MAY 3, 2024

HEART DISEASE MORTALITY CAUSES

BASED ON METADATA AND PATIENT OUTCOMES

LOUPE, PAUL
DATA VISUALIZATION

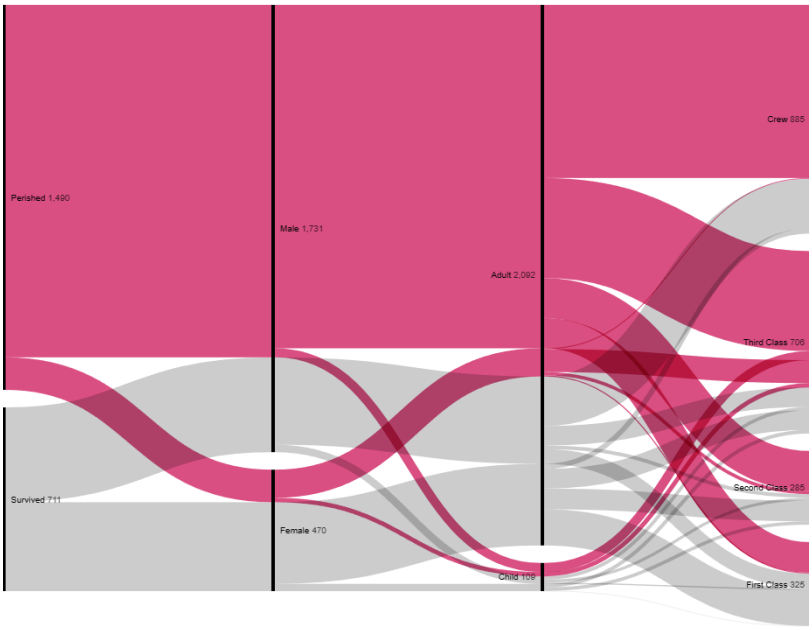
Intro

This project was dedicated to learning more about a disease that my father-in-law has and is currently undergoing a procedure for a heart transplant. I wanted to create a visual similar to what was shown in the UpSet video, where we could easily select and make determinations from the data in excel. The dataset itself was sourced from Kaggle [Cardiovascular Disease Dataset (kaggle.com)](kaggle.com).

The purpose of this visual was initially to discover what characteristics cause higher mortality rates in heart disease related events. However, this program also has the potential to be used as a template for viewing datasets that I plan on incorporating into my workflow and sharing the template on GitHub when finished.

Design Process

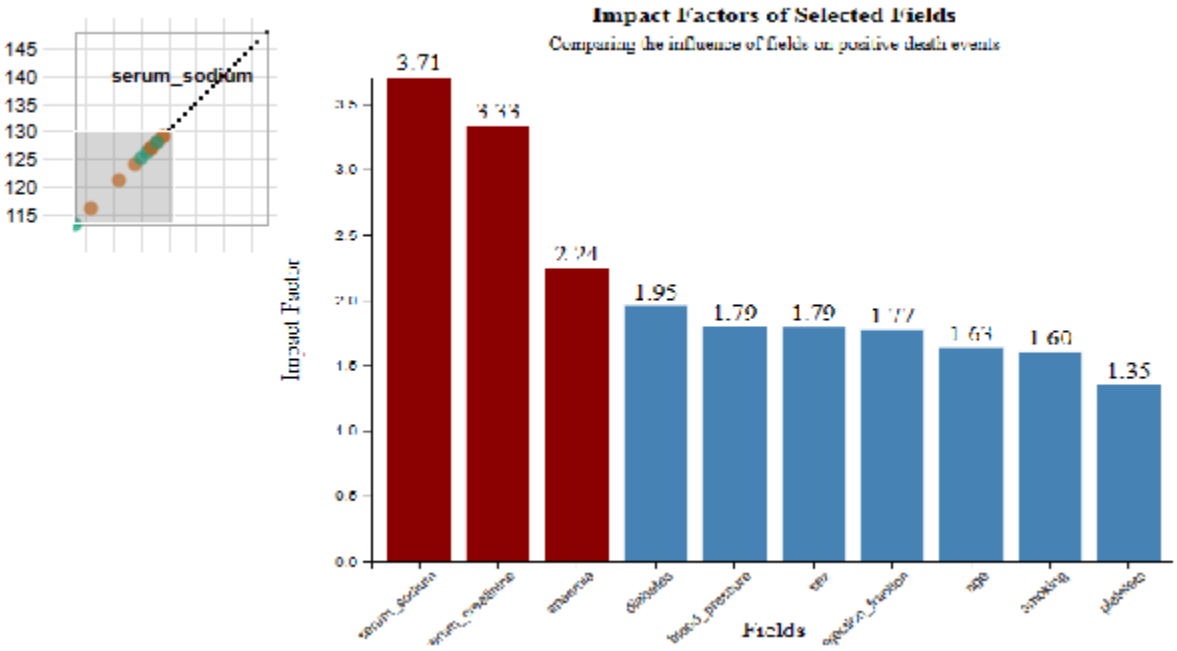
My initial idea was to create a parallel set that shows positive cases and their causes, but it wasn't interactive enough. This would be interesting as it shows good information and looks good as well, but I wanted to have advanced functionality to really be able to dig into our dataset.



So, I looked around at templates and discovered the "Brushable scatterplot matrix". This allowed us to plot all of the data in a matrix, select clusters of data, and see how that data is affected by the selection.

While it is extremely powerful, it is also a bit messy and overwhelming, but I decided to go with this based on the results that I was able to immediately see. For example, while I was messing around with the selections I discovered something that was completely against my bias, what I've heard from doctors, and what my father-in-law is currently doing. This is that sodium levels in the blood,

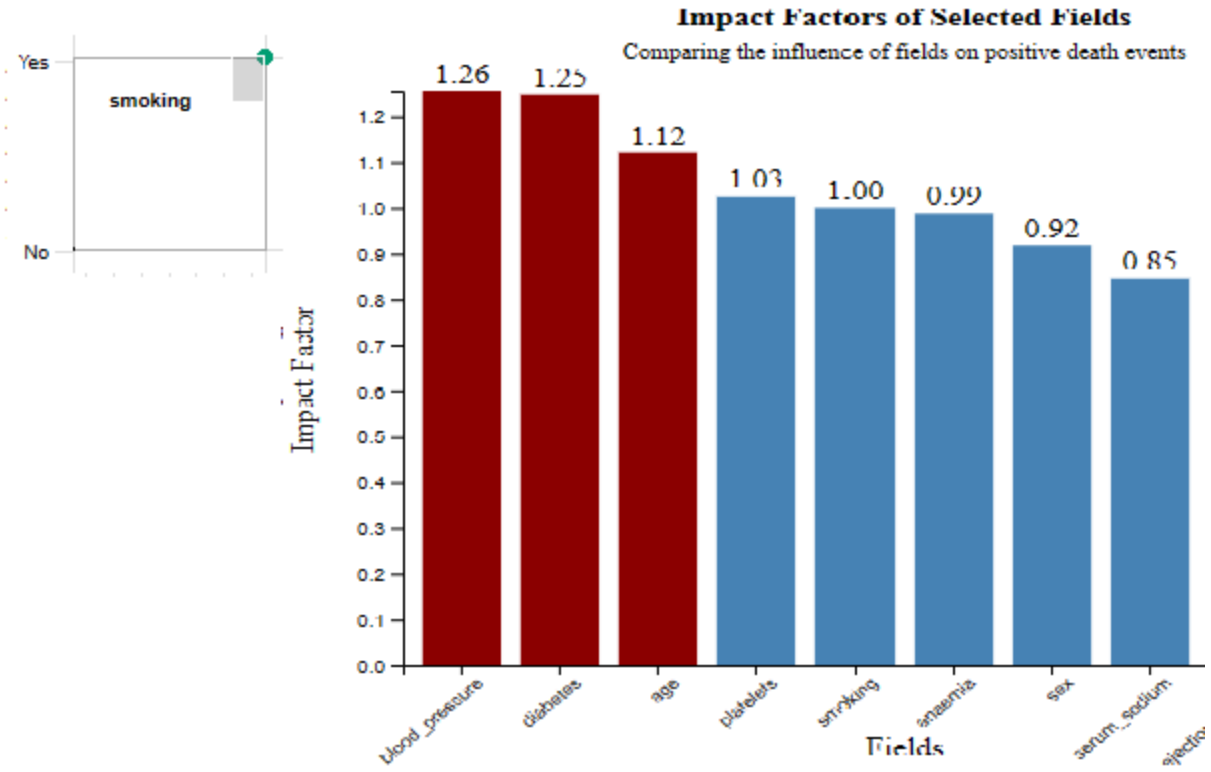
when lower, actually has a very high mortality rate. This goes against what I've heard where doctors encourage patients with heart disease to lower their sodium levels.



Meaningful thresholds were easily visible and verifiable upon selection. For non-death events Age <80, for death-events creatinine > 2, ejection fraction < 30, sodium levels <135.

There was a potential to reduce the amount of matrices shown here in order to clean up the overall look, but I decided against it as a functionality choice to explore our data.

While the visual does a great job of showing us which characteristics are responsible for certain events, I also decided to add in an interactive bar chart that maps highest impact factor from left to right. This allows us to see for sure which characteristics are most highly impacted by our selections without doubt. This also revealed meaningful correlations between certain characteristics, that by themselves, weren't as meaningful. For example, smoking is not as impactful as I thought it would be, but when combining it with high blood pressure and diabetes, you can easily see the negative impact.



Impact factor was chosen as it shows us the increased risk of a death event. Anything over 1 shows increased risk, and under 1 shows a decreased risk.

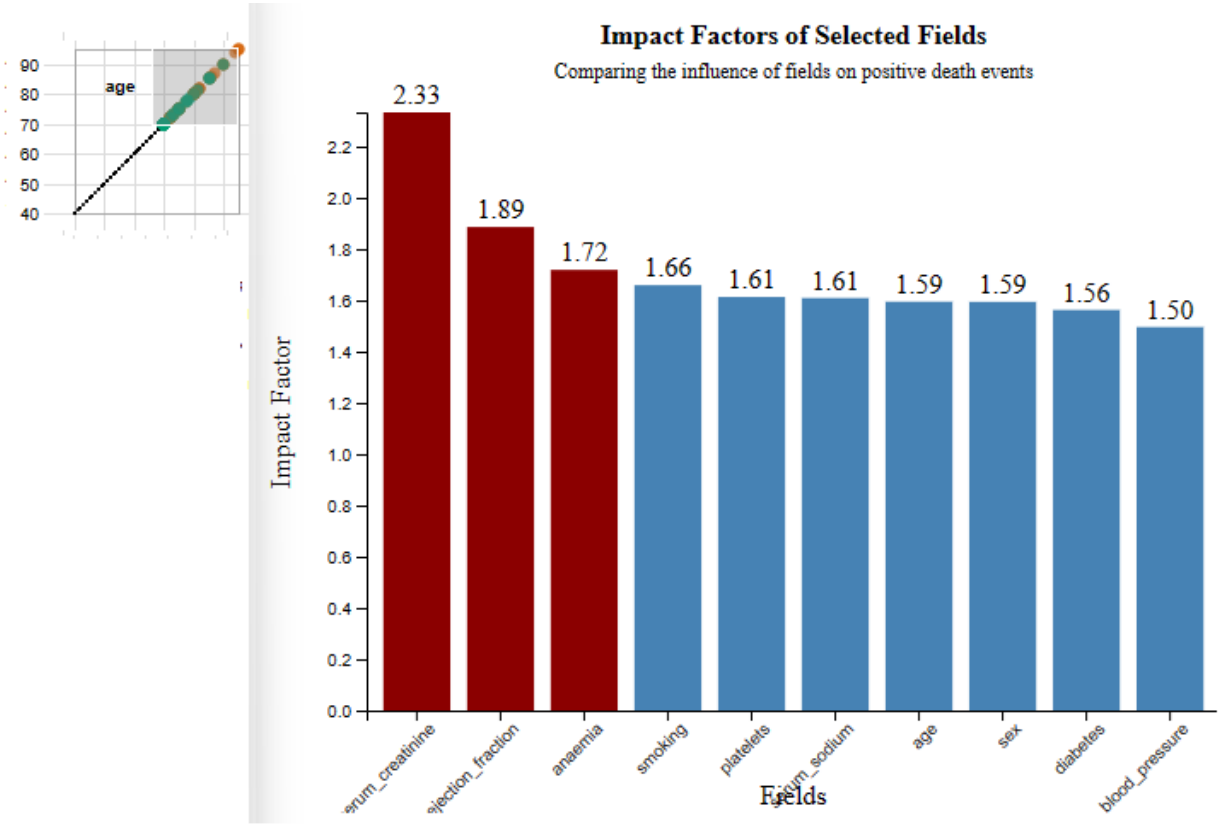
The color scheme is not beautiful, but meaningful as green is a non-death event and orange is a death event. Originally green and red, but considerations were taken for those who are colorblind.

Conclusion

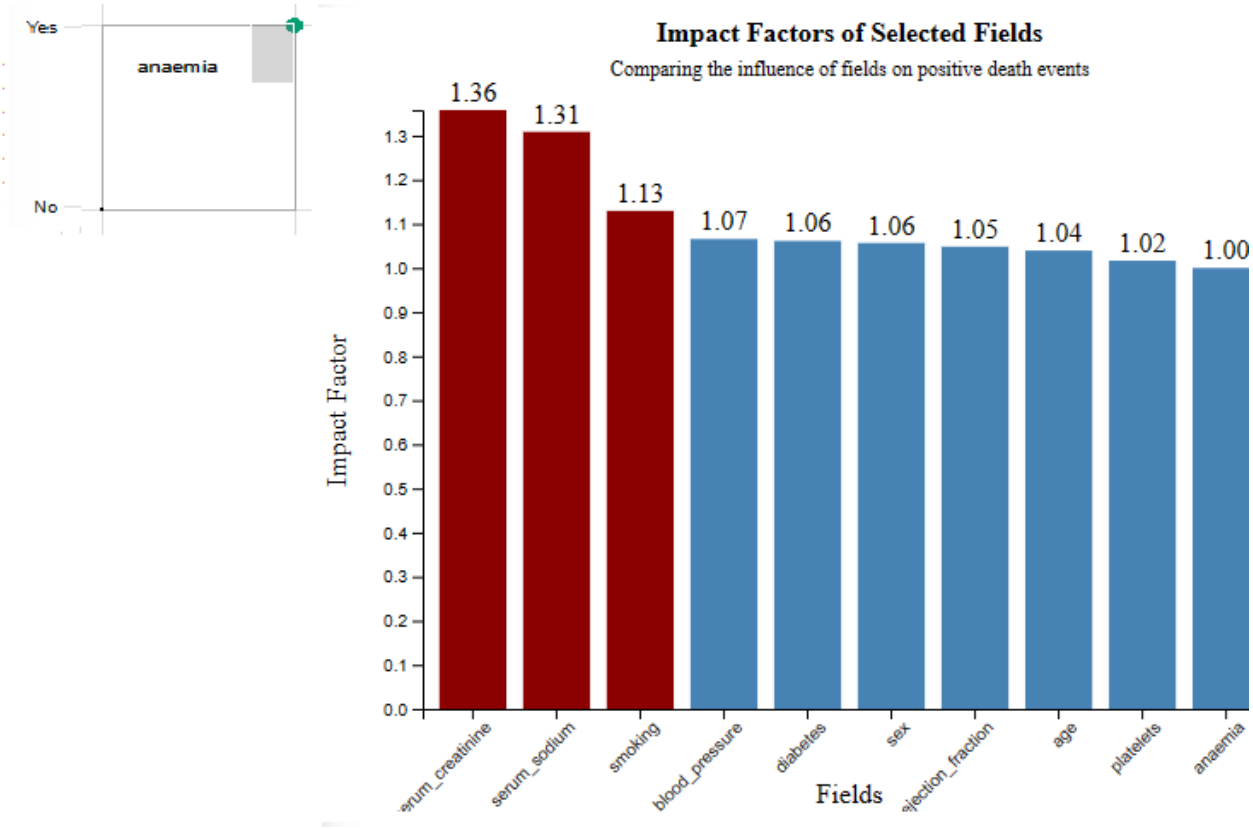
The meaningful information that is readily accessible from this interactive visual “Brushable scatterplot matrix” along with the interactive bar chart that shows the impact factor of the selected data is a powerful tool that we can use to quickly find meaning in data to further prepare our datasets for processing with machine learning, deep learning or other similar applications. For more findings please see the selections and impact factors below.

Visual Snapshots

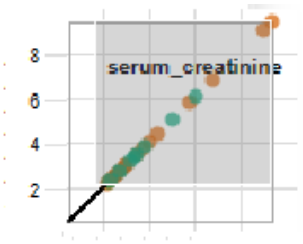
Here are some of the meaningful findings from exploring the data using the interactive visual. Feel free to visit the website and try these selections and others for yourself.

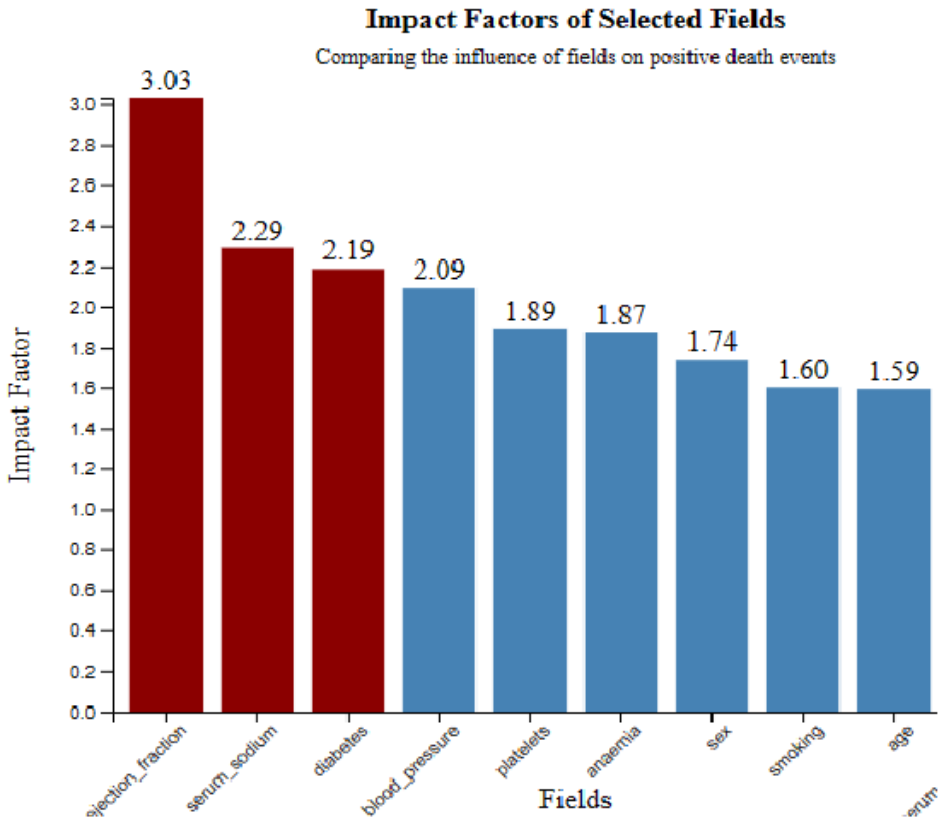


Here we look at impact factors for death events at ages over 70.

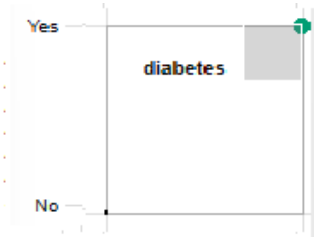


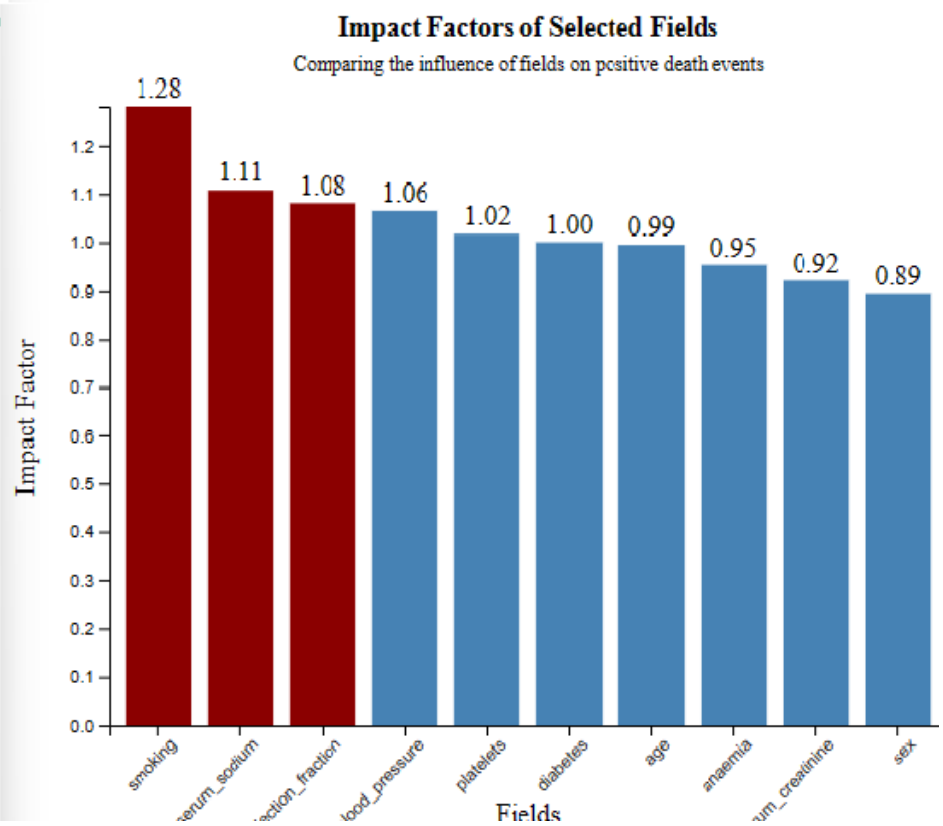
Here are impact factors of death events if the patient has anaemia.

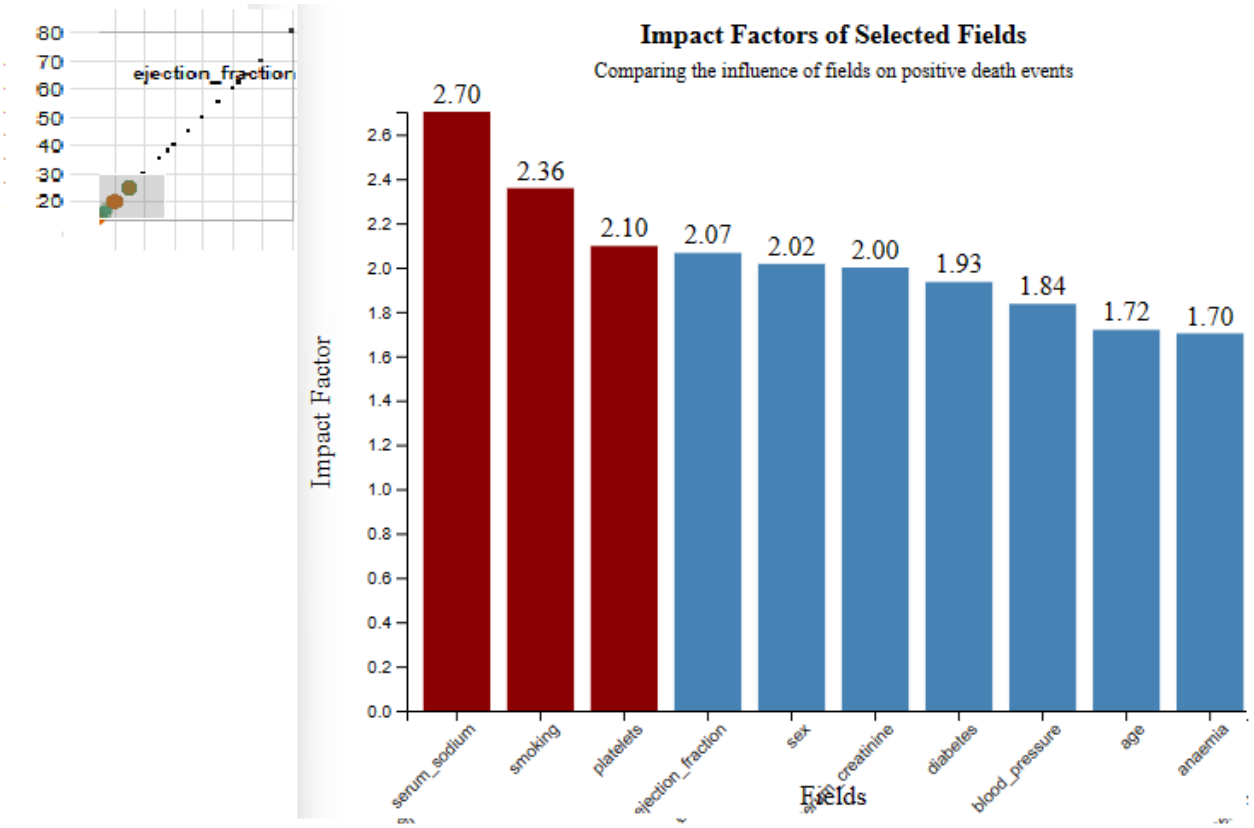




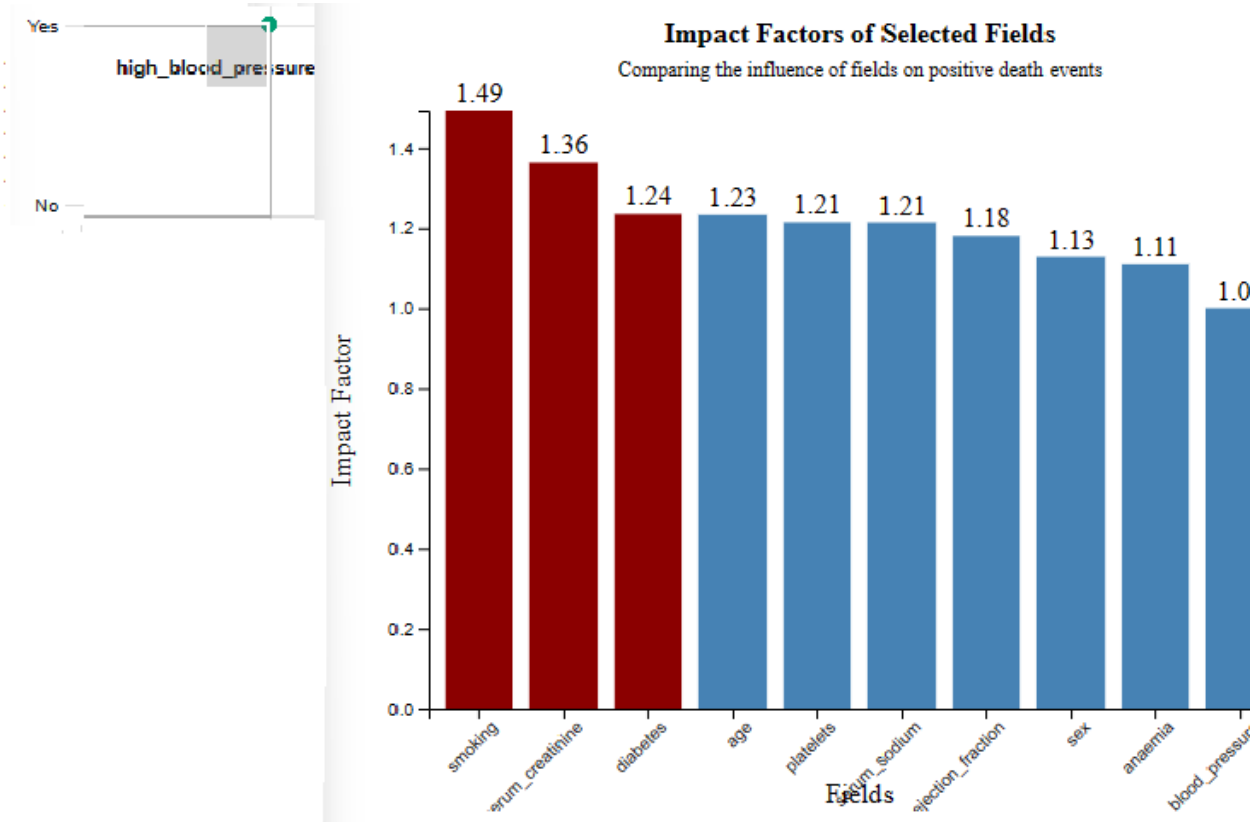
Here we look at impact factors in death events if the patient's serum creatinine is over 2.

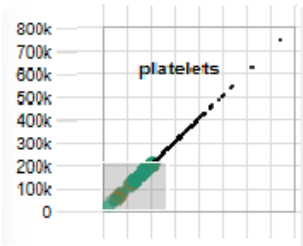


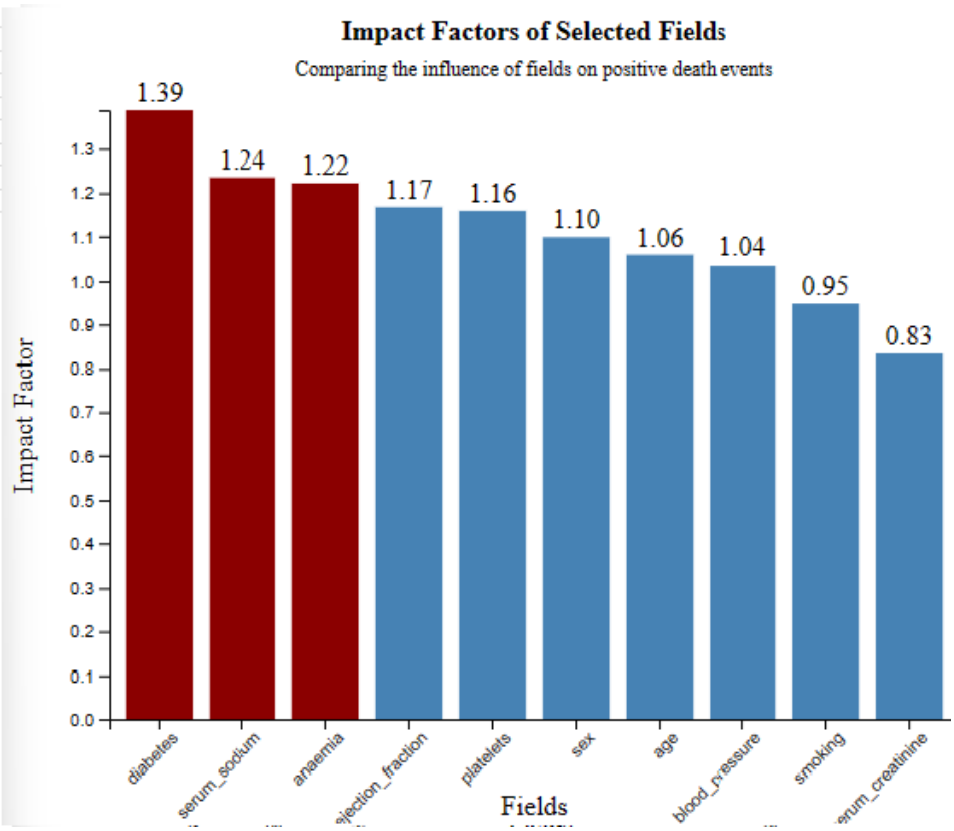




Here are impact factors in death events for patients with ejection fractions below 30.







Here we look at impact factors in death events for patients with platelet counts below 200k.