



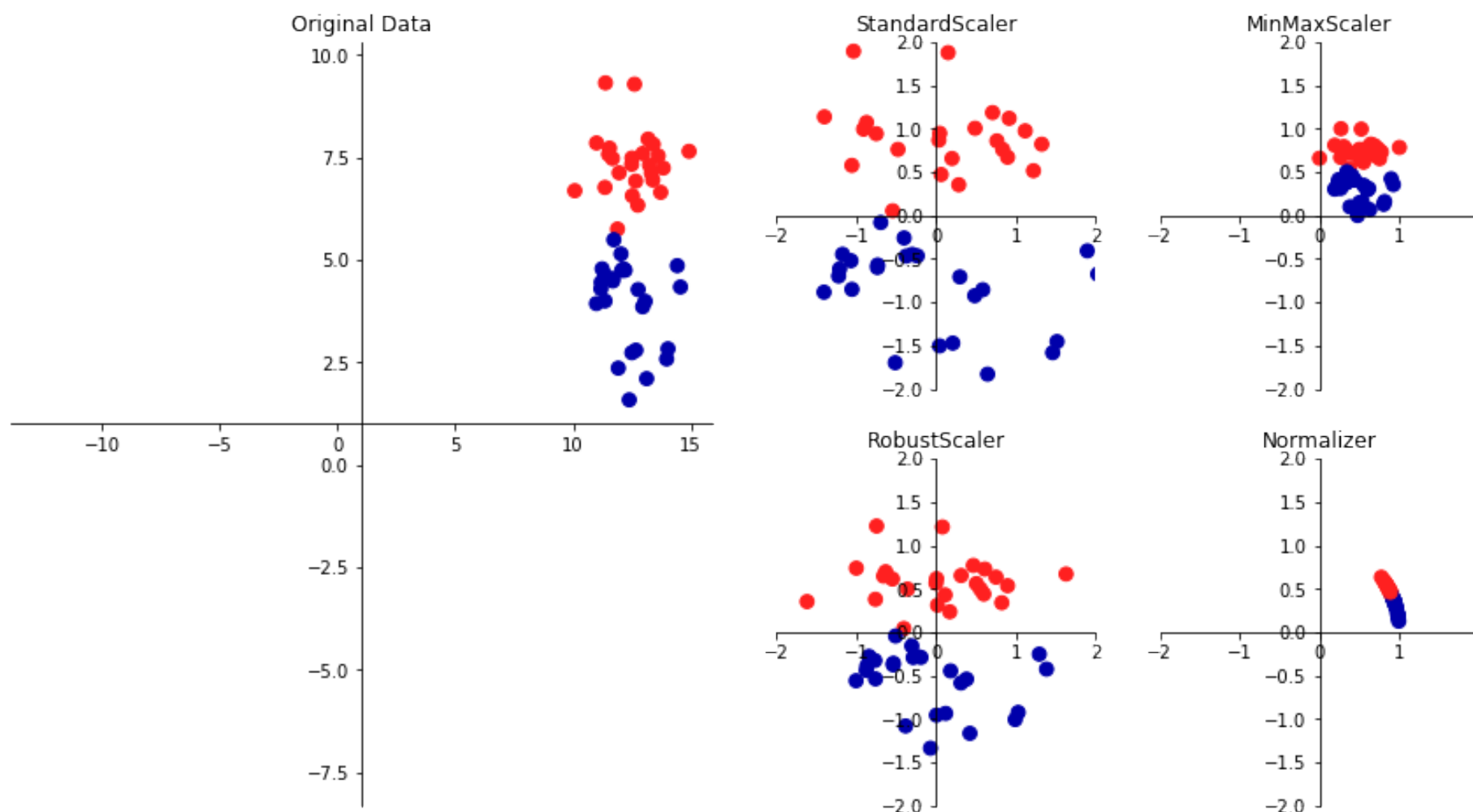
# Введение в машинное обучение (Machine Learning)

проф. кафедры Эконометрики и математической экономики ЭФ  
д.т.н. Шилин Кирилл Юрьевич

РАНХиГС каб. 419/3  
email: [kshilin@ranepa.ru](mailto:kshilin@ranepa.ru)

# Преоброцессинг (перемасштабирование)

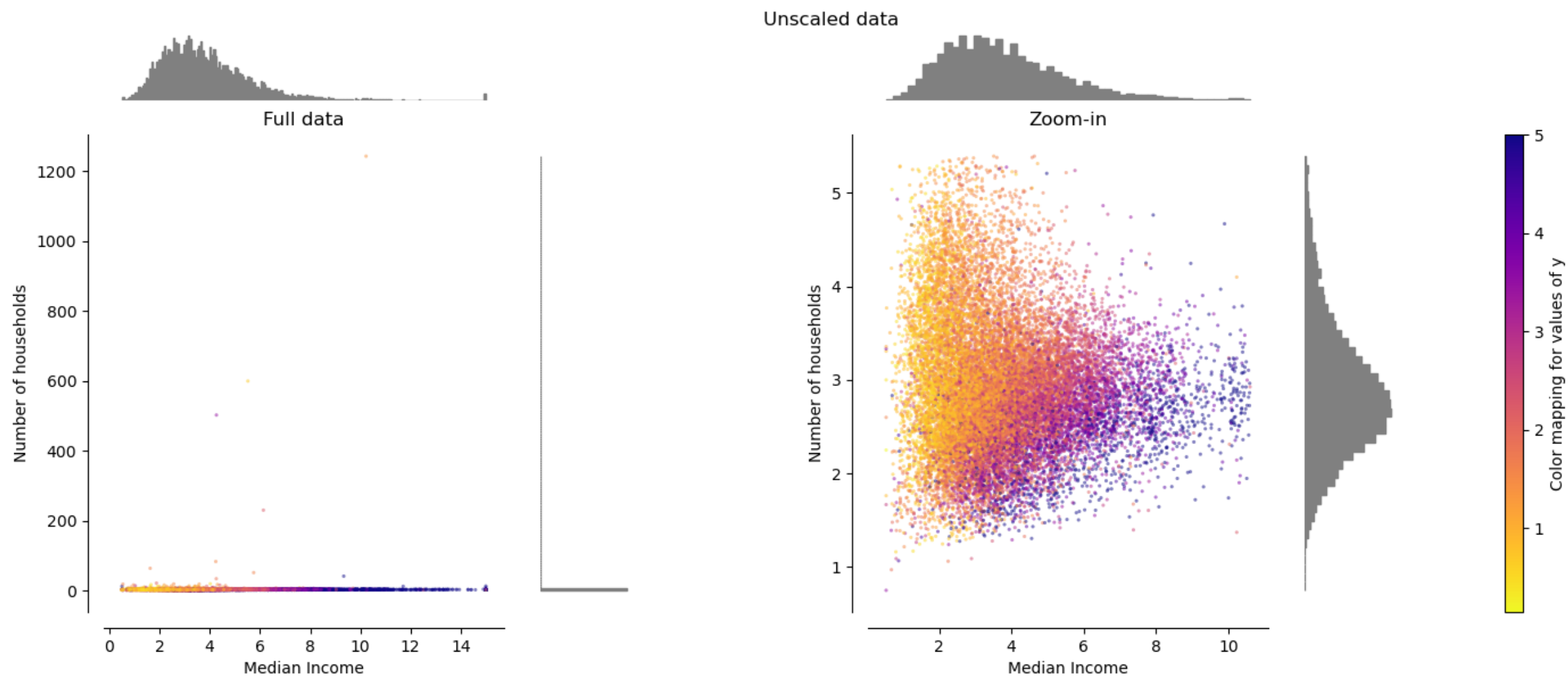
**Рассмотрим основные алгоритмы перемасштабирования для данных в машинном обучении**



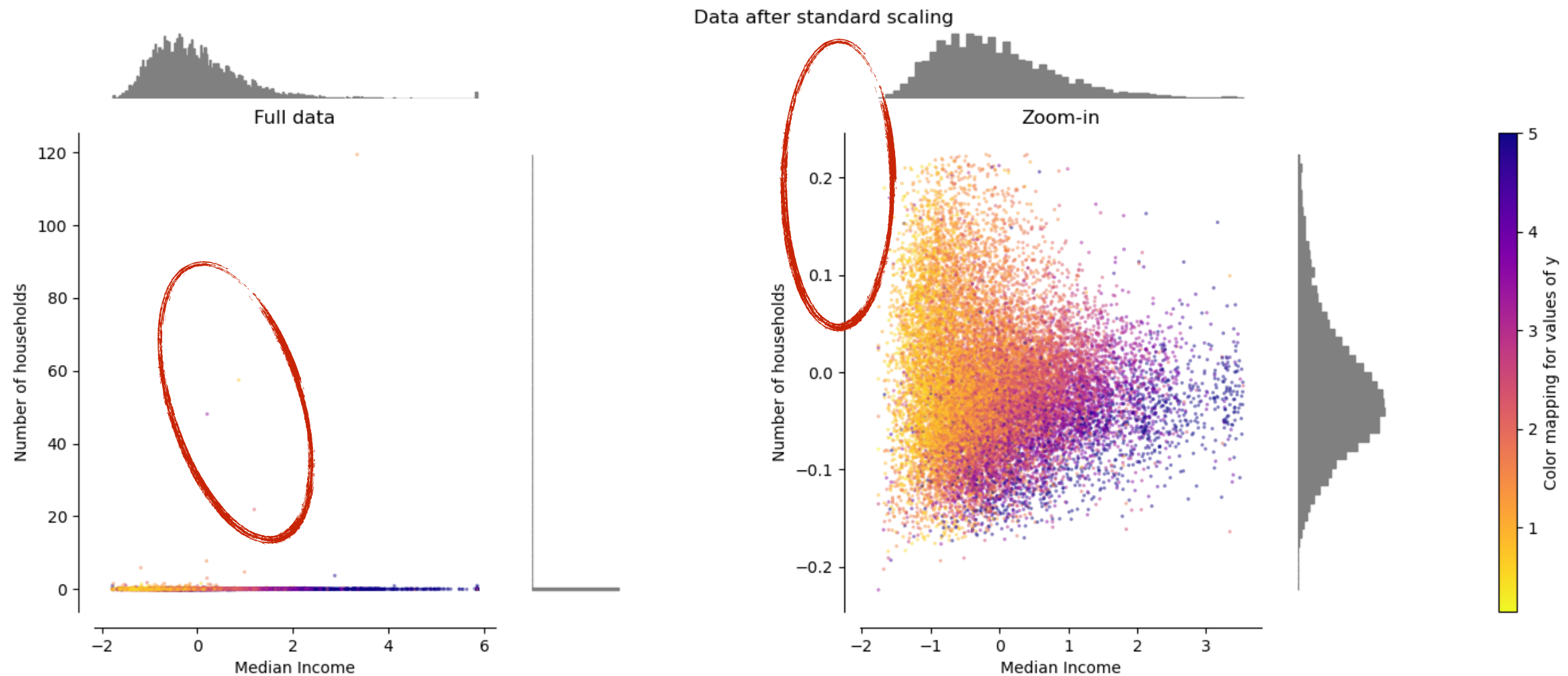
**ВНИМАНИЕ!**

**Масштабирующее преобразование производим ТОЛЬКО на обучающем наборе. Применяем результат и для обучающего и для тестового набора.**

## Оригинальные данные



## StandardScaler

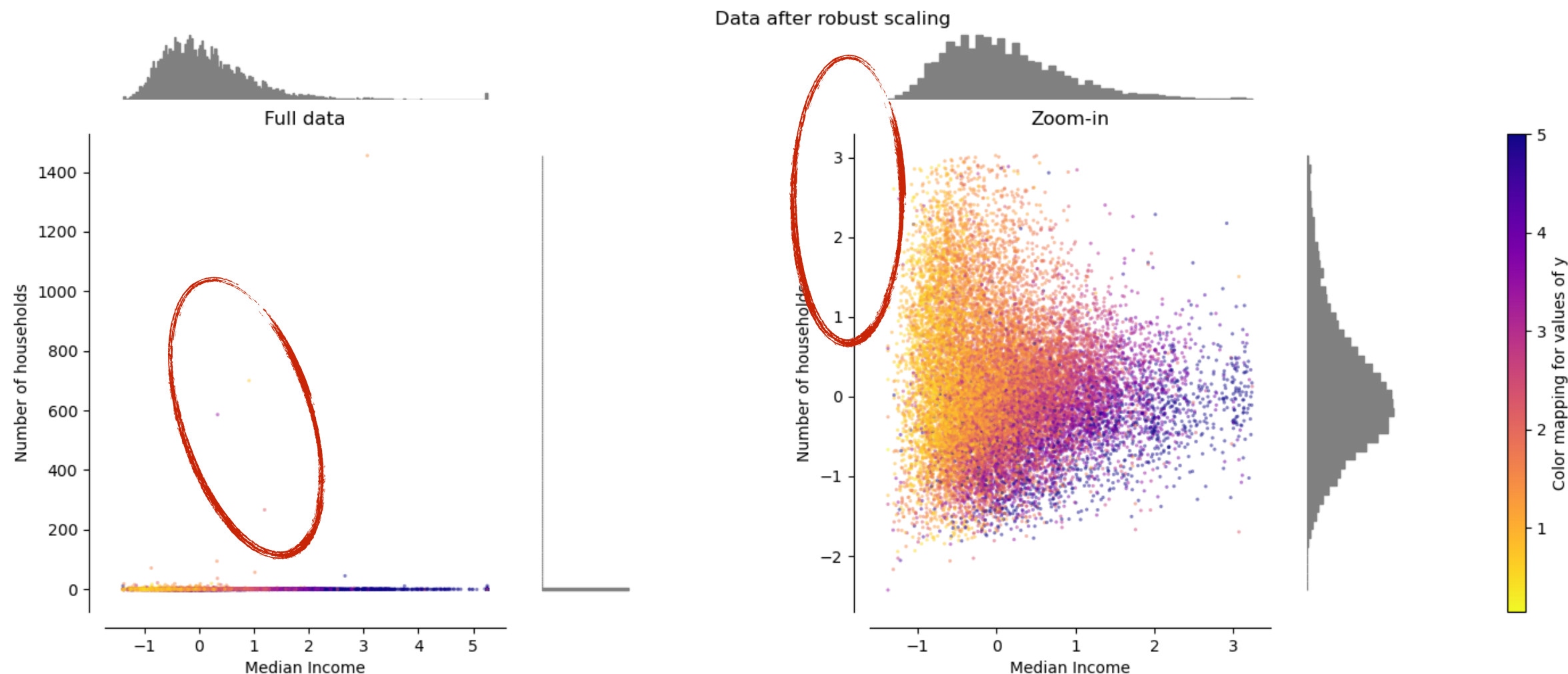


### ВНИМАНИЕ!

Если в данных есть «выбросы» это может стать проблемой, по «умолчанию» - обычно принято считать, что выбросов нет ...

# Алгоритмы

## RobustScaler

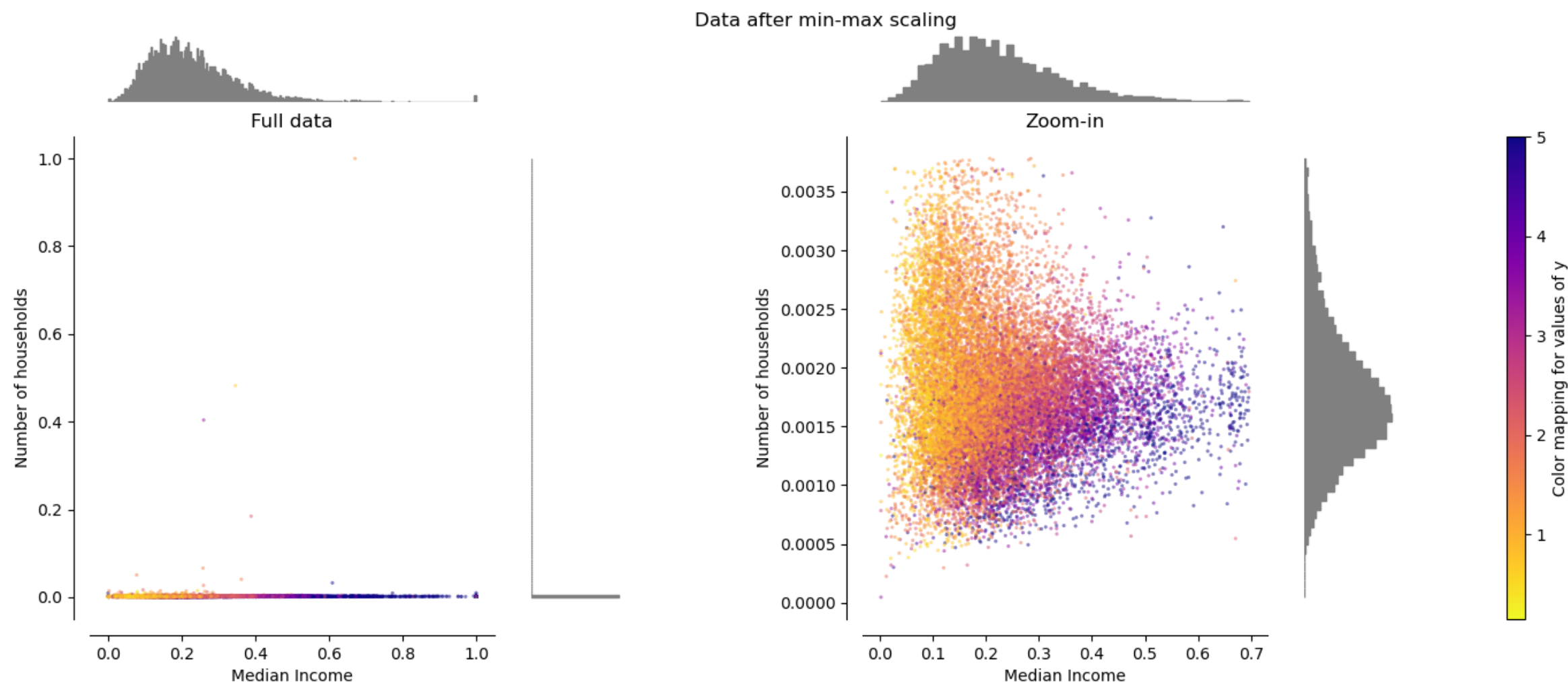


**Этот способ позволяет «избавиться» от выбросов, но выбросы не фильтруются и не удаляются из выборки!**



# Алгоритмы

## MinMaxScaler



### ВНИМАНИЕ!

Если в данных есть «выбросы» это может стать проблемой, используется если нужны только положительные значения.

# Алгоритмы

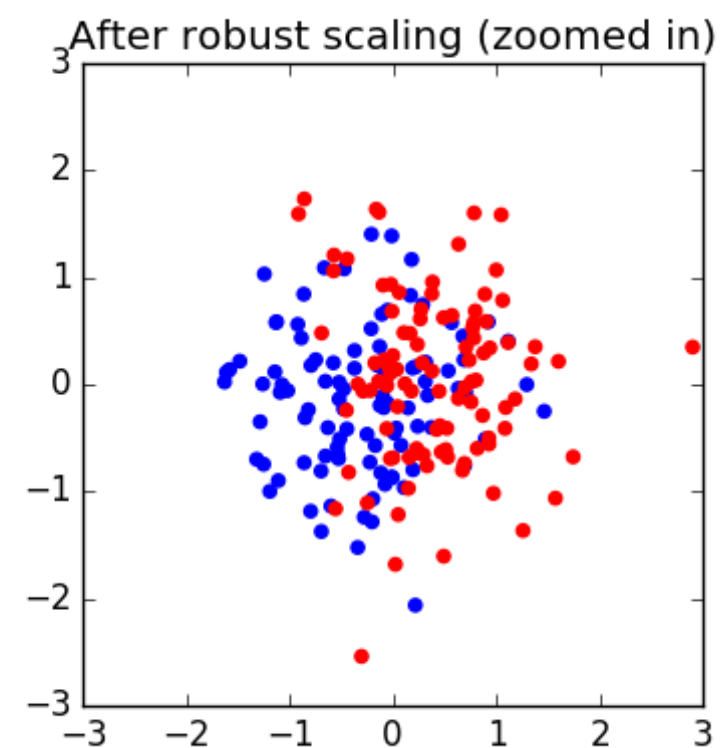
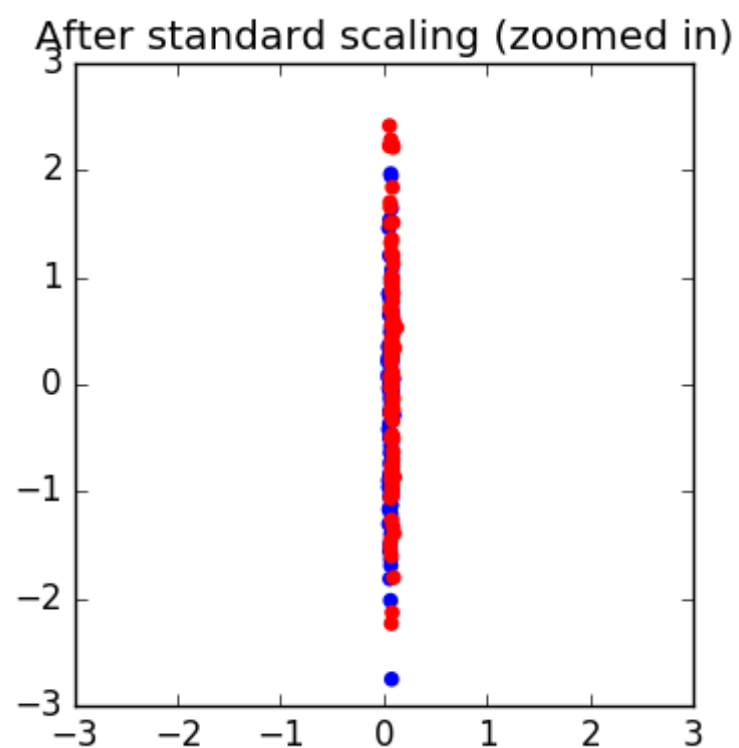
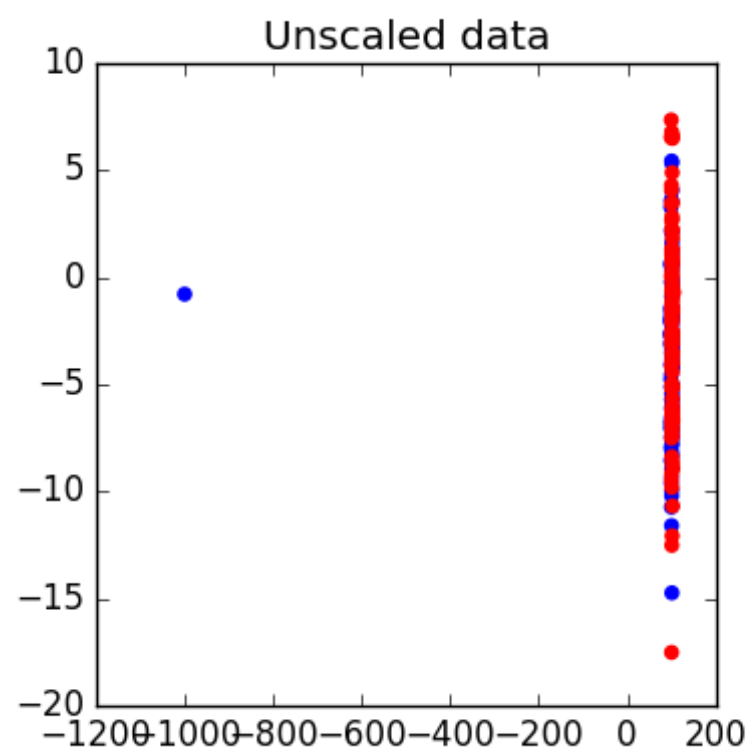
---

## Другие алгоритмы перемасштабирования:

1. PowerTransformer
2. QuantileTransformer (uniform output)
3. QuantileTransformer (Gaussian output)
4. Normalizer¶

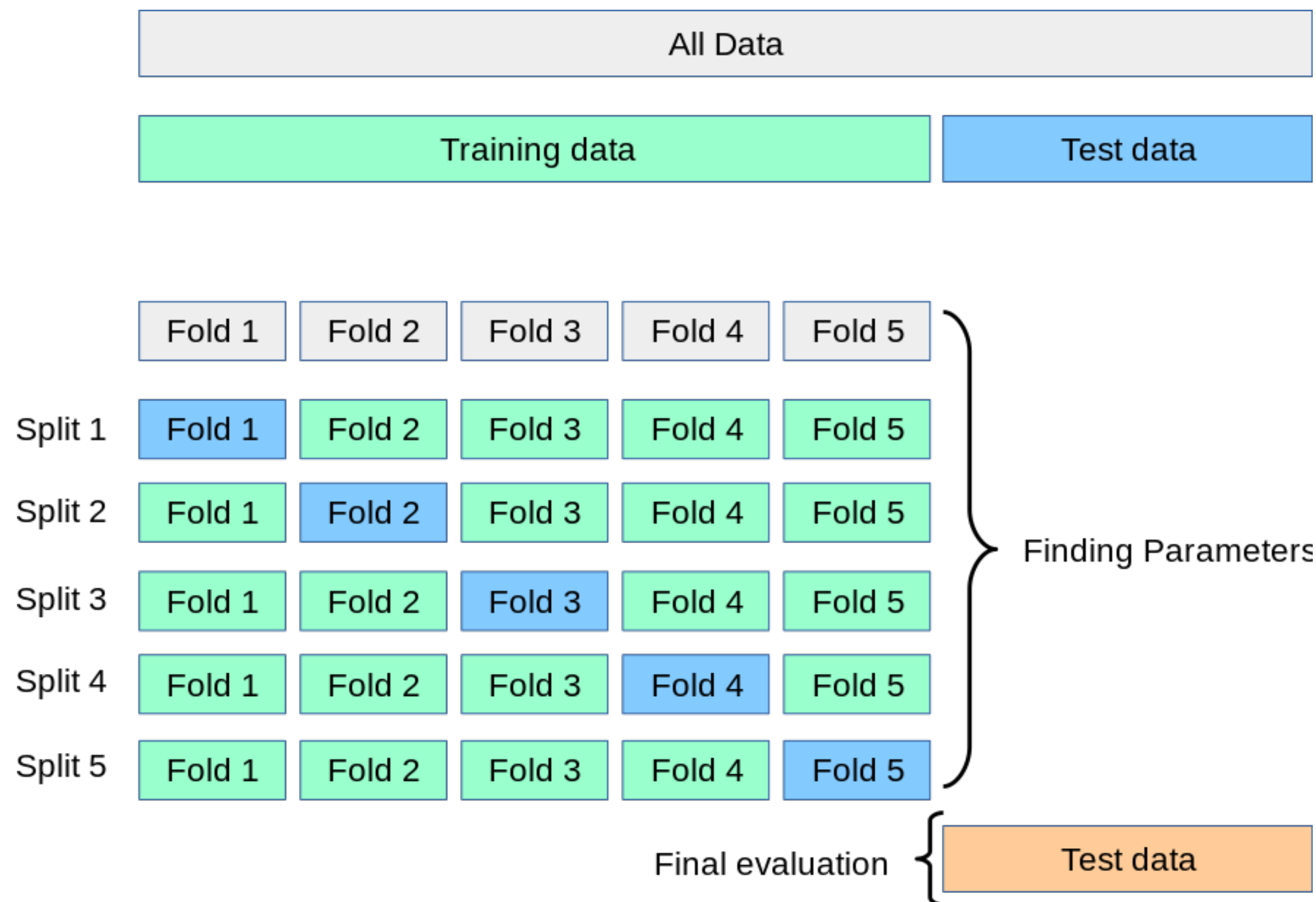
Сравнение алгоритмов и основное описание **можно смотреть тут**

# Проблема «выбросов»

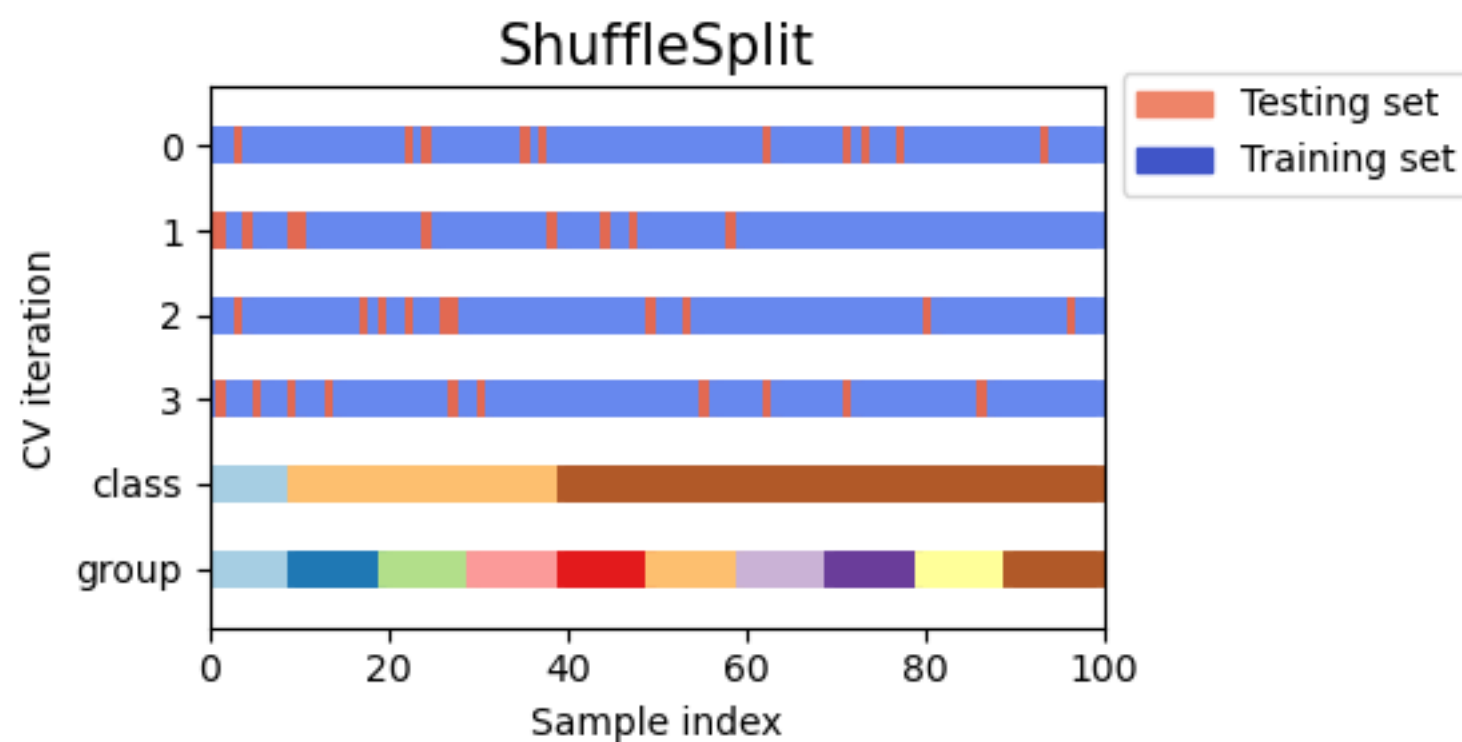
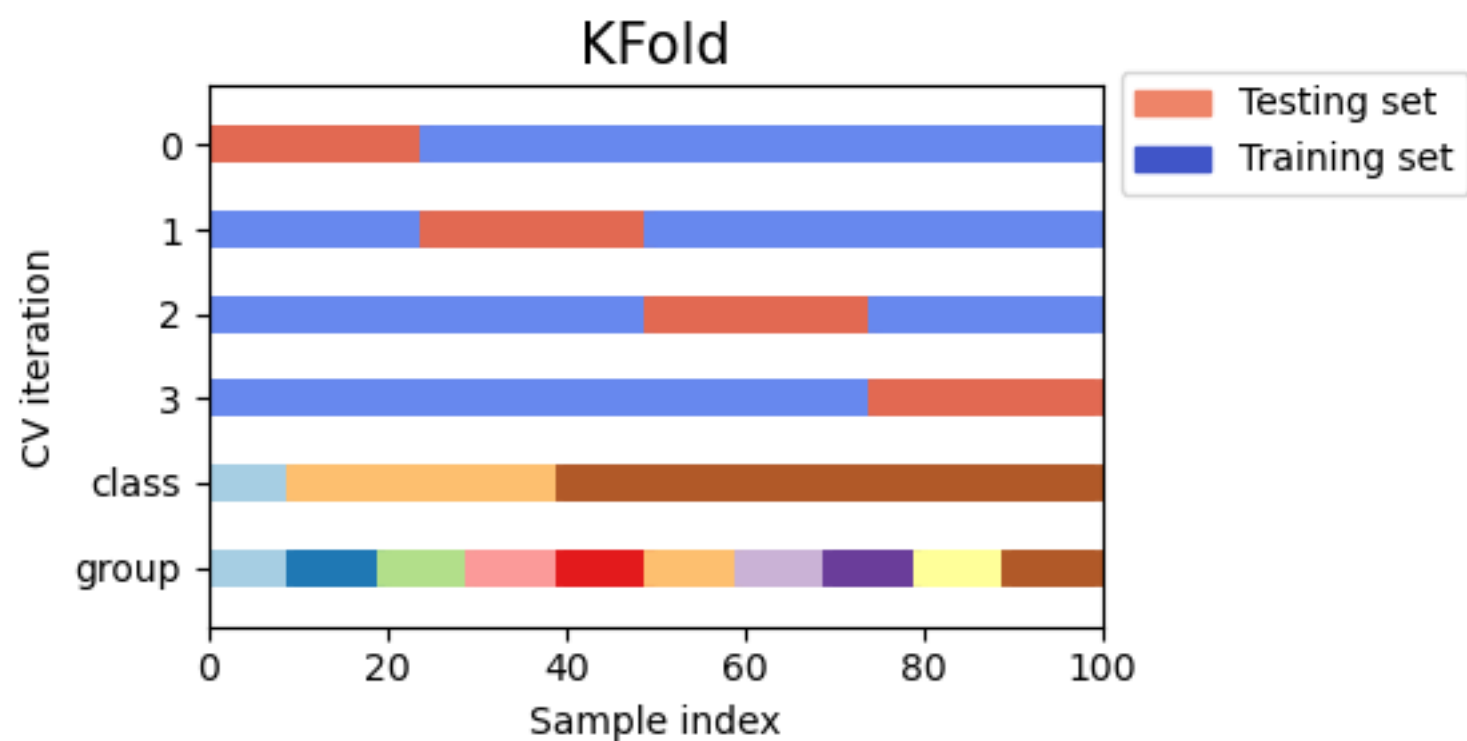




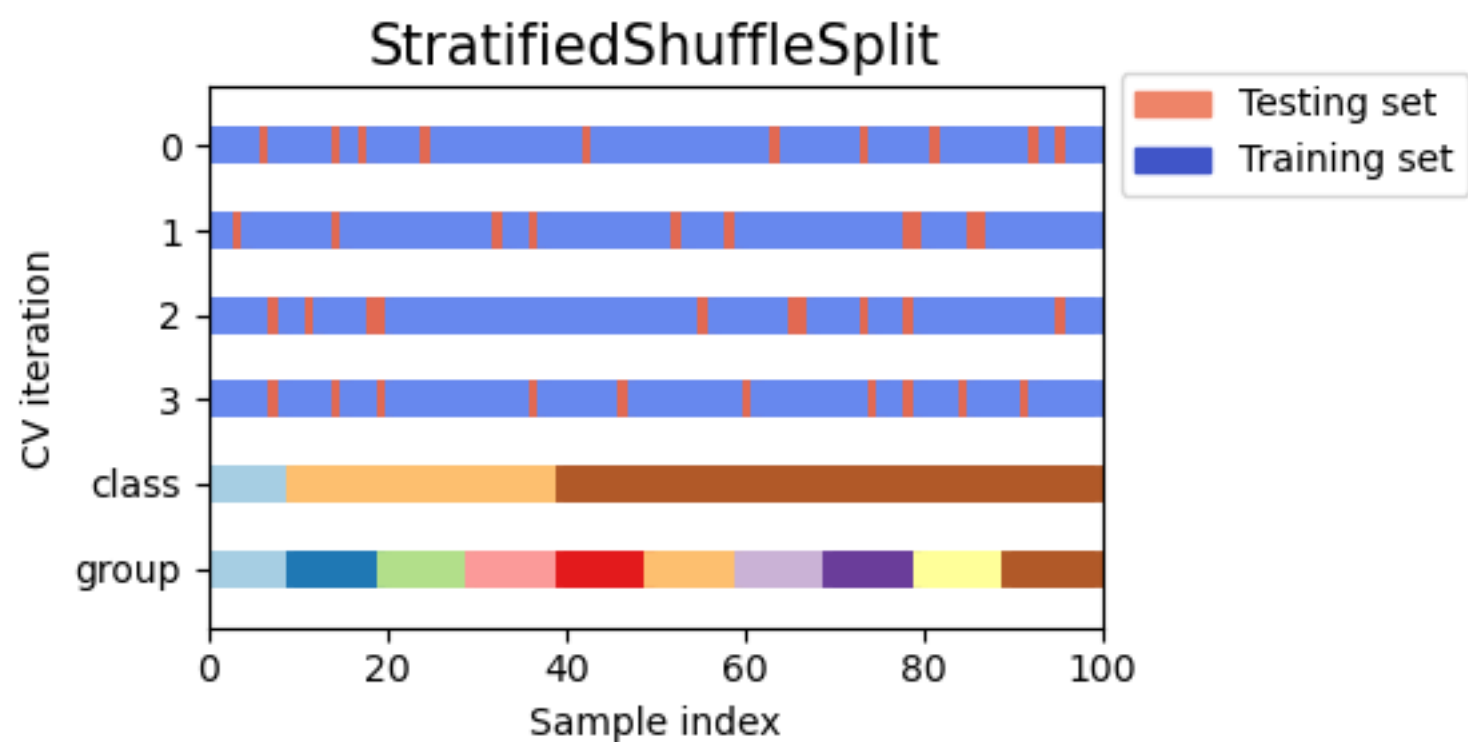
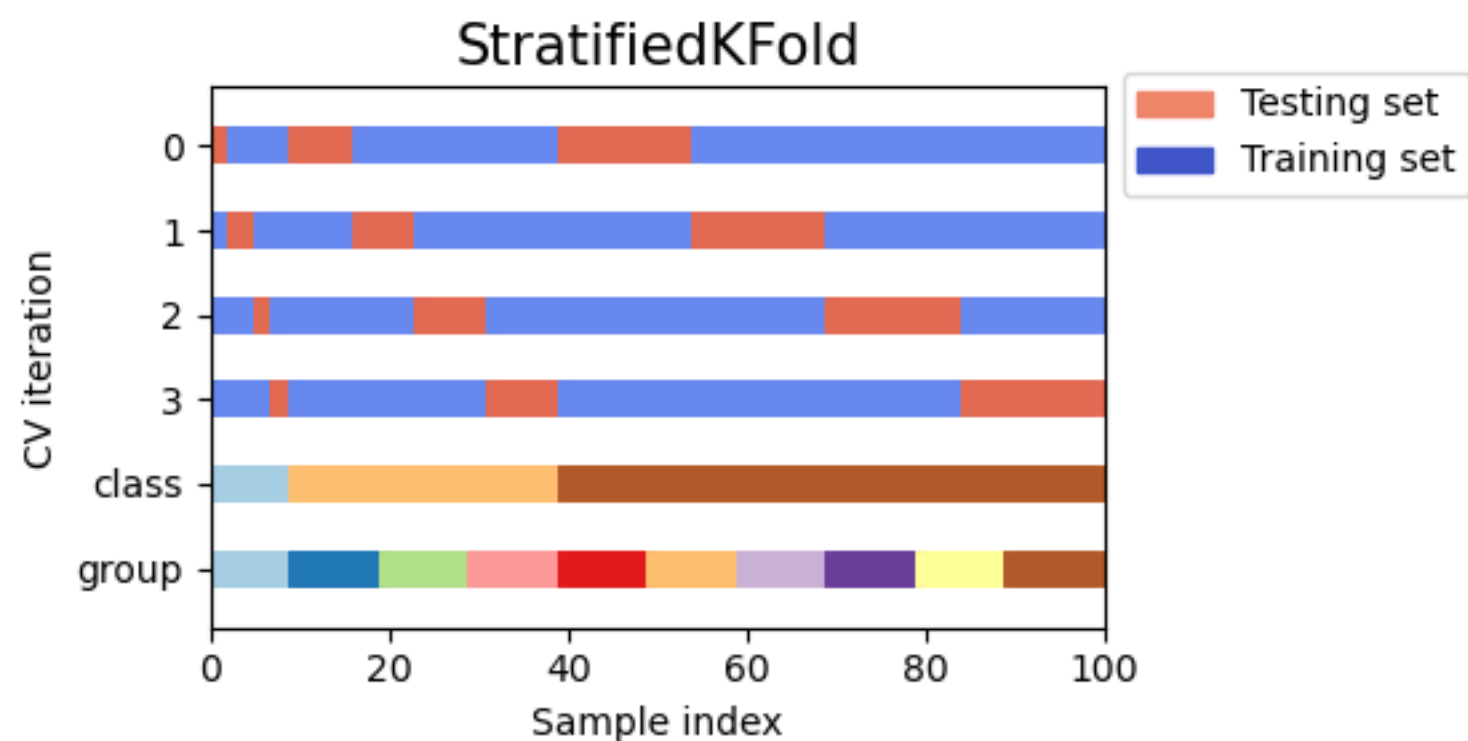
# Кросс-валидация



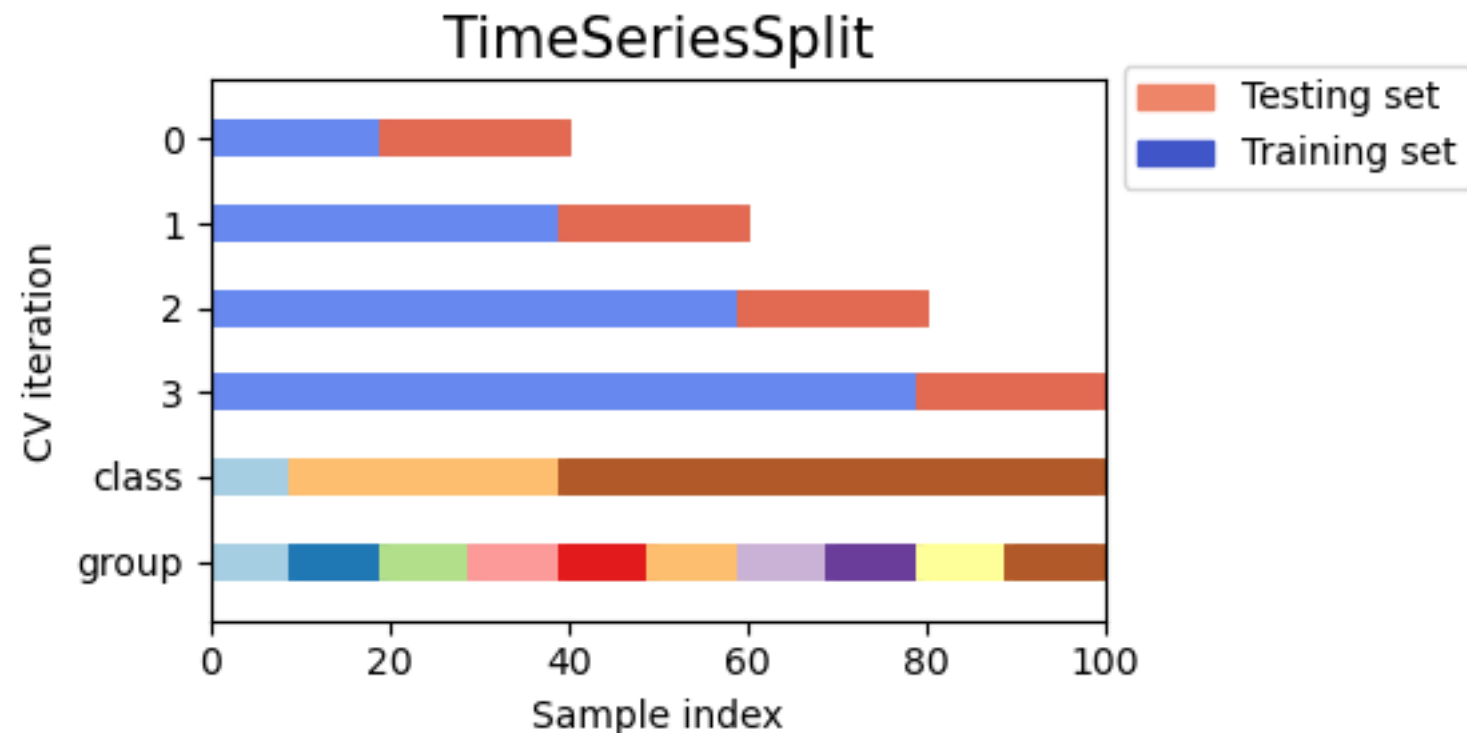
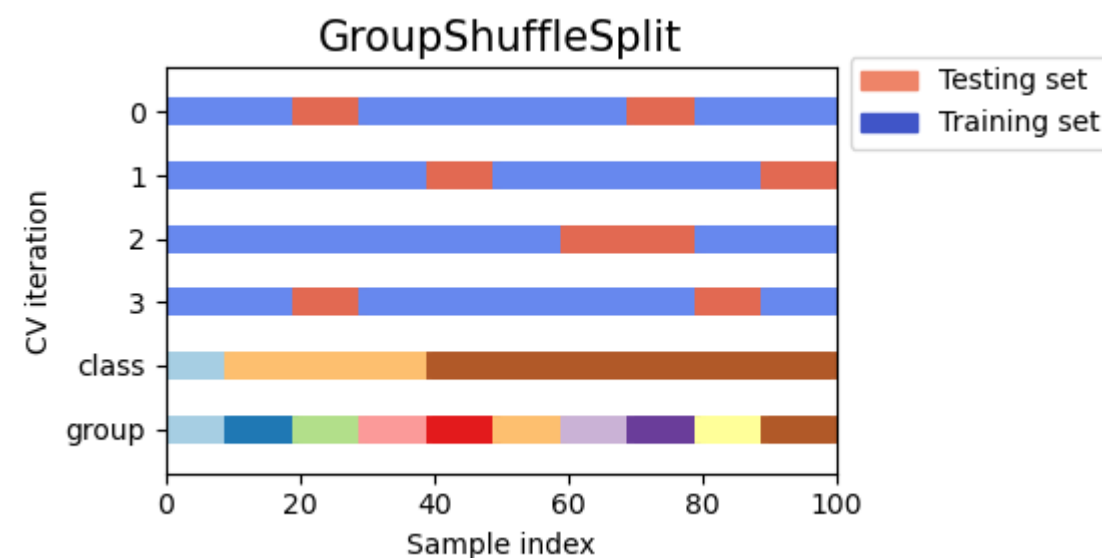
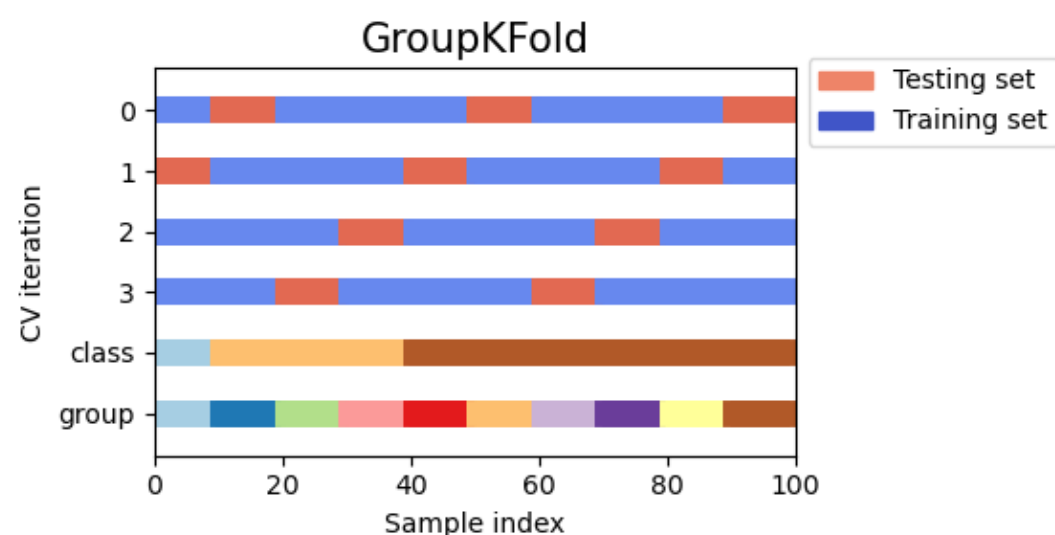
# Кросс-валидация



# Кросс-валидация



# Кросс-валидация



Демонстрацию кросс-валидации [можно посмотреть тут](#)