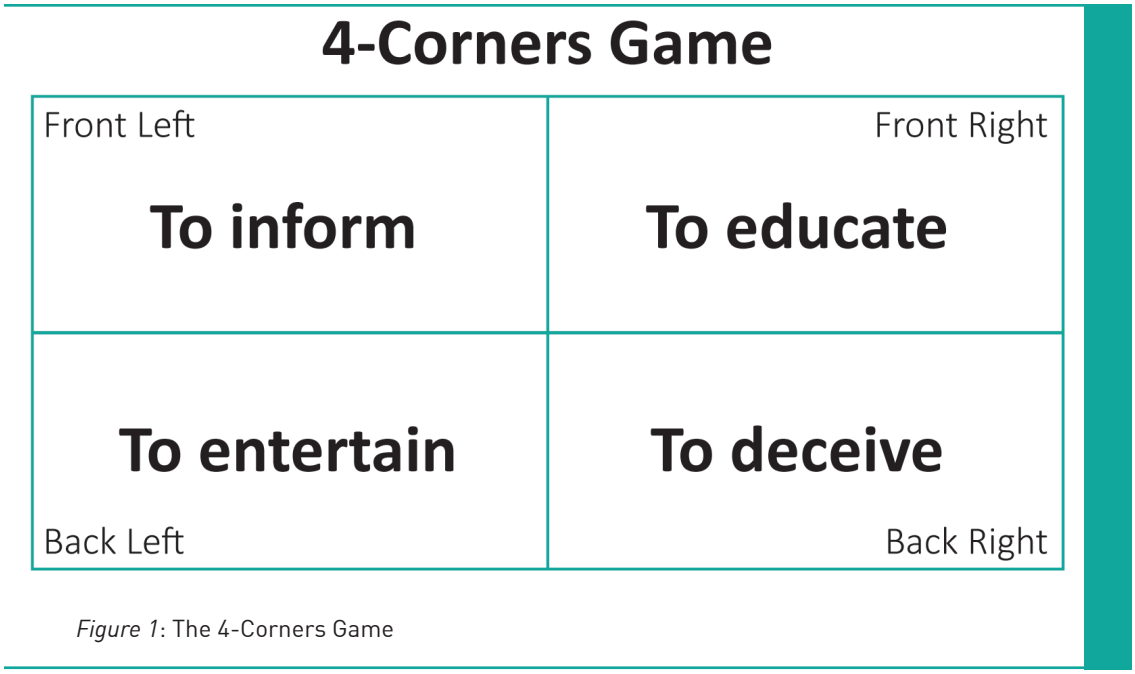


Figure 1: The 4-Corners Game

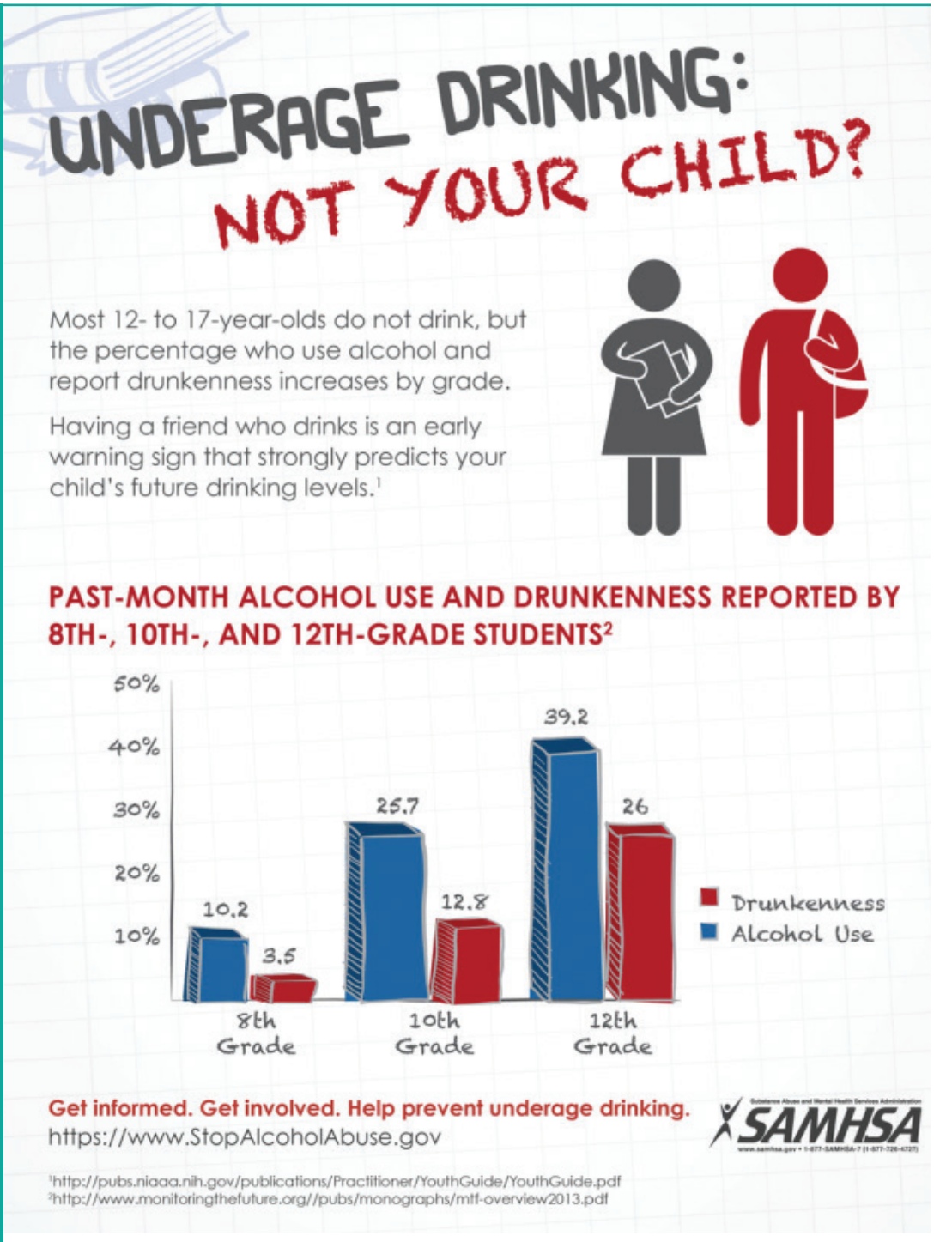


Figure 2: Source: https://www.stopalcoholabuse.gov/resources/infographics/share.aspx?info=13

Who created this?

- **Substance Abuse and Mental Health Services Administration (SAMHSA)** is the lead government agency for StopAlcoholAbuse.gov, the web portal for the Interagency Coordinating Committee on the Prevention of Underage Drinking (ICCPUD)
- **Data comes from MonitoringTheFuture.org's 2014 report on adolescent drug abuse.** Monitoring the Future (MTF) is a non-governmental organization (NGO) that conducts and publishes results from an annual "long-term epidemiological study that surveys trends in legal and illicit drug use among American adolescents and adults as well as personal levels of perceived risk and disapproval for each drug. The survey is conducted by researchers at the University of Michigan's Institute for Social Research, funded by research grants from the National Institute on Drug Abuse, one of the National Institutes of Health." [Wikipedia]
- **Perspective or agenda?** Produced by a government agency that seeks to reduce teen drug use.

What is the message or thesis?

- **Teen drinking increases from grades 9 thru 12.**
- **Having a friend who drinks predicts likelihood that a student will use alcohol.**
- **Seems like the storyteller wants to educate parents about underage drinking.** Audience is parents.
- **What information is missing?** Maybe it would be helpful to know if the data represents urban/rural, low income/high income, students with low or high GPAs, etc. This data is from a 2013 survey; does it represent an increase or decrease from previous years?
- **Does this information make sense?** Can you corroborate the findings elsewhere? Responsibility.org, drugabuse.org, and other sites all refer to this report. Seems like this is a respected source.

When was the chart or graph created? When was the source data collected? Is it different from the publication date of the visualization?

- **No publication or copyright date is listed for the graphic.**
- **The data in the MonitorTheFuture.org report was collected in 2013; the report was published in 2014.**
- **The data seems reasonably current.**

Where did the data come from? Was the data collected by a research team at a university? A non-profit? What else have they published on this topic? Does knowing the source change the context?

- **University of Michigan's Institute for Social Research** conducts the MTF survey each year (1975-2013) on a range of drug use by adolescents.
- **This lends credibility to this infographic** because it is conducted by a major research university and sponsored by several government agencies.

How might I better understand this topic? What context or background information is required? How does the author or organization's perspective affect the interpretation of the data?

- **Read the "Alcohol" section on page 37 of the MTF report** to understand how "drunkenness" and "alcohol use" are defined in this study.
- **Read the "Study Design and Methods" section on page 3 of the MTF report** to answer questions about how many students were surveyed, what sampling methods were used, and what specific questions were asked.
- **University researchers would be likely to use a scientific approach** rather than having a specific agenda on this topic.

Appendix C: Unpacking visual decisions in an infographic

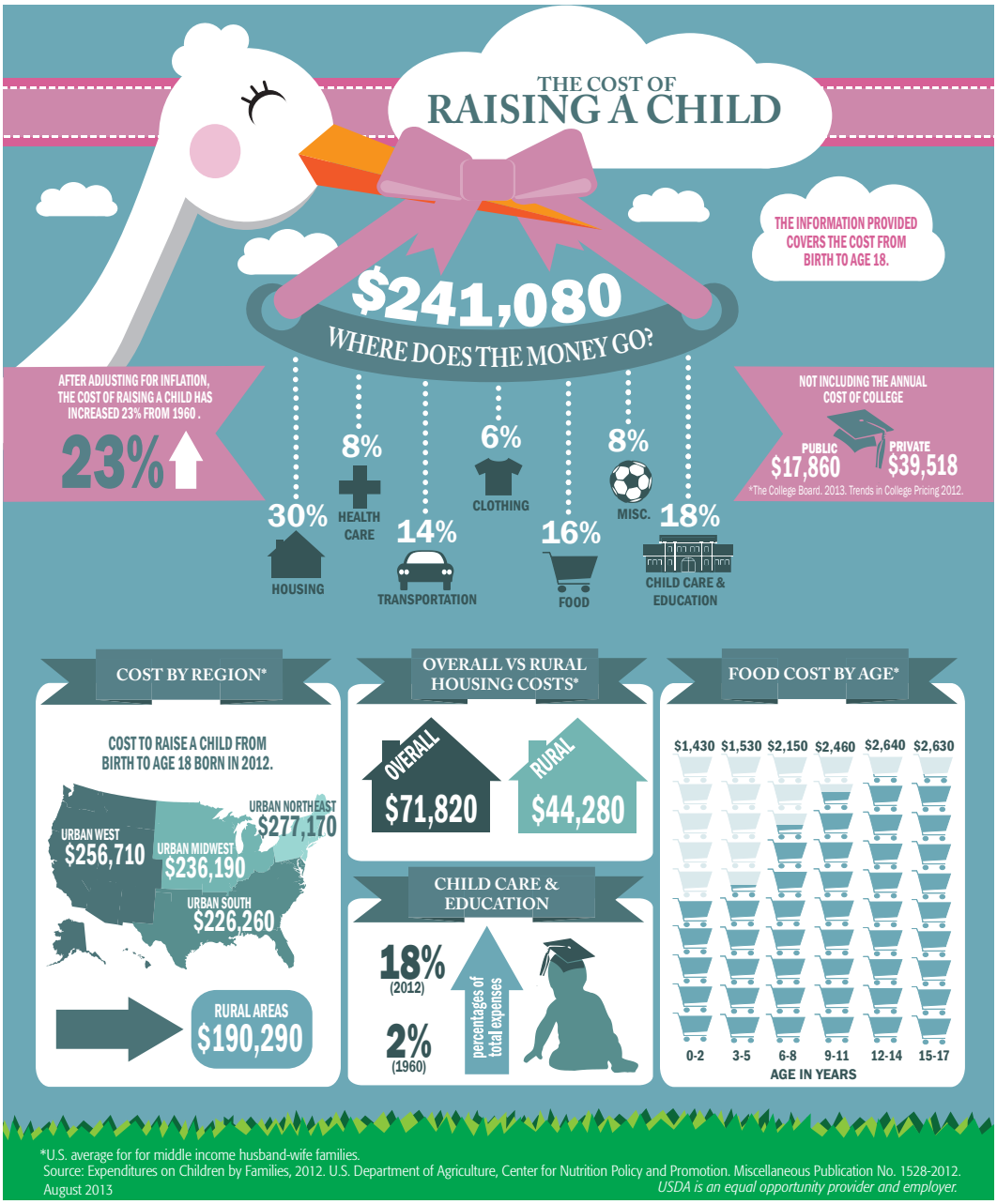


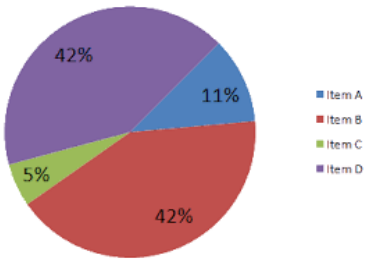
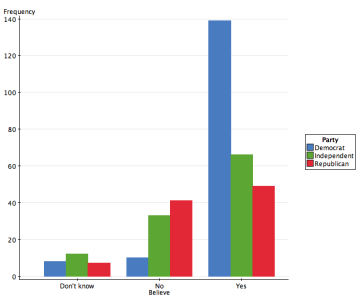
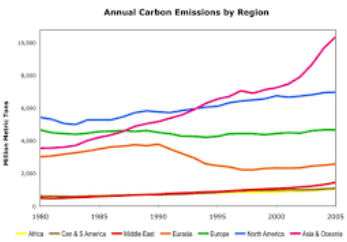
Figure 3: August 2013 infographic by the United States Department of Education showing the anticipated cost of raising a child from birth to age 18. Source: https://www.cnpp.usda.gov/sites/default/files/expenditures_on_children_by_families/CRC2012InfoGraphic.pdf

Activity: Evaluating an infographic's layout

Consider questions like these when evaluating the aesthetic choices in "The Cost of Raising a Child."

1. **Is this the best way to visualize this data?** Could the components of the $241K be displayed to better differentiate between the high cost of housing vs. the relatively low cost of clothing?
2. **Consider questions like these when evaluating the aesthetic choices in "The Cost of Raising a Child."** How does the the color pink make you feel? The rosy-cheeked stork? Is this just about girls? Could cost of raising a girl vs. boy differ?
3. **What stands out?** Does the designer want to focus on some data more than others? Why is the 23% increase since 1960 in such a large font? Why are college costs included when the graphic supposedly depicts costs only from birth to age 18 and most students are in college beyond age 18?
4. **Do the icons accurately represent the data?** Does each shopping cart depict the same cost? Should the urban and rural house icons be the same size?

Appendix D: Choosing effective visualization types

Type of Chart	Best for. . .	
Pie Chart	Showing the relative proportion of variables as parts of a whole. Beware of many small "slices," especially without labels. Pie charts make comparisons difficult.	
Bar Graph (horizontal or vertical)	Illustrating discrete quantities of a number of outcomes. Must include labels and legends.	
Line Graph	Showing change in a variable over time. Labels and units of measure must be clear.	
3D Graphs (any type)	Nothing! 3D graphics are inherently difficult to interpret without error.