Beyond Memorability: Visualization Recognition and Recall

Supplemental Material

Michelle A. Borkin*, *Member, IEEE*, Zoya Bylinskii*, Nam Wook Kim, Constance May Bainbridge, Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, *Senior Member, IEEE*, and Aude Oliva

In this supplemental material we include additional experimental details, extra analyses, and additional visual and textual examples. This extra material provides additional evidence for the main conclusions we make in the paper, and in some cases provides extra insight.

In the document below, for easy navigation, we include content listed under section headings that correspond to the sections in the main text. We also provide relevant context as needed to connect these supplemental material sections to the main text.

For more information about our database and the visualizations included in this supplement, please visit: http://massvis.mit.edu

(Note: Due to potential copyrights of the discussed visualizations, we only include thumbnail-sized images of the visualizations in our paper and this supplement.)

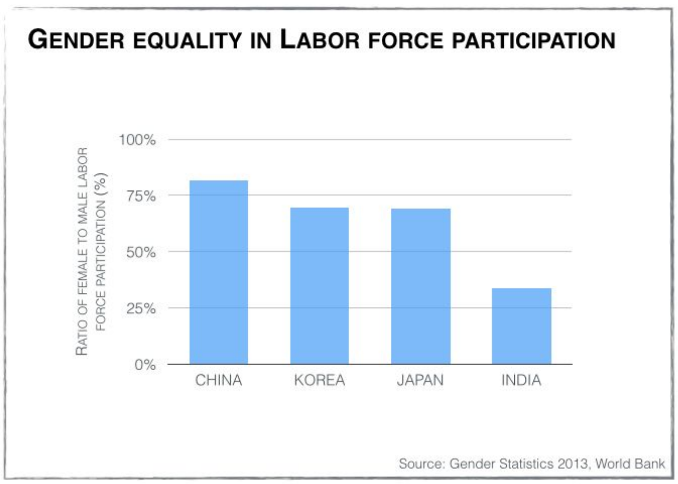
Section 3: Data Collection and Annotation

To gain deeper insight into what elements of a visualization affect its memorability, recognition, and recall, three experts (students who had all completed the Harvard University introductory course in visualization) manually labeled the polygons of each of the visual elements in the 393 target visualizations using the LabelMe system[1]. Labels were reviewed for accuracy and consistency, and corrected by a visualization expert. As part of LabelMe, labels were recorded as polygons in xml and then converted to binary masks for further analysis.

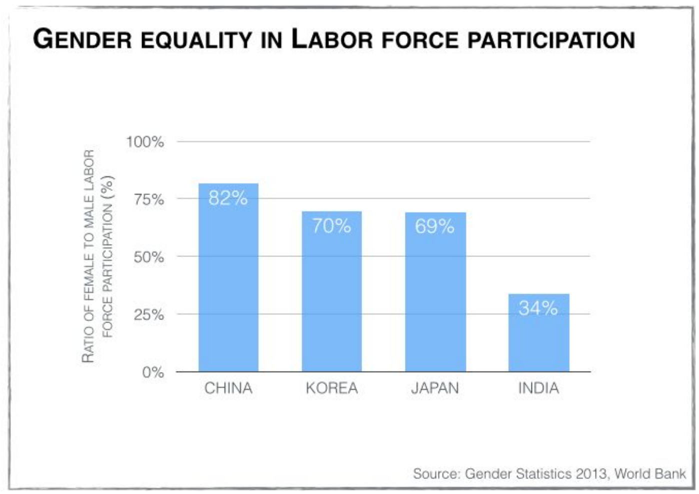
Additionally, we documented whether each of the 393 target visualizations exhibited data and message redundancy. A visualization exhibits data redundancy if the data being presented is visually encoded in more than one way. This can include the addition of quantitative values as labels (e.g., numbers on top of bars in a bar chart or on sectors in a pie chart), or the use of channels such as color, size, or opacity to represent a value already exhibited in a visualization such as the x- or y-axis values. In contrast, a visualization exhibits message redundancy if the main conclusion or message of the visualization is explicitly presented to the viewer in multiple ways. For example, the addition of explanatory annotations, labels, text, and pictures. A visualization can exhibit both data and message redundancy. Also note that our definitions for data and message redundancy do not take into account any task based considerations or requirements. We present examples of how data and message redundancy can be added to the data in the bar chart (Fig. S1), and scatter plot (Fig. S2) below.

---
[1] http://labelme.csail.mit.edu/Release3.0/

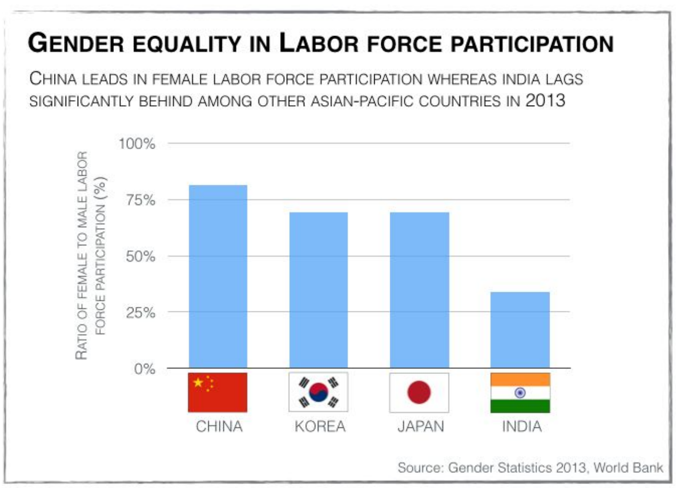
Fig. S1: Examples of how to add redundancy to a bar plot: here, the data redundancy consists of annotated numerical values on top of the bars, and message redundancy consists of flag pictograms to represent the countries on the x-axis.



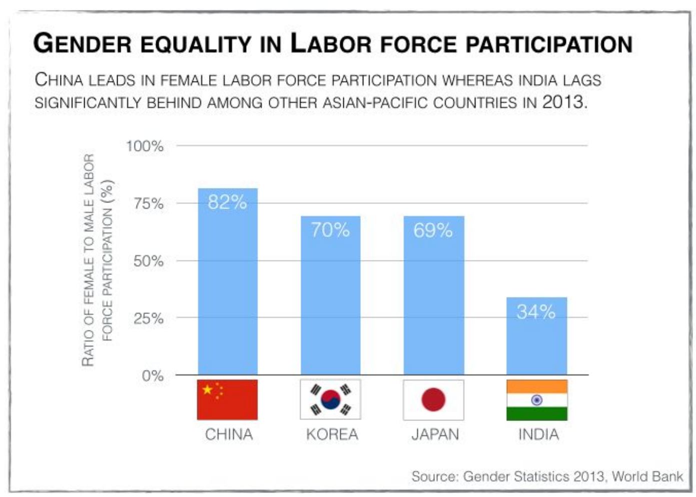
ORIGINAL



DATA REDUNDANCY



MESSAGE REDUNDANCY



DATA & MESSAGE REDUNDANCY

Fig. S2: Examples of how to add redundancy to a scatter plot: here, the data redundancy corresponds to scaling the data markers proportional to the plotted values, and message redundancy corresponds to annotating the trend line with explanatory text.

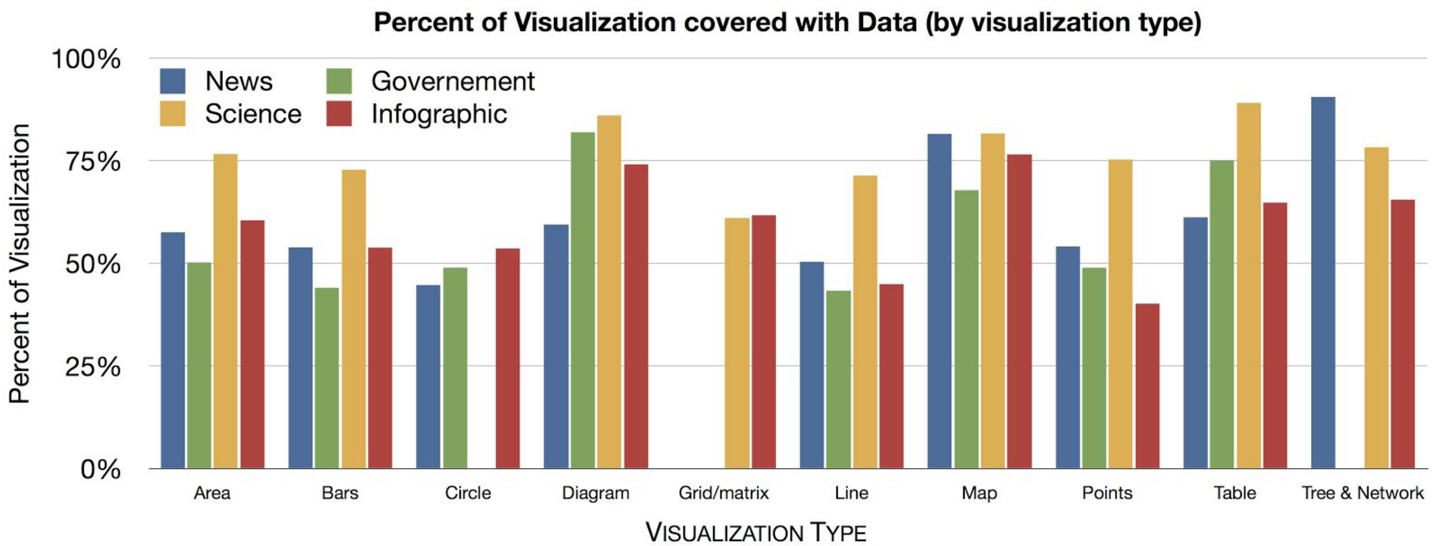
Section 4: Analysis of Labeled Visualizations

In Fig. S3 we present a tagcloud of the human recognizable objects (HRO) annotated in our labeled targets. The size of each word is proportional to its frequency of occurrence. The human recognizable objects are primarily in the form of company logos (McDonalds, Twitter, Apple, etc.), international flags commonly used in the news media visualizations, and pictograms or photographs of human representations and computer/technology depictions.

Fig. S3: Tagcloud of human recognizable objects labeled in our visualization dataset.

In Fig. S4 we plot the percent of average image area covered by the data label for each visualization type per visualization source. The scientific journals in general have the largest image area devoted to displaying data. The visualization types with the largest areas for data display are diagrams, maps, and tables.

Fig. S4: Different sources have different amounts of data covering the visualizations

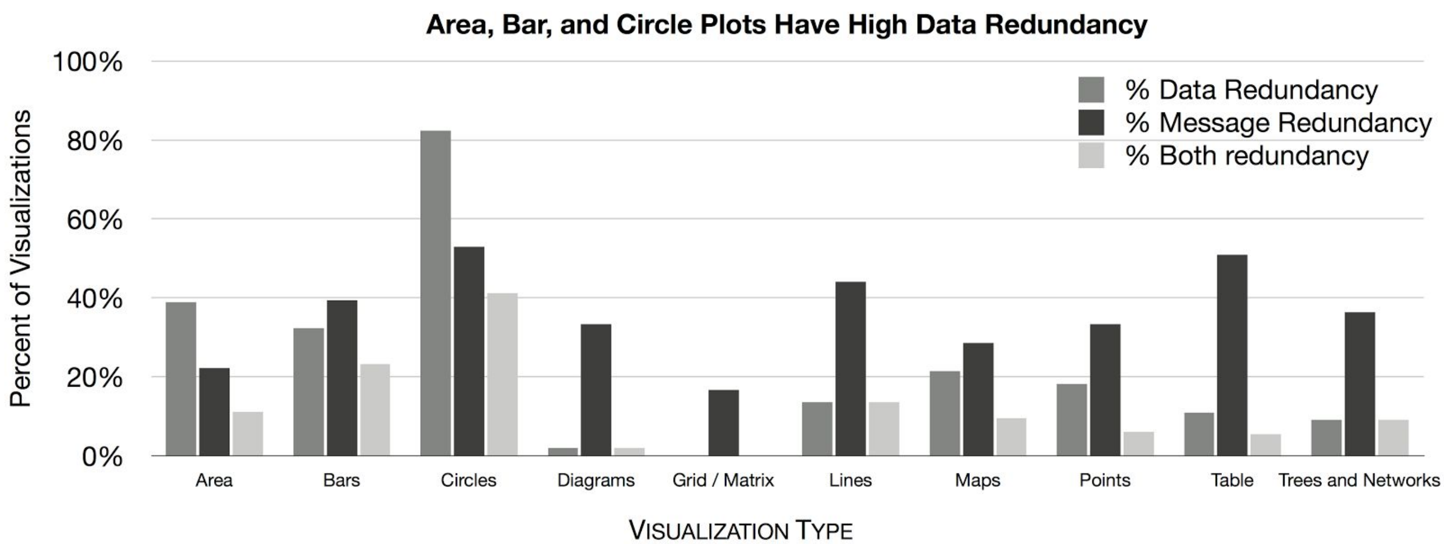


Presence of redundancy across visualizations:

When examining redundancy across visualization types, the message redundancy rates are comparable across all visualization types but highest for circle (53%), table (51%), line (44%), and bar chart (39%) visualizations (see Fig. S5 below). The circle visualizations also have the highest percentage (82%) of data redundancy followed by area (39%) and bar charts (32%). The circle and area visualization types are the only ones that contain a higher percentage of data redundancy than message redundancy visualizations. Overall, infographic visualizations have the most amount of redundancy, and scientific visualizations have the least.

Note that the average percent of the visualization covered with textual elements (e.g., paragraphs of description, annotation, etc.) is significantly less for scientific publications as compared to the other publication venues. This is likely due to the scientific journal context in which the scientific figures occur, accompanied by explanatory text and not expected to be stand-alone. Introducing redundancy into these figures incurs additional space, and thus potential page costs (for journal paper publishing). Thus the context in which visualizations occur (e.g., as part of a paper, a website, a presentation, etc.) is also a relevant factor to take into account when analyzing visualization design. This was beyond the scope of the current paper but offers potential future extensions for study.

Fig. S5: Different visualization types have different amounts of data and message redundancy



Section 5: Experiment Overview

Presented below are additional details about our experimental set-up and design:

Sec. 5.1: Experiment Set-up & Participants

All participants gave written consent to participate in our experiments. The study was run under MIT IRB Protocol #0409000913. A single experiment lasted exactly 1 hour, and participants were compensated $25 for their time.

Each experiment covered ~25% of the target visualizations (98-100 target images), thus each individual could participate in up to four different versions of the experiment (on separate days). Participants who returned to participate would never see the same images as in their previous sessions. On average each participant completed 2 experiment sessions with 9 participants completing all 4 experiment sessions. The selection and permutation of the visualizations were randomized in each case.

Sec. 5.2: Encoding Phase

In this phase, participants examined 100 visualizations for 10 seconds each. After every 10 visualizations, participants were given an opportunity to take a break (e.g., to stretch, rest their eyes, etc.) to reduce and mitigate any possible fatigue. Given the short experimental durations, fatigue was not expected.

To remove any effects of temporal ordering of visualizations (e.g. due to possible fatigue), different participants saw different selections of visualizations in randomly-permuted orderings. Moreover, visualizations seen at this phase were again randomly reordered in the following phase.

Sec. 5.2: Recognition Phase

In this phase, 100 target visualizations (from the previous phase) and 100 additional interspersed filler visualizations were presented in a random permuted order for 2 seconds each with a 0.5 second fixation cross between visualizations. Participants pressed the spacebar anytime they recognized a visualization from the previous experimental phase. Participants could press the spacebar as long as the visualization remained on the screen (before 2 seconds elapsed). During this phase of the experiment, participants could take an optional break every 20 visualizations as needed to mitigate any possible fatigue. Most participants took no breaks across all of experiments.

We collected the fixation locations and durations, and the number of correctly-recognized visualizations (HITs).

Sec. 5.3: Recall Phase

During this phase of the experiment, participants again saw a subset of the target visualizations they correctly recognized in the previous phase. These visualizations were again randomly permuted to remove any temporal effects. Participants' gazes were no longer recorded and participants could sit normally, without a chin-rest. Participants were given 20 minutes in total to write as many descriptions of visualizations as possible. There was no limit to how much time or text length was spent on each visualization, nor any requirement to complete a certain number of visualizations. Each participant worked at his or her own speed and level of detail. Participants were also allowed to skip visualizations for which they could not construct a description.

Sec. 5.5: Performance Metrics

Memorability measures:
In the memorability study presented in [8], targets and fillers were interleaved in sequence. HR and FAR were defined *for targets only*: HR being a correct recognition of a target presented for the second time in the sequence, and FAR being a response on the first presentation of the target. Thus, a dprime measure was used to compute the memorability score by taking into account both HR and FAR for each target image. In the present study, the set-up is different, in that the encoding and recognition experimental phases are separated in time, and participant responses are only collected during the recognition phase. Thus, participants do not have a chance to respond to the first presentation of targets (which occurs during encoding), and thus FAR is not defined for our target images. To make our results comparable to the study in [8], we consider HR instead of dprime as a measure of memorability. Note that HR has more commonly been used to measure memorability [2,11,20] because it is more interpretable than dprime.

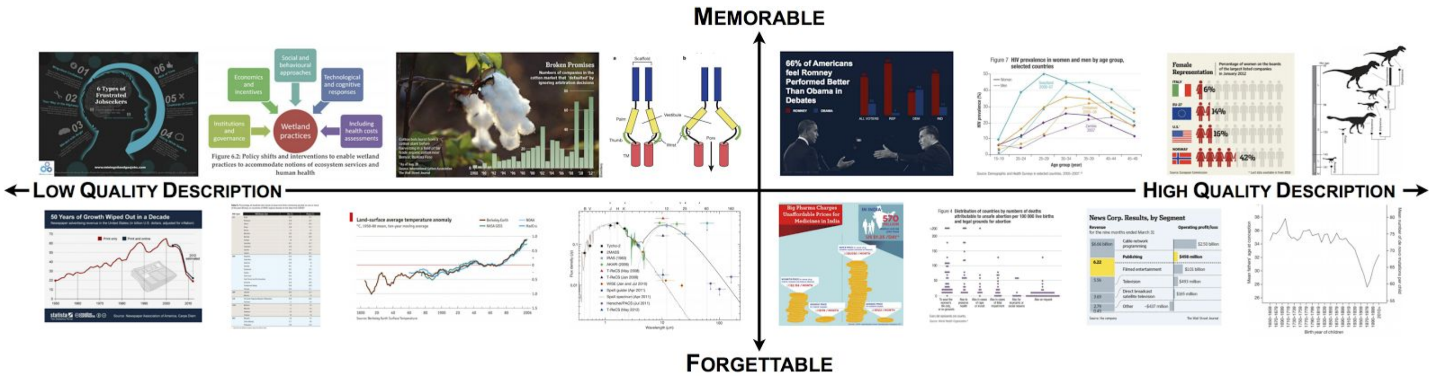
Section 6: Experimental Results and Discussion

Section 6.1.1: Does "at-a-glance" memorability generalize?

The rightmost panel of Figure 6 in the main text is available again in Figure S6 here at a larger resolution. Included are the top and bottom ranked visualizations for each of our four publication source categories across description quality and memorability. The y-axis represents recognition HR, and the x-axis represents average text description quality at recall. In each quadrant, the visualizations from left to right are from Infographic, Government, News Media, and Scientific publication sources.

Note that although this is a sampling of visualizations found at different cross-sections of memorability and description quality, the distribution of visualizations across these quadrants is not uniform. Of the visualizations in the top ⅓ most memorable, 64% are also in the top ⅓ best described, while 19% are in the bottom third most poorly described; of the visualizations in the bottom ⅓ most forgettable, 45% are also in the bottom ⅓ most poorly described, while 24% are in the top ⅓ best described. Thus, memorability and description quality are related (more about this in Section 6.1.4 of the main text).

Fig. S6: Memorability versus description quality

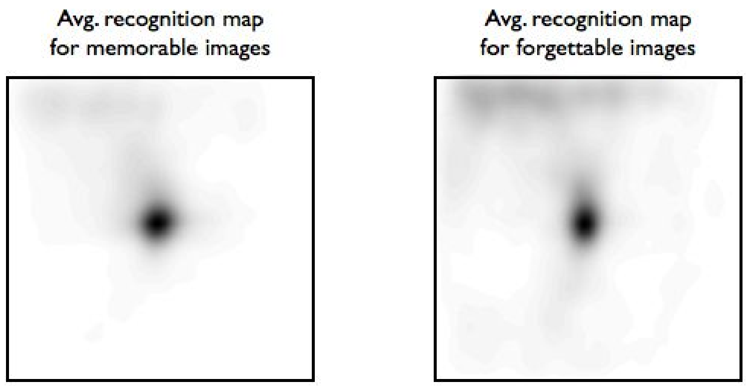


Visualization source links for Fig. S6 plotted visualizations (from left to right, top to bottom):
http://visual.ly/6-types-frustrated-jobseekers
http://www.who.int/water_sanitation_health/publications/2012/review_of_wetlands/en/
http://www.wsj.com/articles/SB10000872396390444772804577623532167565646
http://www.nature.com/nature/journal/v489/n7416/abs/nature11375.html
http://visual.ly/did-romney-outperform-obama-debate-1
http://www.afro.who.int/index.php?option=com_docman&task=doc_download&gid=6495
http://www.wsj.com/articles/SB10001424052702303395604577431832161133916
http://www.nature.com/nature/journal/v484/n7392/full/nature10906.html
http://visual.ly/50-years-growth-wiped-out-decade
http://www.who.int/substance_abuse/publications/global_alcohol_report/en/
http://www.economist.com/blogs/dailychart/2011/10/climate-change
http://www.nature.com/nature/journal/v487/n7405/full/nature11210.html
http://www.msfaccess.org/our-work/hiv-aids/article/1890
http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/
http://www.wsj.com/articles/SB10001424052702303640804577489363802971458
http://www.nature.com/nature/journal/v488/n7412/full/nature11396.html

Section 6.1.2: What are the differences between the most and least recognizable visualizations?

Figure S7 contains the average of the recognition fixation heat maps for the 25 most recognizable visualizations (left) and 25 least recognizable visualizations (right). The central focus on the left indicates quick recognition through visual association, whereas the map on the right probably indicates visual search for semantic associations. The fixations along the top of the heat map for the least recognizable visualizations generally correspond to the location of the title and paragraph text describing the visualization in further detail.

Fig. S7: Average recognition fixation maps for the top 25 most recognizable and bottom 25 least recognizable visualizations: the forgettable images require more visual exploration to recall.



In order to quantify the difference more generally between the fixation patterns for the top 1/3 most recognizable and bottom 1/3 least recognizable visualizations “at-a-glance” [8], we calculated 3 metrics: the average recognition fixation distance viewed away from the image center, the spatial variances in recognition fixation locations, and the number of distinct loci fixated during recognition.

First, we calculated the average distance from the visualization center at which participants fixate. We computed the Euclidean distance between the center of the visualization and each of a participant's recognition fixations on the visualization, and averaged over all of these distances. Participants on average look further away from the center for the least recognizable visualizations (208 pixels) than for the most recognizable visualizations (192 pixels, $t(3189) = 6.29$, $p < 0.001$).

In order to evaluate the variability of the recognition fixations, for each participant we calculated the variances of the x and y fixation locations over all of the recognition fixations. We then took the mean of these two variance values in order to produce an overall variance value. The most recognizable visualizations have a significantly lower fixation variability (20,395-pixel variance) as compared to the least recognizable visualizations (23,424-pixel variance, $t(3189) = 7.17$, $p < 0.001$). Thus participants need to look around less when recognizing the visualizations that are memorable “at-a-glance”.

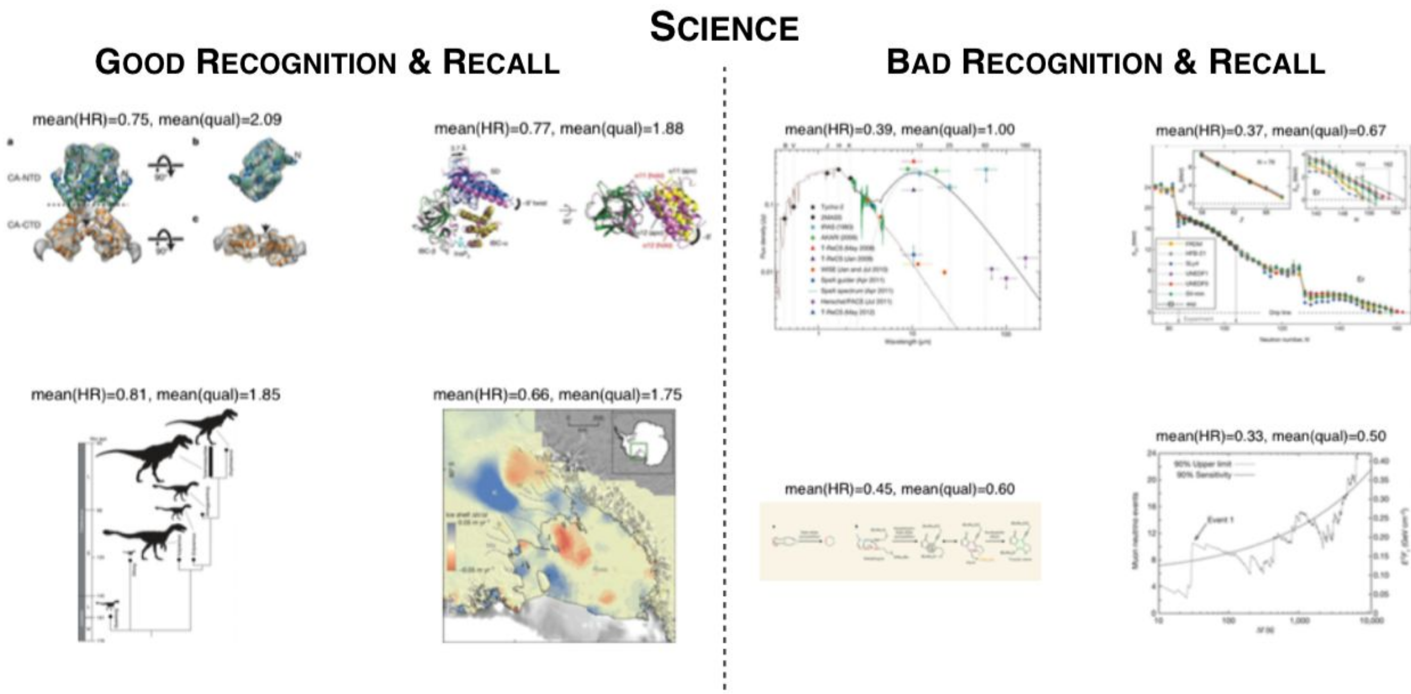
We also compared the average number of distinct foci in the recognition fixations. This is equivalent to looking for discrete regions of a visualization where observers stop and focus their attention when retrieving a visualization from memory. For the 1/3 most recognizable and the 1/3 least recognizable visualizations, we computed recognition fixation maps over all user fixations, applied a threshold of 0.1 and counted the number of connected components (each fixation map was scaled so that the intensity values spanned from 0 to 1). We obtained an average of 2.40 connected components (i.e., distinct foci) in the most recognizable visualization fixations maps and an average of 3.53 connected components in the least recognizable visualizations ($t(240) = 5.38$, $p < 0.001$). Thus the least recognizable visualizations have more foci (i.e. more distinct places that the observer paid attention to) than the most recognizable visualizations, indicative of visual search for an association for recognition.

Over all 3 fixation measures analyzed, there are statistically significant differences between the most and least recognizable visualizations, in that visualizations that are not memorable “at-a-glance” (the least recognizable) require more exploratory fixations before a participant successfully recognizes the visualization.

Section 6.1.4: What do people recall after 10 seconds of encoding?

As a supplement to Fig. 6c in the main text, we provide below quadrants for visualizations with the highest and lowest mean HIT Rates (HR) and the highest and lowest mean description qualities. The visualizations are grouped by source category.

Example Images of Good & Bad Recognition and Recall:

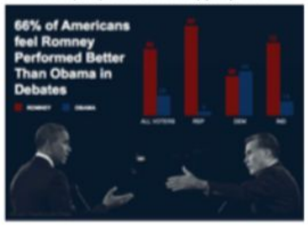


Infographics

Good Recognition & Recall

mean(HR)=0.76, mean(qual)=2.62

mean(HR)=0.86, mean(qual)=2.50



mean(HR)=0.80, mean(qual)=2.43

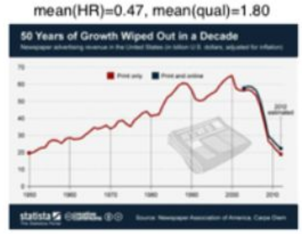
mean(HR)=0.71, mean(qual)=2.38

Bad Recognition & Recall

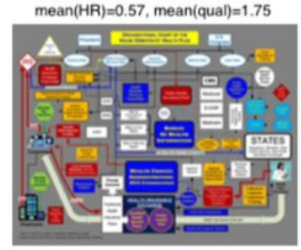
mean(HR)=0.56, mean(qual)=1.88

mean(HR)=0.52, mean(qual)=1.82

mean(HR)=0.47, mean(qual)=1.80



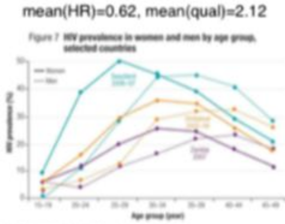
mean(HR)=0.57, mean(qual)=1.75



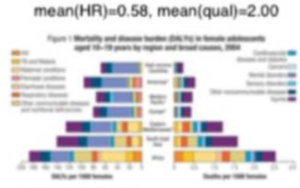
Government

Good Recognition & Recall

mean(HR)=0.62, mean(qual)=2.12

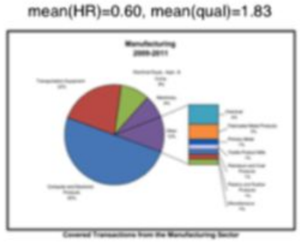


mean(HR)=0.58, mean(qual)=2.00



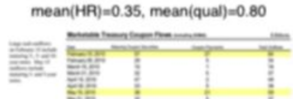
mean(HR)=0.58, mean(qual)=1.86

mean(HR)=0.60, mean(qual)=1.83

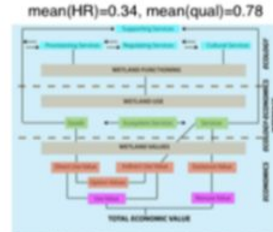


Bad Recognition & Recall

mean(HR)=0.35, mean(qual)=0.80

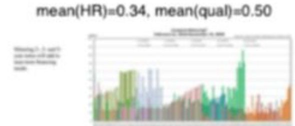


mean(HR)=0.34, mean(qual)=0.78

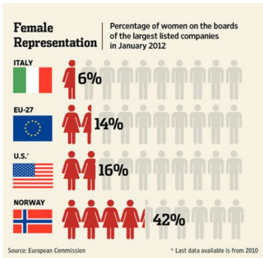


mean(HR)=0.28, mean(qual)=0.67

mean(HR)=0.34, mean(qual)=0.50



To demonstrate the amount of information a study participant was able to encode in 10 seconds and then recall after 20 minutes, we provide the sample visualizations and participant-generated recall text descriptions below. The instructional prompt was “Describe the visualization in as much detail as possible:”



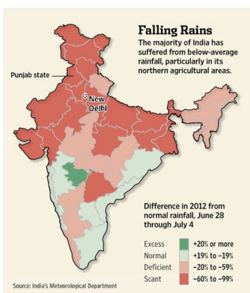
Title: “Female Representation”
Text: “Percentage of women on the boards of the largest listed companies in 2012.”
Source:http://www.wsj.com/articles/SB10001424052702303395604577431832161133916

“Percent of women in government of various countries. Norway led with a bit under 50%”

“The number of women vs. men executives compared for different countries. Norway is the bottom row.”

“Percentage of female executives on the board by country. Highest was norway with 42%”

“Number of women on executive boards of various countries. Norway is represented in the 40%'s the US is next the EU countries are averaged next and Italy has an extremely low representation.”



Title: “Falling Rain”
Text: “The majority of India has suffered from below-average rainfall, particularly in its northern agricultural areas.”
Source: http://www.wsj.com/ (original article no longer available)

“Showing the droughts in India. Only one region had excess amounts of rain compared to previous years but most of the north had scant amounts”

“a map of india that was about rainfall. the red parts mean they didn't get as much rainfall as they usually do and the green parts are where they got more.”

“Northern India isn't getting as much rain as it used to. The one little green bit is getting extra rain.”

“Drought figures for India. Looks like it's agricultural areas (the mostly the north) suffered the most.”

Title: “Foreign Bookings”
Text: “Change from a year ago in bookings of U.S. hotels by Europeans.”
Source: http://www.wsj.com/articles/SB10001424052702303379204577476743470788850

“Change in number of hotel bookings in the U.S. by people of different European nationalities. Belgium and France were on top (book more) and Greece was at the bottom.”

“Foreign Bookings : showing bookings of u.s. hotels by country Belgium was the highest”

“‘Foreign bookings’ a comparison of how bookings of US hotels by Europeans has changed in the past [time period]. Greeks don't (probably can't afford to) go abroad lately but Belgians are starting to.

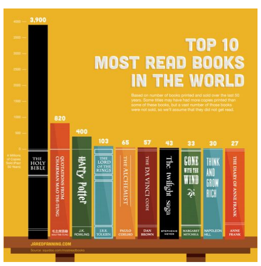
“change in european hotel bookings in the us in one year greece dropped by 41 %”

Section 6.2.1: Titles

Across the 330 visualizations with at least 3 descriptions, the average description quality of visualizations with titles is 1.90 as compared to 1.30 for visualizations without titles (the difference in quality is statistically significant at *$p<0.001$*). These statistics are computed over 2,085 descriptions that were generated for visualizations with titles, and 581 descriptions that were generated for visualizations without titles. The difference between the number of descriptions generated in each case is informative as well. Based on this data, participants were more than 3.5 times more likely to write a description for a visualization that had a title than for one that didn't, which points to the ease (and correspondingly, difficulty) of encoding and retrieving these visualizations from memory.

However, not all titles are equally effective. Below we provide examples of visualizations that are at the high/low extremes of recognition HR, and quality ratings. The examples are split by "good" (i.e., had a large fraction of participants mention or use the title in their recall text description and received high quality scores) and "bad" (i.e., unutilized titles and received low quality scores):

GOOD TITLE EXAMPLES:



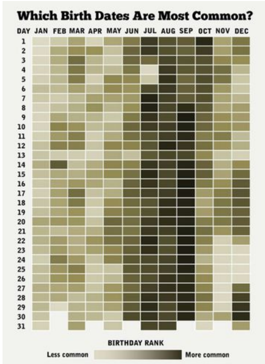
Title: "Top 10 Most Read Books in the World"
Text: "Based on number of books printed and sold over the last 50 years. Some titles may have had more copies printed than some of these books, but a vast number of those books were not sold, so we'll assume that they did not get read."
Source: http://visual.ly/top-10-most-read-books-world

"The most read books globally. I think it went Holy Bilbe something Harry Potter Lord of the Rings The Alchemist .... and The Diary of Anne Frank was last"

"Most popular/ widely read books for example the Holy Bible has the most worldwide readers followed by Harry Potter Lord of the Rings Twilight and other books (not necessarily in that order)"

"Most read books. Holy bible. Quotes of Mao Ze Tung. Harry Potter."

"The most read books in the world. The bible is the first. The Lord of the Rings the Twighlight Saga the Davinci Code are also represented."



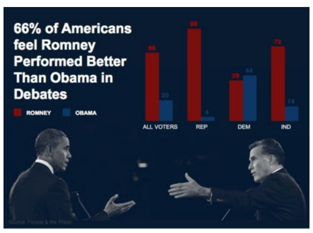
Title: “Which Birth Dates are Most Common?”
Source: http://thedailyviz.com/2012/05/12/how-common-is-your-birthday/ via http://visual.ly/

“Percent of people born on each day of the year. X-axis is month Y-axis is day. Most popular birthdays are in late summer and early fall.”

“This was about which birthdays were most common with darker shades meaning more people being born on that day. September seemed to be the most popular birth month and the end of December had a high number of birthdays.”

“months in which people were born...there's a concentration in the late summer seemingly the highest in September”

“this was a chart of most common birthdays. the darker the color the more common the birthday. september was the darkest month”



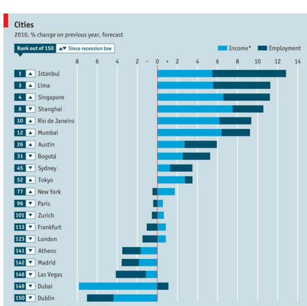
Title: “66% of Americans feel Romney Performed Better Than Obama in Debates”
Source: http://visual.ly/did-romney-outperform-obama-debate-1

“the percentages of people who thought Romney debated better than Obama. (~60% of everyone (~80% Republicans 60% independent and ~35% democrats) At the bottom is a picture of the two gesturing or about to shake hands”

“Percentage of people of different political parties who picked either Romney (red) or Obama (blue) as being better at debate. Romney one in all political party categories except in the Democrats”

“66% of Americans believe Romney did better in the debate is what the quote on the left top reads. Democrats believed Obama did better with a small majority. Other groups were also considered around 4 total. Fox News viewers may have been one category. Obama pic on left lower Romney on right”

BAD TITLE EXAMPLES:



Title: “Cities”
Text: “2010, % change on previous year, forecast”
Source: http://www.economist.com/blogs/dailychart/2010/11/global_cities

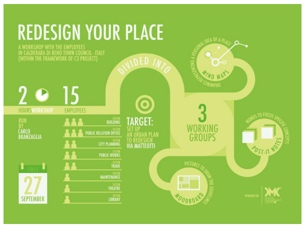
“Countries list employment and GDP?”

“ranking of cities by growth in employment? istanbul was at the top”

“Some countries have gotten better on this metric some worse. More better than worse though.”

“Increase in employment rate and GDP(?) of different cities internationally.”

“Something to do with the economies of developing nations and useable income vs taxed income”



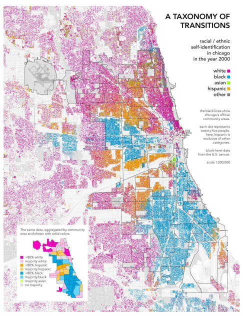
Title: “Redesign Your Place”
Text: “A workshop with the employees in Calderara di Reno town council - Italy (within the framework of C3 project)”
Source: http://visual.ly/redesign-your-place

“Update your place or something like that.a A humanoid figure with tentacles for limbs advertising a product.”

“The square says that something is divided into three groups.”

“something about the italian government city council”

“Somehting about a target perhaps about more effectively running a meeting in Italy or something like that”



Title: “A Taxonomy of Transition”
Text: “racial/ethnic self-identification in Chicago in the year 2010”
Source: http://www.radicalcartography.net/index.html?chicagodots via http://visual.ly/

“The colors are ethnicities and the graph is a map of some area showing who lives in what neighborhood.”

“Hispanic is orange and exclusive of other groups.”

“chicago ethnicity identification population segregation by races”

“florida map of location of races”

“Chicago race/ ethnicity identity map pink is white orange is hispanic blue is black”

“segregation in chicago”

“Chicago population by neighborhood by race”

In Table S1 we include some statistics about where a title was most commonly located across all our publication source categories. In Table S2 we calculate how frequently titles were fixated (across all participants and all visualization in a specific category) and how often they were described depending on where the title was located. In Table S3 we include the average title length across all visualizations in a given source category.

Table S1: Title Placement: most visualizations had the title located at the top of the visualization. Government visualizations had the most number of visualizations with a title at the bottom.

	Title on Top	**Title on Bottom**
Infographics	85.9%	4.4%
News	96.7%	1.6%
Government	82.0%	16.0%
Science	2.5%	0.0%

Table S2: Fixation of title and description of title conditioned on title location. Across all sources, titles were more likely to be fixated and described when found on the top of the visualization than elsewhere.

	Proportion of times title was fixated (at encoding)				Proportion of times title was described (at recall)			
	In total	when on top	when on bottom	when elsewhere	In total	when on top	when on bottom	when elsewhere
Overall	59%	76%	63%	7%	46%	57%	55%	10%
Infographics	80%	83%	79%	57%	72%	74%	71%	57%
Government	81%	85%	62%	41%	55%	59%	46%	17%
News	66%	66%	46%	46%	43%	43%	58%	46%

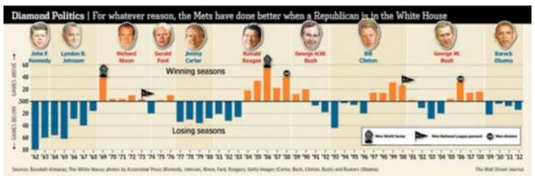
Table S3: Mean title lengths: government visualizations tended to have the longest titles.
(significant pairwise differences between sources at p = 0.001 level

	Title length (averaged over all visualizations)
Infographics	30.0 (SD: 16.9)
News	21.2 (SD: 13.2)
Government	87.6 (SD: 42.7)
Science	1.3 (SD: 8.6)

Section 6.2.2: Pictograms

Pictograms, i.e., human recognizable objects, did not seem to distract participants during encoding. In fact, averaged over all visualization elements, the total fixation time spent on pictograms was less than all the other visualization elements. Visualizations with pictograms also tended to have better quality descriptions written for them during recall (see Table S4). However, not all pictograms add to the effectiveness of a visualization. Below we present visualizations with the highest description quality ratings that utilize pictograms as part of message redundancy ("**good use**"), and visualizations with the lowest description quality ratings that do not utilize pictograms as part of message redundancy ("**bad use**").

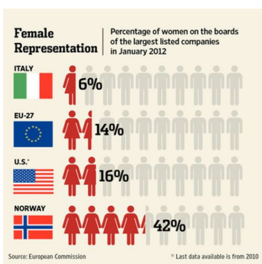
Good Use of Pictograms



Title: "Diamond Politics"
Text: "For whatever reason, the Mets have done better when a Republican is in the White House."
Source: http://www.wsj.com/ (original article no longer available)

Pictographic elements:

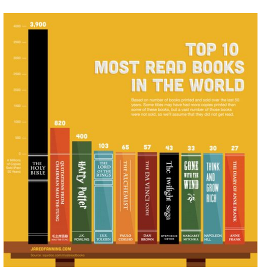
- *Photographic portraits for additional face recognition of each president. (Message redundancy)*
- *Icons to denote championships.*



Title: "Female Representation"
Text: "Percentage of women on the boards of the largest listed companies in 2012."
Source:
http://www.wsj.com/articles/SB10001424052702303395604577431832161133916

Pictographic elements:

- *Flags of each country listed. (Message redundancy)*
- *Pictorial bar chart icons help to reinforce that the percentages are of women on the company boards.*

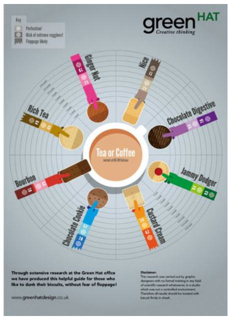


Title: "Top 10 Most Read Books in the World"
Text: "Based on number of books printed and sold over the last 50 years. Some titles may have had more copies printed than some of these books, but a vast number of those books were not sold, so we'll assume that they did not get read."
Source: http://visual.ly/top-10-most-read-books-world

Pictographic elements:

- *Each bar is represented as the actual spine of the book it encodes.*

Bad Use of Pictograms



Title: “Tea or Coffee”
Text: “Through extensive research at the Green Hat office we have produced this helpful guide for those who like to dunk their biscuits, without fear of floppage!”
Source: http://visual.ly/tea-biscuit-guide

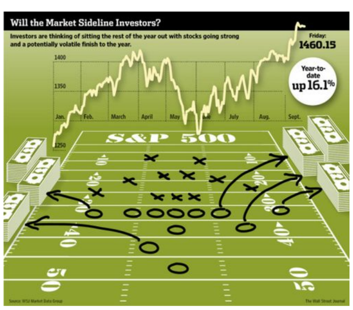
Pictographic elements:

- *Although cookie cartoons are illustrations of the cookies discussed, the hands holding them are stacked bars with no data encoding purpose to the pictograms nor the bar length/stack.*

Title: “The QR Invasion”
Source: http://visual.ly/qr-quick-response-invasion

Pictographic elements:

- *QR codes represent the topic at hand (QR code history).*
- *Godzilla is representative of the country of Japan where the QR code was invented, but serves no message or data redundancy role.*
- *The company logos are representative of companies that use QR codes, but they do not relate to the historical facts in the table.*



Title: “Will the Market Sideline Investors?”
Text: “Investors are thinking of sitting the rest of the year out with stocks going strong and a potentially volatile finish to the year.”
Source: http://www.wsj.com/ (original article no longer available)

Pictographic elements:

- *Football strategy play diagram is representative of “sidelining” players, in this case piles of cash from investment. (The metaphor is weak and easily lost in the visualization composition.)*

Table S4. The effect of pictograms on the description quality of visualizations by source (** = p < 0.001, t-statistic corresponds to comparison of visualizations with and without pictograms). Across all sources, visualizations with pictograms have similar or better quality descriptions than visualizations without pictograms.

	Mean description quality (0-3 scale)		% with pictograms	
	With pictogram	Without pictogram	In top 1/3 images (with good descriptions)	In bottom 1/3 images (bad descriptions)
Overall	2.01**	1.50**	20%	2%
Infographics	2.09	2.07	39%	39%
News	2.10**	1.84**	17%	3%
Government	1.46	1.36	0%	0%
Science	1.52**	1.12**	4%	0%

Section 6.2.3: Other Elements

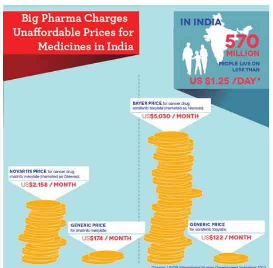
Across all visualizations, the elements that were fixated the most often (refixated) were the legend, table header row (often acting as a title), and title. The elements that were fixated the longest were the paragraph, legend, and header row. See pages 25-29 of this document.

For the government visualizations, the header row is refixated the most and the longest. For infographics, other than pictograms, title followed by legend are refixated the most and the longest. For news media, the header row and legend received the most refixations, while the legend and paragraph were fixated the longest in total. The header row and legend were also refixated the most and fixated the longest for the science visualizations. During encoding across all visualizations, the element refixated the most other than the data itself is the legend.

Section 6.3: The Importance of Redundancy

The average description quality is higher for visualizations that contain message redundancy (1.99) than for visualizations that do not (1.59, *p<0.001*). Similarly, visualizations that contain data redundancy also have better quality descriptions (2.01) than those that do not (1.70, *p<0.001*). Below we present examples of data redundancy from our experiment. In the groupings below, "good" means a high average quality description.

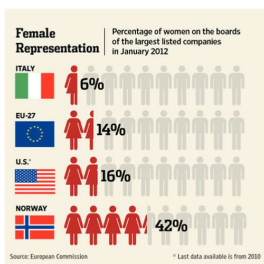
Good *Data & Message* Redundancy Examples



Title: "Big Pharma Charges Unaffordable Prices for Medicines in India"
Source: http://www.msfaccess.org/our-work/hiv-aids/article/1890 via http://visual.ly

Data redundancy:
- *stack heights are annotated with numerical dollar amounts*

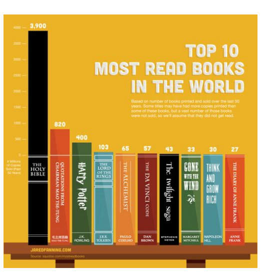
Message redundancy:
- *coin stacks convey a financial quantity being measured*
- *map of India, the topic of the visualization, is included*



Title: "Female Representation"
Text: "Percentage of women on the boards of the largest listed companies in 2012."
Source:
http://www.wsj.com/articles/SB10001424052702303395604577431832161133916

Data redundancy:
- *horizontal bars are annotated with numerical percentages*

Message redundancy:
- *Country names are provided along with corresponding country flags*
- *Female icons (and their proportion relative to male icons) depicts the topic of visualization*

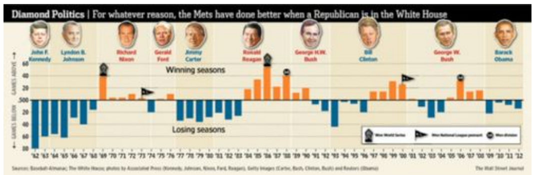


Title: "Top 10 Most Read Books in the World"
Text: "Based on number of books printed and sold over the last 50 years. Some titles may have had more copies printed than some of these books, but a vast number of those books were not sold, so we'll assume that they did not get read."
Source: http://visual.ly/top-10-most-read-books-world

Data redundancy:
- *numerical values are annotated at the top of the bars*

Message redundancy:
- *the graph bars are visually composed out of the spines of the books plotted*

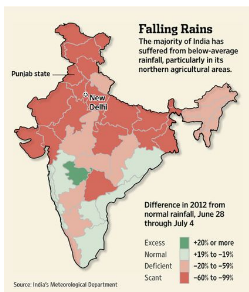
Good *Message-Only* Redundancy Examples



Title: “Diamond Politics”
Text: “For whatever reason, the Mets have done better when a Republican is in the White House.”
Source: http://www.wsj.com/ (original article no longer available)

Message redundancy:

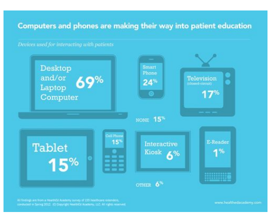
- *the paragraph text conveys the trend that the data is supposed to convey*
- *politicians’ names are associated with photographs of their faces*
- *the colors of the bars correspond to the political parties depicted*



Title: “Falling Rain”
Text: “The majority of India has suffered from below-average rainfall, particularly in its northern agricultural areas.”
Source: http://www.wsj.com/ (original article no longer available)

Message redundancy:

- *the paragraph text conveys the trend that the data is supposed to convey*



Title: “Computers and phones are making their way into patient education”
Text: “Devices used for interaction with patients”
Source: http://visual.ly/computers-and-phones-patient-education

Message redundancy:

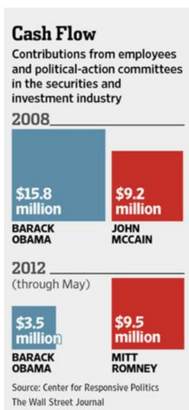
- *icons of technological devices are presented along with their names*

Good *Data-Only* Redundancy Examples

Title: “Foreign Bookings”
Text: “Change from a year ago in bookings of U.S. hotels by Europeans.”
Source:
http://www.wsj.com/articles/SB100014240527023033792045774767434707888850

Data redundancy:

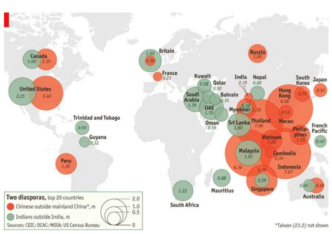
- *numerical annotations alongside bars*



Title: “Cash Flow”
Text: “Contributions from employees and political-action committees in the securities and investment industry”
Source:
http://www.wsj.com/articles/SB100008723963904440979045775352011359121944

Data redundancy:

- *scaling of bars corresponds to the numerical dollar amounts communicated*



Title: “Two diasporas”
Text: “top 20 countries”
Source:
http://www.economist.com/blogs/dailychart/2011/11/diasporas

Data redundancy:

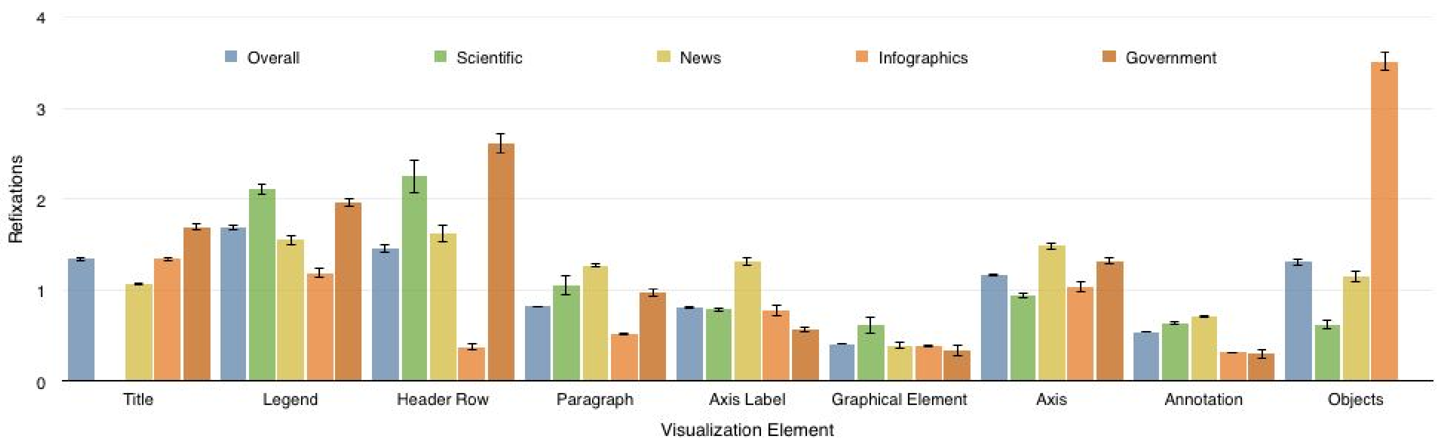
- *scaling of plotted markers corresponds to their numerical values*

Section 6 (cont.):

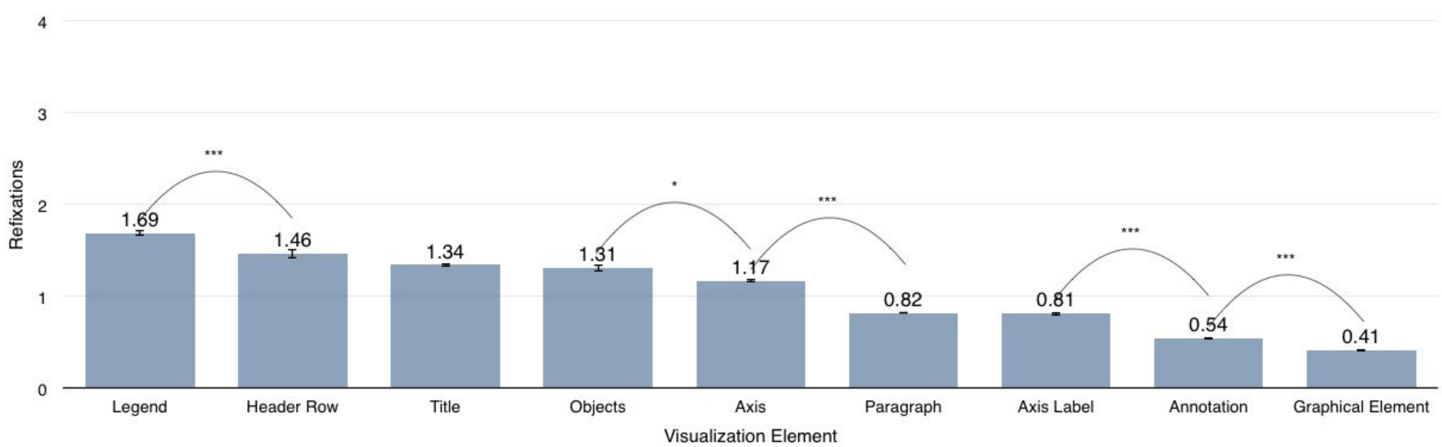
We computed 2 fixation measures by intersecting fixation locations on a visualization with the labeled visual elements to determine when fixations landed within each element: refixations and TFT (total fixation time, measured in ms), defined with the corresponding plots below. Note that a single fixation can land on several elements at once (e.g., an annotation on a graph). In this case, we count the fixation as belonging to all of those elements. We collect all of a viewer's fixations during the encoding phase (10 s), and we discard as noise fixations lasting less than 150 milliseconds. All fixation measures are averaged across viewers and different sets of visualizations, and compared using Bonferonni-corrected t-tests. We annotate some of the main pairwise significance values on our plots below with the following convention: (***) if the comparison was statistically significant at the $p<0.001$ level, (**) if at the $p<0.01$ level, and (*) if at the $p<0.05$ level.

Refixations: The number of times a viewer returns to an element during the entire viewing period (including the first time the element is fixated). Consecutive fixations on the same element are not counted.

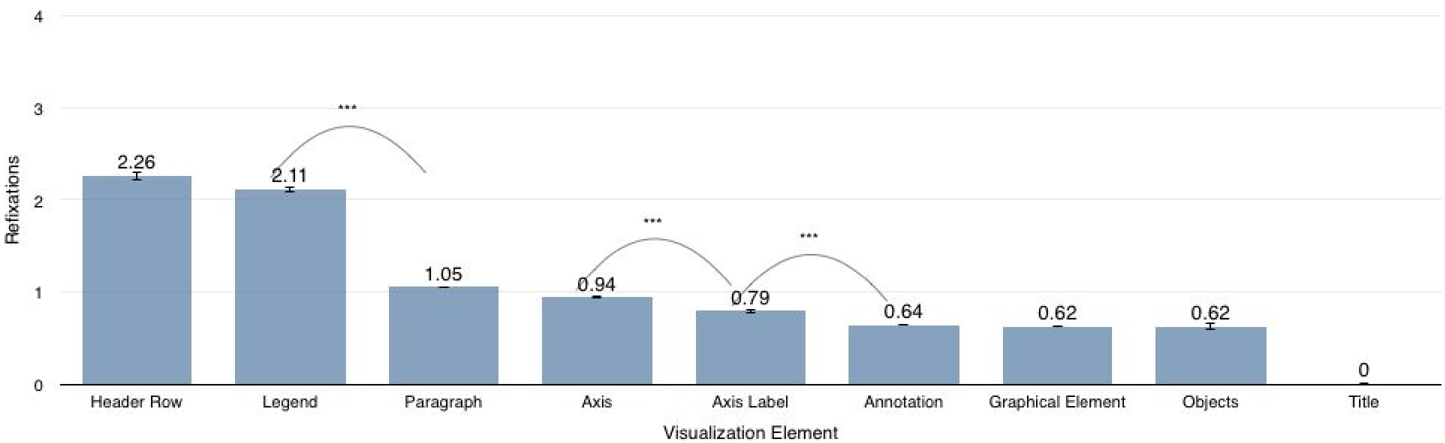
Refixations



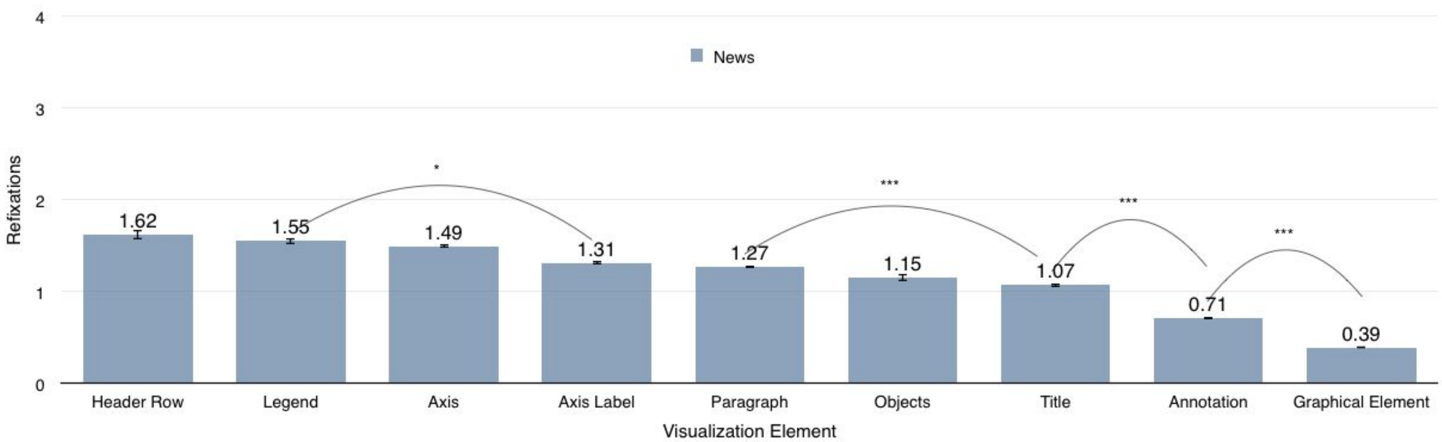
Overall



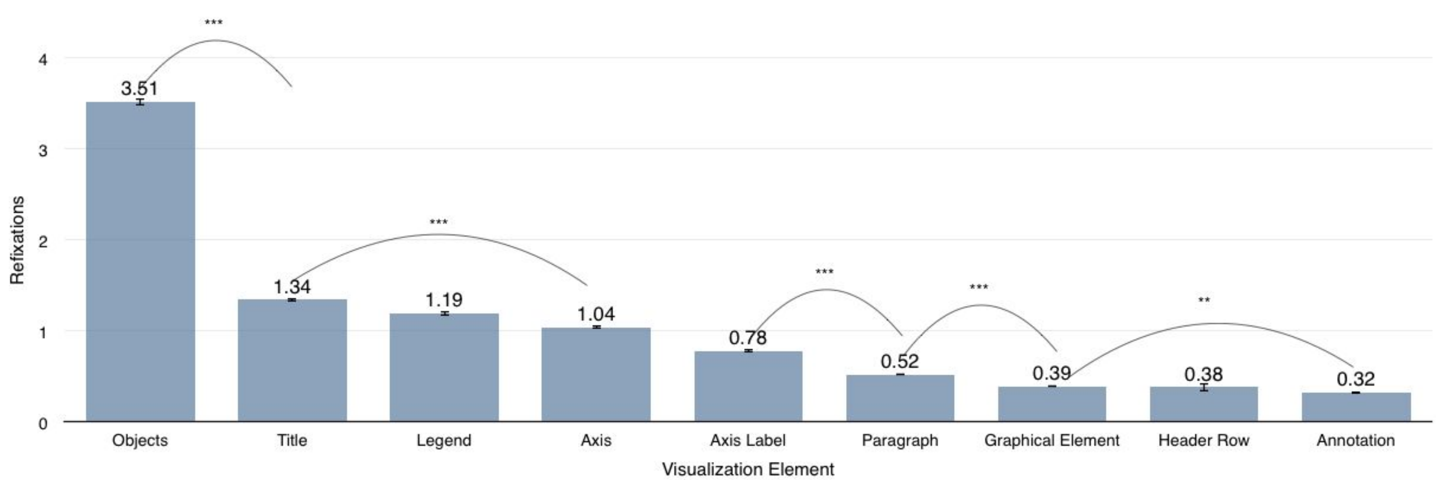
Science



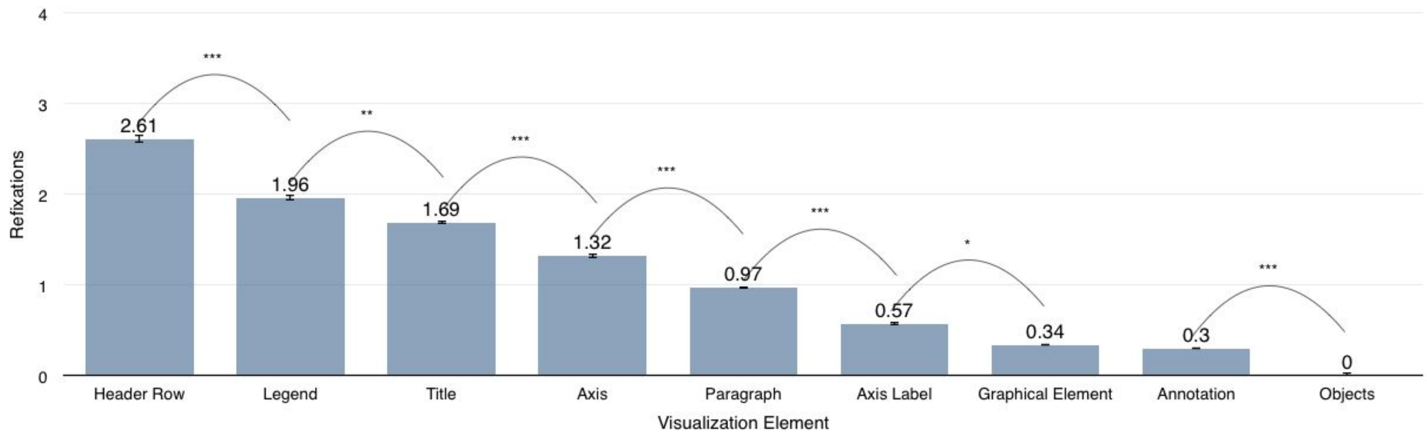
News Media



Infographics

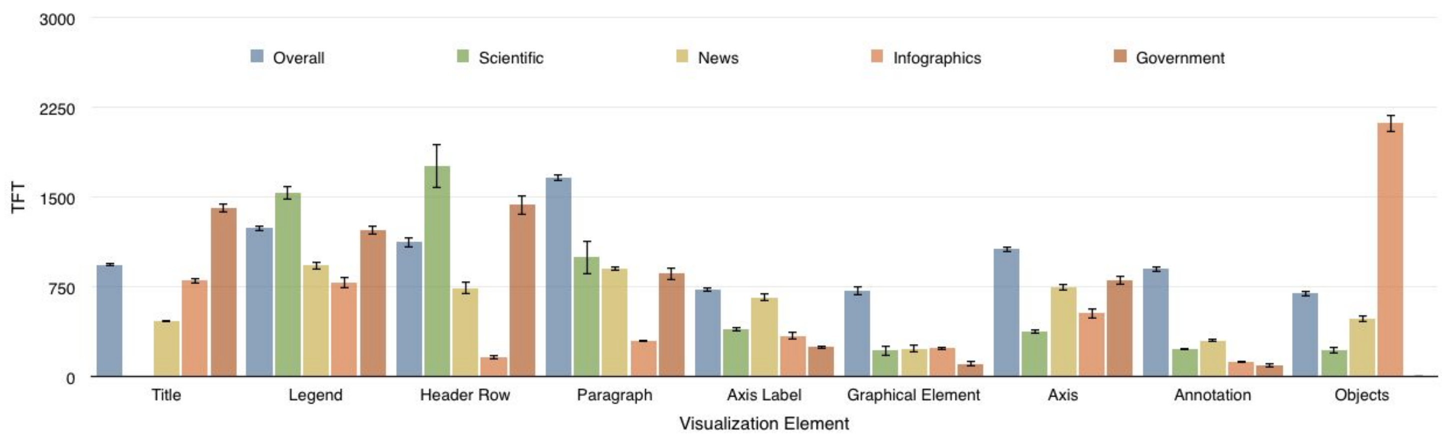


Government

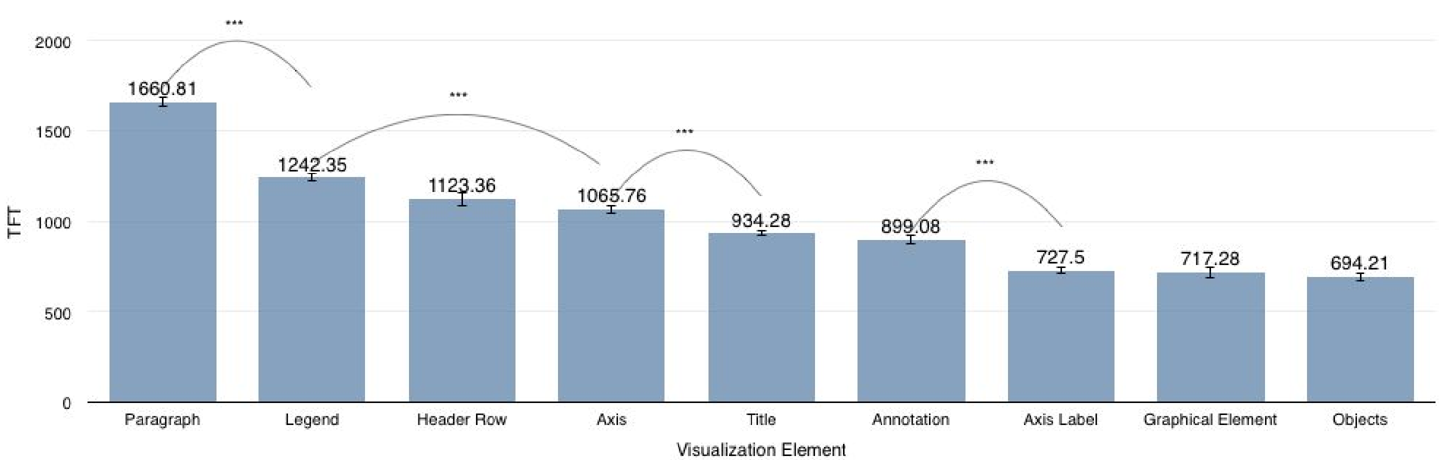


Total fixation time (TFT): Total duration of a viewer's fixations landing on a given visual element throughout the entire viewing period.

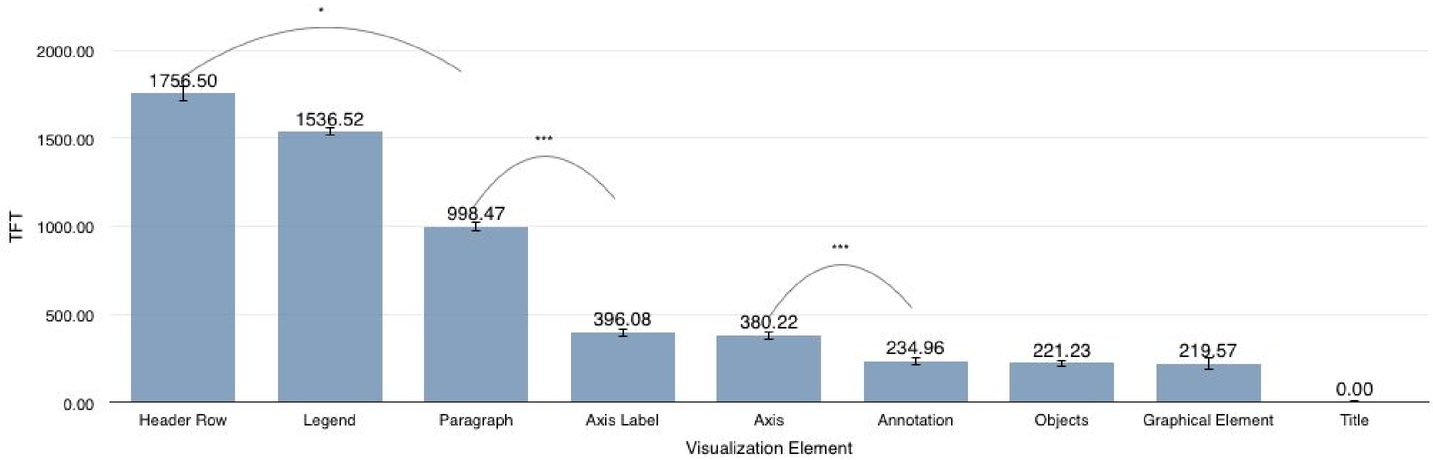
TFT



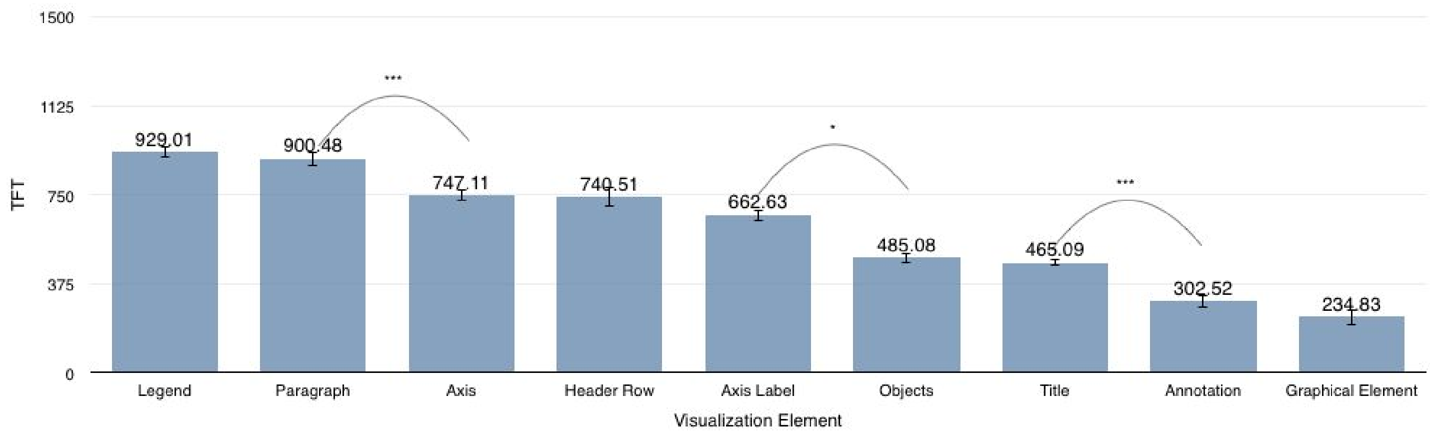
Overall



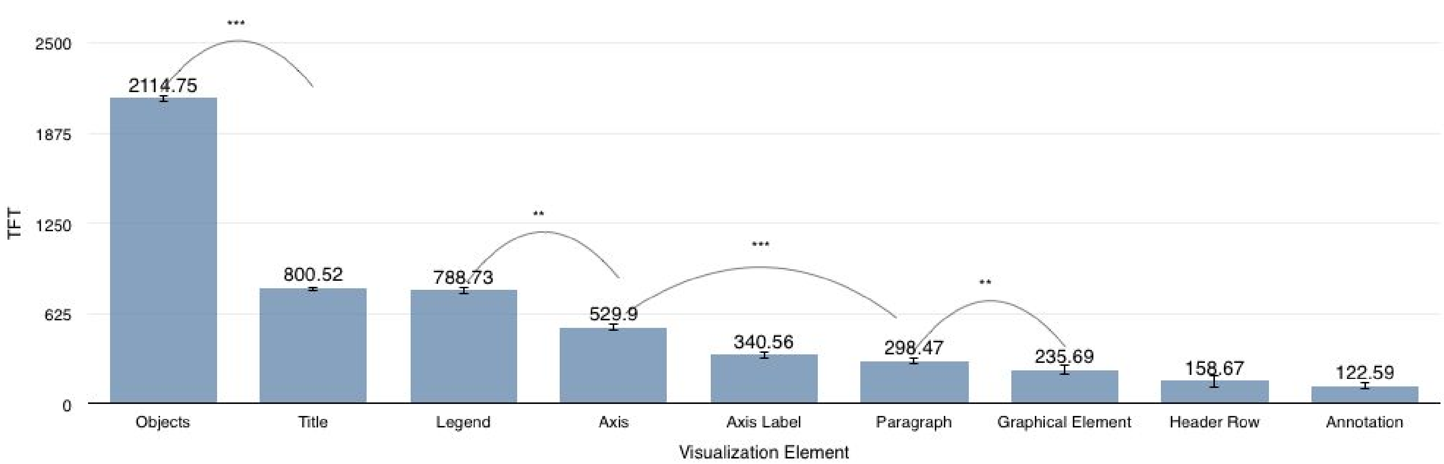
Science



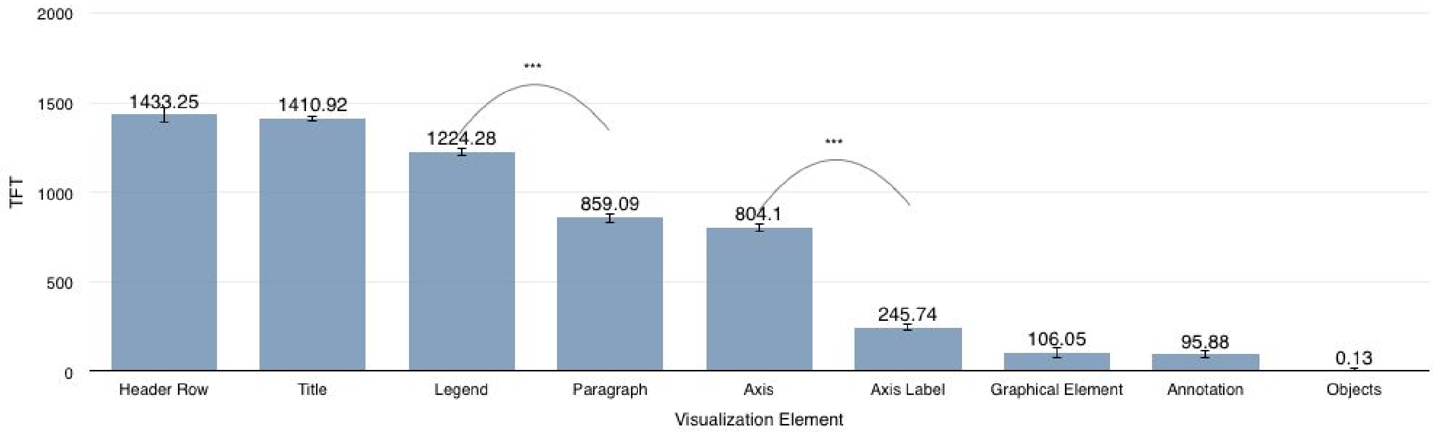
News Media



Infographics



Government


2000
1500
1000
500
0
TFT
1433.25
1410.92
1224.28
859.09
804.1
245.74
106.05
95.88
0.13
***
***
Header Row
Title
Legend
Paragraph
Axis
Axis Label
Graphical Element
Annotation
Objects
Visualization Element