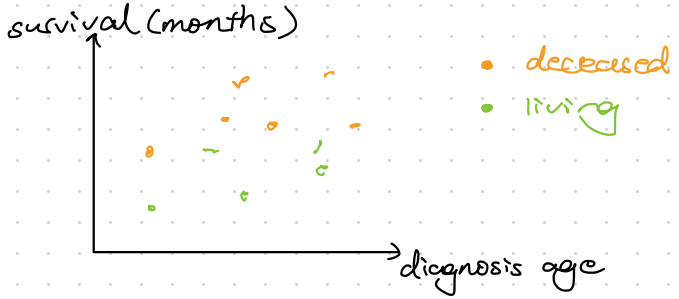
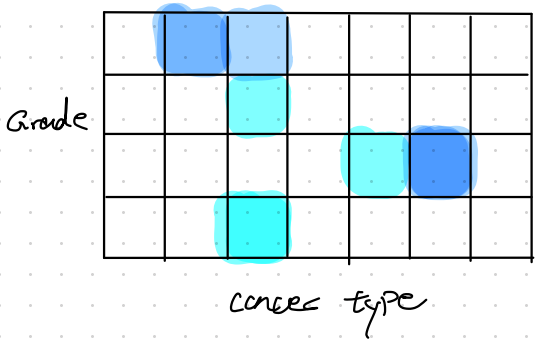
Stage #1 — Brain Storming

Design #1 — diagnosis age & survival status & survival



Survival Month

Design #2 — cancer type & grade & mutation count

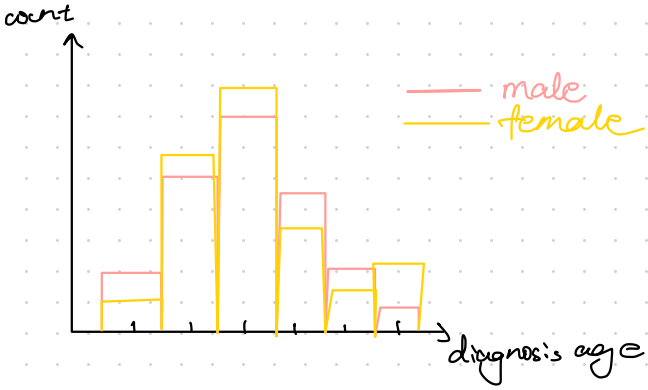


→ high mutation cnt

→ low mutation cnt

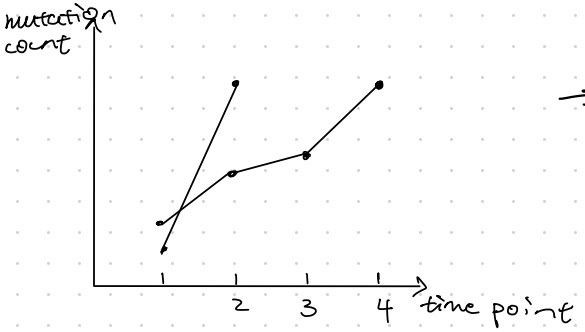
Mutation Count

Design #3 — diagnosis age & sex (overlayed histogram)



Other

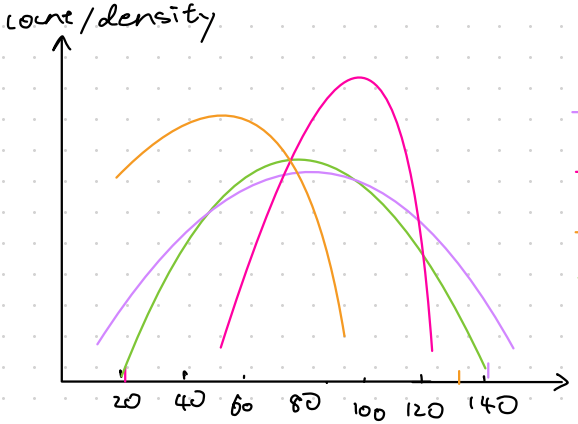
Design #4 —— mutation count change @ different time points



→ hover over to see patient ID

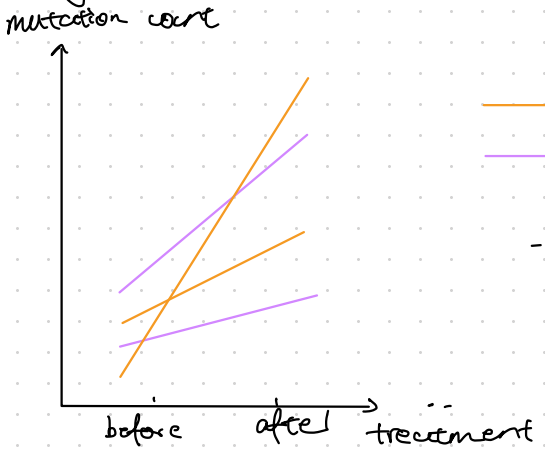
Mutation Count

Design #5 —— grade & DECEASED survival (months) (histogram)



Survival Month

Design #6 —— whether treatment reduce mutation counts.

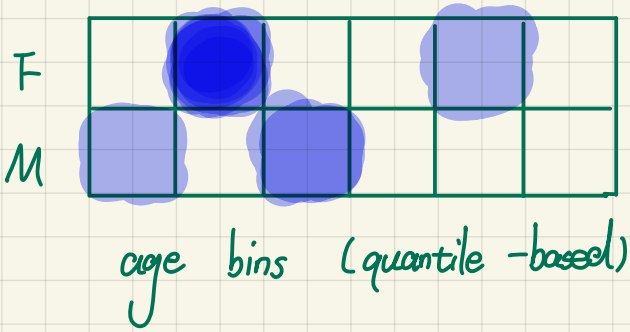


Mutation Count

- maybe grey out the rest & just highlight
  - most significant change
  - average change (☆)

Sheet 1 BrainStorm

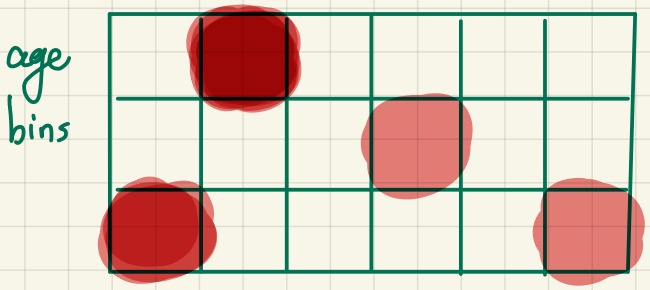
Design 7: Heatmap - living status by age by gender



proportion of deceased.

Living Status

Design 8: Heatmap - living status by cancer type by age.

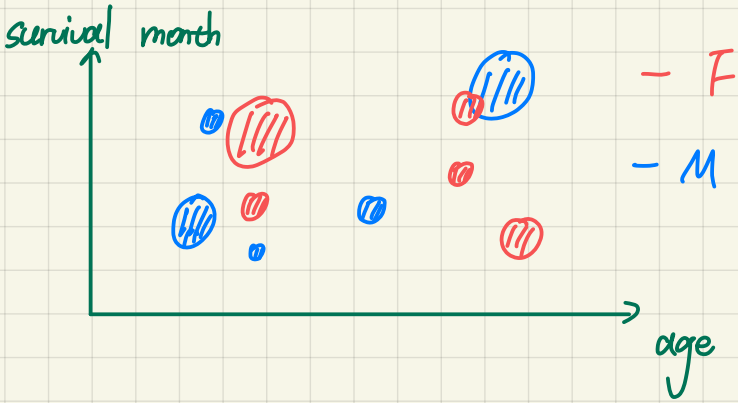


proportion of deceased.

Living Status

Category: survival month

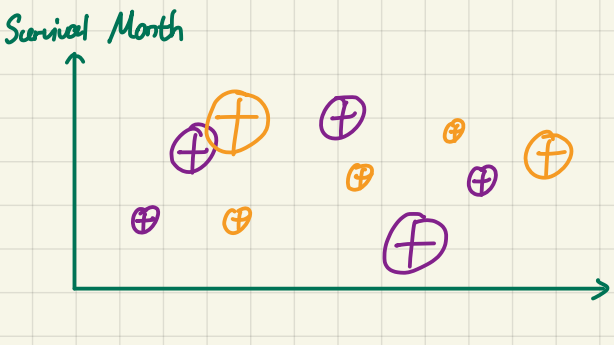
Design 9: Scatter Plot - Survival Month by age by gender, topped with mean mutation counts.



- F
- M

size of the bubble indicates the mean mutation counts, probably in log scale.

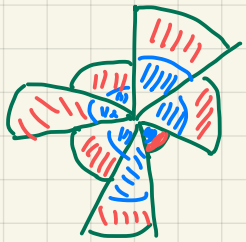
Design 10: Scatter Plot – Survival Month by age by treatment, topped with mean mutation counts.



- Radiation
- Chemotherapy.

Survival Month

Design 11: Rose Chart – Cancer Type by gender

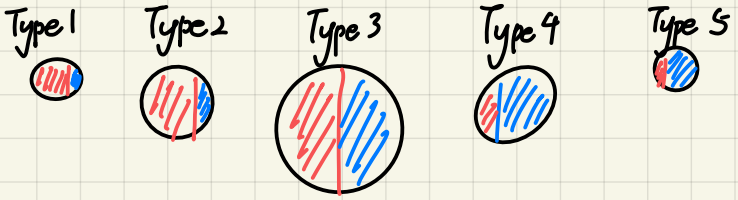


- F
- M

Very Ugly!

Other

Design 12: Bubble – Survival Rate by Cancer Type



- living counts
- death toll.

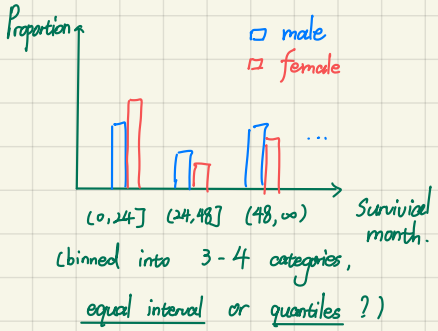
More lethal

less lethal

Living Status

the size of the bubble indicates the # occurrence of that cancer type.

Something fun: Survival month (bin) by gender



(binned into 3-4 categories, equal interval or quantiles ?)

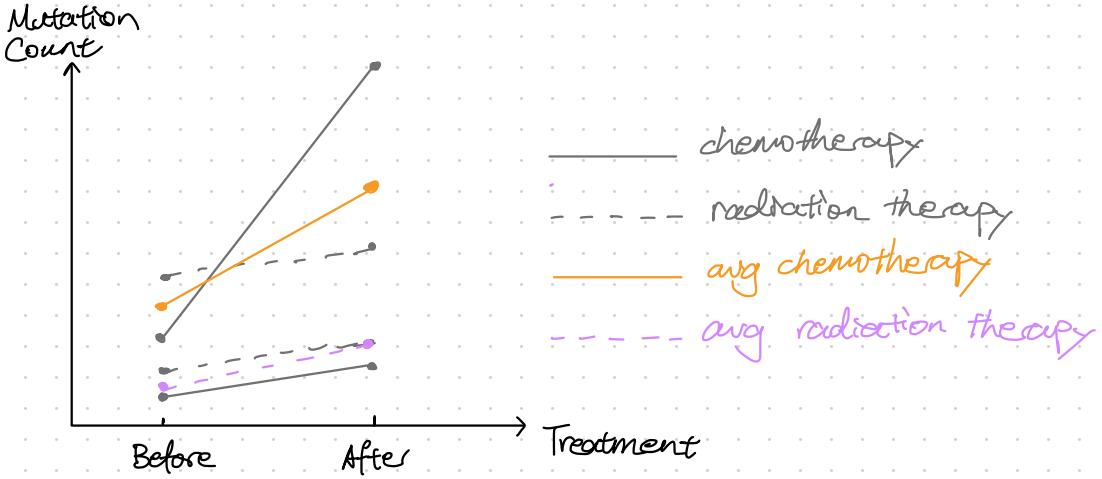
resemblance to TVD in a permutation test.

Stage #2 — Initial Designs

Design #6 — Mutation Count

① Layout

Title: mutation change before & after different treatment

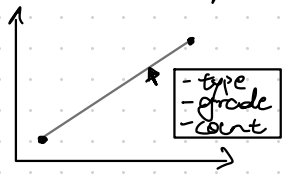


② Focus

- Goal: which treatment is more effective at treating cancer
- Focus/Emphasize: there are many individual lines that can be plotted, but the AVERAGE of the two treatments can be very informative.

③ Operate

- HOVER OVER each line to see individual data:
  - ↳ cancer type, grade, mutation count



④ Discussion

- Advantages: - shows individual information BUT w/ a FOCUS on the averages.
- Disadvantages: - if dataset grows larger, the individual lines could become cluttered.

⑤ Meta-Information

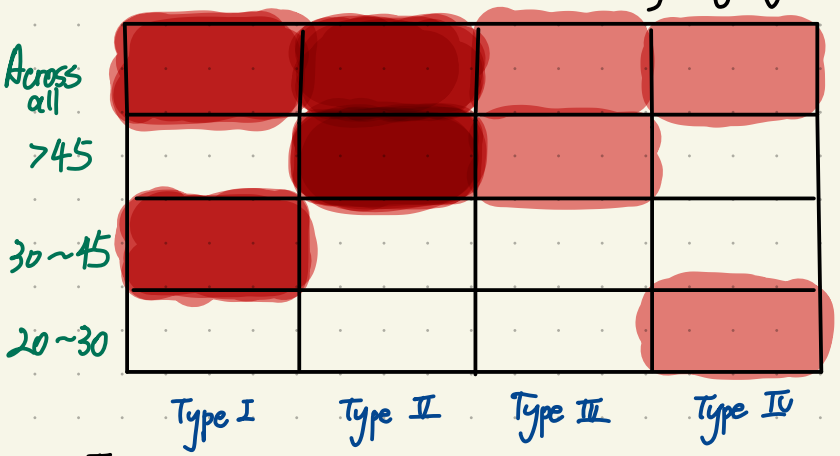
- Title: mutation change before & after different treatment
- Author: BubbleTea
- Date: 05/12/21

Sheet 3

Design 8 — Heatmap : living status by cancer type by age.

① Layout & Information

Title: Survial Rate by Cancer Type by Age



survivial proportion.

Time: ...

Source: ...

Author: ...

② Focus

Goal: which cancer type is most lethal in different age groups.

Emphasis: on cancer type, by highlighting the survival rate across all age groups.

③ Operate

- Hover to display tooltips: tooltips in each region contains
  - cancer type
  - # living
  - # deceased

④ Discussion:

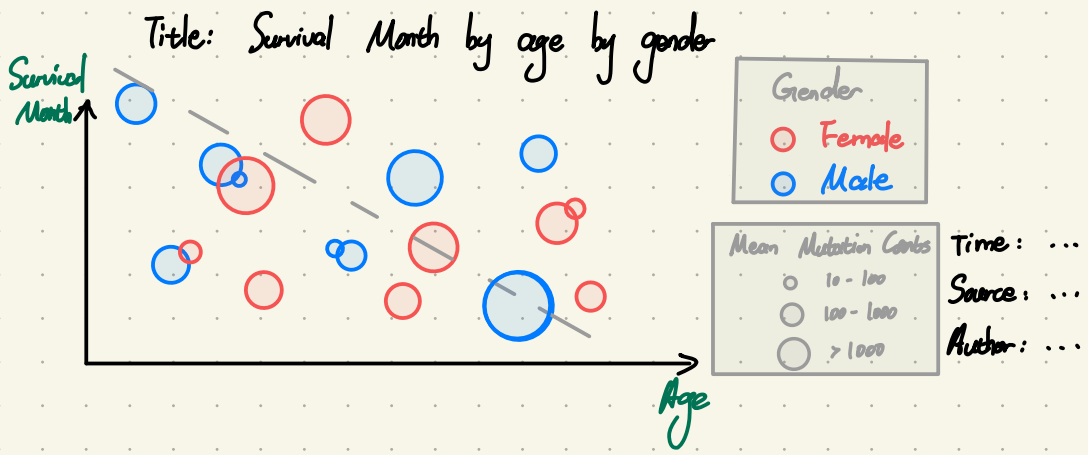
Advantage: could also demonstrate the trend of lethality as people age for each type of cancer.

Disadvantage: could belie the fact of susceptibility of different cancer types : the total number of each cancer is hidden.

Design 9 – Scatter Plot: Survival Month by age by gender, topped with the mutation counts (mean across all timepoint)

① Layout & Information

deceased ↑



② Focus

Goal: discover the correlation between survival month and age.
and see if gender & mutation count have a bearing on this relationship.

Emphasis: display the general trend of correlation between survival month and age through a regression line.

③ Operate:

↳ Hover and display: tooltips for each point include
- Age
- Survival Month
- Gender
- Mutation count at each timepoint.

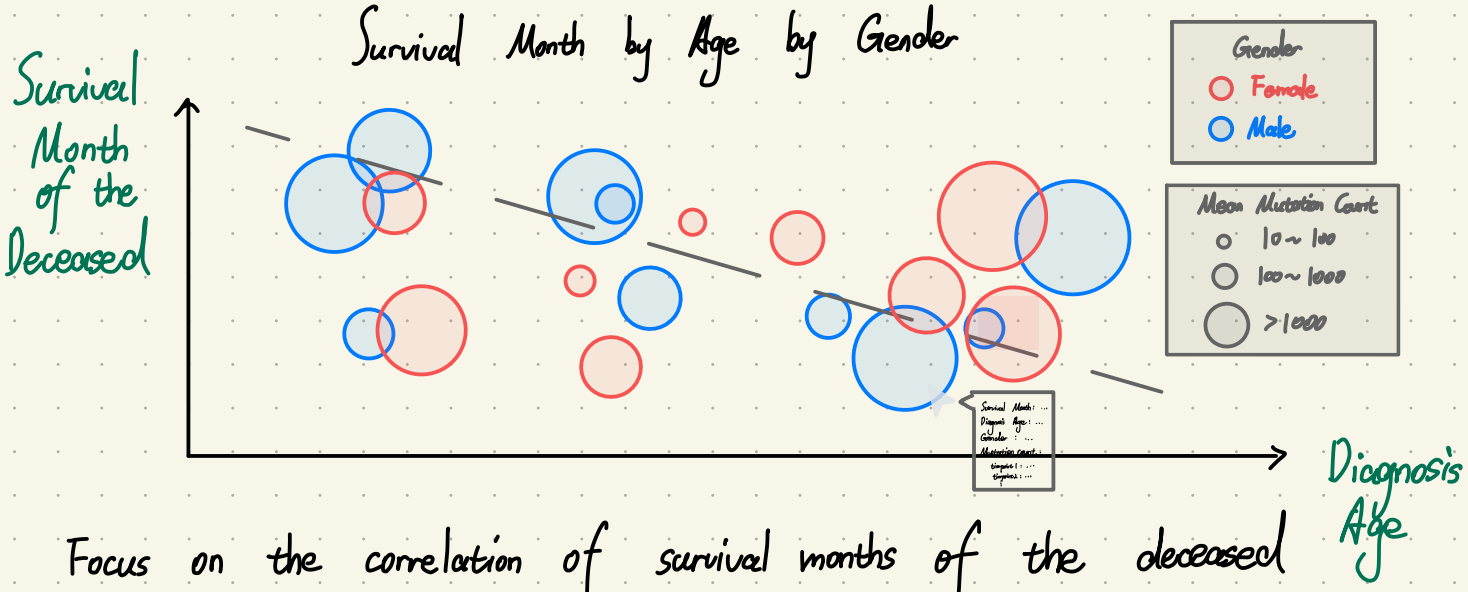
④ Discussion:

Advantage: give access to individual medical history along with the overall trend

Disadvantage: if dataset becomes larger, there may be intense overlappings that is hard for viewers to access individual data.

Sheet 5 Final Design.

① Design:



Focus on the correlation of survival months of the deceased and their diagnosis age. (Support Zooming)

② Math and Algorithm:

Algorithm: - the line is given by a linear regression that minimizes MSE: $\sum_{i=1}^{n} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$ where y_i are the survival month of the deceased and x_i are the diagnosis age.

- The mean mutation count is computed by grouping by patientID and compute the average mutation count for all timpint. The corresponding bubble size is binned into 3 groups: $(10, 100]$, $(100, 1000]$, and $(1000, \infty)$.

③ Software Requirement: Python 3.7, Altair 4.1, Pandas 1.2.4, Numpy 1.20. and related dependencies.

④ Estimation of cost & Time:

It takes a programer less than an hour to build.

For a large dataset, the code should take less than 3 minutes to render the interactive plot.

⑤ Other Requirements:

A high resolution display screen, preferrably 2560 · 1600 or better.