

Программа межкафедрального курса
«Визуализация и анализ географических данных на языке R»
Автор программы: ведущий научный сотрудник *Самсонов Тимофей Евгеньевич*

Цель дисциплины:

Приобретение навыков программирования как необходимой компетенции для решения задач обработки данных, их графической визуализации, автоматизированного построения карт и выполнения пространственного анализа в современных географических исследованиях.

Задачи дисциплины:

- Привить базовые навыки программирования, формализации и решения научно-исследовательских задач средствами языка R, заложить основы дальнейшего самостоятельного и углубленного изучения программирования.
- Обучить приемам чтения, структурного преобразования, статистической обработки, обобщения и экспорта табличных данных средствами языка R
- Сформировать представление об основных методах графической визуализации данных, программных средствах построения графиков на языке R и принципах их использования в приложении к табличным данным.
- Дать необходимые сведения о пространственных данных, моделях и форматах их представления; требованиях к пространственным данным для построения карт и их совместного использования с табличными данными; основных источниках открытых пространственных данных.
- Обучить приемам чтения, соединения с табличными данными и визуализации пространственных данных для построения общегеографических и тематических карт, а также серий карт на языке R.
- Сформировать общее представление о методах и задачах пространственной статистики и ее основных разделов, включая: анализ и моделирование пространственной автокорреляции, пространственную интерполяцию и аппроксимацию, анализ и моделирования пространственных распределений.
- Обучить приемам использования методов пространственной статистики для решения географических задачи средствами языка R.

Форма проведения аудиторных занятий:

Традиционные лекции, интерактивные лекции (демонстрации)

Формы самостоятельной работы студентов:

Решение задач на языке R

В результате освоения дисциплины студент должен:

Знать:

Основные принципы программирования, типы, структуры данных, основные операторы и функции в языке R, подходы к их использованию. Различные типы графиков, связь их графических элементов с исходными данными, методы построения и области применения. Базовые понятия и термины геоинформатики в области пространственных данных. Основные источники открытых пространственных данных. Приемы построения общегеографических и тематических карт на языке R на основе пространственных и табличных данных. Основные положения и методы пространственной статистики, ее разделов, пути их реализации на языке R.

Уметь:

Концептуализировать решение географических задач в терминах программирования. Работать с внешними источниками данных (таблицами и пространственными слоями) в среде R, загружать и сохранять их. Осуществлять манипуляции над структурными элементами табличных данных (строками, столбцами и ячейками) для решения задач анализа и визуализации данных. Строить графики по количественным и качественным данным таблиц. Соединять пространственные и табличные данные по ключевым полям. Строить общегеографические карты, а также тематические карты способами значков, картограмм, картодиаграмм и изолиний с послойной окраской в среде R.

Оценивать географическое соседство и вычислять соответствующие пространственные веса, вычислять индекс Морана для оценки пространственной автокорреляции, строить модели пространственной авторегрессии. Вычислять значения статистических функций и тестов для оценки типов пространственных распределений. Осуществлять пространственную интерполяцию и аппроксимацию данных средствами языка программирования R, давать результатам географическую интерпретацию.

Владеть:

Средствами языка R для работы с табличными и пространственными данными, статистического анализа, автоматизированного построения карт, решения задач пространственной статистики.

Содержание

Введение в программирование на языке R

История разработки языка R, его целевая аудитория, возможности и ограничения. Области применения R в географических исследованиях. Знакомство со средой RStudio. Типы данных в R и операции с ними. Понятие о функциях. Структуры данных в R: векторы, матрицы, массивы, таблицы, списки. Операции со структурами данных: создание, наполнение, модификация, обращение к элементам, преобразование. Циклы и условные операторы. Ввод и вывод данных в консоли R. Пакеты R: установка и использование.

Визуализация и анализ табличных данных на языке R

Чтение табличных данных из текстовых файлов и таблиц Microsoft Excel. Работа со строками и столбцами таблиц: создание, обращение, вычисление, наименование, поиск, фильтрация данных. Функции статистического анализа для табличных данных. Построение основных типов графиков: столбчатые диаграммы, гистограммы и эмпирические функции распределения, бокс-диаграммы, секторные диаграммы, диаграммы рассеяния, линейные графики. Основные параметры функции plot (). Работа с графическими параметрами: цвет, размер, толщина, паттерн, прозрачность. Добавление текста на графики. Создание и настройка легенд, осей, рамок и заголовков. Экспорт графиков и таблиц в файлы.

Визуализация пространственных данных и построение карт на языке R

Введение в пространственные данные. Модели пространственных данных и пространственных объектов. Основные форматы пространственных данных. Пакеты R для работы с пространственными данными. Специализированные типы данных `sp` и `raster`. Чтение, преобразование и визуализация растровых данных. Построение цветовой шкалы и шкалы интервалов для визуализации растровых данных. Чтение, преобразование и визуализация векторных данных. Наложение слоев на карте, добавление подписей. Понятие о соединении таблиц, присоединение табличных данных к пространственным. Знакомство с пакетом `lattice`. Построение картограмм и картодиаграмм в R. Построение серии карт. Экспорт карт в графические форматы. Основные источники открытых пространственных данных

Анализ и моделирование пространственной автокорреляции

Понятие о пространственной автокорреляции и сеточных данных (`lattice data`), особенности их анализа. Проблема модифицируемой территориальной единицы (MAUP). Пространственное соседство и методы его формализации. Понятие о весах пространственных связей и методах их оценки, построение матрицы весов. Основные тесты на пространственную автокорреляцию: индексы Морана, Джири и LISA. Построение и реализация моделей пространственной авторегрессии. Пакет R `spdep` для исследования пространственной автокорреляции, и решение задач с его помощью. Географическая интерпретация результатов.

Пространственная интерполяция и аппроксимация

Методы детерминистической интерполяции по данным в точках. Метод ближайшего соседа, обратно взвешенных расстояний, радиальных базисных функций, иерархических базисных сплайнов. Интерполяция на основе триангуляции. Локальная и глобальная аппроксимация тренда. Восстановление непрерывных полей распределения на основе точечных данных. Пакеты R для пространственной интерполяции и аппроксимации. Введение в геостатистику. Вариография. Вариограмма, ее основные параметры. Модели вариограмм. Кригинг и его основные виды. Предсказание значений в точках с помощью кригинга. Географическая интерполяция полей методом универсального (регрессионного) кригинга. Пакет R `gstat` для

геостатистического анализа и решение задач с его помощью. Географическая интерпретация результатов.

Анализ и моделирование пространственных распределений

Основные задачи анализа пространственных распределений. Методы описания внешней и внутренней конфигураций точек. Различные типы оболочек. Основные типы распределений: регулярное, случайное, кластерное. Статистические тесты на случайность. Свойства I (плотность) и II (расстояния) порядка как индикаторы типа распределения. Расстояние до ближайшего соседа, G-функция и F-функция. Расстояние до границы и W-функция. K-функция Рипли, индекс и диаграмма Моришита. Вычисление и анализ графиков данных функций средствами пакета R spatstat. Построение и реализация стохастических моделей точечных процессов. Понятие о кластеризации и декластеризации. Методы кластеризации K средних, ISODATA, DBSCAN/OPTICS. Полигональная и сеточная декластеризация. Вычисление статистик с учетом декластеризации данных. Географическая интерпретация результатов.

Дополнительные главы (для самостоятельно освоения)

Понятие о стандартах описания картографических проекций (PROJ.4, WKT, EPSG). Чтение, задание и трансформация картографической проекции. Интерактивное взаимодействие пользователя и карты, выделение области и выборка объектов. Вырезание данных с помощью операций оверлея. Знакомство с оператором %>% и пакетом dplyr для эффективной манипуляции табличными данными. Знакомство с пакетами ggplot и ggmap для построения графиков и карт. Создание веб-карт с помощью библиотеки rleaflet.

Предполагаемая форма проведения промежуточной аттестации (зачета):

Тест, защита творческого проекта