

Аннотация дисциплины «Программирование и анализ данных»

На геоинформатику можно смотреть с двух сторон. С одной стороны, это раздел географии, в котором методы информатики применяются для решения задач. С другой стороны — это раздел информатики, в котором география является прикладной областью. В курсе «Программирование и анализ данных» мы занимаем вторую позицию: превращаемся в программистов и учимся мыслить о географических задачах в терминах и концепциях, которые можно превратить в программы, понятные компьютеру. Более того, мы смотрим на эти задачи с позиций Data Science — науки о данных. Рассматриваем данные географические — как непространственные, так и пространственные.

Первая половина курса посвящена непространственным географическим данным. Это различные характеристики географических объектов, которые обычно именуется атрибутами: протяженность линий метро, насыщенность почв органическим углеродом, рождаемость по странам, характеристики тропических ураганов и многое другое. Из этих данных можно извлечь много полезного, не рассматривая их пространственную составляющую. Для этого используются преобразования структуры данных (рядов и таблиц), статистический анализ, прогнозирование на основе временных рядов, создание функций, решающих автономные задачи, построение графиков и анимаций и многое другое.

Вторая половина курса сфокусирована на пространственных географических данных, которые мы привыкли называть «пространственные данные». Достаточно глубоко дается техническая реализация привычных моделей данных — векторной и растровой, рассматриваются операции преобразования этих данных, а также автоматизированного создания общегеографических и тематических карт — со множеством технических нюансов, без которых невозможно достичь высокого качества результата. Заключительный блок курса посвящен пространственному анализу с использованием векторных, растровых и сетевых моделей, а также методов интерполяции.

В процессе прохождения курса студенты изучают и применяют на практике современный язык программирования R, который обладает превосходной экосистемой библиотек для обработки данных и пространственного анализа. По итогам курса каждый слушатель становится уверенным пользователем этого языка, а также приобретает полезные навыки программиста-геоинформатика.