HIV/AIDS through stories

Julie Fortuna  
Stanford University  
jfortuna@stanford.edu

Stephanie Nicholson  
Stanford University  
snicholl1@stanford.edu

ABSTRACT
The Internet contains many HIV/AIDS data visualizations, and many sites where users can post personal stories about HIV/AIDS, but features no resources that display stories in the context of data. In an effort to increase HIV/AIDS awareness and decrease stigma, this visualization provides an integrated view of HIV/AIDS data and stories from the individuals behind the numbers.

Author Keywords
HIV/AIDS, storytelling, narratives.

ACM Classification Keywords
H5.m. Information interfaces and presentation: Miscellaneous.

INTRODUCTION
HIV/AIDS is a worldwide epidemic. As of 2011, 34 million people were living with the disease [1]. The most severely affected region is sub-Saharan Africa, where nearly 1 in 20 people have HIV/AIDS [2]. Many interventions are underway, but HIV/AIDS remains a leading global health crisis. As of 2009, there were approximately 1.1 million cases in the United States [3].

The World Wide Web has a great deal of content about HIV/AIDS. First, many sites attempt to visualize details of the epidemic. These sites mainly come from government sources. For example, the CDC has an interactive map that allows users to select parameters to filter data and view trends on a map [4] (Figure 1). But, the query feature is difficult to find. UNAIDS, the Joint United Nations Programme on HIV/AIDS, also has a map-based visualization of world HIV/AIDS data on its website [5]. The icons used for querying can be confusing to a new user.

Second, there are a number of sites that allow people to share their personal stories about HIV/AIDS with others [6, 7]. These websites vary slightly in organization, but all share a common feature: it’s difficult to filter stories by specific attributes. For example, if a user wanted to look for stories about a white female in her 30s, she would have to read a number of story titles, or even skim the stories themselves, to find stories that matched her search query. Further, stories are not at all integrated with data to present information about HIV/AIDS in context. Rather, if users want to find out about HIV/AIDS stories as well as statistics, they will have to visit different websites. This visualization attempts to solve that problem, through storytelling techniques. By integrating data to tell a story about HIV/AIDS then letting a user explore stories through a graph-based interface, we hope to present a more holistic picture of the epidemic that decreases stigma and increases awareness.

RELATED WORK
Related work can be divided into two different groups: apps that blend stories and data and publications that contained similar visualization methods.

Al Gore’s “Our Choice” app [8] takes on the issue of climate change. People using the app hear Al Gore’s voice telling a story about climate change. Photos, videos, and interactive graphics accompany his words. This format blends storytelling with facts, a goal this project attempts to achieve with HIV/AIDS.

TeachAIDS [9] is a software package that provides “medically-accurate, pedagogically-grounded, and culturally-tailored animated software to optimize learning and retention” of information about HIV/AIDS. There are currently over 50 country-specific versions of the software. This application is similar to our solution in that it attempts to combine HIV/AIDS and interactive computer applications, but more focused on intervention, while our solution is centered on providing information.

Other visualizations in the New York Times informed the design of visualizations for the current project. Mike Bostock’s visualization “How the tax burden has changed” [10], published on November 29, 2012, is comprised of a series of assertions (e.g. “What’s driven the changes? Federal income tax rates have declined . . .”) followed by an interactive visualization to support each assertion. All visualizations are line graphs, and also contain a supporting caption. This method of stepping through a story line by line, and providing a visualization to support each assertion, is adopted in the current project.

A second New York Times visualization inspired the filter view of this project. “The Jobless Rate for People Like You” [11], published in the New York Times on November 6, 2009 allows exploration of how unemployment has affected different demographic groups. People can choose the demographic groups that they fit into best, and see how their demographics’ unemployment rate compares to others, as well as the national unemployment rate. We wanted to include this personalization aspect in our project, so people would feel like HIV/AIDS was more relevant to them.
METHODS
We first outline the design decisions for each stage of the application. Then, we report initial user feedback from pilot studies with the application.

Data Collection
HIV/AIDS statistics were collected from the United States Centers for Disease Control, UNAIDS, AIDS.gov, and the CIA World Factbook [1, 3, 4, 5]. Population statistics were collected from the US Census Bureau’s website. Stories are scraped from publicly available compilations of HIV/AIDS stories on various websites [6, 7].

Design
Design decisions were specific to each of the three views: story, filter, and prevention.

Story view
The story view was designed to give people a brief background on the status of the HIV/AIDS epidemic in the United States and pique their curiosity in the issue. The story view has six panels.

The first panel provides an introduction to the HIV/AIDS epidemic on the world stage. The header text for this segment reads, “HIV/AIDS is a worldwide problem.” To demonstrate this, people can hover over the maps of four countries (including the US) to see the HIV/AIDS prevalence rate each one.

To focus in on the epidemic in the United States, the second panel reads “… and the United States is certainly not exempt. This panel features a simple donut chart displaying the proportions of HIV+/HIV- individuals in the country. Since viewers often struggle to estimate size in charts like these, facts are also provided on the right side of the panel.

Now that the user has been grounded in the proportion of Americans affected by HIV/AIDS, the app demonstrates how HIV/AIDS can affect people in many different demographics, including age, race/ethnicity, location, and sexual orientation. There are many misconceptions about the populations affected by HIV/AIDS, and this portion of the visualization is meant to combat those stereotypes.

Age is visualized first (Figure 2). In the default view, people see a bubble chart proportioned by number of HIV/AIDS cases per age group. People can toggle to see bubbles proportioned by the percentage of the population falling into each age group. A bubble chart with resizable bubbles was chosen to draw attention to the relationship between portion of population and portion of cases for each age group—for some groups, the two proportions are very similar, and for others, they are very different. One drawback of a bubble-style chart is that people struggle to accurately estimate the relative area of objects [13]. Further, since the bubbles are proportional, it is not clear how many people each circle actually represents. These drawbacks are countered through the use of tooltips: when a user hovers over a bubble, he sees a popup showing the age of the group and the number of people falling into this category.

Figure 1: the CDC’s interactive map visualization of the HIV/AIDS prevalence in the United States

Figure 2: panel 3 in the story view, showing HIV/AIDS cases by age group

Figure 3: panel displaying data on sexual orientation and HIV/AIDS

Figure 4: panel displaying data on ethnicity and HIV/AIDS
Next, people see a panel about sexual orientation (Figure 3). In this visualization, instead of toggling between two views of the same graph, they toggle between two different versions of the same pie chart: one for males and one for females. The pie chart shows the proportion of new cases in 2010 from same-sex and heterosexual contact for males and females. The radius of the chart is proportional to the number of new cases. In initial versions of the chart, both the “male” and “female” charts were the same size. However, this gave pilot users the false impression that the same number of men and women contracted HIV/AIDS in 2010 (4 times as many men contracted HIV/AIDS in 2010). Thus, when the user selects “females,” the pie chart shrinks proportionally.

The third point made regards racial/ethnic inequality with regards to HIV/AIDS (Figure 4). Once again, a bubble chart is employed and users can toggle between bubbles proportioned by number of cases per group (the default) and the population size of each group. Putting this graph in the third position was a conscious decision—putting two bubble charts in a row might increase the odds that users confuse the two (thus gleaning the wrong impression), and we want to start and end with different types of visualizations to give the user the impression that they have learned many different things from our story.

After exploring ages, sexual orientations, and races/ethnicities, people finally explore location (Figure 5). People see a map of the United States where states are colored based on the number of new cases per 100,000 reported in that state. Continuous data is broken into quantiles to allow viewers to better distinguish between values [14]. Users can use the slider at the bottom to toggle across years. Up to this point, the application has only shown users data about new cases in 2010. This view is intended to let users get a more longitudinal perspective on the epidemic. It also allows users to look for trends in states that matter to them, foreshadowing the filtering they will be able to do in the filter view.

After viewing this series of visualizations, the user may be left wondering how HIV/AIDS actually affects people. The next panel is headlined “HIV/AIDS is affecting people right now” and it intends to answer that question (Figure 6). It simply shows the number of new cases in the US and the world for the given day. The number of dots is striking; the user has to scroll down to see the entire page. The goal of this visualization is to make people want to learn more, which they can do by clicking the link titled “Meet people like you diagnosed with HIV/AIDS” to move to the filter view.

Filter view
For the filter view (Figure 7), our biggest concern was how we could make the data personalized to each individual user. One early idea we played around with was a quiz-type interface, where people would be asked their current state, sex, ethnicity, and age, and then we would display the results for their demographics’ AIDS diagnosis rate compared to the national AIDS diagnosis rate. However, this wasn’t possible with the data available from the CDC. For confidentiality reasons, the CDC practices cell suppression for groups with a population denominator less than 100, as well as suppressing the data to just two of the three demographic categories at any time: sex, ethnicity, and age [4].

![Figure 5: map displaying state-by-state HIV/AIDS prevalence rates](image)

![Figure 6: visualization showing new HIV/AIDS infections in one day. One dot represents one person.](image)

![Figure 7: Filter view displaying search results](image)
We then decided to use a set of filtering controls at the top of the interface to allow people more direct control over the specific demographic information that went into calculating the AIDS diagnosis rate. One interesting thing we saw with this design is that users would toggle between the different options to see how their choice affected the diagnosis rate. For example, one user selected his state as “California,” then marked his gender as “Male” and watched the diagnosis rate increase. He then marked his ethnicity as “Asian” and watched the rate go back down. We worked around the cell suppression problem in our filtering view by displaying a message that the query is too specific, and we suggest that the user generalize their query a bit more to see results.

The visualization of AIDS diagnosis rates is displayed on a log scale for increased discrimination between the various diagnosis rates. Although we worried that the general public would be unfamiliar with the idea of a log scale, the diagnosis rates are so small for many of the demographics that the same graph on a linear scale was highly clustered at 0%, which made it difficult to see small changes in the diagnosis rate. We also use details on demand on the diagnosis rate visualization, so when people hover over each diagnosis rate line, the demographic information for that rate is displayed above the visualization. We chose not to use tooltips for these details on demand, as we determined that they cluttered the already busy visualization and made it more difficult to target individual lines.

The other key feature of the filter view is the ability to filter both the CDC data and the stories simultaneously. Since our story data is scraped from publicly available HIV/AIDS compilations on the Internet, and many people don’t share their demographic information in these stories, we don’t have stories for every possible combination of demographics. These stories show up with more general filters, like “All ages,” but won’t display otherwise. If there are no matching stories with the demographics the user has selected, we also display a message suggesting that the user try a more general query. For displaying the story itself, we toyed with applying text visualization techniques to the stories, but ultimately decided they would be more powerful in their original form.

**Prevention view**

The goal of the prevention view (Figure 8) was to link statistics to prevention methods, while minimizing the potential for misinforming users. Because data about the effectiveness of different methods of preventing HIV/AIDS does not exist (this would be a highly unethical experiment to run), data for this page is drawn from transmission statistics for 2010. On the left panel of the page, a bar chart displays transmissions categories and the number of men and women who contracted HIV/AIDS from the specific transmission category. When a user holds her mouse over a bar, she can read prevention advice specific to that transmission category. This visualization was designed to give the users access to information that would help them the most. For example, if someone is sharing needles, telling them to use a condom is less important than educating them about the importance of using clean needles.

**User Testing**

We conducted two small user studies. The first was a pilot study with approximately four participants, and the second was a user study with five participants. Qualitative information and quantitative data were collected. Qualitative results were collected by observing people use the system, and also by collecting feedback after people interacted with the system. For quantitative feedback, participants answered a series of survey questions, pre-test and post-test. Survey questions asked about general HIV/AIDS knowledge and beliefs about the epidemic. The full questionnaire is located in Appendix A. Time spent on each page was also recorded.

**RESULTS**

Pilot feedback, qualitative feedback, and quantitative feedback were collected with regards to the study design. Pilot feedback was used to modify the interface before the next round of user testing.

**Pilot feedback**

The first series of pilot users provided feedback on their impressions of the data based on the interface. The interface was modified based on their feedback.

---

**Figure 8: prevention view**

**Figure 9: the default view of the bubble graph for ages before (right) and after (left) incorporating pilot study feedback.**
The bubble-style charts confused users. In the initial prototype, the default view showed the bubbles proportioned by population. Users had to toggle the view in order to show proportions by number of cases. However, this confused users. One reported, “Wow, it’s so sad that so many kids under 15 have HIV/AIDS” interpreting the population-based proportions as case-based. One can guess why pilot user had this impression—when viewing an entire page of visualizations about HIV/AIDS it is easy to assume all visualizations you see reflect data about HIV/AIDS cases. Thus, the default view for both bubble charts was changed to the “cases” view. This way, when the user loads the page they see bubbles proportioned by cases, and only see the bubbles based on proportion if they click the toggle button. Old and new defaults are shown in Figure 9.

In addition to the modifications made to the bubble charts, small changes were made to the “sexuality” display format and legend. The pilot version of this visualization showed three categories: same sex contact, same sex contact and IV drug use, and heterosexual contact. However, this confused people, giving the impression that everyone who contracted HIV/AIDS through IV drug use was also infected through same sex contact. Thus, in the final version, only two transmission categories are displayed: same sex contact and heterosexual contact.

Qualitative observations
Qualitative observations were collected for each of the three views.

Story view
Users seemed intrigued, but also distracted, by the interactive graphics. After the post-test questionnaire, one participant reported that if she’d known she would be asked questions about her knowledge of HIV/AIDS, she would have read more closely, but she had spent more time toggling the visualizations.

The bubble chart showing the ages of people with HIV/AIDS also particularly surprised users. When asked about what part of the visualization surprised them the most, a user in his fifties reported, “I didn’t realize so many people my age had AIDS.” Another user, in her twenties reported, “I’m surprised there are so many older people who have it.”

Filter view
When interacting with the filter view, users appeared to be distracted by the log scale. One artifact of the log scale is the exaggeration of small differences. When some y values are exactly zero and others are not, a zigzag pattern is produced. A handful of users expressed confusion over this pattern.

Users also focused on the graph portion of the filter view first, noticing how if they moved their mouse over the lines their color would change. When they did turn to the filtering portion of the interface, a few expressed confusion about why, after they filtered, there were still fifty lines on the graph. After it was pointed out that they needed to also filter by state, they expressed understanding of the interface design.

Filter view
People spent a relatively brief amount of time on the prevention view. One participant remarked that this page was her favorite because of its simplicity. In general, people tended to hover over all transmission categories, not just one or two.

Quantitative feedback
Users provided quantitative feedback via a questionnaire. Time spent on each page of the site was also tracked.

Before and after interacting with the visualization, participants were asked to rate their knowledge of HIV/AIDS on a scale from 1-5, with 1 representing very little knowledge and 5 representing a great deal of knowledge. Mean knowledge ratings were higher at post-test (mean = 2.9) than at pre-test (mean = 2.0).

Table 1. Pre- and post-test responses to questions about HIV/AIDS knowledge and relevance

<table>
<thead>
<tr>
<th>Question</th>
<th>Pre-test mean</th>
<th>Post-test mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>On a scale of 1 (very little) to 5 (a lot), how much do you think you know about HIV/AIDS?</td>
<td>2.0</td>
<td>2.9</td>
</tr>
<tr>
<td>To what extent do you agree with this statement (1 = not at all, 5 = very strongly): HIV/AIDS just doesn’t affect people like me.</td>
<td>1.5</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Table 2. Pre-test responses to general HIV/AIDS knowledge questions

<table>
<thead>
<tr>
<th>Question</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which racial/ethnic group do you think is most affected by HIV/AIDS?</td>
<td>White: 3 responses Black: 2 responses</td>
</tr>
<tr>
<td>Which age group do you think is most affected by HIV/AIDS?</td>
<td>20-24: 2 responses 25-29: 1 response 30-34: 2 responses</td>
</tr>
</tbody>
</table>
Knowledge of facts about HIV/AIDS was also evaluated before users viewed the site. Users were asked about the ethnicity most affected by HIV/AIDS and the age group most affected. Participant responses are shown in Table 2.

Time spent on each page seemed approximately proportional to page length. Users spent on average 2 minutes and 22 seconds on the story view (standard deviation of 1 minute, 29 seconds), 1 minute and 16 seconds on the filter view (standard deviation of 49 seconds), and 27 seconds on the prevention view (standard deviation of 19 seconds).

All users visited the filter page, but only two out of five users actually clicked on stories to read them. The rest just focused on the graph.

DISCUSSION
In general, participants did not consider themselves very knowledgeable pre-test or post-test, although post-test knowledge ratings were higher. This could be because the visualization gave them additional knowledge about the epidemic. But, the result could also be due to a bias to please the experimenter. In future studies, a better design would be to ask half the participants to rate their knowledge pre-test, and half to rate their knowledge post-test.

Interestingly, while participants did not consider themselves very knowledgeable about the epidemic, they still thought it strongly affected people like themselves. Post-test, there was a small increase in belief that HIV/AIDS affected people like them, but this is likely a result of our extremely small sample size.

Our general knowledge questions affirmed that participants did not know many details about the epidemic in the United States. None of our participants correctly identified the most-affected age group (40-44 and 45-49 year-olds had the most new diagnoses in 2010). Less than half of our participants (2/5) knew that black/African-American individuals had the most new cases of HIV/AIDS of any racial/ethnic group in 2010.

We were surprised that only two users actually read stories from the filter view. Perhaps future versions of the interface could include the start of each individual story directly on the filter view, to entice people to click and read the rest of the story.

The goal of this visualization design was to increase understanding of HIV/AIDS and evoke empathy by creating a compelling series of webpages that told a story about the epidemic and the individuals affected. Our pilot study shows that most of these goals were achieved. Users seemed generally engaged by the site, spending an average of over four minutes looking at it, as well as engaging with all the interactive visualizations. Our quantitative and qualitative results both indicate that the visualization increased users’ understanding of the epidemic in America.

FUTURE WORK
This work is a first step down many possible paths. Future extensions could be to:

- Increase the volume of content, and provide more search parameters. This will allow users to match with more specific stories.
- Incorporate additional functionality into the application. More specifically, allow users to upload their own new stories in addition to reading stories we’ve preloaded the app with.
- Consider applying the same technique to different diseases.
- Give users more personalized feedback based on their location. For example, from the prevention view, provide users with links to their local Planned Parenthood clinic.
- Modify the application for use in other regions in the world. TeachAIDS has already had a great deal of success personalizing its tutorials to different parts of the world.
- Perform more user tests to better understand how this application influences peoples’ beliefs about HIV/AIDS.

ACKNOWLEDGMENTS
We thank Jeff Heer and all the TAs who helped us with this project, plus all our friends who gave us feedback along the way and helped us with user testing.

REFERENCES
1. amFAR. http://www.amfar.org/about_hiv_and_aids/facts_and_statistics_worldwide/.  


APPENDIX A: User pretest & posttest

**Pre-test**
1. What age range do you think is most affected by HIV/AIDS in the United States?
   a. <15  
   b. 15-19  
   c. 20-24  
   d. 25-29  
   e. 30-34  
   f. 35-39  
   g. 40-44  
   h. 45-49  
   i. 50-54  
   j. 55-59  
   k. 60+

2. What group do you think is most affected by HIV/AIDS in the United States?
   a. White  
   b. Black  
   c. Asian  
   d. Latino/a  
   e. Other

3. On a scale of 1 (very little) to 5 (a lot), how much do you think you know about HIV/AIDS?

4. What proportion of the US population do you think has HIV/AIDS?
   a. 0.06%  
   b. 0.6%  
   c. 6%  
   d. 60%

5. To what extent do you agree with this statement (1 = not at all, 5 = very strongly): HIV/AIDS just doesn’t affect people like me.

**Post-test**
1. On a scale of 1 (very little) to 5 (a lot), how much do you think you know about HIV/AIDS?

2. To what extent do you agree with this statement (1 = not at all, 5 = very strongly): HIV/AIDS just doesn’t affect people like me.