

УДК 614.84, 004.8

doi: 10.34987/vestnik.sibpsa.2021.47.69.013

ПРИМЕНЕНИЕ ГЕОИНФОРМАЦИОННЫХ ИНСТРУМЕНТОВ ДЛЯ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ ПРИ АНАЛИЗЕ ПРОСТРАНСТВЕННОГО РАСПРЕДЕЛЕНИЯ ПОЖАРОВ

Бабёнышев С.В., к.ф.-м.н.; Бойко Г.М., к.п.н., доцент; Малютин О.С., Матеров Е.Н. к.ф.-м.н.

ФГБОУ ВО Сибирская пожарно-спасательная академия ГПС МЧС России

Аннотация. В статье рассмотрены возможности языков программирования Python и R в применении к геокодированию и анализу географических данных, связанных с пожарами и их последствиями. В работе рассмотрены: нанесение на статические и интерактивные географические карты исходных данных по пожарам, графика плотности пожаров, данных по времени прибытия подразделений, а также построение зон транспортной доступности и оптимальных маршрутов. Используемые аналитические инструменты могут быть полезны при определении порядка привлечения сил и средств подразделений пожарно-спасательных гарнизонов, в частности, при разработке планов привлечения и расписаний выездов сил и средств гарнизонов для тушения пожаров. Приведенная в работе процедура анализа может быть использована, с учетом специфики, практически для любого населенного пункта Российской Федерации.

Ключевые слова: географические данные, район выезда.

APPLICATION OF GEOINFORMATION TOOLS FOR WORKING WITH BIG DATA IN THE ANALYSIS OF THE SPATIAL DISTRIBUTION OF FIRES

Babenyshev S.V., Ph.D. of Physical and Mathematical Sciences; Boyko G.M., Ph.D. of Pedagogical Sciences; O.S. Malyutin; Materov E.N., Ph.D. of Physical and Mathematical Sciences

FSBEE HE Siberian Fire and Rescue Academy EMERCOM of Russia

Abstract. The paper discusses the possibilities of the Python and R programming languages in application to geocoding and analysis of geographical data on fires and their consequences. The paper considers: drawing of static and interactive geographical maps of the initial data on fires, a graph of the density of fires, data on the time of arrival of units, fire areas, as well as the construction of transport accessibility zones and optimal routes. The analytical tools used here can be useful in determining the order of attraction of forces and means of units of fire and rescue garrisons, in particular, in the development of plans for attraction and schedules of visits of forces and means of garrisons to extinguish fires. The analysis procedure described in this paper can be used, taking into account the specifics, for almost any locality in the Russian Federation.

Keywords: geographical data, jurisdictional area.

Введение

В настоящее время для оперативного мониторинга, управления, сбора, анализа и обобщения информации в целях экстренного реагирования по предупреждению и ликвидации чрезвычайных ситуаций используются самые различные информационные средства. Особой важностью для управления обладает географическая компонента этой информации (геоданные), которая при этом может иметь как постоянный, так и динамически меняющийся характер.

В данной статье рассматривается одно из направлений использования геоданных – создание картографических решений с использованием статистических данных по пожарам для исследования порядка привлечения сил и средств подразделений пожарно-спасательных гарнизонов, в частности, указания предельных областей нормативного прибытия (для городских населенных пунктов норматив составляет 10 минут [1]). Данное направление можно рассматривать как продолжение работы, выполнявшейся в рамках НИР «Разработка геоинформационной системы расчета и оптимизации границ района выезда пожарных подразделений с учетом фактического состояния транспортной сети» [2].

Обработка географических данных требует соответствующих программно-языковых и аналитических средств, реализованных, как правило, только в рамках специализированных и, в основном, коммерческих геоинформационных систем. Общераспространенные языки программирования и программные библиотеки обладают достаточными возможностями для обработки и визуализации геоданных. Например, для определения и визуализации оптимальных границ районов выезда в работе [2] использовались языки программирования C# и JavaScript. В данной статье применяются языки программирования R и, частично, Python, входящие в стандартный стек инструментов анализа данных МЧС России.

С другой стороны, настоящая статья входит в цикл публикаций [3-5], касающихся возможностей языка программирования R в применении к задачам природной и техносферной безопасности.

Исходными данными для анализа в статье являются электронные базы данных учета пожаров и их последствий на основе Приказа МЧС России от 24 декабря 2018 года № 625 (вместе с «Порядком заполнения и представления карточки учета пожара»).

Работа выполнена в рамках НИР «Разработка предложений по применению перспективных методов «Аналитики больших данных» в деятельности МЧС России», в результатах которого анализируются методы, применяемые при работе с «большими данными», которые могут быть использованы в интересах МЧС России.

Задачи, решаемые в статье можно разбить на следующие этапы:

1. сопоставление адресов из карточек учета пожаров с географическими координатами (геокодинг);
2. нанесение отметок пожаров на различные географические карты-подложки, анализ участков с наибольшей плотностью пожаров;
3. отрисовка зон транспортной достижимости с использованием алгоритма кратчайшего пути.

Полученная геоинформация может быть использована в планировании и оптимизации распределения ресурсов пожарно-спасательных подразделений, профилактики, а также при рассмотрении чрезвычайных ситуаций иного рода.

Сопоставление адресов из карточек учета пожаров с географическими координатами

К сожалению, электронные базы данных пожаров и их последствий не содержат информацию о координатах, что делает привязку пожаров к карте очень трудоемкой процедурой. Типичная «ручная» процедура включает в себя обращение к геоинформационному сервису с целью выбора подходящего адреса из предлагаемых сервисом вариантов, определение географических координат и ручной перенос в базу данных.

Проблемы с записями, осложняющие переход к автоматическому геокодингу, даже при наличии справочника адресов с геокоординатами, включают в себя следующие типичные случаи:

- Запись адреса в исходном наборе данных о пожарах не имеет строгой формализации. Так, например, один и тот же адрес «Красный проспект 224» может быть записан как «Кр. проспект 224» или «Красный пр-кт 224».
- Запись может содержать различные дополнительные сведения, не относящиеся к географическому расположению объекта, например, номер квартиры или ориентиры на местности.
- В записи могут быть указаны сразу несколько объектов.
- Отсутствие адреса в справочнике, например, из-за изменений в застройке.
- Записи могут содержать опечатки.

Переход к автоматизированной обработке может потребовать предварительной обработки строковых выражений, содержащих адреса, например, средствами библиотеки `regex` в языке программирования Python:

- Необходимо отделить название улицы от номера дома, тем самым проделав предварительную токенизацию текста. В большинстве случаев номер дома является выражением, находящимся в конце строки адреса, отделенный от остальной части символом пробела. При условии, что номер дома записан без дополнительных пробелов, отделить его от основной части достаточно просто.
- Следует удалить наиболее часто встречающиеся фрагменты строки не несущие смысловой нагрузки в контексте ее машинного распознавания, такие как «ул.», «стр.» и прочие.
- Наиболее часто встречающиеся фрагменты строки, представляющие краткие формы записи отдельных слов, важных для корректного машинного распознавания, следует заменить на полные слова. К таким заменам относятся, например, «пр-кт» – «проспект».

Тем не менее, даже тщательная предварительная обработка строковых данных стандартными методами не гарантирует однозначного определения адресов в задаче автоматического распознавания, так как не позволяет учесть весь спектр особенностей, порождаемых свободной формой записи адреса объекта. Реально возможно распознать лишь наиболее часто встречающиеся шаблоны неформальной записи, при этом нет никакой возможности учесть, например, грамматические ошибки.

Следовательно, для сопоставления адресов требуется применение более интеллектуальных инструментов. В работе таковыми стали специальные библиотеки Python `nlTK` (для работы с натуральным языком) и `fuzzymatcher`, позволяющие оценивать разницу различных строковых выражений по расстоянию Левенштейна (редакционное расстояние). Применение этих библиотек позволило добиться однозначной интерпретации более 60% всех записей адресов на случайной модельной выборке.

Таким образом, в общем виде (полу-)автоматическая процедура сопоставления адресов может быть построена следующим образом:

1. Загружается набор данных выездов пожарных подразделений.
2. Сведения об адресе пожара приводятся к виду, который может быть однозначно интерпретирован при машинной обработке.
3. Загружается база координат объектов, расположенных в районе выезда.
4. Осуществляется сопоставление адреса, указанного в базе выездов с адресами объектов из базы координат.
5. При необходимости осуществляется дополнительный прогон, сопоставляя не найденные ранее адреса с использованием более трудоемких, «жадных» алгоритмов.
6. Полученные значения записываются в результирующую базу данных для последующего использования.

Обработка геоданных в R

Для обработки информации по пожарам и последующего анализа координат используется R – современный, динамично развивающийся язык программирования. R показал себя эффективным инструментом для решения широкого круга задач, связанных с географическими данными. Ввод и выполнение команд в командной строке во многих случаях происходит быстрее, чем работа в графическом интерфейсе пользователя настольной ГИС. Для некоторых задач, таких как пространственная статистика и моделирование, R может быть единственным реалистичным способом выполнения работы. Другие языки, такие как Python, Java и C++, могут быть также использованы для геовычислений, кроме того, существуют отличные ресурсы для изучения геовычислений и без R, но ни один из них не обеспечивает уникального сочетания экосистемы библиотек, статистических возможностей, возможностей визуализации, предлагаемых сообществом R.

В R используется большое количество библиотек для анализа геопространственных данных: `sp`, `raster`, `rgdal`, `sf`, `tmap`, `leaflet`, `ggmap` и других.

Сначала мы рассмотрим работу с сервисом OpenStreetMap на примере библиотеки `osmdata`, которая использует Overpass API. Данные, которые мы будем обрабатывать и визуализировать, относятся к количеству пожаров в Новосибирске за последние несколько лет.

В качестве первого примера, загрузим в R данные по пожарам в Новосибирске, включающие период с 2016 года по 2020 год. Для дальнейшего анализа оставим только значимые переменные: дату пожара, категорию объекта, географические координаты, время прибытия, площадь пожара.

Отметим неоднородность исходных данных (см. рис. 1), а именно, в последние два рассматриваемых года учитывалось гораздо больше пожаров, что видно из графика ниже.

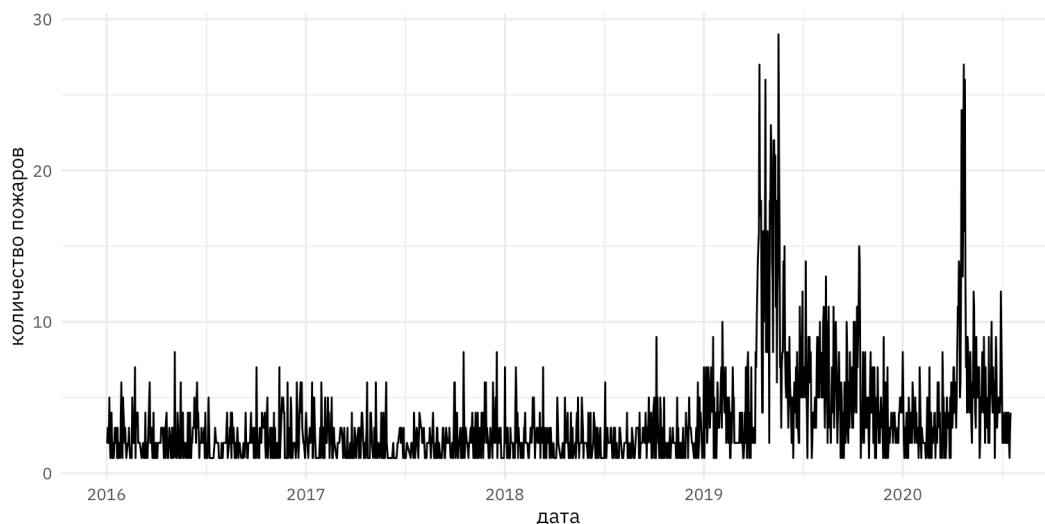


Рис. 1. Количество пожаров в сутки в г. Новосибирске (2016-2020 гг.)

Действительно, в 2019 и в 2020 годах на графике наблюдаются существенные всплески количества пожаров, что можно объяснить переходом на новую законодательную базу с 1 января 2019 года, при котором понятия пожар и загорание были совмещены. Для устранения неоднородности можно сделать фильтрацию данных, исключив такие объекты горения как мусор вне территории жилой зоны, сухую траву, и некоторые категории, например, полосы отчуждения и обочины дорог, но мы этого здесь делать не будем.

На этапе предварительной подготовки, с помощью библиотеки `osmdata` загрузим данные из OpenStreetMap по интересующему нас городу (Новосибирск): крупные и мелкие улицы и реки.

Затем, используя библиотеку `ggplot2`, нанесем на карту точки, где произошли пожары в рассматриваемый период (см. рис. 2).

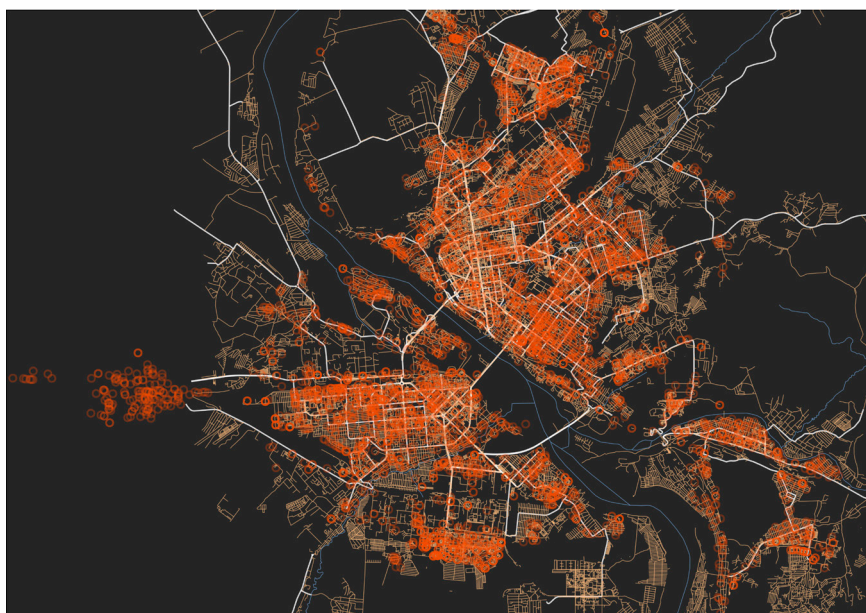


Рис. 2. Визуализация пожаров (2016-2020 гг.) на карте города Новосибирска

Предыдущий рисунок не позволяет сделать качественные выводы об интенсивности в силу наложения точек друг на друга. Для более глубокого анализа хотелось бы получить представление о проблемных участках с максимальным количеством пожаров. Визуально выделить участки с наибольшим количеством пожаров можно, разделив всю область значений на равные интервалы и использовать различную интенсивность цвета для каждого уровня. Возможности библиотеки ggplot2 позволяют разделять участки с различной плотностью пожаров. Полученный результат представлен на рисунке ниже (см. рис. 3). Интенсивность цвета показывает количество пожаров в охватываемом контуре.

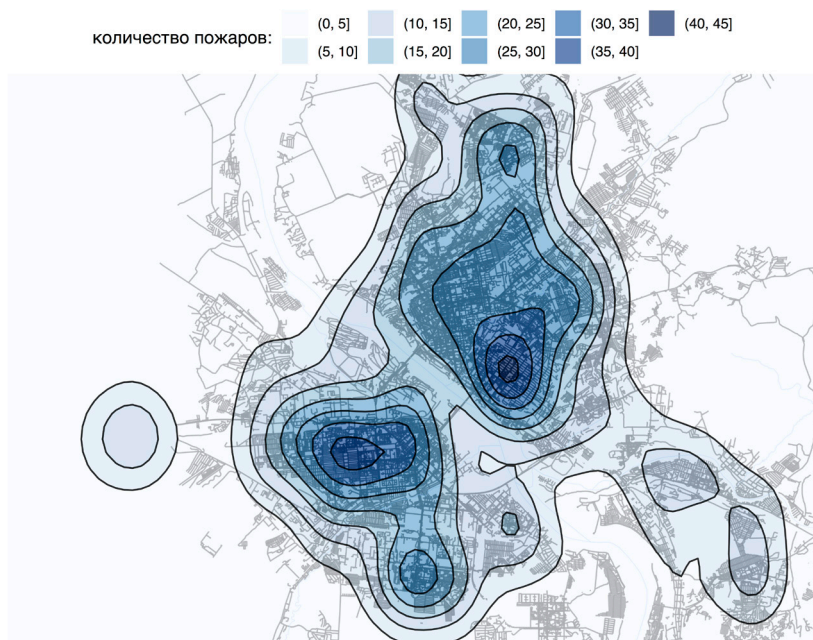


Рис. 3. Плотность пожаров (2016-2020 гг.) на карте города Новосибирска

График плотности пожаров может быть полезен для определения степени пожарной опасности районов города, как при предварительном планировании боевых действий по тушению пожаров и проведению АСР, так и в профилактических целях.

Далее рассмотрим на карте данные по прибытию пожарно-спасательных подразделений, выделив только те события, где время прибытия составило более 10 минут. На рисунке ниже (см. рис. 4) показаны некоторые основные пожарные части города Новосибирска, где точками красного цвета выделены пожары, время прибытия до которых составило более 10 минут.

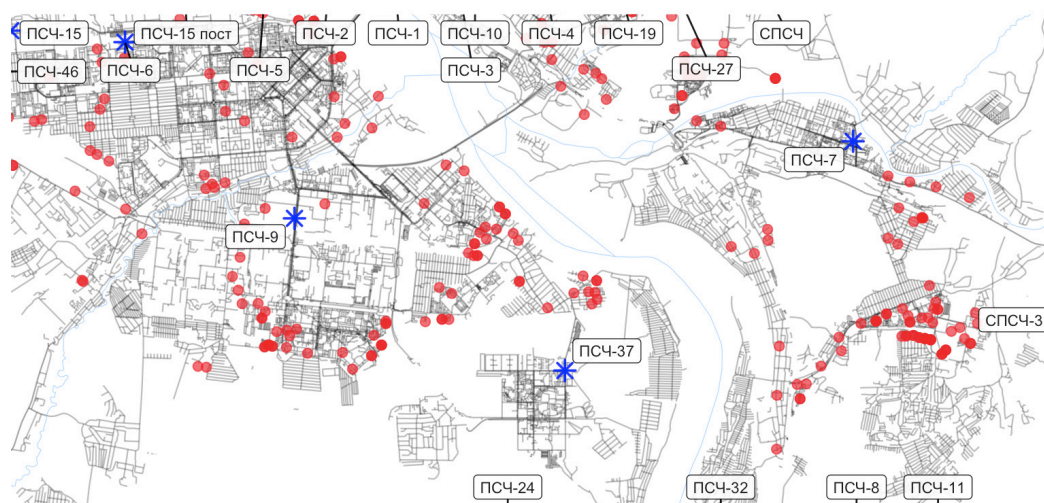


Рис. 4. Фрагмент данных, иллюстрирующих труднодоступные районы прибытия пожарно-спасательных подразделений города Новосибирска

Отметим, что обычно при анализе зон транспортной доступности зоны 10-минутной достижимости из пожарной части не учитывают интенсивность транспортного потока и особенности дорожного рельефа, поэтому карты такого рода можно использовать в качестве дополнительного инструмента при оптимизации ресурсов пожарно-спасательных подразделений.

Создание интерактивных карт

Следующий этап работы с геоданными, рассматриваемый в настоящей статье, – это создание и использование интерактивных карт в R. Одной из самых популярных open-source библиотек для создания интерактивных карт на JavaScript является Leaflet. Соответствующая библиотека Leaflet for R также позволяет легко создавать интерактивные карты и интегрировать Leaflet как HTML-виджеты не прибегая при этом к средствам JavaScript. Пример Leaflet карты, иллюстрирующей пожары в Новосибирске за период с начала 2016 года по середину июля 2020 года, показан на рисунке ниже (см. рис. 5). Для удобства отображения большого количества маркеров, отмечающих точки, где произошли пожары, смежные маркеры автоматически сгруппированы в кластеры.

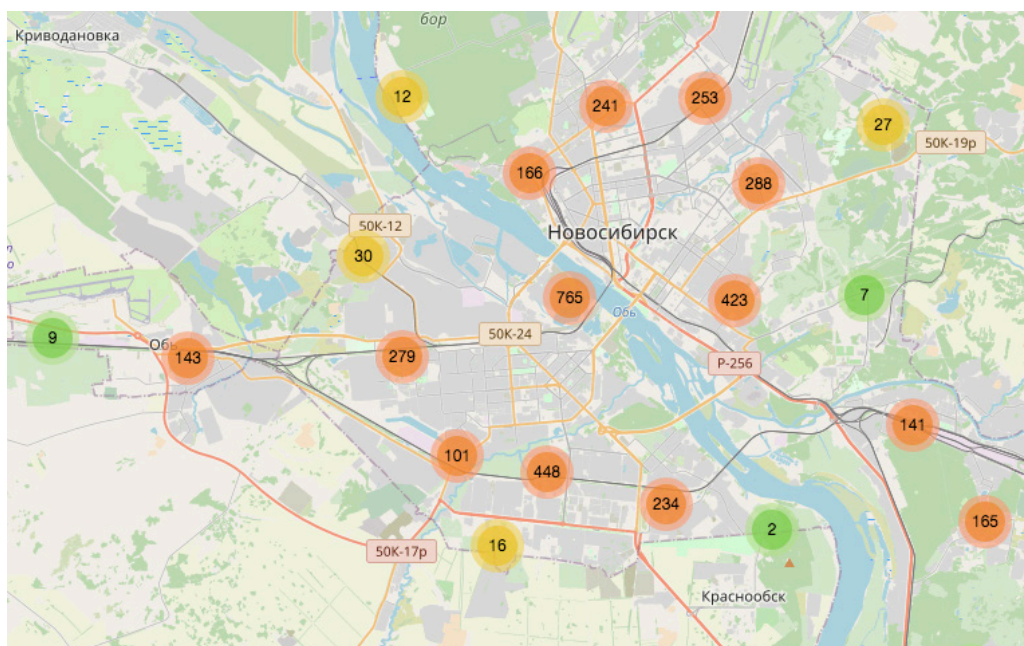


Рис. 5. Пример отображения пожаров на карте города Новосибирска с использованием библиотеки Leaflet

Средства Leaflet позволяют: изменять текст всплывающих подсказок маркеров, представлять маркеры в виде кругов для отображения площади пожара, делать интерактивные легенды, фильтровать данные средствами crosstalk, интегрировать средства Leaflet в динамические отчеты Shiny, и т.д. Важным свойством Leaflet является то, что можно выбирать различные готовые подложки для отображения карт, например, слой OpenFireMap, который наносит на карту пожарные части, гидранты и пожарные водоемы.

Также, мощным средством анализа географических данных, о котором следует упомянуть, является библиотека osrm, представляющая собой интерфейс R к онлайн-библиотеке маршрутизации OSRM, работающей на основе данных OpenStreetMap. Нас будут интересовать зоны транспортной доступности из определенной фиксированной точки на основе изохрон по времени движения (см. также [6, параграф 14.2]). Для этого нам понадобится функция osrmIsochrone из библиотеки osmr. Используя osmr в совокупности с Leaflet, можно получить интерактивную карту покрытия для любой заданной пожарной части. В качестве опции osrm.profile для функции osrmIsochrone можно указать «car», «bike» или «foot» на основе сервиса routing.openstreetmap.de, однако из минусов функции можно отметить тот факт, что при этом нельзя задать исходную скорость движения. Плюсом библиотеки является то, что ведется автоматический учет характера направления для дорог с односторонним движением. Пример интерактивной карты, определяющей районы выезда с градацией от 0 до 10 минут с изохронами, ограничивающими двухминутные районы, показан ниже (см. рис. 6).

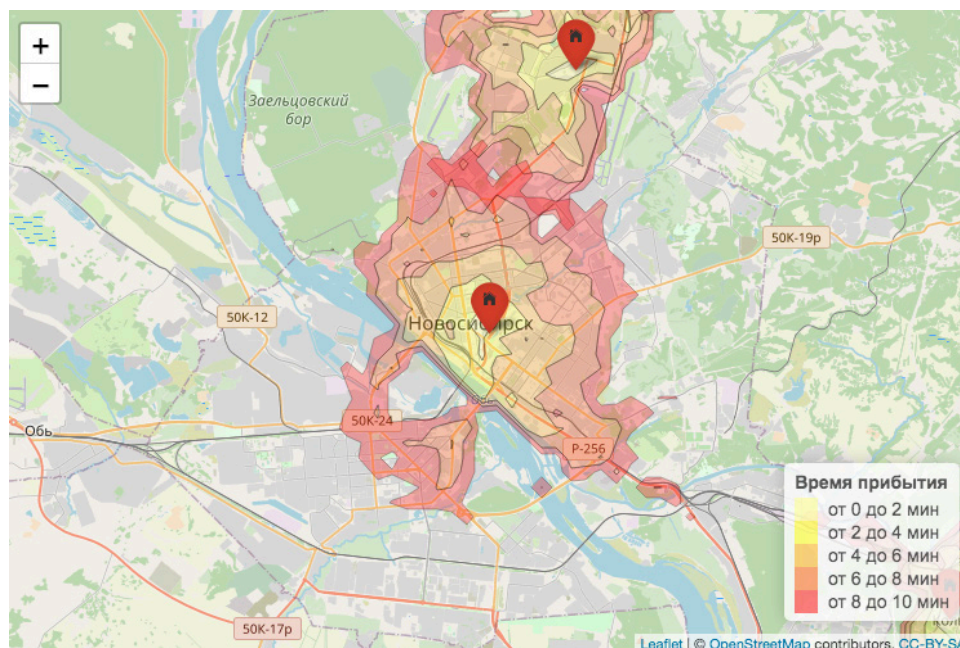


Рис. 6. Пример областей транспортной доступности, ограниченных изохронами, для города Новосибирска

Заключение

В статье были рассмотрены некоторые возможности языка программирования R в применении к анализу географических данных по пожарам и их последствиям. Были рассмотрены нанесение на статические и интерактивные карты:

- исходных данных по пожарам;
- плотности пожаров;
- данных по времени прибытия подразделений.

Также, было рассмотрено построение зон транспортной доступности и оптимальных маршрутов.

Используемые аналитические инструменты и сделанные выводы могут быть полезны при определении порядка привлечения сил и средств подразделений пожарно-спасательных гарнизонов – в частности при разработке планов привлечения и расписаний выездов сил и средств гарнизонов для тушения пожаров и проведения АСР. Приведенная в работе процедура анализа может быть использована, с учетом специфики, практически для любого населенного пункта Российской Федерации и иных чрезвычайных ситуаций, где используются географические данные.

Литература

1. Федеральный закон от 22 июля 2008 г. № 123-ФЗ «Технический регламент о требованиях пожарной безопасности» (с изменениями и дополнениями): [Электронный ресурс] // Информационно-правовой портал Гарант.ру. URL: <https://base.garant.ru/12161584/>. (Дата обращения: 26.03.2021) .
2. Мельник, А.А. Расчетное определение области нормативного обслуживания пожарных частей на территории города Красноярска / Мельник А.А., Антонов А.В., Мартинович Н.В., Малютин О.С. // Научно-аналитический журнал «Сибирский пожарно-спасательный вестник», 2019, №3. – С.61-66. – Режим доступа: http://vestnik.sibpsa.ru/wp-content/uploads/2019/v3/N14_61-66.pdf, свободный. – Загл. с экрана. – Яз. рус., англ.
3. Бабёнышев, С.В. Прогнозирование временных рядов на основе методов машинного обучения в вопросах обеспечения природной и техносферной безопасности / Бабёнышев С.В., Малютин О.С., Матеров Е.Н. // Научно-аналитический журнал «Сибирский пожарно-спасательный вестник», 2021, №1. – С.75-83. – Режим доступа: http://vestnik.sibpsa.ru/wp-content/uploads/2021/v1/N20_75-83.pdf, свободный. – Загл. с экрана. – Яз. рус., англ.

4. Матеров, Е.Н. Использование языка программирования R в вопросах пожарной безопасности: анализ статистики количества пожаров на основе теории временных рядов. Сибирский пожарно-спасательный вестник. 2019. №2 (12). С. 52-57.
5. Матеров, Е.Н. Использование языка программирования R в вопросах пожарной безопасности: анализ главных компонент. Сибирский пожарно-спасательный вестник. 2019. №1 (13). С. 49-53.
6. Самсонов Т.Е. Визуализация и анализ географических данных на языке R. М.: Географический факультет МГУ, 2021. DOI: 10.5281/zenodo.901911.