

Возможности работы с пространственными данными статистического пакета R

[Обсудить в форуме](#) Комментариев — 8

Эта страница опубликована в основном списке статей сайта по адресу <http://gis-lab.info/qa/rspatial.html>

Краткая описание возможностей R по работе с пространственными данными

R - бесплатный, свободно распространяемый и очень активно развивающийся статистический пакет. Благодаря своей гибкости, настраиваемости и открытому коду пакет практически признан стандартом де-факто и используется в огромном количестве проектов, учебных учреждений, компаний. Параллельно основному пакету, энтузиастами разрабатываются сотни дополнительных модулей для каждого направления статистики.

До недавнего времени главной проблемой было наведение "моста" между самими пространственными данными и той массой аналитических возможностей, предоставляемой R. Новые пакеты осуществляющие импорт/экспорт данных максимально облегчили этот переход.

Будучи изначально статистическим пакетом, R предоставляет неизмеримо большие возможности по моделированию и статанализу любых, а в наших случаях и пространственных данных, что делает его интересным инструментом для изучения. Однако, как обычно, необходимо подчеркнуть, что главной фигурой в процессе анализа все же является аналитик, и то насколько он понимает задачу и алгоритм решения как правило будет определять результат.

Задача этой заметки указать и дать некоторое описание функциональности R по работе с пространственными данными. При этом очень сложно избежать дублирования и некоторой избыточности, как сложно их избежать при любой попытке классифицировать средства анализа, управления и создания пространственными данными, особенно одновременно со статистическими алгоритмами, поэтому эта заметка, в угоду упоминанию большего количества модулей к R ставит полноту их описания. Мы оставляем за Читателем необходимость выяснения деталей (как касающихся пространственных технологий, так и статистических методов), но будем стараться создавать более подробные описания по конкретным модулям по мере появления в них необходимости у авторов.

Мы не ставим перед собой невозможную задачу упоминания всех возможных статистических операций с пространственными данными, начиная с гистограмм и заканчивая классификациями типа ISODATA, все это также реализовано в пакете R на базовом уровне и не требует установки дополнительных пакетов. Таким образом главным вопросом на который пытается ответить эта статья является: "Можно ли работать с пространственными данными в R и что с ними можно там делать?". Ответить на которые коротко можно: "ДА" и "ПОЧТИ ВСЕ".

Необходимым напомнить, что использование любого из упомянутых ниже пакетов требует базовых статистических/специальных знаний, а также знакомства с тем как работает R, только в этом случае использование этих пакетов может быть максимально эффективно.

- **Классы пространственных данных:** Так как многие модули сталкиваются с задачей импорта и использования пространственных данных была принята попытка создать библиотеку общих классов и функций построения графиков для таких данных. Задачу унификации пространственных данных, их графического отображения и базовых пространственных операций решает пакет [sp](#). Многие другие пакеты работающие с пространственными данными (такие как [rgdal](#) и [maptools](#)) используют эту библиотеку классов.
- **Импорт и экспорт пространственных данных:** В R можно работать как с векторными так и растровыми данными. Существуют пакеты, позволяющие осуществлять работу напрямую с наборами данных, без

конвертации, например [rgdal](#) - пакет расширяющий возможности R функциями работы с растровыми данными из пакета [GDAL](#) - и векторными из пакета [OGR](#). Этот пакет так же предоставляет поддержку проекционных преобразований векторных данных на базе [PROJ.4](#). Доступ к другим типам векторных данных:

- [maps](#) ([mapdata](#) и [mapproj](#)) - доступ к пространственным базам данных типа S,
- [RArcinfo](#) - чтение бинарных и *.e00 файлов созданных Arcinfo
- [maptools](#) и [shapefiles](#) - чтение и запись шейп-файлов.
- [maptools](#) - дополнительные функции для чтения и записи полигональных файлов в форматах WinBUGS, Mondrian, Stata (команда tmap).
- [spgrass6](#) и [GRASS](#) - доступ к данным GRASS 6 и 5-й версии соответственно.
- **Визуализация**: удобные цветовые шкалы могут создаваться с помощью пакета [RColorBrewer](#), эти шкалы могут изменяться и расширяться с помощью базовой функции R - `colorRampPalette`. Пакет [classInt](#) предоставляет функции помогающие выбрать границы классов для тематических карт.
- **Анализ точечных распределений**: основные функции анализа точечных распределений находятся в пакете [spatial](#) являющемся частью основного пакета R. В дополнение к этому, пакет [spatstat](#) позволяет задавать зону интереса любой конфигурации и makes extensions to marked processes and spatial covariates. Этот пакет отличается хорошими возможностями model-fitting и симуляций, подробно его возможности освещаются на [странице проекта и книге авторов](#). Это единственный пакет который позволяет пользователю to fit inhomogeneous point process models with interpoint interactions. The [splanx](#) package also allows point data to be analysed within a polygonal region of interest, and covers many methods, including 2D kernel densities. The functions for binning points on grids in [ash](#) may also be of interest.
- **Геостатистика**: Пакет [gstat](#) предоставляет широкие возможности однофакторной и многофакторной геостатистики, в том числе для больших наборов данных. Пакеты [geoR](#) и [geoRglm](#) содержат функции геостатистики основанные на моделировании. Подобный широкий набор функций также может быть найден в пакете [fields](#).

Помимо этого, полезными могут оказаться следующие пакеты:

- - [RandomFields](#) симуляция и анализ случайных полей
 - [vardiag](#) - диагностика вариограмм
 - [sgeostat](#) - геостатистика
 - [tripack](#) - триангуляция
 - [akima](#) - сплайн интерполяция
 - [spatialCovariance](#) - вычисление пространственных матриц ковариации
- **Анализ площадных данных**: Пакет [spdep](#) предоставляет базовые функции для построения списков соседства и пространственных весов, тесты на пространственную автокорреляцию полигональных данных, такие как Moran's I, и функции нахождения пространственных регрессионных моделей таких как SAR и CAR, подразумевающих, что пространственная зависимости может быть описана известными весовыми коэффициентами. Пакет [spgwr](#) содержит реализацию метода пространственно взвешенной регрессии для исследования возможной non-stationarity.
- Специализированные задачи:
 - [spsurvey](#) - построения выборок.
 - [trip](#) - доступ и управления данными телеметрии животных
 - [DCluster](#) пакет для идентификации пространственной кластеризации заболеваний
- Экологические исследования (более подробное описание методов - [Environmetrics](#) Task View)
 - [grasper](#) - environmental prediction using GAM
 - [ade4](#) - исследовательский анализ данных и Эвклидовы методы в науках об окружающей среде
 - [adehabitat](#) - анализ выбора местообитаний животными
 - [pastecs](#) - управление, разложение и анализ пространственно-временных серий
 - [vegan](#) - методы ординации

Главным недостатком работы с пространственными данными в R, на текущий момент, является ограничение на работу с большими наборами данных.

Если у вас есть свои интересные примеры использования R в анализе пространственных данных, пожалуйста,

пишите, мы с удовольствием их опубликуем у нас на сайте.

В статье использованы материалы CRAN Task View: Analysis of Spatial Data

Ссылки по теме

- [CRAN Task View: Analysis of Spatial Data](#)
- [Подробнее о пакете sp в R-News](#)

[Обсудить в форуме](#) Комментариев — 8

Последнее обновление: 2014-05-15 01:27

Дата создания: 19.06.2007

Автор(ы): [Максим Дубинин](#)