Минобрнауки России

Санкт-Петербургский государственный политехнический университет

Институт информационных технологий и управления

Кафедра информационных и управляющих систем

КУРСОВОЙ ПРОЕКТ

по дисциплине «Наука о данных и аналитика больших объемов информации»

Анализ звонков в службу спасения

Выполнил

студент гр. 13544/1 Басов С.А.

Руководитель Никифоров И.В.

Санкт-Петербург

2016

Оглавление

[Цель 2](#_Toc469715357)

[Введение 3](#_Toc469715358)

[Задачи 4](#_Toc469715359)

[Реализация 5](#_Toc469715360)

[Визуализация 6](#_Toc469715361)

[Заключение 12](#_Toc469715362)

Цель

Целью данного курсового проекта является обработка и анализ информации по выбранной теме. Выбранной темой является: «Звонки в службу спасения 911 на территории Соединенных Штатов Америки».

# Введение

Для курсового проекта по данному курсу мною была выбрана тема «Звонки в службу спасения 911». Набор данных был взят с сайта [www.kaggle.com](http://www.kaggle.com) в формате .csv. Данная информация является публичной и общедоступной. В данном файле содержится информация о количестве звонков, их частоте, местоположении звонившего (город, zip-код), дате и времени совершенного звонка, тип и подтип (причина) звонка. Размер файла составляет ~25.5мб.

Имея такое относительно большое количество данных, мы можем проанализировать их и сделать выводы о том, в какое время суток совершается наибольшее количество звонков, из какого города звонят чаще всего, а также наиболее распространенные причины звонков в службу спасения.

Проведенный анализ может помочь сотрудникам подобного рода служб быть более подготовленным к вызовам, так как позволяет наглядно отследить зависимость количества вызовов от местоположения и времени суток. Также можно понять какая причина звонка наиболее распространена в том или ином городе.

# Задачи

Построить графики:

* Количество звонков в зависимости от даты на 2016г
* Количество звонков определенного типа в зависимости от даты на 2016г
* Количество звонков в зависимости от месяца
* Количество звонков в зависимости от времени суток
* От дня недели
* Количество звонков определенного подтипа
* Топ 10 zip-кодов (индексов) по количеству звонков
* Топ 10 городов по количеству звонков
* Топ 10 городов по количеству звонко с отображением типа звонка
* С отображением подтипа звонка

# Реализация

Для реализации поставленных задач была выбрана среда разработки RStudio и язык программирования R.

Первое, что необходимо сделать – это подключить набор библиотек, необходимых визуализации данных.

Затем, считать данные из файла 911.csv.

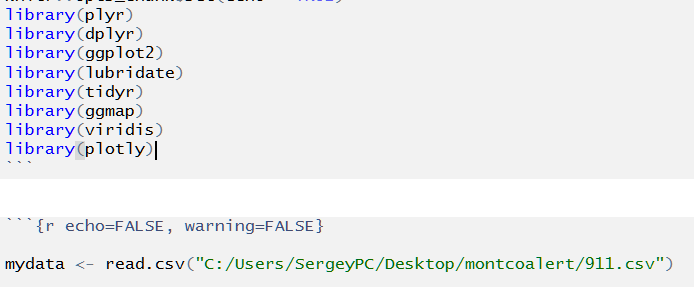


Рис.1 – Считывание данных.

## Визуализация

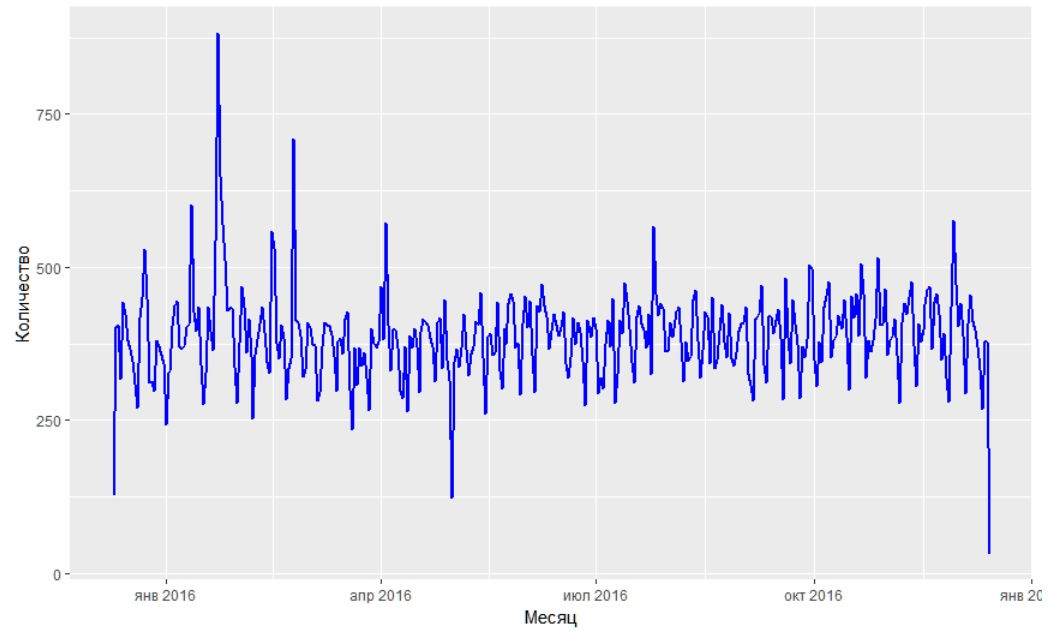


Рис.2 – По датам.

На данном графике можно видеть отображение количества звонков в службу спасения по дате за 2016год. Среднее значение находится на уровне 350-450 звонков. Наибольшее количество звонков было в феврале и составляло около 900 звонков.

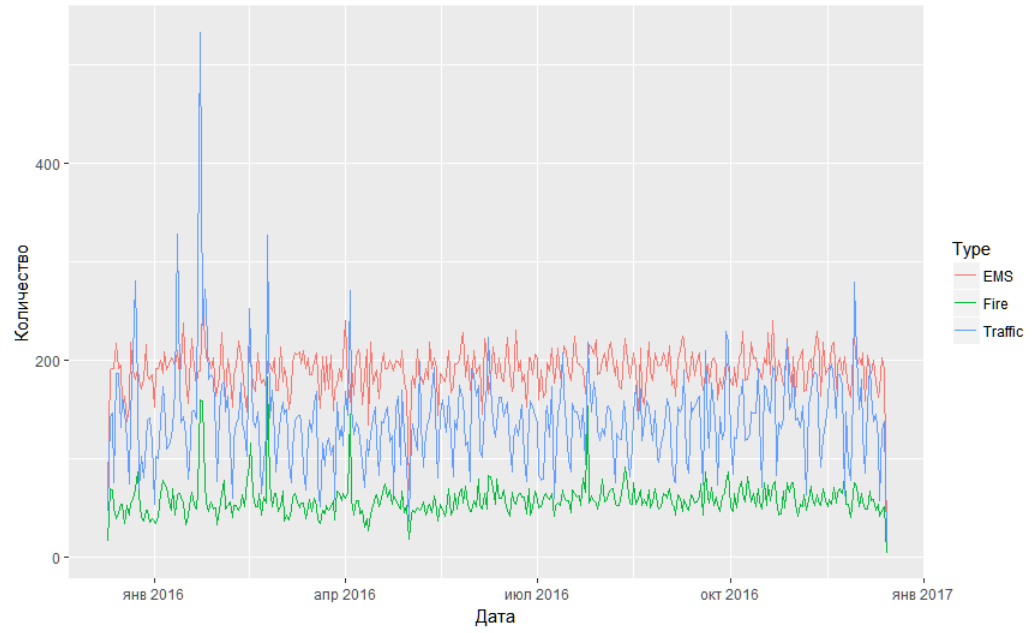


Рис. 3 – По датам и типам.

На данном графике можно видеть количество звонков определенного типа по датам за 2016год. Практически все время лидирует тип обращения EMS, однако, в феврале виден скачок количества звонков с типом Traffic.

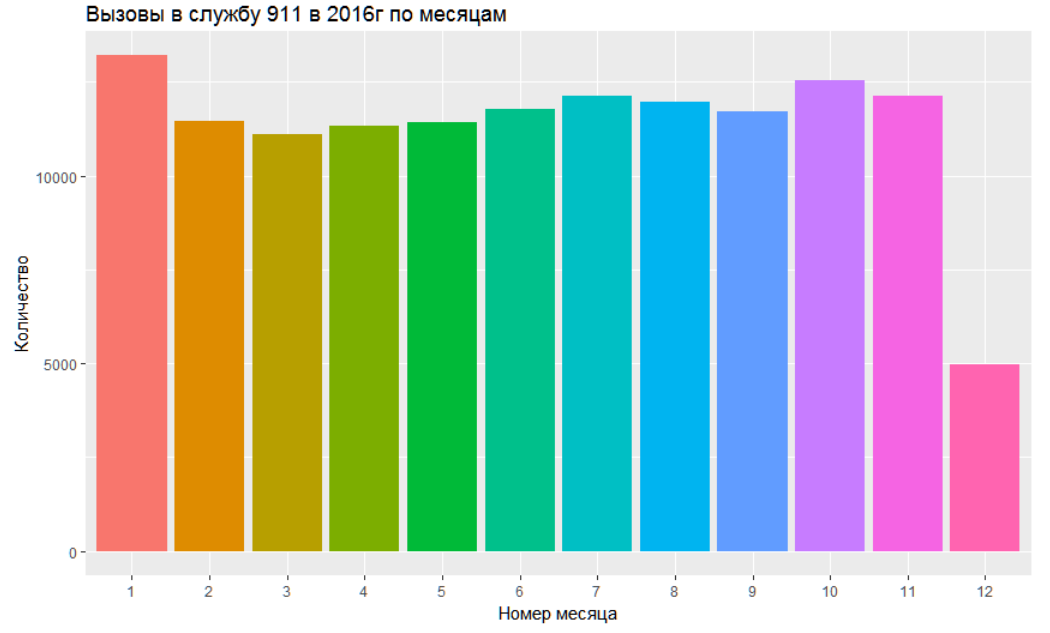


Рис. 4 – По месяцам.

Данная диаграмма отображает количество звонков по месяцам за весь промежуток времени. Можно отметить, что наибольшее количество звонков было совершено в январе, а в декабре наблюдается спад.

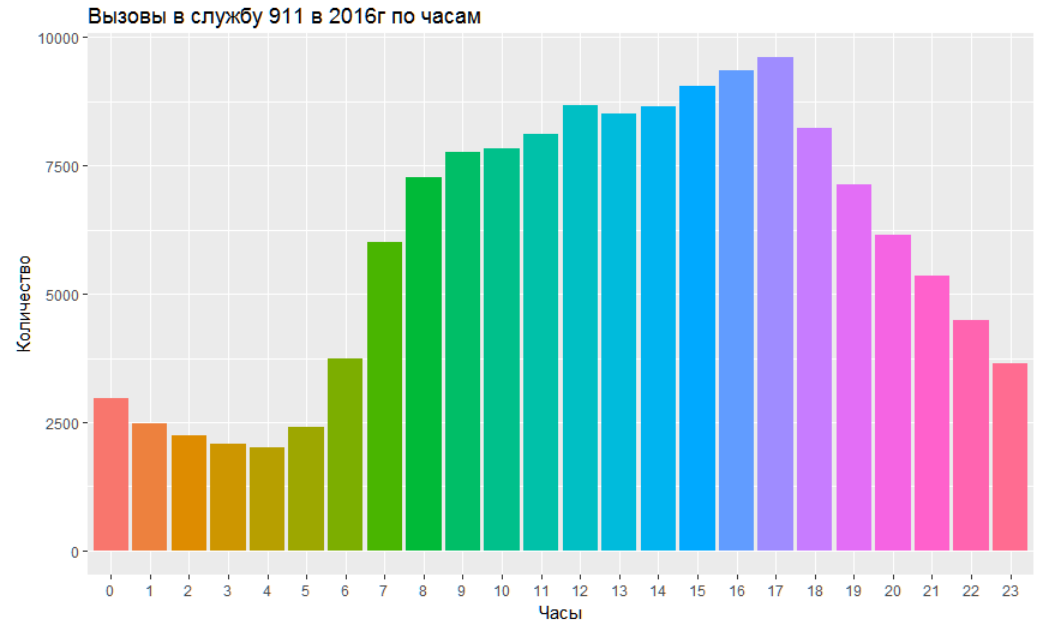


Рис. 5 – по часам.

По данной диаграмме можно определить интенсивность звонков в течении дня. Можно сделать вывод о том, что количество звонков начинает расти после 6 утра, а своего пика достигает к 16-17 часам. Затем наблюдается интенсивный спад к полуночи.

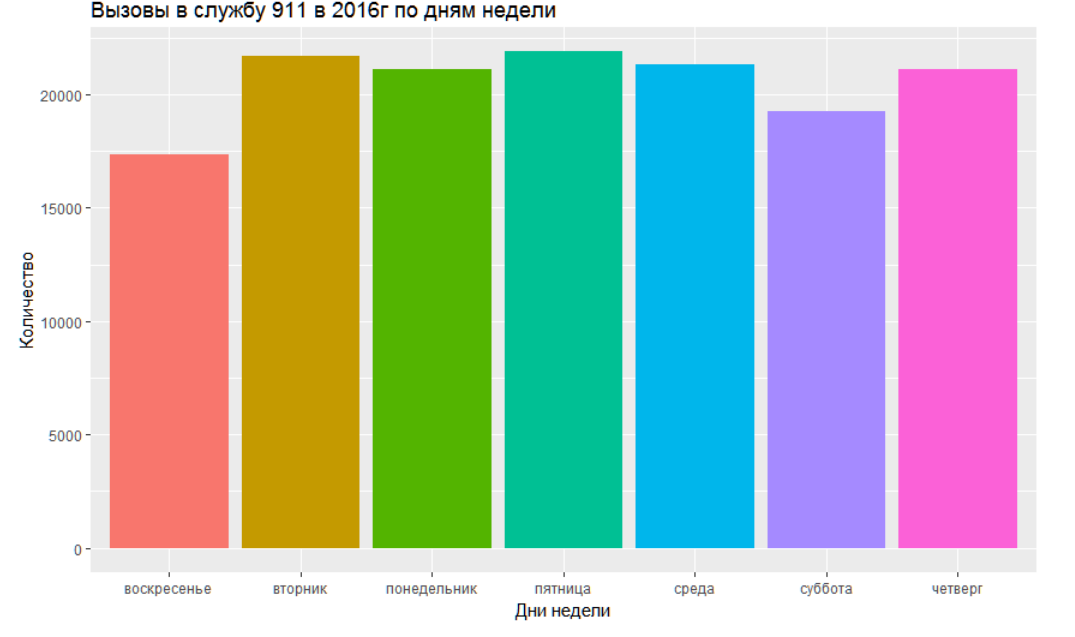


Рис. 6 – по дням недели.

Данная диаграмма визуализирует информацию о количестве звонков в службу спасения в течении недели. Шаг – один день. Наименьшее количество звонков поступало в воскресенье.

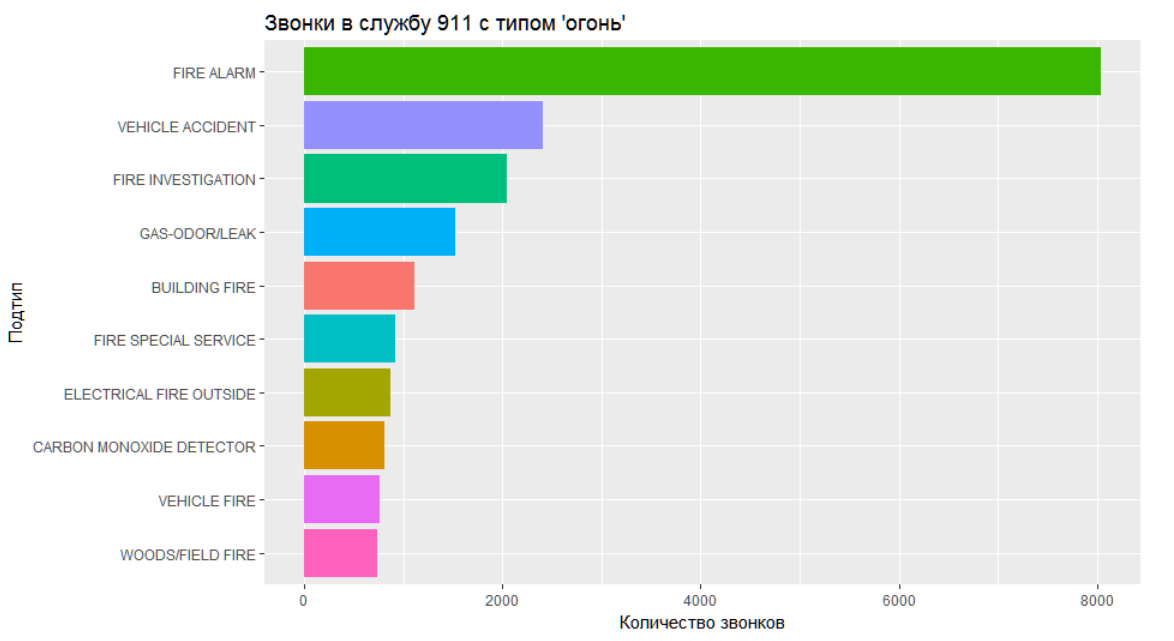


Рис. 7 – По типу.

На диаграмме показано количество звонков за весь промежуток времени типа «возгорание» с подтипами. Преобладает по количеству звонков подтип «Пожарная сигнализация».

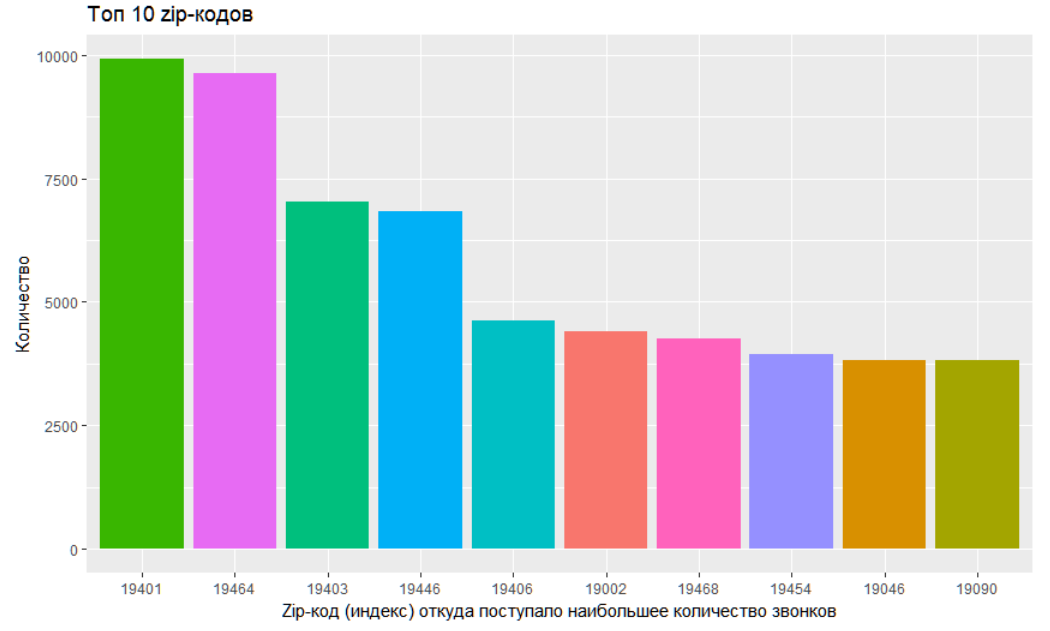


Рис. 8 – по zip-коду.

Данная диаграмма показывает количество звонков относительно местоположения звонившего определяемое zip-кодом (индексом). Наибольшее количество звонков было совершено из районов с кодом 19401 и 19464.

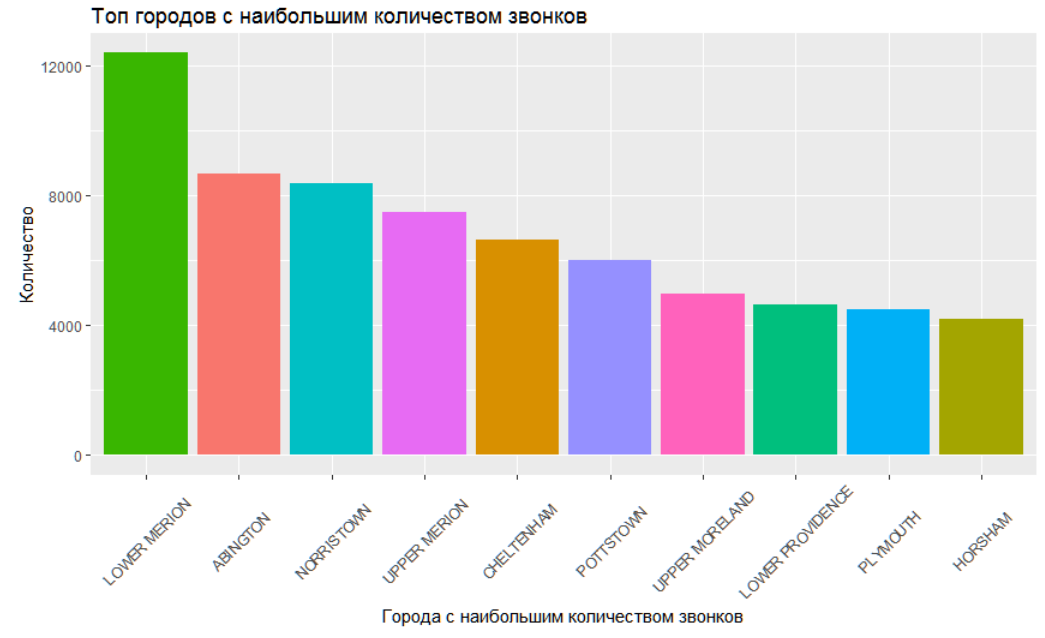


Рис. 9 – по населенным пунктам.

На диаграмме отображено количество звонков из определенного населенного пункта США. Видно, что лидирует по данному критерию Lower Merion.

Таблица 1 – численные показатели по населенным пунктам.

|  | **Twp**  <fctr> | **Total**  <int> | **Perc**  <dbl> |  |
| --- | --- | --- | --- | --- |
| 28 | LOWER MERION | 12400 | 8.625187 |  |
| 2 | ABINGTON | 8656 | 6.020937 |  |
| 37 | NORRISTOWN | 8364 | 5.817828 |  |
| 59 | UPPER MERION | 7485 | 5.206413 |  |
| 8 | CHELTENHAM | 6655 | 4.629082 |  |
| 43 | POTTSTOWN | 6024 | 4.190171 |  |
| 60 | UPPER MORELAND | 4965 | 3.453553 |  |
| 31 | LOWER PROVIDENCE | 4649 | 3.233750 |  |
| 42 | PLYMOUTH | 4505 | 3.133586 |  |
| 21 | HORSHAM | 4204 | 2.924217 |  |

Так же, для большего удобства, данные могут быть представлены в виде таблицы.

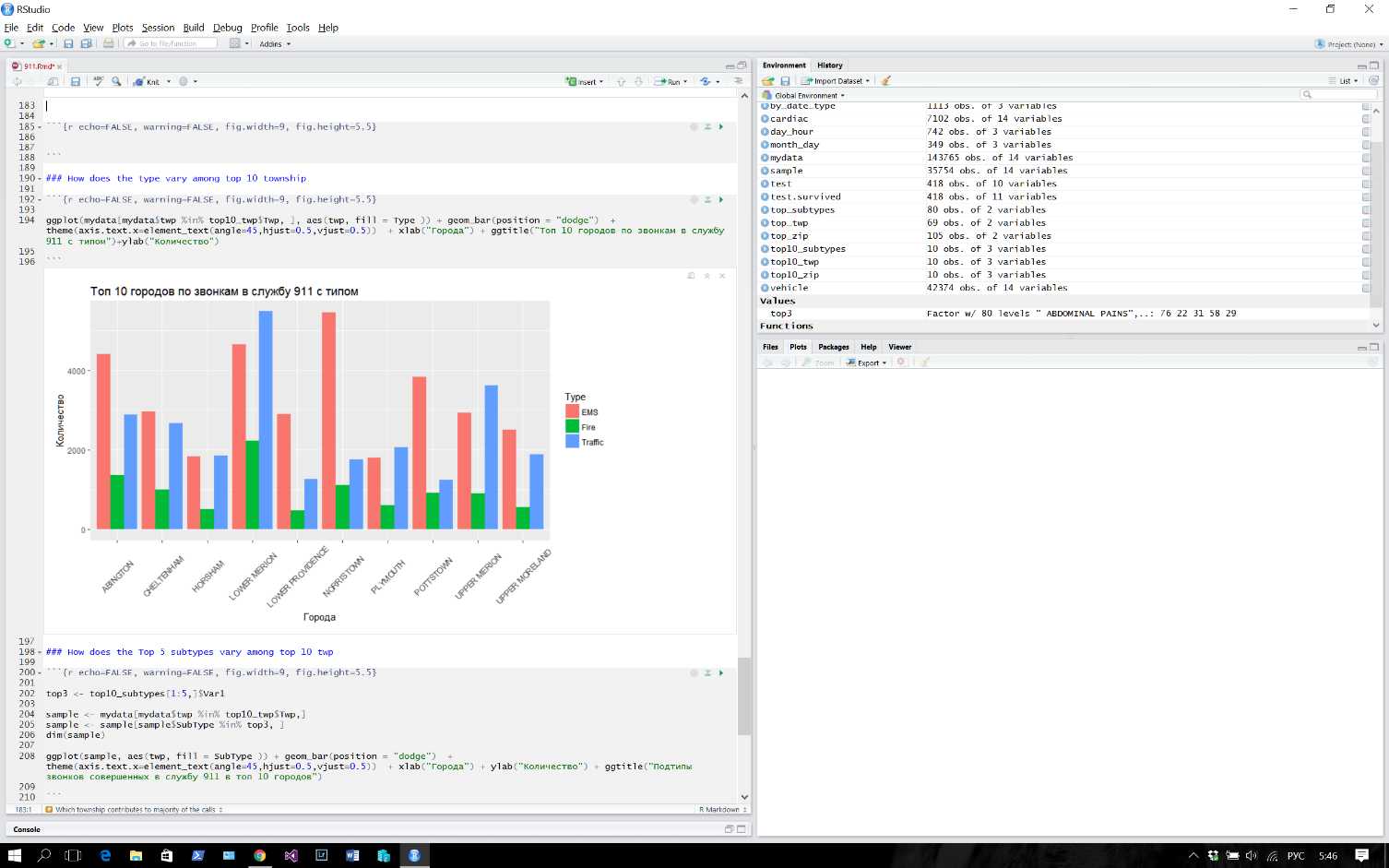


Рис. 10 – по населенным пунктам и типу.

На диаграмме показано количество звонков определенного типа в каждом населенном пункте. Практически во всех населенных пунктах лидируют звонки с типом обращения EMS. Наименьшее количество звонков во всех случаях с типом «возгорание».

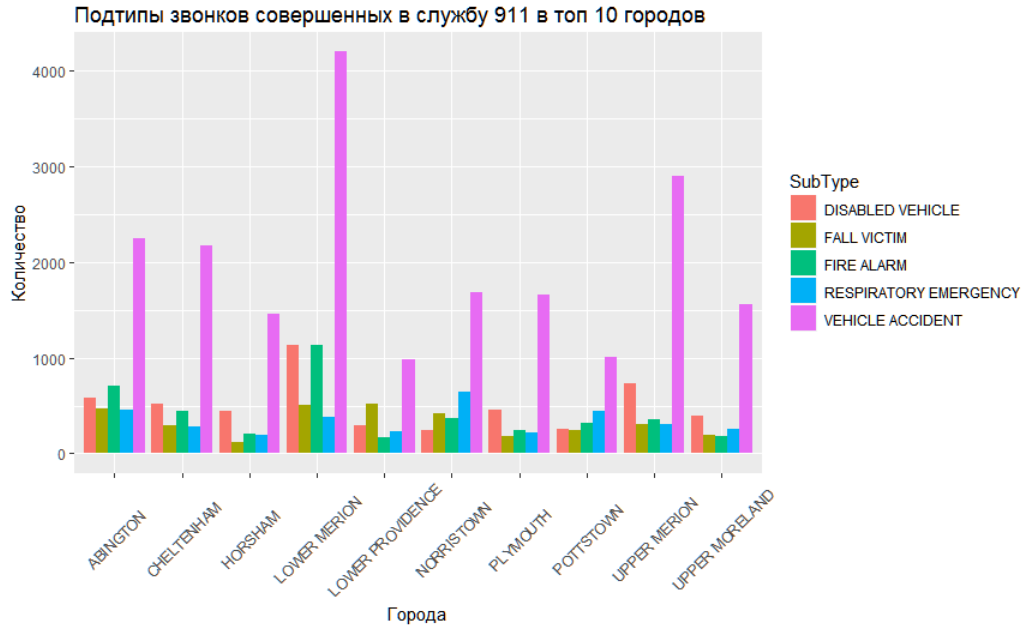


Рис. 11 – подтипы по населенным пунктам.

На последней диаграмме продемонстрировано количество обращений в службу 911 по населенным пунктам с указанием подтипа обращения. Наиболее часто встречаемый подтип во всех случаях – Vehicle Accident (дорожно-транспортные происшествия).

# Заключение

В ходе написания курсового проекта был освоен язык программирования R, а также получены навыки работы в среде программирования RStudio и навыки обработки больших объемов информации с помощью данных инструментов.

Данные навыки являются очень важными в настоящее время, поскольку количество информации постоянно растет и требует определенных умений и знаний при ее обработке.

Касаемо выбранной темы, анализ подобного рода информации может положительно сказаться на работе спецслужб, которые заранее будут готовы к тому что в определенное время суток обычно поступает большее количество звонков, и, возможно, в данное время следует привлечь большее количество сотрудников. Так же можно уследить зависимость причин обращения в службу 911 от населенного пункта, что поможет быстрее реагировать на вызовы.

Ссылка на репозиторий: https://github.com/Warhazar/BigData.git