# Survey

## Unsupervised Feature-Based Visualization Tool Applied on Medical Data

**Research Summary:**

The purpose of our study is to produce a new unsupervised feature-based technique for data visualization inspired from parallel coordinates and hierarchical clustering, specially dendrograms, in order to visualize large datasets effectively and interactively and extract correlations between different features or group of features of the dataset.

Our tool "Mirrored Dendrograms" displays two dendrograms placed in parallel where each is the result of clustering the data based on a feature or a group of features of interest to the user. The dendrograms are linked at similar internal nodes, or sub-clusters, taking into consideration similarity between the structure of clustered data and not only the data itself (on the leaves), to allow extracting the correlation between the chosen features. **In other words, the more the links, the higher is the correlation and vice versa**. Moreover, the tool enables user interaction to zoom in (show all patients and full dendrograms) and out (show clusters of patients at leaves when applicable) the data, depending on the needed level of details and presents the data to the user initially at a default zooming level which is a compromise between data similarity and granularity, meaning that it provides enough information and a good similarity result.

For the sake of this research, we have chosen patients migraine medical data (Electronic Health Records EHRs). The leaf nodes of the dendrograms represent individual patients identified by a number: patient 1 at dendrogram 1 is the same as patient 1 at dendrogram 2. When the visualization is zoomed out, the number becomes an indicator of the number of patients in the cluster. Patients are clustered based on a selected feature: dendrogram 1 reflects patients clustered based on feature 1 and dendrogram 2 reflects patients clustered based on feature 2. The purpose is to extract the correlation between feature 1 and feature 2 from the produced visualization.

**Question 1:**

Please specify your age group.

☐ Under 18 ☐ 18-24 ☐ 25-34 ☐ 35-44

☐ 45-54 ☐ 55-64 ☐ 65+

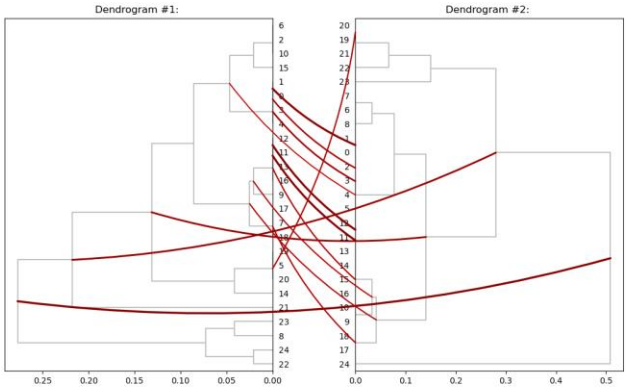
**Question 2:**

Please specify your profession.

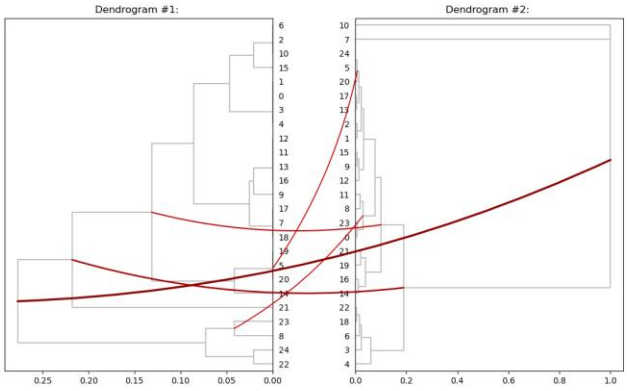
**Question 3:**

The images below show two visualizations of 25 patients produced by our tool “Mirrored Dendrogams”. The first image shows the visualization of two correlated features whereas the second image shows the visualization of two less correlated features. On a scale from 1 to 10, rate the visualization from the perspective of defining the correlation between the mentioned features.

☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10




Correlated features




Less correlated features

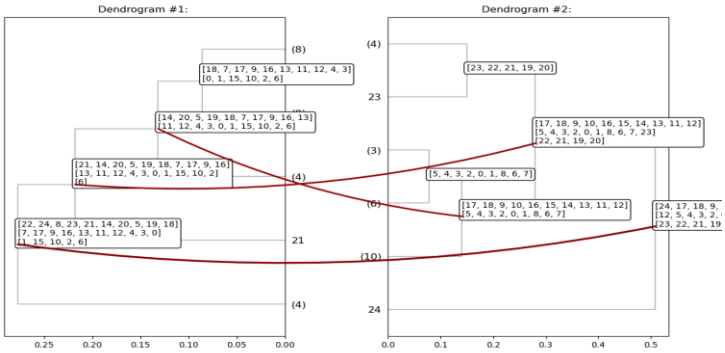
Additional Comments

**Question 4:**

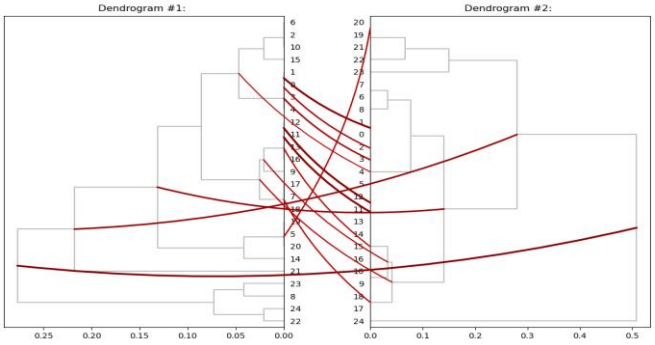
The first image below shows the visualization of 25 patients to extract the correlation between "Days of Migraine" and "Frequency of Abortive Treatment" as it appears "zoomed out" by default by our tool "Mirrored Dendrogams" to reduce complexity. On a scale from 1 to 10, rate this default zoomed out visualization from the perspective of defining the correlation between the mentioned features, its ability to reduce complexity, as compared to the full visualization in the second image.

P.S: The labels at the nodes are the actual patients' identification numbers.

☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10




Default Zooming Level



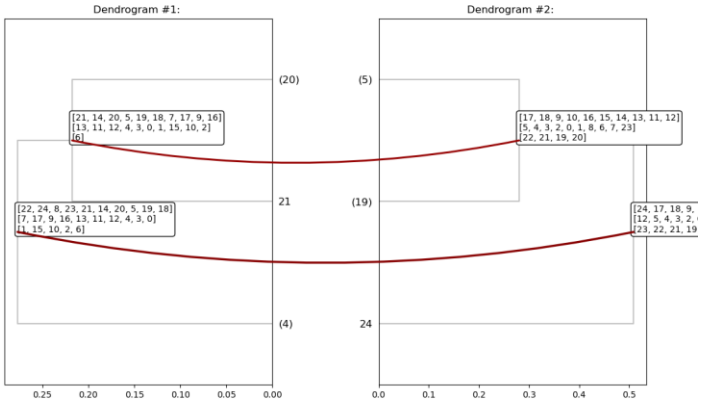

Full Zoomed in Visualization

Additional Comments

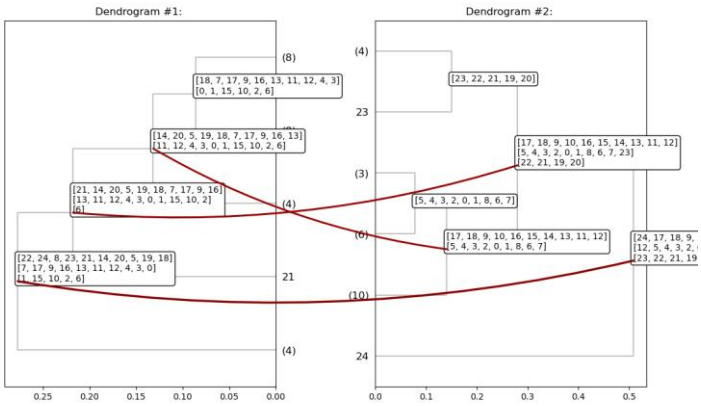
**Question 5:**

The images below show the visualization of 25 patients to extract the correlation between “Days of Migraine” and “Frequency of Abortive Treatment” at different zooming levels. Originally, the more users zoom in, the more details they get regarding clusters and their correlation. On a scale from 1 to 10, rate this variation of zooming and whether it is descriptive of the correlation value between the features with respect to the level of details required.

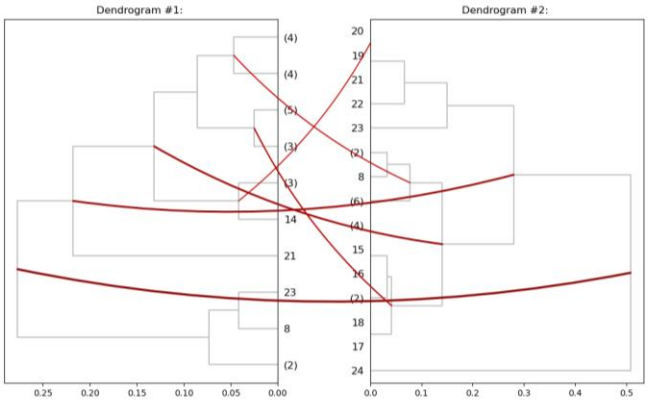
☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10



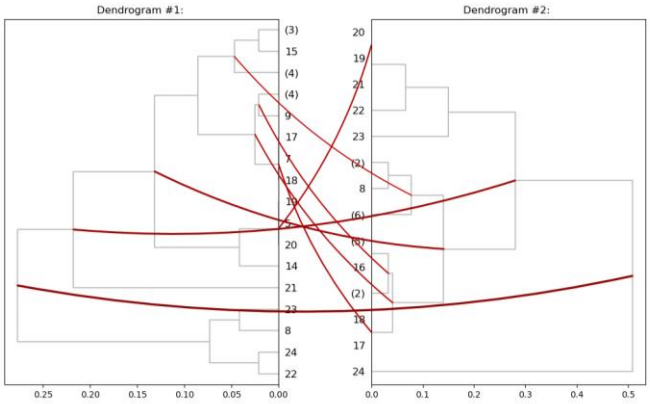

3 Elements vs 3 Elements



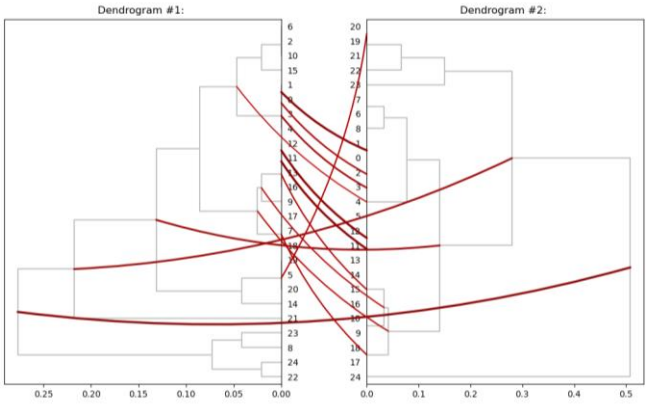

5 Elements vs 6 Elements




10 Elements vs 15 Elements




17 Elements vs 14 Elements



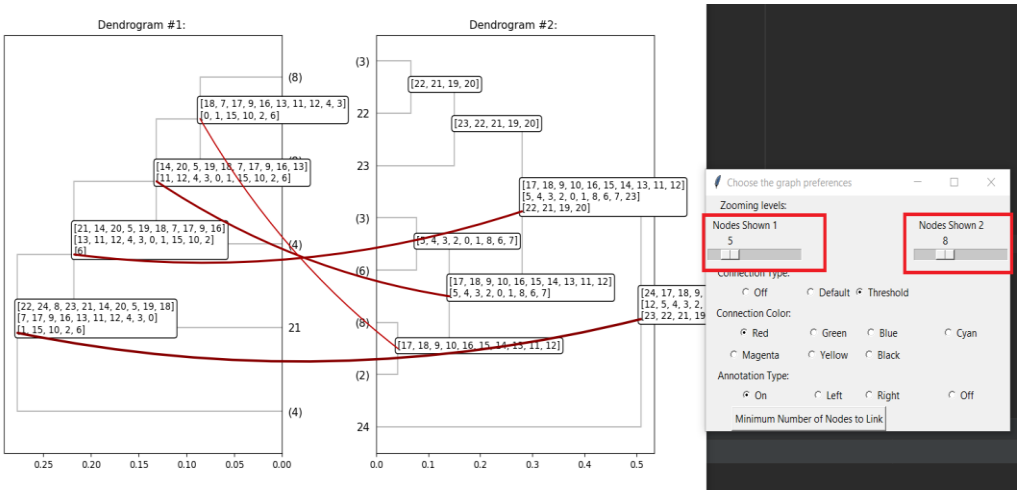

25 Elements vs 25 Elements

## Additional Comments

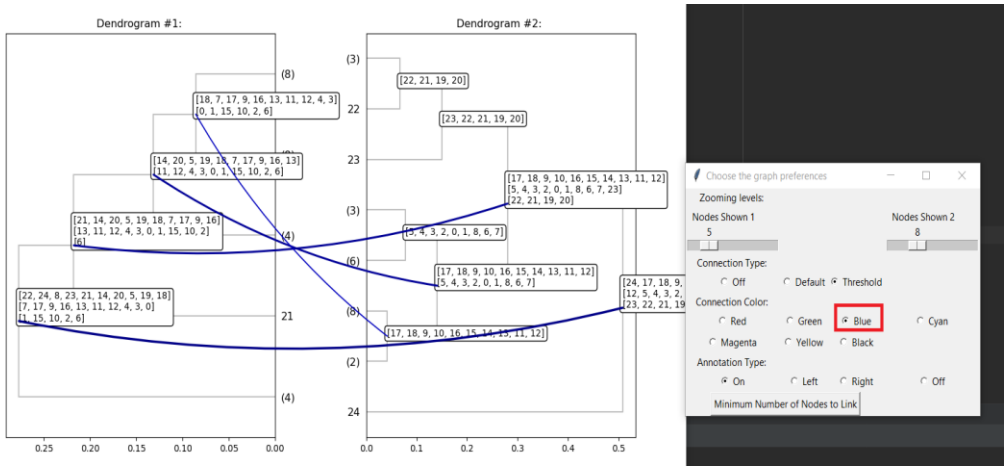
**Question 6:**

The images below show the level of interaction and control the user is able to do in the “Mirrored Dendrograms” tool. On a scale from 1 to 10, rate these interactive capabilities.

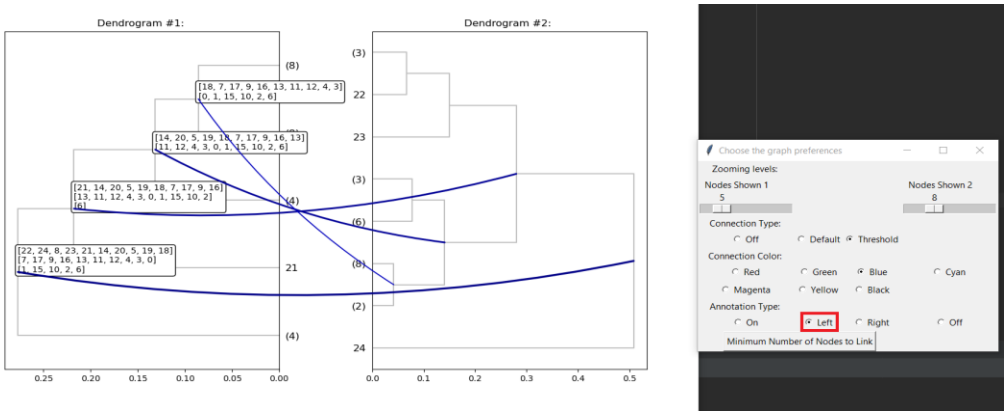
☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10



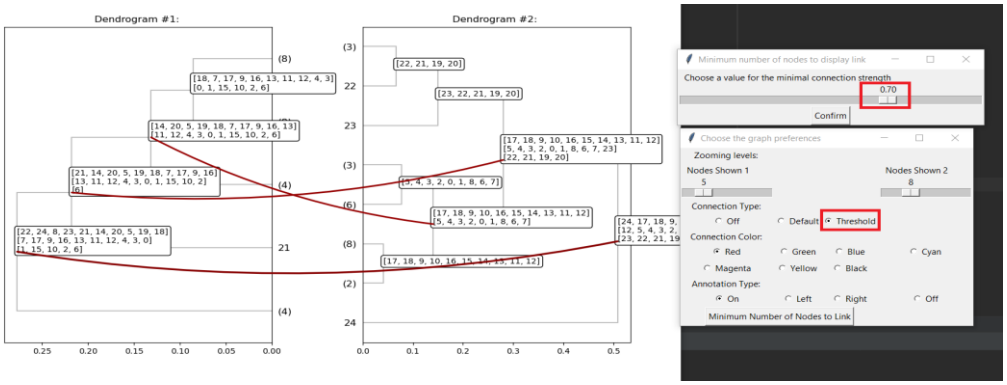

Varying Number of Nodes Displayed



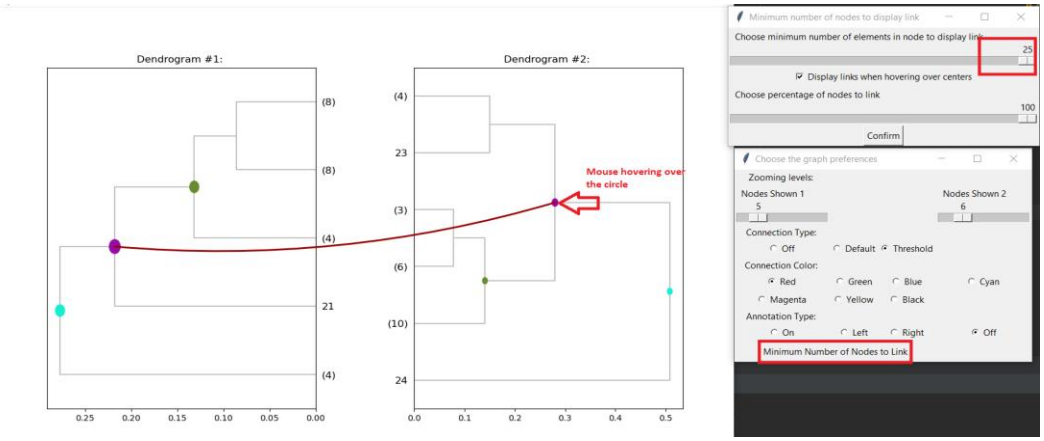

Varying Links Color




Varying Annotations




Varying Similarity Threshold




Replacing Links with Same Color Circles and Hovering Over a Circle to Display the Link

## Additional Comments

**Question 7:**

The images below show three different visualizations of 25 patients to extract the correlation between “days of migraine” and “frequency of abortive treatment”. The first image is produced by our tool “Mirrored Dendrogams” with internal links threshold of 0.5 (50% similarity or more). The second image is produced by “Tanglegram” visualization. The third image is produced by “Heatmap” visualization. On a scale from 1 to 10, rate each visualization in conveying correlation information between the two features.

Mirrored Dendrograms:

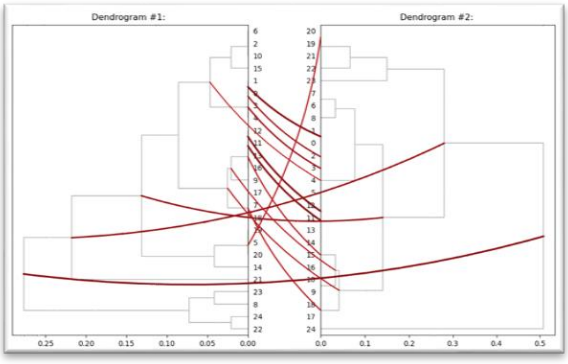
☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Tanglegram:

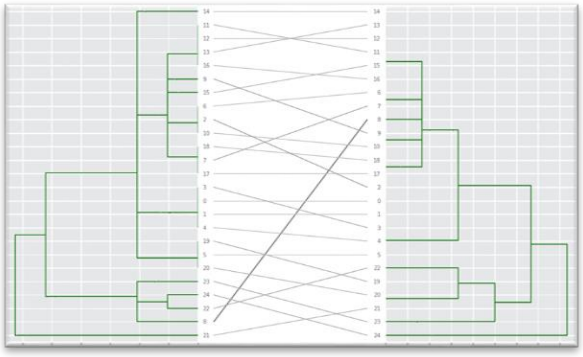
☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Heatmap:

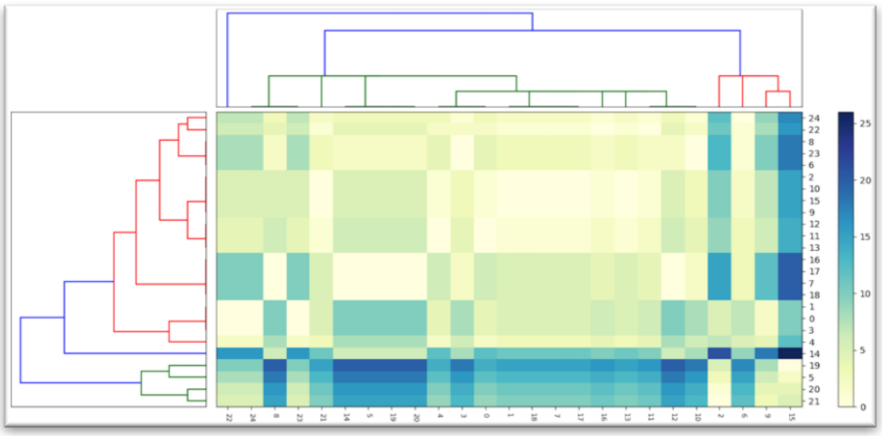
☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10




Mirrored Dendrograms



Tanglegram



Heatmap

## Additional Comments