MPRA

Munich Personal RePEc Archive

Some Statistical Analysis of Greek Crime Data

Tsiamtsouri, Alexandra and Panaretos, John

July 1999

SOME STATISTICAL ANALYSIS OF GREEK CRIME DATA

Alexandra Tsiamtsiouri and John Panaretos

Department of Statistics
Athens University of Economics and Business
76 Patision St., 104 34, Athens, GREECE

Abstract

The increase of criminal activity worldwide calls for cooperation in order to combat crime. All over the world, scientists, including Statisticians, of many seemingly unrelated fields are cooperating in order to find methods to prevent or to reduce crime.

In this paper we analyze Greek crime data. The data are reported crimes per month or per year for the periods 1987-1997 (for the monthly data) and 1982-1997 (for the annual data). Some of the examined crimes are Commonly Dangerous Crimes, Robbery, Theft, Murder, Rape, Arson, Personal Vengeance, Drug Violations, Smuggling of Antiquities and Beggary. Data about the offenders are also examined.

Multivariate methods such as Principal Component Analysis (PCA), and Clustering methods are employed in order to map the crime activity in Greece and point out the areas where certain crimes seem to be increased. Quality Control is employed in order to make inferences about the months when different crimes are more possible to occur.

Keywords: Crime Data, Cluster Analysis, PCA.

Introduction

The increase of criminal activity worldwide calls for the help of statistics in criminal justice. All over the world, scientists of many seemingly unrelated fields are cooperating in order to find methods to prevent or to reduce crime. In a world conference that took place on September 21st -25th

1998 in South Africa, with theme "Modern Criminal Investigation, Organized Crime & Human Rights", it is noted that crime has become more and more sophisticated, technologically advanced and internationalized. It is also noted that this is a global trend and brings in its wake very complicated issues that have, and will continue to have, very severe implications on the maintenance of law and order. Therefore, there is a definite need for cooperation in order to combat crime.

Some analyses of crime data

The first attempt to analyze crime data statistically seems to have been done by Ahamad (1967). The data analyzed by Ahamad were the number of criminal offences of eighteen different types for Great Britain for the years 1950-63. The data were analyzed using Principal Component Analysis (PCA) in order 'to investigate the relationships between different crimes and to determine to what extent the variation in the number of crimes from year to year may be explained by a small number of unrelated factors'. Ahamad, analyzed the (18×18) correlation matrix of the data and found that three components accounted for 92 per cent of the total variation. He interpreted the first component as measuring the general increase in crime due to changes in population structure and social conditions. This interpretation of the first component given by Ahamad coincides with the interpretation by Chatfield and Collins (1980). However, trying to interpret the second component, Ahamad (1967) suggests that this component may 'reflect changes in recording practice by the police over the period'. Chatfield and Collins (1980) comment this interpretation by Ahamad as rather dubious. Moreover, Ahamad (1967) admits that identification of the third component is 'extremely difficult'.

Recently, Kent Borowick in his PHD dissertation entitled "Analysis of low-probability count data with applications in crime analysis" (1997) says that crime was chosen for his research because it has become a measure for quality of life and the public is increasingly concerned with reducing and controlling criminal activity. His dissertation focuses on analytical

techniques, which are useful for low probability count data such as crime data. Three major topics are discussed. The first topic includes possible distributions of crime data along with a general description of the data and collection procedures. A Poisson-Binomial model is introduced which is then modified using Bayesian priors. All models are interpreted in terms of crime data and fit to crime counts for Waco, Texas. Two goodness-of-fit tests are reviewed and used to test the fit of each marginal distribution to the observed data. The second topic reviews nonparametric statistics used to test for significant differences in the location of two populations. The third topic introduces the use of process control techniques to crime analysis. Borowick (1997) suggests that the use of quality control techniques provides a significant opportunity to improve crime analysis methods using a methodology that is widely utilized in business, industry and government.

Applications to Greek crime data

In the sequel, we employ some of the above mentioned techniques, as well as some others, to analyze Greek crime data.

The available data are crimes reported over the periods 1982-1997 for annual counts and 1987-1997 for monthly counts. Data are tabulated both as crime counts by crime and as crime counts by location. However, information about the exact date or time of the crime occurrence was not available.

The offenses used in our analyses are: commonly dangerous crimes, offenses against life, wounding, offenses against property, beggary, illegal possession and usage of arms, illegal possession of explosives, offenses concerning antiquities, offenses concerning drugs, smuggling, arsons, arsons in forests, murder (either premeditated or not), killings by car, injuries by car, suicide, larceny, robbery and rape.

A first examination of the data shows a general increase of crime activity the last years. Commonly dangerous crimes appear to have increased by almost 64% since 1982, although there was a peak in 1988 where the crimes reported where about two and a half times as many as in 1997. Offenses against life seem to be steadily increasing since 1982. Wounding has

shown several fluctuations since 1982 but the general pattern is increasing. Offenses against property are one of the crimes with the largest increase. They have steadily increased in a high rate, so that in 1997 they are almost four times as many as in 1982 (23423 offenses in 1982 and 90583 offenses in 1997). Beggary shows some fluctuation from 1982 to 1991 with no dramatically change, but from 1992 and on it has a clear increasing pattern with an impressive rate. Illegal possession and usage of arms also had an almost steadily increasing pattern from 1982 to 1994 with higher rate of increase since 1995. Illegal possession of explosives after reaching its peak in 1991 where the number of offenses was almost five times larger than the number of offenses in 1982 has a decreasing pattern since then. The offenses concerning antiquities show a similar pattern. After reaching their peak in the period 1993-1994 they are decreasing since then. Another impressive pattern is that of the drugs' offenses. It is clear that there is a dramatic increase since 1993. Smuggling also has an increasing pattern although it shows some fluctuation. Arsons and arsons in forests show similar patterns with more notable that of the arsons in forests. One could say that after most of the Greek forests have been turned to carbon, arsonists experience great difficulties in finding a forest to burn. Premeditated murder has a steadily increasing pattern. Unpremeditated murder shows some fluctuation since 1987 but it has being decreasing since 1989. Killings by car have a clear increasing pattern, although the increase rate changes over time. Car injuries show the most fluctuating pattern with more evidence of periodicity rather than a certain trend. Suicides have an increasing pattern but the last three years they have slightly decreased. Larceny has almost a linear increase since 1982. Robbery was slightly increasing since 1987 but since then the increase is more rapid.

Figures 1 to 7 give time plots of the above mentioned offenses for the annual data during the period 1982-1997. The time plots of the monthly data in figures 9 to 15 give a more detailed picture of the behavior of these crime counts over time.

Figure 1: Time plots of some crime-variables of the annual data for the period 1982-1997
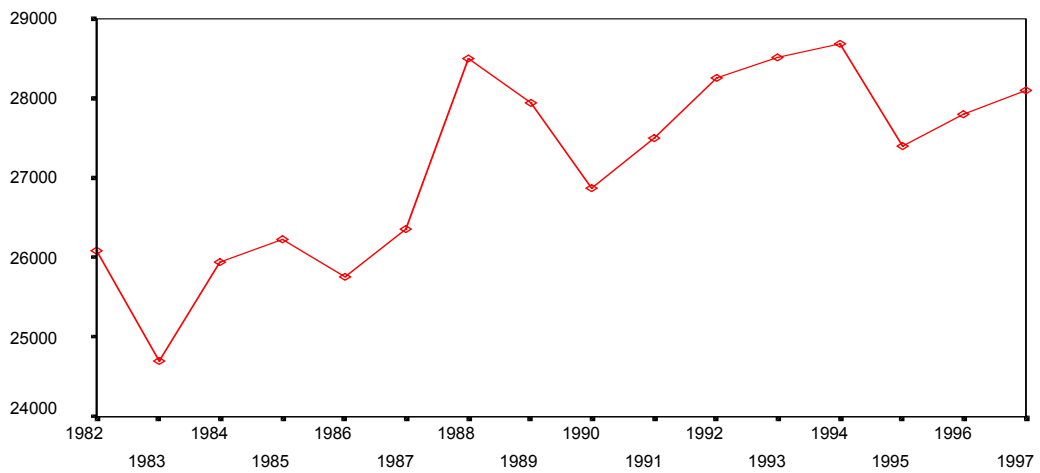

Commonly dangerous crimes
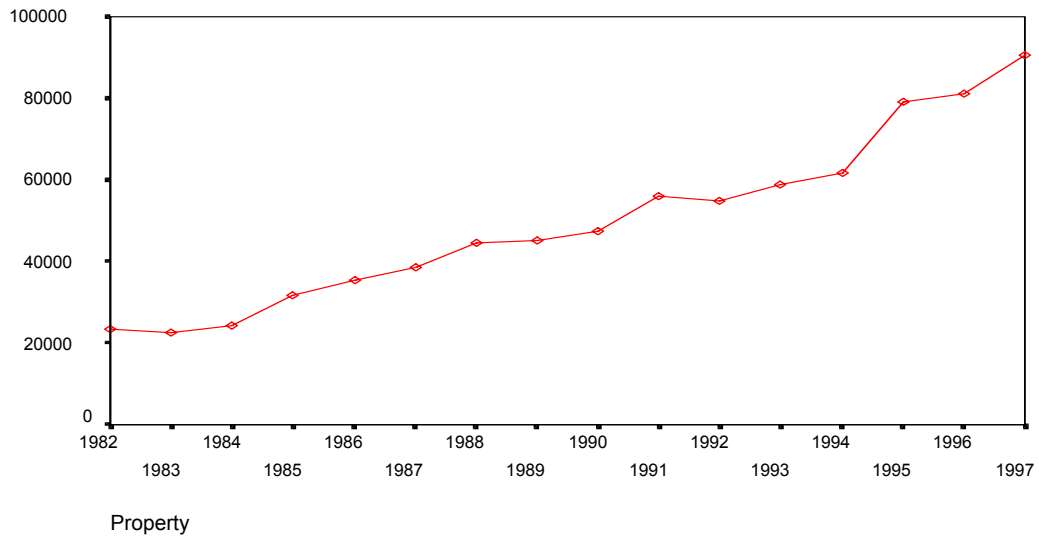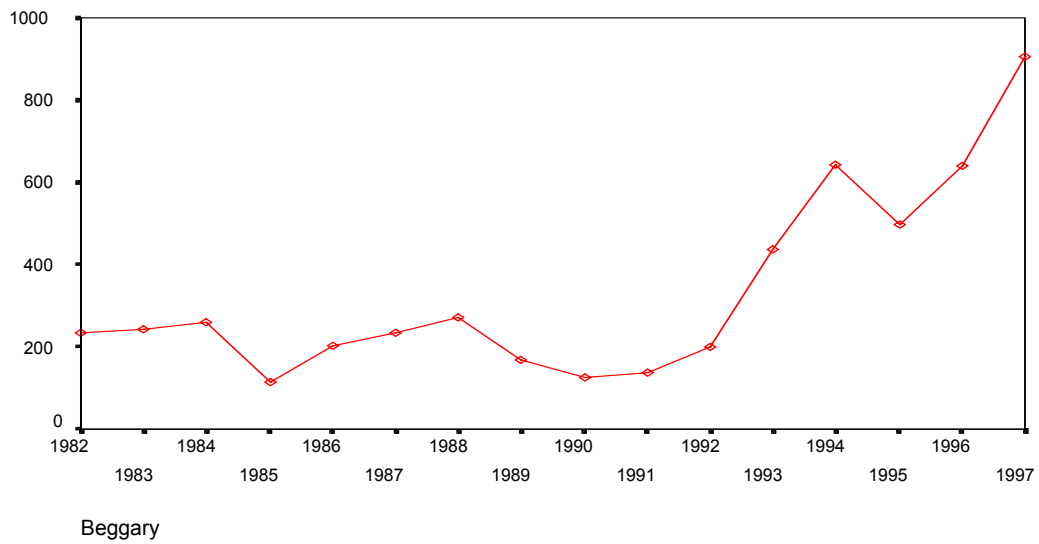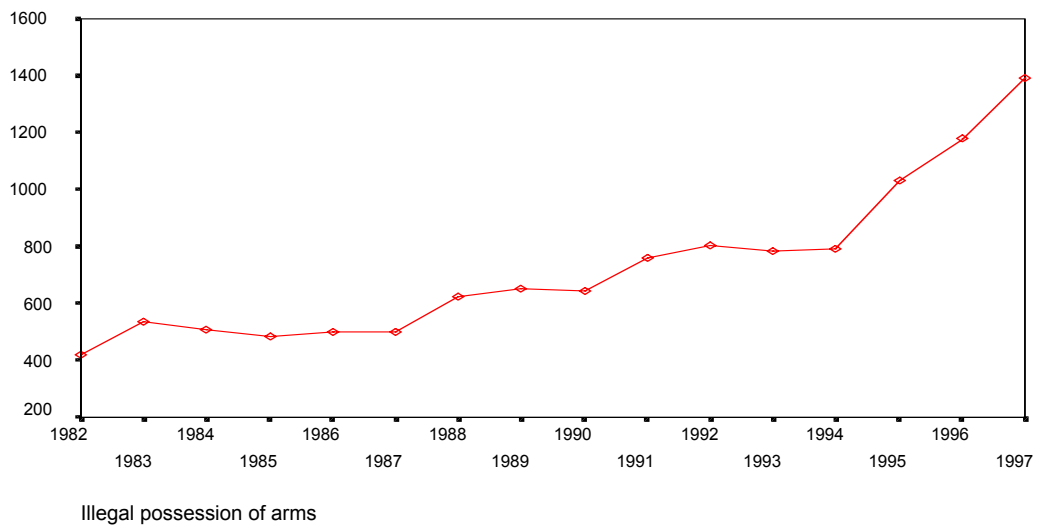
(a)

Offenses against life

(b)
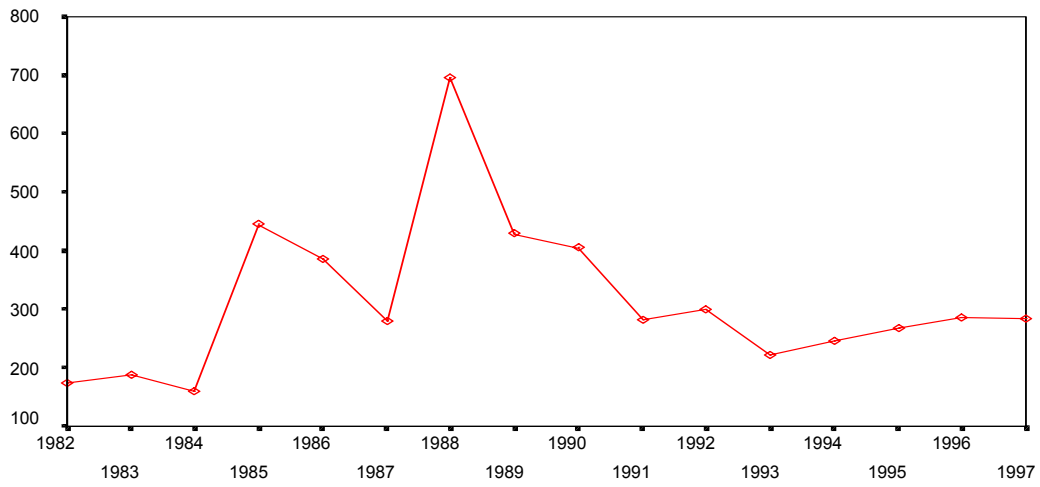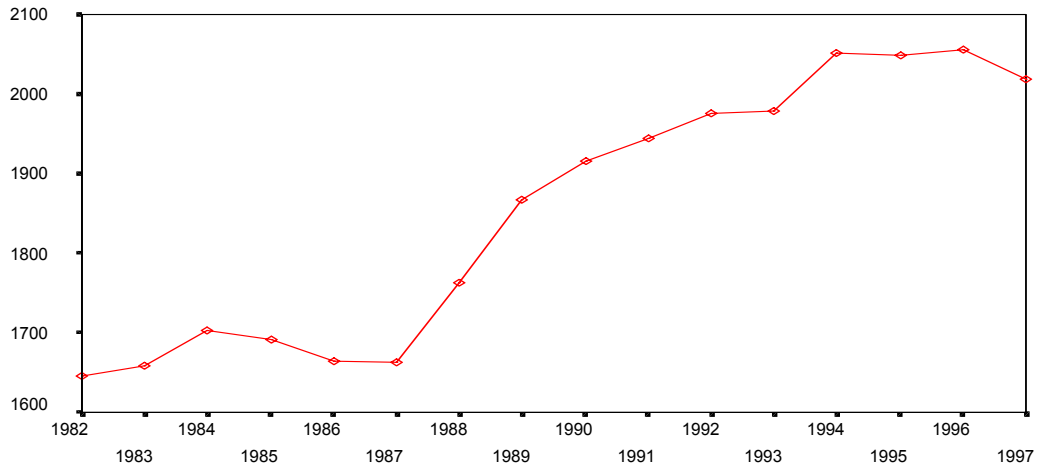
Wounding

(c)

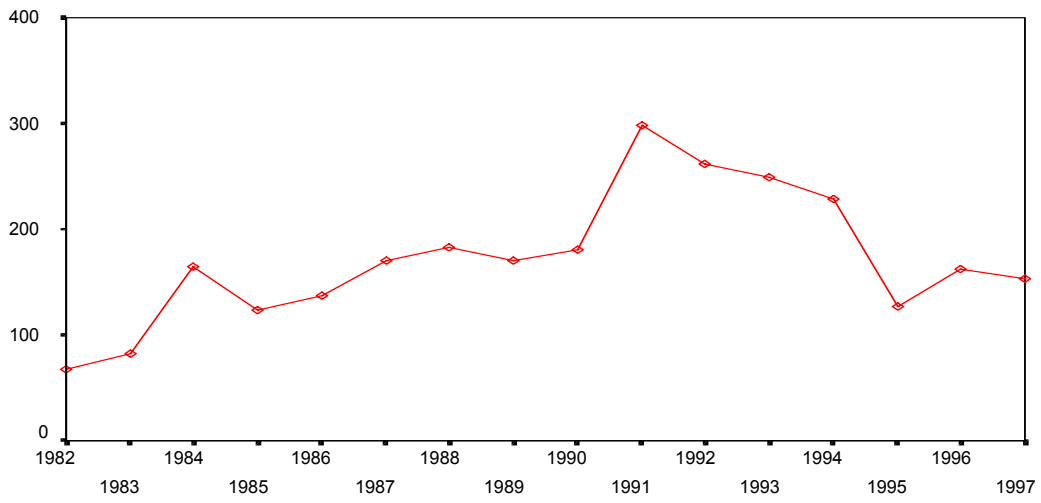**Figure 2: Time plots of some crime-variables of the annual data for the period 1982-1997**

Property

(a)

Beggary

(b)

Illegal possession of arms

(c)

Figure 3: Time plots of some crime-variables of the annual data for the period 1982-1997
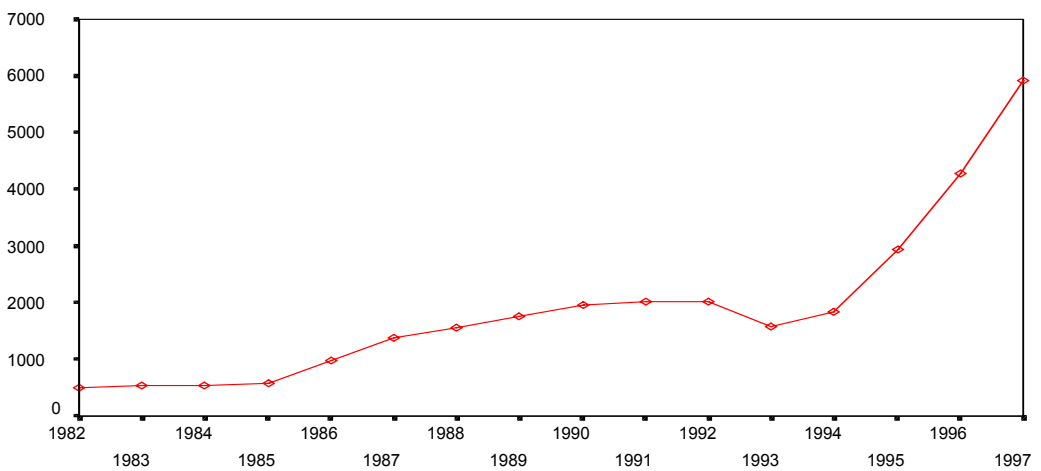

Illegal possession of explosives
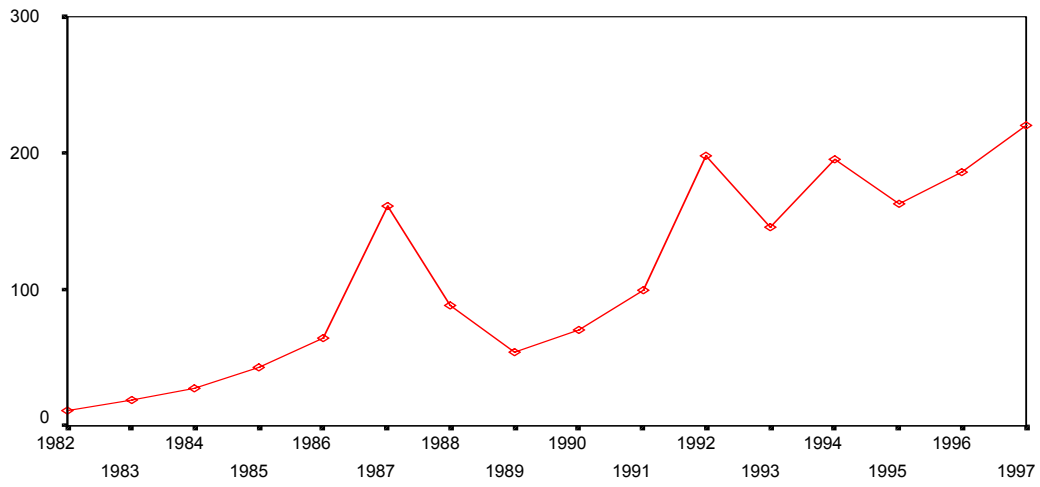
(a)
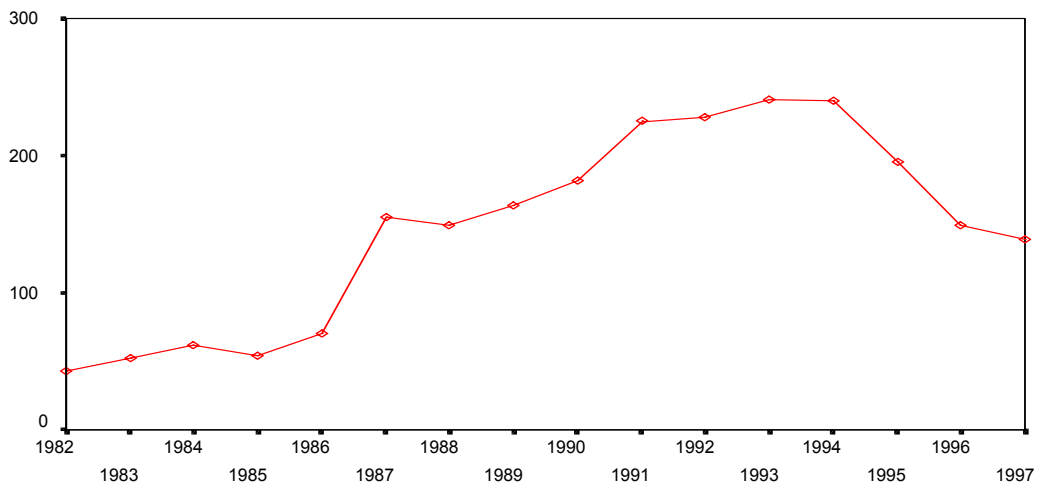
Offenses concerning antiquities

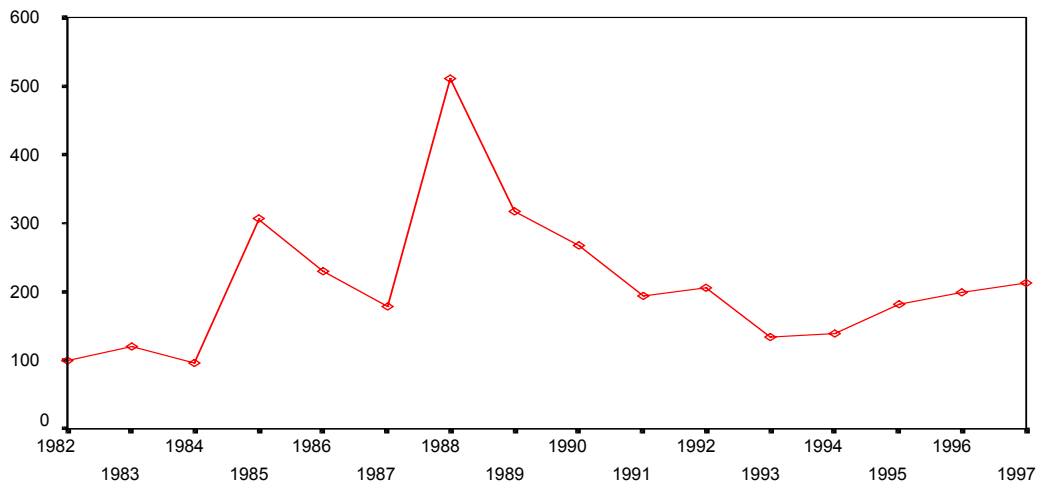(b)

Offenses about drugs' violations

(c)

7

Figure 4: Time plots of some crime-variables of the annual data for the period 1982-1997
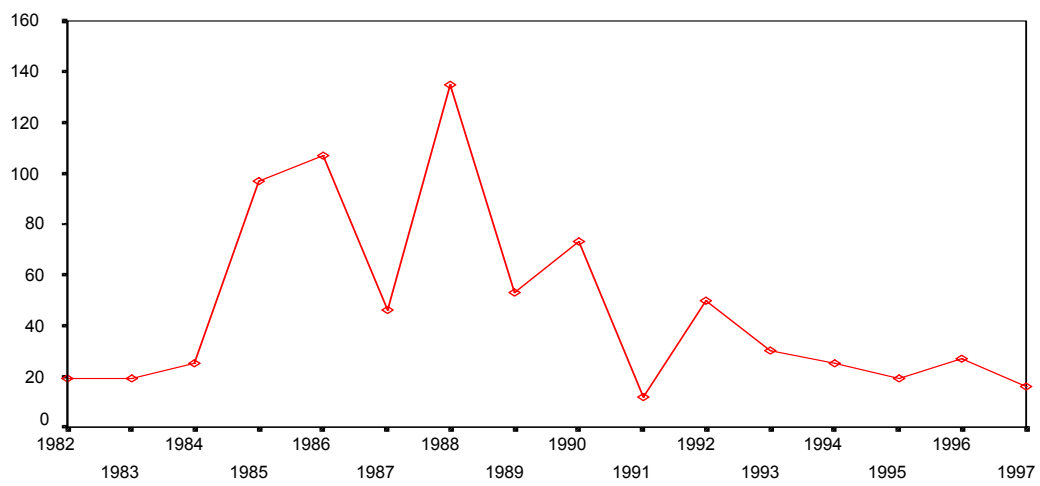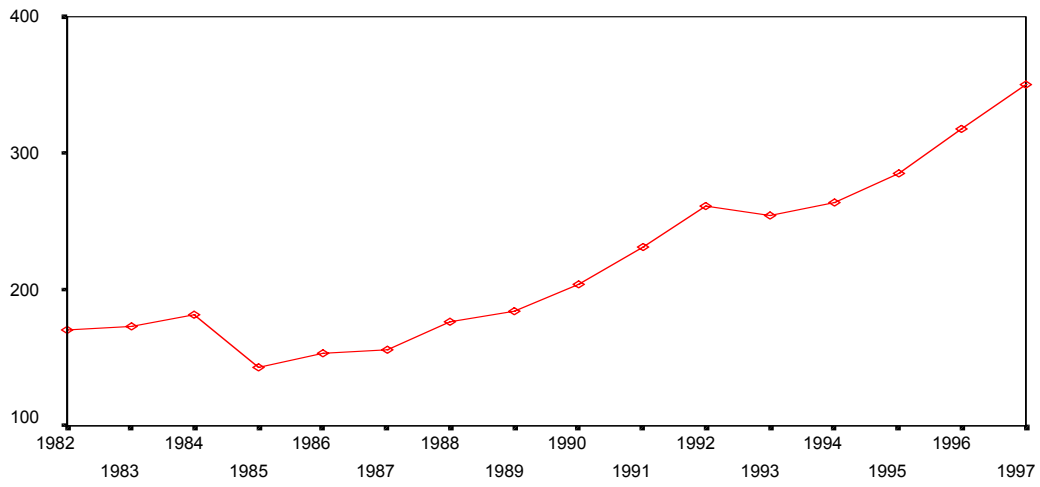

Smuggling

(a)

Arsons
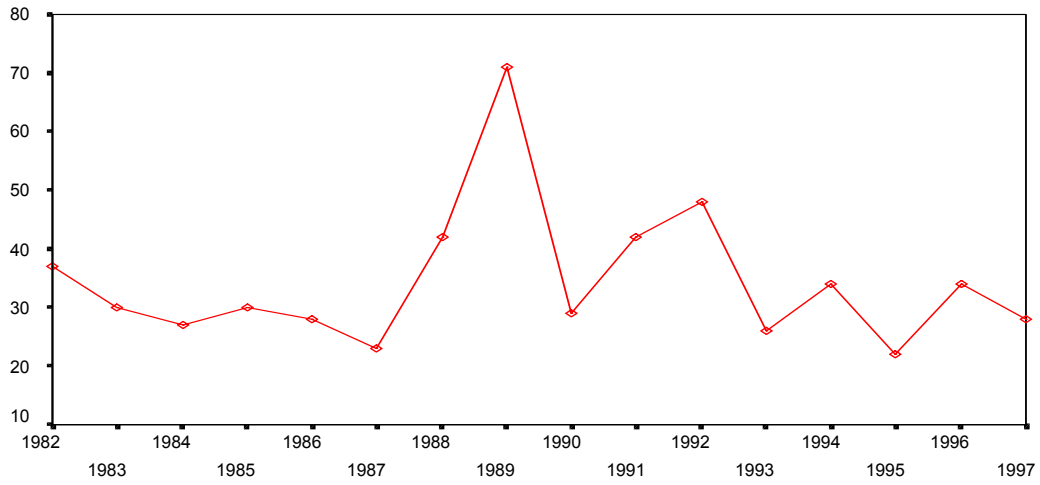
(b)

Arsons in forests

(c)

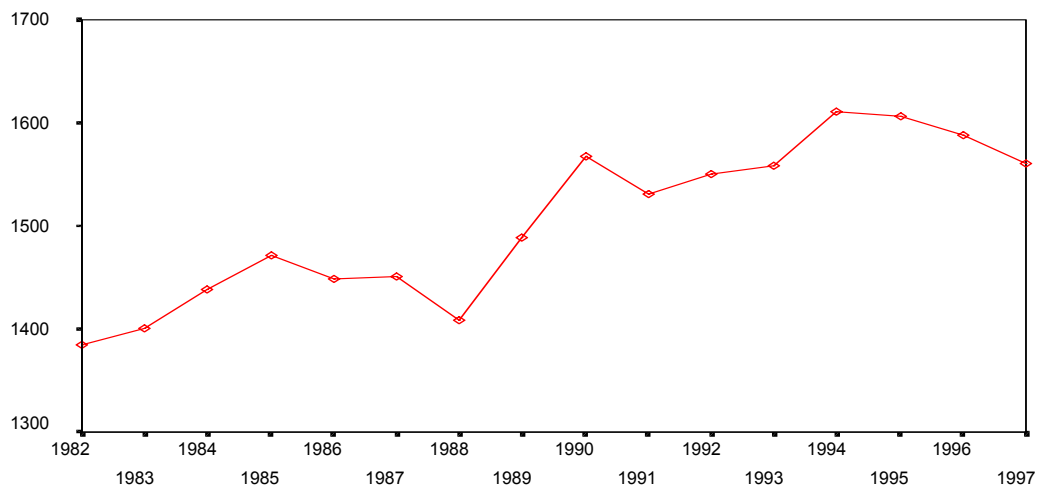**Figure 5: Time plots of some crime-variables of the annual data for the period 1982-1997**

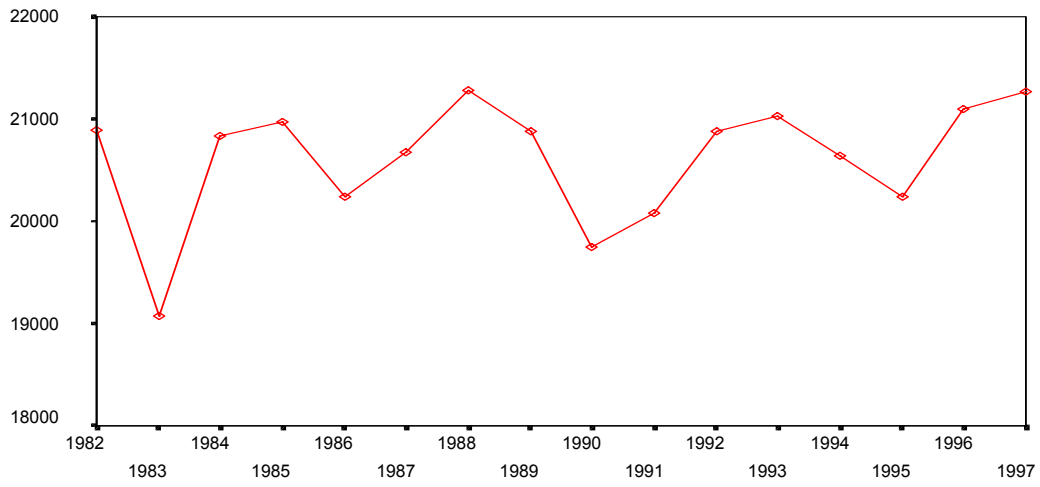Premeditated murders

(a)

Unpremeditated murders

(b)

Deaths due to car accidents

(c)

Figure 6: Time plots of some crime-variables of the annual data for the period 1982-1997

Injuries due to car accidents

(a)
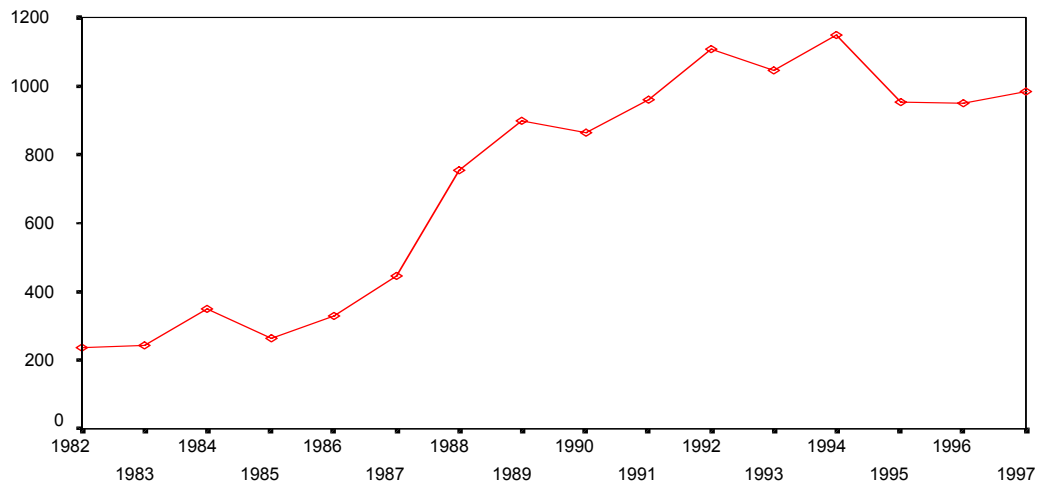
Suicides

(b)
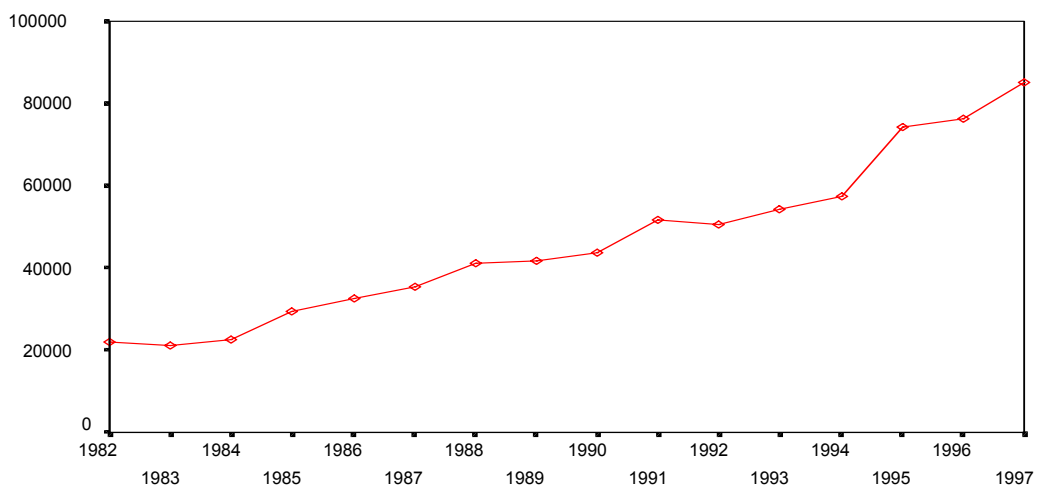
Larceny

(c)

Figure 7: Time plot of robbery (annual data) for the period 1982-1997

Robbery
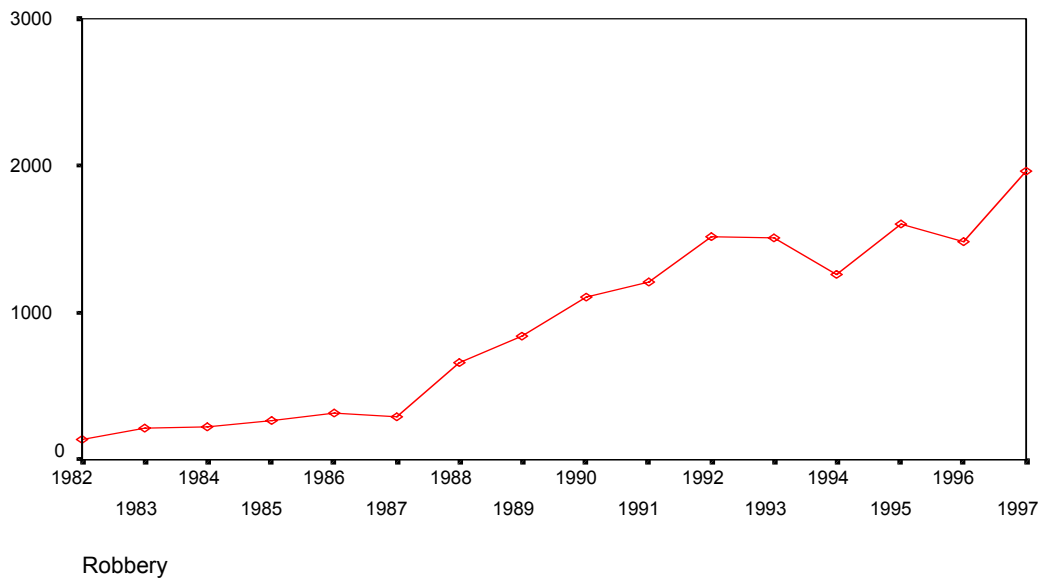
As far as rape is concerned, it has been recorded separately as a crime only after 1988. Therefore, we can only create a picture from 1988 to 1997. A time plot of rapes reported from 1988 to 1997 is given in figure 8.

Figure 8: Time plot of rapes for the period 1982-1997

Rapes

Figure 9: Time plots of some offenses using monthly data for the period 1987-1997

Commonly dangerous crimes

(a)

Offenses against life

(b)

Wounding
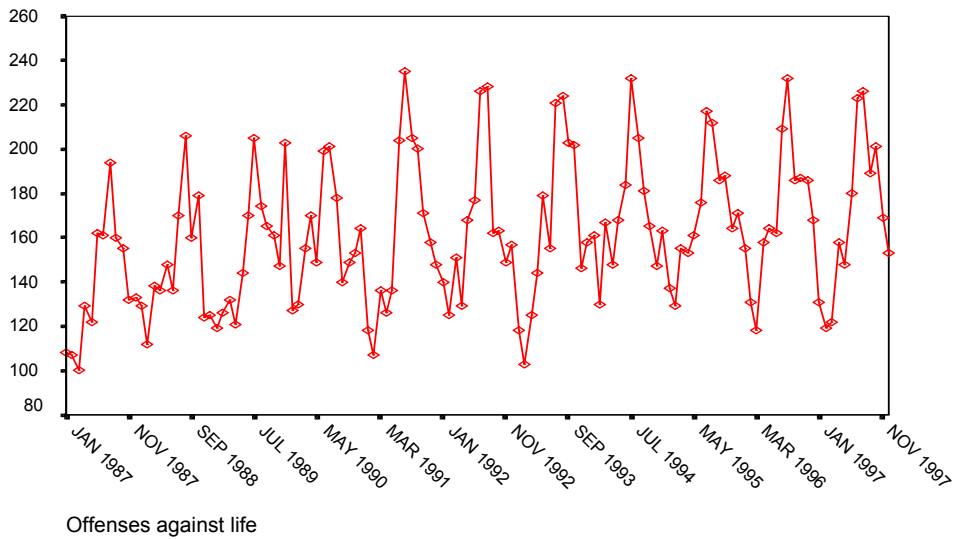
(c)

Figure 10: Time plots of some offenses using monthly data for the period 1987-1997

Property

(a)

Beggary

(b)

Illegal possession of arms

Figure 11: Time plots of some offenses using monthly data for the period 1987-1997

Illegal possession of explosives

(a)

Offenses concerning antiquities

(b)

Drugs' Violations

Figure 12: Time plots of some offenses using monthly data for the period 1987-1997

Smuggling

(a)

Arsons

(b)

Arsons in forests

15
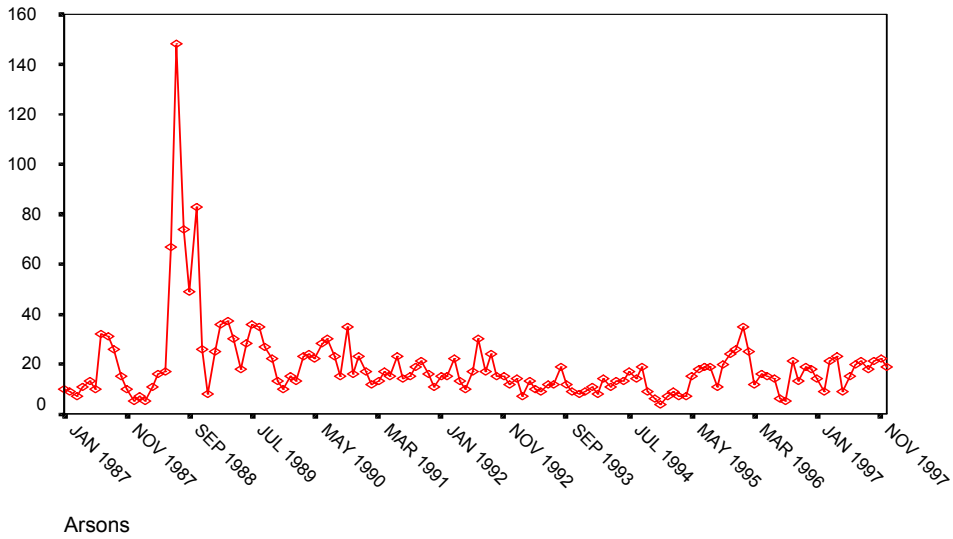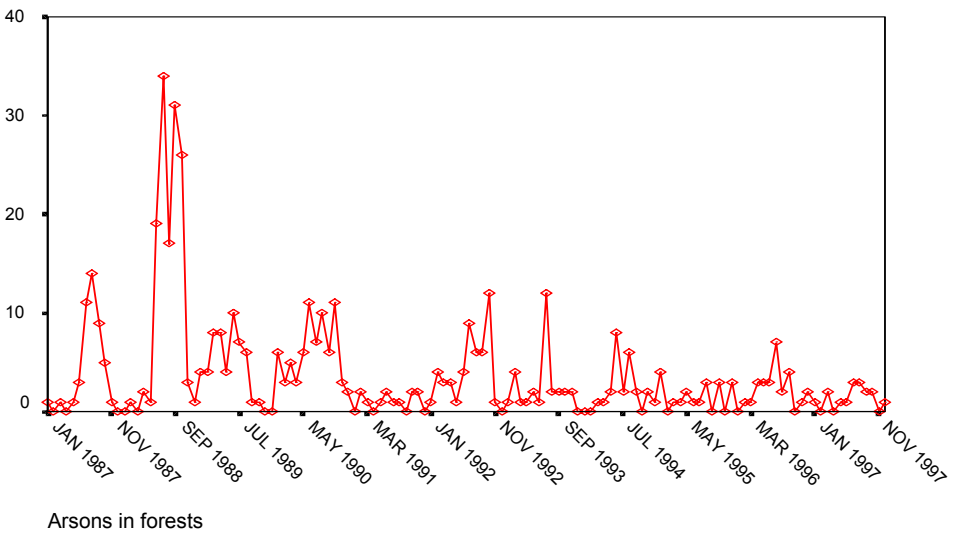
Figure 13: Time plots of some offenses using monthly data for the period 1987-1997

Premeditated murder

(a)

Deaths due to car accidents

(b)

Unpremeditated murder
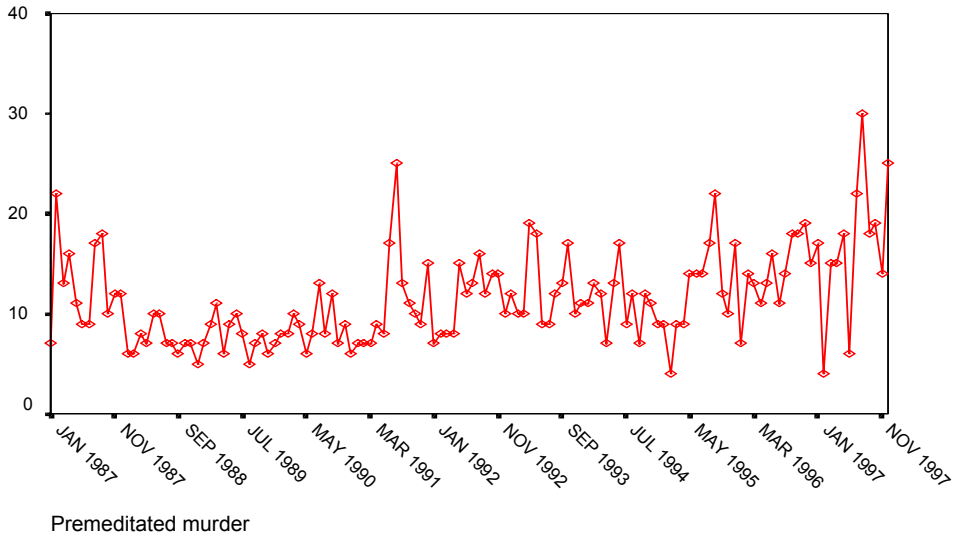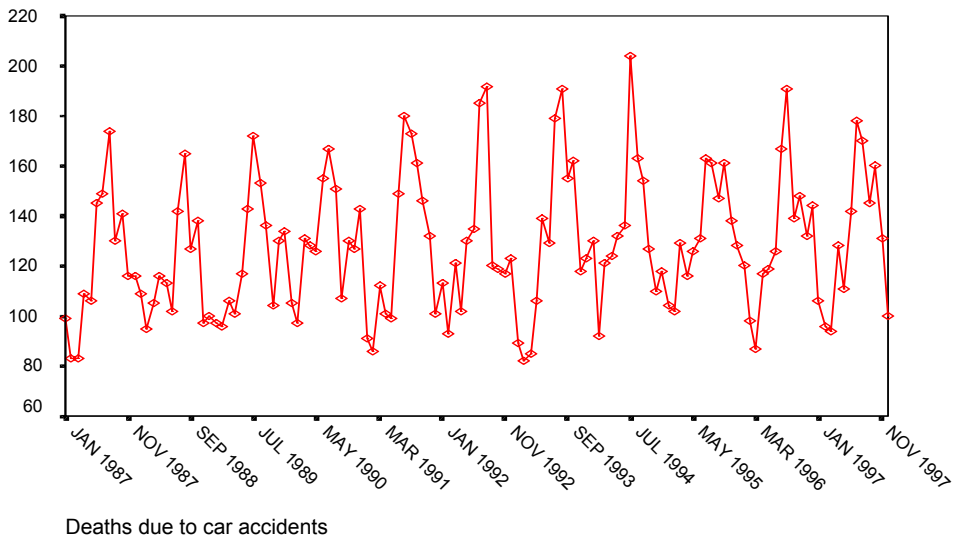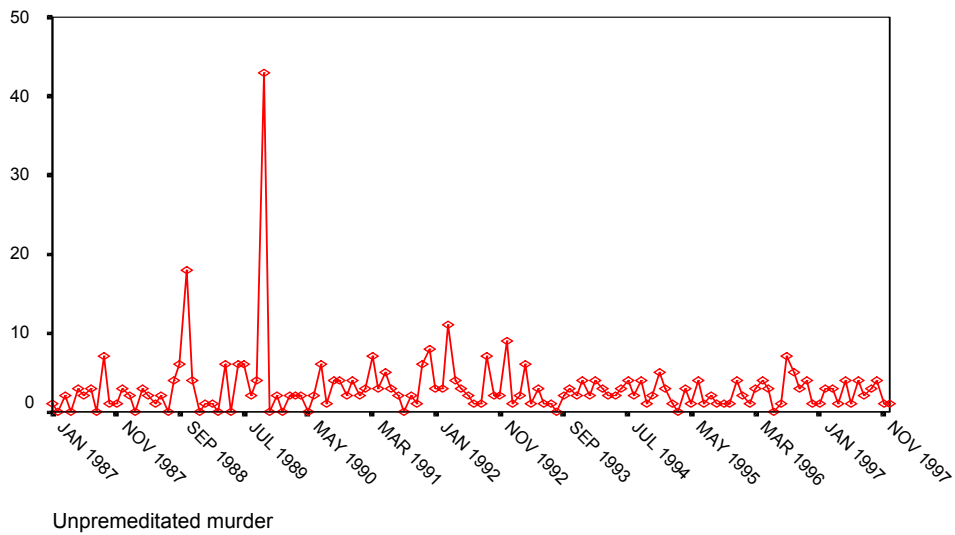
Figure 14: Time plots of some offenses using monthly data for the period 1987-1997

Injuries due to car accidents

(a)

Suicides

(b)

Larceny

17

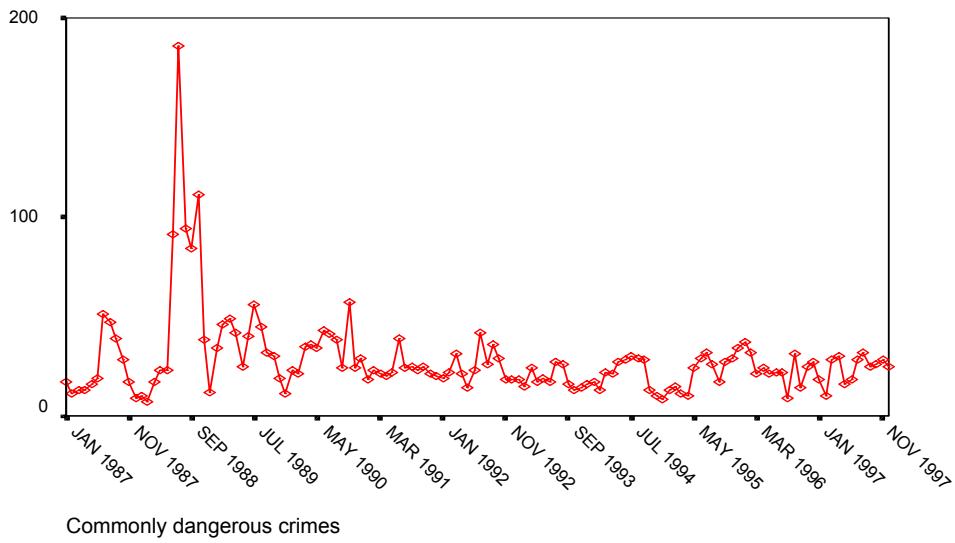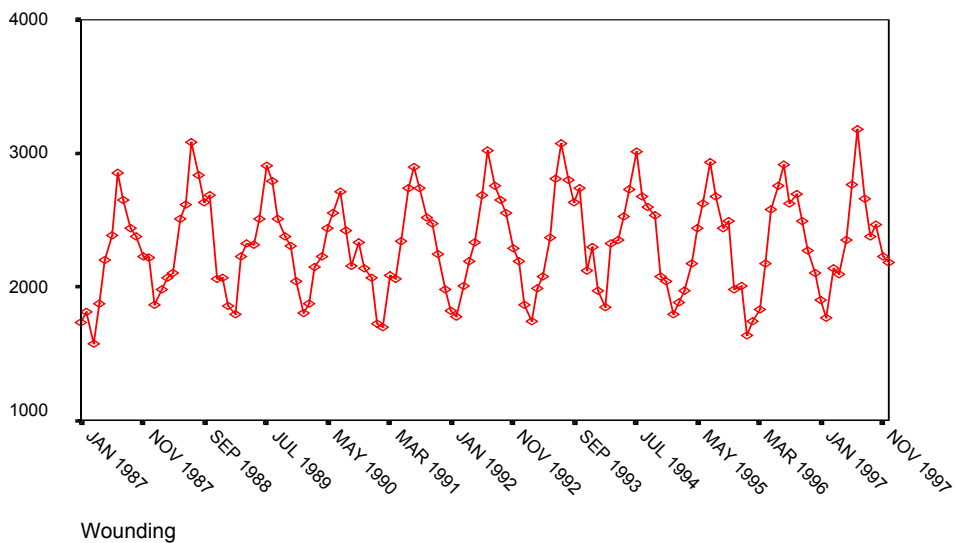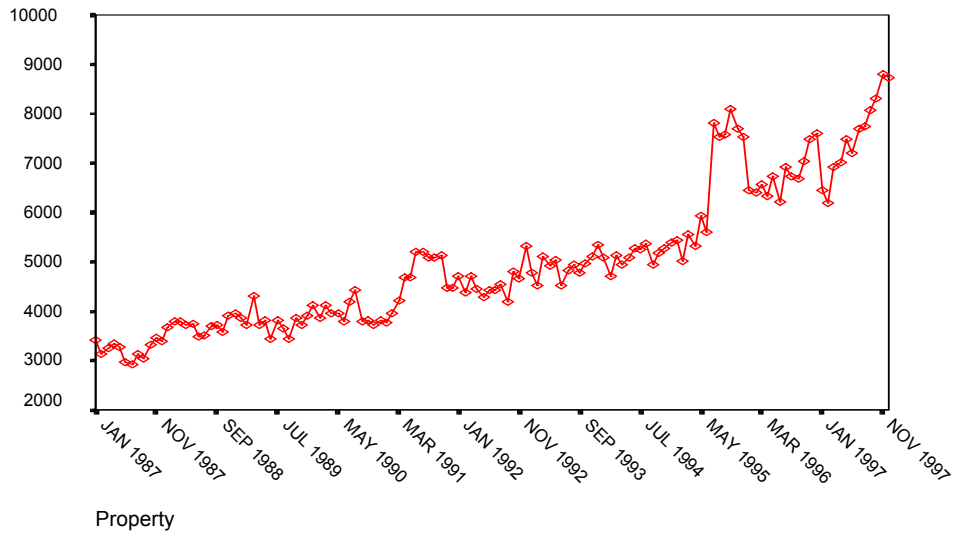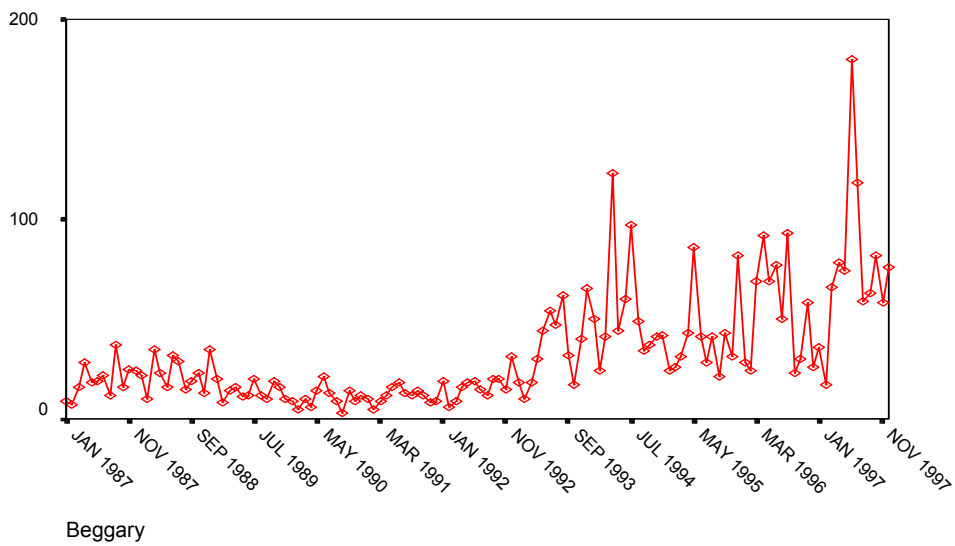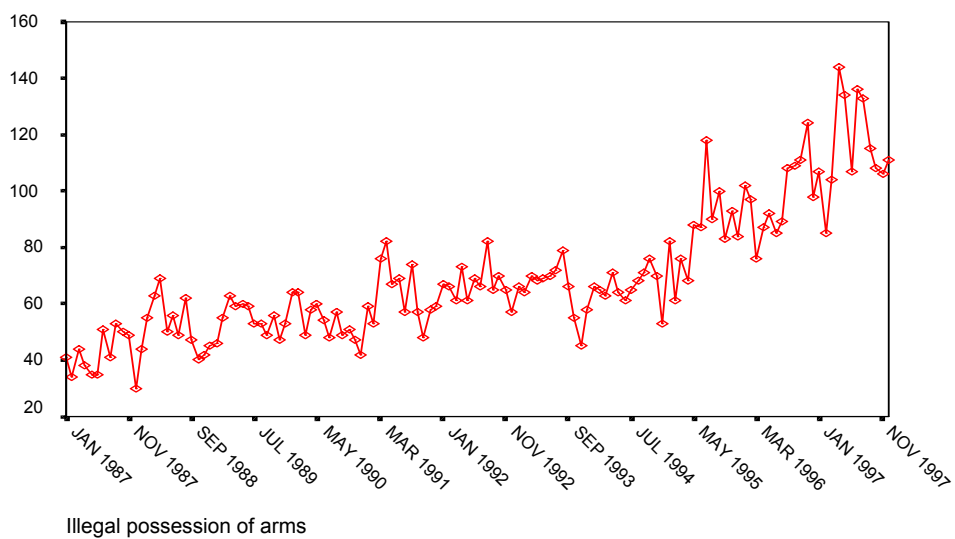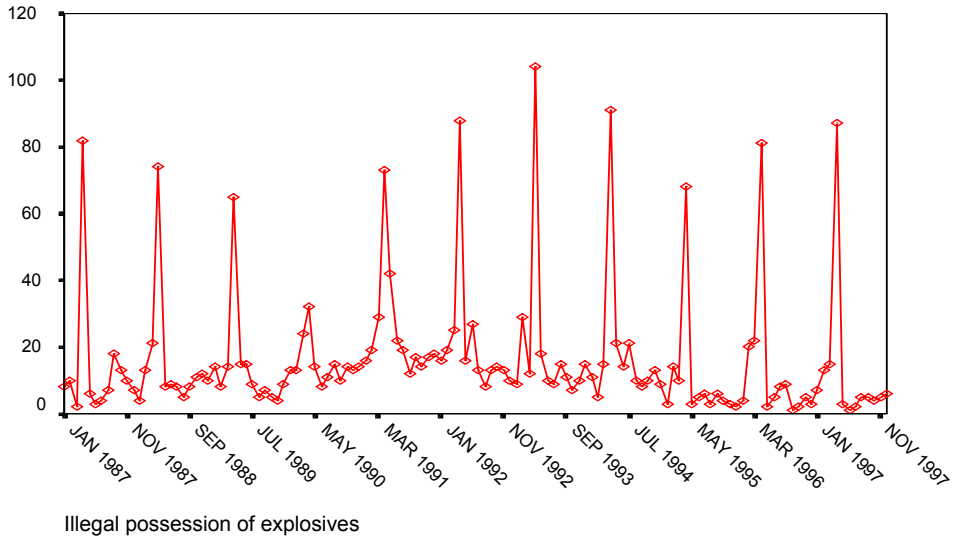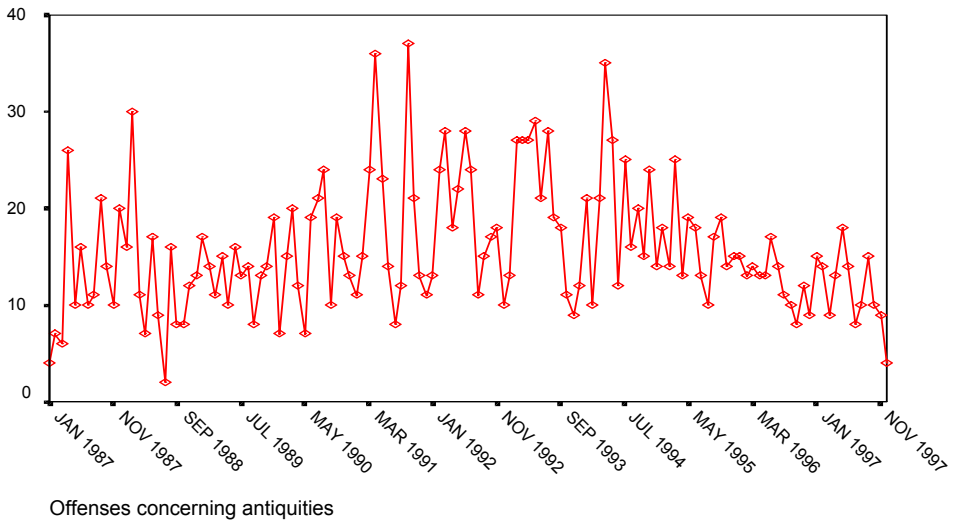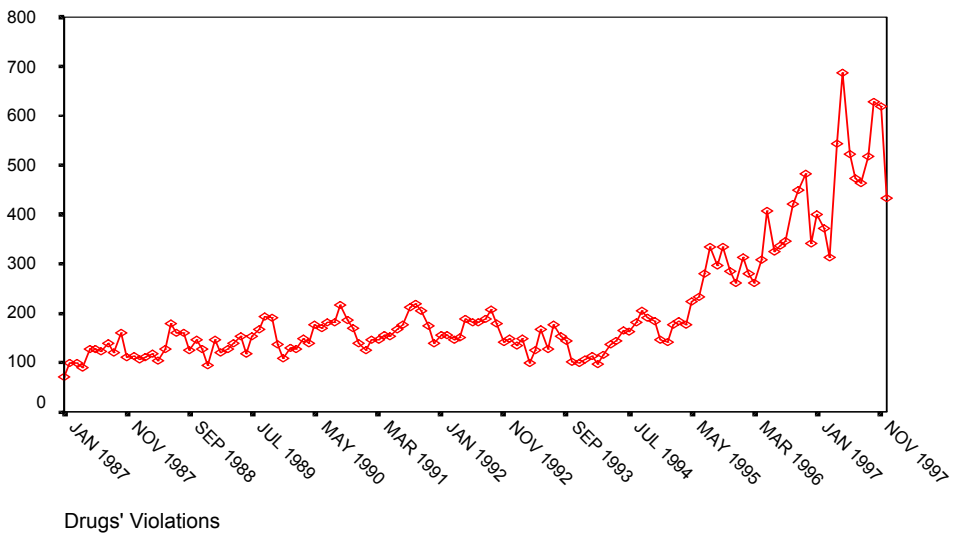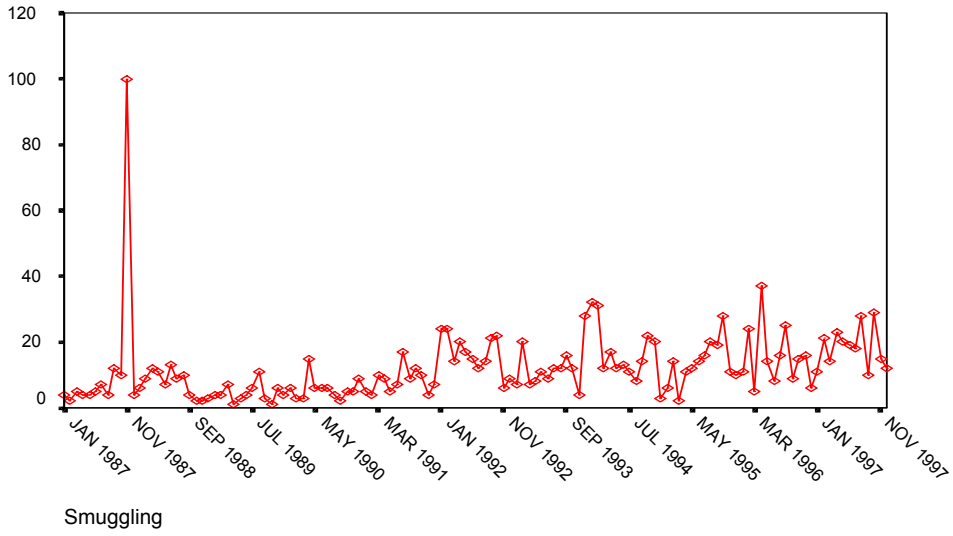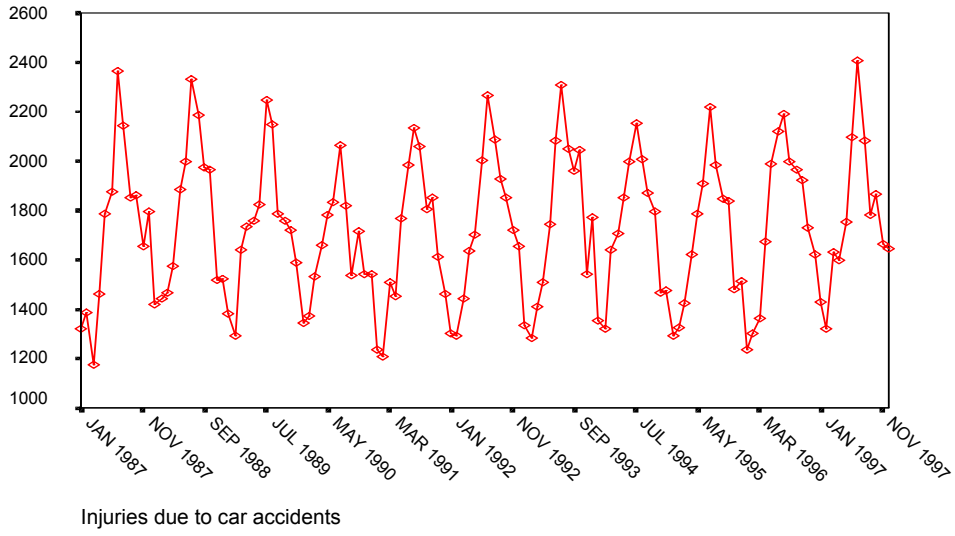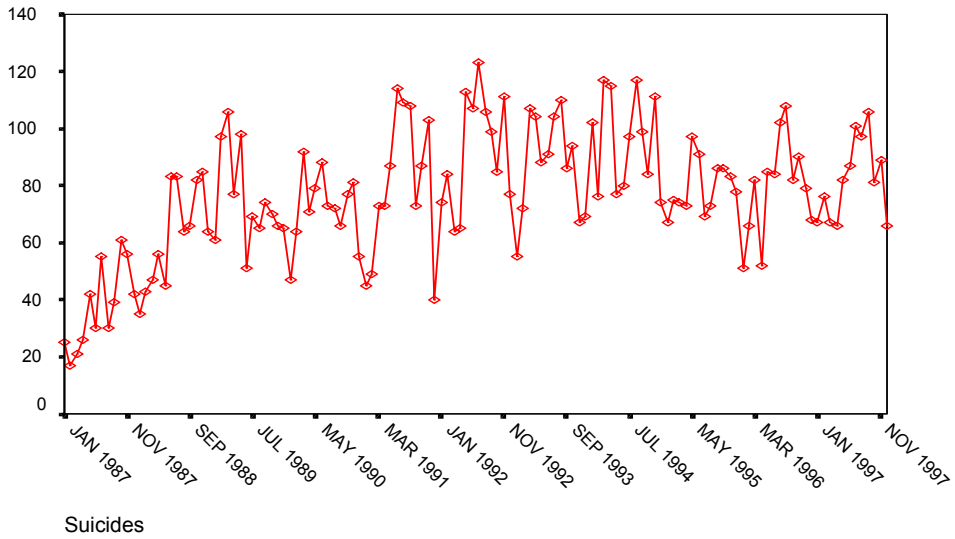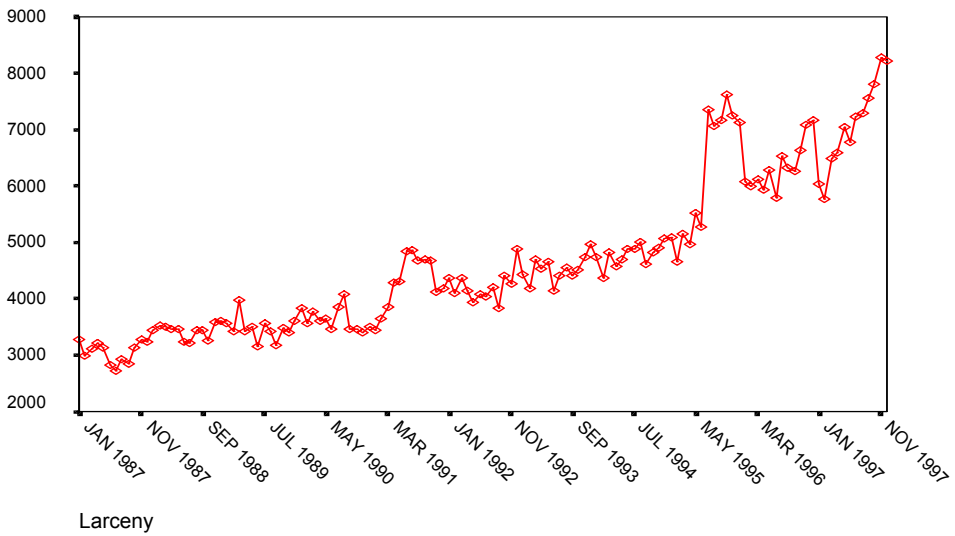**Figure 15: Time plot of robbery using monthly data for the period 1987-1997**

Robbery
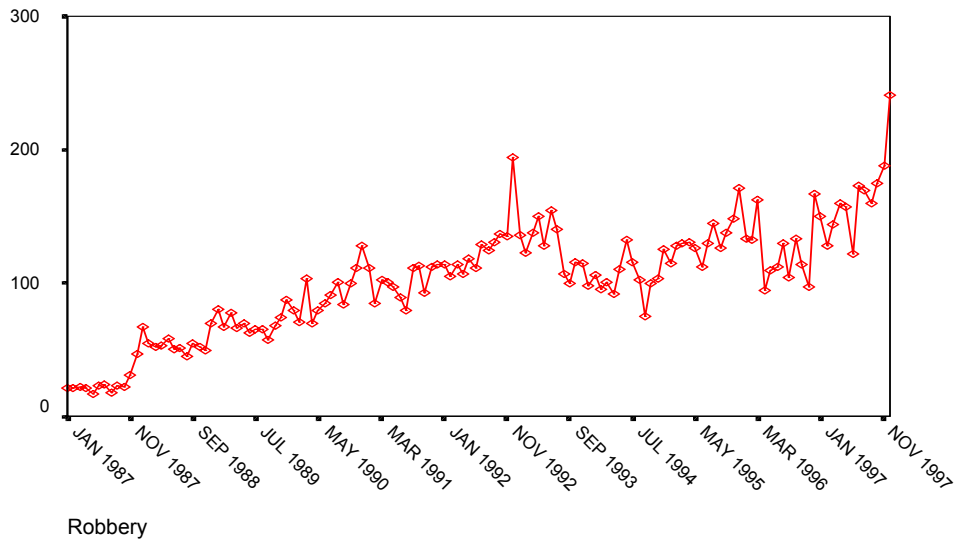
The data were analyzed using PCA analysis. This analysis aims to help us in further analyses. However, its results can give us information about the relationships between the above mentioned crimes. Rape is excluded from this analysis since there are available counts for rape only after 1988. Nevertheless, as far as the monthly data are concerned, a PCA analysis can be employed to all crime-variables since the number of months under consideration (n=120) will still be larger than the number of variables (p=20). Here we present the results of the two PCA analyses (one for the annual and one for the monthly data), with rape excluded, in order to be able to make a comparison.

The PCA analysis of the annual data resulted in four important components, which account for the 88.6% of the total variation. The PCA analysis of the monthly data resulted in five important components, which account for the 74.7% of the total variation. In tables 1 and 2 one can see the percentage of variance explained by each of the components for the annual data and for the monthly data, respectively. The criterion used to determine the number of components is the eigenvalue-one-criterion (see Jobson, J.D., 1992). However, the scree test criterion (see Jobson, 1992) for both the annual and the monthly data indicates that only three components should be retained.

TABLE 1

Variance explained by the four most important components
of the annual data by PCA analysis

Component	Total	% of Variance	Cumulative %
1	11.656	55.5	55.5
2	3.549	16.9	72.4
3	2.224	10.6	83.2
4	1.179	5.6	88.6

TABLE 2

Variance explained by the five most important components
of the monthly data by PCA analysis

Component	Total	% of Variance	Cumulative %
1	6.536	29.7	29.7
2	5.203	23.6	53.4
3	2.377	10.8	64.2
4	1.285	5.8	70.0
5	1.042	4.7	74.7

In tables 3 and 4 the component correlations for the most important components in each analysis are given.

The first component of the annual data is positively correlated with most of the variables. It is negatively correlated with the arsons in forests, and the other physical injuries while it is nearly uncorrelated with the commonly dangerous crimes, the arsons and the homicides by misadventure. In this point it is useful to the reader to note that commonly dangerous crimes include arsons and arsons in forests. Trying to interpret this component we notice that all the crimes except the latter five have been increasing from 1982 since 1997.

TABLE 3

Component correlations for the first four components of the annual data

(1982-1997)

	Component			
	1	**2**	**3**	**4**
Robbery	0.972	-0.019	-0.054	0.097
Offenses against life	0.961	-0.019	-0.163	0.068
Offenses against property	0.961	-0.052	0.211	0.097
Larceny	0.957	-0.065	0.222	0.103
Murders	0.931	-0.279	0.105	0.122
Taking the law into one's hands	0.924	0.223	-0.300	0.018
Illegal possession and usage of arms	0.901	-0.189	0.281	0.221
Non-negligent killing of a person by car	0.886	-0.089	-0.145	-0.064
Smuggling offenses	0.882	-0.053	0.096	-0.148
Simple, unprovoked and dangerous physical injuries	0.854	0.278	-0.344	0.056
Drug offenses	0.838	-0.116	0.402	0.213
Physical injuries	0.823	0.457	-0.094	-0.131
Offenses concerning antiquities	0.779	0.248	-0.500	-0.164
Beggary and Vagrancy	0.728	-0.367	0.411	0.100
Commonly dangerous crimes	-0.045	0.917	0.267	0.231
Arsons	0.003	0.904	0.281	0.282
Arsons in forests	-0.348	0.795	0.295	0.049
Homicides by misadventure	0.035	0.553	-0.331	0.206
Illegal possession and usage of explosives	0.545	0.348	-0.551	-0.365
Physical injury of a person by car	0.330	0.358	0.495	-0.467
Other physical injuries	-0.282	-0.162	-0.537	0.651

On the contrary, commonly dangerous crimes, arsons, homicides by misadventure and especially arsons in forests and other physical injuries have been decreasing after reaching their peak in the years 1988, 1987, 1989, 1987 and 1992, respectively. The value of −0.348 for the arsons in forests is explained by the fact that it is the offense that has shown the most decreasing pattern the last years. Thus, it is obvious that the first component of the annual data measures the general increase of crime.

The second component of the annual data shows the higher correlations with the commonly dangerous crimes, arsons, arsons in forests and homicide by misadventure. Thus, it shows the higher correlations with the offenses that, as mentioned before, have decreasing patterns from a breaking point after. On the other hand, the absolute larger negative correlations appear with beggary, premeditated murder and illegal possession of arms, which have been steadily

increasing and with an increasing rate. Therefore, the second component of the annual data might be interpreted as an index of an offense having a decreasing trend.

The third component of the annual data has the higher correlations, although negative ones, with the illegal possession and usage of explosives, other physical injuries and offenses concerning antiquities, whereas the largest positive correlations, yet not too strong, are those with physical injuries of a person by car, beggary and drug offenses. However, it is not clear what this component represents.

The interpretation of the fourth component is even more difficult and it will not be attempted.

TABLE 4

Component correlations for the first five components of the monthly data (1987-1997)

	Component				
	1	**2**	**3**	**4**	**5**
Offenses against life	0.770	0.436	-0.165	-0.221	0.021
Offenses against property	0.759	-0.557	0.219	0.009	-0.019
Larceny	0.758	-0.557	0.215	0.010	-0.023
Murders	0.757	-0.163	0.041	-0.003	-0.226
Illegal possession and usage of arms	0.713	-0.507	0.241	0.134	0.054
Non-negligent killing by car	0.700	0.488	-0.191	-0.197	-0.007
Drug offenses	0.697	-0.413	0.395	0.013	0.084
Physical injuries of a person by car	0.682	0.653	-0.028	-0.073	0.126
Beggary and vagrancy	0.626	-0.253	0.073	0.242	0.284
Taking the law into one's hand	0.512	0.244	-0.376	0.189	-0.120
Smuggling Offenses	0.477	-0.304	-0.044	0.405	-0.058
Arsons in forests	-0.050	0.739	0.477	0.255	-0.079
Commonly dangerous crimes	-0.028	0.701	0.623	0.273	-0.029
Physical Injuries	0.664	0.688	-0.097	-0.062	0.122
Simple, unprovoked and dangerous physical injuries	0.578	0.633	-0.266	-0.051	0.153
Total cases of rape	0.482	0.544	-0.352	0.124	-0.124
Robbery	0.519	-0.531	-0.091	0.068	-0.153
Offenses concerning antiquities	-0.103	-0.016	-0.676	0.341	0.166
Arsons	-0.045	0.646	0.656	0.267	-0.042
Illegal possession and usage of explosives	-0.212	-0.138	-0.237	0.669	0.381
Homicide by misadventure	-0.136	0.140	0.137	-0.226	0.673
Other simple physical injuries	-0.181	0.333	-0.251	0.249	-0.392

The first component of the monthly data is positively correlated with most of the offenses, as in the analysis of the annual data. The higher correlation coefficients are those of offenses against life, property, larceny, murder, illegal possession of arms, non-negligent killing of a person by car and drug offenses. In our attempt to interpret this component we notice that these offenses are those who have a clear steadily increasing pattern although fluctuation is present. The negative correlations of illegal possession of explosives, homicide by misadventure and other physical injuries can (as in the case of the annual data) be explained because of the fact that these offenses have decreased the last few years (last months as well). Hence, the first component of the monthly data can again be interpreted as measuring the increase of crime activity.

The second component of the monthly data is highly correlated with the commonly dangerous crimes, arsons in forests, physical injuries, arsons and physical injuries of a person by car. All of these offenses have a very fluctuating pattern, with evidence of seasonality. Even arsons and arsons in forests, which have decreased the last years, have a periodical pattern in a monthly base, with higher frequencies for the months June to October. Moreover, it is negatively correlated with offenses that appear to have no seasonality, such as offenses against property, larceny and robbery. Thus, the second component of the monthly data might be interpreted as an index of periodicity in crime activity.

The third component of the monthly data has the largest positive correlations with arsons, commonly dangerous crimes and arsons in forests, whereas the largest negative correlations appear with offenses concerning antiquities, cases of one taking the law into one's hand and rape. However, it is difficult to find an interpretation of this component from these correlations.

The interpretation of the other components is even more difficult and it will not be attempted.

Clustering methods can also be employed in order to classify the above-mentioned offenses according to their behavior over the giving time period. In this paragraph we will only give some hierarchical clustering methods (see Everitt, 1993).

In order to determine the most appropriate number of clusters for the data there are several stopping rules (objective criteria), which indicate where to 'cut' the dendrogram (see Everitt, 1993 and Jobson, 1992). For reasons of simplicity, since these are only preliminary analyses, we will not get deeper into the selection criteria. In order to obtain the appropriate number of clusters we will use the scree-test criterion. A red line on the following dendrograms indicates where one should 'cut' the dendrograms in order to get the most representative partition for the data.

The methods that will be presented in this paragraph are the Complete Linkage, the Centroid, and the Ward's method. The dendrograms obtained for the annual data are given in figures 16, 17 and 18 respectively.

Figure 16: Dendrogram of annual data using Complete Linkage method
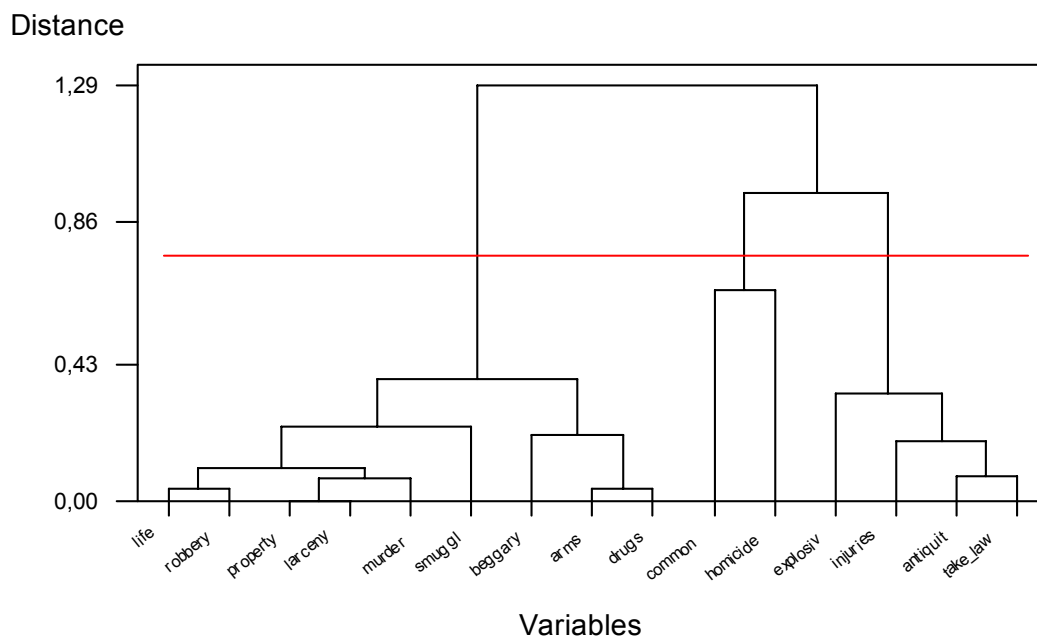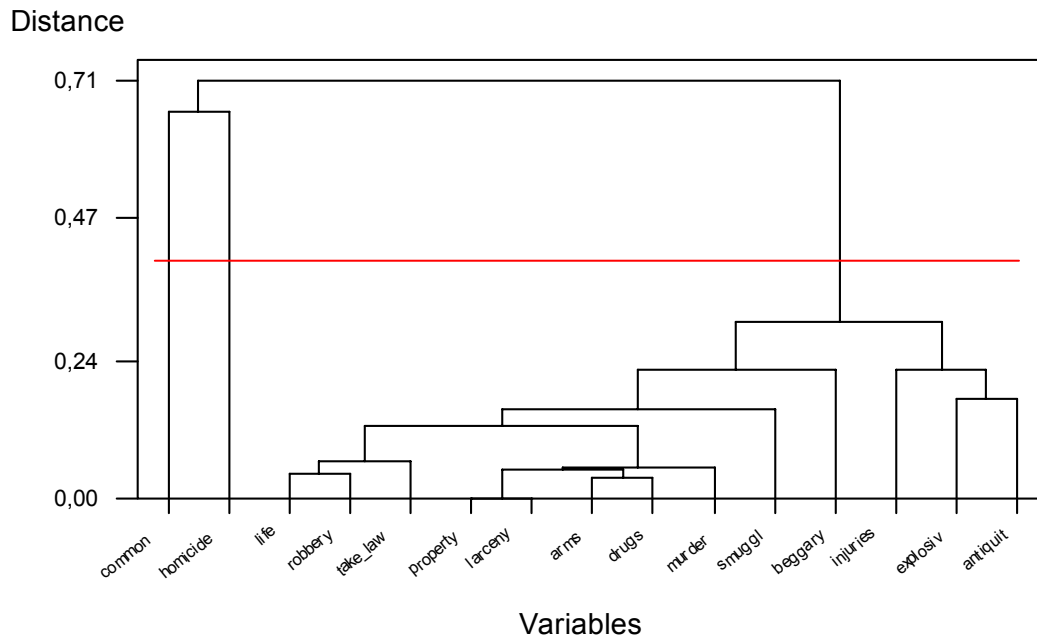

Using Complete Linkage a three-cluster solution seems to be more appropriate as the corresponding scree plot indicates.

The first cluster consists of commonly dangerous crimes and homicides by misadventure. The second cluster consists of offenses against life, offenses against property, beggary and vagrancy, illegal possession and usage of arms, drug offenses, smuggling offenses, murder, larceny and robbery. The third

cluster consists of physical injuries, illegal possession and usage of explosives, offenses concerning antiquities and the cases of one taking the law into one's hand.

Figure 17: Dendrogram of annual data using Centroid method

Using Centroid Method we again obtain a three-cluster solution as it derives from the corresponding scree plot.

The first cluster consists of commonly dangerous crimes. The second cluster consists of homicides by misadventure. The third cluster consists of offenses against life, offenses against property, beggary and vagrancy, illegal possession and usage of arms, drug offenses, smuggling offenses, murder, larceny, robbery, physical injuries, illegal possession and usage of explosives, offenses concerning antiquities and the cases of one taking the law into one's hand.

Using Ward's Method a three-cluster solution seems to be more appropriate.

The first cluster consists of commonly dangerous crimes and homicides by misadventure. The second cluster consists of offenses against life, offenses against property, beggary and vagrancy, illegal possession and usage of arms, drug offenses, smuggling offenses, murder, larceny and robbery. The third

cluster consists of physical injuries, illegal possession and usage of explosives, offenses concerning antiquities and the cases of one taking the law into one's hand.

Figure 18: Dendrogram of annual data using Ward's method
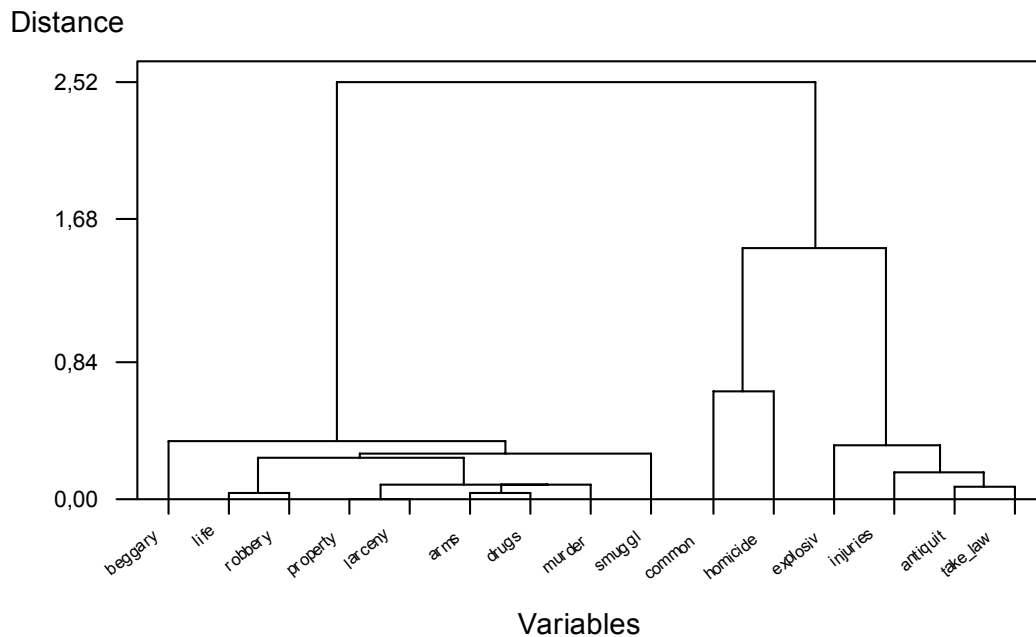
The Complete Linkage and the Ward's methods led to exactly the same clustering of the annual offenses.

Further analyses are yet to be employed in order to obtain a complete picture of the crime activity in Greece. In this report we only commented on some analyses employed to crime counts per year. Employment of such analyses to crime counts per location will determine the relationships between different types of offenses in certain counties of Greece.

Conclusion

Crime data should be confronted with caution. The more detailed the data are, the more analyses can be employed, whose results will help to combat crime. This is only an effort to illustrate that cooperation of the police with statisticians could result in a detailed picture of crime activity in Greece, which would lead to better prevention of crime.

References

Ahamad, B. (1967). An analysis of crimes by the method of principal components. *Applied Statistics,* 16, pp17-35.

Borowick, K. S. (1997). Analysis of low probability count data with applications in Crime Analysis. A dissertation submitted to the *Baylor University,* Waco, Texas.

Chatfield, C. and Collins, A. J. (1980). **Introduction to Multivariate Analysis**. Chapman & Hall.

Everitt, B. S. (1993). **Cluster Analysis**. Halsted Press, New York.

Jobson, J. D. (1992). **Applied Multivariate Data Analysis, Volume II: Categorical and Multivariate Methods.** Springer-Verlag, New York.