BICYCLE THEFT IN SHEFFIELD

A STUDY OF THE UK POLICE DATASET USING R

JANUARY 22, 2018

YUE SHEN

Data Analytics, The University of Sheffield

Table of Contents

Abstract

Background. The literature reveals the situation and the trend of bicycle theft which occurs in Sheffield based on the analysis of the UK Police datasets with three external public datasets about boundary, council CCTV cameras and population. Previous surveys have all indicated bicycle theft is a common problem around the world.

Aims. The study aimed to summarise and visualise the situation of bicycle theft which occurs in Sheffield between December 2014 and November 2017. Besides, the study tried to understand the relation between bicycle theft and the outcome types of the records and aimed at forecasting the prospective trend of bicycle theft in this city as well.

Methods. The study used basic descriptive statistic methods in R to summarise and visualise data and conducted several inferential statistic methods such as Pearson's Chi-squared test and Holt-Winters Smoothing Method to test hypothesis and forecast trend.

Results. The choropleth map shows the most bicycle theft occurs in the city centre and some specific areas and streets has a high incidence. Bicycle thefts are more difficult to be identified the suspects than the other types of crime. In forecasting the Sheffield 2018 trend of bicycle theft, a gradually upward tendency is showed, and according to the forecast time series the highest incidence will occur around September.

Conclusion. The author conclude that people should be alert when parking bicycles in some specific areas and streets. Because of an upward tendency, government and the public should pay more attention to crime prevention to reduce the occurrence of bicycle theft.

1. Problem Definition of Bicycle Theft in Sheffield

On 9 January 2018, Sheffield launched a new bike-sharing scheme called "ofo", however, just one week later the police started to complain about such a hassle from the vandalised and dumped bikes (BBC News, 2018). Because bicycles are widely available, useful and easy to sell, they are very common to become the targets for crimes. A worldwide victim survey indicates that the possibility of experiencing bicycle theft is much higher than the other types of vehicle crimes (Tompson, 2012). Bicycle theft is one of the crime types reported to the UK police and monthly crime records are available in the UK Police Dataset from December 2014. So, what is the situation of bicycle theft like in Sheffield? Through analysing the dataset, we can gain insights about the crime rate and trends about bicycle theft in Sheffield, which are very helpful for us to make decisions about bicycling in the future.

2. Data Description and Pre-processing

To better understand the situation of bicycle theft in Sheffield, all the records of street-level crime from December 2014 to November 2017 in South Yorkshire were collected from the UK Police datasets. First, all the records of Sheffield were extracted through filtering the column of "LSOA name" which had a corresponding city name in its structure and then these records were sorted by the types of crime to get a bicycle theft dataset for further analysis. In addition, three external datasets were used during the research, namely boundary data from UK Data Service, census statistics for LSOA areas and current CCTV location details in Sheffield. Table 1 lists the details of the above four datasets and the URL links are available in the Appendix. In the process of data analysis, all these datasets were joined together over LSOA area identifiers.

Table 1. List of datasets used in this report

Dataset	Source	Region	Time
Street-level crime	DATA.POLICE.UK	South Yorkshire	From Dec. 2014 to Nov. 2017
Boundary	UK Data Service	Sheffield	2011 and later
Population	GOV.UK	England	Sep. 2015
CCTV Location	Sheffield City Council	Sheffield	July 11, 2017

3. Chosen Techniques for Data Analysis

Common descriptive statistic methods were used to summarise and present the meaningful features of dataset, like the percentage of bicycle theft among all crimes types in Sheffield, the distribution of last outcome and the places of high incidence of bicycle theft recent years. To find out whether there is more possibility to end in a no suspect identified outcome for a bicycle theft than the other types of crime, a Pearson's Chi-squared test with Yates' continuity correction was used to analyse the outcome records. Pearson's Chi-squared test is common for analysing categorical data among different groups and which would perform better with Yates' continuity correction for the analysis of 2 × 2 tables (Choi, Blume, & Dupont, 2015; Larose & Larose, 2014). Furthermore, a time series forecasting with Holt-Winters Smoothing Method was performed to predict the trend of bicycle theft based on the monthly frequency of recent years. The exponential and Holt–Winters techniques are sensitive to unusual events or outliers and which are just corresponding to recent Sheffield's bicycle theft records where obvious seasonal effects and erratic changes exist (Gelper, Fried, & Croux, 2010). Besides, a Box-Ljung test and a Shapiro-Wilk normality test were also conducted to test the result of forecasting. To better interpret the results of analysis, many powerful methods were used to visualise data, such as creating an outcome bar chart with *ggplot*, contrasting the bicycle theft locations and the positions of council CCTV cameras with *ggmap*, filling a choropleth map about frequencies in different areas with *tmap* and using *leaflet* to make an interactive map to show the incidents records on specific streets. In the following part, the results from carrying out the techniques mentioned above will be discussed.

4. Results and Discussion

4.1 WHERE: *Where Does Bicycle Theft Occur Most Frequently in Sheffield?*

The crime records of Sheffield were organised according to different types of crime, which showed there were totally 238,347 crime records between December 2014 and November 2017 in Sheffield and 1,903 bicycle theft incidents during this period coming to 0.8% of the total. The details of the comparison are reported in Table 2. The locations of bicycle theft are mapped by contrast with the positions of current council CCTV cameras (Figure 1). The map represents most of the incidents occurring alongside the roads and the cameras just have covered most areas in the city centre.

Table 2. Comparison of crime types in Sheffield between December 2014 and November 2017

Crime type	Count	Percentage
Anti-social behaviour	107,280	45.01%
Violence and sexual offences	30,598	12.84%
Criminal damage and arson	20,801	8.73%
Other theft	16,030	6.73%
Vehicle crime	15,471	6.49%
Burglary	14,513	6.09%
Shoplifting	13,455	5.65%
Public order	7,110	2.98%
Theft from the person	3,165	1.33%
Drugs	2,786	1.17%
Robbery	2,056	0.86%
Bicycle theft	1,903	0.8%
Other crime	1,882	0.79%
Possession of weapons	1,297	0.54%
Total	**238,347**	**100%**

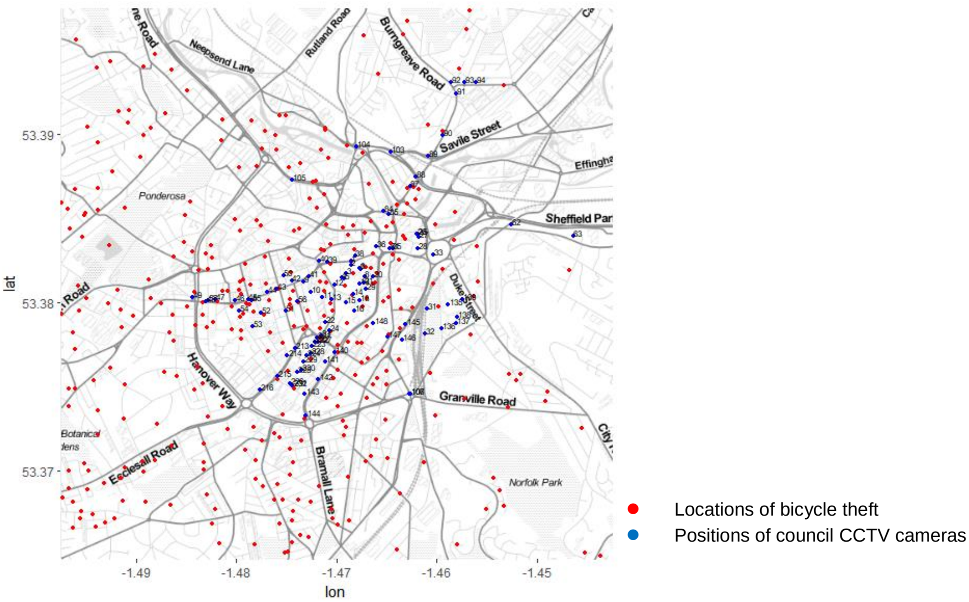
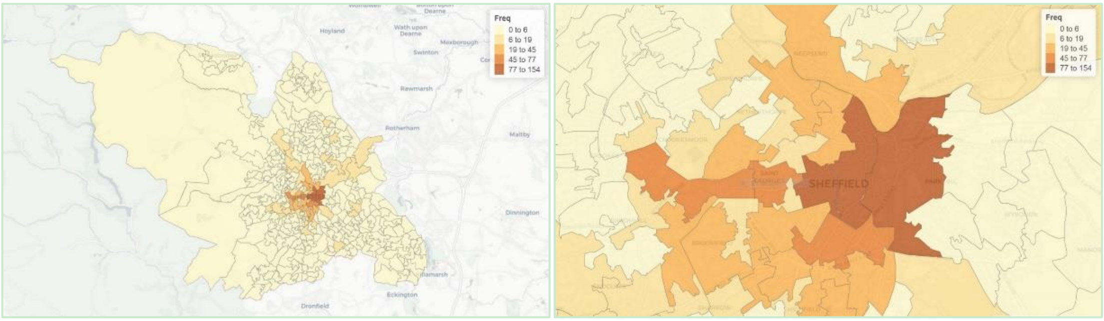


Fig. 1. Locations of bicycle theft contrasted with the positions of council CCTV cameras.

The frequencies of bicycle theft were counted by different LSOA areas in Sheffield and most of the incidences unsurprisingly occurred in the city centre (Figure 2). The crime rate of bicycle theft in Sheffield is about 1 incident per 1000 residents every year and the highest crime rate area is Sheffield 073D where nearly 30 incidents occurred per 1000 residents every year (Figure 3). Some high incidence places (e.g., parking area, shopping area, supermarket, educational building, etc.) and specific streets are listed in Table 3 with the total number of bicycle theft records. The incidents occurring on Bakers Hill, Brook Hill, Backfields and Cross Burgess Street, the four highest incidence streets of bicycle theft, are marked on Figure 4.

Table 3. High incidence areas and streets of bicycle theft

Area	Count	Street	Count
On or near Parking Area	67	On or near Bakers Hill	36
On or near Shopping Area	58	On or near Brook Hill	23
On or near Supermarket	50	On or near Backfields	22
On or near Further/Higher Educational Building	49	On or near Cross Burgess Street	16
On or near Conference/Exhibition Centre	39	On or near Upper Hanover Street	16
On or near Pedestrian Subway	31	On or near Durham Road	15
On or near Sports/Recreation Area	28	On or near Harrow Street	14
On or near Theatre/Concert Hall	25	On or near Sheaf Gardens	14
On or near Hospital	14	On or near Howard Lane	13
On or near Park/Open Space	14	On or near King Street	12



(a) (b)

Fig. 2. (a) Bicycle theft distribution in Sheffield and (b) Bicycle theft distribution in the city centre.

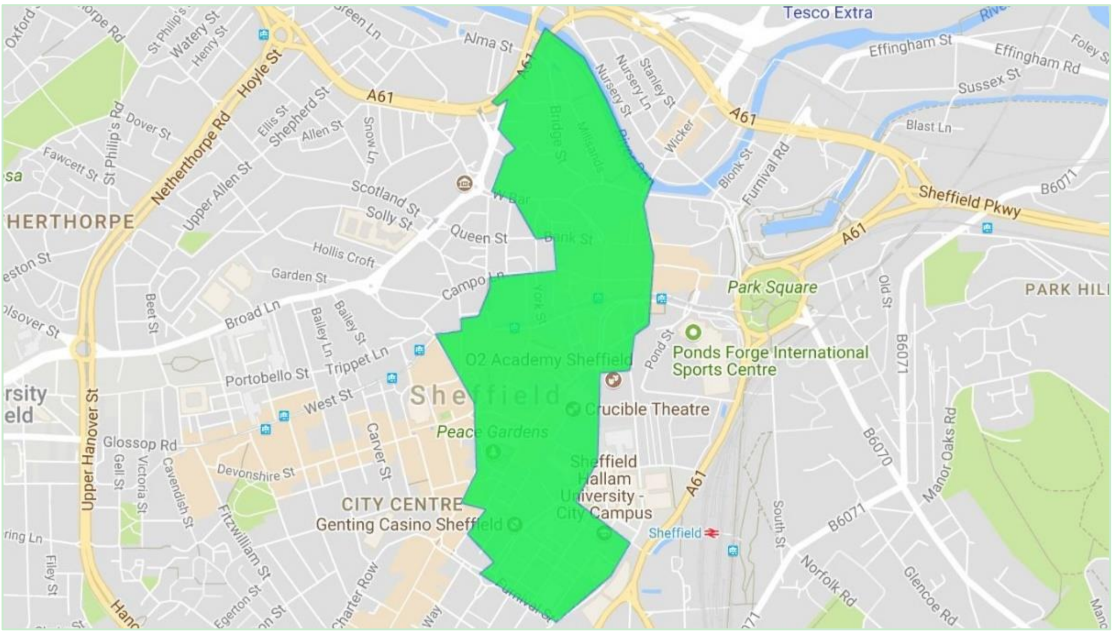


Fig. 3. LSOA name: Sheffield 073D; LSOA code: E01033264.

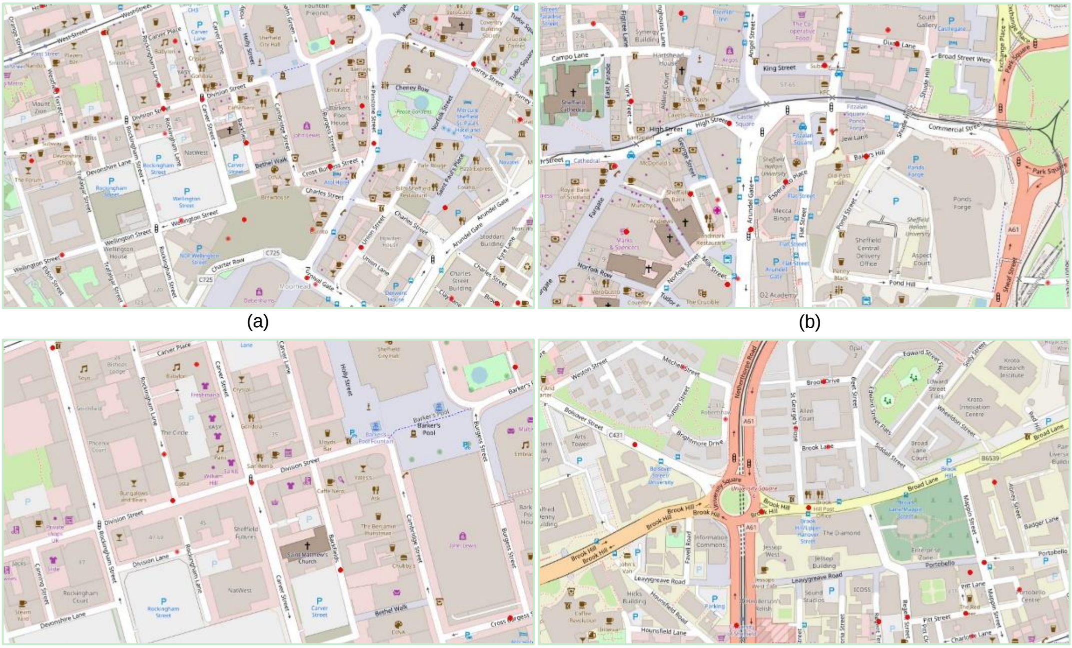


Fig. 4. (a) On or near Bakers Hill, (b) On or near Brook Hill, (c) On or near Backfields and (d) On or near Cross Burgess Street.

4.2 HOW: *How Does Most Outcomes of Bicycle Theft End in Sheffield?*

A bar chart is displayed in which proportions of different categories of last outcomes within different crime types are visualised in different colours and most of the outcomes of bicycle theft are occupied by the colour of green which stands for the outcome of "Investigation complete; no suspect identified" (Figure 5). Because the crime type of "Anti-social behaviour" has no records about last outcome, this category has been filtered out in the chart.

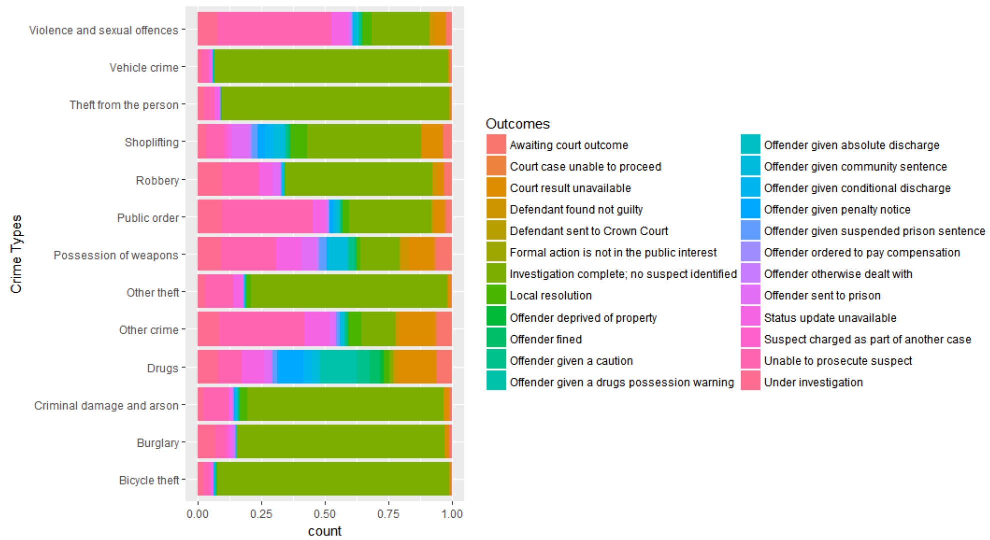


Fig. 5. Proportions of types of last outcomes within different types of crime.

To find out if there is a more significant possibility of bicycle theft ended in an outcome of no suspect identified than the other types of crimes, a chi-square test was used to check the hypothesis. First, the crime type and last outcome categories were simplified into two binary variables and then a chi-square test of independence was performed to examine the relation between the above variables. The result indicated that the relation between these variables was significant, $X^2\ (1, N = 131{,}067) = 879.56, p < .01$. Bicycle theft was more likely to end in an outcome of no suspect identified than the other types of crime in Sheffield. The output of Pearson's Chi-squared test with Yates' continuity correction is reported in Table 4.

Table 4. Crosstabulation of Crime Type and Last Outcome

Last Outcome	Crime Type		Row Total	X^2
	Bicycle Theft	Other Crimes		
Investigation complete; no suspect identified	1,741 (29.68)	74,500 (-29.68)	76,241	**879.56****
Other Outcomes	162 (-29.68)	54,664 (29.68)	54,826	
Colum Total	**1,903**	**129,164**	**131,067**	

Note. ** $= p \leq .01$. Adjusted standardised residuals appear in parentheses below group frequencies.

4.3 WHAT: *What Is the Sheffield 2018 Trend of Bicycle Theft?*

To forecast the trend of bicycle theft, a time series of the monthly incidents number is showed in Figure 6 (a) at first and then appended by a linear regression line as shown in Figure 6 (b) which indicates a gradually upward tendency. Figure (c) decomposes the observed time series into three components (i.e. trend, seasonal and random) and one of them, seasonal, is adjusted from the observed time series which is shown in Figure 6 (d). A Holt-Winters exponential smoothing forecasting method was used to predict the tendency of bicycle theft in the future with trend and additive seasonal component. A comparison between forecast and observed time series is shown in Figure 7 (a) and the forecasting of bicycle theft this year with two different confidence levels both indicates an upward tendency, and which is plotted in Figure 7 (b). To test the result of forecasting, a Box-Ljung test of residuals was performed to examine the randomness based on 20 lags. The serial correlation among residuals was non-significant, $X^2\ (20,\ N = 24) = 11.78, p = .92$, so the residuals were independently distributed. Furthermore, a plot of the residuals autocorrelations and a residuals time series both indicates the forecasting result is acceptable (Figure 8). A Shapiro-Wilk normality test of residuals was conducted ($W = 0.95, p = .23$), and the residuals were significantly normal distribution (*mean = 0*) which strengthened the reliability about the forecasting result further.

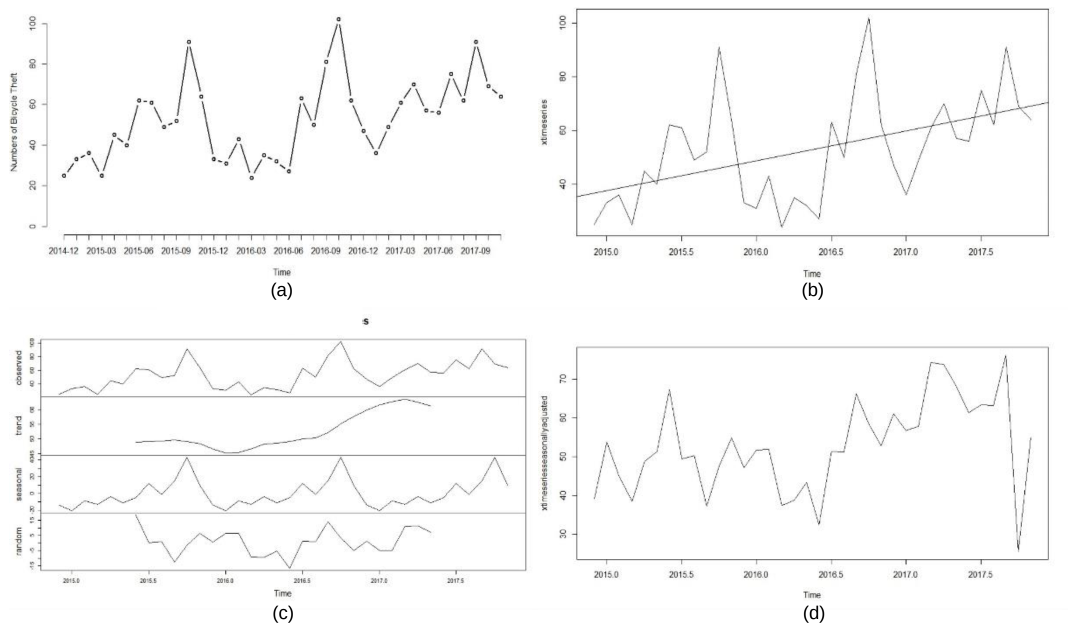
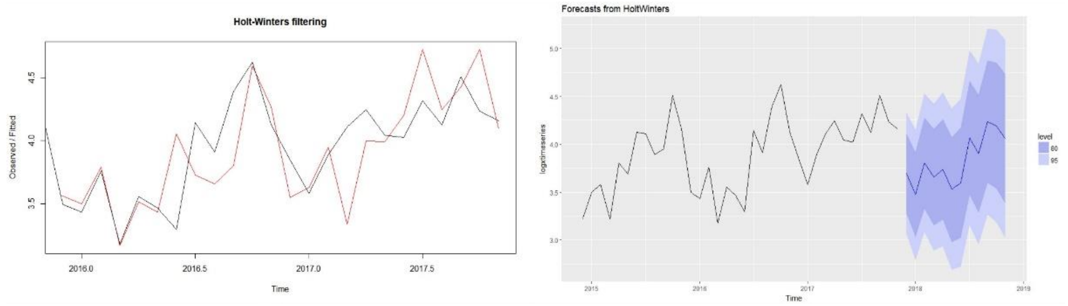


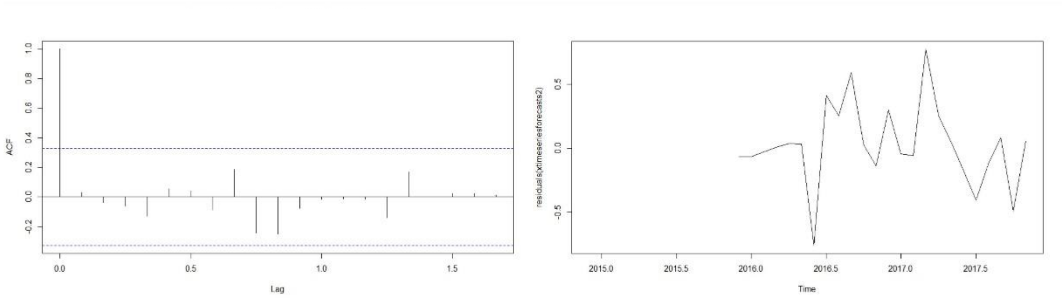
Fig. 6. (a) Numbers of bicycle theft occurred in Sheffield between December 2014 and November 2017,

(b) Time series appended by a linear regression line, (c) Decomposition of additive time series (observed, trend, seasonal and random time series from above to below) and (d) Seasonally adjusted time series.



(a) (b)

Fig. 7. (a) Observed time series appended by fitted forecasting time series in red and (b) Forecast trend of bicycle theft in 2018.



(a) (b)

Fig. 8. (a) Correlogram of the forecasting residuals and (b) Time series of the forecasting residuals.

5. Conclusions

This research has attempted to discover bicycle theft in Sheffield by asking (1) *Where* does bicycle theft occur most frequently in Sheffield; (2) *How* does most outcomes of bicycle theft end in Sheffield and (3) *What* is the Sheffield 2018 trend of bicycle theft. Choropleth map makes clear *where* thefts occur in Sheffield and most of the incidents happened in the parking, shopping or educational building areas of city centre. Some specific streets, such as Bakers Hill, Brook Hill and Backfields, show a very high incidence and these results remind us should be more alert about bicycling around the places mentioned above. Regarding the analysis of *how* most incidents of bicycle theft end, the most striking finding is bicycle theft is more difficult to be identified the suspects, that means if you lost your bike in Sheffield, the chance to recover it is very slim. With respect to forecast *what* the Sheffield 2018 trend of bicycle theft is, a gradually upward tendency of crime rate is showed, and the highest incidence will occur around September. Interventions to reduce bicycle theft are various, not only the police should act to prevent crime, but also the public should be more cautious about parking bikes no matter how short of the parking time is and how close the bike is to you.

References

BBC News. (2018). Police refuse to collect vandalised Sheffield ofo bikes. Retrieved 18 January, 2018, from http://www.bbc.co.uk/news/uk-england-south-yorkshire-42705531

Choi, L., Blume, J. D., & Dupont, W. D. (2015). Elucidating the Foundations of Statistical Inference with 2 × 2 Tables. *PLoS ONE, 10*(4). https://doi.org/10.1371/journal.pone.0121263

Gelper, S., Fried, R., & Croux, C. (2010). Robust Forecasting with Exponential and Holt – Winters Smoothing. *Journal of Forecasting*, *29*(June 2009), 285–300. https://10.1002/for. 1125

Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining* (2nd ed.). Hoboken: John Wiley & Sons.

Sidebottom, A. (2012). *Bicycle (bike) theft. JDiBrief Series*. London: UCL Jill Dando Institute of Security and Crime Science. Retrieved 18 January, 2018, from http://www. jdibrief.com

Appendix A: The URL Links of Datasets

Street-level crime:

https://data.police.uk/data/fetch/65bafe7b-b3ca-4606-8fac-31355fe8c35c/

Boundary:

https://borders.ukdataservice.ac.uk/bookmark.html?userDirectory=154E2E4A6CE9961FB81516393967117264%2F15163939676571788225469151640172&service=BOUNDARIES

CCTV Location:

https://data.sheffield.gov.uk/api/views/4d37-a5ib/rows.csv?accessType=DOWNLOAD

Population:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/467774/File_7_ID_2015_All_ranks__deciles_and_scores_for_the_Indices_of_Deprivation__and_population_denominators.csv

Appendix B: Codes for R Programming

```
library(sp)

library(RCurl)

library(ggplot2)

library(ggmap)

library(tidyr)

library(tmap)

library(rgdal)

library(kml)

library(maptools)

library(sp)
```

```
library(lubridate)
library(plotly)
library(dplyr)
library(TTR)
library(forecast)
library(leaflet)
# data preparation
sys201711 <- read.csv("2017-11\\2017-11-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
sys201710 <- read.csv("2017-10\\2017-10-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
sys201709 <- read.csv("2017-09\\2017-09-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
sys201708 <- read.csv("2017-08\\2017-08-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
sys201707 <- read.csv("2017-07\\2017-07-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
sys201706 <- read.csv("2017-06\\2017-06-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
sys201705 <- read.csv("2017-05\\2017-05-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
sys201704 <- read.csv("2017-04\\2017-04-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
sys201703 <- read.csv("2017-03\\2017-03-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
sys201702 <- read.csv("2017-02\\2017-02-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
sys201701 <- read.csv("2017-01\\2017-01-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
sys201612 <- read.csv("2016-12\\2016-12-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
sys201611 <- read.csv("2016-11\\2016-11-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
sys201610 <- read.csv("2016-10\\2016-10-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
sys201609 <- read.csv("2016-09\\2016-09-south-yorkshire-street.csv",header = TRUE, stringsAsFactors = FALSE)
```

```
sys201608 <- read.csv("2016-08\\2016-08-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201607 <- read.csv("2016-07\\2016-07-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201606 <- read.csv("2016-06\\2016-06-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201605 <- read.csv("2016-05\\2016-05-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201604 <- read.csv("2016-04\\2016-04-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201603 <- read.csv("2016-03\\2016-03-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201602 <- read.csv("2016-02\\2016-02-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201601 <- read.csv("2016-01\\2016-01-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201512 <- read.csv("2015-12\\2015-12-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201511 <- read.csv("2015-11\\2015-11-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201510 <- read.csv("2015-10\\2015-10-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201509 <- read.csv("2015-09\\2015-09-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201508 <- read.csv("2015-08\\2015-08-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201507 <- read.csv("2015-07\\2015-07-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201506 <- read.csv("2015-06\\2015-06-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201505 <- read.csv("2015-05\\2015-05-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201504 <- read.csv("2015-04\\2015-04-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201503 <- read.csv("2015-03\\2015-03-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201502 <- read.csv("2015-02\\2015-02-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys201501 <- read.csv("2015-01\\2015-01-south-yorkshire-street.csv",header = TRUE,
```

```
stringsAsFactors = FALSE)

sys201412 <- read.csv("2014-12\\2014-12-south-yorkshire-street.csv",header = TRUE,
stringsAsFactors = FALSE)

sys <- rbind(sys201711, sys201710, sys201709, sys201708, sys201707, sys201706,
sys201705, sys201704, sys201703, sys201702, sys201701, sys201612, sys201611,
sys201610, sys201609, sys201608, sys201607, sys201606, sys201605, sys201604,
sys201603, sys201602, sys201601, sys201512, sys201511, sys201510, sys201509,
sys201508, sys201507, sys201506, sys201505, sys201504, sys201503, sys201502,
sys201501, sys201412)

sys <- sys %>% extract(LSOA.name, into = c("city","area"), '^(.*)\\s+(.*)$')

sheffield_crime <- sys[sys$city=="Sheffield",]

sheffield_crime <- sheffield_crime[!(sheffield_crime$city == "" |
is.na(sheffield_crime$city)),]

write.csv(sheffield_crime, "sheffield_crime.csv")

# data description

type_table <- table(sheffield_crime$Crime.type)

type_table

prop.table(type_table)

round(prop.table(type_table)*100,2)

crime.type <- as.data.frame(type_table)

names(crime.type)=c("Type","Freq")

crime.type[order(-crime.type$Freq),]

colSums(crime.type["Freq"])

sheffield_crime2 <- subset(sheffield_crime, Crime.type!="Anti-social behaviour")

ggplot(data=sheffield_crime2)+

geom_bar(mapping=aes(x=sheffield_crime2$Crime.type,fill=sheffield_crime2$Last.outcome.
category),position = "fill")+xlab("Crime Types") +labs(fill="Outcomes") +coord_flip()

sheffieldggmap <- get_map(location = "Sheffield, UK", maptype="toner-lite", zoom = 14,
source = "google")

sheffield_crime_bt <- sheffield_crime[sheffield_crime$Crime.type=="Bicycle theft",]

cameras <- read.csv("Current_CCTV_Location_Details_2017.csv", header=TRUE)

head(cameras)

cameras2 <- cameras %>% extract(Coordinates, into=c("lat", "lon"), '^[(](.*),\\s+(.*)[)]$')

head(cameras2)

cameras2$lat<-as.numeric(as.character(cameras2$lat))
```

```
cameras2$lon<-as.numeric(as.character(cameras2$lon))

ggmap(sheffieldggmap) +

  geom_point(data = sheffield_crime_bt, aes(x=Longitude, y=Latitude), color="red", size=1)
+

  geom_point(data=cameras2, aes(x=lon, y=lat), size=1, color="blue") +

  geom_text(data = cameras2, aes(x=lon, y=lat, label=Cam_No), size=2, vjust=0, hjust=-0.1)
+

  guides(fill = FALSE, alpha = FALSE)

sheffieldShape <- readOGR(dsn = "./BoundaryData", layer = "england_lsoa_2011",
stringsAsFactors = FALSE)

codetable <- table(sheffield_crime_bt$LSOA.code)

codetable <- as.data.frame(codetable, stringsAsFactors = FALSE)

names(codetable) = c("Code", "Freq")

sheffieldShape@data <- left_join(sheffieldShape@data, codetable, by=c('code'='Code'))

sheffieldShape$Freq[is.na(sheffieldShape@data$Freq)] <- 0

tmap_mode("view")

tm_shape(sheffieldShape, projection = "rd") +

  tm_fill("Freq", style = "jenks", border.col = "black") +

  tm_borders(alpha = .5)

leaflet() %>%

  addTiles() %>%

  addCircleMarkers(lng = sheffield_crime_bt$Longitude, lat = sheffield_crime_bt$Latitude,

                   popup = sheffield_crime_bt$Month,

                   radius = 2,

                   stroke = TRUE, fillOpacity = 0.75, color = "red")

deprivation2015<-
read.csv("File_7_ID_2015_All_ranks__deciles_and_scores_for_the_Indices_of_Deprivation_
_and_population_denominators.csv", header=TRUE)

View(deprivation2015)

deprivation2015Pop <- deprivation2015 %>% select(LSOA.name..2011., LSOA.code..2011.,
Total.population..mid.2012..excluding.prisoners.)

names(deprivation2015Pop)[names(deprivation2015Pop)=="LSOA.name..2011."]<-
"LSOA_name"

names(deprivation2015Pop)[names(deprivation2015Pop)=="LSOA.code..2011."]<-
"LSOA_code"
```

```
names(deprivation2015Pop)[names(deprivation2015Pop)=="Total.population..mid.2012..excluding.prisoners."]<-"Total_population"

sheffieldShape@data<-left_join(sheffieldShape@data, deprivation2015Pop, by=c('code'='LSOA_code'))

sheffieldShape@data$crime.rate <- sheffieldShape@data$Freq / (sheffieldShape@data$Total_population*3)

colSums(sheffieldShape@data["Total_population"])

crime.rate.average=1903/(557382*3)

1000*crime.rate.average# the crime rate of bicycle theft is 1 incidents per 1000 residents everyyear in Sheffield

summary(sheffieldShape@data$crime.rate)# the highest crime rate place of bicycle theft (S1 1HA) is 30 incidents per 1000 residents everyyear

write.csv(sheffieldShape@data, "crime.rate.csv")

location_bt <- table(sheffield_crime_bt$Location)

location_bt <- as.data.frame(location_bt)

location_bt <- arrange(location_bt, desc(Freq))

View(location_bt)

write.csv(location_bt, "location_bt.csv")# the highest rate of place accured

# Chi-squared test

TO.data <- select(sheffield_crime2, Last.outcome.category, Crime.type)

a <- filter(TO.data, Last.outcome.category=="Investigation complete; no suspect identified" & Crime.type=="Bicycle theft") %>% count()

b <- filter(TO.data, Last.outcome.category!="Investigation complete; no suspect identified" & Crime.type=="Bicycle theft") %>% count()

c <- filter(TO.data, Last.outcome.category=="Investigation complete; no suspect identified" & Crime.type!="Bicycle theft") %>% count()

d <- filter(TO.data, Last.outcome.category!="Investigation complete; no suspect identified" & Crime.type!="Bicycle theft") %>% count()

c(a, b, c, d)

p <- as.table(rbind(c(1741,74500), c(162,54664)))

dimnames(p) <- list(Outcome = c("Investigation complete; no suspect identified","Other Outcomes"),

               Crime = c("Bicycle theft", "Other Crimes"))

(Xsq <- chisq.test(p))  # Prints test summary

Xsq$observed   # observed counts (same as M)

Xsq$expected   # expected counts under the null
```

```
Xsq$residuals  # Pearson residuals
Xsq$stdres     # standardized residuals
 # time series analysis
bt_table <- table(sheffield_crime_bt$Month)
plot(bt_table, type="b", xlab="Time", ylab="Numbers of Bicycle Theft")
bt_table <- as.data.frame(bt_table)
xtimeseries<-ts(bt_table$Freq,frequency = 12,start=c(2014,12))
plot.ts(xtimeseries)
abline(lm(xtimeseries~time(xtimeseries)))
xtimeseriesSMA<-SMA(xtimeseries,n=3)
plot.ts(xtimeseriesSMA)
xtimeseriescomponents<-decompose(xtimeseries)
plot(xtimeseriescomponents)
xtimeseriesseasonallyadjusted<-xtimeseries-xtimeseriescomponents$seasonal
plot(xtimeseriesseasonallyadjusted)
logxtimeseries<-log(xtimeseries)
xtimeseriesforecasts<-HoltWinters(logxtimeseries)
xtimeseriesforecasts
plot(logxtimeseries)
plot(xtimeseriesforecasts)
xtimeseriesforecasts2<-forecast(xtimeseriesforecasts,h=12)
autoplot(xtimeseriesforecasts2)
acf(xtimeseriesforecasts2$residuals,lag.max=20, na.action = na.pass)
Box.test(xtimeseriesforecasts2$residuals,lag=20,type="Ljung-Box")
residuals(xtimeseriesforecasts2)
plot(residuals(xtimeseriesforecasts2))
b <- xtimeseriesforecasts2$residuals
b <- as.data.frame(b)
b <- na.omit(b)
b <- as.numeric(b$x)
shapiro.test(b)
mean(b)
```