ISSN 2278-3091
Volume 6, No.5, September -October 2017
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse02652017.pdf

Data Analysis for Discovering Relationships between CCTV and Crime

Yeonjoon Han[1], Minsoo Lee[2]
[1]Dept. of Computer Science and Engineering, Ewha Womans University, Seoul, Korea, Email: coco7j@gmail.com
[2]Dept. of Computer Science and Engineering, Ewha Womans University, Seoul, Korea, Email: mlee@ewha.ac.kr

ABSTRACT

In public spaces, many CCTV cameras have been installed with the aim of crime prevention and criminal arrest. Many studies show that CCTV could reduce the number of crimes. However, Even though there are a lot of CCTVs, still there are a lot of crimes around us. Through this R project, confirm if there is a relationship between CCTV and crimes. In this paper, Using R program, clustering the areas where CCTV are installed and comparing that result with the areas where the crime cases have been occurred. The Result shows that comparing with the areas where there is no CCTV, the areas where the CCTV cameras are installed have smaller crime occurrence than the others.

Key words : CCTV, Crime, Data Analysis, Data Analysis, R Program.

1. INTRODUCTION

These days, Millions of closed circuit television (CCTV) cameras are installed in streets and businesses for increasing public safety and reducing crime. In public spaces, many CCTV cameras have been installed with the aim of crime prevention and criminal arrest. Because Crime is a big problem everywhere. Such CCTV data is an important example of big data and can be used for meaningful analysis [1]. Many studies about relationship between CCTV and crime are published. The analysis [2] found that surveillance systems were most effective in parking lots, where their use resulted in a 51% decrease in crime. Systems in other public settings had some effect on crime – a 7% decrease in city centers and in significant. In another results of study [3], the number of robberies and thefts in the areas with CCTV installed reduced by 47.4%, while the areas without CCTV showed practically no change in the number of crimes. However, Even though there are a lot of CCTVs, still there are a lot of crimes around us. We wonder why crimes are still occurred and if it is real that the CCTV has effects on crime reduction. Through this R project, we want to confirm if there is a relationship between CCTV and crimes. Using R program, clustering the areas where CCTV are installed and comparing that result with the areas where the crime cases have been occurred.

> cctv

	location	longitude	latitude
1	Park Heights Ave - Fallstaff Rd	-76.70282	39.36495
2	Park Heights Ave - Clarks Ln.	-76.69966	39.36236
3	Park Heights Ave - N.Bancroft Rd	-76.69888	39.36174
4	Park Heights Ave - Bancroft Rd	-76.69785	39.36090
5	Park Heights Ave - Fords Ln.	-76.69657	39.36016
6	Park Heights Ave - W Strathmore Ave	-76.69433	39.35900
7	Park Heights Ave - W Strathmore Ave	-76.69390	39.35878
8	Edmonson Ave - Swan Ave	-76.69383	39.29349
9	Park Heights Ave - Pinkney Rd	-76.69278	39.35818
10	Park Heights Ave - Taney	-76.69122	39.35740
11	Park Heights Ave - Menlo Dr	-76.68969	39.35644
12	Park Heights Ave - Glen Ave	-76.68817	39.35566
13	Park Heights Ave - Trainor Ave	-76.68757	39.35500
14	Park Heights Ave - Trainor Ave(JCC)	-76.68628	39.35390
15	Park Heights Ave - Manhattan Ave	-76.68524	39.35301
16	Park Heights-Rogers	-76.68332	39.35185
17	3500 Northern Pkwy(E&T)	-76.68254	39.35383
18	Edmonson Ave - N Loudon Ave	-76.68246	39.29388
19	W Belvedere Ave - Litchfield Ave	-76.68028	39.34630

Figure 1: CCTV Data set

> crime

	age	sex	race	arrestLocation	neighborhood	longitude	latitude
1	22	M	U	100 W Jeffrey St	Brooklyn	-76.6262	39.24740
2	29	F	B	1000 N Wolfe St	Middle East	-76.5748	39.29390
3	32	M	B	1000 N Wolfe St	Middle East	-76.5748	39.29370
4	33	M	B	1000 W Franklin St	Harlem Park	-76.5694	39.29600
5	29	M	B	1200 E 25 Th St	Coldstream Homestead Montebello	-76.5942	39.32280
6	60	M	B	1200 Shellbanks Rd	Cherry Hill	-76.6181	39.29170
7	34	F	B	1200 W Baltimore St	Poppleton	-76.6089	39.28970
8	48	M	B	1500 Gorsuch Ave	Coldstream Homestead Montebello	-76.6356	39.29290
9	23	M	B	1500 Harlem Ave	Harlem Park	-76.5477	39.36820
10	25	M	B	1500 Harlem Ave	Harlem Park	-76.5489	39.36680
11	38	F	B	1500 Madison Ave	Madison Park	-76.6031	39.30930
12	33	F	B	1500 Madison St	Gay Street	-76.5964	39.30960
13	26	F	W	1600 Joplin St	Broening Manor	-76.5770	39.31590
14	52	M	B	300 Ilchester Ave	Harwood	-76.6296	39.27280
15	62	M	B	4000 Eirman Ave	Arcadia	-76.6353	39.28140
16	49	M	B	3400 Belair Rd	Belair-Edison	-76.5877	39.32470
17	59	M	B	3700 Ravenwood Ave	Belair-Edison	-76.5655	39.28860
18	30	M	W	3700 Saint Paul St	Guilford	-76.6325	39.29500
19	25	M	B	400 Baltimore St	Downtown	-76.5416	39.27630
20	24	M	B	400 N Rose St	McElderry Park	-76.5444	39.31560

Figure 2: Crime Data set

2. DATA SET EXPLANATION

In this paper, we use CCTV and Crime data set in a certain city, Baltimore, Maryland, USA. There are three data sets. First is the CCTV data (Fig. 1) which includes locations where CCTV cameras are installed and CCTV's longitude and latitude information. Second is the Crime data (Fig. 2) which includes offender's information which are age, gender and race. It also includes the location where the crime occurred, neighborhood means that the district of the city. And longitude and latitude data. Third, the Crime_Most data (Fig. 3) includes gender that divide crime data with male and female for each district, location which means that the district of the city where the crime occurred. Longitude and latitude data is the area where the crime has occurred the most in the district. And the count data means that how many crime has occurred by gender for each district.

> most

	gender	location	longitude	latitude	count
1	M	Abell	-76.6821	39.3447	11
2	F	Abell	-76.6821	39.3447	2
3	M	Allendale	-76.6394	39.2845	70
4	F	Allendale	-76.6742	39.3101	14
5	M	Arcadia	-76.5864	39.2946	16
6	M	Arlington	-76.6362	39.3352	314
7	F	Arlington	-76.6219	39.2495	48
8	M	Armistead Gardens	-76.5580	39.2872	36
9	F	Armistead Gardens	-76.6821	39.3478	12
10	M	Ashburton	-76.6825	39.3276	17
11	F	Ashburton	-76.5978	39.3199	3
12	M	Baltimore Highlands	-76.5927	39.3086	261
13	F	Baltimore Highlands	-76.5938	39.3045	87
14	M	Barclay	-76.5964	39.3096	99
15	F	Barclay	-76.5775	39.2922	19
16	M	Barre Circle	-76.6437	39.3081	5
17	F	Barre Circle	-76.6574	39.2889	2
18	M	Bayview	-76.6453	39.2883	44
19	F	Bayview	-76.6659	39.3306	11
20	M	Beechfield	-76.6829	39.3368	20

Figure 3: Crime_Most Data set

3. DATA ANALYSIS METHODOLOGY

The Density Based Clustering of Applications with Noise and Related Algorithms (DBSCAN) is used for data analysis. DBSCAN is in fpc package. Use dbscan like dbscan(x, eps, minPts = 5, weights = NULL, search = "dist", ...). Let's explain DBSCAN's arguments. x is a data matrix or a dist object. Eps is the parameter eps defines the radius of neighborhood around a point x. MinPts is the number of minimum points in the eps region (for core points). Weights is weights for the data points and only needed to perform weighted clustering. Search is nearest neighbor search strategy (linear, dist, etc…). DBSCAN estimates the density around each data point by counting the number of points in a user-specified eps-neighborhood and applies a used-specified MinPts thresholds to identify core, border and noise points. In second step, core points are joined into a cluster if they are density reachable. Finally, border points are assigned to clusters. Algorithm only needs parameters eps and MinPts. [4] Unlike K-means, DBSCAN does not require the user to specify the number of clusters to be generated. Also DBSCAN can find any shape of clusters. The density of points in a cluster is considerably higher than the density of points outside the cluster.

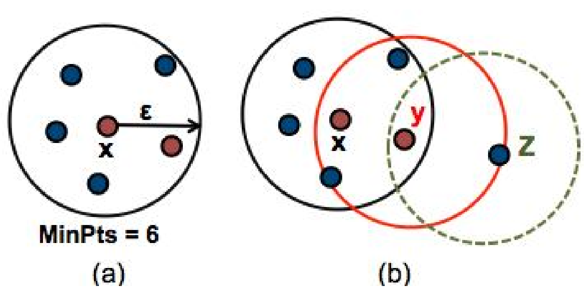

Figure 4: The Algorithm of DBSCAN[3]

In DBSCAN Algorithm (Fig. 4), The goal is to identify dense regions, which can be measured by the number of objects close to a given point. Two important parameters are required for DBSCAN: epsilon ("eps") and minimum points ("MinPts"). The parameter eps defines the radius of neighborhood around a point x. It's called called the ϵ ϵ -neighborhood of x. The parameter MinPts is the minimum number of neighbors within "eps" radius. Any point x in the dataset, with a neighbor count greater than or equal to MinPts, is marked as a core point. We say that x is border point, if the number of its neighbors is less than MinPts, but it belongs to the ϵ ϵ -neighborhood of some core point z. Finally, if a point is neither a core nor a border point, then it is called a noise point or an outlier. The figure below shows the different types of points (core, border and outlier points) using MinPts = 6. Here x is a core point because neighbours ϵ (x)=6 neighboursϵ(x)=6 , y is a border point because neighbours ϵ (y)<MinPts neighboursϵ(y)<MinPts , but it belongs to the ϵ ϵ -neighborhood of the core point x. Finally, z is a noise point. A density-based cluster is defined as a group of density connected points

1. For each point x i xi, compute the distance between x i xi and the other points. Finds all neighbor points within distance eps of the starting point (x i xi). Each point, with a neighbor count greater than or equal to MinPts, is marked as core point or visited. 2. For each core point, if it's not already assigned to a cluster, create a new cluster. Find recursively all its density connected points and assign them to the same cluster as the core point. 3. Iterate through the remaining unvisited points in the dataset. Those points that do not belong to any cluster are treated as outliers or noise. [5]

4. R CODES FOR DATA ANALYSIS

The R codes for the data analysis are given below and the result of ggmap is shown in Fig. 5.

```
setwd("c:\\rtest")
library(fpc)
crime <- read.csv("crime2.csv", header = T)
head(crime)
library(ggmap)
g_m <- get_map("baltimore", zoom=13,
maptype="roadmap")
gang.map <- ggmap(g_m) + geom_point(data=crime,
aes(x=longitude, y=latitude), size=1, alpha=0.7,
color="#993300", shape=15)
gang.map
cctv <- read.csv("crime_camera.csv", header = T)
head(cctv)
gang.map <- ggmap(g_m) + geom_point(data=cctv,
aes(x=longitude, y=latitude), size=2, alpha=0.7,
color="#003366", shape=17)
gang.map
most <- read.csv("crime_most.csv", header = T)
head(most)
```

```
gang.map <- ggmap(g_m) + geom_point(data=most,
aes(x=longitude, y=latitude), size=2, alpha=0.7,
color="#FF0066", shape=15)
gang.map
db <- dbscan(cbind(cctv$latitude, cctv$longitude), eps
=0.004, MinPts = 5)
plot(crime$longitude, crime$latitude, pch = 17, col =
"#336699")
par(new=TRUE)
plot(cctv$longitude, cctv$latitude, col = db$cluster, pch
= 20)
plot(most$longitude, most$latitude, pch = 17, col =
"#336699")
par(new=TRUE)
plot(cctv$longitude, cctv$latitude, col = db$cluster, pch
= 20)
gang.map <- ggmap(g_m) + geom_point(data=cctv,
aes(x=longitude, y=latitude), size=2, alpha=0.7,
color="#006666", shape=17)+geom_point(data=most,
aes(x=longitude, y=latitude), size=2, alpha=0.7,
color="#663366", shape=15)
gang.map
```

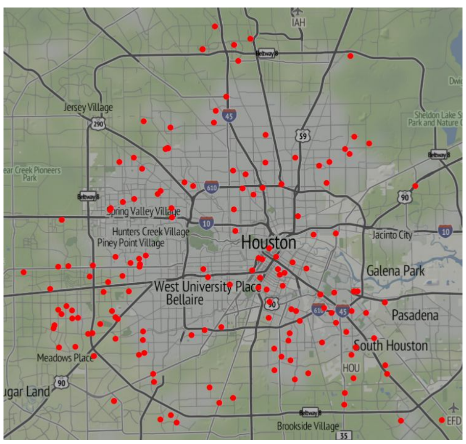

Figure 5: The result of ggmap

Besides using DBSCAN, another package used is ggmap[6]. ggmap is a new tool which enables such visualization by combining the spatial information of static maps from Google Maps, OpenStreetMap, Stamen Maps or CloudMade Maps with the layered grammar of graphics implementation of ggplot2. Use ggmap like ggmap(ggmap, extent = "panel", base_layer, maprange = FALSE, legend = "right", padding = 0.02, darken = c(0, "black"), ...). Let's explain arguments in ggmap. ggmap is an object of class ggmap. Extent is how much of the plot should take up the map.

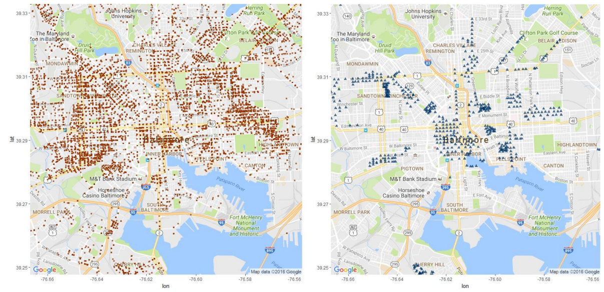

Figure 6: The crime and CCTV data using ggmap(a. crime data, b. CCTV data)

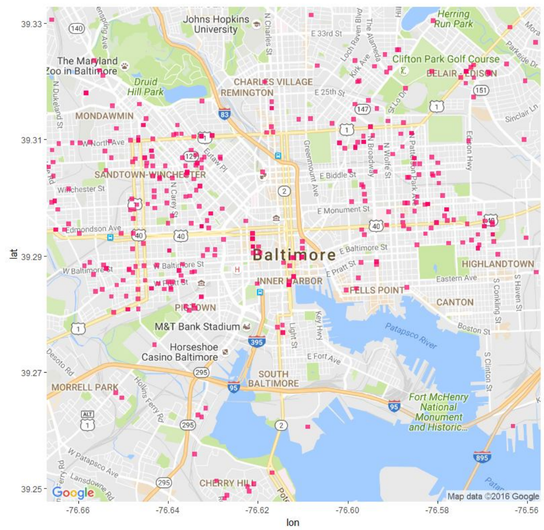

Figure 7: The most frequent crime data using ggmap

5. DATA ANALYSIS RESULTS

When plot the location of crime using library ggmap, there are a lot of crimes in Baltimore. (Fig. 6-a) And also show that CCTV data. Many CCTVs are tend to installed in center of the Baltimore city. Fig. 7 shows that the most frequent crime by gender for each district.

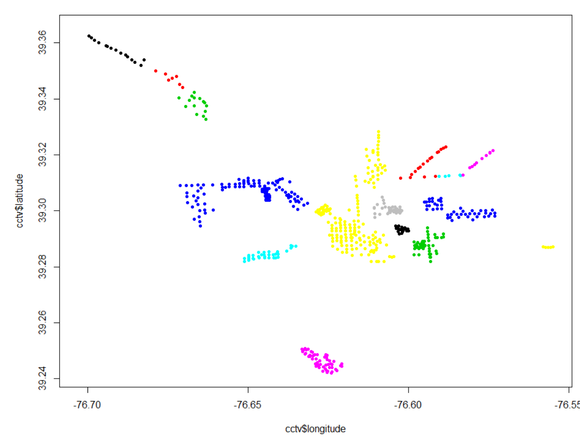

Figure 8: The result of DBSCAN clustering with CCTV data

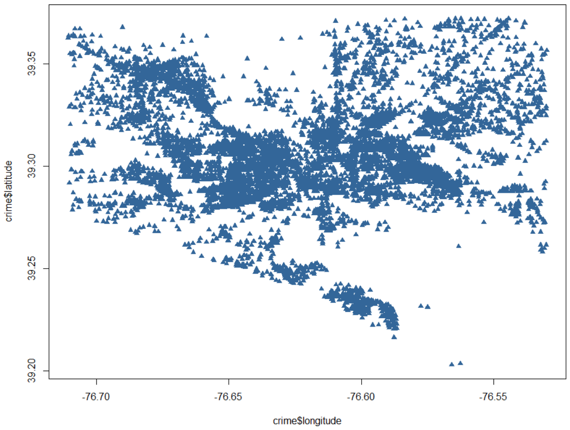

Figure 9: The result of plot with crime data

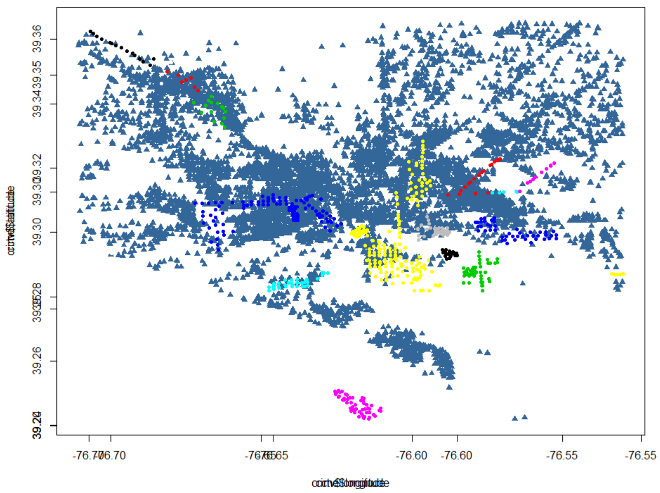

Figure 10: The result of plot with crime data and CCTV clustering data

First make density based clustering using DBSCAN. Data is CCTV data set's latitude and longitude. Using this data, eps is 0.004 and MinPts is 5. This means that the cluster has 5 points at least. The result of DBSCAN clustering is Fig. 8. It makes 15 clusters with CCTV data. And then, to confirm the relationship between CCTV data and crime data (Fig. 9), we have to plot this result with crime data. (Fig. 10). There are a lot of crimes in Baltimore, so we couldn't distinguish the relationship between CCTV and crime data easily.

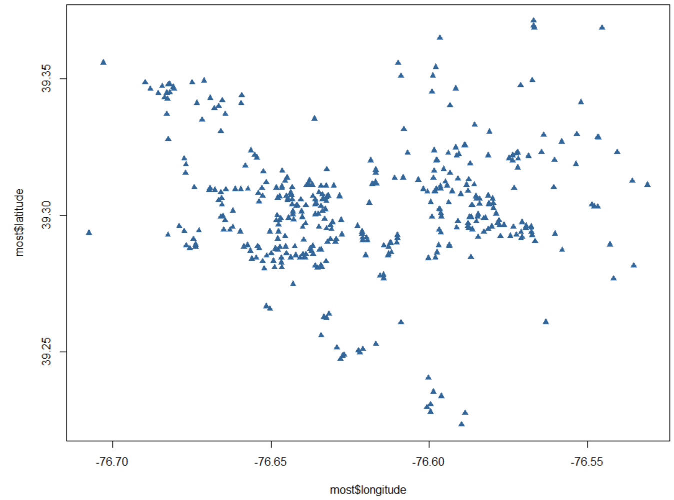

Figure 11: The result of plot with the most frequent crime data

Therefore, we need to make the crime data meaningfully. So that's why we use the most frequent crime data. (Fig. 11) This data is compression of crime data base on the districts of the Baltimore city. It shows that the most frequent crime occurrence area for each district of the city.

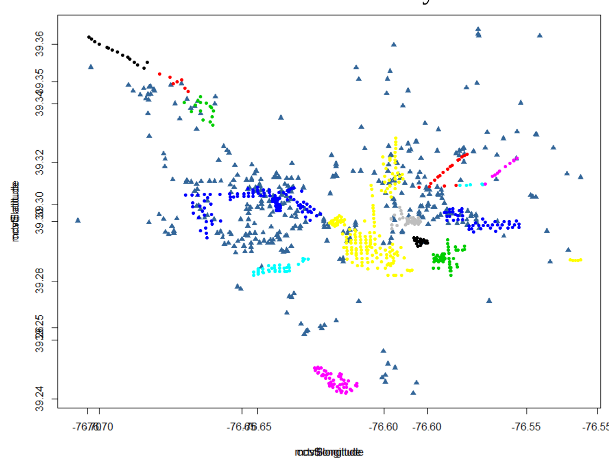

Figure 12: The result of plot the most frequent crime data and CCTV clustering data

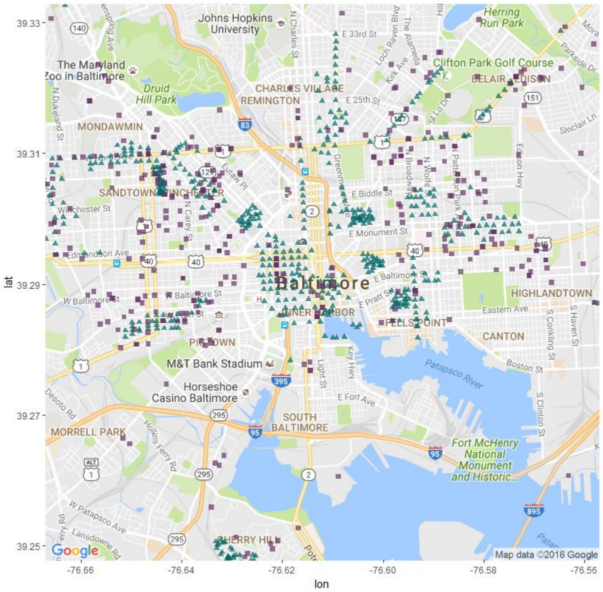

Figure 13: The most frequent crime data and CCTV clustering data using ggmap

To confirm the relationship between CCTV data and the most frequent crime data, we have to plot CCTV data and crime data together. In Fig.12, I found that the crimes tend to occur outside the area of CCTV. Still, some of crimes occurs in the area of CCTV. But much more crimes occurs outside the range of CCTV. It shows that CCTV installation effects the crime reduction. Because CCTV is surveillance system, It always keep watch to make us safe and it helps police to arrest offenders.

6. CONCLUSION

In this paper, we confirm the relationship between CCTV data and crime data. The result shows that many crimes tend to occurs outside the range of CCTV. Through this result, we can find out that CCTV system is very useful to keep our society more safe and it helps police to arrest offenders more easily. The CCTVs provide evidence that the people actually committed the crimes but do not deter them from the same. There are many crimes in our city nevertheless, the more CCTV is installed, the fewer crimes occurred. At the face of it, CCTV appears as one of the most effective ways to prevent occurrence of crimes.

REFERENCES

1. Maqsood, Amir, **Study of Big Data: An Industrial Revolution Review of applications and challenges**, International Journal of Advanced Trends in Computer Science and Engineering, vol. 6. no. 3, pp. 31-34, WARSE, June 2017.
2. Welsh, Brandon C.; Farrington, David P, **Public Area CCTV and Crime Prevention: An Updated Systematic Review and Meta-Analysis**, *Justice Quarterly*, vol. 26, No. 4, October 2009
3. Part, HyeonHo; Oh, GyeongSeok; Paek, SeungYeop, **Measuring the Crime Displacement and Diffusion of Benefit Effects of Open-street CCTV in South Korea**, Issue 3, vol. 40, *International Journal of Law, Crime and Justice*, pp. 179-191, September 2012.
4. Michael Hahsler; Matthew Piekenbrock; Sunil Arya; David Mount,. **Package 'dbscan'**, CRAN, ver 0.9-8, pp.2-4, August 2016.
5. Martin Ester, Hans-Peter Kriegel, Joerg Sander, Xiaowei Xu. **A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise**. Institute for Computer Science, University of Munich. *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining*, 1996.
6. David Kahle; Hadley Wickham. **ggmap: Spatial Visualization with ggplot2**, CONTRIBUTED RESEARCH ARTICLES, *The R Journal* Vol. 5/1, June 2013