

# Конспект по теме «Валидация результатов»

## Увеличиваем группы

Если объект исследования является обособленным, никак не связанным с ближайшими соседями, то для анализа лучше подходит **столбчатый график**, так как он подчёркивает изолированность данных. Столбчатый график строят методом `plot()`, параметром передают тип графика:

```
kind='bar'.
```

```
data.plot(y='column', kind='bar');
```

Однако, если объектов - много и не все они обладают достаточно высокими целевыми показателями, можно произвести их группировку для дальнейшей обработки. Группировку в таком случае можно произвести как выборочное переименование.

Выборочно изменяют значения методом **where()**. Ему передают 2 параметра: булев массив и новые значения. Если в булевом массиве True, соответствующее ему значение не изменится; а если False — значение поменяется на второй параметр метода.

```
data['column'].where(s > control_value, default_value)
```

В некоторых случаях, для визуализации данных уместно использовать круговую диаграмму, т.к. она хороша для показа доли каждого значения от 100% всех. Чтобы получить круговую диаграмму достаточно в `.plot` задать `kind='pie'`. При этом нужно задать столбец с данными параметром `y`:

```
data.plot(y='column', kind='pie');
```

## Разбитые по группам данные

Работать с данными в зависимости от определённых значений в столбце по отдельности стандартными методами не всегда бывает удобно. Для того, чтобы работать с данными по группам, существует метод `groupby()`, который автоматически перебирает уникальные значения. Мы передадим ему столбец, а он вернёт последовательность пар: значение столбца — срез данных с этим значением. И в дальнейшем мы сможем обрабатывать эти срезы в соответствии с нашими нуждами.

```
for column_value, column_slice in data.groupby('column'):
    # do something
```