Лица Чернова (англ. Chernoff faces) — отображение многомерных данных в виде человеческого лица, его отдельных частей.

Анализ информации при помощи такого способа отображения основан на способности человека интуитивно находить сходства и различия в чертах лица.

Американский математик Герман Чернов в 1973 году опубликовал работу, в которой изложил концепцию использования этой способности восприятия лица человека для построения пиктографиков. Их применяют, как правило, в двух случаях:

* когда нужно выявить характерные зависимости или группы наблюдений;
* когда необходимо исследовать предположительно сложные взаимосвязи между несколькими переменными.

Основная идея представления информации в "***лицах Чернова***" состоит в кодировании значений различных переменных в характеристиках или чертах человеческого лица. Для каждого наблюдения рисуется отдельное "лицо". На каждом "лице" относительные значения переменных представлены как формы и размеры отдельных черт лица (например, *длина* и ширина носа, размер глаз).

В первоначальном варианте использовалось 18 параметров: размер глаза, размер зрачка, позиция зрачка, наклон глаза, изгиб брови, верхняя граница волос, нижняя граница волос, обвод лица, нос, размер рта, изгиб рта и т.д. В 1981 году Бернард Флури и Ганс Ридвил (Bernhard Flury and Hans Riedwyl) улучшили концепцию и добавили лицам Чернова асимметрию, что позволило увеличить вдвое количество переменных (до 36).

Флури приводит пример удачного многофакторного анализа с помощью лиц. Он проанализировал 100 реальных и 100 поддельных банкнот по параметрам размера границ, отступов и диагоналей. Поддельные банкноты четко выделились в отдельную группу. Таким образом, анализ позволил выявить различающиеся группы объектов.

Асимметрия позволяет рассматривать объекты в прогрессе. Пример показывает различные факторы у пациентов, к которым применялось лечение. Левая сторона лица показывает значения параметров до, а правая - после лечения. Посмотрите как изменилось состояние параметров. Легко можно понять, кому и насколько стало лучше, даже не вникая в сущность исследуемых параметров.

Рассмотрим следующий пример. В этом случае шесть переменных - уровень разводов, процентная доля населения в возрасте 25 лет и старше, по крайней мере, со степенью бакалавра, процент работающих женщин, уровень преступлений на 100 000 человек, процент безработных и средний семейный доход - служили параметрами, указывающими на качество жизни.

Каждая переменная была разделена на три категории. Шесть параметров были сопоставлены для составления лиц Чернова: цвет лица (степень бакалавра), количество волос (семейный дохода), наклон бровей (работающие женщины), кривизну улыбки (безработица), размер носа (уровень разводов) и размер уха (уровень преступности).

Данная карта лиц Чернова конечно, является лишь примером. Необходимо учитывать много факторов, например, как выбрать элемент лица для конкретной переменной качества жизни, сколько классов следует использовать для каждой переменной и как найти лучший способ компоновки лиц.

**Визуализация в R**

Для начала следует установить пакет  [aplpack](http://cran.r-project.org/web/packages/aplpack/index.html)(Another Plot Package).

-После чего нужно загрузить данные (будем использовать выборку «уровень преступности по штатам»). В выборке 8 столбцов. Первый столбец - это имя штата, за исключением строки для среднего значения для США и округа Колумбия. Остальные столбцы - семь категорий преступлений. После чего просто вызываем функцию faces(crime[,2:8]) c нужными параметрами. Получим следующее:

Можно увидеть, что некоторые лица улыбаются. Для более нейтральных наборов данных, таких как качество жизни или бейсбольная статистика, это имело бы смысл. Функция faces () не позволяет нам выбирать, какие элементы ассоциировать с каждым параметром, поэтому нам нужно найти другое решение. Согласно документации изгиб улыбки применяется к шестому столбцу входной матрицы. Заполним шестой столбец в наших данных одинаковым значением. Таким образом, все улыбки будут нейтральными. Получим лица без улыбок.

Теперь можно ассоциировать каждое лицо со штатом.

жестокое преступление:

* Высота лица (убийство)
* высота рта (нападение при отягчающих обстоятельствах)
* форма лица (грабеж)

имущественное преступление

* ширина глаз (кража автомобилей)
* высота глаз (воровство)
* ширина рта (кража со взломом)

-На этом слайде представлена визуализация параметров выборки mtcars. Лица позволяют легко идентифицировать похожие автомобили в наборе данных - например, Honda Civic, Toyota Corolla и Fiat 128.

Алгоритм прорисовки лица занимает немного времени - я бы не рекомендовала его для ввода данных с более чем 500 наблюдениями. На мой взгляд, все, что больше, чем 10 × 10 тяжело визуально классифицировать.

-Лица Чернова для выборки «Ирисы Фишера». Здесь используется 5 параметров (5й – это вид ириса). На следующем слайде уже 4 параметра, а 5й используется для подписи.

-Проанализируем данные по игре в крикет, и попытаемся сравнить 5 подающих игроков. Можно посмотреть на несколько показателей, таких как скорость удара, средний уровень подачи и т. д. Но глядя на эти показатели в таблице, сделать вывод сложно. Можно создать лицо Чернова и сравнить множество показателей среди этих игроков.

Игроки сравнивались по следующим показателям:

1. Средний уровень подачи

2. Скорость удара

3. Количество четырех очков в матче

4. Количество очков, равных шести, за матч

5. Соотношение подач мяча к общему количеству сыгранных матчей

Чтобы создать лица, эти метрики были сопоставлены с некоторыми чертами лица: 1. Высота лица, 2. Изгиб улыбки, 3. Ширина глаз 4. Высота глаз 5. Ширина лица

Как видно, самое счастливое лицо у Sehwag, так как он имеет наибольшую скорость удара, переменную, отображаемую изгибом улыбки. Кроме того, обратите внимание, что у Dhoni очень вытянутое лицо, это опять-таки связано с тем, что среднее значение удара отображается высотой лица, а Dhoni имеет очень хорошее среднее значение.

Еще одно замечание - это ширина глаз как для Dhoni, так и для Yuvraj, очень маленькая, это говорит о том, что у обоих игроков хоть и очень хорошая скорость удара, но они сделали много пробегов, пробегая между калитками.

**Плюсы и минусы**

* Данным способом нельзя проводить количественные оценки. Можно сделать предварительные выводы исходя из построенных изображений, однако для более глубокого анализа требуется большая работа с данными.
* Сложность заключается в правильном сопоставлении исследуемых переменных.
* Даже не смотря на выборку с данными, наблюдатель может идентифицировать уникальные для каждой конфигурации значений наглядные характеристики объектов.
* Лица Чернова нашли широкое применение для анализа ситуации в самых разных областях. Этот метод позволяет быстро оценивать состояние многофакторных систем (10-30-мерные множества), используя врожденную способность человека быстро оценивать изменения в лице человека по многим параметрам: оценка методов лечения, массива статистических данных, оценки режима на атомных электростанциях, в спорте, находить фальшивые банкноты и др.