Cancer Data Dashboard for Clinical and Hospital Management

Hanke Heun-Johnson, Carlin Cherry, and Kevin Tsang

University of Southern California, Los Angeles CA, USA
heunjohn@usc.edu, ccherry@usc.edu, kktsang@usc.edu

Abstract. In this paper, we describe the development of a data dashboard for a large cancer hospital that serves patients throughout the Los Angeles area. The dashboard design was inspired by state-wide and national cancer data dashboards, charts from published cancer research papers, and consultations with clinicians and hospital management. Using a React framework, we created d3 and Mapbox charts related to cancer characteristics, patient demographics, and geographical distribution of patients in relation to historical cancer mortality data. We additionally provide details on generating mock data to approximate the final dataset. Finally, we suggest improvements to be implemented before deploying the cancer data dashboard at the hospital.

Keywords: Cancer · Dashboard development · Data visualization · JavaScript · d3 · React

1 Introduction

Specialized cancer hospitals in urban areas serve thousands of patients each year, and collect large amounts of data on demographics, clinical characteristics, and short- and long-term outcomes. Providing easy access to these data to clinicians and hospital management teams may improve clinical management of current patients and ensure equity in health care access in the service area of the hospital.

This paper describes the development of an internal hospital dashboard with information about patients that come to a hospital for cancer diagnostic services or treatment. Patients initiate contact at various clinical stages, e.g. for screening, diagnoses, secondary opinions, treatment, or palliative care. We aimed to incorporate information about cancer stages and malignancy, the treatments that are being performed, demographic information about the patients, the types of cancers that are being diagnosed and treated, and the long-term disease progress and outcomes. Other information that is presented is focused on the reach of the hospital in the local catchment area, and the distance of patients to the hospital. These numbers are compared to local cancer mortality rates to determine whether the hospital is treating sufficient numbers of patients in areas where clinical care is needed most.

Generally, hospital data dashboards are not publicly accessible, and thus it is hard to determine whether similar dashboards exist for other hospitals.

Public dashboards with cancer statistics exist [2,12], but do not contain HIPAA-protected data that are useful to determine a hospital's patient mix, clinical benchmarks, and geographical reach. For this specific hospital, the data dashboard will provide a novel approach to combine patients' clinical information with geographical variables and other demographic information.

The intended end users will be clinicians and hospital management staff who can use the dashboard to make decisions about staffing, diagnostic needs, hospital reach and local area representation.

Section 2 contains a description of related visualization efforts of cancer data, section 3 describes the data that were used, section 4 explains the design and development approach, section 5 showcases the main aspects of the website, and section 6 details contributions to the field and next steps.

2 Related Work

Inspiration for our dashboard came from two websites, created by The Centers for Disease Control and Prevention [2] and the Department of Health in Pennsylvania [12]. Both present dashboards that include interactive bar charts with metadata about patient's race and ethnicity distributions. The CDC's dashboard consists of a map of the US with clear color coding of cancer incidence by state that allows for easy geographical comparison. The dashboard specific for Pennsylvania on the other hand is far more complex and contains many layers of information. While the latter dashboard is hard to navigate and understand, we appreciated the in-depth storyline and the use of summary boxes at the top of the overview page, which we implemented in our dashboard as well. To make our dashboard more intuitive to navigate, we opted for a fixed menu on the left of the page.

Another topic we wanted to highlight is the distance of patients to the hospital, and the geographical reach of the hospital in the area. Inspiration for this topic came from a paper by Welke et al.[16], which shows a map with travel distance to different hospitals across the US for surgery, represented by lines in a network. As we only have data for one hospital, lines won't be necessary; a simple dot map would suffice. This also allows us to color code the dots to indicate tumor stages for each of the patients; malignancy status is implemented with a filter menu.

The alluvial diagram on our dashboard was based on figure 1d in a recent paper on pancreatic cancer subtypes [10]. Other papers served as inspiration for future implementations, e.g. temporal line charts [6], 3D charts of tumor growth for each of the patients [13], and decay charts to indicate differential health care access and mobility based on socioeconomic status [7]. All of these charts would require additional temporal data that is available in the hospital's database.

3 Data

Data from the cancer hospital were not readily available because of lengthy Institutional Review Board procedures, and thus we utilized a publicly-available Kaggle dataset [8] augmented with mock data on patient demographics for the development phase of this project. This dataset contains information on physiological tumor characteristics such as radius and compactness, and whether a tumor is malignant or benign. To more closely match the final dataset for this dashboard, we added data points for tumor staging (1-3 for benign, and 1-4 for malignant tumors) loosely based on the size (higher stage for larger tumors), and race/ethnicity (random assignment, based loosely on Los Angeles county demographics). The mock summary statistics can be found in Table 1.

Zip codes were added to observations as follows: a shapefile of all zip codes in California [1] was reduced to zip codes in the vicinity of the hospital using QGIS-LTR software, converted to geojson [3], corrected for "winding issues" [17], and boundary details were reduced to 15 % of the original complexity [15]. The source file also contains population and average age for each zip code. Zip codes with a population of 0 were removed. To assign zip codes logically to the observations in the cancer dataset, we based the frequency of patients in each zip code on the cancer mortality rates for each zip code in 2019 [5], and the latter information was also added to the dataset for each zip code. We choose to focus on cancer deaths in relation to the number of patient, as areas with relatively more cancer deaths per capita could benefit from better health care access to reduce the number of deaths.

Table 1. Patient characteristics in dataset. Asterisks (*) indicate randomized/mock data.

Measure	Outcome
Benign cancer	62.0 %
Malignant cancer	38.0 %
Stage 1*	38.2 %
Stage 2*	24.9 %
Stage 3*	26.8 %
Stage 4*	10.1 %
Avg tumor radius	14.1 mm
Asian*	9.8 %
Black*	11.8 %
Hispanic*	47.3 %
White*	31.6 %
Avg distance to hospital*	28.0 mi

Another statistic that would be important to ensure adequate health care access is the distance of a patient to the hospital. For each zip code, we determined the latitude and longitude of the center and the distance from that zip code to the hospital (as the crow flies to zip code 91010) [9]. Several zip codes were not present in the lookup table; the coordinates were determined with Google maps instead and added to the lookup table. To increase the sample size in each zip code, the dataset was expanded to 1250 observations by randomly sampling from the existing dataset. To prevent all dots in each zip code to aggregate on the map, a random value was added to or subtracted from each latitude and longitude, deviating no more than 2.2km in either direction from the center of each zip code.

4 Approach

The design was first and foremost guided by the needs of clinicians and management of the hospital; during consultations the teams requested visualizations of physiological tumor characteristics by tumor stage, malignancy status by patient demographics, and the geolocation of patients in relation to the hospital. For the implementation, we used the examples described in more detail in section 2, (cancer dashboards by the CDC and the Pennsylvania Department of Health) as main inspiration.

The dashboard navigation follows a bottom-up approach, from physiological characteristics of the patient to the geographical reach of the hospital. First we describe tumor characteristics by tumor stage, then tumor malignancy by demographics (race and ethnicity), then we show where patients reside in relation to the hospital (i.e. how far they need or choose to travel to receive care), and finally we evaluate the number of patients at the hospital in relation to the cancer deaths in the region. The bubble chart indicates which zip codes in the area are underrepresented as patients (any bubbles below the general trend line would benefit from outreach by the hospital). Since our data are artificially generated based on this correlation, mechanically there is little variation from the trend line; real-world data would provide better insights.

On each page, info boxes at the top summarize relevant main statistics at-a-glance. Color coding of ordinal (tumor stages), nominal (cancer deaths) and categorical (benign vs. malignant and race/ethnicity) variables is implemented in color-blind-safe color schemes.

5 System

We used an open source creative commons dashboard template [4]. This enabled us to customize, rather than create, the backbone website on which we wanted to display our custom visualizations. Our site is built and customized using Node.js, React, and Bootstrap. All charts are created with d3 or Mapbox. The Mapbox chart [11] and alluvial diagram [14] were based on templates found online. For performance purposes, we hosted select charts on separate servers.

Highlights of the dashboard include 1) summary boxes at the top of each page that will contain outcomes that are automatically updated when new data are added in the final version of the dashboard; 2) Responsive "card" layouts to organize the charts; 3) Detailed charts on tumor physiology by stage; 4) An interactive chart to explore patient numbers by race/ethnicity; 5) an alluvial diagram to easily determine patient proportions by race/ethnicity, malignancy, and cancer stage; 6) Maps to evaluate hospital performance metrics and treatment accessibility by zip code, and determine under-representation of patients from certain local areas based on cancer mortality data; 7) Color-blind safe color schemes.

6 Conclusion

With this dashboard we created a novel visualization tool for a specific large cancer hospital in the Los Angeles area. The data charts will help clinicians and hospital management with data-driven decision making regarding cancer treatments, establishing an equitable patient mix, and targeting localized outreach efforts, which ultimately could improve patient outcomes and access to care.

Future improvements to the dashboard include:

- Implementation of the final dataset after obtaining Institutional Review Board approval.
- Customization and refinement of visualizations based on underlying trends observed in the real-world data; as we mainly used randomized or mechanically correlated data points to feed into the charts, we were not able to explore and highlight any interesting correlations that may exist.
- Creation of additional charts using temporal data with time trends for individual patients and patient groups.
- Automation of pre-processing steps to generate appropriate input files for each of the charts, and to feed summary statistics into the info boxes at the top of each page.
- Resolving implementation issues for certain charts into the React framework (Mapbox chart and alluvial diagram).

References

1. Census zip code tabulation areas, california, 2010. https://earthworks.stanford.edu/catalog/stanford-dc841dq9031, last accessed November 20, 2021
2. Cdc cancer statistics at a glance. https://gis.cdc.gov/Cancer/USCS/#/AtAGlance/, last accessed November 26, 2021
3. Shapefile to geojson converter. https://products.aspose.app/gis/conversion/shapefile-to-geojson, last accessed November 20, 2021
4. Paper dashboard react. https://www.creative-tim.com/product/paper-dashboard-react, last accessed November 20, 2021
5. California health and human services - 2019 final deaths by year by zip code. https://data.chhs.ca.gov/dataset/death-profiles-by-zip-code/resource/d4711b8e-6eb4-417c-91f5-ee075558adbe, last accessed November 20, 2021
6. DeSantis, C.E., Ma, J., Goding Sauer, A., Newman, L.A., Jemal, A.: Breast cancer statistics, 2017, racial disparity in mortality by state. CA: A Cancer Journal for Clinicians **67**(6), 439–448 (2017). https://doi.org/https://doi.org/10.3322/caac.21412
7. Jia, P., Wang, F., Xierali, I.M.: Differential effects of distance decay on hospital inpatient visits among subpopulations in florida, usa. Environmental monitoring and assessment **191**(2), 1–16 (2019)
8. Kaggle dataset - prediction of benign and malignant breast cancer. https://www.kaggle.com/khotijahs1/prediction-of-benign-and-malignant-breast-cancer, last accessed November 20, 2021

9. Zip code lat/lon and distance excel lookup table. https://www.youtube.com/watch?v=cpUVV8q7WNo, last accessed November 20, 2021
10. Li, Y., Li, Y., Yu, S., Qian, L., Chen, K., Lai, H., Zhang, H., Li, Y., Zhang, Y., Gu, S., Meng, Z., Huang, S., Wang, P.: Circulating EVs long RNA-based subtyping and deconvolution enable prediction of immunogenic signatures and clinical outcome for PDAC. Mol. Ther. Nucleic Acids **26**, 488–501 (Dec 2021)
11. Mapbox template for multiple marker filter. https://docs.mapbox.com/mapbox.js/example/v1.0.0/markers-with-multiple-filters/, last accessed November 20, 2021
12. Pennsylvania department of health, cancer statistics dashboard. https://www.health.pa.gov/topics/HealthStatistics/CancerStatistics/dashboard/Pages/Cancer-Dashboard.aspx, last accessed November 26, 2021
13. Rani, V., Dhenakaran, S.: Classification of ultrasound breast cancer tumor images using neural learning and predicting the tumor growth rate. Multimedia Tools and Applications **79**(23), 16967–16985 (2020)
14. Alluvial diagram template. https://bl.ocks.org/d3noob/06e72deea99e7b4859841f305f63ba85, last accessed November 20, 2021
15. Mapshaper. https://mapshaper.org/, last accessed November 20, 2021
16. Welke, K.F., Pasquali, S.K., Lin, P., Backer, C.L., Overman, D.M., Romano, J.C., Jacobs, J.P., Karamlou, T.: Hospital distribution and patient travel patterns for congenital cardiac surgery in the united states. The Annals of Thoracic Surgery **107**(2), 574–581 (2019)
17. Observable - d3 black box map. https://observablehq.com/@john-guerra/d3-black-box-map, last accessed November 20, 2021