

A3 Story Board

Group Members:

David Caldwell

Shirley Leung

James Wu

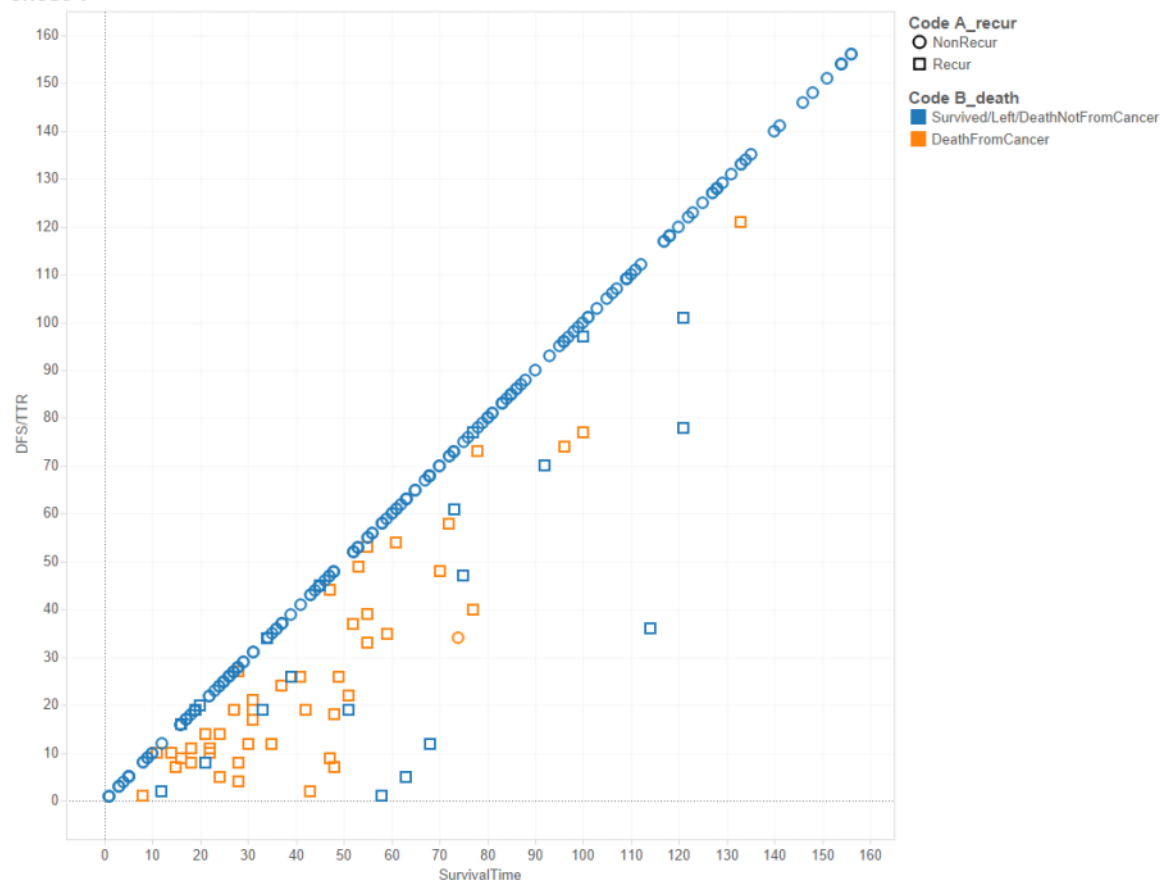
Tableau Exploration

Our first step was to pick a dataset and explore it via Tableau. We started with various global health data sets from the WHO, but decided that due to difficulties with data curation and inconsistencies across data to focus on a "cleaner" dataset. The Wisconsin Prognostic breast cancer database was found via the machine learning database repository link on the website.

The histologic characteristics provided on the cells in this database suggested to us the interesting idea of being able to visualize simultaneously patient cell characteristics as well as patient prognoses.

Our first graphic below in Tableau demonstrates differences in survival time and the recurrence time for cancer for the entire patient subset. We thought from looking at this scatterplot that looking into the cell characteristics that were associated with the severity of cancer would be an informative visualization.

Sheet 1

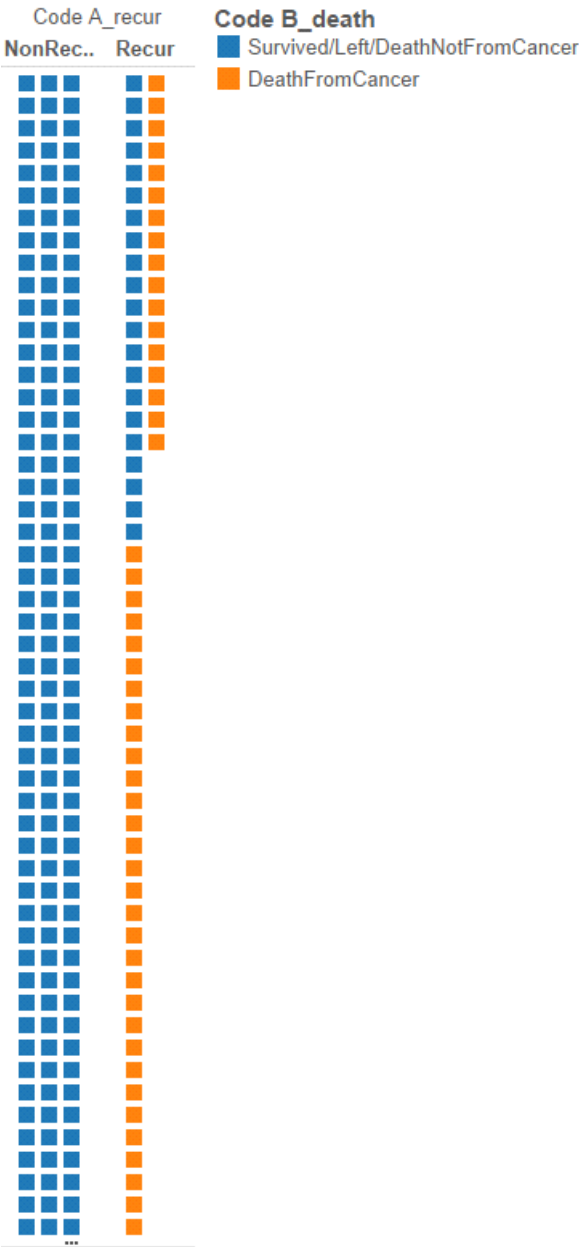


Sum of SurvivalTime vs. sum of DFS/TTR. Color shows details about Code B_death. Shape shows details about Code A_recur. Details are shown for Patient.

Our second visualization was another look at recurrence and survival. From here we started to develop the idea of color coding or highlighting patients visually to provide a broad graphic

summarizing survival.

Sheet 2



Code B_death
(color) broken
down by Code
A_recur. Details
are shown for
Patient.

Our scatterplots below exploring different characteristics of cells, and the potential differences in patients that survived or died from their cancer, suggested using small multiples or parallel coordinates to visualize the multidimensional nature of our data. We decided on using parallel coordinates.

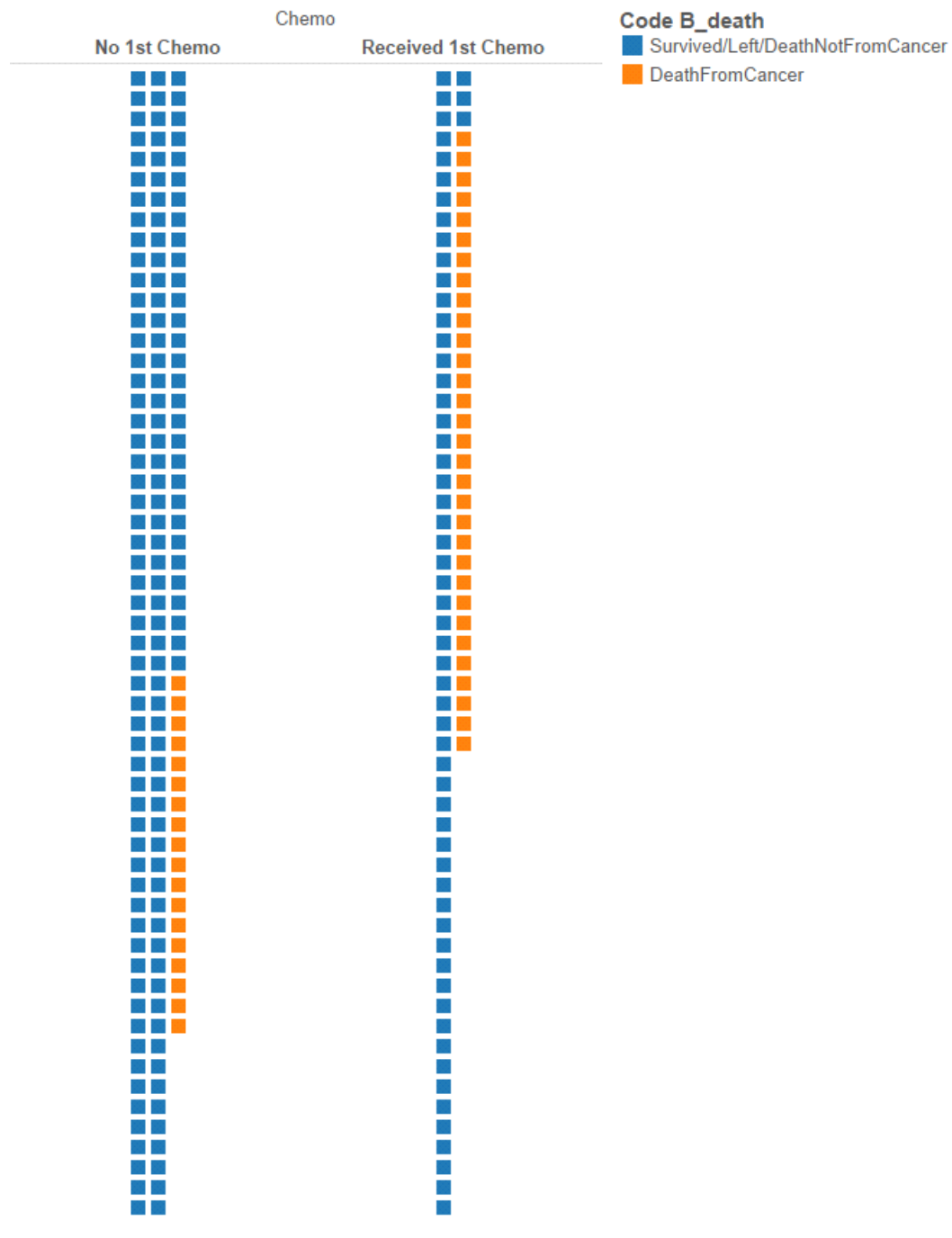
Sheet 3



Sum of Area, sum of Compct and sum of Node All vs. sum of Smooth, sum of Texture and sum of Fract D. Color shows details about Code B_death. Details are shown for Patient.

The graph below demonstrates the differences in patients who either received a 1st round of chemo, or didn't. We ended up using the color outline of patients to represent whether or not they received a 1st round of chemotherapy.

Sheet 4



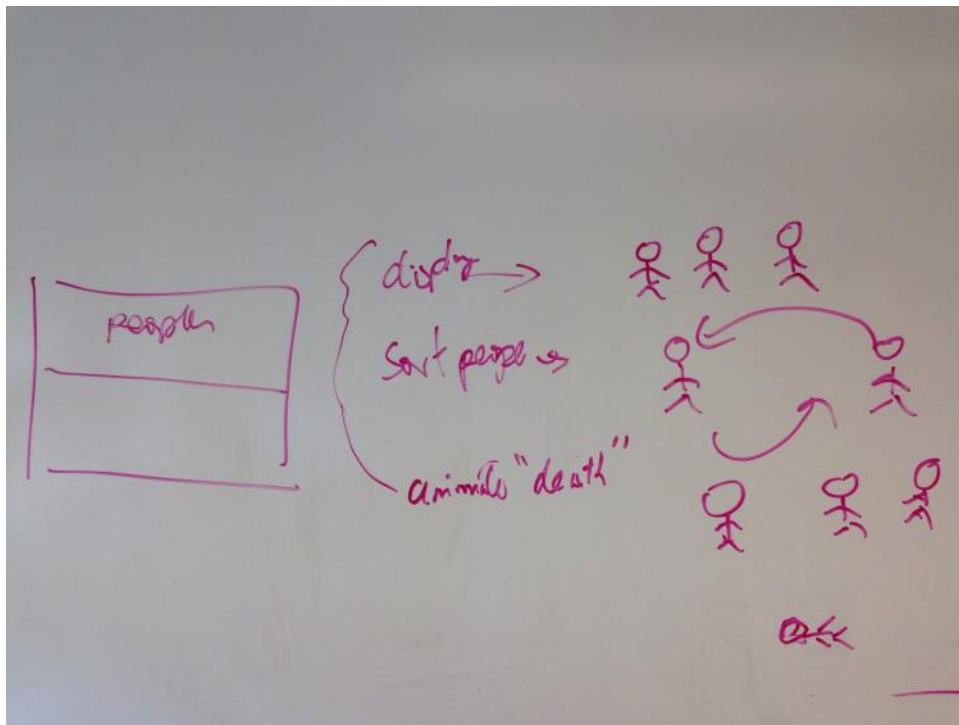
Code B_death (color) broken down by Chemo. Details are shown for Patient.

Visualization design.

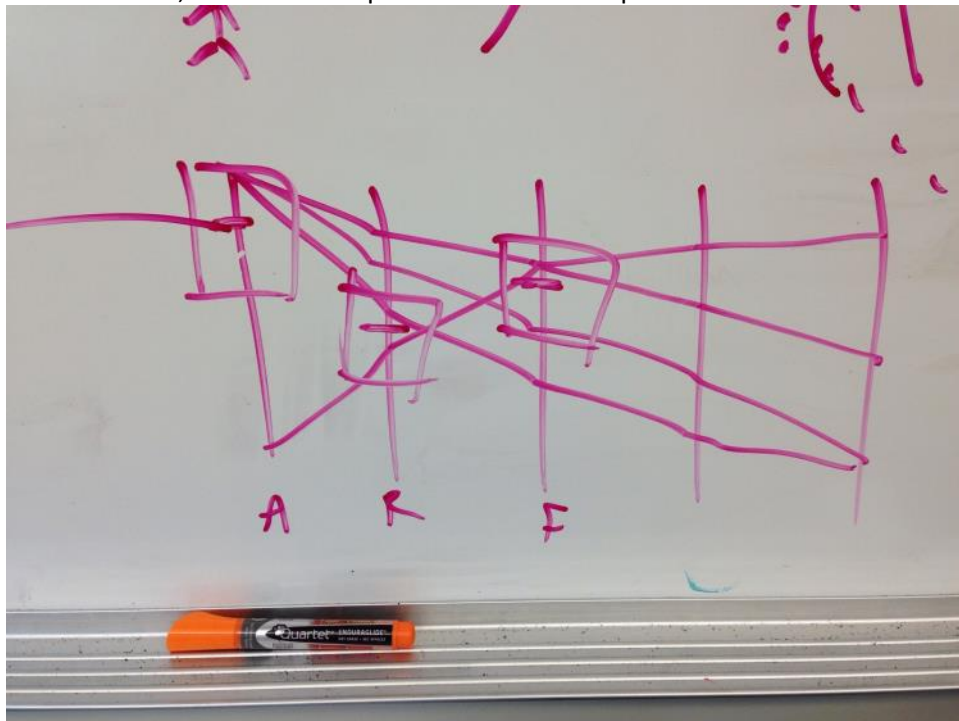
After doing exploration in Tableau, we began sketching out our visualization.

In the whiteboard diagrams below, the initial stages of our visualization can be seen.

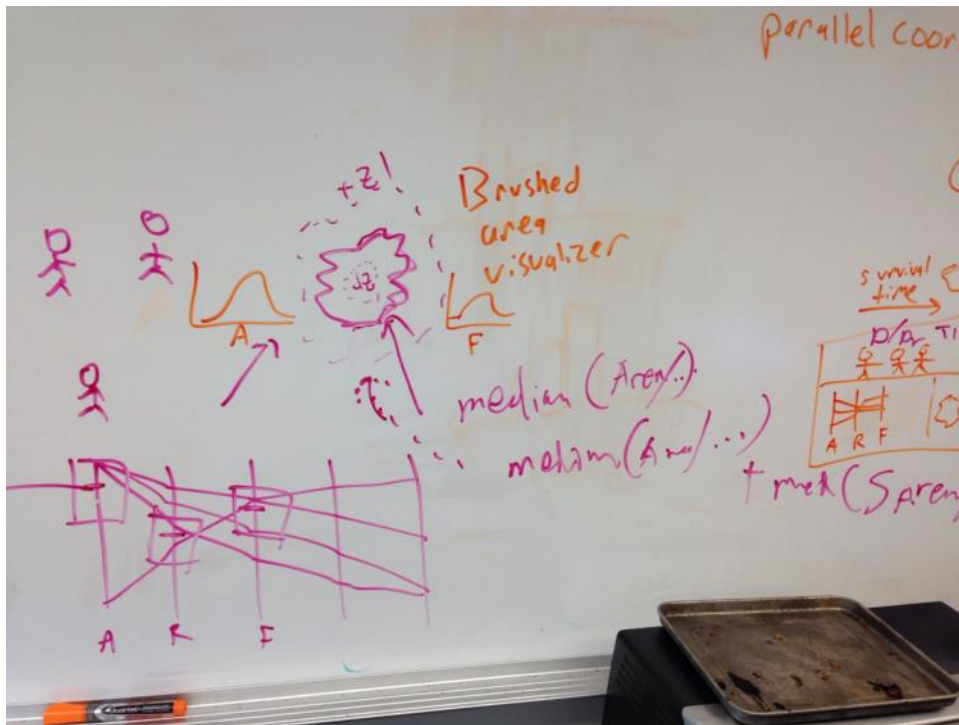
The image below illustrates the idea of stick figure represents of people dynamically updating.



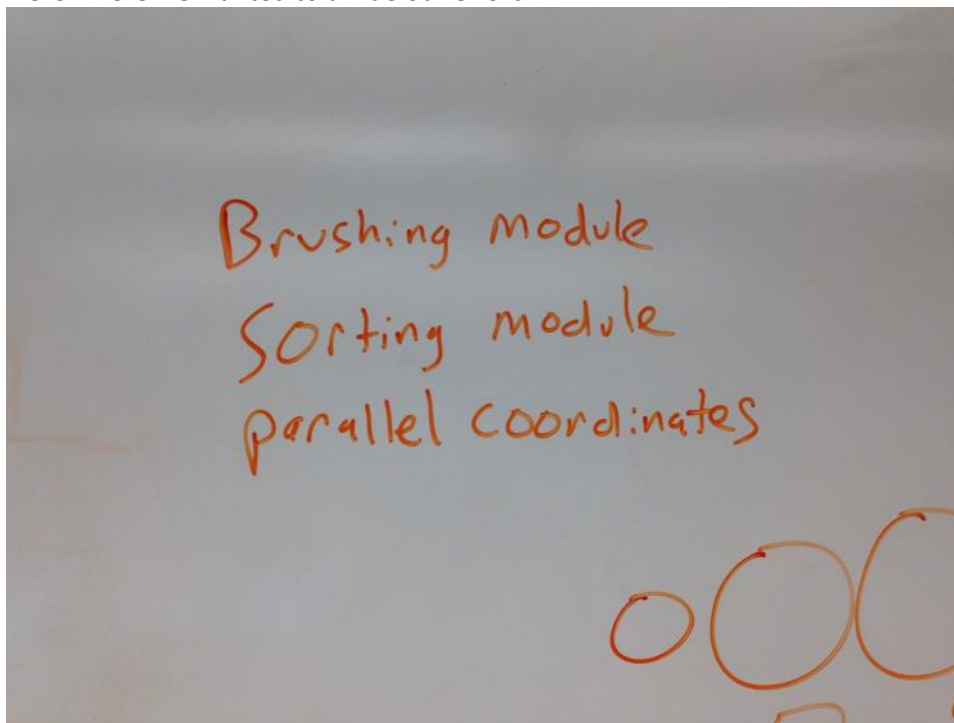
This image below shows our thoughts on using parallel coordinates and brushing to select cell characteristics, and the subsequent visualization of patients based off of these cell characteristics.



Once the cell characteristics were brushed, we envisioned creating an example cell that highlighted some of these characteristics. We also considered the idea of histograms showing the spread of cell characteristics, but decided against this in our final implementation.



The image below highlights the three components that we viewed as critical to our visualization, and were where we wanted to divide our effort.



The image below shows our first thoughts on the integration of all three of these parts.



Final Design and Implantation.

The drawing below illustrates our more refined idea for our visualization. The top part illustrates our parallel coordinates and brushing, that is then extracts data for the subsequent visualization parts.

The second part of the diagram shows the idealized cell representation, which demonstrates the average cell within a brushed selection, plus or minus z-scores. Characteristics to be visualized included area and "fractalness", which is a metric of how "jagged" or rough the perimeter of a cell is.

The third part shows the pictorial representation of patients selected via the brushing. We originally wanted to have clickable and sortable patients via interaction with this part of interaction, and on the individual implementations of the parts we were able to achieve this, but on subsequent integration we decided to simplify our design and have the majority of interaction occur with the parallel coordinates.

① Area
PARALLEL
COORDINATES
(reorderable
axes, z-
score coloring
variable on
click)

Fractal
dimension

Radius

Brushing to
select data for ② + ③

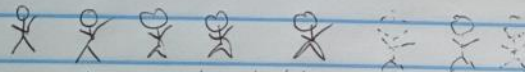
②
IDEALIZED
CELL
REPRESENTATION

- | z-score
area, fractal
dimension

Median or
mean
area,
fractal dim

+ | z-score
fractal dimension

③ TRUE/FALSE
PICTORIAL
PATIENT
CHARACTERISTIC
(SORTED)
(by clicked variable in ①)



User clicks to select how
they want the
people outlines colored

⊙ = chemo yes
⊙ = chemo no

⊙ = death yes
⊙ = death no