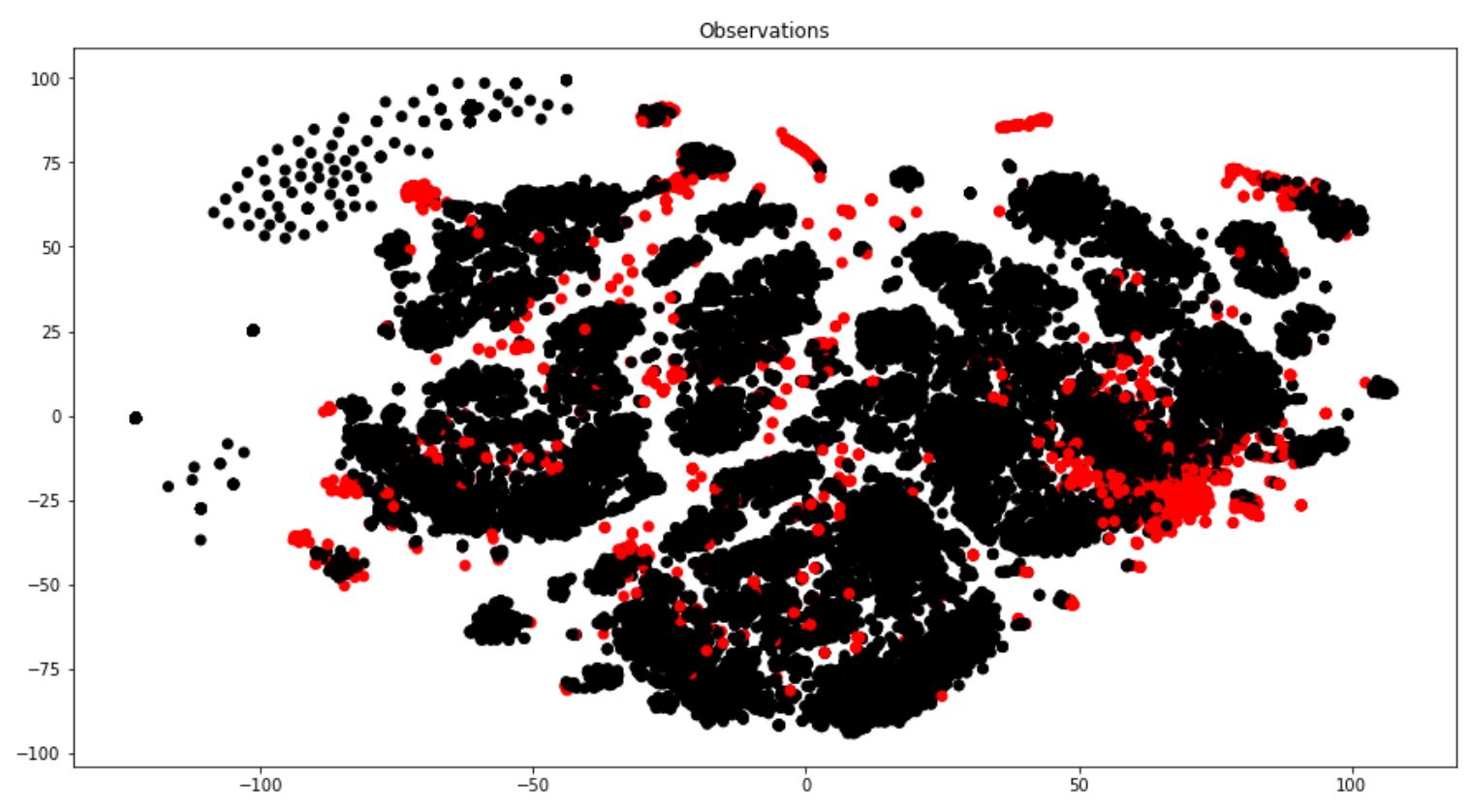
# **Appendix**

Visualization or dimensionality reduction of a high-dimensional data is a real challenge and undisputable necessity in data analysis. There is an opinion that in many practical problems, the high-dimensionality of datasets is an unnecessary exaggeration and proper projection into a low-dimensional space will successfully solve the problem without loss of information. Then, the problem is in the practical solution of the corresponding projections. Manifold Learning generalizes linear dimensionality reduction methods like PCA to be sensitive to non-linear structure in data in an unsupervised manner. The main idea of the projection is the preserving of “distances” between observations in low- and high-dimensional spaces. For visualization purposes, the natural projection is the 2-dimensional space where two observations have almost the same “distance” as in the high-dimensional space.

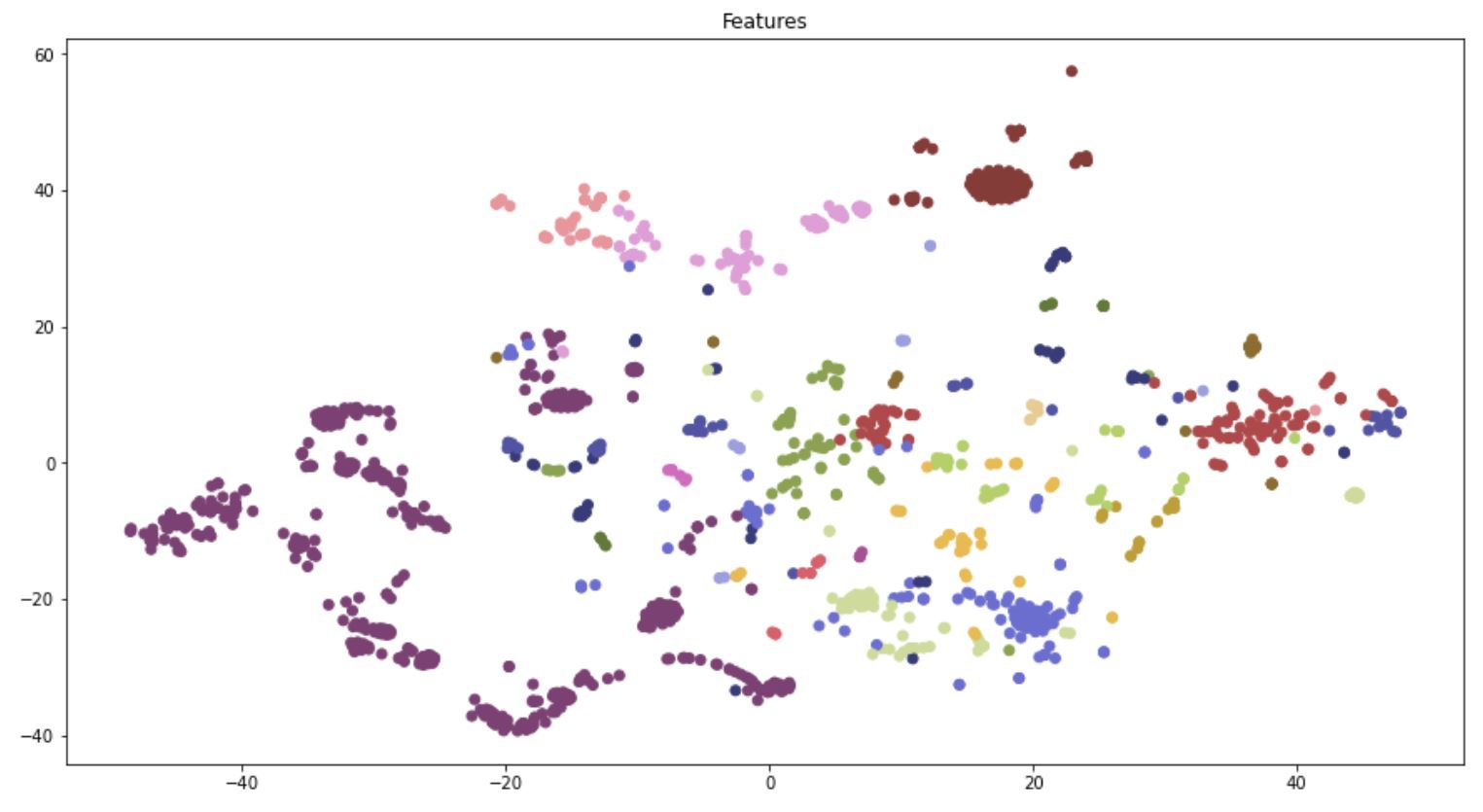
t-SNE (t-distributed Stochastic Neighbor Embedding) [61-65] converts the relationships of data points to probabilities. The similarities in the original space are represented by Gaussian joint probabilities and the affinities in the embedded space are represented by Student’s t-distributions. This allows t-SNE to be particularly sensitive to local structure and has a few other advantages over existing techniques.

Fig. 29 shows the result of application of t-SNE to the samples of our dataset. It shows different groups of samples gathered together probably describing different states and the red points indicate the outliers as in Fig.3. The projection of the high-dimensional dataset into the 2-dimensional visualization seems rather descriptive. Probably, it can be useful for data labeling in cases when they are not available. Moreover, such visualizations reveal possible different states of systems.

Fig. 30 shows the result of application of t-SNE on the features of our dataset. In this case, we used the correlation distance to see highly correlated groups of variables. The colors are obtained via application of the hierarchical clustering described above (feature agglomeration). For more clarity, we used only 20 clusters (instead of 381).



**Figure 29.** t-SNE applied to observations. Red points correspond to outliers (see Fig. 3).



**Figure 30.** t-SNE applied to features (columns of the dataset) with the correlation distance. Different colors correspond to different clusters. For the purposes of visualization, the coloring is performed for 20 clusters.