

Разработка методов визуализации многомерных дискретных данных

Авербух В. Л., Михайлов И. О.

Институт математики и механики имени Н. Н. Красовского УрО РАН
Уральский федеральный университет (УрФУ) имени первого Президента России Б.Н. Ельцина.
Екатеринбург

Шестой Национальный Суперкомпьютерный Форум (НСКФ-2017),
Переславль-Залесский

Многомерные дискретные данные

Многомерные данные содержат информацию о трех или более признаках для каждого объекта. Почти все статистические данные и многие результаты вычислений являются многомерными. Главной проблемой при визуализации многомерных данных является, то что большинство средств визуализации не способны одновременно визуализировать значительное число измерений.

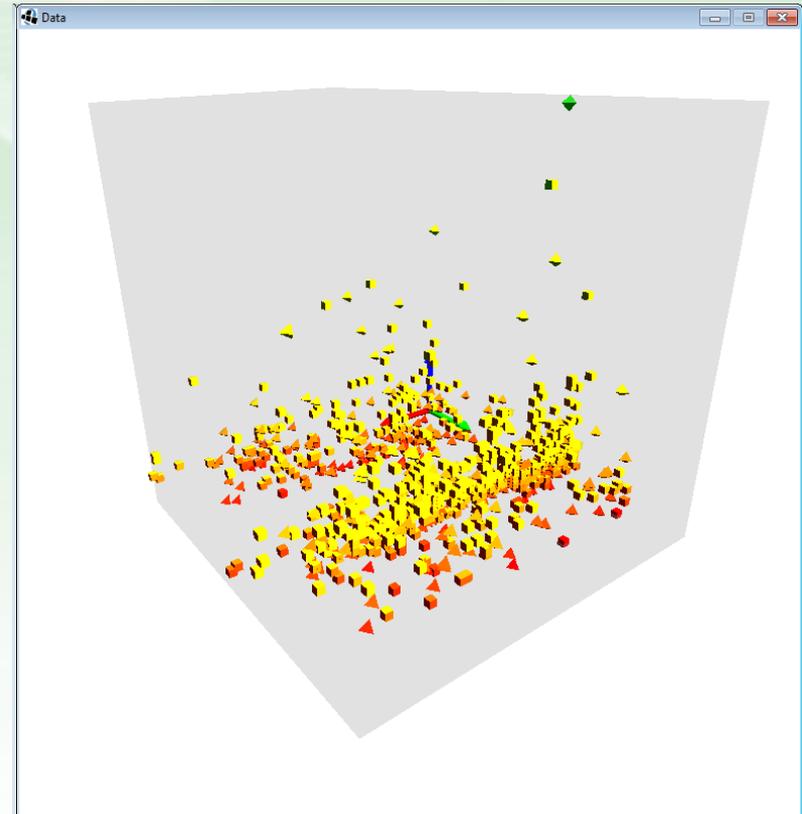
Данная работа рассматривает различные подходы, позволяющие визуализировать значительное число измерений с сохранением информативности. Для этого создан прототип системы визуализации с возможностью интерактивного выбора визуализируемых осей, способа их представления и фильтрации данных. В представленной работе рассмотрена разработка методов визуализации многомерных медицинских данных, собранных медицинской системой «qMS» и предоставленных компанией СП.АРМ.

Планируется глубокая интеграция с создаваемым программным комплексом.

Номер пациента	Пол	Возраст	Диагноз	Длительность эпизода	Рез
10019/C2014	Мужской	24,0	E10.1	15,0	
10046/C2016	Женский	42,0	E10.7	21,0	
10069/C2014	Женский	32,0	E10.7	30,0	
10154/C2014	Женский	56,0	E10.2	11,0	
10179/C2016	Мужской	56,0	E10.7	9,0	
10236/C2015	Женский	24,0	E10.7	13,0	
10238/C2016	Женский	51,0	E10.7	23,0	
10315/C2016	Мужской	28,0	E10.7	13,0	

Методы визуализации многомерных данных

- Координаты элемента в трехмерном пространстве - 3 измерения
- Цвет элемента - от одного до 3х измерений
- Размер элемента - одно измерение
- Форма элемента - одно измерение, ограниченный набор значений
- Прозрачность элемента
- Вращение элемента
- Блеск элемента — бинарное значение

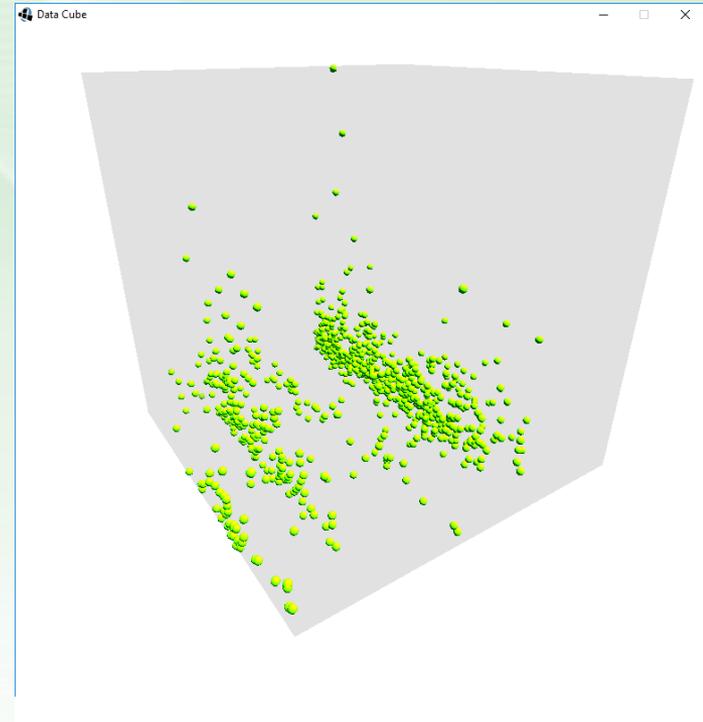


Координаты в трехмерном пространстве

Использование координат элемента как выразительного средства для визуализации данных является одним из самых распространённых подходов. Однако и он требует некоторого анализа для комфортной работы с данными.

- Нормализация данных - для того чтобы они поместились в рассматриваемую область (обычно куб).
- Использование нелинейного масштабирования чтобы избежать скученности элементов, но при этом есть риск искажения данных.

Ответы на эти вопросы сильно зависят от характера визуализируемых данных. Также эти вопросы необходимо учитывать в других методах визуализации рассмотренных ниже

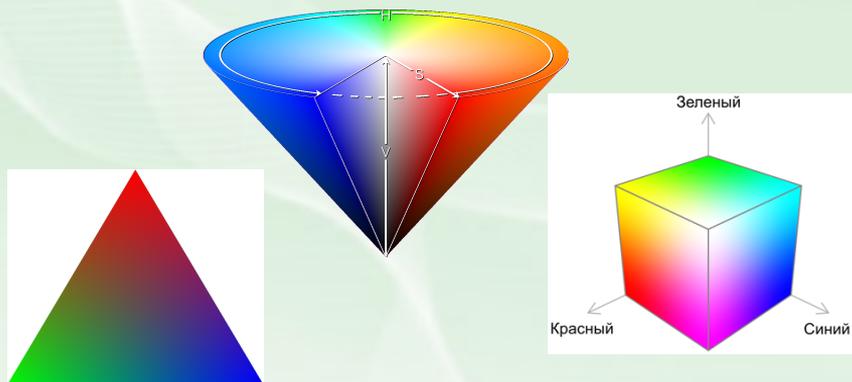


Цвет элемента

Не менее выразительным средством является цвет элемента. Помимо рассмотренных выше вопросов возникают следующие:

- Выбор цветового пространства (RGB, HSV, HSL);
- Количество измерений передаваемых цветом (от одного до трех);
- Проблема совпадения цвета элемента с цветом фона;
- Проблема “агрессивных” цветов.

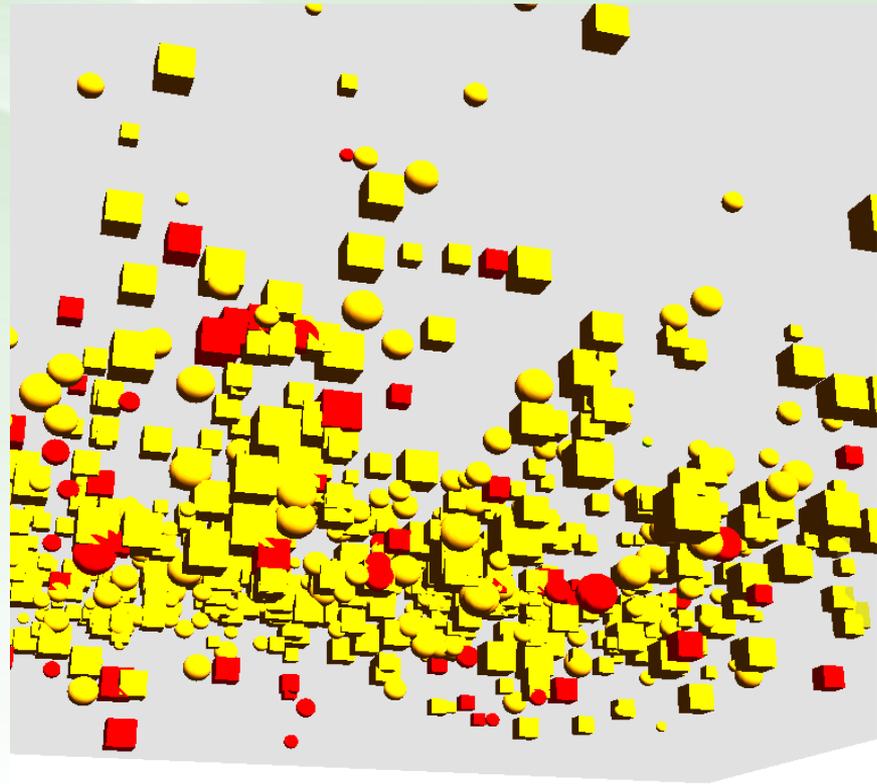
Также важно учитывать пользователей с дальтонизмом, что может сильно ограничивать использование представления данных цветом.



92%	Нормальное зрение	
2.7%	Дейтераномалия	
0.66%	Протаномалия	
0.59%	Протанопия	
0.56%	Дейтеранопия	

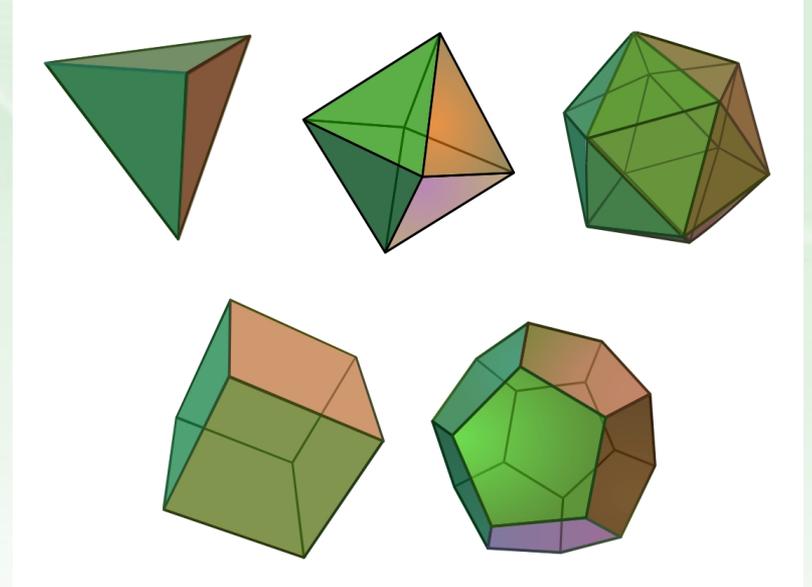
Размер элемента

Изменение размера элемента является менее выразительным средством, чем его положение или размер главным образом из-за ограниченного диапазона изменения этого параметра, так как слишком маленький элемент может быть незаметным, а слишком крупный элемент может перекрывать соседние элементы. Однако в случаях когда пользователю необходимо в первую очередь рассмотреть элементы данных со значениями, близкими к максимальным такое поведение является желаемым. В некоторых случаях для корректной визуализации может потребоваться пересчет значений по выбранному измерению — например, обратно-пропорциональная функция от исходных данных или отклонение от среднего значения.



Форма элемента

Главным ограничением при использовании геометрической формы элемента для визуализации является ограниченное количество легко различимых человеком геометрических форм. В качестве примера такого набора можно рассмотреть правильные многогранники. Также стоит заметить, что и ментально геометрические формы лучше подходят для визуализации качественных измерений разбивающих на некоторые равнозначные группы, чем для визуализации количественных измерений данных.

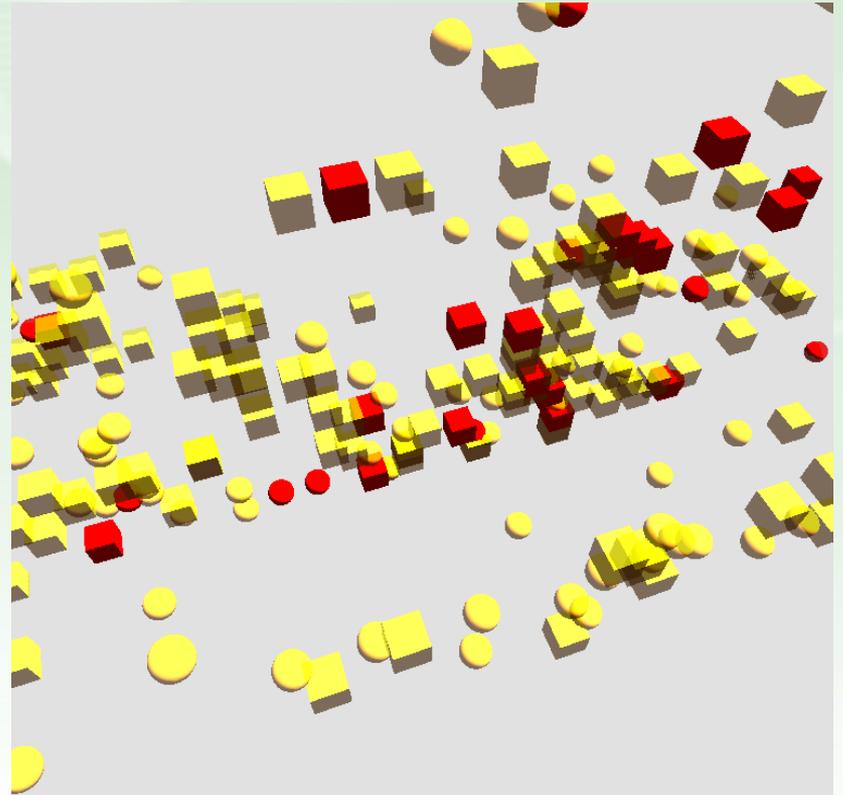


Прозрачность элемента

Изменение прозрачности является нежелательным средством визуализации, так как, подобно изменению размера элементов, может снижать заметность элементов. Однако прозрачность также может приводить к искажению форм и смещению элементов.

Введение полупрозрачных кластеров поверх элементов данных может существенно увеличить выразительность визуализации, при этом степень прозрачности может быть как заданной пользователем, так и рассчитываться по какому-либо параметру кластера, например его разреженности.

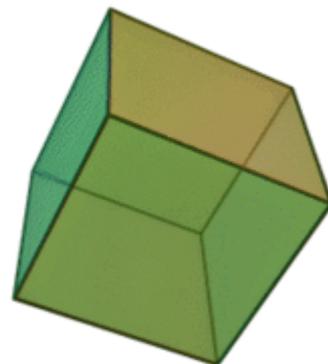
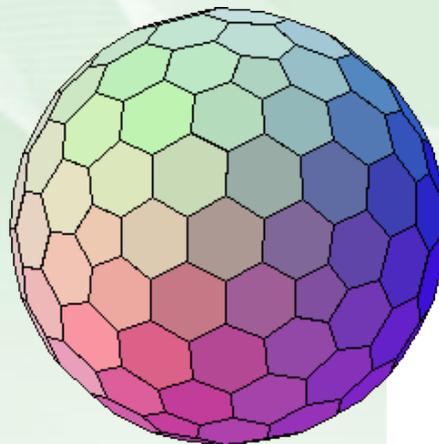
Также полупрозрачность повышает вычислительную сложность визуализации.



Вращение и блеск элемента

Эти подходы при длительной работе вызывают утомление пользователя и другие симптомы cybersickness, особенно при работе в очках виртуальной реальности, поэтому их широкое использование нежелательно.

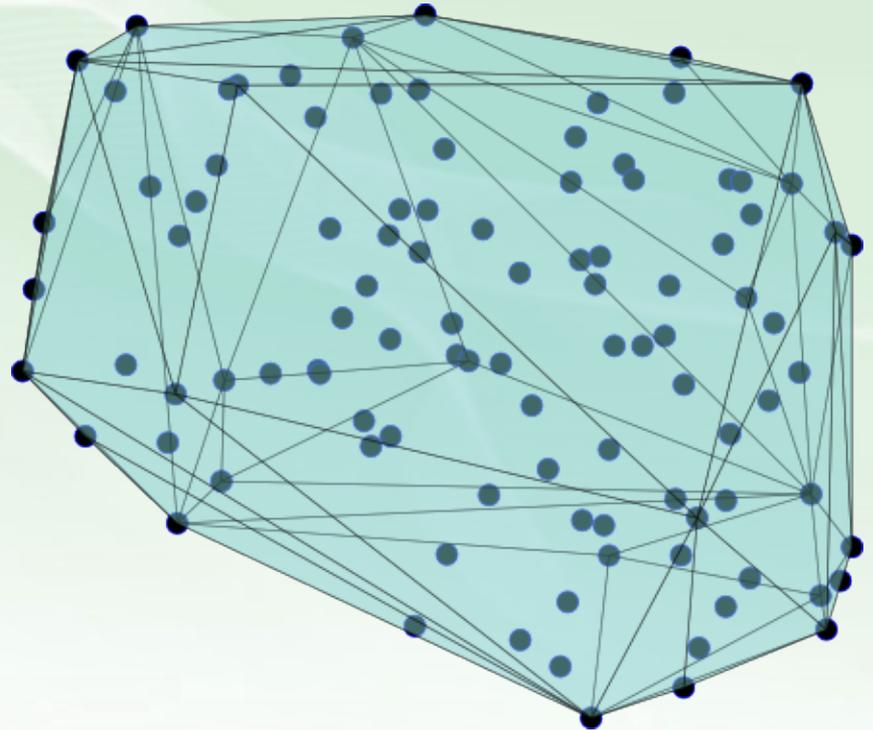
Но данные методы могут быть эффективно использованы для кратковременной подсветки небольшой группы элементов, чтобы привлечь к ним внимание пользователя, не трогая другие визуализируемые измерения.



Кластеры

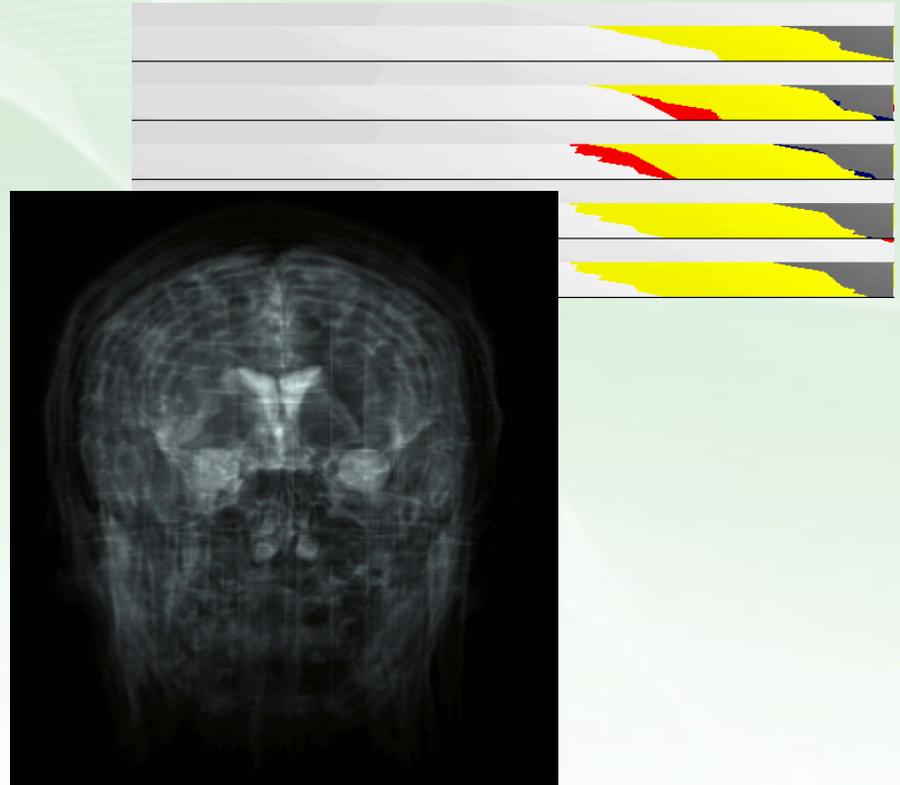
Кластерный анализ может существенно повысить выразительность данных для восприятия человеком. Очевидно что выделение кластеров может быть наиболее эффективно применено к измерениям соответствующим координатам элементов в пространстве, так как при разбиении на кластеры по другим измерениям выпуклые оболочки над кластерами могут быть слишком раздробленными.

Подробнее данный вопрос рассмотрен в статье: «Разработка системы интерактивного визуального анализа многомерных данных»
<http://sv-journal.org/2014-4/08/index.php?lang=ru>

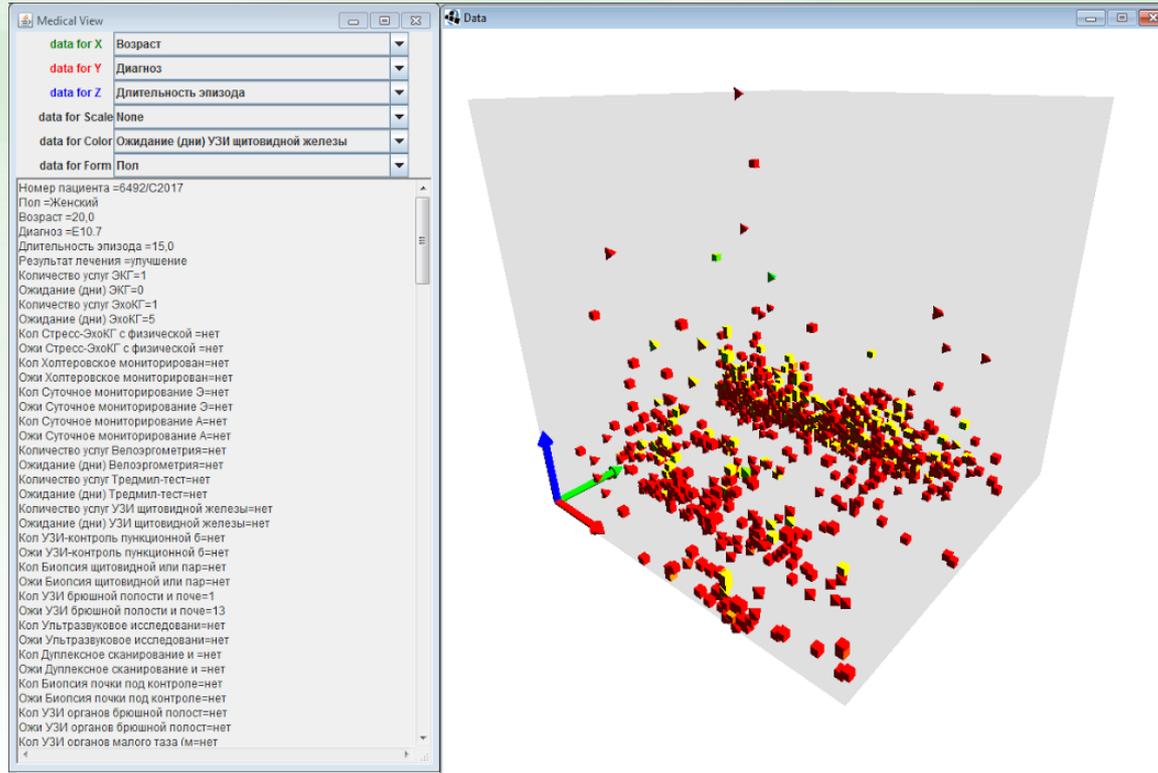


Дискретная и непрерывная визуализация

Рассмотренный подход оптимально подходит для визуализации дискретных данных, то есть тех, в которых можно выделить отдельные несвязанные элементы. Для визуализации непрерывных многомерных множеств значительно лучше подходит объёмная визуализация. Так как в ней отсутствуют отдельные элементы, то исключается передача данных размером и формой элементов, а важность прозрачности возрастает так как внутренние области можно увидеть только применяя срезы или фильтрацию данных.



Визуализация многомерных данных



Выводы

Выводы

Визуализация многомерных данных в общем случае требует как развитого интерактивного пользовательского интерфейса для выбора визуализируемых измерений и способа их визуализации, так и возможности доработки системы визуализации для целей конечного пользователя. Такой подход позволяет визуализировать данные с наибольшей выразительностью существенно облегчая их анализ.

Перспективы

- Интеграция с очками виртуальной реальности;
- Поддержка визуализации непрерывных многомерных данных;
- Добавление внутренних и внешних средств для выделения кластеров;
- Представление изменения данных в течение времени;
- Разработка SQL-подобных запросов к исходным данным для формирования выборки готовых к визуализации данных;
- Создание интерактивного веб-интерфейса.

Спасибо за внимание.
Вопросы?
igormich88@gmail.com