**Review of data classification methods**

* Manual interval.

Используется для создания собственных классов, вручную добавляя разделители классов и диапазоны классов, которые подходили бы для данных.

* Defined interval.

Используется для определения серии классов с одинаковым диапазоном значений. Количество классов определяется автоматически, исходя из размера интервала и максимального размера выборки. Размер интервала должен быть достаточно мал, чтобы соответствовать минимальному разрешенному числу классов, которое равно трем.

* Equal interval.

Используется чтобы разбить диапазон атрибутивных значений на под-диапазоны равного размера. Это позволит задать число интервалов, а разделители классов будут определены автоматически в зависимости от диапазона значений. Метод равных интервалов наиболее подходит для известных диапазонов значений. Данный метод акцентирует внимание на величине значения атрибута относительно других значений.

* Natural breaks (Jenks)

По классификации естественных границ по Дженксу классы основаны на естественных группах, присущих данным. Границы классов создаются таким образом, чтобы наилучшим образом сгруппировать сходные значения и максимизировать различия между классами. Объекты делятся на классы, границы которых устанавливаются там, где встречаются относительно большие различия между значениями данных. Классификация методом естественных границ индивидуальна для конкретных данных и не подходит для сравнения нескольких карт, построенных на различной исходной информации.

* Quantile.

В классификации по квантилям каждый класс содержит одинаковое количество объектов. Такая классификация хорошо подходит для линейно распределенных данных. Этот метод назначает в каждый класс одинаковое количество данных. Здесь не бывает пустых классов или классов, содержащих слишком малое или слишком большое количество значений. Поскольку при классификации методом квантилей объекты сгруппированы по принципу их одинакового количества в каждом классе, полученная карта может ввести в заблуждение. Похожие объекты могут попасть в разные классы, а объекты с существенно различающимися значениями могут оказаться в одном классе. Минимизировать искажение можно путём увеличения числа классов.

* Geometrical interval.

Геометрический интервал строит границы классов, основываясь на интервалах, имеющих геометрическую последовательность. Геометрический коэффициент в этом классификаторе может измениться (на обратный к нему), чтобы оптимизировать диапазоны классов. Алгоритм создает геометрические интервалы путем минимизации суммы квадратов числа элементов в каждом классе. Это позволяет добиться того, чтобы в каждом диапазоне классов было примерно равное количество значений, и чтобы изменения между интервалами были согласующимися. Этот алгоритм был специально разработан для обработки непрерывных данных. Этот метод сочетает в себе преимущества методов равного интервала, естественных границ (Дженкса) и квантиля. Он позволяет корректно разделить средние значения и выбивающиеся из ряда крайние, позволяя получить результат, адекватный с картографической точки зрения и визуально привлекательный.

* Standard deviation.

Среднеквадратичное отклонение показывает, насколько значение атрибута объекта отличается от среднего значения. Границы классов строятся с равными диапазонами значений, пропорциональными стандартному отклонению (обычно в интервалах один, половина, одна треть или одна четверть), используя средние значения и стандартное отклонение от среднего.

Sources:

Data classification methods—ArcGIS Pro | [Documentation](https://pro.arcgis.com/de/pro-app/latest/help/mapping/layer-properties/data-classification-methods.htm)

Machine Learning: [A Probabilistic Perspective](https://www.cs.ubc.ca/~murphyk/MLbook/pml-toc-1may12.pdf), Kevin P. Murphy.