

# Библиотека для визуализации графика в параллельных осях

Тыцкий Владислав

Ноябрь 2020

## Графики в параллельных осях

График в параллельных осях — метод визуализации многомерных данных.

Для отображения векторов в  $n$ -ом пространстве рисуется  $n$  параллельных линий(осей) на равном расстоянии друг от друга. Вектор представляется в виде ломаной кривой, с вершинами на параллельных осях. Точка пересечения линии с  $i$ -ой осью соответствует  $i$ -ой координате объекта.

Возникают естественные вопросы:

- В каком порядке расположить оси?
- В какую сторону направлять ось?
- Какой масштаб выбрать для каждой оси?

Существуют программные реализации графики в параллельных осях: ELKI, GGobi, Mondrian, Orange, ROOT, plotly.

К своему удивлению обнаружил, что в matplotlib из коробки нет parallel coordinates.

Его можно построить только с пакетом pandas(внутри они все равно используют matplotlib).

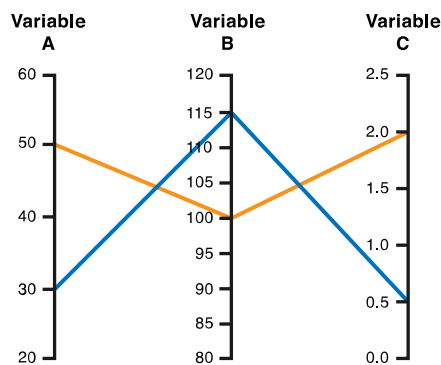


Рис. 1: Пример графика

## Иерархические графики

При визуализации большого количества объектов(векторов) график в параллельных осях становится сложночитаемым. Связано это с тем, что количество линий на графике становится также много, все друг на друга наслаивается и разобраться в этом невозможно.

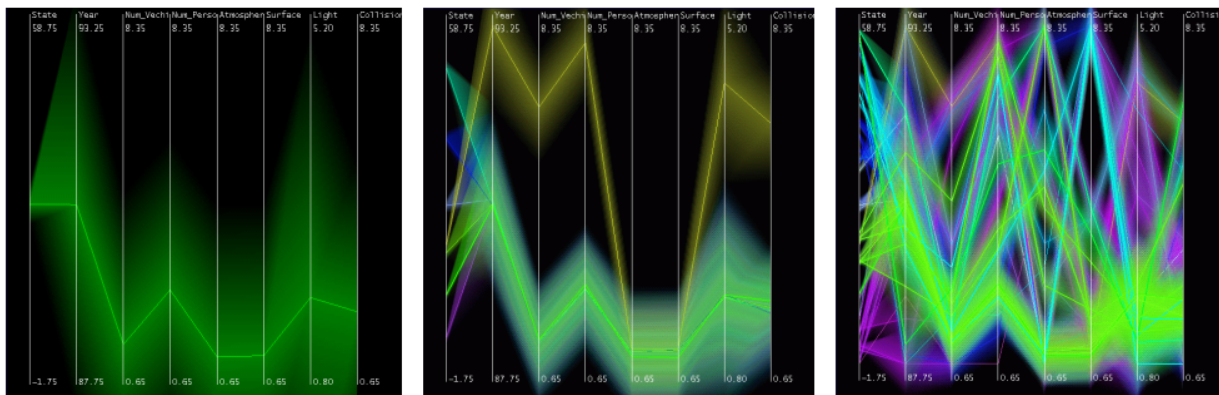


Рис. 2: Изображения графиков начиная с корня и заканчивая большим количеством кластеров

Одним из методов решения этой проблемы является добавление кластеров(можно рассматривать просто метки классов) — каждому кластеру(классу) соответствует определенный цвет, и каждый объект окрашивается в цвет кластера(класса). Такой подход немного помогает, но можно обобщить эту идею до более сложной.

Иерархические графики в параллельных осях представляют собой метод визуализации не объектов, а некоторых иерархических кластерных структур — дендрограмм. Давайте вместо визуализации конкретных объектов будем визуализировать сообщества похожих объектов(меру похожести можно выбрать). Чтобы визуализировать сообщества(кластеры) нужно выбрать некоторые статистики, например среднее и стандартное отклонение. Среднее нарисуем обычной линией, а стандартное отклонение отобразим полупрозрачным градиентом Рис.2. Так график становится более читабельным, а детализацию будем регулировать с помощью включения новых кластеров из дендрограммы.

## Что должно быть в библиотеке

В библиотеке необходимо реализовать:

- построение обычных графиков в параллельных осях
- построение иерархических графиков
  - отрисовка полупрозрачного градиента(заливки)
  - работу с иерархическими кластерами(дендрограммами)
  - цветовое кодирование кластеров
- опционально
  - выделение подмножества линий в диапазоне значений одной из осей
  - нахождение оптимального расположения осей
  - создание иерархических кластеров на основе входящей выборки

## Эстетика в визуализации

Помимо основных вещей, таких как легенда, подписи к осям, отметки на осях, название графика, хотелось бы добавить:

- цветовые схемы
- плавные градиентные переходы
- полупрозрачность линий

Ориентироваться можно на библиотеку [parasoljs](#). Parasol современная интерактивная библиотека для построения parallel coordinates, делающая почти все, что необходимо, кроме отображение кластеров с помощью полупрозрачной градиентной заливки.

## Дизайн библиотеки

Функционал и интерфейс библиотеки хочется сделать приближенным к более высокоуровневым библиотекам(например seaborn).

Пока не продумал прототипы функций, но можно будет за основу взять реализацию обычных графиков в параллельных осях в pandas.