

Big Data: Data Analysis Boot Camp

Looking at Real-World Crime Data

Chuck Cartledge, PhD

21 January 2018

Table of contents (1 of 1)

1 Intro.

2 Data source

- Details

3 Questions

4 Bootstrapping

- R functions

5 Hands-on

6 Q & A

7 Conclusion

8 References

9 Files

What are we going to cover?

- 1 Look at publicly available police incident data from Virginia Beach, VA
- 2 Ask different questions about the data
- 3 Explore different ways to present the data



Data can be downloaded from here:

“This dataset includes information about incidents where the police department responds to an offense and a report of crime is generated. This dataset excludes incidents assigned to 14 of the 152 Incident Based Reporting Codes (IBR). The specific IBR codes excluded are: 'runaway', 'death investigation', 'death, accidental', 'death, drowning', 'death, suicide', 'death, auto fatality', 'attempted suicide', 'officer involved shooting, death', 'officer involved shooting, no death', 'missing person', 'lost property', 'habitual offender', 'other non-reportable offenses', and 'SVU information only.’” VB Staff[1]

The screenshot shows a web browser displaying the 'Police Incident Reports' dataset page. The page title is 'Police Incident Reports' with a sub-tab for 'Public Safety'. A search bar is visible at the top right. Below the title, there is a brief description of the dataset and its update date (December 17, 2017). A table provides details about the dataset, including the number of records (4,294) and the number of attachments (257). The table also lists the categories and tags associated with the data. The page is titled 'About this Dataset' and includes sections for 'Updated', 'Data last updated', 'Data created', 'Version', 'Attachments', 'Topics', and 'Data provided by'.

About this Dataset	
Updated	December 17, 2017
Data last updated	December 17, 2017 June 13, 2017
Data created	September 28, 2016
Version	1.0.0
Records	4,294
Attachments	257
Category	Public Safety
Tags	police, incidents, crime reports, burglary, violent property, assault, drug abuse
Data provided by	Chadler Decker Public Safety Administration

[https://data.vbgov.com/
Public-Safety/Police-
Incident-Reports/iqkq-gr5p](https://data.vbgov.com/Public-Safety/Police-Incident-Reports/iqkq-gr5p)

Same image.

The screenshot shows a web browser window displaying the City of Virginia Beach Data Portal. The URL in the address bar is `data.vbgov.com/Public-Safety/Police-Incident-Reports/iqkq-gr5p`. The page features the City of Virginia Beach logo and a search bar. A navigation menu includes links for 'VBGov Home', 'Catalog', 'Suggest a Dataset', 'About this Site', 'Tutorials', 'Developers', 'Open Performance', and 'Open Budget'. The main content area is titled 'Police Incident Reports' with a 'Public Safety' tag. It includes a description of the dataset, an 'Updated' date of December 17, 2017, and a table of statistics showing 4,294 views and 257 downloads. There are also sections for 'Attachments' (listing documents like 'Police Incidents - Dataset Information Worksheet.docx'), 'Topics' (listing 'Public Safety'), and 'Tags' (listing 'police, incident, crime, reports, burglary, violent, property, assault').

City of Virginia Beach

VBGov Home Catalog Suggest a Dataset About this Site Tutorials Developers Open Performance Open Budget

Police Incident Reports Public Safety

Explore Data Export API Share

This dataset includes information about incidents where the police department responds to an offense and a report of crime is generated. This dataset excludes incidents assigned to 14 of the 152 Incident Based Reporting Codes (IBR). The specific IBR codes excluded are: 'runaway', 'death investigation', 'death, accidental', 'death, drowning', 'death, suicide', 'death, auto fatality'...

Updated December 17, 2017
Data Provided by Nathalie Pastrana

About this Dataset

Updated December 17, 2017

Data Last Updated December 17, 2017
Metadata Last Updated June 13, 2017

Date Created September 28, 2016

Views	Downloads
4,294	257

Data Provided by Nathalie Pastrana
Dataset Owner Open Data Administrator

Attachments

- Police Incidents - Dataset Information Worksheet.docx
- Excluded Offenses.docx
- Status List.docx
- Precincts and Zones Map.pdf

Topics

Category Public Safety

Tags police, incident, crime, reports, burglary, violent, property, assault
Show More

Licensing and Attribution

`https://data.vbgov.com/Public-Safety/Police-Incident-Reports/iqkq-gr5p`

Download a CSV file.

Police Incident Reports | Open Data - Opera

Police Incident Repo x Precincts_and_Zone x +

data.vb.gov.com/PublicSafety/Police-Incident-Reports/iqkq-gr5p

City of Virginia Beach

VBGov Home Catalog Suggest a Dataset About this Site Tutorials Developers Open Performance Open Budget

Police Incident Reports Public Safety

Explore Data Export API Share ...

This dataset includes information about incidents where the police department responds to an offense and a report of crime is generated. This dataset excludes incidents assigned to 14 of the 152 Incident Based Reporting Codes (IBR). The specific IBR codes excluded are: 'runaway', 'death investigation', 'death, accidental', 'death, drowning', 'death, suicide', 'death, auto fatality', 'attempted suicide', 'officer involved shooting, death', 'officer involved shooting, no death', 'missing person', 'lost property', 'habitual offender', 'other non-reportable offenses', and 'SVU information only'. [Less](#)

Download Police Incident Reports

Download Police Incident Reports for offline use in other applications.

CSV CSV for Excel

Additional Formats

[CSV for Excel \(Europe\)](#) [TSV for Excel](#)

[RDF](#) [XML](#)

[BSS](#)

About this Dataset

Updated
December 17, 2017

Data Last Updated December 17, 2017 Metadata Last Updated June 13, 2017

Date Created September 28, 2016

Views **4,294** Downloads **257**

Attachments

- Police Incidents - Dataset Information Worksheet.docx
- Excluded Offenses.docx
- Status List.docx
- Precincts and Zones Map.pdf

Topics

Category Public Safety

6/21

Details

The CSV file has a header line identifying the various data fields:

- | | | | |
|---|--------------------|----|---------------------|
| 1 | Police Case Number | 6 | Offense Description |
| 2 | Date Reported | 7 | Subdivision |
| 3 | Date Occurred | 8 | Zone ID |
| 4 | Date Found | 9 | Case Status |
| 5 | Offense Code | 10 | Location |

Each location cell has three lines.

Lots of questions can be asked.

- 1 How many of the incidents apply to Virginia Beach?
- 2 How many unique incidents are there (some incidents have the same time and location)?
- 3 Is there a seasonal variation (time series)?
- 4 What is the incident per capita ratio? Does it vary across time, or season?
- 5 Where are most of the incidents (geospatial)?
- 6 What is the best location for a police precinct (dendrogram and cut height)?
- 7 What is the best location for new police precinct (minimization problem)?
- 8 What is the best location to buy a new home (local minimum problem)?

R functions

main()

```
1 main <- function ()
2 {
3   savedCrimeData <- file.path("../", "Data", "savedCrimeData.rdat")
4
5   crimeData <- initCrimeData(useExisting=TRUE, savedFile=savedCrimeData ,
6                               cityNumber=1)
7
8   geoPlotCrimeData (crimeData)
9
10  print("The program has ended.")
11 }
```

dumpObject()

```
1 dumpObject <- function(object, comment=" ", verbose=TRUE)
2 {
3   if(verbose == TRUE)
4   {
5     print(sprintf("%s — Dumping the object: %s (of type: %s, class: %s)",
6                 comment,
7                 deparse(substitute(object)),
8                 typeof(object),
9                 class(object)
10                )
11          )
12
13     print(object)
14   }
15 }
```

geoPlotCrimeData()

```
1 geoPlotCrimeData <- function(crimeData, zoom = 10)
2 {
3   lons <- range(crimeData$lon)
4   lats <- range(crimeData$lat)
5
6   crimeData$color <- "red"
7
8   saveFileName <- file.path("..", "Data", sprintf("%f-%f-%f-%f-%f", lons[1],
9     lats[1], lons[2], lats[2], zoom))
10
11  map <- getBaseMap(lats, lons, useSaveExisting=TRUE, saveFileName, zoom)
12
13  foo <- ggmap(map) +
14    scale_x_continuous(limits = lons, expand = c(0, 0)) +
15    scale_y_continuous(limits = lats, expand = c(0, 0)) +
16    geom_point(data=crimeData[,c("lon", "lat")], col=crimeData$color)
17
18  print(foo)
19 }
```

R functions

getBaseMap()

```
1 getBaseMap <- function(lats, lons, useSaveExisting, saveFileName, zoom)
2 {
3   doWork <- TRUE
4   if (useSaveExisting == TRUE)
5   {
6     if (file.exists(saveFileName) == TRUE)
7     {
8       load(saveFileName)
9       doWork <- FALSE
10    }
11  }
12
13  if (doWork == TRUE)
14  {
15    map <- get_map(location = c(lon = mean(lons), lat = mean(lats)),
16                    zoom = zoom,
17                    maptype = "satellite",
18                    source = "google")
19
20    if (useSaveExisting == TRUE)
21    {
22      save(map, file=saveFileName)
23    }
24  }
25  map
26 }
```

initCrimeData() (1 of 3)

```
1 initCrimeData <- function(useExisting=TRUE, savedFile=tempfile(), cityNumber=1)
2 {
3   dataFile <- file.path("../", "Data", "Police_Incident_Reports.csv")
4
5   switch(as.character(cityNumber),
6     "1" = {
7     northernLat <- 37.5
8     southernLat <- 36
9     westernLon <- -76.5
10    easternLon <- -75
11    },
12    "2" = {
13    northernLat <- 90
14    southernLat <- -90
15    westernLon <- -180
16    easternLon <- 180
17    }
18  )
19
20  doWork <- TRUE
21
22  if (useExisting == TRUE)
23  {
24    if (file.exists(savedFile) == TRUE)
25    {
26      load(savedFile)
27      doWork <- FALSE
28    }
29  }
```

R functions

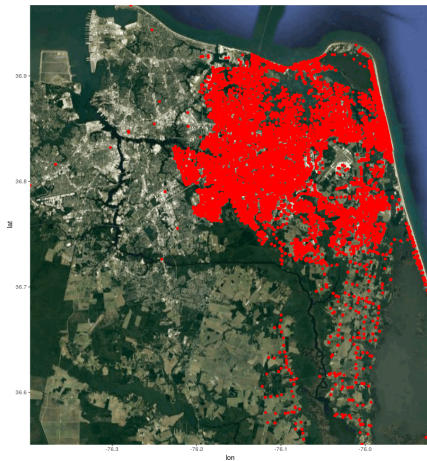
initCrimeData() (2 of 3)

```
29   }
30
31   if (doWork == TRUE)
32   {
33     d <- readLines(dataFile)
34     numberOfRecords <- length(d)
35     ## numberOfRecords <- 100
36     dumpObject(numberOfRecords, comment="Number of crime related data
           records")
37
38     lats <- c()
39     lons <- c()
40     for (i in seq(4, numberOfRecords, by=3))
41     {
42       ## temp <- unlist(regmatches(d[i], gregexpr('\\(?:[0-9.,]+', d[i])))
43       temp <- as.numeric(unlist(regmatches(d[i], gregexpr('-?[0-9.,]+', d[i]
           ]))))
44       if ((is.na(temp[1]) == FALSE) && (is.na(temp[2]) == FALSE))
45       {
46         if ((southernLat <= temp[1] && (temp[1] <= northernLat) &&
47             (westernLon <= temp[2] && (temp[2] <= easternLon))
48             )
49         {
50           lats <- c(lats, temp[1])
51           lons <- c(lons, temp[2])
52         }
53       }
54     }
55   }
```

initCrimeData() (3 of 3)

```
55     returnValue <- data.frame(lat=lats , lon=lons)
56
57     save(returnValue , file=savedFile)
58 }
59
60 dumpObject(nrow(returnValue) , comment="Number of crime locations")
61 returnValue
62
63 }
```

The results.



Now the fun begins!

The screenshot shows the RStudio interface with a script editor on the left and the Environment and Console panes on the right.

Script Editor (backends.R):

```

1 # http://thinkstart.com/twitter-authentication-with-r/
2 # https://nknanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
3
4 rm(list=ls())
5
6 library(openssl)
7 library(httr)
8 library(jsonlite)
9 library(bitops)
10 library(RCurl)
11 # library(rjson)
12 library(RMySQL)
13 library(DBI)
14 library(RPostgreSQL)
15 library(streamR)
16
17
18 source("library.R")
19
20 searcher <- function(dbCon, table, configurationData, refreshStrings, token=NULL)
21 {
22   # https://cran.r-project.org/web/packages/jsonlite/vignettes/json-apis.html
23   # Create your own application key at https://dev.tw
24   # https://dev.twitter.com/rest/reference/get/search/tweets
25
26 }

```

Environment Pane: Global Environment - Environment is empty

Console:

```

R version 3.3.1 (2016-06-21) -- "Bug In Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

```

Q & A time.

Q: How do you shoot a blue elephant?

A: With a blue-elephant gun.

Q: How do you shoot a pink elephant?

A: Twist its trunk until it turns blue, then shoot it with a blue-elephant gun.



What have we covered?

- Looked at publicly available police incident data in different ways
- Applied tools and techniques to real-world data



Next: Wrap up and harder and more interesting questions

References (1 of 1)

- [1] Virginia Beach Staff, Police Incident Reports, <https://data.vbgov.com/Public-Safety/Police-Incident-Reports/iqkq-gr5p>, 2017.

Files of interest

- 1 Bootstrap program 