Course Project for Data Visualization, Spring 2020

3-d Radviz visualisation method

Prepared By: Noopa Jagadeesh, Dipkumar patel, Prasanth varma

Table of Contents

Prepare a 3D-radviz visualization method

Goal

Convert high dimensional data into 3-dimensional points and plot attributes on the sphere and plot the points inside the sphere and than visualize 12 datasets from UCI machine learning repository.

Description

Visualizing a higher dimensional data has been a challenge since so long for researchers and over the time new techniques have been developed to visualize them. RadViz(Hoffman et al. 1997) is one among that technique which converts higher dimensional data into 2 dimension and plot it on the circle. RadViz is a multivariate data visualisation algorithm that plots high-dimensional data on the perimeter of a circle, then plots points inside the circle so that the point normalises its values on the centre-to-arc axes. It helps to visualize clusters of the datasets. It works on following principle:

"Each point kept in a place with springs attached to the variable anchors at the other end. Each spring's stiffness is proportional to the value of the corresponding variable, and the point ends up in the position where the spring forces are in balance. Variable values are normalized between 0 and 1 before visualisation. For those variables, data instances close to a set of variable anchors have higher values than for the others"

Extension can be provided to existing RadViz technique by replacing circle with a sphere and by converting high dimensional data into 3D instead of 2D. RadViz-3D can help visualizing clusters efficiently and can add a dimension to visualize more information which are not possible in current 2 dimension. In order to achieve this, we started looking for some existing research work that has achieved this objective. We came across the paper titled "Three-dimensional Radial Visualization of High-dimensional Continuous or Discrete Data" – written by(Fan Dai, Yifan Zhu and Ranjan Maitra). This paper presented a technique to prepare a 3D-radviz visualization. They proposed MRP (Max Ratio Projection) method to utilize grouping info and present them in lower dimension.

In order to create a 3d-radviz we used a R library named radviz3d. It can be installed using the steps mentioned in the readme file. GitHub library: https://github.com/fanne-stat/radviz3d

This library provides a platform to plot the 3d radviz. Using the configurations we can change the transparency of the sphere, labels location and width of the axis, etc.

1) Iris Data Set: [Dataset Info](#)

The data set contains three classes of 50 instances each, where each class refers to the iris plant type. There are total 4 attributes named sepal length, sepal width, Petal length and petal width. Visualization of this dataset using radviz3D is as shown below:
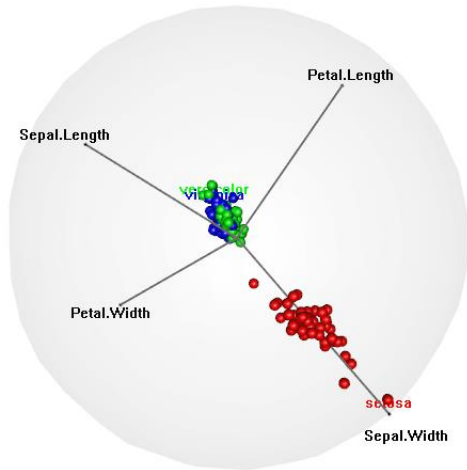

2) Adult dataset: [Dataset Info](#)
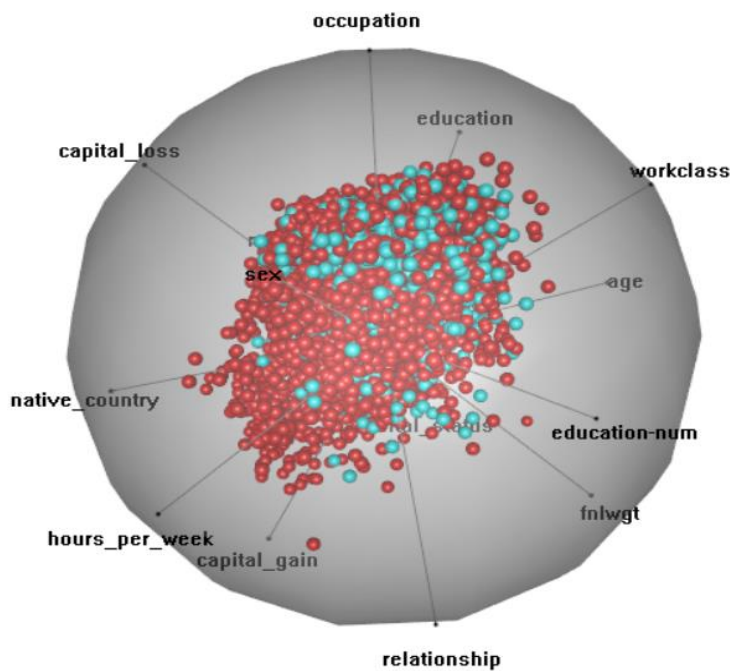
Use: To predict if income will exceed $50k/yr or not
No. of attributes: 14
No. of instances: 48842
Columns: age, work class, fnlwgt, education, education-num, marital-status, occupation, relationship, race, sex, capital-gain, capital-loss, hours-per-week, native-country and salary

Fig: RadViz-3d Visualization of Adult dataset

3) Wine dataset: Dataset Info
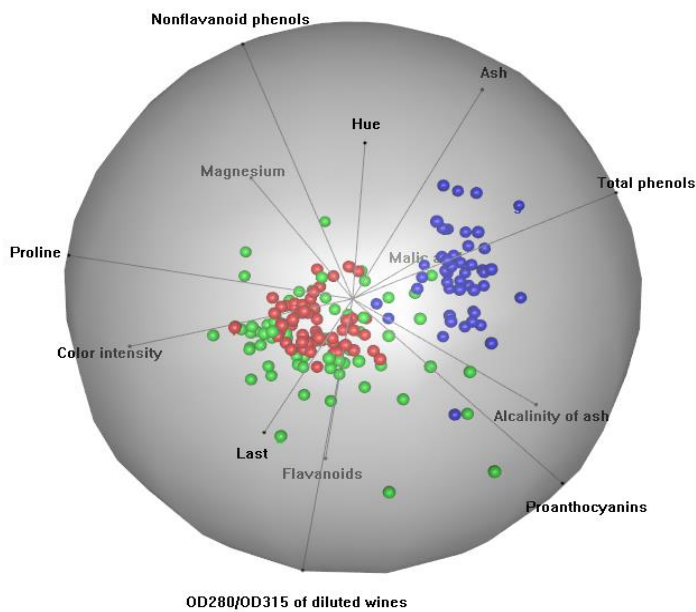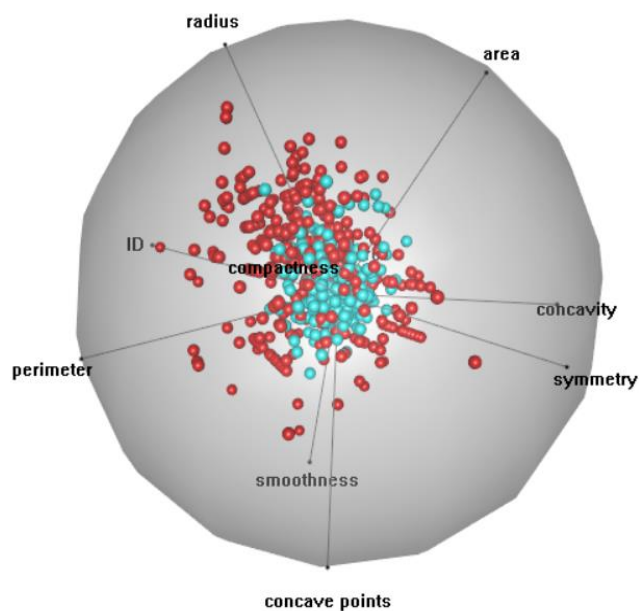

Use: To predict the origin of wine
No. of attributes: 13
No. of instances: 178
Columns: "Alcohol", "Malic acid", "Ash", "Alcalinity of ash", "Magnesium", "Total phenols", "Flavanoids", "Nonflavanoid phenols", "Proanthocyanins", "Color intensity", "Hue", "OD280/OD315 of diluted wines" and "Proline"

Fig: RadViz-3d visualization of Wine dataset

4) Breast Cancer Wisconsin (Diagnostic): Dataset Info
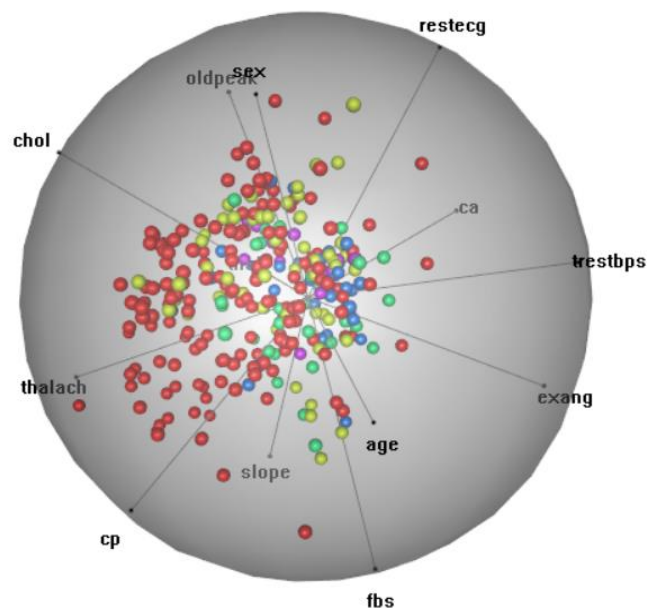

Use: to predict two types of cancer
No. of attributes: 32
No. of instances: 569
Columns: "Alcohol", "Malic acid", "Ash", "Alcalinity of ash", "Magnesium", "Total phenols", "Flavanoids", "Nonflavanoid phenols", "Proanthocyanins", "Color intensity", "Hue", "OD280/OD315 of diluted wines" and "Proline"

Fig: RadViz-3d visualization of breast cancer dataset

4

5) Heart Disease: Dataset Info

Use: to predict heart disease presence
No. of attributes: 75
No. of instances: 303

Fig: RadViz-3d visualization of Heart Disease dataset

6) Wine Quality: Dataset Info

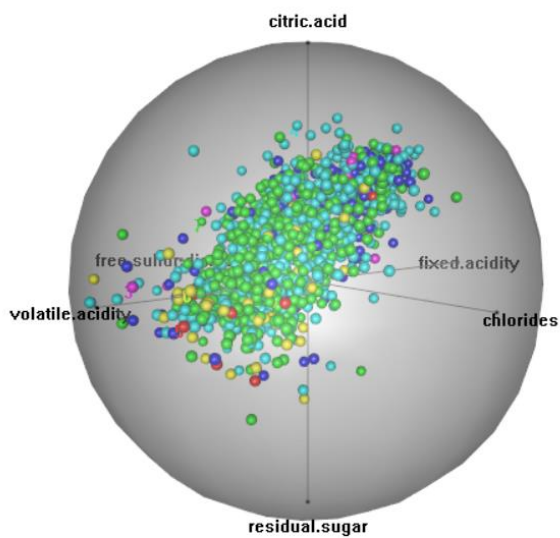This dataset is to predict the wine quality. It has 12 attributes and 4898 instances.

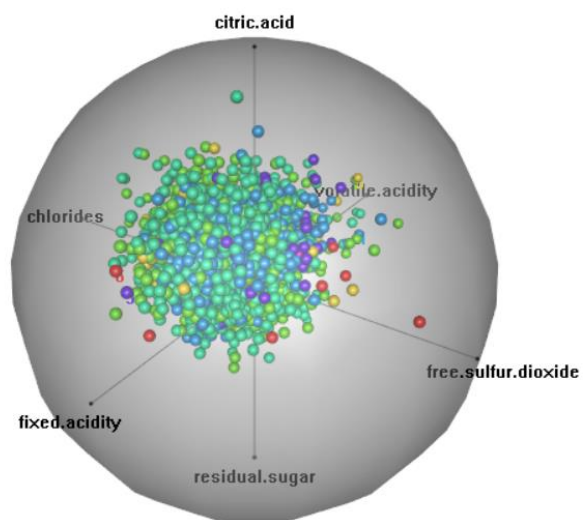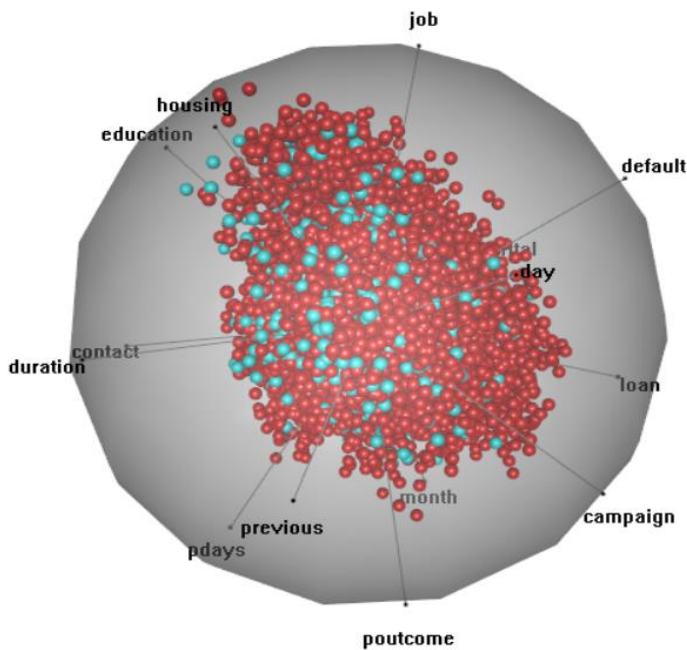Fig: RadViz-3d Red Wine Visualization

Fig: RadViz-3d White Wine Visualization

7) Bank Marketing: Dataset Info

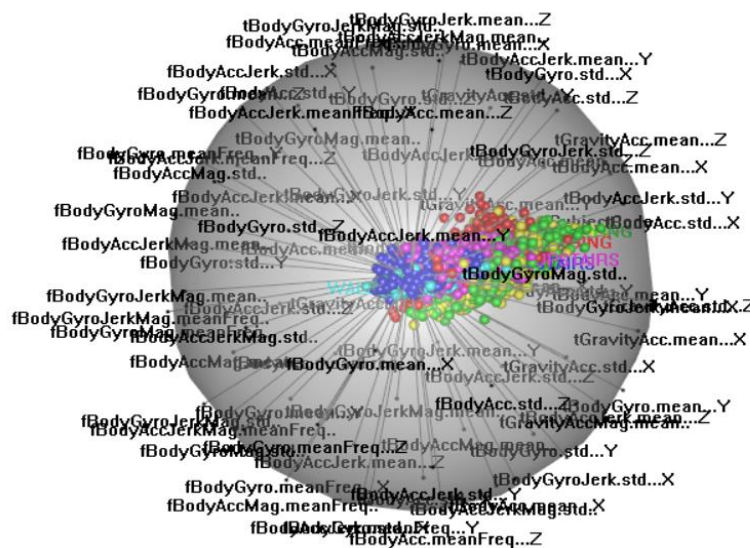**Use:** To predict if the client will subscribe a term deposit
No. of attributes: 17
No. of instances: 45211

Fig: RadViz-3d Visualization of Bank Marketing Dataset

8) Car Evaluation: Dataset Info

9) Human Activity Recognition Using Smartphones: Dataset info

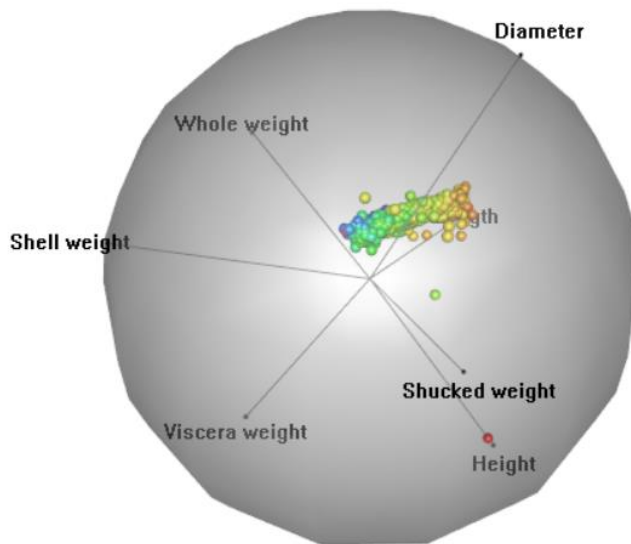**Use:** To predict different human activities using sensors reading
No. of attributes: 561
No. of instances: 10299

Fig: RadViz-3d visualization of human activity recognition using smartphones
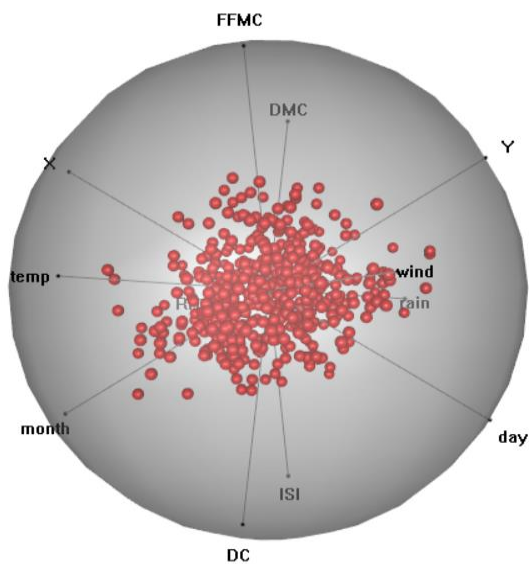
10) Abalone: Dataset Info

Use: To predict the age of abalone
No. of attributes: 8
No. of instances: 4177

Fig: radViz-3d visualization of Abalone dataset

11) Forest Fires: Dataset Info

Use: To predict the burned area of forest fires
No. of attributes: 13
No. of instances: 517

Fig: RadViz-3d visualization of forest fires dataset

12) Student Performance: [Dataset Info](#)

This dataset is to predict student performance. This dataset has 33 instances and 649 attributes.
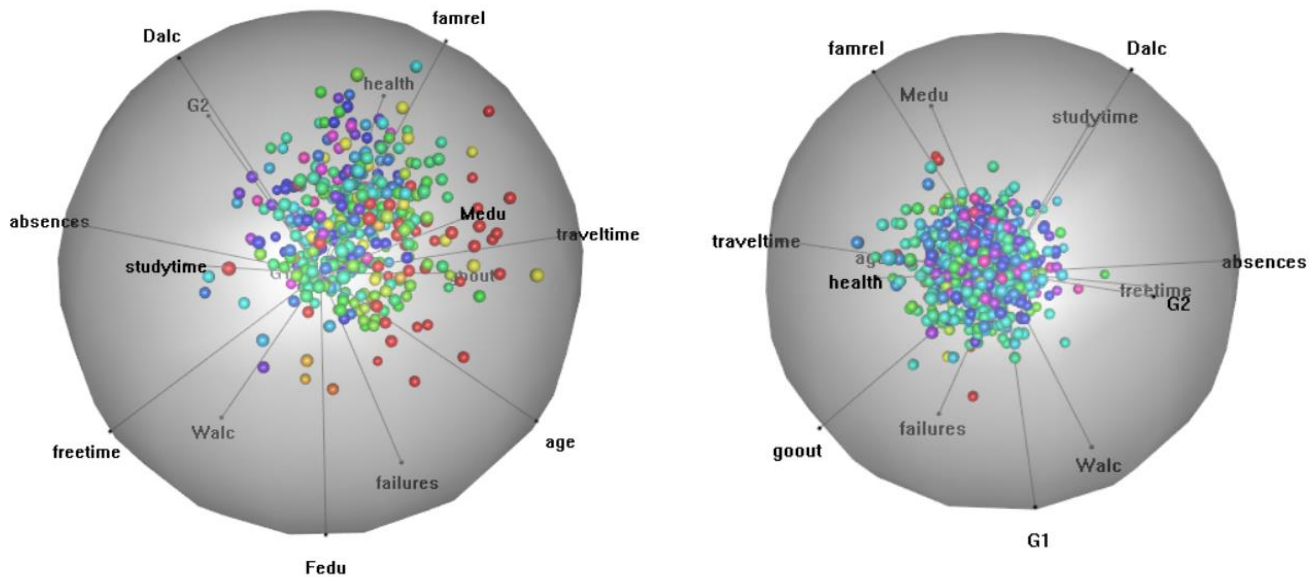

Fig: RadViz-3d visualization of Student Maths and Student Por dataset

Conclusion

By plotting radViz-3d visualization for 12 datasets, we can conclude that RadViz-3d(sphere) provides more insight than 2d radviz. It adds 4th dimension by using colors to show the clustering and in clustering if the attributes are nearby then visualizing the relation becomes better using radviz-3d.

Reference

1. GitHub R library: https://github.com/fanne-stat/radviz3d
2. Research paper: [Three-dimensional Radial Visualization of High-dimensional Continuous or Discrete Data](#)