

# MIMIC-III UMAP Result Overview

Zad Chin

Advisor: Professor Finale Doshi-Velez

Harvard Data to Actionable Knowledge Lab (DtAK)

---

## Overview of Data:

- MIMIC-III data: 10184 patients, 7 vital features included, two type of dataset (with/without outliers)
- Data Constraints: Min 48 hours of stay, First 48 hours of data, Min age of 18
- 7 features: Heart rate, Creatinine, Hematocrit, Fi02, Mean Blood Pressure, Respiratory Rate, Sodium

## Overview of Standardization Methods:

- ***Discrete-Valid:*** Standardized against valid range of features in discrete manner: 0 if within valid range, 1 if over valid range, -1 if below valid range
- ***Discrete-Normal:*** Standardized against normal range of features in discrete manner: 0 if within valid range, 1 if over valid range, -1 if below valid range
- ***Continuous-Valid:*** Standardized against valid range in a continuous manner
- ***Continuous-Normal:*** Standardized against normal range in a continuous manner;

## Our findings so far:

- Different standardizations leads to different clusterings in UMAP analysis, and each of the clusters are separate by a distinct feature.
  - Discrete-Valid: One cluster with few outliers
  - Discrete-Normal: Cluster discriminated by Creatinine and Hematocrit
  - Continuous-Valid: Cluster discriminated by Fraction Inspired Oxygen
  - Continuous-Normal: Just one cluster
- We also have tested out against TSNE and get similar result for each of the standardization methods.
- DBSCAN does not provide useful analysis, even coupled with PCA to help reduce dimensions. Our hypothesis is that the high dimensionality of the data causes DBSCAN to perform poorly.

### Valid and Normal Ranges for each vital

#### **Valid Range:**

Vitals	Valid Low	Valid High
Heart Rate	0	350
Respiratory Rate	0	300
Creatinine	0.10	60
Hematocrit	0	75
Fraction Inspired Oxygen	0.21	1.00
Mean Blood Pressure	14	330
Sodium	50	225

Source: [Github Mimic-pipeline](#)

#### **Normal Range:**

Vitals	Normal Low	Normal High
Heart Rate	60	100
Respiratory Rate	12	18
Creatinine	0.59	1.35
Hematocrit	37	52
Fraction Inspired Oxygen	0.21	0.5
Mean Blood Pressure	70	100
Sodium	135	145

Source: I got it from Sonali last year

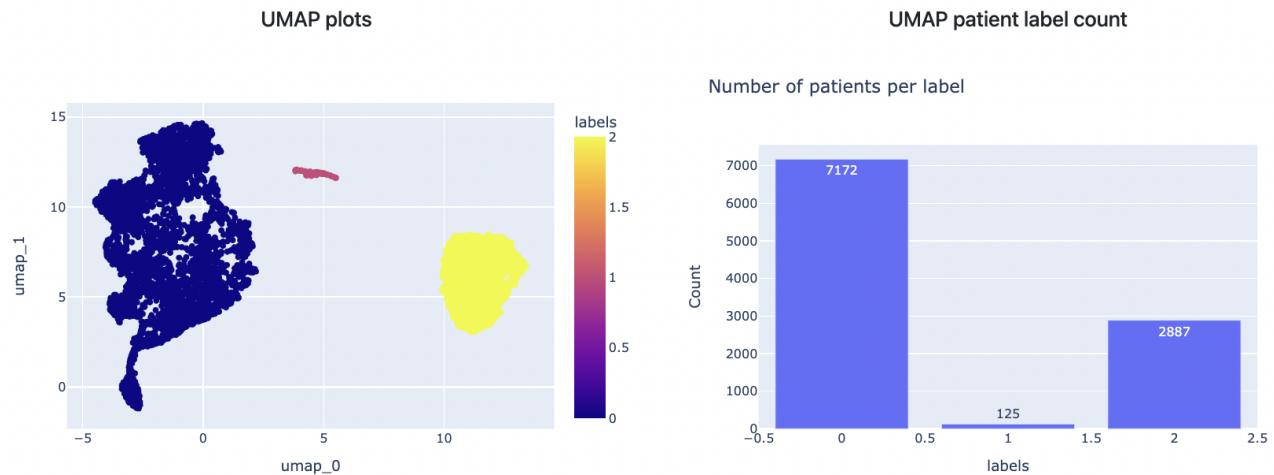
### Experiment 1:

Data: MIMIC with Outliers

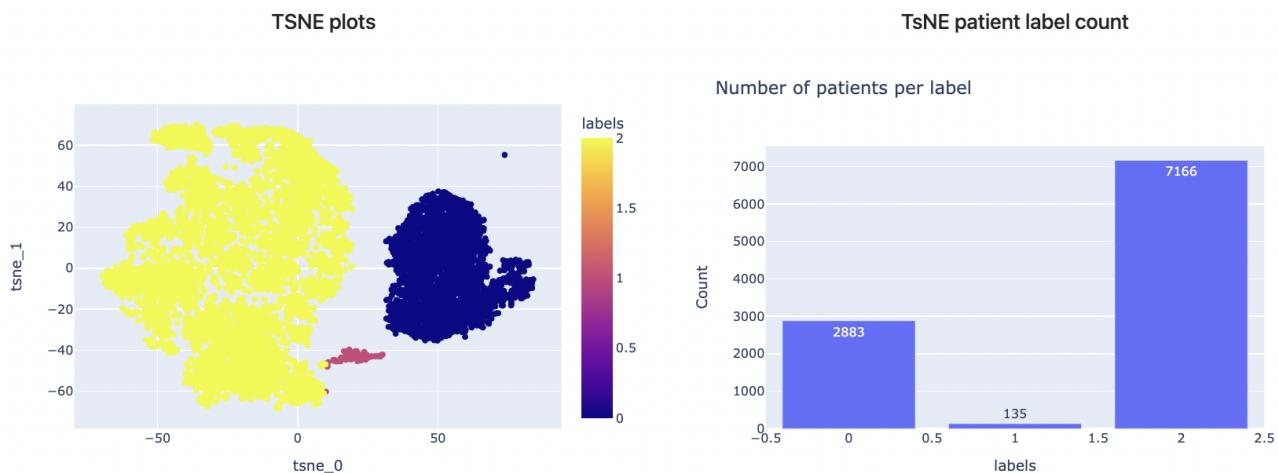
Standardization Method: Continuous-Valid

### Result:

- The result is consistent for TSNE analysis, as in all the labels are coherent (ex: label 0 in UMAP is the same as label 2 in TSNE)
- The third cluster, which made up of 125 patients are outliers in the data.
- The discriminant between Label 0 and Label 2 is Fraction Inspired Oxygen (Refer sample analysis for Experiment 2 below) / Consistent with TSNE

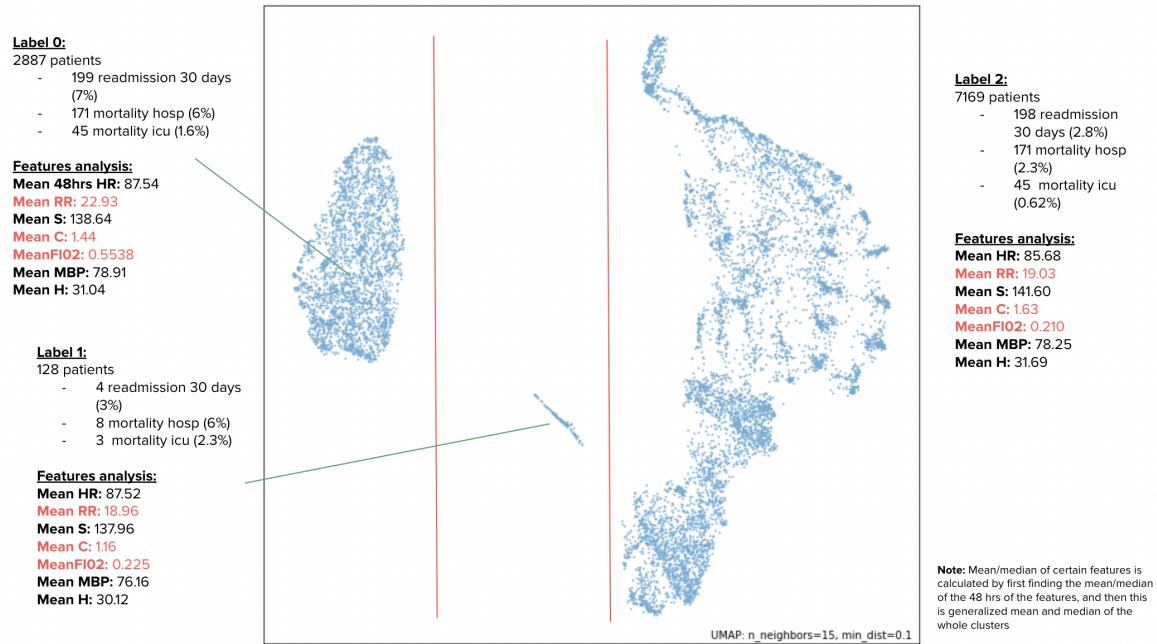


**Figure 1:** UMAP result (Continuous-Valid/With Outliers)

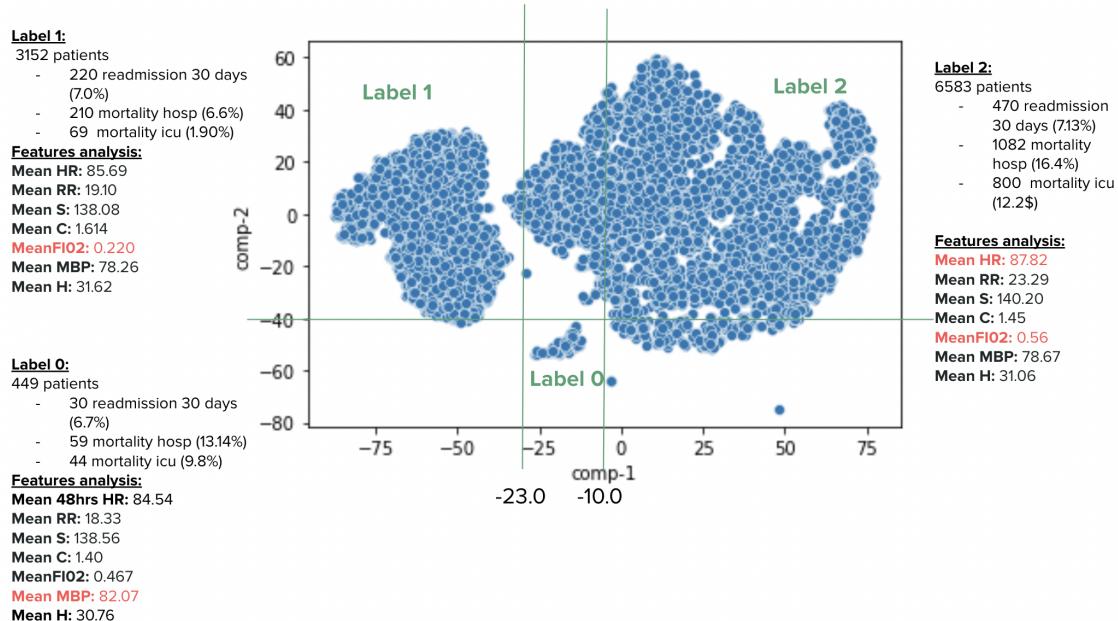


**Figure 2:** TSNE result (Continuous-Valid/With Outliers)

## UMAP Brief Analysis:



## TSNE Brief Analysis:



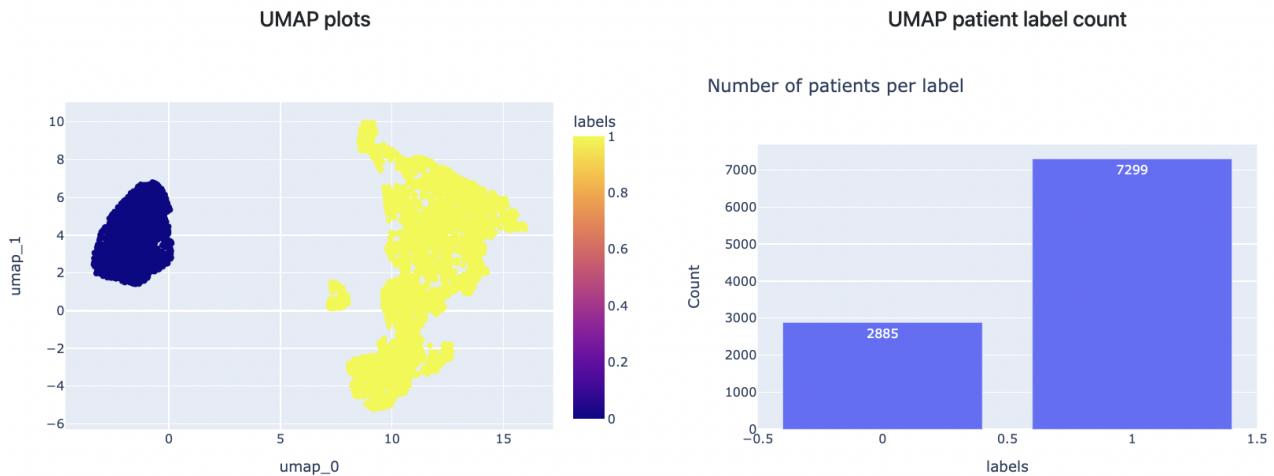
## Experiment 2:

Data: MIMIC without Outliers

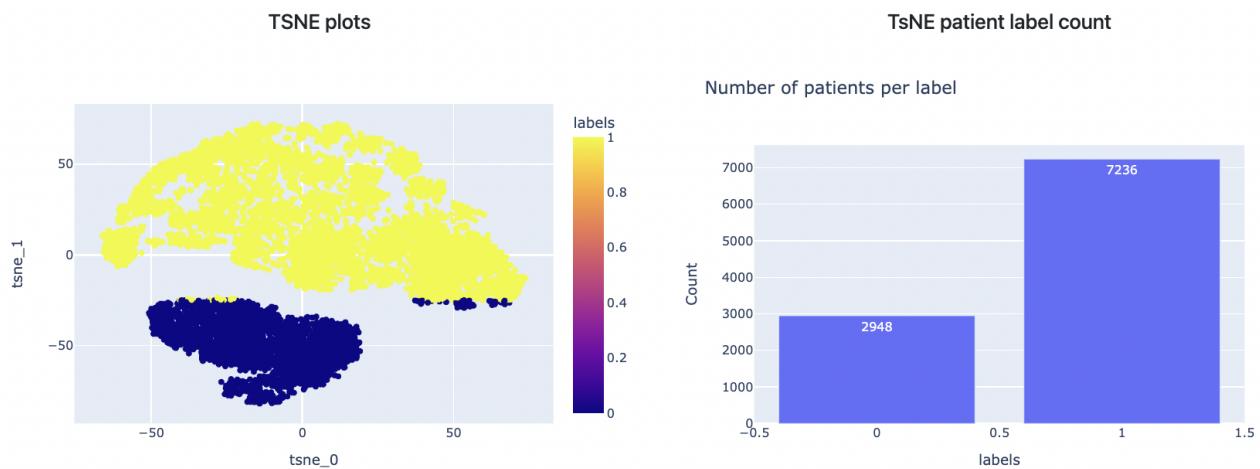
Standardization Method: Continuous-Valid

### Result:

- The result is consistent for TSNE analysis, as in all the labels are coherent (ex: label 0 in UMAP is the same as label 2 in TSNE)
- The discriminant between Label 0 and Label 2 is Fraction Inspired Oxygen  
(Refer sample analysis below)

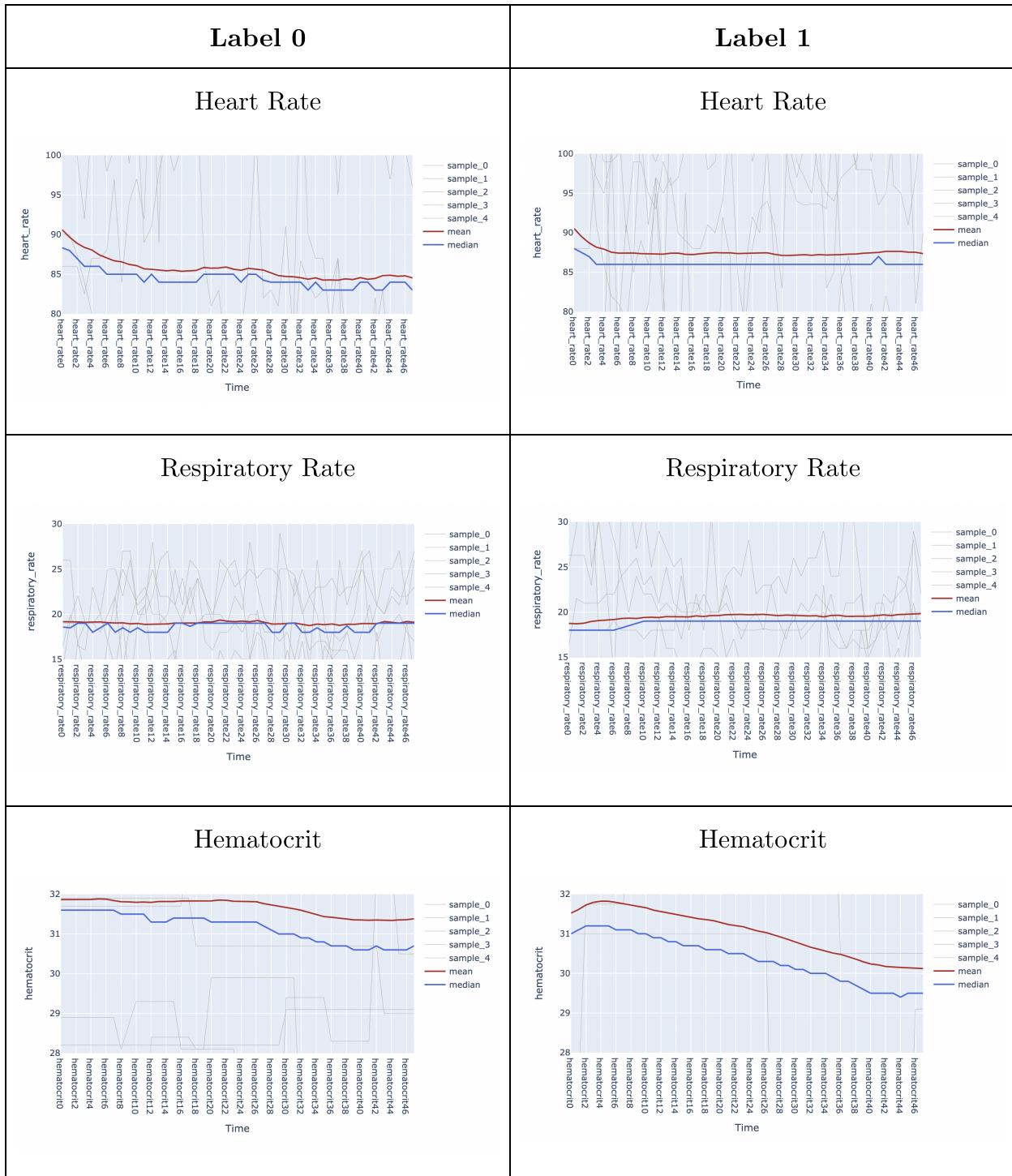


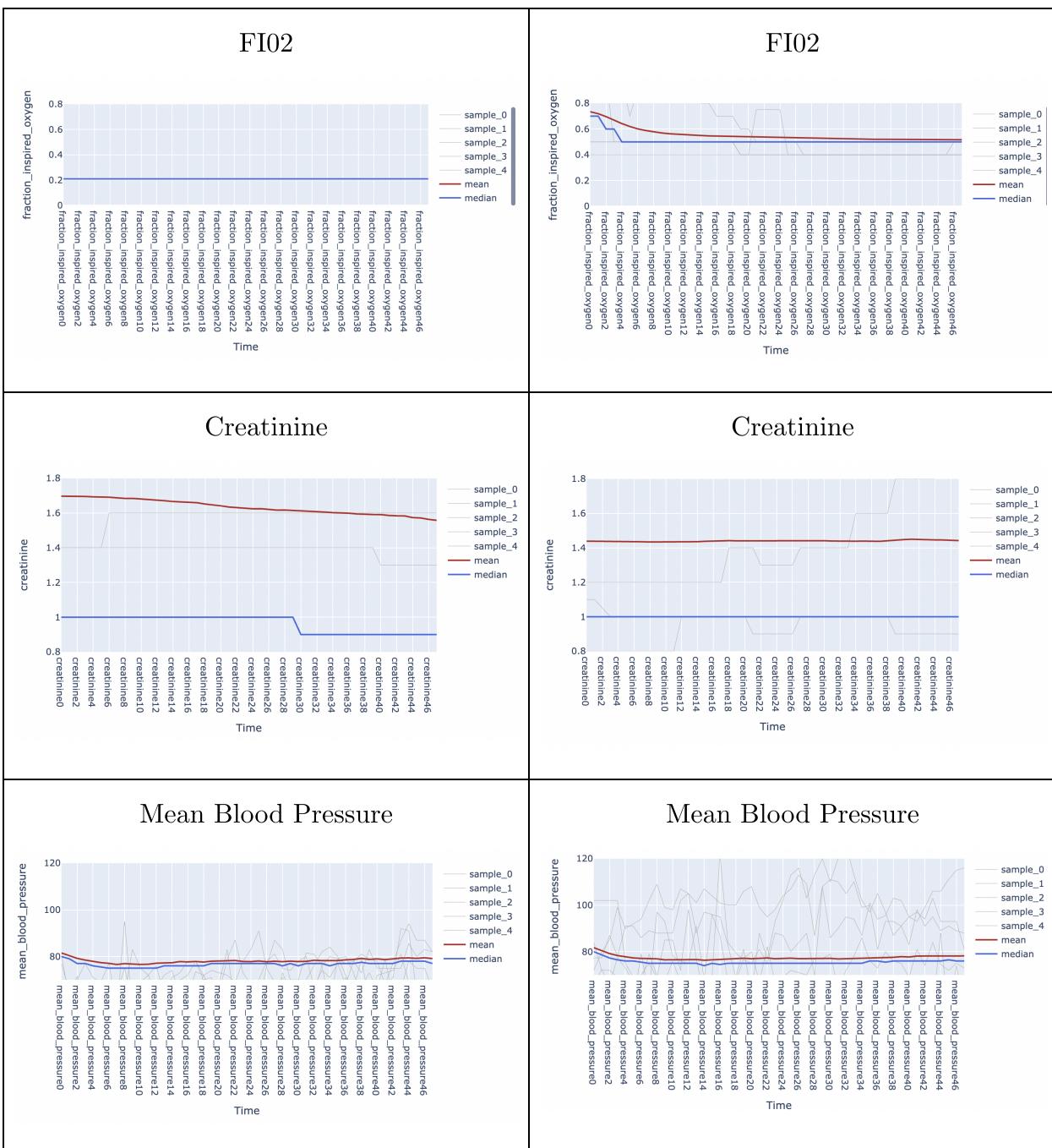
**Figure 3:** UMAP result (Continuous-Valid/Without Outliers)

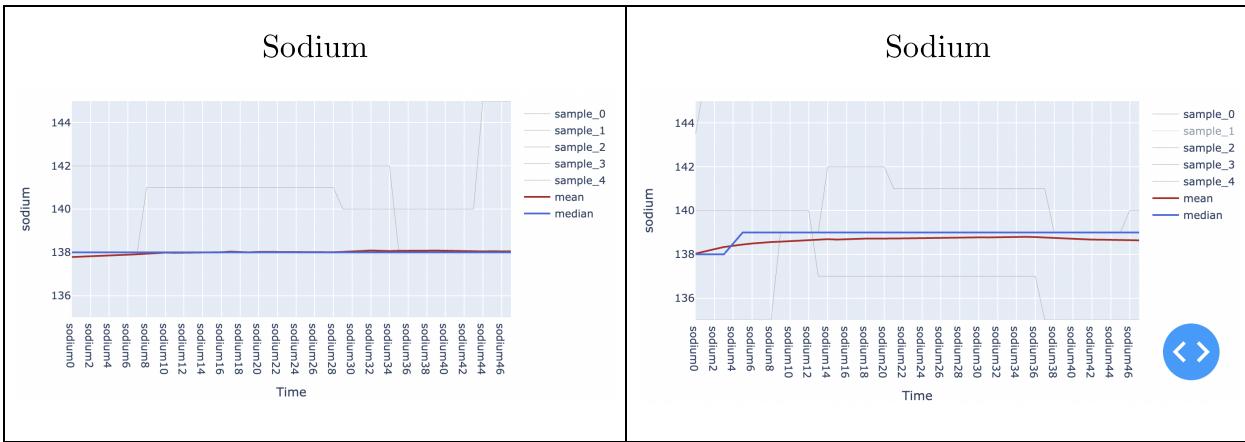


**Figure 4:** TSNE result (Continuous-Valid/Without Outliers)

## Sampling Analysis for Experiment 2:







We plotted the graphs against 5 samples, 10 samples, and 20 samples. The above examples are 5 samples. Feel free to contact me if you need the result from 10 samples or 20 samples.

**Experiment 3:**

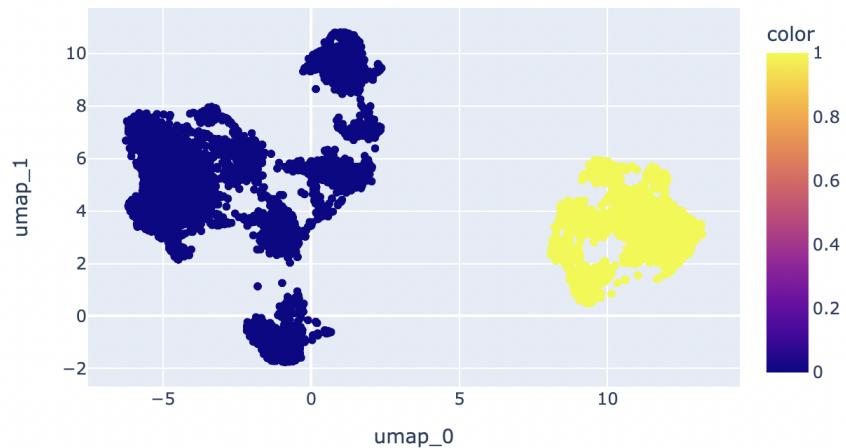
**Data:** MIMIC without Outliers

**Standardization Method:** Discrete-Normal

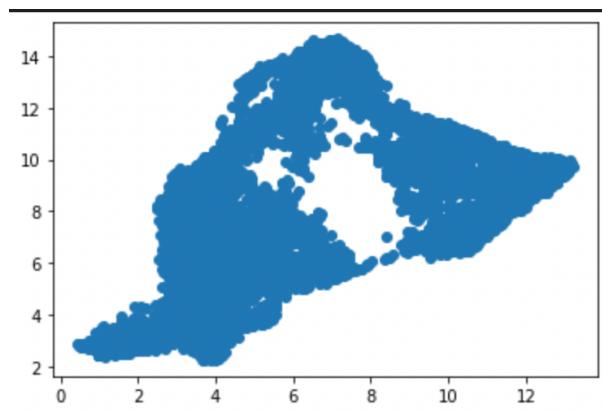
**Result:**

- The discriminant between Label 0 and Label 2 is Creatinine and Hematocrit.  
(Refer sample analysis below): Dropping both creatinine and hematocrit will result in one cluster (as figure below)

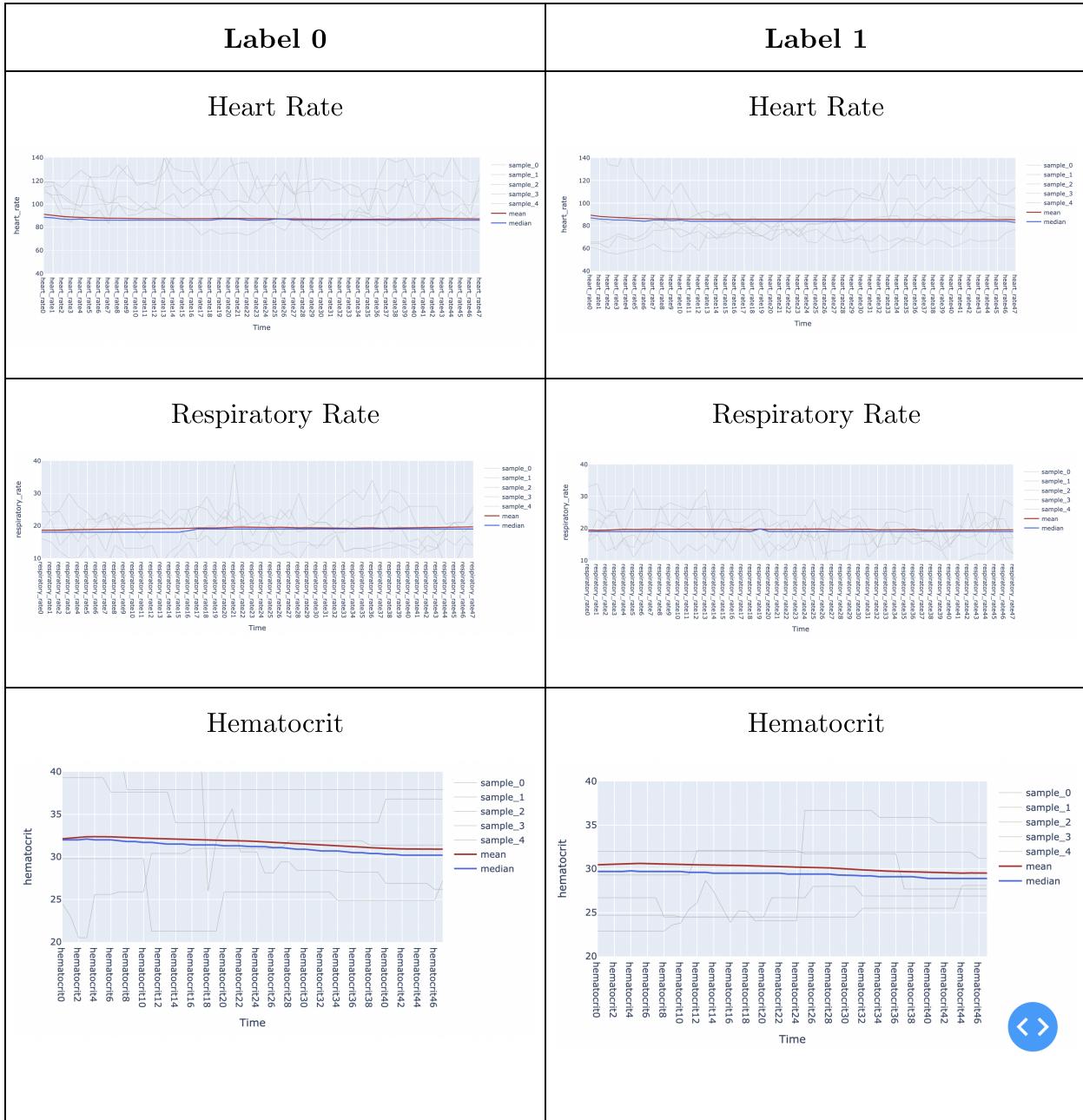
**UMAP result:**

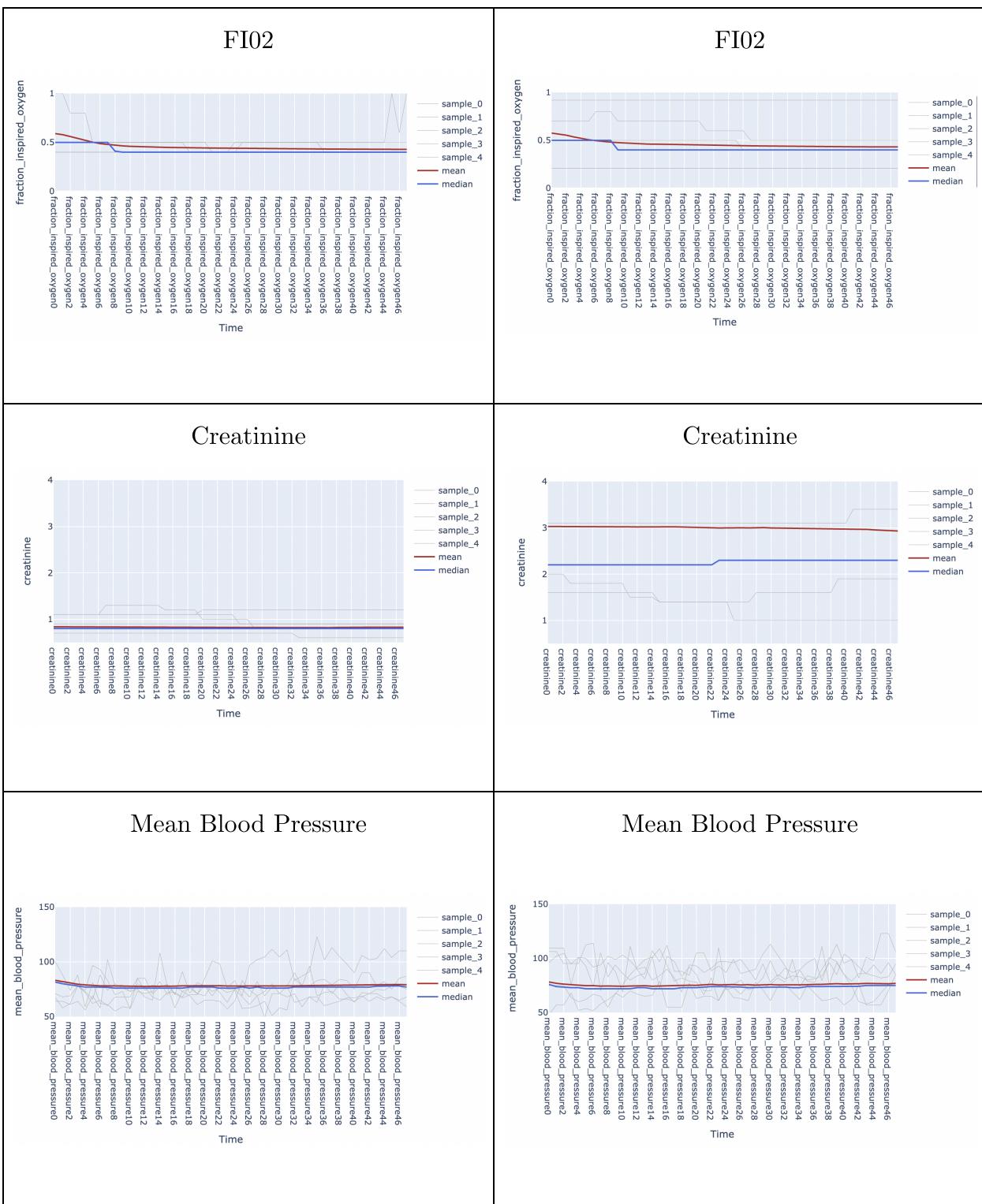


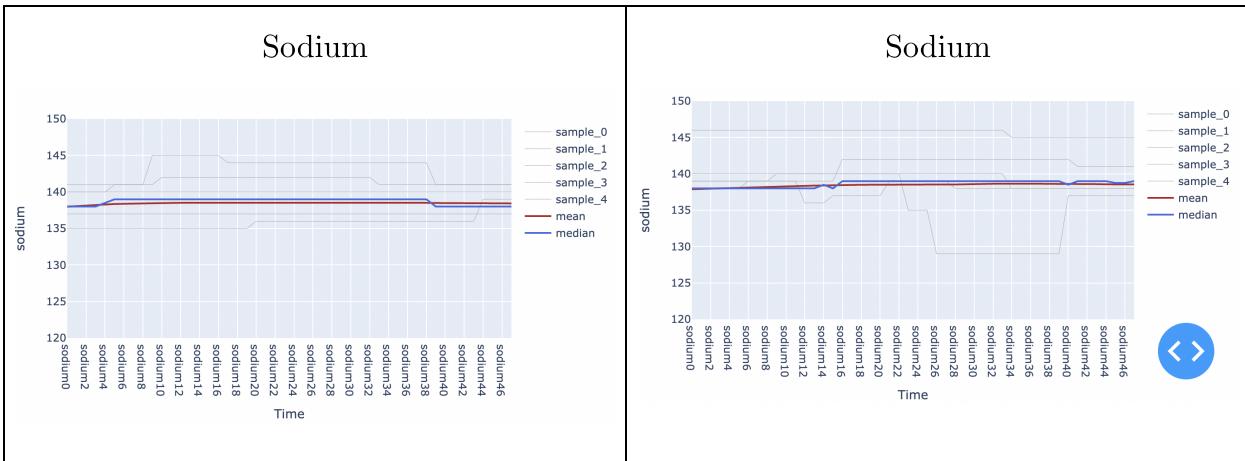
**UMAP Result after dropping Creatinine and Hematocrit:**



## Sampling Analysis for Experiment 3:





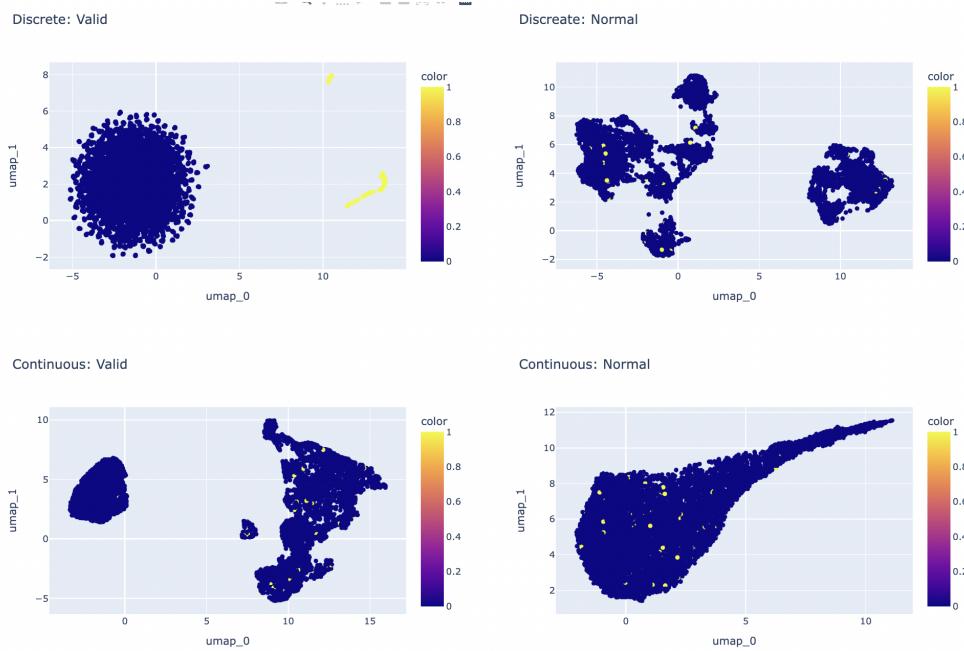


We plotted the graphs against 5 samples, 10 samples, and 20 samples. The above examples are 5 samples. Feel free to contact me if you need the result from 10 samples or 20 samples.

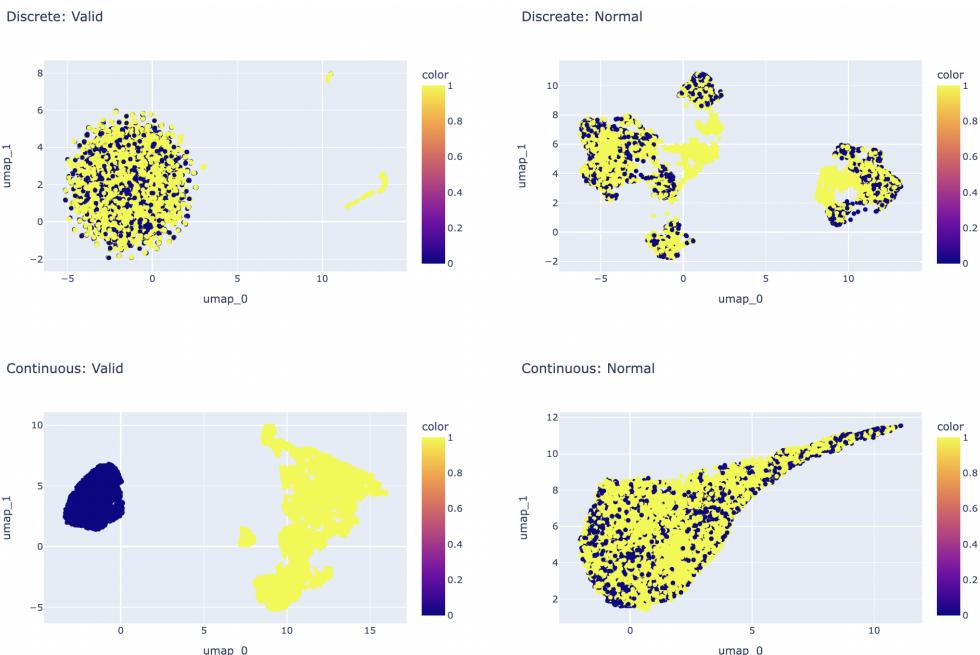


## Overview of different plotting result and their labelling:

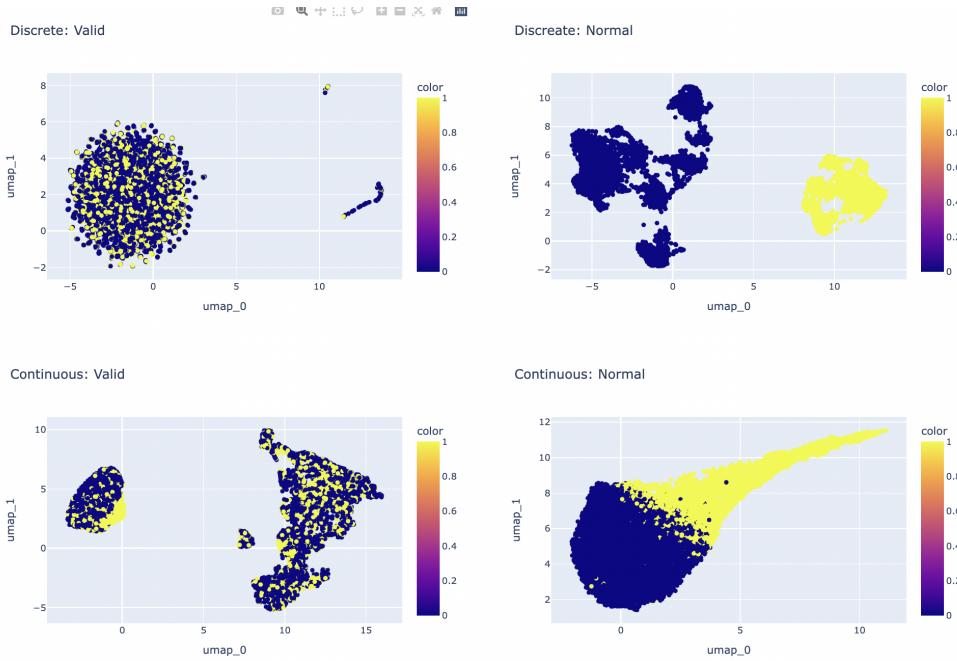
### Label Color: Discrete-Valid



### Label Color: Continuous-Valid



## Label Color: Discrete-Normal



## Label Color: Continuous-Normal

