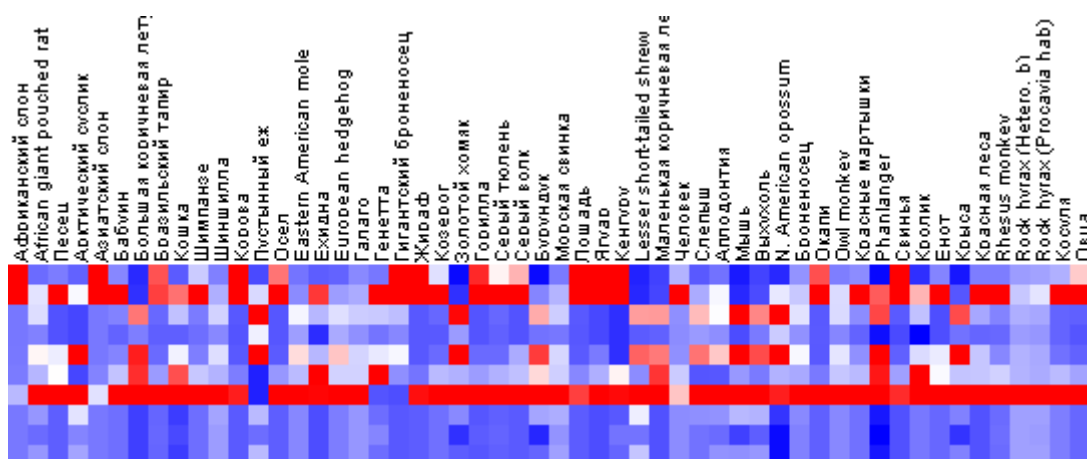


# Визуализация данных

## НСЕ 3.5 – иерархический кластерный анализ данных



В 2002 году профессор компьютерных наук [Ben Shneiderman](#) и его ассистент [Jinwook Seo](#) выпустили первую версию своей бесплатной программы для проведения иерархического кластерного анализа данных (кластерный анализ является одним из разделов статистики касающейся статистической обработки данных) — Hierarchical Clustering Explorer. 11 ноября 2005 года вышел первый релиз НСЕ 3.5. Сегодня рассказ пойдет об основных возможностях и интерфейсе этой программой для визуализации данных. Также будет затронута тема понятия и видов кластерного анализа данных.

С самого начала определимся с терминологией. Кластер — это объединение нескольких однородных элементов, которое может рассматриваться как самостоятельная единица, обладающая определенными свойствами. Так кластером может быть объединение нескольких связанных родственными отношениями людей — семья. Или в качестве кластера вполне может выступать класс, учебная группа или группа контактов в телефонной книге (работа, друзья, девушки). Кластерный анализ — это задача разбиения заданной выборки объектов (ситуаций) на подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались (впервые термин ввел Tryon в 1939 году).

### Кратко об НСЕ

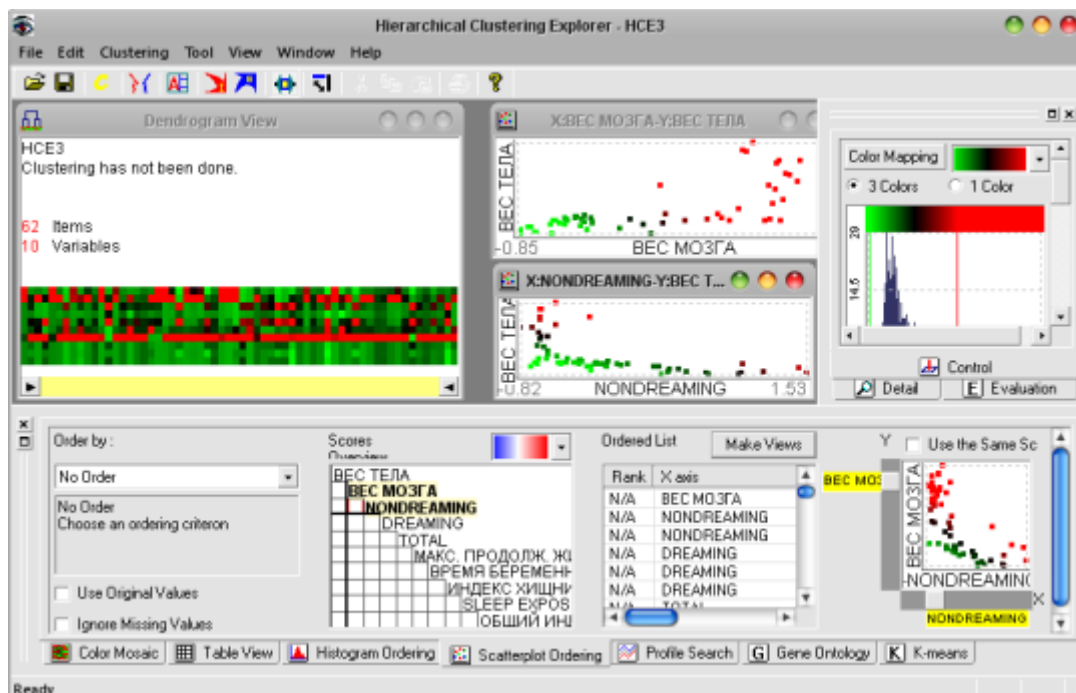
Вес архива с программой — 1,79 Мб.

Разработчики — [Ben Shneiderman](#) и [Jinwook Seo](#).

Условия распространения — бесплатно.

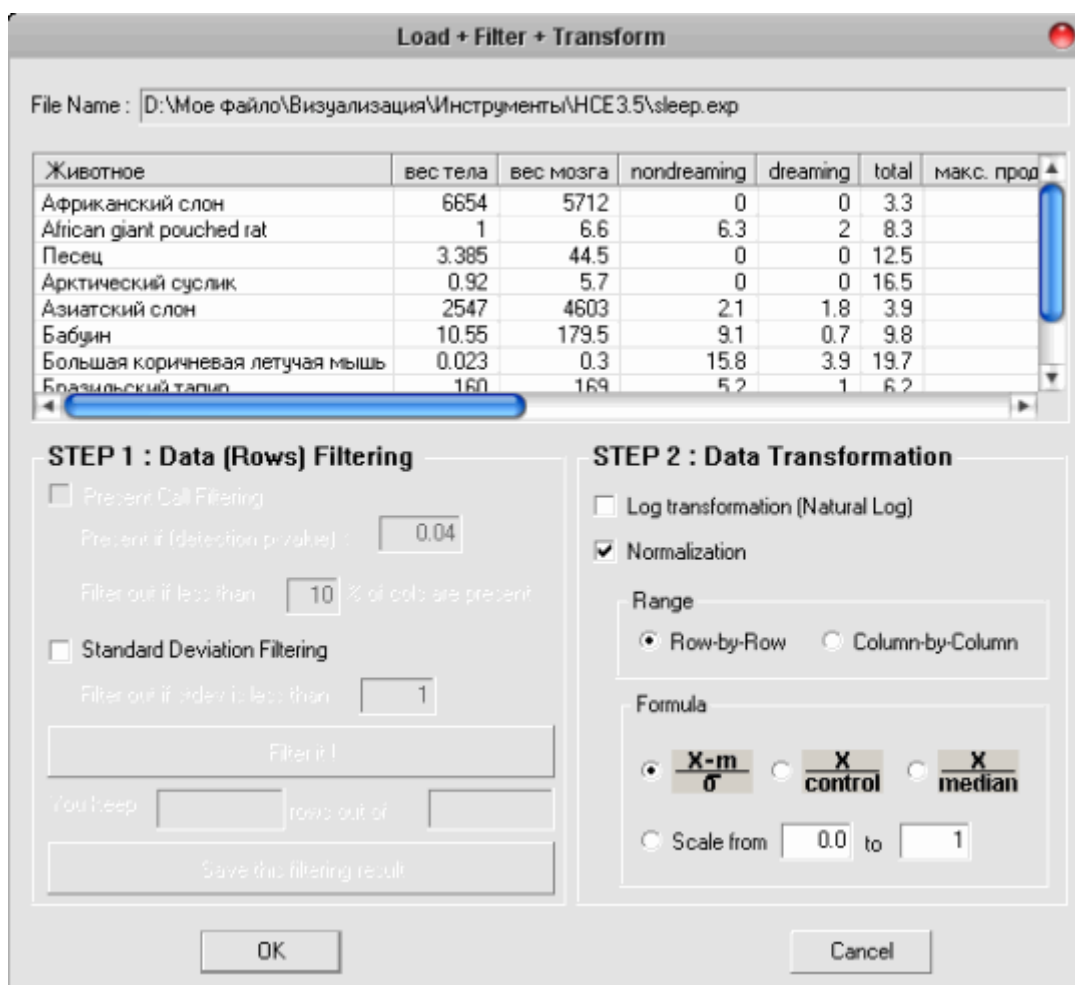
Потребление памяти — 14 Мб оперативной и 7 Мб виртуальной памяти.

Скачать можно с официальной страницы — <http://www.cs.umd.edu/hcil/hce/index.html>.



## Интерфейс

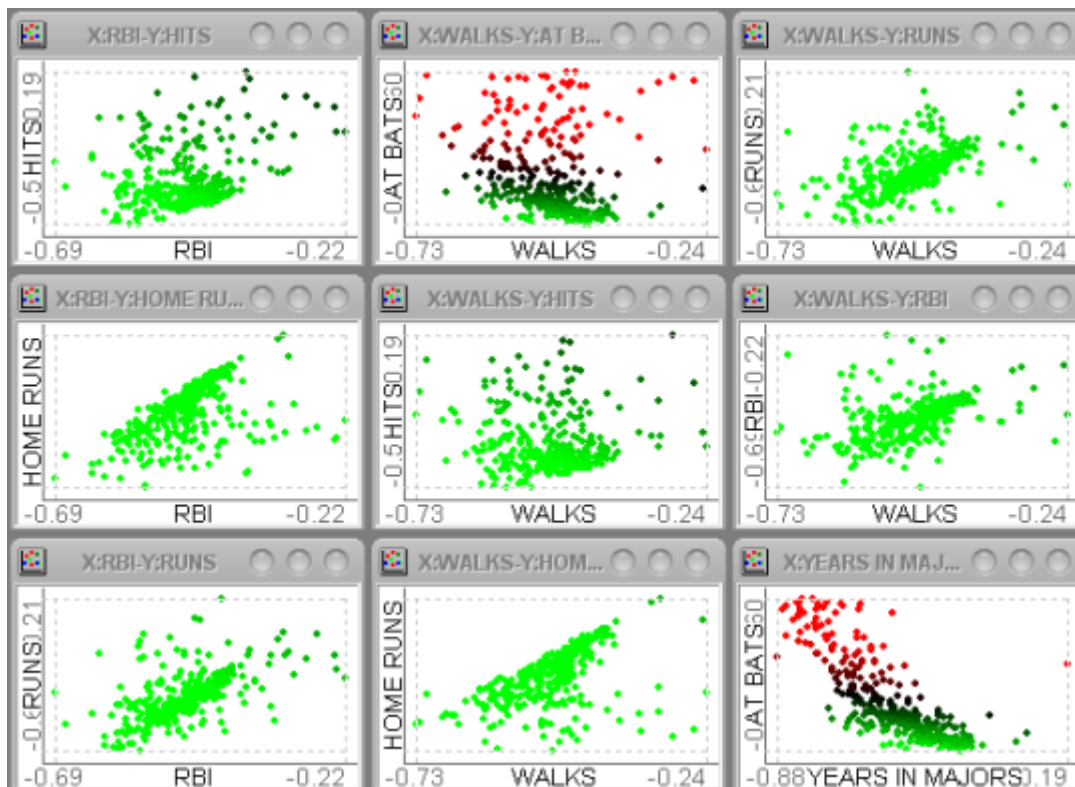
Интерфейс HCE разбит на три области: информационная и контрольная панель, а также основная область. Информационная панель содержит семь вкладок: цветовая мозаика, просмотр данных в виде таблицы, настройка и предварительный просмотр гистограммы, настройка и предварительный просмотр диаграммы рассеивания, поиск по профилю, gene ontology и k-средняя. Во вкладках histogram ordering и scatterplot ordering приведены списки кластеров, характеризующие их параметры и предварительный просмотр гистограммы и диаграммы рассеивания. Основная область предназначена для вывода цветовой мозаики, гистограмм, дендрограмм и диаграмм рассеивания.



## Демонстрация работы

Для демонстрации принципов работы Hierarchical Clustering Explorer откроем файл Sleep. Файл Sleep содержит информацию о нескольких однородных элементах (в данном случае), обладающих определенными свойствами (время сна, вес мозга, «хищнический индекс» и др.).

В процессе работы необходимая визуализация данных настраивается в главном окне по размерам, задается цветовая гамма. При выборе одного из кластеров, указанный объект будет параллельно отражаться на диаграмме рассеивания. При необходимости держать в поле зрения несколько кластеров выделяем нужное правой кнопкой мышки. В настройках программы Tool => Option => Scatterplot задается размер и тип (кружок или квадратик) маркера.



## Вывод

НСЕ 3.5 это прежде всего программа для анализа и однопользовательского исследования данных, ни о какой публикации в веб говорить не придется. Также эта программа создана для профессионалов, которые знакомы со статистикой (К-средняя, дисперсия и др.) и для достижения профессиональных результатов готовы пожертвовать удобствами интерфейса. НСЕ поможет исследовать, находить закономерности и проникать в глубь «численной природы» вещей.