

Análise multivariada sobre bases de dados criminais

Simone Catarina Rodrigues Ferreira

Janeiro 2012

Dissertação submetida à Faculdade de Ciências e Tecnologia da Universidade de Coimbra
para apreciação nas provas de Mestrado em Química Forense.

A meus pais

Agradecimentos

Não encontro melhor forma de prestar homenagem a todas as pessoas que contribuíram para o culminar desta etapa, do que iniciar este espaço a citar Fernando Pessoa “O valor das coisas não está no tempo em que elas duram, mas na intensidade com que acontecem. Por isso existem momentos inesquecíveis, coisas inexplicáveis e pessoas incomparáveis”.

Cada aprendizagem é marcada por estágios e esta só pode ser realizada com o reconhecimento dos ensinamentos dos orientadores e das pessoas com quem trabalhamos e convivemos. Quero mostrar a todos a minha gratidão, pelo apoio e dedicação com o qual fui contemplada ao longo deste trabalho.

Ao Professor Doutor Alberto António Caria Canelas Pais, pela orientação, disponibilidade e paciência, pelos ensinamentos e entusiasmo que me transmitiu; para quem aprender é a única coisa de que a mente nunca se cansa, nunca tem medo e nunca se arrepende.

Ao Departamento de Química, reconhecido pelo exemplo de competência que dá prestígio a esta Faculdade.

A todos os colegas e amigos de curso, obrigado pela cumplicidade, paciência e ajuda que partilhámos. A amizade deixa-nos caminhar à vontade lado a lado, estimula e partilha a nossa maneira de olhar a vida.

À minha família, especialmente aos meus pais que sempre me apoiaram, acreditaram em mim e respeitaram as minhas escolhas.

"Conhecer não é demonstrar nem explicar,
é aceder à visão."

Antoine de Saint Exupery

Resumo

Este trabalho assenta na aplicação de métodos quimiométricos no tratamento de informação multivariada relativa a estatísticas criminais.

Uma das principais contribuições das técnicas de tratamento de informação é claramente auxiliar a interpretação e racionalização dos dados o que passa, em grande medida, pela detecção dos padrões subjacentes. Só quando estes são conhecidos se podem criar modelos de previsão e desenvolver políticas de prevenção e controlo.

A análise multivariada é incontornável quando se está perante informação proveniente de inúmeros objectos, caracterizados por diversas variáveis. Na maioria dos casos, a inspecção do perfil global dos dados é claramente mais informativa que as avaliações parâmetro a parâmetro.

Neste trabalho abordaremos uma variedade de dados relativos a detenções e a delitos cometidos nos Estados Unidos da América durante o período de um ano. Os resultados mostram que três métodos numéricos, a análise de agrupamentos (HCA), a análise de componentes principais (PCA) e o método dos mínimos quadrados não lineares, permitem tirar conclusões sobre a forma como os vários tipos de crime se distribuem e se relacionam. Além disso, servem como base para futuros desenvolvimentos na criação de modelos e análise de factores, também aqui sugeridos.

Especificamente, verificámos que crimes relacionados com a droga desempenham um papel fulcral na distinção entre os diversos estados americanos. Verificámos também que existe uma variação gradual das características criminais desde os estados rurais do centro Norte até à periferia mais rica. Alguma correlação com outros indicadores, nomeadamente o número de habitantes e a riqueza produzida, foram também encontrados. Finalmente sugerimos, tanto quanto é do nosso conhecimento pela primeira vez, uma forma funcional para a dependência da taxa de criminalidade na população das diversas povoações que constituem os Estados Unidos da América.

Abstract

This work shows the usefulness of chemometrics methods in the treatment of multivariate information on crime statistics.

A major strength of these techniques, directed to the processing of information, is that they allow the interpretation and rationalization of the data, which is made, in large part, by the detection of the underlying patterns. Only when these are known, it is possible to create forecasting models and develop prevention and control policies.

Multivariate analysis is compelling when we are faced with information from many objects, characterized by several variables. In most cases, the inspection of the global profile of the data is clearly more informative than the evaluation made parameter by parameter.

In this work we address a variety of data on arrests and offences committed in the United States of America during the period of one year (2009).

The results show that the three multivariate methods, hierarchical cluster analysis (HCA), principal component analysis (PCA) and the method of non-linear least squares (NLLS), allow us to draw conclusions about how the various types of crime are distributed and how they relate. In addition, the analysis serves as a basis for future developments in modelling and factors analysis, also suggested here.

Specifically, we found that drug-related crimes play a central role in the distinction between the various states. We also found that there is a gradual variation on the criminal characteristics from rural states from the centre north to the richer, more populated, periphery. We also found some correlation with other indicators, including the number of inhabitants and the wealth produced. Finally we suggest, to the best of our knowledge for the first time, a functional form for the dependence of the crime rate in the population of the different communities and cities that constitute the United States of America.

Abreviaturas

1. Siglas

BJS - Bureau of Justice Statistics

FBI - Federal Bureau of Investigation

HCA – Hierarchical Cluster Analysis

IACP - Associação Internacional dos Chefes de Polícia

NCVS - National Crime Victimization Survey

NIBRS - Reporting System Incident Based

NSA - National Sheriffs' Association

PC – Principal Component

PCA – Principal Component Analysis

UCR - Uniform Crime Reporting

2. Designação dos estados

Alabama	AL	Nebraska	NE
Alaska	AK	Nevada	NV
Arizona	AZ	New Hampshire	NH
Arkansas	AR	New Jersey	NJ
California	CA	New Mexico	NM
Colorado	CO	New York	NY
Connecticut	CT	North Carolina	NC
Delaware	DE	North Dakota	ND
Georgia	GA	Ohio	OH
Hawaii	HI	Oklahoma	OK
Idaho	ID	Oregon	OR
Illinois	IL	Pennsylvania	PA
Indiana	IN	Rhode Island	RI
Iowa	IA	South Carolina	SC
Kansas	KS	South Dakota	SD
Kentucky	KY	Tennessee	TN
Louisiana	LA	Texas	TX
Maine	ME	Utah	UT
Maryland	MD	Vermont	VT
Massachusetts	MA	Virginia	VA
Michigan	MI	Washington	WA
Minnesota	MN	West Virginia	WV
Mississippi	MS	Wisconsin	WI
Missouri	MO	Wyoming	WY
Montana	MT		

Prefácio

Nesta dissertação iremos utilizar técnicas numéricas e de análise de dados, com utilização muito frequente em química e normalmente associadas à área da Quimiometria, para proceder à análise de uma base de dados disponibilizada pelo *Federal Bureau of Investigation* (FBI) dentro do programa *Uniform Crime Reporting* (UCR).

A utilização deste tipo de técnicas é extremamente frequente na Química Forense e pode ser associada quer aos passos mais usuais da química analítica, quer a problemas mais específicos relacionados com caracterização de materiais apreendidos, detecção de falsificações, e muitos, muitos outros.

Verificamos que, na análise a que iremos proceder, estamos perante um problema ou problemas que podem ser designados como clássicos dentro das aproximações quimiométricas. De facto a base de dados é multivariada e os objectos, na acepção usual, são povoações ou agregados de povoações. As variáveis têm a ver com a incidência de vários tipos de crime. Trata-se de uma informação extensa mas perfeitamente tratável com métodos comuns, dos quais destacaremos aqui a análise hierárquica de agrupamentos e a análise de componentes principais. Procederemos, também, à análise de factores.

A dissertação encontra-se dividida em duas grandes partes. Uma primeira, que designaremos por caracterização geográfica, e uma segunda, de menor extensão, e mais específica em que procuraremos a relação entre população e taxa de criminalidade. A primeira permitir-nos-á estabelecer os padrões que caracterizam os Estados Unidos da América (EUA) em termos dos crimes cometidos estado a estado. Para tal, introduziremos vários perfis de caracterização e tentaremos estabelecer comportamentos comuns, divergentes e identificar as variáveis caracterizadoras. A segunda parte, na nossa opinião igualmente importante, faz a busca de uma relação que se tem mostrado extremamente esquiva.

A estrutura da dissertação, no seu corpo principal, é a seguinte: no primeiro capítulo, de carácter introdutório, faremos uma revisão da literatura no que respeita aos aspectos geográficos da caracterização criminal e dos factores subjacentes. No segundo capítulo será apresentada a base de dados UCR e, no capítulo seguinte serão descritas as técnicas e os métodos utilizados. O capítulo IV apresentará os resultados referentes à análise das detenções efectuadas em 2009, por estado. No capítulo V, abordaremos o contexto segundo o qual deve ser encarada a relação taxa de criminalidade vs. população e apresentaremos as nossas sugestões para essa questão. Finalmente, o capítulo VI reunirá alguns comentários.

Nesta dissertação, e com o propósito de evitar ambiguidades, foram mantidas várias designações no original anglo-saxónico. De entre estas, salientem-se as designações dos diversos tipos de crime, estados e organizações.

Índice

1. Padrões criminais: uma revisão	1
1.1. Aspectos geográficos	1
1.2. Factores	3
2. A base de dados <i>Uniform Crime Reporting</i>	7
2.1. O programa	7
2.1.1. Contexto histórico	8
2.1.2. Evolução	9
2.1.3. Definição dos delitos	9
2.1.4. Organização dos dados: alguns aspectos	10
2.2. Avaliação da fiabilidade	12
3. Técnicas e métodos	15
3.1. Considerações gerais	15
3.2. Quimiometria	16
3.3. Análise de agrupamentos	17
3.3.1. Procedimento hierárquico	19
3.3.2. Semelhança e distância	19
3.3.3. Ligação	21
3.4. Análise de componentes principais	23
3.4.1. A transformação	24
3.4.2. Obtenção dos componentes	24
3.4.3. Redução de dimensionalidade	26
3.4.4. Representações gráficas	28
3.5. Mínimos quadrados não lineares	29
3.5.1. Algoritmo Levenberg - Marquardt	30
4. As detenções em 2009	33
4.1. Os dados em bruto	33
4.1.1. HCA por tipo de crime	33
4.1.2. Análise de componentes principais	40
4.2. Aumentando a especificidade	45
4.2.1. Frequência	45

4.2.2. Taxa de criminalidade	46
4.2.3. Fracção.....	48
4.3. Aumentando a gravidade.....	50
4.3.1. Taxa de criminalidade	50
4.3.2. Fracção.....	51
4.3.3. Cálculo alternativo.....	52
4.4. Factores	54
4.5. Perspectiva geral e discussão	55
5. Taxa de criminalidade e população.....	61
5.1. Contexto.....	61
5.2. Construindo o perfil: da nação ao estado	63
5.3. A dependência funcional	68
5.4. Discussão	70
6. Comentários finais	71
Referências bibliográficas	73
Anexo	83

Índice

1. Padrões criminais: uma revisão	1
1.1. Aspectos geográficos	1
1.2. Factores	3
2. A base de dados <i>Uniform Crime Reporting</i>	7
2.1. O programa	7
2.1.1. Contexto histórico	8
2.1.2. Evolução	9
2.1.3. Definição dos delitos	9
2.1.4. Organização dos dados: alguns aspectos	10
2.2. Avaliação da fiabilidade	12
3. Técnicas e métodos	15
3.1. Considerações gerais	15
3.2. Quimiometria	16
3.3. Análise de agrupamentos	17
3.3.1. Procedimento hierárquico	19
3.3.2. Semelhança e distância	19
3.3.3. Ligação	21
3.4. Análise de componentes principais	23
3.4.1. A transformação	24
3.4.2. Obtenção dos componentes	24
3.4.3. Redução de dimensionalidade	26
3.4.4. Representações gráficas	28
3.5. Mínimos quadrados não lineares	29
3.5.1. Algoritmo Levenberg - Marquardt	30
4. As detenções em 2009	33
4.1. Os dados em bruto	33
4.1.1. HCA por tipo de crime	33
4.1.2. Análise de componentes principais	40
4.2. Aumentando a especificidade	45
4.2.1. Frequência	45

4.2.2. Taxa de criminalidade	46
4.2.3. Fracção.....	48
4.3. Aumentando a gravidade.....	50
4.3.1. Taxa de criminalidade	50
4.3.2. Fracção.....	51
4.3.3. Cálculo alternativo.....	52
4.4. Factores	54
4.5. Perspectiva geral e discussão	55
5. Taxa de criminalidade e população.....	61
5.1. Contexto.....	61
5.2. Construindo o perfil: da nação ao estado	63
5.3. A dependência funcional	68
5.4. Discussão	70
6. Comentários finais	71
Referências bibliográficas	73
Anexo	78

Capítulo 1

Padrões criminais: uma revisão

1.1. Aspectos geográficos

Nesta dissertação iremos proceder a uma caracterização dos padrões criminais dos Estados Unidos da América, recorrendo a dados do programa *Uniform Crime Reporting* do FBI, que será abordado em detalhe no próximo capítulo. Um dos aspectos que torna relevante este tipo de caracterização tem a ver com uma avaliação do custo do crime para a sociedade. Este custo encontra-se na base do desenvolvimento de muitos programas sociais relacionados, por exemplo com o tratamento da dependência ou com o policiamento [1]. Note-se que o custo pode, muitas vezes, ser indirecto. Certas comunidades vêem extremamente reduzida a actividade económica devido à existência de crime violento. No entanto, parece estar demonstrado que diversas áreas de actividade económica podem ter uma sensibilidade diferente ao crime [2].

Outra preocupação tem sido a correspondência entre os índices de criminalidade a nível de cidades e a nível nacional. Os resultados são consistentes com a existência de um padrão, claro e único para os grandes centros urbanos. Esta observação apoia a ideia de que existe uma tendência nacional significativa e sugere a conveniência da continuação do seu estudo. A pergunta permanece “O crime, tal como a política, é uma preocupação local, ou as taxas de criminalidade seguem um padrão nacional mais alargado?” [3]. A realização de uma análise geográfica é essencial, também, para avaliar o impacto de políticas utilizadas quer a nível local, quer a nível global. Um exemplo de uma tal aplicação tem a ver com o chamado *Project Safe Neighborhoods* que motivou um estudo sobre as tendências do crime violento em todas as cidades dos Estados Unidos acima de 100 000 habitantes. A obtenção de uma relação causal não é, no entanto, uma tarefa fácil [4]. Note-se que a caracterização geográfica tem surgido, recentemente, a par com a análise de padrões criminais [5] e permite também, por vezes, a identificação de estruturas espaciais dinâmicas numa análise socioeconómica de crime ou crimes [6].

Os criminologistas sempre tentaram explicar porque é que certos tipos de crime, ou diferentes níveis de criminalidade, são encontrados de forma distinta nas várias comunidades. Apesar da nossa abordagem ter um enfoque mais lato, tem havido alguma tendência, no passado, para reconhecer que as grandes comunidades têm sido o foco primário da teoria e investigação da prevenção criminal. Existe um reconhecimento

crescente da importância de transferir essa atenção para as pequenas comunidades, nas quais os atributos do lugar e as actividades de rotina se combinam para o desenvolvimento de eventos criminais [7]. Daí que certas análises já sejam feitas rua a rua. Torna-se, no entanto, difícil obter padrões coerentes [8]. Tal não impede que a análise seja realizada sobre amostras ainda menores, incluindo não mais do que quarenta residentes, recorrendo a técnicas numéricas sofisticadas [9].

Tem sido também referido que uma avaliação à escala global, dos padrões do crime tem de ser conduzida antes de se tentar focar em zonas específicas, por mais significativas que essas possam ser [10]. Na mesma linha tem surgido algumas preocupações sobre modelos baseados em variáveis reduzidas, que impedem uma avaliação de escala [11]. Em certos casos, o estudo é feito directamente sobre amostras consideradas representativas. Estas amostras podem ser de dimensões razoáveis e sobre elas procede-se, muitas vezes, a um estudo de factores determinantes como seja, por exemplo, o uso de droga [12]. Noutros casos, a amostra é consideravelmente mais reduzida [13]. Naturalmente que, nestes casos, podem ser detectadas variações espaciais de grande amplitude [14], o que leva à necessidade da identificação de microestruturas [15]. Note-se que tais aproximações contrastam com outras em que o foco está na interacção inter-grupos o que determina, em muitos casos, um aumento da zona ou da amostra sob estudo [16]. Tem-se também constatado, recorrendo a argumentos de alguma complexidade, baseados em influências temporárias e permanentes, diferenças nos estudos inter e intra comunidades [17]. Refira-se que a composição da amostra também tem sido uma preocupação dos investigadores, nomeadamente, quando se pretendem analisar factores como a composição racial ou o desemprego. Inconsistências que surjam na amostragem podem produzir falta de consenso em estudos análogos, impedindo o estabelecimento de conclusões gerais [18].

Tem-se também tentado estabelecer uma relação entre características do meio envolvente e o crime. A modelação passa pela utilização de características relacionadas com o nível socioeconómico, a mobilidade residencial, a heterogeneidade racial e a estabilidade familiar [19]. A análise é complexa, assenta num modelo de desorganização social e raras vezes é totalmente conclusiva. O conceito de desorganização social tem também sido combinado com outros aspectos como sejam por exemplo a evolução económica. Uma das conclusões é que o decréscimo de empregos nas indústrias produtivas aumenta as taxas de *aggravated assault*, *larceny-theft* e *burglary* [20]. A complexidade de análises efectuadas com base na desorganização social é bem patente na referência [21].

Um dos aspectos importantes relativos à identificação dos padrões relacionados com o crime tem precisamente a ver com a escolha daqueles que promovem uma mais directa

caracterização desses padrões e com as suas inter-relações. Certos estudos seminais apresentam, por exemplo, relações nulas entre os crimes contra as pessoas, ao mesmo tempo que invocam uma relação mais directa entre os crimes contra a propriedade [22]. Tal significa, também, que uma análise temporal tem de assentar numa avaliação detalhada das variáveis sobre as quais se faz a descrição. Por outras palavras, não é possível encarar a evolução temporal, que estará fora do âmbito deste trabalho, sem analisar em profundidade cada um dos pontos temporais.

Naturalmente que o recurso a métodos computacionais é generalizado. Uma faceta em que os computadores são praticamente indispensáveis está relacionada com a visualização dos padrões espaço-temporais do crime. Uma revisão comparativa de técnicas de visualização está patente na referência [23]. Também têm sido analisadas as características óptimas que o software de análise de crimes deve possuir, nomeadamente para o desenho de estratégias preventivas e de detecção [24].

A análise espacial, à qual daremos especial ênfase nesta dissertação, tem frequentemente como objectivo a previsão de eventos criminais [25-26]. Também têm sido desenvolvidos esforços no sentido da previsão geo-temporal, recorrendo por exemplo a redes neuronais artificiais [27]. Esta análise espacial do crime é bastante semelhante a uma análise epidemiológica e daí que na primeira se recorra frequentemente a métodos empregados na segunda [28].

Técnicas como as que serão utilizadas no nosso trabalho podem ter um foco diferente, e o *data mining* pode ser feito sobre as metodologias utilizadas na perpetração do crime [29].

1.2. Factores

Têm sido extremamente variados os factores escolhidos para a relação, por exemplo, com a taxa de criminalidade. Uma escolha recorrente tem a ver com a estrutura de idades, já que o estabelecimento de uma associação permite a previsão das tendências do crime com base em dados demográficos. Refira-se, como exemplo, que as detenções ocorrem primordialmente sobre adolescentes e jovens adultos. Embora não seja explicitamente utilizada essa informação no presente trabalho, dados relativos à idade constam da informação disponibilizada na UCR. Note-se, no entanto, que esta previsão com base demográfica, embora potencialmente muito atraente, pode ser vista como não adequada. Com efeito, a relação idade/crime surge muitas vezes como mais fraca do que geralmente se considera, dado que este factor é facilmente diluído pelos restantes [30]. Quando se considera robusta a ligação entre idade e crime, pode tentar-se explicar a

etiologia desta relação [31]. Indo mais longe, alguns desenvolvimentos têm tentado estabelecer relações cruzadas entre, por exemplo, a idade e a influência dos pares. Novamente os resultados não são totalmente conclusivos [32].

Outros aspectos demográficos que têm sido examinados são, por exemplo, a densidade populacional ou o número de agentes policiais *per capita*. Enquanto o primeiro parece estar relacionado com o crime contra a propriedade, as relações com o segundo não são tão claras, parecendo haver alguma especificidade no tipo de crime para estabelecimento da relação. De resto, desde os últimos 30 anos que se tem verificado que a identificação de factores é muitas vezes inconclusiva ou só parcialmente conclusiva. [33].

Tem também sido frequentemente objecto de estudo uma avaliação de factores relacionados com a origem étnica ou a composição racial, em paralelo ou não, com aspectos relacionados com a imigração [6].

Também o género tem sido utilizado como base para avaliação de outros factores, nomeadamente estruturais e de relação (casamento, emprego, educação, pobreza). Cite-se, a nível de exemplo, que *larceny-theft* é o crime contra a propriedade mais comum para ambos os sexos, seguido de *burglary* [34]. Noutros casos, avalia-se mesmo um conjunto de factores sócio-demográficos em conjunto com uma separação em género [35]. Realce-se não ter sido possível identificar, nos detidos, diferenças entre os géneros no que diz respeito a idade, nível de educação e fontes de rendimento. Finalmente certos estudos incidem num só género mas as análises são efectuadas a nível nacional [36].

A análise de factores pode ir ainda mais longe e incluir aspectos como o autocontrolo e, talvez mais difícil ainda, a sua previsão [37]. Note-se que este mesmo conceito de auto controlo tem sido utilizado em relações cruzadas com comportamento juvenil violento [38]. Na mesma linha já foi feita a associação entre coeficiente de inteligência e envolvimento em actividades criminais ou de delinquência [39-40]. Vários outros aspectos têm sido examinados em contextos semelhantes [34].

Naturalmente que os factores económicos não tem sido descurados, nomeadamente na sua relação com crimes sobre a propriedade. Uma análise recorrente tem incidido sobre o desemprego. A relação com o desemprego parece, no entanto, ser ainda um problema longe de uma solução clara. Refira-se que um trabalho exigente sobre esta relação exige simultaneamente uma boa medida do desemprego e da criminalidade. Parece estar estabelecido que diversos modos de medir o primeiro afectam o padrão com que surge o segundo [41]. Tem também sido estudada a influência da desigualdade nos rendimentos sobre a taxa de criminalidade. Novamente o resultado é por vezes ambíguo [42], mostrando-se que esta desigualdade está positivamente associada com a taxa de criminalidade na

análise para um determinado ano mas negativamente na evolução temporal. Noutros casos os resultados parecem ser mais concludentes [43]. Na mesma linha, e para crimes específicos, verificou-se que a desigualdade tem efeitos significativos quer em *burglary* quer em *robbery* [44]. Outros indicadores, como o designado índice de miséria foram também utilizados para descrever a relação com a taxa de criminalidade, verificando-se uma relação positiva [45]. Refira-se também que, desde há muito, têm sido desenvolvidos esforços para a obtenção de modelos globais descrevendo a relação entre crime, dissuasão e variáveis sócio económicas [46-47]. Os factores económicos, sociais e políticos surgem frequentemente em estudos onde se analisa uma evolução temporal, por vezes ao longo de um largo período de tempo [48-49].

Outro factor que tem sido analisado é o do nível de penas suspensas (*probation*). Nos Estados Unidos, em 2004, o número de penas suspensas correspondia ao dobro do número de sentenças de prisões efectivas [50]. Novamente aqui parece surgir uma relação positiva entre a liberdade condicional e a taxa de criminalidade, embora não se possa concluir que um aumento no grau de supervisão tenha um efeito benéfico na dissuasão do crime.

Muitas das análises que são actualmente apresentadas focam factores específicos sobre os quais se tenta estabelecer uma relação causal. Um dos exemplos é o exame das relações recíprocas entre crime violento e a estabilidade residencial nas comunidades. Verifica-se que certas comunidades, mais estáveis, parecem exibir um efeito protector contra a violência [51]. As relações entre vários tipos de crime têm também surgido frequentemente na literatura. Cite-se, por exemplo, a relação entre álcool, drogas e crime violento [52] que surge, também, na nossa análise (veja-se capítulo 4. Outro exemplo é a relação entre álcool e *rape* [53].

Tem também sido sugeridas aproximações globais, nomeadamente sob a égide da chamada ciência social evolucionária. Esta analisa factores tão diversos como o género, a poligamia ou a educação na relação com o crime [54].

A análise têm também incidido sobre alguns factores menos óbvios, por exemplo, de carácter ambiental, como sejam a relação entre perfis criminais e exposição ao chumbo [55] e a relação entre o aquecimento global e o crime [56].

O facto de que, como foi referido anteriormente, muitas das relações se apresentarem como menos consequentes faz com que, em trabalho muito recente, se tenha procedido a uma re-identificação de indicadores, especialmente aqueles que apresentam capacidade preditiva [57].

Capítulo 2

A base de dados *Uniform Crime Reporting*

O Departamento de Justiça dos Estados Unidos da América administra dois programas estatísticos para medir a magnitude, a natureza e o impacto da criminalidade no país: o UCR (*Uniform Crime Reporting*) e o NCVS (*National Crime Victimization Survey*). Cada um destes programas produz informações valiosas sobre os vários aspectos do crime no país.

Neste trabalho, como é de resto frequente, apoiar-nos-emos exclusivamente nos dados e procedimentos do programa UCR [58], que passamos a descrever.

2.1. O programa

A Associação Internacional dos Chefes de Polícia (IACP) iniciou a publicação de informações para o programa UCR em 1929, apenas dois anos após a IACP ter estabelecido uma comissão para investigar a uniformidade das estatísticas criminais. Desde Setembro de 1930, o *Federal Bureau of Investigation* (FBI) [59]¹, tem administrado o programa e, actualmente, recolhe informação sobre os seguintes crimes: *murder and nonnegligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny-theft, motor vehicle theft, e arson*. Os organismos de aplicação da lei comunicam os dados relativos às detenções para vinte e uma categorias de crime adicionais.

O programa UCR compila dados de relatórios mensais provenientes destes organismos ou registos de incidentes individuais transmitidos directamente para o FBI ou para agências estaduais centralizadas que reportam ao FBI. O programa examina exhaustivamente cada relatório que recebe para verificação da verosimilhança, exactidão e dos desvios que podem indicar erros. Grandes variações nos níveis de criminalidade podem indicar a existência de procedimentos de registos modificados, relatórios incompletos ou mudanças nos limites de uma jurisdição. Para que possam ser identificadas eventuais flutuações incomuns na contagem de uma agência, o programa compara os relatórios mensais com submissões anteriores da agência e com os de agências similares.

Este programa apresenta a contagem de crimes para o país como um todo, assim como para as regiões, estados, municípios, cidades, vilas, zonas sob a acção da lei tribal e

¹ Veja-se a referência que apresenta alguns dados interessantes sobre esta organização

universidades. Esta vasta informação permite estudos entre as jurisdições vizinhas e entre aquelas com populações semelhantes e outras características comuns.

O programa UCR do FBI é pois um esforço estatístico, a nível nacional, que envolve a cooperação de quase 18 000 cidades, universidades, municípios, estados, organismos tribais e agências policiais federais que reportam voluntariamente os crimes que são trazidos à sua atenção.

Em 2009, as agências para aplicação da lei, activas no programa UCR, representaram mais de 295 milhões de habitantes dos Estados Unidos da América (96.3 % da população total). A cobertura representou 97.1% da população na estatística das áreas metropolitanas, 90.9% da população de cidades fora das áreas metropolitanas e 93.0% da população de municípios não metropolitanos.

2.1.1. Contexto histórico

Tendo reconhecido a necessidade de realizar estatísticas da criminalidade nacional, a IACP formou a comissão dos *Uniform Crime Records*, na década de 1920, para desenvolver um sistema uniforme de estatística criminal. Depois de estudar os códigos penais dos estados e realizar uma avaliação da prática dos registos em uso, a comissão, em 1929, elaborou um plano para relatar o crime que se tornou a base do programa UCR. O plano inclui definições padrão de ofensas para sete classificações de infracções principais, conhecidas como crimes de parte I, com o objectivo de medir as flutuações no volume global e na taxa de crime. Os responsáveis por tal desenvolvimento também instituíram a regra da hierarquia como principal procedimento de notificação para o que hoje é conhecido como o resumo.

A classificação das sete infracções de parte I incluem os crimes violentos de *murder and nonnegligent manslaughter, forcible rape, robbery e aggravated assault* e os crimes contra a propriedade de *burglary, larceny-theft, e motor vehicle theft*. Por ordem do Congresso, em 1979, o crime *arson* foi adicionado como o oitavo a fazer parte das infracções da categoria parte I.

Em Janeiro de 1930, quatrocentas cidades, que representam vinte milhões de habitantes, distribuídos ao longo de quarenta e três estados, começaram a sua participação no programa UCR. Nesse mesmo ano, o congresso aprovou uma autorização para que o Procurador-geral possa recolher informação criminal. O Procurador-Geral, por sua vez, designou o FBI para servir de base de recolha para todos os dados criminais colectados. Todos os anos, desde então, os dados baseados nas classificações uniformizadas e

procedimentos para reportar ofensas e detenções bem definidos, passaram a ser obtidos a partir das agências de aplicação da lei a nível nacional.

2.1.2. Evolução

Embora a recolha de dados tenha permanecido praticamente inalterados ao longo dos anos, na década de 1980 o programa tinha evoluído para uma mais ampla aplicação. Reconhecendo a necessidade de melhoramento da estatística, os organismos de aplicação da lei procederam a um esforço de modernização. O FBI forneceu o seu apoio formulando uma reestruturação do crime em 3 fases. Em primeiro lugar, as agências passariam a usar um sistema baseado no relato de incidentes para reportar crimes e detenções, *Reporting System Incident Based* (NIBRS). Em segundo, o programa nacional UCR passaria a recolher informação a dois níveis (*limited and full participation*) e em terceiro o UCR iria introduzir um programa de garantia da qualidade.

No final da década de 1980, o FBI terminou o desenvolvimento e assumiu a gestão do programa nacional do UCR, NIBRS. Começou a receber, a partir de um número reduzido de agências, os dados provenientes deste sistema em Janeiro de 1989. À medida que as agências de aplicação da lei passaram a estar informadas acerca das vantagens inerentes a esta forma de recolha de dados começaram a aderir ao sistema. Tendo como base as submissões de informação de 2009 concluímos que 44 % das agências participaram sob este sistema.

Considera-se que a inclusão do NIBRS fez desaparecer uma limitação básica do UCR, tradicional, permitindo recolher informação proveniente das vítimas. Note-se que isto permitiu também estabelecer o paralelo entre o NIBRS e o NCVS [60].

2.1.3. Definição dos delitos

O programa do *Uniform Crime Reporting* (UCR) divide as ofensas em dois grupos, crimes de parte I e parte II. Em cada mês, os organismos de aplicação da lei, que participam no programa, submetem a informação dos crimes de parte I dos quais tomaram conhecimento. Para os restantes crimes, parte II, são apenas fornecidas informações acerca das detenções efectuadas.

O programa UCR recolhe os dados sobre os crimes de parte I, com o objectivo de medir o nível e o âmbito dos crimes que ocorrem no país. Os fundadores deste programa escolheram estes crimes por serem os mais graves, por ocorrerem com regularidade em todos os estados e por serem susceptíveis de ser comunicados à polícia.

As infracções de parte I e parte II estão compiladas e definidas no anexo.

2.1.4. Organização dos dados: alguns aspectos

Neste trabalho recorreremos essencialmente a duas tabelas. Uma delas corresponde às detenções efectuadas por cada crime, em cada estado, durante o período de um ano. A segunda corresponde ao número de infracções de que os organismos, participantes no programa, tomaram conhecimento.

Comecemos pela descrição da primeira tabela.

Arrests, by State, 2009 (Table 69) - Estes dados correspondem a 29 tipos de crimes. A tabela fornece os dados relativos às detenções efectuadas em cada Estado, durante o ano de 2009. Discrimina o número total de detenções e as que foram realizadas com pessoas de idade inferiores a 18 anos.

Estes dados representam o número de pessoas detidas, mas deve ser tida em consideração a possibilidade de uma pessoa ser detida mais do que uma vez ao longo do ano. Em alguns casos, o valor presente nesta tabela poderá pois, representar múltiplas detenções da mesma pessoa.

Os dados são disponibilizados sob a forma de folha de cálculo e têm a aparência indicada no fragmento constante da Figura 1.

	A	B	C	D	E	F	G	H	I
1	Table 69								
2	Arrests								
3	by State, 2009								
4	State		Total all classes ¹	Violent crime ²	Property crime ²	Murder and nonnegligent manslaughter	Forcible rape	Robbery	Aggravated assault
5	ALABAMA	Under 18	11,035	620	3,735	28	20	336	2,112
6		Total all ages	193,586	5,759	24,104	292	325	1,657	3,400
7	ALASKA	Under 18	3,780	201	1,077	0	7	37	1,117
8		Total all ages	40,349	2,145	3,937	22	77	283	1,755
9	ARIZONA	Under 18	50,921	1,344	10,980	13	29	370	9,997
10		Total all ages	336,686	9,364	44,116	202	226	2,214	6,700
11	ARKANSAS	Under 18	11,745	352	3,323	5	32	63	2,112
12		Total all ages	125,958	3,988	15,785	113	240	632	3,000
13	CALIFORNIA	Under 18	203,345	15,146	43,399	183	235	6,231	8,400
14		Total all ages	1,474,004	121,369	164,744	1,811	2,044	21,577	95,000
15	COLORADO	Under 18	37,890	814	7,687	17	67	206	4,000
16		Total all ages	213,163	6,470	25,826	166	421	1,088	4,000
17	CONNECTICUT	Under 18	18,665	1,170	4,087	7	20	227	6,000

Figura 1. Representação da folha de cálculo com os dados fornecidos através do programa do *Uniform Crime Reporting* (Table 69 na notação original) e correspondente às detenções efectuadas, para 29 tipos de crime, em cada Estado, durante o ano de 2009.

Para a segunda, *table 8* na notação do programa UCR, temos de ter em conta a sua especificidade para cada estado. Apresenta-se na Figura 2 um exemplo para o estado do Alabama. Vejamos uma descrição mais detalhada.

Offenses Known to Law Enforcement, by State by City, 2009 (Table 8) - Esta tabela fornece o volume de crime violento (*murder and nonnegligent manslaughter, forcible rape, robbery, e aggravated assault*) e o de crimes contra a propriedade (*burglary, larceny-theft, e motor vehicle theft*) relatados para cada cidade do estado, pelos organismos envolvidos.

City	Population	Violent crime	Murder and nonnegligent manslaughter	Forcible rape	Robbery	Aggravated assault	Property
Abbeville	2,932	9	0	1	0	8	
Adamsville	4,704	25	0	1	17	7	
Addison	710	5	0	1	0	4	
Alabaster	30,002	41	0	2	12	27	
Albertville	20,078	82	0	11	12	59	
Alexander City	15,057	132	2	10	31	89	
Aliceville ²	2,456	22	0	0	3	19	
Andalusia	8,994	54	1	5	7	41	
Anniston	23,598	584	5	27	113	439	
Ardmore	1,275	0	0	0	0	0	
Ashford	2,087	5	0	0	0	5	
Ashland	1,873	5	0	0	0	5	
Ashville	2,599	8	0	0	0	8	
Athens	24,261	32	0	3	18	11	
Attalla	6,514	57	1	4	3	49	
Auburn ²	57,342	139	3	8	43	85	
Autaugaville	875	0	0	0	0	0	
Baker Hill	313	0	0	0	0	0	
Bear Creek	995	8	0	0	0	8	
Birmingham	1,184	0	0	0	0	0	

Figura 2. Representação da folha de cálculo com os dados fornecidos através do programa do *Uniform Crime Reporting (Table 8* na notação original) para o volume de crime violento e de crimes contra a propriedade relatados para cada cidade do estado durante o ano de 2009.

2.2. Avaliação da fiabilidade

Os dados mais recentes da UCR não são considerados como sofrendo de vários problemas que foram levantados em anos anteriores [61], muitas vezes relacionados com lapsos de informação originados por uma deficiente cobertura do território. Foi até, nalguns casos, e para colmatar tais deficiências, sugerida a utilização conjunta das duas maiores fontes de informação [62-63], UCR e NCVS. Note-se, no entanto, que divergências entre estas duas fontes foram um tema recorrente na literatura [64-65].

A informação recolhida pelo UCR não está, no entanto, totalmente isenta de críticas. Uma delas tem a ver com o facto de que os dados contêm ainda falhas significativas, com esquemas de preenchimento que são por vezes considerados inadequados e inconsistentes

[66]. Normalmente, essas inconsistências surgem ao nível do *county*, e parecem justificar o aparecimento de tendências contra intuitivas, por exemplo em estudos relacionados com o direito ao uso e porte de armas.

Mais recentemente [67], foi considerado que embora a desagregação a nível do *county* permita examinar erros de medida, considerados omnipresentes em qualquer tipo de dados, não é a existência de erros que é relevante mas até que ponto esse erro é sistemático. Não sendo fácil concluir quão afectados são os dados por estes erros, parece ser de algum consenso que não existe um problema sistemático com os dados a este nível, o que implicitamente os valida para uma agregação a nível estadual.

Capítulo 3

Técnicas e métodos

3.1. Considerações gerais

O conhecimento, na actualidade e em muitas áreas, apenas pode ser atingido após a aquisição e processamento de uma grande quantidade de informação, de diferentes origens. A maior dificuldade não reside muitas vezes na aquisição, mas na localização do fio condutor subjacente à grande quantidade de dados disponível.

A completa compreensão das relações que se estabelecem entre as diversas variáveis faz com que a análise seja complexa. Neste contexto, recorre-se genericamente à análise multivariada, que pode ser definida como o conjunto de métodos e técnicas que utilizam simultaneamente todas as variáveis na interpretação teórica dos dados [68].

O tratamento dos dados envolve a análise de um grande número de variáveis. Muitas vezes, um pequeno número destas contém os aspectos mais relevantes, enquanto as restantes pouco ou nada acrescentam à interpretação dos resultados.

A análise multivariada permite, pois, a redução de dados ou simplificação estrutural, na qual o fenómeno em estudo é representado da forma mais simples possível, sem sacrificar informação relevante, tornando as interpretações mais simples. Permite, assim, a ordenação e agrupamento de objectos² ou variáveis similares, a inspecção da sua interdependência, tarefas de previsão, e, também, a construção e teste de hipóteses [68].

Existem vários métodos de análise multivariada com finalidades distintas entre si. Quando o interesse é verificar como um conjunto de elementos se relaciona, o quanto estes são semelhantes segundo as variáveis utilizadas, destacam-se dois métodos, a análise de agrupamentos hierárquica (*HCA*³) e a análise de componentes principais (*PCA*⁴).

Assim, a redução de variáveis é realizada através de critérios objectivos, permitindo a construção de gráficos bi- ou tridimensionais contendo maior informação estatística, o que pode ser conseguido através da análise de componentes principais. Também é possível construir agrupamentos entre os objectos de acordo com a sua similaridade, utilizando todas

² Objecto é, na prática, um ponto no espaço multivariável.

³ Do inglês *Hierarchical Cluster Analysis*

⁴ Do inglês *Principal Component Analysis*

as variáveis disponíveis, e representá-los de forma bidimensional através de um dendrograma [69].

Estes métodos desempenham um papel importante no tratamento da informação, sendo utilizados em áreas muito distintas que vão das Ciências Sociais à Química, Biologia e Medicina, ou das Ciências Económicas à Engenharia. Podem ser usados em aplicações tão variadas como monitorização ambiental, arqueologia científica, ciências forenses, geoquímica e muitas outras [70].

O presente trabalho assenta na aplicação de algoritmos padrão no tratamento e interpretação da informação mais relevante, em sistemas fora da área da química. A informação a analisar é fornecida por uma base de dados que compila informação sobre vários tipos de crime que ocorrem anualmente nos Estados Unidos da América.

Especificamente, procurámos combinar duas ferramentas quimiométricas (*HCA* e *PCA*) com vista à simplificação e compressão da informação, de modo a permitir a sua caracterização extensiva e a extracção de aspectos relevantes.

3.2. Quimiometria

A Quimiometria corresponde a uma área onde são realizadas análises abrangentes com aplicação num campo interdisciplinar que combina estatística multivariada, modelação matemática, computação e química [71-72].

O termo foi proposto pela primeira vez em 1972 por Bruce Kowalski, da Universidade de Washington, e Svante Wold, da Suécia, os quais formaram dois anos depois a Sociedade Internacional de Quimiometria. São considerados os fundadores desta nova disciplina que envolve a aplicação de métodos multivariados que permitem a compreensão e visualização das relações existentes entre as variáveis e os objectos [73].

Quando se possuem medições numéricas, estas normalmente requerem a aplicação de procedimentos estatísticos básicos, para tratamento dos dados, que incluem cálculos de médias, desvios-padrão, desvios-padrão relativos e testes de significância. A criação de instrumentos analíticos mais complexos e a necessidade de lidar com um maior conjunto de dados experimentais têm exigido novas abordagens [74].

Os métodos quimiométricos possuem um enorme potencial no tratamento de diversos tipos de problemas. No entanto, até à segunda metade dos anos 80, a sua presença na literatura era escassa. A partir dos anos 90 nota-se um aumento significativo das aplicações, levando também ao aumento da capacidade dos investigadores em extrair informações dos dados [75]. Naturalmente, muito deste esforço surgiu em resposta à

necessidade de desenvolver novos métodos matemáticos e estatísticos para lidar com a vasta quantidade de dados produzida pelos instrumentos analíticos modernos [75].

Neste momento esta disciplina permite tratar dados mais complexos, que requerem a utilização de técnicas estatísticas multivariadas, álgebra matricial e análise numérica [76].

A disponibilidade de computadores mais potentes e menos dispendiosos contribuiu para a rápida evolução dos métodos quimiométricos, ao permitir uma análise mais flexível de grandes conjuntos de dados, o desenvolvimento de algoritmos computacionais mais eficientes e a difusão de software quimiométrico.

A Quimiometria envolve diferentes métodos, tais como a optimização e validação de metodologias analíticas, o planeamento experimental, a estimativa de parâmetros, o processamento de sinal, a análise de factores, a calibração multivariada, a utilização de redes neurológicas artificiais, o reconhecimento de padrões ou o processamento de imagem digital [74].

Na literatura, são publicadas regularmente revisões detalhadas sobre a aplicação dos métodos quimiométricos em várias áreas do conhecimento. É difícil descrever em detalhe cada contribuição histórica desde os primeiros anos da quimiometria. Em geral, estes métodos têm sido aplicados com sucesso na visualização dos dados, na classificação, na resolução de curvas multivariadas e na predição em química analítica, química ambiental, engenharia, investigação médica e na indústria [77-80].

Recentemente, algumas abordagens mais complexas têm sido propostas para auxiliar em estudos de desenvolvimento como a genómica, proteómica, bioinformática e a metabolómica [81-83].

Devemos, no entanto, ter em atenção o facto de que, devido à enorme diversidade de algoritmos e suas variantes, à complexidade crescente e à multiplicidade de linguagens usadas, é difícil implementar soluções coerentes, eficazes e inovadoras, perdendo-se alguma informação. Além disso, acreditamos que existe um conjunto de ferramentas padrão que ainda não estão totalmente desenvolvidas, compreendidas e exploradas [71]. Nas secções seguintes concentrar-nos-emos nas técnicas e métodos a que recorreremos no presente trabalho.

3.3. Análise de agrupamentos

A análise de agregados é uma técnica não supervisionada que permite a formação de grupos, para um determinado conjunto de dados, nos quais os objectos dentro de um grupo partilham características semelhantes [74]. Tem por finalidade reunir, por algum

critério de classificação, os objectos em grupos, de tal forma que exista alguma homogeneidade dentro do grupo e heterogeneidade entre grupos.

Esta análise tem sido aplicada em diversas áreas, incluindo a biologia (análise de informação genética), a recuperação de informação (agrupamento de resultados), o clima (encontrar padrões na pressão atmosférica das regiões polares e áreas do oceano), a Psicologia e a Medicina (detecção de padrões na distribuição espacial ou temporal de uma doença) e a Economia (análise adicional e actividades de marketing) [69].

A definição do que é um agregado não é muito precisa, pois varia de acordo com a natureza dos dados e com os resultados desejados. A dificuldade em decidir quais os grupos que podem ser formados pode ser ultrapassada tendo como base o princípio de que as distâncias entre pares de pontos, objectos, no espaço de medição, são inversamente proporcionais ao seu grau de similaridade [73].

Existem diversas versões de análises de agrupamentos [84] sendo a versão por nós escolhida a análise não supervisionada em modo hierárquico, já que o modo de associação é, em grande medida, independente de critérios impostos.

O processo de agrupamento, Figura 3, envolve basicamente duas etapas: a primeira refere-se à estimativa de uma medida de similaridade ou dissimilaridade entre os objectos e a segunda refere-se à adopção de uma técnica para a sequência de formação de grupos [85].

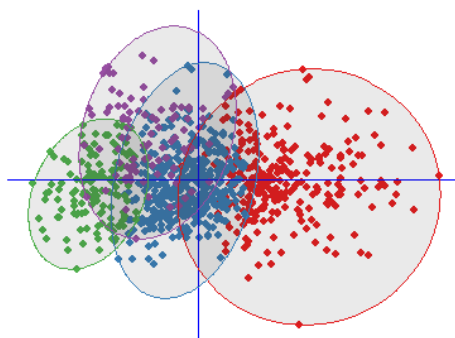


Figura 3. Representação do processo de agrupamento

Existe um grande número de medidas de similaridade ou de dissimilaridade sendo a escolha entre elas baseada na preferência do utilizador. Com a definição da medida a ser utilizada, a etapa seguinte é a adopção do método de agrupamento para formação dos grupos. Para realização desta tarefa, existe um grande número de métodos disponíveis, sobre os quais o utilizador tem de decidir qual o mais adequado ao seu propósito. Note-se que algoritmos diferentes podem levar a diferentes soluções [86].

A opção entre algoritmos exige a tomada de uma série de decisões independentes, que requerem o conhecimento das características dos diversos algoritmos à disposição.

3.3.1. Procedimento hierárquico

A técnica de agrupamento hierárquico consiste numa série de agrupamentos sucessivos entre objectos. Parte-se de n grupos de apenas um objecto, que vão sendo agrupados, sucessivamente, até que se encontre apenas um grupo que incluirá a totalidade dos n objectos.

Este método conduz a uma estrutura que descreve uma hierarquia de agrupamentos sobre os dados. Para um número inicial de n objectos na base de dados, ao todo ocorrem $n - 1$ associações.

A *HCA* interliga os objectos pelas suas associações, produzindo um dendrograma onde os objectos semelhantes, segundo as variáveis escolhidas, são agrupados entre si [69].

O dendrograma corresponde à representação bidimensional do esquema da associação sucessiva dos objectos, atendendo à sua similaridade, até culminar na fusão de todos os grupos num único grupo final. São especialmente úteis na visualização de semelhanças entre objectos representados por pontos no espaço com dimensão maior do que três, onde a representação de gráficos convencionais não é possível.

No que segue encontram-se descritas as etapas do processo de agrupamento, assim como as medidas de distância implementadas.

3.3.2. Semelhança e distância

Para que uma função d seja uma distância – e uma distância é uma medida de dissimilaridade - é necessário que as seguintes condições sejam satisfeitas, para quaisquer objetos i, j, k :

1. $d(i, j) = d(j, i)$ (simétrica);
2. $d(i, j) > 0$, se $i \neq j$;
3. $d(i, j) = 0$, se e somente se, $i = j$;
4. $d(i, j) \leq d(i, k) + d(z, k)$ (desigualdade triangular).

A propriedade (1) implica que todos os elementos da matriz de distância são positivos, a propriedade (2) implica que a sua diagonal é formada por zeros. A propriedade (3), por sua vez, implica que a matriz de distâncias é simétrica em relação à diagonal. Para que um índice de proximidade seja considerado uma métrica, este deve satisfazer, além das

três propriedades anteriores, a propriedade (4) de desigualdade triangular. Qualquer função que satisfaça estas quatro propriedades é chamada de distância.

O ponto de partida do processo de agrupamento corresponde à construção da matriz de distâncias, Figura 4, que é necessário calcular e armazenar,

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & \dots & \dots & d_{1n} \\ \vdots & d_{22} & & & & \vdots \\ \vdots & \vdots & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & \dots & \dots & d_{nn} \end{bmatrix}$$

Figura 4. Matriz D, contém medidas de similaridade ou dissimilaridade ao longo de n objectos.

Nesta matriz, cada elemento descreve o grau de diferença entre cada dois objectos com base nas variáveis escolhidas.

Uma das medidas de distância mais importantes é a que têm por base o coeficiente de correlação (em si uma medida de semelhança),

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}} \quad (1)$$

e que pode ter, por exemplo, a forma $1-|r|$. O coeficiente de correlação calcula a força e a direcção de uma relação linear entre duas variáveis, e varia entre -1 e +1. Um valor próximo de +1 ou -1 indica a existência de uma forte correlação positiva e negativa respectivamente, entre as duas variáveis. Se o valor for próximo de 0 significa que não existe qualquer correlação ou a existir que é extremamente diminuta [70].

Mais utilizada ainda é a distância euclidiana (ou o seu quadrado), cuja forma é

$$d(i, j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{in} - x_{jn}|^2} \quad (2)$$

Esta distância não possui um valor limite, mas é sempre positiva ou nula. A determinação de valores de pequenas dimensões é indicativa da existência de uma forte semelhança entre as variáveis [70].

Outras distâncias de uso comum são a distância de Manhattan

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{in} - x_{jn}| \quad (3)$$

e a distância de Minkowski:

$$d(i, j) = \sqrt[q]{|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \dots + |x_{in} - x_{jn}|^q} \quad (4)$$

onde $q \geq 1$. Logo, a distância de Minkowski abrange tanto a distância euclidiana ($q = 2$) como a distância de Manhattan ($q = 1$).

Uma vez calculadas as distâncias podem ser utilizados diferentes tipos de modos para promover a ligação entre os objectos. A abordagem mais comum é o chamado agrupamento associativo, no qual os objectos individuais são gradualmente ligados uns aos outros em grupos, dos modos que a seguir se descrevem [85]. A ligação é, basicamente, a estratégia para se estabelecerem distâncias objecto-grupo ou grupo-grupo.

3.3.3. Ligação

Método da ligação simples (*single linkage*)

A ligação simples, cujo procedimento se esquematiza na Figura 5, é um dos métodos mais fáceis e de rápida aplicação. A distância entre os grupos é definida como sendo aquela entre os objectos mais próximos, avaliando-se a semelhança através da menor distância entre pontos. Este método tende a formar longas cadeias⁵, é sensível a *outliers*, pois tem tendência a incorporar os *outliers* num grupo já existente, e grupos muito próximos podem não ser identificados [85].

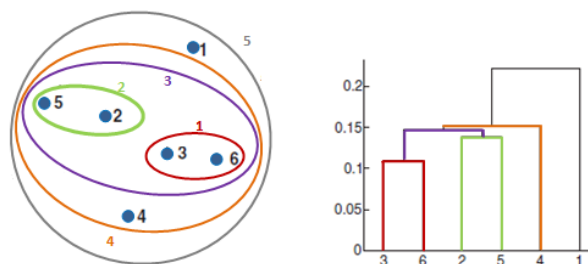


Figura 5. Distância entre grupos calculada através da menor distância (*single linkage*). À direita apresenta-se o respectivo dendrograma [69].

⁵ Situação em que há um primeiro grupo de um ou mais objectos que passa a incorporar um grupo de apenas um objecto, formando uma longa cadeia, onde se torna difícil definir um nível de corte para classificar os objectos em grupos.

Método da ligação completa (*complete linkage*)

Ao contrário do anterior, o método da ligação completa, representado na Figura 6, determina a distância entre dois grupos de acordo com a maior distância entre um par de objectos, sendo cada objecto pertencente a um grupo distinto. Geralmente, leva a grupos compactos e discretos, sendo os seus valores de dissimilaridade relativamente grandes [85].

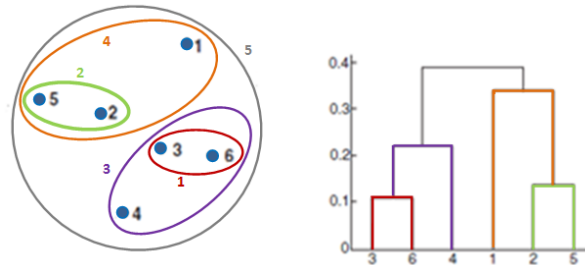


Figura 6. Distância entre grupos através da associação completa (*complete linkage*) [69].

Método da ligação média (*average linkage*)

Neste método, representado na Figura 7, a distância entre dois grupos é definida como a média das distâncias entre todos os pares de objectos em cada grupo [85].

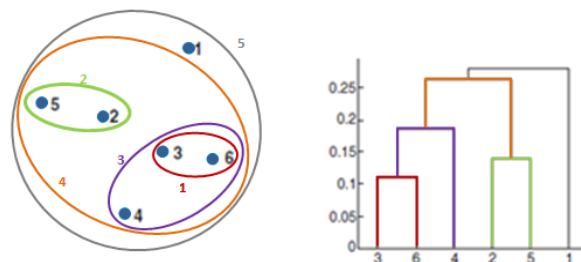


Figura 7. Distância obtida através da média das distâncias entre os objectos (*average linkage*) [69].

Método de variância mínima (*Ward*)

O método da ligação de *Ward*, esquematizado na Figura 8, baseia-se na análise de variância, associando os objectos aos grupos nos quais estes promovem a menor variância intra-grupo. Este algoritmo é altamente eficiente na formação de grupos.

Inicialmente, admite que cada um dos objectos constitui um único grupo. Considerando a primeira reunião de objectos num novo grupo, a soma dos desvios dos pontos representativos dos seus elementos, em relação à média do grupo, é calculada, e dá uma indicação de homogeneidade do grupo formado. Os grupos formados possuem uma

elevada homogeneidade interna. No entanto, pode apresentar resultados insatisfatórios quando o número de elementos em cada grupo é praticamente igual; tem tendência a combinar grupos com poucos elementos e é sensível à presença de *outliers* [85].

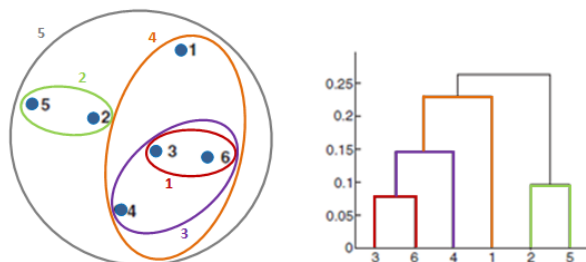


Figura 8. Distância entre grupos obtida através do método da variância mínima (*Ward linkage*) [69].

3.4. Análise de componentes principais

A análise de componentes principais (*PCA*) é provavelmente a mais antiga e mais conhecida das técnicas de análise multivariada [70], e é recomendada como uma ferramenta exploratória para encontrar tendências em dados desconhecidos. É um método simples, não paramétrico, de extração de informações relevantes a partir de conjuntos de dados multivariados [87-88].

O seu objectivo principal consiste na redução da dimensionalidade de grandes matrizes de dados - as m variáveis originais são substituídas por um outro subconjunto de p variáveis não correlacionadas, de menor dimensão, designadas por componentes principais, com a menor perda de informação possível.

Para além da redução de dimensionalidade, outra vantagem deste método é o facto de que as novas variáveis, as componentes principais, não são correlacionadas e, em vez de se analisar um elevado número de variáveis originais com uma estrutura inter-relacional complexa, analisam-se apenas algumas variáveis não correlacionadas [89].

Aplicando este método, é possível efectuar uma simplificação e redução da dimensão original dos dados, modelação, detecção de *outliers*, selecção de variáveis importantes num determinado sistema, classificação e previsão [89].

De um modo geral, este método descobre novas variáveis (componentes principais, PCs) que consigam reunir a maior parte da variabilidade dos dados, sendo estas novas variáveis combinações lineares das originais. De entre todas as possíveis combinações lineares, escolhe-se, em cada caso, a de variância máxima, dado que as componentes principais devem reflectir, tanto quanto possível, as características dos dados, devendo

explicar uma grande parte da variação associada às variáveis iniciais [89]. A representação destas variáveis pode ser observada na Figura 9.

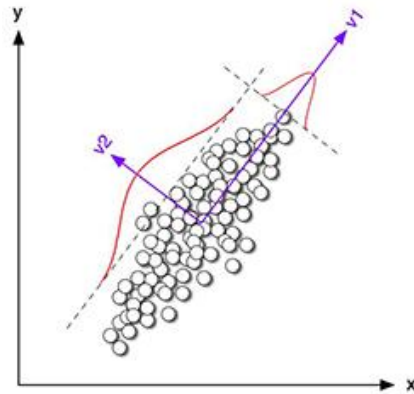


Figura 9. Representação das PCs

3.4.1. A transformação

Seja $X = [X_1 \dots X_p]'$ um vector aleatório com valor médio μ e matriz de covariância Σ ,

$$\Sigma = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \sigma_{pp} \end{bmatrix} \quad (5)$$

onde,

$$\sigma_{ij} = Cov(X_i, X_j) \quad (6)$$

Seja $X_{(n \times p)}$ uma matriz de dados de dimensão n , de observações deste vector aleatório

$$X = \begin{bmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{bmatrix} \quad (7)$$

O resultado do PCA é uma rotação do sistema ortogonal de eixos associados às variáveis iniciais (matriz $X_{(n \times p)}$), de forma que, após a sua aplicação, se disponha de um novo sistema de dados $Y_{(n \times p)}$. As colunas da matriz Y são as novas variáveis Y_j , as componentes principais. Com este passo pretende-se encontrar o novo conjunto de p variáveis (componentes principais, $Y_1 \dots Y_p$) não correlacionadas e que apresentem a máxima variância [90].

3.4.2. Obtenção dos componentes

As componentes principais são combinações lineares das p variáveis da matriz X

$$Y_j = a_{1j}X_1 + a_{2j}X_2 + \dots + a_{pj}X_p \quad (8)$$

onde $j = 1, \dots, p$ e a_{ij} ($i = 1, \dots, p; j = 1, \dots, p$) são constantes.

Os coeficientes destas combinações lineares são determinados de modo a satisfazerem as condições seguintes:

1. $Var(Y_1) \geq Var(Y_2) \geq \dots \geq Var(Y_p)$.
2. Quaisquer duas componentes principais não são correlacionadas, $Corr(Y_i, Y_j) = 0, \forall i, j$.
3. Em qualquer componente principal a soma dos quadrados dos coeficientes que engloba é igual a 1 (para $Y_i: a_{1j}^2 + a_{2j}^2 + \dots + a_{pj}^2 = 1$).

Das condições anteriores retiramos que, Y_1 é a componente principal com maior variância; Y_2 é a componente principal com a segunda maior variância, sujeita à condição de ser não correlacionada com Y_1 ; Y_3 é a componente principal com a terceira maior variância, sujeita à condição de ser não correlacionada com Y_1 e com Y_2 (e assim por diante).

Assim, na Equação (8), a_1, a_2, \dots, a_p são, respectivamente, os p vectores próprios associados aos p maiores valores próprios de Σ ($\lambda_1 > \lambda_2 > \dots > \lambda_p$), com $Var(Y_j) = \lambda_j$.

A covariância entre cada duas componentes principais Y_j e $Y_{j'}$ é, como se disse, imposta como nula, pois todas as componentes principais foram determinadas de forma a serem não correlacionadas. Tem-se então que $Cov(Y_j, Y_{j'}) = a_j' \Sigma a_{j'} = a_j' \lambda_j a_{j'} = \lambda_j a_j' a_{j'} = 0$, que equivale a ter $a_j' a_{j'} = 0$, o que indica que a_j' e $a_{j'}$ (com $j \neq j'$) são vectores ortogonais.

Em muitas situações, as variáveis em estudo não são todas medidas na mesma unidade, na mesma escala, ou são até de natureza distinta. Surge, assim, a necessidade de estabelecer uma certa uniformização, que se consegue através da divisão de cada valor pelo desvio padrão da variável centrada correspondente. Este procedimento conduz à obtenção de variáveis com valor médio nulo e variância unitária. As variáveis em estudo passam a ter todas a mesma variância e a influência das variáveis de variância pequena tende a ser inflacionada enquanto a influência das variáveis de variância elevada tende a ser reduzida.

A matriz de covariância do conjunto destas “novas” variáveis é igual à matriz de correlação do conjunto de variáveis iniciais, dado que:

$$Cov\left(\frac{X_i}{\sigma_i}, \frac{X_j}{\sigma_j}\right) = \frac{Cov(X_i, X_j)}{\sigma_i \sigma_j} = Corr(X_i, X_j) \quad (9)$$

Assim, a análise de componentes principais de um conjunto de dados deste tipo, é efectuada utilizando a matriz de correlação, P . As componentes principais serão

determinadas tendo em conta os valores e vectores próprios da matriz P . Matematicamente, tudo se processa da mesma forma. No entanto, os vectores próprios de P não são iguais aos de Σ , e as componentes principais também não serão as mesmas.

A matriz P define-se como

$$P = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \vdots & \vdots & \dots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{bmatrix} \quad (10)$$

em que

$$\rho_{ij} = \text{Corr}(X_i, X_j) \quad (11)$$

3.4.3. Redução de dimensionalidade

A redução de dimensionalidade atinge-se considerando apenas algumas das componentes principais, i.e., as de maior variância. Dado que, as componentes principais se podem ordenar por ordem decrescente da sua variância e que quanto maior esta for mais representativa dos dados originais será a correspondente componente principal, devemos considerar as primeiras componentes principais.

Assim, a soma das variâncias das componentes principais é dada por

$$\sum_{j=1}^p \text{Var}(Y_j) = \sum_{j=1}^p \lambda_j \quad (12)$$

Além disso, como se sabe, numa matriz simétrica (que é o caso de Σ) a soma dos seus valores próprios é igual ao traço da matriz, pelo que

$$\text{tr}(\Sigma) = \sum_{j=1}^p \text{Var}(X_j) \Rightarrow \sum_{j=1}^p \lambda_j = \sum_{j=1}^p \text{Var}(X_j) \quad (13)$$

de onde

$$\sum_{j=1}^p \text{Var}(Y_j) = \sum_{j=1}^p \text{Var}(X_j) \quad (14)$$

Isto significa que a soma das variâncias das variáveis originais é igual á soma das variâncias das componentes principais (se considerarmos todas as componentes principais explicamos toda a variabilidade). Assim, a proporção da variância total que é explicada pela j -ésima componente principal Y_j e que indica a importância da mesma é dada por

$$\frac{\lambda_j}{\sum_{j=1}^p \lambda_j} = \frac{\lambda_j}{tr(\Sigma)} \quad (15)$$

Existem vários critérios que podem ser usados para a escolha do número de componentes principais. Estes critérios encontram-se descritos no que se segue.

Critério de Pearson (ou regra dos 80%)

Este critério é utilizado quando se utiliza a matriz de covariância. O número de componentes principais é escolhido até recuperarmos mais de 80% da informação total ou variabilidade total. Por outras palavras, devem considerar-se tantas componentes principais quantas as necessárias para que a percentagem de variância por elas explicada seja superior a 80%. Tal consiste em reter as primeiras r componentes principais de modo a que [91]

$$\sum_{j=1}^r \frac{\lambda_j}{\sum_{j=1}^p \lambda_j} = \frac{\sum_{j=1}^r \lambda_j}{\sum_{j=1}^p \lambda_j} \geq 0.80 \quad (16)$$

Critério de Kaiser ($\lambda > 1$)

O critério de Kaiser é utilizado com a matriz de correlação, embora o critério anteriormente descrito também seja uma possibilidade. Neste caso, devem ser consideradas apenas as componentes com valor próprio superior à unidade (note-se que este valor unitário é média do conjunto de valores próprios) [91]

$$\bar{\lambda} = \frac{1}{p} \sum_{j=1}^p \lambda_j \quad (17)$$

Scree plot

Este terceiro critério permite utilizar um gráfico onde se representam os pontos de abcissa j e ordenada igual ao j -ésimo valor próprio ou à percentagem de variância explicada pela j -ésima componente principal, i.e. pontos de coordenadas (j, λ_j) ou $(j, \lambda_j / \sum_{j=1}^p \lambda_j)$, onde se distinguem as contribuições das diversas componentes principais. De acordo com este critério, devem-se considerar as r componentes principais que mais contribuem, destacando-se de forma acentuada das restantes [89].

Scores e loadings

As coordenadas das amostras no novo sistema de referência são designadas por *scores*, enquanto o coeficiente da combinação linear que descreve cada *PC*, isto é, os pesos das variáveis originais em cada *PC*, são denominados por *loadings*.

Nesta altura, sabemos que as componentes principais resultam de uma transformação sobre as variáveis em estudo (combinação linear). Podemos agora pensar em aplicar a mesma transformação aos dados, ou seja, aos vectores de observações x_1, x_2, \dots, x_p (colunas da matriz de dados X) das variáveis X_1, X_2, \dots, X_p , respectivamente.

Obtemos uma nova matriz de dados, a matriz Y , com dimensão $(n \times r)$ em que o ij -ésimo elemento será igual ao *score* do i -ésimo objecto para a j -ésima componente principal

$$y_{ij} = a_{1j}x_{i1} + a_{2j}x_{i2} + \dots + a_{pj}x_{ip} \quad (18)$$

A matriz dos *scores* dos objectos é dada por

$$Y = \begin{bmatrix} y_{11} & \dots & y_{1r} \\ \vdots & \ddots & \vdots \\ y_{n1} & \dots & y_{nr} \end{bmatrix} \quad (19)$$

3.4.4. Representações gráficas

As representações gráficas são um óptimo auxiliar na interpretação dos resultados do PCA. No desenvolvimento deste trabalho, recorreu-se ao programa Gnuplot (versão 4.4) para efectuar estas representações.

Representação das variáveis (*loadings*)

Na representação gráfica das *loadings*, cada ponto representa uma variável e o plano é definido por dois (ou três) eixos correspondentes aos primeiros componentes principais. A cada variável é associado um ponto, cujas coordenadas são as correlações dessa variável com cada uma das componentes principais em causa.

Representação dos objectos no novo sistema de eixos (*scores*)

Na representação gráfica dos *scores*, cada ponto representa um objecto e tal como no caso anterior, o plano é definido por dois (ou três) eixos correspondentes aos componentes principais. A cada objecto é associado um ponto, cujas coordenadas são os

scores desse objecto para cada uma das componentes principais. Em geral escolhem-se preferencialmente as duas primeiras componentes, porque são as que mais contribuem para a explicação da variabilidade dos dados e por razões de clareza na representação.

3.5. Mínimos quadrados não lineares

O método dos mínimos quadrados é talvez a técnica de aproximação mais utilizada na análise numérica. Nesta secção vai ser realizada uma introdução ao método dos mínimos quadrados utilizando conceitos da álgebra linear e de análise numérica. Também vai ser apresentada a forma como pode ser aplicado para o ajuste de curvas. Em 1809, Carl Friedrich Gauss demonstrou que a melhor maneira de determinar um parâmetro desconhecido de uma equação de condições é minimizando a soma dos quadrados dos resíduos, mais tarde chamado de Mínimos Quadrados por Adrien-Marie Legendre.

Um programa de mínimos quadrados utiliza uma minimização da soma S

$$S \equiv \sum_{i=1}^N (y_i^0 - y_i)^2 \quad (20)$$

onde y_i^0 corresponde aos valores observados para y e y_i aos valores calculados de y ou seja, o método dos mínimos quadrados implica o processo de minimização dos quadrados dos resíduos, pressupondo uma distribuição Gaussiana ou normal (Figura 10).

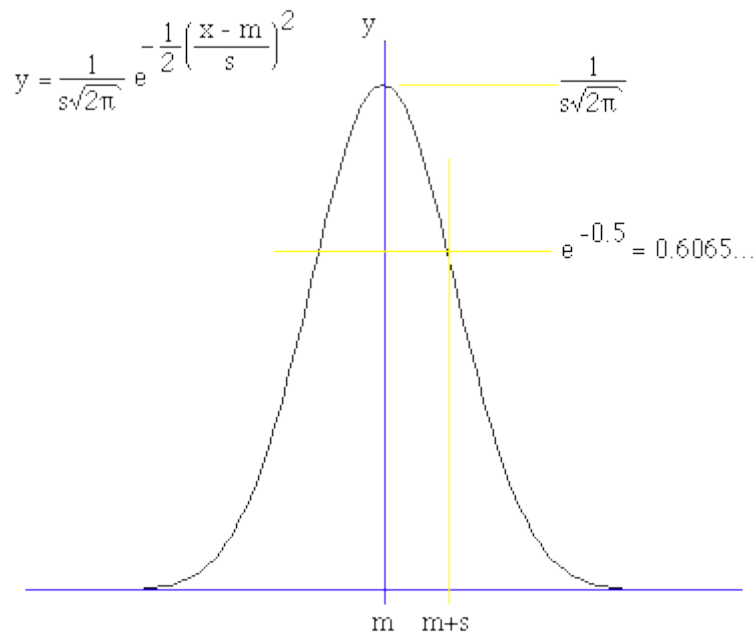


Figura 10. Distribuição gaussiana (normal) para uma variável x , com média m e desvio padrão s [92].

O problema dos mínimos quadrados pode ser formulado através de sistemas lineares. Este problema encontra aplicação na determinação da curva que melhor se ajusta a um conjunto de pontos $\{ (x_1, y_1), \dots, (x_n, y_n) \}$.

Os conceitos e princípios básicos na discussão de mínimos quadrados lineares podem ser adoptados para os casos de ajustes não-lineares. Existem diversos métodos que são usados para sua resolução, sendo aqui discutido o método de Levenberg-Marquardt [93-94].

3.5.1. Algoritmo Levenberg - Marquardt

O algoritmo de Levenberg-Marquardt foi desenvolvido para resolver iterativamente problemas de minimização de funções não-lineares pelo método dos mínimos quadrados. É o algoritmo de optimização mais utilizado hoje em dia, superando outros numa grande variedade de situações [95-97].

Vamos escrever a função a ser minimizada na forma

$$f(x) = \frac{1}{2} \sum_{j=1}^m r_j^2(x) \quad (21)$$

onde $x = (x_1, x_2, \dots, x_n)$ é um vector e cada r_j é uma função de R^n a R . Os valores de r_j são designados como resíduos, numa notação óbvia em mínimos quadrados, sendo assumido que $m \geq n$.

Por simplicidade, f pode ser representada a partir de um vector de resíduos definido por $r(x) = (r_1(x), r_2(x), \dots, r_m(x))$. Assim, a função f pode ser reescrita como $f(x) = \frac{1}{2} \|r(x)\|^2$ e as suas derivadas podem ser expressas, usando a matriz Jacobiana J de r em relação a x , definida como

$$J(x) = \frac{\partial r_j}{\partial x_i}, \quad 1 \leq j \leq m, \quad 1 \leq i \leq n.$$

Vamos considerar em primeiro lugar a situação linear, na qual cada função r_i é linear. Aqui a matriz Jacobiana é constante e podemos representar r de forma a f ser

definida pela função quadrática $f(x) = \frac{1}{2} \|J(x) + r(0)\|^2$. Também podemos obter

$\nabla f(x) = J^T \frac{1}{2} (Jx + r)$ e $\nabla^2 f(x) = J^T J$. Resolvendo a equação para um mínimo, $\nabla f(x) = 0$,

obtemos $x_{\min} = -(J^T J)^{-1} J^T r$, que consiste na solução de um sistema de equações lineares.

Retornemos ao caso geral não linear, nesta situação temos

$$\nabla f(x) = \sum_{j=1}^m r_j(x) \nabla r_j(x) = J(x)^T r(x) \quad (22)$$

$$\nabla^2 f(x) = J(x)^T J(x) + \sum_{j=1}^m r_j(x) \nabla^2 r_j(x) \quad (23)$$

O problema dos mínimos quadrados não lineares apresenta uma característica distinta. Para uma dada matriz Jacobiana, J , podemos obter a matriz Hessiana $\nabla^2 f(x)$, se for possível aproximar $r_{j(x)}$ através de funções lineares pequenas, $\nabla^2 r_j(x)$, ou se os valores dos resíduos $(r_{j(x)})$ também forem diminutos. Neste caso a matriz Hessiana transforma-se simplesmente em

$$\nabla^2 f(x) = J(x)^T J(x) \quad (24)$$

tal como para os casos lineares. A aproximação aqui usada é uma de quase linearidade dos resíduos de forma a que $\nabla^2 r_j(x)$ seja muito pequeno. É importante notar que esta última equação só é válida se os resíduos forem muito pequenos. Problemas que envolvam resíduos de grandes dimensões não podem ser resolvidos através da aproximação quadrática, e conseqüentemente, o desempenho dos algoritmos aqui apresentados diminui.

De uma maneira geral, o algoritmo de Levenberg-Marquardt propõe uma solução intermédia entre o algoritmo do gradiente descendente (*steepest descent*) e o método iterativo de Gauss-Newton. A regra de actualização de parâmetros pode ser descrita como

$$x_{i+1} = x_i - (H + \lambda I)^{-1} \nabla f(x_i) \quad (25)$$

onde x representa o vector de parâmetros e ∇f representa o gradiente do erro médio quadrático, H representa a matriz Hessiana, I a matriz identidade e λ é um factor de ajuste.

Podemos ver que esta regra leva em consideração quer a inclinação da superfície do erro (método do gradiente descendente) quer a curvatura desta superfície (método de Gauss-Newton).

O factor de ajuste indica qual dos dois métodos é o predominante. Para grandes valores do factor de ajuste, predomina o método do gradiente descendente e a actualização

dos pesos ocorre na direcção da inclinação da superfície do erro. Caso contrário, predomina o método de Gauss-Newton e a actualização ocorre no sentido da curvatura da função.

O algoritmo controla o valor do factor de ajuste de forma iterativa. O cálculo é iniciado com um valor arbitrário, procede-se ao cálculo do erro e aplica-se a regra de actualização de pesos. Calcula-se, então, o novo parâmetro. Caso o erro tenha aumentado, deve aumentar-se o factor de ajuste. Em seguida, deve recomeçar-se a iteração. Caso o erro tenha diminuído, deve diminuir-se o factor de ajuste. A direcção do gradiente é tão mais próxima quanto mais distante se estiver do ponto mínimo.

Deve notar-se que podem existir problemas com o algoritmo de Levenberg-Marquardt, os quais levaram Marquardt a propor uma melhoria no algoritmo original que faz com que a parcela relativa ao gradiente descendente passe também a incorporar informação acerca da curvatura da superfície do erro, fazendo com que a actualização seja considerável mesmo quando o gradiente é muito pequeno.

A regra final de actualização do algoritmo de Levenberg-Marquardt é então

$$x_{i+1} = x_i - (H + \lambda \text{diag}[H])^{-1} \nabla f(x_i) \quad (26)$$

Capítulo 4

As detenções em 2009

Nesta secção procederemos à análise da *table 69*, já descrita na secção 2.1.4, que sumaria os dados relativos às detenções efectuadas para 29 tipos de crime em cada estado durante o ano de 2009. Devido à ausência de informação num elevado número de tipos de crime, o estado da *Florida* não foi incluído no estudo.

4.1. Os dados em bruto

4.1.1. HCA por tipo de crime

Numa primeira fase, esta análise não será baseada em quaisquer pressupostos. Como tal, consideraremos cada estado como um objecto, caracterizado pelo número total de cada um dos crimes, organizados sob a forma de um vector. A ausência de pressupostos leva-nos a considerar, numa primeira abordagem, todos os 29 crimes tabelados.

O tratamento dos dados iniciar-se-á sob a forma de uma análise hierárquica de agrupamentos, com base na ligação de *Ward*, já descrita no capítulo 3. O dendrograma obtido consta da Figura 11. Nele são visíveis quatro grupos explicitamente enumerados na Figura 12.

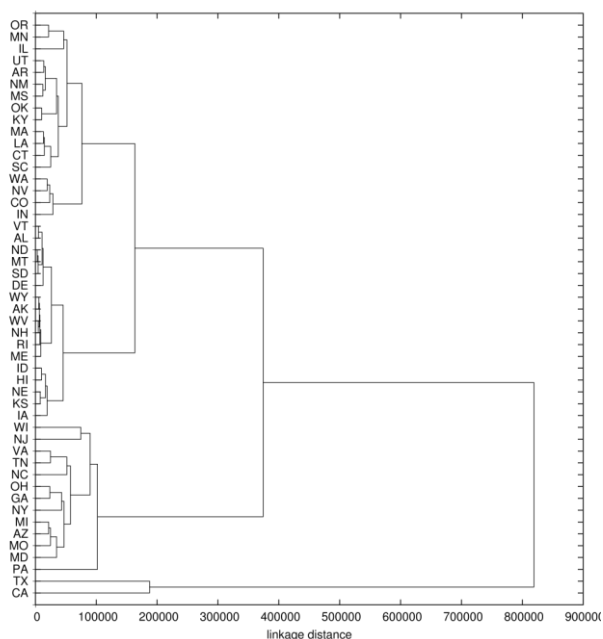


Figura 11. Dendrograma obtido descrevendo cada estado sob a forma de um vector em que os elementos correspondem ao número total de crimes, de acordo com a tabela xxx do FBI UCR para 2009. A ligação assenta no método de *Ward*. É visível uma estrutura com 4 grupos embora o último, englobando *Texas* e *California*, junte dois objectos de fraca semelhança entre si mas claramente distintos dos restantes.

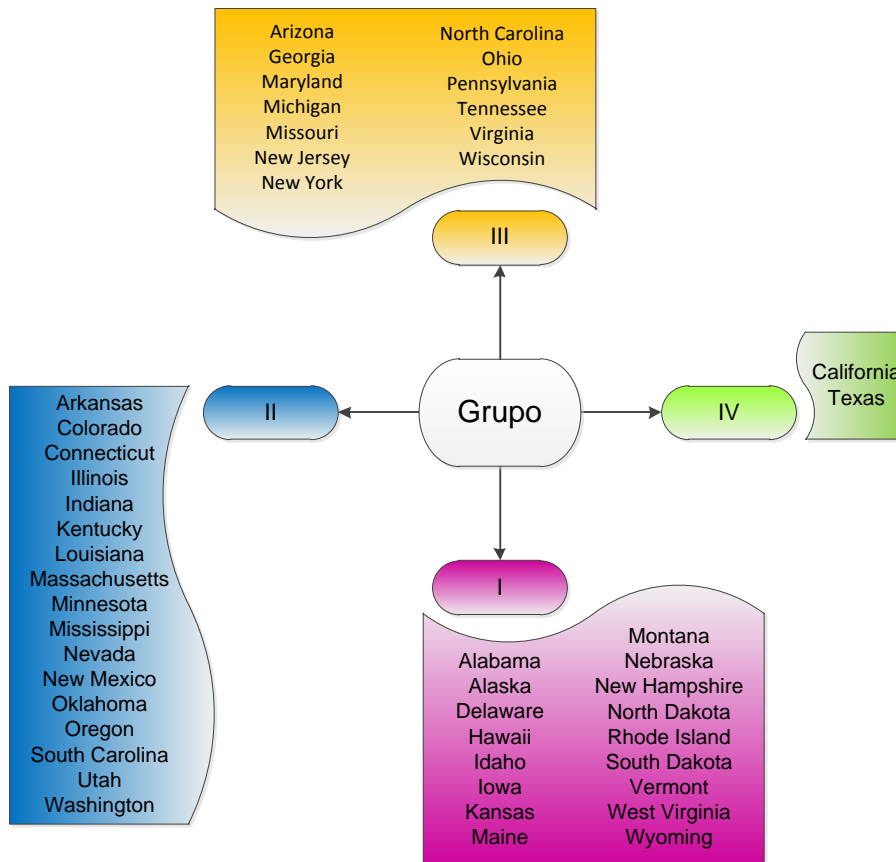


Figura 12. Representação esquemática dos grupos formados através de análise hierárquica de agrupamentos, na qual cada estado é um objecto que é caracterizado pelo número total de cada um dos crimes, com base numa ligação de *Ward*. Para estabelecimento dos grupos foi utilizada uma linha de corte adequada.

Os diversos grupos são compactos, com excepção do grupo IV. Neste último, *California* e *Texas* apresentam notórias diferenças entre si. Apesar disso as diferenças face aos restantes estados transformam-nos num só agregado. Os dois primeiros grupos compreendem 17 estados cada, sendo o terceiro um pouco menor, com 13 estados.

Para permitir uma análise preliminar caracterizadora destes grupos inspeccionemos a Figura 13 que faz a sua representação sobre o mapa dos Estados Unidos.

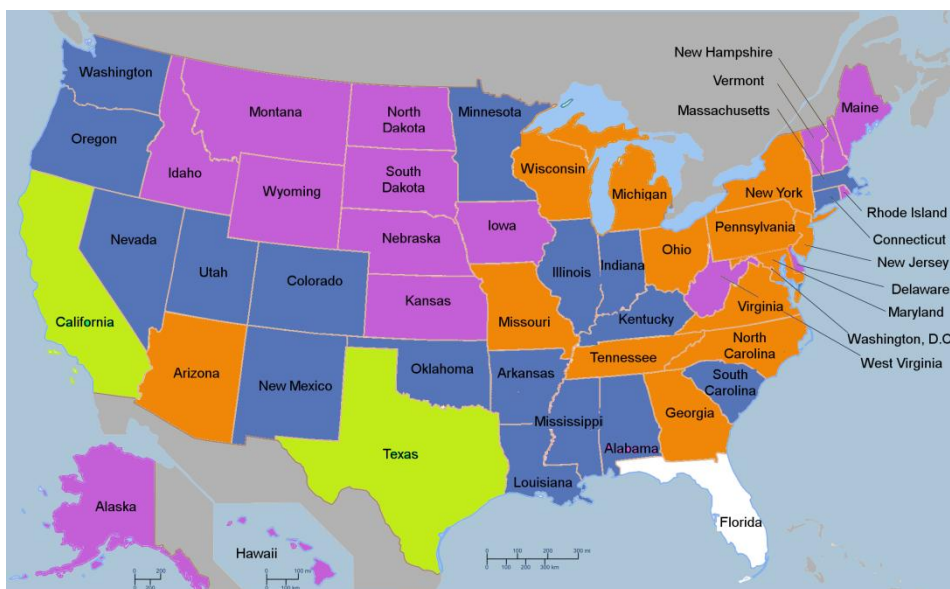


Figura 13. Padrões geográficos obtidos numa análise preliminar. Representação sobre o mapa dos Estados Unidos dos grupos patentes no dendrograma da Figura 11. A rosa, azul, laranja e verde representam-se respectivamente os grupos I, II, III e IV.

Neste mapa é visível uma zona central a partir da fronteira do Canadá e que engloba 8 estados. A *West Virginia* e um conjunto situado na parte nordeste que compreende *New Hampshire*, *Vermont*, *Maine* e *Rhode Island*, e ainda o *Alaska* e *Hawaii* completam este grupo. A referida zona central encontra-se rodeada por uma faixa que se inicia no estado de *Washington* e que se estende quase ininterruptamente até ao *Minnesota*. Alguns estados dispersos como sejam *South Carolina*, *Connecticut* e *Massachusetts* fazem também parte deste grupo. Essencialmente residindo na parte Este, surge o terceiro grupo, no qual apenas o *Arizona* se afasta geograficamente. Finalmente *Texas* e *California*, como já referido formando um só grupo, encontram-se geograficamente afastados, separados por *Arizona* e *New Mexico*.

Ainda à luz desta análise preliminar, a qual será refinada em secções subsequentes, podemos adiantar uma hipótese justificativa da formação desta estrutura. De facto os estados centrais caracterizam-se por possuírem populações na ordem de um milhão de habitantes, registando um número de crimes proporcional a esta população e que iremos considerar como correspondendo a uma baixa criminalidade total. O avanço dessa zona central para a periferia corresponde, genericamente, a uma passagem através de um grupo de criminalidade intermédia, grupo II, atingindo áreas de criminalidade média alta, a Leste ou alta, a Sul e Sudoeste. Este padrão está esquematizado na Figura 13 e é, em grande medida, o resultado do aumento da população estadual no mesmo sentido. A já referida proporcionalidade entre crime e população traduz o padrão criminal num padrão geográfico.

De facto, os estados centrais de menor população estão rodeados por estados de população crescente.



Figura 14. Esquema ilustrativo do perfil geral patente na Figura 13.

Taxa de criminalidade

Prossigamos agora para uma análise semelhante à realizada na secção anterior, mas na qual cada estado é agora definido por um vector em que a frequência de cada tipo de crime é dividida pela população total do estado. O dendrograma obtido através do procedimento geral utilizado neste trabalho está patente na Figura 15. Registam-se substanciais alterações relativamente ao observado com base apenas na frequência de crimes. Podemos agora identificar cinco grupos, descritos na Figura 16. O número de elementos de cada grupo varia entre seis e quinze. Na Figura 17 apresenta-se a distribuição geográfica.

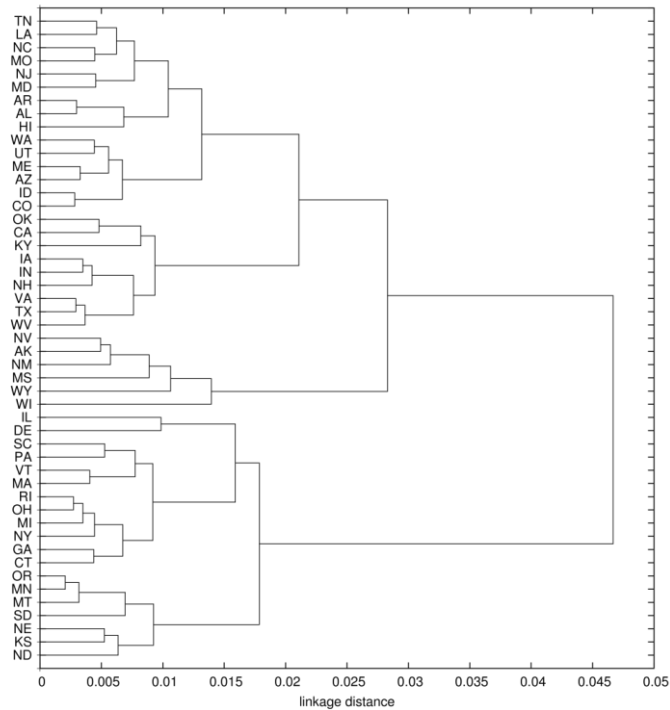


Figura 15. Dendrograma obtido descrevendo cada estado sob a forma de um vetor em que os elementos correspondem à frequência de cada tipo de crime dividida pela população total do estado. A ligação assenta no método de Ward. É visível uma estrutura com 5 grupos na qual o número de elementos de cada grupo varia entre seis e quinze.

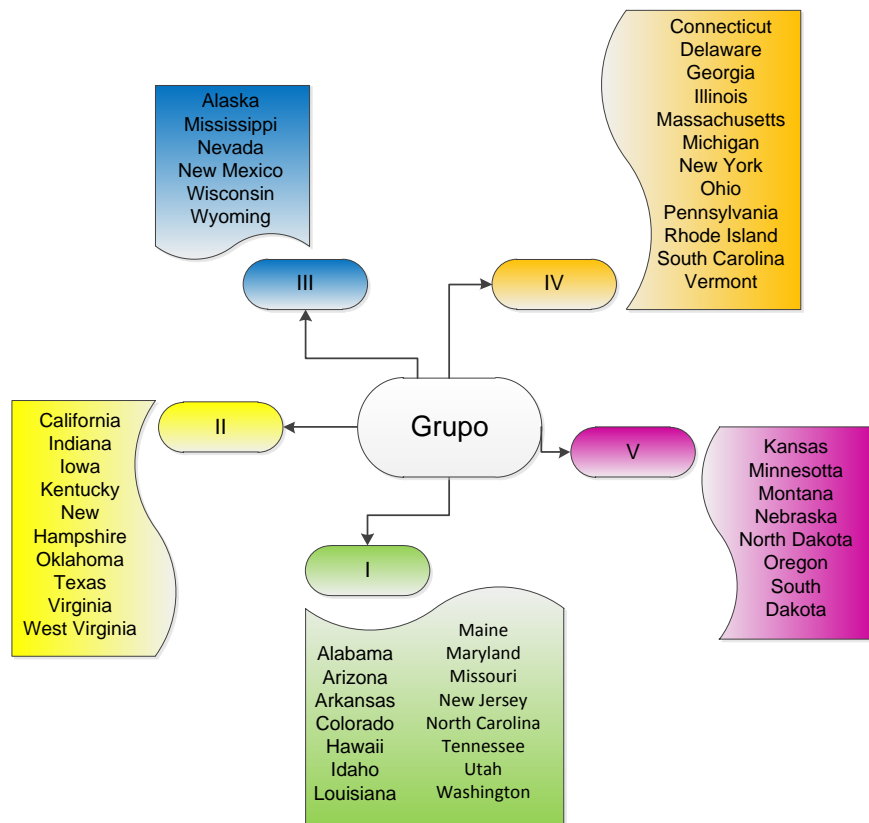


Figura 16. Representação esquemática dos grupos formados através de análise hierárquica de agrupamentos, na qual cada estado é um objecto que é caracterizado pela frequência de cada tipo de crime dividida pela população total do estado, com base numa ligação de *Ward*.

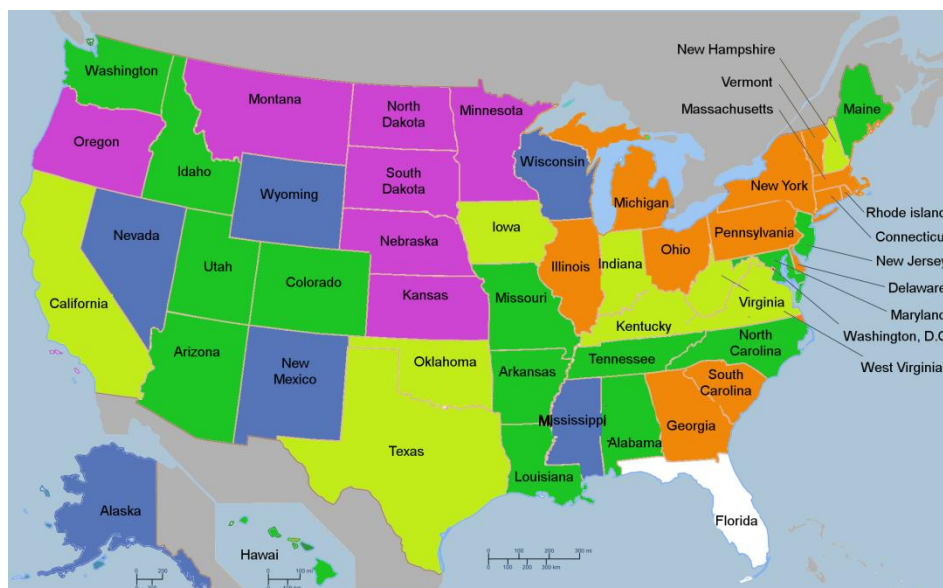


Figura 17. Padrões geográficos obtidos. Representação sobre o mapa dos Estados Unidos dos grupos patentes no dendrograma da Figura 15. A verde, amarelo, azul, laranja e rosa representam-se respectivamente os grupos I, II, III, IV e V.

Existe alguma tendência para que os grupos se estendam ao longo de zonas contíguas, com excepção do grupo III no qual os estados estão dispersos.

Relativamente ao anteriormente observado mantém-se uma zona central até à fronteira com o Canadá, grupo V, alguma prevalência de estados a Este, grupo IV, e *California* e *Texas* ainda fazem parte de um mesmo grupo, agora alargado. Não é trivial apresentar uma hipótese explicativa para este perfil com este nível de análise.

Fracção de crime

Finalmente, inspeccionemos os padrões criados por uma definição de cada estado com base num perfil em que cada elemento do vector é a fracção de cada tipo de crime relativamente ao número total de crimes cometido nesse estado. A representação do dendrograma resultante está patente na Figura 18, estando os grupos formados indicados na Figura 19. A Figura 20 apresenta a distribuição geográfica.

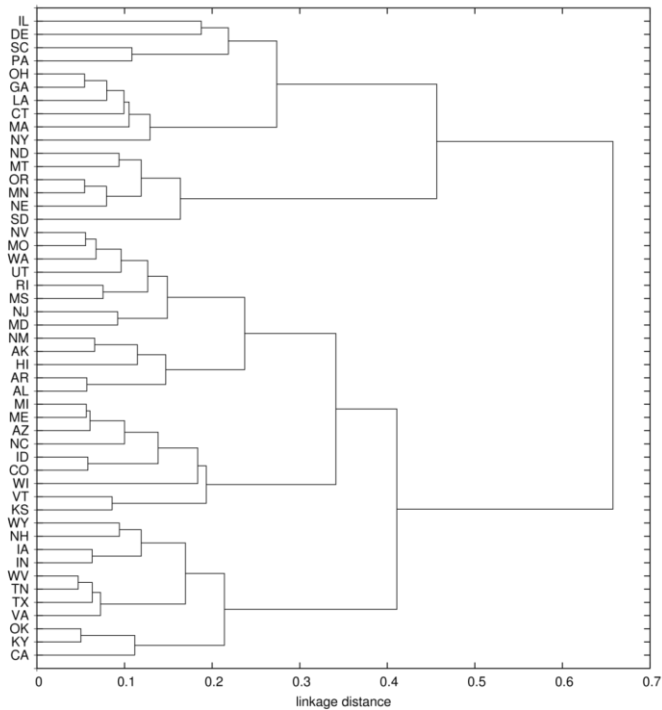


Figura 18. Dendrograma obtido descrevendo cada estado sob a forma de um vector em que cada elemento é representado pela fracção de cada tipo de crime relativamente ao número total cometido nesse estado. A ligação assenta no método de *Ward*. É visível uma estrutura com 5 grupos na qual o número de elementos de cada grupo varia entre seis e doze.

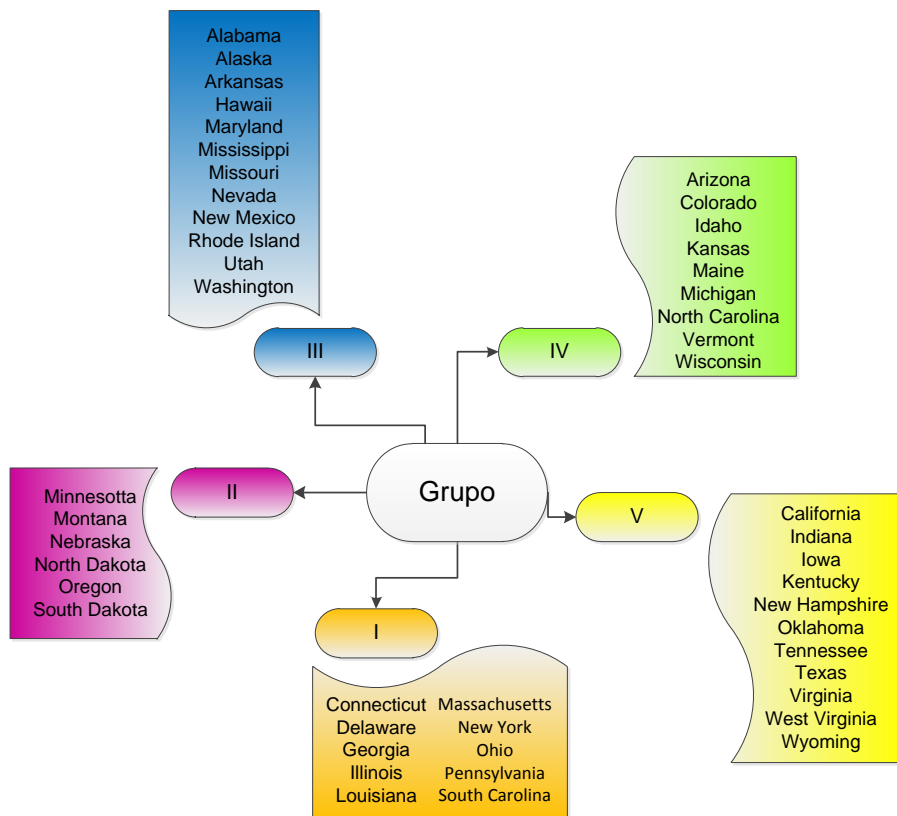


Figura 19. Representação esquemática dos grupos formados através de análise hierárquica de agrupamentos, na qual cada estado é um objecto que é caracterizado pela fracção de cada tipo de crime relativamente ao número total cometido nesse estado, com base numa ligação de *Ward*.



Figura 20. Padrões geográficos obtidos. Representação sobre o mapa dos Estados Unidos dos grupos patentes no dendrograma da Figura 8. A laranja, rosa, azul, verde e amarelo representam-se respectivamente os grupos I, II, III, IV e V.

Destes resultados é visível um padrão semelhante ao obtido na secção anterior, mas agora com os estados mais homogeneamente distribuídos pelos grupos. Um deles é menor, contendo apenas seis elementos enquanto os restantes variam entre 9 e 12 elementos.

O padrão obtido revela novamente um número substancial de estados contíguos pertencentes ao mesmo grupo, mas também aqui adiaremos uma caracterização mais profunda dos grupos formados.

4.1.2. Análise de componentes principais

Nesta secção procederemos a uma identificação das variáveis caracterizadoras dos três sistemas anteriormente sujeitos ao estudo de agrupamentos. Tentaremos também extrair informação que nos permita confirmar ou esclarecer a formação dos grupos identificados. Esta análise assentará na matriz de covariância, embora seja apresentado um exemplo adicional em que se utiliza a matriz de correlação. Lembramos que a matriz de covariância privilegia as grandes variações, em valor absoluto, enquanto a matriz de

correlação considera variações relativas significativas. Nos casos em que o perfil foi normalizado, as discrepâncias atenuam-se.

A frequência de crimes

O resultado da utilização de PCA sobre as frequências relativas a cada tipo de crime revela um padrão de interpretação bastante simples, como já indicado nos resultados da HCA. De facto apenas um componente é suficiente para explicar mais de 90% da variabilidade total. Na Figura 21 representam-se as *loadings* para os dois primeiros componentes.

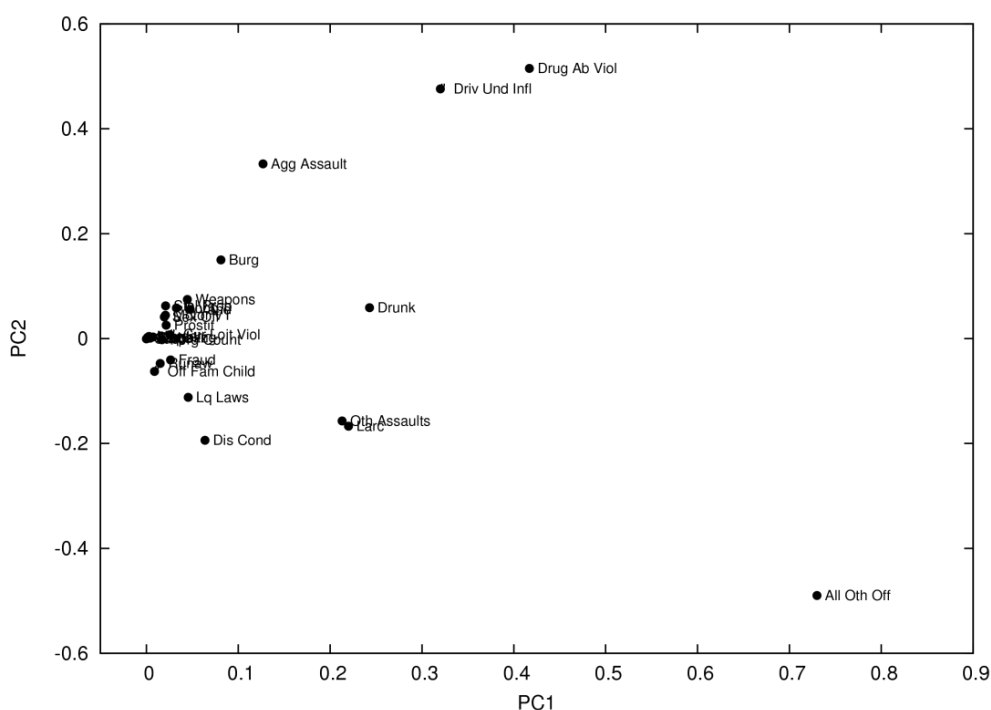


Figura 21. Representação das *loadings* obtidas por análise de componentes principais. Note-se que, embora por facilidade de representação constem os dois primeiros componentes, apenas o primeiro deve ser encarado como relevante.

É visível que todas as *loadings* do primeiro componente são positivas. Tal significa, simplesmente, que avançar ao longo desse eixo corresponde a evoluir de um menor número de crimes total para um maior número. Nota-se que, apesar de esta evolução não ser específica para determinadas variáveis, existem algumas de maior relevância para a caracterização dos dados. Citem-se entre estas *all other offenses*, *drug abuse violations* e

driving under influence. Tratam-se, simplesmente das que ocorrem com maior frequência em cada estado e que, portanto, dominam nesta análise.

Taxa de criminalidade

Recorrendo a uma definição de cada estado com base na respectiva taxa de criminalidade, são necessários três componentes para recuperar cerca de 80% da variabilidade original. Temos, assim, perto de 60% para o primeiro componente e 12% para o segundo. A existência de três componentes relevantes é um claro indicador de que o padrão obtido se reveste de maior complexidade do que na análise da frequência de crimes. Observação semelhante já foi feita aquando da análise de agrupamentos.

Na Figura 22 apresentam-se as *loadings* correspondentes ao perfil sobre análise.

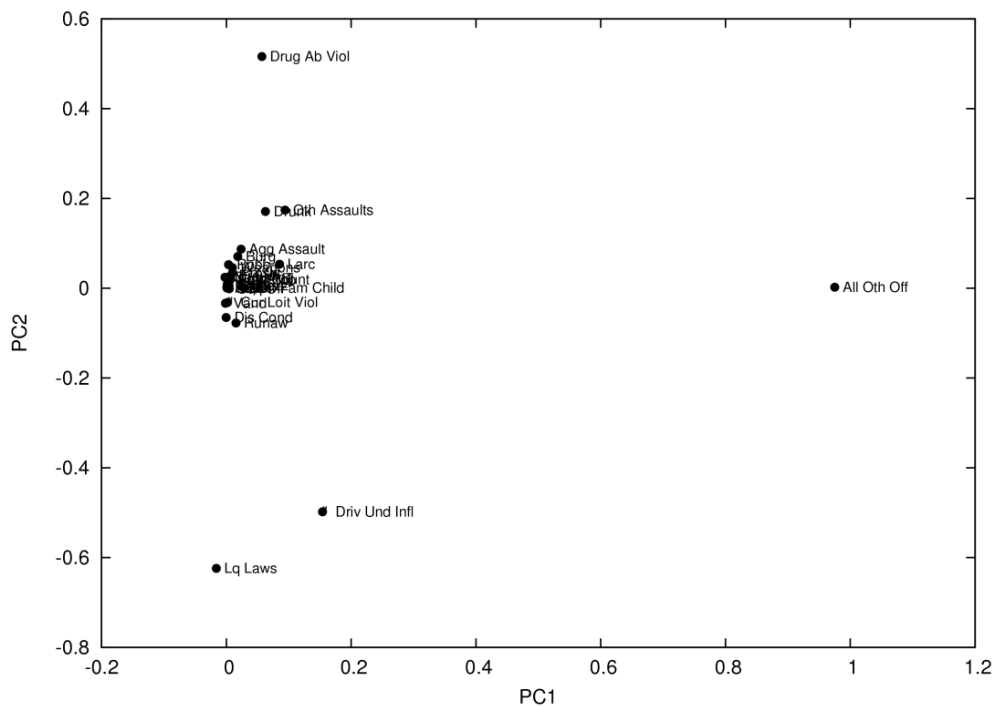


Figura 22. Representação das *loadings* obtidas por análise de componentes principais.

All other offenses dominam claramente na primeira componente enquanto na segunda se destacam crimes relacionados com *liquor laws*, *drug abuse violations* e *driving under influence*. Refira-se que a primeira e a última tem sinal oposto à restante.

Inspeccionemos agora a representação dos objectos no plano dos dois primeiros componentes, Figura 23. Por uma questão de clareza, já que se regista alguma sobreposição, estes grupos serão designados através da cor em que estão representados.

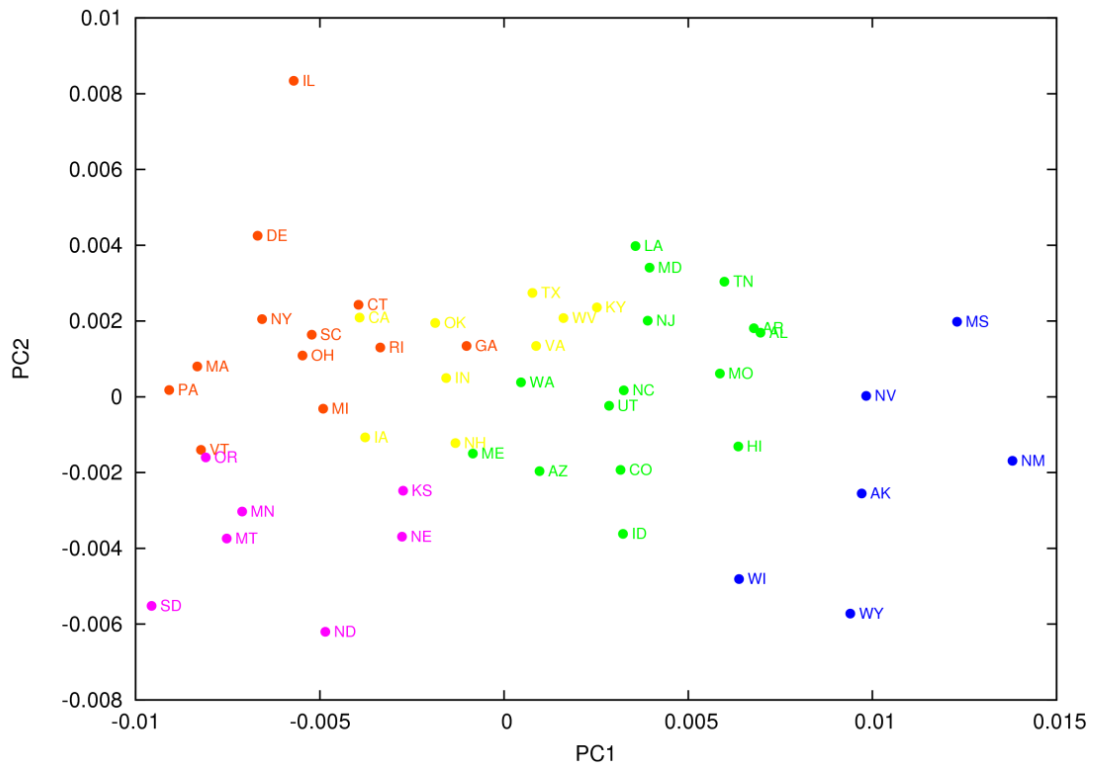


Figura 23. Representação dos objectos no plano PC1/ PC2

Nota-se claramente nesta Figura que a discriminação dos grupos implica a utilização das duas primeiras componentes. Assim sendo, do grupo laranja para o grupo rosa verifica-se um decréscimo das *drug abuse violations* e um aumento das infracções relacionadas com a bebida: *liquor laws* e *driving under influence*. Quando nos movemos destes dois grupos através do grupo amarelo, verde e azul regista-se um aumento das *all other offenses*. Estas são dominantes no grupo azul que apresenta, como já referido, uma grande dispersão a nível geográfico (cf. Figura 17). Esta dispersão pode estar relacionada com a falta de especificidade desta categoria.

Fracção de crime

Utilizando-se para cada estado um perfil resultante do valor relativo da frequência de cada crime, necessitamos de três componentes para recuperar perto de 80% da variabilidade original (na realidade apenas com 4 componentes se excede formalmente este valor). A primeira componente corresponde a cerca de 47%, obtendo-se 68% incluindo a segunda.

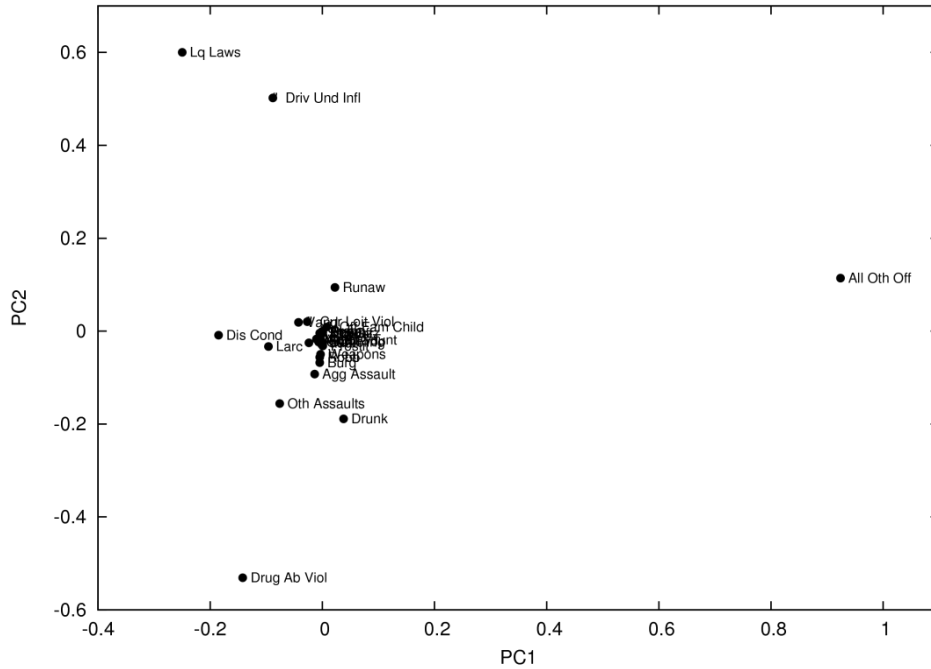


Figura 24. Representação das *loadings* obtidas por análise de componentes principais.

As *loadings* dominantes são muito semelhantes às encontradas no perfil anterior (taxa de criminalidade) com *all other offenses* a dominar na primeira componente e *drug abuse violations* e infracções relacionadas com *liquor laws* e *driving under influence* na segunda.

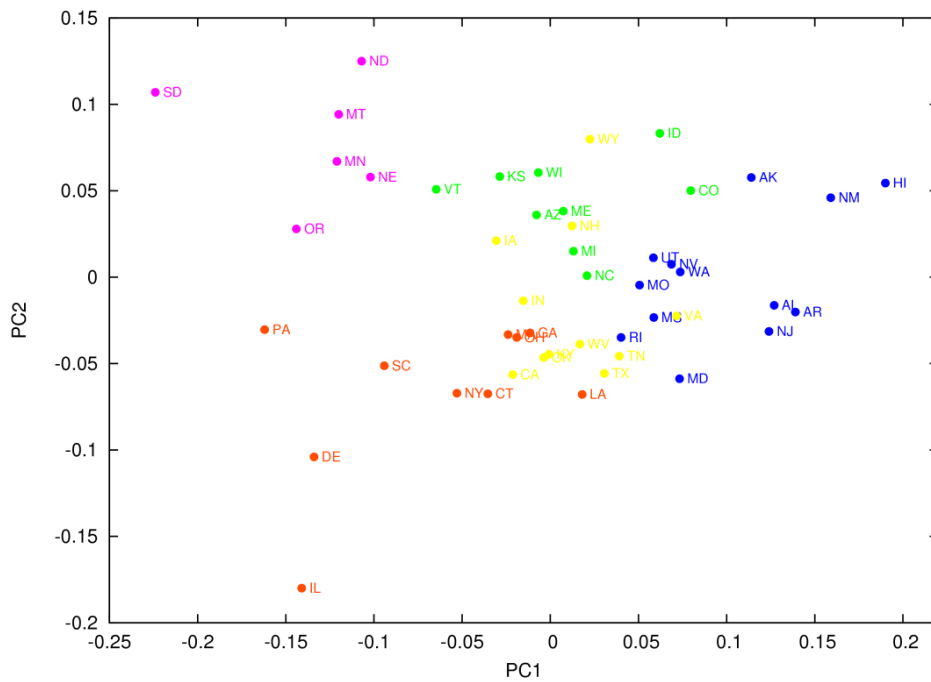


Figura 25. Representação dos objectos no plano PC1/ PC2

Uma vez mais o grupo laranja reflecte uma prevalência de *drug abuse violations* contra *liquor laws* e *driving under influence* no grupo rosa. A discriminação ao longo da primeira componente, como seria de esperar, é menos clara, registando-se uma mistura significativa entre os grupos amarelo e verde, com o azul caracterizado por uma maior intensidade das *all other offenses*. O grupo rosa mantém a coerência geográfica já observada nos perfis anteriores mas os restantes revelam bastante dispersão.

4.2. Aumentando a especificidade

Os resultados expostos na secção anterior revelaram, em qualquer dos três perfis utilizados, uma forte influência de uma variável pouco específica que junta infracções de ordem variada. Na classificação utilizada pelo FBI UCR surge uma outra variável deste tipo, *other assaults*. Vamos agora refazer a análise anterior, omitindo-as. Permitiremos assim que surjam outras tendências menos visíveis devido à sua magnitude, a qual é especialmente marcada para *all other offenses*.

4.2.1. Frequência

A análise hierárquica de agrupamentos não é sensível à omissão de *all other offenses* e *other assaults*. De facto, como este perfil assenta no valor absoluto dos crimes cometidos, os grupos anteriormente identificados mantêm-se na generalidade.

Também nos resultados PCA se mantém uma única componente, recuperando mais de 86% da variabilidade, pelas razões anteriormente aduzidas. No entanto, o padrão relativo às variáveis dominantes na caracterização sofre uma alteração drástica, embora as *loadings* se mantenham todas positivas para o primeiro componente, Figura 26.

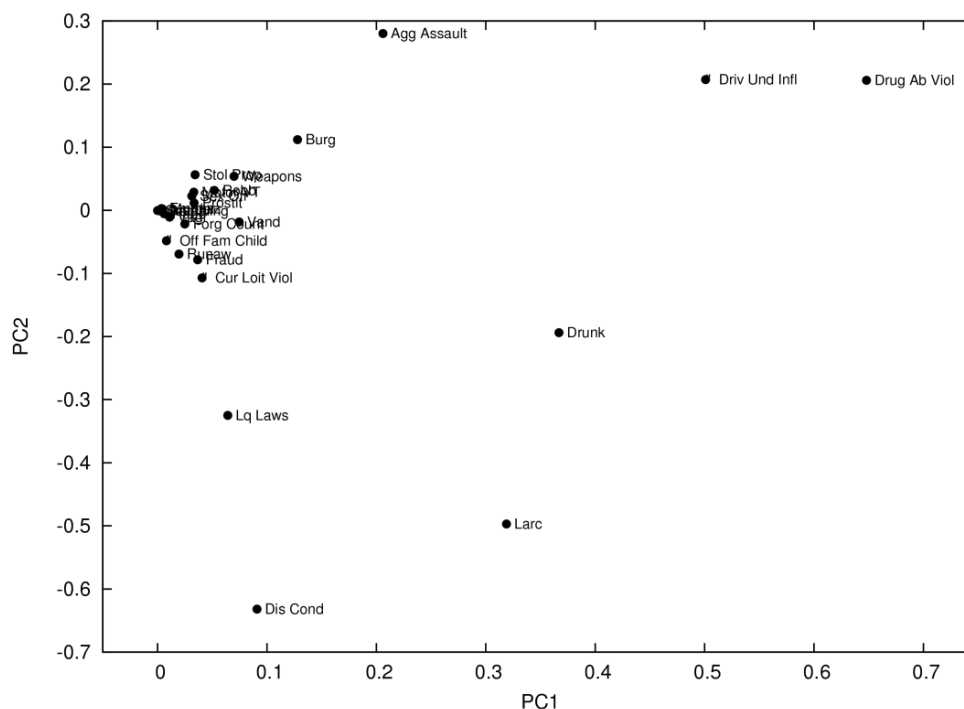


Figura 26. Representação das *loadings* obtidas por análise de componentes principais

Dominam claramente as *drug abuse violations* e *driving under influence*. Surgem depois, em ordem decrescente de importância, *drunkenness* e *larceny-theft*, *aggravated assault* e *burglary*. As restantes variáveis concentram-se para valores reduzidos da primeira componente.

4.2.2. Taxa de criminalidade

Para o perfil estabelecido pela divisão entre a frequência de cada tipo de crime e a população, as alterações registadas pela remoção das duas variáveis de menor especificidade são notórias logo ao nível da análise de agrupamentos. Relativamente à análise anteriormente efectuada e em que foram considerados todos os vinte e nove crimes, verifica-se agora algum crescimento na maior parte dos grupos e o isolamento de Illinois de todos os restantes estados.

No que respeita à análise de componentes principais é notório um aumento do número de componentes necessário para uma boa descrição do sistema. De facto, a primeira componente recupera apenas cerca de 33%, com um pouco mais de 20% para a segunda e 20% para a terceira. São necessárias 4 para exceder os 80%. O *scree plot* da Figura 27 indica também que é necessário um elevado número de componentes para uma descrição aceitável da informação original. Com efeito, só existe estabilização na

variabilidade a partir, aproximadamente, da sétima componente, sugerindo pois a utilização de seis componentes.

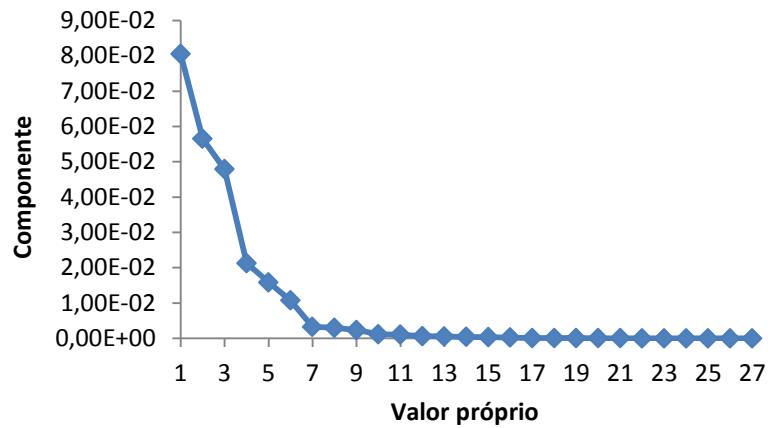


Figura 27. Scree plot para os resultados PCA obtidos com a matriz de variância covariância, para o perfil estabelecido pela divisão entre a frequência de cada tipo de crime e a população.

Quanto às *loadings*, dominam as *liquor laws* e *driving under influence* e, com sinal oposto, *drug abuse violations* e *drunkenness*. As restantes variáveis caracterizadoras são bastante menos significativas.

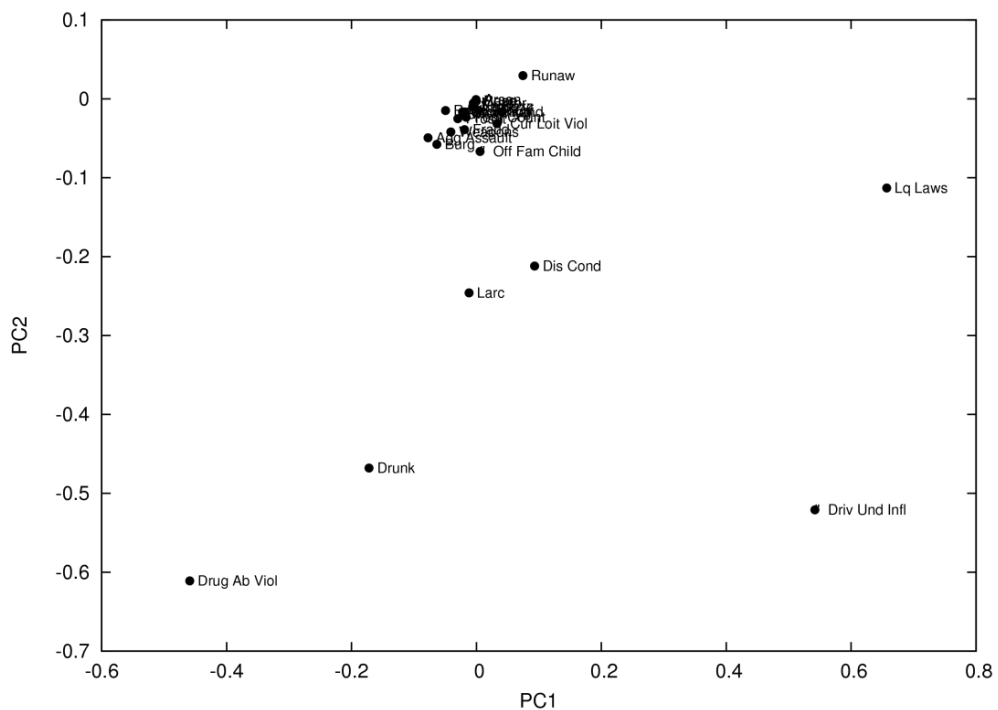


Figura 28. Representação das *loadings* obtidas por análise de componentes principais.

A representação geográfica dos grupos, se combinada com os resultados *PCA* indica-nos que possuímos um centro Norte em que dominam as infracções *liquor laws* e *driving under influence*, e que à medida que avançamos para a periferia, as *drug abuse violations* começam a dominar. *Illinois* aparece como um caso isolado, numa região em que as *drug abuse violations* são bastante relevantes. Trata-se do estado com o maior número de crimes deste último tipo por número de habitantes. A Figura 29 representa estes resultados.

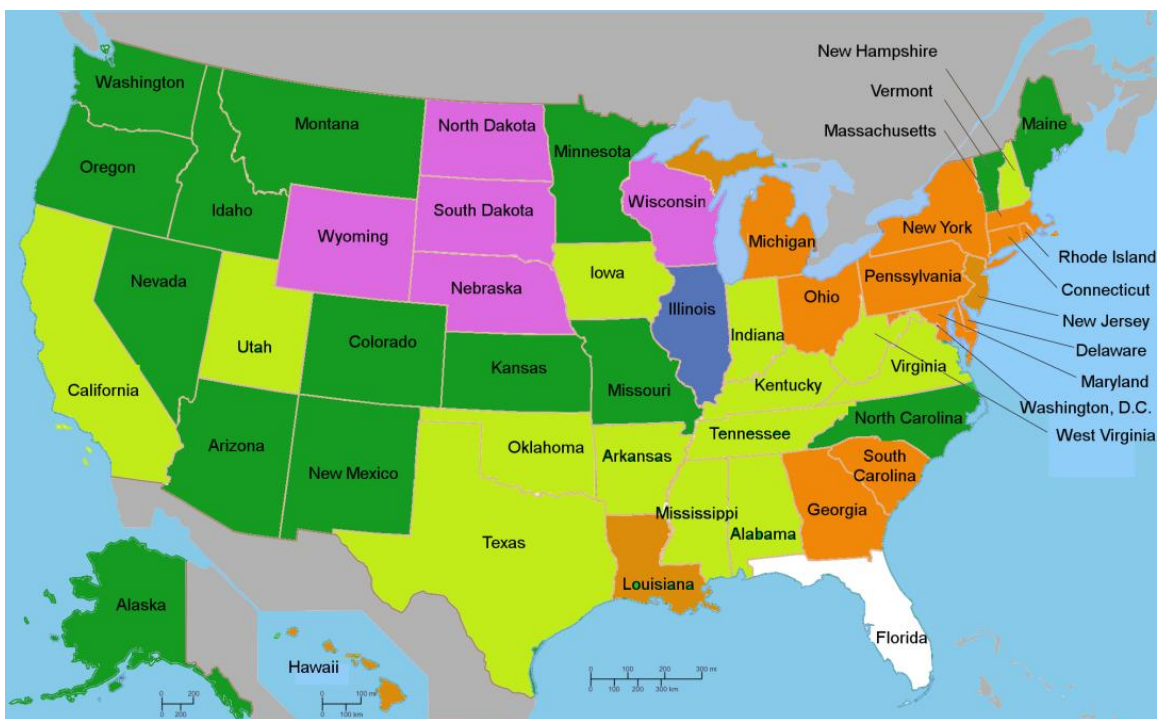


Figura 29. Padrões geográficos obtidos. Representação sobre o mapa dos Estados Unidos dos grupos formados na análise de *HCA*.

4.2.3. Fracção

Para o perfil correspondente à fracção de cada um dos crimes cometidos, vamos proceder novamente a uma análise conjunta *HCA/PCA*. A Figura 30 representa os *scores* no plano *PC1/PC2*, indicando os grupos extraídos de *HCA* sob a forma de cores.

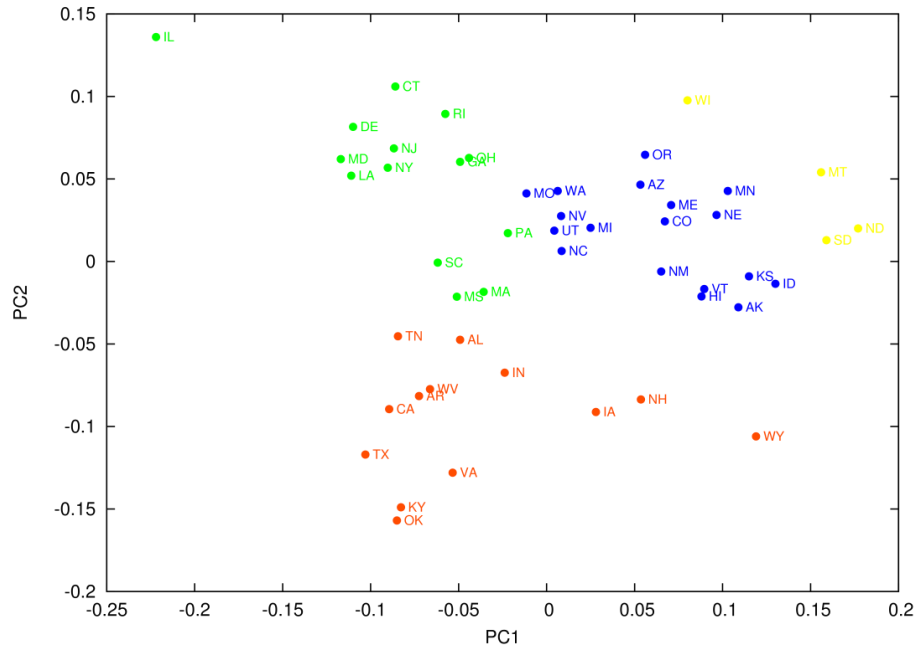


Figura 30. Representação dos scores no plano PC1/PC2 para o perfil correspondente à fração de crime cometido em cada estado.

Existem quatro grupos com três deles (verde, azul e amarelo) evoluindo ao longo da primeira componente, enquanto o quarto, vermelho, se distingue dos restantes três pela sua situação na parte negativa do eixo correspondente à segunda componente. *Illinois* surge no grupo verde, mas claramente isolado.

Observemos, de seguida a Figura 31 que nos indica as *loadings* correspondentes.

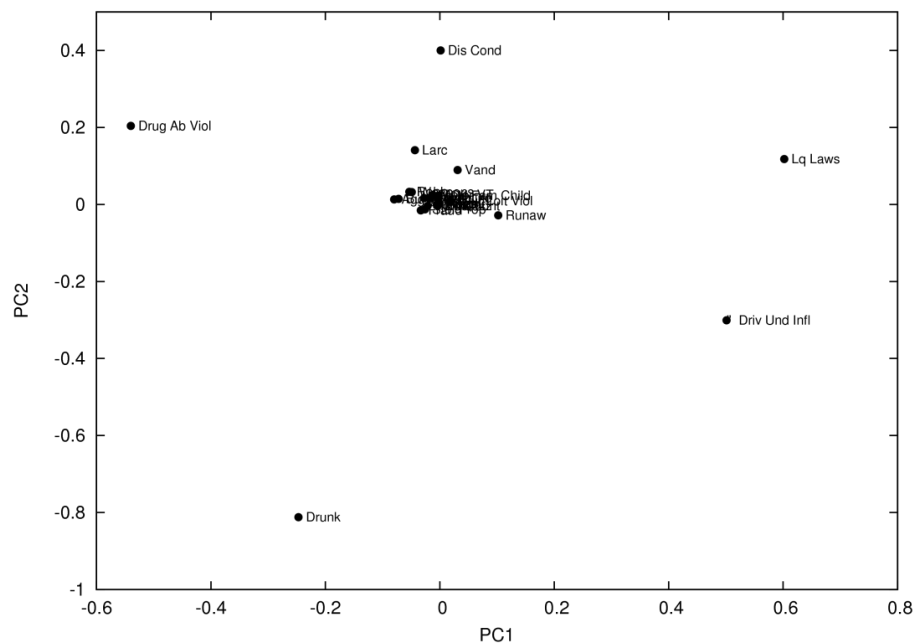


Figura 31. Representação das *loadings* obtidas por análise de componentes principais.

Os três primeiros grupos referidos caracterizam-se por uma evolução em que as *drug abuse violations* vão dando lugar às infracções relacionadas com as *liquor laws* e as *driving under influence* numa evolução ao longo do eixo PC1. Por outro lado a evolução dos grupos (verde, azul e amarelo) para o grupo vermelho faz-se pelo agravamento de infracções relacionadas com a bebida (*drunkenness*).

Illinois está numa posição particular devido à elevada fracção de crimes associados com a droga. Note-se que esta situação já tinha sido detectada anteriormente na análise correspondente sobre a taxa de criminalidade.

4.3. Aumentando a gravidade

Vamos agora proceder a uma redução ainda mais drástica do número de tipos de crimes a considerar na análise. Globalmente, tal corresponderá a incidir sobre crimes de maior gravidade. Esta redução permite-nos, ainda, uma mais fácil ligação ao capítulo seguinte em que a análise é baseada no número de infracções conhecidas, ao contrário da base utilizada até aqui que corresponde ao número de detenções efectuadas. Serão, pois, apenas levados em consideração 16 crimes: *murder and nonnegligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny-theft, motor vehicle theft, arson, forgery and counterfeiting, fraud, embezzlement, stolen property (buying, receiving, possessing), weapons (carrying, possessing), prostitution and commercialized vice, sex offenses* e *drug abuse violations*.

Como temos verificado que os resultados que surgem da utilização do número de crimes em bruto têm sido semelhantes, quaisquer que sejam os crimes seleccionados, focaremos a nossa atenção essencialmente nos perfis correspondentes à taxa de criminalidade e fracção de crime.

Analisemos de seguida os padrões sugeridos pela taxa de criminalidade com os 16 crimes seleccionados.

4.3.1. Taxa de criminalidade

Os resultados da análise de PCA indicam a existência de dois componentes para se atingir um total de 88% da variabilidade original. No entanto, o primeiro componente é dominante e corresponde a perto de 72%. Nesta componente destaca-se uma variável, *drug abuse violations*. Assim, temos a sucessão, na Figura 32, do grupo rosa, vermelho, azul, verde e o estado de *Illinois* (amarelo) num crescendo da taxa de criminalidade referente a

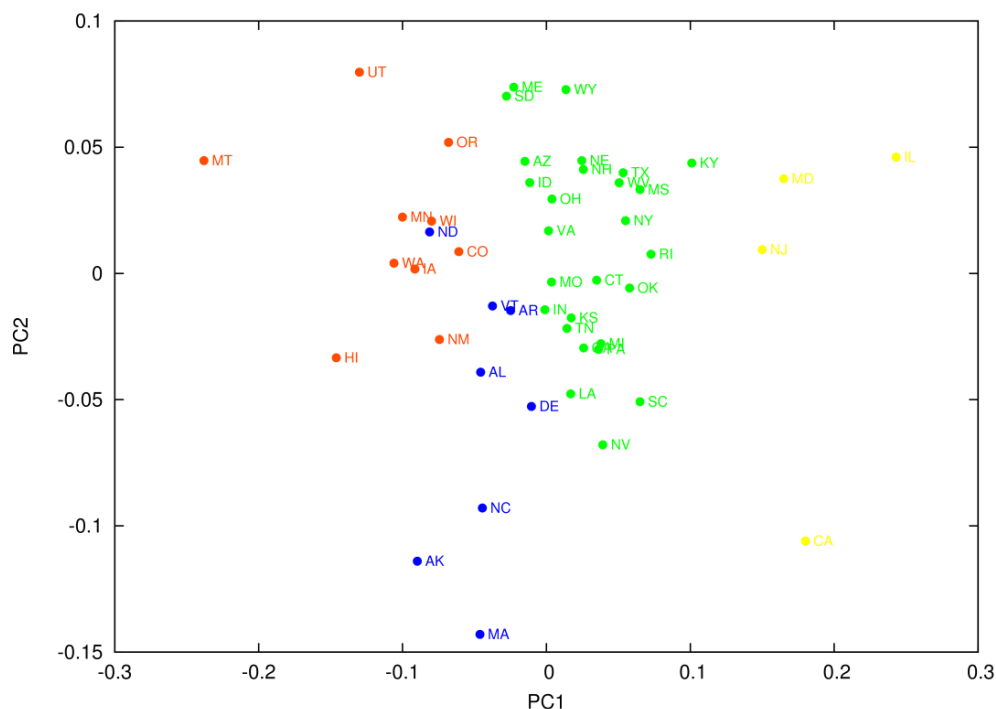


Figura 33. Representação dos *scores* no plano PC1/PC2 para o perfil correspondente à fracção de crime cometido em cada estado.

4.3.3. Cálculo alternativo

Até ao momento, as análises de componentes principais que produzimos assentam na matriz de variância/covariância. Esta dá ênfase às variáveis dominantes, em valor absoluto, e como tal são fáceis de identificar as principais *loadings* e o número de componentes necessário para a recuperação de uma quantidade apreciável da variância original é diminuto. Por outras palavras, esta análise conduz a resultados mais facilmente interpretáveis. Tem, no entanto, uma deficiência óbvia: atribui-se uma importância automaticamente menor a certas variáveis caracterizadoras, dado que variam apreciavelmente de estado para estado, mas que estão associadas a valores absolutos diminutos.

Sem pretensão de sermos exaustivos vamos apresentar, de seguida, um exemplo do que se poderia obter com base numa análise que recorra à matriz de correlação para o perfil definido pela fracção de crimes, no qual se considera o menor conjunto, 16 crimes. Em primeiro lugar, surgem seis valores próprios superiores à unidade, o que sugere a utilização de seis componentes principais. Estes valores próprios tomam, por ordem decrescente, os valores 3.1, 2.6, 1.8, 1.5, 1.3 e 1.1. Nota-se, pois, uma contribuição comparável de todos os correspondentes componentes.

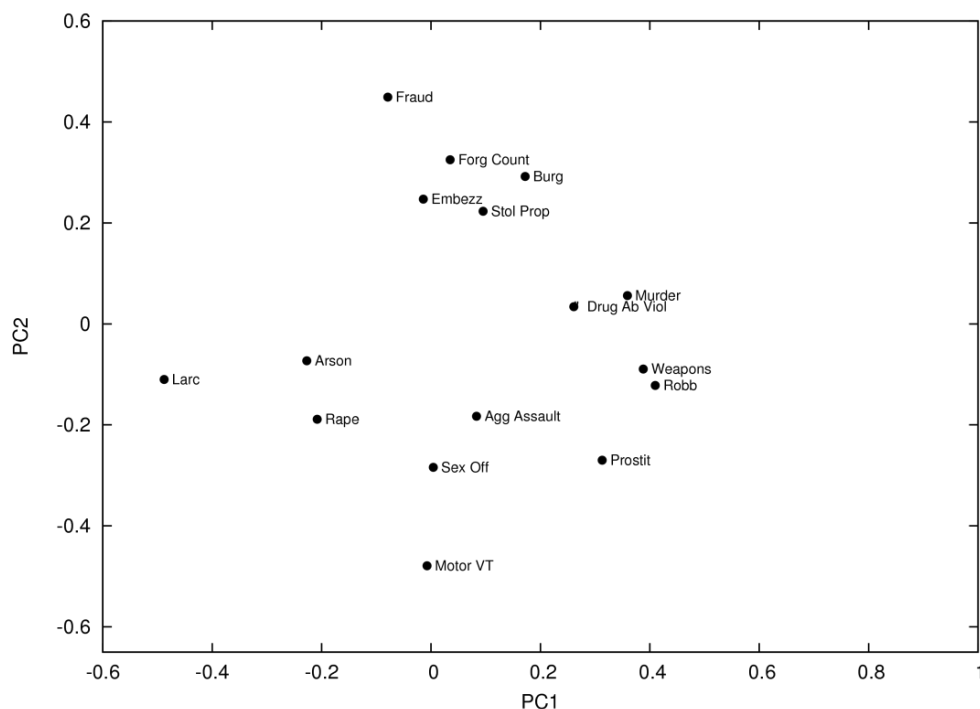


Figura 34. Representação das *loadings* obtidas por análise de componentes principais com a matriz de correlação para o perfil definido pela fracção de crimes, no qual se considera o menor conjunto, 16 crimes.

Observando a Figura 34 verificamos também que o número elevado de variáveis contribui de modo apreciável para o primeiro e segundo componentes. Vejamos o que sucede para o primeiro. No sentido negativo do eixo surge *larceny-theft*, algo destacado, que se opõe a um conjunto formado pelo menos por três variáveis, *murder*, *weapons* e *robbery*. Estas três variáveis já não podem ser consideradas como destacadas daquelas que as precedem. Esta situação pode ser interpretada como resultante da substituição do *larceny-theft* por crimes mais violentos, à medida que avançamos ao longo do primeiro componente. Refira-se que esta evolução tem, na contrapartida da representação de *scores*, *Montana* no início do eixo e *California* e *Illinois* no outro extremo, confirmando observações feitas anteriormente. Para o segundo componente, alguns crimes económicos e contra a propriedade, opõem-se ao *motor vehicle theft*, numa tendência menos trivial.

Apesar de verificarmos que esta nova análise proporciona informação acrescida, temos de concluir que a sua interpretação se reveste de maior complexidade. Não sugerimos que não deva ser considerada, mas no contexto desta dissertação apresentaremos apenas esta ilustração.

4.4. Factores

Sem pretendermos atingir uma profundidade excessiva, analisemos alguns aspectos sugeridos pela rotação *varimax* sobre os quatro principais componentes correspondentes a vinte e nove crimes, com o perfil dado pela fracção de cada tipo de crime. Indica-se na Tabela 1 o eixo sobre o qual se encontra a componente dominante de cada loading.

Tabela 1. Componente dominante da *loading PCA* após rotação *varimax*. A numeração dos eixos é arbitrária.

Eixo	Delito
1	All other offenses; Curfew and loitering law violations
2	Murder and nonnegligent manslaughter; Forcible rape; Robbery; Aggravated assault; Burglary; Motor vehicle theft; Weapons; carrying, possessing, etc; Prostitution and commercialized vice; Sex offenses; Drug abuse violations; Gambling; Liquor laws; Suspicion; Runaways
3	Arson; Stolen property; buying, receiving, possessing; Vandalism; Drunkenness; Disorderly conduct
4	Larceny-theft; Other assaults; Forgery and counterfeiting; Fraud; Embezzlement; Offenses against the family and children; Driving under the influence; Vagrancy

Sobre o eixo 1 projectam-se apenas 2 variáveis, *curfew and loitering law violations* e *all other offenses*. Esta última é pouco específica e a primeira corresponde a um delito de menor gravidade. Depois, surge um eixo 2 que reúne crimes de maior gravidade, embora sobre ele constem também alguns outros cuja gravidade é um pouco menor ou menos bem definida. No eixo 3 encontram-se três delitos contra a propriedade, *arson*, *stolen property* e *vandalism*, assim como dois relacionados com a bebida e conduta desordeira. Finalmente, no eixo 4, reúnem-se crimes de carácter económico, nomeadamente *forgery and*

counterfeiting, fraud e embezzlement. No entanto, outros delitos aparentemente não relacionados surgem também neste eixo, como sejam *driving under influence* e *offenses against family and children*. O relevante *larceny-theft* encontra-se também sobre este eixo. De uma maneira simplista, e salvaguardando várias exceções, podemos sugerir que associado ao eixo 2 está o factor violência, acompanhado ou não de outros factores, nomeadamente o económico. O eixo 3 parece agrupar factores subjacentes à destruição de propriedade. Finalmente o eixo 4 é de carácter económico.

4.5. Perspectiva geral e discussão

Os resultados das secções anteriores, devido à utilização de vários conjuntos de variáveis descritivas (29, 27 e 16 crimes respectivamente) e três diferentes perfis (o número de crimes, a taxa de criminalidade e a fracção de crimes, todas expressas por tipo de crime) tornam complicada uma análise global. No entanto, iremos no que se segue resumir algumas das tendências identificadas e algumas conclusões mais gerais.

Em primeiro lugar, tornou-se patente que o tipo de perfil escolhido para a caracterização é determinante no tipo de resultados obtidos. Assim sendo, quando utilizamos como perfil o número total de crimes encontramos uma distribuição de estados aparentemente governada por factores geográficos. Se tivermos em consideração que o número de crimes é, *grosso modo*, proporcional à população torna-se à primeira vista estranho este padrão. Na realidade, o seu aparecimento tem precisamente a ver com uma população que aumenta a partir dos estados centrais do Norte em direcção à periferia. Não nos parece que os tipos de crime utilizados na descrição alterem significativamente este padrão geral.

Se contrastarmos a Figura 13 com a Figura 35, que nos apresenta a população por estado, verificamos uma semelhança notória que confirma a conclusão anterior. Verifique-se, por curiosidade, que também o produto interno bruto (PIB) estadual segue uma distribuição muito semelhante, Figura 36. Uma conclusão directa relacionando produção com crime não é, no entanto, muito verosímil. Crime e PIB crescem com a população e têm, portanto, uma relação causal comum com esta.

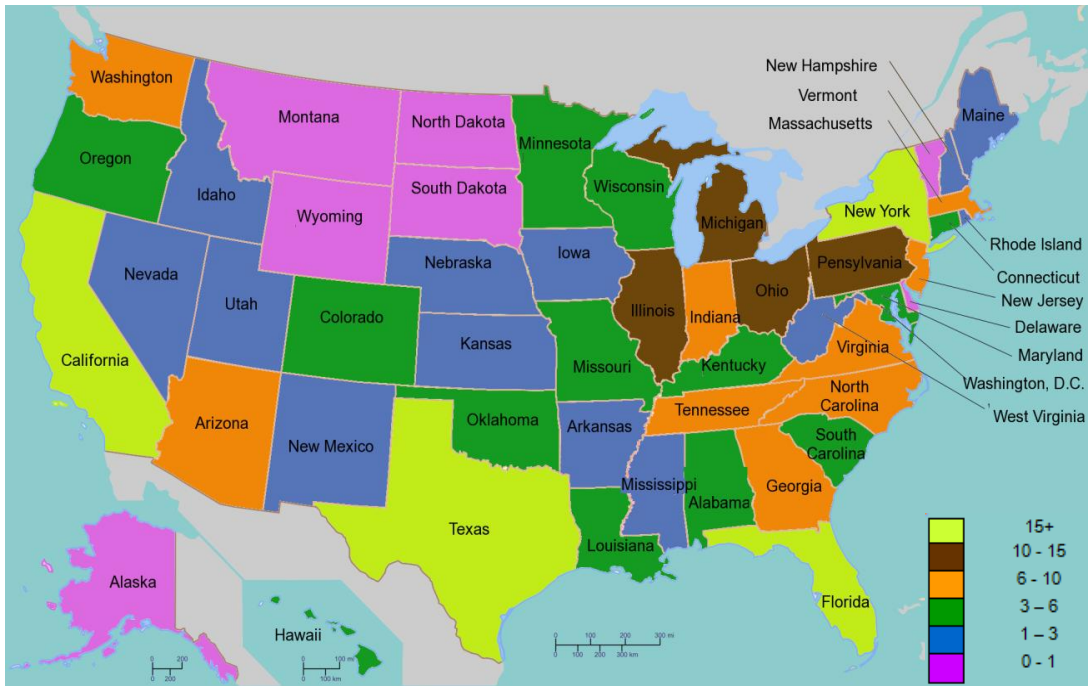


Figura 35. Representação da população, milhões de pessoas, por estado

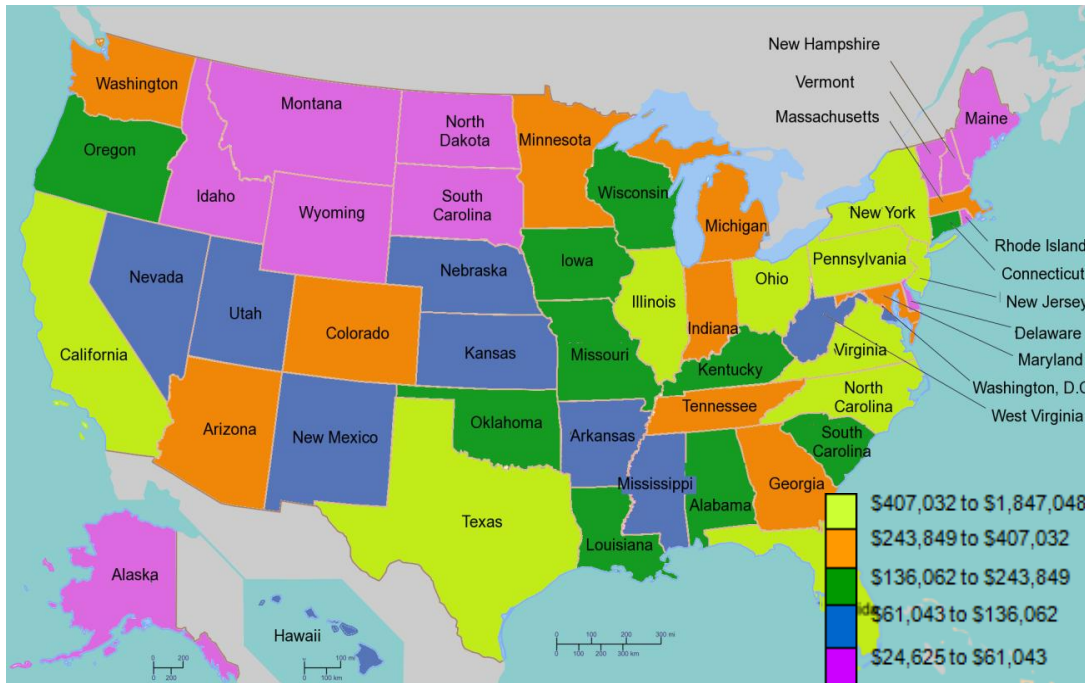


Figura 36. Representação do PIB, (milhões de dólares) por estado.

Quando utilizamos a taxa de criminalidade a situação altera-se, já que o tipo ou tipos de crimes dominantes na caracterização evolui. Com efeito, este perfil, ao contrário do anterior, não assenta numa variação global em que o tipo de crime é irrelevante. Para a totalidade dos vinte e nove crimes, a caracterização é, em grande medida, feita com base

nas *all other offenses*. Trata-se de uma categoria que engloba vários tipos de crime o que, ao mesmo tempo, lhe retira a especificidade. Quando se descarta esta e uma outra variável do mesmo tipo (*other assaults*) verificamos que começam a surgir as *drug abuse violations* e várias infracções relacionadas com o consumo de álcool como dominantes. Misturam-se, no entanto, com *larceny-theft* ou *aggravated assault* num panorama à primeira vista pouco claro. Apesar disso, tem sido intensamente estudado o problema do consumo de droga ligado à violência e crime [98], que são tomados como fortemente relacionados. Note-se ainda que o crime violento e as drogas têm sido considerados como os problemas mais importantes que os Estados Unidos têm de enfrentar [35]. Este tipo de relação entre droga e crimes contra a propriedade parece ser essencialmente resultante da motivação económica, devido ao elevado custo das drogas ilícitas. No entanto outras explicações têm sido oferecidas, nomeadamente a de que comportamentos desviantes que culminam no consumo de drogas e ofensas criminais, ocorrem num contexto de síndrome geral de desvio comportamental. Registe-se, no entanto, que a pesquisa tem mostrado que a principal actividade de muitos consumidores de droga é a sua venda. Estas questões estão discutidas em detalhe na referência [98]. Refira-se, ainda, que nem sempre o crime está associado ao consumo, estando muitas vezes relacionado com a distribuição e outros aspectos. Um facto interessante apresentado num estudo realizado em *New York*, em 1989, foi a apresentação da conclusão de que apenas 5% dos homicídios relacionados com a droga eram um resultado do seu consumo, enquanto três quartos da totalidade eram cometidos por pessoas que roubavam para pagar drogas, estavam envolvidos em disputas territoriais ou pretendiam fazer a colecta de dívidas [52]. Na mesma linha, um outro estudo mostrou que não havia uma clara evidência de relação entre o consumo de droga e o número de homicídios e outros crimes violentos. Em contraste, uma revisão recente de um número elevado de estudos mostrou [99], que a probabilidade de cometer um crime é três a quatro vezes maior para os consumidores do que para os não consumidores. Verificou-se, ainda que o tipo de droga consumida também influencia esta estatística.

Este assunto apresenta-se, portanto, como extremamente complexo, levando um autor a designá-lo como “a puzzle inside an enigma” [100].

Note-se que é também frequentemente assumido que o consumo de álcool conduz à violência. Vários estudos mostraram a associação, não necessariamente causal, entre o uso do álcool e a violência assim como um elevado consumo de álcool pelas pessoas que cometem os crimes violentos [52]. A relação entre consumo de álcool e crime tem, inclusivamente, levado a que seja examinado o impacto sobre o crime de leis restritivas de venda, em certos dias da semana [101].

Reduzindo para os 16 crimes de maior gravidade, o padrão torna-se bastante mais claro, com uma variável dominante, *drug abuse violations*, e uma outra, em posição ortogonal, *larceny-theft*. Embora não se possa tirar uma conclusão definitiva, este posicionamento relativo sugere que estes dois crimes assentam em factores distintos.

Finalmente, quando é considerada a fracção de cada crime, voltamos a detectar o aparecimento, quase dominante, de *all other offenses* quando os vinte e nove crimes são considerados. Quando reduzimos para vinte e sete verificamos, mais claramente, o aparecimento de uma característica directamente resultante deste tipo de perfil. É que, tratando-se de uma fracção, o aumento da relevância de um ou mais crimes é sempre feito à custa da perda de importância de outros. Por exemplo, para o eixo PC1, *drug abuse violations* e infracções relacionadas com as *liquor laws* encontram-se com variações opostas, facto geralmente ausente da literatura. Este tipo de contra variação não tem necessariamente de surgir quando são considerados quer o número total de crimes quer a sua taxa. Aumentando a gravidade, isto é, considerando apenas o conjunto de dezasseis crimes, *drug abuse violations* e *larceny-theft* dominam, variando em oposição quando a fracção de crimes é considerada. Novamente, aqui mostra-se que não há uma relação directa entre as infracções relacionadas com droga e este tipo de atentado à propriedade.

A análise permitiu, também, identificar situações características em três estados: *Texas* e *California*, apresentados como semelhantes em várias das alternativas de abordagem, e *Illinois*. Este último é de explicação mais fácil e por ele iremos começar. Este estado caracteriza-se por um número de crimes relacionado com *drug abuse violations* muito significativo, tendo em conta a sua população. Já no que respeita a *Texas* e *California*, só apresentam uma posição destacada relativamente aos outros estados quando é considerado o número total de crimes. De facto, quando a descrição é baseada na taxa de crimes ou na fracção de crimes ficam englobados em grupos bastante mais vastos e podem até não pertencer necessariamente ao mesmo grupo.

Foquemos agora a nossa atenção, como ilustração, no perfil definido pela fracção de crimes considerando apenas os 16 crimes mais graves. Tínhamos identificado, nos resultados anteriormente apresentados, que *larceny-theft* e *drug abuse violations* eram as duas variáveis mais caracterizadoras. Concluímos também que uma redução na primeira correspondia a um aumento da segunda. A Figura 37 ilustra esta evolução para *Montana*, *Missouri* e *Illinois*. O primeiro encontra-se num extremo do eixo caracterizador, *Missouri* num ponto intermédio e *Illinois* no outro extremo (cf. Figura 33).

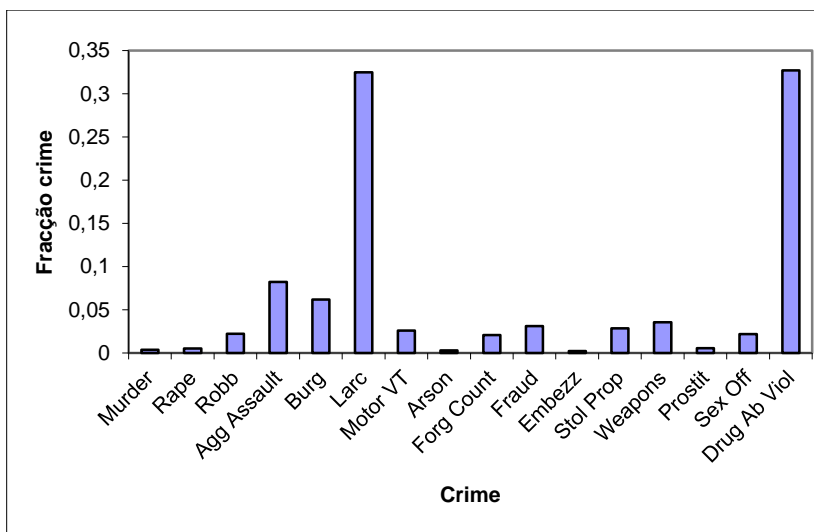
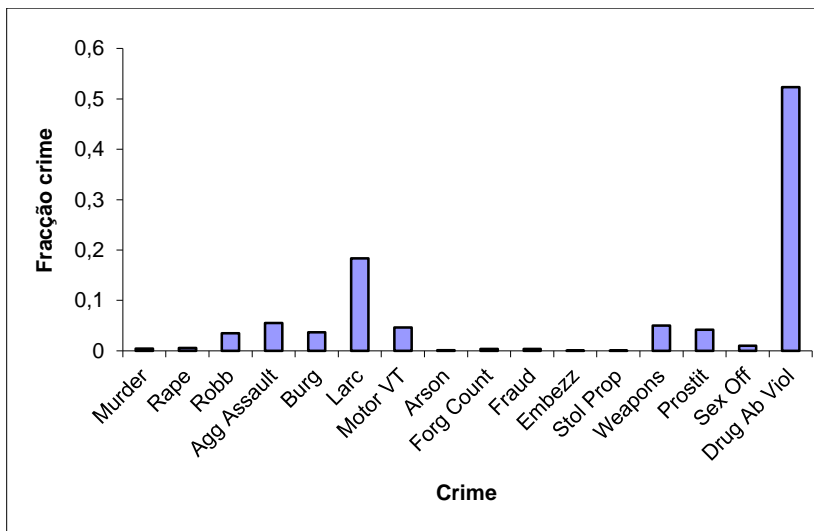
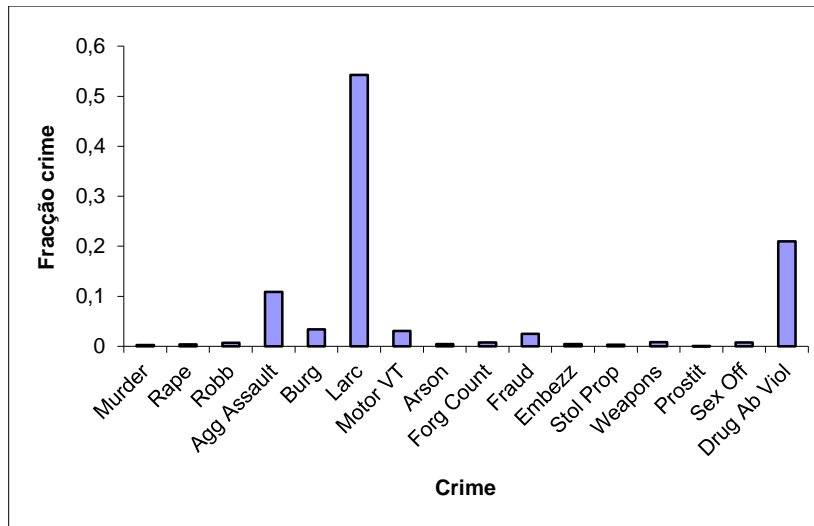


Figura 37. Perfil definido pela fracção de crimes considerando apenas os 16 crimes mais graves. Evolução ilustrada para *Montana* painel (a), *Missouri* painel (b) e *Illinois* painel (c).

Verifica-se que as barras correspondentes a *larceny-theft* e *drug abuse violations* dominam, correspondendo a frações que podem exceder os 50% e dando conta, em conjunto, de cerca de 70% dos crimes. A utilização dos dados FBI UCR desde os anos noventa, mostra uma tendência geral de decréscimo nos casos de *larceny-theft* enquanto regista um aumento nos casos de *drug abuse violations*, o que vem, em certa medida, e numa análise temporal confirmar as nossas observações.

Note-se também, um ainda notório *aggravated assaults*, com muito menor prevalência. Lembre-se que esta variável surge como bastante importante num pouco relevante PC2. Ilustrações do mesmo tipo seriam encontradas para os restantes sistemas considerados.

A importância das *drug abuse violations* é bem conhecida, embora a identificação geográfica da intensidade do problema seja menos clara. No entanto, já em 1996 tinha sido estabelecido que 45% dos detidos no estado de Illinois estavam de acordo com critérios que permitiam estabelecer uma dependência vitalícia da droga [102]. Posteriormente, em 2008, o *Arrestee Drug Abuse Monitoring Program*, mostrou que 87% das pessoas detidas em Chicago testaram positivamente para o uso de drogas! Por outro lado, *Montana* está englobado num grupo de estados essencialmente rurais em que os problemas associados a drogas parecem ser menos severos. Ao longo da secção temos identificado *Montana* fazendo frequentemente grupo com *North Dakota* e *South Dakota*, por exemplo, com o mesmo tipo de características, já previamente assinaladas [102].

A análise patente neste capítulo permitiu, portanto, não só descrever como apresentar relações que surgiam em trabalhos dispersos na literatura, o que parece justificar a nossa abordagem sistemática.

Capítulo 5

Taxa de criminalidade e população

5.1. Contexto

Verificou-se, na secção anterior, que uma enorme parte do esforço desenvolvido em criminologia assenta na identificação de factores. Temos, assim, um sistema em que as respostas correspondem a número de crimes ou uma outra quantidade associada, com variáveis independentes escolhidas de um modo mais ou menos criterioso. Dentro das respostas surgem o número de crimes, aquilo que na literatura anglo-saxónica é designado como *counts*, e a taxa de criminalidade, designada por *crime rate*, ambos já abordados no capítulo anterior. Estes indicadores podem ser globais, considerando todos os delitos, ou podem ser estabelecidos após distinções várias, por exemplo, crime violento e crime contra a propriedade.

Note-se que a escolha da variável dependente tem sido extremamente controversa, tendo sido posta em causa, por exemplo, a validade teórica da utilização de quocientes como os que surgem na taxa de criminalidade. Apesar disso, a taxa de criminalidade representa o número de delitos numa base *per capita*, de tal maneira que se podem fazer comparações entre jurisdições correspondentes a populações diferentes [103], o que estabelece a sua utilidade.

É um facto bem estabelecido que o volume total de crime correlaciona bem com o tamanho da população sobre a qual é determinado. A Figura 38 ilustra claramente este facto, apesar de uma notória curvatura presente para populações mais elevadas.

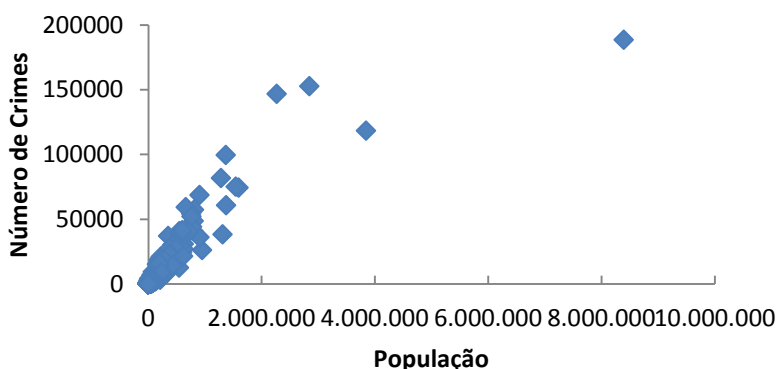


Figura 38. Representação do crime total vs. população, para todas as cidades de todos os estados.

No entanto, a relação entre taxa de criminalidade e tamanho da população é bastante menos clara. Genericamente, considera-se que existe uma evidência considerável, baseada na informação policial, de vitimização, etc., que substancia uma relação positiva entre a população e o crime *per capita*, apesar de algumas excepções [104-105]. Estas surgem para certos delitos, [106-108], em certos pontos temporais, [109-110], ou em certos locais [111-113], tendo, também sido apontada a existência de relações curvilíneas [114]. Foi também sugerido que a dependência pode estar a enfraquecer [104, 115]. Apesar de tudo, a relação positiva entre tamanho de uma povoação e o crime é, até, frequentemente citada como sendo um dos factos da criminologia,[116-119]. As contradições são, no entanto, muitas e, como sumariado na referência [120] a análise multivariada revela que enquanto que o tamanho da população não tem um efeito apreciável na taxa de criminalidade violenta e na taxa de criminalidade dos crimes contra a propriedade, continua a ser, de longe, o melhor factor de previsão para o número de crimes violentos ou contra a propriedade.

Mesmo quando se considera que a taxa de criminalidade aumenta com o tamanho da população, têm sido registadas discrepâncias entre as chamadas análises de secção recta e as longitudinais. Nas primeiras, faz-se o congelamento temporal e analisam-se as várias jurisdições. Espera-se, caso a premissa seja verdadeira, que a representação da taxa de criminalidade vs. população tenha um declive positivo. Na evolução longitudinal, isto é, em função do tempo, espera-se, em contrapartida, que um aumento na população produza um aumento, eventualmente com decalagem, na taxa de criminalidade. Autores que verificaram o comportamento para um determinado espaço temporal, não conseguiram identificar a evolução temporal referida [119]. Um outro aspecto tem a ver com o facto de, aparentemente, a dependência da taxa de criminalidade na população ser variável, dependendo da gama populacional considerada. Numa abordagem em que se focaram grupos de cidades desde as menos populosas até às de elevada população, verificou-se que apenas para estas últimas poderia a taxa de criminalidade decrescer com a população [103].

Verifica-se, então, que os resultados e tratamentos apresentados na literatura são díspares, por vezes mesmo contraditórios e muito raramente conclusivos. Este assunto será por nós abordado a partir da secção 5.3. O propósito é estabelecer, de um modo tanto quanto possível inequívoco, a relação funcional entre taxa de criminalidade e população.

5.2. Construindo o perfil: da nação ao estado

Um dos problemas associados com o estabelecimento de uma dependência funcional da taxa de criminalidade na população tem a ver com a enorme dispersão que a primeira apresenta. Na Figura 39 surge a distribuição da taxa de criminalidade para povoações não excedendo 10 000 habitantes. Verifica-se que esta cobre uma gama de valores entre 0 e 0.1. Estamos apenas a inspeccionar povoações com população relativamente reduzida e temos, desde já, alguma dificuldade em situar a análise.

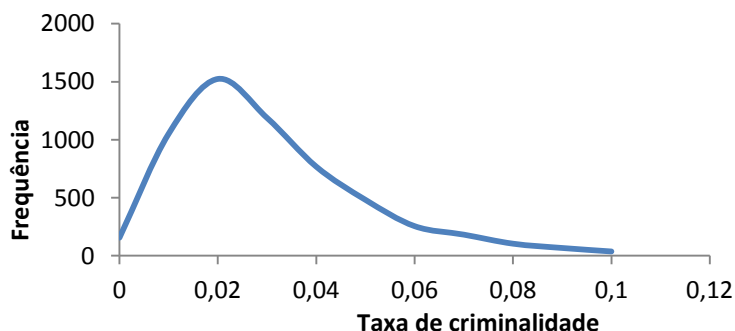


Figura 39. Distribuição da taxa de crime para povoações não excedendo 10 000 habitantes

Vamos, portanto, proceder de um modo sistemático que nos permita obter indicadores da variação da taxa de criminalidade. Para tal vamos procederemos à aglomeração de povoações, definindo para cada aglomerado i uma taxa global T calculada a partir das j povoações que o constituem.

$$T_j = \frac{\sum_j nc_j}{\sum_j pop_j} \quad (27)$$

onde nc_j e pop_j são respectivamente o número de crimes e a população de cada agregado populacional considerado. A divisão em aglomerados será, adicionalmente, baseada numa divisão, tanto quanto possível equitativa, da população total sob análise e que, no que se segue corresponderá à de cada estado ou à população total dos Estados Unidos. Para tal, estabeleceremos um número total n de agregados, e dividiremos a população total por esse número, de tal maneira que, para o agregado j teremos

$$\sum_j pop_j \approx \frac{população\ total}{n} \quad (28)$$

A sequência iniciar-se-á com $j = 1$ para as cidades de menor população, indo aumentando até que em $j = n$ juntaremos as de maior população que verificam a Equação (28). Naturalmente que esta divisão é apenas aproximada, especialmente para o último aglomerado em que a existência de grandes centros populacionais não permite uma divisão exacta.

Analisemos então a Figura 40. Nela está representada a taxa de criminalidade, como definida pela Equação (27), em função da média populacional para o aglomerado considerado. Esta representação utiliza, adicionalmente, $n = 10$ no painel (a) e $n = 5$ no painel (b). A Figura recorre a todos os 9141 aglomerados constantes na base de dados UCR.

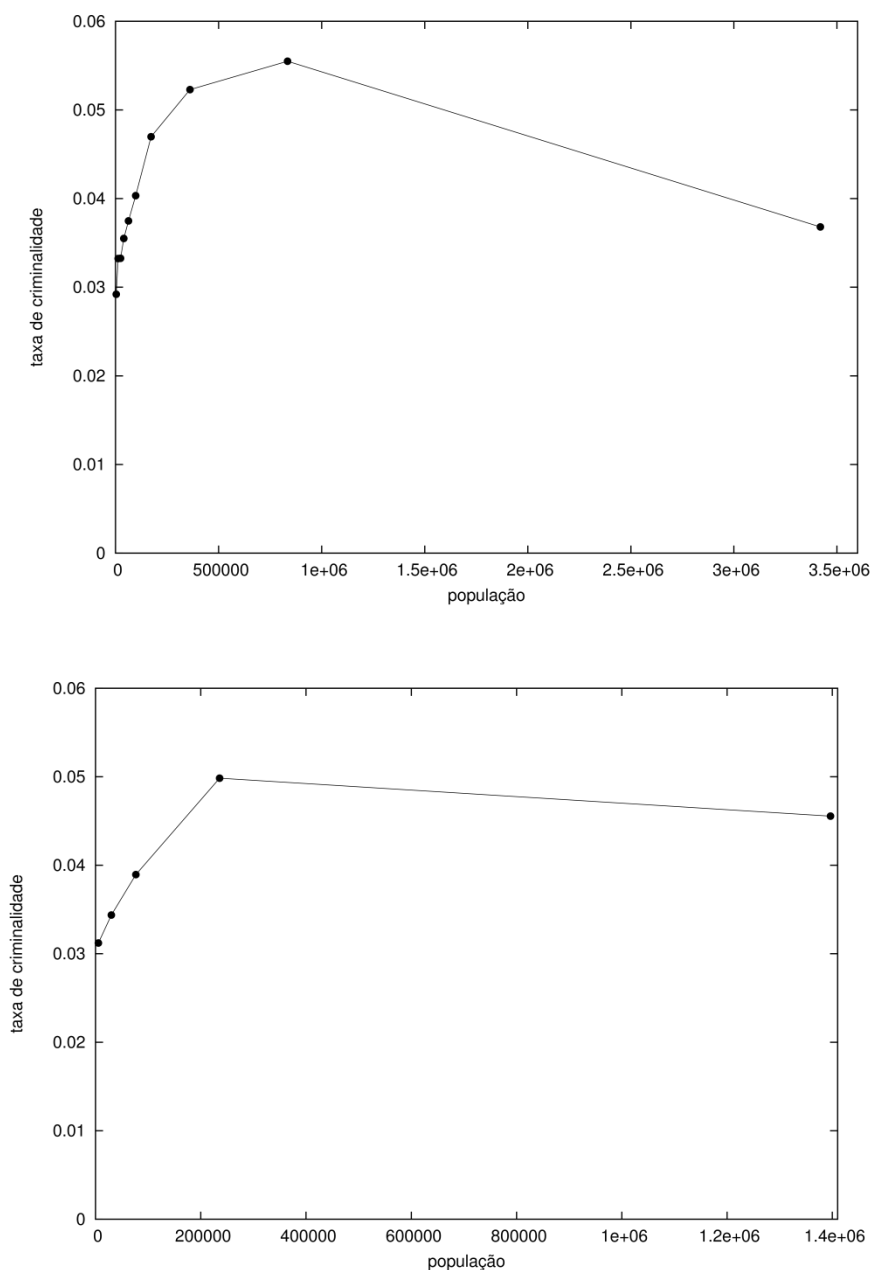
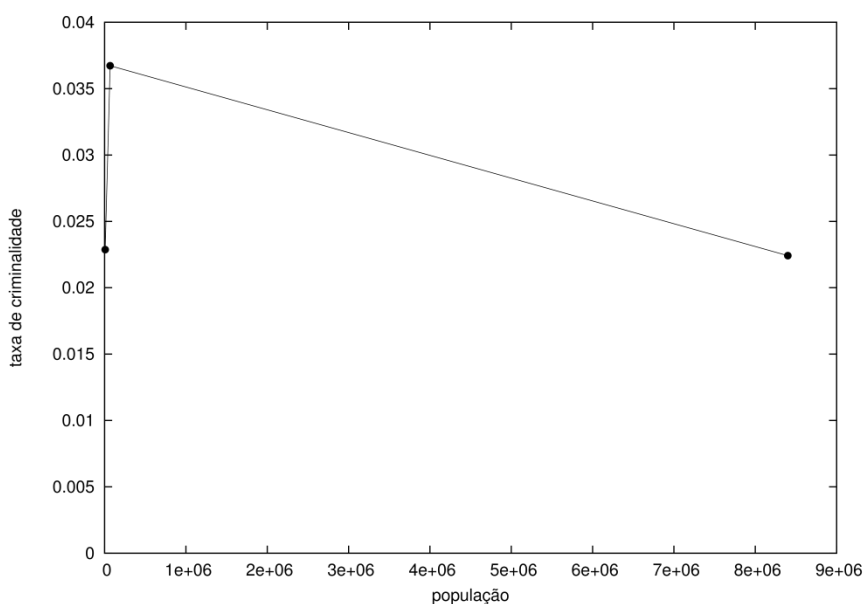


Figura 40. Representação da taxa de criminalidade, como definida pela Equação (1), em função da média populacional para o aglomerado considerado. Esta representação utiliza, adicionalmente, $n = 10$ no painel (a) e $n = 5$ no painel (b).

Notemos, em primeiro lugar, que o número de aglomerados não parece influenciar a forma geral da curva embora para um menor número de habitantes seja visível uma quebra no painel (a). A curva inicia-se, com $n = 5$, em cerca de 0.03 crimes por habitante, para as povoações mais pequenas, atinge um máximo de 0.05 para uma população média de 200 000 habitantes, após o que regista algum decréscimo terminando em cerca de 0.045. Verificamos, assim, uma tendência ascendente inicial que é substituída por uma diminuição quando se consideram centros populacionais de grandes dimensões.

Passemos, de seguida, a uma análise estado a estado que nos permita verificar se esta tendência geral é seguida ou se, pelo contrário, existem padrões específicos. Note-se que, no que se segue, consideraremos apenas cinco aglomerados o que parece conferir uma maior estabilidade à descrição. Na Figura 41 representa-se a taxa de criminalidade vs. população para um conjunto de quatro estados, todos eles incluindo centros populacionais de grandes dimensões: *New York, California, Texas e Illinois*.



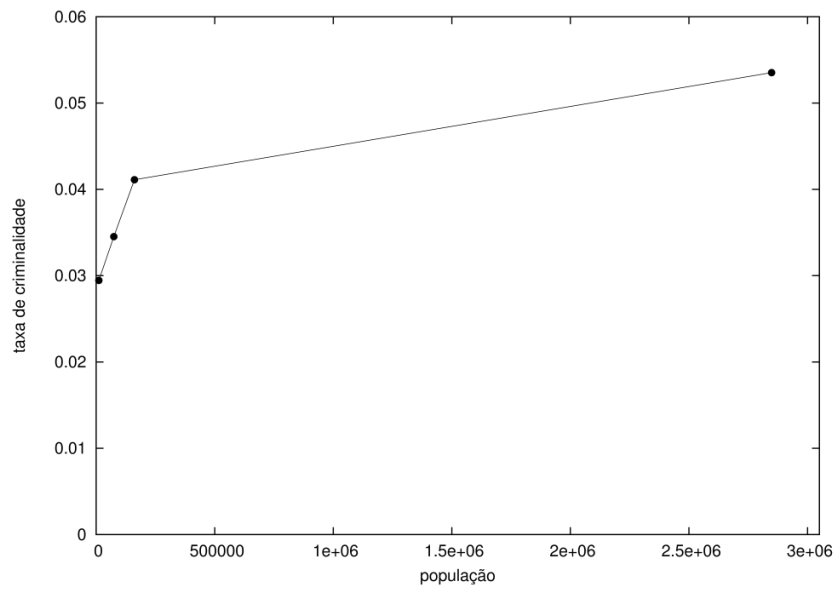
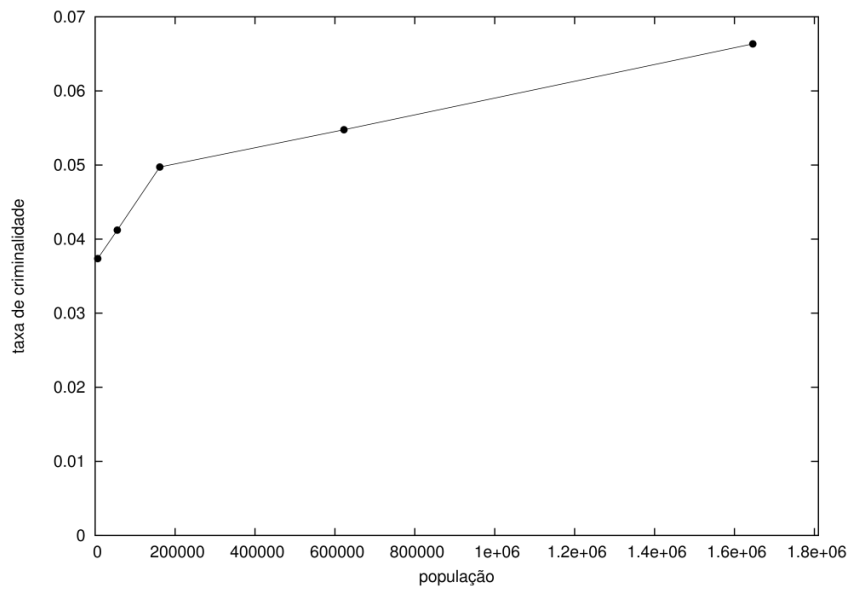
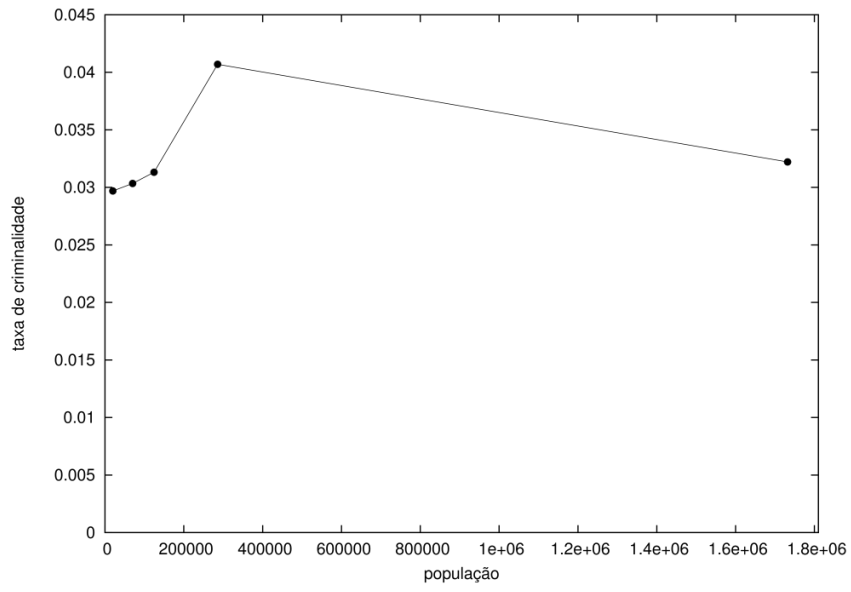


Figura 41. Representação da taxa de criminalidade vs população para um conjunto de quatro estados, todos eles incluindo centros populacionais de grandes dimensões: *New York* no painel (a), *California* no painel (b), *Texas* no painel (c) e *Illinois* no painel (d).

É claramente visível que, para os dois primeiros, a tendência é semelhante à estabelecida a partir da Figura 40, apresentando um máximo. Já para *Texas* e *Illinois*, é identificada uma tendência monotonicamente crescente, com algum grau de estabilização neste último. Também a *Pennsylvania* apresenta centros populacionais de grandes dimensões, verificando-se uma tendência (dados não representados) monotonicamente crescente. Regressando à Figura 40, podemos verificar que a posição do máximo é grandemente afectada pelo número de aglomerados considerado. Apesar disso, podemos afirmar que a inversão na tendência ascendente se dará, tipicamente, acima dos 500 000 habitantes. Tal significa, então, que muitos dos estados se encontrarão apenas na fase ascendente. Outros, contam com uma população de tal modo reduzida que não é possível estabelecer um perfil bem definido. Vamos, então, prosseguir esta análise estado a estado. Passemos então para o outro extremo, isto é, estados com povoações de reduzidas dimensões. Se inspeccionarmos aqueles que, no total, não excedem um milhão de habitantes, verificamos que muitos deles apresentam um perfil irregular. Tal resulta, parcialmente, de uma gama reduzida de valores populacionais, que se situam sempre numa classe geral de agregados pequenos. Entre estes citem-se *Delaware*, *Montana*, *North Dakota*, *Vermont*, *West Virginia* e *Wisconsin*. Note-se que perfis menos definidos podem surgir também quando estão presentes povoações maiores ou estados de maior população total. É o caso de *Arizona*, *Colorado*, *Florida*, *Georgia*, *Iowa*, *Idaho*, *Kansas*, *Maryland*, *North Carolina*, *New Mexico*, *Oklahoma*, *South Carolina* e *Virginia*.

Têm surgido recentemente alguns estudos focando a influência local das tendências nacionais [3]. Nos capítulos anteriores procedemos a uma análise estado a estado, em que o todo foi caracterizado a partir das suas partes. No entanto, a análise feita no presente capítulo permite uma comparação directa entre os resultados globais, veja-se a Figura 40 e os resultados obtidos para cada estado. Note-se que um determinado estado seguir, neste caso, a tendência global, não significa apresentar uma curva análoga à da referida Figura. De facto, muitos dos estados têm populações reduzidas que só lhes permitem, por exemplo, uma descrição da parte ascendente da curva taxa de criminalidade vs. população.

Sumariemos, então, as nossas observações:

- i. Apenas *New York* e *California* apresentam a tendência semelhante à tendência global, iniciando-se num ponto de menor taxa de crime para reduzida população, aumentando em direcção a um máximo que surge por volta do milhão

de habitantes, e com um decréscimo final que mantém, no entanto, a taxa de crime sempre acima do ponto de partida. Outros estados de população total elevada não verificam a tendência global. Citem-se, por exemplo, *Illinois* e *Texas*.

ii. Muitos dos estados pequenos apenas podem contribuir para a descrição da zona inicial, monotonamente crescente, da curva global. De facto, muitos deles apresentam esta tendência crescente, como sejam *Alabama*, *Connecticut*, *Indiana*, *Massachusetts*, *Maine*, *Michigan*, *Minnesota*, *Mississippi*, *Missouri*, *New Hampshire*, *New Jersey*, *Ohio*, *Pennsylvania*, *Rhode Island*, *Texas*, *Utah*, *Washington* e *Wisconsin*, mas muitos outros apresentam perfis irregulares ou de outro tipo. Naturalmente, uma amostra de menores dimensões apresenta-se como menos representativa, o que pode explicar em parte, estes últimos perfis. Esta observação, numa outra forma, foi anteriormente apresentada em termos da evolução temporal de uma determinada povoação. Verificou-se que mesmo descartando agregados com populações inferiores a 100 000 habitantes, a variação anual podia ser muito notória [3]. Naturalmente, tal argumento aplica-se perfeitamente à nossa análise, desta feita no que respeita a enormes flutuações entre povoações com o mesmo número de habitantes.

5.3. A dependência funcional

A enorme dispersão assinalada no início deste capítulo promoveu a construção de agregados de várias povoações para nos permitir analisar a dependência da taxa de criminalidade na população. Verificámos que, em muitos dos casos, o nível de agregação não afecta significativamente a forma geral da curva, embora na análise estado a estado possa promover algum grau de perturbação. Vamos pois basear-nos nos resultados da secção anterior para voltar a desacoplar as diversas povoações, mantendo uma descrição coerente da dependência que procuramos. Os resultados da Figura 24 sugerem, como já amplamente discutido, uma curva inicialmente ascendente, que passa por um máximo e depois decai para valores mais baixos do que os registados inicialmente.

Assim sendo, parece surgir como apropriada a equação para a taxa de criminalidade, T em função da população pop

$$T = T_{\infty} + (T_{\infty} - T_0 + ae^{-b pop}) \quad (29)$$

onde T_{∞} corresponde à taxa de criminalidade assintótica, T_0 à taxa de criminalidade para uma população nula e a e b são parâmetros ajustáveis.

Para a definição dos parâmetros ajustáveis desta função utilizámos todos os dados disponíveis, povoação a povoação, e procedemos a um ajuste global, com base no algoritmo *Levenberg - Marquardt*. Adicionalmente, T_{∞} foi imposta com base no valor da taxa de criminalidade para *New York* e T_0 resultou da média de 5923 povoações, com população abaixo de 10 000 habitantes. Considerando a extrema dispersão registada para baixos valores da população, utilizou-se o valor da população como peso para o ajuste. Concretizando, obtivemos $T_{\infty} = 0.022$, $T_0 = 0.029$, $a = 0.0075 \pm 9 \times 10^{-5}$ e $b = 0.081 \pm 6 \times 10^{-4}$.

Podemos observar a curva ajustada, representada na Figura 42, em conjunto com os pontos que deram origem ao ajuste.

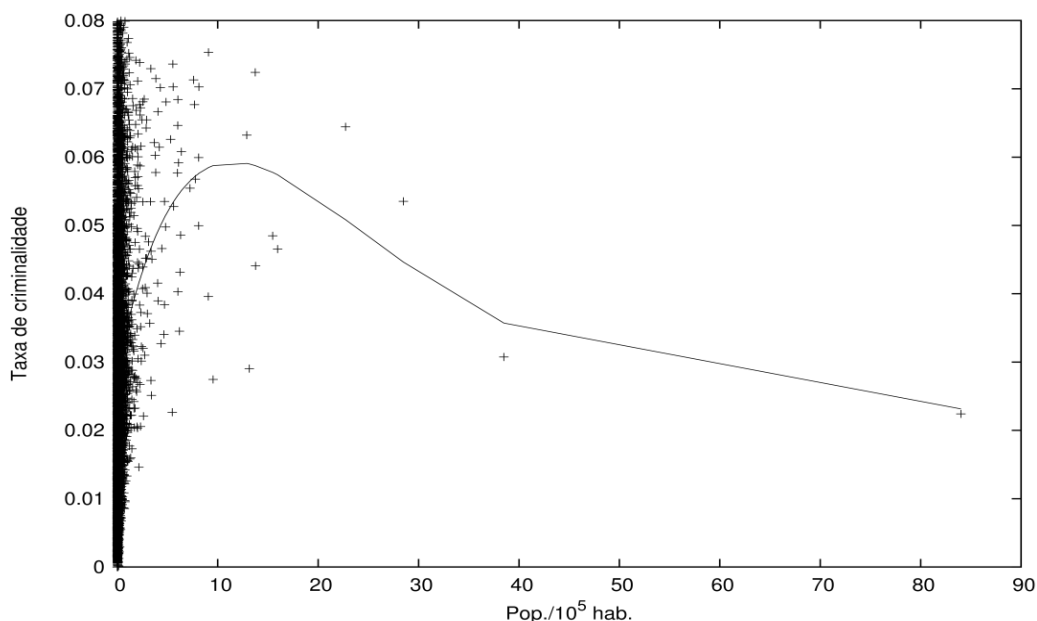


Figura 42. Representação da curva ajustada recorrendo a um ajuste global, com base no algoritmo *Levenberg - Marquardt* da taxa de criminalidade vs população para todos os dados disponíveis, povoação a povoação.

No entanto, e regressando à Figura 39 verificamos que, pelo menos para o conjunto das povoações até 10 000 habitantes, a distribuição parece afastar-se da normalidade. Apesar disso, consideramos que o ajuste por mínimos quadrados não lineares é, ainda, adequado. Com efeito o facto de terem sido utilizados todos os pontos disponíveis faz com que o ajuste seja automaticamente pesado com a distribuição adequada, com uma actuação em conjunto com o peso explicitamente imposto.

5.4. Discussão

Registámos, a partir dos resultados que nos dão a dependência da taxa de criminalidade na população, um conjunto de observações que agora sumariamos. Em primeiro lugar, esta dependência passa por um máximo e decai para maiores valores da população. Este comportamento é compatível com o descrito na referência [103], em que, para povoações com população superior a 250 000 habitantes existe uma tendência geral um decaimento da taxa de criminalidade com a população. No entanto, na nossa função explícita, a parte descendente da curva só se inicia após o milhão de habitantes. Note-se que a observação do presente trabalho e a do artigo citado não são contraditórias, já que este último incide sobre um grupo específico, não pretendendo localizar o ponto exacto de viragem.

Outro aspecto relevante tem a ver com a relação entre o padrão nacional e o padrão local. Este assunto foi abordado na literatura, tendo-se concluído que o padrão nacional é visível a nível das áreas urbanas de maiores dimensões [3]. Também aqui os nossos resultados são, genericamente, consistentes com os da literatura. A variação não monotónica da taxa de criminalidade em função da população é claramente visível nos estados *New York* e *California*. Está, no entanto, ausente em *Illinois*, apesar de neste estado existir o terceiro maior centro urbano dos EUA. Está também ausente no populoso estado do *Texas*. Vários estados, com populações variáveis, e que podem ou não englobar um centro urbano de grandes dimensões apresentam tendências monotonicamente crescentes. Por outras palavras, o padrão local frequentemente não reflecte o comportamento global do país.

A literatura tem também levantado o problema de que o comportamento espacial é incoerente com a evolução temporal [119]. Pensamos que os nossos resultados podem dar alguma contribuição para entender esta discrepância. Por um lado, um dos pressupostos base invocados na literatura tem a ver com uma tendência monotonicamente crescente, mas este pressuposto claramente não se verifica para os grandes centros urbanos, isto é, não podemos esperar um crescimento da taxa de criminalidade com a população que não é visível na perspectiva de *cross section*. Por outro lado, e para as menores povoações o grau de dispersão obscurece a maior parte das observações. Finalmente, quando os centros urbanos possuem populações da ordem do milhão de habitantes estamos situados numa zona de transição que inviabiliza o estabelecimento de qualquer tendência. Naturalmente, numa outra perspectiva, torna-se extremamente difícil fazer previsões já que uma variação na população pode conduzir a evoluções diversas na taxa de criminalidade.

Capítulo 6

Comentários finais

Este trabalho pretendeu realizar uma caracterização global de alguma da informação recolhida no programa UCR do FBI. Debruçámo-nos sobre vários conjuntos de variáveis descritivas (29, 27 e 16 crimes respectivamente) e três perfis diferentes (o número de crimes, a taxa de criminalidade e a fracção de crimes). Tornou-se patente que o tipo de perfil escolhido para a caracterização é determinante no tipo de resultados obtidos.

Quando utilizamos como perfil o número total de crimes encontramos uma distribuição de estados aparentemente governada por factores geográficos. Com a taxa de criminalidade a situação altera-se, já que o tipo ou tipos de crimes dominantes na caracterização evolui. Com efeito, este perfil, ao contrário do anterior, não assenta numa variação global em que o tipo de crime é irrelevante. Finalmente, quando é considerada a fracção de cada crime, temos de ter em atenção que tratando-se de uma fracção, o aumento da relevância de um ou mais crimes é sempre feito à custa da perda de importância de outros.

Globalmente, verificou-se que muita da criminalidade, caracterizadora dos estados, está associada a *drug abuse violations* ou a infracções relacionadas com o álcool. Também o *larceny-theft surge* com alguma relevância.

A representação da taxa de criminalidade em função da população fornece-nos a oportunidade de visualizarmos o perfil que pode ser encontrado a nível nacional. De uma forma geral, a dependência da taxa de criminalidade com a população passa por um máximo e decai para maiores valores da população. Observação que é descrita pela função que sugerimos nesta dissertação.

Outro aspecto relevante tem a ver com a relação entre o padrão nacional e o padrão local. De facto, a variação não monotónica da taxa de criminalidade em função da população é claramente visível nos estados *New York* e *California*. No entanto vários outros estados, com populações variáveis, e que podem ou não englobar um centro urbano de grandes dimensões apresentam tendências monotonicamente crescentes. Por outras palavras, o padrão local raras vezes reflecte o comportamento global do país.

Tanto a dependência estabelecida a nível nacional como as dependências estaduais ajudaram-nos a racionalizar a discrepância frequentemente encontrada entre a evolução espacial e a temporal. Numa outra perspectiva, torna-se extremamente difícil fazer previsões já que uma variação na população pode conduzir a evoluções diversas na taxa de criminalidade.

Referências bibliográficas

1. McCollister, K.E., M.T. French, and H. Fang, *The cost of crime to society: New crime-specific estimates for policy and program evaluation*. Drug and Alcohol Dependence, 2010. 108(1-2): p. 98-109.
2. Rosenthal, S.S. and A. Ross, *Violent crime, entrepreneurship, and cities*. Journal of Urban Economics, 2010. 67(1): p. 135-149.
3. McDowall, D. and C. Loftin, *Do US City Crime Rates Follow a National Trend? The Influence of Nationwide Conditions on Local Crime Patterns*. Journal of Quantitative Criminology, 2009. 25(3): p. 307-324.
4. McGarrell, E., et al., *Project Safe Neighborhoods and Violent Crime Trends in US Cities: Assessing Violent Crime Impact*. Journal of Quantitative Criminology, 2010. 26(2): p. 165-190.
5. Martin A, A., *Testing for similarity in area-based spatial patterns: A nonparametric Monte Carlo approach*. Applied Geography, 2009. 29(3): p. 333-345.
6. Ye, X. and L. Wu, *Analyzing the dynamics of homicide patterns in Chicago: ESDA and spatial panel approaches*. Applied Geography, 2011. 31(2): p. 800-807.
7. Braga, A. and D. Weisburd, *Editors' Introduction: Empirical Evidence on the Relevance of Place in Criminology*. Journal of Quantitative Criminology, 2010. 26(1): p. 1-6.
8. Groff, E., D. Weisburd, and S.-M. Yang, *Is it Important to Examine Crime Trends at a Local "Micro" Level?: A Longitudinal Analysis of Street to Street Variability in Crime Trajectories*. Journal of Quantitative Criminology, 2010. 26(1): p. 7-32.
9. Bernasco, W., *Modeling Micro-Level Crime Location Choice: Application of the Discrete Choice Framework to Crime at Places*. Journal of Quantitative Criminology, 2010. 26(1): p. 113-138.
10. Lu, Y. and X. Chen, *On the false alarm of planar K-function when analyzing urban crime distributed along streets*. Social Science Research, 2007. 36(2): p. 611-632.
11. Cherry, T.L. and J.A. List, *Aggregation bias in the economic model of crime*. Economics Letters, 2002. 75(1): p. 81-86.
12. Lin, W.H. and R. Dembo, *An integrated model of juvenile drug use: A cross-demographic groups study*. West. Criminol. Rev. Western Criminology Review, 2008. 9(2): p. 33-51.
13. Stylianou, S., *Victimless deviance: Toward a classification of opposition justifications*. West. Criminol. Rev. Western Criminology Review, 2010. 11(2): p. 43-56.

14. Keith, H., *Extreme spatial variations in crime density in Baltimore County, MD*. *Geoforum*, 2006. 37(3): p. 404-416.
15. John R, H., *Micro-structure in micro-neighborhoods: A new social distance measure, and its effect on individual and aggregated perceptions of crime and disorder*. *Social Networks*, 2010. 32(2): p. 148-159.
16. Hipp, J., G. Tita, and L. Boggess, *A New Twist on an Old Approach: A Random-Interaction Approach for Estimating Rates of Inter-Group Interaction*. *Journal of Quantitative Criminology*, 2011. 27(1): p. 27-51.
17. Julie A, P., *Explaining discrepant findings in cross-sectional and longitudinal analyses: An application to U.S. homicide rates*. *Social Science Research*, 2006. 35(4): p. 948-974.
18. John L, W., *Racial composition, unemployment, and crime: Dealing with inconsistencies in panel designs*. *Social Science Research*, 2008. 37(3): p. 787-800.
19. I. Y. Sun, R.T., R. R. Gainey, *Neighborhood Characteristics and Crime: A Test of Sampson and Groves' Model of Social Disorganization*. *Western Criminology Review*, 2004. 5((1)): p. 1-16.
20. Oh, J.-H., *Social disorganizations and crime rates in U.S. central cities: Toward an explanation of urban economic change*. *The Social Science Journal*, 2005. 42(4): p. 569-582.
21. Kawachi, I., B.P. Kennedy, and R.G. Wilkinson, *Crime: social disorganization and relative deprivation*. *Social Science & Medicine*, 1999. 48(6): p. 719-731.
22. Jones, E.T., *Crime change patterns in American cities*. *Journal of Criminal Justice*, 1976. 4(4): p. 333-340.
23. Brunsdon, C., J. Corcoran, and G. Higgs, *Visualising space and time in crime patterns: A comparison of methods*. *Computers, Environment and Urban Systems*, 2007. 31(1): p. 52-75.
24. Oatley, G.C. and B.W. Ewart, *Crimes analysis software: 'pins in maps', clustering and Bayes net prediction*. *Expert Systems with Applications*, 2003. 25(4): p. 569-588.
25. Xue, Y. and D.E. Brown, *Spatial analysis with preference specification of latent decision makers for criminal event prediction*. *Decision Support Systems*, 2006. 41(3): p. 560-573.
26. Liu, H. and D.E. Brown, *Criminal incident prediction using a point-pattern-based density model*. *International Journal of Forecasting*. 19(4): p. 603-622.
27. Corcoran, J.J., I.D. Wilson, and J.A. Ware, *Predicting the geo-temporal variations of crime and disorder*. *International Journal of Forecasting*. 19(4): p. 623-634.

28. Sparks, C.S., *Violent crime in San Antonio, Texas: An application of spatial epidemiological methods*. *Spat. Spatio-temporal Epidemiol. Spatial and Spatio-temporal Epidemiology*, 2011. 2(4): p. 301-309.
29. Keyvanpour, M.R., M. Javideh, and M.R. Ebrahimi, *Detecting and investigating crime by means of data mining: a general crime matching framework*. *Procedia Computer Science*, 2011. 3: p. 872-880.
30. Broadhurst, R.G. and R.A. Maller, *Estimating the numbers of prison terms in criminal careers from one-step probabilities of recidivism*. *Journal of Quantitative Criminology*, 1991. 7(3): p. 275-290.
31. Blonigen, D.M., *Explaining the relationship between age and crime: Contributions from the developmental literature on personality*. *Clinical Psychology Review*, 2010. 30(1): p. 89-100.
32. Mears, D.P. and S.H. Field, *Closer Look at the Age, Peers, and Delinquency Relationship*. *Western Criminology Review*, 2002. 4(1): p. 20-29.
33. Shichor, D., D.L. Decker, and R.M. O'Brien, *The relationship of criminal victimization, police per capita and population density in twenty-six cities*. *Journal of Criminal Justice*, 1980. 8(5): p. 309-316.
34. Lo, C.C. and H. Zhong, *Linking Crime Rates to Relationship Factors: The Use of Gender-Specific Data*. *Journal of Criminal Justice*, 2006. 34(3): p. 317-329.
35. Celia C, L., *Sociodemographic factors, drug abuse, and other crimes: How they vary among male and female arrestees*. *Journal of Criminal Justice*. 32(5): p. 399-409.
36. Suzanne, A., *Structural correlates of female homicide: A cross-national analysis*. *Journal of Criminal Justice*. 37(6): p. 576-585.
37. Blackwell, B.S. and A.R. Piquero, *On the relationships between gender, power control, self-control, and crime*. *Journal of Criminal Justice*. 33(1): p. 1-17.
38. Vowell, P.R., *A Partial Test of an Integrative Control Model: Neighborhood Context, Social Control, Self-Control, and Youth Violent Behavior*. *Western Criminology Review*, 2007. 8((2)): p. 1-15.
39. Beaver, K.M. and J.P. Wright, *The association between county-level IQ and county-level crime rates*. *Intelligence*. 39(1): p. 22-26.
40. Bartels, J.M., et al., *Correlations between estimates of state IQ and FBI crime statistics*. *Personality and Individual Differences*, 2010. 48(5): p. 579-583.
41. Canter, D., et al., *Predicting Serial Killers' Home Base Using a Decision Support System*. *Journal of Quantitative Criminology*, 2000. 16(4): p. 457-478.

42. Jesse, B., *Does income inequality lead to more crime? A comparison of cross-sectional and time-series analyses of United States counties*. *Economics Letters*, 2007. 96(2): p. 264-268.
43. Scorzafave, L.G. and M.K. Soares, *Income inequality and pecuniary crimes*. *Economics Letters*, 2009. 104(1): p. 40-42.
44. Jongmook, C., *Income inequality and crime in the United States*. *Economics Letters*, 2008. 101(1): p. 31-33.
45. Tang, C.F. and H.H. Lean, *New evidence from the misery index in the crime function*. *Economics Letters*, 2009. 102(2): p. 112-115.
46. Masih, R., *Modelling the dynamic interactions among crime, deterrence and socio-economic variables: evidence from a vector error-correction model*, in *Selected papers of the MSSA/IMACS 10th biennial conference on Modelling and simulation*. 1995, Elsevier Science Publishers B. V.: Perth, Australia. p. 411-416.
47. Grubestic, T.H., E.A. Mack, and M.T. Kaylen, *Comparative modeling approaches for understanding urban violence*. *Social Science Research*, 2012. 41(1): p. 92-109.
48. McCall, P.L., K.F. Parker, and J.M. MacDonald, *The dynamic relationship between homicide rates and social, economic, and political factors from 1970 to 2000*. *Social Science Research*, 2008. 37(3): p. 721-735.
49. Land, K.C., K.F. Parker, and P.L. McCall, *Heterogeneity in the rise and decline of city-level homicide rates 1976-2005: A latent trajectory analysis*. *Social Science Research*, 2011. 40: p. 363-378.
50. Worrall, J.L., et al., *An analysis of the relationship between probation caseloads and property crime rates in California counties*. *Journal of Criminal Justice*. 32(3): p. 231-241.
51. Boggess, L.N. and J.R. Hipp, *Violent Crime, Residential Instability and Mobility: Does the Relationship Differ in Minority Neighborhoods?* *Journal of Quantitative Criminology*, 2010. 26(3).
52. Sara, M., *Alcohol, Drugs and Violent Crime*. *International Review of Law and Economics*, 2005. 25(1): p. 20-44.
53. Zimmerman, P.R. and B.L. Benson, *Alcohol and rape: An "economics-of-crime" perspective*. *International Review of Law and Economics*, 2007. 27(4): p. 442-473.
54. Nigel, B., *Evolutionary social science: A new approach to violent crime*. *Aggression and Violent Behavior*. 13(3): p. 237-250.
55. Rick, N., *Understanding international crime trends: The legacy of preschool lead exposure*. *Environmental Research*, 2007. 104(3): p. 315-336.

56. Rotton, J. and E.G. Cohn, *Global Warming and U.S. Crime Rates*. Environment and Behavior, 2003. 35(6): p. 802-825.
57. Yearwood, D.L. and G. Koinis, *Revisiting property crime and economic conditions: An exploratory study to identify predictive indicators beyond unemployment rates*. The Social Science Journal, 2011. 48(1): p. 145-158.
58. United States Department of Justice, F.B.o.I., *Crime in the United States, 2009*. September 2010, <http://www.fbi.gov/ucr/09cius.htm>.
59. Steven A, S., *The Federal Bureau of Investigation: Its history, organization, functions and publications*. Government Publications Review (1973), 1979. 6(3): p. 213-239.
60. Chilton, R. and J. Jarvis, *Victims and Offenders in Two Crime Statistics Programs: A Comparison of the National Incident-Based Reporting System (NIBRS) and the National Crime Victimization Survey (NCVS)*. Journal of Quantitative Criminology, 1999. 15(2): p. 193-205.
61. Skogan, W.G., *Measurement problems in official and survey crime rates*. Journal of Criminal Justice, 1975. 3(1): p. 17-31.
62. Maltz, M.D., *Crime statistics: A mathematical perspective*. Journal of Criminal Justice, 1975. 3(3): p. 177-194.
63. Bursik Jr, R.J. and H.G. Grasmick, *The use of multiple indicators to estimate crime trends in American cities*. Journal of Criminal Justice, 1993. 21(5): p. 509-516.
64. O'Brien, R.M., *Comparing detrended UCR and NCS crime rates over time: 1973-1986*. Journal of Criminal Justice, 1990. 18(3): p. 229-238.
65. O'Brien, R.M., *Detrended UCR and NCS crime rates: Their utility and meaning*. Journal of Criminal Justice, 1991. 19(6): p. 569-574.
66. Maltz, M.D. and J. Targonski, *A Note on the Use of County-Level UCR Data*. Journal of Quantitative Criminology, 2002. 18(3): p. 297-318.
67. Lott, J.R. and J. Whitley, *Measurement Error in County-Level UCR Data*. Journal of Quantitative Criminology, 2003. 19(2): p. 185-198.
68. Esbensen, K.H., et al., *Multivariate data analysis: in practice : an introduction to multivariate data analysis and experimental design*. 2002: Camo.
69. Tan, P.N., M. Steinbach, and V. Kumar, *Introduction to data mining*. 2006: Pearson Addison Wesley.
70. Breerton, R.G., *Chemometrics: data analysis for the laboratory and chemical plant*. 2003: Wiley.

71. Steven D, B., *Has the chemometrics revolution ended? Some views on the past, present and future of chemometrics*. *Chemometrics and Intelligent Laboratory Systems*, 1995. 30(1): p. 49-58.
72. Svante, W., *Chemometrics; what do we mean with it, and what do we want from it?* *Chemometrics and Intelligent Laboratory Systems*, 1995. 30(1): p. 109-115.
73. Gemperline, P., *Practical guide to chemometrics*. 2006: CRC/Taylor & Francis.
74. Mutihac, L. and R. Mutihac, *Mining in chemometrics*. *Analytica Chimica Acta*, 2008. 612(1): p. 1-18.
75. Philip K, H., *The evolution of chemometrics*. *Analytica Chimica Acta*, 2003. 500(1-2): p. 365-377.
76. Martens, H. and M. Martens, *Multivariate analysis of quality: an introduction*. 2001: Wiley.
77. Liu, S., S. Kokot, and G. Will, *Photochemistry and chemometrics—An overview*. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, 2009. 10(4): p. 159-172.
78. Roggo, Y., et al., *A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies*. *Journal of Pharmaceutical and Biomedical Analysis*, 2007. 44(3): p. 683-700.
79. Ni, Y. and S. Kokot, *Does chemometrics enhance the performance of electroanalysis?* *Analytica Chimica Acta*, 2008. 626(2): p. 130-146.
80. Piacenti da Silva, M., et al., *Discriminant analysis of trace elements in normal, benign and malignant breast tissues measured by total reflection X-ray fluorescence*. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2009. 64(6): p. 587-592.
81. Madsen, R., T. Lundstedt, and J. Trygg, *Chemometrics in metabolomics—A review in human disease diagnosis*. *Analytica Chimica Acta*, 2010. 659(1-2): p. 23-33.
82. Romualdi, C., et al., *Pattern recognition in gene expression profiling using DNA array: a comparative study of different statistical methods applied to cancer classification*. *Human Molecular Genetics*, 2003. 12(8): p. 823-836.
83. Engreitz, J.M., et al., *Independent component analysis: Mining microarray data for fundamental human gene expression modules*. *Journal of Biomedical Informatics*, 2010. 43(6): p. 932-944.
84. Giacomino, A., et al., *The role of chemometrics in single and sequential extraction assays: A Review. Part II. Cluster analysis, multiple linear regression, mixture resolution, experimental design and other techniques*. *Analytica Chimica Acta*, 2011. 688(2): p. 122-139.

85. Jain, A.K., M.N. Murty, and P.J. Flynn, *Data clustering: a review*. ACM Comput. Surv., 1999. 31(3): p. 264-323.
86. Kaufman, L. and P.J. Rousseeuw, *Introduction*, in *Finding Groups in Data*. 2008, John Wiley & Sons, Inc. p. 1-67.
87. Campanella, L., G. De Angelis, and G. Visco, *Chemometric investigation of the efficiency of different TiO₂ based catalysts as principal components of TOC photochemical sensors under development*. Analytical and Bioanalytical Chemistry, 2003. 376(4): p. 467-475.
88. Kokot, S., et al., *Data Interpretation by some Common Chemometrics Methods*. Electroanalysis, 1998. 10(16): p. 1081-1088.
89. Davies, A.M.C. and T. Fearn, *Back to basics: the principles of principal component analysis*. Spectroscopy Europe. Tony Davies Column. April/May, 2005.
90. Sneath, P.H.A. and R.R. Sokal, *Numerical taxonomy: the principles and practice of numerical classification*. 1973: W. H. Freeman.
91. Jolliffe, I.T., *Principal component analysis*. 2002: Springer-Verlag.
92. Brownlee, K.A., *Statistical theory and methodology in science and engineering*. 1960: Wiley.
93. Anton, H. and C. Rorres, *Álgebra linear com aplicações*. 2002, Porto Alegre: Bookman.
94. Callioli, C.A., H.H. Domingues, and R.C.F. Costa, *Álgebra linear e aplicações*. 1983: Atual.
95. Levenberg, K., *A method for the solution of certain problems in least squares*. Quart. Applied Math., 1944. 2: p. 164-168.
96. Marquardt, D.W., *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*. SIAM Journal on Applied Mathematics, 1963. 11(2): p. 431-441.
97. Nocedal, J. and S.J. Wright, *Numerical optimization*. 1999: Springer.
98. Harrison, L. and J. Gfroerer, *The Intersection of Drug Use and Criminal Behavior: Results from the National Household Survey on Drug Abuse*. Crime & Delinquency, 1992. 38(4): p. 422-443.
99. Bennett, T., K. Holloway, and D. Farrington, *The statistical association between drug misuse and crime: A meta-analysis*. Aggression and Violent Behavior. 13(2): p. 107-118.
100. Simpson, M., *The relationship between drug use and crime: a puzzle inside an enigma*. International Journal of Drug Policy, 2003. 14(4): p. 307-319.

101. Paul, H., *Sunday liquor laws and crime*. Journal of Public Economics, 2012. 96(1-2): p. 42-52.
102. McAuliffe, W.E., et al., *Measuring interstate variations in drug problems*. Drug and Alcohol Dependence, 1999. 53(2): p. 125-145.
103. James J, N., III, *Establishing the statistical relationship between population size and UCR crime rate: Its impact and implications*. Journal of Criminal Justice. 32(6): p. 547-555.
104. Ackerman, W.V., *Socioeconomic Correlates of Increasing Crime Rates in Smaller Communities*. The Professional Geographer, 1998. 50(3): p. 372-387.
105. Archer, D. and R. Gartner, *Violence and Crime in Cross-National Perspective*. 1987: Yale University Press.
106. Conklin, J.E., *Criminology*. 1981, New York: Macmillan Publishing Co.
107. GC, O., *Explaining regional and urban variation in crime: a review of research*. In LaFree G (ed) *The nature of crime: continuity and change*. Department of Justice, Office of Justice Programs, National Institute of Justice, Washington DC. , 2000. 1: p. 261-308.
108. Tittle, C.R., *Sanctions and social deviance: the question of deterrence*. 1980: Praeger.
109. Land, K.C., P.L. McCall, and L.E. Cohen, *Structural Covariates of Homicide Rates: Are There Any Invariances Across Time and Social Space?* American Journal of Sociology, 1990. 95(4): p. 922-963.
110. McCall, P.L., C.L. Kenneth, and L.E. Cohen, *Violent criminal behavior: Is there a general and continuing influence of the South?* Social Science Research, 1992. 21(3): p. 286-310.
111. Berman, Y., *Size of Population and Juvenile Delinquency in Cities in Israel*. Criminology, 1973. 11(1): p. 105-114.
112. Krohn, M.D., L. Lanza-Kaduce, and R.L. Akers, *Community Context and Theories of Deviant Behavior: An Examination of Social Learning and Social Bonding Theories*. Sociological Quarterly, 1984. 25(3): p. 353-371.
113. Berman, Y., *Size of Population and Juvenile Delinquency in Cities in Israel*. Criminology, 1973. 11(1): p. 105-114.
114. Macionis, J.J. and V.N. Parrillo, *Cities and urban life*. 2007: Pearson Prentice Hall.
115. Sandim, H.R.Z., et al., *Profile of social disadvantage in the 100 largest cities of the United States, 1980 to 1990/1993*. Cities, 1998. 15(5): p. 317-326.
116. Braithwaite, J., *Crime, shame, and reintegration*. 1999: Cambridge University Press.

117. Gottfredson, M.R. and T. Hirschi, *A general theory of crime*. 1990: Stanford University Press.
118. Siegel, L.J., *Criminology*. 2003: Wadsworth/Thomson Learning.
119. Rotolo, T. and C. Tittle, *Population Size, Change, and Crime in U.S. Cities*. *Journal of Quantitative Criminology*, 2006. 22(4): p. 341-367.
120. Chamlin, M., J. Cochran, *An Excursus on the Population Size-Crime Relationship*. *Western Criminology Review*, 2004. 5((2)): p. 119-130.

Anexo

Definição das ofensas segundo o programa do *Uniform Crime Reporting*

As infracções designadas como de parte I são as que a seguir se indicam.

Criminal homicide

- a) *Murder and nonnegligent manslaughter*: Consiste no assassinato intencional de um ser humano por outro. Estão incluídas as mortes causadas por negligência, tentativas de homicídio e assaltos com o objectivo de matar. Suicídios e mortes acidentais são excluídos. O programa classifica os homicídios justificáveis de forma separada e limita a sua definição a (1) a morte de um criminoso por um polícia no cumprimento do dever ou (2) a morte de um criminoso, durante a prática de um crime, por um cidadão.
- b) *Manslaughter by negligence*: é definido como a morte de outra pessoa por negligência grave. A morte de pessoas devido à sua própria negligência, mortes acidentais não resultantes de negligência grosseira e fatalidades originadas pela condução não se incluem nesta categoria.

Forcible rape— é definido como o conhecimento carnal de uma mulher à força e contra a sua vontade. Estão incluídos estupros pela força e tentativas ou assaltos para violações, independentemente da idade da vítima. Situações de Infracções legais (não recorrendo a força – vítima menor de idade) são excluídas.

Robbery— consiste em tomar ou tentar tirar algo de valor, que esteja ao cuidado ou no controle de uma pessoa ou pessoas, pela força, ameaça de força ou violência e /ou introduzindo medo na vítima.

Aggravated assault— é definido como um ataque ilegal, de uma pessoa sobre outra, com o objectivo de infligir lesões corporais graves ou agravadas. Este tipo de ataque é normalmente acompanhado pelo uso de uma arma ou por meios susceptíveis de produzir morte ou uma grande lesão corporal. As agressões simples encontram-se excluídas.

Burglary (breaking or entering)— consiste na entrada ilegal numa estrutura para cometer um crime ou um roubo. A tentativa de entrada forçada encontra-se incluída.

Larceny-theft (except motor vehicle theft)— é definido como a toma ilegal e o transporte de um bem que se encontra na posse de outrem. Exemplos deste tipo de ofensa são: roubos de bicicleta, peças e acessórios de veículos automóveis, furtos ou o roubo de qualquer artigo ou bem que não é conduzido pela acção da força, violência ou fraude. As

tentativas de roubo estão incluídas. Os desfalques, jogos de confiança, falsificações, fraudes, cheques, etc. não estão incluídos nesta categoria.

Motor vehicle theft— consiste no roubo ou tentativa de roubo de um veículo motorizado. Um veículo motorizado é por definição um que é auto movível e usa estradas e não carris para se locomover. As lanchas, equipamentos de construção, aviões e equipamentos agrícolas estão excluídos desta categoria.

Arson— é definido como qualquer queima ou tentativa de queima, intencional ou mal intencionada, com ou sem intenção fraudulenta, de bens pessoais ou de outra pessoa tais como uma casa de habitação, edifícios públicos, veículos automóveis, aviões, etc

As infracções Parte II, para as quais apenas são recolhidos os dados das detenções, discriminam-se de seguida.

Other assaults (simple)— assaltos ou tentativas de assalto onde não foi usada qualquer arma e a vítima não sofreu uma lesão grave ou experimentou qualquer lesão. Estão incluídos comportamentos de perseguição, intimidação, coerção e rituais iniciáticos.

Forgery and counterfeiting— consiste na alteração, cópia ou imitação de algo, sem autoridade ou direito, com a intenção de enganar ou defraudar, fazendo passar a cópia ou o item alterado pelo que é original ou único. Também engloba a venda, a compra ou a posse de algo que tenha sido copiado ou imitado com a intenção de enganar ou defraudar. Inclui a forma tentada.

Fraud— definida como a perversão intencional da verdade com o propósito de induzir outra pessoa ou entidade, tendo como base a confiança gerada, a entregar algo de valor ou a conceder um direito legal. Consiste na conversão fraudulenta e obtenção de dinheiro sob falsos pretextos. Os jogos de confiança e os cheques sem fundos, excepto falsificações, estão incluídos nesta categoria.

Embezzlement— consiste na apropriação ilegal ou uso inadequado, por parte de um agressor, para o seu uso próprio de dinheiro, propriedade, ou qualquer outra coisa de valor confiado ao seu cuidado, custódia ou controlo.

Stolen property: buying, receiving, possessing— compra, receptação, posse, venda, ocultação, ou transporte de qualquer propriedade com o conhecimento de que esta foi tomada ilegalmente quer por roubo, fraude, peculato, furto, etc. Inclui a forma tentada.

Vandalism— destruir ou danificar, de forma intencional ou maliciosa, qualquer propriedade pública ou privada, real ou pessoal, sem o consentimento do proprietário ou pessoa que

tenha o controlo ou a custódia, através de acções que impliquem cortar, rasgar, quebrar, marcar, pintar, desenhar, sujar, ou qualquer outra acção especificada na lei local. Também se incluem tentativas.

Weapons: carrying, possessing, etc.— violação de leis ou decretos que proíbam o fabrico, compra, venda, transporte, posse, ocultação ou uso de armas de fogo, instrumentos cortantes, explosivos, dispositivos incendiários ou outras armas mortais. Inclui a forma tentada.

Prostitution and commercialized vice— consiste na promoção ilegal ou participação em actividades sexuais com fins lucrativos, incluindo as tentativas. Procurar clientes ou transporte de pessoas para fins de prostituição; possuir, administrar ou gerir uma habitação ou outro estabelecimento com a finalidade de proporcionar um lugar onde a prostituição é realizada; ou de qualquer outra forma assistir ou promover a prostituição.

Sex offenses (except forcible rape, prostitution, and commercialized vice)— crimes contra a castidade e decência moral. O incesto e o atentado ao pudor fazem parte desta categoria. Também se incluem tentativas.

Drug abuse violations— Violação das leis que proíbem a produção, distribuição e/ou uso de determinadas substâncias controladas. Incluem o cultivo ilegal, o fabrico, a distribuição, a venda, o uso, a posse, o transporte ou importação de qualquer droga ou narcótico. Contabilizam as detenções por violações de leis locais e estaduais, especificamente as relacionadas com a posse ilegal, venda, uso, cultivo, produção e confecção de estupefacientes. As categorias de drogas a controlar são as seguintes: ópio ou cocaína e seus derivados (morfina, heroína, codeína); maconha; narcóticos sintéticos (narcóticos fabricados que podem causar dependência real – demerol, metadona) e medicamentos não narcóticos perigosos (barbitúricos, benzedrina)

Gambling— definido como a aposta ilegal recorrendo ao uso de dinheiro ou qualquer outra coisa de valor. Inclui: ajudar, promover ou explorar um jogo de azar ou algum outro jogo; possuir ou transmitir informações sobre apostas; fabricar, vender, comprar ou possuir qualquer equipamento, dispositivos ou mercadorias de jogo; adulterar o resultado de um evento desportivo ou competição para ficar com vantagem no jogo.

Offenses against the family and children— consistem em actos ilícitos, não violentos, realizados por um membro da família ou por um responsável legal, que ameaçam o bem-estar físico, mental, económico ou moral de um outro membro da família e que não são classificados como outros crimes tais como assalto e ofensas sexuais. Também se incluem tentativas.

Driving under the influence— conduzir, operar um veículo motorizado ou um veículo de transporte comum enquanto se encontrar incapaz física e mentalmente como resultado do consumo de bebidas alcoólicas ou pelo uso de drogas e estupefacientes.

Liquor laws— violação das leis ou decretos estaduais que proibam a produção, a venda, a compra, o transporte, a posse ou o uso de bebidas alcoólicas. Não inclui a condução sob o efeito do álcool. As violações federais também estão excluídas.

Drunkness— consiste na ingestão de bebidas alcoólicas até ao ponto das faculdades mentais e da coordenação motora ficarem substancialmente prejudicadas. Não inclui a condução sob o efeito do álcool.

Disorderly conduct— Qualquer comportamento que tenda a perturbar a ordem pública ou o decoro, escandalize a comunidade ou choque o senso de moralidade pública.

Vagrancy— consiste na violação de uma ordem judicial, regulamento, portaria ou lei que exijam a remoção de pessoas das ruas ou de outras áreas específicas; proibam as pessoas de permanecerem numa determinada área ou lugar de forma ociosa ou sem rumo; proibam as pessoas de vaguear de lugar para lugar sem terem meios visíveis de subsistência.

All other offenses— Todas as violações de leis estaduais ou locais que não estejam identificadas de forma específica como ofensas de Parte I ou Parte II, excepto as violações ao tráfego.

Suspicion— prisão sem que tenha sido cometido um delito específico e libertação sem que tenha sido apresentada uma acusação formal.

Curfew and loitering laws (persons under age 18)— violações ao toque de recolher e às leis de vadiagem por pessoas menores de 18 anos de idade.

Runaways (persons under age 18)— limitado a jovens que foram colocados sob custódia protectora ao abrigo das disposições dos estatutos locais.