

## **Crimes at Garda Station Level 2003-2015.**

Visualisation of crime uses the data file **Crimes at Garda Station Level** which contains nine reported crime categories for each of the 563 Garda Stations in the Republic of Ireland from 2003 to 2015. The data set is published by All-Island Research Laboratory (AIRO) based on data obtained from Central Statistics Office (CSO) and is hosted on the [data.gov.ie](http://data.gov.ie) site under the title *Crimes at Garda Stations Level 2010-15*<sup>1</sup>.

### **Additional Data Sets**

CSO boundary files for Garda stations

AIRO population estimates for each Garda division

### **Format**

The data file contains 563 rows (excluding the header) and 131 columns. Each row contains the following information for each Garda station - id, Station Name, Division, x co-ordinate, y co-ordinate and the number of recorded crimes each year from 2003 to 2015 for each of the twelve crime categories (apart from one crime category which is missing for 2003). The data dictionary associated with this file is shown in Appendix 1.

### **Reason for selecting Crimes at Garda Station Level 2010-15**

The reason Crimes at Garda Station Level was chosen is because of the richness of the data set. The variables include 573 Garda Stations with their associated easting and northing geo-coordinates, 24 divisions, 13 years of data for each of 12 crime categories. This offered a range of options for analysis and visualisation including the use of polygon maps using ESRI shape files, hotspot maps and the use of trellis or facet plots for visualising categorical variables. The availability of data over time offered the opportunity to showcase some graphical ideas that are perhaps not that commonly used like Statistical Process Control (SPC) Charts which can be used as a management tool to assess trends in crime. Finally the availability of 12 crime categories meant that more advanced analytical methods could be employed to segment the country into areas which have similar crime patterns. An example using Principal Component Analysis (PCA) is presented in the progress report.

### **Data Preparation**

The data was in wide format meaning that each of the twelve crimes for each of the 13 years were placed in a total of 117 columns. To facilitate greater flexibility in analysing and visualising this data set the file was transformed to long format. This involved reducing the 117 columns containing the year and crime variables to just two columns namely year and crime category. The transformed data set has dimensions of 87,828 rows and six columns. In this format each row contains the number of reported crimes for a specific year for a specific crime for each of the 563 Garda stations.

The locational data was provided in two fields labelled x and y representing the easting and northing of the Irish Grid co-ordinate system. These easting and northing co-ordinates had to be transformed to longitude and latitude (or ETRF89 GPS Co-ordinates) for mapping the crime data. This was achieved by downloading the Ordnance Survey of Ireland (OSI) application Grid InQuest II. Importing the easting and northing data using this application returned the corresponding longitude and latitude for each station.

Quality control checks were performed on the data which revealed a small number of errors and omissions. The file contained no data for the year 2003 for one crime category which was *Attempts or threats to murder, assaults etc.*

---

<sup>1</sup> contains data from 2003-2015  
Crime at Garda Station Level

The division variable had two labels for Westmeath called *Westmeath* and *Westmeath Division* with 186 and 1,872 cases, respectively.

The data file in [data.gov.ie](http://data.gov.ie) does not include exposure data to allow crime numbers to be adjusted by say population size. However population in each division was sourced from a crime data file created by AIRO and used by the Cork Examiner newspaper [http://www.irishexaminer.com/crime/crime\\_data/](http://www.irishexaminer.com/crime/crime_data/). Population estimates at Garda station level were located in the CSO census 2011 Boundary files repository located at <http://www.cso.ie/en/census/census2011boundaryfiles/>. CSO has boundary files which contain population data for Garda Regions, Garda Divisions, Garda Districts and Garda Sub Districts (station).

The CSO boundary files for Garda stations have two variables for Garda station name - one in English and the other in Irish. However, the [data.gov.ie](http://data.gov.ie) file contains one Garda station name variable which includes both Irish and English language station names. Linking the [data.gov.ie](http://data.gov.ie) with the CSO English language Garda station means that a small number of Irish Garda station names in the [data.gov.ie](http://data.gov.ie) will not be mapped.

### **Software Used**

The data analysis involved the use of the visualisation software Tableau ([www.Tableau.com](http://www.Tableau.com)) used to create interactive visualisation and the statistical exploratory and visualisation software DataDesk ([www.DataDesk.com](http://www.DataDesk.com)) which was used for data preparation and statistical multivariate analysis.

### **Outcome and Impact**

The outcome is the creation of interactive graphical visualisations down to the level of Garda station which can allow for a fuller understanding of crime in Ireland to be obtained.

The combination of visual and statistical concepts could be of assistance to the general public but also for Garda management/regulatory authorities. Most of the visualisations allow for dynamic interaction with the crime data.

### **Timeline for Completion**

The project is effectively completed. Outstanding work is mainly associated with using a standard Garda station name list, some data checks and combining the visualisations into one file with a worksheet for each graphic.

## Progress Report

The project involved the development of a number of interactive visualisations that allows users to gain a comprehensive insight into Garda crime between 2003 and 2015. This progress report highlights six interactive and dynamic visualisations that have been created using the data set **Crimes at Garda Station Level**

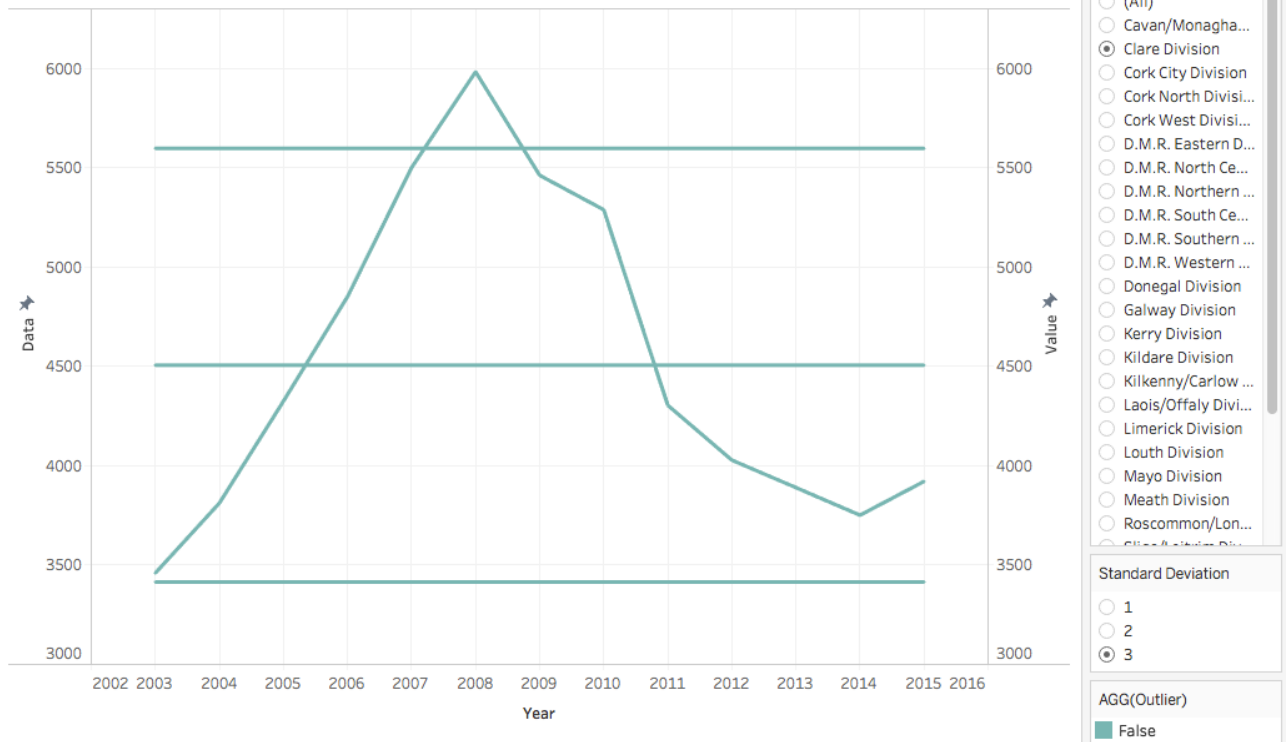
### 1. **Visualisation of Crime Data using an Interactive Statistical Process Control (SPC) Chart**

Statistical Process Control (SPC) has its origins in manufacturing industry where it was first introduced by Walter Shewhart in the 1920s when working for the Bell Telephone company. He outlined the procedure in his classic text *The Economic Control of Manufactured Product* published in 1926 and the most popular SPC charts - Shewhart charts - are named after him. In recent decades SPC has spread from manufacturing industry to many non-manufacturing sectors as we can adapt the production line analogy to a range of sectors which involves the monitoring of some characteristic over time. In this example the number of reported crimes are plotted over time for each division. The objective is to establish if crime is increasing or decreasing. This simply put is determined by assessing if the number of crimes exceed the upper and/or lower limits on the charts. The SPC chart used to visualise Crime data is known as a Mean chart.

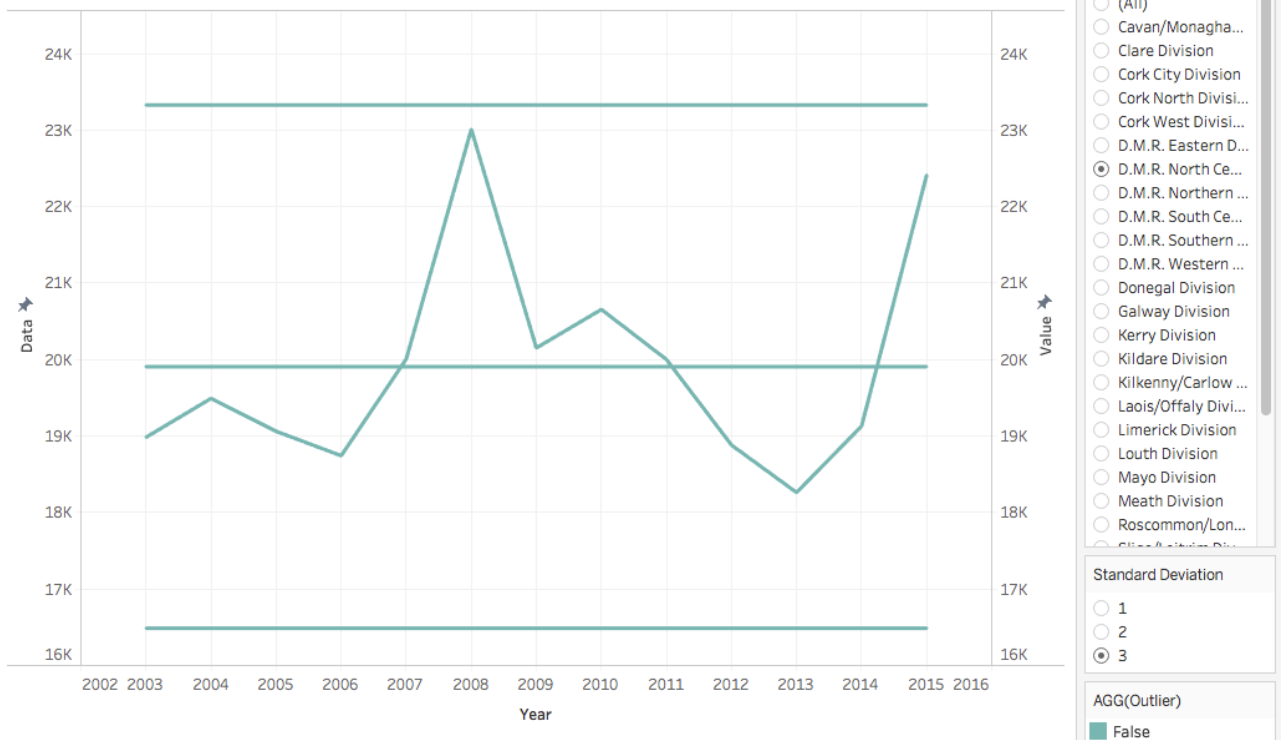
The charts show a similar picture for most divisions with crime increasing from 2003 to reach a peak at around 2007-08 and decreasing to 2015. An example of the SPC chart computed for Clare Division is shown in Figure 1.1 (Top). The exception to this pattern are the two Garda divisions in Dublin inner city - DMR South Central and DMR North Central which report a more stable crime trend over the 13 years. The bottom chart in Figure 1.1 is for DMR North Central. Incidentally, these two inner city divisions report the largest number of crimes and statistical multivariate analysis, which will be introduced at the end of this report, show that these two divisions form a cluster.

As with all statistical studies health warnings come in to play - population could be increasing/decreasing, Garda recording of crime may vary while reporting levels also fluctuate. In addition, it should be noted that SPC charts have a number of assumptions which might not be met in this application like independence of observations, but the intention is to showcase the concept of using charts of this kind when data collected over time is available.

Statistical Process Control Chart



Statistical Process Control Chart



**Figure 1.1:** SPC plot of total annual crime reported to An Garda Siochana between 2003 and 2015 for Clare Division (Top) and Dublin Metropolitan Region North Central (Bottom)

## 2. Visualisation of Garda Crime using Trellis Plots

The format of the data set **Crimes at Garda Station Level** with Garda division and crime both discrete variables allows a range of graphical visualisations to be deployed. One important graphic is a trellis or facet plot of crime data which allows patterns over time to be visualised by crime category for each Garda division. For example, *Controlled Drug Offences* increased from 2003 reaching a maximum in 2008/09 and declined thereafter. *Damage to Property* has generally been declining over the last decade across most divisions. The category *Burglary*, with the exception of perhaps DMR Western, has remained fairly stable over the time period. Other interesting observations are the spike in *Kidnapping and Related Offences* in 2009 for DMR Northern and the substantial increase *Offences against government, justice procedures and organisations* in DMR North Central.

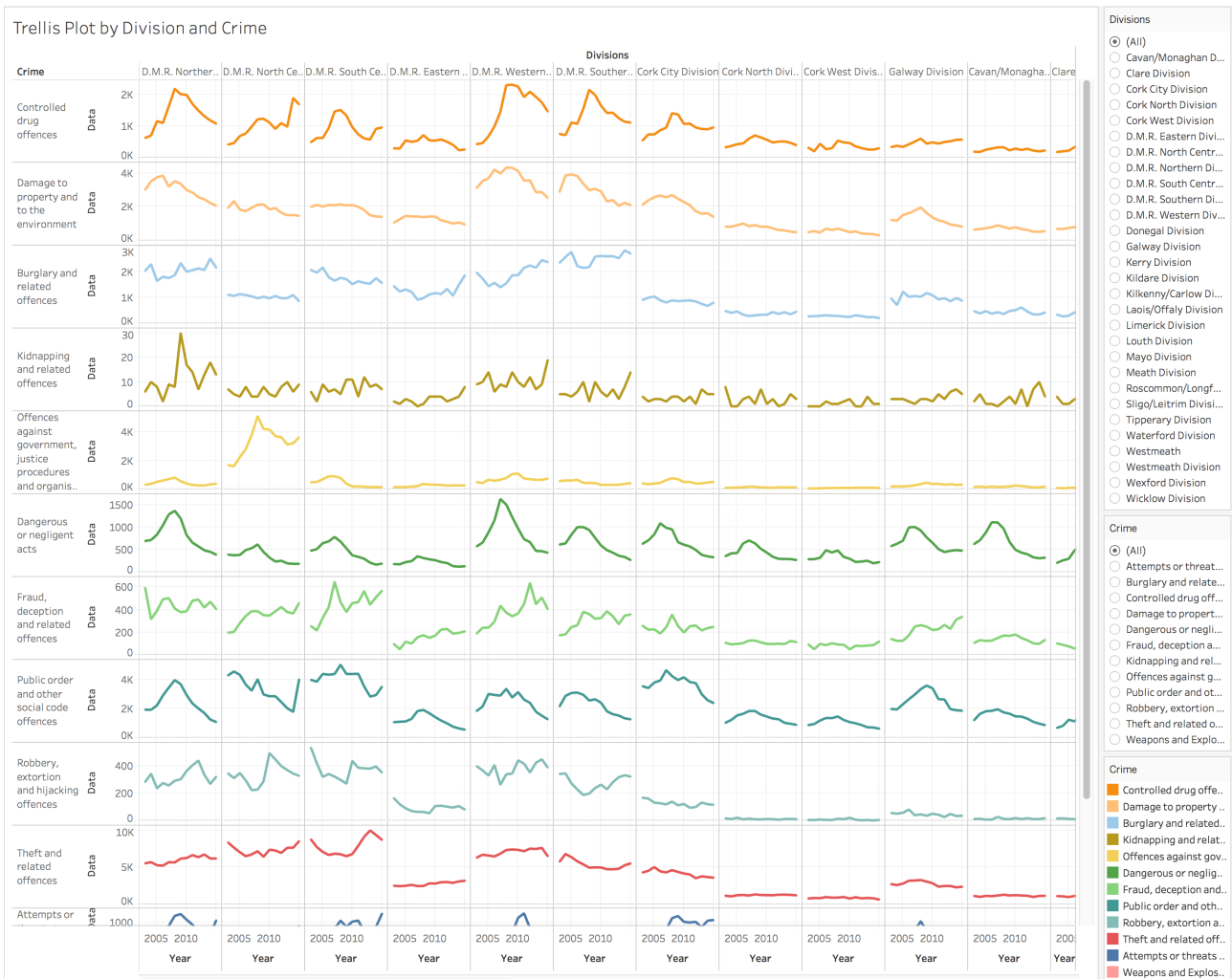


Figure 1.2. Extract of Trellis Plot for Crime Category classified by Garda Division

The number of crimes in this category increased from just over 1,600 in 2003 to over 5,000 offences recorded in 2008. DMR North Central stands out for overall levels of this offence over the 13 year period. One of the early pioneers of data visualisation Jacques Bertin was a strong advocate of the importance of taking sufficient time to examine visualisations and allowing the patterns to unfold. Trellis plots are one such example of the appropriateness of Bertin's viewpoint.

### 3. Highlight Table

The highlight tables shown in Figure 1.3 give an overall view of the distribution of crime by division over the 13 years 2003-2015 by crime category. The user can assess the highest numbers in terms of actual reports (top chart) or adjusted to population per 10,000 (bottom chart) in each division. The number of crimes or crime rate is encoded by colour drawing the user to significant cells in the table. The tables shows that DMR North Central reports high crime rates in most categories followed by DMR South Central. Looking across rows we see the offences with the highest rates in the data set are *Theft and related offences* and *Public order and other social code offences*.

Highlight Table - Unadjusted

Crime	Divisions												
	Cavan/Monaghan	Clare	Cork City	Cork	Cork West	D.M.R. Eastern	D.M.R. North Ce.	D.M.R. Norther.	D.M.R. South Ce.	D.M.R. Souther.	D.M.R. Western	Donegal	Galway
Attempts or threats to murder, assaults, harassments and related offences	7,851	4,298	11,352	4,928	4,467	4,099	8,419	10,929	10,821	9,731	10,325	9,883	9,551
Burglary and related offences	5,802	4,979	11,377	5,113	3,653	16,455	13,535	26,824	22,860	32,868	24,943	5,684	13,000
Controlled drug offences	3,311	4,170	12,569	6,524	4,537	6,008	13,533	18,255	11,811	17,420	20,133	4,056	6,180
Damage to property and to the environment	9,043	8,585	28,353	10,008	6,779	16,305	24,092	39,516	24,909	38,083	46,826	13,556	17,370
Dangerous or negligent acts	8,384	4,026	8,642	5,644	4,083	2,857	4,627	10,544	5,906	8,304	11,397	6,936	8,600
Fraud, deception and related offences	1,779	1,046	3,180	1,402	1,118	2,073	4,529	5,753	5,746	3,972	4,945	1,755	2,900
Kidnapping and related offences	44	22	46	36	15	38	82	155	97	84	135	50	400
Offences against government, justice procedures and organisation of crime	2,315	2,113	6,530	1,679	1,135	3,220	43,122	5,873	5,684	5,817	9,319	1,693	3,900
Public order and other social code offences	19,258	14,148	47,651	17,820	13,543	15,745	43,403	31,708	52,354	29,639	31,998	26,345	33,400
Robbery, extortion and hijacking offences	185	248	1,640	160	97	1,227	4,403	4,107	4,822	3,545	4,901	262	600
Theft and related offences	14,547	14,162	54,271	15,713	10,691	35,566	95,420	77,639	102,392	70,090	91,407	15,786	35,400
Weapons and Explosives Offences	619	776	2,124	702	463	787	3,628	3,199	3,091	2,787	4,358	829	1,000

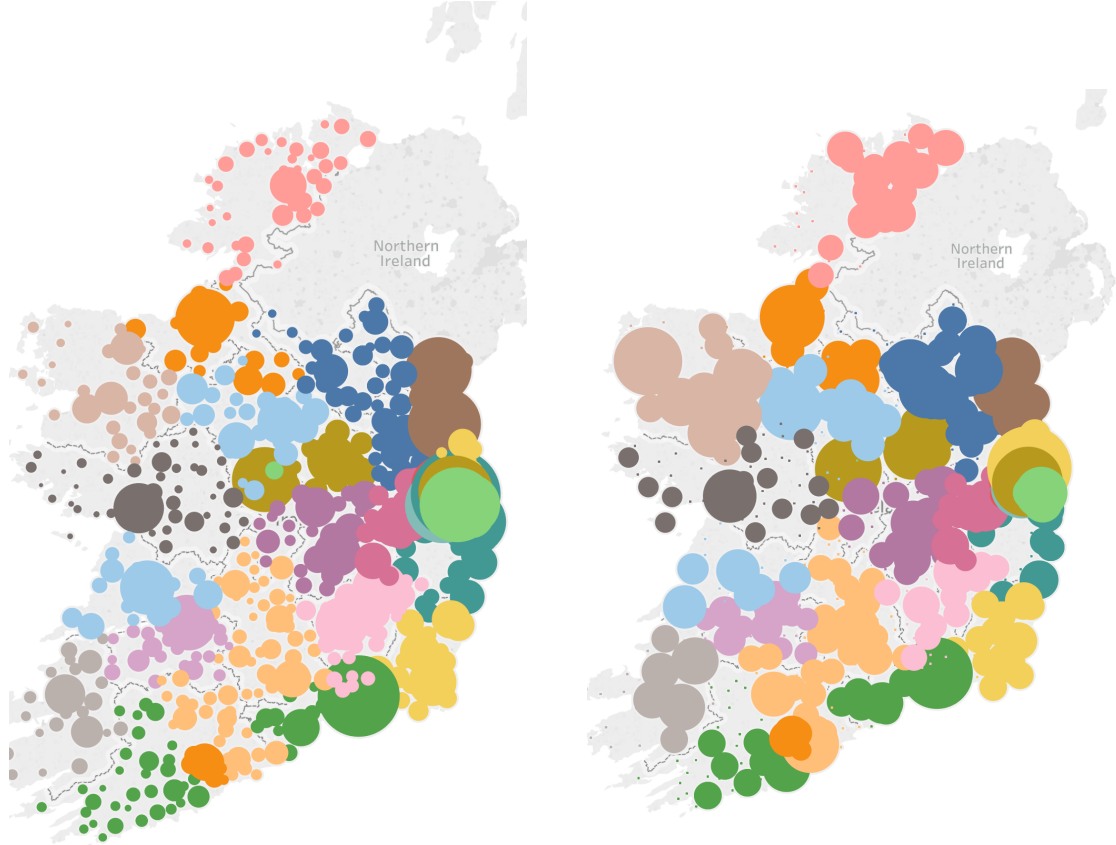
Highlight Table Adjusted by Division Population

Crime	Divisions												
	Cavan/Monaghan	Clare	Cork City	Cork	Cork West	D.M.R. Eastern	D.M.R. North Ce.	D.M.R. Norther.	D.M.R. South Ce.	D.M.R. Souther.	D.M.R. Western	Donegal	Galway
Attempts or threats to murder, assaults, harassments and related offences	587	396	496	321	316	216	965	331	873	407	343		
Burglary and related offences	434	458	497	333	259	866	1,552	813	1,844	1,373	830		
Controlled drug offences	248	384	549	424	321	316	1,552	553	953	728	670		
Damage to property and to the environment	677	790	1,238	651	480	858	2,763	1,198	2,010	1,591	1,558		
Dangerous or negligent acts	627	371	377	367	289	150	531	320	477	347	379		
Fraud, deception and related offences	133	96	139	91	79	109	519	174	464	166	164		
Kidnapping and related offences	3	2	2	2	1	2	9	5	8	4	4		
Offences against government, justice procedures and organisation of crime	173	194	285	109	80	170	4,945	178	459	243	310		
Public order and other social code offences	1,441	1,302	2,080	1,159	959	829	4,977	961	4,224	1,238	1,064		
Robbery, extortion and hijacking offences	14	23	72	10	7	65	505	124	389	148	163		
Theft and related offences	1,088	1,303	2,369	1,022	757	1,872	10,942	2,353	8,261	2,928	3,040		
Weapons and Explosives Offences	46	71	93	46	33	41	416	97	249	116	145		

**Figure 1.3** Highlight table for crime categories by division (unadjusted) (top plot) and adjusted by 10,000 population in each division (bottom plot)

#### 4. Hotspot Map

The availability of longitude and latitude for each Garda station allows the distribution of crime to be visualised as a map display. The plots in Figure 4.1 represent the number of crimes (scaled by division population) reported for burglary (left plot) and kidnapping (right plot) encoded by a circle with the diameter proportional to the crime rate. The different colours in Figure 1.4 represent different divisions. Plots of this kind allows for regional variations within divisions and the country overall to be visualised and contrasted. For example, the distribution of burglaries is weighted towards the east of the country while kidnapping reports hotspots can be observed in north Donegal.

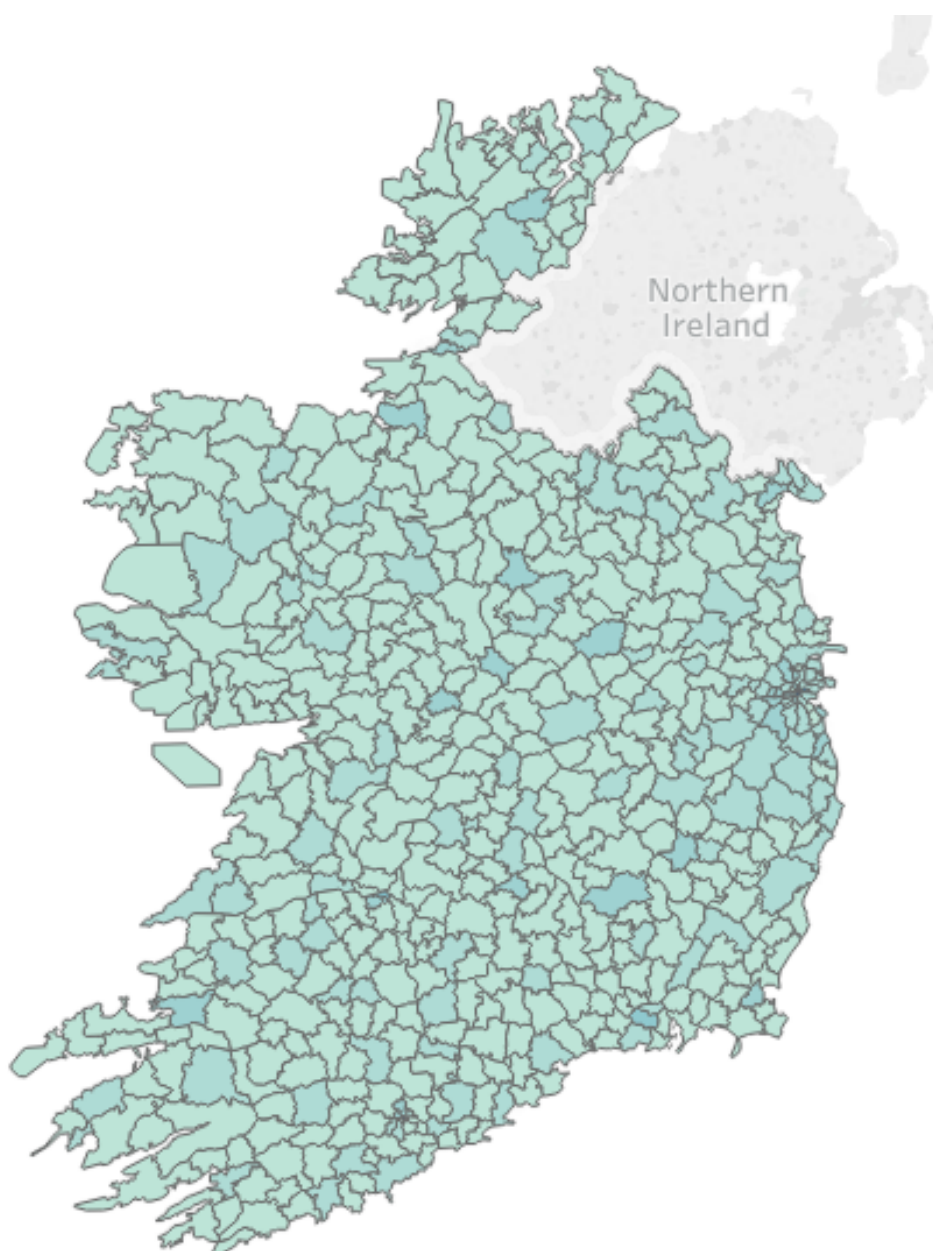


**Figure 1.4** Hot spot map of burglary (left) and kidnapping crime (right)

However plots of this kind have their limitations as 573 circles with size determined by the number of crimes reported to each Garda station can lead to considerable clutter. Another option is to use a polygon map where the boundaries of each Garda station are visualised and which can be encoded by colour corresponding to the crime density. This visualisation will be examined in the next section.

## 5. Polygon Files using Garda Station Boundaries

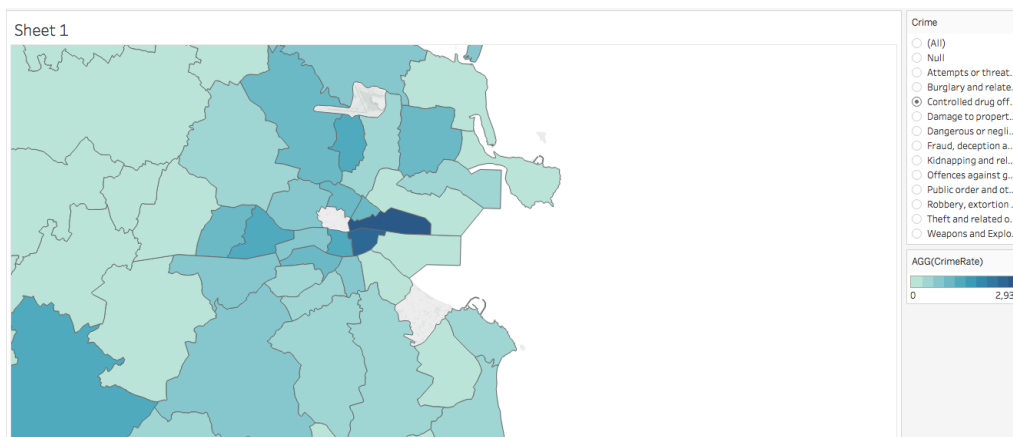
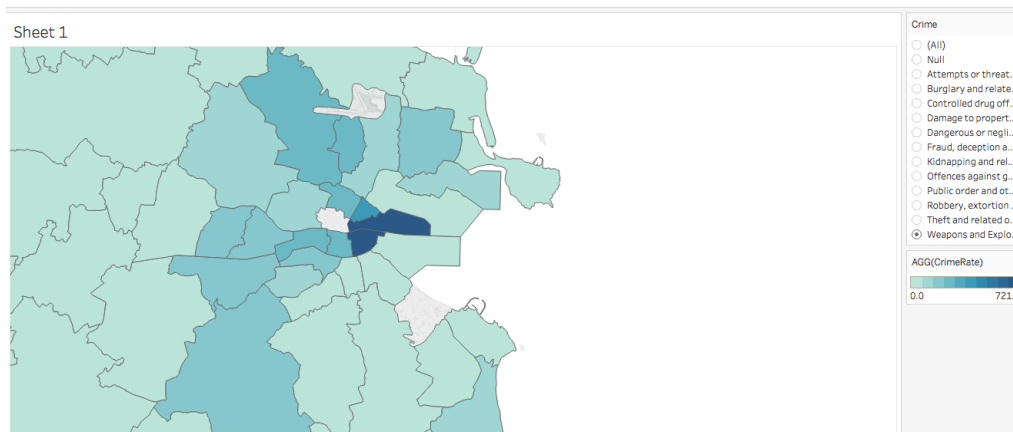
Visualisation of the distribution of crime at Garda station level can also be achieved by using the ESRI shape files from CSO which involve the display of a polygon representing the boundary of each Garda station as shown in Figure 1.5. However, linking this file with the open data crime file leads to a number of Garda stations not being matched. This arises because the open data file Garda station names are a mixture of english and irish language forms whereas station name in the CSO boundary file is either an all english station name variable or an all irish station name variable.



**Figure 1.5** Garda station boundaries, 2011



Visualisation of crime by station boundary is most effective when the map is increased in size to reveal the colour gradients. For example, looking at the Dublin Region the higher crime rates for *weapons and explosives* in the north inner and south inner cities, northwest, west and south west regions contrasts with the experience of Garda stations in south Dublin as shown in Figure 1.6 (top graph).



**Figure 1.6.** Crime rates for *Weapons and Explosives* (top) and *Controlled Drug Offences*, 2003-2015 by Garda station

The elevated rates for *controlled drug offences* can be seen in the outer stations of the city in the bottom plot of Figure 1.6.

## 6. Multivariate Analysis of Crime Data using Principal Component Analysis

One useful statistical technique which can be used to summarise the principal features of the crime data set is Principal Component Analysis (PCA). This technique involves reducing the complexity of a data set by taking into account correlations that may exist between the 12 crime variables. The objective is to derive a small number of components that account for a high percentage of variation. For example instead of using the 12 crime variables we can use a smaller number of principal components to describe the data set. These components are based on linear transformations of the original 12 variables.

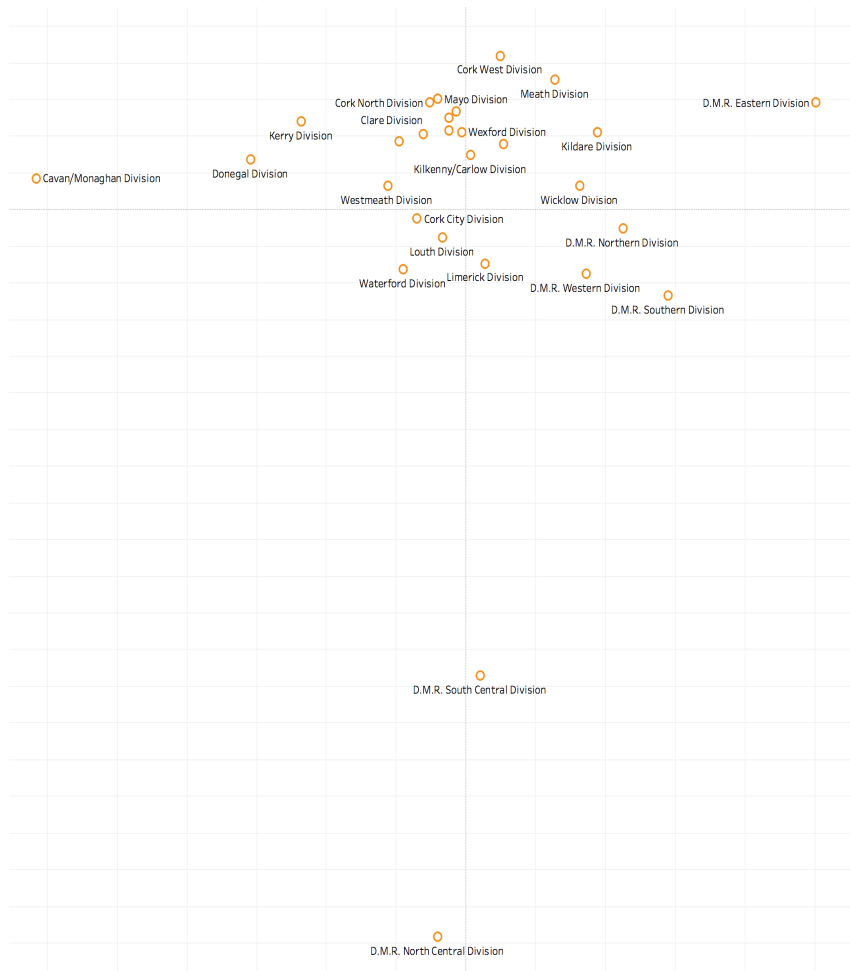
Applying PCA shows that the 12 crime variables can be replaced by two components which account for 90 per cent of the total variation of the data set. The first component is basically a weighted average of each crime category while the second component can be considered a contrast between crimes against property and crimes against people. Plotting the two components as a scatterplot the clustering of divisions with similar crime profiles is evident. In addition a number of divisions stand out as having distinct profiles as shown in Figure 1.7.

A rough visual analysis suggests that divisions with similar profiles are:

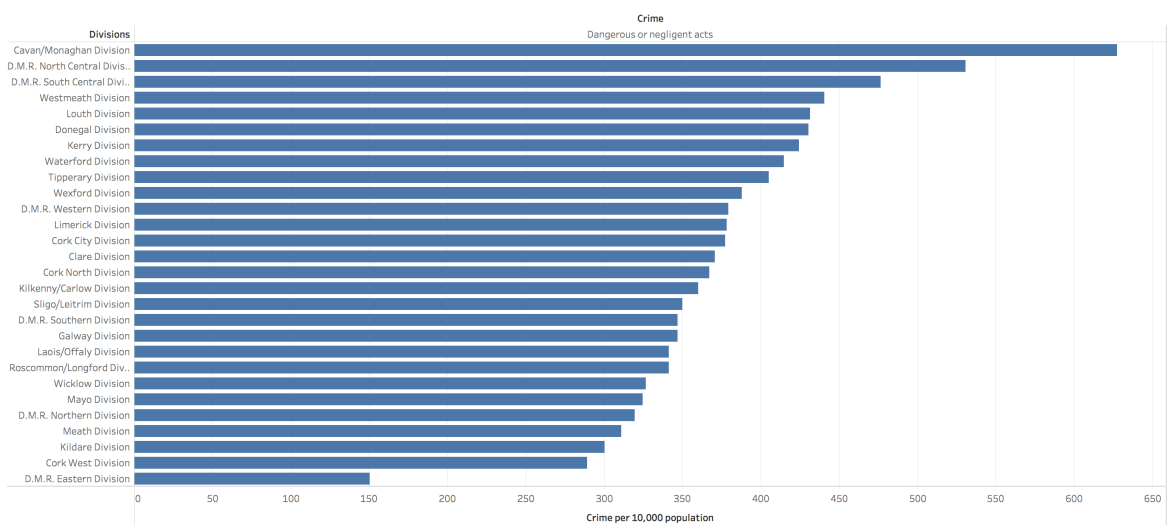
1. Donegal & Kerry Divisions
2. DMR South Central & DMR North Central
3. DMR Northern, DMR Southern, DMR Western & Wicklow, Kildare and Meath
4. Limerick, Waterford, Cork City and Louth
5. Galway, Sligo, Leitrim, Roscommon, Clare, Tipperary, Wexford, Westmeath, Cork North, Cork West, Laois Offaly and perhaps Mayo Division.

The above groupings fall into regional clusters e.g. west of Ireland cluster or greater Dublin area cluster etc. The crime rates that differentiate the clusters could then be further investigated.

Cavan Monaghan and DMR Eastern are two outlier Divisions in the scatterplot which can be explained by their unusual crime rates in the category *Dangerous or negligent acts*. For Cavan/Monaghan the rate at 631 per 10,000 population over the 13 years is the highest of all 28 divisions while DMA Eastern is the lowest at 150 per 10,000 population as shown in Figure 1.8.



**Figure 1.7** Plot of first two principal components



**Figure 1.8** Dangerous or negligent acts per 10,000 population 2003 - 2015 by garda division

## Appendix 1: Data Dictionary

Variable Name	Description
Year	2003-2015
id	
Station	Garda Station Name (573)
Division	Garda Division Name (28) Cavan/Monaghan Clare Cork City Cork North Cork West D.M.R. Eastern D.M.R. North Central D.M.R. Northern D.M.R. South Central D.M.R. Southern D.M.R. Western Donegal Galway Kerry Kildare Kilkenny/Carlow Laois/Offaly Limerick Louth Mayo Meath Roscommon/Longford Sligo/Leitrim Tipperary Waterford Westmeath Wexford Wicklow

Crime Category	Attempts or threats to murder, assaults, harassments and related offences
	Burglary and related offences
	Controlled drug offences
	Damage to property and to the environment
	Dangerous or negligent acts
	Fraud, deception and related offences
	Kidnapping and related offences
	Offences against government, justice procedures and organisation of crime
	Public order and other social code offences
	Robbery, extortion and hijacking offences
	Theft and related offences
	Weapons and Explosives Offences