

График в параллельных осях

Студент: Тыцкий В.И.

Научный руководитель: Майсурадзе А.И.

МГУ имени М. В. Ломоносова, факультет ВМК, кафедра ММП

Оглавление

Введение

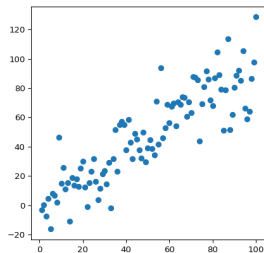
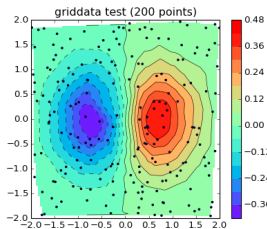
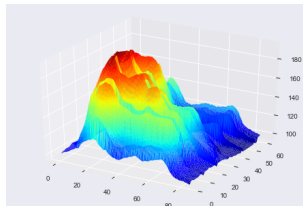
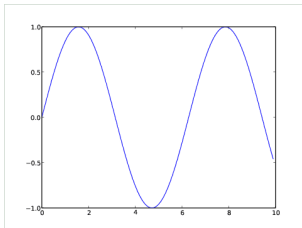
Модификации

Проблемы построения

Методы выбора порядка

О библиотеке

Примеры диаграмм



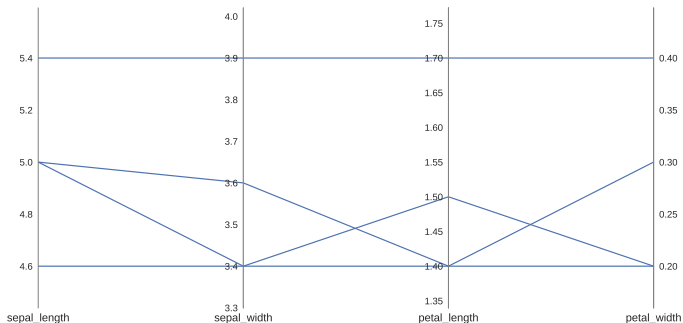
Историческая справка

The value of data visualization is not seeing “zillions” of objects but rather recognizing relations among them.

Alfred Inselberg

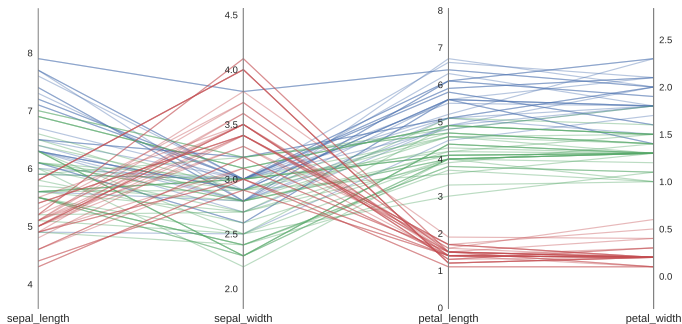
- Параллельные координаты были известны еще в 19-ом веке
- В 1980-ых были популяризированы Альфредом Инсельбергом

Классический график в параллельных осях



Модификации: кластеры

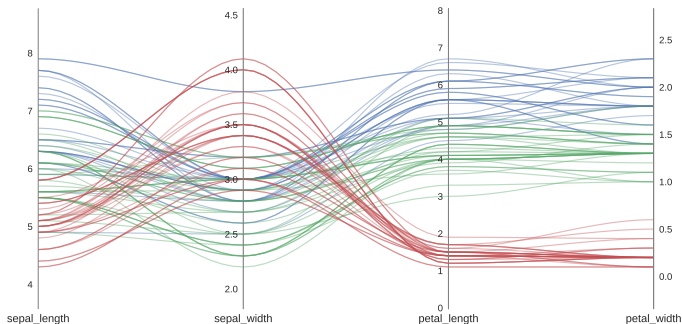
Кластер — класс родственных элементов статистической совокупности.



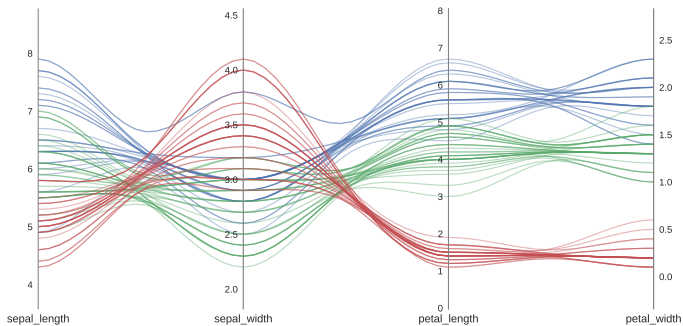
Чаще всего именно в таком виде используют график в параллельных осях.

Сглаживание линий

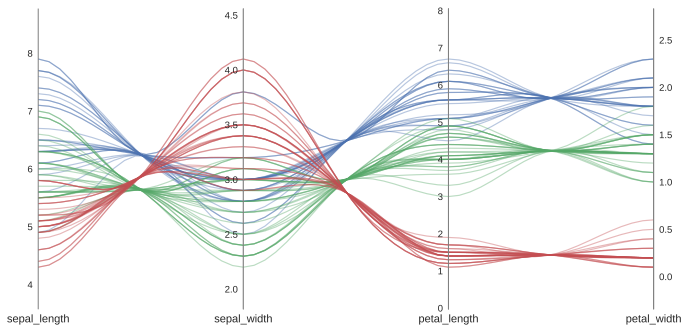
Человеку проще воспринимать гладкие линии, поэтому читаемость графика заметно возрастает.



Связывание линий

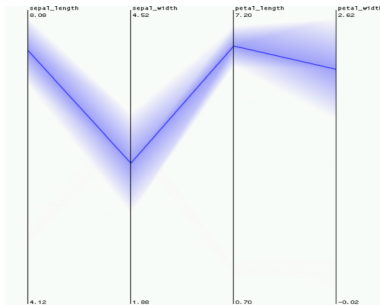


Связывание линий



Иерархические графики

Изображаем статистики распределений соответствующих кластеров (std, min, max, mean) вместо отрисовки каждого объекта.



Иерархические графики

Пусть $X = (x_1, \dots, x_n)$ – выборка, где $x_i \in \mathbb{R}^n$.

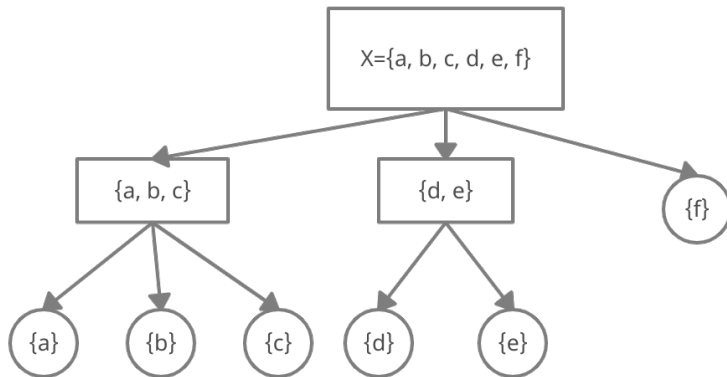
Назовем множество P m -разбиением множества X на m -подмножеств $\{P_1, \dots, P_m\}$ такое, что:

$$1. P_i \cap P_j = \emptyset, \quad \forall i, j = \overline{1, m}$$

$$2. \bigcup_{i=1}^m P_i = X$$

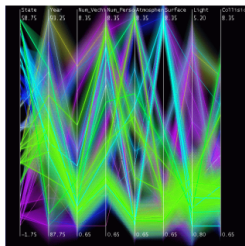
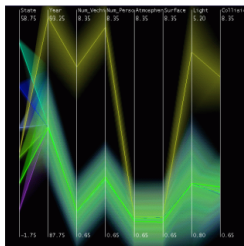
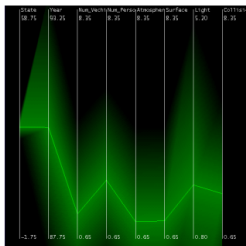
Организуем иерархическую структуру в виде дерева, где корнем является X , а каждая вершина сопоставлена элементу разбиения родительской вершины.

Пример иерархического разбиения

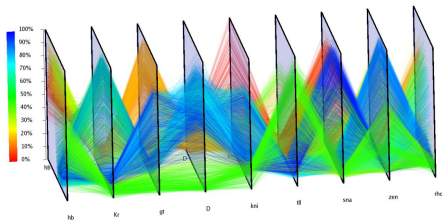


Иерархические графики

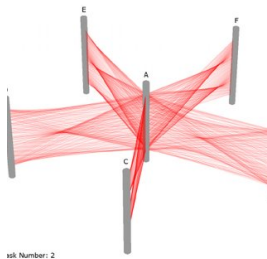
Регулируя глубину, мы добавляем/уменьшаем количество кластеров на графике



3D



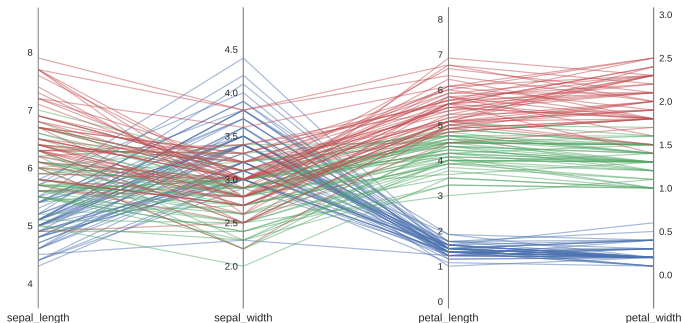
3D parallel coordinates

3D multi-relational
parallel coordinates

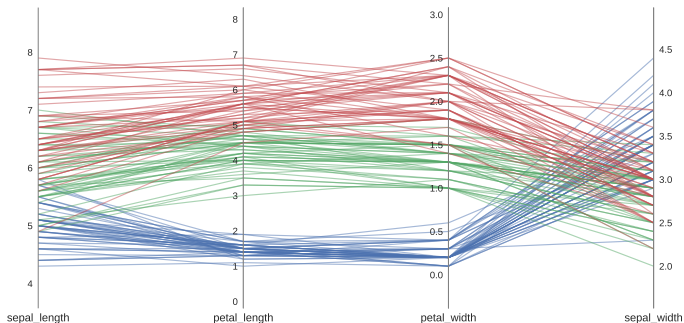
Естественные вопросы при построении

- В каком порядке расположить оси?
- В какую сторону направлять оси?
- Как много объектов отобразить?
- Какой масштаб выбрать для каждой оси?

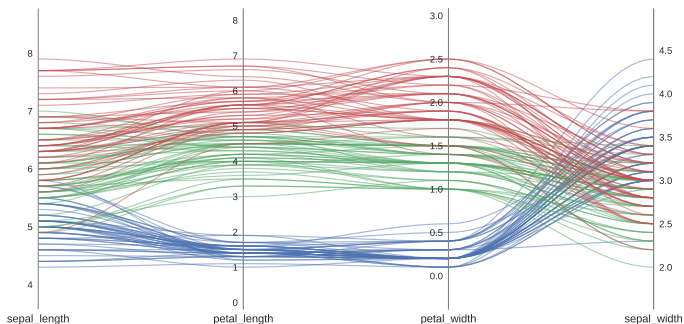
Выбор направлений и порядка осей



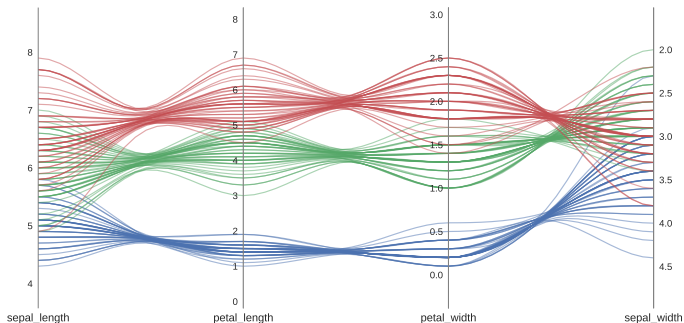
Выбор направлений и порядка осей



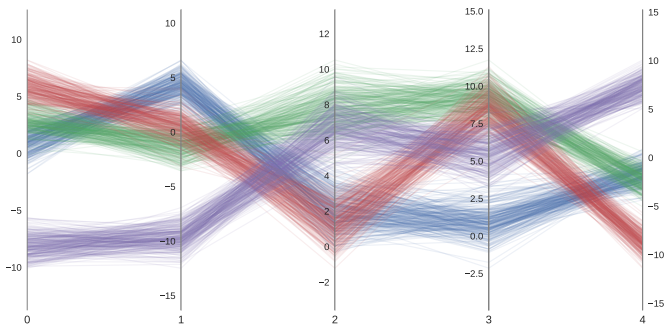
Выбор направлений и порядка осей



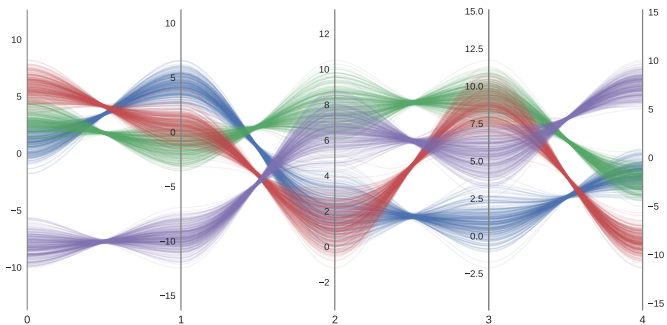
Выбор направлений и порядка осей



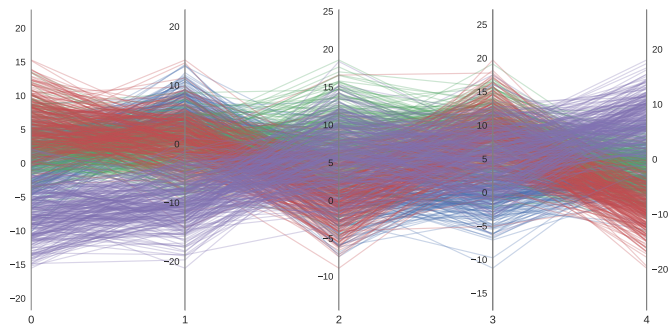
Влияние количества объектов на читаемость



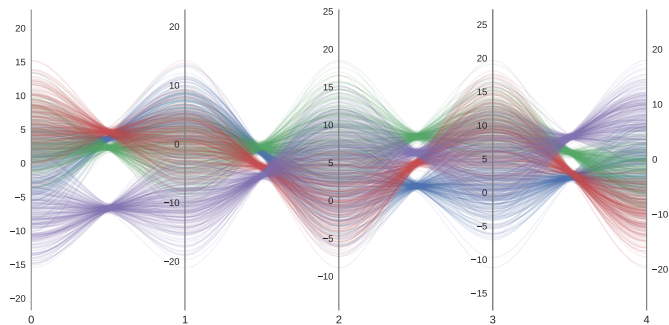
Влияние количества объектов на читаемость



Влияние количества объектов на читаемость



Влияние количества объектов на читаемость



Резюмируя

- Изменение степени прозрачности линий.
- Использование гладких линий.
- Связывание линий в рамках кластеров.
- Отображение лишь части объектов.
- Изменение порядка и направления осей.

Корреляция

Пусть даны две выборки $X = (x_1, \dots, x_n), Y = (y_1, \dots, y_n)$.

Корреляция Пирсона

$$\rho_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad |\rho_{XY}| \leq 1$$

Пусть R_i – ранг наблюдения x_i , S_i – ранг наблюдения y_i

Корреляция Спирмена

$$r_{XY} = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 \sum_{i=1}^n (S_i - \bar{S})^2}}, \quad |r_{XY}| \leq 1$$

$\rho_{XY} = 0, r_{XY} = 1$, где $X = Y^2$ и X симметрично распределена относительно нуля.

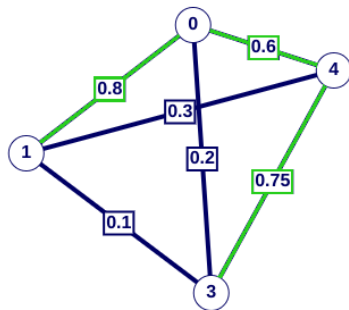
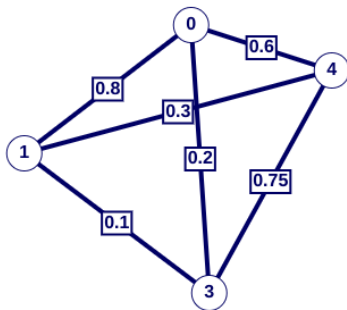
Задача о самом длинном пути

Это задача поиска простого пути максимальной длины в заданном графе. Является NP-трудной и не может быть решена за полиномиальное время для произвольных графов.

Пусть $X = (x_1, \dots, x_n)$ – выборка, где $x_i \in \mathbb{R}^n$.

Построим связный граф $G(V, E)$, где каждая вершина $u^i \in V$ соответствует i -ой координате (i -ой оси на графике), а каждому ребру $\{u^i, u^j\} \in E$ сопоставим вес равный $|r_{x^i, x^j}|$.

- Простейшим перебором задача решается за $O(n!)$
- Можно свести к задаче коммивояжера.
- С помощью методов динамического программирования можно улучшить асимптотику решения.



Обзор текущих средств

- На Python есть простейшая реализация лишь в библиотеке pandas!
- ELKI, GGobi, Mondrian, Orange и ROOT.
- Parcoords.js интерактивная библиотека на JavaScript.

Цели

- Дать возможность исследователям "безболезненно" использовать график в параллельных осях.
- Построение красивых и информативных графиков из "коробки".
- Реализация всевозможных видов данных графиков.

Технические подробности

- Статические графики.
- Библиотека пишется на языке Python на базе matplotlib.
- Простой высокоуровневый интерфейс. Как и в библиотеке seaborn методы могут принимать pandas.DataFrame, обычные numpy массивы или списки – для всего единый интерфейс.

Возможности

- Построение классических графиков в параллельных осях
 - Возможность рисовать гладкие линии.
 - Возможность "связывания" линий кластеров.
 - Возможность "связывания" линий на основе близости.
- Построение иерархических графиков
 - Отрисовка полупрозрачного градиента.
 - Работа с иерархическими кластерами.
 - Изображение распределения с помощью градиента.

Дополнительные возможности

- выделение подмножества линий в диапазоне значений одной из осей.
- нахождение оптимального расположения осей.
- создание иерархических кластеров на основе входящей выборки.

Итоги (после первого семестра)

- Возможность рисовать гладкие линии. **Пока что не добавлен параметр задающий вид кривой.**
- Возможность "связывания" линий кластеров. Добавлен непрерывный параметр задающий степень связывания.
- **Возможность связывания линий на основе близости не реализована**
- Интерфейс для пользователя практически полностью повторяет реализацию seaborn.¹

¹Большинство графиков в презентации нарисованы с помощью данной библиотеки.