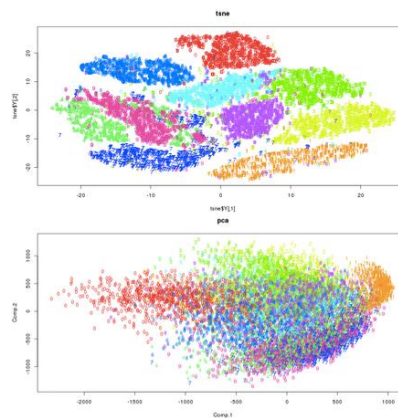


Dimension Reduction for Big Data Visual Analytics

Project Description: Big data come usually as a large set of multidimensional instances. Visualizing this kind of data, such as actigraphy data produced by wearable sensors (e.g. Fitbit), is crucial for instance to discover typical behaviors that would distinguish patients with specific diseases, like diabetes and obesity, two major issues in Qatar.

But visualizing these large and high dimensional data is very challenging. Approaches like Principal Component Analysis generate many artefacts due to their linear assumption, while t-SNE, a well-known nonlinear technique does not scale well to large datasets. We developed a new projection technique (HilMap) to represent multidimensional data as a 2 dimensional multiscale scatterplot. We want to assess its scalability and develop a Python code to make it easy to use and to interpret.



tSNE (top) and PCA (bottom) 2 dimensional projection of originally multidimensional data.

Duties/Activities: The core code of HilMap already exist, but there is still work to be done to make it usable and useful:

- 1) Implement a Python/Bokeh code to allow interactive data analysis based on HilMap able to handle large datasets.
- 2) Compare quality and speed of HilMap with other DR techniques, optimize the code of HilMap

Required Skills: Python

Preferred Intern Academic Level: BSc 3rd year or 4th year

Learning Opportunities:

Python and Bokeh programming language. Advanced visualization and processing techniques

Expected Team Size: 4 interns (2 for each task)

Mentors

Michael Aupetit email: maupetit@hbku.edu.qa

Abdelkader Baggag email : abaggag@hbku.edu.qa

Ala Abuthawabeh (on task 1)