Cervical Cancer: Visualizing a Global Health Crisis

Sarvesh Krishan, Keyi Kang, Yisong Tang
Engineering Computer Science 0163
University of California, Davis
skrishan@ucdavis.edu, kkkang@ucdavis.edu,

I. INTRODUCTION

This report intends to inform the reader on the process of designing and implementing an effective storytelling visualization based around cervical cancer data. The GitHub repository can be accessed by clicking on this LINK here.

II. BACKGROUND & OBJECTIVES

Cervical cancer is cancer that starts in the cells of the cervix. The cervix is the lower, narrow end of the uterus (womb) that connects the uterus to the vagina (birth canal). Anyone with a cervix is at risk for cervical cancer but most often occurs in people over age 30. Human papillomavirus (HPV) is the main cause of cervical cancer, which can be combatted through the HPV vaccine. However, cervical cancer is not an equitable issue. The highest rates of cervical cancer incidence and mortality are in low- and middle-income countries. Our objective is to provide a storytelling visualization to the user that reflects the major inequities driven by lack of access to national HPV vaccination, cervical screening and treatment services, and social and economic determinants and raise awareness about how we can abolish cervical cancer as a global health issue within a generation.

III. THE DATA & THE STORY

Our data is retrieved from Kaggle, which can be accessed by clicking on this LINK here. While the files provide projected data using various theoretical levels of HPV vaccine coverage, I focused on the historical/collected data from 2010 to 2020 since the projected data leaves a little too much ambiguity to be desired. Therefore, the data I used is the file `hpv_past_results.csv`, which contains 2212 observations with 10 variables, which some include country name, year, HPV vaccine coverage, income group, possible cases, and possible deaths. I hypothesized that there is a correlation between income group, HPV vaccine coverage, and possible cases/deaths, so I wrote a Python script to perform exploratory data analysis. After aggregating the data by income groups and averaging HPV vaccine coverage and possible case/deaths, it was evident that there was a positive correlation between income group and HPV vaccine coverage and a negative correlation between income group and possible cases/deaths, as expected from my hypothesis. Thus, given this data, the story focuses on informing the user the utmost importance of addressing the inequity of the HPV vaccine amongst countries with differing levels of income and how this inequity allows this global health crisis to persist and in some cases, grow over time.

IV. DESIGN & METHODOLOGY

The structure that I designed the storytelling visualization around was Martini Glass. The user is pulled in through an eye-catching metric, following with an introduction to cervical cancer. With a few additional metrics

and a visualization depicting the severity of the ratio between cervical cancer incidence and mortality, the stage is set to introduce the solution to the health crisis, which is the HPV vaccine. As mentioned in the previous section, HPV vaccine coverage is appears dependent on a country's income group, which is visualized to the user through a bar plot showing the mean HPV coverage across the four income groups. Now that the user has come to terms that there is an inequity issue around the HPV vaccine, we want to zoom out and show just how much this inequity affects countries of different income groups over a decades worth of time. This is accomplished through two line plots where the mean number of possible cases and deaths by income group is plotted against time (from 2010 to 2020). The user will likely observe through these visualizations that the HPV vaccine coverage inequity causes cases/deaths to increase at differing rates, exacerbating the crisis for low to lower middle income countries. Given that the narrative has been delivered, the storytelling visualization opens up to an interactive map where the user can explore any country of their choosing. It concludes with a call to action, pointing out key takeaways that combat the cervical cancer crisis.

V. IMPLEMENTATION CONSIDERATIONS

To deploy the Martini Glass structure, the storytelling visualization was implemented such that the narrative flows downward on the web page, so the user begins at the top of the page and scrolls until the narrative concludes and opens up for interactive data exploration. The initial supplemental visualization depicting the ratio of case incidence to mortality was created in Google Slides and delivered first to offer a simple yet impactful introduction. The following three static plots were created using the Seaborn library in Python with a few considerations. Since all three plots group the countries data by income group, I encode this via distinct colors and a legend to ensure effective readability. Along with readability, the accuracy of the plots were also considered with the bars of the bar plot annotated with numerical labels and the line plots being paired with tables denoting the slope of the lines, which in context is the average case/death rate from 2010 to 2020. These considerations for prioritizing readability and accuracy for the user extends to the primary visualization, the interactive global map, which was implemented using `leaflet.js` and `d3.js`. The purple shading represents the level of HPV vaccine coverage, but since the narrative emphasizes the relationship between HPV vaccine coverage, income group, and possible cases/deaths, the map also has color-encoded icons to denote a country's income group with legends for both HPV vaccine coverage and income group for readability. Additionally, a hover functionality was implemented for accuracy, so that the user has access to the income group, HPV vaccine coverage, possible cases and deaths data for the country they are hovering over.

VI. CHALLENGES

Most major challenges surrounded the implementation of the interactive map. The initial roadblock was learning the structure and usage of `.geojson` files, which personally I was unfamiliar with before this project. At first, I was under the impression that the map simply needs a `.csv` file of my data with unique identifiers for the countries, such as the commonly used 2 letter

abbreviations or 3 letter abbreviations, but I learned that I need a `.geojson` file where my data is nested as additional properties along with country name in the feature set. Therefore, I downloaded a `.geojson` file that contains the names and geometry of countries, imported as a dataframe into Python using the Geopandas library, merged my data, and saved a new `.geojson` file that contained my data for use on the map. Another major challenge arose with the implementation of color-encoded icons to visually denote the income level of a country on the map. There were multiple instances where I had to debug my code as the icons would not show on the map, but most were caused due to the separation of the `style` function and the insertion of the icons. Once merged together, the issues were resolved, but it took a realization to get there.

VII. RESULTS & FEEDBACK

Feedback was collected from both Professor Liu (course instructor) and Jessica Chen (course teaching assistant) during their Office Hours to evaluate the current effectiveness of the work-in-progress builds. Changes included to increase font size, bold phrases to emphasize takeaways for the reader, increase accuracy of the visualizations through annotated numerical labels, and visual indication of hyperlinks through [LINK]. Thanks to the valuable feedback received, the storytelling visualization iteratively improved for the end user.

VIII. FURTHER WORK

The time constraints did not permit the implementation of everything I had in mind. To improve this visualization storytelling, I would allow the user to pick the year through a drop down menu to depict the bar plot and interactive map with the respective year's data (currently displays the year 2020). I would also change the two line plots from being static to interactive so that the user can hover over the points that comprise the line to see their value (possible cases/deaths) for that given year.

IX. REFERENCES

1. https://www.cdc.gov/cancer/cervical/basic_info/index.htm
2. https://www.who.int/news-room/fact-sheets/detail/cervical-cancer
3. https://www.emro.who.int/noncommunicable-diseases/campaigns/cervical-cancer-awareness-month-2024.html#:~:text=January%20is%20Cervical%20Cancer%20Awareness,cancer%20within%20a%20few%20generations.
4. https://www.cdc.gov/cancer/dcpc/resources/features/cervicalcancer/index.htm

X. DIVISION OF LABOR

I (Sarvesh Krishan) wrote the code for the implementation of the storytelling visualization and this report. My group mate, Keyi Kang, helped me create the presentation slides together. Unfortunately at this time, our group mate, Yisong Tang, has no participation in this project.