

# **Visual Language and Visualization**

High-dimensional Data Visualization

Ting Wang    Xuejun Xie    Luyao Wang

# Task1

## Data Visualization Based on “CO<sub>2</sub> emissions” Data Set

**Description:** This data set is from UNdata, which provides data of 175 countries' CO<sub>2</sub> emissions in 51 years(1960-2010).

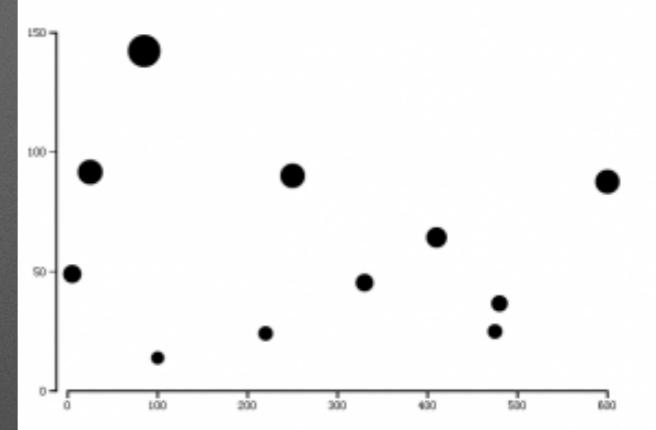
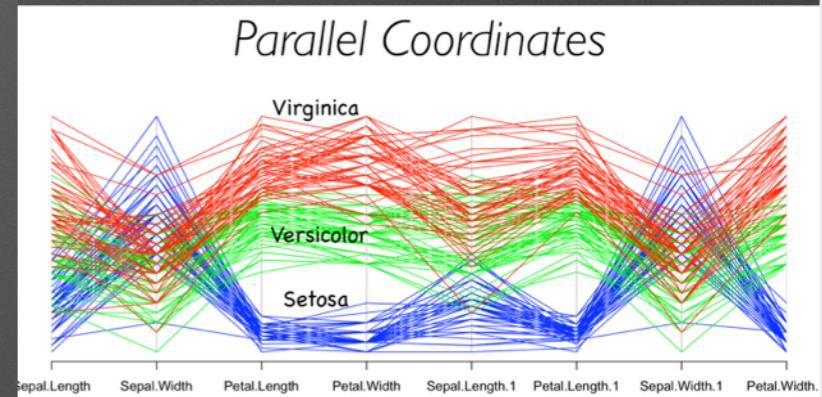
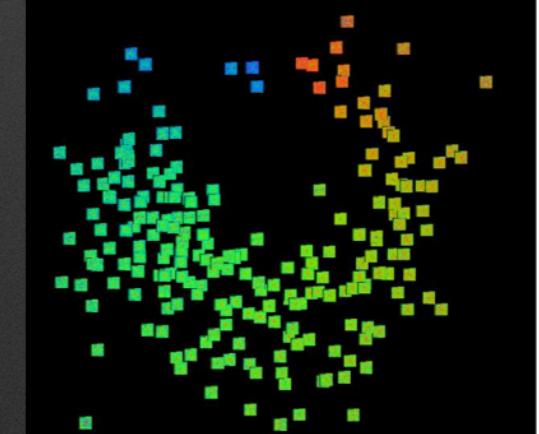
### Size:

- 175 records
- 51 attributes and a label(country's name)
- 368KB

### Task( our work):

- Display CO<sub>2</sub> emissions for each year and country.
- Analysis the variation of CO<sub>2</sub> emissions in each country.
- Look for non-compliance with the laws of abnormal data, and analyze the reasons.

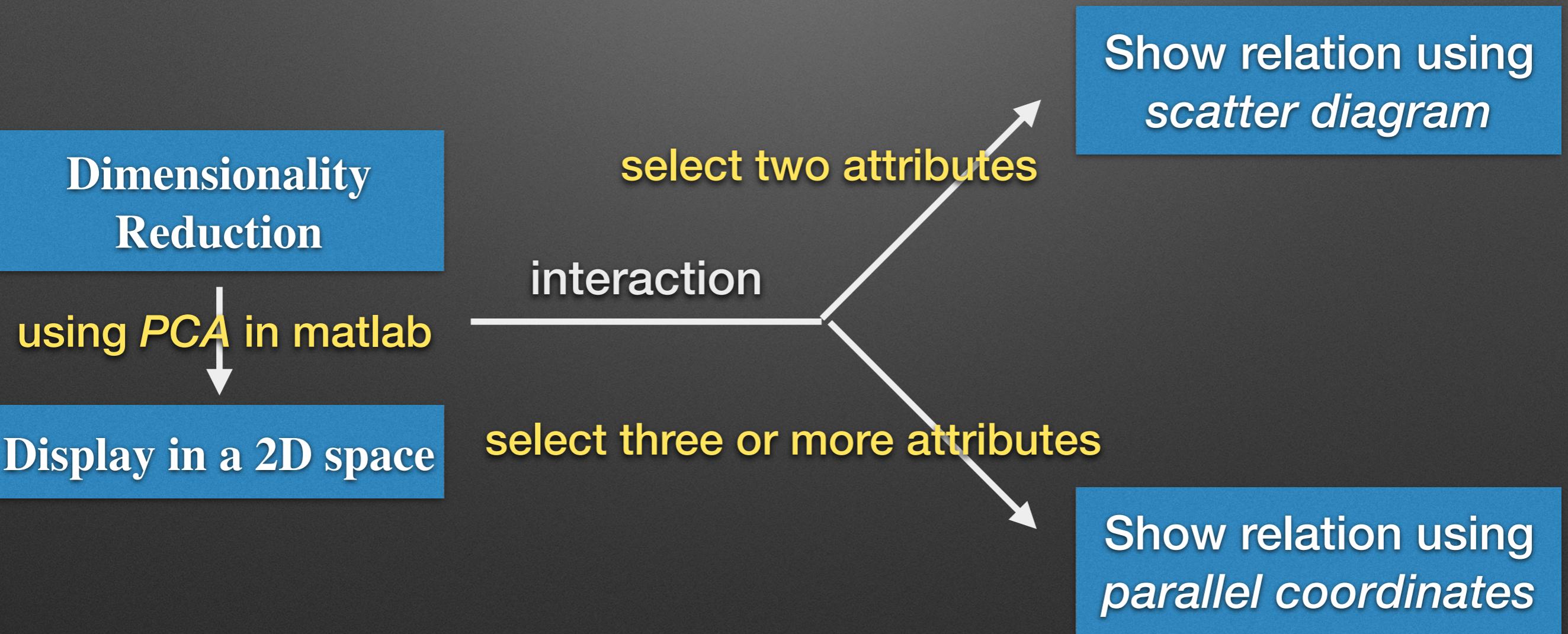
# Recent method for “High-dimensional Data Visualization”

METHOD	ADVANTAGE	DISCRIPTION
SCATTER DIAGRAM	Show relations between two dimensions	
PARALLEL COORDINATES	Show relations among multivariate data	
DIMENSIONALITY REDUCTION	Show the macroscopic of multivariate data	

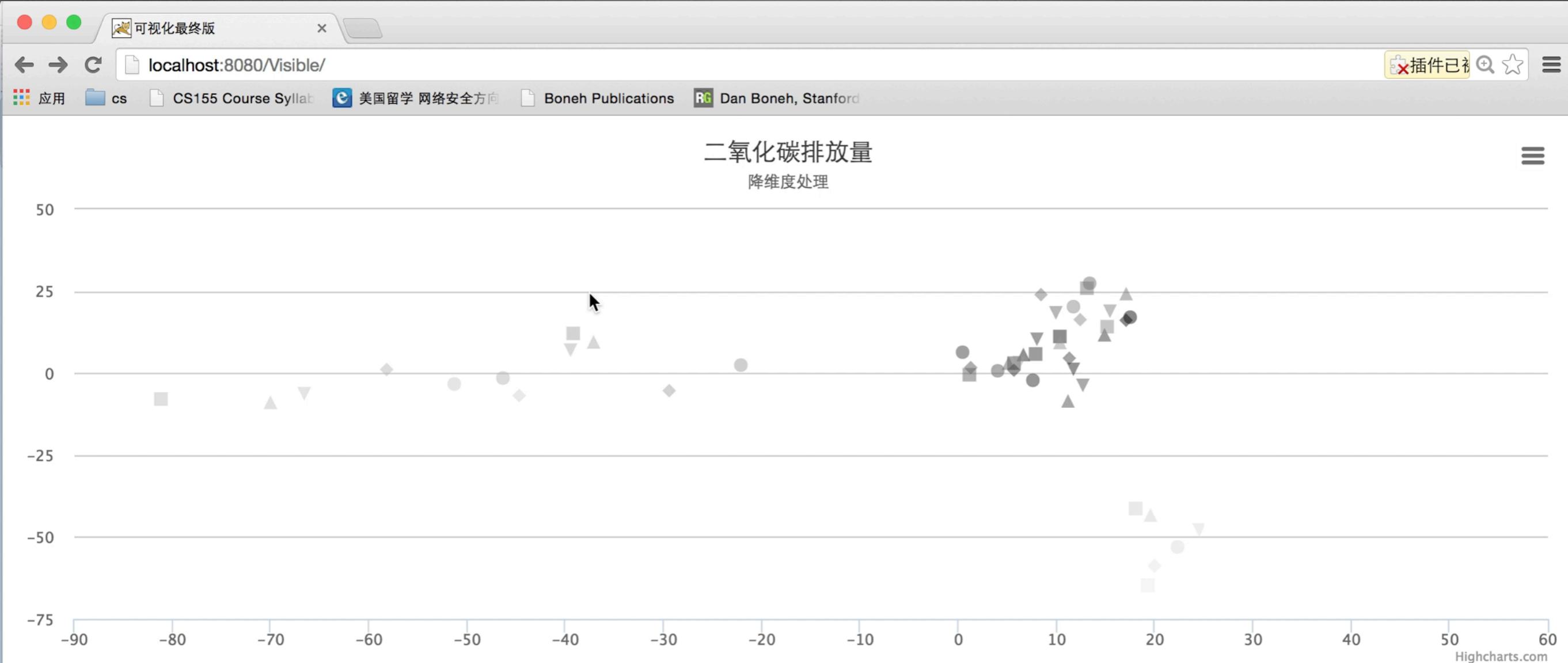
# Our Idea

Combining these three methods and adding interaction

## Realization process



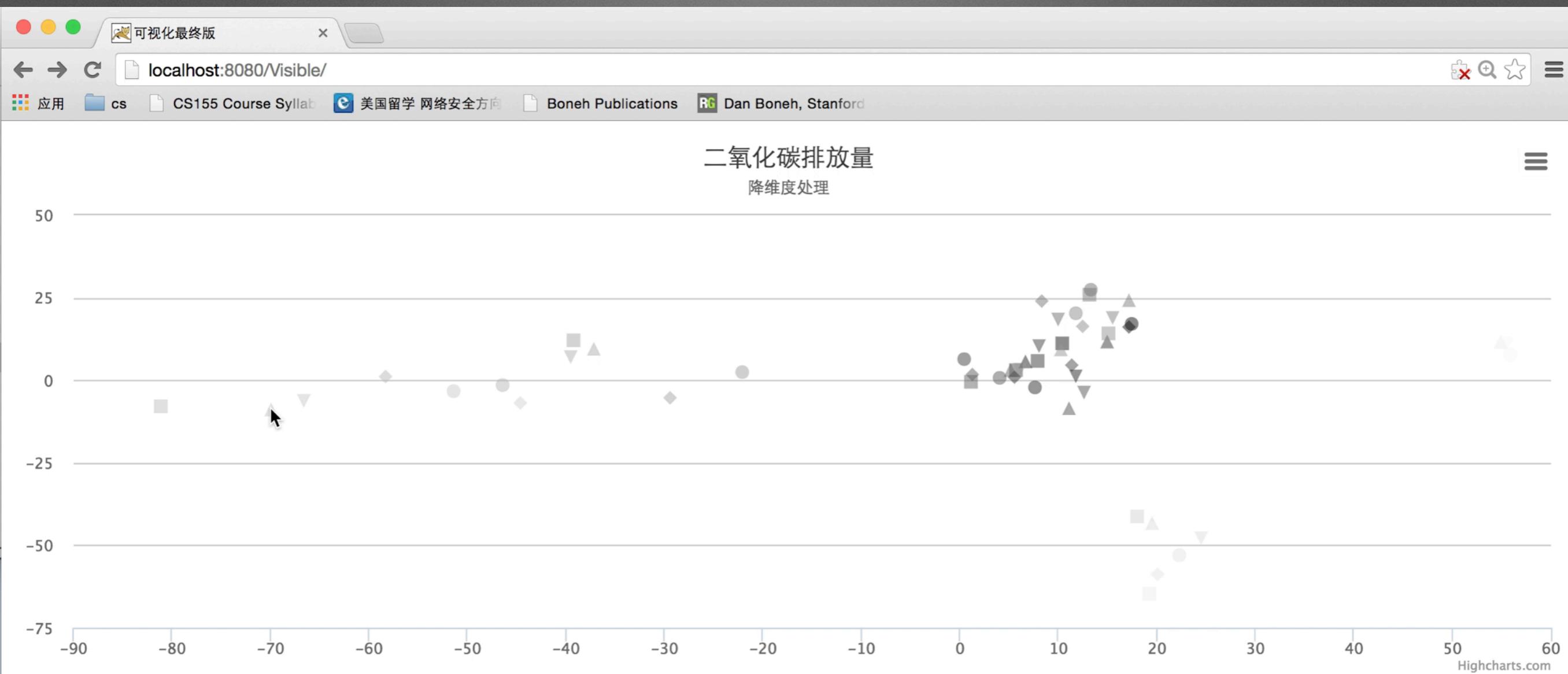
# Demo1 Dimensionality Reduction



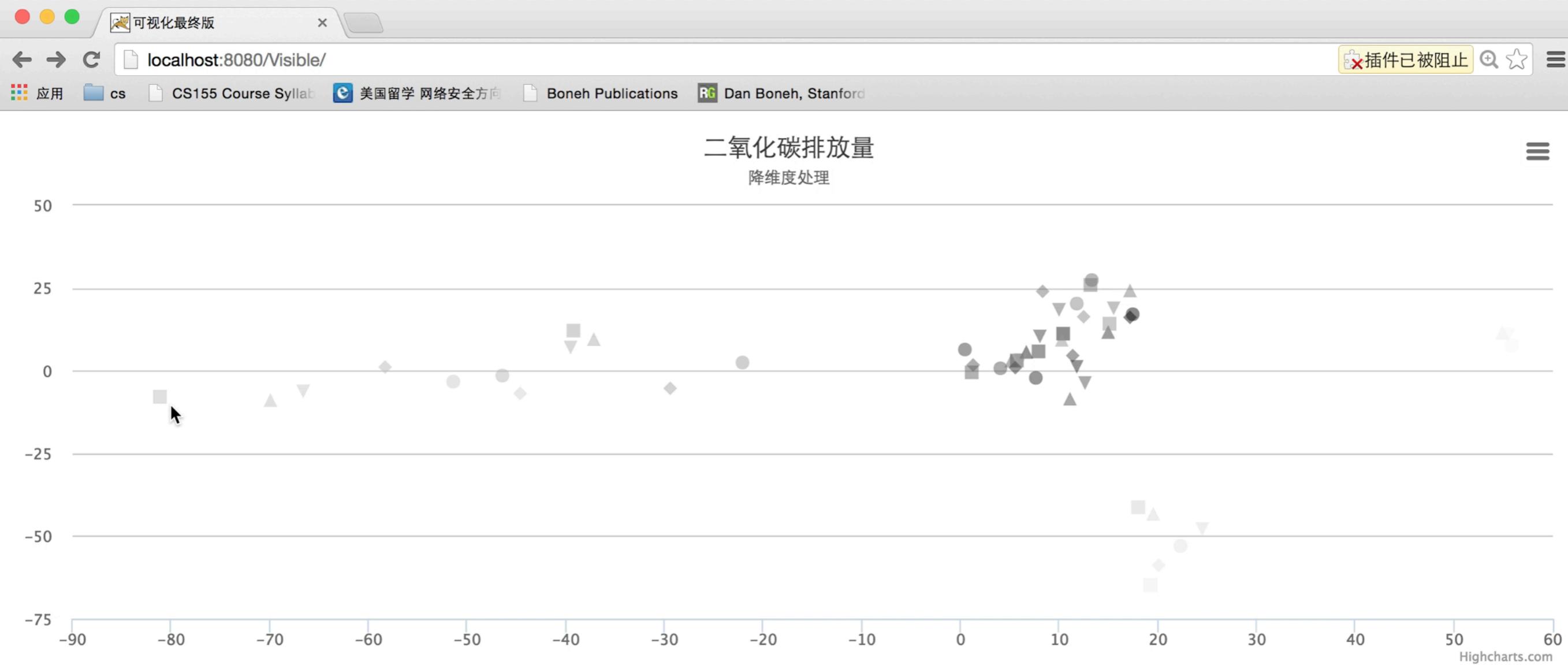
两列散点图=>GO

两列及以上数据处理=>GO

# Demo2 Show relation using scatter diagram



# Demo3 Show relation using *parallel coordinates*



两列散点图=>GO

两列及以上数据处理=>GO

# Problems though showing some information

- **Work for specific datasets**

We have tried some dataset like KDD 99 ..... (**sparse matrix**)  
strange when show in Parallel Coordinates also scatter diagram

- **For similar data**

Some problems..... **overlapping!**

- **Do not work for text**

# Task2

## Data Visualization Based on Customer Data

**Description:** This dataset was used for the Coil 2000 data mining competition. It contains customer data for an insurance company.

### Size:

- 5822 records
- 86 attributes

**Task( for example):** Try to find the correlation among each attributes and try to do some prediction, like predicting whether someone will buy a caravan insurance policy.

# Problems & Our idea:

Problem1: Different data type. Different domain

# Problems & Outcomes

## Problem1: Different Demographic Groups

L2:

- 1 Successful hedonists
- 2 Driven Growers
- 3 Average Family
- 4 Career Loners
- 5 Living well
- 6 Cruising Seniors
- 7 Retired and Religious
- 8 Family with grown ups
- 9 Conservative families
- 10 Farmers

# Problems & Our idea:

Problem1: Different data type. Different domain

Our idea:

- Dim reduction and project the data on map like task1  
**Hope the neighbor attributes have some relation.....**
- Group these attributes considering their correlation.  
**Calculate correlation coefficient or use other methods.**

# Problem

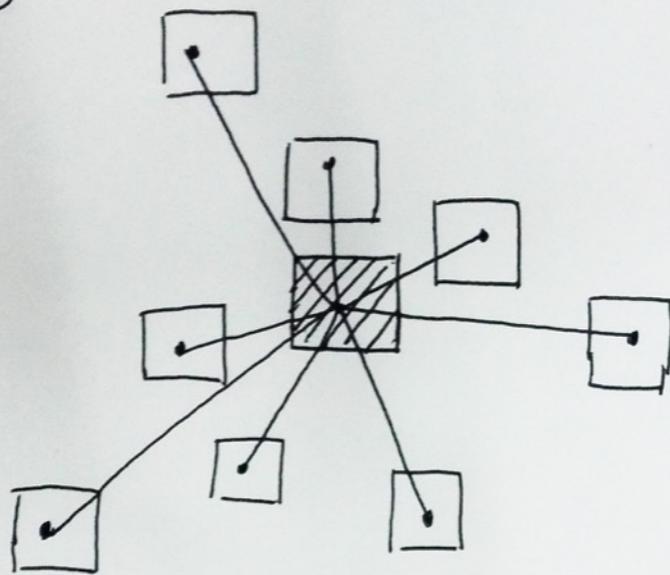
## Problem1:

Our idea:

- Dim reduction
- Group them

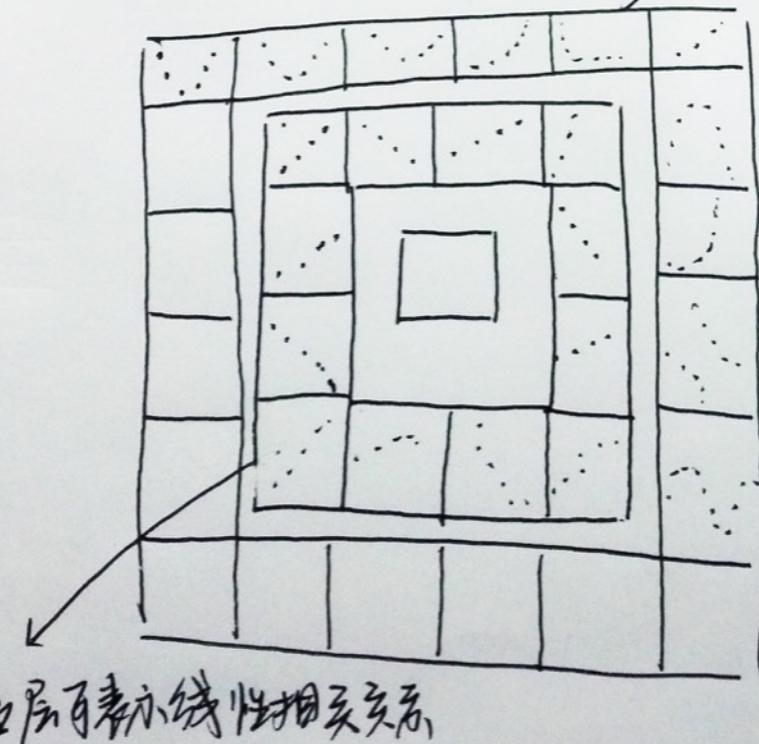
用户选择一个属性(列)为核心，其余属性(列)按照相关系数或相似度  
类型分布在周围。

如：① 这边表示相关程度(如线性相关系数)



② 这边表示相关系数

→ 其他相关系数如指数、三角等。



这层可表示线性相关关系

domain

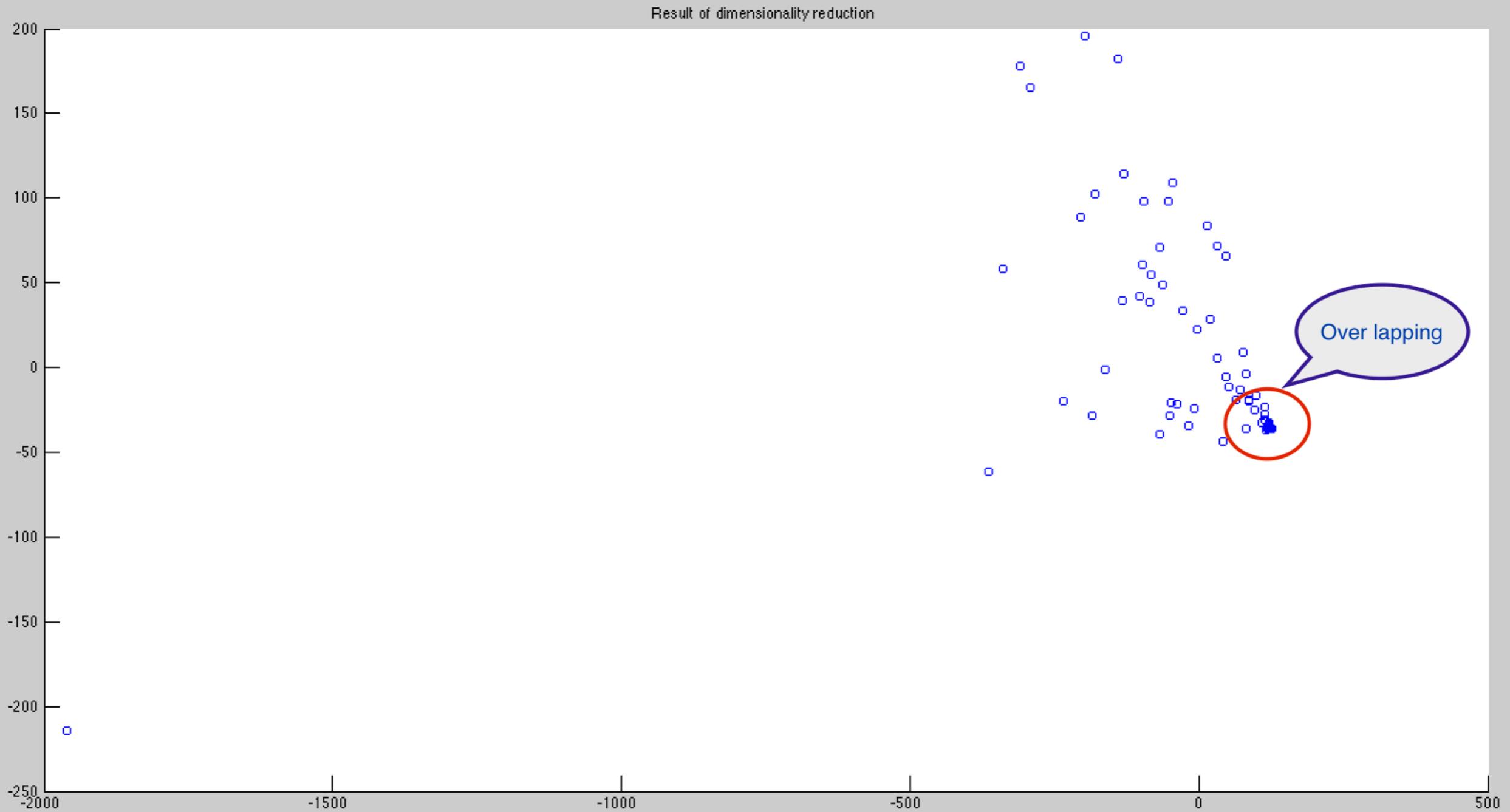
map like task1

me relation.....

correlation.

other methods.

# Problems & Our idea:



# Problem

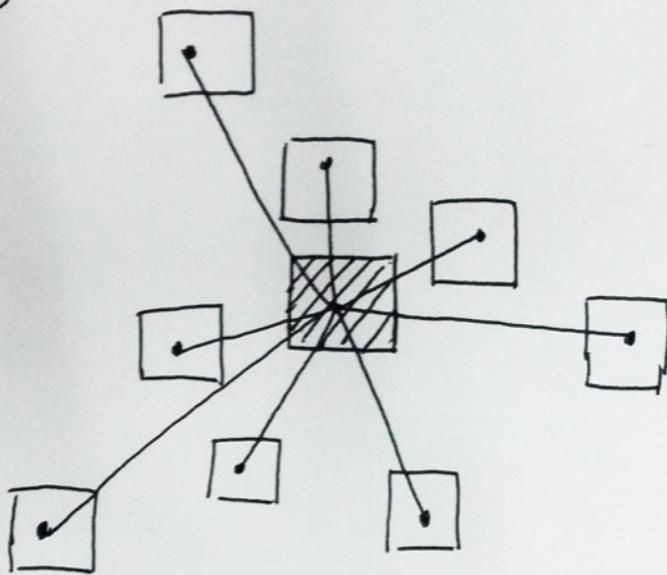
## Problem1:

Our idea:

- Dim reduction
- Group them

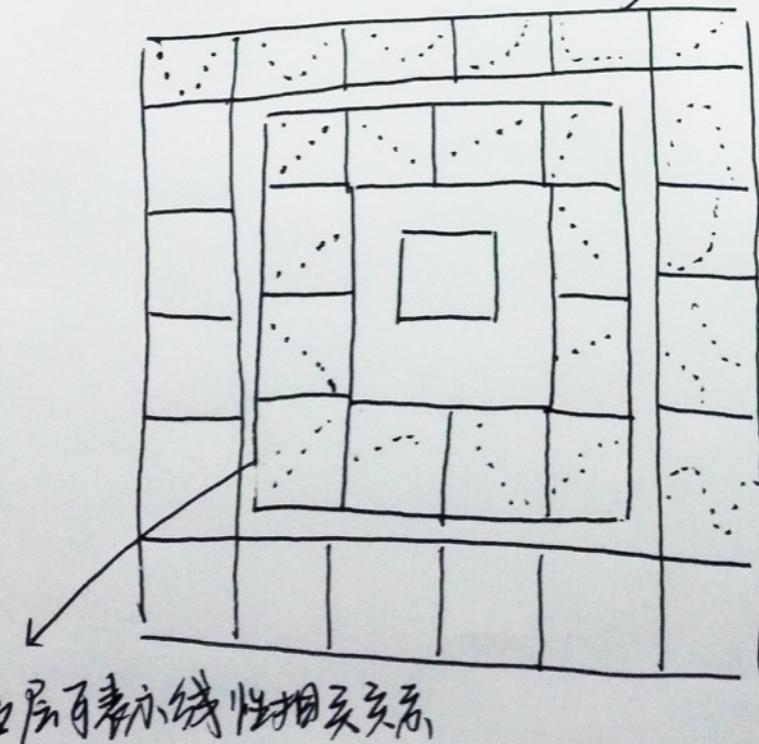
用户选择一个属性(列)为核心，其余属性(列)按照相关系数或相似度  
类型分布在周围。

如：① 这边表示相关程度(如线性相关系数)



② 这边表示相关系数

→ 其他相关系数如指数、三角等。



domain

map like task1

me relation.....

correlation.

other methods.

# Problems & Our idea:

**Problem2:** Approaches do not consider the background knowledge the user may have.

**Our idea:**

- Also project each records on map using PCA.....
- Add different user-defined properties (using some attr.)

**Make user find the best dim reduction result  
(we provide a rich output for different user-defined properties)**

# Problems & Our idea:

Problem3: Some data can not present well using scatter diagram or Parallel Coordinates  
(like  $\{(0,0),(0,0),(0,0),(0,0),(1,0),(0,1),\dots\}$ )

```
,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0  
,0,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0  
,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0  
0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,  
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,  
,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
```

# Problems & Our idea:

**Problem3:** Some data can not present well using scatter diagram or Parallel Coordinates  
(like  $\{(0,0),(0,0),(0,0),(0,0),(1,0),(0,1),\dots\}$ )

**Our idea:**

- Provide different graphs( bar graphs.....)

# Thank You!