Exploring the Types of Messages that Pie Charts Convey in Popular Media

Richard Burns[1], Eric Balawejder[1], Wiktoria Domanowska[1], Stephanie Elzer Schwartz[2], and Sandra Carberry[3]

[1] Dept. of Computer Science, West Chester University, West Chester, PA 19383 USA
rburns@wcupa.edu
[2] Dept. of Computer Science, Millersville University, Millserville, PA 17551 USA
[3] Dept. of Computer Science, University of Delaware, Newark, DE 19716 USA

Abstract. In popular media, information graphics (pie charts, bar charts, line graphs) are frequently used to convey high-level intended messages. This paper focuses on the pie chart graphic type. We have collected a corpus of pie chart information graphics found in popular media, and for each chart, a team of annotators recognized its intended message. In this paper, we report on the types of intended messages that the team of annotators recognized and their inter-annotator agreement. We also briefly survey some of the communicative signals that graphic designers used which helped the annotators recognize these messages.

1 Introduction

Information graphics, such as bar charts and line graphs, are common visual devices frequently incorporated into multimodal documents to achieve a set of communicative goals [6] [5]. In popular media (magazines such as *Time* and newspapers such as *USA Today*), information graphics are sometimes included in an article to convey some additional, supplemental high-level message that transcends supporting data, rather than simply providing low-level data points. For example, the grouped bar chart in Figure 1 ostensibly conveys a high-level message that *"Women are more likely than men to delay medical treatment"*.

The idea that information graphics can be considered a form of language follows Clark [3] who noted that language is any "signal" or lack thereof, where a signal is any deliberate action that is intended to convey a message, including gestures and facial expressions. Thus, we view information graphics as a form of language, where the designer of a graphic is able to deliberately use *communicative signals* to help convey an intended message to the viewer of the graphic.

This paper presents preliminary results in our study of designing a system that can automatically reason about the most likely intended message of a pie chart, using present or absent communicative signals in the graphic as evidence.

It is non-trivial to identify the intended message of an information graphic; Carberry [2] found that a graphic's message is often not contained in the graphic's

caption or in the article accompanying the graphic. Thus, the use of natural language processing techniques only on the graphic's caption or only on surrounding article text cannot be relied on to provide enough evidence to recognize the graphic's high-level message.

Previously, our research group has implemented intended message recognition systems for other kinds of information graphics: simple bar charts [4], line graphs [7], and grouped bar charts [1]. These three implemented systems use a Bayesian network to probabilistically capture the relationships between high-level intended messages and communicative signals that help signal the messages. Because each type of information graphic is able to convey a unique set of possible messages compared to the other information graphic types, the end-result for each of the systems has been very different. Simple bar charts, line graphs, and grouped bar charts each have a different set of message categories, and different communicative signals are utilized by graph designers to help convey the high-level intended messages.

Women more likely to delay medical treatment

Percentage of Americans who postpone doctor's visits because of costs:

37%

21%   21%

17%

Under 49 years old   Over 50 years old

Men

Women

Fig. 1. From *USA Today.*

This work is the first of our knowledge that studies the problem of recognizing the intended high-level message of a pie chart when it is drawn in popular media.

We have collected a set of pie chart information graphics occurring in popular media, and examined these charts to identify (1) the types of high-level messages that graphic designers convey using pie charts, and (2) the kinds of communicative signals present in pie charts that appear likely to assist the recognition of high-level messages. Unsurprisingly, in our preliminary investigation so far, the types of recognized high-level messages and identified communicative signals are different than those in simple bar charts, line graphs, and grouped bar charts.

One application of this research is for sight-impaired individuals who cannot view information graphics. Alternative access screen readers can convert the content of a pie chart to text, but only at the level of low-level raw data: (e.g. *"the first pie chart slice is 18.5%, the second pie chart slice is 7.3%, ..."*). Our research aims to generate the high-level message as text for sight-impaired users.

Section 2 of the paper describes some of the messages categories that we identified and Section 3 presents some of the communicative signals that we found. Section 4 introduces some unexpected properties of pie charts in popular media that could be avenues for interesting future work.

2   Pie Chart Message Categories

We collected 115 pie chart information graphs from popular media.[4]. Of those, we retained 90 of the charts, as the rest appeared to contain *only* data, and did

---

[4] The corpus of pie charts is available at: http://www.cs.wcupa.edu/rburns/piecharts
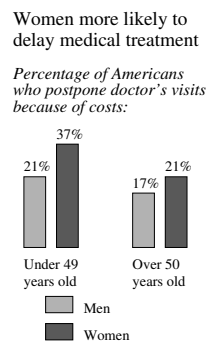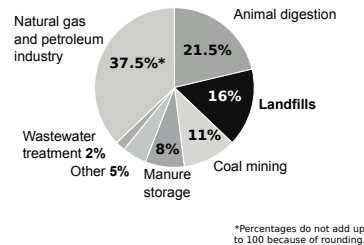
U.S. METHANE EMISSIONS

Fig. 2. From *National Geographic*.
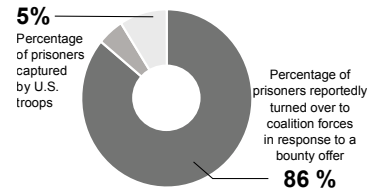
A Dismal Decade

Fig. 3. Graphic from *Time Magazine*.

not appear to the annotators to convey any intended message. (Inter-annotator agreement is discussed later.) We then analyzed the corpus to generalize the kinds of high-level intended messages that we recognized into *message categories*.

There are nine pie chart message categories that we defined. Because of space constraints, we can only present graphical examples for a subset of the message categories. Below, we formally define the name of each category, the number of parameters that messages in each category take, and provide a short description.

SingleSlice($< s >$). Single slice messages recognize a high-level message that involves a single, salient, pie chart slice. Generally, the pie charts that fall within this category seem to be designed so that the graph viewer compares a specific, single slice against the other slices in the pie chart. For example, consider the pie chart in Figure 2. This pie chart ostensibly conveys that *Landfills are a significant source of U.S. methane emissions, the third highest, behind the natural gas and petroleum industry as well as animal digestion.* The parameter $< s >$ in the message category syntax is instantiated with the single pie chart slice that is to be compared against the other slices. That is, this message would be represented as: *SingleSlice(s = Landfills)*.

Versus($< s_1, s_2 >$). Versus messages capture two salient slices, which are compared against each other. In contrast to single slice messages in which a salient pie chart slice is compared against the rest of the slices in the pie chart, the two salient slices in versus messages are compared with each other rather than the other slices. For example, the pie chart in Figure 3 ostensibly conveys the message that *most prisoners were turned over to coalition forces because of bounties, rather than being captured by troops.* The versus message category is instantiated with two parameters: s_1 and s_2, the slices that should be compared with each other.

BiggestSlice(). Biggest slice messages identify a single slice of the pie chart that is larger than all of the other slices. Because only one slice can be the largest (assuming no ties), the biggest slice message category has no parameters. For example, presumably the intended message in the pie chart in Figure 4 is that

there were a greater number of male deaths than female deaths in which illicit fentanyl was detected.

NoMajority(). No majority messages capture that none of the slices in the pie chart are larger than 50%. For example, the pie chart in Figure 5 ostensibly intends to convey the high-level message that *individuals in search of work take a variable range in time in order to find a job.*

Fraction($< s >$). Fraction messages represent that slice $< s >$ is a fractional percentage of the pie chart, such as the messages *juniors make up <u>one third</u> of the class* and *<u>half</u> of the revenue is from Philadelphia.*

AddSlices($< s_1, s_2, ..., s_n >$). Add slices messages recognize the aggregation of multiple slices. Each slice that is added together is a parameter in this category.

TwoTiedForBiggest($< s_1, s_2 >$). Two tied for biggest messages capture that two slices in the pie chart are approximately the same size.

SmallestSlice(). Smallest slice messages identify a single slice of the pie chart that is smaller than all of the other slices.

NumberOfParts(). Finally, number of parts messages capture the quantity of slices in a pie chart, for a message such as, *there are <u>six</u> reasons identified for not working among uninsured adults.*

2.1   Most Frequent Message Categories

The information graphic types of simple bar charts [4], grouped bar charts [1], line graphics [7], and pie charts, each have a different set of message categories though some categories do overlap. As shown in Table 1, the top two most frequent message categories for each graphic type contain around 30-50% of the collected graphics in those corpora. Notably, while the most frequent pie chart messages involve a *single salient slice*, the most frequent simple and grouped bar chart messages are distributed between either a *trend* message or a message
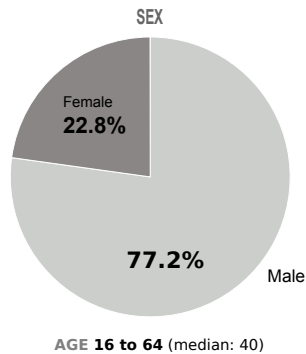
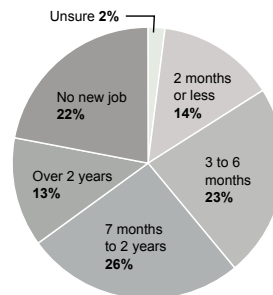

Fig. 4. From *The Philadelphia Inquirer.*

Fig. 5. From *The Philadelphia Inquirer.*

Table 1. Most frequent high-level messages by information graphic type.

Message Category	Number of Occurrences	Percentage
Simple Bar Charts		
Increasing-Trend	26	23.6%
Maximum	25	22.7%
Grouped Bar Charts		
Entity-Comparison	68	20.6%
Rising-Entities-All	36	10.9%
Pie Charts		
Biggest-Slice	24	23.3%
Single-Slice	20	22.2%
Line Graphs		
Rising-Trend	66	27.5%
Change-Trend	58	24.2%

that involves a *single bar entity*. The most frequently occurring messages in line graphs involve *trends*. These results highlight the importance of studying each of the information graphic types separately, and also can be used to inform the process of designing appropriate information graphics.

2.2 Annotation and Inter-Coder Agreement

The annotation of the corpus was performed with the following process: we first individually recognized the intended message for each pie chart and classified it into its appropriate intended message. Then, we conducted a consensus-based annotation by meeting as a group and discussing each of our annotations, revising any annotations if we were strongly swayed. The final annotation for each pie chart was decided by majority vote.

Three coders met and deliberated final annotations for 30 of the pie charts in the corpus. Notably, all of the individual annotators sometimes recognized exactly the same message for a pie chart before any discussion, or a majority of them agreed to exactly the same message after a discussion.[5] This level of agreement is a good result and shows that (1) the recognition of pie chart messages is not as subjective as it may initially appear, and (2) our derived and recognized set of pie chart message categories does capture the types of messages that graphic designers convey in popular media using pie charts. A summary of the inter-annotator agreement is shown in Table 2.

3 Communicative Signals

The presence and absence of communicative signals assist the recognition of a high-level intended message conveyed in a pie chart.

---

[5] Two annotations were only counted as matching if they had: (1) the same message category and (2) the same instantiation. For example, the two messages *SingleSlice(Landfills)* and *SingleSlice(Animal digestion)* would not be a match because their instantiations are not identical.

Table 2. Summary of the annotation agreement between coders. Table rows display *The percentage of pie charts that ...*

Percentage	Description
36.6%	all coders recognized with exactly the same message, before any discussion.
56.6%	a majority of coders recognized with exactly the same message, before any discussion.
63.3%	all coders recognized with exactly the same message, after discussion.
100%	a majority of coders recognized with exactly the same message, after discussion.

Visual Signals. One visual signal that a graphic designer may use to help communicate some intended message is prominence, by coloring a specific pie chart slice a salient coloring, or boldfacing the label of a pie chart slice. An example of this communicative signal is present in Figure 2, which helps signal that *Landfills* should be compared against the other pie chart slices. Another example of a visual signal found in the pie chart corpus is the use of similar colors across multiple pie chart slices. For example in Figure 3, the slices for *Bounty* and *Troops* are colored similarly (though not exactly identical), helping signal that they should be compared, while still contrasting them against the *Unlabeled 9%* slice.[6] Another example of a visual, communicative signal is separation, when one pie chart slice is purposely drawn slightly "separated" or "exploded" away from the center of the pie, drawing additional attention to it.

Linguistic Signals. Although it does not always fully capture a graphic's intended message, the caption text of a pie chart can sometimes serve as a linguistic signal that helps convey its message. For example, in the pie chart in Figure 6, the verb *split* helps signal the intended message that *there is no majority slice amongst the slices: "will", "will not", and "unsure"*. We have also observed instances of the article headline of a multimodal article helping to signal the intended message of a pie chart. Another linguistic clue that can serve as a communicative signal is when one pie chart slice is mentioned in the caption or article headline, while the other slices are not mentioned.

4 Conclusion

There are several avenues of future work that we are exploring: First, we are currently constructing a Bayesian network, which has a top-level node with states that enumerate all possible pie chart messages. This top-level node is linked to children leaf nodes that represent the possible communicative evidence in a graphic. Given our corpus of pie chart graphics, we will train the network to learn the probabilistic relationships between pie chart high-level intended messages and the communicative evidence that is present or absent in the charts.

Second, we have observed numerous instances of *multiple pie charts* drawn adjacent to one another, where the single intended message of the graphic seems

---

[6] In the original graphic, *Bounty* is colored yellow, *Troops* is orange, and the unlabeled slice is gray.
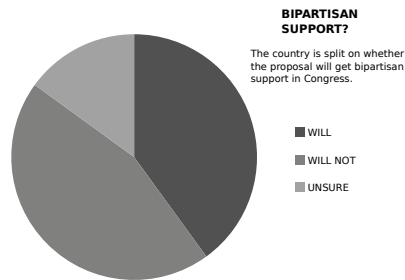
Fig. 6. Graphic from *USA Today*.
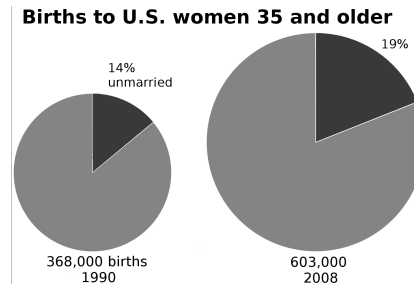
Fig. 7. From *National Geographic*.

to involve *both* pie charts, rather than two individual and separate intended messages. For example in the multiple pie charts shown in Figure 7, the high-level message conveyed is that *the percentage of births to unmarried U.S. women 35 and older increased from 1990 to 2008.* This avenue of future work explores the unique types of messages and communicative signals that can be found when multiple pie charts are purposely drawn adjacent to each other.

Summary. In this paper, we have presented novel research that introduces (1) a corpus of pie charts that we have collected from popular media, (2) a sampling of the types of messages that pie charts are able to convey, and (3) examples of communicative signals that help communicate these messages. These identified messages and communicative signals are unique compared to other types of information graphics that have been previously studied.

References

1. Burns, R., Carberry, S., Elzer, S., Chester, D.: Automatically recognizing intended messages in grouped bar charts. In: Proceedings of the Seventh International Conference on the Theory and Application of Diagrams. pp. 8–22 (2012)
2. Carberry, S., Elzer, S., Demir, S.: Information graphics: An untapped resource of digital libraries. In: Proceedings of 9th International ACM SigIR Conference on Research and Development on Information Retrieval. pp. 581–588 (2006)
3. Clark, H.: Using Language. Cambridge University Press (1996)
4. Elzer, S., Carberry, S., Zukerman, I.: The automated understanding of simple bar charts. Artificial Intelligence 175(2), 526–555 (February 2011)
5. Green, N.L., Carenini, G., Kerpedjiev, S., Mattis, J., Moore, J.D., Roth, S.F.: Auto-brief: an experimental system for the automatic generation of briefings in integrated text and information graphics. International Journal of Human-Computer Studies 61(1), 32–70 (2004)
6. Iverson, G., Gergen, M.: Statistics: The Conceptual Approach. Springer-Verlag, New York (1997)
7. Wu, P., Carberry, S., Elzer, S., Chester, D.: Recognizing the intended message of line graphs. In: Proceedings of the 6th International Conference on Diagrammatic Representation and Inference. pp. 220–234. Diagrams' 10, Springer-Verlag (2010)