Jennifer Rogers
Patient Similarity Final Project

Github repository:
https://github.com/visdesignlab/dataExploration

Assessment and prediction of a patient's outcome following a given medical procedure is a combination of multiple, varied and complex health data. There has been significant interest in the use of visual analysis tools to assist in the synthesis of multivariate data for decision making (Fan et al. 2017). With this, is the challenge of developing meaningful visualization of temporal data (Gotz et al. 2014). This research endeavour is applied to the healthcare field with the development of tools to analyze patient cohorts and meaningful synthesis of EHR data (Bernard et al. 2015).

This project seeks to develop a visual tool for orthopedic surgeons that highlights patterns within patient records to better predict patient trajectories following a given procedure. This can be divided into 2 major tasks:

1. Dynamic formation of patient cohorts. The practitioner will need to be able to adjust filtering filters to form cohorts based on specific criteria. This includes the selection of demographic properties as well as temporal events e.g.- all patients who have had spinal injection followed by physical therapy evaluation.

2. Visualization of collective progression of physical function and pain levels of patient cohort with the ability to align with a given procedure. The practitioner will need to view the progress and physical function of the selected patients following a given procedure. This would track the trajectories of the cohort through VAS and PROMIS Physical function scores, in the hope that practitioners could evaluate the cohort to draw conclusions whether a given procedure had a perceived positive or negative effect on the cohort.

The data used for the project is provided by the Orthopedic research center from their data warehouse. The data is provided in a series of CSV files and will need cleanup. Columns with empty values need to be filled with zeros. The patient entries need to be grouped in a meaningful way as the number of entries for patient visits are hundreds of thousands of lines long. Aggregation by patient ID and repeated codes will assist in the data bloat in the initial dataset. Binning of demographic information, specifically age and BMI groups will assist in developing initial groups within the body of patients. Length of groups and min/max values for patient visits and codes will provide insight on distribution of procedures in the EHR records.

The main achievement for the visualization would allow users to interact with the data to identify patterns within groups of similar patients in a more intuitive way. The visual tools must be able to communicate temporal data as well as information for a large number of patients at once. The visualizations must respond to changes in the similar patient calculations that is done through the interface. The visual aspect of the application consists of 3 major components.

1. Visualization of patient population/ large cohort

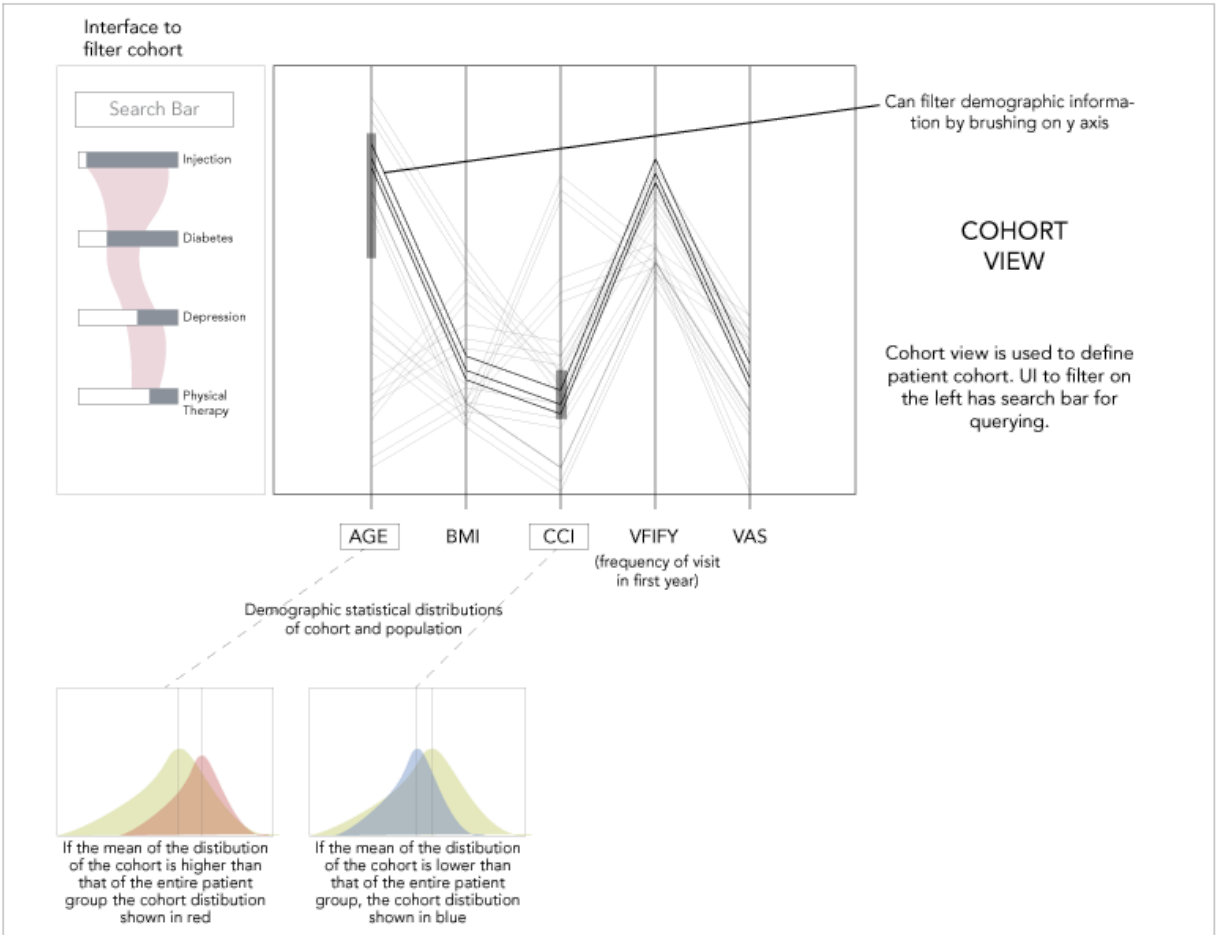
Development of visualization to highlight initial patterns in patient data including trends in demographic and presence of procedures. The visualization will have to show an overview of approximately 6000 patients. The visualization would need multiple views to highlight aggregated distributions of patient orders and codes, as well as multivariate view to show any grouping between multiple attributes within the patient population.

2. Interface to dynamically filter patients by demographic and procedural information.

The application must respond to the changing parameters that the user can adjust in the interface. This controls the event and demographic filtering for the patient cohort. For the first stage, the application will have to visualize multiple variables within a large group of patient records. The user would have the ability to track the visual output of each filter and can add/subtract filters as well as rearrange the sequence of filtering.

3. Graph visualizing time series data regarding a subset of the cohort or patient population.

The second phase will visualize multiple attributes and time series data from the filtered group. Progress of patients is expressed through measured physical function scores known as PROMIS scores. Visual Analogue Scale (VAS) tracks patient self reported pain levels. Head and neck are used in this project. Oswestry Disability Index (ODI) measurements are recorded in an attempt to quantify disability levels of low back pain. These 3 forms of measurement are used to assess the given body of orthopedic patients and will be used to show fluctuation in pain and physical function in the visualization.



Schedule by Week

November 3. Dynamic filtering of patient cohort is implemented. Visualization of entire population is filtered.

November 10. First visualization of the cohort data is complete. Visualization for demographic information.

November 17. Patient function measurement visual is functional.

November 24. Visualization components are fine tuned in user navigation and aesthetic.

November 28. Project regarding the scope of the semester is complete

References

Bernard, Jürgen, et al. "A visual-interactive system for prostate cancer cohort analysis." *IEEE computer graphics and applications*35.3 (2015): 44-55.

Du, Fan, et al. "Finding similar people to guide life choices: Challenge, design, and evaluation." *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017.
Gotz, David, and Harry Stavropoulos. "Decisionflow: Visual analytics for high-dimensional temporal event sequence data." *IEEE transactions on visualization and computer graphics* 20.12 (2014): 1783-1792.

Rind, Alexander, et al. "Interactive information visualization to explore and query electronic health records." *Foundations and Trends® in Human–Computer Interaction* 5.3 (2013): 207-298.