UK CRIMES REPORT

Abstract

Background: Recently R as a data analysis tool has become widely popular among data scientists. However, outputs in R have not always been easy to interpret for the laymen.

Aim: We have aimed to harness the power of R to provide a visual and written analysis of an important dataset regarding crimes in the UK.

Methodology: This report has been produced by using R functions and packages which were appropriate for the aggregation and analysis of the key variables present in our dataset. Comparisons of finding were also made with regards to past reports on crime in the UK.

Conclusions: Like in many other countries, the level of crimes in the UK is seasonal and is concentrated around major cities. It was also found that the data provided by the police forces lack completeness.

Contents

Problem definition

As citizens, we have always been challenging the government for better security policies and operations. We would like to see decisions made with the aid of clear evidence and actual information rather than just instincts. As the data provided by the Home Office is so fragmented, we would like to be able to aggregate all the police files to be able to make the overarching inferences about crime occurrences and discover potential ideas on improving our police operations.

Data Description

The UK Police data has been made available by the Home Office via the site data.police.uk. It is a collection of crimes reported by 44 police forces every month. Please note that while the Metropolitan Police Service deals with the greater London area as well national tasks, the city of London Police is also responsible for the city of London area within London.

```
> freq(crime_2017$Falls.within)
```

	var	frequency	percentage	cumulative_perc
1	Metropolitan Police Service	960836	15.92	15.92
2	Greater Manchester Police	380661	6.31	22.23
3	West Yorkshire Police	294282	4.88	27.11
4	West Midlands Police	262817	4.35	31.46
5	South Yorkshire Police	194582	3.22	34.68
6	Northumbria Police	193304	3.20	37.88
7	Kent Police	192628	3.19	41.07
8	Hampshire Constabulary	190336	3.15	44.22
9	Lancashire Constabulary	177786	2.95	47.17
10	Thames Valley Police	171920	2.85	50.02
11	Avon and Somerset Constabulary	169818	2.81	52.83
12	Essex Police	167843	2.78	55.61
13	Merseyside Police	158208	2.62	58.23
14	Police Service of Northern Ireland	145370	2.41	60.64
15	Sussex Police	139997	2.32	62.96
16	Devon & Cornwall Police	133977	2.22	65.18
17	South Wales Police	132541	2.20	67.38
18	West Mercia Police	119635	1.98	69.36
19	Nottinghamshire Police	117851	1.95	71.31
20	Hertfordshire Constabulary	109853	1.82	73.13
21	Staffordshire Police	109250	1.81	74.94
22	Cheshire Constabulary	96724	1.60	76.54
23	Humberside Police	92369	1.53	78.07
24	Cleveland Police	90711	1.50	79.57
25	Leicestershire Police	89190	1.48	81.05
26	Derbyshire Constabulary	88925	1.47	82.52
27	Surrey Police	88835	1.47	83.99
28	Cambridgeshire Constabulary	80768	1.34	85.33
29	Northamptonshire Police	74989	1.24	86.57
30	Durham Constabulary	73765	1.22	87.79
31	Dorset Police	70583	1.17	88.96
32	Norfolk Constabulary	69375	1.15	90.11
33	North Yorkshire Police	62890	1.04	91.15
34	Bedfordshire Police	61013	1.01	92.16
35	North Wales Police	58924	0.98	93.14
36	Suffolk Constabulary	58623	0.97	94.11
37	Gwent Police	57379	0.95	95.06
38	Wiltshire Police	56383	0.93	95.99
39	Lincolnshire Police	54745	0.91	96.90
40	Gloucestershire Constabulary	54048	0.90	97.80
41	Warwickshire Police	53573	0.89	98.69
42	Dyfed-Powys Police	37772	0.63	99.32
43	Cumbria Constabulary	35809	0.59	99.91
44	City of London Police	5229	0.09	100.00

The data are anonymised and quality assured before being made available publicly (Home Office UK, 2017). There are 12 fixed or standardised columns for all the files irrespective of the police force. Below is the list of the columns in the files. A further explanation is given in the appendix section.

```
> colnames(crime_2017)
 [1] "Crime.ID"               "Month"                 "Reported.by"           "Falls.within"          "Longitude"
 [6] "Latitude"               "Location"              "LSOA.code"             "LSOA.name"             "Crime.type"
[11] "Last.outcome.category"  "Context"
```

The field context is a new feature added by UK police, and is at present always empty.

We have joined all the police forces files for each month and then join all monthly data for the period January – November 2017 using Rstudio.

Chosen Techniques

The biggest challenge was collecting and collating the data in such a way that our goal of scaling up the analysis to a national level could be met. We used the following functions, so we can be efficient in importing multiple datasets at once.

files = list.files(pattern="*.csv")

data_list = lapply(files, read.csv, header = TRUE)

We also used the do.call() function which allow us to use a list instead of calling several arguments in our function. The rbind() function helped us join our datasets together by rows since all the columns were the same in all the datasets.

The other challenge we faced was that the police dataset had categorical variables which were difficult to work with at times. As such, we decided to work on the frequencies of occurrences of the variables and try to present them in a visualisation which is easy to the eye for the public. The funModelling R package provides some functions which help the data scientist in their daily tasks such as efficient data preparation and analysis (Casas, 2017). We have used the function freq() as it allows to get the frequencies tables as well as charts with just one line of code.

Results and discussions

General number of crimes committed in England, Wales and Northern Ireland

Total number of crimes reported by each police force

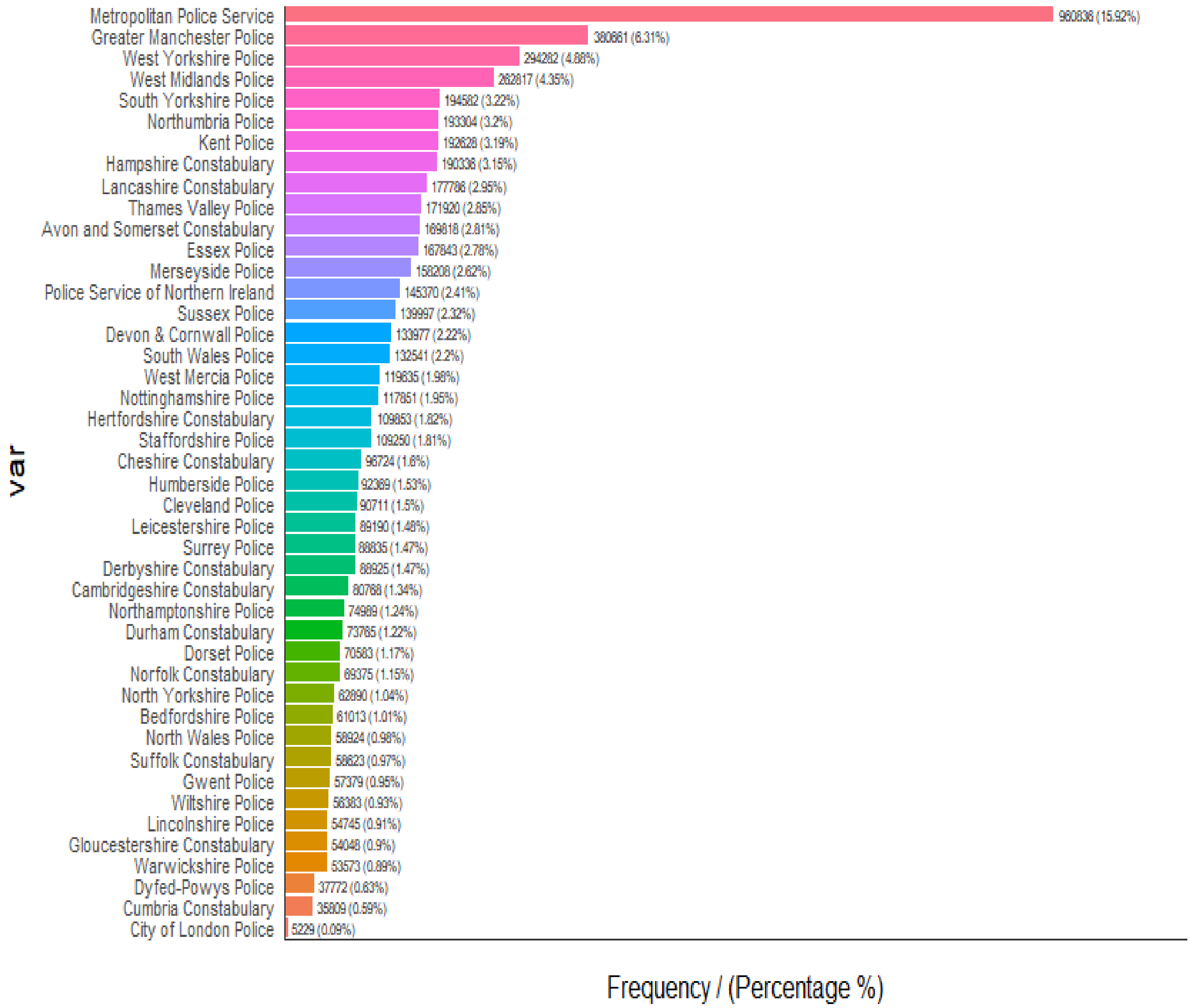

Figure 1 Reported Crimes by Police Forces

A total of 6,036,117 crimes were committed in England, Wales and Northern Ireland between January and November 2017. This figure is considerably higher than the figures provided by (Office for National Statistics, 2017) for the year ending June 2017 which was 5.2 million offences.

We can see from the above chart that the London Metropolitan Service has reported the largest amount of crimes committed between January and November 2017. At 960,836, this represents 15.92% of all the crime committed in the UK between January and November 2017. At the opposite end, the police force with the smallest number crimes reported is the City of London Police with 5,229 which is 0.09 % of the total crimes committed in the UK between January 2017 and November 2017. Crimes in South Yorkshire where we live accounted for 3.22% of the total with 194,582 crimes committed. This is the 5th biggest region in terms of crimes after Greater London, Greater Manchester, West Yorkshire and West Midlands respectively. The following map shows the concentration of crimes in the UK. We are obviously able to deduct visually that the more populous and the regions around major cities have more crimes.

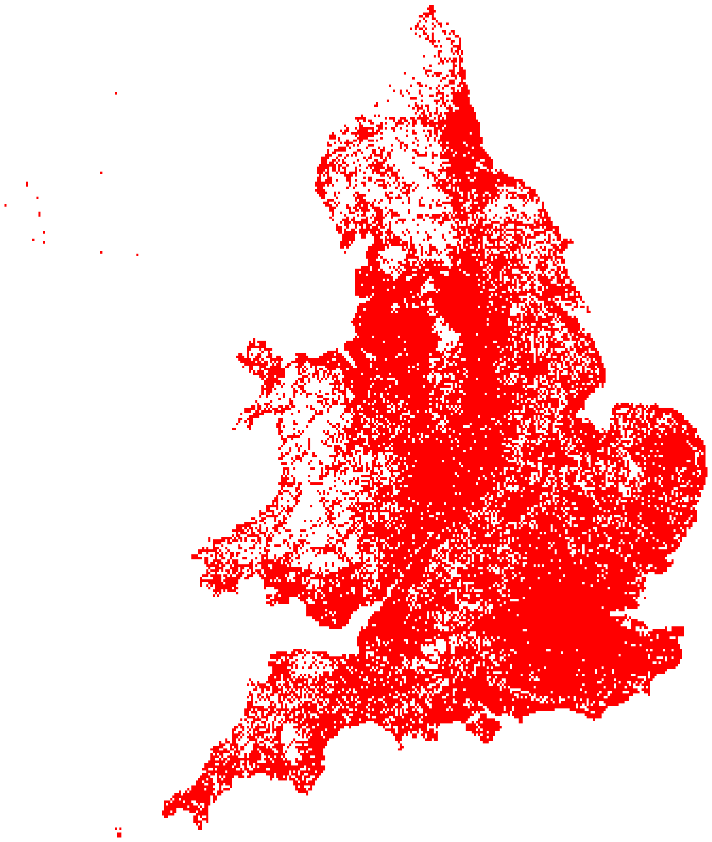

Figure 2 A map of the UK showing the distribution of crimes

What types of crimes are committed in England, Wales and Northern Ireland?

It is worth noting that the data made available by UK Police does not include fraudulent offences as well as digital crimes(Office for National Statistics, 2017). UK Police is currently working to improve the recording of such data and subsequently their sharing with the public(Home Office UK, 2017).

Crimes in the UK are grouped by the Home Office into 14 categories which are:

Anti-social behaviour

Bicycle theft

Burglary

Criminal damage and arson

Drugs

Other crime

Other theft

Possession of weapons

Public order

Robbery

Shoplifting

Theft from the person

Vehicle crime

Violence and sexual offences

A further description of crimes is given in the appendix section of this document. We are unsure about the reasoning by the home office to group both violence and sexual offences together. This makes a bit harder to explore in detail on such an important topic that is sexual offences.

From our dataset, we observe that Anti-Social behaviour offences are the most widespread type of crime in the UK. Anti-social behaviour offences account for 1,602,332 of the total crimes committed or 26,55%. However this figure shows a decrease compared to the figures of the year ending June 2017 of 1.8 million which in turn is also a 1% decrease on the figures of the year ending June 2016(Office for National Statistics, 2017). Violence and sexual offences is now the second highest category of crimes in the UK , owing it to the general rise of sexual crimes and their reporting (Office for National Statistics, 2017). The figure of 1,372,163 for violence and sexual offences is more than double the next highest type of offences which is criminal damage and arson at 559,037.

Possession of weapons is the lowest type of crime in the UK at 0.61% of the total figure. The chart below shows a detailed breakdown of the types of crime in the UK.

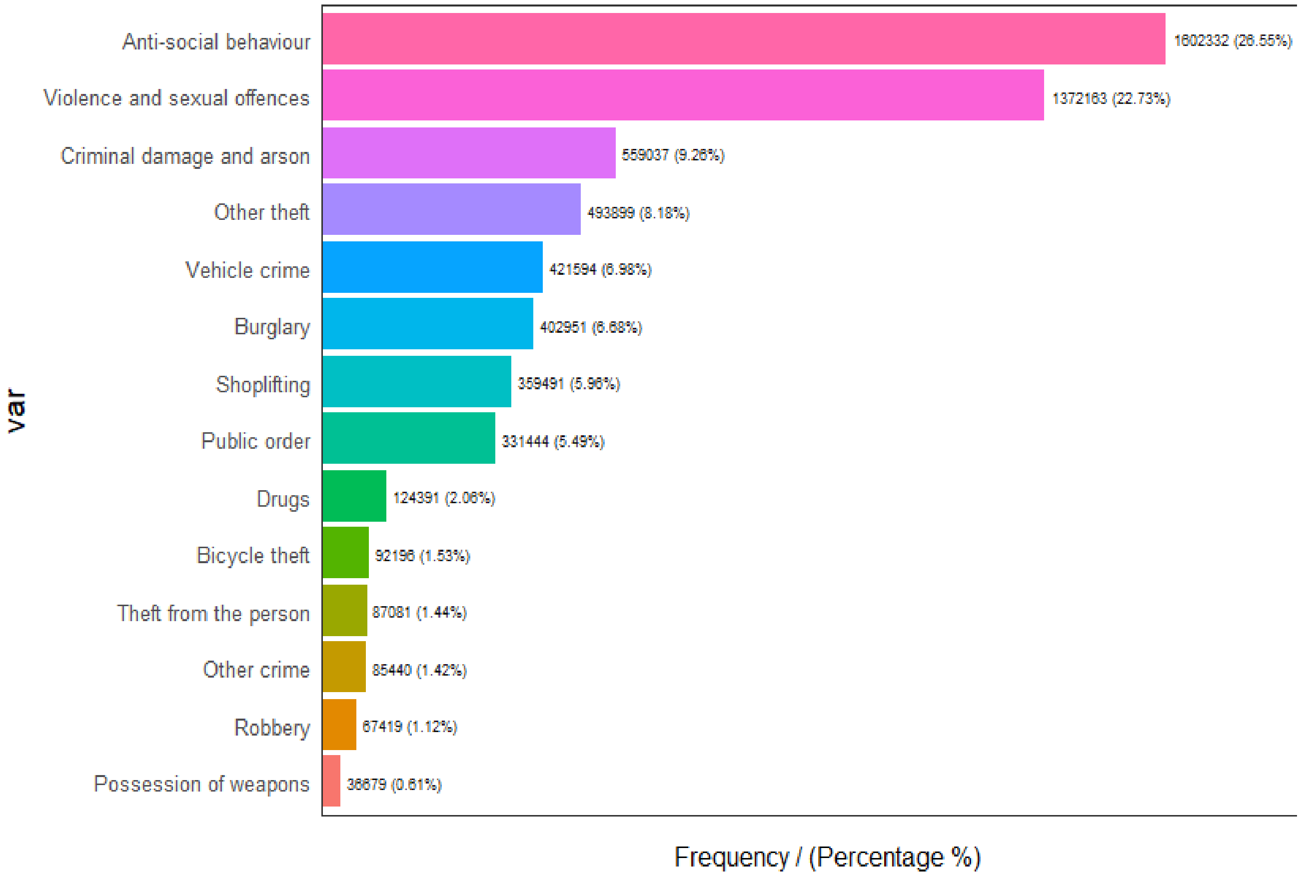

Figure 3 Types of crimes in the UK

Number of crimes in the UK per month between January and November 2017

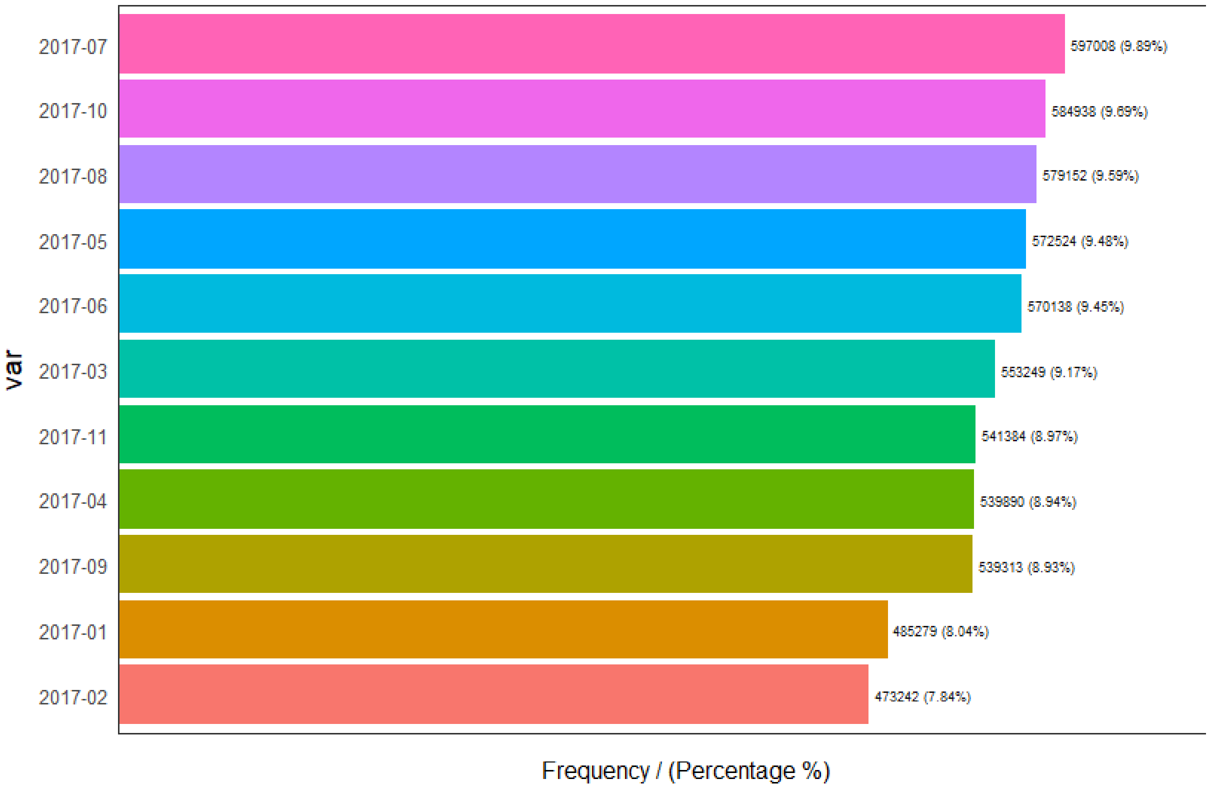

Figure 4 Monthly reported number of crimes in the UK

July 2017 was the busiest month for the police in terms of criminal offences with 597,008 crimes. The least busy month was February with 473,242. This could be due to the fact that February is the shortest month of the year. Looking at our data, it is fair to say that in the warmest months more crimes are committed than the coldest months.

The chart below shows a pattern in terms of the numbers of crimes being committed each month. The numbers rise one month, then decrease the next and then rise again in a very constant fashion, apart from the month of September. This seems to create a subtle prediction in a way that tell us if the police will be busy looking at past and present data.

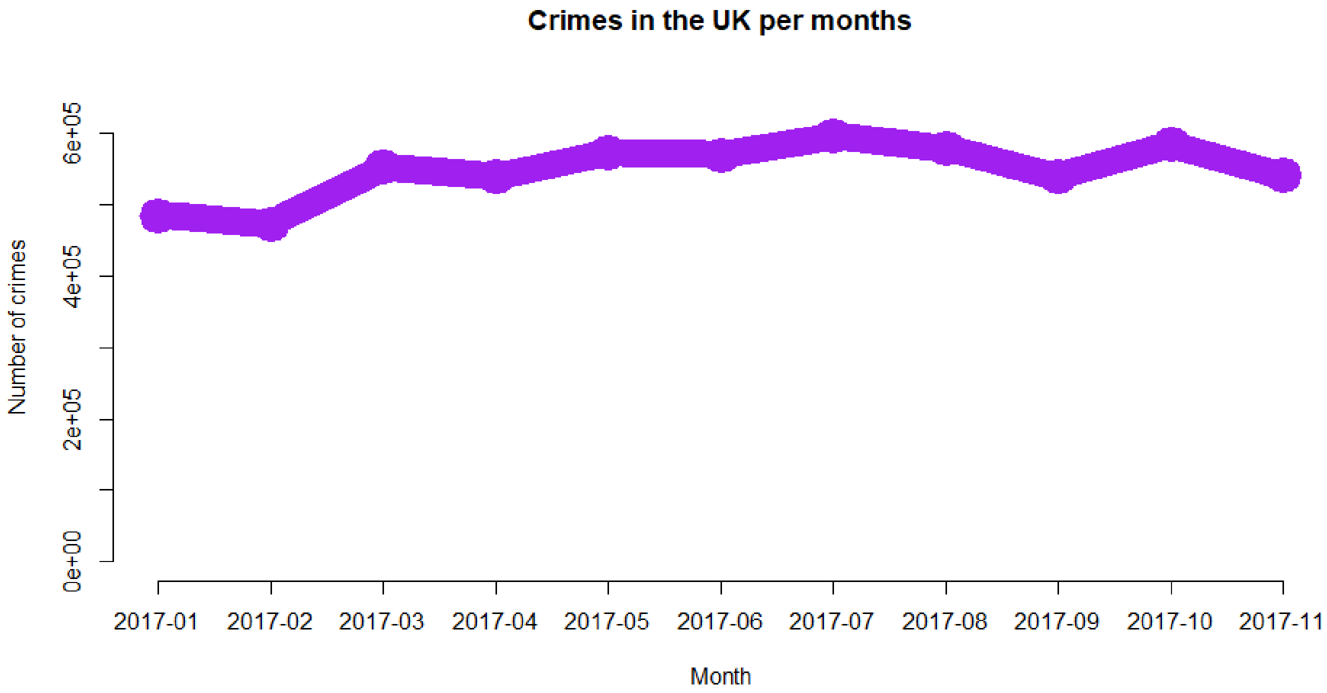

Figure 5 Evolution of the number of crimes overtime in the UK

How are criminal offences dealt with by the police forces?

In 30 percent of the cases, the police would not identify a suspect after completing the investigation. This is a huge percentage and could highlight a problem with the effectiveness of certain police operations as well as resource allocation. The number of police officers and the funding for the police both fell by 14% between 2010 and 2015 (Disney & Simpson, 2017). UK Police have claimed that some missing data may be due to British Transport Police and Northern Ireland not being particularly required to provide their data, (Home Office UK, 2017). However, there is a worrying issue in that we do not know what happen to 25% of the criminal cases. There are less criminal offenders being prosecuted than those being charged or ordered to pay compensation. The following chart gives a comprehensible breakdown of the outcomes after the UK police have responded to criminal incident.

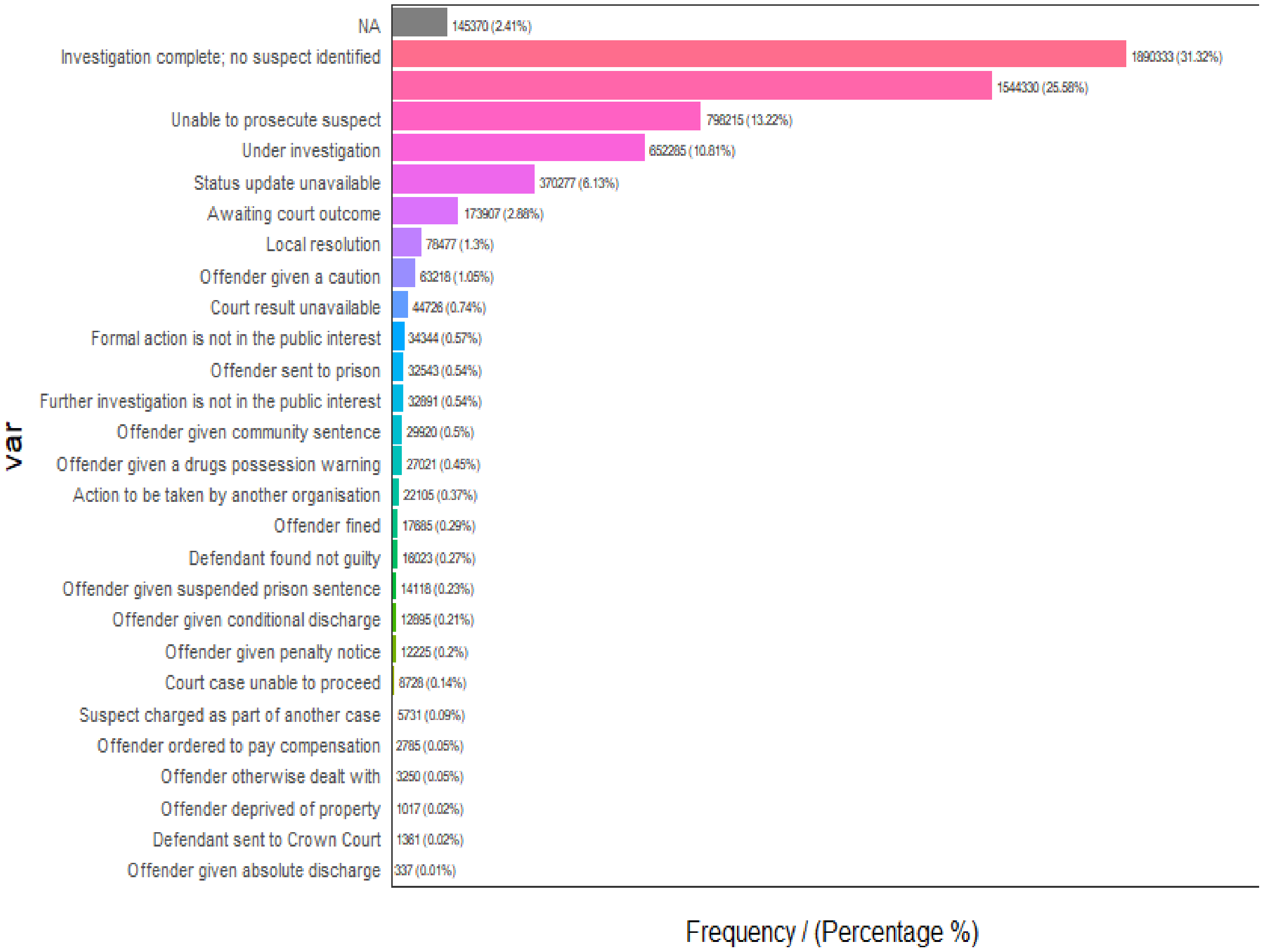

Figure 6 Outcome categories and their numbers after criminal offences

Conclusion

We have been able to aggregate an impressive amount of data from the records made available by the Home Office via data.police.uk. In total we processed 6,036,117 observations which have enable us to get better oversight of the different patterns pertaining to crimes occurrences in the UK as well as how they are recorded.

Furthermore, our Analysis of the data has identified potential sources of problems with regards to the above-mentioned patterns. Regions around largest cities and warmer months attracted more criminal activities. There is also a poor recording system of the outcome as illustrated by the high volumes of missing data.

Finally, we hope the above analysis help improve decision making at both operational and strategic level so, the police may fight better to reduce crime rates in the UK.

References

Casas, P. (2017). *Data Science Live Book*. Retrieved from https://livebook.datascienceheroes.com/

Disney, R., & Simpson, P. (2017). Police workforce and funding in England and Wales IFS Briefing Note BN208. *The Institute for Fiscal Studies*. Retrieved from https://www.ifs.org.uk/uploads/publications/bns/bn208.pdf

Home Office UK. (2017). data.police.uk. Retrieved January 21, 2018, from https://data.police.uk/about/

Office for National Statistics. (2017). Crime in England and Wales - Office for National Statistics. Retrieved January 21, 2018, from https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/june2017

Appendices

Appendix 1 – RStudio script

Let's start by collecting and collating our data.

Our aim is to have an overview of the crimes committed at national level, namely in England, Wales and Northern Ireland.

in a one-year period. When we downloaded the data from the data.police.uk, we received a zip folder of all the information.

#Interestingly and painfully at the same time, we discovered that, there were spreadsheets for each police force

and for each month from January to November. That's roughly 483 spreadsheets since there was no

information for a police force in November.

First, we set the working directory for the month we need:

setwd("C:/Users/romdj/OneDrive/Doing Data Science/Intro to Data Science/INFO627.Rproj/UK Crimes - 17/UK Crimes - Jan 17")

We then import all the spreadsheets which have .csv extensions (all the 44 police forces)

files = list.files(pattern="*.csv")

data_list = lapply(files, read.csv, header = TRUE)

We now join up all the spreadsheets for all the police forces for the month.

uk_crime_jan_17 <- do.call(rbind, data_list)

#Let's check everything is correct by looking at the month column in our data set

head(uk_crime_jan_17$Month)

Let's do the same for all the months for which we obtained data, i.e January to November 2017.

```
setwd("C:/Users/romdj/OneDrive/Doing Data Science/Intro to Data Science/INFO627.Rproj/UK Crimes - 17/UK Crimes - Feb 17")

files = list.files(pattern="*.csv")

data_list = lapply(files, read.csv, header = TRUE)

uk_crime_feb_17 <- do.call(rbind, data_list)

head(uk_crime_feb_17$Month)


setwd("C:/Users/romdj/OneDrive/Doing Data Science/Intro to Data Science/INFO627.Rproj/UK Crimes - 17/UK Crimes - Mar 17")

files = list.files(pattern="*.csv")

data_list = lapply(files, read.csv, header = TRUE)

uk_crime_mar_17 <- do.call(rbind, data_list)

head(uk_crime_mar_17$Month)



setwd("C:/Users/romdj/OneDrive/Doing Data Science/Intro to Data Science/INFO627.Rproj/UK Crimes - 17/UK Crimes - Apr 17")

files = list.files(pattern="*.csv")

data_list = lapply(files, read.csv, header = TRUE)

uk_crime_apr_17 <- do.call(rbind, data_list)

head(uk_crime_apr_17$Month)


setwd("C:/Users/romdj/OneDrive/Doing Data Science/Intro to Data Science/INFO627.Rproj/UK Crimes - 17/UK Crimes - May 17")

files = list.files(pattern="*.csv")

data_list = lapply(files, read.csv, header = TRUE)

uk_crime_may_17 <- do.call(rbind, data_list)

head(uk_crime_may_17$Month)


setwd("C:/Users/romdj/OneDrive/Doing Data Science/Intro to Data Science/INFO627.Rproj/UK Crimes - 17/UK Crimes - Jun 17")
```

```
files = list.files(pattern="*.csv")

data_list = lapply(files, read.csv, header = TRUE)

uk_crime_jun_17 <- do.call(rbind, data_list)

head(uk_crime_jun_17$Month)


setwd("C:/Users/romdj/OneDrive/Doing Data Science/Intro to Data Science/INFO627.Rproj/UK Crimes - 17/UK Crimes - Jul 17")

files = list.files(pattern="*.csv")

data_list = lapply(files, read.csv, header = TRUE)

uk_crime_jul_17 <- do.call(rbind, data_list)

head(uk_crime_jul_17$Month)


setwd("C:/Users/romdj/OneDrive/Doing Data Science/Intro to Data Science/INFO627.Rproj/UK Crimes - 17/UK Crimes - Aug 17")

files = list.files(pattern="*.csv")

data_list = lapply(files, read.csv, header = TRUE)

uk_crime_aug_17 <- do.call(rbind, data_list)

head(uk_crime_aug_17$Month)


setwd("C:/Users/romdj/OneDrive/Doing Data Science/Intro to Data Science/INFO627.Rproj/UK Crimes - 17/UK Crimes - Sep 17")

files = list.files(pattern="*.csv")

data_list = lapply(files, read.csv, header = TRUE)

uk_crime_sep_17 <- do.call(rbind, data_list)

head(uk_crime_sep_17$Month)


setwd("C:/Users/romdj/OneDrive/Doing Data Science/Intro to Data Science/INFO627.Rproj/UK Crimes - 17/UK Crimes - Oct 17")

files = list.files(pattern="*.csv")

data_list = lapply(files, read.csv, header = TRUE)

uk_crime_oct_17 <- do.call(rbind, data_list)

head(uk_crime_oct_17$Month)
```

```
setwd("C:/Users/romdj/OneDrive/Doing Data Science/Intro to Data Science/INFO627.Rproj/UK Crimes - 17/UK Crimes - Nov 17")

files = list.files(pattern="*.csv")

data_list = lapply(files, read.csv, header = TRUE)

uk_crime_nov_17 <- do.call(rbind, data_list)

head(uk_crime_nov_17$Month)


# We now have all the spreadsheets of the police forces, all joined up according to the month.

#Now let's finalise our super dataset by combining all the months' worth of data.


uk_crime_2017 <- do.call(rbind, list(uk_crime_jan_17, uk_crime_feb_17, uk_crime_mar_17, uk_crime_apr_17, uk_crime_may_17, uk_crime_jun_17, uk_crime_jul_17, uk_crime_aug_17, uk_crime_sep_17, uk_crime_oct_17, uk_crime_nov_17))


#That went well, we now have our super data of 483 files called uk_crime_2017

# In order not to lose our new creation when we need it, we will export the data set in a chosen directory and folder.


setwd("C:/Users/romdj/OneDrive/Doing Data Science/Intro to Data Science/INFO627.Rproj")


write.csv(uk_crime_2017, "uk_crime_2017.csv")


# To avoid any fatal errors when working on the data set in R, we will copy it to a new dataset which we will name

# crime_2017.


crime_2017 <- uk_crime_2017


# After all the preparations, it is time to process the structure of our dataset.

# The following function give us a feel of the variables, we will deal with in our dataset.


colnames(crime_2017)
```

The str function gives us a snapshot of what is in our dataset, the types of data in the variables and their levels.

str(crime_2017)

My favourite package for data exploration: funModelling. Calling funModelling will also bring the

other packages such as Hmisc,lattice, survival, Formula and ggplot2.

library(funModeling)

The decribe function helps with the categorical and numerical profiling of the dataset

describe(crime_2017)

An even more interesting way of profiling the data structure. df_status allows me to view the quantity

and percentage of zeros and missing data for all the variables. This helps me make further decision on

data processing such as which column to eliminate, whether I should remove the missing data or replace them.

df_status(crime_2017)

As mentioned above, I feel there is no point keeping the last columns since all the information is missing.

crime_2017 <- crime_2017[, -12]

I am happy keeping all the other columns for now as you never know. I also need to investigate further the

number of NAs for the outcome categories 2.4 of 6 million observations is still huge for such an important variable.

Now we are ready for more discovery and summary of the data. Let's look at the frequency tables and chart

for our important variables. Let ‘start with the total number of crimes reported by each police force.

```
freq(crime_2017$Falls.within)
```

How about looking at the distribution of the crimes on a map? First I need to remove the missing data from the

Longitude and Latitude columns. I then transform the data into spatial data points to make it easier to plot.

```
crime_2017_map <- na.omit(crime_2017)

crime_2017_coords <- cbind(Longitude = as.numeric(as.character(crime_2017_map$Longitude)), latitude = as.numeric(as.character(crime_2017_map$Latitude)))

library("sp")

crime_2017.pts <- SpatialPointsDataFrame(crime_2017_coords, crime_2017_map[ , -(5:6)], proj4string = CRS("+init=epsg:4326"))

plot(crime_2017.pts, pch = ".", col = "red")
```

Looking at the types of crimes committed

```
freq(crime_2017$Crime.type)
```

Looking at the number of crimes per months

```
freq(crime_2017$Month)
```

Number of crimes overtime

```
monthly <- table(crime_2017$Month)
```

```
plot(monthly, col = "purple", main = "Crimes in the UK per months", lwd = 20, type = "b", ylim = c(0, max(monthly)*1.1), ylab = "Number of crimes", xlab = "Month")

# Looking at how the offences are dealt with how the information is recorded

freq(crime_2017$Last.outcome.category)
```

Appendix 2 - CSV Columns

The columns in the CSV files are as follows:

Reported by - The force that provided the data about the crime.

Falls within - At present, also the force that provided the data about the crime. This is currently being looked into and is likely to change in the near future.

Longitude and Latitude - The anonymised coordinates of the crime. See Location Anonymisation for more information.

LSOA code and LSOA name - References to the Lower Layer Super Output Area that the anonymised point falls into, according to the LSOA boundaries provided by the Office for National Statistics.

Crime type - One of the crime types listed in the Police.UK FAQ.

Last outcome category - A reference to whichever of the outcomes associated with the crime occurred most recently. For example, this crime's 'Last outcome category' would be 'Formal action is not in the public interest'.

Context - A field provided for forces to provide additional human-readable data about individual crimes. Currently, for newly added CSVs, this is always empty.

Source: (Home Office UK, 2017)

Appendix 5 – Crime Type

All crime

Total for all categories.

Anti-social behaviour

Includes personal, environmental and nuisance anti-social behaviour.

Bicycle theft

Includes the taking without consent or theft of a pedal cycle.

Burglary

Includes offences where a person enters a house or other building with the intention of stealing.

Criminal damage and arson

Includes damage to buildings and vehicles and deliberate damage by fire.

Drugs

Includes offences related to possession, supply and production.

Other crime

Includes forgery, perjury and other miscellaneous crime.

Other theft

Includes theft by an employee, blackmail and making off without payment.

Possession of weapons

Includes possession of a weapon, such as a firearm or knife.

Public disorder and weapons

Includes offences which cause fear, alarm, distress or a possession of a weapon such as a firearm.

Public order

Includes offences which cause fear, alarm or distress.

Robbery

Includes offences where a person uses force or threat of force to steal.

Shoplifting

Includes theft from shops or stalls.

Theft from the person

Includes crimes that involve theft directly from the victim (including handbag, wallet, cash, mobile phones) but without the use or threat of physical force.

Vehicle crime

Includes theft from or of a vehicle or interference with a vehicle.

Violence and sexual offences

Includes offences against the person such as common assaults, Grievous Bodily Harm and sexual offences.

Source: (Home Office UK, 2017)

Appendix 4 – Data Exploration tables

```
> df_status(crime_2017)
                variable q_zeros p_zeros    q_na   p_na q_inf p_inf    type  unique
1               Crime.ID       0       0       0   0.00     0     0  factor 4314001
2                  Month       0       0       0   0.00     0     0  factor      11
3            Reported.by       0       0       0   0.00     0     0  factor      44
4           Falls.within       0       0       0   0.00     0     0  factor      44
5              Longitude       0       0  102705   1.70     0     0 numeric  603155
6               Latitude       0       0  102705   1.70     0     0 numeric  585866
7               Location       0       0       0   0.00     0     0  factor  237546
8              LSOA.code       0       0  145370   2.41     0     0  factor   34750
9              LSOA.name       0       0  145370   2.41     0     0  factor   34750
10            Crime.type       0       0       0   0.00     0     0  factor      14
11 Last.outcome.category       0       0  145370   2.41     0     0  factor      27
12               Context       0       0 6036117 100.00     0     0 logical       0
```

```
       var frequency percentage cumulative_perc
1  2017-07    597008       9.89            9.89
2  2017-10    584938       9.69           19.58
3  2017-08    579152       9.59           29.17
4  2017-05    572524       9.48           38.65
5  2017-06    570138       9.45           48.10
6  2017-03    553249       9.17           57.27
7  2017-11    541384       8.97           66.24
8  2017-04    539890       8.94           75.18
9  2017-09    539313       8.93           84.11
10 freq(data, str_input = NA, plot = TRUE, na.rm = FALSE, path_out)   .15
11 2017-02    473242       7.84          100.00
```

```
> freq(crime_2017$Falls.within)
                                  var frequency percentage cumulative_perc
1          Metropolitan Police Service    960836      15.92           15.92
2            Greater Manchester Police    380661       6.31           22.23
3                West Yorkshire Police    294282       4.88           27.11
4                 West Midlands Police    262817       4.35           31.46
5                South Yorkshire Police    194582       3.22           34.68
6                   Northumbria Police    193304       3.20           37.88
7                          Kent Police    192628       3.19           41.07
8              Hampshire Constabulary    190336       3.15           44.22
9             Lancashire Constabulary    177786       2.95           47.17
10                Thames Valley Police    171920       2.85           50.02
11      Avon and Somerset Constabulary    169818       2.81           52.83
12                        Essex Police    167843       2.78           55.61
13                   Merseyside Police    158208       2.62           58.23
14 Police Service of Northern Ireland    145370       2.41           60.64
15                       Sussex Police    139997       2.32           62.96
16             Devon & Cornwall Police    133977       2.22           65.18
17                  South Wales Police    132541       2.20           67.38
18                  West Mercia Police    119635       1.98           69.36
19              Nottinghamshire Police    117851       1.95           71.31
20          Hertfordshire Constabulary    109853       1.82           73.13
21                Staffordshire Police    109250       1.81           74.94
22              Cheshire Constabulary     96724       1.60           76.54
23                   Humberside Police     92369       1.53           78.07
24                    Cleveland Police     90711       1.50           79.57
25               Leicestershire Police     89190       1.48           81.05
26             Derbyshire Constabulary     88925       1.47           82.52
27                       Surrey Police     88835       1.47           83.99
28          Cambridgeshire Constabulary     80768       1.34           85.33
29             Northamptonshire Police     74989       1.24           86.57
30                Durham Constabulary     73765       1.22           87.79
31                       Dorset Police     70583       1.17           88.96
32                Norfolk Constabulary     69375       1.15           90.11
33               North Yorkshire Police     62890       1.04           91.15
34                  Bedfordshire Police     61013       1.01           92.16
35                  North Wales Police     58924       0.98           93.14
36                Suffolk Constabulary     58623       0.97           94.11
37                        Gwent Police     57379       0.95           95.06
38                    Wiltshire Police     56383       0.93           95.99
39                  Lincolnshire Police     54745       0.91           96.90
40        Gloucestershire Constabulary     54048       0.90           97.80
41                  Warwickshire Police     53573       0.89           98.69
42                   Dyfed-Powys Police     37772       0.63           99.32
43                Cumbria Constabulary     35809       0.59           99.91
44              City of London Police      5229       0.09          100.00
```

```
> str(crime_2017)
'data.frame':   6036117 obs. of  11 variables:
 $ Crime.ID             : Factor w/ 4314001 levels "","0002ecf6cc11667fa485d7f3cc174cfdde1b8552732f933d2fa8efbf92f8a1e2",..: 9971 6436 199 6158 1413 6467 3293 4116 8069 7751 ...
 $ Month                : Factor w/ 11 levels "2017-01","2017-02",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Reported.by          : Factor w/ 44 levels "Avon and Somerset Constabulary",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Falls.within         : Factor w/ 44 levels "Avon and Somerset Constabulary",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Longitude            : num  -2.49 -2.51 -2.5 -2.51 -2.52 ...
 $ Latitude             : num  51.4 51.4 51.4 51.4 51.4 ...
 $ Location             : Factor w/ 237546 levels "No Location",..: 941 753 1053 246 3555 2549 246 941 683 3555 ...
 $ LSOA.code            : Factor w/ 34750 levels "","E01004509",..: 36 36 36 36 36 36 36 36 36 36 ...
 $ LSOA.name            : Factor w/ 34750 levels "","Bath and North East Somerset 001A",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ Crime.type           : Factor w/ 14 levels "Anti-social behaviour",..: 3 3 4 4 4 4 4 13 13 14 ...
 $ Last.outcome.category: Factor w/ 27 levels "","Action to be taken by another organisation",..: 23 14 23 10 23 23 11 23 23 10 ...
```

```
> freq(crime_2017$Crime.type)
                         var frequency percentage cumulative_perc
1      Anti-social behaviour   1602332      26.55           26.55
2  Violence and sexual offences   1372163      22.73           49.28
3     Criminal damage and arson    559037       9.26           58.54
4                 Other theft    493899       8.18           66.72
5               Vehicle crime    421594       6.98           73.70
6                    Burglary    402951       6.68           80.38
7                 Shoplifting    359491       5.96           86.34
8                Public order    331444       5.49           91.83
9                       Drugs    124391       2.06           93.89
10              Bicycle theft     92196       1.53           95.42
11      Theft from the person     87081       1.44           96.86
12                Other crime     85440       1.42           98.28
13                    Robbery     67419       1.12           99.40
14      Possession of weapons     36679       0.61          100.00
```

```
> freq(crime_2017$Last.outcome.category)
                                                    var frequency percentag
e cumulative_perc
1        Investigation complete; no suspect identified   1890333      31.3
2           31.32
2                                                        1544330      25.5
8           56.90
3                             Unable to prosecute suspect    798215      13.2
2           70.12
4                                    Under investigation    652285      10.8
1           80.93
5                              Status update unavailable    370277       6.1
3           87.06
6                                 Awaiting court outcome    173907       2.8
8           89.94
7                                                   <NA>    145370       2.4
1           92.35
8                                       Local resolution     78477       1.3
0           93.65
9                             Offender given a caution     63218       1.0
5           94.70
10                            Court result unavailable     44726       0.7
4           95.44
11        Formal action is not in the public interest     34344       0.5
7           96.01
12 Further investigation is not in the public interest     32891       0.5
4           96.55
13                             Offender sent to prison     32543       0.5
4           97.09
14                  Offender given community sentence     29920       0.5
0           97.59
15          Offender given a drugs possession warning     27021       0.4
5           98.04
16           Action to be taken by another organisation     22105       0.3
7           98.41
17                                     Offender fined     17685       0.2
9           98.70
18                         Defendant found not guilty     16023       0.2
7           98.97
19            Offender given suspended prison sentence     14118       0.2
3           99.20
20               Offender given conditional discharge     12895       0.2
1           99.41
21                       Offender given penalty notice     12225       0.2
0           99.61
22                         Court case unable to proceed      8728       0.1
4           99.75
23            Suspect charged as part of another case      5731       0.0
9           99.84
24                        Offender otherwise dealt with      3250       0.0
5           99.89
25               Offender ordered to pay compensation      2785       0.0
5           99.94
26                         Defendant sent to Crown Court      1361       0.0
2           99.96
27                        Offender deprived of property      1017       0.0
2           99.98
28                  Offender given absolute discharge       337       0.0
1          100.00
```

```
> describe(crime_2017)
crime_2017

 12  Variables      6036117  Observations
--------------------------------------------------------------------------------
----------------------------------------------------------------
Crime.ID
       n  missing distinct
 6036117        0  4314001

lowest :
0002ecf6cc11667fa485d7f3cc174cfdde1b8552732f933d2fa8efbf92f8a1e2 000942789
42087b7aba920da214aaec5da05bc68e5a4eeabc21af766c912b879 0017b6b8d7f5d9760f
d4e060b8c6e92f91c908e61cefa67714d3c6da2436b0e2 001a8fbf9f69b6960b3161edf1a
ddc1825a1edfad52552fde30b0829eb2ca289
highest: ff689141ca115ff10467df2674eeec18873c8a919e6468353b43942059a7f2d3
ff833efe253cfb0851fa82f22bfd6be6aea53753d56eeab1ccd1f4e2287d2cb4 ffacbc79c
a23ca95c8ed83d4e765d071d8fb14e9d56d5c880619c28085c7dc63 ffe14fa9adb57b5201
8787b2b7c28e6825815eef0138b5561c408bc94547684e fff661aaf72781e612965e959a6
041e1c1897c6b67ba42fd90275a2253ece744
--------------------------------------------------------------------------------
----------------------------------------------------------------
Month
       n  missing distinct
 6036117        0       11

Value      2017-01 2017-02 2017-03 2017-04 2017-05 2017-06 2017-07 2017-08
2017-09 2017-10 2017-11
Frequency   485279  473242  553249  539890  572524  570138  597008  579152
539313  584938  541384
Proportion   0.080   0.078   0.092   0.089   0.095   0.094   0.099   0.096
0.089   0.097   0.090
--------------------------------------------------------------------------------
----------------------------------------------------------------
Reported.by
       n  missing distinct
 6036117        0       44 |

lowest : Avon and Somerset Constabulary Bedfordshire Police          Cam
bridgeshire Constabulary    Cheshire Constabulary          City of London
Police
highest: Warwickshire Police            West Mercia Police             Wes
t Midlands Police          West Yorkshire Police          Wiltshire Polic
e
--------------------------------------------------------------------------------
----------------------------------------------------------------
Falls.within
       n  missing distinct
 6036117        0       44

lowest : Avon and Somerset Constabulary Bedfordshire Police          Cam
bridgeshire Constabulary    Cheshire Constabulary          City of London
Police
highest: Warwickshire Police            West Mercia Police             Wes
t Midlands Police          West Yorkshire Police          Wiltshire Polic
e
--------------------------------------------------------------------------------
----------------------------------------------------------------
Longitude
       n  missing distinct     Info     Mean      Gmd      .05      .10
.25      .50      .75      .90      .95
```

```
 5933412   102705   603155        1    -1.42    1.585  -3.8163  -3.0337  -
2.2295  -1.4040  -0.2177   0.1153   0.5520

lowest : -8.148313 -8.107121 -8.106383 -8.099225 -8.095469, highest:  1.75
7268  1.757344  1.757530  1.758674  1.759519
--------------------------------------------------------------------------------
----------------------------------------------------------------
Latitude
       n  missing distinct     Info     Mean      Gmd      .05      .10
.25      .50      .75      .90      .95
 5933412   102705   585866        1    52.49    1.346    50.83    51.21
51.51    52.38    53.48    53.99    54.69

lowest : 49.91306 49.91315 49.91322 49.91385 49.91392, highest: 55.78407 5
5.78410 55.78813 55.78824 55.78962
--------------------------------------------------------------------------------
----------------------------------------------------------------
Location
       n  missing distinct
 6036117        0   237546

lowest : No Location                  On or near A30              On or nea
r A303              On or near A357              On or near A358
highest: On or near Wellworthy Drive On or near Whitesbury Road  On or nea
r Windflower Road  On or near Witts Lane       On or near Zander Road
--------------------------------------------------------------------------------
----------------------------------------------------------------
LSOA.code
       n  missing distinct
 5890747   145370    34750

lowest :           E01004509 E01007495 E01007602 E01008068, highest: E0101
3388 E01027518 E01030195 E01021188 E01028602
--------------------------------------------------------------------------------
----------------------------------------------------------------
LSOA.name
       n  missing distinct
 5890747   145370    34750

lowest :                                 Bath and North East Somerset 00
1A Bath and North East Somerset 001B Bath and North East Somerset 001C Bat
h and North East Somerset 002A
highest: York 008E                       Northumberland 020C
Suffolk Coastal 005C            Wealden 014A                         South
Oxfordshire 011C
--------------------------------------------------------------------------------
----------------------------------------------------------------
Crime.type
       n  missing distinct
 6036117        0       14

lowest : Anti-social behaviour        Bicycle theft                 Burglar
y                     Criminal damage and arson    Drugs
highest: Robbery                      Shoplifting                   Theft f
rom the person        Vehicle crime                 Violence and sexual off
ences
--------------------------------------------------------------------------------
----------------------------------------------------------------
Last.outcome.category
       n  missing distinct
 5890747   145370       27
```