

Carolyn Vilter
22.10.31

Three-Minute Overview

Since my initial data review earlier this month, data sharing limitations have led me to shift to a different topic. Going forward, I'll be analyzing health data: specifically, a detailed dataset I created over the course of six months in 2019 that captured 7 features of my health (including heart rate, blood oxygen level, and two lung function metrics) twice a day. At the time, I was just beginning to recover from a surprising bout of pneumonia, and I had learned that I had a lung condition that would soon require a significant period of treatment. The experience inspired me to keep close track of my recovery, and my background in data analysis inspired me to do so through rigorous data collection!

The project resurfaced when I was thinking of data sources that might lend themselves to interesting visualizations. Years ago, I had created a few basic line graphs in excel which lived in my health spreadsheet and automatically updated as new data came in each day. Looking back at the graphs, I was struck by the obvious trends, the likelihood of associations between variables, and the potential for visualization to bring context and clarity. The dataset has different dimensions that make it interesting: personal yet dry numeric facts, contrasted with external medical reference points, coupled with human context (like my qualitative records of relevant milestones and feelings, which I noted alongside my daily data). Analyzing it will certainly be illuminating to me, might honestly be useful to pulmonologists, and will hopefully also convey to the average person a little bit of the story of my experience.

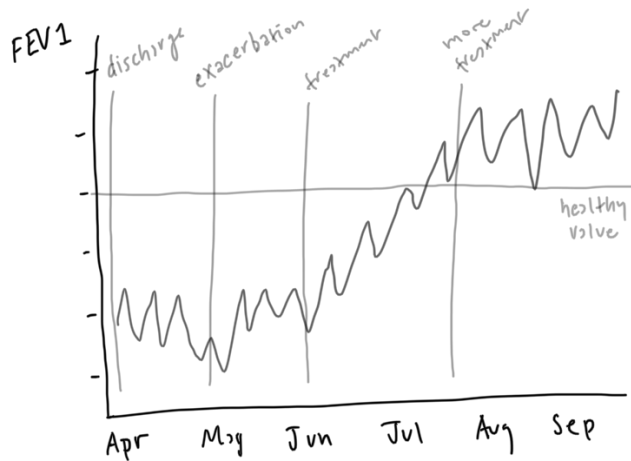
Possible Charts

One high-priority visualization is a line graph of FEV1 (forced expiratory volume, the metric that improved most dramatically) over time

Annotations would provide useful context:

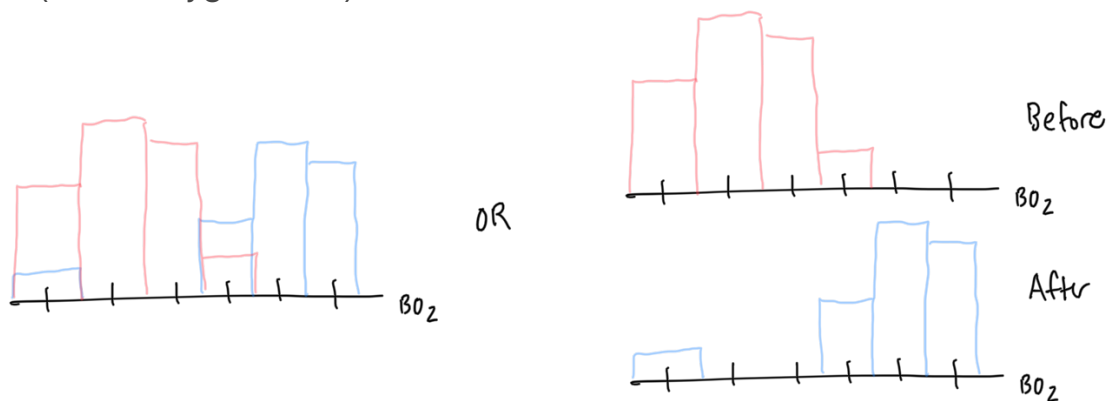
- Vertical lines denoting important dates: return from hospital, beginning of treatment, etc.

- Horizontal line indicating threshold for "healthy" (80% of predicted) values
- At select points on the line: notes from my qualitative dataset on how I was feeling / what I was doing at the time (static or possibly interactive, e.g. hover text)



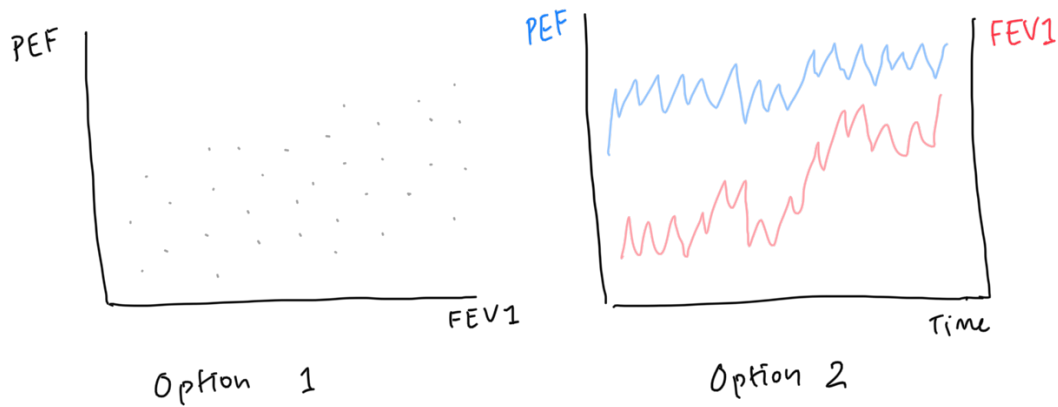
Another approach that would show off meaningful discontinuities in the data:
Two histograms arranged on the same x-axis - one for data before vs. one for data after the day I began treatment - to show the difference in the distributions

- Candidate variables: FEV1 again, PEF (peak expiratory flow), BO2 (blood oxygen level)



Visual representation of the correlation (or lack thereof) between two metrics (likely FEV1 and PEF)

- Option 1: Scatterplot with one metric on the x axis and one on the y axis
- Option 2: Line graph with time on the x axis and two y axes representing the two metrics



A very meta chart:

Bar chart of the intensity of my qualitative data collection over time

- Rationale: as I improved over time, the number of supplemental notes I collected on my feelings and impressions (alongside the objective data points) dropped off significantly
- Suggests that their frequency was a proxy for my general wellbeing

