

ВИЗУАЛИЗАЦИЯ И АНАЛИЗ ГЕОДАННЫХ

Е.В. Мартынюк¹

Научный руководитель: ст. преподаватель, к.т.н. А.А. Орлов²

¹Национальный исследовательский Томский политехнический университет,
Россия, г.Томск, пр. Ленина, 30, 634050

²Томский государственный университет систем управления и радиоэлектроники,
Россия, г.Томск, пр. Ленина, 40, 634034

E-mail: catherine.martyniuk@mail.ru

VISUALIZATION AND ANALYSIS OF GEODATA

E.V. Martyniuk¹

Scientific Supervisor: Senior Lecturer, PhD A.A. Orlov²

¹Tomsk Polytechnic University, Russia, Tomsk, Lenin str., 30, 634050

²Tomsk State University of Control Systems and Radioelectronics,
Russia, Tomsk, Lenin str., 40, 634034

E-mail: catherine.martyniuk@mail.ru

Abstract. *Visualization and analysis of crime data using R language tools and Google Maps by mapping the intensity of any crime. Based on maps built by the intensity, the analysis of the crime situation in the city.*

Введение. В работе рассматривается решение задачи визуализации геоинформационных данных о преступлениях, совершенных в городе Сан-Франциско (США) с использованием языка R [1]. Для каждого зарегистрированного преступления приводится информация о точных координатах и категории преступления (для дальнейшего анализа возможно дата, время и пр.). С использованием дополнительных библиотек языка R: ggmap [2], ggplot2 [3], lubridate [4], dplyr [5], устанавливается интенсивность преступления в зависимости от района его совершения и визуализируется с использованием карт GoogleMaps. Подобный способ визуализации данных позволяет эксперту произвести оценку криминальной обстановки в определенном районе города с целью выработки мероприятий по снижению криминальной активности.

Цель работы. Визуализация и анализ геоинформационных о совершенных преступлениях.

Иллюстративный пример. В качестве исходной выборки использовались данные [6] о преступлениях за период 01.01.2003 по 18.08.2015 в городе Сан-Франциско (США), содержащие 20833 записей. Каждая запись содержит атрибутивные и пространственные поля: категория, описание, дата, время, район, улица, вид наказания, географические координаты. Из полученных данных можно получить информацию о криминальной обстановке в городе.

Путем построения гистограмм зависимости (рис. 1) количества преступлений от месяца (от 1 до 12), числа месяца (от 1 до 31), дня недели (с понедельника до воскресенья) производится статистический анализ выборочных данных, показывающий, что строгой зависимости совершения преступлений от

указанных выше характеристик не наблюдается. Это значит, что совершение преступлений в целом не зависят, например, от дня выдачи заработной платы, или выходного или рабочего дня недели.

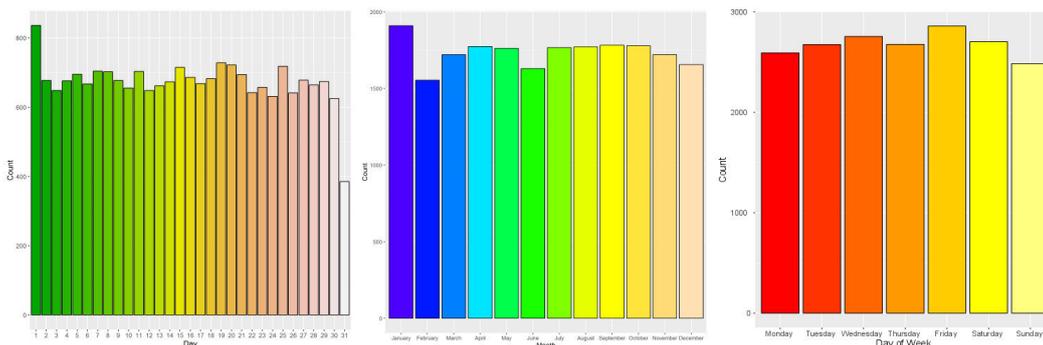


Рис. 1. Гистограммы зависимости преступлений от (слева-направо) дня месяца, месяца, дня недели

Присутствует зависимость от времени суток (рис. 3): интенсивность уменьшается после полуночи и продолжается до раннего утра (6.00), далее снова наблюдается рост числа преступлений и достигает своего пика к 18.00.

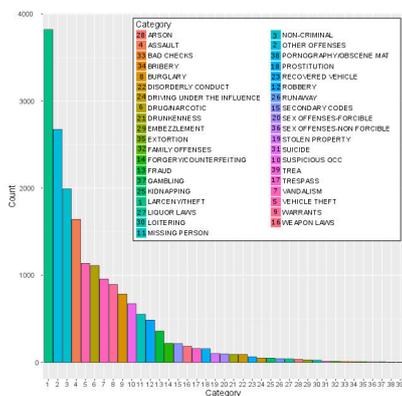


Рис. 2. Гистограмма всех категорий

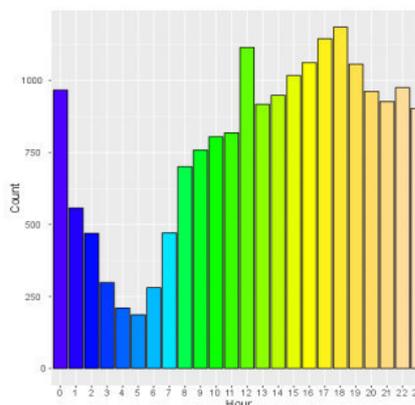


Рис. 3. Гистограмма зависимости преступлений от времени суток

Визуализация данных с использованием карт GoogleMaps в зависимости от категории преступления и места его совершения представлена на рис. 1.

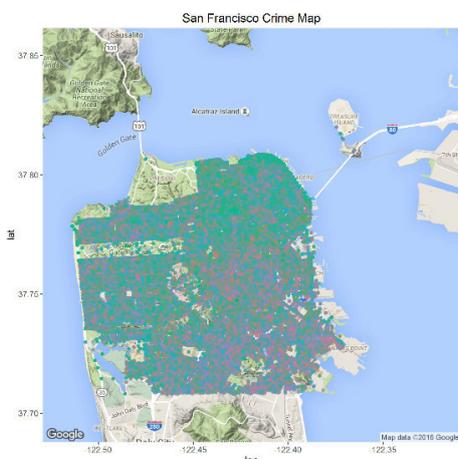


Рис. 4. Карта преступлений

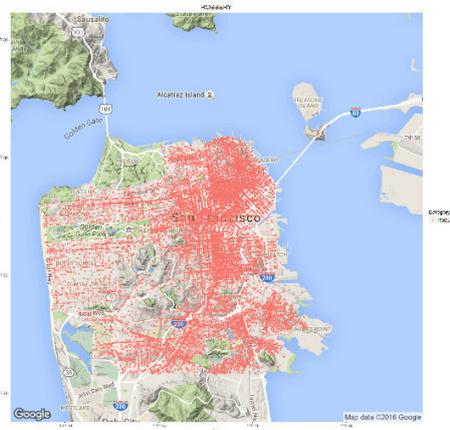


Рис. 5. Карта для категории ROBBERY

Используя карту всех преступлений (рис. 4), получить какую-либо полезную информацию не представляется возможным ввиду большого количества категорий и случаев преступлений. Локализация преступлений определенной категории (рис. 5, для категории ROBBERY (*Ограбление (англ.)*)) дает представление о «доминирующей» части города, но недостаточно объективное и точное. Основываясь на статистике преступлений (рис. 2), для анализа используем самые часто встречающиеся (от 500 случаев) категории преступлений: LARCENY/THEFT, OTHEROFFENSES, NON-CRIMINAL, ASSAULT, VEHICLE THEFT, DRUG/NARCOTIC, VANDALISM, WARRANTS, BURGLARY, SUSPICIOUS OCC, MISSING PERSON, ROBBERY. Большое число случаев говорит о том, что с этими преступлениями стоит бороться в первую очередь, поэтому такие категории представляют наибольший интерес.

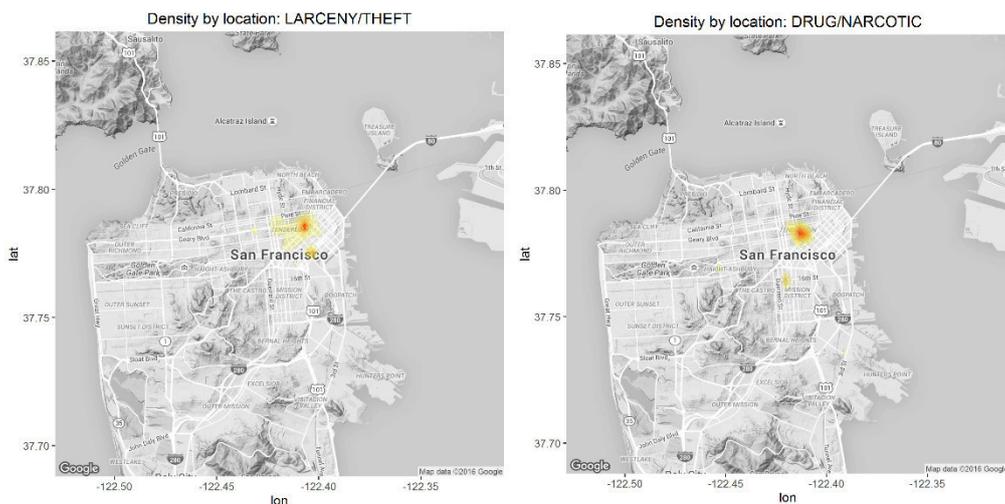


Рис. 6. Интенсивность и локализация категорий DRUG/NARCOTIC, LARCENY/THEFT

Визуализация данных путем построения карты интенсивности преступления (рис. 6, пример для категорий DRUG/NARCOTIC, LARCENY/THEFT) показала, что пик у 10 из 12 заявленных категорий преступлений приходится на район Тендерлоин.

Выводы. По результатам проведенного анализа составлено представление о криминальной обстановке в городе Сан-Франциско. Наблюдается зависимость преступления от места и времени его совершения. Наиболее криминальный район города – Тендерлоин. Однако такие характеристики как день недели, число месяца или месяц не оказывают существенного влияния на интенсивность преступлений. Основным фактором интенсивности является время суток.

СПИСОК ЛИТЕРАТУРЫ

1. The R Project for Statistical Computing [Electronic version]. – Режим доступа: www.r-project.org. – 25.03.16.
2. Kahle D., Wickham H. (2016, January 23). Package ‘ggmap’. Version 2.6.1. CRAN, 60 p.
3. Wickham H., Chang W. (2016, March 1). Package ‘ggplot2’. Version 2.1.0. CRAN, 192 p.
4. Golemund G., Spinu V., Wickham H. (2015, December 3). Package ‘lubridate’. Version 1.5.0. CRAN, 73 p.
5. Wickham H., Francois R. (2015, September 1). Package ‘ggmap’. Version 4.0.3. CRAN, 70 p.
6. SF OpenData [Electronic version]. – Режим доступа: data.sfgov.org. – 25.03.16.